Segmentation simultanée de signaux: application aux données de microarray CGH

Baba Thiam, Stephane Robin

▶ **To cite this version:**

Baba Thiam, Stephane Robin. Segmentation simultanée de signaux: application aux données de microarray CGH. Congrès conjoint de la Société Statistique du Canada et de la Société Française de Statistique, May 2008, Ottawa, Canada. pp.219. ⟨hal-01197547⟩

HAL Id: hal-01197547

https://hal.science/hal-01197547v1

Submitted on 6 Jun 2020

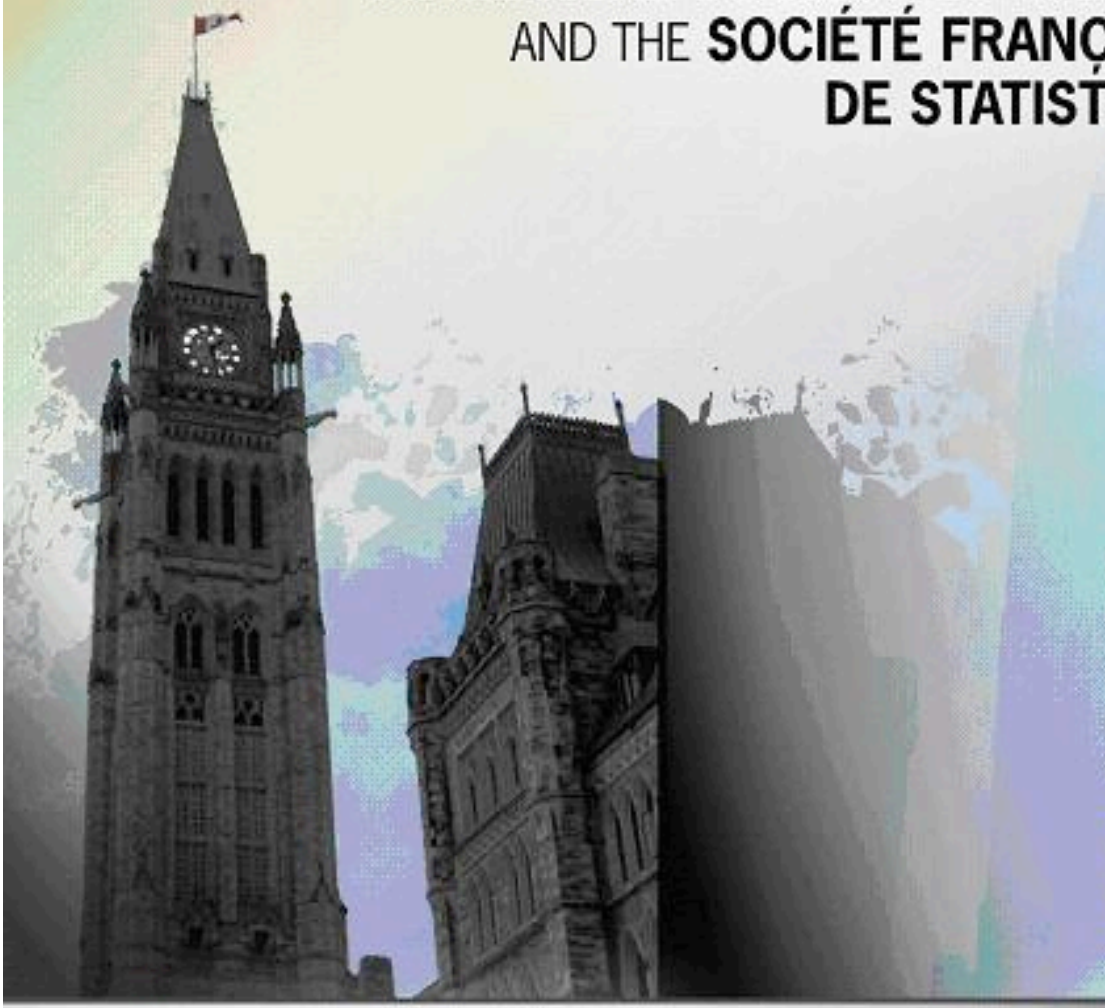
HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

HAL Authorization

CONGRÈS CONJOINT DE LA
SOCIÉTÉ STATISTIQUE DU CANADA
ET DE LA **SOCIÉTÉ FRANÇAISE DE STATISTIQUE**

JOINT MEETING OF THE
STATISTICAL SOCIETY OF CANADA
AND THE **SOCIÉTÉ FRANÇAISE DE STATISTIQUE**



www.ssc.ca
www.sfds.asso.fr

25 AU 29 MAI 2008
CENTRE DES CONGRÈS D'OTTAWA (CANADA)

MAY 25-29 2008
OTTAWA CONGRESS CENTRE (CANADA)

Table of Contents • Table des matières

Bienvenue • Welcome 2

Sponsors • Commanditaires 4

Organizers • Organisateurs 5

Exhibitors • Exposants 5

Language of Presentations • Langue des présentations 6

Simultaneous Translation • Traduction simultanée 6

Translation of Abstracts and Preparation of the Program • Traduction des résumés et préparation du programme 7

On-site Registration • Inscription sur place 7

General Information • Information générale 8

Committees and Meetings • Comités et réunions 13

List of Sessions • Liste des sessions 17

Program • Programmme 24

Abstracts • Résumés 84

Author Index (per session) • Index des auteurs (par session) 288

Author Index • Index des auteurs 298

Bienvenue • Welcome

Chers collègues et amis,

En juillet, il y aura 400 ans qu'une première colonie française permanente voyait le jour au Canada. Pour marquer cet anniversaire, la Société statistique du Canada (SSC) et la Société française de statistique (SFdS) ont voulu tenir ensemble leurs assises annuelles.

Cet audacieux pari se traduit aujourd'hui par un congrès d'une ampleur exceptionnelle, dont le programme scientifique reflète les multiples préoccupations, souvent complémentaires, des statisticiens d'Europe et d'Amérique du Nord. Votre présence à ce congrès contribuera certainement à son succès.

Très tôt, la ville d'Ottawa est apparue aux organisateurs comme le lieu idéal de ce rassemblement historique : bilingue, cosmopolite, raffinée, elle compte plus de statisticiens que nulle autre au pays. Nous espérons que vous prendrez le temps d'explorer la région.

L'organisation d'un événement d'une telle envergure repose sur un grand nombre de bénévoles. Nous leur devons beaucoup, mais plus encore à Bruno Rémillard et Marc Hallin, co-présidents du Comité scientifique, ainsi qu'à Pierre Lavallée, président du Comité d'organisation, et à ses collaborateurs de Statistique Canada.

De part et d'autre de l'Atlantique, nombreux sont ceux qui ont travaillé d'arrache-pied pour harmoniser nos traditions associatives, dont Jean-Michel Marin et Michel Delecroix pour la SFdS, ainsi que Christian Léger et Édouard Chen pour la SSC. Le caractère bilingue du congrès doit aussi beaucoup à Hélène Bérard et à son équipe de traducteurs. Merci à tous !

Maintenant que tout est en place, à vous d'en profiter : amusez-vous, tissez des liens et faites progresser la science. C'est le bonheur que nous vous souhaitons.

Le président de la Société statistique du Canada,
Christian Genest, Ph.D., P.Stat.

Dear colleagues and friends,

Next July will be 400 years since the establishment of the first permanent French settlement in Canada. To commemorate this anniversary, the Statistical Society of Canada (SSC) and the Société française de statistique (SFdS) saw it fit to hold a Joint Annual Meeting.

This daring venture materialises today in a meeting of exceptional breadth, whose scientific program reflects the multiple and often complementary concerns of European and North American statisticians. Your attendance at this meeting will certainly contribute to its success.

Early on, it became apparent to the organisers that the city of Ottawa would be the ideal location for this historical gathering: bilingual, cosmopolitan, sophisticated, and home to more statisticians than any other Canadian city. Please take some time to explore the region.

The organisation of an event of such magnitude depends on a large number of volunteers. We are indebted to all of them but more particularly to Bruno Rémillard and Marc Hallin, Co-Chairs of the Program Committee, as well as Pierre Lavallée, Chair of the Local Arrangements Committee, and his collaborators at Statistics Canada.

On both sides of the Atlantic, many people worked diligently to harmonise the traditions of our associations, including Jean-Michel Marin and Michel Delecroix for the SFdS, as well as Christian Léger and Edward Chen for the SSC. The bilingual character of this meeting also owes much to Hélène Bérard and her team of translators. Thanks to all!

Now that everything is in place, the scene is set for you to enjoy yourself, to meet people, and to advance the science. We hope you have the time of your life.

Christian Genest, Ph.D., P.Stat.
President of the Statistical Society of Canada

Chers collègues,

Pour la première fois de notre histoire, un congrès conjoint est organisé par la Société Française de Statistique et la Société statistique du Canada. Il ne s'agit cependant pas de notre première venue au Canada puisque, déjà en 1996, nous avions organisé nos Journées de Statistique à Québec et qu'il y a trois ans la même ville a accueilli notre Colloque Francophone sur les Sondages. Nous voici donc tous réunis pour cinq jours dans le Centre des congrès de cette ville accueillante qu'est Ottawa, pour deux manifestations scientifiques conjointes : le 36^{e} congrès annuel de la SSC et les 40es Journées de Statistique de la SFdS. Ce congrès, qui regroupe ateliers, séances de communications invitées et séances de communications libres, propose aussi des événements spécifiques à chaque association : pour la SFdS, l'allocution de Davy Paindaveine de l'Université Libre de Bruxelles, invité exceptionnel de la présidence, la conférence « Lucien Le Cam » prononcée cette année par Richard Gill de l'Université de Leiden, et les allocutions de Jean-François Quessy de l'Université du Québec à Trois-Rivières et de Gilles Stoltz de l'École Normale Supérieure à Paris, lauréats du prix « Marie-Jeanne-Laurent-Duhamel » destiné à valoriser la qualité du travail de thèse d'un jeune statisticien francophone. Cette manifestation doit être, grâce à sa dimension internationale, un lieu d'échanges d'expériences et de rencontres privilégié entre chercheurs et utilisateurs de la statistique des milieux académiques, gouvernementaux et industriels des deux continents. Elle vous est entièrement destinée, et vit et tire sa richesse de vos contributions. C'est la raison pour laquelle nous comptons sur votre participation active pour en faire un évènement d'exception. Au nom de la Société Française de Statistique, je vous souhaite la bienvenue pour cette édition originale de nos Journées de Statistique au sein de ce congrès conjoint.

La présidente de la Société Française de Statistique,
Christiane Guinot

Dear Colleagues,

For the first time in our history, a joint meeting is organized by the Société Française de Statistique and the Statistical Society of Canada. However, this is not our first visit to Canada since in 1996 we organized our Journées de Statistique in Québec, and three years ago the same city has welcomed our Conférence Francophone sur les Sondages. We are here all together for five days in the Congress Centre of this welcoming city that is Ottawa, for two joint scientific events: the 36th annual conference of the SSC and the 40th Journées de Statistique of the SFdS. This congress, which includes workshops, invited communications sessions and free papers sessions, also proposes specific events in each association. It includes for SFdS, an address by Davy Paindaveine of the Université Libre de Bruxelles, special invited of the SFdS Presidency, the Conference "Lucien Le Cam" devoted this year to Richard Gill of the University of Leiden, and speeches by Jean-François Quessy of the Université du Québec à Trois-Rivières and Gilles Stoltz of the École Normale Supérieure in Paris, winners of the "Marie-Jeanne-Laurent Duhamel" award. This prize intended to enhance the quality of the doctorate thesis of a young French-speaking statistician. This meeting should be, thanks to its international dimension, a forum for exchanges of experience and exceptional meetings between researchers and users of statistics from academies, governments and industries on both continents. This congress is entirely for you, and lives and derives its richness of your contributions. That is why we wait for your active participation to make it an exceptional event. On behalf of the Société Française de Statistique, I welcome you to this remarkable edition of our Journées de Statistique within this joint meeting.

Christiane Guinot
President of the Société Française de Statistique

Sponsors • Commanditaires

- Ambassade de France
- Canadian Heritage • Patrimoine canadien
- Carleton University
- Centre de recherches mathématiques
- Fields Institute
- Institut des sciences mathématiques
- MITACS
- Pacific Institute for the Mathematical Sciences
- SAS Institute
- Statistics Canada • Statistique Canada
- University of Ottawa • Université d’Ottawa

Organizers • Organisateurs

Local Arrangements Committee • Comité d'organisation

- Pierre Lavallée, Stat. Can. (Chair • président)
- Hélène Bérard, Stat. Can.
- Michel Delecroix, ENSAI
- Diane Galarneau, Stat. Can.
- Carole Jean-Marie, Stat. Can.
- Jean-Michel Marin, INRIA
- Isabelle Michaud, Stat. Can.
- Mike Sirois, Stat. Can.
- Paola Versolato, Stat. Can.
- Carolyn Zirbser, Stat. Can.

Program Committee • Comité scientifique

- Marc Hallin, Université Libre de Bruxelles
- Bruno Rémillard, HÉC Montréal

Exhibitors • Exposants

The exhibitors' booths are located in the West part of the Congress Hall A/E. The exhibitors are listed below.

Les kiosques des exposants sont situés dans la partie ouest du Congress Hall A/E. Les exposants apparaissent ci-dessous.

- Bedford, Freeman
- Chapman and Hall/CRC

- Éditions du renouveau pédagogique
- John Wiley & Sons
- Nelson Education
- Pearson Education
- SAS
- Société Française de Statistique
- Somabec
- Springer
- Statistical Society of Canada • Société statistique du Canada
- Statistical Society of Ottawa • Société statistique d'Ottawa
- Statistics Canada • Statistique Canada

Language of Presentations • Langue des présentations

At the time that they submitted their abstract, speakers were asked to provide the language in which they intend to give their oral presentation as well as the language of their visual aids. Icons are used to provide this information for each paper. For the oral presentation, we have used the icons Ⓔ, Ⓕ, whereas [E], [F] indicate the language of the visual aids. The letter inside identifies the language: E for English and F for French. Please note that the plenary talks are simultaneously translated.

Lorsque les conférenciers ont soumis leur résumé, ils ont spécifié la langue dans laquelle ils comptaient faire leur présentation orale, ainsi que la langue du support visuel. À titre informatif, nous avons inclus cette information à l'aide d'icônes pour chaque présentation. Pour la présentation orale nous avons utilisé les icônes Ⓕ, Ⓔ, alors que [F], [E] indiquent le support visuel. La lettre à l'intérieur identifie la langue : F pour français et E pour anglais (English). Il est important de noter que pour les présentations des séances plénières seront traduites simultanément.

Simultaneous Translation • Traduction simultanée

Simultaneous translation is offered during plenary sessions. Participants that would like to use this facility are asked to get the required earphones at the Registration Desk.

La traduction simultanée est offerte lors des séances plénières. On invite les participants qui désirent utiliser ce service à se procurer les écouteurs à cet effet à la table d'inscription.

Translation of Abstracts and Preparation of the Program • Traduction des résumés et préparation du programme

Translation of all abstracts has been accomplished by Statistics Canada staff under the leadership of Hélène Bérard, Chair of the Bilingualism Committee of the Statistical Society of Canada. Hélène, members of her committee, and Félix Labrecque-Synnot, a student at Université de Montréal, have also participated at this imposing endeavour.

The program with abstracts was produced using a database at the Centre de recherches mathématiques (CRM) under the supervision of Christian Léger. Our special thanks go to Christian and staff at CRM, especially André Montpetit, Daniel Ouimet, and François Cassistat. Printing and preparation of USB keys were accomplished at Statistics Canada.

La traduction de tous les résumés a été faite par du personnel de Statistique Canada sous la gouverne de Hélène Bérard, présidente du comité du bilinguisme de la Société statistique du Canada. Hélène, les membres de son comité ainsi que Félix Labrecque-Synnot, étudiant à l'Université de Montréal, ont également participé à la réalisation de cet imposant travail.

Le programme et les résumés ont été compilés à l'aide d'une base de données du Centre de recherches mathématiques (CRM) sous la supervision de Christian Léger. Nos sincères remerciements à Christian et au personnel du CRM, notamment André Montpetit, Daniel Ouimet et François Cassistat.

On-site Registration • Inscription sur place

Registration will take place in the foyer of the Congress Hall level of the Ottawa Congress Centre. Registration will be from 8:00 to 17:00 on Sunday, May 25, and from 7:30 to 18:00 from Monday, May 26, to Thursday, May 29. The Ottawa Congress Centre is situated at 55 Colonel By Drive, besides the Rideau (shopping) Centre and the Rideau Canal.

L'inscription se fera dans le foyer du niveau Congress Hall du Centre des congrès d'Ottawa. L'inscription sera de 8h00 à 17h00 le dimanche 25 mai, et de 7h30 à 18h00 du lundi 26 mai au jeudi 29 mai. Le Centre des congrès d'Ottawa est situé au 55 de la promenade Colonel By, à côté du Centre (commercial) Rideau et du canal Rideau.

General Information • Information générale

Transportation • Transport

Bus • Autobus

A flyer of OCTranspo is provided in your congress bag.

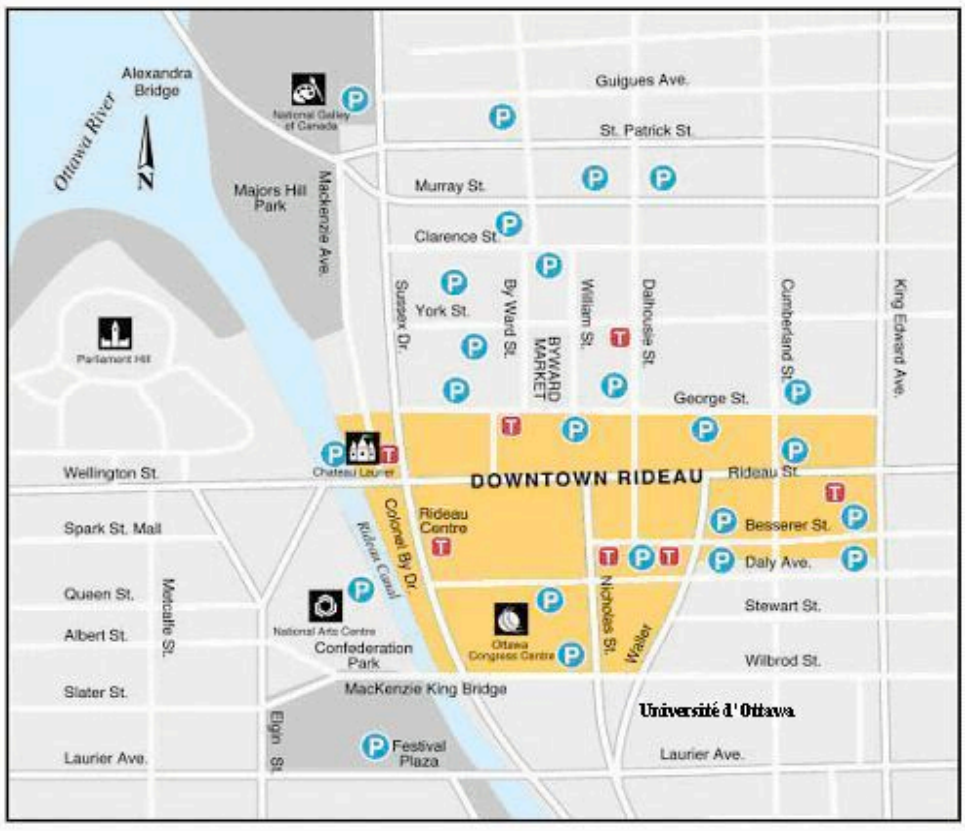
Un dépliant d'OCTranspo est inclus dans le sac du congrès.

Taxi

Blueline 613-238-1111
Capital Taxi 613-744-3333

Parking Close By • Stationnements à proximité

Citipark 613-234-6526
Fare • Tarif 1,25 CAD/$\frac{1}{2}$ h
Passes • Laissez-passer 18 CAD/12 h; 25 CAD/24 h



Rooms • Salles

See the rooms' maps and the Ottawa University's Campus maps on the last pages of the program.

The rooms are situated on the two floors of the Congress Centre. There are the Capital Hall Level (basement) and the Congress Hall Level (1st floor). The registration table, the room for the plenary sessions, the posters sessions, the Internet Access, and the exhibitors will be situated at the Congress Hall Level. The Colonel By Salon (where the welcoming reception will be) and the banquet room are situated at this level. The rooms for the workshops and the ones for the parallel sessions are situated at the Capital Hall Level.

Voir les plans des salles et du campus de l'Université d'Ottawa aux dernières pages du programme.

Les salles sont situées sur les deux étages du Centre des congrès. Il s'agit des niveaux « Capital Hall » (sous-sol) et « Congress Hall » (1er étage). Au niveau Congress Hall, on retrouve la table d'inscription, la salle des séances plénières, les sessions sur affiches, l'accès à Internet, ainsi que les exposants. Le Salon Colonel By (où se tiendra la réception de bienvenue) et la salle de banquet se situent à ce niveau. Au niveau Capital Hall, on retrouve les salles des ateliers et celles des séances en parallèle.

Workshops • Ateliers

The workshops are training activities organized by the SSC Sections. They are held concurrently on Sunday May 25 (no other scientific activities will take place on Sunday). They are targeted at all meeting participants, including students. Registration is necessary and is done while registering for the meeting. The workshops will all take place at the Capital Hall level of the Congress Centre: Regression Modelling Strategies (Biostatistics Section) in 1B from 8:30 to 17:00, Selected Topics in Analysis of Survey Data (Survey Methods Section) in 2B from 9:00 to 16:00, and Seductive Curves and Surfaces: Functional Data Analysis in Action (Business and Industrial Statistics Section) in 3B from 10:00 to 15:30.

Les ateliers sont des activités de formation organisées par les groupes d'intérêt de la SSC. Ils auront lieu concurremment le dimanche 25 mai (aucune autre activité scientifique n'a lieu le dimanche). Ils s'adressent à tous les congressistes, incluant les étudiants. L'inscription est obligatoire et se fait en même temps que pour le congrès. Les ateliers auront tous lieu au niveau Capital Hall du Centre des congrès : Stratégies de modélisation en régression (Groupe de biostatistique) au 1B de 8h30 à 17h00, Sujets choisis en analyse de données d'enquêtes (Groupe de méthodes d'enquêtes) au 2B de 9h00 à 16h00 et Courbes et surfaces séduisantes : l'analyse de données fonctionnelles en action (Groupe de statistique industrielle et de gestion) au 3B de 10h00 à 15h30.

Poster Session • Séance d'affichage

The Poster Sessions are in the same room as the exhibitors and the internet access, i.e., in the West part of the Congress Hall A/E.

This year, we will have three poster sessions, one on each day from Monday through Wednesday. Presenters are expected to be by their poster during the afternoon session, between 14:35 and 16:05. The poster can be set up in the morning and left there until the end of the day at which point it should be removed.

Les séances d'affichage ont lieu dans la même salle que les exposants et l'accès internet, c'est-à-dire dans la partie ouest du Congress Hall A/E.

Cette année, nous aurons trois séances d'affichage, une par jour, du lundi au mercredi. Les présentateurs doivent être devant leur affiche durant la séance de 14h35 à 16h05. Les affiches peuvent être installées le matin pour n'être enlevées qu'à la fin de la journée.

NSERC News Session and Workshop • Séance de nouvelles et atelier offerts par le CRSNG

Monday, May 26 • lundi 26 mai, 18:00–19:00, Ottawa Congress Centre, Capital Hall 6A
Wenesday, May 28 • mercredi 28 mai, 12:45–13:30, Ottawa Congress Centre, Capital Hall 1A

An NSERC news session will take place on Wednesday, May 28 from 12:45 to 13:30 in Capital Hall room 1A. Serge Villemure, Director, Research Grants and Scholarships Directorate at NSERC will make a presentation and lead a discussion on:

- Future Changes to the Discovery Grants Program and Grant Selection Committee Structure: NSERC has commissioned two reviews in order to assess the Discovery Grants Program, including its approach to supporting research in the natural sciences and engineering, as well as the structure and some elements of the operations of the selection committees that review Discovery Grant applications. NSERC would like to discuss the changes envisioned as a result of these reviews with the community prior to preparing the final recommendations on the implementation. Input from all members of the community would be welcomed.
- Latest Federal Government Budget Announcement: Impact on NSERC's budget.
- Discovery Grants Competitions Results 2008: NSERC will present a brief report on the most recent Discovery Grants competition.

Also, there will be an NSERC workshop on Monday from 18:00 to 19:00 in Capital Hall room 6A. Hugh Chipman, Chair of the statistics grant selection committee will lead this workshop. The news session and the workshop are open to all interested researchers. New faculty members are especially welcome!

Une séance d'information du CRSNG aura lieu le mercredi 28 mai, de 12h45 à 13h30 dans la salle Capital Hall 1A. Serge Villemure, le directeur de la Direction des subventions de recherche et bourses du CRSNG fera une présentation et animera une discussion sur :

- Futurs changements au Programme des subventions à la découverte et à la structure des comités de sélection des subventions : le CRSNG a mis sur pied deux comités d'examen pour évaluer son programme de subventions à la découverte, notamment concernant son approche au financement de la recherche dans les sciences naturelles et en génie, ainsi que la structure et quelques éléments opérationnels des comités de sélection qui évaluent les demandes au programme de subventions à la découverte. Le CRSNG souhaiterait discuter avec la communauté les changements envisagés suite aux travaux de ces deux comités avant de préparer les recommandations finales quant à leur mise en œuvre. Les réactions de tous les membres de la communauté sont bienvenu(e)s.
- Dernières annonces budgétaires du Gouvernement fédéral : Impact sur le budget du CRSNG.
- Résultats des compétitions 2008 des subventions à la découverte : Le CRSNG va présenter un bref rapport de la dernière compétition des subventions à la découverte.

Un atelier CRSNG aura également lieu le lundi de 18h00 à 19h00 dans la salle Capital Hall 6A. Hugh Chipman, président du comité de sélection des subventions de recherche en statistiques l'animera. La séance d'information et l'atelier sont ouverts à tous les chercheurs intéressés. Les nouveaux membres du corps professoral sont tout particulièrement les bienvenu(e)s !

Welcome Reception • Réception de bienvenue

The welcoming reception will take place at the Colonel By Salon, which is situated at the Congress Hall level, nearby the registration table and the room for the plenary sessions on Sunday, May 25 from 17:00 to 19:00. A cocktail with appetizers will be served. All conference attendees and companions are warmly invited to come. This event is free.

La réception de bienvenue aura lieu au salon Colonel By qui est situé au niveau Congress Hall, à côté de la table d'inscription et de la salle des séances plénières, le dimanche 25 mai de 17h00 à 19h00. Un coquetel et des amuse-gueules seront servis. Les participants au Congrès et les personnes accompagnantes sont tous chaleureusement invités à venir profiter de la soirée. L'entrée est libre.

Reception at the University of Ottawa • Réception à l'Université d'Ottawa

This reception will be held on Monday, May 26 from 19:00 to 22:00 in the very nice reception room in Tabaret Hall of the University of Ottawa, which is situated close to the Congress Centre. (30 CAD per person, payable activity for all)

Cette réception aura lieu le lundi 26 mai de 19h00 à 22h00 dans la très belle salle de réception du pavillon Tabaret de l'Université d'Ottawa qui est située à deux pas du Centre des congrès. (30 CAD par personne, activité payante pour tous)

Banquet

A sumptuous banquet will take place on Wednesday, May 28 from 19:00 until midnight in Congress Hall of the Congress Centre. After a gourmet meal with soft music in the background, the Big Band Caravane will live things up with music and songs, in French and English (`www.bigbandcaravane.com`). Do not miss it! (Free for registered participants for the meeting who have elected to make a meal choice when they registered, 80 CAD for others).

Un somptueux banquet aura lieu le mercredi 28 mai de 19h00 à minuit au Congress Hall du Centre des congrès. Après un repas gastronomique servi sur fond de musique douce, le Big Band Caravane entrera en piste avec des musiques et chansons jazzées, tant francophones qu'anglophones (`www.bigbandcaravane.com`). À ne pas manquer ! (Gratuit pour les congressistes qui ont fait un choix de repas lorsqu'ils se sont inscrits, 80 CAD pour les autres).

Suggestions for Restaurants near-by • Suggestion de restaurants à proximité

HOTELS	RESTAURANTS
Les Suites Hotel	Presse Café
The ARC Hotel	The ARC Lounge/Restaurant
Quality Hotel	Don Cherry's Sports Grille
Novotel Hotel	Café Nicole (renovating)
Chateau Laurier	Wilfrid's, Lounge Zoé's
Cartier Place	Café Mezzaluna

Other Restaurants • Autres restaurants

Hard Rock Café 73 York Street, 613-241-2442 (American • américain)

Empire Grill 47 Clarence Street, 613-241-1343 (Steak & seafood • steak & fruits de mer)

Courtyard Rest. 21 George Street, 513-241-1516 (Fine dining • fine cuisine)

Mamma Grazzi's Kitchen 25 George Street, 613-241-8656 (Italian • italien)

Métropolitain Brasserie 700 Sussex Drive, 613-562-1160 (French • français)

Oregano's 74 George Street, 613-241-5100 (Italian • italien)

Daly's Westin Hotel, 11 Colonel By Drive, 613-560-7333

Other Shops • Autres commerces

Byward Market 55 Byward Market Square (fresh fruits, vegetables, flowers, souvenirs, crafts, clothing • fruits frais, légumes, fleurs, souvenirs, artisanat, vêtements)

Rideau Centre 50 Rideau Street (180 stores – movie theatres, restaurants, currency exchange, art work, jewellery • 180 boutiques – cinémas, restaurants, bureau de change, œuvres d'art, bijoux)

Pharmacy • Pharmacie in the Rideau Centre, 50 Rideau Street (toiletries, medicine) • dans le Rideau Centre (articles de toilette, médicaments)

Loblaws 363 Rideau Street (groceries • alimentation)

Internet Access • Accès internet

Computers with access to internet are available in West part of the Congress Hall, nearby the exhibitors and the poster sessions.

Des ordinateurs avec accès internet sont disponibles dans la partie ouest du Congress Hall A/E, près des exposants et des sessions sur affiches.

Daycare on Site • Garderie sur place

GiGi's Sitting Service 613-749-1295

Adult babysitters, multilingual, bonded, C.P.R. certified • Gardiennes adultes, polyglottes, sous contrat, certifiées R.C.R.

Reservations required. CASH ONLY. Minimum of four hours, 16 CAD per hour for up-to three (3) children. If more than three children, add 1 CAD per hour. If there are two children under five (5) years of age, then two babysitters are required. Return taxi fare must be paid for the babysitter if services are required past 23:30.

Réservations requises. COMPTANT SEULEMENT. Minimum de quatre (4) heures, 16 CAD de l'heure pour trois (3) enfants ou moins. Si plus de trois enfants, on doit ajouter 1 CAD de l'heure. S'il y a deux enfants de moins de cinq (5) ans, alors deux gardiennes sont requises. Les frais de taxi pour le retour doivent être payés si les services requis dépassent 23h30.

Committees and Meetings
Comités et réunions

Fauteux, Morissette, Montpetit and University Centre are buildings of the University of Ottawa. Capital Hall is the basement of the Ottawa Congress Centre.

Fauteux, Morissette, Montpetit et University Centre sont les noms de pavillons de l'Université d'Ottawa. Capital Hall est le sous-sol du Centre des congrès d'Ottawa.

Calendar Listing • Horaire journalier

Saturday May 24 • Samedi 24 mai

18:30–22:00	Executive Committee • *Comité exécutif*	

Sunday, May 25 • Dimanche 25mai

9:00–11:00	Finance Committee • *Comité des finances*	Capital Hall 5B
9:00–12:00	Heads Meeting • *Rencontre des directeurs de départements*	Fauteux 137
11:00–12:00	Publications Committee • *Comité des publications*	Capital Hall 5B
12:00–16:00	Board of Directors • *Conseil d'administration*	Capital Hall 5B

Monday, May 26 • Lundi 26 mai

12:15–13:30	Probability Section (Section Officers) • *Groupe de probabilité (directeurs du groupe)*	Morissette 250
	Survey Methods Section (Section Officers) • *Groupe des méthodes d'enquête (directeurs du groupe)*	Morissette 251
	Biostatistics Section (Section Officers) • *Groupe de biostatistique (directeurs du groupe)*	Morissette 252
	Business and Industrial Statistics Section (Section Officers) • *Groupe de statistique industrielle et de gestion (directeurs du groupe)*	Morissette 256
	Accredited Members General Meeting (open) • *Assemblée générale des membres accrédités (public)*	Montpetit 204
12:30–14:00	*Réunion du conseil de la SFdF*	Morissette 221
18:00–19:00	Biostatistics Section (General Meeting)• *Groupe de biostatistique (assemblée générale)*	Capital Hall 1A

	Business and Industrial Statistics Section (General Meeting) • *Groupe de statistique industrielle et de gestion (assemblée générale)*	Capital Hall 2A
	Probability Section (General Meeting) • *Groupe de probabilité (assemblée générale)*	Capital Hall 3A
	Survey Methods Section (General Meeting) • *Groupe des méthodes d'enquête (assemblée générale)*	Capital Hall 4A
	Assemblée générale des membres de l'ASSQ	Capital Hall 5A

Tuesday, May 27 • Mardi 27 mai

12:00–14:00	SFdS–SSC Meeting on Ethics • *Réunion conjointe SFdS et SSC sur l'éthique*	Morissette 250
12:15–13:30	Research Committee • *Comité de la recherche*	Morissette 221
	NPCDS • *PNSDC*	Morissette 251
	CJS Editorial Board • *Comité de rédaction de la RCS*	Morissette 252
	Public Relations Committee • *Comité des relations publiques*	Morissette 256
	Bilingualism Committee • *Comité du bilinguisme*	Montpetit 204
18:00–19:00	*Assemblée générale de la SFdS*	Capital Hall 2A
18:00–19:30	SSC Annual General Meeting • *Assemblée générale annuelle de la SSC*	Capital Hall 1A

Wednesday, May 28 • Mercredi 28 mai

12:00–14:00	*Réunion sur l'utilisation du français pendant les congrès internationaux en particulier à l'IIS*	Montpetit 204
12:15–13:00	Biostatistics Section (Section Officers) • *Groupe de biostatistique (directeurs du groupe)*	Morissette 250
12:15–13:30	Committee on Women in Statistics • *Comité sur les femmes en statistique*	Morissette 221
	Liaison	Morissette 251
	Awards Committee • *Comité des prix*	Morissette 252
	Statistical Education Committee • *Comité d'éducation en statistique*	Morissette 256

Thursday, May 29 • Jeudi 29 mai

12:15–13:00	Business and Industrial Statistics Section (Section Officers) • *Groupe de statistique industrielle et de gestion (directeurs du groupe)*	Morissette 251
	Probability Section (Section Officers) • *Groupe de probabilité (directeurs du groupe)*	Montpetit 204
	Survey Methods Section (Section Officers) • *Groupe des méthodes d'enquête (directeurs du groupe)*	Montpetit 207
12:15–13:30	Program Committee • *Comité du programme*	Morissette 250
	Accreditation Committee • *Comité sur l'accréditation*	Fauteux 137
17:30–19:30	Board of Directors • *Conseil d'administration*	University Centre 205
19:30	Executive Committee • *Comité exécutif*	University Centre 205

Alphabetical Listing • Liste alphabétique

Accreditation Committee • *Comité sur l'accréditation*
Thursday, May 29 • jeudi 29 mai, 12:15–13:30, Fauteux 137

Accredited Members General Meeting (open) • *Assemblée générale des membres accrédités (public)*
Monday, May 26 • lundi 26 mai, 12:15–13:30, Montpetit 204

Assemblée générale annuelle des membres de l'ASSQ
lundi 26 mai, 18:00–19:00, Capital Hall 5A

Assemblée génerale de la SFdS
mardi 27 mai, 18:00–19:00, Capital Hall 2A

Awards Committee • *Comité des prix*
Wednesday, May 28 • mercredi 28 mai, 12:15–13:30, Morissette 252

Bilingualism Committee • *Comité du bilinguisme*
Tuesday, May 27 • mardi 27 mai, 12:15–13:30, Montpetit 204

Biostatistics Section (General Meeting) • G*roupe de biostatistique (assemblée générale)*
Monday, May 26 • lundi 26 mai, 18:00–19:00, Capital Hall 1A

Biostatistics Section (Section Officers) • *Groupe de biostatistique (directeurs du groupe)*
Monday, May 26 • lundi 26 mai, 12:15–13:30, Morissette 252
Wednesday, May 28 • mercredi 28 mai, 12:15–13:00, Morissette 250

Board of Directors • *Conseil d'administration*
Sunday, May 25 • dimanche 25 mai, 12:00–16:00, Capital Hall 5B
Thursday, May 29 • jeudi 29 mai, 17:30–19:30, University Centre 205

Business and Industrial Statistics Section (General Meeting) • *Groupe de statistique industrielle et de gestion (assemblée générale)*
Monday, May 26 • lundi 26 mai, 18:00–19:00, Capital Hall 2A

Business and Industrial Statistics Section (Section Officers) • *Groupe de statistique industrielle et de gestion (directeurs du groupe)*
Monday, May 26 • lundi 26 mai, 12:15–13:30, Morissette 256
Thursday, May 29 • jeudi 29 mai, 12:15–13:00, Morissette 251

CJS Editorial Board • *Comité de rédaction de la RCS*
Tuesday, May 27 • mardi 27 mai, 12:15–13:30, Morissette 252

Committee on Women in Statistics • *Comité sur les femmes en statistique*
Wednesday, May 28 • mercredi 28 mai, 12:15–13:30, Morissette 221

Executive Committee • *Comité exécutif*
Saturday May 24 • samedi 24 mai, 18:30–22:00
Thursday May 29 • jeudi 29 mai, 19:30, University Centre 205

Finance Committee • *Comité des finances*
Sunday, May 25• dimanche 25 mai, 9:00–11:00, Capital Hall 5B

Heads Meeting • *Rencontre des directeurs de départements*
Sunday, May 25 • dimanche 25 mai, 9:00–12:00, Fauteux 137

Liaison
Wednesday, May 28 • mercredi 28 mai, 12:15–13:30, Morissette 251

NPCDS • *PNSDC*
Tuesday, May 27• mardi 27 mai, 12:15–13:30, Morissette 251

Probability Section (General Meeting) • *Groupe de probabilité (assemblée générale)*
Monday, May 26 • lundi 26 mai, 18:00–19:00, Capital Hall 3A

Probability Section (Section Officers) • *Groupe de probabilité (directeurs du groupe)*
Monday, May 26 • lundi 26 mai, 12:15–13:30, Morissette 250
Thursday, May 29 • jeudi 29 mai, 12:15–13:00, Montpetit 204

Program Committee • *Comité du programme*
Thursday, May 29 • jeudi 29 mai, 12:15–13:30, Morissette 250

Public Relations Committee • *Comité des relations publiques*
Tuesday, May 27 • mardi 27 mai, 12:15–13:30, Morissette 256

Publications Committee • *Comité des publications*
Sunday, May 25 • dimanche 25 mai, 11:00–12:00, Capital Hall 5B

Réunion du conseil de la SFdS
lundi 26 mai, 12:30–14:00, Morissette 221

Réunion sur l'utilisation du français pendant les congrès internationaux en particulier l'IIS
mardi 28 mai, 12:00–14:00, Montpetit, 204

Research Committee • *Comité de la recherche*
Tuesday, May 27 • mardi 27 mai, 12:15–13:30, Morissette 221

SFdS–SSC Meeting on Ethics • *Réunion conjointe SFdS et SSC sur l'éthique*
Tuesday, May 27 • mardi 27 mai, 12:00–14:00, Morissette 250

SSC General Meeting • *Assemblée générale de la SSC*
Tuesday, May 27 • mardi 27 mai, 18:00–19:30, Capital Hall 1A

Statistical Education Committee • *Comité d'éducation en statistique*
Wednesday, May 28 • mercredi 28 mai, 12:15–13:30, Morissette 256

Survey Methods Section (General Meeting) • *Groupe des méthodes d'enquête (assemblée générale)*
Monday, May 26 • lundi 26 mai, 18:00–19:00, Capital Hall 4A

Survey Methods Section (Section Officers) • *Groupe des méthodes d'enquête (directeurs du groupe)*
Monday, May 26 • lundi 26 mai, 12:15–13:30, Morissette 251
Thursday, May 29 • jeudi 29 mai, 12:15–13:00, Montpetit 207

List of Sessions • Liste des sessions

Monday May 26 • Lundi 26 mai

08:30–08:45

Congress Hall A/E	Session 01A	Inaugural Session *Séance inaugurale*

08:45–10:00

Congress Hall A/E	Session 01B	SSC Presidential Invited Address *Allocution de l'invité de la présidente de la SSC*

10:30–12:00

Capital Hall 1A	Session 02A	Measurement Error Models *Modèles d'erreurs de mesures*
Capital 2A	Session 02B	Session I in Memory of André Dabrowski *Séance I à la mémoire d'André Dabrowski*
Capital 3A	Session 02C	Data Mining *Forage de données*
Capital 4A	Session 02D	Statistical Models for Extreme Values *Modèles statistiques pour les valeurs extrêmes*
Capital 5A	Session 02E	Predictive Models in Business and Industrial Statistics *Modèles de prévisions en statistique industrielle et de gestion*
Capital Hall 6A	Session 02F	Survey Applications *Applications en échantillonnage*
Capital Hall 7A	Session 02G	Nonparametric Regression *Régression non paramétrique*
Capital Hall 8A	Session 02H	Estimation and Hypothesis Tests *Estimation et tests d'hypothèses*
Capital Hall 9A	Session 02I	Teaching of Statistics and Applications *Enseignement de la statistique et applications*

13:30–14:30

Congress Hall A/E	Session 03A	Lucien Le Cam Address *Conférence Lucien Le Cam*

14:35–16:05

Capital Hall 1A	Session 04A	Predicting Modeling Applications in Customer Acquisition and Retention *Prévision de la modélisation des applications en acquisition et rétention de clients*

Capital Hall 2A	Session 04B	Combining Survey and Administrative Data *Combinaison de données administratives et de données d'enquêtes*
Capital Hall 3A	Session 04C	Statistical Analysis of Hydroclimatic Proxy Data: Bayesian Approaches *Analyse statistique des empreintes hydro-climatiques : approches bayésiennes*
Capital Hall 4A	Session 04D	Monte Carlo Markov Chains *Chaînes de Markov Monte Carlo*
Capital Hall 5A	Session 04E	Hypothesis Testing in Biostatistics *Tests d'hypothèse en biostatistique*
Capital Hall 6A	Session 04F	Public Health Studies *Études en santé publique*
Capital Hall 7A	Session 04G	Applications of Functional Estimation *Applications de l'estimation fonctionnelle*
Congress Hall A/E West	Session 04H	Poster Session I *Séance d'affichage I*
Congress Hall A/E West	Session 04I	Case Studies 1: Handling Item Nonresponse in Surveys *Étude de cas 1 : Traitement de la non-réponse à une question dans les enquêtes*
Congress Hall A/E West	Session 04J	Case Study 2: Effects of Genetic Variation on the Relationship between Diet and Cardiovascular Disease Risk *Étude de cas 2 : Effets de la variation génétique sur la relation entre l'alimentation et le risque de maladie cardiovasculaire*

16:30–18:00

Capital Hall 1A	Session 05A	May we, Should we and Why Collect Ethnic Statistics? French and Canadian Points of View *Peut-on, doit-on et pourquoi, collecter des données sur l'origine ethnique ? Les points de vue canadien et français*
Capital Hall 2A	Session 05B	Mixture Models in Statistical Finance *Modèles de mélanges en finance statistique*
Capital Hall 3A	Session 05C	Bayesian Statistics *Statistiques bayésiennes*
Capital Hall 4A	Session 05D	Markov Models in Biostatistics *Modèles markoviens en biostatistique*
Capital Hall 5A	Session 05E	Analysis of Censored Data *Analyse de données censurées*
Capital Hall 6A	Session 05F	Applied Probability *Probabilités appliquée*
Capital Hall 7A	Session 05G	Estimation and Tests for Time Series *Estimation et tests pour séries chronologiques*

18:00–19:00

Capital Hall 6A	NSERC Workshop: Tips to Prepare your Next Discovery Grant Application *Atelier du CRSNG* : *Conseils pour l'élaboration de votre prochaine demande de subvention à la découverte*

Tuesday May 27 • Mardi 27 mai

8:45–10:00

Congress Hall	Session 06A	SFdS Presidential Invited Address *Allocution de l'invité de la présidente de la SFdS*

10:30–12:00

Capital Hall 1A	Session 07A	Data Integration for Genomic Data *Intégration de données sur le génome*
Capital Hall 2A	Session 07B	Session II in Memory of André Dabrowski *Séance II à la mémoire d'André Dabrowski*
Capital Hall 3A	Session 07C	History of Statistical Science *Histoire de la statistique*
Capital Hall 4A	Session 07D	Model Selection *Sélection de modèles*
Capital Hall 5A	Session 07E	Estimation Methods for Surveys Data *Méthodes d'estimation pour données de sondages complexes*
Capital Hall 6A	Session 07F	Semiparametric and Nonparametric Models *Modèles semi-paramétriques et non paramétriques*
Capital Hall 7A	Session 07G	Classification Methods and Data Mining *Méthodes de classification et forage de données*
Capital Hall 8A	Session 07H	Financial Engineering *Ingénierie financière*
Capital Hall 9A	Session 07I	Sensitivity Analysis and Fiability in Business and Industrial Statistics *Analyse de sensibilité et fiabilité en statistique industrielle et de gestion*

13:30–14:30

Congress Hall	Session 08A	SSC Gold Medal Award Address *Allocution du récipiendaire de la médaille d'or de la SSC*

14:35–15:35

Capital Hall 1A	Session 09A	Special Invited Session of the Survey Methods Section *Allocution de l'invité d'honneur du groupe de méthodologie d'enquête*
Capital Hall 2A	Session 09B	Statistical Process Control Methods in Health Care *Méthodes statistiques de contrôle des processus en santé*
Capital Hall 3A	Session 09C	Statistical Analysis of Hydroclimatic Proxy Data: Frequentist Approaches *Analyse statistique des empreintes hydro-climatiques : approches fréquentielles*
Capital Hall 4A	Session 09D	How to Reconcile the Demands for Data Confidentiality and Research Needs? The Role of Secure Access Centers. *Comment concilier les exigences de la confidentialité des données et les besoins de la recherche ? Le rôle des centres d'accès sécurisés*
Capital Hall 5A	Session 09E	Classification and Algorithms *Classification et algorithmes*
Capital Hall 6A	Session 09F	Robustness *Robustesse*

Capital Hall 7A	Session 09G	Analysis of Dependent Data *Analyse de données dépendantes*
Capital Hall 8A	Session 09H	Biostatistical Models *Modèles biostatistiques*
Capital Hall 9A	Session 09I	Probability Laws and Asymptotic Properties *Lois de probabilité et propriétés asymptotiques*
Congress Hall A/E West	Session 09J	Poster Session II *Séance d'affichage II*

16:30–18:00

Capital Hall 1A	Session 10A	Some New Perspectives on Longitudinal Data Analysis *Nouvelles perspectives en analyse de données longitudinales*
Capital Hall 2A	Session 10B	Session in Honour of Denis Bosq *Session en hommage à Denis Bosq*
Capital Hall 3A	Session 10C	Longitudinal Data *Données longitudinales*
Capital Hall 4A	Session 10D	Regression Methods in Biostatistics *Méthodes de régression en biostatistique*
Capital Hall 5A	Session 10E	Copulas and Applications *Copules et applications*
Capital Hall 6A	Session 10F	Limiting Behaviour of Test Statistics *Comportement asymptotique de statistiques de test*
Capital Hall 7A	Session 10G	Survey Methods *Méthodes d'échantillonnage*

Wednesday May 28 • Mercredi 28 mai

8:30–10:00

Congress Hall A/E	Session 11A	Pierre Robillard Award Addresses *Allocutions des récipiendaires du prix Pierre Robillard*

10:30–12:00

Capital Hall 1A	Session 12A	Round Table: Issues in Collaboration in Medical Research *Table ronde : Questionnements sur la collaboration en recherche médicale*
Capital Hall 2A	Session 12B	Monte Carlo Methods *Méthodes Monte Carlo*
Capital Hall 3A	Session 12C	Rank-based Inference *Inférence fondée sur les rangs*
Capital Hall 4A	Session 12D	Censored Data *Données censurées*
Capital Hall 5A	Session 12E	Design of Experiments in Business and Industrial Statistics *Planification d'expériences en statistique industrielle et de gestion*
Capital Hall 6A	Session 12F	Data Validation and Small Area Estimation *Validation de données et estimation d'aires faibles*
Capital Hall 7A	Session 12G	Estimation Methods *Méthodes d'estimation*
Capital Hall 8A	Session 12H	Functional Estimation *Estimation fonctionnelle*

Capital Hall 9A	Session 12I	Extreme Values and Applications *Valeurs extrêmes et applications*

13:30–14:30

Congress Hall A/E	Session 13A	*Canadian Journal of Statistics* Award Address *Allocution du récipiendaire du prix de* La revue canadienne de statistique

14:35–16:05

Capital Hall 1A	Session 14A	Design and Analysis of Industrial Experiments *Planification et analyse d'expériences industrielles*
Capital Hall 2A	Session 14B	Calibration and Variance Estimation in Surveys *Calage et estimation de la variance dans les enquêtes*
Capital Hall 3A	Session 14C	Environmental Statistics *Statistiques en environnement*
Capital Hall 4A	Session 14D	Likelihood Inference *Inférence par vraisemblance*
Capital Hall 5A	Session 14E	Model Assessment and Goodness-of-fit *Évaluation de modèles et adéquation*
Capital Hall 6A	Session 14F	Stochastic Models *Modèles stochastiques*
Capital Hall 7A	Session 14G	Analysis of Microarrays *Analyse de biopuces*
Congress Hall A/E West	Session 14H	Poster Session III *Séance d'affichage III*

16:30–18:00

Capital Hall 1A	Session 15A	Markov Models *Modèles markoviens*
Capital Hall 2A	Session 15B	Actuarial Science *Actuariat*
Capital Hall 3A	Session 15C	Goodness-of-fit Tests *Tests d'adéquation*
Capital Hall 4A	Session 15D	Inference for Linear Models *Inférence pour modèles linéaires*
Capital Hall 5A	Session 15E	Clinical Trials *Essais cliniques*
Capital Hall 6A	Session 15F	Functional Estimation and Censored Data *Estimation fonctionnelle et données censurées*
Capital Hall 7A	Session 15G	Stochastic Models in Biostatistics *Modèles stochastiques en biostatistique*

Thursday May 29 • Jeudi 29 mai

8:30–10:00

Congress Hall A/E	Session 16A	Marie-Jeanne-Laurent-Duhamel Addresses *Allocutions des récipiendaires du prix Marie-Jeanne-Laurent-Duhamel*

10:30–12:00

Capital Hall 1A	Session 17A	Statistical Genomics *Statistique génomique*
Capital Hall 2A	Session 17B	Parametric Statistics on Varieties *Statistique paramétrique sur les variétés*
Capital Hall 3A	Session 17C	Copulas *Copules*
Capital Hall 4A	Session 17D	Empirical Likelihood *Vraisemblance empirique*
Capital Hall 5A	Session 17E	Estimation Methods in Business and Industrial Statistics *Méthodes d'estimation en statistique industrielle et de gestion*
Capital Hall 6A	Session 17F	Mixture Models *Modèles de mélanges*
Capital Hall 7A	Session 17G	Time Series Modelling *Modélisation de séries chronologiques*

13:30–14:30

Congress Hall A/E	Session 18A	Address of the winner of the Centre de Recherches Mathématiques/Statistical Society of Canada Prize *Allocution du récipiendaire du prix Centre de recherches mathématiques/Société statistique du Canada*

14:35–14:35

Capital Hall 1A	Session 19A	Isobel Loutit Invited Address on Business and Industrial Statistics *Allocution Isobel Loutit sur la statistique industrielle et de gestion*

14:35–16:05

Capital Hall 2A	Session 19B	Synthetic Data Methods for Confidentiality Protection and Disclosure Avoidance *Méthodes de données synthétiques pour la protection de la confidentialité et de la non-divulgation*
Capital Hall 3A	Session 19C	Recent Advances in Analysis of Event Data *Développements récents dans l'analyse des données événementielles*
Capital Hall 4A	Session 19D	Time Series *Séries chronologiques*
Capital Hall 5A	Session 19E	Latent Variables *Variables latentes*
Capital Hall 6A	Session 19F	Bayesian Methods *Méthodes bayésiennes*
Capital Hall 7A	Session 19G	Nonparametric Density Estimation *Estimation non paramétrique de la densité*
Capital Hall 8A	Session 19H	Survival Data Analysis *Analyse de données de survie*

16:30–18:00

Capital Hall 1A	Session 20A	Teaching of Statistics: Didactical and Pedagogical Problems *L'enseignement de la statistique : problèmes didactiques et pédagogiques*
Capital Hall 2A	Session 20B	Statistical Methods for Complex Survey Data *Méthodes statistiques pour données de sondages complexes*

Capital Hall 3A	Session 20C	Inference and Applications *Inférence et applications*
Capital Hall 4A	Session 20D	Hypothesis Tests *Tests d'hypothèses*
Capital Hall 5A	Session 20E	Bayesian Methods in Biostatistics *Méthodes bayésiennes en biostatistique*
Capital Hall 6A	Session 20F	Estimation from Biomedical Data *Estimation fondées sur des données biomédicales*
Capital Hall 7A	Session 20G	Estimation and Tests of Serial Dependence *Estimation et tests de dépendance sérielle*
Capital Hall 8A	Session 20H	Multivariate Statistics *Statistique multidimensionnelle*

Program • Programme

Lundi 26 mai • Monday, May 26

08:30 – 08:45 Session 01A — Congress Hall A/E

Inaugural Session
Session inaugurale

Special Session • Conférence spéciale

Special Guest Speakers • Conférenciers invités de marque

Organizer and Chair • Responsable et président: Christian Genest, *Université Laval*

08:45 – 10:00 Session 01B — Congress Hall A/E

SSC Presidential Invited Address
Allocution de l'invité du président de la SSC

Special Session • Conférence spéciale

Organizer and Chair • Responsable et président: Christian Genest, *Université Laval*

08:45 Paul Embrechts, *ETH Zurich* E F

Statistics and Quantitative Risk Management • La statistique et la gestion quantitative des risques [MS-1]

10:00 – 10:30 Pause-café (parrainée par SAS) • Coffee Break (sponsored by SAS)

10:30 – 12:00 Session 02A — Capital Hall 1A

Measurement Error Models
Modèles d'erreur de mesures

Invited Paper Session • Présentations sur invitation

Session sponsored by the *Biostatistics Section*
Session parrainée par le *Groupe de biostatistique*

Organizer and Chair • Responsable et présidente: Grace Yi, *University of Waterloo*

10:30 Naisyin Wang & Erning Li, *Texas A&M University*, Nae-Yuh Wang, *Johns Hopkins University* E E

Effects of Ignoring Measurement-Error Correlations in Longitudinal Covariates • Effets d'ignorer les corrélations de l'erreur de mesure dans les covariables longitudinales [MS-2]

10:52 John Neuhaus & Charles McCulloch, *University of California, San Francisco*

Diagnosing Misspecification of Generalized Linear Mixed Models Using Conditional Likelihood and Covariate Partitioning Methods • Diagnostiquer la mauvaise spécification de modèles linéaires généralisés mixtes au moyen de la vraisemblance conditionnelle et de méthodes de partitionnement des covariables [MS-3]

11:14 Paul Gustafson, *University of British Columbia,* Shahadut Hossain, *BC Cancer Agency*

Bayesian Analysis of Measurement Error Problems • Analyse bayesienne des problèmes d'erreur de mesure [MS-4]

11:36 Richard Cook, *University of Waterloo,* Leilei Zeng, *Simon Fraser University*, Theodore Warkentin, *McMaster University*

Regression Analysis with a Misclassified Covariate from a Current Status Regression Analysis with a Misclassified Covariate from a Current Status Observation Scheme • Analyse de la régression avec une covariable mal classée à partir d'une analyse du statut de régression courant avec une covariable mal classée à partir d'un schème d'observation d'un statut courant [MS-5]

10:30 – 12:00 Session 02B — Capital Hall 2A

Session I in Memory of André Dabrowski
Séance I à la mémoire d'André Dabrowski

Invited Paper Session • Présentations sur invitation

Session sponsored by the *Probability Section*
Session parrainée par le *Groupe de probabilité*

Organizer and Chair • Responsable et président: David McDonald, *Université d'Ottawa*

10:30 Harold Dehling, *University of Bochum*

André Dabrowski's Work on Limit Theorems and Weak Dependence • Les travaux d'André Dabrowski sur les théorèmes limites et la dépendance faible [MS-6]

11:30 David McDonald & Andre Dabrowski, *University of Ottawa*, Uwe Rosler, *Universitat Kiel*

A Reminiscence of our Research Collaboration with Andre Dabrowski • Souvenir de notre collaboration avec Andre Dabrowski [MS-7]

10:30 – 12:00 Session 02C — Capital Hall 3A

Data Mining
Forage de données

Invited Paper Session • Présentations sur invitation

Organizer and Chair • Responsable et président: Hugh Chipman, *Acadia University*

10:30 Charles Bouveyron, *Université Paris 1 Panthéon-Sorbonne*

Classification of Complex Data with Model-based Techniques • Classification de données complexes grâce aux techniques paramétriques [MS-8]

11:00 Steven Wang, *York University,* Weiliang Qiu, *Harvard University*, Ruben Zamar, *University of British Columbia*

An Iterative Non-parametric Clustering Algorithm Based on Local Shrinking • Un algorithme de classification itératif non paramétrique basé sur le rétrécissement local [MS-9]

11:30 Russell Steele, *McGill University*, Anne-Sophie Charest, *Carnegie Mellon University*, Members of the *Canadian Scleroderma Research Group*

Missing Data Approaches for High-dimensional Categorical Data • Approches des données manquantes pour données catégorielles de haute dimension [MS-10]

10:30 – 12:00 Session 02D — Capital Hall 4A

Statistical Models for Extreme Values
Modèles statistiques pour les valeurs extrêmes

Invited Paper Session • Présentations sur invitation

Organizer and Chair • Responsable et présidente: Anne-Laure Fougères, *Université Paris X*

10:30 Arthur Charpentier, *Université Rennes 1*, Johan Segers, *Université Louvain la Neuve* F E

Tails of Archimedean Copulas • Queues des copules archimédiennes [MS-11]

11:00 Étienne Marceau & Hélène Cossette, *Université Laval*, Jean-Phillippe Lemay, *Standard Life Investments Inc.* F E

Actuarial Aspects of Catastrophe Mortality Risk • Acpects actuariels du risque catastrophique de mortalité [MS-12]

11:30 Cécile Mercadier & Gabriela Ciuperca, *Université Lyon 1 (France)* F E

Weighted least Squares Estimation of the Extreme Value Index • Estimation de l'indice des valeurs extrêmes par la méthode des moindres carrés pondérés [MS-13]

10:30 – 12:00 Session 02E — Capital Hall 5A

Predictive Models in Business and Industrial Statistics
Modèles de prévisions en statistique industrielle et de gestion

Contributed Paper Session • Séance de communications libres

Session sponsored by the *Business and Industrial Statistics Section*
Session parrainée par le *Groupe de statistique industrielle et de gestion*

Chair • Président: Christian Derquenne, *Électricité et Gaz de France*

10:30 Jean-Michel Poggi, Michel Misiti, Yves Misiti & Georges Oppenheim, *Université d'Orsay* F E

Optimized Clusters for Disaggregated Electricity Load Forecasting • Classification optimisée pour la prévision désagrégée de la courbe de charge électrique [MS-14]

10:45 Agnès Lagnoux Renaudie & Jean-Marc Azaïs, *LSP - Université Paul Sabatier - Toulouse France*, Sophie Bercu, *EDF France*, Jean-Claude Fort, *LSP - Université Paul Sabatier - Toulouse France*, Pierre Lé, *EDF France* F E

Load Curves Estimation and Simultaneous Confidence Bands • Estimation et intervalles de confiance simultanés de courbes de charge [MS-15]

11:00 Julien Fouquau, *LEO*, Marie Bessec, *CGEMP* F E

The Non-linear Link between Electricity Consumption and Temperature in Europe: a Threshold Panel Approach • Le lien non linéaire entre la consommation électrique et la température en Europe, une approche sur panel à seuil [MS-16]

11:15 Alain Dessertaine, *EDF R&D OSIRIS* F E

Non-parametric Electric Load Forecasting by "Aggregation/Disaggregation" • Prévision non paramétrique par "Agrégation/désagrégation" de la consommation électrique [MS-17]

10:30 – 12:00 Session 02F — Capital Hall 6A

Survey Applications
Applications en échantillonnage

Contributed Paper Session • Séance de communications libres

Session sponsored by the *Survey Methods Section*
Session parrainée par le *Groupe de méthodologie d'enquête*

Chair • Président: Yves Tillé, *Université de Neuchâtel*

10:30 Pierre Daoust, Sylvie Gauthier & Nora Bohossian, *Statistique Canada* (F) [E]

The Quarterly Financial Survey (QFS) for Enterprises • L'enquête financière trimestrielle (EFT) des sociétés [MS-18]

10:45 Xiaohong Peng, Rameshwar Gupta & Renjun Ma, *University of New Brunswick* (E) [E]

Childhood Weight and Cognitive Development: A Study of Young Children in Canada • Poids et développement cognitif durant l'enfance : une étude de la petite enfance au Canada [MS-19]

11:00 Salima Taibi, *LMRS Université de Rouen,* Manase Bezara, *Université de Tamatave Madagascar,* Candice Rouen, *INSA Rouen* (F) [E]

Production Model for Rice in Madagascar • Modèle de rendement du riz à Madagascar [MS-20]

11:15 Javier Oyarzun, *Statistique Canada* (F) [E]

The Experience of Using Unincorporated Business Tax Data • L'utilisation des données fiscales : l'expérience des entreprises non incorporées [MS-21]

11:30 Olga Vasechko, *Research Institute of Statistics, Kyiv, Ukraine,* Michel Grun-Rehomme, *Université Paris2, ERMES* (F) [E]

A New Synthetic Index for Qualitative Short-term Surveys • Un nouvel indicateur synthétique pour les enquêtes d'opinion [MS-22]

11:45 Tracy Tabuchi, *Statistics Canada* (E) [F]

Using Paradata to Evaluate Collection Activities in the Youth in Transition Survey • L'utilisation des paradonnées pour évaluer les activités de collecte dans l'Enquête auprès des jeunes en transition [MS-23]

10:30 – 12:00 Session 02G — Capital Hall 7A

Nonparametric Regression
Régression non paramétrique

Contributed Paper Session • Séance de communications libres

Chair • Président: Eric Matzner-Lober, *Université Rennes II*

10:30 Olivier Faugeras, *LSTA - Université Pierre et Marie Curie - Paris 6* (F) [E]

A Quantile-Copula Approach to Conditional Density Estimation • Estimation de la densité conditionnelle par quantiles et copules [MS-24]

10:45 Jean-Michel Poggi, *Université d'Orsay,* Anestis Antoniadis, *Laboratoire Jean Kuntzmann, Université Joseph Fourier,* Irene Gijbels, *Department of Mathematics and Leuven Statistics, Research Centre (LStat)* (F) [E]

Smoothing Non-Equispaced Heavy Noisy Data with Wavelets • Estimation par ondelettes pour des données fortement bruitées et non equiespacées [MS-25]

11:00 Jean-Yves Brua, *Institut de Recherche Mathématique Avancée*

Asymptotic Behaviour of the Minimax Risk in Nonparametric Heteroscedastic Regression Models • Comportement asymptotique du risque minimax dans des modèles de régression non parametriques et hétéroscédastiques [MS-26]

11:15 Laurent Rouvière, *Université Rennes 2*, Nicolas Hengartner, *Los Alamos National Laboratoty*, Eric Matzner-Lober, *Université Rennes 2*

Multiplicative Bias Corrected Nonparametric Smoothers • Correction multiplicative du biais en régression [MS-27]

11:30 Johan Segers, *Université catholique de Louvain*, Mary C. Meyer, *Colorado State University*

Estimating the Support Function of a Convex Region using Regression Splines • Estimation de la fonction de support d'une partie convexe moyennant les splines de régression [MS-28]

11:45 Zaher Mohdeb, *Université Mentouri de Constantine, Algérie*

Hypothesis Testing for the Equality of Nonparametric Regression Functions • Test d'hypothèse de l'égalité de fonctions de régression non paramétriques [MS-29]

10:30 – 12:00 Session 02H — Capital Hall 8A

Estimation and Hypothesis Tests
Estimation et tests d'hypothèses

Contributed Paper Session • Séance de communications libres

Chair • Président: Alain Latour, *Université de Grenoble*

10:30 Karim Oualkacha & Louis-Paul Rivest, *Université Laval*

Statistical Models for $SE(p)$ Data • Modèles statistiques pour les données de $SE(p)$ [MS-30]

10:45 Angelo Canty & Shaheena Bashir, *McMaster University*

Model-based Bootstrapping for M-estimators • Le bootstrap basé sur des modèles pour des m-estimateurs [MS-31]

11:00 Richmond Oldford, *University of Waterloo*, Catherine Hurley, *National University of Ireland (Maynooth)*

Some Applications of Graph Theory to Statistical Graphics • Certaines applications de la théorie des graphes aux graphiques statistiques [MS-32]

11:15 Maia Lesosky & Peter Kim, *University of Guelph*

Regularized Solutions of Fredholm Integral Equations on the Euclidean Motion Group • Solutions régularisées aux équations de l'intégrale de Fredholm sur le groupe des déplacement euclidiens [MS-33]

11:30 Gun Ho Jang & Michael Evans, *University of Toronto*

Invariant P-values for Model Checking and Checking for Prior-data Conflict • Valeurs-p invariantes pour la vérification de modèles et la détection de conflits données-*a priori* [MS-34]

11:45 Linglong Kong & Ivan Mizera, *University of Alberta*

Construction of Bivariate Charts Using Quantile Contours • Construction de graphiques bivariées en utilisant les contours du quantile [MS-35]

10:30 – 12:00 Session 02I — Capital Hall 9A

Teaching of Statistics and Applications
Enseignement de la statistique et applications

Contributed Paper Session • Séance de communications libres

Chair • Président: Jean-Claude Régnier, *Université de Lyon II*

10:30 François Husson, Julie Josse & Lê Sébastien, *Agrocampus Rennes* E E

FactoMineR an R Package for Exploratory Data Analysis for Teaching and Research • FactoMineR : un ensemble R d'analyses exploratoires des données pour l'enseignement et la recherche [MS-36]

10:45 AnneMichele Millar, *Mount Saint Vincent University* E E

Dispelling Statistical Myths: Consulting and Collaboration. • Dissiper les mythes statistiques : la consultation et la collaboration [MS-37]

11:00 Alain Bihan-Poudec, *Université Catholique de l'Ouest (France)*, François Larose, *Université de Sherbrooke* F E

French Social Science and Humanities Undergraduate Representations of the Statistics • Représentation de la statistique chez des étudiants de premier cycle universitaire en sciences humaines et sociales [MS-38]

11:15 Bethany White & John Braun, *The University of Western Ontario* E E

Service-learning in Undergraduate Statistics Education • Programme d'apprentissage communautaire dans les études de premier cycle en statistique [MS-39]

11:30 Muhammad Islam, *MacEwan College* E E

How Numbers Reveal and Conceal the Truth in Capturing Patterns in Qualitative Data? • De quelle façon les nombres révèlent-ils et cachent-ils la vérité dans la saisie de modèles pour les données qualitatives ? [MS-40]

11:45 Danielle Mihram, *University of Southern California*, G. Arthur Mihram, *Princeton, NJ* E F

Science (I.Statistics; II.Applied Mathematics; III. Pure Mathematics): Viewed via Newton and Buffon • La Science (I. La Statistique ; II. Les Mathématiques Appliquées ; III. Les Mathématiques Pures) : Vue dans les œuvres de Newton et de Buffon [MS-41]

12:00 – 13:30 Dîner • Lunch

13:30 – 14:30 Session 03A — Congress Hall A/E

Lucien Le Cam Address
Conférence Lucien Le Cam

Special Session • Conférence spéciale

Organizer and Chair • Responsable et président: Marc Hallin, *Université Libre de Bruxelles*

13:30 Richard Gill, *Leiden University, Netherlands* E E

Hunting Serial Killer Nurses with Statistics • Poursuite de d'infirmières tueuses en série avec statistique [MS-42]

14:35 – 16:05 Session 04A — Capital Hall 1A

Predicting Modeling Applications in Customer Acquisition and Retention
Prévision de la modélisation des applications en acquisition et rétention de clients

Invited Paper Session • Présentations sur invitation

Session sponsored by the *Business and Industrial Statistics Section*
Session parrainée par le *Groupe de statistique industrielle et de gestion*

Organizer and Chair • Responsable et président: Patrick Turgeon, *Figurs* | Analytics, Insights*

14:35 Daniel Lemire, *Indicia* (F) [E]

Traps to Avoid when Building a Sales Forecasting Model • Pièges à éviter lors de la conception d'un modèle de prévision des ventes [MS-43]

15:05 Olivia Parr-Rud, *OLIVIAGroup* (E) [E]

Measures that Matter - Driving Predictive Model Results to the Bottom Line • Des mesures qui importent - passer des résultats d'un modèle de prévision au résultat net [MS-44]

15:35 Emma Warrillow, *Emma Warrillow & Associates Inc.* (E) [E]

Business-to-Business Applications of Predictive Modeling • Applications de la modélisation de la prévision de compagnie à compagnie [MS-45]

14:35 – 16:05 Session 04B — Capital Hall 2A

Combining Survey and Administrative Data
Combinaison de données administratives et de données d'enquêtes

Invited Paper Session • Présentations sur invitation

Session sponsored by the *Survey Methods Section*
Session parrainée par le *Groupe de méthodologie d'enquête*

Organizer and Chair • Responsable et président: Lenka Mach, *Statistics Canada*

14:35 Philippe Brion, *INSEE* (F) [E]

The Role of the Estimators for the Production of the French Structural Business Statistics • Le rôle des estimateurs dans la production des statistiques françaises structurelles d'entreprises [MS-46]

14:57 Stuart Pursey & François Brisebois, *Statistics Canada* (E) [F]

Estimating Monthly Sales by Combining Survey Data with Administrative Data in the Monthly Survey of Food Services and Drinking Places • L'estimation des ventes mensuelles en combinant les données d'enquête et les données administratives de l'Enquête mensuelle sur les services de restauration et débits de boisson [MS-47]

15:19 Michael Larsen, *Iowa State University*, Partha Lahiri, *University of Maryland* (E) [E]

Regression Analysis of Linked Files • Analyse de régression des fichiers couplés [MS-48]

15:41 Therese Stukel, Doug Manuel & Laura Rosella, *Institute for Clinical Evaluative Sciences* (E) [F]

Predicting New Cases of Diabetes in Canada • Prédiction des nouveaux cas de diabètes au Canada [MS-49]

14:35 – 16:05 Session 04C — Capital Hall 3A

Statistical Analysis of Hydroclimatic Proxy Data: Bayesian Approaches
Analyse statistique des empreintes hydro-climatiques : approches bayésiennes

Invited Paper Session • Présentations sur invitation

Session sponsored by / Session parrainée par
Groupe environnement et statistique de la SFdS

Organizer and Chair • Responsable et président: Éric Parent, *ENGREF*

14:35 Jean-Jacques Boreux, *Université de Liège*, Philippe Naveau, *Laboratoire des Sciences du Climat et de l'Environnement (LSCE)*, Luc Perrault, *IREQ, Hydro-Québec*, Antoine Nicault, *INRS, Québec*, Eric Parent, *Paris AGRO TECH*, Jacques Bernier (F) (E)

A Bayesian Model to Extract a Common Factor from a Set of Tree-rings Width • Un modèle bayésien pour extraire un facteur commun d'une population d'arbres [MS-50]

15:05 Vincent Garreta, *CEREGE, CNRS/Université Aix-Marseille*, Paul Miller, *Geobiosphere Science Center, department of Physical Geography & Ecosystems Analysis, Lund University*, Joël Guiot & Christelle Hély, *CEREGE, CNRS/Université Aix-Marseille* (F) (E)

Temporal Model for Climate Reconstruction using Pollen Data by Dynamic Vegetation Model Inversion • Modèle temporel pour la reconstruction du climat à partir de pollens et en inversant un modèle dynamique de végétation [MS-51]

15:35 Delphine Grancher, *CNRS Bellevue*, Ophélie Guin & Philippe Naveau, *LSCE*, Jean-Jacques Boreux, *Université de Liège* (F) (E)

Hierarchical Model Applied to Dendrochronology • Modèles hiérarchiques appliqués à la dendrochronologie [MS-52]

14:35 – 16:05 Session 04D — Capital Hall 4A

Monte Carlo Markov Chains
Chaînes de Markov Monte Carlo

Invited Paper Session • Présentations sur invitation

Chair • Président: Bruno Rémillard, *HEC-Montréal*

Organizer • Responsable: Jeffrey Rosenthal, *University of Toronto*

14:35 Christophe Andrieu, *University of Bristol* (E) (E)

Some New Results on Controlled MCMC Algorithms • Quelques résultats récents concernant les MCMC contrôles [MS-53]

15:05 Mylène Bédard, *Université de Montréal* (F) (E)

On the Optimal Scaling Problem for Metropolis-Hastings Algorithms • Robustesse de l'échelonnage optimal pour les algorithmes de type Metropolis-Hastings [MS-54]

15:35 Eric Moulines & Gersende Fort, *Institut des Télécommunications/Telecom ParisTech* (E) (E)

What Can we Learn from the Renormalization Methods in MCMC? • Que peut-on apprendre des méthodes de renormalisation pour les algorithmes de MCMC ? [MS-55]

14:35 – 16:05 Session 04E — Capital Hall 5A

Hypothesis Testing in Biostatistics
Tests d'hypothèse en biostatistique

Contributed Paper Session • Séance de communications libres

Session sponsored by the *Biostatistics Section*
Session parrainée par le *Groupe de biostatistique*

Chair • Présidente: Naisying Wang, *Texas A&M University*

14:35 Xu Jianfeng & Yingwei Peng, *Queen's University* E E

Statistical Tests in Selecting Cure Models • Tests statistiques pour la sélection de modèles de guérison [MS-56]

14:50 Alexandre Bureau, Jordie Croteau, Chantal Mérette, Yvon Chagnon, Marc-André Roy & Michel Maziade, *Centre de recherche Université Laval - Robert-Giffard* E E

Statistics for Testing Genetic Linkage under Two-gene Epistatic Models • Statistiques pour tester la liaison génétique sous des modèles épistatiques à deux gènes [MS-57]

15:05 Sameer Parpia & Lehana Thabane, *McMaster University* E E

Investigating the Impact of Different Effect Measures on Power for Trials Using Binary Outcomes • Examen de l'effet des mesures d'effet différents sur la puissance des essais utilisant des résultats binaires [MS-58]

15:20 Yun Joo Yoo, *Samuel Lunenfeld Research Institute, Mount Sinai Hospital*, Lei Sun, *University of Toronto*, Shelley Bull, *Samuel Lunenfeld Research Institute, Mount Sinai Hospital* E E

On Using Prior Information to Improve Power of Genome-wide Studies: Weighted P-values Versus Stratified False Discovery Control • De l'utilisation de l'information *a priori* pour améliorer la puissance des études sur tout le génome : valeurs-p pondérées contre contrôle stratifié des fausses découvertes [MS-59]

15:35 Gulhan Alpargu, *California State University, Fullerton* E E

Missing Data: TDT with Two Affected Offspring • Données manquantes : test du déséquilibre de transmission avec deux enfants affectés [MS-60]

15:50 Mohamedou Sow, *GEMA Université Bordeaux 1 et CNRS*, Laurent Briollais, *Samuel Lunenfeld Research Institute*, Gilles Durrieu, *University Bordeaux 1 et CNRS* F F

Robust Statistical Approaches for Sib-Pair Linkage Analysis of Quantitative Trait Loci (QTLs) • Approche robuste de l'analyse de liaisons génétiques pour des traits quantitatifs [MS-61]

14:35 – 16:05 Session 04F — Capital Hall 6A

Public Health Studies
Études en santé publique

Contributed Paper Session • Séance de communications libres

Chair • Président: Marc Fredette, *HEC-Montréal*

14:35 William Ross & Cunye Qiao, *Health Canada* E E

The Safety of Genetically Modified Foods - Defining Statistical Evidence • La sûreté des aliments génétiquement modifiés - définir la preuve statistique [MS-62]

14:50 David Stephens, *McGill University*, Daniel Graham, *Imperial College London* E E

The Impact of Components of Deprivation on Child Pedestrian Casualties in England • L'effet des

composantes de la défavorisation sur les accidents impliquant des enfants piétons en Angleterre [MS-63]

15:05 Chenglin YE, Gina BROWNE & Lehana THABANE, *McMaster University* E E

Measurement of Integrated Human Service Networks • Mesure des réseaux de service humains intégrés [MS-64]

15:20 Annie RUDNICHI, *Hôpital Hôtel-Dieu, Paris* F E

Use of the Database from the Salaried Workers' French Health Insurance System for the Analysis of the Efficiency of Care Pathways in Chronic Diseases: Preliminary Results • Utilisation des données de la CNAMTS dans l'analyse de l'efficience des parcours de soins dans les maladies chroniques : résultats préliminaires [MS-65]

15:35 Salima TAIBI, *Esitpa - Rouen*, Aurore LAMBERT, *INSA - Rouen*, Patrice LEPELLETIER & Karine LAVAL, *Esitpa - Rouen*, Christian MOUGIN, *INRA - Versailles* F E

Establishing a Soil Quality Index • Mise en place d'un indice la qualité des sols [MS-66]

15:50 Sophie TRICAUD-VIALLE & Elodie PAMBRUN, *Comité de Coordination de l'Evaluation Clinique et de la Qualité en Aquitaine (CCECQA)* F E

From the Indicator to the Honours List : Relevant Use of Graph Chart Representations • De l'indicateur au palmarès : utilisation pertinente des représentations graphiques [MS-67]

14:35 – 16:05 Session 04G — Capital Hall 7A

Applications of Functional Estimation
Applications de l'estimation fonctionnelle

Contributed Paper Session • Séance de communications libres

Chair • Président: Jim RAMSAY, *McGill University*

14:35 Nicolas BRUNEL & Florence D'ALCHÉ-BUC, *laboratoire IBISC, Université d'Evry* F E

Estimation of Differential Equations via Nonparametric Methods: Applications in Systems Biology • Estimation d'équations différentielles par des méthodes non paramétriques : applications en biologie systémique [MS-68]

14:50 Jeanne CASAERT & Denis BOSQ, *Laboratoire de Statistique Théorique et Appliquée (Université Paris 6)*, Rachid ZIANI, *SNCF* F E

Kriging Modelling of the Degradation along the Catenary Contact Wire • Modélisation par krigeage de la dégradation le long du fil de contact caténaire [MS-69]

15:05 Stéphane GERMAIN, Belkacem ABDOUS, Pierre VALOIS & Louis-Paul RIVEST, *Université Laval* F E

Option Characteristic Curves Estimation by Penalized Marginal Maximum Likelihood • Estimations des courbes caractéristiques d'options par maximum de vraisemblance marginale pénalisée [MS-70]

15:20 Benoit CADRE, *ENS Cachan Bretagne* F E

Kernel Estimation of Density Level Sets • Estimation des ensembles de niveaux d'une densité par la méthode du noyau [MS-71]

14:35 – 16:05 Session 04H — Congress Hall A/E West

Poster Session I
Séance d'affichage I

Contributed Paper Session • Séance de communications libres

14:35 Christophe Lalanne, Marianne Mavel, Marina Esposito-Farèse & Pascal Bessonneau, *Centre international d'études pédagogiques* F

Psychometrics for Educational Assessment: The Test de connaissance du français • Approche psychométrique de l'évaluation dans le domaine de l'éducation : le Test de connaissance du français [MS-72]

14:35 Chel Hee Lee & Taraneh Abarin, *University of Manitoba* E

Instrumental Variable Approach to Covariate Measurement Error in Generalized Linear Models • Approche des variables instrumentales pour l'erreur de mesure des covariables dans les modèles linéaires généralisées [MS-73]

14:35 Paul Nguyen & Duncan Murdoch, *University of Western Ontario* E

Approximate Conditional Permutation Tests for Logistic Models • Tests de permutations conditionnels approximatifs pour modèles logistiques [MS-74]

14:35 Huanhuan Wu, *Simon Fraser University*, Nancy Dewith, *The B.C. Ministry of Agriculture &Lands*, Carl Schwarz & Thomas Loughin, *Simon Fraser University* E

Statistical Inference for Minimum Inhibitory Concentration (MIC) Data • Inférence statistique pour données de concentration d'inhibition minimale [MS-75]

14:35 Cathy Maugis & Bertrand Michel, *Université Paris-Sud XI* E

A Non-asymptotic Penalized Criterion for Gaussian Mixture Model Selection • Critère pénalisé non asymptotique pour la sélection d'un modèle de mélange gaussien [MS-76]

14:35 Qiujing Zhou & Brad Johnson, *University of Manitoba* E

Coupon Collecting in the Presence of Rare Coupons • Collection de coupons et la présence de coupons rares [MS-77]

14:35 Taehan Bae & Reg. J. Kulperger, *The University of Western Ontario* E

A Model for Corporate Exit Prediction: Discrete Hazard and Competing Risks Model • Un modèle de prévision de la sortie d'une corporation du marché : modèle à défaillance discrète et à risques compétitifs [MS-78]

14:35 Miao (Mandy) Zhang, *University of Manitoba* E

Optimal Semi-Foldover Plans for Two-Level Fractional Factorial Designs • Plans avec demi-repliement optimaux pour des plans fractionnels factoriels à deux niveaux [MS-79]

14:35 Ashkan Zahedi & Sorana Froda, *Université du Québec à Montréal* E

Simple Tests for Identifying the Holling Type II Predator-prey Model • Quelques tests simples pour identifier le modèle prédateur-proie de Holling type II [MS-80]

14:35 Monica Sirski, *University of Manitoba* E

Diagnostic Procedures for Functional Data Analysis • Procédures diagnostiques pour l'analyse des données fonctionnelles [MS-81]

14:35 Hui Song, Yingwei Peng, Dongsheng Tu & Yi Niu, *Queen's University* E

Testing the Homogeneity in Two-sample Censored Mixture Model when Cure Proportion May Exist in One Group • Tester l'homogénéité dans un modèle de mélange à deux échantillons censurés lorsque qu'une proportion de guérison peut exister dans un groupe [MS-82]

14:35 – 16:05 Session 04I — Congress Hall A/E West

Case Studies 1: Handling Item Nonresponse in Surveys

Étude de cas 1 : Traitement de la non-réponse à une question dans les enquêtes

Case Study • Étude de cas

Organizers • Responsables: David Haziza, *Université de Montréal* & Gordon Kuromi, *Statistics Canada*

14:35 He (Shawn) Gao, Shuang Lu, Beilei Wu & Han Zhang, *University of Calgary* E

14:35 Michael Regier, Alexandra Romann & Samuel Wong, *University of British Columbia* E

14:35 Monica Moreno, *University of Alberta* E

14:35 Josh Gould, Le Li, Vineetha Warriyar & Yingru Celia Huang, *University of Waterloo* E

14:35 – 16:05 Session 04J — Congress Hall A/E West

Case Study 2: Effects of Genetic Variation on the Relationship between Diet and Cardiovascular Disease Risk

Étude de cas 2 : Effets de la variation génétique sur la relation entre l'alimentation et le risque de maladie cardiovasculaire

Case Study • Étude de cas

Organizer • Responsable: Alison Gibbs, *University of Toronto*

14:35 Kelly Burkett, Ji-Hyung Shin & Conghui Qu, *Simon Fraser University* E

14:35 Maggie Chen, Ye Li & Ying Qi, *University of Toronto* E

14:35 Sebastian (Song) Li, *University of Toronto* E

14:35 Stephen Welsh, *UBC-Okanagan* E

14:35 Nathan Sahrmann & Brandon Lewis, *UNBC* E

14:35 Longyang Wu, Baojiang Chen, Zhijian Chen & Lihua Wang, *University of Waterloo* E

14:35 Zhidong Wang, Fang Chang, Yawen Xu & David Kim, *York University* E

16:05 – 16:30 Pause-café • Coffee Break

16:30 – 18:00 Session 05A — Capital Hall 1A

May We, Should We and Why Collect Ethnic Statistics? French and Canadian Points of View

Peut-on, doit-on et pourquoi, collecter des données sur l'origine ethnique ? Les points de vue canadien et français

Invited Paper Session • Présentations sur invitation

Session sponsored by / Session parrainée par le *Groupe statistique et société de la SFdS*

Organizer and Chair • Responsable et président: Jean-Louis Bodin, *ADETEF-France*

16:30 Tina Chui & Hélène Maheux, *Statistics Canada/Statistique Canada* F

Collecting Ethno-cultural Data in a Multicultural Society: the Canadian Experience • La collecte de données ethnoculturelles dans une société multiculturelle : l'expérience canadienne [MS-83]

17:00 Roxane Silberman, *Directrice de recherche au CNRS/Secrétaire générale du Comité de concertation pour les données en sciences humaines et sociales* (F) [F]

Can We Ask Questions about Skin Colour? An Analysis of the New French Controversy • Peut-on poser des questions sur la couleur de la peau ? Une analyse de la nouvelle polémique française [MS-84]

17:30 René Padieu, *SFdS - Commission de déontologie* (F) [E]

Is Collecting Data on Ethnic Origin Legitimate? • Est-il licite de recueillir des données sur l'origine ethnique ? [MS-85]

16:30 – 18:00 Session 05B — Capital Hall 2A

Mixture Models in Statistical Finance
Modèles de mélanges en finance statistique

Invited Paper Session • Présentations sur invitation

Organizer and Chair • Responsable et président: Jiahua Chen, *University of British Columbia*

16:30 Zeng-Hua Lu, *University of South Australia,* Chen, *University of British Columbia*, Trogdon, *RTI International* (E) [E]

Finite Mixture of Binary Logistic Regression Analysis on Labor Supply of Mothers with Disabled Children • Mélange fini de l'analyse de régression logistique binaire sur l'emploi des mères d'enfants handicapés [MS-86]

17:00 Weidong Tian & Carole Bernard, *University of Waterloo* (E) [E]

Insurance Market Effects of Risk Management Metrics • Effets des mesures de la gestion des risques sur le marché de l'assurance [MS-87]

17:30 Yonggan Zhao, *Dalhousie University* (E) [E]

Exchange Rate Dynamics with Regime Switching • Dynamique du taux de change avec changement de régimes [MS-88]

16:30 – 18:00 Session 05C — Capital Hall 3A

Bayesian Statistics
Statistiques bayésiennes

Invited Paper Session • Présentations sur invitation

Organizer and Chair • Responsable et président: Jean-François Angers, *Université de Montréal*

16:30 Raphael Gottardo, Francois Caron & Arnaud Doucet, *University of British Columbia* (E) [E]

On-line Changepoint Detection and Parameter Estimation for Genome-wide Transcript Analysis • Détection du point de changement en ligne et estimation des paramètres pour la transcription de l'analyse pour tout le génome [MS-89]

17:00 James Merleau, *Institut de Recherche d'Hydro-Québec*, Jean-François Angers, *Université de Montréal,* Luc Perreault, *Institut de Recherche d'Hydro-Québec*, Anne-Catherine Favre, *INRS, Eau, Terre et Environnement* (E) [E]

Bayesian Free-knot Regression Splines • Splines de régression bayésiens [MS-90]

17:30 Christian Robert, *Université Paris Dauphine et CREST, INSEE, Paris*, Jean-Michel Marin, *INRIA Futurs, Université Paris Sud Orsay*, Nicolas Chopin, *CREST et ENSAE, INSEE, Paris* (E) [E]

Computational Approaches to Bayesian Model Choice • Méthodes algorithmiques pour la sélection de modèles bayesiens [MS-91]

16:30 – 18:00 Session 05D — Capital Hall 4A

Markov Models in Biostatistics
Modèles markoviens en biostatistique

Contributed Paper Session • Séance de communications libres

Session sponsored by the *Biostatistics Section*
Session parrainée par le *Groupe de biostatistique*

Chair • Président: Gérard Derzko, *Sanofi-Aventis-Montpellier*

16:30 Kelly Burkett, Jinko Graham & Brad McNeney, *Simon Fraser University* (E) E

Sampling Genetic Ancestries Conditional on Observed Genotype Data • Échantillonnage des origines génétiques conditionnellement aux données génotypiques observées [MS-92]

16:45 Fabrice Larribe, *Université du Québec à Montréal*, Sabin Lessard, *Université de Montréal* (E) E

A New Composite-conditional-likelihood Approach for Fine Genetic Mapping • Une nouvelle approche de cartographie génétique basée sur des vraisemblance conditionnelles composite [MS-93]

17:00 Ahmadou Alioum, Pierre Joly, Luc Letenneur & Daniel Commenges, *Centre de Recherche INSERM - Université de Bordeaux 2* (F) E

A Time-varying Coefficients Illness-Death Model for the Effect of Gender on Incidence and Mortality of Dementia • Etude de l'effet du genre sur l'incidence de la démence et de la mortalité à l'aide d'un modèle sain-malade-mort avec covariables et coefficients de régression dépendants du temps [MS-94]

17:15 Gregory Pond, *Everest Clinical Research Services/University of Toronto*, W.Y. Wendy Lou, *University of Toronto* (E) E

Design and Analysis of Sequential Clinical Trials Using a Markov Chain Transition Rate Model with Conditional Power • Planification et analyse d'essais cliniques séquentiels utilisant un modèle de taux de transitions d'une chaîne de Markov à puissance conditionnelle [MS-95]

17:30 Baojiang Chen, Grace Yi & Richard Cook, *University of Waterloo* (E) E

Analysis of Disease Progression via Multi-State Models under a Nonignorable Inspection Process • Analyse de l'évolution de la maladie à l'aide de modèles multi-états sous un processus d'inspection non ignorable [MS-96]

17:45 Tung-Lung Wu & James Fu, *University of Manitoba* (E) E

Finite Markov Chain Imbedding and Its Application to Matching Probability between Two Markov Dependent Biological Sequences • Insertion d'une chaîne de Markov finie et son application à la probabilité d'appariement de deux séquences biologiques markoviennes dépendantes [MS-97]

16:30 – 18:00 Session 05E — Capital Hall 5A

Analysis of Censored Data
Analyse de données censurées

Contributed Paper Session • Séance de communications libres

Session sponsored by the *Biostatistics Section*
Session parrainée par le *Groupe de biostatistique*

Chair • Présidente: Brenda MacGibbon, *Université du Québec à Montréal*

16:30 Zhulin He & Xuewen Lu, *University of Calgary* (E) E

Estimation for the Censored Single-Index Model • Estimation pour le modèle à indice unique censuré [MS-98]

16:45 Amélie Detais & Jean-François Dupuy, *Université Paul Sabatier Toulouse III* (F) [E]

Maximum Likelihood Estimation in a Partially Observed Stratified Regression Model with Censored Data • Estimation du maximum de vraisemblance dans un modèle de régression stratifié partiellement observé avec données censurées [MS-99]

17:00 Cecilia Cotton & Patrick Heagerty, *University of Washington* (E) [E]

Inference for Treatments Targeting Control of an Intermediate Measure • Inférence dans les traitements visant le contrôle d'une mesure intermédiaire [MS-100]

17:15 Liting Zhu, Xianming Tan & Dongsheng Tu, *NCIC Clinical Trial Group, Queen's University* (E) [E]

Testing the Homogeneity of Two Survival Functions against a Mixture Alternative Based on Censored Data • Essai de l'homogénéité de deux fonctions de survie comparé à un mélange alternatif fondé sur des données censorisées [MS-101]

17:30 Lin Liu & Florin Vaida, *University of California at San Diego* (E) [E]

Numeric EM for Mixed Effects Models with Censored Response • Algorithme EM numérique pour des modèles à effets mixtes avec réponse censurée [MS-102]

16:30 – 18:00 Session 05F — Capital Hall 6A

Applied Probability
Probabilités appliquées

Contributed Paper Session • Séance de communications libres

Chair • Présidente: Ayesha Ali, *University of Guelph*

16:30 Michael Rotondi, *University of Western Ontario* (E) [E]

Merry Christmas Markov Chains! • Joyeux Noël Chaînes de Markov ! [MS-103]

16:45 Vivien Rossi, *CIRAD*, Nicolas Desassis & Fabien Campillo, *INRIA* (E) [F]

Individual-based Modelling of Spatio-temporal of Forest Dynamics • Modélisation centrée sur l'individu des dynamiques spatio-temporelle forestières [MS-104]

17:00 Hsing-Ming Chang & James C. Fu, *University of Manitoba* (E) [E]

Priority Modelling of Patient Wait-Times by Using FMCI Technique for Benchmarking and Improving Canadian Health-care System • Modélisation prioritaire du temps d'attente du patient en utilisant la technique FMCI pour établir des données repères et améliorer le système de santé canadien [MS-105]

17:15 Hafedh Faires & Richard Emilion (E) [E]

Continuous Time Dirichlet Hierarchical Models • Les modèles hiérarchiques de Dirichlet continu dans le temps [MS-106]

17:30 Katherine Davies & John Braun, *University of Western Ontario* (E) [E]

Extensions of a Multi-Parameter Renewal Process • Extensions d'un processus de renouvellement multiparamétrique [MS-107]

16:30 – 18:00 Session 05G — Capital Hall 7A

Estimation and Tests for Time Series
Estimation et tests pour séries chronologiques

Contributed Paper Session • Séance de communications libres

Chair • Président: Jean-Michel Zakoian, *Université Lille 3*

16:30 Pierre Lafaye de Micheaux, *LJK, Grenoble Universités* (F) [E]

Neyman Smooth Test of Normality in a Time Series Context • Test lisse de Neyman dans un contexte de séries chronologiques [MS-108]

16:45 Azadeh Moghtaderi, David Thomson & Glen Takahara, *Queen's University* (E) [E]

Estimation of the Evolutionary Spectrum for Nonstationary Processes with Improved Boundary Behavior • Estimation du spectre évolutif pour processus non stationnaires avec comportement de frontière amélioré [MS-109]

17:00 Mireille Getler-Summa & Catherine Pardoux, *Université Paris-Dauphine, Ceremade, CNRS,* Christian Delavelle & David Fayolle, *Aji-Europe* (F) [E]

Modelling Time Series with Constraints: Looking for Similarities; Validation • Modélisation de chroniques sous contraintes : recherche de ressemblances, validation [MS-110]

17:15 Bei Chen & Yulia, R Gel, *University of Waterloo* (E) [E]

Regularized Autoregressive Frequency Estimation • Estimation de fréquences autorégressive régularisée [MS-111]

17:30 Anne Gégout-Petit, *INRIA & Université de Bordeaux 2*, Bernard Bercu, *INRIA & Université de Bordeaux 1,* Benoite de Saporta, *INRIA & Université de Bordeaux 4* (F) [E]

Asymptotical Analysis for Bifurcating Autoregressive Processes via a Martingale Approach • Analyse asymptotique des processus autoregressifs de bifurcation par des méthodes martingales [MS-112]

17:45 Scott Wile, *Dalhousie University* (E) [E]

Parameter Estimation for State Space Models using Sequential Monte Carlo Methods and State Augmentation • Méthodes de Monte Carlo séquentielles pour l'estimation des paramètres dans un modèle à espace d'états [MS-113]

18:00–19:00 NSERC Workshop: Tips to Prepare your Next Discovery Grant Application • Atelier du CRSNG: Conseils pour l'élaboration de votre prochaine demande de subvention à la découverte — Capital Hall 6A

Mardi 27 mai • Tuesday, May 27

08:45 – 10:00 Session 06A — Congress Hall A/E

SFdS Presidential Invited Address
Allocution de l'invité de la présidente de la SFdS

Special Session • Conférence spéciale

Organizer and Chair • Responsable et présidente: Christiane Guinot, *CERIES*

08:45 Davy Paindaveine, *Université Libre de Bruxelles*, Hannu Oja & Klaus Nordhausen, *University of Tampere*, Sara Taskinen, *University of Jyväskylä*

Invariant Methods for Independent Component Models • Méthodes invariantes pour les modèles à composantes indépendantes [MS-114]

10:00 – 10:30 Pause-café • Coffee Break

10:30 – 12:00 Session 07A — Capital Hall 1A

Data Integration for Genomic Data
Intégration de données sur le génome

Invited Paper Session • Présentations sur invitation

Session sponsored by the *Biostatistics Section*
Session parrainée par le *Groupe de biostatistique*

Organizer and Chair • Responsable et président: Laurent Briollais, *University of Toronto*

10:30 Rafal Kustra, *University of Toronto*, Adam Zagdanski, *Wroclaw University of Technology*

Genomic Data Fusion: Balancing Discovery and Interpretability • Fusion de données génomiques : équilibrer découvertes et interprétabilité [MS-115]

10:52 Joseph Beyene, Elena Parkhomenkho & David Tritchler, *University of Toronto*

Integrating High-Throughput Genomic Data Using Sparse Canonical Correlation Analysis • Intégration de données génomiques haut débit par analyse parcimonieuse des corrélations canoniques [MS-116]

11:14 Jinnan Liu, *Samuel Lunenfeld Research Institute, Mount Sinai Hospital, and York University*, Sohee Kang, *Samuel Lunenfeld Research Institute,Mount Sinai Hospital*, Helene Massam, *Samuel Lunenfeld Research Institute, Mount Sinai Hospital, and York University*, Laurent Briollais, *Samuel Lunenfeld Research Institute,Mount Sinai Hospital*

How to Incorporate Expert Biological Knowledge into Conjugate Priors for Discrete Graphical Models? • Comment incorporer les connaissances d'experts en biologie à des *a priori* conjuguées pour des modèles graphiques discrets ? [MS-117]

11:36 Gary Bader, *University of Toronto*

Collecting Functional Genomics Data for Association Studies and Gene Prioritization • Collecte de données génomiques fonctionnelles pour les études d'association et de priorisation des gènes [MS-118]

10:30 – 12:00 Session 07B — Capital Hall 2A

Session II in Memory of André Dabrowski
Séance II à la mémoire d'André Dabrowski

Invited Paper Session • Présentations sur invitation

Session sponsored by the *Probability Section*
Session parrainée par le *Groupe de probabilité*

Chair • Président: Herold Dehling, *Ruhr Universität, Bochum*

Organizer • Responsable: David McDonald, *Université d'Ottawa*

10:30 Rafal Kulik, Andre Dabrowski & Gail Ivanoff, *University of Ottawa*

Poisson Limits for Empirical Point Processes • Limite de Poisson pour les processus de point empiriques [MS-119]

11:00 Mahmoud Zarepour, *University of Ottawa*, Thierry Bedard, *Ontario Teachers' Pension Plan*, Andre Dabrowski, *University of Ottawa*

Return and Value at Risk Using the Dirichlet Process • Rendement et valeur à risque utilisant un processus de Dirichlet [MS-120]

11:30 David McDonald & Andre Dabrowski, *University of Ottawa*, Jiyeon Lee, *Yeungnam University*

Large Deviations of Multitype Queues • Grandes déviations de files d'attente multitype [MS-121]

10:30 – 12:00 Session 07C — Capital Hall 3A

History of Statistical Science
Histoire de la statistique

Invited Paper Session • Présentations sur invitation

Organizer and Chair • Responsable et président: David Bellhouse, *University of Western Ontario*

10:30 Antoine de Falguerolles, *Institut de Mathématiques, Laboratoire de Statistique et Probabilités, Université de Toulouse (III)*

Some Early Uses of the Least-squares Method. • Quelques emplois précoces de la méthode des moindres carrés [MS-122]

11:00 Michael Friendly, *York University*

The Golden Age of Statistical Graphics • L'âge d'or des graphiques statistiques [MS-123]

11:30 Adrian Rice, *Randolph-Macon College*, Eugene Seneta, *University of Sydney*

De Morgan in the Pre-history of Statistical Hypothesis Testing • De Morgan et la préhistoire des tests d'hypothèse statistique [MS-124]

10:30 – 12:00 Session 07D — Capital Hall 4A

Model Selection
Sélection de modèles

Invited Paper Session • Présentations sur invitation

Organizer and Chair • Responsable et président: Pascal Massart, *Université de Paris Sud-Orsay*

10:30 Aurélien Garivier, *CNRS, LTCI, Telecom ParisTech*, Florencia Leonardi, *Universidade de São Paulo* F E

Context Tree Estimation via Algorithm Context and Penalized Maximum Likelihood • Identification d'arbres de contexte : l'algorithme "context" et le maximum de vraisemblance pénalisé [MS-125]

11:00 Claire Lacour, *Université Paris Descartes*, Elodie Brunel, *Université de Montpellier*, Fabienne Comte, *Université Paris Descartes* F E

Adaptive Estimation of the Conditional Distribution Function in the Presence of Censoring • Estimation adaptative de la fonction de répartition conditionelle en présence de censure [MS-126]

11:30 Vincent Rivoirard, *Université Paris Sud et ENS Paris*, Patricia Reynaud-Bouret, *CNRS - ENS Paris* F E

Thresholding Estimation of a Poisson Intensity with Unknown or Infinite Support • Estimation par seuillage de l'intensité d'un processus de Poisson à support inconnu ou infini [MS-127]

10:30 – 12:00 Session 07E — Capital Hall 5A

Estimation Methods for Surveys Data
Méthodes d'estimation pour données de sondages complexes

Contributed Paper Session • Séance de communications libres

Session sponsored by the *Survey Methods Section*
Session parrainée par le *Groupe de méthodologie d'enquête*

Chair • Président: Jean-Claude Deville, *École Nationale de la Statistique et de l'Analyse de l'Information*

10:30 Simon Bonner & Carl Schwarz, *Simon Fraser University* E E

Hierarchical Bayesian Modelling of Two-Stage Capture-Recapture Experiments • Modélisation bayésienne hiérarchique des expérimentations à deux étapes de la capture et de la recapture [MS-128]

10:45 Ivan Carrillo & Changbao Wu, *University of Waterloo* E E

GEE Analysis of Longitudinal Surveys with Imputation for Missing Responses • Analyse par EEG de données d'enquêtes longitudinales avec imputation pour les réponses manquantes [MS-129]

11:00 Zilin Wang, *Wilfrid Laurier University*, Mary Thompson, *University Of Waterloo* E E

A Resampling Approach to Estimate Variance Components of Multilevel Models • Une approche de rééchantillonnage pour estimer les composantes de la variance de modèles multiniveaux [MS-130]

11:15 Olivier Sautory, *CEPE-INSEE*, Jean-Claude Deville, *ENSAI-CREST* F F

Calibration for Surveys using Uncertain Auxiliary Information • Calage d'une enquête sur une information auxiliaire incertaine [MS-131]

11:30 Yves Tillé & Desislava Nedyalkova, *Université de Neuchâtel* F E

Optimal Sampling and Estimation Strategies under Linear Model • Stratégies d'échantillonnage et d'estimation optimales sous un modèle linéaire [MS-132]

11:45 Abdellatif Demnati, *Statistique Canada*, J.N.K. Rao, *Carleton University* F E

Linearized Variance Estimation from Simulated Census Data • Estimation de la variance linéarisée à l'aide de données de recensement simulées [MS-133]

10:30 – 12:00 Session 07F — Capital Hall 6A

Semiparametric and Nonparametric Models
Modèles semi-paramétriques et non paramétriques

Contributed Paper Session • Séance de communications libres

Chair • Président: Denis Bosq, *Université de Paris VI*

10:30 Jureckova Jana, Jan Kalina & Marek Omelka, *Charles University in Prague* E E

Two-Sample Rank Tests for Nonnegative Observations with Multivarite Applications • Tests à deux échantillons pour des observations non-négatives avec des applications multivariées [MS-134]

10:45 Jérôme Saracco, *Université Bordeaux 4*, Benoit Liquet, *ISPED, Bordeaux*, Marie Chavent, *Université Bordeaux 2* F E

Two-step Estimation in a Multivariate Semiparametric Sample Selection Model • Estimation en deux étapes dans un modèle de sélection semi paramétrique multivarié [MS-135]

11:00 Stéphane Gaiffas & Agathe Guilloux, *LSTA - Université Paris 6* F E

Adaptive Estimation by Aggregation for Marked Counting Processes • Estimation adaptative par agrégation pour des processus de comptage marqués [MS-136]

11:15 Geenens Gery &, Léopold Simar, *Institut de Statistique, Université catholique de Louvain* F E

Non- and Semiparametric Tests for Conditional Independence in Two-way Contingency Tables • Tests non-paramétrique et semi-paramétriques d'indépendance conditionnelle dans une table de contingence à deux entrées [MS-137]

11:30 Bernard Colin & Ernest Monga, *Université de Sherbrooke* F E

Semiparametric Test of Indepedence between Random Vectors • Test d'indépendance semi-paramétrique entre vecteurs aléatoires [MS-138]

11:45 Agathe Guilloux & Stéphane Gaïffas, *LSTA- Université Pierre et Marie Curie*, Fabienne Comte, *MAP5 -Université René Descartes* F E

Adaptive Estimation for Marked Counting Processes • Estimation adaptative pour des processus de comptage marqués [MS-139]

10:30 – 12:00 Session 07G — Capital Hall 7A

Classification Methods and Data Mining
Méthodes de classification et forage de données

Contributed Paper Session • Séance de communications libres

Chair • Président: Denis Larocque, *HEC-Montréal*

10:30 Abbas Khalili & Jiahua Chen, *University of British Columbia*, Shili Lin, *Ohio State University* E E

Learning Sparse Finite Mixture of Regression Models in High Dimensional Feature Space • Mélange fini parcimonieux de modèles de régression pour l'apprentissage dans un espace de caractéristiques de grande dimension [MS-140]

10:45 Wu Zhou & Wayne Oldford, *University of Waterloo* E E

Defining a Cluster Tree for a Multivariate Discrete Distribution • Définir un arbre de classification pour une distribution discrète multivariée [MS-141]

11:00 Arnaud Guyader, *Université de Rennes 2*, Gérard Biau, *Université de Paris 6* & Frédéric Cérou, *IRISA/INRIA, Rennes* (F) (E)

Kernel Methods and Nearest Neighbor Classification • Méthodes à noyaux et plus proches voisins [MS-142]

11:15 Abdelaziz El Matouat, *Université du Havre*, Hassania Hamzaoui, *Université de Meknès, Maroc*, Lahcen Kassri, *Université de Fès, Maroc*, Patrick Martin, *LaMIPS, NXP, Ensicaen* (F) (F)

Color Image Segmentation by 2*D* Histogram Classification and Information Criteria • Segmentation d'une image couleur par la classification des histogrammes 2*D* et les critères d'information [MS-143]

11:30 Abdessamad Dine, Denis Larocque & François Bellavance, *HÉC Montréal* (E) (E)

Multivariate Mixed Decision Trees • Arbres de décisions mixtes multivariés [MS-144]

10:30 – 12:00 Session 07H — Capital Hall 8A

Financial Engineering
Ingénierie financière

Contributed Paper Session • Séance de communications libres

Session sponsored by the *Probability Section*
Session parrainée par le *Groupe de probabilité*

Chair • Président: Bruno Rémillard, *HEC-Montréal*

10:30 René Ferland, *Université du Québec à Montréal*, Gaston Giroux (F) (E)

Law of Large Numbers for Dynamic Bargaining Markets • Loi des grands nombres pour des marchés dynamiques avec négociation [MS-145]

10:45 Francois Watier & René Ferland, *Université du Québec à Montréal* (E) (E)

Mean-variance Efficiency in a Market with Zero-coupons • Stratégie moyenne-variance optimale dans un marché avec obligations à escompte [MS-146]

11:00 Romuald Momeya, *Université de Montréal* (F) (F)

Regime Switching Models with Lévy Process • Les modèles à régime variable sous des processus de Lévy [MS-147]

11:15 Adam Metzler & Don L. McLeish, *University of Waterloo* (E) (E)

Calibrating Multivariate First-Passage Models for Credit Risk • Calibrer des modèles de premier passage multivariés pour le risque de crédit [MS-148]

10:30 – 12:00 Session 07I — Capital Hall 9A

Sensitivity Analysis and Fiability in Business and Industrial Statistics
Analyse de sensibilité et fiabilité en statistique industrielle et de gestion

Contributed Paper Session • Séance de communications libres

Session sponsored by the *Business and Industrial Statistics Section*
Session parrainée par le *Groupe de statistique industrielle et de gestion*

Chair • Président: Smiley Cheng, *University of Manitoba*

10:30 Amandine Marrel, *CEA Cadarache*, Béatrice Laurent, *Institut de Mathématiques de Toulouse, INSA, Université de Toulouse*, Bertrand Iooss, *CEA Cadarache* (F) (E)

Use of Gaussian Process for Sensitivity Analysis, Application to a Spatial Output of Computer

Code • Utilisation des processus Gaussiens pour l'analyse de sensibilité : mise en oeuvre sur une sortie spatiale d'un code de calcul [MS-149]

10:45 Gérard D'AUBIGNY, *Département de Statistique, MS*^{3}*-LJK, Université de Grenoble,*, Lambert PIERRAT, *LJ-Consulting, Saint-Martin d'Hères* (F) E

An Heuristic Approximation of the Sampling Distribution of the Coefficient of Variation • Une approximation heuristique de la loi d'échantillonnage du coefficient de variation [MS-150]

11:00 Luan JAUPI, *CNAM-Paris* (E) F

Multivariate SPC Methods to Monitor Complex Product-Process Parameters • Méthodes multivariées pour la maîtrise de processus de fabrication complexes [MS-151]

11:15 Vincent COUALLIER & Léo GERVILLE-RÉACHE, *Institut Mathématique de Bordeaux, Université Bordeaux 2* (F) E

Optimization of the Total Test Time for Global Reliability Demonstration Design in a Competing Failure Modes Model. • Allocation optimale du temps de tests pour une démonstration de fiabilité globale dans un cadre de modes multiples de défaillance [MS-152]

11:30 Eduard IBANEZ TARRADELLAS, *TNS WORLDPANEL* (F) E

From ABC Model to DEF Model • Modèle DEF : une modélisation statistique complémentaire du modèle ABC [MS-153]

12:00 – 13:30 Dîner • Lunch

13:30 – 14:30 Session 08A — Congress Hall A/E

SSC Gold Medal Award Address
Allocution du récipiendaire de la médaille d'or de la SSC

Special Session • Conférence spéciale

Organizer and Chair • Responsable et présidente: Nancy REID, *University of Toronto*

13:30 Don MCLEISH, *University of Waterloo* (E) E

From Martingales to Monte Carlo: A Personal Random Walk • Des Martingales à Monte Carlo : une marche aléatoire personnelle [MS-154]

14:35 – 15:35 Session 09A — Capital Hall A1

Special Invited Session of the Survey Methods Section
Allocution de l'invité d'honneur du groupe de méthodologie d'enquête

Special Session • Conférence spéciale

Session sponsored by the *Survey Methods Section*
Session parrainée par le *Groupe de méthodologie d'enquête*

Organizer and Chair • Responsable et président: Jiahua CHEN, *University of British Columbia*

14:35 David BINDER, *Statistics Canada (Retired)* (E) F

How to Analyze Survey Data Badly: Answers to Criticisms of Using Design-Based Methods for Inference • Comment mal analyser des données d'enquête : réponse aux critiques relatives à l'utilisation des méthodes fondées sur le plan de sondage pour l'inférence [MS-155]

14:35 – 16:05 Session 09B — Capital Hall 2A

Statistical Process Control Methods in Health Care
Méthodes statistiques de contrôle des processus en santé

Invited Paper Session • Présentations sur invitation

Session jointly sponsored by the *Business and Industrial Statistics Section* and the *Biostatistics Section*
Session parrainée conjointement par le *Groupe de statistique industrielle et de gestion* et le *Groupe de biostatistique*

Chair • Président: Richard Cook, *University of Waterloo*

Organizers • Responsables: Richard Cook & Stefan Steiner, *University of Waterloo*

14:35 Vern Farewell, *Medical Research Council Biostatistics Unit, UK* E E

Performance Monitoring in the Public Services • Contrôle du rendement dans les services publics [MS-156]

15:05 Charlie Goldsmith, *McMaster University* E E

Statistical Process Control Methods in Health Care • Méthodes de contrôle du processus statistique dans les soins de santé [MS-157]

15:35 William Woodall, *Virginia Tech*, Olivia Grigg, *Cambridge University*, Howard Burkom, *Johns Hopkins University* E E

Research Issues and Ideas in Health-Related Surveillance • Questions de recherche et idées relatives à la surveillance en matière de santé [MS-158]

14:35 – 16:05 Session 09C — Capital Hall 3A

Statistical Analysis of Hydroclimatic Proxy Data: Frequentist Approaches
Analyse statistique des empreintes hydro-climatiques : approches fréquentielles

Invited Paper Session • Présentations sur invitation

Session sponsored by / Session parrainée par
Groupe environnement et statistique de la SFdS

Organizer and Chair • Responsable et président: Éric Parent, *ENGREF*

14:35 Liliane Bel, *AgroParisTech*, Avner Bar-Hen, *Université Paris 5*, Rachid Cheddadi, *Institut des Sciences de l'Evolution* F E

Climate Reconstruction from Pollens: Climate Impact on the Spatial Beech Genetic Diversity • Reconstitution climatique à partir de pollens : étude de l'impact du climat sur la répartition spatiale de la diversité génétique des hêtres [MS-159]

15:05 Philippe Naveau, *LSCE-CNRS-IPSL*, Caspar Ammann, *NCAR, Colorado*, Deborah Batista, *Colorado University, Denver*, Julie Carreau, *LSCE-CNRS-IPSL, France* E E

Reconstructing Past Volcanic Forcings with a Multivariate Multi-State Space Model • Reconstruire le forçage volcanique à partir d'un modèle d'états multivarié [MS-160]

15:35 Francis Zwiers, *Environment Canada, Climate Research Division*, Min Tsao & Terry Lee, *University of Victoria* E E

Evaluation of Proxy-Based Millennial Reconstruction Methods • Évaluation de méthodes de reconstruction millénaire fondées sur des données indirectes [MS-161]

14:35 – 16:05 Session 09D — Capital Hall 4A

How to Reconcile the Demands for Data Confidentiality and Research Needs? The Role of Secure Access Centers.
Comment concilier les exigences de la confidentialité des données et les besoins de la recherche ? Le rôle des centres d'accès sécurisés.

Invited Paper Session • Présentations sur invitation

Session sponsored by / Session parrainée par
Groupe statistique et société de la SFdS

Chair • Président: Jean-Louis Bodin, *ADETEF-France*

Organizer • Responsable: Jean-François Royer, *CREST-INSEE*

14:35 Roxane Silberman, *Directrice de recherche au CNRS/Secrétaire générale du Comité de concertation pour les données en sciences humaines et sociales* (E) (F)

Processing Confidential Data and the Methods of Protection: The researchers perspective • Le traitement des données confidentielles et les dispositifs de protection : le point de vue des chercheurs. [MS-162]

15:05 Gustave Goldmann, *Statistique Canada* (F) (E)

The RDC Network: From a Seed to a Forest • Le réseau des CDR : d'une graine à la forêt [MS-163]

15:35 Michel Isnard, *INSEE* (F) (E)

Insee's Secure Access Centers : A Way to the Future • Les centres d'accès sécurisés à l'Insee : projets en cours [MS-164]

14:35 – 16:05 Session 09E — Capital Hall 5A

Classification and Algorithms
Classification et algorithmes

Invited Paper Session • Présentations sur invitation

Organizer and Chair • Responsable et président: Gérard Biau, *Université de Paris VI*

14:35 Luc Devroye, *McGill University* (F) (E)

Hierarchical Methods for Multidimensional Data: Convergence, Complexity and Learning • Méthodes hiérarchiques pour la classification multidimensionnelle : convergence, complexité, apprentissage [MS-165]

15:05 Godfried Toussaint, *McGill University*, Binay Bhattacharya, *Simon Fraser University* (E) (E)

Instance-Based Learning with Proximity Graphs • Apprentissage basé sur les occurences avec graphes de proximité [MS-166]

15:35 Nicolas Vayatis, *ENS de Cachan*, Stéphan Clémençon, *Telecom-ParisTech* (F) (E)

Optimization Principles for the Ranking Problem • Principes d'optimisation pour le problème de l'ordonnancement [MS-167]

14:35 – 16:05 Session 09F — Capital Hall 6A

Robustness
Robustesse

Invited Paper Session • Présentations sur invitation

Chair • Présidente: Julie Zhou, *University of Victoria*

Organizer • Responsable: Doug Wiens, *University of Alberta*

14:35 Marc Genton, *Université de Genève and Texas A&M University,* Adelchi Azzalini, *University of Padua* (F) E

Robust Likelihood Methods Based on the Skew-*t* and Related Distributions • Méthodes de vraisemblance robustes fondées sur des distributions *t* asymétriques et autres distributions de la même famille [MS-168]

15:05 Jean-Claude Massé, *Université Laval* (F) E

Robust Efficient Multivariate Trimmed Means • Moyennes tronquées multivariées robustes et efficaces [MS-169]

15:35 Matias Salibian-Barrera, *The University of British Columbia,* Ying Wei, *Columbia University* (E) F

Robust Quantile Regression with Non-Elliptical Covariates • Régression quantile robuste avec covariables non elliptiques [MS-170]

14:35 – 16:05 Session 09G — Capital Hall 7A

Analysis of Dependent Data
Analyse de données dépendantes

Contributed Paper Session • Séance de communications libres

Session sponsored by the *Biostatistics Section*
Session parrainée par le *Groupe de biostatistique*

Chair • Président: Bruno Rémillard, *HÉC Montréal*

14:35 Chloé Friguet, Maéla Kloareg & David Causeur, *Mathématiques Appliquées - Agrocampus Rennes* (E) E

Multiple Tests for High-throughput Data assuming a Factor Modelling of Dependence • Modèle en facteurs pour les procédures de tests multiples en grande dimension [MS-171]

14:50 Ji (Emmy) Cheng, *St. Joseph's Healthcare Hamilton,* Lehana Thabane & Donald Willison, *McMaster University,* Co-investigators, *McMaster University, York University, University of Ottawa* (E) E

Analysis of Correlated and Clustered Data Using a Survey on Privacy and Health Research in Stigmatized Populations as an Example • Analyse des données corrélées et regroupées en utilisant une enquête sur la vie privée et la recherche sur la santé chez les populations stigmatisées à titre d'exemple [MS-172]

15:05 Lisa Lix & Phongsack Manivong, *University of Manitoba* (E) E

Testing Multiple Correlated Outcomes in Clinical Trials with Repeated Measurements: Comparing Adjusted p-Value Approaches • Essai des résultats multi-corrélés dans les essais cliniques ayant des mesures répétées : comparaison des approches de la valeur p ajustée [MS-173]

15:20 Lilia Ramirez Ramirez & Mary Thompson, *University of Waterloo* (E) E

Outbreaks in Populations with Contact Network Structure • Épidémie dans les populations ayant une structure en réseau des contacts [MS-174]

15:35 David Tolusso & Richard Cook, *University of Waterloo* (E) E

A Generalized Estimation Equations Approach to Current Status Data with Dependent Responses • Une approche par équations d'estimation généralisées des données sur l'état courant avec réponses dépendantes [MS-175]

15:50 Ana-Maria Staicu, *University of Bristol*, Ciprian M. Crainiceanu, *Johns Hopkins University*, Raymond J. Carroll, *Texas A&M University* E E

An Approach to the Analysis of Spatially Correlated Multilevel Functional Data • Une approche à l'analyse de données fonctionnelles à plusieurs niveaux corrélées spatialement [MS-176]

14:35 – 16:05 Session 09H — Capital Hall 8A

Biostatistical Models
Modèles biostatistiques

Contributed Paper Session • Séance de communications libres

Session sponsored by the *Biostatistics Section*
Session parrainée par le *Groupe de biostatistique*

Chair • Président: Paul Gustafson, *University of British Columbia*

14:35 Peng Zhang, *University of Alberta*, Zhenguo Qiu, *Alberta Cancer Board*, Peter X.-K. Song, *University of Michigan* E E

Robust Analysis of Longitudinal Proportional Data Using Mixed-Effects Models • Analyse robuste de données longitudinales proportionnelles à l'aide de modèles à effets mixtes [MS-177]

14:50 Karen Leffondré, *Université de Montréal*, Willy Wynant, *Centre de recherche du CHUM*, Zhirong Cao, *Université de Montréal*, Michal Abrahamowicz, *McGill University* E F

The Cox Model for Epidemiological Case-control Data: a Simulation Study • Le modèle de Cox pour les données épidémiologiques cas-témoins : une étude de simulation [MS-178]

15:05 Kenneth Lo, *University of British Columbia*, Ryan Brinkman, *Terry Fox Laboratory, BC Cancer Research Center*, Raphael Gottardo, *University of British Columbia* E E

Model-Based Clustering using *t* Mixtures with Box-Cox Transformation • Classification à base de modèles utilisant des mélanges de loi *t* avec transformation de Box-Cox [MS-179]

15:20 Benjamin Rich, Alina Dyachenko & Antonio Ciampi, *McGill University* E E

Soft Node Trees for Predicting Continous and Multinomial Variables • Arbres à noeuds mous pour prévoir des variables continues et multinomiales [MS-180]

15:35 Erica Moodie, *McGill University*, Thomas Richardson, *University of Washington* E E

Bias Reduction for g-Estimation of Optimal Dynamic Regimes at Exceptional Laws • Réduction de biais pour g-estimation de régimes optiques dynamiques à lois exceptionelles [MS-181]

15:50 Dianliang Deng, *University of Regina*, Sudhir R. Paul, *University of Windsor* E F

Assessing Goodness of Fit of Generalized Linear Models to Sparse Data using Higher Order Moment Corrections • Évaluer la qualité de l'ajustement de modèles linéaires généralisés pour les données dispersées en utilisant des corrections plus élevées de l'ordre du moment [MS-182]

14:35 – 16:05 Session 09I — Capital Hall 9A

Probability Laws and Asymptotic Properties
Lois de probabilité et propriétés asymptotiques

Contributed Paper Session • Séance de communications libres

Session sponsored by the *Probability Section*
Session parrainée par le *Groupe de probabilité*

Chair • Présidente: Delphine Blanke, *Université de Paris VI*

14:35 Efoevi Koudou & Pierre Vallois, *Institut Elie Cartan, Nancy, France* (F) (E)

A Class of Matsumoto-Yor Functions • Une classe de fonctions de Matsumoto-Yor [MS-183]

14:50 Ludovic Menneteau, *Université Montpellier II* (F) (E)

MDP and LIL for the Empirical Covariance of Linear Hilbertian Processes • PDM et LLI pour la covariance de processus linéaires hilbertiens [MS-184]

15:05 Lorenzo Molkorogov, *Université inter-age de Morbier.* (F) (F)

Asymptotic Universal Theorem • Un théorème asymptotique universel [MS-185]

15:20 Adel Settati, *IRMAR - Université Rennes 1* (F) (E)

Strong Approximation of the Function Indexed Empirical Process Under Random Entropy Condition. • Approximation forte du processus empirique indexé par des fonctions sous condition d'entropie aléatoire. [MS-186]

15:35 Aziz Khanchi, *University of Ottawa* (E) (E)

A Ratio Limit Theorem for a Markov Kernel with Unbounded Jumps • Un théorème limite du rapport pour un noyau markovien à sauts non bornés [MS-187]

14:35 – 16:05 Session 09J — Congress Hall A/E West

Poster Session II
Séance d'affichage II

Contributed Paper Session • Séance de communications libres

14:35 Qing Wu, *Mayo Clinic Arizona*, Jeanette Magnus & Bencaz Angelle, *Tulane University* (F)

Depression and Bone Mineral Density: A Meta-analysis of Epidemiological Studies • La dépression et la teneure minérale de l'os : une méta-analyse des études épidémiologiques [MS-188]

14:35 Qi Er Tan, James Fu & Xikui Wang, *University of Manitoba* (E)

Waiting Time for Health Services • Temps d'attente pour des services de santé [MS-189]

14:35 Shuying Sun, Shuying Sun & Dustin Potter, *The Ohio State University* (E)

'Quantile' is a Better Choice when 'Average' is Not Effective in Methylation Microarray Analysis • Un 'quantile' est un meilleur choix lorsque la 'moyenne' n'est pas efficace pour l'analyse de micropuce de méthylation [MS-190]

14:35 Eunhee Kim, *National Institute of Toxicological Research*, Gun-Young Lee, *Korea Fooda and Drug Administration*, Munsin Kim, *National Institute of Toxicological Research*, Ki-Sung Kwon, *Korea Fooda and Drug Administration*, Dongdeuk Jang & Bong-Hyun Nam, *National Institute of Toxicological Research* (E)

Predictive Models for the Growth of Microbial Pathogens • Modèles prédictifs pour la croissance de pathogènes microbiens [MS-191]

14:35 Mohamed Abdolell & Jennifer Payne, *Dalhousie University*, Judy Caines, *Nova Scotia Breast Screening Program*, Wendy Lou, *University of Toronto*, Sam Stewart, *University of Waterloo* (E)

Automated Public Health Surveillance (PHS) Systems: An Open-source Implementation • Systèmes automatisés de surveillance de la santé publique (SSP) : une implémentation en code source ouvert [MS-192]

14:35 Geraldine Lo Siou, Ilona Csizmadi, S. Elizabeth McGregor, Karen A. Kopciuk & Paula J. Robson, *Alberta Cancer Board* (E)

Use of Cluster Analyses to Assess Dietary Patterns: a Comparison of Unsupervised Clustering

Methods. • L'utilisation des méthodes de classification pour identifier des typologies alimentaires : une comparaison des techniques de classification non-supervisées [MS-193]

14:35 Yanqing Yi & Xikui Wang, *University of Manitoba* E

Response Adaptive Designs with a Variance-penalized Criterion • Plans de réponses adaptées ayant un critère de variance pénalisée [MS-194]

14:35 Kyle Vincent, *Simon Fraser University* E

Advantages of Adaptive Sampling Designs for Sampling from a Network Setting • Avantages des plans d'échantillonnage adaptatifs pour l'échantillonnage à partir de l'établissement de réseaux [MS-195]

14:35 Dena Schanzer, *Public Health Agency of Canada*, Trevor Dummer & Joanne Langley, *Dalhousie University*, Samina Aziz, Brian Winchester & Theresa Tam, *Public Health Agency of Canada* E

Geo-temporal Spread of Influenza A, at Different Geographic Scales, Canada and United States • La propagation de la grippe à différentes échelles géographiques au Canada et aux États-Unis [MS-196]

14:35 Judy Anne Chapman, Patti O'Brien, Zhangjun Ding, Lois Shepherd, *NCIC Clinical Trials Group Queen's University*, Margot Burnell, *Atlantic Health sciences Corporation*, Mark Levine, *McMaster University*, Kathleen Pritchard, *Sunnybrook Odette Cancer Centre Sunnybrook Health Sciences Centre*, David Matthews, *University of Waterloo*, Chunsong Qu, *NCIC Clinical Trials Group Queen's University* E

Estimating Efficacy in Clinical Trials: Case Study of NCIC CTG MA.21 Adjuvant Breast Cancer Trial • Estimation de l'efficacité dans les essais cliniques : étude de cas de l'essai de traitement adjuvant du cancer du sein NCIC CTG MA.21 [MS-197]

14:35 Robin Burkhalter, Grace Y. Yi & K. Stephen Brown, *University of Waterloo* E

Analysis of Incomplete Longitudinal Data: an Application to Self-Reported Smoking Status among Adolescents from the Waterloo Smoking Prevention Project • Analyse des données longitudinales incomplètes : une application du statut de fumeur autodéclaré parmi les adolescents à partir du projet de prévention contre le tabac de Waterloo [MS-198]

14:35 Nataliya Dragieva, *Université du Québec à Montréal* F

Confidence Interval and Permutation Test in the Context of MapArg Method • Intervalle de confiance et test de permutation dans le contexte de la méthode MapArg [MS-199]

14:35 Hedy Jiang & Patrick Brown, *Cancer Care Ontario* E

Power Calculation for Detecting Gene-environment Interaction in the Ontario Cancer Cohort • Calcul de la puissance de détection des interactions gènes et environnement dans la *Ontario Cancer Cohort* [MS-200]

14:35 Ruochu Gao, Neil Klar & Allan Donner, *University of Western Ontario* E

Extension of Cochran-Armitage Test for Analysis of Correlated Ordinal Data • Extension du test de Cochran-Armitage pour l'analyse de données ordinales corrélées [MS-201]

14:35 Mohamed Abdolell, *Dalhousie University*, Michael LeBlanc, *Fred Hutchinson Cancer Research Center*, Sam Stewart, *QEII Health Sciences Centre* E

Binary Recursive Partitioning for Continuous Longitudinal Outcomes • Partitionnement binaire récursif pour des résultats continus longitudinaux [MS-202]

14:35 Françoise Tondu, *Danone* E

A Different Approach to Assessing Departure from Normality with the Z-Score • Une approche différente pour juger de l'écart à la nomalité avec le Z-score [MS-203]

14:35 Vaneeta Grover, *Dalhousie University*, David Cole, *University of Toronto*, David Hamilton, *Dalhousie University* E

Likelihood Ratio Test for Association in Discordant Sib-pairs • Test du rapport de vraisemblance pour l'association dans les paires de frères et soeurs discordantes [MS-204]

14:35 Mohammad Islam & Angus McLeod, *University of Western Ontario* E

Improved Class Prediction in Gene Expression Microarray Data • Prédiction de classe améliorée dans l'expression des gènes des données de microréseaux [MS-205]

14:35 Giovani Silva, Cindy Xin Feng & C.B. Dean, *Simon Fraser University* E

Hotspot Detection in Disease Mapping • Détection de points chauds en cartographie des maladies [MS-206]

16:05 – 16:30 Pause-café • Coffee Break

16:30 – 18:00 Session 10A — Capital Hall 1A

Some New Perspectives on Longitudinal Data Analysis
Nouvelles perspectives en analyse de données longitudinales

Invited Paper Session • Présentations sur invitation

Chair • Président: Peng Zhang, *University of Alberta*

Organizer • Responsable: Peter Song, *University of Michigan*

16:30 Michel Chavance, *INSERM, Université Paris XI* E F

Detecting the Omission of a Time-dependent Confounder in Longitudinal Models • Identifier l'oubli d'une variable chronologique dans un modèle longitudinal [MS-207]

17:00 Daniel Commenges, *INSERM* F E

The Stochastic Process Framework for Causal Inference • Inférence causale basée sur les processus stochastiques [MS-208]

17:30 Peter Song, *University of Michigan*, Mingyao Li, *University of Pennsylvania*, Ying Yuan, *University of Texas MD Anderson Cancer Center* E E

Joint Regression Analysis of Correlated Data Using Gaussian Copulas • Analyse de régression conjointe de données corrélées au moyen de copules gaussiennes [MS-209]

16:30 – 18:00 Session 10B — Capital Hall 2A

Session in Honour of Denis Bosq
Session en hommage à Denis Bosq

Invited Paper Session • Présentations sur invitation

Organizer and Chair • Responsable et président: Michel Delecroix, *ISUP-Université de Paris VI*

16:30 Gérard Biau, *Université Paris VI*, Kevin Bleakley, *Université Montpellier II*, Laszlo Györfi & György Ottucsak, *Budapest University of Technology and Economics* F E

Nonparametric Sequential Prediction of Time Series • Prévision séquentielle non-paramétrique des séries temporelles [MS-210]

17:00 Delphine Blanke, *Université Pierre et Marie Curie - Paris 6* F E

Sampling Schemes for Regression Estimation and Applications • Plans d'échantillonnages pour l'estimation de la régression et applications [MS-211]

17:30 Serge Guillas, *University College London*, Ming-Jun Lai, *University of Georgia* F E

Approximation of Functional Regression Models using Bivariate Splines • Approximation de régression linéaire fonctionnelle avec splines bivariées [MS-212]

16:30 – 18:00 Session 10C — Capital Hall 3A

Longitudinal Data
Données longitudinales

Contributed Paper Session • Séance de communications libres

Session sponsored by the *Biostatistics Section*
Session parrainée par le *Groupe de biostatistique*

Chair • Président: Avner Bar-Hen, *Université Paris 5*

16:30 Xiaoqin Xiong & Joel Dubin, *University of Waterloo* E E

Simulation of Mixed Longitudinal Data Measured at Distinct Timepoints • Simulation de données longitudinales mixtes mesurées à des points distincts dans le temps [MS-213]

16:45 Eleanor Pullenayegum, *McMaster University*, Richard Cook, *Waterloo University* E E

Selection Biases in Repeated Measures Analysis of Treatment Effects for Recurring Clinical Events • Biais de sélection dans les analyses de mesures répétées des effets des traitements pour les évènements cliniques récurrents [MS-214]

17:00 Cyntha Struthers & Don McLeish, *University of Waterloo* E E

Modeling Longitudinal Data with Application to the Multicenter Aids Cohort Study • Modélisation de données longitudinales avec application à la Multicenter Aids Cohort Study [MS-215]

17:15 Paul McNicholas, *University of Guelph*, Thomas Brendan Murphy, *University College Dublin* E F

Model-Based Clustering of Longitudinal Data • Classification à base de modèles de données longitudinales [MS-216]

17:30 Bezza Hafidi, *Université Ibn Zohr, Agadir, Maroc*, Nourddine Azzaoui, *Université de Lille 1* F E

Criteria for Longitudinal Data Model Selection Based on Kullback's Symmetric Divergence • Critères de sélection de modèles de données longitudinales fondés sur la divergence symétrique de Kullback [MS-217]

16:30 – 18:00 Session 10D — Capital Hall 4A

Regression Methods in Biostatistics
Méthodes de régression en biostatistique

Contributed Paper Session • Séance de communications libres

Session sponsored by the *Biostatistics Section*
Session parrainée par le *Groupe de biostatistique*

Chair • Président: Jérôme Saracco, *Université de Bordeaux I*

16:30 Li Hua Yue, Duncan Murdoch, Wenqing He & Hristo S. Sendov, *University of Western Ontario* E E

Variable Selection Using Branching LARS • Sélection des variables par une méthode LARS ramifiée [MS-218]

16:45 Binod Neupane, Lehana Thabane & Mark Loeb, *McMaster University*

Reducing Bias in Observational Studies using Propensity Score Methods • Réduire le biais d'études observationnelles par des méthodes de scores de propension [MS-219]

17:00 Bei Jiang & K. C. Carriere, *University of Alberta*

Estimation for Age-Period-Cohort Models: With Application to the Mesothelioma Data in Alberta • Estimation pour modèles âge-période-cohorte : application aux données de Mésothéliome en Alberta [MS-220]

17:15 Jennifer Liu, Lehana Thabane, Emmy Cheng & Roman Jaeschke, *McMaster University*

Approaches in Dealing with Heterogeneity in Meta-analysis Studies • Approches pour composer avec l'hétérogénéité dans les études de méta-analyse [MS-221]

17:30 Shahedul Khan & Grace Chiu, *University of Waterloo*

Longitudinal Bent-Cables for Monitoring Atmospheric CFC Decline and Other Change Models • Cordes pliées longitudinales pour surveiller la baisse des CFC atmosphériques et autres modèles de changement [MS-222]

17:45 Nicolas Molinari, *Université Montpellier 1 - CHU de Nîmes*

Supervised Classification and Free Knot Splines for Biomedical Data • Classification supervisée et spline à noeuds variables pour des données biomédicale [MS-223]

16:30 – 18:00 Session 10E — Capital Hall 5A

Copulas and Applications
Copules et applications

Contributed Paper Session • Séance de communications libres

Chair • Président: Jean-François Quessy, *Université du Québec à Trois-Rivières*

16:30 Taoufik Bouezmarni & Jeroen Rombouts, *HEC-Montréal*

Nonparametric Tests for Conditional Independence via Copula • Tests d'indépendance conditionelle non paramétriques au moyen de copules [MS-224]

16:45 Aristidis Nikoloulopoulos, *University of British Columbia*, Mathieu Fortin, *Ministère des ressources naturelles et de la faune du Québec*, Christian Genest & Louis-Paul Rivest, *Université Laval*

Predicting Dependent Binary Outcomes through Logistic Regressions and Copulas • Prévision de variables binaires dépendantes à l'aide de régressions logistiques et de copules [MS-225]

17:00 Eric Parent, *Équipe Modélisation, Risque, Statistique, Environnement (MORSE), ENGREF, AgroParisTech, Paris*, Anne-Catherine Favre, *INRS-ETE, Québec*

Bayesian Inference of a Copula Model for a Bivariate Hydrological Analysis • Estimation bayésienne d'un modèle de copule pour une analyse hydrologique fréquentielle bivariée [MS-226]

17:15 Simon Guillotte & François Perron, *Université de Montréal*

A Bayesian Estimator for the Dependence Function of a Bivariate Extreme Value Distribution • Un estimateur bayésien pour la fonction de dépendance d'une loi bivariée de valeurs extrêmes [MS-227]

17:30 Yildiz Yilmaz, *University of Waterloo*

Goodness-of-Fit Tests for Copula Models of Bivariate Lifetime Data • Tests d'ajustement pour modèles de copules de données de survie bivariées [MS-228]

16:30 – 18:00 Session 10F — Capital Hall 6A

Limiting Behaviour of Test Statistics
Comportement asymptotique de statistiques de test

Contributed Paper Session • Séance de communications libres

Chair • Présidente: Nancy Reid, *University of Toronto*

16:30 Laura Dumitrescu & Raluca Balan, *University of Ottawa,* Ioana Schiopu-Kratina, *Statistics Canada* (E) [E]

The Asymptotically Optimal Estimating Equation for Longitudinal Data • L'equation estimatrice asymptotiquement optimale pour les données longitudinales [MS-229]

16:45 Yuliya Martsynyuk, *Carleton University* (E) [E]

New Multivariate Central Limit Theorems in Linear Structural and Functional Error-in-Variables Models • De nouveaux théorèmes centraux limite multivariés pour des modèles linéaires structurels et fonctionnels à erreur dans les variables [MS-230]

17:00 Zi Jin & Nancy Reid, *University of Toronto* (E) [E]

On the Asymptotic Properties of the Signed Composite Likelihood Ratio Statistic • Des propriétés asymptotiques de la statistique du ratio de vraisemblance composite signée [MS-231]

17:15 Olivier Wintenberger & Jean Marc Bardet, *Université Paris 1* (F) [E]

Asymptotic Normality of the Quasi-Maximum Likelihood Estimator for Affine Models • Normalité asymptotique de l'Estimateur du Quasi Maximum de Vraisemblance pour les modèles affines [MS-232]

16:30 – 18:00 Session 10G — Capital Hall 7A

Survey Methods
Méthodes d'échantillonnage

Contributed Paper Session • Séance de communications libres

Session sponsored by the *Survey Methods Section*
Session parrainée par le *Groupe de méthodologie d'enquête*

Chair • Président: David *Haziza, Université de Montréal*

16:30 Christian Derquenne, *EDF R&D* (F) [E]

Free Model for Generalized Path Modelling • Modèle libre dans le cadre des modèles à équations structurelles généralisés [MS-233]

16:45 Lionel Qualité, Desislava Nedyalkova & Yves Tillé, *Université de Neuchâtel* (F) [E]

General Framework for the Rotation of Units in Repeated Survey Sampling • Cadre général pour la rotation des individus dans les enquêtes répétées [MS-234]

17:00 Marc Christine & Sébastien Faivre, *Insee-France* (F) [F]

Some Methodological Issues on Drawing Samples from the New French Census. • Questions méthodologiques sur les tirages d'échantillons dans le nouveau recensement français [MS-235]

Mercredi 28 mai • Wednesday, May 28

08:30 – 10:00 Session 11A — Congress Hall A/E

Pierre Robillard Award Addresses
Allocutions des récipiendaires du prix Pierre Robillard

Special Session • Conférence spéciale

Organizer and Chair • Responsable et président: Sudhir R. Paul, *University of Windsor*

08:30 Juli Atherton, *McGill University*

Bayesian Optimal Design for Changepoint Problems • Plans d'expérience optimaux bayésiens pour des problèmes avec point de rupture [MS-236]

09:15 Jingjing Wu, *University of Calgary*

Minimum Hellinger Distance Estimation in Semiparametric Models • Estimation du minimum de la distance d'Hellinger dans les modèles semi-paramétriques [MS-237]

10:00 – 10:30 Pause-café • Coffee Break

10:30 – 12:00 Session 12A — Capital Hall 1A

Round Table: Issues in Collaboration in Medical Research
Table ronde : Questionnements sur la collaboration en recherche médicale

Invited Paper Session • Présentations sur invitation

Organizer and Chair • Responsable et président: Lehana Thabane, *McMaster University*

10:30 Robert Platt, *McGill University*

PeterBacchetti, *University of California - San Francisco*

Rollin Brant, *University of British Columbia*

Karen Kopciuk, *Alberta Cancer Board*

Rhonda Rosychuk, *University of Alberta*

Andrew Willan, *University of Toronto*

Round Table: Issues in Collaboration in Medical Research • Table ronde : enjeux de collaboration dans la recherche médicale [MS-238]

10:30 – 12:00 Session 12B — Capital Hall 2A

Monte Carlo Methods
Méthodes Monte Carlo

Invited Paper Session • Présentations sur invitation

Organizer and Chair • Responsable et président: Bruno Rémillard, *HÉC Montréal*

10:30 Pierre L'Ecuyer, *Université de Montréal*

On the Effectiveness of Randomized Quasi-Monte Carlo for Option Pricing • Sur l'efficacité des méthodes quasi Monte-Carlo randomisées pour la fixation des prix [MS-239]

11:00 Sylvain Rubenthaler, *Université de Nice-Sophia Antipolis*, Tobias Rydén & Magnus Wiktorsson, *Lund University*

Fast Simulated Annealing in R^d and Application to Max-likelihood • Recuit rapide dans R^d et application au max de vraisemblance [MS-240]

11:30 Michael Kouritzin, *University of Alberta*

Sequential Monte Carlo in Climate Change • Méthodes de Monte Carlo séquentielles pour l'étude des changements climatiques [MS-241]

10:30 – 12:00 Session 12C — Capital Hall 3A

Rank-based Inference
Inférence fondée sur les rangs

Invited Paper Session • Présentations sur invitation

Organizer and Chair • Responsable et président: Davy Paindaveine, *Université libre de Bruxelles*

10:30 Denis Larocque, *HÉC Montréal*, Riina Haataja, Jaakko Nevalainen & Hannu Oja, *Tampere School of Public Health*

Two Sample Tests for the Nonparametric Behrens-Fisher Problem with Clustered Data • Tests pour deux échantillons pour le problème de Behrens-Fisher non paramétrique avec données corrélées en grappes [MS-242]

11:00 Angelika Rohde, *Weierstrass Institute Berlin*

Adaptive Goodness-of-Fit Tests Based on Signed Ranks • Tests d'ajustement adaptatifs basés sur les rangs signés [MS-243]

11:30 Thomas Verdebout, Marc Hallin & Davy Paindaveine, *Université libre de Bruxelles*

Rank-based Inference for Principal Components • Inférence fondée sur les rangs pour les composantes principales [MS-244]

10:30 – 12:00 Session 12D — Capital Hall 4A

Censored Data
Données censurées

Invited Paper Session • Présentations sur invitation

Organizer and Chair • Responsable et président: Thierry Duchesne, *Université Laval*

10:30 Laurent Bordes, *Université de Pau et des Pays de l'Adour*, Jean-Yves Dauxois, *Université de Franche-Comté*, Pierre Joly, *Université Bordeaux 2*

Semiparametric Competing Risks with Missing Information • Risques concurrents semi-paramétriques avec données manquantes [MS-245]

11:00 Willard Braun & Xiuli Kang, *University of Western Ontario*

Data Sharpening for Monotonic Local Polynomial Regression with Interval-Censored Responses • Affilage de données en régression monotone polynomiale locale avec réponses censurés par intervalles [MS-246]

11:30 Brenda MacGibbon, *Université du Québec à Montréal*

Hazard Function Modeling • Modélisation de la fonction de risque instantanée [MS-247]

10:30 – 12:00 Session 12E — Capital Hall 5A

Design of Experiments in Business and Industrial Statistics
Planification d'expériences en statistique industrielle et de gestion

Contributed Paper Session • Séance de communications libres

Session sponsored by the *Business and Industrial Statistics Section*
Session parrainée par le *Groupe de statistique industrielle et de gestion*

Chair • Président: John Brewster, *University of Manitoba*

10:30 Robert McLeod & Jenna Tichon, *University of Winnipeg*, William Li, *University of Minnesota* E E

Optimal Nonregular Two-Level Split-Plot Designs with a Few Whole-Plots • Plan en parcelles subdivisées à deux niveaux non régulier optimal avec quelques grandes parcelles [MS-248]

10:45 Walter Tinsson, *Université de Pau et des Pays de l'Adour* F E

Experimental Designs for Qualitative and Quantitative Factors • Plans d'expérience pour facteurs qualitatifs et quantitatifs [MS-249]

11:00 Olivier Vasseur, *ONERA*, Jessica Franco & Bernard Corre, *TOTAL*, Michelle Sergent, *Université Paul Cézanne* F E

Space Filling Designs Qualification in High Dimension by Minimal Spanning Tree. • Qualification des plans Space Filling Designs en grande dimension par l'arbre de longueur minimal [MS-250]

11:15 Ndèye Niang Keita, *Chaire de Statistique Appliquée et CEDRIC, CNAM*, Flavio Fogliatto, *Universidade Federal do Rio Grande do Sul*, Pierre-Louis Gonzalez, *Chaire de Statistique Appliquée et CEDRIC, CNAM* F F

Multivariate Statistical Control of Batch Processes with Varying Duration • Contrôle multivarié de procédés par lots à temps variable [MS-251]

10:30 – 12:00 Session 12F — Capital Hall 6A

Data Validation and Small Area Estimation
Validation de données et estimation d'aires faibles

Contributed Paper Session • Séance de communications libres

Session sponsored by the *Survey Methods Section*
Session parrainée par le *Groupe de méthodologie d'enquête*

Chair • Président: Louis-Paul Rivest, *Université Laval*

10:30 Saad Rais, *Statistics Canada* E E

Outlier Detection Methodology for Price Index Surveys • Méthode de détection des valeurs aberrantes applicable aux enquêtes pour l'établissement des indices des prix [MS-252]

10:45 Holly Hutton, *Statistics Canada* E E

Dynamic Outlier Detection in Price Indexes • Détection dynamique des données aberrantes dans les indices de prix [MS-253]

11:00 Ka Ho Wu, *Chinese University of Hong Kong*, Zhao-Guo Chen, *Statistics Canada* E E

Benchmarking Comparison and Survey Error Modelling • Comparaison de l'étalonnage et modélisation de l'erreur de l'enquête [MS-254]

11:15 Catalin Dochitoiu, Susana Rubin-Bleuer & Carlos Leon, *Statistics Canada* E F

Small Area Estimation to Study the Impact of Globalization • Estimation sur petits domaines pour étudier l'effet de la mondialisation [MS-255]

11:30 Yong You, *Statistics Canada* E E

Small Area Estimation Using Area Level Models with Model Checking and Applications • Estimation sur petits domaines à l'aide de modèles au niveau du domaine avec vérification du modèle et applications [MS-256]

11:45 Yaakov Malinovsky & Yosef Rinott, *Hebrew University of Jerusalem* E E

Prediction of Ordered Random Effects in a Simple Small Area Model • Prévision d'effets aléatoires ordonnés dans un modèle de petit domaine simple [MS-257]

10:30 – 12:00 Session 12G — Capital Hall 7A

Estimation Methods
Méthodes d'estimation

Contributed Paper Session • Séance de communications libres

Chair • Président: Matias Salibian-Barrera, *University of British Columbia*

10:30 Bruno Pelletier, *Université Montpellier II*, Gérard Biau, *Université Paris VI*, Benoît Cadre, *ENS Cachan-Bretagne* F E

Density Support Estimation • Estimation du support d'une densité [MS-258]

10:45 Denis Bosq, *Université Paris 6 - Pierre et Marie Curie* F E

Estimating the Support of a Probability Distribution • De l'estimation du support d'une loi de probabilité [MS-259]

11:00 Thomas Laloë, *Université Montpellier II* F E

L_1 Quantization in Banach Spaces • Quantization L_1 dans un espace de Banach [MS-260]

11:15 Chi Wai Yu & Bertrand Clarke, *Department of Statistics, University of British Columbia, Canada* E E

Median Loss Analysis • Analyse de la perte médiane [MS-261]

11:30 Yogesh Tripathi & Éric Marchand, *Université de Sherbrooke* E E

Estimating Normal Quantiles when the Mean is Bounded • Estimations de quantiles normaux lorsque la moyenne est bornée [MS-262]

11:45 Mohammad Jafari Jozani & Éric Marchand, *Université de Sherbrooke* E E

Estimation and Prediction under Balanced Type loss functions with applications • Estimation et prévision sous des fonctions de perte de type équilibré avec applications [MS-263]

10:30 – 12:00 Session 12H — Capital Hall 8A

Functional Estimation
Estimation fonctionnelle

Contributed Paper Session • Séance de communications libres

Chair • Président: Luc Devroye, *McGill University*

10:30 Alexandre Leblanc, *University of Manitoba* E

Some Asymptotic Properties of Bernstein Estimators of Distribution Functions • Quelques propriétés asymptotiques des estimateurs de Bernstein d'une fonction de répartition [MS-264]

10:45 Stéphane Girard, *INRIA Rhône-Alpes*, Pierre Jacob, *Université Montpellier 2* F E

Frontier Estimation via Regression on High Power-transformed Data • Estimation de contour par régression sur les données élevées à la puissance [MS-265]

11:00 Min Jiang & Serge Provost, *University of Western Ontario* E

Orthogonal Polynomial Density Estimates and their Kernel Representation • Estimations de densité par des polynômes orthogonaux et représentation noyau de ces estimations [MS-266]

11:15 Serge Provost, *The University of Western Ontario*, Hyung-Tae Ha, *Kyungwon University*, Min Jiang, *The University of Western Ontario* F E

On Approximating the Derivative of the Logarithm of a Density by Means of a Rational Function • Sur l'approximation de la dérivée du logarithme d'une densité de probabilité par une fonction rationnelle [MS-267]

10:30 – 12:00 Session 12I — Capital Hall 9A

Extreme Values and Applications
Valeurs extrêmes et applications

Contributed Paper Session • Séance de communications libres

Chair • Président: Étienne Marceau, *Université Laval*

10:30 Mei Ling Huang & Ke Zhao, *Brock University* E E

On Inverse Power Laws and Risk Analysis • Des lois de puissance inverse et de l'analyse de risque [MS-268]

10:45 Jean-Noel Bacro, *Université Montpellier 2*, Liliane Bel, *AgroParisTech*, Christian Lantuéjoul, *Ecole des Mines, Fontainebleau* F E

A Multiple Test Procedure for the Asymptotic Independence of Spatial Extremes • Une procédure de tests multiples pour l'indépendance asymptotique des extrêmes spatiaux [MS-269]

11:00 Abdelaati Daouia, *GREMAQ, Université de Toulouse I*, Jean-Pierre Florens, *GREMAQ/IDEI, Université de Toulouse I*, Léopold Simar, *Université Catholique de Louvain* F E

Frontier Estimation and Extreme Values Theory • Estimation de frontières utilisant la théorie des extrêmes [MS-270]

11:15 Michel Grun-Rehomme, *Université Paris 2, ERMES*, Olga Vasechko, *Research Institute of Statistics, Kyiv, Ukraine*, Noureddine Benlagha, *Université Paris 2, ERMES* F E

Car Insurance Management: A Local Approach for Serious Accidents • Une approche locale de la gestion des sinistres graves en assurance automobile [MS-271]

11:30 Alexandru Asimit, *University of Toronto*, Bruce Jones, *University of Western Ontario* E E

Dependence and the Asymptotic Behavior of Large Claims Reinsurance • Dépendance et comportement asymptotique de grandes réclamations de réassurance [MS-272]

12:00 – 13:30 Dîner • Lunch

12:45 – 13:30 Nouvelles du CRSNG • News from NSERC — Capital Hall 1A

13:30 – 14:30 Session 13A — Congress Hall A/E

Canadian Journal of Statistics Award Address
Allocution du récipiendaire du prix de *La revue canadienne de statistique*

Special Session • Conférence spéciale

Organizer and Chair • Responsable et président: Wayne Oldford, *University of Waterloo*

13:30 Hwashin Shin, *Health Canada*, Glen Takahara, *Queen's University*, Duncan Murdoch, *University of Western Ontario* (E) [E]

Optimal Designs for Calibration of Orientations • Les plans optimaux pour calibrer des mesures d'orientation [MS-273]

14:35 – 16:05 Session 14A — Capital Hall 1A

Design and Analysis of Industrial Experiments
Planification et analyse d'expériences industrielles

Invited Paper Session • Présentations sur invitation

Session sponsored by the *Business and Industrial Statistics Section*
Session parrainée par le *Groupe de statistique industrielle et de gestion*

Organizer and Chair • Responsable et président: Rob McLeod, *University of Winnipeg*

14:35 John Brewster, *University of Manitoba* (E) [E]

Analysis of Data Arising from Industrial Split-Plot Screening Experiments • Analyse de données provenant de tests de dépistage industriels avec subdivision des parcelles [MS-274]

15:05 William Li & C. J. Nachtsheim, *University of Minnesota*, V. Agoboto, *Meharry Medical College* (E) [E]

Screening Designs for Model Selection • Plan de tri pour la sélection de modèles [MS-275]

15:35 Pritam Ranjan, *Acadia University*, Derek Bingham, *Simon Fraser University* (E) [E]

Batch-Sequential Designs for Estimating Process Features of Expensive Simulators • Conception séquencée par lots pour l'estimation des caractéristiques du processus des simulateurs coûteux [MS-276]

14:35 – 16:05 Session 14B — Capital Hall 2A

Calibration and Variance Estimation in Surveys
Calage et estimation de la variance dans les enquêtes

Invited Paper Session • Présentations sur invitation

Session sponsored by the *Survey Methods Section*
Session parrainée par le *Groupe de méthodologie d'enquête*

Organizer and Chair • Responsable et président: David Haziza, *Université de Montréal*

14:35 Guillaume Chauvet, *CREST-ENSAI* (F) [E]

Bootstrap for Multistage Sampling • Bootstrap pour un tirage à plusieurs degrés [MS-277]

15:05 Jean-Claude Deville, *ENSAI/CREST* (E) [F]

Successive or Iterated Calibrations • Calages successifs ou itérés [MS-278]

15:35 Anne Ruiz-Cazen, *TSE (Gremaq et IMT)*, Camélia Goga, *IMB*, Jean-Claude Deville, *ENSAI/CREST* E F

Linearization and Composite Estimation for Two-sample Survey Data • Linéarisation et estimation composite dans les enquêtes à deux échantillons [MS-279]

14:35 – 16:05 Session 14C — Capital Hall 3A

Environmental Statistics
Statistiques en environnement

Invited Paper Session • Présentations sur invitation

Organizer and Chair • Responsable et président: Bruce Smith, *Dalhousie University*

14:35 Grace Chiu, *University of Waterloo*, Peter Guttorp, *University of Washington*, Anton Westveld, *University of Nevada, Las Vegas* E E

Gauging Ecosystem Health with Latent Health Factor Models • Évaluation de la santé des écosystèmes à l'aide de modèles avec déterminant de la santé latent [MS-280]

15:05 Michael Dowd, *Dalhousie University* E E

Estimation for Differential Equation Based Environmental Models • Estimation pour modèles environnementaux basés sur des équations différentielles [MS-281]

15:35 Laure Malherbe, *Institut National de l'Environnement Industriel et des Risques (INERIS)*, Frédérik Meleux, Laurence Rouïl & Anthony Ung, *INERIS*, Chantal Fouquet (de), *Ecole Nationale Supérieure des Mines de Paris / Centre de géosciences* E F

Air Quality Mapping using Simulated and Observed Data • Utilisation de données modélisées et observées pour cartographier la pollution atmosphérique [MS-282]

14:35 – 16:05 Session 14D — Capital Hall 4A

Likelihood Inference
Inférence par vraisemblance

Invited Paper Session • Présentations sur invitation

Chair • Président: Don Fraser, *University of Toronto*

Organizer • Responsable: Nancy Reid, *University of Toronto*

14:35 Anthony Davison, & Claudia Lozada-Can, , *Ecole Polytechnique Fédérale de Lausanne (EPFL)* F E

Accurate Likelihood Inference for Overdispersed Data • Inférence par vraisemblance pour les données sur-dispersées [MS-283]

15:05 Nancy Reid & Jin Zi, *University of Toronto*, Grace Yi, *University of Waterloo* E F

Composite Likelihood • Vraisemblance composite [MS-284]

15:35 Judith Rousseau, *CEREMADE, Université Paris Dauphine et CREST*, Dongchu Sun, *University of Missouri, Columbia* E F

Bayesian Nonparametric Approaches with Partially Improper A Priori • Approches bayesiennes non-paramétriques utilisant des *a priori* partiellement impropres [MS-285]

14:35 – 16:05 Session 14E — Capital Hall 5A

Model Assessment and Goodness-of-fit
Évaluation de modèles et adéquation

Invited Paper Session • Présentations sur invitation

Organizer and Chair • Responsable et président: Richard Lockhart, *Simon Fraser University*

14:35 Martin Bilodeau, *Université de Montréal*, Pierre Lafaye de Micheaux, *Université Pierre Mendès France* E F

A-dependence Statistics for Mutual and Serial Independence of Categorical Variables • Les statistiques de dépendance A pour l'indépendance mutuelle et sérielle entre variables catégorielles [MS-286]

15:05 Jean-Renaud Pycke, *Université d'Evry* E E

Goodness-of-Fit Tests for Distributions on Fractafolds Based on the Sierpinski Gasket • Tests d'adéquation à des distributions sur des "fractafolds" basées sur le tamis de Sierpinski [MS-287]

15:35 Michael Stephens & Richard Lockhart, *Simon Fraser University* E F

On Two Tests for Exponentiality with Different Motivations • Deux tests pour la distribution exponentielle avec motivations différentes [MS-288]

14:35 – 16:05 Session 14F — Capital Hall 6A

Stochastic Models
Modèles stochastiques

Contributed Paper Session • Séance de communications libres

Chair • Président: Bruno Rémillard, *HEC-Montréal*

14:35 David Brillinger, *University of California, Berkeley* E E

Modelling Game Outcomes of the Brazilian 2006 Series A Championship as Ordinal-valued • Modélisation des résultats du championnat brésilien de série A comme valeurs ordinales [MS-289]

14:50 Yasaman Hosseinkashi, Shoja'eddin Chenouri & Christopher G. Small, *University of Waterloo* E E

Inference on Dynamic Graphs • Inférence dans les graphiques dynamiques [MS-290]

15:05 Sévérien Nkurunziza, *University of Windsor* E E

Inference Problem in some Predator-prey Systems • Problème d'inférence dans certains systèmes proie-prédateurs [MS-291]

15:20 Dhouha El Haj Ali, *Université Manar I, Tunisie* F F

Strategic Misspecification in Discrete Choice Models • Erreur de spécification dans les modèles de choix discrets [MS-292]

15:35 Jessica Franco, *TOTAL EP/GSR/TG/G&I*, Xavier Bay, *EMSE/G2I/3MI*, Bernard Corre, *TOTAL EP/GSR/COP/EM*, Delphine Dupuy, *EMSE/G2I/3MI* F E

Strauss Processes: A New Space-filling Design for Computer Experiments • Planification d'expériences numériques à partir du processus ponctuel de Strauss [MS-293]

15:50 Shafiqur Rahman, *Sultan Qaboos University* E E

Age Distribution Model for a Developing Nation • Modèle de répartition par âge pour un pays en développement [MS-294]

14:35 – 16:05 Session 14G — Capital Hall 7A

Analysis of Microarrays
Analyse de biopuces

Contributed Paper Session • Séance de communications libres

Session sponsored by the *Biostatistics Section*
Session parrainée par le *Groupe de biostatistique*

Chair • Président: Michael Newton, *University of Wisconsin at Madison*

14:35 Shaheena Bashir & Angelo Canty, *McMaster University*

Stratified False Discovery Rates and Q-values for Gene Expression Microarrays • Taux de fausses découvertes stratifiées et valeurs de q pour l'expression génétique en microréseaux [MS-295]

14:50 Wenqing He, *University of Western Ontario*, Shelley Bull, *University of Toronto*

Applications of Intra-class Correlation-based Criteria in Microarray Data Analysis • Applications d'une corrélation intra-classe fondée sur l'analyse des données des puce à ADN [MS-296]

15:05 Maela Kloareg & David Causeur, *Agrocampus Rennes*

Improving Supervised Classification for High Dimensional Data by Adding External Information. Application to Microarray data. • Améliorer les procédures de classification supervisée en intégrant de l'information externe. Application à des données génomiques. [MS-297]

15:20 Baba Thiam, Stéphane Robin & Emilie Lebarbier, *UMR INA P-G/ENGREF/INRA MIA 518*, Franck Picard, *UMR CNRS-5558/Laboratoire de Biométrie et Biologie Evolutive, Université Lyon 1*

Joint Segmentation of Processes: Application to the Analysis of Multiple CGH Profiles • Segmentation simultanée de signaux : application aux données de microarrays CGH [MS-298]

15:35 Tamanna Howlader & Yogendra P. Chaubey, *Concordia University*

Improved Estimation for cDNA Microarray Images in Wavelet Domain • Estimation améliorée pour les images de microréseaux d'ADNc dans le domaine des ondelettes [MS-299]

15:50 Jennifer Asimit, *Samuel Lunenfeld Research Institute, University of Toronto*, Neil Klar, *University of Western Ontario*, Shelley Bull, *Samuel Lunenfeld Research Institute, University of Toronto*

Assessing Inter-rater Agreement for Multiple Outcomes in Tissue Microarrays • Évaluer l'accord inter-juges pour des résultats multiples sur une micropuce de tissus [MS-300]

14:35 – 16:05 Session 14H — Congress Hall A/E West

Poster Session III
Séance d'affichage III

Contributed Paper Session • Séance de communications libres

14:35 You Liang, *University of Manitoba*

One-armed Bandit Processes with Observations from the Erlang(k) Distribution • Processus du bandit manchot avec observations provenant de la distribution Erlang(k) [MS-301]

14:35 Yi Hua & Xikui Wang, *University of Manitoba*

Optimal Investment-Consumption Problem with Hidden Markov Model • Problème d'investissement-consommation optimal avec modèle de Markov caché [MS-302]

14:35 Tingting Gou, W. John Braun & Duncan J. Murdoch, *University of Western Ontario* E

Simulation of Exotic Options • Simulation d'options exotiques [MS-303]

14:35 Duncan Murdoch & Tingting Gou, *University of Western Ontario* E

Binary Adaptive Rejection Sampling • Échantillonnage de rejet adaptatif binaire [MS-304]

14:35 Julie Carreau & Philippe Naveau, *Laboratoire des Sciences du Climat et de l'Environnement (LSCE-IPSL)*, Eric Sauquet, *Cemagref - Unité de recherche hydrologie-hydraulique* E

Statistical Rainfall-runoff Models based on a Mixture with Heavy-tailed Components • Modèles statistiques de pluie-débit à base de mélange avec composantes à queues lourdes [MS-305]

14:35 Ian Clara & Cam-Loi Huynh, *University of Manitoba* E

A Monte Carlo Simulation Investigation of Latent Variable Growth Models for Interaction Effects • Une étude par simulation de Monte Carlo de modèles de croissance à variables latentes pour les effets d'interaction [MS-306]

14:35 Charles Bouveyron, *Université Paris 1 Panthéon-Sorbonne*, Julien Jacques, *Université Lille 1* E

Adaptative Linear Models for Regression and Application to Socioeconomic Data • Modèles linéaires adaptatifs en régression et application à des données socioéconomiques [MS-307]

14:35 Rachel Dean & Wayne Oldford, *University of Waterloo* E

Warning Labels for Lotteries: Visual Display of Small Probabilities • Des avertissements pour les jeux de loterie : les images compréhensible des faibles probabilités [MS-308]

14:35 Hugues Massé & Fabrice Larribe, *Université du Québec à Montréal* E

Quasi-exact Estimation in the Coalescent Process • Estimation quasi exacte dans le processus de coalescence [MS-309]

14:35 Hugh Chipman, *Acadia University*, Erika Nahm, *Government of Canada* E

Mining Network Data • Exploration des données de réseau [MS-310]

14:35 Jason Loeppky, *University of British Columbia Okanagan*, Jerome Sacks, *National Institute of Statistical Sciences*, William Welch, *University of British Columbia* E

Choosing the Sample Size of a Computer Experiment • Choisir la taille d'échantillon d'une expérience numérique [MS-311]

14:35 Laurent Donini, *Xerox Research Centre Europe, Laboratoire Jean Kuntzman, INRIA Rhône Alpes*, Victor Ciriza, *Xerox Research Centre Europe*, Jean-Baptiste Durand, *Laboratoire Jean Kuntzman, INRIA Rhône Alpes, Universités de Grenoble*, Stéphane Girard, *Laboratoire Jean Kuntzman, INRIA Rhône Alpes* E

A Statistical Model for Optimizing Power Consumption of Printers • Un modèle statistique pour l'optimisation de la consommation électrique des imprimantes [MS-312]

14:35 Frédéric Bertrand & Myriam Maumy, *IRMA - ULP*, Jean-Pierre Bidon, *Interstat* E

High-performance Method to Predict Sales of Fast-moving Sales Goods • Quelle méthode pour prévoir les ventes de produits sur un marché à évolution rapide [MS-313]

14:35 Manuela Maia, *Universidade Católica Portuguesa*, Paula Vicente & Elizabeth Reis, *ISCTE* E

Indirect Sampling in the Context of Multiples Frames • Sondage Indirect dans le contexte des bases multiples [MS-314]

14:35 Piotr Biernot & Erica Moodie, *McGill University* E

Using Reducts to Guide Variable Selection for Optimal Dynamic Regimes • Utilisation de réduits pour guider la sélection de variables pour des régimes dynamiques optimaux [MS-315]

14:35 Frédéric Bertrand, *IRMA - ULP* E

Weakly Invariant Designs and Polynomial Designs • Plans faiblement invariants et problèmes de construction de type polynomial [MS-316]

14:35 Mohamedou Sow, Gilles Durrieu, Damien Tran, Pierre Ciret & Jean-Charles Massabuau, *GEMA Université Bordeaux 1 & CNRS* F

Development of Non-parametric Models to Analyse Bivalve Behaviour • Développement de modèles non paramétriques pour l'analyse du comportement de bivalve [MS-317]

14:35 Adel Settati, *IRMAR - Université Rennes 1* E

Some Invariance Principle for the Empirical Process Indexed by Function • Quelques principes d'invariance forts pour le processus empirique indexé par des fonctions [MS-318]

14:35 Frédéric Bertrand & Myriam Maumy, *IRMA-ULP*, Jacques Grosmann, *DRAF-SRPV Rhône* E

Using Multitable Analyses to Explore Data Generated by the National Grapevine Wood Diseases Survey • Analyse des données récoltées par l'Observatoire national des maladies du Bois de la vigne à l'aide de techniques multi-tableaux [MS-319]

14:35 Karim Rahim, *Queen's University*, Jeannine St. Jacques, *University of Regina*, David Thomson, *Queen's University* E

Harmonic Analysis of Individual Diatom Counts from Lake Mina Minesota • Analyse harmonique de numérations individuelles de diatomées dans le lac Mina au Minnesota [MS-320]

14:35 François Bellavance, *HEC Montréal*, Assi N'Guessan, *École Polytechnique Universitaire de Lille* E

A Statistical Model to Estimate the Effect of Roadside Billboards on Crash Frequencies in their Immediate Vicinity • Un modèle statistique pour évaluer l'effet des panneaux réclame le long des routes sur le nombre des collisions routières dans leur voisinage immédiat [MS-321]

16:05 – 16:30 Pause-café • Coffee Break

16:30 – 18:00 Session 15A — Capital Hall 1A

Markov Models
Modèles markoviens

Invited Paper Session • Présentations sur invitation

Session sponsored by the *Committee on Women in Statistics* and the *Caucus for Women in Statistics*
Session parrainée par le *Comité de femmes en statistique* et le *Caucus des femmes en statistique*

Organizer and Chair • Responsable et présidente: Julie Horrocks, *University of Guelph*

16:30 Ayesha Ali, *University of Guelph*, Angelique Massie, *University of Washington* E E

Using Graphical Markov Models in Model Selection • Utilisation de modèles markoviens graphiques dans la sélection de modèles [MS-322]

17:00 Gersende Fort, *LTCI, CNRS / TELECOM ParisTech* F E

Fluid Limit of Markov Chain Monte Carlo Samplers • Une approche fluide des échantillonneurs de Monte Carlo par chaînes de Markov [MS-323]

17:30 Leilei Zeng, *Simon Fraser University*, Richard Cook, *University of Waterloo* E E

A Multi-state Model for Bivariate Interval Censored Failure Time Data • Un modèle à multi-états pour données bivariées de temps de défaillance censurées par intervalles [MS-324]

16:30 – 18:00 Session 15B — Capital Hall 2A

Actuarial Science
Actuariat

Invited Paper Session • Présentations sur invitation

Organizer and Chair • Responsable et président: Gary Parker, *Simon Fraser University*

16:30 Jose Garrido, *Concordia University*, Md Sharif Mozumder Mozumder, *University of Dhaka, Bangladesh* (E) E

Estimation of the Jump Size Distribution in a Levy Risk Process • Estimation de la mesure de sauts pour modèles de risque basés sur un processus de Lévy [MS-325]

17:00 Stéphane Loisel & Daniel Serant, *Université Claude Bernard Lyon 1, ISFA* (F) E

In the Core of Longevity Risk: Dependence in Stochastic Mortality Models and Cut-offs in Prices of Longevity Swaps • Au coeur du risque de longévité : dépendance stochastique et impact sur les prix des permutations de longévité [MS-326]

17:30 Gary Parker & Cary C.L. Tsai, *Simon Fraser University* (F) E

Optimal Strategies for Ruin Probabilities and Expected Gains • Stratégies Optimales pour probabilités de ruine et profits espérés [MS-327]

16:30 – 18:00 Session 15C — Capital Hall 3A

Goodness-of-fit Tests
Tests d'adéquation

Contributed Paper Session • Séance de communications libres

Chair • Président: Bruno Rémillard, *HEC-Montréal*

16:30 Alain Desgagné, *Université du Québec à Montréal*, Pierre Lafaye de Micheaux, *Université Pierre-Mendès-France*, Alexandre Leblanc, *University of Manitoba* (E) E

Goodness-of-Fit Test for Normality Against Generalized Exponential Power Alternatives • Test de normalité contre des alternatives de la famille de puissance d'exponentielles généralisée [MS-328]

16:45 John Spinelli, *B.C. Cancer Research Centre*, Michael Stephens, *Simon Fraser University* (E) E

Testing the Gamma and Weibull Distributions with Grouped Data • Tester les distribution de Weibull et de Gamma avec les données groupées [MS-329]

17:00 Gilles Ducharme & Céline Vincent, *Université Montpellier II* (F) E

Change Point Test to Detect Vortical Structures in Stretches of Water • Un test de rupture pour détecter des structures tourbillonnaires dans les plans d'eau [MS-330]

17:15 Imen Kammoun & Jean-Marc Bardet, *Université Paris 1 - Panthéon Sorbonne* (F) E

Detecting Changes in the Fluctuations of a Gaussian Process and an Application to Heartbeat Time Series • Détection de ruptures pour un processus gaussien et une application aux séries des fréquences cardiaques [MS-331]

17:30 Ahmad Zghoul, *Qatar University* (E) E

A Goodness-of-Fit Test for Exponentiality • Test de la qualité de l'ajustement pour l'exponentialité [MS-332]

16:30 – 18:00 Session 15D — Capital Hall 4A

Inference for Linear Models
Inférence pour modèles linéaires

Contributed Paper Session • Séance de communications libres

Chair • Présidente: Jana Jureckova, *Charles University Prague*

16:30 Ali Ghodsi, Babak Alipanahi & Michael Biggs, *University of Waterloo*

Distance Metric Learning Vs. Fisher Discriminant Analysis • Apprentissage par mesure de distance contre analyse discriminante de Fisher [MS-333]

16:45 Saumen Mandal & Mrityunjay Samanta, *University of Manitoba*

A Unified Approach to Efficient Estimation in Simple Linear Regression • Une approche unifiée à l'estimation efficace en régression linéaire simple [MS-334]

17:00 Karelyn Davis, *Statistics Canada & Carleton University*, Chul-Gyu Park & Sanjoy Sinha, *Carleton University*

Constrained Inference in Generalized Linear and Mixed Models • Inférence avec contraintes des modèles linéaires généralisés et mixtes [MS-335]

17:15 Gérard d'Aubigny & Catherine d'Aubigny, *MS^3-LJK, Université de Grenoble*, Ali Shadrock, *Université de Téhéran, Iran*

Testing for Null Partial Correlation: a Formal Comparison of Freedman-Lane's and Kennedy's Permutation Test Strategies • Tests de nullité d'une corrélation partielle : une comparaison formelle des stratégies permutationnelles de Freedman-Lane et Kennedy [MS-336]

17:30 Marc Bourdeau, *École Polytechnique de Montréal*

Note on the Respective Widths of Confidence Intervals for Means with Known and Unknown Variances • Notes sur les largeurs respectives des intervalles de confiance pour les moyennes avec variances connues et inconnues [MS-337]

17:45 Bezza Hafidi, *Université Ibn Zohr, Agadir, Maroc*, Abdallah Mkhadri, *Université Cadi-Ayyad, Marrakech,Maroc*

Kullback Information Criterion for Mixture Regression Models • Critère d'Information de Kullback pour le mélange des modèles de régression [MS-338]

16:30 – 18:00 Session 15E — Capital Hall 5A

Clinical Trials
Essais cliniques

Contributed Paper Session • Séance de communications libres

Chair • Président: John Neuhaus, *University of California, San Francisco*

16:30 Xianming Tan, *NCIC Clinical Trials Group, Cancer Research Institute, Queen's University*, Glen Takahara, *Queen's University*, Dongsheng Tu, *NCIC Clinical Trials Group, Cancer Research Institute, Queen's University*

Two-Stage Design for Phase II Cancer Clinical Trials with Responses and Early Progression as Coprimary Endpoints • Design en deux étapes pour les essais cliniques de phase II en cancérologie avec la réponse ou la progression précoce comme co-critères de jugement principaux [MS-339]

16:45 Timothy Ramsay, *University of Ottawa*, Kaveh Shojiana, *The Ottawa Hospital*

Early Stopping Rules and Bias in RCTs • Règles précoces d'arrêt et biais dans les études randomisées contrôlées (ERC) [MS-340]

17:00 Zhijian Chen, Grace Y. Yi & Changbao Wu, *University of Waterloo* E E

Analysis of Clustered Data with Response Measurement Error • Analyse des données agrégées ayant une erreur de mesure de la réponse [MS-341]

17:15 Yun Hee Choi & Laurent Briollais, *Samuel Lunenfeld Research Institute at Mount Sinai Hospital* E E

Modeling Multistage Sampling of Family Data with Missing Information • Modélisation de l'échantillonnag à plusieurs niveaux de données familiales avec des informations manquantes [MS-342]

16:30 – 18:00 Session 15F — Capital Hall 6A

Functional Estimation and Censored Data
Estimation fonctionnelle et données censurées

Contributed Paper Session • Séance de communications libres

Chair • Président: Lauren Bordes, *Université de Pau*

16:30 Jean-François Plante, *University of Toronto* F E

On a Weighted Kaplan-Meier Estimate • À propos d'un estimateur de Kaplan-Meier pondéré [MS-343]

16:45 Olivier Bouaziz, *Laboratoire de Statistique Théorique et Appliquée*, Olivier Lopez, *Crest-Ensai, Irmar and Weierstrass Institute* E F

Conditional Density Estimation in a Censored Single-index Regression Model • Estimation de la densité conditionnelle dans un modèle à direction révélatrice unique en présence de censures [MS-344]

17:00 Elodie Brunel, *Université Montpellier 2*, Fabienne Comte, *Université Paris Descartes* F E

Distribution Function Estimation under Interval Censoring Case 1 • Estimation de la fonction de répartition dans le modèle de censure par intervalle de type 1 [MS-345]

17:15 Eve Leconte & Sandrine Casanova, *TSE (GREMAQ), Université Toulouse 1* F E

Using Conditional Quantiles to Estimate the Cumulative Distribution Function of a Censored Variable in a Small Area • Estimation sur un petit domaine de la fonction de répartition de l'utilisation de quantiles conditionnels variable censurée à l'aide de quantiles conditionnels [MS-346]

17:30 Philippe Saint-Pierre & Agathe Guilloux, *Laboratoire de Statistique Théorique et Appliquée (LSTA), Université Pierre et Marie Curie - Paris 6* F E

Bivariate Cumulative Distribution Function Estimate in Case of Right and Left Censoring • Estimateur de la fonction de répartition bivariée avec censures à droite et à gauche [MS-347]

17:45 Imad Bou-Hamad, *HEC Montreal* E E

Discrete-Time Survival Trees • Arbres de survie à temps discret [MS-348]

16:30 – 18:00 Session 15G — Capital Hall 7A

Stochastic Models in Biostatistics
Modèles stochastiques en biostatistique

Contributed Paper Session • Séance de communications libres

Session sponsored by the *Biostatistics Section*
Session parrainée par le *Groupe de biostatistique*

Chair • Président: Daniel Commenges, *Université Victor Segalen Bordeaux 2*

16:30 Patrick Brown & Todd Norwood, *Cancer Care Ontario*, Steve Fan & Jamie Stafford, *University of Toronto* E

Small-Scale Spatial Modelling of Cancer Risk in Ontario • Analyse spatiale a courte échelle des taux de cancer en Ontario [MS-349]

16:45 Mahmoud Torabi & Rhonda J. Rosychuk, *University of Alberta* E

Spatio-temporal Modeling of Disease Rate Mapping • Modélisation spatio-temporelle de la cartographie du taux d'incidence d'une maladie [MS-350]

17:00 Jeffrey Dowden, Khokan Sikdar & John Knight, *Newfoundland and Labrador Centre for Health Information* E

An Application of Modeling Count Data using Linked Administrative Healthcare Databases • Une application du dénombrement des données modélisées en utilisant des bases de données administratives sur les soins de santé [MS-351]

17:15 Alejandro Murua, *Université de Montréal* E

Modeling Replicated Time-course Evolution of Gene-expression Profiles • La modélisation de l'évolution dans le temps de profils d'expression génétique [MS-352]

17:30 Alain Latour, *Université Pierre-Mendès-France/MS3-LJK*, Lionel Truquet, *Université Paris 1 / CREST* F

A Positive and Negative Integer-valued Bilinear Type Process • Un processus bilinéaire à valeurs entières positives et négatives [MS-353]

Jeudi 29 mai • Thursday, May 29

08:30 – 10:00 Session 16A — Congress Hall A/E

Marie-Jeanne-Laurent-Duhamel Addresses
Allocutions des récipiendaires du prix Marie-Jeanne-Laurent-Duhamel

Special Session • Conférence spéciale

Chairs • Présidents: Gérard Biau, *Université Paris 6* & Avner Bar-Hen, *Université Paris 5*

Organizer • Responsable: Jean-Jacques Droesbeke, *Université libre de Bruxelles*

08:30 Gilles Stoltz, *CNRS - Ecole normale supérieure - HEC Paris* (F) E

Sequential Aggregation of Predictors, with Aapplication to Air-Quality Forecasting • Agrégation séquentielle de prédicteurs, avec application à la prévision de la qualité de l'air [MS-354]

09:15 Jean-François Quessy, *Université du Québec à Trois-Rivières*, Christian Genest, *Université Laval*, Bruno Rémillard, *HEC Montréal*, Marie-Hélène Toupin & Mhamed Mesfioui, *Université du Québec à Trois-Rivières* (F) E

Copulas : Goodness-of-fit and Independence Tests • Copules : tests d'adéquation et d'indépendance [MS-355]

10:00 – 10:30 Pause-café • Coffee Break

10:30 – 12:00 Session 17A — Capital Hall 1A

Statistical Genomics
Statistique génomique

Invited Paper Session • Présentations sur invitation

Session sponsored by the *Biostatistics Section*
Session parrainée par le *Groupe de biostatistique*

Organizer and Chair • Responsable et président: Raphaël Gottardo, *University of British Columbia*

10:30 David Bickel, *University of Ottawa*, Mary Beatty, Shai Lawit & Nicholas Bate, *Pioneer Hi-Bred international, Inc.* (E) F

Gene-Gene Influences Inferred from Microarray Time Course Data • Influences gène-gène inférées d'après des données de microréseau recueillies au cours du temps [MS-356]

11:00 Michael Newton & Chung Lisa, *University of Wisconsin* (E) E

A Dynamic Programming Algorithm to Evaluate Dirichlet Ordering Probabilities, with Application to Multigroup Gene Expression Analysis • Un algorithme dynamique pour évaluer les probabilités d'arrangements de variables Dirichlet, avec application à l'analyse d'expression génétique à plusieurs groupes [MS-357]

11:30 Aurélie Labbe, *Université Laval*, Alexandre Bureau, Arafat Tayeb & Chantal Merette, *Centre de recherche Université Laval Robert Giffard* (E) E

Generalization to Extended Pedigrees of a Latent Class Model with Familial Dependence • Généralisation aux pedigrees étendus d'un modèle de classes latentes avec dépendance familiale [MS-358]

10:30 – 12:00 Session 17B — Capital Hall 2A

Parametric Statistics on Varieties
Statistique paramétrique sur les variétés

Invited Paper Session • Présentations sur invitation

Session jointly sponsored by the *Probability Section* and the *Mathematical Statistics Section of SFdS*
Session conjointement parrainée par le *Groupe de probabilité* et le *Groupe de statistique mathématique de la SFdS*

Chair • Président: Laurent Younès, *Johns Hopkins University*

Organizer • Responsable: Philippe Berthet, *Université de Rennes I*

10:30 Peter Kim, *University of Guelph*, Peter Bubenik, *Cleveland State University* E E

A Statistical Approach to Persistent Homology • Une approche statistique de l'homologie persistante [MS-359]

11:00 Victor Patrangenaru, *Florida State University*, Rabi Bhattacharya, *University of Arizona* E E

Nonparametric Statistical Analysis on 3D Shape Manifolds • Analyse statistique non paramétrique sur variétés de formes en trois dimensions [MS-360]

11:30 Laurent Younes, *Johns Hopkins University* F E

Statistical Analysis of Surface Shape • Analyse statistique dans des espaces de surfaces [MS-361]

10:30 – 12:00 Session 17C — Capital Hall 3A

Copulas
Copules

Invited Paper Session • Présentations sur invitation

Organizer and Chair • Responsable et président: Arthur Charpentier, *Université Rennes 1*

10:30 Anne-Catherine Favre, *Institut national de la recherche scientifique, Centre Eau, Terre et Environnement*, Christian Genest, *Université Laval*, Julie Béliveau, *Ministère de l'Éducation, du Loisir et du Sport*, Christiane Jacques, *Ministère des Affaires Municipales et des Régions* F E

Meta-elliptical Copulas and their Applications • Les copules méta-elliptiques et leurs applications [MS-362]

11:00 Stéphane Loisel, *Université Claude Bernard Lyon 1*, Arthur Charpentier, *Université de Rennes 1* E E

On Some Classes of Hierarchical Archimedean Copulas • Étude de certaines classes de copules archimédiennes hiérarchiques [MS-363]

11:30 Johanna Neslehova, *ETH Zurich*, Christian Genest, *Université Laval* E F

A Primer on Copulas for Count Data • Une introduction aux copules dans le contexte de données de dénombrement [MS-364]

10:30 – 12:00 Session 17D — Capital Hall 4A

Empirical Likelihood
Vraisemblance empirique

Invited Paper Session • Présentations sur invitation

Chair • Président: Johan Segers, *Université catholique de Louvain*

Organizer • Responsable: Ingrid VAN KEILEGOM, *Université Lille 3*

10:30 Ingrid VAN KEILEGOM, *Université catholique de Louvain*, Elisa MOLANES-LOPEZ, *Universidad Carlos III de Madrid*, Noel VERAVERBEKE, *Universiteit Hasselt* (F) E

Empirical Likelihood for Non-smooth Criterion Functions • La vraisemblance empirique pour des fonctions de critère non-lisses [MS-365]

11:00 Dan NORDMAN & Petrutza CARAGEA, *Iowa State University* (E) E

A Spatial Empirical Likelihood for Variogram Estimation • Une vraisemblance spatiale empirique pour l'estimation de variogrammes [MS-366]

11:30 Liang PENG, *Georgia Institute of Technology* (E) E

Conditional Variance Estimation in Heteroscedastic Regression Models • Estimation conditionelle de la variance en régression hétéroscédastique [MS-367]

10:30 – 12:00 Session 17E — Capital Hall 5A

Estimation Methods in Business and Industrial Statistics
Méthodes d'estimation en statistique industrielle et de gestion

Contributed Paper Session • Séance de communications libres

Session sponsored by the *Business and Industrial Statistics Section*
Session parrainée par le *Groupe de statistique industrielle et de gestion*

Chair • Président: François BELLAVANCE, *HEC-Montréal*

10:30 Sidonie LEFEBVRE, Gérard DURAND, Antoine ROBLIN & Suzanne VARET, *ONERA The French Aerospace Lab* (F) E

Statistical Estimation of Aircraft Infrared Signature Dispersion • Estimation statistique de la dispersion de la signature infrarouge d'un aéronef [MS-368]

10:45 David GINSBOURGER & Olivier ROUSTANT, *Ecole des Mines de Saint-Etienne*, Yann RICHET, *Institut de Radioprotection et de Sûreté Nucléaire* (F) E

Kriging with Heterogeneous Nugget Effect for the Approximation of Noisy Simulators with Tunable Fidelity • Krigeage avec effet de pépite hétérogène pour l'approximation de simulateurs bruités à fidélité réglable [MS-369]

11:00 Fateh CHEBNA & Taha, B.M.J. OUARDA, *Institut National de la Recherche Scientifique* (E) F

Depth and Homogeneity in Regional Flood Frequency Analysis • Profondeur et homogénéité dans l'analyse régionale de la fréquence des crues [MS-370]

11:15 Shojaeddin CHENOURI, *University of Waterloo*, Asokan MULAYATH VARIYATH, *Texas A&M University*, Stefan H. STEINER, *University of Waterloo* (E) E

A Multivariate Robust Control Chart for Individual Observations • Graphique de contrôle robuste multivarié pour les observations individuelles [MS-371]

11:30 Mark REESOR & Wei XIA, *University of Western Ontario* (E) F

Risk Analysis of a Retail Bond Sales Strategy • Analyse du risque dans la stratégie de la vente d'obligations d'épargne au détail [MS-372]

11:45 David GINSBOURGER, Xavier BAY & Laurent CARRARO, *Ecole des Mines de Saint-Etienne* (F) E

Covariance Kernels for Spatial Interpolation of Symmetrical Functions • Noyaux de covariance pour le krigeage de fonctions symétriques [MS-373]

10:30 – 12:00 Session 17F — Capital Hall 6A

Mixture Models
Modèles de mélanges

Contributed Paper Session • Séance de communications libres

Chair • Président: Yonggan Zhao, *Dalhousie University*

10:30 Othmane Kortbi, *Université de Sherbrooke* (F) [F]

Bayes Minimax Estimators of the Mean of a Scale Mixture of Multivariate Normal Distributions • Les estimateurs de Bayes minimax de la moyenne des distributions mélangées de lois normales [MS-374]

10:45 Jean-Patrick Baudry, *Université Paris-Sud*, Gilles Celeux, *INRIA* (F) [E]

Combining Gaussian Components for Clustering: Mixtures of Mixtures • Classifier en agrégeant des composantes gaussiennes : mélanges de mélanges [MS-375]

11:00 Catherine Trottier, Christian Lavergne & Mariem Zouch, *I3M - Montpellier* (F) [E]

Weighted Poisson Mixed Model for Underdispersed Longitudinal Count Data • Modèle de Poisson pondéré mixte pour données longitudinales de comptage sous-dispersées [MS-376]

11:15 John Bentley, Richard Lockhart & Michael Stephens, *Simon Fraser University* (E) [E]

On a Mixture of Von Mises and Uniform Distributions • À propos d'un mélange de loi de von Mises et de loi unifrome [MS-377]

11:30 Zhengmin Zhang, A.K.Md. Ehsanes Saleh & Patrick J. Farrell, *Carleton University* (E) [E]

Moment Method Estimation of Mixing Distributions • Estimation d'un mélange de distributions par la méthode des moments [MS-378]

11:45 Pengfei Li & Jiahua Chen, *University of British Columbia* (E) [E]

Homogeneity Test in Normal Mixture Models: the EM Approach • Test d'homogénéité pour des modèles de mélanges de normales : l'approche EM [MS-379]

10:30 – 12:00 Session 17G — Capital Hall 7A

Time Series Modelling
Modélisation de séries chronologiques

Contributed Paper Session • Séance de communications libres

Chair • Président: Pierre Duchesne, *Université de Montréal*

10:30 Nicolas Bousquet, Thierry Duchesne, Louis-Paul Rivest & Jérôme Lemay, *Université Laval* (F) [E]

Definition and Estimation of Biological Reference Points for Halieutic Resource Management in Stochastic Frameworks. • Définition et estimation de repères biologiques pour la gestion de ressources halieutiques dans des cadres d'étude stochastiques [MS-380]

10:45 Vincent Lefieux, *RTE/DMA - CREST ENSAI* (F) [E]

Application of Semiparametric Modelling to Times Series Forecasting. Case of the Electricity Consumption • Modèles semi-paramétriques appliqués à la prévision des séries temporelles. Cas de la consommation d'électricité [MS-381]

11:00 Elizabete Almeida, Michael Dowd & Joanna Flemming, *Dalhousie University*, William Li, *Fisheries and Oceans Canada* (E) [E]

An Analysis of Biogeochemical Variables from Bedford Basin, NS, Canada Using a Trend Plus

Cycle State Space Model • Une analyse des variables bio-géo-chimique de Bedford Basin, NE, Canada en utilisant un cycle de tendance plus un modèle d'espace état [MS-382]

11:15 Virginie Dordonnat, *EDF R&D, France / VU University Amsterdam*, Siem Jan Koopman & Marius Ooms, *VU University Amsterdam*, Alain Dessertaine, *EDF R&D France* (F) E

Dynamic Factors in Periodic Time-varying Regression Models • Facteurs dynamiques dans des modèles de régression périodiques à coefficients variables dans le temps [MS-383]

12:00 – 13:30 Dîner • Lunch

13:30 – 14:30 Session 18A — Congress Hall A/E

Address of the winner of th Centre de Recherches Mathématiques/Statistical Society of Canada Prize
Allocution du récipiendaire du prix Centre de recherches mathématiques/Société statistique du Canada

Special Session • Conférence spéciale

Organizer and Chair • Responsable et président: Louis-Paul Rivest, *Université Laval*

13:30 Paul Gustafson, *University of British Columbia* (E) E

The In-between World of Proper Priors and Nonidentified Models • L'entre deux mondes des lois *a priori* intégrables et des modèles non-identifiables [MS-384]

14:35 – 15:35 Session 19A — Capital Hall 1A

Isobel Loutit Invited Address on Business and Industrial Statistics
Allocution Isobel Loutit sur la statistique industrielle et de gestion

Special Session • Conférence spéciale

Session sponsored by the *Business and Industrial Statistics Section*
Session parrainée par le *Groupe de statistique industrielle et de gestion*

Organizer and Chair • Responsable et président: Roman Viveros-Aguilera, *McMaster University*

14:35 Vijay Nair, *University of Michigan* (E) E

Industrial Statistics: The Gap between Research and Practice • Statistiques industrielles : l'écart entre la recherche et la pratique [MS-385]

14:35 – 16:05 Session 19B — Capital Hall 2A

Synthetic Data Methods for Confidentiality Protection and Disclosure Avoidance
Méthodes de données synthétiques pour la protection de la confidentialité et de la non-divulgation

Invited Paper Session • Présentations sur invitation

Session sponsored by the *Survey Methods Section*
Session parrainée par le *Groupe de méthodologie d'enquête*

Organizer and Chair • Responsable et président: Michael Larsen, *Iowa State University*

14:35 Sam Hawala, *US Census Bureau* (E) E

Data Synthesis via Expert Knowledge, Modeling, and Matching • Construction de données synthétiques au moyen de connaissances, de modélisation et d'appariement [MS-386]

15:05 Jerome Reiter, *Duke University* (E) [E]

Synthetic Data Methods: A Discussion of the Most Pressing Challenges • Méthodes de données synthétique : une discussion sur les défis les plus pressants [MS-387]

15:35 Jennifer Huckett & Michael Larsen, *Iowa State University* (E) [E]

Combining Methods to Create Synthetic Microdata: Quantile Regression, Hot Deck, and Rank Swapping • Combiner des méthodes pour créer des microdonnées synthétiques : régression quantile, hot deck et permutation de rangs [MS-388]

14:35 – 16:05 Session 19C — Capital Hall 3A

Recent Advances in Analysis of Event Data
Développements récents dans l'analyse des données événementielles

Invited Paper Session • Présentations sur invitation

Session sponsored by / Session parrainée par
Groupe biopharmacie et santé de la SFdS

Organizer and Chair • Responsable et président: François Aubin, *Agence de la Biomédecine*

14:35 Michal Abrahamowicz, *McGill University*, Amel Mahboubi & Christine Binquet, *Université de Bourgogne*, Roch Giorgi, *Université de la Méditerranée*, Claire Bonithon-Kopp & Catherine Quantin, *Université de Bourgogne* (E) [F]

Flexible Modeling of Time-dependent and Non-linear Effects of Continuous Covariates in Relative Survival • Modélisation flexible des effets dépendants du temps et non loglinéaires des variables continues en survie relative [MS-389]

15:05 Gérard Derzko, *Sanofi-Aventis R&D* (E) [E]

Nonparametric Estimation with Recurrent Events – Application to Safety Data in Clinical Trials • Estimation non-paramétrique pour des événements récurrents – Application aux données de tolérance dans les études cliniques [MS-390]

15:35 Marc Fredette, *HEC Montréal* (E) [E]

Forecasting of Recurrent Events: a Bayesian Approach • Approche Bayésienne pour la prédiction d'événements récurrents [MS-391]

14:35 – 16:05 Session 19D — Capital Hall 4A

Time Series
Séries chronologiques

Invited Paper Session • Présentations sur invitation

Organizer and Chair • Responsable et président: Christian Francq, *Université de Lille 3*

14:35 Pierre Duchesne & Eugen Ursu, *Université de Montréal* (F) [E]

On Modeling and Diagnostic Checking Vector Periodic Autoregressive Time Series Models • De la modélisation et l'ajustement des modèles de séries chronologiques autorégressives périodiques vectorielles [MS-392]

15:05 Bruno Rémillard, *HEC Montréal*, Christian Genest, *Université Laval*, Kilani Ghoudi, *United Arab Emirates University* (F) [E]

Tests of Serial Dependance • Tests de dépendance sérielle [MS-393]

15:35 Jean-Michel Zakoian, *CREST et Université Lille 3*, Christian Francq, *Université Lille 3* (F) [E]

Relative Efficiencies of Tests on GARCH Coefficients • Efficacités relatives de tests sur les coefficients GARCH [MS-394]

14:35 – 16:05 Session 19E — Capital Hall 5A

Latent Variables
Variables latentes

Invited Paper Session • Présentations sur invitation

Organizer and Chair • Responsable et présidente: Mary Thompson, *University of Waterloo*

14:35 David Flora, *York University* (E) [E]

Two-part Latent Variable Models for Semicontinuous Longitudinal Data • Modèles à variable latente pour les données longitudinales semi-continues [MS-395]

15:05 James Ramsay, *McGill University* (E) [E]

Latent Variables: Cautionary Remarks and Alternative Approaches • Variables latentes : avertissements et approches alternatives [MS-396]

15:35 Mary Thompson, *University of Waterloo* (E) [E]

Latent Variable Measurement Models: Equivalence, Invariance and Identifiability • Modèles de mesure de variables latentes : équivalence, invariance et identifiabilité [MS-397]

14:35 – 16:05 Session 19F — Capital Hall 6A

Bayesian Methods
Méthodes bayésiennes

Contributed Paper Session • Séance de communications libres

Chair • Président: Éric Moulines, *Ecole Nationale Supérieure des Télécommunications*

14:35 Longhai Li, *University of Saskatchewan*, Radford Neal, *University of Toronto* (E) [E]

Compressing Parameters in Bayesian High-order Models • Paramètres compressants dans les modèles bayesiens à ordre élevé [MS-398]

14:50 Éric Marchand, *Université de Sherbrooke*, William E. Strawderman, *Rutgers University*, Keven Bosa, *Statistique Canada*, Aziz Lmoudden, *Université de Sherbrooke* (F) [E]

On the Behaviour of Bayesian Credible Intervals for some Restricted Parameter Space Problems: Recent Results • Sur le comportement de divers estimateurs bayésiens par intervalles pour des espaces paramétriques contraints : résultats récents [MS-399]

15:05 Aline Tabet, Arnaud Doucet & Kevin Murphy, *University of British Columbia* (E) [E]

Bayesian Inference in the Multivariate Probit Model • Inférence bayésienne dans le modèle probit multivarié [MS-400]

15:20 Christophe Abraham, *Montpellier SupAgro* (F) [E]

A Computation Method in Robust Bayesian Decision Theory • Une méthode de calcul d'après la théorie de la décision bayésienne robuste [MS-401]

15:35 Marylène Troupé & Jean Vaillant, *Université des Antilles et de la Guyane*, Evans Gouno, *Université Bretagne-Sud* (F) [F]

Bayesian Inference for Auto-excited Point Processes Observed Partially • Inférence bayésienne pour des processus ponctuels auto-excités partiellement observés [MS-402]

14:35 – 16:05 Session 19G — Capital Hall 7A

Nonparametric Density Estimation
Estimation non paramétrique de la densité

Contributed Paper Session • Séance de communications libres

Chair • Président: Benoît Cadre, *Université de Montpellier*

14:35 Eric Matzner-Lober, *IRMAR, Université Rennes 2*, Nicolas Hengartner, *Los Alamos National Laboratory*

Asymptotic Unbiased Density Estimators • Estimateurs asymptotiquement sans biais de la densité [MS-403]

14:50 Pierre-André Cornillon, *Montpellier SupAgro*, Nicolas Hengartner, *Los Alamos National Laboratory*, Eric Matzner-Lober, *IRMAR Université Rennes 2*

Recursive Bias Estimation and L2 Boosting • Réduction itérative de biais et L2 Boosting [MS-404]

15:05 Mostafa Filali, *Université de Bourgogne*, Rachid Sabre, *Laboratoire MAIS, ENESAD*

Asymptotic Normality and Test for Deconvolution of Mixed Measure • Normalité asymptotique et tests pour des estimations de déconvolution d'une mesure mixte [MS-405]

15:20 Alain Celisse & Stéphane Robin, *AgroParisTech/INRA MIA*

Model Selection by Cross-validation • Sélection de modèle par validation-croisée [MS-406]

15:35 Shan Jiang, *Queen's University*, Yaohua Wu, *University of Science and Technology of China*, Dongsheng Tu, *Queen's University*

Bandwidth Selection for Nonparametric Confidence Intervals of a Probability Density Function • Sélection du paramètre de lissage pour les intervalles de confiance non paramétriques d'une fonction de densité [MS-407]

15:50 Reza Pakyari, *Arak University*

Nonparametric Methods for Density Estimation in Multivariate Mixtures • Méthodes non paramétriques d'estimation de densités de mélanges multivariés [MS-408]

14:35 – 16:05 Session 19H — Capital Hall 8A

Survival Data Analysis
Analyse de données de survie

Contributed Paper Session • Séance de communications libres

Chair • Présidente: Ingrid van Keilegom, *Université catholique de Louvain*

14:35 Hyun Lim & Hyun Lim, *University of Saskatchewan*, Xu Zhang, *Georgia State University*

Analysis of Recurrent Gap Time Data with Application to the Emergency Department Visit Rates among Violently Injured Pediatric Patients • Analyse de données de temps d'écart avec application aux taux de visite aux urgences chez les patients pédiatriques violemment blessés [MS-409]

14:50 Pierre Joly, *Inserm, U897*, Fabien Corbiere, *INRA, UMR1225*

Parametric Frailty Approach in a Mixture Cure Model in Survival Analysis. • Modèles de mélanges paramétriques, avec fraction non à risque, en analyse de survie pour données groupées [MS-410]

15:05 Keyue Ding & Zhangjun Ding, , *Queen's University*

Genes Selection and Components Retention in Building a Supervised Survival Prediction Model • Sélection de gènes et rétention des composantes lors de la conception supervisée d'un modèle prédictif de survie [MS-411]

15:20 Douglas Schaubel, *University of Michigan* (E) [E]

Modeling Treatment Effects on the Recurrent Event Mean and Rate in the Presence of a Terminating Event • Modéliser les effets de traitement sur la moyenne et le taux d'évènements récurrents en présence d'un évènement final [MS-412]

15:35 Sony Brar, *University of Calgary*, Penny Brasher, *University of British Columbia* (E) [E]

Comparison of Failure Probabilities for Local-regional Recurrence • Comparaison des probabilités de défaillance pour évaluer la récidive locale ou régionale [MS-413]

16:05 – 16:30 Pause-café • Coffee Break

16:30 – 18:00 Session 20A — Capital Hall 1A

Teaching of Statistics: Didactical and Pedagogical Problems
L'enseignement de la statistique : problèmes didactiques et pédagogiques

Invited Paper Session • Présentations sur invitation

Session sponsored by / Session parrainée par
Groupe d'enseignement de la statistique de la SFdS

Organizer and Chair • Responsable et président: Jean-Claude Régnier, *Université de Lyon II*

16:30 Linda Gattuso, *Université du Québec à Montréal* (F) [E]

School Mathematics in Statistical Treatment and Teacher's Training • Les mathématiques scolaires dans les traitements statistiques et la formation des enseignants [MS-414]

17:00 Jean-Claude Régnier, *Groupe d'enseignement de la statistique (SFDS)* (F) [F]

Mathematical Knowledge and Statistical Skills: Obstacles in Statistical Learning among Non-specialists • Connaissances mathématiques et Compétences statistiques : quelques obstacles à l'apprentissage de la statistique chez les non-spécialistes [MS-415]

17:30 Jean-Claude Oriol, *Université de Lyon II* (F) [E]

Place of Simulation and Training of Inquiries in the Didactic of Statistics • Place de la simulation et de la pratique d'enquête dans la didactique de la statistique [MS-416]

16:30 – 18:00 Session 20B — Capital Hall 2A

Statistical Methods for Complex Survey Data
Méthodes statistiques pour données de sondages complexes

Invited Paper Session • Présentations sur invitation

Session sponsored by the *National Program on Complex Data Structures*
Session parrainée par le *Programme sur les structures de données complexes*

Organizer and Chair • Responsable et président: Changbao Wu, *University of Waterloo*

16:30 Xin Feng & Randy Sitter, *Simon Fraser University* (E) [E]

Confidence Intervals for Proportion and Quantile Estimations with Focus on the US National Health and Nutrition Examination Survey • Intervalles de confiance pour les estimations de proportions et de quantiles en mettant l'accent sur la National Health and Nutrition Examination Survey américaine [MS-417]

17:00 Taslim Mallick & Brajendra Sutradhar, *Memorial University*, Milorad Kovacevic, *Statistics Canada*

Analysis of Incomplete Longitudinal Binary Data Subject to Missing at Random with Application to the SLID Data • Analyse de données binaires longitudinales incomplètes sous un mécanisme de données manquant au hasard avec application aux données de l'EDTR [MS-418]

17:30 Qian Zhou, *University of Waterloo*, Yong You, *Statistics Canada*

Hierarchical Bayes Small Area Estimation for the Canadian Community Health Survey • Estimation bayésienne hiérarchique de petits domaines pour l'Enquête sur la santé dans les collectivités canadiennes [MS-419]

16:30 – 18:00 Session 20C — Capital Hall 3A

Inference and Applications
Inférence et applications

Contributed Paper Session • Séance de communications libres

Chair • Président: Marc Genton, *Université de Genève*

16:30 Bifeng Xie, *University of Western Ontario*

Measuring the Mortality Risk Based on the Lee-Carter Model • Mesurer le risque de mortalité à partir du modèle Lee-Carter [MS-420]

16:45 Jaiwook Baik, *Korea National Open UnivERSITY*

Several Approaches to Warranty Data • Plusieurs approches à l'égard des données sur la garantie [MS-421]

17:00 Wan-chen Lee, *University of Manitoba*

Discovering Frequent and Rare Patterns in a Nucleotide Sequence • Découverte de structures fréquentes et rares dans une suite de nucléotides [MS-422]

17:15 Suwannee Adsavakulchai, *University of the Thai Chamber of Commerce*

Risk Assessment of Blood Lead Level from Lead Contamination in the Upper Mae-Klong Watershed, Kanchanaburi Province, Thailand • Évaluation des risques liés à la concentration de plomb dans le sang résultant de la contamination par le plomb dans le bassin supérieur du Mae-Klong (province de Kanchanaburi, Thaïlande) [MS-423]

16:30 – 18:00 Session 20D — Capital Hall 4A

Hypothesis Tests
Tests d'hypothèses

Contributed Paper Session • Séance de communications libres

Chair • Président: Gilles Ducharme, *Université de Montpellier*

16:30 James Hanley, *McGill University / Université McGill*, Marilyse Julien & Erica Moodie, *McGill University*

Student's *z*, *t*, and *s*: What if Gosset had *R*? • Les statistiques *z*, *t*, et *s* de Student : et si Gosset avait *R* ? [MS-424]

16:45 Yadolah Dodge, *University of Neuchâtel*, Iraj Yadegari, *Tarbiate Modares Universtiy*

On Direction Dependence • De la dépendance de la direction [MS-425]

17:00 Joseph Amuah, Michael Eliasziw & Gordon Fick, *University of Calgary*

Hypothesis Tests for the Concurrent Assessment for Interrater Agreement and Intrarater Reliability • Tests d'hypothèses pour l'évaluation simultané de l'accord inter-juges et de la fiabilité inter-juges [MS-426]

17:15 Bernard Goldfarb & Mireille Gettler-Summa, *Université Paris-Dauphine, Ceremade CNRS*

A Homogeneity Measurement For Functional Data Clusters • Mesure d'homogénéité de classes issues d'un ensemble de données fonctionnelles [MS-427]

17:30 Atsu Dorvlo, *Sultan Qaboos University, Sultanate of Oman*

A Test for Dominance • Un test de dominance [MS-428]

16:30 – 18:00 Session 20E — Capital Hall 5A

Bayesian Methods in Biostatistics
Méthodes bayésiennes en biostatistique

Contributed Paper Session • Séance de communications libres

Session sponsored by the *Biostatistics Section*
Session parrainée par le *Groupe de biostatistique*

Chair • Présidente: Judith Rousseau, *Université de Paris-Dauphine*

16:30 Geneviève Lefebvre & Paul Gustafson, *University of British Columbia*

Comparing traditional regression and inverse-probability-weighting for estimating causal effects: a Bayes risk approach • Comparaison par risque bayésien des approches de régression traditionnelle et de pondération pour l'estimation d'effets causals [MS-429]

16:45 Sohee Kang, Laurent Briollais & Liu Jinnan, *Samuel Lunenfeld Research Institue*

MCMC Model Determination for Bayesian Graphical Models • Détermination de modèle MCMC pour modèles graphiques bayésiens [MS-430]

17:00 Stephanie Powers & James Stamey, *Baylor University*

Bayesian Variable Selection for Logistic Regression with Misclassification • Sélection de la variables bayesienne pour la régression logistique avec erreurs de classement [MS-431]

17:15 Lihui Zhao, *Simon Fraser University*, Allan Becker, *University of Manitoba*, Moira Chan-Yeung, *University of British Columbia*

Bayesian Analysis of Longitudinal Binary Data Subject to Misclassification • Analyse bayesienne des données binaires longitudinales sujettes aux erreurs de classement [MS-432]

17:30 Fabrice Elegbede, Benoît Marçais & Jean-Claude Pierrat, *Institut Nationale de le Recherche Agronomique / ENGREF*

Statistical Modelling for Epidemiology of Alder Decline • Modélisation statistique pour l'épidémiologie du dépérissement des aulnes glutineux [MS-433]

17:45 Hyang Mi Kim, *University of Calgary*, A.K.Md.Ehsanes Saleh, *Carleton University*, K.M. Hassanein & R.S. Hassanein, *KUMC, Kansas*

Quasi-empirical Bayes Methodology for Improving Meta-analysis • Méthodologie de Bayes quasi-empirique pour améliorer la méta-analyse [MS-434]

16:30 – 18:00 Session 20F — Capital Hall 6A

Estimation from Biomedical Data
Estimation fondées sur des données biomédicales

Contributed Paper Session • Séance de communications libres

Session sponsored by the *Biostatistics Section*
Session parrainée par le *Groupe de biostatistique*

Chair • Président: Avner Bar-Hen, *Université Paris 5*

16:30 Mohammed Haddou & Louis-Paul Rivest, *Université Laval*

Estimating the Rotation Axes of the Ankle joint: A Directional Model Using Non-Linear Mixed-Effects Theory • Estimation des axes de rotation de la cheville : un modèle directionnel utilisant la théorie des modèles non-linéaires à effets mixtes [MS-435]

16:45 Laura Faye, *University of Toronto*, Shelley Bull, *University of Toronto; Samuel Lunenfeld Research Institute*, Lei Sun, *University of Toronto; Hospital for Sick Children Research Institute*

Reducing Selection Bias: Efficiency and Robustness of Parametric & Non-parametric Approaches • Réduction du biais de sélection : efficacité et robustesse des approches paramétrique et non paramétrique [MS-436]

17:00 Kevin Bleakley & Marie-Paule Lefranc, *Université Montpellier 2*, Gérard Biau, *Université Pierre et Marie Curie - Paris VI*

Recovering True Probabilities for Nucleotide Trimming Processes in T-cell Receptor TRA and TRG Junction Formation with IMGT Tools • Dérivation des vraies distributions de probabilité pour le processus de grignotage aléatoire des récepteurs T à l'aide d'outils automatiques (d'IMGT) [MS-437]

17:15 Krishna Saha, *Central Connecticut State University*

Semiparametric Estimation for the Dispersion Parameter in the Analysis of Over or Under Dispersed Count Data • Estimation semi-paramétrique pour le paramètre de dispersion dans l'analyse des données chiffrées surdispersées ou sous dispersées [MS-438]

17:30 Jiajia Zhang, *University of South Carolina*, Yingwei Peng, *Queen's University*

Estimation Method of the Semiparametric Mixture Cure Gamma Frailty Model • Méthode d'estimation du modèle de mélange semi-paramétrique gamma de fragilité de la guérison [MS-439]

17:45 Andreas Futschik & Florian Gach, *University of Vienna*

On the Inadmissibility of Watterson's Estimator • À propos de l'inadmissibilité de l'estimateur de Watterson [MS-440]

16:30 – 18:00 Session 20G — Capital Hall 7A

Estimation and Tests of Serial Dependence
Estimation et tests de dépendance sérielle

Contributed Paper Session • Séance de communications libres

Chair • Président: Bruno Rémillard, *HEC-Montréal*

16:30 David Thomson, *Queen's University*

Estimation of Autocorrelation and Trend-in-Time Series • Estimation d'autocorrélation et de tendances dans des séries chronologiques [MS-441]

16:45 Ye Sun, *University of Toronto*, Augustine Wong, *York University*, Marie Rekkas, *Simon Fraser University*

Improved Inference for First Order Autocorrelation Using Likelihood Analysis • Inférence améliorée pour l'autocorrélation de premier ordre utilisant l'analyse de vraisemblance [MS-442]

17:00 Mohammad Roknossadati & Mahmoud Zarepour, *University of Ottawa*

M-estimation for a Spatial Unilateral Autoregressive Model with Infinite Variance Innovations • M-estimation pour un modèle spatial unilatéral autorégressif avec innovations à variance infine [MS-443]

17:15 Melody Ghahramani, *University of Winnipeg*, Aerambamoorthy Thavaneswaran, *University of Manitoba*

Identification Method for Time Series Models Using Estimating Functions • Méthode d'identification pour les modèles de séries chronologiques utilisant les fonctions d'estimation [MS-444]

17:30 Melody Ghahramani, *University of Winnipeg*, Aerambamoorthy Thavaneswaran, *University of Manitoba*

Fuzzy Volatility Forecasting • Prévision de volatilités floues [MS-445]

16:30 – 18:00 Session 20H — Capital Hall 8A

Multivariate Statistics
Statistique multidimensionnelle

Contributed Paper Session • Séance de communications libres

Chair • Président: Martin Bilodeau, *Université de Montréal*

16:30 Julie Josse, François Husson & Jérôme Pagès, *Agrocampus Rennes*

Missing Values in Multiple Multivariate Analysis • Données manquantes dans les analyses multivariées multiples [MS-446]

16:45 Rafik Abdesselam, *Université de Caen Basse-Normandie*

Discriminant Multiple Correspondence Analysis • Analyse des correspondances multiples discriminantes [MS-447]

17:00 Merlin Mbuembue Njoya & Louis-Paul Rivest, *Université Laval*

Multivariates Non-linear Models for Estimating the Biomass of Forest Trees • Modèles non linéaires multivariés pour l'estimation de la biomasse forestière [MS-448]

Abstracts • Résumés

Session 01A Lun. 26 mai • Mon. May 26, 08:30 – 08:45 Congress Hall A/E

Inaugural Session
Session inaugurale

Session 01B Lun. 26 mai • Mon. May 26, 08:45 – 10:00 Congress Hall A/E

SSC Presidential Invited Address
Allocution de l'invité du président de la SSC

[MS-1]
Statistics and Quantitative Risk Management
La statistique et la gestion quantitative des risques

Paul Embrechts, *ETH Zurich*

Mathematics and Finance experience an intensive and fruitful exchange of ideas. This is obvious in the field of Mathematical Finance, where especially the pricing and hedging of financial derivatives leads to numerous interesting developments, both in theory as well as in the applications. In this talk, I will highlight a similar interplay between Statistics and Quantitative Risk Management. The latter field is fairly new and poses problems very different from standard Mathematical Finance and Econometrics. Topics to be discussed include: risk measure estimation, risk aggregation and concentration, risk diversification. Several examples will be given and areas of future statistical research highlighted.

Les mathématiques et la finance génèrent des échanges d'idées intensifs et fructueux. Un exemple évident se trouve dans le domaine des mathématiques financières. En particulier, le calcul de prix et de couvertures de produits dérivés engendre de nombreux développements intéressants aussi bien en théorie que dans la pratique. Je voudrais présenter de telles interactions qui existent entre la statistique et la gestion quantitative des risques. Ce dernier domaine est assez récent et soulève des problèmes très différents de ceux que l'on rencontre en mathématiques financières ou en économétrie. Les sujets abordés seront : l'estimation de mesures des risques, l'agrégation et la concentration des risques, la diversification des risques. Plusieurs exemples seront traités et des futurs sujets de recherche en statistique discutés.

Session 02A Lun. 26 mai • Mon. May 26, 10:30 – 12:00 Capital Hall 1A

Measurement Error Models
Modèles d'erreur de mesures

[MS-2]

Effects of Ignoring Measurement-Error Correlations in Longitudinal Covariates
Effets d'ignorer les corrélations de l'erreur de mesure dans les covariables longitudinales

Naisyin Wang & Erning Li, *Texas A&M University*, Nae-Yuh Wang, *Johns Hopkins University*

There is a long history of investigators using measurement error methodologies to analyze data from a setting where covariates in the primary regression are latent variables of a secondary longitudinal process. In this talk, I will use numerical outcomes and examples to illustrate potential concerns when one ignores the measurement-error correlations in the longitudinal covariates. I will also discuss how to extend existing methodologies to accommodate such scenarios and present an analysis of a Johns Hopkins Precursors Study dataset that investigates the impacts of initial average and increment slope of an individual's BMI on their hypertension status later in life.

Depuis longtemps, les chercheurs utilisent des méthodes basées sur l'erreur de mesure pour analyser les données dans des conditions où les covariables comprises dans la régression primaire sont des variables latentes d'un processus longitudinal secondaire. Des résultats numériques et des exemples seront utilisés dans la communication pour illustrer des problèmes potentiels lorsque l'on ignore les corrélations des erreurs de mesure dans les covariables longitudinales. La communication comprendra aussi une discussion de la manière d'étendre les méthodes existantes pour tenir compte de ce genre de scénarios et la présentation de l'analyse d'un ensemble de données provenant de l'étude des précurseurs réalisée à Johns Hopkins en vue d'évaluer les effets de la valeur moyenne initiale et de l'incrément initial de pente de l'IMC d'un sujet sur sa situation d'hypertension dans le futur.

[MS-3]

Diagnosing Misspecification of Generalized Linear Mixed Models Using Conditional Likelihood and Covariate Partitioning Methods
Diagnostiquer la mauvaise spécification de modèles linéaires généralisés mixtes au moyen de la vraisemblance conditionnelle et de méthodes de partitionnement des covariables

John Neuhaus & Charles McCulloch, *University of California, San Francisco*

Large discrepancies between parameter estimates from the fits of generalized linear mixed models, conditional likelihood methods, and models that partition covariates into between- and within-cluster components are possible. Using analytic results, simulation studies and fits to example data, this talk assesses the utility of comparing these fits as a method for diagnosing model misspecification. We show that simple discrepancies may not be indicative of a particular type of model failure, as has been suggested in the literature. We provide modifications of conditional and partitioning methods to enhance the diagnostic utility of comparing fits.

Il est possible d'observer des écarts importants entre les estimés des paramètres obtenus en ajustant des modèles linéaires généralisés, par méthodes de vraisemblance conditionnelle, et avec des modèles partitionnant les covariables en composantes inter- et intra-grappes. Par des résultats analytiques, des études de simulation, et des ajustements à des jeux de données, cet exposé évalue l'utilité de comparer ces ajustements comme méthode de diagnostic de mauvaise spécification de modèle. Nous montrons que des écarts simples peuvent ne pas être indicateurs d'un type particulier de défaillance du modèle, tel que suggéré dans la littérature. Nous apportons des modifications aux méthodes conditionnelles et par partitionnement pour améliorer l'utilité diagnostique de la comparaison des ajustements.

[MS-4]
Bayesian Analysis of Measurement Error Problems
Analyse bayesienne des problèmes d'erreur de mesure

Paul Gustafson, *University of British Columbia*, Shahadut Hossain, *BC Cancer Agency*

A common challenge in inferring regression relationships is poor measurement of one or more regressors. There is considerable literature on methodology which acknowledges, or 'adjusts for' such measurement error. One broad issue in this literature is the extent to which parametric procedures should be favoured or disfavoured, particularly in modelling the distribution of the true but unobserved regressor. Another broad issue is the extent to which strong assumptions about the measurement error mechanism may be required or desirable in order to ensure model identification. Within the rubric of Bayesian analysis, we will discuss both these issues.

Une difficulté rencontrée fréquemment dans l'inférence des relations de régression est celle de la mauvaise mesure d'une ou de plusieurs variables explicatives. Les auteurs de nombreuses études méthodologiques reconnaissent ce genre d'erreurs de mesure ou « font des corrections pour » ces erreurs. Une grande question soulevée dans cette littérature est celle de savoir dans quelle mesure il convient de favoriser ou de défavoriser les méthodes paramétriques, particulièrement dans la modélisation de la distribution de la variable explicative réelle, mais inobservée. Une autre grande question est celle de savoir dans quelle mesure des hypothèses fortes concernant le mécanisme d'erreur de mesure pourraient être requises ou souhaitables pour pouvoir reconnaître le modèle. Nous discuterons de ces deux questions dans le contexte de l'analyse bayésienne.

[MS-5]
Regression Analysis with a Misclassified Covariate from a Current Status Regression Analysis with a Misclassified Covariate from a Current Status Observation Scheme
Analyse de la régression avec une covariable mal classée à partir d'une analyse du statut de régression courant avec une covariable mal classée à partir d'un schème d'observation d'un statut courant

Richard Cook, *University of Waterloo*, Leilei Zeng, *Simon Fraser University*, Theodore Warkentin, *McMaster University*

In orthopedic studies interest lies in examining the effect of a serological response on risk of deep vein thrombosis. The serological response is an indicator of whether the patient developed antibodies following exposure to a antithrombotic drug but seroconversion status is available only at discharge from hospital. The seroconversion time is therefore subject to a current status observation scheme, of Case I interval censoring leading to potential misclassification. We develop likelihood based approach for fitting regression models that account for misclassification of seroconversion status due to early testing using parametric and nonparametric estimates of the seroconversion time distribution.

Dans les études sur l'orthopédie on s'intéresse à l'examen de l'effet de la réponse sérologique sur le risque de la thrombose veineuse profonde. La réponse sérologique est un indicateur dans le cas où le patient développe des anticorps suivant une exposition à un médicament antithrombotique, mais l'état de séroconversion se produit seulement lors du congé de l'hôpital. Le moment de séroconversion est d'ailleurs sujet à un statut d'observation courant, d'intervalle de censure de Cas 1 menant à une possible erreur de classement. Nous proposons une approche basée sur la vraisemblance pour ajuster les modèles de régression qui tiennent compte de l'erreur de classement du statut de séroconversion due au premier test utilisant les estimations paramétriques et non paramétriques de la distribution du temps de la séroconversion.

Session 02B Lun. 26 mai • Mon. May 26, 10:30 – 12:00 Capital Hall 2A

Session I in Memory of André Dabrowski
Séance I à la mémoire d'André Dabrowski

[MS-6]
André Dabrowski's Work on Limit Theorems and Weak Dependence
Les travaux d'André Dabrowski sur les théorèmes limites et la dépendance faible

Harold DEHLING, *University of Bochum*

I will highlight Andre's work on weak dependence and limit theorems focusing on ideas more than on specific results. I hope to tell you about André Dabrowski, the human being behind the mathematics. In addition to being a highly original mathematician, André was an inspiring teacher and a respected administrator. But above all, André was a deeply caring colleague and friend to many of us. Personally, I had the special privilege of more than 25 years of collaboration with André, and this has been one of the most rewarding experiences of my entire career. I have lost a great friend.

Pendant toute sa carrière, André Dabrowski a contribué par des idées originales à la théorie de la dépendance faible et les théorèmes limites. De surcroît, André a toujours su trouver des applications surprenantes et non simplistes de cette théorie. Dans ma présentation je mettrai en relief ses travaux sur la dépendance faible et les théorèmes limites.

Également, j'espère décrire l'homme derrière les mathématiques, car au-delà de sa carrière de mathématicien original, André était un professeur qui inspirait ses étudiants ainsi qu'un administrateur respecté. Mais surtout, André était pour moi et beaucoup de ses collègues, un ami sympathique. J'ai eu le privilège de collaborer avec André pendant 25 ans et cette collaboration a été parmi les plus fructueuses de ma carrière. J'ai perdu un grand ami !

[MS-7]
A Reminiscence of our Research Collaboration with Andre Dabrowski
Souvenir de notre collaboration avec Andre Dabrowski

David MCDONALD & Andre DABROWSKI, *University of Ottawa*, Uwe ROSLER, *Universitat Kiel*

Ion channels are nature's transistors. The evolutionary history of ion channels is so ancient that one can find common genes in plants and animals. These tiny channels open or close to allow relatively large currents of ions to flow (10^{-9} amperes for human nerve cells) in response to tiny electrical, chemical or mechanical signals. One of the early techniques to analyse these structures was the patch clamp experiment where a tiny portion of cellular membrane was stretched across the tip of a tiny pipette and subject to stimulus. Sensitive equipment was developed to measure the current across single channels. Unfortunately, it is hard to avoid getting more than one ion channel in a patch, so the observed current is often a superposition of different currents. I will discuss our work on deconvolving and analyzing these superimposed currents.

Les canaux ioniques font office de transistors biologiques. L'histoire évolutionnaire est tellement ancienne qu'on trouve des structures communes entre les plantes et les animaux. Ces canaux s'ouvrent pour faire passer un courant relativement grand comme réponse a un petit signal électrique, chimique ou mécanique. En laboratoire on découpe la membrane d'une cellule et on mesure le courant qui traverse la membrane en réponse à un signal. Malheureusement, on observe souvent la superposition des courants de plusieurs canaux ioniques. Je parlerai de nos travaux de déconvolution de cette superposition.

Session 02C Lun. 26 mai • Mon. May 26, 10:30 – 12:00 Capital Hall 3A

Data Mining
Forage de données

[MS-8]
Classification of Complex Data with Model-based Techniques
Classification de données complexes grâce aux techniques paramétriques

Charles Bouveyron, *Université Paris 1 Panthéon-Sorbonne*

Modern data give rise to new problems in statistical learning because of their complex nature. Indeed, the data provided by modern applications are usually high-dimensional or structured and classical statistical learning techniques are not able to deal with such data in a satisfying way. We will present in this talk model-based methods designed for the classification of complex data. Applications of these methods to image analysis and social network analysis will be developed as well.

Les données modernes posent des problèmes nouveaux en apprentissage statistique du fait de leur nature complexe. En effet, les données fournies par les applications modernes sont usuellement de grande dimension ou structurées et les méthodes classiques d'apprentissage statistique ne permettent pas de résoudre de façon satisfaisante les problèmes posés. Nous présenterons dans cet exposé des techniques paramétriques adaptées à la classification de données complexes. Nous présenterons également des applications de ces méthodes à l'analyse d'images et à l'analyse des réseaux sociaux.

[MS-9]
An Iterative Non-parametric Clustering Algorithm Based on Local Shrinking
Un algorithme de classification itératif non paramétrique basé sur le rétrécissement local

Steven Wang, *York University*, Weiliang Qiu, *Harvard University*, Ruben Zamar, *University of British Columbia*

We will present a non-parametric clustering method based on non-parametric local shrinking. Each data point is transformed in such a way that it moves a specific distance toward a cluster center. The direction and the associated size of each movement are determined by the median of its K-nearest neighbors. The optimal value of the number of neighbors is determined by optimizing some commonly used index functions that measure the strengths of clusters generated by the algorithm. The number of clusters and the final partition are determined automatically without any input parameter except the stopping rule for convergence.

Nous présentons une méthode de classification non paramétrique basée sur le rétrécissement local non paramétrique. Chaque observation est transformée de façon à ce qu'elle se déplace d'une distance spécifique vers le centre d'une grappe. La direction et la taille associée à chaque déplacement sont déterminées par la médiane des K plus proches voisins. La valeur optimale du nombre de voisins est déterminée en optimisant des fonctions indices fréquemment utilisées et mesurant la force des grappes générées par l'algorithme. Le nombre de grappes et la partition finale sont déterminés automatiquement sans paramètre en entrée outre le critère d'arrêt pour la convergence.

[MS-10]
Missing Data Approaches for High-dimensional Categorical Data
Approches des données manquantes pour données catégorielles de haute dimension

Russell Steele, *McGill University*, Anne-Sophie Charest, *Carnegie Mellon University*, Members of the *Canadian Scleroderma Research Group*

We compare two methods that use flexible mixtures of discrete distributions to approximate the

Nous comparons deux méthodes utilisant des mélanges flexibles de distributions discrètes pour approximer la distri-

posterior distribution of high-dimensional missing categorical data. First, we have implemented the Latent Class method of van Ginkel et al. (2007), which uses a Bayesian latent class model which assumes that the discrete variables are conditionally independent given an unknown latent class variable. We compare the results from this standard method to a new, non-parametric categorical clustering method by Zhang et al. (2006). These two methods are used for imputing missing data from the Canadian Scleroderma Research Group Registry.

bution *a posteriori* de données catégorielles manquantes de haute dimension. Premièrement, nous avons mis en œuvre la méthode de classe latente de van Ginkel *et al*. (2007), qui utilise un modèle de classe latente bayésien supposant que les variables discrètes sont indépendantes conditionnellement à une variable de classe latente inconnue. Nous comparons les résultats de cette méthode standard à une nouvelle méthode de classification catégorielle non paramétrique développée par Zhang *et al*. (2006). Ces deux méthodes sont utilisées pour imputer des données manquantes du registre du Groupe de recherche canadien sur la sclérodermie.

Session 02D Lun. 26 mai • Mon. May 26, 10:30 – 12:00 Capital Hall 4A

Statistical Models for Extreme Values
Modèles statistiques pour les valeurs extrêmes

[MS-11]
Tails of Archimedean Copulas
Queues des copules archimédiennes

Arthur Charpentier, *Université Rennes 1*, Johan Segers, *Université Louvain la Neuve*

A complete and user-friendly directory of tails of Archimedean copulas is presented which can be used in the selection and construction of appropriate models with desired properties. The results are synthesized in the form of a decision tree: Given the values of some readily computable characteristics of the Archimedean generator, the upper and lower tails of the copula are classified into one of three classes each, one corresponding to asymptotic dependence and the other two to asymptotic independence. For a long list of single-parameter families, the relevant tail quantities are computed so that the corresponding classes in the decision tree can easily be determined. In addition, new models with tailor-made upper and lower tails can be constructed via a number of transformation methods. The frequently occurring category of asymptotic independence turns out to conceal a surprisingly rich variety of tail dependence structures.

Un répertoire complet et convivial des queues de copules archimédiennes est présenté. Il peut être utilisé pour la sélection et la construction de modèles appropriés avec les propriétés voulues. Les résultats sont synthétisés sous forme d'arbre de décision : compte tenu des valeurs de certaines caractéristiques du générateur archimédien, chacune des queues inférieures et supérieures de la copule est classée dans une des trois catégories, l'une de ces catégories correspondant à la dépendance asymptotique et les deux autres à l'indépendance asymptotique. Pour une longue liste de familles à paramètre unique, les quantités pertinentes de queues sont calculées afin que les catégories correspondantes dans l'arbre de décision puissent être facilement déterminées. De plus, de nouveaux modèles avec des queues inférieures et supérieures personnalisées peuvent être créés grâce à un certain nombre de méthodes de transformation. On constate que la catégorie de l'indépendance asymptotique, qui se présente souvent, masque une diversité étonnamment riche de structures de dépendance de queue.

[MS-12]
Actuarial Aspects of Catastrophe Mortality Risk
Acpects actuariels du risque catastrophique de mortalité

Étienne Marceau & Hélène Cossette, *Université Laval*, Jean-Phillippe Lemay, *Standard Life Investments Inc.*

Catastrophe mortality risk may arise e.g. from natural catastrophes, extreme climatic conditions. Catastrophe mortality risk may impose important problems for a life insurer's solvency. We consider a general model to describe the occurrences of certain future catastrophic events and their intensities. The stochastic force of mortality of an insured is defined in terms of the occurrence and the intensity of the catastrophe. We examine the evaluation of the future costs for life insurance and annuity contracts. We use risk measures (e.g. VaR and TVaR) to assess the impact of mortality catastrophes on a life insurer's portfolio.

Le risque catastrophique de mortalité peut résulter de catastrophes naturelles ou de conditions climatiques extrêmes. Le risque catastrophique de mortalité peut causer des problèmes importants pour la solvabilité d'un assureur. On considère un modèle général pour décrire l'avènement de catastrophes futures et de leurs intensités. La force de mortalité stochastique d'un assuré est modélisée en fonction de l'avènement et de l'intensité d'une catastrophe. On examine l'évaluation des coûts pour des produits d'assurance vie et de rente. On utilise des mesures de risque (p.ex. : VaR, TVaR) pour quantifier l'impact financier du risque catastrophique sur le portefeuille d'un assureur.

[MS-13]
Weighted least Squares Estimation of the Extreme Value Index
Estimation de l'indice des valeurs extrêmes par la méthode des moindres carrés pondérés

Cécile Mercadier & Gabriela Ciuperca, *Université Lyon 1 (France)*

One of the most important problems in univariate extreme value theory is the estimation of the extreme value index (denoted by γ). We restrict the study to underlying distribution that belongs to the max-domain of attraction of the Fréchet distribution ($\gamma > 0$). We present (constrained and unconstrained) weighted least squares estimators of γ and derive estimators of extreme quantiles.

Their "quality" depends heavily on the number of order statistics considered in the procedure. Since the optimal threshold (according to the asymptotic mean square error) depends on a second order parameter, its estimation is also developed in this presentation.

L'estimation de l'indice des valeurs extrêmes (noté γ) est un des problèmes majeurs en statistique des extrêmes univariés. Pour cette étude, nous nous plaçons dans le domaine d'attraction de la loi de Fréchet ($\gamma > 0$). Des estimateurs de γ (contraints ou pas) construits par la méthode des moindres carrés pondérés sont présentes et utilisées pour l'estimation des quantiles extrêmes.

La « qualité » de chacune de ces procédures dépend fortement du nombre de statistiques d'ordre utilisées. Comme le seuil optimal (défini pour minimiser l'erreur en moyenne quadratique asymptotique) dépend d'un paramètre ρ dit de second ordre, l'estimation de ρ est également présentée dans cet exposé.

Session 02E Lun. 26 mai • Mon. May 26, 10:30 – 12:00 Capital Hall 5A

Predictive Models in Business and Industrial Statistics
Modèles de prévisions en statistique industrielle et de gestion

[MS-14]
Optimized Clusters for Disaggregated Electricity Load Forecasting
Classification optimisée pour la prévision désagrégée de la courbe de charge électrique

Jean-Michel Poggi, Michel Misiti, Yves Misiti & Georges Oppenheim, *Université d'Orsay*

In order to take into account the variation of the EDF (the French electrical company) portfolio due to the liberalization of the electrical market, it is essential to be able to conveniently disaggregate the global signal. The objective is to suitably disaggregate the global load curve obtained by summing up all the individual curves in order to improve the prediction performance. The strategy is to optimize a preliminary clustering of individual load curves with respect to a previsibility index. The optimized clustering scheme directed by forecasting is based on a cross-prediction dissimilarity index and a discrete gradient type algorithm.

Afin de prendre en compte la variation du portefeuille d'EDF (la société d'électricité française) induite par la libéralisation du marché électrique, il est crucial de pouvoir désagréger le signal global. L'objectif est de désagréger convenablement la courbe de charge globale obtenue en sommant toutes les consommations individuelles afin d'améliorer la qualité de la prévision. La stratégie proposée est d'optimiser, par rapport à un indice de prévisibilité, une première classification des courbes individuelles. Cette optimisation est basée sur un algorithme de type gradient discret dirigé par la prévision au moyen d'un indice de dissimilarité déduit d'un critère de prévision croisée.

[MS-15]
Load Curves Estimation and Simultaneous Confidence Bands
Estimation et intervalles de confiance simultanés de courbes de charge

Agnès Lagnoux Renaudie & Jean-Marc Azaïs, *LSP - Université Paul Sabatier - Toulouse France*, Sophie Bercu, *EDF France*, Jean-Claude Fort, *LSP - Université Paul Sabatier - Toulouse France*, Pierre Lé, *EDF France*

The load curve, a time series corresponding to power consumption, allows EDF to have a better knowledge about customer electricity use. This load curve is only available for customers with automated meter reading. For the others, EDF must estimate it to know the profitability of a customer and offer him an appropriate price. For that purpose, we construct a "decomposed" model (including the periodical behaviors) to estimate precisely the load curves and give simultaneous confidence intervals for these curves. This model will permit us to predict the load curve of a given customer based on the explanatory variables.

La courbe de charge (évolution de la puissance appelée par un client au cours du temps) permet à EDF de mieux connaître ses modes de consommation. Cependant, cette courbe n'est disponible que pour certains clients ayant un compteur. Pour les autres, EDF cherche à l'estimer afin de connaître leur rentabilité et leur offrir une tarification appropriée. Pour estimer au mieux ces courbes, nous construisons un modèle « décomposé » tenant compte des comportements périodiques et donnant des intervalles de confiance simultanés. Il permettra de prédire la courbe annuelle d'un individu quelconque à partir de ses seules variables explicatives.

[MS-16]

The Non-linear Link between Electricity Consumption and Temperature in Europe: a Threshold Panel Approach

Le lien non linéaire entre la consommation électrique et la température en Europe, une approche sur panel à seuil

Julien Fouquau, *LEO*, Marie Bessec, *CGEMP*

This paper investigates the relationship between electricity demand and temperature in the European Union. We address this issue by means of a panel threshold regression model on 15 European countries over the last two decades. Our results confirm the non-linearity of the link between electricity consumption and temperature found in more limited geographical areas in previous studies. By distinguishing between Northern and Southern countries, we also find that this non-linear pattern is more pronounced in the warm countries. At last, rolling regressions highlight the significant impact of climate change on electricity use in Europe.

Cet article étudie la relation entre la demande d'électricité et la température dans les pays de l'Union européenne. Notre démarche se situe sur un modèle à seuil sur 15 pays européens sur les vingt dernières années. Notre résultat confirme le lien non linéaire entre la consommation électrique et la température obtenu précédemment dans des zones géographiques plus limitées. En faisant la distinction entre les pays du nord et du sud, nous trouvons également que le comportement non linéaire est plus prononcé dans les pays chauds. Finalement, les régressions roulantes mettent en lumière l'impact significatif du changement climatique sur l'électricité en Europe.

[MS-17]

Non-parametric Electric Load Forecasting by "Aggregation/Disaggregation"

Prévision non paramétrique par "Agrégation/désagrégation" de la consommation électrique

Alain Dessertaine, *EDF R&D OSIRIS*

The short-term load forecasting by "aggregation/disaggregation" uses load individual curves after an appropriate elaboration of a clustering of these curves. In this framework, we will present a non-parametric kernel forecasting methodology which aggregates forecasts of each curve using its own past and the past of each curve of its own cluster. The weighting will be based on a kernel applied to the similarities of the reference window with all same size windows in the explored past, and the similarity of the curve with the other curves in its cluster.

La prévision court-terme par « agrégation/désagrégation » permet l'utilisation de courbes individuelles de consommation électrique via l'élaboration adaptée de partition de ces courbes. Dans ce cadre, nous présenterons une méthodologie de prévision non paramétrique à noyau afin d'agréger les prévisions de chaque courbe utilisant son propre passé et celui de chaque courbe de sa classe d'appartenance. La pondération sera basée sur un noyau appliqué sur la similarité de la fenêtre de référence avec toutes les fenêtres de même taille sur les historiques explorés, et sur la ressemblance globale de la courbe avec les autres courbes de sa classe.

Session 02F Lun. 26 mai • Mon. May 26, 10:30 – 12:00 Capital Hall 6A

Survey Applications
Applications en échantillonnage

[MS-18]
The Quarterly Financial Survey (QFS) for Enterprises
L'enquête financière trimestrielle (EFT) des sociétés

Pierre Daoust, Sylvie Gauthier & Nora Bohossian, *Statistique Canada*

The QFS is the main source of up-to-date information about corporate profits in Canada. Quarterly estimates include several hundred of the standard accounting variables. They are used by governments, banks, and other institutions, and they provide essential data for the Canadian System of National Accounts. The survey design includes a take-all stratum of the largest corporations, and a sampling of medium-sized corporations. The results are benchmarked to the Annual financial and Taxation Statistics figures, a census of financial data available from corporate income tax returns. In this presentation, we discuss the methodology of the QFS and future survey improvements.

L'EFT est la source principale d'information à jour relativement aux profits des sociétés canadiennes. Les estimations trimestrielles comprennent plusieurs centaines des variables types de comptabilité. Elles sont utilisées par les gouvernements, les banques, et autres établissements, et sont des données essentielles pour le Système de comptabilité nationale du Canada. Le plan de sondage de l'enquête comprend une strate à tirage complet des entreprises les plus importantes, et un sondage à tirage partiel des entreprises de taille moyenne. Les résultats sont étalonnés aux statistiques financières et fiscales annuelles (SFFA), un recensement de données financières est disponible grâce aux déclarations de revenus des sociétés. Lors de cette présentation, nous discuterons de la méthodologie de l'EFT et des améliorations futures de l'enquête.

[MS-19]
Childhood Weight and Cognitive Development: A Study of Young Children in Canada
Poids et développement cognitif durant l'enfance : une étude de la petite enfance au Canada

Xiaohong Peng, Rameshwar Gupta & Renjun Ma, *University of New Brunswick*

The prevalence of childhood and adolescent obesity has increased dramatically in Canada and worldwide in the most recent two decades. At the same time, there are still unacceptable proportions of children who are underweight in Canada. This study investigates the association between childhood weight and cognitive development among young children aged 4 and 5 years in Canada based on National Longitudinal survey of Children and Youth (NLSCY). This study finds that both seriously underweight and seriously obese children tend to have poorer cognitive development than their peers. Early interventions to these children's cognitive development may have an important impact on their future success.

La prévalence de l'obésité durant l'enfance et à l'adolescence a augmenté spectaculairement au Canada et ailleurs dans le monde au cours des deux dernières décennies. Parallèlement, le Canada compte encore une proportion inacceptable d'enfants ayant un poids insuffisant. Nous étudions l'association entre le poids et le développement cognitif chez les enfants de quatre et cinq ans au Canada en nous fondant sur des données provenant de l'Enquête longitudinale nationale sur les enfants et les jeunes (ELNEJ). Nous constatons que les enfants présentant une insuffisance pondérale grave ainsi que ceux présentant une obésité grave ont tendance à manifester un moins bon développement cognitif que leurs pairs. Une intervention précoce chez ces enfants concernant le développement cognitif pourrait avoir d'importants effets sur leur réussite future.

[MS-20]
Production Model for Rice in Madagascar
Modèle de rendement du riz à Madagascar

Salima Taibi, *LMRS Université de Rouen*, Manase Bezara, *Université de Tamatave Madagascar*, Candice Rouen, *INSA Rouen*

The purpose of this study is to determine a list of criteria for explaining rice production in the province of Tamatave in Madagascar. To find these criteria, we have based our study on the surveys that have been conducted. A number of techno-economical indicators have been observed, such as the global situation, production criteria for growing rice, commercialization trends, as well as the operation's trajectory, expectations and level of wealth. In total, 1,300 questionnaires were used. We applied various statistical and automatic learning methods to this issue. We have used the bagging method. A typology of these operations was also deduced.

Le but de l'étude est de déterminer un ensemble de critères permettant d'expliquer le rendement de riz dans la province de Tamatave à Madagascar. Pour trouver ces critères, nous nous sommes basés sur les enquêtes réalisées. Plusieurs indicateurs technicoéconomiques ont été retenus tels que la situation globale, les critères de productivité rizicole, les modes de commercialisation, la trajectoire de l'exploitation, les attentes et le niveau de richesse. Au total 1300 questionnaires ont été exploités. Nous avons appliqué diverses méthodes statistiques ou d'apprentissage automatique. Nous avons utilisé la méthode « bagging ». Une typologie de ces exploitations en a aussi été déduite.

[MS-21]
The Experience of Using Unincorportated Business Tax Data
L'utilisation des données fiscales : l'expérience des entreprises non incorporées

Javier Oyarzun, *Statistique Canada*

Statistics Canada has begun relying more and more on the use of administrative data for its various survey programs. T1 fiscal data provided by unincorporated enterprises are an example of the kinds of data used at SC in order to produce estimates for different economic sectors. At the same time, that sort of database requires a certain degree of statistical processing, such as outlier detection, imputation of missing variables and classification using statistical models. This presentation will describe the essential aspects involved in resolving methodological problems related to estimate production for the unincorporated enterprise population.

Statistique Canada (SC) accorde de plus en plus d'importance à l'utilisation des données administratives dans ses différents programmes d'enquêtes. Les données fiscales T1 provenant des entreprises non incorporées sont un exemple d'utilisation à SC afin de produire des estimations pour différents secteurs économiques. Toutefois, une telle base de données demande un certain traitement statistique, telle que la détection de valeurs aberrantes, l'imputation de variables manquantes et la classification à l'aide de modèles statistiques. Cet exposé décrira les aspects essentiels dans la résolution de problèmes méthodologiques pour la production des estimations pour la population des entreprises non incorporées.

[MS-22]
A New Synthetic Index for Qualitative Short-term Surveys
Un nouvel indicateur synthétique pour les enquêtes d'opinion

Olga Vasechko, *Research Institute of Statistics, Kyiv, Ukraine*, Michel Grun-Rehomme, *Université Paris 2, ERMES*

In this paper, we propose a new method of calculating a robust aggregative index for temporal repetitive surveys. This method use rank data with the median and a lexicographic ordering relation of

De nombreuses enquêtes d'opinion (comme les enquêtes qualitatives conjoncturelles) comportent des questions qualitatives ordonnées et corrélées pour estimer un facteur inobservable. On propose un nouveau mode de calcul d'un in-

the individual data. A confidence interval is constructed around the aggregative index. The indicator's advantage is in simplicity of calculation. This paper studies the effectiveness of this approach. We compare our method with two other often-used approaches. Using a unique set of business and consumer data from the Ukraine, we compare a synthetic index with the dynamic factor approach for business data and with the score method for consumer data.

dice synthétique robuste, qui utilise une statistique de rang avec la médiane et qui est fondée sur une relation d'ordre lexicographique à partir des données individuelles. Un intervalle de confiance est construit autour de cette statistique. Il est mis en œuvre d'une part sur des données conjoncturelles ukrainiennes dans l'industrie et d'autre part sur des données d'une enquête d'opinion. Il est comparé aux autres méthodes (analyse factorielle dynamique, score).

[MS-23]

Using Paradata to Evaluate Collection Activities in the Youth in Transition Survey
L'utilisation des paradonnées pour évaluer les activités de collecte dans l'Enquête auprès des jeunes en transition

Tracy Tabuchi, *Statistics Canada*

The Youth in Transition Survey is a longitudinal survey that gathers information regarding school-work transitions and factors that affect them. Collection occurs every two years via a Computer Assisted Telephone Interview (CATI). The complexity of the questionnaire can lead to misinterpretation either by the interviewer or respondent; this can in turn affect data quality. Audit Trail files, which record all actions performed during a given CATI session, were examined to assess common errors, edit pop-ups and application functionality. Areas for potential improvement were identified in hopes of improving data quality for future collections. Results are presented using Cycle 4 data.

L'Enquête auprès des jeunes en transition est une enquête longitudinale qui rassemble de l'information sur les transitions école-marché du travail et les facteurs qui les affectent. La collecte est effectuée tous les deux ans par l'entremise de l'interview téléphonique assistée par ordinateur (ITAO). La complexité du questionnaire peut mener à de fausses interprétations tant du point de vue de l'interviewer que des répondants ; ce qui peut en retour affecter la qualité des données. Les dossiers de vérifications à rebours, qui enregistrent toutes les actions effectuées au cours d'une séance ITAO, ont été examinés pour évaluer les erreurs communes, vérifier les fenêtres instantanées et les fonctionnalités d'application. Les secteurs où des améliorations seraient possibles ont été identifiés dans l'espoir d'améliorer la qualité des données pour les activités de collecte futures. Les résultats sont présentés à partir des données du Cycle 4.

Session 02G Lun. 26 mai • Mon. May 26, 10:30 – 12:00 Capital Hall 7A

Nonparametric Regression
Régression non paramétrique

[MS-24]

A Quantile-Copula Approach to Conditional Density Estimation
Estimation de la densité conditionnelle par quantiles et copules

Olivier Faugeras, *LSTA - Université Pierre et Marie Curie - Paris 6*

We present a new non-parametric estimator of the conditional density of the kernel type. It is based on an efficient transformation of the data by quantile transform. By use of the copula representa-

Nous présentons un nouvel estimateur non paramétrique à noyau de la densité conditionnelle. Il est basé sur une transformation efficace des données par transformation de quantile. L'utilisation de la fonction de copule permet alors

tion, it turns out to have a remarkable product form, compared to the quotient-shaped competitors of the Nadaraya-Watson type, based on nonparametric regression. We study its asymptotic properties and perform a comparative analysis of its bias and variance.

d'avoir un estimateur qui a une forme remarquable en produit alors que les estimateurs basés sur la régression de type Nadaraya-Watson ont une forme en quotient. Nous étudions ses propriétés asymptotiques et comparons son biais et sa variance aux estimateurs concurrents.

[MS-25]

Smoothing Non-Equispaced Heavy Noisy Data with Wavelets
Estimation par ondelettes pour des données fortement bruitées et non equiespacées

Jean-Michel Poggi, *Université d'Orsay*, Anestis Antoniadis, *Laboratoire Jean Kuntzmann, Université Joseph Fourier*, Irene Gijbels, *Department of Mathematics and Leuven Statistics, Research Centre (LStat)*

We consider a nonparametric noisy data model observed on deterministic, not necessarily equispaced design points. The unknown signal belongs to a wide range of function classes, while the additive zero-median noise distribution is unknown and possibly heavy tailed. As in Brown, Cai, Zhou (2007), using local medians, we construct new variables structured as a Gaussian nonparametric regression, but contrary to these authors, the resulting data are not equispaced. Using a wavelet block penalizing procedure we construct an estimator attaining the optimal rate of convergence over a wide range of Besov classes. The performances of our robust procedure are evaluated on datasets.

On considère un modèle de régression non paramétrique. Le signal inconnu appartient à une large classe d'espaces fonctionnels. Le bruit additif est de loi inconnue pouvant être à queues lourdes. Comme dans Brown, Cai, Zhou (2007), des médianes locales construisent de nouvelles variables structurées comme un modèle de régression non paramétrique gaussienne mais, contrairement à ces auteurs, les nouvelles données ne sont pas équi-espacées. Par une méthode d'ondelettes de pénalisation par blocs, on construit un estimateur atteignant la vitesse de convergence optimale pour une large classe d'espaces de Besov. Les performances de notre procédure robuste sont évaluées sur des données.

[MS-26]

Asymptotic Behaviour of the Minimax Risk in Nonparametric Heteroscedastic Regression Models
Comportement asymptotique du risque minimax dans des modèles de régression non paramétriques et hétéroscédastiques

Jean-Yves Brua, *Institut de Recherche Mathématique Avancée*

In nonparametric heteroscedastic regression models, we are interested in the asymptotic behaviour of the minimax risk as the regression function is estimated pointwise for the absolute error loss. The smoothness of the regression function is initially assumed to be known. An asymptotically efficient kernel estimator is proposed when the noises are Gaussian or have unknown distribution. Then, in the case where the smoothness of the regression function remains unknown, an adaptive estimator which attains the minimax rate is constructed.

Dans des modèles de régression non paramétriques et hétéroscédastiques, nous nous intéressons au comportement asymptotique du risque minimax, utilisant la fonction de perte absolue quand la fonction est estimée en un point fixe. La régularité de la fonction de régression est supposée connue dans un premier temps. Un estimateur à noyau asymptotiquement efficace est alors proposé dans le cas où les bruits sont gaussiens ou ont une distribution inconnue. Par la suite, dans le cas où la régularité de la fonction est inconnue, on construit un estimateur à noyau adaptatif qui atteint la vitesse minimax.

[MS-27]
Multiplicative Bias Corrected Nonparametric Smoothers
Correction multiplicative du biais en régression

Laurent Rouvière, *Université Rennes 2,* Nicolas Hengartner, *Los Alamos National Laboratoty*, Eric Matzner-Lober, *Université Rennes 2*

Bias estimation is an important question in nonparametric estimation in density or regression. We propose a simple method which allows us to reduce the bias of regression estimators. The method consists of applying a multiplicative bias correction to a pilot estimator. Under classical conditions on the regularity of the regression function, we show that the estimate has zero bias and exactly the same asymptotic variance as the kernel estimate. We illustrate the performances of our estimate with the help of a finite sample simulation study.

Les estimateurs non paramétriques de la densité ou de la fonction de régression sont en général asymptotiquement biaisés. Nous étudions une méthode simple à mettre en œuvre et permettant de réduire le biais d'estimateurs non paramétriques de la fonction de régression. L'approche consiste à appliquer une correction multiplicative du biais d'un estimateur pilote. Sous des hypothèses classiques concernant la régularité de la fonction de régression, l'estimateur proposé possède un biais asymptotique nul et une variance identique à celle de l'estimateur à noyau. Les résultats présentés seront illustrés sur des données simulées.

[MS-28]
Estimating the Support Function of a Convex Region using Regression Splines
Estimation de la fonction de support d'une partie convexe moyennant les splines de régression

Johan Segers, *Université catholique de Louvain,* Mary C. Meyer, *Colorado State University*

The problem of reconstructing a convex region in the plane given noisy measurements of its support function has applications in medical imaging and robotic vision. From a statistical point of view, this is a nonparametric regression problem under the shape constraint that the regression function is the support function of some convex region. We propose an estimator using regression splines. While the estimator satisfies the shape constraint, at the same time it achieves the optimal point-wise convergence rate for the unconstrained case. The problem is formulated as a projection onto a convex cone, which provides an efficient algorithm.

Le problème de la reconstruction d'une partie convexe du plan euclidien étant donné des observations de sa fonction de support est pertinent dans les sciences médicales et la vision robotique. Du point de vue statistique, il s'agit d'un problème de régression non paramétrique avec la contrainte que la fonction de régression soit la fonction de support d'une partie convexe. Un estimateur est proposé et est basé sur les splines de régression. Tandis que l'estimateur vérifie la contrainte de forme, en même temps sa vitesse de convergence est optimale dans le cas où il n'y a pas de contraintes. En formulant l'estimateur comme une projection sur un cône convexe, un algorithme efficace est obtenu.

[MS-29]
Hypothesis Testing for the Equality of Nonparametric Regression Functions
Test d'hypothèse de l'égalité de fonctions de régression non paramétriques

Zaher Mohdeb, *Université Mentouri de Constantine, Algérie*

We propose a new procedure for testing the comparison of two regression functions f_1 and f_2 in the case of homoscedastic error structure and a fixed design. Our approach is based on the empirical Fourier coefficients of the regression functions f_1 and f_2 respectively. As our main results, we obtain the asymptotic distribution of the test statistic un-

Nous proposons une nouvelle procédure des test d'hypothèses de comparaison deux fonctions de régression f_1 et f_2 dans le cas où les erreurs sont homoscédastiques et un échantillonnage fixé. Notre approche est basée sur les coefficients de Fourier respectivement des fonctions de régression f_1 et f_2. Nous obtenons la loi asymptotique de la statistique de test sous l'hypothèse nulle $f_1 = f_2$ et sous les alternatives

der the null hypothesis $f_1 = f_2$ and local and global alternatives. A simulation study is conducted to investigate the finite sample performance of our test.

globales ainsi que les alternatives locales. Une étude par simulation est menée pour vérifier la performance des résultats obtenus pour les petites tailles d'échantillons.

Session 02H Lun. 26 mai • Mon. May 26, 10:30 – 12:00 Capital Hall 8A

Estimation and Hypothesis Tests
Estimation et tests d'hypothèses

[MS-30]
Statistical Models for $SE(p)$ Data
Modèles statistiques pour les données de $SE(p)$

Karim Oualkacha & Louis-Paul Rivest, *Université Laval*

This talk investigates a new statistical model for data in $SE(p)$, the set of all rigid displacements in $\mathbb{R}^p$. A displacement involves a rotation and a translation. The new model expresses the displacement M as a true displacement M_0 perturbed by experimental errors. Some errors specifications are discussed. We propose an estimator of the mean rotation that uses both the rotation and the translation data. This estimator is compared to the rotation only estimator when the rotation errors have the symmetric matrix Fisher-Von Mises distribution, when $p = 2, 3$. A numerical example is also presented.

Cet exposé porte sur un nouveau modèle statistique pour les déplacements dans $SE(p)$, l'ensemble de tous les déplacements dans $\mathbb{R}^p$. Un déplacement implique une rotation et une translation. Le nouveau modèle exprime le déplacement M comme un vrai déplacement M_0 perturbé par des erreurs expérimentales. Quelques spésifications des erreurs sont discutées. Nous proposons un estimateur de la rotation moyenne qui utlise à la fois les données rotation et translation. Cet estimateur est comparé avec l'estiamteur obtenu par des rotations seulement quand les erreurs de rotation sont distribuées selon la distribution de Fisher-Von Mises, quand $p = 2, 3$. Un exemple numérique est ainsi présenté.

[MS-31]
Model-based Bootstrapping for M-estimators
Le bootstrap basé sur des modèles pour des M-estimateurs

Angelo Canty & Shaheena Bashir, *McMaster University*

In small sample situations, the bootstrap can be used to get standard errors and confidence intervals for parameters of the linear model estimated using M-estimation. Much of the literature on this topic suggests a case-resampling approach. For designed experiments, however, the use of model-based residual resampling is more appropriate. As with least squares estimation this involves standardization of the robust residuals to have equal variance. We examine some difficulties that arise in the implementation of the bootstrap and residual standardization and compare a number of different approaches for the bootstrap in this context.

En présence de petits échantillons, le bootstrap peut être utilisé afin d'obtenir des écarts-types et intervalles de confiance pour les paramètres du modèle linéaire estimé au moyen de la M-estimation. Une part importante de la littérature à ce sujet suggère une approche de rééchantillonnage de cas. Pour des expériences planifiées, cependant, l'utilisation de rééchantillonnage de résidus à base de modèles est plus appropriée. Comme dans le cas de l'estimation par moindres carrés, ceci implique de standardiser les résidus robustes afin qu'ils soient de variance égale. Nous examinons certaines difficultés qui surviennent dans la mise en œuvre du bootstrap et de la standardisation des résidus et comparons différentes approches pour le bootstrap dans ce contexte.

[MS-32]
Some Applications of Graph Theory to Statistical Graphics
Certaines applications de la théorie des graphes aux graphiques statistiques

Richmond Oldford, *University of Waterloo*, Catherine Hurley, *National University of Ireland (Maynooth)*

In this presentation, we show how Hamiltonian paths, Eulerian tours, and Hamiltonian decompositions of complete graphs have a role to play in statistical graphics. Examples are used to illustrate their application. In particular, we present a new plot for multiple comparisons, one which is both simpler and more informative than those currently available. A second example addresses the use of glyphs for visualizing high dimensional data. Some of the relevant mathematics of graph theory will be reviewed.

Dans cette présentation nous montrons comment les chemins hamiltoniens, les cycles eulériens et les décompositions hamiltoniennes de graphes complets ont un rôle à jouer dans la visualisation graphique statistique. Les exemples sont utilisés pour illustrer leur application. Plus précisément nous présentons un nouveau graphique pour les comparaisons multiples, une qui est plus simple et plus instructif que ceux qui sont actuellement disponibles. Un deuxième example traite de l'utilisation de *glyphs* pour visualiser des données de hautes dimensions. Certaines des théories mathématiques pertinentes des graphes seront reconsidérées.

[MS-33]
Regularized Solutions of Fredholm Integral Equations on the Euclidean Motion Group
Solutions régularisées aux équations de l'intégrale de Fredholm sur le groupe des déplacement euclidiens

Maia Lesosky & Peter Kim, *University of Guelph*

This work solves a deconvolution problem on the Euclidean motion group. Regularization techniques are used to handle the ill-posed nature of the problem. Upper bounds for the mean integrated squared error are calculated for a number of cases.

Ce travail résout un problème de déconvolution sur le groupe des déplacements euclidiens. Des techniques régularisées sont utilisées pour traiter la nature mal posée du problème. Des bornes supérieures pour l'erreur quadratique intégrée moyenne sont calculées pour un certain nombre de cas.

[MS-34]
Invariant P-values for Model Checking and Checking for Prior-data Conflict
Valeurs-p invariantes pour la vérification de modèles et la détection de conflits données-*a priori*

Gun Ho Jang & Michael Evans, *University of Toronto*

Model checking and checking for prior-data conflict involve comparing the observed value of a statistic against some known distribution. If it is a surprising value, then we have evidence against the sampling model or evidence against no prior-data conflict, respectively. The use of P-values seems appropriate for measuring surprise, but there are some ambiguities associated with this approach. We aim at a resolution of some of these problems by providing a method for computing P-values based upon measurement characteristics of the observed data. These P-values are applied in several problems.

La vérification de modèles et la détection de conflits données-*a priori* impliquent souvent de comparer la valeur observée d'une statistique à une certaine distribution connue. S'il s'agit d'une valeur surprenante, nous avons alors une justification contre le modèle échantillonnal ou contre l'absence de conflit données-*a priori*, respectivement. L'utilisation de valeurs-p semble appropriée pour mesurer la surprise, mais il y a certaines ambiguïtés dans cette approche. Nous cherchons à résoudre certains de ces problèmes en fournissant une méthode de calcul des valeurs-p basée uniquement sur des caractéristiques de mesure des données observées. Ces valeurs-p sont appliquées à plusieurs problèmes.

[MS-35]
Construction of Bivariate Charts Using Quantile Contours
Construction de graphiques bivariées en utilisant les contours du quantile

Linglong Kong & Ivan Mizera, *University of Alberta*

Quantiles are useful in lots of fields and have been studied in many contexts. In this talk, we will present a method to construct quantile contours. The proposed multivariate quantiles are a direct extension of their univariate counterparts, and hence share many common merits, like immediate data-analytic interpretation, easy computation (in bivariate cases) and so on. The performance of this approach is demonstrated through an application to joint height-weight screening. We construct the bivariate growth charts and use them to locate an individual subject's percentile rank. The constructed chart provides more informative readings than conventional univariate growth, BMI and Rohrer Index charts.

Les quantiles sont utiles dans de nombreux domaines et ont été étudiés dans de multiples contextes. Nous présentons une méthode de construction des contours de quantile. Les quantiles multivariés proposés sont une extension directe de leur version univariée et, donc, ont en commun avec celle-ci de nombreux avantages, tels que l'interprétation analytique directe des données, la facilité de calcul (dans les cas bivariés) et ainsi de suite. Nous illustrons la performance de cette approche à l'aide d'une application de tri conjoint selon la taille et le poids. Nous construisons les graphiques de croissance bivariée et les utilisons pour repérer le rang centile d'un sujet particulier. Leeu graphique construit fournit des mesures plus informatives que les graphiques classiques de croissance univariée, d'IMC et d'indice de Rohrer.

Session 02I Lun. 26 mai • Mon. May 26, 10:30 – 12:00 Capital Hall 9A

Teaching of Statistics and Applications
Enseignement de la statistique et applications

[MS-36]
FactoMineR an R Package for Exploratory Data Analysis for Teaching and Research
FactoMineR : un ensemble R d'analyses exploratoires des données pour l'enseignement et la recherche

François Husson, Julie Josse & Lê Sébastien, *Agrocampus Rennes*

We will show how to use the FactoMineR package for teaching. This package allows us to analyse multivariate data taking into account several types of variables (continuous and/or categorical), several types of structures on the data (structure on the individuals, structure on the variables or hierarchy on the variables). Classical methods of exploratory data analysis (Principal Component Analysis, Correspondence Analysis and Multiple Correspondence Analysis) as well as advanced methods (Multiple Factor Analysis, Hierarchical Multiple Factor Analysis, Dual Multiple Factor Analysis) are implemented. A Graphical User Interface makes this package accessible to non-users of R and allows us to work on easily exportable interactive graphs.

Nous montrerons comment utiliser la librairie de fonctions FactoMineR en enseignement. Cette librairie permet d'analyser des données multidimensionnelles en prenant en compte différents types de variables (qualitative ou quantitative), différents types de structure sur les données (structure sur les individus ou les variables ou hiérarchie sur les variables). Les principales méthodes d'analyse des données (ACP, AFC, ACM) ainsi que des méthodes avancées (analyse factorielle multiple, analyse factorielle multiple hiérarchique, analyse factorielle multiple duale) sont implémentées. Une interface graphique permet de rendre accessible cette librairie à des non-utilisateurs de R et de travailler sur des graphiques interactifs facilement exportables.

[MS-37]
Dispelling Statistical Myths: Consulting and Collaboration.
Dissiper les mythes statistiques : la consultation et la collaboration

AnneMichele Millar, *Mount Saint Vincent University*

Often the statistician is called in at the end of the day — the participants for the study have been chosen, treatments imposed, and the resulting data has been both collected and inspected. While some of these studies are well designed and well implemented, a misunderstanding of statistical issues, or worse, a strong belief in a statistical fallacy, can mean that we take on the role of rescuing rather than analyzing data. The speaker will relate this to her experience as a consultant and collaborator in an academic environment.

Souvent, le statisticien est appelé en bout de ligne, quand les participants à l'étude ont été choisis, les traitements, imposés et les données résultantes, recueillies et inspectées. Quoique certaines de ces études soient bien conçues et bien mises en Suvre, une mauvaise compréhension de questions statistiques, ou pire, une ferme croyance dans un faux raisonnement statistique, peut signifier que nous nous retrouvons dans le rôle de sauveteur plutôt que d'analyste des données. La conférencière discutera de ce qui précède dans le contexte de son expérience en tant que conseillère et de collaboratrice dans un milieu universitaire.

[MS-38]
French Social Science and Humanities Undergraduate Representations of the Statistics
Représentation de la statistique chez des étudiants de premier cycle universitaire en sciences humaines et sociales

Alain Bihan-Poudec, *Université Catholique de l'Ouest (France)*, François Larose, *Université de Sherbrooke*

In the last years, Statistics' education has appeared progressively as a research domain. If many authors underline the importance of the affects and their impacts on learning in statistics, only a few considered Shaughnessy's (2006) proposal to base teaching on learners'reality. It is on the very basis of this proposal that we conducted a survey with 700 social sciences and humanities undergraduate students in France. The survey was aimed toward the identification of their social representation of statistics. In this presentation, we will see up to which point these representations rely upon their experiences in previous studies, particularly in statistics.

Dans les dernières années, l'enseignement de la statistique s'est progressivement constitué comme champ de recherche. Si nombre de travaux soulignent l'importance de la dimension affective ou en pointent les nombreuses erreurs dans cet apprentissage, peu répondent à la suggestion de Shaughnessy (2006) de partir des apprenants eux-mêmes. Dans cette perspective a été réalisée en France une enquête auprès de 700 étudiants de premier cycle universitaire en sciences humaines et sociales afin de connaître les représentations qu'ils se faisaient de la statistique. La communication vise à les préciser et à voir si elles dépendent du type d'études suivies et de l'expérience antérieure de la statistique.

[MS-39]
Service-learning in Undergraduate Statistics Education
Programme d'apprentissage communautaire dans les études de premier cycle en statistique

Bethany White & John Braun, *The University of Western Ontario*

Service-learning is an experiential educational tool which involves the use of community-based projects to enhance learning by giving students the opportunity to learn through experience and therefore demonstrating the value of the course material. Several American schools have introduced

L'apprentissage par le service communautaire est un outil pédagogique expérientiel qui comprend le recours à des projets communautaires pour rehausser l'apprentissage en donnant aux étudiants la possibilité d'apprendre par l'expérience et, par conséquent, de saisir la valeur de la matière des cours. Plusieurs écoles américaines ont introduit l'apprentissage

service-learning in introductory statistics courses (Anderson & Sungur 1999, Root & Thorme 2001, Nordmoe 2007, and Hydorn 2007) and have observed positive effects on student engagement. The purpose of this talk is to share our experiences in implementing and evaluating a service-learning project in a third-year statistics course. The benefits and challenges of this endeavour and student reaction will be discussed.

par le service dans les cours de statistique de niveau élémentaire (Anderson et Sungur, 1999 ; Root et Thorme, 2001 ; Nordmoe, 2007 ; Hydorn. 2007) et ont observé des effets positifs sur la motivation des étudiants. Le but de la communication est de faire part de nos expériences de mise en Suvre et d'évaluation d'un projet d'apprentissage par le service dans un cours de statistique de troisième année. Les avantages de cette entreprise, les défis qu'elle a posés et les réactions des étudiants seront discutés.

[MS-40]

How Numbers Reveal and Conceal the Truth in Capturing Patterns in Qualitative Data?
De quelle façon les nombres révèlent-ils et cachent-ils la vérité dans la saisie de modèles pour les données qualitatives ?

Muhammad Islam, *MacEwan College*

The materials in this paper originate from disagreement of a solution of a lab assignment in teaching an introductory statistics course. Display tools such as stacked bar graphs and mosaic plots are important tools to capture patterns in qualitative data. In this paper, the visual representation of categorical data is discussed by the use of areas. To this end, some examples are presented to show how areas and not the frequencies reveal patterns of relation using published data. It is also shown that graphical displays should be meaningful when they are used in conjunction with a probability model.

Le sujet de la présente communication émane d'un désaccord concernant la solution d'un devoir de laboratoire donné dans le cadre d'un cours d'introduction à la statistique. Les outils d'affichage, tels que les graphiques à barres empilées et les diagrammes en mosaïque, sont importants pour dégager les modèles dans les données qualitatives. La communication porte sur la représentation visuelle de données catégoriques à l'aide d'aires. Des exemples sont présentés afin de montrer comment les aires et non les fréquences révèlent les modèles de relation dans des données publiées. Nous montrons aussi que les affichages graphiques utilisés en même temps qu'un modèle probabiliste devraient être sensés.

[MS-41]

Science (I.Statistics; II.Applied Mathematics; III. Pure Mathematics): Viewed via Newton and Buffon
La Science (I. La Statistique ; II. Les Mathématiques Appliquées ; III. Les Mathématiques Pures) : Vue dans les œuvres de Newton et de Buffon

Danielle Mihram, *University of Southern California*, G. Arthur Mihram, *Princeton, NJ*

Science is that human activity devoted to the search for the very explanation of (for the truth about) any particular naturally occurring phenomenon. How does each of the activities (I; II; III) qualify as science?: I. Do regression theory and hypothesis-testing deceive us as Science, yet they encourage practitioners to model before having observed? II. Do applied mathematicians presume (unjustifiably) that the Scientific Method is an exact mimicry of theorem-proving (assumptions; deduction; conclusion), the activity of the pure mathematician (III)? Illustrations: the respective successes of Newton and Buffon, the latter for both the binomial theo-

La science est cette activité humaine vouée à la recherche de l'explication exacte (c'est-à-dire, à la recherche de la vérité)sur n'importe quel phénomène précis qui se produit naturellement. Comment caractériser scientifiquement chacune des activités (I ; II ; III) ? I. Est-ce que la régression et le test d'hypothèse nous dupent en passant comme Science, tout en encourageant les praticiens à créer des modèles avant d'avoir observé le phénomène en question ? II. Est-ce que les mathématiciens en mathématiques appliquées présument (injustifiablement) que la méthode scientifique est une imitation exacte de la preuve de théorèmes (hypothèses ; déduction ; conclusion), l'activité des mathématiciens en mathématiques pures (III) ? Illustrations : Les succès respectifs de Newton

rem and his role as the Royal Gardener?

et de Buffon, et, pour ce dernier, non seulement le théorème binomial, mais aussi son rôle d'Intendant du Jardin Royal des Plantes de Paris.

Session 03A Lun. 26 mai • Mon. May 26, 13:30 – 14:30 Congress Hall A/E

Lucien Le Cam Address
Conférence Lucien Le Cam

[MS-42]
Hunting Serial Killer Nurses with Statistics
Poursuite de d'infirmières tueuses en série avec statistique

Richard Gill, *Leiden University, Netherlands*

I will analyse scientific evidence in the case of the Dutch nurse Lucia de Berk, serving a life sentence for 7 murders and 3 attempts in hospitals in The Hague during 1996 – 2001. I will argue that medical evidence for the conviction was, for one part, wrong, and for the other part, contaminated by bad statistics; statistical evidence was, for one part, wrong, and for the other part, contaminated by incorrect medical information. No other serious evidence exists. This logical circle is locked down by the legal institutions of the Netherlands, and it appears that they have thrown away the key.

J'analyserai la preuve scientifique dans le cas de l'infirmière hollandaise Lucia de Berk, purgeant une condamnation à perpétuité pour 7 meurtres et 3 tentatives de meurtre aux hôpitaux à La Haye entre 1996 et 2001. Je soutiendrai que la preuve médicale de la condamnation était fausse dans une partie et contaminées par la mauvaise statistique dans une autre partie ; la preuve statistique était fausse dans une partie et contaminée dans une autre partie par des informations médicales incorrectes. Il n'existe aucune autre preuve sérieuse. Ce cercle logique est verrouillé par les institutions légales des Pays-Bas et il semble qu'ils aient jeté la clé.

Session 04A Lun. 26 mai • Mon. May 26, 14:35 – 16:05 Capital Hall 1A

Predicting Modeling Applications in Customer Acquisition and Retention
Prévision de la modélisation des applications en acquisition et rétention de clients

[MS-43]
Traps to Avoid when Building a Sales Forecasting Model
Pièges à éviter lors de la conception d'un modèle de prévision des ventes

Daniel Lemire, *Indicia*

The presentation will focus the geo-marketing application of statistical modeling. Over the years, it was observed that companies have been unable to build effective sales forecasting models. In fact, many business people are cynical about using statistics to analyze their retail/distribution network and predict profitability.

After discussions with corporate managers, identification of common elements about statistical modeling was conducted. It was observed that similar mistakes were repeated from one company

La communication portera sur l'application de la modélisation statistique en géomarketing. Au fil des ans, on a constaté que les sociétés n'arrivaient pas à construire de bons modèles de prévision des ventes. En fait, nombre de gens d'affaires ont une vision cynique de l'utilisation des statistiques pour analyser leurs réseaux de vente au détail ou de distribution et pour prédire leur rentabilité.

Après plusieurs discussions avec des directeurs d'entreprise, nous avons relevé les éléments communs dans la modélisation statistique. Nous avons constaté que les mêmes erreurs se répétaient d'une société à l'autre. La communication

to the next. The presentation will cover a list of the most common errors committed in the creation of forecasting models used to evaluate the viability of future business sites.

portera sur une liste des erreurs les plus fréquemment commises dans la création des modèles de prévision utilisés pour évaluer la viabilité de futurs emplacements commerciaux.

[MS-44]

Measures that Matter — Driving Predictive Model Results to the Bottom Line
Des mesures qui importent – passer des résultats d'un modèle de prévision au résultat net

Olivia Parr-Rud, *OLIVIAGroup*

Traditionally, predictive models are evaluated using measures such as the KS, Gini and/or the Chi-square Goodness of Fit test. However, these statistics have limitations when evaluating models for profit-driven business applications such as acquisition and retention. As general measures of fit, they fail to allow for comparison between models that optimize for specific targets, channels, timing or revenue goals. Fortunately, several powerful revenue and ROI-driven model evaluation methods exist that allow the user to fine tune the selection process. In this presentation, we will explore these methods and provide industry examples for selecting the optimal model for a variety of applications.

Traditionnellement, les modèles de prévision sont évalués par des mesures telles les tests de manque d'ajustement khideux, Gini ou KSD. Cependant, ces statistiques ont des limites lorsqu'elles évaluent des modèles pour des applications d'entreprises motivées par le profit, telles l'acquisition et la rétention. Comme mesures générales d'ajustement, elles ne permettent pas de comparer deux modèles optimisés pour des cibles, conduits ou buts de revenus spécifiques, ou pour la détermination de moments propices. Heureusement, plusieurs puissantes méthodes d'évaluation de modèles centrées sur le revenu et le rendement du capital investi existent et permettent à l'utilisateur d'ajuster le processus de sélection. Dans cette présentation, nous explorons ces méthodes et présentons des exemples de l'industrie de sélection dle modèle optimal pour une variété d'applications.

[MS-45]

Business-to-Business Applications of Predictive Modeling
Applications de la modélisation de la prévision de compagnie à compagnie

Emma Warrillow, *Emma Warrillow & Associates Inc.*

Predictive modelling has been extensively used in the business-to-consumer space for many years. However, its use in business-to-business is more recent, but growing in importance.

In this session, Emma will provide some examples of what companies — in a variety of industries — are doing to use predictive modelling to target their dealings with business customers. She will highlight some challenges they face and the impact of their efforts.

La modélisation prédictive a été largement utilisée dans un contexte entreprise-consommateur depuis plusieurs années. Cependant, son utilisation dans un contexte interentreprises est plus récente, mais d'importance croissante.

Dans cette session, Emma apportera quelques exemples de ce que les entreprises – dans une variété d'industries – font afin d'utiliser la modélisation prédictive afin de cibler leurs activités auprès de leur clientèle commerciale. Elle soulignera certains défis auxquels elles font face et l'impact de leurs efforts.

Session 04B Lun. 26 mai • Mon. May 26, 14:35 – 16:05 Capital Hall 2A

Combining Survey and Administrative Data
Combinaison de données administratives et de données d'enquêtes

[MS-46]

The Role of the Estimators for the Production of the French Structural Business Statistics
Le rôle des estimateurs dans la production des statistiques françaises structurelles d'entreprises

Philippe Brion, *INSEE*

INSEE is planning to redesign the French structural business statistics, using administrative data (tax data, data about employees and wages, customs data) in an intensive way and completing them with a survey conducted on a sample. The paper presents the method that is intended to be used to produce estimates — especially sector-based estimates —, which is based on the two types of data (administrative and coming from the survey), but also the classifying of enterprises within the business register.

L'INSEE mène un projet de remaniement des statistiques structurelles d'entreprises françaises s'appuyant sur une utilisation intensive des données administratives (données fiscales, données d'emploi et de salaires, données douanières) complétée par une enquête menée sur un échantillon. L'étude présente la méthode de production d'estimations préconisée, en particulier pour les statistiques sectorielles, qui s'appuie sur les deux types de données (administratives et obtenues par l'enquête), mais également sur le classement des unités dans le répertoire d'entreprises.

[MS-47]

Estimating Monthly Sales by Combining Survey Data with Administrative Data in the Monthly Survey of Food Services and Drinking Places
L'estimation des ventes mensuelles en combinant les données d'enquête et les données administratives de l'Enquête mensuelle sur les services de restauration et débits de boisson

Stuart Pursey & François Brisebois, *Statistics Canada*

The Monthly Survey of Food Services and Drinking Places of Statistics Canada provides estimates of monthly sales and the number of locations. Recently this survey has been redesigned and a major element of the redesign is the extensive use of administrative data. The goal of our discussion is to describe the approaches used to combine collected sample survey data from this survey with data from the Goods and Service Tax (GST), thus providing a way of maintaining data quality while reducing survey costs and reducing the burden on respondents.

L'Enquête mensuelle sur les services de restauration et débits de boissons de Statistique Canada fournit des estimations des ventes mensuelles et du nombre d'emplacements. Récemment, cette enquête a subi un remaniement dont un des éléments majeurs est l'utilisation étendue des données administratives. L'objectif de notre discussion est de décrire les approches utilisées pour combiner les échantillons de données d'enquête recueillis avec les données de la taxe sur les produits et services (TPS), et ainsi fournir un moyen de maintenir la qualité des données tout en réduisant les coûts liés à l'enquête et le fardeau du répondant.

[MS-48]

Regression Analysis of Linked Files
Analyse de régression des fichiers couplés

Michael Larsen, *Iowa State University*, Partha Lahiri, *University of Maryland*

Record linkage can be used to join together two files that contain information on the same individ-

Le couplage d'enregistrements peut être utilisé pour réunir deux fichiers qui contiennent des renseignements sur les

uals, but lack unique personal identification codes. The possibility of errors in linkage causes problems for estimating the relationships between variables on the two files. The effect is analogous to the impact of measurement error. If the model is a linear regression model and linkage error probabilities can be estimated, then an unbiased estimator of the regression coefficients and its standard error can be computed. Methods for estimating linkage probabilities and adjusting regression coefficients are described. Extensions of the method are suggested.

mêmes personnes, mais où ne figure pas de code d'identification personnel unique. La possibilité que des erreurs aient lieu dans le couplage pose des problèmes pour l'estimation des relations entre les variables des deux fichiers. L'effet est analogue à celui d'une erreur de mesure. Si l'on a affaire à un modèle de régression linéaire et que l'on peut estimer les probabilités des erreurs de couplage, il est possible de calculer un estimateur sans biais des coefficients de régression et son erreur-type. Nous décrivons des méthodes d'estimation des probabilités de couplage et d'ajustement des coefficients de régression. Enfin nous proposons des extensions de la méthode.

[MS-49]
Predicting New Cases of Diabetes in Canada
Prédiction des nouveaux cas de diabètes au Canada

Therese Stukel, Doug Manuel & Laura Rosella, *Institute for Clinical Evaluative Sciences*

We have developed a population risk model to predict future risk of diabetes in Canada based on age, sex and body mass index by individually linking Ontario respondents to the Statistics Canada National Population Health Survey (NPHS) to a chart-validated diabetes registry using Ontario health administrative data. We validated using Manitoba NPHS and Ontario Canadian Community Health Survey data. We estimated trends by applying the model to national health surveys, using bootstrap replicate weights to obtain confidence intervals. We estimated 1.8M new diabetic cases will develop by 2015. This model can assist in planning future diabetes prevention and management strategies.

Nous avons élaboré un modèle de risque à l'échelle de la population pour prédire le risque futur de diabète au Canada d'après l'âge, le sexe et l'indice de masse corporelle en couplant les données sur les Ontariens qui ont participé à l'Enquête nationale sur la santé de la population (ENSP) réalisée par Statistique Canada à celles d'un registre du diabète validé d'après les dossiers médicaux en utilisant des données administratives sur la santé en Ontario. Nous avons validé le modèle en nous servant des données sur les participants manitobains à l'ENSP et de celles sur les participants ontariens à l'Enquête sur la santé dans les collectivités canadiennes. Nous avons estimé les tendances par l'application du modèle à des enquêtes nationales sur la santé en utilisant les poids de rééchantillonnage bootstrap pour obtenir les intervalles de confiance. Nous avons estimé que 1,8 million de nouveaux cas de diabète se manifesteront d'ici à 2015. Ce modèle peut faciliter la planification des futures stratégies de prévention et de gestion du diabète.

Session 04C Lun. 26 mai • Mon. May 26, 14:35 – 16:05 Capital Hall 3A

Statistical Analysis of Hydroclimatic Proxy Data: Bayesian Approaches
Analyse statistique des empreintes hydro-climatiques : approches bayésiennes

[MS-50]

A Bayesian Model to Extract a Common Factor from a Set of Tree-rings Width
Un modèle bayésien pour extraire un facteur commun d'une population d'arbres

Jean-Jacques Boreux, *Université de Liège*, Philippe Naveau, *Laboratoire des Sciences du Climat et de l'Environnement (LSCE)*, Luc Perreault, *IREQ, Hydro-Québec*, Antoine Nicault, *INRS, Québec*, Eric Parent, *Paris AGRO TECH*, Jacques Bernier

A tree is a bio-indicator of its climate and the model's task is to extract a climatic signal contained in the rings of trees living on a given site. Here, we present a Bayesian model, i.e. hypotheses, parameters, and error terms. The common factor appears as a latent variable in a hierarchical structure. Its median has a strong linear correlation to the signal obtained through dendrochronologists' best practices. In addition to the quantification of uncertainty, our model delivers a series of information such as the contribution of each tree to the common signal. A Gibbs sampling via a Matlab code is available.

L'arbre est un bio-indicateur du climat et la tâche du modélisateur est d'extraire un signal climatique enfoui dans les cerles des arbres vivant sur un même site. Nous présentons ici un modèle bayésien, c'est-à-dire des hypothèses, des paramètres et des termes d'erreur. Le facteur commun apparaît sous la forme d'une variable latente dans une structure hiérarchique. Sa médiane est fortement linéairement corrélée au signal obtenu par les bonnes pratiques des dendrochronologues. Outre la quantification des incertitudes, notre modèle délivre une série d'informations comme la contribution de chaque arbre au signal commun. L'échantillonnage de Gibbs sous Matlab est disponible.

[MS-51]

Temporal Model for Climate Reconstruction using Pollen Data by Dynamic Vegetation Model Inversion
Modèle temporel pour la reconstruction du climat à partir de pollens et en inversant un modèle dynamique de végétation

Vincent Garreta, *CEREGE, CNRS/Université Aix-Marseille*, Paul Miller, *Geobiosphere Science Center, department of Physical Geography & Ecosystems Analysis, Lund University*, Joél Guiot & Christelle Hély, *CEREGE, CNRS/Université Aix-Marseille*

Identification and counting of pollens from a lake sediment core provide high quality information for past vegetation and climate reconstruction. The statistical model that we present describes relations between climate, vegetation and pollens including LPJ-GUESS vegetation model. It's a temporal hierarchical model which contains two hidden levels: climate and vegetation to be reconstructed. The originality of this model is its included time link. Stochastic aspect of the vegetation model associated to the temporal structure of the statistical model requires the use of a particle filter for inference. We present a climate reconstruction at a single site over the last 10,000 years.

La reconnaissance et le comptage des pollens contenus dans les sédiments lacustres fournissent une information de haute qualité pour reconstruire végétation et climat passé. Le modèle statistique que nous présentons décrit les relations entre climat, végétation et pollen en incluant le modèle dynamique de végétation LPJ-GUESS. Il est temporel hiérarchique et contient deux couches cachées qui sont climat et végétation à reconstruire. L'originalité de ce modèle est le lien temporel qu'il contient. La nature aléatoire de LPJ-GUESS et l'aspect temporel du modèle statistique nous obligent à utiliser un filtre particulaire pour l'inférence. Nous présentons une reconstruction climatique en un site sur 10 000 ans.

[MS-52]
Hierarchical Model Applied to Dendrochronology
Modèles hiérarchiques appliqués à la dendrochronologie

Delphine Grancher, *CNRS Bellevue*, Ophélie Guin & Philippe Naveau, *LSCE*, Jean-Jacques Boreux, *Université de Liège*

Continuing the work previously descbribed by J.-J. Boreux et al., we have adapted a Bayesian model for dendrochronological data on cembro pine in the French Alps. These data have been distributed by a number of distinct sites, and we have studied the site effect through the increasing complexity of the original model. We have compared many hierarchical models in order to better explore and understand the possible contributions of the different climate and water signs that are represented at the level of the hidden explanatory variables.

Dans la continuité des travaux précédemment décrits par J.-J. Boreux *et al.*, nous avons adapté un modèle hiérarchique bayésien à des données dendrochronologiques de pin cembro dans les Alpes françaises. Ces données étant réparties sur plusieurs sites distincts nous avons étudié l'effet site dans la complexification du modèle original. Nous avons comparé plusieurs modèles hiérarchiques afin de mieux explorer et comprendre les possibles contributions des différents signaux climatiques et hydriques qui sont représentés au niveau des variables explicatives cachées.

Session 04D Lun. 26 mai • Mon. May 26, 14:35 – 16:05 Capital Hall 4A

Monte Carlo Markov Chains
Chaînes de Markov Monte-Carlo

[MS-53]
Some New Results on Controlled MCMC Algorithms
Quelques résultats récents concernant les MCMC contrôles

Christophe Andrieu, *University of Bristol*

Controlled MCMC algorithms are emerging simulation techniques that aim at helping the design of efficient algorithms by optimising their properties with limited human intervention. In this talk we will present some recent results that complete the existing theory (and in particular explain what is observed in practice) as well as some new algorithms and applications.

Les MCMC contrôlées constituent un récent champ de recherche dans le domaine des techniques de simulation, dont le but est d'aider le développement d'algorithmes efficaces par le truchement d'une optimisation de leurs propriétés qui limite l'intervention de l'utilisateur. Dans cet exposé nous présenterons quelques résultats nouveaux concernant la théorie de ces algorithmes ainsi que de nouveaux algorithmes et applications.

[MS-54]
On the Optimal Scaling Problem for Metropolis – Hastings Algorithms
Robustesse de l'échelonnage optimal pour les algorithmes de type Metropolis-Hastings

Mylène Bédard, *Université de Montréal*

Metopolis – Hastings algorithms, an important class of MCMC algorithms, allow us to generate samples from highly complex probability distributions. In order to implement these algorithms, it is necessary to select a proposal distribution and to tune the scale parameter of this distribution such

Les algorithmes de type Metropolis-Hastings, une classe importante d'algorithmes MCMC, permettent de générer des échantillons provenant de distributions complexes. En appliquant ces algorithmes, il est nécessaire de choisir une distribution instrumentale et de sélectionner le paramètre d'échelle de cette distribution, afin d'optimiser la perfor-

as optimizing the efficiency of the method. Results about the optimal scaling of the proposal distribution and the optimal acceptance rate of the algorithm for the special case of multidimensional target distributions with independent components are available in the literature. I consider to which extent these results could be generalized to more complex target distributions.

mance de l'algorithme. Des résultats sur l'échelonnage optimal de la distribution instrumentale et sur l'optimalisation du taux d'acceptation des algorithmes sont disponibles dans la littérature pour les distributions cibles multidimensionnelles aux composantes i.i.d. Je considère dans quelle mesure ces résultats peuvent être généralisés à des distributions cibles plus complexes.

[MS-55]
What Can we Learn from the Renormalization Methods in MCMC?
Que peut-on apprendre des méthodes de renormalisation pour les algorithmes de MCMC ?

Eric Moulines & Gersende Fort, *Institut des Télécommunications/Telecom ParisTech*

Several renomalization methods have been introduced to understand the dynamics of complex MCMC algorithms. There are several options that can be considered, depending upon the particular characteristic of the algorithm that we wish to assess. Fluid methods allow us to understand the behavior of the Markov chain "out-of-the equilibrium," i.e., when the initial state is far from the main body of the distribution. Diffusion limits rather consider the behavior of large dimensional MCMC in the "stationary regime." We show how these different scalings may be combined to investigate several aspects of the MCMC methods.

Différentes méthodes de renormalisation ont été introduites pour comprendre le comportement de méthodes de simulations MCMC complexes. Différentes méthodes peuvent être considérées suivant les caractéristiques particulières de l'algorithme que nous voulons mettre en évidence. Les limites fluides permettent de comprendre le comportement des méthodes de la chaîne de Markov « hors d'équilibre ». Les « limites diffusives » permettent d'étudier le comportement à l'équilibre des MCMC dans des espaces de grandes dimensions. Nous montrons dans ce travail comment ces différentes limites peuvent être combinées pour étudier d'autres aspects du comportement des algorithmes.

Session 04E Lun. 26 mai • Mon. May 26, 14:35 – 16:05 Capital Hall 5A

Hypothesis Testing in Biostatistics
Tests d'hypothèse en biostatistique

[MS-56]
Statistical Tests in Selecting Cure Models
Tests statistiques pour la sélection de modèles de guérison

Xu Jianfeng & Yingwei Peng, *Queen's University*

We consider two popular cure models for survival data with a cure fraction in this work: the mixture cure model and the promotion time cure model. We investigate some statistical tests that are useful in selecting an appropriate model from the two types of models. Simulation studies are provided to examine the null distributions and the powers of the tests. The tests are applied to leukaemia data sets.

Dans ce travail, nous considérons deux modèles de guérison populaires pour données de survie avec une fraction guérie : le modèle de mélange de guérison et le modèle de guérison à temps de promotion. Nous étudions certains tests statistiques utiles dans la sélection d'un modèle approprié parmi ces deux types de modèles. Des études de simulation permettant d'examiner les distributions nulles et la puissance des tests sont présentées. Les tests sont appliqués à un jeu de données de leucémie.

[MS-57]

Statistics for Testing Genetic Linkage under Two-gene Epistatic Models
Statistiques pour tester la liaison génétique sous des modèles épistatiques à deux gènes

Alexandre Bureau, Jordie Croteau, Chantal Mérette, Yvon Chagnon, Marc-André Roy & Michel Maziade, *Centre de recherche Université Laval - Robert-Giffard*

Given that interactions between genes (epistasis) are involved in the aetiology of complex diseases, two-gene epistatic models should facilitate the detection of genetic linkage between markers and genes in extended families. We review likelihood-based statistics (LOD scores) that can be constructed to test for linkage under two-gene models, with a focus on the conditional LOD score to test for linkage to a gene given the linkage information at another gene. Simulations and an application to schizophrenia and bipolar disorder in Eastern Quebec families reveal that the gain in signal can be important.

Sachant que des interactions entre gènes (épistasies) sont impliquées dans l'étiologie des maladies complexes, des modèles épistatiques à deux gènes devraient faciliter la détection de la liaison génétique entre des marqueurs et des gènes dans des familles étendues. Nous présentons les statistiques basées sur la vraisemblance (score LOD) que l'on peut construire pour tester la liaison sous des modèles à deux gènes, en particulier le score LOD conditionnel pour tester la liaison à un gène sachant l'information de liaison à un autre gène. Des simulations et une application à la schizophrénie et à la maladie bipolaire dans des familles de l'est du Québec révèlent que le gain de signal peut être considérable.

[MS-58]

Investigating the Impact of Different Effect Measures on Power for Trials Using Binary Outcomes
Examen de l'effet des mesures d'effet différents sur la puissance des essais utilisant des résultats binaires

Sameer Parpia & Lehana Thabane, *McMaster University*

The most common ways of expressing the effect of an intervention in trials using binary outcomes include relative risk (RR), odds ratio (OR), risk difference (RD), or the number needed to treat. Many trials often assume RD as the effect measure at the design stage, but adopt RR or OR at the analysis stage — which may affect the ability to detect the target effect size. In this presentation, we present the results of a simulation study to investigate the impact of using an alternate binary outcome measure in analysis from that used in the design on statistical power.

Les moyens les plus fréquents d'exprimer l'effet d'une intervention dans les essais dont les résultats sont binaires comprennent le risque relatif (RR), le rapport de cotes (RC), la différence des risques (DR) ou le nombre de patients à traiter. Dans de nombreux essais, il arrive que l'on suppose à l'étape de la conception que la DR sera la mesure de l'effet, mais que l'on adopte au stade de l'analyse le RR ou le RC, ce qui peut avoir une incidence sur la capacité de déceler la grandeur de l'effet étudié. Dans la communication, nous présentons les résultats d'une étude par simulation réalisée pour étudier l'effet qu'a sur la puissance l'utilisation dans l'analyse d'une autre mesure du résultat binaire que celle employée à l'étape de la conception.

[MS-59]

On Using Prior Information to Improve Power of Genome-wide Studies: Weighted P-values Versus Stratified False Discovery Control
De l'utilisation de l'information *a priori* pour améliorer la puissance des études sur tout le génome : valeurs-p pondérées contre contrôle stratifié des fausses découvertes

Yun Joo Yoo, *Samuel Lunenfeld Research Institute, Mount Sinai Hospital*, Lei Sun, *University of Toronto*, Shelley Bull, *Samuel Lunenfeld Research Institute, Mount Sinai Hospital*

The power of genome-wide studies can be improved by utilizing prior information, and two

La puissance des études sur l'ensemble du génome peut être améliorée en utilisant de l'information *a priori*, et deux

non-Bayesian methods have been proposed in the framework of false discovery rate control: the weighted p-value method and the stratification method. We show theoretically that the stratification method can be formulated as a weighted p-value method. The simulation results confirm our theoretical derivation and show that when the prior is not informative, unlike the weighted p-value method, the stratification method does not lose power. When the prior is indeed informative, both methods improve power with a marginally better performance by the weighted p-value method.

méthodes non bayésiennes ont été proposées dans le cadre du contrôle du taux de fausses découvertes : la méthode de la valeur-p pondérée et la méthode de la stratification. Nous montrons théoriquement que la méthode de la stratification peut être formulée comme une méthode de valeur-p pondérée. Les résultats de simulation confirment nos résultats théoriques et montrent que lorsque l'information *a priori* n'est pas informative, contrairement à la méthode de la valeur-p pondérée, la méthode de la stratification ne perd pas de puissance. Lorsque l'information *a priori* est informative, les deux méthodes améliorent la puissance, et la performance de la méthode de la valeur-p pondérée est marginalement meilleure.

[MS-60]
Missing Data: TDT with Two Affected Offspring
Données manquantes : test du déséquilibre de transmission avec deux enfants affectés

Gulhan Alpargu, *California State University, Fullerton*

The transmission disequilibrium test (TDT) detects linkage between a marker and a disease-susceptibility locus in the presence of linkage disequilibrium between two loci. The TDT requires complete genotypes. Several methods have been proposed to deal with missing genotypes in parents but not in offspring. We propose a simple method (MI-TDT) to impute missing genotypes in any members of the family with two affected offspring. We illustrate the MI-TDT to identify the significant genes susceptible to type 1 diabetes for the incomplete SNP data from U.K. Warren families.

Le test du déséquilibre de transmission (TDT) détecte les liens entre un marqueur et un locus de vulnérabilité à la maladie en présence de déséquilibre de liaisons entre deux locus. Le TDT requiert des génotypes complets. Plusieurs méthodes ont été proposées pour traiter des génotypes manquant chez les parents et complets chez les enfants. Nous proposons une méthode simple (MI-TDT) pour imputer les génotypes manquants chez n'importe quel membre de la famille avec deux enfants affectés. Nous illustrons l'utilisation du MI-TDT pour identifier les gènes significatifs dans la vulnérabilité au diabète de type I avec données de SNP incomplets provenant de familles Warren au Royaume-Uni.

[MS-61]
Robust Statistical Approaches for Sib-Pair Linkage Analysis of Quantitative Trait Loci (QTLs)
Approche robuste de l'analyse de liaisons génétiques pour des traits quantitatifs

Mohamedou Sow, *GEMA Université Bordeaux 1 et CNRS,* Laurent Briollais, *Samuel Lunenfeld Research Institute,* Gilles Durrieu, *University Bordeaux 1 et CNRS*

Study of quantitative traits loci (QTLs) associated with complex diseases may help to dissect their underling genetic mechanisms. Haseman – Elston (HE) least-square estimator is appropriate for detecting linkage of chromosomal regions to quantitative traits in candidate gene or genome-wide search but may lack some robustness in the presence of outliers or non-normal trait distributions. We propose several alternatives to the HE approach

L'étude de traits quantitatifs intermédiaires associés à des maladies complexes peut aider à mettre en évidence des gènes impliqués dans ces maladies. La méthode de Haseman et Elston (HE), basée sur l'estimateur des moindres carrés est sensible aux valeurs aberrantes et aux hypothèses de distributions des traits. Nous proposons des alternatives à l'approche de HE basée sur des estimateurs robustes et une transformation des traits. Les simulations montrent qu'en présence de perturbations de l'hypothèse de base de la mé-

based on robust estimators and transformation of QTLs. Simulation studies with perturbations of the HE assumptions using bivariate skew-t distribution, show that robust approaches have better performance than the classical HE approach.

thode HE par la distribution t asymétrique, l'approche robuste se comporte mieux que l'approche classique de HE.

Session 04F Lun. 26 mai • Mon. May 26, 14:35 – 16:05 Capital Hall 6A

Public Health Studies
Études en santé publique

[MS-62]

The Safety of Genetically Modified Foods — Defining Statistical Evidence
La sûreté des aliments génétiquement modifiés – définir la preuve statistique

William Ross & Cunye Qiao, *Health Canada*

In general, regulators require that foods derived from genetically modified crops "are as safe as those derived from conventionaly bred crops. . . " — OECD. To date, no statistical theory has been developed for quantifying evidence for the "equivalence" of GM and non-GM foods. We provide an analysis of this problem in the light of recent controversies regarding the introduction of GM corn to the European market. Both traditional and Bayesian approaches to testing are examined along with the testing equivalence by the wrong statistical hypothesis. Alternate characterisations of statistical evidence are also considered.

En général, les autorités demandent que les produits alimentaires dérivés de cultures génétiquement modifiés "soient aussi sécuritaires que ceux dérivés de cultures conventionnelles. . ." – OCDE, traduction libre. À ce jour, aucune théorie statistique quantifiant la preuve pour l'« équivalence » de produits alimentaires GM et non-GM n'a été développée. Nous apportons une analyse de ce problème à la lumière de récentes controverses relatives à l'introduction de maïs GM sur le marché européen. Des procédures de test bayésiennes et traditionnelles sont examinées. Des caractérisations alternatives de la preuve statistique sont aussi prises en considération.

[MS-63]

The Impact of Components of Deprivation on Child Pedestrian Casualties in England
L'effet des composantes de la défavorisation sur les accidents impliquant des enfants piétons en Angleterre

David Stephens, *McGill University*, Daniel Graham, *Imperial College London*

The relationship between deprivation in an area and the rate of child pedestrian casualties (CPC), including children killed or seriously injured (CKSI), has been long suggested. In this talk, I will outline an analysis that relates rates of CKSI and CPC in a census district to components of deprivation (Housing Score, Crime Score etc.) from a recent study in England. Important in the analysis is the construction of an appropriate mean model, a model for spatial dependence, and correction for confounding. I will demonstrate the use of propensity score methods that produce unbiased estimation of the dependence on deprivation.

L'existence d'une relation entre la défavorisation dans une région et le taux d'accidents impliquant des enfants piétons, y compris les enfants tués ou gravement blessés, est évoquée de longue date. La communication décrira une analyse qui relie les taux d'enfants tués ou gravement blessés et d'accidents impliquant des enfants piétons dans un district de recensement aux composantes de la défavorisation (score de logement, score de criminalité, etc.) tirées d'une étude réalisée récemment en Angleterre. Un aspect important de l'analyse est la construction d'un modèle de moyenne approprié, d'un modèle de dépendance spatiale et de corrections pour tenir compte des variables confusionnelles. L'utilisation de méthodes fondées sur les scores de propension qui produisent une estimation sans biais de la relation de dépendance avec la privation sera illustrée.

[MS-64]
Measurement of Integrated Human Service Networks
Mesure des réseaux de service humains intégrés

Chenglin Ye, Gina Browne & Lehana Thabane, *McMaster University*

An integration network has been advanced as a strategy for delivering multiple human services that have traditionally been provided by autonomous agencies independently. In this presentation we use data from a randomized controlled trial to evaluate the effect of the Children's Disability Treatment Network in improving human services and reducing wait-lists. In analysis we use baseline data to measure congruence among agencies, its possible association with quality of the network and how these relate to the volume capacity of services at the agency level. We also explore different graphical approaches of displaying measures of congruence among agencies.

Un réseau d'intégration a été proposé comme stratégie de fourniture de services humains multiples qui étaient traditionnellement procurés indépendamment par des organismes autonomes. La communication décrit l'utilisation de données provenant d'un essai randomisé contrôlé pour évaluer l'effet du réseau de traitement de l'incapacité chez les enfants en ce qui concerne l'amélioration des services humains et la réduction des listes d'attente. Dans l'analyse, nous utilisons des données de référence pour mesurer la congruence entre les organismes, son association éventuelle avec la qualité du réseau et la façon dont ces facteurs sont reliés à la capacité de fourniture de services au niveau de l'organisme. Nous examinons aussi diverses approches graphiques de présentation des mesures de la congruence entre les organismes.

[MS-65]
Use of the Database from the Salaried Workers' French Health Insurance System for the Analysis of the Efficiency of Care Pathways in Chronic Diseases: Preliminary Results
Utilisation des données de la CNAMTS dans l'analyse de l'efficience des parcours de soins dans les maladies chroniques : résultats préliminaires

Annie Rudnichi, *Haute Autorité de Santé*

Exploratory research in a representative sample of the French population in the data base aimed to extract a subpopulation of non-insulino treated diabetic patients without complication. A few patients (less than 1%) follow the cares pathway recommended by the French National Authority for Health in the guideline elaborated for long-lasting disorders, whereas the annual amount of ambulatory health expenses observed in this group (first decile= 417€ – last decile= 2,383€) are close to the theoretical one calculated (Minimum : 440€ – maximum : 1,875€). The bulk of reimbursed services are partially similar to the recommended.

Une recherche exploratoire sur un échantillon représentatif de la population française affiliée au régime général a cherché, à partir d'une population de patients diabétiques de type 2 en affection longue durée, à sélectionner une cohorte de patients non compliqués, non insulino-traités. Dans celle-ci, le parcours de soins ambulatoires préconisé dans le référentiel élaboré par la Haute Autorité de Santé s'est révélé suivi par peu de patients (moins de 1%) alors que le coût annuel observé en 2006 est superposable (1er décile : 417€ – 9e décile : 2 383€) au théorique calculé (Minimum : 440€ – maximum : 1 875€).

[MS-66]
Establishing a Soil Quality Index
Mise en place d'un indice la qualité des sols

Salima Taibi, *Esitpa - Rouen,* Aurore Lambert, *INSA - Rouen,* Patrice Lepelletier & Karine Laval, *Esitpa - Rouen,* Christian Mougin, *INRA - Versailles*

The purpose of the study is to develop an index on the biological state of soil, in order to better eval-

L'objectif de l'étude est d'élaborer un indice de l'état biologique des sols permettant de mieux évaluer la qualité d'un

uate the quality of a given soil. Microbial, fungal and bacterial biomasses, as well as physical chemical properties have been estimated using descriptive variables. Data collection will take place over three years, including a final analysis of the amount of copper in the soil. The bio-indicator is thought to be a combination of information originating from each of the compartments listed above, that accounts for the heterogeneity of the data observed. The techniques used are discriminate analysis and longitudinal regression, as well as modeling based on random forests.

sol. Les biomasses microbiennes, fongiques et bactériennes ainsi que les caractéristiques physico-chimiques ont été estimées par des variables descriptives. La collecte des données s'est déroulée sur trois années, dont une dernière avec apport de cuivre. Le bio-indicateur est pressenti comme étant une combinaison d'informations issues de chacun des compartiments cités ci-dessus et en tenant compte de l'hétérogénéité des données observées. Les techniques utilisées sont des analyses discriminantes, des régressions logistiques et une modélisation basée sur les forêts aléatoires.

[MS-67]

From the Indicator to the Honours List : Relevant Use of Graph Chart Representations
De l'indicateur au palmarès : utilisation pertinente des représentations graphiques

Sophie Tricaud-Vialle & Elodie Pambrun, *Comité de Coordination de l'Evaluation Clinique et de la Qualité en Aquitaine (CCECQA)*

What kind of graphical representation should be used for comparative information based on health care quality indicators? Two graphical methods are presented: the league table, which presents a hierarchy, and a funnel plot, which divides the establishments into groups (good, average, poor). The different interpretations identified are illustrated through the "Score Logistique ©Saphora-MCO" values of 21 health establishments. The funnel plots, which quickly identify outlier elements, are recommended. Since the indicator's ultimate goal is to lead the establishments in implementing the required improvements, is it necessary to establish an Honours list?

Quelle représentation graphique privilégier pour les informations comparatives réalisées à partir d'indicateurs de qualité des soins ? Deux méthodes graphiques sont présentées : le league table qui présente un classement, et le funnel plot qui répartit les établissements en groupes (bon, moyen, mauvais) ; les différences d'interprétation identifiées sont illustrées à partir des valeurs du « Score Logistique ©Saphora-MCO » de 21 établissements de santé. Les funnel plots, permettant l'identification rapide des éléments ayant une performance divergente, sont recommandés : l'indicateur ayant pour finalité d'amener l'établissement à mettre en place les actions d'amélioration requises, est-il nécessaire d'établir un palmarès ?

Session 04G Lun. 26 mai • Mon. May 26, 14:35 – 16:05 Capital Hall 7A

Applications of Functional Estimation
Applications de l'estimation fonctionnelle

[MS-68]

Estimation of Differential Equations via Nonparametric Methods: Applications in Systems Biology
Estimation d'équations différentielles par des méthodes non paramétriques : applications en biologie systémique

Nicolas Brunel & Florence d'Alché-Buc, *laboratoire IBISC, Université d'Evry*

Differential equations are widely used in experimental sciences. Nevertheless, the estimation of the parameters indexing the equations still gen-

Les équations différentielles sont des modèles courants en sciences expérimentales. Pour autant, l'estimation des paramètres indexant ces équations pose toujours des difficul-

erates important computational difficulties (huge computational times, local maxima) when using classical methods (least squares, likelihood). We study the use of nonparametric estimators of the solution and of its derivative for the construction of M-estimators that are easier to compute. We show that the parametric estimator is consistent and asymptotically normal under simple conditions, but the rate of convergence remains nonparametric. We propose then some procedures for ameliorating the asymptotic behavior and for integrating prior knowledge.

tés calculatoires importantes (temps de calcul élevés, optima locaux) avec les méthodes classiques (moindres carrés, vraisemblance). Nous étudions l'utilisation d'estimateurs non paramétriques de la solution et de sa dérivée pour la construction de M-estimateurs plus simples à déterminer. Nous montrons alors que l'estimateur paramétrique est consistant et asymptotiquement normal sous des conditions simples, mais que la vitesse de convergence reste non paramétrique. Nous proposons aussi des procédures pour améliorer le comportement asymptotique et intégrer des connaissances *a priori*.

[MS-69]
Kriging Modelling of the Degradation along the Catenary Contact Wire
Modélisation par krigeage de la dégradation le long du fil de contact caténaire

Jeanne Casaert & Denis Bosq, *Laboratoire de Statistique Théorique et Appliquée (Université Paris 6)*, Rachid Ziani, *SNCF*

As part of the preventive maintenance of the contact wire, we want to calculate the likelihood of not detecting a critical point (point that is too thin) We know the thickness of all the points spaced at 20 cm and we would like to estimate the thickness of points spaced at 2 cm. The kriging interpolation is used in order to model the degradation along the contact wire thanks to the known thicknesses. As a complement, the non-parametric estimation of the density of the thicknesses is computed conditionally to the thickness of the two flanked known points.

Dans le cadre de la maintenance préventive des fils de contact caténaire, nous cherchons à calculer le risque de ne pas détecter un point critique du fil (point avec une épaisseur trop faible). Nous connaissons l'épaisseur des points tous les 20 cm et voulons estimer l'épaisseur des points tous les 2 cm. Nous utilisons l'interpolation par krigeage pour modéliser la dégradation le long du fil de contact à l'aide des épaisseurs connues. Afin de compléter les résultats obtenus, nous estimons de façon non paramétrique la densité conditionnelle des épaisseurs de points sachant l'épaisseur du couple de points connus encadrant.

[MS-70]
Option Characteristic Curves Estimation by Penalized Marginal Maximum Likelihood
Estimations des courbes caractéristiques d'options par maximum de vraisemblance marginale pénalisée

Stéphane Germain, Belkacem Abdous, Pierre Valois & Louis-Paul Rivest, *Université Laval*

In item response theory, the responses of subjects to a questionnaire could be seen as random vectors whose probability law depends on a latent random variable. With the supposition of independent items conditional upon the latent variable, we use the penalized marginal likelihood to construct a new estimator of the characteristic curves. We will show some properties of this estimator and its implementation based on wavelets decomposition and the EM algorithm.

En théorie de réponse aux items, les réponses d'individus à un questionnaire peuvent être perçues comme des vecteurs aléatoires dont la loi de probabilité dépend d'une variable aléatoire latente. En supposant que les items du questionnaire sont indépendants conditionnellement à la variable latente et en utilisant la pénalisation de la vraisemblance marginale, nous construisons un nouvel estimateur des courbes caractéristiques. Nous présenterons quelques propriétés de cet estimateur ainsi que sa mise en œuvre en pratique basée sur une décomposition en ondelettes combinée avec l'algorithme EM.

[MS-71]
Kernel Estimation of Density Level Sets
Estimation des ensembles de niveaux d'une densité par la méthode du noyau

Benoit Cadre, *ENS Cachan Bretagne*

One of the most important application of density level sets estimation is in unsupervized cluster analysis : here, one tries to break a complex data set into a series of piecewise similar groups or structures, each of which may then be regarded as a separate class of data, thus reducing overall data complexity. In our approach, we consider that these groups are the connected components of a density level set. In this talk, we present some results on the exact rate of convergence of a plug-in estimator of the density level set. We also consider the important case where the level corresponds to a fixed probability for the common law of the data.

L'estimation de l'ensemble de niveau associé à une densité intervient comme réponse possible à de nombreux problèmes de statistique, souvent liés au besoin de classer les observations. Par exemple, la problématique de l'analyse des groupes consiste à décomposer l'échantillon des données en plusieurs groupes. Une façon de construire ces groupes est de considérer qu'ils sont les composantes connexes d'un ensemble de niveau associé à la densité des observations. Dans cet exposé, nous présentons des résultats de vitesse de convergence exacte pour l'estimateur d'insertion (plug-in) de l'ensemble de niveau de la densité. Nous considérons également le cas important où le niveau correspond à une probabilité fixe pour les données consensuelles.

Session 04H Lun. 26 mai • Mon. May 26, 14:35 – 16:05 Congress Hall A/E West

Poster Session I
Séance d'affichage I

[MS-72]
Psychometrics for Educational Assessment: The Test de connaissance du français
Approche psychométrique de l'évaluation dans le domaine de l'éducation : le Test de connaissance du français

Christophe Lalanne, Marianne Mavel, Marina Esposito-Farèse & Pascal Bessonneau, *Centre international d'études pédagogiques*

We propose a general overview of the psychometrical procedures used to calibrate and link the Test de connaissance du français (TCF®) to the Common European Framework of Reference for language testing (European Union Council). The TCF comprises 80 multiple choice items designed to assess listening and reading proficiency in French as a foreign language. Discriminative power and difficulty estimates of items are obtained using methods derived from Classical Test Theory and Item Response Theory, while test structure is analysed through factorial analysis. Preliminary results on the use of a mixed model approach for examining Differential item functioning are also discussed.

Nous exposons le schéma général des analyses psychométriques réalisées pour calibrer le Test de connaissance du français (TCF®) avec le Cadre européen de référence pour les langues (Conseil de l'Europe). Le TCF est composé de 80 items à choix multiples de réponse et permet d'évaluer le niveau à l'écrit et à l'oral en français langue étrangère. Le pouvoir discriminant et l'estimation de la difficulté des items sont estimés dans le cadre de la Théorie classique des tests et de la Théorie de la réponse à l'item, tandis que la structure du test est analysée à l'aide d'analyses factorielles. Des résultats préliminaires sur l'utilisation des modèles mixtes pour l'étude du fonctionnement différentiel des items sont également discutés.

[MS-73]

Instrumental Variable Approach to Covariate Measurement Error in Generalized Linear Models
Approche des variables instrumentales pour l'erreur de mesure des covariables dans les modèles linéaires généralisées

Chel Hee Lee & Taraneh Abarin, *University of Manitoba*

The poster presents method of moment estimation for generalized linear measurement error models using the instrumental variable approach. The measurement error has a parametric distribution that is not necessarily normal, while the distributions of the unobserved covariates are nonparametric. We also propose two simulation-based estimators to overcome the possible computational difficulties. The proposed estimators are strongly consistent and asymptotically normally distributed under some regularity conditions. Finite sample performances of the estimators are investigated through simulation studies.

L'affiche présente l'estimation par la méthode du moment pour les modèles d'erreur de mesure linéaire généralisée utilisant l'approche par variable instrumentale. L'erreur de mesure a une distribution paramétrique qui n'est pas nécessairement normale alors que les distributions des covariables inobservées sont non paramétriques. Nous proposons également deux estimateurs fondés sur la simulation pour surmonter les possibles difficultés computationnelles. Les estimateurs proposés sont fortement cohérents et distribués normalement et asymptotiquement selon certaines conditions de régularité. Les performances des estimateurs sont examinées au moyen d'études de simulation.

[MS-74]

Approximate Conditional Permutation Tests for Logistic Models
Tests de permutations conditionnels approximatifs pour modèles logistiques

Paul Nguyen & Duncan Murdoch, *University of Western Ontario*

Conditional permutation tests permute the observations while holding the observed sufficient statistics under the null hypothesis fixed. In the presence of continuous nuisance parameters, it is common that no permutations other than the observed one keep the sufficient statistics fixed, resulting in a degenerate conditional distribution for the test statistics. In this paper, we present a test that uses a weighted permutation distribution, where the weights decline with increasing distance from the observed data, to condition approximately on the values of the sufficient statistics for the nuisance parameters.

Les tests de permutations conditionnels permutent les observations tout en fixant la valeur des statistiques exhaustives observées sous l'hypothèse nulle. En présence de paramètres de nuisance continus, il est fréquent qu'aucune permutation autre que celle observée ne conserve la valeur des statistiques exhaustives fixe, résultant en une distribution conditionnelle dégénérée pour les statistiques de test. Dans cet article, nous présentons un test utilisant une distribution de permutations pondérées, où les poids décroissent alors que la distance par rapport aux données observées augmente, pour conditionner approximativement sur les valeurs des statistiques exhaustives pour les paramètres de nuisance.

[MS-75]

Statistical Inference for Minimum Inhibitory Concentration (MIC) Data
Inférence statistique pour données de concentration d'inhibition minimale

Huanhuan Wu, *Simon Fraser University*, Nancy Dewith, *The B.C. Ministry of Agriculture &Lands*, Carl Schwarz & Thomas Loughin, *Simon Fraser University*

The log-normal distribution is often used for modeling MIC (Minimum Inhibitory Concentration) data. We consider a model for which the unknown mean value of logMIC is related to some explanatory variables. The MIC value is interval censored

La distribution log-normale est souvent utilisée pour modéliser des données de concentration d'inhibition minimale (CIM). Nous considérons un modèle pour lequel la valeur moyenne inconnue de logCIM est reliée à certaines variables explicatives. La valeur CIM est censurée par intervalles et les

so ordinary least squares cannot be used. For models containing random effects, MLE (maximum likelihood estimates) methods are infeasible and Bayesian approaches are required. Estimates from the two analysis methods are discussed and compared. Four real data sets were analyzed using the methodology we developed.

moindres carrés ordinaires ne peuvent donc pas être utilisés. Pour des modèles contenant des effets aléatoires, les estimateurs du maximum de vraisemblance ne peuvent être utilisés et une approche bayésienne est nécessaire. Des estimations obtenues au moyen de deux méthodes d'analyse sont comparés. Quatre jeux de données réelles ont été analysés avec la méthodologie que nous avons développée.

[MS-76]

A Non-asymptotic Penalized Criterion for Gaussian Mixture Model Selection
Critère pénalisé non asymptotique pour la sélection d'un modèle de mélange gaussien

Cathy Maugis & Bertrand Michel, *Université Paris-Sud XI*

Specific Gaussian mixtures are considered to simultaneously solve variable selection and clustering problems. A non-asymptotic penalized criterion is proposed to select the number of mixture components and the relevant variable subset. Because of the non-linearity of the Kullback-Leibler contrast on Gaussian mixtures, a general model selection theorem for MLE is used to obtain the penalty function form, requiring the control of bracketing entropy of Gaussian mixture families. In practice, a slope heuristics is applied to calibrate the unknown multiplicative constant for defining efficient penalty functions from the data.

Des mélanges gaussiens sont considérés pour obtenir simultanément une classification des données et établir le rôle des variables. Un critère pénalisé non asymptotique est établi pour sélectionner le nombre de composantes du mélange et les variables significatives pour la classification. Le contraste Kullback-Leibler ayant un comportement non linéaire sur ces mélanges, un théorème général de sélection de modèles est utilisé pour obtenir la forme des pénalités. Ce théorème exige le contrôle de l'entropie à crochets des familles de mélanges gaussiens étudiées. En pratique, une heuristique de pente est utilisée pour déterminer la constante multiplicative inconnue dans la fonction de pénalité.

[MS-77]

Coupon Collecting in the Presence of Rare Coupons
Collection de coupons et la présence de coupons rares

Qiujing Zhou & Brad Johnson, *University of Manitoba*

Suppose we purchase packs of (distinct) baseball cards where the majority of cards in the set occur with the same relative frequency and a few of the cards are rare. Of interest is the expected number of packs required in order to collect a full set of cards. This is a version of the coupon subset collection problem. In this paper, we will review some existing results for this problem and present a method of approximating this expectation in some special cases.

Supposons que nous achetons des paquets de cartes de baseball (distinctes) et que la majorité des cartes formant le jeu complet surviennent avec la même fréquence relative tandis que quelques-unes sont rares. Nous voulons déterminer le nombre attendu de paquets de cartes requis pour obtenir un jeu complet de cartes. Il s'agit d'une version du problème de collecte d'un sous-ensemble de coupons. Dans la communication, nous passerons en revue certains résultats existants pour ce problème et présenterons une méthode d'approximation de l'espérance susmentionnée dans certains cas particuliers.

[MS-78]

A Model for Corporate Exit Prediction: Discrete Hazard and Competing Risks Model
Un modèle de prévision de la sortie d'une corporation du marché : modèle à défaillance discrète et à risques compétitifs

Taehan Bae & Reg. J. Kulperger, *The University of Western Ontario*

Publicly traded companies can leave a public system by bankruptcy or merger/acquisition. Discrete sub-hazard functions are modeled with a multinomial logit function. Time-varying baseline hazard functions may capture unobservable or missing macroeconomic information which affect the firms globally. Multivariate mean reverting autoregressive processes are assumed to be a prior process for the baseline hazard functions and their posterior mode estimates are considered. In sample and out of sample predictions are studied. The Poisson approximation and entropy related methods are developed and proposed for prediction accuracy measures. The methodologies are applied to 11,802 U.S. industrial firms spanning 1987 to 2006.

Les sociétés ouvertes peuvent quitter un système public par faillite ou par fusion/acquisition. Les fonctions de sous-défaillance sont modélisées par une fonction logit multinomiale. Des fonctions de défaillance de base variant avec le temps peuvent contenir de l'information macroéconomique manquante ou non observable affectant les firmes globalement. Des processus multivariés autorégressifs de retour à la moyenne sont supposés être un processus *a priori* pour les fonctions de défaillance de base et les estimateurs basés sur leur mode *a posteriori* sont considérés. Les prévisions dans et hors de l'échantillon sont étudiées. L'approximation de Poisson et des méthodes relatives à l'entropie sont développées et proposées comme mesures de précision des prévisions. Les méthodes sont appliquées à 11 802 firmes des États-Unis de 1987 à 2006.

[MS-79]

Optimal Semi-Foldover Plans for Two-Level Fractional Factorial Designs
Plans avec demi-repliement optimaux pour des plans fractionnels factoriels à deux niveaux

Miao (Mandy) Zhang, *University of Manitoba*

Two-level fractional factorial designs are commonly used in industrial screening experiments. One strategy for de-aliasing effects of interest is to run a second experiment, using a foldover design. This can be an expensive option, because it involves doubling the number of runs. A less expensive option is to use a semi-foldover design, which uses only one-half of the additional runs. The combined design is then a non-regular three-quarter fractional factorial design. For a given initial fractional factorial design, we consider the determination of the optimal semi-foldover plan, using optimality criteria commonly used for non-regular designs.

Les plans factoriels fractionnaire à deux niveaux sont fréquemment utilisés dans des expériences de dépistage industriel. Une stratégie pour déplier les effets d'intérêts est d'effectuer une seconde expérience, utilisant un plan à repliement. Cette option peut être coûteuse, car elle implique de doubler le nombre de passages. Une option moins coûteuse est d'utiliser un plan à semi-repliement, qui utilise seulement la moitié des passages supplémentaires. Le plan combiné est alors un plan factoriel irrégulier aux trois quarts fractionnaire. Pour un plan factoriel fractionnaire donné, nous considérons la détermination du plan à semi-repliement optimal, par rapport à des critères d'optimalité fréquemment utilisés pour des plans irréguliers.

[MS-80]

Simple Tests for Identifying the Holling Type II Predator-prey Model
Quelques tests simples pour identifier le modèle prédateur-proie de Holling type II

Ashkan Zahedi & Sorana Froda, *Université du Québec à Montréal*

We propose a method to discriminate among several predator-prey dynamical systems which ex-

On propose une méthode qui permet de discriminer entre divers systèmes dynamiques qui modélisent l'interaction

hibit oscillatory behaviour. One of the oldest of such sytems is the Holling type II model, which states that the solutions can spiral around an equilibrium point in the phase-plane; we base our inference on some geometric features of this solution. In order to decide whether this type of model holds, we consider observational errors and perform a few elementary tests, in particular on a regression slope and on the difference between two group means. Finally, we illustrate our method on simulated and real data.

proie-prédateur et présentent un comportement oscillatoire. Un des modèles les plus anciens est le modèle de Holling (type II) dont les solutions peuvent spiraler autour d'un point d'équilibre dans l'espace des phases ; notre inférence a comme point de départ la géométrie de cette solution. Afin de décider si ce type de modèle est valable, on considère des erreurs d'observation et on effectue des tests élémentaires, en particulier sur la pente d'une droite et sur la différence de deux moyennes. On applique notre méthode à des données simulées et réelles.

[MS-81]

Diagnostic Procedures for Functional Data Analysis

Procédures diagnostiques pour l'analyse des données fonctionnelles

Monica Sirski, *University of Manitoba*

Functional data analysis is the analysis of data where the response is a curve. Such data commonly arise in industrial settings, in both physical and computer experiments. Shen and Xu (2007) proposed several adaptations of diagnostic statistics to check model assumptions in functional situations: studentized residuals, jackknife residuals, and Cook's distance. Using data from a designed computer experiment, our goal is to investigate various diagnostic tools for functional data, including those proposed by Shen and Xu. We also examine the effect of unequally-spaced time points and the consequences of smoothing the response curves.

L'analyse de données fonctionnelles est l'analyse de données où la réponse est une courbe. Les données de ce genre sont généralement recueillies dans des conditions industrielles, dans le cadre d'expériences physiques ainsi que numériques. Shen et Xu (2007) ont proposé plusieurs adaptations des statistiques de diagnostic pour vérifier les hypothèses de modélisation dans des situations de données fonctionnelles, à savoir les résidus studentisés, les résidus jackknife et la distance de Cook. À l'aide de données provenant d'une expérience numérique planifiée, nous étudions divers outils diagnostiques applicables aux données fonctionnelles, y compris ceux proposés par Shen et Xu. Nous examinons aussi l'effet de points temporels non équidistants et les conséquences du lissage des courbes de réponse.

[MS-82]

Testing the Homogeneity in Two-sample Censored Mixture Model when Cure Proportion May Exist in One Group

Tester l'homogénéité dans un modèle de mélange à deux échantillons censurés lorsque qu'une proportion de guérison peut exister dans un groupe

Hui Song, Yingwei Peng, Dongsheng Tu & Yi Niu, *Queen's University*

In the censored mixture model, to test the homogeneity of two samples, we have to face a problem of testing infinity of parameter when cured component may exist in one group and the conditional distribution comes from exponential distribution. In this work we propose a method based on modified likelihood ratio test. As a simulation result, the MLRT is powerful when the censoring rate is not extremely heavy, then compare it with log-rank test and likelihood ratio test.

Dans le modèle de mélanges censuré, pour tester l'homogénéité de deux échantillons, nous faisons face au problème du test de l'infinité d'un paramètre lorsqu'une composante guérie peut exister dans un groupe et que la distribution conditionnelle est exponentielle. Dans ce travail, nous proposons une méthode basée sur un test du rapport des vraisemblances modifié (TRVM). Des résultats de simulation indiquent que le TRVM est puissant lorsque le taux de guérison n'est pas très élevé. Nous comparons la performance du TRVM, du test des log-rangs et du TRV.

Session 04I Lun. 26 mai • Mon. May 26, 14:35 – 16:05 Congress Hall A/E West

Case Studies 1: Handling Item Nonresponse in Surveys

Étude de cas 1 : Traitement de la non-réponse à une question dans les enquêtes

Case Studies 1: Handling Item Nonresponse in Surveys
Étude de cas 1 : Traitement de la non-réponse à une question dans les enquêtes

David Haziza, *Université de Montréal* & Gordon Kuromi, *Statistics Canada*

Despite the best efforts to maximize response, it is almost certain that some degree of nonresponse will occur in large scale surveys. The purpose of this case study is to find imputation strategies that can reduce the nonresponse bias as much as possible and to consider the problem of variance estimation in the presence of imputed data.

Malgré tous les efforts engagés pour maximiser les réponses, il est presque certain que les enquêtes à grande échelle comporteront un certain degré de non-réponse. L'objectif de cette étude de cas est de découvrir des stratégies d'imputation permettant de réduire le biais de la non-réponse dans la mesure du possible et d'étudier le problème de l'estimation de la variance en la présence de données imputées.

Session 04J Lun. 26 mai • Mon. May 26, 14:35 – 16:05 Congress Hall A/E West

Case Study 2: Effects of Genetic Variation on the Relationship between Diet and Cardiovascular Disease Risk

Étude de cas 2 : Effets de la variation génétique sur la relation entre l'alimentation et le risque de maladie cardiovasculaire

Case Study 2: Effects of Genetic Variation on the Relationship between Diet and Cardiovascular Disease Risk
Étude de cas 2 : Effets de la variation génétique sur la relation entre l'alimentation et le risque de maladie cardiovasculaire

Alison Gibbs, *University of Toronto*

Cardiovascular disease is the most common cause of death in Canada. High levels of LDL cholesterol and low levels of HDL cholesterol are known to be associated with increased risk. In this case study, students will analyse data collected during a recent clinical trial of subjects with type-2 diabetes. The purpose of the case study is to examine the effects of genetic variation in the subjects on the relationship between changes in diet and blood cholesterol level.

La maladie cardiovasculaire est la cause de décès la plus commune au Canada. Les taux élevés de cholestérol LDL et les taux réduits de cholestérols HDL sont depuis longtemps associés à un risque accru. Dans cette étude de cas, les étudiants analyseront des données collectées au cours d'un récent essai clinique de patients souffrant de diabète de type 2. L'objectif de l'étude de cas est d'examiner les effets de la variation génétique chez ces patients sur la relation entre la modification de leur alimentation et le niveau de cholestérol sanguin.

Session 05A Lun. 26 mai • Mon. May 26, 16:30 – 18:00 Capital Hall 1A

May We, Should We and Why Collect Ethnic Statistics? French and Canadian Points of View

Peut-on, doit-on et pourquoi, collecter des données sur l'origine ethnique ? Les points de vue canadien et français

[MS-85]

Collecting Ethno-cultural Data in a Multicultural Society: the Canadian Experience
La collecte de données ethnoculturelles dans une société multiculturelle : l'expérience canadienne

Tina Chui & Hélène Maheux, *Statistics Canada/Statistique Canada*

Canada has collected information on the country's ethnic origins for over 100 years. The measurement of the concept evolved over time. This presentation will discuss why Statistics Canada collects the information, outline the challenges of collecting the data, highlight the key changes to the measurement of the concept, provide information on the related concepts and their measurements as well as present the latest results from the 2006 Census.

Le Canada recueille des renseignements au sujet de l'origine ethnique de ses habitants depuis plus de 100 ans. La mesure du concept a évolué au fil du temps. Au cours de la présentation, nous examinerons les raisons pour lesquelles Statistique Canada recueille ces renseignements, nous ferons état des défis rencontrés lors de la collecte des données, nous ferons ressortir les changements majeurs qui ont été apportés à la mesure du concept, nous fournirons des renseignements sur les concepts connexes et leurs mesures et nous présenterons les derniers résultats du recensement de 2006.

[MS-86]

Can We Ask Questions about Skin Colour? An Analysis of the New French Controversy
Peut-on poser des questions sur la couleur de la peau ? Une analyse de la nouvelle polémique française

Roxane Silberman, *Directrice de recherche au CNRS/Secrétaire générale du Comité de concertation pour les données en sciences humaines et sociales*

France has recently encountered a heated debate about ethnic statistics, which was begun by a Constitutional Council decision which invalidated an amendment to a law from 1978 and by the removal of a question about skin-colour from a survey project by Ined and Insee. What are the terms and who are the key players in this debate? Is this the same debate that was raised a few years ago about another survey? We will present the chronology of this controversy to explain the context, the different aspects and the ways it has evolved.

La France vient de connaître un violent débat à propos des statistiques ethniques qui s'est soldé par une décision du Conseil constitutionnel invalidant un amendement à une loi de 1978 et par le retrait d'une question sur la couleur de la peau dans un projet d'enquête de l'Ined et de l'Insee. Quels sont les termes et les acteurs de ce débat ? Le débat est-il le même que celui surgi il y a quelques années à propos d'une autre enquête ? On présentera la chronologie de cette polémique pour en éclairer le contexte, les différents aspects et les points d'évolution.

[MS-87]

Is Collecting Data on Ethnic Origin Legitimate?
Est-il licite de recueillir des données sur l'origine ethnique ?

René Padieu, *SFdS - Commission de déontologie*

For many years in France, the inclusion or exclusion of ethnic data in statistical inquiries has been

Inclure ou non des éléments d'ethnicité dans les enquêtes statistiques fait l'objet d'une vive controverse en France de-

a matter of sharp dispute. This is rooted in national history and the present social situation (immigration, unemployment ,etc.). Discrimination and integrating immigrants are two major political issues, for which some people call for statistics. Others argue this would infringe on the principle according to which no distinctions should be made based on origin. What, then, are statisticians supposed to do? Should they refrain from creating such data? Form debates with adequate information? Disentangle misinterpretations? Allow for legitimate uses of the data while simultaneously preventing abuses?

puis plusieurs années. Cela s'explique à la fois par l'histoire du pays et par le contexte récent (immigration, chômage, etc.). L'intégration des immigrés et discriminations sont deux des thèmes politiques. Certains réclament des statistiques sur ces thèmes. Pour d'autres, ce serait contraire à l'interdiction d'établir des distinctions selon l'origine. Que doit faire le statisticien ? S'interdire ces variables ? Éclairer les débats par une information appropriée ? Expliquer les confusions de sens ? Permettre les utilisations légitimes tout en empêchant les autres ?

Session 05B Lun. 26 mai • Mon. May 26, 16:30 – 18:00 Capital Hall 2A

Mixture Models in Statistical Finance
Modèles de mélanges en finance statistique

[MS-88]

Finite Mixture of Binary Logistic Regression Analysis on Labor Supply of Mothers with Disabled Children
Mélange fini de l'analyse de régression logistique binaire sur l'emploi des mères d'enfants handicapés

Zeng-Hua Lu, *University of South Australia*, Jiahua Chen, *University of British Columbia*, Justin Trogdon, *RTI International*

This paper studies the economic effect of child disability on mother's labor supply. Examination of the population heterogeneity forms important part of our studies. We formulate an econometric model, which allows for the possible existence of heterogeneous groups of mothers, whose status in labor force are attributed to different explanatory mechanisms in terms of different sets of explanatory factors and different interactions among factors. Our findings based on the recent Australian Bureau of Statistics' survey data present strong evidence on existence of such the heterogeneity. Some aspects of statistical inference concerning our model are studied.

Cet article étudie l'incidence économique de l'invalidité des enfants sur l'emploi de leur mère. L'examen de l'hétérogénéité de la population occupe une place importante dans nos études. Nous élaborons un modèle économétrique, permettant l'existence possible de groupes hétérogènes de mères, dont le statut dans la main-d'œuvre est attribué à différents mécanismes explicatifs déterminés par différents ensembles de facteurs explicatifs et par différentes interactions entre ces facteurs. Nos conclusions, fondées sur les données de l'enquête menée récemment par le Bureau de la statistique de l'Australie, présentent des preuves solides de l'existence d'une telle hétérogénéité. Certains aspects d'induction statistique concernant notre modèle sont étudiés.

[MS-89]

Insurance Market Effects of Risk Management Metrics
Effets des mesures de la gestion des risques sur le marché de l'assurance

Weidong Tian & Carole Bernard, *University of Waterloo*

Economic efficiency is improved in the presence of an implemented risk management program of the insurer. Risk management requirements are imposed by regulators to reduce the insurers' in-

La mise en place d'un programme de gestion des risques par l'assureur accroît l'efficacité économique. Les exigences en matière de gestion des risques sont imposées par les organismes de réglementation en vue de réduire le risque

solvency risk, as well as to improve the insurance market stability. We extend the classical analysis on optimal insurance design to the case when the insurer implements regulatory requirements. Our analysis reveals that insured are better protected in the event of greater loss irrespective of the optimal design from either the insured or the insurer perspective. Therefore the overall insurance market becomes more stable.

d'insolvabilité des assureurs et d'accroître la stabilité du marché de l'assurance. Nous étendons l'analyse classique du plan d'assurance optimal au cas où l'assureur applique les exigences réglementaires. Notre analyse révèle que les assurés sont mieux protégés en cas d'une perte plus importante, indépendamment du plan optimal tant du point de vue de l'assuré que de l'assureur. Par conséquent, le marché global de l'assurance devient plus stable.

[MS-90]
Exchange Rate Dynamics with Regime Switching
Dynamique du taux de change avec changement de régimes

Yonggan Zhao, *Dalhousie University*

This paper presents a dynamic model of optimal currency returns with a hidden Markov regime switching process. We postulate a weak form of interest rate parity that the hedged risk premiums on currency investments are identical within each regime across all currencies. Both the in-sample and the out-of-sample data for the period of January 2002 – March 2005 strongly support this hypothesis.

Cet article présente un modèle dynamique de retours de devises optimaux avec un processus markovien caché de changement de régimes. Nous postulons une forme faible de parité de taux d'intérêts : les primes de risque couvertes sur les investissements en devises sont identiques à l'intérieur de chaque régime pour toutes les devises. Les données dans et hors de l'échantillon pour la période de janvier 2002 à mars 2005 soutiennent fortement cette hypothèse.

Session 05C Lun. 26 mai • Mon. May 26, 16:30 – 18:00 Capital Hall 3A

Bayesian Statistics
Statistiques bayésiennes

[MS-91]
On-line Changepoint Detection and Parameter Estimation for Genome-wide Transcript Analysis
Détection du point de changement en ligne et estimation des paramètres pour la transcription de l'analyse pour tout le génome

Raphael Gottardo, François Caron & Arnaud Doucet, *University of British Columbia*

We consider the problem of identifying novel RNA transcripts using tiling arrays. Standard approaches to this problem rely on the calculation of a sliding window statistic or on simple changepoint models. These methods suffer from several drawbacks including the need to determine a threshold to label transcript regions and/or specify the number of transcripts. Here, we propose a Bayesian multiple changepoint model to simultaneously identify the number of transcripts, the transcript boundaries and their associated levels. We present a computationally efficient online algorithm which allows us to jointly estimate both the changepoint locations and the model parameters.

Nous examinons le problème de la reconnaissance de nouvelles transcriptions de l'ARN à l'aide de réseaux d'ADNc. Les approches standards de la résolution de ce problème s'appuient sur le calcul d'une statistique à fenêtre mobile ou sur de simples modèles de points de changement. Ces méthodes souffrent de plusieurs défauts, dont la nécessité de déterminer un seuil pour étiqueter les régions de transcription ou de spécifier le nombre de transcriptions. Ici, nous proposons un modèle bayésien à points de changements multiples pour déterminer simultanément le nombre de transcriptions, les bornes de transcription et leurs niveaux connexes. Nous présentons un algorithme en ligne informatiquement efficace qui nous permet d'estimer conjointement les emplacements des points de changement et les paramètres du modèle.

[MS-92]
Bayesian Free-knot Regression Splines
Splines de régression bayésiens

James Merleau, *Institut de Recherche d'Hydro-Québec*, Jean-François Angers, *Université de Montréal*, Luc Perreault, *Institut de Recherche d'Hydro-Québec*, Anne-Catherine Favre, *INRS, Eau, Terre et Environnement*

Bayesian free-knot regression splines constitute an effective and parsimonious method to model curves. In this contribution, we extend its use to situations where observations arise from a class of continuous distributions which includes those of the exponential family. We consider approximations of the marginal distribution of the data since it plays a key role in the MCMC algorithm used to explore the space of knot configurations. Furthermore, we propose a strategy to discriminate which statistical distribution seems the most adequate for a given data set. We apply some of these ideas to a hydrological sample of curves.

Les splines de régression bayésiens représentent une méthode efficace et parcimonieuse pour modéliser des courbes. Dans cette contribution, nous prolongeons cette approche lorsque les observations proviennent d'une classe de distributions continues qui inclut celles de la famille exponentielle. Nous considérons des approximations de la distribution marginale des observations puisque celle-ci joue un rôle important dans l'algorithme MCMC utilisé pour explorer l'espace des configurations nodales. De plus, nous proposons une stratégie pour discriminer la distribution statistique qui semble la plus adéquate pour une collection d'observations donnée. Nous appliquons la méthodologie à un échantillon de courbes hydrologiques.

[MS-93]
Computational Approaches to Bayesian Model Choice
Méthodes algorithmiques pour la sélection de modèles bayésiens

Christian Robert, *Université Paris Dauphine et CREST, INSEE, Paris*, Jean-Michel Marin, *INRIA Futurs, Université Paris Sud Orsay*, Nicolas Chopin, *CREST et ENSAE, INSEE, Paris*

In this talk, we will cover recent developments of ours in the computation of marginal distributions for the comparison of statistical models in a Bayesian framework. While the introduction of reversible jump MCMC by Green in 1995 is rightly perceived as the "second MCMC revolution," its implementation is often too complex for the problems at stake. When the number of models under study is of a reasonable magnitude, there exist computational alternatives that avoid model exploration with a reasonable efficiency and we will discuss here the pros and cons of several of those methods.

Cet exposé s'attache à exposer des avancées récentes dans l'approximation des probabilités a posteriori de modèles dans un contexte bayésien. Même si l'introduction de la méthode à sauts réversibles par Green (1995) est une seconde révolution MCMC, sa mise en œuvre est souvent démesurée à l'aune des problèmes traités. Si le nombre de modèles à comparer demeure raisonnable, il existe des alternatives permettant un traitement modèle par modèles et nous en étudions plusieurs en montrant leurs avantages et leurs limites.

Session 05D Lun. 26 mai • Mon. May 26, 16:30 – 18:00 Capital Hall 4A

Markov Models in Biostatistics
Modèles markoviens en biostatistique

[MS-94]
Sampling Genetic Ancestries Conditional on Observed Genotype Data
Échantillonnage des origines génétiques conditionnellement aux données génotypiques observées

Kelly Burkett, Jinko Graham & Brad McNeney, *Simon Fraser University*

For understanding the association of genetic variability with disease outcomes, it can be useful to model the latent genetic ancestry giving rise to the sample's genetic variability. Incorporating ancestry into genetic association analysis requires Monte Carlo methods to sample from the ancestry space as this multidimensional space is too large to enumerate. This presentation describes our implementation of a Markov Chain Monte Carlo sampler for the topology, node times and data at internal nodes of the ancestral tree, conditional on the observed sequence data sampled at present. Potential applications for the sampler will also be described.

Afin de comprendre l'association entre la variabilité génétique et les issues de la maladie, il peut être utile de modéliser l'origine génétique latente qui donne lieu à la variabilité génétique de l'échantillon. L'intégration de l'origine dans l'analyse de l'association génétique requiert des méthodes de Monte-Carlo pour tirer des échantillons dans l'espace des origines, car cet espace multidimensionnel est trop grand pour être dénombré. La présente communication décrit notre mise en œuvre d'un échantillonneur de Monte-Carlo par chaîne de Markov pour la topologie, le nœud Temps et les données aux nœuds internes de l'arbre des origines, conditionnellement aux données sur les séquences observées échantillonnées au temps présent. Les applications éventuelles de l'échantillonneur seront également décrites.

[MS-95]
A New Composite-conditional-likelihood Approach for Fine Genetic Mapping
Une nouvelle approche de cartographie génétique basée sur des vraisemblance conditionnelles composite

Fabrice Larribe, *Université du Québec à Montréal*, Sabin Lessard, *Université de Montréal*

The coalescent process is frequently used to model genealogies of populations in order to do fine genetic mapping. Inferences of the coalescent's parameters are based on MCMC methods and work only for a small amount of data (as far as the number of genetic markers goes, for example). We propose to use a new approach called composite-conditional-likelihood (a weighted product of conditional likelihood which is defined on subsets of consecutive genetic markers) as a solution to obtain a reliable estimate of the likelihood. This enables us to extend the approach to any genetic distance and to a large number of genetic markers.

Le processus de coalescence peut être utilisé pour modéliser des généalogies à des fins de cartographie génétique. Les inférences des paramètres sont basées sur des méthodes MCMC et ne sont possibles que pour de petits ensembles de données (en nombre de marqueurs génétiques, par exemple). Nous proposons d'utiliser une nouvelle approche basée sur des vraisemblances conditionnelles composites (calculées comme un produit de vraisemblance conditionnelle définie sur des ensembles de marqueurs génétiques contigus), afin d'obtenir un estimé fiable de la vraisemblance, et d'être capable d'étendre l'approche à un grand nombre de marqueurs génétiques, répartis possiblement sur une grande distance génétique.

[MS-96]

A Time-varying Coefficients Illness-Death Model for the Effect of Gender on Incidence and Mortality of Dementia

Étude de l'effet du genre sur l'incidence de la démence et de la mortalité à l'aide d'un modèle « sain-malade-mort » avec covariables et coefficients de régression dépendants du temps

Ahmadou Alioum, Pierre Joly, Luc Letenneur & Daniel Commenges, *Centre de Recherche INSERM - Université de Bordeaux 2*

We propose a time-varying coefficients illness-death model to investigate differences in incidence and mortality of dementia between males and females, adjusted for other covariates. Earlier works suggest that the effect of gender on incidence of dementia and mortality rates may change with increasing age. To accommodate such situations, we consider a non-homogeneous Markov model with piecewise constant transition intensities and use regression splines to allow the intensity ratios to depend on time. The proposed extension is applied to the Paquid data. The performance of the method is evaluated using a simulation study.

Nous proposons un modèle multi-états « sain-malade-mort » avec des coefficients de régression dépendants du temps pour étudier l'effet du genre sur l'incidence de la démence et sur la mortalité. Des études précédentes ont montré que l'incidence de la démence et de la mortalité en fonction de l'âge est différente chez les hommes et les femmes. Pour modéliser ce phénomène, nous considérons un modèle de Markov non homogène avec intensités constantes par morceaux et prise en compte de variables explicatives, les risques relatifs étant modélisés par des fonctions splines. La méthode proposée est illustrée sur les données de l'étude Paquid. Le rendement de cette méthode est évalué en utilisant une étude de simulation.

[MS-97]

Design and Analysis of Sequential Clinical Trials Using a Markov Chain Transition Rate Model with Conditional Power

Planification et analyse d'essais cliniques séquentiels utilisant un modèle de taux de transitions d'une chaîne de Markov à puissance conditionnelle

Gregory Pond, *Everest Clinical Research Services/University of Toronto*, W.Y. Wendy Lou, *University of Toronto*

There exists an ongoing need for statistical designs of Phase II oncology clinical trials [P2OCT] that incorporate various types of outcome data. We introduce the finite Markov-chain imbedding technique for the design and analysis of such trials, allowing for the study of patterns and sets of patient outcomes. We performed a simulation study using two novel R-language computer codes and clinical-trial data, calculating p-values and conditional power for various Markov-chain models. Although written for the context of P2OCT, these codes could be used for any similar Markov process, potentially increasing trial efficiency and improving our statistical understanding of study results.

Il existe un besoin actuel de plans statistique d'essais cliniques oncologiques de phase II (ECOP2) incorporant différents types de réponses. Nous présentons la méthode d'emboîtement de chaîne de Markov finie pour la planification et l'analyse de telles expériences, permettant l'étude de tendances et d'ensembles de résultats chez les patients. Nous avons effectué une étude de simulation en utilisant deux nouveaux programmes R (et des données d'essai clinique) calculant les valeurs-p et la puissance conditionnelle de divers modèles à chaînes de Markov. Bien qu'écrit pour le contexte d'ECOP2, le code pourrait être utilisé pour n'importe quel processus de Markov similaire, augmentant potentiellement l'efficacité des essais cliniques et améliorant notre compréhension statistique des résultats de ces études.

[MS-98]

Analysis of Disease Progression via Multi-State Models under a Nonignorable Inspection Process
Analyse de l'évolution de la maladie à l'aide de modèles multi-états sous un processus d'inspection non ignorable

Baojiang Chen, Grace Yi & Richard Cook, *University of Waterloo*

This paper concerns progressive multi-state models for time-dependent life history data. We cast the problem under the convenient framework concerning longitudinal missing data problems. The maximum likelihood method based on an EM algorithm is used to estimate parameters and variance estimation is provided using the Louis's method. Simulation studies demonstrate that the proposed method works well under a variety of situations. A real dataset arising from a psoriatic arthritis study is analyzed with the proposed method.

La présente communication traite de modèles multi-états progressifs applicables à des données sur l'histoire de vie dépendantes du temps. Nous modélisons le problème sous le cadre commode des problèmes de données manquantes longitudinales. Nous utilisons la méthode du maximum de vraisemblance basée sur un algorithme EM pour estimer les paramètres et nous fournissons une estimation de la variance produite par la méthode de Louis. Des études par simulation démontrent que la méthode proposée donne de bons résultats dans diverses situations. Nous analysons un ensemble de données réelles tirées d'une étude sur la polyarthrite psoriasique par la méthode proposée.

[MS-99]

Finite Markov Chain Imbedding and Its Application to Matching Probability between Two Markov Dependent Biological Sequences
Insertion d'une chaîne de Markov finie et son application à la probabilité d'appariement de deux séquences biologiques markoviennes dépendantes

Tung-Lung Wu & James Fu, *University of Manitoba*

We consider a matching problem between two DNA sequences. Matching DNA sequences provides important information for exploring the relationship between or within species in biological researches and evolutionary processes. Usually the elements in a sequence are assumed to be distributed independently and identically. Nevertheless, in reality they are dependent and are often modeled as Markov models. In this presentation, we will first introduce a Finite Markov Chain Imbedding approach. Then we can easily obtain the distribution of the length of the longest matching contiguous pattern between two independent Markov dependent sequences via the Finite Markov Chain Imbedding technique.

Nous envisageons un problème d'appariement de deux séquences d'ADN. Les séquences d'ADN appariées fournissent des renseignements précieux pour l'étude de la relation entre diverses espèces ou au sein d'une espèce dans le contexte de la recherche biologique et des processus d'évolution. Habituellement, on suppose que les éléments d'une séquence sont indépendants et identiquement distribués. Néanmoins, en réalité, ils sont dépendants et souvent modélisés à l'aide de modèles de Markov. Dans la présente communication, nous commencerons par présenter une approche d'insertion de chaîne de Markov finie. Ensuite, nous pouvons obtenir facilement la distribution de la longueur du motif apparié le plus long entre deux séquences markoviennes indépendantes par la technique d'insertion d'une chaîne de Markov finie.

Session 05E Lun. 26 mai • Mon. May 26, 16:30 – 18:00 Capital Hall 5A

Analysis of Censored Data
Analyse de données censurées

[MS-100]

Estimation for the Censored Single-Index Model
Estimation pour le modèle à indice unique censuré

Zhulin He & Xuewen Lu, *University of Calgary*

We propose two estimation methods for the censored single-index model: minimum weighted average conditional variance estimation (MWAVE) and outer product of weighted gradients(OPWG). These methods use the Kaplan – Meier weights to account for censoring. The MWAVE method estimates parameters by minimizing the overall weighted approximation errors which are calculated from the response variable and the estimated smooth link function. The OPWG method works on the eigenvector corresponding to the largest eigenvalue from the outer product of weighted gradients of the estimated link function. Simulations and real data examples are used to illustrate the proposed methods.

Nous proposons deux méthodes d'estimation pour le modèle à indice unique censuré, à savoir l'estimation de la variance conditionnelle moyenne pondérée minimale (EVCMPM) et le produit externe des gradients pondérés (PEGP). Ces méthodes s'appuient sur les pondérations de Kaplan-Meier pour tenir compte de la censure. La méthode EVCMPM estime le paramètre en minimisant les erreurs globales pondérées d'approximation qui sont calculées d'après la variable de réponse et la fonction lien lisse estimée. La méthode PEGP porte sur le vecteur propre correspondant à la valeur propre la plus grande résultant du produit externe des gradients pondérés de la fonction lien estimée. Des simulations et des exemples fondés sur des données réelles sont utilisés pour illustrer les méthodes proposées.

[MS-101]

Maximum Likelihood Estimation in a Partially Observed Stratified Regression Model with Censored Data
Estimation du maximum de vraisemblance dans un modèle de régression stratifié partiellement observé avec données censurées

Amélie Detais & Jean-François Dupuy, *Université Paul Sabatier Toulouse III*

The model studied in this paper is a generalization of Cox's proportional hazard model: individuals are divided into a finite number of strata sharing the same covariate influence, each stratum having its own baseline intensity rate. Some strata belongings are assumed to be missing at random. Non-parametric maximum likelihood (NPML) is used to estimate regression parameters in this mixture model with censored data. This paper establishes the existence, consistency and asymptotic normality results for the NPML estimates. Furthermore, we derive consistent variance estimates for both parametric and non-parametric components.

Le modèle étudié dans cet article généralise le modèle de Cox à risques proportionnels : les individus sont divisés en un nombre fini de strates sur lesquelles les covariables ont la même influence, mais ayant des intensités de base différentes. L'appartenance à une strate est aléatoirement manquante. La méthode du maximum de vraisemblance non paramétrique permet d'estimer les paramètres de régression dans ce modèle de mélange à données censurées. Cet article montre l'existence, la consistance et la normalité asymptotique des estimateurs du maximum de vraisemblance non paramétrique. De plus, des estimateurs consistants de la variance asymptotique sont obtenus, à la fois pour les composantes paramétriques et non paramétriques du modèle.

[MS-102]
Inference for Treatments Targeting Control of an Intermediate Measure
Inférence dans les traitements visant le contrôle d'une mesure intermédiaire

Cecilia Cotton & Patrick Heagerty, *University of Washington*

We consider the problem of comparing the causal effects of multiple treatment regimes on survival using observational data and allowing for the possibility that subjects may be adherent to multiple regimes at once. Our approach is to duplicate or clone each subject and examine their adherence to each regime. Then inverse probability of censoring weights can be used to adjust for nonadherence and a marginal structural model (MSM) or log rank test with an appropriate variance estimate can be used to compare survival. These methods are illustrated by comparing epoetin therapy regimes with different target hemoglobin ranges in hemodialysis subjects.

Nous considérons le problème de la comparaison des effets causatifs des schémas thérapeutiques multiples sur la survie en utilisant des données d'observation et en tenant compte du fait que les sujets pourraient suivre plusieurs schémas thérapeutiques en même temps. Notre approche consiste à copier, ou à cloner, chacun des sujets et à examiner leur adhésion à chaque schéma thérapeutique. Puis, nous pouvons utiliser la probabilité inverse des poids de censure afin de corriger pour la non-observation et un modèle structurel marginal (MSM) ou un test du log-rank conjugué à une estimation appropriée de la variance pour comparer la survie. Nous illustrons ces méthodes en comparant les schémas de traitement par l'époétine avec diverses fourchettes cibles d'hémoglobine chez les sujets en hémodialyse.

[MS-103]
Testing the Homogeneity of Two Survival Functions against a Mixture Alternative Based on Censored Data
Essai de l'homogénéité de deux fonctions de survie comparé à un mélange alternatif fondé sur des données censorisées

Liting Zhu, Xianming Tan & Dongsheng Tu, *NCIC Clinical Trial Group, Queen's University*

In comparing the survival times between a treatment and placebo where only an (unknown) portion of patients in the treatment group will respond to the treatment, the traditional parametric methods ignoring the existence of mixture components in the treatment group or the nonparametric methods may not be powerful. Recently, Fu, Chen and Kalbfleisch (2006) developed a modified likelihood ratio test (MLRT) for testing homogeneity in a two sample problem with uncensored data. We extend this test to censored data and compare the type I error and power of the MLRT with that of nonparametric and parametric tests through Monte-Carlo simulations.

Si l'on compare les temps de survie chez un groupe recevant un traitement et un groupe recevant un placebo dans des conditions où une partie seulement (inconnue) des patients compris dans le groupe de traitement réagiront à ce dernier, les méthodes paramétriques classiques ne tenant pas compte de l'existence de composantes mixtes dans le groupe de traitement ou les méthodes non paramétriques pourraient ne pas être suffisamment puissantes. Récemment, Fu, Chen et Kalbfleisch (2006) ont élaboré un test du rapport de vraisemblance modifié (TRVM) pour tester l'homogénéité dans le cas d'un problème à deux échantillons avec données non censurées. Nous étendons ce test à des données censurées et comparons l'erreur de type I et la puissance du TRVM à celles des tests non paramétriques et paramétriques dans des simulations de Monte-Carlo.

[MS-104]
Numeric EM for Mixed Effects Models with Censored Response
Algorithme EM numérique pour des modèles à effets mixtes avec réponse censurée

Lin Liu & Florin Vaida, *University of California at San Diego*

We propose an EM algorithm for computing the maximum likelihood and residual maximum likeli-

Nous proposons un algorithme EM pour le calcul du maximum de vraisemblance et du maximum de vraisemblance

hood for linear and non-linear mixed effects models with censored response. In contrast with previous developments, this algorithm uses closed-form expressions at the E-step, as opposed to Monte Carlo simulation. These expressions rely on formulas for the mean and variance of a truncated multinormal distribution, and can be computed using available software. This leads to an improvement in the speed of computation of up to an order of magnitude. We apply the methodology to the analysis of the longitudinal viral load of a cohort of HIV-infected subjects.

résiduel applicable à des modèles linéaires et non linéaires à effets mixtes lorsque la réponse peut être censurée. Contrairement aux développements antérieurs, cet algorithme utilise des expressions explicites à l'étape E au lieu de simulations de Monte-Carlo. Ces expressions s'appuient sur des formules de la moyenne et de la variance d'une distribution multinormale tronquée et peuvent être calculées en utilisant les logiciels disponibles. Elles produisent un accroissement de la vitesse de calcul allant jusqu'à un ordre de grandeur. Nous appliquons la méthode à l'analyse de la charge virale longitudinale d'une cohorte de sujets infectés par le VIH.

Session 05F Lun. 26 mai • Mon. May 26, 16:30 – 18:00 Capital Hall 6A

Applied Probability
Probabilités appliquées

[MS-105]
Merry Christmas Markov Chains!
Joyeux Noël Chaînes de Markov !

Michael Rotondi, *University of Western Ontario*

A musical theme is composed of various notes (and durations) arranged in a melodic sequence. Within this presentation, the note transitions of 25 traditional Christmas carols will be explored using Markov chains. Although this approach is simplistic, approximately 50% to 70% of the melodic transitions can be predicted using only the preceding note. While accounting for note duration, limiting probabilities illustrate the natural melodic tendency towards the tonic, mediant and dominant notes of the major scale. Visual representations also show that the majority of note transitions are indeed by step.

Un thème musical est composé de diverses notes (et durées) arrangées en une série mélodique. Dans la communication, nous examinons les transitions entre les notes de 25 chants de Noël traditionnels à l'aide de chaînes de Markov. Bien que cette approche soit simpliste, de 50% à 70% environ des transitions mélodiques peuvent être prédites en utilisant uniquement la note précédente. En tenant compte de la durée de la note, les probabilités limites illustrent la tendance mélodique naturelle vers les notes toniques, médiante et dominante de la gamme majeure. Les représentations visuelles montrent aussi que la majorité des transitions entre les notes se font effectivement par échelon.

[MS-106]
Individual-based Modelling of Spatio-temporal of Forest Dynamics
Modélisation centrée sur l'individu des dynamiques spatio-temporelle forestières

Vivien Rossi, *CIRAD*, Nicolas Desassis & Fabien Campillo, *INRIA*

Faced with contemporary environmental concerns, modelling of the dynamics of natural forests is one of the major issues of computational forestry. We propose an individual-based model of the natural forest dynamics. Our model is explicitly spatialized in terms of spatio-temporal Markovian pro-

La préservation des forêts est au cSur des préoccupations environnementales actuelles. La modélisation des dynamiques des forêts naturelles est un élément incontournable pour parvenir à gérer durablement les forêts. Nous proposons un modèle individuel centré sur les dynamiques des forêts naturelles. Notre modèle est spatialement explicité par un proces-

cesses. The state variable includes the position of an individual together with other characteristics. The competition among individuals for access to resources is taken into account. We explicitly describe the Markovian dynamic and propose an efficient Monte Carlo simulation procedure. The model will be inferred with a numerical Bayesian approach on data from the Paracou experimental site (French Guiana).

sus markovien spatio-temporel. La variable d'état contient la position de l'individu plus d'autres caractéristiques. Nous explicitons la dynamique markovienne et proposons une méthode efficace de simulation type Monte-Carlo. L'inférence du modèle sera réalisée suivant une approche bayésienne numérique sur des données du site expérimental de Paracou (Guyane française).

[MS-107]

Priority Modelling of Patient Wait-Times by Using FMCI Technique for Benchmarking and Improving Canadian Health-care System

Modélisation prioritaire du temps d'attente du patient en utilisant la technique FMCI pour établir des données repères et améliorer le système de santé canadien

Hsing-Ming Chang & James C. Fu, *University of Manitoba*

Strengthening Canadian health-care by reducing patients' wait times to therapy has been for many years a major issue. Evidence based results on estimates of various wait times may not be readily available, and no benchmark can be made to compare with desirable targets of the variables.

If hospital emergency departments are considered to be like priority-queuing systems with thresholds, the Finite Markov Chain Imbedding technique can always be used to find exact probability distributions of such systems. Actual and theoretical estimation of wait times can be evaluated for making analyses and measurements to improve our health-care system.

Consolider les soins de santé canadiens en réduisant le temps d'attente des patients pour recevoir des soins est un enjeu crucial depuis un bon nombre d'années. Les preuves fondées sur l'estimation de plusieurs temps d'attente ne sont pas facilement disponibles et aucune donnée repère ne peut être obtenue pour comparer avec les objectifs souhaitables des variables.

Si les urgences des hôpitaux sont considérées comme les systèmes de files d'attente prioritaire ayant des seuils, la technique Finite Markov Chain Imbedding technique [chaîne de Markov finie combinatoire] peut toujours être utilisée pour trouver la probabilité exacte des distributions de tels systèmes. L'estimation réelle et théorique du temps d'attente peut être évaluée pour faire des analyses et des mesures pour améliorer notre système de santé.

[MS-108]

Continuous Time Dirichlet Hierarchical Models

Les modèles hiérarchiques de Dirichlet continu dans le temps

Hafedh Faires & Richard Emilion

We consider Dirichlet processes whose parameter is a measure proportional to the distribution of a continuous-time process, such as a Brownian motion or a Markov jump process. We use them to propose some Bayesian hierarchical models based on stochastic differential equations in a random environment. We propose a method for estimating the parameters of such models and illustrate it on the Black – Scholes equation in a random environment.

Nous étudions les processus de Dirichlet dont le paramètre est une mesure proportionnelle à la loi d'un processus temporel, par exemple un mouvement brownien ou un processus de saut markovien. Nous les utilisons pour proposer des modèles hiérarchiques bayésiens basés sur des équations différentielles stochastiques en milieu aléatoire. Nous proposons une méthode pour estimer les paramètres de tels modèles et nous l'illustrons sur l'équation de Black-Scholes en milieu aléatoire.

[MS-109]
Extensions of a Multi-Parameter Renewal Process
Extensions d'un processus de renouvellement multiparamétrique

Katherine Davies & John Braun, *University of Western Ontario*

Ivanoff and Merzbach (2006) and Ivanoff (2007) have introduced a spatial renewal process, and suggested its possible application to the modelling of the spread of forest fires. In this talk, we consider two extensions of this model. The first extension incorporates marking. We will discuss simulation and parameter estimation for this model. The second extension is essentially a hybrid of the Ivanoff – Merzbach process with a deterministic forest fire simulation model: Prometheus. Our goal is to develop a truly stochastic fire spread simulator which retains the underlying physical attributes of the Prometheus model.

Ivanoff et Merzbach (2006), et Ivanoff (2007) ont introduit un processus de renouvellement spatial et proposé son application éventuelle à la modélisation de la propagation des incendies de forêt. Dans la présente communication, nous considérons deux extensions de ce modèle. La première intègre le marquage. Nous discuterons de la simulation et de l'estimation des paramètres pour ce modèle. La deuxième extension est essentiellement un hybride du processus d'Ivanoff-Merzbach et du simulateur Prometheus qui est un modèle de simulation déterministe des incendies de forêt. Notre objectif est de développer un simulateur de propagation des incendies vraiment stochastique qui retient les attributs physiques sous-jacents du modèle Prometheus.

Session 05G Lun. 26 mai • Mon. May 26, 16:30 – 18:00 Capital Hall 7A

Estimation and Tests for Time Series
Estimation et tests pour séries chronologiques

[MS-110]
Neyman Smooth Test of Normality in a Time Series Context
Test lisse de Neyman dans un contexte de séries chronologiques

Pierre Lafaye de Micheaux, *LJK, Grenoble Universités*

In this presentation, we propose a goodness-of-fit test of normality for the errors of an ARMA(p,q) model with unknown mean. The orders p and q are supposed to be known and will be fixed. The test is based on the data-driven Neyman smooth test approach and is simple to perform. A simulation study will show how the test performs for small to moderate sample sizes. We will try to show how this procedure can be extended to other situations, such as, for example, more sophisticated models or to a multivariate frameork.

Dans cet exposé, nous proposons un test d'ajustement de la normalité des erreurs d'un modèle ARMA(p,q) de moyenne inconnue. Les ordres p et q du modèle sont supposés connus et seront fixés. Le test est basé sur l'approche des tests lisses adaptatifs de Neyman et est simple à mettre en œuvre. Une étude par simulations montrera comment le test se comporte pour des tailles d'échantillon petites ou moyennes. Nous tenterons de montrer comment cette procédure peut être étendue à d'autres situations, par exemple des modèles plus élaborés ou encore un cadre multivarié.

[MS-111]
Estimation of the Evolutionary Spectrum for Nonstationary Processes with Improved Boundary Behavior
Estimation du spectre évolutif pour processus non stationnaires avec comportement de frontière amélioré

Azadeh Moghtaderi, David Thomson & Glen Takahara, *Queen's University*

The evolutionary spectrum, a time-dependent analogue of the spectrum of a stationary process, was

Le spectre évolutif, un analogue dépendant du temps du spectre d'un processus stationnaire, a été développé par

developed by Priestley (1965). Existing estimators of the evolutionary spectrum suffer from bias problems in the boundary region of the time-frequency domain, due to filtering effects. Estimation near the boundary is important for spectral prediction methods and more accurate estimation of the covariance function. We propose a new evolutionary spectrum estimator which mitigates these problems. Our estimator is based on extrapolation of the evolutionary spectrum in the time and frequency directions. We apply our estimator to realizations of simulated non-stationary processes with known evolutionary spectra.

Priestley (1965). Les estimateurs existant de la théorie du spectre évolutif souffrent de problèmes de biais dans la région frontière du domaine temps-fréquence, dus à des effets de filtrage. L'estimation près de la frontière est importante pour les méthodes de prévisions spectrales et pour une estimation plus précise de la fonction de covariance. Nous proposons un nouvel estimateur du spectre évolutif qui atténue ces problèmes. Notre estimateur est basé sur l'extrapolation du spectre évolutif dans les directions du temps et de la fréquence. Nous appliquons notre estimateur à des réalisations de processus non stationnaires simulés ayant des spectres évolutifs connus.

[MS-112]
Modelling Time Series with Constraints: Looking for Similarities; Validation
Modélisation de chroniques sous contraintes : recherche de ressemblances, validation

Mireille Getler-Summa & Catherine Pardoux, *Université Paris-Dauphine, Ceremade, CNRS*, Christian Delavelle & David Fayolle, *Aji-Europe*

A set of m_i exogenous time series corresponds to each endogenous variable i in a set of time series without seasonal cycles. Models are built with less than s_i exogenous variables $\{s_i < m_i\}$. Time series have been transformed when necessary in order to be defined on the same dates without lag. The issue consists of reducing the number of target models as well as the combinatory heuristics for the exogenous variables. Therefore a 'pyramidisation' of an ascending hierarchy is proposed, the L^2 distance being the initial dissimilarities between two functions. The application deals with the French domestic transport forecasting (1985 – 2005), by road, railway and river, per product.

Pour chaque variable endogène i d'un ensemble de chroniques sans saisonnalité, on dispose de m_i chroniques exogènes. On construit des modèles avec au plus s_i variables exogènes ($s_i < m_i$). Après transformations si nécessaires, les séries sont définies sur les mêmes dates, sans retard. Le problème est de réduire le nombre des modèles recherchés et des heuristiques dans la combinatoire des variables exogènes. On propose pour cela une « pyramidisation » d'une hiérarchie ascendante, les dissimilarités initiales étant les distances L^2 entre courbes. L'application concerne la prévision du transport domestique de marchandises en France entre 1985 et 2005 par mode (routier, ferroviaire, fluvial), par produits.

[MS-113]
Regularized Autoregressive Frequency Estimation
Estimation de fréquences autorégressive régularisée

Bei Chen & Yulia, R Gel, *University of Waterloo*

Many real life processes can be modeled as a noised sum of sinusoids. Unknown frequencies can be obtained from approximating generalized spectral density of the process by an autoregressive (AR) model. AR model has a simple structure, and its estimation and asymptotic properties are well-investigated. We propose a regularized AR (RAR) approximation, which enables to estimate parame-

De nombreux processus réels peuvent être modélisés sous la forme d'une somme bruitée de sinusoïdes. Les fréquences inconnues peuvent être obtenues par approximation de la densité spectrale généralisée du processus à l'aide d'un modèle autorégressif (AR). Le modèle AR possède une structure simple, et son estimation et ses propriétés asymptotiques ont été bien étudiées. Nous proposons une approximation AR régularisée (ARR), qui permet d'estimer les pa-

ters with different level of accuracy. Therefore, the repeated model selection and parameter estimation are avoided as sample size increases. The regularizer is chosen by cross-validation. In this talk, we discuss asymptotic properties of RAR, links to other regularization techniques and present numerical studies.

ramètres avec divers degrés de précision. Par conséquent, la sélection répétée du modèle et l'estimation du paramètre à mesure que croît la taille de l'échantillon est évitée. Le régularisateur est choisi par validation croisée. Dans la présente communication, nous discutons des propriétés asymptotiques du modèle ARR et des liens à d'autres techniques de régularisation, et présentons des études numériques.

[MS-114]

Asymptotical Analysis for Bifurcating Autoregressive Processes via a Martingale Approach
Analyse asymptotique des processus autoregressifs de bifurcation par des méthodes martingales

Anne Gégout-Petit, *INRIA & Université de Bordeaux 2*, Bernard Bercu, *INRIA & Université de Bordeaux 1*, Benoite de Saporta, *INRIA & Université de Bordeaux 4*

We establish asymptotic results for the least-squares estimators of the parameters of the dynamic of a first-order bifurcating autoregressive process. The model is the following:

$$\begin{cases} X_{2n} &= a_0 + b_0 X_n + \varepsilon_{2n} \\ X_{2n+1} &= a_1 + b_1 X_n + \varepsilon_{2n+1} \end{cases}$$

where $(\varepsilon_{2n}, \varepsilon_{2n+1})_n \in N^*$ are i.i.d. We estimates a_0, b_0, a_1, b_1, $\mathrm{var}(\varepsilon_i)$ and $\mathrm{cov}(\varepsilon_{2n}, \varepsilon_{2n+1})$. We give almost sure results, a CLT and a law of the iterated logarithm. The use of martingales convergence results allows us to extend the results established by J. Guyon [2007].

Nous établissons des résultats asymptotiques pour les estimateurs des moindres carrés des paramètres de la dynamique d'un processus autorégressif de bifurcation. Le modèle est le suivant :

$$\begin{cases} X_{2n} &= a_0 + b_0 X_n + \varepsilon_{2n} \\ X_{2n+1} &= a_1 + b_1 X_n + \varepsilon_{2n+1} \end{cases}$$

avec $(\varepsilon_{2n}, \varepsilon_{2n+1})_n \in N^*$ variables i.i.d. Nous estimons a_0, b_0, a_1, b_1, $\mathrm{var}(\varepsilon_i)$ et $\mathrm{cov}(\varepsilon_{2n}, \varepsilon_{2n+1})$. Nous établissons des résultats de convergence presque sûre, un TCL et une loi du log itéré. L'utilisation des résultats de convergence de martingales nous permet d'étendre les résultats de J. Guyon [2007] et de simplifier ses hypothèses.

[MS-115]

Parameter Estimation for State Space Models using Sequential Monte Carlo Methods and State Augmentation
Méthodes de Monte-Carlo séquentielles pour l'estimation des paramètres dans un modèle à espace d'états

Scott Wile, *Dalhousie University*

Parameter estimation in stochastic dynamic systems is an important problem in many scientific fields. The development and improvement of several filtering methods for sequential Bayesian analysis has advanced their effectiveness in estimating time- varying states. However, the need for algorithms which deal with the simultaneous estimation of state variables and parameters remains. The focus of this thesis is on the estimation of unknown parameters of stochastic dynamic systems (also referred to as state space models) using Monte Carlo methods. First the maximum likelihood approach is used in conjunction with the Kalman filter for parameter estimation in the linear, Gaussian setting. For the nonlinear and possibly non-Gaussian

L'estimation des paramètres d'un système dynamique stochastique est un problème important dans un grand nombre de domaines scientifiques. Le développement des méthodes de filtrage a conduit à une amélioration des performances de l'analyse bayésienne séquentielle des systèmes à état variable dans le temps. Cependant, le besoin d'algorithmes permettant l'estimation simultanée des variables d'état et des paramètres subsiste. L'accent est mis, dans cette présentation, sur l'estimation par les méthodes de Monte-Carlo. Dans un premier temps, une approche du type maximum de vraisemblance est adoptée en conjonction avec une méthode de filtre de Kalman dans un contexte linéaire et gaussien. Le cas non linéaire et éventuellement non gaussien est abordé au moyen d'approximations fondées sur les filtres à particules. Les méthodes de type *sequential importance sampling*

setting, approximate solutions to the filtering problem are introduced in the form of particle filters. Both the Sequential Importance Resampling filter and Ensemble Kalman filter are described in detail. The self-organizing state space model of Kitagawa (1998) is implemented with both particle filters for simultaneous estimation of the state and static parameters. The utility of this approach is illustrated with the use of a simple state space model, for which the Kalman filter yields optimal estimates, and a highly nonlinear augmented state space model.

filters et les filtres Kalman d'ensembles sont décrits en détail. Le modèle de *self-organizing state space* de Kitagawa (1998) est mis en œuvre avec des filtres à particules pour l'estimation des états ainsi que pour celle des paramètres. L'utilité de cette approche est illustrée au moyen de deux modèles : un modèle simple, pour lequel le filtre de Kalman fournit des estimateurs optimaux, et un modèle hautement non linéaire.

Session 06A Mar. 27 mai • Tues. May 27, 08:45 – 10:00 Congress Hall A/E

SFdS Presidential Invited Address
Allocution de l'invité de la présidente de la SFdS

[MS-116]
Invariant Methods for Independent Component Models
Méthodes invariantes pour les modèles à composantes indépendantes

Davy Paindaveine, *Université Libre de Bruxelles*, Hannu Oja & Klaus Nordhausen, *University of Tampere*, Sara Taskinen, *University of Jyväskylä*

Although they enjoy a tremendous success in engineering (in particular, in blind source separation problems), independent component models have not attracted much attention in the statistical community. In this talk, we argue that this is very puzzling since these models (i) provide an extension of multivariate Gaussian models that is often more satisfactory than the one associated with elliptic models, and (ii) can be treated as efficiently as the latter by means of invariant semiparametric techniques. Our results are illustrated in various hypothesis testing problems (one-sample location tests, tests for multivariate independence, etc.)

Bien qu'ils connaissent un succès très important auprès des ingénieurs (notamment pour ce qui est des problèmes de séparation aveugle de sources), les modèles à composantes indépendantes restent boudés par les statisticiens. Dans cet exposé, nous démontrons que cela est réellement surprenant, car ces modèles (i) fournissent des extensions des modèles gaussiens multivariés souvent plus satisfaisantes que les modèles elliptiques, et (ii) peuvent être traités aussi bien que ces derniers au moyen de méthodes semi-paramétriques invariantes. Nous illustrons nos résultats sur divers problèmes de test d'hypothèse (le test de position à un échantillon, le test d'indépendance entre deux vecteurs aléatoires, etc.)

Session 07A Mar. 27 mai • Tues. May 27, 10:30 – 12:00 Capital Hall 1A

Data Integration for Genomic Data
Intégration de données sur le génome

[MS-117]
Genomic Data Fusion: Balancing Discovery and Interpretability
Fusion de données génomiques : équilibrer découvertes et interprétabilité

Rafal Kustra, *University of Toronto,* Adam Zagdanski, *Wrocław University of Technology*

Multivariate modeling of genomics data has two distinct challenges: very high dimensionality of the problem, and wealth of systematic biological information external to the experimental data at hand. In this talk we explore one way of utilizing the information in the GeneOntology project when clustering genes in a large microarray experiment. We show that "biasing" the clustering procedure towards GeneOntology reduces the variability of clustering results and improves their biological relevance. We will also show some tools for choosing a fair balance between experimental data and prior knowledge.

La modélisation multivariée de données génomiques comporte deux défis distincts : la très haute dimensionnalité du problème, et la richesse de l'information biologique systématique externe aux données expérimentales. Dans cet exposé, nous explorons une façon d'utiliser l'information du projet GeneOntology lors de la classification de gènes dans une expérience à micropuce de haute dimension. Nous montrons que « biaiser » la procédure de classification vers GeneOncology réduit la variabilité des résultats de la classification et augmente leur pertinence biologique. Nous allons également présenter quelques outils permettant de choisir un équilibre juste entre les données expérimentales et les connaissances *a priori*.

[MS-118]
Integrating High-Throughput Genomic Data Using Sparse Canonical Correlation Analysis
Intégration de données génomiques haut débit par analyse parcimonieuse des corrélations canoniques

Joseph Beyene, Elena Parkhomenkho & David Tritchler, *University of Toronto*

In the analysis of high-throughput data, linear combinations resulting from standard canonical correlation analysis may lack biological interpretability. Moreover, small sample size may lead to computational problems and poor generalizability. Methods such as Sparse Principal Component Analysis consider each type of variables separately and focus on the correlation within each set of measurements rather than between sets. Here we propose Sparse Canonical Correlation Analysis (SCCA) in the context of high-throughput data integration. SCCA maximizes the correlation between the subsets of variables of different types while performing variable selection. We assess the method using simulated data and illustrate with real data.

Dans l'analyse des données haut débit, les combinaisons linéaires qui résultent de l'analyse classique des corrélations canoniques manquent parfois d'interprétabilité biologique. En outre, la petite taille des échantillons peut donner lieu à des problèmes de calcul et à une capacité de généralisation médiocre. Des méthodes telles que l'analyse en composantes principales parcimonieuse prennent en considération chaque type de variables séparément et se concentrent sur la corrélation dans chaque ensemble de mesures plutôt qu'entre les ensembles. Ici, nous proposons une analyse parcimonieuse des corrélations canoniques dans le contexte de l'intégration de données haut débit. Cette analyse maximise la corrélation entre les sous-ensembles de variables de divers types, tout en exécutant la sélection des variables. Nous évaluons la méthode en utilisant des données simulées et nous l'illustrons à l'aide de données réelles.

[MS-119]

How to Incorporate Expert Biological Knowledge into Conjugate Priors for Discrete Graphical Models?
Comment incorporer les connaissances d'experts en biologie à des *a priori* conjuguées pour des modèles graphiques discrets ?

Jinnan Liu, *Samuel Lunenfeld Research Institute, Mount Sinai Hospital, and York University*, Sohee Kang, *Samuel Lunenfeld Research Institute,Mount Sinai Hospital*, Helene Massam, *Samuel Lunenfeld Research Institute, Mount Sinai Hospital, and York University*, Laurent Briollais, *Samuel Lunenfeld Research Institute,Mount Sinai Hospital*

Data collected in genetic studies often comes in as a multi-dimensional contingency table, where each dimension corresponds to a gene or disease status. Bayesian graphical models offer a general framework to identifying possible links between the genes and disease studied. It is of interest to specify a prior that is mathematically convenient and also incorporates expert biological knowledge. We propose a conjugate prior — the hyper Dirichlet distribution, under a different parameterization that easily reflects the expert believes on the biological process. We will contrast the use of informative and non-informative prior using both simulated and real data sets.

Les données recueillies en études génétiques ont souvent la forme de tableaux de contingence multidimensionnels, où chaque dimension correspond à un gène où un état de maladie. Les modèles graphiques bayésiens offrent un cadre général pour identifier les liens possibles entre les gènes et la maladie étudiée. Il est d'intérêt de spécifier une loi *a priori* qui est mathématiquement pratique et qui incorpore la connaissance d'experts. Nous proposons une loi conjuguée – la distribution hyper Dircihlet, sous une paramétrisation différente qui reflète facilement les connaissances de l'expert sur le processus biologique. Nous contrastons l'utilisation d'*a priori* informatifs et non informatifs en utilisant des jeux de données réels et simulés.

[MS-120]

Collecting Functional Genomics Data for Association Studies and Gene Prioritization
Collecte de données génomiques fonctionnelles pour les études d'association et de priorisation des gènes

Gary Bader, *University of Toronto*

Increasingly, the power of genetic association studies to make new discoveries depends on having enough sample data to identify significant marker (e.g. SNP or gene) combinations. Biologists know that gene products work together in processes, so new methods that choose relevant marker combinations based on prior biological knowledge may help reduce the number of tested combinations. "Systems biology" technologies have generated huge amounts of this data, but unfortunately, this information is still difficult to combine and use. I will review sources of prior information including integrated functional association networks, molecular interactions and biological pathways.

De plus en plus souvent, le pouvoir qu'ont les études d'association génétique de mener à de nouvelles découvertes dépend de l'existence d'une quantité suffisante de données échantillonnées pour identifier des combinaisons significatives de marqueurs (p.ex. SNP ou gène). Les biologistes savent que les produits géniques fonctionnent de concert dans les processus, si bien que de nouvelles méthodes permettant de choisir les combinaisons pertinentes de marqueurs en se fondant sur des connaissances biologiques préalables permettraient peut-être de réduire le nombre de combinaisons testées. Les technologies de la « biologie des systèmes » ont produit d'énormes quantités de données de ce genre, mais malheureusement, cette information est encore difficile à combiner et à utiliser. Je passerai en revue les sources d'information a priori, y compris les réseaux intégrés d'associations fonctionnelles, les interactions moléculaires et les voies biologiques.

Session 07B Mar. 27 mai • Tues. May 27, 10:30 – 12:00 Capital Hall 2A

Session II in Memory of André Dabrowski
Séance II à la mémoire d'André Dabrowski

[MS-121]

Poisson Limits for Empirical Point Processes
Limite de Poisson pour les processus de point empiriques

Rafał Kulik, André Dabrowski & Gail Ivanoff, *University of Ottawa*

I will discuss Andre's last paper written jointly with Gail Ivanoff and Rafał Kulik.

Define the scaled empirical point process on an independent and identically distributed sequence as the random point measure with masses at $a_n^{-1}Y_i$. For suitable scaling we obtain the weak limit of these processes through a novel use of a dimension-free method based on the convergence of compensators of multiparameter martingales. The method extends previous results in several directions. The joint limit of the empirical process evaluated at distinct points is given by independent Poisson processes. These results also hold for multivariate Y_i with little additional effort. Applications are provided both to nearest-neighbour density estimation in high dimensions, and to the asymptotic behaviour of multivariate extremes such as those arising from bivariate normal copulas.

Also, I will discuss connections between this paper and the other Andre's papers on point processes.

Je présenterai le dernier article d'André (avec Gail Ivanoff et Rafał Kulik).

Je définis le processus de point empirique basé sur une distribution indépendante et identique de variable aléatoire. En escaladant nous obtenons la limite de ces processus. Nous utilisons une méthode fondée sur la convergence de martingales de multiparamètres. La méthode étend des résultats précédents. La limite du processus empirique évalue à des points distincts par les processus de Poisson indépendants. Ces résultats sont aussi vrais pour le cas multivariés avec peu d'effort supplémentaire. Nous donnons des applications pour l'estimation de densité, et de comportement asymptotique d'extrêmes multivariés tels que ceux se produisant à partir de copules normales bivariées.

[MS-122]

Return and Value at Risk Using the Dirichlet Process
Rendement et valeur à risque utilisant un processus de Dirichlet

Mahmoud Zarepour, *University of Ottawa*, Thierry Bedard, *Ontario Teachers' Pension Plan*, André Dabrowski, *University of Ottawa*

This paper introduces an alternative methodology to model-based simulation by using a Monte Carlo simulation of the Dirichlet process to estimate the value at risk. The model is constructed in a non-parametric Bayesian framework. A notable advantage of this model is that on average the random draws are sampled from a mixed distribution which consists of a prior guess of an expert and the empirical process based on a random sample of historical asset returns. The method is can be counted as a machine learning tool and the estimate is updated as new data arrive.

Cet article présente une méthode alternative à la simulation à base de modèles en utilisant une simulation de Monte-Carlo du processus de Dirichlet pour estimer la valeur à risque. Le modèle est construit dans un cadre bayésien non paramétrique. Un avantage notable de ce modèle est qu'en moyenne les tirages aléatoires sont échantillonnés d'une distribution mixte consistant en une prévision *a priori* faite par un expert et un processus empirique basé sur un échantillon aléatoire de rendements d'actifs historiques. Cette méthode peut être considérée comme un outil d'apprentissage automatique et l'estimé est mis à jour lorsque de nouvelles données sont disponibles.

[MS-123]
Large Deviations of Multitype Queues
Grandes déviations de files d'attente multitype

David McDonald & Andre Dabrowski, *University of Ottawa*, Jiyeon Lee, *Yeungnam University*

Consider a multitype queue where customers are served in their order of arrival at a rate which depends on the customer type. The stationary distribution is not of product form but defines the Markov additive process with the type served being the Markovian part and the size of the queue being the additive part. We derive sharp asymptotics of the steady state distribution by finding a harmonic function h for this Markov additive process. The h-transformation produces a change of measure that increases the arrival rate and decreases the departure rate, thus making large deviations common.

Imaginons une file d'attente ou les clients sont servis dans l'ordre d'arrivée à une vitesse qui dépend du type du client. La distribution stationnaire n'a pas une forme produite on peut représenter ce système comme un processus de Markov additif ou le type du client est la composante markovienne et le nombre de clients est la composante additive. Nous obtenons la distribution asymptotique lorsque la file devient grande en trouvant une fonction harmonique h pour ce processus de Markov additif. La transformation-h effectue un changement de mesure où le taux d'arrivée croît et le taux de service décroît ce qui augmente la probabilité de grandes déviations.

Session 07C Mar. 27 mai • Tues. May 27, 10:30 – 12:00 Capital Hall 3A

History of Statistical Science
Histoire de la statistique

[MS-124]
Some Early Uses of the Least-squares Method.
Quelques emplois précoces de la méthode des moindres carrés

Antoine de Falguerolles, *Institut de Mathématiques, Laboratoire de Statistique et Probabilités, Université de Toulouse (III)*

Adrien-Marie Legendre (1752 – 1833) is the author of the first publication on the method of least-squares, and it was he who coined the name (1805, 1806). In his publication, Legendre presents an illustrative example in geodetics where he considers two quantitative regressors, and assumes correlated measurement errors.

How was the method received by engineers? This presentation has two objectives. The first is to recall Legendre's data modeling in present-day notation. The second is to illustrate early acceptances of the method of least-squares in France (1834) and Spain (1858) in situations where both quantitative and qualitative (a group factor) regressors are simultaneously considered.

Adrien-Marie Legendre (1752-1833) est l'auteur de la première publication sur la méthode des moindres carrés, dont il a forgé le nom (1805, 1806). Dans sa publication, Legendre présente un exemple illustratif en géodésie où il considère deux régresseurs quantitatifs, et suppose des erreurs de mesures corrélées.

Comment la méthode fut-elle reçue par les ingénieurs ? Cet exposé a deux objectifs. Le premier est de rappeler la modélisation de Legendre dans des notations actuelles. Le second est d'illustrer des recours précoces à la méthode des moindres carrés en France (1834) et en Espagne (1858) dans des situations où un régresseur quantitatif et un régresseur qualitatif (un facteur groupe) sont simultanément considérés.

[MS-125]
The Golden Age of Statistical Graphics
L'âge d'or des graphiques statistiques

Michael Friendly, *York University*

Statistical graphics and data visualization have long histories, but their modern forms began only in the early 1800s. Between roughly 1850 to 1900, there occurred an explosive growth in the use of graphic methods and the range of topics to which they were applied. Innovations were prodigious, some of the most exquisite graphics ever produced appeared, resulting in the "Golden Age of Statistical Graphics."

I trace the origins of this period in terms of the infrastructure required to produce this explosive growth. To illustrate, I describe some specific contributions that give rise to the appellation "Golden Age".

Les graphiques statistiques et la visualisation des données ont une longue histoire. Toutefois, leur représentation moderne date du début de 1800. Entre 1850 et 1900, il s'est produit une importante croissance de l'utilisation des méthodes graphiques et de l'étendue des sujets auxquels elles s'appliquaient. Les innovations étaient prodigieuses et certain des graphiques les plus détaillés ont alors fait leur apparition. C'était l'âge d'or des graphiques statistiques.

Je retracerai les origines de cette période en ce qui concerne les infrastructures requises qui ont mené à cette importante croissance. À titre d'illustrations, je décrirai certaines contributions qui ont fait naître l'appellation « âge d'or ».

[MS-126]
De Morgan in the Pre-history of Statistical Hypothesis Testing
De Morgan et la préhistoire des tests d'hypothèse statistique

Adrian Rice, *Randolph-Macon College*, Eugene Seneta, *University of Sydney*

Although the research of the 19th-century mathematician Augustus De Morgan in formal logic is fairly familiar to historians of mathematics, his work in probability is largely unknown to the modern reader. For this reason, few would be aware that this work contains a self-admitted error in probabilistic reasoning. This mistake is particularly intriguing because it also appears to be an early example of hypothesis testing, a topic of much controversy in the development of mathematical statistics in the 20th century. This paper examines the mathematical and historical details of De Morgan's error.

Bien que la recherche du mathématicien du 19e siècle Augustus De Morgan en logique formelle est relativement connue des historiens des mathématiques, ses travaux sur la probabilité son largement méconnus du lectorat moderne. C'est pourquoi, peu de mathématiciens seraient au courant que ces travaux contiennent des erreurs qu'il a admises au sujet du raisonnement probabiliste. Ces erreurs sont particulièrement intrigantes, car elles semblent également être un exemple précoce du test d'hypothèse, un sujet des plus controversé dans les développements des statistiques mathématique au 20^{e} siècle. Cet article examine les détails mathématiques et historiques de l'erreur du professeur Morgan.

Session 07D Mar. 27 mai • Tues. May 27, 10:30 – 12:00 Capital Hall 4A

Model Selection
Sélection de modèles

[MS-127]
Context Tree Estimation via Algorithm Context and Penalized Maximum Likelihood
Identification d'arbres de contexte : l'algorithme "context" et le maximum de vraisemblance pénalisé

Aurélien Garivier, *CNRS, LTCI, Telecom ParisTech*, Florencia Leonardi, *Universidade de São Paulo*

Stochastic chains with memory of variable length (VLMC, or Context Tree Sources) constitute a class of processes including Markov Chains, but are potentially much more parsimonious. We show how to identify a VLMC model from data thanks to two low-complexity estimators: one resulting from the Context and Penalized Maximum Likelihood algorithm. The analysis of their consistency results from self-normalized exponential concentration inequalities for martingales.

Les chaînes stochastiques d'ordre variable (VLMC, ou sources à arbres de contexte) sont une classe de processus qui généralisent les chaînes de Markov tout en étant potentiellement bien plus parcimonieux. Nous montrons dans cet exposé comment identifier un modèle de VLMC grâce à deux estimateurs facilement calculables : celui que fournit l'algorithme « context » et le maximum de vraisemblance pénalisé. L'analyse de leur consistance est faite au moyen des inégalités de concentration exponentielles autonormalisées pour martingales.

[MS-128]
Adaptive Estimation of the Conditional Distribution Function in the Presence of Censoring
Estimation adaptative de la fonction de répartition conditionelle en présence de censure

Claire Lacour, *Université Paris Descartes*, Elodie Brunel, *Université de Montpellier*, Fabienne Comte, *Université Paris Descartes*

We consider a sample of n observations (X_i, Y_i) independent and identically distributed. The aim of this talk is to estimate the conditional cumulative distribution function of Y given X. Estimators are built via the minimization of an original contrast (which is a least squares contrast in the x-direction and a projection contrast in the y-direction) on anisotropic spaces. The procedure is adapted to positive censored random variables, i.e., when only $Z_i = inf(Y_i, C_i)$ and $\delta_i = IY_i \leq C_i$ are observed with C_i an i.i.d. sequence independent of (X_i, Y_i). A selection model method allows us to obtain an adaptive estimator. Some simulations illustrate the talk.

On considère un échantillon de n observations (X_i, Y_i) indépendantes et identiquement distribuées. Le but de cet exposé est d'estimer la fonction de répartition conditionnelle F de Y sachant X. On construit des estimateurs en minimisant sur des espaces anisotropes un contraste original, de type moindres carrés en abcisse et de type projection en ordonnée. On étend cette procédure au cas de données censurées lorqu'on observe seulement $Z_i = inf(Y_i, C_i)$ et $\delta_i = IY_i \leq C_i$ avec C_i une suite i.i.d. indépendante de (X_i, Y_i). Une méthode de sélection de modèles permet d'obtenir un estimateur adaptatif. Quelques simulations complèteront cet exposé.

[MS-129]
Thresholding Estimation of a Poisson Intensity with Unknown or Infinite Support
Estimation par seuillage de l'intensité d'un processus de Poisson à support inconnu ou infini

Vincent Rivoirard, *Université Paris Sud et ENS Paris*, Patricia Reynaud-Bouret, *CNRS - ENS Paris*

Our goal is to estimate the intensity f of a Poisson process, where the support of f is unknown and

Notre but est d'estimer l'intensité f d'un processus de Poisson. En utilisant une décomposition par ondelettes de f, nous

may be infinite. The intensity f is decomposed on a wavelet basis and we propose a random thresholding procedure. We first show the optimality of our procedure in the oracle and minimax settings. The second part of our talk deals with optimal calibration of the threshold and for this purpose we establish oracle inequalities for special functional classes. These theoretical results are extended by simulations performed on a wide range of signals.

proposons une procédure par seuillage aléatoire qui ne suppose pas connu le support de f. En particulier, ce dernier peut être infini. Nous montrons l'optimalité de notre procédure dans le cadre des approches oracle et minimax. La deuxième partie de notre exposé porte sur la calibration optimale du seuil de notre procédure. Ce travail s'appuie sur des inégalités oracle établies pour des classes fonctionnelles particulières. Il est prolongé par des simulations réalisées sur une large gamme de signaux.

Session 07E Mar. 27 mai • Tues. May 27, 10:30 – 12:00 Capital Hall 5A

Estimation Methods for Surveys Data
Méthodes d'estimation pour données de sondages complexes

[MS-130]
Hierarchical Bayesian Modelling of Two-Stage Capture-Recapture Experiments
Modélisation bayésienne hiérarchique des expérimentations à deux étapes de la capture et de la recapture

Simon Bonner & Carl Schwarz, *Simon Fraser University*

Populations of salmon are commonly assessed by simple two stage capture-recapture experiments. Each day, a sample of individuals is marked and returned to the population, and a second sampled, containing marked and unmarked individuals, is later obtained.

A key assumption in models of such experiments is that all individuals have the same probability of inclusion in the second sample. Standard methods may produce inaccurate estimates when this is violated.

Our method combines hierarchical Bayesian models of the capture probabilities with smoothing of the daily counts. This allows us to identify anomalies in the capture probabilities and correct the population estimates.

Habituellement, les populations de saumon sont évaluées à l'aide de simples expériences de capture-recapture en deux étapes. Chaque jour, un échantillon d'individus est marqué et remis dans la population et un deuxième échantillon, contenant des individus marqués et non marqués, est obtenu plus tard.

Une hypothèse fondamentale des modèles de ce genre d'expériences est que tous les individus ont la même probabilité d'inclusion dans le deuxième échantillon. Les méthodes classiques peuvent produire des estimations inexactes quand cette hypothèse est violée.

Notre méthode combine des modèles hiérarchiques bayésiens des probabilités de capture avec lissage des dénombrements quotidiens. Cela nous permet de repérer les anomalies dans les probabilités de capture et de corriger les estimations de population.

[MS-131]
GEE Analysis of Longitudinal Surveys with Imputation for Missing Responses
Analyse par EEG de données d'enquêtes longitudinales avec imputation pour les réponses manquantes

Ivan Carrillo & Changbao Wu, *University of Waterloo*

We present results on the consistency and the variance of estimators obtained through the Generalized Estimating Equation (GEE) approach, using complex longitudinal survey data with imputation for missing response values. We discuss the use of the hot-deck, both weighted and unweighted,

Nous présentons les résultats sur la convergence et la variance des estimateurs obtenus par l'approche des équations d'estimation généralisée (EEG), en utilisant des données provenant d'enquêtes longitudinales complexes avec imputation pour les valeurs de réponse manquantes. Nous discutons de l'utilisation des méthodes d'imputation hot-deck,

and the nearest neighbour imputation methods under the current setting, and report some simulation results using data from the National Longitudinal Survey of Children and Youth (NLSCY).

pondérées ainsi que non pondérées, et d'imputation par le plus proche voisin dans les conditions courantes et présentons les résultats de certaines simulations à l'aide de données provenant de l'Enquête longitudinale nationale sur les enfants et les jeunes (ELNEJ).

[MS-132]

A Resampling Approach to Estimate Variance Components of Multilevel Models

Une approche de rééchantillonnage pour estimer les composantes de la variance de modèles multiniveaux

Zilin Wang, *Wilfrid Laurier University*, Mary Thompson, *University Of Waterloo*

With the presence of unequal probability sampling in a multilevel model, the weight inflated estimators for variance components can be biased even though the use of survey weights results in design consistent estimators of the parameters. In this paper, we propose a resampling method, and with this resampling method, we derive bias correction terms to improve the accuracy of the underlying estimators of the variance components. The small sample properties of the estimators of parameters for a multilevel model from this resampling method are studied through simulations using populations generated from a simple random model.

En présence d'échantillonnage avec probabilités inégales dans un modèle multiniveaux, les estimateurs pondérés des composantes de la variance peuvent être biaisés, même si l'utilisation des poids de sondage produit des estimateurs des paramètres convergents sous le plan de sondage. Dans la présente communication, nous proposons une méthode de rééchantillonnage grâce à laquelle nous dérivons des termes de correction du biais en vue d'améliorer la précision des estimateurs sous-jacents des composantes de la variance. Nous étudions les propriétés en petit échantillon des estimateurs des paramètres d'un modèle multiniveaux découlant de cette méthode de rééchantillonnage à l'aide de simulations utilisant des populations générées à partir d'un modèle aléatoire simple.

[MS-133]

Calibration for Surveys using Uncertain Auxiliary Information

Calage d'une enquête sur une information auxiliaire incertaine

Olivier Sautory, *CEPE-INSEE*, Jean-Claude Deville, *ENSAI-CREST*

Calibration methods allow the improvement of the accuracy of the survey estimates using exact auxiliary information. In practice, it happens that the auxiliary totals are known with errors. Jean-Claude Deville gave theoretical developments dealing with this context, in terms of estimators and variances (Statistics Canada Symposium, 1999). This presentation will show how we can apply theses results to household surveys conducted at INSEE, which often use auxiliary information coming from a "large" survey (labour force survey).

Les méthodes de calage permettent d'améliorer la précision des estimateurs provenant d'une enquête par sondage en utilisant une information auxiliaire exacte. Dans la pratique, il arrive que les totaux auxiliaires soient connus avec une certaine erreur. Jean-Claude Deville a apporté des éléments théoriques adaptés à ce contexte, en termes d'estimateurs et de leur précision (symposium de Statistique Canada de 1999). La communication montrera comment peuvent s'appliquer ces résultats dans le cadre des enquêtes auprès des ménages de l'INSEE, qui utilisent souvent une « grande » enquête (l'enquête emploi) comme source d'information auxiliaire.

[MS-134]
Optimal Sampling and Estimation Strategies under Linear Model
Stratégies d'échantillonnage et d'estimation optimales sous un modèle linéaire

Yves Tillé & Desislava Nedyalkova, *Université de Neuchâtel*

In survey sampling, model-based and model-assisted inferences can lead to very different estimators. These two paradigms are however far from being so different if we search for an optimal strategy rather than just an optimal estimator. One strategy is the use of a couple composed of a sampling design and an estimator. We will show that, under a linear model, the optimal model-assisted and model-based strategies are often the same. It is then possible to construct an inference for both the design-based and model-based frameworks.

En statistique d'enquête, l'inférence basée sur le modèle et l'inférence assistée par un modèle peuvent conduire à des estimateurs radicalement différents. Ces deux paradigmes sont pourtant loin d'être si différents si nous cherchons une stratégie optimale plutôt qu'un estimateur optimal, une stratégie étant un couple composé d'un plan de sondage et d'un estimateur. Nous allons montrer que, sous un modèle linéaire, les stratégies optimales sont souvent identiques qu'elles soient basées sur un modèle ou seulement assistées par un modèle. Il est alors possible de construire une inférence valide à la fois sous le modèle et sous le plan de sondage.

[MS-135]
Linearized Variance Estimation from Simulated Census Data
Estimation de la variance linéarisée à l'aide de données de recensement simulées

Abdellatif Demnati, *Statistique Canada*, J.N.K. Rao, *Carleton University*

Simulated census data are often generated from a probability sample by imputing for the non-sampled units and sample non-respondents, using auxiliary variables available for the population units. Composite imputation with two or more different imputation methods is also used; for example, the value from an administrative file when it is available and regression imputation otherwise. A naïve estimator of a total based on the simulated census total and a design consistent estimator are studied, and associated linearization variance estimators are obtained by using the unified variance estimation approach of Demnati and Rao (Survey Methodology, 2004). Simulation results are also presented.

Il est fréquent de produire des données de recensement simulées à l'aide d'un échantillon probabiliste en faisant une imputation pour les unités non échantillonnées et les non-répondants dans l'échantillon, à l'aide des variables auxiliaires disponibles pour les unités de population. L'imputation composite regroupant deux méthodes d'imputation ou plus est également utilisée ; par exemple, l'utilisation de la valeur provenant d'un fichier administratif quand elle est disponible et l'imputation par la régression autrement. Nous étudions un estimateur naïf d'un total fondé sur le total de recensement simulé et un estimateur convergent sous le plan, et nous obtenons les estimateurs de variance par linéarisation connexes en utilisant l'approche d'estimation de la variance unifiée de Demnati et Rao (Techniques d'enquête, 2004). Nous présentons aussi les résultats d'études par simulation.

Session 07F Mar. 27 mai • Tues. May 27, 10:30 – 12:00 Capital Hall 6A

Semiparametric and Nonparametric Models
Modèles semi-paramétriques et non paramétriques

[MS-136]

Two-Sample Rank Tests for Nonnegative Observations with Multivarite Applications
Tests à deux échantillons pour des observations non-négatives avec des applications multivariées

Jureckova Jana, Jan Kalina & Marek Omelka, *Charles University in Prague*

Two-sample rank tests are studied along with their power functions under the contiguous Lehmann alternatives. As an application, we suggest multivariate versions of the Wilcoxon, Kolmogorov – Smirnov, Psi- and Savage tests, that are not only consistent, but also locally and asymptotically unbiased and distribution free under the hypothesis as well as under the Lehmann alternatives.

Des tests à deux échantillons sont étudiés ainsi que leurs fonctions de puissance selon des alternatives adjacentes de Lehman. En tant qu'application, nous suggérons les versions multivariées des tests de Wilcoxon, Kolmogorov-Smirnov, Psi- et Savage. Ils ne sont pas compatibles, mais ils sont non biaisés localement et asymptotiquement et sont non paramétrique sous l'hypothèse ainsi que sous les alternatives de Lehmann.

[MS-137]

Two-step Estimation in a Multivariate Semiparametric Sample Selection Model
Estimation en deux étapes dans un modèle de sélection semi-paramétrique multivarié

Jérôme Saracco, *Université Bordeaux 4*, Benoit Liquet, *ISPED, Bordeaux*, Marie Chavent, *Université Bordeaux 2*

We consider a multivariate semiparametric sample selection model and we develop a geometric approach to the estimation of the slope vectors in the outcome equation and in the selection equation. Contrary to most existing methods, we deal symmetrically with both slope vectors. The estimation method is link-free and distribution-free, and it works in two main steps: a multivariate Sliced Inverse Regression step, and a Canonical Analysis step. We establish root n-consistency and asymptotic normality of the estimates. We give results from a simulation study in order to illustrate the estimation method.

Nous considérons un modèle semi-paramétrique multivarié de sélection et nous développons une approche géométrique pour l'estimation des vecteurs des paramètres de l'équation de sélection et de l'équation d'observation. Contrairement à la plupart méthodes existantes, la méthode proposée traite symétriquement les deux vecteurs de paramètres ; elle fonctionne en deux étapes : une étape de régression SIR multivariée, suivie d'une étape d'analyse canonique. Nous avons établi la convergence en probabilité et la normalité asymptotique de nos estimateurs. Nous illustrons le bon fonctionnement pratique de la méthode sur des simulations numériques.

[MS-138]

Adaptive Estimation by Aggregation for Marked Counting Processes
Estimation adaptative par agrégation pour des processus de comptage marqués

Stéphane Gaïffas & Agathe Guilloux, *LSTA - Université Paris 6*

We present an adaptive aggregation estimator of the intensity of a marked counting process. The idea is to mix penalized empirical risk minimizers working in different models (with or without reduction of the dimension, for instance single-index). We

Nous présentons un estimateur adaptatif de l'intensité d'un processus de comptage marqué obtenu par agrégation. L'idée est d'agréger (mélanger) des estimateurs obtenus par minimisation du risque empirique pénalisé (avec ou sans réduction de la dimension, par exemple single-index). Nous

show that this provides an estimator that possesses strong adaptation properties. Indeed, the mixing (or aggregated) estimator is adaptive over several smoothness classes and for several model structures simultaneously.

montrons que l'estimateur ainsi obtenu a de fortes propriétés d'adaptation. En effet, l'estimateur agrégé est adaptatif pour différentes classes de régularité et différentes structures de modèle simultanément.

[MS-139]

Non- and Semiparametric Tests for Conditional Independence in Two-way Contingency Tables
Tests non-paramétrique et semi-paramétriques d'indépendance conditionnelle dans une table de contingence à deux entrées

Geenens Gery &, Léopold Simar, *Institut de Statistique, Université catholique de Louvain*

The classical chi-square test of independence between two categorical variables R and S suffers from the implicitly assumed homogeneity of the population. Yet some characteristics of each individual can be associated with R and S, and influence the dependence structure of the underlying contingency table. Along with this, a generalization of the chi-square test is proposed, testing for the conditional independence of R and S given a vector of covariates. The conditional distributions of R and S are nonparametrically estimated, and a divergence criterion is built. Also, based on a Single-Index assumption, the test is adapted to a semiparametric setting.

Le test du khi-carré classique d'indépendance entre deux variables catégorielles R et S souffre de l'hypothèse implicite d'homogénéité de la population. Or, certaines caractéristiques propres à chaque individu sont probablement associées à R et S, et influencent la structure de dépendance de la table de contingence sous-jacente. En ce sens, une généralisation du test du khi-carré est proposée, testant l'indépendance conditionnelle entre R et S connaissant un vecteur de covariables. Les distributions conditionnelles de R et S sont estimées non paramétriquement, et un critère de divergence est construit. Aussi, en se basant sur une hypothèse de direction révélatrice unique, le test est adapté à un contexte semi-paramétrique.

[MS-140]

Semiparametric Test of Indepedence between Random Vectors
Test d'indépendance semi-paramétrique entre vecteurs aléatoires

Bernard Colin & Ernest Monga, *Université de Sherbrooke*

We propose, in the framework of a semi-parametric approach, a procedure leading to a test of the hypothesis of independence between n vectors of finite dimension. Some interesting results have already been obtained in a case of a semi-parametric test tackling the problem of the independence of components of a random vector. We'll adapt those results to the case of the hypothsesis of independence of finite blocks of a random vector. A test resulting from these considerations will be proposed and illustrated using examples involving copulas. Some simulations will be performed in order to obtain an empirical evaluation of the power of the proposed test.

Étant donné un ensemble de n vecteurs aléatoires de dimensions finies, nous proposons, dans le cadre d'une approche semi-paramétrique, une procédure de test afin de déterminer si l'hypothèse d'indépendance entre ces vecteurs est réaliste. Nous appuyant sur des résultats analogues, obtenus dans le cas d'un test semi-paramétrique d'indépendance entre les composantes d'un vecteur aléatoire, nous exploitons ces derniers dans le cadre indiqué ci-dessus. Une procédure de test découlant de ces considérations sera proposée et illustrée à l'aide d'exemples portant sur des copules. Quelques simulations seront également présentées dans le but d'évaluer empiriquement la puissance du test proposé.

[MS-141]
Adaptive Estimation for Marked Counting Processes
Estimation adaptative pour des processus de comptage marqués

Agathe Guilloux & Stéphane Gaïffas, *LSTA – Université Pierre et Marie Curie*, Fabienne Comte, *MAP5 – Université René Descartes*

We propose a new general method for the nonparametric estimation of the intensity of marked counting processes satisfying the Aalen mulplicative intensity model. A new empirical risk is introduced, which will allow us to develop adaptive estimators by projection or aggregation. General minimax bounds are provided.

Nous présentons une nouvelle méthode générale d'estimation de l'intensité des processus de comptage marqués vérifiant le modèle d'intensité multiplicative d'Aalen. Un nouveau risque empirique est introduit, il permettra de définir des estimateurs adaptatifs par projection ou agrégation. Les bornes minimax seront présentées.

Session 07G Mar. 27 mai • Tues. May 27, 10:30 – 12:00 Capital Hall 7A

Classification Methods and Data Mining
Méthodes de classification et forage de données

[MS-142]
Learning Sparse Finite Mixture of Regression Models in High Dimensional Feature Space
Mélange fini parcimonieux de modèles de régression pour l'apprentissage dans un espace de caractéristiques de grande dimension

Abbas Khalili & Jiahua Chen, *University of British Columbia*, Shili Lin, *Ohio State University*

Modern technology has led to the generation of high dimensional data in many areas of scientific research. In applications, most often the dimensionality of the feature vector $x = (x_1, x_2, \ldots, x_p)$, exceeds the number of observations (x_i, Y_i)'s, i.e., $n < p$. In such a situation, we can hope to identify a small number of the features x_j's that have the most effect on the variation of Y. In this talk, we discuss the problem of feature selection in high dimensional data when a sparse mixture of regression models effectively describes the relationship between Y and x.

La technologie moderne a donné lieu à la production de données de grande dimension dans de nombreux domaines de recherche scientifique. Dans les applications, le plus souvent, la dimensionnalité du vecteur de caractéristiques $x = (x_1, x_2, \ldots, x_p)$ est supérieure au nombre d'observations (x_i, Y_i), c.-à-d. $n < p$. Dans une telle situation, nous pouvons espérer reconnaître un petit nombre des caractéristiques x_j qui ont le plus d'effet sur la variation de Y. Dans la communication, nous discuterons du problème de la sélection des caractéristiques dans le cas des données de grande dimension, quand un mélange parcimonieux de modèles de régression décrit effectivement la relation entre Y et x.

[MS-143]
Defining a Cluster Tree for a Multivariate Discrete Distribution
Définir un arbre de classification pour une distribution discrète multivariée

Wu Zhou & Wayne Oldford, *University of Waterloo*

Cluster trees are one means of identifying the goal of clustering for data from continuous distributions. Extending this concept to multivariate discrete distributions is an open problem and presents interesting conceptual and computational challenges.

Les arbres de classification sont une méthode pour identifier le but de la classification pour des données de distributions continues. L'élargissement de ce concept à des distributions multivariées discrètes est un problème ouvert et comporte des défis tant conceptuels qu'au niveau des calculs.

We describe these and present an extension for 2d discrete distributions based on a simple spring model of an elastic sheet. The spring model algorithm constructs undirected graphs in which each edge represents the continuity between two points in the distribution at a certain level.

Defining a discrete distribution cluster tree in higher dimensional space will be discussed as well.

Nous décrivons ceux-ci et présentons un élargissement à des distributions discrètes à deux dimensions basé sur un modèle simple d'un ressort sur une feuille de plastique. L'algorithme du modèle du ressort construit des graphes non orientés dans lesquels chaque arête représente la continuité entre deux points dans la distribution à un certain niveau.

Nous discutons également de la définition d'un arbre de classification pour un distribution discrète dans un espace à plus haute dimension.

[MS-144]
Kernel Methods and Nearest Neighbor Classification
Méthodes à noyaux et plus proches voisins

Arnaud Guyader, *Université de Rennes 2*, Gérard Biau, *Université de Paris 6* & Frédéric Cérou, *IRISA/INRIA, Rennes*

Let S be a set of objects, $(X_1, \ldots, X_n)$ in S whose labels $(Y_1, \ldots, Y_n)$ have values in $\{0, 1\}$. A positive definite kernel $K(X, X')$ is given and allows us to measure the similarity between two objects in S. Thanks to the kernel trick, one can then define a distance d based on S. If a new object X has to be classified, we propose to apply the nearest neighbor method, which consists of looking at the nearest neighbors of X for the metric d and choosing 0 or 1 for its label according to the majority vote.

Considérons un ensemble S et des éléments $(X_1, \ldots, X_n)$ auxquels sont associés les étiquettes $(Y_1, \ldots, Y_n)$ à valeurs dans $\{0, 1\}$. On dispose de plus d'un noyau défini positif $K(X, X')$ permettant de mesurer la similarité entre deux objets de S. Grâce à l'astuce du noyau, on peut ainsi construire une distance d sur S à partir de K. Supposons alors qu'on ait à classer un nouvel objet X dans l'une des deux classes 0 ou 1, on peut maintenant appliquer la méthode des plus proches voisins au moyen de la métrique d.

[MS-145]
Color Image Segmentation by 2*D* Histogram Classification and Information Criteria
Segmentation d'une image couleur par la classification des histogrammes 2*D* et les critères d'information

Abdelaziz El Matouat, *Université du Havre*, Hassania Hamzaoui, *Université de Meknès, Maroc*, Lahcen Kassri, *Université de Fès, Maroc*, Patrick Martin, *LaMIPS, NXP, Ensicaen*

In this paper we present a color image segmentation which allows us to develop image clusters by an unsupervised classification of pixels in the RGB color space. We consider a classification by the information criterion Phi_Beta of the RG and GB histograms and the G histogram separately, and we deduce a global classification in the RGB space. We also propose a quantification function which evaluates the segmentation quality. Finally, the efficiency of the proposed method is clearly shown on the tested images "house," "peppers" and "monarch."
Keywords: Histogram, Fuzzy set, Phi_Beta criterion, Fuzzy segmentation algorithm.

Dans cet article, nous présentons une méthode de segmentation d'images couleur permettant la construction des régions de l'image par une classification non supervisée des pixels dans l'espace couleur RVB. Nous considérons une classification par le critère d'information Phi_Beta des histogrammes RV et VB et l'histogramme V séparément, et nous déduisons une classification globale dans l'espace RVB. Nous proposons également une fonction de quantification d'une segmentation qui mesure la qualité de la segmentation. L'efficacité de cette technique est finalement mise en évidence sur les images naturelles « House », « Peppers » et « Monarch ».
Mots clés : histogramme, sous-ensemble flou, critère d'information Phi_Beta, algorithme de segmentation flou.

[MS-146]
Multivariate Mixed Decision Trees
Arbres de décisions mixtes multivariés

Abdessamad Dine, Denis Larocque & François Bellavance, *HÉC Montréal*

We propose a tree-based method for multivariate outcomes consisting in a mixture of categorical and continuous responses. The split function for tree-growing is derived from a likelihood based approach for a general location model (GLOM). One situation where the new approach should be appealing is when mixed types multiple outcomes are used as surrogates for an unobserved latent outcome. An illustration of the application of the new method is given with health care data extracted from an administrative database.

Nous proposons une méthode basée sur des arbres pour des résultats multivariés consistant en un mélange de réponses catégorielles et continues. La fonction de division pour l'accroissement des arbres est obtenue à partir d'une approche basée sur la vraisemblance pour un modèle de position général (MPOG). Une situation où la nouvelle approche devrait être attrayante est lorsque des résultats multiples de types mixtes sont utilisés comme substituts pour une réponse latente non observée. Une illustration de l'application de la nouvelle méthode utilisant des données de soins de santé provenant d'une banque de données administrative est donnée.

Session 07H Mar. 27 mai • Tues. May 27, 10:30 – 12:00 Capital Hall 8A

Financial Engineering
Ingénierie financière

[MS-147]
Law of Large Numbers for Dynamic Bargaining Markets
Loi des grands nombres pour des marchés dynamiques avec négociation

René Ferland, *Université du Québec à Montréal*, Gaston Giroux

We describe the random meeting motion of a finite number of investors in dynamic bargaining markets as a Markov pure jump process with interactions. Using a sequence of these we prove a functional law of large numbers relating the large motions with the market of the so-called continuum of agents.

Nous décrivons les rencontres aléatoires d'investisseurs dans un marché dynamique avec négociation par un processus interactif markovien de sauts purs. Nous démontrons une loi des grands nombres fonctionnelle qui fait le lien entre ces processus et un modèle de marché comportant un continuum d'agents.

[MS-148]
Mean-variance Efficiency in a Market with Zero-coupons
Stratégie moyenne-variance optimale dans un marché avec obligations à escompte

François Watier & René Ferland, *Université du Québec à Montréal*

We study a mean-variance optimal investment problem in a continuous-time framework where the interest rates follow Cox – Ingersoll – Ross dynamics. We construct an optimal portfolio through the solutions of backward stochastic differential equations. We also give sufficient conditions under which an explicit analytical expression is available for the optimal wealth of the investor.

Nous considérons un problème de gestion de portefeuilles en temps continu où les taux d'intérêt se comportent selon un modèle de type Cox-Ingersoll-Ross. Nous élaborons une stratégie optimale au moyen de solutions d'équations différentielles stochastiques rétrogrades. Enfin, nous fournissons des critères permettant d'obtenir une expression explicite du capital accumulé de l'investisseur.

[MS-149]
Regime Switching Models with Lévy Process
Les modèles à régime variable sous des processus de Lévy

Romuald Momeya, *Université de Montréal*

Due to the insufficiencies of the classic Black – Scholes model, notably the hypotheses of the normality of the log-returns and the continuity of the trajectories of the underlying data, the literature in mathematical finance grew rich these last twenty years from an outfit of models which attempt, at many points, to generalize the approach of Black – Scholes. The regime switching models constitute one of these attempts which ally the presence of the jumps in the underlying's dynamics, while at the same time allowing the possibility of a stochastic volatility governed by a hidden Markov chain. We attempt in our presentation to show the potentialities of this kind of modeling, notably in the field of the Derivative Pricing theory.

Face aux lacunes du modèle classique de Black-Scholes, notamment les hypothèses de normalité des log-rendements et de la continuité des trajectoires des données sous-jacentes, la littérature de la mathématique financière a connu ces vingt dernières années une criossance des modèles qui tentent de généraliser l'approche de Black-Scholes. Les modèles à régime variable constituent une de ces tentatives qui allient à la fois la présence des sauts dans les processus sous-jacents alors qu'ils permettent la possibilité d'une volatilité stochastique gouvernée par une chaîne de Markov. Nous tenterons ici de faire voir les potentialités de ce genre de modélisation notamment dans le domaine de l'évaluation des options.

[MS-150]
Calibrating Multivariate First-Passage Models for Credit Risk
Calibrer des modèles de premier passage multivariés pour le risque de crédit

Adam Metzler & Don L. McLeish, *University of Waterloo*

In this talk we investigate a general framework for multivariate first passage models in credit risk. We allow for both stochastic volatility in firm value and a stochastic default threshold. This results in a setting where defaults occur upon first passage of time-changed Brownian motions to stochastic barriers. We discuss two models from this framework in detail, both of which calibrate successfully to market quotes for synthetic CDO tranches.

Dans cet exposé nous étudions une cadre général pour les modèles multivariés de premier passage dans le risque de crédit. Nous permettons une volatilité stochastique dans la valeur de la firme et un seuil de défaut stochastique. Ceci résulte en un contexte où le défaut de crédit peut survenir au premier passage par des mouvements browniens changeant avec le temps de barrières stochastiques. Nous discutons en détail de deux modèles pour ce cadre, tous deux ayant été calibrés avec succès au moyen de cours du marché de tranches d'obligations synthétiques adossées à des actifs.

Session 07I Mar. 27 mai • Tues. May 27, 10:30 – 12:00 Capital Hall 9A

Sensitivity Analysis and Fiability in Business and Industrial Statistics
Analyse de sensibilité et fiabilité en statistique industrielle et de gestion

[MS-151]

Use of Gaussian Process for Sensitivity Analysis, Application to a Spatial Output of Computer Code
Utilisation des processus Gaussiens pour l'analyse de sensibilité : mise en œuvre sur une sortie spatiale d'un code de calcul

Amandine Marrel, *CEA Cadarache*, Béatrice Laurent, *Institut de Mathématiques de Toulouse, INSA, Université de Toulouse*, Bertrand Iooss, *CEA Cadarache*

In the context of the pollutant transfer numerical modelling, the required computer codes are often computer time expensive. They depend on a high number of input parameters and compute, in output, spatial maps of concentration. To measure the input influence on the response variability, variance-based sensitivity indices are estimated with a limited number of code simulations. The proposed methodology consists of decomposing each simulated map over an orthogonal wavelet basis. Then, the main coefficients of the projection are modelled by a Gaussian process metamodel. This functional metamodel of the inputs is finally used to estimate sensitivity maps.

Dans le cadre de la simulation numérique du transfert de polluants, les codes de calcul utilisés sont souvent coûteux, dépendent d'un grand nombre de paramètres et fournissent, en sortie, des cartes spatiales de concentration. Pour déterminer l'influence des variables d'entrée incertaines sur la sortie du code, on souhaite estimer, à partir d'un nombre limité de simulations du code, des indices de sensibilité (indices de Sobol). La méthodologie proposée consiste à projeter chaque carte simulée sur une base d'ondelettes puis à modéliser les principaux coefficients de la projection par un modèle de processus gaussien, fonction des variables d'entrée. Ce métamodèle fonctionnel est ensuite utilisé pour estimer les cartes de sensibilité.

[MS-152]

An Heuristic Approximation of the Sampling Distribution of the Coefficient of Variation
Une approximation heuristique de la loi d'échantillonnage du coefficient de variation

Gérard d'Aubigny, *Département de Statistique, MS^3-LJK, Université de Grenoble,*, Lambert Pierrat, *LJ-Consulting, Saint-Martin d'Hères*

The applicability of the stress-strength method to mechanical dimensioning necessitates some knowledge about the statistical distribution of two characteristics. Product fiability results from their sensitivity to variability, measured in practice by the coefficient of variation. The sampling distribution of its estimator is quite cumbersome. Thus, in the normal case, one should use the non-central Student distribution. We present a simple analytical approximation, which tries to avoid such complexities in the small sample context. Our proposal is based on the derivation of Gaussian quantiles, and seems accurate enough for practical purposes in the range of values generally met in practice.

L'application de la méthode résistance-contrainte au dimensionnement mécanique implique de connaître deux distributions statistiques. La fiabilité des produits dépend de leur sensibilité à toute variabilité, mesurée en pratique par un coefficient de variation. La loi d'échantillonnage de son estimateur est assez complexe. Ainsi, dans le cas gaussien, on doit utiliser la distribution de Student décentrée. Nous présentons une approximation analytique simple, destinée à contourner cette difficulté et valable pour des petits échantillons. Notre proposition utilise le calcul de quantiles gaussiens et semble suffisamment précis pour les besoins concrets, dans le domaine de valeurs généralement rencontrées en pratique.

[MS-153]
Multivariate SPC Methods to Monitor Complex Product-Process Parameters
Méthodes multivariées pour la maîtrise de processus de fabrication complexes

Luan Jaupi, *CNAM-Paris*

Multivariate quality control problems involve the evaluation of a process based on the simultaneous behavior of quality characteristics and process parameters. Generally, not all of these quality characteristics are equally important. Multivariate methods for monitoring processes with multivariate measurements in both product quality characteristics space and process parameters space are considered. Our approach to build up such control charts consists of monitoring the stable level of variability of the process according to the directions determined by the eigenvectors of a generalized PCA based on the leading variables.

Généralement dans un processus de fabrication complexe les valeurs de plusieurs caractéristiques de qualité et de paramètres « process » sont relevées. Lorsque le nombre de caractéristiques de qualité est élevé, généralement elles ne sont pas de même importance. Certaines sont essentielles pour décrire la qualité des unités produites et d'autres moins importantes. La communication portera sur la construction de cartes de contrôle multidimensionnel pour la maîtrise et le suivi de processus/produit complexes. Notre approche pour construire de telles cartes de contrôle consiste à suivre le niveau stable de la variabilité du procédé selon les directions déterminées par les vecteurs propres d'une ACP généralisée basée sur les caractéristiques de qualité les plus importantes.

[MS-154]
Optimization of the Total Test Time for Global Reliability Demonstration Design in a Competing Failure Modes Model.
Allocation optimale du temps de tests pour une démonstration de fiabilité globale dans un cadre de modes multiples de défaillance

Vincent Couallier & Léo Gerville-Réache, *Institut Mathématique de Bordeaux, Université Bordeaux 2*

Let X be the random lifetime of a system submitted to p failure modes. Then X is the minimum of $(U_1, \ldots, U_p)$ where U_i is the lifetime corresponding to the ith failure mode. This competing failure modes model assumes that $(U_1, \ldots, U_p)$ are independent. The main issue is to find the optimal test time in an accelerated life model for each failure mode which minimizes, for instance, the total test time. The constraint is to design this multi-tests procedure in order to achieve the overall reliability demonstration. A Bayesian framework will also be considered.

Soit X une variable aléatoire modélisant la durée de vie d'un système. Si plusieurs modes de défaillance concurrents sont possibles, X est définie comme le minimum de $(U_1, \ldots, U_p)$ où U_i est la durée de vie liée au i^{e} mode de défaillance. Dans ce modèle à risque compétitif, les variables $(U_1, \ldots, U_p)$ sont supposées indépendantes. L'objectif est d'allouer des valeurs pour les temps de test propre à chaque mode de défaillance en minimisant par exemple le temps total de test pour une démonstration de fiabilité globale à partir de modèles de vies accélérées. Un modèle bayesien pour les paramètres inconnus est envisageable.

[MS-155]
From ABC Model to DEF Model
Modèle DEF : une modélisation statistique complémentaire du modèle ABC

Eduard Ibanez Tarradellas, *TNS WORLDPANEL*

Taking the main ideas of ABC model applied like a cost-method in managerial accounting fields, we develop a new model named DEF model (Decision-Effective Factors) that uses statistical issues in order to create a useful decision-method when we are

Le modèle ABC entendu comme méthode de coût dans l'environnement comptable peut être le point de départ d'une nouvelle modélisation statistique appelée modèle DEF (Decision-Effective Factors) laquelle utilisent des problèmes statistiques afin de créer une méthode de décision effcicae

comparing the "census" world with the "statistical" world. This topic is a key point when we use statistical data to understand markets, and we validate correction of this data by making comparisons with internal data from these markets provided in different ways. It is a very difficult thing to ensure equality between both sources of information. DEF model is developed from few proposals but with a lot of implications and powerful consequences in the environment of managerial fields when we use statistical data to know markets.

lorsqu'il faut comparer des données réelles avec des données statistiques. Ce sujet est un aspect essentiel et peut être très utile quand on veut comprendre l'évolution des marchés. Nous validons les corrections de ces données en faisait des comparaisons entre les données internes de ces marchés qui sont fournies de différentes façons. Il est très difficile d'assurer la qualité entre les deux sources d'information. Le modèle DEF est développé à partir de quelques propositions, mais qui ont beaucoup d'implications et de conséquences dans l'environnement de gestion patronale lorsque des données statistiques sont utilisées pour connaître les marchés.

Session 08A Mar. 27 mai • Tues. May 27, 13:30 – 14:30 Congress Hall A/E

SSC Gold Medal Award Address
Allocution du récipiendaire de la médaille d'or de la SSC

[MS-156]
From Martingales to Monte Carlo: A Personal Random Walk
Des martingales à Monte-Carlo : une marche aléatoire personnelle

Don McLeish, *University of Waterloo*

Arguably, our discipline has one foot firmly planted on a continent of elegant mathematical techniques and the second foot equally solidly on another continent, populated by massive amounts of data and immense computational power. This can be a wholly uncomfortable position in the presence of continental drift. Here, I provide examples of this sometimes tenuous and occasionally uncomfortable, but often fruitful relationship between theoretical development and computational advances. One example is the use of optimal importance sampling to simulate rare events and diffusion processes in finance or biostatistics, both fields where rapid development is fuelled partly by Monte Carlo Methods.

On pourrait soutenir que notre discipline a un pied fermement planté sur un continent de méthodes mathématiques élégantes et le second tout aussi solidement planté sur un autre continent, peuplé de quantités énormes de données et d'une puissance informatique immense. En cas de dérive des continents, cette position peut être tout à fait inconfortable. Ici, je donne des exemples de cette relation parfois ténue et à l'occasion inconfortable, mais souvent fructueuse entre le développement théorique et les progrès informatiques. L'un de ces exemples est le recours à un échantillonnage préférentiel optimal pour simuler des événements et des processus de diffusion rares en finance ou en biostatistique, deux domaines dont le développement rapide est sous-tendu en partie par les méthodes de Monte-Carlo.

Session 09A Mar. 27 mai • Tues. May 27, 14:35 – 15:35 Capital Hall A1

Special Invited Session of the Survey Methods Section
Allocution de l'invité d'honneur du groupe de méthodologie d'enquête

[MS-157]

How to Analyze Survey Data Badly: Answers to Criticisms of Using Design-Based Methods for Inference
Comment mal analyser des données d'enquête : réponse aux critiques relatives à l'utilisation des méthodes fondées sur le plan de sondage pour l'inférence

David Binder, *Statistics Canada (Retired)*

In recent years, there has been much discussion on whether or not it is appropriate to incorporate the survey weights and other design information when making inferences about model parameters from survey data collected under a complex design. The main schools of thought are either to use a design-based approach, or to use a model-based approach. In this paper, we discuss some general theory under a model-design-based randomization framework, and we address some of the criticisms of the design-based approach within this framework. We see some examples where a model-based approach can give better results than a pure design-based approach.

Ces dernières années, on a beaucoup discuté de la question de savoir s'il est approprié ou non d'intégrer les poids de sondage et d'autres informations sur le plan quand on fait des inférences au sujet des paramètres d'un modèle à partir de données d'enquête recueillies selon un plan de sondage complexe. Les principales écoles de pensée sont celles qui défendent l'approche fondée sur le plan de sondage ou l'approche fondée sur un modèle. Dans la présente communication, nous discutons d'une théorie générale dans un cadre de randomisation sous un modèle et sous le plan, et nous examinons dans ce cadre certaines critiques de l'approche fondée sur le plan. Nous découvrons certains exemples où une approche fondée sur un modèle peut donner de meilleurs résultats qu'une approche fondée purement sur le plan.

Session 09B Mar. 27 mai • Tues. May 27, 14:35 – 16:05 Capital Hall 2A

Statistical Process Control Methods in Health Care
Méthodes statistiques de contrôle des processus en santé

[MS-158]

Performance Monitoring in the Public Services
Contrôle du rendement dans les services publics

Vern Farewell, *Medical Research Council Biostatistics Unit, UK*

In many countries efforts are being made to improve the quality of public services through performance management. Some background will be given to these efforts. A discussion will be given of some specific methodology that has been devised, with a particular focus on longitudinal monitoring of performance in health related contexts. The adaptation of methods used in other contexts to incorporate risk adjustment will be emphasized. Issues likely to be relevant to future efforts to improve performance in the public services will be highlighted.

De nombreux pays s'efforcent d'améliorer la qualité des services publics en recourant à la gestion du rendement. Nous donnerons le contexte général de ces efforts. Puis nous discuterons de méthodes particulières qui ont été élaborées, en nous concentrant surtout sur la surveillance longitudinale du rendement dans des contextes reliés à la santé. L'accent sera mis sur l'adaptation de méthodes utilisées dans d'autres contextes en vue d'intégrer l'ajustement du risque. Nous mettrons en relief les questions qui se poseront vraisemblablement dans le cadre de futurs efforts en vue d'améliorer le rendement dans les services publics.

[MS-159]
Statistical Process Control Methods in Health Care
Méthodes de contrôle du processus statistique dans les soins de santé

Charlie Goldsmith, *McMaster University*

In healthcare, the most common method of process control consists of control charts. A summary of control charts for use with various data types will be provided. A survey of health science journals will be used to show how control charts are used to improve health care quality. Biostatisticians should work more closely with health professionals to encourage the processes in health care to be monitored with control charts to detect when the process is out of control, and when it needs improvement. Examples of patient, clinical and hospital data will be used to illustrate the points in this talk.

Dans le domaine des soins de santé, la méthode la plus fréquente de contrôle des processus consiste à utiliser des cartes de contrôle. Nous présenterons un sommaire des cartes de contrôle qu'il convient d'utiliser pour divers types de données. Nous nous servirons d'une enquête sur les revues spécialisées en sciences de la santé pour illustrer comment les cartes de contrôle sont utilisées pour améliorer la qualité des soins. Les biostatisticiens devraient travailler en collaboration plus étroite avec les professionnels de la santé en vue de promouvoir la surveillance des processus dans le domaine des soins de santé à l'aide de cartes de contrôle afin de détecter quand le processus est hors de contrôle et quand il doit être amélioré. Dans la communication, nous nous servons d'exemples de données sur les patients, de données cliniques et de données hospitalières pour illustrer les points susmentionnés.

[MS-160]
Research Issues and Ideas in Health-Related Surveillance
Questions de recherche et idées relatives à la surveillance en matière de santé

William Woodall, *Virginia Tech*, Olivia Grigg, *Cambridge University*, Howard Burkom, *Johns Hopkins University*

In this overview paper, some of the surveillance methods and metrics used in health-related applications are described and contrasted with those used in industrial practice. Many of the health-related surveillance methods are based on the concepts and methods of statistical process control. Public health data often include spatial information as well as temporal information, and in this and other regards, public health applications could be considered more challenging than industrial applications. Avenues of research into various topics in health-related monitoring are suggested.

Key words: Recurrence interval; Risk adjustment; Spatiotemporal surveillance; Statistical process control; Time-to-signal.

Cette étude d'ensemble a pour but de décrire certaines méthodes de surveillance et mesures utilisées dans les applications dans le domaine de la santé et de les comparer à celles utilisées dans la pratique industrielle. Nombre de méthodes de surveillance en matière de santé sont fondées sur les concepts et les méthodes du contrôle statistique des processus. Les données sur la santé publique comprennent souvent des renseignements spatiaux ainsi que temporels, si bien qu'à cet égard et à d'autres, les applications au domaine de la santé publique pourraient être considérées comme posant plus de difficultés que les applications industrielles. Nous proposons des pistes de recherche dans divers domaines de la surveillance en matière de santé.

Session 09C Mar. 27 mai • Tues. May 27, 14:35 – 16:05 Capital Hall 3A

Statistical Analysis of Hydroclimatic Proxy Data: Frequentist Approaches
Analyse statistique des empreintes hydro-climatiques : approches fréquentielles

[MS-161]

Climate Reconstruction from Pollens: Climate Impact on the Spatial Beech Genetic Diversity
Reconstitution climatique à partir de pollens : étude de l'impact du climat sur la répartition spatiale de la diversité génétique des hêtres

Liliane Bel, *AgroParisTech,* Avner Bar-Hen, *Université Paris 5,* Rachid Cheddadi, *Institut des Sciences de l'Evolution*

The aim of this presentation is modeling the relationship between genetic diversity in the European beech forests (represented by a positive number) and curves of temperature and precipitation reconstructed from pollen in the European Pollen Database. Our model links the genetic measure to the climate curves through a spatio-temporal functional regression. We show that diversity is higher when the change in climate during the Holocene was more important. Statistical properties of the model are derived.

Dans ce travail nous quantifions l'influence des variations climatiques, reconstitués à partir des pollens de l'European Pollen Database des 15 000 dernières années, sur la diversité génétique actuelle du hêtre. À l'aide d'un modèle spatio-temporel fonctionnel multivarié, nous montrons que la diversité est plus riche dans les régions où le réchauffement observé lors de l'holocène a été plus marqué. Enfin nous étudions les propriétés statistiques du modèle proposé.

[MS-162]

Reconstructing Past Volcanic Forcings with a Multivariate Multi-State Space Model
Reconstruire le forçage volcanique à partir d'un modèle d'états multivarié

Philippe Naveau, *LSCE-CNRS-IPSL,* Caspar Ammann, *NCAR, Colorado,* Deborah Batista, *Colorado University, Denver,* Julie Carreau, *LSCE-CNRS-IPSL, France*

To understand the cause of climate variability, it is important to attribute past climate variations to particular forcing factors. Our main focus is to introduce an automatic procedure to estimate the magnitude of strong but short-lived perturbations. Our extraction algorithm handles multivariate time series with a common but unknown forcing. This statistical procedure is based on a multivariate multi-state space model and it can provide an accurate estimator of the timing and duration of the deposition of volcanic sulfate. The flexibility, robustness and limitations of our approach are discussed by applying our method to simulated and real multivariate time series.

Pour comprendre la cause de la variabilité du climat, il est important d'attribuer les variations climatiques passées à des facteurs de forçage particuliers. Notre principal objectif est d'introduire une procédure automatique pour estimer l'ordre de grandeur de perturbations fortes, mais de courte durée. Notre algorithme d'extraction traite les séries chronologiques multivariées ayant un forçage commun, mais inconnu. Cette procédure, qui est fondée sur un modèle multi-états multivarié, peut fournir un estimateur exact du moment et de la durée des dépôts de sulfates volcaniques. Nous discutons de la souplesse, de la robustesse et des limites de notre approche en appliquant la méthode à des séries chronologiques multivariées simulées et réelles.

[MS-163]
Evaluation of Proxy-Based Millennial Reconstruction Methods
Évaluation de méthodes de reconstruction millénaire fondées sur des données indirectes

Francis Zwiers, *Environment Canada, Climate Research Division,* Min Tsao & Terry Lee, *University of Victoria*

A range of statistical approaches for reconstructing historical temperature variations from proxy data are compared using both climate model data and real world paleo-climate proxy data. A new method for reconstruction based on a state-space time series model and Kalman filter algorithm is also proposed. The state space modelling approach and the recently developed RegEM method generally perform better than their competitors on inter-annual time scales. The state space modelling approach can incorporate additional, non-temperature, information, such as the estimated response to greenhouse gas forcing, thereby permitting simultaneous reconstruction and detection analysis as well as future projection.

Nous comparons une gamme d'approches statistiques de la reconstruction des variations historiques de température d'après des données indirectes en utilisant des données de modèles climatiques, ainsi que des données indirectes paléoclimatiques réelles. Nous proposons une nouvelle méthode de reconstruction fondée sur un modèle chronologique d'espace d'états et l'algorithme du filtre de Kalman. L'approche de modélisation de l'espace d'états et la méthode RegEM développée récemment donnent généralement de meilleurs résultats que leurs concurrentes sur les échelles temporelles interannuelles. L'approche de la modélisation de l'espace d'états peut introduire de l'information supplémentaire, n'ayant pas trait à la température, telle que la réponse estimée au forçage par les gaz à effet de serre, ce qui permet d'effectuer simultanément la reconstruction et l'analyse de détection, ainsi qu'une projection future.

Session 09D Mar. 27 mai • Tues. May 27, 14:35 – 16:05 Capital Hall 4A

How to Reconcile the Demands for Data Confidentiality and Research Needs? The Role of Secure Access Centers.

Comment concilier les exigences de la confidentialité des données et les besoins de la recherche ? Le rôle des centres d'accès sécurisés.

[MS-164]
Processing Confidential Data and the Methods of Protection: The researchers perspective
Le traitement des données confidentielles et les dispositifs de protection : le point de vue des chercheurs.

Roxane Silberman, *Directrice de recherche au CNRS/Secrétaire générale du Comité de concertation pour les données en sciences humaines et sociales*

Secure centres with residential or remote access allow researchers to process personal and confidential data. These systems bring into question certain aspects of the difficult evolution towards free access to data that began in the 1960s. This evolution was made a reality by Data Archives, and, at the same time, blurred the boundary between public statisticians and researchers. Researchers are still not in agreement about these systems in terms of costs, methods of working, procedures for access, and control measures. However, they have begun, more and more, to recognize the necessity of implementing safeguards.

Les centres sécurisés en accès résidentiel ou à distance permettent le traitement par les chercheurs de données individuelles confidentielles. Ces dispositifs remettent en question certains aspects de l'évolution difficile entamée depuis les années 60 vers un accès libre aux données, concrétisée par les Data Archives en même temps qu'ils affaiblissent la frontière entre statisticiens publics et chercheurs. Les chercheurs sont encore loin d'être unanimes sur ces dispositifs au regard des coûts, des façons de travailler, des procédures d'accès, des instances en charge du contrôle. Mais ils reconnaissent de plus en plus la nécessité d'organiser des garanties.

[MS-165]
The RDC Network: From a Seed to a Forest
Le réseau des CDR : d'une graine à la forêt

Gustave Goldmann, *Statistique Canada*

The impact of the network of research data centres (RDC) on quantitative research on Canadian Society has grown remarkably since 2000. The contribution of the RDCs can be measured by: the number of young scholars (the next generation of researchers) that are trained, the number of research products that are published, the advancement of analytical methods, etc. Regardless of what measure is used, the contribution of the RDC Network to the quantitative social sciences is obvious. This presentation will provide an overview of the success of the Network and the challenges it faces in the coming years.

Depuis 2000, le réseau des centres de données de recherche (CDR) a connu une croissance remarquable de son impact dans le domaine de la recherche quantitative sur la société canadienne. La contribution des CDR peut se mesurer de plusieurs façons : par la formation de la prochaine génération de chercheurs quantitatifs ; par le nombre de produits de recherche publiés ; par l'avancement des méthodes d'analyse, etc. Quelle que soit la dimension, la contribution du réseau des CDR aux sciences humaines quantitatives au Canada est évidente. Cette communication fera le survol des succès du réseau et des défis qui l'attendent dans les prochaines années.

[MS-166]
Insee's Secure Access Centers : A Way to the Future
Les centres d'accès sécurisés à l'Insee : projets en cours

Michel Isnard, *INSEE*

Allowing a broader use of statistical data by researchers has always been a strong guideline of the Insee, the French national institute of statistics. After the access, through "Quetelet network", of anonymised individual files (fichiers de production et de recherche), Insee is working, both on the dispositions of the statistical law and on practical issues, on the installation of secure access centre(s), which should allow researchers to access not directly identifiying individual files. The presentation will describe the present situation, the pilot study and ways for the future.

Permettre une plus large utilisation des données statistiques par les chercheurs a toujours été une ligne directrice forte de l'Insee. Après la mise à disposition, au moyen du réseau Quetelet, de fichiers individuels anonymisés (fichiers de production et de recherche), l'Insee travaille, tant sur le plan juridique que sur le plan pratique, à la mise en place de centre(s) d'accès sécurisé qui devrai(en)t permettre aux chercheurs d'avoir accès à des fichiers individuels indirectement nominatifs. La présentation décrira la situation actuelle, le projet pilote et les différentes voies en cours d'exploration.

Session 09E Mar. 27 mai • Tues. May 27, 14:35 – 16:05 Capital Hall 5A

Classification and Algorithms
Classification et algorithmes

[MS-167]
Hierarchical Methods for Multidimensional Data: Convergence, Complexity and Learning
Méthodes hiérarchiques pour la classification multidimensionnelle : convergence, complexité, apprentissage

Luc Devroye, *McGill University*

Breiman, Friedman, Gordon and Stone recognized that tree classifiers would be very valuable to prac-

Quelques idées ont profondément influencées les algorithmes de classification : la méthode du plus proche voisin,

ticing statisticians. Their CART algorithm became very popular indeed. Designing tree-based classifiers, however, has its pitfalls. It is easy to make them too simple or too complicated so that Bayes risk consistency is compromised. In this talk, we explore the relationship between algorithmic complexity of tree-based methods and performance.

la discrimination linéaire, la minimisation du risque empirique, et les méthodes hiérarchiques. On peut aussi combiner un ou plusieurs classificateurs simples pour se construire un autre. Par exemple, on sait que la méthode du plus proche voisin n'est pas optimale, mais si on échantillonne un petit ($o(n)$, $\omega(n)$) sous-ensemble des données et si on prend la majorité des "votes" des plus proches voisins, un vote par échantillon, le classificateur devient convergent au sens de Bayes. On discutera de ce principe pour les classificateurs basés sur des partitions aléatoires de l'espace à partir d'arborescences.

[MS-168]
Instance-Based Learning with Proximity Graphs
Apprentissage basé sur les occurences avec graphes de proximité

Godfried Toussaint, *McGill University*, Binay Bhattacharya, *Simon Fraser University*

One of the most attractive non-parametric decision rules is the K-NN rule, in which a pattern is classified into the majority class among its K nearest neighbors in the training set. Several questions arise: How large should K be? Should all K neighbors be weighted equally? Can the storage of the training set be reduced without degrading the performance of the decision rule? Gabriel graphs offer elegant solutions to these problems, but pose the problem of computing them efficiently in high dimensions. In this talk we discuss recent work on the application of Gabriel graphs in this area.

Une des règles de décision non paramétrique les plus attrayantes est la règle K-NN, où une structure est classée avec ses K plus proches voisins dans l'ensemble d'apprentissage. Plusieurs questions surviennent : de quelle taille K devrait-il être ? Est-ce que tous les K voisins devraient être pondérés identiquement ? Est-ce que le stockage de l'ensemble d'apprentissage peut être réduit sans dégrader la performance de la règle de décision ? Les graphes de Gabriel apportent une solution élégante à ces problèmes, mais en posent un nouveau : comment les calculer efficacement en dimension élevée ? Dans cet exposé nous discutons de récents travaux sur l'application des graphes de Gabriel dans ce domaine.

[MS-169]
Optimization Principles for the Ranking Problem
Principes d'optimisation pour le problème de l'ordonnancement

Nicolas Vayatis, *ENS de Cachan*, Stéphan Clémençon, *Telecom-ParisTech*

In many applications such as scoring or search engines, the information needs to be filtered and sorted. Hence, relevant instances have to be presented in a ranked list where the more relevant instances would appear at the top. A decision rule with a good classification rate could, at the same time, present poor ranking performance. The approach presented here consists of searching for rules defined as optimizers of specific criteria dedicated to ranking. Statistical learning theory was successfully developed for classification and we will set the statistical grounds for similar inference principles in ranking applications.

Dans certaines applications, comme le scoring ou les moteurs de recherche, l'information doit être filtrée, mais aussi ordonnée. Ainsi, les instances pertinentes doivent se présenter sous forme de liste ordonnée où les instances les plus pertinentes seraient placées en tête. Une règle de décision ayant un bon taux de reconnaissance des instances pertinentes peut très bien présenter des performances désastreuses en termes d'ordonnancement. Le point de vue développé ici consiste à la recherche de règles optimisant directement le critère de performance dédié à l'application d'ordonnancement. On montrera que la théorie de l'apprentissage statistique développée avec succès dans le cadre de la classification nécessite de nombreuses modifications pour l'élaboration de principes inférentiels statistiquement fondés.

Session 09F Mar. 27 mai • Tues. May 27, 14:35 – 16:05 Capital Hall 6A

Robustness
Robustesse

[MS-170]
Robust Likelihood Methods Based on the Skew-*t* and Related Distributions
Méthodes de vraisemblance robustes fondées sur des distributions *t* asymétriques et autres distributions de la même famille

Marc Genton, *Université de Genève and Texas A&M University*, Adelchi Azzalini, *University of Padua*

The robustness problem is tackled by adopting a parametric class of distributions flexible enough to match the behaviour of the observed data. In a variety of practical cases, one reasonable option is to consider distributions which include parameters to regulate their skewness and kurtosis. As a specific representative of this approach, the skew-t distribution is explored in more detail and reasons are given to adopt this option as a sensible general-purpose compromise between robustness and simplicity, both for treatment and for interpretation of the outcome. Some theoretical arguments, outcomes of a few simulation experiments and various wide-ranging examples are provided.

Le problème de la robustesse est attaqué en adoptant une classe paramétrique de distributions qui sont suffisamment flexibles pour représenter le comportement des observations. Dans une variété de cas pratiques, une option raisonnable est de considérer des distributions qui comprennent des paramètres pour régler leur asymétrie et leur aplatissement. Comme représentant spécifique de cette approche, la distribution t asymétrique est explorée plus en détail et des raisons sont apportées pour adopter cette option comme un compromis judicieux et à tous usages entre la robustesse et la simplicité du traitement et de l'interprétation des résultats. Quelques arguments théoriques, les résultats de simulations et divers exemples sur des données réelles sont fournis.

[MS-171]
Robust Efficient Multivariate Trimmed Means
Moyennes tronquées multivariées robustes et efficaces

Jean-Claude Massé, *Université Laval*

Among univariate location statistics, the trimmed mean has long been regarded as a robust and efficient alternative to the sample mean. A multivariate analogue calls for a measure of centrality for the sample points. Depth functions provide a convenient tool for doing that with respect to a multivariate probability distribution. This talk will examine two types of multivariate trimmed means based on the Tukey depth function. A parallel will be drawn between these location estimators and the classical trimmed mean.

En raison de ses propriétés de robustesse et d'efficacité, la moyenne tronquée est souvent préférée à la moyenne échantillonnale comme estimateur de localisation univariée. Toute notion analogue de moyenne tronquée multivariée fait appel à une mesure de centralité des points de l'échantillon. Dans cet exposé, on présentera deux types de moyennes tronquées multivariées obtenues en mesurant la centralité par la fonction profondeur de Tukey. On fera un parallèle entre ces estimateurs de localisation et les moyennes tronquées classiques.

[MS-172]
Robust Quantile Regression with Non-Elliptical Covariates
Régression quantile robuste avec covariables non elliptiques

Matias Salibian-Barrera, *The University of British Columbia*, Ying Wei, *Columbia University*

Although quantile regression estimators are robust against the presence of observations with atypically large responses, they can be seriously affected by a few outlying points in the covariate space. This problem can be alleviated by downweighting observations with high-leverage. Unfortunately, when the covariates are not elliptically distributed, Mahalanobis distances may not correctly identify atypical points. We discuss the use of weighted quantile regression estimators using a leverage measure based on Rosenblatt's multivariate transformation. We show that this measure is able to reflect non-elliptical structures and apply this methodology to growth measurements on Finnish children.

Bien que les estimateurs par la régression quantile soient robustes à la présence d'observations contenant des réponses anormalement grandes, ils peuvent être sérieusement affectés par l'existence de quelques points aberrants dans l'espace des covariables. Ce problème peut être atténué par la sous-pondération des observations dont l'effet de levier est important. Malheureusement, quand les covariables ne sont pas distribuées elliptiquement, les distances de Mahalanobis ne permettent pas toujours de repérer correctement les points atypiques. Nous discutons de l'utilisation des estimateurs par la régression quantile pondérée comportant une mesure de l'effet de levier fondée sur la transformation multivariée de Rosenblatt. Nous montrons que cette mesure peut refléter les structures non elliptiques et nous appliquons la méthodologie à des mesures de la croissance chez les enfants finlandais.

Session 09G Mar. 27 mai • Tues. May 27, 14:35 – 16:05 Capital Hall 7A

Analysis of Dependent Data
Analyse de données dépendantes

[MS-173]
Multiple Tests for High-throughput Data assuming a Factor Modelling of Dependence
Modèle en facteurs pour les procédures de tests multiples en grande dimension

Chloé Friguet, Maéla Kloareg & David Causeur, *Mathématiques Appliquées - Agrocampus Rennes*

Dependence between variables is known to generate instability in multiple testing procedures. Therefore, an increasing interest, particularly in microarray data analysis, is to control actual error rates rather than expected ones. We propose to assume a factor-analysis model for the covariance matrix that allows us to identify common and specific variability of data. Close form expressions for the variance of error rates are derived. A new test statistic is also defined, taking into account the factor structure. It is shown to reduce the variance of error rates, improving power in simultaneous tests.

La corrélation entre variables entraîne des problèmes d'instabilité dans les procédures de tests multiples. Ainsi, en particulier dans l'analyse de données de biopuces, il est important de contrôler les taux d'erreurs réels plutôt que leurs espérances. Nous proposons de considérer un modèle d'analyse en facteur qui suppose la décomposition de la variance en une partie commune et une partie spécifique. Une formulation exacte de la variance des taux d'erreurs en découle. Une nouvelle statistique de test est définie également, tenant compte de la structure en facteurs proposée. Cela permet de réduire la variance des taux d'erreurs, améliorant la puissance des tests simultanés.

[MS-174]

Analysis of Correlated and Clustered Data Using a Survey on Privacy and Health Research in Stigmatized Populations as an Example

Analyse des données corrélées et regroupées en utilisant une enquête sur la vie privée et la recherche sur la santé chez les populations stigmatisées à titre d'exemple

Ji (Emmy) Cheng, *St. Joseph's Healthcare Hamilton*, Lehana Thabane & Donald Willison, *McMaster University*, Co-investigators, *McMaster University, York University, University of Ottawa*

Choosing among possible analytic strategies for dependent outcomes in the presence of clustering is not often simple. We conducted a national survey of the Canadian public on privacy and research use of personal information. The primary aim was to compare responses between respondents with several stigmatizing disease conditions. The data structure includes multiple responses on consent preferences across several scenarios which results in clustering within respondents. In this presentation, we compare different approaches including multivariate analysis, univariate analysis adjusted for multiple comparisons and other methods for clustered data such as generalized estimating equations and fixed-effects model to analyse the data.

Faire un choix parmi les stratégies d'analyse qui peuvent être appliquées à des résultats dépendants en présence de regroupement n'est généralement pas simple. Nous avons réalisé un sondage national sur la vie privée et l'utilisation de renseignements personnels en recherche auprès des membres du public au Canada. Le principal objectif était de comparer les réponses des personnes présentant diverses maladies stigmatisantes. La structure des données comporte des réponses multiples au sujet des préférences de consentement pour plusieurs scénarios qui se traduisent par une mise en grappes des répondants. Dans la communication, nous comparons diverses approches, dont l'analyse multivariée, l'analyse univariée corrigée pour tenir compte des comparaisons multiples et d'autres méthodes applicables aux données regroupées, telles que les équations d'estimation généralisées et les modèles à effets fixes pour analyser les données.

[MS-175]

Testing Multiple Correlated Outcomes in Clinical Trials with Repeated Measurements: Comparing Adjusted p-Value Approaches

Essai des résultats multi-corrélés dans les essais cliniques ayant des mesures répétées : comparaison des approches de la valeur p ajustée

Lisa Lix & Phongsack Manivong, *University of Manitoba*

In many clinical trials, treatment and control groups are compared across repeated measures means on each of several outcomes. It is often desired to control the probability of making a false assertion while ensuring maximum power to detect effects. Monte Carlo simulations were used to evaluate procedures that account for the dependence structure among test statistics when adjusting p-values for multiple testing. Procedures due to Dubey, Westfall – Young, and Roy – Bargman are extended to repeated measures designs. The Roy – Bargman procedure was frequently conservative, while the Westfall-Young procedure could be substantially more powerful depending on the repeated measures means configuration and correlation structure.

Dans de nombreux essais cliniques, les groupes de traitement et de contrôle sont comparés en se fondant sur les moyennes de mesures répétées effectuées pour plusieurs résultats. Il est souvent désirable de tenir compte de la probabilité de faire une affirmation fausse tout en veillant à ce que la puissance de détection des effets soit maximale. Nous avons utilisé des simulations de Monte Carlo pour évaluer des méthodes qui tiennent compte de la structure de dépendance entre les statistiques de test lorsque l'on ajuste les valeurs p pour des tests multiples. Nous étendons les procédures élaborées par Dubey, Westfall-Young et Roy-Bargman à des plans d'expérience avec mesures répétées. La méthode de Roy-Bargman donne souvent des résultats conservateurs, tandis que celle de Westfall-Young pourrait être considérablement plus puissante selon la configuration et la structure de corrélation des moyennes des mesures répétées.

[MS-176]
Outbreaks in Populations with Contact Network Structure
Épidémie dans les populations ayant une structure en réseau des contacts

Lilia Ramirez Ramirez & Mary Thompson, *University of Waterloo*

Recent epidemic models incorporate the heterogeneous exposure that individuals have to an infective agent spreading in a population. This heterogeneity is introduced with a contact structure that is described in terms of random graphs.

Random graphs or networks have been widely used to model the relationship of several systems, including the World-Wide Web, language and have been observed to model social relationships like the number of sexual partners. In this talk we present results derived for epidemic models based on two level networks (network of sub networks) with special emphasis in the Reed – Frost and measles data.

Les modèles épidémiologiques récents intègrent l'exposition hétérogène des individus à un agent infectieux qui se propage dans une population. Cette hétérogénéité est prise en compte dans le modèle en introduisant dans celui-ci la structure des contacts, qui est décrite à l'aide de graphes aléatoires.

Les graphes aléatoires ou réseaux ont été utilisés à grande échelle pour modéliser la relation entre plusieurs systèmes, y compris les langages, le World-Wide Web, et ont été observés pour modéliser les relations sociales, comme le nombre de partenaires sexuels. Nous présentons dans la communication les résultats découlant de modèles épidémiologiques fondés sur des réseaux à deux niveaux (réseau de sous-réseaux) en insistant particulièrement sur les modèles de Reed-Frost et les données sur la rougeole.

[MS-177]
A Generalized Estimation Equations Approach to Current Status Data with Dependent Responses
Une approche par équations d'estimation généralisées des données sur l'état courant avec réponses dépendantes

David Tolusso & Richard Cook, *University of Waterloo*

Consider a study where subjects are not observed continuously in time, rather they are seen only at one assessment at which it is determined whether or not they have a condition of interest. Such data is known as "current status" data and is an extreme form of interval censoring. Suppose, in addition, that subjects are not independent, but grouped in clusters. We show, using generalized estimating equations, how parameter estimates may be obtained with a dependence induced by a copula. First and second order GEE's are considered. The methods are applied to a study of patients undergoing orthopedic surgery.

Considérons une étude où, au lieu d'être observés continuellement au cours du temps, les sujets ne sont vus qu'à l'occasion d'une seule évaluation durant laquelle on détermine s'ils présentent ou non un état d'intérêt. Ce genre de données, appelées données sur « l'état courant », représentent une forme extrême de censure par intervalle. Supposons en outre que les sujets ne sont pas indépendants, mais groupés en grappes. Nous montrons, en utilisant des équations estimatrices généralisées, comment les estimations des paramètres peuvent être obtenues en induisant une dépendance à l'aide d'une copule. Nous considérons des équations d'estimations généralisées de premier et de deuxième ordre. Nous appliquons les méthodes à une étude portant sur des patients qui subissent une intervention chirurgicale orthopédique.

[MS-178]

An Approach to the Analysis of Spatially Correlated Multilevel Functional Data
Une approche à l'analyse de données fonctionnelles à plusieurs niveaux corrélées spatialement

Ana-Maria Staicu, *University of Bristol*, Ciprian M. Crainiceanu, *Johns Hopkins University*, Raymond J. Carroll, *Texas A&M University*

We describe the framework and inferential tools for multilevel functional data where the functions at the lowest hierarchy level are spatially correlated. Previous literature to analyze such settings includes a Bayesian wavelet-based procedure and a Bayesian semiparametric method, both computationally intensive techniques. We propose to model the functions nonparameterically using multilevel eigenfunction bases, which appear in a multilevel functional principal component scenario, plus a weakly stationary process to account for the spatial correlation. The proposed procedure is illustrated with a simulation study and p27 measurements data over relative cell position for a sample of rodents, in a colon carcinogenesis study.

Nous décrivons le cadre de référence et les outils inférentiels pour des données fonctionnelles à plusieurs niveaux où les fonctions au plus bas niveau hiérarchique sont spatialement corrélées. La littérature analysant de tels contextes comprend une procédure bayésienne à base d'ondelettes et une méthode bayésienne semi-paramétrique, toutes deux lourdes en termes de calculs. Nous proposons de modéliser les fonctions de façon non paramétrique en utilisant des bases de fonctions propres à plusieurs niveaux, que l'on retrouve dans un scénario de composantes principales fonctionnelles à plusieurs niveaux, et d'un processus faiblement stationnaire pour tenir compte de la corrélation spatiale. La procédure proposée est illustrée à l'aide d'une étude de simulation et de mesures p27 sur les positions relatives de cellules d'un échantillon de rongeurs, dans une étude de carcinogenèse du colon.

Session 09H Mar. 27 mai • Tues. May 27, 14:35 – 16:05 Capital Hall 8A

Biostatistical Models
Modèles biostatistiques

[MS-179]

Robust Analysis of Longitudinal Proportional Data Using Mixed-Effects Models
Analyse robuste de données longitudinales proportionnelles à l'aide de modèles à effets mixtes

Peng Zhang, *University of Alberta*, Zhenguo Qiu, *Alberta Cancer Board*, Peter X.-K. Song, *University of Michigan*

Longitudinal proportional data, commonly collected in medical practice, need some special statistical methodology. We extend logistic-normal distribution to logistic-*t* distribution to provide robust logic linear mixed models against outliers for proportional data via utilization of multivariate *t* distributions. Different models are constructed to handle proportional data containing outliers in random effects, error terms or both. Estimation and inference of the proposed models are discussed and illustrated by a real data example. Simulation studies show better robustness against different outliers by logistic-*t* models. Data analysis indicates the proposed robust models obtain consistent conclusions for data with or without outliers.

Les données longitudinales proportionnelles recueillies couramment dans la pratique médicale requièrent une méthodologie statistique spéciale. Nous étendons la loi normale logistique à la loi *t* logistique pour fournir des modèles linéaires mixtes logiques, robustes aux valeurs aberrantes par la voie de l'utilisation de lois *t* multivariées. Nous construisons divers modèles pour traiter les données proportionnelles contenant des valeurs aberrantes dans les effets aléatoires, les termes d'erreur ou les deux. Nous discutons de l'estimation et de l'inférence des modèles proposés et nous les illustrons à l'aide d'exemples fondés sur des données réelles. Les études par simulation révèlent une plus grande robustesse à diverses valeurs aberrantes dans le cas des modèles *t* logistiques. L'analyse des données indique que les modèles robustes proposés produisent des conclusions cohérentes pour les données avec ou sans valeurs aberrantes.

[MS-180]

The Cox Model for Epidemiological Case-control Data: a Simulation Study

Le modèle de Cox pour les données épidémiologiques cas-témoins : une étude de simulation

Karen Leffondré, *Université de Montréal*, Willy Wynant, *Centre de recherche du CHUM*, Zhirong Cao, *Université de Montréal*, Michal Abrahamowicz, *McGill University*

Many exposures investigated in epidemiological case-control studies may vary over time. The effects of these exposures are usually estimated using logistic regression, which does not directly account for changes in covariate values over time. By contrast, the Cox model with time-dependent covariates directly accounts for these changes over time. However, the over-sampling of cases relative to controls requires using different weighting for cases and controls in each risk set. In this talk, we propose different weighting schemes for cases and controls, and we evaluate their performance for different types of time-varying covariates through a comprehensive simulation study.

De nombreuses expositions explorées dans les études épidémiologiques cas-témoins peuvent varier au cours du temps. Les effets de ces expositions sont habituellement estimés en utilisant la régression logistique qui ne tient pas directement compte des changements au cours du temps des valeurs des covariables. Par opposition, le modèle de Cox avec covariables dépendantes du temps tient directement compte de ces changements. Cependant, le sur-échantillonnage des cas par rapport aux témoins requiert d'utiliser des pondérations différentes pour les cas et les témoins dans chaque ensemble de sujets à risque. Dans cette présentation, nous proposons différents schémas de pondération des cas et des témoins, et nous évaluons leur performance pour différents types de covariables dépendantes du temps au moyen d'une étude de simulation.

[MS-181]

Model-Based Clustering using *t* Mixtures with Box-Cox Transformation

Classification à base de modèles utilisant des mélanges de loi *t* avec transformation de Box-Cox

Kenneth Lo, *University of British Columbia*, Ryan Brinkman, *Terry Fox Laboratory, BC Cancer Research Center*, Raphael Gottardo, *University of British Columbia*

The most commonly used model-based clustering approach is based on finite Gaussian mixtures. However, its normality assumption is not often realistic. Furthermore, Gaussian mixture models are not robust against outliers. To address these issues, we propose a *t* mixture model with Box – Cox transformation, which provides a unified framework to handle outlier identification and data transformation simultaneously. We describe an EM algorithm to estimate model parameters (including transformation selection). We will also examine several variants of EM to speed up the algorithm. To demonstrate the proposed methodology, we discuss a cluster analysis on cell population identification in flow cytometry data.

L'approche de classification à base de modèles la plus fréquemment utilisée est basée sur des mélanges gaussiens finis. Cependant, l'hypothèse de normalité n'est pas souvent réaliste. De plus, les modèles de mélanges gaussiens ne sont pas robustes aux valeurs aberrantes. Pour répondre à ces difficultés, nous proposons un modèles de mélange de lois *t* avec transformation de Box-Cox, qui fournit un cadre unifié pour identifier les valeurs aberrantes et simultanément transformer les données. Nous décrivons un algorithme EM pour estimer les paramètres du modèle (incluant la sélection de la transformation). Nous examinons aussi plusieurs variantes de EM afin d'augmenter la vitesse de l'algorithme. Pour faire la démonstration de la méthodologie proposée, nous discutons d'une analyse de classification en identification de populations de cellules avec des données de cytométrie en flux.

[MS-182]
Soft Node Trees for Predicting Continous and Multinomial Variables
Arbres à nœuds mous pour prévoir des variables continues et multinomiales

Benjamin Rich, Alina Dyachenko & Antonio Ciampi, *McGill University*

A new implementation of the soft node tree model is presented, in which a response is allowed to be multinomial or continuous, as in CART. The soft node tree is a model that differs from CART in that a split on a continuous variable is represented by a sigmoid function, giving the probability of belonging to either branch. It retains the advantages of CART, namely increased flexibility (over say logistic regression) through the nonlinearity introduced by the tree structure, and ease of interpretation. We have produced an extensible and reusable framework, with emphasis on software engineering and performance.

Une nouvelle mise en œuvre du modèle d'arbre à nœuds mous est présentée dans laquelle une réponse peut être multinomiale ou continue, comme dans les CART. L'arbre à nœuds mous est un modèle qui diffère des CART par le fait qu'une division pour une variable continue est représentée par une fonction sigmoïde, donnant la probabilité d'appartenir à chacune des deux branches. Il conserve les avantages des CART, soient une flexibilité accrue (par rapport à, par exemple, la régression logistique) grâce à la non-linéarité induite par la structure d'arbre, et l'aisance d'interprétation. Nous avons produit un cadre réutilisable et pouvant être élargi, en mettant l'emphase sur le génie logiciel et la performance.

[MS-183]
Bias Reduction for g-Estimation of Optimal Dynamic Regimes at Exceptional Laws
Réduction de biais pour g-estimation de régimes optiques dynamiques à lois exceptionelles

Erica Moodie, *McGill University*, Thomas Richardson, *University of Washington*

A dynamic regime is a function that takes treatment and patient-specific covariate history as inputs and returns a treatment decision. Robins (1986) introduced g-estimation and recently showed that this can be used to make inference about the optimal dynamic regime (Robins, 2002; 2004). This method is always consistent, but can be asymptotically biased under the so-called exceptional laws. In this talk, I will explain the problem of exceptional laws and introduce a new approach to g-estimation that shares all of the asymptotic properties of ordinary g-estimates at non-exceptional laws while providing substantial reduction in the asymptotic bias at exceptional laws.

Un régime de traitement dynamique est une fonction qui prend l'histoire de traitements et de covariables des patients comme données et rend une décision de traitement. Robins (1986) a présenté la g-estimation et a récemment montré que cela peut être utilisé pour faire l'inférence du meilleur régime dynamique (Robins, 2002 ; 2004). Cette méthode converge toujours, mais peut être asymptotically biaisée aux lois exceptionnelles. Dans cette présentation, j'expliquerai le problème de lois exceptionnelles et présenterai une nouvelle approche à la g-estimation qui partage toutes les propriétés asymptotiques des g-estimations ordinaires aux lois non exceptionnelles en fournissant la réduction substantielle à biais asymptotique aux lois exceptionnelles.

[MS-184]
Assessing Goodness of Fit of Generalized Linear Models to Sparse Data using Higher Order Moment Corrections
Évaluer la qualité de l'ajustement de modèles linéaires généralisés pour les données dispersées en utilisant des corrections plus élevées de l'ordre du moment

Dianliang Deng, *University of Regina*, Sudhir R. Paul, *University of Windsor*

We will present the procedure of assessing goodness-of-fit properties of the conditional distribution of the Pearson statistic for non-canonical general-

Nous présenterons la méthode d'évaluation des propriétés d'ajustement de la distribution conditionnelle de la statistique de Pearson pour des modèles linéaires généralisés non

ized linear models with data that are extensive but sparse, by Edgeworth approximation of p-values using higher order moment corrections. Some simulations are conducted to compare the performance of the classical Pearson statistic, a standardized modified Pearson statistic, a standardized modified deviance statistic, a modified Pearson statistic based on Edgeworth approximation with the first three conditional moments and a modified Pearson statistic based on Edgeworth approximation with the first four conditional moments. Two examples and a discussion are presented.

canoniques et des données de grande portée, mais dispersées à l'aide de l'approximation d'Edgeworth des valeurs p en utilisant des corrections de moments d'ordre plus élevé. Nous effectuons certaines simulations pour comparer les propriétés de la statistique classique de Pearson, d'une statistique de Pearson modifiée normalisée, d'une statistique de déviance modifiée normalisée, d'une statistique de Pearson modifiée fondée sur l'approximation d'Edgeworth avec les trois premiers moments conditionnels et d'une statistique de Pearson modifiée fondée sur l'approximation d'Edgeworth avec les quatre premiers moments conditionnels. Nous donnerons deux exemples et présenterons une discussion.

Session 09I Mar. 27 mai • Tues. May 27, 14:35 – 16:05 Capital Hall 9A

Probability Laws and Asymptotic Properties
Lois de probabilité et propriétés asymptotiques

[MS-185]
A Class of Matsumoto – Yor Functions
Une classe de fonctions de Matsumoto-Yor

Efoevi Koudou & Pierre Vallois, *Institut Elie Cartan, Nancy, France*

We define Matsumoto – Yor functions as decreasing functions from $(0,\infty)$ onto $(0,\infty)$ with the following property: there exist independent, positive random variables X and Y such that the variables $f(X+Y)$ and $f(X)-f(X+Y)$ are independent. An example of such a function follows from works by Matsumoto, Yor and Letac: $f(x)=1/x$, the law of X being necessarily inverse-Gaussian while Y is gamma-distributed. We investigate a class of sufficiently regular Matsumoto-Yor functions with a prescribed behaviour at 0 and infinity. This leads, up to constants, to the functions defined by $f(x)=1/x$, $f(x)=1/(e^x-1)$ and their inverse functions. The associated densities are provided.

Nous définissons les fonctions de Matsumoto-Yor : les fonctions f strictement décroissantes de $(0,\infty)$ dans $(0,\infty)$ pour lesquelles il existe des variables aléatoires positives X, Y indépendantes telles que les variables $f(X+Y)$ et $f(X)-f(X+Y)$ soient indépendantes. D'après des travaux de Matsumoto, Yor, Letac, $f(x)=1/x$ a cette propriété et les lois associées sont nécessairement gaussienne inverse pour X, gamma pour Y. Nous caractérisons les fonctions de Matsumoto-Yor suffisamment régulières, avec un comportement spécifique en 0 et à l'infini. On trouve, modulo des constantes, les fonctions définies par $f(x)=1/x$, $f(x)=1/(e^x-1)$ et leurs réciproques. Les lois correspondantes sont précisées.

[MS-186]
MDP and LIL for the Empirical Covariance of Linear Hilbertian Processes
PDM et LLI pour la covariance de processus linéaires hilbertiens

Ludovic Menneteau, *Université Montpellier II*

In this talk, the moderate deviation principle and the law of the iterated logarithm for the empirical covariance of Hilbertian linear processes are considered. Several applications are provided: Moderate deviation principles and laws of the iterated log-

On s'intéresse ici au principe de déviations modérées (PDM) et à la loi du logarithme itérée (LLI) de la covariance empirique d'un processus linéaire hilbertien. Plusieurs applications sont considérées : Dans le cadre de l'analyse en composante principale, on obtient le PDM et la LLI des valeurs

arithm for the eigenvalues and associated projectors of the empirical covariance are obtained. Moderate deviation principle and law of the iterated logarithm for the Yule Walker estimator of the coefficient of a finite dimensional valued autoregressive process is also considered.

propres et des projecteurs associés de l'opérateur de covariance empirique. En dimension finie, le PDM et la LLI de l'estimateur de Yule Walker du coefficient d'un processus autorégressif d-dimensionnel sont aussi considérés.

[MS-187]
Asymptotic Universal Theorem
Un théorème asymptotique universel

Lorenzo Molkorogov, *Université inter-age de Morbier.*

We have provided results on the weak convergence of a series of random variables, which are all sums of independent random variables and are all distributed identically towards a specific random variable. Our result, called the "Universal Asymptotic Theorem," involves a sum of random variables whose number leads to infinity. These variables are believed to be independent and follow the same rule. To achieve a final result, the sum must be centered by subtracting its average and dividing it by its standard deviation. Under relatively broad conditions, the law of probability converges towards a normal unit law.

Nous donnons un résultat sur la convergence faible d'une suite de variables aléatoires. Toute somme de variables aléatoires indépendantes et identiquement distribuées tend vers une certaine variable aléatoire. Notre résultat, appelé « théorème asymptotique universel », concerne une somme de variables aléatoires dont le nombre tend vers l'infini. Ces variables sont supposées indépendantes et de même loi. Pour obtenir un résultat fini, il faut centrer cette somme en lui soustrayant sa moyenne et la réduire en la divisant par son écart-type. Sous des conditions assez larges, la loi de probabilité converge alors vers une loi normale unitaire.

[MS-188]
Strong Approximation of the Function Indexed Empirical Process Under Random Entropy Condition.
Approximation forte du processus empirique indexé par des fonctions sous condition d'entropie aléatoire.

Adel Settati, *IRMAR - Université Rennes 1*

We obtain some rates of strong approximation of the function indexed empirical process by a sequence of Brownian bridges. We work under random entropy conditions and adapt the recent technique of Berthet and Mason (2006). In view of Giné and Zinn (1984) and Talagrand (1987) our weakest condition is close to necessary for the Donsker property, but however guaranty a rate $(\log n)^{-a}$ which significatively improves the universal $(\log\log n)^{1/2}$ of Dudley and Philipp (1983). Our strongest condition leads to polynomial rates of Gaussian approximation. We also investigate results of Koltchinskii (1994) and Rio (1994) when their additional analytical assumptions are relaxed.

Nous obtenons des vitesses d'approximation forte du processus empirique par une suite de ponts browniens dans le cadre indexé par des fonctions. Nous travaillons sous des conditions d'entropie aléatoire et adaptons la méthode de Berthet et Mason (2006). Au vu de Giné et Zinn (1984) et Talagrand (1987) notre condition la plus faible est quasiment nécessaire pour la propriété de Donsker, mais garantie néanmoins une vitesse $(\log n)^{-a}$ qui améliore significativement la vitesse universelle $(\log\log n)^{1/2}$ de Dudley et Philipp (1983). Notre condition la plus forte conduit à des vitesses d'approximation gaussienne polynomiales. Nous revisitons également les résultats de Koltchinskii (1994) et Rio (1994) lorsque leurs hypothèses analytiques supplémentaires sont retirées.

[MS-189]

A Ratio Limit Theorem for a Markov Kernel with Unbounded Jumps
Un théorème limite du rapport pour un noyau markovien à sauts non bornés

Aziz Khanchi, *University of Ottawa*

We provide the limiting distribution of the Markovian part of a Markov additive chain when its additive part passes a threshold which is tending to infinity. We extend the results in [H. Kesten, Ann. Probab. **2** (1974) 355 – 386] to cover the cases where the Markovian part of the associated twisted chain is not positive recurrent. The twisted chain is obtained by using Doob's *h*-transform. Moreover, a ratio limit theorem will be established for the level crossing kernel. This is an interesting result since the level crossing kernel has unbounded jumps and does not satisfy the known sufficient conditions.

Nous donnons la distribution limite de la partie markovienne d'une chaîne de Markov additive lorsque sa partie additive dépasse un seuil qui tend vers l'Infini. Nous élargissons les résultats de [H. Kesten, Ann. Probab. **2** (1974) 355-386] pour couvrir les cas ou la partie markovienne de la chaîne torsadée associée n'est pas récurrente positive. La chaîne torsadée est obtenue en utilisant la transformée-*h* de Doob. En outre, un théorème limite du rapport sera établi pour le noyau de passage à niveau. Ceci est un résultat intéressant car le noyau est à sauts non bornés et ne satisfait pas les conditions suffisantes connues.

Session 09J Mar. 27 mai • Tues. May 27, 14:35 – 16:05 Congress Hall A/E West

Poster Session II
Séance d'affichage II

[MS-190]

Depression and Bone Mineral Density: A Meta-analysis of Epidemiological Studies
La dépression et la teneure minérale de l'os : une méta-analyse des études épidémiologiques

Qing Wu, *Mayo Clinic Arizona*, Jeanette Magnus & Bencaz Angelle, *Tulane University*

This study presents results of a meta-analysis of depression and BMD. English-language articles published before May 2006 were used as data sources. A total of 11 studies met our inclusion criteria. Using a standard protocol, two team members independently abstracted the data. In a random-effects model, data from each study were pooled and weighted by the inverse of the total variance. Our study shows that depression is significantly associated with low BMD in hip ($p = 0.012$), however it does not show the association is significant in spine ($p = 0.13$). This study provides a thorough review of depression and BMD.

Cette étude présente les résultats d'une méta-analyse de la dépression et de la teneure minérale de l'os. Les articles publiés en anglais avant mai 2006 ont été utilisés comme source des données. Onze études en tout ont satisfait nos critères d'inclusion. En utilisant un protocole standard, deux membres de l'équipe ont fait le compte rendu des données d'une manière indépendante. Dans un modèle à effet aléatoire, les données de chaque étude ont été rassemblées et pondérées par l'inverse de la variance totale. Notre étude montre que la dépression est significativement associée avec une faible teneure minérale de l'os pour ce qui est de la hanche ($p = 0{,}012$). Cependant, cela ne montre pas que l'association est significative pour ce qui est de la colonne vertébrale ($p = 0{,}13$). Cette étude présente une revue en profondeur de la dépression et de la teneure minérale de l'os.

[MS-191]
Waiting Time for Health Services
Temps d'attente pour des services de santé

Qi Er Tan, James Fu & Xikui Wang, *University of Manitoba*

We analyze the waiting time for key medical diagnostics and treatment service, using the Canadian Community Health Survey (CCHS) database. The waiting time is modeled by the generalized exponential distribution which is specified by a vector of covariates of patients' backgrounds. Important covariates are selected, and the best fitted distribution function is determined and the parameters of the fitted distribution function are estimated. Furthermore, the model is validated by different statistical test procedures.

Nous analysons le temps d'attente pour les services de diagnostics et de traitements médicaux clés en utilisant la base de données de l'Enquête sur la santé dans les collectivités canadiennes (ESCC). Le temps d'attente est modélisé par la distribution exponentielle généralisée laquelle est précisée par un vecteur de covariables des antécédents des patients. Des covariables importantes sont sélectionnées et la meilleure fonction de distribution ajustée est déterminée tout comme les paramètres de la fonction de distribution sont estimés. De plus, le modèle est validé par différentes procédures de tests statistiques.

[MS-192]
'Quantile' is a Better Choice when 'Average' is Not Effective in Methylation Microarray Analysis
Un 'quantile' est un meilleur choix lorsque la 'moyenne' n'est pas efficace pour l'analyse de micropuce de méthylation

Shuying Sun, Shuying Sun & Dustin Potter, *The Ohio State University*

DNA methylation is thought to play an important role in gene silencing. In order to identify methylation signature, high throughput microarrays have been generated for 50+ breast cancer cell lines. Methylation microarray analysis is very challenging due to the existence of many measured and unmeasured confounding factors. To overcome this challenge, we use the quantile regression to model methylation level of probes on CpG islands. The performance of our method is good as it can correctly identify known methylated and unmethylated genes. With this method, we have identified regions/genes that are commonly methylated among breast cancer cell lines.

La méthylation d'ADN est supposée jouer un rôle dans le processus réduisant un gène au silence. Afin d'identifier la signature de méthylation, des micropuces à haut débit ont été générées pour plus de 50 lignées cellulaires du cancer du sein. L'analyse de méthylation de micropuces est très difficile à cause de la présence d'un grand nombre de facteurs confondus mesurés et non mesurés. Pour surmonter ce défi, nous utilisons la régression quantile afin de modéliser le niveau de méthylation de sondes sur des îles CpG. Notre méthode est performante car elle peut identifier correctement les gènes méthylés et non méthylés. Avec cette méthode, nous avons identifié des régions/gènes qui sont fréquemment métylés parmi les lignées cellulaires du cancer du sein.

[MS-193]
Predictive Models for the Growth of Microbial Pathogens
Modèles prédictifs pour la croissance de pathogènes microbiens

Eunhee Kim, *National Institute of Toxicological Research*, Gun-Young Lee, *Korea Fooda and Drug Administration*, Munsin Kim, *National Institute of Toxicological Research*, Ki-Sung Kwon, *Korea Fooda and Drug Administration*, Dongdeuk Jang & Bong-Hyun Nam, *National Institute of Toxicological Research*

This study illustrated the statistical inference process for establishing the predictive models for the growth of microbial pathogens, using the general primary models and the secondary models which

Cette étude illustre le processus d'inférence statistique pour l'établissement des modèles prédictifs pour la croissance de pathogènes microbiens en utilisant les modèles primaires généraux et les modèles secondaires qui sont utilisés dans

are used in the fields of Quantitative Microbial Risk Assessment (QMRA). The modeling processes for microbial growth consist of estimating the parameters of the primary and secondary models and combining two models. Several issues from the combination of the two models were considered and the real data were applied to the predictive model for the growth of Bacillus cereus in ready-to-eat foods.

les champs du Quantitative Microbial Risk Assessment (QMRA) [Évaluation quantitative du risque microbien]. Les processus de modélisation pour la croissance microbienne se composent de l'estimation des paramètres des modèles primaires et secondaires et de la combinaison de deux modèles. Plusieurs questions provenant de la combinaison de deux modèles ont été considérées et les données réelles ont été appliquées au modèle prédictif pour la croissance du Bacillus cereus dans les aliments prêt-à-manger.

[MS-194]

Automated Public Health Surveillance (PHS) Systems: An Open-source Implementation
Systèmes automatisés de surveillance de la santé publique (SSP) : une implémentation en code source ouvert

Mohamed Abdolell & Jennifer Payne, *Dalhousie University*, Judy Caines, *Nova Scotia Breast Screening Program*, Wendy Lou, *University of Toronto*, Sam Stewart, *University of Waterloo*

The Nova Scotia Breast Screening Program screening and diagnostic information systems are unique nationally in that patient trajectories are captured in real-time, for all mammographically-screened women province-wide. We outline the automation process for the generation of the NSBSP Annual Report exclusively using GPL software including R, LaTeX, Sweave, Perl, and several other helper scripting languages on Ubuntu 7.10. A key goal is to explore the feasibility of integrating SPC methods in PHS for the early detection of emerging trends. We discuss key features of PHS and lessons learned as we plan the development of a fully automated surveillance system.

Les systèmes de dépistage et d'information diagnostique du Programme de dépistage du cancer du sein de la Nouvelle-Écosse sont uniques au Canada en ce sens que les trajectoires des patientes sont saisies en temps réels, pour toutes les femmes de la province qui subissent une mammographie de dépistage. Nous décrivons dans les grandes lignes le processus d'automation pour la production des rapports annuels du Programme en utilisant exclusivement des logiciels en langage de programmation universel, dont R, LaTeX, Sweave, Perl et plusieurs autres langages de script auxiliaires tournant sur plateforme Ubuntu 7.10. Un objectif clé consiste à explorer la faisabilité de l'intégration des méthodes de contrôle statistique du processus (CSP) dans la SSP pour permettre la détection précoce de nouvelles tendances. Nous discutons des caractéristiques importantes de la SSP, ainsi que des leçons apprises en prévision du développement d'un système de surveillance entièrement automatisé.

[MS-195]

Use of Cluster Analyses to Assess Dietary Patterns: a Comparison of Unsupervised Clustering Methods.
L'utilisation des méthodes de classification pour identifier des typologies alimentaires : une comparaison des techniques de classification non-supervisées

Geraldine Lo Siou, Ilona Csizmadi, S. Elizabeth McGregor, Karen A. Kopciuk & Paula J. Robson, *Alberta Cancer Board*

In nutritional epidemiology, relationships between diet and chronic disease risk can be explored by focusing on overall dietary patterns rather than on individual nutrients. This study focuses on unsupervised cluster analyses. Although K-means and Ward's minimum variance are the most widely used clustering methods in nutritional epidemiol-

En épidémiologie nutritionnelle, les associations entre alimentation et maladies chroniques peuvent être étudiées en considérant l'alimentation dans sa globalité plutôt qu'en nutriments isolés. Cette étude se limite aux méthodes de classification non supervisées. Bien que les méthodes avec les centres mobiles et Ward soient les plus couramment utilisées en épidémiologie nutritionnelle, d'autres algorithmes

ogy, other algorithms could be more appropriate. Using dietary intake data (55 food groups) reported by 6,445 men and 10,299 women in The Tomorrow Project®, a longitudinal cohort study in Alberta, Canada, we will evaluate the consistency of the results obtained from applying various unsupervised clustering techniques.

pourraient être plus appropriés. Utilisant les données sur la consommation alimentaire (55 groupes alimentaires) rapportées par 6 445 hommes et 10 299 femmes participant au The Tomorrow Project®, une étude longitudinale de cohorte en Alberta, Canada, nous évaluerons la cohérence des résultats obtenus en appliquant diverses techniques de classification non supervisées.

[MS-196]

Response Adaptive Designs with a Variance-penalized Criterion
Plans de réponses adaptées ayant un critère de variance pénalisée

Yanqing Yi & Xikui Wang, *University of Manitoba*

We consider a response adaptive design of clinical trials with a variance-penalized criterion. It is shown that this criterion evaluates the performance of a response adaptive design based on both the number of patients assigned to the better treatment and the power of the statistical test. A new proportion of treatment allocation is proposed. Under reasonable assumptions, the proposed design is demonstrated to generate an asymptotic variance of allocation proportions which is smaller than that of the drop-the-loser design. Simulation comparisons of the proposed design with some existing designs are presented.

Nous considérons une conception adaptative en fonction de la réponse d'essais cliniques comportant un critère de variance pénalisé. Nous montrons que ce critère évalue la performance d'un plan adaptatif en fonction de la réponse d'après le nombre de patients affectés au meilleur traitement ainsi que la puissance du test statistique. Nous proposons une nouvelle proportion pour la répartition entre les traitements. Sous des hypothèses raisonnables, nous démontrons que le plan proposé produit une variance asymptotique des proportions de répartition plus faible que celle du plan avec abandon du traitement perdant (drop-the-loser). Nous présentons des comparaisons par simulation du plan proposé à certains plans existants.

[MS-197]

Advantages of Adaptive Sampling Designs for Sampling from a Network Setting
Avantages des plans d'échantillonnage adaptatifs pour l'échantillonnage à partir de l'établissement de réseaux

Kyle Vincent, *Simon Fraser University*

With the increasing prevalence of diseases like HIV, efficient methods of sampling from hidden human populations are desired to help reduce infection. Wherever an abundance of observations of interest are encountered, closely associated individuals are more likely to reveal a higher than expected incidence rate. It is usually more feasible to adaptively include these associated units in the sample, as opposed to randomly selecting new units from the population. This study evaluates the use of new adaptive designs in network and spatial settings, and determines when efficiency gains over conventional sampling techniques are likely to be seen in certain populations.

Étant donné la prévalence croissante de maladies telles que l'infection par le VIH, de bonnes méthodes d'échantillonnage des populations humaines cachées sont souhaitables afin de réduire la propagation de l'infection. Quand on se trouve face à une abondance d'observations d'intérêt, les individus étroitement associés aux sujets observés sont plus susceptibles de révéler un taux d'incidence plus élevé que prévu. Il est généralement plus commode d'inclure de façon adaptative ces unités associées dans l'échantillon que de sélectionner aléatoirement de nouvelles unités dans la population. La présente étude a pour but d'évaluer l'utilisation de nouveaux plans d'échantillonnage adaptatifs dans des conditions de réseau et de dimension spatiale, et de déterminer quand il est probable que des gains d'efficacité par rapport aux méthodes d'échantillonnage classique seront réalisés dans certaines populations.

[MS-198]

Geo-temporal Spread of Influenza A, at Different Geographic Scales, Canada and United States
La propagation de la grippe à différentes échelles géographiques au Canada et aux États-Unis

Dena Schanzer, *Public Health Agency of Canada*, Trevor Dummer & Joanne Langley, *Dalhousie University*, Samina Aziz, Brian Winchester & Theresa Tam, *Public Health Agency of Canada*

Recent modeling efforts simulating to the spread of pandemic influenza have focussed on airline travel, however empirical evidence is limited. Using weekly reports of laboratory confirmed influenza (FluWatch, Canada and CDC Flu, US) and other Canadian data with finer spatial details, we will characterize the geo-temporal spread of influenza over the 1997/98 – 2005/06 period at both the large and small scale. Variation in direction of spread suggests that it is unlikely that travel alone is responsible for the dynamics of influenza transmission. Influenza appears to be geographically widespread early in the epidemic, though important local differences can occur.

À l'heure actuelle, on ne sait pas comment la grippe se transmet à travers de vastes régions géographiques en peu de temps chaque année. Récemment on a mis l'accent sur la modélisation des voyages aériens bien que des données empiriques soient limitées. Des rapports hebdomadaires du nombre de cas confirmés de la grippe (FluWatch, Canada et CDC Flu, États-Unis) nous ont permis de caractériser la propagation de la grippe de 1997 à 2006 à grande échelle. Au Canada, des données plus raffinées spatialement nous donnent une image de la transmission régionale. La direction de propagation varie, ce qui suggère qu'il est peu probable que les voyages aériens soient le facteur le plus important pour la dynamique de transmission de la grippe. Les cas de la grippe semblent bien étendus géographiquement au début de l'épidémie, même si d'importantes différences locales peuvent survenir.

[MS-199]

Estimating Efficacy in Clinical Trials: Case Study of NCIC CTG MA.21 Adjuvant Breast Cancer Trial
Estimation de l'efficacité dans les essais cliniques : étude de cas de l'essai de traitement adjuvant du cancer du sein NCIC CTG MA.21

Judy Anne Chapman, Patti O'Brien, Zhangjun Ding, Lois Shepherd, *NCIC Clinical Trials Group Queen's University*, Margot Burnell, *Atlantic Health sciences Corporation*, Mark Levine, *McMaster University*, Kathleen Pritchard, *Sunnybrook Odette Cancer Centre Sunnybrook Health Sciences Centre*, David Matthews, *University of Waterloo*, Chunsong Qu, *NCIC Clinical Trials Group Queen's University*

Estimates of efficacy in clinical trials are frequently obtained from unadjusted Kaplan – Meier (K-M) or Cumulative Incidence (CI) curves. NCIC CTG MA.21 is an adjuvant breast cancer trial of women 60 years or less, with node positive or high risk node negative disease, which closed following the first Interim Analyses due to significant treatment differences between study therapies. We compared 3-year estimated efficacy using K-M, unadjusted and adjusted CI, unadjusted and adjusted Cox survivor models. Differences between statistical methods exceeded intra-method differences by treatment arm.

Les estimations de l'efficacité dans les essais cliniques sont souvent obtenues à l'aide des courbes non corrigées de Kaplan-Meier (K-M) ou d'incidence cumulative (IC). L'étude NCIC CTG MA.21, un essai de traitement adjuvant du cancer du sein ayant pour sujets des femmes de 60 ans ou moins présentant une tumeur avec ganglions positifs ou une tumeur à haut risque avec ganglions négatifs, a été interrompue après les premières analyses intérimaires, à cause de différences significatives entre les résultats des traitements étudiés. Nous comparons l'efficacité estimée sur trois ans à l'aide des modèles de survie de K-M, IC non corrigé et corrigé, et de Cox non corrigé et corrigé. Les différences entre les méthodes statistiques sont supérieures aux différences intra-méthode selon le groupe de traitement.

[MS-200]

Analysis of Incomplete Longitudinal Data: an Application to Self-Reported Smoking Status among Adolescents from the Waterloo Smoking Prevention Project

Analyse des données longitudinales incomplètes : une application du statut de fumeur autodéclaré parmi les adolescents à partir du projet de prévention contre le tabac de Waterloo

Robin Burkhalter, Grace Y. Yi & K. Stephen Brown, *University of Waterloo*

This paper presents the analyses of longitudinal data from the Waterloo Smoking Prevention Project (WSPP) that follows a cohort of youth through seven years and measures the self-reported smoking status and associated variables. The WSSP data typically contain missing observations. We address this feature with several modeling strategies. Weighted Generalized Estimating Equations (GEE) approach is specifically employed in the analyses to adjust for the bias induced by missing data.

La communication a pour but de présenter les analyses des données longitudinales provenant du Waterloo Smoking Prevention Project (WSPP) qui consistent à suivre une cohorte de jeunes pendant sept ans et à mesurer la situation autodéclarée d'usage du tabac et les variables connexes. En général, les données du WSSP présentent des observations manquantes. Nous adoptons plusieurs stratégies de modélisation pour résoudre ce problème. Nous choisissons spécifiquement l'approche des équations d'estimation généralisées (EEG) pondérées pour corriger le biais dû aux données manquantes dans les analyses.

[MS-201]

Confidence Interval and Permutation Test in the Context of MapArg Method

Intervalle de confiance et test de permutation dans le contexte de la méthode MapArg

Nataliya Dragieva, *Université du Québec à Montréal*

The genetics cartographic method MapArg calculates the likelihood estimator of the potential mutation position via an ancestral recombination graph. In order to test the quality of this estimator in the context of MapArg method, a confidence interval is built using the bootstrap method and significance thresholds for the likelihood function are set by the permutation test. For the simulated datasets, the bias of the estimator (MapArg) is corrected in the cases when the initial method gives inaccurate estimations.

La méthode MapArg (Larribe *et al*., 2002) de cartographie génétique fournit un estimateur à maximum de vraisemblance pour la position probable de la mutation (par « graphe de recombinaison ancestrale »). Afin de tester la qualité de cet estimateur, dans le contexte de la méthode MapArg un intervalle de confiance pour l'estimateur obtenu est construit par la méthode du « bootstrap » et des seuils de signification pour la fonction de la vraisemblance sont établis par un test de permutation. Pour les ensembles de données simulés, le biais de l'estimateur (MapArg) était corrigé avec succès dans certains cas où la méthode initiale donne de mauvaises estimations.

[MS-202]

Power Calculation for Detecting Gene-environment Interaction in the Ontario Cancer Cohort

Calcul de la puissance de détection des interactions gènes et environnement dans la *Ontario Cancer Cohort*

Hedy Jiang & Patrick Brown, *Cancer Care Ontario*

The Ontario Cancer Cohort will recruit individuals aged 35 – 69, taking base-line measurements, tracking risk factors and assessing the community-level environmental variables in Ontario. As is the case with most large studies, a power analysis is needed to inform and justify the sample size chosen. The cohort was simulated at the individual

L'Ontario Cancer Cohort est une initiative qui consistera à recruter des sujets de 35 à 69 ans, à prendre des mesures de référence, à faire le suivi des facteurs de risque et à évaluer les variables environnementales au niveau communautaire en Ontario. Comme cela est le cas dans la plupart des grandes études, une analyse de puissance est nécessaire pour décider de la taille d'échantillon requise et justifier ce choix.

level from the distribution of the Ontario population. Random effects were generated at individual and community levels, and cancer risk computed by modifying Ontario's observed incidence rates. A Cox proportional hazards model was then fit into the simulated datasets, resulting in powers of detecting gene-environment interactions and community level effects.

La cohorte a été simulée au niveau individuel en partant de la répartition de la population ontarienne. Les effets aléatoires ont été générés aux niveaux individuel et communautaire, et les risques de cancer ont été calculés en modifiant les taux d'incidence observés en Ontario. Un modèle à risques proportionnels de Cox a été ajusté aux ensembles de données simulés, ce qui a permis de déterminer la puissance de détection des interactions gènes-environnement et des effets au niveau communautaire.

[MS-203]

Extension of Cochran – Armitage Test for Analysis of Correlated Ordinal Data
Extension du test de Cochran-Armitage pour l'analyse de données ordinales corrélées

Ruochu Gao, Neil Klar & Allan Donner, *University of Western Ontario*

Considerable attention in the literature has been given to the analyses of independent ordinal outcomes. For example, the Cochran – Armitage statistic may be used to test for the effect of intervention using ordinal data from an individually randomized trial. We describe analytic derivations and use a simulation study to compare extensions of the Cochran – Armitage statistic for analyses of correlated ordinal outcomes. Our discussion focuses on application to cluster randomized trials. Methods are illustrated using data from a school-based smoking prevention program.

Beaucoup d'attention a été accordée dans la littérature aux analyses de résultats ordinaux indépendants. Par exemple, la statistique de Cochran-Armitage peut être utilisée pour tester l'effet de l'intervention en utilisant des données ordinales provenant d'un essai randomisé individuellement. Nous décrivons les dérivations analytiques et réalisons une étude par simulation afin de comparer les extensions de la statistique de Cochran-Armitage pour les analyses de résultats ordinaux corrélés. Notre discussion est axée sur l'application à des essais randomisés par grappes. Les méthodes sont illustrées à l'aide de données provenant d'un programme de prévention du tabagisme mis en place dans les écoles.

[MS-204]

Binary Recursive Partitioning for Continuous Longitudinal Outcomes
Partitionnement binaire récursif pour des résultats continus longitudinaux

Mohamed Abdolell, *Dalhousie University*, Michael LeBlanc, *Fred Hutchinson Cancer Research Center*, Sam Stewart, *QEII Health Sciences Centre*

A binary recursive partitioning algorithm that models continuous longitudinal outcomes is implemented in the longRpart R package, in which a wrapper function is written for the rpart function, thus leveraging all of rpart's tree growing and plotting tools. The deviance measure (Mahalanobis distance) is derived from the -2 log-likelihood and passed to the rpart function to evaluate the goodness-of-split criteria; nine correlation structures can be specified, several of which are AR(1), ARMA, compound symmetry, exponential spatial, and Gaussian spatial. A plotting function is developed that graphically annotates nodes. longRpart is applied to repeated CD4 counts data from an AIDS clinical trial.

Un algorithme de partitionnement binaire récursif qui modélise des réponses longitudinales continues est mis en œuvre dans la librairie R longRpart, où une fonction enveloppante est écrite pour la fonction rpart, tirant ainsi avantage de tous les outils de rpart relatifs à la croissance et au traçage d'arbres. La mesure de déviance (la distance de Mahalanobis) est obtenue à partir de -2 log-vraisemblance et est envoyée à la fonction rpart pour évaluer le critère d'ajustement. Neuf structures de corrélation peuvent être spécifiées, dont AR(1), ARMA, symétrique composée, spatiale exponentielle et spatiale gaussienne. Une fonction de traçage est développée qui permet d'annoter graphiquement les nœuds. longRpart est appliquée à des données de dénombrement CD4 répétées provenant d'un essai clinique sur le SIDA.

[MS-205]

A Different Approach to Assessing Departure from Normality with the Z-Score
Une approche différente pour juger de l'écart à la nomalité avec le Z-score

Françoise Tondu, *Danone*

In a nutrition-health clinical study, a 2 group-parallel study design, and ordinal qualitative data (from 1 to 5, Likert scoring) are studied during the 2-week study of product consumption with baseline and final visits. An assessment of departure from normality has been studied with a different approach using z-score, which is built up with mean and standard deviation. Results of the distribution analysis for Baseline and Final levels including z-score have been shown with box-plot for decision. As departures from normality of distribution have been assessed, non-parametric analysis has been considered more robust than parametric analysis when comparing both groups.

Dans une étude clinique en nutrition santé, 2 groupes parallèles, une donnée qualitative ordinale (de 1 à 5, échelle de Likert) a été recueillie au début et en fin d'étude après 2 semaines de consommation du produit. L'écart à la normalité du paramètre a été exploré avec une approche différente en utilisant le Z-score, calculé avec la moyenne et l'écart-type de la variable. Les distributions des valeurs brutes et les z-scores au baseline et en fin d'étude ont été présentés en box-plot. Comme une distribution normale n'a pas pu être considérée, l'analyse non paramétrique a été retenue comme plus robuste que l'analyse paramétrique, pour comparer les deux groupes.

[MS-206]

Likelihood Ratio Test for Association in Discordant Sib-pairs
Test du rapport de vraisemblance pour l'association dans les paires de frères et sœurs discordantes

Vaneeta Grover, *Dalhousie University*, David Cole, *University of Toronto*, David Hamilton, *Dalhousie University*

A likelihood ratio test is proposed to assess association between a marker or candidate gene and a disease. The likelihood is based on joint genotypic probabilities for a discordant sib-pair under a general disease model. Validity and power of the test are investigated by computer simulation. The proposed method is compared with case-control and sib-TDT methods of examining association and linkage between a candidate locus and a disease.

Un test du rapport de vraisemblance est proposé pour évaluer l'association entre un gène marqueur ou un gène candidat et une maladie. La vraisemblance est basée sur les probabilités génotypiques conjointes pour une paire frère-sœur discordante sous un modèle général de maladie. La validité et la puissance du test sont étudiées par simulation informatique. La méthode proposée est comparée à des méthodes sib-TDT (test de déséquilibre de transmission chez les frères et sœurs) d'examen de l'association et du lien entre un locus de gène candidat et une maladie.

[MS-207]

Improved Class Prediction in Gene Expression Microarray Data
Prédiction de classe améliorée dans l'expression des gènes des données de microréseaux

Mohammad Islam & Angus McLeod, *University of Western Ontario*

A successful diagnosis and treatment of cancer depend on efficient classification of tumors. Micoarray enables us to monitor expression levels of thousands of genes simultaneously in a whole genome. Thus it allows us to meet the challenge of precisely distinguishing tumor classes, which is one of the important focuses of data-mining of microarray data analysis. In this research we develop an im-

Le succès du diagnostic et du traitement du cancer dépend de la bonne classification des tumeurs. Les micropuces nous permettent de surveiller simultanément les niveaux d'expression de milliers de gènes dans un génome complet. Ils nous forcent donc à relever le défi de distinguer précisément les classes de tumeurs, ce qui constitue l'un des aspects importants de l'exploration des données dans l'analyse des données micropuces. L'étude porte sur

proved classification rule, which outperforms the existing classification methods. The application of the proposed method to simulated as well as real gene expression data confirms the claim.

l'établissement d'une règle améliorée de classification qui donne de meilleurs résultats que les méthodes de classification existantes. L'application de la méthode proposée à des données simulées ainsi que réelles sur l'expression des gènes confirme cette allégation.

[MS-208]
Hotspot Detection in Disease Mapping
Détection de points chauds en cartographie des maladies

Giovani Silva, Cindy Xin Feng & C.B. Dean, *Simon Fraser University*

In disease mapping, Bayesian hierarchical models have often been used to provide smoothed mortality and incidence rate estimates. We consider the use of these models for identifying extreme rates or relative risks, so-called "hotspots." In addition and for comparison, we also consider the weighted rank squared error loss function (Wright et al., 2003) for the estimation of extrema. Simulation studies compare different approaches for making inference on regional ranks. Finally, an analysis of lung cancer incidence in Ontario over the period 1995 – 2002 illustrates the methods developed.

En cartographie des maladies, les modèles bayésiens hiérarchiques ont souvent été utilisés pour fournir des estimateurs lissés des taux d'incidence et de mortalité. Nous considérons l'utilisation d'un tel modèle pour identifier des taux ou risques relatifs extrêmes, dits « points chauds ». De plus, et à des fins de comparaisons, nous considérons aussi la fonction de perte d'erreur quadratique des rangs pondérés (Wright *et al.*, 2003) pour l'estimation des extremums. Des études de simulation comparent différentes approches pour faire l'inférence sur les rangs régionaux. Finalement, une analyse de l'incidence du cancer du poumon en Ontario de 1995 à 2002 illustre les méthodes développées.

Session 10A Mar. 27 mai • Tues. May 27, 16:30 – 18:00 Capital Hall 1A

Some New Perspectives on Longitudinal Data Analysis
Nouvelles perspectives en analyse de données longitudinales

[MS-209]
Detecting the Omission of a Time-dependent Confounder in Longitudinal Models
Identifier l'oubli d'une variable chronologique dans un modèle longitudinal

Michel Chavance, *INSERM, Université Paris XI*

In longitudinal models, omitting a time-dependent confounder can bias the estimate of the time-dependant variable of interest. It is well known that in such a situation, the between-subjects and the within-subjects effects of this variable can differ. However, it is less well known that estimators based on different assumptions about the covariance structure of the outcomes can also converge to different values. Using real and simulated data, we illustrate this phenomenon and we consider how it can be used to detect the omission of a confounder.

Dans un modèle longitudinal, l'oubli d'une variable chronologique entraîne un risque de biais sur la mesure de l'effet de la variable d'intérêt à laquelle elle est liée. Il est bien connu qu'on peut alors observer, pour cette variable, des effets intersujets et intrasujets différents. On sait moins que les estimateurs obtenus en supposant différentes structures de covariance peuvent converger vers des valeurs différentes. Nous utilisons des données réelles et simulées pour illustrer ce phénomène. Nous envisageons comment il peut être utilisé pour déceler l'omission d'une variable de confusion.

[MS-210]
The Stochastic Process Framework for Causal Inference
Inférence causale basée sur les processus stochastiques

Daniel Commenges, *INSERM*

We present a general dynamic model as a framework for possible causal interpretation. We first state a criterion of local independence in terms of measurability of processes involved in the Doob – Meyer decomposition; then we define direct and indirect influence. This framework makes it possible to link descriptive and explicative statistical models, and encompasses quantitative processes and events. We give a dynamic representation of a conventional joint model for HIV load and CD4 counts. We show its inadequacy in capturing causal influences while, on the contrary, known mechanisms of HIV infection can be expressed through a system of differential equations.

Nous présentons un modèle dynamique général comme support à des interprétations causales. Nous donnons un critère d'indépendance locale en terme de mesurabilité des processus impliqués dans la décomposition de Doob-Meyer. Puis nous définissons l'influence directe et indirecte. Ce cadre permet de faire le lien entre modèles descriptifs et explicatifs, et peut être appliqué aux événements aussi bien qu'aux marqueurs quantitatifs. Nous donnons une représentation dynamique d'un modèle conjoint classique de la charge virale et du nombre de lymphocytes CD4. Nous montrons que le modèle classique ne représente pas bien les influences causales qui peuvent être exprimées dans un modèle basé sur un système d'équations différentielles.

[MS-211]
Joint Regression Analysis of Correlated Data Using Gaussian Copulas
Analyse de régression conjointe de données corrélées au moyen de copules gaussiennes

Peter Song, *University of Michigan*, Mingyao Li, *University of Pennsylvania*, Ying Yuan, *University of Texas MD Anderson Cancer Center*

This talk concerns a new joint modeling approach for correlated data analysis. Utilizing Gaussian copulas, we present a unified and flexible machinery to integrate separate one-dimensional generalized linear models (GLMs) into a joint regression analysis of continuous, discrete and mixed correlated outcomes. The availability of joint probability models enables us to develop a full maximum likelihood inference. Numerical illustrations are focused on regression models for discrete correlated data, including multi-dimensional logistic regression models and a joint model for mixed normal and binary outcomes. Simulation studies and real-world data examples are considered in the illustration.

Cet exposé traite d'une nouvelle approche de modélisation conjointe pour l'analyse de données corrélées. En utilisant des copules gaussiennes, nous présentons une machinerie unifiée et flexible pour intégrer des modèles linéaires généralisés (MLGs) unidimensionnels séparés en une analyse de régression conjointe de résultats corrélés discrets, continus et mixtes. La disponibilité de modèles de probabilités conjointes nous permet de développer une inférence par maximum de vraisemblance complète. Nous présentons des illustrations numériques centrées sur les modèles de régression pour données discrètes corrélées, incluant les modèles de régression logistique multidimensionnels et un modèle conjoint pour données mixtes normales et binaires. Des études de simulation et des exemples utilisant des données réelles sont considérés dans l'illustration.

Session 10B Mar. 27 mai • Tues. May 27, 16:30 – 18:00 Capital Hall 2A

Session in Honour of Denis Bosq
Session en hommage à Denis Bosq

[MS-212]
Nonparametric Sequential Prediction of Time Series
Prévision séquentielle non paramétrique des séries temporelles

Gérard Biau, *Université Paris VI,* Kevin Bleakley, *Université Montpellier II,* Laszlo Györfi & György Ottucsak, *Budapest University of Technology and Economics*

Time series prediction covers a vast field of everyday statistical applications in medical, environmental and economic domains. In this talk, we develop nonparametric prediction strategies based on the combination of a set of "experts" and show the universal consistency of these strategies under a minimum of conditions. We perform an in-depth analysis of real-world data sets and show that these nonparametric strategies are more flexible, faster and generally outperform ARMA methods in terms of normalized cumulative prediction errors.

La prévision des séries temporelles revêt une importance toute particulière dans les applications statistiques relatives aux domaines médicaux, environnementaux ou économiques. Dans cette conférence, nous présentons des stratégies de prévision non paramétriques basées sur la combinaison d'un ensemble d'« experts » élémentaires et établissons leur convergence universelle sous des hypothèses minimales. Nous illustrons l'ensemble de nos résultats en analysant en détail des jeux de données réels, qui mettent en évidence le fait que ces stratégies non-paramétriques sont plus flexibles, plus rapides et meilleures (en termes d'erreur de prédiction cumulée normalisée) que les traditionnelles méthodes ARMA.

[MS-213]
Sampling Schemes for Regression Estimation and Applications
Plans d'échantillonnages pour l'estimation de la régression et applications

Delphine Blanke, *Université Pierre et Marie Curie - Paris 6*

For continuous-time processes observed at high frequency, we give thresholds (that one has to left between two successive observations) over which nonparametric kernel regression estimators still reach the i.i.d. rate. These thresholds depend on both the dimension and regularity of the underlying sample paths. The results obtained are applied to nonparametric prediction and interpolation of missing values.

Pour des processus à temps continu, observés à haute fréquence, nous donnons les seuils (devant séparer deux observations consécutives) au-dessus desquels les estimateurs à noyau de la régression atteignent la vitesse du cas i.i.d. Ces seuils dépendent à la fois de la dimension et de la régularité des trajectoires sous-jacentes. Les résultats obtenus sont appliqués à la prévision non paramétrique et à l'interpolation de valeurs manquantes.

[MS-214]
Approximation of Functional Regression Models using Bivariate Splines
Approximation de régression linéaire fonctionnelle avec splines bivariées

Serge Guillas, *University College London*, Ming-Jun Lai, *University of Georgia*

We consider the functional linear regression model where the explanatory variable is a random surface and the response is a real random variable, with

Nous considérons le modèle de régression linéaire fonctionnelle où la variable explicative est une surface aléatoire. La réponse est une variable aléatoire réelle, avec bruit borné

bounded or normal noise. Bivariate splines over triangulations represent the random surfaces. We use this representation to construct least squares estimators of the regression function with or without a penalization term. Under the assumptions that the regressors in the sample are bounded and span a large enough space of functions, bivariate splines approximation properties yield the consistency of the estimators. Simulations illustrate the quality of the asymptotic properties on a realistic domain.

ou normal. Des splines bivariées sur une triangulation représentent les surfaces aléatoires. Nous utilisons cette représentation pour construire des estimateurs des moindres carrés de la fonction de régression avec ou sans terme de pénalisation. Sous l'hypothèse selon laquelle les variables de l'échantillon sont bornées et couvrent un espace assez grand de fonctions, des propriétés d'approximation de splines bivariées fournissent la convergence des estimateurs. Des simulations illustrent la qualité des propriétés asymptotiques sur un domaine réaliste.

Session 10C Mar. 27 mai • Tues. May 27, 16:30 – 18:00 Capital Hall 3A

Longitudinal Data
Données longitudinales

[MS-215]

Simulation of Mixed Longitudinal Data Measured at Distinct Timepoints
Simulation de données longitudinales mixtes mesurées à des points distincts dans le temps

Xiao qin Xiong & Joel Dubin, *University of Waterloo*

Traditional models cannot be applied to longitudinal datasets where the response is not measured at the same timepoints as time-dependent covariates. A binning approach is proposed to smooth this data before performing any modelling. To explore the properties and evaluate the method of binning, we conducted a simulation study. We will discuss our findings, as well as the challenges of conducting a simulation of this type of longitudinal dataset.

Keywords: longitudinal, binning, smoothing, simulation

Les modèles classiques ne peuvent pas être appliqués aux ensembles de données longitudinales dans lesquels la réponse n'est pas mesurée aux mêmes points dans le temps que les covariables variant avec le temps. Nous proposons une approche de groupement par classe (binning) pour lisser ces données avant de procéder à toute modélisation. Pour explorer les propriétés et évaluer la méthode de groupement par classe, nous avons réalisé une étude par simulation. Nous discuterons de nos résultats, ainsi que des défis que pose une simulation de ce type d'ensemble de données longitudinales.

Mots clés : longitudinale, groupement par classe, lissage, simulation

[MS-216]

Selection Biases in Repeated Measures Analysis of Treatment Effects for Recurring Clinical Events
Biais de sélection dans les analyses de mesures répétées des effets des traitements pour les évènements cliniques récurrents

Eleanor Pullenayegum, *McMaster University*, Richard Cook, *Waterloo University*

Many chronic disease processes feature acute episodic conditions which warrant therapeutic intervention to alleviate symptoms or to reduce the risk of further complications. In randomized clinical trials, patients suffering from such diseases are often randomized to one of several treatments and followed over a fixed period, during which any episodes are treated with the assigned treatment.

De nombreux processus morbides chroniques présentent des conditions aiguës épisodiques qui justifient une intervention thérapeutique afin de soulager les symptômes ou de réduire le risque d'autres complications. Dans les essais cliniques randomisés, les patients souffrant de ce genre de maladies sont souvent affectés aléatoirement à l'un de plusieurs traitements et suivis au cours d'une période fixe durant laquelle tout épisode est traité conformément au traitement assigné.

The data structure is then similar to that of longitudinal data arising from studies with prescheduled follow-up assessments, and analyses are often based on corresponding methodology. We argue that this is inappropriate in general, and only rarely are such analyses justified.

La structure des données est alors semblable à celle de données longitudinales provenant d'études comportant des évaluations de suivi planifiées, et les analyses sont souvent basées sur la méthodologie correspondante. Nous soutenons que cette approche est généralement inappropriée, et que ce genre d'analyses n'est que rarement justifié.

[MS-217]
Modeling Longitudinal Data with Application to the Multicenter Aids Cohort Study
Modélisation de données longitudinales avec application à la Multicenter Aids Cohort Study

Cyntha Struthers & Don McLeish, *University of Waterloo*

Motivated by an example in the book "Analysis of Longitudinal Data" by Diggle, Heagerty, Liang and Zeger, we analyze longitudinal data from the 2002 release of the Multicenter Aids Cohort Study. The response (CD4 count for seroconverters) is fitted using a stochastic process with parameters which are functions of time-dependent covariates. There are extensive missing values in these data which are partially accommodated in our analysis. The parameters are estimated using both Bayesian and frequentist methods. We obtain the asymptotic covariance matrix of the estimators and Bayesian credible regions for the parameters, the latter by Gibbs sampling.

Motivés par un exemple mentionné dans le livre intitulé « Analysis of Longitudinal Data » publié par Diggle, Heagerty, Liang et Zeger, nous avons analysé des données longitudinales provenant de la diffusion de 2002 des données de la Multicenter Aids Cohort Study. La réponse (numération des CD4 chez les sujets ayant fait une séroconversion) est ajustée par un processus stochastique dont les paramètres sont des fonctions de covariables variant en fonction du temps. Ces données présentent un grand nombre de valeurs manquantes dont il est partiellement tenu compte dans notre analyse. Les paramètres sont estimés selon des méthodes bayésiennes ainsi que fréquentistes. Nous obtenons la matrice de covariance asymptotique des estimateurs et les régions de crédibilité bayésiennes, ces dernières par échantillonnage de Gibbs.

[MS-218]
Model-Based Clustering of Longitudinal Data
Classification à base de modèles de données longitudinales

Paul McNicholas, *University of Guelph*, Thomas Brendan Murphy, *University College Dublin*

Model-based clustering is a method of clustering data based on mixture modeling. A Gaussian mixture model is frequently used and a common approach exploits an eigenvalue decomposition of the group covariance matrices to give a variety of covariance structures. These ideas are extended to facilitate application to longitudinal data via a modified Cholesky decomposition of the covariance matrices. For each mixture component, a variety of constraints are imposed to give a family of mixture models. This new family of mixture models is applied to data on the weight of rats on one of three different diets, giving excellent cluster-capturing performance.

La classification à base de modèles est une méthode de classification de données basées sur des modèles de mélanges. Un modèle de mélange gaussien est souvent utilisé et une approche commune exploite la décomposition des matrices de covariance du groupe en valeurs propres pour donner une variété de structures de covariance. Ces idées sont élargies pour en faciliter l'application à des données longitudinales au moyen d'une décomposition de Cholesky modifiée des matrices de covariance. Pour chaque composante du mélange, une variété de contraintes sont imposées pour donner une famille de modèles de mélanges. Cette nouvelle famille de mélanges est appliquée à des données de poids de rats suivant trois diètes différentes, donnant d'excellents résultats.

[MS-219]

Criteria for Longitudinal Data Model Selection Based on Kullback's Symmetric Divergence
Critères de sélection de modèles de données longitudinales fondés sur la divergence symétrique de Kullback

Bezza Hafidi, *Université Ibn Zohr, Agadir, Maroc*, Nourddine Azzaoui, *Université de Lille 1*

Longitudinal data analysis is widely exploited in clinical trials, epidemiology, agriculture... Recently, Azari et al. (2006) showed that (AIC) criterion and its corrected versions cannot be applied to model selection for longitudinal data with correlated errors. They proposed two selection criteria, AICc and RICc, by applying likelihood and residual likelihood approaches. These two criteria are estimators of the Kullback – Leibler's divergence which is asymmetric. In this work, we apply the likelihood and residual likelihood approaches to propose new criteria, suitable for small samples longitudinal data, based on the Kullback's symmetric divergence. Their performance is examined in a large simulation study.

L'analyse des données longitudinales est d'une grande utilité dans les domaines des essais cliniques, d'épidémiologie, d'agriculture, etc. Récemment, Azari *et al*. (2006) ont montré que le critère (AIC) ainsi que ses versions corrigées ne peuvent pas être directement appliqués aux données longitudinales avec des erreurs corrélées. Ils ont proposé deux critères, AICc et RICc, en utilisant la notion de vraisemblance et de vraisemblance résiduelle. Leurs critères sont des estimations de la distance asymétrique de divergence de Kullback-Leibler. Dans ce travail, nous proposons deux nouveaux critères adaptés aux échantillons de petites tailles de données longitudinales en se basant sur la divergence symétrique de Kulback et les approches de maximum de vraisemblance et de vraisemblance résiduelle. Les performances de ces critères sont examinées dans une étude de simulations.

Session 10D Mar. 27 mai • Tues. May 27, 16:30 – 18:00 Capital Hall 4A

Regression Methods in Biostatistics
Méthodes de régression en biostatistique

[MS-220]

Variable Selection Using Branching LARS
Sélection des variables par une méthode LARS ramifiée

Li Hua Yue, Duncan Murdoch, Wenqing He & Hristo S. Sendov, *University of Western Ontario*

Identification of cost-efficient diagnostic factors is of interest to health researchers. The modern diagnostic technique may involve group costs: including one factor which invokes the cost of the whole group of factors. The trade-off between statistical significance and the cost of collecting data for the statistical model is our focus. We extend the Least Angle Regression (LARS) variable selection method to incorporate various costs of factors. A branch and bound search method combined with LARS will be employed to select cost-efficient factors.

L'identification de facteurs de diagnostic rentables présente un intérêt pour les chercheurs du domaine de la santé. La technique moderne de diagnostic pourrait comporter un coût de groupe : inclure un facteur appelle le coût du groupe complet de facteurs. Nous nous intéressons au compromis entre la signification statistique et le coût de la collecte des données pour le modèle statistique. Nous étendons la méthode fondée sur l'algorithme Least Angle Regression (LARS) de sélection des variables afin d'y intégrer divers coûts des facteurs. Nous employons une méthode de recherche ramifiée et bornée combinée à l'algorithme LARS pour choisir des facteurs rentables.

[MS-221]

Reducing Bias in Observational Studies using Propensity Score Methods
Réduire le biais d'études observationnelles par des méthodes de scores de propension

Binod Neupane, Lehana Thabane & Mark Loeb, *McMaster University*

Multivariable regression adjustments and propensity score methods are two popular methods of reducing biases in the results of observational studies. Some theoretical and simulative studies suggest that the propensity score methods outperform multivariable regression adjustments in reducing biases under variety of situations. Also, some carefully conducted propensity score analyses confirm the results from randomized experiments. In this talk, we will present a review of arguments about how a propensity score model should be specified for unbiased analysis and review of different propensity score techniques and comparisons of propensity score methods with multivariable regression techniques in reducing biases.

Les ajustements par régression multivariable et les méthodes de scores de propension sont deux méthodes populaires de réduction de biais des résultats d'études observationnelles. Certaines études théoriques et de simulation suggèrent que les méthodes de scores de propension donnent de meilleurs résultats que les ajustements par régression multivariable dans la réduction de biais sous diverses conditions. Aussi, certaines analyses de scores de propension soigneusement conduites confirment les résultats d'expériences aléatoires. Dans cet exposé, nous présentons comment un modèle de scores de propension devrait être spécifié pour une analyse sans biais et révisons différentes techniques de scores de propension et comparons les méthodes de scores de propensions aux techniques de régression multivariable en ce qui a trait à la réduction du biais.

[MS-222]

Estimation for Age-Period-Cohort Models: With Application to the Mesothelioma Data in Alberta
Estimation pour modèles âge-période-cohorte : application aux données de mésothéliome en Alberta

Bei Jiang & K. C. Carriere, *University of Alberta*

Age-Period-Cohort (APC) models have been widely used by epidemiologists to analyze the trends of disease incidence and mortality. The linear dependency: "Cohort = Period – Age" results in the well-known "non-identifiability" problem. We review the "solutions" that are built on estimable functions (Holford, 1983). We present how restricted regression splines can be used to extend the classical framework of factorial APC models (Heuer, 1997). Further, we consider how P-splines can be applied to smooth the APC model and how the "non-identifiability" problem can be dealt with accordingly (Currie et al., 2006). Finally, these methods are illustrated to the mesothelioma data in Alberta.

Les modèles Âge-Période-Cohorte (APC) ont été largement utilisés par les épidémiologistes pour analyser les tendances de l'incidence et de la mortalité de maladies. La dépendance linéaire « Cohorte = Période – Âge » a pour résultat le célèbre problème de « non-identifiabilité ». Nous révisons les "solutions" fondées sur des fonctions estimables (Holford, 1983). Nous présentons comment les splines de régression restreintes peuvent être utilisées afin d'élargir le cadre traditionnel des modèles factoriels APC (Heuer 1997). De plus, nous considérons comment les P-splines peuvent être appliquées pour lisser le modèle APC et comment le problème de « non-identifiabilité » peut alors être traité (Currie *et al.*, 2006). Finalement, ces méthodes sont illustrées à l'aide de données de mésothéliome en Alberta.

[MS-223]

Approaches in Dealing with Heterogeneity in Meta-analysis Studies
Approches pour composer avec l'hétérogénéité dans les études de méta-analyse

Jennifer Liu, Lehana Thabane, Emmy Cheng & Roman Jaeschke, *McMaster University*

Heterogeneity in meta-analysis of clinical studies can be approached in many ways including ex-

Le problème de l'hétérogénéité dans la méta-analyse des études cliniques peut être abordé de nombreuses façons, dont

plaining it through meta-regression or subgroup analysis. We use data from a systematic review of the medical literature to show that many investigators test heterogeneity first and explain subsequently using meta-regression or subgroup analysis. In our presentation, we offer arguments against this approach and suggest reasons why heterogeneity should be explained first based on a priori hypotheses. We illustrate the approach using data from a systematic review of trials to assess the safety of long acting beta agonists among patients with asthma using inhaled corticosteroids.

son explication par la méta-régression ou par l'analyse de sous-groupes. Nous utilisons des données provenant d'une revue systématique de la littérature médicale pour montrer que de nombreux chercheurs testent d'abord l'hétérogénéité, puis l'explique en utilisant la méta-régression ou l'analyse de sous-groupes. Dans la communication, nous présentons des arguments à l'encontre de cette approche et avançons les raisons pour lesquelles l'hétérogénéité devrait être expliquée en premier lieu, en se basant sur des hypothèses a priori. Nous illustrons l'approche à l'aide de données provenant d'une revue systématique des essais destinés à évaluer la sécurité des bêta-agonistes à effet prolongé chez les asthmatiques utilisant des corticostéroïdes inhalés.

[MS-224]

Longitudinal Bent-Cables for Monitoring Atmospheric CFC Decline and Other Change Models
Cordes pliées longitudinales pour surveiller la baisse des CFC atmosphériques et autres modèles de changement

Shahedul Khan & Grace Chiu, *University of Waterloo*

The classic bent-cable regression models the gradual transition of a single profile by comprising two linear segments, joined by a quadratic bend. For longitudinal data, each profile exhibits a different random cable, but all profiles come from the same population cable. For example, the abundances of the chlorofluorocarbons in the atmosphere have been monitored from numerous monitoring stations around the globe, and each station has its own bent-cable coefficients. This motivated us to extend the classic bent-cable regression in the framework of random-effects model. The model is illustrated by simulation results and with an application to the chlorofluorocarbons data.

La régression à cordes pliées classique modélise la transition graduelle d'un profil simple en englobant deux segments linéaires, joints par un coude quadratique. Pour des données longitudinales, chaque profil exprime une corde aléatoire distincte, mais tous les profils proviennent de la même corde de la population. Par exemple, les concentrations de chlorofluorocarbones dans l'atmosphère ont été surveillées par plusieurs stations d'observations autour du globe, et chaque station a ses propres coefficients de cordes repliées. Ceci nous a motivés à étendre la régression à cordes repliées classique dans le cadre des modèles à effets aléatoires. Le modèle est illustré au moyen de résultats de simulation et d'une application aux données de chlorofluorocarbone.

[MS-225]

Supervised Classification and Free Knot Splines for Biomedical Data
Classification supervisée et spline à nœuds variables pour des données biomédicale

Nicolas Molinari, *Université Montpellier 1 - CHU de Nîmes*

Data in many different fields come to practitioners through a process naturally described as functional. We propose a classification procedure of oxidation curves. Our algorithm is based on two stages: fitting the functional data by linear splines with free knots and classifing the estimated knots which estimate useful oxidation parmeters (the lag time and the decomposition time). A real data set on 57 oxidation curves is used to illustrate our approach.

Les données fonctionnelles sont couramment rencontrées dans les applications. Nous proposons une approche de classification de courbes d'oxydation. L'algorithme est décomposé en deux étapes : ajustement des données fonctionnelles par des fonctions splines linéaires à nœuds variables et classification. La position des nœuds peut être interprétée le temps de latence et de décomposition de l'oxydation. Un jeu de données réelles permettra d'illustrer l'approche.

Session 10E Mar. 27 mai • Tues. May 27, 16:30 – 18:00 Capital Hall 5A

Copulas and Applications
Copules et applications

[MS-226]
Nonparametric Tests for Conditional Independence via Copula
Tests d'indépendance conditionelle non paramétriques au moyen de copules

Taoufik Bouezmarni & Jeroen Rombouts, *HEC-Montréal*

This paper proposes nonparametric tests for conditional independence and particularly for Granger non-causality. The proposed test is based on the comparison of copula densities using Hellinger distance. The Bernstein density copula is used to estimated the unknown copulas. For β-mixing data, we investigate the asymptotic proprieties of the Bernstein density copula: we give its asymptotic bias and variance, the uniform strong consistency and asymptotic normality of Bernstein copula. We show the asymptotic normality of the conditional independence test statistic. A detailed simulation study investigates the performance of the test. Applications using financial data are provided.

Cet article propose des tests non paramétriques d'indépendance conditionnelle et, plus particulièrement, de non-causalité de Granger. Le test proposé est basé sur la comparaison de densités de copules au moyen de la distance d'Hellinger. La copule de densité de Bernstein est utilisée pour estimer les copules inconnues. Pour des données de β-mélange, nous étudions les propriétés asymptotiques de la copule de densité de Bernstein : nous donnons son biais et sa variance asymptotique, et montrons sa cohérence uniforme forte et sa normalité asymptotique. Nous montrons la normalité asymptotique de la statistique du test d'indépendance conditionnelle. Une étude de simulation détaillée se penche sur la performance du test. Nous présentons des applications utilisant des données financières.

[MS-227]
Predicting Dependent Binary Outcomes through Logistic Regressions and Copulas
Prévision de variables binaires dépendantes à l'aide de régressions logistiques et de copules

Aristidis Nikoloulopoulos, *University of British Columbia*, Mathieu Fortin, *Ministère des ressources naturelles et de la faune du Québec*, Christian Genest & Louis-Paul Rivest, *Université Laval*

Gauvreau & Pagano (1997) showed how to model dependent binary variables by using a Farlie – Gumbel – Morgenstern copula to combine logistic regressions for the margins. An extension of this approach will be considered. First, a copula-specific score statistic will be introduced for testing whether the outcomes are conditionally independent. A composite likelihood will then be proposed for estimation and prediction. As will be seen, elliptical copulas are convenient in this context: besides allowing a wide range of dependence, they substantially reduce the computational burden associated with the IFM estimation method of Joe (2005). Forestry data will be used to illustrate the approach.

Gauvreau & Pagano (1997) ont montré comment modéliser des variables binaires dépendantes en se servant d'une copule de Farlie-Gumbel-Morgenstern pour combiner les régressions logistiques des marges. Une généralisation de cette approche sera présentée. On proposera d'abord un test d'indépendance conditionnelle des variables binaires fondé sur une statistique score spécifique à la copule. On suggérera ensuite une vraisemblance composite aux fins d'estimation et de prévision. Comme on le verra, les copules elliptiques sont bien adaptées à ce contexte : en plus d'offrir une large gamme de dépendances, elles réduisent substantiellement l'effort de calcul lié à l'implantation de la méthode d'estimation IFM de Joe (2005). Une application en foresterie illustrera le propos.

[MS-228]

Bayesian Inference of a Copula Model for a Bivariate Hydrological Analysis
Estimation bayésienne d'un modèle de copule pour une analyse hydrologique fréquentielle bivariée

Eric Parent, *Équipe Modélisation, Risque, Statistique, Environnement (MORSE), ENGREF, AgroParisTech, Paris,*
Anne-Catherine Favre, *INRS-ETE, Québec*

Large spring floods in the Quebec region exhibit correlated peakflows, duration and volume. Consequently, traditional univariate hydrological analyses are unable to provide a meaningful design flood level as requested in hydraulogical engineering. We study 47 years of a peak/volume dataset for the Romaine River with a parametric copula model. The margins are modelled with a norma (or gamma) two component mixture and the dependance is depicted through a parametric family of copulas (Clayton or Arch 12). Parameter joint inference and model selection is performed under the Bayesian paradigm.

Au Québec, les grandes crues de printemps sont caractérisées par la pointe, le volume et la durée. Ces variables étant corrélées, les analyses fréquentielles univariées traditionnelles ne sont pas en mesure de fournir une évaluation complète pour une crue de projet, utile en ingénierie hydraulique. Nous analysons 47 années de couples pointe-volume de crues de la rivière Romaine à l'aide d'un modèle paramétrique de copule. Les marges sont modélisées par un mélange (gamma ou normal), et la dépendance par une famille paramétrique de copules (Arch12 et Clayton). L'inférence et la sélection de modèles sont alors réalisées sous la perspective bayésienne.

[MS-229]

A Bayesian Estimator for the Dependence Function of a Bivariate Extreme Value Distribution
Un estimateur bayésien pour la fonction de dépendance d'une loi bivariée de valeurs extrêmes

Simon Guillotte & François Perron, *Université de Montréal*

A bivariate distribution with continuous margins can be decomposed via a copula and its marginal distributions. In the case of extreme value distributions, the copula is characterized by a dependence function while each of the margins depends on three parameters. In this talk, we propose a Bayesian approach for the simultaneous estimation of the dependence function and the parameters defining the marginal distributions. A nonparametric model is constructed for the dependence function and a reversible jump MCMC algorithm is proposed for numerical evaluations of the Bayesian estimator. We illustrate our method on a hydrological data set and make predictions.

Toute distribution bivariée dont les marges sont continues peut être décomposée via sa copule et ses distributions marginales. Dans le cas des distributions de valeurs extrêmes, la copule est caractérisée par une fonction de dépendance tandis que chacune des marges est définie par trois paramètres. Dans cet exposé, nous proposons une approche bayésienne pour l'estimation simultanée de la fonction de dépendance et des lois marginales. Un modèle non paramétrique est construit pour la fonction de dépendance et un algorithme MCMC à sauts réversibles pour le calcul de l'estimateur bayésien est proposé. Nous illustrerons l'utilité de notre approche en l'appliquant sur un jeu de données hydrologiques pour faire des prévisions.

[MS-230]

Goodness-of-Fit Tests for Copula Models of Bivariate Lifetime Data
Tests d'ajustement pour modèles de copules de données de survie bivariées

Yildiz Yilmaz, *University of Waterloo*

We discuss the fit of a copula family for bivariate censored data with covariates. The adequacy of the proposed copula family is checked by embedding it in an expanded family of copulas. When

Nous discutons de l'ajustement d'une famille de copules pour des données bivariées censurées avec covariables. L'adéquation de la famille de copules proposée est vérifiée en l'emboîtant dans une famille de copules élargie. Lorsque

the proposed and the expanded copula models are estimated by maximum likelihood estimation, the likelihood ratio test is used. However, when they are estimated by two-stage estimation technique, model checking is performed by a pseudo likelihood ratio test. Power comparisons of the likelihood ratio and the pseudo likelihood ratio tests with some other existing goodness-of-fit tests are performed. The methods are considered in real life applications.

les familles proposées et élargies sont estimées par la methode du maximum de vraisemblance, le test du rapport de vraisemblance est utilisé. Cependant, lorsqu'elles sont estimées par une méthode à deux étapes, l'évaluation du modèle est faite au moyen d'un test du pseudo rapport de vraisemblance. Nous effectuons la comparaison des puissances du test du rapport de vraisemblance, du test du pseudo rapport de vraisemblance et de d'autres tests d'ajustement existants. Les méthodes sont considérées dans des applications réelles.

Session 10F Mar. 27 mai • Tues. May 27, 16:30 – 18:00 Capital Hall 6A

Limiting Behaviour of Test Statistics
Comportement asymptotique de statistiques de test

[MS-231]
The Asymptotically Optimal Estimating Equation for Longitudinal Data
L'equation estimatrice asymptotiquement optimale pour les données longitudinales

Laura Dumitrescu & Raluca Balan, *University of Ottawa*, Ioana Schiopu-Kratina, *Statistics Canada*

In this talk, we introduce a marginal model for longitudinal data, in which the residuals form a martingale difference sequence. This model allows us to consider a class of estimating equations, which contains the GEE proposed by Liang and Zeger (1986). Using the approach of Heyde (1997), we identify a sequence of "asymptotically optimal" equations within this class. Next, we prove the existence of a sequence of strongly consistent estimators for the regression parameter, defined as roots of the asymptotically optimal equations.

Dans cette présentation, nous introduisons un modèle marginal pour les données longitudinales dans lesquelles les résidus forment une séquence de différence de martingale. Ce modèle nous permet de considérer une classe d'équations d'estimation lesquelles contiennent l'EEG (équations d'estimation généralisées) proposé par Liang et Zeger (1986). En utilisant l'approche de Heyde (1987), nous identifions une séquence d'équations « asymptotiquement optimale » à l'intérieur de cette classe. Ensuite, nous prouvons l'existence d'une séquence d'estimateurs solidement cohérents pour le paramètre de régression, définis comme les racines des équations asymptotiquement optimales.

[MS-232]
New Multivariate Central Limit Theorems in Linear Structural and Functional Error-in-Variables Models
De nouveaux théorèmes centraux limite multivariés pour des modèles linéaires structurels et fonctionnels à erreur dans les variables

Yuliya Martsynyuk, *Carleton University*

We deal simultaneously with linear structural and functional error-in-variables models (SEIVM and FEIVM). New joint CLTs are established for modified least squares estimators of the slope and intercept under some first time, so far the most general, model conditions. Due to them being in Studentized forms to begin with, the obtained CLTs, unlike related ones in the literature, are a priori

Nous traitons simultanément avec des modèles linéaires structuraux et fonctionnels à erreurs dans les variables (MLSEDV et MFEDV). De nouveaux TLC sont établis pour les estimateurs des moindres carrés modifiés de la pente et de l'ordonnée à l'origine sous des conditions jusqu'ici les plus générales sur les modèles. Étant donnée leur forme déjà studentisée, les TLC obtenus, contrairement à des théorèmes similaires décrits dans la littérature, sont *a priori* ba-

data-based and free of any unknown parameters associated with the explanatory and error variables. The very forms of our CLTs are universal for the SEIVM and FEIVM. This extends a previously known interplay between a SEIVM and a FEIVM.

sés sur les données et libres de tous paramètres inconnus associés aux variables explicatives et aux variables d'erreur. La forme même de nos TLC est universelle pour les MLSEDV et MFEDV. Ceci élargit une interaction connue entre un MLSEDV et un MFEDV.

[MS-233]

On the Asymptotic Properties of the Signed Composite Likelihood Ratio Statistic
Des propriétés asymptotiques de la statistique du ratio de vraisemblance composite signée

Zi Jin & Nancy Reid, *University of Toronto*

Composite likelihood has received increased attention in handling large dimensional data sets when the computation of full likelihood is highly complex. In particular, we concentrate on marginal composite likelihood, which arises by considering marginal densities. Under the scenario where consistency and normality of composite likelihood estimates are satisfied, we summarize different composite test statistics with their asymptotic distributions, and outline higher order asymptotic properties of the signed composite likelihood ratio statistic. Some examples are analyzed in detail, and simulation studies are presented to further explore the efficiency of composite likelihood and asymptotic performance of the signed composite likelihood ratio statistic.

La vraisemblance composite a reçu une attention accrue pour soutenir des données de grande dimensionnalité, lorsque le calcul de la vraisemblance totale est hautement complexe. En particulier, nous nous concentrons sur la vraisemblance marginale composite qui apparaît lorsque l'on considère la densité marginale. Assumant la consistance et la normalité de la vraisemblance composite, nous énumérons différentes statistiques de test composite avec leur distribution asymptotique et décrivons les propriétés d'ordre supérieur de la statistique du ratio de vraisemblance composite. Plusieurs exemples sont analysés en détail et des simulations sont présentées en vue d'explorer l'efficience de la vraisemblance composite ainsi que la performance asymptotique de la statistique du ratio de vraisemblance signée.

[MS-234]

Asymptotic Normality of the Quasi-Maximum Likelihood Estimator for Affine Models
Normalité asymptotique de l'Estimateur du Quasi Maximum de Vraisemblance pour les modèles affines

Olivier Wintenberger & Jean-Marc Bardet, *Université Paris 1*

The strong consistency and asymptotic normality of the Quasi-Maximum Likelihood Estimator are proven in the general framework of multidimensional affine models. The conditions are often better than those known for classic models (for example, Garch models, Arch, ARMA-Garch, one-dimensional or otherwise). The results are also applicable to new models, and or extensions of classic models.

La consistance forte et la normalité asymptotique de l'estimateur du quasi-maximum de vraisemblance sont prouvées dans le cadre général des modèles affines multidimensionnels. Les conditions sont souvent meilleures que celles connues pour les modèles classiques (par exemple les modèles GARCH, ARCH, ARMA-GARCH unidimensionnels ou non). Les résultats sont aussi applicables à de nouveaux modèles, extensions des modèles classiques.

Session 10G Mar. 27 mai • Tues. May 27, 16:30 – 18:00 Capital Hall 7A

Survey Methods
Méthodes d'échantillonnage

[MS-235]
Free Model for Generalized Path Modelling
Modèle libre dans le cadre des modèles à équations structurelles généralisés

Christian Derquenne, *EDF R&D*

This paper introduces a new method to build a graphical model of categorical variables (a free model) in frame of structural equation models. The process is the following: clustering of variables, calculation of latent variable by cluster, searching for statistical links between latent variables and expert decision to position these previous links. Finally, this free model is estimated by the Regression on First Generalized Principal Components approach (Derquenne, 2005). This new method is applied to real data in marketing, listing the advantages and drawbacks of this approach with respect to others, notably the Bayesian networks. Lastly, we discuss further potential research.

Ce papier introduit une nouvelle méthode pour construire un modèle graphique (modèle libre) sur variables catégorielles dans le cadre des modèles à équations structurelles. La démarche est la suivante : classification des variables, calcul d'une variable latente par classe, recherche de liens statistiques entre variables latentes, décision de l'expert du domaine d'application d'orienter ces liens. Enfin, ce modèle libre est estimé avec l'approche Régression sur premières composantes principales généralisées (Derquenne, 2005). Cette méthode a été appliquée en marketing, puis les forces et les faiblesses sont établies par rapport à un réseau bayesien et des voies potentielles de recherches sont discutées.

[MS-236]
General Framework for the Rotation of Units in Repeated Survey Sampling
Cadre général pour la rotation des individus dans les enquêtes répétées

Lionel Qualité, Desislava Nedyalkova & Yves Tillé, *Université de Neuchâtel*

The negative coordination of samples is one of the main theoretical problems faced by statistical institutes. There is a collection of existing solutions that try to attend to it. Our aim is to provide the core of a theory that includes the main existing sampling designs. We review the main existing negative coordination methods: the Brewer method, the method of permutation of random numbers, and the burden method. We show that it is possible to compute the cross-sectional and longitudinal sampling designs and in some cases even the joint sampling design. Finally, we give a general method that allows us to apply almost any cross-sectional design with a given longitudinal systematic design.

La coordination négative d'échantillons est un problème important pour les instituts de Statistique. Il existe une gamme de méthodes qui tentent de le résoudre. Notre but est de réunir ces méthodes et de les décrire dans un cadre unifié. Nous présentons les principales méthodes de coordination négative existantes : la méthode de Brewer, la méthode des numéros aléatoires permanents et la méthode de charges. Nous montrons qu'il est possible de calculer les plans transversaux et longitudinaux associés, et parfois le plan complet. Enfin, nous donnons une méthode qui permet de choisir presque librement les plans transversaux, pour un plan longitudinal systématique donné.

[MS-237]
Some Methodological Issues on Drawing Samples from the New French Census.
Questions méthodologiques sur les tirages d'échantillons dans le nouveau recensement français

Marc Christine & Sébastien Faivre, *Insee-France*

The new rotative Census settled in France leads us to review the building of samples for household surveys conducted by Insee. First, in order to draw samples from the last part of the territory covered by the Census, in fixed areas, it has been necessary to build specific primary units. Then, drawing on those units — some of them devoted to the national Master sample, the others to regional extensions — requires a two-phase method, using both balancing conditions and fixed probabilities of selection. Finally, different kinds of estimators are studied, with impact on weightings and quality of the final samples of dwellings resulting from this process.

Le nouveau recensement rotatif mis en place en France conduit à revoir la constitution des échantillons des enquêtes ménages de l'Insee. D'une part, pour impacter chaque année la fraction recensée la plus récente au sein de zones géographiques fixes, il a fallu construire des unités primaires spécifiques. Ensuite, le tirage de ces unités, articulant l'échantillon-maître national et le complément destiné aux extensions régionales, a nécessité l'élaboration d'une méthode en plusieurs phases, respectant des conditions d'équilibrage et des probabilités d'inclusion fixées. Enfin, les choix d'estimateurs conduisent à revisiter la problématique des pondérations des unités finales (logements) et à vérifier leur qualité.

Session 11A Mer. 28 mai • Wed. May 28, 08:30 – 10:00 Congress Hall A/E

Pierre Robillard Award Addresses
Allocutions des récipiendaires du prix Pierre Robillard

[MS-238]
Bayesian Optimal Design for Changepoint Problems
Plans d'expérience optimaux bayésiens pour des problèmes avec point de rupture

Juli Atherton, *McGill University*

The thesis concerns optimal designs for changepoint problems where there is a possible change in mean at an unknown location. Through the introduction of a design measure, different in structure but similar in spirit to that of Kiefer, very general optimal design results were obtained to estimate the before- and after-change means and to test for a change.

La thèse propose des plans d'expérience optimaux pour des situations où la moyenne d'une loi peut changer en un point indéterminé. L'introduction d'une mesure de plan, différente mais semblable à celle de Kiefer, permet d'obtenir dans des conditions très générales des plans optimaux pour l'estimation de la moyenne avant et après un point de rupture, dont l'existence peut être testée.

[MS-239]
Minimum Hellinger Distance Estimation in Semiparametric Models
Estimation du minimum de la distance d'Hellinger dans les modèles semi-paramétriques

Jingjing Wu, *University of Calgary*

The successful application of minimum Hellinger distance (MHD) approach to fully parametric models is well known. The MHD estimators achieve efficiency at the model density and simultaneously have desirable robustness properties (Beran, 1977).

Il est bien connu qu'on peut appliquer avec succès l'approche du minimum de la distance d'Hellinger (MDH) à des modèles pleinement paramétriques. Les estimateurs MDH sont efficaces à la densité du modèle et présentent en même temps des propriétés de robustesse souhaitables

In this thesis, I extend this approach to semiparametric models. Asymptotic properties such as consistency, asymptotic normality, efficiency and adaptivity of the proposed MHD estimators are investigated. The presence of infinite dimensional nuisance parameters adds an extra degree of complexity to the construction of MHD estimators and to the development of asymptotic results. The robustness and small sample properties are also presented using Monte Carlo studies.

(Beran, 1977). Dans cette thèse, j'étends cette approche aux modèles semi-paramétriques. J'examine les propriétés asymptotiques, comme la cohérence, la normalité asymptotique, l'efficience et l'adaptabilité des estimateurs MDH proposés. La présence de paramètres de nuisance de dimension infinie ajoute une couche supplémentaire de complexité à la confection des estimateurs MDH et à l'obtention de résultats asymptotiques. Nous présentons aussi les propriétés de la robustesse et des petits échantillons à l'aide d'études Monte-Carlo.

Session 12A Mer. 28 mai • Wed. May 28, 10:30 – 12:00 Capital Hall 1A

Round Table: Issues in Collaboration in Medical Research
Table ronde : Questionnements sur la collaboration en recherche médicale

[MS-240]
Round Table: Issues in Collaboration in Medical Research
Table ronde : enjeux de collaboration dans la recherche médicale

Andrew Willan, *University of Toronto*

Background. Effective statistical collaboration in a multidisciplinary health research environment requires skills not taught in theoretical statistics courses. Most statisticians often learn such skills through trial-and-error. The purpose of this session is to provide a forum to discuss the strategies for acquiring the important skills needed for effective collaboration in medical research.
Topics covered in the workshop. The discussions will cover topics that include: Discussing the role of statisticians in collaborative health research Opportunities and challenges How to communicate with non-statisticians: verbal and written communication How to deal with grant proposals Statistics and ethics: case studies of ethical dilemmas faced by many practicing biostatisticians How to negotiate your role as a biostatistical collaborator How to negotiate compensation and acknowledgement for your contributions Mentoring of new biostatisticians How to report statistics in medical research How to report a sample size calculation in a protocol for granting agency or Research Ethics submission, abstract for a scientific meeting, and manuscript for publication How to turn a clinical problem into a researchable question including key outcomes

Pour bien collaborer à l'activité statistique dans un milieu multidisciplinaire de recherche en santé, il faut posséder des compétences que n'offrent pas les cours de statistique théorique. La plupart des statisticiens acquièrent de telles compétences par « tâtonnement ». La présente séance offre aux participants l'occasion de discuter des stratégies à employer pour acquérir les compétences importantes nécessaires à la collaboration efficace en recherche médicale.
Thèmes de l'atelier. Les discussions porteront, notamment, sur les thèmes suivants : Le rôle des statisticiens en recherche coopérative sur la santé Occasions et défis La communication orale et écrite avec de non-statisticiens Les propositions de subventions Statistique et éthique : études de cas de dilemmes éthiques auxquels font face de nombreux biostatisticiens en exercice La négociation du rôle de collaborateur en biostatistique La négociation de la rémunération et de la reconnaissance de son apport Le mentorat des nouveaux statisticiens La présentation de statistiques en recherche médicale La présentation du calcul de la taille d'un échantillon dans un protocole adressé à un organisme subventionnaire, une présentation sur l'éthique en recherche, un résumé destiné à une assemblée scientifique ou un texte à publier La transformation d'un problème clinique en une question qui se prête à la recherche, y compris principaux résultats

Session 12B Mer. 28 mai • Wed. May 28, 10:30 – 12:00 Capital Hall 2A

Monte Carlo Methods
Méthodes Monte-Carlo

[MS-241]
On the Effectiveness of Randomized Quasi-Monte Carlo for Option Pricing
Sur l'efficacité des méthodes quasi Monte-Carlo randomisées pour la fixation des prix

Pierre L'Ecuyer, *Université de Montréal*

Option pricing in finance requires estimation of high-dimensional integrals and this is often done by the Monte Carlo (MC) method. However, randomized quasi-Monte Carlo (RQMC) often outperforms MC in the sense that it can provide a much more accurate estimator with an equivalent amount of work. We will review recent theoretical explanations behind this success, will make the connection with uniformity criteria for selecting the RQMC point sets, and will provide several concrete numerical illustrations showing large performance gains.
xs

L'évaluation d'options en finance requiert l'estimation d'intégrales en plusieurs dimensions. Cette estimation se fait souvent par la méthode Monte-Carlo (MC). Mais les méthodes quasi Monte-Carlo randomisées sont souvent plus performantes que MC dans ce contexte, dans le sens qu'elles fournissent un estimateur sans biais et à plus faible variance pour un budget de calcul équivalent. Nous passerons en revue les explications théoriques récentes de ce succès, ferons le lien avec des critères de sélection des ensembles de points pour RQMC, et donnerons plusieurs exemples numériques où les gains de performance sont très importants.

[MS-242]
Fast Simulated Annealing in R^d and Application to Max-likelihood
Recuit rapide dans R^d et application au max de vraisemblance

Sylvain Rubenthaler, *Université de Nice-Sophia Antipolis*, Tobias Rydén & Magnus Wiktorsson, *Lund University*

We study the convergence of a variation of simulated annealing named fast annealing. This optimisation procedure is well suited to finding the max-likelihood in the case where the likelihood cannot be computed exactly.

Nous étudions ici la convergence d'une variante du recuit simulé appelée recuit rapide. Cette procédure d'optimisation est bien adaptée à la recherche du maximum de vraisemblance dans le cas où la vraisemblance ne peut pas être calculée de manière exacte.

[MS-243]
Sequential Monte Carlo in Climate Change
Méthodes de Monte-Carlo séquentielles pour l'étude des changements climatiques

Michael Kouritzin, *University of Alberta*

Polar ice caps contain Calcium dust evidence of temperatures dating back hundreds of thousands of years. Physicists have developed stochastic models for the evolution of temperature and dust, accounting for yearly random variations. Some popular models utilize heavy-tailed processes to switch between ice age and non-ice age eras. Considering measurement imperfections in the dust component of the model, we obtain a filtering model. However, the classical finite energy condition required for the filtering equations is not satisfied. We ex-

Les calottes polaires contiennent de la poussière de calcium donnant des preuves des températures qui régnaient il y a des centaines de milliers d'années. Les physiciens ont élaboré des modèles stochastiques de l'évolution de la température et de la poussière qui tiennent compte des variations annuelles aléatoires. Certains modèles répandus utilisent des processus à queue lourde pour osciller entre les ères glaciaires et non glaciaires. En tenant compte des imperfections des mesures de la composante poussière du modèle, nous obtenons un modèle de filtration. Cependant, la condition classique d'énergie finie requiert que les équations de filtration

plain how to extend the filtering equations to handle this model and use Sequential Monte Carlo techniques to track temperature.

ne soient pas satisfaites. Nous expliquons comment étendre les équations de filtration afin de traiter ce modèle et utilisons des méthodes de Monte-Carlo séquentielles pour suivre la température.

Session 12C Mer. 28 mai • Wed. May 28, 10:30 – 12:00 Capital Hall 3A

Rank-based Inference
Inférence fondée sur les rangs

[MS-244]

Two Sample Tests for the Nonparametric Behrens – Fisher Problem with Clustered Data
Tests pour deux échantillons pour le problème de Behrens-Fisher non paramétrique avec données corrélées en grappes

Denis Larocque, *HÉC Montréal*, Riina Haataja, Jaakko Nevalainen & Hannu Oja, *Tampere School of Public Health*

In this talk, we consider the nonparametric Behrens – Fisher problem with cluster correlated data. A class of weighted Mann – Whitney test statistics is introduced, studied and compared with existing procedures. The new tests are valid when the distributions do not have the same scales and/or shapes under the null hypothesis.

Dans cette présentation, nous considérons le problème de Behrens-Fisher non paramétrique avec données corrélées en grappes. Nous proposons une classe de tests de Mann-Whitney pondérés et les comparons à des procédures existantes. Les nouveaux tests sont valides en présence d'hétéroscédasticité et aussi lorsque les distributions n'ont pas la même forme sous l'hypothèse nulle.

[MS-245]

Adaptive Goodness-of-Fit Tests Based on Signed Ranks
Tests d'ajustement adaptatifs basés sur les rangs signés

Angelika Rohde, *Weierstrass Institute Berlin*

Within the nonparametric regression model with unknown regression function l and independent, symmetric errors, a new multiscale signed rank statistic is introduced and a conditional multiple test of the simple hypothesis $l = 0$ against a nonparametric alternative is proposed. This test is distribution-free and exact for finite samples even in the heteroscedastic case. It is rate-optimal and adaptive in the minimax sense against Hölder alternatives under sup-norm loss. Furthermore, its relative asymptotic efficiency with respect to an asymptotically minimax optimal test is close to one in case of homoscedastic Gaussian errors within a broad range of Hölder classes simultaneously.

Dans le modèle de régression non paramétrique avec fonction de régression inconnue l et erreurs indépendantes et symétrique, une statistique de rangs signés à plusieurs échelles est présentée, et un test conditionnel multiple de l'hypothèse simple $l = 0$ contre une alternative non paramétrique est proposé. Ce test est libre de toute distribution et exact pour échantillons finis même dans le cas hétéroscédastique. Il atteint le taux de convergence optimal et est adaptatif dans le sens minimax face aux alternatives de Hölder sous la perte norme sup. En outre, son efficacité asymptotique relative à un test asymptotiquement minimax est près de un dans le cas d'erreurs gaussiennes homoscédastiques, et ce, simultanément à l'intérieur d'une grande étendue de classes de Hölder.

[MS-246]
Rank-based Inference for Principal Components
Inférence fondée sur les rangs pour les composantes principales

Thomas Verdebout, Marc Hallin & Davy Paindaveine, *Université libre de Bruxelles*

One of the most classical tools of multivariate analysis, Principal Components is among the oldest, since that hundred-year-old technique has been first introduced by Pearson (1901). Today, this well-known method is used a lot in social sciences, economy, biology and geography among many other disciplines. We construct rank-based test statistics for hypothesis testing problems related to Principal Component Analysis. Test procedures based on these statistics enjoy a number of attractive properties; they are valid under the family of elliptical distributions, they are efficient at some specified density and they exhibit high asymptotic relative efficiencies with respect to their competitors.

Parmi les outils les plus classiques d'analyse multivariée, les composantes principales sont aussi un des plus anciens puisqu'elles furent introduites il y a cent ans par Pearson (1901). Aujourd'hui, cette méthode est abondamment utilisée en sciences sociales, en économie, en biologie et en géographie pour ne citer que quelques disciplines. Nous construisons des statistiques de test basées sur les rangs pour des problèmes de test d'hypothèse en relation avec l'analyse en composantes principales. Les procédures de test basées sur ces statistiques possèdent de nombreuses propriétés attractives ; elles sont valides sous la famille des lois elliptiques, elles sont efficaces sous une densité spécifiée et elles possèdent de bonnes efficacités asymptotiques relatives par rapport à leurs concurrentes.

Session 12D Mer. 28 mai • Wed. May 28, 10:30 – 12:00 Capital Hall 4A

Censored Data
Données censurées

[MS-247]
Semiparametric Competing Risks with Missing Information
Risques concurrents semi-paramétriques avec données manquantes

Laurent Bordes, *Université de Pau et des Pays de l'Adour*, Jean-Yves Dauxois, *Université de Franche-Comté*, Pierre Joly, *Université Bordeaux 2*

We consider a nonparametric and a semiparametric (in presence of covariates) additive hazards rate competing risks model with censoring and failure cause possibly missing completely at random. Estimators of the unknown parameters are proposed in order to satisfy some optimality criteria. Large sample results are given for all the estimators. Our nonparametric method is applied to a real data set and the behaviour of the semiparametric estimators are analyzed through a Monte Carlo study.

On considère un modèle non paramétrique, ou, lorsque des variables explicatives sont présentes, un modèle semiparamétrique à risques additifs pour modéliser des risques concurrents en présence de censure à droite. L'éventualité de ne pas connaître la cause de défaillance est prise en compte dans le modèle. Des estimateurs des paramètres inconnus sont proposés et doivent vérifier un critère d'optimalité asymptotique. Les propriétés des estimateurs sont présentées et leurs comportements sont analysés au moyen des simulations. Le modèle non paramétrique est appliqué à des données réelles.

[MS-248]
Data Sharpening for Monotonic Local Polynomial Regression with Interval-Censored Responses
Affilage de données en régression monotone polynomiale locale avec réponses censurés par intervalles

Willard Braun & Xiuli Kang, *University of Western Ontario*

Constrained data sharpening is a method in which data are perturbed so that application of a given nonparametric estimator will satisfy a given constraint. We consider an L^2 form of data sharpening in local constant and linear regression with interval-censored responses subject to a monotonicity constraint. Interval censoring is handled using iterative conditional expectations which together with the sharpening method are evaluated by iterating through a sequence of quadratic programs. The performance of this sharpening method is evaluated by simulations; in particular, we consider a mean squared error criterion for both the regression function and its first derivative.

L'affilage de données sous contrainte est une méthode où les données sont perturbées dans l'ordre de façon à ce que l'application d'un estimateur non paramétrique donné satisfasse à une certaine contrainte. Nous considérons une forme L^2 d'affilage de données en régression linéaire et localement constante avec réponses censurées par intervalles sujette à une contrainte de monotonicité. La censure par intervalles est traitée en utilisant des espérances conditionnelles itératives qui, avec la méthode d'affilage, sont évaluées par itération à travers une suite de programmes quadratiques. La performance de cette méthode d'affilage est évaluée par simulation ; en particulier, nous considérons le critère de l'erreur quadratique moyenne pour la fonction de régression et sa première dérivée.

[MS-249]
Hazard Function Modeling
Modélisation de la fonction de risque instantanée

Brenda MacGibbon, *Université du Québec à Montréal*

Estimation of the hazard function in the random right censorship model is consdiered here. Both a Bayesian and a frequentist nonparametric framework are discussed. The utility of wavelets in modeling the hazard is also shown; in particular, their use can sometimes lead to a simpler parametric model. The methods are applied to some interesting data sets, including one from a leukemia study.

Dans cet article, nous étudions l'estimation de la fonction de risque instantanée pour le modèle de censure aléatoire. Nous examinons également deux approches non paramétriques, bayésienne et fréquentiste. Nous montrons comment utiliser les ondelettes pour modéliser le risque instantané et que, parfois, l'utilisation des ondelettes conduit à un modèle paramétrique plus simple. Les méthodes sont appliquées à des ensembles intéressants de données incluant un cas d'une étude sur la leucémie.

Session 12E Mer. 28 mai • Wed. May 28, 10:30 – 12:00 Capital Hall 5A

Design of Experiments in Business and Industrial Statistics
Planification d'expériences en statistique industrielle et de gestion

[MS-250]
Optimal Nonregular Two-Level Split-Plot Designs with a Few Whole-Plots
Plan en parcelles subdivisées à deux niveaux non régulier optimal avec quelques grandes parcelles

Robert McLeod & Jenna Tichon, *University of Winnipeg*, William Li, *University of Minnesota*

Split-plot experiments may arise when it is impractical to completely randomize the treatment combinations of a designed experiment. Presently, the

Les expériences à l'aide de la méthode des parcelles subdivisées peuvent avoir lieu lorsqu'il est impossible de rendre aléatoire les combinaisons de traitement d'une expérience

literature almost exclusively focuses upon the construction and selection of optimal two-level regular split-plot designs. To provide more flexible design choices in the split-plot setting, we explore the use of nonregular two-level split-plot designs having 12, 16, 20 and 24 runs and between one and three whole-plot factors. We extend the notions of word length, indicator functions and generalized minimum aberration to the nonregular split-plot setting and propose five ranking schemes for assessing optimality.

conçue. Actuellement, la littérature se concentre presque exclusivement sur la construction et la sélection de plan en parcelles subdivisées régulier à deux niveaux optimaux. Pour fournir un choix de plan plus flexible dans les paramètres des parcelles subdivisées, nous explorons l'utilisation de plan en parcelles subdivisée non régulier à deux niveaux qui ont 12, 16, 20 et 24 passages et entre un à trois facteurs à grandes parcelles. Nous étendons les notions de longueur de mot, fonctions indicatrices et aberration minimale généralisée au paramètre des parcelles subdivisées non régulier et proposons cinq systèmes de classement par ordre d'importance pour évaluer l'optimalité.

[MS-251]
Experimental Designs for Qualitative and Quantitative Factors
Plans d'expérience pour facteurs qualitatifs et quantitatifs

Walter Tinsson, *Université de Pau et des Pays de l'Adour*

A main practical problem is the construction of efficient designs for both quantitative and qualitative factors. We consider designs adapted to a linear model additive for the qualitative factors and quadratic for the quantitative factors. Such a model is a generalization of the well-known blocked effects model. Extending the notion of blocks, we prove that it is possible to obtain designs such that the qualitative effects are orthogonal to the quantitative ones. All the effects can also be estimated in a very easy way. If the experiments are expensive, then D-optimal subdesigns can be obtained by an exchange algorithm.

Un important problème pratique est celui de la construction de plans d'expérience efficaces pour facteurs qualitatifs et quantitatifs. Nous considérons ici des plans adaptés à un modèle linéaire additif pour les facteurs qualitatifs et quadratiques pour les facteurs quantitatifs. Un tel modèle généralise le modèle à effets de blocs classique. En étendant la notion de bloc on montre qu'il est possible d'obtenir des plans tels que les effets des facteurs qualitatifs et quantitatifs sont orthogonaux. Tous ces effets peuvent de plus être estimés de manière très simple. Si les expériences sont très coûteuses alors des plans D-optimaux de petite taille peuvent être extraits avec un algorithme d'échange.

[MS-252]
Space Filling Designs Qualification in High Dimension by Minimal Spanning Tree.
Qualification des plans Space Filling Designs en grande dimension par l'arbre de longueur minimal

Olivier Vasseur, *ONERA*, Jessica Franco & Bernard Corre, *TOTAL*, Michelle Sergent, *Université Paul Cézanne*

In the field of computer experiments, when the relations between the response and the entries of the numerical code are complex, the experimental designs (Space Filling Designs) are used for the exploration of the code or the construction of response surfaces. The points of these plans must fill space as well as possible. In our talk, we present the advantages of the use of criteria based on the Minimal Spanning Tree (MST) to qualify these plans. The results will be illustrated by various types of plans and will point out the interest of WSP plans when the number of parameters is higher than 5.

Dans le domaine de l'expérimentation numérique, lorsque les relations entre la réponse et les entrées du code de calcul sont complexes, les plans d'expériences (Space Filling Designs) sont utilisés pour l'exploration du code ou la construction de métamodèles. L'objectif est alors que les points de ces plans remplissent au mieux l'espace. Dans notre exposé nous présentons les avantages qu'offre l'utilisation de critères basés sur l'arbre de longueur minimal (ALM) pour qualifier ces plans. Les résultats seront illustrés par différents types de plans et souligneront l'intérêt des plans WSP lorsque le nombre de paramètres est supérieur à 5.

[MS-253]
Multivariate Statistical Control of Batch Processes with Varying Duration
Contrôle multivarié de procédés par lots à temps variable

Ndèye Niang Keita, *Chaire de Statistique Appliquée et CEDRIC, CNAM*, Flavio Fogliatto, *Universidade Federal do Rio Grande do Sul*, Pierre-Louis Gonzalez, *Chaire de Statistique Appliquée et CEDRIC, CNAM*

The purpose of this paper is to provide a new quality control strategy for unsynchronized batch processes, which cannot be directly monitored using classical control charts based on multiway principal component analysis. In our proposition, data arising from such batch processes are completed using a straightforward scheme to preserve all information on the variability in batch profiles along the time axis. The completed data set is reduced using a three-way method and monitoring of batch performance is accomplished directly on principal plane graphs, from which non-parametric off-line and on-line control charts are derived. Examples using real or simulated data illustrate the proposed method.

Cette communication propose une stratégie utilisant des méthodes d'analyse de données évolutives pour le monitoring de procédés par lots à temps variable que les techniques classiques, basées sur l'analyse des données fonctionnelles, ne permettent pas de contrôler. L'approche consiste à compléter les données en utilisant un modèle préservant la variabilité temporelle. On construit alors des cartes de contrôles non paramétriques, pour le contrôle en fin et en cours de fabrication, à partir des plans factoriels issus de l'application de méthodes d'analyse conjointe de tableaux. L'approche est illustrée sur des données simulées et réelles.

Session 12F Mer. 28 mai • Wed. May 28, 10:30 – 12:00 Capital Hall 6A

Data Validation and Small Area Estimation
Validation de données et estimation d'aires faibles

[MS-254]
Outlier Detection Methodology for Price Index Surveys
Méthode de détection des valeurs aberrantes applicable aux enquêtes pour l'établissement des indices des prix

Saad Rais, *Statistics Canada*

Survey data may be contaminated with anomalous observations known as outliers, which must be identified and treated. Statistics Canada's Consumer Price Index (CPI) processing system is undergoing a redesign, part of which involves employing an automated and reliable outlier detection method. Three such outlier detection methods have received significant attention in the world of price index survey methodology: The Tukey Algorithm, the Quartile Method, and the Hidiroglou-Berthelot Method. We study and compare these methods using CPI data by examining the tolerance intervals resulting from each of the methods, and by observing the impact of these methods on the index.

Les données d'enquête peuvent être contaminées par des observations anormales, appelées valeurs aberrantes, qu'il convient de repérer et de traiter. Le système de traitement de l'Indice des prix à la consommation (IPC) de Statistique Canada fait à l'heure actuelle l'objet d'un remaniement qui consiste en partie à employer une méthode automatisée et fiable de détection des valeurs aberrantes. Dans le domaine de la méthodologie des enquêtes pour l'établissement des indices des prix, trois méthodes de détection des valeurs aberrantes ont suscité beaucoup d'intérêt, à savoir l'algorithme de Tukey, la méthode des quartiles et la méthode d'Hidiroglou-Berthelot. Nous étudions et comparons ces méthodes à l'aide de données de l'IPC en examinant les intervalles de tolérance qu'elles produisent et en observant leur effet sur l'indice.

[MS-255]
Dynamic Outlier Detection in Price Indexes
Détection dynamique des données aberrantes dans les indices de prix

Holly Hutton, *Statistics Canada*

One aspect of analyzing survey data is measuring the effect of single observations on the overall estimate, especially outliers. Generally, a systematic approach is employed to identify and modify observations falling outside a set threshold. Dynamic outlier detection is such that, with the aid of an analytical user interface, analysts can use systematic outlier detection methods and manual data updates to immediately see the impact on the price index. The modifications can be kept or undone during the process. This paper highlights a new interface for analysts to test different outlier detection methods and details the underlying theory involved.

Un des aspects de l'analyse de données d'enquête est de mesurer l'effet de certaines observations sur l'estimation totale, surtout lorsqu'il s'agit de données aberrantes. Généralement, une approche systématique est employée pour identifier et modifier les observations qui se trouvent au-delà d'un certain seuil. La détection dynamique de données aberrantes est telle, qu'avec l'aide d'une interface analytique les analystes peuvent utiliser des méthodes systématiques de détection de données aberrantes et changer manuellement les données aberrantes pour voir immédiatement l'effet sur l'indice de prix. Les modifications peuvent être conservées ou annulées pendant le processus. Cet article présente un nouvel outil graphique destiné aux analystes dont l'objectif est de tester différentes méthodes de détection de données aberrantes et de fournir les détails de la théorie sous-jacente.

[MS-256]
Benchmarking Comparison and Survey Error Modelling
Comparaison de l'étalonnage et modélisation de l'erreur de l'enquête

Ka Ho Wu, *Chinese University of Hong Kong*, Zhao-Guo Chen, *Statistics Canada*

For a target socio-economic variable, two sources of data with different frequencies and precisions may be available. A benchmarking procedure is using the less frequent and more reliable data to adjust the more frequent and less reliable data. For advanced benchmarking methods, the survey error model is needed but usually unavailable. In this talk, we compare three benchmarking methods, namely the Denton, regression, and signal-extraction methods. Assuming the survey-error series follows an AR(1) model, we compare the benchmarking results by using default models and a survey error modelling procedure. This study is useful for developing and applying a benchmarking software.

Deux sources de données de précisions différentes et avec des fréquences de collecte différentes peuvent être disponibles pour une variable socioéconomique. Une procédure d'étalonnage consiste à utiliser les données moins fréquentes et plus fiables pour corriger les données plus fréquentes et moins fiables. Pour les méthodes d'étalonnage avancé, un modèle d'erreur d'échantillonnage est nécessaire mais généralement non disponible. Dans cet article, nous comparons trois méthodes d'étalonnage, à savoir la méthode de Denton, la méthode basée sur un modèle de régression et la méthode basée sur l'extraction du signal. En supposant que la série d'erreurs d'échantillonnage suit un processus autorégressif d'ordre 1, nous comparons les résultats de l'étalonnage à l'aide de modèles par défaut et d'un processus de modélisation des erreurs d'échantillonnage. Cette étude est utile pour mettre au point et appliquer un logiciel d'étalonnage.

[MS-257]
Small Area Estimation to Study the Impact of Globalization
Estimation sur petits domaines pour étudier l'effet de la mondialisation

Catalin Dochitoiu, Susana Rubin-Bleuer & Carlos Leon, *Statistics Canada*

We study the feasibility of producing relevant statistics to measure the impact of globalization in the Canadian economy. For this, totals of key economic variables for the wholesale industry at the level of "Trade Group" by "Province" by "Globalization Indicators" are required. We consider small area estimates using area level models with random effects under a "frequentist" approach. We use penalized splines in order to accommodate departures from linearity. We derive the properties of the estimators from the General Linear Mixed Model framework using the methodology described in (Rao, 2003).

Nous étudions la faisabilité de produire de produire des statistiques pertinentes pour mesurer l'effet de la mondialisation sur l'économie canadienne. Pour ce faire, les totaux de variables économiques pour l'industrie du commerce de gros au niveau du « groupe de commerce » par « province » et par « indicateur de mondialisation » sont requises. Nous examinons des estimations sur petits domaines en utilisant des modèles au niveau du domaine avec effets aléatoires selon une approche « fréquentiste ». Afin de tenir compte des écarts par rapport à la linéarité, nous utilisons des « splines » pénalisées. Nous dérivons les propriétés des estimateurs du cadre du modèle linéaire mixte généralisé en utilisant la méthodologie décrite dans Rao (2003).

[MS-258]
Small Area Estimation Using Area Level Models with Model Checking and Applications
Estimation sur petits domaines à l'aide de modèles au niveau du domaine avec vérification du modèle et applications

Yong You, *Statistics Canada*

Various area level models have been developed to produce adequate model-based estimates for small areas. In this paper, we will discuss area level models including the basic Fay-Herriot model, unmatched models, cross-sectional and time series models and area level spatial models. We will also discuss the problem of sampling variance modeling for area level models using direct variance estimates and smoothing techniques including the methods of generalized variance functions and design effects smoothing. Then we will discuss the problem of model checking using posterior predictive distributions for area level models. Finally, we will provide applications of area level modeling.

Divers modèles au niveau du domaine ont été élaborés en vue de produire des estimations fondées sur un modèle adéquat pour des petits domaines. Dans la présente communication, nous discutons des modèles au niveau du domaine, dont le modèle élémentaire de Fay-Herriot, les modèles non appariés, les modèles transversaux et chronologiques, ainsi que les modèles spatiaux. Nous discuterons également du problème de la modélisation de la variance d'échantillonnage pour des modèles au niveau du domaine en utilisant des estimations directes de la variance et des techniques de lissage, y compris les méthodes fondées sur des fonctions de variance généralisées et le lissage des effets de plan. Puis, nous aborderons le problème de la vérification des modèles en utilisant des lois prédictives a posteriori pour les modèles au niveau du domaine. Enfin, nous présenterons des applications de la modélisation au niveau du domaine.

[MS-259]
Prediction of Ordered Random Effects in a Simple Small Area Model
Prévision d'effets aléatoires ordonnés dans un modèle de petit domaine simple

Yaakov Malinovsky & Yosef Rinott, *Hebrew University of Jerusalem*

Estimation of ordered parameters arises naturally in the context of Small Area Estimation (SAE). For

L'estimation de paramètres ordonnés survient naturellement dans le contexte d'estimations de petits domaines (EPD). Par

example, one may want to estimate the parameter associated with the best or worst area, or a certain percentile. We use a simple SAE model to show that estimation of ordered parameters by the corresponding ordered estimates of each area separately does not yield good results with respect to Mean Square Error (MSE). Shrinkage-type estimates are considerably better, and their performance is close to that of the optimal estimates, which cannot in general be computed explicitly.

exemple, on peut s'intéresser à l'estimation du paramètre associée au meilleur ou au pire domaine, ou un certain centile. Nous utilisons un modèle d'EPD simple afin de montrer que l'estimation de paramètres ordonnés par les estimés ordonnés correspondant de chaque domaine ne donne pas de bons résultats d'après le critère de l'erreur quadratique moyenne (EQM). Les estimateurs par rétrécissement sont considérablement meilleurs, et leur performance est proche de celles des estimateurs optimaux, qui ne peuvent généralement pas être calculés explicitement.

Session 12G Mer. 28 mai • Wed. May 28, 10:30 – 12:00 Capital Hall 7A

Estimation Methods
Méthodes d'estimation

[MS-260]
Density Support Estimation
Estimation du support d'une densité

Bruno Pelletier, *Université Montpellier II*, Gérard Biau, *Université Paris VI*, Benoît Cadre, *ENS Cachan-Bretagne*

Given an unknown multivariate probability density f with compact support, and an i.i.d. random sample drawn from f, we study the estimator of the support of f defined as unions of balls centered at the observations and of common radius r_n. To measure the quality of the estimation, we use a general criterion based on the volume of the symetric difference. Under some mild assumptions, and using tools from differential geometry, we establish the exact convergence rate of the estimator, then we discuss consequences of our results from a statistical perspective.

Étant donné une densité de probabilité multivariée inconnue f de support compact, et un échantillon i.i.d. issu de f, nous étudions l'estimateur du support de f défini par l'union de boules de rayon r_n centrées sur les observations. Afin de mesurer la qualité de l'estimation, nous utilisons un critère général fondé sur le volume de la différence symétrique. Sous quelques hypothèses non restrictives, et en utilisant des outils de la géométrie différentielle, nous établissons la vitesse de convergence exacte de l'estimateur du support, puis nous discutons des conséquences de ce résultat dans une perspective statistique.

[MS-261]
Estimating the Support of a Probability Distribution
De l'estimation du support d'une loi de probabilité

Denis Bosq, *Université Paris 6 - Pierre et Marie Curie*

This communication deals with the estimation of the support S of an unknown probability distribution from observations. First, we briefly recall some known results. Then, we study the rate of convergence of the Geffroy estimator of the edge function g when S is not compact. We also construct and study a smooth estimator of g based on a nonparametric density estimator. Finally we derive an estimator of a manifold based on the "explosion" of a functional estimator.

Cet exposé est consacré à l'estimation du support S d'une loi de probabilité à partir d'un échantillon de cette loi. Nous commençons par rappeler brièvement quelques résultats connus, puis nous étudions la vitesse de convergence de l'estimateur de Geffroy de la fonction frontière g lorsque S n'est pas compact. Nous construisons également un estimateur lisse de g, basé sur un estimateur non paramétrique de la densité. Enfin, si S est une variété, nous utilisons l'explosion d'un estimateur fonctionnel pour estimer S.

[MS-262]
L^1 Quantization in Banach Spaces
Quantization L^1 dans un espace de Banach

Thomas Laloë, *Université Montpellier II*

Let X be a random variable with distribution μ taking values in a general Banach space H. First, we establish the existence of an optimal quantization of μ with respect to the L^1 distance. Seconds, we propose several estimators of the optimal quantizer in the potentially infinite-dimensional space H, with the associated algorithms. Finally, we discuss practical result obtained from simulated and real-life data sets.

Soit X une variable aléatoire de loi μ à valeurs dans un espace de Banach général H. Nous commençons par établir l'existence d'une « quantification » optimale pour μ par rapport à une distance L^1. Ensuite nous proposons plusieurs estimateurs du « quantificateur » optimal dans l'espace (pouvant être de dimension infinie) H, ainsi que les algorithmes permettant de les calculer. Enfin nous examinons les résultats pratiques obtenus avec des données réelles ou simulées.

[MS-263]
Median Loss Analysis
Analyse de la perte médiane

Chi Wai Yu & Bertrand Clarke, *Department of Statistics, University of British Columbia, Canada*

In place of the expected loss in the conventional statistical decision theory, we justify using the median of the loss. The resulting estimates can be derived in many standard examples. The estimates yield improved predictive performance, have a higher resistance to outliers than the usual expectation-based estimates, and are resistant to the specific loss used to form them. In several simulations, we systematically replace the usual expectation-based methods with our median-based approach. In particular, we use a median based cross-validation under squared error loss and verify that this give better results for model selection outside the normal error case.

Au lieu de la perte espérée dans la théorie classique de la décision statistique, nous justifions l'utilisation de la médiane de la perte. Les estimations résultantes peuvent être calculées dans de nombreux exemples standards. Les estimations ont de meilleures propriétés prédictives, ont une plus forte résistance aux valeurs aberrantes que les estimations fondées sur l'espérance habituelle et sont plus résistantes à la perte particulière utilisée pour les former. Dans plusieurs simulations, nous remplaçons systématiquement les méthodes fondées sur l'espérance habituelle par notre approche de la médiane. Nous utilisons notamment une validation croisée basée sur la médiane sous une fonction de perte égale au carré de l'erreur et vérifions qu'elle donne de meilleurs résultats pour le choix du modèle en dehors du cas d'une erreur normale.

[MS-264]
Estimating Normal Quantiles when the Mean is Bounded
Estimations de quantiles normaux lorsque la moyenne est bornée

Yogesh Tripathi & Éric Marchand, *Université de Sherbrooke*

We consider the problem of estimating normal quantiles under the constraints that the mean is bounded. The loss function is scale invariant squared error. Some commonly known estimators for the unrestricted problem are MLE, UMVUE and BEE. In the present article, these estimators are improved by those which capitalize on the prior information available on the parameter space.

Nous considérons le problème d'estimation des quantiles normaux sous la contrainte que la moyenne est bornée. La fonction de perte est l'erreur quadratique invariante aux changements d'échelle. Des estimateurs bien connus pour le problème non restreint sont les EMV, ESBVUM et MEE. Dans le présent article, ces estimateurs sont améliorés par ceux tirant profit de de l'information *a priori* disponible sur l'espace paramétrique. De plus, en utilisant l'approche

Also, using IERD approach of Kubokawa (1994) a class of estimators dominating BEE is developed. It is shown that this class contains the generalized Bayes estimator with respect to the uniform prior on the resteicted parameter space.

EIDR de Kubokawa (1994), une classe d'estimateurs dominant les MEE est développée. Nous montrons que cette classe contient l'estimateur de Bayes généralisé par rapport à la loi *a priori* uniforme sur l'espace paramétrique restreint.

[MS-265]

Estimation and Prediction under Balanced Type loss functions with applications
Estimation et prévision sous des fonctions de perte de type équilibré avec applications

Mohammad Jafari Jozani & Éric Marchand, *Université de Sherbrooke*

In this paper a general class of balanced loss functions (BLF) is introduced and the Bayesian estimator of an unknown parameter θ is derived under the corresponding unbalanced loss. The issues of Bayesianity, admissibility, dominance, and minimaxity under balanced squared error loss function(BSELF) will be discussed and demonstrated how results under SELF serves directly to derive results under BSELF. The obtained results will be generalized to the Bayesian prediction problem. Some applications of BLF in credibility theory and Bayesian prediction of record values will be given.

Dans cet article, une classe générale de fonctions de perte équilibrées (FPÉ) est présentée et l'estimateur bayésien d'un paramètre inconnu θ est obtenu. Nous montrons comment l'estimateur de Bayes sous la FPÉ peut être obtenu en utilisant l'estimateur de Bayes sous la perte non équilibrée correspondante. Nous discutons des problèmes de Bayésiennenté, d'admissibilité, dominance, et minimaxité sous la fonction de perte d'erreur quadratique équilibrée (FPEQÉ), et montrons comment les résultats sous la FPEQ servent directement à l'obtention de résultats sous la FPEQÉ. Les résultats obtenus sont généralisés au problème de la prévision bayésienne. Quelques applications des FPÉ en théorie de la crédibilité et en prévision bayésienne de valeurs record sont présentées.

Session 12H Mer. 28 mai • Wed. May 28, 10:30 – 12:00 Capital Hall 8A

Functional Estimation
Estimation fonctionnelle

[MS-266]

Some Asymptotic Properties of Bernstein Estimators of Distribution Functions
Quelques propriétés asymptotiques des estimateurs de Bernstein d'une fonction de répartition

Alexandre Leblanc, *University of Manitoba*

It is well known that kernel estimators of distribution functions can asymptotically outperform the empirical distribution function in terms of (integrated) mean square error. In this talk, we discuss the fact that this is also true for Bernstein estimators of distribution functions. Specifically, we introduce a higher order expansion for the asymptotic (integrated) mean square error of Bernstein estimators and examine the relative deficiency (cf. Hodges and Lehman, 1970, Ann. Math. Stat. **41**, 783 – 801) of the empirical distribution function with respect to these estimators.

Il est bien connu que les estimateurs à noyaux de fonctions de répartition peuvent asymptotiquement dominer la fonction de répartition empirique en termes d'écart quadratique moyen (intégré). Durant cette présentation, nous discuterons du fait que les estimateurs d'une fonction de répartition basés sur les polynômes de Bernstein possèdent aussi cette propriété. Spécifiquement, nous présenterons un développement en série d'ordre supérieur de l'écart quadratique moyen (intégré) des estimateurs de Bernstein et examinerons la déficience relative (cf. Hodges et Lehman, 1970, Ann. Math. Stat. **41**, 783-801) de la fonction de répartition empirique par rapport à ces estimateurs.

[MS-267]
Frontier Estimation via Regression on High Power-transformed Data
Estimation de contour par régression sur les données élevées à la puissance

Stéphane Girard, *INRIA Rhône-Alpes*, Pierre Jacob, *Université Montpellier 2*

We present a new method for estimating the frontier of a multidimensional sample. The estimator is based on a regression of the power-transformed data. Two estimators of the regression function are considered: kernel and local polynomials estimators. We assume that the exponent of the transformation goes to infinity while the smoothing parameter goes to zero. We give conditions on these two parameters to obtain complete convergence and asymptotic normality. The optimality of the speed of convergence is established for multidimensional, Lipschitzian frontiers. The good performance of the estimator is illustrated on some finite sample situations.

Nous présentons un nouvel estimateur du contour d'un échantillon multidimensionnel. L'estimateur est basé sur une régression effectuée sur les données transformées par élévation à la puissance. Deux estimateurs de la fonction de régression sont considérés : un estimateur à noyau et un estimateur par polynômes locaux. Nous supposons que l'exposant de la transformation tend vers l'infini alors que le paramètre de lissage tend vers zéro. La convergence presque complète et la normalité asymptotique de l'estimateur sont établies et l'optimalité de la vitesse de convergence est obtenue pour des contours multidimensionnels et lipschitziens. Les bonnes performances de l'estimateur sont illustrées sur simulations.

[MS-268]
Orthogonal Polynomial Density Estimates and their Kernel Representation
Estimations de densité par des polynômes orthogonaux et représentation noyau de ces estimations

Min Jiang & Serge Provost, *University of Western Ontario*

Density estimates which are based on sample moments and expressible as the product of a weight function and a certain linear combination of the corresponding orthogonal polynomials are being considered. A kernel representation of such density estimates is also provided. Particular attention is given to estimates involving the Legendre, Jacobi, Laguerre and Hermite classical orthogonal polynomials. A criterion is proposed for determining the optimal number of sample moments to be utilized in conjunction with the proposed approach. The methodology is applied to samples generated from mixtures of known distributions and several well-know data sets as well.

Cet article porte sur les estimations de densité fondées sur les moments de l'échantillon et pouvant être exprimées comme le produit d'une fonction de pondération et d'une certaine combinaison linéaire des polynômes orthogonaux correspondants. Une représentation noyau de ces estimations de densité est également fournie. Une attention particulière est accordée aux estimations impliquant les polynômes orthogonaux classiques de Legendre, de Jacobi, de Laguerre et d'Hermite. Un critère est proposé pour déterminer le nombre optimal de moments de l'échantillon à utiliser en association avec l'approche proposée. La méthode est appliquée à des échantillons générés à partir de mélanges de distributions connues ainsi que d'ensembles de données célèbres.

[MS-269]
On Approximating the Derivative of the Logarithm of a Density by Means of a Rational Function
Sur l'approximation de la dérivée du logarithme d'une densité de probabilité par une fonction rationnelle

Serge Provost, *The University of Western Ontario*, Hyung-Tae Ha, *Kyungwon University*, Min Jiang, *The University of Western Ontario*

A density function approximation methodology, according to which the first derivative of the logarithm of the density approximant is expressed as a

L'on propose une méthodologie pour approximer une densité de probabilité selon laquelle la première dérivée du logarithme de l'approximant est exprimée comme un rapport

ratio of polynomials, is being proposed. The polynomial coefficients of the numerator and denominator of the rational function are determined by solving the system of linear equations which results from matching the moments of the approximant to those of the distribution being approximated. Several illustrative examples will be presented.

de polynômes. Les coefficients polynomiaux du numérateur et du dénominateur de la fonction rationnelle sont déterminés en résolvant le système d'équations linéaires obtenu en posant les moments de l'approximant égaux à ceux de la distribution approximée. Plusieurs exemples seront présentés à titre illustratif.

Session 12I Mer. 28 mai • Wed. May 28, 10:30 – 12:00 Capital Hall 9A

Extreme Values and Applications
Valeurs extrêmes et applications

[MS-270]
On Inverse Power Laws and Risk Analysis
Des lois de puissance inverse et de l'analyse de risque

Mei Ling Huang & Ke Zhao, *Brock University*

Statistical inverse power laws have been developed during the last decade. The Pareto distribution is a power-tailed distribution having many applications. This paper proposes an optimal estimation method of the truncated Pareto distribution. For the new estimator, an exact efficiency function relative to the classical estimators and its properties are derived. Monte Carlo simulation results confirm the theoretical conclusions. Applications of hurricane and forest fire loss data have been analyzed based on these estimation methods. The studies show that the new estimation method is more efficient relative to the classical methods and fits data well in risk analysis.

Des lois statistiques de puissance inverse ont été développées au cours de la dernière décennie. La distribution de Pareto est une distribution à longue queue ayant plusieurs applications. Cet article propose une méthode d'estimation optimale pour la distribution de Pareto tronquée. Une fonction exacte d'efficacité relative aux estimateurs classiques du nouvel estimateur, ainsi que ses propriétés, sont obtenus. Des résultats de simulation de Monte-Carlo confirment les conclusions théoriques. Des applications à des données de pertes dues aux ouragans et aux feux de forêts ont été analysées à l'aide de ces méthodes d'estimation. Les études montrent que la nouvelle méthode est plus efficace que les méthodes classiques et ajuste bien les données en analyse de risque.

[MS-271]
A Multiple Test Procedure for the Asymptotic Independence of Spatial Extremes
Une procédure de tests multiples pour l'indépendance asymptotique des extrêmes spatiaux

Jean-Noel Bacro, *Université Montpellier 2*, Liliane Bel, *AgroParisTech*, Christian Lantuéjoul, *Ecole des Mines, Fontainebleau*

For spatial processes such as environmental ones, there is interest in features of the dependence of extreme spatial events. In the spatial context, asymptotic independence of extremes corresponds to extreme values that are not spatially structured. Focusing on pairwise dependence of maxima from data recorded in space, we propose to test the spatial independence of maxima through the madogram function, one of the functions usually used

L'étude de processus environnementaux spatiaux amène à s'intéresser à la dépendance des extrêmes de processus spatiaux. La notion usuelle d'indépendance asymptotique des extrêmes correspond à des processus ne présentant plus de structures spatiales dans leurs réalisations extrêmes. Nous proposons une procédure permettant de tester l'indépendance asymptotique spatiale de tout couple de maxima mesurés en deux sites de l'espace. Notre statistique de test est construite à partir de la fonction madogramme,

in geostatistics. A procedure to take into account the induced multiple tests within each class of distances is proposed. Illustrations on simulated and real data sets are given.

fonction couramment utilisée en géostatistique. Une prise en compte de la multiplicité des tests effectués à classe de distance fixée est proposée. L'approche est illustrée sur données simulées et réelles.

[MS-272]
Frontier Estimation and Extreme Values Theory
Estimation de frontières utilisant la théorie des extrêmes

Abdelaati Daouia, *GREMAQ, Université de Toulouse I*, Jean-Pierre Florens, *GREMAQ/IDEI, Université de Toulouse I*, Léopold Simar, *Université Catholique de Louvain*

In this paper, we investigate the problem of monotone frontier estimation from an extreme-values theory perspective. This allows us to revisit the asymptotic theory of the popular FDH estimator, to derive new and asymptotically Gaussian estimators and to provide useful asymptotic confidence bands for the monotone boundary. The study of the asymptotic properties of the frontier estimators is carried out by relating them to an original dimensionless random sample and then applying standard extreme-values theory. The finite sample behavior of the estimators is explored through Monte Carlo experiments.

Dans ce travail, on établit un lien simple entre l'estimation de la frontière monotone d'un support multivarié et la théorie standard des extrêmes. Ce qui permet de déterminer les conditions nécessaires et suffisantes pour la convergence en loi de l'estimateur usuel FDH et de spécifier sa loi asymptotique ainsi que sa vitesse de convergence. Aussi, on propose deux nouveaux estimateurs avec une loi asymptotique normale permettant de construire des intervalles de confiance pour la frontière ce qui n'est pas le cas de l'estimateur FDH. Les propriétés de convergence des estimateurs sont illustrées par une étude empirique.

[MS-273]
Car Insurance Management: A Local Approach for Serious Accidents
Une approche locale de la gestion des sinistres graves en assurance automobile

Michel Grun-Rehomme, *Université Paris 2, ERMES*, Olga Vasechko, *Research Institute of Statistics, Kyiv, Ukraine*, Noureddine Benlagha, *Université Paris 2, ERMES*

This article studies the stability of the premium in each risk class for car insurance. This stability enables a better match between the cost of accidents and insurance pricing in strongly competing markets. The presence of serious accidents disturbs this hypothesis of collective risk differentiation from one class to another and the temporal stability of the premium indicator. The approach suggested for the inlier's detection (local approach) uses the premium variance estimation. A practical numerical application to insurance data is presented.

Cet article se propose d'étudier la stabilité temporelle des indicateurs de risque (prime pure) dans les classes (supposées homogènes) d'assurés en assurance automobile. Cette stabilité est nécessaire pour obtenir une bonne adéquation entre la sinistralité et les cotisations des assurés dans ce contexte d'un marché fortement concurrentiel. La présence de sinistres graves vient perturber cette hypothèse de différenciation du risque collectif d'une classe à l'autre et la stabilité temporelle de cet indicateur de prime pure. La démarche proposée pour détecter les « inliers » (point de vue local au niveau des classes), est basée sur une estimation de la variance de l'indicateur de prime pure, pour une précision souhaitée (calculée sur la différence entre classes successives) et un risque d'erreur fixé. Une application numérique sur des données réelles d'assurance est présentée pour mettre en pratique cette démarche.

[MS-274]
Dependence and the Asymptotic Behavior of Large Claims Reinsurance
Dépendance et comportement asymptotique de grandes réclamations de réassurance

Alexandru Asimit, *University of Toronto*, Bruce Jones, *University of Western Ontario*

Insurance companies often use reinsurance as a mechanism for sharing risk, particularly when there is the possibility of catastrophic losses. Two appealing reinsurances are ECOMOR (excédent du coût moyen relatif) and LCR (largest claims reinsurance). Under these treaties, the amount paid by the reinsurer is based on the upper order statistics of the claims. An extension of the classical compound Poisson risk model is considered, where the inter-arrival time and forthcoming claim are dependent. Asymptotic tail probabilities of the reinsurance amount under those treaties are obtained. The speed of convergence of our results is examined via a simulation study.

Les compagnies d'assurances recourent souvent à la réassurance comme moyen de partager les risques, particulièrement quand pourraient se produire des pertes catastrophiques. Deux types de traité de réassurance intéressants sont ECOMOR (pour excédent du coût moyen relatif) et LCR (largest claims reinsurance – pour réassurance des sinistres les plus coûteux). Aux termes de ces traités, le montant versé par le réassureur est fondé sur les statistiques d'ordre supérieur des sinistres. Nous considérons une extension du modèle classique de risque composé de Poisson, où l'intervalle de temps d'arrivée et le sinistre à venir sont dépendants. Nous obtenons les probabilités asymptotiques de queue de distribution du montant de la réassurance aux termes de ces traités. Nous examinons la vitesse de convergence de nos résultats au moyen d'une étude par simulation.

Session 13A Mer. 28 mai • Wed. May 28, 13:30 – 14:30 Congress Hall A/E

Canadian Journal of Statistics Award Address
Allocution du récipiendaire du prix de *La revue canadienne de statistique*

[MS-275]
Optimal Designs for Calibration of Orientations
Les plans optimaux pour calibrer des mesures d'orientation

Hwashin Shin, *Health Canada*, Glen Takahara, *Queen's University*, Duncan Murdoch, *University of Western Ontario*

This paper concerns designed experiments involving observations of orientations following the models of Prentice(1989) and Rivest & Chang(2006). The authors state minimal conditions on the designs for consistent least squares estimation of the matrix parameters in these models. The conditions are expressed in terms of the axes and rotation angles of the design orientations. The authors show that designs satisfying $U_1+\cdots+U_n = 0$ are optimal in the sense of minimizing the estimation error average angular distance. The authors give constructions of optimal n-point designs when n is at least 4 and they compare the performance of several designs through approximations and simulation.

Cet article concerne la planification d'expériences dans le cadre des modèles de Prentice (1989) et de Rivest & Chang (2006) pour des mesures d'orientation. Les auteurs donnent des conditions minimales qu'un plan expérimental doit respecter pour que l'estimation des moindres carrés des paramètres matriciels de ces modèles soit convergente. Ces conditions s'expriment en fonction des axes et des angles des matrices de rotation du plan. Les auteurs montrent que les plans tels que $U_1 + \cdots + U_n = 0$ sont optimaux au sens où la distance angulaire moyenne des erreurs d'estimation est minimisée. Ils précisent la forme des plans optimaux à n points, où n est supérieur ou égal à 4, et ils comparent la performance de divers plans au moyen d'approximations et de simulations.

Session 14A Mer. 28 mai • Wed. May 28, 14:35 – 16:05 Capital Hall 1A

Design and Analysis of Industrial Experiments
Planification et analyse d'expériences industrielles

[MS-276]
Analysis of Data Arising from Industrial Split-Plot Screening Experiments
Analyse de données provenant de tests de dépistage industriels avec subdivision des parcelles

John Brewster, *University of Manitoba*

Fractional factorial designs are often used for screening purposes. If some factors are harder to vary than others, then these designs may have a split-plot structure. With the addition of center points, estimates of the variance components of interest may be obtained. However, these estimates are often based on few degrees of freedom, particularly at the whole-plot level, which may lead to negative variance estimates and other inferential problems. We show that adjusting the degrees of freedom in the analysis of variance table, based on the data, can lead to procedures with good frequentist properties and a Bayesian interpretation.

Les plans fractionnels factoriels sont souvent utilisés à des fins de dépistage. Si certains facteurs sont plus difficiles à faire varier que d'autres, alors ces plans peuvent avoir une structure à subdivision de parcelles. Avec l'ajout de points centraux, des estimateurs des composantes de la variance d'intérêt peuvent être obtenus. Cependant, ces estimateurs sont souvent basés sur peu de degrés de liberté, particulièrement au niveau des parcelles entières, ce qui peut mener à des estimés négatifs de la variance et à d'autres problèmes inférentiels. Nous montrons qu'ajuster les degrés de liberté dans la table d'analyse de variance, en se basant sur les données, peut mener à des procédures ayant de bonnes propriétés fréquentistes et une interprétation bayésienne.

[MS-277]
Screening Designs for Model Selection
Plan de tri pour la sélection de modèles

William Li & C. J. Nachtsheim, *University of Minnesota*, V. Agoboto, *Meharry Medical College*

We introduce new criteria for model discrimination and use these and existing criteria to evaluate standard orthogonal designs and to construct optimal two-level designs for screening experiments. Results indicate that, for a given sample size and number of desired factors, not all orthogonal designs are model-discriminating designs for the model spaces considered. It is also shown that nonorthogonal optimal designs can be constructed that out-perform orthogonal designs for the criteria considered. Simulation studies demonstrate that the constructed optimal designs can lead to a higher probability of choosing the correct model in the analysis procedure than using competing orthogonal designs.

Nous introduisons de nouveaux critères pour la discrimination entre divers modèles et les utilisons avec les critères existants pour évaluer des plans orthogonaux standard et pour construire des plans optimaux à deux niveaux pour des expériences de tri. Les résultats indiquent que, pour une taille donnée d'échantillon et un nombre donné de facteurs souhaités, les plans orthogonaux ne permettent pas tous de faire la discrimination entre les modèles pour les espaces de modélisation considérés. Nous montrons aussi qu'il est possible de construire des plans non orthogonaux optimaux qui donnent de meilleurs résultats que les plans orthogonaux pour les critères envisagés. Des études par simulation démontrent que les plans optimaux construits peuvent aboutir à une probabilité plus élevée de choisir le modèle correct dans la procédure d'analyse que l'utilisation des plans orthogonaux concurrents.

[MS-278]
Batch-Sequential Designs for Estimating Process Features of Expensive Simulators
Conception séquencée par lots pour l'estimation des caractéristiques du processus des simulateurs coûteux
Pritam Ranjan, *Acadia University*, Derek Bingham, *Simon Fraser University*

In many branches of physical science, when complex physical phenomena are very expensive or time-consuming to observe, observations from their corresponding computer models are used instead. In such experiments, scientists are often interested in identifying different pre-specified process features (e.g., process optima, threshold values and quantiles). In this talk, I will outline batch sequential design strategies for selecting trials which are aimed at the estimation of these pre-specified process features. Furthermore, I will discuss the involved optimization issues, where parallel programming approach turns out to be beneficial.

Dans de nombreuses branches des sciences physiques, quand l'observation de phénomènes physiques complexes coûte très cher et demande beaucoup de temps, on utilise plutôt des observations provenant des modèles informatiques correspondants. Dans ce genre d'expérience, les scientifiques cherchent souvent à découvrir différentes caractéristiques préspécifiées du processus (p.ex., optima, valeurs seuils et quantiles du processus). La communication décrira dans les grandes lignes les stratégies de conception séquentielle par lots pour la sélection d'essais dont le but est d'estimer les caractéristiques préspécifiées des processus. Seront aussi abordées les questions d'optimisation soulevées pour lesquelles une approche de programmation parallèle s'avère favorable.

Session 14B Mer. 28 mai • Wed. May 28, 14:35 – 16:05 Capital Hall 2A

Calibration and Variance Estimation in Surveys
Calage et estimation de la variance dans les enquêtes

[MS-279]
Bootstrap for Multistage Sampling
Bootstrap pour un tirage à plusieurs degrés
Guillaume Chauvet, *CREST-ENSAI*

We propose a Bootstrap technique for multistage sampling. Our algorithm consists of a correction of the method originally proposed by Gross (1980) for simple random sampling, and known not to be consistent for multistage sampling. The idea consists of building a pseudo-population of pseudo Primary Sampling Units, based on the original sample. The first stage drawing is performed, but the second stage drawing is corrected. Two-stage sampling with simple random sampling at each stage and self-weighted two-stage sampling are covered by the proposed Bootstrap technique. The method is evaluated through a limited set of simulations.

Nous proposons une méthode de bootstrap permettant de traiter le cas d'un échantillonnage à plusieurs degrés. Elle adapte la méthode proposée pour le sondage aléatoire simple par Gross (1980). L'idée consiste à construire à partir de l'échantillon une population constituée de pseudo unités primaires dans laquelle on reproduit le tirage du 1er degré, mais pour laquelle le tirage du 2^e degré est modulé. Le bootstrap proposé couvre les cas particuliers importants du tirage avec sondage aléatoire simple à chaque degré et celui du tirage autopondéré. La méthode est évaluée à l'aide de simulations.

[MS-280]
Successive or Iterated Calibrations
Calages successifs ou itérés

Jean-Claude Deville, *ENSAI/CREST*

Sometimes several successive calibration reweightings are performed for estimating survey data. We make precise how, at each step, the effects of the preceding calibrations are damaged. It is also possible to iterate the successive calibrations by performing the process again. A simple sufficient condition is given for convergence. The converged estimator benefits from the information contained in all auxiliary variables. A necessary and sufficient condition allows detecting the cases of non-convergence where variance can increase indefinitely.

Dans une procédure d'estimation, on procède parfois à plusieurs repondérations successives. On précise en quoi chaque calage détériore les améliorations de précision recherchées par les précédents. On peut aussi itérer les calages en recommençant la procédure sur le premier ensemble de variables, puis le second. Si le processus converge, l'estimation bénéficie de l'information apportée par toutes les variables auxiliaires. On établit une condition suffisante simple pour qu'il en soit ainsi. Une condition nécessaire et suffisante plus complexe permet de déceler les cas défavorables de non-convergence et de variance croissante.

[MS-281]
Linearization and Composite Estimation for Two-sample Survey Data
Linéarisation et estimation composite dans les enquêtes à deux échantillons

Anne Ruiz-Cazen, *TSE (Gremaq et IMT)*, Camélia Goga, *IMB*, Jean-Claude Deville, *ENSAI/CREST*

In two-sample surveys, interest may lie in estimating change of nonlinear functions of finite population totals such as ratios, correlation coefficients or measures of inequality like the Gini index. For one-sample surveys, the linearization method based on the influence function is a powerful tool for variance estimation of nonlinear statistics. In the two-sample case, estimating the variance for such complex statistics has to take into account the existence of overlapping samples. Our presentation consists of giving a two-sample linearization technique based on partial influence functions and of studying composite estimation that take into account overlapping samples.

Dans une enquête portant sur deux échantillons, l'intérêt peut porter sur l'estimation de l'évolution de fonctions de totaux de population telles que des ratios, des coefficients de corrélation ou des mesures d'inégalités comme l'indice de Gini. Dans le cas d'un seul échantillon, la méthode de linéarisation basée sur la fonction d'influence est un outil puissant pour l'estimation de variance de statistiques non linéaires. Dans le cas de deux échantillons, l'estimation de la variance de ces statistiques complexes doit prendre en compte la présence d'échantillons chevauchants. Notre présentation consiste à proposer une méthode de linéarisation basée sur la fonction d'influence partielle et à étudier l'estimation composite qui tient compte des échantillons chevauchants.

Session 14C Mer. 28 mai • Wed. May 28, 14:35 – 16:05 Capital Hall 3A

Environmental Statistics
Statistiques en environnement

[MS-282]
Gauging Ecosystem Health with Latent Health Factor Models
Évaluation de la santé des écosystèmes à l'aide de modèles avec déterminant de la santé latent

Grace Chiu, *University of Waterloo*, Peter Guttorp, *University of Washington*, Anton Westveld, *University of Nevada, Las Vegas*

We propose a model-based approach for constructing ecological health indices through statistical inference. Our latent health factor index (LHFI) is obtained by estimating an unobservable health factor term in a mixed-effects ANOCOVA that directly models the relationship among indicator variables (or metrics) and health. Unlike conventional indices (e.g., IBI and O/E index) that rely on domain-specific calibrations of metrics against reference conditions whose non-constancy is largely unaccounted for, our methodology (a) involves no explicit reference conditions while metrics are intrinsically "calibrated" in the context of multiple comparisons, and (b) naturally incorporates spatio-temporal influences on calibration schemes.

Nous proposons une approche fondée sur un modèle de la construction d'indices de la santé écologique par inférence statistique. Notre indice à déterminant latent de la santé (LHFI pour Latent Health Factor Index) est obtenu en estimant un terme de déterminant de la santé inobservable dans une analyse de covariance avec effets mixtes qui modélise directement la relation entre les variables indicatrices et la santé. Contrairement aux indices classiques (par exemple IBI et indice O/E) qui s'appuient sur des étalonnages des mesures par domaine en fonction de conditions de référence dont la convergence n'est, en grande partie, pas prise en compte, notre méthode a) ne fait intervenir aucune condition de référence explicite, les mesures étant « étalonnées » intrinsèquement dans le contexte de comparaisons multiples et b) intègrent naturellement les influences spatio-temporelles sur les scénarios d'étalonnage.

[MS-283]
Estimation for Differential Equation Based Environmental Models
Estimation pour modèles environnementaux basés sur des équations différentielles

Michael Dowd, *Dalhousie University*

In this talk I consider the estimation problem for differential equation (DE) based environmental models. The particular emphasis is DE based numerical models of ocean biogeochemistry using available non-Gaussian measurements. The specific problem considered is estimating the system state and parameters for time dependent nonlinear stochastic dynamics. These estimation, or data assimilation, methods rely on Monte Carlo solutions implemented via sampling or ensemble methods. State estimation uses resampling/bootstrap and MCMC (or combinations thereof); the associated parameter estimation problem is treated with likelihood and state augmentation approaches. Challenges for adaptation to large dimension systems are discussed.

Dans cet exposé, je considère le problème d'estimation de modèles à base d'équations différentielles (ED), plus particulièrement des modèles numériques à base d'ED modélisant la biogéochimie océanique utilisant des mesures disponibles non gaussiennes. Le problème spécifique considéré est l'estimation de l'état et des paramètres du système avec une dynamique stochastique non linéaire dépendant du temps. Ces méthodes d'estimation ou d'assimilation de données s'appuient sur des solutions de Monte-Carlo mises en œuvre au moyen de méthodes d'échantillonnage ou d'ensemble. L'estimation de l'état du modèle s'appuie sur le rééchantillonnage/bootstrap ou les MCMC ; le problème associé d'estimation des paramètres est traité avec des approches de vraisemblance et d'augmentation d'état. Nous discutons des défis propres à l'adaptation des méthodes à des systèmes de dimension élevée.

[MS-284]

Air Quality Mapping using Simulated and Observed Data

Utilisation de données modélisées et observées pour cartographier la pollution atmosphérique

Laure Malherbe, *Institut National de l'Environnement Industriel et des Risques (INERIS)*, Frédérik Meleux, Laurence Rouïl & Anthony Ung, *INERIS*, Chantal de Fouquet, *Ecole Nationale Supérieure des Mines de Paris / Centre de géosciences*

Since 2003, the modelling system PREV'Air, which is a co-operative initiative of several French public institutions, has been delivering daily short-term forecasts and simulated maps of atmospheric pollutants over Europe and France. Every morning, the maps of ozone and PM10 simulated by the chemistry-transport model CHIMERE are adjusted using the observation data coming from the monitoring stations. The analysis technique used consists of adding a corrective field estimated by geostatistical interpolation to the model output. This study investigates some methodological issues raised by this process and aims to develop a correction model appropriate to each pollutant.

Depuis 2003, le système de modélisation PREV'Air, fruit d'une collaboration entre plusieurs organismes publics, diffuse quotidiennement des prévisions à court terme et des cartographies de polluants atmosphériques sur l'Europe et la France. Chaque matin, les cartographies d'ozone et de PM10 de la veille simulées par le modèle de chimie-transport CHIMERE sont ajustées à l'aide des données de stations de mesure. La technique d'analyse employée consiste à ajouter au modèle un champ correctif estimé par interpolation géostatistique. Ce travail s'attache aux problèmes de méthode posés par cette procédure et cherche à définir un modèle de correction adapté à chaque polluant.

Session 14D Mer. 28 mai • Wed. May 28, 14:35 – 16:05 Capital Hall 4A

Likelihood Inference

Inférence par vraisemblance

[MS-285]

Accurate Likelihood Inference for Overdispersed Data

Inférence par vraisemblance pour les données sur-dispersées

Anthony Davison, & Claudia Lozada-Can, , *Ecole Polytechnique Fédérale de Lausanne (EPFL)*

Data that are overdispersed relative to a simple parametric model arise in many applications, and there is a large body of literature on suitable models for such data. Inference in such cases typically relies on large-sample, first order likelihood theory. This talk will outline an approach to higher order likelihood inference for such data, based on a form of approximate semiparametric likelihood constructed from a mixture model. Applications will be described and some examples given.

Dans la pratique on rencontre souvent des données qui sont surdispersées par rapport à un modèle paramétrique simple, et il existe beaucoup d'approches pour traiter cette surdispersion. Dans ce cas l'inférence statistique se base souvent sur les approximations de premier ordre pour des quantités standards de vraisemblance. Dans cette intervention on considèrera la construction des inférences plus précises en utilisant une vraisemblance semiparamétrique approximée. Des applications et des exemples des idées seront exposées.

[MS-286]
Composite Likelihood
Vraisemblance composite

Nancy Reid & Jin Zi, *University of Toronto*, Grace Yi, *University of Waterloo*

In complex multivariate settings, it is sometimes much simpler to construct a 'likelihood' from lower dimensional marginal or conditional distributions: the general construction was called composite likelihood in Lindsay (1988). Under mild conditions the resulting composite maximum likelihood estimate is consistent and asymptotically normal, with asymptotic variance of a 'sandwich' form. The likelihood ratio type statistic has as limiting distribution a weighted sum of chi-squares. Composite likelihood can be surprisingly efficient; I will discuss ongoing research to understand this through asymptotic analysis. Special attention will be focussed on the so-called pairwise likelihood formed by taking the product of bivariate marginals.

Dans des paramètres multivariés complexes, il est parfois beaucoup plus simple de construire une « vraisemblance » à partir de distributions marginales ou conditionnelles de dimension moindre : la construction générale était nommée vraisemblance composite chez Lindsay (1988). Dans des conditions bénignes, l'estimation du maximum de vraisemblance composite résultant est asymptotiquement normale et convergente avec une variance asymptotique de forme « intercalée ». La statistique de type de rapport de vraisemblance statistique a comme distribution limite une somme pondérée des chi carrés. La vraisemblance composite peut être étonnamment efficace. Il sera question de recherches en cours en vue de comprendre cela par l'entremise d'une analyse asymptotique. Une attention particulière sera accordée à ce qu'on appelle la vraisemblance par paire formée en utilisant le produit des distributions marginales bivariées.

[MS-287]
Bayesian Nonparametric Approaches with Partially Improper A Priori
Approches bayésiennes non paramétriques utilisant des *a priori* partiellement impropres

Judith Rousseau, *CEREMADE, Université Paris Dauphine et CREST*, Dongchu Sun, *University of Missouri, Columbia*

In Bayesian nonparametric approaches, it is usually highly recommended to use proper priors since a finite number of observations has little chance to overcome the impropriety of a prior constructed on an infinite dimensional space to make it proper. However, in some cases the parameter space is naturally divided into an infinite dimensional space and a finite dimensional space (as in semi-parametric models). It might then be of interest to use improper priors on the finite dimensional space. In this work, we give general conditions to obtain the same rate of concentration of the posterior distribution around the true value as in the case of a fully proper, well-behaved prior. The special case of spline approaches is studied.

Il est souvent vivement recommandé, dans les approches bayésiennes non paramétriques, de n'utiliser que des lois a priori propres. Cependant, dans les modèles de type semi-paramétrique, il peut être intéressant d'envisager des lois impropres sur la partie paramétrique du paramètre. Dans ce travail, nous déterminons des conditions générales sur le modèle et sur la loi a priori pour que la loi a posteriori se comporte bien. Nous étudions la vitesse de concentration de la loi a posteriori autour de la vraie valeur lorsqu'une loi partiellement impropre est utilisée. Nous illustrons nos résultats en étudiant les modèles splines.

Session 14E Mer. 28 mai • Wed. May 28, 14:35 – 16:05 Capital Hall 5A

Model Assessment and Goodness-of-fit
Évaluation de modèles et adéquation

[MS-288]

A-dependence Statistics for Mutual and Serial Independence of Categorical Variables
Les statistiques de dépendance A pour l'indépendance mutuelle et sérielle entre variables catégorielles

Martin Bilodeau, *Université de Montréal*, Pierre Lafaye de Micheaux, *Université Pierre Mendès France*

The Möbius transformation of probability cells in a multi-way contingency table is used to partition the Pearson chi-square test of mutual independence into A-dependence statistics. A similar partition is proposed for a universal and consistent test of serial independence in a stationary sequence of a categorical variable. The partition proposed can be adapted whether using estimated or theoretical marginal probabilities. With the aim of detecting a dependence of high order in a long sequence, A-dependence terms of the partition measuring increasing lagged dependences can be combined in a Box – Pierce type test of serial independence. A real data analysis of a nucleotide sequence using the Box – Pierce type test is provided.

La transformation de Möbius des probabilités des cellules d'un tableau de contingence multidimensionnel est utilisée pour obtenir une décomposition du khi-deux de Pearson, pour l'indépendance mutuelle, en composantes mesurant la dépendance A. Une décomposition similaire donne un test universel et consistant pour l'indépendance sérielle d'une suite stationnaire d'une variable catégorielle. Cette décomposition peut être utilisée que les probabilités marginales soient théoriques ou estimées. Pour les suites longues exhibant une dépendance sérielle d'un ordre élevé, on propose un test similaire au test de Box-Pierce. Ce test est effectué sur une longue série de nucléotides lors d'une application en génétique.

[MS-289]

Goodness-of-Fit Tests for Distributions on Fractafolds Based on the Sierpinski Gasket
Tests d'adéquation à des distributions sur des "fractafolds" basées sur le tamis de Sierpinski

Jean-Renaud Pycke, *Université d'Evry*

Watson's goodness-of-fit test for uniformity on a circle is closely related to the spectrum of the Laplacian of the corresponding underlying Riemannian manifold. Some key properties of this quite simple spectrum can be more clearly understood when we see the circle as the limit of circle graphs and using the so-called spectral decimation method. This interpretation is fruitful since it enables us to adapt some goodness-of-fit tests on the circle to less trivial fractafolds as those based on the Sierpinski gasket.

Le test de Watson pour l'adéquation à l'uniformité sur le cercle est lié au spectre du laplacien de la variété Riemannienne sous-jacente correspondante. Des propriétés clées de ce spectre assez simple peuvent être mieux comprises en voyant le cercle comme limite de graphes circulaires et en utilisant la méthode dite de décimation spectrale. Cette interprétation est féconde, car elle permet d'adapter des tests d'adéquation sur le cercle à des « fractafolds » moins triviales telles celles construites à partir du tamis de Sierpinski.

[MS-290]
On Two Tests for Exponentiality with Different Motivations
Deux tests pour la distribution exponentielle avec motivations différentes

Michael Stephens & Richard Lockhart, *Simon Fraser University*

A method of testing for the exponential distribution is based on a comparison of the empirical distribution function (EDF) of the sample with the fitted cumulative distribution function (CDF). This method will be called the classical method. An alternative method arises in a natural way in reliability and survival analysis-called the total time on test method. It has been observed that these apparently different procedures give very similar powers. This phenomenon is examined.

Une méthode de test pour la distribution exponentielle est fondée sur une comparaison entre la fonction de distribution empirique de l'échantillon et la fonction de répartition ajustée. Nous l'appellerons méthode classique. Une autre méthode, appelée méthode du temps d'observation, se dégage d'une manière naturelle en analyse de fiabilité et de survie. Il a été constaté que ces procédures en apparence différentes donnent des puissances fortes semblables. Nous examinons ce phénomène.

Session 14F Mer. 28 mai • Wed. May 28, 14:35 – 16:05 Capital Hall 6A

Stochastic Models
Modèles stochastiques

[MS-291]
Modelling Game Outcomes of the Brazilian 2006 Series A Championship as Ordinal-valued
Modélisation des résultats du championnat brésilien de série A comme valeurs ordinales

David Brillinger, *University of California, Berkeley*

"In football, everything is complicated by the presence of the opposing team."
Sartre, J.-P. (2004). *Critique of Dialectical Reason*, Volume One. NY: Verson. (Originally written in French in 1960).

« Au football, tout est compliqué par la présence de l'équipe adverse. »
J.-P. Sartre (1960). *Critique de la raison dialectique*. Gallimard, Paris.

[MS-292]
Inference on Dynamic Graphs
Inférence dans les graphiques dynamiques

Yasaman Hosseinkashi, Shoja'eddin Chenouri & Christopher G. Small, *University of Waterloo*

A stochastic graph is a sequence of random graphs evolving over time. Many efforts have been made to model features of real world dynamic graphs. This work considers the estimation for these models. The dynamic properties are formulated with parameters for vertex and degree evolution. The method of moments and pseudo-likelihood are used to estimates the parameters of these models.

Un graphe stochastique est une série de graphes aléatoires qui évoluent au cours du temps. De nombreux efforts ont été déployés en vue de modéliser les caractéristiques des graphiques dynamiques de situations réelles. La présente étude porte sur l'estimation de ces modèles. Les propriétés dynamiques sont formulées à l'aide de paramètres pour l'évolution du degré des sommets. Nous utilisons la méthode des moments et la pseudo-vraisemblance pour estimer les paramètres de ces modèles.

[MS-293]
Inference Problem in some Predator-prey Systems
Problème d'inférence dans certains systèmes proie-prédateurs

Sévérien Nkurunziza, *University of Windsor*

We consider an inference problem concerning the interaction parameters of a Lotka-Volterra system of ordinary differential equations (L-V-ODE). As in Froda and Nkurunziza (2007), we consider the stochastic model where the actual population sizes are viewed as random perturbations of the solutions to the deterministic L-V-ODE. To this end, these perturbations are assumed to follow correlated Ornstein-Uhlenbeck processes. Thanks to a certain re-parametrisation of L-V-ODE, we establish a likelihood ratio test. Moreover, we study the asymptotic properties of this test. The performance of this test is illustrated through real data sets and some simulation studies.

On considère un problème d'inférence pour les paramètres d'interaction du système équations différentielles ordinaire de Lotka-Volterra (EDO-L-V). Comme dans Froda et Nkurunziza (2007), nous considérons le modèle stochastique où des tailles de population réelles sont perçues comme des perturbations aléatoires des solutions de l'EDO-L-V. À ce sujet, ces perturbations sont supposées suivre des processus d'Ornstein-Uhlenbeck corrélés. À l'aide d'une reparamétrisation de l'EDO-L-V, nous proposons le test du rapport de vraisemblance. De plus, nous étudions les propriétés asymptotiques de ce test. La mise en application dudit test est illustrée à l'aide d'une analyse des données réelles et simulées.

[MS-294]
Strategic Misspecification in Discrete Choice Models
Erreur de spécification dans les modèles de choix discrets

Dhouha El Haj Ali, *Université Manar I, Tunisie*

Strategic interaction is a fundamental consideraion in the study of making choices. Multinomial and binomial logit and probit econometric models have become commonplace in studies on decision making. These models have been strongly criticized by most of the recent econometric works that deal with discrete choice models. By showing that discrete choice binary model regression — where the latent variable is a linear function related to the estimation parameters and the regressors — is structurally non-compatible with strategic interaction. We characterize the specification error resulting from the use of traditional models in order to analyse strategic data. We construct a strategic model where the probability of occurrence for each action is monotone compared to the regressors. We demonstrate that even under these ideal conditions for the use of logit and probit, — in the analysis of strategic actions — these classic models provide non-convergent estimaters.

L'interaction stratégique est une notion fondamentale dans l'étude de la prise de décision. Les modèles économétriques utilisés dans la plupart des études de comportement sont le probit et le logit binaires ou multinomiaux. Ces modèles ont été fortement critiqués dans la plupart des travaux économétriques récents qui portent sur les modèles de choix discrets. En montrant que la régression des modèles de choix discrets binaires – où la variable latente est une fonction linéaire par rapport aux paramètres à estimer et aux regrésseurs – est structurellement non compatible avec l'interaction stratégique. Nous caractérisons l'erreur de spécification résultante de l'utilisation des modèles traditionnels pour analyser des données stratégiques. Nous construisons un modèle stratégique ou toute probabilité de réalisation de chaque action est monotone par rapport aux regrésseurs. Nous montrons que même sous ces conditions idéales d'utilisation de logit ou probit,– en cas d'analyse des actions stratégiques – ces modèles classiques fournissent des estimateurs non convergents.

[MS-295]
Strauss Processes: A New Space-filling Design for Computer Experiments
Planification d'expériences numériques à partir du processus ponctuel de Strauss

Jessica Franco, *TOTAL EP/GSR/TG/G&I*, Xavier Bay, *EMSE/G2I/3MI*, Bernard Corre, *TOTAL EP/GSR/COP/EM*, Delphine Dupuy, *EMSE/G2I/3MI*

In the computer experiments setting, Space-Filling Designs (SFDs) are widely used to explore the complex relationship between inputs and outputs. In this communication, a new SFD is initially defined with the help of the Strauss process. Through Markov chain Monte-Carlo (MCMC) methods, more general Gibbs processes can be used to achieve different goals. We will see that it is easy to sample over the entire range of each input variable as Latin hypercubes do it. Moreover, non-homogeneous designs can be constructed to account for a priori information.

Dans le contexte de la planification d'expériences numériques où la relation précise entre les entrées et les sorties du code est inconnue, il est préférable d'utiliser des « Space-Filling Designs » (SFDs). Nous présenterons un nouveau type de SFD basé sur le processus ponctuel de Strauss. Ces plans obtenus à partir de méthodes Monte-Carlo par chaîne de Markov (McMC) sont facilement modulables et peuvent ainsi répondre à divers objectifs. Signalons en particulier les bonnes propriétés de projection des expériences sur les axes factoriels à l'image des hypercubes latins. Il est possible également de générer des points de manière hétérogène si l'on dispose d'informations *a priori*.

[MS-296]
Age Distribution Model for a Developing Nation
Modèle de répartition par âge pour un pays en développement

Shafiqur Rahman, *Sultan Qaboos University*

Efficient estimates of the proportions of populations at different age levels are essential for making a quality budget for any developing or developed nation. These are obtained from the best-fitted model. Applying chi-square goodness of fit test and model selection criteria it is observed that the age distribution of the PNG population is exponential. Age distribution of PNG population is compared with that of other developed nations and observed that the age distribution of PNG population did not change significantly over the last five decades and the median age of PNG population is about half of that of other developed nations.

L'estimation efficace des proportions de population appartenant à divers groupes d'âge est essentielle à une bonne budgétisation pour tout pays en développement ou développé. Ces proportions sont obtenues à l'aide du modèle le mieux ajusté. L'application du test de qualité de l'ajustement du khi-deux et des critères de sélection du modèle permet de voir que la répartition par âge de la population de la Papouasie–Nouvelle-Guinée (PNG) est exponentielle. Nous comparons la répartition par âge de la population de la PNG à celle d'autres pays développés et constatons qu'elle n'a pas varié de manière significative au cours des cinq dernières décennies et que l'âge médian de cette population est environ deux fois plus faible que celui observé pour d'autres pays développés.

Session 14G Mer. 28 mai • Wed. May 28, 14:35 – 16:05 Capital Hall 7A

Analysis of Microarrays
Analyse de biopuces

[MS-297]

Stratified False Discovery Rates and Q-values for Gene Expression Microarrays
Taux de fausses découvertes stratifiées et valeurs de q pour l'expression génétique en microréseaux

Shaheena Bashir & Angelo Canty, *McMaster University*

The multiple testing problem in gene expression micorarrays is often handled using a False Discovery Rate (FDR) approach. In many genetic studies, there can be a natural stratification of the hypotheses to be tested. FDR can then be controlled separately within each stratum. This can lead to increased power to detect differential expression and a refined estimate of FDR and q-values in the stratum of primary interest which is expected to have a higher number of true positives. We present a simulation study and applications of the method to rodent models for Type 1 Diabetes using congenic strains.

Les multiples problèmes de tests dans l'expression génétique en microréseaux sont souvent traités en utilisant l'approche du taux de fausses découvertes (TFD). Dans un bon nombre d'études génétiques, il y a une stratification naturelle de l'hypothèse à tester. Le TFD peut être contrôlé séparément dans chaque strate. Cela peut mener à une augmentation de la puissance de détection des expressions différentielles et à une estimation perfectionnée du TFD et des valeurs de q dans la strate d'intérêt primaire dont on s'attend à avoir un grand nombre de vrais positifs. Nous présentons une étude de simulation et les applications de la méthode à des modèles rongeurs pour le diabète de type 1 en utilisant des souches congéniques.

[MS-298]

Applications of Intra-class Correlation-based Criteria in Microarray Data Analysis
Applications d'une corrélation intra-classe fondée sur l'analyse des données des puce à ADN

Wenqing He, *University of Western Ontario*, Shelley Bull, *University of Toronto*

Microarray technology has become a widely-used tool in medical research. Microarray experiments consist of complicated procedures, and errors can be introduced in any of the steps. Such errors and quality assessment information of the experiments need to be taken in to account when analyzing the data. Microarray data are essentially measurements of genomic features, such as gene expression, made via genetic markers. The intra-class correlation (ICC) can fruitfully be employed in microarray data analysis. In this talk, we will present modified ICC criteria and demonstrate how they can be used in different stages of microarray data analysis.

La technologie des microréseaux est devenue un outil très répandu en recherche médicale. Les expériences portant sur ces microréseaux comprennent des procédures compliquées et des erreurs peuvent être commises à n'importe quelle étape. Ce genre d'erreurs et les données d'évaluation de la qualité des expériences doivent être prises en compte dans l'analyse des données. Les données de microréseaux sont essentiellement des mesures de caractéristiques génomiques, comme l'expression des gènes, faites à l'aide de marqueurs génétiques. La corrélation intra-classe (CIC) peut être employée fructueusement dans l'analyse des données de microréseaux. Dans la communication, nous présenterons des critères de corrélation intra-classe modifiés et montrerons comment ils peuvent être utilisés à diverses étapes de l'analyse des données de microréseaux.

[MS-299]
Improving Supervised Classification for High Dimensional Data by Adding External Information. Application to Microarray data.
Améliorer les procédures de classification supervisée en intégrant de l'information externe. Application à des données génomiques.

Maela Kloareg & David Causeur, *Agrocampus Rennes*

Supervised classification is a common problem in high dimensional and low sample size situations as, for example, microarray data. Accuracy of the standard techniques is compromised in the case of highly correlated variables. We compare the results of a few classification methods on simulated data under various correlation schemes by focusing on the misclassification rate and suggest improvements based on external information. Finally we apply this method to a real life microarray data set.

La construction de règles d'affectation d'un individu à un groupe prédéfini est une problématique récurrente dans le cas de données en grande dimension, telles que les données postgénomique par exemple, en particulier dans le cas d'une structure de dépendance entre les variables. Nous comparons ici les résultats obtenus par plusieurs méthodes sur des données simulées en étudiant les taux de mauvais classement et proposons d'améliorer ces procédures par l'intégration d'information externe. Enfin, nous mettons en application ces méthodes sur un jeu de données réel obtenu à partir de biopuces.

[MS-300]
Joint Segmentation of Processes: Application to the Analysis of Multiple CGH Profiles
Segmentation simultanée de signaux : application aux données de microarrays CGH

Baba Thiam, Stéphane Robin & Emilie Lebarbier, *UMR INA P-G/ENGREF/INRA MIA 518,* Franck Picard, *UMR CNRS-5558/Laboratoire de Biométrie et Biologie Evolutive, Université Lyon 1*

The purpose of the CGH experiments (Comparative Genomic Hybridation) is to detect chromosomal aberrations. A CGH profile can be viewed as a succession of segments. The purpose of the statistical analysis is to detect them. The objective of the CGH arrays experiments now has to deal with several samples simultaneously. To analyze the profiles, we propose a linear model with breakpoints which takes into account a position effect. The model parameters are estimated by maximum likelihood combining least squares with a dynamic programming algorithm. An application is done for the analysis of CGH data on bladder cancer.

L'objectif des expériences de microréseaux CGH (Comparative Genomic Hybridation) est de détecter des aberrations chromosomiques. Un profil CGH peut être vu comme une succession de segments. L'analyse statistique consiste donc à les détecter. Les nouvelles expériences de microréseaux CGH amènent désormais à analyser plusieurs échantillons biologiques. Pour analyser simultanément ces profils, nous proposons un modèle linéaire avec ruptures qui prend en compte un effet position. Les paramètres sont estimés par maximum de vraisemblance alternant les moindres carrés avec un algorithme de programmation dynamique. Cette méthode est appliquée pour l'analyse de données CGH du cancer de la vessie.

[MS-301]
Improved Estimation for cDNA Microarray Images in Wavelet Domain
Estimation améliorée pour les images de microréseaux d'ADNc dans le domaine des ondelettes

Tamanna Howlader & Yogendra P. Chaubey, *Concordia University*

Complementary DNA (cDNA) microarray experiments involve steps that result in high level of noise in the red and green channel images. Noise removal is crucial for obtaining accurate log-intensity ratios. The wavelet transform has shown signifi-

Les expériences portant sur des micropuces à ADN complémentaire (ADNc) comprennent des étapes qui produisent un bruit élevé dans les canaux rouge et vert des images. L'élimination de ce bruit est essentielle à l'obtention de ratios des logarithmes d'intensité exacts. La transformée en

cant success in the denoising of cDNA microarray images. However, existing wavelet-based methods process each image individually. This paper describes a noise reduction technique that exploits the dependency between wavelet transform coefficients of the two channels by using a locally-adaptive joint statistical model. Experiments show that the proposed method provides improved denoising performance in terms of mean squared error and yields more accurate log-intensity ratios.

ondelettes a donné des résultats prometteurs dans le débruitage des images de micropuces à ADNc. Toutefois, dans les méthodes d'ondelettes existantes, chaque image est traitée individuellement. La présente communication décrit une méthode de réduction du bruit qui tire parti de la dépendance entre les coefficients d'ondelettes des transformées des deux canaux en utilisant un modèle statistique conjoint localement adaptatif. Des expériences montrent que la méthode proposée a de meilleures propriétés de débruitage pour ce qui est de l'erreur quadratique moyenne et produit des ratios des logarithmes d'intensité plus exacts.

[MS-302]

Assessing Inter-rater Agreement for Multiple Outcomes in Tissue Microarrays
Évaluer l'accord inter-juges pour des résultats multiples sur une micropuce de tissus

Jennifer Asimit, *Samuel Lunenfeld Research Institute, University of Toronto,* Neil Klar, *University of Western Ontario,* Shelley Bull, *Samuel Lunenfeld Research Institute, University of Toronto*

Studies of agreement between measurements taken from the same subject commonly occur in medical and psychiatric research. An interest in inter-rater agreement for multiple outcomes stems from the study of tissue microarrays (TMAs), which involve binary measurements taken from multiple genetic markers. We propose a method for assessing inter-rater agreement between two raters across a number of multiple binary outcomes measured on each subject, which allows for covariate-adjusted estimates. Based on the model, tests of homogeneity are performed and a decision on the appropriateness of pooling can be made. The properties of this method are examined via a simulation study.

Les études d'accord entre les mesures prises sur le même sujet sont fréquentes en recherche médicale et psychiatrique. L'étude de micropuces de tissus (MPT), qui implique des mesures binaires prises sur plusieurs marqueurs génétiques, est une source d'intérêt pour l'accord inter-juges pour des résultats multiples. Nous proposons une méthode pour évaluer l'accord inter-juges entre deux juges sur plusieurs résultats binaires mesurés sur chaque sujet, permettant des estimés ajustés par les covariables. En se basant sur ce modèle, des tests d'homogénéité sont effectués et une décision sur la justesse de la mise en commun peut être prise. Les propriétés de cette méthode sont évaluées au moyen d'une étude de simulation.

Session 14H Mer. 28 mai • Wed. May 28, 14:35 – 16:05 Congress Hall A/E West

Poster Session III
Séance d'affichage III

[MS-303]

One-armed Bandit Processes with Observations from the Erlang(k) Distribution
Processus du bandit manchot avec observations provenant de la distribution Erlang(k)

You Liang, *University of Manitoba*

We consider the one-armed bandit process with observations from an Erlang(k) distribution. By the Bayesian approach, an algorithm is presented to calculate the sequence of break-even values characterizing the optimal selections of arms. An opti-

Nous considérons le processus du bandit manchot avec observations provenant d'une loi d'Erlang(k). Un algorithme d'approche bayésienne est présenté pour calculer la suite des seuils de rentabilité caractérisant la sélection optimale de bras. Le problème de la décision optimale pour un pro-

mal decision problem of a special one-armed bandit process with unrestricted switching times is also investigated, whose random sampling times have a negative exponential distribution and sampling values have an Erlang(2) distribution. The monotonicity of the Gittins index of this process is discussed, based on which the optimal decision problem is simplified to an optimal stopping problem.

cessus du bandit manchot particulier à temps de changement non restreint est aussi étudié. Les temps observés dans ce problème ont une distribution exponentielle négative et les valeurs observées ont une distribution Erlang(2). Nous discutons de la monotonicité de l'indice de Gittins de ce processus, qui ramène le problème de la décision optimale à un problème d'arrêt optimal.

[MS-304]
Optimal Investment-Consumption Problem with Hidden Markov Model
Problème d'investissement-consommation optimal avec modèle de Markov caché

Yi Hua & Xikui Wang, *University of Manitoba*

The investment-consumption problem modeled by the hidden Markov process is studied. A financial market with one risky and one risk-free assets is considered. Decisions are made at the beginning of each period on the proportion of the remaining wealth to be invested in the risky asset. Risk-free asset return is assumed to be constant, but random return of the risky asset depends on the market environment, which is characterized by a hidden discrete-time Markov chain. We formulate the investor's decisions as an optimal stochastic control problem. The solution of the optimal strategy is characterized and some simulation results are shown.

Nous étudions le problème d'investissement-consommation modélisé par le processus markovien caché. Nous considérons un marché financier constitué d'un actif risqué et d'un actif sans risque. Au début de chaque période, des décisions sont prises quant à la proportion de la richesse restante qui sera investie dans l'actif risqué. Nous supposons que le rendement de l'actif sans risque est constant, mais que le rendement aléatoire de l'actif risqué dépend de la conjoncture du marché, qui est caractérisée par une chaîne de Markov à temps discret cachée. Nous formulons les décisions de l'investisseur sous forme d'un problème de contrôle stochastique optimal. Nous caractérisons la solution de la stratégie optimale et présentons certains résultats de simulation.

[MS-305]
Simulation of Exotic Options
Simulation d'options exotiques

Tingting Gou, W. John Braun & Duncan J. Murdoch, *University of Western Ontario*

We consider the problem of simulating the times of events such as extremes and barrier crossings in diffusion processes. We develop a rejection sampler based on Shepp (1979) for simulating an extreme of a Brownian motion, and use it in a general recursive scheme for more complex simulations, including simultaneous simulation of the minimum and maximum and application to more general diffusions. We price exotic options that are difficult to price analytically: a lookback option that includes discounting at the risk-free interest rate, and a chooser option where the choice is made at the time of a barrier crossing.

Nous considérons le problème de la simulation de temps d'événements tels les extrêmes et les franchissements de barrière dans des processus de diffusion. Nous développons un échantillonneur acceptation/rejet basé sur Shepp (1979) pour simuler un extrême du mouvement brownien, et l'utilisons dans un schéma récursif général pour des simulations plus complexes, incluant la simulation simultanée du maximum et du minimum et une aplication à des diffusions plus générales. Nous fixons le prix d'options exotiques dont il est difficile de fixer le prix analytiquement : une option rétroviseur incluant l'actualisation au taux d'intérêt sans risque, et une option double où le choix est fait au temps de franchissement de barrière.

[MS-306]
Binary Adaptive Rejection Sampling
Échantillonnage de rejet adaptatif binaire

Duncan Murdoch & Tingting Gou, *University of Western Ontario*

In this paper, we consider the problem of sampling from a given probability density function where traditional rejection sampling is inefficient. Given bounds on the density in intervals (or in higher dimensions, bounds within hyper-rectangles), our algorithm adapts its proposals to reduce the rejection rate. We illustrate with several examples, including simultaneous simulation of extremes and their times in a Brownian motion.

Nous considérons le problème de l'échantillonnage à partir d'une densité de probabilité dans le cas où l'échantillonnage par rejet classique est inefficace. Sachant les bornes sur la densité dans les intervalles (ou, pour des dimensions plus grandes, les bornes à l'intérieur d'hyper-rectangles), notre algorithme adapte ses propositions de manière à réduire le taux de rejet. Nous donnons plusieurs exemples, y compris la simulation simultanée des valeurs extrêmes et de leurs temps dans un mouvement brownien.

[MS-307]
Statistical Rainfall-runoff Models based on a Mixture with Heavy-tailed Components
Modèles statistiques de pluie-débit à base de mélange avec composantes à queues lourdes

Julie Carreau & Philippe Naveau, *Laboratoire des Sciences du Climat et de l'Environnement (LSCE-IPSL)*, Eric Sauquet, *Cemagref - Unité de recherche hydrologie-hydraulique*

Rainfall-runoff models are used to predict the distribution of river flows given appropriate meteorological variables. We propose a class of stochastic latent models that better accounts for extreme floods while still giving a sensible model for small to medium discharge. Latent states can capture the different modes of runoff processes. Given the latent state value, the runoff is modelled with a flexible distribution, based on the Generalized Pareto distribution. We evaluate the proposed class of models on the Orgeval basin data. The latent state and the runoff distributions depend on rainfall measurements from the surrounding area.

Les modèles pluie-débit servent à prédire le flux d'une rivière étant donné l'observation de variables météorologiques. Nous proposons une classe de modèles stochastiques avec états latents qui permet de mieux prendre en compte les crues extrêmes tout en modélisant adéquatement les débits faibles et moyens. Les états latents permettent de capter les différents modes des processus de débit. Étant donné la valeur de l'état latent, le débit est modélisé à l'aide d'une distribution flexible qui s'appuie sur la Pareto généralisée. Nous évaluons cette classe de modèles sur les données du basin de l'Orgeval. Les distributions de l'état latent et du débit dépendent des précipitations aux alentours du site.

[MS-308]
A Monte Carlo Simulation Investigation of Latent Variable Growth Models for Interaction Effects
Une étude par simulation de Monte-Carlo de modèles de croissance à variables latentes pour les effets d'interaction

Ian Clara & Cam-Loi Huynh, *University of Manitoba*

Latent growth curves are an effective tool for describing the change or growth of an attribute over time. The current Monte Carlo study investigated three methods (Duncan et al., 1999; Wen et al., 2000; Schumacker, 2000) of representing an interaction effect in latent growth models. The Duncan and Wen models showed a small upward estima-

Les courbes de croissance latente constituent un outil efficace pour décrire l'évolution ou la croissance d'un attribut au cours du temps. L'étude de Monte-Carlo décrite ici avait pour but d'examiner trois méthodes (Duncan *et al.*, 1999 ; Wen *et al.*, 2000 ; Schumacker, 2000) de représentation d'un effet d'interaction dans les modèles de croissance latente. Les modèles de Duncan et de Wen présentaient un

tion bias of the interaction estimate and produced good model fit indices, but showed some problems with converging onto a solution. The Schumacker model showed a downward bias in the latent slope interaction estimate, and also showed a poor fit to the data.

léger biais par excès dans l'estimation de l'interaction et produisaient de bons indices d'ajustement du modèle, mais posaient certains problèmes de convergence vers une solution. Le modèle de Schumacker présentait un biais par défaut dans l'estimation de l'interaction de la pente latente, ainsi qu'un ajustement médiocre aux données.

[MS-309]

Adaptative Linear Models for Regression and Application to Socioeconomic Data
Modèles linéaires adaptatifs en régression et application à des données socioéconomiques

Charles Bouveyron, *Université Paris 1 Panthéon-Sorbonne*, Julien Jacques, *Université Lille 1*

The general setting of regression analysis is to identify a relationship between a response variable Y and one or several explanatory variables X by using a learning sample. In a prediction framework, the main assumption for predicting Y on a new sample is that the regression model $Y = f(X) + \epsilon$ is still valid. Unfortunately, this assumption is not always true in practice. We propose to adapt the original regression model to the new sample by estimating a transformation between the original regression function $f(X)$ and the new one $f^*(X)$. The interest of this work is illustrated by an application in the socioeconomic context.

Le but de la régression est d'identifier une relation entre une variable réponse Y et une ou plusieurs variables explicatives X en utilisant un échantillon d'apprentissage. Une des hypothèses principales pour pouvoir faire de la prédiction sur un nouvel échantillon est que le modèle de régression $Y = f(X) + \epsilon$ est toujours valide. Cependant, cette hypothèse n'est pas toujours vérifiée dans la pratique. Nous proposons dans ce travail d'adapter le modèle de régression initial au nouvel échantillon en estimant une transformation entre la fonction de régression initiale $f(X)$ et la nouvelle $f^*(X)$. Nous présenterons une application de notre approche à des données socioéconomiques.

[MS-310]

Warning Labels for Lotteries: Visual Display of Small Probabilities
Des avertissements pour les jeux de loterie : les images compréhensible des faibles probabilités

Rachel Dean & Wayne Oldford, *University of Waterloo*

Lottery players often over-estimate their chances of winning. Problem gamblers tend to have extraordinarily distorted impressions of these chances. Even though the probability of winning is available to lottery players, its magnitude is rarely appreciated, likely because these values have little grounding in common experience. Standard statistical graphics are insufficient, largely because of space restrictions. We develop and present a number of new visual displays to help provide the necessary grounding. These are shown for a number of popular lotteries and gaming machines. We recommend these displays as warning labels, much like those attached to cigarette packages.

Les joueurs de loterie surestiment leurs chances de gagner. Les fanatiques du jeu ont des impressions tout à fait déformées de leurs chances de gagner. Même si les probabilités de gagner sont facilement disponibles, leur importance n'est que rarement comprise ; probablement parce que les nombres ne sont pas facilement compréhensibles pour le commun des mortels. Les graphiques statistiquement standard sont insuffisants, surtout en raison des restrictions d'espace. Nous développons et nous présentons quelques images compréhensibles qui aideront à visualiser les faibles probabilités. Nous créons les images pour quelques jeux de loterie et les machines de jeu les plus connues. Nous recommandons que ces images soient utilisées comme avertissements, semblables à ceux que l'on retrouve sur les paquets de cigarettes.

[MS-311]
Quasi-exact Estimation in the Coalescent Process
Estimation quasi exacte dans le processus de coalescence

Hugues Massé & Fabrice Larribe, *Université du Québec à Montréal*

The application of Griffiths – Tavaré's equations for parameters estimation in the coalescent process usually requires the use of Monte Carlo methods. We explore a new approach allowing a different use of these recurrence equations to obtain a quasi-exact estimation of the likelihood without simulations. This new method is possible by combining properties of the coalescent process with composite likelihoods. We apply this new idea in a fine-mapping genetic problem where we estimate the position of a mutation causing a simple genetic disease.

L'application des équations de Griffiths-Tavaré afin d'estimer des paramètres du processus de coalescence requiert habituellement l'utilisation de méthodes Monte Carlo. Nous explorons une nouvelle approche permettant d'utiliser ces équations de récurrence de façon différente pour obtenir une estimation quasi exacte de la vraisemblance sans avoir recours aux simulations. Cette nouvelle méthode est rendue possible en combinant les propriétés du processus de coalescence avec des vraisemblances composites. Nous appliquons cette nouvelle idée dans le cadre d'un problème de cartographie génétique fine où nous voulons estimer la position d'une mutation causant une maladie génétique simple.

[MS-312]
Mining Network Data
Exploration des données de réseau

Hugh Chipman, *Acadia University*, Erika Nahm, *Government of Canada*

Data arising from networks can be rich and complex. For example, the 'header' information in email messages specifies a sender, recipient(s) and a transaction time. Each email address can be viewed as a node in a graph, and email as transactions along the edges. The availability of large databases of such transactional network data presents challenging opportunites for the development of new models. The usual 'statistical learning' (supervised or unsupervised) can only be attempted after developing novel techniques to deal with this complex structure. This can be further complicated if the network structure is thought to change over time.

Les données provenant de réseaux peuvent être riches et complexes. Par exemple, l'information qui figure dans l'« en-tête » des courriels spécifie un expéditeur, le ou les destinataires et l'heure de la transaction. Chaque adresse de courrier électronique peut être considérée comme un nœud dans un graphe et les courriels, comme des transactions le long des arêtes. L'existence de grandes bases de données contenant ce genre de données de réseau transactionnelles offre des possibilités intéressantes pour le développement de nouveaux modèles. L'« apprentissage statistique » habituel (supervisé ou non supervisé) ne peut être tenté qu'après avoir établi des techniques novatrices pour faire face à cette structure complexe. Celle-ci peut devenir encore plus compliquée si l'on pense que la structure du réseau évolue au cours du temps.

[MS-313]
Choosing the Sample Size of a Computer Experiment
Choisir la taille d'échantillon d'une expérience numérique

Jason Loeppky, *University of British Columbia Okanagan*, Jerome Sacks, *National Institute of Statistical Sciences*, William Welch, *University of British Columbia*

In recent years, computer models are supplementing or even replacing physical experiments. These models are often computationally demanding, thus limiting the number of code runs available to emu-

Ces dernières années, les modèles numériques ont complété, voire même remplacé, les expériences physiques. Ces modèles requièrent souvent des ressources informatiques considérables, ce qui limite le nombre d'exécutions du code dis-

late the code output using a statistical approximation. We produce evidence supporting the informal rule of choosing a runs size that is 10 times the number of input dimensions. Our arguments quantify characteristics of computer codes that affect the sample size required for a desired level of accuracy when approximating the code via a Gaussian process. The evidence supporting these arguments stems primarily from a simulation study and via specific codes modeling climate.

ponibles pour émuler la sortie du code en utilisant une approximation statistique. Nous fournissons des preuves à l'appui de la règle officieuse consistant à choisir une taille des exécutions qui est égale à dix fois le nombre de dimensions d'entrée. Nos arguments quantifient les caractéristiques des codes informatiques qui ont une incidence sur la taille d'échantillon requise pour un niveau souhaité d'exactitude quand le code est obtenu de façon approximative par un processus gaussien. Les preuves à l'appui de ces arguments découlent principalement d'une étude par simulation et de codes particuliers modélisant le climat.

[MS-314]

A Statistical Model for Optimizing Power Consumption of Printers

Un modèle statistique pour l'optimisation de la consommation électrique des imprimantes

Laurent Donini, *Xerox Research Centre Europe, Laboratoire Jean Kuntzman, INRIA Rhône Alpes,* Victor Ciriza, *Xerox Research Centre Europe,* Jean-Baptiste Durand, *Laboratoire Jean Kuntzman, INRIA Rhône Alpes, Universités de Grenoble,* Stéphane Girard, *Laboratoire Jean Kuntzman, INRIA Rhône Alpes*

We propose to determine the optimal wait interval that a printer must respect before entering sleep mode. To this end, we introduce a criterion defined by a compromise between power consumption and user impact. The optimal wait interval is inferred by minimizing this criterion. Our approach requires an appropriate statistical modeling of inter-print submission events. Initially, inter-print intervals are supposed to be independent and the optimal wait interval is inferred accordingly. Following this, a hidden Markov chain is used to model print event dependencies due to specific activity states like business hours or night periods.

Nous proposons de déterminer le temps d'attente optimal qu'une imprimante doit respecter avant d'entrer en veille. Pour cela, nous définissons un critère de compromis entre la consommation électrique de l'imprimante et la gêne utilisateur. Le temps d'attente est estimé par minimisation de ce critère. Notre approche repose sur une modélisation statistique des durées interrequêtes. Dans un premier temps, l'estimation est basée sur l'hypothèse d'indépendance de ces durées. Dans un second temps, compte tenu de différents états d'activité comme les heures de bureau ou de repos, une modélisation par chaîne de Markov cachée permet d'introduire des dépendances entre les durées inter-requêtes.

[MS-315]
High-performance Method to Predict Sales of Fast-moving Sales Goods
Quelle méthode pour prévoir les ventes de produits sur un marché à évolution rapide

Frédéric Bertrand & Myriam Maumy, *IRMA - ULP*, Jean-Pierre Bidon, *Interstat*

To anticipate marketing actions ahead of market changes and to predict futures sales becomes one of the main stakes of managers. They need a high-performance forecasting tool that accounts for market dynamics. But, what is the most high-performance method to predict fast moving sales? We will try to answer this question by relying on a comparative study of different forecasting techniques, from the most usual to "state of the art" more sophisticated tools such as state-space models. We will apply those models to standard time series taken out of literature and to data sets derived from real marketing context.

Anticiper les décisions marketing liées aux changements du marché ou prévoir les ventes futures est devenu l'un des enjeux principaux lors des prises de décisions. Les décideurs ont besoin d'outils de prévision ayant les meilleures performances possibles et qui puissent prendre en compte la dynamique propre du marché. Quelle méthode choisir pour prévoir les ventes sur un marché à évolution rapide ? Notre réponse se basera sur l'étude comparative des performances en prévision de plusieurs techniques des plus simples aux plus récentes et complexes comme les modèles espace-états sur des séries temporelles classiques et à des jeux de données réels provenant du marketing.

[MS-316]
Indirect Sampling in the Context of Multiples Frames
Sondage Indirect dans le contexte des bases multiples

Manuela Maia, *Universidade Católica Portuguesa*, Paula Vicente & Elizabeth Reis, *ISCTE*

In a survey, a sample should be selected from a sampling frame. Frequently this frame does not exist or it may not correspond to the target population, and hence, contain only some elements that are somehow related to the required sample. In Indirect Sampling, links between the target population and the sampling frame are used to produce an estimate for the target population. The Generalized Weight Share Method, developed by Lavallée (1995) and Lavallée (2002) can be used to provide a final weight to the surveyed units. Here we will propose a GWSM approach applicable in the context of multiple frames.

Dans une enquête on doit choisir un échantillon parmi une base de sondage. Souvent cette base n'existe pas ou elle n'a pas une correspondance avec la population cible et, donc, ne contiet que quelques éléments reliés, d'une certaine façon à l'échantillon exigé. Au sondage indirect les liens entre la population cible et la base sont employés pour produire une estimation de la population cible. La méthode généralisée du partage des poids, développée par Lavallée (1995 et 2002), peut être employée pour produire un poids final aux unités considérées. Nous proposons, dans cette présentation, une analyse de MGPP applicable dans le contexte des bases multiples.

[MS-317]
Using Reducts to Guide Variable Selection for Optimal Dynamic Regimes
Utilisation de réduits pour guider la sélection de variables pour des régimes dynamiques optimaux

Piotr Biernot & Erica Moodie, *McGill University*

In estimating optimal treatment strategies, the variables used to tailor treatments to patient profiles are typically hand-picked by experts or selected without heed to correlation between attributes. Are those variables optimal for decision making? To answer this question, we develop a new approach

Dans l'estimation de stratégies de traitement optimales, les variables utilisées pour calibrer les traitements aux profils des patients sont habituellement triées à la main par des experts ou sélectionnées sans porter attention à la corrélation entre attributs. Ces variables sont-elles optimales pour la prise de décision ? Pour répondre à cette question, nous

to tailoring variable selection. We use rough sets to suggest collections of useful attributes, then rank the collections using a criterion of Gunter et al (2007). The approach has several advantages: primarily, objectivity and the ability to account for correlation among tailoring variables. We demonstrate our approach in the context of STAR*D, a multistage trial of depression.

développons une nouvelle approche à la sélection des variables de calibration. Nous utilisons des ensembles bruts pour suggérer des collections d'attributs utiles, et classons ensuite les collections en utilisant un critère de Gunter *et al*. (2007). Cette approche a plusieurs avantages : principalement, l'objectivité et la possibilité de tenir compte de la corrélation entre les variables de calibration. Nous illustrons notre approche dans le contexte de STAR*D, une étude de la dépression à plusieurs niveaux.

[MS-318]
Weakly Invariant Designs and Polynomial Designs
Plans faiblement invariants et problèmes de construction de type polynomial

Frédéric Bertrand, *IRMA - ULP*

A design is said to be polynomial if the coordinates of the points supporting the design are the solutions of a system of polynomial equalities or inequalities; such a system can always be solved using semidefinite programming or Gröbner bases. Many sought-after design properties, such as alphabetic optimality and orthogonal blocking, can be easily stated in the framework of polynomial designs. The same holds for G-weakly invariant designs, G being any compact group of matrices, since we show G-weak invariance boils down to H-weak invariance, where H is a subgroup of the orthogonal group O_v of $\mathbb{R}^v$.

Un plan expérimental est solution d'un problème de construction de type polynomial si les coordonnées des points support du plan sont les solutions d'un système d'équations et d'inéquations polynomiales, système que l'on peut toujours résoudre à l'aide de la programmation semidéfinie positive ou des bases de Gröbner. De nombreuses propriétés recherchées, comme l'optimalité alphabétique ou le blocage orthogonal, se formulent naturellement ainsi. Nous obtenons le même résultat pour la recherche de dispositifs G-faiblement invariants, pour G un groupe de matrices compactes quelconques, en réduisant la G-invariance faible à la H-invariance faible où H est un sous-groupe du groupe orthogonal O_v de $\mathbb{R}^v$.

[MS-319]
Development of Non-parametric Models to Analyse Bivalve Behaviour
Développement de modèles non paramétriques pour l'analyse du comportement de bivalve

Mohamedou Sow, Gilles Durrieu, Damien Tran, Pierre Ciret & Jean-Charles Massabuau, *GEMA Université Bordeaux 1 & CNRS*

Valvometry (measurement of valve activity in bivalve molluscs) allows us to record different aspects of molluscan ecophysiology. This application in an environmental context provides us with a global analysis of water quality, that runs 24h a day. It uses the bivalve's ability to close its shell when exposed to a contaminant as an alarm system. We developed a set of non-invasive recording devices, high-tech electronic and mathematical analysis of the signal to estimate parameters describing the animal's behaviour. The data and the statistical modelling approach based on kernel estimator will be presented. The results are automatically updated daily in `http://www.domino.u-bordeaux.fr/molluscan_eye`.

La valvométrie (mesure de l'activité valvaire) permet de décrire différents aspects de l'écophysiologie d'un bivalve. Dans un contexte environnemental, c'est un outil d'analyse 24h/24 de la qualité de l'eau. Les fermetures du bivalve exposé à un contaminant peuvent servir de système d'alarme. Nous avons développé un nouvel ensemble d'enregistrement et d'analyse électronique et mathématique du signal permettant d'estimer différents paramètres du comportement animal. Ce comportement est modélisé par un modèle non paramétrique basé sur l'estimateur de type noyau. Les résultats de ces analyses sont automatiquement mis à jour chaque jour sur le site web `http://www.domino.u-bordeaux.fr/molluscan_eye`.

[MS-320]

Some Invariance Principle for the Empirical Process Indexed by Function
Quelques principes d'invariance forts pour le processus empirique indexé par des fonctions

Adel Settati, *IRMAR - Université Rennes 1*

If $X_1, \ldots, X_n$ are i.i.d. with law P on $\mathcal{X}$ then the empirical process $\alpha_n = \sqrt{n}(P_n - P)$ satisfies a strong invariance principle for a pre-Gaussian class of functions $\mathcal{F}$ with rates v_n if one can construct a sequence of Brownian bridges B_n indexed by $\mathcal{F}$ such that

$$\sup_{f\in\mathcal{F}} |\alpha_n(f) - B_n(f)| = O(v_n) \quad [\mathbb{P}]\text{-a.s.}$$

We have extended the recent method of Berthet and Mason to obtain strong invariance principles under various kinds of hypotheses. The first set of assumptions is based on random entropy and analytical structure. The second concerns mixing conditions when $X_1, \ldots, X_n$ are weakly dependent.

Soit $X_1, \ldots, X_n$ des variables i.i.d. de loi P. On dit que le processus empirique $\alpha_n = \sqrt{n}(P_n - P)$ vérifie le principe d'invariance fort pour une classe de fonctions $\mathcal{F}$ prégaussienne avec une vitesse de convergence v_n, si il existe une suite de ponts Browniens B_n indexés par la classe $\mathcal{F}$, tel que

$$\sup_{f\in\mathcal{F}} |\alpha_n(f) - B_n(f)| = O(v_n) \quad [\mathbb{P}]\text{-p.s.}$$

Nous avons étendu la méthode récente de Berthet et Mason pour obtenir des principes d'invariance forts sous divers types d'hypothèses. Le premier jeu d'hypothèses est basé sur l'entropie aléatoire et la structure analytique de la classe. Le deuxième jeu concerne les conditions de mélange lorsque $X_1, \ldots, X_n$ sont faiblement dépendandes

[MS-321]

Using Multitable Analyses to Explore Data Generated by the National Grapevine Wood Diseases Survey
Analyse des données récoltées par l'Observatoire national des maladies du Bois de la vigne à l'aide de techniques multi-tableaux

Frédéric Bertrand & Myriam Maumy, *IRMA-ULP*, Jacques Grosmann, *DRAF-SRPV Rhône*

Since 2001 and the banning of the use of sodium arsenite in every sector of agriculture in France then in Europe, there is no longer any authorized means of curing grapevine trunk diseases, namely the eutypa dieback, the esca decline and the black dead arm. Accounting for the stability of the incidence rates of these diseases and the association patterns between cultivators and diseases is therefore the main goal of the National Grapevine Wood Diseases Survey. We have fially collected enough data to perform multitable analyses on this dataset, such as the ACT-Statis, partial triadic or multiple coinertia analyses.

L'objectif de l'Observatoire National des Maladies du Bois de la Vigne est de dresser un état des lieux de l'expression des symptômes foliaires des maladies du bois. En effet, suite à l'interdiction de l'utilisation de l'arsénite de soude, en France en 2001 puis en Europe, les viticulteurs ne disposent plus d'aucune méthode de lutte chimique « curative » homologuée contre les principales maladies du bois de la vigne que sont le black dead arm, l'esca et l'eutypiose. Nous disposons enfin de suffisamment d'années de mesures pour utiliser des méthodes multi tableaux comme la méthode Statis ou l'analyse de coinertie multiple.

[MS-322]

Harmonic Analysis of Individual Diatom Counts from Lake Mina Minesota
Analyse harmonique de numérations individuelles de diatomées dans le lac Mina au Minnesota

Karim Rahim, *Queen's University*, Jeannine St. Jacques, *University of Regina*, David Thomson, *Queen's University*

We present a spectral analysis of diatom counts from Lake Mina, Minnesota. In the analysis, we find common periodicities in different species, and discuss phase relations between high abundance

Nous présentons une analyse spectrale de numérations de diatomées faites dans le lac Mina, au Minnesota. Nous constatons l'existence de périodicités communes à diverses espèces et examinons les relations de phase entre les espèces

species. Evidence of periodic solar cycles in the higher prevalent species of diatoms is discussed, specifically multitaper spectrograms reveal changing periodic cycles likely attributable to solar and or lunar influences. Evidence of a reduction in periodic components is found in certain species of diatoms near the Maunder Minimum is indicated. Problems encountered in the analysis of this short series with potentially irregularly-spaced data are discussed.

très abondantes. Nous examinons les preuves de l'existence de cycles solaires périodiques chez les espèces les plus répandues de diatomées, en particulier le fait que les spectrogrammes de type « multitaper » (fenêtres de pondération multiples) révèlent des cycles périodiques évolutifs vraisemblablement attribuables aux influences solaire et lunaire. Nous dégageons des signes d'une réduction des composantes périodiques chez certaines espèces de diatomées, près du minimum de Maunder. Enfin, nous discutons des problèmes rencontrés dans l'analyse de cette courte série avec données éventuellement espacées irrégulièrement.

[MS-323]

A Statistical Model to Estimate the Effect of Roadside Billboards on Crash Frequencies in their Immediate Vicinity

Un modèle statistique pour évaluer l'effet des panneaux réclame le long des routes sur le nombre des collisions routières dans leur voisinage immédiat

François Bellavance, *HEC Montréal,* Assi N'Guessan, *École Polytechnique Universitaire de Lille*

A statistical approach is proposed to estimate the average impact of roadside billboards on the frequency of different types of crashes (fatal, severe, minor, property damage only). The number of crashes of each type for the period before and after the installation of billboards follows a multinomial distribution where the probabilities include a parameter measuring the average effect of billboards on the incidence of crashes after their installation and the information from control zones to allow for potential confounding factors. The results show no statistically significant effect of roadside billboards on the frequency of crashes in their immediate vicinity.

Nous proposons un modèle statistique pour évaluer l'effet moyen des panneaux réclame le long des routes sur le nombre de collisions et leur gravité (mortelle, sévère, mineure, dommage matériel seulement). Le nombre de collisions de chaque type avant et après l'apparition des panneaux réclame est modélisé à l'aide d'une distribution multinomiale où les probabilités incluent un paramètre pour l'effet moyen des panneaux réclame sur l'occurrence de collisions après leur installation ainsi que de l'information provenant de zones contrôles pour prendre en considération des facteurs confondants potentiels. Les résultats montrent que les panneaux réclame le long des routes n'ont aucun effet statistiquement significatif sur le nombre de collisions dans leur voisinage immédiat.

Session 15A Mer. 28 mai • Wed. May 28, 16:30 – 18:00 Capital Hall 1A

Markov Models
Modèles markoviens

[MS-324]

Using Graphical Markov Models in Model Selection

Utilisation de modèles markoviens graphiques dans la sélection de modèles

Ayesha Ali, *University of Guelph,* Angelique Massie, *University of Washington*

Graphical Markov models have been used in diverse fields ranging from psychology to engineering to artificial intelligence to gene networking. In many problems, the main objective is to discover

Les modèles markoviens graphiques ont été utilisés dans divers domaines allant de la psychologie aux réseaux de gènes en passant par le génie et l'intelligence artificielle. Dans de nombreux problèmes, le principal objectif est de découvrir

the underlying structure of some process (i.e. the graph representing the process) given data. In this talk we will focus on processes that can be represented with a directed acyclic graph (DAG). We will provide a brief review of DAGs, and discuss model search methods across Markov equivalent DAGs. We will then discuss the implications when some of the variables in the graph are unobserved, or latent.

la structure sous-jacente d'un processus particulier (p.ex. le graphe représentant le processus), sachant les données. Dans la communication, nous mettrons l'accent sur les processus qui peuvent être représentés par des graphes acycliques orientés ou DAG (pour directed acyclic graph). Nous décrirons brièvement les graphes acycliques orientés et discuterons des méthodes de recherche de modèles parmi des graphes acycliques orientés équivalents au sens de Markov. Puis, nous discuterons des incidences quand certaines variables du graphe sont inobservées, ou latentes.

[MS-325]
Fluid Limit of Markov Chain Monte Carlo Samplers
Une approche fluide des échantillonneurs de Monte-Carlo par chaînes de Markov

Gersende Fort, *LTCI, CNRS / TELECOM ParisTech*

We propose a transformation of some Markov chains which will allow us to define its fluid limit: by renormalization in time, space, and initial value of the chain, we exhibit a time-continuous process which governs the dynamic of the initial chain. The goal is to identify the quantities that govern the ergodic behavior of the Markov chain, by showing their impact on the dynamics of the associated fluid process which, by definition, gives information on the transient steps of the chain. We will consider applications of these techniques to the choice of the design parameters of some MCMC samplers.

Nous présenterons une transformation de chaînes de Markov qui permet de définir sa limite fluide : par la renormalisation en temps, en espace et dans le point initial de la chaîne, on montre un processus à temps continu qui gouverne la dynamique de la chaîne initiale. L'objectif est de mettre en évidence les facteurs responsables de l'ergodicité de la chaîne de Markov via leur impact sur la dynamique du processus fluide associé qui, par définition, porte des informations sur le régime transitoire de la chaîne. Nous considérerons l'application de ces techniques au choix des paramètres de mise en œuvre de certains échantillonneurs MCMC.

[MS-326]
A Multi-state Model for Bivariate Interval Censored Failure Time Data
Un modèle à multi-états pour données bivariées de temps de défaillance censurées par intervalles

Leilei Zeng, *Simon Fraser University*, Richard Cook, *University of Waterloo*

Interval censored life history data arise when the events of interest are only detectable at periodic assessments. When interest lies in the occurrence of two such events, bivariate interval censored event time data are obtained. We describe how to fit a four state Markov model useful for characterizing the association between two interval censored event times. The approach treats the two events symmetrically and enables one to fit multiplicative intensity models which give estimates of covariate effects as well as relative risks characterizing the association between the two events.

Des données d'état de santé censurées par intervalles surviennent lorsque les événements d'intérêt ne sont détectables que lors d'évaluations périodiques. Lorsque nous nous intéressons à la réalisation de deux tels événements, nous obtenons des données bivariées censurées par intervalles. Nous décrivons comment ajuster un modèles markovien à quatre états utile pour caractériser l'association entre deux temps d'événements censurés par intervalles. Cette approche traite les deux événements symétriquement et permet d'ajuster des modèles d'intensité multiplicatifs donnant des estimés des effets des covariables ainsi que les risques relatifs caractérisant l'association entre les deux événements.

Session 15B Mer. 28 mai • Wed. May 28, 16:30 – 18:00 Capital Hall 2A

Actuarial Science
Actuariat

[MS-327]

Estimation of the Jump Size Distribution in a Levy Risk Process
Estimation de la mesure de sauts pour modèles de risque basés sur un processus de Lévy

Jose Garrido, *Concordia University*, Md Sharif Mozumder Mozumder, *University of Dhaka, Bangladesh*

The aggregate claims S=S(t);t>0 are modeled with a Lévy process (a subordinator, to be precise). Positive jumps larger than a given threshold ϵ represent claims, while the more frequent smaller fluctuations model other, non-insurance uncertainty (changes in portfolio composition, other expenses/income). The problem is to estimate the jump measure from a truncated sample path of S. We propose a non-parametric estimator of the subordinator's Lévy measure and of the Poisson rate for the number of jumps larger than ϵ. Examples show how the estimator discriminates between different Lévy processes, such as the gamma, inverse Gaussian or generalized inverse Gaussian process.

Les réclamations totales S=S(t) ;t>0 sont modélisées avec un processus de Lévy (un subordinateur, pour être précis). Les sauts positifs supérieurs à un certain seuil ϵ représentent les réclamations, alors que les nombreuses fluctuations plus petites modélisent d'autres incertitudes (changements de la composition du portefeuille, d'autres dépenses/revenu, etc.). Le problème est l'estimation de la mesure de saut à partir d'un chemin tronqué de S. Nous proposons un estimateur non paramétrique de la mesure de Lévy du subordinateur et du taux de Poisson pour le nombre de sauts supérieurs à ϵ. Les exemples montrent comment l'estimateur distingue entre divers Lévy, comme le processus gamma, gaussien inverse ou gaussien inverse généralisé.

[MS-328]

In the Core of Longevity Risk: Dependence in Stochastic Mortality Models and Cut-offs in Prices of Longevity Swaps
Au cœur du risque de longévité : dépendance stochastique et impact sur les prix des permutations de longévité

Stéphane Loisel & Daniel Serant, *Université Claude Bernard Lyon 1, ISFA*

We propose and calibrate a new stochastic mortality model that takes inter-age correlations into account. The so-called stochastic logit's Deltas model is based on the study of the multivariate time series of the differences of logits of yearly mortality rates. These correlations are important and we illustrate our study on a real-life portfolio. We determine their impact on the price of a pure longevity swap, in which most of the financial risk is taken by a third party. The hypotheses of our model are statistically tested and various measures of risk of the present value of liabilities are computed.

Nous proposons et calibrons un modèle de mortalité stochastique qui prend en compte les corrélations inter-âges. Ce modèle appelé Stochastic logit's Deltas model est fondé sur l'étude de la série multivariée des différences des logits des taux de mortalité annuels. Ces corrélations sont importantes et nous étudions leur impact sur le prix d'un swap pur de longévité (pour lequel le risque financier est transféré à un tiers). Nos hypothèses sont testées statistiquement et le modèle est calibré sur un portefeuille réel.

[MS-329]
Optimal Strategies for Ruin Probabilities and Expected Gains
Stratégies Optimales pour probabilités de ruine et profits espérés

Gary Parker & Cary C.L. Tsai, *Simon Fraser University*

We study strategies to reduce ruin probabilities and enhance expected profits in the classical discrete time surplus process. Various distributions of claim sizes and frequencies are considered. We analyze a pricing method where the renewal premiums are based on Buhlmann's credibility theory. Then we add two policy provisions, a deductible and a policy limit. We propose an index and a VaR measure to select optimal strategies.

Nous étudions des stratégies réduisant les probabilités de ruine et qui améliorent les profits espérés dans le modèle classique de processus de surplus en temps discret. Diverses distributions pour le montant et la fréquence des réclamations sont considérées. Nous analysons une méthode de tarification avec primes de renouvellements basées sur la théorie de la crédibilité de Buhlmann. Puis nous ajoutons deux provisions, une franchise et une limite de couverture. Nous proposons un index et une mesure de risque, VaR, pour choisir des stratégies optimales.

Session 15C Mer. 28 mai • Wed. May 28, 16:30 – 18:00 Capital Hall 3A

Goodness-of-fit Tests
Tests d'adéquation

[MS-330]
Goodness-of-Fit Test for Normality Against Generalized Exponential Power Alternatives
Test de normalité contre des alternatives de la famille de puissance d'exponentielles généralisée

Alain Desgagné, *Université du Québec à Montréal*, Pierre Lafaye de Micheaux, *Université Pierre-Mendès-France*, Alexandre Leblanc, *University of Manitoba*

The generalized exponential power (GEP) density has a large range of tails behavior, which may be exponential, polynomial or logarithmic. Taking this family of densities as an alternative, a normality test based on Rao's score test statistic is proposed. Since the normal density is a special case of the GEP density, this is equivalent to testing a simple null hypothesis about the GEP density parameters. The main interest in this approach is to provide a powerful test against alternatives with non-normal tails. A Monte Carlo simulation study is performed to evaluate the level and power of the test.

La famille de puissance exponentielle généralisée (GEP) contient une grande variété de comportements d'ailes, qui peuvent être exponentielles, polynomiales ou logarithmiques. En considérant cette famille de densités comme alternative, un test de normalité basé sur la statistique du score de Rao est présenté. Puisque la densité normale est un cas particulier de la densité GEP, cela est équivalent à tester une hypothèse simple sur les paramètres de la densité GEP. Le principal intérêt dans cette approche est de fournir un test puissant contre des alternatives avec des ailes non normales. Une étude de simulation est effectuée afin d'évaluer le niveau et la puissance du test.

[MS-331]
Testing the Gamma and Weibull Distributions with Grouped Data
Tester les distribution de Weibull et de Gamma avec les données groupées

John Spinelli, *B.C. Cancer Research Centre*, Michael Stephens, *Simon Fraser University*

Some geological data has been presented with the observations clustered into cells so that the data has

Certaines données géologiques ont été présentées en regroupant les observations dans des cellules de sorte que les

been grouped. This presents problems in finding the parameters of the fitted distributions (gamma or Weibull) using maximum likelihood. These problems are addressed and goodness-of-fit procedures are then given based on the grouped data.

données ont été groupées. Ce groupement pose des problèmes pour le calcul des paramètres des distributions ajustées (gamma ou Weibull) par la méthode du maximum de vraisemblance. Nous examinons ces problèmes, puis nous donnons des procédures de vérification de la qualité de l'ajustement fondées sur les données groupées.

[MS-332]

Change Point Test to Detect Vortical Structures in Stretches of Water

Un test de rupture pour détecter des structures tourbillonnaires dans les plans d'eau

Gilles Ducharme & Céline Vincent, *Université Montpellier II*

In stretches of water, vortical structures retain nutrients and algae, which make them prone to producing toxins. Detection of vortexes is thus part of an early warning system. Starting from a field of random directional vectors, we are developing a test to detect such structures. A scan window explores the points of a grid. At each point, a change point test determines the presence / absence of a vortex in the scanning window. The asymptotic distribution of the test statistic is derived. Simulations show that it approximates the true distribution rather well for reasonable sample sizes.

Dans les plans d'eaux, les structures tourbillonnaires retiennent et mélangent nutriments et algues, ce qui peut les amener à émettre des toxines. La détection de tourbillons est donc un élément d'un système d'alerte. Partant d'un champ de vecteurs aléatoires directionnels, nous développons un test de détection de tourbillon. Une fenêtre de scan explore les points d'une grille. En chaque point, un test de détection de rupture détermine la présence ou l'absence d'un tourbillon dans la fenêtre. La loi asymptotique de cette statistique de test est obtenue. Des simulations montrent que cette dernière se rapproche bien de la loi exacte pour des tailles raisonnables.

[MS-333]

Detecting Changes in the Fluctuations of a Gaussian Process and an Application to Heartbeat Time Series

Détection de ruptures pour un processus gaussien et une application aux séries des fréquences cardiaques

Imen Kammoun & Jean-Marc Bardet, *Université Paris 1 - Panthéon Sorbonne*

We propose a method of detection of multiple abrupt changes of the long-range dependence (respectively self-similarity, local fractality) parameters for Gaussian processes. From a wavelet analysis, the estimator of the m change instants (m is supposed to be known) is proved to satisfy a limit theorem with an explicit convergence rate. Moreover, a central limit theorem is established for an estimator of each parameter. A goodness-of-fit test is also built in each time domain without change and is proved to asymptotically follow a chi-square distribution. Such statistics are applied to heart rate data of marathon runners and lead to interesting conclusions.

Nous proposons une méthode de détection de multiples ruptures des paramètres de longue mémoire (respectivement d'autosimilarité, de fractalité locale) pour des processus gaussiens. À partir de méthodes d'analyse par ondelettes, un estimateur des m points de changement (m est supposé connu) est construit et on montre qu'il vérifie un théorème limite avec une vitesse de convergence explicite. Un théorème de la limite centrale est également établi pour l'estimateur de chaque paramètre et un test d'ajustement est mis en place dans chaque zone où le paramètre est inchangé. Appliquée aux données de fréquences cardiaques qui sont relevées sur les coureurs du marathon, cette méthode conduit à des conclusions intéressantes.

[MS-334]
A Goodness-of-Fit Test for Exponentiality
Test de la qualité de l'ajustement pour l'exponentialité

Ahmad Zghoul, *Qatar University*

A goodness-of-fit test based on the increments of the order statistics is developed for the exponential distribution. An exact distribution of the test is derived and its power is computed for several alternatives. Compared to other known test statistics, the test is performing well.

Nous élaborons un test de la qualité de l'ajustement fondé sur les incréments des statistiques d'ordre pour la loi exponentielle. Nous calculons la distribution exacte de la statistique de test et sa puissance pour plusieurs options. Comparativement à d'autres statistiques de test, notre test donne de bons résultats.

Session 15D Mer. 28 mai • Wed. May 28, 16:30 – 18:00 Capital Hall 4A

Inference for Linear Models
Inférence pour modèles linéaires

[MS-335]
Distance Metric Learning Vs. Fisher Discriminant Analysis
Apprentissage par mesure de distance contre analyse discriminante de Fisher

Ali Ghodsi, Babak Alipanahi & Michael Biggs, *University of Waterloo*

In many fundamental problems in statistical-computing, the Euclidean distances between data points do not represent the desired topology that we are trying to capture. Distance metric learning methods address this problem by constructing a Mahalanobis distance over the input space and use it in place of Euclidean distances.

Most of the proposed algorithms in this area are iterative. In this paper, we propose a closed-form solution to this problem and provide a new problem setup in which the algorithm performs better than many standard methods. Furthermore, we show a strong relationship between these methods and the Fisher Discriminant Analysis (FDA).

Dans plusieurs problèmes fondamentaux en calcul statistique, les distances euclidiennes entre les données ne représentent pas la topologie désirée. Des méthodes d'apprentissage de métriques de distance répondent à ce problème en construisant une distance de Mahalanobis sur l'espace des intrants et l'utilisent à la place de distances euclidiennes.

La plupart des algorithmes proposés dans ce domaine sont itératifs. Dans cet article, nous proposons une solution analytique à ce problème et présentons une nouvelle formulation du problème où l'algorithme performe mieux que plusieurs méthodes usuelles. En outre, nous montrons qu'il y a une relation forte entre ces méthodes et l'analyse discriminante de Fisher.

[MS-336]
A Unified Approach to Efficient Estimation in Simple Linear Regression
Une approche unifiée à l'estimation efficace en régression linéaire simple

Saumen Mandal & Mrityunjay Samanta, *University of Manitoba*

We develop a unified approach to efficient estimation in a simple linear regression without solving the likelihood equations when the error distribution is known. We construct efficient estimators of the parameters using linear combinations of order statistics of random samples drawn from the pop-

Nous développons une approche unifiée à l'estimation efficace en régression linéaire simple sans résoudre les équations de vraisemblance lorsque la distribution des erreurs est connue. Nous construisons des estimateurs efficaces des paramètres en utilisant des combinaisons linéaires des statistiques d'ordre d'échantillons aléatoires tirés de la popula-

ulation. We also construct efficient estimators of the parameters for symmetric type-II censoring and symmetric error distribution using linear combinations of available order statistics with additional weights to the smallest and largest order statistics. It is shown that our estimators are asymptotically normally distributed. Finally, we present examples for several error distributions, to illustrate the theoretical calculations.

tion. Nous construisons aussi des estimateurs efficaces des paramètres avec censure symétrique de type II et distribution d'erreurs symétrique en utilisant des combinaisons linéaires des statistiques d'ordre disponibles et en ajoutant des poids aux plus grandes et plus petites statistiques d'ordre. Nous montrons que nos estimateurs sont asymptotiquement de loi normale. Finalement, nous présentons des exemples pour plusieurs distributions d'erreurs, afin d'illustrer les calculs théoriques.

[MS-337]

Constrained Inference in Generalized Linear and Mixed Models
Inférence avec contraintes des modèles linéaires généralisés et mixtes

Karelyn Davis, *Statistics Canada & Carleton University*, Chul-Gyu Park & Sanjoy Sinha, *Carleton University*

In the past half-century, statisticians have recognized the improvement in efficiency of many inference problems as a result of implementing the prior ordering of parameters or restrictions in the analysis. As it is often the case that observations are not normally distributed and are sometimes observed in a cluster, Generalized Linear Models (GLMs) or Generalized Linear Mixed Models (GLMMs) are employed. This presentation will extend estimation and hypothesis testing methods for such models under inequality constraints using the Gradient Projection (GP) algorithm. Results of simulation studies and applications will also be discussed.

Au cours des dernières années, les statisticiens ont reconnu l'amélioration en précision aux problèmes d'inférence lorsqu'on tient compte de l'ordre des paramètres ou de restrictions dans l'analyse. Souvent les données n'ont pas une distribution normale ou se retrouvent en grappes, alors les modèles linéaires généralisés (GLM) ou linéaires généralisés mixtes (GLMM) sont utilisés. La présentation examinera des méthodes d'estimation et de tests d'hypothèses de tels modèles sous des contraintes d'inégalité en appliquant l'algorithme de projection du gradient (GP). Les résultats de simulations et d'applications seront aussi discutés.

[MS-338]

Testing for Null Partial Correlation: a Formal Comparison of Freedman-Lane's and Kennedy's Permutation Test Strategies
Tests de nullité d'une corrélation partielle : une comparaison formelle des stratégies permutationnelles de Freedman-Lane et Kennedy

Gérard d'Aubigny & Catherine d'Aubigny, *MS^3-LJK, Université de Grenoble*, Ali Shadrock, *Université de Téhéran, Iran*

Anderson and Legendre (1999) compared by simulations permutation methods for testing the significance of one partial regression coefficient in a linear setting. They studied the type I error and power of four tests: the permutation of raw data (Manly, 1991), the permutation under the full model (ter Braak, 1992), and two alternative methods permutating residuals under the reduced model (Freedman – Lane, 1983; Kennedy, 1995). They concluded that the Freedman – Lane test presented the most consistent and reliable re-

Anderson et Legendre (1999) comparent par simulations les méthodes permutationnelles de test d'un coefficient régression linéaire partielle. Ils étudient l'erreur de première espèce et la puissance de quatre tests : permutation des données brutes (Manly 1991), permutation des résidus dans le modèle maintenu (ter Braak, 1992) et deux méthodes permutant ceux du modèle réduit (Freedman et Lane, 1983 ; Kennedy, 1995). Ces auteurs jugent que Freedman-Lane fournit les résultats les plus cohérents et fiables parmi les quatre. Nous avons établi les propriétés de biais-variance permutationnelles des tests de Kennedy et de Freedman-Lane, puis

sults among the four. We derived and compared the permutational bias-variance properties of the Kennedy and Freedman-Lane tests, and discussed their consequences on interpretations of the Anderson – Legendre simulations.

discuté leurs conséquences pour l'interprétation des simulations de Anderson et Legendre.

[MS-339]

Note on the Respective Widths of Confidence Intervals for Means with Known and Unknown Variances
Notes sur les largeurs respectives des intervalles de confiance pour les moyennes avec variances connues et inconnues

Marc Bourdeau, *École Polytechnique de Montréal*

Most professors of statistics assume implicitly that confidence intervals (CI) for means of normal distributions with unknown variances are wider than the ones where the variances are assumed to be known. The reason is that less knowledge brings more uncertainty, confirmed by the T-scores, which surely entails wider CI. This is far from always being the case. This simple fact follows from elementary mathematical considerations. This fact was suggested by the observations of an animation designed for CI learning. We will also comment on the use of NT for the learning of the discipline.

Les enseignants en statistique admettent souvent que les intervalles de confiance (IC) pour les moyennes des populations gaussiennes sont plus grands lorsqu'on suppose que les variances sont inconnues que lorsqu'elles sont présumées connues : moins d'information entraînant sûrement plus d'incertitude. Des considérations mathématiques élémentaires montrent que ce n'est pas vrai pour une bonne partie des cas. Ces développements ont été suggérés par l'observation d'une animation destinée à faire comprendre le concept des IC aux étudiants. On commentera aussi l'utilité des TIC dans l'apprentissage de la discipline.

[MS-340]

Kullback Information Criterion for Mixture Regression Models
Critère d'Information de Kullback pour le mélange des modèles de régression

Bezza Hafidi, *Université Ibn Zohr, Agadir, Maroc*, Abdallah Mkhadri, *Université Cadi-Ayyad, Marrakech,Maroc*

We consider the problem of jointly selecting the number of components and variables in finite mixture regression models. The Akaike information criterion (AIC) and the Kullback information criterion (KIC) are unsatisfactory because they overestimate the number of components which lead to incorrect variables being retained in the model. Naik et al. (2007) proposed a new mixture regression criterion (MRC) which simultaneously determines the number of components and variables.

Similarly to MRC and in the same context, we propose a new information criterion, MRC_s. We show that MRC_s performs well than MRC in large simulation study.

Nous examinons le problème de la sélection conjointe des nombres de composantes et des variables dans les mélanges finis de modèles de régression. Le critère d'information d'Akaike (AIC) et le critère d'information de Kullback (KIC) sont insatisfaisants, parce qu'ils surestiment le nombre de composantes, de sorte que des variables inappropriées sont retenues dans le modèle. Naik *et al.*(2007) ont proposé un nouveau critère de mélange de régressions (MRC) qui permet de déterminer simultanément les nombres de composantes et de variables. Dans le même ordre d'idée que le MRC et dans le même contexte, nous proposons un nouveau critère d'information, appelé MRC_s. Nous montrons que ce dernier donne d'aussi bons résultats/de meilleurs résultats que le MRC dans une grande étude par simulation.

Session 15E Mer. 28 mai • Wed. May 28, 16:30 – 18:00 Capital Hall 5A

Clinical Trials
Essais cliniques

[MS-341]

Two-Stage Design for Phase II Cancer Clinical Trials with Responses and Early Progression as Co-primary Endpoints

Design en deux étapes pour les essais cliniques de phase II en cancérologie avec la réponse ou la progression précoce comme co-critères de jugement principaux

Xianming Tan, *NCIC Clinical Trials Group, Cancer Research Institute, Queen's University*, Glen Takahara, *Queen's University*, Dongsheng Tu, *NCIC Clinical Trials Group, Cancer Research Institute, Queen's University*

In Phase II cancer clinical trials of newly developed cytotoxic or molecularly targeted agents, it is proposed to use both response and early progression rates as co-primary endpoints of the trials. An agent is considered promising if either its response rate is high or its early progression rate is low. In this paper, we present some strategies for the search for optimal two-stage designs by using the downhill simplex algorithm, which minimizes the expected total sample size and has its type I and II errors bounded above by pre-specified values. Examples are given for some scenarios appearing in clinical practice.

Dans les essais cliniques de cancérologie qui testent les nouvelles thérapies moléculaires ciblées ou cytotoxiques, la réponse et la progression précoce sont souvent utilisées comme cocritères de jugement principaux. Un agent est considéré comme prometteur si le taux de réponse est important ou si le taux de progression précoce est bas. Dans cet article, nous présentons des stratégies pour la recherche du design en deux étapes en utilisant l'algorithme descendant du simplexe qui minimise l'effectif total attendu, et dont les erreurs de type I et II sont limitées par des valeurs préspécifiées. Des exemples seront donnés pour plusieurs cas cliniques possibles.

[MS-342]

Early Stopping Rules and Bias in RCTs

Règles précoces d'arrêt et biais dans les études randomisées contrôlées (ERC)

Timothy Ramsay, *University of Ottawa*, Kaveh Shojiana, *The Ottawa Hospital*

The fact that early stopping rules lead to conditional bias in randomized controlled trials has been acknowledged in statistical literature since at least 1958, but it's implications have been poorly appreciated by statisticians and are largely unknown by epidemiologists. I will present results, both analytic and from a simulation study, to demonstrate that the conditional bias present when a study stops early can be so large as to render the study largely uninformative. A apparently large effect size can easily correspond to a small, clinically unimportant true effect.

Le fait que les règles d'arrêt précoce mènent à des biais d'accoutumance dans les études randomisées contrôlées est reconnu dans la documentation statistique depuis au moins 1958, mais ses incidences ont mal été étudiées par les statisticiens en plus de demeurer largement méconnues des épidémiologistes. Les résultats, à la fois analytiques et découlant d'une étude de simulation, seront présentés en vue de démontrer que les biais d'accoutumance présents lorsque l'on met fin de façon précoce à une étude peuvent faire en sorte que, en raison de leur importance, l'étude ne permette pas d'en apprendre beaucoup. Une taille de l'effet semblant être importante peut facilement correspondre à un faible effet réel sans importance sur le plan clinique.

[MS-343]
Analysis of Clustered Data with Response Measurement Error
Analyse des données agrégées ayant une erreur de mesure de la réponse

Zhijian Chen, Grace Y. Yi & Changbao Wu, *University of Waterloo*

Measurement error has been a long standing concern in medical research and survey sampling. Much research has been directed to measurement error in covariates. There is relatively little work concerning measurement error in response. In this talk, we describe analysis methods to handle response measurement error. We specifically focus the discussion on clustered data under generalized random effects models. Numerical studies will be presented to assess the performance of the proposed methods.

L'erreur de mesure est une source de préoccupation de longue date en recherche médicale et en échantillonnage. Une foule d'études ont eu pour sujet l'erreur de mesure dans les covariables. Par contre, assez peu de travaux ont porté sur l'erreur de mesure dans la réponse. La communication décrira des méthodes d'analyse pour traiter cette dernière. Nous nous attarderons tout spécialement à la discussion des données groupées sous des modèles généralisés à effets aléatoires. Nous présenterons des études numériques pour évaluer la performance des méthodes proposées.

[MS-344]
Modeling Multistage Sampling of Family Data with Missing Information
Modélisation de l'échantillonnage à plusieurs niveaux de données familiales avec des informations manquantes

Yun Hee Choi & Laurent Briollais, *Samuel Lunenfeld Research Institute at Mount Sinai Hospital*

In family-based genetic studies, multistage sampling permits the allocation of resources to families that are most informative for a given objective while allowing population-based inference. A common problem in this design is the presence of missing genetic information among family members. To analyze multistage family designs, we propose a composite likelihood approach that we further extend to account for missing genetic information using an EM algorithm. We then show an application to a study of early-onset breast cancer among BRCA mutation carriers where data are collected from several cancer family registries using a multistage sampling.

Dans les études génétiques familiales, l'échantillonnage à plusieurs niveaux permet l'allocation des ressources aux familles les plus informatives pour un objectif donné tout en rendant possible une inférence à la population entière. Un problème courant pour ce plan d'échantillonnage est la présence d'informations génétiques manquantes chez les membres de la famille. Pour analyser les plans d'échantillonnage familiaux à plusieurs niveaux, nous proposons une approche de vraisemblance composite que nous généralisons ensuite afin de tenir compte des informations génétiques manquantes en utilisant un algorithme EM. Nous montrons ensuite une application à une étude du cancer du sein précoce chez les porteurs d'une mutation de BRCA à partir de données recueillies dans plusieurs registres de cancer familiaux utilisant un plan d'échantillonnage à plusieurs niveaux.

Session 15F Mer. 28 mai • Wed. May 28, 16:30 – 18:00 Capital Hall 6A

Functional Estimation and Censored Data
Estimation fonctionnelle et données censurées

[MS-345]
On a Weighted Kaplan-Meier Estimate
À propos d'un estimateur de Kaplan-Meier pondéré

Jean-François Plante, *University of Toronto*

The MAMSE (minimum averaged mean square error) nonparametric adaptive weights were developed in the context of the weighted likelihood to infer about a population of interest while using samples from $m - 1$ other sources as well. We propose a generalized version of the MAMSE weights for right-censored data and use them to define the weighted Kaplan – Meier Estimate (WKME). The WKME converges uniformly in probability to the target distribution and simulations show that its performance on finite samples exceeds that of the usual Kaplan-Meier estimate in many cases.

Les poids EQMIM (erreur quadratique moyenne intégrée minimale) ont été développés dans le contexte de la vraisemblance pondérée afin d'inférer au sujet d'une population donnée, tout en utilisant l'information provenant de $m - 1$ autres sources. Nous présentons une généralisation des poids EQMIM qui s'accommode des données censurées à droite et nous en servons afin de définir un estimateur de Kaplan-Meier pondéré (EKMP). L'EKMP converge uniformément en probabilité vers la distribution ciblée et des simulations indiquent qu'il offre des performances supérieures à celle de l'estimateur de Kaplan-Meier habituel dans plusieurs cas.

[MS-346]
Conditional Density Estimation in a Censored Single-index Regression Model
Estimation de la densité conditionnelle dans un modèle à direction révélatrice unique en présence de censures

Olivier Bouaziz, *Laboratoire de Statistique Théorique et Appliquée*, Olivier Lopez, *Crest-Ensai, Irmar and Weierstrass Institute*

Under a single-index regression assumption, we introduce a new semiparametric procedure to estimate a conditional density of a censored response. The regression model can be seen as a generalization of Cox regression model, and also as a profitable tool for performing dimension reduction under censoring. We derive the consistency and asymptotic normality of our estimator from the index parameter. Furthermore, we provide a new adaptive procedure which allows us both to choose the smoothing parameter involved in our approach, and to circumvent the weak performances of Kaplan – Meier estimator in the right-tail of the distribution.

Dans un modèle à direction révélatrice unique, nous introduisons une nouvelle méthode d'estimation semi-paramétrique pour estimer la densité conditionnelle de données censurées. Ce modèle de régression peut-être vu à la fois comme une généralisation du modèle de Cox et aussi comme une méthode efficace de réduction de la dimension en présence de censures. Nous obtenons des résultats de consistance et de normalité asymptotique pour le paramètre d'intérêt de la régression. De plus, nous utilisons une procédure adaptative pour choisir le paramètre de lissage et pour éviter les problèmes d'estimation dans les queues de distribution de l'estimateur de Kaplan-Meier.

[MS-347]

Distribution Function Estimation under Interval Censoring Case 1

Estimation de la fonction de répartition dans le modèle de censure par intervalle de type 1

Elodie Brunel, *Université Montpellier 2*, Fabienne Comte, *Université Paris Descartes*

Let X be a survival time with unknown distribution function F. In the interval censoring case 1 model, the survival time X is not observed. Instead, the pair of variables (U, δ) is observed where U is an examination time and δ is the indicator function of the event $(X \leq U)$. Two types of adaptive projection estimators of the distribution function are proposed. The first estimator is a ratio estimator while the second one results from a direct mean square regression contrast. Both estimators are proved to automatically achieve the optimal rate associated with the unknown regularity of the function.

Dans le modèle de censure par intervalle de type 1, la durée de vie X n'est pas observée. Seul, le couple de variables (U, δ) est observé où U représente la durée écoulée jusqu'à un instant d'observation et δ est l'indicatrice de l'événement $(X \leq U)$. Deux estimateurs adaptatifs de la fonction de répartition F de X sont étudiés. L'un est construit comme quotient d'estimateurs alors que l'autre est obtenu directement par minimisation d'un contraste des moindres carrés. On montre que les deux estimateurs sont adaptatifs dans le sens où ils atteignent la vitesse non paramétrique optimale associée à la régularité inconnue de la fonction. Une étude de simulations compare les deux approches.

[MS-348]

Using Conditional Quantiles to Estimate the Cumulative Distribution Function of a Censored Variable in a Small Area

Estimation sur un petit domaine de la fonction de répartition de l'utilisation de quantiles conditionnels variable censurée à l'aide de quantiles conditionnels

Eve Leconte & Sandrine Casanova, *TSE (GREMAQ), Université Toulouse 1*

We consider estimation of the cumulative distribution function of a possibly censored variable Y on small areas. This estimation is usually based on auxiliary information related to the variable of interest and information is "borrowed" from the other areas. In this setting, to take censoring into account, prediction of the Y value for non-sampled individuals is performed non-parametrically by conditional quantiles, whose orders are estimated by generalized Kaplan – Meier estimators. Estimation of the c.d.f. in the area is derived using Kaplan – Meier estimator. The obtained estimators are compared by simulations with Horvitz – Thompson estimator and applied to real data.

On souhaite estimer la fonction de répartition d'une variable d'intérêt Y, éventuellement censurée, sur de petits domaines. Cette estimation se base habituellement sur les échantillons de tous les domaines et sur une information auxiliaire. Dans ce cadre, pour prendre en compte la censure, nous prédisons non paramétriquement la valeur de Y d'un individu hors échantillon par des quantiles conditionnels d'ordres estimés par l'estimateur de Kaplan-Meier généralisé. Un estimateur de la f.d.r. du domaine s'en déduit par Kaplan-Meier. Des simulations comparent cette méthode avec l'estimateur de Horvitz-Thompson. Un exemple d'application est étudié.

[MS-349]

Bivariate Cumulative Distribution Function Estimate in Case of Right and Left Censoring

Estimateur de la fonction de répartition bivariée avec censures à droite et à gauche

Philippe Saint-Pierre & Agathe Guilloux, *Laboratoire de Statistique Théorique et Appliquée (LSTA), Université Pierre et Marie Curie - Paris 6*

The purpose of this work is to estimate the bivariate cumulative distribution function (CDF) when one variable is left-censored and the other one is right-

L'objet de ce travail est l'estimation de la fonction de répartition (f.d.r.) bivariée dans le cas où une variable est censurée à gauche et l'autre censurée à droite. Cette question ne semble

censored. This issue does not seem to be studied in the literature, while several studies deal with the case where the two variables are right-censored. The representation of the bivariate survival function as a product integral of the cumulative hazard function, obtained in the right censoring case, is generalized to left- and right-censoring cases. A nonparametric estimator of CDF is obtained. The strong consistency of the estimator is investigated.

pas être étudiée dans la littérature bien que plusieurs travaux traitent du cas où les deux variables sont censurées à droite. La méthode proposée s'inspire d'une approche permettant, dans le cas bivarié avec deux censures à droite, d'exprimer la f.d.r. comme un produit intégral de la fonction de hasard cumulée. Un estimateur non paramétrique de la f.d.r. est obtenu. La forte convergence de l'estimateur est étudiée.

[MS-350]
Discrete-Time Survival Trees
Arbres de survie à temps discret

Imad Bou-Hamad, *HEC Montreal*

Tree-based methods are frequently used in studies with censored survival time. The existing methods are tailor-made to deal with a survival time variable that is measured continuously. However, survival variables measured on a discrete scale are often encountered in practice. We propose a new tree construction method specifically adapted to such discrete-time survival variables. A simulation study shows that single trees and bagged-trees perform well compared to a parametric model. A real data example investigating the usefulness of personality dimensions in predicting early onset of cigarette smoking is presented.

Les méthodes d'analyse basées sur les arbres sont fréquemment utilisées dans les études portant sur des temps de survie censurés. Les méthodes existantes sont adaptées au cas par cas pour traiter une variable de temps de survie qui est mesurée de manière continue. Cependant, en pratique, on rencontre souvent des variables de survie mesurées sur une échelle discrète. Nous proposons une nouvelle méthode de construction d'arbres spécialement adaptée à ce genre de variables. Une étude par simulation montre que les arbres uniques et les arbres agrégés par bootstrap (bagged-trees) se comparent bien à un modèle paramétrique. Un exemple fondé sur des données réelles qui étudie l'utilité des dimensions de la personnalité dans la prédiction du début du tabagisme est présenté.

Session 15G Mer. 28 mai • Wed. May 28, 16:30 – 18:00 Capital Hall 7A

Stochastic Models in Biostatistics
Modèles stochastiques en biostatistique

[MS-351]
Small-Scale Spatial Modelling of Cancer Risk in Ontario
Analyse spatiale a courte échelle des taux de cancer en Ontario

Patrick Brown & Todd Norwood, *Cancer Care Ontario*, Steve Fan & Jamie Stafford, *University of Toronto*

Markov random field models are a standard tool for disease mapping, and there is no methodological barrier to applying them to fine resolution data afforded by postal codes obtained from the Ontario Cancer Registry. Some practical considerations do arise, such as a mismatch between postal code regions and census regions; suppression of covariates such as income levels in some regions with small

Le champ de Markov est un outil standard pour la cartographie de la maladie, et la méthodologie pourrait facilement être appliquée aux codes postaux obtenus du Registre du cancer de l'Ontario malgré la résolution très fin qu'ils accordent. Un certain nombre de considérations pratiques se posent, comme les différences entre les codes postaux et les régions de recensement ; répression des variables telles que les niveaux de revenus dans certaines régions à faible popu-

population; irregularity of regions' different sizes in urban and rural areas; and changes in boundaries over time. This talk will show a spatial analysis of lung cancer rates in Essex county, Ontario, and discuss to the above problems.

lation ; irrégularité des régions, surtout entre zones urbaines et les zones rurales, et le changement des régions de recensement d'un recensement à l'autre. Cette présentation démontrera une analyse spatiale des taux de cancer du poumon dans le comté d'Essex, en Ontario, et discutera quelques résolutions pour les problèmes présentés.

[MS-352]

Spatio-temporal Modeling of Disease Rate Mapping
Modélisation spatio-temporelle de la cartographie du taux d'incidence d'une maladie

Mahmoud Torabi & Rhonda J. Rosychuk, *University of Alberta*

In this talk, we propose generalized linear mixed models (GLMMs) for the analysis of geographic and temporal variability of disease rates. This class of models adopts spatially correlated random effects and random temporal components. The model also accommodates the interaction between space and time. We introduce the seasonal effects that have not been previously addressed and estimate model parameters through generalized estimating equations. The model provides smoothed maps of disease risk and eliminates the instability of estimates in low-population areas while maintaining geographic resolution. We illustrate our approach using a real dataset.

Dans cet exposé, nous proposons des modèles linéaires généralisés mixtes pour analyser la variabilité géographique et temporelle de taux d'incidence de maladies. Cette classe de modèles comprend des effets aléatoires spatialement corrélés et des composantes aléatoires temporelles. Le modèle accommode aussi l'interaction entre l'espace et le temps. Nous présentons des effets saisonniers qui n'étaient considérés antérieurement et estimons les paramètres du modèle au moyen d'équations estimatrices généralisées. Le modèle produit des cartes lisses du risque associée à la maladie et élimine l'instabilité des estimés dans des régions de faible population sans perdre de résolution géographique. Nous illustrons notre approche avec un jeu de données réelles.

[MS-353]

An Application of Modeling Count Data using Linked Administrative Healthcare Databases
Une application du dénombrement des données modélisées en utilisant des bases de données administratives sur les soins de santé

Jeffrey Dowden, Khokan Sikdar & John Knight, *Newfoundland and Labrador Centre for Health Information*

Count data occur frequently in healthcare studies, and include number of hospitalizations and days spent in hospital. Such data are routinely modeled using Poisson regression. In many instances, data are overdispersed or contain a higher frequency of zeros than expected under the Poisson distribution. In this presentation, we examine the Poisson model and three alternatives: Negative binomial, zero-inflated Poisson, and Poisson hurdle. These models are fitted to linked administrative healthcare data to assess predictors of hospitalizations among diabetic seniors in the province of Newfoundland and Labrador. Analytical and graphical approaches are used to aid in model selection and interpretation.

Les dénombrements sont fréquents dans les études sur les soins de santé et comprennent le nombre d'hospitalisations et de jours passés à l'hôpital. Ces données sont ordinairement modélisées par une régression de Poisson. Or, dans de nombreux cas, les données sont surdispersées ou présentent une fréquence de valeurs nulles plus élevée que celle prévue sous la loi de Poisson. Dans la communication, nous examinons le modèle de Poisson et trois autres options, à savoir le modèle binomial négatif, le modèle de Poisson à surplus de zéros (zero-inflated Poisson) et le modèle du « saut » de Poisson (Poisson hurdle). Ces modèles sont ajustés à des données administratives sur les soins de santé afin d'évaluer les prédicteurs de l'hospitalisation chez les personnes âgées diabétiques de Terre-Neuve-et-Labrador. Des approches analytiques et graphiques sont utilisées pour faciliter le choix et l'interprétation des modèles.

[MS-354]
Modeling Replicated Time-course Evolution of Gene-expression Profiles
La modélisation de l'évolution dans le temps de profils d'expression génétique

Alejandro Murua, *Université de Montréal*

We present a statistical model for time-course studies including independent replicates. Our model deals with smoothed time-courses through a Gaussian intrinsic autoregressive process. The smoothed time-courses are obtained by penalizing the size of their second derivatives. Parameters are estimated via Iterated Conditional Modes. If replicates are available, we estimate the smoothing parameter using cross-validation. This methodology is applied to the analysis and clustering of gene-expression data with the goal of studying the development of the retina in mice. We show that working only with the modeled signal time-courses leads to better clustering results.

Nous présentons un modèle statistique pour des études d'évolution dans le temps y compris des répétitions indépendantes. Notre modèle traite le lissage de signaux d'évolution temporelle en utilisant un processus autorégressif gaussien intrinsèque. Le lissage des signaux est obtenu en pénalisant la taille de leur deuxième dérivée. Les paramètres sont estimés via les maximaux conditionnels itérés. Lorsque des répétitions sont disponibles, nous estimons le paramètre de lissage à l'aide de la validation croisée. Cette méthodologie est appliquée à l'analyse et au groupement des données d'évolution d'expression génétique dans le but d'élucider le développement de la rétine chez les souris. Nous montrons que l'utilisation des signaux lissés par notre modèle conduit aux meilleurs groupements des données.

[MS-355]
A Positive and Negative Integer-valued Bilinear Type Process
Un processus bilinéaire à valeurs entières positives et négatives

Alain Latour, *Université Pierre-Mendès-France/MS3-LJK,* Lionel Truquet, *Université Paris 1 / CREST*

With the aim of generalizing a bilinear type counting model, we redefine a more elaborated model than the one proposed by Doukhan et al. (2006). This new model, which can be useful in a context of inventory management or epidemiology, is based on a thinning operator allowing for negative values. In a simple way, we demonstrate the existence of such a process. The estimation of the parameters is tackled with the help of a quasi-maximum likelihood approach (referred to as QMLE). Under certain hypotheses, we prove that these QMLE estimators are strongly convergent. With some additionnal assumptions, we obtain asymptotically normal estimators.

Dans le but de généraliser un modèle de comptage de type bilinéaire, nous redéfinissons un modèle plus élaboré que celui de Doukhan *et al*. (2006). Ce nouveau modèle pouvant être utile dans un contexte de gestion des inventaires et de médecine sociale et préventive est basé sur un opérateur d'amincissement permettant des valeurs négatives. On démontre de manière simple l'existence d'un tel processus. L'estimation des paramètres est étudiée avec une approche de quasi-vraisemblance maximale (EQVM). Sous certaines hypothèses, on démontre la convergence presque sûre des estimateurs EQVM. En ajoutant quelques hypothèses supplémentaires, on obtient une estimation asymptotiquement normale.

Session 16A Jeu. 29 mai • Thu. May 29, 08:30 – 10:00 Congress Hall A/E

Marie-Jeanne-Laurent-Duhamel Addresses
Allocutions des récipiendaires du prix Marie-Jeanne-Laurent-Duhamel

[MS-356]
Sequential Aggregation of Predictors, with Aapplication to Air-Quality Forecasting
Agrégation séquentielle de prédicteurs, avec application à la prévision de la qualité de l'air

Gilles Stoltz, *CNRS - Ecole normale supérieure - HEC Paris*

Individual sequences are concerned with (1) sequential prediction of outcomes that are not the realizations of some underlying stochastic process, (2) given some base predictions formed by experts. One aims at predicting almost as well as the best expert. Typical good forecasters maintain one weight per expert, update these weights depending on the past performances, and output at each step a linear combination of the experts' advices according to these weights. In this talk, I will describe some efficient forecasters and study their performances for predicting ozone peaks over Europe.

On parle de suites individuelles (1) lorsqu'il s'agit de prédire séquentiellement des observations qui ne sont pas des réalisations d'un certain processus stochastique sous-jacent, (2) avec, à cet effet, l'aide d'experts formant chacun une prédiction propre. Le but est de prédire presque aussi bien que le meilleur expert. Les bons algorithmes de prédiction utilisent typiquement un vecteur de poids (un pour chaque expert), qu'ils mettent à jour séquentiellement, et utilisent pour former une prédiction finale donnée par une combinaison linéaire des prédictions propres selon ce vecteur. Je décrirai dans cet exposé quelques tels algorithmes et étudierai leurs performances dans le cadre de la prédiction de pics d'ozone à l'échelle de l'Europe.

[MS-357]
Copulas : Goodness-of-fit and Independence Tests
Copules : tests d'adéquation et d'indépendance

Jean-François Quessy, *Université du Québec à Trois-Rivières*, Christian Genest, *Université Laval*, Bruno Rémillard, *HEC Montréal*, Marie-Hélène Toupin & Mhamed Mesfioui, *Université du Québec à Trois-Rivières*

The first part of my presentation will be dedicated to a survey of the results obtained in my thesis. In particular, I will talk about new goodness-of-fit testing procedures for copula families and asymptotic relative efficiency of some tests of multivariate independence. In the second part, I will mention some recent results that can be seen as extensions of those in my thesis and I will describe new avenues of research that merit to be explored in the future.

La première partie de mon exposé sera dédiée à une revue des résultats obtenus dans ma thèse. En particulier, je parlerai de nouvelles procédures pour tester l'adéquation à des familles de copules et de l'efficacité relative asymptotique de quelques tests d'indépendance multivariée. Dans la deuxième partie, je mentionnerai quelques résultats plus récents qui font suite à ma thèse et je décrirai de nouvelles avenues de recherche qui méritent d'être explorées dans le futur.

Session 17A Jeu. 29 mai • Thu. May 29, 10:30 – 12:00 Capital Hall 1A

Statistical Genomics
Statistique génomique

[MS-358]

Gene-Gene Influences Inferred from Microarray Time Course Data
Influences gène-gène inférées d'après des données de microréseau recueillies au cours du temps

David Bickel, *University of Ottawa,* Mary Beatty, Shai Lawit & Nicholas Bate, *Pioneer Hi-Bred international, Inc.*

Unlike earlier instruments measuring gene expression, microarrays represent a large-enough portion of the genome to enable computation of an informative upper bound of how much confidence one may reasonably place in causal relationships between genes, at least in principle. To determine the extent to which physical influences between genes can be obtained from microarray data, we perturbed a simple plant cell system and measured the gene expression over time. Model parameters were inferred for 25 downstream genes of interest using a thoroughly studied approximation of the posterior distribution, accounting for model uncertainty through Bayesian model averaging.

Contrairement aux instruments antérieurs de mesure de l'expression des gènes, les micropuces représentent une part suffisamment grande du génome pour permettre le calcul d'une borne supérieure informative du degré de confiance que l'on peut, du moins en principe, accorder aux relations causales entre gènes. Pour déterminer la mesure dans laquelle des données de micropuces peuvent renseigner sur les influences physiques entre les gènes, nous avons perturbé un système cellulaire végétal simple et avons mesuré l'expression des gènes au cours du temps. Nous avons inféré les paramètres du modèle pour 25 gènes d'intérêt situés en amont en utilisant une approximation minutieusement étudiée de la loi *a posteriori*, en tenant compte de l'incertitude du modèle par le moyennage de modèles bayésiens.

[MS-359]

A Dynamic Programming Algorithm to Evaluate Dirichlet Ordering Probabilities, with Application to Multigroup Gene Expression Analysis
Un algorithme dynamique pour évaluer les probabilités d'arrangements de variables Dirichlet, avec application à l'analyse d'expression génétique à plusieurs groupes

Michael Newton & Chung Lisa, *University of Wisconsin*

The Dirichlet distribution appears in many applications. Though its density is trivial to compute, integrals recording probabilities of potential interest may be difficult to evaluate. Analysis allows us to recast the probability of each vector ordering as a probability on independent negative binomial variables, and thus to construct a dynamic programming algorithm for efficient computation. The development was motivated by multi-group gene expression analysis, in which vector orderings correspond to structures relating mean expression among different groups of tissue samples. Posterior structure probabilities are computed by the proposed algorithm. We demonstrate the resulting model-based clustering in several time-course microarray examples.

La distribution de Dirichlet apparaît dans plusieurs applications. Bien que sa densité soit triviale à évaluer, les intégrales correspondant à des probabilités d'intérêt peuvent être difficiles à évaluer. L'analyse nous permet de ré-exprimer la probabilité de chaque arrangement de vecteurs comme une probabilité de variables indépendantes de loi binomiale négative, et donc de construire un algorithme de programmation dynamique pour des calculs efficaces. Ce développement était motivé par l'analyse d'expression génétique à plusieurs groupes, où les arrangements de vecteurs correspondants à des structures reliant l'expression moyenne parmi différents groupes d'échantillons de tissus. Les probabilités de structure *a posteriori* sont calculées par l'algorithme proposé. Nous montrons la classification à base de modèles résultante dans plusieurs exemples de micropuces à décours temporel.

[MS-360]
Generalization to Extended Pedigrees of a Latent Class Model with Familial Dependence
Généralisation aux pedigrees étendus d'un modèle de classes latentes avec dépendance familiale

Aurélie Labbe, *Université Laval,* Alexandre Bureau, Arafat Tayeb & Chantal Merette, *Centre de recherche Université Laval Robert Giffard*

Considering clinical data arising from a collection of pedigrees, we developed a method to identify homogeneous disease sub-types based on multivariate disease measurements. We previously extended latent class analysis to allow dependence between the latent disease class status of relatives within nuclear families. Here, we present an algorithm to perform computations under our model in extended pedigrees, which recursively processes nuclear families like the Elston – Stewart algorithm in genetic linkage analysis. We present simulations illustrating the behaviour of our approach on various pedigree structures. We also present an illustration of our approach in genetic linkage analysis.

Considérant des données cliniques provenant d'un ensemble d'antécédents, nous avons développé une méthode permettant d'identifier des sous-types homogènes de maladies, basés sur un ensemble de mesures. Nous avons auparavant étendu les modèles de classes latentes en considérant la dépendance entre les classes latentes de maladie et les individus d'une même famille nucléaire. Ici, nous présentons un algorithme qui permet d'effectuer les calculs nécessaires à notre modèle, dans le cadre de familles élargies. Cet algorithme découpe chaque famille en familles nucléaires, similairement à l'algorithme de Elston-Stewart en analyse de liaison génétique. Nous présentons des simulations illustrant le comportement de notre approche en fonction de différentes structures des antécédents. Nous présentons aussi une illustration de notre approche au moyen d'une analyse de liaison génétique.

Session 17B Jeu. 29 mai • Thu. May 29, 10:30 – 12:00 Capital Hall 2A

Parametric Statistics on Varieties
Statistique paramétrique sur les variétés

[MS-361]
A Statistical Approach to Persistent Homology
Une approche statistique de l'homologie persistante

Peter Kim, *University of Guelph,* Peter Bubenik, *Cleveland State University*

Assume that a finite set of points is randomly sampled from some subspace. Recent advances in computational topology have provided several approaches to recovering the geometric and topological properties of the underlying space. In this paper, we take a statistical approach to this problem. We assume that the data is randomly sampled from an unknown probability distribution. We define two filtered complexes with which we can calculate the persistent homology of a probability distribution. Using statistical estimators for samples from certain families of distributions, we show that we can recover the persistent homology of the underlying distribution.

Supposons qu'un ensemble fini de points est échantillonné aléatoirement à partir d'un sous-espace. Des progrès récents en topologie numérique ont fourni plusieurs approches de recouvrement des propriétés géométriques et topologiques de l'espace sous-jacent. Dans la présente communication, nous adoptons une approche statistique de ce problème. Nous supposons que les données sont échantillonnées aléatoirement à partir d'une loi de probabilité inconnue. Nous définissons deux complexes filtrés à l'aide desquels nous pouvons calculer l'homologie persistante d'une loi de probabilité. En utilisant des estimateurs statistiques pour les échantillons tirés de certaines familles de lois, nous montrons que nous pouvons recouvrer l'homologie persistante de la loi sous-jacente.

[MS-362]
Nonparametric Statistical Analysis on 3D Shape Manifolds
Analyse statistique non paramétrique sur variétés de formes en trois dimensions

Victor Patrangenaru, *Florida State University*, Rabi Bhattacharya, *University of Arizona*

Nonparametric analysis on 3D shape manifolds arises in 3D scene identification in medical imaging, bioinformatics, pattern recognition, defense, etc. The type of shape depends on the method of image acquisition: 3D similarity shape analysis is useful in medical imaging and proteomics, while a statistical analysis on 3D projective shape manifolds or on 3D affine shape manifolds is the recommended approach in high level image analysis of regular digital camera outputs or of high resolution aerial or satellite images. After identifying the geometry structure of various shape manifolds, we give an extrinsic methodology for inference on 3D shape manifolds, and examples.

L'analyse non paramétrique des variétés de formes en trois dimensions (3D) intervient dans l'identification de scènes en 3D en imagerie médicale, en bio-informatique, en reconnaissance de motifs, en défense, etc. Le type de formes dépend de la méthode d'acquisition de l'image : l'analyse de similarité des formes en 3D est utile en imagerie médicale et en protéomique, tandis que l'analyse statistique sur variétés de formes projectives en 3D ou sur variétés de formes affines en 3D est l'approche recommandée dans l'analyse de haut niveau des images produites par des appareils photographiques numériques ordinaires ou d'images aériennes ou de satellite à forte résolution. Après avoir déterminé la structure géométrique de diverses variétés de formes, nous présentons une méthodologie extrinsèque pour l'inférence sur des variétés de formes en 3D, ainsi que des exemples.

[MS-363]
Statistical Analysis of Surface Shape
Analyse statistique dans des espaces de surfaces

Laurent Younes, *Johns Hopkins University*

We will discuss methods for the statistical analysis of datasets of surfaces in 3D spaces where the element of interest is the shape. The main domain of application is with anatomical shapes.

From a statistical point of view, the approach is relatively standard: we use an exponential chart around a well-chosen exemplar, or template. The implementation of this program is more challenging, however, starting from the definition of an infinite dimensional Riemannian manifold of surfaces, and carrying on with the study and numerical solutions of the basic equations in this manifold, including geodesic equations, parallel translation and Jacobi fields).

Nous présentons une méthodologie pour l'analyse statistique de données composées de surfaces dans $\mathbb{R}^3$. Nous discuterons en particulier l'analyse de leur forme, dans le but principal de traiter des données anatomiques.

D'un point de vue statistique, l'approche est relativement classique : elle est basée sur une approximation par une carte exponentielle autour d'une forme centrale convenablement choisie. La mise en œuvre de ce programme est quant à elle plus difficile, en commençant par la définition d'une structure de variété riemannienne sur les surfaces, et en allant jusqu'à l'étude et la résolution numérique d'équations de base dans cette variété, comme les équations géodésiques, le transport parallèle et les champs de Jacobi.

Session 17C Jeu. 29 mai • Thu. May 29, 10:30 – 12:00 Capital Hall 3A

Copulas
Copules

[MS-364]
Meta-elliptical Copulas and their Applications
Les copules méta-elliptiques et leurs applications

Anne-Catherine Favre, *Institut national de la recherche scientifique, Centre Eau, Terre et Environnement*, Christian Genest, *Université Laval*, Julie Béliveau, *Ministère de l'Éducation, du Loisir et du Sport*, Christiane Jacques, *Ministère des Affaires Municipales et des Régions*

Meta-elliptical copulas are introduced as a flexible tool for modeling multivariate data. These models are not new but their properties are scattered in the literature and inference procedures for this class of copulas are still in their early stages of development. The properties of this broad class of dependence functions are reviewed, along with associated rank-based procedures for copula parameter estimation and goodness-of-fit testing. A new graphical tool is also proposed for selecting an appropriate meta-elliptical copula. Hydrological data, namely peak, volume and duration of the annual spring flood for the Romaine River (Québec, Canada) are used for illustration purposes.

Les copules méta-elliptiques sont présentées comme un outil flexible pour modéliser des données multivariées. Ces modèles ne sont pas nouveaux, mais leurs propriétés sont disséminées dans la littérature et les procédures d'inférence n'en sont qu'à leurs balbutiements. Les propriétés de cette large classe de fonctions de dépendance sont revues ainsi que les procédures basées sur les rangs permettant d'estimer les paramètres de la copule et d'en tester l'adéquation. Un nouvel outil graphique est aussi proposé pour sélectionner une copule méta-elliptique appropriée. Des données hydrologiques composées de la pointe, du volume et de la durée de la crue maximale printanière de la rivière Romaine (Québec, Canada) sont utilisées à des fins illustratives.

[MS-365]
On Some Classes of Hierarchical Archimedean Copulas
Étude de certaines classes de copules archimédiennes hiérarchiques

Stéphane Loisel, *Université Claude Bernard Lyon 1*, Arthur Charpentier, *Université de Rennes 1*

In many fields of applications, like in credit risk, asymmetric dependence may be present in large dimensions. We consider extensions of Archimedean copulas called hierarchical Archimedean copulas that satisfy some bivariate dependence assumptions. We also study copulas "defined by blocks" and compare them with hierarchical ones. Statistical issues, in particular estimation of parameters, are addressed as well.

Dans plusieurs champs d'application, tel le risque de crédit, une dépendance asymétrique peut être présente en dimensions élevées. Nous considérons une extension des copules archimédiennes appelées copules archimédiennes hiérarchiques qui satisfont certains présupposés de dépendance bivariée. Nous étudions aussi des copules 'définies par blocs' et les comparons aux copules hiérarchiques. Nous répondons également à certaines difficultés statistiques, en particulier en estimation paramétrique.

[MS-366]
A Primer on Copulas for Count Data
Une introduction aux copules dans le contexte de données de dénombrement

Johanna Neslehova, *ETH Zurich*, Christian Genest, *Université Laval*

Various facts will be reviewed about copulas linking discrete distributions. It will be shown that the

Nous rappelons différents faits à propos des copules liant des distributions discrètes. Nous montrons que la possibi-

possibility of ties resulting from atoms in the probability distribution invalidates various familiar relations that lie at the root of copula theory in the continuous case. The talk will highlight some of the dangers and limitations of undiscriminating transposition of modeling and inference practices from the continuous setting to the discrete one.

lité d'égalités résultant d'atomes dans la fonction de masse invalide plusieurs relations familières au cœur de la théorie des copules dans le cas continu. L'exposé soulignera certains dangers et limitations de la transposition sans discernement de pratiques d'inférence et de modélisation du cas continu au cas discret.

Session 17D Jeu. 29 mai • Thu. May 29, 10:30 – 12:00 Capital Hall 4A

Empirical Likelihood
Vraisemblance empirique

[MS-367]
Empirical Likelihood for Non-smooth Criterion Functions
La vraisemblance empirique pour des fonctions de critère non-lisses

Ingrid Van Keilegom, *Université catholique de Louvain*, Elisa Molanes-Lopez, *Universidad Carlos III de Madrid*, Noel Veraverbeke, *Universiteit Hasselt*

Suppose that $X, X_1, \ldots, X_n$ is a sequence of independent and identically distributed random vectors. Let $\mu \in \mathbb{R}^p$ be a parameter of interest and $\nu \in \mathbb{R}^q$ be some nuisance parameter. The unknown, true parameters (μ_0, ν_0) are uniquely determined by the system of equations $E\{g(X, \mu_0, \nu_0)\} = 0$, where $g = (g_1, \ldots, g_{p+q})$ is a vector of $p + q$ functions. In this paper we develop an empirical likelihood method to do inference for the parameter μ_0. The results in this paper are valid under very mild conditions on the vector of criterion functions g. In particular, we do not require that $g_1, \ldots, g_{p+q}$ are smooth in μ or ν.

Supposons que $X, X_1, \ldots, X_n$ soit une séquence de vecteurs aléatoires indépendants et identiquement distribués. Soit $\mu \in \mathbb{R}^p$ un paramètre d'intérêt et $\nu \in \mathbb{R}^q$ un paramètre de nuisance. Les vrais paramètres inconnus (μ_0, ν_0) sont déterminés d'une façon unique par le système d'équations $E\{g(X, \mu_0, \nu_0)\} = 0$, où $g = (g_1, \ldots, g_{p+q})$ est un vecteur de $p + q$ fonctions. Dans ce papier nous développons une méthode de vraisemblance empirique pour faire de l'inférence pour le paramètre μ_0. Les résultats de ce papier sont valables sous des conditions très faibles sur le vecteur de fonctions de critère g. En particulier, nous ne supposons pas que $g_1, \ldots, g_{p+q}$ sont lisses en μ ou ν.

[MS-368]
A Spatial Empirical Likelihood for Variogram Estimation
Une vraisemblance spatiale empirique pour l'estimation de variogrammes

Dan Nordman & Petrutza Caragea, *Iowa State University*

The talk will describe a method for estimating variogram parameters with empirical likelihood (EL), which has close connections to least squares variogram fitting. The EL is formulated from blocks of spatial data lying on a grid. The EL method allows nonparametric confidence regions for variogram parameters to be calibrated with likelihood-ratio statistics and provides point estimators that are asymptotically as efficient as the best least squares estimators. Some simulation evidence is provided

Cet exposé décrira une méthode d'estimation de paramètres d'un variogramme par vraisemblance empirique, qui a de forts liens avec l'ajustement de variogramme par moindres carrés. La vraisemblance empirique est formulée à partir de blocs de données spatiales disposées sur une grille. La méthode par vraisemblance empirique permet aux régions de confiance non paramétriques pour les paramètres du variogramme d'être calibrées avec des statistiques de rapport de vraisemblance et produit des estimateurs ponctuels asymptotiquement aussi efficaces que les meilleurs estimateurs

to compare EL variogram estimation to other least squares approaches.

par moindres carrés. Des résultats de simulations comparant l'estimation de variogrammes par vraisemblance empirique aux autres approches par moindres carrés sont présentés.

[MS-369]
Conditional Variance Estimation in Heteroscedastic Regression Models
Estimation conditionelle de la variance en régression hétéroscédastique

Liang Peng, *Georgia Institute of Technology*

First, we propose a new method for estimating the conditional variance in heteroscedasticity regression models. For heavy tailed innovations, this method is in general more efficient than either of the local linear and local likelihood estimators. Secondly, we apply a variance reduction technique to improve the inference for the conditional variance. Thirdly, we applied empirical likelihood methods to construct interval estimation.

Premièrement, nous proposons une nouvelle méthode pour estimer la variance conditionnelle dans des modèles de régression hétéroscédastique. Pour des innovations à ailes relevées, cette méthode est en général plus efficace que les estimateurs localement linéaires où à vraisemblance locale. Deuxièmement, nous appliquons une technique de réduction de la variance pour améliorer l'inférence pour la variance conditionnelle. Troisièmement, nous appliquons des méthodes de vraisemblance empirique pour construire un intervalle d'estimation.

Session 17E Jeu. 29 mai • Thu. May 29, 10:30 – 12:00 Capital Hall 5A

Estimation Methods in Business and Industrial Statistics
Méthodes d'estimation en statistique industrielle et de gestion

[MS-370]
Statistical Estimation of Aircraft Infrared Signature Dispersion
Estimation statistique de la dispersion de la signature infrarouge d'un aéronef

Sidonie Lefebvre, Gérard Durand, Antoine Roblin & Suzanne Varet, *ONERA The French Aerospace Lab*

Existing models of aircraft infrared signature (IRS) do not account for the dispersion induced by uncertainty on input data (presentation angles, weather, etc.). As a result, they are of little use when estimating the detection performance of optronic systems. We thus propose a methodology for predicting IRS dispersion: – use of experiment designs to identify input variables with negligible influence on IRS and set them at a constant value, – Quasi-Monte Carlo estimation of the IRS dispersion, – set-up of a metamodel.We illustrate our approach with a typical air to ground attack example.

Les modèles existants de signature infrarouge d'aéronef (SIR) ne prennent pas en compte la dispersion induite par la méconnaissance des données d'entrée (angles de présentation, météorologie). Par conséquent, ils sont insuffisants pour évaluer les performances de détection des systèmes optroniques. Aussi proposons-nous une démarche pour prédire la dispersion de la SIR : – utilisation de plans d'expériences pour identifier les variables d'entrée d'influence négligeable sur la SIR, qui sont fixées à une valeur constante, – estimation quasi Monte-Carlo de la dispersion de la SIR, – conception d'un métamodèle.Nous illustrerons notre approche pour une attaque air-sol type.

[MS-371]

Kriging with Heterogeneous Nugget Effect for the Approximation of Noisy Simulators with Tunable Fidelity
Krigeage avec effet de pépite hétérogène pour l'approximation de simulateurs bruités à fidélité réglable

David Ginsbourger & Olivier Roustant, *Ecole des Mines de Saint-Etienne,* Yann Richet, *Institut de Radioprotection et de Sûreté Nucléaire*

Kriging is intensively used to interpolate costly deterministic computer experiments. Since many recent advances in simulation rely on probabilistic methods (Monte Carlo, etc...), it becomes necessary to take their randomness into account. Here we present an adaptation of the "Nugget Effect," originally developed in geostatistics to model observational noise, to the frame of computer experiments with tunable probabilistic noise. We propose a statistical model for such simulators, give a derivation of the associated Kriging equations (KEPH), and discuss the estimation of the covariance parameters. We finally consider perspectives such as the use of KEPH for noisy-simulator-based uncertainty propagation or optimization strategies.

Le krigeage est aujourd'hui fréquemment employé pour interpoler des résultats de simulation numérique déterministe. L'explosion des techniques stochastiques de simulation rend nécessaire la prise en considération d'aléas propres aux simulateurs. Nous présentons ici une adaptation de l'effet de pépite – originellement développé en géostatistique pour rendre compte d'irrégularités locales- pour le cadre des expériences numériques à fidélité variable. Nous proposons une modélisation statistique de tels simulateurs, donnons le détail des équations de krigeage associées (KEPH), et discutons l'estimation des paramètres de covariance. Nous présentons enfin des perspectives, en particulier au sujet de l'usage du KEPH pour l'optimisation et la propagation d'incertitudes.

[MS-372]

Depth and Homogeneity in Regional Flood Frequency Analysis
Profondeur et homogénéité dans l'analyse régionale de la fréquence des crues

Fateh Chebna & Taha, B.M.J. Ouarda, *Institut National de la Recherche Scientifique*

In regional flood frequency analysis, the restriction to a particular region corresponds to a loss of some information and the definition of a region generates a border effect. The proposed method is based on depth functions. The regression model is considered, where its parameters are estimated using an iterative procedure. Traditional approaches represent special cases of the depth-based approach. Along with flexibility and generality of the proposed method, it leads to estimation results that are superior to those obtained by the Canonical Correlation Analysis approach. Finally, the method is general and can be useful where the regression model is applicable.

Dans l'analyse régionale de la fréquence des crues, la restriction à une région particulière correspond à une perte d'information et la définition d'une région produit un effet de frontière. La méthode proposée est fondée sur des fonctions de profondeur. Nous considérons le modèle de régression dont les paramètres sont estimés par une méthode itérative. Les approches classiques représentent les cas particuliers de l'approche fondée sur la profondeur. Outre sa souplesse et sa généralité, la méthode proposée produit de meilleures estimations que celles obtenues par l'approche de l'analyse des corrélations canoniques. Enfin, la méthode est générale et peut s'avérer utile quand le modèle de régression est applicable.

[MS-373]
A Multivariate Robust Control Chart for Individual Observations
Graphique de contrôle robuste multivarié pour les observations individuelles

Shojaeddin Chenouri, *University of Waterloo*, Asokan Mulayath Variyath, *Texas A&M University*, Stefan H. Steiner, *University of Waterloo*

To monitor a multivariate process classical Hotelling's T^2 control chart is often used. However, it is well known that such control charts are very sensitive to the presence of outlying observations in the historical Phase I data used to set up its control limits. In this paper, we propose a robust Hotelling's T^2 type control chart for individual observations based on highly robust and efficient estimators of location and scatter known as re-weighted minimum covariance determinant estimators. We illustrate how to set the control limit for the proposed new control chart, and study its performance using simulations.

Pour contrôler un processus multivarié, le graphique de contrôle classique T^2 de Hotelling est souvent utilisé. Toutefois, il est de notoriété publique que ces graphiques de contrôle sont très sensibles à la présence d'observations se démarquant de la distribution dans les données historiques de Phase I utilisées pour définir leurs limites de contrôle. Dans cet article, nous proposons un graphique de contrôle robuste de type T^2 de Hotelling pour les observations individuelles, fondé sur des estimateurs extrêmement robustes et efficaces de localisation et de dispersion appelés estimateurs MCD (covariance de déterminant minimum) pondérés. Nous montrons comment fixer la limite de contrôle pour le nouveau graphique de contrôle proposé et étudions son efficacité par le biais de simulations.

[MS-374]
Risk Analysis of a Retail Bond Sales Strategy
Analyse du risque dans la stratégie de la vente d'obligations d'épargne au détail

Mark Reesor & Wei Xia, *University of Western Ontario*

Each year as part of their overall borrowing program, many governments sell savings bonds to retail investors. Examples of these bonds include Canada Savings Bonds and Ontario Savings Bonds. Each month of a sales campaign proceeds as follows: (i) the bond coupon stream is set at the start of the month; (ii) sales of this bond occur during the entire month; and (iii) the bonds sold that month are issued at the end of the month. This strategy subjects the issuer to interest-rate risks that we analyse using Monte Carlo.

Chaque année, dans le cadre de leur programme d'emprunt global, de nombreuses administrations publiques vendent des obligations d'épargne à des investisseurs privés. Les obligations d'épargne du Canada et les obligations d'épargne de l'Ontario en sont des exemples. Chaque mois, la campagne de vente se déroule de la façon suivante : (i) la série de coupons est établie au début du mois, (ii) les ventes de l'obligation ont lieu durant le mois complet et (iii) les obligations vendues durant le mois sont émises à la fin de celui-ci. Cette stratégie expose l'émetteur à un risque de taux d'intérêt que nous analysons par une méthode de Monte-Carlo.

[MS-375]
Covariance Kernels for Spatial Interpolation of Symmetrical Functions
Noyaux de covariance pour le krigeage de fonctions symétriques

David Ginsbourger, Xavier Bay & Laurent Carraro, *Ecole des Mines de Saint-Etienne*

Learning a deterministic function using a Gaussian process relies on the selection of a covariance kernel. When some prior information is available concerning symmetries of the function to be approximated, it is clearly unreasonable not to use it in the choice of the kernel. We propose a characterization

L'apprentissage statistique d'une fonction déterministe par un processus gaussien nécessite de sélectionner un noyau de covariance. Lorsque l'on dispose *a priori* d'informations sur les symétries de la fonction que l'on souhaite approximer, il est fort dommageable de ne pas les utiliser à l'étape du choix du noyau. Nous proposons une caractérisation des noyaux de

of the kernels which associated Gaussian processes have their paths invariant under the action of a finite group of transformations. We then give an example of such symmetrical processes, built on the basis of stationary Gaussian processes, and having interesting regularity properties.

covariance dont les processus gaussiens associés possèdent des réalisations invariantes par l'action d'un groupe fini de transformations. Nous donnons ensuite un exemple de tels processus symétriques, construits sur la base de processus gaussiens stationnaires, et jouissant d'intéressantes propriétés de régularité.

Session 17F Jeu. 29 mai • Thu. May 29, 10:30 – 12:00 Capital Hall 6A

Mixture Models
Modèles de mélanges

[MS-376]

Bayes Minimax Estimators of the Mean of a Scale Mixture of Multivariate Normal Distributions
Les estimateurs de Bayes minimax de la moyenne des distributions mélangées de lois normales

Othmane Kortbi, *Université de Sherbrooke*

Bayes estimation of the mean of a variance mixture of multivariate normal distributions is considered under sum of squared errors loss. We find broad class of priors which result in proper and generalized Bayes minimax estimators. This paper extends the results of Strawderman and Maruyama but somewhat more in the spirit of Fourdrinier et al. for the normal case, in the sense that we construct classes of priors giving rise to minimaxity. A main condition for minimaxity is that the mixing distributions of the sampling distribution and the prior distribution satisfy a monotone likelihood ratio property with respect to a scale parameter.

Dans ce travail, nous considérons l'estimation bayesienne de la moyenne d'une distribution de mélanges de lois normales multivariées selon le coût quadratique. Nous avons trouvé une grande classe de lois a priori (généralisées) qui génèrent des estimateurs minimax. Ce travail est principalement une extension du travail de Strawderman et de Maruyama et les techniques utilisées sont standards. Cependant, la condition clé de régularité de la minimaxité est décrite d'une manière élégante par une propriété de monotonie du rapport de vraisemblance des deux densités de mélanges de la distribution d'échantillonnage et de la distribution *a priori*.

[MS-377]

Combining Gaussian Components for Clustering: Mixtures of Mixtures
Classifier en agrégeant des composantes gaussiennes : mélanges de mélanges

Jean-Patrick Baudry, *Université Paris-Sud*, Gilles Celeux, *INRIA*

Model-based clustering consists of fitting a mixture model to data and identifying each cluster with one of its components. Gaussian components are typically used. But clusters of interest might be poorly fitted through Gaussian distributions. The approximation properties of Gaussian mixtures, provided the number of components is sufficient, make them attractive nevertheless. Consequently, each cluster is identified to a mixture. We suggest selecting the total number of Gaussian components through BIC and then combining some of them hierarchically according to an entropy criterion so as to get a sensible number of clusters. This method will be discussed through numerical experiments.

La classification statistique consiste à identifier chaque classe à une composante d'un modèle de mélange – gaussien typiquement – ajusté à des données. Mais une classe intéressante est parfois mal approchée par une loi gaussienne. Les mélanges gaussiens restent cependant intéressants pour leurs bonnes propriétés d'approximation lorsque le nombre de composantes est suffisant. Nous proposons donc d'identifier chaque classe à un mélange. Le nombre total de composantes est choisi par le critère BIC, puis certaines d'entre elles sont agrégées selon un critère d'entropie jusqu'à obtenir un nombre judicieux de classes. Nous commenterons cette approche hiérarchique en nous appuyant sur des expérimentations numériques.

[MS-378]
Weighted Poisson Mixed Model for Underdispersed Longitudinal Count Data
Modèle de Poisson pondéré mixte pour données longitudinales de comptage sous-dispersées

Catherine Trottier, Christian Lavergne & Mariem Zouch, *I3M - Montpellier*

When analysing the elongation of leafy axes, a modelisation of the number of new leaves that appeared during successive time periods is sought. For this type of longitudinal count data, a clear under-dispersion is often noticed. We propose a model based on weighted Poisson distribution WPD(λ, r, a), where parameter r expresses the over- or under-dispersion and the effect of covariates is introduced via parameter λ. With fixed a, taking advantage of the bi-dimensional exponential structure, we propose a Fisher type estimation algorithm and study its performances in simulations.

Lors de l'analyse de l'allongement d'axes feuillés, on cherche à modéliser le nombre de nouvelles feuilles apparues sur des périodes de temps successives. Ce type de données longitudinales de comptage présente souvent une nette sous-dispersion. Nous proposons une modélisation à l'aide d'une distribution de Poisson pondérée WPD(λ, r, a), où le paramètre r exprime le lien à la sur-dispersion ou à la sous-dispersion et l'effet des covariables est introduit au niveau du paramètre λ. En tirant parti, à a fixé de la structure exponentielle bi-dimensionnelle, nous proposons un algorithme d'estimation de type Fisher et étudions ses performances sur simulations.

[MS-379]
On a Mixture of Von Mises and Uniform Distributions
À propos d'un mélange de loi de von Mises et de loi unifrome

John Bentley, Richard Lockhart & Michael Stephens, *Simon Fraser University*

The von Mises distribution is often used for modelling directional data around a circle, exhibiting a single cluster, for instance, movements of insects reacting to light or flights of humming birds. However, some data sets appear to contain a group of points which might be uniformly distributed around the circle. We consider a model of a mixture of von Mises and uniform distributions. Maximum likelihood estimation is used to estimate parameters for the mixture model. Goodness-of-fit test procedures are presented for selecting the appropriate model (uniform, von Mises, or mixture) and determining its adequacy.

La loi de von Mises est souvent utilisée pour modéliser autour d'un cercle des données directionnelles qui présentent une grappe unique, par exemple, les mouvements d'insectes qui réagissent à la lumière ou des vols de colibris. Cependant, certains ensembles de données semblent contenir un groupe de points qui pourraient être distribués uniformément autour du cercle. Nous considérons un modèle de mélange d'une loi de von Mises et d'une loi uniforme. Nous utilisons l'estimation du maximum de vraisemblance pour estimer les paramètres du modèle de mélange. Nous présentons des procédures de test de la qualité de l'ajustement pour choisir le modèle approprié (uniforme, von Mises ou mélange) et pour déterminer son adéquation.

[MS-380]
Moment Method Estimation of Mixing Distributions
Estimation d'un mélange de distributions par la méthode des moments

Zhengmin Zhang, A.K.Md. Ehsanes Saleh & Patrick J. Farrell, *Carleton University*

We establish the identifiability of a negative binomial mixture, and the identifiability of a normal mixture. By the moment method, we construct a consistent nonparametric estimate of the mixing distribution in these two mixture models.

Nous établissons l'identifiabilité d'un mélange de lois binomiales négatives et celle d'un mélange de lois normales. Par la méthode des moments, nous construisons une estimation non paramétrique convergente du mélange de distributions dans ces deux modèles de mélange.

[MS-381]
Homogeneity Test in Normal Mixture Models: the EM Approach
Test d'homogénéité pour des modèles de mélanges de normales : l'approche EM

Pengfei Li & Jiahua Chen, *University of British Columbia*

The class of normal mixture distributions is arguably most important mixture models, yet it is technically most challenging, its likelihood function is unbounded based on a set of random samples if no artificial bounds are placed on its component variance parameter; it is not strongly identifiable and it has infinite Fisher information with respect to mixing proportions. We show that an EM-test for homogeneity is effective at overcoming many challenges in the context of finite normal mixtures. Its limiting distribution is very simple and can be easily used in applications. A real example is used to illustrate the application of the EM-test.

La classe des mélanges de distributions normales est vraisemblablement le modèle de mélange le plus important, bien qu'elle représente un grand défi technique : sa fonction de vraisemblance basée sur un ensemble d'échantillons aléatoires n'est pas bornée si nous ne plaçons pas de bornes artificielles sur ses paramètres de composantes de la variance ; elle n'est pas fortement identifiable et elle a une information de Fisher infinie par rapport aux proportions du mélange. Nous montrons un Test EM pour l'homogénéité qui permet de surmonter plusieurs difficultés dans le contexte des mélanges finis de lois normales. Sa distribution limite est très simple et peut aisément être utilisée dans plusieurs applications. Un exemple réel est utilisé afin d'illustrer l'application du test EM.

Session 17G Jeu. 29 mai • Thu. May 29, 10:30 – 12:00 Capital Hall 7A

Time Series Modelling
Modélisation de séries chronologiques

[MS-382]
Definition and Estimation of Biological Reference Points for Halieutic Resource Management in Stochastic Frameworks.
Définition et estimation de repères biologiques pour la gestion de ressources halieutiques dans des cadres d'étude stochastiques

Nicolas Bousquet, Thierry Duchesne, Louis-Paul Rivest & Jérôme Lemay, *Université Laval*

When setting up exploitation quotas, managers of halieutic resources may face statistical issues, some of which are investigated here. The frequent choice of deterministic, parametric state-space models for modeling the temporal variation of the resources provides explicit biological reference points, such as the maximum sustainable yield (MSY) which appears as an ideal objective for the exploiters. Parameter estimation use noisy observations of indirect measures of abundance. To improve their relevance, environmental errors are incorporated into the models. This makes parameter estimation more difficult. A key requirement to avoid overoptimistic harvesting strategies is to redefine the reference points in these stochastic frameworks.

Nous examinons quelques problèmes statistiques rencontrés par les gestionnaires de ressources halieutiques pour l'établissement des quotas d'exploitation. Le choix fréquent de modèles paramétriques à espaces d'état pour représenter la variation temporelle d'une ressource permet de définir des indicateurs biologiques explicites. Ainsi, le rendement maximal admissible (MSY), c'est-à-dire le plus grand prélèvement moyen d'une ressource qui ne compromette pas son renouvellement, constitue un idéal. L'estimation résulte d'observations bruitées indirectes. Pour gagner en pertinence, des erreurs environnementales sont intégrées aux modèles. Si l'estimation devient plus délicate, les indicateurs biologiques doivent d'abord être redéfinis dans ce cadre stochastique sous peine d'être dangereusement optimistes.

[MS-383]

Application of Semiparametric Modelling to Times Series Forecasting. Case of the Electricity Consumption
Modèles semi-paramétriques appliqués à la prévision des séries temporelles. Cas de la consommation d'électricité

Vincent Lefieux, *RTE/DMA - CREST ENSAI*

RTE, in charge of operating the French electric transportation grid, needs an accurate forecast of the power consumption in order to operate it correctly. Nonparametric forecasting methods have already been tested without real success, due to the 'curse of dimensionality'. Recently, semi-parametric methods which improve on the pure non-parametric approach have been proposed to estimate a regression function. Based on the concept of 'index', one of those methods called MAVE can apply to time series. We propose a semi-linear and semi-parametric model, based on the MAVE method, which allows us to simultaneously take into account the autoregressive aspect of the problem and the exogenous variables.

Une prévision correcte de la consommation d'électricité est fondamentale pour le bon fonctionnement du réseau électrique français, dont RTE a la charge. Des méthodes de prévision non paramétriques ont déjà été testées sans succès véritable, notamment à cause du « fléau de la dimension ». On a proposé récemment des méthodes semi-paramétriques d'estimation d'une régression qui améliorent l'approche non paramétrique pure. L'une d'elles, basée sur la notion de « directions révélatrices » appelée MAVE, peut s'appliquer aux séries temporelles. Nous proposons un modèle semi-paramétrique semi-linéaire, basé sur la méthode MAVE, qui prend simultanément en compte l'aspect autorégressif du problème, et l'introduction de variables exogènes.

[MS-384]

An Analysis of Biogeochemical Variables from Bedford Basin, NS, Canada Using a Trend Plus Cycle State Space Model
Une analyse des variables bio-géo-chimique de Bedford Basin, NE, Canada en utilisant un cycle de tendance plus un modèle d'espace état

Elizabete Almeida, Michael Dowd & Joanna Flemming, *Dalhousie University*, William Li, *Fisheries and Oceans Canada*

Abundance levels for plankton and other biogeochemical variables were recorded between 1992 and 2006 in Bedford Basin, NS, Canada as part of an ongoing oceanographic monitoring program. Results are presented from time series analysis of these data. Particularly, a trend plus cycle state space model is developed and applied to these data. This allows for the seasonal cycle of the abundance levels to be adaptive, thereby allowing us to examine interannual changes in the seasonal cycle (in terms of timing and level). Parameter estimation for this model is considered. Finally, a trend analysis is conducted on the optimal state estimates.

Les niveaux d'abondance du plancton et d'autres variables biogéochimiques ont été enregistrés entre 1992 et 2006 dans le bassin de Bedford, en Nouvelle-Écosse, au Canada, dans le cadre d'un programme permanent de surveillance océanographique. Nous présentons les résultats de l'analyse chronologique de ces données. En particulier, nous élaborons d'un modèle d'espace d'état avec tendance-cycle que nous appliquons à ces données. Le cycle saisonnier des niveaux d'abondance peut ainsi être adaptatif, ce qui nous permet d'examiner ses variations interannuelles (en ce qui concerne le moment dans le temps et le niveau). Nous examinons l'estimation des paramètres de ce modèle. Enfin, nous effectuons une analyse de tendance sur les estimations d'état optimal.

[MS-385]
Dynamic Factors in Periodic Time-varying Regression Models
Facteurs dynamiques dans des modèles de régression périodiques à coefficients variables dans le temps

Virginie Dordonnat, *EDF R&D, France / VU University Amsterdam,* Siem Jan Koopman & Marius Ooms, *VU University Amsterdam,* Alain Dessertaine, *EDF R&D France*

We consider dynamic multivariate periodic regression modelling for high frequency data. The dependent univariate time series is transformed to a lower frequency multivariate time series for periodic regression modelling. For hourly series, we specify one equation per hour of the day. The regression coefficients differ across equations and vary stochastically over time. Since the unrestricted model contains many unknown parameters, we develop a methodology within the state-space framework to model dynamic factors in the coefficients, with common coefficient dynamics across equations. We apply our method to French national hourly electricity loads with weather variables and calendar variables as regressors.

Nos travaux portent sur des modèles multivariés de régressions dynamiques, périodiques pour des données à granularité temporelle fine. La série temporelle univariée à modéliser est transformée en une série multivariée à granularité moins fine. Les coefficients de régression sont différents selon l'équation et varient dans le temps de manière stochastique. Le modèle sans contraintes dépendant de nombreux paramètres, nous développons une méthodologie dans le cadre des modèles à espaces d'états pour modéliser les coefficients à travers les équations à l'aide de facteurs dynamiques communs. Nous appliquons notre méthode à la consommation horaire d'électricité en France, avec des variables explicatives calendaires et météorologiques.

Session 18A Jeu. 29 mai • Thu. May 29, 13:30 – 14:30 Congress Hall A/E

Address of the winner of th Centre de Recherches Mathématiques/Statistical Society of Canada Prize

Allocution du récipiendaire du prix Centre de recherches mathématiques/Société statistique du Canada

[MS-386]
The In-between World of Proper Priors and Nonidentified Models
L'entre deux mondes des lois *a priori* intégrables et des modèles non-identifiables

Paul Gustafson, *University of British Columbia*

An identified model and root-n consistent estimation of parameters is as good as it gets statistically. Conversely, a nonidentified model and no prior information seems to be as bad as it gets. This talk examines the in-between world of nonidentified models with proper prior distributions. We discuss how nonidentified models arise naturally, and even necessarily, particularly in observational-study contexts. Then we examine the performance of Bayesian estimators arising from such models, showing how sometimes this depends less on the choice of prior distribution than one might guess. Practical implications and recommendations will be discussed.

Un modèle identifié et des estimateurs cohérents à l'ordre $\sqrt{n}$ des paramètres est la meilleure situation possible pour un statisticien. Réciproquement, un modèle non identifié, sans information *a priori* sur les paramètres semble être la pire situation possible. Cet exposé examine le cas intermédiaire de modèles non identifiés avec distributions *a priori* propres. Nous discutons de l'émergence naturelle, et même nécessaire, de modèles non identifiés, particulièrement dans le contexte d'études observationnelles. Nous examinons ensuite la performance d'estimateurs bayésiens associés à de tels modèles, et montrons comment celle-ci est parfois moins dépendante du choix de la distribution *a priori* que l'on pourrait croire. Nous discutons finalement d'implications pratiques et de recommandations.

Session 19A Jeu. 29 mai • Thu. May 29, 14:35 – 15:35 Capital Hall 1A

Isobel Loutit Invited Address on Business and Industrial Statistics
Allocution Isobel Loutit sur la statistique industrielle et de gestion

[MS-387]
Industrial Statistics: The Gap between Research and Practice
Statistiques industrielles : l'écart entre la recherche et la pratique

Vijay Nair, *University of Michigan*

Statistical methods have contributed extensively to the development of automotive, electronic, chemical, telecommunications, and other industries over the past century. The industrial applications have in turn driven developments in statistical methodology. In recent years, however, there appears to be an increasing disconnect between research and practice in industrial statistics. In this talk, I will use specific examples to make this case (on both sides of the coin) and, with help from the audience, explore possible reasons and remedies.

Les méthodes statistiques ont contribué considérablement au développement des industries de l'automobile, de l'électronique, des produits chimiques, des télécommunications et d'autres au cours du dernier siècle. Les applications industrielles ont à leur tour dicté les progrès en méthodologie statistique. Cependant, ces dernières années, le fossé semble s'être creusé entre la recherche et la pratique en statistique industrielle. La communication fournira des exemples précis en vue de défendre cette thèse (en examinant les deux côtes de la question) et, avec l'aide des participants, en explorera les raisons et les remèdes éventuels.

Session 19B Jeu. 29 mai • Thu. May 29, 14:35 – 16:05 Capital Hall 2A

Synthetic Data Methods for Confidentiality Protection and Disclosure Avoidance
Méthodes de données synthétiques pour la protection de la confidentialité et de la non-divulgation

[MS-388]
Data Synthesis via Expert Knowledge, Modeling, and Matching
Construction de données synthétiques au moyen de connaissances, de modélisation et d'appariement

Sam Hawala, *US Census Bureau*

The focus of the paper is a method to produce synthetic data through the combined use of expert knowledge, model fitting, and matching using the model predicted values. All three elements play an important role in the successful reproduction of the aggregate behavior and the main features of a data set. Analysts at the U.S. Census Bureau provide expert knowledge which we incorporate into the synthesis so that the synthetic data passes all edit requirements. The Bureau relies on the synthetic data method in conjunction with traditional disclosure avoidance techniques to protect data confidentiality.

Le point central de mon article est une méthode pour produire des données synthétiques grâce à l'utilisation combinée de l'expertise sur le genre des données en question, des modèles appropriés, et l'utilisation des valeurs prédites pour trouver des donateurs. Ces trois éléments jouent un rôle important dans le succès de reproduction du comportement général et les caractéristiques principales de l'ensemble des données. Les analystes de l'US Census Bureau fournissent des connaissances spécialisées qu'on introduit dans la synthèse afin que les données synthétiques réussissent la phase d'éditions des données. Le Bureau s'appuie sur les données synthétiques en combinaison avec des techniques traditionnelles pour la confidentialité des données.

[MS-389]
Synthetic Data Methods: A Discussion of the Most Pressing Challenges
Méthodes de données synthétique : une discussion sur les défis les plus pressants

Jerome Reiter, *Duke University*

Many statistical agencies are considering the use of multiple imputations to protect confidentiality, also called synthetic data approaches. While much of the inferential methods for synthetic data have been developed, there remain many challenges to implementing synthetic data approaches successfully. This talk reviews the main challenges and discusses potential approaches to address those challenges. These challenges represent some of the main areas of research in synthetic data methodology.

Un grand nombre d'organismes statistiques envisagent de recourir à l'imputation multiple, également appelée approche des données synthétiques, pour protéger les renseignements confidentiels. Bien qu'une grande part des méthodes d'inférence applicables aux données synthétiques aient été élaborées, la mise en Suvre fructueuse de l'approche des données synthétiques continue de poser de nombreux défis. La communication comprendra l'examen des principaux défis et une discussion des approches qui pourraient être adoptées pour les relever. Ces défis comptent parmi les principaux domaines de recherche relatifs aux méthodes des données synthétiques.

[MS-390]
Combining Methods to Create Synthetic Microdata: Quantile Regression, Hot Deck, and Rank Swapping
Combiner des méthodes pour créer des microdonnées synthétiques : régression quantile, hot deck et permutation de rangs

Jennifer Hucketт & Michael Larsen, *Iowa State University*

Government agencies must release data while upholding confidentiality. We propose combining disclosure limitation methods. Conditional quantile regression models describe complex relationships. Hot deck selection of donors also captures complex relationships. Rank swapping perturbs data. We propose to simulate values for confidential variables using quantile regression predictions at randomly selected quantiles. We impute values for additional variables using hot deck imputation with rank swapping. The resulting microdata have low disclosure risk and high data utility. Developments include simulation methods and measures of disclosure risk and inferential value. Applications include U.S. Census veterans' and Iowa Department of Revenue tax returns data.

Les organismes gouvernementaux doivent diffuser des données tout en veillant au respect de la confidentialité. Nous proposons de combiner des méthodes de contrôle de la divulgation. Les modèles de régression quantile conditionnelle décrivent des relations complexes. La sélection hot deck de donneurs reflète aussi des relations complexes. La perturbation de rangs perturbe les données. Nous proposons de simuler des valeurs pour des variables confidentielles en utilisant des prédictions par régression quantile pour des quantiles choisis au hasard. Nous imputons les valeurs pour des variables supplémentaires en utilisant l'imputation hot deck avec permutation de rangs. Les microdonnées résultantes présentent un faible risque de divulgation et une utilité élevée. Les travaux de développement comprennent des méthodes de simulation, ainsi que des mesures du risque de divulgation et de la valeur inférée. Les applications s'appuient sur des données sur les vétérans du recensement des États-Unis et des données provenant des déclarations de revenus de l'Iowa Department of Revenue.

Session 19C Jeu. 29 mai • Thu. May 29, 14:35 – 16:05 Capital Hall 3A

Recent Advances in Analysis of Event Data
Développements récents dans l'analyse des données événementielles

[MS-391]

Flexible Modeling of Time-dependent and Non-linear Effects of Continuous Covariates in Relative Survival
Modélisation flexible des effets dépendants du temps et non loglinéaires des variables continues en survie relative

Michal Abrahamowicz, *McGill University,* Amel Mahboubi & Christine Binquet, *Université de Bourgogne,* Roch Giorgi, *Université de la Méditerranée,* Claire Bonithon-Kopp & Catherine Quantin, *Université de Bourgogne*

We address 3 challenges encountered in studies of prognostic factors for mortality: (1) the unknown cause of death; (2) time-dependent effects; (3) non-linear effects. We extend the Esteve's additive relative survival model, and model disease-specific mortality hazard as: $hd(t) = \exp(\gamma(t)) \times \exp[\alpha(x)\beta(t)]$, where: $\gamma(t)$ is the baseline hazard, $\alpha(x)$ is a non-linear transformation of X, and $\beta(t)$ represents the time-dependent effect of X. All 3 functions $\gamma(t)$, $\alpha(x)$, and $\beta(t)$, are estimated with regression splines using the 3-step alternating conditional algorithm. We apply the method to re-assess the effects of prognostic factors on survival in colon cancer.

Nous abordons trois défis rencontrés dans les études pronostiques de mortalité : (1) la cause de décès est souvent inconnue ; les effets de covariables sont (2) dépendants du temps et (3) non linéaires. Afin de pallier ces problèmes, nous proposons une extension flexible du modèle de survie relative d´Estève, et modélisons la mortalité spécifique à la maladie comme : $hd(t) = \exp[\gamma(t)] \exp[\alpha(x)\beta(t)]$ Les 3 fonctions : $\gamma(t)$ = risque de base, $\alpha(x)$ = transformation non linéaire de X, $\beta(t)$ = effet dépendant du temps, sont estimées avec des splines, par un algorithme itératif conditionel de 3 étapes. Le modèle est appliqué pour estimer la survie dans cancer du colon.

[MS-392]

Nonparametric Estimation with Recurrent Events — Application to Safety Data in Clinical Trials
Estimation non paramétrique pour des événements récurrents – Application aux données de tolérance dans les études cliniques

Gérard Derzko, *Sanofi-Aventis R&D*

In long-term clinical trials, the safety of studied drugs is mainly assessed through the point and accuracy estimations of a variety of time functions associated with the marked point processes of recurrent adverse events. Even under complex conditions, including competing risks, censoring and terminal events, some of these functions satisfy appropriately defined differential systems. Nonparametric consistent process functions estimators are obtained through discretization and simple solving algorithms of these systems. Kernel Smoothed Histograms (KSH) allow for curve estimation of derivative process functions. Nonparametric simultaneous confidence bands, accounting for the common over-dispersion due to population heterogeneity, are best obtained through bootstrapping.

Dans les études cliniques à long terme, la tolérance des produits étudiés est principalement évaluée par des estimations centrales et de précision de diverses fonctions du temps associées aux processus ponctuels marqués des effets adverses récurrents. Sous des conditions mêmes complexes, telles que des risques compétitifs, censure ou événements terminaux, certaines de ces fonctions satisfont à des systèmes différentiels simples et discrétisables, dont les solutions fournissent des estimateurs non paramétriques convergents. L'estimation de fonctions dérivées nécessite un lissage à noyau discret (KSH). Des bandes de confiance simultanées intégrant le phénomène usuel de surdispersion causé par l'hétérogénéité des populations, sont obtenues par bootstrap.

[MS-393]
Forecasting of Recurrent Events: a Bayesian Approach
Approche bayésienne pour la prédiction d'événements récurrents

Marc Fredette, *HEC Montréal*

We will discuss methods for predicting the eventual total number of recurrent events observed over a period of time. As recurrent events data accumulate, methods based on Poisson processes with random effects often give appropriate frequentist prediction intervals. However, these methods fail to provide adequate intervals when they are based on early or limited data. Extensions incorporating experts' knowledge and other types of prior information will be discussed. An application to the prediction of automobile warranty claims will be considered but possible applications in biostatistics will also be discussed.

Nous présenterons des méthodes afin de prédire le nombre total de manifestations récurrentes au sein d'une population donnée. Lorsque de tels événements récurrents sont observés dans le temps, des méthodes utilisant des processus de Poisson avec des effets aléatoires mènent habituellement à des intervalles de prédiction (fréquentistes) appropriés. Toutefois, ces méthodes sont inadéquates lorsqu'une faible fraction du nombre total des événements est observée. Nous présenterons donc des méthodes utilisant de l'information *a priori* afin d'améliorer ces intervalles. Une application à la prévision du nombre de réparation pour des véhicules sous garantie sera considérée. Cependant, quelques applications possibles en biostatistique seront également discutées.

Session 19D Jeu. 29 mai • Thu. May 29, 14:35 – 16:05 Capital Hall 4A

Time Series
Séries chronologiques

[MS-394]
On Modeling and Diagnostic Checking Vector Periodic Autoregressive Time Series Models
De la modélisation et l'ajustement des modèles de séries chronologiques autorégressives périodiques vectorielles

Pierre Duchesne & Eugen Ursu, *Université de Montréal*

Vector periodic autoregressive time series models form an important class of models. We derive the asymptotic distributions of the least squares estimators of the parameters, allowing the parameters in a given season to satisfy linear constraints. We also obtain the asymptotic distribution of the residual autocovariance matrices, and we deduce the asymptotic distribution of the residual autocorrelation matrices. Portmanteau test statistics designed for diagnosing the adequacy of vector periodic models are introduced and we study their asymptotic distributions. The proposed test statistics are illustrated in a small simulation study, and an application with bivariate quarterly West German data is presented.

Les modèles de séries chronologiques autorégressifs périodiques et vectoriels forment une classe importante de modèles. Nous obtenons les distributions asymptotiques des estimateurs des moindres carrés des paramètres, permettant les paramètres d'une saison donnée de satisfaire des contraintes linéaires. Nous obtenons aussi la distribution asymptotique des matrices d'autocovariances résiduelles, et nous déduisons la distribution asymptotique des matrices d'autocorrélations résiduelles. Des statistiques de test de type portemanteau désignées pour diagnostiquer l'ajustement des modèles périodiques vectoriels sont introduites et nous étudions leur distribution asymptotique. Les statistiques de test proposées sont illustrées dans une petite étude de simulation, et une application avec des données bivariées de l'Allemagne de l'Ouest est présentée.

[MS-395]
Tests of Serial Dependance
Tests de dépendance sérielle

Bruno Rémillard, *HEC Montréal*, Christian Genest, *Université Laval*, Kilani Ghoudi, *United Arab Emirates University*

New tests of independence for the innovations of time series are proposed. These tests are based on functionals of empirical processes constructed from the residuals or their normalized ranks. Under some conditions, the asymptotic behavior under the hypothesis of independence is distribution-free. Many test statistics are proposed, including the well-known BDS and its rank-based counterpart. Monte Carlo experiments are used to illustrate that the new tests are quite powerful under several dependence alternatives.

On propose de nouveaux tests d'indépendance pour les innovations dans les modèles de séries chronologiques. Ces tests sont basés sur des fonctionelles de processus empiriques construits à partir de résidus ou des rangs des résidus. Sous certaines conditions, leur comportement asymptotique ne dépend ni des marges ni des paramètres du modèle. Plusieurs statistiques sont déduites de ces processus dont la statistique BDS, ainsi que son analogue basé sur des rangs. Des expériences Monte Carlo sont utilisées pour montrer que ces nouveaux tests sont très puissants par rapport à plusieurs hypothèses alternatives de dépendance.

[MS-396]
Relative Efficiencies of Tests on GARCH Coefficients
Efficacités relatives de tests sur les coefficients GARCH

Jean-Michel Zakoian, *CREST et Université Lille 3*, Christian Francq, *Université Lille 3*

This presentation is concerned with testing the nullity of coefficients in GARCH models. The problem is non standard because the quasi-maximum likelihood estimator is subject to positivity constraints. We establish the asymptotic null and local alternative distributions of Wald, score, and quasi-likelihood ratio tests. Efficiency comparisons under fixed alternatives are also considered. Two cases of special interest are: tests of the null hypothesis of one coefficient equal to zero, and tests of the null hypothesis of no conditional heteroscedasticity. The theoretical results are illustrated by means of simulation experiments and an empirical application is proposed.

L'exposé porte sur les tests de nullité de coefficients dans les modèles GARCH. Le problème est non standard en raison des contraintes de positivité. Nous établissons les distributions asymptotiques, sous l'hypothèse nulle et sous des alternatives locales, des tests de Wald, du score et du quasi-rapport de vraisemblance. Des comparaisons d'efficacité sous des alternatives fixes sont aussi effectuées. Deux cas particuliers importants d'hypothèse nulle sont : la nullité d'un seul coefficient et l'absence d'hétéroscédasticité conditionnelle. Les résultats théoriques sont illustrés par des simulations et une application empirique est proposée.

Session 19E Jeu. 29 mai • Thu. May 29, 14:35 – 16:05 Capital Hall 5A

Latent Variables
Variables latentes

[MS-397]
Two-part Latent Variable Models for Semicontinuous Longitudinal Data
Modèles à variable latente pour les données longitudinales semi-continues

David Flora, *York University*

Semicontinuous variables with a large proportion of observations equal to a single value (typically zero) with the remaining observations continuously distributed (above zero) are common in longitudinal behavioural research. After presenting the basic latent growth model, I describe Olsen and Schafer's (2001) joint two-part model for semicontinuous variables; Part 1 is a latent growth model for a binary outcome (zero vs. non-zero) and Part 2 is a latent growth model for a continuous outcome (given non-zero response in Part 1). This model is then compared to an alternative two-part model where Part 1 is instead a discrete-time hazard model.

Les variables semi-continues ayant une grande proportion d'observations égales à une valeur simple (habituellement zéro) et ayant des observations restantes distribuées de façon continue (au dessus de zéro) sont communes dans la recherche longitudinale sur le comportement. Après avoir présenté le modèle de croissance latente de base, je décrirai le modèle conjoints à deux parties pour les variables semi-continues de Olsen et Schafer (2001). La partie 1 est un modèle de croissance latente pour un résultat binaire (zéro comparé à différent de zéro) et la partie 2 est un modèle de croissance latente pour un résultat continue (étant donné la réponse différente de zéro dans la partie 1). Ce modèle est alors comparé à une modèle alternatif à deux parties où la partie 1 est plutôt un modèle à temps discret à risque.

[MS-398]
Latent Variables: Cautionary Remarks and Alternative Approaches
Variables latentes : avertissements et approches alternatives

James Ramsay, *McGill University*

The latent variable formulations of factor analysis and test theory in psychometrics has led to unfortunate misinterpretations by researchers, and has perhaps even limited the developments of these fields. Alternatively, a latent variable can be seen as a functional parameter that defines a space curve or higher dimensional manifold within observation space. This idea is illustrated for these two fields, and the re-interpretation of a latent variable as one among an infinite number of charting maps on the manifold avoids the dangers of misinterpretation inherent in the "variable" part of the phrase.

Les formulations des variables latentes en analyse factorielle et en théorie des tests dans un cadre psychométrique ont mené à de malencontreuses mauvaises interprétations par des chercheurs, et ont peut-être même limité le développement de ces champs d'études. Alternativement, une variable latente peut être vue comme un paramètre fonctionnel définissant une courbe gauche où une variété de plus haute dimension dans l'espace des observations. Cette idée est illustrée pour ces deux champs d'études, et la ré-interprétation d'une variable latente comme une parmi un nombre infini de cartes d'une variété permet d'éviter les dangers de mauvaises interprétations inhérentes à la partie « variable » de la phrase.

[MS-399]
Latent Variable Measurement Models: Equivalence, Invariance and Identifiability
Modèles de mesure de variables latentes : équivalence, invariance et identifiabilité

Mary Thompson, *University of Waterloo*

In surveys conducted for analytic purposes, social scientists often measure participant characteristics, experiences and attitudes using questionnaire items and combinations of these. If such a survey is conducted in more than one society, in more than one language, or with more than one data collection mode, the latent underlying variables take on increased conceptual importance. I will discuss with examples latent variable models put forward in comparative analyses, and associated calibrations of measurement, tests for cross-cultural invariance, and conditions for identifiability of parameters.

Dans les enquêtes construites à des fins analytiques, les spécialistes des sciences sociales mesurent souvent les caractéristiques, expériences et attitudes des participants au moyen d'items de questionnaires et de combinaisons de ces items. Si une telle enquête est conduite dans plus d'une société, dans plus d'une langue, ou avec plus d'une méthode de collecte de données, les variables latentes sous-jacentes prennent une importance conceptuelle accrue. Je discuterai, avec des exemples, de modèles de variables latentes mis en œuvre dans des analyses comparatives, des calibrations devant être apportée aux mesures, des tests pour l'invariance interculturelle, et des conditions nécessaires à l'identifiabilité des paramètres.

Session 19F Jeu. 29 mai • Thu. May 29, 14:35 – 16:05 Capital Hall 6A

Bayesian Methods
Méthodes bayésiennes

[MS-400]
Compressing Parameters in Bayesian High-order Models
Paramètres compressants dans les modèles bayesiens à ordre élevé

Longhai Li, *University of Saskatchewan*, Radford Neal, *University of Toronto*

Bayesian classification and regression with high order interactions is largely infeasible because Markov chain Monte Carlo (MCMC) would need to be applied with a great many parameters. In this talk we show how to make it feasible by effectively reducing the number of parameters, exploiting the fact that many interactions have the same values for all training cases. Our method uses a single "compressed" parameter to represent the sum of all parameters associated with a set of patterns that have the same value for all training cases. We show in detail how to compress parameters for logistic sequence prediction models.

La classification bayesienne et la régression avec des interactions à ordre élevé est pratiquement infaisable parce que la méthode de Monte-Carlo par chaînes de Markov (MCMC) devrait alors être appliquée avec un beaucoup de paramètres. Dans cette présentation, nous montrons une façon de rendre cette méthode faisable en réduisant efficacement le nombre de paramètres, et ce, en tirant partie du fait qu'un bon nombre d'interactions ont la même valeur pour tous les cas. Notre méthode utilise un paramètre « compressé » simple pour représenter la somme de tous les paramètres associés avec un ensemble de modèles qui ont la même valeur pour tous les cas. Nous montrons en détail la façon de compresser les paramètres pour les modèles de prédiction à séquence logistique.

[MS-401]

On the Behaviour of Bayesian Credible Intervals for some Restricted Parameter Space Problems: Recent Results

Sur le comportement de divers estimateurs bayésiens par intervalles pour des espaces paramétriques contraints : résultats récents

Éric Marchand, *Université de Sherbrooke*, William E. Strawderman, *Rutgers University*, Keven Bosa, *Statistique Canada*, Aziz Lmoudden, *Université de Sherbrooke*

To estimate a restricted parameter, Zhang and Woodroofe (2003), Roe and Woodroofe (2000), as well as Marchand and Strawderman (2006) investigate HPD credible sets associated with priors obtained as the truncations of non-informative priors onto the restricted parameter space. Namely, attractive frequentist properties of these procedures are seen to be applicable in many situations. Here, we present very recent results, in particular in the form of improved lower bounds for the frequentist coverage of these procedures.

Pour l'estimation d'un paramètre sous contraintes, des résultats récents de Zhang et Woodroofe (2003), Roe et Woodroofe (2000) et Marchand et Strawderman (2006) témoignent de belles et intéressantes propriétés fréquentistes d'intervalles de confiance bayesiens associés à des lois a priori obtenus comme projections sur l'espace paramétrique de lois a priori non informatives. Lors de cet exposé, nous présentons des résultats récents, notamment sous forme de meilleures bornes inférieures pour la probabilité fréquentiste de recouvrement de ces intervalles bayésiens.

[MS-402]

Bayesian Inference in the Multivariate Probit Model

Inférence bayésienne dans le modèle probit multivarié

Aline Tabet, Arnaud Doucet & Kevin Murphy, *University of British Columbia*

The multivariate Probit model is a popular class of models particularly suitable for the analysis of correlated binary data. Difficulties with this model are due to computation as the likelihood of the observed discrete data is obtained by integrating over a multidimensional constrained space of latent variables. We adopt a Bayesian approach and develop an an efficient Markov chain Monte Carlo algorithm for estimation in MVP models under the full correlation and the structured correlation assumptions. Our algorithm has many advantages over previous approaches, namely it handles identifiability and uses a marginally uniform prior on the correlation matrix directly.

Le modèle Probit multivarié est une classe populaire de modèles particulièrement adaptée à l'analyse de données binaires corrélées. Les principales difficultés liées à ce modèle sont dues au calcul, la vraisemblance des données discrètes observées étant obtenue en intégrant sur un espace multidimensionnel restreint de variables latentes. Nous adoptons une approche bayésienne et développons un algorithme efficace de chaînes de Markov Monte-Carlo pour l'estimation en présence de modèles PMV sous les hypothèses de corrélation complète et de corrélation structurée. Notre algorithme a plusieurs avantages par rapport aux approches précédentes, notamment son traitement de l'identifiabilité et son utilisation d'une loi *a priori* marginalement uniforme sur la matrice de corrélation.

[MS-403]

A Computation Method in Robust Bayesian Decision Theory

Une méthode de calcul d'après la théorie de la décision bayésienne robuste

Christophe Abraham, *Montpellier SupAgro*

According to Bayesian decision theory, an optimal decision is a decision which maximizes the expectation of a utility function. In practice, many utility functions can fit the preferences of the decision

D'après la théorie de la décision bayésienne, une décision optimale est une décision pour laquelle l'espérance de l'utilité est maximum. En pratique, plusieurs fonctions d'utilité peuvent s'ajuster aux préférences du décideur. On

maker. We provide a method for computing the extreme optimal decisions when the utility ranges over a class U. The infinite dimensional class U is made up of continuous functions with constraints on values and the shape. The method enables us to approximate the optimal decision for any u in U. Then, the extreme optimal decisions are computed by the simulated annealing algorithm. Theoretical results on the accuracy of the approximations and an application are provided.

propose une méthode pour calculer les décisions optimales extrêmes lorsque l'utilité parcourt une classe U. La classe U, de dimension infinie, est constituée des fonctions continues avec des contraintes de valeurs et de formes. La méthode proposée permet d'approcher la décision optimale pour tout u dans U. Les décisions optimales extrêmes sont calculées par l'algorithme du recuit simulé. Nous donnons des résultats théoriques sur les approximations ainsi qu'une application.

[MS-404]

Bayesian Inference for Auto-excited Point Processes Observed Partially
Inférence bayésienne pour des processus ponctuels auto-excités partiellement observés

Marylène Troupé & Jean Vaillant, *Université des Antilles et de la Guyane*, Evans Gouno, *Université Bretagne-Sud*

Let N be an n-dimensional time point process for which k components ($k < n$) are driving the $n - k$ other components. We consider the situation where a single realization of N over a time interval $[0, T]$ is observed in the following way: we have counting observations at fixed dates for the driven components whereas the driving components are observed continuously on the time interval $[0, T]$. Using Bayesian techniques, we develop tools for statistical inference about the effect parameter vector and the action duration parameter. As an illustration, Markovian, Hawkes, Poisson and General Yule driving processes are presented.

Soit N un processus ponctuel en dimension n, ayant k composantes dirigeant les $n - k$ autres ($k < n$). Nous considérons le cas où une unique réalisation du processus N est observée de la façon suivante : nous avons des données de comptage à des dates fixées pour les composantes dirigées, tandis que les composantes dirigeantes sont observées de façon continue dans l'intervalle $[0, T]$. Des outils d'inférence bayésienne sont développés, concernant le paramètre vectoriel d'effet, ainsi que le paramètre de durée d'action. Le cas de processus directeurs de Hawkes, Poisson et Yule sont présentés.

Session 19G Jeu. 29 mai • Thu. May 29, 14:35 – 16:05 Capital Hall 7A

Nonparametric Density Estimation
Estimation non paramétrique de la densité

[MS-405]

Asymptotic Unbiased Density Estimators
Estimateurs asymptotiquement sans biais de la densité

Eric Matzner-Lober, *IRMAR, Université Rennes 2*, Nicolas Hengartner, *Los Alamos National Laboratory*

This talk will present a computationally tractable density estimator that has the same asymptotic variance as the classical Nadaraya – Watson density estimator but whose asymptotic bias is zero. We achieve this result using a two stage estimator that applies a multiplicative bias correction to an oversmooth pilot estimator. Simulations show that our asymptotic results are available for samples as low

Cet exposé présentera un estimateur facile à mettre en œuvre admettant la même variance asymptotique que l'estimateur classique de Nadaraya-Watson mais dont le biais asymptotique est nul. Nous obtenons ce résultat en utilisant un estimateur à deux étapes : la première consiste à choisir un premier estimateur de la densité trop lisse donc biaisé, la seconde consiste alors à appliquer à cet estimateur pilote une correction multiplicative du biais. Les simulations montrent

as $n = 50$, where we see an improvement of as much as 20.

que les résultats asymptotiques de notre estimateur se retrouvent déjà pour des échantillons de taille $n = 50$. Nous pouvons obtenir des améliorations de l'ordre de 20.

[MS-406]
Recursive Bias Estimation and L_2 Boosting
Réduction itérative de biais et L_2 Boosting

Pierre-André Cornillon, *Montpellier SupAgro,* Nicolas Hengartner, *Los Alamos National Laboratory,* Eric Matzner-Lober, *IRMAR Université Rennes 2*

A general iterative bias correction procedure for regression smoothers is presented. This bias reduction schema is shown to correspond operationally with the L_2 Boosting algorithm and provides a new statistical interpretation for L_2 Boosting. We analyze the behavior of the Boosting algorithm applied to common smoothers S which we show depend on the spectrum of $I - S$. We present examples of a common smoother for which Boosting generates a divergent sequence. The statistical interpretation suggests combining the algorithm with an appropriate stopping rule for the iterative procedure. Finally we illustrate the practical performances via simulations.

Un schéma général de correction itérative de biais est présenté. Cette procédure est applicable pour les lisseurs S dans un modèle standard de régression. Cette procédure itérative de réduction de biais admet un lien évident avec le L_2 boosting ce qui permet une nouvelle interprétation statistique du L_2 boosting. Une analyse de la stabilité de la procédure est fournie avec des illustrations pour des choix classiques de lisseurs : noyaux, splines de lissage et k plus proches voisins. Enfin, des critères d'arrêt de cette procédure sont présentés et ainsi que leurs comportements sur des simulations.

[MS-407]
Asymptotic Normality and Test for Deconvolution of Mixed Measure
Normalité asymptotique et tests pour des estimations de déconvolution d'une mesure mixte

Mostafa Filali, *Université de Bourgogne,* Rachid Sabre, *Laboratoire MAIS, ENESAD*

The density estimate by deconvolution method has been widely studied, where the variable is observed with random errors. In this work, we consider the case where the probability measure is the sum of an absolutely continuous part with respect to the Lebesgue measure and a discrete measure. We combine deconvolution and double kernel methods for giving a density estimate of the continuous part and the amplitudes estimate of discrete part. The asymptotic normality has been studied and used for testing the existence and location of jump points.

L'estimation par la méthode de déconvolution de la densité de probabilité d'une variable aléatoire observée avec des erreurs aléatoires a été largement étudiée dans la littérature. Dans ce travail nous considérons le cas où la mesure de cette variable aléatoire possède une certaine mixture : la somme d'une mesure absolument continue et d'une mesure discrète. Nous combinons la méthode de déconvolution et celle du double noyau pour donner des estimateurs de la densité de la partie continue et les amplitudes aux points de sauts. La normalité asymptotique de ces estimateurs a été étudiée permettant de tester l'existence et localiser les points de sauts.

[MS-408]
Model Selection by Cross-validation
Sélection de modèle par validation-croisée

Alain Celisse & Stéphane Robin, *AgroParisTech/INRA MIA*

In the density estimation framework, we use a cross-validation method named leave-p-out (LPO) to estimate the quadratic risk of histograms. Unlike what usually happens, closed-form expressions are derived for this risk estimator and its moments. We define the "optimal" value of the parameter p (of cross-validation) as the minimizer of the Mean Square Error of the LPO estimator. Optimality for the procedure is expressed in terms of convergence of both the risk estimator and the optimal p. An oracle inequality and adaptivity are obtained for the resulting histogram. More insight is also given about the relationship between p and over-fitting.

En estimation de densité, nous utilisons une méthode de validation-croisée appelée leave-p-out (LPO), afin d'estimer le risque quadratique d'un histogramme. Dans ce cas particulier, l'estimateur du risque ainsi que ses moments admettent des expressions explicites qui nous permettent de définir la valeur « optimale » du paramètre p (de la validation-croisée) comme le minimiseur de l'erreur quadratique moyenne de l'estimateur LPO. L'optimalité de la procédure est établie en termes de convergence, à la fois pour l'estimateur du risque et pour le p optimum. Nous obtenons une inégalité oracle ainsi qu'une propriété d'adaptativité pour l'histogramme résultant. Nous tentons également de mieux comprendre la relation entre le choix de p et le surajustement.

[MS-409]
Bandwidth Selection for Nonparametric Confidence Intervals of a Probability Density Function
Sélection du paramètre de lissage pour les intervalles de confiance non paramétriques d'une fonction de densité

Shan Jiang, *Queen's University*, Yaohua Wu, *University of Science and Technology of China*, Dongsheng Tu, *Queen's University*

It was shown by Hall (1991) that undersmoothing with a bandwidth smaller than that minimizing mean square error is required to construct a correct confidence interval for a probability density function based on nonparametric kernel estimates. For a given dataset, determination of extent for undersmoothing is still a difficult problem. In this talk, we present a solution to this problem by deriving a formula of bandwidth selection based on the asymptotic expansion for the coverage probability of the undersmoothed conference interval. The performance of the nonparametric confidence intervals based on kernel estimates with proposed bandwidth is evaluated by Monte-Carlo simulations.

Il a été montré par Hall (1991) que sous-lisser avec un paramètre de lissage plus petit que celui minimisant l'erreur quadratique moyenne est nécessaire afin de construire un intervalle de confiance correct pour une fonction de densité basée sur des estimateurs à noyaux non paramétriques. Pour un jeu de données spécifique, déterminer l'étendue du sous-lissage est toujours un problème difficile. Dans cet exposé, nous présentons une solution à ce problème en obtenant une formule de sélection du paramètre de lissage basée sur le développement asymptotique pour la probabilité de couverture de l'intervalle de confiance sous-lissé. La performance des intervalles de confiance non paramétrique basée sur des estimateurs à noyaux avec le paramètre de lissage proposé est évaluée par des simulations de Monte-Carlo.

[MS-410]
Nonparametric Methods for Density Estimation in Multivariate Mixtures
Méthodes non paramétriques d'estimation de densités de mélanges multivariés

Reza Pakyari, *Arak University*

A new nonparametric estimation method for the component density functions of a multivariate mix-

Une nouvelle méthode d'estimation non paramétrique pour les fonctions de densités des composantes d'un modèle de

ture model is proposed. The estimation procedure is based on inverting the mixture model that is expressed in terms of densities. A Monte Carlo simulation study shows that the nonparametric estimator performs relatively better than its parametric competitor when estimating the most probable component and when the value of the mixing proportion is close to one.

mélange multivariée est proposée. La procédure d'estimation est basée sur l'inversion du modèle de mélange exprimé en termes de densités. Une étude de simulation Monte-Carlo montre que l'estimateur non paramétrique performe relativement mieux que son compétiteur paramétrique lors de l'estimation de la composante la plus probable et lorsque la valeur de la proportion de mélange est près de un.

Session 19H Jeu. 29 mai • Thu. May 29, 14:35 – 16:05 Capital Hall 8A

Survival Data Analysis
Analyse de données de survie

[MS-411]

Analysis of Recurrent Gap Time Data with Application to the Emergency Department Visit Rates among Violently Injured Pediatric Patients

Analyse de données de temps d'écart avec application aux taux de visite aux urgences chez les patients pédiatriques violemment blessés

Hyun Lim & Hyun Lim, *University of Saskatchewan*, Xu Zhang, *Georgia State University*

Sequentially ordered multivariate failure time data or recurrent event time data are commonly observed in biomedical longitudinal studies, and gap times in such data are often of interest. Generally, standard hazard regression methods cannot be applied to the gap times because of induced dependent censoring. In multivariate failure time data, the multiplicative and additive hazards models are examples of such model in a counting process setting. Using the Emergency Department Visit Study, we consider the additive hazards model for recurrent gap time data. The differences in estimates obtained by the multiplicative and additive hazards models are examined.

Des données multivariées de temps de défaillance séquentiellement ordonnées ou des données de temps d'événements récurrents sont souvent observées dans des études longitudinales biomédicales, et les temps d'écart dans de telles données sont souvent d'intérêt. Généralement, les méthodes usuelles de régression de défaillance ne peuvent être appliquées aux temps d'écart à cause de la censure dépendante induite. Avec des données de temps de défaillance multivariées, les modèles de défaillance additifs et multiplicatifs sont des exemples d'un tel modèle dans le contexte d'un processus de dénombrement. En utilisant une étude sur les visites aux urgences, nous considérons le modèle des risques additifs pour données de temps d'écarts récurrents. Les différences entre les estimateurs obtenus avec les modèles de défaillance additifs et multiplicatifs sont examinées.

[MS-412]

Parametric Frailty Approach in a Mixture Cure Model in Survival Analysis.

Modèles de mélanges paramétriques, avec fraction non à risque, en analyse de survie pour données groupées

Pierre Joly, *Inserm, U897*, Fabien Corbiere, *INRA, UMR1225*

Mixture cure models have been developed in survival analysis to take into account the possibility that some (cured) subjects may never undergo the event of interest. We generalize this approach to shared frailties model to analyze grouped data. A piecewise exponential proportional hazards model

Les modèles de survie de mélange permettent de prendre en compte la possibilité de ne pas être à risque de connaître l'évènement d'intérêt. Nous généralisons ces modèles à l'analyse de données groupées. La distribution des temps de survie des individus à risque est modélisée par un modèle exponentiel par morceaux à risque proportionnels et fragilités

with random effects is used to modelize the survival function of uncured subjects and a logistic regression model with random effects is used to modelize the cured fraction. Independent or correlated random effects are considered. The proposed approach is applied to a multi-flock study to investigate risk factors associated with classical scrapie.

partagées. L'appartenance au groupe à risque est modélisée par un modèle logistique à fragilités partagées. Différentes structures de variance-covariance des effets aléatoires sont envisagées. Cette approche est motivée par l'analyse des facteurs de risque liés à la tremblante classique du mouton dans plusieurs élevages.

[MS-413]

Genes Selection and Components Retention in Building a Supervised Survival Prediction Model
Sélection de gènes et rétention des composantes lors de la conception supervisée d'un modèle prédictif de survie

Keyue Ding & Zhangjun Ding, , *Queen's University*

Pretreatment gene expression profile of cancer possess the information about patient's disease prognosis and sensitivity to cancer therapy, and the information can be extracted and be used for tailoring individualized treatment. In building supervised survival prediction model, informative genes selection, and method of synthesizing information across the informative genes play a key role. We propose a method in determining genes included in model building, and rule in determining the number of components retention in the survival prediction model, and demonstrate the method in prediction of patient's survival and for treatment selection with a NCIC CTG cancer clinical trial.

Le profil d'expression génique avant le traitement du cancer contient de l'information sur le pronostic de la maladie et sur la sensibilité du patient au traitement anticancéreux. Cette information peut être extraite et utilisée pour individualiser le traitement. La sélection de gènes informatifs et la méthode de synthèse de l'information fournie par les divers gènes informatifs jouent un rôle dans la conception supervisée d'un modèle de prédiction de la survie. Nous proposons une méthode de détermination des gènes qu'il convient d'inclure dans la conception du modèle et une règle de détermination du nombre de composantes à retenir dans le modèle de prédiction de la survie, et nous illustrons la méthode de prédiction de la survie des patients et de choix du traitement dans le contexte d'un essai clinique réalisé par le Groupe des essais cliniques de l'Institut national du cancer du Canada (GEC INCC).

[MS-414]

Modeling Treatment Effects on the Recurrent Event Mean and Rate in the Presence of a Terminating Event
Modéliser les effets de traitement sur la moyenne et le taux d'évènements récurrents en présence d'un évènement final

Douglas Schaubel, *University of Michigan*

Often in biomedical studies, the event of interest is recurrent (e.g., hospitalization) and the recurrent event sequence is subject to being stopped by a terminating event (e.g., death). In comparing treatment options, a marginal method is preferred when interest is focused on the overall public health impact of a treatment. We propose semiparametric methods for use in observational studies to estimate the cumulative treatment effect on the marginal mean number of recurrent events. Large-sample properties are derived, with their applicability in finite samples assessed through simulation. The proposed methods are applied to kidney failure data.

Souvent, en études biomédicales, l'événement d'intérêt est récurrent (p.ex., l'hospitalisation) et la suite d'événements récurrents est sujette à un arrêt par un événement final (p.ex., la mort). En comparant des options de traitement, une méthode marginale est préférée lorsque l'intérêt est centré sur l'impact global sur la santé publique d'un traitement. Nous proposons des méthodes semi-paramétriques pour estimer l'effet cumulatif du traitement sur le nombre marginal d'événements récurrents dans des études observationnelles. Nous obtenons des propriétés asymptotiques et évaluons leur applicabilité en présence d'échantillons finis au moyen de simulations. Les méthodes proposées sont appliquées à des données d'insuffisance rénale.

[MS-415]
Comparison of Failure Probabilities for Local-regional Recurrence
Comparaison des probabilités de défaillance pour évaluer la récidive locale ou régionale

Sony Brar, *University of Calgary*, Penny Brasher, *University of British Columbia*

In the field of cancer, clinical studies are often interested in the long-term outcome, local-regional recurrence. For research purposes and treatment specification, it is important that this outcome be appropriately summarized. This presentation discusses the theoretical limitations in the traditional use of the Kaplan-Meier technique to estimate survival probabilities for specific events such as local-regional recurrence. More specifically, the emphasis will be to explain the methods that deal with cause-specific events by comparing probabilities of failure. The true probability of failure for a specific event can only be assessed by simultaneously considering the incidence of failure for all other causes.

En cancérologie, les études cliniques portent souvent sur le résultat à long terme, c'est-à-dire la récidive locale ou régionale. Pour les besoins de la recherche et de la spécification du traitement, il est important que ce résultat soit résumé de manière appropriée. La présente communication traite des limites théoriques de l'utilisation habituelle de la méthode de Kaplan-Meier pour estimer les probabilités de survie pour des événements particuliers, tels que la récidive locale ou régionale. Plus précisément, nous tâchons surtout d'expliquer les méthodes axées sur les événements par cause comparativement à des probabilités de défaillance. Dans le cas d'un événement particulier, la probabilité réelle de défaillance ne peut être évaluée qu'en considérant simultanément l'incidence de la défaillance pour toutes les autres causes.

Session 20A Jeu. 29 mai • Thu. May 29, 16:30 – 18:00 Capital Hall 1A

Teaching of Statistics: Didactical and Pedagogical Problems
L'enseignement de la statistique : problèmes didactiques et pédagogiques

[MS-416]
School Mathematics in Statistical Treatment and Teacher's Training
Les mathématiques scolaires dans les traitements statistiques et la formation des enseignants

Linda Gattuso, *Université du Québec à Montréal*

Statistics is a necessary part of the education of today's citizens and many think it should start gradually from primary school. However, statistical thinking is in many ways different from mathematical thinking. Attention to Statistics is seen to take away time teachers prefer to reserve for Mathematics. On the contrary, Statistics offers an ideal and concrete environment to illustrate and reinforce many mathematical concepts. Present and future teachers must be made aware of this fact. We will illustrate how this can be done at different school levels.

La statistique doit faire partie de la formation de tout citoyen et selon certains, son étude devrait commencer graduellement à partir de l'école primaire. Or, la pensée statistique se distingue de la pensée mathématique en plusieurs points. Pour qu'elle se développe, les enseignants doivent y consacrer du temps qu'ils préfèrent consacrer aux mathématiques. La statistique offre toutefois un environnement idéal et très concret pour illustrer et renforcer de nombreux concepts mathématiques. Les futurs enseignants et ceux en exercice doivent en être conscients. Nous illustrerons comment on peut le faire à divers niveaux d'enseignements.

[MS-417]

Mathematical Knowledge and Statistical Skills: Obstacles in Statistical Learning among Non-specialists
Connaissances mathématiques et Compétences statistiques : quelques obstacles à l'apprentissage de la statistique chez les non-spécialistes

Jean-Claude Régnier, *Groupe d'enseignement de la statistique (SFDS)*

Using a production analysis of 3rd and 4th year university students in Early or Continued Education Sciences in France (sample of 392 copies of the January 2008 exam), we intend to explain the difficulties and obstacles faced by students who are not specializing in statistics, and who do not have a strong science background, when learning about statistics and developing their statistical skills. By exploring the challenges encountered during training that encourages a statistical mindset, based on witness-accounts, we intend to put forth teaching solutions that are adapted to the large-group learning conditions within the university environment.

Par l'analyse de productions d'étudiants de 3e et 4e années d'Université en sciences de l'éducation en formation initiale ou continue en France (échantillon de 392 copies d'examen de janvier 2008), nous tenterons d'expliciter les difficultés et les obstacles auxquels ces étudiants non spécialistes de statistique et avec une formation scientifique peu élevée sont confrontés, dans leur apprentissage de la statistique et dans le développement de leurs compétences statistiques. En explicitant les défis rencontrés dans une formation visant le développement de l'esprit statistique, à partir de témoignages, nous attendons que ressortent des propositions pédagogiques adaptées aux conditions universitaires d'enseignement en grands groupes.

[MS-418]

Place of Simulation and Training of Inquiries in the Didactic of Statistics
Place de la simulation et de la pratique d'enquête dans la didactique de la statistique

Jean-Claude Oriol, *Université de Lyon II*

Our work highlights the fact that statistical teaching holds in itself a primitive and epistemological breach between data collecting and processing on the one hand, and, on the other hand, an approach based on the calculation of probabilities more or less remote from social practices. It seems therefore that questionnaire-surveys constitute a means of reducing this primitive gap, by allowing the subject to build a sense of learning. Furthermore, we have studied how the constructions of simulations performed by the students enables them to obtain the invariants of the variability present in phenomena whose study is at the core of statistical situations.

Notre travail expose que l'enseignement de la statistique porte en lui une césure à la fois primitive et épistémologique, entre d'une part la collecte et le traitement des données, et d'autre part une approche fondée sur le calcul des probabilités plus ou moins éloignée des pratiques sociales, et que la réalisation d'enquêtes est un moyen d'atténuer la trace de cet écart primitif, permettant au sujet de construire le sens de ses apprentissages. D'une façon complémentaire nous étudierons comment la construction de simulations par les apprenants leur permet de dégager des invariants de la variabilité présente dans toute situation statistique.

Session 20B Jeu. 29 mai • Thu. May 29, 16:30 – 18:00 Capital Hall 2A

Statistical Methods for Complex Survey Data
Méthodes statistiques pour données de sondages complexes

[MS-419]

Confidence Intervals for Proportion and Quantile Estimations with Focus on the US National Health and Nutrition Examination Survey

Intervalles de confiance pour les estimations de proportions et de quantiles en mettant l'accent sur la National Health and Nutrition Examination Survey américaine

Xin Feng & Randy Sitter, *Simon Fraser University*

It has been noted that the usual confidence interval for proportions does not perform well for extreme proportions. In surveys, the issue is complicated by the survey design and issues of whether to use design effects, effective sample size and effective degrees of freedom arise. The question is which of the many possible confidence intervals should be recommended for the U.S. National Health and Nutrition Examination Surveys users and what cautions should be given. In addition, the issues may be different if the interval is being used in combination with Woodruff's method to form confidence intervals for extreme quantiles.

Il a été souligné que les intervalles de confiance habituellement calculés pour les proportions ne donnent pas de bons résultats en cas de proportions extrêmes. Dans le contexte des enquêtes, le problème se complique du fait du plan de sondage et des questions se posent quant à l'utilisation des effets de plans, de la taille effective d'échantillon et des nombres effectifs de degrés de liberté. Il s'agit de savoir, parmi les nombreux intervalles de confiance possible, lequel il convient de recommander aux utilisateurs des données de la National Health and Nutrition Examination Surveys américaine et quels avertissements devraient être formulés. En outre, les problèmes pourraient être différents si l'intervalle est utilisé en combinaison avec la méthode de Woodruff en vue de produire des intervalles de confiance pour les quantiles extrêmes.

[MS-420]

Analysis of Incomplete Longitudinal Binary Data Subject to Missing at Random with Application to the SLID Data

Analyse de données binaires longitudinales incomplètes sous un mécanisme de données manquant au hasard avec application aux données de l'EDTR

Taslim Mallick & Brajendra Sutradhar, *Memorial University*, Milorad Kovacevic, *Statistics Canada*

In a longitudinal study, responses of an individual may be missing in some follow-ups. These missing responses can be completely random (MCAR) or random depending on individual's history (MAR). The purpose of this study is to analyse SLID data with binary response variable assuming MAR type missing responses. This analysis is carried out through a Weighted Generalized Quasilikelihood (WGQL) approach and the results are compared with the existing Weighted Generalized Estimating Equation (WGEE) approach mainly under MAR assumption. A simulation study based on SLID data is also performed to examine the effect of sampling on the estimation methods.

Dans une étude longitudinale, il arrive que les réponses d'une personne manquent pour certains suivis. Ces réponses peuvent manquer entièrement au hasard (MCAR pour missing completely at random) ou au hasard (MAR pour missing at random) selon les antécédents de la personne. Le but de la présente étude est d'analyser les données de l'EDTR avec une variable de réponse binaire en supposant que les données manquent selon un mécanisme MAR. Cette analyse est exécutée en suivant une approche de quasi-vraisemblance généralisée pondérée (QVGP) et les résultats sont comparés à ceux de l'approche existante des équations d'estimation généralisées pondérées (EEGP), principalement sous l'hypothèse MAR. Nous réalisons aussi une étude par simulation fondée sur des données de l'EDTR

pour examiner l'effet de l'échantillonnage sur les méthodes d'estimation.

[MS-421]
Hierarchical Bayes Small Area Estimation for the Canadian Community Health Survey
Estimation bayésienne hiérarchique de petits domaines pour l'Enquête sur la santé dans les collectivités canadiennes

Qian Zhou, *University of Waterloo*, Yong You, *Statistics Canada*

We propose four full hierarchical Bayes (HB) models which relax the two strong assumptions of Fay-Herriot area-level models by constructing Gaussian conditional autoregressive (CAR) models on the area-specific effects to induce spatial correlation, and/or assuming the sampling variances unknown. Through analyzing the survey data from Canadian Community Health Survey, we make comparison among the HB model-based estimates and direct design-based estimates for the rate of asthma for the 20 health regions in BC province. Our results have shown that the model-based estimates perform better than the direct estimates. Moreover, the proposed CAR models have smaller CVs than the Fay-Herriot model.

Nous proposons quatre modèles bayésiens complètement hiérarchiques (BH) qui relâchent deux hypothèses fortes des modèles de domaines de Fay-Herriot en construisant des modèles gaussiens conditionnels autorégressifs (CAR) sur les effets spécifiques aux domaines pour induire une corrélation spatiale, ou en supposant les variances échantillonnales inconnues. En analysant les données de l'Enquête sur la santé dans les collectivités canadiennes, nous comparons les estimés du taux d'asthme pour les 20 régions le la Colombie-Britannique basés sur le modèle BH et ceux basés directement sur le plan d'expérience. Nos résultats montrent que les estimés basés sur le modèle performent mieux que les estimés directs. De plus. les modèles CAR proposés ont des variances conditionnelles plus petites que le modèle Fay-Herriot.

Session 20C Jeu. 29 mai • Thu. May 29, 16:30 – 18:00 Capital Hall 3A

Inference and Applications
Inférence et applications

[MS-422]
Measuring the Mortality Risk Based on the Lee-Carter Model
Mesurer le risque de mortalité à partir du modèle Lee-Carter

Bifeng Xie, *University of Western Ontario*

The Lee-Carter model is a well-accepted model in mortality forecasting because of its capability of providing interval forecasting for mortality rates. Residual-based semiparametric bootstrap methods have been proposed by Koissi et al. (2005) and Brouhns et al. (2002), for example, in order to quantify the uncertainty of quantities of actuarial interest (such as life- expectancy, net single premium of a life annuity, and so on). In this study, we assess these bootstrap methods and compare their accuracy with a block bootstrap.

Le modèle de Lee-Carter est un modèle bien accepté en prévision de la mortalité grâce à sa capacité de fournir des intervalles de prévision pour les taux de mortalité. Des méthodes de bootstrap semi-paramétrique basées sur les résidus ont été proposées dans Koissi *et al*. (2005) et Brouhns *et al*. (2002) pour, par exemple, quantifier l'incertitude de quantités d'intérêt actuariel (telles l'espérance de vie, la prime unique nette d'une rente viagère, etc.). Dans cette étude, nous évaluons ces méthodes de bootstrap et comparons leur précision avec le bootstrap par blocs.

[MS-423]
Several Approaches to Warranty Data
Plusieurs approches à l'égard des données sur la garantie

Jaiwook Baik, *Korea National Open UnivERSITY*

Warranty data can be modeled using two different approaches: two-dimensional point process and one-dimensional point process with usage as a function of age. The first approach has three different models: bivariate model, marked point process and 'counting and matching model on both directions independently. But the models or variations of them need more investigation to be applicable to real warranty claim data.

Les données sur les garanties peuvent être modélisées suivant deux approches, à savoir un processus ponctuel bidimensionnel et un processus ponctuel unidimensionnel, en utilisant l'usage comme fonction de l'âge. La première approche comprend trois modèles, à savoir le modèle bivarié, le processus ponctuel marqué et le modèle de comptage et d'appariement indépendamment dans les deux directions. Toutefois, les modèles ou les variantes de ceux-ci devront être étudiés plus en détail pour pouvoir être appliqués aux données réelles sur les réclamations au titre de la garantie.

[MS-424]
Discovering Frequent and Rare Patterns in a Nucleotide Sequence
Découverte de structures fréquentes et rares dans une suite de nucléotides

Wan-chen Lee, *University of Manitoba*

Kleffe and Borodovsky(1992) provided an approximation for first and second moment of counts of words. Fu and Koutras(1994) addressed FMCI for the distribution theory of runs. We modify the approximation and give a new idea and procedure. Applied these two methods in a sequence to get approximate and exact expectation and variance of counts of all triplet patterns. Via standardizing, transformed observed frequents to z-score and classified frequent and rare words by setting up a threshold. Given the same criterion, we show the mis-classification rate by comparing results from approximate method with from exact distribution.

Kleffe et Borodovsky (1992) ont donné une approximation pour le premier et le second moment de dénombrements de mots. Fu et Koutras (1994) ont discuté de l'ICMF dans le contexte de la théorie de la distribution de séquences homogènes. Nous modifions l'approximation et présentons une nouvelle idée et procédure. Nous appliquons ces deux méthodes à une suite pour obtenir des valeurs exactes et approximées de l'espérance et de la variance du nombre de tous triplets possibles. Nous standardisons les fréquences observées en scores-z et catégorisons les mots comme étant "rares" ou "fréquents" en fixant un seuil. À partir du même critère, nous trouvons le taux de mauvaise classification en comparant les résultats obtenus au moyen de l'approximation et de la méthode exacte.

[MS-425]
Risk Assessment of Blood Lead Level from Lead Contamination in the Upper Mae-Klong Watershed, Kanchanaburi Province, Thailand
Évaluation des risques liés à la concentration de plomb dans le sang résultant de la contamination par le plomb dans le bassin supérieur du Mae-Klong (province de Kanchanaburi, Thaïlande)

Suwannee Adsavakulchai, *University of the Thai Chamber of Commerce*

The 84 of blood samples comprise 57 children (68% with age of 0 – 15 years and 27 adults (32% with age of > 15 years. In children, the samples with Pb blood level < 25μg/dL make up 35.1

Les 84 échantillons de sang comprennent 57 enfants (68%) âgés de 0 à 15 ans, et 27 adultes (32%), âgés de plus de 15 ans. Chez les enfants, les échantillons dont le taux de plombémie est inférieur à 25 μg/dL représentent 35,1% des échantillons, alors que ceux de taux supérieur à 25 μg/dL re-

présentent 44%. Chez les adultes, 96.3% des échantillons ont un taux de plombémie inférieur à 40 μg/dL. Nous concluons que les activités humaines sont un facteur important dans la contamination au plomb dans la région du bassin supérieur du Mae-Klong.

Session 20D Jeu. 29 mai • Thu. May 29, 16:30 – 18:00 Capital Hall 4A

Hypothesis Tests
Tests d'hypothèses

[MS-426]
Student's *z*, *t*, and *s*: What if Gosset had *R*?
Les statistiques *z*, *t*, et *s* de Student : et si Gosset avait *R* ?

James Hanley, *McGill University / Université McGill*, Marilyse Julien & Erica Moodie, *McGill University*

This year marks the 100th anniversary of Student's *The Probable Error of a Mean*. Despite its seminal nature, textbooks give Gosset, and the article, short shrift. Few are aware of the z statistic whose sampling distribution he actually derived, the derivation, his simulations to check his work, his *z* table, Fisher's derivation, or the switch from *z* to *t*. We present some of these aspects, and rework his simulations. We hope that the next generation of statisticians comes to know more about the man and his work, and appreciates that not all statistical distributions are derived in a single pass.

L'année 2008 marque le 100^{e} anniversaire de la publication de « The Probable Error of a Mean » par William Sealy Gosset (Student). Malgré l'avancement généré par ses réalisations, les livres de statistiques d'aujourd'hui lui offrent peu de reconnaissance. Ainsi, peu d'entre nous connaissent la statistique *z* dont il a dérivé la distribution, ses calculs et ses simulations, ainsi que la preuve incomplète terminée par Fisher – un des nombreux statisticiens avec lesquels il a collaboré – et qui le mena à la célèbre statistique *t*. Nous présentons certains de ces aspects et reprenons ses simulations en utilisant la technologie du 21e siècle.

[MS-427]
On Direction Dependence
De la dépendance de la direction

Yadolah Dodge, *University of Neuchâtel*, Iraj Yadegari, *Tarbiate Modares Universtiy*

Under the hypothesis of the existence of a linear relationship between two random variables, a series of new formulas are presented in this article. One of these formulas, the relationship between the correlation coefficient and the kurtosis will be used to determine the direction of the dependent variable in linear relationship.

Sous l'hypothèse de l'existence d'une relation linéaire entre deux variables aléatoires, une série de nouvelles formules est présentée dans cet article. Une de ces formules décrit la relation entre le coefficient de corrélation et la kurtose et sera utilisée pour déterminer la direction de la variable dépendante dans la relation linéaire.

[MS-428]

Hypothesis Tests for the Concurrent Assessment for Interrater Agreement and Intrarater Reliability
Tests d'hypothèses pour l'évaluation simultané de l'accord inter-juges et de la fiabilité inter-juges

Joseph Amuah, Michael Eliasziw & Gordon Fick, *University of Calgary*

For continuous measurements, agreement is often estimated by the intraclass correlation coefficient. Simultaneous estimation of inter-rater agreement and intra-rater reliability had been developed previously using a Method of Moments approach but hypothesis tests using this approach did not yield valid results.

This research aims were to derive valid testing procedures and evaluate their properties using Monte Carlo simulation by reporting empirical Type I error rates. Interval estimation is also evaluated based on interval widths and coverage rates.

The results indicate that the newly-derived methods yield valid Type I error rates and coverage levels.

Pour des mesures continues, l'accord est souvent estimé par le coefficient de corrélation intraclasse. L'estimation simultanée de l'accord inter-juges et de la fiabilité intra-juges a été développée précédemment au moyen de l'approche de la méthode des moments, mais les tests d'hypothèse utilisant cette approche n'ont pas donné de résultats valides.

Le but de cette recherche était d'obtenir des procédures de test valides et d'évaluer leurs propriétés par une simulation de Monte-Carlo en comparant les taux d'erreur de type I empiriques.

Les résultats indiquent que la nouvelle méthode donne des taux d'erreur de type I et des niveaux de couverture valides.

[MS-429]

A Homogeneity Measurement For Functional Data Clusters
Mesure d'homogénéité de classes issues d'un ensemble de données fonctionnelles

Bernard Goldfarb & Mireille Gettler-Summa, *Université Paris-Dauphine, Ceremade CNRS*

We propose a homogeneity measure for clusters in order to compare, in a functional data framework, two sets of temporal series according to their consistencies. A sequential procedure allowing an interactive software application relies on the change-point detection for a time series and on the tracking of the sequences shared by all the series, in order to calculate, after allowable transformations, the Kullback – Leibler divergences (KLD) between every couple of curves. A V-test deduced from the significance of the KLD associated test is applied. The number of significant V-tests, at a fixed level, measures the cluster heterogeneity.

Pour comparer, dans un cadre de données fonctionnelles, deux classes de séries temporelles par rapport à leur consistance, nous proposons une mesure d'homogénéité de classes. Un schéma séquentiel permettant une implémentation logicielle interactive associe la détection des points de rupture d'une chronique et le repérage de séquences communes aux séries pour calculer, après les transformations nécessaires, les divergences de Kullback-Leibler (DKL) des couples de courbes intraclasses. On applique un V-test déduit de la significativité du test associé à la DKL. Le nombre de V-tests significatifs pour un seuil fixé mesure l'hétérogénéité d'une classe.

[MS-430]

A Test for Dominance
Un test de dominance

Atsu Dorvlo, *Sultan Qaboos University, Sultanate of Oman*

In some selection situations, a simple majority is not enough to indicate that one item dominates the others. In this paper we develop a simple test procedure for testing the dominance of one item with

Dans certaine situations données, la majorité simple n'est pas suffisante pour montrer qu'un élément domine les autres. Dans cet article, nous développons un test simple pour tester la dominance d'un élément par rapport aux autres dans une

respect to others in a multi option situation. Expressions are given for computing the relevant p-values. The procedure is illustrated.

situation à options multiples. Les expressions sont données pour calculer les valeurs p appropriées.

Session 20E Jeu. 29 mai • Thu. May 29, 16:30 – 18:00 Capital Hall 5A

Bayesian Methods in Biostatistics
Méthodes bayésiennes en biostatistique

[MS-431]

Comparing traditional regression and inverse-probability-weighting for estimating causal effects: a Bayes risk approach

Comparaison par risque bayésien des approches de régression traditionnelle et de pondération pour l'estimation d'effets causals

Geneviève Lefebvre & Paul Gustafson, *University of British Columbia*

Estimating causal treatment effects with observational data requires adjustment for confounding at the analysis stage. In this talk, we present the results of a simulation study that compares the performances of the usual regression estimator and two inverse-probability-weighting (IPW) estimators for estimating treatment effects. We adopted a decision-theoretic viewpoint and compared the performances of the estimators across the parameter space for different scenarios of model misspecification. We found that for small-to-moderate sample sizes, the regression estimator compares appreciatively to the IPW estimators, including the so-called "doubly-robust" IPW estimator.

L'estimation de l'effet causal d'un traitement à partir de données d'observation requiert un ajustement pour les facteurs confondants au stade de l'analyse. Dans cette présentation, nous discuterons des résultats d'une étude de simulation visant à comparer l'approche de régression traditionnelle à deux approches utilisant une technique de pondération (IPW), pour l'estimation de tels effets. Dans ce travail, les performances des estimateurs sont comparées à travers l'espace des paramètres et pour plusieurs scénarios de mauvaises spécifications de modèles. Nos résultats indiquent que pour des tailles d'échantillon de petits à modérés, l'estimateur de régression est supérieur aux estimateurs pondérés, comprenant l'estimateur pondéré « doublement robuste ».

[MS-432]

MCMC Model Determination for Bayesian Graphical Models

Détermination de modèle MCMC pour modèles graphiques bayésiens

Sohee Kang, Laurent Briollais & Liu Jinnan, *Samuel Lunenfeld Research Institue*

For higher dimensional contingency table, the number of possible models increases exponentially. Determining the best final graph in such a context is very challenging. We present here two methodologies for Bayesian model determination in discrete decomposable graphical models: MCMC (Madigan & York, 1993) and Reversible Jump MCMC (Green, 1995). Through simulations of theoretical graphs with certain significant associations between edges, we assessed the performance of these two methods and studied their sensitivity to the hyper-parameter specification. We finally show

Pour un tableau de contingence de plus haute dimension, le nombre de modèles possibles augmente exponentiellement. Déterminer le meilleur graphe final dans un tel contexte est très difficile. Nous présentons ici deux méthodologies pour la détermination de modèles bayésienne dans le contexte de modèles graphiques décomposables : MCMC (Madigan & York, 1993) et MCMC à sauts réversibles (Green, 1995). En simulant des graphes théoriques avec certaines associations significatives entre les arêtes, nous avons évalué la performance de ces deux méthodes et étudié leur sensibilité à la spécification des hyper paramètres. Nous montrons finalement une application de ces deux méthodes à une étude sur

an application of these methods to a breast cancer study investigating the association with multiple SNPs in the DNA repair pathway.

le cancer du sein se penchant sur l'association avec de multiples SNP sur la voie de réparation de l'ADN.

[MS-433]

Bayesian Variable Selection for Logistic Regression with Misclassification
Sélection de la variables bayesienne pour la régression logistique avec erreurs de classement

Stephanie Powers & James Stamey, *Baylor University*

Response misclassification is a problem in many studies with binary outcomes. In a logistic regression where the binary response is subject to misclassification, we investigate the impact ignoring misclassification has on determining a subset of significant covariates. We then develop a Bayesian approach to select significant covariates while accounting for the misclassification. The procedure is applied to a real data example of a smoking cessation study and extensions to potential pharmaceutical problems are discussed.

La classification incorrecte des réponses est un problème que posent de nombreuses études dont les résultats sont binaires. Dans une régression logistique où la réponse binaire est sujette à une erreur de classification, nous étudions l'effet de la non-prise en compte de l'erreur de classification sur la détermination d'un sous-ensemble de covariables significatives. Puis, nous élaborons une méthode bayésienne de sélection des covariables significatives qui tient compte de l'erreur de classification. À titre d'exemple, nous appliquons la procédure à des données réelles provenant d'une étude sur le renoncement au tabac et discutons de son extension à des problèmes pharmaceutiques éventuels.

[MS-434]

Bayesian Analysis of Longitudinal Binary Data Subject to Misclassification
Analyse bayesienne des données binaires longitudinales sujettes aux erreurs de classement

Lihui Zhao, *Simon Fraser University*, Allan Becker, *University of Manitoba*, Moira Chan-Yeung, *University of British Columbia*

In an asthma prevention study to evaluate the effectiveness of a multifaceted intervention program, high-risk infants were randomized and tested for asthma at 1, 2 and 7 years of age. The diagnosis procedure was known to be imperfect but the exact misclassification rate was unknown. We present a Bayesian method to model the true longitudinal response by generalized linear model, and incorporate the unknown misclassification rate through prior distribution. Markov Chain Monte Carlo methods are applied to draw inference for the parameters of treatment effect as well as misclassification rate. The asthma prevention study is used as an illustration.

Dans une étude de la prévention de l'asthme conçue pour évaluer l'efficacité d'un programme d'interventions multiples, des nourrissons à haut risque ont été randomisés et soumis à un examen diagnostic de l'asthme à l'âge d'un an, de deux ans et de sept ans. On savait que la méthode diagnostique était imparfaite, mais on ignorait le taux exact d'erreurs de classification. Nous présentons une méthode bayésienne de modélisation de la réponse longitudinale réelle en utilisant un modèle linéaire généralisé et en intégrant le taux inconnu d'erreurs de classification au moyen d'une loi a priori. Nous appliquons deux méthodes de Monte Carlo par chaîne de Markov pour inférer les paramètres d'effet du traitement, ainsi que de taux d'erreurs de classification. L'étude de prévention de l'asthme est utilisée à titre d'exemple.

[MS-435]
Statistical Modelling for Epidemiology of Alder Decline
Modélisation statistique pour l'épidémiologie du dépérissement des aulnes glutineux

Fabrice Elegbede, Benoît Marçais & Jean-Claude Pierrat, *Institut Nationale de le Recherche Agronomique / ENGREF*

To study the impact of a new lethal disease on alders, a compartmental model was proposed on the basis of previous studies conducted over a survey area with 3,000 charted trees set up along the Sarre River (Tributary of Moselle River in northeastern France). Parameters estimation in this model was done using a statistical Bayesian method based on a type of Markov Chain Monte Carlo (MCMC) algorithms. Also, this study allows us to check the existence of a density threshold to monitor the epidemic. Furthermore, along with a description of point processes, estimation of infection spatial risk was performed, among other things.

Pour étudier l'impact d'une maladie létale sur la démographie des aulnes, un modèle à compartiments a été proposé sur la base d'études préalables menées sur un dispositif d'environ 3 000 arbres recensés et cartographiés le long de la Sarre (affluent de la Moselle). L'estimation des paramètres du modèle a été réalisée par utilisation de méthodes bayésiennes associées à des algorithmes de type MCMC. Aussi, cette étude a permis de vérifier l'existence d'un éventuel seuil de densité en dessous duquel l'épidémie pourrait être contrôlée. Par ailleurs, à l'aide de processus ponctuels, une estimation spatialisée du risque d'infection a été entre autres réalisée.

[MS-436]
Quasi-empirical Bayes Methodology for Improving Meta-analysis
Méthodologie de Bayes quasi empirique pour améliorer la méta-analyse

Hyang Mi Kim, *University of Calgary*, A.K.Md.Ehsanes Saleh, *Carleton University*, K.M. Hassanein & R.S. Hassanein, *KUMC, Kansas*

This article addresses the problem of heterogeneity among various studies to be combined in a meta-analysis. We adopt quasi-empirical Bayes methodology to predict the odds ratios for each study. As a result, the predicted odds ratios are pulled toward the estimated common odds ratio of the various studies under consideration. With strong heterogeneity among the studies, we jointly consider the display of the 95 CIs of the ORs and a Dixon's test (1950) for outliers to exclude the extreme estimated OR. We demonstrate the effectiveness of our methodology based on the data analyzed by Thompson and Pocok (1987).

La communication porte sur le problème de l'hétérogénéité de diverses études qui doivent être combinées dans une méta-analyse. Nous adoptons une méthode bayésienne quasi empirique pour prédire les rapports de cotes pour chaque étude. Par conséquent, les rapports de cotes prévus sont rapprochés du rapport de cotes estimé des diverses études examinées. Sous une forte hétérogénéité entre les études, nous considérons conjointement l'affichage des intervalles de confiance à 95% des rapports de cotes et un test de Dixon (1950) pour les valeurs aberrantes afin d'exclure les rapports de cotes estimés extrêmes. Nous démontrons l'efficacité de notre méthode en nous fondant sur les données analysées par Thompson et Pocok (1987).

Session 20F Jeu. 29 mai • Thu. May 29, 16:30 – 18:00 Capital Hall 6A

Estimation from Biomedical Data
Estimation fondées sur des données biomédicales

[MS-437]

Estimating the Rotation Axes of the Ankle joint: A Directional Model Using Non-Linear Mixed-Effects Theory

Estimation des axes de rotation de la cheville : un modèle directionnel utilisant la théorie des modèles non-linéaires à effets mixtes

Mohammed Haddou & Louis-Paul Rivest, *Université Laval*

We consider the problem of estimating the directions of the two rotation axes of the human ankle. The geometry of an ankle carries useful clinical information. The data set for analysis is a time ordered sequence of 3×3 rotation matrices giving the ankle's orientations as the foot moves with respect to the lower leg. These rotation matrices are assumed to follow von Mises – Fischer distributions. The rotation axes are first estimated by fitting a "fixed-effects" directional model to the sequence of 3x3 rotation matrices. The ankle movement is, usually, not wide enough to allow an effective estimation of all parameters involved. A second model using random effects is then adjusted to the data. The talk will focus on some statistical issues regarding the estimation of the two axes. Data analyses will illustrate the main findings.

On désire estimer les axes de rotation de la cheville humaine. La géométrie d'une cheville apporte de l'information clinique utile. L'ensemble de données d'analyse se compose d'une suite ordonnée dans le temps de matrices de rotation 3 par 3 donnant les orientations de la cheville relativement au bas du pied. On suppose que ces matrices de rotation suivent les distributions Von Mises-Fischer. Les axes de rotations sont d'abord estimés en ajustant un modèle directionnel à effets fixes à la suite des matrices de rotation 3 par 3. Le mouvement de la cheville n'étant pas assez ample pour permettre une estimation efficace de tous les paramètres. Un second modèle, utilisant des effets aléatoires, est alors proposé. Cet exposé mettra l'accent sur certains problèmes entourant l'estimation des deux axes de rotation. Des analyses des données illustreront les principales découvertes.

[MS-438]

Reducing Selection Bias: Efficiency and Robustness of Parametric & Non-parametric Approaches

Réduction du biais de sélection : efficacité et robustesse des approches paramétrique et non paramétrique

Laura Faye, *University of Toronto,* Shelley Bull, *University of Toronto; Samuel Lunenfeld Research Institute,* Lei Sun, *University of Toronto; Hospital for Sick Children Research Institute*

Model selection bias is prevalent in genome-wide genetic studies because the same sample is often used for both gene discovery and gene-effect estimation. The resulting upward-biased effect estimates lead to optimistic study design with inadequate sample size and deceptively high power. We investigate the efficiency and robustness of the resampling-based method of Sun and Bull (2005) and the likelihood-based approach of Zollner and Pritchard (2007) through simulation. While both methods substantially reduce the estimation bias, the bootstrap and likelihood methods tend to under- and over-correct respectively. Corresponding con-

Le biais de sélection du modèle est répandu dans les études génétiques à l'échelle du génome, parce que le même échantillon est souvent utilisé pour la découverte des gènes ainsi que pour l'estimation de leurs effets. Les estimations des effets entachées d'un biais par excès résultantes donnent lieu à des plans d'études optimistes dont la taille d'échantillon est inadéquate et la puissance, trompeusement élevée. Nous étudions par simulation l'efficacité et la robustesse de la méthode fondée sur le rééchantillonnage de Sun et Bull (2005) et de l'approche fondée sur la vraisemblance de Zollner et Pritchard (2007). Bien que ces deux méthodes réduisent considérablement le biais d'estimation, celle du bootstrap a tendance à le sous-corriger et celle de la vraisemblance, à

fidence intervals for both methods are typically wider than nominal.

le surcorriger. Les intervalles de confiance correspondants pour les deux méthodes sont généralement plus grands que l'intervalle nominal.

[MS-439]

Recovering True Probabilities for Nucleotide Trimming Processes in T-cell Receptor TRA and TRG Junction Formation with IMGT Tools

Dérivation des vraies distributions de probabilité pour le processus de grignotage aléatoire des récepteurs T à l'aide d'outils automatiques (d'IMGT)

Kevin Bleakley & Marie-Paule Lefranc, *Université Montpellier 2*, Gérard Biau, *Université Pierre et Marie Curie - Paris VI*

Immunoglobulins and T-cell receptors are the backbones of our immune system. They are created in billions of different forms by an amazing series of poorly understood random physical processes. Given a large data set of T-cell receptor DNA sequences, we use automatic tools (from IMGT) to output a (biased) empirical probability distribution for several of these physical processes. We then show how to remove the bias, expose good estimates of the true probability distribution and then predict the true nature of one of these physical processes: the random trimming process.

Les immunoglobulines et les récepteurs T représentent une partie fondamentale de notre système immunitaire. Ils sont créés en milliards de formes différentes par une série remarquable de processus physiques encore mal compris. Pour un grand ensemble donné de séquences ADN des récepteurs T, on utilise des outils automatiques (d'IMGT) pour calculer des distributions de probabilité empiriques (biaisées) pour plusieurs de ces processus physiques. Ensuite, on montre comment enlever le biais, extraire des distributions de probabilité améliorées et prédire la vraie nature d'un des processus physiques : celui de grignotage aléatoire.

[MS-440]

Semiparametric Estimation for the Dispersion Parameter in the Analysis of Over or Under Dispersed Count Data

Estimation semi-paramétrique pour le paramètre de dispersion dans l'analyse des données chiffrées surdispersées ou sous dispersées

Krishna Saha, *Central Connecticut State University*

Several semiparametric estimators of the dispersion parameter in the analysis of over or under dispersed count data when there is no likelihood available are investigated. We consider the double extended quasi-likelihood (DEQL), the pseudo-likelihood, and the optimal quadratic estimating (OQE) equations method, and compare them with the ML method, the method of moments, and the extended quasi-likelihood through simulation study. Simulations show evidence that the estimators based on the DEQL and OQE have superior bias and efficiency property for moderate and large sample size, and for small sample size, respectively. Three real life datasets arising in biostatistical practices are analyzed.

Plusieurs estimateurs semi-paramétriques des paramètres de dispersion dans l'analyse des données chiffrées surdispersées ou sous dispersées lorsqu'il n'y a pas de vraisemblance disponibles sont examinés. Nous considérons la quasi-vraisemblance double étendue, la pseudo-vraisemblance et les équations d'estimation quadratique optimale et les comparons avec la méthode du maximum de vraisemblance, la méthode des moments et la quasi-vraisemblance étendue par l'entremise d'une étude de simulation. Les simulations prouvent que les estimateurs fondés sur la double quasi-vraisemblance étendue et les estimations quadratiques optimales ont des biais et des propriétés de rendement supérieurs pour les tailles d'échantillon grande et modérée ainsi que pour une petite taille d'échantillon respectivement. Trois ensembles de données inhérentes aux pratiques en biostatistique sont analysés.

[MS-441]

Estimation Method of the Semiparametric Mixture Cure Gamma Frailty Model
Méthode d'estimation du modèle de mélange semi-paramétrique gamma de fragilité de la guérison

Jiajia Zhang, *University of South Carolina*, Yingwei Peng, *Queen's University*

Mixture cure frailty model has been proposed to analyze censored survival data with a cured fraction and unobservable information among the uncured patients. In this paper, we extend the mixture cure frailty model by incorporating covariates into both the cure rate and the latency distribution, and propose semiparametric estimation methods based on the EM algorithm and the multiple imputation method. We show that both estimation methods work well and apply them to a data set of failure times from bone marrow transplant patients.

Un modèle de mélange de fragilité avec guérison a été proposé pour analyser des données de survie censurés avec une fraction guérie et de l'information non observable chez les patients non guéris. Dans cet article, nous étendons le modèle de mélange de fragilité avec guérison en incorporant des covariables au taux de traitement et à la distribution de latence, et proposons des méthodes d'estimation semi paramétriques basées sur l'algorithme EM et la méthode d'imputation multiple. Nous montrons que les deux méthodes d'estimation fonctionnent bien et les appliquons à un jeu de données de temps de défaillance de patients ayant reçu une transplantation de moelle osseuse.

[MS-442]

On the Inadmissibility of Watterson's Estimator
À propos de l'inadmissibilité de l'estimateur de Watterson

Andreas Futschik & Florian Gach, *University of Vienna*

We consider the estimation of the scaled mutation parameter, which is one of the parameters of key interest in population genetics. We provide a general result showing when such estimators can be improved using shrinkage when taking the mean squared error as the measure of performance. As a consequence, we show that Watterson's estimator is inadmissible, and propose an alternative shrinkage-based estimator that is easy to calculate and has a smaller mean squared error. We then derive improved versions for other estimators, including the MLE. Simulations provide information about the amount of improvement achieved by our alternative estimators.

Nous considérons l'estimation du paramètre de mutation mis à l'échelle (scaled mutation parameter), qui est l'un des principaux paramètres d'intérêt en génétique des populations. Nous donnons un résultat général qui montre quand ce genre d'estimateur peut être amélioré par rétrécissement (shrinkage) en prenant l'erreur quadratique moyenne comme mesure de performance. Par conséquent, nous montrons que l'estimateur de Watterson est inadmissible et proposons un estimateur de rechange fondé sur le rétrécissement qui est facile à calculer et possède une plus petite erreur quadratique moyenne. Puis, nous dérivons des versions améliorées pour d'autres estimateurs, y compris l'estimateur du maximum de vraisemblance. Des simulations fournissent des renseignements sur l'importance de l'amélioration réalisée à l'aide de nos estimateurs de rechange.

Session 20G Jeu. 29 mai • Thu. May 29, 16:30 – 18:00 Capital Hall 7A

Estimation and Tests of Serial Dependence
Estimation et tests de dépendance sérielle

[MS-443]
Estimation of Autocorrelation and Trend-in-Time Series
Estimation d'autocorrélation et de tendances dans des séries chronologiques

David Thomson, *Queen's University*

The classical Bartlett estimate of the autocovariance of a nominally stationary time series is inadmissable. Various improved estimates have been proposed, but most are based on an underlying parametric model. A general alternative proposed in Thomson (1982) Thomson et al. (2007) is to estimate a multitaper spectrum estimate and use the Wiener – Khintchine theorem to obtain the corresponding autocorrelation estimates. These estimates are positive-definite and have much lower bias than standard estimates. In a similar manner, estimates of trend can be made by a low–frequency projection. Dependence of the estimator on parameters such as the time-bandwidth product are discussed.

L'estimateur de Bartlett classique pour l'autocovariance d'une série chronologique nominalement stationnaire est inadmissible. Plusieurs estimateurs améliorés ont été proposés, mais la plupart sont basés sur un modèle paramétrique sous-jacent. Une alternative générale proposée dans Thomson (1982) Thomson *et al*. (2007) est d'estimer un estimateur de spectre à filtres multiples et d'utiliser le théorème de Wiener-Khintchine pour obtenir les estimateurs d'autocorrélation correspondants. Ces estimateurs sont définis positifs et ont un biais beaucoup plus petit que celui des estimateurs usuels. Da façon similaire, des estimateurs de tendance peuvent être obtenus par une projection à basses fréquences. Nous discutons de la dépendance de l'estimateur par rapport aux paramètres tels le produit temps-lissage.

[MS-444]
Improved Inference for First Order Autocorrelation Using Likelihood Analysis
Inférence améliorée pour l'autocorrélation de premier ordre utilisant l'analyse de vraisemblance

Ye Sun, *University of Toronto*, Augustine Wong, *York University*, Marie Rekkas, *Simon Fraser University*

Testing for first-order autocorrelation in small samples using the standard asymptotic test can be seriously misleading. Recent methods in likelihood asymptotics are used to derive more accurate p-value approximations for testing the autocorrelation parameter in a regression model. The methods are based on conditional evaluations and are thus specific to the particular data obtained. A numerical example and simulations are provided to show that this new likelihood method provides higher order improvements and is superior in terms of central coverage even for autocorrelation parameter values close to unity.

Tester pour l'autocorrélation de premier ordre avec le test asymptotique standard peut être sérieusement trompeur en présence de petits échantillons. Des méthodes récentes en analyse asymptotique de la vraisemblance sont utilisées pour obtenir des approximations plus précises de la valeur p pour tester le paramètre d'autocorrélation dans un modèle de régression. Ces méthodes sont basées sur des évaluations conditionnelles et sont donc spécifiques aux données obtenues. Un exemple numérique et des simulations montrent que cette nouvelle méthode basée sur la vraisemblance apporte des améliorations d'ordre supérieur et est supérieure en termes de convergence centrale même pour des valeurs du paramètre d'autocorrélation près de l'unité.

[MS-445]
M-estimation for a Spatial Unilateral Autoregressive Model with Infinite Variance Innovations
M-estimation pour un modèle spatial unilatéral autorégressif avec innovations à variance infine

Mohammad Roknossadati & Mahmoud Zarepour, *University of Ottawa*

We study the limiting behavior of the M-estimators of parameters for a spatial n-lateral autoregressive model $Z_{ij} = \alpha Z_{i-1,j} + \beta Z_{i,j-1} - \alpha\beta Z_{i-1,j-1} + \epsilon_{ij}$ with independent and identically distributed innovations in the domain of attraction of a stable law with index $\alpha_0 \in (0, 2]$. Both stationary and unit root models and some extensions are considered. It is shown that only if innovations belong to the domain of attraction of a normal law, M-estimators are asymptotically normal. A numerical example and a simulation study are also given.
Keywords: Spatial autoregression; Stable process; M-estimates; Itô integral.

Nous étudions le comportement limite des M-estimateurs de paramètres d'un modèle spatial n-latéral autorégressif $Z_{ij} = \alpha Z_{i-1,j} + \beta Z_{i,j-1} - \alpha\beta Z_{i-1,j-1} + \epsilon_{ij}$ avec innovations indépendantes et identiquement distribuées dans le domaine d'attraction d'une loi stable avec indice $\alpha_0 \in (0, 2]$. Des modèles stationnaires et à racine unitaire, ainsi que quelques extensions, sont considérés. Nous montrons que l'appartenance des innovations au domaine d'attraction d'une loi normale est une condition suffisante à la normalité asymptotique des M-estimateurs. Un exemple numérique et une étude de simulation sont aussi présentés.
Mots-clés : autorégression spatiale, processus stables, M-estimateurs, intégrale d'Itô.

[MS-446]
Identification Method for Time Series Models Using Estimating Functions
Méthode d'identification pour les modèles de séries chronologiques utilisant les fonctions d'estimation

Melody Ghahramani, *University of Winnipeg*, Aerambamoorthy Thavaneswaran, *University of Manitoba*

Financial returns are often modelled as autoregressive time series with innovations having conditional heteroscedastic variances, especially with Generalized Autoregressive Heteroscedastic (GARCH) processes. The conditional distribution in GARCH models is assumed to follow a parametric distribution. Typically, the error distribution is selected without justification. In this talk, the asymptotic correlation between the sample mean and sample median is obtained as an application of combining the least squares and least absolute deviation estimating functions. We then show how the combined estimating function result can be used to identify the error distribution in linear and nonlinear time series models including GARCH models.

Les rendements financiers sont souvent modélisés sous forme de séries temporelles autorégressives dans lesquelles les innovations ont des variances hétéroscédastiques conditionnelles, en particulier dans les processus autorégressifs conditionnellement hétéroscédastiques généralisés (modèles GARCH). La distribution conditionnelle dans les modèles GARCH est censée suivre une distribution paramétrique. En règle générale, la distribution des erreurs est sélectionnée sans justification. Dans cette discussion, la corrélation asymptotique entre la moyenne d'échantillon et la médiane d'échantillon est obtenue en appliquant la combinaison des fonctions d'estimation par les moindres carrés et par les moindres écarts absolus. Nous montrons ensuite comment le résultat de la fonction d'estimation combinée peut être utilisé pour déterminer la distribution des erreurs dans les modèles de séries temporelles linéaires et non linéaires, notamment dans les modèles GARCH.

[MS-447]
Fuzzy Volatility Forecasting
Prévision de volatilités floues

Melody Ghahramani, *University of Winnipeg*, Aerambamoorthy Thavaneswaran, *University of Manitoba*

In this talk, we introduce a class of fuzzy AR-GARCH (autoregressive processes with GARCH

Dans cet exposé, nous présentons une classe de AR-GARCH (processus autorégressifs avec erreurs GARCH) flous pour

errors) for volatility modeling and study their moment properties. The superiority of fuzzy forecasts over minimum mean-square (MMSE) forecasts will also be discussed in some detail. The parameters of the AR-GARCH model are modeled using triangular fuzzy numbers where Buckley's data-driven method is used to determine the spread of the triangular fuzzy numbers by making use of the standard errors of the estimated parameters. Fuzzy forecasts will be illustrated with examples. Fuzzy forecast intervals turn out to be narrower than the corresponding MMSE forecast intervals.

la modélisation de volatilités et étudions les propriétés de leurs moments. Nous discutons de la supériorité des prévisions floues par rapport aux prévisions d'erreur quadratique moyenne minimale (EQMM). Les paramètres du modèle AR-GARCH sont modélisés au moyen de nombres triangulaires flous où la méthode de Buckley basée sur les données est utilisée pour déterminer l'étendue des nombres triangulaires flous en se servant des écarts types des paramètres estimés. Les prévisions floues seront illustrées avec des exemples. Les intervalles prédictifs flous sont plus courts que les intervalles prédictifs EQMM.

Session 20H Jeu. 29 mai • Thu. May 29, 16:30 – 18:00 Capital Hall 8A

Multivariate Statistics
Statistique multidimensionnelle

[MS-448]
Missing Values in Multiple Multivariate Analysis
Données manquantes dans les analyses multivariées multiples

Julie Josse, François Husson & Jérôme Pagès, *Agrocampus Rennes*

In this presentation, we focus on handling missing values in multiple multivariate datasets (individuals described by several sets of variables) and more precisely in Multiple Factor Analysis (Escoufier – Pagès). First, we will quickly review how missing values are taken into account in the framework of PCA. Then, we extend Bishop and Tipping's work, which showed that principal axes emerge as a maximum likelihood parameters estimation in a latent variable model closely related to Factor Analysis. An EM algorithm to obtain the principal subspace iteratively can be used and offers a proper and convenient way to deal with missing values.

Dans cette présentation, nous nous intéresserons à la gestion des données manquantes en tableaux multiples (individus décrits par plusieurs ensembles de variables) dans le cadre exploratoire de l'analyse factorielle multiple AFM). Nous rappellerons comment réaliser une ACP en présence de données manquantes. Puis nous étendrons les travaux de Bishop et Tipping (1999) qui ont montré que les axes principaux de l'ACP peuvent être obtenus comme une solution du maximum de vraisemblance d'un modèle à variables latentes apparentées à un modèle d'analyse en facteur. Cette approche nous permettra de proposer un algorithme itératif EM pour réaliser une AFM sur données incomplètes.

[MS-449]
Discriminant Multiple Correspondence Analysis
Analyse des correspondances multiples discriminantes

Rafik Abdesselam, *Université de Caen Basse-Normandie*

The proposed approach leads to analyzing the associations between a set of quantitative variables and several qualitative variables measured on a same set of individuals. In a decision-making context, the proposed method can be considered as a gen-

L'approche proposée consiste à analyser les associations entre un ensemble de variables quantitatives et plusieurs variables qualitatives mesurées sur un même ensemble d'individus. Dans un contexte d'aide à la décision, la méthode proposée peut être considérée comme une généralisa-

eralization of discriminant analysis to the multiple group variables case. It's described as a principal component analysis of the centre of gravity tables. The decomposition of its duality diagram on a double diagram illustrates the link and the passing properties between multiple correspondence analysis and discriminant analysis. An application resulting from real data illustrates the utility of the discrimination model thus defined.

tion de l'analyse discriminante à plusieurs variables cibles-groupes. Elle est présentée comme une analyse en composantes principales des tableaux des centres de gravité. La décomposition de son schéma de dualité en un double schéma illustre le lien et les propriétés de passage entre l'analyse des correspondances multiples et l'analyse discriminante. Une application sur données réelles illustre l'utilité du modèle de discrimination ainsi défini.

[MS-450]

Multivariates Non-linear Models for Estimating the Biomass of Forest Trees
Modèles non linéaires multivariés pour l'estimation de la biomasse forestière

Merlin Mbuembue Njoya & Louis-Paul Rivest, *Université Laval*

This presentation considers a set of non-linear equations which was produced for estimating the four biomass components of a tree: leaf, branch, bark and wood. Nowadays, due to the possibility of using a satellite's data and laser, the interest in obtaining accurate forest biomass predictions is increasing. We will expose an approach with multivariate non-linear models for this purpose. The idea is to use the flexibility of multivariate models, by taking account not only of the correlation between trees sharing a same plot of a forest, but also of the correlation between the four biomass components.

Cet exposé considère un système d'équations non linéaires qui a été produit pour estimer les quatre composantes de biomasse d'un arbre : feuille, branche, écorce et bois. Aujourd'hui, grâce aux laser et données satellitaires, l'intérêt de mieux prédire la biomasse forestière est grandissant. Ainsi, on va présenter une approche d'estimation des équations de biomasse par des modèles non linéaires multivariés. L'idée est d'exploiter la flexibilité des modèles multivariés, afin de tenir compte non seulement de la corrélation entre les quatre composantes de biomasse, mais aussi de la dépendance entre les arbres qui partagent une même parcelle.

The translation team did its best. Please forgive us for our misinterpretations, we are neither professional translators nor specialists in all areas of statistics.

L'équipe de traduction a fait de son mieux. Excusez-nous pour les erreurs de traduction, nous ne sommes ni des traducteurs professionnels ni des spécialistes dans tous les domaines de la statistique.

Author Index (per session)
Index des auteurs (par session)

Abarin, Taraneh 04H
Abdesselam, Rafik 20H
Abdolell, Mohamed 09J
Abdous, Belkacem 04G
Abraham, Christophe 19F
Abrahamowicz, Michal 09H, 19C
Adsavakulchai, Suwannee 20C
Agoboto, V. 14A
Ali, Ayesha 15A
Alioum, Ahmadou 05D
Alipanahi, Babak 15D
Almeida, Elizabete 17G
Alpargu, Gulhan 04E
Ammann, Caspar 09C
Amuah, Joseph 20D
Andrieu, Christophe 04D
Angelle, Bencaz 09J
Angers, Jean-François 05C
Antoniadis, Anestis 02G
Asimit, Alexandru 12I
Asimit, Jennifer 14G
Atherton, Juli 11A
Azaïs, Jean-Marc 02E
Aziz, Samina 09J
Azzalini, Adelchi 09F
Azzaoui, Nourddine 10C

Bacchetti, Peter 12A
Bacro, Jean-Noel 12I
Bader, Gary 07A
Bae, Taehan 04H
Baik, Jaiwook 20C
Balan, Raluca 10F
Bardet, Jean-Marc 10F, 15C
Bar-Hen, Avner 09C
Bashir, Shaheena 02H, 14G
Bate, Nicholas 17A
Batista, Deborah 09C
Baudry, Jean-Patrick 17F
Bay, Xavier 14F, 17E
Beatty, Mary 17A
Becker, Allan 20E
Bédard, Mylène 04D
Bedard, Thierry 07B
Bel, Liliane 09C, 12I
Béliveau, Julie 17C
Bellavance, François 07G, 14H
Benlagha, Noureddine 12I
Bentley, John 17F
Bercu, Bernard 05G
Bercu, Sophie 02E
Bernard, Carole 05B
Bernier, Jacques 04C
Bertrand, Frédéric 14H
Bessec, Marie 02E
Bessonneau, Pascal 04H
Beyene, Joseph 07A
Bezara, Manase 02F
Bhattacharya, Binay 09E
Bhattacharya, Rabi 17B
Biau, Gérard 07G, 10B, 12G, 20F
Bickel, David 17A
Bidon, Jean-Pierre 14H
Biernot, Piotr 14H
Biggs, Michael 15D
Bihan-Poudec, Alain 02I
Bilodeau, Martin 14E
Binder, David 09A
Bingham, Derek 14A
Binquet, Christine 19C
Blanke, Delphine 10B
Bleakley, Kevin 10B, 20F
Bohossian, Nora 02F
Bonithon-Kopp, Claire 19C
Bonner, Simon 07E
Bordes, Laurent 12D
Boreux, Jean-Jacques 04C
Bosa, Keven 19F

Bosq, Denis 04G, 12G
Bouaziz, Olivier 15F
Bouezmarni, Taoufik 10E
Bou-Hamad, Imad 15F
Bourdeau, Marc 15D
Bousquet, Nicolas 17G
Bouveyron, Charles 02C, 14H
Brant, Rollin 12A
Brar, Sony 19H
Brasher, Penny 19H
Braun, W. John 02I, 05F, 12D, 14H
Brewster, John 14A
Brillinger, David 14F
Brinkman, Ryan 09H
Briollais, Laurent 04E, 07A, 15E, 20E
Brion, Philippe 04B
Brisebois, François 04B
Brown, K. Stephen 09J
Brown, Patrick 09J, 15G
Browne, Gina 04F
Brua, Jean-Yves 02G
Brunel, Elodie 07D, 15F
Brunel, Nicolas 04G
Bubenik, Peter 17B
Bull, Shelley 04E, 14G, 20F
Bureau, Alexandre 04E, 17A
Burkett, Kelly 04J, 05D
Burkhalter, Robin 09J
Burkom, Howard 09B
Burnell, Margot 09J

Cadre, Benoît 12G
Cadre, Benoit 04G
Caines, Judy 09J
Campillo, Fabien 05F
Canty, Angelo 02H, 14G
Cao, Zhirong 09H
Caragea, Petrutza 17D
Caron, Francois 05C
Carraro, Laurent 17E
Carreau, Julie 09C, 14H
Carriere, K. C. 10D
Carrillo, Ivan 07E
Carroll, Raymond J. 09G
Casaert, Jeanne 04G
Casanova, Sandrine 15F
Causeur, David 09G, 14G
Celeux, Gilles 17F
Celisse, Alain 19G
Cérou, Frédéric 07G
Chagnon, Yvon 04E
Chang, Fang 04J
Chang, Hsing-Ming 05F
Chan-Yeung, Moira 20E
Chapman, Judy Anne 09J
Charest, Anne-Sophie 02C
Charpentier, Arthur 02D, 17C
Chaubey, Yogendra P. 14G
Chauvet, Guillaume 14B
Chavance, Michel 10A
Chavent, Marie 07F
Chebana, Fateh 17E
Cheddadi, Rachid 09C
Chen, Baojiang 04J, 05D
Chen, Bei 05G
Chen, Jiahua 05B, 07G, 17F
Chen, Maggie 04J
Chen, Zhao-Guo 12F
Chen, Zhijian 04J, 15E
Cheng, Emmy 09G, 10D
Chenouri, Shoja'eddin 14F, 17E
Chipman, Hugh 14H
Chiu, Grace 10D, 14C
Choi, Yun Hee 15E
Chopin, Nicolas 05C
Christine, Marc 10G
Chui, Tina 05A
Ciampi, Antonio 09H
Ciret, Pierre 14H
Ciriza, Victor 14H
Ciuperca, Gabriela 02D
Clara, Ian 14H
Clarke, Bertrand 12G
Clémençon, Stéphan 09E
Cole, David 09J
Colin, Bernard 07F
Commenges, Daniel 05D, 10A
Comte, Fabienne 07D, 07F, 15F
Cook, Richard 02A, 05D, 09G, 10C, 15A
Corbiere, Fabien 19H
Cornillon, Pierre-André 19G
Corre, Bernard 12E, 14F
Cossette, Hélène 02D
Cotton, Cecilia 05E
Couallier, Vincent 07I
Crainiceanu, Ciprian M. 09G
Croteau, Jordie 04E
Csizmadi, Ilona 09J

Dabrowski, Andre 02B, 07B
d'Alché-Buc, Florence 04G
Daouia, Abdelaati 12I
Daoust, Pierre 02F
d'Aubigny, Catherine 15D
d'Aubigny, Gérard 07I, 15D
Dauxois, Jean-Yves 12D
Davies, Katherine 05F
Davis, Karelyn 15D
Davison, Anthony 14D
Dean, C.B. 09J
Dean, Rachel 14H
de Falguerolles, Antoine 07C
de Fouquet, Chantal 14C
Dehling, Harold 02B
Delavelle, Christian 05G
Demnati, Abdellatif 07E
Deng, Dianliang 09H
Derquenne, Christian 10G
Derzko, Gérard 19C
de Saporta, Benoite 05G
Desassis, Nicolas 05F
Desgagné, Alain 15C
Dessertaine, Alain 02E, 17G
Detais, Amélie 05E
Deville, Jean-Claude 07E, 14B
Devroye, Luc 09E
Dewith, Nancy 04H
Dine, Abdessamad 07G
Ding, Keyue 19H
Ding, Zhangjun 09J, 19H
Dochitoiu, Catalin 12F
Dodge, Yadolah 20D
Donini, Laurent 14H
Donner, Allan 09J
Dordonnat, Virginie 17G
Dorvlo, Atsu 20D
Doucet, Arnaud 05C, 19F
Dowd, Michael 14C, 17G
Dowden, Jeffrey 15G
Dragieva, Nataliya 09J
Dubin, Joel 10C
Ducharme, Gilles 15C
Duchesne, Pierre 19D
Duchesne, Thierry 17G
Dumitrescu, Laura 10F
Dummer, Trevor 09J
Dupuy, Delphine 14F
Dupuy, Jean-François 05E
Durand, Gérard 17E
Durand, Jean-Baptiste 14H
Durrieu, Gilles 04E, 14H
Dyachenko, Alina 09H

Elegbede, Fabrice 20E
El Haj Ali, Dhouha 14F
Eliasziw, Michael 20D
El Matouat, Abdelaziz 07G
Embrechts, Paul 01B
Emilion, Richard 05F
Esposito-Farèse, Marina 04H
Evans, Michael 02H

Faires, Hafedh 05F
Faivre, Sébastien 10G
Fan, Steve 15G
Farewell, Vern 09B
Farrell, Patrick J. 17F
Faugeras, Olivier 02G
Favre, Anne-Catherine 05C, 10E, 17C
Faye, Laura 20F
Fayolle, David 05G
Feng, Cindy Xin 09J
Feng, Xin 20B
Ferland, René 07H
Fick, Gordon 20D
Filali, Mostafa 19G
Flemming, Joanna 17G
Flora, David 19E
Florens, Jean-Pierre 12I
Fogliatto, Flavio 12E
Fort, Gersende 04D, 15A
Fort, Jean-Claude 02E
Fortin, Mathieu 10E
Fouquau, Julien 02E
Franco, Jessica 12E, 14F
Francq, Christian 19D
Fredette, Marc 19C
Friendly, Michael 07C
Friguet, Chloé 09G
Froda, Sorana 04H
Fu, James C 05D, 05F, 09J
Futschik, Andreas 20F

Gach, Florian 20F
Gaïffas, Stéphane 07F
Gaiffas, Stéphane 07F
Gao, He (Shawn) 20
Gao, Ruochu 09J

Garivier, Aurélien 07D
Garreta, Vincent 04C
Garrido, Jose 15B
Gattuso, Linda 20A
Gauthier, Sylvie 02F
Gégout-Petit, Anne 05G
Gel, Yulia, R 05G
Genest, Christian 10E, 16A, 17C, 19D
Genton, Marc 09F
Germain, Stéphane 04G
Gerville-Réache, Léo 07I
Gery, Geenens 07F
Gettler-Summa, Mireille 05G, 20D
Ghahramani, Melody 20G
Ghodsi, Ali 15D
Ghoudi, Kilani 19D
Gijbels, Irene 02G
Gill, Richard 03A
Ginsbourger, David 17E
Giorgi, Roch 19C
Girard, Stéphane 12H, 14H
Giroux, Gaston 07H
Goga, Camélia 14B
Goldfarb, Bernard 20D
Goldmann, Gustave 09D
Goldsmith, Charlie 09B
Gonzalez, Pierre-Louis 12E
Gottardo, Raphael 05C, 09H
Gou, Tingting 14H
Gould, Josh 04I
Gouno, Evans 19F
Graham, Daniel 04F
Graham, Jinko 05D
Grancher, Delphine 04C
Grigg, Olivia 09B
Grosmann, Jacques 14H
Grover, Vaneeta 09J
Grun-Rehomme, Michel 02F, 12I
Guillas, Serge 10B
Guillotte, Simon 10E
Guilloux, Agathe 07F, 15F
Guin, Ophélie 04C
Guiot, Joél 04C
Gupta, Rameshwar 02F
Gustafson, Paul 02A, 18A, 20E
Guttorp, Peter 14C
Guyader, Arnaud 07G
Györfi, Laszlo 10B

Ha, Hyung-Tae 12H
Haataja, Riina 12C
Haddou, Mohammed 20F
Hafidi, Bezza 10C, 15D
Hallin, Marc 12C
Hamilton, David 09J
Hamzaoui, Hassania 07G
Hanley, James 20D
Hassanein, K.M. 20E
Hassanein, R.S. 20E
Hawala, Sam 19B
He, Wenqing 10D, 14G
He, Zhulin 05E
Heagerty, Patrick 05E
Hély, Christelle 04C
Hengartner, Nicolas 02G, 19G
Hossain, Shahadut 02A
Hosseinkashi, Yasaman 14F
Howlader, Tamanna 14G
Hua, Yi 14H
Huang, Mei Ling 12I
Huang, Yingru Celia 04I
Huckett, Jennifer 19B
Hurley, Catherine 02H
Husson, François 20H, 02I
Hutton, Holly 12F
Huynh, Cam-Loi 14H

Ibanez Tarradellas, Eduard 07I
Iooss, Bertrand 07I
Islam, Mohammad 09J
Islam, Muhammad 02I
Isnard, Michel 09D
Ivanoff, Gail 07B

Jacob, Pierre 12H
Jacques, Christiane 17C
Jacques, Julien 14H
Jaeschke, Roman 10D
Jafari Jozani, Mohammad 12G
Jana, Jureckova 07F
Jang, Dongdeuk 09J
Jang, Gun Ho 02H
Jaupi, Luan 07I
Jérémie, Moïse 10F
Jianfeng, Xu 04E
Jiang, Bei 10D
Jiang, Hedy 09J
Jiang, Min 12H
Jiang, Shan 19G
Jin, Zi 10F

Jinnan, Liu 20E
Johnson, Brad 04H
Joly, Pierre 05D, 12D, 19H
Jones, Bruce 12I
Josse, Julie 02I, 20H
Julien, Marilyse 20D

Kalina, Jan 07F
Kammoun, Imen 15C
Kang, Sohee 07A, 20E
Kang, Xiuli 12D
Kassri, Lahcen 07G
Khalili, Abbas 07G
Khan, Shahedul 10D
Khanchi, Aziz 09I
Kim, David 04J
Kim, Eunhee 09J
Kim, Hyang Mi 20E
Kim, Munsin 09J
Kim, Peter 02H, 17B
Klar, Neil 09J, 14G
Kloareg, Maéla 09G, 14G
Knight, John 15G
Kong, Linglong 02H
Koopman, Siem Jan 17G
Kopciuk, Karen A. 09J, 12A
Kortbi, Othmane 17F
Koudou, Efoevi 09I
Kouritzin, Michael 12B
Kovacevic, Milorad 20B
Kulik, Rafal 07B
Kulperger, Reg. J. 04H
Kustra, Rafal 07A
Kwon, Ki-Sung 09J

Labbe, Aurélie 17A
Lacour, Claire 07D
Lafaye de Micheaux, Pierre 05G, 14E, 15C
Lagnoux Renaudie, Agnès 02E
Lahiri, Partha 04B
Lai, Ming-Jun 10B
Lalanne, Christophe 04H
Laloë, Thomas 12G
Lambert, Aurore 04F
Langley, Joanne 09J
Lantuéjoul, Christian 12I
Larocque, Denis 07G, 12C
Larribe, Fabrice 05D, 14H
Larsen, Michael 04B, 19B
Latour, Alain 15G
Laurent, Béatrice 07I
Laval, Karine 04F
Lavergne, Christian 17F
Lawit, Shai 17A
Lé, Pierre 02E
Lebarbier, Emilie 14G
LeBlanc, Michael 09J
Leblanc, Alexandre 12H, 15C
Leconte, Eve 15F
L'Ecuyer, Pierre 12B
Lee, Chel Hee 04H
Lee, Gun-Young 09J
Lee, Jiyeon 07B
Lee, Terry 09C
Lee, Wan-chen 20C
Lefebvre, Geneviève 20E
Lefebvre, Sidonie 17E
Leffondré, Karen 09H
Lefieux, Vincent 17G
Lefranc, Marie-Paule 20F
Lemay, Jérôme 17G
Lemay, Jean-Phillippe 02D
Lemire, Daniel 04A
Leon, Carlos 12F
Leonardi, Florencia 07D
Lepelletier, Patrice 04F
Lesosky, Maia 02H
Lessard, Sabin 05D
Letenneur, Luc 05D
Levine, Mark 09J
Lewis, Brandon 04J
Li, Erning 02A
Li, Le 04I
Li, Longhai 19F
Li, Mingyao 10A
Li, Pengfei 17F
Li, Sebastian (Song) 04J
Li, William 12E, 14A, 17G
Li, Ye 04J
Liang, You 14H
Lim, Hyun 19H
Lin, Shili 07G
Liquet, Benoit 07F
Lisa, Chung 17A
Liu, Jennifer 10D
Liu, Jinnan 07A
Liu, Lin 05E
Lix, Lisa 09G
Lmoudden, Aziz 19F

Lo, Kenneth 09H
Lockhart, Richard 14E, 17F
Loeb, Mark 10D
Loeppky, Jason 14H
Loisel, Stéphane 15B, 17C
Lopez, Olivier 15F
Lo Siou, Geraldine 09J
Lou, W.Y.Wendy 05D, 09J
Loughin, Thomas 04H
Lozada-Can, Claudia 14D
Lu, Shuang 04I
Lu, Xuewen 05E
Lu, Zeng-Hua 05B

Ma, Renjun 02F
MacGibbon, Brenda 12D
Magnus, Jeanette 09J
Mahboubi, Amel 19C
Maheux, Hélène 05A
Maia, Manuela 14H
Malherbe, Laure 14C
Malinovsky, Yaakov 12F
Mallick, Taslim 20B
Mandal, Saumen 15D
Manivong, Phongsack 09G
Manuel, Doug 04B
Marçais, Benoît 20E
Marceau, Étienne 02D
Marchand, Éric 12G, 19F
Marin, Jean-Michel 05C
Marrel, Amandine 07I
Martin, Patrick 07G
Martsynyuk, Yuliya 10F
Massabuau, Jean-Charles 14H
Massam, Helene 07A
Massé, Hugues 14H
Massé, Jean-Claude 09F
Massie, Angelique 15A
Matthews, David 09J
Matzner-Lober, Eric 02G, 19G
Maugis, Cathy 04H
Maumy, Myriam 14H
Mavel, Marianne 04H
Maziade, Michel 04E
Mbuembue Njoya, Merlin 20H
McCulloch, Charles 02A
McDonald, David 02B, 07B
McGregor, S. Elizabeth 09J
McLeish, Don L. 07H, 08A, 10C
McLeod, Angus 09J
McLeod, Robert 12E
McNeney, Brad 05D
McNicholas, Paul 10C
Meleux, Frédérik 14C
Menneteau, Ludovic 09I
Mercadier, Cécile 02D
Mérette, Chantal 04E, 17A
Merleau, James 05C
Mesfioui, Mhamed 16A
Metzler, Adam 07H
Meyer, Mary C. 02G
Michel, Bertrand 04H
Mihram, Danielle 02I
Mihram, G. Arthur 02I
Millar, AnneMichele 02I
Miller, Paul 04C
Misiti, Michel 02E
Misiti, Yves 02E
Mizera, Ivan 02H
Mkhadri, Abdallah 15D
Moghtaderi, Azadeh 05G
Mohdeb, Zaher 02G
Molanes-Lopez, Elisa 17D
Molinari, Nicolas 10D
Molkorogov, Lorenzo 09I
Momeya, Romuald 07H
Monga, Ernest 07F
Moodie, Erica 09H, 14H, 20D
Moreno, Monica 04I
Mougin, Christian 04F
Moulines, Eric 04D
Mozumder, Md Sharif Mozumder 15B
Mulayath Variyath, Asokan 17E
Murdoch, Duncan J. 04H, 10D, 13A, 14H
Murphy, Kevin 19F
Murphy, Thomas Brendan 10C
Murua, Alejandro 15G

Nachtsheim, C. J. 14A
Nahm, Erika 14H
Nair, Vijay 19A
Nam, Bong-Hyun 09J
Naveau, Philippe 04C, 09C, 14H
Neal, Radford 19F
Nedyalkova, Desislava 07E, 10G
Neslehova, Johanna 17C
Neuhaus, John 02A
Neupane, Binod 10D

Nevalainen, Jaakko 12C
Newton, Michael 17A
N'Guessan, Assi 14H
Nguyen, Paul 04H
Niang Keita, Ndèye 12E
Nicault, Antoine 04C
Nikoloulopoulos, Aristidis 10E
Niu, Yi 04H
Nkurunziza, Sévérien 14F
Nordhausen, Klaus 06A
Nordman, Dan 17D
Norwood, Todd 15G

O'Brien, Patti 09J
Oja, Hannu 06A, 12C
Oldford, Richmond 02H
Oldford, Wayne 07G, 14H
Omelka, Marek 07F
Ooms, Marius 17G
Oppenheim, Georges 02E
Oriol, Jean-Claude 20A
Ottucsak, György 10B
Oualkacha, Karim 02H
Ouarda, Taha, B.M.J. 17E
Oyarzun, Javier 02F

Padieu, René 05A
Pagès, Jérôme 20H
Paindaveine, Davy 06A, 12C
Pakyari, Reza 19G
Pambrun, Elodie 04F
Pardoux, Catherine 05G
Parent, Eric 04C, 10E
Park, Chul-Gyu 15D
Parker, Gary 15B
Parkhomenkho, Elena 07A
Parpia, Sameer 04E
Parr-Rud, Olivia 04A
Patrangenaru, Victor 17B
Paul, Sudhir R. 09H
Payne, Jennifer 09J
Pelletier, Bruno 12G
Peng, Liang 17D
Peng, Xiaohong 02F
Peng, Yingwei 04E, 04H, 20F
Perreault, Luc 04C, 05C
Perron, François 10E
Picard, Franck 14G
Pierrat, Jean-Claude 20E
Pierrat, Lambert 07I
Plante, Jean-François 15F
Platt, Robert 12A
Poggi, Jean-Michel 02E, 02G
Pond, Gregory 05D
Potter, Dustin 09J
Powers, Stephanie 20E
Pritchard, Kathleen 09J
Provost, Serge 12H
Pullenayegum, Eleanor 10C
Pursey, Stuart 04B
Pycke, Jean-Renaud 14E

Qi, Ying 04J
Qiao, Cunye 04F
Qiu, Weiliang 02C
Qiu, Zhenguo 09H
Qu, Chunsong 09J
Qu, Conghui 04J
Qualité, Lionel 10G
Quantin, Catherine 19C
Quessy, Jean-François 16A

Rahim, Karim 14H
Rahman, Shafiqur 14F
Rais, Saad 12F
Ramirez Ramirez, Lilia 09G
Ramsay, James 19E
Ramsay, Timothy 15E
Ranjan, Pritam 14A
Rao, J.N.K. 07E
Reesor, Mark 17E
Regier, Michael 04I
Régnier, Jean-Claude 20A
Reid, Nancy 10F, 14D
Reis, Elizabeth 14H
Reiter, Jerome 19B
Rekkas, Marie 20G
Rémillard, Bruno 16A, 19D
Reynaud-Bouret, Patricia 07D
Rice, Adrian 07C
Rich, Benjamin 09H
Richardson, Thomas 09H
Richet, Yann 17E
Rinott, Yosef 12F
Rivest, Louis-Paul .. 02H, 04G, 10E, 17G, 20F, 20H
Rivoirard, Vincent 07D
Robert, Christian 05C
Robin, Stéphane 14G, 19G
Roblin, Antoine 17E
Robson, Paula J. 09J

Rohde, Angelika 12C
Roknossadati, Mohammad 20G
Romann, Alexandra 04I
Rombouts, Jeroen 10E
Rosella, Laura 04B
Rosler, Uwe 02B
Ross, William 04F
Rossi, Vivien 05F
Rosychuk, Rhonda 12A
Rosychuk, Rhonda J. 15G
Rotondi, Michael 05F
Rouen, Candice 02F
Rouïl, Laurence 14C
Rousseau, Judith 14D
Roustant, Olivier 17E
Rouvière, Laurent 02G
Roy, Marc-André 04E
Rubenthaler, Sylvain 12B
Rubin-Bleuer, Susana 12F
Rudnichi, Annie 04F
Ruiz-Cazen, Anne 14B
Rydén, Tobias 12B

Sabre, Rachid 19G
Sacks, Jerome 14H
Saha, Krishna 20F
Sahrmann, Nathan 04J
Saint-Pierre, Philippe 15F
Saleh, A.K.Md. Ehsanes 17F, 20E
Salibian-Barrera, Matias 09F
Samanta, Mrityunjay 15D
Saracco, Jérôme 07F
Sauquet, Eric 14H
Sautory, Olivier 07E
Schanzer, Dena 09J
Schaubel, Douglas 19H
Schiopu-Kratina, Ioana 10F
Schwarz, Carl 04H, 07E
Sébastien, Lê 02I
Segers, Johan 02D, 02G
Sendov, Hristo S. 10D
Seneta, Eugene 07C
Serant, Daniel 15B
Serge, University of Western Ontario 12H
Sergent, Michelle 12E
Settati, Adel 09I, 14H
Shadrock, Ali 15D
Shepherd, Lois 09J
Shin, Hwashin 13A
Shin, Ji-Hyung 04J
Shojiana, Kaveh 15E
Sikdar, Khokan 15G
Silberman, Roxane 05A, 09D
Silva, Giovani 09J
Simar, Léopold 07F, 12I
Sinha, Sanjoy 15D
Sirski, Monica 04H
Sitter, Randy 20B
Small, Christopher G. 14F
Song, Hui 04H
Song, Peter X.-K. 09H, 10A
Sow, Mohamedou 04E, 14H
Spinelli, John 15C
Stafford, Jamie 15G
Staicu, Ana-Maria 09G
Stamey, James 20E
Steele, Russell 02C
Steiner, Stefan H. 17E
Stephens, David 04F
Stephens, Michael 14E, 15C, 17F
Stewart, Sam 09J
St. Jacques, Jeannine 14H
Stoltz, Gilles 16A
Strawderman, William E. 19F
Struthers, Cyntha 10C
Stukel, Therese 04B
Sun, Dongchu 14D
Sun, Lei 04E, 20F
Sun, Shuying 09J
Sun, Ye 20G
Sutradhar, Brajendra 20B

Tabet, Aline 19F
Tabuchi, Tracy 02F
Taibi, Salima 02F, 04F
Takahara, Glen 05G, 13A, 15E
Tam, Theresa 09J
Tan, Qi Er 09J
Tan, Xianming 05E, 15E
Taskinen, Sara 06A
Tayeb, Arafat 17A
Thabane, Lehana 04E, 04F, 09G, 10D
Thavaneswaran, Aerambamoorthy 20G
Thiam, Baba 14G
Thompson, Mary 07E, 09G, 19E
Thomson, David 05G, 14H, 20G
Tian, Weidong 05B
Tichon, Jenna 12E

Tillé, Yves 07E, 10G
Tinsson, Walter 12E
Tolusso, David 09G
Tondu, Françoise 09J
Torabi, Mahmoud 15G
Toupin, Marie-Hélène 16A
Toussaint, Godfried 09E
Tran, Damien 14H
Tricaud-Vialle, Sophie 04F
Tripathi, Yogesh 12G
Tritchler, David 07A
Trogdon, Justin 05B
Trottier, Catherine 17F
Troupé, Marylène 19F
Truquet, Lionel 15G
Tsai, Cary C.L. 15B
Tsao, Min 09C
Tu, Dongsheng 04H, 05E, 15E, 19G

Ung, Anthony 14C
Ursu, Eugen 19D

Vaida, Florin 05E
Vaillant, Jean 19F
Valois, Pierre 04G
Vallois, Pierre 09I
Van Keilegom, Ingrid 17D
Varet, Suzanne 17E
Vasechko, Olga 02F, 12I
Vasseur, Olivier 12E
Vayatis, Nicolas 09E
Veraverbeke, Noel 17D
Verdebout, Thomas 12C
Vicente, Paula 14H
Vincent, Céline 15C
Vincent, Kyle 09J

Wang, Lihua 04J
Wang, Nae-Yuh 02A
Wang, Naisyin 02A
Wang, Steven 02C
Wang, Xikui 09J, 14H
Wang, Zhidong 04J
Wang, Zilin 07E
Warkentin, Theodore 02A
Warrillow, Emma 04A
Warriyar, Vineetha 04I
Watier, Francois 07H
Wei, Ying 09F
Welch, William 14H
Welsh, Stephen 04J
Westveld, Anton 14C
White, Bethany 02I
Wiktorsson, Magnus 12B
Wile, Scott 05G
Willan, Andrew 12A
Willison, Donald 09G
Winchester, Brian 09J
Wintenberger, Olivier 10F
Wong, Augustine 20G
Wong, Samuel 04I
Woodall, William 09B
Wu, Beilei 04I
Wu, Changbao 07E, 15E
Wu, Huanhuan 04H
Wu, Jingjing 11A
Wu, Ka Ho 12F
Wu, Longyang 04J
Wu, Qing 09J
Wu, Tung-Lung 05D
Wu, Yaohua 19G
Wynant, Willy 09H

Xia, Wei 17E
Xie, Bifeng 20C
Xiong, Xiaoqin 10C
Xu, Yawen 04J

Yadegari, Iraj 20D
Ye, Chenglin 04F
Yi, Grace Y. 05D, 09J, 14D, 15E
Yi, Yanqing 09J
Yilmaz, Yildiz 10E
Yoo, Yun Joo 04E
You, Yong 12F, 20B
Younes, Laurent 17B
Yu, Chi Wai 12G
Yuan, Ying 10A
Yue, Li Hua 10D

Zagdanski, Adam 07A
Zahedi, Ashkan 04H
Zakoian, Jean-Michel 19D
Zamar, Ruben 02C
Zarepour, Mahmoud 07B, 20G
Zeng, Leilei 02A, 15A
Zghoul, Ahmad 15C
Zghoul, ahmad 15C
Zhang, Han 04I
Zhang, Jiajia 20F

Zhang, Miao (Mandy) 04H
Zhang, Peng 09H
Zhang, Xu 19H
Zhang, Zhengmin 17F
Zhao, Ke 12I
Zhao, Lihui 20E
Zhao, Yonggan 05B
Zhou, Qian 20B
Zhou, Qiujing 04H
Zhou, Wu 07G
Zhu, Liting 05E
Zi, Jin 14D
Ziani, Rachid 04G
Zouch, Mariem 17F
Zwiers, Francis 09C

Author Index • Index des auteurs

Abarin, Taraneh 117
Abdesselam, Rafik 286
Abdolell, Mohamed 172, 176
Abdous, Belkacem 115
Abraham, Christophe 265
Abrahamowicz, Michal 166, 260
Adsavakulchai, Suwannee 275
Agoboto, V. 208
Ali, Ayesha 229
Alioum, Ahmadou 127
Alipanahi, Babak 234
Almeida, Elizabete 256
Alpargu, Gulhan 111
Ammann, Caspar 157
Amuah, Joseph 277
Andrieu, Christophe 108
Angelle, Bencaz 170
Angers, Jean-François 125
Antoniadis, Anestis 96
Asimit, Alexandru 207
Asimit, Jennifer 220
Atherton, Juli 191
Azaïs, Jean-Marc 91
Aziz, Samina 174
Azzalini, Adelchi 161
Azzaoui, Nourddine 183

Bacchetti, Peter 192
Bacro, Jean-Noel 205
Bader, Gary 138
Bae, Taehan 119
Baik, Jaiwook 275
Balan, Raluca 188
Bardet, Jean-Marc 189, 233
Bar-Hen, Avner 157
Bashir, Shaheena 98, 218
Bate, Nicholas 245
Batista, Deborah 157
Baudry, Jean-Patrick 253
Bay, Xavier 217, 252
Beatty, Mary 245
Becker, Allan 279
Bédard, Mylène 108
Bedard, Thierry 139
Bel, Liliane 157, 205
Béliveau, Julie 248
Bellavance, François 150, 229
Benlagha, Noureddine 206
Bentley, John 254
Bercu, Bernard 135
Bercu, Sophie 91
Bernard, Carole 123
Bernier, Jacques 107
Bertrand, Frédéric 226–228
Bessec, Marie 92
Bessonneau, Pascal 116
Beyene, Joseph 137
Bezara, Manase 94
Bhattacharya, Binay 160
Bhattacharya, Rabi 247
Biau, Gérard 149, 180, 201, 282
Bickel, David 245
Bidon, Jean-Pierre 226
Biernot, Piotr 226
Biggs, Michael 234
Bihan-Poudec, Alain 101
Bilodeau, Martin 214
Binder, David 155
Bingham, Derek 209
Binquet, Christine 260
Blanke, Delphine 180
Bleakley, Kevin 180, 282
Bohossian, Nora 93
Bonithon-Kopp, Claire 260
Bonner, Simon 143
Bordes, Laurent 195
Boreux, Jean-Jacques 107, 108
Bosa, Keven 265
Bosq, Denis 115, 201
Bouaziz, Olivier 239
Bouezmarni, Taoufik 186
Bou-Hamad, Imad 241
Bourdeau, Marc 236

Bousquet, Nicolas 255
Bouveyron, Charles 88, 223
Brant, Rollin 192
Brar, Sony 271
Brasher, Penny 271
Braun, W. John 101, 133, 196, 221
Brewster, John 208
Brillinger, David 215
Brinkman, Ryan 166
Briollais, Laurent 111, 138, 238, 278
Brion, Philippe 105
Brisebois, François 105
Brown, K. Stephen 175
Brown, Patrick 175, 241
Browne, Gina 113
Brua, Jean-Yves 96
Brunel, Elodie 142, 240
Brunel, Nicolas 114
Bubenik, Peter 246
Bull, Shelley 110, 218, 220, 281
Bureau, Alexandre 110, 246
Burkett, Kelly 121, 126
Burkhalter, Robin 175
Burkom, Howard 156
Burnell, Margot 174

Cadre, Benoît 116, 201
Caines, Judy 172
Campillo, Fabien 131
Canty, Angelo 98, 218
Cao, Zhirong 166
Caragea, Petrutza 249
Caron, François 124
Carraro, Laurent 252
Carreau, Julie 157, 222
Carriere, K. C. 184
Carrillo, Ivan 143
Carroll, Raymond J. 165
Casaert, Jeanne 115
Casanova, Sandrine 240
Causeur, David 162, 219
Celeux, Gilles 253
Celisse, Alain 268
Cérou, Frédéric 149
Chagnon, Yvon 110
Chang, Fang 121
Chang, Hsing-Ming 132
Chan-Yeung, Moira 279
Chapman, Judy Anne 174
Charest, Anne-Sophie 88
Charpentier, Arthur 89, 248
Chaubey, Yogendra P. 219
Chauvet, Guillaume 209
Chavance, Michel 178
Chavent, Marie 146
Chebana, Fateh 251
Cheddadi, Rachid 157
Chen, Baojiang 121, 128
Chen, Bei 134
Chen, Jiahua 123, 148, 255
Chen, Maggie 121
Chen, Zhao-Guo 199
Chen, Zhijian 121, 238
Cheng, Ji (Emmy) 163, 184
Chenouri, Shojaeddin 215, 252
Chipman, Hugh 224
Chiu, Grace 185, 211
Choi, Yun Hee 238
Chopin, Nicolas 125
Christine, Marc 191
Chui, Tina 122
Ciampi, Antonio 167
Ciret, Pierre 227
Ciriza, Victor 225
Ciuperca, Gabriela 90
Clara, Ian 222
Clarke, Bertrand 202
Clémençon, Stéphan 160
Cole, David 177
Colin, Bernard 147
Commenges, Daniel 127, 179
Comte, Fabienne 142, 148, 240
Cook, Richard 86, 128, 164, 181, 230
Corbiere, Fabien 269
Cornillon, Pierre-André 267
Corre, Bernard 197, 217
Cossette, Hélène 90
Cotton, Cecilia 130
Couallier, Vincent 153
Crainiceanu, Ciprian M. 165
Croteau, Jordie 110
Csizmadi, Ilona 172

Dabrowski, André 87, 139, 140
d'Alché-Buc, Florence 114
Daouia, Abdelaati 206
Daoust, Pierre 93
d'Aubigny, Catherine 235

d'Aubigny, Gérard 152, 235
Dauxois, Jean-Yves 195
Davies, Katherine 133
Davis, Karelyn 235
Davison, Anthony 212
Dean, C.B. 178
Dean, Rachel 223
de Falguerolles, Antoine 140
de Fouquet, Chantal 212
Dehling, Harold 87
Delavelle, Christian 134
Demnati, Abdellatif 145
Deng, Dianliang 167
Derquenne, Christian 190
Derzko, Gérard 260
de Saporta, Benoite 135
Desassis, Nicolas 131
Desgagné, Alain 232
Dessertaine, Alain 92, 257
Detais, Amélie 129
Deville, Jean-Claude 144, 210
Devroye, Luc 159
Dewith, Nancy 117
Dine, Abdessamad 150
Ding, Keyue 270
Ding, Zhangjun 174, 270
Dochitoiu, Catalin 200
Dodge, Yadolah 276
Donini, Laurent 225
Donner, Allan 176
Dordonnat, Virginie 257
Dorvlo, Atsu 277
Doucet, Arnaud 124, 265
Dowd, Michael 211, 256
Dowden, Jeffrey 242
Dragieva, Nataliya 175
Dubin, Joel 181
Ducharme, Gilles 233
Duchesne, Pierre 261
Duchesne, Thierry 255
Dumitrescu, Laura 188
Dummer, Trevor 174
Dupuy, Delphine 217
Dupuy, Jean-François 129
Durand, Gérard 250
Durand, Jean-Baptiste 225
Durrieu, Gilles 111, 227
Dyachenko, Alina 167

Elegbede, Fabrice 280
El Haj Ali, Dhouha 216
Eliasziw, Michael 277
El Matouat, Abdelaziz 149
Embrechts, Paul 84
Emilion, Richard 132
Esposito-Farèse, Marina 116
Evans, Michael 99

Faires, Hafedh 132
Faivre, Sébastien 191
Fan, Steve 241
Farewell, Vern 155
Farrell, Patrick J. 254
Faugeras, Olivier 95
Favre, Anne-Catherine 125, 187, 248
Faye, Laura 281
Fayolle, David 134
Feng, Cindy Xin 178
Feng, Xin 273
Ferland, René 150
Fick, Gordon 277
Filali, Mostafa 267
Flemming, Joanna 256
Flora, David 263
Florens, Jean-Pierre 206
Fogliatto, Flavio 198
Fort, Gersende 109, 230
Fort, Jean-Claude 91
Fortin, Mathieu 186
Fouquau, Julien 92
Franco, Jessica 197, 217
Francq, Christian 262
Fredette, Marc 261
Friendly, Michael 141
Friguet, Chloé 162
Froda, Sorana 119
Fu, James C. 128, 132, 171
Futschik, Andreas 283

Gach, Florian 283
Gaïffas, Stéphane 146, 148
Gao, He (Shawn) 121
Gao, Ruochu 176
Garivier, Aurélien 142
Garreta, Vincent 107
Garrido, Jose 231
Gattuso, Linda 271
Gauthier, Sylvie 93
Gégout-Petit, Anne 135
Gel, Yulia, R 134

Genest, Christian 186, 244, 248, 262
Genton, Marc 161
Germain, Stéphane 115
Gerville-Réache, Léo 153
Gery, Geenens 147
Gettler-Summa, Mireille 134, 277
Ghahramani, Melody 285
Ghodsi, Ali 234
Ghoudi, Kilani 262
Gibbs, Alison 121
Gijbels, Irene 96
Gill, Richard 103
Ginsbourger, David 251, 252
Giorgi, Roch 260
Girard, Stéphane 204, 225
Giroux, Gaston 150
Goga, Camélia 210
Goldfarb, Bernard 277
Goldmann, Gustave 159
Goldsmith, Charlie 156
Gonzalez, Pierre-Louis 198
Gottardo, Raphael 124, 166
Gou, Tingting 221, 222
Gould, Josh 121
Gouno, Evans 266
Graham, Daniel 112
Graham, Jinko 126
Grancher, Delphine 108
Grigg, Olivia 156
Grosmann, Jacques 228
Grover, Vaneeta 177
Grun-Rehomme, Michel 94, 206
Guillas, Serge 180
Guillotte, Simon 187
Guilloux, Agathe 146, 148, 240
Guin, Ophélie 108
Guiot, Joél 107
Gupta, Rameshwar 93
Gustafson, Paul 86, 257, 278
Guttorp, Peter 211
Guyader, Arnaud 149
Györfi, Laszlo 180

Ha, Hyung-Tae 204
Haataja, Riina 194
Haddou, Mohammed 281
Hafidi, Bezza 183, 236
Hallin, Marc 195
Hamilton, David 177
Hamzaoui, Hassania 149
Hanley, James 276
Hassanein, K.M. 280
Hassanein, R.S. 280
Hawala, Sam 258
Haziza, David 121
He, Wenqing 183, 218
He, Zhulin 129
Heagerty, Patrick 130
Hély, Christelle 107
Hengartner, Nicolas 97, 266, 267
Hossain, Shahadut 86
Hosseinkashi, Yasaman 215
Howlader, Tamanna 219
Hua, Yi 221
Huang, Mei Ling 205
Huang, Yingru Celia 121
Huckett, Jennifer 259
Hurley, Catherine 99
Husson, François 100, 286
Hutton, Holly 199
Huynh, Cam-Loi 222

Ibanez Tarradellas, Eduard 153
Iooss, Bertrand 152
Islam, Mohammad 177
Islam, Muhammad 102
Isnard, Michel 159
Ivanoff, Gail 139

Jacob, Pierre 204
Jacques, Christiane 248
Jacques, Julien 223
Jaeschke, Roman 184
Jafari Jozani, Mohammad 203
Jana, Jureckova 146
Jang, Dongdeuk 171
Jang, Gun Ho 99
Jaupi, Luan 153
Jianfeng, Xu 109
Jiang, Bei 184
Jiang, Hedy 175
Jiang, Min 204
Jiang, Shan 268
Jin, Zi 189
Jinnan, Liu 278
Johnson, Brad 118
Joly, Pierre 127, 195, 269
Jones, Bruce 207
Josse, Julie 100, 286

Julien, Marilyse 276

Kalina, Jan 146
Kammoun, Imen 233
Kang, Sohee 138, 278
Kang, Xiuli 196
Kassri, Lahcen 149
Khalili, Abbas 148
Khan, Shahedul 185
Khanchi, Aziz 170
Kim, David 121
Kim, Eunhee 171
Kim, Hyang Mi 280
Kim, Munsin 171
Kim, Peter 99, 246
Klar, Neil 176, 220
Kloareg, Maéla 162, 219
Knight, John 242
Kong, Linglong 100
Koopman, Siem Jan 257
Kopciuk, Karen A. 172, 192
Kortbi, Othmane 253
Koudou, Efoevi 168
Kouritzin, Michael 193
Kovacevic, Milorad 273
Kulik, Rafał 139
Kulperger, Reg. J. 119
Kuromi, Gordon 121
Kustra, Rafal 137
Kwon, Ki-Sung 171

Labbe, Aurélie 246
Lacour, Claire 142
Lafaye de Micheaux, Pierre 133, 214, 232
Lagnoux Renaudie, Agnès 91
Lahiri, Partha 105
Lai, Ming-Jun 180
Lalanne, Christophe 116
Laloë, Thomas 202
Lambert, Aurore 113
Langley, Joanne 174
Lantuéjoul, Christian 205
Larocque, Denis 150, 194
Larose, François 101
Larribe, Fabrice 126, 224
Larsen, Michael 105, 259
Latour, Alain 243
Laurent, Béatrice 152
Laval, Karine 113
Lavergne, Christian 254
Lawit, Shai 245
Lé, Pierre 91
Lebarbier, Emilie 219
LeBlanc, Michael 176
Leblanc, Alexandre 203, 232
Leconte, Eve 240
L'Ecuyer, Pierre 193
Lee, Chel Hee 117
Lee, Gun-Young 171
Lee, Jiyeon 140
Lee, Terry 158
Lee, Wan-chen 275
Lefebvre, Geneviève 278
Lefebvre, Sidonie 250
Leffondré, Karen 166
Lefieux, Vincent 256
Lefranc, Marie-Paule 282
Lemay, Jérôme 255
Lemay, Jean-Phillippe 90
Lemire, Daniel 103
Leon, Carlos 200
Leonardi, Florencia 142
Lepelletier, Patrice 113
Lesosky, Maia 99
Lessard, Sabin 126
Letenneur, Luc 127
Levine, Mark 174
Lewis, Brandon 121
Li, Erning 85
Li, Le 121
Li, Longhai 264
Li, Mingyao 179
Li, Pengfei 255
Li, Sebastian (Song) 121
Li, William 196, 208, 256
Li, Ye 121
Liang, You 220
Lim, Hyun 269
Lin, Shili 148
Liquet, Benoit 146
Lisa, Chung 245
Liu, Jennifer 184
Liu, Jinnan 138
Liu, Lin 130
Lix, Lisa 163
Lmoudden, Aziz 265
Lo, Kenneth 166
Lockhart, Richard 215, 254
Loeb, Mark 184

Loeppky, Jason 224
Loisel, Stéphane 231, 248
Lopez, Olivier 239
Lo Siou, Geraldine 172
Lou, W.Y.Wendy 127, 172
Lou, Wendy 172
Loughin, Thomas 117
Lozada-Can, Claudia 212
Lu, Shuang 121
Lu, Xuewen 129
Lu, Zeng-Hua 123

Ma, Renjun 93
MacGibbon, Brenda 196
Magnus, Jeanette 170
Mahboubi, Amel 260
Maheux, Hélène 122
Maia, Manuela 226
Malherbe, Laure 212
Malinovsky, Yaakov 200
Mallick, Taslim 273
Mandal, Saumen 234
Manivong, Phongsack 163
Manuel, Doug 106
Marçais, Benoît 280
Marceau, Étienne 90
Marchand, Éric 202, 203, 265
Marin, Jean-Michel 125
Marrel, Amandine 152
Martin, Patrick 149
Martsynyuk, Yuliya 188
Massabuau, Jean-Charles 227
Massam, Helene 138
Massé, Hugues 224
Massé, Jean-Claude 161
Massie, Angelique 229
Matthews, David 174
Matzner-Lober, Eric 97, 266, 267
Maugis, Cathy 118
Maumy, Myriam 226, 228
Mavel, Marianne 116
Maziade, Michel 110
Mbuembue Njoya, Merlin 287
McCulloch, Charles 85
McDonald, David 87, 140
McGregor, S. Elizabeth 172
McLeish, Don 154, 182
McLeish, Don L. 151
McLeod, Angus 177
McLeod, Robert 196
McNeney, Brad 126
McNicholas, Paul 182
Meleux, Frédérik 212
Menneteau, Ludovic 168
Mercadier, Cécile 90
Mérette, Chantal 110, 246
Merleau, James 125
Mesfioui, Mhamed 244
Metzler, Adam 151
Meyer, Mary C. 97
Michel, Bertrand 118
Mihram, Danielle 102
Mihram, G. Arthur 102
Millar, AnneMichele 101
Miller, Paul 107
Misiti, Michel 91
Misiti, Yves 91
Mizera, Ivan 100
Mkhadri, Abdallah 236
Moghtaderi, Azadeh 133
Mohdeb, Zaher 97
Molanes-Lopez, Elisa 249
Molinari, Nicolas 185
Molkorogov, Lorenzo 169
Momeya, Romuald 151
Monga, Ernest 147
Moodie, Erica 167, 226, 276
Moreno, Monica 121
Mougin, Christian 113
Moulines, Eric 109
Mozumder, Md Sharif Mozumder 231
Mulayath Variyath, Asokan 252
Murdoch, Duncan J. 117, 183, 207, 221, 222
Murphy, Kevin 265
Murphy, Thomas Brendan 182
Murua, Alejandro 243

Nachtsheim, C. J. 208
Nahm, Erika 224
Nair, Vijay 258
Nam, Bong-Hyun 171
Naveau, Philippe 107, 108, 157, 222
Neal, Radford 264
Nedyalkova, Desislava 145, 190
Neslehova, Johanna 248
Neuhaus, John 85
Neupane, Binod 184
Nevalainen, Jaakko 194

Newton, Michael 245
N'Guessan, Assi 229
Nguyen, Paul 117
Niang Keita, Ndèye 198
Nicault, Antoine 107
Nikoloulopoulos, Aristidis 186
Niu, Yi 120
Nkurunziza, Sévérien 216
Nordhausen, Klaus 136
Nordman, Dan 249
Norwood, Todd 241

O'Brien, Patti 174
Oja, Hannu 136, 194
Oldford, Richmond 99
Oldford, Wayne 148, 223
Omelka, Marek 146
Ooms, Marius 257
Oppenheim, Georges 91
Oriol, Jean-Claude 272
Ottucsak, György 180
Oualkacha, Karim 98
Ouarda, Taha, B.M.J. 251
Oyarzun, Javier 94

Padieu, René 122
Pagès, Jérôme 286
Paindaveine, Davy 136, 195
Pakyari, Reza 268
Pambrun, Elodie 114
Pardoux, Catherine 134
Parent, Eric 107, 187
Park, Chul-Gyu 235
Parker, Gary 232
Parkhomenkho, Elena 137
Parpia, Sameer 110
Parr-Rud, Olivia 104
Patrangenaru, Victor 247
Paul, Sudhir R. 167
Payne, Jennifer 172
Pelletier, Bruno 201
Peng, Liang 250
Peng, Xiaohong 93
Peng, Yingwei 109, 120, 283
Perreault, Luc 107, 125
Perron, François 187
Picard, Franck 219
Pierrat, Jean-Claude 280
Pierrat, Lambert 152
Plante, Jean-François 239
Platt, Robert 192
Poggi, Jean-Michel 91, 96
Pond, Gregory 127
Potter, Dustin 171
Powers, Stephanie 279
Pritchard, Kathleen 174
Provost, Serge 204
Pullenayegum, Eleanor 181
Pursey, Stuart 105
Pycke, Jean-Renaud 214

Qi, Ying 121
Qiao, Cunye 112
Qiu, Weiliang 88
Qiu, Zhenguo 165
Qu, Chunsong 174
Qu, Conghui 121
Qualité, Lionel 190
Quantin, Catherine 260
Quessy, Jean-François 244

Rahim, Karim 228
Rahman, Shafiqur 217
Rais, Saad 198
Ramirez Ramirez, Lilia 164
Ramsay, James 263
Ramsay, Timothy 237
Ranjan, Pritam 209
Rao, J.N.K. 145
Reesor, Mark 252
Regier, Michael 121
Régnier, Jean-Claude 272
Reid, Nancy 189, 213
Reis, Elizabeth 226
Reiter, Jerome 259
Rekkas, Marie 284
Rémillard, Bruno 244, 262
Reynaud-Bouret, Patricia 142
Rice, Adrian 141
Rich, Benjamin 167
Richardson, Thomas 167
Richet, Yann 251
Rinott, Yosef 200
Rivest, Louis-Paul 98, 115, 186, 255, 281, 287
Rivoirard, Vincent 142
Robert, Christian 125
Robin, Stéphane 219, 268
Roblin, Antoine 250
Robson, Paula J. 172
Rohde, Angelika 194

Roknossadati, Mohammad 285
Romann, Alexandra 121
Rombouts, Jeroen 186
Rosella, Laura 106
Rosler, Uwe 87
Ross, William 112
Rossi, Vivien 131
Rosychuk, Rhonda J. 192, 242
Rotondi, Michael 131
Rouen, Candice 94
Rouïl, Laurence 212
Rousseau, Judith 213
Roustant, Olivier 251
Rouvière, Laurent 97
Roy, Marc-André 110
Rubenthaler, Sylvain 193
Rubin-Bleuer, Susana 200
Rudnichi, Annie 113
Ruiz-Cazen, Anne 210
Rydén, Tobias 193

Sabre, Rachid 267
Sacks, Jerome 224
Saha, Krishna 282
Sahrmann, Nathan 121
Saint-Pierre, Philippe 240
Saleh, A.K.Md. Ehsanes 254
Saleh, A.K.Md.Ehsanes 280
Salibian-Barrera, Matias 162
Samanta, Mrityunjay 234
Saracco, Jérôme 146
Sauquet, Eric 222
Sautory, Olivier 144
Schanzer, Dena 174
Schaubel, Douglas 270
Schiopu-Kratina, Ioana 188
Schwarz, Carl 117, 143
Sébastien, Lê 100
Segers, Johan 89, 97
Sendov, Hristo S. 183
Seneta, Eugene 141
Serant, Daniel 231
Sergent, Michelle 197
Settati, Adel 169, 228
Shadrock, Ali 235
Shepherd, Lois 174
Shin, Hwashin 207
Shin, Ji-Hyung 121
Shojiana, Kaveh 237
Sikdar, Khokan 242
Silberman, Roxane 122, 158
Silva, Giovani 178
Simar, Léopold 147, 206
Sinha, Sanjoy 235
Sirski, Monica 120
Sitter, Randy 273
Small, Christopher G. 215
Song, Hui 120
Song, Peter 179
Song, Peter X.-K. 165
Sow, Mohamedou 111, 227
Spinelli, John 232
Stafford, Jamie 241
Staicu, Ana-Maria 165
Stamey, James 279
Steele, Russell 88
Steiner, Stefan H. 252
Stephens, David 112
Stephens, Michael 215, 232, 254
Stewart, Sam 172, 176
St. Jacques, Jeannine 228
Stoltz, Gilles 244
Strawderman, William E. 265
Struthers, Cyntha 182
Stukel, Therese 106
Sun, Dongchu 213
Sun, Lei 110, 281
Sun, Shuying 171
Sun, Ye 284
Sutradhar, Brajendra 273

Tabet, Aline 265
Tabuchi, Tracy 95
Taibi, Salima 94, 113
Takahara, Glen 133, 207, 237
Tam, Theresa 174
Tan, Qi Er 171
Tan, Xianming 130, 237
Taskinen, Sara 136
Tayeb, Arafat 246
Thabane, Lehana 110, 113, 163, 184
Thavaneswaran, Aerambamoorthy 285
Thiam, Baba 219
Thompson, Mary 144, 164, 264
Thomson, David 133, 228, 284
Tian, Weidong 123
Tichon, Jenna 196
Tillé, Yves 145, 190

Tinsson, Walter 197
Tolusso, David 164
Tondu, Françoise 177
Torabi, Mahmoud 242
Toupin, Marie-Hélène 244
Toussaint, Godfried 160
Tran, Damien 227
Tricaud-Vialle, Sophie 114
Tripathi, Yogesh 202
Tritchler, David 137
Trogdon, Justin 123
Trottier, Catherine 254
Troupé, Marylène 266
Truquet, Lionel 243
Tsai, Cary C.L. 232
Tsao, Min 158
Tu, Dongsheng 120, 130, 237, 268

Ung, Anthony 212
Ursu, Eugen 261

Vaida, Florin 130
Vaillant, Jean 266
Vallois, Pierre 168
Valois, Pierre 115
Van Keilegom, Ingrid 249
Varet, Suzanne 250
Vasechko, Olga 94, 206
Vasseur, Olivier 197
Vayatis, Nicolas 160
Veraverbeke, Noel 249
Verdebout, Thomas 195
Vicente, Paula 226
Vincent, Céline 233
Vincent, Kyle 173

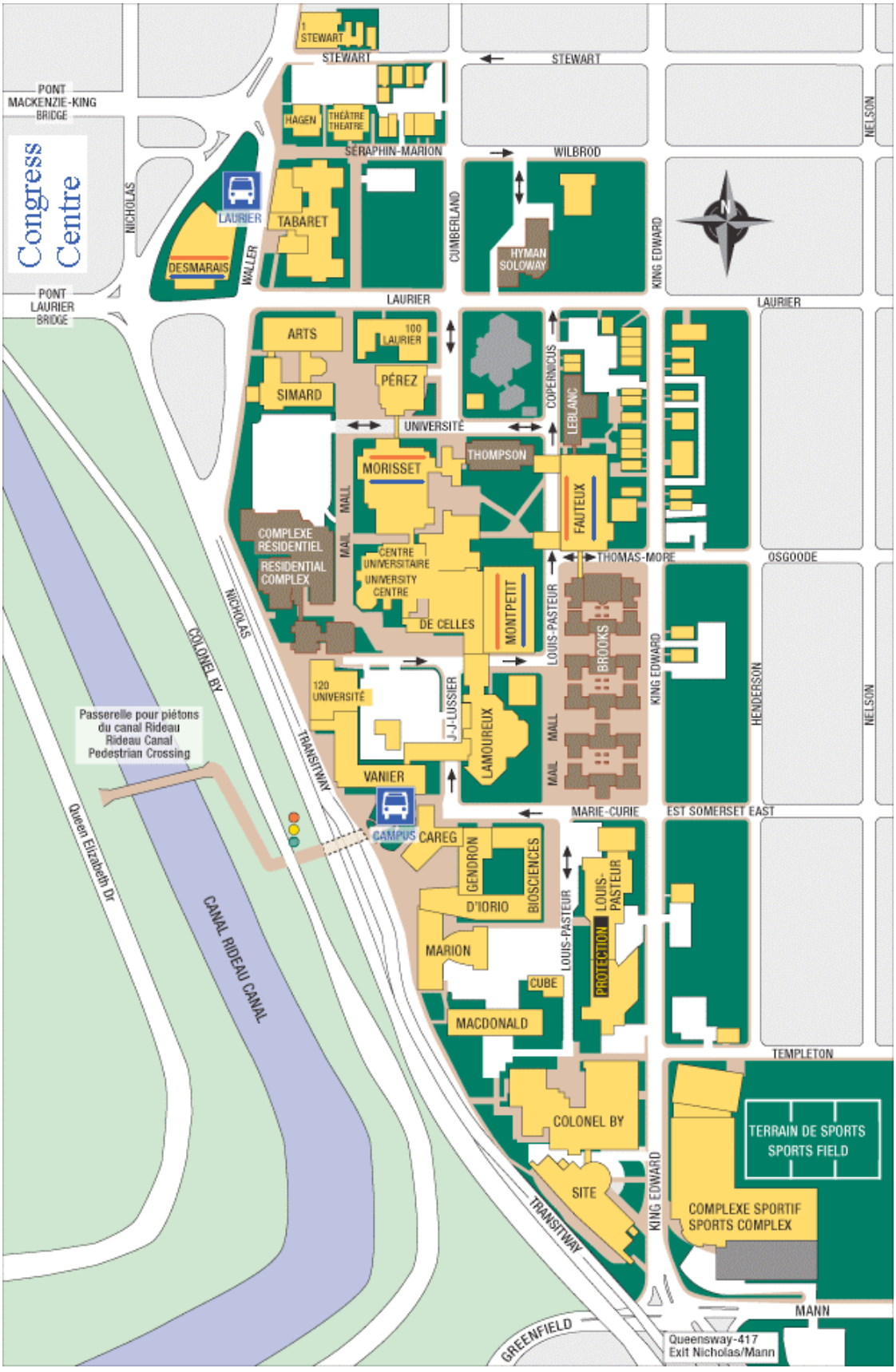
Wang, Lihua 121
Wang, Nae-Yuh 85
Wang, Naisyin 85
Wang, Steven 88
Wang, Xikui 171, 173, 221
Wang, Zhidong 121
Wang, Zilin 144
Warkentin, Theodore 86
Warrillow, Emma 104
Warriyar, Vineetha 121
Watier, François 150
Wei, Ying 162
Welch, William 224
Welsh, Stephen 121
Westveld, Anton 211
White, Bethany 101
Wiktorsson, Magnus 193
Wile, Scott 135
Willan, Andrew 192
Willison, Donald 163
Winchester, Brian 174
Wintenberger, Olivier 189
Wong, Augustine 284
Wong, Samuel 121
Woodall, William 156
Wu, Beilei 121
Wu, Changbao 143, 238
Wu, Huanhuan 117
Wu, Jingjing 191
Wu, Ka Ho 199
Wu, Longyang 121
Wu, Qing 170
Wu, Tung-Lung 128
Wu, Yaohua 268
Wynant, Willy 166

Xia, Wei 252
Xie, Bifeng 274
Xiong, Xiaoqin 181
Xu, Yawen 121

Yadegari, Iraj 276
Ye, Chenglin 113
Yi, Grace 128, 175, 213, 238
Yi, Yanqing 173
Yilmaz, Yildiz 187
Yoo, Yun Joo 110
You, Yong 200, 274
Younes, Laurent 247
Yu, Chi Wai 202
Yuan, Ying 179
Yue, Li Hua 183

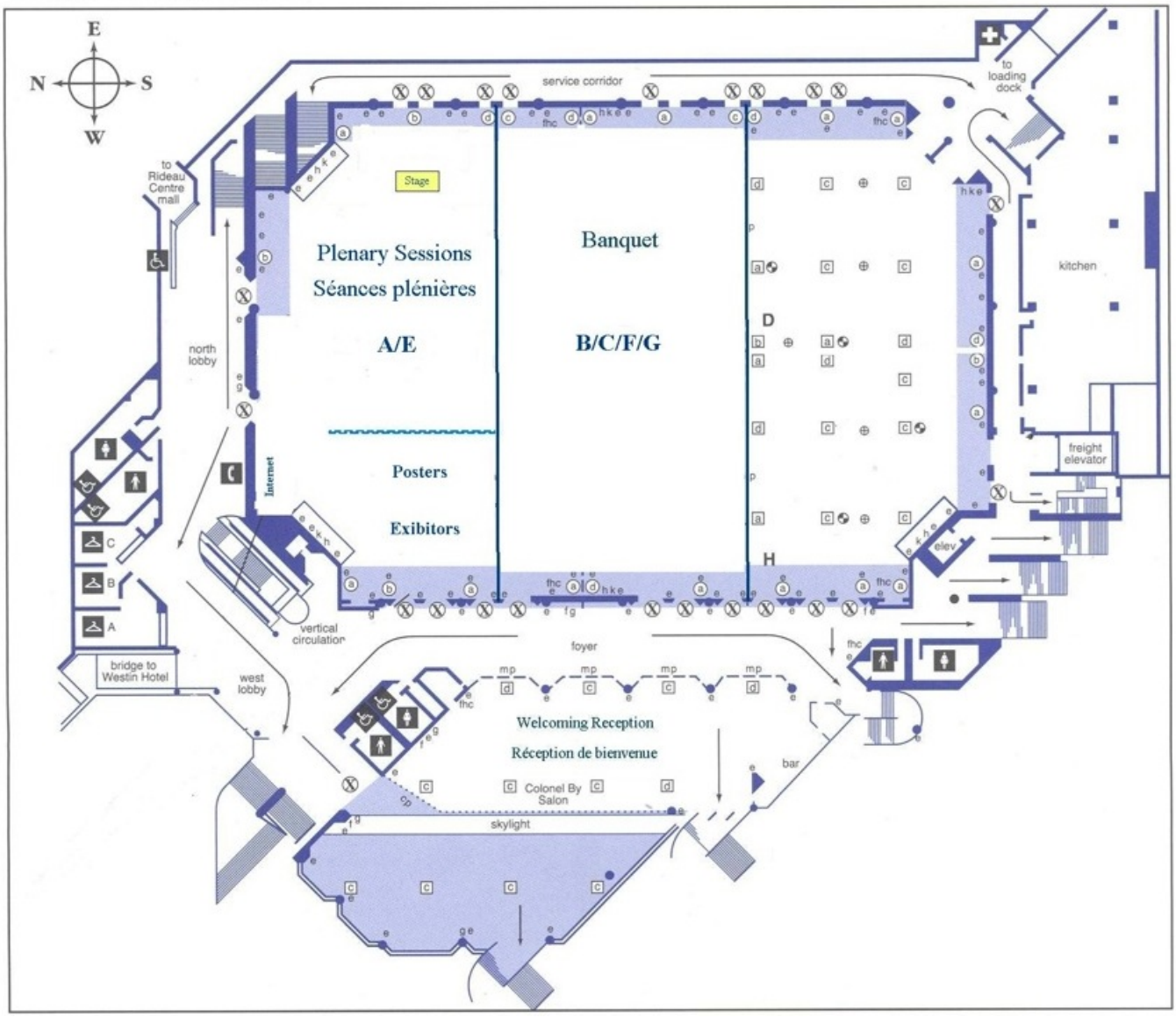
Zagdanski, Adam 137
Zahedi, Ashkan 119
Zakoian, Jean-Michel 262
Zamar, Ruben 88
Zarepour, Mahmoud 139, 285
Zeng, Leilei 86, 230
Zghoul, Ahmad 234
Zhang, Han 121
Zhang, Jiajia 283
Zhang, Miao (Mandy) 119
Zhang, Peng 165

Zhang, Xu 269
Zhang, Zhengmin 254
Zhao, Ke 205
Zhao, Lihui 279
Zhao, Yonggan 124
Zhou, Qian 274
Zhou, Qiujing 118
Zhou, Wu 148
Zhu, Liting 130
Zi, Jin 213
Ziani, Rachid 115
Zouch, Mariem 254
Zwiers, Francis 158

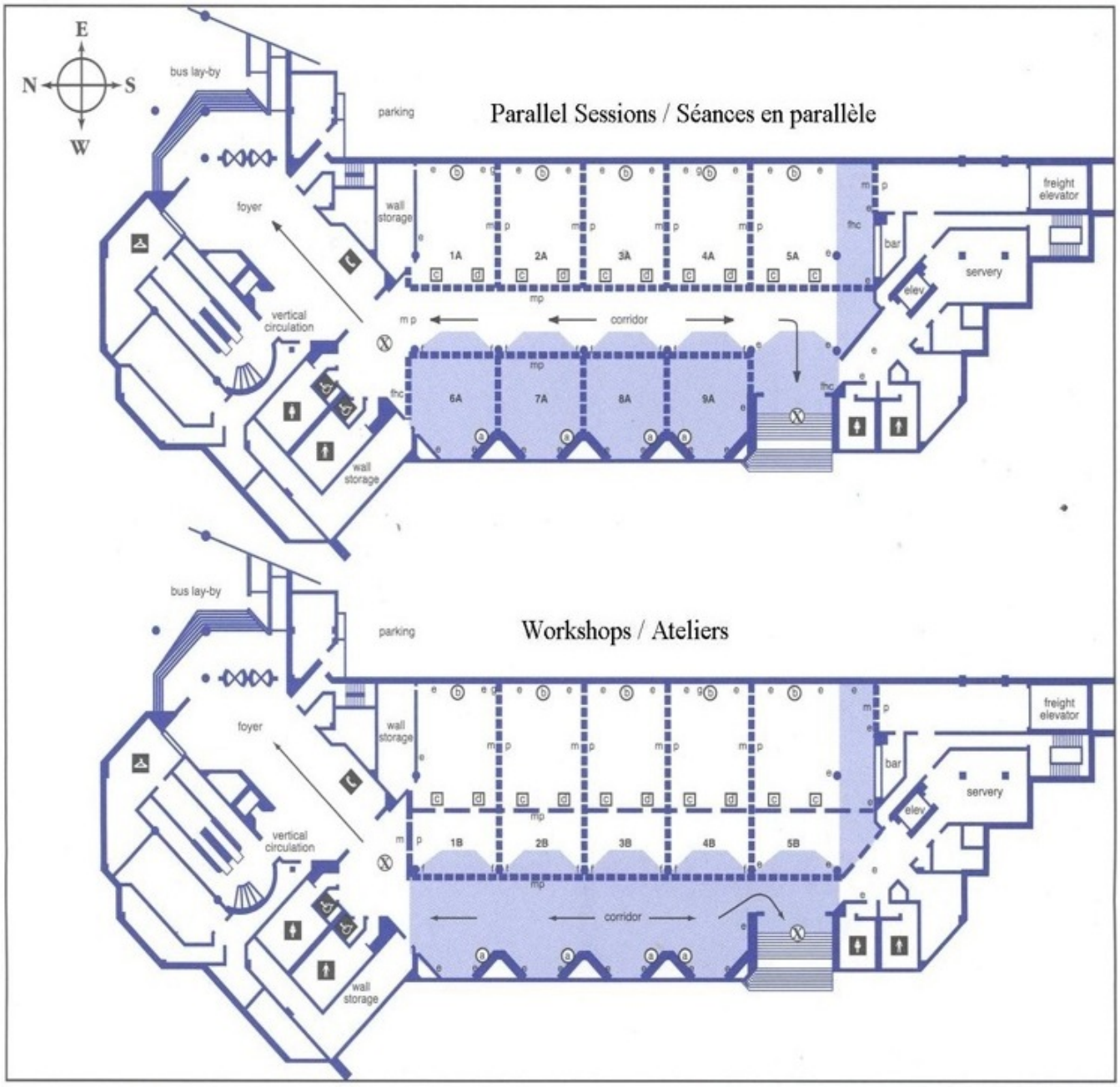


Ottawa University's Campus • Campus de l'Université d'Ottawa

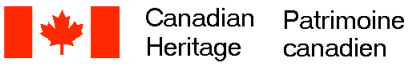
Congress Hall Level



Capital Hall Level



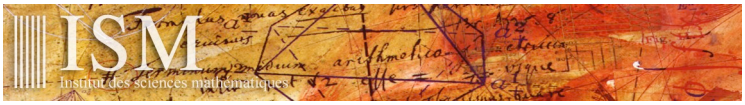
Liberté • Égalité • Fraternité
RÉPUBLIQUE FRANÇAISE


Canadian Heritage
Patrimoine canadien

Carleton
UNIVERSITY
Canada's Capital University

CENTRE
DE RECHERCHES
MATHÉMATIQUES

FIELDS


ISM
Institut des sciences mathématiques

MITACS

sas
Superior software that gives you
THE POWER TO KNOW.

STATISTICAL SOCIETY OF OTTAWA
CHAPTER - ASA
REGIONAL ASSOCIATION - SSC
SOCIETÉ STATISTIQUE D'OTTAWA
SECTION - ASA
ASSOCIATION RÉGIONALE - SSC

Pacific Institute for the Mathematical Sciences

Statistics Canada
Statistique Canada

uOttawa
L'Université canadienne
Canada's university