

ANNALES DE L'I. H. P.

G. PÓLYA

Sur quelques points de la théorie des probabilités

Annales de l'I. H. P., tome 1, n° 2 (1930), p. 117-161

http://www.numdam.org/item?id=AIHP_1930__1_2_117_0

© Gauthier-Villars, 1930, tous droits réservés.

L'accès aux archives de la revue « Annales de l'I. H. P. » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

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

Sur quelques points de la théorie des probabilités

PAR

G. PÓLYA

Les pages qui suivent reproduisent, abstraction faite de quelques altérations, les Leçons que j'ai eu l'honneur de faire à l'Institut Henri Poincaré en mars 1929. Je diviserai ici les matières en deux chapitres. Dans le premier, il s'agit d'une propriété caractéristique de la loi de Gauss et dans le second de quelques autres lois de fréquence. Ces dernières ressemblent à la loi de Gauss en ce qu'elles résultent, elles aussi, de la superposition d'un très grand nombre de petits effets fortuits, mais elles diffèrent de la loi de Gauss en ce que les effets fortuits dont elles résultent ne sont pas indépendants. Dans le premier chapitre, je donnerai les démonstrations en détail. Dans le second je m'efforcerai surtout de présenter les problèmes d'une manière intuitive et, quant aux démonstrations, je me bornerai quelquefois à une indication de la marche à suivre.

Qu'il me soit permis d'exprimer ici tous mes remerciements à la direction de l'Institut Henri Poincaré et en particulier à M. Émile Borel de l'honneur de leur invitation qui m'a permis d'exposer mes remarques devant un auditoire d'élite.

I. — Sur une propriété caractéristique de la loi de Gauss

I

Envisageons des mesures pour lesquelles la probabilité d'une erreur ne dépend que de la valeur numérique de l'erreur. Désignons par $\varphi(x)$ la loi d'erreurs ; c'est-à-dire que la probabilité qu'une erreur tombe entre les limites a et b est donnée par

$$\int_a^b \varphi(x) dx.$$

Supposons que n observations indépendantes d'une grandeur physique dont la « vraie valeur » est l ont donné les valeurs

$$l_1, l_2, l_3, \dots, l_n.$$

Les erreurs commises sont $l - l_1, l - l_2, \dots, l - l_n$ et la probabilité de les avoir rencontrées ensemble est proportionnelle à

$$\varphi(l - l_1) \varphi(l - l_2) \dots \varphi(l - l_n).$$

En réalité on ne connaît pas l et nous devons trouver une valeur plausible de l en connaissant les valeurs l_1, l_2, \dots, l_n fournies par l'observation et en supposant connue la loi d'erreurs $\varphi(x)$. Il est naturel de prendre comme point de départ la distribution de probabilités dont la densité au point x est proportionnelle à

$$(I) \quad \varphi(x - l_1) \varphi(x - l_2) \dots \varphi(x - l_n).$$

J'admets que nous étions d'accord jusqu'ici. Mais maintenant il s'agit de nous décider entre plusieurs chemins possibles.

a) On peut choisir comme valeur plausible de la grandeur mesurée *la valeur la plus probable*, c'est-à-dire la valeur $x = l'$ qui rend (I) maximum. C'est le choix fait par GAUSS.

b) On peut aussi choisir *la valeur probable*, c'est-à-dire la valeur l'' déterminée par l'équation

$$\int_{-\infty}^{+\infty} (x - l'') \varphi(x - l_1) \varphi(x - l_2) \dots \varphi(x - l_n) dx = 0.$$

SUR QUELQUES POINTS DE LA THÉORIE DES PROBABILITÉS

C'est ce que BERTRAND a fait remarquer ⁽¹⁾ en critiquant le choix de GAUSS.

c) On pourrait choisir aussi la *valeur médiane*, c'est-à-dire la valeur l''' déterminée par

$$\int_{-\infty}^{l'''} \varphi(x - l_1) \cdots \varphi(x - l_n) dx = \int_{l'''}^{+\infty} \varphi(x - l_1) \cdots \varphi(x - l_n) dx.$$

On pourrait encore choisir les abscisses d'autres points remarquables de la fonction (I), mais arrêtons-nous ici.

POINCARÉ, en critiquant la critique de BERTRAND, a remarqué que les deux chemins *a*) et *b*) mènent au même but, si la loi d'erreurs $\varphi(x)$ est celle de GAUSS ⁽²⁾. C'est une propriété très commode de la loi de GAUSS : elle nous épargne une hésitation pénible. On est ainsi amené à se demander *quelle est la loi d'erreurs la plus générale qui donne*

$$l' = l'',$$

l_1, l_2, \dots, l_n étant quelconques ?

Voilà la question à laquelle je tâcherai de répondre dans ce qui suit ⁽³⁾.

II

Nous demandons de la loi d'erreurs qu'elle nous fournisse le même résultat par le chemin *a*) et par le chemin *b*), quels que soient les nombres donnés l_1, l_2, \dots, l_n . C'est une chose bien compliquée, tâchons d'en démêler les postulats essentiels.

Tout d'abord, la méthode *a*) doit être applicable. Nous posons donc comme premier postulat :

(I) *Etant donnés des nombres réels quelconques l_1, l_2, \dots, l_n , la fonction (I) atteint son maximum pour une seule valeur de x .*

Considérons maintenant $n\phi$ mesures

$$l_1, l_1, \dots, l_1; l_2, l_2, \dots, l_2; \dots, l_n, l_n, \dots, l_n$$

où chacune des valeurs l_1, l_2, \dots, l_n est répétée ϕ fois. En cherchant la

(1) J. BERTRAND, Calcul des probabilités, 2^e éd. (Paris, 1907). Voir p. 171-172.

(2) H. POINCARÉ, Calcul des probabilités, 2^e éd. (Paris, 1912). Voir p. 237-240 et aussi p. 176 et les suivantes.

(3) La réponse est formulée au n^o VI. J'ai communiqué le résultat au Congrès international de Bologne, le 5 septembre 1928. Voir les Actes de ce Congrès.

valeur plausible pour cette combinaison de mesures, il faut d'après la méthode *a*) chercher le maximum du produit (I) élevé à la *p*-me puissance. La situation de ce maximum *ne dépend pas de p*. Donc le résultat fourni par la méthode *b*) doit être, lui aussi, indépendant de *p* ; prenons cela comme second postulat :

(II) *Etant donnés des nombres réels quelconques* l_1, l_2, \dots, l_n , *l'équation*

$$(2) \quad \int_{-\infty}^{+\infty} (x - l) [\varphi(x - l_1) \varphi(x - l_2) \cdots \varphi(x - l_n)]^p dx = 0$$

donne la même valeur pour l quel que soit l'entier positif p.

J'ajoute encore les hypothèses supplémentaires que voici :

(I') *L'abscisse du maximum de* $\varphi(x)$ *divise l'axe des x en deux parties ; dans chacune d'elles* $\varphi(x)$ *est monotone.*

(III) *La fonction* $\varphi(x)$ *possède une dérivée seconde continue,*
 $\varphi(x) > 0$ *et l'intégrale* $\int_{-\infty}^{+\infty} x\varphi(x)dx$ *existe.*

Le postulat que $\varphi(x)$ ne possède qu'un seul maximum absolu est contenu en (I), comme cas particulier ; (I') demande encore que $\varphi(x)$ ne possède qu'un seul maximum *relatif*. (III) demande en somme une certaine régularité de $\varphi(x)$. Il est bien possible que les postulats (I') et (III) puissent être remplacés par d'autres moins restrictifs sans nuire au résultat final tandis que (I) et (II) me paraissent essentiels.

Notre but est de chercher la fonction $\varphi(x)$ la plus générale satisfaisant à (I), (I'), (II) et (III).

III

J'arrive à deux lemmes dont nous aurons besoin plus tard.

Lemme A. — Supposons que $f(x)$ *et* $g(x)$ *soient des fonctions continues, que* $g(\xi) > 0$ *et que* $f(x)$ *atteigne son maximum au point* ξ *et nulle part ailleurs, que* $f(x) \geq 0$ *et* $f(x)$ *tende vers 0 pour* $x \rightarrow \pm \infty$ *et enfin que l'intégrale*

$$\int_{-\infty}^{+\infty} |g(x)| f(x) dx$$

soit convergente. Alors

$$\lim_{n \rightarrow \infty} \sqrt[n]{\int_{-\infty}^{+\infty} g(x) f(x)^n dx} = f(\xi).$$

Ce qui est essentiel dans ce lemme est bien connu. Les hypothèses entraînent l'existence de trois nombres positifs ε , α et β , $\alpha < \beta$, tels que

$$\begin{aligned} f(x) < f(\xi) - 2\varepsilon & \quad \text{pour} \quad |x - \xi| > \beta, \\ g(x) > 0 & \quad \text{pour} \quad |x - \xi| < \beta, \\ f(x) > f(\xi) - \varepsilon & \quad \text{pour} \quad |x - \xi| < \alpha. \end{aligned}$$

On déduit le lemme A de l'inégalité double

$$\begin{aligned} (f(\xi) - \varepsilon)^n \int_{\xi - \alpha}^{\xi + \alpha} g(x) dx - (f(\xi) - 2\varepsilon)^{n-1} \left(\int_{-\infty}^{\xi - \beta} + \int_{\xi + \beta}^{+\infty} \right) |g(x)| f(x) dx \\ \leq \int_{-\infty}^{+\infty} g(x) f(x)^n dx \leq f(\xi)^{n-1} \int_{-\infty}^{+\infty} |g(x)| f(x) dx. \end{aligned}$$

Lemme B. — Admettons que la fonction $f(x)$ est continue, que $0 < f(x) < f(\xi)$ pour $(x - \xi)^2 > 0$ et que $f(x)$ est croissante pour $x < \xi$ et décroissante pour $x > \xi$. Alors on peut conclure de

$$(3) \quad \int_{-\infty}^{+\infty} x f(x)^p dx = 0 \quad \text{pour} \quad p = 1, 2, 3, \dots$$

que $f(x)$ est une fonction paire.

Avant de commencer la démonstration observons que de la seule hypothèse (3) on ne peut pas conclure que $f(-x) = f(x)$ comme le montre l'exemple de la fonction $f(x)$ qui est égale à 1 dans les trois intervalles $(-8, -5)$, $(1, 4)$ et $(5, 7)$ et égale à 0 partout ailleurs : elle satisfait à (3) sans être paire.

L'hypothèse faite sur la forme de la courbe $y = f(x)$ et l'existence de l'intégrale mentionnée dans (3) ($p = 1$ suffit ici) entraînent $f(x) \rightarrow 0$ pour $x \rightarrow \pm \infty$.

Si l'abscisse ξ mentionnée dans le lemme B n'était pas 0, on pourrait déduire du lemme A que le premier membre de (3) a le signe de ξ pour n suffisamment grand, ce qui contredirait l'hypothèse. Donc $\xi = 0$, le maximum de $f(x)$ est $f(0)$.

Soient donnés les nombres y et ε ,

$$(4) \quad 0 < \varepsilon < y < f(0).$$

Déterminons la fonction $\Psi(t) = \Psi(t; y, \varepsilon)$ en posant

$$t \Psi(t; y, \varepsilon) = \begin{cases} 0 & \text{pour } 0 \leq t \leq y - \varepsilon \\ 1 & \text{pour } y \leq t \leq f(0) \end{cases}$$

et en supposant que $t\Psi(t)$ est linéaire pour $y - \varepsilon \leq t \leq y$. D'après le théorème de WEIERSTRASS sur l'approximation des fonctions continues on peut déterminer des polynômes $\mathfrak{P}_1(t), \mathfrak{P}_2(t), \dots$

$$\mathfrak{P}_n(t) = a_{n1} + a_{n2}t + a_{n3}t^2 + \dots + a_{nn}t^{n-1}$$

tels qu'on ait uniformément pour $0 \leq t \leq f(0)$.

$$(5) \quad \lim_{n \rightarrow \infty} \mathfrak{P}_n(t) = \Psi(t).$$

En utilisant l'hypothèse (3) pour $p = 1, 2, \dots, n$ on obtient

$$(6) \quad \int_{-\infty}^{+\infty} x f(x) \mathfrak{P}_n(f(x)) dx = \int_{-\infty}^{+\infty} x \{ a_{n1}f(x) + a_{n2}f(x)^2 + \dots + a_{nn}f(x)^n \} dx = 0.$$

Les polynômes $\mathfrak{P}_n(t)$ étant uniformément bornés pour $0 \leq t \leq f(0)$, l'intégrale (6) a une majorante de la forme

$$(7) \quad \int_{-\infty}^{+\infty} |x| f(x) l dx,$$

où l est une constante positive. On conclut ainsi de (5) et de (6) que

$$(8) \quad \int_{-\infty}^{+\infty} x f(x) \Psi(f(x); y, \varepsilon) dx = 0.$$

SUR QUELQUES POINTS DE LA THÉORIE DES PROBABILITÉS

Cette dernière intégrale est encore majorée par (7) (avec $l = 1$) si on y fait tendre ε vers 0. Désignons par x_1 la plus petite et par x_2 la plus grande racine de l'équation

$$f(x) = y.$$

Ces racines existent d'après (4) et puisque la fonction continue $f(x)$ tend vers 0 quand $x \rightarrow \pm \infty$; on voit aussi que

$$(9) \quad x_1 < 0 < x_2.$$

D'après l'hypothèse faite sur la forme de la courbe $y = f(x)$

$$f(x) \geq y \quad \text{pour} \quad x_1 \leq x \leq x_2$$

et $f(x) < y$ si x est extérieur à l'intervalle (x_1, x_2) . On obtient donc

$$0 = \lim_{\varepsilon \rightarrow 0} \int_{-\infty}^{+\infty} x f(x) \Psi(f(x); y, \varepsilon) dx = \int_{x_1}^{x_2} x \cdot 1 \cdot dx = x_2^2 - x_1^2.$$

Tenant compte de (9), nous avons

$$x_1 = -x_2$$

et puisque c'est vrai pour un y quelconque entre 0 et $f(0)$ la fonction $f(x)$ est paire.

IV

Le lemme B étant démontré, retournons à la recherche des fonctions $\varphi(x)$ satisfaisant aux postulats (I), (I'), (II) et (III).

Observons qu'une substitution linéaire opérée sur x est sans influence sur les postulats (I)-(III). En changeant x en $x + a$, où a est une constante facile à déterminer, on peut donc supposer que

$$(10) \quad \int_{-\infty}^{+\infty} x \varphi(x) dx = 0.$$

D'après l'hypothèse fondamentale (II), on conclut de (10) que

$$\int_{-\infty}^{+\infty} x \varphi(x)^p dx = 0 \quad \text{pour} \quad p = 1, 2, 3, \dots$$

Ces équations, en nombre infini, entraînent en vertu du lemme B — dont les autres conditions sont remplies d'après (I') et (III) — que $\varphi(x)$ est paire

$$(II) \quad \varphi(-x) = \varphi(x).$$

Soit donné un nombre c quelconque. Observons qu'en vertu de (II) la fonction

$$(I2) \quad \begin{aligned} \varphi(x-c)\varphi(x+c) &= \varphi(-x+c)\varphi(-x-c) \\ &= \varphi(-x-c)\varphi(-x+c) \end{aligned}$$

est paire. Elle n'atteint son maximum qu'une fois, d'après l'hypothèse fondamentale (I), elle doit donc l'atteindre au point $x = 0$. On a ainsi, en faisant usage encore une fois de (II)

$$(I3) \quad \varphi(c-x)\varphi(c+x) < \varphi(c)^2 \quad \text{pour} \quad x^2 > 0.$$

En posant

$$(I4) \quad \begin{aligned} c-x &= x_1, & c+x &= x_2 \\ \log \varphi(x) &= \Psi(x) \end{aligned}$$

et en observant que x et c étant arbitraires x_1 et x_2 le sont aussi, nous pouvons exprimer (I3) par

$$(I5) \quad \frac{\Psi(x_1) + \Psi(x_2)}{2} < \Psi\left(\frac{x_1 + x_2}{2}\right) \quad \text{pour} \quad x_1 \neq x_2.$$

(I5) exprime que la courbe $y = \Psi(x)$ est partout *concave vue d'en bas*.

Nous obtenons de (I5) que, pour $x_1 \neq x_2$,

$$\frac{1}{2} \left\{ \sum_{v=1}^n \Psi(x_1 - l_v) + \sum_{v=1}^n \Psi(x_2 - l_v) \right\} < \sum_{v=1}^n \Psi\left(\frac{x_1 + x_2}{2} - l_v\right)$$

ce qui exprime que la courbe

$$(I6) \quad y = \Psi(x - l_1) + \Psi(x - l_2) + \dots + \Psi(x - l_n),$$

elle aussi, est partout concave vue d'en bas. En particulier, nous obtenons ainsi, que cette courbe est *croissante à gauche et décroissante à droite de son maximum unique*. Observons que nous savons maintenant davantage que nous n'avons postulé au commencement en posant (I). Nous savons maintenant assez de l'allure de la fonction (I)

pour rendre applicable le lemme B à l'équation (2) valable pour $p = 1, 2, 3, \dots$. Nous obtenons que la fonction (1) est une fonction paire de $(x - l)$.

En somme, nous avons obtenu la propriété suivante des fonctions $\varphi(x)$ satisfaisant aux postulats (I), (I'), (II) et (III) : *Etant donnés des nombres réels quelconques l_1, l_2, \dots, l_n il existe un nombre l , tel que $\varphi(x - l_1) \varphi(x - l_2) \dots \varphi(x - l_n)$ soit une fonction paire de $(x - l)$ et une fonction décroissante de $(x - l)^2$.*

Si une loi d'erreurs $\varphi(x)$ jouit de cette propriété et l_1, l_2, \dots, l_n sont des mesures, elle fournit évidemment la même valeur avec les deux méthodes a) et b) expliquées au n° 1 : la courbe $y = \varphi(x - l_1) \dots \varphi(x - l_n)$ est symétrique par rapport à l'axe vertical $x = l$ qui contient aussi bien son maximum unique que son centre de gravité. Maintenant nous pouvons être assurés que les postulats (I), (II), (I') et (III) embrassent toute la question du n° 1, ce qui n'était pas évident *a priori*.

Il y a plus. Une loi d'erreurs $\varphi(x)$ jouissant de la dite propriété de symétrie fournira la même valeur aussi avec la méthode c) ou avec toute autre méthode raisonnable.

Les déductions suivantes qui sont encore nécessaires pour déterminer la forme des fonctions $\varphi(x)$ cherchées s'appuieront sur la propriété de symétrie que nous venons d'établir.

V

Jusqu'ici nous n'avons pas fait usage de la première partie de l'hypothèse (III) concernant les dérivées de $\varphi(x)$. Il serait désirable de continuer le raisonnement sans faire appel à ces dérivées ; cela conviendrait mieux au caractère approximatif de la théorie des probabilités qu'on ne devrait pas perdre des yeux en faisant des déductions, comme M. BOREL, a si justement fait remarquer. Toutefois nous ferons usage des dérivées de $\varphi(x)$ dans ce qui suit, parce que c'est plus commode.

Occupons-nous de la fonction

$$\Psi(x) = \log \varphi(x).$$

En vertu des hypothèses réunies sous (III), $\Psi'(x)$ et $\Psi''(x)$ existent et sont continues. L'inégalité (15) entraîne, comme on voit facilement, que $\Psi'(x)$ est une fonction décroissante au sens stricte, le cas où $\Psi'(x)$

est constante dans un intervalle est exclu comme le signe d'égalité dans (15). On conclut de (11), (14) que

$$(17) \quad \Psi'(-x) = -\Psi'(x)$$

et puis, $\Psi'(x)$ étant strictement monotone,

$$(18) \quad \Psi'(x) < 0 \quad \text{pour} \quad x > 0.$$

Soient donnés un nombre positif y et un entier m supérieur à 1. Considérons la fonction de x

$$\Psi'(-y) + m\Psi'(x) = h(x).$$

En vertu de (17) et de (18)

$$h(0) = \Psi'(-y) > 0, \quad h(y) = (m-1)\Psi'(y) < 0.$$

Donc il existe un η

$$0 < \eta < y$$

tel que $h(\eta) = 0$; donc

$$(19) \quad \Psi'(-y) + m\Psi'(\eta) = 0.$$

C'est-à-dire que la dérivée de la fonction de x

$$(20) \quad \varphi(x-y) \varphi(x+\eta)^m$$

s'annule pour $x = 0$. Mais puisque $\Psi'(x)$ est strictement monotone, la dérivée de (20) ne peut s'annuler qu'une fois. L'abscisse $x = 0$ est donc pour la fonction (20) l'abscisse de symétrie dont l'existence a été établie au n° IV, la fonction (20) est paire. On a donc en faisant usage de (11)

$$(21) \quad \varphi(x-y) \varphi(x+\eta)^m = \varphi(-x-y) \varphi(-x+\eta)^m \\ = \varphi(x+y) \varphi(x-\eta)^m.$$

En éliminant m entre (19) et (21) et en utilisant (17) on obtient

$$\frac{\Psi'(y)}{\Psi'(\eta)} \log \frac{\varphi(x+\eta)}{\varphi(x-\eta)} = \log \frac{\varphi(x+y)}{\varphi(x-y)}$$

ce qu'on peut écrire

$$(22) \quad 2\Psi'(y) \frac{\Psi(x+\eta) - \Psi(x-\eta)}{2\eta} = [\Psi(x+y) - \Psi(x-y)] \frac{\Psi'(\eta)}{\eta}.$$

Regardons y comme constante et faisons tendre l'entier m vers ∞ .

|

SUR QUELQUES POINTS DE LA THÉORIE DES PROBABILITÉS

On conclut de (19) d'abord que $\Psi'(\eta) \rightarrow 0$, puis, $\Psi'(x)$ étant strictement monotone, que $\eta \rightarrow 0$. En définitive (22) nous fournit pour $m \rightarrow \infty$

$$(23) \quad 2\Psi'(x) \Psi'(y) = [\Psi(x+y) - \Psi(x-y)]\Psi''(0).$$

En différentiant (23) par rapport à x et à y nous obtenons

$$(24) \quad 2\Psi''(x) \Psi''(y) = [\Psi''(x+y) + \Psi''(x-y)]\Psi''(0).$$

Observons que $\Psi''(0) = 0$ est exclu par (23) puisque $\Psi'(x)$ ne s'annule pas identiquement. En posant

$$(25) \quad \frac{\Psi''(x)}{\Psi''(0)} = f(x)$$

nous obtenons finalement

$$(26) \quad 2f(x) f(y) = f(x+y) + f(x-y).$$

C'est une des équations fonctionnelles classiques traitées par CAUCHY (4). Observons qu'en vertu de l'hypothèse (III) la fonction $f(x)$ donnée par (25) est continue ; elle est paire, voir (17), et se réduit à 1 pour $x = 0$. Dans ces conditions la solution ne peut avoir d'après CAUCHY qu'une des trois formes suivantes :

$$(27) \quad \frac{\Psi''(x)}{\Psi''(0)} = f(x) = \begin{cases} 1 \\ \cosh bx = \frac{1}{2}(e^{bx} + e^{-bx}) \\ \cos bx \end{cases}$$

b étant une certaine constante positive. La troisième forme, $\cos bx$ est exclue par (17) et (18), la première n'est qu'un cas limite de la seconde, pour $b \rightarrow 0$. En observant que $\Psi''(0)$ doit être négative et $\Psi'(x)$ impaire, voir (17), (18), on obtient par deux intégrations

$$(28) \quad \Psi(x) = \frac{c}{b^2} [1 - \cosh (bx)] + \log d,$$

ce qui contient comme cas limite ($b \rightarrow 0$)

$$(28^*) \quad \Psi(x) = -\frac{cx^2}{2} + \log d;$$

(4) A. L. CAUCHY, Cours d'Analyse de l'Ecole Royale Polytechnique (Paris 1821). Voir p. 114-122.

b, c, d sont des constantes positives. En appliquant la transformation opérée au commencement du n° IV dans la direction inverse on trouve la forme définitive de $\varphi(x)$.

VI

Voici le résultat :

Chaque fonction $\varphi(x)$ satisfaisant aux postulats (I), (I'), (II) et (III) est nécessairement d'une des deux formes suivantes

$$(29) \quad \varphi(x) = de^{cb^{-2} [1 - \cosh b(x-a)]}$$

$$(29^*) \quad \varphi(x) = de^{-\frac{c(x-a)^2}{2}}$$

b, c, d étant des constantes positives et a une constante réelle quelconque ; (29*) est le cas limite de (29) pour $b \rightarrow 0$.

On peut ajouter qu'inversement a, b, c étant donnés, $b > 0, c > 0, a$ quelconque, on peut déterminer d de manière que $\varphi(x)$, donnée par (29) ou (29*) soit une loi d'erreur. Cette loi d'erreur fournira la même valeur pour la grandeur mesurée avec les trois méthodes a, b, c expliquées au n° I, quelles que soient les mesures.

En effet, $\varphi(x)$ étant positive, pour qu'elle puisse être admise comme loi d'erreur elle doit encore satisfaire à la condition

$$\int_{-\infty}^{+\infty} \varphi(x) dx = 1$$

ce qui n'exige qu'un choix convenable de d .

Puis, l_1, l_2, \dots, l_n étant donnés, l'équation en l

$$(30) \quad \sinh b(l - l_1) + \sinh b(l - l_2) + \dots + \sinh b(l - l_n) = 0$$

aura une unique racine réelle, vu que la fonction

$$\sinh x = \frac{e^x - e^{-x}}{2}$$

est strictement monotone et varie de $-\infty$ à $+\infty$ avec x . (On peut même dire que la racine l sera comprise entre le plus petit et le plus grand

SUR QUELQUES POINTS DE LA THÉORIE DES PROBABILITÉS

des nombres l_1, l_2, \dots, l_n) $\Psi(x)$ étant donnée par (28) — je prends, ce qui est sans importance, $a = 0$ dans (29) — on a, en vertu de (30)

$$\begin{aligned} & \sum_{v=1}^n \Psi(x - l_v) - n(cb^{-2} + \log d) \\ = & -cb^{-2} \sum_{v=1}^n [\cosh b(x - l) \cosh b(l - l_v) + \sinh b(x - l) \sinh b(l - l_v)] \\ = & -cb^{-2} \cosh b(x - l) \sum_{v=1}^n \cosh b(l - l_v). \end{aligned}$$

C'est bien une fonction paire de $(x - l)$ et une fonction décroissante de $(x - l)^2$.

On peut vérifier encore plus facilement que la fonction (29*) possède la même propriété de symétrie ; dans ce cas-là, même

$$\varphi(a_1(x - l_1)) \varphi(a_2(x - l_2)) \dots \varphi(a_n(x - l_n))$$

est une fonction paire de $(x - l^*)$; $a_1, a_2, \dots, a_n, l_1, l_2, \dots, l_n$ étant quelconques et l^* étant choisi convenablement. (Cette propriété caractérise d'ailleurs complètement (29*) puisqu'elle ne convient pas à (29) ; cela a une certaine importance quand on envisage des mesures qui n'ont pas toutes la même précision.)

En somme nous avons complètement déterminé les lois d'erreur qui mènent au même résultat avec les deux méthodes *a*) et *b*). Elles forment la famille de courbes (29) à 3 paramètres, contenant la courbe de GAUSS (29*) qui ne dépend que de 2 paramètres. En supposant (10) c'est-à-dire qu'il n'y ait pas d'erreur constante, on obtient $a = 0$, donc le nombre des paramètres se réduit d'une unité.

VII

Je vais ajouter quelques remarques sur la loi d'erreurs cyclique, dont la notion est due à M. VON MISES ⁽⁵⁾. Je n'ai pas à examiner ici si cette loi est applicable ou non au problème certainement passionnant de la « Ganzzahligkeit » (problème des valeurs entières tirées de l'observation) que son auteur a en vue ; je m'occuperai d'elle comme

(5) Physikalische Zeitschrift (1919), p. 490-500.

d'une notion mathématique apte à éclaircir par analogie les questions correspondantes sur la loi d'erreurs ordinaire.

Admettons qu'en mesurant des angles la probabilité d'une erreur ne dépend que de sa valeur numérique. (Les « angles » sont des valeurs réelles quelconques envisagées mod. 2π .) Comme loi d'erreurs cyclique je désignerai une fonction $\varphi(x)$ telle que la probabilité qu'une erreur tombe entre les limites α et β est

$$\int_{\alpha}^{\beta} \varphi(x) dx \quad (\alpha < \beta \leq \alpha + 2\pi).$$

On doit avoir

$$(31) \quad \varphi(x + 2\pi) = \varphi(x), \quad \varphi(x) \geq 0, \quad \int_0^{2\pi} \varphi(x) dx = 1.$$

Si n observations indépendantes d'un angle dont la vraie valeur est λ ont donné les mesures $\lambda_1, \lambda_2, \dots, \lambda_n$, la probabilité de les avoir rencontrées ensemble est proportionnelle à

$$\varphi(\lambda - \lambda_1) \varphi(\lambda - \lambda_2) \dots \varphi(\lambda - \lambda_n).$$

Si on ne connaît pas la vraie valeur de l'angle mesuré, seulement les mesures $\lambda_1, \lambda_2, \dots, \lambda_n$ fournies par l'observation et la loi d'erreurs $\varphi(x)$, on peut se proposer de former une conjecture plausible sur la vraie valeur. Mais ici surgit une nouvelle difficulté, inhérente au nouveau problème. Il y a évidemment des cas où on ne peut pas former une conjecture raisonnable, comme par exemple lorsque $\lambda_1, \lambda_2, \dots, \lambda_n$ désignent les directions de n rayons d'un cercle qui sont dirigés vers les sommets d'un polygone régulier de n côtés inscrit au cercle. Il faut éviter de demander trop. Bornons-nous au cas où les mesures couvrent moins qu'un demi-cercle, donc où on a

$$(32) \quad \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n < \lambda_1 + \pi.$$

Nous n'exigeons des méthodes suivantes un résultat déterminé que dans le cas (32).

Il est naturel d'envisager la distribution de probabilités sur le cercle unité dont la densité au point x est

$$(33) \quad \varphi(x - \lambda_1) \varphi(x - \lambda_2) \dots \varphi(x - \lambda_n).$$

(Le rayon passant par le « point x » forme l'angle x avec une direction donnée). On peut proposer deux méthodes.

a) On prend comme valeur plausible de l'angle cherché la valeur qui rend (33) maximum. C'est la méthode proposée par M. VON MISES, analogue à celle de GAUSS.

b) On cherche le centre de gravité d'une masse, répartie le long du cercle unité dont la densité au point x est (33) et on détermine l'angle cherché par le rayon qui passe par ce centre de gravité. Cela revient à prendre comme valeur plausible de l'angle cherché celle des deux valeurs λ satisfaisant à l'équation

$$\int_0^{2\pi} \sin(x - \lambda) \varphi(x - \lambda_1) \varphi(x - \lambda_2) \cdots \varphi(x - \lambda_n) dx = 0$$

qui est située dans le *plus petit* des deux arcs déterminés par λ_1 et λ_n ; n'oublions pas que nous avons supposé (32). Cette méthode n'était pas proposée par M. VON MISES mais est suggérée par ses considérations.

Nous avons un problème analogue à celui du n° I si nous cherchons les *lois d'erreurs cycliques pour lesquelles les méthodes a) et b) mènent nécessairement au même résultat*. Il sera suffisant d'esquisser la marche générale du raisonnement.

Comme au n° II, nous pouvons détacher les 2 postulats suivants :

(I). Quels que soient $\lambda_1, \lambda_2, \dots, \lambda_n$ satisfaisant à (32), la fonction (33) n'a qu'un maximum absolu entre 0 et 2π .

(II). Quels que soient $\lambda_1, \lambda_2, \dots, \lambda_n$, satisfaisant à (32), la solution de l'équation en λ

$$\int_0^{2\pi} \sin(x - \lambda) [\varphi(x - \lambda_1) \varphi(x - \lambda_2) \cdots \varphi(x - \lambda_n)]^p dx = 0$$

est indépendante de p , $p = 1, 2, 3, \dots$

Pour pouvoir appliquer les méthodes des nos III-V j'ai dû en outre faire les hypothèses supplémentaires suivantes :

(I''). La fonction (33) n'admet qu'un maximum *relatif* entre 0 et 2π , c'est-à-dire qu'elle est monotone entre son minimum et son maximum, les deux étant univoquement déterminés mod. 2π .

(III). Les dérivées $\varphi'(x)$ et $\varphi''(x)$ existent et sont continues : $\varphi(x) > 0$.

Avec des modifications faciles, les raisonnements des nos III-V mènent à l'équation fonctionnelle (26) dont il faudra maintenant prendre la troisième solution, inutilisable pour le problème précédent (voir (27)). Plus précisément, vu le postulat (I) et la périodicité exigée par (31), il faudra la prendre avec $b = 1$. D'où on obtient la loi d'erreurs cyclique

$$(34) \quad \varphi(x) = de^{c \cos x}$$

C'est la loi qui a été déduite par M. VON MISES *l. c.* d'un postulat différent, analogue au postulat dont GAUSS a déduit sa loi. Il me paraît intéressant que cette loi cyclique (34) conserve avec la loi de GAUSS une analogie aussi au point de vue de la comparaison des méthodes *a)* et *b)*.

Beaucoup plus loin l'analogie n'ira pas, me semble-t-il. En cherchant une loi d'erreurs cyclique qui provient de la superposition d'une multitude de petites erreurs indépendantes, on doit arriver à la loi de GAUSS « enroulée autour du cercle » ; c'est la loi

$$\varphi(x) = \frac{1}{\sqrt{2\pi c}} \sum_{n=-\infty}^{+\infty} e^{-\frac{(x - 2n\pi)^2}{2c}}$$

représentée, ce qui est assez curieux, par une série théta.

II. — Sur quelques courbes de fréquence résultant de la superposition d'effets fortuits interdépendants

I

La situation privilégiée de la courbe de GAUSS dans la statistique mathématique s'explique par deux raisons, l'une expérimentale et l'autre théorique. L'expérience nous montre que, dans des domaines très divers, il y a des courbes statistiques qui, avec une approximation raisonnable, peuvent être assimilées à la courbe de GAUSS. La théorie nous apprend que c'est la courbe de GAUSS qui doit résulter de la superposition d'un très grand nombre de petits effets fortuits indépendants. La constatation théorique élucide le fait expérimental ; plusieurs grandeurs dont la statistique nous est familière, comme les erreurs d'observation, les dimensions des organes des êtres vivants, etc., subissent des modifications dues à l'action d'une multitude de petites

causes incontrôlables qu'on peut raisonnablement considérer comme « dues au hasard ».

Mais dans tous les domaines d'application on rencontre aussi des courbes statistiques qui s'écartent considérablement de la courbe de GAUSS et une tâche très importante de la théorie est de chercher d'autres courbes théoriques, auxquelles les courbes expérimentales pourraient être raisonnablement assimilées. Ce problème paraît, par sa nature, indéterminé. On connaît les travaux de PEARSON, BRUNS, CHARLIER, KAPTEYN et d'autres qui en présentent des solutions très différentes, chaque solution généralisant la courbe de GAUSS dans une autre direction.

Il me paraît qu'il ne faut pas considérer ce problème d'une manière formelle et chercher des familles de courbes quelconques qui se déduisent de la courbe de GAUSS par une opération analytique quelconque ou partagent avec la courbe de GAUSS une propriété analytique quelconque. Ce qu'il faudrait élargir c'est la base théorique de la courbe de GAUSS et c'est le chemin suivant qui semble se présenter le plus naturellement. Gardons la conception fondamentale que la courbe théorique cherchée résulte de la sommation d'un très grand nombre d'effets dûs au hasard, mais *abandonnons l'hypothèse supplémentaire* que ces effets sont *indépendants*. Le problème mathématique qui s'impose de cette façon peut être formulé ainsi : étudier les lois limites provenant de la somme d'un grand nombre de variables aléatoires interdépendantes.

C'est ce problème vaste, trop vaste peut-être, que j'entreprendrai d'éclaircir par quelques exemples choisis avec soin. Ils nous feront entrevoir quelques conclusions générales et nous mèneront à quelques lois limites de fréquences, qui ne seront pas simplement des formules commodes pour l'interpolation, mais des formules susceptibles d'interprétation.

II

Considérons une suite de variables aléatoires ⁽⁶⁾

$$X_1, X_2, X_3, \dots, X_n, X_{n+1}, \dots$$

et concevons-les comme réalisées par des expériences successives,

(6) Les variables aléatoires seront désignées par des caractères gras.

x_1 faisant le commencement et x_{n+1} venant après x_n . Le cas général d'interdépendance consiste en ceci que *les chances de x_{n+1} sont déterminées par l'ensemble des valeurs que les expériences précédentes ont attribuées à x_1, x_2, \dots, x_n .*

Un cas particulier important est celui où les chances de x_{n+1} ne dépendent que de la somme $x_1 + x_2 + \dots + x_n$, deux combinaisons x_1, x_2, \dots, x_n de même somme donnant les mêmes chances pour x_{n+1} . Dans ce cas-ci je parlerai d'*influence globale*.

Un autre cas particulier important est celui où les chances de x_{n+1} ne dépendent que de x_n , deux combinaisons x_1, x_2, \dots, x_n avec la même valeur de x_n donnant les mêmes chances pour x_{n+1} . Dans ce cas-là je parlerai de l'*influence du prédécesseur*. Ce cas a été introduit par A. MARKOFF qui en a fait une étude approfondie sous le nom d'*épreuves liées en chaîne* (7).

On dit que les variables aléatoires $x_1, x_2, \dots, x_n, x_{n+1}, \dots$ sont *indépendantes*, si les chances de x_{n+1} ne dépendent pas du tout du résultat des épreuves précédentes, deux combinaisons quelconques x_1, x_2, \dots, x_n donnant les mêmes chances pour x_{n+1} .

L'interdépendance des variables aléatoires quelconques $x_1, x_2, \dots, x_n, x_{n+1}, \dots$ consiste en ceci que les chances de x_n sont modifiées si on fixe les valeurs d'autres variables $x_j, x_k, \dots, x_p, x_q, \dots$ qui peuvent en partie précéder x_n et en partie la suivre. La distinction faite entre influence générale, influence globale, influence du prédécesseur et indépendance a été basée sur la considération des chances de x_{n+1} quand toutes les variables précédentes x_1, x_2, \dots, x_n sont fixées et aucune des variables suivantes x_{n+1}, x_{n+2}, \dots ne l'est. Il est encore important de considérer les chances de x_n lorsqu'aucune variable n'est fixée, ni x_1 , ni x_2, \dots ni x_{n+1}, \dots , les chances de x_n « avant la première épreuve » que je nommerai les chances *à priori* de x_n . (Une confusion avec la signification demi-philosophique du terme « probabilité *à priori* » me semble facile à éviter.) Si les variables aléatoires x_1, x_2, x_3, \dots sont indépendantes, les chances de x_n après la fixation des valeurs d'autres variables quelconques $x_j, x_k, \dots, x_p, x_q, \dots$ sont égales à ses chances *à priori*.

(7) A. A. MARKOFF : a) *Wahrscheinlichkeitsrechnung* (Leipzig-Berlin, 1912), p. 272-298 ; b) Supplément à la 3^e édition russe de l'ouvrage précité ; traduction française parue séparément (St.-Petersbourg, 1913), p. 44-66.

SUR QUELQUES POINTS DE LA THÉORIE DES PROBABILITÉS

Je dirai que la suite $x_1, x_2, \dots, x_n, \dots$ est *homogène* si les chances à *priori* de toutes les variables sont les mêmes.

Les variables $x_1, x_2, \dots, x_n, \dots$ peuvent varier d'une manière continue ou discontinue, elles peuvent aussi être des vecteurs dans un espace d'un nombre quelconque de dimensions.

Nous avons le cas le plus simple, lorsque chaque épreuve ne présente qu'une alternative, c'est-à-dire lorsque x_n n'est susceptible que de deux valeurs différentes, disons de 0 et de 1 ($n = 1, 2, 3, \dots$). Dans ce cas je parlerai d'épreuves ou de variables *alternatives*.

Le cas le plus simple, le plus classique et aussi le plus important, est celui où les épreuves sont à la fois alternatives, homogènes et indépendantes. Chaque épreuve devant présenter une alternative, chacune la même alternative, avec des chances indépendantes du résultat des autres, l'épreuve en question peut être représentée par un tirage d'une urne qui ne contient que deux espèces de boules, disons rouges et noires, l'urne présentant à tous les tirages la même composition. Il s'agit donc de tirages réitérés d'une urne dont on maintient la composition constante, en remettant chaque fois la boule tirée. La suite de tirages ne dépend que d'un paramètre, de la probabilité de tirer rouge de l'urne en question que nous désignerons par ρ . C'est le schéma d'urnes le plus classique à propos duquel Jacques BERNOULLI a trouvé son théorème et DE MOIVRE a rencontré la loi de fréquences qu'on nomme généralement la loi de GAUSS.

C'est le schéma d'urne que je veux modifier essentiellement, mais aussi simplement que possible. Je considérerai des épreuves alternatives, homogènes mais non-indépendantes et je fixerai le degré d'interdépendance par un seul paramètre, qui s'ajoutera au paramètre ρ , probabilité à *priori* de tirer rouge, la même pour toutes les épreuves. On pourrait ajouter ce second paramètre d'une infinité de manières, mais je ne considérerai que deux structures d'interdépendance, essentiellement différentes : je prendrai un exemple de l'influence globale et un autre de l'influence du prédécesseur.

III

Un phénomène familier où il s'agit d'une influence globale (avec des à peu près inévitables en statistique) est le progrès d'une épidémie. Nous jugeons les chances d'éviter l'épidémie surtout d'après le *chiffre*

total des cas de maladie ou de mort. Le fait élémentaire dans le progrès d'une épidémie est la production des germes ; celui qui tombe malade produit de nouveaux germes, en nombre beaucoup supérieur et augmente par cela les chances de son entourage de tomber malade. En réduisant ce fait à son expression la plus simple et en y ajoutant une certaine symétrie, propice au traitement mathématique, nous sommes amenés au schéma d'urnes suivant :

Une urne contient originalement N boules, dont R sont rouges et S noires, $R + S = N$. Nous faisons de l'urne des tirages successifs en ajoutant à l'urne, après chaque tirage, à la place de la boule tirée $1 + \Delta$ boules *de la même couleur*. Si Δ est positif, le nombre des boules augmente après chaque tirage, chaque succès obtenu favorise les chances des succès à obtenir, chaque insuccès gêne encore les chances des épreuves suivantes, le succès ainsi que l'insuccès sont *contagieux*. Ajouter un nombre négatif signifie enlever ; donc si Δ est négatif, le nombre des boules diminue au cours des tirages, chaque succès obtenu diminue les chances de succès ultérieurs ; mais aussi les insuccès sont de la même nature et chaque coup tend à amener un revirement de fortune. Si $\Delta = 0$, les épreuves sont indépendantes ; c'est le cas classique de la boule remise. Le cas $\Delta = -1$, celui de la boule non-remise est aussi classique.

Associons aux tirages successifs les variables aléatoires x_1, x_2, x_3, \dots le premier tirage correspondant à x_1 , le second à x_2 , etc. Soit $x_n = 1$ si le tirage correspondant fait sortir une boule rouge et soit $x_n = 0$ s'il fait sortir une noire.

Après n tirages l'urne contiendra un nombre total de $N + n\Delta$ boules, dont $R + (x_1 + x_2 + \dots + x_n)\Delta$ rouges. La probabilité de tirer une boule rouge au $(n + 1)$ -*me* tirage est donc

$$(1) \quad \frac{R + (x_1 + x_2 + \dots + x_n)\Delta}{N + n\Delta} = \frac{\rho + (x_1 + x_2 + \dots + x_n)\delta}{1 + n\delta}.$$

Je pose ici comme dans ce qui suit

$$(2) \quad \rho = \frac{R}{N}, \quad \sigma = \frac{S}{N} = 1 - \rho, \quad \delta = \frac{\Delta}{N}.$$

(1) est la probabilité de l'équation $x_{n+1} = 1$. Elle ne dépend que de la somme $x_1 + x_2 + \dots + x_n$ des variables aléatoires précédentes ; c'est bien le cas de l'influence globale.

La probabilité (1) contient deux paramètres, ρ et δ ; ρ est la pro-

SUR QUELQUES POINTS DE LA THÉORIE DES PROBABILITÉS

babilité *a priori* de l'équation $x_1 = 1$, δ exprime par son signe la direction et par sa valeur absolue l'intensité de l'interdépendance. On peut attribuer à ρ une valeur quelconque de l'intervalle $0 < \rho < 1$. Etant donné ρ , si la série comprend n tirages, δ doit satisfaire à la condition

$$(3) \quad \rho + (n - 1)\delta \geq 0;$$

le premier membre est le numérateur de la probabilité d'avoir $x_n = 1$ après $x_1 = x_2 = \dots = x_{n-1} = 1$. La condition (3) est l'unique à laquelle δ doit satisfaire pourvu que $\rho \leq \sigma$.

La structure d'interdépendance que je viens d'expliquer et qui peut être caractérisée soit par le schéma d'urnes considéré, soit par la formule (1), sera nommée brièvement la structure de la « contagion » (8).

IV

Le cas le plus simple de l'influence du prédécesseur est représenté par le schéma d'urnes suivant.

Pour faire une série de tirages, on dispose de trois urnes, $\mathcal{U}_1, \mathcal{U}_\rho, \mathcal{U}_\sigma$ chacune contenant des boules rouges et noires. De l'urne \mathcal{U}_1 on ne fait qu'un tirage, le premier de la série. Si la boule sortie au *n-me* tirage était rouge, le $(n + 1)$ -me tirage se fera de \mathcal{U}_ρ , si elle était noire, il se fera de \mathcal{U}_σ . On maintient constante la composition de \mathcal{U}_ρ ainsi que celle de \mathcal{U}_σ en remettant chaque fois la boule puisée. Les probabilités sont données par le tableau suivant :

urne	\mathcal{U}_1	\mathcal{U}_ρ	\mathcal{U}_σ
probabilité de boule rouge	ρ_1	ρ_ρ	ρ_σ
probabilité de boule noire	σ_1	σ_ρ	σ_σ

Donc ρ_ρ est la probabilité du « rouge après rouge », σ_ρ du « noir après rouge », etc. On a

$$(4) \quad \rho_1 + \sigma_1 = 1, \quad \rho_\rho + \sigma_\rho = 1, \quad \rho_\sigma + \sigma_\sigma = 1.$$

(8) Voir F. EGGENBERGER et G. PÓLYA : a) Zeitschrift f. angewandte Mathematik u. Mechanik, t. III (1923), p. 279-289 ; b) Comptes-rendus, t. 187 (1928), p. 870-872 ; F. EGGENBERGER, Thèse, Zurich, 1924 ; A. A. MARKOFF, Bulletin de l'Académie Imp. des Sciences, Pétrograd (1917) p. 177-186 (M. Serge BERNSTEIN a eu l'amabilité de me signaler cet ouvrage écrit en russe). Essentiellement la même structure de probabilité a été envisagée par Léon BRILLOUIN en connection avec un problème de la théorie des quanta, Annales de Physique, t. 7 (1927), p. 315-331.

Associons aux tirages successifs les variables aléatoires x_1, x_2, x_3, \dots de la même manière qu'au n° III. La probabilité que la boule sortie au $(n + 1)$ -me tirage soit rouge ou, ce qui est la même chose, la probabilité d'avoir $x_{n+1} = 1$ est donnée par

$$(5) \quad \rho_\sigma + (\rho_\rho - \rho_\sigma)x_n = \rho_\sigma + \tau x_n$$

où j'ai posé

$$(6) \quad \tau = \rho_\rho - \rho_\sigma = \sigma_\sigma - \sigma_\rho = \begin{vmatrix} \rho_\rho & \sigma_\rho \\ \rho_\sigma & \sigma_\sigma \end{vmatrix}$$

La probabilité (5) ne dépend que de x_n ; c'est bien le cas de l'influence du prédécesseur.

Si τ est positif, la structure des tirages favorise la *ressemblance* de chaque épreuve à la précédente ; elle favorise la *dissemblance* si τ est négatif. Dans le premier cas, la valeur absolue de τ exprime la force de la « tradition », dans le second la force de l'« opposition ».

Un exemple aussi simple qu'ingénieux de cette structure d'interdépendance est due à MARKOFF lui-même ⁽⁹⁾. Il est familier à tout le monde : c'est un texte quelconque dans lequel les lettres, consonnes et voyelles, se suivent. Entendons par $\rho_\rho, \rho_\sigma, \sigma_\rho, \sigma_\sigma$ les quatre probabilités suivantes déterminées expérimentalement par dénombrement direct : « consonne après consonne », « consonne après voyelle », « voyelle après consonne », « voyelle après voyelle ». Nous aurons un τ négatif, donc le cas de la dissemblance et le schéma d'urnes expliqué correspondra approximativement aux faits observables ⁽¹⁰⁾. On pourrait penser que le cas de τ positif, celui de la ressemblance, se trouve réalisé dans une longue lignée de descendants successifs. Les dispositions sont transmises de père en fils ; les influences ancestrales de prédécesseurs très éloignés peuvent se faire sentir, mais seulement indirectement, par l'intermédiaire du *prédécesseur immédiat* ⁽¹¹⁾. Quoiqu'il en soit, simplement pour avoir une expression brève, je parlerai du cas discuté des épreuves liées en chaîne, caractérisée par (5), comme de la structure de l'« hérédité ».

(9) Voir l. c. 7) b) p. 56-66.

(10) Mais seulement approximativement : la $(n - 1)$ -me lettre semble exercer sur la $(n + 1)$ -me une influence *directe* non négligeable bien que moins forte que celle qu'elle exerce indirectement par l'intermédiaire de la n -me ; et il y a encore la $(n - 2)$ -me lettre, etc. Voir les résultats expérimentaux de MARKOFF l. c.

(11) Cette idée a été émise, sous une forme plus précise par Serge BERNSTEIN, *Mathematische Annalen*, t. 97 (1926), p. 1-59, voir p. 40-41.

V

Je vais calculer quelques espérances mathématiques, utiles à l'étude de la structure de la « contagion ». Ce calcul sera des plus simples si nous nous appuyons sur le fait suivant :

La structure de la « contagion » est homogène, c'est-à-dire que toutes les variables aléatoires $x_1, x_2, \dots, x_n, \dots$ ont les mêmes probabilités à priori. Outre cela, l'interdépendance entre deux variables quelconques x_k et x_l est la même.

Cela est plausible puisque la probabilité (1) est une fonction symétrique des variables qui précèdent x_{n+1} , mais cela ne va pas de soi. Le résultat énoncé contient une propriété importante de la structure de la contagion qui sera démontrée sous une forme un peu plus générale au n° VII et dont je ferai usage dès maintenant.

Valeur probable de x_n . L'espérance mathématique de x_n pour n quelconque est la même que celle de x_1 , à cause de l'homogénéité. x_1 ne prend que deux valeurs, 1 et 0, avec la probabilité ρ et σ respectivement ; on a donc (12)

$$(7) \quad \mathbb{E}x_n = \mathbb{E}x_1 = \rho \cdot 1 + \sigma \cdot 0 = \rho.$$

Ecart de x_n . Le carré de l'écart moyen quadratique de x_n est, lui aussi, indépendant de n . On a

$$(8) \quad \mathbb{E}(x_n - \rho)^2 = \mathbb{E}(x_1 - \rho)^2 = \rho(1 - \rho)^2 + \sigma(0 - \rho)^2 = \rho\sigma.$$

Coefficient de corrélation de x_k et x_l . Il suffit de calculer, d'après le principe avancé, le coefficient de corrélation de x_1 et de x_2 , celui de x_k et x_l étant le même. Commençons par calculer la probabilité d'avoir les deux équations simultanées

$$x_1 = 1, \quad x_2 = 1,$$

c'est-à-dire que les deux premiers tirages donnent rouge. La probabilité d'avoir $x_1 = 1$ est ρ . Étant donné $x_1 = 1$, la probabilité d'avoir $x_2 = 1$ est, d'après (1), $\frac{\rho + \delta}{1 + \delta}$. La probabilité cherchée est la probabilité composée

$$\rho \cdot \frac{\rho + \delta}{1 + \delta}.$$

(12) E désigne l'espérance mathématique.

En calculant de la même manière encore trois probabilités analogues, on obtient la table de corrélation de x_1 et de x_2

	x_1		
x_2		I	0
I		$\frac{\rho(\rho + \delta)}{I + \delta}$	$\frac{\rho\sigma}{I + \delta}$
0		$\frac{\sigma\rho}{I + \delta}$	$\frac{\sigma(\sigma + \delta)}{I + \delta}$

Observons en passant que cette table est symétrique par rapport à sa diagonale principale, donc x_2 dépend de la même manière de x_1 que x_1 de x_2 . En l'utilisant on trouve

$$(9) \quad E(x_1 - \rho)(x_2 - \rho) = \frac{\rho(\rho + \delta)}{I + \delta}(I - \rho)(I - \rho) + \frac{\rho\sigma}{I + \delta}(0 - \rho)(I - \rho) + \dots \\ = \frac{\rho\sigma\delta}{I + \delta} = E(x_k - \rho)(x_l - \rho).$$

Le coefficient de corrélation entre x_k et x_l est, en vertu de (8) et de (9),

$$(10) \quad \frac{E(x_k - \rho)(x_l - \rho)}{\sqrt{E(x_k - \rho)^2 E(x_l - \rho)^2}} = \frac{\delta}{I + \delta};$$

il a le signe de δ et il croît avec δ , s'approchant de l'unité quand δ est positif et très grand, c'est-à-dire lorsque la contagion est forte.

Valeur probable de $r = x_1 + x_2 + \dots + x_n$. — La variable aléatoire r est le nombre, sujet au hasard, des boules rouges sorties aux n premiers tirages. Sa valeur probable est

$$(11) \quad E r = E x_1 + E x_2 + \dots + E x_n = n\rho$$

en vertu de (7).

Ecart de r . — On obtient son carré, en utilisant (8) et (9)

$$(12) \quad E(r - n\rho)^2 = E(x_1 - \rho + x_2 - \rho + \dots + x_n - \rho)^2 \\ = nE(x_1 - \rho)^2 + n(n - 1)E(x_1 - \rho)(x_2 - \rho) \\ = n\rho\sigma\left(I + \frac{(n - 1)\delta}{I + \delta}\right).$$

Coefficient de dispersion. — On obtient le carré de ce coefficient (qui sera comme d'usage, désigné par Q) en divisant $E(r - n\rho)^2$ par $n\rho\sigma$, ce qui serait la valeur de (12) si x_1, x_2, \dots, x_n étaient indé-

pendants. Donc

$$(13) \quad Q^2 = \frac{1 + n\delta}{1 + \delta}.$$

On a $Q > 1$ ou $Q < 1$ selon que $\delta > 0$ ou $\delta < 0$. C'est-à-dire que la dispersion est hypernormale ou hyponormale selon que les chances augmentent ou diminuent avec le succès. Cela se comprend.

VI

La structure de l'« hérédité », telle qu'elle a été définie au n° IV, n'est pas nécessairement homogène. Désignons par ρ_k la probabilité à priori (voir n° II) d'avoir $x_k = 1$ et par σ_k celle d'avoir $x_k = 0$. On a

$$(14) \quad \rho_k + \sigma_k = 1.$$

Cherchons la relation entre ρ_k et ρ_{k+1} .

Que la boule tirée au $(k + 1)$ -*me* tirage soit rouge peut arriver de deux manières : elle est sortie de \mathcal{U}_ρ ou de \mathcal{U}_σ selon que la boule précédente a été rouge ou noire. La probabilité totale est donc

$$\rho_{k+1} = \rho_k \rho_\rho + \sigma_k \rho_\sigma$$

ou, utilisant (6) et (14)

$$(15) \quad \rho_{k+1} = \tau \rho_k + \rho_\sigma.$$

Déterminons ρ par

$$(16) \quad \rho = \tau \rho + \rho_\sigma.$$

On aura

$$\rho_{k+1} - \rho = \tau(\rho_k - \rho)$$

et, en répétant le raisonnement,

$$(17) \quad \rho_{k+m} - \rho = \tau^m(\rho_k - \rho).$$

En posant $k = 1$ dans (17) on voit qu'en supposant

$$(18) \quad \rho_1 = \rho, \quad \sigma_1 = 1 - \rho$$

on obtient $\rho_1 = \rho_2 = \rho_3 = \dots = \rho$ c'est-à-dire que la série des variables aléatoires x_1, x_2, x_3, \dots est homogène.

A partir d'ici, j'introduis la nouvelle hypothèse (18) comme partie essentielle de la définition de « l'hérédité » dont la structure devient

par cela homogène. Observons encore, que d'après la définition complétée, cette structure ne dépend que de deux paramètres, ρ et τ . En effet en posant

$$(19) \quad \sigma = 1 - \rho$$

on tire de (4), (6), (16) et (19)

$$(20) \quad \begin{cases} \rho_\rho = \rho + \tau\sigma, & \sigma_\rho = \sigma - \tau\sigma, \\ \rho_\sigma = \rho - \tau\rho, & \sigma_\sigma = \sigma + \tau\rho. \end{cases}$$

En vertu de (18), (19), (20), la composition des trois urnes $\mathfrak{u}_1, \mathfrak{u}_\rho, \mathfrak{u}_\sigma$, est exprimée par ρ et τ . On peut choisir ρ dans l'intervalle $0 < \rho < 1$. Étant donné ρ , τ doit être tel que $\rho_\rho, \rho_\sigma, \sigma_\rho, \sigma_\sigma$, données par (20), soient comprises entre 0 et 1. Si l'on suppose $\rho \leq \sigma$ il est nécessaire (et aussi suffisant) que

$$(21) \quad -1 \leq -\frac{\rho}{1-\rho} < \tau < 1.$$

Calculons encore la probabilité d'avoir les deux équations simultanées

$$\mathbf{x}_k = 1, \quad \mathbf{x}_{k+m} = 1$$

c'est-à-dire que le k -me et le $(k+m)$ -me tirage donnent rouge les deux. La probabilité d'avoir $\mathbf{x}_k = 1$ est ρ . Étant donné $\mathbf{x}_k = 1$, on obtient la probabilité d'avoir $\mathbf{x}_{k+m} = 1$ comme le ρ_{k+m} de la formule (17) à condition d'y mettre $\rho_k = 1$. La probabilité cherchée est la probabilité composée

$$\rho(\rho + \tau^m\sigma).$$

En calculant de la même manière encore trois probabilités analogues, on obtient la table de corrélation de \mathbf{x}_k et de \mathbf{x}_{k+m} :

	\mathbf{x}_k		
		1	0
\mathbf{x}_{k+m}			
1		$\rho(\rho + \tau^m\sigma)$	$\sigma(\rho - \tau^m\rho)$
0		$\rho(\sigma - \tau^m\sigma)$	$\sigma(\sigma + \tau^m\rho)$

Observons que cette table étant symétrique par rapport à sa diagonale principale, la variable \mathbf{x}_k est liée de la même manière à \mathbf{x}_{k+m} que \mathbf{x}_{k+m} l'est à \mathbf{x}_k . Observons encore que les probabilités qui y sont

SUR QUELQUES POINTS DE LA THÉORIE DES PROBABILITÉS

contenues ne dépendent que de m , valeur absolue de la différence des indices k et $k + m$.

En résumé :

La structure de l'« hérédité » est homogène, c'est-à-dire que toutes les variables aléatoires $x_1, x_2, \dots, x_n, \dots$ ont les mêmes probabilités a priori. L'interdépendance entre deux variables quelconques x_k et x_l ne dépend que de leur distance $|k - l|$.

Nous pouvons maintenant calculer tout comme au n° V, les espérances mathématiques les plus essentielles à l'étude de la structure de l'hérédité.

Valeur probable de x_n . A cause de l'homogénéité

$$(22) \quad \mathbb{E}x_n = \mathbb{E}x_1 = \rho \cdot 1 + \sigma \cdot 0 = \rho.$$

Ecart de x_n . Le carré de cet écart est

$$(23) \quad \mathbb{E}(x_n - \rho)^2 = \mathbb{E}(x_1 - \rho)^2 = \rho\sigma.$$

Coefficient de corrélation de x_k et x_{k+m} . La table de corrélation trouvée ci-dessus donne

$$(24) \quad \mathbb{E}(x_k - \rho)(x_{k+m} - \rho) = \rho\sigma\tau^m.$$

On obtient de (23) et (24) le coefficient de corrélation

$$(25) \quad \frac{\mathbb{E}(x_k - \rho)(x_{k+m} - \rho)}{\sqrt{\mathbb{E}(x_k - \rho)^2 \mathbb{E}(x_{k+m} - \rho)^2}} = \tau^m.$$

Il est positif dans le cas de la ressemblance ($\tau > 0$) ; dans le cas de la dissemblance, il est négatif pour $m = 1$ et son signe alterne quand m parcourt successivement les nombres 1, 2, 3, ... ; dans les deux cas sa valeur absolue diminue si $|\tau|$ diminue ou si m augmente. Tout cela se comprend, si l'on envisage l'exemple des voyelles et des consonnes ou celui des descendants successifs, fils, petit-fils, arrière petit-fils, etc.

Valeur probable de $r = x_1 + x_2 + \dots + x_n$. — C'est le nombre moyen des boules rouges sorties en n tirages. C'est $\mathbb{E}r = n\rho$.

Ecart de r . — On obtient pour son carré, en utilisant (23) et (24)

$$(26) \quad \begin{aligned} \mathbb{E}(r - n\rho)^2 &= \mathbb{E}(x_1 - \rho + x_2 - \rho + \dots + x_n - \rho)^2 \\ &= n\mathbb{E}(x_1 - \rho)^2 + 2(n-1)\mathbb{E}(x_1 - \rho)(x_2 - \rho) + \dots + 2\mathbb{E}(x_1 - \rho)(x_n - \rho) \\ &= n\rho\sigma + 2(n-1)\rho\sigma\tau + 2(n-2)\rho\sigma\tau^2 + \dots + 2\rho\sigma\tau^{n-1}. \end{aligned}$$

G. POLYA

Coefficient de dispersion. On a

$$(27) \quad Q^2 = \frac{E(r-n)^2}{n\sigma^2} = 1 + \frac{2}{n} [(n-1)\tau + (n-2)\tau^2 + (n-3)\tau^3 + \dots + \tau^{n-1}] \\ = \frac{1+\tau}{1-\tau} - 2\tau \frac{1-\tau^n}{n(1-\tau)^2}.$$

On voit, encore mieux par la première des deux expressions (27), que la dispersion est hypernormale ($Q > 1$) s'il y a « tradition » ($\tau > 0$) et hyponormale ($Q < 1$) s'il y a « opposition » ($\tau < 0$).

VII

Retournons à la structure de la « contagion » pour démontrer la proposition avancée au début du n° V.

Calculons pour commencer la probabilité q_r d'avoir

$$(28) \quad x_1 = x_2 = \dots = x_r = 1, \quad x_{r+1} = x_{r+2} = \dots = x_n = 0$$

c'est à dire la probabilité que parmi n tirages successifs les r premiers donnent rouge et les s derniers noir, où

$$r + s = n.$$

q_r est une probabilité composée dont on trouve les r premiers facteurs en mettant $n = 0, 1, 2, \dots, r-1$ dans (1) et en donnant les justes valeurs à x_1, x_2, \dots ; on trouve les s derniers facteurs en soustrayant la valeur fournie par (1) de l'unité. La probabilité cherchée est donc

$$(29) \quad q_r = \frac{\rho(\rho+\delta)(\rho+2\delta)\dots(\rho+(r-1)\delta)}{1(1+\delta)(1+2\delta)\dots(1+(r-1)\delta)} \frac{\sigma(\sigma+\delta)\dots(\sigma+(s-1)\delta)}{(1+\rho\delta)(1+(\rho+1)\delta)\dots(1+(n-1)\delta)}.$$

La probabilité que parmi les $n = r + s$ variables aléatoires

$$x_1, x_2, x_3, \dots, x_n$$

r déterminées soient égales à 1 et les autres s à 0, est également q_r , quel que soit l'ordre prescrit pour les 1 et les 0, parce que la probabilité en question est composée de n facteurs dont les dénominateurs sont les mêmes que de ceux qui composent q_r et les numérateurs ne sont changés qu'en ordre.

SUR QUELQUES POINTS DE LA THÉORIE DES PROBABILITÉS

Introduisons n indéterminées (variables ordinaires) z_1, z_2, \dots, z_n et calculons l'espérance mathématique

$$(30) \quad \mathbb{E} z_1^{x_1} z_2^{x_2} z_3^{x_3} \dots z_n^{x_n}.$$

Le terme $z_1 z_2 \dots z_r$ sera multiplié par la probabilité d'avoir les équations simultanées (28), donc par q_r . Mais un terme quelconque $z_{i_1} z_{i_2} \dots z_{i_r}$, produit de r facteurs différents, sera également multiplié par q_r . La somme de tous les termes analogues est la r -me fonction symétrique élémentaire

$$z_1 z_2 \dots z_r + z_1 z_2 \dots z_{r-1} z_{r+1} + \dots + z_{n-r+1} \dots z_{n-1} z_n = Z_r.$$

Donc

$$(31) \quad \mathbb{E} z_1^{x_1} z_2^{x_2} \dots z_n^{x_n} = q_0 + q_1 Z_1 + q_2 Z_2 + \dots + q_n Z_n = \mathcal{F}(z_1, z_2, z_3, \dots, z_n)$$

où $\mathcal{F}(z_1, z_2, \dots, z_n)$ est un polynôme symétrique de n variables.

La connaissance de $\mathcal{F}(z_1, z_2, \dots, z_n)$ suffit pour répondre à une question de probabilité quelconque concernant x_1, x_2, \dots, x_n . Par exemple la probabilité d'avoir $x_k = 1$ est donnée par

$$\mathbb{E} x_k = \left(\frac{\partial \mathcal{F}(z_1, z_2, \dots, z_n)}{\partial z_k} \right)_{z_1 = z_2 = \dots = z_n = 1},$$

celle d'avoir $x_k = 1$ et $x_l = 1$ par

$$\mathbb{E} x_k x_l = \left(\frac{\partial^2 \mathcal{F}(z_1, z_2, \dots, z_n)}{\partial z_k \partial z_l} \right)_{z_1 = z_2 = \dots = z_n = 1}$$

etc. Mais $\mathcal{F}(z_1, z_2, \dots, z_n)$ est *symétrique* en z_1, z_2, \dots, z_n . Par conséquent *toutes les probabilités qui dépendent de x_1, x_2, \dots, x_n sont parfaitement symétriques par rapport à ces n variables aléatoires*. Cela contient la proposition à démontrer et davantage ; par exemple l'interdépendance entre trois variables quelconques x_k, x_l et x_m est la même, quels que soient les indices, etc.

Les calculs faits nous seront utiles encore pour d'autres questions. Par exemple la probabilité $p_{r,s}$ d'obtenir r fois rouge et s fois noir sur $n = r + s$ tirages dans un ordre quelconque ou d'avoir, ce qui est la même chose, $r = x_1 + x_2 + \dots + x_n = r$, est évidemment

$$(32) \quad p_{r,s} = \binom{n}{r} q_r = \binom{r+s}{r} \frac{\rho(\rho+\delta) \dots (\rho+(r-1)\delta) \sigma(\sigma+\delta) \dots (\sigma+(s-1)\delta)}{1(1+\delta) \dots (1+(n-1)\delta)}.$$

En mettant $z_1 = z_2 = \dots = z_n = e^z$ dans (31) on obtient la fonction

caractéristique de la somme $r = x_1 + x_2 + \dots + x_n$, c'est-à-dire la fonction

$$(33) \quad \varphi_n(z) = \mathbb{E}e^{rz} = \mathbb{E}e^{(x_1 + x_2 + \dots + x_n)z} = \sum_{r=0}^n q_r \binom{n}{r} e^{rz} = \sum_{r=0}^n p_{r, n-r} e^{rz}$$

$$= \frac{\Gamma\left(\frac{1}{\delta}\right)}{\Gamma\left(\frac{\rho}{\delta}\right) \Gamma\left(\frac{\sigma}{\delta}\right)} \int_0^1 x^{\frac{\rho}{\delta}-1} (1-x)^{\frac{\sigma}{\delta}-1} [xe^z + (1-x)]^n dx.$$

La dernière expression de $\varphi_n(z)$ s'obtient à partir de (32) à l'aide de l'expression classique de l'intégrale eulérienne B par des transformations faciles que je n'explique pas ici.

VIII

Nous devons encore connaître la fonction caractéristique de $r = x_1 + x_2 + \dots + x_n$ pour la structure de l'« hérédité ». Pour gagner du temps, je cite ici sans démonstration la formule suivante de MARKOFF (13) : Si $p_{r,s}$ désigne la probabilité de tirer rouge r fois et noir s fois sur $n = r + s$ tirages ou, ce qui est la même chose, d'avoir $r = r$ on a

$$(34) \quad \sum_{r=0}^{\infty} \sum_{s=0}^{\infty} p_{r,s} u^r v^s = \frac{1 - \tau(\sigma u + \rho v)}{1 - (\rho + \sigma\tau)u - (\sigma + \rho\tau)v + \tau uv} = \mathcal{F}(u, v).$$

La fonction caractéristique cherchée est

$$(35) \quad \mathbb{E}e^{rz} = \sum_{r=0}^n p_{r, n-r} e^{rz} = \frac{1}{2\pi i} \oint \frac{\mathcal{F}(e^z w, w)}{w^{n+1}} dw,$$

la ligne d'intégration dans le plan de la variable complexe w étant un cercle parcouru dans le sens positif dont le centre est $w = 0$ et le rayon aussi petit que, z étant donné, la fonction analytique $\mathcal{F}(e^z w, w)$ de w soit régulière à l'intérieur et sur le cercle. Cette forme n'est pas très explicite, mais tout à fait suffisante pour les passages à la limite dont on aura besoin.

(13) Voir l. c. 7) b), p. 49, formule (10). On peut déduire de cette formule le cas limite traité par MARKOFF ; pour une méthode simple, voir l'indication qui sera donnée à la fin du n° IX.

IX

Rapprochons maintenant les deux structures d'interdépendance, celle de la « contagion » et celle de l' « hérédité » que nous avons traitées séparément jusqu'ici.

Considérons pour les deux structures la loi de la variable aléatoire r , donc la probabilité $p_{r,n-r}$ d'obtenir exactement r fois rouge sur n tirages, $r = 0, 1, 2, \dots, n$. Cette probabilité est déterminée par la fonction génératrice respective, (33) ou (35), même déterminée explicitement par (32) pour la première des deux structures. En portant les $n + 1$ probabilités

$$(36) \quad p_{0,n}, \quad p_{1,n-1}, \quad p_{2,n-2}, \quad \dots \quad p_{n,0}$$

comme ordonnées équidistantes dans un système rectangulaire, nous obtenons une distribution (une « courbe ») de probabilités ; elle dépend de 3 paramètres dans les deux cas, de ρ, δ, n , respectivement de ρ, τ, n . Comme en statistique nous n'attribuons de valeur qu'aux observations portant sur un grand nombre d'individus, il est naturel de chercher les formes limites résultant de cette distribution, quand n , le nombre des tirages, tend vers ∞ . Il y a différentes formes limites, la courbe de Gauss n'étant que la plus connue, puisque ρ et δ comme ρ et τ peuvent varier différemment avec n .

On peut classifier les formes limites de deux points de vue différents : d'abord d'après l'ordre de grandeur de ρ , c'est-à-dire d'après la fréquence de l'événement considéré, puis d'après l'ordre de grandeur de δ respectivement de τ , c'est-à-dire d'après le degré d'interdépendance entre les événements individuels.

Concernant la fréquence, je considère deux cas.

a) Je parle d'événements *usuels* si ρ garde une valeur invariable (entre 0 et 1, limites exclues) lorsque n tend vers l'infini. Dans ce cas $n\rho$, la valeur probable des tirages rouges, tend vers ∞ .

b) Je parle d'événements *rares* si la valeur probable $n\rho$ garde une valeur positive invariable quand $n \rightarrow \infty$. Dans ce cas ρ tend vers zéro.

Je distinguerai trois degrés d'interdépendance. C'est Q , le coefficient de dispersion qui sera la base essentielle de la classification.

1) Les événements *indépendants* sont caractérisés par $\delta = 0$, respectivement $\tau = 0$. Dans ce cas le coefficient de dispersion Q est égal à

l'unité et, pour les deux structures envisagées, la réciproque est vraie : $Q = 1$ implique $\delta = 0$, respectivement $\tau = 0$ en vertu de (13) et de (27).

2) Je parle de *contagion faible* si δ tend vers zéro, de manière que $n\delta$ garde une valeur invariable (différente de 0 et supérieure à -1) et de *faible ressemblance* si τ garde une valeur invariable ($\tau \neq 0$, $-1 < \tau < 1$) quand $n \rightarrow \infty$. Dans ces deux cas d'*interdépendance faible* Q tend vers une limite finie positive et différente de 1.

3) Je parle de *contagion forte* si δ garde une valeur positive invariable et de *forte ressemblance* si τ tend vers 1 de manière que $n(1 - \tau)$ garde une valeur positive invariable quand $n \rightarrow \infty$. Dans ces deux cas d'*interdépendance forte* $\frac{Q^2}{n}$ tend vers une limite positive finie.

Les trois degrés d'interdépendance ont été distingués d'après l'allure de Q , brièvement

$$1) Q = 1 \quad 2) Q \text{ fini} \quad 3) \frac{Q}{\sqrt{n}} \text{ fini.}$$

Nous aurions pu aussi les distinguer d'après l'allure du coefficient de corrélation entre x_1 et x_n , les deux tirages *les plus éloignés* ; c'est $\frac{\delta}{1 + \delta}$ respectivement τ^{n-1} . Nous avons dans les trois cas respectivement

$$1) \frac{\delta}{1 + \delta} = 0 \quad 2) \frac{\delta}{1 + \delta} \rightarrow 0 \quad 3) \frac{\delta}{1 + \delta} \text{ fini}$$

$$\tau^{n-1} = 0 \quad \tau^{n-1} \rightarrow 0 \quad \tau^{n-1} \text{ fini}$$

(en admettant $n(1 - \tau) = \lambda$, on a $\tau^{n-1} \rightarrow e^{-\lambda}$).

La manière qui est la plus commode au point de vue du calcul, de caractériser les cas 1), 2), 3) ainsi que *a)* et *b)* est celle adoptée dans leur définition. C'est d'assigner à un des paramètres ρ , δ ou τ ou à une combinaison simple d'un paramètre avec n une valeur constante quand $n \rightarrow \infty$. Je réunis ces constantes, caractéristiques des cas respectifs, avec la notation dont je me servirai plus tard, dans la tablelle suivante :

$$a) \rho \quad b) n\rho = h$$

$$1) \delta = 0 \quad 2) n\delta = d \quad 3) \delta$$

$$\tau = 0 \quad \tau \quad n(1 - \tau) = \lambda.$$

Observons à propos de *a)* et *b)* qu'il aurait été inutile d'introduire un cas auquel $\rho \rightarrow 1$ puisque l'on peut supposer $\rho \leq \sigma$. Observons à

SUR QUELQUES POINTS DE LA THÉORIE DES PROBABILITÉS

propos du cas 3) que $Q \rightarrow \infty$ implique $Q^2 > 1$ donc $\delta > 0$, $\tau > 0$. Observons encore qu'on pourrait introduire un quatrième degré d'interdépendance, celui de l'interdépendance « ultraforte » caractérisé par

$$\begin{aligned} 4) \quad \delta &= +\infty \\ \tau &= 1 \end{aligned}$$

ou par ce que le coefficient de corrélation $\frac{\delta}{1+\delta}$ respectivement τ^{n-1} est exactement 1. Ce cas est instructif comme terme de comparaison mais sans intérêt en lui-même : Le premier excepté, les coups sont forcés. Chaque tirage à partir du second suit nécessairement son prédécesseur, de manière que la probabilité d'avoir $r = 0$ est σ , celle d'avoir $r = n$ est ρ , tandis que celle d'avoir une valeur intermédiaire quelconque est 0.

D'après ce qui a été dit, il sera clair, ce qu'il faut entendre par le cas 1a) par exemple. C'est le cas où les conditions 1) et a) sont réunies, le cas d'événements usuels indépendants (ρ est fixé, $\delta = 0$, respectivement $\tau = 0$), le cas classique de BERNOULLI menant à la loi normale de GAUSS.

Le cas 3b) c'est-à-dire le cas d'événements rares à interdépendance forte est celui où

$$\rho = \frac{h}{n}, \quad \tau = 1 - \frac{\lambda}{n}$$

h, λ, δ sont des constantes positives. Dans ce cas on trouve facilement la distribution limite en envisageant $p_{0,n}$, la probabilité d'avoir $r = 0$ c'est-à-dire de ne tirer que des boules noires pendant les n premiers tirages. Pour la structure de la « contagion » on a d'après (32)

$$\begin{aligned} p_{0,n} &= \frac{\sigma(\sigma + \delta) \cdots (\sigma + (n-1)\delta)}{1(1+\delta) \cdots (1+(n-1)\delta)} \\ &= \left(1 - \frac{h}{n}\right) \left(1 - \frac{h}{n(1+\delta)}\right) \cdots \left(1 - \frac{h}{n[1+(n-1)\delta]}\right) \\ &> 1 - \frac{h}{n} \left(1 + \frac{1}{1+\delta} + \frac{1}{1+2\delta} + \cdots + \frac{1}{1+(n-1)\delta}\right) \rightarrow 1, \end{aligned}$$

pour celle de l'« hérédité » on trouve

$$\begin{aligned} p_{0,n} &= \sigma \sigma^{n-1} = (1-\rho) (1-\rho(1-\tau))^{n-1} \\ &= \left(1 - \frac{h}{n}\right) \left(1 - \frac{h\lambda}{n^2}\right)^{n-1} \rightarrow 1. \end{aligned}$$

Donc la distribution limite est la même pour les deux structures et peu intéressante comme distribution de probabilité : $r = 0$ a la probabilité 1 et toutes les autres valeurs de r la probabilité 0. Donc le résultat est le même que si tous les tirages donnaient forcément noir dès le commencement.

Pour les autres cas la distribution limite est plus intéressante mais moins facile à calculer. Prenons par exemple le cas 3a), le cas d'événements usuels à contagion forte, si nous nous bornons à la première des deux structures. Dans ce cas ρ et δ sont des constantes, $0 < \rho < 1$, $\delta > 0$. Introduisons la nouvelle variable aléatoire $x = \frac{r}{n}$ qui prend la valeur $\frac{r}{n}$ avec la même probabilité que r la valeur r . Sa fonction caractéristique est donc, d'après (33)

$$\mathbb{E}e^{xz} = \mathbb{E}e^{\frac{r}{n}z} = \varphi_n\left(\frac{z}{n}\right) = \frac{\Gamma\left(\frac{1}{\delta}\right)}{\Gamma\left(\frac{\rho}{\delta}\right)\Gamma\left(\frac{\sigma}{\delta}\right)} \int_0^1 x^{\frac{\rho}{\delta}-1} (1-x)^{\frac{\sigma}{\delta}-1} \left[xe^{\frac{z}{n}} + (1-x)\right]^n dx.$$

En observant que

$$\left[xe^{\frac{z}{n}} + (1-x)\right]^n = \left[1 + \frac{xz}{n} + \frac{xz^2}{2n^2} + \dots\right]^n \rightarrow e^{xz}$$

pour $n \rightarrow \infty$ on trouve la limite de la fonction caractéristique. On en déduit (14) que la distribution des probabilités de la variable aléatoire $x = \frac{r}{n}$ devient continue à la limite, la densité de probabilité au point $x = x$ étant

$$(37) \quad \frac{\Gamma\left(\frac{1}{\delta}\right)}{\Gamma\left(\frac{\rho}{\delta}\right)\Gamma\left(\frac{\sigma}{\delta}\right)} x^{\frac{\rho}{\delta}-1} (1-x)^{\frac{\sigma}{\delta}-1}.$$

C'est de cette manière, en partant des fonctions génératrices (33) et (35) qu'on peut calculer les distributions limites pour les deux

(14) Dans ce cas comme dans tous ceux qui figurent dans ces leçons, le théorème que j'ai donné sur la limite de la fonction caractéristique (*Mathematische Zeitschrift*, t. VIII (1920), p. 171-181) et le théorème analogue que M. Paul LÉVY a donné un peu plus tard (*Comptes-rendus*, t. 175 (1922) p. 854-856) sont également applicables.

SUR QUELQUES POINTS DE LA THÉORIE DES PROBABILITÉS

structures dans tous les cas qu'on va énumérer au n° suivant. L'exemple que j'ai traité est un des plus simples ; en particulier le calcul de (35) exige un choix judicieux de la ligne d'intégration. Mais je n'ai pas le temps de m'arrêter aux détails de ces calculs.

X

Les distributions de probabilités que nous allons considérer, résultent toutes de la loi de distribution de la variable aléatoire r . On se rappelle que r prend les valeurs

$$0, \quad 1, \quad 2, \quad \dots \quad n$$

avec les probabilités respectives

$$p_{0,n}, \quad p_{1,n-1}, \quad p_{2,n-2}, \quad \dots \quad p_{n,0}$$

et que r n'est autre chose que le nombre des boules rouges sorties aux n premiers tirages, le schéma d'urne pouvant être celui de la « contagion » ou celui de l'« hérédité ». Quelquefois nous devons introduire une nouvelle variable aléatoire $x = ar + b$, transformée linéaire de r avec des coefficients a, b qui dépendent de n ; x sera à la limite une variable aléatoire continue, c'est-à-dire qu'elle prendra toutes les valeurs d'un certain intervalle ; mais la loi de probabilité de x peut être discontinue. Quelquefois nous appellerons r la variable de la loi limite. Mais alors r pourra prendre des valeurs entières non-négatives quelconques

$$0, \quad 1, \quad 2, \quad \dots \quad m, \quad \dots$$

avec des probabilités qui seront désignées par

$$p_0, \quad p_1, \quad p_2, \quad \dots \quad p_m, \quad \dots$$

de manière qu'on ait

$$p_m = \lim_{n \rightarrow \infty} p_{m,n-m}.$$

Nous allons considérer 9 différents passages à la limite. Ce sont les cas 1a) 1b) 2a) 2b) 3a) expliqués au n° précédent, le cas t) ou cas de

transition qui sera expliqué plus loin et les cas 3b) 4a) et 4b), triviaux en eux-mêmes mais instructifs pour la comparaison (15).

1a) *Evénements usuels indépendants*. ρ constante, $0 < \rho < 1$, $\delta = \tau = 0$, les deux structures envisagées coïncident dans le cas particulier commun de l'indépendance. En posant

$$(38) \quad x = \frac{r - n\rho}{\sqrt{n\rho\sigma}}$$

x prendra à la limite toutes les valeurs de l'intervalle $-\infty, +\infty$. La densité de la probabilité au point $x = x$ est

$$\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

C'est la courbe normale ou de GAUSS. Ce résultat classique, la « Loi des grands nombres » est dû à DE MOIVRE.

1b) *Evénements rares indépendants*. La valeur probable $n\rho = h$ est constante, $\delta = \tau = 0$, c'est encore le cas d'indépendance, cas particulier commun aux deux structures. La probabilité p_m d'avoir $r = m$ à la limite est donnée comme coefficient de la fonction génératrice

$$(39) \quad \begin{aligned} E z^n &= p_0 + p_1 z + \dots + p_m z^m + \dots \\ &= e^{-h} + \frac{e^{-h} h}{1!} z + \dots + \frac{e^{-h} h^m}{m!} z^m + \dots = e^{h(z-1)}. \end{aligned}$$

Ce résultat est encore classique, c'est la « Loi des petits nombres » due à POISSON.

2a) *Evénements usuels à interdépendance faible*. ρ constante, $0 < \rho < 1$, $n\delta = d$ constante, τ constante, $\rho + d \geq 0$ à cause de (3) et les quatre expressions (29) sont positives. En introduisant x par (38), x varie à la limite de $-\infty$ à $+\infty$. La densité de probabilité au point $x = x$ est

$$\frac{1}{\sqrt{2\pi\mu}} e^{-\frac{x^2}{2\mu^2}};$$

$$\mu^2 = 1 + d \quad \text{ou} \quad \mu^2 = \frac{1 + \tau}{1 - \tau}$$

(15) Les cas 1a) 1b) sont classiques, ainsi qu'un cas particulier de 2a) (celui de la boule non remise). Pour les autres cas limites non triviaux voir : a) Structure de la « contagion » : tous les cas chez MARKOFF l. c. 8) et EGGENBERGER et PÓLYA l. c. 8) a) et b). Les courbes limites continues 3a) et t) furent rencontrées déjà avant par PEARSON ; voir n° XI. 3) Structure de l'« hérédité » : 2a) MARKOFF l. c. 7) a) b) ; 2b) A. AEPPLI, Thèse, Zurich 1924 ; 3a) et t) ont été indiqués la première fois, à ma connaissance, dans ces leçons.

SUR QUELQUES POINTS DE LA THÉORIE DES PROBABILITÉS

selon qu'il s'agit de « contagion » ou d'« hérédité ». C'est la courbe de GAUSS, comme au cas 1a), mais la dispersion diffère de la normale, le signe de la différence étant celui de d ou de τ , selon la structure.

2b) *Evénements rares à interdépendance faible.* $n\rho = h$, $n\delta = d$ et τ sont des constantes positives, $\tau < 1$. ($d < 0$ est exclu par (3), $\tau < 0$ par (21).) La probabilité p_m d'avoir $r = m$ à la limite est donnée par

$$(40) \quad \mathbb{E}z^r = p_0 + p_1z + \dots + p_mz^m + \dots = (1 - (z - 1)d)^{-\frac{h}{d}}$$

s'il s'agit de « contagion » et par

$$(41) \quad \mathbb{E}z^r = p_0 + p_1z + \dots + p_mz^m + \dots = e^{\frac{h(1-\tau)(z-1)}{(1-\tau z)}}$$

s'il s'agit d'« hérédité ».

3a) *Evénements usuels à interdépendance forte.* ρ , δ , $n(1 - \tau) = \lambda$ sont des constantes positives, $0 < \rho < 1$. On introduit la variable

$$(42) \quad x = \frac{r}{n}$$

qui prendra à la limite toutes les valeurs de l'intervalle $0 \leq x \leq 1$.

Pour la structure de la contagion on a une distribution limite continue ; nous avons calculé auparavant sa densité au point $x = x$, voir (37).

Pour la structure de l'« hérédité » on a une distribution limite en partie continue, en partie discontinue. Les valeurs extrêmes

$$x = 0, \quad x = 1$$

ont des probabilités finies, respectivement

$$\sigma e^{-\rho\lambda}, \quad \rho e^{-\sigma\lambda},$$

les valeurs intermédiaires n'ont que des probabilités infiniment petites, la densité de la probabilité étant une fonction continue possédant toutes les dérivées. On déduit ces propriétés d'une discussion appropriée de la fonction caractéristique

$$\mathbb{E}e^{xz} = e^{\frac{1}{2}(z-\lambda)} \left[\cosh\left(\frac{1}{2}\Delta\right) + (\lambda + (2\rho - 1)z) \frac{\sinh\left(\frac{1}{2}\Delta\right)}{\Delta} \right]$$

où

$$\Delta = \sqrt{z^2 + 2(2\rho - 1)\lambda z + \lambda^2}$$

et, comme d'usage,

$$2 \cosh x = e^x + e^{-x}, \quad 2 \sinh x = e^x - e^{-x}.$$

Je note encore ici le carré de l'écart de x ⁽¹⁶⁾.

$$E(x - \rho)^2 = \frac{2\rho\sigma}{\lambda^2} (e^{-\lambda} - 1 + \lambda).$$

3b) *Evénements rares à interdépendance forte.* $n\rho$, δ , $n(1 - \tau) = \lambda$ sont des constantes positives. Le cas a été traité à la fin du n° IX. La distribution limite peut être conçue de deux manières. On peut prendre r comme variable ; alors à la limite la probabilité d'avoir $r = 0$ est 1 et celle d'avoir $r = m$ est 0 si $m > 0$. Ou bien en posant $x = ar$, x variera d'une manière continue dans un intervalle fini ou infini, dont l'extrémité gauche est le point $x = 0$; ce point aura la probabilité 1, tous les autres points ensemble la probabilité 0.

4a) *Evénements usuels à interdépendance ultraforte.* ρ constante, $0 < \rho < 1$, $\delta = +\infty$ et $\tau = 1$, les coups sont forcés à partir du second ; les deux structures coïncident tout à fait dans ce cas particulier. x étant défini par (42), la probabilité d'avoir $x = 0$ est σ , celle d'avoir $x = 1$ est ρ et celle d'avoir $0 < x < 1$ est 0.

4b) *Evénements rares à interdépendance ultraforte.* $\rho = 0$, $\delta = +\infty$, $\tau = 1$, les coups sont forcés à partir du premier. Comme distribution limite c'est la même chose que 3b).

t) *Cas de transition.* Dans ce cas

$$n^\theta \rho, \quad n^\theta \delta, \quad n^{1-\theta}(1 - \tau)$$

sont des constantes positives ; θ est une fraction constante, $0 < \theta < \frac{1}{2}$ ⁽¹⁷⁾.

Ce cas est, aussi bien au point de vue de la fréquence qu'au point de vue de l'interdépendance, la transition entre les cas 3a) et 2b) comme on le voit du tableau suivant qui montre les puissances de n auxquelles $n\rho$ et Q^2 sont proportionnels.

$$\begin{array}{lll} n\rho : & a)n & t)n^{1-\theta} & b)n^0 \\ Q^2 : & 3)n & t)n^{1-\theta} & 2)n^0. \end{array}$$

(16) Pour une connexion possible de cette formule avec des questions de physique voir G. J. TAYLOR, Diffusion by continuous movements, Proceedings of the London M. S. 2^e série, t. XX (1921), p. 196-210, en particulier p. 198-201.

(17) On peut arriver à la même loi limite sous des conditions un peu moins restrictives.

SUR QUELQUES POINTS DE LA THÉORIE DES PROBABILITÉS

Posons

$$\frac{\rho}{\delta} = \rho n(1 - \tau) = \alpha,$$

$$\frac{r}{n\delta} = (1 - \tau) r = x.$$

Pour les deux structures, x variera à la limite continuellement de 0 à ∞ .

Pour la structure de la contagion, on a une distribution limite continue ; sa densité au point x est

$$\frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-x}.$$

Pour la structure de l'hérédité, on a une distribution limite en partie continue, en partie discontinue. La probabilité d'avoir $x = 0$ est finie, $= e^{-\alpha}$. Les valeurs positives de x n'ont qu'une probabilité infiniment petite, la densité de probabilité au point x étant

$$e^{-\alpha-x} \frac{\sqrt{\alpha} J_1(2\sqrt{\alpha x})}{i\sqrt{x}} = e^{-\alpha-x} \sum_{n=0}^{\infty} \frac{\alpha^{n+1} x^n}{(n+1)! n!}$$

où $J_1(x)$ désigne la fonction de BESSEL d'indice 1.

Je termine cette énumération par une table et un diagramme que voici.

Cas	Limites	p	\mathcal{P}
1a	$-\infty, \infty$	0	2
1b	—	1	—
2a	$-\infty, \infty$	1	2
2b	—	2	—
3a	0, 1	2	4
(3b)	—	0	—
(4a)	0, 1	1	3
(4b)	—	0	—
t	0, ∞	1	3

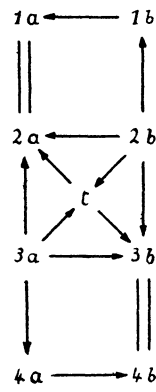


Figure 1.

La première colonne de la table contient le nom du cas limite. Il faut se rappeler que 1, 2, 3, 4 y désignent des degrés croissants d'interdépendance, a les événements usuels, b les rares. Les cas tri-

viaux sont entre parenthèses. La seconde colonne donne des limites de la variation de la variable aléatoire, si cette variation est continue ; un trait horizontal désigne les cas où cette variable ne prend que les valeurs 0, 1, 2, 3, ... ; les cas 3*b*) et 4*b*) sont traités ici comme discontinus. La 3-me colonne donne p , le nombre des paramètres de la distribution sous sa forme considérée dans l'énumération, et la 4-me colonne, pour les distributions à variable continue, le nombre total \mathcal{P} des paramètres qu'on peut y introduire en faisant usage d'une transformation linéaire arbitraire.

Dans le diagramme toutes les distributions, aussi les discontinues, sont considérées comme admettant une transformation linéaire arbitraire de la variable aléatoire. Alors la différence entre 1*a*) et 2*a*) disparaît ; c'est pour cela qu'ils sont réunis par un signe = (mis verticalement) de même que 4*b*) et 3*b*). La flèche allant de 2*b*) à 1*b*) indique que 1*b*) est cas limite de 2*b*). En effet, en faisant $d \rightarrow 0$ dans (40) ou en mettant $\tau = 0$ dans (41) on obtient les deux fois (39). Inversement, la distribution 2*b*) qui contient 2 paramètres ne peut pas être cas limite de 1*b*) qui n'en contient qu'un. Les autres flèches ont une signification analogue. Le cas 3*b*) est considéré ici comme continu. Il est à remarquer que la courbe de GAUSS, 1*a*) ou 2*a*), est cas limite de 1*b*) 2*b*) *t*) et 3*a*), donc de toutes les autres distributions non triviales. Qu'on observe aussi la situation de cas *t*), intermédiaire entre 3*a*) et 2*b*) mais, d'une autre manière, aussi entre 3*a*) et 2*a*).

C'est l'analogie des deux structures de la « contagion » et de « l'hérédité » qui me paraît le plus remarquable en tout ceci. Elle est tellement complète qu'en construisant la table et le diagramme, il était superflu de séparer les structures.

XI

Les résultats énumérés au n° précédent ouvrent la voie à d'autres recherches. Je dois me borner à quelques indications.

1. L'énumération des cas limites est-elle *complète* ? Je ne saurais donner une réponse définitive (18). J'esquisserai une réponse partielle

(18) En tous cas il manque le cas où la variable aléatoire X étant déterminée par (42), $X = \rho$ a la probabilité 1 à la limite. Ce cas limite, trivial en soi-même, peut être atteint d'une manière intéressante, par exemple lorsqu'il y a une « opposition extrême » dans la structure de l'« hérédité » :

$$\rho = \sigma = \frac{1}{2}, \quad \tau = -1, \quad Q^2 = \frac{1 - (-1)^n}{2n} \rightarrow 0.$$

en me servant d'une modification de la méthode par laquelle M. Karl PEARSON a obtenu primitivement les distributions qui proviennent ici, aux cas 3a) et t), de la structure de la « contagion ». Pour cette structure, on obtient de (32), en écrivant pour abrégé p_r à la place de $p_{r,n-r}$,

$$(43) \quad \frac{p_{r+1} - p_r}{p_{r+1}} = - \frac{(1 - 2\delta)(r + 1) - (n + 1)(\rho - \delta)}{(r\delta + \rho)(n - r)},$$

$$(44) \quad \frac{p_r - p_{r-1}}{p_{r-1}} = - \frac{(1 - 2\delta)r - (n + 1)(\rho - \delta)}{r(n\delta + \sigma - r\delta)}.$$

Si nous considérons r comme variable, le dénominateur de la fraction dans le second membre est un polynôme de second degré dont les racines sont réelles :

$$r = -\frac{\rho}{\delta}, \quad n \quad \text{pour (43),}$$

$$r = 0, \quad n + \frac{\sigma}{\delta} \quad \text{pour (44).}$$

Le domaine de variabilité de la variable aléatoire r est compris entre 0, première racine pour (44) et n , seconde racine pour (43).

Une transformation linéaire $x = ar + b$ changera (43) et (44) en équations analogues avec des propriétés correspondantes des racines. Admettons que dans un cas limite où la variable aléatoire est x , il y ait une densité de probabilité $y = \varphi(x)$ qui est une fonction dérivable de x ; admettons en outre que la convergence vers le cas limite soit « suffisamment bonne ». Alors (43) et (44) *coïncideront* à la limite dans une même équation de la forme

$$(45) \quad \frac{1}{y} \frac{dy}{dx} = \frac{a_0 + x}{b_0 + b_1x + b_2x^2}$$

où le dénominateur $b_0 + b_1x + b_2x^2$ n'a pas de racines imaginaires. Observons que le degré du dénominateur peut s'abaisser à la limite si l'une ou l'autre des racines devient infinie. On voit que le domaine de variation de x est compris entre les deux racines, entre ∞ (ou $-\infty$) et l'unique racine ou enfin entre $-\infty$ et $+\infty$ selon que le degré du dénominateur est 2, 1, ou 0.

Faisant abstraction d'une transformation linéaire, on peut admettre que le dénominateur du second membre de (45) est $x(1 - x)$ ou x ou 1.

Ces trois cas figurent effectivement dans l'énumération du n° X. On obtient en partant de (43) ou de (44)

$$\frac{1}{y} \frac{dy}{dx} = \left(\frac{\rho}{\delta} - 1 \right) \frac{1}{x} + \left(\frac{\sigma}{\delta} - 1 \right) \frac{1}{x-1} \quad \text{au cas 3a)}$$

$$\frac{1}{y} \frac{dy}{dx} = -1 + \frac{x-1}{x} \quad \text{au cas t)}$$

$$\frac{1}{y} \frac{dy}{dx} = -\frac{x}{1+d} \quad \text{au cas 2a).}$$

Les racines de $b_0 + b_1x + b_2x^2$ étant supposées réelles, une discussion facile des intégrales y de l'équation (45) montre que, abstraction faite toujours d'une transformation linéaire, ces trois cas 3a) t) 2a) sont les seuls où $\int y dx$ converge entre les limites admissibles (racines et $\pm \infty$ comme spécifié ci-dessus).

D'après cela, en tant qu'il s'agit des distributions limites *continues* de la structure de la « contagion », notre énumération est complète.

2. Comme il a été dit, les considérations précédentes ne sont qu'une modification de la première déduction que M. PEARSON a donnée de ses courbes de fréquence bien connues provenant de (45) ⁽¹⁹⁾.

Je n'ai pas à discuter ici, bien entendu, l'intérêt des applications que M. PEARSON et son école ont faites de ces courbes. Tout ce que je dirai se rapportera à la première introduction citée, entre laquelle et le présent traitement, il convient de signaler les différences suivantes:

(a) Je considère ici (43) et (44) à la place du rapport $\frac{2(\hat{p}_{r+1} - \hat{p}_r)}{\hat{p}_{r+1} + \hat{p}_r}$ considéré par M. PEARSON.

(b) Le point de départ est ici le schéma d'urnes de la « contagion », schéma plus général que celui de la boule non-remise dont s'est servi M. PEARSON.

(c) Les cas considérés ici sont des cas limites ; pour chacun d'eux j'ai indiqué soigneusement de quelle manière il faut varier les paramètres du schéma initial pour y arriver et quelle est la signification du passage à la limite pour la structure des probabilités. Les cas considérés par M. PEARSON proviennent de l'équation (45), bien sûr, mais ils ne proviennent pas pour cela d'une structure de probabilités, puisque M. PEARSON fait varier les paramètres d'une façon

(19) Philosophical Transactions, t. 186 (1895), p. 343-414.

SUR QUELQUES POINTS DE LA THÉORIE DES PROBABILITÉS

quelconque, sans se soucier, comme d'ailleurs il le dit explicitement lui-même, si ses formules restent ou ne restent pas susceptibles d'interprétation dans les termes du problème initial.

La dernière remarque explique pourquoi nous n'avons pas obtenu ici toutes les courbes de M. PEARSON bien que nous soyons partis d'un schéma d'urnes plus général ⁽²⁰⁾.

3. Notre déduction donne à chacune des courbes limites obtenues une signification, la rattache à une structure d'interdépendance qu'on pourra, avec une approximation plus ou moins grande, reconnaître au matériel statistique fourni par l'observation.

Un exemple suffira pour montrer le genre d'application que j'ai en vue. Envisageons la statistique d'une épidémie rare, comme la petite vérole ou le choléra le sont de nos jours. Les individus observés sont les habitants d'un pays. On enregistre les cas mortels par suite de l'épidémie par mois. A la fin d'une certaine période on aura constaté M_0 mois sans cas mortels, M_1 avec exactement 1, ... M_r avec exactement r cas mortels. Il s'agit de donner une expression théorique à laquelle la suite

$$(46) \quad M_0, M_1, M_2, \dots, M_r, \dots$$

pourra être assimilée.

Observons d'abord que l'épidémie est *rare*. Donc il est naturel d'examiner si la suite (46) est proportionnelle ou non à la suite des probabilités

$$(47) \quad p_0, p_1, p_2, \dots, p_r, \dots$$

données par une des formules (39) (40) (41) relatives aux événements rares (cas 1b) et 2b)).

Observons ensuite que les cas d'épidémie ne semblent pas être indépendants : un mois avec beaucoup de cas est un mauvais présage pour le suivant. Cela élimine à peu près *a priori* la formule (39) du cas 1b) d'événements rares indépendants.

Reste à décider entre les deux formules (40) et (41) du cas 2b) ; la première provient de la structure de la « contagion », la seconde de celle de l'« hérédité ». Dans la première structure l'influence mutuelle mesurée par le coefficient de corrélation est la même entre deux indi-

⁽²⁰⁾ Que les autres courbes de M. PEARSON ne sont pas cas limites du schéma d'urnes de la contagion, a été montré encore d'une autre manière, plus rigoureuse par M. EGGENBERGER.

vidus quelconques de la population, dans la seconde l'influence mutuelle décroît rapidement avec la distance croissante des individus (supposés rangés en série linéaire). On dirait que la réalité est quelque part entre les deux structures. Si l'on envisage les schéma d'urnes, c'est la première structure qui semble être plus près de la réalité (voir le commencement du n° III).

Les calculs numériques exécutés ⁽²¹⁾ montrent que (39) (indépendance) est décidément en désaccord avec les faits; (41) (« hérédité ») donne une approximation meilleure et (40) (« contagion ») va encore mieux. S'il s'agit de la statistique de cas mortels par suite de certains accidents industriels, l'indépendance est encore en désaccord, et les deux structures de la « contagion » et de l'« hérédité » fournissent des approximations à peu près égales.

4. Il y a des explications plausibles, il n'y en a pas d'obligatoires, au moins pas en statistique. Un exemple frappant de ce principe est donné par la structure de la « contagion » au cas 3a); si $\rho = \sigma = \delta = \frac{1}{2}$, la densité de probabilité correspondante, donnée par (37) devient égale à 1 pour $0 < x < 1$, c'est-à-dire que la distribution de probabilités devient uniforme dans l'intervalle 0,1. Personne n'a trouvé nécessaire d'expliquer une répartition uniforme de probabilité par quelque chose d'autre, mais nous voyons qu'elle *peut* être expliquée par une structure compliquée d'interdépendance. Sans doute elle pourrait être expliquée encore d'une infinité d'autres manières ⁽²²⁾.

Voici un autre exemple. Si on rencontrait des séries d'événements alternatifs et homogènes montrant une dispersion anormale, mais réparties d'après la loi de GAUSS, la statistique correspondrait à notre cas 2a). On pourrait l'expliquer tout aussi bien en supposant une structure de « contagion » qu'une structure d'« hérédité », bien que ces structures soient différentes; mais elles mènent, dans le cas 2a), à la même répartition des fréquences.

5. Les deux structures envisagées, celle de la « contagion » et celle de l'« hérédité » sont différentes, mais leurs cas limites montrent un parallélisme remarquable, comme on a vu aux n°s IX et X. Dans

(21) Pour les calculs avec (39) et (40) voir EGGENBERGER *l. c.* 8). Les calculs concernant (41) ne sont pas encore publiés.

(22) Des remarques analogues s'appliquent au cas $\sigma = \rho = \delta$, n fini, plus intéressant encore à certains égards.

SUR QUELQUES POINTS DE LA THÉORIE DES PROBABILITÉS

quels cas ce parallélisme va-t-il, malgré la différence des structures, jusqu'à la coïncidence des lois limites ?

Il faut écarter ici le cas de l'indépendance ainsi que celui de l'interdépendance ultraforte, donc les cas

$$1a) \quad 1b) \quad 4a) \quad 4b)$$

parce que dans ces cas particuliers communs les deux structures ne sont pas effectivement différentes. Elles sont effectivement différentes mais donnent la même loi limite dans les cas

$$2a) \quad 3b)$$

et donnent des lois limites différentes dans les cas

$$3a) \quad t) \quad 2b)$$

Laissons encore de côté le cas 3b), trivial et pas trop différent de 4b). Alors le cas 2a) reste le seul cas où la même loi limite a pu s'imposer aux deux structures, malgré leur différence effective, et cette loi limite commune est la loi de GAUSS.

Ce fait me semble symptomatique ; il nous fait entrevoir que la loi de GAUSS jouit d'une situation exceptionnelle non seulement dans la théorie des probabilités indépendantes, mais aussi dans celle des interdépendantes. En effet, en poursuivant les recherches de MARKOFF, M. Serge BERNSTEIN a réussi à montrer dans un travail très important⁽²³⁾ que c'est la loi de GAUSS qui doit résulter de la superposition d'un grand nombre d'événements interdépendants, pourvu que l'interdépendance soit suffisamment « faible ». (Je n'ai énoncé qu'une des conditions et d'une manière assez vague.)

Du résultat de M. BERNSTEIN on ne peut pas conclure qu'en dehors de la loi de GAUSS il n'y a pas d'autres lois qui puissent s'imposer en même temps à des structures d'interdépendance très différentes. On ne peut pas conclure cela des exemples traités dans ce travail non plus, mais c'est ce qui est rendu plausible par tout ce qui a été dit. Il serait important de préciser ces remarques et de trouver quelque propriété de la loi de GAUSS qui caractérise sa place dans la théorie des probabilités interdépendantes comme elle a été caractérisée dans celle des indépendantes. Ce problème est encore bien obscur ; toutefois j'ai tenu à le signaler avant de terminer ces leçons.

(23) *l. c.*, II).