



Sur l'extension de certaines évaluations statistiques au cas de petits échantillons

Author(s): Maurice Frechet

Source: *Revue de l'Institut International de Statistique / Review of the International Statistical Institute*, Vol. 11, No. 3/4 (1943), pp. 182-205

Published by: International Statistical Institute (ISI)

Stable URL: <http://www.jstor.org/stable/1401114>

Accessed: 21/12/2009 05:40

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/action/showPublisher?publisherCode=isi>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



International Statistical Institute (ISI) is collaborating with JSTOR to digitize, preserve and extend access to *Revue de l'Institut International de Statistique / Review of the International Statistical Institute*.

<http://www.jstor.org>

SUR L'EXTENSION DE CERTAINES EVALUATIONS STATISTIQUES AU CAS DE PETITS ECHANTILLONS

par Maurice Fréchet.

Introduction.

Ce mémoire ¹⁾ est consacré à l'extension au cas de petits échantillons de la méthode de détermination empirique d'un paramètre basée sur le principe de la moindre dispersion et à sa comparaison avec les méthodes basées sur le principe de la valeur dominante et sur celui de la plus grande plausibilité.

Si nous nous en étions tenus aux démonstrations, nous aurions pu abrégér sensiblement ce mémoire. Mais il nous a paru nécessaire d'entrer dans plus de détails qu'on ne le fait généralement, afin de séparer plus nettement des déductions mathématiques, les hypothèses et les conventions sur lesquelles elles reposent et dont le choix, aussi plausible que possible, n'a cependant rien de nécessaire.

Notations. — Soient X_1, X_2, \dots, X_n , n valeurs prises par une variable aléatoire X au cours de n épreuves indépendantes. On se limitera, dans la suite, au cas où la loi de répartition de X peut s'exprimer par une probabilité élémentaire δdx et où, de plus, la densité de probabilité δ en un point x est une fonction d'une forme connue $f(x, \theta)$, dépendant d'un paramètre dont la valeur vraie θ_0 est inconnue.

On se propose d'évaluer θ_0 connaissant d'une part, la forme $f(x, \theta)$ de δ et, d'autre part, les résultats de n épreuves qui ont donné les valeurs numériques x_1, x_2, \dots, x_n à X_1, X_2, \dots, X_n . Sous cette forme stricte, le problème *ne peut être résolu* par de simples déductions mathématiques.

Il s'agit donc de fixer *certaines conventions plausibles* qui assigneront à tout „échantillon” de n valeurs x_1, x_2, \dots, x_n de X une valeur déterminée t , laquelle sera prise comme valeur empirique de la valeur vraie θ_0 . t est donc une certaine fonction convenablement choisie de x_1, x_2, \dots, x_n

$$(1) \quad t = H_n(x_1, \dots, x_n).$$

On voit que chaque échantillon détermine t , de sorte que

$$(2) \quad T_n = H_n(X_1, \dots, X_n)$$

est une variable aléatoire dont chaque échantillon détermine une valeur. (Quand, dans nos raisonnements, n sera fixe, nous écrirons, pour simplifier, H au lieu de H_n et T au lieu de T_n).

¹⁾ Le contenu de ce mémoire a formé une partie de notre cours de statistique mathématique à l'Institut Henri Poincaré pendant l'hiver 1939—1940. Il constitue l'un des chapitres du deuxième cahier (en préparation) de nos „Leçons de Statistique Mathématique”, dont le premier cahier „Introduction: Exposé préliminaire de Calcul des Probabilités” (119 pages in-quarto, dactylographiées) vient de paraître au „Centre de Documentation Universitaire”, Tournais et Constans, Paris.

Choix de la fonction $H(X_1, \dots, X_n)$ qui déterminera la valeur empirique de θ_0 . — Il est important d'analyser les raisons qui peuvent guider dans le choix de la fonction H pour en faire éclater le caractère inductif, pour faire voir que s'il est possible de le rendre plausible, il ne peut être question de le considérer comme nécessaire.

Nous ferons cette analyse pour trois des méthodes connues.

PREMIÈRE PARTIE.

Méthode de la plus petite dispersion.

Première condition imposée à la fonction H . On espère que la valeur $t = H(x_1 \dots x_n)$ déduite de l'échantillon $x_1 \dots x_n$, sera voisine de θ_0 .

Afin de choisir H de façon à réaliser ce désir, on pourra convenir d'abord de choisir H de façon que θ_0 soit une „valeur typique” ¹⁾ de la variable aléatoire $T = H(X_1, \dots, X_n)$. Par exemple, θ_0 sera une valeur probable de T , ou bien une dominante. etc. Dans ce qui suit, nous choisirons comme valeur typique de T , sa valeur moyenne; nous imposerons donc à la fonction H la condition

$$\mathfrak{M} H(X_1, \dots, X_n) = \theta_0,$$

et pour rappeler que c'est la valeur moyenne calculée en supposant que dans $f(x, \theta)$, on doit prendre $\theta = \theta_0$, nous écrivons

$$(3) \quad \mathfrak{M}_{\theta_0} H(X_1, \dots, X_n) = \theta_0.$$

On voit qu'ici nous nous inspirons de l'idée simple suivante. Supposant (3) vérifiée, nous avons à calculer empiriquement une valeur typique θ_0 de T , connaissant *une seule* valeur de T , soit $H(x_1, \dots, x_n)$. La seule ressource dans ce cas, c'est de prendre comme valeur approchée de θ_0 , la seule valeur connue, t , de T . On voit combien nous sommes peu assurés que cette valeur soit réellement approchée.

Seconde condition. Cependant, la condition (3) ne détermine pas H , elle délimite seulement une classe de fonctions que nous pourrions appeler „admissibles”. Nous pouvons profiter de cette indétermination pour améliorer le choix de H . En effet, il est clair que la valeur approchée t de θ_0 risquera d'autant moins d'être éloignée de θ_0 , que la dispersion de T sera plus petite. Par conséquent, on choisira, parmi les fonctions H admissibles, celle ou celles dont la dispersion est la plus faible. Ici encore nous pouvons traduire cette condition de façons diverses en mesurant la dispersion de T par son écart moyen, son écart probable, etc.

Voici donc le principe d'une méthode de choix de la fonction H : on restreint le choix de H en le limitant aux fonctions telles que θ_0 soit une valeur typique de $T = H(X_1, \dots, X_n)$ et parmi celles-là, on prend celle dont la dispersion est la plus petite.

Non seulement, nous ne sommes pas assurés que pour la fonction H choisie, la valeur empirique $t = H(x_1, \dots, x_n)$ sera voisine de θ_0 , mais même, ce choix n'est pas unique, puisqu'il serait modifié selon ce que l'on conviendrait de choisir pour valeur typique ou pour dispersion.

¹⁾ La terminologie et les notations de ce mémoire sont celles du premier cahier de nos „Leçons de Statistique Mathématique” signalées en note, p. 182.

Troisième condition. Mais une autre difficulté se présente, même quand on fait un choix déterminé de la valeur typique et de la dispersion, quand par exemple, on choisit, comme nous allons le faire, la valeur moyenne et l'écart quadratique moyen.

Il devrait suffire qu'on ait assujetti la fonction cherchée H'' aux deux conditions

$$(4) \quad \theta_0 = \mathfrak{N}_{\theta_0} H''$$

et

$$(5) \quad \mathfrak{N} (H'' - \theta_0)^2 \leq \mathfrak{N}_{\theta_0} (H - \theta_0)^2$$

pour toutes les fonctions H telles que

$$(6) \quad \theta_0 = \mathfrak{N}_{\theta_0} H.$$

Or, parmi les fonctions H vérifiant (4) figure la fonction réduite à une constante égale à θ_0 et pour cette valeur particulière $\mathfrak{N}_{\theta_0} (H - \theta_0)^2$ est nulle. Donc, d'après (5), $\mathfrak{N} (H'' - \theta_0)^2$ doit être nulle: on sait qu'alors $H'' (X_1, \dots, X_n)$ doit être presque certainement égale aussi à θ_0 . La fonction cherchée H'' qui serait choisie pour déterminer une valeur empirique de θ_0 , ne serait donc connue que si, *préalablement*, on connaissait θ_0 , c'est-à-dire, précisément ce qu'on cherche.

Devant cette difficulté, on va assujettir la fonction cherchée à une nouvelle condition qui s'impose moins naturellement que les précédentes. Tout en conservant les conditions précédentes, nous allons supposer qu'au lieu de les imposer seulement en prenant pour θ_0 la valeur vraie cherchée, on les suppose vérifiées en remplaçant θ_0 par n'importe quelle valeur θ . Alors, d'une part, elles auront bien lieu comme il le fallait pour la valeur particulière θ_0 de θ qui nous intéresse; et d'autre part, il ne sera plus nécessaire de connaître précisément θ pour les imposer.

C'est là une méthode qui a été appliquée depuis longtemps pour les grands échantillons, et que nous allons mettre en oeuvre pour les petits échantillons ou plus précisément, pour une valeur arbitraire de n . Les premiers des résultats qui suivent ont été déjà obtenus par Dooob. Nous les avons retrouvés (et enseignés en 1939) sans avoir remarqué à ce moment l'antériorité de cet auteur.

Position du problème. Il s'agit de déterminer une fonction $H' (X_1, \dots, X_n)$ telle que l'on ait, quel que soit θ :

$$(7) \quad \theta = \mathfrak{N}_{\theta} H' (X_1, \dots, X_n)$$

et

$$(8) \quad \mathfrak{N}_{\theta} [H' (X_1, \dots, X_n) - \theta]^2 \leq \mathfrak{N}_{\theta} [H (X_1, \dots, X_n) - \theta]^2$$

pour toutes les fonctions H telles que

$$(9) \quad \theta = \mathfrak{N}_{\theta} H (X_1, \dots, X_n).$$

Remarque. Supposons qu'on admette parmi les fonctions H des fonctions dépendant de θ ; il ne paraît pas a priori impossible, en effet, de déterminer la valeur empirique t de θ_0 , au moyen de l'équation implicite

$$t = H' (x_1, x_2, \dots, x_n, t)$$

Mais dans ce cas, la condition (9) est évidemment vérifiée en prenant

$$H (x_1, \dots, x_n, \theta) \equiv \theta$$

et alors d'après (8), $\varpi_{\theta} [H'(X_1, \dots, X_n, \theta) - \theta]^2$ devrait être nul et par suite $H'(X_1, \dots, X_n, \theta)$ serait presque „certainement” égal à θ . Alors l'équation (10) se réduirait à $t = t$ et ne permettrait pas de déterminer t .

Ainsi, dans ce qui précède, les fonctions H et H' doivent être supposées essentiellement *indépendantes de θ* , limitation qui a comme nous le verrons (p. 194) des conséquences très importantes.

BORNE INFÉRIEURE DE LA DISPERSION DE LA VALEUR EMPIRIQUE DU PARAMÈTRE.

Dans la suite, on devra rendre σ_T minimum, il est donc intéressant d'en chercher une borne inférieure et de montrer que cette borne inférieure (nécessairement ≥ 0) n'est pas, en général, nulle. De plus, si l'on veut se limiter aux grands échantillons, ce qui est souvent utile, il est bon d'avoir une expression asymptotique (en fonction de n) de σ_{T_n} ou plutôt de sa borne inférieure.

La résolution de ces différents problèmes s'opère au moyen de la formule

$$(10) \quad \sigma_T \geq \frac{1}{\sigma_A \sqrt{n}}$$

et de la formule

$$(11) \quad \left(\text{plus petite limite de } \sqrt{n} \sigma_{T_n} \right) \underset{n \rightarrow \infty}{\geq} \frac{1}{\sigma_A}$$

où A est la variable aléatoire

$$(12) \quad A = \frac{1}{f(X, \theta)} \frac{\partial f(X, \theta)}{\partial \theta}.$$

Ces deux formules — ou des formes équivalentes de ces deux formules — sont connues depuis un mémoire publié en 1898 par Karl Pearson et Filon. Mais il faut observer que les premiers auteurs qui ont écrit la première, (10), l'ont démontrée seulement asymptotiquement, c'est-à-dire l'ont écrite en tant qu'écriture abrégée, valable seulement et alors approximativement, pour les grandes valeurs de n . C'est dire qu'ils n'ont réellement démontré que la seconde formule, soit (11). C'est ce qu'ont fait aussi Edgeworth ¹⁾ puis R. A. Fisher ²⁾. C'est Doob ³⁾ qui semble avoir obtenu le premier la formule non asymptotique (10) et sans l'hypothèse de normalité faite par ses prédécesseurs.

L'objet de ce paragraphe consiste à montrer que sous des hypothèses beaucoup plus générales que celles des premiers auteurs, on peut établir la formule asymptotique (11). Et que si l'on rend plus strictes nos hypothèses, on peut même prouver la formule $\sigma_T \geq \frac{1}{\sigma_A \sqrt{n}}$ pour toute valeur de n .

On verra, en outre, qu'on peut mettre les démonstrations des premiers auteurs sous une forme plus familière aux étudiants de Calcul des Probabilités. Ce faisant, la démonstration, — essentiellement différente des démonstrations antérieures, *qui cessent d'être valables* dans le cas plus général où nous nous plaçons, — n'en devient pas pour cela plus compliquée, tout au contraire, et fournit non

¹⁾ Journal of the Royal Statistical Society, 1908.

²⁾ On the mathematical foundations of theoretical statistics (Philosophical Transactions, 1922, A CCXXII, pp. 309—368).

³⁾ Statistical Estimation (Transactions American Math. Soc., t. 39, 1936, p. 410-421)

seulement (11) mais (10). Elle semble avoir été donnée pour la première fois par Doob.

Sous les hypothèses faites dans les publications antérieures à Doob, la variable aléatoire T_n était supposée devenir „normale” autour de θ pour les grandes valeurs de n . (Rigoureusement parlant, cela voulait dire que la densité de probabilité au point y , de la variable réduite: $Y = \frac{T_n - \theta}{\sigma_{T_n}}$ convergeait vers: $\frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}$).

Au lieu de supposer ainsi connue la distribution réduite limite de T_n et d'imposer de cette façon, au choix de la valeur empirique, une condition *que rien ne suggère à l'intuition*, nous nous contenterons d'imposer une condition qui est une conséquence particulière de cette hypothèse de distribution. Il résulte en effet de celle-ci que la valeur moyenne de T_n tendait vers θ avec $\frac{1}{n}$.

Mais il n'est nullement nécessaire de passer par l'hypothèse trop stricte de la „normalité”. Il est en effet très naturel d'imposer directement au choix de la fonction H la condition, comme on l'a vu plus haut, que θ soit égal à une valeur typique de H et, en particulier, comme nous le ferons ici, à sa valeur moyenne.

Il est vrai que, d'après les hypothèses des auteurs précédents. $\mathfrak{N}_\theta T_n$ était supposé seulement *tendre* vers θ quand $n \rightarrow \infty$. Mais alors leurs résultats n'étaient utilisables que pour de grandes valeurs de n . Si nous voulons obtenir un résultat valable même pour de petits échantillons, il y a donc lieu de supposer que $\mathfrak{N}_\theta T$ est précisément égal à θ .

Ce sera notre premier cas. Nous examinerons ensuite comme second cas, l'hypothèse où $\mathfrak{N}_\theta T_n$ tend vers θ quand n croît.

Premier cas. Nous supposons que $\mathfrak{N}_\theta T = \theta$, c'est-à-dire que

$$(16) \quad \theta = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} H(x_1, \dots, x_n) \prod_{j=1}^n f(x_j, \theta) dx_j^{(1)}$$

Les fonctions H et f sont donc soumises à la condition (16) et à la condition

$$(17) \quad \int_{-\infty}^{+\infty} f(x, \theta) dx = 1 \quad \text{ou}$$

$$(18) \quad \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} \prod [f(x_j, \theta) dx_j] = 1$$

La condition (18) est inévitable. La condition (16) est seulement une condition plausible, mais moins stricte d'ailleurs que celle de la „normalité”. Enfin, nous supposerons (uniquement pour la commodité des calculs), que les fonctions H et f sont assez continues et régulières et que leur comportement à l'infini est assez rapide pour qu'on ait le droit d'appliquer dans ce qui va suivre les règles de dérivation d'une intégrale par rapport à un paramètre.

Question d'existence. Il est important pour la première partie et aussi pour la seconde de prouver qu'il est possible de satisfaire à toutes ces conditions.

Par exemple, supposons que $f(x, \theta)$ soit de la forme $g(x - \theta)$ avec:

$$(19) \quad \int_{-\infty}^{+\infty} g(t) dt = 1; \quad \int_{-\infty}^{+\infty} t g(t) dt = 0$$

¹⁾ En posant, en général, comme d'habitude $\prod_{j=1}^n u_j = u_1 u_2 \dots u_j \dots u_n$.

Posons de plus $\sigma^2 = \int_{-\infty}^{+\infty} t^2 g(t) dt$. (Telle est, entre autres, la fonction :

$$g_\sigma(t) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{t^2}{2\sigma^2}}.$$

Alors (17) est vérifié. En prenant $H(x_1, \dots, x_n) = \frac{x_1 + \dots + x_n}{n}$ on peut s'assurer que l'équation (16) sera aussi satisfaite.

On verrait qu'il en est de même si l'on prenait par exemple,

$$(20) \quad H(x_1, x_2, \dots, x_n) = \frac{\sum x_i}{n} + \frac{\sum (x_i - x_k)^2}{n-1} - n \sigma^2.$$

Plus généralement, soit $h(x)$ une fonction à dérivée partout de même signe et qui varie de $-\infty$ à $+\infty$, quand x croît de $-\infty$ à $+\infty$, et supposons que

$$f(x, \theta) \equiv g(h(x) - \theta) h'_x,$$

$g(t)$ vérifiant encore les conditions (19). On pourrait alors prendre par exemple

$$H(x_1, \dots, x_n) = \frac{\sum h(x_i)}{n}; \text{ ou}$$

$$H(x_1, \dots, x_n) = \frac{\sum h(x_i)}{n} + \frac{\sum [h(x_i) - h(x_k)]^2}{n-1} - n \sigma^2.$$

Démonstration de l'inégalité (10). La démonstration de (10) est très courte et se réduit à l'utilisation des hypothèses faites par la dérivation en θ des conditions qui les expriment et à l'application aux formules obtenues de l'inégalité généralisée de Schwarz (cette dernière qui s'écrit

$$(29) \quad [\mathfrak{N}(Z T)]^2 \leq \mathfrak{N} Z^2 \mathfrak{N} T^2$$

est valable, comme on sait, pour tout couple Z, T de variables aléatoires définies sur la même catégorie d'épreuves).

En dérivant (18) par rapport à θ ; on a

$$(30) \quad \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} [\sum A(x_i)] \Pi [f(x_i, \theta) dx_i] = 0$$

où $A(x_i)$ est une valeur particulière de la variable aléatoire

$$A = A(X) = \frac{1}{f(X, \theta)} \frac{\partial f(X, \theta)}{\partial \theta}.$$

En posant $U = \sum A(X_i)$, on peut donc écrire, d'après (30)

$$(31) \quad \mathfrak{N}_\theta U = 0.$$

D'autre part, en dérivant (16) après l'avoir mis sous la forme équivalente

$$(32) \quad \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} [H(x_1, \dots, x_n) - \theta] \Pi [f(x_i, \theta) dx_i] = 0,$$

on obtient

$$\int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} [H(x_1, \dots, x_n) - \theta] [\sum A(x_i)] \Pi [f(x_i, \theta) dx_i] - 1 = 0$$

ou

$$\mathfrak{N} [(T - \theta)U] = 1$$

et puisque

$\mathfrak{N} T = \theta$, $\mathfrak{N} U = 0$, on aura

$$(33) \quad \mathfrak{N} [(T - \mathfrak{N} T) (U - \mathfrak{N} U)] = 1,$$

d'où, par l'inégalité généralisée de Schwarz (29):

$$(34) \quad 1 \leq [\mathfrak{N} (T - \mathfrak{N} T)^2] [\mathfrak{N} (U - \mathfrak{N} U)^2] = (\sigma_T \sigma_U)^2$$

On peut d'ailleurs simplifier en observant que U est la somme de variables indépendantes; de sorte qu'en vertu de l'égalité de Bienaymé, on a:

$$(\sigma_U)^2 = \Sigma [\sigma_{A(X_i)}]^2 = n(\sigma_A)^2$$

D'où, finalement, l'inégalité annoncée

(10)

$$(\sigma_T)^2 \geq \frac{1}{n(\sigma_A)^2} \text{ avec } T = H(X_1, \dots, X_n), \quad A = \frac{1}{f(X, \theta)} \frac{\partial f(X, \theta)}{\partial \theta}$$

inégalité remarquable où le second membre ¹⁾ est indépendant du choix de la fonction H définissant la „valeur empirique” T et où dans le premier membre, on peut prendre pour T toute valeur empirique $T = H(X_1, \dots, X_n)$ assujettie à la seule condition (16), c'est-à-dire: $\mathfrak{N}_\theta T = \theta$ quel que soit θ .

Second cas. Si l'on est disposé à limiter l'utilisation de la formule obtenue aux grandes valeurs de n , on peut considérer un cas plus général que le précédent.

Au lieu d'assujettir le choix de la fonction $T = H$ à la condition $\mathfrak{N}_\theta T = \theta$, contentons-nous de supposer que la quantité $\varepsilon(\theta, n) = \mathfrak{N}_\theta T_n - \theta$ converge vers 0 avec $\frac{1}{n}$ ou plutôt, pour faciliter la démonstration, supposons que $\varepsilon(\theta, n)$ est

dérivable par rapport à θ et que $\omega(\theta, n) = \frac{\partial \varepsilon(\theta, n)}{\partial \theta}$ converge vers zéro quand $n \rightarrow \infty$. Dans cette hypothèse, plus générale que la précédente (et qui abandonne encore une fois l'hypothèse de la „normalité” à la limite de la loi de distribution de $T = H(X_1, \dots, X_n)$), la formule (10) n'est plus, en général, exacte, mais elle l'est à la limite, sous la forme (11). La démonstration est aussi courte que la précédente dont elle suit la même ligne générale.

Démonstration. On a encore la condition (18) dont on déduit (31) par dérivation. La condition (16) est remplacée par $\mathfrak{N}_\theta T_n - \theta = \varepsilon(\theta, n)$, d'où l'on déduit, comme plus haut, par dérivation

$$\mathfrak{N} [(T_n - \mathfrak{N} T_n) (U - \mathfrak{N} U)] = 1 + \frac{\partial \varepsilon}{\partial \theta} (\theta, n)$$

D'après l'inégalité généralisée de Schwarz et l'égalité de Bienaymé, on en déduit, comme plus haut

$$n(\sigma_{T_n})^2 (\sigma_A)^2 \geq \left[1 + \frac{\partial \varepsilon}{\partial \theta} (\theta, n) \right]^2 \quad \sigma_{T_n} \sqrt{n} \geq \frac{1}{\sigma_A} \left| 1 + \frac{\partial \varepsilon}{\partial \theta} \varepsilon(\theta, n) \right|.$$

¹⁾ Observons qu'en vertu de $\mathfrak{N} A = 0$, on peut écrire:

$$(\sigma_A)^2 = \int_{-\infty}^{+\infty} \left[\frac{1}{f} \frac{\partial f}{\partial \theta} \right]^2 f dx, \quad \text{ou: } (\sigma_A)^2 = \int_{-\infty}^{+\infty} \frac{1}{f(x, \theta)} \left[\frac{\partial f(x, \theta)}{\partial \theta} \right]^2 dx.$$

D'où :

$$(11) \quad [\text{plus petite limite de } (\sigma_{T_n}) \sqrt{n}] \geq \frac{1}{\sigma_A}.$$

DETERMINATION DE LA DENSITE MINIMISANTE.

De même que l'inégalité généralisée de Schwarz (29) nous a permis d'obtenir simplement l'inégalité connue (11) et l'inégalité assez récente (10), de même la condition classique ¹⁾ pour que l'inégalité généralisée de Schwarz devienne une égalité va nous permettre de déterminer dans quel cas σ_T atteint sa borne inférieure $\frac{1}{\sigma_A \sqrt{n}}$.

Premier cas. Nous traiterons seulement le cas de n quelconque et, par suite, nous supposons la valeur empirique T choisie de sorte que $\mathfrak{N}_\theta T = \theta$ quel que soit θ . On a vu qu'alors, on a :

$$(10) \quad \sigma_T \geq \frac{1}{\sigma_A \sqrt{n}}.$$

Ici σ_T et σ_A dépendent en général de θ . Si, pour une valeur déterminée de θ , et une fonction H' particulière parmi les fonctions H admissibles, les deux membres de (10) sont égaux, c'est que le second membre est non seulement *une* borne inférieure du premier, mais même est son minimum et que ce minimum est atteint pour $H \equiv H'$.

Or, on sait ¹⁾ que l'inégalité généralisée de Schwarz (29) ne se réduit à une égalité que dans le cas où il existe deux nombres α, β non tous deux nuls tels que $\alpha Z + \beta T$ soit „presque certainement” égal à zéro. L'inégalité (10) ou (34) ne devient donc une égalité que si α, β existent (non aléatoires et non tous deux nuls) tels que $\alpha(H' - \theta) + \beta U = 0$ (sauf en un cas de probabilité nulle que nous négligerons dans la suite). D'ailleurs, en supposant évidemment que f dépend effectivement de θ , U n'est pas $= 0$, donc $\alpha \neq 0$ et par suite

$$(40) \quad H' = \theta + \lambda' U, \text{ où } \lambda' \text{ est un nombre certain.}$$

D'ailleurs, en vertu de (40) et (33), on a

$$1 = \lambda' \mathfrak{N}_\theta U^2 = \lambda' n \mathfrak{N}_\theta A^2, \text{ d'où } \lambda' = \frac{1}{n \mathfrak{N}_\theta A^2} \text{ et par suite}$$

$$H' = \theta + \frac{1}{n \mathfrak{N}_\theta A^2} \Sigma \left[\frac{\frac{\partial f(X_i, \theta)}{\partial \theta}}{f(X_i, \theta)} \right], \text{ ou}$$

$$(41) \quad H' = \theta + \frac{\Sigma_i \frac{\partial}{\partial \theta} [L f(X_i, \theta)]}{n \int_{-\infty}^{+\infty} \left[\frac{\partial f(x, \theta)}{\partial \theta} \right]^2 \frac{dx}{f(x, \theta)}}. \quad 2)$$

D'ailleurs, il est clair que, d'après (40) et (31) $\mathfrak{N}_\theta H' = \theta + \lambda' \mathfrak{N}_\theta U = \theta$. H' serait donc bien une des fonctions admissibles si H' était indépendant de θ , condition dont nous avons constaté la nécessité page 184.

¹⁾ Voir, par exemple, p. 20 du premier cahier de nos „Leçons de Statistique Mathématique”.

²⁾ $L f$ désigne ici et dans la suite le logarithme népérien de f , $\log_e f$.

Or, c'est ce qui n'a pas lieu en général. Observons, en effet que, d'après (41)

$$(42) \quad H' = \frac{1}{n} \sum h(X_i, \theta) \text{ avec}$$

$$(43) \quad h(X, \theta) = \theta + \lambda(\theta) \frac{\partial}{\partial \theta} L f(X, \theta), \text{ où}$$

$$(44) \quad \lambda(\theta) = \frac{1}{\int_{-\infty}^{+\infty} \left[\frac{\partial f(x, \theta)}{\partial \theta} \right]^2 \frac{dx}{f(x, \theta)}}.$$

Pour que H' soit indépendant de θ , il faut et il suffit que $h(X, \theta)$ soit indépendant de θ . Or, on trouve facilement des exemples où il n'en est pas ainsi. Reprenons, pour le montrer, l'exemple de la page 186. Si $f(x, \theta) = g(x - \theta)$, on a

$$\frac{1}{\lambda(\theta)} = \int_{-\infty}^{+\infty} [g'(x - \theta)]^2 \frac{1}{g(x - \theta)} dx = \int_{-\infty}^{+\infty} [g'(t)]^2 \frac{1}{g(t)} dt = \frac{1}{\lambda}$$

quantité indépendante de θ . D'où l'interprétation de (43):

$$h(x, \theta) = \theta + \lambda \frac{\partial}{\partial \theta} L g(x - \theta)$$

Si l'on suppose $h(x, \theta)$ indépendant de θ et égal à $h(x)$, on aura, en intégrant en θ

$$k(x) - \frac{[h(x) - \theta]^2}{2\lambda} = L g(x - \theta)$$

où $k(x)$ est une fonction arbitraire de x . Ainsi $L g(x - \theta)$ est une fonction du second degré en θ . En remplaçant x, θ , par $-\theta, -x$, $g(x - \theta)$ ne change pas, donc $L g(x - \theta)$ est aussi du second degré en x et le terme de plus haut degré est $-\frac{x^2}{2\lambda}$.

Donc $L g(x - \theta)$ est de la forme: $-\frac{(x - \theta)^2}{2\lambda} + \alpha \theta x + \beta \theta + \gamma x + \delta$

et puisque $L g(x - \theta)$ est une constante pour $x = \theta$, on a

$$\alpha = 0, \beta + \gamma = 0. \text{ D'où}$$

$$L g(x - \theta) = -\frac{(x - \theta)^2}{2\lambda} + \gamma(x - \theta) + \delta$$

avec $\lambda > 0$, ou

$$L g(x - \theta) = -\frac{(x - \theta - \lambda \gamma)^2}{2\lambda} + \delta'$$

$$f(x, \theta) = g(x - \theta) = \rho e^{-\frac{(x - \theta - \lambda \gamma)^2}{2\lambda}};$$

$$g(t) = \rho e^{-\frac{(t - \lambda \gamma)^2}{2\lambda}}.$$

Or, d'après (19), $g(t)$ est la densité de probabilité d'une variable aléatoire ayant pour moyenne 0 et pour écart quadratique moyen σ . Donc $\rho = \frac{1}{\sqrt{2\lambda\pi}}$, $\lambda \gamma = 0$, $\lambda = \sigma^2$, d'où enfin

$$(45) \quad f(x, \theta) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\theta)^2}{2\sigma^2}}.$$

On a vu (p. 187) que si $f(x, \theta) = g(x - \theta)$ où $g(t)$ vérifie les conditions (19), il y a une infinité de fonctions H indépendantes de θ , telles que $\mathfrak{N} H(X_1, \dots, X_n) = \theta$. Mais, parmi toutes les variables aléatoires X correspondantes, il n'y a que celles qui sont des variables „laplaciennes” pour chacune desquelles un choix d'une fonction $H(x_1, \dots, x_n)$ indépendante de θ (et vérifiant encore $\mathfrak{N}_\theta H = \theta$) peut donner un écart σ_T atteignant la borne inférieure

$$\frac{1}{\sigma_A \sqrt{n}} \quad (\text{et ceci quel que soit } \theta).$$

Etude des densités distinguées. Appelons (provisoirement, dans ce mémoire) *densité distinguée*, toute densité de probabilité $f(x, \theta)$ telle que la fonction

$$(46) \quad \theta + \frac{\frac{\partial L f(x, \theta)}{\partial \theta}}{\int_{-\infty}^{+\infty} \left[\frac{\partial}{\partial \theta} f(x, \theta) \right]^2 \frac{dx}{f(x, \theta)}}$$

soit indépendante de θ .

Pour ces densités distinguées, on va pouvoir déterminer la fonction minimisante $H'(X_1, \dots, X_n)$ et étendre au cas des petits échantillons la comparaison des méthodes d'estimation faites par divers auteurs dans le cas des grands échantillons. Il vaut donc la peine de chercher la forme générale de $f(x, \theta)$ pour cette catégorie de variables.

Mais auparavant constatons que dans le cas d'une densité distinguée et dans ce cas seulement σ_{T_n} peut atteindre sa borne $\frac{1}{\sigma_A \sqrt{n}}$ pour au moins un choix parmi les fonctions H telles que $\mathfrak{N}_\theta H = \theta$ et qu'alors un tel choix ne peut fournir qu'une seule fonction (indépendante de θ) à savoir la fonction H donnée par (41). Pour une telle variable, la fonction (46) ne dépend, par hypothèse, qu'en apparence de θ . En appelant $h(x)$ cette fonction, on voit qu'on a l'identité de la forme

$$(47) \quad \lambda(\theta) \frac{\partial}{\partial \theta} L f(x, \theta) = h(x) - \theta$$

où $\lambda(\theta) > 0$. On peut considérer $\frac{1}{\lambda(\theta)}$ comme la dérivée seconde d'une fonction $\mu(\theta)$; d'où $\frac{\partial}{\partial \theta} L f(x, \theta) = \mu_\theta''(\theta) [h(x) - \theta]$.

Par suite $L f(x, \theta) - \mu'_\theta [h(x) - \theta] - \mu(\theta)$ est une quantité indépendante de θ que nous pouvons représenter par $l(x)$.

Ainsi toute densité distinguée, $f(x, \theta)$, est de la forme

$$(48) \quad f(x, \theta) = e^{\mu'_\theta [h(x) - \theta] + \mu(\theta) + l(x)}$$

D'ailleurs $\mu(\theta)$, $h(x)$ et $l(x)$ ne peuvent y être pris arbitrairement, car on doit avoir :

$$(49) \quad \int_{-\infty}^{+\infty} f(x, \theta) dx = 1.$$

Ces deux conditions nécessaires (48) et (49) sont suffisantes. En effet, soient,

réciroquement, trois fonctions $\mu(\theta)$, $h(x)$, $l(x)$ telles qu'on ait quel que soit θ

$$(50) \quad \int_{-\infty}^{+\infty} e^{\mu' \theta [h(x) - \theta] + \mu(\theta) + l(x)} dx = 1.$$

Alors, la fonction (46) est distinguée. En effet, si $\lambda(\theta)$ est une fonction de θ quelconque, l'expression $\theta + \lambda(\theta) \frac{\partial}{\partial \theta} L f(x, \theta)$ se réduit, d'après (48), à

$$(51) \quad \theta + \lambda(\theta) \mu''_{\theta} [h(x) - \theta]$$

et il suffit de prouver que $\lambda(\theta) \mu''_{\theta}$ se réduit à 1 quand

$$(52) \quad \frac{1}{\lambda(\theta)} = \int_{-\infty}^{+\infty} \left[\frac{\partial}{\partial \theta} L f(x, \theta) \right]^2 f(x, \theta) dx = (\sigma_A)^2$$

pour montrer que l'expression (46) se réduit à une fonction $h(x)$ indépendante de θ . Le second membre de (52) est ici :

$$(53) \quad \frac{1}{\lambda(\theta)} = \int_{-\infty}^{+\infty} \mu''^2 (h - \theta)^2 e^{\mu'(h - \theta) + \mu + l} dx.$$

Or, (50) étant vérifiée quel que soit θ , dérivons-là en θ , on aura :

$$\int_{-\infty}^{+\infty} e^{\mu' \theta (h - \theta) + \mu + l} \mu''_{\theta} (h - \theta) dx = 0$$

On peut diviser par μ''_{θ} qui est indépendant de x . Si l'on dérive alors encore en θ , on aura :

$$(54) \quad \int_{-\infty}^{+\infty} e^{\mu' \theta (h - \theta) + \mu + l} \mu'' (h - \theta)^2 dx = \int_{-\infty}^{+\infty} e^{\mu' \theta (h - \theta) + \mu + l} dx = 1$$

d'où en combinant (53) et (54)

$$(52\text{bis}) \quad \lambda \mu'' = 1.$$

Incidentement, puisque, d'après (52), $\lambda(\theta)$ est positif, il en résulte aussi que $\mu'' \left(= \frac{1}{\lambda(\theta)} \right)$ est aussi positif.

On peut d'ailleurs préciser d'une manière plus directe que par (50), le choix des fonctions $\mu(\theta)$, $h(x)$, $l(x)$: on peut prendre arbitrairement $h(x)$ et $l(x)$ ¹⁾ et alors $\mu(\theta)$ est déterminé par (50) ou même mieux par une formule explicite. En effet, (50) peut s'écrire

$$e^{\theta \mu' - \mu} = \int_{-\infty}^{+\infty} e^{\mu' \theta h(x) + l(x)} dx.$$

Donnons-nous alors arbitrairement $h(x)$ et $l(x)$ et soit s une variable arbitraire : la fonction

$$\int_{-\infty}^{+\infty} e^{s h(x) + l(x)} dx \quad 1)$$

sera une fonction positive connue que nous pourrons représenter par $e^{\Psi(s)}$. On voit alors que $\mu(\theta)$ sera défini par

$$\theta \mu' - \mu = \Psi(\mu')$$

¹⁾ $h(x)$ et $l(x)$ sont arbitraires, mais tels bien entendu que l'intégrale représentée ci-dessous par $e^{\Psi(s)}$ soit convergente et dérivable en θ .

$$(55) \quad \mu = \theta \mu' - \psi(\mu')$$

c'est-à-dire une équation de Clairaut. La solution $\mu' = \text{constante}$ réduirait $f(x, \theta)$, d'après (48) à une fonction indépendante de θ , cas où le problème n'aurait plus de sens. μ est donc donné par la solution singulière de (55), qui est unique et s'obtient en éliminant s entre $\mu = \theta s - \psi(s)$ et $\theta = \psi'(s)$ ou encore entre

$$(55\text{bis}) \quad \begin{aligned} e^{\theta s - \mu} &= \int_{-\infty}^{+\infty} e^{s h(x) + l(x)} dx \text{ et} \\ \int_{-\infty}^{+\infty} e^{s h(x) + l(x)} [h(x) - \theta] dx &= 0. \end{aligned}$$

Si l'on veut, $\mu(\theta)$ est donné par la relation

$$e^{-\mu} = e^{-\theta s} \int_{-\infty}^{+\infty} e^{s h(x) + l(x)} dx$$

où s est donné en fonction de θ par la relation implicite (55bis).

Fonction H' minimisante. Quelle est alors quand on connaît la densité distinguée $f(x, \theta)$, et qu'on l'a mise sous la forme (48), celle, H' , des fonctions $H(X_1, \dots, X_n)$ vérifiant $\theta = \mathfrak{N}_\theta H$ et telle que σ_H atteigne pour chaque valeur de θ , un minimum absolu, égal à $\frac{1}{\sigma_A \sqrt{n}}$.

C'est, d'après ce qui précède, c'est-à-dire d'après (42)

$$(56) \quad H'(X_1, \dots, X_n) = \frac{1}{n} [h(X_1) + \dots + h(X_n)].$$

Cette fonction est bien déterminée connaissant $f(x, \theta)$ quand f est de la forme (48).

Pour déterminer H' connaissant f , il n'est, cependant, pas nécessaire de mettre d'abord f sous la forme (48) afin de calculer d'abord $h(x)$. Il suffit d'écrire H' sous la forme (41) qui ne nécessite que la connaissance de f . Si $f(x, \theta)$ est une fonction distinguée, le second membre ne dépendra qu'en apparence de θ et se réduira de lui-même à une expression de la forme (56) ce qui sera une manière de déterminer h connaissant f .

De même, on peut calculer la valeur empirique $t = H'(x_1, \dots, x_n)$ correspondant à un échantillon donné, soit par la formule simple $t = \frac{\sum h(x_i)}{n}$ quand on connaît h , soit seulement connaissant f .

En effet, H' est aussi donnée sous la forme (40). De sorte que

$$t = \theta + \lambda(\theta) \sum \frac{\partial}{\partial \theta} L f(x_i, \theta)$$

où le second membre ne dépend qu'en apparence de θ . En prenant $\theta = t$, on a donc, puisque $\lambda(\theta) > 0$,

$$(57) \quad \sum_i \frac{\partial}{\partial t} L f(x_i, t) = 0$$

Ainsi, quand $f(x, \theta)$ est une densité distinguée, la valeur empirique t de θ correspondant à un échantillon x_1, \dots, x_n est une racine de l'équation (57) en t .

D'ailleurs, cette équation a *une racine et une seule* quand X est une variable distinguée. Car alors, d'après (48)

$$\Sigma \frac{\partial}{\partial t} L f(x_i, t) = \mu_t'' \left[\frac{\Sigma h(x_i)}{n} - t \right]$$

avec $\mu'' > 0$. On retrouve la racine (unique) $t = \frac{\Sigma h(x_i)}{n}$

qui serait obtenue si l'on avait employé directement pour H' l'expression (56).

Expression de la valeur empirique. On a vu que: pour que σ_{T_n} atteigne la borne inférieure $\frac{1}{\sigma_A \sqrt{n}}$, il faut et il suffit que $f(x, \theta)$ soit une densité distinguée.

Constatons alors que s'il en est ainsi, il n'y a qu'une seule fonction H' , une seule évaluation empirique de θ fournie par la méthode de la plus petite dispersion. Cette fonction est donnée, connaissant f , par la formule (41); mais quand on a pu mettre f sous la forme (48), on constate que H' est de la forme

$$T \equiv H'(X_1, \dots, X_n) = \frac{1}{n} \Sigma h(X_i)$$

Cette fonction H' ne peut donc avoir une forme arbitraire: non seulement, c'est une somme de fonctions ne dépendant chacune que d'une des quantités X_1, \dots, X_n , mais c'est même la *moyenne arithmétique des n valeurs observées $h(X_i)$ d'une même variable aléatoire auxiliaire $Y = h(X)$.*

Expression de la dispersion. Quand la densité $f(x, \theta)$ est distinguée, l'écart quadratique moyen de T_n est donné d'après ce qui précède, par

$$(\sigma_{T_n})^2 = \frac{1}{n(\sigma_A)^2} = \frac{1}{n \int_{-\infty}^{+\infty} \left[\frac{\partial f(x, \theta)}{\partial \theta} \right]^2 \frac{dx}{f(x, \theta)}}$$

Quand on a pu mettre f sous la forme (48), on peut obtenir σ_{T_n} sans intégration, explicitement en fonction de θ . En effet, dans ce cas nous avons établi les formules (52) et (52bis), d'où résulte immédiatement:

$$(58) \quad \boxed{(\sigma_{T_n})^2 = \frac{1}{n \mu''_{\theta^2}}}$$

Loi de probabilité de la valeur empirique. En posant $Y = h(X)$, on voit que T_n étant la moyenne arithmétique de n valeurs (indépendantes) de Y , sa loi de probabilité réduite tendra quand $n \rightarrow \infty$ vers la loi de Laplace sous des conditions peu limitatives à imposer à Y ¹⁾, ou, par suite, à X . La valeur moyenne θ de T_n étant indépendante de n , la valeur moyenne de la variable réduite $\xi_n = \frac{T_n - \theta}{\sigma_A} \sqrt{n}$ restera nulle et son écart quadratique moyen restera égal à 1. Dès lors la probabilité élémentaire de ξ_n tendra vers

$$\frac{1}{\sqrt{2\pi}} e^{-\frac{\xi^2}{2}} d\xi$$

De sorte que l'expression asymptotique de la probabilité élémentaire de T_n sera

$$\sqrt{n} \frac{1}{\sigma_A \sqrt{2\pi}} e^{-\frac{n(t-\theta)^2}{2\sigma_A^2}} dt \quad \text{avec } (\sigma_A)^2 = \mu''_{\theta^2}.$$

¹⁾ Il suffit de supposer que Y a une valeur moyenne et une fluctuation finies.

L'écart moyen de ξ_n tendra vers $\sqrt{\frac{2}{\pi}}$ et par suite celui de T_n aura pour expression asymptotique

$$\sqrt{\frac{2}{n\pi}} \sigma = \frac{1}{\sqrt{n}} \left[\sigma_A \sqrt{\frac{2}{\pi}} \right].$$

Généralisation. Soit une fonction quelconque $\alpha(\tau)$. Si l'on pose

$$\varphi(x, \tau) = f(x, \alpha(\tau))$$

on voit qu'on peut dire aussi bien que la densité élémentaire de X est $\varphi(x, \tau)$ que $f(x, \theta)$. La méthode précédente d'évaluation approchée de θ_0 était au fond une méthode d'évaluation de la vraie densité $\delta(x) = f(x, \theta_0)$. Or si $\theta_0 = \alpha(\tau_0)$, on a $\delta(x) = \varphi(x, \tau_0)$. La méthode précédente appliquée à $f(x, \theta)$ et à $\varphi(x, \tau)$ doit donner chaque fois une évaluation approchée de la même fonction $\delta(x)$. Mais nous avons vu qu'appliquée à $f(x, \theta)$, on était conduit à la restreindre au cas où $f(x, \theta)$ a une forme spéciale. Nous sommes alors conduit à poser la question sous une forme plus générale qui va élargir le champ de validité de la méthode. Il s'agit de chercher les formes de $f(x, \theta)$ telles qu'il existe au moins un changement de paramètre $\theta = \alpha(\tau)$ après lequel la densité de probabilité $\varphi(x, \tau) = f(x, \alpha(\tau))$, devienne distinguée.

On pourra alors prendre pour valeur empirique de τ des fonctions $K(X_1, \dots, X_n)$ telles que $\mathfrak{N}_\theta K$ — qui est aussi une fonction de τ : $\mathfrak{N}_{\alpha(\tau)} K(X_1, \dots, X_n)$ — soit précisément égal à τ et non plus à $\theta = \alpha(\tau)$. Parmi ces fonctions K indépendantes de τ , on choisira celle pour laquelle $\mathfrak{N}_\theta (K - \tau)^2 = \mathfrak{N}_{\alpha(\tau)} (K - \tau)^2$ est égal à sa borne inférieure. Une telle fonction ne sera indépendante de τ et par suite de θ que si $\varphi(x, \tau)$ est de la forme (48), soit ici

$$(60) \quad \varphi(x, \tau) = e^{\nu_\tau [k(x) - \tau] + \nu(\tau) + m(x)}.$$

Il s'agit de voir quelle expression de $f(x, \theta)$ en fonction de θ en résultera.

Puisqu' inversement τ est fonction de θ , $\tau = \alpha(\theta)$, on aura pour f une expression de la forme

$$(60) \quad f(x, \theta) = \varphi(x, \alpha(\theta)) = e^{e^{(\theta)} k(x) + \chi(\theta) + m(x)} \quad 1).$$

Ainsi, pour pouvoir appliquer notre méthode plus générale il sera nécessaire que $L f(x, \theta)$ soit de la forme

$$(61) \quad L f(x, \theta) = \varrho(\theta) k(x) + \chi(\theta) + m(x)$$

où $\varrho, \chi; k, m$ sont certaines fonctions de θ et de x respectivement, choisies de telle façon que l'on ait

$$(62) \quad \int_{-\infty}^{+\infty} f(x, \theta) dx = 1.$$

Les conditions nécessaires (60) et (62) suffisent-elles? Autrement dit, peut-on choisir α, l, μ , tels que ϱ, k, χ, m vérifiant (60) et (62), on puisse écrire

¹⁾ Après avoir obtenu cette solution de notre problème, nous avons remarqué qu'elle coïncide avec la solution, obtenue par G. Darmois, d'un autre problème: trouver la forme générale que doit avoir $f(X, \theta)$ pour qu'il existe une fonction $K(X_1, \dots, X_n)$ constituant un „résumé exhaustif“ (sufficient statistic, dirait R. A. Fisher, qui a introduit cette notion) de l'échantillon X_1, \dots, X_n relativement à θ et à $f(x, \theta)$. Voir, par exemple, G. Darmois, Résumés exhaustifs d'un ensemble d'observations (Bull. Inst. Int. de Stat. XXXIX: 2, Athènes 1937, pp. 288—293).

l'identité

$$\varrho(\theta) k(x) + \chi(\theta) + m(x) = \nu'_{\tau} [k(x) - \tau] + \nu(\tau) + m(x)?$$

Il suffira de prendre

$$(62\text{bis}) \quad \nu'_{\tau} = \varrho(\theta), \quad \nu(\tau) - \tau \nu'_{\tau} = \chi(\theta).$$

Or cela est possible, car en différentiant ces deux relations et divisant membre à membre, on aura :

$$(63) \quad \tau = \frac{-\chi'_{\theta}}{\varrho'_{\theta}}$$

d'où, en portant dans la seconde relation :

$$(64) \quad \nu = \frac{-\varrho}{\varrho'_{\theta}} \chi'_{\theta} + \chi.$$

L'expression de τ en fonction de θ est fournie par (63) et la fonction $\nu(\tau)$ s'obtient en éliminant θ entre (63) et (64).

On retiendra, en particulier, que si la densité f est de la forme (60), d'une part, il y a au moins un changement de variable $\tau = a(\theta)$ qui transforme $f(x, \theta)$ en une densité distinguée $\varphi(x, \tau)$; et d'autre part, il n'y a qu'un tel changement de variable: celui défini par (63).

Toutefois, dans les raisonnements précédents, nous avons implicitement supposé possible de résoudre l'équation (63) en θ . A ce sujet, il est intéressant d'établir une relation qualitative entre ϱ' et $\frac{\chi'_{\theta}}{\varrho'_{\theta}}$. En dérivant l'équation (62), on obtient

$$(65) \quad \int_{-\infty}^{+\infty} f(x, \theta) [\varrho'_{\theta} k(x) + \chi'_{\theta}] dx = 0.$$

Observons d'abord que ϱ'_{θ} ne peut être $\equiv 0$, sans quoi, ϱ serait indépendant de θ , et alors aussi $\chi(\theta)$, d'après (62bis) et par suite f serait d'après (61) indépendant de θ , cas où le problème de la détermination de θ_0 n'aurait plus de sens. On peut diviser par ϱ'_{θ} , d'où :

$$\int_{-\infty}^{+\infty} f(x, \theta) \left[k(x) + \frac{\chi'_{\theta}}{\varrho'_{\theta}} \right] dx = 0.$$

D'où en dérivant encore une fois en θ

$$(66) \quad \varrho' \int_{-\infty}^{+\infty} f(x, \theta) \left[k + \frac{\chi'_{\theta}}{\varrho'_{\theta}} \right]^2 dx + \frac{d}{d\theta} \frac{\chi'_{\theta}}{\varrho'_{\theta}} = 0.$$

Nous voyons que ϱ' et $\frac{d}{d\theta} \frac{\chi'_{\theta}}{\varrho'_{\theta}}$ sont de signes contraires. Par suite: il suffit que ϱ'_{θ} ne s'annule pas pour qu'on soit assuré que l'équation (63) n'a pas plus d'une racine en θ (Si les limites de variation de $\left(-\frac{\chi'_{\theta}}{\varrho'_{\theta}}\right)$, finies ou non, sont τ_1 et τ_2 , l'équation (63) aura une solution et une seule pour $\tau_1 < \tau < \tau_2$).

Nombre des fonctions arbitraires dans $f(x, \theta)$. On pourra appeler densité quasi-distinguée une densité de probabilité $f(x, \theta)$ s'exprimant sous la forme (60) où k, m, ϱ, χ sont seulement assujettis à la condition (62).

On observe d'ailleurs qu'on peut prendre ϱ , k , m , arbitrairement ¹⁾, χ étant alors déterminé par la relation

$$(69) \quad e^{-\chi(\theta)} = \int_{-\infty}^{+\infty} \varrho e^{\varrho(\theta)k(x)+m(x)} dx.$$

Dans ces conditions, la valeur empirique τ' de $\tau_0 = a(\theta_0)$ serait fournie par

$$\tau' = \frac{\sum k(x_i)}{n}$$

et par suite celle, θ' , de θ par

$$(70) \quad \boxed{\frac{-\chi'_{\theta'}}{\varrho'_{\theta'}} = \frac{\sum k(x_i)}{n}.}$$

Dans le cas actuel, on a

$$\begin{aligned} \Sigma \frac{\partial}{\partial \theta} L f(x_i, \theta) &= \Sigma \frac{\partial}{\partial \theta} [\varrho(\theta)k(x_i) + m(x_i) + \chi(\theta)] \\ &= n \varrho'_{\theta} \left[\frac{\sum k(x_i)}{n} + \frac{\chi'_{\theta}}{\varrho'_{\theta}} \right], \end{aligned}$$

de sorte que la valeur θ' cherchée est encore ici une racine de l'équation en θ , (57), comme on l'avait prouvé dans le cas particulier où f est distinguée.

Quant au nombre de racines de cette équation, on voit que si $\frac{\chi'_{\theta}}{\varrho'_{\theta}}$ est une fonction monotone, et si ϱ'_{θ} ne s'annule pas, l'équation ne pourra avoir plusieurs racines. Cette condition peut être simplifiée, comme précédemment: si ϱ'_{θ} ne s'annule pas, on a vu que $\frac{\chi'_{\theta}}{\varrho'_{\theta}}$ est alors monotone. Donc: il suffit que ϱ'_{θ} ne s'annule pas pour que l'équation (57) ne possède pas plus d'une seule racine. Elle en possèdera certainement une si l'on a

$$\tau_1 < \frac{\sum k(x_i)}{n} < \tau_2$$

Remarque. Si l'on représente par $F[u(x, \theta)]$ l'opération

$$(71) \quad F[u(x, \theta)] = \frac{\partial^2}{\partial x \partial \theta} L | u(x, \theta)|$$

on voit immédiatement que la condition nécessaire et suffisante pour qu'une densité $f(x, \theta)$ soit quasi-distinguée, — c'est-à-dire de la forme (61) — est que la fonction f vérifie l'équation aux dérivées partielles du quatrième ordre

$$(72) \quad \frac{\partial^2}{\partial x \partial \theta} L \left| \frac{\partial^2}{\partial x \partial \theta} L f(x, \theta) \right| = 0$$

c'est-à-dire $F[F[f(x, \theta)]] = 0$ qu'on peut écrire

$$(73) \quad F^2[f(x, \theta)] = 0.$$

Expression de la valeur empirique de θ . On a vu que si la densité est quasi-distinguée, la valeur empirique d'un certain nouveau paramètre, τ , est la moyenne arithmétique des valeurs observées d'une certaine nouvelle variable aléatoire Y .

¹⁾ Ici encore l'arbitraire dans le choix de ϱ , k , m est limité qualitativement par la condition que l'intégrale (69) soit convergente et dérivable en θ .

Quand on a pu mettre f sous la forme (60), ces nouvelles quantités τ , Y sont liées aux anciennes θ , X , par les relations :

$$\tau = -\frac{\chi'_{\theta}}{\varrho'_{\theta}}, \quad Y = k(X).$$

de sorte que H' est déterminée par l'équation implicite

$$-\frac{\chi'(H(X_1, \dots, X_n))}{\varrho'(H(X_1, \dots, X_n))} = \frac{1}{n} \sum k(X_i)$$

Expression de la dispersion. La dispersion de la valeur empirique du nouveau paramètre τ , c'est-à-dire de $V = \frac{1}{n} \sum k(X_i)$ est, en posant encore $\varphi(x, \tau) = f(x, \alpha(\tau))$, encore donnée par

$$(\sigma_V)^2 = \frac{1}{n \int_{-\infty}^{+\infty} \left[\frac{\partial}{\partial \tau} \varphi(x, \tau) \right]^2 \frac{dx}{\varphi(x, \tau)}}$$

En vertu de (58), on peut l'écrire avec les notations de la page 196

$$(\sigma_V)^2 = \frac{1}{n \nu''_{\tau}}$$

et d'après (63), (64)

$$(\sigma_V)^2 = -\frac{\frac{\partial}{\partial \theta} \frac{\chi'_{\theta}}{\varrho'_{\theta}}}{n \varrho'_{\theta}}$$

Loi de probabilité de la valeur empirique. La densité $\varphi(x, \tau)$ étant distinguée, la loi de probabilité réduite de la valeur empirique $V = \frac{\sum k(X_i)}{n}$ du nouveau paramètre τ tendra à la limite vers la loi de Laplace: la probabilité élémentaire pour que V soit compris entre t et $t + dt$ aura une expression asymptotique de la forme (59), du moins sous des conditions peu limitatives pour $f(x, \theta)$.

Mais il n'en sera pas de même pour l'ancien paramètre T : la loi réduite d'une fonction $T = \alpha(V)$ d'une variable aléatoire V dont la loi réduite tend vers la loi de Laplace, ne tend pas en général vers cette loi de Laplace.

DEUXIEME PARTIE

Comparaison avec d'autres méthodes de détermination empirique d'un paramètre dans le cas des petits échantillons.

Nous venons d'appliquer aux petits échantillons un certain principe de choix qu'on pourrait brièvement appeler „principe de la plus petite dispersion” et qui permet de déterminer une valeur plausible du paramètre θ , au moins quand la densité de probabilité $f(x, \theta)$ est une fonction quasi-distinguée. A côté de ce principe, déjà antérieurement pratiqué pour les grands échantillons (sous sa

forme restreinte où la valeur typique choisie est la valeur moyenne et où la dispersion est évaluée par l'écart quadratique moyen), d'autres principes ont été proposés que nous allons comparer au premier pour constater que, dans le cas où le premier est applicable, ils donnent le même résultat. Cette comparaison, déjà faite antérieurement pour les grands échantillons, se trouvera donc étendue ici aux petits échantillons.

Principe de la valeur dominante.

Supposons que l'on connaisse, non seulement la forme de la fonction $f(x, \theta)$, — telle que les échantillons considérés correspondent à une densité de probabilité élémentaire $f(x, \theta_0)$ —, mais encore que l'on connaisse la loi de probabilité a-priori de θ , définie, par exemple, par une densité de probabilité $\psi(\theta)$.

Il y a deux parties dans cette nouvelle hypothèse: 1°) on suppose précisé en quel sens θ est considérée comme une variable aléatoire, 2°) on donne sa densité de probabilité. La première partie est la plus délicate. Nous avons affaire à une variable aléatoire X relative à une catégorie C_{θ_0} bien déterminée d'épreuves où elle a une densité de probabilité $f_0(x) = f(x, \theta_0)$. Nous allons maintenant supposer que C_{θ_0} appartient à une famille de catégories C_{θ} d'épreuves, famille qui, elle, n'est pas donnée et dont la réalisation matérielle n'est pas fixée, mais est telle que dans la catégorie C_{θ} d'épreuves, la densité de probabilité de X soit, par hypothèse, $f(x, \theta)$.

Ceci étant, soient x_1, \dots, x_n , n valeurs possibles de X , données d'avance ainsi que dx_1, \dots, dx_n . Considérons les opérations suivantes: on a choisi au hasard la catégorie C_{θ_1} , puis ayant observé un échantillon X_1, \dots, X_n correspondant à θ_1 , sans avoir pu déterminer exactement les valeurs de X_1, \dots, X_n , on a pu constater que

$$(80) \quad x_1 < X_1 < x_1 + dx_1, \dots, x_n < X_n < x_n + dx_n.$$

On veut calculer la probabilité $\delta d\theta$ pour que, cette constatation ayant eu lieu, la catégorie C_{θ_1} ait été choisie de sorte que $\theta < \theta_1 < \theta + d\theta$.

D'après la formule de Bayes-Laplace, en posant:

$$(81) \quad D(x_1, \dots, x_n, \theta) \equiv \prod_{i=1}^n f(x_i, \theta), \text{ on aura}$$

$$\delta d\theta = \frac{D(x_1, \dots, x_n, \theta) dx_1 \dots dx_n \psi(\theta) d\theta}{\int_{-\infty}^{+\infty} D(x_1, \dots, x_n, \theta) dx_1 \dots dx_n \psi(\theta) d\theta}$$

d'où l'on tire, pour δ , une expression de la forme

$$\delta(x_1, \dots, x_n, \theta) = \frac{D(x_1, \dots, x_n, \theta) \psi(\theta)}{I(x_1, \dots, x_n)}$$

Une fois admises — et c'est là le point douteux — les hypothèses précédentes, cette valeur de δ est incontestable. Il s'agit maintenant d'en conclure la valeur approchée θ' qu'on se propose d'attribuer à θ_0 connaissant x_1, \dots, x_n . On décide, — et il faut bien marquer qu'on va ainsi faire une convention assurément plausible mais non démontrable, — qu'on prendra pour cette valeur θ' , la valeur de θ la plus probable ou plus exactement sa valeur dominante, c'est-à-dire celle, pour laquelle la densité δ de probabilité est la plus grande. On est donc ramené pour calculer θ' à chercher la valeur de θ qui rend maximum la quantité $\delta(x_1, \dots, x_n, \theta)$ ou, puisque x_1, \dots, x_n , sont donnés, la quantité $D(x_1, \dots, x_n, \theta) \psi(\theta)$, c'est-

à-dire $f(x_1, \theta) \dots f(x_n, \theta) \psi(\theta)$. Nous voyons que pour appliquer la méthode, il a d'abord fallu faire des hypothèses et des conventions qui ne s'imposent pas inexorablement et il faut enfin maintenant préciser la valeur qu'on donnera à $\psi(\theta)$.

On a reproché aux méthodes qui, dans le problème actuel ou dans d'autres tout différents, s'appuient sur la formule de Bayes-Laplace, de faire appel à trop d'hypothèses qu'on ne peut justifier.

Cela est tout à fait exact; mais c'est, par contre, le mérite de ces méthodes, de mettre bien en évidence les hypothèses que l'on fait. Quand, comme on le fait bien souvent, on suppose uniforme la distribution de la probabilité à priori de θ , on sait combien cette hypothèse est hasardeuse et on n'attache pas à la valeur qui en résulte pour θ' une certitude qui ne serait pas justifiée. Dans d'autres méthodes comme celle que nous étudierons ensuite, il est plus difficile de se rendre compte de la conformité plus ou moins approchée des faits aux hypothèses admises, celles-ci n'apparaissant pas explicitement dans l'énoncé du principe qu'on applique.

Quand on suppose uniforme la distribution de θ dans un intervalle fini a, b on aura $\psi(\theta) = \text{constante} = \frac{1}{b-a}$ entre a et b et $\psi(\theta) = 0$ en dehors. Dans ce cas, θ' correspondra aussi au maximum de

$$D(x_1, \dots, x_n, \theta) = f(x_1, \theta) \dots f(x_n, \theta)$$

Deux incertitudes de la méthode. Nous avons signalé une première difficulté dans l'application de la méthode de la valeur dominante, qui consiste en ce que pour l'appliquer, il faudrait connaître la fonction $\psi(\theta)$ qu'en général on ignore. Il en résulte une incertitude dans le choix de θ' qui ne résulte pas de la formule de Bayes ni de la méthode elle-même, mais de l'incertitude de nos connaissances. Nous allons maintenant signaler une autre incertitude qui réside dans le principe même de la méthode, dans l'adoption pour θ' , d'une valeur dominante.

Changement de paramètre. Supposons, ici encore, qu'on fasse un changement de paramètre $\theta = a(\tau)$, $\tau = a(\theta)$ en supposant que $a(\tau)$ et $a(\theta)$ ont des dérivées partout positives (pour $a(\tau)$ dans l'intervalle, fini ou infini, de variation de τ). Alors la valeur $\tau' = a(\theta')$ qui correspond à la valeur θ' cherchée, est telle qu'elle rend maximum l'expression $D(x_1, \dots, x_n, a(\tau)) \psi(a(\tau))$.

Mais supposons qu'on ait directement appliqué la méthode à $\psi(x, \tau) = f(x, a(\tau))$. On aurait introduit la probabilité élémentaire à priori de τ , soit $\Psi(\tau) d\tau = \psi(\theta) d\theta = \psi(a(\tau)) a'(\tau) d\tau$ et il aurait fallu rendre maximum

$$D(x_1, \dots, x_n, a(\tau)) \Psi(\tau) = D(x_1, \dots, x_n, a(\tau)) \psi(a(\tau)) a'(\tau).$$

On aurait donc choisi une valeur $\tau'' = a(\theta'')$ telle que

$$D(x_1, \dots, x_n, \theta) \psi(\theta) a'(a(\theta))$$

soit maximum pour $\theta = \theta''$. Ainsi la valeur cherchée serait, suivant qu'on cherche θ directement ou par l'intermédiaire de τ , la valeur θ' qui rend maximum le produit $\bar{\omega}(\theta) = D(x_1, \dots, x_n, \theta) \psi(\theta)$ ou la valeur θ'' qui rend maximum le produit

$$\bar{\omega}_1(\theta) = \bar{\omega}(\theta) a'(a(\theta)) = \frac{\bar{\omega}(\theta)}{a'_\theta}$$

Or en général, ces deux valeurs peuvent être distinctes. Et comme, à priori, nous n'avons pas de raison d'exprimer la probabilité élémentaire de X en prenant comme paramètre plutôt θ que τ , nous sommes contraints à constater une incertitude dans le choix de θ qui ne tient pas à l'emploi du théorème de Bayes, mais

à celui du principe même de la dominante. On peut s'expliquer cette incertitude ainsi: une valeur dominante n'est pas une valeur plus probable mais une valeur centre du petit intervalle de longueur *donnée* qui est la plus probable. Or quand on change le paramètre, des intervalles de variations égales de l'ancien paramètre deviennent inégaux avec le nouveau paramètre; de sorte que la valeur dominante de l'ancien paramètre, non seulement n'est pas égale à la valeur dominante du nouveau, mais elle ne lui correspond même pas.

Elimination de l'incertitude. Il existe un moyen d'éliminer à la fois de la méthode les deux sortes d'incertitude que nous avons signalées. Il consiste à n'appliquer la méthode qu'en choisissant parmi les différentes représentations paramétriques $\varphi(x, \tau)$ de la densité de probabilité, un paramètre particulier, un paramètre préférentiel, et en le choisissant en outre de façon à connaître sa distribution de probabilité à priori. Or cela est possible de la façon suivante.

Nous avons vu qu'il fallait rendre maximum $\bar{\omega}_1(\theta) = D(x_1, \dots, x_n, \theta) \frac{\psi(\theta)}{a'(\theta)}$

Supposons qu'on prenne $a(\theta)$ de sorte que $\frac{\psi(\theta)}{a'(\theta)}$ soit une constante, par exemple, l'unité. De sorte que $a(\theta) = \int \psi(\theta) d\theta$ ou $d\tau = \psi(\theta) d\theta$.

C'est ce qu'on obtiendra en prenant pour $a(\tau)$ la fonction de répartition à priori de θ : $a(\tau) = \text{Prob. } \{\theta < \tau\} = \int_{-\infty}^{\tau} \psi(\theta) d\theta$

$\tau = a(\theta)$ est dans ce cas un paramètre privilégié: c'est celui parmi les paramètres (fonctions de θ) possibles qui a une répartition uniforme ¹⁾ des probabilités à priori. Si l'on restreint l'application du principe de la dominante à ce paramètre privilégié, on voit que la valeur correspondante du paramètre θ donné, sera encore obtenue comme valeur rendant maximum $D(x_1, \dots, x_n, \theta)$ c'est-à-dire $f(x_1, \theta) \dots f(x_n, \theta)$, ou encore rendant maximum

$$(82) \quad \Sigma L f(x_i, \theta).$$

Quand $f(x, \theta)$ est dérivable en θ , θ' sera donc racine de l'équation en θ déjà rencontrée

$$(57) \quad \sum_{i=1}^n \frac{\partial}{\partial \theta} L f(x_i, \theta) = 0$$

On ne peut dire qu'inversement, une racine de (57) correspond nécessairement à un maximum de (82). Mais considérons le cas où $-\frac{\partial^2}{\partial \theta^2} L f(x, \theta)$ ne change pas de signe quand x varie. Alors, ce signe sera négatif: en effet, on a

$$\begin{aligned} \int_{-\infty}^{+\infty} f(x, \theta) dx &= 1, \text{ d'où en dérivant en } \theta \\ \int_{-\infty}^{+\infty} f'_{\theta}(x, \theta) dx &= 0, \text{ qu'on peut écrire} \\ \int_{-\infty}^{+\infty} \left[\frac{\partial}{\partial \theta} L f(x, \theta) \right] f(x, \theta) dx &= 0; \end{aligned}$$

¹⁾ En réalité, toute fonction linéaire de τ aurait aussi une répartition uniforme, mais conduirait à la même conclusion.

d'où en dérivant en θ

$$\int_{-\infty}^{+\infty} \left[\frac{\partial^2}{\partial \theta^2} L f(x, \theta) \right] f(x, \theta) dx + \int_{-\infty}^{+\infty} \left[\frac{\partial}{\partial \theta} f(x, \theta) \right]^2 \frac{dx}{f(x, \theta)} = 0$$

Ce qui établit notre remarque.

Dès lors, $\Sigma \frac{\partial}{\partial \theta} L f(x_i, \theta)$ passe du signe $+$ au signe $-$ quand θ croît en passant par une racine de (57). Et par suite: lorsque $\frac{\partial^2}{\partial \theta^2} L f(x, \theta)$ garde un signe constant, l'équation (57) ne peut avoir plus d'une racine en θ , et, si elle a une racine, celle-ci correspond nécessairement à un maximum absolu de (82), c'est-à-dire est égale à la valeur θ' cherchée.

Dans le cas particulier, où $f(x, \theta)$ est une fonction distinguée, le principe de la plus petite dispersion et le principe de la valeur la plus probable (appliqué dans l'hypothèse de l'uniformité des probabilités a priori des valeurs de θ) conduisent chacun à une seule valeur approchée de θ_0 , et ces deux principes fournissent la même valeur approchée, θ' , de θ_0 , laquelle valeur est l'unique racine de (57).

Quand $f(x, \theta)$ est une fonction quasi-distinguée, le même résultat subsiste au moins dans le cas où, en écrivant $f(x, \theta)$ sous la forme (60), la fonction ϱ'_θ ne change pas de signe quand θ varie.

Principe de la plus grande plausibilité.

Dans un travail consacré à la détermination empirique de l'écart quadratique moyen σ d'une variable aléatoire X , Helmhert avait calculé la probabilité élémentaire δds pour que l'écart quadratique moyen empirique de l'échantillon,

$$\sqrt{\frac{\Sigma \left(x_i - \frac{\Sigma x_i}{n} \right)^2}{n}},$$

soit compris entre s et $s + ds$, en supposant que X vérifie la loi dite normale. Il a alors posé a priori le principe ¹⁾ que la meilleure valeur, s , à choisir pour σ connaissant l'échantillon x_1, \dots, x_n , était celle pour laquelle la densité correspondante est maximum.

R. A. Fisher ²⁾ a généralisé ensuite le principe de Helmhert — d'ailleurs, avant d'en avoir connu l'existence — au cas d'une variable aléatoire ayant une densité de probabilité de forme quelconque mais dépendant d'un paramètre inconnu. Il a donné à ce principe généralisé le nom de „method of maximum likelihood”, que nous traduirons librement par l'expression „application du principe de la plus grande plausibilité”. Il est essentiel d'observer que, dans le cas de Helmhert comme dans celui de Fisher, il ne s'agit pas d'une nouvelle méthode appliquant des axiomes ou principes connus du Calcul des Probabilités mais de l'application d'un nouveau principe qu'on est libre d'accepter ou de rejeter et dont la valeur ne pourra être établie que par de nombreux succès obtenus au cours d'investigations dans des domaines variés. Ces succès consisteront à trouver en appliquant ce principe des valeurs suffisamment approchées de paramètres dont par ailleurs on a découvert la valeur par d'autres moyens.

Précisons sa forme générale appliquée au cas d'une variable aléatoire possédant une densité de probabilité $f(x, \theta_0)$ où l'on connaît la fonction à deux variables $f(x, \theta)$ sans connaître θ_0 .

¹⁾ *Astronomische Nachrichten*, 88, no. 2096, 122, 1876.

²⁾ *Messenger of Mathematics*, 41, 1912, p. 155—160.

Si l'on connaît un échantillon x_1, \dots, x_n , la probabilité que l'on ait à la fois $x_1 < X_1 < x_1 + dx_1, \dots, x_n < X_n < x_n + dx_n$ est donnée par la probabilité élémentaire

$$\prod_{i=1}^n [f(x_i, \theta) dx_i] = D(x_1, \dots, x_n, \theta) dx_1 \dots dx_n.$$

Le principe de la plus grande plausibilité consiste à affirmer qu'on obtiendra une valeur approchée θ'' de θ_0 en prenant pour θ la valeur, θ'' , qui rend maximum, quand θ varie, la densité de probabilité $D(x_1, \dots, x_n, \theta)$. Ainsi il ne s'agit pas de supposer que l'échantillon x_1, \dots, x_n , est un échantillon „dominant” en ce sens qu'il a une plus grande densité de probabilité que les autres échantillons. C'est, au contraire, la valeur choisie θ'' , du paramètre et non l'ensemble x_1, \dots, x_n définissant l'échantillon, qui a une position dominante en ce sens que le même échantillon a une plus grande densité de probabilité pour la „population” correspondant à la valeur θ'' qu'on se propose de choisir que pour les autres valeurs de θ .

À la lumière de cette distinction, le principe apparaît bien comme un principe de choix assez arbitraire qui ne peut guère être justifié que par le succès, à vérifier, de ses applications.

Appliquons cette méthode dans le cas de la „première loi de Laplace” avec un écart moyen donné μ , mais une valeur probable inconnue θ . On aura

$$f(x, \theta) = \frac{1}{2\mu} e^{-\frac{|x-\theta|}{\mu}}.$$

Donc :

$$D(x_1, \dots, x_n, \theta) = \frac{1}{(2\mu)^n} e^{-\frac{1}{\mu} \sum_i |x_i - \theta|}$$

Le maximum de D quand θ varie sera obtenu quand $\sum_i |x_i - \theta|$ sera minimum, c'est-à-dire, comme on sait, quand θ est égal à la médiane (ou à une des médianes) des valeurs observées de $X : x_1, \dots, x_n$.

Cette application montre qu'il n'est pas nécessaire, pour pouvoir utiliser le principe, de supposer que $f(x, \theta)$ est dérivable en θ quels que soient x et θ .

Faisons cependant cette hypothèse; on aura alors, dans ce cas général, à chercher le maximum de $D(x_1, \dots, x_n) = \prod f(x_i, \theta)$, ou encore celui de

$$L D = \sum_i L f(x_i, \theta).$$

La valeur empirique θ'' cherchée devra donc vérifier la relation plusieurs fois déjà rencontrée

$$(57) \quad \sum \frac{\partial}{\partial \theta} L f(x_i, \theta) = 0.$$

Nous retrouvons donc l'équation à laquelle nous étions arrivés plus haut par la méthode de la valeur dominante.

Summary:

Writer hopes that the reader will find that some of the ideas or of the theorems of this memoir are new to him. However, as its contents were first intended as a chapter of a course on mathematical statistics, this paper must be judged from the didactic or pedagogical, as well as from the scientific point of view.

It is devoted to a critical study and to an extension of some of the methods of estimation of parameters of a priori adopted probability laws, when a small sample has been drawn. More precisely: Let X_1, \dots, X_n be a system of n random variables, and assume the elementary probability $f(x, \theta) dx$ of X to be known, leaving only undetermined the value of a parameter θ . The problem is to estimate θ by means of n observed values x_1, \dots, x_n of X_1, \dots, X_n . The estimation taken as an approximate value of θ shall be one function $t = H(x_1, \dots, x_n)$ of the known values x_1, \dots, x_n . Different methods have been proposed for the purpose of fixing which function H shall be chosen.

First, writer deals with what is sometimes called the best unbiased estimate of the parameter. H will be called an „unbiased” estimate of θ , if the average value of $H(x_1, \dots, x_n)$ is equal to θ . It is called the „best” unbiased estimate, if the variance of $H(x_1, \dots, x_n)$ — i.e. the average of $[H(x_1, \dots, x_n) - \theta]^2$ — is minimum. Writer makes clear that these definitions — though natural enough — are conventional and arbitrary ones. Most of the previous writers had applied this definition in the hypothesis of n being great and X satisfying the laplacian (so called normal) probability law.

n arbitrary: Writer showed in a course of lectures in 1939, how to extend their formula to the general case by using Schwarz' inequality. He first proved that $\sqrt{n} \sigma_H \geq \frac{1}{\sigma_A}$, where σ_v is written for the variance of v and $A = \frac{1}{f(x, \theta)} \frac{\partial}{\partial \theta} f(x, \theta)$. Later on, he happened to hear that Doob had previously obtained the same result by the same method.

n great: When, instead of supposing the average value of H to be θ , it is only assumed that this is so at the limit when $n \rightarrow \infty$, writer shows that the lowest limit of $\sqrt{n} \sigma_H$ is $\geq \frac{1}{\sigma_A}$.

Then, reverting to the first case, the question is to find, whether σ_H may be equal to its lower bound $\frac{1}{\sigma_A \sqrt{n}}$. Writer points out that, generally, this does not happen. For example, when $f(x, \theta) = g(x - \theta)$, the only case in which a function H may be found such that $\sigma_H = \frac{1}{\sigma_A \sqrt{n}}$ is that one when X is a laplacian (so called normal) variable.

Writer shows that a necessary and sufficient condition that one of the σ_H be equal to $\frac{1}{\sigma_A \sqrt{n}}$ is that functions $h(x), l(x), \mu(\theta)$ exist, such that $f(x) = e^{u'(\theta)[h(x) - \theta] + \mu(\theta) + l(x)}$. (We will then say that $f(x)$ is a „distinguished function”). Then the minimizing function H is $H'(x_1, \dots, x_n) = \frac{1}{n} [h(x_1) + \dots + h(x_n)]$; so that the estimate of θ is $t = \frac{1}{n} \sum_i h(x_i)$. In addition: $(\sigma_H)^2 = \frac{1}{n \mu''_{\theta}}$. When f is known, but h is not, t is the unique root of $\sum_i \frac{\partial}{\partial t} L f(x_i, t) = 0$.

Writing $T_n = H'(x_1, \dots, x_n)$ and $\xi_n = \frac{T_n - \theta}{\sigma_A} \sqrt{n}$, ξ_n tends to be a laplacian variable.

Then writer raises the question: which change of parameter $t = a(\tau)$ — whence $f(x, a(\tau)) = \varphi(x, \tau)$ — is such that $\varphi(x, \tau)$ is a distinguished function. And he finds that

in this case, functions $\varrho(\theta), \chi(\theta), k(x), m(x)$ must exist, such that $f(x, \theta) = e^{\varrho(\theta)k(x) + \lambda(\theta) + m(x)}$ (x, θ , being in such a case a „quasi“-distinguished function).

It happens that G. Darmonis previously came to the same class of functions by solving a different problem, mentioned in the present memoir p. 195.

In the second part, a comparison is made of the previous method with the method of the most probable value and with the method of maximum likelihood.

The first one is based on the Bayes-Laplace formula. It is pointed out that the main objection to the use of this formula (that generally we do not know the a priori probability of θ) is balanced by the advantage of our knowing exactly the hypotheses on which the solution is based and our therefore also knowing the weakness of the solution, a weakness which is just as real, but not so apparent in the other methods. It is then shown that the best unbiased estimate and the most probable value (when all values of θ are assumed to be equally probable) coincide when $f(x, \theta)$ is a distinguished function (and even when it is a quasi-distinguished function, provided ϱ'_{θ} does not change sign).

As for the method of maximum likelihood, it is pointed out that it is founded on a plausible, but arbitrary principle which can only be justified by the successes it will achieve in those cases, where verification is possible.
