



**HAL**  
open science

## Estimation selon le transport optimal et les fonctions caractérisantes

Sonia Fourati, Henri Heinich, Mhamed Itmi

► **To cite this version:**

Sonia Fourati, Henri Heinich, Mhamed Itmi. Estimation selon le transport optimal et les fonctions caractérisantes. Annales de l'ISUP, 2015, 59 (3), pp.47-80. hal-03604879

**HAL Id: hal-03604879**

**<https://hal.science/hal-03604879>**

Submitted on 10 Mar 2022

**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.

## Estimation selon le transport optimal et les fonctions caractérisantes

SONIA FOURATI \* HENRI HEINICH † ET MHAMED ITMI ‡

### Résumé

*Nous commençons par améliorer certains estimateurs d'un modèle  $\{P_\theta\}_{\theta \in \mathbb{R}^d}$ , puis nous donnons deux méthodes permettant de trouver de nouveaux estimateurs. La première, axée sur le transport optimal, conduit aux estimateurs de Monge. En exemples nous examinons les lois de Lévy, de Weibull et Log-Logistique. La seconde méthode, basée sur des fonctions "caractérisantes", permet de nouvelles estimations que nous développons pour le mélange de gaussiennes, les lois Gamma et les lois Stables.*

### Abstract

Estimation according to the optimal transportation and characterizing functions.

*We begin by improving some estimators of a model  $\{P_\theta\}_{\theta \in \mathbb{R}^d}$ , then we give two methods allowing to find new estimators. The first one with the optimal transportation leads to Monge estimators. As examples we examine the laws of Levy, Weibull and Log-Logistic. The second method, based on characterizing functions, also allows new estimators that we develop for the Gaussian-mixture, Gamma and Stable-distribution laws.*

AMS 2000 Subject Classifications: 62F10

Key words and phrases: Best estimators, median, optimal transportation, Weibull, Probability Mixture, Gamma, Lévy, Stable Distribution and Log-Logistic estimations.

## 1 Introduction

### 1.1 Plan

La partie 2 est consacré à l'amélioration d'un estimateur 1-linéaire d'un paramètre de translation, voir aussi [15]. Nous examinons ensuite le cas d'un estimateur 1-affine,  $\mathcal{E}$ , et donnons des conditions de domination de  $\mathcal{E}$  par  $a.\mathcal{E}(X)+b$ ,  $(a, b) \in (\mathbb{R}^d)^2$ .

\*Normandie Université, INSA de Rouen, LMI, Avenue de l'Université 76801 Saint-tienne-du-Rouvray Cedex France. E-mail: Sonia.Fourati@insa-rouen.fr

†Normandie Université, INSA de Rouen, LMI, Avenue de l'Université 76801 Saint-tienne-du-Rouvray Cedex France. E-mail: heinich@insa-rouen.fr

‡Normandie Université, INSA de Rouen, LITIS, Avenue de l'Université, 76801 Saint-tienne-du-Rouvray Cedex. E-mail: mhamed.itmi@insa-rouen.fr

La partie principale de ce travail est consacrée à la recherche de nouveaux estimateurs. À cette fin deux méthodes sont utilisées.

- \* La première basée sur le transport optimal conduit aux estimateurs dits de Monge.

- \* La seconde utilise des fonctions "caractérisantes", comme la fonction caractéristique, donne d'autres nouveaux estimateurs consistants.

- Dans le paragraphe 3, nous utilisons l'approche du transport optimal, voir par exemple [11], dans un cadre statistique pour la recherche d'estimateurs. Cette méthode a été amorcée dans [7]. Outre le cas d'un paramètre de translation, nous abordons le cas d'un paramètre d'échelle en considérant une distance logarithmique, invariante par multiplication. Nous déterminons, pour cette distance, la notion de médiane, l'estimateur de Monge et le risque associé, voir le paragraphe 3.6.

Les résultats sont illustrés par des exemples classiques comparant des estimateurs de Monge aux estimateurs habituels.

- La partie 4 est consacrée aux estimateurs de Monge des lois de Lévy, de Weibull et Log-Logistique à deux paramètres. Nous donnons la forme "explicite" de ces estimateurs et des résultats avec les simulations numériques de ces lois.

- La partie 5 développe la méthode basée sur des fonctions caractérisantes, première et seconde fonction caractéristique et fonction génératrice des moments. Différents estimateurs sont trouvés selon la fonction ou la distance choisie.

- La partie 6 explicite ces procédures pour des exemples comme le mélange de gaussiennes, la loi Gamma ainsi que des simulations numériques pour les lois stables de Lévy.

- Dans la partie 7 nous suggérons quelques autres possibilités et développements.

## 1.2 Quelques notations

Le cadre général de la partie 2 est constitué par les espaces de Köthe  $\mathcal{E}$  de variables aléatoires (v.a.) de norme  $\|\cdot\|$ . Ainsi des v.a.  $U$  et  $V$  de  $\mathcal{E}$  de même loi ont même norme  $\|U\| = \|V\|$ , cf.[16].

Citons comme exemples les espaces d'Orlicz avec leurs normes de Luxemburg ou d'Orlicz, et donc, en particulier, les espaces  $\mathcal{L}^p$ ,  $1 \leq p < +\infty$ , cf.[13]. Pour simplifier nous nous limiterons souvent aux espaces  $\mathcal{L}^1$  et  $\mathcal{L}^2$  par la suite.

L'objectif de la partie 2 est d'améliorer certains estimateurs, principalement ceux d'un paramètre de translation et d'un paramètre d'échelle.

Pour cela nous utilisons les notions de moyenne et de médiane qui sont des paramètres de localisation d'une variable aléatoire réelle. Leurs extensions au cas vectoriel est classique, citons pour les médianes le "survey" [19]. Pour des v.a. à valeurs  $\mathbb{R}^d$  nous utilisons la notion de médiane suivante :

Soit  $\mathcal{E}$  un espace de Köthe formé de v.a. à valeurs  $\mathbb{R}^a$  et  $X \in \mathcal{E}$ , on montre, cf.[10], l'existence d'un point  $m(X) \in \mathbb{R}^d$  tel que

$$\|X - m(X)\|_{\mathcal{E}} \leq \|X - x\|_{\mathcal{E}}, \forall x \in \mathbb{R}^d.$$

Un tel point est, par définition, une *médiane* de  $X$ .

Par exemple, si  $\mathcal{E} = \mathbb{L}^2$ ,  $m(X)$  est l'espérance de  $X$ , notée  $E[X]$  et, si  $d = 1$  et  $\mathcal{E} = \mathbb{L}^1$ ,  $m(X)$  est une médiane au sens usuel.

Il est possible d'étendre l'inégalité définissant la médiane aux v.a. à valeurs dans un espace vectoriel de dimension infinie et même à valeurs dans des espaces métriques cf.[6] et [10]. Ces médianes servent à améliorer des estimateurs.

Considérons une famille  $\{P_\theta\}$ , de probabilités dépendant d'un paramètre de translation, i.e.  $P_\theta(\cdot) = P_0(\cdot - \theta)$ , on écrit aussi plus rapidement  $P$  pour la valeur centrale du paramètre, ici  $P = P_0$ . Soit  $X$  une v.a. de loi  $P_\theta$ , i.e.  $\mathcal{L}(X) = P_\theta$ , alors  $X - \theta$  est de loi  $P$ , cf.[8].

Convenons de réserver les notations  $Id$  pour l'identité et  $Y$  pour la v.a. de loi  $P$ . Ainsi,  $\|f(X)\|_\theta$  ou  $\|f(\cdot)\|_\theta$  est la norme de  $f(X)$  lorsque  $\mathcal{L}(X) = P_\theta$  et  $\|f(Y)\|$  ou  $\|f(\cdot)\|$  lorsque  $\mathcal{L}(Y) = P$ . Par conséquent, en notant  $\triangleq$  la relation servant de définition, pour un paramètre de translation,  $\|f(X)\|_\theta \triangleq \|f(Y - \theta)\|$  ou plus rapidement,  $\|f\|_\theta \triangleq \|f(\cdot - \theta)\|$ .

## 2 Amélioration d'un estimateur

### 2.1 Cas d'un estimateur de translation 1-affine.

Plaçons nous dans le cadre d'une famille de probabilités  $\{P_\theta\}$ ,  $\theta \in \mathbb{R}^d$ , déduites de  $P$  par translation. Pour un point  $x = (x_1, \dots, x_n)$  de  $(\mathbb{R}^d)^n$  et  $\theta \in \mathbb{R}^d$  convenons d'écrire  $x - \theta = (x_1 - \theta, \dots, x_n - \theta)$ .

Un estimateur  $\mathcal{E}$  d'un  $n$ -échantillon est dit *1-linéaire* si  $\mathcal{E}$  est une application linéaire de  $(\mathbb{R}^d)^n$  dans  $\mathbb{R}^d$  telle que  $\mathcal{E}(x - \theta) = \mathcal{E}(x) - \theta$ ,  $\forall \theta \in \mathbb{R}^d$ .

Un estimateur *1-affine* est le translaté (par un point de  $\mathbb{R}^d$ ) d'un estimateur 1-linéaire.

L'exemple standard d'un estimateur 1-linéaire est  $\mathcal{E}(x) = \frac{1}{n} \sum_1^n x_i$ .

Soit  $\mathcal{E}(X)$  un estimateur 1-affine d'un  $n$ -échantillon  $X = (X_1, \dots, X_n)$ ,  $X_i$  i.i.d. de loi commune  $P_\theta$ . Comme  $Y_i = X_i - \theta$  a pour loi  $P$ , le risque de cet estimateur est :

$$R(\theta, \mathcal{E}(X)) \triangleq \|\mathcal{E}(X) - \theta\|_\theta = \|\mathcal{E}(Y)\|.$$

Cette relation montre que la fonction de risque est constante, il est donc particulièrement simple de comparer des estimateurs 1-affines.

Si maintenant  $m = m(\mathcal{E}(Y))$  est une médiane de  $\mathcal{E}(Y)$  pour la norme  $\|\cdot\|$  ( $Y$  est de loi  $P$ ) et  $X$  de loi  $P_\theta$ , alors le risque de l'estimateur affine  $\mathcal{E}^*(X) = \mathcal{E}(X) - m$ , est

$$R(\theta, \mathcal{E}^*(X)) = \|\mathcal{E}(X) - \theta - m\|_\theta = \|\mathcal{E}(Y) - m\| \leq \|\mathcal{E}(Y)\| = \|\mathcal{E}(X) - \theta\|_\theta = R(\theta, \mathcal{E}(X)).$$

Par suite, si  $m$  est non nulle alors  $\mathcal{E}^*(X)$  est meilleur que  $\mathcal{E}(X)$ .

Notons aussi que si  $m_\theta$  est une médiane de  $X$ ,  $\mathcal{L}(X) = P_\theta$ , alors  $m_\theta = m + \theta$ .

Prenons par exemple les translatés d'une loi exponentielle  $\mathcal{E}(1)$ , pour la perte  $\mathbb{L}^1$ , i.e.  $E_\theta[|X - \theta|]$ , l'estimateur de  $\theta$ ,  $X - \log(2)$  est meilleur que  $X$ .

## 2.2 Amélioration locale

Sur  $\mathbb{R}^d$  on considère une norme compatible avec l'ordre habituel. La relation entre la norme et l'ordre d'un espace de Köthe de v.a. exprime que si  $|X| \leq |Z|$  alors

$$\|X\| = \| |X| \| \leq \|Z\|.$$

Pour  $(x, a)$  un couple de points de  $\mathbb{R}^d$  notons  $a.x \triangleq (a_i x_i)_{1 \leq i \leq d}$ . Ainsi, pour  $a$ ,  $0 \leq |a| < 1$  i.e.  $0 \leq |a_i| < 1$  tout  $i$ , on a  $\|a.X\| < \|X\|$ , si  $\|X\| \neq 0$ .

Lorsque  $P_\theta(\cdot) = P(\cdot - \theta)$ ,  $\theta \in \mathbb{R}^d$ , la version locale d'un estimateur 1-affine,  $\mathcal{E}(X)$ , est  $\mathcal{E}^*(X) = a.\mathcal{E}(X) + b$  où  $a$  et  $b$  appartiennent à  $\mathbb{R}^d$ .

Comparons le risque  $R(\theta, \mathcal{E}^*) = \|\mathcal{E}^*(X) - \theta\|_\theta$  de cet estimateur à celui de  $\mathcal{E}(X)$ , qui est  $R(\theta, \mathcal{E}) = \|\mathcal{E}(X) - \theta\|_\theta = \|\mathcal{E}(Y)\|$ .

En notant indifféremment par 1 la constante réelle ou le vecteur de coordonnées égales à 1, il vient :  $R(\theta, \mathcal{E}^*) = \|a.\mathcal{E}(X) - \theta - \theta(1 - a) + b\|_\theta = \|a.\mathcal{E}(Y) - \theta(1 - a) + b\|$ .

**Proposition 2.1** *Soit  $\mathcal{E}$  un estimateur 1-affine de  $\theta$ , pour une famille  $\{P_\theta(\cdot) = P(\cdot - \theta)\}$ . Si  $m$  est une médiane de  $\mathcal{E}$  pour  $P$ , alors l'estimateur  $\mathcal{E}^*(X) = \mathcal{E}(X) - m$  est aussi bon que  $\mathcal{E}(X)$  et meilleur lorsque la médiane est unique et non nulle.*

*De plus, pour tout  $a \in \mathbb{R}^d$ ,  $0 < |a| < 1$  et pour tout  $\theta_0$  il existe un voisinage  $V$  de  $\theta_0$  tel que l'estimateur  $\mathcal{E}^*(X) = a.\mathcal{E}(X) + \theta_0.(1 - a)$  est meilleur que  $\mathcal{E}(X)$  sur  $V$ .*

**Preuve** Soient  $\theta_0$  un point de  $\mathbb{R}^d$ ,  $a$  un point de l'intervalle  $]0, 1[$  de  $\mathbb{R}^d$  et  $b = \theta_0.(1 - a)$ . Le risque de l'estimateur  $\mathcal{E}^*(X) = a.\mathcal{E}(X) + b$  au point  $\theta_0$ , est  $R(\theta_0, \mathcal{E}^*) = \|a.\mathcal{E}(X) + \theta_0 - a\theta_0 - \theta_0\|_{\theta_0}$ . Il s'écrit aussi  $R(\theta_0, \mathcal{E}^*) = \|a.(\mathcal{E}(X) - \theta_0)\|_{\theta_0} = \|a.\mathcal{E}(Y)\| < \|\mathcal{E}(Y)\| = \|\mathcal{E}(X) - \theta_0\|_{\theta_0} = R(\theta_0, \mathcal{E})$ .

Pour une v.a.  $Z$  donnée, l'application  $\theta \rightarrow R(\theta, Z)$  est continue, donc pour tout  $\theta$  appartenant à un voisinage adéquat de  $\theta_0$ , on a  $R(\theta, \mathcal{E}^*) < R(\theta, \mathcal{E})$ .  $\square$

Notons aussi que l'unicité de la médiane est assurée par la stricte convexité de la norme ou, dans le cas de  $\mathbb{L}^1$ , par la stricte croissance de la fonction de répartition de  $P$ .

## 3 Estimateurs de Monge

Comme nous l'avons dit dans l'introduction, il existe relativement peu de méthodes permettant de trouver un estimateur, le transport optimal en donne une nouvelle.

Pour une famille de probabilités  $\{P_\theta\}$ ,  $\theta \in \mathbb{R}^d$  et si l'on observe  $x = (x_1, \dots, x_n)$ , un choix (a priori) de  $\theta$  est donné par  $\text{Arg min}_\theta d(P_\theta, \mu_n)$  où  $\mu_n$  est la mesure empirique du  $n$ -échantillon et  $d(\cdot, \cdot)$  une distance entre les probabilités.

L'application qui à l'observation  $x = (x_1, \dots, x_n)$ , associe un  $\theta$  réalisant  $\text{Arg min}$  précédent est appelé *estimateur de Monge* (pour la distance  $d$ ).

Dans [7] on trouve quelques propriétés comme la consistance de ces estimateurs pour le cas d'un paramètre de translation et celui de l'écart-type.

Dans le but de déterminer l'estimateur de Monge, complétons nos notations.

Pour un  $n$ -échantillon ordonné  $x = (x_{(1)}, \dots, x_{(n)})$ ,  $x_{(k)} < x_{(k+1)}$ , on note  $\mu_n$  la loi empirique et  $(t_{0,n}, \dots, t_{n,n})$ ,  $t_{0,n} = -\infty$ ,  $t_{n,n} = +\infty$ , les points définissant les  $n$ -quantiles de la fonction de répartition  $F$  :  $F(t_{1,n}) = F(t_{2,n}) - F(t_{1,n}) = \dots = F(t_{n,n}) - F(t_{n-1,n}) = \frac{1}{n}$ .

Si  $F$  est continue et pour la norme  $\mathbb{L}^p$ , on sait que  $\text{Arg min}_{\mathcal{L}(X,Z)=(P,\mu_n)} \|X - Z\|_{\mathbb{L}^p}$  est atteint

pour une v.a. de la forme  $Z = Z(X)$ , où  $Z = \sum_1^n x_{(k)} \mathbb{I}_{(t_{k-1,n}, t_{k,n})}$ , cf. [11].

Pour rendre les calculs plus aisés, nous limitons cette étude aux espaces  $\mathbb{L}^1$  et  $\mathbb{L}^2$ .

**Remarque.** Nous ne pouvons pas obtenir des formules générales et explicites dans le cas général. En fait, cette difficulté se retrouve dans la plupart des estimateurs, par exemple pour celui du maximum de vraisemblance •

### 3.1 Cas de deux paramètres : translation et échelle.

Cherchons l'estimateur de Monge pour un  $n$ -échantillon d'une famille  $\{P_{m,\sigma}\}$ ,  $P = P_{0,1}$  où  $P_{m,\sigma}$  se déduit de  $P$  par la transformation  $x \rightarrow \sigma x + m$ .

Nous ne considérons, pour l'instant, que la distance quadratique, le cas de la norme  $\mathbb{L}^1$  i.e. de la valeur absolue est abordé ultérieurement dans la partie 3.6.

En prenant  $Z = \sum_1^n x_{(k)} \mathbb{I}_{(t_{k-1,n}, t_{k,n})}$  comme v.a. de loi  $\mu_n = \frac{1}{n} \sum_1^n \delta_{x_{(k)}}$  par rapport à  $P$ , la distance quadratique de  $P_{m,\sigma}$  à  $\mu_n$  est

$$E_{P_{m,\sigma}} \left[ \left( x - Z \left( \frac{x-m}{\sigma} \right) \right)^2 \right] = E_P \left[ (\sigma x + m - Z(x))^2 \right].$$

Par suite si  $Y$  a pour loi  $P$ , le minimum est atteint pour

$$\bar{m} = E[Z] - \bar{\sigma} E[Y] \quad \text{et} \quad \bar{\sigma} = \frac{1}{E[Y^2]} (E_P[xZ(x)] - \bar{m} E[Y]).$$

Ce qui donne

$$\begin{aligned} \bar{\sigma} &= \frac{1}{\text{var}(Y)} (E_P[xZ(x)] - E[Z]E[Y]) \\ \bar{m} &= E[Z] - \frac{E[Y]}{\text{var}(Y)} (E_P[xZ(x)] - E[Z]E[Y]). \end{aligned}$$

L'estimateur de Monge du couple  $(m, \sigma)$  est le couple formé par :

$$\bar{m}^{\mathcal{M}} = \frac{1}{n} \sum_1^n X_k - \frac{E[Y]}{\text{var}(Y)} \left[ \sum_1^n X_{(k)} \int_{t_{k-1,n}}^{t_{k,n}} x dP(x) - \frac{E[Y]}{n} \sum_1^n X_k \right]$$

$$\bar{\sigma}^{\mathcal{M}} = \frac{1}{\text{var}(Y)} \left[ \sum_1^n X_{(k)} \int_{t_{k-1,n}}^{t_{k,n}} x dP(x) - \frac{E[Y]}{n} \sum_1^n X_k \right].$$

Nous avons supposé que  $E[Y]=0$ ,  $E[Y^2]=1$  ce qui conduit, plus simplement, a :

$$\bar{m}^{\mathcal{M}} = \frac{1}{n} \sum_1^n X_k \quad \text{et} \quad \bar{\sigma}^{\mathcal{M}} = \sum_{k=1}^n X_{(k)} \int_{t_{k-1,n}}^{t_{k,n}} x dP(x). \quad (1)$$

Nous allons restreindre, provisoirement, notre étude au seul cas d'un paramètre d'échelle.

### 3.2 Estimateur "additif" d'un paramètre d'échelle.

Notons  $t > 0$  un paramètre d'échelle,  $P_t$  est la loi définie par  $P_t(A) = P(\frac{1}{t}.A)$ . La multiplication par un nombre réel  $t$  est liée à la variance, l'écart-type ou à un moment de la v.a. Elle intervient aussi dans la normalisation d'une v.a. :  $t = \frac{1}{\|X\|}$ .

Pour la famille  $\{P_t\}$  on a  $d(P_t, \mu_n) = \|x - Z_t(x)\|_t$  où  $Z_t$  est une v.a. adéquate de loi  $\mu_n$  relativement à  $P_t$ , i.e.  $Z_t(P_t) = \mu_n$ , il est clair que  $Z_t(\cdot) = Z_t(\frac{\cdot}{t})$ .

L'estimateur de Monge pour la distance quadratique est obtenu en reprenant le calcul précédent avec un paramètre de translation nul :  $m = 0$ . Soit, pour un  $n$ -échantillon,

$$\mathcal{E}_{Mg}^+ = \mathcal{E}_{Mg}^+(X_1, \dots, X_n) = \frac{1}{E[Y^2]} \sum_1^n X_{(k)} \int_{t_{k-1,n}}^{t_{k,n}} x dP(x). \quad (2)$$

Explicitons cette formule pour deux exemples classiques.

#### 3.2.1 Cas d'une loi gaussienne

Avec une loi  $\mathcal{N}(0, 1)$ , dont la médiane est nulle, l'estimateur de Monge d'un  $n$ -échantillon de  $\mathcal{N}(0, t^2)$  est

$$\mathcal{E}_{Mg}^+(\mathcal{N}) = \mathcal{E}_{Mg}^+(\mathcal{N})(X_1, \dots, X_n) = \frac{1}{\sqrt{2\pi}} \sum_1^n X_{(k)} \left( \exp\left[-\frac{t_{k-1,n}^2}{2}\right] - \exp\left[-\frac{t_{k,n}^2}{2}\right] \right).$$

Les quantiles  $t_{k,n}$  sont tabulés et ainsi  $\mathcal{E}_{Mg}^+(\mathcal{N})$  a une expression simple.

Par exemple pour un 2-échantillon, l'estimateur s'écrit

$$\mathcal{E}_{Mg}^+(\mathcal{N}) = \frac{1}{\sqrt{2\pi}} |X_1 - X_2|.$$

### 3.2.2 Cas des lois uniformes

Prenons une famille  $\{P_t\}$  de lois uniformes sur  $]0, t]$ .

L'estimateur de Monge, dans le cas quadratique, se déduit de l'équation 2 avec  $t_{k,n} = \frac{k}{n}$  :

$$\mathcal{E}_{Mg}^+(\mathcal{U}) = \mathcal{E}_{Mg}^+(\mathcal{U})(X_1, \dots, X_n) = \frac{3}{2n^2} \sum_1^n (2k-1)X_{(k)}.$$

Rappelons que pour un  $n$ -échantillon de la loi uniforme sur  $]0, t]$ , l'estimateur du maximum de vraisemblance (*emv*) est  $Z_{emv} = \max(X_1, \dots, X_n)$ .

Il est clair que l'estimateur de Monge peut donner des valeurs plus grandes que l'*emv* :

Pour un 2-échantillon de v.a. positives :  $\frac{3}{8} [\inf(X_1, X_2) + 3 \sup(X_1, X_2)] > \sup(X_1, X_2)$ .

Enfin lorsque  $n$  est grand,  $\mathcal{E}_{Mg}^+(\mathcal{U})$  se comporte comme  $3 \frac{\sum_1^n kX_{(k)}}{n^2}$  et est consistant en moyenne quadratique.

### 3.3 Risque d'un estimateur d'échelle (multiplicatif).

De manière analogue au cas linéaire, un estimateur  $\mathcal{E}$  est dit *d'échelle ou multiplicatif* si  $\mathcal{E}(tx) = t\mathcal{E}(x)$ ,  $x \in \mathbb{R}^n$  et  $t \in \mathbb{R}_+$ . Les estimateurs de Monge précédents et, bien sûr, celui du maximum de vraisemblance (*emv*) sont des estimateurs multiplicatifs.

Pour un  $n$ -échantillon, le risque d'un estimateur multiplicatif est

$$\begin{aligned} R(t, \mathcal{E}) &= \int_{\mathbb{R}^n} (\mathcal{E}(x) - t)^2 dP_t^{\otimes n}(x) = \int_{\mathbb{R}^n} (\mathcal{E}(tx) - t)^2 dP^{\otimes n}(x) \\ &= t^2 \int_{\mathbb{R}^n} (\mathcal{E}(x) - 1)^2 dP^{\otimes n}(x) = t^2 R(1, \mathcal{E}). \end{aligned}$$

Ainsi la comparaison du risque de tels estimateurs est particulièrement simple, elle se réduit à comparer les risques au point  $t = 1$ .

Si  $\mathcal{E}$  est un estimateur d'échelle,  $\mathcal{E} - a$  perd son caractère multiplicatif et la comparaison des risques devient plus difficile.

#### 3.3.1 Quelques comparaisons des risques.

Comparons le risque de  $\mathcal{E}_{Mg}^+$  à celui d'autres estimateurs d'échelles naturels .

– Cas d'une loi normale :  $\mathcal{N}(0, t^2)$ .

Limitons nous à un 2-échantillon. Comme  $\frac{X_1 - X_2}{\sqrt{2}}$  suit une loi  $\mathcal{N}(0, 1)$ , on en déduit aisément que  $R(1, \mathcal{E}_{Mg}^+(\mathcal{N})) \simeq 0.417994$ .

L'estimateur naturel de l'écart-type d'une loi centrée, l'*emv*,  $Z_{emv} \triangleq \sqrt{\frac{X_1^2 + X_2^2}{2}}$  a pour risque  $R(1, Z_{emv}) = E_{P^{\otimes 2}} \left[ \left( \sqrt{\frac{x^2 + y^2}{2}} - 1 \right)^2 \right] \simeq 0.22754\dots$

\* L'estimateur naturel (*emv*) est meilleur que celui de Monge  $\mathcal{E}_{Mg}^+$ .

On obtient le même résultat avec une loi de Laplace à la place de la gaussienne.

On peut améliorer ces estimateurs en introduisant les médianes.

Comme pour la norme  $\mathbb{L}^2$  la médiane coïncide avec la moyenne, on trouve pour l'*emv* :

$$Z_{emv} = \sqrt{\frac{X_1^2 + X_2^2}{2}}, \quad E[Z_{emv}] = \frac{\sqrt{\pi}}{2} \text{ et, pour } \mathcal{E}_{Mg}^+(\mathcal{N}) = \frac{1}{\sqrt{2\pi}} |X_1 - X_2|, \quad E[\mathcal{E}_{Mg}^+(\mathcal{N})] = \frac{\sqrt{2}}{\pi}.$$

#### - Cas d'une loi uniforme

Soit  $\{P_t\}$  une famille de loi uniforme sur  $]0, t[$ , pour un  $n$ -échantillon, le risque quadratique de l'estimateur de Monge,  $\mathcal{E}_{Mg}^+(\mathcal{U}) = \mathcal{E}_{Mg}^+(\mathcal{U})(X_1, \dots, X_n) = \frac{2}{3n^2} \sum_{k=1}^n (2k-1)X_{(k)}$  est

$$R(\mathcal{E}_{Mg}^+(\mathcal{U})) \triangleq R(1, \mathcal{E}_{Mg}^+(\mathcal{U})) = \int_{[0,1]^n} (\mathcal{E}_{Mg}^+(\mathcal{U})(x) - 1)^2 dx.$$

Ainsi le risque d'un 2-échantillon est :

$$R(\mathcal{E}_{Mg}^+) = \iint_{[0,1]^2} \left( \frac{3}{8} [\inf(x, y) + 3 \sup(x, y)] - 1 \right)^2 dx dy \simeq 0.11718499\dots$$

Tandis que celui du maximum de vraisemblance est :

$$R(Z_{emv}) = \iint_{[0,1]^2} [\sup(x, y) - 1]^2 dx dy = \frac{1}{6}.$$

\* Pour un 2-échantillon, l'estimateur de Monge est meilleur que l'*emv*.

Remarquons que la distance usuelle invariante par translation est donc, a priori, bien adaptée à un paramètre de translation. Alors que pour estimer un paramètre d'échelle, comme l'écart-type, nous sommes en présence d'une multiplication.

Ceci justifie l'introduction du cadre spécifique que nous allons développer.

### 3.4 Existence d'une médiane dans le cas multiplicatif.

Pour un couple  $(x, y)$  de nombres réels strictement positifs posons

$$\ell(x, y) = |\log(x) - \log(y)|.$$

On vérifie sans problème que  $\ell$  est une distance sur  $\mathbb{R}_+$ . Cette distance est invariante par multiplication :  $\ell(ax, ay) = \ell(x, y)$ . De façon naturelle, pour une norme  $\|\cdot\|$ , l'extension aux v.a.r. est  $\|\ell(X, Y)\|$  lorsque les v.a. sont positives.

Pour une probabilité  $P$  telle que  $P(]0, \infty[) = 1$  et  $P_t(\cdot) = P(\frac{\cdot}{t})$  on a  $\|\ell(X, t)\|_t = \|\ell(Y, 1)\|$ , où, bien sûr,  $Y$  a pour loi  $P$ .

Cette relation est analogue à celle du cas "translation" :  $\|X - \theta\|_\theta = \|Y\|$ .

Comme auparavant, on voit que pour une distance invariante par multiplication le risque d'un estimateur d'échelle (multiplicatif) est une constante car  $\mathcal{E}(tX) = t\mathcal{E}(X)$ .

### 3.5 Expression de la médiane multiplicative

L'adaptation de la notion de médiane pour la distance  $\ell$  donne un point  $m^\times = m^\times(X)$  strictement positif tel que :

$$\|\ell(X, m^\times)\| \leq \|\ell(X, x)\|, \quad \forall x > 0.$$

Lorsque la norme est strictement convexe, on a l'unicité de cette médiane. Il en est de même pour  $\mathbb{L}^1$  si, par exemple,  $P$  est à densité strictement positive.

**Proposition 3.1** Soient  $\{P_t\}$  une famille de probabilités où  $t$  est un paramètre d'échelle i.e.  $P_t(A) = P(\frac{1}{t}.A)$ ,  $P(]0, \infty[) = 1$  et  $\ell$  une distance "logarithmique" invariante par multiplication.

Le risque d'un estimateur  $\mathcal{E}$  d'échelle est une constante  $R(t, \mathcal{E}) = \|\ell(\mathcal{E}, t)\|_t = \|\ell(\mathcal{E}, 1)\|$ .

La médiane d'une v.a.  $Y$  de loi  $P$  est la médiane ordinaire pour la norme  $\ell$ - $\mathbb{L}^1$  (la valeur absolue) et  $\exp(E[\log(Y)])$  pour le cas  $\ell$ -quadratique i.e.  $\ell$ - $\mathbb{L}^2$ .

**Preuve** Notons aussi que pour un estimateur  $\mathcal{E}$  d'échelle,  $\frac{1}{m^\times}\mathcal{E}$  est aussi bon que  $\mathcal{E}$  et meilleur si  $m^\times$  est différent de 1. En effet,

$$R(t, \frac{\mathcal{E}}{m^\times}) = \|\ell(\frac{\mathcal{E}}{m^\times}, t)\|_t = \|\ell(\frac{\mathcal{E}}{m^\times}, 1)\|_1 = \|\ell(\mathcal{E}, m^\times)\|_1 \leq \|\ell(\mathcal{E}, 1)\|_1 = \|\ell(\mathcal{E}, t)\|_t = R(t, \mathcal{E}).$$

Déterminons cette médiane, pour les normes de  $\mathbb{L}^1$  et  $\mathbb{L}^2$ , cf [6] pour le cas général.

- Pour la norme  $\mathbb{L}^1$ . Supposons  $P$  à densité  $f$ ,  $P(]0, \infty[) = 1$  et posons

$$\begin{aligned} G(x) &= E_P[\ell(\text{Id}, x)] = \int_0^x \log\left(\frac{x}{t}\right) dP(t) + \int_x^\infty \log\left(\frac{t}{x}\right) dP(t) \\ &= \log(x)P(]0, x]) - \int_0^x \log(t) dP(t) - \log(x)P([x, +\infty[) + \int_x^\infty \log(t) dP(t). \end{aligned}$$

Avec la fonction de répartition  $F$ , la dérivée s'écrit

$$\begin{aligned} G'(x) &= \frac{F(x)}{x} + \log(x).f(x) - \log(x).f(x) - \frac{1 - F(x)}{x} + \log(x).f(x) - \log(x).f(x) \\ &= \frac{2F(x) - 1}{x}. \end{aligned}$$

Ainsi  $G$  atteint son minimum en un point, la médiane,  $m_1^\times = m_1^\times(P)$ , tel que

$$F(m_1^\times) = \frac{1}{2}. \quad (3)$$

Pour une v.a.  $Y$  de fonction de répartition  $F$  la médiane vérifie  $F(m_1^\times(Y)) = \frac{1}{2}$ , ce qui correspond à la médiane classique.

- Pour la norme  $\mathbb{L}^2$ . Le calcul est simple, la dérivée en  $t$  de  $G(t) = E[(\ell(Id, t))^2] = \int_0^\infty (\log(x) - \log(t))^2 dP(x)$  est  $G'(t) = 2 \frac{\log(t)}{t} - 2 \frac{1}{t} \int_0^\infty \log(x) dP(x)$ .

Ainsi, la médiane du cas quadratique est

$$m_2^\times = m_2^\times(P) = \exp \left[ \int_0^\infty \log(x) dP(x) = \exp [E[\log(Y)]] \right] \quad (4)$$

où  $Y$  est une v.a. strictement positive de loi  $P$ .  $\square$

### 3.6 Estimateur de Monge pour un paramètre d'échelle.

Nous recherchons un estimateur de  $t$  par la méthode du transport optimal et pour les coûts  $\ell\text{-}\mathbb{L}^1$  et  $\ell\text{-}\mathbb{L}^2$  i.e. relatifs à la distance  $\ell$ .

Nous devons trouver  $\text{Arg min}_t \left[ \min_{Z(P_t)=\mu_n} E_{P_t} [|\ell(Id, Z)|^p] \right]$  pour  $p=1, 2$  et où  $\mu_n$  est la loi empirique d'un  $n$ -échantillon.

Dans un premier temps, pour  $n$  points  $\{x_1, \dots, x_n\}$ ,  $x_i < x_{i+1}$ , il faut déterminer une v.a.  $Z_t = \sum_1^n x_i \mathbb{1}_{A_t^i}$ , où les  $A_t^i$  forment une partition avec  $P_t(A_t^i) = 1/n$  et telle que

$$Z_t \in \text{Arg min}_{Z(P_t)=\mu_n} \int [\ell(x, Z(x))]^p dP_t(x), \quad \mu_n = \frac{1}{n} \sum_1^n \delta_{x_i}.$$

On sait que  $Z_t$  est croissante donc de la forme  $Z_t = \sum_1^n x_i \mathbb{1}_{I_t^i}$  où les  $I_t^i$  sont des intervalles.

Ainsi un estimateur de Monge d'un paramètre d'échelle est un point  $\mathcal{E}^\times$  tel que

$$E_{\mathcal{E}^\times} [(\ell(Id, Z_{\mathcal{E}^\times}))^p] \leq \min_t \{ E_t [(\ell(Id, Z_t))^p], Z_{\mathcal{E}^\times}(P_{\mathcal{E}^\times}) = Z_t(P_t) = \mu \}.$$

La v.a.  $Z_{\mathcal{E}^\times}$  étant croissante et en outre le caractère multiplicatif permet d'écrire

$E_{P_t}[h(x)] = E_P[h(tx)]$ ,  $f_t(x) = \frac{1}{t} f\left(\frac{x}{t}\right)$ ,  $F_t(x) = F\left(\frac{x}{t}\right)$ . Par suite, si  $m_t^{k,n}$  est le  $k^{\text{ème}}$

quantile de  $P_t$ , i.e.  $F_t(m_t^{k,n}) = \frac{k}{n} = F\left(\frac{m_t^{k,n}}{t}\right)$ , on en déduit que  $m_t^{k,n} = tm_1^{k,n} = tm^{k,n}$ .

Nous notons  $I^{k,n}$  l'intervalle  $(m^{k-1,n}, m^{k,n})$  avec les conventions  $m^{0,n} = 0$  et  $m^{n,n} = 1$ .

### 3.6.1 Cas $\ell$ -quadratique.

Le carré de la distance de  $P_t$  à la loi empirique d'un  $n$ -échantillon réordonné,  $\{x_{(k)}\}$ , est :

$$\begin{aligned} \sum_1^n E_{P_t} \left[ (\log(x) - \log(x_{(k)}))^2 \mathbb{I}_{I^{k,n}}(x) \right] &= \sum_1^n E_P \left[ \left( \log(x) + \log\left(\frac{t}{x_{(k)}}\right) \right)^2 \mathbb{I}_{I^{k,n}}(x) \right] \\ &= \int \log^2(x) dP + \sum_1^n 2 \log\left(\frac{t}{x_{(k)}}\right) \int_{I^{k,n}} \log(x) dP + \frac{1}{n} \sum_1^k \log^2\left(\frac{t}{x_{(k)}}\right) \end{aligned}$$

dont la dérivée en  $t$ ,  $2 \frac{1}{t} E_P[\log(Id)] + 2 \frac{1}{t} \frac{1}{n} \sum_1^n \log\left(\frac{t}{x_{(k)}}\right)$ , s'annule pour

$$\log(t) = -E_{P_1}[\log(x)] + \frac{1}{n} \sum_1^n \log(x_k).$$

Ainsi pour une v.a.  $Y$  de loi  $P$  l'estimateur de Monge d'échelle est

$$\mathcal{E}_{2,Mg}^\times \triangleq \mathcal{E}_{2,Mg}^\times(X_1, \dots, X_n) = \left( \prod_1^k X_k \right)^{\frac{1}{n}} \exp(-E[Y]). \quad (5)$$

On vérifie, sans peine, que la médiane (pour  $\mathbb{L}^2$ ) de cet estimateur est  $m_2^\times(\mathcal{E}_{2,Mg}^\times) = 1$ . Ce qui est conforme avec l'équation 4.

Explicitons avec nos deux exemples classiques.

- Pour  $P$  de loi uniforme sur  $]0, 1]$ ,  $E_P[\log(x)] = \int_0^1 \log(x) dx = -1$ , d'où

$$\mathcal{E}_{2,Mg}^\times(\mathcal{U}) = \exp[1] \sqrt[n]{X_1 \dots X_n}$$

- Pour une famille  $\{P_t\}$  où  $P_t$  est la valeur absolue d'une loi normale  $\mathcal{N}(0, t^2)$ , i.e.  $P$  a pour densité  $f(x) = \sqrt{\frac{2}{\pi}} \exp\left(-\frac{x^2}{2}\right) \mathbb{I}_{\mathbb{R}_+}(x)$ . L'estimateur déduit d'un  $n$ -échantillon et  $Y$  de loi  $P$  est

$$\mathcal{E}_{2,Mg}^\times(|\mathcal{N}|) = \exp[-E[\log(Y)]] \sqrt[n]{X_1 \dots X_n}.$$

Numériquement on a  $\exp[-E_P[\log(x)]] \simeq 1.88736$ .

On remarque que l'estimateur  $\mathcal{E}_{2,Mg}^\times(|\mathcal{N}|)(X_1 \dots X_n)$  est meilleur que  $\sqrt[n]{X_1 \dots X_n}$ .

La médiane ordinaire est solution de  $\int_0^m f(x) dx = \frac{1}{2}$  i.e.  $m_1^\times(\mathcal{E}_{2,Mg}^\times(|\mathcal{N}|))$ , est le 3<sup>ème</sup> fractile de la loi  $\mathcal{N}(0, 1)$ , soit approximativement 0.67448...

Pour des calculs complémentaires, on peut se référer à [20] qui donne la loi de  $X_1 X_2$ .

### 3.6.2 Cas de la $\ell$ -valeur absolue.

Pour une probabilité à densité, on utilise la relation :

$$\left( \int_m^{\frac{a}{t}} |\log[x] - \log[a/t]| dP(x) \right)' = \left( \int_{\frac{a}{t}}^m |\log[x] - \log[a/t]| dP(x) \right)' = \frac{1}{t} [F(m) - F(a/t)].$$

- Dans le cas d'un 2-échantillon et en prenant pour  $m$  la médiane, on trouve que l'estimateur de Monge :

$$\begin{aligned} \text{Arg min}_t \left[ \int_0^{mt} |\log(x/x_{(1)})| dP_t + \int_{mt}^{\infty} |\log(x/x_{(2)})| dP_t \right] = \\ \text{Arg min}_t \left[ \int_0^m |\log(x) - \log(x_{(1)}/t)| dP + \int_m^{\infty} |\log(x) - \log(x_{(2)}/t)| dP \right] \end{aligned}$$

est atteint pour  $\mathcal{E}_{1,Mg}^\times$ , solution en  $t$  de l'équation

$$F\left[\frac{X_1}{t}\right] + F\left[\frac{X_2}{t}\right] = 1. \quad (6)$$

- Pour un 3-échantillon un calcul analogue donne :

$$2\left(F\left[\frac{X_1}{t}\right] + F\left[\frac{X_2}{t}\right] + F\left[\frac{X_3}{t}\right]\right) = 3. \quad (7)$$

Ces équations admettent une unique solution,  $\mathcal{E}_{1,Mg}^\times$ , lorsque  $P$  est à densité strictement positive sur  $\mathbb{R}_+$ .

La résolution des deux précédentes équations est simple lorsque  $F$  est la fonction de répartition de la loi uniforme sur  $(0, 1)$ . Un 2-échantillon donne  $\mathcal{E}_{1,Mg}^\times(\mathcal{U}) = X_1 + X_2$ .

Et pour un 3-échantillon on obtient,  $\mathcal{E}_{1,Mg}^\times(\mathcal{U}) = \frac{2}{3}(X_1 + X_2 + X_3)$

On voit que la médiane (ordinaire)  $m_1^\times$  de  $\mathcal{E}_{1,Mg}^\times(\mathcal{U})$  est 1 et ceci pour un 2 ou 3-échantillon.

Nous n'abordons pas les tailles supérieures de l'échantillon, elles conduisent à des expressions plus complexes où l'ordre intervient.

### 3.7 $\ell$ -Risques des estimateurs de Monge d'échelles.

Le risque  $\ell$ -quadratique de l'estimateur de Monge  $\mathcal{E}_{2,Mg}^\times = \exp[-E[\log(Y)]] \sqrt[n]{X_1 \cdots X_n}$  où  $Y$  est de loi  $P$ , est  $R_2(1, \mathcal{E}_{2,Mg}^\times) = E_P[(\log[\mathcal{E}_{2,Mg}^\times])^2]$ . En développant il vient

$$R_2(1, \mathcal{E}_{2,Mg}^\times) = (E[\log(Y)])^2 - 2(E[\log(Y)])^2 + \frac{n-1}{n}(E[\log(Y)])^2 + \frac{1}{n}E[\log^2(Y)].$$

Ce qui se simplifie en :

$$R_2(\mathcal{E}_{2,Mg}^\times) \triangleq R_2(1, \mathcal{E}_{2,Mg}^\times) = \frac{1}{n} [E[\log^2(Y)] - (E[\log(Y)])^2] = \frac{1}{n} \text{var}(\log(Y)). \quad (8)$$

Explicitons avec nos exemples.

### 3.7.1 Loi uniforme

– Risques associés à  $\mathcal{E}_{2,Mg}^\times(\mathcal{U})$ .

Pour une famille  $P_t$  de loi uniforme sur  $]0, t]$  et  $Y$  de loi  $P$  on a  $\text{var}(\log(Y)) = 1$ .

Le risque d'un  $n$ -échantillon est

$$R_2(\mathcal{E}_{2,Mg}^\times(\mathcal{U})) = \frac{1}{n}.$$

Le risque de l'estimateur du maximum de vraisemblance,  $\mathcal{E}_{emv}$ , est

$$R_2(\mathcal{E}_{emv}) = \int_{[0,1]^n} (\log(\max[x_1, \dots, x_n]))^2 dx_1 \cdots dx_n = n \int_0^1 (\log(x))^2 x^{n-1} dx = \frac{2}{n^2}.$$

\* Pour la perte quadratique le risque de l'emv est inférieur à celui de Monge  $\mathcal{E}_{2,Mg}^\times(\mathcal{U})$ .

Toujours avec  $\mathcal{E}_{2,Mg}^\times(\mathcal{U})$ , calculons son risque pour perte la valeur absolue (norme  $\mathbb{L}^1$ ).

Le risque  $R_1$  de l'estimateur précédent  $\mathcal{E}_{2,Mg}^\times(\mathcal{U}) = \exp[1] \sqrt[n]{X_1 \cdots X_n}$  est

$$\begin{aligned} R_1(\mathcal{E}_{2,Mg}^\times(\mathcal{U})) &= E_P\left[\left|1 + \frac{1}{n} \sum_1^n \log(X_i)\right|\right] = \frac{1}{(n-1)!} \int_0^1 \left|1 + \frac{1}{n} \log(x)\right| \left(\log\left(\frac{1}{x}\right)\right)^{n-1} dx \\ &= \frac{2}{(n-1)!} \int_{e^{-n}}^1 \left(1 + \frac{1}{n} \log(x)\right) \left(\log\left(\frac{1}{x}\right)\right)^{n-1} dx = \frac{-2}{(n-1)!} \int_0^{e^{-n}} \left(1 + \frac{1}{n} \log(x)\right) \left(\log\left(\frac{1}{x}\right)\right)^{n-1} dx. \end{aligned}$$

Alors que le risque de l'emv est  $R_1(\mathcal{E}_{emv}(\mathcal{U})) = n \int_0^1 |\log(x)| x^{n-1} dx = \frac{1}{n}$ .

L'évaluation numérique de ces expressions conduit à :

\* Pour la perte  $\mathbb{L}^1$  le risque de l'estimateur emv est inférieur à celui de Monge  $\mathcal{E}_{2,Mg}^\times(\mathcal{U})$ .

On obtient le même résultat en remplaçant  $\exp[-E[\log(Y)]]$  par la médiane relative à

$\mathbb{L}^1$ . L'estimateur devenant  $\frac{\mathcal{E}_{2,Mg}^\times}{m_{1,n}}$  où  $m_{1,n}$  est la médiane i.e.  $P(X_1 \cdots X_n \leq m_{1,n}) = 0.5$ .

– Risques associés à  $\mathcal{E}_{1,Mg}^\times(\mathcal{U})$ .

Ici, comme vu précédemment, nous ne considérons que des échantillons de taille 2 ou 3.

Les risques sont donnés dans le tableau suivant où  $R_i(\mathcal{E}) = R_i(1, \mathcal{E})$ .

$$R_1(\mathcal{E}_{1,Mg}^\times(\mathcal{U})(X_1, X_2)) = E_P[|\log(X_1 + X_2)|] \simeq 0.386294$$

$$R_2(\mathcal{E}_{1,Mg}^\times(\mathcal{U})(X_1, X_2)) = E_P[(\log(X_1 + X_2))^2] \simeq 0.302023$$

$$R_1(\mathcal{E}_{emv} = \sup(X_1, X_2)) = 0.5$$

$$R_1(\mathcal{E}_{1,Mg}^\times(\mathcal{U})(X_1, X_2, X_3)) = E_P\left[\left|\log\left(\frac{2}{3}(X_1 + X_2 + X_3)\right)\right|\right] \simeq 0.296167$$

$$R_2(\mathcal{E}_{1,Mg}^\times(\mathcal{U})(X_1, X_2, X_3)) = E_P\left[\left(\log\left(\frac{2}{3}(X_1 + X_2 + X_3)\right)\right)^2\right] \simeq 0.161179$$

$$R_1(\mathcal{E}_{emv} = \sup(X_1, X_2, X_3)) = 0.33333\dots$$

\* Le risque de l'estimateur de Monge  $\mathcal{E}_{1,Mg}^{\times}(\mathcal{U})$  est inférieur à celui de l'emv, tant pour la perte quadratique que pour la valeur absolue.

### 3.7.2 Cas de la valeur absolue d'une gaussienne, $|\mathcal{N}|$ .

Comparons le risque de Monge, d'un 2-échantillon et celui de l'estimateur naturel :

\* Le risque de l'estimateur de Monge est inférieur à celui de l'estimateur empirique.

En effet, si  $m = \sqrt{\frac{2}{\pi}} \int_0^{\infty} \log(x) \exp(-\frac{x^2}{2}) dx$ , le risque quadratique de l'estimateur de Monge d'un  $n$ -échantillon de la valeur absolue d'une gaussienne est, avec l'équation 8,

$$R_2(1, \mathcal{E}_{2,Mg}^{\times}(|\mathcal{N}|)) = \frac{1}{n} \sqrt{\frac{2}{\pi}} \int_0^{\infty} (\log(x) - m)^2 \exp(-\frac{x^2}{2}) dx.$$

Soit numériquement,  $R_2(\mathcal{E}_{2,Mg}^{\times}(|\mathcal{N}|)) \simeq \frac{1.2337}{n}$ . Pour  $n=2$ , le risque est  $\simeq 0.61685$ .

Toujours pour un 2-échantillon, le risque de l'estimateur empirique,  $\sqrt{Z_{nat}}$  obtenu à partir de l'estimation de l'écart-type  $Z_{nat} = (X_1 - \frac{X_1 + X_2}{2})^2 + (X_2 - \frac{X_1 + X_2}{2})^2$ , i.e.  $Z_{nat} = \frac{1}{2}(X_1 - X_2)^2$ . D'où  $\sqrt{Z_{nat}} = \frac{|X_1 - X_2|}{\sqrt{2}}$  avec  $X_i = |Y_i|$ , les deux v.a.  $Y_i$  étant indépendantes de même loi  $\mathcal{N}(0, 1)$ . Le risque quadratique de cet estimateur est

$$R_2(\sqrt{Z_{nat}}) = E_{P_t}[(\ell(\sqrt{Z_{nat}}, t))^2] = E_P[(\log(\sqrt{Z_{nat}}))^2] \simeq 2.58008..$$

Tandis que pour la valeur absolue,  $\mathbb{L}^1$ , on a, pour un 2-échantillon,

$$R_1(\mathcal{E}_{2,Mg}^{\times}(|\mathcal{N}|)) = \frac{1}{2} \sqrt{\frac{2}{\pi}} \int_0^{\infty} |\log(x) - m| \exp(-\frac{x^2}{2}) dx \simeq 0.83264$$

$$\text{et } R_1(\sqrt{Z_{nat}}) = E_P[|\log(\sqrt{Z_{nat}})|] \simeq 1.26563$$

Ce qui montre notre assertion.

L'estimateur du maximum de vraisemblance pour un  $n$ -échantillon est  $Z_{emv} = \sqrt{\frac{\sum X_i^2}{n}}$ . La loi de  $\sum X_i^2$  est celle d'un  $\chi^2(n)$  et comme  $Z_{emv}$  est multiplicatif son risque est aisé à calculer.

Par exemple pour un 2-échantillon,  $R_2(Z_{emv}) \simeq 0.494528$  et  $R_1(Z_{emv}) \simeq 1.01598$

\* Pour la perte quadratique le risque de l'emv est inférieur à celui de l'estimateur de Monge  $\mathcal{E}_{2,Mg}^{\times}(|\mathcal{N}|)$  et c'est le contraire pour la perte valeur absolue.

Toujours à cause de la complexité des équations 6 et 7, le calcul des risques associés à  $\mathcal{E}_{1,Mg}^{\times}(|\mathcal{N}|)$  ne peut être que numérique et sur des exemples, même pour des 2 ou 3-échantillons.

## 4 Estimateurs de Monge de quelques exemples classiques.

Sortant des exemples simples de lois uniformes ou gaussiennes, nous allons déterminer les estimateurs de Monge pour les paramètres de différentes lois. L'intérêt de nos estimateurs réside, outre leurs existences, en leur simplicité.

### 4.1 Estimation pour la loi de Weibull

Commençons par chercher la distance de Monge-Wasserstein entre la loi de Weibull et un  $n$ -échantillon de cette loi. Pour cela nous utilisons les  $n$ -quantiles  $\{t_{k,n}\}$  de la loi exponentielle en convenant que  $t_{0,n} = 0$  et  $t_{n,n} = \infty$ , ils sont déterminés par  $\int_0^{t_{k,n}} e^{-x} dx = \frac{k}{n}$ , i.e.

$$t_{k,n} = \log\left(\frac{n}{n-k}\right), \quad 0 \leq k \leq n.$$

La loi de Weibull à 2 paramètres  $P_{(\theta,\alpha)}$  a pour densité  $f_{(\theta,\alpha)} = \theta \cdot \alpha \cdot x^{\alpha-1} \exp(-\theta x^\alpha) \mathbb{1}_{\mathbb{R}_+}(x)$ .

Elle se déduit de la loi exponentielle  $\mathcal{E}(1)$ , ici  $P$ , par le changement de variable  $x \rightarrow \left(\frac{x}{\theta}\right)^{\frac{1}{\alpha}}$ .

Pour un  $n$ -échantillon réordonné  $\{x_k\}$ , la distance  $\mathbb{L}^p$  de  $P_{(\theta,\alpha)}$  à  $\mu_n$  est,

$$\text{Arg min}_{(\theta,\alpha)} \sum_1^n \int_{I_{k,n}(\theta,\alpha)} |x - x_k|^p f_{(\theta,\alpha)}(x) dx$$

où  $I_{k,n}(\theta, \alpha)$  est l'intervalle  $(a_{k-1}(\theta, \alpha), a_k(\theta, \alpha))$ ,  $a_0 = 0, a_n = \infty$ , et  $P_{(\theta,\alpha)}(I_{k,n}(\theta, \alpha)) = \frac{1}{n}$ .

Comme  $t_{k,n} = \left(\frac{a_k(\theta, \alpha)}{\theta}\right)^{\frac{1}{\alpha}}$ , cette distance s'écrit

$$\text{Arg min}_{(\theta,\alpha)} \sum_1^n \int_{t_{k-1,n}}^{t_{k,n}} \left| \left(\frac{x}{\theta}\right)^{\frac{1}{\alpha}} - x_k \right|^p \exp(-x) dx.$$

- Pour la distance quadratique i.e.  $p = 2$ , on doit trouver

$$\text{Arg min}_{(\theta,\alpha)} \left[ F(\theta, \alpha) = \int \left(\frac{x}{\theta}\right)^{\frac{2}{\alpha}} e^{-x} dx - 2 \sum_1^n \int_{t_{k-1,n}}^{t_{k,n}} \left(\frac{x}{\theta}\right)^{\frac{1}{\alpha}} x_k e^{-x} dx + \frac{1}{n} \sum x_k^2 \right].$$

Pour simplifier, posons  $v = \frac{1}{\theta}$  et  $u = \frac{1}{\alpha}$ , l'expression devient à une constante près

$$F(v, u) = v^u \left[ \int_0^\infty v^u x^{2u} e^{-x} dx - 2 \sum_1^n x_k \int_{t_{k-1,n}}^{t_{k,n}} x^u e^{-x} dx \right].$$

La dérivée en  $v$  s'annule pour

$$v^u \int_0^\infty x^{2u} e^{-x} dx = \sum_1^n x_k \int_{t_{k-1,n}}^{t_{k,n}} x^u e^{-x} dx \quad \text{et donc}$$

$$\theta = \left( \frac{\int_0^\infty x^{2u} e^{-x} dx}{\sum_1^n x_k \int_{t_{k-1,n}}^{t_{k,n}} x^u e^{-x} dx} \right)^\alpha.$$

Tandis que la dérivée en  $u$  s'annule pour  $u$  solution de :

$$v^u \int_0^\infty x^{2u} \cdot \log(vx) \cdot e^{-x} dx = \sum_1^n x_k \cdot \int_{t_{k-1,n}}^{t_{k,n}} x^u \cdot \log(vx) \cdot e^{-x} dx.$$

Si  $\alpha$  est connu, le paramètre  $\theta$  est aisément identifiable. Néanmoins nous ne poursuivrons pas cette voie.

– Cherchons un estimateur de Monge avec la méthode du coût logarithmique.

La distance  $\ell\text{-}\mathbb{L}^p$  de  $P_{(\theta,\alpha)}$  à  $\mu_n$  est

$$\begin{aligned} \ell^p(P_{(\theta,\alpha)}, \mu_n) &\triangleq \text{Arg min}_{(\theta,\alpha)} \sum_1^n \int_{I_{k,n}(\theta,\alpha)} |\log(x) - \log(x_k)|^p f_{(\theta,\alpha)}(x) dx \\ &= \text{Arg min}_{(\theta,\alpha)} \left\{ F^\ell(\theta, \alpha) = \sum_1^n \int_{t_{k-1,n}}^{t_{k,n}} \left| \frac{1}{\alpha} \log\left(\frac{x}{\theta}\right) - \log(x_k) \right|^p dP \right\} \end{aligned}$$

Pour  $p = 2$  et en notant  $Y$  une v.a. de loi  $\mathcal{E}(1)$ ,

$$F^\ell(\theta, \alpha) = \frac{1}{\alpha^2} E[\log^2\left(\frac{Y}{\theta}\right)] - 2 \sum_1^n \frac{1}{\alpha} \log(x_k) \int_{t_{k-1,n}}^{t_{k,n}} \log\left(\frac{x}{\theta}\right) dP + \frac{1}{n} \sum \log^2(x_k).$$

Simplifions les écritures en posant  $v = \frac{1}{\theta}$  et  $u = \frac{1}{\alpha}$ ,

$$F^\ell(v, u) = u^2 E[\log^2(vY)] - 2u \sum_1^n \log(x_k) \int_{t_{k-1,n}}^{t_{k,n}} \log(vx) dP + \frac{1}{n} \sum \log^2(x_k).$$

En développant et à une constant près,

$$\begin{aligned} F^\ell(v, u) &= u^2 [\log^2(v) + E[\log^2(Y)] + 2 \log(v) E[\log(Y)]] \\ &\quad - 2u \log(v) \frac{1}{n} \sum_1^n \log(x_k) - 2u \sum_1^n \log(x_k) \int_{t_{k-1,n}}^{t_{k,n}} \log(x) e^{-x} dx. \end{aligned}$$

Désignons par  $\{X_{(k)}\}$ ,  $k=1 \dots n$  le  $n$ -échantillon réordonné et posons  $M_n = \frac{1}{n} \sum_1^n \log(X_k)$ .

La dérivée en  $v$  s'annule pour  $\log(v) = \frac{1}{u} M_n - E[\log(Y)]$  où  $Y$  est de loi  $\mathcal{E}(1)$ ,

Et, en écrivant  $A_n = \sum_1^n \log(X_{(k)}) \int_{t_{k-1,n}}^{t_{k,n}} \log(x) dP$ , la dérivée en  $u$  s'annule pour

$$u(\log^2(v) + E[\log^2(Y)] + 2\log(v)E[\log(Y)]) = M_n \log(v) + A_n.$$

En définitive l'estimateur de Monge  $(\mathcal{E}_{\log(\theta)}^\times(W), \mathcal{E}_\alpha^\times(W))$  du couple  $(\log(\theta), \alpha)$  est :

$$\mathcal{E}_{\log(\theta)}^\times(W) \triangleq \frac{M_n E[\log^2(Y)] - A_n E[\log(Y)]}{M_n E[\log(Y)] - A_n} \quad (9)$$

$$\mathcal{E}_\alpha^\times(W) \triangleq \frac{\text{var}(\log(Y))}{A_n - M_n E[\log(Y)]} \quad (10)$$

L'estimateur  $(\mathcal{E}_\theta^\times(W))$  se déduit en prenant  $\exp[(\mathcal{E}_{\log(\theta)}^\times(W))]$ .

Ces estimateurs sont aisés à trouver, ils n'exigent pas les calculs complexes de l'emv ni ceux de l'estimateur de Monge du cas "purement" quadratique vu ci-dessus.

Par exemple, simulons une répétition de 10 fois un 1000-échantillon d'une loi de Weibull de paramètres  $\theta = 2$  et  $\log(\alpha) = 3$  = 1.09861. Nous obtenons pour le couple  $(\mathcal{E}_\theta^\times, \mathcal{E}_{\log(\alpha)}^\times)$  respectivement comme moyenne et variance  $\simeq (2.00315; 1.10218)$  et  $\simeq (0.00010; 0.00006)$ .

Pour une répétition de 10 fois d'un 2000-échantillon de paramètres  $\theta = 3$  et  $\alpha = 2$ . Nous avons, pour le couple  $(\mathcal{E}_\theta^\times, \mathcal{E}_\alpha^\times)$ , respectivement comme moyenne et variance :  $\simeq (3.01634; 2.00562)$  et  $\simeq (0.000786; 0.00959)$ .

Le tableau suivant est inspiré de celui de [3] où l'auteur compare diverses méthodes classiques d'estimation. Comme on ne considère qu'un échantillon de taille  $n$  et que les vraies valeurs sont connues l'auteur adopte le critère  $\text{TD} = \left| \frac{\text{est}(\beta) - \beta}{\beta} \right| + \left| \frac{\text{est}(\log(\alpha)) - \log(\alpha)}{\log(\alpha)} \right|$  que nous calculons. Nous reprenons les mêmes paramètres mais ne disposant pas des valeurs numériques de son échantillon nous ne pouvons comparaître les résultats.

$\beta$	$\alpha    \log(\alpha)$	n	est( $\beta$ )	est ( $\alpha$ )    est(log[ $\alpha$ ])	TD
1	10    2.30259	20	1.11238	9.73921    2.27616	0.139161
		40	0.906592	8.7161    2.16517	0.24071
		60	1.03047	13.142    2.57582	0.269558
		80	1.01877	10.6915    2.36945	0.0834427
		100	1.01473	10.4259    2.3443	0.0555806
2.3	145    4.97673	20	2.36728	156.365    5.05219	0.101935
		40	2.20687	125.693    4.833849	0.194098
		60	2.27347	139.206    4.93595	0.0531595
		80	2.25354	129.604    4.86448	0.138994
		100	2.29179	142.235    4.95748	0.0230073
2.9	357    5.87774	20	2.77849	277.924    5.62735	0.326423
		40	2.88124	342.243    5.83552	0.0495869
		60	2.93388	369.503    5.91216	0.0455178
		80	2.87977	336.437    5.81841	0.0680947
		100	2.89319	344.818    5.84302	0.0376774
3.5	1270    7.14677	20	3.77249	1652.4    7.40998	0.309276
		40	3.3999	933.004    6.83841	0.389794
		60	3.38391	1005.21    6.91296	0.296581
		80	3.56179	1409.54    7.25102	0.116652
		100	3.44845	1101.64    7.00456	0.167555
1.9	872    6.77079	20	1.99449	1058.83    6.96492	0.22618
		40	1.86172	770.15    6.6466	0.152384
		60	1.96415	1050.56    6.95708	0.203728
		80	1.91845	911.257    6.81483	0.0527917
		100	1.89273	835.541    6.72808	0.0474641

De même nous ne cherchons pas de comparaisons avec les méthodes classiques, cf.[17], dont les risques sont très difficilement calculables.

## 4.2 Loi de Lévy

La densité d'une loi stable d'ordre  $\frac{1}{2}$  de Lévy de paramètre  $\theta$  -voir aussi la partie 6.4- est

$$f_{\theta}(x) = \sqrt{\frac{\theta}{2\pi}} \frac{\exp\left[-\frac{\theta}{2x}\right]}{x^{3/2}} \mathbb{1}_{\mathbb{R}_+}(x).$$

Rappelons aussi que la probabilité associée  $P_{\theta}$  ne possède pas de moyenne, ce qui limite les méthodes d'estimation du paramètre  $\theta$ .

Ainsi nous ne pouvons pas calculer le risque, soit en moyenne quadratique soit en valeur absolue, de l'estimateur du maximum de vraisemblance,  $Z_{emv} = \frac{n}{\sum_1^n \frac{1}{X_k}}$ .

Reste la possibilité de choisir une fonction de perte (ou de coût) logarithmique.

Pour trouver des estimateurs, utilisons la méthode du transport optimal.

Notons  $P$  la probabilité de densité  $f_1$  et remarquons que  $P_\theta$  se déduit de  $P$  par  $x \rightarrow \theta x$ .

Reprenons les notations habituelles,  $Y$  v.a. de loi  $P$ ,  $\mu_n$  la loi empirique,  $Z_\theta$  v.a. croissante,  $Z_\theta(P_\theta) = \mu_n$ ,  $X_\theta$  une v.a. de loi  $P_\theta$  et  $(0, t_{k,n,\theta})$  l'intervalle tel que  $P_\theta(0, t_{k,n,\theta}) = \frac{k}{n}$ , avec les conventions  $t_{0,n,\theta} = 0$  et  $t_{n,n,\theta} = \infty$ . Pour  $\theta=1$  on écrit plus simplement  $t_{k,n}$ .

La distance  $\ell$ -quadratique de  $P_\theta$  à  $\mu_n$  est

$$\ell^2(P_\theta, \mu_n) \triangleq E[\ell^2(X_\theta, Z_\theta)] = \int_0^\infty (\log(x) - \log(x_{(k)}))^2 \mathbb{I}_{(t_{k-1,n,\theta}, t_{k,n,\theta})}(x) f_\theta(x) dx.$$

En faisant le changement de variable  $x \rightarrow x/\theta$  on arrive à

$$\begin{aligned} \ell^2(P_\theta, \mu_n) &= \sum_1^n \int_{t_{k-1,n}}^{t_{k,n}} (\log(x\theta) - \log(x_{(k)}))^2 f_1(x) dx. \\ &= E[\log^2(Y)] + \log^2(\theta) + 2\log(\theta)E[\log(Y)] + \frac{1}{n} \sum_1^n \log^2(x_k) \\ &\quad - 2 \sum_1^n \log(x_{(k)}) \int_{t_{k-1,n}}^{t_{k,n}} \log(x) f_1(x) dx - 2 \frac{\log(\theta)}{n} \sum_1^n \log(x_k). \end{aligned}$$

Dont la dérivée en  $\theta$  s'annule pour  $\theta$  solution de

$$\log(\theta) + E[\log(Y)] = \frac{1}{n} \sum_1^n \log(x_k).$$

On en déduit l'estimateur de Monge :

$$\mathcal{E}_{2,Mg}^\times(\mathcal{L})(X_1, \dots, X_n) = \exp \left[ \frac{1}{n} \sum_1^n \log(X_k) - E[\log(Y)] \right]. \quad (11)$$

Remarquons que les estimateurs  $Z_{emv}$  et  $\mathcal{E}_{2,Mg}^\times(\mathcal{L})$  sont des estimateurs d'échelles i.e. multiplicatifs.

Comparons leur risque avec la fonction de perte  $\ell$ -quadratique :

$$\begin{aligned} R_2(\theta, Z_{emv}) &= R_2(1, Z_{emv}) = \int_{\mathbb{R}_+^n} \log^2 \left( \frac{1}{n} \sum_1^n \frac{1}{x_k} \right) \prod_1^n f_1(x_k) dx_1 \cdots dx_n \\ R_2(\theta, \mathcal{E}_{2,Mg}^\times(\mathcal{L})) &= R_2(1, \mathcal{E}_{2,Mg}^\times(\mathcal{L})) = \int_{\mathbb{R}_+^n} \left( \frac{1}{n} \sum_1^n \log(x_k) - E[\log(Y)] \right)^2 \prod_1^n f_1(x_k) dx_1 \cdots dx_n \end{aligned}$$

Ce qui nous permet d'affirmer que

\* *L'estimateur de Monge est meilleur que l'emv pour la perte  $\ell$ -quadratique.*

### 4.3 Loi Log-Logistique

Comme dernier exemple abordons la loi Log-Logistique  $P_{\alpha,\beta}$  où  $\alpha$  et  $\beta$  sont deux paramètres strictement positifs. Sa fonction de répartition est  $F_{\alpha,\beta}(x) = \frac{x^\beta}{\alpha^\beta + x^\beta} \mathbb{1}_{\mathbb{R}_+}(x)$  et sa densité  $f_{\alpha,\beta}(x)$ .

La loi  $P_{\alpha,\beta}$  se déduit de la probabilité  $P = P_{1,1}$  par le changement de variable  $y = \alpha x^{\frac{1}{\beta}}$ .

L'emv est solution de  $\sup_{\alpha,\beta} \left\{ n \log\left(\frac{\beta}{\alpha}\right) + (\beta - 1) \sum_1^n \log\left(\frac{x_k}{\alpha}\right) - 2 \sum_1^n \log\left(1 + \left(\frac{x_k}{\alpha}\right)^\beta\right) \right\}$

On obtient un système de relations dont la résolution en  $\alpha$  et  $\beta$  est loin d'être simple.

L'estimateur de Monge quadratique est obtenu comme minimum de

$$\alpha^2 E_P \left[ x^{\frac{2}{\beta}} \right] + \frac{1}{n} \sum_1^n x_{(k)}^2 - 2\alpha \sum_1^n x_{(k)} \int_{t_{k-1,n}}^{t_{k,n}} x^{\frac{1}{\beta}} dP(x)$$

La dérivée en  $\alpha$  de cette expression s'annule pour

$$\alpha = \frac{\sum_1^n x_{(k)} \int_{t_{k-1,n}}^{t_{k,n}} x^{\frac{1}{\beta}} f_{1,1}(x) dx}{E_P \left[ x^{\frac{2}{\beta}} \right]}$$

Celle en  $\frac{1}{\beta}$  s'annule pour

$$\alpha E_P \left[ x^{\frac{2}{\beta}} \log(x) \right] = \sum_1^n x_{(k)} \int_{t_{k-1,n}}^{t_{k,n}} x^{\frac{1}{\beta}} \log(x) f_{1,1}(x) dx.$$

Comme dans le cas de l'emv on obtient un système trop complexe pour être explicité ici.

Passons à la distance logarithmique  $\ell$ -quadratique  $\ell^2(P_{\alpha,\beta}, \mu_n)$  est

$$\begin{aligned} E_P \left[ \sum_1^n (\log(\alpha x^{\frac{1}{\beta}}) - \log(x_{(k)}))^2 \mathbb{1}_{t_{k-1,n}, t_{k,n}} \right] &= \log^2(\alpha) + \frac{1}{\beta^2} \sum_1^n \int_{t_{k-1,n}}^{t_{k,n}} \log^2(x) dP(x) \\ &+ \frac{1}{n} \sum_1^n \log^2(x_{(k)}) + 2 \frac{1}{\beta} \log(\alpha) \sum_1^n \int_{t_{k-1,n}}^{t_{k,n}} \log(x) dP(x) \\ &- 2 \log(\alpha) \frac{1}{n} \sum_1^n \log(x_{(k)}) - \frac{2}{\beta} \sum_1^n \log(x_{(k)}) \int_{t_{k-1,n}}^{t_{k,n}} \log(x) dP(x). \end{aligned}$$

La dérivée en  $\alpha$  s'annule pour

$$\frac{1}{n} \sum_1^n \log(x_{(k)}) = \log(\alpha) + \frac{1}{\beta} \sum_1^n \int_{t_{k-1,n}}^{t_{k,n}} \log(x) dP(x).$$

Celle en  $\frac{1}{\beta}$  s'annule pour

$$\frac{1}{\beta} \sum_1^n \int_{t_{k-1,n}}^{t_{k,n}} \log^2(x) dP(x) = \sum_1^n \log(x_{(k)}) \int_{t_{k-1,n}}^{t_{k,n}} \log(x) dP(x) - \log(\alpha) \sum_1^n \int_{t_{k-1,n}}^{t_{k,n}} \log(x) dP(x).$$

**Récapitulons** Les  $n$ -quantiles de  $P$  sont donnés par les intervalles  $(t_{k-1,n}, t_{k,n})$ ,  $k=1 \dots n$  où  $t_{k,n} = \frac{k}{n-k}$ . Pour un  $n$ -échantillon  $(X_1, \dots, X_n)$ , en posant

$$M_n = \frac{1}{n} \sum_1^n \log(X_k), \quad A_n = \sum_1^n \int_{\frac{k-1}{n-k+1}}^{\frac{k}{n-k}} \frac{\log(x)}{(1+x)^2} dx, \quad B_n = \sum_1^n \int_{\frac{k-1}{n-k+1}}^{\frac{k}{n-k}} \left( \frac{\log(x)}{1+x} \right)^2 dx$$

$$\text{et } C_n = \sum_1^n \log(X_{(k)}) \int_{\frac{k-1}{n-k+1}}^{\frac{k}{n-k}} \frac{\log(x)}{(1+x)^2} dx,$$

l'estimateur de Monge du couple  $(\alpha, \beta)$  pour un coût  $\ell$ -quadratique est le couple

$$\mathcal{E}_{2,Mg}^{\times}(\mathcal{L}\mathcal{L})(X_1, \dots, X_n) = \mathcal{E}_{2,Mg}^{\times}(\mathcal{L}\mathcal{L}) = (\mathcal{E}_{2,Mg}^{\alpha}(\mathcal{L}\mathcal{L}), \mathcal{E}_{2,Mg}^{\beta}(\mathcal{L}\mathcal{L})) \quad \text{où}$$

$$\mathcal{E}_{2,Mg}^{\beta}(\mathcal{L}\mathcal{L}) = \frac{B_n - A_n^2}{C_n - A_n M_n} \quad (12)$$

$$\mathcal{E}_{2,Mg}^{\alpha}(\mathcal{L}\mathcal{L}) = \exp \left[ M_n - \frac{A_n}{\mathcal{E}_{2,Mg}^{\beta}(\mathcal{L}\mathcal{L})} \right]. \quad (13)$$

Les estimateurs sont particulièrement maniables et les calculs aisés.

Par exemple, en faisant trois simulations d'un 1000-échantillon de  $P_{3,2}$ , on obtient :

$$(\mathcal{E}_{2,Mg}^{\alpha} = 3.00626, \mathcal{E}_{2,Mg}^{\beta} = 2.01027), (\mathcal{E}_{2,Mg}^{\alpha} = 3.00156, \mathcal{E}_{2,Mg}^{\beta} = 2.00175) \text{ et}$$

$$(\mathcal{E}_{2,Mg}^{\alpha} = 3.00865, \mathcal{E}_{2,Mg}^{\beta} = 2.00188).$$

Plus généralement, 50 simulations d'un 1000-échantillon donnent,

- pour  $\alpha$  une moyenne de 3.01108 et une variance de 0.000031
- et, pour  $\beta$ , une moyenne de 2.00406 et une variance de 0.0001237.

## 5 Estimation à partir de fonctions caractérisantes.

Pour une famille  $\{P_{\theta}\}$  nous cherchons un estimateur de  $\theta$  en utilisant des fonctions caractérisant ces lois.

Dans cette partie nous considérons les fonctions suivantes, dites *caractérisantes* :

- La fonction caractéristique (f.c.),  $\phi(u)$ , qui dépend éventuellement d'un paramètre  $\theta$ . La f.c. empirique d'un  $n$ -échantillon est  $\phi_n(u)$  en omettant le paramètre.

- La seconde fonction caractéristique notée  $\psi(u, \theta) = \log(\phi(u, \theta))$  et  $\psi_n(u)$  pour celle d'un  $n$ -échantillon.

- Enfin, lorsqu'elle existe, nous utilisons aussi la fonction génératrice des moments (f.g.m.) :  $g(u, \theta) = E_{P_\theta}[u^X]$ ,  $u \in \mathbb{R}_+$ ,  $X$  de loi  $P(\theta)$  et  $g_n(u)$  pour un  $n$ -échantillon.

L'idée d'introduire des fonctions caractéristiques est naturelle et déjà abondamment utilisée (d'une autre façon) voir par exemple [21] où on trouve une abondante bibliographie, citons aussi [4] et, pour les cumulants, [18].

Rappelons que l'estimation de  $\theta$  est basée sur le principe suivant :

*On cherche le  $\theta$  tel qu'une fonction caractéristique de la loi  $P(\theta)$  correspondante, soit la plus proche possible de celle de la loi empirique.*

Lorsque l'on prend comme critère de proximité une distance entre des lois on trouve les estimateurs de Monge.

Pour les fonctions précédentes nous allons étudier diverses distances. Par exemple pour les f.c. la distance liée à la convergence simple ou de manière équivalente (avec un théorème de Lévy) à la convergence uniforme sur tout compact.

Cette procédure se heurte à la difficulté de trouver  $\min_{\theta} \left[ \sup_{u \in K} |\phi(u, \theta) - \phi_n(u)| \right]$ .

En effet, pour obtenir des expressions simples, on souhaite une relation comme

$$\frac{\partial}{\partial \theta} \sup_{u \in K} (|\phi(u, \theta) - \phi_n(u)|) = \sup_{u \in K} \left| \frac{\partial}{\partial \theta} (\phi(u, \theta) - \phi_n(u)) \right|.$$

Malheureusement elle est peu probable.

C'est pourquoi nous introduisons, a priori, une mesure ou une probabilité, notée dans cette partie,  $\mu$  et une distance liée à cette mesure. Nous étudions

- La distance quadratique (ou son carré),  $\|\phi_1 - \phi_2\|_2 \triangleq (\int |\phi_1 - \phi_2|^2 d\mu)^{\frac{1}{2}}$

- La distance log-quadratique correspondant à  $\ell_2(\phi_1, \phi_2) \triangleq \|\log(\phi_1) - \log(\phi_2)\|_2^2$ .

Dans ces différents cas, il est plus aisé de vérifier la véracité de

$$\frac{\partial}{\partial \theta} E_{\mu}[\phi(u, \theta)] = E_{\mu} \left[ \frac{\partial \phi(u, \theta)}{\partial \theta} \right].$$

De plus, nous voulons que les estimateurs de  $\theta \in \mathbb{R}^d$  soient consistants. Pour cela nous devons faire quelques hypothèses minimales et quasiment toujours réalisées .

- L'application  $\theta \rightarrow P(\theta)$  est injective et continue pour la convergence en loi.

- Pour chaque  $\theta$  fixé, une fonction caractéristique est déterminée par ses valeurs sur un ensemble dénombrable dense dans un intervalle.

- Les mesures  $\mu$  considérées sont diffuses ou dans le cas discret on prend une suite de mesures  $\{\mu_n\}$  chacune uniforme sur  $D^n = \{u_k^n\}$ ,  $1 \leq k \leq K(n)$  et  $D^n$  converge vers un

ensemble dense dans un intervalle déterminant pour les fonctions caractérisantes de  $P(\theta)$ ;

- Enfin, nous supposons que pour  $\mu$  (ou les  $\mu_n$ ) la relation  $\frac{\partial}{\partial}(E_\mu[\cdot]) = E_\mu[\frac{\partial}{\partial}(\cdot)]$  est satisfaite.

**Proposition 5.1** *Sous les hypothèses précédentes, l'estimateur déduit par une fonction caractérisante,  $\theta_n$  est p.s. consistant i.e.  $\theta_n \xrightarrow{n \rightarrow \infty} \theta_0$ ,  $\theta_0$  vraie valeur du paramètre.*

**Preuve.** La loi des grands nombres assure que  $\phi_n(u) \rightarrow \phi(u, \theta_0)$  uniformément sur tout compact. Faisons la preuve pour  $\mu$  : la distance entre  $\phi_n$  et  $\phi(u, \theta_n)$  est inférieure à celle entre  $\phi_n$  et  $\phi(u, \theta_0)$  qui tend vers 0. Donc  $\theta_n \rightarrow \theta$ .

Pour une suite de mesures uniformes discrètes  $\{\mu_n\}$  ( $\mu_n$  ne désigne plus nécessairement des lois empiriques des échantillons) l'estimateur  $\theta_n$  est obtenu en minimisant  $E_{\mu_n}[d(\phi(u, \theta), \phi_n(u))]$  et l'adaptation de la partie précédente est évidente.  $\square$

Dans le cas discret et pour une taille suffisamment grande d'observations, une suite de mesures se distingue naturellement : Si  $\{x_n\}$  est la suite des valeurs observées, on peut prendre pour  $\mu_n$  la mesure uniforme sur  $\{x_k\}$ ,  $k \leq n$ .

## 5.1 Équations des estimateurs.

### 5.1.1 Distance quadratique et fonction caractéristique.

Écrivons la f.c. empirique d'un  $n$ -échantillon sous la forme  $\phi_n(u) = A_n(u) + iB_n(u)$  avec

$$A_n(u) = \frac{1}{n} \sum_1^n \cos(ux_k), \quad B_n(u) = \frac{1}{n} \sum_1^n \sin(ux_k) \text{ et pour } P_\theta, \quad \phi(u, \theta) = A(u, \theta) + iB(u, \theta).$$

Il vient  $|\phi - \phi_n|^2 = (A - A_n)^2 + (B - B_n)^2$ , ainsi  $E_\mu[\frac{\partial}{\partial\theta}|\phi - \phi_n|^2] = 0$  équivaut à

$$E_\mu \left[ A \frac{\partial}{\partial\theta}(A) + B \frac{\partial}{\partial\theta}(B) \right] = E_\mu \left[ A_n \frac{\partial}{\partial\theta}(A) + B_n \frac{\partial}{\partial\theta}(B) \right]. \quad (14)$$

### 5.1.2 Distance logarithmique et seconde fonction caractéristique.

Notons la f.c. empirique d'un  $n$ -échantillon sous la forme  $\phi_n = \rho_n(u) \exp(iD_n(u))$  et celle de la loi paramétrée  $\phi(u, \theta) = \rho(u, \theta) \exp(iD(u, \theta))$ .

Le carré du module de la différence des logarithmes s'écrit

$$\begin{aligned} \ell^2(\phi, \phi_n) &= |\log(\rho(u, \theta)) - \log(\rho_n(u))|^2 + |D_n(u) - D(u, \theta)|^2 \\ &= \log^2(\rho(u, \theta)) + \log^2(\rho_n(u)) + D_n^2(u) + D^2(u, \theta) \\ &\quad - 2D_n(u)D(u, \theta) - 2\log(\rho(u, \theta))\log(\rho_n(u)). \end{aligned}$$

Le minimum correspondant est solution de

$$\begin{aligned} E_\mu \left[ \log(\rho(u, \theta)) \frac{\partial \log(\rho(u, \theta))}{\partial \theta} + D(u, \theta) \frac{\partial D(u, \theta)}{\partial \theta} \right] \\ = E_\mu \left[ \frac{\partial \log(\rho(u, \theta))}{\partial \theta} \log(\rho_n(u)) + D_n(u) \frac{\partial D(u, \theta)}{\partial \theta} \right] \end{aligned} \quad (15)$$

Le développement en série entière de la loi empirique donne la variante suivante, où  $K_k^n$  est le  $k^{\text{ème}}$  cumulante de la loi d'un  $n$ -échantillon.

$$\log(\rho_n(u)) = \sum_k (-1)^k \frac{u^{2k}}{(2k)!} K_{2k}^n \quad \text{et} \quad D_n(u) = \sum_k (-1)^k \frac{u^{2k+1}}{(2k+1)!} K_{2k+1}^n.$$

Le minimum du carré des logarithmes satisfait l'équation :

$$\begin{aligned} E_\mu \left[ \log(\rho(u, \theta)) \frac{\partial \log(\rho(u, \theta))}{\partial \theta} + D(u, \theta) \frac{\partial D(u, \theta)}{\partial \theta} \right] \\ = \sum_k (-1)^k \frac{K_{2k}^n}{(2k)!} E_\mu \left[ \frac{\partial \log(\rho(u, \theta))}{\partial \theta} u^{2k} \right] + \sum_k (-1)^k \frac{K_{2k+1}^n}{(2k+1)!} E_\mu \left[ u^{2k+1} \frac{\partial D(u, \theta)}{\partial \theta} \right]. \end{aligned} \quad (16)$$

Cette expression, parfois plus maniable, offre une autre alternative pour la recherche des estimateurs.

### 5.1.3 Distance avec la fonction génératrice des moments

Soient  $g_n(u)$  la fonction génératrice des moments de la loi empirique d'un  $n$ -échantillon et  $g(u, \theta)$  celle de la loi de  $P_\theta$ , lorsqu'elle existe. L'estimateur de  $\theta$  pour la distance quadratique est solution de  $\frac{\partial}{\partial \theta} \|g(u, \theta) - g_n(u)\|_2^2 = 0$  i.e.

$$E_\mu \left[ g(u, \theta) \frac{\partial}{\partial \theta} (g(u, \theta)) \right] = E_\mu \left[ g_n(u) \frac{\partial}{\partial \theta} (g(u, \theta)) \right]. \quad (17)$$

Une expression similaire se déduit avec la distance logarithmique.

## 6 Exemples d'estimateurs caractérisants.

### 6.1 Loi avec un paramètre de forme.

Considérons le cas d'une f.c. de la forme i.e.  $\phi(u, \theta) = (\phi(u))^\theta$ , où  $\theta$  est le paramètre de forme, posons  $\log(\phi(u)) = \log(\rho(u)) + iD(u)$ , ainsi  $\rho(u, \theta) = (\rho(u))^\theta$  et  $D(u, \theta) = \theta D(u)$ . L'équation 15 donne comme estimateur de  $\theta$

$$\hat{\theta} = \frac{E_\mu \left[ \log(\rho(u)) \log(\rho_n(u)) + D(u) D_n(u) \right]}{E_\mu \left[ \log^2(\rho(u)) + D^2(u) \right]}. \quad (18)$$

## 6.2 Autour des Gaussiennes.

### 6.2.1 Variance d'une gaussienne.

Commençons par estimer la variance d'une gaussienne centrée avec l'équation 14.

Comme  $\theta = t^2$ ,  $\phi = \exp(-u^2\theta/2) = A(u, t)$  et  $B(u, t) = 0$ , on obtient :

$$E_\mu [u^2 \exp(-u^2\theta)] = \frac{1}{n} \sum_1^n E_\mu [u^2 \exp(-\frac{u^2\theta}{2}) \cos(ux_k)].$$

Choisissons différentes mesures a priori.

\* Lorsque  $\mu$  est la loi gaussienne normalisée, alors

$$E_\mu [u^2 \exp(-u^2\theta)] = \frac{1}{\sqrt{2\pi}} \int u^2 \exp(-\frac{u^2}{2}(1+2\theta)) du = (1+2\theta)^{-\frac{3}{2}}.$$

Par suite, la relation 14 se simplifie en

$$(1+2\theta)^{-\frac{3}{2}} = \frac{1}{n\sqrt{2\pi}} \sum_1^n \int u^2 \exp(-\frac{u^2}{2}(1+\theta)) \cos(ux_k) du.$$

Une approche numérique permet de résoudre cette équation.

\* Choisissons pour  $\mu$  la mesure de Lebesgue  $\lambda$ , on arrive à

$$\int u^2 \exp(-u^2\theta) du = \frac{\sqrt{\pi}}{2\theta^{\frac{3}{2}}} = \frac{1}{n} \sum_1^n \int u^2 \exp(-\frac{u^2\theta}{2}) \cos(ux_k) du. \text{ La solution } \theta \text{ vérifie}$$

$$\frac{n\sqrt{\pi}}{2} = \sum_{k=1}^n \int u^2 \exp(-\frac{u^2}{2}) \cos(\frac{x_k u}{\sqrt{\theta}}) du.$$

Utilisons maintenant l'équation 18 lorsque  $\phi$  est la f.c. d'une gaussienne  $\mathcal{N}(0, \theta)$ , on arrive à l'estimateur  $\bar{\theta}_n$  :

$$\bar{\theta}_n E_\mu \left[ \frac{u^4}{2} \right] = -E_\mu [u^2 \log(\rho_n(u))].$$

Cette expression est aisément calculable mais reste plus compliquée que les choix classiques. On obtient encore une autre variante en utilisant l'équation 16.

### 6.2.2 Mélange de gaussiennes.

La fonction caractéristique d'un mélange de gaussiennes s'écrit

$$\phi(u, \theta) = \sum_{k=1}^N \alpha_k \cos(um_k) \exp(-\frac{u^2 t_k^2}{2}) + i \sum_{k=1}^N \alpha_k \sin(um_k) \exp(-\frac{u^2 t_k^2}{2}) = A(u, \theta) + iB(u, \theta)$$

où  $\theta = (N, \{\alpha_k\}, \{t_k\}, \{m_k\}, k = 1 \dots N)$  avec  $0 \leq \alpha_k$ ,  $\sum \alpha_k = 1$ ,  $m_k \in \mathbb{R}$ ,  $0 \leq t_k$ .

En prenant successivement  $\frac{\partial}{\partial \alpha_p}(\cdot) = 0$ ,  $\frac{\partial}{\partial t_p^2}(\cdot) = 0$  et  $\frac{\partial}{\partial m_p}(\cdot) = 0$ , l'équation 14, conduit à

résoudre, pour  $p$  variant de 1 à  $N$ , les systèmes suivants :

$$\begin{aligned} E_\mu \left[ \cos(um_p) \exp\left(-\frac{u^2 t_p^2}{2}\right) (A(u, \theta) - A_n(u)) \right] &= E_\mu \left[ \sin(um_p) \exp\left(-\frac{u^2 t_p^2}{2}\right) (B_n(u) - B(u, \theta)) \right] \\ E_\mu \left[ u^2 \cos(um_p) \exp\left(-\frac{u^2 t_p^2}{2}\right) (A(u, \theta) - A_n(u)) \right] &= E_\mu \left[ u^2 \sin(um_p) \exp\left(-\frac{u^2 t_p^2}{2}\right) (B_n(u) - B(u, \theta)) \right] \\ E_\mu \left[ u \sin(um_p) \exp\left(-\frac{u^2 t_p^2}{2}\right) (A(u, \theta) - A_n(u)) \right] &= E_\mu \left[ u \cos(um_p) \exp\left(-\frac{u^2 t_p^2}{2}\right) (B(u, \theta) - B_n(u)) \right]. \end{aligned}$$

La variable entière,  $N$ , est déterminée de façon standard par écarts successifs.

En remplaçant  $A$ ,  $A_n$ ,  $B$  et  $B_n$  par leurs expressions, ces équations deviennent :

$$\begin{aligned} \sum_{k=1}^n E_\mu \left[ \alpha_k \cos[u(m_p - m_k)] \exp\left[-\frac{u^2}{2}(t_p^2 + t_k^2)\right] \right] &= \frac{1}{n} \sum_{k=1}^n E_\mu \left[ \cos[u(m_p - x_k)] \exp\left[-\frac{u^2}{2}t_p^2\right] \right] \\ \sum_{k=1}^n E_\mu \left[ \alpha_k u^2 \cos[u(m_p - m_k)] \exp\left[-\frac{u^2}{2}(t_p^2 + t_k^2)\right] \right] &= \frac{1}{n} \sum_{k=1}^n E_\mu \left[ u^2 \cos[u(m_p - x_k)] \exp\left[-\frac{u^2}{2}t_p^2\right] \right] \\ \sum_{k=1}^n E_\mu \left[ \alpha_k u \sin[u(m_p - m_k)] \exp\left[-\frac{u^2}{2}(t_p^2 + t_k^2)\right] \right] &= \frac{1}{n} \sum_{k=1}^n E_\mu \left[ u \sin[u(m_p - x_k)] \exp\left[-\frac{u^2}{2}t_p^2\right] \right] \end{aligned}$$

\* Lorsque l'on choisit pour mesure  $\mu$  une suite adéquate de mesures discrètes  $\{\mu_n\}$  formées à partir d'une suite  $\{u_k^n\}$ ,  $k \leq K(n)$ , la première des équations simplifiées précédentes devient, toujours pour tout  $p = 1 \dots N$ ,

$$\sum_{\substack{k \leq K(n) \\ l \leq N}} \alpha_l \cos[u_k^n(m_l - m_p)] \exp\left[-\frac{(u_k^n)^2}{2}(t_l^2 + t_p^2)\right] = \frac{1}{n} \sum_{\substack{k \leq K(n) \\ l \leq N}} \cos[u_k^n(x_l - m_p)] \exp\left[-\frac{(u_k^n)^2}{2}t_p^2\right].$$

Le reste des expressions est laissé au lecteur. Une approche numérique est indispensable.

### 6.3 Loi Gamma à deux paramètres.

La fonction caractéristique d'une loi Gamma à deux paramètres positifs  $\theta = (t, v)$  est  $\phi(u, \theta) = (1 - itu)^{-v} = (1 + t^2 u^2)^{-\frac{v}{2}} \exp[iv \arctan(ut)]$ .

Son logarithme s'écrit  $\psi(u, \theta) = \log(\phi(u, \theta)) = -\frac{v}{2} \log(1 + t^2 u^2) + iv \arctan(tu)$ .

Avec successivement les dérivées partielles par rapport à  $v$  et à  $t$ , l'équation 15 conduit à résoudre les équations :

$$v E_\mu \left[ \frac{1}{4} \log^2(1 + u^2 t^2) + \arctan^2(ut) \right] = E_\mu \left[ -\frac{1}{2} \log(1 + u^2 t^2) \log(\rho_n(u)) + D_n(u) \arctan(ut) \right]$$

$$\begin{aligned} \text{et } vE_\mu \left[ \frac{tu^2}{1+u^2t^2} \left[ \frac{1}{2} \log(1+u^2t^2) + \arctan(ut) \right] \right] \\ = E_\mu \left[ \frac{u}{1+u^2t^2} \left[ -tu \log(\rho_n(u)) + D_n(u) \right] \right] \end{aligned}$$

La solution de ce système nécessite une approche numérique.

\* En utilisant le développement en série entière de la f.c. de la loi empirique, on arrive aux équations suivantes :

$$\begin{aligned} vE_\mu \left[ \frac{1}{4} \log^2(1+u^2t^2) + \arctan^2(ut) \right] \\ = - \sum_k \frac{(-1)^k}{(2k)!} K_{2k}^n E_\mu \left[ u^{2k} \frac{1}{2} \log(1+t^2u^2) \right] \\ + \sum_k \frac{(-1)^k}{(2k+1)!} K_{2k+1}^n E_\mu \left[ u^{2k+1} \arctan(ut) \right] \\ \text{et } vE_\mu \left[ \frac{tu^2}{1+u^2t^2} \left[ \frac{1}{2} \log(1+u^2t^2) + \arctan(ut) \right] \right] \\ = - \sum_k \frac{(-1)^k}{(2k)!} K_{2k}^n E_\mu \left[ u^{2k+2} \frac{t}{1+t^2u^2} \right] \\ + \sum_k \frac{(-1)^k}{(2k+1)!} K_{2k+1}^n E_\mu \left[ u^{2k+2} \frac{1}{1+t^2u^2} \right] \end{aligned}$$

Là encore, il semble qu'une approche numérique soit nécessaire.

Néanmoins il convient de noter les remarques suivantes :

- Les cumulants  $K_k^n$  sont calculables pour tout échantillon, indépendamment de la loi paramétrée.

- Si l'intégrale par rapport à  $\mu$  d'une fonction impaire est nulle, les équations précédentes se simplifient encore.

- La primitive en  $u$  de  $\frac{u^k}{1+t^2u^2}$  est connue.

\* Enfin, pour utiliser le log de la f.g.m.  $\psi(u, \theta) = \log(g(u)) = -v \log(1-ut)$  et la distance  $\ell$ -quadratique, il faut supposer, a priori, que  $0 < ut < 1$ . Notons  $\psi_n(u, \theta)$  le log de la f.g.m. empirique. On doit résoudre  $E_\mu \left[ \frac{\partial}{\partial v} [(-v \log(1-ut)) - \psi_n(u)]^2 \right] = 0$  i.e.

$$v = - \frac{E_\mu [\psi_n(u) \log(1-ut)]}{E_\mu [\log^2(1-ut)]} \quad (*)$$

et  $E_\mu \left[ \frac{\partial}{\partial t} [(-v \log(1-ut)) - \psi_n(u)]^2 \right] = 0$  i.e.  $vE_\mu \left[ \frac{u \log(1-ut)}{1-ut} \right] = -E_\mu \left[ \frac{u \psi_n(u)}{1-ut} \right]$

Donc  $E_\mu[\psi_n(u) \log(1-ut)] E_\mu\left[\frac{u}{1-ut} \log(1-ut)\right] = E_\mu[\log^2(1-ut)] E_\mu\left[\frac{u}{1-ut} \psi_n(u)\right]$ .

Supposons  $0 < t < 1$  et prenons pour  $\mu$  la loi uniforme sur  $[0, 1]$ , on doit trouver  $t$  satisfaisant la relation :

$$\int_0^1 \psi_n(u) \log(1-ut) du \int_0^1 \frac{u}{(1-ut)} \log(1-ut) du = \int_0^1 \log^2(1-ut) du \int_0^1 \frac{u}{1-ut} \psi_n(u) du.$$

On connaît, en fonction de  $t$ , les primitives en  $u$  de  $\frac{u}{(1-ut)} \log(1-ut)$  et  $\log^2(1-ut)$ .

De plus les fonctions  $\psi_n(u)$  se développent en série entière en fonction des moments de la loi empirique  $\psi_n(u) = \sum_k u^k a_k^n$  où les  $a_k^n$  sont déterminés par l'échantillon  $\{x_i\}$ ,  $1 \leq i \leq n$  et sont connus. On arrive à une équation du type

$$f(t) \sum_k a_k^n \int_0^1 u^k \log(1-ut) du = g(t) \sum_k a_k^n \int_0^1 u^{k+1} \frac{1}{1-ut} dx.$$

Remarquons que les primitives en  $u$  de  $\frac{u^{k+1}}{1-ut}$  et de  $u^k \log(1-ut)$  sont connues en fonction de  $t$  et  $k$ . On est conduit ainsi à une équation du type  $h(t) = 0$ , la fonction  $h$  étant explicite. Ce système est, numériquement, plus simple à résoudre. La solution en  $t$  donne celle de  $v$  grâce à l'équation (\*) précédente.

En définitive, en donnant a priori une borne supérieure au paramètre  $t$ , on obtient une détermination "simple" des paramètres de la loi Gamma.

## 6.4 Loi stable de Paul Lévy

Cette loi dépend de quatre paramètres  $\theta = (\alpha, a, b, d)$  qui sont assujettis aux conditions  $0 < \alpha \leq 2$ ,  $-1 \leq b \leq 1$ ,  $a \in \mathbb{R}_+$  et  $d \in \mathbb{R}$ .

Pour  $\alpha \neq 1$ , la seconde f.c. de cette loi est

$$\psi(u, \theta) = -a|u|^\alpha + i[du + ab|u|^\alpha \text{signe}(u) \tan\left(\frac{\pi\alpha}{2}\right)].$$

Le cas  $\alpha = 1$  se traite séparément en remplaçant  $\tan\left(\frac{\pi\alpha}{2}\right)$  par  $\frac{2}{\pi} \log(|u|)$  dans  $\psi(u, \theta)$ .

Pour simplifier l'écriture, nous limitons l'étude à des mesures  $\mu$  ne chargeant que  $\mathbb{R}_+$  ce qui permet de supprimer les valeurs absolues et le signe intervenant dans  $\psi$ .

Écrivons  $\psi(u, \theta) = -au^\alpha + i[du + abu^\alpha \tan\left(\frac{\pi\alpha}{2}\right)] = -au^\alpha + iD(u, \theta)$

Pour un  $n$ -échantillon  $\{x_k\}$ , considérons le log de la f.c. de la loi empirique :

$$\psi_n(u) = \log \left[ \frac{1}{n} \sum_{k=1}^n \exp(iux_k) \right], \text{ que nous écrivons } \psi_n(u) = \gamma_n(u) + iD_n(u).$$

La distance  $\ell$ -quadratique conduit à minimiser  $E_\mu [(au^\alpha + \gamma_n(u))^2 + (D_n(u) - D(u, \theta))^2]$ .  
D'où le système d'équations :

$$\begin{aligned} \frac{\partial}{\partial a} E_\mu[\cdot] = 0 &\iff E_\mu[au^{2\alpha} + u^\alpha \gamma_n(u)] = b \tan\left(\frac{\pi\alpha}{2}\right) E_\mu[u^\alpha (D_n(u) - D(u, \theta))] \\ \frac{\partial}{\partial d} E_\mu[\cdot] = 0 &\iff E_\mu[u(D_n(u) - D(u, \theta))] = 0 \\ \frac{\partial}{\partial b} E_\mu[\cdot] = 0 &\iff E_\mu[u^\alpha (D_n(u) - D(u, \theta))] = 0 \\ \frac{\partial}{\partial \alpha} E_\mu[\cdot] = 0 &\iff E_\mu[a \log(u) u^\alpha (au^\alpha + \gamma_n(u))] \\ &= ab E_\mu\left[u^\alpha \left(\tan\left(\frac{\pi\alpha}{2}\right) \log(u) + \frac{\pi}{2} \cos^{-2}\left(\frac{\alpha\pi}{2}\right)\right) (D_n(u) - D(u, \theta))\right]. \end{aligned}$$

Ces équations se simplifient en :

$$\begin{aligned} E_\mu[u^\alpha \gamma_n(u)] &= -a E_\mu[u^{2\alpha}] \\ E_\mu[u D_n(u)] &= d E_\mu[u^2] + ab \tan\left(\frac{\pi\alpha}{2}\right) E_\mu[u^{\alpha+1}] \\ E_\mu[u^\alpha D_n(u)] &= ab \tan\left(\frac{\pi\alpha}{2}\right) E_\mu[u^{2\alpha}] + d E_\mu[u^{\alpha+1}] \\ a E_\mu[u^{2\alpha} \log(u)] + E_\mu[u^\alpha \log(u) \gamma_n(u)] &= b \tan\left(\frac{\pi\alpha}{2}\right) E_\mu[u^\alpha (D_n(u, \theta) - D(u)) \log(u)]. \end{aligned}$$

Système à résoudre numériquement lorsque que  $\mu$  est fixée.

Des simplifications apparaissent avec un choix judicieux de la mesure  $\mu$ , par exemple, prenons pour  $\mu$  la mesure de Lebesgue sur  $[0, 1]$ . Les équations deviennent

$$\begin{aligned} \frac{a}{2\alpha + 1} &= - \int_0^1 u^\alpha \gamma_n(u) du \\ \frac{d}{3} + \frac{ab}{\alpha + 2} \tan\left(\frac{\pi\alpha}{2}\right) &= \int_0^1 u D_n(u) du \\ \frac{d}{\alpha + 2} + \frac{ab}{2\alpha + 1} \tan\left(\frac{\pi\alpha}{2}\right) &= \int_0^1 u^\alpha D_n(u) du \end{aligned}$$

et pour la dernière équation :

$$\begin{aligned} \left(\frac{a}{2\alpha + 1}\right)^2 - a \int_0^1 u^\alpha \log(u) \gamma_n(u) du \\ = ab \tan\left(\frac{\pi\alpha}{2}\right) \left[ \int_0^1 u^\alpha \log(u) D_n(u) du + \frac{d}{(2 + \alpha)^2} + \frac{ab}{(1 + 2\alpha)^2} \tan\left(\frac{\pi\alpha}{2}\right) \right]. \end{aligned}$$

Posons,  $T = ab \tan\left(\frac{\pi\alpha}{2}\right)$  et, en oubliant les indices  $n$ ,  $A = \int_0^1 u\psi_n(u)du$ ,  $B = \int_0^1 u^\alpha\psi_n(u)(u)du$  et  $C = \int_0^1 u^\alpha\psi_n(u)du$ , on arrive au système :

$$\begin{aligned} \frac{a}{2\alpha+1} &= -Re[B] \\ \frac{d}{3} + \frac{T}{\alpha+2} &= Im[A] \\ \frac{d}{\alpha+2} + \frac{T}{2\alpha+1} &= Im[B] \\ \left(\frac{a}{2\alpha+1}\right)^2 - a \int_0^1 u^\alpha \log(u) \gamma_n(u) du &= T \left[ Im[C] + \frac{d}{(2+\alpha)^2} + \frac{T}{(1+2\alpha)^2} \right]. \end{aligned}$$

On commence par évaluer  $a$  puis  $d$  et  $T$  en fonction de  $\alpha$ . La dernière équation donne une solution pour l'estimateur  $\hat{\alpha}$ . On remarque que la recherche est simplifiée par la monotonie de la relation en  $\alpha$  (de part et d'autre de 1). On en déduit les autres estimations :  $\hat{a}$ ,  $\hat{b}$ ,  $\hat{d}$ .

Soulignons que la principale difficulté provient du calcul de l'intégrale du log de la f.c.

$$\int_0^1 u^x \log\left(\frac{1}{n} \sum_1^n \exp(iux_k)\right) du, \quad 0 \leq x \leq 2, \quad \text{dès que } n \text{ est grand.}$$

Aussi pour calculer cette intégrale nous utilisons une suite équirépartie, [5] et [14]. Nous prenons d'abord la suite  $\{k\pi\}$  modulo 1 et, plus intéressante la suite de Van Der Corput qui est à discrédance faible.

Dans le tableau suivant nous n'avons pris que 10 échantillons de tailles 2000 et des suites équiréparties de longueur 10000, ce qui se ressent dans la précision des paramètres.

<b>k*π</b>	<b>α</b>	<b>d</b>	<b>a</b>	<b>b</b>
valeur réelle	<b>1.7</b>	<b>0.5</b>	<b>1</b>	<b>0.3</b>
Moyenne	1.7472	0.47075	1.05558	0.353557
Écart type	0.0380856	0.0547958	0.0384082	0.144249
<b>Van Der Corput</b>	<b>α</b>	<b>d</b>	<b>a</b>	<b>b</b>
valeur réelle	<b>1.7</b>	<b>0.5</b>	<b>1</b>	<b>0.3</b>
Moyenne	1.68967	0.503734	0.980865	0.311434
Écart type	0.0271028	0.0439332	0.0215607	0.0788849
valeur réelle	<b>0.4</b>	<b>0.5</b>	<b>1</b>	<b>0.8</b>
Moyenne	0.398395	0.503011	0.997661	0.805392
Écart type	0.0100238	0.0489287	0.0208583	0.0335414

On amorce une comparaison avec le tableau 3 de [4]. On a simulé 25 fois une 250-échantillon et utilisé les 7000 premiers termes de la suite de Van Der Corput :

valeur réelle	$\alpha = 1.5$	$d = 0.$	$a = 1.$	$b = 0.$
Moyenne	1.50821	0.0284442	0.991901	- 0.0249287
Écart type	0.0754775	0.207524	0.109914	0.263947

Enfin, lorsque  $\mu = \mu_n$  est la loi uniforme sur  $\{u_k^n\}$ ,  $1 \leq k \leq K(n)$ , les estimateurs sont solutions d'une équation  $H(\theta) = 0$  qui se résoud par des méthodes numériques ou probabilistes classiques [12].

Le mémoire [2] contient de nombreuses références sur la simulation et d'autres méthodes pour l'estimation des paramètres des lois stables.

## 7 Extensions.

Rappelons au préalable, que nous n'avons pas comparé les estimateurs de Monge de la partie 3.1 avec ceux déduits de l'*emv* classique.

Les résultats obtenus concernent essentiellement les espaces  $\mathbb{L}^2$  et  $\mathbb{L}^1$ , néanmoins il est possible de passer à  $\mathbb{L}^p$  ou aux espaces d'Orlicz et même à ceux de Köthe. Dans le cadre général les calculs deviennent trop complexes.

*Voyons brièvement quelques pistes pouvant être suivies.*

– Pour une famille  $P_{\theta, \sigma}$  déduite de  $P$  par la transformation  $x \rightarrow \sigma x + \theta$ , pour calculer un risque on peut prendre "l'écart" :  $\mathbf{d}_p((\theta, \sigma), (\theta^*, \sigma^*)) \triangleq (|\theta - \theta^*|^p + (\ell(\sigma, \sigma^*))^p)^{\frac{1}{p}}$  où  $\ell$  est la distance logarithmique.

Dans ce cas et pour  $\mathcal{E}(x) = (\mathcal{E}_1(x), \mathcal{E}_2(x))$  estimateur du couple  $(\theta, \sigma)$ , un choix naturel du risque relatif à une norme  $\mathbb{L}^p$  est  $R_p((\theta, \sigma), \mathcal{E}) = \int (|\mathcal{E}_1(x) - \theta|^p + (\ell(\mathcal{E}_2(x), \sigma))^p) dP_{\theta, \sigma}$ .

Il s'en suit, lorsque  $\mathcal{E}_1$  est linéaire et  $\mathcal{E}_2$  multiplicatif, que  $R_p((\theta, \sigma), \mathcal{E}) = R_p((0, 1), \mathcal{E})$ .

C'est le cas de l'estimateur de Monge du couple trouvé dans la partie 3.1 et pour l'*emv*.

– Lorsqu'une famille  $\{P_t\}$  se déduit d'une probabilité  $P$  par  $P_t(\cdot) = P((\cdot)^{1/t})$  i.e. si  $Y$  est de loi  $P$ ,  $X_t = Y^t$  est de loi  $P_t$ , nous pouvons adopter un schéma analogue à celui du paramètre d'échelle ou du paramètre de forme, voir la partie 6.1. Il suffit de considérer la distance

$$\ell\ell : \ell\ell(x, y) \triangleq |\log(\log(x)) - \log(\log(y))|$$

où évidemment  $x, y$  et le paramètre  $t$  sont strictement positifs et alors,  $\ell\ell(x^t, y^t) = \ell\ell(x, y)$ .

– De même, pour la loi de Weibull à 2 paramètres qui, nous l'avons vu est obtenue à partir de loi exponentielle de paramètre 1, par la transformation  $x \rightarrow (\frac{x}{\theta})^{1/\alpha}$ ; on peut

s'intéresser à l'écart  $\mathbf{d}((\theta, \alpha), (\theta^*, \alpha^*)) \triangleq \ell(\theta, \theta^*) + \ell(\alpha, \alpha^*)$ .

Comme précédemment, si un estimateur  $\mathcal{E} = (\mathcal{E}_1, \mathcal{E}_2)$  du couple  $(\theta, \alpha)$  est multiplicatif pour sa première coordonnée et "puissance" pour la seconde, son risque vérifie

$$R((\theta, \alpha), \mathcal{E}) = R((1, 1), \mathcal{E}).$$

Ce qui simplifie les comparaisons et permet d'obtenir d'autres estimations et le calcul des risques dans le paragraphe 4.1

– Une limitation de la distance logarithmique  $\ell$  est qu'elle n'a été définie que sur  $\mathbb{R}_+$ . Par conséquent, nous n'avons plus considéré que des v.a. strictement positives.

Pour lever cette restriction une possibilité est, par exemple, de prendre pour  $xy \neq 0$ ,

$$\ell^*(x, y) = \ell(|x|, |y|) \text{ si } xy > 0 \text{ et } \ell^*(x, y) = \ell(|x|, |y|) + 1 \text{ si } xy < 0.$$

L'inégalité triangulaire  $\ell^*(x, y) \leq \ell^*(x, z) + \ell^*(z, y)$  est évidente ainsi que  $\ell^*(x, y) = \ell^*(y, x)$  et, enfin,  $\ell^*(x, y) = \ell^*(tx, ty)$  pour tout  $t > 0$ .

Cette distance permet d'aborder les v.a. à valeurs dans  $\mathbb{R} \setminus \{0\}$ .

– Reste la difficulté de trouver l'estimateur de Monge en général, voir par exemple [1] et [9]. Un autre point délicat est de trouver les fractiles d'une loi dépendant d'un paramètre. Pour cela nous avons utilisé le fait que les paramètres se déduisent à partir d'une valeur "centrale". C'est loin d'être toujours le cas.

– Nous laissons au lecteur la possibilité de poursuivre dans ces directions.

## References

- [1] ABDELLAOUI, T. (1998) Optimal solution of a Monge-Kantorovitch transportation problem. *Journal of Computational and Applied Mathematics* **96**, 149–161.
- [2] ADAM, E. (2001) L'analyse Fractale des marchés financiers. *Mémoire ISUP*. <http://www.multimania.com/eadam2>
- [3] AL-FAWZAN, M. A. (2000) *Methods for Estimating the Parameters for the Weibull Distribution*, <http://interstat.statjournals.net/>
- [4] ALVAREZ, A., OLIVARES, P. (2005) Méthodes d'estimation pour les lois stables avec des applications en finances *Journal de la Société Française de Statistique*, **146**, 4, 23–54.
- [5] BEJIAN, R., FAURE, H. (1977) Discrépance de la suite de Van der Corput. *C. R. Acad. Sci.*, **286**. Série A, 313–316.
- [6] BELILI, N., HEINICH, H. (2001) Median for metric spaces, *Appl. Mathematicae*, **28**, 2, 191–209.

- [7] BELILI, N., HEINICH, H. (2004) Application de la distance de Wasserstein au mélange, *C. R. Acad. Sci.*, **339**, 1, 653–658.
- [8] BERGER, J.O. (1985) *Statistical Decision Theory and Bayesian Analysis*. Second ed. Springer, NewYork.
- [9] BONNEU, F., DAOUIA, A. (2010) Mass transportation and the consistency of the empirical optimal conditional locations. *Ann Oper Res* **181** 159–170.
- [10] BRU, B., HEINICH, H. (1985) Meilleures approximations et médianes conditionnelles. *Ann. Inst. Henri Poincaré*, **21**, 3, 197–224.
- [11] CUESTA-ALBERTOS, J. A. ; MATRÁN, C.; RACHEV, S.T. AND RÜSCHENDORF, L. (1996) Mass transportation problems in probability theory. *Math. Scientist*, **21**, 34–72.
- [12] DUFLO, M. (1990) *Méthodes récursives aléatoires*. Masson.
- [13] KRASNOSEL'SKI, RUTICKIL, Y.B. (1961) *Convex function and Orlicz spaces*. Noordhoff, Gronigen.
- [14] KUIPER, I. AND NIEDERREITER, H. (1974) *Uniform Distribution*. John Wiley, New-York.
- [15] LEHMANN, E.L., CASELLA, G. (1998) *Theory of Point Estimation*. Second ed, Springer, NewYork.
- [16] LINDENSTRAUSS, J. AND TZAFRIRI, L. (1996) *Classical Banach spaces*. Springer, Berlin.
- [17] MORICE, E. (1968) Quelques problèmes d'estimation relatifs à la loi de Weibull. *Revue de Statistique Appliquée*, **16**, 3, 43–63.
- [18] NICOLAS, J-M. (2002) Introduction aux statistiques de deuxième espèce. *Traitement du Signal*, **19**, 3, 139–167.
- [19] SMALL, C.G. (1999) A survey of multidimensional medians. *Internat. Statist. Review*, **58**, 263–277.
- [20] SORIN, B., THIONET, P. (1968) Lois de probabilités de Bessel. *Revue de Statistique Appliquée*, **16**, 4, 65–72.
- [21] YU, J. (2004) Empirical Characteristic Function Estimation and Its Applications. *Econometric Reviews*, **23**, 2, 93–123.