BULLETIN DE LA S. M. F.

P. LÉVY

Théorie des erreurs. La loi de Gauss et les lois exceptionnelles

Bulletin de la S. M. F., tome 52 (1924), p. 49-85

<http://www.numdam.org/item?id=BSMF_1924__52__49_1>

© Bulletin de la S. M. F., 1924, tous droits réservés.

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

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
http://www.numdam.org/

THÉORIE DES ERREURS.
LA LOI DE GAUSS ET LES LOIS EXCEPTIONNELLES;

PAR M. PAUL LÉVY.

PREMIÈRE PARTIE.

LA LOI DE GAUSS.

1. Les erreurs accidentelles, dans l'emploi des instruments de mesure, obéissent à la loi de Gauss, et l'on peut considérer comme bien connu aujourd'hui que l'explication de cette circonstance repose sur ce théorème fondamental : « *La somme d'un grand*

nombre d'erreurs indépendantes et très petites obéit à la loi de Gauss, ou se ramène par addition d'une constante à une erreur obéissant à la loi de Gauss. »

Pour savoir ce que vaut cette explication, il faut se demander si les restrictions contenues dans l'énoncé de ce théorème sont bien contenues dans la notion d'erreur accidentelle.

Si une cause d'erreur déterminée a un effet appréciable, que l'on puisse étudier et peut-être évaluer en fonction de certaines données, l'erreur qui en résulte est systématique. Une erreur accidentelle appréciable ne peut donc exister que par l'accumulation de plusieurs causes d'erreur, dont chacune est trop peu importante pour que l'on puisse distinguer son effet. Ces erreurs partielles sont nécessairement indépendantes, car on ne peut concevoir de dépendance entre elles que si elles sont dans une certaine mesure l'effet d'une même cause, et l'effet de cette cause devrait alors être rattaché à l'erreur accidentelle. Les restrictions contenues dans les mots « erreurs indépendantes et très petites » sont donc bien contenues dans la notion d'erreur accidentelle.

Poincaré semble avoir pensé qu'il fallait de plus que les erreurs partielles soient symétriques (1). Une pareille restriction, si elle avait réellement été nécessaire, aurait réduit à rien la valeur de la justification cherchée, dont le principe est qu'on peut arriver à une conclusion sur l'erreur totale sans rien admettre sur les erreurs partielles qui ne paraisse être effectivement vérifié en pratique; on ne peut évidemment admettre d'hypothèse aussi précise que celle de la symétrie de chaque erreur partielle. Il est surprenant que Poincaré n'ait as cherché à s'affranchir d'une hypothèse aussi inadmissible. Il y serait aisément parvenu. Quelle que soit en effet la méthode par laquelle on cherche à démontrer le théorème fondamental, on constate aisément qu'on peut remplacer l'hypothèse de la symétrie par celle que les différentes erreurs partielles aient leurs valeurs probables nulles, et la suppression de cette dernière hypothèse conduit seulement à ce résultat que l'erreur totale, si elle n'obéit pas à la loi de Gauss, diffère par une constante d'une erreur obéissant à la loi de Gauss.

D'ailleurs cette constante ne saurait être rattachée à l'erreur

(1) Voir *La Science et l'Hypothèse*, p. 241.

accidentelle. Si, en effet, elle est par exemple positive, la valeur probable de l'erreur est positive, et il ne faut pas perdre de vue que ce résultat, s'il est exact, est un résultat valable *a priori*, que l'on peut affirmer sans savoir quelle sera effectivement la valeur de l'erreur dans tel ou tel cas particulier. Or si l'on peut affirmer *a priori* l'exactitude de ce résultat, c'est évidemment qu'il a une cause, et par là même la constante que nous considérons se rattache à l'erreur systématique.

Un exemple le fera bien comprendre. Dans les observations astronomiques, la réfraction du rayon lumineux dans l'atmosphère est une cause d'erreur, et la déviation totale apparaît comme la somme des déviations dans les couches successives de l'atmosphère; l'erreur apparaît ainsi comme une somme d'erreurs partielles qu'on ne peut étudier individuellement, faute de données suffisantes sur l'état des différentes couches de l'atmosphère. Mais il y a une raison pour que ces erreurs partielles n'aient pas leurs valeurs moyennes nulles; en raison de la densité croissante des diverses couches de l'atmosphère que traverse le rayon lumineux, celui-ci se rapproche de la verticale. Ce n'est qu'après avoir mis de côté la déviation qui en résulte, que l'erreur signalée peut être considérée comme accidentelle, et comme l'on mesurera la déviation systématique en faisant la moyenne des déviations observées dans un grand nombre de cas, l'erreur accidentelle aura nécessairement sa valeur moyenne nulle.

Il reste une dernière restriction, contenue dans l'énoncé de notre théorème fondamental. Nous avons parlé de la *somme* des erreurs partielles. Mais il n'est pas évident que les différentes erreurs s'ajoutent simplement. Si x est la quantité à mesurer, peut-être faut-il ajouter les erreurs sur $\log x$, ou sur $\frac{1}{x}$; d'autres formules plus compliquées sont possibles. Il est évident ici que nous nous trouvons en présence d'une restriction qui n'est pas contenue dans la notion d'erreur accidentelle, car autrement, en raisonnant successivement sur x et sur $\log x$, on arriverait au résultat que les erreurs sur x et sur $\log x$ obéissent toutes deux à la loi de Gauss, ce qui est manifestement absurde.

Heureusement la difficulté disparaît si l'erreur totale (et non

seulement chaque erreur partielle) est assez petite pour qu'on puisse négliger son carré. Dans ces conditions toutes les formules possibles au sujet de la composition des erreurs partielles reviennent au même, et le résultat obtenu, en admettant que ces erreurs s'ajoutent, est exact.

Ainsi, rien qu'en précisant ce qui est incontestablement contenu dans la notion d'erreur accidentelle, nous voyons que le théorème fondamental indiqué au début conduit à cette conclusion : « *Dans toute mesure suffisamment précise, l'erreur accidentelle obéit à la loi de Gauss.* »

2. Ce qui précède s'adresse à quiconque a l'habitude de raisonner avec un peu de bon sens sur les questions scientifiques. C'est maintenant l'affaire du mathématicien de démontrer le théorème fondamental, et d'abord de préciser ce qui dans son énoncé n'est pas encore suffisamment précis : « Quand dirons-nous qu'une erreur est très petite ? »

Nous supposerons dans la suite (1) que les erreurs partielles aient leur valeur probable nulle; d'après ce qui précède, cette restriction ne diminue pas la portée du résultat. Il est alors naturel de prendre la valeur quadratique moyenne d'une erreur pour évaluer son ordre de grandeur. Mais le théorème qu'on serait ainsi conduit à énoncer est inexact.

M. Lindeberg (2) a eu alors l'idée de faire l'hypothèse plus restrictive qu'aucune des erreurs partielles ne dépassait en aucun cas une valeur déterminée très petite (par rapport à la valeur quadratique moyenne de l'erreur totale, que l'on peut supposer prise comme unité). En partant de cette hypothèse, il démontre que la loi de probabilité de l'erreur totale diffère très peu de celle de Gauss. C'est ce qu'il appelle le théorème fondamental et ce que nous appellerons le premier théorème de M. Lindeberg. L'hypothèse faite paraissant absolument admissible, c'est avec raison que

(1) Nous considérons dans la deuxième partie des erreurs dont la valeur absolue a sa valeur probable infinie; bien entendu, dans ce cas, l'hypothèse faite n'a pas de sens et ne s'applique pas.

(2) *Annales Academiæ Scientiarum Fennicæ*, série A., t. XVI (1920). *Voir* aussi la communication du même auteur au V^e^ Congrès des Mathématiciens scandinaves (Helsingfors, 1922).

l'auteur voit dans ce théorème une explication tout à fait satisfaisante, probablement la première en date, du rôle de la loi de Gauss.

Si ce théorème suffit pour la théorie des erreurs, il ne résout pas d'une manière satisfaisante la question mathématique qui se pose. L'hypothèse faite est trop restrictive, puisqu'elle n'est pas vérifiée dans le cas où les erreurs partielles obéissent déjà à la loi de Gauss, et que dans ce cas le théorème est évidemment vrai. Il faut alors chercher à établir le théorème fondamental en admettant pour les erreurs partielles des lois de probabilité telles que l'erreur ne soit pas limitée supérieurement en valeur absolue, mais que la probabilité des grandes valeurs de l'erreur soit suffisamment petite. Liapounoff avait dès 1900 obtenu des résultats importants dans cet ordre d'idées, mais en introduisant des hypothèses trop restrictives. M. Lindeberg, dans un Mémoire publié en 1922 (*Math. Zeitschrift*, vol. XV), et moi-même, dans des Notes publiées la même année dans les *Comptes rendus* de l'Académie des Sciences, avons résolu la question en n'introduisant que des hypothèses qui semblent être réellement dans la nature des choses. La première partie du présent travail est destinée à compléter les indications données dans mes Notes.

On m'excusera d'ajouter quelques mots pour indiquer l'indépendance de mes travaux et de ceux de M. Lindeberg. C'est à l'occasion de mon enseignement à l'École Polytechnique que j'ai été conduit à m'occuper du calcul des probabilités, et, le résultat essentiel que j'ai obtenu, ainsi que le principe de la méthode employée, se trouvent indiqués pour la première fois dans mon cours professé à l'automne de 1920, quoique avec assez peu de précision, le caractère de l'enseignement dont j'étais chargé ne me permettant pas de développer cette question. Quoique ayant constaté l'insuffisance, au point de vue qui nous occupe, des démonstrations contenues dans les Ouvrages classiques, je n'ai pas pensé d'abord qu'un résultat aussi aisé à obtenir, comme on le verra par la suite, par des méthodes classiques, pût être nouveau. J'étais occupé à cette époque par d'autres travaux, et ce n'est qu'en 1922, après avoir complété mes premiers résultats par d'autres concernant les lois exceptionnelles (celles pour lesquelles le théorème fondamental est en défaut), que je commençai à les publier, dans

des Notes présentées à l'Académie des Sciences, sans connaître encore les travaux de M. Lindeberg.

3. Si nous examinons maintenant les méthodes par lesquelles on a entrepris jusqu'ici de mettre en évidence que la loi de Gauss s'introduit nécessairement dans la *composition* d'un grand nombre d'erreurs (¹), il semble que trois méthodes doivent surtout retenir l'attention.

La première, qui est classique, repose sur l'étude du jeu de pile ou face. La courbe de Gauss s'introduit tout naturellement, parce qu'elle est dessinée à la limite, après un changement d'échelle convenable, par la succession des points qui représentent les coefficients du binome. Il semble que cette méthode donne l'explication la plus naturelle et la plus élémentaire du rôle de la loi de Gauss. Mais sa portée est insuffisante. Il est sans doute facile de traiter par cette méthode le cas d'un jeu comportant un nombre quelconque d'éventualités inégalement probables. Mais, pour arriver au théorème fondamental qui nous occupe, il faudrait d'une part supposer que la loi de probabilité dont dépend le gain du joueur est quelconque, d'autre part qu'elle peut varier d'un coup à l'autre. La méthode classique ne semble pas permettre cette double extension.

Avant d'indiquer le principe des autres méthodes, précisons la manière dont il y a lieu de définir une loi de probabilité à une variable. Il est clair que la notion d'une telle loi est liée à la notion de répartition, sur l'axe des x, de masses positives et de somme égale à l'unité; la probabilité pour que la variable considérée soit dans un certain intervalle est égale à la somme des masses situées dans cet intervalle dans la répartition liée à la loi considérée. Or il est bien connu, depuis les travaux de M. Lebesgue, qu'il peut exister trois sortes de masses :

1° Des masses réparties sur l'axe des x avec une densité sommable $f(x)$;

2° Des masses finies situées en certains points; ces points peu-

(¹) Nous appellerons dans la suite composition de plusieurs erreurs indépendantes (ou des lois de probabilités auxquelles elles obéissent), l'addition de ces erreurs.

vent être en nombre fini, ou constituer un ensemble dénombrable absolument quelconque;

3° Des masses réparties avec une densité infinie dans un ensemble de mesure nulle (mais non dénombrable), sans qu'aucun point contienne de masse finie. Un exemple donné par M. Lebesgue fait bien comprendre ce que peut être une telle répartition (1).

Si l'on veut définir la loi de probabilité la plus générale, il ne peut être alors question de le faire par la donnée de la fonction $f(x)$, qu'on appelle souvent fonction de probabilité et que je propose d'appeler *fonction des probabilités élémentaires*. On peut au contraire la définir par la donnée de la fonction $F(x)$, égale à la somme des masses situées dans l'intervalle $(-\infty, x)$; s'il y a une masse finie située exactement au point d'abscisse x, nous admettrons pour fixer les idées que la moitié de cette masse est comptée dans la somme $F(x)$. Cette fonction représente donc la probabilité pour que la variable soit inférieure à x, augmentée de la moitié de la probabilité pour qu'elle soit égale à x. Pour éviter toute confusion entre $F(x)$ et $f(x)$ qui est sa dérivée, nous l'appellerons *fonction des probabilités totales*. La valeur probable d'une fonction $\omega(x)$ doit alors être représentée par l'intégrale de Stieltjes (2)

$$\int \omega(x)\, dF(x).$$

Si l'on définit une loi de probabilité variable et une loi fixe par leurs fonctions des probabilités totales respectives $F(x)$ et $G(x)$, on doit naturellement considérer que la loi variable tend vers la loi fixe si $F(x)$ tend vers $G(x)$, sauf peut-être aux points de dis-

---

(1) H. LEBESGUE, *Leçons sur l'Intégration*, p. 55. *Voir* aussi mes *Leçons d'Analyse fonctionnelle*, p. 39.

(2) Au début de mon enseignement à l'École Polytechnique, je considérais ces notions comme banales. Les ayant exposées incidemment, en 1921, au cours d'une conférence faite à la Société de Philosophie, et ayant remarqué que les auditeurs au courant de la question considéraient comme nouvelle cette application au Calcul des Probabilités de notions classiques, je la donnai à ce moment comme étant de moi. J'ignorais à cette époque que cette application avait été faite systématiquement par M. Mises (*Math. Zeitschrift*, t. 8). D'après M. Lindeberg, c'est à M. Mises que revient la priorité sur cette question.

continuité de $G(x)$. La convergence est d'ailleurs nécessairement uniforme si la fonction $G(x)$ est continue. Une loi de probabilité tend alors vers la loi de Gauss *réduite* (¹) si la valeur limite de $F(x)$ est

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt.$$

Or la fonction des probabilités totales ne se prête pas bien à l'étude de l'addition des erreurs, à cause de la nécessité d'introduire une intégrale d'ordre $n-1$ pour définir la fonction des probabilités totales relative à une erreur, somme de n erreurs données. Il est alors indiqué de chercher des manières *indirectes* de définir une loi de probabilité.

La méthode la plus simple me paraît être celle qui repose sur la considération de la *fonction caractéristique* $\varphi(z)$, valeur probable de e^{izx}.

Poincaré, dans la deuxième édition de son *Calcul des Probabilités*, appelle fonction caractéristique la valeur de e^{zx}, c'est-à-dire la fonction $\varphi\left(\frac{z}{i}\right)$. Il est certainement préférable d'introduire la fonction $\varphi(z)$, tant parce qu'elle est toujours parfaitement définie quelle que soit la loi de probabilité considérée, que parce que la donnée de cette fonction, pour z réel, définit parfaitement cette loi. Dans le cas d'une loi pouvant être définie par sa fonction des probabilités élémentaires $f(x)$, cela résulte immédiatement de la formule connue de Fourier

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-izx} \varphi(z)\, dz.$$

Pour le cas général, je renvoie à ma Note du 13 décembre 1922.

Si une loi variable et une loi fixe ont des fonctions caractéristiques désignées respectivement par $\varphi(z)$ et $\omega(z)$, on peut consi-

---

(¹) Nous disons qu'une loi est réduite si la valeur probable de la variable est nulle et si sa valeur quadratique moyenne est égale à l'unité. Toute loi pour laquelle la valeur quadratique moyenne est finie peut être réduite par un changement d'origine et un changement d'unité. On remarquera que, pour qu'une loi tende vers la loi de Gauss, il est *suffisant, mais non nécessaire*, que la loi réduite correspondante tende vers la loi de Gauss réduite.

dérer que la loi variable tend vers la loi fixe si $\varphi(z)$ tend vers $\omega(z)$, et cela uniformément dans tout intervalle fini. Cette définition de la limite d'une loi de probabilité est exactement équivalente à la précédente, comme je l'ai montré dans ma Note citée. La loi variable tend donc vers la loi de Gauss réduite si $\varphi(z)$ tend vers $e^{-\frac{z^2}{2}}$, et cela uniformément dans tout intervalle fini.

L'application de la notion de fonction caractéristique simplifie singulièrement l'étude de la composition des lois de probabilités. Si en effet les lois de probabilités de n erreurs indépendantes ont pour fonctions caractéristiques respectives $\varphi_1(z), \varphi_2(z), \ldots, \varphi_n(z)$, la loi de probabilité à laquelle obéit leur somme a pour caractéristique

$$\varphi(z) = \varphi_1(z)\,\varphi_2(z)\ldots\,\varphi_n(z),$$

comme on le vérifie aisément. Cette formule si simple résout immédiatement tous les problèmes relatifs à la composition des erreurs, à cela près qu'il n'est pas toujours facile de voir la signification, au point de vue de la loi de probabilité considérée, d'une condition imposée à $\varphi(z)$. Cauchy, qui a dès 1853 appliqué cette méthode, ne semble pas s'en être aperçu. Malgré cette difficulté, il me semble que la méthode en question est celle par laquelle il est le plus naturel de traiter les problèmes de la théorie des erreurs. C'est par son application systématique que j'ai obtenu des résultats, publiés en partie depuis 1922 dans les *Comptes rendus*, et dont les principaux seront exposés dans la suite (¹).

Indiquons avant cela le principe de la troisième méthode qui paraît devoir retenir l'attention, celle de M. Lindeberg.

J'appellerai *opération fonctionnelle de M. Lindeberg* l'opération qui consiste à passer d'une fonction $\omega(x)$ à la fonction

$$\Omega(x) = \int_{-\infty}^{+\infty} \omega(x-t)\,d\mathrm{F}(t).$$

A chaque loi de probabilité correspond donc une telle opération, et l'on constate sans peine qu'effectuer successivement les opérations

(¹) Un exposé complet ne pouvant trouver place dans les limites de ce Mémoire, je pense revenir sur ces questions dans un Ouvrage sur le Calcul des Probabilités.

fonctionnelles relatives à plusieurs erreurs indépendantes équivaut à effectuer l'opération fonctionnelle relative à l'erreur totale. Ces transformations fonctionnelles se multiplient donc tout aussi simplement que les fonctions caractéristiques, et si la méthode de M. Lindeberg me paraît moins simple que la précédente, c'est parce que la notion de transformation fonctionnelle est moins simple que celle de fonction, et en fait ses calculs sont plutôt moins simples que les miens.

Par contre, la définition de l'opération fonctionnelle considérée donne évidemment une définition surabondante de la loi de probabilité, et l'on peut en profiter, en choisissant la fonction $\omega(x)$ de manière à mettre immédiatement en évidence la fonction des probabilités totales. Si $\omega(x) = 0$ pour x négatif et 1 pour x positif, on a évidemment $\Omega(x) = F(x)$. Comme il faut que $\omega(x)$ soit continu, et qu'il est même utile que cette fonction ait des dérivées des trois premiers ordres continues, M. Lindeberg introduit un raccord convenable entre l'intervalle où cette fonction est nulle et celui où elle est égale à l'unité, et se trouve ainsi en présence d'une méthode très simple et d'une remarquable portée (1).

4. Étudions donc la méthode basée sur l'emploi de la fonction caractérisque. Voici d'abord à peu près ce que l'on trouve dans Poincaré :

Il suppose la fonction caractéristique $\varphi(z)$ développable en série de Maclaurin. Si la loi considérée est réduite, on a évidemment

$$\varphi(0) = 1, \qquad \varphi'(0) = 0, \qquad \varphi''(0) = 1,$$

et, par suite,

$$(1) \qquad \psi(z) = \log\varphi(z) = -\frac{z^2}{2} + \alpha_3 z^3 + \alpha_4 z^4 + \dots.$$

Si n erreurs indépendantes obéissent à cette même loi, leur somme

(1) Pour bien comprendre l'esprit de la méthode de M. Lindeberg, on peut remarquer que si la fonction $\omega(x)$ définit une loi de probabilité (que ce soit sa fonction des probabilités totales), l'opération fonctionnelle effectuée sur cette fonction n'est autre que la composition de la loi de probabilité qu'elle définit avec celle qui s'agit d'étudier: la loi résultante étant nécessairement continue si une seule des lois composantes l'est. on s'explique qu'il puisse y avoir intérêt à remplacer la loi étudiée par le résultat de sa composition avec une loi continue. J'ai utilisé la même idée dans ma Note du 13 novembre 1922.

obéit à une loi pour laquelle le logarithme de la fonction caractéristique est $n\psi(z)$, et pour la loi réduite correspondante, le logarithme de la fonction caractéristique est évidemment

$$(2) \qquad \Psi(z) = n\psi\left(\frac{z}{\sqrt{n}}\right) = -\frac{z^2}{2} + \alpha_3\frac{z^3}{\sqrt{n}} + \alpha_4\frac{z^4}{n} + \ldots.$$

Pour n infini, cette fonction tend vers $-\frac{z^2}{2}$, de sorte que la loi considérée tend vers celle de Gauss.

Or l'hypothèse que $\varphi(z)$ soit développable en série ne sert à rien. Sous la seule hypothèse que la valeur quadratique moyenne soit finie, ce qui permet de raisonner sur les lois réduites, on a

$$(3) \qquad \psi(z) = -\frac{z^3}{2}[1 + \omega(z)],$$

la fonction $\omega(z)$ s'annulant avec z. L'expression (2) devient alors

$$-\frac{z^2}{2}\left[1 + \omega\left(\frac{z}{\sqrt{n}}\right)\right].$$

Elle tend vers $-\frac{z^2}{2}$ pour n infini, et cela uniformément dans tout intervalle fini, de sorte que la loi considérée tend vers la loi de Gauss.

Si maintenant l'on veut généraliser les résultats précédents en étudiant la somme d'erreurs indépendantes obéissant à des lois différentes, il y a, comme nous allons le voir, quelques précautions à prendre, mais il n'y a aucune difficulté essentielle, et l'on peut être surpris que les résultats qui suivent ne soient pas connus depuis longtemps, le problème ayant retenu l'attention des plus grands savants, et le principe de la méthode qui permet de le résoudre étant connu depuis longtemps.

5. Introduisons quelques définitions qui rendront nos raisonnements plus intuitifs.

Nous appellerons loi $\mathcal{L}_2$ une loi de probabilité pour laquelle la valeur quadratique moyenne de la variable soit finie. Nous définirons les lois que nous aurons à considérer par la fonction $\psi(z)$,

logarithme de la fonction caractéristique $\varphi(z)$ (1), et nous ne considérerons que des *erreurs*, ou variables dont la valeur probable est nulle. Pour les lois $\mathcal{L}_2$ réduites, la fonction $\psi(z)$ est de la forme (3); pour les lois non réduites correspondant à une telle loi par un changement d'unité, le logarithme de la fonction caractéristique est $\psi(mz)$, m étant la valeur quadratique moyenne.

Nous appellerons *famille normale de lois* $\mathcal{L}_2$ *réduites* un ensemble de lois pour lesquelles, la fonction $\psi(z)$ étant mise sous la forme (3), on ait, lorsque $|z|$ est inférieur à un nombre positif ζ,

$$|\omega(z)| \leqq h(z),$$

$h(z)$ désignant une fonction paire de z, nulle pour $z = 0$, et ne décroissant jamais lorsque $|z|$ croît de 0 à ζ; bien entendu ζ et $h(z)$ sont les mêmes pour toutes les lois de la famille considérée. La condition que $h(z)$ soit une fonction paire, et non décroissante lorsque z croît, n'est d'ailleurs nullement restrictive, puisque autrement on pourrait remplacer $h(z)$ par le maximum de $h(t)$ quand t varie de $-z$ à $+z$.

Une famille de lois $\mathcal{L}_2$ non réduite sera dite normale si les lois réduites correspondantes constituent une famille normale.

Si l'on représente chaque loi de probabilité $\mathcal{L}_2$ réduite par un point d'un espace idéal, il est commode de considérer qu'une famille normale correspond à des points d'une région bornée de cet espace, et qu'une famille non réduite comprend des points s'éloignant à l'infini. Il ne faut d'ailleurs pas chercher à donner un sens très précis à cette image, destinée seulement à rendre les raisonnements plus intuitifs.

Nous dirons d'autre part qu'une loi de probabilité *appartient au domaine d'attraction de la loi de Gauss* si la somme de n erreurs indépendantes obéissant à cette loi, divisée par un facteur convenable, obéit à une loi tendant vers celle de Gauss si n augmente indéfiniment.

D'après le n° 4, toute loi $\mathcal{L}_2$ appartient au domaine d'attraction

(1) On remarquera que, $\varphi(z)$ pouvant s'annuler et changer de signe, $\psi(z)$ peut devenir infini ou imaginaire; mais cette fonction est toujours réelle et finie pour z assez petit, et cela suffit pour qu'il n'y ait aucune difficulté pour la suite.

de la loi de Gauss. Le résultat que nous allons maintenant établir peut s'énoncer en disant qu'on est assuré de trouver la loi de Gauss à la limite, les valeurs quadratiques moyennes des erreurs partielles étant bien entendu supposées très petites, si les lois de probabilités correspondant à ces erreurs constituent une famille normale, mais que des circonstances différentes peuvent se produire si ces lois correspondent à des points de l'espace idéal considéré s'éloignant à l'infini.

6. Le théorème fondamental est donc le suivant :

THÉORÈME. — *Si* $x_1, x_2, \ldots, x_n, \ldots$ *sont des variables indépendantes obéissant à des lois* $\mathcal{L}_2$ *réduites qui constituent une famille normale, et si* $m_1, m_2, \ldots, m_n, \ldots$ *sont des nombres positifs tels que,* μ_n *désignant le plus grand des nombres* $m_1, m_2, \ldots, m_n$ *et* M_n^2 *étant le nombre positif tel que*

$$M_n^2 = m_1^2 + m_2^2 + \ldots + m_n^2,$$

le rapport $\frac{\mu_n}{M_n}$ *tende vers zéro pour n infini, alors la variable*

$$X_n = m_1 x_1 + m_2 x_2 + \ldots + m_n x_n$$

obéit à une loi de probabilité tendant pour n infini vers la loi de Gauss.

Définissons la loi de probabilité à laquelle obéit x_i par le logarithme $\psi_i(z)$ de sa fonction caractéristique. En posant

$$\psi_i(z) = -\frac{z^2}{2}[1 + \omega_i(z)],$$

on a, par hypothèse,

$$|\omega_i(z)| \leq h(z), \qquad [|z| < \zeta], \tag{4}$$

$h(z)$ étant une fonction paire, nulle pour $z = 0$ et ne décroissant pas quand $|z|$ croît de 0 à ζ. La loi de probabilité à laquelle obéit x_n a le logarithme de sa fonction caractéristique évidemment égal à

$$\Psi_n(z) = \psi_1\left(\frac{m_1 z}{M_n}\right) + \psi_2\left(\frac{m_2 z}{M_n}\right) + \ldots + \psi_n\left(\frac{m_n z}{M_n}\right)$$
$$= -\frac{z^2}{2}[1 + \Omega_n(z)],$$

en posant

$$\Omega_n(z) = \frac{m_1^2 \omega_1\left(\frac{m_1}{M_n} z\right) + \ldots + m_n^2 \omega_n\left(\frac{m_n}{M_n} z\right)}{M_n^2}.$$

D'après l'inégalité (4), et l'hypothèse que la fonction $h(z)$ ne puisse décroître quand z croît, on a

$$|\Omega_n(z)| < h\left(\frac{\mu_n}{M_n}|z|\right).$$

Le rapport $\frac{\mu_n}{M_n}$ tendant vers zéro, il résulte de cette inégalité que cette fonction tend vers zéro, et cela uniformément dans tout intervalle fini. Alors $\Psi_n(z)$ tend vers $-\frac{z^2}{2}$, et cela uniformément dans tout intervalle fini. Il résulte alors des résultats rappelés au n° 3 que la loi de probabilité à laquelle obéit X_n tend vers la loi de Gauss. C. Q. F. D.

7. Montrons maintenant par un exemple que l'introduction de la notion de famille normale est essentielle.

M. Lindeberg considère n variables indépendantes, obéissant à une même loi de probabilité $\mathfrak{P}_n$ définie de la manière suivante : les valeurs -1, 0, 1, seules possibles, ont pour probabilités respectives $\frac{1}{2n}$, $1 - \frac{1}{n}$, $\frac{1}{2n}$. La somme de ces n variables obéit à une loi que nous désignerons par $\mathfrak{P}_n^n$. M. Lindeberg signale que, pour n infini, cette loi ne tend pas vers celle de Gauss. Il suffit, pour s'en convaincre, d'observer que la probabilité de la valeur zéro est manifestement supérieure à $\left(1 - \frac{1}{n}\right)^n$, donc à la limite au moins égale à $\frac{1}{e}$ (1).

D'ailleurs notre méthode permet de trouver d'une manière précise la limite de cette loi. La loi $\mathfrak{P}_n^n$ ayant pour fonction caractéris-

(1) Il faut bien remarquer que le résultat de M. Lindeberg n'est pas en contradiction avec le fait que, si l'on prend une loi $\mathfrak{P}_n$ bien déterminée et qu'on considère la somme d'un nombre indéfiniment croissant N de variables qui obéissent à cette loi, son quotient par $\sqrt{\frac{N}{n}}$ obéit à la limite de la loi de Gauss.

tique

$$\left(1-\frac{1-\cos z}{n}\right)^n$$

tend vers une loi $\mathfrak{P}$ dont la fonction caractéristique est

$$e^{\cos z-1}.$$

C'est une loi pour laquelle les valeurs entières sont seules possibles, chacune des valeurs entières $+p$ et $-p$ ayant pour probabilité

$$\frac{1}{2\pi}\int_0^{2\pi} e^{\cos z-1}\cos pz\,dz.$$

Ces probabilités peuvent s'exprimer à l'aide de la fonction de Bessel.

Observons d'autre part que, sans modification essentielle dans tout ce qui précède, on peut énoncer notre théorème fondamental sous la forme suivante :

Si n variables $x_1, x_2, \ldots, x_n$ obéissent à des lois $\mathfrak{L}_2$ réduites appartenant à une famille normale bien déterminée, et si $m_1, m_2, \ldots, m_n$ étant des coefficients tels que

$$m_1^2+m_2^2+\ldots+m_n^2=1,$$

la loi de probabilité à laquelle obéit l'expression

$$m_1x_1+m_2x_2+\ldots+m_nx_n$$

diffère de la loi de Gauss d'aussi peu qu'on le veut, pourvu que le plus grand des coefficients $m_1, m_2, \ldots, m_n$ soit suffisamment petit (ce qui implique évidemment que n soit grand).

Étant donné cet énoncé, la loi $\mathfrak{P}_n^n$ de l'exemple de M. Lindeberg ne peut tendre, pour n infini, vers une limite différente de la loi de Gauss que parce que l'ensemble des lois $\mathfrak{P}_n$ ne constitue pas une famille normale. L'importance de cette notion de famille normale est donc établie.

Au lieu de modifier l'énoncé du théorème fondamental, on peut modifier l'exemple de M. Lindeberg de manière à arriver au même résultat. Considérons une suite indéfinie de variables

$$y_1,\ y_2,\ \ldots,\ y_n,\ \ldots,$$

obéissant respectivement aux lois

$$\mathfrak{P}_1^1, \quad \mathfrak{P}_2^2, \quad \ldots, \quad \mathfrak{P}_n^n, \quad \ldots,$$

et étudions la somme

$$Y_n = a_1 y_1 + a_2 y_2 + \ldots + a_n y_n, \tag{5}$$

les a_n étant des coefficients croissants, et tels que $\sqrt{n}\frac{a_{n-1}}{a_n}$ augmente indéfiniment avec n. On peut considérer y_n comme la somme de n variables

$$x_{N-n+1}, \quad x_{N-n+2}, \quad \ldots \quad x_N, \qquad \left[N = \frac{n(n+1)}{2}\right],$$

obéissant à loi $\mathfrak{P}_n$. Dans ces conditions, Y_n prend la forme

$$X_N = m_1 x_1 + m_2 x_2 + \ldots + m_N x_N, \tag{6}$$

les coefficients m_i vérifiant la condition indiquée dans le premier énoncé du théorème fondamental, c'est-à-dire que chacun des termes de la somme (6) n'aura qu'une influence négligeable sur la loi de probabilité à laquelle obéit cette somme. Si donc cette loi ne tend pas vers celle de Gauss, ce sera parce que les lois $\mathfrak{P}_n$ ne constituent pas une famille normale.

Prenons alors

$$a_n = a^n, \qquad (a > 1).$$

Le dernier terme de la somme (5), constitué par la réunion de n termes de la somme (6), obéit à une loi tendant vers la loi $\mathfrak{P}$, et n'est évidemment pas négligeable devant l'ensemble des autres. Il en résulte évidemment que X_n n'obéit pas à la limite à la loi de Gauss (si en effet la loi à laquelle obéit X_{n-1} différait peu de celle de Gauss, celle à laquelle obéit X_n en différerait).

D'ailleurs on voit aisément que la loi à laquelle obéit $\frac{X_n}{a_n}$ tend vers une loi limite, pour laquelle le logarithme de la fonction caractéristique est

$$-2\left[\sin^2\frac{z}{2} + \sin^2\frac{z}{2a} + \ldots + \sin^2\frac{z}{2a^n} + \ldots\right].$$

On peut obtenir une limite différente en prenant pour les a_n une suite de croissance plus rapide. Si par exemple le rapport $\frac{a_{n+1}}{a_n}$

augmente indéfiniment (mais moins rapidement que $\sqrt{n}$), la loi à laquelle obéit le rapport $\frac{X_n}{a_n}$ tend vers la loi $\mathfrak{P}$.

Si au contraire on prend une suite de croissance moins rapide, et telle que $\frac{a_{n+1}}{a_n}$ tende vers l'unité, on obtient à la limite la loi de Gauss.

En repensant alors à l'espace idéal dont chaque point représente une loi $\mathfrak{L}_2$, nous pouvons énoncer les résultats sous la forme suivante, en considérant l'addition de deux erreurs comme la *composition de deux points*, effectuée en attribuant à chacun d'eux un *poids* égal à la valeur probable du carré de l'erreur correspondante : alors l'opération de composition conduit à se rapprocher du point O qui représente la loi de Gauss, et si l'on compose un grand nombre de points pris dans une région bornée, les poids étant tels que le plus grand soit négligeable devant la somme des autres, on obtient un point très voisin du point O. Il n'en sera plus de même en général si les points choisis constituent une suite de points s'éloignant à l'infini; mais même dans ce cas on peut obtenir le point O à la limite si les poids correspondant aux différents points ne croissent pas trop rapidement.

8. Il reste à se demander ce que devient la notion de famille normale si l'on définit les lois de probabilités considérées sans faire intervenir leurs fonctions caractéristiques. Sans résoudre complètement ce problème, nous établirons le théorème suivant :

Pour qu'un ensemble de lois $\mathfrak{L}_2$ réduites constituent une famille normale, il suffit que, quelque petit que soit ε, *on puisse déterminer* X *tel que*

$$\int_{-X}^{+X} x^2 dF(x) > 1 - \varepsilon, \tag{7}$$

$F(x)$ *désignant la fonction des probabilités totales de n'importe laquelle des lois considérées.*

Soit en effet $\varphi(z)$ la fonction caractéristique correspondant à la fonction des probabilités totales $F(x)$. L'intégrale

$$\varphi''(z) = -\int_{-}^{+\infty} x^2 e^{izx} dF(x)$$

admettant comme majorante l'intégrale

$$\int_{-\infty}^{+\infty} x^2\, dF(x),$$

on commet une erreur inférieure à 2ε en remplaçant $\varphi''(z)+1$ par

$$\int_{-X}^{+X} x^2(1-e^{izx})\, dF(x).$$

Cette expression elle-même s'annulant pour $z=0$ et ayant sa dérivée inférieure en valeur absolue à

$$\int_{-X}^{+X} |x^3|\, dF(x) \leqq X \int_{-\infty}^{+\infty} x^2\, dF(x) = X$$

est inférieure à $X|z|$, et l'on a, quel que soit z réel,

$$|\varphi''(z)+1| < 2\varepsilon + X|z|,$$

d'où, par deux intégrations successives,

$$\left|\varphi(z) - 1 + \frac{z^2}{2}\right| < \varepsilon z^2 + X\frac{|z^3|}{6}.$$

Or, si a et b sont supérieurs à $\frac{1}{2}$, on a évidemment

$$\left|\log b - \log a\right| = \left|\int_a^b \frac{dx}{x}\right| < 2\,|b-a|.$$

Si z est assez petit, $\varphi(z)$ et $1-\frac{z^2}{2}$ sont certainement supérieurs à $\frac{1}{2}$, et il vient

$$\left|\log\varphi(z) - \log\left(1-\frac{z^2}{2}\right)\right| < 2\left|\varphi(z) - 1 + \frac{z^2}{2}\right| < 2\varepsilon z^2 + \frac{X}{3}|z^3|,$$

et par suite

$$\frac{2}{z^2}\left|\log\varphi(z) + \frac{z^2}{2}\right| < 5\varepsilon + \frac{2}{3}X|z|.$$

Le nombre ε pouvant être pris aussi petit qu'on le veut, le second membre de cette inégalité peut lui-même être rendu aussi petit qu'on le veut à condition que z soit assez petit. Cette inégalité

donne donc

$$\frac{2}{z^2}\left|\log\varphi(z)+\frac{z^2}{2}\right|<h(z),$$

$h(z)$ étant une fonction déterminée de z, s'annulant avec z, c'est-à-dire enfin

$$|\omega(z)|<h(z),$$

en posant

$$\log\varphi(z)=-\frac{z^2}{2}[1+\omega(z)].$$

Les lois de probabilité considérées constituent donc une famille normale. C. Q. F. D.

On peut dire que la condition (7) exprime que, pour toutes les lois considérées, supposées réduites, les intégrales

$$\int_{-\infty}^{+\infty} x^2\,dF(x),$$

nécessairement égales à l'unité, *convergent également* vers cette limite. Si les variables obéissant à ces lois sont limitées supérieurement, cette condition est sûrement vérifiée. Dans le cas contraire, il faut que les probabilités des grandes valeurs de la variable soient suffisamment petites.

Cette condition d'égale convergence, suffisante pour que les lois considérées constituent une famille normale, l'est donc pour l'application du théorème fondamental.

9. Le résultat précédent ne coïncide pas exactement avec celui de M. Lindeberg, qui raisonne sur les lois de probabilités auxquelles obéissent les erreurs partielles, et non sur les lois réduites correspondantes. Il suffit de modifier très légèrement notre raisonnement pour retrouver exactement le résultat de M. Lindeberg.

Soit $F_i(x)$ la fonction des probabilités totales d'une telle loi; soit m_i l'erreur quadratique moyenne, la valeur probable de l'erreur étant nulle. Posons

$$\int_{-X}^{+X} x^2\,dF_i(x)=m_i^2-\varepsilon_i.$$

Un calcul analogue à celui qui précède donne, pour la fonction

caractéristique $\varphi_i(z)$, l'inégalité

$$\left|\varphi_i(z) - 1 + m_i^2 \frac{z^2}{2}\right| < \varepsilon_i z^2 + \frac{1}{6} X m_i^2 |z^3|,$$

et pour son logarithme $\psi_i(z)$,

$$\left|\psi_i(z) + m_i^2 \frac{z^2}{2}\right| < 3\varepsilon_i z^2 + \frac{1}{3} X m_i^2 |z^3|,$$

cette inégalité étant valable, non plus seulement pour z petit, mais dans un intervalle d'autant plus grand que m_i est plus petit.

Si n est le nombre des erreurs partielles, l'addition des inégalités analogues donne, en supposant $\Sigma m_i^2 = 1$, et en désignant par $\Psi(z)$ le logarithme de la fonction caractéristique de la loi à laquelle obéit l'erreur totale,

$$\left|\Psi(z) + \frac{z^2}{2}\right| < 3z^2 \Sigma \varepsilon_i + \frac{1}{3} X |z^3|.$$

L'hypothèse que fait M. Lindeberg revient à supposer que l'on peut rendre X et $\Sigma\varepsilon_i$ simultanément très petits (¹); dans ces conditions, le plus grand des nombres m_i étant également supposé très petit, l'inégalité précédente s'applique, et montre que, dans n'importe quel intervalle fini, $\Psi(z)$ diffère peu de $-\frac{z^2}{2}$. Alors la loi de probabilité à laquelle obéit l'erreur totale diffère peu de la loi de Gauss, ce qui démontre le deuxième théorème de M. Lindeberg, et par suite aussi son premier théorème, qui en est un cas particulier.

SECONDE PARTIE.

LES LOIS EXCEPTIONNELLES.

10. L'objet de cette seconde partie est d'étudier les lois *excep-*

(¹) La somme $\Sigma\varepsilon_i$ est une fonction de X qui décroît de 1 à 0 quand X croît à partir de 0. L'hypothèse considérée revient à dire qu'elle décroît suffisamment vite, ou encore que, dans le calcul de Σm_i^2, les petites valeurs de la variable x ont seules de l'importance. Cette hypothèse étant sûrement réalisée si toutes les erreurs partielles restent inférieures à un nombre très petit, le premier théorème de M. Lindeberg est un cas particulier de son second théorème.

tionnelles, c'est-à-dire n'appartenant pas au domaine d'attraction de la loi de Gauss. Pour une telle loi, l'erreur quadratique moyenne est nécessairement infinie.

Nous mettrons en évidence l'existence de certaines lois qui jouent, pour un ensemble convenable de lois, le même rôle que la loi de Gauss pour l'ensemble des lois L_2. Il est évidemment nécessaire pour cela qu'une telle loi soit *stable*, c'est-à-dire que, x_1, x_2, ..., x_n étant des variables indépendantes obéissant à cette loi, et a_1, a_2, ..., a_n des coefficients positifs quelconques, on puisse déterminer un nombre positif A, fonction de a_1, a_2, ..., a_n, tel que la variable

$$X = \frac{a_1 x_1 + a_2 x_2 + \ldots + a_n x_n}{A}$$

obéisse à la même loi; il suffit d'ailleurs évidemment, pour qu'une loi soit stable, que cette propriété soit vérifiée pour $n = 2$.

Si l'on cherche à définir une telle loi par le logarithme $\psi(z)$ de sa fonction caractéristique, la propriété indiquée s'exprime par la relation

$$\psi(a_1 z) + \psi(a_2 z) = \psi(A z),$$

d'où l'on déduit aisément

$$\psi(z) = -k\,|z^\alpha| \tag{7}$$

et

$$A^\alpha = a_1^\alpha + a_2^\alpha,$$

ou, si n est quelconque,

$$A^\alpha = a_1^\alpha + a_2^\alpha + \ldots + a_n^\alpha. \tag{8}$$

Dans la formule (7), k est une quantité qui peut avoir une valeur constante pour z positif, et une autre valeur constante pour z négatif. Nous n'étudierons ici que le cas des lois symétriques, pour lesquelles ces deux constantes ont une même valeur réelle et positive (pour le cas des lois dissymétriques, voir ma Note des *Comptes rendus* du 7 mai 1923). Par un changement convenable d'unité, on peut donner à k telle valeur positive que l'on voudra; nous prendrons la valeur $\frac{1}{\Gamma(\alpha+1)}$, et désignerons par L_α la loi ainsi définie, pour laquelle

$$\psi(z) = -\frac{|z^\alpha|}{\Gamma(\alpha+1)}.$$

D'ailleurs, pour une loi de probabilité quelconque, $\psi(0) = 0$, et le rapport $\frac{1-\varphi(z)}{z^2}$ a pour $z = 0$ une limite bien déterminée, finie ou infinie suivant que l'erreur quadratique moyenne est finie ou infinie, mais jamais nulle. L'exposant α est donc positif et ne peut être supérieur à 2 ; on remarque que, pour $\alpha = 2$, on retrouve la loi de Gauss.

Ces lois L_α sont évidemment des lois continues, dont les fonctions des probabilités élémentaires $f_\alpha(x)$ sont données par la formule de Fourier

$$f_\alpha(x) = \frac{1}{\pi}\int_0^{+\infty} \cos xz \; e^{-\frac{z^\alpha}{\Gamma(\alpha+1)}} dz.$$

Mais leur existence n'est pas évidente; pour l'établir il reste à montrer que $f_\alpha(x)$ est toujours positif ou nul; cette condition ne peut être vérifiée pour $\alpha > 2$ (car autrement on serait conduit à la conclusion que la loi L_α existe, contrairement aux remarques qui précèdent); mais il n'est pas évident qu'elle l'est pour $\alpha < 2$, ou du moins cela n'est évident que pour $\alpha = 1$, au cas où la fonction $f(x)$, facile à calculer, a la valeur

$$f_1(x) = \frac{1}{\pi}\frac{1}{1+x^2}.$$

Cauchy, qui connaissait déjà les propriétés des lois L_α, n'a pas remarqué cette difficulté. M. G. Pólya l'a remarquée et résolue pour $\alpha < 1$. On va voir que les lois L_α existent effectivement quand α est quelconque entre 0 et 2, comme je l'ai énoncé dans ma Note du 27 mars 1922 et démontré dans celle du 23 avril 1923 (1).

11. Appelons loi $\mathcal{L}_\alpha$ toute loi pour laquelle on ait

$$\psi(z) = -\frac{k}{\Gamma(\alpha+1)}|z^\alpha|[1+\omega(z)],$$

la fonction $\omega(z)$ s'annulant avec z. Une telle loi sera dite *réduite*

(1) En publiant la première de ces Notes, j'ignorais que les lois L_α aient déjà été considérées. M. G. Pólya, en me signalant ses propres travaux, a eu l'amabilité d'attirer mon attention sur ceux de Cauchy, publiés dans les *Comptes rendus* de 1853.

si $k = 1$; elle peut toujours être réduite par un changement d'unité. Nous dirons que les lois $\mathcal{L}_\alpha$ réduites constituent une *famille normale* si toutes les fonctions $\omega(z)$ correspondantes sont majorées par une même fonction $h(z)$, définie et continue si z est assez petit, et s'annulant avec z.

En partant de ces définitions, on peut généraliser le théorème fondamental du n° 6, qui correspond au cas où $\alpha = 2$. Cette généralisation est tellement immédiate qu'il ne paraît pas utile de reprendre la démonstration. On voit en particulier que toute loi $\mathcal{L}_\alpha$ appartient au domaine d'attraction de la loi L_α correspondante, et que, si n erreurs indépendantes obéissent à une loi $\mathcal{L}_\alpha$, le quotient de leur somme par $N = n^{\frac{1}{\alpha}}$ obéit à une loi qui tend, pour n infini, vers la loi L_α.

Rien d'ailleurs dans ces raisonnements ne suppose l'existence de la loi L_α établie *a priori*. Pour établir cette existence, il nous suffit donc d'établir l'existence d'une loi $\mathcal{L}_\alpha$.

Considérons d'abord à cet effet une loi dissymétrique, pour laquelle la fonction des probabilités élémentaires $f(x)$, nulle pour $x < X$ (X étant positif), soit égale pour $x > X$ à

$$\frac{C}{x^{\alpha+1}}, \quad (C = \alpha X^\alpha).$$

Pour cette loi, la fonction caractéristique est, pour z positif,

$$\varphi(z) = C\int_X^\infty \frac{e^{izx}\,dx}{x^{\alpha+1}} = Cz^\alpha \int_{zX}^\infty \frac{e^{i\xi}\,d\xi}{\xi^{\alpha+1}}$$

(en posant $zx = \xi$).

Le calcul qui suit étant valable si α est un nombre positif non entier, nous nous placerons d'abord dans ce cas, et appellerons p le plus grand nombre entier inférieur à α, et $P(x)$ le polynome de degré p constitué par les $p+1$ premiers termes du développement de e^x en série de Maclaurin. La formule précédente donne immédiatement

$$(9) \qquad \varphi(z) = Cz^\alpha \int_{zX}^\infty \frac{P(i\xi)\,d\xi}{\xi^{\alpha+1}} - Cz^\alpha \int_0^{zX} \frac{e^{i\xi} - P(i\xi)}{\xi^{\alpha+1}}\,d\xi + Cz^\alpha \int_0^\infty \frac{e^{i\xi} - P(i\xi)}{\xi^{\alpha+1}}\,d\xi,$$

ce qui montre que $\varphi(z)$ est la somme d'une fonction entière et de $\lambda C z^{\alpha}$, λ ayant la valeur

$$\lambda = \int_0^{\infty} \frac{e^{i\xi} - P(i\xi)}{\xi^{\alpha+1}} d\xi.$$

Il est d'ailleurs facile de calculer cette intégrale en faisant les deux transformations suivantes : 1° poser $i\xi = x$, et intégrer de o à $+\infty$ (au lieu d'intégrer de o à $-\infty\, i$), ce qui ne change pas la valeur de l'intégrale (l'intégrale le long d'un quart de cercle de très grand rayon étant nulle); 2° intégrer $p+1$ fois par parties, en dérivant le numérateur et intégrant le dénominateur. Il vient ainsi

$$\lambda = i^{-\alpha} \frac{\Gamma(-\alpha)}{\Gamma(p+1-\alpha)} \int_0^{\infty} \frac{e^{-x}}{x^{\alpha-p}} dx = i^{-\alpha}\Gamma(-\alpha),$$

Posons maintenant

$$f(x) = f_0(x) + f_1(x),$$

$f_0(x)$ étant pair et $f_1(x)$ impair; $f_0(x)$ est donc une fonction nulle si $|x| < X$ et égale à $\frac{C}{2|x|^{\alpha+1}}$ si $|x| > X$. La loi de probabilité ayant $f_0(x)$ pour fonction des probabilités élémentaires admet évidemment pour fonction caractéristique la partie réelle de celle que nous venons de former, et comme cette fonction est nécessairement paire, elle est la somme d'une fonction entière paire et de l'expression

$$(10) \qquad C\cos\frac{\pi}{2}\alpha\,\Gamma(-\alpha)|z|^{\alpha} = -\frac{\pi}{\sin\frac{\pi}{2}\alpha}\frac{C}{2}\frac{|z|^{\alpha}}{\Gamma(\alpha+1)}\ (^1).$$

Si α est entier, il faut prendre $p = \alpha - 1$, et la deuxième inté-

---

(1) On remarque la simplification qui se produirait si nous avions introduit $\frac{x^{-\alpha-1}}{\Gamma(-\alpha)}$ au lieu de $x^{-\alpha-1}$. Nous ne l'avons pas fait, tant parce que le résultat sous la forme indiquée dans le texte s'étend sans modification au cas où α est un entier impair, que parce qu'il y a avantage à donner pour la forme réduite des lois $\mathcal{L}_{\alpha}$ une définition qui coïncide avec celle déjà donnée dans le cas où $\alpha = 2$, et en conséquence à introduire l'expression $\frac{|z|^{\alpha}}{\Gamma(\alpha+1)}$.

grale de la formule (9) introduit le terme

$$-C\frac{i^{\alpha}}{\Gamma(\alpha+1)}z^{\alpha}\log z,$$

ce qui conduit à une modification essentielle du résultat précédent si α est pair; mais si α est impair la modification, ne portant que sur la partie imaginaire de $\varphi(z)$, disparaît lorsque, considérant une loi symétrique, on n'a qu'à prendre la partie réelle de cette fonction.

Dans le cas qui nous intéresse, α est compris entre 0 et 2; le résultat obtenu est toujours applicable, et comme $\varphi'(0)=0$ si $\alpha>1$, l'expression (10) constitue la partie principale, pour z infiniment petit, de $\varphi(z)-1$, et par suite de $\psi(z)$, $\varphi(z)$ étant la fonction caractéristique de la loi considérée et $\psi(z)$ son logarithme. Cette loi est donc une loi $\mathcal{L}_{\alpha}$ $\left(\text{réduite si } C=\frac{2}{\pi}\sin\frac{\pi}{2}\alpha\right)$; l'existence des lois $\mathcal{L}_{\alpha}$, et par suite celle des lois L_{α}, est ainsi établie pour les valeurs de α comprises entre 0 et 2.

12. En pratique, toutes les fois qu'ayant à mesurer une ou plusieurs grandeurs inconnues, on a obtenu des données surabondantes, on emploie des méthodes de compensation des erreurs faciles à justifier si l'on admet que les erreurs obéissent à la loi de Gauss. Il n'est pas sans intérêt d'observer que ces méthodes devraient être modifiées si les erreurs obéissaient à une loi L_{α} $(\alpha<2)$.

Pour faciliter la discussion qui va suivre, nous introduirons quelques notions préliminaires indispensables.

Nous appellerons *paramètre de précision*, ou simplement *paramètre*, pour une erreur dont la valeur probable est nulle, un nombre qui indique l'ordre de grandeur de l'erreur à laquelle on doit s'attendre (en valeur absolue). Si l'on arrive à préciser convenablement la définition de ce nombre, une méthode de mesure sera d'autant plus satisfaisante que le paramètre de l'erreur est plus petit.

Mais il est essentiel de se rendre compte de ce qu'il y a forcément d'arbitraire dans la définition d'un tel nombre. Imaginons deux expérimentateurs ayant à mesurer une même quantité, mais en vue d'applications différentes, et supposons que le premier

tienne surtout à éviter une erreur supérieure à un nombre ε (en valeur absolue), tandis que le second tiendra à ce que l'erreur ne dépasse pas une valeur ε' différente de ε; tel sera le cas s'ils sont candidats à deux examens différents et doivent être reçus, l'un si l'erreur ne dépasse pas ε, et l'autre si l'erreur ne dépasse pas ε'. Il est clair que s'ils ont à comparer deux méthodes de mesure, celle qui sera la meilleure pour le premier expérimentateur, parce qu'elle rendra maxima la probabilité d'une erreur inférieure à ε, pourra ne pas l'être pour le second. L'appréciation de la valeur d'une méthode de mesure est donc essentiellement relative, et il est vain de vouloir définir par un seul nombre, qu'on appellerait paramètre, la valeur d'une méthode de mesure.

On considère souvent dans ce but l'erreur quadratique moyenne m. Bien des raisons peuvent être données pour justifier le rôle que l'on fait jouer à cette quantité, et il semble qu'en effet on ne puisse pas choisir un meilleur paramètre; mais il y a nécessairement tout de même dans ce choix quelque chose d'arbitraire, puisque nous avons vu que deux observateurs différents peuvent avec raison avoir des opinions différentes sur la valeur relative de deux méthodes de mesure. La seule chose que l'on puisse dire est que pour n'importe quel observateur l'inconvénient d'une erreur x, pour les erreurs d'un signe déterminé, peut être défini par une fonction $g(x)$ croissant avec $|x|$; on est alors conduit à considérer comme la meilleure la méthode qui rend minima l'intégrale

$$\int_{-\infty}^{+\infty} g(x)\,dF(x),$$

$F(x)$ étant la fonction des probabilités totales. Il est assez naturel d'admettre, par raison de symétrie, que la fonction $g(x)$ est paire, et par raison d'homogénéité que $f(x) = |x|^{\beta}$; mais déjà cela est arbitraire et notre expérimentateur qui tient surtout à ce que l'erreur ne dépasse pas en valeur absolue un nombre déterminé ε n'admettrait pas cette conclusion. Pour justifier le choix de l'erreur quadratique moyenne pour définir la valeur d'une méthode de mesure, il faut admettre de plus que $\beta = 2$, de sorte que $g(x) = x^2$. Quelle que soit la valeur pratique des considérations qui conduisent à ce résultat, comment ne pas voir ce qu'elles ont nécessairemeut d'ar-

bitraire? D'ailleurs, dans le cas des lois L_α, l'erreur quadratique moyenne est infinie; la méthode est en défaut pour la comparaison de deux lois de cette nature, et elle conduit à considérer une mesure où l'erreur obéit à une telle loi comme moins bonne que n'importe quelle mesure où l'erreur quadratique moyenne est finie, ce qui dans certains cas est manifestement contraire au bon sens.

La seule remarque que l'on puisse faire sans rien admettre d'arbitraire, est que, si plusieurs erreurs obéissent à des lois réductibles à une même loi par un changement d'unité, on peut prendre pour paramètres les nombres par lesquels il faut diviser ces erreurs pour que tous les quotients obéissent à la même loi. Ces paramètres sont ainsi définis à un même facteur près, facteur sans doute arbitraire, mais sans importance tant qu'on ne compare qu'entre elles les erreurs considérées.

Si plusieurs mesures indépendantes d'une même quantité donnent des valeurs $X_1, X_2, \ldots, X_p$, on peut considérer comme une nouvelle méthode de mesure celle qui consiste à faire la moyenne de ces différentes valeurs; soit X le résultat ainsi obtenu. Il est naturel d'ailleurs d'attribuer un plus grand poids à celles des mesures dans lesquelles on a le plus de confiance; pour préciser jusqu'à quel point on avantagera certaines mesures, on devra chercher pour quel système de valeur de ces poids l'erreur commise en adoptant la moyenne X ainsi calculée a le plus petit paramètre. Nous sommes conduits ainsi à la notion de *poids* relatifs de différentes mesures. Mais cette notion, reposant sur celle de paramètre, a le même caractère arbitraire.

Il n'y a qu'un cas dans lequel on échappe à ce caractère arbitraire; c'est le cas où les différentes erreurs obéissent à des lois réductibles par des changements d'unité à une même loi *stable*. Alors il en est de même de l'erreur obtenue en faisant la moyenne X et il n'y a plus rien d'arbitraire dans la comparaison des différents systèmes de poids que l'on peut adopter pour faire la moyenne; la notion du meilleur poids à adopter est alors très claire.

On sait que, si la loi considérée est celle de Gauss, les poids que l'on est ainsi conduit à donner aux différentes mesures sont inversement proportionnels aux carrés des paramètres. Pour les lois L_α, α étant compris entre 1 et 2, les poids sont inversement propor-

tionnels aux puissances des paramètres d'exposant $\frac{\alpha}{\alpha - 1}$. Le résultat reste vrai à la limite pour $\alpha = 1$, ce qui indique qu'il ne faut tenir compte que des mesures les plus précises; si plusieurs mesures sont également précises, leurs poids peuvent être choisis arbitrairement, et la moyenne ainsi formée constitue une mesure ayant la même précision que chaque mesure partielle. Si $\alpha < 1$, on arrive à un résultat plus paradoxal encore; non seulement il ne faut tenir compte que des mesures les plus précises, mais si n mesures sont également précises, il vaut mieux se contenter de la première mesure faite que de prendre la moyenne arithmétique de ces mesures.

La somme des n erreurs a en effet, d'après la formule (8), un paramètre $N = n^{\frac{1}{\alpha}}$ fois plus grand que celui des erreurs partielles, et, par suite, leur moyenne a un paramètre $\frac{N}{n}$ fois plus grand, et comme $N > n$ si $\alpha < 1$, on voit que le calcul de la moyenne diminue la précision.

Ce résultat paradoxal tient à ce que, sur n mesures, il y aura fatalement, si n est grand, quelques mesures très erronées qui risqueront de fausser complètement la moyenne. Mais il n'est pas pour cela impossible d'utiliser le grand nombre des mesures pour augmenter la précision. La courbe des fréquences différant peu de celle des probabilités si n est grand, il est bien certain qu'en prenant un nombre tel que sur les n mesures faites il y ait autant de nombres plus grands que lui que de nombres plus petits, on a un procédé de mesure dont la précision augmente indéfiniment avec n. On peut aussi écarter, dans une proportion déterminée, les plus grands et les plus petits nombres trouvés, et prendre la moyenne des nombres conservés. C'est à mon avis ce qu'on devrait faire toutes les fois que l'on n'est pas sûr d'être dans les conditions qui justifient l'application de la loi de Gauss. Tel est le cas pour le réglage du tir de l'artillerie (du moins dans les conditions défectueuses de fabrication des munitions ayant existé pendant la guerre), parce que l'on est réduit à traiter comme étant des erreurs accidentelles des erreurs qui sont en réalité des erreurs systématiques que l'on ne sait pas corriger.

Dans le cas où l'erreur obéit à la loi L_α, une variante de la

méthode précédente consiste à ranger les nombres trouvés par ordre de grandeur, et faire leur moyenne en les affectant de poids convenables, qui réduisent l'importance des mesures extrêmes; on peut par exemple prendre les poids proportionnels aux coefficients du binome. Cette méthode diffère essentiellement de celle indiquée d'abord où chaque poids était défini *a priori*, d'après la confiance qu'inspirait la méthode de mesure; ici il dépend du résultat trouvé.

Ainsi les méthodes habituelles de compensation des erreurs, applicables avec quelques modifications dans le cas des lois L_α, α étant compris entre 1 et 2, doivent être complètement modifiées si α ne dépasse pas l'unité. On voit par là ce qu'il faut penser des théories qui prétendent justifier les méthodes classiques de compensation indépendamment de la loi de probabilité de l'erreur.

Ces théories ont d'ailleurs leur origine dans des travaux de Gauss et de Laplace. Mais leur fausseté aurait dû apparaître lorsqu'en 1853 Cauchy attira l'attention sur les lois L_α, et en particulier sur la loi L_1. Or Bienaymé, dans un Mémoire publié en 1867 dans le *Journal de Mathématiques*, se fait l'avocat de ces théories, refusant de croire que l'on puisse rencontrer en pratique, dans une bonne méthode de mesure, une erreur dont la valeur quadratique moyenne soit infinie, et telle que la moyenne de n mesures ne mérite pas plus de confiance que chacune d'elles.

L'argument est exact; mais si on l'examine d'un peu plus près, il se retourne contre la thèse de l'auteur. Pourquoi, en effet, ces circonstances paradoxales ne se présentent-elles pas dans les bonnes méthodes de mesure? Parce que, dans ces méthodes, toutes les erreurs systématiques étant corrigées, il ne peut y avoir d'erreur appréciable que par l'accumulation de plusieurs causes d'erreur dont chacune est négligeable. Mais alors l'erreur obéit à la loi de Gauss et c'est seulement dans ce cas que les méthodes de calcul classiques sont justifiées.

Sans doute est-il possible, tout en faisant les hypothèses suffisantes pour justifier l'application de la loi de Gauss, de ne pas se préoccuper de la loi de probabilité de l'erreur, et justifier directement l'application des méthodes de calculs classiques. Mais on ne peut considérer une pareille justification comme indépendante de la loi de Gauss. Suivre une pareille méthode serait d'ailleurs

masquer la nature des choses. On n'a rien compris à la théorie des erreurs si l'on n'a pas compris d'abord pourquoi les erreurs accidentelles obéissent à la loi de Gauss; ensuite les méthodes de calcul à employer en pratique s'obtiennent sans difficulté.

L'erreur de Bienaymé se retrouve dans des travaux plus récents, et même sous une forme plus grave.

Dans un Mémoire publié dans les *Mathematische Annalen* de 1915, MM. Baer et F. Bernstein cherchent à justifier les méthodes de calcul classiques en considérant comme des notions premières celles que nous avons définies plus haut sous les noms de *poids* et *paramètre de précision.* Ils donnent comme avantage de leur méthode le fait qu'elle soit indépendante de la loi de Gauss, et même de la notion de probabilité.

Nous savons ce qu'il faut penser du premier de ces prétendus avantages. Nous serons conduits à des conclusions plus sévères encore si nous discutons ce que vaut le second; il sera impossible de ne pas conclure à une véritable erreur de principe.

Considérer la notion de paramètre de précision comme intuitive, c'est admettre qu'on peut définir par un seul nombre les avantages d'une méthode de mesure. Cela se fait d'ailleurs couramment. On me dit : « Tel instrument de mesure donne le résultat au centimètre près. » Je comprends que je ne dois pas compter sur la précision du millimètre et que l'erreur dépassera rarement le centimètre; je considère ce langage comme admissible si l'on ne prétend pas donner une signification mathématique au paramètre qui définit la précision. Sous cette condition, on peut considérer la notion de paramètre comme intuitive, et elle est fondamentale au point de vue pratique; un expérimentateur doit savoir l'ordre de grandeur de l'erreur à laquelle il doit s'attendre. Mais si l'on veut donner la valeur exacte de ce paramètre, on retombe dans les difficultés signalées tout à l'heure. Lorsque l'on cherche à préciser une notion intuitive, il arrive souvent que l'on s'aperçoive que l'intuition est un excellent guide. Dans le cas qui nous occupe, la prétendue intuition, qui conduit à définir par un seul nombre la précision d'un instrument de mesure, est trompeuse. Pour peu qu'on réfléchisse, on sera conduit à penser que les différentes erreurs possibles se présentent avec des fréquences déterminées,

c'est-à-dire que la précision de la mesure ne peut être définie que par la donnée de la loi de probabilité de l'erreur. Il faut donc une fonction pour définir cette précision, et, comme nous l'avons déjà indiqué, ce n'est que par une grossière approximation que l'on peut en général définir par un nombre unique les propriétés essentielles de cette fonction.

C'est donc une grave erreur que de croire qu'on puisse se passer de la notion de probabilité, et s'il paraît assez naturel d'admettre que les notions de poids et de paramètres de précision ont un sens, et d'admettre de plus, par raison d'homogénéité, que le poids est inversement proportionnel à une puissance du paramètre (d'exposant β), les considérations exposées au début de ce numéro montrent que cela ne peut se justifier qu'en faisant l'hypothèse que l'erreur obéisse à une loi stable. Si c'est la loi de Gauss, $\beta = 2$; dans les autres cas $\beta > 2$. Comme il faut que $\beta = 2$ pour que l'on arrive à justifier les méthodes classiques de compensation des erreurs, MM. Baer et Bernstein introduisent un axiome qui revient à admettre que $\beta = 2$; c'est le plus artificiel de leurs axiomes. Les autres peuvent paraître intuitifs et il faut y regarder d'assez près pour s'apercevoir que dans ce cas l'intuition est trompeuse. Celui-là ne peut se justifier que par le désir d'arriver au résultat, et j'aimerais mieux que MM. Baer et Bernstein nous disent franchement : « Les méthodes classiques de compensation des erreurs sont bonnes, car nous considérons comme un axiome qu'elles le sont. »

Poincaré a critiqué à plusieurs reprises, dans ses ouvrages philosophiques, ceux qui croient avoir fait quelque chose lorsqu'ils ont fait reposer une démonstration sur une définition arbitraire qu'ils appellent axiome. La théorie ainsi construite ne vaut quelque chose que si des raisons de bons sens incontestables nous obligent à considérer ces axiomes comme exacts en pratique. La géométrie d'Euclide et celle d'Hilbert ne valent rien si l'on ne soumet d'abord à la critique du bon sens les axiomes qui sont à leur base, et les disciples d'Hilbert, séduits par la rigueur impeccable de ses raisonnements, mais oubliant l'importance de cette discussion des axiomes, ont souvent fait une œuvre, parfaite au point de vue logique, mais sans aucune valeur au point de vue de la justification des résultats qu'ils croient établir.

13. Nous allons maintenant compléter nos remarques sur la composition d'un grand nombre d'erreurs par l'étude de quelques cas ne rentrant pas dans ceux étudiés jusqu'ici.

Considérons d'abord le cas de n erreurs très petites, sur lesquelles n_1 obéissent à la loi L_α, et les n_2 autres à la loi L_β,

$$(n_1 + n_2 = n,\ 0 < \alpha < \beta \leqq 2),$$

et, en introduisant des coefficients $a_1, a_2, \ldots, a_n$ étudions la somme

$$x = a_1 x_1 + a_2 x_2 + \ldots a_n x_n.$$

Elle obéit à une loi de probabilité pour laquelle le logarithme de la fonction caractéristique est de la forme

$$\psi(z) = -|az|^\alpha - |bz|^\beta.$$

Si d'ailleurs les coefficients $a_1, a_2, \ldots, a_n$ sont tous du même ordre de grandeur et si n_1 et n_2 sont du même ordre de grandeur, a sera de l'ordre de grandeur de $n^{\frac{1}{\alpha}}$ et b sera de l'ordre de grandeur de $n^{\frac{1}{\beta}}$, de sorte qu'à la limite le second terme pourra être négligé, et, en divisant x par a, on obtiendra comme loi limite la loi L_α. Tel sera le cas général.

Mais on peut aussi s'arranger pour que, n augmentant indéfiniment, n_1^β et n_2^α soient du même ordre de grandeur. Alors a et b sont du même ordre de grandeur, et l'on peut s'arranger pour avoir à la limite une loi pour laquelle

$$\psi(z) = -A|z|^\alpha - B|z|^\beta, \tag{11}$$

A et B ayant telles valeurs positives que l'on voudra.

Il est à remarquer que la convergence vers cette limite de la loi de probabilité à laquelle obéit le quotient de x par un nombre convenable est une *semi-convergence ;* on n'obtiendrait pas la même limite en changeant l'ordre des termes; si, par exemple, ce changement de l'ordre était tel que n_1 et n_2 soient du même ordre de grandeur, nous savons qu'on obtiendrait à la limite la loi L_α. Il est bien évident d'ailleurs, d'une manière générale, que si l'addition d'un grand nombre d'erreurs très petites conduit à la limite à une loi qui n'est pas stable, comme c'est le cas pour celle définie

par la formule (11), la convergence vers cette loi ne peut être qu'une semi-convergence.

Si maintenant nous considérons la somme d'une infinité d'erreurs obéissant à des lois L_α d'indices

$$\alpha_1, \alpha_2, \ldots, \alpha_p,$$

chacun de ces indices correspondant à une infinité d'erreurs, le terme correspondant au plus petit de cet indices (soit α_1, pour fixer les idées) l'emportera en général sur les autres, et l'on obtiendra à la limite la loi L_{α_1}. Mais on peut aussi s'arranger pour avoir à la limite une loi pour laquelle

$$\psi(z) = -A_1|z|^{\alpha_1} - A_2|z|^{\alpha_2} - \ldots - A_p|z|^{\alpha_p},$$

les coefficients $A_1, A_2, \ldots, A_p$ étant positifs, et à cela près quelconques.

Ce résultat reste vrai si la suite des indices correspondant aux différentes erreurs est quelconque, et l'on peut ainsi obtenir à la limite n'importe quelle loi pour laquelle $\psi(z)$ soit de la forme

$$-\Sigma A_i|z|^{\alpha_i},$$

les A_i étant positifs et les α_i compris entre 0 et 2, mais la somme étant à cela près quelconque, et pouvant même se réduire à une intégrale définie (ordinaire ou de Stieltjes).

14. Les lois L_α, que nous avons étudiées jusqu'ici, sont caractérisées par le fait que $\psi(z)$, ou ce qui revient au même $\varphi(z) - 1$, a une valeur principale de la forme $-A|z|^\alpha$. Nous allons maintenant étudier des cas pour laquelle $\psi(z)$ n'est pas de cette forme. Supposons d'abord que l'on ait

$$\psi(z) = -A|z|^\alpha \left(\log \frac{1}{|z|}\right)^{-\beta} [1 + \omega(z)], \tag{12}$$

A désignant $\frac{1}{\Gamma(\alpha+1)}$, et $\omega(z)$ s'annulant avec z. La somme de n erreurs obéissant à une telle loi, divisée par un *coefficient de réduction* N tel que

$$N^\alpha (\log N)^\beta = n; \tag{13}$$

obéit à une loi pour laquelle le logarithme de la fonction caractéristique est

$$-nA\left|\frac{z}{N}\right|\left(\log\frac{N}{|z|}\right)^{-\beta}\left[1+\omega\left(\frac{z}{N}\right)\right]\sim -A|z|^{\alpha},\qquad (n\to\infty),$$

et qui par suite tend vers la loi L_α. Ainsi l'introduction du facteur logarithmique change seulement l'ordre de grandeur du coefficient de réduction, mais non la forme de la loi obtenue à la limite.

Introduisons maintenant une fonction $\lambda(x)$, paire, et telle que $x\frac{\lambda'(x)}{\lambda(x)}$ tende vers zéro pour x infini. Une telle fonction est à croissance *lente et régulière;* d'une manière précise, elle croît ou décroît plus lentement que n'importe quelle puissance de x, et $\lambda(x)$ et $\lambda(kx)$, k étant constant, sont équivalents pour x infini. On vérifie aisément que la substitution de $\lambda\left(\frac{1}{z}\right)$ à $\log\frac{1}{|z|}$ ne change rien au calcul qui précède, de sorte que toute loi pour laquelle on ait, pour z infiniment petit,

$$\psi(z)\sim -A|z|^{\alpha}\lambda\left(\frac{1}{z}\right) \tag{14}$$

appartient au domaine d'attraction de la loi L_α.

D'ailleurs, ces lois existent effectivement. Pour que la formule (13) soit vérifiée, il suffit, du moins si la croissance de $\lambda(x)$ est assez régulière, que la loi considérée soit définie par une fonction des probabilités élémentaires $f(x)$, telle que l'on ait, pour x infini,

$$f(x)\sim\frac{1}{\pi}\sin\frac{\pi}{2}\alpha\frac{\lambda(x)}{|x|^{\alpha+1}}. \tag{15}$$

Ce résultat se rattache à cette circonstance générale que l'allure de $\varphi(z)$ à l'origine dépend de l'allure de $f(x)$ à l'infini, une modification de $f(x)$ à distance finie ayant pour effet d'ajouter à $\varphi(z)$ une fonction entière. On peut préciser cette relation soit par une méthode analogue à celle du n° 11, soit par une méthode d'un caractère plus transcendant, reposant sur une généralisation de la notion de dérivée, indiquée dans ma Note du 22 mai 1923 et développée dans un Mémoire à l'impression dans le *Bulletin des Sciences mathématiques* (¹). Je n'insisterai pas davantage sur

(¹) C'est par cette méthode que j'ai d'abord établi le résultat du n° 11. C'est

cette question, n'ayant pas jusqu'ici réussi à démontrer d'une manière suffisamment simple que la formule (15) entraîne la formule (14).

15. Les considérations qui précèdent s'étendent avec quelques modifications au cas où $\alpha = 0$. Il existe des lois de probabilité pour lesquelles $\psi(z)$ s'annule avec z moins rapidement que n'importe quelle puissance de z, étant par exemple de la forme (12), avec $\alpha = 0$ et $\beta = 1$. En prenant alors le coefficient de réduction $N = e^n$ qui résulte de la formule (13), on trouve à la limite, pour le quotient par N de la somme de n erreurs obéissant à la loi considérée, la loi L_0, c'est-à-dire la loi pour laquelle $\psi(z) = 0$. La variable est nécessairement nulle; ce n'est pas réellement une loi de probabilité.

On peut d'abord être tenté de croire que cette circonstance est due à ce qu'on a pris pour N une fonction trop rapidement croissante de n. Il n'en est rien. En prenant pour N une fonction moins rapidement croissante, on ne peut pas trouver une loi limite déterminée.

Il est facile de s'expliquer cette circonstance. Considérons p nombres $\alpha_1, \alpha_2, \ldots, \alpha_p$ croissant de 0 à 1, et désignons par $\xi_1, \xi_2, \ldots, \xi_p$ des nombres tels que α_i représente la probabilité pour que la somme de n erreurs ne dépasse pas ξ_i en valeur absolue. En général, n augmentant indéfiniment, les différents nombres ξ_i augmentent indéfiniment, et sont du même ordre de grandeur; alors l'erreur totale X, divisée par une fonction de n ayant le même ordre de grandeur, obéit à une loi tendant vers une limite déterminée. Mais dans le cas considéré, les différents nombres ξ_i sont d'ordres de grandeur différents; chacun des rapports $\frac{\xi_{i+1}}{\xi_i}$ augmente indéfiniment avec n, et le rapport $\frac{|X|}{\xi_i}$ obéit à une loi pour laquelle il y a à la limite deux valeurs possibles seulement, 0 et ∞, de probabilités respectives α_i et $1 - \alpha_i$. Si, comme on le fait en général, on choisit le coefficient de réduction de manière à

à la suite d'une lettre de M. G. Pólya, me communiquant son impression que l'introduction de la dérivée généralisée n'était pas nécessaire, que j'ai cherché une démonstration plus élémentaire et trouvé celle adoptée au n° 11.

rendre nulle la probabilité des valeurs infinies, on arrive nécessairement à cette conclusion que la seule valeur possible à la limite est zéro. Il est même impossible de choisir le coefficient de réduction de manière que la probabilité que $\frac{|X|}{N}$ soit compris entre deux nombres positifs fixes ne tende pas vers zéro.

16. Il reste à étudier le cas où la fonction $\psi(z)$, ou ce qui revient au même $\varphi(z)-1$, est à décroissance irrégulière lorsque z est infiniment petit. Nous considérons seulement le cas où $\psi(z)$ est de la forme $-|z|^{\alpha}\psi_1(z)$, la fonction $\psi_1(z)$ variant une infinité de fois entre deux nombres positifs.

Cette question est liée à un problème posé par M. G. Pólya; il s'agit de trouver les lois de probabilités telles que, x_1 et x_2 étant deux erreurs indépendantes obéissant à une telle loi, on puisse trouver des nombres a_1, a_2 et A (définis seulement à un facteur commun près), tels que $\frac{a_1x_1+a_2x_2}{A}$ obéisse à la même loi. Nous dirons qu'une telle loi est *semi-stable*. Le problème diffère de celui de la recherche des lois stables parce que dans ce cas a_1 et a_2 sont quelconques. Dans le cas des lois semi-stables, la propriété indiquée, qui s'exprime à notre point de vue par la relation

$$\psi(a_1z)+\psi(a_2z)=\psi(Az), \tag{16}$$

n'a lieu que pour des valeurs convenables de a_1 et a_2. On peut donc concevoir qu'une loi soit semi-stable sans être stable.

Nous supposerons pour simplifier $a_1=a_2$; on se rend aisément compte que cette restriction n'est pas essentielle. On vérifie alors l'équation (16) en prenant

$$\psi(z)=-|z|^{\alpha}\psi_1\left(\frac{\alpha\log|z|}{\log 2}\right), \tag{17}$$

la fonction ψ_1 étant uue fonction périodique de période égale à l'unité. Si l'on considère la somme de n erreurs obéissant à la loi définie par une telle fonction, divisée par le coefficient de réduction $N=n^{\frac{1}{\alpha}}$, on constate qu'elle obéit à une loi de probabilité qui dépend seulement de la partie entière de $\frac{\log n}{\log 2}$. On n'a pas une loi limite déterminée, mais une infinité de lois limites, pour les-

quelles

$$\psi(z) = -|z|^{\alpha}\psi_1\left(\frac{\alpha \log k|z|}{\log 2}\right),$$

k étant un coefficient positif arbitraire. Il en résulte qu'une loi semi-stable ne peut pas appartenir au domaine d'attraction d'une loi stable (1).

Mais ces lois semi-stables existent-elles effectivement? Nous savons que $\frac{\psi(z)}{z^2}$ tend, pour $z=0$, vers une limite déterminée, finie ou infinie, ce qui exclut la possibilité de la formule (17) si $\alpha = 2$. Mais ce raisonnement ne s'applique pas si $\alpha < 2$ et M. G. Pólya a montré que les lois semi-stables existent effectivement, du moins si $\alpha < 1$. Il résulte de sa démonstration qu'elles dépendent même d'une fonction arbitraire (2).

Tenant compte de ce résultat, nos conclusions seront les suivantes : il existe des lois stables, dépendant d'un paramètre si l'on ne considère que les lois symétriques, mais de deux paramètres si l'on considère les lois dissymétriques. Chacune de ces lois a un domaine d'attraction contenant des lois dépendant d'une fonction arbitraire. Il existe des lois n'appartenant à aucun de ces domaines d'attraction, et dépendant aussi d'une fonction arbitraire.

---

(1) Ce résultat est d'ailleurs bien évident. Sans préciser la forme de $\psi(z)$ on remarque, en supposant toujours a_1 et a_2 égaux, que la loi de probabilité à laquelle obéit la somme de 2^h erreurs obéissant à la loi considérée, ne se distinguant de celle-ci que par un changement d'unité, ne peut tendre vers une loi d'un type différent.

(2) *Voir* G. Pólya, *Herleitung des Gausschen Fehlergesetzes aus einer Funktionaleichung* (*Math. Zeitschrift*, Mémoire daté du 9 août 1922).