ANNALES DE L'INSTITUT FOURIER

ALFRED RÉNYI

Sur un théorème général de probabilité

Annales de l'institut Fourier, tome 1 (1949), p. 43-52

<http://www.numdam.org/item?id=AIF_1949__1__43_0>

SUR UN THÉORÈME GÉNÉRAL DE PROBABILITÉ

par Alfred RÉNYI (*) (Budapest).

Le but du présent article est de donner une généralisation d'un théorème que j'ai exposé dans un travail précédent (1). En même temps que ce théorème présente un intérêt en lui-même, c'est-à-dire comme un théorème du calcul des probabilités, et il a des applications à la statistique mathématique, sur lesquelles j'espère revenir, je ferai remarquer qu'il doit son origine à la théorie des nombres, notamment à mes recherches concernant l'hypothèse de Goldbach (2). En effet, ce théorème peut être considéré comme une généralisation du « grand crible » de M. Linnik (3) qui à son tour peut être déduit de ce théorème général. Cependant dans le présent article je n'insiste pas sur ces applications (4).

Soit donné un ensemble E, dont les éléments seront dénotés par ξ. Soit donné un corps F borélien des parties de E. Les parties de E appartenant à F seront notées A, B, ... et appelées des événements. Soit donnée une fonction absolument additive d'ensemble non-négative, définie pour toute partie A de E appartenant à F, qui sera notée P(A) et appelée la probabilité de l'événement A ; supposons finalement que $P(E) = 1$. Autrement dit nous envisageons un champ de probabilité au sens de M, Kolmogoroff (5).

(*) Conférence faite à l'Institut Fourier de l'Université de Grenoble, le 1er juillet 1949.

(1) A. Rényi, *Un nouveau théorème concernant les fonctions indépendantes etc.* (Journal de Math. XXVIII, Fasc. 2, 1949, p. 137-149).

(2) A. Rényi, *Sur la représentation des nombres pairs comme la somme d'un nombre premier et d'un nombre presque premier* (Bull. Acad. Sci. U. R. S. S. série math. 12, n° 1, 1948, p. 57-78).

(3) U. V. Linnik, *The large sieve* (C. R. Acad. Sci. U. R. S. S., 30, n° 4, 1941, p. 290-292).

(4) Paraîtra prochainement. Voir aussi *l. c.* (1).

(5) A. Kolmogoroff, *Grundbegriffe der Wahrscheinlichkeitsrechnung* (Ergebnisse d. Math. u. Grenzgebiete, 1933).

Introduisons d'abord quelques définitions bien connues. Une fonction à valeurs réelles $x = x(\xi)$ définie sur E sera appelée une variable aléatoire si elle est mesurable par rapport à P. La valeur probable $e(x)$ d'une variable aléatoire x est, par définition, l'intégrale de x par rapport à la mesure P et étendu sur E, c'est-à-dire

$$(1) \qquad e(x) = \int_E x \, d\mathrm{P}.$$

La valeur probable d'une variable aléatoire x relative à l'événement A ayant une probabilité $\mathrm{P(A)} > 0$, sera notée $e_{\mathrm{A}}(x)$ et définie par

$$(2) \qquad e_{\mathrm{A}}(x) = \frac{1}{\mathrm{P(A)}} \int_{\mathrm{A}} x \, d\mathrm{P}.$$

Notons $\sigma^2(x)$ le carré de l'écart moyen de x, défini par

$$(3) \qquad \sigma^2(x) = e\left[\left(x - e(x)\right)^2\right].$$

Soit pour t réel quelconque $\mathrm{A}^x(t)$ l'ensemble des éléments ξ de E pour lesquels $x(\xi) < t$, c'est-à-dire soit $\mathrm{A}^x(t)$ l'événement $x < t$, et posons $\mathrm{V}^x(t) = \mathrm{P}\left(\mathrm{A}^x(t)\right)$, en d'autres mots, soit $\mathrm{V}^x(t)$ la fonction de distribution de x. Soit $\mathrm{I} = [a, b]$ l'intervalle $a \leq t < b$ de l'axe réel R, $-\infty \leq a < b \leq +\infty$. Pour notre but il est plus commode d'utiliser au lieu de $\mathrm{V}^x(t)$ la fonction d'intervalle $\mathrm{V}^x[\mathrm{I}]$ définie pour $\mathrm{I} = [a, b]$ par

$$(4) \qquad \mathrm{V}^x[\mathrm{I}] = \mathrm{V}^x(b) - \mathrm{V}^x(a).$$

Quand nous parlons dans ce qui suit de la fonction de distribution de x, nous entendons toujours la fonction d'intervalle $\mathrm{V}^x[\mathrm{I}]$. Evidemment $\mathrm{V}^x[\mathrm{I}]$ est une fonction d'intervalle additive et non-négative, et nous avons $\mathrm{V}^x[\mathrm{R}] = 1$ où R désigne tout l'axe réel. Si $\mathrm{A}^x[\mathrm{I}]$ désigne l'ensemble des éléments ξ de E pour lesquels la valeur $x(\xi)$ est contenue dans l'intervalle I, c'est-à-dire si $\mathrm{A}^x[\mathrm{I}]$ désigne l'événement $x \in \mathrm{I}$, nous avons évidemment $\mathrm{P}(\mathrm{A}^x[\mathrm{I}]) = \mathrm{V}^x[\mathrm{I}]$. Nous appelons les variables x et y indépendantes si pour toutes les paires d'intervalles I_1, I_2, nous avons

$$(5) \qquad \mathrm{P}\left(\mathrm{A}^x[\mathrm{I}_1] \cdot \mathrm{A}^y[\mathrm{I}_2]\right) = \mathrm{P}\left(\mathrm{A}^x[\mathrm{I}_1]\right) \cdot \mathrm{P}\left(\mathrm{A}^y[\mathrm{I}_2]\right).$$

Il s'ensuit que si x et y sont indépendantes, alors pour tout I

$$(6) \qquad e_{\mathrm{A}^x[\mathrm{I}]}(y) = e(y).$$

Maintenant nous allons donner quelques définitions nouvelles. Dans un article antérieur nous avons introduit la notion de fonction de distribution $V_B^x[I]$ de x relative à l'événement B (où nous supposons $P(B) > 0$) définie par

$$V_B^x[I] = \frac{P(A^x[I] \,.\, B)}{P(B)} \tag{7}$$

et aussi une quantité mesurant l'écart entre $V_B^x[I]$ et $V^x[I]$. Cette dernière quantité était définie (à un facteur près) par la formule

$$D_B^2(x) = \frac{P(B)}{1 - P(B)} \int \frac{\left(V_B^x[I] - V^x[I]\right)^2}{V^x[I]}. \tag{8}$$

L'intégrale dans (8) est l'intégrale de Burkill ([6]) de la fonction d'intervalle qui figure sous le signe de l'intégrale, et est étendue à tout l'axe réel R. Dans ce qui suit nous appellerons la quantité $D_B(x)$ — supposée non-négative — la discrépance de x par rapport à l'événement B. La notion de discrépance — qui s'est montrée bien utile — peut être considérée comme une généralisation de la notion de coefficient de corrélation. En effet, si la variable x est la variable caractéristique d'un événement A (bien entendu $P(A) > 0$) c'est-à-dire si $x(\xi) = 1$ pour $\xi \epsilon A$ et $x(\xi) = 0$ dans le cas contraire, alors $D_B(x)$ devient égal à l'expression

$$\frac{|P(AB) - P(A) \,.\, P(B)|}{\sqrt{P(A)\left(1 - P(A)\right)P(B)\left(1 - P(B)\right)}}$$

c'est-à-dire sera égal au coefficient de corrélation des événements A et B. Dans ce qui suit nous aurons besoin d'étendre la notion de discrépance; nous allons envisager la discrépance d'une variable aléatoire x par rapport à une autre variable aléatoire y. Pour arriver à une généralisation naturelle de la quantité (8) pour ce cas, nous transformons la formule (8) comme suit : soit y_B la variable caractéristique de l'événement B, alors nous avons

$$V_B^x[I] = \frac{V^x[I] \,.\, e_{A^x[I]}(y_B)}{e(y_B)}. \tag{9}$$

Il s'ensuit que

$$D_B^2(x) = \frac{1}{\sigma^2(y_B)} \int_B \left[e_{A^x[I]}(y_B) - e(y_B)\right]^2 V^x[I] \tag{10}$$

([6]) Voir p. e. S. Kempisty, *Fonctions d'intervalle non additives* (Act. Sci. et Ind.).

en vertu de $\sigma^2(y_B) = P(B)(1 - P(B))$. Alors, en posant au lieu de y_B dans (10) une variable aléatoire y quelconque (nous supposons seulement que y n'est pas presque sûrement constante, c'est-à-dire que $\sigma^2(y) > 0$) autrement dit, en posant

$$(11) \qquad D_y^2(x) = \frac{1}{\sigma^2(y)} \int_R [e_{A^x[I]}(y) - e(y)]^2 V^x[I]$$

nous obtenons une définition de la discrépance $D_y(x)$ de x par rapport à y, qui est une généralisation directe de la formule (8) à laquelle elle se réduit pour $y = y_B$, c'est-à-dire si y est une variable caractéristique, et par conséquent $D_y(x)$ se réduit au coefficient de corrélation si non seulement y mais aussi x est une variable caractéristique.

Notons quelques propriétés de la discrépance $D_y(x)$ définie par la formule (11). Si x et y sont indépendantes, nous avons (en vertu de (6)) $D_y(x) = 0$. Si x et y sont presque sûrement égales (c'est-à-dire si la probabilité de $x = y$ est égale à 1), nous avons $D_y(x) = 1$. En effet, supposons que x ne prend qu'un nombre fini de valeurs distinctes $x_1 < x_2 < \cdots < x_n$ (autrement dit, supposons que x est une « variable en escalier ») et posons $P(x = x_i) = p_i$, alors nous avons

$$D_x^2(x) = \frac{1}{\sigma^2(x)} \sum_{i=1}^{n} p_i (x_i - e(x))^2 = \frac{\sigma^2(x)}{\sigma^2(x)} = 1$$

d'où vient que $D_x(x) = 1$ pour tout x (bien entendu $\sigma^2(x) > 0$) parce que chaque variable x peut être approchée en mesure par des variables en escalier. En général nous avons $0 \leqq D_y(x) \leqq 1$. Pour démontrer la seconde inégalité, notons d'abord que $D_y(x)$ reste invariant si nous remplaçons y par $y - c$ où c est un constant (où c est presque sûrement constant), parce que $e(y - c) = e(y) - c$ et aussi $e_A(y - c) = e_A(y) - c$ pour tout A. Alors si nous posons $Y = y - e(y)$, nous avons en vertu de $e(Y) = 0$,

$$(12) \qquad D_y^2(x) = D_Y^2(x) = \frac{1}{e(Y^2)} \int_R e^2_{A^x[I]}(Y) V^x[I].$$

Soit $z(x, I)$ la variable caractéristique d'évévement $A^x(I)$, alors nous avons

$$(13) \qquad e_{A^x[I]}(Y) = \frac{1}{V^x[I]} \int_E Y \cdot z(x, I)\, dP.$$

En appliquant l'inégalité bien connue de Schwarz, nous obtenons

$$e^2_{\mathrm{A}x[\mathrm{I}]}(\mathrm{Y}) \leqq \frac{1}{\left(\mathrm{V}^x[\mathrm{I}]\right)^2}\int_{\mathrm{E}} \mathrm{Y}^2 z(x, \mathrm{I})\, d\mathrm{P} \int_{\mathrm{E}} z(x, \mathrm{I})\, d\mathrm{P}.$$

Ainsi en utilisant les relations évidentes

$$\int_{\mathrm{E}} z(x, \mathrm{I})\, d\mathrm{P} = \mathrm{V}^x[\mathrm{I}] \qquad \text{et} \qquad \int_{\mathrm{R}} z(x, \mathrm{I}) = 1,$$

nous obtenons que

$$\mathrm{D}^2_{\mathrm{Y}}(x) \leqq \frac{1}{e(\mathrm{Y}^2)}\int_{\mathrm{R}}\int_{\mathrm{E}} \mathrm{Y}^2 . z(x, \mathrm{I})\, d\mathrm{P} = \frac{\int_{\mathrm{E}} \mathrm{Y}^2\, d\mathrm{P}}{e(\mathrm{Y}^2)} = \frac{e(\mathrm{Y}^2)}{e(\mathrm{Y}^2)} = 1,$$

ce qui démontre l'assertion $\mathrm{D}_y(x) \leqq 1$. Remarquons encore, que $\mathrm{D}_y(x)$ ne dépend pas des valeurs numériques de x mais seulement de la distribution de ces valeurs. Autrement dit si $f(t)$ est une fonction mesurable de la variable réelle t à valeurs réelles, de plus univalente, c'est-à-dire $f(t_1) \neq f(t_2)$ pour $t_1 \neq t_2$, et si nous posons $z(\xi) = f(x(\xi))$, nous avons $\mathrm{D}_y(z) = \mathrm{D}_y(x)$; en particulier nous pouvons remplacer x par $\mathrm{X} = x - e(x)$. Pour démontrer cette proposition, il suffit, eomme plus haut, de le démontrer pour x variable en escalier, qui prend seulement les valeurs x_i $(i = 1, 2, \ldots n)$. Posons $\mathrm{P}(x = x_i) = p_i$, alors nous avons

$$\mathrm{D}^2_y(x) = \frac{1}{\sigma^2(y)}\sum_{i=1}^{n} p_i\left[e_{(x=x_i)}(y) - e(y)\right]^2$$

et on voit immédiatement que $\mathrm{D}_y(x)$ reste invariant en remplaçant x par $z = f(x)$ parce que l'ensemble des éléments ξ de E pour lesquels $x = x_i$ est identique à l'ensemble des éléments ξ pour lesquels $z = z_i = f(x_i)$. Si nous laissons tomber la restriction sur l'univalence de $f(x) = z$, alors nous avons $\mathrm{D}_y(z) \leqq \mathrm{D}_y(x)$. Ceci découle du fait que $\mathrm{Q}(\mathrm{I}) = \left(e_{\mathrm{A}x[\mathrm{I}]}(y) - e(y)\right)^2 \mathrm{V}^x[\mathrm{I}]$ figurant sous le signe d'intégrale dans la définition de $\mathrm{D}_y(x)$ est une fonction sous additive d'intervalle, c'est-à-dire que si I_1 et I_2 sont des intervalles sans point commun, $\mathrm{Q}(\mathrm{I}_1 + \mathrm{I}_2) \leqq \mathrm{Q}(\mathrm{I}_1) + \mathrm{Q}(\mathrm{I}_2)$. En effet, en vertu de (13), Q(I) peut se mettre sous la forme $\frac{\left(\mathrm{Q}_1(\mathrm{I})\right)^2}{\mathrm{Q}_2(\mathrm{I})}$ ou $\mathrm{Q}_1(\mathrm{I})$ et $\mathrm{Q}_2(\mathrm{I})$ sont des fonctions additives d'intervalle, et la sous-additivité de Q(I) est une conséquence de l'inégalité élémentaire $\frac{(a+b)^2}{c+d} \leq \frac{a^2}{c} + \frac{b^2}{d}$ valable pour $c > 0$, $d > 0$, a, b réels.

Nous avons encore besoin d'une notion nouvelle, celle de suite de variables aléatoires « presque indépendantes deux-à-deux ». Soient x et y deux variables aléatoires et soit $d(x, y)$ la borne supérieure de

$$\left| \frac{P(A^x[I_1]A^y[I_2])}{P(A^x[I_1])P(A^y[I_2])} - 1 \right| \tag{14}$$

pour toutes les paires d'intervalles I_1 et I_2 pour lesquelles

$$P(A^x[I_1]) \cdot P(A^y[I_2]) \neq 0.$$

Nous appellerons la suite $\{x_n\}$ $(n = 1, 2, \ldots)$ une suite de variables presque indépendantes deux-à-deux si la forme quadratique

$$\sum_{n=1}^{\infty} \sum_{\substack{m=1 \\ n \neq m}}^{\infty} d(x_n, x_m) t_n t_m$$

est bornée, c'est-à-dire s'il existe une constante Δ telle que

$$\left| \sum_{n=1}^{\infty} \sum_{\substack{m=1 \\ n \neq m}}^{\infty} d(x_n, x_m) t_n t_m \right| \leq \Delta \sum_{n=1}^{\infty} t_n^2 \tag{15}$$

pour toute suite de nombres réels $t_n (n = 1, 2, \ldots)$ pour lesquels $\sum_{n=1}^{\infty} t_n^2 < \infty$. Dans ce cas nous appellerons Δ le module de dépendance de la suite $\{x_n\}$. Il est clair que si les variables x_n sont indépendantes deux-à-deux, alors $d(x_n, x_m) = 0$ pour tout $n \neq m$ et ainsi (15) aura lieu avec $\Delta = 0$.

Cela posé nous pouvons énoncer notre

THÉORÈME : *Soit $\{x_n\}$ $(n = 1, 2, \ldots)$ une suite de variables presque indépendantes deux-à-deux, soit Δ le module de dépendance de la suite $\{x_n\}$, et soit y une variable aléatoire quelconque. Dans ces conditions l'inégalité*

$$\sum_{n=1}^{\infty} D_y^2(x_n) \leq (1 + \Delta)\left(1 + \left(\frac{e(y)}{\sigma(y)}\right)^2\right) \tag{16}$$

aura lieu.

L'inégalité (16) est exacte pour les suites de variables indépendantes, car dans ce dernier cas la variable aléatoire y peut être choisie d'une telle manière que dans (16) l'égalité aura lieu. En effet posons $y = x_1 - e(x_1)$, alors $D_y(x_n) = 0$ pour $n = 2, 3, \ldots$, $D_y(x_1) = 1$ et $e(y) = 0$, et par conséquent (16) se réduit à $1 = 1$. Notons encore que dans l'article (1) nous avons démontré (16) dans un cas spécial, notamment pour y égale à une variable caractéristique d'un événement, et avec une définition plus restreinte de presque indépendance.

Pour la démonstration du théorème énoncé, nous avons besoin du lemme suivant, qui exprime le fait que l'inégalité connue de Bessel peut être généralisée pour les systèmes de fonctions quasi-orthogonales. Ce lemme — pour le cas des fonctions d'une variable réelle — est dû à M. Boas, Jr. (7).

Nous appelons une suite de variables aléatoires $\{\varphi_n\}$ $(n = 1, 2, \ldots)$ définies sur E un système quasi-orthogonal, si en posant $c_{nm} = \int_E \varphi_n \varphi_m dP$, la forme quadratique $\sum_1^\infty \sum_1^\infty c_{nm} t_n t_m$ est bornée, c'est-à-dire s'il existe un constant K tel que

$$\left| \sum_1^\infty \sum_1^\infty c_{nm} t_n t_m \right| \leqq K \sum_1^\infty t_n^2 \qquad \text{pour} \qquad \sum_1^\infty t_n^2 < \infty.$$

Nous appellerons K la borne du système quasi-orthogonal $\{\varphi_n\}$. Alors nous avons le suivant

Lemme : *Si* $\{\varphi_n\}$ *est un système quasi-orthogonal avec la borne* K, *et si pour une variable aléatoire* y *quelconque nous posons* $\int_E y \varphi_n dP = \gamma_n$, *alors nous avons* $\sum_1^\infty \gamma_n^2 \leqq K \,.\, e(y^2)$.

La démonstration de ce lemme est très simple. En effet pour tout N entier, nous avons

$$\int_E \left(y - \frac{1}{K} \sum_1^N \gamma_n \varphi_m \right)^2 dP = e(y^2) - \frac{2}{K} \sum_1^N \gamma_n^2 + \frac{1}{K^2} \sum_1^N \sum_1^N c_{nm} \gamma_n \gamma_m \geqq 0.$$

(7) R. P. Boas Jr., *A general moment problem* (Amer. Journ. Math., 1941, p. 361-370).

Mais par hypothèse $\sum_1^N\sum_1^N c_{nm}\gamma_n\gamma_m \leqq K\sum_1^N \gamma_n^2$, et ainsi on obtient

$$e(y^2) \geqq \frac{1}{K}\sum_1^N \gamma_n^2.$$

Faisons tendre N vers ∞, nous obtenons notre Lemme.

Démontrons maintenant notre théorème. Choisissons pour tout $n = 1, 2, \ldots$ une décomposition arbitraire de l'axe réel en un nombre fini d'intervalles I_{ni} $(i = 1, 2, \ldots N_n)$ sans point commun. Soit $A_i^{x_n}$ la partie de E telle que $x_n(\xi)\epsilon I_{ni}$ pour $\xi \epsilon A_i^{x_n}$. Posons $P(A_i^{x_n}) = P_{ni}$, évidemment il vient

$$\sum_{i=1}^{N_n} P_{ni} = 1 \tag{17}$$

parce que par hypothèse nous avons $I_{n_1} + I_{n_2} + \cdots I_{nN_n} = R$ où R désigne tout l'axe réel. Soit Φ_{ni} la variable caractéristique de $A_i^{x_n}$ et posons

$$\varphi_{ni} = \frac{\Phi_{ni} - P_{ni}}{\sqrt{P_{ni}}} \tag{18}$$

(bien entendu, sans restreindre la généralité, nous envisageons seulement les décompositions $\{I_{ni}\}$ pour lesquelles tous les P_{ni} sont positives). Soit

$$c_{nmij} = \int_E \varphi_{ni}\varphi_{mj}\,dP. \tag{19}$$

Nous avons pour $n \neq m$ l'inégalité

$$|c_{nmij}| = \sqrt{P_{ni}P_{mj}}\left|\frac{P(A_i^{x_n}A_j^{x_m})}{P(A_i^{x_n}A_j^{x_m})} - 1\right| \leqq \sqrt{P_{ni}P_{mj}}\,.\,d(x_n, x_m) \tag{20}$$

puis pour $i \neq j$

$$c_{nnij} = -\sqrt{P_{ni}P_{nj}} \tag{21}$$

et finalement

$$c_{nnii} = 1 - P_{ni}. \tag{22}$$

Utilisant ces formules nous allons montrer que la suite à double entrée $\{\varphi_{ni}\}$ ($n = 1, 2, \ldots; i \leqq N_n$) est un système quasi-orthogonal, ayant la borne $K = 1 + \Delta$, ou Δ est le module de dépendance de la suite presque indépendant $\{x_n\}$.

En effet, en posant tout de suite t_{ni} à double entrée

$$(n = 1, 2, \ldots; \quad i = 1, 2, \ldots N_n)$$

$$(23) \qquad S = \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} \sum_{i=1}^{N_n} \sum_{j=1}^{N_m} c_{nmij} t_{ni} t_{mj} \qquad \text{et} \qquad \theta_n = \sum_{i=1}^{N_n} t_{ni} \sqrt{P_{ni}}$$

nous obtenons, en vertu des relations (20), (21) et (22):

$$(24) \qquad |S| \leqq \sum_{\substack{n=1 \\ n \neq m}}^{\infty} \sum_{m=1}^{\infty} d(x_n, x_m) \theta_n \theta_m + \sum_{n=1}^{\infty} \left(\sum_{i=1}^{N_n} t_{ni}^2 - \theta_n^2 \right).$$

En appliquant l'inégalité de Schwarz, et en posant

$$(25) \qquad T_n = \sum_{i=1}^{N_n} t_{ni}^2,$$

puis tenant compte de (17), on obtient $\theta_n \leqq T_n^{\frac{1}{2}}$ et

$$(26) \qquad |S| \leqq \sum_{\substack{n=1 \\ n \neq m}}^{\infty} \sum_{m=1}^{\infty} d(x_n, x_m) T_n^{\frac{1}{2}} T_m^{\frac{1}{2}} + \sum_{n=1}^{\infty} T_n.$$

Ainsi, d'après la définition du module de dépendance Δ, il vient

$$(27) \qquad |S| \leqq (1 + \Delta) \sum_{n=1}^{\infty} T_n = (1 + \Delta) \sum_{n=1}^{\infty} \sum_{i=1}^{N_n} t_{ni}^2$$

ce qui démontre que le système $\{\varphi_{ni}\}$ est quasi-orthogonal, avec $K = 1 + \Delta$. Alors, notre lemme nous donne, si nous posons

$$(28) \qquad \gamma_{ni} = \int_E y \varphi_{ni} dP \quad (n = 1, 2, \ldots; \; i \leqq N_n),$$

l'inégalité

$$(29) \qquad \sum_{n=1}^{\infty} \sum_{i=1}^{N_n} \gamma_{ni}^2 \leqq (1 + \Delta) e(y^2).$$

En vertu de (18) un calcul simple donne

$$(30)\qquad \gamma_{ni}^2 = \left(e_{A_i^{x_n}}(y) - e(y)\right)^2 V^{x_n}[I_{ni}]\,;$$

alors faisant usage de (29) et (30), on obtient, en posant

$$(31)\qquad D_n = \sum_{i=1}^{N_n} \gamma_{ni}^2 = \sum_{i=1}^{N_n} \left(e_{A_i^{x_n}}(y) - e(y)\right)^2 V^{x_n}[I_{ni}]$$

que

$$(32)\qquad \sum_{n=1}^{\infty} D_n \leqq (1+\Delta)e(y^2).$$

D'autre part, pour une fonction sous-additive d'intervalle F(I) l'intégrale de Burkill $\int F(I)$ est égale à la borne supérieure de la somme $\Sigma F(I_k)$, pour toute décomposition $\{I_k\}$ de l'axe réel en un nombre fini d'intervalles sans points communs (voir (6)). Ainsi la borne supérieure de D_n est égale à $\sigma^2(y)D_y^2(x_n)$ et tenant compte du fait que les décompositions sont tout à fait arbitraires, nous obtenons

$$(33)\qquad \sigma^2(y)\sum_{n=1}^{\infty} D_y^2(x_n) \leqq (1+\Delta)e(y^2).$$

Comme évidemment $\dfrac{e(y^2)}{\sigma^2(y)} = 1 + \left(\dfrac{e(y)}{\sigma(y)}\right)^2$, (33) est exactement l'énoncé de notre théorème, qui est ainsi démontré.

(Manuscrit reçu en août 1949.)