il en résulte que $V_n^{(2)}$ est dense dans Φ. Donc $V_n^{(3)}$ est un ensemble ouvert dans Φ et dense dans Φ et l'ensemble

$$\Phi - \prod_{n=1}^{\infty} V_n^{(3)} \tag{25}$$

est de première catégorie. D'autre part, si l'on a (21) et si f^* satisfait à (19) resp. (20), alors f satisfait à (17) resp. (18). Donc $U_k^{(3)} \subset U_k^{(1)}$, $V_n^{(3)} \subset V_n^{(1)}$. Enfin, si $f \in \prod_{n=1}^{\infty} V_n^{(1)}$, alors on a pour tout x et pour une infinité d'entiers k l'une au moins des inégalités (17), (18), — l'intégrale (I) ne saurait donc être convergente pour aucun x. Il en résulte

$$\Phi - N \subset \Phi - \prod_{n=1}^{\infty} V_n^{(1)} \subset \Phi - \prod_{n=1}^{\infty} V_n^{(3)}, \tag{26}$$

c. à d. $\Phi - N$ est de première catégorie, c. q. f. d.

(Reçu par la Rédaction le 31. 3. 1931).

Sur les séries dont les termes sont des variables éventuelles indépendantes

par

M. PAUL LÉVY (Paris).

Introduction.

§ 1. Notations et remarques préliminaires.

Nous désignerons par x_1, $x_2, \ldots, x_n, \ldots$ des variables dépendant de lois de probabilité indépendantes les unes des autres; par S_n la somme $x_1 + x_2 + \ldots + x_n$; par $\mathcal{E}\{x\}$ la valeur probable de x, et par $\mathcal{D}\{x\}$ celle de $[x - \mathcal{E}\{x\}]^2$; par $\mathcal{P}\{E\}$ la probabilité d'un évènement E; par $\mathcal{P}\{E, E'\}$ celle de la réalisation de E et E'; par $\mathcal{R}\{E_n\}$ la probabilité que E_n soit réalisé pour une infinité de valeurs de l'entier n.

Rappelons que, si les probabilités $\mathcal{P}\{E_n\}$ sont indépendantes, $\mathcal{R}\{E_n\}$ est égal à 0 ou 1 suivant que la somme $\Sigma \mathcal{P}\{E_n\}$ est finie ou infinie, toute autre valeur étant exclue. Ce résultat est dû à M. Emile Borel.

La loi de probabilité dont dépend une variable x étant définie par la *fonction des probabilités totales*

$$F(\xi) = \mathcal{P}\{x < \xi\}, \tag{1}$$

il y a lieu, aux points de discontinuité de cette fonction, de distinguer les deux expressions

$$\begin{cases} \lim F(\xi - \varepsilon) = \mathcal{P}\{x < \xi\}, \\ \lim F(\xi + \varepsilon) = \mathcal{P}\{x \leqslant \xi\} \end{cases} \tag{2}$$

(ε désignant un infiniment petit positif). Toutefois, pour simplifier les formules, il nous arrivera d'employer la notation (1) pour désigner indifféremment l'une ou l'autre des expressions (2) ou n'importe quelle valeur intermédiaire. Il en résultera que certaines formules écrites en égalant deux valeurs de $F(\zeta)$ pourront être inexactes aux points de discontinuité de cette fonction; mais cela sera sans importance, les valeurs exactes de $\mathcal{P}\{x<\xi\}$ et $\mathcal{P}\{x\leqslant\xi\}$ pouvant toujours s'obtenir par le passage à la limite indiqué par les formules (2).

On sait que, si les lois de probabilité de deux variables indépendantes x et y sont définies par les fonctions des probabilités totales $F(x)$ et $G(y)$, celle dont dépend la somme $x+y=z$ est définie par la fonction

$$H(z)=\int_{-\infty}^{+\infty} G(z-x)\,dF(x)=\int_{-\infty}^{+\infty} F(z-y)\,dG(y)\,, \tag{3}$$

l'identité de ces deux expressions pouvant se vérifier par une intégration par parties. Toutefois ces expressions cessent d'avoir un sens pour les points de discontinuité de $H(z)$, car pour ces valeurs de z les fonctions $G(z-x)$ et $F(x)$ sont discontinues pour une même valeur de x; on pourrait, par des conventions convenables, préciser leur signification; mais cela est inutile; la loi dont dépend z est bien définie si l'on sait calculer $H(z)$ en tout point où cette fonction est continue.

§ 2. Notions générales et résumé du présent travail.

Bien que les sommes telles que S_n aient été étudiées depuis longtemps, les recherches effectuées à leur sujet ont d'abord eu pour objet l'étude de la loi de probabilité dont dépend une telle somme pour une valeur très grande de n considérée indépendamment des autres, et l'on a pendant longtemps négligé de considérer dans leur ensemble la suite des valeurs de S_n. Ce point de vue avait pu être considéré incidemment par certains auteurs, mais son étude systématique n'a été entreprise qu'à une date très récente; on peut trouver l'origine de ces recherches dans un Mémoire publié en 1909 par M. Emile Borel [1]), puis dans le travail de M. Cantelli [2]) qui a démontré en 1917 un théorème

[1]) Emile Borel, Sur les probabilités dénombrables et leurs applications arithmétiques, Rendiconti del circolo matematico di Palermo 26 (1909) p. 247—271.

[2]) F. P. Cantelli, Sulla probabilità come limite della frequenza, Rendiconti della r. Accademia dei Lincei V 26 (1917) p. 39—45.

qu'on a appelé depuis la loi forte des grands nombres. En 1925, MM. Khintchine et Kolmogoroff [3]) ont montré que la probabilité de la convergence de la série Σx_n ne pouvait être que 0 ou 1, et donné le moyen de reconnaître si l'on se trouve dans l'un ou l'autre de ces deux cas; en 1928, M. Kolmogoroff [4]) a donné une démonstration simplifiée de ce théorème fondamental.

Le Chapitre I du présent travail se rattache à ce théorème, dont nous donnons une démonstration qui, pour la seconde partie du théorème, est plus simple que celles antérieurement publiées. En outre, nous nous sommes efforcé de rendre les raisonnements plus intuitifs, en dégageant certaines notions qui jouent un grand rôle dans cette étude, notamment la notion de *dispersion* d'une loi de probabilité. Cette notion est essentielle si l'on veut, non seulement étudier la probabilité de la convergence de la série Σx_n, mais chercher s'il existe une suite de constantes $a_1, a_2, \ldots, a_n, \ldots$ telles que la probabilité de la convergence de la série $\Sigma(x_n-a_n)$ soit l'unité.

D'autre part nous montrons par un raisonnement *a priori* très simple et indépendant du critère donné par MM. Khintchine et Kolmogoroff que la probabilité de la convergence ne peut être que 0 ou 1. Des raisonnements de ce genre joueront à plusieurs reprises un grand rôle au cours de ce travail. Ajoutons que, mis au courant des principaux résultats obtenus par nous dans cet ordre d'idées, M. Kolmogoroff nous a fait savoir par une lettre du 28 Mars 1931 qu'il connaissait, sans l'avoir publié, le principe de ces raisonnements; mais certains des résultats que nous en avons déduits, au Chapitre II notamment, lui étaient inconnus.

Le Chapitre II a pour objet l'étude du cas de divergence. Nous rappelons et complétons sur certains points les résultats connus sur le rôle asymptotique de la loi de Gauss et des autres

[3]) A. Khintchine et A. Kolmogoroff, Über Konvergenz von Reihen, deren Glieder durch den Zufall bestimmt werden, Recueil Math. de Moscou 32 (1925) p. 668—677. Ce travail suppose que chacun des x_n prend au plus une infinité dénombrable de valeurs, restriction qui en réalité n'est pas nécessaire, et qui est supprimée dans celui de M. Kolmogoroff (loc. cit. [4]).

[4]) A. Kolmogoroff, Über die Summen durch den Zufall bestimmter unabhängiger Größen, Mathematische Annalen 99 (1928) p. 309—319.

lois stables. Nous montrons ensuite que, quelle que soit la suite (A) de constantes $A_1, A_2, \ldots, A_n, \ldots$ la probabilité $\mathcal{R}\{S_n > A_n\}$ ne peut être que 0 ou 1; nous avons déjà indiqué ce résultat dans le cas de Bernoulli [5]). Il en résulte que la suite des lois de probabilité dont dépendent les x_n définit dans l'ensemble des suites (A) une coupure analogue à celle définie par la convergence d'une série telle que Σe^{-A_n}. Des résultats de MM. Khintchine [6]) et Kolmogoroff [7]), que nous avons nous-même légèrement précisés dans le cas de Bernoulli [8]), définissent d'une manière assez approchée la place de cette coupure, dans les cas classiques; nous montrons que, dans le cas des lois stables autres que celles de Gauss, les formules des savants russes doivent être remplacées par des formules tout à fait différentes.

Dans le Chapitre III, nous étudions au contraire différentes questions se rattachant au cas de convergence. Nous montrons d'abord que, même s'il s'agit de semi-convergence, la loi de probabilité dont dépend la somme S est indépendante de l'ordre des termes, et nous donnons la condition pour que cette loi soit continue. Le cas de convergence présentant d'ailleurs un grand nombre de cas particuliers intéressants, nous en étudions quelques-uns.

Quelques-uns des résultats établis au cours de ce travail ont été énoncés dans une Note présentée à l'Académie des Sciences le 16 Mars 1931 [9]).

Chapitre I.

La probabilité de la convergence.

§ 3. Définitions.

Nous appellerons *oscillation* de S_ν entre n et N, et désignerons par $\sigma_{n,N}$ le maximum de $|S_{\nu'} - S_\nu|$ pour ν, $\nu' = n+1$, $n+1, n+2, \ldots, N$.

Nous poserons

$$\sigma_n = \lim_{N\to\infty} \sigma_{n,N}, \qquad \sigma = \lim_{n\to\infty} \sigma_n .$$

L'existence de ces limites est d'ailleurs certaine, $\sigma_{n,N}$ ne pouvant que croître avec N et décroître quand n croît. Le nombre σ définit *l'oscillation-limite* de la série Σx_n; $\sigma = 0$ caractérise les séries convergentes, et σ positif et borné caractérise les séries divergentes à sommes bornées.

Nous poserons

$$P_{n,N}(l) = \mathcal{P}\{\sigma_{n,N} < l\} .$$

On sait que, si une infinité dénombrable d'ensembles E_n $(n = 1, 2, \ldots)$ sont tels que chacun contienne le suivant, la mesure de leur partie commune est bien définie comme limite de la mesure de E_n; il en est de même, si chaque E_n est au contraire contenu dans le suivant, de la mesure de l'ensemble constitué par leur réunion. On en déduit que, sauf peut-être aux points de discontinuité des fonctions considérées,

$$\begin{aligned} P_n(l) &= \mathcal{P}\{\sigma_n < l\} = \lim_{N\to\infty} P_{n,N}(l), \\ P(l) &= \mathcal{P}\{\sigma < l\} = \lim_{n\to\infty} P_n(l), \\ P &= \mathcal{P}\{\sigma > 0\} = \lim_{l\to 0} P(l), \end{aligned} \tag{4}$$

la limite P ainsi définie est la probabilité de la convergence de la série Σx_n [10]). Remarquons que $P = 1$ implique $P(l) = 1$ pour tout l positif.

---

[5]) Paul Lévy, Sur un théorème de M. Khintchine, Bulletin des Sciences Mathématiques II 55 (1931) p....

[6]) A. Khintchine, Ein Satz der Wahrscheinlichkeitsrechnung, Fundamenta Mathematicae 6 (1924) p. 9—20.

[7]) A. Kolmogoroff, Gesetz des iterierten Logarithmus, Mathematische Annalen 101 (1929) p. 126—135.

[8]) Loc. cit. [1]).

[9]) Paul Lévy, Quelques théorèmes sur les probabilités dénombrables, Comptes rendus 192 (1931) p. 658.

[10]) M. Kolmogoroff, loc. cit. [4]), définit P par ce triple passage à la limite. Les considérations du texte, évidemment sous-entendues par ce savant, montrent que cette définition est parfaitement correcte et rend inutiles dans la plupart des cas les méthodes souvent employées, pour démontrer que $P = 1$, qui consistent à se donner deux suites de nombres ε_p et η_p tels que $\varepsilon = \Sigma \varepsilon_p$ et $\eta = \Sigma \eta_p$ soient arbitrairement petits, et déterminer ensuite les n_p de manière qu'entre n_p et n_{p+1} l'oscillation soit $< \varepsilon_p$, sauf dans des cas de probabilité $< \eta_p$. Toutefois nous avons dû utiliser encore une méthode de ce genre au § 9.

Remarquons que la définition de P s'applique même si les x_n ne sont pas ndépendants.

§ 4. La probabilité de la convergence.

Théorème I. *La probabilité P de la convergence de Σx_n, dans le cas d'indépendance des x_n, ne peut être que 0 ou 1.*

Par définition de P, η étant arbitrairement petit, on peut déterminer

$$\varepsilon, \varphi(l), \text{ et } \psi(l, n) > n,$$

de manière que

$$l < \varepsilon, \; n > \varphi(l), \; N > \psi(l, n)$$

entraînent

$$|P_{n,N}(l) - P| < \eta.$$

Prenons alors

$$l < \frac{\varepsilon}{2}, \quad \text{et } n, N, N'$$

tels que

$$n > \varphi(l), \quad n > \varphi(2l), \quad N > \psi(l, n),$$
$$N' > \psi(l, N), \quad N' > \varphi(2l, n).$$

Alors $P_{n,N}(l)$, $P_{N,N'}(l)$, $P_{n,N'}(l)$, $P_{n,N'}(2l)$ sont compris entre $P-\eta$ et $P+\eta$. D'ailleurs $\sigma_{n,N'}$ est compris entre $\sigma_{n,N}+\sigma_{N,N'}$ et le plus grand terme de cette somme, de sorte que

$$\mathfrak{P}\{\sigma_{n,N'} < l\} \leqslant \mathfrak{P}\{\sigma_{n,N} < l, \; \sigma_{N,N'} < l\} \leqslant \mathfrak{P}\{\sigma_{n,N'} < 2l\},$$

c'est-à-dire

$$P_{n,N'}(l) \leqslant P_{n,N}(l)\, P_{N,N'}(l) \leqslant P_{n,N'}(2l).$$

Ces quatre nombres différant très peu de P, on a $P = P^2$, c. q. f. d.

De ce théorème résulte immédiatement que:

Théorème II. *Ou bien il est possible de déterminer une suite de constantes a_n ($n = 1, 2, \ldots$) telles que la probabilité de la convergence de $\Sigma(x_n - a_n)$ soit l'unité, et dans ce cas a_n est déterminé au terme près d'une série convergente, ou bien cette probabilité est nulle quelles que soient les constantes a_n.*

§ 5. La dispersion des sommes S_n.

Nous appellerons *fonction de concentration* (ou de concentration *maxima*) d'une variable éventuelle x la fonction

$$\alpha = f(l) = \operatorname{Max} \mathfrak{P}\{x' < x < x''\} \qquad (x'' - x' = l)\ ^{11}),$$

[11]) La fonction $f(l)$ est une fonction non décroissante, variant de $f(0) \geqslant 0$ à $f(\infty) = 1$; mais, pour la fonction limite définie plus loin, on peut

et *fonction de dispersion*, ou plus simplement *dispersion*, la fonction inverse $l = \varphi(\alpha)$; elle définit donc la longuer du plus petit intervalle auquel corresponde une probabilité égale à α. D'après (3), la fonction de concentration de la somme de deux variables indépendantes est au plus égale à celle de chacune de ces variables. Par l'effet de l'addition des nouvelles variables, la concentration ne peut que diminuer; la dispersion ne peut qu'augmenter [12]).

Désignons alors par $f_{n,N}(l)$ la fonction de concentration relative à $S_N - S_n$. Elle ne peut que diminuer si N augmente, et au contraire augmenter avec n, il existe donc une fonction de concentration limite $f(l)$, bien définie par les

$$f_n(l) = \lim_{N\to\infty} f_{n,N}(l), \quad f(l) = \lim_{n\to\infty} f_n(l).$$

et une fonction de dispersion limite, $l = \varphi(\alpha)$, inverse de $f(l)$.

Lorsqu'on remplace la série Σx_n par $\Sigma(x_n - a_n)$, la fonction $f_{n,N}(l)$ ne change pas; la fonction $P_{n,N}(l)$ changeant au contraire, nous désignerons par $Q_{n,N}(l)$ son maximum lorsqu'on fait varier les constantes a_n.

Nous allons montrer que les ordres de grandeur des fonctions $f_{n,N}(l)$ et $Q_{n,N}(l)$ sont liés.

§ 6. Relation entre l'oscillation et la dispersion dans le cas d'un nombre fini de termes.

Nous raisonnerons sur la somme S_n, et écrirons, pour le présent § seulement, S au lieu de S_n, et σ, $P(l)$, $Q(l)$, $f(l)$ au lieu

avoir $f(\infty) < 1$. D'autre part, si un intervalle $(x, x + nl)$ correspond à la probabilité $f(nl)$, à l'un au moins des intervalles

$$(x, x+l), \; (x+l, x+2l), \ldots, \; [x+(n-1)l, \; x+nl],$$

correspond une probabilité au moins égale à $\frac{1}{n} f(nl)$. Donc

$$\frac{f(l)}{l} \geqslant \frac{f(nl)}{nl} \qquad (n = 2, 3, \ldots \infty).$$

[12]) Dans le cas de deux variables interdépendantes x et y, si, x étant donné, la fonction de concentration $f_x(l)$ relative à y a une borne supérieure $f(l)$, indépendante de x, cette borne s'applique aussi à la fonction de concentration relative à $x + y$.

de $\sigma_{o,n}, \ldots, f_{o,n}(l)$; nous désignerons par T le plus grand des $|S_\nu|$ $(\nu = 1, 2, \ldots, n)$; on a évidemment

(5) $$T \leqslant \sigma \leqslant 2T.$$

De la définition de $P(l)$ et $f(l)$ résulte que

(6) $$P(l) \leqslant \mathcal{P}\{|S_n| < l\} \leqslant f(2l),$$

et par suite

(7) $$Q(l) \leqslant f(2l).$$

D'autre part, par définition même de $f_{o,\nu}(2l)$, il est possible, en retranchant de $x_1, x_2, \ldots, x_n$ des constantes convenables, de faire en sorte que l'on ait

(8) $$\mathcal{P}\{|S_\nu| < l\} = f_{o,\nu}(2l) \geqslant f(2l) \quad (\nu = 1, 2, \ldots, n),$$

et par suite

(9) $$\mathcal{P}\{|S - S_\nu| < 2l\} \geqslant f^2(2l).$$

Considérons maintenant l'hypothèse $T \geqslant 3l$. Si sa probabilité était grande, on aurait peu de chances d'aboutir à une valeur finale de $|S|$ inférieure à l, ce qui serait en contradiction avec (8). Pour préciser cette remarque, observons que cette hypothèse peut être réalisée de n manières différentes, suivant la valeur de ν réalisant pour la première fois l'inégalité $S_\nu \geqslant 3l$, et, dans chacun de ces cas, la probabilité de $|S| < l$ est, d'après (9), au plus égale à $1 - f^2(2l)$. Il en est donc de même de la probabilité de $|S| < l$ lorsqu'on sait que $T \geqslant 3l$ sans savoir dans lequel de ces cas on se trouve, et par suite on a

$$\mathcal{P}\{T \geqslant 3l, |S| < l\} \leqslant \mathcal{P}\{T \geqslant 3l\}[1 - f^2(2l)].$$

D'autre part on a

$$\mathcal{P}\{T < 3l, |S| < l\} \leqslant \mathcal{P}\{T < 3l\} = 1 - \mathcal{P}\{T \geqslant 3l\},$$

et par suite, en ajoutant ces deux formules et tenant compte de (8),

(10) $$f(2l) = \mathcal{P}\{|S| < l\} \leqslant 1 - f^2(2l)\,\mathcal{P}\{T \geqslant 3l\}.$$

Or, d'après (5),

$$\mathcal{P}\{T \geqslant 3l\} \geqslant \mathcal{P}\{\sigma \geqslant 6l\} = 1 - P(6l) \geqslant 1 - Q(6l).$$

On a donc enfin

(11) $$1 - Q(6l) \leqslant \frac{1 - f(2l)}{f^2(2l)}.$$

Les formules (7) et (11) constituent les relations cherchées. Elles montrent que la connaissance de $\alpha = f(2l)$, si cette probabilité est assez voisine de l'unité, permet de limiter supérieurement $Q(l)$ et inférieurement $Q(6l)$. En termes peu précis, en appelant borne supérieure probable d'une quantité une borne qui a peu de chance d'être dépassée, on peut dire que l'oscillation et la dispersion ont des bornes supérieures probables du même ordre de grandeur, à condition que l'on ait préalablement retranché des x_n des constantes choisies de manière à rendre la première de ces deux bornes aussi petite que possible.

Le résultat précédent est à rapprocher d'un résultat remarquable obtenu par M. Kolmogoroff [13]. Supposons les $\mathcal{E}\{x_\nu\}$ nuls et les $\mathcal{E}\{x_\nu^2\}$ bornés, et posons

$$\mathcal{E}\{S^2\} = \Sigma\{x_\nu^2\} = M^2.$$

On connaît l'inégalité de Bienaymé-Tchebycheff

(12) $$\mathcal{P}\{S \geqslant cM\} \leqslant \frac{1}{c^2}.$$

M. Kolmogoroff a montré que l'on a de même

(13) $$\mathcal{P}\{T \geqslant cM\} \leqslant \frac{1}{c^2},$$

de sorte que la connaissance de M permet de limiter supérieurement à la fois les probabilités des grandes valeurs de S et de T (et par suite de l'oscillation $\sigma \leqslant 2T$). Cela n'empêche pas qu'il était nécessaire d'indiquer la relation précise qui existe entre $f(l)$ et $Q(l)$.

§ 7. La condition de convergence de MM. Khintchine et Kolmogoroff.

Revenons aux séries infinies; $f(l)$, $P(l)$ et $Q(l)$ désigneront de nouveau des fonctions limites; d'ailleurs les formules (6), (7) et (11) restent vraies à la limite, de sorte que $f(l) = 0$ pour tout $l > 0$ entraîne $Q(l) = 0$, et $f(l) = 1$ équivaut à $Q(l) = 1$.

Supposons d'abord $\mathcal{E}\{x_\nu\} = 0$ $(\nu = 1, 2, \ldots)$ et la série $\Sigma\mathcal{D}\{x_\nu\} = \Sigma\mathcal{E}\{x_\nu^2\}$ convergente. Etant donnés l et ε arbitrairement petits, on peut déterminer n de manière que

[13]) Loc. cit. [4]), Satz I. La méthode de M. Kolmogoroff est tout à fait analogue à celle qui nous a conduit à la formule (10).

$$\sum_{n+1}^{\infty} \mathcal{E}\{x_\nu^2\} < l^2 \varepsilon^2 ,$$

et par suite, d'après (5) et (13), pour tout $N > n$

$$\mathcal{P}\{\sigma_{n,N} \geqslant 2l\} \leqslant \varepsilon^2 .$$

On en déduit que $P(l) = 1$, et par suite $P = 1$.

Cette démonstration est celle de M. Kolmogoroff. On peut remarquer que l'on peut substituer l'emploi des inégalités (11) et (12) à celui de l'inégalité (13); l'inégalité (12) montre en effet que, sous les hypothèses indiquées, on a $f(l) = 1$, et par (11) on en conclut que $Q(l) = 1$, ce qui caractérise le cas de convergence prévu par le théorème II [14]). D'ailleurs le raisonnement de M. Kolmogoroff est un peu plus précis et montre, non seulement la convergence probable de $\Sigma(x_\nu - a_\nu)$ pour une détermination convenable des a_ν, mais celle de Σx_ν.

Naturellement, si l'on suprime la condition $\mathcal{E}\{x_\nu\} = 0$, la condition que $\Sigma \mathcal{D}\{x_\nu\}$ converge reste suffisante pour qu'on soit dans le cas de convergence prévu par le théorème II, et même, d'une manière plus précise, on peut prendre $a_n = \mathcal{E}\{x_n\}$. Mais cette condition n'est pas nécessaire. On le voit aisément en introduisant la notion de *suites équivalentes*, due à M. Khintchine.

Deux suites de variables x_ν et $\overline{x}_\nu$ sont équivalentes si $\mathcal{P}\{x_\nu = \overline{x}_\nu\}$ est le terme général d'une série convergente [15]). Il y a dans ce cas une probabilité égale à l'unité que les deux suites ne diffèrent que par un nombre fini de termes; on peut donc les substituer l'une à l'autre pour toutes les questions dont la solution n'est pas modifiée par le changement d'un nombre fini de termes, par exemple pour les questions de convergence.

[14]) Ce point sera précisé au début du § 9. Il est presque évident; mais il s'agit de justifier le passage à la limite, en montrant qu'on peut déterminer les a_n de manière que $P(l) = Q(l)$, ce qui est vrai par définition pour un nombre fini de termes, et une valeur particulière de n.

[15]) Cela implique (sauf dans le cas où la suite des x_n est équivalente à une suite de constantes) que x_n et $\overline{x}_n$ ne soient pas indépendants. Mais leur interdépendance est sans importance pour la suite, et il n'y aurait rien à changer si l'on remplaçait $\overline{x}_n$ par une variable y_n indépendante de x_n mais dépendant de la même loi de probabilité que $\overline{x}_n$.

Il suffit alors, pour que l'on soit dans le cas de convergence pour l'une des séries $\Sigma(x_\nu - a_\nu)$, que $\Sigma \mathcal{D}\{\overline{x}_\nu\}$ converge. Mais il faut remarquer qu'on ne pourra pas toujours prendre $a_\nu = \mathcal{E}\{x_\nu\}$, même si cette quantité est bornée; ainsi supposons que x_ν ait pour valeurs possibles 0 et ν^2, leurs probabilités respectives étant $1 - \frac{1}{\nu^2}$ et $\frac{1}{\nu^2}$, de sorte que $\mathcal{E}\{x_\nu\} = 1$. Il y a une probabilité égale à l'unité pour que la série Σx_ν converge (et même que tous ses termes soient nuls à partir d'un certain rang, et $\Sigma(x_\nu - 1)$ ne peut pas converger.

Ce qui précède établit, par la voie même suivie par M. Kolmogoroff, la première partie du théorème de MM. Khintchine et Kolmogoroff, qui s'énonce comme il suit:

Théorème III. *Pour qu'on soit dans le cas de convergence prévu dans le Théorème II, il faut et il suffit qu'il existe une suite équivalente à la suite des x_ν et telle que, pour les variables $\overline{x}_\nu$ de cette suite, $\Sigma \mathcal{D}\{\overline{x}_\nu\}$ converge.*

Nous savons déjà que cette condition est suffisante. Pour montrer qu'elle est nécessaire, supposons que la probabilité de la convergence de Σx_ν soit l'unité (les a_ν pouvant sans restriction être supposés nuls). Cela implique la convergence de

$$\Sigma \mathcal{P}\left\{ |x_\nu| > \frac{\varepsilon}{2} \right\},$$

ε étant arbitrairement petit (autrement il y aurait une probabilité égale à l'unité que x_ν ne tende pas vers zéro). On obtient donc une suite équivalente à celle des x_ν en posant $\overline{x}_\nu = x_\nu$ si $|x_\nu| \leqslant \frac{\varepsilon}{2}$ et $\overline{x}_\nu = 0$ dans le cas contraire. Nous poserons

$$y_\nu = \overline{x}_\nu - \mathcal{E}\{\overline{x}_\nu\},$$

de sorte que l'on a toujours $|y_\nu| \leqslant \varepsilon$. Il suffit, pour établir le résultat annoncé, d'établir la convergence de

$$\sum \mathcal{D}\{\overline{x}_\nu\} = \sum \mathcal{E}\{y_\nu^2\} ;$$

pour cela nous ferons l'hypothèse contraire et montrerons qu'elle aboutit à une contradiction.

D'après les hypothèses faites, l et η étant arbitrairement petits, et n arbitrairement grand, on peut choisir n et N de manière que, en posant

$$S' = \sum_{n+1}^{N} y_\nu, \quad A = -\sum_{n+1}^{N} \mathcal{E}\{\bar{x}_\nu\},$$

on ait

(14) $$\mathcal{E}\{S'^2\} = \sum_{n+1}^{N} \mathcal{E}\{y_\nu^2\} = M^2 > m^2,$$

(15) $$\mathcal{P}\{|S' - A| > l\} < \eta\ ^{16)}.$$

Considérons alors l'expression

$$M'^4 = \mathcal{E}\{S'^4\} = 3M^4 - 2\sum_{n+1}^{N} \mathcal{E}\{y_\nu^4\};$$

compte tenu de $|y_\nu| \leqslant \varepsilon$, on a

(16) $$\left|\frac{M'^4}{M^4} - 3\right| \leqslant \frac{2\varepsilon^2}{M^2}.$$

D'autre part, d'après (15)

$$M^2 \geqslant (1-\eta)(A-l)^2,$$

et par suite, en prenant pour fixer les idées $\eta = 10^{-3}$, $l = \frac{M}{5}$, on a $A < \frac{5}{4}M$. Dans la somme

$$\mathcal{E}\{(S'-A)^2\} \geqslant M^2$$

distinguons les termes provenant des valeurs de $|S' - A|$ qui sont $\leqslant l$, comprises entre l et $10\,M$, ou supérieures à $10\,M$. Les deux premières sommes partielles sont au plus égales à $l^2 = \frac{M^2}{25}$ et $100\,\eta\, M^2 = \frac{M^2}{10}$. La dernière dépasse donc $\frac{8}{10}M^2$. Or, pour ces termes, on a

$$|S'| \geqslant 8M, \quad |S'| \geqslant \frac{8}{10}|S'-A|$$

et par suite

$$|S'^4| > 40\,M^2\,|S'-A|^2,$$

et enfin

$$M'^4 = \mathcal{E}\{S'^4\} > 40\,M \cdot \frac{8}{10} M^2 = 32\,M^4,$$

ce qui est en contradiction avec la formule (16). Le théorème est donc complètement démontré.

§ 8. Remarques. 1° Nous avons même obtenu un résultat plus précis que le résultat énoncé. On est en effet sûr, dans le cas de convergence de Σx_ν, d'obtenir la suite des $\bar{x}_\nu$ en faisant abstraction des valeurs de x_ν dépassant un nombre $\frac{\varepsilon}{2}$ choisi d'une manière arbitraire (ε n'intervenant que par le rapport $\frac{\varepsilon}{M}$, on peut prendre par exemple $\frac{\varepsilon}{2} = 1$, pourvu que M soit grand).

On est donc conduit aux règles pratiques suivantes:

a) Pour que l'on soit dans le cas de convergence pour la série Σx_ν elle-même, il faut et il suffit que $\Sigma\mathcal{P}\{|x_\nu| > 1\}$ converge, ainsi que $\Sigma\mathcal{E}'\{x_\nu^2\}$, $\mathcal{E}'$ indiquant que dans le calcul de $\mathcal{E}$ on fait abstraction des valeurs de x_ν dont le module dépasse l'unité.

b) Pour que l'on soit dans le cas de convergence pour l'une des séries $\Sigma(x_\nu - a_\nu)$, il faut et il suffit, en désignant par a'_ν un nombre pour lequel la probabilité $\mathcal{P}\{|x_\nu - a'_\nu| > 1\}$ atteigne son minimum u_ν (ou en diffère peu), que u_ν soit le terme d'une série convergente, et qu'il en soit de même de $u'_\nu = \mathcal{D}'\{x_\nu\}$, $\mathcal{D}'$ indiquant que pour le calcul de $\mathcal{D}$ on remplace par a'_ν les valeurs de x_ν extérieures à l'intervalle $(a'_\nu - 1, a'_\nu + 1)$.

On fait ainsi à chaque loi de probabilité correspondre deux nombres déterminés u_ν et u'_ν, et, pour qu'on soit dans le cas de convergence pour une des séries $\Sigma(x_\nu - a_\nu)$, il faut et il suffit que Σu_ν et $\Sigma u'_\nu$ convergent.

2° Il est à peine besoin de faire remarquer que l'extension des théorèmes I à III aux séries complexes est immédiate. Considérons par exemple le cas où

$$x_\nu = c_\nu\, e^{i\theta_\nu},$$

[16]) On peut remarquer que, $\frac{\varepsilon}{M}$ étant très petit, $\frac{s'}{M}$ dépend d'une loi peu différente de celle de Gauss, ce qui est en contradiction avec l'hypothèse d'une probabilité supérieure à $1-\eta$ dans un intervalle de longueur très petite $2l$. Le théorème est donc démontré. Le raisonnement du texte a pour objet de donner une démonstration élémentaire indépendante du rôle que joue la loi de Gauss; naturellement, au lieu des moments d'ordres 2 et 4, on pourrait faire intervenir deux moments quelconques.

c_ν étant une constante, et θ_ν étant choisi au hasard suivant une loi telle que $\mathcal{E}\{e^{i\theta_\nu}\}=0$ [17]); la condition nécessaire et suffisante pour que la probabilité de la convergence de Σx_ν soit l'unité est que Σc_ν^2 converge. Si c_ν, au lieu d'être une constante, est une variable positive choisie au hasard indépendamment de θ_ν, la condition de convergence est que $\Sigma \mathcal{P}\{c_\nu > 1\}$ et $\Sigma \mathcal{E}\{\overline{c}_\nu^2\}$ convergent (en posant $\overline{c}_\nu = c_\nu$ si $c_\nu \leqslant 1$ et $\overline{c}_\nu = 0$ si $c_\nu > 1$).

§ 9. Théorème IV. *Dans le cas de convergence prévu dans le théorème II, $f(l)=1$; dans le cas de divergence, $f(l)=0$.*

Montrons d'abord que le cas de convergence est caractérisé par $f(l)=1$, ce qui, d'après (7) et (II), équivaut à $Q(l)=1$. Il est évident d'abord que dans le cas de convergence, on a $P(l)=1$ pour celle des séries $\Sigma(x_\nu - a_\nu)$ qui converge, et par suite $Q(l)=1$. Réciproquement, si $Q(l)=1$, on peut, étant donnée une suite de nombres ε_p et η_p tels que $\Sigma\varepsilon_p$ et $\Sigma\eta_p$ convergent, déterminer une suite d'entiers croissants n_p tels que

$$Q_{n_p,\, n_{p+1}}(\varepsilon_p) > 1-\eta_p,$$

et par suite, pour un choix convenable des a_ν,

$$P_{n_p,\, n_{p+1}}(\varepsilon_p) > 1-\eta_p,$$

d'où, pour $N > n > n_p$,

$$P_{n,N}(\varepsilon_p+\varepsilon_{p+1}+\ldots) > 1-(\eta_p+\eta_{p+1}+\ldots);$$

par suite, quelque petits que soient ε et η, on a $P_{n,N}(\varepsilon) > 1-\eta$ pour n assez grand, d'où enfin $P(l)=1$ et $P=1$ pour la série $\Sigma(x_\nu - a_\nu)$ considérée.

Pour achever la démonstration du théorème IV, il reste à montrer qu'il n'y a pas d'autre hypothèse possible que $f(l)=0$ et $f(l)=1$. A cet effet, nous allons écarter successivement trois hypothèses qui, compte tenu de ce que $f(l)$ est une fonction non décroissante comprise entre 0 et 1, épuisent toutes les autres

[17]) Dans le cas particulier où la loi dont dépend θ_ν est celle qui donne à chaque intervalle $d\theta$ (entre 0 et 2π) la probabilité $\frac{d\theta}{2\pi}$, ces séries ont été étudiées par M. H. Steinhaus (Sur la probabilité de la convergence de séries, Studia Mathematica 2 (1930) p. 21—40), puis par MM. Paley et Zygmund (On some series of functions, Cambridge Philosophical Society 26 (1930) p. 337—357 et 458—474).

possibilités. Le principe commun des trois parties de la démonstration est analogue à celui de la démonstration du théorème I: par définition de $f(l)$, on peut choisir trois nombres n, N, N' (ou même une succession de nombres n_p), tels que les fonctions de concentration relatives à $X=S_N-S_n$, à $Y=S_{N'}-S_N$, et à leur somme $S_{N'}-S_n$, différent peu de $f(l)$, et nous montrerons que cela est en contradiction avec le principe de la dispersion croissante indiqué au début du § 5.

Première hypothèse. Il existe un L positif et fini tel que $f(l)<1$ pour $l<L$ et $f(l)=1$ pour $l>L$

Cela signifie que, pour chacune des variables X et Y, les valeurs possibles (en négligeant des valeurs très peu probables) se répartissent dans un intervalle de longueur voisine de L, les valeurs voisines de chacune des extrémités ayant des probabilités qui ne sont pas négligeables. Alors les valeurs possibles pour $X+Y$ se répartiront dans les mêmes conditions dans un intervalle de longueur double, ce qui est en contradiction avec l'hypothèse que cette somme ait une fonction de concentration peu différente de $f(l)$.

Deuxième hypothèse. Pour l infini, $f(l)$ tend vers l'unité par valeurs plus petites que l'unité.

On peut alors choisir l tel que $f(l)=\beta>\frac{2}{3}$; posons $f(2l)=1-\gamma$. Les fonctions de concentration relatives à X, à Y, et à $Z=X+Y$ doivent avoir, pour l et $2l$, des valeurs très voisines de β et $1-\gamma$.

Considérons alors la formule (3) écrite pour les fonctions de probabilité de ces variables. Elle donne

$$H(z+l)-H(z)=\int_{-\infty}^{+\infty}[F(x+l)-F(x)]\,dG(z-x).$$

Désignons par ξ la valeur de x pour laquelle $F(x+l)-F(x)$ atteint son maximum β', par hypothèse peu différent de $\beta>\frac{2}{3}$; pour les valeurs de x extérieures à l'intervalle $(\xi-l,\xi+l)$, cette différence est forcément $<\frac{1}{3}<\frac{\beta'}{2}$. Or

$$\int_{z-\xi-l}^{z-\xi+l} d\,G(x) = 1-\gamma'' \leqslant 1-\gamma',$$

γ' étant peu différent de γ. On en déduit

$$H(z+l)-H(z) < \beta'(1-\gamma'') + \frac{1}{2}\beta'\gamma'' \leqslant \beta'\left(1-\frac{1}{2}\gamma'\right),$$

c'est-à-dire que la fonction de concentration relative à Z ne saurait, pour la valeur l considérée, dépasser un maximum peu différent de $\beta\left(1-\frac{1}{2}\gamma\right)$, donc inférieur à β. Cela est en contradiction avec les hypothèses faites.

T r o i s i è m e h y p o t h è s e. $f(\infty)=\alpha$ est compris entre 0 et 1.

Remarquons d'abord que si, pour chacune des variables X et Y, on a la probabilité α concentrée en un point, cette circonstance peut se reproduire pour leur somme Z; mais cela n'est possible que si $\alpha=\frac{1}{h}$, h étant entier, si X admet h valeurs possibles $X_1, X_2, \ldots, X_h$, de même probabilité α, et si enfin $X_1+Y_1=X_2+Y_2=\ldots=X_h+Y_h$. Dans ce cas, Z admet encore une valeur de probabilité α, mais une seule, et cette circonstance ne peut pas se reproduire si l'on ajoute une troisième variable Y' analogue à X et Y.

De même, si la loi de probabilité de chacune des variables X, Y, Y' comporte une probabilité α pour un intervalle de longueur très petite l, une probabilité plus grande ne pouvant se trouver que dans un intervalle de longueur beaucoup plus grande L, la loi dont dépend X, Y peut comporter encore une probabilité α pour un intervalle de longueur très petite ($2l$ et non plus l), mais cela n'est plus possible pour la somme $X+Y+Y'$. D'ailleurs ce qui importe n'est pas que l soit petit d'une manière absolue, mais par rapport à L.

Or, l'hypothèse $f(\infty)=\alpha$ implique que, ε étant arbitrairement petit, on puisse déterminer l, n', $\varphi(n)$ de manière que, pour $n>n'$ et $N>\varphi(n)$, on ait $f_{n,N}(l)>\alpha-\varepsilon$, une probabilité supérieure à $\alpha+\varepsilon$, pour la loi dont dépend $X=S_N-S_n$, ne pouvant, si N est assez grand, correspondre qu'à un intervalle de longueur L beaucoup plus grande que l. Il en sera de même, si N' et N'' sont assez grands, de $Y=S_{N'}-S_n$ et $Y'=S_{N''}-S_{n'}$.

Pour la somme $X+Y=S_{N'}-S_n$, on pourra avoir une probabilité voisine de α dans un intervalle de longueur finie (d'ailleurs $2l$ et non l), mais pour $X+Y+Y'=S_{N''}-S_n$, on ne pourra plus avoir une probabilité voisine de α que dans un intervalle de longueur beaucoup plus grande. Cela montre que dans ce cas encore la fonction $f(l)$ considérée ne peut être une fonction de concentration limite.

Le théorème IV est donc démontré. On voit que *en dehors du cas de convergence, l'addition des variables x_n produit une dispersion indéfiniment croissante, une probabilité non infiniment petite ne pouvant être attribuée à un intervalle de longueur finie.*

Chapitre II.

Le cas de divergence.

§ 10. L e r ô l e a s y m p t o t i q u e d e l a l o i d e G a u s s.

On sait que dans le cas de divergence, sous des conditions assez peu restrictives, S_n, divisé par un facteur convenable, dépend d'une loi qui diffère très peu de la loi de Gauss, si n est grand, et tend vers elle pour n infini. M. Lindeberg[18]) a établi en 1920, en supposant $\mathcal{E}\{x_\nu\}=0$ et

$$\sum_1^n \mathcal{E}\{x_\nu^2\} = M^2$$

fini, que si $|x_\nu|<\varepsilon M$, ε étant très petit, $\frac{S_n}{M}$ dépend d'une loi très peu différente de celle de Gauss. En 1922, complétant son premier résultat, il a établi[19]) que la conclusion subsiste pourvu que

$$\sum_1^n \mathcal{E}'\{x_\nu^2\} = M'^2 > (1-\eta)M^2, \tag{17}$$

$\mathcal{E}'\{x_\nu^2\}$ désignant le résultat obtenu en remplaçant par 0, dans le calcul de $\mathcal{E}\{x_\nu^2\}$, les valeurs de x_ν, supérieures en module à εM; ε et η sont deux nombres très petits.

[18]) Annales Academiae Scientiarum Fennicae A 16 (1920).
[19]) Mathematische Zeitschrift 15 (1922) p. 211—261.

Bien que par la méthode de M. Lindeberg son second théorème ne soit pas beaucoup plus difficile à établir que son premier, il n'est peut-être pas inutile d'observer que la notion de suite équivalente permet de passer sans difficulté du premier théorème au second.

Les infiniment petits ε et η étant indépendants, on peut supposer ε et ε' très petits, puis $\eta < \varepsilon^2 \varepsilon'$. La formule (17) implique que

$$\sum_1^n \mathcal{P}\{|x_\nu| > \varepsilon M\} < \varepsilon', \tag{18}$$

de sorte que la probabilité d'une erreur sur S_n est inférieure à ε' si l'on remplace chaque x_ν par une quantité $\bar{x}_\nu$ égale à x_ν si $|x_\nu| \leqslant \varepsilon M$ et à zéro dans le cas contraire; au point de vue d'un passage à la limite, la suite des $\bar{x}_\nu$ est équivalente à celle des x_ν. Le premier théorème de M. Lindeberg, appliqué à $\Sigma[\bar{x}_\nu - \mathcal{E}\{\bar{x}_\nu\}] = \Sigma y_\nu$ donne le résultat cherché.

On remarque d'ailleurs que la condition (17) est plus restrictive qu'il n'est nécessaire. Par contre la condition (18) ne l'est pas assez; pour l'application à Σy_ν du premier théorème de M. Lindeberg, il faut en effet que

$$M'^2 = \sum_1^n \mathcal{E}'\{x_\nu^2\},$$

quoique $< M^2$, reste très grand par rapport à $\varepsilon^2 M^2$ [20], et la condition (18) ne suffit pas pour cela. Mais en remplaçant (17) par la condition que $\frac{\varepsilon M}{M'}$ soit très petit, nous obtenons un théorème plus général que celui de M. Lindeberg. En changeant les notations, on peut l'exprimer comme il suit:

Théorème V. *Pour que S_n soit de la forme $A_n + B_n s_n$, A_n et B_n étant des constantes et s_n dépendant d'une loi tendant vers celle de Gauss, il suffit qu'il existe une suite équivalente à celle*

[20]) On remarque d'ailleurs que l'existence de M n'est pas nécessaire, εM intervenant seul dans la formule (18). L'hypothèse $\mathcal{E}\{x_n^2\}$ borné, essentielle pour les théorèmes de M. Lindeberg, n'intervient plus dans notre théorème V.

étudiée et à laquelle on puisse appliquer le premier théorème de M. Lindeberg. Il suffit donc qu'on puisse déterminer c de manière que $\Sigma \mathcal{P}\{|x_\nu| > c\}$ converge et que $\Sigma \mathcal{E}\{\bar{x}_\nu^2\}$ diverge, $\bar{x}_\nu$ étant égal à x_ν si $|x_\nu| \leqslant c$ et à zéro dans le cas contraire.

Cette condition n'est, bien entendu, pas nécessaire. En effet elle implique que l'ordre de grandeur probable de chaque terme soit très petit par rapport à celui de la somme, et cette condition n'est pas nécessaire pour un terme qui obéirait précisément à la loi de Gauss. Cette remarque conduit à généraliser le théorème V par le suivant, dans lequel L_n est un nombre qui indique l'ordre de grandeur de la dispersion de S_n, par exemple la dispersion correspondant à la probabilité $\frac{1}{2}$:

Théorème VI. *La conclusion du théorème V subsiste si*

$$x_n = a_n + b_n \xi_n + \eta_n + \eta'_n,$$

a_n et b_n étant des constantes, ξ_n dépendant de la loi de Gauss, $|\eta_1|, |\eta_2|, \ldots, |\eta_n|$ ayant des bornes supérieures dont la plus grande est infiniment petite par rapport à L_n, et la série $\Sigma \mathcal{P}\{\eta'_n \neq 0\}$ étant convergente. (Les quantités η_n et η_n peuvent dépendre de x_n, mais les x_n sont indépendants les uns des autres, de sorte qu'il en est de même des η_n; nous supposons bien entendu qu'on est dans le cas de divergence, de sorte que L_n augmente indéfiniment).

La démonstration ne présente aucune difficulté. Il n'est peut-être pas impossible d'obtenir dans cet ordre d'idées une condition nécessaire et suffisante; mais son application pratique serait très difficile. Remarquons que ces considérations conduisent, en faisant abstraction du terme η'_n, et supposant η_n indépendant de ξ_n, à se poser le problème suivant, lui-même difficile, mais dont la solution serait d'un assez grand intérêt: *étant donnée une variable x telle que $\mathcal{E}\{x\} = 0$, chercher le plus grand nombre a tel que x puisse se mettre sous la forme $a\xi + \eta$, ξ dépendant de la loi de Gauss réduite, et η étant indépendant de ξ.* En introduisant la *fonction caractéristique* $\varphi(t) = \mathcal{E}\{e^{itx}\}$, cela revient à chercher le plus grand nombre a tel que $\Phi(t) = \varphi(t)\, e^{\frac{a^2 t^2}{2}}$ soit une fonction caractéristique. On doit avoir

$$\Phi(t) \leqslant 1, \quad a^2 \leqslant \mathcal{E}\{x^2\}$$

(la seconde de ces inégalités résultant d'ailleurs de la première, écrite pour t très petit), ce qui donne des bornes supérieures de a; mais la valeur exacte du maximum reste difficile à trouver.

§ 11. Les lois stables en général.

On sait[21]) qu'il existe d'autres lois que la loi de Gauss jouissant de cette propriété que, si x et y sont deux variables indépendantes dépendant d'une telle loi, il en est de même, à un facteur constant près, de $ax+by$, a et b étant des constantes quelconques. Ces lois sont dites *stables*, et dépendant de deux paramètres, sans compter celui qui correspond à un changement d'unité. Chacune de ces lois L a un *domaine d'attraction*, composé de lois $\mathcal{L}$ jouissant de la propriété suivante: si $x_1, x_2, \ldots, x_n$ sont des variables indépendantes obéissant à la même loi $\mathcal{L}$, leur somme, divisée par une fonction convenable de n, dépend d'une loi tendant vers la loi L. Toutes les lois pour lesquelles $\mathcal{E}\{x\} = 0$, $\mathcal{E}\{x^2\} = 1$, appartiennent au domaine d'attraction de la loi de Gauss; en ne distinguant pas deux lois réductibles l'une à l'autre par un changement linéaire, toutes les lois pour lesquelles $\mathcal{E}\{x^2\}$ est borné, et même probablement toutes celles pour lesquelles $\mathcal{E}\{|x|^{2-\varepsilon}\}$ est borné quelque petit que soit ε, appartiennent au domaine d'attraction de la loi de Gauss. Rappelons enfin qu'un changement ne portant que sur les valeurs de x inférieures en module à une borne arbitrairement grande ne change pas le domaine d'attraction auquel se rattache une loi $\mathcal{L}$; par un changement ne portant que sur les valeurs de x très grandes en module et de probabilité arbitrairement petite, on peut donc, d'une loi $\mathcal{L}$ appartenant au domaine d'attraction d'une loi stable L, faire une loi $\mathcal{L}'$ appartenant au domaine d'attraction d'une autre loi L'.

Il en résulte évidement qu'on peut trouver une suite de variables $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n, \ldots$, équivalente à la suite des x_n, mais dépendant de lois $\mathcal{L}'_1, \mathcal{L}'_2, \ldots, \mathcal{L}'_n, \ldots$ du domaine d'attraction de L', tandis que les x_n dépendent de la loi $\mathcal{L}$, du domaine d'attraction de L. Alors la somme

$$\sum_1^n \bar{x}_\nu,$$

divisée par un facteur convenable, dépend d'une loi résultante tendant vers la loi L, bien que chacune des lois composantes appartienne au domaine d'attraction de L'.

Cette remarque montre que l'existence et la nature de la loi limite ne dépendent pas des propriétés individuelles des lois composantes, mais des propriétés de la famille de lois considérées, et complète les résultats que nous avons indiqués ailleurs au sujet des *familles normales de lois* L [22]).

[21]) Paul Lévy, Calcul des probabilités, p. 252 à 263.

§ 12. La borne supérieure des S_n.

Nous dirons qu'une suite de constantes A_n borne supérieurement les S_n pour n infini si l'on a, sauf peut-être un nombre fini de fois, $S_n > A_n$. Il s'agit d'étudier la probabilité $P = \mathcal{R}\{S_n > A_n\}$ de cette circonstance (notation indiquée au § 1); en désignant par $P_{n,N}$ la probabilité d'au moins une réalisation de $S_\nu > A_\nu$ pour les valeurs de ν comprises entre $n+1$ et N, et observant que cela revient au même de dire qu'un évènement E est réalisé pour une infinité de valeurs de ν ou qu'il l'est au moins une fois pour les valeurs de ν supérieures à n'importe quel entier n, on voit qu'on peut définir P par la formule

$$P = \lim_{n\to\infty} \lim_{N\to\infty} P_{n,N}$$

analogue à celles du § 3. Nous allons montrer le

Théorème VII. *Dans le cas de divergence, quelles que soient les constantes* A_n, *la probabilité* $P = \mathcal{R}\{S_n > A_n\}$ *ne peut pas avoir d'autre valeur que 0 ou 1* [23]).

[22]) Loc. cit. [21]).

[23]) Pour les inégalités indépendantes $x_n > A_n$, le même énoncé est une conséquence immédiate du principe de M. Borel rappelé au début du § 1.

Nous pouvons supposer chacun des x_n borné; dans le cas contraire, en effet, on peut remplacer la suite étudiée par une suite équivalente remplissant cette condition.

Nous supposerons $P > P' > 0$; il s'agit de montrer que, dans ces conditions, $P = 1$.

Par définition de P, ou peut déterminer n' et $\varphi(n)$ de manière que $n > n'$ et $N > \varphi(n)$ entraînent $P_{n,N} > P'$. Considérons alors une suite de nombres n_p tels que $n_0 > n'$ et $n_{p+1} > \varphi(n_p)$. Dans chacun des intervalles définis par ces nombres, la probabilité d'au moins une réalisation de $S_\nu > A_\nu$ est supérieure à P'. La conclusion $P = 1$ en résulterait si ces circonstances étaient indépendantes (d'après le principe de M. Borel rappelé au § 1).

Nous allons montrer qu'on peut déterminer $\psi(p)$ de manière que pour $q > \psi(p)$, le résultat des n_p premières expériences ne peut pas modifier de plus de ε la probabilité Ω d'une réalisation au moins de $S_\nu > A_\nu$ pour les valeurs de ν comprises entre $n_q + 1$ et n_{q+1}. En choisissant alors une suite de valeurs p_h telles que $p_{h+1} > \psi(p_h + 1)$, on aura, parmi les intervalles séparés par les nombres n_p, distingué une infinité d'intervalles tels que, pour chacun d'eux, indépendamment des autres, la probabilité d'une réalisation au moins de $S_\nu > A_\nu$ soit $\geqslant P - \varepsilon$, et la conclusion $P = 1$ sera établie.

Ecrivons n et N au lieu de n_p et n_q. Désignons par $F(x)$ la probabilité de $S_N < X$, et par $\omega(X)$ la probabilité dans l'hypothèse $S_N = X$, d'au moins une réalisation de $S_\nu > A_\nu$ pour les valeurs de ν comprises entre $N+1$ et n; c'est une fonction non-décroissante de X. La formule

$$\Omega = \int_{-\infty}^{+\infty} \omega(X)\, dF(X) = 1 - \int_{-\infty}^{+\infty} F(X)\, d\omega(X) \text{ [24]},$$

montre alors que l'erreur sur Ω ne peut pas dépasser l'erreur maxima sur $F(X)$.

[24]) Pour la légitimité de cette intégration par parties, voir la remarque finale du § 1. Il suffit de modifier très peu chaque A_n pour que l'intégration par parties soit légitime, et l'on voit aisément qu'une modification suffisamment petite ne risque pas de changer la probabilité étudiée.

Désignons par l la différence des valeurs extrêmes de S_n. $F(X)$ étant une probabilité appréciée après connaissance des résultats des n premières expériences, ses valeurs extrêmes correspondent aux valeurs extrêmes de S_n et la variation possible de $F(X)$, par suite celle de Ω, provenant d'une modification des résultats des n premières expériences, ne peuvent pas dépasser le maximum de la variation de $F(X)$, due à l'addition d'une constante l à S_n, c'est-à-dire (S_N étant la somme des termes indépendants S_n et $S_N - S_n$) la fonction de concentration $f_{n,N}(l)$ définie au § 5. D'après le théorème IV, appliqué à la série

$$\sum_{n+1}^{\infty} x_\nu ,$$

on a $f_n(l) = 0$, et $f_{n,N}(l) < \varepsilon$ pour N assez grand, c. q. f. d.

Ce résultat généralise celui que nous avons établi antérieurement pour le cas de Bernoulli, [25]), et prouve que la donnée des lois de probabilité dont dépendent les x_n définit une coupure dans l'ensemble des suites de constantes A_n. Nous avons établi dans le travail cité (pour le cas de Bernoulli, mais la généralisation ne présente aucune difficulté) deux propriétés de cette coupure, montrant son analogie avec celles résultant de la convergence ou de la divergence d'une série dont les termes dépendent des A_n.

§ 13. Résultats plus précis concernant la borne supérieure des S_n.

Dans le cas de Bernoulli, où x_n a les valeurs possibles $1-\alpha$ et $-\alpha$, leurs probabilités respectives étant α et $1-\alpha$, M. Khintchine [26]) a montré que

$$\mathcal{R}\{ S_n > c\sqrt{2\alpha(1-\alpha)\, n \log\log n}\}$$

a la valeur 0 si $c > 1$ et 1 si $c < 1$. Nous avons montré [27]) que cette probabilité est encore 1 si $c = 1$, et donné même un ré-

[25]) loc. cit. [5]).
[26]) loc. cit. [6]).
[27]) loc. cit. [5]).

sultat plus précis. On peut se demander si l'hypothèse suivante, tout à fait précise, ne serait pas exacte pour toutes les fonctions $\lambda(t)$ telles que $\lambda(kt) \sim \lambda(t)$ pour t infini: „La probabilité

$$\mathcal{R}\{S_n > \sqrt{2\,\alpha\,(1-\alpha)\,n\,\log\lambda(n)}\}$$

est 0 si la série $\Sigma \frac{1}{n\lambda(n)}$ converge et 1 si cette série diverge". Mais la démonstration d'un tel résultat, s'il est exact, semble exiger des méthodes très différentes de celles employées jusqu'ici.

M. Kolmogoroff [28]) a étendu le théorème de M. Khintchine à des cas beaucoup plus généraux; dans cette extension, $\alpha(1-\alpha)\,n$ est remplacé par

$$\sum_1^n \mathcal{D}\{x_\nu\}.$$

Sans chercher s'il est possible de généraliser encore, nous observerons que cette extension est liée au rôle de la loi de Gauss, et ne comprend même pas tous les cas où cette loi s'applique. Ainsi, si tous les x_n dépendent d'une même loi, telle que $\mathcal{E}\{x^2\}$ soit infini, que $\mathcal{E}\{\frac{x^2}{\log(1+x^2)}\}$ soit fini, et que $\mathcal{E}\{x\}=0$, S_n, divisé par un facteur convenable, dépend d'une loi tendant vers la loi de Gauss, et pourtant la formule de M. Kolmogoroff ne s'applique pas à ce cas. Mais, dans le cas des lois stables autres que celle de Gauss, la différence avec le cas classique est bien plus grande, comme nous allons le montrer.

Nous appellerons loi L_α la loi définie par la fonction caractéristique

$$\mathcal{E}\{e^{itx}\} = e^{-|t|^\alpha}, \qquad (0<\alpha<2).$$

Elle est symétrique, c'est-à-dire que les valeurs x et $-x$ sont également probables, et l'on a, pour l infini

$$\mathcal{P}\{|s_n|>X\} \sim \frac{C}{l^\alpha}, \tag{19}$$

C étant une constante positive. De plus elle est *stable* [29]); chacune des variables indépendantes x_n dépendant de cette loi, il en est de même de $n^{-\frac{1}{\alpha}}S_n$. On a donc, pour $\frac{X^\alpha}{n}$ infini,

$$\mathcal{P}\{|s_n|>X\} \sim \frac{Cn}{X^\alpha}. \tag{20}$$

Nous allons montrer que:

Théorème VIII. *Chacune des variables indépendantes x_n dépendant de la loi L_α, $\lambda(t)$ désignant une fonction telle que l'oscillation de $\log\lambda(t)$ entre t et $2t$ soit infiniment petite pour t infini, la probabilité d'une infinité de réalisations de*

$$|S_\nu| > [\nu\,\log\nu\,\lambda(\log\nu)]^{\frac{1}{\alpha}} \tag{21}$$

est 0 si $u_p = \frac{1}{p\lambda(p)}$ est le terme général d'une série convergente et 1 dans le cas contraire.

Supposons d'abord la série Σu_p convergente, et cherchons à limiter supérieurement les probabilités

$$v_p = \mathcal{P}\{|S_n| > \frac{1}{4}[n\log n\,\lambda(\log n)]^{\frac{1}{\alpha}}\},$$

$$v'_p = \mathcal{P}\{\underset{\nu=n+1}{\overset{2n}{\mathrm{Max}}}|S_\nu - S_n| > \frac{3}{4}[n\log n\,\lambda(\log n)]^{\frac{1}{\alpha}}\} \qquad (n=2_p).$$

D'après (20), et l'hypothèse faite sur $\lambda(t)$, on a

$$v_p \sim \frac{C'}{p\,\lambda(p)}, \qquad \left(C' = \frac{4^\alpha C}{\log 2}\right).$$

La limitation supérieure ainsi obtenue pour $|s_n|$ s'applique à tous les $|s_\nu - s_n|$ $(n<\nu\leqslant 2n)$. Le calcul qui nous a conduit de (8) à (10) donne alors

$$v'_p \leqslant \frac{v_p}{(1-v_p)^2},$$

de sorte que la convergence de Σu_p entraîne celle de $\Sigma(v_p+v'_p)$. Or $v_p+v'_p$ limite supérieurement la probabilité d'une réalisation de

[28]) loc. cit. [7]).
[29]) loc. cit. [21]).

$$|S_\nu| > [n \log n \lambda(\log n)]^{\frac{1}{\alpha}},$$

et a fortiori celle d'une réalisation de l'inégalité (21), pour $2_p < \nu \leqslant 2_{p+1}$. La convergence de $\Sigma(v_p + v'_p)$ prouve donc que la probabilité d'une infinité de réalisations de cette inégalité est nulle, ce qui démontre la première partie du théorème VIII.

Supposons maintenant Σu_p divergent; il en est de même de Σv_p, et cette conclusion subsiste si dans la définition de v_p on remplace S_n par $S_{2n} - S_n$ et $\frac{1}{4}$ par un coefficient k supérieur à $1 + 2^{\frac{1}{\alpha}}$. Or, pour $n = 2, 4, \ldots, 2^p, \ldots$, les valeurs de $S_{2n} - S_n$ sont indépendantes les unes des autres. Compte tenu de

$$\log 2n\lambda(\log 2n) \sim \log n \lambda \log n),$$

la divergence de Σv_p prouve donc qu'il y a une probabilité égale à 1 que l'on ait une infinité de fois

$$|S_{2n} - S_n| > [n \log n \lambda(\log n)]^{\frac{1}{\alpha}} + [2n \log 2n \lambda(\log 2n)]^{\frac{1}{\alpha}},$$

et par suite (21) pour l'une des valeurs n et $2n$ de ν, ce qui démontre la seconde partie du théorème.

Naturellement ce résultat s'applique séparément aux valeurs positives et aux valeurs négatives de S_n.

Il s'étend sans difficulté, en ce qui concerne les valeurs positives, à la loi définie par la formule

$$(22)\quad \log \mathcal{E}\{e^{itx}\} = -(\cos\frac{\pi}{2}\alpha - i \sin\frac{\pi}{2}\alpha)t^\alpha \quad (t > 0, 0 < \alpha < 1).$$

C'est une loi stable, comme la loi L_α, mais elle s'en distingue essentiellement parce que les valeurs positives de t sont seules possibles[30]. Pour ces valeurs, les formules (19) et (20) et le théorème VIII s'étendent sans difficulté.

De ce résultat, nous allons déduire le théorème suivant:

Théorème IX. *Si les variables indépendantes x_n dépendent de lois telles que, pour $\xi > X$, on ait*

$$(23)\quad \mathcal{P}\{|x_n| > \xi\} < \frac{C}{\xi^\alpha}, \qquad \mathcal{P}\{x_n > \xi\} > \frac{c}{\xi^\alpha} \qquad (0 < \alpha < 1),$$

[30]) loc. cit. [21]).

α, X, c, C étant indépendants de n, la conclusion du théorème VIII subsiste (et s'applique aussi aux valeurs positives de S_n mais non aux valeurs négatives).

Supposons d'abord les x_n sûrement positifs. On peut considérer x_n comme une fonction non décroissante d'une variable ξ_n dépendant de la loi définie par la formule (22), et écrire

$$a\xi_n - b < x_n < A\xi_n + B$$

a, b, A, B étant des constantes positives. On en déduit

$$a\sum_1^n \xi_\nu - nb < S_n < A\sum_1^n \xi_\nu + nB,$$

et le résultat obtenu pour la loi dont dépendent les ξ_n s'étend par cette double inégalité à celles dont dépendent les x_n. On remarque d'ailleurs que, pour obtenir la limite supérieure de S_n, la première des conditions (23) est seule utilisée; rappelons d'autre part que, si le résultat obtenu prouve la possibilité de grandes valeurs de $n^{-\frac{1}{\alpha}} S_n$, ces valeurs sont très peu probables.

Supprimons maintenant l'hypothèse $x_n > 0$. Le choix de x_n peut alors se décomposer en deux opérations, la première déterminant le signe de x_n (la probabilité β_n de $x_n > 0$ est évidemment supérieure à $\beta = c/X^\alpha$), la seconde déterminant ensuite la valeur exacte de x_n. Le résultat précédent s'applique à la somme S'_n des x_ν positifs [compte tenu de ce que le nombre des x_ν positifs, pour $\nu = 1, 2, \ldots, n$ n'est pas $o(n)$], et à la somme $-S''_n$ des x_ν négatifs, en ce qui concerne la borne supérieure de S''_n. D'ailleurs S'_n et S''_n, une fois le signe de chacun des x_n connu, sont indépendants, de sorte que les valeurs de S''_n grandes par rapport à $n^{\frac{1}{\alpha}}$, très peu probables, ont très peu de chances de coïncider avec les grandes valeurs de S'_n. Les résultats obtenus pour S'_n s'appliquent donc à S_n, c. q. f. d.

L'extension du théorème IX au cas où $1 \leqslant \alpha < 2$ est plus difficile. Il faut en tout cas, si $\alpha > 1$, ajouter la condition restrictive $\mathcal{E}\{x_n\} = 0$; mais, même avec cette condition, nous ne pouvons affirmer que la conclusion subsiste.

§ 14. Cas des séries complexes.

Nous désignerons par $R(x)$ la partie réelle de x. Nous allons montrer le

Theorème X. *Pour que l'on ait*

$$\mathcal{R}\{|S_n| > A_n\} = 0 \tag{24}$$

il faut que, quel que soit θ, on ait

$$\mathcal{R}\{R(S_n e^{-i\theta}) > A_n\} = 0,$$

et il suffit que, quel que soit θ, on ait

$$\mathcal{R}\{R(S_n e^{-i\theta}) > k A_n\} = 0$$

($k = \cos\varphi$ étant un nombre inférieur à 1).

La première partie du théorème est évidente. La seconde résulte de ce qu'on peut diviser le cercle trigonométrique en un nombre fini d'arcs de longueurs inférieures à 2φ. Si alors l'inégalité $|S_n| > A_n$ était réalisée une infinité de fois, l'un au moins de ces arcs contiendrait les arguments d'une infinité de S_n réalisant cette inégalité, et, θ désignant l'argument du milieu de cet arc, on aurait une infinité de fois

$$R(S_n e^{-i\theta}) \geqslant |S_n| \cos\varphi > k A_n.$$

Si cette circonstance a une probabilité nulle, il en est a fortiori de même de l'hypothèse d'une infinité de réalisations de $|S_n| > A_n$, c. q. f. d.

De ce théorème résulte évidemment le

Corollaire. *Si chacun des x_n dépend d'une loi de probabilité qui ne change pas par une rotation des axes, pour que (24) soit vrai, il faut que*

$$\mathcal{R}\{R(S_n) > A_n\} = 0,$$

et il suffit que, pour une valeur de k inferieure à 1, on ait

$$\mathcal{R}\{R(S_n) > k A_n\} = 0\text{ [31]}.$$

[31]) On remarque que les résultats obtenus depuis le début du § 14 ne supposent pas que l'on soit dans le cas de divergence probable de Σ_n; ils ne supposent même pas l'indépendance des x_n; mais, dans le cas d'interdépendance de ces variables, il faut pour le corollaire préciser que la loi dont dépend l'ensemble des variables $x_1, x_2, \ldots, x_n$ ne change pas par une rotation des axes.

Si en particulier on a $|x_n| \leqslant 1$, et que

$$M_n^2 = \sum_1^n \mathcal{E}\{|x_\nu|^2\}$$

augmente indéfiniment avec n, il résulte immédiatement des résultats de M. Kolmogoroff rappelés au début du § 13 que

$$\mathcal{R}\{|S_n| > c M_n \sqrt{\log\log M_n}\}$$

est 0 si $c > 1$ et 1 si $c < 1$.

Ce résultat s'applique en particulier aux séries

$$\sum e^{i\theta_n}, \quad \sum c_n e^{i\theta_n},$$

étudiées par M. Steinhaus, puis par M. M. Paley et Zygmund [32]); (θ_n est une variable choisie entre 0 et 2π, chaque intervalle $d\theta$ ayant une probabilité $\frac{d\theta}{2\pi}$; les c_n sont bornés et Σc_n^2 diverge). Il coïncide, pour la première de ces séries, avec un des théorèmes de ces derniers auteurs; mais notre méthode de démonstration est tout à fait différente.

Il est remarquable que, pour de telles séries, les grandes valeurs de $|S_n|$, à un facteur près infiniment voisin de l'unité, soient les mêmes que celles de $R(S_n)$; on aurait pu s'attendre à l'introduction d'un facteur $\sqrt{2}$. Mais les grandes valeurs de la partie réelle et de la partie imaginaire de S_n ne sont pas réalisées simultanément, et ce facteur n'intervient pas [33]).

Naturellement tous les résultats de ce § s'étendent sans difficulté à l'addition de vecteurs dans l'espace à p dimensions. Dans ce cas, M_n^2 étant défini comme ci-dessus, les grandes va-

[32]) loc. cit. [17]).

[33]) On peut aussi voir dans cette circonstance une application de la remarque que nous avons faite dans notre Mémoire „Sulla legge forte dei grandi numeri", Giornale dell' Istituto Italiano degli Attuari, Anno II (1931) p. 17, l. 8—13. Cette remarque concerne d'ailleurs les cas se rattachant à la loi de Gauss. Pour ceux qui se rattachent aux lois L_α ($\alpha < 2$), et permettent l'application du théorème IX, le corollaire du théoréme X se précise, par le fait que le facteur constant k ne change rien, et les inégalités $|S_n| > A_n$ et $R(S_n) > A_n$ ont la même probabilité d'être vérifiées une infinité de fois.

leurs de la longueur du vecteur représentant la somme S_n seront de l'ordre de grandeur de

$$M_n \sqrt{\frac{2}{p} \log\log M_n} \;.$$

Chapitre III.

Le cas de convergence[34]).

§ 15. Semi-convergence et convergence absolue.

Le théorème I s'appliquant à la série $\Sigma|x_n|$, la probabilité de la convergence absolue de Σx_n, comme celle de la convergence simple, ne peut être que 0 ou 1. Les remarques qui suivent visent surtout le cas où il y a une probabilité égale à 1 que Σx_n converge, sans converger absolument. Dans ce cas, on pourrait croire que la loi de probabilité dont dépend la somme S de la série, que l'on obtient comme limite de la loi dont dépend S_n, dépend de l'ordre des termes. Nous allons montrer qu'il n'en est rien:

Théorème XI. *La loi de probabilité dont dépend la somme S est indépendante de l'ordre des termes.*

Première démonstration. On sait qu'une loi de probabilité est bien définie par la donnée, pour t réel, de la fonction caractéristique

$$\varphi(t) = \mathcal{E}\{e^{itx}\}$$

et que la correspondance ainsi établie est satisfaisante au point de vue de la continuité, de sorte que, la loi dont dépend S étant limite de celle dont dépend S_n, on a

$$\mathcal{E}\{e^{its}\} = \prod_1^\infty \mathcal{E}\{e^{itx_n}\} \;.$$

Or, chacun des facteurs étant en module $\leqq 1$, le produit est indépendant de l'ordre des termes, c. q. f. d.

[34]) Dans tout ce Chapitre, nous supposons qu'on ait ajouté à chacun des x_n une constante convenable, de manière que ce soit la série Σx_n elle-même qui ait une probabilité de convergence égale à l'unité.

Deuxième démonstration. Nous pouvons sans restriction supposer $\mathcal{E}\{x_n\} = 0$ et la série $\Sigma\mathcal{E}\{x_n^2\}$ convergente; le cas général se ramène en effet à celui-là par application de la notion de suite équivalente. Désignons par S_n et S'_p deux sommes partielles de la série S et d'une série S' déduite de S par un changement de l'ordre des termes. Etant donné ε arbitrairement petit, choisissons n assez grand pour que

$$\sum_{n+1}^\infty \mathcal{E}\{x_\nu^2\} < \varepsilon^2,$$

et p assez grand pour que S'_p contienne tous les termes de S_n. On a alors

$$\mathcal{E}\{(S'_p - S_n)^2\} < \varepsilon^2,$$

et par suite, d'après l'inégalité de Bienaymé-Tchebycheff,

$$\mathcal{P}\{(S'_p - S_n)^2 > \varepsilon\} < \varepsilon.$$

A la limite, les lois de probabilité dont dépendent S et S' sont donc confondues, c. q. f. d.

Le raisonnement qui précède nous conduit d'ailleurs à un résultat plus précis que le premier raisonnement. Choisissons en effet au hasard un changement déterminé de l'ordre des termes, *la loi de probabilité dont dépend ce choix étant bien définie, mais absolument quelconque.* Indépendamment de ce choix, effectuons le choix successif des x_n. La probabilité que Σx_n change par le changement considéré de l'ordre des termes est nulle. Donc:

Théorème XII. *En exceptant des séries Σx_n de probabilité nulle, bien que la série Σx_n soit en général semi-convergente, il y a une probabilité égale à l'unité pour qu'un changement de l'ordre des termes choisi au hasard ne change pas la somme de la série.*

§ 16. Nature de la loi limite.

Nous dirons qu'une loi de probabilité est *continue* si la fonction des probabilités totales est continue, et, dans le cas contraire, qu'elle est *discontinue* ou *mixte*, suivant que la somme des variations de cette fonction aux points de discontinuité est $= 1$ ou < 1; cette somme sera appelée *discontinuité totale* de la loi considérée.

Théorème XIII. *S'il n'existe aucune suite de constantes équivalentes à la suite des x_n, la loi dont dépend S est continue. S'il existe une telle suite, cette loi est continue, discontinue, ou mixte, suivant que la plus petite des discontinuités totales des lois composantes est égale à* 0, *à* 1, *ou comprise entre* 0 *et* 1. (En d'autres termes, dans le second cas considéré, la règle est la même que pour l'addition d'un nombre fini de termes).

Pour démontrer la première partie du théorème, nous supposerons qu'une valeur S' de S ait une probabilité $\alpha > 0$, et nous montrerons qu'il existe une suite de constantes équivalentes à la suite des x_n.

A cet effet, ε étant un nombre positif arbitrairement petit, déterminons l de manière que

$$\mathcal{P}\{|S-S'|\leqslant 2l\} < \alpha+\varepsilon \tag{25}$$

puis n' de manière que pour $n > n'$ on ait

$$\mathcal{P}\{|S-S_n| > l\} < \varepsilon. \tag{26}$$

On a alors

$$\mathcal{P}\{|S-S_n|\leqslant l, |S_n-S'|\leqslant l\} \leqslant \mathcal{P}\{|S-S'|\leqslant 2l\} < \alpha+\varepsilon,$$

et par suite

$$\mathcal{P}\{|S_n-S'|\leqslant l\} < \frac{\alpha+\varepsilon}{1-\varepsilon}. \tag{27}$$

D'autre part la formule

$$\alpha = \mathcal{P}\{S=S'\} = \sum \mathcal{P}\{S_n = s\}\ \mathcal{P}\{S-S_n = S'-s\}$$

montre que, parmi les valeurs de s comprises entre $S'-l$ et $S'+l$, il en existe au moins une pour laquelle $\mathcal{P}\{S_n = s\} > \alpha-\varepsilon$ (autrement, d'après (26), le second membre serait au plus égal à $\varepsilon+(1-\varepsilon)(\alpha-\varepsilon) < \alpha$). Soit S'_n cette valeur; on a, compte tenu de (27),

$$|S'-S'_n|\leqslant l, \qquad \alpha-\varepsilon < \alpha_n = \mathcal{P}\{S_n = S'_n\} < \frac{\alpha+\varepsilon}{1-\varepsilon}. \tag{28}$$

Considérons alors la probabilité

$$\mathcal{P}\{S=S',\ S_n \neq S'_n\} = \alpha-\alpha_n \mathcal{P}\{S-S_n = S'-S'_n\} = \alpha-\alpha_n\beta_n;$$

en distinguant deux cas, suivant que $|S_n-S'|\leqslant l$ ou $> l$, et tenant compte de (26) et (27), on a

$$\alpha-\alpha_n\beta_n < \left(\frac{\alpha+\varepsilon}{1-\varepsilon}-\alpha\right)+\varepsilon = \varepsilon\left(1+\frac{1+\alpha}{1-\varepsilon}\right),$$

et, compte tenu de (28),

$$\beta_n = \mathcal{P}\{S-S_n = S'-S'_n\} > \frac{\alpha-2\varepsilon-2\varepsilon\alpha+\varepsilon^2}{\alpha+\varepsilon} > 1-3\varepsilon,$$

c'est-à-dire que, pour la loi de probabilité dont dépend $S-S_n$, et *a fortiori* pour toutes celles dont dépendent les sommes finies S_N-S_n $(N > n)$, la concentration maxima en un point est supérieure à $1-3\varepsilon$.

Donnons-nous alors une suite de nombres ε_p tels que $\Sigma\varepsilon_p$ converge, et déterminons les entiers n_p de manière que

$$\mathcal{P}\left\{\sum_{n_p+1}^{n_{p+1}} x_\nu \neq 0\right\} < 3\varepsilon_p.$$

D'après (11), les fonctions $Q_p(0)$ relatives à ces sommes partielles vérifient l'inégalité

$$1-Q_p(0) < \frac{3\varepsilon_p}{(1-3\varepsilon_p)^2},$$

c'est-à-dire que, pour un choix convenable des a_n, la probabilité d'une oscillation nulle de la somme $\Sigma(x_\nu-a_\nu)$ pour le groupe de termes $n_p < \nu \leqslant n_{p+1}$ est le terme général d'une série convergente. Il y a donc une probabilité égale à l'unité que l'on ait à partir d'un certain moment $x_n = a_n$; la suite des x_n est donc équivalente à celle des a_n.

Pour démontrer la seconde partie du théorème, nous pouvons supposer les a_n nuls, de sorte que la probabilité de $x_n \neq 0$ est le terme général d'une série convergente. Pour n assez grand, on aura donc

$$\mathcal{P}\{S \neq S_n\} < \varepsilon,$$

et par suite, quel que soit s,

$$(1-\varepsilon)\,\mathcal{P}\{S_n = s\} < \mathcal{P}\{S=s\} < \mathcal{P}\{S_n = s\}+\varepsilon \tag{29}$$

Ces formules montrent que tout point de probabilité positive pour S l'est pour S_n, pour n assez grand; que tout point de probabilité positive pour S_n l'est pour S; que si tous les x_ν, et par suite tous les S_n, dépendent de lois discontinues, la discontinuité totale de la loi dont dépend S est l'unité. Si tous les x_ν dépendent de lois discontinues ou mixtes, l'une au moins étant mixte, S_n, et par suite S, dépendent de lois mixtes (d'une part il y a au moins une valeur à probabilité positive; d'autre part la discontinuité totale de S, somme des quantités indépendantes x_ν et $S-x_\nu$, ne peut dépasser la plus grande des discontinuités des x_ν). Enfin, si un seul des x_ν dépend d'une loi continue, il est bien évident que toutes les sommes comprenant ce terme, dépendent de lois continues. Le théorème est ainsi démontré dans tous les cas.

§ 17. Comparaison des S_n avec une suite de constantes.

C'est le problème traité aux §§ 12 et 13 pour le cas de divergence. Pour le cas de convergence, sa solution est élémentaire. Nous distinguerons deux cas, suivant que la loi dont dépend S est continue ou non.

Premier cas. *La loi dont dépend S est continue.*

Soit alors A la limite inférieure d'indétermination des constantes données A_n. La probabilité de $S=A$ est nulle, et l'on a évidement

(31) $$\mathcal{R}\{S_n > A_n\} = \mathcal{P}\{S > A\}.$$

Deuxième cas. *La loi dont dépend S est discontinue.*

Nous pouvons dans ce cas, en retranchant des x_n des constantes convenables, supposer que $\Sigma\mathcal{P}\{x_n \neq 0\}$ converge.

Si A n'est pas un point de discontinuité pour la loi dont dépend S, la formule (31) subsiste. Si c'est un point de discontinuité, il faut distinguer les trois possibilités $S<A$, $S=A$, $S>A$. Si $F(X)$ désigne la fonction des probabilités totales de la loi dont dépend S, on peut désigner leurs probabilités respectives par

$$\lim F(A-\varepsilon) = F(A),\ \lim[F(A+\varepsilon)-F(A-\varepsilon)] = \alpha,$$

et

$$1-\alpha-F(\alpha).$$

Dans le premier cas, on a pour tout n assez grand $S_n < A_n$; le contraire a lieu dans le troisième. Dans le second, la probabilité d'une infinité de réalisations de $S_n < A_n$ est 1 s'il y a une infinité de constantes A_n inférieures à A et 0 dans le cas contraire. La probabilité $\mathcal{R}\{S_n > A_n\}$ est donc $F(A)+\alpha$ dans le premier de ces cas et $F(A)$ dans le second. Elle est ainsi connue dans tous les cas.

§ 18. Exemples divers.

Des exemples bien connus de séries absolument convergentes montrent bien que des lois continues peuvent résulter, à la limite, de la composition de lois continues, conformément au théorème XIII.

Soit d'abord x_n une variable égale à $0, 1, \ldots,$ ou 9, chacune de ces valeurs ayant la probabilité 1/10. La somme

$$S = \Sigma\frac{x_n}{10^n}$$

représente un nombre choisi au hasard entre 0 et 1, d'après le principe même de la numération décimale. Il en est de même, d'après le principe de la numération dyadique, de $\frac{S+1}{2}$, si

(32) $$S = \Sigma\frac{\pm 1}{2^n}$$

Considérons maintenant, plus généralement, la somme

(33) $$S = \Sigma\frac{\pm 1}{q^n} \qquad (q > 1).$$

Si $q>2$, il est bien connu qu'elle donne un exemple de loi continue pour laquelle les valeurs possibles forment un ensemble de mesure nulle. Le cas de $q=3$ a été considéré par M. H. Lebesgue. Dans le cas où $q=4$, S' et S'' désignant deux variables indépendantes dépendant de la loi obtenue, $2S'+S''$ représente une variable choisie au hasard entre -1 et $+1$, chaque valeur étant obtenue une fois et une seule (à l'exception d'une infinité dénombrable de valeurs). Si alors on fait correspondre respectivement aux variables S' et S'' les sommes s' et s'' de la série $\Sigma\frac{\pm 1}{2^n}$ obtenues avec les mêmes choix de signes, la relation entre le point s', s'' et la somme $2S'+S''$ fait correspondre le carré $|s'|<1, |s''|<1$ et le segment de droite $|2S'+S''|<1$.

Si au contraire $q<2$, chaque valeur de S de module inférieur à $\frac{1}{q-1}$ est obtenue d'une infinité de manières. Si par exemple $q=\sqrt{2}$, on peut écrire

$$\Sigma\frac{\pm 1}{(\sqrt{2})^n}=\sqrt{2}\,\Sigma\frac{\pm 1}{2_n}+\Sigma\frac{\pm 1}{2_n}\sim\sqrt{2}X'+X'',$$

X' et X'' étant indépendants l'un de l'autre et chacun choisi au hasard entre -1 et $+1$, et le signe $\sim$ indiquant que les deux membres dépendent de la même loi de probabilité.

Comme exemple de nature différente considérons la série

$$\Sigma\frac{\pm 1}{n^s}\sim\Sigma\left[\frac{1}{(2p+1)^s}\Sigma\frac{\pm 1}{2^{ks}}\right] \qquad (s>1/2),$$

chaque choix de signe dépendant bien entendu aussi bien de p que de h. Le cas où $s=1$ a été considéré dès 1924 par M. Norbert Wiener [35]). Dans ce cas on peut encore écrire

$$\Sigma\frac{\pm 1}{n}\sim\Sigma\frac{2x_p}{2p+1}, \tag{34}$$

chaque x_p étant choisi au hasard entre -1 et $+1$.

Toutes ces lois sont naturellement faciles à définir par leur fonctions caractéristiques $\mathcal{E}\{e^{itS}\}$. Dans le cas de la formule (32), en égalant les fonctions caractéristiques des deux membres, on retrouve la formule connue

$$\mathcal{E}\{e^{itS}\}=\frac{\sin t}{t}=\cos\frac{t}{2}\cos\frac{t}{4}\cdots\cos\frac{t}{2^n}\cdots,$$

et dans le cas de la formule (34), il vient

$$\mathcal{E}\{e^{itS}\}=\cos\frac{t}{1}\cos\frac{t}{2}\cdots\cos\frac{t}{n}\cdots$$

$$=\frac{\sin 2t}{2t}\,\frac{\sin\frac{2t}{3}}{\frac{2t}{3}}\cdots\frac{\sin\frac{2t}{2p+1}}{\frac{2t}{2p+1}}\cdots.$$

[35]) Publications of the Massachusetts Institute of Technology II 77 (1924).

Il peut y avoir aussi intérêt à considérer des produits de la forme

$$\Pi\left(1\pm\frac{1}{p^s}\right),\quad \Pi\left(1\pm\frac{1}{p^s}\pm\frac{1}{p^{2s}}\cdots\right),$$

les p étant les nombres premiers; chacun de ces produits, effectué, donne une somme de la forme

$$\Sigma\pm\frac{\mu(n)}{n^s},\quad \Sigma\frac{\pm 1}{n^s};$$

mais il y a lieu de remarquer que les signes de ces dernières sommes ne sont pas indépendants, de sorte qu'il ne semble guère possible de tirer parti de l'étude de ces produits pour la théorie des nombres premiers.

(Reçu par la Rédaction le 1. 5. 1931).