BULLETIN DE LA S. M. F.

JACQUES CHAPELON

Sur l'inégalité fondamentale du calcul des probabilités

Bulletin de la S. M. F., tome 65 (1937), p. 100-108

<http://www.numdam.org/item?id=BSMF_1937__65__100_0>

© Bulletin de la S. M. F., 1937, tous droits réservés.

L'accès aux archives de la revue « Bulletin de la S. M. F. » (http://smf.emath.fr/Publications/Bulletin/Presentation.html) implique l'accord avec les conditions générales d'utilisation (http://www.numdam.org/conditions). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme Numérisation de documents anciens mathématiques

http://www.numdam.org/

SUR L'INÉGALITÉ FONDAMENTALE DU CALCUL DES PROBABILITÉS

PAR M. JACQUES CHAPELON.

1. Supposons que l'on ait une loi de probabilité, par exemple à variate continue, $p(x)\,dx$, c'est-à-dire une distribution de la masse unité, de densité $p(x)$, portée par l'axe des x. Si m_1 est l'abscisse du centre d'inertie G de la distribution, $\mu_2=\sigma^2$ le moment d'inertie de la distribution par rapport au centre d'inertie, on sait que si l'on porte de part et d'autre de G une longueur $t\sigma$, la somme P des masses supportée par ce segment de longueur $2\,t\sigma$ est supérieure à $1-\frac{1}{t^2}$, la démonstration résultant immédiatement de ce que μ_2 est supérieur ou égal à $t^2\sigma^2(1-\mathrm{P})$.

Je me suis proposé de généraliser ce résultat pour une loi de probabilité à plusieurs variates.

2. On fera les démonstrations dans le cas de deux dimensions seulement, les procédés utilisés étant généraux.

Soit donc une distribution de la masse unité avec la densité superficielle $p(x_1, x_2)$. Supposons qu'elle ait un centre d'inertie G, de coordonnées m_1 et m_2 et qu'elle soit rapportée à des axes de coordonnées parallèles aux axes donnés et passant par le centre d'inertie. Les moments du second ordre de la distribution sont alors définis par les trois intégrales

$$\mu_{ik}=\int\!\!\int_{w} x_i x_k\, p(x_1, x_2)\,dx_1\,dx_2 \qquad (i, k=1, 2),$$

qu'on suppose convergentes et qui sont étendues au domaine d'existence des masses w.

Posons

$$g(u_1, u_2)=\int\!\!\int_{w}(u_1 x_1+u_2 x_2)^2\,dx_1\,dx_2=\mu_{11}u_1^2+2\mu_{12}u_1u_2+\mu_{22}u_2^2.$$

La conique d'équation tangentielle

$$(c) \qquad g(u_1, u_2) = 1$$

est une ellipse. On l'appellera *ellipse type*.

On peut également introduire un *rectangle type*, (r) : c'est le rectangle de côtés parallèles aux axes et circonscrit à l'ellipse type. Ses côtés ont respectivement pour longueurs $2\sigma_1$ et $2\sigma_2$, avec $\sigma_1^2 = \mu_{11}$, $\sigma_2^2 = \mu_{22}$.

Dans le cas d'une distribution de masses à n dimensions, $p(x_1, x_2, \ldots, x_n)\, dx_1 dx_2, \ldots, dx_n$, rapportée à son centre d'inertie G, il s'introduit une *quadrique type*, du genre ellipsoïde et d'équation tangentielle

$$(Q) \qquad \sum_{ik} \mu_{ik} u_i u_k = 1 \qquad (i, k = 1, 2, \ldots, n),$$

et un *parallélépipède type* (Π), circonscrit à la quadrique et coupant l'axe Gx_i aux points $x_i = \pm \sigma_i$ avec $\sigma_i^2 = \mu_{ii}$, $(i = 1, 2, \ldots n)$.

3. Prenons l'ellipse concentrique à (c) et homothétique dans le rapport t. Son équation est

$$(c_t) \qquad t^2 g(u_1, u_2) = 1 \qquad (t \geqq \sqrt{2}).$$

Le rectangle (r_t), circonscrit à (c_t) est concentrique à (r), et son homothétique dans le rapport t. Soit ABCD le rectangle (r_t). Décomposons le domaine d'existence des masses en trois parties : 1° le rectangle (r_t), qui contient la masse P et donne, dans le calcul de μ_{11} une contribution μ'_{11}; 2° les deux demi-bandes parallèles à Gx_2 et ayant pour base AB et CD; elles contiennent la masse P′ et donnent une contribution μ''_{11} dans le calcul de μ_{11}; 3° le domaine restant qui contient donc la masse $1 - P - P'$ et donne une contribution μ'''_{11}.

D'ailleurs, μ_{11} est le moment du second ordre de la distribution obtenue en projetant toutes les masses sur Gx_1 parallèlement à Gx_2. Donc, $\mu'_{11} + \mu''_{11}$ correspond à une distribution de masses totale $P + P'$ portée par la projection de AB sur Gx_1 et μ'''_{11}, à une distribution de masses portée par les deux segments indéfinis de l'axe des abscisses et (n° 1), on sait que

$$P + P' \geqq 1 - \frac{1}{t^2}.$$

Raisonnant sur le moment μ_{22} comme on vient de le faire sur le moment μ_{11}, il faudra introduire une masse Q′ contenue dans les demi-bandes de bases AD et BC et parallèles à Ox_1 et l'on aura, de même,

$$P + Q' \geqq 1 - \frac{1}{t^2}.$$

Donc

$$2P + P' + Q' \geqq 2 - \frac{2}{t^2},$$

et comme

$$1 \geqq P + P' + Q',$$

il en résulte

$$P \geqq 1 - \frac{2}{t^2}.$$

Plus généralement, dans le cas de n variates :

Si l'on considère une distribution de la masse unité dans un espace à n dimensions, la somme des masses intérieures à un parallélépipède concentrique au parallélépipède type et homothétique dans le rapport t, est supérieure à

$$1 - \frac{n}{t^2} \qquad (t \geqq \sqrt{n}).$$

4. On peut obtenir un résultat plus précis en raisonnant de la manière suivante.

Reprenons l'ellipse (c) et opérons sur le point $M(x_1, x_2)$ une transformation affine l'amenant en $M'(X_1, X_2)$ de telle sorte que la masse attachée au point M passe au point M′ ou, plus précisément, si $p(x_1, x_2)\,dx_1\,dx_2$ est la masse de l'élément $dx_1\,dx_2$ en M, que la masse de l'élément correspondant soit la même.

Alors, si

$$\begin{aligned} x_1 &= a X_1 + b X_2 \\ x_2 &= a' X_1 + b' X_2 \end{aligned} \qquad (ab' - ba' \neq 0),$$

on aura

$$p(aX_1 + bX_2,\ a'X_1 + b'X_2)\,(ab' - ba')\,dX_1\,dX_2 = p(x_1, x_2)\,dx_1\,dx_2.$$

Si pour la nouvelle distribution on emploie les mêmes notations pour les moments, mais en prenant des grandes lettres, on aura, en tenant compte de l'identité

$$u_1 x_1 + u_2 x_2 = U_1 X_1 + U_2 X_2$$

$$\sum_{ik} \mu_{ik} u_i u_k = \sum_{ik} M_{ik} U_i U_k \qquad (i, k = 1, 2),$$

de telle sorte que la conique (c) se transforme en la conique d'équation tangentielle

$$(C) \qquad G(U_1, U_2) = M_{11} U_1^2 + 2 M_{12} U_1 U_2 + M_{22} U_2^2 = 1.$$

Or, on peut toujours choisir la transformation affine de telle sorte que (C) soit un cercle (Γ). Alors $M_{11} = M_{22} = \Sigma^2$ et $M_{12} = 0$.

Puis la conique (c_t) devient, par cette transformation affine particulière, le cercle (Γ_t) d'équation tangentielle

$$(\Gamma_t) \qquad t^2 \Sigma^2 (U_1^2 + U_2^2) = 1,$$

et d'équation ponctuelle

$$X_1^2 + X_2^2 = t^2 \Sigma^2.$$

Ceci posé, soit P la masse intérieure à la conique (c_t). C'est aussi la masse intérieure au cercle (Γ_t). Or, pour ce cercle, il est aisé d'évaluer une limitation de P. Car, le moment d'inertie de la distribution par rapport à G est la somme des moments d'inertie par rapport aux axes de coordonnées, donc il est égal à $2\Sigma^2$. Puis, dans le calcul de ce moment d'inertie, la masse extérieure au cercle (Γ_t) donne au moins la contribution

$$(1 - P) t^2 \Sigma^2,$$

et la donne effectivement quand toute la masse extérieure, $1 - P$, est étalée sur la circonférence du cercle (Γ_t). Négligeant la contribution de la masse intérieure au cercle (Γ_t), on obtient l'inégalité

$$2 \Sigma^2 \geqq (1 - P) t^2 \Sigma^2,$$

donc

$$P \geqq 1 - \frac{2}{t^2}.$$

Plus généralement, dans le cas de n variates :

Si l'on considère une distribution de la masse unité dans un espace à n dimensions, la somme des masses intérieures à une quadrique (Q_t), *concentrique à la quadrique type* (Q), *et* *homothétique dans le rapport t est supérieure à*

$$1 - \frac{n}{t^2} \qquad (t \geqq \sqrt{n}).$$

Le théorème du n° 3 est évidemment contenu dans celui-ci.

5. Si, au lieu de prendre comme quadrique de référence la quadrique type (Q), on avait pris la quadrique $(Q_{\sqrt{n}})$, en appelant $(Q_{\tau\sqrt{n}})$ la quadrique contenant la masse P, on aurait eu

$$P \geqq 1 - \frac{1}{\tau^2} \qquad (\tau \geqq 1).$$

Les quadriques $(Q_{\tau\sqrt{n}})$ ont un intérêt géométrique particulier. Appelons $s_1, s_2, \ldots, s_n$ les racines de l'équation en s de la forme quadratique $\sum_{ik} \mu_{ik} u_i u_k$: ce sont les carrés des demi-axes de la quadrique (Q).

Alors, le volume de la quadrique $(Q_{\tau\sqrt{n}})$ est, en posant

$$g = \sqrt[n]{s_1 s_2 \ldots s_n},$$

$$V = \frac{(n\pi)^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2} + 1\right)} g^{\frac{n}{2}} \tau^n,$$

où g est la moyenne géométrique des carrés des demi-axes. Supposons g borné supérieurement, et inférieurement par un nombre positif.

La formule de Stirling donne, si n est grand,

$$V \sim \frac{1}{\sqrt{\pi}} e^{\frac{n}{2}\log 2\pi e g \tau^2 - \frac{1}{2}\log n}$$

et, par suite, V est petit, pour n grand, si $2\pi e g\tau^2 \leqq 1$, et au contraire est grand, pour n grand, si $2\pi e g \tau^2 > 1$.

Comme, dans l'application du théorème, on suppose $\tau \geqq 1$, on voit qu'on a la valeur critique $g = \frac{1}{2\pi e}$ séparant des modes possibles différents de distribution de la matière.

6. **Théorèmes du type de Tchebychef.** — On obtient immédiatement les théorèmes du type de Tchebychef pour des systèmes de variates indépendantes ou non en appliquant l'inégalité fondamentale et tenant compte des propositions qui suivent.

On sait que si l'on a une loi de probabilité $p(x_1, x_2, \dots, x_n)$ $dx_1, dx_2, \dots, dx_n$, sa fonction caractéristique, fonction génératrice des moments est, si E désigne le symbole de la moyenne

$$\mathrm{E}e^{i(u_1x_1+u_2x_2+\dots+u_nx_n)} = \varphi(u_1, u_2, \dots, u_n).$$

(i). Si les variates sont stochastiquement indépendantes, la fonction φ se décompose en un produit de n fonctions $\varphi_j(u_j)$, $(j = 1, 2, \dots, n)$, et réciproquement si $\varphi_j(u_j)$ est une fonction caractéristique, un produit $\prod_j \varphi_j(u_j)$ est une fonction caractéristique d'une loi de probabilité où les variates sont indépendantes.

(ii). Si l'on annule $n-k$ variates, par exemple $u_{k+1}, u_{k+2}, \dots, u_n$, on obtient la fonction caractéristique de la loi de probabilité correspondant à la projection des masses sur l'hyperplan $Ox_1\,x_2, \dots, x_k$, parallèlement à l'hyperplan $Ox_{k+1}\,x_{k+2}, \dots, x_n$.

(iii). Si l'on égale toutes les variables, on obtient une fonction $\varphi(u, u, \dots, u)$, qui est la fonction caractéristique de la somme

$$z = x_1 + x_2 + \dots + x_n.$$

(iv). Si l'on a deux lois de probabilité, l'une dont les variates sont $x_1, x_2, \dots, x_n$ et de fonction caractéristique $\varphi(u_1, u_2, \dots, u_n)$, l'autre dont les variates sont $y_1, y_2, \dots, y_n$ et de fonction caractéristique $\psi(v_1, v_2, \dots, v_n)$, la fonction $\varphi(u_1, u_2, \dots, u_n)\psi(v_1, v_2, \dots, v_n)$ est la fonction caractéristique de la loi dont les $n+m$ variates sont $x_1, x_2, \dots, x_n, y_1, y_2, \dots, y_n$.

(v). Supposons $n = m$. On reconnaît immédiatement que $\varphi(u_1, u_2, \dots, u_n)\,\psi(u_1, u_2, \dots, u_n)$ est la fonction caractéristique de la loi de probabilité des variates $z_1 = x_1 + y_1$, $z_2 = x_2 + y_2, \dots$, $z_n = x_n + y_n$, et ceci se généralise pour N lois.

(vi). En particulier, si ces N lois sont les mêmes, on a le théorème suivant : si n variates $x_1^{(k)}, x_2^{(k)}, \dots, x_n^{(k)}$ sont stochastiquement dépendantes et si leur loi de probabilité est la même quand k varie de 1 à N, la fonction caractéristique des N variates

$$z_1 = \sum_1^{\mathrm{N}} x_1^{(k)}, \qquad z_2 = \sum_1^{\mathrm{N}} x_2^{(k)}, \qquad \dots, \qquad z_n = \sum_1^{\mathrm{N}} x_n^{(k)},$$

est la N^{ième} puissance de la fonction caractéristique des n variates initiales.

(vii). Prenons encore une loi à trois variates (x, y, z) dont la fonction caractéristique est $\varphi(u, v, w)$. On pose

$$\begin{aligned} X &= Ax + A'y + A''z, \\ Y &= Bx + B'y + B''z, \end{aligned}$$

et l'on demande de trouver la fonction caractéristique $\Phi(U, V)$ des variates X et Y.

On a

$$\begin{aligned} \Phi(U, V) &= Ee^{i(UX+VY)} = Ee^{i[U(Ax+A'y+A''z)+V(Bx+B'y+B''z)]} \\ &= Ee^{i[(AU+BV)x+(A'U+B'V)y+(A''U+B''V)z]}, \end{aligned}$$

donc

$$\Phi(U, V) = \varphi(AU + BV, A'U + B'V, A''U + B''V).$$

On obtiendra alors les théorèmes de Tchebychef en remarquant que le premier membre de l'équation de la quadrique type s'obtient en supposant la loi envisagée rapportée à un point moyen et en prenant le double, changé de signe des termes du second degré dans le développement en série de Maclaurin de la fonction caractéristique.

7. On se bornera à faire deux applications.

1° Supposons d'abord qu'on ait une loi à N variates stochastiquement dépendantes, que les moments du premier ordre m_k soient bornés en valeur absolue et que m_N ait une limite bornée en valeur absolue quand N augmente indéfiniment. Alors $\frac{1}{N}\sum_1^N m_k$ a également une limite qui sera la limite m_z de la moyenne de

$$z = \frac{x_1 + x_2 + \ldots + x_N}{N},$$

quand N augmente indéfiniment.

Supposons donc la loi à N variates rapportée à son point moyen et soit $\varphi(u_1, u_2, \ldots, u_N)$ sa fonction caractéristique. Alors la fonction caractéristique de z est $\varphi\left(\frac{u}{N}, \frac{u}{N}, \ldots, \frac{u}{N}\right)$. Ici, la quadrique type se réduit à deux points

$$\frac{u^2}{N^2}\sum_{ik}\mu_{ik} = 1.$$

D'ailleurs, la quadrique type de la loi donnée est

$$\sum_{ik} \mu_{ik} u_i u_k = 1.$$

Si $x_1 + x_2 + \ldots + x_N = \lambda$ est l'équation d'un plan tangent à cette quadrique, on a

$$\sum_{ik} \mu_{ik} = \lambda^2.$$

La distance δ de ce plan à l'origine est $\frac{\lambda}{\sqrt{N}}$. Donc

$$\delta^2 = \frac{1}{N} \sum_{ik} \mu_{ik}.$$

Si donc les demi-axes de la quadrique type restent toujours bornés par un nombre α, on aura $\delta < \alpha$, donc

$$\sum_{ik} \mu_{ik} < N\alpha^2,$$

et la déviation type de la loi de z étant

$$\sigma = \frac{1}{N} \sqrt{\sum_{ik} \mu_{ik}},$$

on aura

$$\sigma < \frac{\alpha}{\sqrt{N}}.$$

Donc, par application de l'inégalité fondamentale, la valeur de z tend stochastiquement vers la limite de la moyenne arithmétique des moyennes des variates quand N augmente indéfiniment.

Il suffirait même, pour obtenir ce résultat, d'être sûr que α est d'un ordre de grandeur inférieur à l'orde de $\sqrt{N}$.

2° Comme autre exemple, partons d'une loi à trois variates (x, y, z), de fonction caractéristique $\varphi(u, v, w)$ et formons les sommes

$$\xi_1 = \frac{x_1 + x_2 + \ldots + x_N}{N},$$

$$\zeta_2 = \frac{y_1 + y_2 + \ldots + y_N}{N},$$

$$\xi_3 = \frac{z_1 + z_2 + \ldots + z_N}{N}.$$

La loi de probabilité des variates $x_1, y_1, z_1, x_2, y_2, z_2, \ldots, x_N, y_N, z_N$ a pour fonction caractéristique

$$\varphi(u_1, v_1, w_1)\,\varphi(u_2, v_2, w_2)\ldots\varphi(u_N, v_N, w_N),$$

et, par suite, la loi de probabilité des variates ξ_1, ξ_2, ξ_3 a pour fonction caractéristique $\left[\varphi\left(\frac{u}{N}, \frac{v}{N}, \frac{w}{N}\right)\right]^N$.

Si, enfin, on pose

$$X_1 = a_1^1\,\xi_1 + a_1^2\,\xi_2 + a_1^3\,\xi_3,$$
$$X_2 = a_2^1\,\xi_1 + a_2^2\,\xi_2 + a_2^3\,\xi_3,$$

la fonction caractéristique des variates X_1, X_2 sera

$$\left[\varphi\left(\frac{a_1^1\,u_1 + a_2^1\,u_2}{N}, \frac{a_1^2\,u_1 + a_2^2\,u_2}{N}, \frac{a_1^3\,u_1 + a_2^3\,u_3}{N}\right)\right]^N.$$

Si donc (m_1, m_2, m_3) est le point moyen de la loi en x, y, z, on aura comme point moyen de la loi en X_1, X_2

$$X_1^0 = a_1^1\,m_1 + a_2^2\,m_2 + a_1^3\,m_3,$$
$$X_2^0 = a_2^1\,m_1 + a_2^2\,m_2 + a_2^3\,m_3.$$

Si la distribution x, y, z est rapportée à un point moyen, il en est de même de la distribution X_1, X_2. Supposons qu'il en soit ainsi, alors la conique type de la distribution X_1, X_3 est

$$\frac{1}{N}\sum_{ik}\mu_{ik}(a_1^k a_1 + u_2^k u_2)(a_1^i\,u_1 + a_2^i\,u_2) = 1 \qquad (i, k = 1, 2, 3),$$

dont le premier membre tend vers zéro quand N augmente indéfiniment. L'ellipse type tendant vers zéro, on voit que X_1, X_2 tendent stochastiquement vers X_1^0, X_2^0.

Plus généralement, si l'on fait N observations d'un phénomène dépendant de n variations stochastiquement dépendantes et qu'on en déduise la position d'un point (X_1, X_2) d'un plan, la position du point tend stochastiquement vers le point moyen quand le nombre N des observations augmente indéfiniment.