



**HAL**  
open science

## La perte d'information par sondage (Calcul des probabilités) - (Première partie)

Pierre Thionet

► **To cite this version:**

Pierre Thionet. La perte d'information par sondage (Calcul des probabilités) - (Première partie). Annales de l'ISUP, 1959, VIII (4), pp.267-375. hal-04094688

**HAL Id: hal-04094688**

**<https://hal.science/hal-04094688>**

Submitted on 11 May 2023

**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.

Pierre THIONET

**LA PERTE D'INFORMATION  
PAR SONDAGE**

(Calcul des probabilités)

1. The first part of the paper is devoted to a discussion of the theory of the Turing machine and its application to the theory of computation. It is shown that the Turing machine is a model of the human mind and that the theory of computation is a theory of the human mind.

2. The second part of the paper is devoted to a discussion of the theory of the Turing machine and its application to the theory of computation. It is shown that the Turing machine is a model of the human mind and that the theory of computation is a theory of the human mind.

3. The third part of the paper is devoted to a discussion of the theory of the Turing machine and its application to the theory of computation. It is shown that the Turing machine is a model of the human mind and that the theory of computation is a theory of the human mind.

4. The fourth part of the paper is devoted to a discussion of the theory of the Turing machine and its application to the theory of computation. It is shown that the Turing machine is a model of the human mind and that the theory of computation is a theory of the human mind.

5. The fifth part of the paper is devoted to a discussion of the theory of the Turing machine and its application to the theory of computation. It is shown that the Turing machine is a model of the human mind and that the theory of computation is a theory of the human mind.

**LA PERTE D'INFORMATION**  
**PAR SONDAGE**  
 (Calcul des probabilités)

This section contains the main body of the paper, including the detailed discussion of the theory of the Turing machine and its application to the theory of computation. It is shown that the Turing machine is a model of the human mind and that the theory of computation is a theory of the human mind.

The text is dense and technical, covering various aspects of the theory of computation and the theory of the human mind. It includes several mathematical proofs and diagrams.

The paper concludes with a summary of the main results and a discussion of the implications of the theory of the Turing machine and its application to the theory of computation.

## INTRODUCTION

L'objet de la présente thèse est l'introduction du concept d'information au sens de Schutzenberger dans la théorie des sondages. On est parti de l'idée de la validité, dans ses grandes lignes, de la technique des sondages (telle qu'elle est enseignée dans les manuels de Yates - traduit en français sous la direction de M.G. Darmois, - de Deming, de Hansen, Hurvitz et Madow, de Sukhatme, de Cochran, etc., dans les cours de Tang, Kellerer, Žarcović, etc. Lévy-Bruhl, Chartier, Desabie, et nous-même). On a admis toutefois que les concepts théoriques qui sont à la base de ces méthodes pouvaient gagner à un examen nouveau, - et que la technique s'en trouverait consolidée.

Malgré l'accord qui règne entre tous ces auteurs, il existait en fait deux écoles : celle de Rothamsted pour laquelle les échantillons et populations sont extraits d'univers d'effectifs infinis, où les variables ont des distributions volontiers continues et dotées de densités; celle du Bureau of the CENSUS pour laquelle les échantillons sont extraits de populations-univers d'effectifs finis (ce qui conduit notamment à considérer un coefficient de corrélation intraclasse compris pour les uns entre 1 et 0, et pour les autres entre 1 et  $-1/\nu - 1$ ). Nous nous plaçons d'emblée dans l'école américaine, sous réserve qu'on n'oublie pas le cas limite (essentiel en pratique) où  $\nu$  devient infiniment grand,

L'existence de lois de distribution dont le moment d'ordre 2 n'existe pas ne posera pas de problème bien grave (car toutes les informations ne sont définies qu'à un facteur près).

On nous reprochera sans doute davantage d'avoir défini la perte d'information à partir d'un univers complètement connu dont les éléments échappent un à un à notre vue (contrairement aux méthodes d'analyse progressive), au lieu de définir le gain d'information à partir d'un état initial de la connaissance, comme le fait Lindley (suivant de près la définition que Shannon donne de l'information dans la théorie des communications). Lindley n'obtient de résultats qu'en supposant connues et la loi de distribution (école de Rothamsted) et la probabilité a

priori des paramètres à estimer, considérés comme des variables aléatoires (méthode d'estimation d'après Bayes). Cette théorie ne paraît pas, au moins sous sa forme actuelle, applicable aux techniques de sondage (Ch.VIII). Par exemple une quantité d'information qui devient infinie avec le nombre de tirages (avec remise) est inapplicable aux sondages à plusieurs degrés, stratifiés et en grappe.

Au contraire nous avons retrouvé, aux premières places des pertes d'information, la variance d'échantillonnage, dont fait grand usage la technique. Nous avons d'ailleurs critiqué, pour les estimateurs biaisés, le procédé courant qui consiste à ajouter à la variance le carré du biais; ce total n'est plus une perte d'information. Nous avons également signalé que l'usage de la variance n'était pas approprié au cas de l'estimation d'une médiane.

De toute façon la variance est une perte d'information particulière, à savoir symétrique; elle suppose qu'on attribue le même poids à deux erreurs d'échantillonnage, l'une positive, l'autre négative, de même ampleur. - On en a trouvé bien d'autres.

Les expressions de la perte d'information dans le cas de distributions quelconques (Ch.IV) n'ont été trouvées que sous des conditions de régularité très restrictives; il est possible que ces conditions puissent être réduites; en outre il y aurait lieu d'étudier plus systématiquement, les possibilités que peut présenter le cas où la dérivée première n'existe pas quand la perte élémentaire s'annule (ou d'autres cas encore).

Nous n'avons pas été en mesure, d'ailleurs, d'interpréter toute perte d'information au moyen d'estimateurs et de quantités à estimer; il semble que ce soit impossible. - Une autre lacune dans nos recherches serait une étude de l'estimation du mode de l'univers par celui de l'échantillon; il y aurait un réel intérêt pratique à savoir exprimer (et minimiser) la perte d'information correspondante.

Pour l'estimation de la plus grande valeur ou de l'étendue (range), la répartition optimum de l'échantillon sera celle qui donne la moindre perte d'information, par une extension naturelle de la théorie de la moindre variance. Le total [coût d'enquête, plus perte d'information] est à minimiser, ce qui est assez dans les idées de Wald (1950); le multiplicateur de Lagrange joue le rôle d'unité monétaire.

Chemin faisant, nous avons complété la théorie de l'estimation par sondage dans le cas biaisé. On aurait pu croire qu'alors toute statistique pouvait servir à estimer tout paramètre. Du moment qu'on définit l'information par une récurrence entre estimations sur une suite de plans de sondage de tailles décroissantes, l'estimation Z' n'informe sur

$\zeta$  que si elle informe sur tous les  $Z$  qui s'intercalent entre  $Z'$  et  $\zeta$ . La perte d'information ne peut être définie que pour les estimateurs uniformément biaisés; ce qui signifie qu'il existe une fonctionnelle  $\Delta[Z'Z]$  telle qu'on ait  $\mathcal{E}\Delta = 0$ .

Ceci conduit à calculer la perte d'information dans le cas où il existe une fonctionnelle  $\psi$  telle que :

$$\mathcal{E} \psi[Z] = \psi[\zeta]$$

En particulierisant encore, on a introduit l'estimateur isomorphe qui permet de retrouver les cas pratiques courants d'estimation biaisée; il correspond au cas où  $\zeta = f(a, b, c)$   $Z = f(A, B, C)$ , la fonction  $f$  étant la même,  $A, B, C$  étant des aléatoires calculés sur l'échantillon, leurs espérances mathématiques étant  $a, b, c$ . Les indicateurs que calculent les techniciens ont été retrouvés, par exemple pour les ratios; ce qu'on considérait jusqu'ici comme des valeurs approchées des carrés des coefficients de variation sont de véritables pertes d'information.

A cette même occasion on a retrouvé l'information de Fisher, comme perte d'information de l'estimation du maximum de vraisemblance - Ch. IV).

Les plans de sondage réels étant notablement plus compliqués que l'urne de Bernoulli, on a mené séparément et parallèlement l'étude de ces deux cas (Ch. I, II). On a eu l'idée de représenter les sondages dans l'urne unique par des points d'un axe et les sondages plus généraux par les points d'un treillis, les distances étant comptées sur les droites du dit treillis (Ch. III). Cet éclatement des difficultés suivant leur nature, topologique ou analytique, s'est révélé fructueux. - La représentation métrique des plans de sondage dans un espace euclidien à un nombre donné de dimensions a paru facile pour les sondages à plusieurs degrés et même à plusieurs phases. - Notre compréhension des formules de variance (apparemment compliquées) s'en est trouvée améliorée (Ch. V, VI).

La perte d'information étant supposée prendre place dans l'enseignement des Sondages, doit-on continuer à parler des informations supplémentaires dont l'emploi permet d'améliorer les estimateurs ou les plans de sondage? Autrement dit, que devient l'information quand on compare plusieurs itinéraires  $\zeta Z$ , au lieu de se déplacer le long d'un seul itinéraire? Tel est l'objet du Chapitre VII.

Au total nous pensons que les statisticiens ne devraient pas dire d'un plan de sondage qu'il apporte plus d'information qu'un autre, mais qu'il fait perdre moins d'information que l'autre. Une perte négative

est un gain, par rapport à un sondage de référence. Les diverses théories définissent des niveaux d'information; pour nous le niveau le plus bas correspond à un sondage ( $\bar{\omega}$ ) de taille minimum; pour d'autres il correspond à des renseignements antérieurs permettant d'écrire une loi de probabilité a priori. Quant au niveau le plus élevé, pour nous il est nécessairement à distance finie; d'autres l'autorisent à s'éloigner à l'infini. - Cela mis à part, l'équation de récurrence est la même; la nature des solutions diffère; car le symbole Espérance Mathématique n'a pas du tout le même sens chez Lindley qu'ici même. La topologie serait la même, si nous n'avions trouvé utile de faire jouer à l'infini un rôle spécial : il représente l'absence de toute information utilisable; et il est conforme aux habitudes des techniciens de penser qu'on n'a pas d'information utilisable tant qu'il reste une strate non sondée (Ch. VIII).

On peut penser finalement que la technique classique des sondages (y compris l'estimation du maximum de vraisemblance de Fisher, sort consolidée de notre analyse.

## CHAPITRE I

# TIRAGES ÉQUIPROBABLES DE BOULES DANS UNE URNE

### I - INTRODUCTION -

#### 1/ - Point de départ concret.

Les plans de sondage sont établis en pratique pour estimer certaines grandeurs physiques (économiques, sociales, médicales, etc.) au moyen d'observations limitées à un échantillon désigné par des tirages au sort plus ou moins compliqués.

Dans la plupart des cas, ces grandeurs sont des valeurs totales (superficie cultivée en jute, nombre de personnes travaillant moins de quarante heures par semaine, etc.) ou (ce qui revient au même pour la statistique pratique) des valeurs moyennes ou des "pourcentages" par unité de sondage. On "extrapole", c'est-à-dire qu'on passe des pourcentages ou valeurs moyennes à l'estimation du total, en multipliant par un facteur connu.

Ainsi la technique des sondages est axée sur les problèmes d'estimation de moyennes - compte tenu d'un plan de sondage plus ou moins compliqué.

L'estimation des ratios est un autre problème pratique très important qui ne pourra être étudié dans les premiers chapitres. On la trouvera au chapitre IV.

On commence par l'étude du plan de sondage le plus simple : tirages équiprobables de boules dans une urne, sans remise puis avec remise.

#### 2/ - Définitions.

Considérons une urne (ou univers)  $U$  renfermant  $v$  boules numérotées  $1, 2, \dots, i, \dots, v$  portant chacune un nombre réel  $x_i$ . Désignons par  $X$  l'ensemble de ces  $v$  valeurs.

La moyenne des  $x_i$  sera notée :

$$\sum_i x_i / v = \bar{x}$$

L'écart-type des  $x_i$  est  $\sigma$ , défini par :

$$\sum_i (x_i - \bar{x})^2 / v = \sigma^2$$

On tire au sort une boule de l'urne,  $n$  fois de suite sans jamais remettre les boules tirées. Pour un échantillon donné de  $n$  boules, la moyenne des  $n$  valeurs de  $x_i$  ainsi connues sera notée :

$$S_i x_i / n = \bar{X}_n \text{ ou } \bar{X}$$

$\bar{X}_n$  est une variable aléatoire lorsque ( $n$  étant donné) on envisage les  $m$  échantillons également probables à extraire de l'urne, avec chacun desquels on peut calculer une valeur de  $\bar{X}_n$ .

Cas des tirages sans remise ou exhaustifs ..  $m = v! / n! (v - n)!$

Cas des tirages avec remise ou bernoulliens  $m = v^n$

Lorsque  $\bar{x}$  n'est pas connu, on l'"estime" par une certaine valeur de  $\bar{X}_n$ , à savoir celle calculée sur l'échantillon effectivement tiré.

On dit que  $\bar{X}_n$  est estimateur "sans biais" de  $x$ , parce qu'on a :

$$\mathcal{E} \bar{X}_n = \bar{x}$$

où l'opérateur  $\mathcal{E}$  désigne l'espérance mathématique (étendue aux  $m$  valeurs également probables de  $\bar{X}_n$ ).

Ce qu'on dit pour  $\bar{x}$  et  $\bar{X}$  s'étend aux moments d'ordre quelconque de  $X$  calculés sur l'univers et sur échantillon (avec la même origine pour  $X$ ).

Plus généralement on considèrera l'estimateur sans biais  $Z$  de l'estimé  $\zeta$  définis sur l'échantillon et sur l'univers respectivement, avec  $\mathcal{E} Z = \zeta$ .

Erreur d'échantillonnage :

On désigne ainsi l'aléatoire  $\bar{X} - \bar{x}$ , ou encore  $Z - \zeta$ .

On a coutume de repérer l'ampleur des erreurs d'échantillonnage par la variance de l'estimateur,  $\mathcal{V} Z = \mathcal{E} (Z - \zeta)^2$ . On sait que :

$$\mathcal{V} \bar{X}_n \begin{cases} = (\sigma^2/n) (v - n/v - 1) & \text{(si tirages exhaustifs)} \\ = \sigma^2/n & \text{(si tirages bernoulliens)} \end{cases}$$

On emploie également beaucoup la racine carrée de cette variance :

$$\sigma(Z) = \sqrt{\mathcal{V}(Z)}$$

ou écart-type de  $Z$ , et la valeur relative de cet écart-type

$$\sigma(Z)/\zeta = C.V(Z) = \gamma$$

ou coefficient de variation (souvent exprimé en pour cent).

## II - LEMMES -

### Lemme 1.

Le carré de l'écart-type de  $v$  valeurs de  $(x)$  peut s'écrire :

$$\sigma^2 = \frac{\sum \delta^2}{v^2}$$

$\delta^2$  désignant le carré de la différence de deux quelconques des valeurs de  $x$ ,

$\sum$  désignant la somme des  $\frac{v(v-1)}{2}$  valeurs distinctes de  $\delta^2$ .

Autre forme du Lemme :  $\frac{v\sigma^2}{v-1} = \frac{\Delta^2}{2}$

où  $\Delta^2 = \sum \delta^2 / \frac{v(v-1)}{2}$  désigne la moyenne des écarts carrés.

### Démonstration du Lemme.

$$\begin{aligned} \delta^2 &= \sum_{i=1}^v \left[ (v-1)x_i - \sum_{j \neq i} x_j \right]^2 / v^3 \\ &= \sum_i \left[ \sum_{j \neq i} (x_i - x_j) \right]^2 / v^3 \\ &= \left[ \sum_{i=1}^v \sum_{j \neq i} (x_i - x_j)^2 + \sum_{i=1}^v \sum_{\substack{j \neq h \\ j \neq h}} (x_i - x_j)(x_i - x_h) \right] / v^3 \end{aligned}$$

Les carrés distincts  $(x_i - x_j)^2$  sont au nombre de  $v(v-1)/2$

$\sum_{j \neq i} (x_i - x_j)^2$  comprend  $(v-1)$  termes

$\sum_{i=1}^v \sum_{j \neq i} (x_i - x_j)^2$  comprend  $(v - 1)$  termes

$\sum_{\substack{j \\ j \neq i \\ i \neq h}} \sum_{\substack{h \\ h \neq j \\ i \neq h}} (x_i - x_j)(x_i - x_h)$  comprend  $(v - 1)(v - 2)$  produits.

En sommant cette dernière expression par rapport à  $i$ , on obtient  $(v - 1)(v - 2)$  produits.

Or les  $(x_i - x_j)(x_i - x_h)$  comprennent  $\frac{v(v - 1)(v - 2)}{2}$  valeurs distinctes, de sorte que

$\sum_i \sum_j \sum_h (x_i - x_j)(x_i - x_h)$  est égal à deux fois la somme de tous les produits distincts.

Mais :

$$\begin{aligned} (x_i - x_j)(x_i - x_h) + (x_j - x_h)(x_j - x_i) &= (x_i - x_j)(x_i - x_h + x_h - x_j) \\ &= (x_i - x_j)^2 \end{aligned}$$

Les  $(v - 1)(v - 2)$  produits donnent donc naissance à  $(v - 1)(v - 2) / 2$  carrés, c'est-à-dire  $(v - 2)$  fois chaque carré distinct, c'est-à-dire

$$(v - 2) \sum \delta^2$$

Ainsi :

$$\begin{aligned} v^3 \delta^2 &= 2 \sum \delta^2 + (v - 2) \sum \delta^2 \\ &= v \sum \delta^2 \end{aligned}$$

d'où

$$\sigma^2 = \sum \delta^2 / v^2 \quad \text{c. q. f. d.}$$

En particulier si la variable  $x$  prend les valeurs  $a$  et  $b$ , le carré de l'écart-type de  $x$  peut s'écrire :

$$\begin{aligned} \sigma^2 &= \left[ \left( a - \frac{a+b}{2} \right)^2 + \left( b - \frac{a+b}{2} \right)^2 \right] / 2 \\ &= (a - b)^2 / 4 \end{aligned}$$

### Lemme 2 sur la variance d'un échantillon exhaustif.

Pour l'échantillon d'effectif  $n$ , la moyenne des différences carrées sans répétition est :

$$D = 2S(x_i - x_j)^2/n(n-1)$$

où S désigne la sommation étendue à l'échantillon. Il est évident que

$$1/ \quad E(D) = \Delta$$

où E désigne l'espérance mathématique lorsque n est donnée.

$$2/ \quad \frac{n-1}{2n} D = s^2 = S_i \frac{(x_i - \bar{X})^2}{n}$$

Il en résulte que :

$$E \left( \frac{ns^2}{n-1} \right) = \frac{v\sigma^2}{v-1}$$

Tel est le lemme 2. (Résultat connu, parfois oublié).

### Lemme 3 sur la variance d'un échantillon bernoullien.

Faisons tendre  $v$  vers l'infini; il vient à la limite :

### Lemme 3 :

$$E' \left( \frac{ns'^2}{n-1} \right) = \sigma^2$$

E' désignant l'espérance mathématique relative aux tirages avec remise. (Résultat bien connu).

## III - ETUDE DES SUITES DE PLANS DE SONDAGE -

### 1/ - Les deux points de vue :

Au lieu d'envisager, comme d'habitude, des tirages successifs dans l'urne, au besoin jusqu'à épuisement ( $n = v$ ), - on pourra, en sens inverse, partir d'une urne parfaitement connue ( $v = n$ ) c'est-à-dire totalement vide (les boules étant placées dans une urne auxiliaire) et extraire de cette dernière, par tirage au sort, successivement

$$1, 2, \dots (v-n), \dots (v-1), v \text{ boules,}$$

qu'on replongera dans l'urne primitive.

### 2/ - Suite de plans de sondage : notion d'information.

#### Cas des sondages exhaustifs.

On peut symboliser la famille de sondages en question par des points consécutifs numérotés  $0, 1, 2, \dots, n, \dots, v$  sur un axe D, le point  $n$  représentant l'ensemble des sondages exhaustifs d'effectif  $n$  (pour la population et la variable  $X$  considérées).

Une échelle arithmétique sur l'axe D peut d'abord être employée

$$(D) \quad \begin{array}{ccccccccccc} 0 & 1 & 2 & 3 & & n & & v-1 & v \\ : & : & : & : & : & : & : & : & : & : \end{array}$$

Si l'on se déplace de (0) vers (v), on connaît de mieux en mieux l'urne; (0) représente l'absence totale de renseignements et (v) figure la connaissance parfaite de l'univers. (On désignera plutôt un tel point par un (U).

On peut dire qu'on "gagne de l'information" en passant de (n) à (n + 1) et qu'on "perd de l'information" en passant de n à (n - 1).

Il serait intéressant que la distance (v n) sur l'axe D représentât la perte d'information résultant du fait qu'on a limité les observations à n boules sur v (en ce qui concerne la connaissance de l'urne U). A cet égard l'échelle arithmétique ne nous paraît pas satisfaisante quantitativement; il ne serait pas conforme à notre concept d'information de dire qu'on perd la même quantité d'information chaque fois que l'effectif de l'échantillon diminue d'une unité. Bien au contraire notre sentiment est que :

- connaître toutes les boules, ou toutes sauf une, n'est pas très différent ;
- connaître une boule, ou n'en connaître aucune, est très différent;
- connaître deux boules au lieu d'une constitue un gain d'information bien plus grand que connaître v boules au lieu de v - 1.

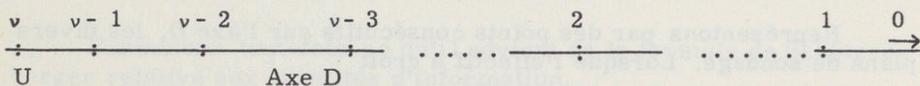
En revanche portons sur un axe D les valeurs pour  $n = 0, 1, 2, \dots, v - 1$ , de la variance de  $\bar{X}_n$

$$v \bar{X}_n = \frac{\sigma^2}{n} \frac{v-n}{v-1}$$

c'est-à-dire :

$$v \bar{X}_n = \frac{v\sigma^2}{v-1} \left( \frac{1}{n} - \frac{1}{v} \right)$$

Le point (0) est rejeté à l'infini, tandis que le point (v) est l'origine des abscisses. Quand n croît, chaque pas fait vers la connaissance parfaite est bien inférieur au précédent.



Ainsi notre intuition sera satisfaite si l'on choisit la variance  $\mathcal{V} \bar{X}$  pour mesurer la perte d'information imputable au sondage.  $\mathcal{V} \bar{X}$  possède une autre propriété.

THEOREME -

La variance  $\mathcal{V} \bar{X}$  satisfait à la définition de l'information suivant l'axiomatique de Schützenberger (Sch. page 45).

En effet, posons :

$$p(v, n) = \mathcal{V} \bar{X}_n$$

symbole rappelant que n unités sur v sont connues; et montrons que :

$$p(v, n - j) - p(v, n) = E P(n, n - j)$$

où P a la même définition que p mais est aléatoire.

Pour cela formons :

$$\begin{aligned} p(v, n - j) - p(v, n) &= \frac{v\sigma^2}{v-1} \left( \frac{1}{n-j} - \frac{1}{v} - \frac{1}{n} + \frac{1}{v} \right) \\ &= \frac{v\sigma^2}{v-1} \left( \frac{1}{n-j} - \frac{1}{n} \right) \end{aligned}$$

et posons :

$$P(n, n - j) = \frac{ns^2}{n-1} \left( \frac{1}{n-j} - \frac{1}{n} \right)$$

où  $s^2$  désigne la variance d'un échantillon (exhaustif) de n boules :

$$s^2 = S(x_i - \bar{X})/n$$

Le théorème découle immédiatement du Lemme 2 :

$$E(ns^2/n - 1) = v\sigma^2/v - 1$$

3/ - Cas des sondages Bernoulliens.

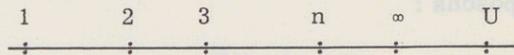
Le nombre  $n$  de tirages peut (théoriquement au moins) être pris aussi grand qu'on veut (et bien entendu dépasser  $v$ ).

Représentons par des points consécutifs sur l'axe  $D$ , les divers plans de sondage. Lorsque l'effectif  $n$  croît

$$n = 1, 2, 3 \dots$$

le point figuratif ( $n$ ) part d'une position (1) et se rapproche constamment sans jamais y parvenir d'un point  $U$  figurant la connaissance parfaite de l'urne.

Le point ( $n$ ) a donc un point d'accumulation  $\infty$  et il n'est pas évident que celui-ci doive coïncider avec le point  $U$  figurant la connaissance parfaite.



Nous ferons par la suite le Postulat  $\infty = U$ , c'est-à-dire que la perte d'information, par rapport à  $U$ , tend vers 0 quand  $n$  tend vers l'infini.

4/ - Accord entre figurations des sondages exhaustif et bernoullien.

Superposons les axes  $D$  relatifs à un sondage exhaustif et à un sondage bernoullien dans la même population.

Il est évident que les deux points (1) doivent coïncider, puisque sondages bernoullien et exhaustif sont indiscernables tant qu'on ne tire qu'une boule. De même le point ( $v$ ) de l'axe "exhaustif" doit coïncider avec le point  $U$  de l'axe bernoullien; et, avec le Postulat ( $\infty = U$ ), il coïncide avec le point  $\infty$ .

En outre, il est intuitif que le tirage de la  $n^{\text{ème}}$  boule apporte moins d'information en sondage bernoullien qu'en sondage exhaustif, ce qui doit se traduire par :



Toutes ces conditions sont remplies si la perte d'information en sondage bernoullien est choisie égale à

$$\frac{\sigma^2}{n}$$

quand la perte en sondage exhaustif est elle-même (à effectif égal)

$$\frac{\sigma^2}{n} \frac{v-n}{v-1}$$

Examinons toutefois ce qu'il advient de la formule de Schützenberger relative aux quantités d'information.

#### 5/ - THEOREME -

La variance  $\vartheta \bar{X}$  en sondage bernoullien satisfait à la définition de l'information suivant l'axiomatique de Schützenberger.

En effet, posons :

$$p'(v, n) = \frac{\sigma^2}{n}$$

il vient

$$p'(v, n-1) - p'(v, n) = \frac{\sigma^2}{n(n-1)}$$

Calculons ensuite

$$P'(n, n-1)$$

Soit  $s'^2$  la variance des  $x_i$  de l'échantillon tiré de taille  $n$ . Pour passer de là à l'échantillon d'effectif  $n-1$ , il faut tirer au sort l'une des  $n$  boules et l'exclure, ce qui équivaut à conserver un échantillon exhaustif de  $(n-1)$  boules; la perte d'information correspondante est donc :

$$P'(n, n-1) = \frac{s'^2}{n-1} \frac{n-(n-1)}{n-1} = \frac{s'^2}{(n-1)^2}$$

Il faut enfin prendre l'espérance mathématique de  $s'^2$ , la sommation étant étendue à tous les échantillons bernoulliens d'effectif  $n$ . Appliquons la lemme 3, il vient :

$$\mathcal{E}' \frac{s'^2}{(n-1)^2} = \frac{1}{(n-1)^2} \times \frac{n-1}{n} \sigma^2 = \frac{\sigma^2}{n(n-1)}$$

D'où

$$p'(v, n-1) - p'(v, n) = \mathcal{E}' P'(n, n-1) \quad \text{c. q. f. d.}$$

6/ - Extension à la variance de tout estimateur sans biais.

Soit  $Z_n$  et  $Z_{n'}$  les estimateurs de  $\zeta$  sur échantillons de tailles  $n$  et  $n'$

$$\mathcal{E}(Z_n) = \zeta, \quad \mathcal{E}(Z_{n'}) = \zeta, \quad \underline{n > n'}$$

Posons

$$p(v, n) = \mathcal{E}(Z_n - \zeta)^2$$

$$p(v, n') = \mathcal{E}(Z_{n'} - \zeta)^2$$

$$P(n, n') = E(Z_{n'} - Z_n)^2$$

où  $E$  désigne l'espérance mathématique de  $Z_{n'}$  sur les échantillons de taille  $n'$  tirés au sort dans un certain échantillon de taille  $n$ . Il vient :

$$\begin{aligned} p(v, n') - p(v, n) &= \mathcal{E} E [(Z_{n'} - Z_n)^2] \\ &= \mathcal{E} E \{ (Z_{n'} - Z_n) [(Z_{n'} + Z_n - 2\zeta) - (Z_{n'} - Z_n)] \} \\ &= 2 \mathcal{E} [(Z_n - \zeta) E(Z_{n'} - Z_n)] \end{aligned}$$

qui est nul si  $Z_{n'}$  est estimateur sans biais de  $Z_n$ .

$$E Z_{n'} = Z_n \quad (\text{quel que soit } n, v \geq n \geq n')$$

Cette condition renferme  $\mathcal{E} Z_{n'} = \zeta$  et  $\mathcal{E} Z_n = \zeta$ .

7/ - Définition.

Nous dirons alors que l'estimateur est uniformément sans biais.

Alors :

$$\forall Z = \mathcal{E}(Z - \zeta)^2$$

est perte d'information (au sens de l'axiomatique de Schützenberger).

Remarque.

Il n'est pas exclu que la relation de récurrence puisse être vérifiée sous des conditions moins strictes,  $n$  et  $n'$  ne prenant que certaines valeurs entières par exemple.

Cas des estimations isomorphes.

Si  $\zeta$  et  $Z$  sont des expressions isomorphes<sup>(1)</sup> (des  $x_i$  de l'univers et de l'échantillon), l'absence de biais est uniforme.

-----  
(1) C'est-à-dire ont même expression analytique.

Exemples.

$\bar{x}$  et  $\bar{X}$  sont uniformément sans biais; de même  $v\sigma^2/v-1$  et  $nS^2/n-1$

Exemple d'estimateur sans biais.

Si on estime la moyenne d'une distribution symétrique par la médiane de l'échantillon ( $n$  et  $n'$  impairs), on a bien

$$\mathcal{E} Z_n = \zeta, \quad \mathcal{E} Z_{n'} = \zeta.$$

mais on n'a pas en général

$$E Z_{n'} = Z_n$$

et la variance  $\mathcal{V} Z_n$  n'est pas une perte d'information.

Exception.

$X = a, b, c, d$  ou  $e$ , avec  $\zeta = c$ ,  $\overline{ce} = -\overline{ca}$ ,  $\overline{cd} = -\overline{cb}$ ,  $\overline{ca} = 7\overline{cb}/2$  avec  $v = 5$ ;  $n = 3$ ;  $n' = 1$ ; on a bien

$$\mathcal{E}(Z_3 - \zeta)(E Z_1 - Z_3) = 0$$

d'où 
$$\mathcal{V} Z_1 - \mathcal{V} Z_3 = \mathcal{E}(Z_1 - Z_3)^2$$

D'une manière générale on évitera de s'intéresser à des distributions trop particulières qui peuvent avoir des propriétés exceptionnelles sans portée pratique.

8/ - Cas de l'Estimation de la variance.

Partant de la définition de  $\sigma^2$ , du Lemme 1

$$\sigma^2 = \sum_i \frac{(x_i - \bar{x})^2}{v}$$

ou mieux :

$$\frac{v\sigma^2}{v-1} = \sum_i \sum_j \frac{(x_i - x_j)^2/2}{v(v-1)} = \mathcal{E}(x_i - x_j)^2/2$$

et de la population des  $x_1, x_2, \dots, x_1, \dots, x_v$ , on en déduit la "super-population" des couples  $x_i x_j (i \neq j)$ .

L'ensemble des  $(i, j)$  étant décrit par le seul indice  $k$

$$1 \leq k \leq \frac{v(v-1)}{2}$$

on posera :

$$y_k = (x_i - x_j)^2/2$$

Tout ceci amène à deux définitions :

(1) de la variance  $\sigma^2$

$$v\sigma^2/v - 1 = \bar{y} = \sum_k y_k / [v(v-1)/2]$$

(2) de l'estimation  $S^2$  de  $v\sigma^2/v - 1$  :

$$nS^2/n - 1 = \bar{Y} = S_k y_k / [n(n-1)/2]$$

avec

$$S \bar{Y} = \bar{y}$$

Variance de  $\bar{Y}$ .

a) Cas des tirages bernoulliens

Le calcul direct de  $\mathcal{V} \bar{Y}$  (Réf. T3 par exemple) donne facilement :

$$\mathcal{V} \bar{Y} = \frac{\mu_4^0}{n} - \frac{(n-3)\sigma^4}{n(n-1)} \equiv \frac{\mu_4^0 - 3\sigma^4}{n} + \frac{2\sigma^4}{n(n-1)}$$

où  $\mu_4^0$  est le moment des  $(x_i - x_j)^4$ .

$\mathcal{V} \bar{Y}$  est donc la somme de deux composantes, qu'on peut représenter par des divisions sur deux axes :

- l'une homothétique de  $\mathcal{V} \bar{X}$  (1, 1/2, 1/3, ... 1/n ...)
- l'autre nouvelle ( $\infty$ , 1, 1/3, 1/6, 1/10, ... 2/n(n-1) ...)

b) Cas des tirages exhaustifs (Réf. T. CR. 1, 1957).

Cette variance a été déjà calculée par de nombreux auteurs, dont Tukey (Jour. Am. Stat. Assoc. 1950, p. 501 & Ann. Math. Stat. 1956, p. 722), Chartier (Bull. Inf. I. N. S. E. E., Mars 1957, p. 31)...

Ces calculs se simplifient notablement si on a recours à la définition suivante de la variance

$$\sigma^2 = \sum (x_i - x_j)^2 / v^2$$

ou, pour abrégé  $\sigma^2 = \sum (a-b^2) / v^2 = \frac{v-1}{2v} \mathcal{E}(a-b)^2$

et si, par ailleurs, on remarque que, de l'identité

$$(a-b)^2(a-c)^2 + (b-c)^2(b-a)^2 + (c-a)^2(c-b)^2 \equiv \frac{1}{2} [(a-b)^4 + (b-c)^4 + (c-a)^4]$$

il suit qu'on a :

$$\mathcal{E}[(a-b)^2(a-c)^2] = \frac{1}{2} \mathcal{E}(a-b)^4$$

Dénombrons alors les termes du développement de  $(\sigma^2)^2 = \sigma^4$

$$\text{Termes en } \left\{ \begin{array}{l} (a-b)^4 \quad ; \quad (a-b)^2(a-c)^2 \quad ; \quad (a-b)^2(c-d)^2 \\ M = \frac{v(v-1)}{2} \left\{ \begin{array}{l} M \quad + \quad 2M(v-2) \quad + \quad M \frac{(v-2)(v-3)}{2} \end{array} \right. \end{array} \right.$$

et, en remplaçant  $v$  par  $n$  et  $M$  par  $m$ , on passera du développement de  $[\sum (a-b)^2]^2$  à celui de  $[S(a-b)^2]^2 = (n^2 S^2)^2$ . D'où les deux formules

$$v^4 \sigma^4 = \frac{v(v-1)^2}{2} \mathcal{E}(a-b)^4 + \frac{v(v-1)(v-2)(v-3)}{4} \mathcal{E}[(a-b)^2(c-d)^2]$$

$$\mathcal{E}[n^2 S^2]^2 = \frac{n(n-1)^2}{2} \mathcal{E}(a-b)^4 + \frac{n(n-1)(n-2)(n-3)}{4} \mathcal{E}[(a-b)^2(c-d)^2]$$

b. 1) En éliminant  $\mathcal{E}[(a-b)^2(c-d)^2]$  entre ces formules et en écrivant :

$$\begin{aligned} \frac{1}{v^2} \sum_i \sum_j (x_i - x_j)^4 &\equiv 2(\mu_4 - 4\mu_3\mu_1 + 3\mu_2^2) \\ &\equiv 2(\mu_4^0 + 3\sigma^4) \end{aligned}$$

d'où

$$\mathcal{E}(a-b)^4 \equiv \frac{v}{v-1} (\mu_4^0 + 3\sigma^4) ,$$

en considérant par ailleurs les  $n$  valeurs  $z_i$  :

$$z_i = (x_i - \mu_1)^2$$

dont la moyenne  $\bar{Z} = S(x_i - \mu_1)^2 / n$  estime  $\zeta = \sigma^2$  avec une variance égale à :

$$\mathcal{V} \bar{Z} = \frac{\sigma_z^2}{n} \frac{v-n}{v-1}$$

avec

$$\sigma_z^2 = \mu_4^0 - \sigma^4$$

on retrouve l'expression habituelle de  $\mathcal{V} \bar{Y}$ 

$$\mathcal{V} \bar{Y} = A \mathcal{V} \bar{Z} + B \sigma^4$$

avec

$$A = \frac{v(v-1)}{(v-2)(v-3)} \left[ 1 - \frac{2}{(n-1)(v-1)} \right]$$

avec

$$B = \frac{v(v-n)}{(v-1)(v-3)} \left[ 1 - \frac{n}{v-1} \right] \frac{2}{n(n-1)}$$

b. 2) Il est plus intéressant d'écrire :

$$\begin{aligned} \mathcal{V} \bar{Y} &= \mathcal{V} \left[ \frac{nS^2}{n-1} \right] = \frac{\mathcal{E} n^4 S^4}{n^2(n-1)^2} - \frac{v^4 \sigma^4}{v^2(v-1)^2} \\ &= \frac{1}{2} \left( \frac{1}{n} - \frac{1}{v} \right) \mathcal{E} (a-b)^4 + \frac{1}{4} \left[ \frac{(n-2)(n-3)}{n(n-1)} - \frac{(v-2)(v-3)}{v(v-1)} \right] \mathcal{E} [(a-b)^2(c-d)^2] \end{aligned}$$

$$\text{et comme } \frac{(n-2)(n-3)}{n(n-1)} \equiv 1 - \frac{4}{n} + \frac{2}{n(n-1)}$$

$$\mathcal{V} \bar{Y} \equiv \left[ \frac{1}{2} \mathcal{E} (a-b)^4 - \mathcal{E} [(a-b)^2(c-d)^2] \right] \left( \frac{1}{n} - \frac{1}{v} \right) + \frac{1}{2} \mathcal{E} [(a-b)^2(c-d)^2] \left[ \frac{1}{n(n-1)} - \frac{1}{v(v-1)} \right]$$

On retrouve deux composantes, avec (lorsque  $n$  varie) les mêmes divisions que dans le cas Bernoullien (qui correspond à  $v$  infini). On verra que c'est là un phénomène général.

#### Remarque.

$\bar{Y}$  n'est défini que si  $n \geq 2$ . On dira que c'est une perte d'information d'ordre 2.

#### 9/ - Extension au cas de plusieurs variables.

a) Considérons  $X$  et  $X'$  deux variables quelconques (et leur distribution sur les unités de sondage (boules) avec :

$$\begin{aligned} \mathcal{E} x_i &= \mu_1 & \mathcal{E} x'_i &= \mu'_1 \\ \bar{X} &= S x_i / n, & \bar{X}' &= S x'_i / n \end{aligned}$$

On a formellement :

$$\mathcal{V}(\bar{X} + t \bar{X}') = \frac{\alpha}{n} \cdot \frac{v-n}{v-1}$$

en posant :

$$\begin{aligned} \alpha &= \mathcal{E} [x_i - \mu_1 + t(x'_i - \mu'_1)]^2 \\ &= \mathcal{E}(x_i - \mu_1)^2 + 2t \mathcal{E}[(x_i - \mu_1)(x'_i - \mu'_1)] + t^2 \mathcal{E}(x'_i - \mu'_1)^2 \end{aligned}$$

On dira que  $\mathcal{V}$  est une perte d'information correspondant à l'estimation simultanée de  $\mu_1$  et  $\mu'_1$  par l'échantillon d'effectif  $n$ .

b) Le terme rectangle de  $\alpha$  possède (comme  $\alpha$  lui-même et chacun des termes carrés) deux expressions, avec les écarts aux moyennes d'une part, avec les écarts  $(x_i - x_j)$  et  $(x'_i - x'_j)$  d'autre part :

$$2t \mathcal{E}(x_i - \mu_1)(x'_i - \mu'_1) = \alpha - \mathcal{E}(x_i - \mu_1)^2 - t^2 \mathcal{E}(x'_i - \mu'_1)^2$$

$$\frac{2t}{v^2} \sum_{ij} (x_i - x_j)(x'_i - x'_j) = \alpha - \frac{1}{v^2} \sum \sum (x_i - x_j)^2 - \frac{t^2}{v^2} \sum \sum (x'_i - x'_j)^2$$

Il en résulte pour la covariance de  $x_i$  et  $x'_i$  des Lemmes analogues à ceux de la variance.

c) Comparons les deux formules (où les  $x_i$  et  $x'_i$  sont supposés essentiellement positifs),

$$\mathcal{E}(x_i - \mu_1)^2 = \mathcal{E}x_i^2 - \mu_1^2$$

$$\mathcal{E}(x_i - \mu_1)(x'_i - \mu'_1) = \mathcal{E}(x_i x'_i) - \mu_1 \mu'_1 = \mathcal{E}(\sqrt{x_i x'_i})^2 - (\sqrt{\mu_1 \mu'_1})^2$$

La covariance est donc une perte d'information pour l'estimation de  $\sqrt{\mu_1 \mu'_1}$ .

d) On peut employer  $\mathcal{V}(Z + t Z')$  ou  $\mathcal{V}(uZ + vZ' + wZ'')$  comme perte d'information dans des cas beaucoup plus généraux : estimateurs sans biais  $Z Z' Z'' \dots$  de caractéristiques  $\zeta, \zeta', \zeta'' \dots$  d'une distribution à une ou plusieurs variables.

### Exemple.

#### Estimation simultanée des deux premiers moments :

avec 
$$= \mathcal{E}[(x_i^2 - \mu_2) + t(x_i - \mu_1)]^2$$

$$\alpha = \mathcal{E}(x_i^2 - \mu_2)^2 + 2t\mathcal{E}(x_i^2 - \mu_2)(x_i - \mu_1) + t^2\mathcal{E}(x_i - \mu_1)^2$$

### 10/ - Urne avec plus de deux couleurs de boules.

Lorsqu'on estime les proportions (p q r) d'unités de sondage possédant les caractères  $\alpha \beta \mathcal{C}$  s'excluant mutuellement ( $p + q + r = 1$ ), il serait utile de savoir calculer une "distance" entre les points (pqr) et (PQR) dans le plan d'équation

$$x + y + z = 1$$

Pour cela on formera :

$$\begin{aligned} V &= \mathcal{V}(uP + vQ + wR) \\ &= \mathcal{V}[(u - w)P + (v - w)Q] \\ &= \frac{v - n}{n(v - 1)} [(u - w)^2 p(1 - p) + (v - w)^2 q(1 - q)] + 2(u - w)(v - w) [\mathcal{E}PQ - pq] \end{aligned}$$

Mais on sait que :

$$\frac{v}{v - 1} \mathcal{E}PQ = \frac{n}{n - 1} pq$$

$$\text{D'où : } V = \frac{v}{v - 1} [p(1 - p) - 2tpq + t^2q(1 - q)] \left( \frac{1}{n} - \frac{1}{v} \right)$$

(avec  $t = v - w/u - w$ ).

### 11/ - Extension à certaines estimations biaisées.

#### Moyennes généralisées.

A chaque unité  $i$  de la population, faisons correspondre  $y_i$ ,  $y_i = f(x_i)$ ,  $f(x)$  désignant une fonction connue arbitraire.

Utilisons le concept de moyenne généralisée (GINI)<sup>(1)</sup>; soit  $(x^*)$  la moyenne généralisée des  $x_i$  relativement à la fonction  $f$ , définie par l'équation :

$$f(x^*) = \bar{y} \quad \text{ou} \quad x^* = f^{-1}(\bar{y})$$

L'échantillon fournit une estimation  $(X^*)$  de  $(x^*)$ , définie par l'équation :

(1) Réf. : par exemple T2.

$$f(X^*) = \bar{Y} \quad \text{ou} \quad X^* = f^{-1}(\bar{Y})$$

$$\mathcal{V} \bar{Y} = \frac{v}{v-1} \sigma_y^2 \left( \frac{1}{n} - \frac{1}{v} \right)$$

est la perte d'information relative à l'estimation de  $\bar{y}$  par  $\bar{Y}$ , autrement dit : de  $\sum f(x_i)/v$  par  $S f(x_i)/n$ .

Trois questions se posent :

- Peut-on dire également que c'est la perte d'information relative à l'estimation de  $x^*$  par  $X^*$  ?
- Ne doit-on au contraire considérer que  $\mathcal{V} X^*$  est la perte d'information relative à l'estimation de  $x^*$  par  $X^*$  ?
- N'existe-t-il donc pas finalement plusieurs façons distinctes de représenter ladite perte d'information ?

Elles tiennent au fait que  $X^*$  n'est pas un estimateur sans biais de  $x^*$  quand  $\mathcal{E} \bar{Y} - \bar{y} = 0$ , sauf si  $f$  est linéaire.

12/ - Cas des estimateurs biaisés en général.

Le biais étant  $b = Z - \zeta$

il vient :

$$\begin{aligned} \mathcal{V} Z &= \mathcal{E} [Z - \zeta + \zeta - \mathcal{E} Z]^2 \\ &= \mathcal{E} (Z - \zeta)^2 + b^2 - 2b^2 \end{aligned}$$

Lorsque l'effectif varie, on voit facilement que ni  $\mathcal{E} Z$ , ni  $\mathcal{E} (Z - \zeta)^2$  ne satisfont à la relation de récurrence des informations.  $\mathcal{E} Z^2 - \zeta^2$  conviendrait manifestement, mais rien ne prouve que ce soit là une perte d'information imputable à l'estimateur  $Z$ , pour l'estimation de  $\zeta$  (voir § II Ch. IV).

Si l'on convient de mesurer la perte d'information relative à  $X^*$  par  $\mathcal{V} \bar{Y}$ , on n'est pas en possession d'une expression analogue pour les autres estimations biaisées (souvent beaucoup plus usuelles).

Il est courant dans la pratique des sondages - lorsque  $v$  est de l'ordre du million et  $n$  de l'ordre du millier - que  $\mathcal{V} Z$  et  $b^2$  ne diffèrent pratiquement pas d'infiniment petits de la forme

$$k/n \text{ et } k'/n^2;$$

de sorte que  $b^2$  est négligeable à côté de  $\mathcal{V} Z$ . En confondant  $x^*$  et  $\mathcal{E} X^*$  ou  $\zeta$  et  $\mathcal{E} Z \dots$ ,  $\mathcal{V} X^*$  ou  $\mathcal{V} Z$  est approximativement perte d'information.

On pourrait se demander s'il n'y a pas lieu de n'attacher la notion de perte d'information qu'aux estimateurs sans biais, vu que les estimateurs "biaisés" ne sont justement employés que dans les cas où on juge  $b^2$  négligeable à côté de  $\mathcal{V}X^*$ . On reviendra sur cette question au §II, Ch. IV.

### 13/ - Cas d'une population infiniment nombreuse.

Revenons sur le cas des sondages avec remise (§ 3°, 4°, 5°) et supposons  $v$  infiniment grand: rien ne s'en trouvera changé. En particulier: La variable  $x$  peut avoir une loi de distribution telle que  $\sigma^2$  n'existe pas,  $\mathcal{V}\bar{X}$  étant infiniment grand. Alors la notion de perte d'information semble en défaut.

On peut voir une raison de ne pas accepter la présente théorie. On cherchera alors à substituer à  $\mathcal{V}Z = E(Z - \zeta)^2$  une expression de la forme  $EW(Z, \zeta)$  qui reste finie cette fois. Mais on verra au chapitre IV que pareille expression n'a aucune raison d'être une information au sens de Schützenberger; si une même forme de  $W$  doit rester valable pour toutes les lois de distribution, et si  $Z$  estime  $\zeta$  sans biais, on verra que (sous certaines conditions de régularité pour  $W$ ),  $(Z - \zeta)^2$  est à un facteur près la seule forme possible pour  $W$ .

Ceci n'exclut pas l'emploi de pertes d'information n'ayant pas la forme  $EW$  (comme on va le voir au chapitre II), non plus d'ailleurs que le recours à des expressions  $EW$ , où la forme même de  $W$  dépendrait étroitement de la loi de distribution.

Il semble plus logique d'appeler formellement  $\mathcal{V}Z$  perte d'information, alors même que  $\sigma^2$  serait infini, toutes les pertes (d'ordre I) étant définies (à une similitude près) par la division  $1, \frac{1}{2}, \frac{1}{3}, \dots, \frac{1}{v}$ .

## CHAPITRE II

### LES PERTES D'INFORMATION DANS LE CAS DE L'URNE DE BERNOULLI <sup>(1)</sup>

#### I - DEFINITION GENERALE DE LA PERTE EN SONDAGE EXHAUSTIF -

Les variables  $X, Y, \dots$  prenant les valeurs  $(x_i, y_i, \dots)$  sur la "boule"  $i$  de l'urne ( $v$ ) d'effectif  $v$ , considérons un échantillon de  $n$  boules (tirées au sort sans remise avec d'égales probabilités) et une expression  $p(v, n)$  des  $x_i$  (de l'univers et de l'échantillon),

nulle si  $n = v$

définie positive pour  $n$  entier  $v > n \geq r$

strictement croissante avec  $(v - n)$ .

On dira de toute expression  $p(v, n)$  qu'elle représente une perte d'information due au fait qu'on ne connaît de l'urne ( $v$ ) que l'échantillon ( $n$ ), si elle est liée à  $p(v, n')$  correspondant à l'échantillon ( $n'$ ),  $n > n'$ , et à  $P(n, n')$  c'est-à-dire à la perte aléatoire d'information sur l'urne aléatoire ( $n$ ) avec l'échantillon ( $n'$ ), par la relation de récurrence :

$$p(v, n') = p(v, n) + \mathcal{E} P(n, n')$$

#### Remarque.

Il s'en suit que la variance  $\mathcal{V}X$  est une perte d'information à ce nouveau point de vue (plus large que celui du chapitre I puisqu'il ne comporte explicitement aucune estimation d'un  $\zeta$  par un certain  $Z$ ).

#### II - LES PERTES D'INFORMATION D'ORDRE 1 -

1/ - Partons de

$$\frac{v\sigma^2}{v-1} \left( \frac{1}{n'} - \frac{1}{v} \right) = \frac{v\sigma^2}{v-1} \left( \frac{1}{n} - \frac{1}{v} \right) + \mathcal{E} \left[ \frac{nS^2}{n-1} \left( \frac{1}{n'} - \frac{1}{n} \right) \right]$$

(1) Réf. T. C. R. 2.

relation établie au chapitre I.

Il résulte de la définition non courante de  $\sigma^2$  et de  $S^2$

$$\sigma^2 = \sum (x_i - x_j)^2 / v^2, \quad S^2 = S(x_i - x_j)^2 / n^2$$

que la relation de récurrence se conserve si l'on y remplace  $\frac{1}{2}(x_i - x_j)^2$  par toute fonction non négative symétrique  $f(x_i, x_j)$  telle que  $f(x_i, x_j)$  soit nul.

D'où, pour que  $p(v, n)$  soit une perte d'information il suffit que :

(Théorème)

$$\begin{aligned} p(v, n) &= \frac{2v}{v-1} \sum \frac{f(x_i, x_j)}{v^2} \left( \frac{1}{n} - \frac{1}{v} \right) \\ &= \mathcal{E}f(x_i, x_j) \cdot \left( \frac{1}{n} - \frac{1}{v} \right) \\ &= \gamma \left( \frac{1}{n} - \frac{1}{v} \right), \quad \text{avec } \gamma = \mathcal{E}f(x_i, x_j) \end{aligned}$$

Remarque.

Dans le cas particulier de

$$f = [g(x_i) - g(x_j)]^2$$

cette nouvelle perte d'information n'est autre que  $\mathcal{V}\bar{Y}$ , avec

$$y = g(x)$$

mais une interprétation aussi élémentaire fait défaut dans le cas où par exemple

$$f = g[(x_i - x_j)^2]$$

## 2/ - Notion d'ordre d'une perte d'information.

Dans le cas de l'estimation de la moyenne  $\bar{x}$  de l'urne par celle  $\bar{X}$  de l'échantillon, un échantillon formé d'une seule boule apporte déjà des renseignements. Aussi dirons-nous que  $\mathcal{V}\bar{X}$ , ainsi que les expressions

$$p(v, n) = \gamma \left( \frac{1}{n} - \frac{1}{v} \right)$$

avec

$$\gamma = \mathcal{E}f(x_i, x_j)$$

sont des pertes d'information du 1er ordre.

Les variances de l'estimation sans biais de  $\sigma^2$  ou du coefficient de corrélation, par exemple, sont des pertes d'information, mais ne sont pas d'ordre 1.

### 3/ - Représentation graphique des pertes d'ordre 1.

Si on porte, sur un axe  $\sqrt{X}$  ou toute autre perte  $p(v, n)$  ainsi définie, on obtient un "axe d'information" ( $\Delta$ ) où, pour une urne ( $v$ ) donnée, les échantillons des diverses tailles sont figurés par une division. Ces divisions sont semblables entre elles, quelles que soient la variable  $X$ , la fonction  $f$  et la taille  $v$  de l'urne. A l'échelle près, il n'existe donc qu'un axe ( $\Delta$ ).

### 4/ - Extension de la notion de perte d'ordre 1.

THEOREME -

Si, par hypothèse,  $p(v, 1)$  est de la forme  $\theta(v) \mathcal{E}f(x_i, x_j)$ , on a nécessairement :

$$p(v, n) = [\theta(v) - \theta(n)] \mathcal{E}f(x_i, x_j)$$

En effet à :

$$p(v, 1) = \theta(v) \cdot \mathcal{E}^{(v)} f$$

correspond :

$$P(v-1, 1) = \theta(v-1) \mathcal{E}^{(v-1)} f$$

etc

Passons de l'état :  $v$  boules connues à l'état : 1 boule connue. On peut le faire :

- soit d'un seul coup (perte d'information  $p(v, 1)$ );

- soit en deux temps en passant par l'état :  $(v-1)$  boules connues, d'où une perte d'information  $p(v, v-1)$  suivie de la perte de  $(v-2)$  boules sur  $(v-1)$ , c'est-à-dire espérance mathématique de la perte d'information

$$P(v-1, 1)$$

relative à une urne d'effectif  $(v-1)$  mais de composition aléatoire.

D'où :

$$p(v, 1) = p(v, v - 1) + \mathcal{E}P(v - 1, 1)$$

c'est-à-dire :

$$\begin{aligned} p(v, v - 1) &= p(v, 1) - \mathcal{E}P(v - 1, 1) \\ &= [\theta(v) - \theta(v - 1)] \gamma \end{aligned}$$

Pour passer du cas où  $n = v - 1$  au cas général où  $n$  est quelconque, il suffit de donner à l'échantillon des effectifs décroissants :  $(v - 1), (v - 2), \text{etc.} \dots (n + 2), (n + 1), n$ .

Les pertes d'information s'ajoutent, et il vient :

$$p(v, n) = [\theta(v) - \theta(n)] \cdot \gamma$$

Remarque.

Pour  $n = 1$ , on doit avoir nécessairement  $\theta(1) = 0$ .

En particulier : avec  $\theta(v) = \frac{v - 1}{v}$  ; on retrouve les pertes précédentes :

$$\theta(v) - \theta(n) = \frac{1}{n} - \frac{1}{v}$$

Toute fonction positive décroissante  $\theta(v)$  fournit des pertes d'information, qui ne diffèrent pas essentiellement entre elles et seront dites encore d'ordre 1, à condition que  $\theta(v)$  soit définie pour toute valeur entière de  $v$ , avec  $\theta(1) = 0$ .

On remarquera pourtant que :

$$\left[ 1 - \frac{1}{v} \right] \mathcal{E}f = \frac{2}{v^2} \sum_{i > j} f(x_i, x_j) = \frac{1}{v^2} \sum_i \sum_j f(x_i, x_j) = \mathcal{E}'f$$

$\mathcal{E}'$  désignant l'espérance mathématique étendue à toutes les combinaisons 2 à 2 avec répétition des  $x_i, x_j$ , c'est-à-dire aux échantillons bernoulliens de deux boules.

On verra (§ V) que c'est la nécessité d'un raccord avec les sondages bernoulliens qui détermine les fonctions  $\theta$  et, pour  $r = 1$ , conduit à prendre  $\theta = 1 - \frac{1}{v}$ .

## III - LES PERTES D'INFORMATION D'ORDRE 2 -

1/ - Exemple de perte du 2ème ordre.

Avec un échantillon de taille  $n \geq 2$ , on peut estimer  $\zeta = v\sigma^2/v - 1$ , par  $Z = nS^2/n - 1$ ; on sait exprimer  $\mathcal{V}Z$  comme combinaison linéaire de  $\mathcal{E}(x_i - x_j)^4$  et  $\mathcal{E}[(x_i - x_j)^2(x_i - x_k)^2]$ ; à savoir :

$$\mathcal{E} \frac{(n^2 S^2)^2}{n^2(n-1)^2} = \frac{1}{n} \mathcal{E} \frac{(a-b)^4}{2} + \frac{(n-2)(n-3)}{n(n-1)} \mathcal{E} \frac{(a-b)^2 \cdot (c-d)^2}{4}$$

$$\left( \frac{v\sigma^2}{v-1} \right)^2 = \frac{1}{v} \mathcal{E} \frac{(a-b)^4}{2} + \frac{(v-2)(v-3)}{v(v-1)} \mathcal{E} \frac{(a-b)^2 \cdot (c-d)^2}{4}$$

$$\begin{aligned} \mathcal{V}Z &= \left( \frac{1}{n} - \frac{1}{v} \right) \mathcal{E} \frac{(a-b)^4}{2} - \left[ \frac{2n-3}{n(n-1)} - \frac{2v-3}{v(v-1)} \right] \mathcal{E} \frac{(a-b)^2(c-d)^2}{2} \\ &= \left( \frac{1}{n} - \frac{1}{v} \right) \left[ \mathcal{E} \frac{(a-b)^4}{2} - \mathcal{E} (a-b)^2(c-d)^2 \right] + \left[ \frac{1}{n(n-1)} - \frac{1}{v(v-1)} \right] \mathcal{E} \frac{(a-b)^2(c-d)^2}{2} \end{aligned}$$

$\mathcal{V}Z$  est une perte d'information qui sera dite du 2ème ordre. C'est la somme d'une perte du 1er ordre et d'une du 2ème ordre : (car ni  $(2n-3)/n(n-1)$ , ni  $1/n(n-1)$  ne sont définis pour  $n < 2$ ).

2/ - Obtention de pertes plus générales.

Si l'on substitue  $f(x_i, x_j)$  à  $(x_i - x_j)^2$  avec :  $f(x_i, x_j) = f(x_j, x_i) > 0$ ,  $f(x_i, x_i) = 0$ , le transformé de  $\mathcal{V}Z$  est toujours une perte d'information d'ordre 2, puisque de la forme  $A \left( \frac{1}{n} - \frac{1}{v} \right) + B \left( \frac{1}{n(n-1)} - \frac{1}{v(v-1)} \right)$ .

3/ - Représentation graphique.

On considèrera deux vecteurs colinéaires, représentant le terme en  $1/n$  et le terme en  $1/n(n-1)$ , et par conséquent deux axes  $\Delta 1$  et  $\Delta 2$ , portant respectivement les divisions en  $1/n$  et  $1/n(n-1)$ , leurs échelles dépendant de la variable  $X$  et leurs origines de la taille  $v$  de l'urne.

4/ - Les cas où  $n = 3$  et  $n = 2$ .

Allant de soi avec  $n \geq 4$ , le calcul de  $\mathcal{V}Z$  serait en défaut pour  $n = 3$  ou  $2$ , si l'on ne raisonnait par continuité.

Pour  $n = 3$ , il vient :

$$\lim \mathcal{E} (a-b)^2(c-d)^2 \equiv \frac{2}{3} [(a-b)^2(a-c)^2 + (b-c)^2(b-a)^2 + (c-a)^2(c-b)^2]$$

(en faisant tendre  $d$  vers  $a$ , ou  $c$  avec des probabilités  $1/3$ ).

Etablissons l'identité :

$$(a-b)^2(a-c)^2 + (b-c)^2(b-a)^2 + (c-a)^2(c-b)^2 \equiv \frac{1}{2} [(a-b)^4 + (b-c)^4 + (c-a)^4]$$

Démonstration.

Posons  $\alpha = a - b$ ;  $\beta = b - c$ ;  $\gamma = c - a$  et développons  $(\alpha + \beta + \gamma)^4 \equiv 0$   
Il vient :

$$\alpha^4 + \beta^4 + \gamma^4 + (4 + 2)(\alpha^2\beta^2 + \beta^2\gamma^2 + \gamma^2\alpha^2) + (8 + 4)(\alpha^2\beta\gamma + \beta^2\gamma\alpha + \gamma^2\alpha\beta) + 4[\alpha^3(\beta + \gamma) + \beta^3(\gamma + \alpha) + \gamma^3(\alpha + \beta)] \equiv 0$$

Mais 
$$\begin{cases} \beta + \gamma = -\alpha & \text{etc} \\ \alpha^2\beta\gamma + \dots = \alpha\beta\gamma(\alpha + \beta + \gamma) = 0 \end{cases}$$

Il reste ainsi :

$$(1 - 4)(\alpha^4 + \beta^4 + \gamma^4) + 6(\alpha^2\beta^2 + \beta^2\gamma^2 + \gamma^2\alpha^2) \equiv 0$$

D'où 
$$\alpha^2\beta^2 + \beta^2\gamma^2 + \gamma^2\alpha^2 \equiv \frac{1}{2}(\alpha^4 + \beta^4 + \gamma^4)$$

d'où 
$$\lim \mathcal{E}(a - b)^2(c - d)^2 \equiv \mathcal{E}(a - b)^4$$

Pour  $n = 2$  : cette dernière identité conserve un sens.

5/ - Conclusion.

$\mathcal{V}Z$  est proportionnelle à  $\mathcal{E}(a - b)^4$  pour  $n = 3$  et  $n = 2$ , ce que le calcul direct montre (bien entendu) en faisant le changement de population :

$$y_k = (x_i - x_j)^2/2 \quad \begin{matrix} 1 \leq i \leq v \\ 1 \leq k \leq \frac{v(v-1)}{2} \end{matrix}$$

On estime  $v\sigma^2/v - 1$  par la moyenne  $\bar{Y}$  des  $y$  échantillon, ce qui permet de trouver pour  $n = 2$

$$\left(\frac{v-1}{v}\right)^2 \mathcal{E}(\bar{Y}^2) = \frac{v-1}{2v} \frac{2 \sum (x_i - x_j)^4}{v^2}$$

La variance s'en déduit en retranchant  $\sigma^4$ .

Pour  $n = 3$ , on trouve de même facilement :

$$\left(\frac{v-1}{v}\right)^2 \mathcal{E}(\bar{Y}^2) = \frac{v-1}{3v} \frac{2 \sum (x_i - \bar{x})^4}{v^2}$$

Pour les pertes plus générales, il n'en est pas ainsi; on a seulement :

$$\lim \mathcal{E}[f(a, b) f(c, d)] = 2 \mathcal{E}[f(a, b) f(a, c)]$$

les composantes en  $f(a, b)^2$  et  $f(a, b) f(a, c)$  ne fusionnant pas; car on ne peut transposer l'identité :

$$(a-b)^2(a-c)^2 + (b-c)^2(b-a)^2 + (c-a)^2(c-b)^2 = 2 [(a-b)^4 + (b-c)^4 + (c-a)^4]$$

Mais néanmoins les pertes sont définies jusqu'à  $n = 2$  inclus.

#### IV - LES PERTES D'INFORMATION D'ORDRE QUELCONQUE $r$ -

##### 1/ - THEOREME 1 -

Si  $p(v, n)$  est défini pour  $n \geq r$ , et si  $p(v, r)$  est de la forme  $\theta(v) \cdot \gamma$  où  $\theta(v)$  croît avec  $v$  à partir de  $\theta(r) = 0$ , on a nécessairement :

$$p(v, n) = [\theta(v) - \theta(n)] \gamma,$$

lorsque l'expression  $C$ , isomorphe de  $\gamma$  pour l'échantillon ( $\gamma$  étant définie sur l'univers), satisfait à la relation :

$$\mathcal{E} C = \gamma$$

##### Démonstration.

Refaire le calcul du § I ci-dessus.

##### Exemple.

$\gamma = \mathcal{E}[(a-b)^2 \cdot (c-d)^2]$  avec  $n \geq 4$  et passage à la limite pour  $n=3$  et 2.

2/ - Il est évident que :

##### THEOREME 2 -

- Si  $p$  est une perte d'information d'ordre  $r$ , il en est de même de  $\lambda p$ ,  $\lambda$  étant une constante.

- Si  $p$  et  $p'$  sont des pertes d'information d'ordres respectifs  $r$  et  $r'$ ,  $r > r'$ ,  $(p + q)$  est une perte d'information d'ordre  $r$ .

- Toute combinaison linéaire de pertes d'information est

une perte d'information; et son ordre est le plus grand des ordres des composantes. Toutefois, si certains coefficients sont négatifs, il est nécessaire que :

- ladite combinaison linéaire ne devienne pas négative;
- les termes de rang le plus élevé ne se détruisent pas.

### Réciproque du théorème 2.

On verra (§ III du Ch. IV) que, pour des plans de sondage très généraux (dont fait partie l'urne de Bernoulli) les pertes d'information sont nécessairement de la forme

$$\sum_i \sum_j [\theta_{ij}(v_i) - \theta_{ij}(n_i)] \gamma_{ij}$$

### 3/ - Sur quelques difficultés relatives à l'ordre d'une perte d'information.

a) Il est clair que l'ordre est déterminé par la nature de la fonction  $\theta(v)$ ; mais l'expression  $C$  elle-même n'est définie qu'autant que  $n$  et  $v$  ne descendent pas en-dessous d'une limite  $s$  qui ne coïncide pas nécessairement avec  $r$ .

Ainsi  $(a-b)^2(c-d)^2$  correspondait à  $s=4$ . Par continuité, on est passé de là à  $n=3$ , puis  $n=2$ ; étant entendu que  $\theta(n)$  serait infini si l'on allait jusqu'à  $n=1$  (donc  $r=2$ ).

b)  $\gamma = \mathcal{E}[(a-b)^2(a-c)^2 + (b-c)^2(b-a)^2 + (c-a)^2(c-b)^2]$  est en réalité (on l'a vu) :  $\gamma = \frac{1}{2} \mathcal{E}(a-b)^4$ ; de même

$$\mathcal{E}[(a-b)(a-c) + (b-c)(b-a) + (c-a)(c-b)]$$

n'est autre que :

$$\gamma = \frac{1}{2} \mathcal{E}(a-b)^2,$$

car on a :

$$(a-b)(a-c) + (b-c)(b-a) + (c-a)(c-b) \equiv \frac{1}{2} [(a-b)^2 + (b-c)^2 + (c-a)^2]$$

Considérons, au contraire, l'expression :

$$|a-b| \cdot |b-c| + |b-c| \cdot |c-a| + |c-a| \cdot |a-b|$$

Si l'on suppose  $a > b > c$ , elle n'est autre que :

$$(a - c)^2 + (a - b)(b - c)$$

Ainsi la valeur de  $s$  n'est pas toujours évidente; tandis qu'aucun doute n'existe pour  $r$ , une fois la fonction  $\theta$  choisie.

c) Mode d'obtention des expressions C et  $\gamma$  ci-dessus.

Considérons  $p$  valeurs de  $x$  (4, par exemple) obtenues par tirage exhaustif, désignées par  $x_1, x_2, \dots, x_p$ .

Désignons par  $f(i, j, \dots, 1)$  une fonction positive symétrique des  $x_1, x_2, \dots, x_p$ ;

Désignons par  $\mathcal{E}_p$  l'espérance mathématique étendue aux combinaisons  $p$  à  $p$  des  $v$  boules;

Désignons par  $E$  l'espérance mathématique étendue aux combinaisons  $p$  à  $p$  de  $(n)$  boules, elles-mêmes tirées exhaustivement dans l'urne.

Posons :  $\gamma = \mathcal{E}_p f(i, j, \dots, 1)$  et  $C = E f(i, j, \dots, 1)$ . Alors il est bien clair qu'on a :

$$\mathcal{E}(C) = \gamma$$

du moins tant que  $n \geq p$  (et pour  $r \leq n < p$ , on obtient par continuité les limites des pertes d'information).

d) Remarque.

On était naturellement conduit à définir, outre les pertes d'information du 1er ordre, des pertes d'information :

- du 2ème ordre, à base de fonctions  $f(i, j, k)$
- du 3ème ordre, à base de fonctions  $f(i, j, k, l)$ ,
- etc.

Tenons-nous-en au 2ème ordre. Nous avons commencé comme suit :

On définissait  $f(i, j, k)$  - perte d'information lorsque deux des trois boules sont connues - comme une fonction positive symétrique de  $x_i, x_j, x_k$  :

$$f(i, j, k) = f(i, k, j) = \dots \text{ etc.}$$

Il y a une chance sur trois pour que  $x_i$  ne soit pas tiré,  $x_j$  et  $x_k$

ayant été tirés; la perte d'information correspondante serait notée :

$$(i; jk) = (i; kj)$$

On aurait donc postulé que :

$$f(i j k) = \frac{1}{3} [(i; jk) + (j; ki) + (k; ij)]$$

Après quoi il était fort tentant de poser, pour un univers d'effectif quelconque :

$$\gamma = \mathcal{E}f(i j k)$$

On aurait donc eu :

- si  $x$  prend les valeurs  $a, b, c$ , (pour  $v = 3$ );
- $a, b, c, d$ , (pour  $v = 4$ );

$$\gamma = f(abc) \quad \text{dans le 1er cas;}$$

$$\gamma = \frac{f(abc) + f(abd) + f(acd) + f(bcd)}{4} \quad \text{dans le 2ème cas.}$$

Mais nous avons rencontré l'expression

$$(a - b)^2(c - d)^2 + (a - c)^2(b - d)^2 + (a - d)^2(b - c)^2$$

qui est une certaine perte d'information correspondant au cas où  $v = 4$ ,  $n = 2$  et qu'on ne peut mettre sous la forme :

$$f(abc) + f(abd) + f(acd) + f(bcd)$$

La base de départ était donc trop étroite.

#### e) Image géométrique.

De même que, pour l'ordre 1,  $f(x_i, x_j)$  représente une distance entre  $x_i$  et  $x_j$ , il est naturel de voir dans  $f(ijk)$  le demi-périmètre du triangle  $(ijk)$ , qui se réduit à la distance  $f(ik)$  si  $i$  et  $j$  viennent se confondre. Ce serait une erreur de voir dans  $f(ijk)$  l'aire d'un triangle, dans  $f(ijkl)$  le volume d'un tétraèdre ; car cela conduirait à postuler :

$$f(i i k) = 0 \quad \text{etc.}$$

L'image du périmètre incite pourtant à écrire :

$$\begin{aligned}
 f(ijk) &= \frac{1}{2} [(ij) + (jk) + (ki)] \\
 &= \frac{3}{2} \cdot \frac{(ij) + (jk) + (ki)}{3}
 \end{aligned}$$

avec  
(ij) = f(ijk), etc.

expressions de pertes du 1er ordre, fallacieusement présentées sous un aspect du 2ème ordre, ce qui ferait croire que tout f(ijk) peut se mettre sous cette forme.

#### V - TIRAGES AVEC REMISE DANS L'URNE -

1/ - La définition de la perte, posée dans le cas des tirages exhaustifs, se modifie avec les tirages bernoulliens comme suit; cette expression  $p'(v, n)$  :

- est définie positive pour  $n$  entier,  $n \geq r$ ;
- est strictement décroissante quand  $n$  croît;
- a une limite nulle pour  $n$  infini;
- satisfait à la relation de récurrence suivante :

$$p'(v, n) - p'(v, n + 1) = \mathcal{E}'P(n + 1, n)$$

où  $\mathcal{E}'$  désigne l'espérance en tirage bernoullien, et  $P$  la perte d'information dans le cas où l'on connaît  $n$  boules sur  $(n + 1)$  d'une urne (de composition aléatoire);  $P$  est calculée aux paragraphes ci-dessus.

En substituant à  $P(n + 1, n)$  l'expression

$$[\theta(n + 1) - \theta(n)]C_n$$

avec  $\mathcal{E}C = \gamma$ ,  $\theta(r) = 0$ ,  $\theta(n) < \theta(n + 1)$ , il vient

$$p'(v, n) = \sum_n^{+\infty} [\theta(h + 1) - \theta(h)] \mathcal{E}'C_n$$

#### 2/ - Calcul de $\mathcal{E}'C_n$ .

##### a) Pertes d'ordre 1.

$$C_n = Ef(x_i, x_j)$$

$$\mathcal{E}'C_n = \mathcal{E}'Ef = E \mathcal{E}'f$$

$$= \mathcal{E}'f = \frac{v-1}{v} \gamma$$

d'où :

$$p'(v, n) = [\theta(\infty) - \theta(n)] \frac{v-1}{v} \gamma$$

En particulier 
$$p'(v, 1) = \theta(\infty) \frac{v-1}{v} \gamma$$

Il faut se souvenir en outre que le tirage bernoullien est indiscernable du tirage exhaustif pour  $n = 1$  :

$$p'(v, 1) \equiv p(v, 1)$$

d'où  $\frac{v}{v-1} \theta(\infty) = \theta(v)$ , c'est-à-dire [si l'on pose  $\theta(\infty) = 1$ ]

$$\theta(v) \equiv 1 - \frac{1}{v}$$

b) Cas général.

THEOREME -

Si les espérances mathématiques d'une statistique donnée Y, pour des tirages avec et sans remise dans l'urne,  $\mathcal{E}'Y$  et  $\mathcal{E}Y$  ont des limites, quand l'effectif de l'urne tend vers l'infini, ces limites coïncident en général.

En effet : En n tirages équiprobables avec remise, la probabilité de tirer n boules distinctes est  $\left(1 - \frac{1}{v}\right) \left(1 - \frac{2}{v}\right) \dots \left(1 - \frac{n-1}{v}\right)$  qui tend vers 1 si v tend vers l'infini, n restant constant. La probabilité des valeurs non nulles de  $(\mathcal{E}'Y - \mathcal{E}Y)$  tend vers zéro. Si l'on suppose que Y reste fini on voit que, si  $\mathcal{E}Y$  tend vers une limite,  $\mathcal{E}'Y$  tendra vers la même; et vice versa.

Corollaire 1.

Partant d'une infinité d'urnes(v) de même composition, et mélangeant 1, 2, 3, ...  $\lambda$  ... urnes identiques, tirant chaque fois n boules avec remise d'une part, sans remise de l'autre, - calculant  $\mathcal{E}'C_n$  et  $\mathcal{E}C_n$ , avec

$$C_n = \text{Ef}(i j \dots 1);$$

alors, si  $\alpha = \mathcal{E}'f$  est une constante en  $\lambda$ , on a nécessairement

$$\lim_{v \rightarrow \infty} \mathcal{E}C_n = \alpha \text{ (sous réserve que } C_n < \infty \text{)}$$

Pour l'établir on notera que  $\mathcal{E}'C_n = \mathcal{E}'(Ef) = E(\mathcal{E}'f)$   
 $= \mathcal{E}'f = \alpha$

Corollaire 2.

Dans les mêmes conditions, si  $\mathcal{E}C_n$  a une limite  $\alpha$  lorsque  $\lambda$  tend vers l'infini et si  $f$  ne dépend pas de  $\nu$ , on a nécessairement :

$$\mathcal{E}'f = \lim_{\nu \rightarrow \infty} \mathcal{E}C_n \quad (\text{sous réserve que } C_n < \infty)$$

Conséquence.

De 
$$p(\nu, n) = \sum_n^{\nu-1} [\theta(h+1) - \theta(h)] \mathcal{E}C_h = [\theta(\nu) - \theta(n)] \gamma$$

on déduit 
$$\sum_n [\theta(h+1) - \theta(h)] \lim_{\lambda \rightarrow \infty} \mathcal{E}C_h = [\theta(\infty) - \theta(n)] \lim_{\lambda \rightarrow \infty} \gamma$$

D'où 
$$p'(\nu, n) = \lim_{\lambda \rightarrow \infty} p(\lambda\nu, n) = [\theta(\infty) - \theta(n)] \alpha \quad \text{avec } \lim_{\lambda \rightarrow \infty} \gamma = \alpha$$

Exemple.

$3C = f(a, b) f(c, d) + f(a, c) f(b, d) + f(a, d) f(b, c)$  (en supposant  $f < \infty$ ,  $f$  ne dépend pas de  $\nu$ ).

$\alpha = \mathcal{E}'C = \mathcal{E}'[f(a, b) f(c, d)]$

$$= \left(\frac{\nu-1}{\nu}\right) \left(\frac{\nu-2}{\nu}\right) \left(\frac{\nu-3}{\nu}\right) \mathcal{E}[f(a, b) f(c, d)] + \bar{\omega} \mathcal{E}[f(a, b) f(a, c)] + \dots$$

Remplaçons  $\nu$  par  $\lambda\nu$  et faisons tendre  $\lambda$  vers l'infini.  $C$  ne dépend pas de l'effectif de l'urne; de même  $\mathcal{E}'C$ .

Il reste :

$$\alpha = \mathcal{E}'C = \lim_{\lambda \rightarrow \infty} \mathcal{E}[f(a, b) f(c, d)]$$

3/ - Relation entre tirages bernoulliens et tirages exhaustifs.

A effectif égal  $n$  pour l'échantillon, la perte  $p'$  est nécessairement supérieure à  $p$ .

$$p'(\nu, n) - p(\nu, n) > 0,$$

ou 
$$[\theta(\infty) - \theta(n)] \alpha - [\theta(\nu) - \theta(n)] \gamma > 0$$

Pour  $n = r$ , il n'y a pas lieu d'identifier  $p$  et  $p'$ ; car il n'est pas

exact que la perte d'information soit la même avec  $r$  unités suivant qu'on les sait distinctes ou non.

Il en résulte apparemment qu'on a plus de latitude dans le choix de la fonction  $\theta$  que pour l'ordre 1.

En pratique, on a vu (au § III) que, pour  $r = 2$ , avec

$$\gamma = E(a - b)^2(c - d)^2$$

l'expression suivante était la seule qui subsistât :

$$\theta(v) - \theta(n) = \frac{1}{n(n-1)} - \frac{1}{v(v-1)}$$

ce qui donne, pour obtenir une fonction  $\theta(v)$  croissante, s'annulant avec  $v = 2$ ,

$$\theta_2(v) = \frac{1}{1.2} - \frac{1}{v(v-1)} = \frac{v(v-1) - 2}{2v(v-1)}$$

On est arrivé à pareille expression par des calculs de nombres de termes. Plus généralement, au moins chaque fois qu'on se limitera à des tirages équiprobables, le calcul de  $\theta$  fera intervenir l'analyse combinatoire, et sera une fraction rationnelle à coefficients entiers de variable  $v$ .

Une fraction rationnelle, croissante ( $v \geq r$ ), admettant le zéro ( $v = r$ ) et les pôles  $v = r - 1, r - 2, \dots, 0$ , et dont les 2 termes sont de degré minimum ( $r$ ); est

$$\theta_r(v) = \frac{1}{r!} - \frac{1}{v(v-1)\dots(v-r+1)} = \frac{1}{r!} - \frac{(v-r)!}{v!}$$

Une fraction rationnelle dont les 2 termes ne sont pas de même degré ne convient pas (du moment qu'on postule l'existence de  $\theta(\infty)$ ).

Toute fraction rationnelle admettant ces pôles peut être décomposée en éléments simples du type précédent (plus un polynôme), soit :

$$a_0 + a_1 \theta_1(v) + a_2 \theta_2(v) \dots + a_r \theta_r(v)$$

(on le voit en identifiant). Le polynôme disparaît en supposant le numérateur de degré  $r$ . La fraction admet ( $v = r$ ) pour zéro si les  $a_i$  sont liés par

$$0 = a_0 + a_1 \left(1 - \frac{1}{r}\right) + a_2 \left(\frac{1}{2!} - \frac{1}{r(r-1)}\right) + \dots + a_{r-1} \left(\frac{1}{r-1!} - \frac{1}{r!}\right)$$

relation qui détermine  $a_0$ . Finalement la forme générale de  $\theta(v)$  pour l'ordre  $r$  est cette combinaison  $\sum a_i \theta_i(v) + a_0$  dans les hypothèses du présent chapitre.

Une théorie n'envisageant de tirages que dans une population infiniment nombreuses (comme celle de Lindley, voir chapitre VIII) n'exclurait pas, en outre, la présence d'un polynôme, rendant infini  $\theta(\infty)$ .

#### 4/ - Cas d'une population infiniment nombreuse.

Il n'est pas exclu que  $\lim_{v \rightarrow \infty} \gamma(v)$  n'existe plus pour certaines distributions (voir n°13, Ch.I), suivant qu'on a choisi telle ou telle fonction  $f(i, j, \dots)$ ; car des expressions uniformément bornées  $|f(i, j)| < M$  n'auraient guère d'intérêt pratique.

relation du déterminant  $\Delta$  et l'absence de l'ordre des variables de (2) pour l'ordre  $\gamma$  est cette combinaison  $\sum_{i=1}^n a_i \gamma_i$  dans les hypothèses du présent chapitre.

Une théorie d'analyse de l'impact des taxes sur la population finit par être développée (comme celle de Lindley, voir chapitre VIII) n'exclut pas, au contraire, la possibilité d'un développement infini (voir l'annexe I).

4) - Cas d'une population infiniment nombreuse. Il n'est pas exclu que l'on ait  $\lim_{n \rightarrow \infty} \gamma_i = \gamma_i$  pour certaines distributions (voir n. 12, Ch. II), suivant qu'on a choisi telle ou telle fonction  $f(\gamma)$ , pour des approximations successives  $f(\gamma) \approx M$  n'est pas exclu d'ailleurs.

On remarque que, dans le cas d'une population infiniment nombreuse, on a  $\lim_{n \rightarrow \infty} \gamma_i = \gamma_i$  pour certaines distributions (voir n. 12, Ch. II), suivant qu'on a choisi telle ou telle fonction  $f(\gamma)$ , pour des approximations successives  $f(\gamma) \approx M$  n'est pas exclu d'ailleurs.

$$\frac{\partial \ln L(\gamma)}{\partial \gamma_i} = \frac{1}{L(\gamma)} \frac{\partial L(\gamma)}{\partial \gamma_i} = \frac{1}{L(\gamma)} \sum_{j=1}^n \frac{\partial L(\gamma)}{\partial \gamma_j} \frac{\partial \gamma_j}{\partial \gamma_i} = \frac{1}{L(\gamma)} \sum_{j=1}^n \frac{\partial L(\gamma)}{\partial \gamma_j} \delta_{ij} = \frac{1}{L(\gamma)} \frac{\partial L(\gamma)}{\partial \gamma_i}$$

On remarque que, dans le cas d'une population infiniment nombreuse, on a  $\lim_{n \rightarrow \infty} \gamma_i = \gamma_i$  pour certaines distributions (voir n. 12, Ch. II), suivant qu'on a choisi telle ou telle fonction  $f(\gamma)$ , pour des approximations successives  $f(\gamma) \approx M$  n'est pas exclu d'ailleurs.

On remarque que, dans le cas d'une population infiniment nombreuse, on a  $\lim_{n \rightarrow \infty} \gamma_i = \gamma_i$  pour certaines distributions (voir n. 12, Ch. II), suivant qu'on a choisi telle ou telle fonction  $f(\gamma)$ , pour des approximations successives  $f(\gamma) \approx M$  n'est pas exclu d'ailleurs.

$$\frac{\partial \ln L(\gamma)}{\partial \gamma_i} = \frac{1}{L(\gamma)} \frac{\partial L(\gamma)}{\partial \gamma_i} = \frac{1}{L(\gamma)} \sum_{j=1}^n \frac{\partial L(\gamma)}{\partial \gamma_j} \frac{\partial \gamma_j}{\partial \gamma_i} = \frac{1}{L(\gamma)} \sum_{j=1}^n \frac{\partial L(\gamma)}{\partial \gamma_j} \delta_{ij} = \frac{1}{L(\gamma)} \frac{\partial L(\gamma)}{\partial \gamma_i}$$

On remarque que, dans le cas d'une population infiniment nombreuse, on a  $\lim_{n \rightarrow \infty} \gamma_i = \gamma_i$  pour certaines distributions (voir n. 12, Ch. II), suivant qu'on a choisi telle ou telle fonction  $f(\gamma)$ , pour des approximations successives  $f(\gamma) \approx M$  n'est pas exclu d'ailleurs.

On remarque que, dans le cas d'une population infiniment nombreuse, on a  $\lim_{n \rightarrow \infty} \gamma_i = \gamma_i$  pour certaines distributions (voir n. 12, Ch. II), suivant qu'on a choisi telle ou telle fonction  $f(\gamma)$ , pour des approximations successives  $f(\gamma) \approx M$  n'est pas exclu d'ailleurs.

$$\frac{\partial \ln L(\gamma)}{\partial \gamma_i} = \frac{1}{L(\gamma)} \frac{\partial L(\gamma)}{\partial \gamma_i} = \frac{1}{L(\gamma)} \sum_{j=1}^n \frac{\partial L(\gamma)}{\partial \gamma_j} \frac{\partial \gamma_j}{\partial \gamma_i} = \frac{1}{L(\gamma)} \sum_{j=1}^n \frac{\partial L(\gamma)}{\partial \gamma_j} \delta_{ij} = \frac{1}{L(\gamma)} \frac{\partial L(\gamma)}{\partial \gamma_i}$$

On remarque que, dans le cas d'une population infiniment nombreuse, on a  $\lim_{n \rightarrow \infty} \gamma_i = \gamma_i$  pour certaines distributions (voir n. 12, Ch. II), suivant qu'on a choisi telle ou telle fonction  $f(\gamma)$ , pour des approximations successives  $f(\gamma) \approx M$  n'est pas exclu d'ailleurs.

$$\frac{\partial \ln L(\gamma)}{\partial \gamma_i} = \frac{1}{L(\gamma)} \frac{\partial L(\gamma)}{\partial \gamma_i} = \frac{1}{L(\gamma)} \sum_{j=1}^n \frac{\partial L(\gamma)}{\partial \gamma_j} \frac{\partial \gamma_j}{\partial \gamma_i} = \frac{1}{L(\gamma)} \sum_{j=1}^n \frac{\partial L(\gamma)}{\partial \gamma_j} \delta_{ij} = \frac{1}{L(\gamma)} \frac{\partial L(\gamma)}{\partial \gamma_i}$$

## CHAPITRE III

# LA PERTE D'INFORMATION AVEC LES PLANS DE SONDAGE QUELCONQUES

### I - DEFINITIONS FONDAMENTALES -

#### A - DEFINITION D'UN PLAN DE SONDAGE.

##### 1/ - Structure de l'univers sondé.

Le problème essentiel des sondages est l'évaluation de certaines caractéristiques d'une population donnée, composée d'éléments.

Par exemple, pour une population humaine discontinue, l'élément est l'être humain; on rencontre aussi des populations "continues", tel le territoire agricole, pour qui on peut envisager l'are, le mètre carré, le centimètre carré, ... comme élément constitutif.

Ces caractéristiques sont (le plus souvent) des grandeurs additives, des rapports de grandeurs additives, etc. ce qui autorise les remaniements suivants.

A cette population en quelque sorte brute, la théorie des sondages substitue une population dotée d'une certaine structure. On sera déjà dans un cas très général en envisageant la structure suivante :

Un premier découpage A, en segments portant les numéros : 1, 2, ...  
 $i \dots a$ ;

Un second découpage B, à l'intérieur de chacun des segments précédents, d'où constitution de sous-segments numérotés :  $i_1, i_2, \dots, i_j \dots i_b$ , à l'intérieur du segment  $n^o i$ ;

puis un troisième, un quatrième découpage, etc. pour aboutir en dernier lieu aux "éléments" de la population portant un numéro :  $i, j, k, \dots l$ .

avec

$$1 \leq i \leq a$$

$$1 \leq j \leq b_i$$

$$1 \leq k \leq c_{ij}$$

.....

$$1 \leq l \leq d_{ijk}$$

En pratique la structure est en partie imposée, en partie le résultat du choix plus ou moins judicieux du statisticien.

## 2/ - Cadre et structure de l'échantillonnage.

Une fois fixée la structure de la population à sonder, on donne le cadre de l'échantillonnage; dans le cas très général décrit ci-dessus, cela signifiera qu'on détermine le nombre des segments, sous-segments, ... éléments à conserver à l'échantillon, dans chacun des découpages.

Pratiquement on se donne en général :

- soit les effectifs  $n_A, n_B \dots$  des unités de sondages à conserver;

- soit les fractions de sondage  $f_A, f_B \dots$  proportions d'unités de sondage à conserver.

Les problèmes réels nécessitent qu'on distingue une notion intermédiaire entre la structure de la population et le cadre de l'échantillon; on conviendra de dire que : Fixer la structure de l'échantillon signifie décider quelles fractions de sondage sont égales à 1, les autres fractions restant à déterminer (entre 0 et 1, bornes exclues).

Pour abrégé, on parlera des effectifs ou de la taille d'un échantillon, pour désigner les paramètres  $n$  ou  $f$  restant à préciser lorsque la structure de l'échantillon est fixée.

## 3/ - Terminologie courante.

a) Si la fraction de sondage  $f_A$  retenue pour le premier découpage est égale à l'unité, la fraction  $f_B$  étant inférieure à 1, le découpage A est appelé stratification et chaque segment est appelé strate.

Chaque parcelle résultant du découpage B est appelée unité primaire de sondage (U. P. S. en abrégé) ou unité du 1er ordre.

Si, au contraire, la fraction  $f_A$  est inférieure à l'unité, c'est le découpage A qui procure les unités primaires de sondage.

b) Si (par exemple)  $f_b$  étant inférieur à 1,  $f_c$  est égal à 1, le découpage C est appelé sous-stratification; et chaque parcelle C est une sous-strate.

Chaque parcelle résultant du découpage D est appelée unité de sondage du 2ème degré.

Si, au contraire, la fraction  $f_c$  est inférieure à 1, le découpage C donne naissance aux unités du 2ème degré.

On définit de même les unités de sondage d'ordre 3, voire d'ordre quelconque.

Cette terminologie peu rationnelle s'explique par le fait que, dans la majorité des sondages réels, on a :

$$f_a = 1, \quad f_b < 1$$

#### 4/ - Désignation de l'échantillon.

Le cadre de l'échantillonnage étant fixé, - pour définir le plan d'échantillonnage il reste à se donner le mode de désignation de l'échantillon.

On ne s'intéressera ici qu'aux échantillons tirés au sort<sup>(1)</sup>, c'est-à-dire constitués comme suit :

- unités de sondage du 1er ordre tirées au sort (à l'intérieur des strates);
- unités de sondage du 2ème ordre tirées au sort parmi celles qui constituent les unités du 1er ordre tirées (et à l'intérieur des sous-strates);
- unités de sondage du 3ème ordre tirées au sort, etc.

Il convient de préciser le mode de tirage au sort et ceci pour chaque degré de sondage. Les modes les plus courants sont :

- sondage bernoullien (à la manière des boules de l'urne de Bernoulli) ou avec remise;
- sondage exhaustif (sans remettre les boules tirées dans l'urne);

(1) On évite de parler d'un échantillon aléatoire, en raison du sens péjoratif de ce mot en français courant.

- sondage bernoullien ou exhaustif avec probabilités inégales de tirage pour les diverses unités;

- sondage équilibré ("balanced sampling"); plus généralement le tirage au sort peut être assorti de clauses restrictives, destinées à écarter l'apparition des cas extrêmes.

5/ - Estimateurs et erreurs d'estimation.

Soit  $x_{ijk\dots 1}$  la valeur que prend une variable  $X$  sur l'élément  $(ijk\dots 1)$ .

Soit  $\zeta$  tout paramètre à estimer sur un échantillon d'éléments  $(ijk\dots 1)$ . L'estimateur  $Z$  de  $\zeta$  est une certaine fonction des  $x$  échantillons.

6/ - Coût d'une enquête sur échantillon.

Le coût des enquêtes est un facteur essentiel dans la pratique. Aussi introduit-on dans la théorie des sondages une fonction de coût<sup>(1)</sup> (plus ou moins réaliste), où entrent les divers effectifs de l'échantillon (à côté d'autres paramètres).

B - BUT DE LA THEORIE DES SONDAGES.

La théorie des sondages vise à permettre le choix du "meilleur" plan de sondage et du meilleur estimateur, sous des conditions restrictives telles que : pour un coût total donné, - ou pour certaines composantes du coût déterminées.

1/ - A la recherche du concept de "meilleur plan de sondage".

Toutes choses égales d'ailleurs, le meilleur plan de sondage serait celui qui fournit en moyenne les erreurs d'échantillonnage les plus faibles. A chaque  $\zeta$  correspond, pour chaque  $Z$ , un écart  $(Z - \zeta)$  erreur d'échantillonnage (ou d'estimation); on cherche à se faire une idée de l'ordre de grandeur de  $(Z - \zeta)$  en définissant (et évaluant numériquement) des indicateurs d'erreur ( $\zeta Z$ ).

a)  $Z - \zeta$  est le "biais" (bias) et représente l'erreur systématique d'échantillonnage.

Un estimateur présentant un biais important et inconnu (de façon à exclure toute correction de biais) serait inutile.

b) Avec un biais nul ou très faible, on est assuré que l'estimation sera "bien centrée"; mais il peut se faire qu'elle soit très

(1) Voir par exemple T. 7.

imprécise si des écarts  $Z - \zeta$  très importants ont des chances non négligeables de se produire. C'est pourquoi on définit des indicateurs de précision.

Le plus utilisé est la variance  $\mathcal{V}Z = \mathcal{E}(Z - \zeta)^2$  ou  $\mathcal{E}(Z - \zeta)^2$  lorsque  $Z$  est sans biais.

Concurremment avec la variance on emploie encore l'erreur-type, le coefficient de variation, l'erreur absolue moyenne (voir ci-dessus, § I du Chap. I).

Au moins depuis Neyman (1934), les plans de sondage optimum de la théorie classique sont définis de façon à rendre minimum certaines variances d'échantillonnage. (minimum lié).

Nous dirons plus généralement que le meilleur plan de sondage est celui qui fait perdre le moins d'information (toutes choses égales d'ailleurs).

## 2/ - Caractères généraux de la perte d'information.

On convient d'exiger a priori d'une perte d'information les caractères suivants :

- 1) être nulle si l'échantillon coïncide avec la population;
- 2) être infinie si l'échantillon a un effectif nul;
- 3) être déterminée par le plan d'échantillonnage et le choix de l'estimateur;
- 4) être croissante lorsque, pour une structure donnée, la taille de l'échantillon décroît.

La théorie des sondages définit déjà certains indicateurs d'erreur attachés à un estimateur donné.

Il est naturel de se demander si ces indicateurs ( $\zeta Z$ ) ne mesurent pas également la perte d'information.

Le biais, la variance, l'écart-type, le coefficient de variation, l'écart moyen présentent les quatre caractères ci-dessus.

Un cinquième caractère va décider si ce sont ou non des pertes d'information.

Donnons-nous la structure de l'échantillon, le mode de tirage et

l'estimateur. On a une famille (F) de plans de sondage ne différant entre eux que par la taille. Considérons deux plans de sondage (1) et (2) tels que la taille de (2) soit inférieure à celle de (1); ce qui signifie que l'un au moins des effectifs de (2) est inférieur à celui de (1) correspondant et qu'aucun n'est supérieur. Ainsi l'indicateur d'erreur ( $\zeta Z_2$ ) est plus grand que l'indicateur ( $\zeta Z_1$ ), mais la différence ( $\zeta Z_2$ ) - ( $\zeta Z_1$ ) n'a pas elle-même, en général, de signification intrinsèque.

Supposons qu'on ait, en outre, la relation de Chasles, plus précisément :

$$(\zeta Z_2) \equiv (\zeta Z_1) + \mathcal{E}_1(Z_1 Z_2)$$

$\mathcal{E}_1$  désignant l'espérance mathématique étendue à toutes les valeurs de l'aléatoire  $Z_1$  substitué à  $\zeta$ . Alors il est naturel de dire : ( $\zeta Z$ ) est perte d'information relative à l'estimation de  $\zeta$  (axiomatique de Schutzberger).

Toutefois si, dans le cas de tirages dans une urne, avec  $\mathcal{E}Z = \zeta$   $\forall Z = (\zeta Z)$  possède les cinq caractères ci-dessus, cela cesse d'être vrai pour le plus banal des sondages à deux degrés.

#### Cas d'un sondage à deux degrés et deux paramètres.

Les unités étant tirées avec d'égales probabilités et exhaustivement, à chacun des degrés de sondage, - les tirages au second degré sont supposés indépendants du résultat au premier degré.

Soit  $x_{ij}$  la valeur prise par la variable  $x$  sur l'unité (ij).

Supposons qu'on ait :  $1 \leq i \leq \mu$  ;  $1 \leq j \leq \bar{v}$  et qu'on tire au sort  $m$  unités primaires (sur  $\mu$ ), puis  $\bar{n}$  unités du second degré dans chaque unité primaire.

On sait alors que la moyenne générale des unités du 2ème degré

$$\bar{\bar{X}} = \frac{\sum_i \sum_j x_{ij}}{\mu \bar{v}}$$

admet comme estimation sans biais

$$\bar{X} = S_i S_j x_{ij} / m \bar{n}$$

dont la variance est :

$$\sigma_{\bar{X}}^2 = \sum_i \frac{1}{m\mu} \frac{\bar{v} \sigma_i^2}{\bar{v} - 1} \left( \frac{1}{\bar{n}} - \frac{1}{\bar{v}} \right) + \frac{\mu \sigma_0^2}{\mu - 1} \left( \frac{1}{m} - \frac{1}{\mu} \right)$$

avec  $\sigma_i^2 = \mathcal{V}(x_{ij}/i); \quad \bar{x}_i = \sum_j x_{ij}/\bar{v}$   
 $\sigma_0^2 = \mathcal{V}(\bar{x}_i).$

Supposons que cette variance représente une perte d'information. Considérons la population formée par l'échantillon de taille  $(m, \bar{n})$ . Tirons-en un échantillon de taille  $(m', \bar{n}')$ .

$$m' < m, \quad \bar{n}' < \bar{n}$$

La perte d'information supplémentaire serait :

$$\frac{1}{m'm} S_i \frac{\bar{n} s_i^2}{\bar{n} - 1} \left( \frac{1}{\bar{n}'} - \frac{1}{\bar{n}} \right) + \frac{m s_0^2}{m - 1} \left( \frac{1}{m'} - \frac{1}{m} \right)$$

Prenons l'espérance mathématique de cette expression pour tous les échantillons de taille  $(m, \bar{n})$ . On a :

$$E_2(\bar{n} s_i^2/\bar{n} - 1) = \bar{v} \sigma_i^2/\bar{v} - 1; \quad \mathcal{E}(m s_0^2/m - 1) = \mu \sigma_0^2/\mu - 1$$

$$\mathcal{E}_1 \left[ \frac{1}{m'm} S_i \frac{\bar{v} \sigma_i^2}{\bar{v} - 1} \right] = \frac{1}{m'\mu} \sum_i \frac{\bar{v} \sigma_i^2}{\bar{v} - 1}$$

Ainsi la nouvelle perte (totale) d'information serait :

$$\frac{1}{\mu} \sum_i \frac{\bar{v} \sigma_i^2}{\bar{v} - 1} \left[ \frac{1}{m'} \left( \frac{1}{\bar{n}'} - \frac{1}{\bar{n}} \right) + \frac{1}{m} \left( \frac{1}{\bar{n}} - \frac{1}{\bar{v}} \right) \right] + \frac{\mu \sigma_0^2}{\mu - 1} \left( \frac{1}{m'} - \frac{1}{m} \right)$$

au lieu de :

$$\frac{1}{\mu} \sum_i \frac{\bar{v} \sigma_i^2}{\bar{v} - 1} \frac{1}{m'} \left( \frac{1}{\bar{n}'} - \frac{1}{\bar{v}} \right) + \frac{\mu \sigma_0^2}{\mu - 1} \left( \frac{1}{m'} - \frac{1}{m} \right)$$

### Conclusion.

Quand on change simultanément le nombre des unités primaires échantillon et celui des unités du 2ème degré, la variance  $\bar{X}$  ne serait plus une perte d'information.

Le vice de raisonnement est dans l'expression supposée de la perte d'information supplémentaire, qui découle d'une interprétation trop simpliste de la transformation

$$(\zeta Z) \longrightarrow (Z_1 Z_2)$$

L'interprétation correcte de  $(\zeta Z)$  deviendra très naturelle après les considérations géométriques qu'on trouvera au § II.

## C - REMARQUE - LA THEORIE DE BASU.

Basu (Ba 1, 1952; Ba 2, 1954) nous a précédé, et notre indicateur ( $\zeta$  Z) est proche parent de sa fonction de risque  $\mathcal{E}W$ . Nous divergerons vite, car :

a) nous avons cherché quelles fonctions de risque étaient solution de la relation de Chasles, devenue équation fonctionnelle (de Chasles);

b) nous avons recours à des considérations géométriques;

c) finalement nous sommes obligé d'élargir la définition de Basu pour définir la perte d'information dans des cas pourtant fort courants.

Basu a tels soucis de rigueur et de généralité que nous citerons in extenso le début de son premier article.

"Soit  $X_1 X_2 \dots X_n$  un système de variables aléatoires dont on sait que la fonction de distribution  $F(x_1, x_2, \dots, x_n)$  appartient à un certain sous-espace  $\Omega$  de l'espace de toutes les fonctions de distribution possibles  $F$ . Par exemple, il arrivera qu'on sache que les  $X$  sont des distributions identiques mais indépendantes, de sorte que  $\Omega$  est le sous-espace de toutes les fonctions de distribution de la forme :

$$F = G(x_1) \cdot G(x_2) \dots G(x_n) \quad (1.1)$$

"G étant une quelconque fonction de distribution d'une seule variable .

"Le problème de l'estimation ponctuelle est celui de l'estimation d'une certaine caractéristique  $\theta = \mu(F)$  de la population, [ $\mu(F)$  étant une fonctionnelle à valeurs réelles définie pour tout  $F \in \Omega$ ] à l'aide d'un estimateur  $t = t(x_1, x_2, \dots, x_n)$  où  $x_i$  est une observation aléatoire de la variable  $X_i$ .

"Désignons par  $W(t, F)$  pour tout  $R \in \Omega$  fixé, les divers poids que le statisticien attache aux différentes valeurs de  $t$  en tant qu'estimations de  $\mu(F)$ ; et soit

$$r(F/t) = \int_R W(t, F) dF \quad (1.2)$$

"(où  $R$  est l'espace échantillon à  $n$  dimensions) la fonction de risque associée à l'estimateur  $t$ .

"Nous supposons qu'il existe des estimateurs  $t$  pour lesquels l'intégrale (1.2) est convergent pour tout  $F \in \Omega$ .

"Si  $r(F/t_1) \leq r(F/t_2)$  pour tout  $F \in \Omega$ , le signe = étant valable pour "au moins un  $F$ , alors on dit que  $t_1$  est uniformément plus puissant "que  $t_2$ .

"L'estimateur  $t_0$  peut être dit admissible s'il n'existe aucun estimateur  $t$  uniformément plus puissant que  $t_0$ .

"Dans le présent article nous nous limiterons aux seules fonctions poids qui sont des fonctions convexes (en décroissant) de  $t$  pour "tout  $F \in \Omega$  c'est-à-dire  $2W(t_1 + t_2/2, F) \leq W(t_1, F) + W(t_2, F)$  pour "tous les  $t_1$  et  $t_2$ . Si le signe = ne vaut que pour  $t_1 = t_2$ , on dira que la "fonction est strictement convexe".

Tout ceci reste valable ici, mais pour un autre usage. Basu se donne a priori une forme d'estimateur (par exemple linéaire pour estimer la moyenne, quadratique pour estimer la variance); et il se donne aussi l'expression du "risque" élémentaire  $W$ . Considérant alors le "risque"  $r = \mathcal{E}W$ , il détermine les paramètres des estimateurs de telle sorte que le "risque" soit minimum. Comme nous l'avons dit ailleurs (Réf. T. 11, 1956) lorsqu'on tire des boules d'une urne, avec

$$W = (Z - \zeta)^2,$$

le risque étant  $r = \mathcal{E}(Z - \zeta)^2 = \mathcal{V}Z + b^2$ ,

(où  $b$  désigne le biais), la politique du moindre risque ne conduit pas du tout à prendre  $b = 0$  puis  $\mathcal{V}Z$  minimum; elle conduit à prendre des estimateurs biaisés.

Par exemple, avec :

$$\zeta = \bar{x}, \quad Z = c\bar{X}, \quad \mathcal{V}Z = c^2\mathcal{V}\bar{X}, \quad \gamma^2 = \mathcal{V}\bar{X}/\bar{x}^2,$$

le moindre risque ne correspond pas à  $c = 1$ , mais à

$$c = 1/(1 + \gamma^2)$$

Lorsque  $\gamma^2$  est inconnu, Basu (suivant les méthodes de Wald) définit tout l'intervalle  $0 < c < 1$  comme fournissant les estimateurs uniformément les plus puissants.

L'intérêt du cas banal  $c = 1$  n'apparaît que si l'on se souvient qu'en pratique  $\gamma^2$  est très petit, donc  $c$  très peu inférieur à 1.

Ce résultat troublant disparaît si, au lieu de définir ainsi le "risque" trop arbitrairement, on lui impose d'être une perte d'information - au sens où Schutzenberger entend information.

Alors  $\mathcal{E}(Z - \zeta)^2$  par exemple n'est une perte d'information qu'avec un estimateur sans biais (uniformément);  $\mathcal{E}Z^2 - \zeta^2$  serait une perte d'information (sous des conditions plus larges) mais ce n'est plus un "risque" au sens de Wald, c'est-à-dire qu'on n'a pas en général

$$\mathcal{E}Z^2 - \zeta^2 = \mathcal{E}W \geq 0$$

(avec  $W \geq 0$ ). Les estimations biaisées amènent donc une rupture entre le point de vue de Basu et le nôtre (voir Chapitres IV et VII).

## II - TOPOLOGIE DES SONDAGES -

### A - L'ESPACE TOPOLOGIQUE.

L'intérêt (voire la nécessité) d'introduire des considérations de topologie dans les sondages n'apparaît que quand on dépasse le stade de l'urne unique, un peu avec le sondage stratifié, beaucoup avec le sondage à plusieurs degrés. On évoquera ici quelques treillis sans recourir à aucun théorème de topologie. On a seulement voulu préciser où finissait le besoin d'un espace topologique, où commençait celui d'une métrique.

#### 1/ - Figuration des échantillons au moyen d'un arbre<sup>(1)</sup>.

Ce n'est pas que le procédé classique (en théorie des Communications) de figuration par un arbre, d'un phénomène aléatoire ne soit applicable aux sondages, - mais il devient vite assez lourd. Décrivons-le cependant dans le cas le plus simple :

#### Exemple : Cas des tirages sans remise dans une urne d'effectif $v$ .

Un point figure la connaissance complète de l'urne (urne vidée). De ce point partent  $v$  branches aboutissant à  $v$  points dont chacun représente un échantillon d'effectif  $v - 1$ .

- chaque branche se ramifie en  $v - 1$  sous-branches, qui se rejoignent deux par deux, en  $v(v - 1)/2$  points représentant les échantillons d'effectif  $v - 2$ ;

- chaque sous-branche se ramifie en  $v - 2$  rameaux convergeant trois par trois, etc.

On arrive ainsi à  $C_v^{v-2}$  point (sondages d'effectif  $n = 2$ ), d'où partent deux rameaux; ceux-ci se rejoignent par  $(v - 1)$  à la fois en  $v$  points figurant les échantillons d'effectif  $n = 1$ .

(1) En topologie, "arbre" a un sens moins large qu'ici; "graphe" serait trop général.

On observera que les échantillons d'effectif 0 ne figurent pas sur l'arbre; ils n'apportent aucune "information"; mais il arrivera que les échantillons d'effectif 1, 2, ... ne renseignent pas non plus sur tel aspect de la population sondée qui justement nous intéresse; par exemple un échantillon d'effectif  $n = 1$ , ne permet pas d'estimer la variance.

Remarque.

On peut également définir "l'arbre" dans l'autre sens, à partir de  $v$  échantillons d'effectif 1, - donnant  $v(v - 1)/2$  échantillons d'effectif 2, etc. si les tirages sont exhaustifs, - ou  $v^2$  échantillons d'effectif 2, etc. si les tirages sont bernoulliens (auquel cas l'arbre ne se referme jamais).

2/ - Figuration (topologique) des tirages dans une urne.

On va symboliser l'arbre du n°1 ci-dessus sur un axe d'information.

Soit  $v$  l'effectif de l'urne,  $n$  l'effectif de l'échantillon; on peut représenter par une suite de points consécutifs ( $n$ ) sur un demi-axe, les sondages d'effectif 1, 2, ...  $n$  ...

Le point ( $v$ ) représentant la connaissance parfaite de l'urne sera pris comme origine du demi-axe.

Il peut être commode de distinguer un point (O) représentant l'absence d'échantillon. On peut le supposer rejeté à l'infini.

En faisant tendre  $v$  vers l'infini, on figure le cas d'une population de base infinie. Les tirages avec remise seront figurés de même. Enfin il n'a pas été spécifié que les tirages étaient équiprobables ou non, et les tirages (bernoulliens) à probabilités inégales sont figurés par le même axe.

3/ - Représentation topologique d'un plan de sondage.

Etant donné la population découpée en strates, unités de sondages, etc., autrement dit une structure de sondage, un plan de sondage appartenant à cette famille de sondages se définit en fixant un certain nombre de paramètres : soit effectifs de l'échantillon, soit fractions de sondage. Chaque plan de sondage de la famille étant représenté par un point d'un espace topologique, ce point occupe les divers sommets d'un treillis lorsqu'on donne aux paramètres toutes les valeurs admissibles.

Quant aux échantillons eux-mêmes, ils seraient représentés par l'arbre; mais celui-ci sera refermé sur son axe, chaque point ou som-

met du treillis figurant tous les échantillons qu'un même plan de sondage peut procurer.

Une fois définis les points de cet espace, les autres éléments essentiels en sont les droites ou "géodésiques". Mais par deux points du treillis ne passe pas nécessairement une droite (et une seule); et nous préférons axes ou génératrices à géodésiques.

L'existence d'un "côté" ou segment entre deux points du treillis est un fait concret; elle signifie qu'on passe du plan de sondage  $\pi$  au plan de sondage  $\pi'$  par une opération matérielle de tirage au sort (per-tes d'une ou plusieurs unités de sondage). En revanche le problème est de savoir si 3 points (et davantage) du treillis sont ou non alignés sur une même génératrice.

Il est naturel de le résoudre en introduisant sur le treillis une métrique, qui est l'"information"<sup>(1)</sup>

La distance entre deux points étant ainsi définie, l'alignement des 3 points A B C s'obtient en écrivant

$$AB + BC = AC$$

Mais on peut objecter que les axes du treillis dépendent alors de la métrique et peuvent changer quand on change de métrique.

Une autre difficulté est l'existence de certains triangles ABC sur le treillis, la distance AC étant alors égale à  $AB + BC$  sans qu'il y ait alignement.

#### Exemple de triangle sur le treillis.

Considérons un sondage stratifié avec deux strates d'effectifs  $(v_1, v_2)$ . Par exemple  $v_1 = 20, v_2 = 100$ . Au grillage relatif au sondage  $(n_1, n_2)$  se superpose un demi-axe relatif aux 20 sondages à fractions sondées égales  $f$ , soit :

$$f = 1/20, 1/10, 3/20, 1/5, 1/4, \text{ etc.}$$

$$n = n_1 + n_2 = 6, 12, 18, 24, 30, 36, \text{ etc.}$$

concernant les couples  $(n_1, n_2)$  suivants :

(1) L'information est une "valuation de treillis", Schützenberger, Thèse, p. 40.

(1, 5) (2, 10) (3, 15) (4, 20) (5, 25) (6, 30) etc.

On a donc par exemple le "triangle" suivant :

$$(3, 15) = (\pi); (3, 10) = (\pi'); (2, 10) = (\pi'')$$

C'est pourquoi on prendra les précautions suivantes :

a) Définition des arêtes telle qu'il n'existe plus de triangles.

On figurera par une arête du treillis tout tirage dans une seule urne, opération faisant passer d'un sommet à un autre, d'un plan de sondage à un autre de taille inférieure. Il est clair qu'on peut toujours (et au moins d'une façon) cheminer sur les arêtes, de l'origine (point figurant la population prise en totalité) à tout sommet du treillis.

b) Définition d'axes d'information invariants si la métrique change.

Par définition sont alignés les points représentant les plans de sondage ( $v, v-1, v-2, v-3 \dots$ ) tels qu'on expulse d'une urne donnée ( $0, 1, 2, 3, \dots$ ) unités de sondage, toutes choses égales d'ailleurs.

Ceci revient à dire que, quelle que soit la perte d'information, les plans de sondage, qui sont des tirages dans une urne unique, sont figurés par un axe.

Les plans de sondage (si compliqués soient-ils) étant des combinaisons de tirages dans une urne, on possède donc une définition "opérationnelle" de certains axes du treillis; l'existence d'autres axes reste possible.

4/ - Exemples de treillis de sondage.

a) Cas des sondages stratifiés (avec 2 strates seulement et 2 paramètres).

Le grillage ( $n_1, n_2$ ) représente (on l'a vu) un sondage à 2 strates (fractions sondées distinctes); le point  $U(v_1, v_2)$  représente l'état de connaissance parfaite des urnes. Les points représentant les sondages à fractions sondées égales  $f$  auraient en outre été alignés sur un demi-axe d'origine  $U$ ; mais il s'agit en fait d'un deuxième treillis distinct du premier, correspondant à un estimateur de  $x$  dépendant d'un seul paramètre  $f$ .

En convenant que les arêtes n'existent qu'avec les pertes d'une boule à la fois, nous perdons le deuxième treillis et les triangles qu'il dessinait avec le grillage.

La convention faite sur les axes nous montre que tous les "barreaux" du grillage sont axes d'information.

b) Cas des sondages à 2 degrés et 2 paramètres.

Soit un univers où les  $\mu$  unités du 1er degré comprennent des nombres égaux d'unités du 2ème degré : l'échantillon est défini par le nombre  $m$  d'unités du 1er degré tirées et le nombre  $\bar{n}$  d'unités du 2ème degré extraites de chacune des précédentes; on le représentera par les sommets du grillage ( $m, \bar{n}$ ).

Le point  $(\mu, \bar{v})$  correspond à une connaissance parfaite; l'axe  $(m, \bar{v})$  représente un sondage en grappes et l'axe  $(\mu, \bar{n})$  un sondage stratifié (à fractions sondées égales), ce sont deux axes d'information.

La définition opérationnelle des axes nous donne les axes ( $m$  constant,  $\bar{n}$  variable) (1er système).

Il existe un second procédé de sondage à 2 degrés qui consiste à tirer d'abord  $\bar{n}$  unités secondaires dans chaque unité primaire, puis à désigner au sort  $m$  sur  $\mu$  unités primaires (2<sup>e</sup> système d'axes).

Les deux procédés sont équivalents parce que ( $\bar{n}$  étant constant) les tirages du 1er et du 2ème degrés sont indépendants. On dispose donc ici encore de deux systèmes d'axes d'information croisés.

c) Cas d'un sondage à 2 phases (et 2 paramètres).

On tire d'abord  $m$  unités (avec ou sans remise) d'une urne d'effectif  $\mu$ . On les trie en trois strates d'effectifs  $m/3$  et on tire  $\bar{n}$  unités de chaque strate. Le treillis formé des points  $(m, \bar{n})$  admet, d'une part, l'axe  $(m, m/3)$ , d'autre part, les axes ( $m$  constant,  $\bar{n}$  variable). Il n'existe pas de second système d'axes croisé avec le premier.

d) Cas d'un sondage à 2 degrés avec dépendance du 2ème degré.

Si l'on sonde un univers où les unités primaires sont de tailles très inégales, on a parfois intérêt à se donner l'effectif total  $n$  au 2ème degré, à répartir entre les  $m$  unités du 1er degré échantillon, au prorata de leurs tailles  $v_i$ . (C'est un cas d'estimation biaisée).

On a ainsi  $n_i/n = v_i/S v_i$  et le 2ème degré de sondage dépend stochastiquement du 1er.

On ne peut tirer d'abord les unités du 2ème degré puis celles du 1er degré. Il n'existe qu'un système d'axes d'information, s'appuyant sur un axe unique transversal.

5/ - Opérations sur les treillis de sondage.

En passant en revue les problèmes courants de sondage, on constate qu'il est souvent maladroit de représenter directement les plans de sondage de paramètres  $f_1, f_2, f_3, \dots$  par le treillis  $(f_1, f_2, f_3, \dots)$ . Il peut être plus commode de considérer un treillis plus vaste : soit qu'on en découpe une partie seule utile, soit qu'on y figure les échantillons possibles avec leurs probabilités respectives, le treillis jouant alors un rôle analogue à celui de l'arbre défini au début.

Plus généralement, le treillis primitif sera découpé en un sous-ensemble des points exclus et divers autres sous-ensembles dont chacun représente un nouveau plan de sondage.

Donnons des exemples.

a) Sondage stratifié avec fractions de sondage égales f.

Du treillis  $(n_1, n_2, \dots, n_h)$  on conserve les points  $(f v_1, f v_2, \dots, f v_h)$ , sous-ensemble décrit quand  $f$  (ou  $n = f v$ ) varie.

b) Sondage stratifié "optimum".

Du même treillis on conserve les points  $\left( \frac{\lambda v_1 \sigma_1}{\sqrt{c_1}}, \frac{\lambda v_2 \sigma_2}{\sqrt{c_2}}, \dots, \frac{\lambda v_h \sigma_h}{\sqrt{c_h}} \right)$  sous-ensemble décrit quand  $\lambda$  (ou  $n$ ) varie (avec  $n = \lambda \sum v_h \sigma_h / \sqrt{c_h}$ ).

c) Stratification a posteriori.

Le même treillis est découpé en sous-ensembles tels que  $\sum n_h = n$ .

Chaque point d'un sous-ensemble donné est affecté d'une probabilité hypergéométrique (tirage exhaustif) ou multinomiale (tirage bernoullien).

Toutefois, on aura pris soin d'éviter qu'aucun des points  $(n_h)$  ne corresponde à une perte d'information infinie introduite artificiellement par un estimateur inadéquat. L'apparition de certains  $n_h$  nuls est un événement de probabilité non nulle; et avant tirage de l'échantillon il faut convenir de l'estimation à employer si cela se produit.

## B - PASSAGE DE L'ESPACE TOPOLOGIQUE A L'ESPACE METRIQUE.

1/ - Perte d'information.

On attache à présent à chaque sommet du treillis une expression  $p$  - la perte d'information -

- nulle au sommet U représentant la connaissance parfaite de l'univers sondé;
- infinie aux points représentant un état d'absence complète de connaissance;
- satisfaisant à la relation de récurrence, à savoir :

$$p(\pi') = p(\pi) + \mathcal{E}P(\pi'/\pi)$$

- en désignant par  $\pi$  et  $\pi'$  2 plans de sondages représentés par 2 points (ou sommets) du treillis non pas quelconques mais reliés par une arête (ou côté);
- et par  $P(\pi'/\pi)$  une expression qui se réduirait à  $p(\pi')$  si l'échantillon du plan de sondage ( $\pi$ ) recouvrait toute la population; autrement dit :

$$P(\pi'/U) = p(\pi')$$

En employant une majuscule P, nous entendons spécifier qu'il s'agit d'une grandeur aléatoire.

Quand on écrit  $\mathcal{E}P(\pi'/\pi)$ , on désigne l'espérance mathématique de  $P(\pi'/\pi)$  étendue à tous les échantillons du plan de sondage ( $\pi$ ).

Partant du point U, et cheminant exclusivement le long des côtés du treillis, on obtient finalement l'information perdue par chaque plan de sondage de la famille; lorsque ( $\pi$ ) peut être atteint par plusieurs itinéraires, la fonction de treillis  $p(\pi)$  a (par hypothèse) une valeur indépendante de l'itinéraire (à condition qu'il s'agisse bien du même plan de sondage  $\pi$ ).

## 2/ - Métrique.

Simultanément on introduit une métrique dans cet espace, en définissant la distance entre deux points situés sur une même géodésique comme égale à la variation de la perte d'information.

$$(\text{distance } \pi\pi') = p(\pi') - p(\pi) = \mathcal{E}P(\pi'/\pi)$$

## 3/ - Conséquence.

Relation entre les informations perdues par 3 plans  $\pi$   $\pi'$   $\pi''$  de la famille de plans de sondage considérée :

Désignant par ( $\pi$ ) ( $\pi'$ ) ( $\pi''$ ) 3 plans de sondage de la famille considérée, représentés par trois sommets du treillis reliés deux à deux par des côtés du treillis,

on a 
$$p(\pi') = p(\pi) + \mathcal{E}P(\pi'/\pi)$$

$$p(\pi'') = p(\pi') + \mathcal{E}P(\pi''/\pi')$$

donc 
$$p(\pi'') = p(\pi) + \mathcal{E} \left[ P(\pi''/\pi') + P(\pi'/\pi) \right]$$

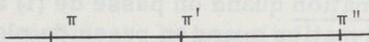
Mais si  $\pi''$  et  $\pi$  sont également sur un côté, on a :

$$p(\pi') - p(\pi) = \mathcal{E}P(\pi'/\pi)$$

D'où :

$$\mathcal{E}P(\pi''/\pi) = \mathcal{E}P(\pi''/\pi') + \mathcal{E}P(\pi'/\pi)$$

Cette condition signifie (en l'absence de triangles sur le treillis) que  $\pi$   $\pi'$   $\pi''$  sont alignés sur un axe d'information portant à la fois le côté  $\pi\pi'$ , le côté  $\pi'\pi''$  et le côté  $\pi\pi''$ .



On sait que cette relation doit subsister quel que soit P.

4/ - Relation entre  $p(\pi)$  et  $p(\pi'')$ .

$\pi$  et  $\pi''$  étant quelconques, on aura :

$$p(\pi'') = p(\pi) + \mathcal{E}P(\pi'_1/\pi) + \mathcal{E}_1P(\pi'_2/\pi'_1) + \dots + \mathcal{E}_kP(\pi''/\pi'_k)$$

à condition qu'on puisse s'arranger pour que  $\pi$   $\pi'_1$   $\pi'_2$  ...  $\pi'_k$   $\pi''$  forme une ligne brisée composée de portions d'axe d'information.

Application.

De la famille F des plans ( $\pi$ ) représentée par le treillis, extrayons une suite de plans de sondage 1, 2, 3, ... n, telle que "la taille de l'échantillon décroisse constamment" et que  $Z_i$   $Z_{i+1}$  (consécutifs) soient toujours sur un même axe d'information. On a par conséquent :

$$(\zeta Z_1) + \mathcal{E}_1(Z_1Z_2) + \mathcal{E}_1\mathcal{E}_2(Z_2Z_3) + \dots + \mathcal{E}_1 \dots \mathcal{E}_{n-1}(Z_{n-1}Z_n) \equiv (\zeta Z_n)$$

En posant :

$$\mathcal{E}_1 \dots \mathcal{E}_i (Z_i Z_{i+1}) = \overline{Z_i Z_{i+1}}$$

la formule s'écrit :

$$\overline{\zeta Z_1} + \overline{Z_1 Z_2} + \overline{Z_2 Z_3} + \dots + \overline{Z_{n-1} Z_n} = \overline{\zeta Z_n}$$

C'est la formule de Chasles (généralisée).

#### Conventions.

a) Jusqu'ici d'après nos conventions, les termes  $\overline{\zeta Z_1}$ ,  $\overline{Z_1 Z_2}$ , etc. sont positifs.

Pour pouvoir parler de relation de Chasles, il est nécessaire de définir aussi les termes négatifs. On va donc convenir que :

$$\text{si } \overline{Z_i Z_j} \text{ est négatif, il est égal à } - \overline{Z_j Z_i}$$

où  $\overline{Z_j Z_i}$  (positif) désigne  $\mathcal{E}_1 \dots \mathcal{E}_j (Z_j Z_i)$ .

#### b) Convention de langage.

On appellera  $\overline{Z_j Z_i}$  (positif) perte d'information quand on passe de (j) à (i) et gain d'information quand on passe de (i) à (j). En particulier ( $\zeta Z_1$ ) est le gain d'information quand on passe du plan d'échantillonnage (1) à une connaissance complète de la population.

#### Conséquences.

a) La "formule de Chasles" s'étend formellement à tous les plans de sondages de F, sans qu'il soit encore nécessaire de supposer : les tailles décroissantes, les plans de sondage consécutifs placés sur une même géodésique ; mais les symboles  $\overline{Z_i Z_1}$  perdent leur signification primitive.

b) Le long de tout circuit fermé, s'il en existe sur les axes du treillis de F, on obtiendra :

$$\overline{Z_1 Z_2} + \overline{Z_2 Z_3} + \dots + \overline{Z_k Z_1} = 0$$

#### Conclusion.

La perte d'information apparaît comme la somme  $\sum X_h$  de composantes dont chacune est la "longueur" d'une portion d'axe du treillis.

En étudiant des plans de sondage concrets, on découvrira les conséquences de l'existence de circuits sur le treillis (sondage à plusieurs degrés) et on s'apercevra de l'existence de composantes sous-tractives (sondage à plusieurs phases).

Mais, au préalable, on va chercher (au Ch. IV) à résoudre l'équation fonctionnelle de Chasles, c'est-à-dire à connaître la forme analytique des P.

Remarque. Les changements de découpage et l'information.

Il est en pratique très fréquent que les  $\zeta$  à estimer soient indépendants du découpage : c'est, par exemple, le cas si on veut estimer la superficie en blé, en France.

La perte d'information est définie à l'aide d'un treillis, qui traduit une certaine structure de sondage. Le treillis se conserve si l'on garde la structure tout en modifiant le découpage du territoire en unités, en strates, ... ; les valeurs de la perte d'information et la représentation métrique du sondage se modifient; ceci traduit les changements internes de l'estimateur (apparemment immuable), pour l'estimation d'un paramètre qui, lui, est invariant dans les transformations du découpage.

Supposons à présent qu'au lieu de la superficie en blé on veuille estimer la variance  $\xi$  entre unités primaires de sondage;  $\xi$  n'est pas invariante quand on modifie le découpage du territoire en unités de sondage.

On estime simultanément : la superficie en blé  $\zeta$  (à l'aide de  $Z$ ), et la perte d'information  $\varphi Z$  (à l'aide d'un certain  $W$ ); en supposant  $Z$  sans biais. L'expression de  $\varphi Z$  renferme  $\xi$ . On s'intéresse parfois à la perte d'information relative à  $W$  (c'est-à-dire une perte d'ordre 2). Lorsque le découpage est modifié, les variations de cette dernière perte ne signifient plus rien : car  $\varphi Z$  et  $W$  varient simultanément, leur "distance" se modifie sous une double cause.  $Z$  estime sans biais  $\zeta$  malgré le découpage changé, alors qu'on ne sait plus trop quel  $\varphi Z$  est estimé par  $W$ .

Nous pensons donc que la théorie de la perte d'information doit être faite avec un ou plusieurs découpages bien déterminés, qu'elle ne supporte pas les changements de découpage.

Application aux sondages à plusieurs degrés.

Empruntons à une étude encore inédite de M. Fonsagrive un cas intéressant de sondages à plusieurs degrés. On suppose que (comme c'est souvent le cas en pays sous-développé) le découpage des unités du 2ème degré et la sous-stratification n'ont lieu que dans les unités échantillon du 1er degré, en fonction des singularités que présentent ces dernières et en vue de réduire le plus possible la variance (à l'intérieur de l'échantillon du 1er degré); ainsi la vraie structure de sondage reste-t-elle en grande partie indéterminée.

Sans doute est-il encore permis de parler de l'ensemble des découpages possibles, et de concevoir un estimateur sans biais, sa va-

riance et l'estimateur sans biais de cette variance pour chacun de ces découpages; mais l'expérience pratique se limite à un seul découpage, en partie embryonnaire, et un seul échantillon; et la variance ainsi estimée manque de signification précise vis-à-vis de toute la partie non spécifiée du découpage.

#### Autres applications.

Avec les procédés courants de stratification a posteriori ou de sondage à deux phases, on rencontre la même difficulté : le découpage en strates n'est arrêté qu'après tirage de l'échantillon, ou du grand échantillon. Si ledit tirage avait donné un autre résultat, on aurait modifié les strates et, par conséquent, la variance d'échantillonnage. L'estimation de cette variance, sur l'échantillon effectivement tiré, a donc une signification limitée.

Pour qu'il en soit autrement, il faudrait (pour éviter toute difficulté) se donner à l'avance des règles prévoyant (dans tous les cas possibles) le découpage en strates à effectuer et l'estimateur à employer .

Nous nous en tiendrons ici à des plans de sondage déterminés par un nombre fini de paramètres, c'est-à-dire à des cadres rigides de sondage.

## CHAPITRE IV

### LA FORME ANALYTIQUE DES PERTES D'INFORMATION

On va rechercher quelles formes les expressions  $p(\pi)$  et  $P(\pi'/\pi)$  sont susceptibles de prendre lorsqu'on ne sait absolument rien de la forme de la distribution sondée, hypothèse très réaliste en théorie des sondages.

#### I - PERTES AU SENS DU CHAPITRE I -

##### 1/ - Introduction.

Au sens du Chapitre I, la perte d'information concerne plus spécialement l'estimation d'une certaine caractéristique  $\zeta$  par un estimateur  $Z$ , supposé sans biais (appréciable).

On va supposer la perte d'information de la forme  $\mathcal{E} W(Z)$  c'est-à-dire fonction de risque (au sens de Wald); cherchons quelles conditions doit satisfaire  $W$  pour que la fonction de risque  $\mathcal{E} W(Z)$  soit perte d'information.

Par hypothèse on a  $W(\zeta) = 0$ ,  $W(Z) > 0$  si  $Z \neq \zeta$ .

On supposera d'abord  $W$  fonction de  $Z$ ; on sera amené ensuite à considérer  $W$  comme fonctionnelle de  $Z$ . Cette fonction de  $Z$  sera supposée convexe; mais nous tenons à conserver le cas très important où  $W = |Z - \zeta|$ ; il ne faut donc pas supposer  $W$  "strictement convexe".

Le problème est à présent de trouver  $W$  tel que  $\mathcal{E} W$  satisfasse à l'équation de Chasles.

Il existe des solutions évidentes de cette équation fonctionnelle qui peut s'écrire :

$$\text{avec } \begin{cases} \mathcal{E} W(Z_1, \zeta) = \sum_h p_1^h W(z_1^h, \zeta) \\ \mathcal{E} W(Z_2, \zeta) = \sum_k p_2^k W(z_2^k, \zeta) \end{cases}$$

$$\sum_k p_2^k W(z_2^k, \zeta) = \sum_h p_1^h W(z_1^h, \zeta) + \sum_h p_1^h \sum_k p_2^k W(z_2^k, z_1^h)$$

2/ - Une solution particulière symétrique de l'équation de Chasles.

Ne retenons que les cas où il est possible d'invertir  $z_1^h$  et  $z_2^k$  dans les formules.

En effet il serait gênant qu'on ait :

$$\sum_h \sum_k p_1^h p_2^k W(z_1^h, z_2^k) = \overline{Z_1 Z_2}$$

alors que  $\sum_k \sum_h p_2^k p_1^h W(z_2^k, z_1^h)$  aurait une autre signification. Ainsi le cas particulier où  $W(z, \zeta)$  est symétrique semble-t-il présenter le plus d'intérêt.

Les exemples de fonction  $W$  donnés par Basu sont justement des fonctions symétriques en  $\zeta$  et  $Z$ .

On a alors par définition :

$$\overline{Z_1 Z_2} = \begin{cases} \mathcal{E}W(Z_2, Z_1) & \text{si Taille (1) > Taille (2)} \\ \mathcal{E}W(Z_1, Z_2) & \text{si Taille (2) > Taille (1)} \end{cases}$$

(le cas où les tailles ne sont pas directement comparables est exclu).

a) L'équation de Chasles

$$\mathcal{E}W(Z_2, \zeta) = \mathcal{E}W(Z_1, \zeta) + \mathcal{E}W(Z_2, Z_1)$$

admet comme solution particulière :

$$W(Z, \zeta) = (Z - \zeta)^2$$

à condition que :

$$\mathcal{E}[(\zeta Z_1)(Z_1 Z_2)] = 0 ;$$

: dans l'espace (A) où on représente l'aléatoire  $Z_1$  par un point, les vecteurs

$$\overrightarrow{\zeta Z_1} \text{ et } \overrightarrow{Z_1 E Z_2}$$

doivent donc être orthogonaux; à moins qu'on n'ait (estimateur uniformément sans biais)

$$E Z_2 = Z_1 \quad (\text{donc } \mathcal{E} Z_1 = \zeta)$$

b) On étend le résultat à :

$$W(Z, \zeta) = [g(z) - g(\zeta)]^2$$

où  $g(z)$  est une fonction quelconque; toujours sous réserve qu'on ait :

$$\mathcal{E} \{ [g(Z_1) - g(\zeta)] [g(Z_2) - g(Z_1)] \} = 0$$

3/ - Une solution non symétrique de l'équation de Chasles.

L'absence de biais de l'estimateur  $Z$  correspond à la condition

$$2 \mathcal{E}(\zeta - Z) = 0$$

c'est-à-dire  $\frac{\partial}{\partial \zeta} \mathcal{E}(\zeta - Z)^2 = 0$ , c'est-à-dire  $\frac{\partial}{\partial \zeta} \mathcal{E}W = 0$  avec  $W = (\zeta - Z)^2$ .

On peut écrire aussi

$$\mathcal{E}W \equiv \mathcal{E}Z^2 - 2\zeta \mathcal{E}Z + \zeta^2 \equiv \underline{\underline{\mathcal{E}Z^2 - \zeta^2}}$$

Sous cette dernière forme, il est clair que la relation de récurrence sera vérifiée.

De même l'absence de biais peut s'écrire encore :

$$\mathcal{E}[\varphi(\zeta)(\zeta - Z)] = 0$$

Posons  $W(Z, \zeta) = \int \varphi(\zeta)(\zeta - Z)d\zeta$  avec

$$\int \varphi(\zeta)d\zeta = \varphi_1(\zeta); \quad \int \zeta \varphi(\zeta)d\zeta = \varphi_2(\zeta)$$

$$W \equiv \varphi_2(\zeta) - \varphi_2(Z) - Z [\varphi_1(\zeta) - \varphi_1(Z)]$$

$$\mathcal{E}W \equiv \mathcal{E} [Z \varphi_1(Z) - \varphi_2(Z)] - [\zeta \varphi_1(\zeta) - \varphi_2(\zeta)]$$

Là encore il apparaît que la relation de récurrence sera vérifiée.

1ere application :  $\varphi(\zeta) = 2\zeta$  avec  $\zeta > 0$ ,

$$\text{d'où} \quad W = (\zeta - Z)^2 \left( \zeta - \frac{\zeta - Z}{3} \right) \equiv \underline{\underline{\frac{1}{3} (2\zeta^3 - 3\zeta^2 Z + Z^3)}}$$

$$\mathcal{E}W = \frac{1}{3} (2\zeta^3 - 3\zeta^2 \mathcal{E}Z + \mathcal{E}Z^3)$$

$$\mathcal{E}W \equiv \frac{1}{3} (\mathcal{E}Z^3 - \zeta^3)$$

2ème application.  $\phi(\zeta) = \zeta^2$

$$\text{d'où} \quad W = (\zeta - Z)^2 \zeta^2 - \frac{1}{2} (\zeta - Z)^4 - \frac{2Z}{3} (\zeta - Z)^3$$

$$= \frac{1}{4} \zeta^4 - \frac{2}{3} \zeta^3 Z + \frac{1}{6} Z^4$$

$$\mathcal{E}W \equiv \frac{1}{6} (\mathcal{E}Z^4 - \zeta^4)$$

Ces exemples mettent sur la voie de résultats plus généraux.

#### 4/ - THEOREME I -

$\mathcal{E}W$  est nécessairement de la forme  $\mathcal{E}\phi(Z) - \phi(\zeta)$ , (où  $\phi$  est une fonction), que l'estimateur  $Z$  soit sans biais ou qu'il soit biaisé, - du moment de  $W$  est fonction de  $\zeta$ .

En effet partons de la relation de récurrence des pertes d'information :

$$(1) \quad \mathcal{E}W(Z_2, \zeta) - \mathcal{E}W(Z_1, \zeta) = \mathcal{E}EW(Z_2, Z_1).$$

On voit que le second membre ne dépend pas de  $\zeta$ ; il en résulte que  $\mathcal{E}W(Z_1, \zeta) = f_1(\zeta)$  et  $\mathcal{E}W(Z_2, \zeta) = f_2(\zeta)$  sont nécessairement de la forme  $k_1 - \phi(\zeta)$  et  $k_2 - \phi(\zeta)$ ,  $\phi(\zeta)$  s'éliminant par différence.

Si l'on substitue  $Z_1$  à  $\zeta$  dans  $W(Z_2, \zeta)$ , il suit de là que

$$\mathcal{E}W(Z_2, \zeta) = k_2 - \phi(\zeta)$$

entraîne  $EW(Z_2, Z_1) = K_{21} - \phi(Z_1)$

tandis que  $\mathcal{E}W(Z_1, \zeta) = \mathcal{E}W(Z_2, \zeta) - \mathcal{E}EW(Z_2, Z_1)$

implique  $\mathcal{E}W(Z_1, \zeta) = k_2 - \phi(\zeta) - [\mathcal{E}K_{21} - \mathcal{E}\phi(Z_1)]$

ou  $\mathcal{E}W(Z_1, \zeta) = [k_2 - \mathcal{E}K_{21}] + \mathcal{E}\phi(Z_1) - \phi(\zeta)$

où le crochet comprend deux termes dépendant du sondage n°2.

Mais par hypothèse  $W$  est fonction de  $\zeta$ , donc  $\mathcal{E}W(Z_1, \zeta)$  ne dépend pas des sondages (2), d'où l'obligation pour  $k_2 - \mathcal{E}K_{21}$  de disparaître.

Donc  $\mathcal{E}W$  est nécessairement de la forme  $\mathcal{E}\phi(Z) - \phi(\zeta)$ . cqfd.

Réciproque du Théorème I.

$$\begin{cases} \mathcal{E}W(Z_1, \zeta) = \mathcal{E}\phi(Z_1) - \phi(\zeta) \\ \mathcal{E}W(Z_2, Z_1) = \mathcal{E}\phi(Z_2) - \phi(Z_1) \end{cases}$$

$$\begin{aligned} \mathcal{E}W(Z_1, \zeta) + \mathcal{E}W(Z_2, Z_1) &= \mathcal{E}\phi(Z_2) - \phi(\zeta) \\ &= \mathcal{E}W(Z_2, \zeta) \end{aligned}$$

A condition que  $[\mathcal{E}\phi(Z) - \phi(\zeta)]$  ait un signe constant (qu'on rend au besoin positif en changeant  $\phi$  de signe) il s'en suit bien que c'est une perte d'information. Mais la restriction qu'on vient de formuler a beaucoup d'importance.

Remarque 1.

$[\mathcal{E}\phi(Z) - \phi(\zeta)]$  est le biais quand on estime  $\phi(\zeta)$  par  $\phi(Z)$ , [alors que  $Z$  estime  $\zeta$  sans biais].

Remarque 2:

Par exemple de la relation  $\mathcal{V}(\bar{X}) = \mathcal{E}(\bar{X}^2) - \bar{x}^2$ , on déduit  $\mathcal{E}(\bar{X}^2) - \bar{x}^2 \geq 0$  ou  $\phi(Z) - \phi(\zeta) \geq 0$  avec  $\zeta = \bar{x}$ ,  $\phi(Z) = Z^2$ ,  $Z = \bar{X}$ .

Plus généralement : toute fonction monotone convexe  $\phi$  satisfait à notre condition de signe; la distribution des points (avec  $\mathcal{E}Z = \zeta$ )  $[Z, \phi(Z)]$  admet le point  $[\zeta, \mathcal{E}\phi(Z)]$  comme barycentre, point situé dans la concavité de la courbe.

C'était le cas pour  $\phi = \zeta^3$  (avec  $\zeta > 0$ ) et  $\phi = \zeta^4$ , aux exemples du n°3.

Le cas des estimateurs biaisés sera étudié au paragraphe II ci-après.

Remarque 3.

Si  $W$  est une fonction dérivable en  $\zeta$ , on peut poser  $\omega = \partial W / \partial \zeta$  et dériver les deux membres de l'équation de Chasles par rapport à  $\zeta$  en posant

$$\mathcal{E}W = \sum_{\top} p^i W(z^i, \zeta)$$

et en traitant les  $p^i$  comme des constantes.

On a d'abord

$$\sum_k p_2^k \omega(\zeta, z_2^k) = \sum_h p_1^h \omega(\zeta, z_1^h)$$

d'où il suit que l'expression :

$$\sum_i p_1^i \omega(\zeta, z_j^i), \quad \text{avec Prob}(Z_j = z_j^i) = p_j^i$$

doit être indépendante du rang  $j$  que l'estimateur  $Z_j$  occupe dans la suite  $\zeta Z_1 Z_2 \dots$  d'estimateurs relatifs à des échantillons de tailles décroissantes. Autrement dit :

$$\sum_i p_j^i \omega(\zeta, z_j^i) = f(\zeta)$$

Supposons  $f(\zeta) \neq 0$  :

Posons : 
$$-\phi(\zeta) = \int f(\zeta) d\zeta$$

et intégrons la dérivée de l'équation de Chasles (avec les  $p^i$  constants). Il viendra

$$\sum_k p_2^k W(\zeta, z_2^k) = \int f(\zeta) d\zeta - a_2 = -\phi(\zeta) - a_2 \quad (\text{et analogue})$$

avec 
$$-\sum_h p_1^h \phi(z_1^h) = a_1 \quad (\text{et analogue})$$

d'où : 
$$-\phi(\zeta) + \sum_k p_2^k \phi(z_2^k) = -\phi(\zeta) + \sum_h p_1^h \phi(z_1^h) + \sum_h p_1^h [-\phi(z_1^h) + \sum_k p_2^k \phi(z_2^k)]$$

donc : 
$$\mathcal{E} W(\zeta, Z) = \mathcal{E} \phi(Z) - \phi(\zeta)$$

On retrouve le théorème I. Il faut noter pourtant que  $f(\zeta)$  n'est pas  $\mathcal{E} \omega$ , car on n'a pas le droit de remplacer

$$\frac{\partial}{\partial \zeta} \mathcal{E} W \quad \text{par} \quad \mathcal{E} \frac{\partial W}{\partial \zeta} .$$

En effet : si par exemple  $W$  est de la forme

$$A_0(\zeta) + A_1(\zeta)Z + A_2(\zeta)Z^2 + \dots$$

on aurait

$$\frac{\partial}{\partial \zeta} \mathcal{E} W = A_0'(\zeta) + A_1'(\zeta) \mathcal{E} Z + A_2'(\zeta) \mathcal{E} Z^2 + \dots + A_1(\zeta) \frac{\partial}{\partial \zeta} (\mathcal{E} Z) + A_2(\zeta) \frac{\partial}{\partial \zeta} (\mathcal{E} Z^2) + \dots$$

contre 
$$\mathcal{E} \frac{\partial W}{\partial \zeta} = A_0'(\zeta) + A_1'(\zeta) \mathcal{E} Z + A_2'(\zeta) \mathcal{E} Z^2 + \dots$$

avec par exemple :  $\mathcal{E} Z = \zeta$ , donc  $\frac{\partial}{\partial \zeta} (\mathcal{E} Z) = 1$ , etc.

On a supposé  $f(\zeta) \neq 0$ .

Au contraire on a  $\mathcal{E} \omega = 0$ , c'est-à-dire  $\mathcal{E}(\zeta - Z) = 0$  (absence de biais).

Finalement il est superflu de supposer  $W$  dérivable; et la première démonstration du théorème I est moins discutable que celle-ci.

4/ - THEOREME II a -

La perte d'information étant  $\mathcal{E}W$ , où  $W$  est une fonction de  $Z$  et  $\zeta$ , il existe nécessairement entre estimateur  $Z$  et estimé  $\zeta$  une relation de la forme

$$\mathcal{E} \Delta(Z, \zeta) = 0$$

En effet posons  $W(Z, \zeta) = \Delta(Z, \zeta) + \phi(Z) - \phi(\zeta)$  avec le  $\phi$  du théorème I ; il vient de la sorte

$$\mathcal{E} \Delta(Z, \zeta) = 0 \quad \text{cqfd.}$$

Cas particulier :  $\mathcal{E}Z - \zeta = 0$  (estimation sans biais).

THEOREME II b -

S'il existe entre estimateur  $Z$  et estimé  $\zeta$  une relation de la forme

$$\mathcal{E} \omega(Z, \zeta) = 0$$

on a nécessairement  $\Delta(Z, \zeta) \equiv \psi(\zeta) \cdot \omega(Z, \zeta)$  ( $\Delta$  étant celui défini au théorème II (a), sous la double condition :

- que la perte d'information soit suffisamment régulière;
  - que la distribution sondée soit suffisamment générale.
- (condition qui sera supposée remplie dans les numéros suivants sauf avis contraire).

En effet si par exemple  $\omega$  et  $\Delta$  sont deux fonctions de  $Z$  développables en série entière

$$\omega = \sum_0^{+\infty} a_i Z^i, \quad \Delta = \sum_0^{+\infty} b_i Z^i$$

les deux équations en  $\zeta$

$$\begin{cases} \mathcal{E} \omega \equiv \sum_0^{+\infty} a_i \mathcal{E}(Z^i) = 0, \\ \mathcal{E} \Delta \equiv \sum_0^{+\infty} b_i \mathcal{E}(Z^i) = 0, \end{cases}$$

déterminent le même  $\zeta$  si (et seulement si)  $b_i/a_i$  est une constante, soit  $\psi(\zeta)$  (en supposant donnés la distribution à sonder, le plan de sondage et l'estimateur et en excluant toute relation particulière entre  $\mathcal{E}Z$ ,  $\mathcal{E}Z^2$ ,  $\mathcal{E}Z^3$ , etc.).

Cas particulier :  $\Phi(\zeta) = \zeta^2$ ;

$$\Delta = (Z - \zeta)^2 + \zeta^2 - Z^2 = 2\zeta^2 - 2\zeta Z = 2\zeta(\zeta - Z)$$

$$\omega = \zeta - Z ; \psi = 2\zeta$$

#### 5/- Conséquence : Relations entre pertes conjuguées d'information

Lorsqu'on permute  $\zeta$  et  $Z$ , la perte d'information n'est pas conservée si  $W$  n'est pas symétrique en  $Z$  et  $\zeta$  ; on dira que ces deux pertes d'information sont conjuguées.

Désignant par une astérisque les fonctions où  $Z$  et  $\zeta$  sont permutés, on a ( $\psi$  désignant  $\psi(\zeta)$  et  $\psi^*$  désignant  $\psi(Z)$ ) :

$$W = \Phi(Z) - \Phi(\zeta) + \Delta$$

$$W^* = \Phi(\zeta) - \Phi(Z) + \Delta^*$$

$$\text{d'où} \quad W + W^* = \Delta + \Delta^* = \psi\omega + \psi^*\omega^*$$

$$W - W^* = 2[\Phi(Z) - \Phi(\zeta)] + (\Delta - \Delta^*)$$

Le cas où  $W (= W^*)$  est symétrique est plus spécialement intéressant. On a alors :

$$\Phi(\zeta) - \Phi(Z) = (\Delta - \Delta^*)/2 = (\psi\omega - \psi^*\omega^*)/2$$

$$W = (\Delta + \Delta^*)/2 = (\psi\omega + \psi^*\omega^*)/2$$

#### 6/- Résolution de l'équation de Chasles avec un estimateur sans biais

#### THEOREME III -

a) Si  $W$  est symétrique, on a

$$W = \lambda(Z - \zeta)^2, \quad \Phi = \lambda\zeta^2$$

En effet on a  $\omega = \zeta - Z$  ; d'où :  $\omega^* = -\omega$

$$\Phi(\zeta) - \Phi(Z) \equiv (\psi\omega - \psi^*\omega^*)/2$$

ou 
$$2\Phi(\zeta) - 2\Phi(Z) = (\zeta - Z)\psi(\zeta) - (Z - \zeta)\psi(Z)$$

c'est-à-dire :

$$[2\Phi(Z) - Z\psi(Z)] + [\zeta\psi(Z) - Z\psi(\zeta)] - [2\Phi(\zeta) - \zeta\psi(\zeta)] = 0$$

où chacun des crochets doit s'annuler séparément à cause de la grande généralité imposée aux distributions (donc à celle de Z); d'où :

$$2 \frac{\Phi(Z)}{Z} = \psi(Z); \quad \frac{\psi(Z)}{Z} = \frac{\psi(\zeta)}{\zeta}; \quad 2 \frac{\Phi(\zeta)}{\zeta} = \psi(\zeta)$$

d'où

$$\frac{\Phi(Z)}{Z^2} = \frac{\Phi(\zeta)}{\zeta^2}$$

imposé quels que soient Z et  $\zeta$ , d'où

$$\Phi(Z) = \lambda Z^2, \quad \Phi'(\zeta) = \lambda \zeta^2, \quad \psi(\zeta) = 2\zeta$$

$$W \equiv [2\lambda\zeta(\zeta - Z) + 2\lambda Z(Z - \zeta)]/2 \equiv \lambda(Z - \zeta)^2 \quad \text{cqfd.}$$

b) Si W est asymétrique, W est nécessairement de la forme

$$W = \Phi(Z) - \Phi(\zeta) + \Phi'(\zeta) \cdot (\zeta - Z)$$

où  $\Phi$  est arbitraire, à part la condition  $\Phi'' > 0$ .

En effet  $\partial W/\partial Z = \Phi'(Z) - \psi(\zeta)$  doit être nul pour  $Z = \zeta$ ; et  $\partial^2 W/\partial Z^2 = \Phi''(Z)$  doit être positif, de façon à s'assurer que  $W(Z, \zeta)$  autour de  $W(\zeta, \zeta) = 0$ , est positif et croît avec  $|Z - \zeta|$ <sup>(1)</sup>. Il suit de là que

$$(\partial W/\partial Z)_0 = \Phi'(\zeta) - \psi(\zeta) = 0,$$

d'où le théorème énoncé.

7/ - Extension immédiate à un certain estimateur biaisé (moyennes à la Gini).

f étant une fonction arbitraire, substituant  $f(\zeta)$  à  $\zeta$  et  $f(Z)$  à  $Z$ ,

(1) Nota :  $\Phi'' > 0$  équivaut (lorsque  $\Phi''$  existe) à la convexité requise pour  $\Phi$ ; voir réciproque du théorème I.

il vient les expressions suivantes de  $\omega$ , de  $W$  et de  $\phi$  dans le cas où  $W$  est symétrique en  $Z$  et  $\zeta$ , - de  $W$  dans le cas où  $W$  n'est pas symétrique (avec  $\phi'' > 0$ )<sup>(1)</sup> :

$$\omega = f(\zeta) - f(Z)$$

$$a) \quad W = W^* = \lambda [f(Z) - f(\zeta)]^2, \quad \phi(\zeta) = \lambda [f(\zeta)]^2$$

$$b) \quad W = \phi[f(Z)] - \phi[f(\zeta)] + \phi'[f(\zeta)] \cdot [f(\zeta) - f(Z)]$$

## II - LES PERTES D'INFORMATION DANS LE CAS DES ESTIMATIONS BIAISEES -

### 1/ - Introduction.

On vient d'étendre à l'estimation biaisée de  $\zeta$  par  $Z$ , fruit de l'estimation sans biais de  $f(\zeta)$  par  $f(Z)$ , le théorème sur la perte de l'estimation sans biais. Ceci ne présente qu'un intérêt pratique médiocre, car les estimateurs biaisés les plus employés ne sont pas de cette forme; mais l'intérêt est d'ordre théorique.

On va voir que les moyens mis en œuvre jusqu'ici sont alors en échec. Si la distribution de  $Z$  n'est absolument pas particularisée et si les fonctions considérées sont parfaitement régulières (et d'abord si ce sont bien des fonctions, non des fonctionnelles) le théorème qui suit montre qu'il n'y a rien au-delà de l'estimation sans biais de  $f(\zeta)$  par  $f(Z)$ .

### THEOREME 1 -

Pour qu'une fonction  $W(Z, \zeta)$  existe, il est nécessaire que  $\mathcal{E}\omega = 0$  soit de la forme

$$\psi(\zeta) [\mathcal{E}f(Z) - f(\zeta)] = 0$$

En effet  $\omega(Z_1, \zeta) = \Omega$  s'écrit  $Z_1 = g(\Omega, \zeta)$

$$\omega(Z_2, Z_1) = \Omega_1 \quad \text{s'écrit} \quad Z_2 = h(\Omega_1, Z_1) = h_1(\Omega_1, \Omega, \zeta)$$

d'où  $\omega(Z_2, \zeta) = h_2(\Omega_1, \Omega, \zeta)$

Or  $\mathcal{E}\omega(Z_2, Z_1) = \mathcal{E}\Omega_1 = 0$  détermine  $Z_1$  en fonction de la distribution de  $Z_2$ ; de même  $\mathcal{E}\omega(Z_1, \zeta) = \mathcal{E}\Omega = 0$  détermine  $\zeta$ . Ces deux relations entraînent donc

$$\mathcal{E}\omega(Z_2, \zeta) = 0$$

(1) Voir note de la page précédente.

Ceci n'est possible que si l'on a :

$$\omega(Z_2, \zeta) = \lambda_1 \Omega_1 + \lambda \Omega$$

(en supposant  $h_2$  suffisamment régulière et les distributions de  $\Omega_1$  et  $\Omega$  assez générales); soit

$$\omega(Z_2, \zeta) = \lambda_1 \omega(Z_2, Z_1) + \lambda \omega(Z_1, \zeta)$$

qui n'est possible qu'avec  $\lambda = \lambda_1 = 1$  (faire  $\zeta = Z_1$  ou  $Z_1 = Z_2$ ); ce qui implique

$$\omega(Z_2, Z_1) = f(Z_2) - f(Z_1) \quad \text{cqfd.}$$

## 2/ - Cas où la fonction $W(Z)$ ne serait pas fonction de $\zeta$ .

Puisqu'on renonce à se limiter aux  $W$  symétriques, il est naturel de supposer que  $W$  est seulement fonction de  $Z$ , avec  $W(\zeta) = 0$ . D'ailleurs le texte de Basu cité à la fin du chapitre III, § I nous y invite. On constate alors ce qui suit.

### THEOREME 2.

Si  $W$  n'est pas fonction de  $\zeta$ , et est fonction continue et dérivable de  $Z$ , l'équation de Chasles est vérifiée, à la condition (nécessaire et suffisante) que  $W(z)$  soit une constante.

En effet la relation de Chasles prend alors la forme :

$$\mathcal{E} W_0(Z_1) = \mathcal{E} W_0(Z_2) - \mathcal{E} W_h(Z_2)$$

où l'indice 0 indique que la distribution de base est la population (des  $x_i$ ); alors que l'indice (h) désigne le grand échantillon fournissant l'estimation  $Z_1 = z_1^h$  et sur lequel est calculée la perte d'information  $W_h(z_2^k)$  pour chaque petit échantillon (k).

$\mathcal{E} W_0(Z_1)$  est fonction des  $z_1^h$  et indépendante des  $z_2^k$ .

Mais si  $W_0$  n'est pas fonction de  $\zeta$ ,  $W_h$  n'est pas fonction de  $z_1^h$ ; et ils'en suit que le second membre a sa dérivée nulle par rapport à  $z_1^h$ ; donc aussi le premier membre.

Par conséquent  $W_0(z_1^h)$  resterait constant, au moins lorsqu'on donne à  $z_1^h$  de petites variations.

Si nous excluons le cas où  $W(Z)$  serait discontinue, on a bien :

$$W_0(z_1^h) = \lambda^h, \quad \text{donc } \mathcal{E} W_0(z_1^h) = \mathcal{E}(\lambda^h) = \Lambda_1$$

$$\mathcal{E} W_0(z_2^k) = \Lambda_2$$

$$\mathcal{E} E W_h(z_2^k) = \Lambda_2 - \Lambda_1$$

Remarque.

Supposer  $W(z)$  continue, c'est écarter le cas où  $W(z)$ , tout en étant "globalement" variable, resterait constante dans des intervalles, autrement dit augmenterait par saut quand  $z$  s'éloigne de  $\zeta$ . En effet, on peut considérer qu'un tel indicateur de la perte d'information serait défectueux du point de vue conceptuel (tout en acceptant en revanche que la courbe (convexe) représentant  $W(z)$  présente des points anguleux - par exemple soit polygonale).

3/ - Le cas exceptionnel d'estimation biaisée.

On rencontre très facilement le cas suivant, qui correspond à une distribution de  $Z$  tout à fait particulière :

THEOREME -

Si l'erreur d'échantillonnage a un signe constant, tant pour  $Z_2 - Z_1$  que pour  $Z_1 - \zeta$ , le module du biais est perte d'information, ainsi que tout  $|g(Z) - g(\zeta)|$ , où  $g$  désigne une fonction monotone.

Réciproquement, si  $\mathcal{E} Z - \zeta$  a un signe constant, il est nécessaire que l'erreur ait un signe constant pour qu'on puisse écrire :  $W = Z - \zeta$ .

En d'autres termes,  $Z_1$  ne prend que des valeurs  $z_1^h$  disons inférieures ou égales à  $\zeta$ ;  $Z_2(z_1^h)$  ne prend que des valeurs inférieures ou égales à  $z_1^h$ ; etc.

Bref l'équation de Chasles est vérifiée, pour la seule raison qu'on a (terme à terme) :

$$W(z_2^k, \zeta) = W(z_1^h, \zeta) + W(z_2^k, z_1^h)$$

en posant  $\pm W(z, \zeta) = \zeta - z$

voire  $\pm W(z, \zeta) = g(\zeta) - g(z)$ .

Bref la perte d'information est anti symétrique. En voici deux applications très simples.

On s'est limité aux sondages exhaustifs.

4/ - Estimation de la plus grande valeur des  $x_i$ .

Soit  $a > b > c \dots$  les valeurs des  $x_i$ . Soit  $f(x)$  une fonction croissante (sur tout l'intervalle des  $x_i$ ) avec, par définition :

$$W(x_i, a) = (ax_i) = f(a) - f(x_i)$$

par exemple  $(ax_i) = a - x_i$

Application.

Pour  $n = v - 1$ , il y a une chance sur  $v$  que l'échantillon ne renferme pas (a), mais alors il renferme forcément (b) et la perte d'information est égale à

$$\& W = \frac{1}{v} (ab)$$

Pour  $n = v - 2$  : parmi  $C_v^{v-2} = C_v^2$  combinaisons, dont  $(v - 1)$  renfermant la lettre a, les combinaisons ne renfermant pas (a) mais (b) sont au nombre de  $(v - 2)$ ; et une combinaison ne renferme ni (a) ni (b) mais (c). Ainsi la perte d'information est :

$$\& W = [(v - 1) (ab) + (ac)] / C_v^2$$

Pour n quelconque : sur  $C_v^n$  combinaisons :  $C_{v-1}^n$  renfermant a,  $C_{v-2}^n$  renfermant b (sans a);  $C_{v-3}^n$  renfermant c (sans a ni b); etc.

D'où la perte d'information :

$$\left[ C_{v-2}^n (ab) + C_{v-3}^n (ac) + \dots \right] / C_v^n$$

Remarques.

1) La théorie précédente se transpose immédiatement au cas de l'estimation de la plus petite des valeurs  $x_i$ .

2) Dans les deux cas, l'estimateur  $Z$  de  $\zeta$  est toujours biaisé, - par défaut s'il s'agit de la plus grande valeur, - par excès dans l'autre alternative; l'écart  $\zeta Z$  est toujours de même signe.

3) A mesure que l'échantillon devient plus petit,  $Z$  ne peut que s'éloigner davantage de  $\zeta$ ; autrement dit  $\zeta Z_n$  et  $Z_n Z_{n-1}$  sont de même signe.

Si la plus grande valeur connue pour un certain échantillon(n) est (b), la plus grande valeur pour les n échantillons d'effectif  $(n - 1)$  qu'il engendre sera b, ou c, ou d ... mais en aucun cas ne peut devenir a.

5/ - Estimation de l'étendue (ou amplitude) ("range").

L'étendue ( $r$ ), écart entre le plus grand et le plus petit des  $x_i$ , ne peut être estimée que si l'échantillon comprend au moins deux boules,  $n \geq 2$ , en quoi elle diffère de la plus grande ( $a$ ) des valeurs de  $x_i$ , qui demande seulement qu'on ait :  $n \geq 1$ .

Cela mis à part,  $r$  et  $a$  sont des propriétés tout à fait analogues : lorsque l'effectif  $n$  de l'échantillon diminue, depuis  $v$  jusqu'à 2, l'écart entre le plus grand et le plus petit des  $x_i$  - échantillons est au plus égal à  $r$  et ne peut jamais augmenter. Il fournit donc de  $r$  un estimateur biaisé par défaut.

La perte d'information correspondante peut être définie de façon comparable.

Soit  $a > b \dots > y > z$ , les valeurs de  $x_i$ .

Pour  $n = v - 1$ , il y a  $v - 2$  chances sur  $v$  que  $(a'z') = (az)$ .

$$W = \frac{1}{v} [(ab) + (yz)]$$

Pour  $n = v - 2$  :

$$W = \frac{1}{C_v^n} \{ 0 \cdot C_{v-2}^2 + [(ab) + (yz)] C_{v-2}^1 + [(ac) - (xz)] \}$$

Pour  $n$  quelconque :

$$W = \frac{1}{C_v^n} \{ 0 \cdot C_{v-2}^n + [(ab) + (yz)] [C_{v-3}^{n-1} + C_{v-4}^{n-2}] + [(ac) + (xz)] [C_{v-3}^{n-2} + \dots + C_{v-6}^{n-4}] + \dots \}$$

6/ - Retour au cas général de l'estimation biaisée.

Alors que ni  $VZ = \mathcal{E}Z^2 - \zeta^2 + [\zeta^2 - (\mathcal{E}Z)^2]$ ,

$$\text{ni } \mathcal{E}(Z - \zeta)^2 = \mathcal{E}Z^2 - \zeta^2 + 2\zeta(\zeta - \mathcal{E}Z)$$

ne sont plus des pertes d'information, il en existe d'évidentes, en ce sens que la relation de récurrence est manifestement vérifiée : telles sont

$$\mathcal{E}Z^2 - \zeta^2 \quad \text{ou} \quad \mathcal{E}\Phi(Z) - \Phi(\zeta)$$

où  $\Phi$  est une fonction arbitraire, sous réserve (a) qu'on ait bien :

$$\Phi(\zeta) < \mathcal{E}\Phi(Z_1) < \mathcal{E}\Phi(Z_2) \dots$$

avec les notations du théorème I ( §I), et (b) que  $Z_2$  se réduise bien à  $Z_1$ , lorsque l'échantillon de  $\pi_2$  recouvre celui de  $\pi_1$  (les plans de sondage  $\pi_1$  et  $\pi_2$  appartenant à une suite de plans à effectifs décroissants pris sur la famille  $\mathcal{F}$ ).

La restriction (a) n'est plus une banale concavité puisque  $\zeta$  n'est plus la moyenne de  $Z_1$ . La restriction (b) est une condition d'estimation "cohérente" au sens des sondages (laquelle ne comporte ni existence de tirages infiniment nombreux ni convergence en probabilité). L'estimateur  $Z$  (biaisé) sera donc supposé à présent "uniformément cohérent", c'est-à-dire cohérent tout au long des suites de plans telles que

$$[U \pi_1 \pi_2 \dots]$$

On s'intéressera ici plus spécialement aux estimateurs isomorphes du paramètre à estimer. La notion d'estimation uniformément cohérente est un peu plus large que l'estimation isomorphe (l'expression de  $\zeta$  à laquelle se réduit  $Z$  lorsque l'échantillon de  $\pi$  recouvre l'univers  $U$  pouvant être finalement plus simple que celle de  $Z$ ). Toute fonction arbitraire  $g(Z)$  est d'ailleurs estimateur isomorphe de  $g(\zeta)$ , ce qui explique l'arbitraire laissé à la fonction  $\Phi$ .

La première difficulté est celle de l'affectation de telle ou telle perte d'information à tel ou tel estimateur. Pourquoi  $(\mathcal{E}Z^2 - \zeta^2)$  serait-elle nécessairement affectée à l'estimation de  $\zeta$  par  $Z$ . Quand on estime le ratio  $\zeta = a/b$  par  $Z = A/B$  avec  $\mathcal{E}A = a$ ;  $\mathcal{E}B = b$ , on sait que les termes en  $(n^{-1})$  de  $\mathcal{E}(Z - \zeta)^2$  et du biais sont :

$$\mathcal{E}(Z - \zeta)^2 \# (\gamma^2 - 2\rho\gamma\gamma' + \gamma'^2) \cdot n^{-1}$$

$$\mathcal{E}Z - \zeta \# (\gamma'^2 - \rho\gamma\gamma') \cdot n^{-1}$$

Notons que, de toute façon,  $V(Z)$  et  $\mathcal{E}(Z - \zeta)^2$  sont équivalents :

$$\mathcal{E}(Z - \zeta)^2 \equiv V(Z) + (\mathcal{E}Z - \zeta)^2$$

En revanche  $\mathcal{E}Z^2 - \zeta^2 \equiv \mathcal{E}(Z - \zeta)^2 + 2\zeta \mathcal{E}(Z - \zeta)$

de sorte que :

$$\mathcal{E}Z^2 - \zeta^2 \# [\gamma^2 - 2\rho\gamma\gamma'(1 + \zeta) + \gamma'^2(1 + 2\zeta)] \cdot n^{-1}$$

Nous serons amené plus loin à définir une perte d'information

pour  $Z$  qui est en désaccord avec ce résultat; autrement dit nous n'attachons ( $\mathcal{E}Z^2 - \zeta^2$ ) à  $Z$  que lorsque le biais est nul (ou d'ordre supérieur au premier). Ceci montre qu'une certaine prudence est nécessaire pour rester cohérent.

### 7/ - Théorie de l'estimation isomorphe.

Désignons par  $W[Z, \zeta]$  la fonctionnelle qu'on entend substituer à la fonction  $W(Z, \zeta)$ .

Posons d'une manière générale :

$$\mathcal{E}W[Z, \zeta] = \mathcal{E}\phi[Z] - \phi[\zeta]$$

$$W[Z, \zeta] = \phi[Z] - \phi[\zeta] + \Delta[Z, \zeta]$$

(les crochets désignant des fonctionnelles). On aura nécessairement :

$$\mathcal{E}\Delta[Z, \zeta] = 0, \quad W[\zeta, \zeta] = 0, \quad W[Z, \zeta] > 0$$

plus une condition de convexité pour  $W$ .

#### 1er problème.

Traduire par une condition de la forme

$$\mathcal{E}\Delta[Z, \zeta] = 0$$

la relation entre estimateur  $Z$  et estimé  $\zeta$ .

S'il s'agit de  $\zeta = a/b$  et  $Z = A/B$ , avec

$$\mathcal{E}A - a = \mathcal{E}B - b = 0,$$

en raisonnant comme pour établir le théorème 1, on montrera qu'on a nécessairement (si  $\Delta$  est fonction très régulière de  $A$  et  $B$ ) :

$$\Delta[Z, \zeta] = \psi_1[\zeta] (a - A) + \psi_2[\zeta] (b - B)$$

D'une façon générale on va supposer que  $\zeta = f(a, b, c, \dots)$  et  $Z = f(A, B, C, \dots)$  avec

$$\mathcal{E}A = a; \quad \mathcal{E}B = b; \quad \mathcal{E}C = c \dots$$

- l'isomorphisme de  $Z$  et  $\zeta$  est traduit par la présence d'une fonction commune  $f$ ;

- les  $A B C \dots$  sont calculés sur l'échantillon et les  $a b c \dots$  sur l'univers;

- A est isomorphe de a, B de b, C de c, etc.

Alors on a nécessairement (pourvu que  $\Delta$  soit une fonctionnelle très régulière) :

$$\Delta[Z, \zeta] = \psi_1[\zeta] \cdot (a - A) + \psi_2[\zeta] \cdot (b - B) + \psi_3[\zeta] \cdot (c - C) + \dots$$

d'où

$$W[Z, \zeta] \equiv \phi[Z] - \phi[\zeta] + \psi_1[\zeta] (a - A) + \psi_2[\zeta] (b - B) + \psi_3[\zeta] (c - C) + \dots$$

2ème problème : Cas où  $W[Z, \zeta]$  est symétrique en Z et  $\zeta$ .

Le calcul du théorème III se transpose :  $W = W^*$  ; posons  $\psi[\zeta] = \psi$ ,  $\psi[Z] = \psi^*$ .

$$2\{\phi[\zeta] - \phi[Z]\} \equiv (a - A)(\psi_1 + \psi_1^*) + (b - B)(\psi_2 + \psi_2^*) + (c - C)(\psi_3 + \psi_3^*) \dots$$

$$\begin{cases} 2\phi[\zeta] \equiv a\psi_1 + b\psi_2 + c\psi_3 \dots \\ 0 \equiv (a\psi_1^* - A\psi_1) + (b\psi_2^* - B\psi_2) + (c\psi_3^* - C\psi_3) \dots \\ 2\phi[Z] \equiv A\psi_1^* + B\psi_2^* + C\psi_3^* \dots \end{cases}$$

$$\frac{\psi_1}{a} = \frac{\psi_1^*}{A} = \alpha ; \quad \frac{\psi_2}{b} = \frac{\psi_2^*}{B} = \beta ; \quad \frac{\psi_3}{c} = \frac{\psi_3^*}{C} = \gamma ; \dots$$

$$\text{d'où } 2\phi[Z] \equiv \alpha A^2 + \beta B^2 + \gamma C^2 \dots$$

$$2W[Z, \zeta] = \alpha(A - a)^2 + \beta(B - b)^2 + \gamma(C - c)^2 + \dots$$

Cette solution n'est pas contestable; mais il se trouve que ce n'est pas celle qui est employée dans la routine des sondages.

3ème problème.

Traduire par une relation entre  $\phi$  et les  $\psi_i$  la convexité de W.  
On a :

$$\begin{aligned} \frac{\partial W}{\partial A} &\equiv \frac{\partial \phi}{\partial A} + \frac{\partial \Delta}{\partial A} \\ &\equiv \frac{\partial \phi}{\partial A} - \psi_1[\zeta] \end{aligned}$$

en supposant que  $W[Z]$  est en réalité une fonction dérivable  $W(A, B, C \dots)$ .

On va écrire que, pour  $\zeta = Z$  (c'est-à-dire  $a = A$ ,  $b = B$ ,  $c = C$ ), on a :

$$\partial W / \partial A = 0, \quad \partial W / \partial B = 0, \quad \partial W / \partial C = 0 \dots$$

$$\text{d'où : } \psi_1 = \partial \Phi [\zeta] / \partial a, \quad \psi_2 = \partial \Phi [\zeta] / \partial b, \quad \psi_3 = \partial \Phi [\zeta] / \partial c \dots$$

$$\text{et } W = \Phi [Z] - \Phi [\zeta] + \frac{\partial \Phi [\zeta]}{\partial a} (a - A) + \frac{\partial \Phi [\zeta]}{\partial b} (b - B) + \frac{\partial \Phi [\zeta]}{\partial c} (c - C) \dots$$

relation transposée de celle du théorème III. De même on doit écrire

$$\frac{\partial^2 \Phi}{\partial a^2} \geq 0, \quad \frac{\partial^2 \Phi}{\partial b^2} \geq 0, \quad \frac{\partial^2 \Phi}{\partial c^2} \geq 0 \dots$$

avec

$$\frac{\partial^2 \Phi}{\partial a^2} \frac{\partial^2 \Phi}{\partial b^2} - \left( \frac{\partial^2 \Phi}{\partial a \partial b} \right)^2 \geq 0 \text{ etc.}$$

En particulier en supposant symétrique en  $\zeta$ ,  $Z$  la fonctionnelle  $W [Z, \zeta]$  on retrouve le résultat précédent.

#### Autre solution particulière.

$$\text{Posons } \psi_1 = \psi \alpha, \quad \psi_2 = \psi \beta, \quad \psi_3 = \psi \gamma \dots$$

Il vient :

$$2(\Phi [\zeta] - \Phi [Z]) = (\psi + \psi^*) \{ \alpha (a - A) + \beta (b - B) + \gamma (c - C) \dots \}$$

en supposant que  $\alpha \beta \gamma \dots$  ne sont pas transposés car constants; on posera :

$$\alpha (a - A) + \beta (b - B) + \gamma (c - C) \dots = \omega [Z, \zeta]$$

avec

$$\omega^* = -\omega$$

et le calcul du théorème III donne cette fois

$$W = \Phi [Z] - \Phi [\zeta] + 2 \sqrt{\Phi [\zeta]} \{ \sqrt{\Phi [\zeta]} - \sqrt{\Phi [Z]} \}$$

avec

$$\sqrt{\Phi [Z]} = \alpha A + \beta B + \gamma C = \psi [Z]$$

$$\sqrt{\Phi [\zeta]} = \alpha a + \beta b + \gamma c = \psi [\zeta]$$

La perte d'information est :

$$\mathcal{E}W = \mathcal{E}[\alpha(A - a) + \beta(B - b) + \gamma(C - c)]^2$$

Choix de  $\alpha\beta\gamma$ .

Jusqu'ici le choix de  $\alpha\beta\gamma$  était resté arbitraire. Les conditions de convexité étaient bien vérifiées ainsi :

$$\partial\Phi/\partial a = 2\alpha(\alpha a + \beta b + \gamma c)$$

d'où

$$\frac{\partial^2\Phi}{\partial a^2} = 2\alpha^2 > 0 \quad \left(\frac{\partial^2\Phi}{\partial a\partial b}\right)^2 = 4\alpha^2\beta^2 = \frac{\partial^2\Phi}{\partial a^2} \cdot \frac{\partial^2\Phi}{\partial b^2}$$

On va particulariser  $\alpha\beta\gamma$  de façon à rejoindre la définition du §1 lorsque  $(A - a), (B - b), (C - c)$  sont infiniment petits. Pour cela on différencie  $\zeta = f(a, b, c \dots)$ ; il vient :

$$d\zeta = f'a \cdot da + f'b \cdot db + f'c \cdot dc \dots$$

et on choisit  $\alpha, \beta, \gamma$  proportionnels à  $f'a, f'b, f'c$ . Ainsi on substitue à

$$\psi(\zeta) = 2\lambda \cdot \zeta \quad (\text{théorème III})$$

la relation

$$d\psi[\zeta] = 2\lambda \cdot d\zeta$$

D'où la perte d'information (particulière, non symétrique) :

$$\mathcal{E}W = \lambda^2 \mathcal{E}[f'a \cdot (A - a) + f'b \cdot (B - b) + f'c \cdot (C - c)]^2$$

### 8/ - Application et conséquences.

#### 1) Estimation du ratio.

$$\zeta = a/b \quad \text{d'où} \quad \log \zeta = \log a - \log b. \quad \frac{d\zeta}{\zeta} = \frac{da}{a} - \frac{db}{b}$$

$$\mathcal{E}W = \lambda^2 \zeta^2 \cdot \mathcal{E}\left(\frac{A - a}{b} - \frac{B - b}{b}\right)^2 = \lambda^2 \zeta^2 \mathcal{E}\left(\frac{A}{a} - \frac{B}{b}\right)^2$$

L'expression usuelle des statisticiens est  $\mathcal{E}\left(\frac{A}{a} - \frac{B}{b}\right)^2$ , premier terme du développement de

$$v \left(\frac{A}{B}\right)$$

#### 2) Estimation du coefficient de corrélation.

$$\rho = \text{Cov}(x y) / \sigma(x) \cdot \sigma(y)$$

On posera

$$a = N\sigma^2(x)/N - 1; \quad b = N\sigma^2(y)/N - 1; \quad c = N \text{Cov}(x, y)/N - 1$$

dont les estimateurs isomorphes A B et C sont sans biais.

Soit  $\rho^2 = \zeta$ . Ainsi à  $[a b \zeta - c^2 = 0]$  correspond bien  $[A B Z - C^2 = 0.]$

$$d\zeta/\zeta = 2(dc/c) - da/a - (db/b)$$

$$\text{d'où} \quad \mathcal{E}W = \lambda^2 \zeta^2 \mathcal{E} \left( 2 \frac{C}{c} - \frac{A}{a} - \frac{B}{b} \right)^2$$

à un facteur près (indépendant de Z).

Là encore il s'agit d'une première approximation de la variance.

#### Notation.

Au lieu de  $\mathcal{E}W = \mathcal{E}(Z - \zeta)^2$  on peut donc écrire

$$\mathcal{E}W = \mathcal{E}(\delta Z - \delta \zeta)^2$$

Parmi toutes les pertes d'information (dépendant d'une fonctionnelle arbitraire  $\phi$ ) on a donc choisi plus particulièrement celle qui (quadratique en A B C) est asymptotique à la variance (lorsque Z est très voisin de z).

#### Remarque (complétant le § 6 ci-dessus).

C'est pour une raison de convexité que les estimateurs biaisés Z n'admettent pas  $Z^2 - \zeta^2$  comme perte d'information.

Il suffit de le vérifier dans le cas du ratio :

$$Z = A/B, \quad \zeta = a/b, \quad \mathcal{E}A - a = \mathcal{E}B - b = 0$$

$$\text{Posons} \quad \phi(\zeta) = \zeta^2, \quad \mathcal{E}W = \mathcal{E}Z^2 - \zeta^2$$

$$W = \left(\frac{A}{B}\right)^2 - \left(\frac{a}{b}\right)^2 + \Delta, \quad \mathcal{E}\Delta = 0$$

$$\text{d'où} \quad \Delta = \lambda(a - A) + \mu(b - B)$$

Imposons à  $W[Z] = W(A, B)$  d'être représentée par une surface tangente (au point  $A = a, B = b$ ) au plan  $\Delta = 0$ ; il vient

$$\frac{\partial W}{\partial A} = 2 \frac{A}{B^2} - \lambda, \quad \frac{\partial W}{\partial B} = - \frac{2 A^2}{B^3} - \mu$$

d'où  $\lambda = 2 a/b^2, \quad \mu = - 2 a^2/b^3$

$$\Delta = \frac{2 a^2}{b^2} \left( \frac{B}{b} - \frac{A}{a} \right)$$

Montrons que la surface représentant  $W(A, B)$  n' est pas convexe.  
On a :

$$\frac{\partial^2 W}{\partial A^2} \cdot \frac{\partial^2 W}{\partial B^2} - \left( \frac{\partial^2 W}{\partial A \partial B} \right)^2 = \frac{2}{B^2} \frac{6 A^2}{B^4} - \left( \frac{4 A}{B^3} \right)^2 = - \frac{4 A^2}{B^6} < 0, \text{ cqfd}$$

D'ailleurs l'expression exacte de  $\mathcal{E} Z^2$  est souvent fort compliquée; et au point de vue des calculs pratiques on gagne à renoncer à  $\mathcal{E} Z^2 - \zeta^2$ .

9/ - Estimation du maximum de vraisemblance.

a) On sait que, pour une variable  $Z$  dont la loi de distribution admet la densité

$$g(z, \zeta)$$

$\zeta$  étant le paramètre à estimer, la quantité d'information de Fisher (pour  $n$  tirages sans remise) est  $nH$ , avec

$$Y = \frac{\partial}{\partial \zeta} \log g(Z, \zeta), \quad H = \mathcal{E}(Y^2)$$

b) En fait  $(H/n)$  est une perte d'information; car c'est la variance d'une moyenne arithmétique  $\bar{Y}$  d'espérance mathématique nulle. En effet, posons

$$y_i = \frac{\partial}{\partial \zeta} \log g(z_i, \zeta), \quad \bar{Y} = S y_i / n.$$

En dérivant (par rapport à  $\zeta$ ) sous le signe  $\int$

$$\int_{-\infty}^{+\infty} g(z, \zeta) dZ = 1$$

(à supposer qu'on ait le droit de le faire) on trouve

$$\mathcal{E} Y = 0; \quad \text{d'où } \mathcal{E} \bar{Y} = 0. \quad \text{cqfd}$$

c) On peut considérer que l'estimation  $Z^0$  du maximum de vraisemblance est un estimateur isomorphe de  $\zeta$ . Plus précisément  $\zeta$  et  $Z^0$  sont racines de deux équations isomorphes, à savoir :

$$\psi(\zeta) \equiv \mathcal{E} y(\zeta) = 0;$$

$$\Psi(Z) = S y_i(Z^0)/n = 0;$$

et, si par exemple leurs premiers membres sont développables en série entière :

$$\psi(t) \equiv \mathcal{E} y(t) \equiv a_0 + a_1 t + a_2 t^2 + \dots$$

$$\Psi(t) = \frac{1}{n} S y_i(t) \equiv A_0 + A_1 t + A_2 t^2 + \dots$$

on a (quel que soit t)  $\mathcal{E} \Psi = \psi$ , et par conséquent  $\mathcal{E} A_j = a_j$ .

On se trouve donc dans un cas à peine plus général qu'aux n° 1 et 2 : à A B C ... est substituée l'infinité des  $A_j$ .

d) Ainsi la perte d'information serait de la forme  $\mathcal{E} W [Z^0]$ , avec

$$W [Z^0] = \phi [Z^0] - \phi [\zeta] + \sum \lambda_j (a_j - A_j)$$

e) A  $\zeta$  et g donnés correspond  $Z^0$ ; donc  $\zeta$  est fonctionnelle de  $Z^0$ ; donc en faisant  $t = \zeta$ ,  $\Psi(\zeta)$  est une fonctionnelle de  $Z^0$  qu'on peut appeler  $\sqrt{\phi} [Z^0]$ .

Alors  $\sqrt{\phi}[\zeta]$  n'est autre que  $\psi(\zeta)$ , c'est-à-dire 0. D'où

$$\mathcal{E} \phi [Z^0] - \phi [\zeta] = \mathcal{E} \bar{Y}^2 - 0 = H/n$$

L'information de Fisher est donc le produit par  $n^2$  d'une perte d'information relative à l'estimation (biaisée, isomorphe) du maximum de vraisemblance.

Nota - Cette variance ne pouvait représenter la perte d'information relative à l'estimation de  $\mathcal{E} Y$  puisque  $\mathcal{E} Y = 0$ .

#### 10/ - Généralisation : Variances de statistiques hors estimation.

Revenons sur le cas de  $n$  tirages équiprobables dans une urne unique de taille  $v$ . Soit  $Y_n$  une statistique quelconque (calculée sur l'échantillon);

$$\mathcal{E} Y_n = y(n, v)$$

dépend en général de  $n$ ; et on ne peut donc dire de  $Y_n$  que c'est un estimateur (sans biais, d'un quelconque  $\eta$ ). En revanche  $\mathcal{V} Y_n$  est perte d'information du moment que la relation de récurrence est satisfaite, c'est-à-dire, avec  $n > n'$ , dès que

$$\mathcal{E}\{[Y_n - y(n, v)] E_n[Y_n - Y_n]\} = 0$$

a) Supposer  $E_n[Y_n] = Y_n$  (uniformément) reviendrait à faire dépendre  $Y_n$  et  $y$  des  $n' < n$ , ce qui est à rejeter. En revanche rien n'empêche d'admettre que la différence

$$E_n Y_n - Y_n$$

est constante (ne dépend pas de  $Y_n$ ) pour  $n'$  donné; ceci implique pourtant que :

$$E_n Y_n - Y_n = y(n', v) - y(n, v),$$

ou 
$$E_n\{Y_n - y(n', v)\} = Y_n - y(n, v).$$

#### Conclusion.

Il y a lieu de supposer la variable

$$Y_n - y(n, v)$$

uniformément sans biais, pour que la perte d'information  $\forall Y_n$  existe (sans être encore attachée à quelque  $\eta$ ).

Alors se pose le même problème qu'au n°9 : Peut-on rattacher

$$\lambda \{ \forall Y_n - y^2(n, v) \}$$

à l'estimation biaisée d'un certain  $\zeta$  par un certain  $Z$  ?

Refaisant à l'envers le calcul du n°9, il vient

$$\psi(\zeta) = 0,$$

$$\Psi(\zeta) = Y_n - y(n, v)$$

Si  $Y_n$  dépend d'un paramètre  $\zeta$  à estimer, tel que

$$\mathcal{E}Y_n(\zeta) - y(n, v, \zeta) = 0$$

l'estimateur "isomorphe" (au sens du n°9) est la racine de l'équation

$$Y_n(Z) - y(n, v, Z) = 0$$

b) Supposons plus simplement que  $Y_n = S y_i / n$ ;

$$\mathcal{E} y_i = y = 0$$

où  $y_i$  dépend d'un paramètre  $\zeta$ ;  $\forall Y_n - \bar{y}^2$  est alors une perte d'information. Il vient

$$\Psi(\zeta) = \frac{S y_i}{n}$$

Posons : 
$$\Psi(Z) = \frac{S y_i(Z)}{n} = 0$$

Il y a isomorphisme entre les équations  $\&y = 0$  et  $S y_i / n = 0$ , et (au même sens qu'au n°9) entre  $\zeta$  et son estimation  $Z$ .

Remarques.

1) Dans quels cas la variable  $Y_n - y(n, \nu)$  est-elle uniformément sans biais ?

Considérons par exemple  $Y_n = S^2$ ,

$$y(n, \nu) = \frac{(n-1)\nu}{(\nu-1)n} \sigma^2, \quad (Y_n = S^2).$$

On n'est pas dans le cas recherché, vu que

$$E \left[ S^2 - \frac{(n-1)\nu}{(\nu-1)n} \sigma^2 \right] = \frac{(n-1)n}{(n-1)n'} \left[ S^2 - \frac{(n-1)\nu}{(\nu-1)n} \sigma^2 \right]$$

La seule possibilité est d'avoir  $E[Y_n] = Y_n + k(n')$ ,  $k$  étant constant (quand  $n'$  est donné) c'est-à-dire indépendant de  $Y_n$ ; il en est ainsi quand  $Y_n$  est de la forme

$$\bar{X}_n + \sum_{\nu=1}^n \delta(\nu, i). \quad (\text{condition suffisante})$$

2) On peut (sous des conditions assez larges) écrire un développement tel que :

$$y(n, \nu) = \zeta + a_1 \left( \frac{1}{n} - \frac{1}{\nu} \right) + a_2 \left( \frac{1}{n} - \frac{1}{\nu} \right)^2 + \dots$$

et on n'a pas en général  $a_2 = \dots = 0$ . En outre un cas encore plus intéressant serait celui où

$$y(n, \nu) = \zeta + a_0 \left( \frac{1}{n} - \frac{1}{\nu} \right)^{1/2} + \dots$$

11/ - Quelques considérations générales sur l'estimation.

a) Il n'est pas question d'exiger de tout estimateur l'absence de biais; mais on ne peut accepter pour estimer  $\zeta$  n'importe quelle statistique  $Z$ .

En pratique on a rencontré trois cas :

- isomorphisme et absence de biais (exemple :  $nS^2/n - 1$  pour  $v\sigma^2/v - 1$ )
- isomorphisme avec biais (exemple :  $S^2$  pour  $\sigma^2$ )
- absence de biais sans isomorphisme (exemple :  $\frac{n(v - 1)S^2}{(n - 1)v}$  pour  $\sigma^2$ )

b) Les exemples ci-dessus relèvent du cas où,  $\eta(v)$  étant isomorphe de l'aléatoire  $Y$ , on a

$$\mathcal{E}Y \equiv y(v, n) = \eta(v) \cdot a(v, n)$$

Un cas également simple est celui où on aurait :

$$\mathcal{E}Y = \eta(v) + b(v, n);$$

on poserait  $\zeta = \eta(v)$ ,  $Z = Y - b(v, n)$ . Le cas mixte serait  $\mathcal{E}Y = a\eta + b$ , d'où  $\zeta = \eta(v)$  pour  $Z = [Y - b(v, n)]/a(v, n)$ .

c) Mais en général il n'est pas question d'éliminer totalement le biais. Par exemple, que  $\sigma$  soit estimé au moyen de  $S$ , ou de  $S\sqrt{n(v - 1)/v(n - 1)}$ , il y a biais.

De ce qui précède, il résulte qu'on mesure la perte d'information relative à  $\frac{nS^2}{n - 1}$  estimateur de  $\frac{v\sigma^2}{v - 1}$ , par  $\mathcal{V}\left(\frac{nS^2}{n - 1}\right)$  relative à  $S^2$  estimateur de  $\frac{(n - 1)v}{(v - 1)n}\sigma^2$ , par  $\mathcal{V}S^2$ ; celle relative à  $S^2$ , estimateur isomorphe biaisé de  $\sigma^2$ , ne sera pas

$$\begin{aligned} \mathcal{E}(S^2 - \sigma^2)^2 &\equiv \mathcal{V}S^2 + \left[ \frac{(n - 1)v}{(v - 1)n}\sigma^2 - \sigma^2 \right]^2 \\ &\equiv \mathcal{V}S^2 + \left[ \frac{v\sigma^2}{v - 1} \right]^2 \left( \frac{1}{n} - \frac{1}{v} \right)^2 \end{aligned}$$

mais  $\lambda \mathcal{V}\left(\frac{nS^2}{n - 1}\right)$

La perte d'information n'est jamais définie qu'à un facteur constant près; et il y a lieu de considérer que  $nS^2/n - 1$ ,  $S^2$ , et aussi  $S$ , ont pour perte d'information symétrique  $\lambda \mathcal{E}\left(\frac{nS^2}{n - 1} - \frac{v\sigma^2}{v - 1}\right)^2$ . Le choix de  $\lambda$  demande une convention supplémentaire (voir Chapitre VII);

d) On a vu (Ch. III, § I, C) que Basu trouvait moins de risque à estimer  $\bar{x}$  par  $c\bar{X}$  ( $c < 1$ ) que par  $\bar{X}$ . Au point de vue "perte d'information", il n'y aurait pas de différence entre  $\bar{X}$  et  $c\bar{X}$  (ou plutôt  $c\bar{X}$  n'informe pas sur  $\bar{x}$ ).

On verra au chapitre VII que certains estimateurs permettent de réduire les pertes d'information (théorie des informations supplémentaires); ce ne serait pas le cas de  $c\bar{X}$ .

On ne saurait pourtant se passer d'un critère pour le choix de l'estimateur ( $c = 1$ , par exemple); critère comprenant comme cas particulier l'absence de biais.

e) La coïncidence complète des formes analytiques de  $Z$  et  $\zeta$  (source du mot "isomorphisme") n'est pas nécessaire en soi; elle assure a fortiori la "cohérence" quand l'échantillon grandit jusqu'à recouvrir toute la population; mais la cohérence (consistency) elle-même lorsque l'échantillon est petit, n'est pas toujours exigée (et peut empêcher l'élimination d'un biais, auquel cas on renonce à la cohérence).

f) En revanche la relation de récurrence de l'axiomatique de l'information adoptée ici suppose (si  $Z$  et  $Z'$  sont estimateurs de  $\zeta$  dans une suite de plans de sondage de tailles décroissantes) que  $Z'$  informe sur  $\zeta$  lorsque  $Z$  informe sur  $\zeta$  et  $Z'$  sur  $Z$ . Plus précisément : la relation  $E Z = \zeta$  est un cas particulier d'une correspondance générale

$$e Z = \zeta$$

qui possède les propriétés suivantes (transitivité).

Soit  $Z_\nu = \zeta, Z_{\nu-1}, Z_{\nu-2}, \dots, Z_n, \dots, Z_r$  ( $r \geq 1$ ) les expressions de  $Z$  pour toutes les tailles d'échantillons successifs (de  $\nu - 1$  à  $r$ );

$$(1) \quad e Z_n = \zeta$$

entraîne  $e' Z_n = Z_m$  pour tout  $m > n$ .

(2) Si  $e Z_n = \zeta$  et  $e' Z_n = Z_m$  on a  $e e' Z_n = \zeta$  et l'opérateur  $(e e')$  fait partie de la même famille que  $e$ .

(3) En particulier, si  $Z_n$  est calculable pour  $n = r = 1$ , on a :

$$Z_1 = f(X, Y, \dots)$$

$$D'où : \quad e[f(X, Y, \dots)] = \zeta, \quad e'[f(X, Y, \dots)] = Z_m$$

qui donnent les définitions de l'estimateur  $Z$  et du paramètre  $\zeta$  de la population  $U$ .

#### Application.

Reprenons le cas où  $\zeta = \bar{x}$ ,  $Z = c\bar{X}$  (Basu; voir IV, I, C ci-dessus).

Soit  $Z_{\nu-1} = a\bar{X}$  :  $\mathcal{E}Z_{\nu-1} = a\zeta$ ; de proche en proche, il vient  $Z_n = a^{\nu-n}\bar{X}$ .  
Pour  $n = 1$  :  $Z_1 = a^{\nu-1} \cdot X$ .

De la même façon, on pourrait avoir  $Z_{\nu-1} = \bar{X} + b$ , d'où

$$Z_n = \bar{X} + (\nu - n)b$$

Rien n'empêche de rendre les coefficients  $a$  et  $b$  fonctions du rang  $n$  de tirage.

#### Remarques.

1) Soit à estimer  $\zeta$  par  $Z$ , tel que  $\mathcal{E}Z = f(\zeta)$ . La transitivité n'est possible qu'avec une relation  $f$  linéaire.

2) On verra encore au n° 14 que le calcul de la perte d'information n'est possible qu'autant que  $\zeta$  et chacun des  $Z$  ont une expression commune, en ce sens qu'on doit pouvoir calculer une certaine fonctionnelle  $\phi[Z]$  pour chaque  $Z$  et en particulier pour  $\zeta$ .

Par exemple, on ne peut admettre  $\zeta = \frac{a + b + c}{3}$ , si  $Z$  prend l'une des trois valeurs  $\sqrt{ab}$ ,  $\sqrt{ac}$ ,  $\sqrt{bc}$  (pour  $n = 2$ ,  $\nu = 3$ ).

#### 12/ - THEOREME 1 bis -

Si  $W[Z, \zeta]$  est fonctionnelle de  $Z$  et  $\zeta$ , on a nécessairement  $\mathcal{E}W = \mathcal{E}\phi[Z] - \phi[\zeta]$ ,  $\phi[\ ]$  étant elle-même fonctionnelle.

En effet  $\mathcal{E}W[Z_2, Z_1]$  n'est pas fonctionnelle de  $\zeta$ ; or ce n'est autre que

$$\mathcal{E}W[Z_2, \zeta] - \mathcal{E}W[Z_1, \zeta]$$

$$\text{d'où : } \mathcal{E}W[Z_2, \zeta] = k_2 - \phi[\zeta], \quad \mathcal{E}W[Z_2, Z_1] = K_{21} - \phi[Z_1]$$

$$\text{d'où : } \mathcal{E}W[Z_1, \zeta] = [k_2 - \mathcal{E}K_{21}] + \mathcal{E}\phi[Z_1] - \phi[\zeta]$$

qui ne doit pas dépendre du sondage n°2; comme quoi

$$k_2 = \mathcal{E}K_{21}$$

Remarque.

$W [Z, \zeta]$  peut dépendre non seulement de  $Z$  et  $\zeta$  mais de la distribution de base; le calcul suppose que ces paramètres restent constants quand on remplace  $W [Z_2, \zeta]$  par  $W [Z_2, Z_1]$ . Par exemple, pour l'estimation du ratio, on doit avoir

$$W [Z_2, \zeta] = \left( \frac{A_2 - a}{a} - \frac{B_2 - b}{b} \right)^2 ; \quad W [Z_1, \zeta] = \left( \frac{A_1 - a}{a} - \frac{B_1 - b}{b} \right)^2$$

$$W [Z_2, Z_1] = \left( \frac{A_2 - A_1}{a} - \frac{B_2 - B_1}{b} \right)^2$$

et non pas 
$$W [Z_2, Z_1] = \left( \frac{A_2 - A_1}{A_1} - \frac{B_2 - B_1}{B_1} \right)^2$$

On dira que l'estimateur est uniformément biaisé si on a

$$E_{\omega} [Z_2, Z_1] = 0$$

condition qui entraîne  $\mathcal{E}_{\omega} [Z, \zeta] = 0$ . Ceci suppose donc que  $Z_1$  est lié à la distribution de  $Z_2$  de façon transitive : c'est le cas pour :

$$E [\alpha(A_2 - A_1) + \beta(B_2 - B_1)] = 0$$

si  $\alpha$  et  $\beta$  sont constants.

13/ - Autre remarque.

La nécessité de considérer tantôt une fonction  $\Phi(Z)$  et tantôt une fonctionnelle  $\Phi[Z]$  peut être mise en évidence en considérant la perte d'information attachée à un point.

Considérons le point  $M(\bar{X}, \bar{Y})$ , estimateur sans biais du point  $m(\bar{x}, \bar{y})$  et la perte d'information qui entache cette estimation, soit

$$V(u \bar{X} + v \bar{Y}) = u^2 V \bar{X} + 2 u v \text{Cov}(\bar{X} \bar{Y}) + v^2 V \bar{Y}$$

Considérons les nombres complexes  $Z = \bar{X} + i \bar{Y}$  et  $\zeta = \bar{x} + i \bar{y}$  affixes de  $M$  et  $m$ ; avec  $\mathcal{E} Z = \zeta$  par définition

$$V Z = V \bar{X} + 2 i \text{Cov} \bar{X} \bar{Y} + i^2 V \bar{Y}$$

Passons en coordonnées polaires  $R e^{i\theta} = Z$ ,  $r e^{i\theta} = \zeta$ . La perte

$$V Z = V(R e^{i\theta}) = \mathcal{E}(R e^{i\theta} - r e^{i\theta})^2$$

n'est manifestement pas de la forme (où cop signifie perte en commun).

$$\lambda_1 p(R) + 2 \lambda_2 \text{cop}(R, \Theta) + \lambda_3 p(\Theta).$$

On peut isoler dans VZ la perte d'information

$$(\mathcal{E}R^2 - r^2)$$

qui n'est pas une variance (car R n'est pas en général estimateur sans biais de r).

En tous cas on ne peut y isoler une expression de la forme

$$\mathcal{E}\phi(\Theta) - \phi(\theta)$$

Si l'on remarque que  $\theta = \text{arc tg } \tau$ ,  $\Theta = \text{arc tg } T$ , où  $\tau$  et  $T$  sont les ratios (univers et échantillon), on comprend que l'emploi d'une fonctionnelle soit nécessaire pour  $(\tau, T)$  aussi bien que pour  $(\theta, \Theta)$ .

Au contraire supposons les unités-échantillons tirées avec des probabilités uniformes pour le rayon vecteur et l'angle polaire; alors on a :

$$\mathcal{E}R = r, \quad \mathcal{E}\Theta = \theta;$$

la perte d'information attachée au point moyen est, disons

$$V(uR + v\Theta)$$

mais les pertes relatives à l'estimation de  $\bar{x}$  par  $\bar{X}$ , de  $\bar{y}$  par  $\bar{Y}$  ne sont plus séparément de la forme

$$\mathcal{E}\phi(\bar{X}) - \phi(\bar{x}), \quad \mathcal{E}\phi(\bar{Y}) - \phi(\bar{y}).$$

14/ - Les pertes d'information des estimateurs uniformément biaisés.

THEOREME II bis -

$$a) \text{ Si } \mathcal{E}W = \mathcal{E}\phi[Z] - \phi[\zeta]$$

on peut poser :  $W = \phi[Z] - \phi[\zeta] + \Delta[Z, \zeta]$

et on a :  $\mathcal{E}\Delta[Z, \zeta] = 0$

$$b) \text{ Pour que les deux égalités } \mathcal{E}\Delta[Z, \zeta] = 0$$

et  $E\Delta [Z', Z] = 0$

entraînent  $\mathcal{E}\Delta [Z', \zeta] = 0$

il suffit qu'on ait :  $\Delta [Z', \zeta] = \lambda \Delta [Z', Z] + \mu \Delta [Z, \zeta]$

(où  $\lambda$  et  $\mu$  sont des constantes).

c) Mais, pour que cette relation reste vérifiée lorsque  $Z$  vient en  $\zeta$ , il faut que  $\lambda = 1$ ; et pour que  $Z'$  puisse se confondre avec  $Z$  il faut que  $\mu = 1$ .

D'où  $\Delta [Z', \zeta] = \Delta [Z', Z] + \Delta [Z, \zeta]$

L'estimateur est uniformément biaisé. Pour cela il suffit qu'on ait :

$$\Delta [Z, \zeta] = \rho \cdot \{\psi [\zeta] - \psi [Z]\} \quad \text{avec } \rho = \rho [\zeta]$$

THEOREME III bis -

a) Si l'on veut  $W [Z, \zeta]$  symétrique, il suffit de poser

$$\begin{aligned} W [Z, \zeta] &= \{\psi [Z] - \psi [\zeta]\}^2 \\ &= \psi^2 [Z] - \psi^2 [\zeta] + 2\psi [\zeta] \{\psi [\zeta] - \psi [Z]\} \end{aligned}$$

d'où  $\Phi [Z] = \psi^2 [Z]$

Réciproquement :  $W = W^*$  avec  $\Delta = \rho (\psi - \psi^*)$

$$W \equiv (\rho - \rho^*)(\psi - \psi^*)/2$$

impliquent  $\Phi^{**} - \Phi \equiv -(\rho + \rho^*)(\psi - \psi^*)/2$

$$\equiv \frac{\rho^* \psi^*}{2} - \frac{\rho \psi}{2} + \frac{1}{2} (\rho \psi^* - \rho^* \psi)$$

d'où  $\rho \psi^* - \rho^* \psi = 0$ ,  $\frac{\rho^*}{\psi^*} = \frac{\rho}{\psi} = 2\lambda$  (constante)

d'où  $\Phi \equiv \lambda \cdot \psi^2$

b) En dehors du cas symétrique, on peut vouloir exiger seulement de  $W$  d'être positive (sauf  $W [\zeta, \zeta] = 0$ ), croissante avec  $|Z - \zeta|$ . On a

$$W \equiv (\Phi^* - \rho \psi^*) - (\Phi - \rho \psi)$$

Posons

$$\Omega = \Phi - \psi^2; \quad \text{faisons } \rho = 2\psi;$$

d'où

$$W = (\psi^* - \psi)^2 + \Omega^* - \Omega$$

avec

$$\Omega = \Omega[\zeta], \quad \Omega^* = \Omega[Z]$$

En exigeant  $\Omega^* > \Omega$  pour  $|Z - \zeta| > 0$ , on remplit les conditions voulues.

### Conclusion.

Lorsque l'estimateur  $Z$  de  $\zeta$  est biaisé, la perte d'information peut se calculer comme pour un estimateur sans biais, à condition : que  $Z$  soit "uniformément" biaisé :  $E_{\omega}[Z_2, Z_1] = 0$ ; qu'on se contente de conditions suffisantes, là où on avait des conditions nécessaires et suffisantes moyennant une régularité suffisante des fonctions.

D'ailleurs en supposant que les fonctionnelles  $\Phi[Z]$  ou  $\psi[Z]$  sont des fonctions de point (le nombre de dimensions de l'espace  $Z$  pouvant même être infini) et en supposant l'existence des dérivées partielles, on peut aller plus loin - on l'a vu aux n°7, 8.

Enfin l'existence des fonctionnelles  $\omega$ ,  $\Phi$  ou  $\psi$  pour tous les  $Z$  ( $y$  compris  $\zeta$ ) sous-entend en fait que  $Z$  et  $\zeta$  aient une expression commune (sans aller peut-être jusqu'à l'isomorphisme des n°7-8).

### III - PERTES AU SENS DU CHAPITRE II -

La perte d'information envisagée ici concerne un plan de sondage bien déterminé et a un certain ordre; il ne lui correspond plus spécialement telle ou telle estimation. On connaît déjà les pertes d'ordre  $r$  (et leurs combinaisons linéaires) relatives à un plan de sondage dans une seule urne (Chapitre II). On se propose de montrer que (sous des conditions relativement peu restrictives) il n'existe pas d'autres solutions de l'équation de Chasles que les dites combinaisons linéaires, avec un plan de sondage cette fois très général.

On suppose invariable la structure du sondage, c'est dire que les plans de sondage ( $\pi$ ) de la famille ( $F$ ) sont figurés par les sommets d'un certain treillis.

Quand on sonde une urne unique, l'axe d'information possède un certain point ( $\omega$ ) correspondant à la plus grande perte d'information (non infinie) (c'est-à-dire que  $\omega$  est le point  $n = 1$  pour estimer  $\bar{x}$ ,  $n = 2$  pour estimer  $\sigma^2$  et  $\bar{x}$ ,  $n = 3$  pour estimer  $\text{cov}(xy)$ ,  $\sigma_x^2$ ,  $\sigma_y^2$ ,  $\bar{x}$  et  $\bar{y}$  etc.).

Ce point  $\omega$  sera l'horizon.

L'horizon d'un treillis de sondage.

Soit une famille  $F$  de plans de sondage de treillis

$$(\lambda_1, \lambda_2 \dots \lambda_k)$$

avec  $\lambda_i \geq r$ . Le point  $(r \ r \dots r)$  en constitue l'horizon.

Exemples.

Sondage stratifié  $r \leq n_h \leq v_h$

Sondage à 2 degrés  $r \leq m \leq \mu, \quad r \leq n_i \leq v_i$

Cas d'un treillis découpé en sous-ensembles.

Considérons le cas d'un échantillon stratifié a posteriori - s'il est tiré sans remise et d'ordre  $r$ , figuré sur un axe par les points  $(n \geq r)$ , le point  $(\omega)$  ( $n = r$ ) est l'horizon. En revanche si l'on considère le treillis du sondage stratifié normal dont il dérive (treillis sur lequel on a convenu par exemple que l'estimateur banal, - non stratifié - était affecté aux points tels qu'un  $n_h$  au moins est inférieur à  $r$ ) il existe un sous-ensemble  $\omega$  ( $n = r$ ) et non un point unique  $\omega$ .

Conventions.

Pour simplifier on supposera qu'il existe sur le treillis un horizon et un seul, où aboutissent tous les trajets  $[U \pi \pi' \dots]$  issus de  $U$  (point figurant la connaissance complète); en appelant "trajet" une ligne brisée (empruntant exclusivement les côtés du treillis) le long de laquelle la taille de l'échantillon décroît (c'est-à-dire que l'un des "effectifs" de l'échantillon décroît).

Si d'autres cas se présentent, on ne retiendra que ceux qui peuvent se ramener aux conventions ci-dessus.

Exemple : Sondage à deux phases.

Le treillis  $(m, n_1 \dots n_h)$  comprend un axe (représentant la première phase de sondage) sur qui s'appuie une famille de génératrices  $(m)$  portant les points figurant la seconde phase. Il n'existe donc pas d'horizon ( $\omega$ ). On sort des hypothèses fixées. (Toutefois ce treillis peut être remplacé par un autre, comme on le verra au chapitre VI; et avec des composantes négatives, on pourrait étendre les résultats qui vont suivre).

## THEOREME I -

Soit I et J deux échantillons du plan de sondage  $\pi''$ . Soit  $g(I, J)$  une fonction symétrique de I et J, nulle quand I et J sont confondus, positive en tout autre cas;  $g$  pouvant d'ailleurs dépendre également de la distribution de base (qu'on désigne par U). Si la perte de  $\pi''$  est de la forme :

$$p(\pi'') = \mathcal{E} g(I, J)$$

et si  $\pi'$  précède  $\pi''$  sur les trajets  $[U\pi'']$  (au sens précédent),  $\pi'$  et  $\pi''$  étant sur le même axe d'information, on a par définition :

$$p(\pi'') = p(\pi') + \mathcal{E} P(\pi''/\pi')$$

d'où

$$p(\pi') = p(\pi'') - \mathcal{E} P(\pi''/\pi')$$

ou

$$p(\pi') = \mathcal{E}_u g(I, J) - \mathcal{E}_u \mathcal{E}_n g(I, J)$$

de la forme

$$p(\pi') = \mathcal{E}_u g(I, J) \cdot [1 - \varphi(\pi''/\pi'')]$$

De même, si  $\pi$  précède  $\pi'$  sur les trajets  $[U\pi']$ , il précède aussi  $\pi''$  a fortiori; et on a :

$$p(\pi) = \mathcal{E}_u g(I, J) - \mathcal{E}_u \mathcal{E}_n g(I, J)$$

La condition d'alignement étant (chapitre III, § II) pour les trois points  $\pi$   $\pi'$   $\pi''$  :

$$\mathcal{E} P(\pi''/\pi) - \mathcal{E} P(\pi''/\pi') = \mathcal{E} P(\pi'/\pi),$$

il s'en suit qu'on peut écrire

$$p(\pi) = \mathcal{E}_u g(I, J) [1 - \varphi(\pi/\pi'')]$$

$$p(\pi') - p(\pi) = \mathcal{E}_u g(I, J) [\varphi(\pi/\pi'') - \varphi(\pi'/\pi'')]$$

## THEOREME 2 -

Soit I, J, K trois échantillons du plan de sondage  $\pi''$  et  $g(I, J, K)$  une fonction symétrique de I, J, K, nulle quand I, J et K sont tous trois confondus et positive en tout autre cas. Supposons la perte  $p(\pi'')$  de la forme

$$p(\pi'') = \mathcal{E}_u g(I, J, K)$$

Si  $\pi'$  précède  $\pi''$  et  $\pi$  précède  $\pi'$  sur un même axe d'information, on a encore :

$$p(\pi) = \mathcal{E}_U g(I, J, K) \cdot [1 - \omega(\pi/\pi'')] ]$$

$$p(\pi') - p(\pi) = \mathcal{E}_U g(I, J, K) \cdot [\omega(\pi/\pi'') - \omega(\pi/\pi'')] ]$$

Ceci s'étend à un nombre quelconque d'échantillons. Dans ce qui suit, ces extensions à plus de deux échantillons ne seront plus explicitement formulées, encore qu'elles demeurent possibles en principe.

Cas où  $\pi''$  et  $\omega$  coïncident.

Supposons en particulier que le plan de sondage  $\pi''$  soit tel que la perte d'information ne soit plus définie pour aucun trajet  $[U\pi]$  où  $\pi''$  précède  $\pi$ ; ce qui revient à dire que  $\pi''$  est l'horizon  $\omega$ . On a donc :

THEOREME 3 -

$$p(\pi) = \mathcal{E}_U g(I, J) [1 - \theta(\pi)] \quad \text{avec } \theta(\omega) = 0,$$

pour tous les plans de sondage  $\pi$  qui sont sur des axes d'information aboutissant au point  $\omega$ , pourvu que la perte relative au plan de sondage ( $\omega$ ) soit elle-même de la forme

$$\mathcal{E}_U g(I, J)$$

Combinaisons linéaires de fonctions  $g(I, J)$ .

THEOREME 4 -

Des fonctions  $g$  différentes (représentant chacune une distance entre deux échantillons) peuvent être envisagées simultanément et combinées linéairement, ce qui donne :

$$p(\omega) = \sum_i \lambda_i \mathcal{E}_U g_i(I, J)$$

$$p(\pi') - p(\pi) = \sum_i \lambda_i \mathcal{E}_U g_i(I, J) [\theta_i(\pi) - \theta_i(\pi')]$$

Les coefficients  $\lambda$  sont des constantes, à moins qu'ils ne dépendent de la distribution de base  $U$ . Il n'est pas nécessaire qu'ils soient tous positifs, mais la combinaison  $p(\pi)$  doit toujours rester positive (avec  $p(U) = 0$ ).

Remarque. Cas de pertes d'information d'ordres divers.

La difficulté suivante se présentera. On tire des boules d'une urne unique.

D'une part soit  $r = 2$ . On tire  $I(x_h, x_k)$  et  $J(x_{h'}, x_{k'})$  avec quoi on forme par exemple :

$$g_1(I, J) = [(x_h - x_k)^2 - (x_{h'} - x_{k'})^2]^2;$$

d'autre part on choisit  $r = 1$ , on tire  $i(x_i)$  et  $j(x_j)$  avec quoi on forme, disons :

$$g_2(i, j) = (x_i - x_j)^4.$$

Il n'est pas question de former  $\lambda_1 g_1 + \lambda_2 g_2$ .

On adopte comme horizon  $\varpi$  le plan de sondage d'effectif  $n = 2$ ; et l'expression  $g_2$  ne saurait être  $(x_i - x_j)^4$  mais par exemple :

$$(x_h + x_k - x_{h'} - x_{k'})^4$$

ou bien encore  $(x_h - x_{h'})^4 + (x_h - x_{k'})^4 + (x_k - x_{h'})^4 + (x_k - x_{k'})^4$

Telle est la signification du théorème 4.

Réciproque (ou THEOREME 5).

Toute perte  $p(\varpi)$  est nécessairement de la forme :

$$p(\varpi) = \sum_{\mu} g(I, J \dots) [\theta(\mu) - \theta(r)]$$

(sous réserve qu'il existe un horizon  $\varpi$ ).

L'énoncé de ce résultat très important et très général va être précisé au cours de la démonstration.

a) Considérons les arêtes du treillis, ayant leur extrémité en  $\varpi$  et leur origine en des points  $\pi^0$  (qui précèdent immédiatement  $\varpi$  sur le trajet  $[\nu \dots \varpi]$ ). Considérons l'un de ces ( $\pi^0$ ) et soit  $E^0$  l'un de ses échantillons.

$E^0$  renferme plusieurs échantillons de ( $\varpi$ ), qu'on désignera par  $ij \dots$ , - et leur donne naissance quand on tire des échantillons de plus en plus petits (en tirant au sort les unités chassées de l'échantillon).

b) La perte d'information supplémentaire, en l'occurrence  $P(\varpi/E^0)$ , est de la forme  $g'(ij \dots)$ .

En effet : elle est nulle si tous les  $ij \dots$  sont identiques entre eux puis-qu'alors on ne perd rien à en perdre un; elle est positive en tous autres cas; elle dépend symétriquement des  $ij \dots$  si ceux-ci sont équiprobables.

c) Par suite l'expression  $P(\varpi/\pi^0)$  est de la forme  $g'(ij \dots)$

et, en posant :

$$g' \equiv g_1 [\theta(\pi^0) - \theta(\varpi)] \quad \text{avec } \theta(\varpi) = 0, \theta \text{ croissante}$$

$$\text{on a} \quad p(\varpi) - p(\pi^0) \equiv \mathcal{E}g_1 \cdot [\theta(\pi^0) - \theta(\varpi)]$$

Remontons alors l'axe  $\pi^0\varpi$  sur toute sa longueur; on aura (théorème 3 ci-dessus) :

$$p(\varpi) - p(\pi) \equiv \mathcal{E}g_1 [\theta(\pi) - \theta(\varpi)]$$

$\theta$  étant une fonction croissante quand la taille de l'échantillon augmente. Repérons les plans de sondage  $\pi$  de cet axe par des indices  $r$  pour  $\varpi$ ,  $(r+1)$  pour  $\pi^0$ ,  $(r+2)$  etc.  $m$  (pour  $\pi$ ). La suite des  $m$  est, soit bornée par un certain  $\mu$ , soit infinie (tirages avec remise et on arrive alors à un point d'accumulation  $\mu^\omega$ ); dans les deux cas on pourra écrire, arrivé à l'extrémité de l'axe opposée à  $\varpi$  :

$$p(\varpi) - p(\pi_\mu) \equiv \mathcal{E}g_1 [\theta_\mu - \theta_r]$$

Or on a supposé (par hypothèse) que tout  $\pi$  pouvait être atteint par un trajet issu de  $U$  et aussi que tous ces trajets aboutissaient en  $\varpi$ . Il existe donc au moins un trajet  $[U \pi_\mu \varpi]$  qu'on va remonter de  $\pi_\mu$  vers  $U$  sur un autre axe (le premier ayant été épuisé). Axe qui d'ailleurs comme le premier peut n'être qu'une arête unique.

Le long de cet axe,  $p(\pi_\mu) - p(\pi)$  est encore de la forme

$$\mathcal{E}g_2 [\omega_n - \omega_s]$$

où  $\omega$  désigne une fonction croissante quand la taille de l'échantillon augmente;  $g_2$  désigne une fonction symétrique d'un certain nombre d'échantillons de  $(\pi_\mu)$ .

De proche en proche, on arrivera ainsi au point  $U$ ,  $p(\varpi) - p(\pi)$  deviendra

$$p(\varpi) - p(U) \equiv p(\varpi)$$

de la forme

$$p(\varpi) = \mathcal{E}g_1(\theta_\mu - \theta_r) + \mathcal{E}g_2(\omega_\nu - \omega_s) + \dots + \mathcal{E}g_\lambda(\rho_\tau - \rho_\tau) \quad \text{cqfd}$$

### Corollaire.

Toute perte  $p(\pi)$  est nécessairement de la forme précédente où  $\theta_m$  remplace  $\theta_r$ . (Démonstration analogue en suivant un trajet  $[U \pi]$  à l'envers, en remontant toujours les axes d'information (jusqu'à leur origine par exemple). C'est le théorème annoncé au Ch. II, § IV, Réciproque du Th. 2.

Remarque.

1) La décomposition de  $p(\pi)$  ou  $p(\varpi)$  en  $\lambda$  composantes n'est pas unique. On verra plus loin qu'avec les sondages à  $\lambda$  strates on peut définir strictement  $\lambda$  composantes, tandis que pour les sondages à  $\lambda$  degrés on pourra en définir strictement  $2^\lambda - 1$  (si chaque degré de sondage est indépendant des autres).

2) Les fonctions  $\theta, \omega, \dots, \rho, \psi$  considérées plus haut ne dépendent pas seulement de la taille des échantillons. Lorsque tous les tirages au sort sont équiprobables, il est bien clair que ces fonctions se déduisent d'un calcul facile d'analyse combinatoire. Mais considérons par exemple une stratification a posteriori en deux strates d'un échantillon bernoullien de taille  $n$ : l'expression de  $p(n)$  n'est pas aussi simple. Par exemple on aurait ( $p_1$  et  $p_2$  désignant les proportions de la population dans chaque strate).

$$\left\{ \begin{array}{l} n_1 \neq 0, \quad n_2 \neq 0 \\ \\ n_1 n_2 = 0 \end{array} \right. \left\{ \begin{array}{l} \bar{X} = p_1 \bar{X}_1 + (1 - p_1) \bar{X}_2; \quad 1 - p_1 = p_2 \\ V \bar{X} = p_1^2 \sigma_1^2 \mathcal{E} \left( \frac{1}{n_1} \right) + p_2^2 \sigma_2^2 \mathcal{E} \left( \frac{1}{n - n_1} \right) \\ \\ \bar{X} = (n_1 \bar{X}_1 + n_2 \bar{X}_2) / n; \quad V \bar{X} = \sigma^2 / n. \end{array} \right.$$

On peut poser  $\mathcal{E}(1/n_1) = \theta(\infty) - \theta(n)$ , mais ne pas oublier qu'une telle expression dépend de  $p_1$ .

Le problème de la forme des fonctions  $g(I, J)$ .

Soit  $I, J$  un couple d'échantillons tiré au sort et  $\mathcal{E}$  l'espérance mathématique des tirages avec remise. On sait que

$$V Z = \mathcal{E}(Z - \zeta)^2$$

peut s'écrire, quel que soit le plan de sondage :

$$V Z = \frac{1}{2} \mathcal{E}(Z_i - Z_j)^2 = \mathcal{E}g(I, J)$$

avec  $g(I, J) = (Z_i - Z_j)^2 / 2$  (en supposant  $\mathcal{E}Z = \zeta$ )

Ainsi pour une suite de plans de sondage de tailles décroissantes extraite d'une famille  $F$ , la perte d'information  $V Z$  conserve à toutes les étapes la forme  $\mathcal{E}g(I, J)$  avec la même fonction  $g$ .

En particulier si la perte est d'ordre 1, l'estimateur  $z(X)$  prend pour  $n = 1$  les valeurs  $z(x_i)$  et  $z(x_j)$  sur le couple échantillon  $x_i, x_j$ ; et on a :

$$g(i, j) = [z(x_i) - z(x_j)]^2 / 2$$

(en supposant  $\mathcal{E}z(X) = \zeta$ ).

Il n'y a aucune raison pour qu'on ait (avec d'autres fonctions  $g$ ) :

$$\mathcal{E}_u g(I, J) \equiv \mathcal{E}_u g(i, j) - \mathcal{E}_u \mathcal{E}_\pi g(i, j)$$

et on a peu de chances de se tromper en voyant dans cette conservation de la forme une propriété de la seule variance, parmi toutes les pertes d'information au point de vue du chapitre II.

Mais on n'a pas trouvé le chemin qui conduirait à un résultat aussi général (qui est assez analogue aux problèmes d'agrégation posés par d'autres théories).

#### IV - PERTE D'INFORMATION D'UN SONDAGE OU D'UNE ESTIMATION ?

##### 1/ - Enoncé du problème.

Dans quel cas est-il possible d'expliquer une perte d'information au sens du Chapitre II (concernant un plan de sondage) en montrant que c'est en fait une perte d'information au sens du Chapitre I (relative à l'estimation d'un paramètre ou d'un groupe de paramètres) ?

Nous sommes loin d'avoir résolu ce problème, mais allons donner quelques résultats fragmentaires le concernant.

##### 2/ - Un principe général.

S'il n'était d'estimations qu'estimations sans biais, il serait possible déjà de poser un principe, d'une application d'ailleurs délicate :

Toute perte au sens du Chapitre I peut se mettre sous la forme :

$$\mathcal{E}\Phi(Z) - \Phi(\zeta)$$

et correspond alors à l'estimation sans biais de  $\sqrt{\Phi(\zeta)}$  par  $\sqrt{\Phi(Z)}$  avec un poids  $W$  symétrique, plus des estimations diverses avec poids dissymétrique.

Avec les estimations biaisées, limitées au cas isomorphe comme on l'a fait ici, le même principe fait intervenir des fonctionnelles :

$$\mathcal{E}\Phi[Z] - \Phi[\zeta]$$

Enfin les théorèmes qu'on évoque ici ne sont vérifiés qu'autant que les fonctions ou fonctionnelles auxquelles on a recours possèdent des dérivées pour  $\zeta = Z$ ; l'existence des dérivées de tous ordres est

requis pour le développement en série; une telle condition est peut-être trop restrictive; mais le théorème laisse échapper à coup sûr le cas de  $W = |Z - \zeta|$ , où la dérivée pour  $Z = \zeta$  n'existe pas; or on sait qu'il existe des pertes  $\mathcal{E}|Z - \zeta|$ , au moins pour les estimations biaisées du "range" et de la plus grande (ou petite) valeur - Les conditions de régularité nécessaires pour établir les théorèmes généraux laissent donc échapper des cas intéressants.

3/ - Un exemple d'interprétation.

Limitons-nous aux tirages équiprobables dans une urne unique; et étudions le cas de

$$p(v, n) = \mathcal{E}(x_i - x_j)^4 \cdot [\theta(v) - \theta(n)]$$

Cette perte d'information au sens du Chapitre II correspond-elle à une estimation? On précisera ceci ultérieurement.

Calcul de  $\alpha$  =  $\sum_i \sum_j (x_i - x_j)^4 / v^2$ .

Partant de  $\frac{1}{v} \sum_1^v (\xi - x_j)^4 = \xi^4 - 4 \xi^3 \mu_1 + 6 \xi^2 \mu_2 - 4 \xi \mu_3 + \mu_4$  et en y faisant successivement :  $\xi = x_1, x_2 \dots x_v$ , sommant et divisant par  $v$ , il vient :

$$\alpha = 2(\mu_4 - 4\mu_3\mu_1 + 3\mu_2^2)$$

les  $\mu$  désignant les moments non-centrés des  $x_i$ ;  $\mu_1 = \bar{x}$ ; exprimé avec les moments centrés, il vient :

$$\alpha = 2(\mu_4^0 + 3\sigma^4)$$

Divers essais pour mettre  $\alpha$  sous la forme  $\mathcal{E}W(z_i, \zeta)$  échouent; mais on trouve autre chose.

Première interprétation.

Il était bien naturel qu'on rapprochât  $\alpha$  et plus généralement  $\sum \sum (x_i - x_j)^{2r} / v^2$ , avec le minimum de la fonction

$$y(\xi) = \sum_i (x_i - \xi)^{2r}$$

lorsque  $\xi$  décrit l'axe des  $x$ ; -minimum qui a lieu pour

$$z(\xi) = \sum_i (x_i - \xi)^{2r-1} = 0$$

équation qui admet une et une seule racine positive, vu que  $z'(\xi)$  est nécessairement positive; en particulier, pour  $r = 1$ , c'est  $\bar{x}$ .

Dans le cas plus particulier de  $r = 2$ , désignons par le minimum de  $y/v$ , atteint pour  $\xi = \xi_0$ . On a :

$$y/v = \xi^4 - 4\mu_1\xi^3 + 6\mu_2\xi^2 - 4\mu_3\xi + \mu_4$$

$$0 = [\xi_0^3 - 3\mu_1\xi_0^2 + 3\mu_2\xi_0 - \mu_3] \xi_0$$

$$\eta = \xi_0^4 - 4\mu_1\xi_0^3 + 6\mu_2\xi_0^2 - 4\mu_3\xi_0 + \mu_4$$

d'où 
$$\eta = (\mu_4 - \mu_1\mu_3) - 3\xi_0(\mu_3 - \mu_1\mu_2) + 3\xi_0^2(\mu_2 - \mu_1^2)$$

Rapprochons cette expression de celle de

$$\frac{1}{2}\alpha = (\mu_4 - \mu_1\mu_3) - 3\mu_1(\mu_3 - \mu_1\mu_2) + 3\mu_2(\mu_2 - \mu_1^2)$$

Pour qu'elles coïncident, il faut et il suffit qu'on ait :

$$\mu_1^2 = \xi_0^2 = \mu_2$$

ce qui est exclu. Mais la ressemblance formelle entre les deux formules justifie qu'on tienne  $\alpha$  pour une distance entre les  $x_i$  et un point fictif où  $\xi_0$  est remplacé par  $\mu_1$  et  $\xi_0^2$  par  $\mu_2$ .

Il est clair que ceci s'étend à toutes les sommes de puissances paires des  $(x_i - x_j)$ .

### Seconde interprétation.

La première interprétation ne permet d'ailleurs pas de retrouver les estimations sans biais de  $\mu_1$  et  $\mu_2$ ; car la variance simultanée s'écrit

$$\begin{aligned} \mathcal{E}[(x_i^2 - \mu_2 + t(x_i - \mu_1))]^2 &= \mathcal{E}(x_i^2 - \mu_2)^2 + 2t\mathcal{E}(x_i^2 - \mu_2)(x_i - \mu_1) + t^2\mathcal{E}(x_i - \mu_1)^2 \\ &= (\mu_4 - \mu_2^2) + 2t(\mu_3 - \mu_1\mu_2) + t^2(\mu_2 - \mu_1^2) \end{aligned}$$

alors qu'on a 
$$\frac{\alpha}{2} = (\mu_4 - \mu_2^2) - 4\mu_1(\mu_3 - \mu_1\mu_2) + 4\mu_2(\mu_2 - \mu_1^2)$$

Or il n'est pas possible de faire  $t^2 = 4\mu_2$ ,  $t = -2\mu_1$  simultanément, car :

$$\mu_2 \neq \mu_1^2$$

En revanche on peut remplacer  $t$  par  $(-2x_j)$ , où  $x_j$  est supposé indépendant de  $x_i$ .

Considérons alors la variable (où  $x_i$  et  $x_j$  sont indépendants)

$$y_k = x_i^2 - 4 x_i x_j + x_j^2$$

on a

$$\mathcal{E} y_k = 2\mu_2 - 4\mu_1^2$$

et

$$\mathcal{V} y_k = \frac{\alpha}{2}$$

1) Il s'agit de tirages bernoulliens.

2) Il s'agit d'une perte d'ordre 2, de sorte que  $p$  n'est perte que si la fonction  $\theta$  est d'ordre 2 (contrairement à ce qu'on avait pu penser, la seule connaissance de  $\alpha$  ne permet donc pas de trancher la question).

Remarque.

D'ailleurs le calcul de  $\mathcal{V} S^2$  ( $nS^2/n - 1$ , estimateur de  $v\sigma^2/v - 1$ ) conduit à une composante en  $\alpha$ . Les combinaisons linéaires de plusieurs variances conservent les propriétés de récurrence. Bien des informations au sens du Chapitre II ne sont que des composantes des informations au sens du Chapitre I; c'est peut-être la véritable "explication" de leur existence.

4/ - Une méthode de calcul (tirages dans une seule urne).

Considérons une perte d'information d'ordre 1

$$\mathcal{E} g(i, j) \left( \frac{1}{n} - \frac{1}{v} \right)$$

où  $g$  est de la forme  $g(z_i, z_j)$ , avec  $z = h(x)$ ,

$$g > 0, \text{ sauf } g(i, i) = 0.$$

Pour  $n = 1$ , on imagine un estimateur  $Z$  prenant les valeurs  $z_i$ ; on ne peut poser

$$W = g(Z, \zeta)$$

En revanche :

THEOREME -

On peut toujours poser

$$W(Z, \zeta) = \mathcal{E}_j g(Z, z_j) - \mathcal{E}_j g(\zeta, z_j) + g(Z, \zeta)$$

Effectivement on vérifie que  $\mathcal{E} W(Z, \zeta)$  est identique à

$$\mathcal{E}_{ij} g(z_i, z_j)$$

et que  $W(\zeta, \zeta) = 0$ , quel que soit  $\zeta$ .

#### Choix de $\zeta$ .

Lorsque  $Z = \zeta + \varepsilon$ ,  $g(Z, \zeta)$  est positif très petit;  $g(\zeta + \varepsilon, z_j) - g(\zeta, z_j)$  est du signe de  $\varepsilon$  ( $\varepsilon - z_j$ ); si l'on veut être assuré que  $W$  est toujours positif (hors le cas  $Z = \zeta$ ), il faut donc que

$$\mathcal{E}_j g(\zeta, z_j)$$

soit le minimum de  $\mathcal{E}_j g(\zeta + \varepsilon, z_j)$ . Ceci conduit à choisir  $\zeta$  tel que  $\min \mathcal{E}_j g(Z, z_j) = \mathcal{E}_j g(\zeta, z_j)$ .

#### Remarque.

Tout ceci suppose  $n = 1$ . Passons au cas où  $n$  est quelconque. Par hypothèse

$$p(v, n) + \mathcal{E}P(n, 1) = p(v, 1)$$

$$p(v, n) = p(v, 1) - \mathcal{E}P(n, 1)$$

On vient d'interpréter  $p(v, 1) = \mathcal{E}W(Z_1, \zeta)$  et on a de même

$$P(n, 1) = EW(Z_1, Z_n)$$

$Z_n$  étant défini dans l'urne aléatoire de taille  $n$  par la même condition que  $\zeta$  l'est dans l'univers.

#### Convention.

On peut alors convenir que  $p(v, n)$  représente la distance entre  $\zeta$  et  $Z_n$ , la perte d'information sur  $\zeta$  quand on connaît  $Z_n$ . Ceci est purement conventionnel cette fois, car

$$p(v, n) = \mathcal{E}W(Z_1, \zeta) - \mathcal{E}EW(Z_1, Z_n)$$

n'implique pas  $p(v, n) = \mathcal{E}W(Z_n, \zeta)$  ni même  $\mathcal{E}W[Z_n, \zeta]$ .

$g$  est une fonction;  $W(Z_1, \zeta)$  et  $W(Z_1, Z_n)$  sont des fonctions.

Nous n'avons aucune raison de choisir une fonctionnelle déterminée  $W[Z_n, \zeta]$ .

Application.  $g = |z_i - z_j|^a$ .

En écrivant que  $g(Z, z_j)$  est minimum pour  $Z = \zeta$ , on retrouve une propriété bien connue de la médiane (ou de l'intervalle médian) lorsque  $a = 1$ ; tandis que, pour  $a = 2$  on a (bien entendu) la moyenne. Pour une valeur quelconque de  $a$ , il faudrait d'abord discuter l'existence et l'unicité de  $\zeta$ .

#### 5/ - Etude directe de l'estimation de la médiane.

##### Entrée en matières.

Pour simplifier on supposera impair l'effectif de la population (soit  $2\nu + 1$ ) et celui de chaque échantillon (soit  $2n + 1$ ).

Partant d'une population entièrement connue, on passera à des échantillons d'effectif décroissant par deux unités à la fois  $n = 2\nu - 1, 2\nu - 3, \dots, 5, 3, 1$ .

Désignons les médianes par  $\zeta = Z_{2\nu+1}; Z_{2\nu-1}; Z_{2\nu-3}; \dots, Z_5, Z_3, Z_1$ .

En particulier (si les  $x_i$  sont tous distincts) :

a)  $Z_{2\nu-1}$  coïncide avec  $\zeta$  si les  $x_i$  inconnues sont l'un plus grand, l'autre plus petit que  $\zeta$ ; dans le cas contraire  $Z_{2\nu-1}$  coïncide avec la valeur  $x_i$  immédiatement supérieure, - ou inférieure, - à  $\zeta$ .

b)  $Z_1$  a une chance sur  $(2\nu + 1)$  de coïncider avec chacune des valeurs  $x_i$ .

c) Les cas intermédiaires sont moins simples; supposons  $\nu = 5$ .

Soit  $a > b > c > d > e$  ( $\zeta = c$ ) les valeurs  $x_i$  considérées. On voit que :

$$\text{pour } n = 3, \quad Z_3 = \begin{cases} c & (\text{probabilité } 4/10) \\ b \text{ ou } d & (\text{probabilité } 3/10) \end{cases}$$

Sauf lorsque la distribution des  $x_i$  est symétrique,  $Z$  est un estimateur biaisé de  $\zeta$ .

##### Perte d'information liée à l'estimation de la médiane.

a) La moyenne et la médiane présentent des analogies importantes. Soit  $x$  un point courant de l'axe,  $\bar{x}$  la moyenne des  $x_j$  et  $\tilde{x}$  leur médiane.

$$\text{On sait que :} \quad \begin{aligned} \min \sum |x_j - x| &= \sum |x_j - \tilde{x}| \\ \min \sum (x_j - x)^2 &= \sum (x_j - \bar{x})^2 \end{aligned}$$

On est donc tenté de voir dans  $\sum |x_i - \bar{x}|/v$  une perte d'information; en fait cette expression ne possède pas les propriétés nécessaires.

Montrons-le sur l'exemple précédent des cinq boules.

La perte d'information devrait s'évaluer, pour  $n' = 1$  :

soit directement par :  $[|ca| + |cb| + |cd| + |ce|]/5$

soit en deux étapes, par :  $[|cb| + |cd|] \frac{3}{10}$  pour  $n = 3$

pour  $n' = 1$  ( $n = 3$ ), par :

10 combinaisons 3 à 3 de 5 lettres	}	acd	+	[ ca  +  cd ]/30
		ace	+	[ ca  +  ce ]/30
		bcd	+	[ cb  +  cd ]/30
		...		.....
		cde	+	[ dc  +  de ]/30

L'identification des deux expressions de l'information donne :

$$\frac{2}{10} [|ae| + |bd|] \equiv \frac{3}{10} |bd| + \left[ \frac{2}{10} |ae| + \frac{1}{10} |bd| \right]$$

c'est-à-dire  $|bd| = 0$

(b, c, d confondus) ce qui est exclu.

b) On est tenté de même de remplacer  $(x_i - x_j)^2/2$  par  $|x_i - x_j|$  dans l'expression de la perte d'information relative à  $\bar{X}$ , pour obtenir celle relative à la médiane échantillon  $\tilde{X}$ , soit :

$$p(v, n) = \frac{\sum |x_i - x_j|}{v(v-1)} \left( \frac{1}{n} - \frac{1}{v} \right) = \frac{1}{n} \frac{\sum \sum |x_i - \bar{x}_j|}{v^2} \cdot \frac{v-n}{v-1}$$

C'est incontestablement une perte d'information. Donnons une expression de

$$\alpha = \frac{1}{v^2} \sum \sum |x_i - x_j|$$

qui rappelle beaucoup celle du paragraphe a, ci-dessus.

Calcul de  $\alpha$ .

Désignons par  $x_n$  la suite ordonnée :

$$x_1 \leq x_2 \leq \dots \leq x_v$$

Il vient :

$$1 \leq h \leq v - 1 \quad \sum_h (x_\nu - x_h) + (x_h - x_1) = (v - 1)(x_\nu - x_1)$$

$$2 \leq h \leq v - 2 \quad \sum_h (x_{\nu-1} - x_h) + (x_h - x_2) = (v - 3)(x_{\nu-1} - x_2)$$

etc.

etc.

Si v est impair, on termine sur  $2 \left[ x_{\frac{\nu+3}{3}} - x_{\frac{\nu-1}{2}} \right]$

Si v est pair, on termine sur  $1. \left[ x_{\frac{\nu+1}{2}} - x_{\frac{\nu}{2}} \right]$

Pour v impair, on a donc :

$$\alpha = \frac{2}{v^2} \left[ (v - 1)(x_\nu - x_1) + (v - 3)(x_{\nu-1} - x_2) + \dots + 2(x_{\frac{\nu+3}{2}} - x_{\frac{\nu-1}{2}}) \right]$$

Avec

$$\begin{cases} x_\nu - x_1 \equiv |x_\nu - \tilde{x}| + |\tilde{x} - x_1| \\ x_{\nu-1} - x_2 \equiv |x_{\nu-1} - \tilde{x}| + |\tilde{x} - x_2| \end{cases} \quad \text{etc.}$$

On peut écrire également :

$$\alpha = \frac{4}{v^2} \sum_h h(x_h - \tilde{x})$$

en désignant cette fois par  $x_h$  la suite  $\dots x_{-3} x_{-2} x_{-1} \tilde{x} x_1 x_2 x_3 \dots$

Remarque 1.

Affecter aux points  $x_i$  des poids  $\varpi_i$  symétriques par rapport à leur médiane ne modifie en rien celle-ci; on a donc encore :

$$\min \sum \varpi_i |x_i - x| = \sum \varpi_i |x_i - \tilde{x}|$$

c'est le cas en particulier pour

$$\frac{\varpi_\nu}{v-1} = \frac{\varpi_{\nu-1}}{v-3} = \frac{\varpi_{\nu-2}}{v-5} = \dots = \frac{\varpi_1}{v-1}$$

(le point médian étant affecté d'un poids arbitraire).

Remarque 2.

Si  $\alpha$  est la perte d'information  $p(v, 1)$  relative à l'estimation de  $\tilde{x}$  par l'un quelconque des  $x_j$  de la population,  $P(n, 1)$  relative à la médiane  $\tilde{X}$  d'un échantillon donné lui est isomorphe; c'est-à-dire :

$$A_n = \frac{4}{n^2} \sum h(x_h - \tilde{X}) \equiv \frac{1}{n^2} \sum |x_i - x_j|$$

De l'identité

$$\mathcal{E} \left( \frac{n A_n}{n-1} \right) \equiv \frac{v\alpha}{v-1}$$

généralisation du Lemme 2, découle bien :

$$\mathcal{E} P(n, 1) = \alpha \frac{v(n-1)}{n(v-1)} = \alpha \left( 1 - \frac{v-n}{n(v-1)} \right)$$

c'est-à-dire :  $p(v, n) = \frac{\alpha}{n} \frac{v-n}{v-1}$ .

En fait il s'agit d'une convention arbitraire.

Exemple.

$$n = v - 2; \quad p(v, v-2) = \alpha \frac{2}{(v-1)(v-2)}$$

est incompatible avec  $p(v, v-2) = \mathcal{E} W(\tilde{X}_{v-2}, \tilde{x})$  où  $\tilde{X}_{v-2}$  désigne les valeurs que prend la médiane d'un échantillon d'effectif  $n = v - 2$ .

$\tilde{X}_{v-2} = x_{\frac{v+1}{2}} = \tilde{x}$	Probabilité = $\left( \frac{v-1}{2} \right)^2 / \frac{v(v-1)}{2} = \frac{v-1}{2v}$
$\tilde{X}_{v-2} = x_{\frac{v-1}{2}} \text{ ou } x_{\frac{v+3}{2}}$	Probabilité = $\left( 1 - \frac{v-1}{2v} \right) / 2 = \frac{v+1}{4v}$

Si  $W$  était fonction de  $\tilde{x}$  et  $\tilde{X}$ ,  $p$  serait fonction seulement des trois valeurs centrales de  $X$ ; alors que si  $p$  est proportionnelle à  $\alpha$ , elle dépend de toutes les valeurs de  $X$  sauf la médiane.

La contradiction disparaîtrait si l'on convenait que  $W[\tilde{X}, \tilde{x}]$  est fonctionnelle de toute la distribution  $X$  : nulle quand  $X$  prend la valeur  $\tilde{x}$ , positive autrement, croissante avec  $|\tilde{x} - \tilde{X}|$ , soit ici :

$$\frac{v+1}{4v} W \left[ x_{\frac{v-1}{2}}, \tilde{x} \right] + \frac{v+1}{4v} W \left[ x_{\frac{v+3}{2}}, \tilde{x} \right] \equiv \frac{2}{v-1} \frac{\alpha}{v-2}$$

Mais nous ne savons pas partager  $(8v\alpha/(v^2-1)(v-2))$  entre les expressions

$$W \left[ x_{\frac{v-1}{2}}, \tilde{x} \right] \text{ et } W \left[ x_{\frac{v+3}{2}}, \tilde{x} \right].$$

6/ - Quelques indications sur le cas où les modules remplacent les carrés.

a) De même qu'on a substitué  $|x_i - x_j|$  à  $(x_i - x_j)^2$  pour trouver la perte d'information attachée à l'estimation de la médiane de la population par la médiane échantillon, étudions l'expression

$$\delta = |u(x_i - x_j) + v(y_i - y_j)| = |(ux_i + vy_i) - (ux_j + vy_j)|$$

qui remplace  $V(uX + vY)$

On peut l'interpréter de plusieurs façons :

Si  $u$  et  $v$  sont considérés comme des vecteurs de base,  $\delta$  est le module du vecteur de composantes  $(x_i - x_j)$ ,  $(y_i - y_j)$ ;

Si  $u$  et  $v$  sont considérés comme des scalaires et  $(u, v)$  comme les composantes rectangulaires d'un vecteur,

$$\delta / \sqrt{u^2 + v^2}$$

est la projection du vecteur  $(x_i - x_j, y_i - y_j)$  sur la direction  $(u, v)$ .

A ce second point de vue

pour  $u = 1, v = 0$  on retrouve le problème de la médiane des  $x_i$ ;

pour  $u = 0, v = 1$  on a celui de la médiane des  $y_i$ ;

pour  $u^2 + v^2 = 1$ , on a celui de la médiane des projections des points  $(x_i, y_i)$  sur  $(u, v)$ .

#### Remarque.

Il est exceptionnel que le nuage des points  $(x, y)$  ait un point médian  $M_0$ ; ceci suppose le nuage décomposé en deux demi-nuages  $(M_h)$   $(M'_h)$   $M_h$  et  $M'_h$  étant alignés avec  $M_0$  et de part et d'autre de  $M_0$  (quel que soit  $h$ ).

Dans le cas général si on fait tourner la direction  $(u, v)$ , la projection de tout point  $(x_i, y_i)$  a l'occasion d'être une médiane (à tour de rôle).

#### b) Cas du produit vectoriel.

Faisant pendant au produit scalaire  $(\vec{X}, \vec{Y})$  caché dans  $V(uX + vY)$ , on peut reconnaître le produit vectoriel  $(\vec{X} \wedge \vec{Y})$  derrière l'expression

$$\alpha = \mathcal{E}|x_i y_j - x_j y_i|$$

Il est clair que  $\alpha$  nul équivaut à  $\vec{X}$  et  $\vec{Y}$  colinéaire (toutes ses

composantes étant nulles); autrement dit  $(x_1, y_1) \dots (x_i, y_i) \dots$  points alignés avec l'origine.

$$\frac{1}{2} |x_i y_j' - x_j y_i|$$

est l'aire du triangle de sommets  $(0,0)$   $(x_i, y_i)$   $(x_j, y_j)$  ou  $OM_i M_j$ .

Ainsi  $\alpha$  mesure en quelque sorte l'éventail entre les divers rayons vecteurs  $OM_i$  autour d'un rayon vecteur central.

c) Sur le cas  $|x_i - x_j|^3 \equiv g(i, j)$ .

Pour simplifier supposons que la variable  $x_i$  prenne quatre valeurs

$$a > b > c > d$$

On aura donc

$$\begin{aligned} \delta &= \frac{2}{4^2} [(a-b)^3 + (a-c)^3 + (a-d)^3 + (b-c)^3 + (b-d)^3 + (c-d)^3] \\ &= \frac{1}{8} [3a^3 + b^3 - c^3 - 3d^3 - 3ab(a-b) - 3ac(a-c) - 3ad(a-d) \\ &\quad - 3bc(b-c) - 3bd(b-d) - 3cd(c-d)] \end{aligned}$$

Le terme  $(3a^3 + b^3 - c^3 - 3d^3)$  est analogue au terme  $(3a + b - c - 3d)$  qui apparaît quand on traite le cas de la première puissance, c'est

$$\sum_n h(x_n^3 - x_n^3)$$

Les autres termes ont une structure comparable, on reconnaît :

$$\sum x_i x_j |x_i - x_j|$$

Au total  $\delta$  est la somme des deux termes

$$\frac{2}{\sqrt{2}} \sum |x_i - x_j|^3 \equiv \frac{2}{\sqrt{2}} \sum |x_i^3 - x_j^3| - \frac{6}{\sqrt{2}} \sum |x_i^2 x_j - x_j^2 x_i|$$

Il n'a pas paru qu'en général cette expression pouvait représenter une perte d'information attachée à l'estimation (biaisée) d'un paramètre. Mais il ne paraît pas absurde d'y voir une perte d'information relative à l'estimation simultanée de deux caractéristiques (médiane et rayon vecteur central).



## TABLE DES MATIÈRES

---

Fasc. 1 — Contribution à la Théorie des Valeurs extrêmes (Deuxième partie). J. GEFFROY .....	3
— Le problème du mouvement brownien et ses généralisations. G. MALECOT .....	67
Fasc. 2 — L'Information en Statistique Mathématique et dans la théorie des communications. D.-D. JOSHI .....	83
— Sur l'inégalité de SLUTSKY. A. FUCHS .....	161
— Analyse d'ouvrages .....	165
Fasc. 3 — Les modèles stochastiques en Génétique de population. G. MALECOT	173
— Conception stochastique de coefficients multiplicateurs dans l'ajustement linéaire des séries temporelles. H. THEIL et L.-B. MENNES	211
— Fonctions de répartition à N dimensions et leurs marges. M. SKLAR	229
— Un modèle économétrique : La Projection à long terme de Verdorn.	233
Fasc. 4 — Sur quelques points de la théorie des tests. G. MORLAT .....	253
— La Perte d'information par sondage (calcul des probabilités). P. THIONET .....	267

