ANNALES DE L'I. H. P.

A. KOLMOGOROV

Sur les propriétés des fonctions de concentrations de M. P. Lévy

Annales de l'I. H. P., tome 16, nº 1 (1958), p. 27-34
<http://www.numdam.org/item?id=AIHP_1958__16_1_27_0>

© Gauthier-Villars, 1958, tous droits réservés.

L'accès aux archives de la revue « Annales de l'I. H. P. » 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 les propriétés des fonctions de concentrations de M. P. Lévy [1]

par

M. A. KOLMOGOROV.

Partout dans la suite nous désignons par

$$\xi = \xi_1 + \ldots + \xi_n$$

la somme des variables aléatoires indépendantes ξ_k, par

$$Q(l) = \sup_x P\{x \leqq \xi \leqq x + l\}$$

la fonction de concentration de la somme ξ et par $Q_k(l)$ la fonction de concentration du ξ_k. Le théorème 48 de la monographie bien connue de M. P. Lévy, *Théorie de l'addition des variables aléatoires* peut être formulé de manière suivante :

Quels que soient $1 \geqq \varepsilon > 0$ *et* $p > 0$ *on peut trouver deux constantes* $\delta > 0$ *et* N *telles que pour* $n \geqq N$ *l'inégalité*

$$Q_k(l) \leqq 1 - \varepsilon$$

implique l'inégalité

$$Q(\delta l \sqrt{n}) \leqq \beta. \tag{1}$$

Récemment j'ai été amené à la nécessité d'avoir une certaine généralisation de l'inégalité (1). J'ai pu me contenter du théorème suivant [1] :

---

(1) Les Publications de l'Institut de Statistique de l'Université de Paris ont consacré deux numéros à l'impression d'articles écrits en hommage à M. Paul Lévy. Le Mémoire suivant est arrivé trop tard pour leur être joint.

THÉORÈME. — *Il existe une constante* C *telle que* (²) *les inégalités*

$$L \geqq l, \qquad L^2 \geqq l^2 \log s$$

ou

$$s = \sum_{k=1} [1 - Q_k(l)]$$

impliquent l'inégalité

$$Q(L) \leqq \frac{cL}{l\sqrt{s}}. \tag{2}$$

Dans cette Note je me propose de donner la démonstration de ce théorème. Une autre *précision* de l'inégalité (1) a été publiée sans démonstration détaillée par W. Dœblin dans son article [2]. *L'endroit* correspondant dans [2] n'est pas formulé de façon bien nette et contient une faute d'impression (³).

Néanmoins, il paraît évident qu'en appliquant la terminologie des fonctions de concentration, le résultat de M. Dœblin peut être formulé de la manière suivante :

Quels que soient $\varepsilon > 0$ *et* $\beta > 0$ *il existe des constantes* $\delta > 0$ *et* N *telles que sous la condition*

$$\max l_k \leqq \frac{L}{N}, \qquad L = \sqrt{\sum_{k=1}^{n} l_k^2}$$

l'inégalité

$$Q_k(l_k) \leqq 1 - \varepsilon$$

implique l'inégalité

$$Q(\delta l) \leqq \beta. \tag{3}$$

Il serait intéressant de trouver de telles inégalités qui comprendraient (2) et (3) comme cas particuliers naturels. Je voudrais indiquer d'une manière plus générale que le développement ultérieur des méthodes élémentaires des raisonnements directs du calcul des probabilités développé en France d'une manière si brillante par P. Lévy et W. Dœblin continue de rester, comme il le paraît, actuel tout autant que le développement des méthodes classiques ou bien celles d'analyse

---

(²) Dans la suite on prend les logarithmes de base 2.

(³) Au lieu de $\sum_i l_i^2$ il est imprimé $\sum_i^{?} l$.

fonctionnelle. En tout cas, je ne réussis pas à démontrer les résultats communiqués dans ma Note [1] précitée, sans l'usage de ces méthodes élémentaires directes. Il est probable que les mathématiciens qui possèdent bien les propriétés fines des fonctions caractéristiques pourront tôt ou tard démontrer et même généraliser les théorèmes de [1] par des méthodes purement analytiques comme c'était déjà le cas pour les résultats démontrés d'abord par les méthodes directes du calcul des probabilités. Il semble cependant que nous restons toujours dans une période où la compétition de ces deux directions conduit aux résultats les plus féconds. Si M. P. Lévy qui possède également bien ces deux directions réalise ce parallèle dans ses propres œuvres, il faudrait souhaiter qu'après la mort prématurée de M. Dœblin, les probabilistes de la jeune génération, soit par complaisance pour les méthodes purement analytiques, soit grâce à l'engouement pour la puissance — d'ailleurs parfaitement justifiée — des méthodes liées aux distributions dans les espaces fonctionnels, n'oublient point les méthodes directes.

Utilisons les notations suivantes :

$$F(x) = P\{\xi < x\}, \qquad a = M\xi, \qquad \sigma^2 = D\xi.$$

On connaît le lemme suivant :

LEMME 1. — *Il existe une constante c_1 telle que l'inégalité*

$$|\xi_k - a_k| \leqq l,$$

où

$$a_k = M\xi_k \qquad (k = 1, 2, \ldots, n)$$

implique l'inégalité

$$|F(x) - G_{a,\sigma}(x)| \leqq c_1 \frac{l}{\sigma},$$

$G_{a,\sigma}$ étant la distribution normale pour les paramètres a, σ.

Le lemme suivant représente la généralisation du lemme 48.2 de la *Théorie de l'addition des variables aléatoires.*

LEMME 2. — *Il existe une constante c_2 telle que les relations*

$$P\{\xi_k = x_k\} = P(\xi_k = -x_k) = \frac{1}{2}, \qquad x_k \geqq l \qquad (k = 1, \ldots, n),$$

$$L \geqq l\sqrt{\log n} \qquad (n \geqq 2)$$

impliquent l'inégalité

$$Q(L) \leqq \frac{c_2 L}{l\sqrt{n}}.$$

Pour la démonstration il faut distinguer les trois cas suivants :

1° $L > l\sqrt{n}$;

2° $L \leqq l\sqrt{n}$, mais il existe $r \geqq 1$ tel que le nombre n_r des constantes x_k dans l'intervalle

$$(4) \qquad 4^{r-1} l \leqq x_k < 4^r l$$

satisfait aux inégalités

$$n_r \geqq \frac{n}{4^r}, \qquad n_r \geqq \frac{n}{4 \log n};$$

3° $L \leqq l\sqrt{n}$ et la quantité r qui jouirait des propriétés indiquées dans le cas 2° n'existe pas.

Dans le premier cas

$$Q(L) \leqq 1 < \frac{L}{l\sqrt{n}}.$$

Dans le deuxième cas il est naturel de considérer la somme ξ_* des ξ_k pour lesquelles (4) a lieu. Comme les termes de cette somme satisfont aux inégalités

$$|\xi_k| \leqq l_* = 4^r l, \qquad \sigma_* = \sqrt{D\xi_*} \geqq 4^{r-1} l \sqrt{n_r} \geqq \frac{l}{2}\sqrt{n},$$

pour la distribution correspondante $F_*(x)$ grâce au lemme 1, on a

$$(5_1) \qquad |F_*(x) - G_{a_*,\sigma_*}(x)| \leqq c_1 \frac{l_*}{\sigma_*} \leqq c_1 \frac{4}{\sqrt{n_r}} \leqq 2c_1 \frac{\sqrt{\log n}}{\sqrt{n}} \leqq c_1 \frac{8}{l\sqrt{n}}.$$

Puisque la fonction de concentration $Q_{a,\sigma}(L)$ (4) de distribution normale $G_{a,\sigma}(x)$ vérifie l'inégalité

$$Q_{a,\sigma}(L) \leqq \frac{L}{\sqrt{2\pi}\sigma},$$

nous avons définitivement dans le deuxième cas [$Q_*(L)$ étant la fonction de concentration pour ξ_*] :

$$(5_2) \qquad Q(4) \leqq Q_*(4) \leqq \frac{4}{\sqrt{2\pi}\sigma_*} + 2c_1 \frac{4}{l\sqrt{n}} \leqq \left(\sqrt{\frac{2}{\pi}} + 16 c_1\right) \frac{L}{l\sqrt{n}}.$$

Dans le troisième cas pour tout $r \geqq 1$ une des deux inégalités suivantes a lieu

$$n_r < \frac{n}{4^r}, \tag{6$_1$}$$

ou bien

$$n_r < \frac{n}{4 \log n}. \tag{6$_2$}$$

La somme n' des n_r vérifiant l'inégalité (6_1) est au plus

$$n \sum_{r=1}^{\infty} \frac{1}{4^r} = \frac{n}{3}.$$

Comme la somme de tous les n_r est égale à n, la somme n'' des n_r vérifiant l'inégalité (6_2) n'est pas inférieure à $\frac{n}{2}$. Il est clair que parmi les n_r il se trouve au moins $2 \log n$ nombres différents de zéro. Comme le nombre des r qui satisfont l'inégalité

$$4^r l < L \leqq l \sqrt{n}$$

est inférieur à $\log n$, il existe une quantité non inférieure à r, tel que

$$n_r > 0, \qquad 4^r l \geqq L.$$

En choisissant parmi ces r seulement les nombres pairs, ou seulement les impairs par ordre de croissance, nous obtiendrons la suite

$$r_1 < r_2 < \ldots < r_s, \qquad s \geqq \frac{\log n}{2},$$

pour laquelle

$$4^{r_1} l \geqq L, \qquad r_{+1} \geqq r + 2.$$

Choisissons les termes $\xi_{k_1}, \xi_{k_2}, \ldots, \xi_{k_s}$ de telle façon qu'on ait

$$4^{r_i - 1} l \leqq x_{k_i} < 4^{r_i} l$$

et étudions leur somme ξ'. Comme on a

$$\xi_{k_i} \leqq 4^{r_1} l = 4_1 \geqq 4,$$
$$4 L_{i-1} \leqq |\xi_{k_i}| \leqq 4_i = 4^{r_i} l \qquad (i = 2, 3, \ldots, s),$$

le raisonnement tout à fait élémentaire qui peut être réalisé par le lecteur montre que la somme ξ' peut se trouver dans un segment quelconque de longueur L pour une seule combinaison bien définie (si $n > 1$) des signes des termes ξ_{k_i}, c'est-à-dire avec une probabilité au plus égale à

$$\frac{1}{2^s} \leqq \frac{1}{\sqrt{n}}.$$

Par conséquent, dans le troisième cas, grâce à la condition $L \geqq l$, on a

$$(5_3) \qquad Q(4) \leqq Q'(4) = \frac{1}{\sqrt{n}} \leqq \frac{L}{l\sqrt{n}}.$$

En rapprochant les inégalités (5_1), (5_2), (5_3) obtenues dans les trois cas, on voit que le lemme 2 se trouve démontré.

Passons à la démonstration du théorème. La relation

$$(7) \qquad F_k(u_k(y)) \leqq y \leqq F_k(u_k(y)+0)$$

[$F_k(x)$ étant la fonction de distribution de l'aléatoire ξ_k, définit d'une façon unique les valeurs de la fonction $u_k(y)$ pour tous les y de l'intervalle $0 < y < 1$, sauf peut-être pour un ensemble dénombrable de points y (4)]. On peut la considérer comme une fonction inverse

$$u(y) = F^{-1}(y)$$

dont la définition est précisée d'une manière convenable pour les points $\underset{\sim}{y}$ qui correspondent aux points de discontinuité de la fonction $F(x)$. Si l'on suppose que les variables aléatoires η_k sont distribuées uniformément sur l'intervalle $(0, 1)$ et qu'elles sont mutuellement indépendantes, les quantités

$$\xi'_k = u_k(\eta_k)$$

seront mutuellement également indépendantes et elles seront soumises aux distributions $F_k(x)$. Il est évident qu'on peut supposer exprimées de la même façon par η_k les variables aléatoires ξ_k originellement données, sans que la généralité du théorème en soit diminuée.

On peut supposer que toutes les fonctions $Q_k(l)$ sont inférieures à 1, puisque, ayant exclu de notre raisonnement les termes ξ_k pour lesquels $Q_k(l) = 1$ et ayant obtenu pour la somme des autres l'inégalité (2), on peut de nouveau introduire les termes exclus, ce qui ne conduit pas à l'augmentation de $Q(L)$.

---

(4) L'unicité n'a pas lieu pour les points $\underset{\sim}{y}$ qui correspondent aux intervalles où $F(x)$ est constante, mais on voit facilement que la fonction $u(y)$ peut être définie sans contradiction en vérifiant les conditions (7) sur l'intervalle entier $(0, 1)$.

Posons

$$\begin{aligned} 4\varepsilon_k &= 1 - Q_k(l), \\ x'_k &= u_k(\varepsilon_k), \\ x''_k &= u_k(1 - \varepsilon_k). \end{aligned}$$

On voit facilement que

$$x''_k - x'_k > l.$$

Désignons par $k_1, k_2, \ldots, k_m$ les indices k pour lesquels

$$\eta_k < \varepsilon_k \quad \text{ou} \quad \eta_k > 1 - \varepsilon_k.$$

Fixons les nombres

$$Z_r = \begin{cases} \eta_{k_r} & \text{pour } \eta_{k_r} < \varepsilon_{k_r}, \\ 1 - \eta_{k_r} & \text{pour } \eta_{k_r} > 1 - \varepsilon_{k_r}. \end{cases}$$

Il est facile de voir que la distribution conditionnelle simultanée des variables aléatoires $\xi_{k_1}, \xi_{k_2}, \ldots, \xi_{k_m}$ sera telle que celles-ci resteront mutuellement indépendantes et que chacune d'elles aura la distribution

$$P\{\xi_{k_r} = a_r + x_r\} = P\{\xi_{k_2} = a_r - x_r\} = \frac{1}{2}$$

ou

$$\begin{aligned} x_r &= \frac{u_{k_r}(-Z_2) - u_{k_2}(Z_r)}{2}, \\ a_r &= \frac{u_{k_r}(-Z_r) + u_{k_r}(Z_r)}{2}. \end{aligned}$$

En appliquant (pour k_r et Z_r fixés) le lemme 2 aux quantités $\xi'_r = \xi_{u_r} - a_r$ pour lesquelles

$$|\xi'_r| \geq l' = \frac{l}{2}$$

nous obtiendrons, lorsque $L \geq l' \sqrt{\log m}$ l'inégalité

$$Q(L) \leq \frac{c_2 L}{c' \sqrt{m}} = \frac{2 c_2 4}{l \sqrt{m}}. \tag{8}$$

L'inégalité (8) est démontrée pour la fonction $Q(L)$ conditionnelle (quand m, k_r, a_r, Z_r sont fixés et $L \geq l' \sqrt{\log m}$). Mais il s'ensuit immédiatement l'inégalité

$$Q(L) \leq p + \frac{2 c_2 L}{l \sqrt{\frac{s}{4}}}$$

pour la fonction $Q(L)$ absolue, où

$$p = 1 - P\left\{\frac{s}{4} \leqq m \leqq s\right\}.$$

Il ne nous reste qu'à évaluer cette dernière probabilité. Comme on a

$$M\,m = \sum_k 2\,\varepsilon_k = \frac{s}{2}, \qquad D\,m = \sum_k 2\,\varepsilon_k(1 - 2\,\varepsilon_k) \leqq \frac{s}{2}$$

on aura, suivant l'inégalité de Tchebycheff, pour $k = \frac{s}{4}$,

$$1 - P\left\{\frac{s}{4} \leqq m \leqq s\right\} \leqq P\left\{\,|\,m - M\,m\,| \geqq k\right\} \leqq \frac{D\,m}{k^2} = \frac{8}{s}.$$

Ainsi nous aurons définitivement

$$Q(L) \leqq \frac{8}{s} + \frac{4\,c_2\,L}{l\sqrt{s}} \qquad \text{pour} \quad s \geqq 1$$

ou bien puisque toujours $Q(L) \leqq 1$ et que d'après nos suppositions $L \geqq l$,

$$Q(L) \leqq \frac{c\,L}{l\sqrt{s}}, \qquad \text{où} \quad c = 8 + 4\,c_2.$$

BIBLIOGRAPHIE.

[1] A. M. Kolmogorov, *Deux théorèmes limites uniformes pour des sommes des variables aléatoires indépendantes* (en russe) (**Теория вероямносмей и ее лрименения** t. 1, p. 426-436).

[2] W. Dœblin, *Sur les sommes d'un grand nombre des variables aléatoires indépendantes* (*Bull. Sc. Math.*, t. 63, 1939, p. 23-32 et 35-64.).