DANIEL DUGUÉ

Application des propriétés de la limite au sens du calcul des probabilités à l'étude de diverses questions d'estimation

Thèses de l'entre-deux-guerres, 1937
<http://www.numdam.org/item?id=THESE_1937__196__1_0>

N° D'ORDRE :
2625
SÉRIE A.
N° DE SÉRIE :
1759

THÈSES

PRÉSENTÉES

A LA FACULTÉ DES SCIENCES DE PARIS

POUR OBTENIR

LE GRADE DE DOCTEUR ÈS SCIENCES MATHÉMATIQUES

PAR

M. Daniel DUGUÉ

1re THÈSE. — Application des propriétés de la limite au sens du calcul des probabilités a l'étude de diverses questions d'estimation.

2e THÈSE. — Fonctions fuschsiennes.

Soutenues le 12 NOV 1937 devant la Commission d'examen.

MM. E. BOREL, *Président.*
R. FRÉCHET } *Examinateurs.*
G. DARMOIS }

PARIS
GAUTHIER-VILLARS, ÉDITEUR
LIBRAIRE DU BUREAU DES LONGITUDES, DE L'ÉCOLE POLYTECHNIQUE
Quai des Grands-Augustins, 55

1937

FACULTÉ DES SCIENCES DE L'UNIVERSITÉ DE PARIS

MM.

Doyen honoraire........ M. MOLLIARD.
Doyen.................. C. MAURAIN, *Professeur*, Physique du Globe.

Professeurs honoraires

H. LEBESGUE.	FREUNDLER.	VESSIOT.	CH. FABRY.
A. FERNBACH.	AUGER.	PORTIER.	LÉON BERTRAND.
ÉMILE PICARD.	BLAISE.	MOLLIARD.	WINTREBERT.
LÉON BRILLOUIN.	DANGEARD.	LAPICQUE.	DUBOSCQ.
GUILLET.	LESPIEAU.	G. BERTRAND.	BOHN.
PÉCHARD.	MARCHIS.	ABRAHAM.	

PROFESSEURS

M. CAULLERY......	T	Zoologie (Évolution des êtres organisés).
G. URBAIN........	T	Chimie générale.
EMILE BOREL.......	T	Calcul des probabilités et Physique mathématique.
JEAN PERRIN.......	T	Chimie physique.
E. CARTAN.........	T	Géométrie supérieure.
A. COTTON.........	T	Recherches physiques.
J. DRACH..........	T	Analyse sup^{re} et Algèbre sup^{re}.
CHARLES PÉREZ.....	T	Zoologie.
E. RABAUD.........	T	Biologie expérimentale.
M. GUICHARD......		Chimie minérale.
PAUL MONTEL......	T	Théorie des fonctions et Théorie des transformations.
L. BLARINGHEM...	T	Botanique.
G. JULIA...........	T	Mécanique analytique et Mécanique céleste.
C. MAUGUIN........	T	Minéralogie.
A. MICHEL-LÉVY...	T	Pétrographie.
H. BÉNARD.........	T	Mécanique expérimentale des fluides.
A. DENJOY.........	T	Application de l'Analyse à la Géométrie.
L. LUTAUD.........	T	Géographie physique et Géologie dynamique.
EUGÈNE BLOCH......	T	Physique théorique et Physique céleste.
G. BRUHAT.........		Physique.
E. DARMOIS........		Enseignement de Physique.
A. DEBIERNE......	T	Physique générale et Radioactivité.
A. DUFOUR.........	T	Physique (P. C. B.).
L. DUNOYER.......		Optique appliquée.
A. GUILLIERMOND.	T	Botanique.
M. JAVILLIER......		Chimie biologique.
L. JOLEAUD........		Paléontologie.
ROBERT-LÉVY......	T	Physiologie comparée.
F. PICARD..........		Zoologie (Évolution des êtres organisés).
HENRI VILLAT......	T	Mécanique des fluides et applications.
CH. JACOB..........	T	Géologie.
P. PASCAL..........	T	Chimie minérale.
M. FRÉCHET.......	T	Calcul différentiel et Calcul intégral.
E. ESCLANGON.....	T	Astronomie.
M^{me} RAMART-LUCAS	T	Chimie organique.
H. BÉGHIN.........	T	Mécanique physique et expérimentale.
FOCH...............		Mécanique expérimentale des fluides.
DE BROGLIE........	T	Théories physiques.
CHRÉTIEN..........		Optique appliquée.
P. JOB.............		Chimie générale.
LABROUSTE........		Physique du Globe.
PRENANT..........		Zoologie.
VILLEY.............		Mécanique physique et expérimentale.
COMBES............	T	Physiologie végétale.
GARNIER...........	T	Mathématiques générales.
PÉRÈS..............		Mécanique théorique des fluides.
HACKSPILL.........		Chimie (P. C. B.).
LAUGIER...........	T	Physiologie générale.
TOUSSAINT.........		Technique aéronautique.
M. CURIE...........		Physique (P. C. B.).
G. RIBAUD.........	T	Hautes températures.
CHAZY.............	T	Mécanique rationnelle.
GAULT.............		Chimie (P. C. B.).
CROZE.............		Recherches physiques.
DUPONT............	T	Théories chimiques.
LANQUINE..........		Géologie.
VALIRON...........		Mathématiques générales.
BARRABÉ..........		Géologie structurale et Géologie appliquée.
MILLOT............		Zoologie (P. C. B.).
F. PERRIN..........		Théories physiques.
VAVON.............		Chimie organique.
G. DARMOIS........		Calcul des probabilités et Physique mathématique.
CHATTON...........	T	Biologie maritime.
AUBEL..............		Chimie biologique.
JACQUES BOURCART.		Géographie physique et Géologie dynamique.
M^{me} JOLIOT-CURIE.		Physique générale et Radioactivité.
PLANTEFOL........		Biologie végétale (P.C.B.).
CABANNES.........		Enseignement de Physique.
GRASSÉ............		Biologie animale (P.C.B.).
PREVOST...........		Chimie (P.C.B.).

Secrétaire..................... A. PACAUD.
Secrétaire honoraire.......... D. TOMBECK.

A MES PARENTS

A Monsieur Georges DARMOIS

PROFESSEUR A LA SORBONNE

PREMIÈRE THÈSE.

APPLICATION DES PROPRIÉTÉS DE LA LIMITE

AU SENS DU CALCUL DES PROBABILITÉS

A L'ÉTUDE DE DIVERSES QUESTIONS D'ESTIMATION.

INTRODUCTION.

Le problème de l'estimation est la première de toutes les questions qu'est amené à se poser un physicien expérimental :

De quelle manière une série de mesures d'une même grandeur M, ayant donné des résultats $x_1, x_2, \ldots, x_n$ peut-elle renseigner sur sa vraie valeur? Depuis longtemps, on admettait, et cela était justifié (comme on le verra dans la suite) par le rôle prédominant que joue la loi de Gauss dans le Calcul des Probabilités, que la meilleure approximation était donnée par la moyenne arithmétique des résultats obtenus, soit $\frac{\Sigma x_i}{n}$.

Le but de ce travail a été d'éclaircir dans une certaine mesure les questions mathématiques soulevées par ce problème physique. Signalons, toutefois, que les procédés employés permettent d'atteindre non seulement la grandeur que l'on cherche à connaître, mais également la précision avec laquelle l'instrument de mesure fournit cette grandeur, et plus généralement un paramètre quelconque, figurant dans la loi de probabilité qui donne la répartition des résultats.

Soit, pour fixer les idées, une mesure obéissant à la loi de Cauchy :

$$\frac{1}{\pi}\frac{1}{1+\frac{(x-m)^2}{s^2}}\frac{dx}{s},$$

on pourra tout aussi bien évaluer M, abscisse du centre de la dispersion que s, échelle de cette dispersion. Soit encore la loi de Pareto, donnant la fréquence des individus possédant un revenu fixé d'avance (compris entre x_1 et x_2) :

$$y = K\int_{x_1}^{x_2} \frac{dx}{x^\alpha},$$

$\left(\text{K étant une constante telle que } K\int_{r_1}^{+\infty} \frac{dx}{x^\alpha} = 1.\ r_1 \text{ étant le revenu minimum}\right.$

$\left.\text{permettant l'existence}\right)$. Toute une série d'expériences, c'est-à-dire un certain nombre de revenus, pris au hasard dans une population donnée, pourra nous renseigner sur la valeur de l'exposant, et dont la signification est évidemment plus abstraite que celle des paramètres précédents.

Enfin, envisageons une loi uniforme de répartition $\frac{dx}{l}$, la variable aléatoire x étant comprise entre les limites o et l.

Nous étudierons également la meilleure manière d'arriver à être fixé sur la valeur de l. Il ne s'agit plus ici d'une grandeur à mesurer, mais de la limite que peut atteindre cette grandeur.

Les méthodes que nous emploierons consistent dans la formation d'une suite de fonctions $f_1(x_1), \ldots, f_n(x_1, \ldots, x_n)$ des résultats des expériences.

Chacune de ces fonctions a une loi de probabilité, c'est-à-dire qu'à deux nombres α et β ($\alpha < \beta$) on peut associer une quantité $P^i_{\alpha,\beta}$, mesure d'ordre i de l'ensemble des points de l'espace à i dimensions tels que $\alpha \leqq f_i(x_1, \ldots, x_i) < \beta$, la mesure d'ordre un ou linéaire de l'ensemble des points d'abscisses x, tels que $\gamma \leqq x < \delta$, étant, par définition, prise égale à la probabilité du même événement, soit $F(\delta) - F(\gamma)$, le signe F désignant la loi de probabilité totale.

Conformément aux notations employées par le statisticien britannique Mr. R.-A. Fischer, auteur de nombreux mémoires sur cette question, nous donnerons à ces fonctions le nom de « statistiques ».

Il est particulièrement intéressant d'étudier le comportement à la limite quand n croît indéfiniment de la loi de probabilité de f_n. C'est ainsi que nous serons amenés à distinguer entre deux sortes de « statistiques » :

Les premières sont telles que l'on puisse trouver un rang i à partir duquel $P^i_{m-\varepsilon,m+\varepsilon}$ est aussi voisin de l'unité qu'on le désire, ε étant choisi arbitrairement petit. C'est ce que nous appellerons des estimations correctes, ou selon l'expression de R.-A. Fischer des statistiques « consistent ». On peut traduire ce fait en termes physiques, en disant qu'elles sont dépourvues d'erreur systématique et, en langage du Calcul des probabilités, en parlant de la convergence en probabilité de ces fonctions vers m, grandeur dont on cherche l'estimation.

Les autres seront toutes celles ne jouissant pas de cette propriété, ce qui peut se produire de deux manières différentes :

Ou bien il y a convergence en probabilité, mais vers une autre valeur que m, soit une fonction de m, soit même vers un nombre indépendant de m (dans ce cas il y a erreur systématique), ou bien il n'y a pas convergence en probabilité.

Un exemple classique de ce dernier cas est fourni par la loi de Cauchy. On sait que la moyenne arithmétique de n variables aléatoires soumises à cette loi, suit encore la même loi avec les mêmes coefficients : prendre $\frac{\Sigma x_i}{n}$ comme « statistique » revient donc à choisir l'une des variables x_i, au hasard.

Les deux premiers Chapitres de ce mémoire sont consacrés à la recherche des estimations correctes dans le cas de un ou plusieurs paramètres.

Nous nous sommes, en particulier, efforcé de mettre en évidence les conditions suffisantes sous lesquelles l'estimation par la méthode du maximum de vraisemblance (maximum of likelihood) converge en probabilité vers m.

Ce fait et un certain nombre d'autres propriétés du maximum de vraisemblance avaient été énoncés par R.-A. Fisher.

Le Chapitre III est occupé par l'étude des lois limites suivies par ces fonctions statistiques. Il s'agit d'exprimer analytiquement le comportement de ce que nous avons appelé l'infiniment petit aléatoire en $\frac{1}{n}$: $f_n - m$. Sous une condition de plus que celles qui sont mises en évidence dans les deux premiers Chapitres, à savoir l'existence, quelles que soient les limites de la variable aléatoire de la valeur moyenne de $\left[\frac{\partial \log f(x, m)}{\partial m}\right]^2$, [$f(x, m)$ étant la densité de probabilité], nous pouvons montrer que l'estimation de R.-A. Fisher suit à la limite une loi de Gauss de la forme

$$\frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-\frac{(\xi - m)^2}{2\sigma^2}\right],$$

et nous donnons une expression de σ.

Mr. J.-L. Doob avait donné de ce théorème, par des voies différentes, une démonstration moins étendue.

Une méthode absolument analogue à celle suivie par le maximum de vraisemblance nous permet d'obtenir les statistiques gaussiennes à la limite. Nous rejoignons ici et nous donnons une extension de certains travaux récents de MM. Paul Lévy et Richard von Mises.

L'étude des limites des valeurs moyennes de φ_n et de φ_n^2 permet de donner une solution aux équations de Fredholm et de Volterra de première espèce, à une infinité de variables dans le cas où le noyau se présente sous une forme simple.

Il est bien évident que toutes ces estimations ne se valent pas.

Dans le cas où la limite est une loi de Gauss, l'estimation la plus précise sera celle dont l'écart-type est le plus petit. Nous sommes arrivé au résultat suivant, qui avait été prévu par R.-A. Fisher et qui est exposé dans le Chapitre IV :

Toutes les estimations dont la loi-limite est de Gauss et telles que leur écart-type converge vers l'écart-type limite, ont une précision bornée supérieurement par celle de la méthode du *maximum of likelihood*.

Cet énoncé peut s'étendre au cas où plusieurs paramètres sont à estimer. Il est alors susceptible d'une interprétation géométrique fort simple. Ces résultats valent à la fois dans le cas d'expériences indépendantes les unes des autres, et dans le cas où les résultats sont liés en chaînes simples ou multiples. Cette généralisation est fondée sur des travaux de Markoff et de M. B. Hostinsky.

Il est également possible de donner diverses extensions dans le cas où la loi-limite n'est pas une loi de Gauss. Comme dans ce cas, il peut ne pas y avoir d'écart-type, les théorèmes porteront sur ce que dans des recherches récentes on appelle la « dispersion » des variables aléatoires.

On peut démontrer que, pour chaque type de loi-limite, la dispersion est inférieurement bornée.

Dans le Chapitre suivant sont étudiées les lois de probabilité liée des différentes estimations d'un même paramètre. Soient deux suites de statistiques f_n et g_n, toutes deux correctes au sens que nous avons précisé, de m. Si l'on suppose que f_n a une valeur numérique comprise entre deux nombres α et β, il s'ensuit que $x_1, \ldots, x_n$ doivent être choisis entre certaines limites, et par conséquent la loi de probabilité de g_n doit en être modifiée.

Une des manières les plus simples d'observer cette modification est de calculer la valeur moyenne du produit $f_n g_n$ ou coefficient de corrélation. Elle est insuffisante en général, mais suffisante si f_n et g_n suivent à la limite une loi de Gauss à deux variables.

On arrive ainsi à mettre en évidence une inégalité limite valable pour n infini dans laquelle entrent, outre cette grandeur, les deux écarts-types limites de f_n et g_n.

On en déduit que si ces deux écarts-types sont égaux à l'écart-type minimum, le coefficient de corrélation est égal à l'unité, ce qui implique que f_n et g_n ne doivent différer que sur un ensemble de probabilité nulle.

En résumé, sous des conditions assez peu restrictives qui seront exposées au long de cette thèse, la statistique du maximum de vraisemblance :

1° converge en probabilité vers le paramètre m;

2° a une loi de répartition-limite de Gauss;

3° est la plus précise de toutes les estimations gaussiennes;

4° elle est la seule à jouir de cette propriété du maximun de précision.

Ces résultats avaient été prévus sans démontration complète par R.-A. Fisher, qui avait, avec raison, on le voit, donné à cette statistique le nom « d'optimum ».

Enfin, prend place l'examen des lois de probabilité permettant ce que M. Georges Darmois a appelé des statistiques exhaustives (« sufficient » selon l'expression des probabilistes anglais).

Il a été possible de faire une étude approfondie de ces différents cas, et l'on vérifie que ces estimations fournissent le maximum de précision.

Cet énoncé comporte une réciproque assez curieuse, en ce sens que, pour une loi quelconque, l'estimation par la méthode du maximum de vraisemblance est à la limite une statistique exhaustive. Toutefois, elle n'a pas la propriété d'extraction *totale* de l'information, elle n'extrait que la partie principale.

Je désire en terminant cette introduction exprimer ici ma reconnaissance à M. Georges Darmois qui m'a indiqué le sujet de ce mémoire et a si aimablement dirigé mon travail, à M. Émile Borel qui a bien voulu présenter à l'Académie des Sciences des notes résumant mes résultats, et à M. Arnaud Denjoy qui a bien voulu se joindre à eux pour constituer le jury.

Qu'il me soit permis de remercier également l'Administration du *Journal de l'École Polytechnique* qui a bien voulu accepter de publier cet article.

NOTATIONS.

Nous appellerons dans la suite $E(x)$ (espérance mathématique) la valeur moyenne de la variable aléatoire x, soit $\int x\,dF(x)$. Si m est sa valeur, on appellera écart-type σ de la variable x la quantité $\sqrt{E(x-m)^2}$. La probabilité pour que la valeur absolue de $x - m$ dépasse $t\sigma$ est inférieure à $\frac{1}{t^2}$ (théorème de Tchebycheff). La fonction caractéristique $\varphi(t)$ relative à la loi de probabilité totale $F(x)$ est $\int e^{itx} dF(x)$.

On voit facilement qu'à toute loi de probabilité correspond une fonction caractéristique (la réciproque n'est pas exacte). Étant donnée une fonction caractéristique, la loi de probabilité est donnée par la formule de réciprocité de Fourier :

$$F(x) - F(0) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \varphi(t) \frac{1 - e^{-itx}}{it} dt.$$

Si $\varphi_1(t)$ et $\varphi_2(t)$ sont les fonctions caractéristiques de deux variables indépendantes x_1 et x_2, la fonction caractéristique de $x_1 + x_2$ est $\varphi_1(t)\varphi_2(t)$; c'est une propriété fondamentale.

CHAPITRE I.

CONVERGENCE EN PROBABILITÉ. STATISTIQUES "CONSISTENT".

Nous nous occuperons tout d'abord du cas où l'on a des variables aléatoires indépendantes les unes des autres et dont les bornes sont des nombres fixes, certains et connus (pour plus de généralité nous les supposerons infinis).

Tout d'abord, rappelons un théorème dont la démonstration a été donnée par M. Khintchine (*Comptes rendus de l'Académie des Sciences*, séance du 11 février 1929) :

Étant donnée une variable aléatoire x dont la valeur moyenne a une existence bien déterminée a, la probabilité pour que $\frac{\Sigma x_i}{n} - a$ soit en module supérieure à ε tend vers o quand n croît indéfiniment, ε étant un nombre positif fixé d'avance.

Cette convergence en probabilité de $\frac{\Sigma x_i}{n}$ vers a, pour employer l'expression utilisée par M. Fréchet (Metron 15-VI-1930), était connue depuis Tchebycheff, qui supposait toutefois l'existence du moment d'ordre 2 : $E(x^2)$. La nouveauté de l'énoncé de M. Khintchine est qu'il ne suppose que l'existence de $E(x)$.

La démonstration est fondée sur le fait que lorsque dans tout intervalle fini la fonction caractéristique converge uniformément vers la fonction e^{ita}, la variable aléatoire converge en probabilité vers a. La fonction caractéristique de $\frac{\Sigma x_i}{n}$ étant $\left[\varphi\left(\frac{t}{n}\right)\right]^n$ si $\varphi(t)$ est celle de x, il suffit de démontrer que $n \log \varphi\left(\frac{t}{n}\right)$ tend vers iat uniformément dans tout intervalle borné de t quand n croît indéfiniment; ou en posant l'égalité $\varphi\left(\frac{t}{n}\right) = 1 + \psi\left(\frac{t}{n}\right)$ que $n\psi\left(\frac{t}{n}\right) \to iat$ dans les mêmes conditions.

$$\psi\left(\frac{t}{n}\right) = \int_{-\infty}^{+\infty} \left(e^{i\frac{t}{n}x} - 1\right) dF(x).$$

Partageons cette intégrale I en trois parties I_1 de $-\infty$ à $-\varepsilon n$, I_2 de $-\varepsilon n$ à $+\varepsilon n$ et I_3 de $+\varepsilon n$ à $+\infty$.

$$|I_1| = \left|\int_{-\infty}^{-\varepsilon n} \left(e^{i\frac{t}{n}x} - 1\right) dF(x)\right|$$
$$< \int_{-\infty}^{-\varepsilon n} \left|e^{i\frac{t}{n}x} - 1\right| dF(x) < 2\int_{-\infty}^{-\varepsilon n} dF(x) < -\frac{2}{\varepsilon n}\int_{-\infty}^{-\varepsilon n} x\, dF(x).$$

Donc, $n|I_1|$ est inférieur à $-\frac{2}{\varepsilon}\int_{-\infty}^{-\varepsilon n} x\, dF(x)$. Comme on a supposé l'existence de $E(x)$, il en résulte que ε étant fixé, on peut prendre n assez grand pour que $n|I_1|$ soit inférieur à un nombre quelconque $\frac{\eta}{3}$, par exemple. La même démonstration s'applique à $n|I_3|$.

Examinons la valeur limite de I_2. On sait que

$$\left|-\int_{-\varepsilon n}^{+\varepsilon n} i\frac{t}{n} x\, dF(x) + \int_{-\varepsilon n}^{+\varepsilon n} \left(e^{i\frac{t}{n}x} - 1\right) dF(x)\right|$$
$$< \int_{-\varepsilon n}^{+\varepsilon n} \left|e^{i\frac{t}{n}x} - 1 - i\frac{t}{n}x\right| dF(x) < C\int_{-\varepsilon n}^{+\varepsilon n} \frac{t^2}{n^2} x^2\, dF(x),$$

C étant une constante déterminée quand on a fixé les bornes de variation de t. Or,

$$\frac{Ct^2}{n^2}\int_{-\varepsilon n}^{+\varepsilon n} x^2\, dF(x) < \frac{Ct^2}{n^2}\varepsilon n\int_{-\varepsilon n}^{+\varepsilon n} |x|\, dF(x) = \frac{Ct^2}{n}\varepsilon\int_{-\varepsilon n}^{+\varepsilon n} |x|\, dF(x).$$

D'après la convergence de $E(x)$ l'intégrale est bornée. Donc

$$\left|nI_2 - it\int_{-\varepsilon n}^{+\varepsilon n} x\, dF(x)\right| < K\varepsilon$$

dans tout intervalle borné de t. On peut donc choisir ε de manière que le premier membre de l'inégalité soit inférieur à $\frac{\eta}{3}$. Ayant ensuite choisi $n > N$ de façon que $n|I_1|$ et $n|I_3|$ soient inférieurs également à $\frac{\eta}{3}$, il s'ensuit que l'on a :

$$n|I_1| + n|I_3| + \left| nI_2 - it\int_{-\varepsilon n}^{+\varepsilon n} x\,dF(x) \right| < \eta,$$

et, par conséquent,

$$\left| nI - it\int_{-\varepsilon n}^{+\varepsilon n} x\,dF(x) \right| < \eta.$$

En prenant encore $n > N'$ d'après les hypothèses $\int_{-\varepsilon n}^{+\varepsilon n} x\,dF(x)$ est aussi voisin de a qu'on le désire. Par conséquent ε étant choisi et n supérieur au plus grand des deux nombres N et N′,

$$|nI - iat| < \eta'.$$

Par conséquent, nI converge uniformément vers iat dans tout intervalle où t est borné, ce qui était nécessaire pour établir la proposition. Remarquons que cette démonstration exige que les lois de probabilité de toutes les variables x_i soient identiques. Nous donnerons plus loin une généralisation, sous certaines conditions, de ce résultat.

Considérons la loi de probabilité élémentaire $f(x, m)$ avec $\int f(x, m)\,dx = 1$. Admettons que l'on puisse dériver sous le signe $\int$, ce qui exige les conditions suivantes pour la fonction $f(x, m)$:

1° $f(x, m)$ doit avoir une dérivée par rapport à m dans tout l'intervalle de variation de x *sauf peut-être pour un ensemble de probabilité nulle*. Cette dernière restriction rend valables les résultats que nous allons énoncer pour des lois du type $e^{-|x-m|}\,dx$ (première loi de Laplace).

Cette condition suffit dans le cas où x est compris entre deux limites finies. Sinon on doit en plus supposer que :

2° $\int_0^C \frac{\partial f(x, m)}{\partial m}\,dx$ converge uniformément vers sa limite quand $|C|$ croît indéfiniment, ceci dans tout l'intervalle où m peut être choisi.

Sous ces conditions on a évidemment

$$\int \frac{\partial f(x, m)}{dm}\,dx = 0,$$

ce qui peut s'écrire

$$E\left[\frac{1}{f(x, m)}\,\frac{\partial f(x, m)}{\partial m}\right] = 0.$$

D'après le théorème précédent il en résulte que

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{f(x_i, m)}\frac{\partial f(x_i, m)}{\partial m}$$

converge en probabilité vers zéro quand n croît indéfiniment.

Soit maintenant l'équation en T

$$\sum_{i=1}^{n}\frac{1}{f(x_i, \mathrm{T})}\frac{\partial f(x_i, \mathrm{T})}{\partial m}=0.$$

Étudions la probabilité pour que $|\mathrm{T}-m|<\varepsilon$ quand n croît indéfiniment. Supposons que $\frac{1}{f(x, m)}\frac{\partial f(x, m)}{\partial m}$ soit *uniformément* continue par rapport à m au point m, *pour toutes les valeurs de x*. Il en résulte que l'on peut trouver m tel que $|\mathrm{T}-m|$ soit inférieur à ε avec

$$\left|\frac{1}{f(x_i, m)}\frac{\partial f(x_i, m)}{\partial m}-\frac{1}{f(x_i, \mathrm{T})}\frac{\partial f(x_i, \mathrm{T})}{\partial m}\right|<\eta,$$

ceci pour toute valeur de x_i. On aura donc

$$\left|\frac{1}{n}\sum_{i=1}^{n}\frac{1}{f(x_i, m)}\frac{\partial f(x_i, m)}{\partial m}\right|<\frac{1}{n}\sum_{i=1}^{n}\left|\frac{1}{f(x_i, m)}\frac{\partial f(x_i, m)}{\partial m}-\frac{1}{f(x_i, \mathrm{T})}\frac{\partial f(x_i, \mathrm{T})}{\partial m}\right|<\eta.$$

Réciproquement, mais à la condition de plus que

$$\sum_{i=1}^{n}\frac{1}{f(x_i, m)}\frac{\partial f(x_i, m)}{\partial m}=0$$

n'ait qu'une racine en m (ou dans l'intervalle de recherche), la condition

$$\left|\frac{1}{n}\sum_{i=1}^{n}\frac{1}{f(x_i, m)}\frac{\partial f(x_i, m)}{\partial m}\right|<\eta \tag{1}$$

entraîne que la racine T de l'équation

$$\sum_{i=1}^{n}\frac{1}{f(x_i, \mathrm{T})}\frac{\partial f(x_i, \mathrm{T})}{\partial m}=0$$

est telle que $|\mathrm{T}-m|<\varepsilon$.

La probabilité de cette dernière inégalité est donc celle de l'inégalité (1).

Elle peut par conséquent être rendue aussi voisine de l'unité qu'on le désire.

On peut encore exclure de toutes les valeurs de x $\left(\text{ensemble sur lequel } \frac{1}{f(x, m)} \frac{\partial f(x, m)}{\partial m} \text{ doit être uniformément continue}\right)$ un ensemble de probabilité nulle; ce qui permet d'étendre le résultat indiqué à la loi de Laplace.

Le théorème de M. Khintchine n'exige aucune hypothèse sur la continuité de $F(x)$. Ses résultats peuvent par conséquent être appliqués au cas des probabilités discontinues. Dans ce cas, encore la racine de l'équation

$$\sum_{i=1}^{n} \frac{1}{p_i(T)} \frac{\partial p_i(T)}{\partial m} = 0$$

converge en probabilité vers m quand n croît indéfiniment sous les conditions indiquées, $p_i(m)$ étant la probabilité qu'a une variable aléatoire susceptible de prendre les valeurs $x_1, \ldots, x_i, \ldots$, d'être égale à x_i. Ces résultats peuvent s'exprimer simplement si l'on considère l'espace à une infinité de dimensions dont il a déjà été question. Les théorèmes que l'on vient de formuler expriment que ce que l'on peut appeler le *voisinage* de la surface $T(x_i, \ldots, x_n, \ldots) = m$ de cet espace occupe *presque* tout cet espace. Cette propriété assez curieuse de la mesure dans l'espace fonctionnel est d'ailleurs connue.

Nous venons de mettre en évidence une généralisation du théorème des grands nombres formulé par Bernoulli. Il est facile, en effet, de voir que la loi de Cauchy elle-même n'y échappe pas puisque

$$\frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{2(x-m)}{1+(x-m)^2} \frac{dx}{1+(x-m)^2}$$

est convergent.

Supposons maintenant que m soit une des limites d'intégration. On aura donc $\int_a^m f(x, m)\,dx = 1$. Si, en plus de la première condition fixée plus haut $\left(\text{dérivabilité sous le signe } \int\right)$, $f(x, m)$ est continue par rapport à x pour $x = m$, on a

$$f(m, m) + \int_a^m \frac{\partial f(x, m)}{\partial m} dx = 0$$

ou

$$E\left[f(m, m) + \frac{1}{f(x, m)} \frac{\partial f(x, m)}{\partial m}\right] = 0.$$

Il est bien évident que nous avons laissé de côté les lois de la forme $\frac{f(x)}{F(m)}$ avec $F(x)=\int_a^x f(x)\,dx$ $\left(\text{dans lesquelles la parenthèse s'écrit } \frac{f(m)}{F(m)} - \frac{f(m)}{F(m)}\right)$. Leur étude sera faite plus tard.

De la même façon que dans le cas précédent, on voit que la racine T de l'équation

$$(2) \qquad f(T, T) + \frac{1}{n}\sum_{i=1}^{n} \frac{1}{f(x_i, T)} \frac{\partial f(x_i, T)}{\partial m} = 0$$

converge en probabilité vers m à la condition que :

1° Comme plus haut $\frac{1}{f(x_i, m)} \frac{\partial f(x_i, m)}{\partial m}$ soit uniformément continue par rapport à m;

2° $f(m, m)$ soit continue, ce qui est déjà réalisé.

Il faut, de plus, que la racine de l'équation (2) soit uniforme tout au moins dans l'intervalle de recherche.

Nous avons tout à l'heure fait remarquer que le théorème de M. Khintchine suppose que toutes les variables suivent la même loi de probabilité. Il est possible de s'affranchir de cette restriction. Toutefois, il est indispensable de faire sur les lois suivies par les variables aléatoires certaines hypothèses. En reprenant les notations du début, on aura à étudier

$$\sum_{i=1}^{n} I_1^i + \sum_{i=1}^{n} I_2^i + \sum_{i=1}^{n} I_3^i = \sum_{i=1}^{n} \varphi_i\left(\frac{t}{n}\right)$$

au lieu de

$$nI_1 + nI_2 + nI_3.$$

En faisant les mêmes majorations que ci-dessus, il apparaît que

$$|\Sigma I_1^i| < \Sigma |I_1^i| < -\frac{2}{\varepsilon n}\Sigma\int_{-\infty}^{-\varepsilon n} x_i\, dF_i(x_i).$$

On voit que si tous les $E(x_i)$ convergent *également* vers leurs limites respectives, on pourra trouver N tel que pour tout $n > N$, ε étant fixé $\int_{-\infty}^{-\varepsilon n} x_i dF_i(x_i)$ soit inférieur à α quel que soit i, et par conséquent $|\Sigma I_1^i|$ sera inférieur à $\frac{2}{\varepsilon}\alpha$ qui pourra être rendu inférieur à $\frac{\eta}{3}$.

De même pour $|\Sigma \mathrm{I}_3^i|$.

On aura encore

$$\left| \mathrm{I}_2^i - i\frac{t}{n}\int_{-\varepsilon n}^{+\varepsilon n} x_i \, d\mathrm{F}_i(x_i) \right| < \mathrm{C}\frac{t^2}{n}\varepsilon \int_{-\varepsilon n}^{+\varepsilon n} |x_i| \, d\mathrm{F}_i(x_i),$$

et évidemment à cause de l'égale convergence des $\mathrm{E}(x_i)$

$$\left| \left| \Sigma \mathrm{I}_2^i - it\frac{\Sigma a_i}{n} \right| - t\zeta \right| < \Sigma \left| \mathrm{I}_2^i - i\frac{t}{n}\int_{-\varepsilon n}^{+\varepsilon n} x_i \, d\mathrm{F}_i(x_i) \right| < \mathrm{C}\frac{t^2}{n}\varepsilon \Sigma \int_{-\varepsilon n}^{+\varepsilon n} |x_i| \, d\mathrm{F}_i(x_i),$$

ζ étant un nombre qui peut être rendu aussi petit qu'on le veut et qui provient de la différence $\mathrm{E}(x_i) - \int_{-\varepsilon n}^{+\varepsilon n} x_i d\mathrm{F}_i(x_i)$. Les $\mathrm{E}(|x_i|)$ existent tous d'après l'hypothèse initiale. Si, de plus, ils sont tels que $\frac{\Sigma \mathrm{E}(|x_i|)}{n}$ soit borné (il suffit par exemple que tous les $\mathrm{E}(|x_i|)$ soient inférieurs à un nombre fini, mais ce n'est pas nécessaire). $\left| \left| \Sigma \mathrm{I}_2^i - it\frac{\Sigma a_i}{n} \right| - t\zeta \right|$ peut être rendu inférieur à $\frac{\eta'}{3}$ et dans tout intervalle où t est borné on en conclura que $\left| \Sigma \mathrm{I}_2^i - it\frac{\Sigma a_i}{n} \right|$ est inférieur à $\frac{\eta}{3}$ et comme plus haut que $\left| \Sigma \mathrm{I}^i - it\frac{\Sigma a_i}{n} \right|$ converge vers zéro d'une manière uniforme partout où t est fini.

On peut donc affirmer que la variable aléatoire $\frac{\Sigma x_i}{n}$ converge vers $\frac{\Sigma a_i}{n}$. Si $\frac{\Sigma a_i}{n}$ converge au sens habituel du mot vers un nombre A $\left(\frac{\Sigma a_i}{n}\right.$ étant au plus égal à $\frac{\Sigma \mathrm{E}(|x_i|)}{n}$ doit d'ailleurs être borné$\Big)$, $\frac{\Sigma x_i}{n}$ converge en probabilité vers ce nombre fixe A.

Il est possible après ces démonstrations d'étendre le théorème des grands nombres à divers cas.

Soient $u_i(x, m)$ des fonctions telles que

$$\int u_i(x, m) f(x, m) \, dx = 0.$$

1° Si toutes ces intégrales convergent également vers zéro [il suffit pour cela que l'on puisse trouver X tel que pour $|x| > \mathrm{X}$ tous les $u_i(x, m)$ soient en valeur absolue inférieurs à $\varphi(x)$, l'intégrale $\int_{-\infty}^{+\infty} \varphi(x) f(x, m)\, dx$ étant convergente];

2° Si l'ensemble dc tous les $\int |u_i(x, m)|\, f(x, m)\, dx$ est borné on en déduit que $\frac{1}{n} \Sigma u_i(x_i, m)$ converge en probabilité vers zéro et comme précédemment, mais à la condition que tous les $u_i(x, m)$ soient *également* uniformément continues au point m que la racine en T de $\Sigma u_i(x_i, \mathrm{T}) = 0$ converge en probabilité vers zéro.

Supposons que l'on ait $u_i(x, m) = u_i(x) - \varphi_i(m)$, on a donc

$$\int u_i(x) f(x, m)\, dx = \varphi_i(m).$$

Si les $u_i(x)$ satisfont aux conditions précédentes qui sont notablement simplifiées (les questions d'uniformité de la continuité en m par rapport à x ne se posent plus) $\frac{1}{n} \Sigma u_i(x_i) - \frac{1}{n} \Sigma \varphi_i(m)$ converge en probabilité vers zéro, et l'on en déduira une statistique *consistent*.

On voit que par ce procédé on peut former une infinité d'estimations correctes. Un cas particulier intéressant comme forme est celui où tous les φ_i sont indentiques. Les u_i le seront aussi. On a une estimation correcte en prenant la racine (il faut qu'il n'y en ait qu'une) de l'équation

$$\varphi(\mathrm{T}) = \frac{\Sigma u(x_i)}{n},$$

soit

$$\mathrm{T} = \varphi^{-}\left(\frac{\Sigma u(x_i)}{n}\right),$$

φ^{-} désignant la fonction inverse.

CHAPITRE II.

DIVERS COMPLÉMENTS DES THÉORÈMES PRÉCÉDENTS.

Appliquons les résultats obtenus au cas où les variables se dispersent autour du paramètre suivant une loi de Gauss de précision $\frac{1}{\sigma}$. On a donc ici

$$f(x, m) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-\frac{(x-m)^2}{2\sigma^2}\right].$$

La solution de l'équation du maximum de vraisemblance est alors

$$T_l = \frac{\Sigma x_i}{n}.$$

C'est le procédé que l'on emploie généralement pour estimer une grandeur dont on connaît les mesures. Si l'on veut connaître σ il faudra former $\sqrt{\frac{\Sigma(x_i - m)^2}{n}}$ (on suppose alors que l'on sait la valeur de m). Dans le cas de la loi de Laplace $\frac{1}{2}e^{-|x-m|}$, on a alors à prendre la médiane de n observations. Pour le cas de la loi de Cauchy, la solution ne se présente pas sous une forme simple. On aura à résoudre l'équation $\sum \frac{x_i - T_l}{1+(x_i - T_l)^2} = 0$ (T_l ne peut être obtenu que par approximations successives). Nous verrons plus loin une manière un peu moins compliquée d'arriver à ce résultat. Il est facile de voir que dans tous ces cas les conditions que nous avons exigées sont satisfaites. Ces statistiques convergent donc en probabilité vers m.

Nous allons maintenant présenter diverses extensions des théorèmes du Chapitre précédent. Tout d'abord nous avons supposé que l'équation $\sum \frac{1}{f}\frac{\partial f}{\partial m}(x_i, T_l) = 0$ n'a qu'une racine en T_l. Il est possible de remplacer cette condition par une autre peut-être un peu plus restrictive dans la forme, mais néanmoins assez souvent remplie et dont il est facile de s'assurer si une loi de probabilité la remplit ou non. Sous des restrictions énoncées au Chapitre I (sauf celle qui est en question) nous avons établi que

$$\frac{1}{n}\sum \frac{1}{f}\frac{\partial f}{\partial m}(x_i, T) = L(T),$$

une fois fixés $x_1, x_2, \ldots, x_n$, est une fonction continue de T. Il suffit donc, pour que l'équation $L(T) = 0$ ait une seule racine en T au voisinage de m, que $L(T)$ ait une dérivée qui soit d'un signe constant dans ce voisinage. Cette dérivée sera

$$\frac{1}{n}\sum \left[\frac{1}{f}\frac{\partial^2 f}{\partial m^2} - \left(\frac{1}{f}\frac{\partial f}{\partial m}\right)^2\right](x_i, T),$$

si $\frac{\partial f}{\partial m}$ est dérivable par rapport à m quel que soit x. Supposons cette fonction continue pour T voisin de m. Si $\int \frac{\partial^2 f}{\partial m^2}dx = 0$ $\Big($même condition de limite uniformément atteinte que dans le Chapitre I, pour l'intégrale $\int_0^t \frac{\partial^2 f}{\partial m^2}dx\Big)$

et si $\mathrm{E}\left[\frac{1}{f}\frac{\partial f}{\partial m}\right]^2$ existe, il en résulte que $\frac{d\mathrm{L}}{d\mathrm{T}}$ converge en probabilité vers $-\mathrm{E}\left[\frac{1}{f}\frac{\partial f}{\partial m}\right]^2+\eta$ (η étant petit si T reste dans le voisinage de m). Donc si l'on reste dans le voisinage de m il y a une chance aussi faible qu'on le désire que $\mathrm{L}(\mathrm{T})=0$ ait plus d'une racine; comme la dérivée seconde de $\Sigma\log f(x_i, m)$ par rapport à m converge en probabilité vers un nombre négatif pour la racine en m de la dérivée première, on voit que $\Sigma\log f(x_i, m)$ passera très probablement par son maximum pour cette valeur.

Tous les résultats exposés dans le premier Chapitre sont fondés sur le fait que si $\mathrm{E}(x)$ a une valeur déterminée, $\frac{\Sigma x_i}{n}$ converge en probabilité vers cette valeur. Il nous a été possible de donner une condition moins restrictive à cette convergence. Cette condition est d'ailleurs non seulement suffisante mais encore nécessaire. En effet, pour que $\frac{\Sigma x_i}{n}$ converge en probabilité vers a, il est nécessaire et suffisant que la fonction caractéristique de cette variable aléatoire, soit $\left[\varphi\left(\frac{t}{n}\right)\right]^n$ converge uniformément dans tout intervalle borné de t vers e^{iat}. En écrivant $\varphi(t)$ sous la forme $\exp\{\psi(t)\}$, nous voyons que $n\psi\left(\frac{t}{n}\right)$ doit tendre uniformément vers iat (avec $|t|<\mathrm{T}$), la condition étant suffisante et par conséquent il est nécessaire et suffisant pour la convergence demandée que $\frac{\psi\left(\frac{t}{n}\right)}{\frac{t}{n}}$ tende vers ia uniformément pour t compris en module entre ε et T (ε étant arbitrairement petit et T arbitrairement grand) pour des valeurs de n augmentant indéfiniment.

Comme $\psi(0)$ est nul, cette condition est équivalente à l'existence à l'origine d'une dérivée de $\psi(t)$, donc de $\varphi(t)$ égale à ia.

Si la variable aléatoire admet une valeur moyenne on voit facilement que cette dérivée existe, mais la réciproque n'est pas exacte. L'exemple suivant le prouve; soit X une variable aléatoire susceptible de prendre toutes les valeurs extérieures à l'intervalle $-2, +2$, avec la probabilité $\frac{\mathrm{K}\,dx}{x^2[\log|x|]^\alpha}$, K étant une constante telle que $\mathrm{K}\int_{\mathrm{E}}\frac{dx}{x^2[\log|x|]^\alpha}$ soit égal à l'unité (¹). Si $\alpha\leqq 1$ on voit que $\mathrm{E}(x)$

(¹) Le symbole $\int_{\mathrm{E}}$ représente l'intégrale étendue à tout l'ensemble sur lequel la variable aléatoire est définie.

n'existe pas, l'intégrale $\int_{E}^{+\infty} \frac{dx}{x[\log|x|]^{\alpha}}$ étant divergente. En raison de la symétrie de la loi de probabilité élémentaire la fonction caractéristique sera égale à $2\,K\int_{2}^{+\infty} \frac{\cos tx\,dx}{x^2[\log x]^{\alpha}}$. S'il y a une dérivée à l'origine, cette dérivée doit être nulle par raison de symétrie. Il suffit donc de démontrer que

$$\lim_{\Delta t=0}\int_{2}^{+\infty}\frac{\cos\Delta t x - 1}{\Delta t}\,\frac{dx}{x^2(\log x)^{\alpha}} = 0.$$

Nous partageons cette intégrale en deux parties I_1 de 2 à $\frac{2\pi}{\Delta t}$ et I_2 de $\frac{2\pi}{\Delta t}$ à $+\infty$.

$$|I_2| = \int_{\frac{2\pi}{\Delta t}}^{+\infty}\left|\frac{\cos\Delta t x - 1}{\Delta t}\right|\frac{dx}{x^2(\log x)^{\alpha}} < \frac{2}{\Delta t}\int_{\frac{2\pi}{\Delta t}}^{+\infty}\frac{dx}{x^2(\log x)^{\alpha}} = \frac{Y}{\pi}\int_{Y}^{+\infty}\frac{dx}{x^2(\log x)^{\alpha}}.$$

En intégrant de Y à hY, ..., $h^{n-1}Y$ à h^nY, ... (avec $h > 1$), il est facile de trouver une majoration de cette intégrale

$$Y\int_{Y}^{+\infty}\frac{dx}{x^2(\log x)^{\alpha}} = Y\sum_{n=1}^{\infty}\int_{h^{n-1}Y}^{h^nY}\frac{dx}{x^2(\log x)^{\alpha}}$$

$$< Y\sum_{n=1}^{\infty}\frac{(h-1)h^{n-1}Y}{(h^{n-1}Y)^2(\log h^{n-1}Y)^{\alpha}} = (h-1)\sum_{n=1}^{\infty}\frac{1}{h^{n-1}(\log h^{n-1}Y)^{\alpha}}.$$

Cette quantité étant inférieure à $\frac{h-1}{(\log Y)^{\alpha}}\sum_{n=1}^{\infty}\frac{1}{h^{n-1}}$ tend vers 0 avec $\frac{1}{Y}$.

De même I_1 : on a, avec le même procédé de majoration,

$$|I_1| = \int_{2}^{\frac{2\pi}{\Delta t}}\left|\frac{\cos\Delta t\,x - 1}{\Delta t}\right|\frac{dx}{x^2(\log x)^{\alpha}}$$

$$< \frac{\Delta t}{2}\int_{2}^{\frac{2\pi}{\Delta t}}\frac{dx}{(\log x)^{\alpha}} = \frac{\pi}{Y}\int_{2}^{Y}\frac{dx}{(\log x)^{\alpha}} = \frac{\pi}{2h^n}\sum_{i=1}^{i=n}\int_{2h^{i-1}}^{2h^i}\frac{dx}{(\log x)^{\alpha}}.$$

Et

$$\frac{1}{2h^n}\sum_{i=1}^{n}\int_{2h^{i-1}}^{2h^i}\frac{dx}{(\log x)^{\alpha}} < \frac{1}{2h^n}\sum_{i=1}^{n}\frac{2(h-1)h^{i-1}}{(\log 2h^{i-1})^{\alpha}}$$

$$< \frac{h-1}{h^n}\left[\frac{1}{(\log 2)^{\alpha}} + \sum_{i=2}^{n}\frac{h^{i-1}}{(i-1)^{\alpha}(\log h)^{\alpha}}\right].$$

Cette quantité tendra vers o si $\frac{1}{h^n}\sum_{i=1}^{n=1}\frac{h^i}{i^\alpha}$ tend vers o. Il est facile de s'en assurer. En effet, en posant $n = p + q$

$$\frac{1}{h^n}\sum_{i=1}^{n-1}\frac{h^i}{i^\alpha} < \frac{1}{h^{p+q}}\left[\sum_{i=1}^{p-1} h^i + \sum_{i=p}^{n-1}\frac{h^i}{p^\alpha}\right] = \frac{1}{h-1}\left[\frac{h^p-1}{h^{p+q}} + \frac{h^p}{p^\alpha}\,\frac{h^q-1}{h^{p+q}}\right].$$

Si l'on s'arrange pour que p et q tendent tous deux vers l'infini, on a démontré que I_1 tend vers o. L'existence d'une dérivée de la fonction caractéristique est donc établie, bien que la loi de probabilité n'ait pas de premier moment. On peut, en vertu de la remarque faite sur la convergence en probabilité, remplacer dans le cas de l'étude du maximum of likelihood la condition de dérivabilité sous le signe d'intégration par l'existence pour $t = 0$ de la dérivée de la fonction

$$\int \exp\left\{ it\frac{1}{f}\,\frac{\partial f}{\partial m}\right\} f\,dx.$$

C'est une condition évidemment compliquée, mais moins restrictive que la précédente. Revenons à l'étude de cette dérivabilité. Comme il s'agit d'une convergence d'intégrale il est à prévoir que l'on ne pourra pas donner de critères nécessaires et suffisants d'après le théorème de du Bois-Reymond. La dérivabilité à l'origine de la fonction caractéristique est liée au comportement à l'infini de la loi de probabilité; la transformation de Fourier $\int e^{itx}\,dF(x)$ remplit en effet dans le domaine fonctionnel le même rôle que l'inversion en géométrie. Nous considérerons tout d'abord des lois de probabilité $p(x)\,dx$ de la forme $\frac{K\,dx}{x^\alpha}$ $(\alpha > 1)$ pour x suffisamment grand en module. Nous aurons donc à examiner les deux intégrales

$$I = \int_A^{+\infty}\frac{1-\cos tx}{t}\,\frac{dx}{x^\alpha} \qquad \text{et} \qquad J = \int_A^{+\infty}\frac{\sin tx}{t}\,\frac{dx}{x^\alpha},$$

quand t tend vers o (on peut se borner au cas de t positif en raison de la symétrie).

Le changement de variable $tx = u$ permet de constater le résultat suivant. Si $\alpha < 2$, I croît indéfiniment, si $\alpha = 2$, I a une limite non nulle et si $\alpha > 2$, I a une limite nulle. J est infini à la limite quand α est inférieur ou *égal* à 2 et nul à la limite quand α est supérieur à 2. On voit par conséquent que

$$\int \cos txp(x)\,dx$$

a une dérivée nulle si $\alpha > 2$ deux nombres dérivés égaux en valeur absolue et de signe contraire si $\alpha = 2$ et une dérivée infinie si $\alpha < 2$.

Mais $\int \sin txp(x)\,dx$ n'est dérivable au sens précis du terme que s'il y a un moment d'ordre un ($\alpha > 2$) ou si $p(x)$ est symétrique pour x suffisamment grand en valeur absolue, auquel cas on a

$$\int_{-\infty}^{+\infty} \sin txp(x)\,dx \equiv \int_{-M}^{+M} \sin txp(x)\,dx,$$

cette quantité étant évidemment dérivable.

Si pour $|x|$ assez grand on a $p(x) = \frac{1}{x^2\varphi(x)}$, $\varphi(x)$ étant une fonction croissante, comme $\int_A^{+\infty} \frac{1-\cos tx}{t}\,\frac{dx}{x^2\varphi(x)}$ est inférieure à $\frac{1}{\varphi(A)}\int_0^{+\infty}\frac{1-\cos u}{u^2}\,du$, on voit que $\int \cos txp(x)\,dx$ est dérivable. Si la condition de symétrie énoncée au paragraphe ci-dessus est réalisée, on voit que $\int e^{itx}p(x)\,dx$ est dérivable, la valeur de la dérivée étant $i\mathfrak{M}_c$ ($\mathfrak{M}_c$ est la moyenne au sens de Cauchy). On peut, en prenant pour $\varphi(x)$ une fonction croissant suffisamment lentement, s'arranger de telle sorte que le premier moment soit divergent. C'est un exemple de cette fonction que l'on a donné plus haut en prenant

$$\varphi(x) = [\log|x|]^\alpha \qquad \text{avec} \quad \alpha \leq 1.$$

Nous avons montré que $\frac{\Sigma x_i}{n}$ converge pourtant en probabilité vers $\mathfrak{M}_c$.

Dans le cas où l'on a affaire à une somme de variables aléatoires n'obéissant pas à la même loi de probabilité (cas à envisager pour l'étude des estimations par les fonctions v_i) il est nécessaire et suffisant pour la convergence en probabilité de cette somme que $\frac{\Sigma\psi_i\left(\frac{t}{n}\right)}{t}$ tende uniformément ($\varepsilon < |t| < T$) vers ia. On peut remplacer cette condition assez peu maniable par un ensemble de conditions suffisantes mais non nécessaires.

Supposons que les $\psi_i(t)$ soient dérivables pour $t = 0$, les dérivées ayant les valeurs ia_i. Si quel que petit que soit ζ on peut trouver η tel que pour $|t| < \eta$ on ait

$$\left|\frac{\psi_i(t)}{t} - ia_i\right| < \zeta$$

quel que soit l'indice i [nous pouvons appeler cette condition l'égalité de la dérivabilité des $\psi_i(t)$] il est facile de voir que $\dfrac{\Sigma\psi_i\left(\frac{t}{n}\right)}{t}$ tend vers $i\dfrac{\Sigma a_i}{n}$ uniformément pour $\varepsilon < |t| < T$. Si la suite $a_1, a_2, \ldots, a_n$ converge au sens de Cesaro vers a, la moyenne des variables aléatoires converge en probabilité vers cette valeur. Il est bien évident qu'un nombre fini de fonctions $\psi_j(t)$ peut échapper à cette condition de dérivabilité. Il faut toutefois que les nombres dérivés supérieurs soient bornés : car alors $\dfrac{1}{n}\dfrac{\Sigma\psi_j\left(\frac{t}{n}\right)}{\frac{t}{n}}$ tend vers o. On pourrait même admettre une infinité de fonctions non dérivables à l'origine. Il faudrait alors leur imposer des conditions de régularité plus compliquées.

Estimation du paramètre figurant dans les limites de la variable aléatoire dans le cas où la loi de probabilité élémentaire est $\frac{dF(x)}{F(m)}$. — Nous avons déjà vu que les procédés généraux ne peuvent s'appliquer. Nous ferons tout de suite le changement de variable aléatoire $F(x) = X$ et du paramètre $F(m) = M$. Si $F(m)$ est continue et a une dérivée non nulle, ce qui permet d'écrire $m = \varphi(M)$ (φ étant continue), un théorème connu permet d'affirmer que si $P(X_1, \ldots, X_n)$ converge en probabilité vers M, $\varphi[P(X_1, \ldots, X_n)]$ converge en probabilité vers m. Nous sommes donc amenés à étudier la loi de probabilité homogène $\frac{dX}{M}$ $(o < X < M)$. Plusieurs fonctions de $X_1, \ldots, X_n$ convergent en probabilité vers M (le double de la moyenne, le double de la médiane). Dans le cas général, comme nous le verrons plus loin, la distribution de ces variables aléatoires tend vers une loi de Gauss. Ce sont, comme nous le définirons plus loin, des infiniment petits aléatoires d'ordre $\frac{1}{2}$ en $\frac{1}{N}$. Il est possible de donner dans ce cas très particulier une estimation beaucoup plus précise. Considérons la plus grande valeur du groupe des n variables aléatoires. Il est facile de voir que sa loi de probabilité élémentaire est $n\left(\frac{X}{M}\right)^{n-1}\frac{dX}{M}$. Si l'on pose

$n(X - M) = \mu$, on voit que la loi de probabilité élémentaire de μ tend vers $e^{\frac{\mu}{M}} \frac{d\mu}{M}$. La probabilité pour que $n(X - M)$ soit compris entre deux nombres (négatifs évidemment) donnés tend vers une limite finie. Cette estimation est donc infiniment plus précise que toutes celles dont nous nous occuperons par la suite et où la variable $\sqrt{n}(X - M)$ joue le rôle que joue ici $n(X - M)$. Il est possible de rattacher cette grandeur au maximum de vraisemblance : la vraisemblance de l'expérience $(X_1, \ldots, X_n)$ sera en effet $\frac{1}{M^n}$. Il faudra donc prendre pour M la plus petite valeur possible, c'est-à-dire le plus grand des nombres $(X_1, \ldots, X_n)$.

Estimation de plusieurs paramètres. — Dans le cas où la loi de probabilité élémentaire contient plusieurs paramètres (deux pour fixer les idées), il est évident que sous les mêmes conditions que dans le cas d'un seul paramètre on peut écrire

$$E\left[\frac{1}{f}\frac{\partial f}{\partial a}\right] = 0 \quad \text{et} \quad E\left[\frac{1}{f}\frac{\partial f}{\partial b}\right] = 0.$$

Par conséquent

$$\frac{1}{n}\sum_{i=1}^{n} \frac{1}{f(x_i, a, b)} \frac{\partial f(x_i, a, b)}{\partial a} \quad \text{et} \quad \frac{1}{n}\sum_{i=1}^{n} \frac{1}{f(x_i, a, b)} \frac{\partial f(x_i, a, b)}{\partial b}$$

convergent en probabilité vers o. On voit que si le système

$$\sum \frac{1}{f(x_i, T_a, T_b)} \frac{\partial f(x_i, T_a, T_b)}{\partial a} = 0 \quad \text{et} \quad \sum \frac{1}{f(x_i, T_a, T_b)} \frac{\partial f(x_i, T_a, T_b)}{\partial b} = 0$$

n'a qu'une racine en T_a, T_b dans l'intervalle de recherche et si les deux fonctions $\frac{\partial \log f(x, a, b)}{\partial a}$ et $\frac{\partial \log (x, a, b)}{\partial b}$ sont uniformément continues par rapport à *l'ensemble* (a, b) sur tout l'intervalle de variation de x, au point a, b les deux nombres T_a et T_b convergent en probabilité vers a et b. Il sera encore plus difficile en général de résoudre ce système que dans le cas d'un paramètre. On peut encore dans ce cas trouver une infinité de statistiques « consistent ».

Nous en donnerons un exemple particulièrement simple. Soient $\varphi_1(x)$ et $\varphi_2(x)$ deux fonctions telles que : $E[\varphi_1(x)] = \varphi_1(a, b)$ et $E[\varphi_2(x)] = \varphi_2(a, b)$ soient deux fonctions continues en (a, b) au point a, b. Si le système

$$z_1 = \varphi_1(a, b), \qquad z_2 = \varphi_2(a, b)$$

n'admet qu'une solution $a = \alpha(z_1, z_2)$ et $b = \beta(z_1, z_2)$, les quantités

$$\alpha\left[\frac{\Sigma\varphi_1(x_i)}{n}, \frac{\Sigma\varphi_2(x_i)}{n}\right] \quad \text{et} \quad \beta\left[\frac{\Sigma\varphi_1(x_i)}{n}, \frac{\Sigma\varphi_2(x_i)}{n}\right]$$

convergent en probabilité respectivement vers a et b. On n'a aucune difficulté à étendre le raisonnement à un nombre fini de paramètres. Il est de même bien évident que rien ne serait changé si au lieu d'une seule variable aléatoire x, on en avait un nombre fini : $x_1, \ldots, x_n$.

Cas d'une infinité dénombrable de paramètres. — Supposons maintenant que dans la loi de probabilité figure une infinité dénombrable de paramètres. C'est ce qui se produira si l'on a une suite de résultats dont on ignore la forme de la loi de distribution, celle-ci appartenant toutefois à un domaine fonctionnel fixé *a priori* (on considérera par exemple toutes les lois admettant un développement infini d'une forme donnée). Trouver la forme de la loi reviendra alors à estimer une suite de paramètres; soit $f(x, a_1, \ldots, a_n, \ldots)$ une telle loi. Sous les mêmes conditions que précédemment on aura, quel que soit i, $E\left[\frac{1}{f}\frac{\partial f}{\partial a_i}\right] = 0$, et par conséquent *chacune* des variables

$$X_1 = \frac{1}{n}\sum \frac{1}{f}\frac{\partial f}{\partial a_1}, \quad \ldots, \quad X_i = \frac{1}{n}\sum \frac{1}{f}\frac{\partial f}{\partial a_i}, \quad \ldots$$

convergera en probabilité vers 0. Mais pour que l'ensemble de ces moyennes converge en probabilité vers 0 il faut une condition supplémentaire. On peut donner une condition suffisante et non nécessaire de cette convergence. On sait que

$$\sigma_{X_1} = \frac{1}{\sqrt{n}}\sqrt{E\left[\frac{1}{f}\frac{\partial f}{\partial a_1}\right]^2}, \quad \ldots, \quad \sigma_{X_i} = \frac{1}{\sqrt{n}}\sqrt{E\left[\frac{1}{f}\frac{\partial f}{\partial a_i}\right]^2}, \quad \ldots.$$

En vertu de l'inégalité de Tchebycheff, on a

$$P\{|X_1| > \varepsilon\} < \frac{\sigma_{X_1}^2}{\varepsilon^2}, \quad \ldots, \quad P\{|X_i| > \varepsilon\} < \frac{\sigma_i^2}{\varepsilon^2},$$

et par conséquent la probabilité que l'une quelconque des quantités $X_1, \ldots, X_i$ soit en valeur absolue supérieure à ε sera inférieure à $\frac{\sum_1^i \sigma_i^2}{\varepsilon^2}$. Si la série $\Sigma \mathrm{E}\left[\frac{1}{f}\frac{\partial f}{\partial a_i}\right]^2$ est convergente, on voit que aussi petit que soit ε cette probabilité pourra être rendue arbitrairement petite. La fin de la démonstration sera calquée sur les précédentes. Il faut supposer que le système n'a qu'une solution

$$a_1 = \alpha_1(x_1, \ldots, x_i, \ldots), \qquad \ldots, \qquad a_i = \alpha_i(x_1, \ldots, x_i, \ldots),$$

et de plus que toutes les fonctions $\frac{1}{f}\frac{\partial f}{\partial a_i}$ sont continues en $a_1, \ldots, a_i, \ldots$, au sens suivant : l'ensemble de conditions $|\alpha_i - a_i| < \eta$ (quel que soit i) entraîne

$$\left|\frac{1}{f}\frac{\partial f}{\partial a_i}(x, \alpha_1, \ldots, \alpha_i, \ldots) - \frac{1}{f}\frac{\partial f}{\partial a_i}(x, a_1, \ldots, a_i, \ldots)\right| < \varepsilon$$

(ε ne dépendant ni de x, ni de i). C'est ce que l'on peut appeler la continuité forte en $a_1, \ldots, a_i, \ldots$. Ces conditions sont évidemment très restrictives.

On peut, comme précédemment, évaluer chacun des $a_1, \ldots, a_i, \ldots$, par des fonctions auxiliaires sur lesquelles on doit faire des restrictions qui sont exactement du même ordre que celles faites pour le maximum de vraisemblance. Mais, de toute façon, en général, on aura à résoudre un système d'une infinité d'équations à une infinité d'inconnues. Le problème n'est donc pas susceptible d'être résolu pratiquement. Nous donnerons un exemple dans lequel il est particulièrement simple et où la solution n'exige pas de conditions de validité. Supposons que l'on sache la loi de probabilité élémentaire développable en série de fonctions quasi-octogonales; prenons le système fourni par la loi de Gauss $\mathrm{G}(x)$. En posant

$$\frac{d^n \mathrm{G}(x)}{dx^n} = \sqrt{n!}\,\mathrm{G}(x)\,\mathrm{P}_n(x),$$

on sait que

$$\int_{-\infty}^{+\infty} \mathrm{G}(x)\,\mathrm{P}_n(x)\,\mathrm{P}_m(x)\,dx = \begin{cases} 0 & \text{si } n \neq m, \\ 1 & \text{si } m = n, \end{cases}$$

on aura alors

$$f(x) = \mathrm{G}(x)\left[1 + \alpha_1 \mathrm{P}_1(x) + \ldots + \alpha_n \mathrm{P}_n(x) + \ldots\right],$$

on voit que : $\mathrm{E}[\mathrm{P}_i(x)] = \alpha_i$.

Le théorème de M. Khintchine montre que $\frac{1}{n}\sum_{j=1}^{n} \mathrm{P}_i(x_j)$ converge en probabilité vers α_i. L'estimation par la méthode du maximum de vraisemblance conduit à résoudre le système

$$\sum_{j=1}^{n} \frac{P_1(x_j)}{1 + \alpha_1 P_1(x_j) + \ldots + \alpha_i P_i(x_j) + \ldots} = 0,$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots,$$
$$\sum_{j=2}^{n} \frac{P_i(x_j)}{1 + \alpha_1 P_1(x_j) + \ldots + \alpha_i P_i(x_j) + \ldots} = 0,$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots,$$

qui est beaucoup moins simple.

Convergence presque certaine. — M. Kolmogoroff a démontré récemment (la démonstration a paru dans le traité de M. Fréchet : *Calcul des probabilités : compléments théoriques*) que l'existence de $E(x)$ entraîne non seulement la convergence en probabilité mais la convergence presque certaine de $\frac{\Sigma x_i}{n}$ vers $E(x)$. Un des aspects de la convergence presque certaine est la convergence forte. Ce qui signifie que la probabilité de l'ensemble des inégalités

$$\left| \frac{1}{n} \sum_{i=1}^{n} x_i - E(x) \right| < \varepsilon,$$
$$\left| \frac{1}{n+1} \sum_{i=1}^{n+1} x_i - E(x) \right| < \varepsilon,$$
$$\ldots\ldots\ldots\ldots\ldots\ldots,$$
$$\left| \frac{1}{n+p} \sum_{i=1}^{n+p} x_i - E(x) \right| < \varepsilon,$$
$$\ldots\ldots\ldots\ldots\ldots\ldots$$

tend vers 1, aussi petit que soit ε quand n croît indéfiniment. Dans le cas où l'on prend pour variable aléatoire $\frac{1}{f}\frac{\partial f}{\partial m}$, on voit que si f est dérivable par rapport à m sous le signe $\int$ la quantité $\frac{1}{n}\sum_{i=1}^{n}\frac{1}{f(x_i, m)}\frac{\partial f(x_i, m)}{\partial m}$ converge *presque certainement* vers 0. Le raisonnement fait au début du Chapitre I montre alors que la probabilité de l'ensemble des inégalités

$$|T_n - m| < \varepsilon, \qquad |T_{n+1} - m| < \varepsilon, \qquad \ldots, \qquad |T_{n+p} - m| < \varepsilon, \qquad \ldots$$

tend vers 1 avec $\frac{1}{n}$ tendant vers 0, sous les mêmes conditions que celles imposées alors.

CHAPITRE III.

LOIS-LIMITES D'ESTIMATION : CRITÈRES DE « GAUSSIVITÉ ».

Il est donc établi que sous certaines conditions la racine de l'équation du « maximum de vraisemblance » converge en probabilité vers le paramètre à estimer. On a pu également former des estimations convergentes en probabilité vers ce paramètre quand le nombre d'expériences croît indéfiniment. Proposons-nous maintenant de chercher les lois-limites auxquelles satisfont ces estimations et de calculer les coefficients déterminant ces lois-limites et en particulier leur écart-type.

Examinons le cas de l'estimation T_l par la méthode de M. Fisher. T_l sera la racine de l'équation

$$\sum_{i=1}^{n} \frac{1}{f(x_i, T_l)} \frac{\partial f(x_i, T_l)}{\partial m} = 0$$

quand n croît indéfiniment. Ce qui peut encore s'écrire

$$\sum_{i=1}^{n} \frac{1}{f} \frac{\partial f}{\partial m}(x_i, T_l - m + m) = \sum_{i=1}^{n} u(x_i, T_l - m + m) = 0,$$

en posant le changement de notation $u = \frac{1}{f} \frac{\partial f}{\partial m}$.

Développons le premier nombre d'après la formule des accroissements finis; on a

$$\sum_{i=1}^{n} u(x_i, m) + (T_l - m) \sum_{i=1}^{n} \frac{\partial}{\partial m} u[x_i, m + \theta(T_l - m)] = 0 \qquad \text{avec} \quad 0 < \theta < 1.$$

Donc

$$T_l - m = -\frac{\dfrac{1}{n} \displaystyle\sum_{i=1}^{n} u(x_i, m)}{\dfrac{1}{n} \displaystyle\sum_{i=1}^{n} \frac{\partial u}{\partial m}[x_i, m + \theta(T_l - m)]}.$$

Si T_l converge en probabilité vers m, le dénominateur converge en probabilité de la même façon que $\frac{1}{n}\sum\frac{\partial u}{\partial m}(x_i, m)$ à condition que $\frac{\partial u}{\partial m}(x_i, m)$ soit uniformément continue en m dans tout l'intervalle de variation de x. Si $E\left[\frac{\partial u}{\partial m}\right]$ existe, le théorème de M. Khintchine s'applique et $\frac{1}{n}\sum_{i=1}^{n}\frac{\partial u(x_i, m)}{\partial m}$ converge en probabilité vers $E\left[\frac{\partial u}{\partial m}\right]$. $\frac{\partial u}{\partial m}$ est égal à $\frac{1}{f}\frac{\partial^2 f}{\partial m^2} - \frac{1}{f^2}\left(\frac{\partial f}{\partial m}\right)^2$.

L'existence de $E\left[\frac{\partial u}{\partial m}\right]$ est donc liée à l'existence de $E\left[\frac{1}{f}\frac{\partial f}{\partial m}\right]^2$ à la condition que $\frac{\partial^2 f}{\partial m^2}$ existe et soit intégrable, auquel cas $\int\frac{\partial^2 f}{\partial m^2}dx = 0$ bien entendu; $E\left[\frac{\partial u}{\partial m}\right]$ est alors égal à $-E\left[\frac{1}{f}\frac{\partial f}{\partial m}\right]^2 = -E[u]^2$.

Ceci posé, nous allons tout d'abord établir un lemme :

Soit une variable aléatoire z_n quotient de y_n par x_n. Supposons que x_n converge en probabilité vers a, différent de 0, alors que la loi de probabilité totale de y_n, $\Phi_n(y)$ tend vers $\Phi(y)$.

Examinons la limite de la loi de probabilité totale de z_n, $F_n(z_n)$.

Une représentation graphique simplifiera le raisonnement :

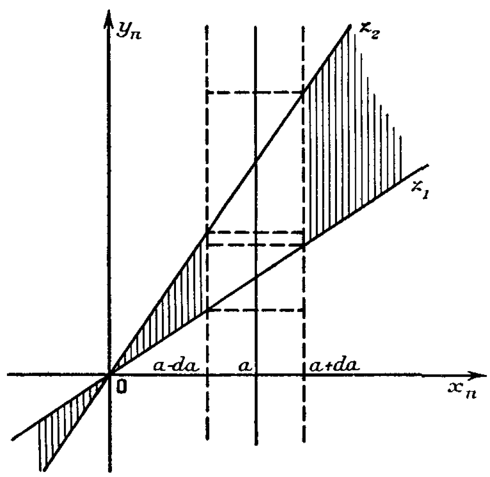

Fig. 1.

Appelons $P_n(x, y_1, y_2)$ la loi de probabilité totale de x_n avec $y_1 < y_n < y_2$. S'il y a une probabilité non nulle pour que l'on ait $y_1 < y_n < y_2$, il est évident que même avec ces liaisons x_n convergera en probabilité vers a, et par suite on pourra trouver N tel que pour $n > \mathrm{N}$ on ait

$$(\alpha) \qquad \mathrm{P}_n(a + da,\ y_1,\ y_2) - \mathrm{P}_n(a - da,\ y_1,\ y_2) > 1 - \varepsilon$$

aussi petits que soient da et ε.

Supposons que z_1 et z_2 soient deux nombres positifs ($z_2 > z_1$), cas qui correspond à la figure. La probabilité pour que l'on ait $z_1 \leqq z < z_2$ sera $\mathrm{F}_n(z_2) - \mathrm{F}_n(z_1)$. On voit facilement qu'elle sera supérieure à la probabilité pour que le point de coordonnées x_n et y_n soit dans le rectangle ayant ses sommets sur les droites d'abscisses $a - da$ et $a + da$ et sur celles d'ordonnées $z_1(a + da)$ et $z_2(a - da)$, et inférieure à la probabilité pour qu'il se trouve *soit* dans le rectangle ayant ses sommets sur les mêmes parallèles à Oy que le précédent et sur les droites d'ordonnées $z_1(a - da)$ et $z_2(a + da)$, *soit* dans la région hachurée. En vertu de la convergence en probabilité de x_n vers a (quel que soit y_n cette fois-ci), il y a une probabilité inférieure à ε' pour que l'on ait un point dans cette région si n est suffisamment grand.

On peut donc écrire l'inégalité

$$\begin{gathered}\{\ \mathrm{P}_n[a + da,\ z_2(a - da),\ z_1(a + da)] \\ - \mathrm{P}_n[a - da,\ z_2(a - da),\ z_1(a + da)]\}\{\Phi_n[z_2(a - da)] - \Phi_n[z_1(a + da)]\}\end{gathered}$$

inférieure à $\mathrm{F}_n(z_2) - \mathrm{F}_n(z_1)$.

Et cette même quantité sera inférieure à

$$\begin{gathered}\{\ \mathrm{P}_n[a + da,\ z_2(a + da),\ z_1(a - da)] \\ - \mathrm{P}_n[a - da,\ z_2(a + da),\ z_1(a - da)]\}\{\Phi_n[z_2(a + da)] - \Phi_n[z_1(a - da)]\} + \varepsilon'.\end{gathered}$$

Il est évident que si z_1 était négatif on aurait à écrire des inégalités analogues mais différentes. La relation (α) permet d'ailleurs de renforcer ces deux inégalités en écrivant

$$\begin{gathered}\Phi_n[z_2(a - da)] - \Phi_n[z_1(a + da)] - \varepsilon'' \\ \leqq \mathrm{F}_n(z_2) - \mathrm{F}_n(z_1) \leqq \Phi_n[z_2(a + da)] - \Phi_n[z_1(a - da)] + \varepsilon'\end{gathered}$$

si $n > \mathrm{N}$.

Si $\Phi_n(y)$ tend vers $\Phi(y)$, on sait que la convergence est obligatoirement uniforme en dehors des points de discontinuité de $\Phi(y)$. Donc, quel que soit y pour $n > \nu$, on aura $|\Phi_n(y) - \Phi(y)| < \eta$ quel que soit η; n étant supérieur au plus grand des nombres N et ν, on aura donc

$$\begin{gathered}\Phi[z_2(a - da)] - \Phi[z_1(a + da)] - 2\eta - \varepsilon'' \\ \leqq \mathrm{F}_n(z_2) - \mathrm{F}_n(z_1) \leqq \Phi[z_2(a + da)] - \Phi[z_1(a - da)] + 2\eta + \varepsilon'.\end{gathered}$$

Si donc, pour $y = az_2$ et az_1 $\Phi(y)$ est continue, on aura

$$\lim [F_n(z_2) - F_n(z_1)] = \Phi(az_2) - \Phi(az_1).$$

Considérons maintenant la variable aléatoire $\sqrt{n}(T_l - m)$ qui est égale à

$$-\frac{\frac{1}{\sqrt{n}}\Sigma u(x_i, m)}{\frac{1}{n}\sum \frac{\partial u}{\partial m}[x_i, m + \theta(T_l - m)]}.$$

Si $E(u^2)$ existe la loi de probabilité totale du numérateur ξ tend vers la loi de Gauss $\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\xi} e^{-\frac{x^2}{2\sigma^2}}dx$ dans laquelle $\sigma = \sqrt{E(u^2)}$. Le dénominateur convergeant probablement vers $-E(u^2)$, il s'ensuit d'après le lemme établi que la loi de probabilité de $\xi = \sqrt{n}(T_l - m)$ tend vers la loi de Gauss d'écart-type $\frac{1}{\sqrt{E(u^2)}}$ soit $\frac{1}{\sqrt{E\left(\frac{1}{f}\frac{\partial f}{\partial m}\right)^2}}$. C'est du moins ce qu'un changement de variables très simple permet de mettre en évidence.

Nous avons donc une expression de la précision du « maximum de vraisemblance ». Les résultats précédents peuvent s'étendre aux statistiques obtenues dans les deux premiers Chapitres. Soient $\varphi_i(x_i, m)$ les fonctions satisfaisant aux conditions suffisantes pour former une estimation « consistent » T_c.

On aura ici encore

$$\sqrt{n}(T_c - m) = \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varphi_i(x_i, m)}{-\frac{1}{n}\sum_{i=1}^{n}\frac{\partial \varphi_i}{\partial m}[x_i, m + \theta(T_c - m)]}.$$

Si les $\frac{\partial \varphi_i}{\partial m}$ sont *également uniformément* continus en m pour toutes les valeurs de x_i, le dénominateur de $\sqrt{n}(T_c - m)$ converge vers la même limite probable que $\frac{1}{n}\sum \frac{\partial \varphi_i}{\partial m}(x_i, m)$. Cette convergence a été étudiée dans les deux premiers Chapitres. Le numérateur est le quotient par $\sqrt{n}$ d'une somme de variables indépendantes. Le problème de sa loi-limite a fait l'objet de nombreuses études. Si les φ_i sont identiques quel que soit i, il suffit que $E(\varphi^2)$ existe pour que la loi limite soit de Gauss.

Sinon, divers critères peuvent être utilisés. Liapounoff a montré que si l'on a

$$\lim_{n\to\infty} \frac{\Sigma E[|\varphi_i|^{2+\delta}]}{[\Sigma E(\varphi_i^2)]^{1+\delta}} = 0,$$

dans lequel δ est un nombre positif aussi petit que l'on veut, la loi-limite est une loi de Gauss de moyenne nulle et d'écart-type $\lim \sqrt{\frac{\Sigma E(\varphi_i^2)}{n}}$.

Dans ce cas la loi-limite de $\sqrt{n}(T_l - m)$ serait encore une loi de Gauss, dont l'écart-type serait

$$\sigma = \frac{\lim \sqrt{\frac{\Sigma E(\varphi_i^2)}{n}}}{\lim \frac{1}{n} \left| \Sigma E\left(\frac{\partial \varphi_i}{\partial m}\right) \right|}.$$

En dehors des conditions énoncées il est donc nécessaire pour qu'il y ait une loi-limite que la fraction indiquée ait une limite pour n infini. Il suffit que chacune des deux suites $E(\varphi_i^2)$ et $E\left(\frac{\partial \varphi_i}{\partial m}\right)$ tende vers une limite finie pour i infini, mais ce n'est nullement nécessaire. Dans le cas particulier que nous avons indiqué où $\varphi_i(x, m)$ est identique à $u(x) - \varphi(m)$, la condition de Liapounoff est vérifiée (tous les φ_i sont identiques); la formule trouvée donne

$$\sigma = \frac{\sqrt{\int u^2(x) f(x, m)\, dx - \varphi(m)^2}}{|\varphi'(m)|}.$$

La condition de la limite de σ pour n infini est toute résolue.

Nous pouvons maintenant donner plus facilement diverses définitions relatives aux *infiniment petits probables*.

Une quantité aléatoire X_n convergeant probablement vers o pour n infini sera dite d'ordre α, si l'on peut trouver deux nombres A et B certains, finis, et de même signe, tels que la probabilité pour que l'on ait $A \leqq n^\alpha X_n < B$ soit une quantité qui reste supérieure à un nombre positif quand n augmente indéfiniment. On voit facilement que si $n^\alpha X_n$ a une loi-limite de probabilité (différente de la loi caractérisant la convergence vers o), X_n est d'ordre α. En particulier $T_l - m$ et $T_c - m$ sont d'ordre $\frac{1}{2}$. Il convient de préciser que si X_n est d'ordre α, il ne saurait être d'ordre β différent de α. En effet, soient $F_n(x)$ les lois de probabilités totales des variables $n^\alpha X_n$ pour chaque valeur de n. On sait que cet ensemble de loi converge uniformément vers au moins une fonction non décroissante (qui peut d'ailleurs ne pas être une loi de probabilité totale). Il peut également y avoir plusieurs et même une infinité de fonctions limites; soit $\Phi(x)$ l'une d'entre elles. Supposons tout d'abord $\beta > \alpha$; la probabilité pour que $C \leqq n^\beta X_n < D$ est égale à $F_n\left(\frac{D}{n^{\beta-\alpha}}\right) - F_n\left(\frac{C}{n^{\beta-\alpha}}\right)$. Cette quantité peut être rendue inférieure à ε pour une infinité de valeurs de n. Pour cela, nous choisirons tous les n dans la suite rendant $F_n(x)$ convergeant vers $\Phi(x)$. Deux cas peuvent se présenter :

1° On peut trouver ξ tel que pour $o < x < \xi$, $\Phi(x)$ soit continue. $F_n(x)$ tend alors vers $\Phi(x)$ d'une manière uniforme dans cet intervalle et $F_n\left(\frac{D}{n^{\beta-\alpha}}\right) - F_n\left(\frac{C}{n^{\beta-\alpha}}\right)$ très peu différent de $\Phi\left(\frac{D}{n^{\beta-\alpha}}\right) - \Phi\left(\frac{C}{n^{\beta-\alpha}}\right)$ peut être rendu aussi petit que l'on veut.

2° Un tel nombre ξ n'existe pas. Il y a donc au voisinage de l'origine une infinité obligatoirement dénombrable de points pour lesquels $\Phi(x)$ est discontinue. Si $\frac{D}{n^{\beta-\alpha}}$ et $\frac{C}{n^{\beta-\alpha}}$ sont des points de discontinuité de $\Phi(x)$, on peut, pour chaque valeur de n, trouver deux nombres a_n et b_n avec

$$b_n > \frac{D}{n^{\beta-\alpha}} > \frac{C}{n^{\beta-\alpha}} > a_n.$$

a_n et b_n tendant vers o et tels que $\Phi(x)$ soit continu aux points a_n et b_n. On a

$$F_n(b_n) - F_n(a_n) \geqq F_n\left(\frac{D}{n^{\beta-\alpha}}\right) - F_n\left(\frac{C}{n^{\beta-\alpha}}\right),$$

$F_n(a_n)$ et $F_n(b_n)$ convergent uniformément vers $\Phi(a_n)$ et $\Phi(b_n)$. $\Phi(x)$ étant une fonction non décroissante, on peut trouver θ tel que pour $o < x < \theta$, on ait $\Phi(x) - \Phi(+o) < \varepsilon$. $F_n(b_n) - F_n(a_n)$ pouvant être rendu arbitrairement voisin de $\Phi(b_n) - \Phi(a_n)$ peut donc être rendu inférieur à ε.

De même $F_n\left(\frac{D}{n^{\beta-\alpha}}\right) - F_n\left(\frac{C}{n^{\beta-\alpha}}\right)$, ceci quels que soient C et D. Si β est inférieur à α la probabilité pour que $C \leqq n^\beta X_n < D$ est égale à $F_n(D n^{\alpha-\beta}) - F_n(C n^{\alpha-\beta})$; le même raisonnement que pour $\beta > \alpha$ s'appliquera, mais cette fois ce sera autour du point $x = +\infty$ au lieu d'être autour de l'origine.

L'ordre d'un infiniment petit ayant une valeur donnée n'est donc pas susceptible d'en prendre une autre. Comme dans l'analyse des variables certaines il serait facile de donner des exemples d'infiniment petits probables d'ordre inférieur ou supérieur à tout nombre donné, ou d'ordre indéterminé compris entre deux valeurs.

On appellera encore *infiniment petits probables équivalents*, deux infiniment petits probables X_n et Y_n tels que l'inégalité $\left| 1 - \frac{X_n}{Y_n} \right| < \varepsilon$ ait une probabilité tendant vers l'unité d'être réalisée. On pourrait déduire de ces définitions toute une série de théorèmes. Les résultats énoncés nous suffisent ; nons allons en donner l'application à deux cas particulièrement intéressants.

1° *Résolution de l'équation du maximum de vraisemblance.* — Nous avons déjà constaté la difficulté de la résolution de cette équation dans le cas général. Étant donnée son importance qui sera mise en évidence dans le Chapitre suivant, il est indispensable de fournir de cette équation une solution *approchée* en un sens que nous préciserons. Nous savons que nous pouvons avoir une infinité de solutions *consistent* ; T_c est l'une d'entre elles. T_l sera donc racine de

$$\Sigma u(x_i, T_l - T_c + T_c) = 0,$$

avec les mêmes notations que celles employées jusqu'ici. Sous les mêmes hypothèses que celles faites au début de ce Chapitre, on peut développer le premier membre. On a donc

$$\Sigma u(x_i, T_c) + (T_l - T_c) \sum \frac{\partial u}{\partial m} [x_i, T_c + \theta(T_l - T_c)] = 0.$$

D'où l'on tire

$$T_l - T_c = - \frac{\frac{1}{n} \Sigma u(x_i, T_c)}{\frac{1}{n} \sum \frac{\partial u}{\partial m} [x_i, T_c + \theta(T_l - T_c)]},$$

$T_l - T_c$ est évidemment un infiniment petit probable. Considérons la quantité $\dfrac{\frac{1}{n} \Sigma u(x_i, T_c)}{E\left(\frac{1}{f} \frac{\partial f}{\partial m}\right)^2}$. Il est facile de voir que le quotient de cette quantité par $T_l - T_c$ est très probablement peu différent de l'unité, car $\frac{1}{n} \sum \frac{\partial u}{\partial m}$ converge en probabilité vers $-E\left(\frac{1}{f} \frac{\partial f}{\partial m}\right)^2$. En se souvenant de ce que $E\left(\frac{1}{f} \frac{\partial f}{\partial m}\right)^2$ est l'inverse du carré de l'écart-type du maximum de vraisemblance, nous pouvons donc écrire

$$T_l = T_c + \frac{\sigma_l^2}{n} \Sigma u(x_i, T_c)$$

aux infiniment petits aléatoires d'ordre supérieur à $\frac{1}{2}$ près, sous des conditions identiques à celles exigées pour la gaussivité de T_l.

Dans le cas de la loi de Cauchy on sait que la médiane μ est une estimation correcte du centre de dispersion; l'estimation de M. Fisher pourra donc être approchée par la formule $\mu + \frac{\sigma_l^2}{n} \sum \frac{2(x_i - \mu)}{1 + (x_i - \mu)^2}$.

2° *Estimation par la méthode du* χ^2. — Prenons le cas d'une variable aléatoire discontinue. Soient $p_1, \ldots, p_i, \ldots$ les probabilités qu'a cette variable de prendre respectivement les valeurs $1, \ldots, i, \ldots$; sur N expériences, N_1 donnent un résultat 1, N_i le résultat i. On appelle χ^2 la quantité $\sum_{i=1}^{\infty} \frac{(N_i - Np_i)^2}{Np_i}$. Si les p_i sont fonctions du paramètre m, une méthode d'estimation consistera à prendre la valeur de m rendant χ^2 minimum. Soit T_χ cette estimation; T_χ sera racine de

$$\sum \frac{N_i}{p_i(T_\chi)} \frac{\partial p_i(T_\chi)}{\partial m} \frac{N_i}{Np_i(T_\chi)} = 0.$$

Nous allons tout d'abord montrer la correction de cette statistique. On voit facilement que

$$E(N_i^2) = N(N-1)p_i^2 + Np_i,$$

et

$$E(N_i^4) = N(N-1)(N-2)(N-3)p_i^4 + 6N(N-1)(N-2)p_i^3 + 7N(N-1)p_i^2 + Np_i.$$

Si l'on considère la variable aléatoire $\frac{1}{N}\sum \frac{N_i}{p_i(m)} \frac{\partial p_i}{\partial m} \frac{N_i}{Np_i(m)}$, sa valeur moyenne est $\frac{1}{N}\sum \frac{1}{p_i(m)} \frac{\partial p_i}{\partial m}$ et son écart-type un infiniment petit en $\frac{1}{\sqrt{N}}$.

Le théorème de Bienaymé-Tchebycheff montre que la variable aléatoire converge en probabilité vers $\frac{1}{N}\sum \frac{1}{p_i} \frac{\partial p_i}{\partial m}$ et par suite vers 0 quand N augmente indéfiniment $\left(\text{à condition que } \sum \frac{1}{p_i} \frac{\partial p_i}{\partial m} \text{ soit convergent}\right)$. Le même raisonnement qu'au Chapitre I permet d'en déduire que T_χ converge en probabilité vers m sous la condition que les $\frac{1}{p_i^2} \frac{\partial p_i}{\partial m}$ soient également continus au point m et que l'équation n'ait qu'une seule racine. D'après la formule obtenue au paragraphe précédent on peut donc écrire

$$T_l - T_\chi = \frac{\sigma_l^2}{N} \sum \frac{N_i}{p_i(T_\chi)} \frac{\partial p_i(T_\chi)}{\partial m}$$

aux infiniment petits aléatoires d'ordre supérieur près, T_l étant la racine de l'équation

$$\sum \frac{N_i}{p_i(T_l)} \frac{\partial p_i(T_l)}{\partial m} = 0.$$

Proposons-nous de calculer la loi-limite de probabilité de la variable aléatoire $\sqrt{n}(T_\chi - m)$.

Il nous faudra ici encore employer le même procédé, T_χ sera racine de l'équation

$$\begin{aligned}\sum \frac{N_i^2}{N} \frac{1}{p_i^2} \frac{\partial p_i}{\partial m}(T_\chi - m + m) \\ = \sum \frac{N_i^2}{N} \frac{1}{p_i^2} \frac{\partial p_i}{\partial m}(m) + (T_\chi - m) \sum \frac{N_i^2}{N} \left[\frac{1}{p_i^2} \frac{\partial^2 p_i}{\partial m^2} - \frac{2}{p_i^3}\left(\frac{\partial p_i}{\partial m}\right)^2\right] [m + \theta(T_\chi - m)] = 0.\end{aligned}$$

Donc

$$\sqrt{n}(T_\chi - m) = \frac{\frac{1}{\sqrt{N}} \sum \frac{N_i^2}{N} \frac{1}{p_i^2} \frac{\partial p_i}{\partial m}(m)}{-\frac{1}{N} \sum \frac{N_i^2}{N} \left[\frac{1}{p_i^2} \frac{\partial^2 p_i}{\partial m^2} - \frac{2}{p_i^3}\left(\frac{\partial p_i}{\partial m}\right)^2\right] [m + \theta(T_\chi - m)]}.$$

Nous supposerons maintenant pour simplifier les hypothèses que le nombre de catégories i est fini. $\frac{N_i}{N}$ converge en probabilité vers p_i. Donc la quantité aléatoire $\frac{N_i^2}{N^2}$ converge en probabilité vers p_i^2. Si les fonctions de m inscrites au dénominateur sont continues en m quel que soit i, on voit que ce dernier convergera en probabilité vers $-\Sigma p_i^2\left[\frac{1}{p_i^2} \frac{\partial^2 p_i}{\partial m^2} - \frac{2}{p_i^3}\left(\frac{\partial p_i}{\partial m}\right)^2\right]$, c'est-à-dire vers $2\sum \frac{1}{p_i}\left(\frac{\partial p_i}{\partial m}\right)^2$. Il faut maintenant chercher la loi de probabilité de la variable aléatoire X_N qui figure au numérateur. Nous écrirons

$$X_N = \frac{1}{\sqrt{N}} \sum \frac{(N_i - Np_i)^2}{Np_i^2} \frac{\partial p_i}{\partial m} + \frac{2}{\sqrt{N}} \sum \frac{N_i}{p_i} \frac{\partial p_i}{\partial m}.$$

$\sum \frac{(N_i - Np_i)^2}{Np_i^2} \frac{\partial p_i}{\partial m}$ a à la limite une loi analogue à celle du χ^2. Donc, la première partie de X_n est un infiniment petit aléatoire d'ordre $\frac{1}{2}$ et converge vers 0. $\frac{2}{\sqrt{N}} \sum \frac{N_i}{p_i} \frac{\partial p_i}{\partial m}$ a pour fonction caractéristique

$$\left[\Sigma p_i \exp\left(t \frac{2}{\sqrt{N}} \frac{1}{p_i} \frac{\partial p_i}{\partial m}\right)\right]^N.$$

En prenant le logarithme et en développant par rapport à t il est facile de voir que cette expression converge uniformément pour tout $|t|$ borné vers $\exp\left\{-\frac{t^2}{2}\sum\frac{4}{p_i}\left(\frac{\partial p_i}{\partial m}\right)^2\right\}$. La seconde partie de X_N suit donc à la limite une loi de Gauss d'écart-type $2\sqrt{\sum\frac{1}{p_i}\left(\frac{\partial p_i}{\partial m}\right)^2}$. Les deux variables dont la somme constitue X_N sont liées, mais il est facile de voir que la première converge en probabilité vers o, quelle que soit la seconde.

Nous établirons le théorème suivant :

Étant données deux suites de variables aléatoires x_N *et* y_N *dont les premières convergent probablement vers* o, *les deuxièmes étant telles que leurs lois de probabilités totales* $G_N(y_N)$ *convergent vers une loi* $G(y)$, *les lois des variables* $X_N = x_N + y_N$ *convergent vers* $G(X)$, *à condition que cette fonction* $G(X)$ *soit continue.*

Si $F_N(x_N)$ est la loi de probabilité de x_N, la loi de X_N sera

$$-\int_{-\infty}^{+\infty} F_N(x_N)\,dG_N(X_N-x_N) = -\left[\int_{-\infty}^{-\varepsilon}+\int_{-\varepsilon}^{+\varepsilon}+\int_{+\varepsilon}^{+\infty} F_N(x_N)\,dG_N(X_N-x_N)\right].$$

La première intégrale pourra évidemment être rendue inférieure à η en valeur absolue, la seconde sera comprise entre

$$(1-\eta)[G_N(X_N-\varepsilon)-G_N(X_N+\varepsilon)] \quad \text{et} \quad \eta[G_N(X_N-\varepsilon)-G_N(X_N+\varepsilon)],$$

et la troisième sera comprise entre

$$-(1-\eta)G_N(X_N-\varepsilon) \quad \text{et} \quad -G_N(X_N-\varepsilon).$$

On a donc

$$\begin{aligned}&\eta[G_N(X_N+\varepsilon)-G_N[X_N-\varepsilon)]+(1-\eta)G_N(X_N+\varepsilon)\\ &\quad<-\int_{-\infty}^{+\infty}F_N(x_N)\,dG_N(X_N-x_N)<\eta+(1-\eta)[G_N(X_N+\varepsilon)-G_N(X_N-\varepsilon)]+G_N(X_N-\varepsilon).\end{aligned}$$

Les deux termes extrêmes peuvent être rendus aussi voisins de $G_N(X_N+\varepsilon)$ qu'on le désire. Si $G(X)$ est continue, $G_N(X_N)$ tend vers $G(X)$ d'une manière uniforme, et $G_N(X_N+\varepsilon)$ est arbitrairement rapproché de $G(X_N)$. Le théorème est donc établi. En se fondant sur lui et sur le résultat obtenu au début du Chapitre, on voit que T_χ suivra à la limite une loi de Gauss où $\sigma = \dfrac{1}{\sqrt{\sum\frac{1}{p_i}\left(\frac{\partial p_i}{\partial m}\right)^2}}$, ce qui est le σ du maximum de vraisemblance. Nous établirons au Chapitre V que dans ces conditions $T_\chi - T_l$ est obligatoirement d'un ordre supérieur à $\frac{1}{2}$. Donc dans ce cas $\frac{1}{N}\sum \frac{N_i}{p_i(T_\chi)}\frac{\partial p_i(T_\chi)}{\partial m}$ est d'ordre supérieur à $\frac{1}{2}$.

Lois-limites dans les autres cas examinés aux Chapitres I et II. — Si le paramètre m figure dans la limite de la variable, on voit facilement, T_l étant toujours l'estimation qui dans ce cas joue le rôle du maximum de vraisemblance, que si $u(x, m) = \frac{1}{f}\frac{\partial f}{\partial m}$ on a

$$\sqrt{n}(T_l - m) = \frac{\frac{1}{\sqrt{n}}\Sigma[u(x_i, m) + f(m, m)]}{-\left[\left(\frac{\partial f}{\partial x} + \frac{\partial f}{\partial m}\right)[m + \theta(T_l - m)] + \frac{1}{n}\sum \frac{\partial u}{\partial m}[x_i, m + \theta(T_l - m)]\right]}.$$

La loi-limite du numérateur est encore une loi de Gauss et son écart type est $E\left[\frac{1}{f}\frac{\partial f}{\partial m}\right]^2 - f(m, m)^2$.

Si $\frac{\partial \log f(x, m)}{\partial m}$ est uniformément continu quel que soit x en m, le dénominateur converge en probabilité vers

$$E\left[\frac{1}{f}\frac{\partial f}{\partial m}\right]^2 + \frac{\partial f}{\partial m}(m, m),$$

à condition que l'on puisse dériver deux fois sous le signe d'intégration la fonction de m $\int_0^m f(x, m)dx$. L'écart-type de la loi-limite sera donc ici

$$\sigma = \frac{\sqrt{E\left[\frac{1}{f}\frac{\partial f}{\partial m}\right]^2 - f(m, m)^2}}{E\left[\frac{1}{f}\frac{\partial f}{\partial m}\right]^2 + \frac{\partial f}{\partial m}(m, m)}.$$

Si l'on a $f(m, m) = 0$ quel que soit m, on voit que l'équation du maximum de vraisemblance a la même forme que dans le cas des limites fixes. Si de plus $\frac{\partial f}{\partial m}(m, m) = 0$, l'écart-type de la loi-limite a également la même forme.

Dans le cas de p paramètres $a_1, \ldots . a_p$ le maximum de vraisemblance fournissant les estimations $A_1, \ldots, A_p$, on peut écrire

$$\xi_1 = \frac{1}{\sqrt{n}} \sum \frac{1}{f} \frac{\partial f}{\partial a_1}(x_i, a_1, \ldots, a_p)$$
$$= -\left[\sqrt{n}(A_1 - a_1) \frac{1}{n} \sum \frac{\partial}{\partial a_1}\left(\frac{1}{f} \frac{\partial f}{\partial a_1}\right) + \ldots + \sqrt{n}(A_p - a_p) \frac{1}{n} \sum \frac{\partial}{\partial a_p}\left(\frac{1}{f} \frac{\partial f}{\partial a_1}\right)\right],$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$\xi_p = \frac{1}{\sqrt{n}} \sum \frac{1}{f} \frac{\partial f}{\partial a_p}(x_i, a_1, \ldots, a_p)$$
$$= -\left[\sqrt{n}(A_1 - a_1) \frac{1}{n} \sum \frac{\partial}{\partial a_1}\left(\frac{1}{f} \frac{\partial f}{\partial a_p}\right) + \ldots + \sqrt{n}(A_p - a_p) \frac{1}{n} \sum \frac{\partial}{\partial a_p}\left(\frac{1}{f} \frac{\partial f}{\partial a_p}\right)\right].$$

Dans les seconds membres on remplace dans les dérivées $a_1, \ldots, a_p$ par $a_1 + \theta(A_1 - a_1), \ldots, a_p + \theta(A_p - a_p)$. On voit que sous les mêmes conditions de continuité que pour un seul paramètre les coefficients du second membre convergent en probabilité vers les coefficients du tableau symétrique :

$$\begin{array}{llll}
E\left[\frac{1}{f} \frac{\partial f}{\partial a_1}\right]^2, & E\left[\frac{1}{f^2} \frac{\partial f}{\partial a_1} \frac{\partial f}{\partial a_2}\right], & \ldots, & E\left[\frac{1}{f^2} \frac{\partial f}{\partial a_1} \frac{\partial f}{\partial a_p}\right], \\
\ldots\ldots & \ldots\ldots, & \ldots & \ldots\ldots, \\
E\left[\frac{1}{f^2} \frac{\partial f}{\partial a_p} \frac{\partial f}{\partial a_1}\right], & E\left[\frac{1}{f^2} \frac{\partial f}{\partial a_p} \frac{\partial f}{\partial a_2}\right], & \ldots, & E\left[\frac{1}{f} \frac{\partial f}{\partial a_p}\right]^2.
\end{array}$$

D'après la théorie des moyennes des variables aléatoires indépendantes définies sur une même catégorie d'épreuves, comme $E\left[\frac{1}{f} \frac{\partial f}{\partial a_i}\right]^2$ et $E\left[\frac{1}{f_2} \frac{\partial f}{\partial a_i} \frac{\partial f}{\partial a_j}\right]$ existent quels que soient i et j, la loi de probabilité limite de l'ensemble $\xi_1, \ldots, \xi_p$ est une loi de Gauss :

$$K \exp\{G(\xi_1, \ldots, \xi_p)\} d\xi_1 \ldots d\xi_p$$

dont la fonction caractéristique est :

$$\exp - \frac{1}{2}\left\{\Sigma u_i^2 E\left[\frac{1}{f} \frac{\partial f}{\partial a_i}\right]^2 + 2\Sigma\Sigma u_i u_j E\left[\frac{1}{f^2} \frac{\partial f}{\partial a_i} \frac{\partial f}{\partial a_j}\right]\right\}.$$

La forme quadratique $G(\xi_1, \ldots, \xi_p)$ est la forme réciproque du logarithme ψ de la fonction caractéristique. Un raisonnement semblable à celui qui a été fait pour une variable aléatoire, montre que la loi-limite de l'ensemble $\alpha_1 = \sqrt{n}(A_1 - a_1), \ldots, \alpha_p = \sqrt{n}(A_p - a_p)$ s'obtient en faisant sur la loi de

$\xi_1, \ldots, \xi_p$ le changement de variables :

$$\xi_1 = \alpha_1 \mathrm{E}\left[\frac{1}{f}\frac{\partial f}{\partial a_1}\right]^2 + \ldots + \alpha_p \mathrm{E}\left[\frac{1}{f^2}\frac{\partial f}{\partial a_1}\frac{\partial f}{\partial a_p}\right],$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots,$$
$$\xi_p = \alpha_1 \mathrm{E}\left[\frac{1}{f^2}\frac{\partial f}{\partial a_p}\frac{\partial f}{\partial a_1}\right] + \ldots + \alpha_p \mathrm{E}\left[\frac{1}{f}\frac{\partial f}{\partial a_p}\right]^2.$$

Puisque ce tableau est symétrique par rapport à la diagonale principale, la fonction caractéristique des v_i attachés aux α_i s'obtiendra en faisant le changement de variable inverse sur les u_i (les v_i correspondant à ξ_i et les u_i à α_i). La loi de probabilité élémentaire des α_i s'obtiendra donc en cherchant la forme réciproque de celle obtenue en faisant la substitution indiquée. En posant, pour abréger l'écriture,

$$\mathrm{E}\left[\frac{1}{f}\frac{\partial f}{\partial a_i}\right]^2 = c_{ii} \qquad \text{et} \qquad \mathrm{E}\left[\frac{1}{f^2}\frac{\partial f}{\partial a_i}\frac{\partial f}{\partial a_j}\right] = c_{ij} \qquad (c_{ij} = c_{ji}),$$

on voit que l'on a

$$v_i = \sum_{j=1}^{p} u_j c_{ij} = -\frac{\partial\psi}{\partial u_i} \qquad (\text{avec } i = 1, \ldots, p).$$

La résolution de ce système par rapport à u_j donne le système suivant :

$$u_j = \frac{1}{\Delta}\sum_{i=1}^{p} v_i \mathcal{C}_{ji} \qquad (j = 1, \ldots, p),$$

$\mathcal{C}_{ji}$ étant le mineur relatif à c_{ji} dans Δ déterminant des coefficients. Pour avoir la forme réciproque de ψ il faudra résoudre en v_i le système

$$\alpha_i = \frac{\partial\psi}{\partial v_i} = \sum_{i=1}^{p} \frac{\partial\psi}{\partial u_j}\frac{\partial u_j}{\partial v_i} \qquad (i = 1, \ldots, p)$$

et remplacer les v_i par leur valeur dans la forme $\psi(v_i)$ transformée de $\psi(u_i)$. Or, $\frac{\partial\psi}{\partial u_j} = -v_j$; donc

$$\alpha_i = -\sum_{j=1}^{p} v_j \frac{\mathcal{C}_{ji}}{\Delta} = -u_i.$$

Il suffira donc, puisque ψ est une forme paire, de former $\psi(\alpha_i)$ et la loi de probabilité des $\alpha_1, \ldots, \alpha_p$ sera

$$K \exp\{\psi(\alpha_i)\} d\alpha_1 \ldots d\alpha_p.$$

Un résultat bien connu sur la loi de Gauss montre que $K = \sqrt{\frac{\Delta}{(2\pi)^p}}$.

Les divers moments $E(\alpha_i^2)$, $E(\alpha_i\alpha_j)$ de cette loi sont les dérivées $\frac{\partial^2\psi}{\partial v_i^2}$ et $\frac{\partial^2\psi}{\partial v_i \partial v_j}$ changées de signe.

Comme $\frac{\partial\psi}{\partial v_i}$ est égal à α_i on voit immédiatement que $E(\alpha_i\alpha_j)$ est égal à $\frac{\mathcal{C}_{ij}}{\Delta}$. Si l'on veut la loi de probabilité d'un nombre r de $\alpha_i (r < p)$ en supposant les autres paramètres *connus*, il suffira dans la fonction caractéristique de rendre nuls les $p - r$, v_i correspondants des α_i négligés. La loi de probabilité des α_i considérés s'obtiendra en annulant les termes qui contiennent les autres α_i, la constante K devenant évidemment $\sqrt{\frac{\Delta'}{(2\pi)^r}}$, Δ' étant le discriminant de la nouvelle forme obtenue. La fonction qui figure dans l'exposant de e se construit donc de proche en proche. Ce fait aura de l'importance au cours du Chapitre IV.

Nous devrions maintenant examiner le cas de résultats liés en chaînes; nous préférons en réserver l'étude au moment où nous aborderons le maximum de précision.

Nous ne nous sommes préoccupé jusqu'à présent que des lois-limites. Nous allons maintenant étudier certaines grandeurs attachées à ces lois. En particulier, les deux premiers moments et la valeur médiane. Il est facile de voir que si des lois de probabilités totales convergent vers une loi-limite, les valeurs médianes μ_n de chacune des lois convergent vers la valeur médiane μ de la loi-limite, à condition que cette valeur soit unique. En effet, à cause de la convergence uniforme on a l'inégalité $|F_n(\mu_n) - F(\mu_n)| < \varepsilon$ pour n suffisamment grand et par conséquent $\frac{1}{2} - \varepsilon < F(\mu_n) < \frac{1}{2} + \varepsilon$. La non-décroissance de la fonction F entraîne que μ_n doit se trouver dans le voisinage de μ seule solution de $F(\mu) = \frac{1}{2}$. Dans le cas présent, la loi-limite de $\sqrt{n}(T_c - m)$ étant une loi de Gauss, la valeur médiane de $\sqrt{n}(T_c - m)$ tend vers zéro et la valeur médiane de T_c tend vers m, ceci sans conditions restrictives supplémentaires.

Il n'en est plus de même dans le cas des moments. M. Fréchet a, dans son Mémoire de « Metron » déjà cité, étudié la question plus générale de la valeur moyenne d'une fonction continue d'une variable aléatoire convergeant en probabilité vers une autre variable aléatoire. Nous rappellerons sans en fournir la démonstration certains de ses résultats qui sont valables en dehors du cas de la convergence, quand une variable aléatoire a une loi de probabilité totale $F_n(x)$ qui tend vers une loi de probabilité $F(x)$: $\varphi(x)$ étant une fonction continue pour que $\int_{-\infty}^{+\infty} \varphi(x)\,dF_n(x)$ tende vers $\int_{-\infty}^{+\infty} \varphi(x)\,dF(x)$, il suffit que les quantités $\int_a^b \varphi(x)\,dF_n(x)$ tendent *uniformément* quel que soit n vers leurs limites quand a et b croissent indéfiniment. Si $\varphi(x)$ est constamment positif, cette condition d'uniformité est de plus nécessaire.

Dans les estimations que nous avons construites $\sqrt{n}(T_e - m)$ se présente sous la forme d'un produit de deux variables aléatoires y_n et x_n, y_n étant égal à $\frac{1}{\sqrt{n}}\,\Sigma\varphi_i(x_i, m)$ et x_n à $\dfrac{1}{\frac{1}{n}\sum \frac{\partial \varphi_i}{\partial m}[x_i, m + \theta(T_e - m)]}$; x_n et y_n sont liées et la loi de probabilité élémentaire des deux variables est $f_n(x_n, y_n)\,dx_n dy_n$. Cherchons la limite de $E[\sqrt{n}(T_e - m)]^2$, c'est-à-dire de $\iint x_n^2 y_n^2 f_n(x_n, y_n)\,dx_n dy_n$.

On sait que $\iint y_n^2 f_n(x_n, y_n)\,dx_n dy_n$ existe quel que soit n et tend vers une limite qui est l'écart type de la loi de Gauss suivie à la limite par y_n. Donc, d'après le théorème de M. Fréchet $\iint y_n^2 f_n(x_n, y_n)\,dx_n\,dy_n$ tend uniformément, quel que soit n, vers sa limite pour y_n croissant indéfiniment. On sait d'autre part que x_n converge en probabilité vers $\frac{1}{a}\cdot$

Supposons que x_n soit borné supérieurement en valeur absolue par M. Nous allons dans ce cas montrer que

$$\lim \iint x_n^2 y_n^2 f_n(x_n, y_n)\,dx_n\,dy_n$$

tend vers la quantité

$$\frac{1}{a^2} \lim \iint y_n^2 f_n(x_n, y_n)\,dx_n\,dy_n.$$

Une représentation graphique aidera encore à suivre le raisonnement :

Nous partagerons la quantité

$$\iint y_n^2 \left(x_n^2 - \frac{1}{a^2}\right) f_n(x_n, y_n)\,dx_n\,dy_n$$

en trois parties, I_1 pour y_n supérieur en valeur absolue à N tel que

$$\iint y_n^2 f_n(x_n, y_n)\, dx_n\, dy_n$$

soit inférieur à ε pour $|y_n| > N$, ceci quel que soit n, I_2 pour $|y_n| < N$ et x_n

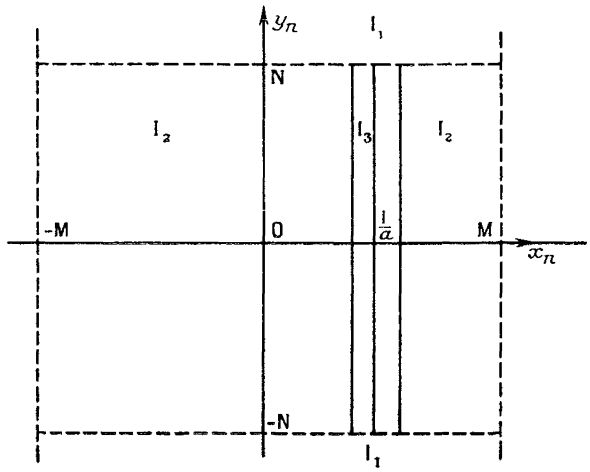

Fig. 2.

extérieur à l'intervalle $\frac{1}{a} - \eta$, $\frac{1}{a} + \eta$, I_3 pour $|y_n| < N$ et x_n compris dans l'intervalle $\frac{1}{a} - \eta$, $\frac{1}{a} + \eta$. On a les inégalités

$$I_1 < \left(M^2 - \frac{1}{a^2}\right) \iint_{|y_n|>N} y_n^2 f_n(x_n, y_n)\, dx_n\, dy_n < \left(M^2 - \frac{1}{a^2}\right) \varepsilon = \varepsilon_1,$$

$$I_2 < \left(M^2 - \frac{1}{a^2}\right) N^2 \left[\int_{-N}^{+N} \int_{-M}^{\frac{1}{a}-\eta} f_n(x_n, y_n)\, dx_n\, dy_n + \int_{-N}^{+N} \int_{\frac{1}{a}+\eta}^{+M} f_n(x_n, y_n)\, dx_n\, dy_n\right]$$
$$< \left(M^2 - \frac{1}{a^2}\right) N^2 \varepsilon' = \varepsilon_2,$$

en vertu de la convergence en probabilité de x_n vers $\frac{1}{a}$. Et

$$I_3 < N^2 \left[\left(\frac{1}{a} + \eta\right)^2 - \frac{1}{a^2}\right] \int_{-N}^{+N} \int_{\frac{1}{a}-\eta}^{\frac{1}{a}+\eta} f_n(x_n, y_n)\, dx_n\, dy_n < N^2 \left[\left(\frac{1}{a} + \eta\right)^2 - \frac{1}{a^2}\right] = \varepsilon_3.$$

Sous l'hypothèse que $\frac{1}{n} \sum \frac{\partial v_i}{\partial m}$ reste quel que soit n supérieur en valeur absolue à un nombre donné pour toute valeur de x et de m $\left(\text{ce qui est réalisé si l'on a uniformément } \frac{\partial v_i}{\partial m} > \mathrm{A} \text{ ou } \frac{\partial v_i}{\partial m} < \mathrm{B}, \text{ avec } \mathrm{A} > 0 \text{ et } \mathrm{B} < 0\right)$, on voit que la limite de $\mathrm{E}[\sqrt{n}(\mathrm{T}_c - m)]^2$ est l'écart-type de la loi-limite de Gauss correspondante. Cette condition est, bien entendu, loin d'être nécessaire. On voit immédiatement qu'elle entraîne que la limite de $\mathrm{E}[\sqrt{n}(\mathrm{T}_c - m)]$ est bornée puisque cette quantité est en valeur absolue inférieure à $\sqrt{\lim \mathrm{E}[\sqrt{n}(\mathrm{T}_c - m)]^2}$. Par conséquent $\mathrm{E}(\mathrm{T}_c) = m$.

Considérons l'équation intégrale suivante :

$$\int \cdots \int \cdots \varphi(x_1, \ldots, x_n, \ldots) f(x_1, m) \ldots f(x_n, m) \ldots dx_1 \ldots dx_n \ldots = m.$$

On voit que la méthode que nous avons employée permet d'obtenir toute une famille de solutions différentes de la solution banale :

$$\varphi = \frac{\lim \Sigma u(x_i)}{n} \qquad \text{avec} \qquad \int u(x) f(x, m)\, dx = m.$$

CHAPITRE IV.

PRÉCISION DE CES DIVERSES STATISTIQUES.

Dans ce Chapitre sera examinée la précision avec laquelle peut être atteint le paramètre inconnu du problème de l'estimation, ce que l'on peut appeler la partie principale de l'infiniment petit probable en $\frac{1}{n}$ qu'est la quantité $\mathrm{T}_n - m$.

Nous introduirons tout d'abord une représentation commode imaginée par R.-A. Fisher. Il s'agit de la notion d'espace statistique. A une expérience ayant fourni les résultats $x_1, x_2, \ldots, x_n$ sera associé dans l'espace à n dimensions un point de coordonnées $x_1, x_2, \ldots, x_n$. Cet espace, quand n augmente indéfiniment, sera l'espace « statistique ». On peut remarquer sa structure analogue à celle de l'espace de Hilbert, sous certaines conditions : si $\mathrm{E}(x^2)$ existe et est égal à 1 pour fixer les idées, la probabilité pour que $\frac{\Sigma x_i^2}{n}$ soit très différente de 1 tend vers zéro quand n augmente indéfiniment. Par conséquent, sous la condition indiquée, l'espace dont les points ont pour coordonnées

$\frac{x_1}{\sqrt{n}}, \frac{x_2}{\sqrt{n}}, \ldots, \frac{x_n}{\sqrt{n}}$ converge en probabilité vers l'espace de Hilbert quand n croît indéfiniment.

Effectuer une estimation consiste à choisir une suite de fonctions $f_1(x_1), \ldots, f_n(x_1, \ldots, x_n), \ldots$; les résultats fournis par une suite E_0 de n expériences étant $x_1^0, \ldots, x_n^0, \ldots$, la valeur $f_n(x_1^0, \ldots, x_n^0)$ sera celle adoptée pour m. Plaçons-nous dans le cas de n expériences. Les surfaces $m_1 = f_n(x_1, \ldots, x_n)$ seront appelées surfaces « isostatistiques », ce seront celles tout le long desquelles l'estimation restera la même. Surface doit être pris ici dans un sens très large. Il se peut que l'on ait affaire à un ensemble discontinu de points (ce sera le cas si la variable aléatoire ne peut prendre qu'un nombre fini, ou une infinité discontinue de valeurs).

La probabilité pour que m_1 soit une valeur donnée sera $d\Phi(m_1, m)$ (m restant constant). C'est la masse comprise entre deux surfaces isostatistiques infiniment voisines de la surface m_1. On aura évidemment :

$$\int d\Phi(m_1, m) = 1, \tag{A}$$

quel que soit m.

La position du point $M(x_1, x_2, \ldots, x_n)$ sur cette surface sera donnée par la connaissance de $n-1$ paramètres $\xi_1, \xi_2, \ldots, \xi_{n-1}$, ces $n-1$ variables pouvant d'ailleurs être quelques-unes des coordonnées $x_1, x_2, \ldots, x_n$. La probabilité liée pour que M occupe une position déterminée sur la surface, m_1 étant fixé, sera $d^{n-1}H(\xi_1, \ldots, \xi_{n-1}, m_1, m)$ (m_1 et m étant constants). On a également

$$\underbrace{\int \cdots \int}_{n-1} d^{n-1}H(\xi_1, \xi_2, \ldots, \xi_{n-1}, m_1, m) = 1 \tag{B}$$

pour toutes valeurs de m_1 et m.

Si l'on suppose pour simplifier l'écriture que Φ a une dérivée φ en m_1 et H une dérivée h d'ordre $n-1$ en $\xi_1, \ldots, \xi_{n-1}$, on peut écrire l'identité suivante :

$$f(x_1, m) \ldots f(x_n, m) = \varphi(m_1, m)\, h(\xi_1, \ldots, \xi_{n-1}, m_1, m) \tag{1}$$

en exprimant de deux façons différentes la probabilité qu'a le point M d'occuper une position déterminée. Le passage d'une forme à l'autre s'effectue en substituant dans le premier membre les expressions de $x_1, x_2, \ldots, x_n$ en fonction de $m_1, \xi_1, \ldots, \xi_{n-1}$ et en multipliant le tout par une fonction de $m_1, \xi_1, \ldots, \xi_{n-1}$

indépendante de m qui sera dans le cas de variables continues le déterminant fonctionnel $\frac{D(x_1, x_2, \ldots, x_n)}{D(m_1, \xi_1, \ldots, \xi_{n-1})}$.

Supposons que f, φ, h soient dérivables par rapport à m, quelles que soient les autres variables (sauf peut-être en excluant un ensemble de mesure nulle de tout leur champ de variation). Si, de plus, les conditions de convergence analogues à celles que nous avons supposées précédemment sont satisfaites, alors

$$\text{(C)} \qquad \int \frac{\partial f(x_i, m)}{\partial m} dx_i = 0, \qquad \int \frac{\partial \varphi(m_1, m)}{\partial m} dm_1 = 0, \qquad \underbrace{\int \cdots \int}_{n-1} \frac{\partial h}{\partial m} d\xi_1 \ldots d\xi_{n-1} = 0.$$

Prenons la dérivée logarithmique par rapport à m des deux membres de l'identité (1) et élevons au carré, il viendra, en raison de l'absence de m dans le déterminant fonctionnel,

$$\sum_{i=1}^{n} \left(\frac{1}{f(x_i, m)} \frac{\partial f(x_i, m)}{\partial m} \right)^2 + 2 \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{1}{f(x_i, m)} \frac{1}{f(x_j, m)} \frac{\partial f(x_i, m)}{\partial m} \frac{\partial f(x_j, m)}{\partial m}$$
$$= \left(\frac{1}{\varphi} \frac{\partial \varphi}{\partial m} \right)^2 + \left(\frac{1}{h} \frac{\partial h}{\partial m} \right)^2 + \frac{2}{\varphi h} \frac{\partial \varphi}{\partial m} \frac{\partial h}{\partial m}.$$

Si nous multiplions membre à membre cette identité par la relation (1), il est facile de voir en vertu de toutes les égalités (A), (B), (C) que

$$\sum_{i=1}^{n} \mathrm{E} \left[\frac{1}{f(x_i, m)} \frac{\partial f(x_i, m)}{\partial m} \right]^2 = n g(m) = \mathrm{E} \left[\frac{1}{\varphi} \frac{\partial \varphi}{\partial m} \right]^2 + \mathrm{E} \left[\frac{1}{h} \frac{\partial h}{\partial m} \right]^2,$$

car les espérances mathématiques de tous les termes rectangles sont nulles. Il en résultera le fait important suivant :

$$\text{(R)} \qquad g(m) \geqq \frac{1}{n} \mathrm{E} \left[\frac{1}{\varphi} \frac{\partial \varphi}{\partial m} \right]^2,$$

l'égalité n'étant possible que si $\mathrm{E}\left(\frac{1}{h} \frac{\partial h}{\partial m}\right)^2$ est nulle, ce qui entraîne $\frac{\partial h}{\partial m} \equiv 0$. Ce cas qui est celui des statistiques exhaustives ou « sufficient » sera examiné plus loin.

Supposons que quel que soit n la statistique m_1 suive la loi de Gauss d'écart-type s_n. Il est facile de voir en effectuant les calculs que $\mathrm{E}\left[\frac{1}{\varphi} \frac{\partial \varphi}{\partial m}\right]^2 = \frac{1}{s_n^2}$, et par conséquent on aura : $s_n^2 \geq \frac{1}{ng(m)}$. Pour chaque valeur de n l'écart-type de la loi est borné inférieurement. On ne peut dépasser une précision limite. Mais le caractère continuellement gaussien est trop exceptionnel pour que nous puissions borner là nos résultats. Nous allons nous efforcer de trouver un énoncé valable pour le cas où la loi-limite seule est gaussienne, et pour cela effectuons tout d'abord un changement de variables. Posons $\mu = \sqrt{n}(m_1 - m)$, la loi $\varphi(m_1, m)\,dm_1$ devient $\frac{1}{\sqrt{n}}\,\varphi\left(m + \frac{\mu}{\sqrt{n}}, m\right)d\mu$.

On a donc

$$\frac{1}{\varphi}\frac{\partial\varphi}{\partial m} = \frac{1}{\varphi}\frac{\delta\varphi}{\delta m} - \sqrt{n}\,\frac{1}{\varphi}\frac{\delta\varphi}{\delta\mu}$$

(le signe δ indiquant la dérivation après le changement de variables).

Il s'ensuit que

$$\frac{1}{n}\mathrm{E}\left[\frac{1}{\varphi}\frac{\partial\varphi}{\partial m}\right]^2 = \frac{1}{n}\mathrm{E}\left[\frac{1}{\varphi}\frac{\delta\varphi}{\delta\mu}\right]^2 + \mathrm{E}\left[\frac{1}{\varphi}\frac{\delta\varphi}{\delta\mu}\right]^2 - \frac{2}{\sqrt{n}}\mathrm{E}\left[\frac{1}{\varphi^2}\frac{\delta\varphi}{\delta\mu}\frac{\delta\varphi}{\delta m}\right].$$

Si $\mathrm{E}\left[\frac{1}{\varphi^2}\frac{\delta\varphi}{\delta\mu}\frac{\delta\varphi}{\delta m}\right]$ reste borné quel que soit n, le troisième terme du second membre peut être rendu inférieur à ε. On aura donc l'inégalité $\Big($puisque $\mathrm{E}\left[\frac{1}{\varphi}\frac{\delta\mu}{\delta m}\right]^2$ est évidemment positif$\Big)$

$$\lim \mathrm{E}\left[\frac{1}{\varphi}\frac{\delta\varphi}{\delta\mu}\right]^2 \leqq \lim \frac{1}{n}\mathrm{E}\left[\frac{1}{\varphi}\frac{\partial\varphi}{\partial m}\right]^2 \leqq g(m).$$

On peut se dispenser d'exiger la condition de convergence que nous avons énoncée : il suffit qu'à partir d'une certaine valeur de n, $\mathrm{E}\left[\frac{1}{\varphi^2}\frac{\delta\varphi}{\delta\mu}\frac{\delta\varphi}{\delta m}\right]$ soit constamment négatif.

Dans $\mathrm{E}\left[\frac{1}{\varphi}\frac{\delta\varphi}{\delta\mu}\right]^2$ intervient seulement la dérivée par rapport à la variable et non plus par rapport au paramètre, ce qui est une notable simplification. Écrivons $\varphi(\mu)$ sous la forme

$$\frac{1}{\sqrt{2\pi}\sigma}\left\{\exp\left[-\frac{\mu^2}{2\sigma^2}\right]\right\}[1 + \mathrm{A}_n(\mu)] = \mathrm{G}(\mu)[1 + \mathrm{A}_n(\mu)] \qquad \text{avec} \qquad \int \mathrm{G}(\mu)[1 + \mathrm{A}_n(\mu)]\,d\mu = 1$$

pour toutes valeurs de n.

On aura alors l'identité

$$\frac{1}{\varphi}\frac{\delta\varphi}{\delta\mu} = \frac{1}{\mathrm{G}}\frac{\partial\mathrm{G}}{\partial\mu} + \frac{1}{1 + \mathrm{A}_n(\mu)}\frac{\partial\mathrm{A}_n}{\partial\mu} = -\frac{\mu}{\sigma^2} + \frac{1}{1 + \mathrm{A}_n}\frac{\partial\mathrm{A}_n}{\partial\mu}$$

et l'on verra facilement en effectuant les calculs que

$$\mathrm{E}\left[\frac{1}{\varphi}\frac{\partial\varphi}{\partial\mu}\right]^2 = \frac{1}{\sigma^2} + \int \frac{\mu^2}{\sigma^4}\mathrm{GA}_n\,d\mu + 2\int \frac{\partial \mathrm{G}}{\partial\mu}\frac{\partial \mathrm{A}_n}{\partial\mu}\,d\mu + \mathrm{E}\left[\frac{\frac{\partial \mathrm{A}_n}{\partial\mu}}{1+\mathrm{A}_n}\right]^2.$$

En intégrant par parties

$$\int_{-\infty}^{+\infty} \frac{\partial \mathrm{G}}{\partial\mu}\frac{\partial \mathrm{A}_n}{\partial\mu}\,d\mu = \left[\mathrm{A}_n\frac{\partial \mathrm{G}}{\partial\mu}\right]_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} \mathrm{A}_n\frac{\partial^2 \mathrm{G}}{\partial\mu^2}\,d\mu.$$

Or,

$$\frac{\partial^2 \mathrm{G}}{\partial\mu^2} = -\frac{1}{\sigma^2}\mathrm{G}(\mu) + \frac{\mu^2}{\sigma^4}\mathrm{G}(\mu).$$

Comme $\int \mathrm{A}_n \mathrm{G}\,d\mu = 0$ quel que soit n, on a, toutes simplifications faites,

$$\mathrm{E}\left[\frac{1}{\varphi}\frac{\partial\varphi}{\partial\mu}\right]^2 = \frac{1}{\sigma^2} - \int_{-\infty}^{+\infty} \mu^2\mathrm{GA}_n\,d\mu + \left[-\frac{\mu}{\sigma^2}\mathrm{A}_n\mathrm{G}\right]_{-\infty}^{+\infty} + \mathrm{E}\left[\frac{\frac{\partial \mathrm{A}_n}{\partial\mu}}{1+\mathrm{A}_n}\right]^2.$$

Supposons que le moment d'ordre 2 de μ ait pour limite σ^2. On aura donc

$$\lim \int_{-\infty}^{+\infty} \mu^2\mathrm{GA}_n\,d\mu = 0.$$

Du fait que cette intégrale a un sens à partir d'un certain rang n, il résulte que le moment d'ordre 1 existe au moins à partir de ce rang, et par conséquent que la limite de $\mu\mathrm{A}_n\mathrm{G}$ pour μ infini en valeur absolue est nulle. Donc, on a

$$\frac{1}{\sigma^2} \leqq \lim \mathrm{E}\left[\frac{1}{\varphi}\frac{\partial\varphi}{\partial\mu}\right]^2 \leqq g(m).$$

On aurait, d'ailleurs, encore le droit d'écrire la première inégalité, même si le moment d'ordre 2 n'avait pas pour limite σ^2. On peut voir que si, à partir d'un certain rang, ce moment est *inférieur* à σ^2, l'inégalité est encore valable. On en conclut donc $\sigma^2 \geqq \frac{1}{g(m)}\cdot$ Si l'on remarque que cette quantité $\frac{1}{g(m)}$ est précisément l'écart-type de la loi-limite de la statistique du maximum de vraisemblance, il en résulte le théorème extrêmement important :

Sous certaines conditions, les lois-limites des estimations gaussiennes ont leur précision bornée supérieurement par celle du « maximum of likelihood ».

C'est donc avec juste raison que R.-A. Fisher a donné aussi à cette dernière le nom d'*optimum*.

Cas de l'estimation de plusieurs paramètres. — On peut également formuler un résultat analogue dans le cas où plusieurs paramètres, deux pour simplifier les calculs, sont l'objet de l'estimation.

Soit $f(x, a, b)$ dans ce cas la loi élémentaire de probabilité avec

$$\int f(x, a, b)\, dx = 0 \tag{1}$$

et soient a_1 et b_1 les deux statistiques

$$a_1 = a_1(x_1, x_2, \ldots, x_n). \qquad b_1 = b_1(x_1, x_2, \ldots, x_n).$$

Ici encore nous effectuerons le changement de variables $a_1, b_1, \xi_1, \xi_2, \ldots, \xi_{n-2}$. Cela suppose dans le cas où l'on prend pour ξ les $n-2$ derniers résultats $x_2, \ldots, x_n$ que $\frac{D(a_1, b_1)}{D(x_1, x_2)}$ n'est pas identiquement nul. Comme les deux variables x_1 et x_2 sont deux quelconques d'entre toutes, il ne faut donc pas que l'on ait $g(a, b) = 0$, et cela est suffisant.

On trouve comme précédemment pour la loi de probabilité de la série d'expériences

$$P\, d\omega = \prod_1^n f(x_i, a, b)\, d\omega \equiv \varphi(a_1, b_1, a, b)\, da_1\, db_1\, h[\xi_1, \ldots, \xi_{n-2}, a_1, b_1, a, b]\, d\omega_1,$$

$d\omega$ et $d\omega_1$ étant les deux différentielles appropriées.

On aura encore

$$\iint \varphi(a_1, b_1, a, b)\, da_1\, db_1 = \int h\, d\omega_1 = 1 \qquad \text{(quels que soient } a \text{ et } b). \tag{1'}$$

Si l'on pose

$$df = \frac{\partial f}{\partial a} da + \frac{\partial f}{\partial b} db$$

et les égalités analogues pour P, φ et h, on voit que $E\left[\frac{dP}{P}\right]^2 = nE\left[\frac{df}{f}\right]^2$ à cause de (1).

$E\left[\frac{df}{f}\right]^2$ est une forme quadratique en da et db que nous appellerons $g(da, db)$. A cause de (1'), on aura de même

$$E\left[\frac{d\varphi}{\varphi}\right]^2 \leqq E\left[\frac{dP}{P}\right]^2 = ng, \tag{R}$$

pour tout da et db, l'égalité n'ayant lieu que pour les statistiques exhaustives, dans lesquelles h ne contient ni a, ni b, et pour celles-là seulement.

Comme plus haut on effectuera le changement de variables aléatoires

$$\alpha = \sqrt{n}(a_1 - a) \qquad \text{et} \qquad \beta = \sqrt{n}(b_1 - b).$$

Il convient de faire maintenant complètement les calculs, car ils diffèrent légèrement du cas d'un seul paramètre

$$\frac{\partial \varphi}{\partial a} = \frac{\delta \varphi}{\delta a} - \sqrt{n}\frac{\delta \varphi}{\delta \alpha},$$

et de même pour b et β, δ ayant la même signification que ci-dessus.

Si l'on pose

$$\delta\varphi = \frac{\delta \varphi}{\delta a}da + \frac{\delta \varphi}{\delta b}db \qquad \text{et} \qquad \Delta\varphi = \frac{\delta \varphi}{\delta \alpha}da + \frac{\delta \varphi}{\delta \beta}db.$$

On peut écrire

$$\frac{1}{n}\mathrm{E}\left[\frac{d\varphi}{\varphi}\right]^2 = \frac{1}{n}\mathrm{E}\left[\frac{\delta\varphi}{\varphi}\right]^2 + \mathrm{E}\left[\frac{\Delta\varphi}{\varphi}\right]^2$$
$$- \frac{2}{\sqrt{n}}\left\{\mathrm{E}\left[\frac{1}{\varphi^2}\frac{\delta\varphi}{\delta a}\frac{\delta\varphi}{\delta\alpha}\right]da^2 + \mathrm{E}\left[\frac{1}{\varphi^2}\frac{\delta\varphi}{\delta b}\frac{\delta\varphi}{\delta\beta}\right]db^2 + \mathrm{E}\left[\frac{1}{\varphi^2}\left(\frac{\delta\varphi}{\delta a}\frac{\delta\varphi}{\delta\beta} + \frac{\delta\varphi}{\delta b}\frac{\delta\varphi}{\delta\alpha}\right)\right]da\,db\right\},$$

$\mathrm{E}\left[\frac{\delta\varphi}{\varphi}\right]^2$ étant une forme positive il suffira que les trois espérances mathématiques coefficients de da^2, db^2, $dadb$ restent bornées pour que l'on puisse écrire

$$\mathrm{E}\left[\frac{\Delta\varphi}{\varphi}\right]^2 \leqq g(da, db).$$

Supposons que $\varphi(\alpha, \beta)$ soit continuellement identique à la loi de Gauss

$$\mathrm{G}(\alpha, \beta) = \frac{d\alpha\, d\beta}{2\pi\sqrt{1 - r^2}\sigma_\alpha\sigma_\beta}\exp\left\{-\frac{1}{2(1 - r^2)}\left(\frac{\alpha^2}{\sigma_\alpha^2} + \frac{\beta^2}{\sigma_\beta^2} - 2r\frac{\alpha\beta}{\sigma_\alpha\sigma_\beta}\right)\right\}.$$

On aura alors

$$\mathrm{E}\left[\frac{\Delta\varphi}{\varphi}\right]^2 = \frac{1}{1 - r^2}\left(\frac{da^2}{\sigma_\alpha^2} + \frac{db^2}{\sigma_\beta^2} - 2r\frac{da\,db}{\sigma_\alpha\sigma_\beta}\right) \leqq g(da, db).$$

Sinon, nous écrirons $\varphi(\alpha, \beta)$ sous la forme

$$\varphi(\alpha, \beta) = \mathrm{G}(\alpha, \beta)[1 + \mathrm{A}_n(\alpha, \beta)],$$

d'où l'on tire

$$\frac{\Delta\varphi}{\varphi} = \frac{\Delta\mathrm{G}}{\mathrm{G}} + \frac{\Delta\mathrm{A}_n(\alpha, \beta)}{1 + \mathrm{A}_n}$$

et

$$\mathrm{E}\left[\frac{\Delta\varphi}{\varphi}\right]^2 = \gamma(da, db) + \iint\left(\frac{\Delta\mathrm{G}}{\mathrm{G}}\right)^2\mathrm{G}\mathrm{A}_n\,d\alpha\,d\beta + 2\iint\Delta\mathrm{G}\,\Delta\mathrm{A}_n\,d\alpha\,d\beta + \mathrm{E}\left[\frac{\Delta\mathrm{A}_n}{1 + \mathrm{A}_n}\right]^2,$$

en appelant $\gamma(da, db)$ la forme quadratique que nous avons obtenue ci-dessus. $\mathrm{E}\left[\frac{\Delta \mathrm{A}_n}{1+\mathrm{A}_n}\right]^2$ étant positif, seul nous intéresse le comportement des deux intégrales. En se servant encore de l'intégration par parties et en se rappelant que l'on a, quel que soit n, $\iint \mathrm{GA}_n\, d\alpha\, d\beta = 0$, il vient

$$\begin{aligned}
&\iint \left(\frac{\Delta \mathrm{G}}{\mathrm{G}}\right)^2 \mathrm{GA}_n\, d\alpha\, d\beta + 2\iint \Delta\mathrm{G}\, \Delta\mathrm{A}_n\, d\alpha\, d\beta \\
&= \frac{-1}{1-r^2}\left\{\iint \left(\frac{\alpha^2}{\sigma_\alpha^2} da^2 + \frac{\beta^2}{\sigma_\beta^2} db^2 + \frac{2\alpha\beta}{\sigma_\alpha^2\sigma_\beta^2} da\, db\right) \mathrm{GA}_n\, d\alpha\, d\beta\right\} \\
&\quad + 2\left\{ da^2 \int_{-\infty}^{+\infty} \left(\mathrm{A}_n \frac{\partial \mathrm{G}}{\partial \alpha}\right)_{\alpha=-\infty}^{\alpha=+\infty} d\beta + db^2 \int_{-\infty}^{+\infty} \left(\mathrm{A}_n \frac{\partial \mathrm{G}}{\partial \beta}\right)_{\beta=-\infty}^{\beta=+\infty} d\alpha \right. \\
&\qquad \left. - da\, db \left[\int_{-\infty}^{+\infty} \left(\mathrm{A}_n \frac{\partial \mathrm{G}}{\partial \alpha}\right)_{\beta=-\infty}^{\beta=+\infty} d\alpha + \int_{-\infty}^{+\infty} \left(\mathrm{A}_n \frac{\partial \mathrm{G}}{\partial \beta}\right)_{\alpha=-\infty}^{\alpha=+\infty} d\beta\right]\right\}.
\end{aligned}$$

Si $\mathrm{E}(\alpha^2) \to \sigma_\alpha^2$, $\mathrm{E}(\beta^2) \to \sigma_\beta^2$ et $\mathrm{E}(\alpha\beta) \to r\sigma_\alpha\sigma_\beta$, on voit facilement que la première parenthèse tend vers zéro, quels que soient da et db donnés. Examinons les diverses intégrales de la seconde parenthèse

$$\int_{-\infty}^{+\infty} \left(\mathrm{A}_n \frac{\partial \mathrm{G}}{\partial \alpha}\right)_{\alpha=-\infty}^{\alpha=+\infty} d\beta = -\frac{1}{(1-r^2)\sigma_\alpha}\left[\frac{\alpha}{\sigma_\alpha}\int_{-\infty}^{+\infty} \mathrm{A}_n \mathrm{G}\, d\beta - \frac{r}{\sigma_\beta}\int_{-\infty}^{+\infty} \beta \mathrm{A}_n \mathrm{G}\, d\beta\right]_{\alpha=-\infty}^{\alpha=+\infty}.$$

Le fait que $\mathrm{E}(\alpha^2)$ tende vers σ_α^2 entraîne l'existence à partir d'un certain rang n_1 de $\mathrm{E}(\alpha^2)$, et par suite celle de $\mathrm{E}(\alpha)$. De même $\mathrm{E}(\beta)$ converge à partir d'un certain rang, n_2. Donc pour $n >$ p. g. (n_1, n_2) $\iint \alpha \mathrm{GA}_n\, d\alpha\, d\beta$ et $\iint \beta \mathrm{GA}_n\, d\alpha\, d\beta$ ont un sens. Ces deux intégrales peuvent se mettre sous la forme

$$\int_{-\infty}^{+\infty} \alpha\, d\alpha \int_{-\infty}^{+\infty} \mathrm{GA}_n\, d\beta \quad \text{et} \quad \int_{-\infty}^{+\infty} d\alpha \int_{-\infty}^{+\infty} \beta \mathrm{GA}_n\, d\beta.$$

L'existence de ces deux expressions entraîne, pour α infini en valeur absolue, la nullité de

$$\alpha \int_{-\infty}^{+\infty} \mathrm{GA}_n\, d\beta \quad \text{et} \quad \int_{-\infty}^{+\infty} \beta \mathrm{GA}_n\, d\beta.$$

Le coefficient de da^2 est donc nul, ainsi bien entendu que ceux de db^2 et de $da\,db$. Ici encore on peut donc écrire

$$(\mathrm{A}) \qquad \gamma(da, db) = \frac{1}{1-r^2}\left[\frac{da^2}{\sigma_\alpha^2} + \frac{db^2}{\sigma_\beta^2} - 2r\frac{da\, db}{\sigma_\alpha\sigma_\beta}\right] \leqq g(da, db).$$

On peut, comme dans le cas d'un seul paramètre, exprimer cette inégalité finale sous des conditions moins restrictives que celles que nous nous sommes imposées : bien qu'analogues elles seraient toutefois moins intuitives et c'est pourquoi il nous a semblé inutile de les exposer.

Les conditions exprimées s'étendent maintenant sans difficulté au cas où l'on a plus de deux paramètres. Nous allons donner une interprétation géométrique simple du résultat obtenu. Considérons les deux coniques

$$\gamma(\alpha, \beta) = \frac{1}{1-r^2}\left[\frac{\alpha^2}{\sigma_\alpha^2} + \frac{\beta^2}{\sigma_\beta^2} - 2r\frac{\alpha\beta}{\sigma_\alpha\sigma_\beta}\right] = 1$$

et

$$g(\alpha, \beta) = \alpha^2 E\left[\frac{1}{f}\frac{\partial f}{\partial a}\right]^2 + \beta^2 E\left[\frac{1}{f}\frac{\partial f}{\partial b}\right]^2 + 2\alpha\beta E\left[\frac{1}{f^2}\frac{\partial f}{\partial a}\frac{\partial f}{\partial b}\right] = 1.$$

Ce sont deux ellipses en vertu de la relation classique de Schwarz pour la seconde, et du fait que r est inférieur à 1 en module pour la première. $\Big($L'existence de la seconde conique est seulement liée à l'existence de $E\left[\frac{1}{f}\frac{\partial f}{\partial a}\right]^2$ et de $E\left[\frac{1}{f}\frac{\partial f}{\partial b}\right]^2$, celle de $E\left[\frac{1}{f}\frac{\partial f}{da}\frac{\partial f}{db}\right]$ découlant des deux premières$\Big)$. L'inégalité (A) exprime que la première ellipse ne peut avoir de points intérieurs à la seconde,

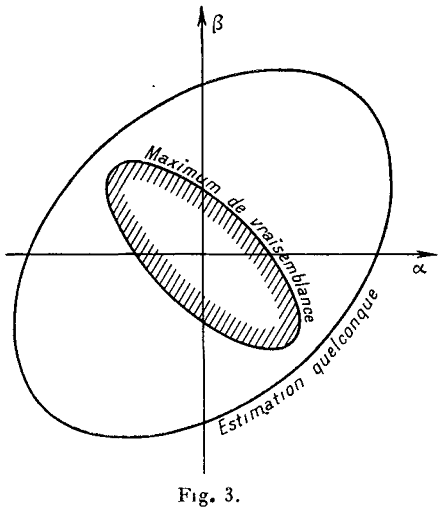

Fig. 3.

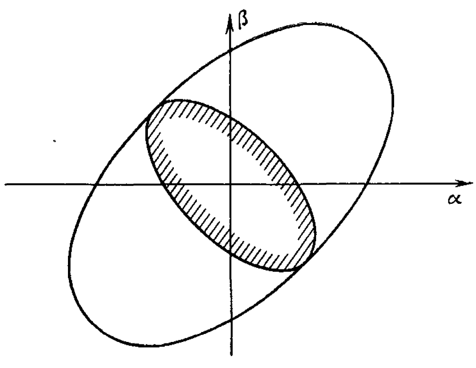

Fig. 4.

qui représente la distribution du maximum de vraisemblance. On sait que la probabilité pour que le point α, β satisfaisant à une loi de Gauss

K $\exp\left\{-\frac{1}{2}\varphi(\alpha,\beta)\right\}$ se trouve à l'intérieur de l'ellipse $\varphi(\alpha,\beta)=1$, est une quantité qui ne dépend pas de la forme de φ.

Dans l'estimation par la méthode de M. Fisher, le point α, β a donc une probabilité de se trouver dans une certaine ellipse égale à celle qu'il aurait de se trouver dans une ellipse tout entière extérieure, si l'on faisait l'estimation par une autre méthode. La proposition de M. Fisher vaut donc pour plusieurs paramètres.

Le plus souvent d'ailleurs on s'arrange pour qu'une combinaison linéaire de α et β, $P\alpha+Q\beta$ fournisse le maximum de précision, et les deux ellipses ont la disposition de la figure 4.

Cas d'une limite non gaussienne. — Il est possible, dans le cas où l'estimation ne suit pas à la limite une loi de Gauss (ce qui peut se présenter en vertu de l'étude faite au Chapitre précédent), d'affirmer néanmoins certains résultats concernant sa précision. Une fois effectué le changement de variable $\mu=\sqrt{n}(m_1-m)$, on se trouve sous des conditions assez larges, ramené à l'inégalité

$$E\left[\frac{1}{F_n}\frac{\partial F_n}{\partial\mu}\right]^2 \leqq g = E\left[\frac{1}{f}\frac{\partial f}{\partial m}\right]^2.$$

Supposons qu'il y ait une loi-limite H telle que $E\left[\frac{1}{H}\frac{\partial H}{\partial\mu}\right]^2$ existe et posons encore $F_n=H(1+A_n)$

$$E\left[\frac{1}{F_n}\frac{\partial F_n}{\partial\mu}\right]^2 = E\left[\frac{1}{H}\frac{\partial H}{\partial\mu}\right]^2 + \int\left(\frac{1}{H}\frac{\partial H}{\partial\mu}\right)^2 HA_n\,d\mu$$
$$+2\int\frac{\partial H}{\partial\mu}\frac{\partial A_n}{\partial\mu}\,d\mu+\int\left(\frac{\frac{\partial A_n}{\partial\mu}}{1+A_n}\right)^2 H(1+A_n)\,d\mu.$$

Le dernier terme étant constamment positif on aura évidemment

$$E\left[\frac{1}{H}\frac{\partial H}{\partial\mu}\right]^2+\int\left(\frac{1}{H}\frac{\partial H}{\partial\mu}\right)^2 HA_n\,d\mu+2\int\frac{\partial H}{\partial\mu}\frac{\partial A_n}{\partial\mu}\,d\mu\leqq g.$$

Supposons que toutes les quantités $E\left[\frac{1}{F_n}\frac{\partial\mu}{\partial F_n}\right]^2$ tendent *également* vers leurs valeurs pour $|\mu|$ grand quel que soit n, on voit que l'on peut borner les intervalles d'intégration à deux nombres finis et écrire

$$E\left[\frac{1}{H}\frac{\partial H}{\partial\mu}\right]^2+\int_{-M'}^{+M}\left(\frac{1}{H}\frac{\partial H}{\partial\mu}\right)^2 HA_n\,d\mu+2\int_{-M'}^{+M}\frac{\partial H}{\partial\mu}\frac{\partial A_n}{\partial\mu}\,d\mu\leqq g+\varepsilon,$$

ε étant aussi petit qu'on le désire. Si F_n tend vers H, de telle sorte que A_n tende uniformément vers zéro dans un intervalle fini de μ,

$$\int_{-M'}^{+M} \left(\frac{1}{H}\frac{\partial H}{\partial \mu}\right)^2 HA_n\, d\mu < \eta E\left(\frac{1}{H}\frac{\partial H}{\partial \mu}\right)^2$$

et

$$\int_{-M'}^{+M} \frac{\partial H}{\partial \mu}\frac{\partial A_n}{\partial \mu}\, d\mu = \left[A_n \frac{\partial H}{\partial \mu}\right]_{-M}^{+M} - \int_{-M'}^{+M} A_n \frac{\partial^2 H}{\partial \mu^2}\, d\mu < 2\eta \left[\frac{\partial H}{\partial \mu}\right]_{-M'}^{+M}.$$

On en déduit donc à la limite que

$$\text{(1)} \qquad E\left[\frac{1}{H}\frac{\partial H}{\partial \mu}\right]^2 \leqq g.$$

On peut donner des exemples de lois de probabilité obéissant à cette double condition d'égale convergence et de limite uniforme. Supposons que l'on ait

$$F_n = H\left[1 + \sum_{i=1}^{p} f_i(n)\, A_i\right],$$

les A_i ne contenant pas n, et les $f_i(n)$ tendant vers zéro avec $\frac{1}{n}$. A partir d'une certaine valeur de n tous les $f_i(n)$ seront donc inférieurs à η, et il en résulte que les deux intégrales

$$\int_{-M'}^{+M} \left(\frac{1}{H}\frac{\partial H}{\partial \mu}\right)^2 H\left[\sum_{i=1}^{p} f_i(n) A_i\right] d\mu \quad \text{et} \quad \int_{-M'}^{+M} \frac{\partial H}{\partial \mu}\left[\sum_{i=1}^{p} f_i(n) \frac{\partial A_i}{\partial \mu}\right] d\mu$$

tendent également en n vers leur limite quand M et M′ croissent indéfiniment et que $\sum_{i=1}^{p} f_i(n)\, A_i$ tend uniformément vers zéro sur l'intervalle $-M'$, $+M'$.

La quantité $E\left[\frac{1}{H}\frac{\partial H}{\partial \mu}\right]^2$ joue pour les lois qui n'ont pas d'écart-type le même rôle que cette grandeur joue pour les lois qui en ont. En effet, si l'on effectue sur la loi un changement de variable $\mu = a\nu$, ce qui conduit à la loi I, on a

$$a^2 E\left[\frac{1}{H}\frac{\partial H}{\partial \mu}\right]^2 = E\left[\frac{1}{I}\frac{\partial I}{\partial \nu}\right]^2.$$

De plus, c étant un nombre donné

$$\text{(2)} \qquad \int_{-c}^{+c} H\, d\mu = \int_{-\frac{c}{a}}^{+\frac{c}{a}} I\, d\nu.$$

Donc, pour un type de loi déterminé (nous appelons type un ensemble de lois dérivant les unes des autres par le changement de variable $\mu = a\nu$) la grandeur $E\left[\frac{1}{H}\frac{\partial I}{\partial \nu}\right]^2$ croît quand croît la précision, puisque d'après l'égalité (2) à probabilité égale la variable ν est contenue dans un intervalle a fois plus petit que la variable μ. L'inégalité (1) prouve donc que pour un type déterminé, la précision est bornée. En particulier, si le maximum of likelihood ne suit pas à la limite une loi de Gauss $\Big($cela entraîne des irrégularités dans la continuité de $f(x, m)$, puisqu'on suppose $E\left[\frac{1}{f}\frac{\partial f}{\partial m}\right]^2$ existant$\Big)$ les estimations suivant à la limite une loi de probabilité du même type que lui, ont une précision moindre.

Il est facile d'étendre ces résultats au cas de plusieurs paramètres, on aura alors avec les mêmes notations que ci-dessus

$$\begin{aligned} E\left[\frac{1}{H(\alpha, \beta)}\frac{\partial H}{\partial \alpha}\right]^2 \alpha^2 + E\left[\frac{1}{H}\frac{\partial H}{\partial \beta}\right]^2 \beta^2 + 2E\left[\frac{1}{H^2}\frac{\partial H}{\partial \alpha}\frac{\partial H}{\partial \beta}\right]\alpha\beta \\ \leqq E\left[\frac{1}{f}\frac{\partial f}{\partial a}\right]^2 \alpha^2 + E\left[\frac{1}{f}\frac{\partial f}{\partial b}\right]^2 \beta^2 + 2E\left[\frac{1}{f^2}\frac{\partial f}{\partial a}\frac{\partial f}{\partial b}\right]\alpha\beta. \end{aligned}$$

Cas où le paramètre figure parmi les limites de la variable aléatoire. — Dans ce cas encore on a le droit d'écrire l'égalité fondamentale

$$\Pi f(x_i, m) = \varphi(m_1, m)\, h(\xi_1, \ldots, \xi_{n-1}, m_1, m)$$

en exprimant de deux façons différentes la probabilité d'un résultat $(x_1, \ldots, x_n)$.

En prenant la dérivée logarithmique des deux membres, on a toujours

$$\sum \frac{1}{f}\frac{\partial f}{\partial m}(x_i, m) = \frac{1}{\varphi}\frac{\partial \varphi}{\partial m} + \frac{1}{h}\frac{\partial h}{\partial m}.$$

Mais si on élève au carré et si l'on prend la valeur moyenne du résultat, on constate que les termes rectangles égaux à $[f(m, m)]^2$ ne sont en général pas nuls dans le premier membre. Ce terme est nul dans le second, si les limites de m_1 ne dépendent pas de m. Si l'on a $f(m, m) = 0$, le premier membre se réduit comme dans les cas précédents à une somme de carrés $\Big[$pour pouvoir prendre la dérivée logarithmique, il faut que $\frac{\partial f}{\partial m}(m, m)$ soit nul$\Big]$.

Le même raisonnement que ci-dessus prouve l'existence d'un maximum de précision, et sous les hypothèses que nous venons d'indiquer $\Big[f(m,m) = \frac{\partial f}{\partial m}(m,m) = 0,$

et estimation dont les limites ne dépendent pas de $m\big]$, on voit que ce maximum de précision est fourni par le maximum de vraisemblance dont l'écart-type est alors $\dfrac{1}{\sqrt{E\left[\frac{1}{f}\frac{\partial f}{\partial m}\right]^2}}$.

Remarque. — Si $f(m, m) = 0$ et $\frac{\partial f}{\partial m}(m, m) = 0$, on a évidemment $\frac{\partial f}{\partial x}(m, m) = 0$,

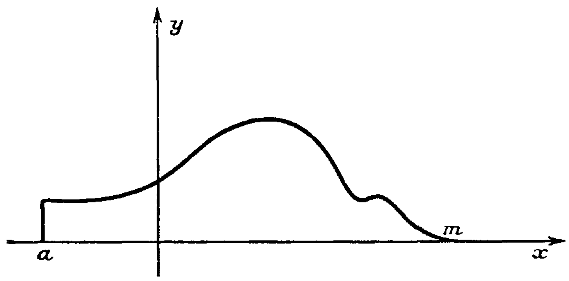

Fig. 5.

ce qui signifie que la courbe des probabilités élémentaires $f(x, m)$ doit être tangente à l'axe des x au point $x = m$.

Cas de résultats liés en chaînes. — L'étude que nous avons faite jusqu'ici est celle du comportement des quantités attachées à la relation (R), quand le nombre n d'expériences augmente indéfiniment.

Nous ne nous sommes jusqu'à présent préoccupé que des différentes formes qu'est susceptible de prendre cette relation, suivant la régularité des surfaces isostatistiques. Les paragraphes qui suivent sont consacrés à l'étude des problèmes que soulèvent les divers cas de probabilités en chaînes.

Supposons que les résultats $x_1, \ldots, x_n$, qui jusqu'ici restaient indépendants les uns des autres, soient liés en une chaîne que nous supposerons tout d'abord simple ; c'est-à-dire que la probabilité d'obtenir le résultat x_i à la $i^{\text{ième}}$ expérience quand le résultat de $i - 1^{\text{ième}}$ a été x_{i-1} est $f(x_{i-1}, x_i)\,dx_i$ avec

$$(1) \qquad \int f(x_{i-1}, x_i)\,dx_i = 1 \qquad \text{(quel que soit } x_{i-1}\text{)}.$$

Le résultat x_1 à la première mesure est fixé *a priori*. Dans ces conditions, la probabilité d'une suite de résultats $x_1, \ldots, x_n$ sera

$$f(x_1, x_2)\,f(x_2, x_3)\ldots f(x_{n-1}, x_n)\,dx_2\ldots dx_n.$$

Appelons $\mathcal{E}[\psi(x_i, \ldots, x_j)]$ l'expression suivante, fonction de x_1

$$\underbrace{\int \cdots \int}_{j-1} \psi(x_i, \ldots, x_j)\, f(x_1, x_2) \ldots f(x_{j-1}, x_j)\, dx_i \ldots dx_j,$$

j étant l'indice le plus élevé figurant dans ψ.

Nous supposerons toujours que chaque probabilité est fonction du paramètre m à estimer.

En employant la même méthode qu'au début de ce Chapitre, il est facile de voir que

$$\sum_{i=1}^{n} \frac{1}{f(x_{i-1}, x_i, m)} \frac{\partial f(x_{i-1}, x_i, m)}{\partial m} = \frac{1}{\varphi} \frac{\partial \varphi}{\partial m} + \frac{1}{h} \frac{\partial h}{\partial m}.$$

En élevant les deux membres au carré, en multipliant respectivement le premier par $f(x_1, x_2) \ldots f(x_{n-1}, x_n)$ et le second par φh et en intégrant entre les limites de variation, on retrouve évidemment au second membre la quantité $\mathrm{E}\left[\frac{1}{\varphi} \frac{\partial \varphi}{\partial m}\right]^2 + \mathrm{E}\left[\frac{1}{h} \frac{\partial h}{\partial m}\right]^2$, et au premier, en raison de la relation (1), une somme d'expressions de la forme

$$\mathcal{E}\left[\frac{1}{f(x_{i-1}, x_i, m)} \frac{\partial f(x_{i-1}, x_i, m)}{\partial m}\right]^2$$

et

$$2\,\mathcal{E}\left[\frac{1}{f(x_{j-1}, x_j, m)} \frac{\partial f(x_{j-1}, x_j, m)}{\partial m} \frac{1}{f(x_{k-1}, x_k, m)} \frac{\partial f(x_{k-1}, x_k, m)}{\partial m}\right].$$

Si la relation (1) est dérivable sous le signe $\int$ on aura évidemment

$$\int \frac{\partial f(x_{i-1}, x_i, m,)}{\partial m}\, dx_i = 0,$$

et par conséquent il est facile de voir que

$$\mathcal{E}\left[\frac{1}{f(x_{j-1}, x_j, m)} \frac{\partial f(x_{j-1}, x_j, m)}{\partial m} \frac{1}{f(x_{k-1}, x_k, m)} \frac{\partial f(x_{k-1}, x_k, m)}{\partial m}\right]$$

est nul ; comme dans le cas de variables indépendantes, on en déduira que

$$\frac{1}{n} \mathrm{E}\left[\frac{1}{\varphi} \frac{\partial \varphi}{\partial m}\right]^2 \leqq \frac{1}{n} \sum_{i=2}^{n} \mathcal{E}\left[\frac{1}{f(x_{i-1}, x_i, m)} \frac{\partial f(x_{i-1}, x_i, m)}{\partial m}\right]^2, \tag{2}$$

sous la condition que $\mathcal{E}\left[\frac{1}{f(x_{i-1}, x_i, m)} \frac{\partial f(x_{i-1}, x_i, m)}{\partial m}\right]^2$ existe quel que soit i.

Il a été démontré par Markoff (*Bulletin Soc. Physico-Math.*, *Kasan*, 2e série, t. 15, n° 4, 1907) et par M. B. Hostinsky (*C. R. Acad. Sc.*, t. 189, p. 78-80) pour les variables continues que $\mathcal{E}[\psi(x_{n-1}, x_n)]$ tend vers une limite indépendante de x_1 quand n croît indéfiniment. Appelons $g(m)$ cette limite au cas où $\psi(x_{n-1}, x_n)$ se trouve être

$$\left[\frac{1}{f(x_{n-1}, x_n, m)} \frac{\partial f}{\partial m}(x_{n-1}, x_n, m)\right]^2.$$

On voit que la relation (2) deviendra à la limite

$$\lim \frac{1}{n} \mathrm{E}\left[\frac{1}{\varphi} \frac{\partial \varphi}{\partial m}\right]^2 \leqq g(m).$$

L'étude du premier membre ne diffère évidemment pas du cas des variables indépendantes. Il y a donc encore un maximum de précision. Ce maximum est encore fourni par la méthode de M. Fisher sous les mêmes hypothèses que celles qui suffisent à établir le caractère gaussien à la limite de l'estimation par les fonctions $u_i(x_i, m)$, dans le cas des variables indépendantes. En effet, on a, en posant $\frac{1}{f} \frac{\partial f}{\partial m}(x_{i-1}, x_i, m) = u(x_{i-1}, x_i, m)$,

$$\sqrt{n}\,(\mathrm{T}_l - m) = \frac{\dfrac{1}{\sqrt{n}} \displaystyle\sum_{i=2}^{n} u(x_{i-1}, x_i, m)}{-\dfrac{1}{n} \displaystyle\sum_{i=2}^{n} \frac{\partial u}{\partial m}[x_{i-1}, x_i, m + \theta(\mathrm{T}_l - m)]}.$$

Si $\frac{\partial u}{\partial m}(x_{i-1}, x_i, m)$ est uniformément continu en m quels que soient x_{i-1}, x_i, la limite probable du dénominateur sera la même que celle de $\frac{1}{n} \sum_{i=2}^{n} \frac{\partial u}{\partial m}(x_{i-1}, x_i, m)$. On a affaire à des variables aléatoires qui suivent des lois de probabilités différentes. Sous les conditions du Chapitre I, le dénominateur converge donc en probabilité vers $g(m)$.

Markoff et M. S. Bernstein ont montré que les résultats de Liapounoff sont valables dans le cas de variables liées en chaînes. Donc, sous les conditions déjà énoncées, la loi-limite du numérateur est une loi de Gauss d'écart-type $\sqrt{g(m)}$. L'écart-type de la loi-limite de $\sqrt{n}(\mathrm{T}_l - m)$ est donc bien $\frac{1}{\sqrt{g(m)}}$.

Ces démonstrations s'étendent sans plus de difficulté au cas où l'on a affaire à une chaîne multiple, chaque variable étant liée à un nombre *fini* de variables la précédant.

Cas de résultats contagieux au sens de M. Polya. — M. Polya a imaginé un schéma ingénieux de tirages dans une urne, dans lequel on a une suite de variables aléatoires formant une chaîne, mais dans laquelle toutes les variables précédant x_{n+1} figurent dans sa loi de probabilité. Dans ce cas $x_1, x_2, \ldots, x_n$ étant les n premiers résultats, la probabilité pour que x_{n+1} soit compris entre x_{n+1} et $x_{n+1}+dx_{n+1}$, est égale à $f\left(\frac{x_1+\ldots+x_n}{n}, x_{n+1}\right)dx_{n+1}$. Nous supposerons ici encore que cette fonction contient un paramètre inconnu m. Ce sera le cas où l'opérateur se laisse influencer dans son $n+1^{\text{ième}}$ résultat par l'allure globale (représentée par la moyenne arithmétique) des n premiers. On aura, quels que soient $m, x_1, \ldots, x_n$, la relation

$$\int f\left(\frac{x_1+\ldots+x_n}{n}, x_{n+1}, m\right) dx_{n+1} = 1.$$

Si cette relation est dérivable par rapport à m sous le signe $\int$ et même dans des cas plus étendus, d'après l'étude du début du Chapitre II, l'estimation obtenue en appliquant la méthode du maximum de vraisemblance, converge en probabilité vers m sous les mêmes hypothèses qu'au début. On peut chercher à étendre à ce cas les résultats relatifs aux lois-limites et au maximum de précision. Ce qui empêche de le faire directement c'est que le théorème de Markoff-Hostinsky suppose que les chaînes sont *finies*. Dans le cas actuel, la première expérience figure par son résultat dans la loi de probabilité de la $n^{\text{ième}}$. Effectuons le changement de variables $\frac{\Sigma x_i}{n} = \xi_n$. Comme un changement de variables aléatoires quelconque laisse invariant le maximum de vraisemblance, nous pourrons raisonner sur les variables ξ_n. Ces variables sont liées en chaîne simple, la probabilité d'une valeur ξ_{n+1} suivant une valeur ξ_n étant

$$(n+1) f[\xi_n, (n+1)\xi_{n+1} - n\xi_n, m]\, d\xi_{n+1}.$$

Mais la loi de probabilité dépend de n, ce qui crée une différence avec le cas précédent. En employant la même méthode qu'au cours du Chapitre (dérivation logarithmique par rapport au paramètre, de la probabilité d'une expérience exprimée de deux façons différentes, élévation au carré et valeur moyenne des résultats), on voit que les termes rectangles sont encore nuls. Appelons $A_n(x_1, m)$ la valeur moyenne de

$$\left[\frac{\partial}{\partial m}\log f[\xi_n,\ (n+1)\xi_{n+1}-n\xi_n,\ m)]\right]^2.$$

Le même raisonnement sera valable si $\dfrac{\sum_{i=1}^{n} A_i(x_1, m)}{a}$ tend vers une limite indépendante de x_1 quand n croît indéfiniment. Ce qui dans ce cas complique les calculs, c'est que la fonction dont on a à prendre la valeur moyenne change à chaque valeur de i.

Sous ces restrictions, il y a encore un maximum de précision, et l'on peut, en imposant à la loi de probabilité des conditions évidemment plus restrictives que précédemment, faire en sorte que ce maximum soit atteint par la méthode du maximum de vraisemblance.

Cas d'une infinité dénombrable de paramètres. — Nous nous bornerons aux lois de probabilités développables en série de fonctions quasi-orthogonales

$$f(x) = G(x)[1 + a_1 P_1(x) + a_2 P_2(x) + \ldots + a_n P_n(x) + \ldots].$$

Supposons que l'on ait évalué les p premiers coefficients :

1° Par la méthode de M. Fisher qui donne les résultats

$$A_1^l,\quad \ldots,\quad A_p^l;$$

2° Par une méthode quelconque donnant les résultats

$$A_1^c,\quad \ldots,\quad A_p^c.$$

Nous considérerons toujours les variables aléatoires $\alpha_i^l = \sqrt{n}(A_i^l - a_i)$ et $\alpha_i^c = \sqrt{n}(A_i^c - a_i)$, n étant le nombre d'expériences.

Le Chapitre III nous apprend que si l'on considère comme connus les autres coefficients, la loi des α_i^l est

$$K \exp - \frac{1}{2}\left\{\Sigma(\alpha_i^l)^2 E\left[\frac{1}{f}\frac{\partial f}{\partial a_i}\right]^2 + 2\Sigma\Sigma\alpha_i^l\alpha_j^l E\left[\frac{1}{f^2}\frac{\partial f}{\partial a_i}\frac{\partial f}{\partial a_j}\right]\right\}.$$

Si la loi-limite des α_i^c est une loi de Gauss

$$K' \exp - \frac{1}{2}\{\Sigma(\alpha_i^c)^2 c_{ii} + 2\Sigma\Sigma\alpha_i^c\alpha_j^c c_{ij}\}$$

et si cette loi est atteinte avec une régularité suffisante pour que les conditions énoncées dans ce Chapitre $\left(E\left[\frac{1}{\varphi^2}\frac{\partial\varphi}{\partial a_i}\frac{\partial\varphi}{\partial\alpha_i}\right]\right.$ borné et moments du second ordre tendant vers les moments de la loi-limite$\left.\right)$ soient réalisées, on sait que l'on a l'inégalité

$$\psi_c = \Sigma X_i^2 c_{ii} + 2\Sigma\Sigma X_i X_j c_{ij} \leqq \Sigma X_i^2 E\left[\frac{1}{f}\frac{\partial f}{\partial a_i}\right]^2 + 2\Sigma\Sigma X_i X_j E\left[\frac{1}{f^2}\frac{\partial f}{\partial a_i}\frac{\partial f}{\partial a_j}\right] = \psi_l$$

quels que soient les X_i, ce qui entraîne, puisque $K = \sqrt{\frac{\Delta}{(2\pi)^p}}$ et $K' = \sqrt{\frac{\Delta'}{2(\pi)^p}}$ $K > K'$. Passons en coordonnées polaires dans l'espace à p dimensions. On a :

$$\begin{array}{ll} \alpha_1^c = \rho_p^c \sin\theta_1, & \alpha_1^l = \rho_p^l \sin\theta_1, \\ \ldots\ldots\ldots\ldots, & \ldots\ldots\ldots\ldots, \\ \alpha_p^c = \rho_p^c \cos\theta_1 \cos\theta_2 \ldots \sin\theta_{p-1}, & \alpha_p^l = \rho_p^l \cos\theta_1 \cos\theta_2 \ldots \sin\theta_{p-1}. \end{array}$$

Cherchons les lois de probabilité de ρ_p^c et de ρ_p^l, soient $P_c(\rho_p^c)$ et $P_l(\rho_p^l)$,

$$P_c(\rho_p^c) = \int^{\rho_p^c} (\rho_p^c)^{p-1}\, d\rho_p^c \int \cdots \int K'(A_c)^{\psi_c(\theta_1,\ldots\theta_{p-1})}\, d\theta_1 \ldots d\theta_{p-1} \qquad \left[A_c = \exp\left\{-\frac{1}{2}(\rho_p^c)^2\right\}\right],$$

$$P_l(\rho_p^l) = \int^{\rho_p^l} (\rho_p^l)^{p-1}\, d\rho_p^l \int \cdots \int K\,(A_l)^{\psi_l(\theta_1,\ldots\theta_{p-1})}\, d\theta_1 \ldots d\theta_{p-1} \qquad \left[A_l = \exp\left\{-\frac{1}{2}(\rho_p^l)^2\right\}\right].$$

Si l'on appelle $J_c(\rho_p^c)$ et $J_l(\rho_p^l)$ les intégrales portant sur $\theta_1, \ldots, \theta_{p-1}$, on voit que $J_c(\rho_p^c) < J_l(\rho_p^l)$ pour toutes les valeurs de la variable.

Donc $P_c(\rho_p^c) < P_l(\rho_p^l)$. A probabilité égale ρ_p^l est inférieur à ρ_p^c, quelle que soit la méthode d'estimation ; il est facile d'avoir une interprétation de ces deux grandeurs.

Si l'on appelle $f_p(x)$ la fonction obtenue en se bornant aux p premiers paramètres et $f_p^l(x)$, $f_p^c(x)$ les estimations de $f_p(x)$ obtenues par les deux méthodes, il est facile de voir que

$$(\rho_p^c)^2 = n\int \frac{[f_p^c(x) - f_p(x)]^2}{G(x)}\, dx \qquad \text{et} \qquad (\rho_p^l)^2 = n\int \frac{[f_p^l(x) - f_p(x)]^2}{G(x)}\, dx.$$

Les grandeurs sont donc une mesure de l'écart de l'estimation, à la fonction. On voit que quel que soit p l'écart qui a des chances d'être le plus petit est celui donné par le maximum de vraisemblance.

CHAPITRE V.

CORRÉLATION ENTRE DIVERSES ESTIMATIONS.

Du fait que toutes deux convergent en probabilité vers un même nombre, deux estimations différentes d'un même paramètre ne sauraient être indépendantes l'une de l'autre.

Reprenons les mêmes notations. Soient m le paramètre et m_1, m_2 les deux statistiques. Nous allons calculer

$$\frac{E(m_1-m)(m_2-m)}{\sqrt{E(m_1-m)^2 E(m_2-m)^2}},$$

et nous verrons dans quelle mesure cette quantité renseigne sur la corrélation entre deux variables aléatoires. Par suite de l'égalité

$$f(x_1, m)\ldots f(x_n, m) = \varphi(m_1, m_2, m)\, h(\xi_1, \ldots, \xi_{n-2}, m_1, m_2, m),$$

on aura comme plus haut l'inégalité

$$\frac{1}{n} E\left[\frac{1}{\varphi}\frac{\partial \varphi}{\partial m}\right]^2 \leqq E\left[\frac{1}{f}\frac{\partial f}{\partial m}\right]^2 = \frac{1}{\sigma_l^2}.$$

Nous effectuerons le même changement de variables que précédemment

$$\mu_1 = \sqrt{n}(m_1 - m) \qquad \text{et} \qquad \mu_2 = \sqrt{n}(m_2 - m).$$

On voit que

$$\frac{1}{\varphi}\frac{\partial \varphi}{\partial m} = \frac{1}{\varphi}\frac{\delta \varphi}{\delta m} - \sqrt{n}\frac{1}{\varphi}\frac{\delta \varphi}{\delta \mu_1} - \sqrt{n}\frac{1}{\varphi}\frac{\delta \varphi}{\delta \mu_2},$$

et, par suite,

$$\begin{aligned}\frac{1}{n} E\left[\frac{1}{\varphi}\frac{\partial \varphi}{\partial m}\right]^2 = \frac{1}{n} E\left[\frac{1}{\varphi}\frac{\delta \varphi}{\delta m}\right]^2 &- \frac{2}{\sqrt{n}} E\left[\frac{1}{\varphi^2}\frac{\delta \varphi}{\delta m}\frac{\delta \varphi}{\delta \mu_1}\right] - \frac{2}{\sqrt{n}} E\left[\frac{1}{\varphi^2}\frac{\delta \varphi}{\delta m}\frac{\delta \varphi}{\delta \mu_2}\right] \\ &+ E\left[\frac{1}{\varphi}\frac{\delta \varphi}{\delta \mu_1}\right]^2 + E\left[\frac{1}{\varphi}\frac{\delta \varphi}{\delta \mu_2}\right]^2 + 2E\left[\frac{1}{\varphi^2}\frac{\delta \varphi}{\delta \mu_1}\frac{\delta \varphi}{\delta \mu_2}\right].\end{aligned}$$

Si donc $E\left[\frac{1}{\varphi^2}\frac{\delta \varphi}{\delta m}\frac{\delta \varphi}{\delta \mu_1}\right]$ et $E\left[\frac{1}{\varphi^2}\frac{\delta \varphi}{\delta m}\frac{\delta \varphi}{\delta \mu_2}\right]$ sont bornées, il en résultera que $E\left[\frac{1}{\varphi}\frac{\delta \varphi}{\delta \mu_1}\right]^2 + E\left[\frac{1}{\varphi}\frac{\delta \varphi}{\delta \mu_2}\right]^2 + 2E\left[\frac{1}{\varphi^2}\frac{\delta \varphi}{\delta \mu_1}\frac{\delta \varphi}{\delta \mu_2}\right]$ est inférieur ou égal à $\frac{1}{\sigma_l^2}$.

Si la limite de $\varphi\left(m + \frac{\mu_1}{\sqrt{n}}, m + \frac{\mu_2}{\sqrt{n}}, m\right)$ est une loi de Gauss aux coefficients

σ_1^2, σ_2^2 et r, on en déduit comme plus haut et sous les mêmes conditions

$$\lim E(\mu_1^2) = \sigma_1^2, \qquad \lim E(\mu_2^2) = \sigma_2^2, \qquad \lim E(\mu_1 \mu_2) = r\sigma_1\sigma_2,$$

que

$$\frac{1}{1 - r^2}\left[\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} - \frac{2r}{\sigma_1\sigma_2}\right] \leqq \frac{1}{\sigma_l^2}.$$

Appelons E_1 et E_2 efficacité de chacune des statistiques les rapports $\frac{\sigma_l^2}{\sigma_1^2}$ et $\frac{\sigma_l^2}{\sigma_2^2}$, quotient de l'écart-type optimum par chacun des écarts-types limites. Sous les conditions énoncées au Chapitre précédent ces deux quantités sont au plus égales à l'unité. Plus elles en seront voisines et meilleures seront les estimations. Nous pouvons donc écrire l'inégalité limite de la façon suivante

$$(1) \qquad r^2 - 2\sqrt{E_1 E_2}\, r - [1 - (E_1 + E_2)] \leqq 0.$$

Le discriminant du premier membre étant égal à $(1 - E_1)(1 - E_2)$ est positif ou nul. Donc si deux statistiques du même paramètre sont données, leur coefficient de corrélation limite (la quantité r) devra obligatoirement être compris entre deux limites r_1 et r_2. La corrélation pourra être nulle si l'on a $E_1 + E_2 \leqq 1$ (deux statistiques assez éloignées de l'optimum). Supposons que l'on ait $E_2 = 1\ (\sigma_2^2 = \sigma_l^2)$, la statistique m_2 est pleinement efficace. On voit qu'une seule valeur est alors possible pour r, soit $\sqrt{E_1}$. Si les deux estimations sont pleinement efficaces, le coefficient de corrélation limite est égal à l'unité. Il s'agit de trouver quelles conséquences on peut en déduire relativement à m_1 et m_2 et μ_1 et μ_2. Comme $E(\mu_1^2)$, $E(\mu_2^2)$ et $E(\mu_1\mu_2)$ tendent respectivement vers σ_1^2, σ_2^2 et $r\sigma_1\sigma_2$ d'après les hypothèses, la limite de $\frac{E(\mu_1\mu_2)}{\sqrt{E(\mu_1^2)E(\mu_2^2)}}$ sera astreinte aux mêmes inégalités que r, et dans le dernier cas envisagé elle sera égale à l'unité. Dans ce cas, la probabilité de l'ensemble des points pour lesquels $(\mu_1 - \mu_2)$ est supérieur à ε peut être rendue inférieure à η. En effet, considérons l'intégrale

$$\iint (\mu_1 - \mu_2)^2 \varphi_n(\mu_1, \mu_2)\, d\mu_1\, d\mu_2 = E(\mu_1^2) + E(\mu_2^2) - 2E(\mu_1\mu_2).$$

Elle peut être rendue inférieure à un nombre fixé d'avance pour $n > N$ et par conséquent $\iint_{|\mu_1 - \mu_2| > \varepsilon} \varphi_n(\mu_1, \mu_2)\, d\mu_1\, d\mu_2$ peut, dans les mêmes conditions, être rendue inférieure à η. Comme $\mu_1 - \mu_2$ est égal à $\sqrt{n}(m_1 - m_2)$, on peut exprimer ce fait en disant que $m_1 - m_2$ est d'ordre supérieur à $\frac{1}{2}$. Ces infiniment petits ont, dans les Chapitres précédents, été considérés comme négligeables. On peut donc dire qu'il n'y a qu'une seule estimation possédant la propriété du minimum d'écart-type, l'estimation optimum de Fisher.

Nous allons maintenant étudier la réciproque des résultats obtenus. E_1 et E_2 étant deux nombres inférieurs à 1 et r restant une quantité inférieure à 1 satisfaisant à la relation (1), est-il possible de construire deux statistiques « gaussiennes à la limite » d'efficacités respectives E_1 et E_2 et ayant une corrélation égale à r?

La relation (1) qui entraîne la condition $|r| \leqq 1$ est-elle suffisante pour définir r? Nous allons montrer qu'il en est ainsi.

Soient T_1 une statistique « consistent » et E_1 son efficacité. On a, par hypothèse,

$$\lim E[\sqrt{n}(T_1 - m)]^2 = \frac{\sigma_l^2}{E_1}.$$

Soit T_l la statistique optimum. Formons l'estimation $T_l + \alpha(T_l - T_1)$ que nous appellerons T_2. Elle est évidemment « consistent » et gaussienne à la limite. On voit, en se souvenant de ce que le coefficient de corrélation limite de T_1 et de T_l est $\sqrt{E_1}$, que

$$\lim E[\sqrt{n}(T_2 - m)]^2 = \frac{\sigma_l^2}{E_2} = \sigma_l^2\left[1 - \alpha^2 + \frac{\alpha^2}{E_1}\right]$$

et

$$\lim E[\sqrt{n}(T_2 - m)\sqrt{n}(T_1 - m)] = \sigma_l^2\left(\alpha + 1 - \frac{\alpha}{E_1}\right).$$

Donc

$$r = \frac{\alpha + 1 - \dfrac{\alpha}{E_1}}{\dfrac{1}{E_1}\sqrt{1 - \alpha^2 + \dfrac{\alpha^2}{E_1}}}.$$

A une valeur de E_2 correspondent deux valeurs égales et opposées de α. Si l'on exprime α en fonction de E_2, on voit que

$$r = \sqrt{E_1 E_2} \pm \sqrt{(1 - E_1)(1 - E_2)}.$$

Ces deux valeurs sont les valeurs extrêmes permises par la formule (1), E_1 et E_2 étant donnés. Pour deux écarts-types donnés, on a donc pu former deux couples d'estimation ayant leur coefficient de corrélation aussi grand ou aussi petit que le permet la formule (1).

Nous allons maintenant construire des estimations donnant des valeurs intermédiaires au coefficient de corrélation : Soient T_1 et T_2 deux estimations d'efficacité E_1 et E_2, et dont le rapport de corrélation est égal à la plus grande valeur possible, soit $\sqrt{E_1 E_2} + \sqrt{(1 - E_1)(1 - E_2)}$. Il s'agit de construire T_3 ayant même écart-type que T_1 et ayant avec T_2 un coefficient de corrélation compris entre

$$\sqrt{E_1 E_2} + \sqrt{(1 - E_1)(1 - E_2)} \quad \text{et} \quad \sqrt{E_1 E_2} - \sqrt{(1 - E_1)(1 - E_2)}.$$

Nous emploierons toujours la même méthode et nous définirons T_3 comme une combinaison simple de T_1 et T_2. Comme il y a deux équations à résoudre, nous introduirons deux coefficients et T_3 s'écrira

$$T_3 = T_1 + \alpha(T_1 - T_2) + \beta\sqrt{n}(T_1 - T_2)^2.$$

Sous les conditions indiquées $T_1 - T_2$ et $\sqrt{n}(T_1 - T_2)^2$ sont des infiniment petits aléatoires d'ordre $\frac{1}{2}$ en $\frac{1}{n}$. Il faudra donc résoudre en α, β le système

$$\lim_{n=\infty} E\left\{\sqrt{n}\left[T_1 - m + \alpha(T_1 - T_2) + \beta\sqrt{n}(T_1 - T_2)^2\right]\right\}^2 = \lim_{n=\infty} E\left\{\sqrt{n}(T_1 - m)\right\}^2$$

et

$$\lim_{n=\infty} E\left\{\sqrt{n}\left[T_1 - m + \alpha(T_1 - T_2) + \beta\sqrt{n}(T_1 - T_2)^2\right]\sqrt{n}(T_2 - m)\right\}$$
$$= \lim_{n=\infty} E\left\{\sqrt{n}(T_1 - m)\sqrt{n}(T_2 - m)\right\} + P\sigma_1\sigma_2,$$

σ_1 et σ_2 étant les écarts-types de T_1 et T_2. En développant, on a

$$\alpha^2 E\left[\sqrt{n}(T_1 - T_2)\right]^2 + \beta^2 E\left[\sqrt{n}(T_1 - T_2)\right]^4 + 2\alpha E\left[\sqrt{n}(T_1 - m)\sqrt{n}(T_1 - T_2)\right]$$
$$+ 2\alpha\beta E\left[\sqrt{n}(T_1 - T_2)\right]^3 + 2\beta E\left\{\sqrt{n}(T_1 - m)\left[\sqrt{n}(T_1 - T_2)\right]^2\right\} = 0$$

et

$$\alpha E\left[\sqrt{n}(T_1 - T_2)\sqrt{n}(T_2 - m)\right] + \beta E\left\{\sqrt{n}(T_2 - m)\left[\sqrt{n}(T_1 - T_2)\right]^2\right\} = P\sigma_1\sigma_2.$$

Si T_1 et T_2 sont tels que les moments du troisième ordre tendent vers les moments de la loi-limite de T_1 et T_2, il est évident que les coefficients des termes en $\alpha\beta$ et β dans la première équation, en β dans la seconde tendent vers o. On voit facilement que

$$\lim E\left[\sqrt{n}(T_1 - T_2)\right]^2 = \sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2,$$
$$\lim E\left[\sqrt{n}(T_1 - T_2)\sqrt{n}(T_1 - m)\right] = \sigma_1^2 - \rho\sigma_1\sigma_2,$$
$$\lim E\left[\sqrt{n}(T_1 - T_2)\sqrt{n}(T_2 - m)\right] = \rho\sigma_1\sigma_2 - \sigma_2^2.$$

On en tire

$$\alpha = \frac{P\sigma_1\sigma_2}{\rho\sigma_1\sigma_2 - \sigma_2^2}$$

et

$$\beta^2 \lim E[\sqrt{n}(T_1 - T_2)]^4 = -\frac{P^2\sigma_1^2\sigma_2^2(\sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2)}{(\rho\sigma_1\sigma_2 - \sigma_2^2)^2} - \frac{2P\sigma_1\sigma_2(\sigma_1^2 - \rho\sigma_1\sigma_2)}{\rho\sigma_1\sigma_2 - \sigma_2^2}.$$

Il est nécessaire que $E[\sqrt{n}(T_1 - T_2)]^4$ ait une limite, mais cette limite peut parfaitement être différente du moment d'ordre quatre de $T_1 - T_2$ sur la loi-limite. On peut donc calculer β sous la réserve que l'expression inscrite au second membre soit positive. Ce qui impose si l'on remplace ρ par sa valeur

$$-2\sqrt{(1 - E_1)(1 - E_2)} < P < 0.$$

Sous la réserve que le coefficient de corrélation donné soit compris entre les deux limites trouvées, le problème est donc résolu; mais il nous a fallu cette fois supposer la convergence des moments vers les moments de la loi-limite pour le troisième ordre.

CHAPITRE VI.

ESTIMATIONS EXHAUSTIVES.

Tous les résultats des Chapitres précédents supposaient un nombre infini de résultats x_i et étaient relatifs aux limites vers lesquelles tendent les lois de probabilité ou certaines quantités attachées à ces lois.

Nous allons maintenant aborder l'examen de certaines distributions permettant de formuler des théorèmes pour un nombre *fini* d'expériences. L'étude en a été faite par M. Darmois dans une Note publiée aux *Comptes rendus de l'Académie des Sciences* (200, 1935, p. 1265). M. Fisher a appelé les estimations qui en découlent des estimations « sufficient ». Dans la relation fondamentale du Chapitre IV

$$nE\left[\frac{1}{f}\frac{\partial f}{\partial m}\right]^2 \geqq E\left[\frac{1}{\varphi}\frac{\partial \varphi}{\partial m}\right]^2,$$

l'égalité n'a lieu pour n fini que si l'on a $E\left[\frac{1}{h}\frac{\partial h}{\partial m}\right]^2 = 0$, ce qui exige que l'on ait quels que soient ξ_1, ξ_{n-1}, m : $\frac{1}{h}\frac{\partial h}{\partial m} \equiv 0$. La fonction de distribution liée de $\xi_1, \ldots, \xi_{n-1}$, ne doit donc pas contenir m. On veut donc trouver toutes les fonctions $f(x, m)$ telles que

$$\prod_{1}^{n} f(x_i, m) \equiv \mathrm{F}[\varphi(x_1, \ldots, x_n), m]\,\mathrm{H}(x_1, \ldots, x_n).$$

On suppose que les variables $\xi_1, \ldots, \xi_{n-1}$ sont identiques aux $x_1, \ldots, x_{n-1}$. Posons

$$\log f(x, m) = p(x, m), \qquad \log \mathrm{F}(\varphi, m) = \mathrm{P}(\varphi, m) \qquad \text{et} \qquad \log \mathrm{H} = \mathcal{H}.$$

On a donc

$$\sum \frac{\partial p(x_i, m)}{\partial m} = \frac{\partial \mathrm{P}}{\partial m} \qquad \text{et} \qquad \frac{\partial^2 p(x_i, m)}{\partial m\, \partial x_i} = \frac{\partial^2 \mathrm{P}}{\partial m\, \partial \varphi} \frac{\partial \varphi}{\partial x_i}.$$

Cette dernière quantité doit donc être indépendante de x_j, quel que soit j ; d'où

$$\frac{\partial^3 \mathrm{P}}{\partial \varphi^2\, \partial m} \frac{\partial \varphi}{\partial x_j} \frac{\partial \varphi}{\partial x_i} + \frac{\partial^2 \mathrm{P}}{\partial \varphi\, \partial m} \frac{\partial^2 \varphi}{\partial x_i\, \partial x_j} = 0$$

quel que soit m. Donc, en dérivant par rapport à m

$$\frac{\partial^4 \mathrm{P}}{\partial \varphi^2\, \partial m^2} \frac{\partial \varphi}{\partial x_j} \frac{\partial \varphi}{\partial x_i} + \frac{\partial^3 \mathrm{P}}{\partial \varphi\, \partial m^2} \frac{\partial^2 \varphi}{\partial x_i\, \partial x_j} = 0.$$

En éliminant $\frac{\partial \varphi}{\partial x_j} \frac{\partial \varphi}{\partial x_i}$ et $\frac{\partial^2 \varphi}{\partial x_i \partial x_j}$ entre ces deux équations, on a l'équation aux dérivées partielles, à laquelle doit satisfaire P

$$\frac{\partial^4 \mathrm{P}}{\partial \varphi^2\, \partial m^2} \frac{\partial^2 \mathrm{P}}{\partial \varphi\, \partial m} - \frac{\partial^3 \mathrm{P}}{\partial \varphi^2\, \partial m} \frac{\partial^3 \mathrm{P}}{\partial \varphi\, \partial m^2} = 0,$$

dont l'intégrale générale est $a(\varphi)b(m) + c(m) + d(\varphi)$, a, b, c, d étant quatre fonctions arbitraires. Il convient de remarquer que pour que les calculs précédents soient valables, la fonction P doit avoir *partout* les dérivées dont on a besoin. C'est encore une différence avec les résultats précédents, pour lesquels on pouvait admettre que les conditions de régularité exigées ne soient pas remplies sur un certain ensemble à condition que cet ensemble ait une probabilité nulle ; a, b, c, d doivent donc avoir partout des dérivées d'ordre deux. La forme de $p(x, m)$ s'obtient maintenant facilement. Comme

$$\frac{\partial^2 p}{\partial m\, \partial x_i} = \frac{\partial^2 \mathrm{P}}{\partial \varphi\, \partial m} \frac{\partial \varphi}{\partial x_i},$$

en dérivant deux fois cette égalité par rapport à m et par rapport à x_i, on trouve que

$$\frac{\partial^4 p}{\partial m^2\,\partial x_i^2}\,\frac{\partial^2 p}{\partial m\,\partial x_i}-\frac{\partial^3 p}{\partial m\,\partial x_i^2}\,\frac{\partial^3 p}{\partial m^2\,\partial x_i}=\left[\frac{\partial^4 \mathrm{P}}{\partial \varphi^2\,\partial m^2}\,\frac{\partial^2 \mathrm{P}}{\partial m\,\partial \varphi}-\frac{\partial^3 \mathrm{P}}{\partial m\,\partial \varphi^2}\,\frac{\partial^3 \mathrm{P}}{\partial m^2\,\partial \varphi}\right]\left(\frac{\partial \varphi}{\partial x_i}\right)^2=0.$$

La loi de probabilité $f(x, m)$ est donc exp. $\{u(x)v(m)+r(x)+s(m)\}$, u, v, r, s étant deux fois dérivables partout et étant telles que l'expression indiquée soit une loi de probabilité.

On peut se poser le même problème quand on a affaire à plusieurs variables $x_1, \ldots, x_k$ et à plusieurs paramètres $\alpha_1, \ldots, \alpha_h$. Si $p(x_1, \ldots, x_k; \alpha_1, \ldots, \alpha_h)$ est encore le logarithme de la loi de probabilité élémentaire, il est facile de vérifier que p satisfait encore à l'ensemble d'équations

$$\frac{\partial^4 p}{\partial \alpha_i^2\,\partial x_j^2}\,\frac{\partial^2 p}{\partial \alpha_i\,\partial x_j}-\frac{\partial^3 p}{\partial \alpha_i\,\partial x_j^2}\,\frac{\partial^3 p}{\partial \alpha_i^2\,\partial x_j}=0$$

quels que soient i et j. On aura donc

$$\begin{aligned} & f(x_1, \ldots, x_k; \alpha_1, \ldots, \alpha_h) \\ & \quad =\exp\left\{\sum_{i=1}^{h} \mathrm{U}_i(x_1, \ldots, x_k)\,\mathrm{A}_i(\alpha_1, \ldots, \alpha_h)+\mathrm{V}(x_1, \ldots, x_k)+\mathrm{B}(\alpha_1, \ldots, \alpha_h)\right\}, \end{aligned}$$

U_i et V ainsi que A_i et B sont encore des fonctions dérivables deux fois.

On peut se poser la question de savoir quelle est la forme des lois permettant une estimation exhaustive pour un paramètre dans le cas où ce paramètre est, soit un paramètre d'estimation au sens précis du terme, soit un paramètre de spécification, selon le vocabulaire de M. Fisher.

Dans le premier cas, on aura $f(x, m)=f(x-m)$ et dans le second $f(x, \sigma)=\frac{1}{\sigma}f\left(\frac{x}{\sigma}\right)$. En posant $x-m=t$ ou $\frac{x}{\sigma}=t$, on transformera l'équation aux dérivées partielles trouvée, en équation différentielle en t. Dans le premier cas on trouve

$$-\frac{d^4 p}{dt^4}\,\frac{d^2 p}{dt^2}+\left(\frac{d^3 p}{dt^3}\right)^2=0,$$

et dans le second

$$\begin{aligned} & -\left(\frac{d^2 p}{dt^2}\,t+\frac{dp}{dt}\right)\left[\frac{d^4 p}{dt^4}\,t^2+6\,\frac{d^3 p}{dt^3}\,t+6\,\frac{d^2 p}{dt^2}\right] \\ & +\left[\frac{d^3 p}{dt^3}\,t+2\,\frac{d^2 p}{dt^2}\right]\left[\frac{d^3 p}{dt^3}\,t^2+4\,\frac{d^2 p}{dt^2}\,t+2\,\frac{dp}{dt}\right]=0. \end{aligned}$$

L'intégrale générale de la première est $\frac{\mathrm{K}}{c^2}e^{ct}+\mathrm{A}t+\mathrm{B}$. Il y a toute une famille d'intégrales singulières $\frac{Ke^2}{4}t^2 + At + B$. Si x varie de $-\infty$ à $+\infty$ la première forme est à rejeter, car elle ne pourrait donner une loi de probabilité que si l'on introduisait la valeur absolue de t, ce qui l'empêcherait d'être partout dérivable deux fois. Dans ce cas, la seule loi acceptable pour $f(x-m)$ est donc $\exp.\left\{-\frac{1}{2\sigma^2}(x-m)^2 + A(x-m) + B\right\}$. C'est-à-dire la loi de Gauss. On peut accepter la première forme si x est susceptible de prendre seulement des valeurs positives (ou si prenant des valeurs des deux signes, il ne peut devenir infini que pour des valeurs d'un seul signe).

Dans le second cas l'intégrale générale est de la forme

$$A \log t + B t^\alpha + D.$$

Il y a encore une famille d'intégrales singulières qui sont

$$\lambda(\log t)^2 + \mu \log t + \nu.$$

Si x varie $-\infty$ à $+\infty$, la seconde forme est à rejeter; la première donne la loi de probabilité

$$K\left(\frac{x}{\sigma}\right)^A \exp\left\{B\left(\frac{x}{\sigma}\right)^\alpha\right\}\frac{dx}{\sigma},$$

ce qui entraîne $A \geqq 0$, $B < 0$ et α entier et pair. On voit que si $A = 0$, $B = -1/2$ et $\alpha = 2$. On retrouve la loi de Gauss. Si x ne peut prendre que des valeurs positives, d'autres formes sont acceptables. En particulier l'intégrale générale permet de mettre en évidence les lois

$$\frac{1}{\Gamma(h)} e^{-\frac{x}{l}}\left(\frac{x}{l}\right)^{h-1}\frac{dx}{l}.$$

Dans le cas des estimations « sufficient » la méthode de M. Fisher conduit à une formule simple pour l'optimum. Il suffit de résoudre

$$-\frac{s'(T)}{\nu'(T)} = \frac{\Sigma u(x_i)}{n}.$$

Considérons une des lois trouvées et faisons le double changement de variable aléatoire et de paramètre suivant : $u(x) = \xi$ et $\nu(m) = \alpha$. La loi deviendra

$$\exp\{\xi\alpha + X(\xi) + A(\alpha)\}\, d\xi.$$

Il convient de remarquer que en général $X(\xi)$ ne sera pas dérivable deux fois partout, car on aura été obligé d'introduire $|u'(x)|$. On n'aura pas une loi « sufficient » en ξ. Il est facile de trouver la fonction caractéristique définie comme $E(e^{it\xi})$. Ce sera ici exp. $\{A(\alpha) - A(\alpha + t)\}$. La fonction caractéristique de $\frac{\Sigma\xi_i}{n}$ sera exp. $\left\{n\left[A(\alpha) - A\left(\alpha + \frac{t}{n}\right)\right]\right\}$. Ce sera la fonction correspondant à la loi de Gauss si $A(\alpha) = a\alpha^2 + b\alpha + c$. (On doit avoir $a < 0$). Si $A(\alpha)$ n'est pas de cette forme, on peut voir que, à condition que $A''(\alpha)$ soit continue, la fonction caractéristique de $\frac{\Sigma[\xi_i + A'(\alpha)]}{\sqrt{n}}$ tend vers exp. $\left\{-\frac{A''(\alpha)}{2}t^2\right\}$ qui correspond à la loi de Gauss d'écart-type $-A''(\alpha)$. Si l'on revient aux anciennes variables x et m on voit que : la loi de $\sqrt{n}\left[\frac{\Sigma u(x_i)}{n} + \frac{s'(m)}{v'(m)}\right]$ tend vers une loi de Gauss d'écart-type

$$\Sigma^2(m) = \frac{v''(m)\,s'(m) - s''(m)\,v'(m)}{v'^3},$$

à condition que cette quantité soit continue en m. Ce quotient doit être positif. Si l'on considère l'estimation obtenue en résolvant : $-\frac{s'(T)}{v'(T)} = \frac{\Sigma u(x_i)}{n}$, on a l'estimation par la méthode du maximum de vraisemblance.

Il est facile de vérifier ici que cette estimation a, à la limite, un comportement gaussien dans lequel l'écart-type est égal à $\frac{v''s' - s''v'}{v'}$, à condition que $\frac{s'}{v'}$ ait une dérivée continue en m, ce qui est réalisé étant données les conditions que nous nous sommes imposées. Cette quantité est égale à $E\left[\frac{1}{f}\frac{\partial f}{\partial m}\right]^2$ comme on le voit en dérivant deux fois sous le signe d'intégration.

Si l'on se reporte au Chapitre III, on voit que la condition *suffisante* pour que $E[\sqrt{n}(T - m)]^2$ soit à la limite égal à $\frac{v''s' - s''v'}{v'}$, et que $E(T)$ soit à la limite égal à m, est que $v''s' - s''v'$ garde un signe constant quel que soit m, ce qui entraînera, étant donné ce que nous avons supposé plus haut, que $v'(m)$ ait un signe constant.

Une fois fait le changement de variable $u(x) = \xi$, la probabilité d'une série de résultats $\xi_1, \ldots, \xi_n$ devient

$$e^{v(m)\Sigma\xi_i + \Sigma\chi(\xi_i) + ns(m)}\, d\zeta_1 \ldots d\zeta_n.$$

La loi de probabilité de $\Xi = \frac{\Sigma\xi_i}{n}$ sera

$$\exp\{n\Xi v(m) + ns(m) + \varphi_n(\Xi)\}\, n\, d\Xi,$$

$\varphi_n(\Xi)$ étant une fonction qui se définit par récurrence

$$\exp\{\varphi_n(\Xi)\} = \int \exp\{\varphi_{n-1}(x)\}\exp\{X(\Xi - x)\}\,dx.$$

La loi de probabilité liée des ξ_i, Ξ étant fixé, sera donc exp. $\{\Sigma X(\xi_i) - \varphi_n(\Xi)\}$ et ne contiendra pas m. Il en sera de même pour la loi de probabilité d'une fonction quelconque de $\frac{\Sigma\xi_i}{n}$. L'estimation par la méthode du maximum de vraisemblance est donc telle que quel que soit n, on ait dans ce cas

$$E\left(\frac{1}{f}\frac{\partial f}{\partial m}\right)^2 = \frac{1}{n}E\left[\frac{1}{\varphi}\frac{\partial \varphi}{\partial m}\right]^2.$$

Dans le cas des lois n'admettant pas une estimation exhaustive, cette égalité est encore vraie mais seulement à la limite. C'est en ce sens que l'on peut dire que l'estimation par la méthode du maximum de vraisemblance est une estimation exhaustive.

Les conditions que nous avons énoncées pour obtenir la forme générale des lois permettant une estimation exhaustive sont suffisantes, mais beaucoup trop restrictives.

L'existence en particulier de la dérivée seconde partout n'est nullement nécessaire. Un exemple permettra de s'en rendre compte. Soit la première loi de Laplace : $\frac{1}{2}e^{-\frac{|x|}{l}}\frac{dx}{l}$. Elle n'entre pas dans la forme générale énoncée puisque $|x|$ n'est pas dérivable partout. Néanmoins, elle admet pour l une évaluation exhaustive. Cherchons la loi de probabilité de la statistique fournie par le maximum de vraisemblance, ici $\frac{\Sigma|x_i|}{n} = \lambda$. Il s'agit de trouver la masse comprise entre les deux surfaces $\frac{\Sigma|x_i|}{n} = \lambda$ et $\frac{\Sigma|x_i|}{n} = \lambda + d\lambda$, la densité étant égale à $\prod_{i=1}^{n}\frac{1}{2^n}\frac{1}{l^n}e^{-\frac{|x_i|}{l}}$, soit dans le cas présent à $\frac{1}{2^n l^n}e^{-\frac{n\lambda}{l}}$; elle est donc constante dans tout le volume considéré. Les plans $\frac{\Sigma|x_i|}{n} = \lambda$ sont par suite des différents signes que peuvent prendre les x_i, au nombre de 2^n. Considérons la pyramide formée par l'un d'eux et les arêtes du « trièdre » de référence. Son volume sera $\frac{(n\lambda)^n}{n!}$, celui compris entre deux plans parallèles infiniment voisins sera $\frac{n^n\lambda^{n-1}}{(n-1)!}d\lambda$, et le volume total de la région considérée sera $2^n\frac{n^n\lambda^{n-1}}{(n-1)!}d\lambda$.

La loi de probabilité élémentaire de λ sera donc

$$\frac{e^{-\frac{n\lambda}{l}}}{l^n} \frac{n^n}{(n-1)!} \lambda^{n-1}\, d\lambda.$$

On vérifiera facilement que sa limite est une loi de Gauss : En revenant à la formule du début du Chapitre VI on voit que $h(x_1, \ldots, x_n)$, loi de probabilité liée du point $x_1, \ldots, x_n$ sur la surface isostatistique $\frac{\Sigma |x_i|}{n} = \lambda$, est dans ce cas : $\frac{(n-1)!}{2^n n^n \lambda^n} = \frac{(n-1)!}{2^n (\Sigma |x_i|)^n} \cdot h$ ne contient donc pas l. Il n'en aurait pas été de même si au lieu de $e^{-\frac{|x|}{l}}$ on avait eu la loi $e^{-|x-l|}$, l étant toujours le paramètre à estimer; l aurait alors figuré dans h, la loi de la forme indiquée n'étant pas susceptible d'une estimation exhaustive (l'application de la méthode de R.-A. Fisher conduit à prendre la médiane des n résultats).

Une autre loi de probabilité permet l'estimation exhaustive. C'est la loi de répartition uniforme $\frac{dx}{l}(0 < x < l)$. Considérons l'estimation $\lambda =$ plus grande valeur $(x_1, \ldots, x_n)$, nous avons vu qu'elle convergeait en probabilité vers l. Sa loi de probabilité élémentaire est $n\left(\frac{\lambda}{l}\right)^{n-1} \frac{d\lambda}{l}$.

Ici $h(x_1, \ldots, x_n)$ sera $\frac{1}{n\lambda^{n-1}}$ et par conséquent ne contiendra pas l. On a donc affaire encore à une estimation exhaustive.

M. R.-A. Fisher a signalé une propriété importante dont devaient jouir les statistiques exhaustives. Considérons n résultats $x_1, \ldots, x_n$ et $\lambda_n(x_1, \ldots x_n)$ l'estimation qu'ils fournissent; soient m autres résultats $x_{n+1}, \ldots, x_{n+m}$ et $\lambda_m(x_{n+1}, \ldots, x_{n+m})$ l'estimation qui leur est propre. L'ensemble $x_1, \ldots, x_{n+m}$ peut fournir une estimation λ_{n+m}; comme λ_n et λ_m résument toute l'information que l'on peut extraire respectivement des $x_1, \ldots, x_n$ et de $x_{n+1}, \ldots, x_{n+m}$, on doit avoir quels que soient m et n, $\lambda_{m+n} = f(\lambda_n, \lambda_m)$. Dans le dernier exemple que nous avons cité l'estimation satisfait bien à cette équation de définition. Il n'en serait pas de même dans le cas d'une estimation par la valeur médiane. Les lois dont nous avons donné la forme générale plus haut fournissent bien une estimation de ce genre, puisque $\lambda_n = \varphi\left(\frac{\Sigma u(x_i)}{n}\right)$. On a alors

$$\lambda_{m+n} = \varphi\left[\frac{n\varphi^-(\lambda_n) + m\varphi^-(\lambda_m)}{m+n}\right],$$

φ^- désignant la fonction inverse de φ.

La formule générale trouvée fournit donc des statistiques exhaustives, mais un certain nombre de ces statistiques échappent à la formule générale.

BIBLIOGRAPHIE.

G. DARMOIS. — Sur les lois de probabilités à estimation exhaustive (*Comptes rendus de l'Académie des Sciences*, **200**, 1935, p. 1265).

J.-L. DOOB. — Probability and Statistics (*Transactions of American Mathematical Society*, **36**, 1934, p. 759).

R.-A. FISHER. — Theory of Statistical estimation (*Proceedings of Cambridge*, t. **22**, p. 700).

FRÉCHET. — Sur la convergence en probabilité (*Metron*, **15**, VI, 1930, p. 3).

KHINTCHINE. — Sur la loi des grands nombres (*Comptes rendus de l'Académie des Sciences*, **188**, 1929, p. 477).

LÉVY. — Propriétés asymptotiques des sommes de variables aléatoires (*Journal de Mathématiques pures et appliquées*, **100**, 1935, p. 347).

MARKOFF. — Extension de la loi des grands nombres aux événements dépendants les uns des autres (*Bulletin de la Société physico-mathématique de Kasan*, 15, 1907, p. 135).

POLYA. — Sur quelques points de la théorie des Probabilités (*Annales de l'Institut Henri-Poincaré*, **1**, 1931-1932, p. 117).

A. KOLMOGOROFF. — *Grundbegriffe der Wahrscheinlichkeitsrechnung*.

Vu et approuvé :

Paris, le 29 octobre 1937.

LE DOYEN DE LA FACULTÉ DES SCIENCES,

C. MAURAIN.

Vu et permis d'imprimer :

Paris, le 29 octobre 1937.

LE RECTEUR DE L'ACADÉMIE DE PARIS,

ROUSSY.

TABLE DES MATIÈRES.