ANNALES DE L'I. H. P.

R. DE MISÉS

Les lois de probabilité pour les fonctions statistiques

Annales de l'I. H. P., tome 6, nº 3-4 (1936), p. 185-212
<http://www.numdam.org/item?id=AIHP_1936__6_3-4_185_0>

Les lois de probabilité pour les fonctions statistiques

PAR

R. DE MISÈS
à Istanbul

En poursuivant certaines recherches dont j'avais donné un premier résultat dans mon *Cours des Probabilités* publié en 1931 (1), je fus amené à établir deux nouveaux théorèmes de limite, qui ne me semblent pas sans intérêt. Je signalais les énoncés essentiels de ces théorèmes dans une note insérée au premier volume de la *Revue de la Faculté des Sciences à Istanbul* (2). D'autre part un article du volume offert en hommage à M. *W. Wirtinger* (3) s'occupait d'une conséquence particulière découlant des nouveaux théorèmes, savoir de la « loi des grands nombres pour les fonctions statistiques ». Dans ce qui suit je me propose de donner la démonstration complète du premier des théorèmes en vue.

Pour se former une idée de ce que je vais prouver, on doit se rappeler le théorème classique concernant la somme de n variables aléatoires, théorème qui, jusqu'à nos jours, ne cessa de susciter les plus importantes recherches de nombreux savants. Étant donnée une suite infinie de lois de probabilité $V_1\ (x_1)$, $V_2\ (x_2)$, $V_3\ (x_3)$,... soumises à certaines restrictions très faibles, on démontre que la distribution (= loi de probabilité) pour la somme $x_1 + x_2 + \ldots\ x_n$ de ces variables tend vers

(1) Vorlesungen aus dem Gebiete der angewandten Mathematik, Bd. I : « Wahrscheinlichkeitsrechnung. » Wien u. Leipzig, Deuticke 1931, S. 192-197, § 7, 6 « Ergänzungssätze zu den Gesetzen der grossen Zahlen ».

(2) *Revue de la Faculté des Sciences de l'Université d'Istanbul*, t. I, 1935, fasc. I, p. 61-80. « Deux nouveaux théorèmes de limite dans le calcul des probabilités. »

(3) Monatshefte für Math. u. Physik, Wirtinger-Festband, Wien 1936, S. 105-128 : Die Gesetze der grossen Zahl für statistische Funktionen.

la fonction de Gauss, si n augmente infiniment. Il va de soi qu'on peut remplacer dans cet énoncé la somme par la moyenne arithmétique. Mais nous allons voir que le théorème s'étend à une classe bien plus générale de fonctions des $x_1, x_2, \ldots x_n$.

La moyenne arithmétique jouit évidemment des propriétés suivantes : 1° Elle est une fonction symétrique, c'est-à-dire elle ne dépend pas de l'ordre dans lequel on range les variables ; 2° Elle ne change pas si l'on remplace le nombre n de variables par un multiple entier $m.\,n$ et si dans le nouvel ensemble de variables toute valeur ancienne se retrouve m-fois. On dira d'une telle fonction qu'elle ne dépend que de la *répartition* des variables $x_1, x_2, \ldots x_n$. Notamment nous entendons par « répartition de n variables » une fonction $S(x)$ définie par les $x_1, x_2, \ldots x_n$ de la façon suivante. Pour un x réel quelconque $nS(x)$ est égal au nombre de celles parmi les variables $x_1, x_2, \ldots x_n$ dont la valeur ne surpasse pas x. Il est évident que la répartition $S(x)$ est une fonction non-décroissante, représentée par une ligne-escalier montant de zéro à un sur des marches dont les hauteurs sont des multiples entiers de $1/n$. Toute fonction f des $x_1, x_2, \ldots x_n$ qui s'exprime par $S(x)$ remplit les conditions signalées ci-dessus pour la moyenne arithmétique et elle sera nommée une « *fonction statistique* » $f\{S(x)\}$. Le théorème dont la démonstration fait l'objet du présent mémoire dit en substance :

Pour toute fonction statistique $f\{S(x)\}$ — naturellement sous certaines restrictions liant les $V_\nu(x)$ et f — *la distribution (loi de probabilité) tend vers la Gaussienne, si n tend vers l'infini.*

Les conditions précises que les distributions données $V_\nu(x)$ doivent remplir, seront indiquées aux §§ 3 et 5. Ici je me borne à donner quelques exemples de fonctions statistiques afin de justifier le nom que j'ai donné à cette classe de fonctions. Tout d'abord le moment de $m^{\text{ème}}$ ordre est une fonction statistique, car il s'exprime sous forme d'une intégrale de Stieltjes [1]

$$\int \varphi(x)dS(x) = \frac{1}{n}\sum_{\nu=1}^{n}\varphi(x_\nu) \tag{1}$$

pour $\varphi(x) = x^m$. Pour de telles fonctions statistiques *linéaires* l'exten-

(1) Le simple signe $\int$ désignera toujours l'intégration de $-\infty$ à $+\infty$, le signe $\int_a$ l'intégration de a à ∞, etc.

sion du théorème classique ne présente aucune difficulté. Mais on peut envisager aussi les moments rapportés à la moyenne arithmétique α (les écarts)

$$M_m = \int (x-\alpha)^m dS(x), \qquad \alpha = \int x dS(x), \tag{2}$$

le plus simple cas de fonctions statistiques non-linéaires. D'une façon plus générale, on peut s'occuper de fonctions quelconques des moments, par exemple du coefficient dit de *Lexis*

$$L = \frac{NM_2}{\alpha(N-\alpha)} \tag{3}$$

où N désigne un nombre entier positif donné, α la moyenne arithmétique et M_2 le moment de deuxième ordre suivant la définition (2). Un autre exemple se présente dans le coefficient de « concentration » ou de « disparité » introduit par *Gini*

$$G = \frac{1}{\alpha} \int_0 S(1-S)dx \tag{4}$$

où l'on suppose toutes les valeurs de x positives. Enfin la notion de fonction statistique s'étend aussi aux problèmes à plusieurs dimensions. Si nous désignons par α, β resp. $M_{\iota\varkappa}$ les moyennes arithmétiques et les moments d'ordre ι, $\varkappa$

$$\alpha = \frac{1}{n}\sum_{\nu=1}^{n} x_\nu, \quad \beta = \frac{1}{n}\sum_{\nu=1}^{n} y_\nu\,; \quad M_{\iota\varkappa} = \frac{1}{n}\sum_{\nu=1}^{n}(x_\nu-\alpha)^\iota (y_\nu-\beta)^\varkappa,$$

le coefficient bien connu de « corrélation »

$$C = \frac{M_{11}}{\sqrt{M_{02}M_{20}}} \tag{5}$$

est une telle fonction statistique. Mais nous nous bornerons dans ce qui suit, au cas d'une seule variable.

La démonstration que je vais établir sera basée sur un lemme élémentaire qui se trouve exposé au § 1. Au § 2 je cite quelques formules, connues en substance, concernant les probabilités dans le problème d'épreuves répétées. Le § 3 donne la démonstration complète du nouveau théorème pour le cas spécial où toutes les distributions données $V_\nu(x)$ sont discontinues et limitées. En ce cas la répartition d'un nombre quelconque de variables s'exprime par un nombre borné de « fréquences relatives », de sorte que la notion *générale* de fonction statistique n'y entre pas en jeu. Le § 4 est consacré aux définitions et aux précisions nécessaires fixant cette notion et celle de la première et deuxième dérivée ainsi que du développement Taylorien d'une fonction statistique. Enfin on trouvera au § 5 la démonstration définitive qui est le but principal de ce mémoire.

§ 1. Lemme préliminaire

1. Le lemme que nous allons exposer et qui servira de base pour tout ce qui suit, exprime en substance : Si, étant données deux fonctions A et B, la distribution de A tend vers la Gaussienne et si en même temps l'espérance mathématique de $|A\text{-}B|$ tend vers zéro, la distribution de B tendra vers la distribution Gaussienne elle aussi. Afin de préciser et de démontrer cet énoncé il faut introduire les notions suivantes.

Nous envisageons une suite infinie de collectifs C_1, C_2, C_3 ... à distributions quelconques. Soient α_n le nombre de dimensions dans C_n, puis $x_1, x_2, \dots x_{\alpha_n}$ les composantes du caractère distinctif (de la variable aléatoire) et $V_n(x_1, x_2, \dots x_{\alpha_n})$ la distribution dans C_n. En particulier $V_n(X_1, X_2, \dots X_{\alpha_n})$ signifie la probabilité pour que dans C_n la valeur de x_ν ne surpasse pas X_ν ($\nu = 1, 2, \dots \alpha_n$). Si A_n, B_n sont de fonctions quelconques des x_ν on appelle les intégrales de Stieltjes, étendues à l'espace entier

$$\int A_n dV_n = E\{A_n\}, \quad \int B_n dV_n = E\{B_n\} \tag{6}$$

les espérances mathématiques de A_n resp. B_n. D'autre part on définit les « distributions » P_n et Q_n de A_n resp. B_n moyennant les intégrales

$$\int_{(A_n \leqq X)} dV_n = P_n(X), \quad \int_{(B_n \leqq X)} dV_n = Q_n(X) \tag{7}$$

La première de ces intégrales, étendue à tous les points de « l'espace caractéristique » pour lesquels la valeur de A_n ne surpasse pas X, donne la probabilité de l'inégalité $A_n \leqq X$, etc.

Pour fixer les idées on peut imaginer une série infinie d'urnes dont chacune est remplie de billets numérotés de 1 à k. L'élément du collectif C_n sera un tirage effectué dans les n premières de ces urnes et son caractère distinctif l'ensemble des chiffres sortis, donc $\alpha_n = n$. La distribution V_n à n dimensions se trouve en ce cas d'après la règie de multiplication des probabilités, si l'on suppose données les probabilités des chiffres 1 à k pour chaque urne. Comme exemple de fonctions A_n, B_n on peut envisager la somme des chiffres tirés ou la somme de leurs carrés, etc. Pour chacune de telles fonctions existe une distribution à une seule dimension, définie par (7). Mais notre hypothèse générale comprend aussi le cas où, par exemple, les billets dans certaines urnes sont pourvus de plusieurs chiffres, etc.

2. Maintenant nous sommes à même de formuler notre lemme : *Soient A_n et B_n deux fonctions du caractère distinctif (de la variable aléatoire) de C_n dans une suite infinie de collectifs C_1, C_2, C_3 ..., soient $P_n(X)$ et $Q_n(X)$ leurs distributions (lois de probabilité) respectives, enfin*

$$\int |A_n - B_n| \, dV_n = E\{|A_n - B_n|\} = E_n \tag{8}$$

l'espérance mathématique de leur différence absolue ; alors les équations

$$\lim_{n \to \infty} E_n = 0, \quad \lim_{n \to \infty} Q_n(X) = F(X) \tag{9}$$

où $F(X)$ est une fonction à dérivée bornée

$$|F'(X)| < M \tag{9'}$$

entraînent l'équation

$$\lim_{n \to \infty} P_n(X) = F(X). \tag{9''}$$

3. En voilà la démonstration très simple. Nous désignons, pour abréger, l'intégrale de dV_n, étendue à un domaine dont les points satisfont à certaines inégalités, comme par exemple $A_n \leqq X$, par $\text{Prob}\{A_n \leqq X\}$ etc. Or on a

$$P_n(X) = \text{Prob}\{A_n \leqq X\}, \qquad Q_n(X) = \text{Prob}\{B_n \leqq X\}$$

donc, en soustrayant de chaque terme $\text{Prob}\{A_n \leqq X, B_n \leqq X\}$

$$P_n(X) - Q_n(X) = \text{Prob}\{A_n \leqq X, B_n > X\} - \text{Prob}\{A_n > X, B_n \leqq X\} \tag{10}$$

Si l est un nombre positif quelconque on voit immédiatement que $\text{Prob}\{A_n \leqq X, B_n > X\} \leqq \text{Prob}\{A_n - B_n \leqq -l\} + \text{Prob}\{X < B_n \leqq X + l\}$

car le domaine cité au premier membre est compris dans la somme des deux domaines cités au second (voir fig. 1). Le dernier terme est égal à $Q_n(X+l) - Q_n(X)$. En vertu de (9') il s'ensuit donc pour des n assez grands :

$$\text{(11)} \qquad \text{Prob}\{A_n \leqq X, B_n > X\} \leqq \text{Prob}\{|A_n - B_n| \geqq l\} + Ml + \varepsilon_1$$

où ε_1 tend vers zéro, si n augmente infiniment.

Ici le domaine $A_n - B_n \leqq -l$ a été remplacé par le plus grand

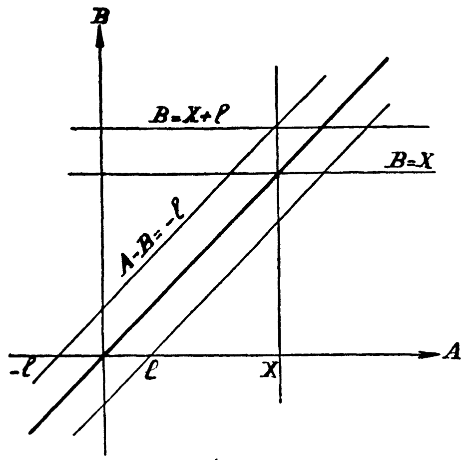

Fig. 1

$|A_n - B_n| \geqq l$. La même inégalité se déduit pour le second terme au deuxième membre de (10) et, étant donné que les probabilités sont des nombres non-négatifs, la valeur absolue de la différence ne peut surpasser la plus grande des deux valeurs à soustraire l'une de l'autre. Donc on tire de (10) et (11) :

$$\text{(12)} \qquad |P_n(X) - Q_n(X)| \leqq \text{Prob}\{|A_n - B_n| \geqq l\} + Ml + \varepsilon_1$$

D'autre part l'espérance mathématique E_n définie par (8) est supérieure ou égale à l'intégrale de $|A_n - B_n|\, dV_n$, si cette intégrale n'est étendue qu'au domaine $|A_n - B_n| \geqq l$:

$$E_n \geqq \int_{|A_n - B_n| \geqq l} |A_n - B_n|\, dV_n \geqq l \int_{|A_n - B_n| \geqq l} dV_n = l\, \text{Prob}\{|A_n - B_n| \geqq l\}$$

En portant cette valeur dans (12) nous recevons

$$|P_n(X) - Q_n(X)| \leqq \frac{E_n}{l} + Ml + \varepsilon_1 \tag{13}$$

Pour un ε positif quelconque on prend $l = \varepsilon/2M$, puis on augmente n autant que $2\,M\,E_n$ (en vertu de la première équation (9)) reste au-dessous de $\varepsilon^2/2$, ce qui donne finalement :

$$|P_n(X) - Q_n(X)| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} + \varepsilon_1. \tag{14}$$

Ainsi le lemme est démontré.

4. Faisons remarquer qu'on pourrait remplacer dans l'énoncé de notre lemme E_n par l'espérance mathématique d'une puissance positive quelconque de $|A_n - B_n|$. La démonstration reste la même sauf que dans l'équation (13) le dénominateur l est à remplacer par la puissance respective de l.

§ 2. Quelques formules concernant les épreuves répétées

1. Nous aurons besoin dans ce qui suit de quelques formules se rapportant au problème des épreuves répétées. Je vais rappeler brièvement ces formules.

Soit $C'_1, C'_2, C'_3, \ldots$ une suite infinie de collectifs dont les caractères distinctifs (les variables aléatoires) ne prennent que k valeurs distinctes et connues d'avance $f_1, f_2, \ldots f_k$. La probabilité pour que dans C'_ν la variable prenne la valeur $f_\varkappa$ sera donnée par $p'_{\nu\varkappa}$. Une expérience effectuée sur les n premiers de ces collectifs formera l'élément du collectif composé C_n. Dans C_n la probabilité d'un résultat tel que

$$f_{\alpha_1} f_{\alpha_2} \ldots f_{\alpha_n} \quad (\alpha_1, \alpha_2, \ldots \alpha_n = 1, 2, \ldots k)$$

se trouve égale au produit des $p'_{\nu\varkappa}$:

$$p'_{1\alpha_1} p'_{2\alpha_2} \ldots p'_{n\alpha_n}. \tag{15}$$

Ce qui nous intéresse c'est la probabilité dans C_n de tous les résultats, différents quant à l'ordre des $f_\varkappa$ et tels que la valeur f_1 s'y trouve $\rho_1\,n$-fois, la valeur f_2 s'y trouve $\rho_2\,n$-fois, etc. Cette probabilité est égale à la somme des expressions (15), étendue à un domaine qui correspond aux chiffres $\rho_1, \rho_2, \ldots \rho_k$. Nous la désignons par $P_n\,(\rho_1, \rho_2, \ldots \rho_k)$ ou simplement par $P_n(\rho)$ et ne chercherons que les moments de premier et de deuxième ordre de $P_n(\rho)$, c'est-à-dire les espérances mathématiques et les dispersions (écarts types) des ρ_k.

2. On connaît bien le fait, facile à démontrer, que si l'on passe de distributions (lois de probabilité) données quelconques à la distribution de la *somme* des variables envisagées (et supposées indépendantes), l'espérance mathématique et la dispersion de cette somme se trouvent par l'*addition* des grandeurs correspondantes primordiales. Or, si nous remplaçons pour le moment dans tous nos collectifs $C'_1, C'_2, \dots C'_n$ la valeur f_1 de la variable aléatoire par 1 et toutes les autres $f_\varkappa$ ($\varkappa \neq 1$) par 0, $n\rho_1$ sera la somme des chiffres sortis. Appelons x la variable modifiée ne prenant que les valeurs 0 et 1 dans C'_ν. La probabilité du résultat $x = 1$ sera $p'_{\nu 1}$, celle de $x = 0$ sera $1 - p'_{\nu 1}$, de sorte que l'espérance mathématique se trouve égale à

$$0.(1 - p'_{\nu 1}) + 1.p'_{\nu 1} = p'_{\nu 1}$$

et la dispersion ou l'espérance mathématique de $(x - p'_{\nu 1})^2$ égale à

$$(0 - p'_{\nu 1})^2(1 - p'_{\nu 1}) + (1 - p'_{\nu 1})^2 p'_{\nu 1} = p'_{\nu 1}(1 - p'_{\nu 1}).$$

D'après ce que nous venons de dire, l'espérance mathématique et la dispersion de $n\rho_1$ sont les sommes de ces expressions, sommes étendues de $\nu = 1$ à n. Donc, d'une façon générale :

$$\sum \rho_\varkappa P_n(\rho) = E_n\{\rho_\varkappa\} = \frac{1}{n} E_n\{n\rho_\varkappa\} = \frac{1}{n}(p'_{1\varkappa} + p'_{2\varkappa} + \dots p'_{n\varkappa}) \tag{16}$$

et d'autre part, en désignant le second membre par $p_{n\varkappa}$

$$\sum [\rho_\varkappa - E_n\{\rho_\varkappa\}]^2 P_n(\rho) = \frac{1}{n^2} E_n\{(n\rho_\varkappa - np_{n\varkappa})^2\} = \frac{1}{n^2}\sum_{\nu=1}^{n} p'_{\nu\varkappa}(1 - p'_{\nu\varkappa}) \tag{17}$$

Ici nous avions posé

$$p_{n\varkappa} = \frac{1}{n}(p'_{1\varkappa} + p'_{2\varkappa} + \dots p'_{n\varkappa}) \tag{18}$$

On en déduit facilement l'identité

$$\frac{1}{n^2}\sum_{\nu=1}^{n} p'_{\nu\varkappa}(1 - p'_{\nu\varkappa}) = \frac{1}{n} p_{n\varkappa}(1 - p_{n\varkappa}) - \frac{1}{n^2}\sum_{\nu=1}^{n}(p'_{\nu\varkappa} - p_{n\varkappa})^2 \tag{19}$$

ce qui nous montre que

$$E_n\{(\rho_\varkappa - p_{n\varkappa})^2\} \leqq \frac{1}{n} p_{n\varkappa}(1 - p_{n\varkappa}) \leqq \frac{1}{n} p_{n\varkappa}. \tag{20}$$

C'est en substance cette formule dont nous nous servirons pour notre démonstration.

3. Si l'on désigne par ρ le vecteur à k dimensions dont les composantes sont $\rho_1, \rho_2, \ldots \rho_k$ et également par p_n le vecteur aux composantes $p_{n1}, p_{n2}, \ldots p_{nk}$ on a

$$(\rho - p_n)^2 = \sum_{\varkappa=1}^{k} (\rho_\varkappa - p_{n\varkappa})^2 \tag{21}$$

Étant donné que

$$\sum_{\varkappa=1}^{k} p_{n\varkappa} = 1 \tag{22}$$

on déduit de (20) et (21) immédiatement l'inégalité

$$E_n \{ (\rho - p_n)^2 \} \leqq \frac{1}{n} \tag{23}$$

Nous reviendrons sur ce résultat au § 4.

4. Jusqu'ici nous avons supposé que les distributions données des C'_ν fussent discontinues. Mais on peut étendre les résultats au cas de distributions continues de la manière suivante.

Soit $V'_\nu(x)$ la probabilité pour que dans C'_ν la variable aléatoire ne dépasse pas la valeur x. De l'autre côté, dans le collectif composé C_n nous désignons par $nS(x)$ le nombre des résultats dont la valeur est inférieure ou égale à x. Prenons une valeur fixe pour x et remplaçons dans tous les C'_ν les valeurs de la variable aléatoire inférieures ou égales à x par 1, toutes les autres par 0. En ce cas $V'_\nu(x)$ et $1\text{-}V'_\nu(x)$ sont les probabilités du résultat 1 resp. 0 dans C'_ν et $nS(x)$ est la *somme* des résultats qui se produisent dans une épreuve effectuée sur C_n. On peut donc appliquer ici la formule (16) en posant $S(x)$ pour $\rho_\varkappa$ et $V'_\nu(x)$ pour $p_{\nu\varkappa}$

$$E_n \{ S(x) \} = \frac{1}{n} [V'_1(x) + V'_2(x) + \ldots V'_n(x)] \tag{24}$$

Introduisons $V_n(x)$ pour le deuxième membre :

$$V_n(x) = \frac{1}{n} [V'_1(x) + V'_2(x) + \ldots V'_n(x)] \tag{25}$$

nous tirerons de (20) l'inégalité

$$E_n \{[S(x) - V_n(x)]^2\} \leqq \frac{1}{n} V_n(x) [1 - V_n(x)] \tag{26}$$

qui est valable pour toute valeur de x.

Soit maintenant $\psi(x)$ une fonction non-négative de x et telle que l'intégrale de $-\infty$ à $+\infty$

$$J = \int \psi(x) [S(x) - V_n(x)]^2 dx \tag{27}$$

existe. Il s'ensuit de (26) que

$$E_n\{J\} = E_n\left\{\int \psi(x) [S(x) - V_n(x)]^2 dx\right\} \leqq \frac{1}{n} \int \psi(x) V_n(x) [1 - V_n(x)] dx \tag{28}$$

Ce résultat nous servira dans le cas général de notre démonstration au § 5.

§ 3. Distributions arithmétiques

1. Avant d'aborder le problème général il sera utile d'étudier le cas spécial de distributions discontinues ou arithmétiques. Si les variables aléatoires ne peuvent prendre qu'une de k valeurs fixées d'avance, la répartition de n résultats d'une épreuve composée est entièrement déterminée par k fréquences relatives $\rho_1, \rho_2, \ldots \rho_k$ dont la somme est l'unité. En ce cas une *fonction statistique des résultats est simplement une fonction de k variables* $\rho_1, \rho_2, \ldots \rho_k$.

On peut partir d'une suite de collectifs $C'_1, C'_2, C'_3 \ldots$ à distributions quelconques pourvu que ces distributions soient définies toutes dans le même espace caractéristique. Imaginons par exemple les tirages effectués dans une suite d'urnes dont chacune est remplie de billets portant des chiffres réels quelconques. Nous subdivisons l'espace caractéristique — dans l'exemple l'axe réel — en k parties $L_1, L_2, \ldots L_k$ et désignons par $p'_{\nu\varkappa}$ la probabilité pour que le caractère distinctif (la variable aléatoire) de C'_ν tombe dans $L_\varkappa$ ($\varkappa = 1, 2, \ldots k$; $\nu = 1, 2, 3 \ldots$). De cette façon les distributions primordiales des C'_ν sont réduites à des distributions discontinues, définies par les $p'_{\nu\varkappa}$. Le théorème que nous allons établir dans le présent paragraphe porte sur tout système de collectifs à distributions arithmétiques.

Nous désignons par C_n la composition de $C'_1, C'_2, \ldots C'_n$ (donc la suite de tirages effectués dans les n premières urnes) et par $p_{n\varkappa}$ la moyenne arithmétique des $p'_{\nu\varkappa}$, voir (18). Le résultat d'un tirage composé, ou un élément de C_n, comprend $\rho_1 n$ valeurs tombant dans L_1, puis $\rho_2\, n$ tombant dans L_2,... enfin $\rho_k\, n$ valeurs qui appartiennent à L_k. Les variables $\rho_\varkappa$ varient dans un domaine D défini par

$$(29) \qquad D : \rho_1 \geqq 0,\ \rho_2 \geqq 0,\ \ldots\ \rho_k \geqq 0\,;\ \rho_1 + \rho_2 + \ldots \rho_k = 1$$

A côté du point $\rho_1, \rho_2, \ldots \rho_k$ nous étudions les points dont les coordonnées sont $p_{n1}, p_{n2} \ldots p_{nk}$ ($n = 1, 2, \ldots$). Ils se trouvent dans D pour toute valeur de n. A partir d'un certain n ces points ne rempliront qu'une certaine partie D_1 de D qui, dans des cas spéciaux, pourra se réduire de plus en plus à un seul point. Pour abréger nous écrirons aussi ρ au lieu de $\rho_1, \rho_2, \ldots \rho_k$ et p_n au lieu de $p_{n1}, p_{n2}, \ldots p_{nk}$, quand il ne s'agit que d'indiquer les variables dont une fonction dépend. La dérivée d'une fonction $f\,(\rho)$ par rapport à $\rho_\varkappa$ au point p_n (c'est-à-dire pour $\rho_\varkappa = p_{n\varkappa}$, $\varkappa = 1, 2, \ldots k$) sera désignée par $f_\varkappa$.

2. Notre théorème de limite s'exprime, pour le cas de distributions arithmétiques, dans ces termes :

Soit $f\,(\rho)$ *une fonction des fréquences relatives jouissant des propriétés suivantes :*

1° $f\,(\rho)$ *est borné dans le domaine D défini par* (29) ;

2° $f\,(\rho)$ *admet des dérivées continues et bornées de premier et deuxième ordre, en chaque point de* D_1 *à partir d'un certain n ;*

3° *Il existe deux indices* α, β *différents l'un de l'autre et un nombre positif* η *tels qu'à partir d'un certain n :*

$$(30) \qquad p'_{n\alpha} > \eta,\ p'_{n\beta} > \eta\,;\ |f_\alpha - f_\beta| > \eta$$

Sous ces conditions la distribution des probabilités pour $f(\rho)$ *tend vers la distribution de Gauss, si n augmente infiniment. En d'autres termes : La probabilité* P_n (X) *de l'inégalité*

$$(31) \qquad H_n\,[f(\rho) - f(p_n)] \leqq X$$

satisfait, uniformément pour toutes les valeurs de X l'équation

$$(32) \qquad \lim_{n \to \infty} P_n(X) = \Phi(X) = \frac{1}{\sqrt{\pi}} \int^{X} e^{-u^2}\, du$$

où H_n *est défini par*

$$(33) \qquad \frac{1}{2H^2_n} = \frac{1}{n^2} \sum_{\nu=1}^{n} \left[\sum_{\varkappa=1}^{k} f^2_\varkappa p'_{\nu\varkappa} - \left(\sum_{\varkappa=1}^{k} f_\varkappa p'_{\nu\varkappa} \right)^2 \right]$$

3. Quant à la démonstration faisons remarquer d'abord que la fonction $f(\rho)$ admet, pour des n assez grands, en vertu des hypothèses 1° et 2° l'emploi de la *formule de Taylor* de deuxième ordre au point $\rho_\varkappa = p_{n\varkappa}$:

$$f(\rho) - f(p_n) = \sum_{\varkappa=1}^{k} (\rho_\varkappa - p_{n\varkappa}) f_\varkappa + \mathrm{R} \tag{34}$$

$$\mathrm{R} = \frac{1}{2} \sum_{\varkappa,\lambda}^{1\ldots k} (\rho_\varkappa - p_{n\varkappa})(\rho_\lambda - p_{n\lambda}) f_{\varkappa\lambda}(\rho') \tag{35}$$

Ici les $f_{\varkappa\lambda}(\rho')$ sont, d'après l'hypothèse 1°, des fonctions bornées, point sur lequel nous reviendrons ci-dessous.

En nous reportant au lemme établi au § 1 nous posons

$$\mathrm{A}_n = \mathrm{H}_n[f(\rho) - f(p_n)], \qquad \mathrm{B}_n = \mathrm{H}_n \sum_{\varkappa=1}^{k} (\rho_\varkappa - p_{n\varkappa}) f_\varkappa, \tag{36}$$

de sorte que $\mathrm{A}_n - \mathrm{B}_n$ devient égal à H_n R. Evidemment A_n et B_n sont des fonctions dépendant du résultat d'une épreuve qu'on effectue sur les n collectifs $C'_1, C'_2, \ldots C'_n$. Nous chercherons à démontrer :

1° Que la distribution Q_n (X) de B_n tend vers la Gaussienne Φ (X) ;

2° Que l'espérance mathématique de la valeur absolue de la différence $\mathrm{A}_n - \mathrm{B}_n = \mathrm{H}_n$ R tend vers zéro, n augmentant à l'infini.

Ensuite le lemme de § 1 fournira immédiatement le résultat en vue, savoir que P_n (X) tend vers la distribution de Gauss.

4. Pour prouver la première de ces deux propositions fixons le caractère distinctif dans les collectifs $C'_1, C'_2, \ldots C'_n$ de la façon suivante. A toute valeur de la variable aléatoire tombant dans L_1 nous attachons le nombre f_1, à toute valeur tombant dans L_2 le nombre f_2, etc. Les $f_1, f_2, \ldots f_k$ (premières dérivées de f au point $\rho_\varkappa = p_{n\varkappa}$) ne dépendent que de n, ce sont donc des constantes dans l'étude de C_n, collectif composé des $C'_1, C'_2, \ldots C'_n$.

Dans le collectif C'_ν existe maintenant une distribution discontinue où la probabilité $p'_{\nu\varkappa}$ correspond à la valeur $f_\varkappa$ de la variable aléatoire.

La moyenne a_ν et la dispersion $r^2{}_\nu$ (écart type) de cette distribution sont données par

$$(37)\quad a_\nu = \sum_{\varkappa=1}^{k} f_\varkappa p'_{\nu\varkappa}\,;\; r^2{}_\nu = \sum_{\varkappa=1}^{k} (f_\varkappa - a_\nu)^2 p'_{\nu\varkappa} = \sum_{\varkappa=1}^{k} f^2{}_\varkappa p'_{\nu\varkappa} - \left(\sum_{\varkappa=1}^{k} f_\varkappa p'_{\nu\varkappa}\right)^2$$

Si l'on effectue une expérience sur les n collectifs C'_1, C'_2, ... C'_n la *somme* x des variables aléatoires (la somme de différentes valeurs $f_\varkappa$) est égale à

$$(38)\qquad x = n \sum_{\varkappa=1}^{k} \rho_\varkappa f_\varkappa$$

D'après des règles bien connues l'espérance mathématique et la dispersion de cette somme sont, dans C_n, égales aux sommes des a_ν resp. des $r^2{}_\nu$. En tenant compte de (18) et (33) nous trouvons

$$(39)\qquad \sum_{\nu=1}^{n} a_\nu = \sum_{\nu=1}^{n} \sum_{\varkappa=1}^{k} f_\varkappa p'_{\nu\varkappa} = n \sum_{\varkappa=1}^{k} f_\varkappa p_{n\varkappa}\,; \qquad \sum_{\nu=1}^{n} r^2{}_\nu = \frac{n^2}{2H^2{}_n}.$$

De l'autre côté on sait que la distribution d'une somme de n variables indépendantes s'approche, sous certaines conditions, de la fonction de Gauss. Citons comme conditions suffisantes l'existence d'une borne supérieure des moments absolus d'ordre trois et l'existence d'une borne inférieure, différente de zéro, des dispersions (1). Ces deux conditions sont remplies dans notre cas de distributions arithmétiques. Car, en vertu de la deuxième hypothèse les f_k sont bornés d'où on déduit pour $f_\varkappa < M$:

$$(40)\qquad \sum_{\varkappa=1}^{k} |f_\varkappa|^3 p'_{\nu\varkappa} < M^3$$

tandis qu'en vertu de la troisième hypothèse on trouve

$$(41)\qquad \begin{aligned} r^2{}_\nu &\geqq p'_{\nu\alpha}(f_\alpha - a_\nu)^2 + p'_{\nu\beta}(f_\beta - a_\nu)^2 \\ &\geqq 2\eta\left[\left(\frac{f_\alpha - f_\beta}{2}\right)^2 + \left(\frac{f_\alpha + f_\beta}{2} - a_\nu\right)^2\right] > \frac{\eta^3}{2}. \end{aligned}$$

Donc la somme x, définie en (38) *aura pour n infini une distribution Gaussienne* dont la moyenne et la dispersion sont données par (39).

(1) Naturellement on pourrait atténuer ces conditions et, par conséquent, notre hypothèse 3°. Mais cette extension n'est pas de grande importance.

Mais notre fonction B_n introduite dans (36) est une fonction linéaire de x :

$$(42)\qquad B_n = \frac{H_n}{n}\left[x - n\sum_{\varkappa=1}^{k} f_\varkappa p_{n\varkappa}\right] = \frac{H_n}{n}\left[x - \sum_{\nu=1}^{n} a_\nu\right]$$

Il s'ensuit que la distribution $Q_n(X)$ de B_n sera une Gaussienne dont la moyenne s'annule et dont la dispersion est $(H_n : n)^2$ - fois celle de x, donc d'après (39) égale à $\frac{1}{2}$. Ainsi nous avons démontré que

$$(43)\qquad \lim_{n\to\infty} Q_n(X) = \frac{1}{\sqrt{\pi}}\int^{X} e^{-u^2}\,du.$$

5. Quant à la deuxième des propositions du nº 3 nous écrivons le reste de la formule de Taylor (35) sous la forme

$$(44)\qquad R = \frac{1}{2}(\rho - p_n)^2 f''(\rho')$$

Ici ρ et p_n désignent, comme ci-dessus, les vecteurs dont les composantes sont les $\rho_\varkappa$ resp. les $p_{n\varkappa}$ et $f''(\rho')$ la deuxième dérivée de f prise sur la « droite » qui mène du point p_n au point ρ. D'après notre hypothèse 2º les dérivées de second ordre sont bornées, donc on peut supposer $|f''(\rho')| < N$ ce qui donne

$$(45)\qquad |R| < \frac{1}{2}N(\rho - p_n)^2$$

La formule (23) déduite au § 2 nous donne une borne supérieure pour l'espérance mathématique du deuxième membre. Vu que, d'après (34) et (36) la différence $A_n - B_n$ est égale à H_nR, il résulte de (23)

$$(46)\qquad E_n\{|A_n - B_n|\} = H_n E_n\{R\} \leqq \frac{1}{2}H_n N E_n\{(\rho - p_n)^2\} \leqq \frac{H_n N}{2n}$$

Mais la deuxième des équations (39) et l'inégalité (41) nous montrent que

$$(47)\qquad \frac{n^2}{H_n^2} = 2\sum_{\nu=1}^{n} r_\nu^2 > n\eta^3, \qquad \frac{H_n}{n} < \frac{1}{\sqrt{n\eta^3}}$$

donc

$$(48)\qquad \lim_{n\to\infty} E_n\{|A_n - B_n|\} \leqq \lim_{n\to\infty}\frac{N}{2}\frac{H_n}{n} = 0$$

Ainsi les deux conditions du lemme sont remplies *et notre théorème est démontré.*

6. Ajoutons une remarque. Souvent on s'intéresse non autant à la distribution complète d'une fonction telle que $f(\rho)$, mais à sa moyenne b_n et sa dispersion s^2_n seulement. Plusieurs auteurs ont indiqué comme valeurs approchées, valables pour des n assez grands :

$$(49) \qquad b_n \sim f(p_n)\,; \qquad s^2_n \sim \frac{1}{n^2} \sum_{\nu=1}^{n} \left[\sum_{\varkappa=1}^{k} f^2_{\varkappa} p'_{\nu\varkappa} - \left(\sum_{\varkappa=1}^{k} f_{\varkappa} p'_{\nu\varkappa} \right)^2 \right]$$

Mais je n'ai jamais trouvé une déduction correcte de ces formules. Notre démonstration précédente prouve que — sous les conditions 1° à 3° pour les f et les V_ν — les seconds membres de (49) sont la moyenne et la dispersion de la distribution-limite de $f(\rho)$.

D'autre part on peut arriver, sans tenir compte de la condition 3°, à établir directement les formules (49) de la façon suivante. La première expression (39) nous montre que l'espérance mathématique de $\Sigma\,(\rho_\varkappa - p_{n\varkappa})\,f_\varkappa$ s'annule. Étant donné que l'espérance mathématique du deuxième membre de (45) tend vers zéro comme $\frac{1}{n}$ d'après (23), la formule (34) fournit :

$$(50) \qquad E_n \left\{ f(\rho) \right\} = f(p_n) + O\left(\frac{1}{n}\right).$$

Il s'ensuit que

$$s^2_n = E_n \left\{ [f - E_n(f)]^2 \right\} = E_n \left\{ [f(\rho) - f(p_n)]^2 \right\} + O\left(\frac{1}{n^2}\right)$$

Le carré du deuxième membre de (34) se compose de trois termes dont le premier, le carré de $\Sigma\,(\rho_\varkappa - p_{n\varkappa})\,f_\varkappa$, admet comme espérance mathématique exactement le second membre de la deuxième équation (49). Le troisième terme est le carré de R, expression soumise à l'inégalité (45). Mais on peut déduire d'une façon analogue, bien que plus compliquée à celle qui fut employée au § 2 que

$$(51) \qquad E_n \left\{ (\rho - p_n)^4 \right\} \leqq \frac{1}{n^2}$$

ce qui montre que l'espérance mathématique de R^2 est d'ordre $\frac{1}{n^2}$.

Enfin, en appréciant le terme mixte $2 \, R \, \Sigma \, (\rho_\varkappa - p_{n\varkappa}) \, f_\varkappa$ d'après l'inégalité de Schwarz on obtient

$$(52) \qquad s^2_n = \frac{1}{n^2} \sum_{\nu=1}^{n} \left[\sum_{\varkappa=1}^{k} f^2_\varkappa p'_{\nu\varkappa} - \left(\sum_{\varkappa=1}^{k} f_\varkappa p'_{\nu\varkappa} \right)^2 \right] \left[1 + O\left(\frac{1}{n}\right) \right] + O\left(\frac{1}{n^2}\right).$$

Cette équation prouve que le second membre de la deuxième formule (49) donne une approximation pour s^2_n sous la condition que sa valeur dépasse sensiblement $\frac{1}{n^2}$. Dans le cas où notre hypothèse 3° est vérifiée cette valeur est au-dessus de $\eta^3 : 2\, n$. Mais si la somme sur ν est convergente, la formule peut tomber en défaut.

§ 4. Les fonctions statistiques

1. Une « distribution » $V(x)$, on le sait bien, est définie comme fonction monotone, non-décroissante et prenant les valeurs 0 resp. 1 pour $x = \mp \infty$

$$(53) \qquad V(-\infty) = 0, \qquad V(\infty) = 1$$

Nous ajoutons la condition que $V(x)$ soit continue du côté droit, c'est-à-dire qu'en chaque point $V(x)$ soit la limite des valeurs $V(x + \xi)$, si ξ tend vers zéro par valeurs positives. Dans un collectif à une seule dimension la probabilité pour que le caractère distinctif (la variable aléatoire) ne surpasse pas x est toujours donnée par une telle distribution $V(x)$.

D'autre part nous envisageons un ensemble de n nombres réels quelconques, par exemple les résultats de n expériences effectuées sur n collectifs quelconques. J'appelle « *répartition* » de ces n nombres une fonction $S(x)$ définie par le fait que $nS(x)$ est égale au nombre de celles parmi les valeurs $x_1, x_2, \ldots x_n$ qui ne surpassent pas x. Évidemment $S(x)$ est représentée par une ligne-escalier dont les marches sont des multiples entiers de $1/n$. Pour toute valeur de n la répartition de n chiffres réels est un cas particulier d'une distribution.

Nous aurons affaire dans ce qui suit à certains ensembles J de distributions. Nous supposerons toujours qu'un tel ensemble comprenne des répartitions à valeur quelconque de n et au surplus certaines distributions d'autre nature. Soient $V(x)$ et $V_1(x)$ deux distributions appartenant à J. Si l'on fait varier t entre o et 1, les distributions

$$V_1(x) + t[V(x) - V_1(x)], \qquad 0 \leqq t \leqq 1 \tag{54}$$

constituent le « segment de droite » de $V_1(x)$ à $V(x)$. Un ensemble qui comprend tous les segments déterminés par deux de ses éléments s'appellera un *ensemble « convexe »*. Un tel ensemble convexe J pourra être défini par exemple par toutes les distributions $V(x)$ pour lesquelles, dans un certain point $x = x_1$, la valeur $V(x_1)$ est comprise entre deux valeurs données a et b, positives et inférieures à l'unité. Ou bien un ensemble convexe est constitué par les distributions $V(x)$ pour lesquelles le produit $V(1 - V)$ s'annule pour $|x|$ infini comme une certaine puissance négative de $|x|$, etc.

2. Soit J un ensemble de distributions suivant les explications données et attachons à chaque fonction $V(x)$ de J une certaine valeur f : je dirai que f est une *fonction statistique*, définie sur J, et j'écrirai $f\{V(x)\}$.

En effet, toutes les fonctions dont on s'occupe dans la statistique générale, les moyennes, écarts quadratiques ou moments d'ordre quelconque, les coefficients de corrélation, le quotient de Lexis, etc., sont de telles fonctions statistiques. Si, dans l'expression de $f\{V(x)\}$ on substitue pour $V(x)$ une répartition $S(x)$, on peut regarder f comme une fonction des n variables $x_1, x_2, \dots x_n$, fonction jouissant des deux propriétés suivantes : 1) Elle est symétrique, c'est-à-dire la valeur de f ne change pas si l'on remplace x_χ par x_λ et simultanément x_λ par x_χ ; 2) la valeur de f ne change pas, si l'on passe de n à un multiple entier $2n$, $3n$,... et que toute valeur $x_1, x_2, \dots x_n$ se retrouve 2 fois, 3 fois,... parmi les nouvelles variables.

Si l'on n'admet comme « variable indépendante » de f que des fonctions $V(x)$ et $S(x)$ qui restent constantes partout sauf certains points connus au préalable $x = a_1, a_2, \dots a_k$, f s'exprime comme fonction de k variables ordinaires, à savoir comme fonction des fréquences relatives $\rho_1, \rho_2, \dots \rho_\chi$ dans le cas de S, ou comme fonction des probabilités ponctuelles $p_1, p_2, \dots p_k$ au cas de V. Ainsi on revient au problème traité dans le paragraphe précédent où il ne s'agissait que de distributions arithmétiques et de répartitions à points de saut fixes.

3. L'exemple le plus simple d'une fonction statistique est la moyenne arithmétique définie par l'intégrale de Stieltjes $\int x\, d\,\mathrm{V}(x)$. Comme cas plus général examinons la fonction statistique *linéaire*, donnée sous forme d'une intégrale.

$$f\{\mathrm{V}(x)\} = \int \psi(x) d\mathrm{V}(x) \tag{55}$$

où $\psi(x)$ est une fonction continue quelconque. Si nous prenons pour $\mathrm{V}(x)$ la répartition $\mathrm{S}(x)$ des nombres $x_1, x_2, \ldots x_n$ on a

$$f\{\mathrm{S}(x)\} = \frac{1}{n} \sum_{\nu=1}^{n} \psi(x_\nu) \tag{56}$$

Si ψ est borné, f est défini pour toute distribution $\mathrm{V}(x)$, au cas contraire l'ensemble J où f existe est limité par une certaine condition à remplir par les $\mathrm{V}(x)$ dans l'infini. Toujours nous supposerons *absolument* convergentes les intégrales généralisées que nous rencontrons dans les définitions de fonctions statistiques.

Comme exemple d'une fonction statistique *non-linéaire* signalons les moments ou écarts d'ordre m :

$$\mathrm{M}_m = \int (x-a)^m d\mathrm{V}(x), \qquad a = \int x d\mathrm{V}(x) \tag{57}$$

D'une façon plus générale, toute fonction ressortant d'une combinaison d'intégrales de Stieltjes comme (55) est une fonction statistique, par exemple le quotient dit de Lexis (N constant)

$$\mathrm{L} = \frac{\mathrm{NM}_2}{a(\mathrm{N}-a)} \tag{58}$$

Mais il y a encore d'autres cas qui ne se ramènent pas à cette forme, par exemple une intégrale double

$$f\{\mathrm{V}(x)\} = \iint \psi(x,y) d\mathrm{V}(x) d\mathrm{V}(y) \tag{59}$$

Évidemment, on peut étendre la définition des fonctions statistiques aux cas de collectifs à plusieurs dimensions. On montrera dans cet ordre d'idées que le coefficient bien connu de corrélation et de même beaucoup d'autres coefficients caractéristiques d'une distribution sont des fonctions statistiques. Mais je me bornerai dans le présent mémoire au cas d'une seule dimension.

4. Du point de vue de l'analyse les fonctions statistiques sont des « *fonctions de ligne* », notion qui, pour la première fois, fut étudiée de plus près par M. *Vito Volterra*. Toutefois le fait que dans nos cas la « variable indépendante » est restreinte à un ensemble de distributions, comporte certaines modifications. Nous suivons, en substance, M. Volterra en introduisant la notion de *dérivée* d'une fonction statistique de la façon suivante.

Soit $f\{V(x)\}$ définie sur un ensemble convexe J et soit $V_1(x)$ une distribution déterminée appartenant à J. Nous dirons que $f\{V(x)\}$ est *dérivable* au « point » $V_1(x)$, si les deux conditions sont remplies :

1° La fonction de t qui s'exprime par

$$f\{V_1(x) + t(V - V_1)(x)\} \tag{60}$$

est dérivable par rapport à t pour $t = 0$, quelle que soit la distribution $V(x)$ de J ;

2° Cette dérivée qui dépendra de $V(x)$ et $V_1(x)$ s'exprime par une intégrale de Stieltjes

$$\frac{d}{dt} f\{V_1(x) + t(V - V_1)(x)\}_{t=0} = \int f'\{V_1(x), y\}\, d(V - V_1)(y) \tag{61}$$

où f' dépend de $V_1(x)$ et d'une variable y, *mais pas de* $V(x)$.

Ces conditions remplies nous appelons $f'\{V(x), y\}$ la *dérivée* de la fonction statistique « au point » $V(x)$. On voit bien que (59) comporte l'existence d'une sorte de « théorème de la différentielle totale » pour les fonctions statistiques dites dérivables.

La dérivée d'une fonction *linéaire*, donnée sous la forme (55) ne dépend pas de $V_1(x)$ et est égale à $\psi(y)$:

$$\frac{d}{dt} \int \psi(x) d[V_1 + t(V - V_1)](x)_{t=0} = \int \psi(y) d(V - V_1)(y) \tag{62}$$

Si f est fonction de plusieurs intégrales de Stieltjes

$$f = F(A, B, C, \ldots); \qquad A = \int \alpha(x) dV(x), \qquad B = \int \beta(x) dV(x), \ldots \tag{63}$$

la dérivée se trouve sous la forme

$$f'\{V(x), y\} = \frac{\partial F}{\partial A}\alpha(y) + \frac{\partial F}{\partial B}\beta(y) + \ldots \tag{64}$$

Pour l'écart d'ordre m, défini par (57), la dérivée est

$$f'\{V(x),y\} = (y-a)^m - mM_{m-1}y \tag{65}$$

La fonction non-linéaire (59) admet la dérivée

$$f'\{V(x),y\} = \int [\psi(x,y) + \psi(y,x)]dV(x) \tag{66}$$

Faisons observer enfin qu'une constante additive (expression ne dépendant pas de y) ajoutée à f' n'a aucune importance, vu que $\int d(V - V_1)$ s'annule toujours.

5. Nous aurons besoin dans ce qui suit, aussi de la *deuxième dérivée* d'une fonction statistique $f\{V(x)\}$. Elle sera désignée par $f''\{V(x), y, z\}$ et définie par

$$\begin{aligned} &\frac{d^2}{dt^2} f\{V_1(x) + t(V-V_1)(x)\}_{t=0} \\ &= \iint f''\{V(x),y,z,\}\, d(V-V_1)(y) \,.\, d(V-V_1)(z) \end{aligned} \tag{67}$$

Pour une fonction linéaire, exprimée par une intégrale de Stieltjes sous la forme (55) la deuxième dérivée s'annule. Si f dépend de plusieurs expressions linéaires, voir (63), on trouve

$$f''\{V(x),y,z\} = \frac{\partial^2 F}{\partial A^2}\alpha(y)\alpha(z) + 2\frac{\partial^2 F}{\partial A \partial B}\alpha(y)\beta(z) + \dots \tag{68}$$

Le moment ou l'écart de m^ème ordre, défini par (57) admet comme deuxième dérivée

$$f''\{V(x),y,z,\} = -2mz(y-a)^{m-1} + m(m-1)M_{m-2}yz \tag{69}$$

En particulier pour $m = 2$ la seconde dérivée de l'écart quadratique est égale à

$$f''\{V(x),y,z\} = -2yz. \tag{70}$$

Ici nous avons supprimé le terme $4az$, puisqu'une expression ne dépendant que d'une seule des deux variables y et z est sans importance dans f'', vu que l'intégrale $\int d(V - V_1)(x)$ s'annule. Aussi est-il permis d'échanger les variables y, z dans chaque terme de $f''\{V(x), y\, z\}$ sans que (67) ne soit altéré.

Pour la fonction définie par l'intégrale double (59) la deuxième dérivée se trouve indépendante de $V(x)$ et égale à

$$f''\{V(x),y,z\} = 2\psi(y,z) \tag{71}$$

6. Enfin nous allons étudier la *formule de Taylor* pour une fonction statistique, en nous bornant à l'ordre deux. Dans le cas ordinaire d'une fonction d'une seule variable $F(t)$ la formule dont il s'agit est

$$F(1) - F(0) = F'(0) + \frac{1}{2}F''(\vartheta), \qquad 0 \leqq \vartheta \leqq 1 \tag{72}$$

Étant données une distribution fixe $V_1(x)$ et une distribution variable $V(x)$ nous prenons pour $F(t)$ la fonction indiquée par (60)

$$F(t) = f\{V_1(x) + t(V - V_1)(x)\}$$

Le premier membre de (72) et le premier terme du second membre de (72) deviennent égaux à

$$f\{V\} - f\{V_1\} \qquad \text{resp.} \int f'\{V_1(x),y\}\, d(V - V_1)(y) \tag{73}$$

Quant au dernier terme de (72) nous posons

$$V_1(x) + \vartheta(V - V_1)(x) = V_2(x), \qquad \frac{t - \vartheta}{1 - \vartheta} = t' \tag{74}$$

ce qui donne

$$V_1(x) + t(V - V_1)(x) = V_2(x) + t'(V - V_2)(x) \tag{75}$$

La deuxième dérivée par rapport à t au point $t = \vartheta$ est la deuxième dérivée par rapport à t' au point $t' = 0$, divisée par $(1 - \vartheta)^2$:

$$F''(\vartheta) = \frac{1}{(1 - \vartheta)^2} \frac{d^2}{dt'^2} \{V_2(x) + t'(V - V_2)(x)\}_{t'=0} \tag{76}$$

Vu que

$$V - V_2 = (1 - \vartheta)(V - V_1) \tag{77}$$

on arrive à

$$F''(\vartheta) = \iint f''\{V_2(x),y,z\}\, d(V - V_1)(y) \,.\, d(V - V_1)(z) \tag{78}$$

On peut exprimer ce résultat comme suit : *Si la fonction statis-*

tique $f\{V(x)\}$ *est deux fois dérivable sur tout le « segment » menant de* $V_1(x)$ *à* $V(x)$, *une formule « Taylorienne » subsiste :*

$$(79)\qquad f\{V(x)\}-f\{V_1(x)\}=\int f'\{V_1(x),y\}\,d(V-V_1)(y)$$
$$+\frac{1}{2}\int f''\{V_2(x),y,z\}\,d(V-V_1)(y)\,.\,d(V-V_1)(z)$$

où $V_2(x)$ *est définie par* (74) *pour une valeur de* ϑ *entre* 0 *et* 1.

Cette formule en combinaison avec le lemme déduit au § 1 nous fournira le moyen de démontrer notre théorème général. Ajoutons qu'il peut arriver que le premier terme du deuxième membre de (79) devient identiquement nul. C'est le cas p. e. pour la fonction

$$(80)\qquad \omega^2\{V(x)\}=\int \lambda(x)[V(x)-V_1(x)]^2dx$$

que j'ai introduite il y a quelque temps comme un « critère » dans la statistique mathématique (1).

§ 5. Le théorème général

1. Il s'agit du même problème qui était exposé au § 3, seulement seront admises maintenant les distributions continues. On donne une suite infinie de collectifs C'_1, C'_2, C'_3... chacun à une seule dimension. Les distributions correspondantes sont $V'_1(x)$, $V'_2(x)$, $V'_3(x)$... Une expérience effectuée sur les n premiers de ces collectifs fournit n nombres réels x_1, x_2,... x_n dont la répartition sera désignée par $S_n(x)$. Si $f\{V(x)\}$ est une fonction statistique, l'expérience envisagée fait ressortir une certaine valeur $f\{S_n(x)\}$. D'après les règles élémentaires du calcul des probabilités la valeur de $f\{S_n(x)\}$ est soumise à une distribution (loi de probabilité) P_n (X) qui est entièrement définie par $V_1'(x)$, $V'_2(x)$,... $V'_n(x)$. Notamment, H_n et K_n étant deux constantes, P_n (X) signifiera la probabilité pour que la valeur de $H_n[f\{S_n(x)\}-K_n]$, déduite de l'expérience sur les n premiers collectifs, ne surpasse pas X. Nous nous proposons de démontrer que, sous certaines conditions, P_n (X) tend vers la distribution de *Gauss* pour n infini.

(1) Voir le livre cité au début, p. 316.

Le théorème classique de *Laplace-Tchebychef* arrête le même fait pour le cas où f se réduit à la moyenne arithmétique $\int x d S_n(x)$. On voit immédiatement que le théorème s'étend sans difficulté à une fonction statistique *linéaire* quelconque. Or, ce qui nous intéresse c'est surtout le cas de fonctions non-linéaires.

Il peut arriver que les répartitions qui se produisent au cours d'une expérience sont soumises à une certaine restriction, par exemple que leurs points de saut ne se trouvent que dans la partie positive de l'axe des x, etc. En tout cas il y aura un certain ensemble de répartitions admissibles pour toute valeur de n. D'autre part les distributions données $V'_\nu(x)$ donnent lieu à une autre suite de distributions, les moyennes arithmétiques définies par

$$(81) \qquad V_n(x) = \frac{1}{n}[V'_1(x) + V'_2(x) + \dots V'_n(x)], \qquad n = 1, 2, \dots$$

Nous supposerons dans ce qui suit, que $f\{V(x)\}$ soit définie dans un ensemble convexe J qui comprend toutes les répartitions admissibles et tous les $V_n(x)$, au moins à partir d'un certain n.

2. Maintenant je vais donner l'énoncé précis de notre premier théorème de limite.

Soit f $\{V(x)\}$ *une fonction statistique satisfaisant aux conditions suivantes :*

1° f $\{V(x)\}$ *est deux fois dérivable dans un ensemble convexe J qui comprend toutes les répartitions* $S_n(x)$ *qui peuvent se présenter au cours des expériences, et toutes les distributions moyennes* $V_n(x)$ *définies par* (81), *au moins à partir d'un certain* n ;

2° *La première dérivée* f$'\{V_n(x), z\}$ *remplit des conditions suffisantes pour la validité du théorème classique sur l'application de la loi de Gauss.* Par exemple, si pour $\nu = 1, 2, 3,\dots$

$$(82) \quad a_\nu = \int f'\{V_n(x),z\}\, dV'_\nu(z), \qquad r^2_\nu = \int [f'\{V_n(x),z\} - a_\nu]^2 dV'_\nu(z)$$

$$C_\nu = \int |f'\{V_n(x),z\} - a_\nu|^{2+\varepsilon} dV'_\nu(z)$$

il suffit que $s_n^2 = r_1^2 + r_2^2 + \dots r_n^2$ divisé par $n^{\frac{2}{2+\varepsilon}}$ ($\varepsilon > 0$) tende vers l'infini et que les C_ν soient bornés ;

3° *Pour la seconde dérivée* f″$\{V(x), y, z\}$ *il existe une fonction positive* ψ (x) *telle que pour* $T_n(x) = V_n(x) - S_n(x)$, *quelle que soit la distribution* $V(x)$ *de* J, *l'inégalité*

$$\left| \iint f''\{V(x),y,z\}\, dT_n(y)dT_n(z) \right| \leqq \int \psi(x)T^2_n(x)dx \tag{83}$$

entraîne la relation

$$\lim_{n\to\infty} \frac{1}{s_n}\int \psi(x)V_n(x)[1 - V_n(x)]dx = 0. \tag{84}$$

Sous ces conditions la distribution des probabilités de f$\{S_n(x)\}$ *tend vers la distribution de Gauss pour* n *infini. En d'autres termes : Si* $P_n(X)$ *signifie la probabilité de l'inégalité*

$$H_n\,[f\{S_n(x)\} - f\{V_n(x)\}] \leqq X \tag{85}$$

on a, uniformément pour toutes les valeurs de X :

$$\lim_{n\to\infty} P_n(X) = \frac{1}{\sqrt{\pi}}\int^{X} e^{-u^2}du = \Phi(X) \tag{86}$$

où H_n *est déterminé par*

$$\frac{1}{2H^2_n} = \frac{s^2_n}{n^2} = \tag{87}$$

$$= \frac{1}{n^2}\sum_{\nu=1}^{n}\left[\int f'^2\{V_n(x),y\}\, dV'_\nu(y) - \left(\int f'\{V_n(x),y\}\, dV'_\nu(y)\right)^2\right]$$

3. La démonstration se fonde sur la formule de Taylor que nous avons déduite au paragraphe précédent. Si l'on porte dans (79) $S_n(x)$ à la place de $V(x)$ et $V_n(x)$ à la place de $V_1(x)$, on obtient, en multipliant par H_n :

$$H_n\,[f\{S_n(x)\} - f\{V_n(x)\}] = H_n\int f'\{V_n(x),y\}\, dT_n(y) \tag{88}$$
$$+ \frac{1}{2}H_n\iint f''\{V(x),y,z\}\, dT_n(y)dT_n(z)$$

Ici T_n est la différence $S_n - V_n$ et $V(x)$ au dernier terme signifie une distribution située sur le « segment » de S_n à V_n.

Écrivons A_n pour le premier membre, B_n pour le premier terme du deuxième, il vient

$$A_n = H_n\,[f\{S_n(x)\} - f\{V_n(x)\}], \qquad B_n = H_n\int f'\{V_n(x),y\}\, dT_n(y) \tag{89}$$

(90) $$A_n - B_n = \frac{1}{2} H_n \iint f'' \{ V(x), y, z \} \, dT_n(y) dT_n(z)$$

L'emploi de (79) est permis en vertu de la première hypothèse sur $f \{ V(x) \}$.

Selon le lemme établi au § 1 il suffira maintenant de démontrer :

1° que la distribution $Q_n(X)$ de B_n tend vers la Gaussienne Φ (X) ;

2° Que l'espérance mathématique de la valeur absolue de la différence $|A_n - B_n|$ tend vers zéro, si n augmente infiniment.

On en déduira, en vertu du lemme cité, que la distribution P_n (X) de A_n tend vers la Gaussienne, elle aussi, et ainsi notre théorème général sera prouvé.

4. Quant à la première de ces deux propositions, introduisons dans (89) $T_n x) = S_n(x) - V_n(x)$, en tenant compte de (81) et de la définition d'une intégrale comme $\int \psi dS_n(x)$, voir (55), (56). Ainsi on obtient

(91) $$B_n = \frac{H_n}{n} \sum_{\nu=1}^{n} f' \{ V_n(x), x_\nu \} - \frac{H_n}{n} \sum_{\nu=1}^{n} \int f' \{ V_n(x), y \} \, dV'_\nu(y)$$

Ici, pour un n fixe, les $x_1, x_2 \dots, x_n$ signifient les résultats immédiats de n épreuves effectuées sur les n premiers des collectifs donnés $C'_1, C'_2, \dots C'_n$. Imaginons qu'on change le caractère distinctif de C'_ν de sorte que la valeur x soit remplacée par $f' \{ V_n(x), x_\nu \}$. En ce cas la première somme dans (91) est une somme de n variables aléatoires indépendantes, et chaque terme de la deuxième somme est égal à la valeur moyenne d'une de ces variables. Si nous écrivons, pour abréger, f_ν pour $f' \{ V_\nu(x), x_\nu \}$, en nous servant en même temps de la notation a_ν introduite par la première des équations (82), nous arrivons à l'expression simple

(92) $$B_n = \frac{H_n}{n} \sum_{\nu=1}^{n} (f_\nu - a_\nu)$$

Abstraction faite du facteur constant $\frac{H_n}{n}$ nous voyons ici une somme de n variables aléatoires dont les valeurs moyennes s'annulent. La deuxième hypothèse sur les fonctions $f \{ V(x) \}$ garantit que la distribution de cette somme tend vers une distribution de Gauss. D'après la deuxième des équations (82) la dispersion de la $\nu^{\text{ième}}$ de nos variables aléatoires est égale à r^2_ν, donc la dispersion de leur somme égale à $s_n^2 = r_1^2 + r_2^2 + \dots r_n^2$, enfin la dispersion de l'expression complète B_n égale à

$$(93) \qquad \left(\frac{H_n}{n}\right)^2 (r^2_1 + r^2_2 + \dots r^2_n) = \frac{H^2_n s^2_n}{n^2} = \frac{1}{2}$$

en vertu de (87). Ainsi il est démontré : *Si $Q_n(X)$ signifie la probabilité pour que B_n ne surpasse pas X, on a*

$$(94) \qquad \lim_{n \to \infty} Q_n(X) = \Phi(X) = \frac{1}{\sqrt{\pi}} \int^X e^{-u^2} du.$$

5. Examinons maintenant la différence $A_n - B_n$ définie par (90). D'après la troisième hypothèse sur les fonctions $f\{V(x)\}$ admises, voir l'équation (83), on a

$$(95) \qquad |A_n - B_n| \leqq \frac{H_n}{2} \int \psi(x)[S_n(x) - V_n(x)]^2 dx.$$

Il s'agit de trouver l'espérance mathématique du deuxième membre. Pour y arriver il ne faut qu'utiliser les formules établies au § 2.

Les équations (27) et (28) nous donnent immédiatement :

$$(96) \qquad |A_n - B_n| \leqq \frac{H_n}{2} J$$

donc

$$(97) \qquad E_n\{|A_n - B_n|\} \leqq \frac{H_n}{2} E_n(J) \leqq \frac{H_n}{2n} \int \psi(x) V_n(x)[1 - V_n(x)] dx$$

Mais le facteur $H_n : 2n$ s'exprime à l'aide de (87) :

$$(98) \qquad \frac{H_n}{2n} = \frac{1}{2 s_n \sqrt{2}}$$

D'après l'équation (84) de l'hypothèse 3° il s'ensuit donc que

$$(99) \qquad \lim_{n \to \infty} E_n\{|A_n - B_n|\} = 0$$

Or, les deux conditions (9), (9') du lemme du § 1 étant remplies, on est arrivé au résultat définitif : *Si $P_n(X)$ signifie la probabilité pour que la valeur de A_n, définie par* (89) *ne surpasse pas X, on a*

$$(100) \qquad \lim_{n \to \infty} P_n(X) = \Phi(X) = \frac{1}{\sqrt{\pi}} \int^X e^{-u^2} du$$

Ainsi le théorème général est établi.

6. Nous ajoutons quelques remarques sur les fonctions $f\{V(x)\}$ qui satisfont à nos hypothèses. Tout d'abord, il n'y a aucune question pour les fonctions linéaires $\int \alpha(x)dV(x)$. Leur deuxième dérivée s'annule, la première est égale à $\alpha(x)$, indépendante de $V(x)$. Les seules conditions à remplir sont, quant aux $\alpha(x)$, les conditions nécessaires pour que le théorème classique subsiste pour les variables aléatoires $\alpha(x)$ soumises aux lois de distribution $V'_\nu(x)$.

Quant aux fonctions non-linéaires le type qui se présente le plus souvent est celui qui était signalé dans (63), (64) et (68), c'est-à-dire une fonction F de plusieurs intégrales de Stieltjes A, B, C,...

$$(63) \qquad f = F(A, B, C, \ldots)\,;\; A = \int \alpha(x)dV(x),\; B = \int \beta(x)dV(x), \ldots$$

Ici on pourra supposer que les dérivées de premier et de second ordre de F par rapport aux A, B, C,... soient bornées, si $V(x)$ est restreinte aux distributions $V_n(x)$ à partir d'un certain n. D'après (64) la première dérivée de $f\{V(x)\}$ se compose de termes comme

$$\frac{\partial F}{\partial A}\alpha(y), \qquad \frac{\partial F}{\partial B}\beta(y), \ldots$$

Notre hypothèse 2° sera donc remplie, si chacune des fonctions $\alpha(x)$, $\beta(x)$,... satisfait aux conditions du théorème classique, au sens indiqué ci-dessus.

Pour examiner la troisième hypothèse constatons d'abord que, selon (68), la seconde dérivée de f se compose de termes comme

$$\frac{\partial^2 F}{\partial A^2}\alpha(y)\alpha(z), \qquad \frac{\partial^2 F}{\partial A \partial B}\alpha(y)\beta(z), \ldots$$

Étant donné que les $\partial^2 F/\partial A^2$ etc. restent bornés, le premier membre de (83) dépendra d'expressions comme

$$(101) \qquad \iint \alpha(y)\beta(z)dT_n(y)dT_n(z) = \int \alpha(x)dT_n(x) \,.\, \int \beta(x)dT_n(x).$$

Supposons les $\alpha(x)$, $\beta(x)$,... dérivables et telles que les premières dérivées $\alpha'(x)$, $\beta'(x)$,... possèdent la majorante $\psi_1(x)$

$$(102) \qquad |\alpha'(x)| < \psi_1(x),\; |\beta'(x)| < \psi_1(x), \ldots$$

On trouve en intégrant par parties

$$\int \alpha(x)dT_n(x) = -\int \alpha'(x)T_n(x)dx$$

et, en appliquant l'inégalité de Schwarz pour un $\psi(x) > 0$

$$(103 \qquad \left|\int \alpha'(x) T_n(x) dx\right|^2 \leqq \int \frac{\alpha'^2(x)}{\psi(x)} dx \,.\, \int \psi(x) T^2_n(x) dx$$

Il s'ensuit en vertu de (101) et (102) que

$$(104) \qquad \left|\iint \alpha(y)\beta(z) dT_n(y) dT_n(z)\right| \leqq \int \frac{\psi_1^2(x)}{\psi(x)} dx \,.\, \int \psi(x) T^2_n(x) dx$$

Donc pour que notre hypothèse 3° soit remplie, il suffit que l'intégrale

$$(105) \qquad \int \frac{\psi_1^2}{\psi} dx$$

soit convergente pour une fonction $\psi(x)$ pour laquelle (84) subsiste.

Si par exemple $f\{V(x)\}$ est une fonction de la moyenne et de la dispersion (comme le coefficient de Lexis) on a $\alpha(x) = x$, $\beta(x) = (x\text{-}a)^2$ où a est restreint à un intervalle fini. Il existe donc une constante $c > 1$ de sorte que $\psi_1(x) = c + 2|x|$ surpasse $|\alpha'(x)|$ et $|\beta'(x)|$. L'intégrale (105) sera convergente, si ψ augmente dans l'infini comme $|x|^{3+\varepsilon}$ pour un $\varepsilon > 0$. Donc, si les distributions données $V'_\nu(x)$ remplissent la condition que le produit $|x|^{4+\varepsilon}\, V'_\nu(x)$ pour $x = -\infty$ et le produit $|x|^{4+\varepsilon}[1 - V'_\nu(x)]$ pour $x = \infty$ restent au-dessous d'un nombre indépendant de ν, les intégrales (84) seront convergentes uniformément par rapport à n et, par conséquent, l'hypothèse 3° sera vérifiée. D'une façon plus générale, si f dépend de moments jusqu'à l'ordre m, il suffit que les $V'_\nu(x)$ resp. les $1 - V'_\nu(x)$ s'annulent dans l'infini au moins comme $|x|^{-2m-2-\varepsilon}$. D'ailleurs la seconde hypothèse demande, en substance, que les intégrales comme $\int \alpha^{2m+\varepsilon} dV'_\nu$, etc., convergent, donc que les V'_ν resp. $1 - V'_\nu$ s'annulent dans l'infini au moins comme $|x|^{-2m-1-\varepsilon}$ pour un ε positif.

D'autre part pour la fonction ω^2, mentionnée dans (80), on voit que notre hypothèse 2° n'est aucunement remplie si l'on prend $V_n(x)$ à la place de $V_1(x)$. Ici la première dérivée $f'\{V_n(x), y\}$ est identiquement nulle, comme il était dit à la fin du § 4, donc tous les r^2_ν s'annulent, etc. M. *N. Smirnoff* (1) vient de publier un résultat très intéressant concernant la distribution-limite de ω^2.

(1) C. R. Paris, t. CCII, 1936, p. 449.