60 fr

LEÇONS PROFESSÉES A L'OFFICE NATIONAL MÉTÉOROLOGIQUE DE FRANCE

R. DE MONTESSUS DE BALLORE

DOCTEUR ÈS SCIENCES
LAURÉAT DE L'INSTITUT

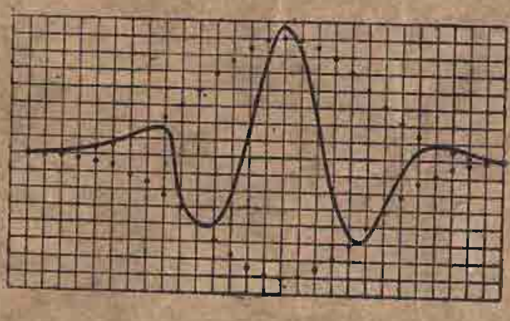
PROBABILITÉS ET STATISTIQUES

PRÉFACE

DE

M. ALLIAUME

PROFESSEUR
A L'UNIVERSITÉ DE LOUVAIN



PARIS

LIBRAIRIE SCIENTIFIQUE HERMANN ET C^{ie}

6, RUE DE LA SORBONNE, 6

1931

PROBABILITÉS
ET
STATISTIQUES

OUVRAGES DU MÊME AUTEUR

Leçons élémentaires sur le Calcul des probabilités. 20 fr.

Exercices et leçons de mécanique analytique professées à la Faculté Libre des Sciences de Lille. . . 30 fr.

Leçons sur les fonctions elliptiques, Cours libre professé à la Faculté des Sciences de Paris. . . . . 35 fr.

Introduction à la théorie des courbes gauches algébriques *(rare)*, Cours libre professé à la Faculté des Sciences de Paris. . . . . 40 fr.

M. Alliaume a bien voulu écrire la Préface de ce livre : je l'en remercie de tout cœur. Il a suivi avec attention mes travaux, sur le Calcul des probabilités ; dans une circonstance particulièrement difficile, il m'a donné un conseil de haute valeur : il était particulièrement désigné pour présenter cet ouvrage au lecteur.

Mes remerciements vont aussi à M. F. J. Duarte qui, à un talent d'analyste incontesté, révélé dans ses *Tables de factorielles*, joint une maîtrise rare pour les calculs numériques, dont j'ai tiré le plus grand profit.

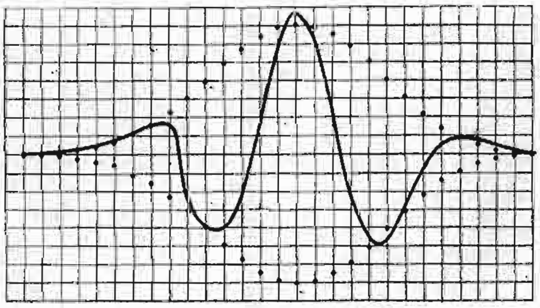
R. M. B.

Septembre 1930.

R. DE MONTESSUS DE BALLORE
DOCTEUR ÈS SCIENCES
LAURÉAT DE L'INSTITUT

PROBABILITÉS
ET
STATISTIQUES

PRÉFACE
DE
M. ALLIAUME
PROFESSEUR
A L'UNIVERSITÉ DE LOUVAIN



PARIS
LIBRAIRIE SCIENTIFIQUE HERMANN ET Cie
6, RUE DE LA SORBONNE, 6

1931

PRÉFACE

M. de Montessus de Ballore me fait l'honneur de me demander d'écrire la préface de son nouveau livre, *Probabilités et Statistiques*.

Sans doute le devoir du préfacier est-il principalement de faire apprécier la portée, l'utilité, l'opportunité de l'ouvrage qu'il présente.

L'intervention de M. de Montessus de Ballore dans la statistique mathématique est assez heureuse pour que cette tâche soit extrêmement facile.

En vue des applications du Calcul des probabilités à la Statistique, le problème essentiel de ce Calcul est le suivant.

Dans une épreuve, un événement a une probabilité p, dont le complément à l'unité est q. On a l'intention de faire m épreuves. Un nombre x est proposé, auquel on donne le nom d'écart, moindre que mp et que mq, et tel que $mp - x$ (et par conséquent aussi $mq + x$) soit un nombre entier. On demande la probabilité y que le nombre de fois que l'événement arrive soit $mp - x$.

La solution rigoureuse de ce problème est

$$y = \frac{m!}{(mp - x)!\,(mq + x)!}\, p^{mp-x} q^{mq+x}.$$

Les calculs de factorielles ont été considérés jusqu'ici comme trop pénibles. Aussi les théoriciens introduisent-ils dans l'énoncé de ce problème certaines hypothèses simplificatrices.

1° Le nombre m d'épreuves est assez grand pour que des factorielles telles que $m!$, $(mp)!$, $(mq)!$, et par conséquent $(mp - x)!$,

$(mq + x)!$, puissent être calculées par la formule approchée de Stirling,

$$n! = n^n e^{-n} \sqrt{2\pi n} \ ;$$

ceci suppose que, eu égard à l'approximation adoptée, p et q sont suffisamment proches de $\frac{1}{2}$. Cette première hypothèse simplificatrice transforme y dans

$$y' = \frac{1}{\sqrt{2\pi mpq}} \left(1 - \frac{x}{mp}\right)^{-mp + x - \frac{1}{2}} \left(1 + \frac{x}{mq}\right)^{-mq - x - \frac{1}{2}}.$$

2° Le premier ordre de petitesse étant celui du rapport $\frac{1}{\sqrt{m}}$, et le rapport $\frac{x}{m}$ étant assez petit pour que $\frac{1}{x}$ ne soit pas une quantité petite, on se contente d'obtenir y aux quantités petites du quatrième ordre près. Cette deuxième hypothèse simplificatrice réduit y' à

$$y'' = \frac{1}{\sqrt{2\pi mpq}} e^{\frac{x(q-p)}{2mpq} - \frac{x^2}{2mpq}}.$$

3° Les auteurs cherchent à justifier une nouvelle simplification de manière à pouvoir écrire

$$y''' = \frac{1}{\sqrt{2\pi mpq}} e^{-\frac{x^2}{2mpq}},$$

ce qui est la formule de Laplace.

Leur raisonnement semble être le suivant.

Au moyen de l'écart relatif λ défini par l'égalité

$$m\lambda = x,$$

l'exposant de e dans y'' a pour valeur

$$\frac{\lambda(q-p)}{2pq} - \frac{m\lambda^2}{2pq},$$

et le nombre m est assez grand pour que le terme qui ne renferme pas m soit négligeable devant celui qui le contient. Mais ce raisonnement ne tient pas compte de la petitesse de $\frac{x}{m} = \lambda$ qui se présente au second degré dans le terme conservé et seulement au premier degré dans le terme supprimé.

L'expression classique y''' n'est donc légitime (et moyennant les deux premières hypothèses simplificatrices) que lorsque p et q sont extrêmement voisins de $\frac{1}{2}$.

Nous voici ramenés à l'expression y''. Mais celle-ci n'est pas beaucoup plus commode que la formule rigoureuse y, et les deux premières hypothèses simplificatrices ne constituaient guère qu'un artifice didactique par lequel nous nous laisserions mener de l'expression rigoureuse à la formule de Laplace.

C'est à cet artifice, qui vient de perdre presque tout avantage, que renonce M. de Montessus de Ballore. Il montre comment il n'y a pas lieu de se laisser rebuter par le calcul des factorielles, et que c'est la loi rigoureuse, la loi binomiale, qui s'applique efficacement à l'étude des statistiques.

En quoi consiste donc, dans le sens qui lui est donné ici, l'étude d'une statistique ?

L'observation nous met entre les mains, sous la forme d'un tableau en deux colonnes, une collection de nombres, x et Y, se correspondant deux à deux. N étant la somme des nombres Y, il s'agit de trouver une équation

$$\mathrm{Y} = \mathrm{N} \cdot f(x, a, b, c, \cdots)$$

à laquelle, à une approximation donnée et déclarée satisfaisante, et moyennant certaines valeurs numériques de paramètres a, b, c,... aussi peu nombreux que possible, satisfassent les couples (x, Y) des nombres du tableau proposé.

Certes, dans son extrême simplicité, et avec son seul paramètre mpq, l'équation de Laplace

$$\mathrm{Y} = \frac{\mathrm{N}}{\sqrt{2\pi mpq}}\, e^{-\frac{x^2}{2mpq}}$$

est bien tentante à appliquer. Mais, — l'auteur de ce livre en fait la preuve, — la statistique ne se plie à pareille simplicité que dans des circonstances tout à fait exceptionnelles : parmi de multiples conditions, ne faut-il pas que, dans la mise en diagramme du tableau statistique, les points représentatifs des couples de nombres (x, Y) se distribuent le long d'une ligne admettant une droite de symétrie parallèle à l'axe des Y ?

Au contraire, c'est à la loi binomiale

$$Y = N \frac{m!}{(mp - x)!\, mq + x)!} p^{mp-x} q^{mq+x},$$

avec ses trois paramètres m, p, q, qu'obéissent (pour une position convenable de l'origine des x) les statistiques les plus intéressantes, les plus instructives, celles dont l'étude est vraiment efficace, — du moins dans l'état actuel des ressources mathématiques dont nous disposons.

Dans quelle mesure considérable nous devons ces ressources à M. de Montessus de Ballore, on le verra dans ce livre, où l'auteur a réuni, sous une forme didactique, les travaux qu'il a fait paraître, depuis sept ans, principalement dans les *Annales de la Société scientifique de Bruxelles*. — Il était temps de les réunir. Cependant la question est loin d'être épuisée, et de nouveaux problèmes apparaissent, si intéressants, et si complexes, que leur énoncé et leur classification sont déjà bien méritoires.

Lorsque la statistique obéit à la loi binomiale, le calcul des valeurs qu'y prennent les paramètres de cette loi constitue toute son étude, et nous possédons désormais des méthodes sûres pour ce calcul. On ose même envisager les causes des faits observés en disant qu'alors « un seul phénomène est en jeu », tandis que, si quelques mesures font exception, un « phénomène principal est en jeu » ainsi que « des phénomènes secondaires ».

Au contraire si la loi binomiale est insuffisante, on dit bien que « des phénomènes du même ordre d'intensité se superposent », mais il reste, précisément, à les dégager les uns des autres, en déterminant les fonctions qui expriment les effets de chacun d'eux et en calculant les valeurs actuelles des paramètres que ces fonctions renferment. — Par un essai, qui réussit, de superposition de deux lois binomiales, M. de Montessus de Ballore montre que telle est bien la voie des prochaines recherches.

M. Alliaume,
Professeur à l'Université de Louvain.

PROBABILITÉS ET STATISTIQUES

CHAPITRE PREMIER

PROBABILITÉS DIRECTES

I. — PRINCIPE DE BERNOULLI

1. Une urne contient des boules, toutes identiques les unes aux autres, sauf que les unes sont *rouges* et les autres sont *noires*.

L'urne renferme n boules noires et r boules rouges.

On tire une boule, après s'être bandé les yeux, pour s'interdire le choix entre une noire et une rouge ; on note la couleur de la boule sortie ; on remet la boule sortie dans l'urne ; on mélange les boules ; on tire une seconde boule, on la remet après avoir noté sa couleur, on tire une troisième boule, etc...

Après N coups, on a tiré n' noires et r' rouges,

$$n' + r' = \mathrm{N}.$$

Ici, nous allons énoncer un principe :

Principe de Bernoulli : *Le rapport* $\frac{n'}{r'}$ *est, en gros, d'autant plus voisin de* $\frac{n}{r}$, *que* N *ou* $n' + r'$ *est plus grand.*

2. Le mot « *en gros* », *que nous choisissons à dessein,* veut dire que la différence

$$\frac{n}{r} - \frac{n'}{r'}$$

est, *le plus souvent*, moindre que

$$\frac{n}{r} - \frac{n''}{r''}$$

quand

$$n' + r' > n'' + r'' ;$$

autrement dit,

$$\frac{n}{r} - \frac{n'}{r'}$$

tend vers zéro, mais tend *irrégulièrement* vers zéro quand $n' + r'$ augmente indéfiniment.

3. $\frac{n}{r} - \frac{n'}{r'}$ tend vers zéro, mais $n' - r'$ augmente indéfiniment : *cela n'a rien de contradictoire* : par exemple les nombres

$$124, \quad 1\,248, \quad 12\,497, \quad 124\,996, \; \cdots$$

s'éloignent progressivement de

$$125, \quad 1\,250, \quad 12\,500, \quad 125\,000, \; \cdots$$

mais les différences

$$\frac{1}{8} - \frac{124}{1\,000}, \quad \frac{1}{8} - \frac{1\,248}{10\,000}, \quad \frac{1}{8} - \frac{12\,497}{100\,000}, \; \cdots$$

diminuent de plus en plus.

4. *Le principe de Bernoulli peut être regardé* soit *comme une loi naturelle,* soit *comme un postulat,* soit *comme une définition,* au même titre que, par exemple, le principe de l'action et de la réaction.

On peut le démontrer, explicitement ou implicitement, en le remplaçant par un principe équivalent.

II. — DÉFINITION DE LA PROBABILITÉ

5. *Cas favorables.* — Pour une raison ou pour une autre, j'ai intérêt à tirer une boule *noire* de l'urne : le tirage d'une boule noire est alors appelé « événement favorable », ou « *cas favorable* ».

6. Probabilité. — Par définition, *la probabilité est le rapport du nombre des cas favorables au nombre des cas possibles.*

Dans l'urne de tout à l'heure, le nombre des boules noires est n : il y a n cas favorables *si* l'on désire tirer une boule noire. Le nombre total des boules est $n + r$: il y a $n + r$ cas possibles. **Par définition**, la probabilité de tirer une noire est

$$\frac{n}{n + r}.$$

7. Soient deux événements E, E′ *exclusifs*, c'est-à-dire que, seules, deux alternatives sont possibles : E se produit, ou bien E′ se produit, tel est le cas de boules de deux couleurs contenues dans une urne : la boule tirée est rouge, c'est l'événement E, ou bien elle est noire, c'est l'événement E′.

Si p est la *probabilité* de E, c'est que N étant le nombre de cas favorables et P le nombre de cas possibles, on a

$$p = \frac{\mathrm{N}}{\mathrm{P}}.$$

N est par exemple le nombre de boules noires, P est le nombre total de boules noires et rouges, c'est-à-dire que s'il a en tout N boules noires et R boules rouges, $\mathrm{P} = \mathrm{N} + \mathrm{R}$.

La probabilité de tirer une rouge est, désignons-la par q,

$$q = \frac{\mathrm{R}}{\mathrm{P}} = \frac{\mathrm{P} - \mathrm{N}}{\mathrm{P}} = 1 - \frac{\mathrm{N}}{\mathrm{P}}:$$

on voit que

$$q = 1 - p \qquad \text{ou} \qquad p + q = 1.$$

q est la *probabilité complémentaire* de p et p est la *probabilité complémentaire* de q.

8. *Nombre probable d'arrivées d'un événement.* — Si la probabilité d'un événement E est p, et si l'on fait N épreuves (si l'on

tire N boules de l'urne, successivement), le nombre probable d'arrivées de E est

$$Np.$$

Par exemple, si une urne renferme 6 boules noires et 7 rouges, la probabilité de sortie d'une noire est $\frac{6}{13}$; si l'on tire une boule 26 fois de suite, en la remettant chaque fois et en mélangeant les boules, le nombre probable de sorties d'une noire est

$$26 \times \frac{6}{13} = 12.$$

Cela veut dire que si l'on fait 500 tirages de 26 boules chacun, — au lieu de 500 on pourrait choisir un nombre quelconque, mais de préférence assez grand — le nombre de tirages où il y aura 12 boules noires pour 26 boules tirées, sera ordinairement *plus grand* que le nombre de tirages où il y aura 11 noires pour 26 tirées, 13 noires pour 26 tirées, etc.

9. *Probabilité nulle, certitude.* — Si l'urne, au lieu de renfermer, comme tout à l'heure, des boules rouges et des boules noires, ne renferme plus que des boules rouges, c'est-à-dire si $n = 0$ la probabilité de tirer une noire, $\frac{0}{0+r}$ est nulle; comme il est impossible ici de tirer une noire, *la probabilité* **zéro** *correspond à l'impossible.*

D'autre part, dans les mêmes conditions, il y a certitude de tirer une rouge et la probabilité de tirer une rouge est $\frac{r}{0+r} = 1$: *la probabilité* **un** *correspond donc à la certitude.*

Parconséquent la probabilité est une fraction positive, comprise entre zéro et un, et pouvant atteindre les limites zéro (impossibilité) *et un* (certitude).

Il n'y a pas lieu de chercher à interpréter les probabilités extérieures à l'ensemble des nombres compris entre zéro et un.

III. — PROBABILITÉS DIRECTES. LEUR CALCUL EFFECTIF

10. Toutes les fois qu'on peut évaluer exactement les nombres de cas possibles et de cas favorables, leur rapport est une *probabilité* que nous appellerons *directe.*

L'évaluation des nombres de cas possibles et des nombres de cas favorables est plus ou moins facile.

Exemple I : *Cas d'un jeu de 32 cartes.* — Le nombre des cas possibles est 32, puisqu'il y a 32 cartes.

Si la sortie d'un as est un cas favorable, le nombre des cas favorables est 4, puisqu'il y a 4 as.

La probabilité de tirer un as est donc $\frac{4}{32}$ ou $\frac{1}{8}$.

Exemple II : *Le jeu de dés.* — Deux dés sont dans un cornet. Chacun des deux dés est marqué des points 1, 2, 3, 4, 5, 6 sur ses faces. On jette les deux dés sur une table. Quelle est la probabilité d'amener un point fixé à l'avance, par exemple le point 9 ?

Les combinaisons possibles sont au nombre de 36 ; les voici :

$1+1=2$	$2+1=3$	$3+1=4$	$4+1=5$	$5+1=6$	$6+1=7$
$1+2=3$	$2+2=4$	$3+2=5$	$4+2=6$	$5+2=7$	$6+2=8$
$1+3=4$	$2+3=5$	$3+3=6$	$4+3=7$	$5+3=8$	$6+3=9$
$1+4=5$	$2+4=6$	$3+4=7$	$4+4=8$	$5+4=9$	$6+4=10$
$1+5=6$	$2+5=7$	$3+5=8$	$4+5=9$	$5+5=10$	$6+5=11$
$1+6=7$	$2+6=8$	$3+6=9$	$4+6=10$	$5+6=11$	$6+6=12$

Le point 5 peut être amené par 1 et 4, 2 et 3, 3 et 2, 4 et 1 ; sa probabilité est $\frac{4}{36}$; le point 8 peut être amené par 2 et 6, 3 et 5, 4 et 4,5 et 3, 6 et 2, sa probabilité est $\frac{5}{36}$. La probabilité du point 2 est $\frac{1}{36}$ seulement. La probabilité du point 9 est aussi $\frac{4}{36}$.

IV. — COMBINAISONS DE PROBABILITÉS ÉLÉMENTAIRES

Il est nécessaire de savoir combiner des probabilités, comme on sait combiner par exemple des Forces, en Mécanique.

A) Probabilité totale

11. *Si les cas favorables à un événement se partagent en plusieurs groupes,* les groupes réunis contenant tous les cas favorables, et figurant une seule fois dans l'ensemble des groupes, *la probabilité de* l'événement, dite *probabilité totale, sera la somme des probabilités pour qu'il appartienne à chaque groupe.*

En effet, les probabilités partielles, relatives à chaque groupe, sont des fractions de même dénominateur

$$\frac{a_1}{D}, \quad \frac{a_2}{D}, \quad \frac{a_3}{D}, \ldots \frac{a_n}{D}.$$

La probabilité résultante, que l'on appelle **probabilité totale** est donc

$$\frac{a_1 + a_2 + \cdots a_n}{D}.$$

Exemple.— Supposons qu'une urne renferme n boules blanches, n' boules rouges et n'' boules noires, avec un nombre quelconque de boules d'autres couleurs, de manière que le nombre total de boules soit N.

Quelle est la probabilité de tirer une boule de l'une des trois couleurs blanche, rouge, noire, indifféremment ?

Le nombre des cas favorables est $n + n' + n''$. Le nombre des cas possibles est N.

La probabilité cherchée est donc

$$\frac{n + n' + n''}{N} \quad \text{ou} \quad \frac{n}{N} + \frac{n'}{N} + \frac{n''}{N}.$$

On peut adopter cet énoncé : *lorsqu'un événement peut se produire dans diverses hypothèses dont les probabilités sont différentes, la probabilité de cet événement est la somme des probabilités des hypothèses favorables* et la probabilité de l'événement est, par définition, une *probabilité totale.*

Voici un autre exemple de probabilité totale.

Quelle est la probabilité, en jetant deux dés, numérotés chacun de 1 à 6 sur leurs faces, d'amener un point total supérieur à 8 ?

Les cas favorables sont les sorties des points 9, 10, 11, 12. D'après le Tableau du n° 10,

la probabilité de sortie du point	9	est	$\frac{4}{36}$;
« «	10		$\frac{3}{36}$;
« «	11		$\frac{2}{36}$;
« «	12		$\frac{1}{36}$;

la probabilité cherchée est donc

$$\frac{4}{36}+\frac{3}{36}+\frac{2}{36}+\frac{1}{36}=\frac{4+3+2+1}{36}=\frac{10}{36}.$$

B) Probabilité composée

12. *La probabilité* composée *se rapporte à la coïncidence de plusieurs événements favorables,* **indépendants les uns des autres.**

Tirer un roi d'un jeu de cartes, et amener le point 7 avec deux dés, est la coïncidence de deux événements indépendants ; la probabilité de cette coïncidence est dite *composée.*

La situation dans le temps de l'événement final n'entre pas en ligne de compte. La probabilité est la même :

que l'on tire d'abord une carte et que l'on jette ensuite les dés,

que l'on jette d'abord les dés et que l'on tire ensuite une carte,

que par un dispositif quelconque, l'on tire une carte, en jetant en même temps les dés.

Soit p' la probabilité du premier événement, et soit p'' la probabilité du second événement. On a

$$p'=\frac{N'}{P'},\qquad p''=\frac{N''}{P''};$$

N′, N″ sont les nombres de cas favorables, et P′, P″ sont les nombres de cas possibles.

Le nombre total de cas possibles est P′ P″, car à *chaque cas* possible de la série P′ on peut adjoindre tous les cas possibles, au nombre de P'', de la deuxième série. De même, le nombre total de cas favorables est N' N''. Donc la probabilité composée p est

$$p = \frac{N'N''}{P'P''}.$$

Ainsi, pour l'exemple choisi, il y a 32 cartes et (nº 10) 36 cas possibles avec deux dés :

$$P' = 32 \qquad P'' = 36 \qquad P'P'' = 32 \times 36\,;$$

il y a 4 rois et (nº 10) 6 manières d'amener le point 7 :

$$N' = 4 \qquad N'' = 7 \qquad N'N'' = 4 \times 7\,;$$

donc

$$p = \frac{N'N''}{P'P''} = \frac{4 \times 7}{32 \times 36}.$$

Il résulte de l'identité suivante

$$p = \frac{N'N''}{P'P''} = \frac{N'}{P'} \times \frac{N''}{P''} = p'p''$$

que *la probabilité composée est le produit des probabilités composantes.*

Ceci étant vrai de deux catégories d'événements, le sera aussi de plusieurs catégories.

Quand les événements simples sont indépendants, le nombre N des cas résultants possibles est le produit des nombres de cas possibles

$$P',\ P'',\ P''',\ \cdots$$

de chaque catégorie. De même, le nombre P de cas résultants favorables est le produit des nombres de cas favorables

$$N',\ N'',\ N''',\ \cdots$$

de chaque catégorie.

La probabilité composée est donc

$$\frac{N' \cdot N'' \cdot N''' \cdot \ldots}{P' \cdot P'' \cdot P''' \cdot \ldots} = \frac{N'}{P'} \times \frac{N''}{P''} \times \frac{N'''}{P'''}, \cdots$$

Nous nous bornons à ces indications sommaires sur la probabilité *totale* et la probabilité *composée*, qui sont étudiées en grand détail dans les ouvrages concernant le Calcul des Probabilités. Ce que nous en disons ici sufit aux applications usuelles de la statistique.

C) Probabilité renforcée

13. *La probabilité totale, la probabilité composée ne sont pas les seules probabilités résultantes de probabilités simples : il existe d'autres résultantes de probabilités simples.*

La **probabilité renforcée** apparaît dans le problème suivant.

Pierre est atteint d'une maladie où se manifestent des symptômes S *et des symptômes* S'.

Quand les symptômes S se manifestent seuls, la probabilité de guérison est p : on la suppose connue *a priori*. La probabilité de non guérison est donc $1 - p$; il est en effet certain (probabilité 1) qu'il y aura guérison ou non guérison et on a

$$p + (1 - p) = 1.$$

On sait aussi *a priori* que : quand les symptômes S' se manifestent seuls, la probabilité de guérison est p'.

On demande *quelle est la probabilité π de guérison quand les symptômes* S *et* S' *se manifestent ensemble.*

Considérons N cas.

Les symptômes S se présentent dans un nombre de cas que nous désignons par NP où P est inconnu (n° 8).

Les symptômes S et S' se manifestent ensemble dans un nombre de cas que nous désignons par NPP', où P' est inconnu.

Pour ces NPP' cas, où S et S' se manifestent ensemble, la probabilité de guérison résulte de S et de S' ; puisque S et S' sont indépendants, elle est (n° 12)

$$pp'.$$

De même, la probabilité de non guérison pour ces NPP' cas où S et S' se manifestent ensemble est

$$(1 - p)(1 - p').$$

Il n'y a pas lieu de faire entrer en ligne de compte la probabilité

$$p(1 - p')$$

de guérison d'après les symptômes S et de non guérison d'après les symptômes S′ ; on ne fera pas non plus entrer en ligne de compte la probabilité

$$(1 - p)p'$$

de non guérison d'après les symptômes S, mais de guérison d'après les symptômes S′.

Le nombre probable des cas de guérison n° **8**,

$$\mathrm{NPP'}pp'$$

d'une part, et le nombre probable des cas de non guérison d'autre part,

$$\mathrm{NPP'}(1 - p)(1 - p')$$

ont pour somme

$$\mathrm{NPP'}[pp' + (1 - p)(1 - p')] = \mathrm{NPP'}(pp' + qq').$$

Par conséquent, la probabilité de guérison quand S et S′ se manifestent ensemble est

$$\frac{\mathrm{NPP'}pp'}{\mathrm{NPP'}(pp' + qq')} ;$$

donc

$$(1) \qquad \pi = \frac{pp'}{pp' + qq'}.$$

C'est l'expression de la **probabilité renforcée.**

Elle se présente dans certains phénomènes météorologiques et peut-être, dans des maladies causées par des associations de microbes.

Ecrivons

$$\pi(p, p') = \frac{pp'}{pp' + (1 - p)(1 - p')} = \frac{pp'}{1 - (p + p') + 2pp'}.$$

Zéro est le signe de l'impossibilité (n° **9**) ; si $p' = 0$, la probabilité de guérison est nulle quand les symptômes S′ se manifestent, que les symptômes S se manifestent ou non. On doit donc avoir

$$\pi(p, 0) = 0 :$$

il en est bien ainsi.

De même si $p = 0$, on doit avoir — et on a en effet — :

$$\pi(0, p') = 0.$$

On doit encore avoir, puisque **1** est le signe de la certitude (nº **9**).

$$\pi(pp') + \pi(1 - p, 1 - p') = 1 :$$

il est facile de vérifier qu'il en est ainsi [1].

D) Autres probabilités résultantes

14. *Il existe d'autres probabilités résultantes.*

Voici un exemple.

On sait que : quand un nuage A, d'aspect donné, vient de l'W, la probabilité de pluie est p ; on sait aussi que : quand un nuage B, d'aspect donné encore, vient du S, la probabilité de pluie est p'.

Quelle est la probabilité de pluie quand les deux nuages sont simultanément en vue ?

La probabilité totale

$$p + p'$$

comprend les cas suivants :

A déclanche la pluie à lui seul ; B déclanche la pluie à lui seul ; A et B déclanchent la pluie par action de l'un sur l'autre ; ce dernier cas, dont la probabilité est

$$pp'$$

rentre dans les deux précédents ; donc $p + p'$ renferme pp' ; pp' doit être retranché de $p + p'$, pour ne pas être compté deux fois ; la probabilité cherchée est donc

$$p + p' - pp'. \tag{2}$$

On a

$$q = 1 - p ; \qquad q' = 1 - p' ;$$

par conséquent

$$p + p' - pp' = 2 - q - q' - (1 - q)(1 - q') = 1 - qq' ;$$

[1] La probabilité renforcée, *Annales Soc. Scient. de Bruxelles*, t. XLIII, 1re partie, 1923.

le produit de $q < 1$ par $q' < 1$ est plus petit que 1 ; donc

$$0 < 1 - qq' < 1$$

et, comme il est nécessaire,

$$0 < p + p' - pp' < 1.$$

Nous arrêtons ici cette question des probabilités résultantes : *elle mérite d'être approfondie.* Nous en reparlerons au n° **16.**

V. — APPLICATIONS

15. M. L. Besson a dressé les Tableaux suivants [1] qui sont des statistiques portant sur environ 1.600 journées des mois de décembre, janvier et février, où il ne pleuvait pas à 9 heures du matin.

Tableau A

Pression en mm.	Nombre de cas de pluie	Nombre de cas total	Prob. p' de pluie
725	1	1	1
730	7	11	0,64
735	18	21	0,86
740	44	60	0,73
745	88	121	0,73
750	119	197	0,60
755	153	293	0,52
760	138	325	0,42
765	89	299	0,30
770	36	188	0,19
775	6	26	0,23
780	0	4	0

Dans le tableau A, sont notées le nombre de journées où le baromètre a marqué par exemple 730 millimètres à 9 heures du

(1) *Annales de l'Observatoire municipal*, Gauthier-Villars, Paris, 1905.

matin, et où il a plu avant minuit, ainsi que le quotient $\frac{7}{11} = 0{,}64$ qui est la probabilité de pluie correspondante.

TABLEAU B

Direction du vent	Nombre de cas de pluie	Nombre de cas total	Prob. p de pluie
N	13	55	0,24
NNE	25	75	0,33
NE	24	96	0,25
ENE	12	79	0,15
E	17	69	0,25
ESE	19	56	0,34
SE	32	86	0,37
SSE	46	115	0,40
S	84	146	0,58
SSW	89	131	0,68
SW	110	148	0,74
WSW	66	88	0,75
W	57	104	0,55
WNW	34	96	0,46
NW	26	65	0,40
NNW	11	37	0,30
Calme	24	100	0,24

Dans le tableau B, on a compté 69 journées où le vent venait de l'est à 9 heures du matin ; sur ces 69, il en est 17 où il a plu avant minuit ; la probabilité de pluie est ici $\frac{17}{69} = 0{,}25$.

Dans le tableau C, les chiffres $\frac{2}{23}$ par exemple indiquent qu'on a relevé 23 journées où le vent venait à 9 heures du matin d'ENE et où, en même temps, le baromètre marquait 770 millimètres ; sur ces 23 journées, on en a compté 2 où il a plu avant minuit ; la probabilité de pluie $\frac{2}{23}$ n'a pas été indiquée en décimales dans le Tableau.

Les probabilités *observées* du Tableau C concordent assez bien

TABLEAU C

	N	NNE	NE	ENE	E	ESE	SE	SSE	S	SSW	SW	WSW	W	WNW	NW	NNW
775mm et au-dessus	.	.	1	.	1	.	2	.	.	.	.	.	1	1	.	.
	2	4	6	.	2	1	3	1	.	.	1	.	3	4	.	.
770	0	3	5	2	2	0	2	3	1	5	2	3	0	4	1	1
	3	12	19	23	18	4	11	8	10	14	8	3	6	11	11	7
765	1	5	1	.	.	.	4	5	13	8	9	8	9	10	5	2
	11	18	19	16	10	6	16	25	29	17	17	14	24	19	17	5
760	.	5	6	4	1	7	7	8	13	24	13	13	12	11	7	1
	13	14	20	17	13	15	23	24	28	30	20	19	21	26	11	7
755	4	6	7	3	3	2	4	9	23	16	31	15	15	8	8	1
	7	16	20	13	10	12	15	20	35	23	40	19	23	18	3	7
750	5	4	2	1	3	5	5	11	17	11	25	7	5	8	3	5
	11	9	7	7	9	12	8	18	23	13	27	10	10	13	7	7
745	1	2	2	1	5	2	6	7	9	15	11	10	8	1	4	1
	4	2	3	1	5	3	8	12	13	19	13	12	10	3	8	2
740	2	.	.	1	1	2	2	1	6	8	9	6	3	1	2	.
	4	.	2	2	2	2	2	4	6	10	9	7	3	2	2	1
735 et au-dessous	.	.	.	.	.	1	.	2	2	5	6	5	4	.	1	.
	.	.	.	.	.	1	.	3	2	8	9	5	4	.	1	.

avec les probabilités renforcées déduites des tableaux A et B [1].

Par exemple, colonne S et ligne **755**,

$$\text{prob.} = \frac{23}{35} = 0{,}66\,;$$

dans A, pour **755**,

$$p = 0{,}51\,;$$

dans B, pour S,

$$p' = 0{,}58\,;$$

ici

$$pp' = 0{,}30, \qquad qq' = (1-p)(1-p') = 0{,}31,$$

$$\pi = \frac{0{,}30}{0{,}30+0{,}21} = \frac{0{,}30}{0{,}51} = 0{,}59 \quad \text{(à comparer avec 0,66 observé)}$$

Puis, colonne SSW et ligne **745**

$$\text{prob.} = \frac{15}{19} = 0{,}79\,;$$

dans A, pour **745**,

$$p = 0{,}73\,;$$

dans B, pour SSW,

$$p' = 0{,}68\,;$$

$$pp' = 0{,}50\,; \qquad qq' = 0{,}27 \times 0{,}32 = 0{,}09,$$

$$\pi = \frac{0{,}50}{0{,}50+0{,}09} = \frac{0{,}50}{0{,}59} = 0{,}81 \quad \text{(à comparer avec 0,79 observé).}$$

La conséquence est la suivante :

Si la statistique B *concorde avec les nombres déduits de* A *et* B *comme nous venons de le faire, le vent et la hauteur barométrique sont deux conséquences d'un même phénomène.*

Sinon, p et p' seraient peut-être combinés par la formule

$$p + p' - pp'\,;$$

or dans le 1er cas,

$$p + p' - pp' = 0{,}79, \text{ au lieu de } 0{,}66 \text{ obs.}$$

(1) Sur la prévision locale du Temps, *C. R. des Acad. de Paris*, t. CLXXVI, 4 juin 1923 ; *ibid.*, 25 juin 1923.

et dans le second cas,

$$p + p' - pp' = 0{,}91, \text{ au lieu de } 0{,}79 \text{: obs.}$$

Ce n'est donc pas cette dernière formule qui est en jeu, *le vent et la hauteur barométrique ne sont pas deux phénomènes indépendants l'un de l'autre.*

Si la probabilité renforcée est reconnue applicable, *par calcul direct* (*p. e.* pour les nombres *élevés* du Tableau C, comme nous l'avons reconnu pour deux cas), elle permet de substituer des nombres obtenus par le calcul aux nombres observés *faibles*, donc peu sûrs, du même tableau C.

Ainsi pour **N** et **740**$^{\text{mm}}$, le tableau C donne **2** et **4**, nombres très faibles, qui conduisent à la probabilité incertaine $p'' = \frac{2}{4} = \mathbf{0{,}5}$. Mais le tableau A donne $p = \mathbf{0{,}73}$ pour **740**$^{\text{mm}}$ et le tableau B donne $p' = \mathbf{0{,}24}$ pour **N**. On en déduit, *avec beaucoup de sûreté par le tableau* C :

$$p'' = \frac{0{,}73 \times 0{,}24 + 0{,}27 \times 0{,}76}{0{,}73 \times 0{,}24} = 0{,}47 :$$

0,47 est, plus exactement que **0,5**, la probabilité de pluie, dans les conditions où s'est placé **M. Besson**, quand le vent est **N** et le baromètre à **740**$^{\text{mm}}$. Il y aurait lieu d'étudier le cas, où trois phénomènes (ou quatre, ou plus) P_1, P_2, P_3 sont tels que les probabilités résultantes de P_1 et P_2, de P_2 et de P_3 soient soumises à la probabilité renforcée, et la possibilité de *calculer* la probabilité finale, P_1, P_2, P_3 entrant simultanément en jeu, car il serait alors difficile de dresser un tableau **D** en triple entrée, où les observations tiendraient compte des trois phénomènes.

Le statisticien pourra rechercher si la loi de combinaison de p et de p' est de la forme

$$(3) \quad \frac{pp'}{pp' + qq'} + \theta\left(p + p' - pp' - \frac{pp'}{pp' + qq'}\right), \qquad 0 < \theta < 1 ;$$

cette formule devient

$$\frac{pp'}{pp' + qq'}, \qquad p + p' - pp'$$

pour $\theta = \mathbf{0}$, $\theta = \mathbf{1}$;

s'il trouve pour θ une valeur non nulle, c'est qu'il y aura dépendance partielle entre les deux phénomènes :

hauteur barométrique, direction du vent.

16. Voici un exemple de statistique où la composition des probabilités partielles ne peut être faite ni par la probabilité composée, ni par la probabilité renforcée, ni par une combinaison simple de l'une et de l'autre.

Statistique des mariages suivant le sexe et l'âge combinés des époux en 1926, en France (1)

Age des époux	Age des épouses								Totaux
	Moins de 20 ans	20 24 ans	25 29 ans	30 34 ans	35 39 ans	40 49 ans	50 59 ans	60 ans et plus	
Moins de 20 ans.	2.893	1.952	206	38	9	—	2	—	5.100
20 à 24 »	35.801	86.360	16.194	2.366	468	127	10	1	141.327
25 à 29 »	16.478	58.531	29.066	7.180	1.754	547	21	4	113.581
30 à 34 »	1.811	10.589	11.212	6.684	2.356	934	57	8	33.651
35 à 39 »	362	2.770	4.756	4.959	3.128	1.627	164	9	17.775
40 à 49 »	151	1.053	2.560	4.480	4.671	5.579	1.069	89	19.652
50 à 59 »	21	131	454	941	1.528	4.192	2.337	382	9.986
60 ans et plus.	6	31	89	195	289	1.110	1.609	1.014	4.343
Totaux........	57.523	161.417	64.537	26.843	14.203	14.116	5.269	1.507	345.415

D'après ce Tableau, la probabilité concernant les mariages d'hommes de **30** à **34** ans est par exemple

$$p = \frac{33\,651}{345\,415} = 0{,}099\,;$$

La probabilité concernant les mariages de jeunes filles ou femmes de **25** à **29** ans par exemple est

$$p' = \frac{64\,537}{345\,415} = 0{,}188\,;$$

(1) *Statistique générale de la France. Annuaire statistique*, 1928, Imprimerie, Nationale, p. 9.

La probabilité concernant les mariages d'hommes de 30 à 34 ans et de jeunes filles ou femmes de 25 à 29 ans est

$$p'' = \frac{11\,212}{345\,415} = 0{,}032.$$

Ici

$$pp' = 0{,}018\,;$$

puis

$$qq' = 0{,}732\,;$$

donc

$$\frac{pp'}{pp' + qq'} = \frac{0{,}018}{0{,}732} = 0{,}025\,:$$

la probabilité p'' ne résulte donc de p et p' ni comme probabilité totale ni comme probabilité renforcée ; ce n'est pas non plus une combinaison linéaire de l'une et l'autre ; si l'on pose en effet

$$pp' + \theta\left(\frac{pp' + qq'}{pp'} - pp'\right) = 0{,}032,$$
$$0{,}018 + \theta \times 0{,}007 = 0{,}032,$$

on a

$$\theta = \frac{0{,}014}{0{,}007} > 1\,;$$

or, on ne peut faire de telles combinaisons que si

$$0 < \theta < 1.$$

Enfin p'' n'est pas non plus une combinaison de l'espèce (3), car en prenant θ pour inconnue, on trouve (3)

$$\theta = 0{,}029\,;$$

or, si l'on part du nombre 4 959 du tableau, si l'on prend encore θ de la formule (3) pour inconnue, on trouve ici

$$\theta = 0{,}075,$$

valeur nettement différente de la précédente.

CHAPITRE II

PROBABILITÉ DANS LES ÉPREUVES RÉPÉTÉES

I. — FORMULE DE BASE

17. *Les probabilités respectives de deux évènements contradictoires* E, E′ *étant* p, p', *quelle est la probabilité que, dans une série de* m *épreuves, l'évènement* E *se produira* α *fois et que l'évènement* E′ *se produira* $m - \alpha$ *fois ?*

La probabilité que la première épreuve amènera E est p ; la probabilité que E sera suivi de E′ est la probabilité composée (nº 12) pq ; la probabilité de la succession E, E′, E est pqp. Plus généralement, la probabilité d'amener α fois E et $m - \alpha$ fois E′, *quand on fixe lesordres d'arrivées* de E, E′, est

$$p^\alpha q^{m-\alpha} ;$$

si rien n'est fixé quant à cet ordre, le nombre des cas possibles, c'est-à-dire le nombre des suites de m événements dont α sont de l'espèce E et $m - \alpha$ de l'espèce E′, étant le nombre de combinaisons possibles

$$\frac{m\,!}{\alpha\,!\,(m-\alpha)\,!},$$

de m objets dont α sont E et dont $m - \alpha$ sont E', la probabilité cherchée T_α est

$$T_\alpha = \frac{m!}{\alpha!\,(m-\alpha)!} p^\alpha q^{m-\alpha}. \tag{1}$$

Par exemple, les probabilités sur 2, 4, 6, 8, 10 coups joués à la roulette d'amener aussi souvent la rouge que la noire sont

$$\frac{2!}{1!\,1!} \times \frac{1}{2^2} = \frac{1}{2}; \quad \frac{4!}{2!\,2!} \times \frac{1}{2^4} = \frac{3}{8}; \quad \frac{6!}{3!\,3!} \times \frac{1}{2^6} = \frac{5}{16};$$

$$\frac{8!}{4! \times 4!} \times \frac{1}{2^8} = \frac{35}{128}; \quad \frac{10!}{5!\,5!} \times \frac{1}{2^{10}} = \frac{63}{256},$$

ou

$$\frac{128}{256}, \quad \frac{96}{256}, \quad \frac{80}{256}, \quad \frac{70}{256}, \quad \frac{63}{256}:$$

on voit que la probabilité décroît quand le nombre d'épreuves croît.

18. La probabilité (1) peut s'écrire

$$T_\alpha = C_m^\alpha p^\alpha q^{m-\alpha}:$$

T_α est le terme en $p^\alpha q^{m-\alpha}$ du développement de $(p + q)^\alpha$.

Par conséquent, si on envisage toutes les valeurs possibles de α,

$$\Sigma T_\alpha = (p + q)^\alpha = 1^\alpha, \qquad \Sigma T_\alpha = 1. \tag{2}$$

Ce résultat peut être généralisé :

Les probabilités respectives de n événements s'excluant les uns les autres, E_1, E_1, ..., E_n, *étant* p_1, p_1, ..., p_n (*il est nécessaire que* $p_1 + p_1 + ... + p_n = 1$), *la probabilité que dans une suite de m épreuves,* E_1, *se produira* α_1 *fois,* E_2 : α_2 *fois* ; E_3 : α_3 fois ; ... ; E_n : α_n *fois* ($\alpha_1 + \alpha_2 + ... + \alpha_n = m$) *est donnée par le terme en* $p_1^{\alpha_1}\, p_2^{\alpha_2}\, p_3^{\alpha_3} \ldots p_n^{\alpha_n}$ *du développement de*

$$(p_1 + p_2 + \cdots + p_n)^m. \tag{3}$$

19. La formule (1) est la base de la théorie des épreuves répétées.

On peut l'écrire

$$T_\alpha = C_m^{m-\alpha} \times p^\alpha q^{m-\alpha}; \tag{4}$$

II. — NOMBRE D'ARRIVÉES LE PLUS PROBABLE

20. Comparons $T_{\alpha-1}$, T_α, $T_{\alpha+1}$. On a (n° **17**, formule **1**)

$$T_{\alpha-1} = \frac{m\,!}{(\alpha-1)\,!\,(m-\alpha+1)\,!}\,p^{\alpha-1}q^{m-\alpha+1}$$

$$T_\alpha = \frac{m\,!}{\alpha\,!\,(m-\alpha)\,!}\,p^\alpha q^{m-\alpha}$$

$$T_{\alpha+1} = \frac{m\,!}{(\alpha+1)\,!\,(m-\alpha-1)\,!}\,p^{\alpha+1}q^{m-\alpha-1}.$$

Donc

$$\frac{T_\alpha}{T_{\alpha-1}} = \frac{m-\alpha+1}{\alpha}\times\frac{p}{q},\qquad \frac{T_\alpha}{T_{\alpha+1}} = \frac{\alpha+1}{m-\alpha}\times\frac{q}{p}.$$

Ecrivons que T_α est plus grand, à la fois, que $T_{\alpha-1}$ et $T_{\alpha+1}$. Il vient,

$$\frac{m-\alpha+1}{\alpha}\times\frac{p}{q}>1,\qquad \frac{\alpha+1}{m-\alpha}\times\frac{q}{p}>1,$$

ou puisque $q = 1 - p$,

$$\frac{m-\alpha+1}{\alpha}\times\frac{p}{1-p}>1,\qquad \frac{\alpha+1}{m-\alpha}\times\frac{1-p}{p}>1,$$

ou encore

$$mp+p-1<\alpha<mp+p$$

ce qui revient à

$$(1)\qquad mp-q<\alpha<mp+p.$$

On remarquera que les deux limites de α diffèrent *exactement* de 1 unité puisque

$$mp+p-(mp-q) = p+q = 1.$$

Il résulte de (1) *que le nombre* α(α est forcément entier) *d'arrivées le plus probable de* E (dont la probabilité est p), *quand on fait* m *épreuves, est le nombre entier compris entre* $mp-q$ et $mp+p$.

21. On doit évidemment avoir par symétrie et semblablement (formule 1), quand on considère les événements **E'**,

$$(2) \qquad mq - p < m - \alpha < mq + q,$$

car p et q sont échangés et α est changé en $m - \alpha$. Cette double inégalité (2) est en effet une conséquence de (1) qu'on peut écrire successivement

$$m(1 - q) + p > \alpha > m(1 - q) - q$$
$$m - mq + p > \alpha > m - mq - q$$
$$- mq + p > - m + \alpha > - mq - q$$
$$mq - p < m - \alpha < mq + q.$$

22. Soit par exemple

$$m = 80, \qquad p = 0{,}33, \qquad q = 0{,}67\ ;$$

les formules (1, 2) donnent, ce qui revient au même :

$$25{,}73 < \alpha < 26{,}73$$
$$53{,}27 < 80 - \alpha < 54{,}27.$$

La probabilité *maximum* est que **E** aura lieu **26** fois et que **E'** aura lieu **80 — 26** ou **54** fois. Cette probabilité est

$$\frac{80\,!}{26\,!\,54\,!} \times 0{,}33^{26} \times 0{,}67^{54} = 0{,}0944.$$

Les probabilités

$$\frac{80\,!}{25\,!\,55\,!}\,0{,}33^{25} \times 0{,}67^{55} = 0{,}0944 \times \frac{25}{55} \times \frac{67}{33} = 0{,}0944 \times \frac{1742}{1815},$$
$$\frac{80\,!}{27\,!\,53\,!}\,0{,}33^{27} \times 0{,}67^{53} = 0{,}0944 \times \frac{54}{27} \times \frac{33}{57} = 0{,}0944 \times \frac{1782}{1809}$$

que **E** aura lieu **25** fois (et **E'** **55** fois), que **E** aura lieu **27** fois (et **E'** **53** fois) sont moindres que la précédente.

23. *Distinction entre* $p < q$ et $p > q$.

I. $p < q$. Soit par exemple $m = 364$, $p = 0{,}1$, $q = 0{,}9$; on a

$$mp = 36{,}4, \qquad mp - q = 35{,}5, \qquad mp + p = 36{,}5\ ;$$

le nombre d'arrivées le plus probable de E est (n° **20**) le nombre entier **36** compris entre **35,5** et **36,5** :

$$36 = mp - 0{,}4 = 36{,}4 - 0{,}4.$$

II. $p > q$. Reprenons $m = \mathbf{364}$, mais soit $p = \mathbf{0{,}9}$ et $q = \mathbf{0{,}1}$. **Ici,**

$$mp = 327{,}6, \qquad mp - q = 327{,}5, \qquad mp + p = 328{,}5\ ;$$

le nombre d'arrivées le plus probable est **328**. Ici

$$328 = mp + 0{,}4 = 327{,}6 + 0{,}4.$$

En conséquence, *si mp n'est pas un nombre entier, le nombre d'arrivées le plus probable est le nombre entier* $mp = s$ *défini comme il suit* :

1° : s est *positif* si $p > q$; s est *négatif si* $p < q$.

2° : s est compris entre -1 et $+1$;

Cas de $p = q$ *et de m pair* : le nombre d'arrivées le plus probable est $\frac{m}{2}$.

Cas de $p = q$ *et de m impair* :

il y a ici deux nombres d'arrivées les plus probables ; ce sont

$$\frac{m}{2} + 0{,}5 \quad \text{et} \quad \frac{m}{2} - 0{,}5$$

24. Ces distinctions sont importantes.

Néanmoins, *nous aurons bien rarement à nous inquiéter si* pm *est entier ou n'est pas entier.*

La considération de boules tirées d'une urne, qui nécessite l'adjonction d'une fraction s à mp quand mp n'est pas entier, ne figure ici qu'à titre d'introduction à l'étude de la fonction de α

$$\frac{m\,!}{\alpha\,!\,(m-\alpha)\,!}\,p^{\alpha}q^{m-\alpha}. \tag{3}$$

III. — LA FONCTION DE PROBABILITÉ SIMPLE

25. La notation (1 du n° 17) est incommode.

Posons

$$\alpha = mp - x \text{ d'où } m - \alpha = m - (mp - x) = m(1 - p) + x = mq + x.$$

(3 n° 24) devient

$$\frac{m\,!}{(mp - x)\,!\,(mq + x)\,!}\,p^{mp - x}q^{mq + x},$$

et nous allons considérer la fonction

$$(4) \qquad y_x = \frac{m\,!}{(mp - x)\,!\,(mq + x)\,!}\,p^{mp - x}q^{mq + x},$$

que nous appellerons **Fonction de probabilité simple** on la désigne aussi sous le nom de *fonction binomiale;* mais son rôle est *si important* que nous préférons la première dénomination.

Dans le cas de boules rouges et noires contenues dans une urne, dans la proportion p et q, y_x est la probabilité de tirer $mp - x$ boules rouges et $mq + x$ boules noires, m étant le nombre de tirages et mp étant un nombre entier.

Si mp n'est pas entier, nous lui adjoignons, quand nous voulons spécifier qu'il s'agit de boules tirées d'une urne, *cas exceptionnel,* la fraction s définie au n° 23 et nous écrivons

$$(4') \qquad y_x = \frac{m\,!}{(mp - s - x)!\,(mq + s + x)!}\,p^{mp - s - x}q^{mq + s + x} \begin{cases} s > 0 \text{ si } p < q \\ s < 0 \text{ si } p > q \end{cases}.$$

Dans (4) et (4′), x **a les valeurs entières** $0, \pm 1, \pm 2, \pm 3, \ldots\ldots$ soumises bien entendu à l'une des conditions

$$0 \leqslant mp - x \text{ et } 0 \leqslant mq + x\,; \quad 0 \leqslant mp - s - x \text{ et } 0 \leqslant mq + s + x.$$

La probabilité maximum est, dans le cas de (4), (mp et mq entiers)

$$y_0 = \frac{m\,!}{mp\,!\,mq\,!}\,p^{mp}q^{mq}$$

et dans le cas (4′) :

$$y_0 = \frac{m\,!}{(mp - s)\,!\,(mq + s)\,!}\,p^{mp - s}q^{mq + s}.$$

Les valeurs de y voisines de y_0 sont dans le cas de (4)

$$y_{-2} = \frac{m\,!}{(mp + 2)\,!\,(mq - 2)\,!}\,p^{mp+2} \times q^{mq - 2}\,;$$

$$y_{-1} = \frac{m\,!}{(mp + 1)\,!\,(mq - 1)\,!}\,p^{mp + 1} \times q^{mq - 1}\,;$$

$$y_0 = \frac{m\,!}{mp\,!\,mq\,!}\,p^{mp} \times q^{mq}\,;$$

$$y_1 = \frac{m\,!}{(mp-1)\,!\,(mq+1)\,!}\,p^{mp-1} \times q^{mq+1}\,;$$

$$y_2 = \frac{m\,!}{(mp-2)\,!\,(mq+2)\,!}\,p^{mp-2} + q^{mq+2}\,;$$

et dans le cas de (4′) les valeurs voisines de y_0 sont :

$$y_{-2} = \frac{m\,!}{(mp-s+2)\,!\,(mq+s-2)\,!}\,p^{mp-s+2} \times q^{mq+s-2}\,;$$

$$y_{-1} = \frac{m\,!}{(mp-s+1)\,!\,(mq+s-1)\,!} \times p^{mp-s+1} \times q^{mq+s-1}\,;$$

$$y_0 = \frac{m\,!}{(mp-s)\,!\,(mq+s)\,!}\,p^{mp-s} \times q^{mq+s}\,;$$

$$y_1 = \frac{m\,!}{(mp-s-1)\,!\,(mq+s+1)\,!}\,p^{mp-s-1} \times q^{mq+s+1}\,;$$

$$y_2 = \frac{m\,!}{(mp-s-2)\,!\,(mq+s+2)\,!}\,p^{mp-s-2} \times q^{mq+s+2}\,;$$

26. Les *Courbes de probabilité* simple (notation y au lieu de y_x)

$$y = \frac{m\,!}{(mp-x)!\,(mq+x)!}\,p^{mp-x}q^{mq+x}$$

sont dissymétriques par rapport à l'ordonnée dont l'abscisse est nulle, mais cette dissymétrie n'est jamais très forte [1] (on suppose mp et mq entiers, $s = 0$; rien à modifier si $s \neq 0$, formule 4′).

Exemple $m = 10\,000$; $p = 0{,}001$; $q = 0{,}999$ (fig. 1) :

$$y = \frac{10\,000\,!}{(10-x)!\,(9990+x)!}\,0{,}001^{10-x} \times 0{,}999^{9\,990+x}.$$

Valeurs de x	y	Valeurs de x	y
0	0,12 517	+ 1	0,12 516
— 1	0,11 379	2	0,11 262
— 2	0,09 482	3	0,09 007
— 3	0,07 292	4	0,06 302
— 4	0,05 207	5	0,03 780
— 5	0,03 470	6	0,01 889
— 6	0,02 168	7	0,00 755
— 7	0,01 274	8	0,00 226
— 8	0,00 707	9	0,00 045
— 9	0,00 372	10	0,00 005
— 10	0,00 185		

[1] Le cas de dissymétrie le plus élevé rencontré est représenté fig. 9. p. 134.

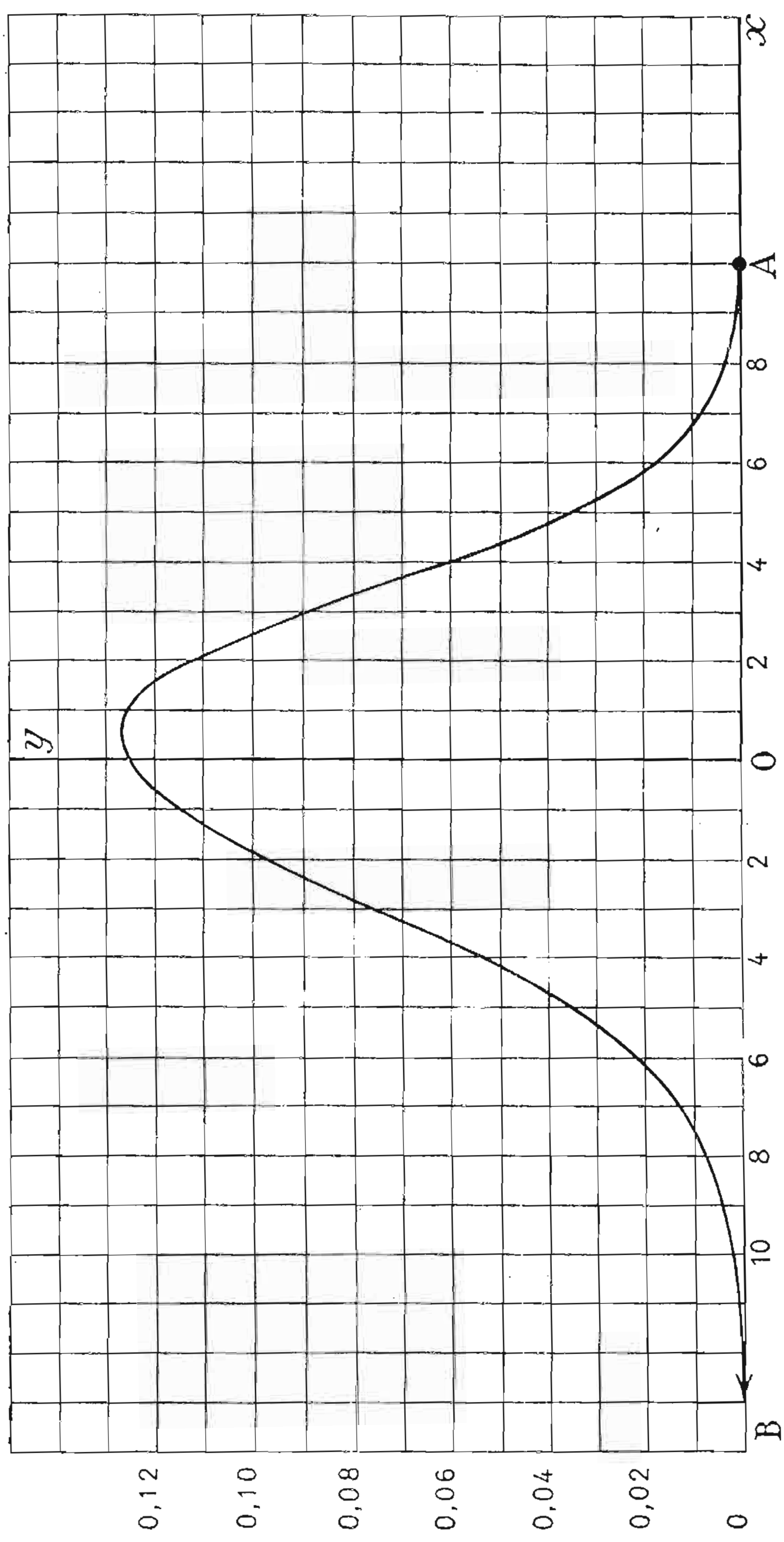


Fig. 1. — Exemple de courbe de probabilité simple $n = 10.000$; $p = 0,001$; $q = 0,999$.
Malgré la grande différence entre p et q, la dissymétrie est *faible*.
Une dissymétrie plus accentuée peut avoir lieu (fig. 9, p. 134).

La courbe a *deux points d'arrêt* A et B, qui ont pour abscisses mp et mq, ici :

$$+10 \quad \text{et} \quad -9990$$

Il en est de même, dissymétrie et points d'arrêt, pour les courbes (4') où mp, mq sont entiers ou fractionnaires, où s est fractionnaire :

$$y = \frac{m!}{(mp - s - x)!\,(mq + s + x)!} p^{mp-s-x} \times q^{mq+s+x}.$$

IV. — DÉVELOPPEMENT EN SÉRIE DE Log y_x VALEURS APPROCHÉES DE LA FONCTION y_x

27. Rappelons tout d'abord la formule de *Stirling*. On a, en *logarithmes décimaux* (particularité trop souvent passée sous silence)

$$\log n! = \log\sqrt{2\pi} + \left(n + \frac{1}{2}\right)\log n - Mn$$
$$+ \frac{M}{12\,n} - \frac{M}{360\,n^3} + \frac{M}{1\,260\,n^5} - \frac{M}{1\,680\,n^7} + \cdots$$

$$M = \log e = 0{,}434294\;481903\;25...$$

$$\log M = \bar{1}{,}637\;7843 \quad \text{ou} \quad \bar{1}{,}63\;778.$$

Dans les calculs à 7 décimales, les termes qui suivent $\frac{M}{360\,n^3}$ sont presque toujours négligeables.

Dans les calculs qui concernent les *satistiques*, on négligera même $\frac{M}{360\,n^3}$ et les termes suivants.

La série ne converge pas. Mais elle donne un grand nombre de décimales exactes, au moins 35, en prenant ses premiers termes en nombre convenable, pour n entier compris entre 0 et 3.000.

La série définit Log $n!$, dans la mesure où elle converge, pour n fractionnaire.

Appliquons cette formule à y_x (form. 4, n° 25).

$$(4) \qquad y_x = \frac{m!}{(mp - x)!\,(mq + x)!} p^{mp-x} q^{mq+x} :$$

pour l'appliquer à la formule (4'), on remplacerait dans ce qui suit mp par $mp - s$.

On trouve, pour la formule (4) :

$$\begin{aligned}
\log y_x &= \log m\,! - \log (mp - x)\,! - \log (mq + x)\,! \\
&\quad + (mp - x) \log p + (mq + x) \log q \\
&= \log\sqrt{2\pi} + \left(m + \frac{1}{2}\right) \log m - \mathrm{M}m + \frac{\mathrm{M}}{12\,m} - \frac{\mathrm{M}}{360\,m^3} \\
&\quad - \log\sqrt{2\pi} - \left(mp - x - \frac{1}{2}\right) \log (mp - x) + \mathrm{M}(mp - x) \\
&\quad - \frac{\mathrm{M}}{12(mp - x)} + \frac{\mathrm{M}}{360(mp - x)^3} \\
&\quad - \log\sqrt{2\pi} - \left(mq + x - \frac{1}{2}\right) \log (mq + x) + \mathrm{M}(mq + x) \\
&\quad - \frac{\mathrm{M}}{12(mq + x)} + \frac{\mathrm{M}}{360(mq + x)^3}
\end{aligned}$$

ou

$$(5) \quad \left\{ \begin{aligned}
\log y_x = &- \log\sqrt{2\pi} + \frac{1}{2} \log \frac{m}{(mp - x)(mq + x)} \\
&+ (mp - x) \log \frac{mp}{mp - x} + (mq + x) \log \frac{mq}{mq + x} \\
&+ \frac{\mathrm{M}}{12m} - \frac{\mathrm{M}}{12} \frac{m}{(mp - x)(mq + x)} \\
&- \frac{\mathrm{M}}{360m^3} + \frac{\mathrm{M}}{360} \left(\frac{1}{(mp - x)^3} + \frac{1}{(mq + x)^3} \right).
\end{aligned} \right.$$

avec

$$- \log \sqrt{2\pi} = \bar{1},600\,8101$$

Comme nous le verrons, la formule (5) est d'une *grande importance* au point de vue pratique.

Si
$$mp < mp - x,$$
on remplace $\quad (mp - x) \log \dfrac{mp}{mp - x} \quad$ par $\quad -(mp - x) \log \dfrac{mp - x}{mp}$;

si
$$mq < mq + x,$$
on remplace $\quad (mq + x) \log \dfrac{mq}{mq + x} \quad$ par $\quad -(mq + x) \log \dfrac{mq + x}{mq}$.

La formule (5) permet de développer *log y* en série. Il suffit d'appliquer la formule de Mac Laurin à chacun de ses termes. On a

$$
(6)\quad \left\{
\begin{aligned}
\log y_x = & -\log\sqrt{2\pi} - \frac{1}{2}\log mpq - \frac{\mathrm{M}}{12}\left(\frac{1}{mpq} - \frac{1}{m}\right) \\
& + \frac{\mathrm{M}}{360}\left(\frac{p^3+q^3}{(mpq)^3} - \frac{1}{m^3}\right) - \cdots \\
& + \mathrm{M}\left[\frac{1}{2}\frac{q-p}{mpq} - \frac{1}{12}\frac{q^2-p^2}{(mpq)^3} + \frac{1}{120}\frac{q^4-p^4}{(mpq)^4} - \cdots\right]x \\
& + \mathrm{M}\left[-\frac{q+p}{mpq} + \frac{1}{2}\frac{q^2+p^2}{(mpq)^2} - \frac{1}{6}\frac{q^3+p^3}{(mpq)^3} + \cdots\right]\frac{x^2}{2} \\
& + \mathrm{M}\left[-\frac{q^2-p^2}{(mpq)^2} + \frac{q^3-p^3}{(mpq)^3} - \frac{1}{2}\frac{q^4-p^4}{(mpq)^4} + \cdots\right]\frac{x^3}{6} \\
& + \mathrm{M}\left[-2\frac{q^3+p^3}{(mpq)^3} + 3\frac{q^4+p^4}{(mpq)^4} + \cdots\right]\frac{x^4}{24} + \cdots
\end{aligned}
\right.
$$

où tous les termes en p, q ; p^2, q^2 ; p^3, q^3, p^4, q^4 des coefficients de x, x^2, x^3, x^4 et du terme indépendant de x sont écrits [1].

28. Cas particuliers.

I. $$y_0 = \frac{1}{\sqrt{2\pi mpq}}.$$

II. $$y_x = \frac{1}{\sqrt{2\pi mpq}}\, e^{-\frac{x^2}{2mpq}}$$ LAPLACE.

III. $$y_x = \frac{1}{\sqrt{2\pi mpq}}\, e^{\frac{q-p}{2mpq}x - \frac{x^2}{2mpq} - \frac{q-p}{6(mpq)^2}x^3}.$$

IV. $$y_0 = \frac{1}{\sqrt{2\pi mpq}}\, e^{-\frac{1}{12}\left(\frac{1}{mpq} - \frac{1}{m}\right)}.$$

On voit avec quelle facilité les formules I à IV peuvent être établies [1]. On notera que *la formule* (6) *en dehors des cas* I, II, IV, *n'a aucun intérêt pour les calculs pratiques.*

[1] *Annales Soc. Scientifique de Bruxelles*, 1927.

V. — CALCULS NUMÉRIQUES

A) Emploi des Tables de factorielles et de la formule (5)

29. Soit à calculer, par exemple,

$$y_0 = \frac{100\,!}{35\,!\;65\,!} \times 0{,}35^{35} \times 0{,}65^{65},$$

formule (5) où $m = \mathbf{100}$; $p = \mathbf{0{,}35}$; $q = \mathbf{0{,}65}$.

Les Tables de factorielles (1) donnent, avec 8 décimales,

$$\begin{aligned} \log 100\,! &= 157{,}97000\,365 \\ \log\ \ 35\,! &= 40{,}01423\,265 \\ \log\ \ 65\,! &= 90{,}91633\,025. \end{aligned}$$

On a ensuite, avec 10 décimales,

$$\begin{aligned} &\log 35\,! = 40{,}01423\,26484 \\ &\log 34\,! = \underline{38{,}47016\,46040} \\ \text{par soustraction : } &\log 35 \ \ = \ \ 1{,}54406\,80444. \\ &\log 65\,! = 90{,}91633\,02540 \\ &\log 64\,! = \underline{89{,}10341\,68973} \\ \text{par soustraction : } &\log 65 \ \ = \ \ 1{,}81291\,33567. \end{aligned}$$

Connaissant log 35 et log 65 avec 10 décimales, on aura, avec 8 décimales,

$$\begin{aligned} \log 35^{35} &= 47{,}09407\,416 \\ \log 65^{65} &= 117{,}83936\,819 \end{aligned}$$

En remplaçant $0{,}35^{35} \times 0{,}65^{65}$ par

$$\frac{35^{35} \times 65^{65}}{100^{100}},$$

Nous avons les éléments nécessaires et suffisants pour calculer $\log y_0$ avec 7 décimales exactes. On trouve

$$\begin{aligned} \log y_0 &= \bar{2}{,}921\,1905, \\ y_0 &= 0{,}083\,4047 : \end{aligned}$$

(1) F. J. DUARTE : *Nouvelles Tables de Log n ! à 33 décimales depuis n = 1 jusqu'à n = 3000*. Imprimerie A. Kundig, Genève, 1927.

y_0 est la probabilité cherchée, *avec 7 décimales exactes.*

30. On trouverait les mêmes valeurs de log y_0 et de y_0 en employant la formule (5).

31. *Observations relatives à la formule* (5).

I. Les multiplications

$$(mp - x) \log \frac{mp}{mp - x}, \qquad (mq + x) \log \frac{mq}{mq + x}$$

peuvent entraîner des erreurs sur les deux dernières décimales de log y_x.

II. La formule (5) s'applique quand les nombres

$$m, \qquad mp - x, \qquad mq + x$$

sont fractionnaires : mais on ne pourrait alors calculer y_x par les Tables de factorielles.

B) Formules de récurrence

32. y_x et y_{x-1} peuvent se déduire l'un de l'autre par la formule

$$(mpq + px)y_x = (mpq - qx + q)y_{x-1} \tag{7}$$

qui est une conséquence immédiate des formules (4)

$$y_x = \frac{m!}{(mp - x)!\,(mq + x)!} p^{mp-x}q^{mp+x}$$

$$y_{x-1} = \frac{m!}{(mp - x + 1)!\,(mq + x - 1)!} p^{mp-x+1}q^{mq+x-1}$$

Au point de vue pratique cela est *fort important.* Faisons $x = 1, 2, 3,...$ dans la formule (7) ; il vient

$$y_1 = \frac{mpq}{mpq + p} y_0; \quad y_2 = \frac{mpq - q}{mpq + 2p} y_1; \quad y_3 = \frac{mpq - 2q}{mpq + 3p} y_2; \cdots$$

Faisons maintenant dans (7), $x = -1, -2, -3$; on a

$$y_{-1} = \frac{mpq}{mpq + q} y_0; \quad y_{-2} = \frac{mpq - p}{mpq + 2q} y_{-1}; \quad y_{-3} = \frac{mpq - 2p}{mpq + 3q} y_{-2}; \cdots$$

le calcul des diverses valeurs de y, à partir de y_1 est ainsi fait rapidement.

A ces formules, on doit substistuer les suivantes, quand il s'agit de calculs numériques :

$$(8)\quad \left\{ \begin{array}{l} y_1 = \dfrac{mp}{mq+1} \times \dfrac{q}{p} y_0 ; \qquad y_2 = \dfrac{mp-1}{mq+2} \times \dfrac{q}{p} y_1 ; \\ \qquad\qquad y_3 = \dfrac{mp-2}{mq+3} \times \dfrac{q}{p} y_2 ; \dots \end{array} \right.$$

$$(9)\quad \left\{ \begin{array}{l} y_{-1} = \dfrac{mq}{mp+1} \times \dfrac{p}{q} y_0 ; \qquad y_{-2} = \dfrac{mq-1}{mp+2} \times \dfrac{p}{q} y_{-1} ; \\ \qquad\qquad y_3 = \dfrac{mq-2}{mp+3} \times \dfrac{p}{q} y_{-2} ; \dots \end{array} \right.$$

VI. — ROLE DES FORMULES APPROCHÉES

33. — La formule II du nº **28** donne une assez bonne représentation approchée de la fonction binomiale ou *fonction de probabilité simple* (nº **25**), quand

$$p = q = 0{,}5.$$

Cette formule II est à la base de l'emploi de la fonction Θ (Chapitre III) dont le rôle est très important.

On a tenté de représenter la fonction de probabilité simple quand $p \neq q$ par la formule III du nº **28** ou par une formule dérivant de celle-ci.

On est arrivé à des résultats satisfaisants (¹). Mais des travaux plus récents, exposés aux Chapitres VI et suivants, ont rendu cette représentation peu utile.

(¹) *Annales Soc. Scient. de Bruxelles*, t. XLV et XLVI, 1926.

CHAPITRE III

LA FONCTION Θ

I. — LES TABLES DE LA FONCTION Θ

34. La fonction Θ est définie par

$$\Theta(u) = \frac{2}{\sqrt{\pi}}\int_0^u e^{-t^2}dt. \tag{1}$$

On a construit des Tables de cette fonction [1].

On notera que

$$\Theta(\infty) = 1.$$

Voici les problèmes usuels qui se posent à propos de la fonction Θ. Ce sont uniquement des problèmes d'interpolation. Comme les différences tabulaires varient très lentement quand on se borne à 4 décimales, ce qui nous suffira dans tous les cas que nous étudierons, l'interpolation se fera toujours par parties proportionnelles.

I. Avant d'étudier les deux problèmes qui peuvent se poser, remarquons que

$$\frac{2}{\sqrt{\pi}}\int_0^{-u} e^{-t^2}dt = -\frac{2}{\sqrt{\pi}}\int_0^u e^{-t^2}dt\,;$$

[1] Le lecteur trouvera une table de la fonction Θ dans : R. DE MONTESSUS DE BALLORE, *Leçons élémentaires sur le Calcul des probabilités*, Gauthier-Villars, à Paris.

donc

$$\Theta(-u) = -\Theta(u). \tag{2}$$

Les cas de u négatif et de $\Theta(u)$ négatif se ramènent donc de suite au cas de u et $\Theta(u)$ positifs.

II. *On connaît u, calculer* $\Theta(u)$. La solution du problème est immédiate si u se trouve dans les Tables; par exemple

$$\Theta(0{,}93) = 0{,}8116.$$

Envisageons donc le cas où u ne figure pas dans les Tables. Soit par exemple

$$u = 0{,}9372.$$

On a

$$\Theta(0{,}93) = 0{,}8116;$$

quand on ajoute 0,01 à 0,93, le nombre 0,8116 est augmenté de la différence tabulaire d

$$d = 0{,}0047 = \Theta(0{,}94) - \Theta(0{,}93);$$

on doit donc augmenter 0,8116 de

$$0{,}72 \times 0{,}0047 = 0{,}0034,$$

quand on ajoute 0,0072 à u.

Ainsi

$$\begin{array}{rl} \Theta(0{,}93) = & 0{,}8116, \qquad d = 0{,}0047, \qquad 0{,}72\,d = 0{,}0034 \\ & \underline{0{,}0034} \\ \Theta(0{,}9372) = & 0{,}8150. \end{array}$$

III. — *On connaît* $\Theta(u)$, *calculer* u.

Si $\Theta(u)$ est dans la Table, celle-ci donne immédiatement u.

Supposons que $\Theta(u)$ ne figure pas dans la Table. Soit par exemple, $\Theta(u) = 0{,}8150$; u est compris entre 0,93 et 0,94, puisque

$$\Theta(0{,}93) = 0{,}8116, \qquad \Theta(0{,}94) = 0{,}8163.$$

On lit dans la Table

$$\Theta(0{,}94) - \Theta(0{,}93) = d = 0{,}0047.$$

De plus,

$$\Theta(u) - \Theta(0{,}93) = 0{,}8150 - 0{,}8116 = 0{,}0034;$$

à 93 on doit donc ajouter

$$x = \frac{0{,}0034}{0{,}0047} = 0{,}72$$

et 93 devient 93,72 ; donc 0,93, devient 0,9372 ; autrement dit

$$u = 0{,}93 + \frac{1}{100} \times \frac{34}{47} = 0{,}93 + 0{,}0072 = 0{,}9372.$$

Nous allons utiliser la fonction Θ pour le calcul *approché* d'expressions numériques

$$\frac{m\,!}{(mp - x)\,!\,(mq + x)} p^{mp-x} q^{mq+x}.$$

II. — LA FONCTION DE PROBABILITÉ SIMPLE ET LA FONCTION Θ (1)

Cas de $p = q$

35. Quand $p = q = 0{,}5$, on a *approximativement* (n° 28, formule **II**), où $p = q = \frac{1}{2}$:

$$y = \frac{1}{\sqrt{\pi m : 2}} e^{-\frac{2x^2}{m}}.$$

Posons :

$$k^2 = \frac{1}{m : 2};$$

il vient

(3) $$y = \frac{k}{\sqrt{\pi}} e^{-k^2x^2},$$

formule fondamentale, *approchée*, donnant une valeur de la fonction de probabilité simple (n° 25, formule 4) quand $p = q = 0{,}5$.

(1) La fonction Θ dans le Calcul des Probabilités, *Ann. Soc. Scient. de Bruxelles*, 1929; Quelques points obscurs du Calcul des Probabilités; *Revue Gén. des Sciences*, 15 avril 1929.

Elle est asymptotiquement exacte quand m tend vers l'infini et quand x tend vers zéro.

Nous disons bien : *cette formule est approchée*. Nous ajoutons : *c'est à tort que des mathématiciens ont pris cette formule comme base du Calcul des probabilités, comme permettant de définir la probabilité* : il ne paraît pas possible de justifier l'hypothèse ainsi faite [1].

A. *Cas de m pair*. — Ici, $\frac{m}{2}$ est un nombre entier, la valeur maximum de y est y_0 (nº 23) et a pour expression

$$y_0 = \frac{m\,!}{\left(\frac{m}{2}\right)!\left(\frac{m}{2}\right)!}\,0{,}5^m\,;$$

de plus

$$y_{-x} = y_x,$$

car

$$y_{-x} = \frac{m\,!}{\left(\frac{m}{2}+x\right)!\left(\frac{m}{2}-x\right)!}\,0{,}5^m, \qquad y_x = \frac{m\,!}{\left(\frac{m}{2}-x\right)!\left(\frac{m}{2}+x\right)!}\,0{,}5^m.$$

Envisageons la somme

$$y_{-x} + y_{-x+1} + \cdots + y_{-1} + y_0 + y_1 + \cdots + y_x$$
$$= \frac{k}{\sqrt{\pi}}\left(e^{-k^2x^2} + e^{-k^2(x-1)^2} + \cdots + e^{-k^2} + 1 + e^{-k^2} + \cdots + e^{-k^2x^2}\right).$$

Nous remplacerons le second membre par l'*intégrale* dite de Laplace (Cf. nº 41 *in finem*) :

$$\frac{k}{\sqrt{\pi}}\int_{-f(x)}^{+f(x)} e^{-k^2x^2}dx = \frac{2k}{\sqrt{\pi}}\int_0^{f(x)} e^{-k^2x^2}dx,$$

en nous proposant de trouver *a posteriori* une expression convenable de la fonction inconnue $f(x)$. Il est naturel d'essayer pour $f(x)$ une fonction linéaire de la forme $bx + c$.

Faisons $x = 0$; on aura, puisque $f(x) = bx + c$,

$$y_0 = \frac{2k}{\sqrt{\pi}}\int_0^c e^{-k^2x^2}dx,$$

équation qui déterminera c. Introduisons la fonction Θ ; posons

$$kx = t\,;$$

[1] Cf. note de la page 60.

on a

$$y_0 = \frac{2}{\sqrt{\pi}} \int_0^c e^{-t^2} dt = \Theta(c) :$$

cette relation permet le calcul de c.

On a ensuite, pour $x = 1$,

$$\begin{aligned} y_0 + 2y_1 &= \frac{2k}{\sqrt{\pi}} \int_0^{b+c} e^{-k^2x^2} dx \\ &= \frac{2}{\sqrt{\pi}} \int_0^{b+c} e^{-t^2} dt \\ &= \Theta(b + c), \end{aligned}$$

relation qui détermine b.

Des considérations que nous ne rapporterons pas conduisent à prendre $b = k$.

On a par conséquent, *mais approximativement*,

$$y_{-x} + y_{-x+1} + \cdots + y_{-1} + y_0 + y_1 + \cdots + y_x = \frac{2k}{\sqrt{\pi}} \int_0^{kx+c} e^{-k^2x^2} dx$$

ou

$$\begin{aligned} & y_{-x} + y_{-x+1} + \cdots + y_{-1} + y_0 + y_1 + \cdots + y_x \\ & \quad = y_0 + 2y_1 + 2y_2 + \cdots + 2y_x = \frac{2}{\sqrt{\pi}} \int_0^{kx+c} e^{-t^2} dt = \Theta(kx + c) \end{aligned}$$

avec

$$y_0 = \Theta(c).$$

Soit $m = 20$. Le calcul direct pour $p = q = 0{,}5$, donne (nº 29)

$$y_0 = 0{,}1762.$$

Les Tables de la fonction Θ donnent à leur tour, en écrivant

$$\begin{aligned} 0{,}1762 &= \Theta(c) : \\ c &= 0{,}1575 ; \end{aligned}$$

or,

$$k = \sqrt{2 : m} = \sqrt{2 : 20} = \sqrt{1 : 10} = 0{,}3162,$$

$$\frac{k}{2} = 0{,}1581 :$$

on voit que, ici, $\frac{k}{2}$ diffère peu de c.

C'est un fait général, que $\frac{k}{2}$ diffère peu de c. Pour $m = 1.000$, le calcul direct ($p = q = 0{,}5$) donne (n° 29)

$$y_0 = 0{,}0252\ ;$$

si l'on écrit

$$0{,}0252 = \Theta(c),$$

les Tables de la fonction Θ donnent

$$c = 0{,}0223\ ;$$

or,

$$k = \sqrt{2 : m} = \sqrt{0{,}002} = 0{,}0447 = 2 \times 0{,}0223.$$

On est ainsi conduit à prendre dans tous les cas :

$$c = \frac{k}{2}.$$

Il en résulte

$$(4)\qquad \begin{cases} y_{-x} + y_{-x+1} + \cdots + y_{-1} + y_0 + y_1 + \cdots + y_x \\ \qquad = y_0 + 2y_1 + \cdots + 2y_x = \Theta\left(kx + \frac{k}{2}\right), \end{cases}$$

et, en particulier

$$y_0 = \Theta\left(\frac{k}{2}\right) = \Theta\left(\frac{1}{2}\sqrt{2 : m}\right).$$

On déduit ensuite de la formule (4)

$$(5)\qquad y_x = \frac{1}{2}\left[\Theta\left(kx + \frac{k}{2}\right) - \Theta\left(kx - \frac{k}{2}\right)\right].$$

Il est difficile d'évaluer analytiquement [1] l'approximation que donnent les formules (3) et (5), mais on peut vérifier que, d'une part :
la somme de toutes les valeurs possibles de y est 1 ; d'autre part, pour x infini (n° 34)

$$\Theta\left(kx + \frac{k}{2}\right) = 1\ ;$$

cette concordance était *nécessaire.* Elle va se retrouver dans les cas numériques que nous traiterons.

[1] Cf. D. Mirimanoff, *Le jeu de pile ou face et les formules de Laplace et de J. Eggenberger,* Commentarii Mathematici Helvetici, vol. 2, 1930, fasc. 2.

36. Examinons quelques cas concrets ($p = q = 0{,}5$).

I. $m = 20$.

On trouve

	Valeurs vraies	Valeurs calculées (form. 5)		Valeurs vraies	Valeurs calculées (form. 5)
y_0	0,1762	0,1769	y_0	0,1762	0,1769
$y_0 + 2y_1$	4966	4976	y_1	1602	1603
$y_0 + 2y_1 + 2y_2$	7368	7364	y_2	1201	1194
$y_0 + \cdots + 2y_3$	8846	8824	y_3	0739	0730
$y_0 + \cdots + 2y_4$	9586	9558	y_4	0370	0367
$y_0 + \cdots + 2y_5$	9882	9857	y_5	0148	0150
$y_0 + \cdots + 2y_6$	9974	9963	y_6	0046	0053
$y_0 + \cdots + 2y_7$	9996	9992	y_7	0011	0014
$y_0 + \cdots + 2y_8$	1,0000	9998	y_8	0002	0004

II. $m = 1.000$.

On trouve

	Valeurs vraies	Valeurs calculées (form. 5)		Valeurs vraies	Valeurs calculées (form. 5)
y_0	0,0252	0,0251	y_0	0,0252	0,0251
$y_0 + 2y_1$	0756	0755	y_1	0252	0252
$y_0 + 2y_1 + 2y_2$	1256	1255	y_2	0250	0250
....	....	....	...	...	...
$y_0 + \cdots + 2y_{10}$	4933	4932	...	...	...
$y_0 + \cdots + 2y_{11}$	5331	5328	y_{11}	0,0199	0,0198
....	....	....	...	....	....
$y_0 + \cdots + 2y_{49}$	0,99827	0,99825	...	....	....
$y_0 + \cdots + 2y_{50}$	99861	99860	y_{50}	0,00017	0,00017

On voit de quel ordre est l'approximation. *Elle est d'autant plus grande que m est plus grand : la formule est asymptotique.*

37. B. *Cas de m impair.* — Nous écrivons comme il suit la formule 4 (n° 25) correspondant à ce cas ($p = q = 0{,}5$) :

$$y_x = \frac{m\,!}{\left(\frac{m}{2} - x\right)!\left(\frac{m}{2} + x\right)!}\, 0{,}5^m ;$$

nous devons donner à x les valeurs $\pm$ **0,5** ; $\pm$ **1,5** ; $\pm$ **2,5**; etc..., (n° **25**, formule 4', cas où mp n'est pas entier).
Ainsi

$$y_{-0,5} = y_{0,5} = \frac{m\,!}{\left(\frac{m}{2} - 0{,}5\right)!\left(\frac{m}{2} + 0{,}5\right)!}\,0{,}5^m,$$

$$y_{-1,5} = y_{1,5} = \frac{m\,!}{\left(\frac{m}{2} - 1{,}5\right)!\left(\frac{m}{2} + 1{,}5\right)!}\,0{,}5^m.$$

Ici, nous avons à évaluer la somme

$$\begin{aligned} & y_{-x-0,5} + y_{-x-0,5+1} + \cdots + y_{-1,5} + y_{-0,5} + y_{0,5} + y_{1,5} + \cdots + y_{x+0,5} \\ &= 2y_{0,5} + 2y_{0,5} + \cdots + 2y_{x+0,5} \\ &= \frac{2k}{\sqrt{\pi}}\left[e^{-k^2 \times 0,5^2} + e^{-k^2 \times 1,5^2} + e^{-k^2 \times 2,5^2} + \cdots + e^{-k^2(x+0,5)^2}\right]. \end{aligned}$$

Nous remplacerons le second membre par

$$\frac{2k}{\sqrt{\pi}} \int_0^{f(x)} e^{-[k \times (x+0,5)]^2} dx,$$

où $f(x)$ est à déterminer.

Posons

$$k(x + 0{,}5) = t \qquad \text{d'où} \qquad kdx = dt\,;$$

il vient

$$2y_{0,5} + 2y_{1,5} + \cdots + 2y_{x+0,5} = \frac{2}{\sqrt{\pi}} \int_0^{f(x)} e^{-t^2} dt = \Theta f(x).$$

Prenons, comme dans le cas de m pair,

$$f(x) = kx + c$$

et faisons $x = o$; c est défini par la relation

$$2y_{0,5} = \Theta(c).$$

Par exemple, si $m = 19$, le calcul direct (n° 29 et suiv.) donne

$$y_{0,5} = 0{,}1762\,;$$

si l'on écrit

$$\Theta(c) = 2 \times 0{,}1762 = 0{,}3524,$$

on a, pour les Tables de la fonction Θ :

$$c = 0{,}3233.$$

Or,

$$k = \sqrt{2 : m} = \sqrt{2 : 19} = 0{,}3244,$$

valeur qui diffère peu de 0,3233.

Cela conduit à prendre

$$c = k = \sqrt{2 : m}$$

et à écrire, *approximativement* bien entendu,

$$2y_{0,5} + 2y_{1,5} + 2y_{2,5} + \cdots + 2y_{x+0,5} = \Theta[(x+1)k]$$

ou

$$y_{0,5} + y_{1,5} + y_{2,5} + \cdots + y_{x+0,5} = \frac{1}{2}\,\Theta[(x+1)k],$$

avec $$k = \sqrt{2 : m}.$$

On en déduit immédiatement

$$y_{x+5,5} = \frac{1}{2}\,\Theta[(x+1)k] - \frac{1}{2}\,\Theta(x+k).$$

Changeons dans cette formule $x + 0{,}5$ en x ; elle devient

$$(5\ bis) \qquad y_x = \frac{1}{2}\left[\Theta\left(kx + \frac{k}{2}\right) - \frac{1}{2}\,\Theta\left(kx - \frac{k}{2}\right)\right] :$$

C'est la formule (5) qu'on a trouvée pour m pair (n° 35).

Ainsi la formule 5 *bis* :

$$y_x = \frac{1}{2}\,\Theta\left(kx + \frac{k}{2}\right) - \frac{1}{2}\,\Theta\left(kx - \frac{k}{2}\right)$$

s'applique aux deux cas :

$$m \text{ pair}, \; m \text{ impair},$$

avec la différence que voici :

si m est *pair*, on donne à x les valeurs 0 ; $\pm$ 1, $\pm$ 2 ; $\pm$ 3 ;...

si m est *impair*, on donne à x les valeurs $\pm$ 0,5 ; $\pm$ 1,5 ; $\pm$ 2,5...

Application à un cas numérique quand m est impair. Soit $m = 19$; on trouve

	Valeurs vraies	Valeurs calculées (form. 5 *bis*)		Valeurs vraies	Valeurs calculées (form. 5 *bis*)
$2y_{0,5}$	0,3524	0,3535	$y_{0,5}$	0,1762	0,1767
$2y_{0,5} + 2y_{1,5}$	6408	6411	$y_{1,5}$	1442	1438
$2y_{0,5} + \cdots + 2y_{2,5}$	8330	8313	$y_{2,5}$	0951	0951
$2y_{0,5} + \cdots + 2y_{3,5}$	9364	9335	$y_{3,5}$	0517	0511
....	....	....	....	....	....

38. Les formules que nous venons de donner sont basées sur l'emploi de la fonction Θ, et nous permettront de ce chef de traiter le problème de l'interpolation (n^os^ 41 et suiv.). Elles donnent à peu près la même approximation que les formules suivantes, d'après M. Mirimanoff [1] :

A. *Formule de Laplace :*

$$y_0 + 2\sum_1^x y_x = \Theta(x) + \sqrt{\frac{2}{\pi m}}\, e^{-x^2}.$$

B. *Formule d'Eggenberger* [2] :

$$y_0 + 2\sum_1^x y_x = \Theta(x'), \qquad x' = x + \frac{1}{\sqrt{2m}}.$$

Ces formules A et B sont des cas particuliers d'une formule indiquée par M. Mirimanoff [1], qui a discuté et indiqué son approximation très grande :

$$y_0 + 2\sum_1^x y_x = \Theta(x) + \sqrt{\frac{2}{\pi m}}\, e^{-x^2} + \mathrm{E}_1,$$

(1) Se reporter à la note de la page 38.

(2) J. Eggenberger, *Beiträge zur Darstellung des Bernoulli'schen Theorems der Gammafunktion und des Laplace'schen Integrals*, Thèse, Berne, 1893 et Berner Mitteilung, t. 50, 1894, Zeitschrift für Math. und Physik, t. 45, 1900; cette formule avait déjà été donnée par De Morgan en 1838, d'après L. Bendersky (*Sur une formule de permutations*, Bulletin des Sciences mathématiques, 2^o^ série, t. LII, déc. 1928).

Le lecteur pourra consulter aussi un Mémoire de M. L. Bendersky : *Sur la sommation des termes du binome de probabilité*, Bull. des Sc. math., 2^e^ série, t. LIV, janv. 1930. Voir note p. 108.

avec

$$E_1 = \frac{e^{-x^2}}{3\sqrt{\pi}}\left(-\frac{7}{2}x + x^2\right)\frac{1}{m} + \frac{e^{-x^2}}{\sqrt{\pi}}\left(-\frac{1}{4}+x^2-\frac{x^4}{3}\right)\frac{1}{m}\sqrt{\frac{2}{m}}$$

$$-\frac{\lambda x + \varepsilon\sqrt{2:m}}{m^2}$$

$$|\lambda| < 0{,}75 \qquad |\varepsilon| < 0{,}1.$$

Cas de $p \neq q$.

39. Quand p diffère notablement de q, la fonction Θ ne donne plus une aussi bonne approximation. On peut cependant l'utiliser pour certains problèmes, comme nous le verrons.

Il s'agit de la fonction

$$y = \frac{m\,!}{(mp - x)\,!\,(mq + x)\,!}\,p^{mp-x}q^{mq+x}$$

où $p \neq q$. Soit par exemple m **= 80 ;** p **= 0,33 ;** q **= 0,67** (n° 22). **Ici**

$$y = \frac{80\,!}{(26{,}4 - x)\,!\,(53{,}6 + x)\,!}\,0{,}33^{26{,}4-x} \times 0{,}67^{53{,}6+x}.$$

Les nombres

$$26{,}4 - x\,;\ 53{,}6 + x$$

doivent être *entiers*, si l'on envisage le tirage de boules d'une urne (n° 25 : cas où mp n'est pas entier). Dans le cas présent, les nombres de boules rouges et noires de l'urne seraient proportionnels à **33** et **67**.

Si l'on veut mettre en évidence que **26,4** — x et **53,6** + x sont entiers, on peut écrire

$$y_x = \frac{80\,!}{(26{,}4 - x)\,!\,(53{,}6 + x)\,!} \times 0{,}33^{26{,}4-x} \times 0{,}67^{53{,}6+x}$$

et *spécifier* qu'on donnera à x les valeurs

$$\begin{array}{lll} 0{,}4\,; & 1{,}4\,; & 2{,}4\,; \cdots \\ -0{,}6\,; & -1{,}6\,; & -2{,}6\,; \cdots \end{array}$$

On a par calcul *direct* (n° 31 et 32)

$$y_{0,4} = \frac{80\,!}{26\,!\,54\,!}\,0{,}33^{26} \times 0{,}67^{54} = 0{,}0944$$

$$y_{1,4} = y_{0,4} \times \frac{26}{55} \times \frac{0{,}67}{0{,}33} = 0{,}0906 \qquad y_{-0,6} = y_{0,4} \times \frac{54}{27} \times \frac{0{,}33}{0{,}67} = 0{,}0930$$

$$y_{2,4} = y_{1,4} \times \frac{25}{56} \times \frac{0{,}67}{0{,}33} = 0{,}0821 \qquad y_{-1,6} = y_{-0,6} \times \frac{53}{28} \times \frac{0{,}33}{0{,}67} = 0{,}0851$$

$$\ldots\ldots\ldots\ldots \qquad y_{-2,6} = y_{-1,6} \times \frac{52}{29} \times \frac{0{,}33}{0{,}67} = 0{,}0766$$

$$\ldots\ldots\ldots\ldots$$

Introduisons la fonction Θ.

Nous sommes en présence d'une fonction y, que nous désignerons par (m, p, q), définie par la formule (4) du n° 25. Nous allons construire une autre fonction semblable $(m'; 0{,}5 ; 0{,}5)$ ayant même ordonnée y_0 que la fonction (m,p,q). La fonction $(m' ; 0{,}5 ; 0{,}5)$ évaluée par lafonction Θ, sera substituée à la fonction (m, p, q).

On a

$$y_x = \frac{m\,!}{(mp-x)!\,(mq+x)\,!}\,p^{mp-x}q^{mq+x}\,; \qquad y_x' = \frac{m'\,!}{\left(\frac{m'}{2}-x\right)!\,\left(\frac{m'}{2}+x\right)!}\,0{,}5^{m'};$$

puis (n° 28, formule 1)

$$y_0 = \frac{1}{\sqrt{2\pi mpq}}\,; \qquad y_0' = \frac{1}{\sqrt{2\pi m' \times \frac{1}{2} \times \frac{1}{2}}}.$$

L'introduction de y_0, y_0' suppose que mp, $\frac{m'}{2}$ sont entiers. Nous *conviendrons* de les introduire même si mp, $\frac{m'}{2}$ sont fractionnaires.

Nous écrivons en conséquence

$$\frac{1}{\sqrt{2\pi mpq}} = \frac{1}{\sqrt{2\pi m' \times \frac{1}{2} \times \frac{1}{2}}}$$

ce qui donne

$$m' = 4mpq.$$

Par conséquent (n° 35, A)

$$k' = \sqrt{2 : m'}$$

a ici la valeur

$$k' = \sqrt{2 : 4\, mpq}$$

ou

$$k' = \frac{1}{\sqrt{2mpq}}.$$

Considérons maintenant la valeur approchée (n° 28, II)

$$(\alpha) \qquad \frac{1}{\sqrt{2\pi mq}}\, e^{-\frac{x^2}{2mpq}}$$

de la fonction

$$\frac{m\,!}{(mp - x)\,!\,(mq + x)\,!}\, p^{mp-x} q^{mq+x}\,;$$

remplaçons p et q par **0,5** et m par m'; (α) devient

$$\frac{1}{\sqrt{\pi m' : 2}}\, e^{-\frac{m'}{x^2}}$$

et on a (n° 35, formule 5; n° 36 formule 5 *bis* où l'on fait

$$k = k' = \sqrt{2mpq}):$$

$$2y_x' = \Theta\left(k'x + \frac{k'}{2}\right) - \Theta\left(k'x - \frac{k'}{2}\right):$$

les y_x' seront des valeurs approchées des y_x.

Pour l'exemple choisi

$$m = 80\,; \qquad p = 0{,}33\,; \qquad q = 0{,}67,$$

on a

$$k' = \frac{1}{\sqrt{2 \times 80 \times 0{,}33 \times 0{,}67}} = 0{,}1681$$

et nous devons calculer $y'_{0,4}$; $y'_{1,4}$;... ; $y_{-0,3}$; $y_{-1,6}$;... qui seront les valeurs *approchées* des valeurs déjà calculées de

$$y_{0,4}\,; \qquad y_{1,4}\,; \cdots; \qquad y_{-1,6}\,; \qquad y_{-1,6}\,; \cdots$$

On a ici :

$$2y'_{0,4} = \Theta\left(0{,}4k' + \frac{k'}{2}\right) - \Theta\left(0{,}4k' - \frac{k'}{2}\right) = \Theta(0{,}9k') - \Theta(-0{,}1k')$$

$$= \Theta(0{,}9k') + \Theta(0{,}1k') = 0{,}1691 + 0{,}0193 = 0{,}1884$$

$$2y'_{-0,6} = \Theta\left(-0{,}6k' + \frac{k'}{2}\right) - \Theta\left(-0{,}6k' - \frac{k'}{2}\right)$$

$$= \Theta(-0{,}1k') - \Theta(-1{,}1k')$$

$$= \Theta(1{,}1k') - \Theta(0{,}1k') = 0{,}2063 - 0{,}0193 = 0{,}1870$$

$$2y'_{-1,6} = \Theta(2,1k') - \Theta(1,1k') = 0,3824 - 0,2063 = 0,1761$$

$$2y'_{1,4} = \Theta\left(1,4k' + \frac{k'}{2}\right) - \Theta\left(1,4k' - \frac{k'}{2}\right) = 0,3485 = 0,1691 = 0,194$$

. . . .

En fin de compte :

. . . .		. . . .	
$y_{-2,6} =$	$0,0766$	$y'_{-2,6} =$	$0,0782$
$y_{-1,6} =$	0862	$y'_{-1,6} =$	0880
$y_{-0,6} =$	0930	$y'_{-0,6} =$	0935
$y_{0,4} =$	0944	$y'_{0,4} =$	0942
$y_{1,4} =$	0906	$y'_{1,4} =$	0897
$y_{2,4} =$	0821	$y'_{2,4} =$	0804
. . . .		. . . .	

40. La représentation de la fonction (m, p, q) par la fonction $(m'\ ;\ 0,5\ ;\ 0,5)$ telle que nous venons de l'effectuer, appelle les remarques suivantes :

I. Elle est d'autant moins bonne que p diffère davantage de q ; elle est moins bonne par exemple pour $p = 0,1$; $q = 0,9$ que pour $p = 0,33$; $q = 0,67$.

II. Elle est d'autant moins bonne qu'on s'écarte davantage de y_0.

III. La formule (5) convient non seulement pour les valeurs de x rendant $mp - x$, $mq + x$ entiers, mais encore pour des valeurs *quelconques* de x, pourvu qu'elles diffèrent toutes de 1 unité, quand on passe de l'une à l'autre ; ces valeurs seront par exemple

$$\cdots -2,6\,; \quad -1,6\,; \quad -0,6\,; \quad 0,4\,; \quad 1,4\,; \quad 2,4\,; \cdots$$

comme tout-à-l'heure. Mais il ne peut plus être question de boules tirées d'une urne, si $mp - x$, $mq + x$ ne sont pas entiers.

IV. Au lieu d'identifier les ordonnée y_0, y_0' on aurait pu identifier deux ordonnées de même rang, par exemple :

$$y_{-1} \quad \text{et} \quad y'_{-1} \quad \text{ou} \quad y_{1,4} \quad et \quad y'_{1,4}\ :$$

nous le ferons à propos de l'interpolation (n° 42).

III. — INTERPOLATION

41. Considérons la fonction très simple

$$y = x^2 = \mathrm{F}(x + 1) - \mathrm{F}(x). \tag{6}$$

Il est évident que $\mathrm{F}(x)$ est de la forme

$$\mathrm{F}(x) = ax^3 + bx^2 + cx + d. \tag{7}$$

Si l'on porte dans (6), on trouve par identification

$$\mathrm{F}(x) = \frac{1}{3}\,x^3 - \frac{1}{2}\,x^2 + \frac{1}{6}\,x.$$

La dérivée de $\mathrm{F}(x)$ étant $x_2 - x + \frac{1}{6}$, on peut écrire

$$y = x^2 = \int_x^{x+1} \left(x^2 - x + \frac{1}{6}\right) dx. \tag{8}$$

Posons maintenant

$$\varphi(x) = 1^2 + 2^2 + 3^2 + \cdots + x^2, \qquad x \text{ entier positif.} \tag{9}$$

On a, d'après (8) :

$$\varphi(x) = \int_0^{x+1} \left(x^2 - x + \frac{1}{6}\right) dx \qquad \text{id.} \tag{10}$$

On peut *convenir* que (10) définit $\varphi\,(x)$ pour des valeurs de x non entières, mais positives, définition que (9) ne donne pas.

Par exemple pour $x = 3 + \frac{1}{4}$, on définira $\varphi\left(3 + \frac{1}{4}\right)$ par

$$\varphi\left(3 + \frac{1}{4}\right) = \int_0^{3+\frac{1}{4}} \left(x^2 - x + \frac{1}{6}\right) dx = \frac{1}{3}\left(\frac{17}{4}\right)^3 - \frac{1}{2}\left(\frac{17}{4}\right)^2 + \frac{1}{6}\left(\frac{17}{4}\right).$$

Si on calcule

$$\int_x^{x+\frac{1}{4}}, \qquad \int_{x+\frac{1}{4}}^{x+\frac{1}{2}}, \qquad \int_{x+\frac{1}{2}}^{x+\frac{3}{4}}, \qquad \int_{x+\frac{3}{4}}^{x+1}$$

on interpole (*cette manière d'interpoler est spéciale au calcul des probabilités*) la fonction (9) ; on y remplace x^2 par 4 termes

$$\int_x^{x+\frac{1}{4}} \left(x^2 - x + \frac{1}{6}\right) dx, \quad \int_{x+\frac{1}{4}}^{x+\frac{1}{2}} (-) dx, \quad \int_{x+\frac{1}{2}}^{x+\frac{3}{4}}, \quad \int_{x+\frac{3}{4}}^{x+1}$$

dont la somme, $\int_x^{x+1}$, est x^2.

C'est une interpolation *au quart*.

Les nombres $\quad \int_x^{x+\frac{1}{10}}, \quad \int_{x+\frac{1}{10}}^{x+\frac{2}{10}}, \quad \int_{x+\frac{2}{10}}^{x+\frac{3}{10}}, \ldots, \quad \int_{x+\frac{3}{10}}^{x+1}$

réalisent une interpolation au *dixième*, et les nombres

$$\int_x^{x+\frac{1}{100}}, \quad \int_{x+\frac{1}{100}}^{x+\frac{2}{100}}, \ldots, \quad \int_{x+\frac{99}{100}}^{x+1}$$

réalisent une interpolation au *centième*.

Il arrive qu'on veut interpoler une fonction F(x), comme nous avons interpolé φ (x), mais qu'on ne puisse pas mettre $\varphi(x)$ sous forme d'intégrale définie telle que (10). On écrit alors parfois

$$f(1) + f(2) + \cdots + f(x) = (\text{approximativement}) \int_0^x f(x)\, dx,$$

quitte à se rendre compte de l'approximation par des calculs numériques : c'est ce que nous avons fait quand nous avons écrit la formule de Laplace (n° **35**, A) qui peut être établie, mais non justifiée, par des considérations mathématiques, pour adapter ensuite cette formule à la représentation de la fonction de probabilité simple (n$^{\text{os}}$ 35 à 39).

De même que nous avons interpolé la fonction

$$1^2 + 2^2 + 3^2 + \cdots + x^2$$

par la formule (10), *nous pouvons actuellement interpoler la fonction de probabilité simple* par la fonction Θ.

42. Ecarts. — abcisses x ont reçu le nom d'*écarts*. Si $p = q$ et si m pair, la formule (4) du n° 35 nous permet de calculer

$$y_{-x} + y_{-x+1} + \cdots + y_{-1} + y_0 + y_1 + \cdots + y_x$$

pour des valeurs de x non entières.

Nous connaissons ainsi *la probabilité*

$$y_{-x} + y_{-x+1} + \cdots + y_{-1} + y_0 + y_1 + \cdots + y_x$$

que l'écart x sera compris entre — x et + x, en pouvant atteindre les limites — x, + x quand x est *fractionnaire.*

Cette probabilité a pour valeur approchée

$$\Theta\left(kx + \frac{k}{2}\right).$$

Par exemple la probabilité que l'écart sera compris entre — 0,3 (en pouvant atteindre — 0,3) et + 0,3 (en pouvant atteindre 0,3) sera

$$\Theta\left(0{,}3\,k + \frac{k}{2}\right).$$

Ces probabilités d'écarts fractionnaires n'ont *aucun sens* quand il s'agit de boules tirées d'une urne, puisqu'on envisage alors seulement les nombres y_x où x a les valeurs *entières* ± 1, ± 2, $\cdots$, et la valeur zéro. Nous verrons qu'elles ont au contraire un sens en statistique.

Si $p = q$ et m impair, nous interpolons par la formule (5 *bis*), identique à la formule (5).

Par exemple $m = 19$ avec $p = q = 0{,}5$. Quelle est la probabilité d'un écart positif, compris entre 0 et 2,7, pouvant atteindre 0 et 2,7 ? Cette probabilité est :

$$\frac{1}{2}\,\Theta\left(kx + \frac{k}{2}\right) \qquad \text{pour } x = 2{,}7$$

ou

$$\frac{1}{2}\,\Theta(3{,}2\,k) \qquad k = \frac{1}{\sqrt{m:2}} = 0{,}3244$$

ou

$$\frac{1}{2}\,\Theta(1{,}038) = 0{,}429.$$

On *notera* que y_0 a été partagé en deux : une moitié de ce nombre est ajoutée à $y_1 + y_2 + \cdots$; l'autre moitié est ajoutée à $y_{-1} + y_{-2} + y_{-3} + \cdots$

Cas de $p \neq q$. Nous calculons

$$\frac{1}{2} y_0 + y_1 + y_2 + \cdots + y_n$$

soit par la fonction Θ (nº 37 *in finem*) soit directement, selon le degré d'approximation que nous voulons obtenir.

Pour interpoler entre n et $n + 1$, nous procéderons comme il va être dit.

Nous allons traiter le cas de

$$m = 100, \qquad p = 0,1, \qquad q = 0,9$$

et nous proposons d'interpoler entre $x = 2$ et $x = 3$. ***Pour avoir*** *un critère numérique*, nous interpolerons entre $x = 2$ et $x = 4$.

Ici

$$y = \frac{100\,!}{(10 - x)\,!\,(90 + x)\,!}\, 0,1^{10-x} \times 0,9^{90+x}.$$

On a, en valeurs *vraies*,

$$\begin{array}{lll}
y_0 = 0,1319 & \frac{1}{2} y_0 & = 0,0659 \\
y_1 = 0,1304 & \frac{1}{2} y_0 + y_1 & = 0,1963 \\
y_2 = 0,1148 & \frac{1}{2} y_0 + y_1 + y_2 & = 0,3111 \\
y_3 = 0,0889 & \frac{1}{2} y_0 + y_1 + y_2 + y_3 & = 0,4000 \\
y_4 = 0,0596 & \frac{1}{2} y_0 + y_1 + y_2 + y_3 + y_4 & = 0,4596 \\
y_5 = 0,0339 & \frac{1}{2} y_0 + \cdots + y_5 & = 0,4953 \\
\dots\dots & \dots\dots\dots\dots & \dots
\end{array}$$

Ecrivons, comme si p était égal à q (nº 35, formule 4) :

$$\frac{1}{2} y_0 + y_1 + y_2 = 0,3111 = \frac{1}{2}\,\Theta\left(2k + \frac{k}{2}\right) = \frac{1}{2}\,\Theta\left(\frac{5k}{2}\right);$$

un calcul rapide donne

$$\frac{5k}{2} = 0,6237 \qquad k = 0,2495 \qquad \frac{k}{2} = 0,1247.$$

En partant de cette valeur de k, on trouvera

$$\text{valeur appr. de } \frac{1}{2} y_0 + y_1 + y_2 + y_3 + y_4 = \frac{1}{2} \Theta\left(4k + \frac{k}{2}\right)$$

$$= \frac{1}{2} \Theta\left(\frac{9k}{2}\right) = \frac{1}{2} \Theta(1{,}1227) = \frac{1}{2}\, 0{,}8512 = 0{,}4438$$

qui diffère de 0,0158 avec 0,4596, *valeur vraie*, donnée un peu plus haut.

Nous écrirons en conséquence, comme formule d'interpolation valable de $x = 2$ à $x = 4$.

$$(1) \qquad \frac{1}{2} y_0 + \sum_{1}^{x} y_x = \frac{1}{2} \Theta\left(kx + \frac{1}{2}\right) + (x - 2) \times 0{,}0079$$

avec

$$k = 0{,}2495: \qquad \frac{k}{2} = 0{,}1247\,;$$

le nombre 0,0079 a été choisi de mnière que l'égalité (1) restitue

$$0{,}3111 \qquad \text{et} \qquad 0{,}4596$$

pour $x = 2$ et $x = 4$.

Le critérium nous est donné en faisant $x = 3$; on trouve

$$\frac{1}{2} \Theta\left(3k + \frac{k}{2}\right) + (3 - 2) \times 0{,}0079 = 0{,}3999$$

au lieu de 0,4000, valeur vraie.

La formule (1) permet donc d'interpoler avec une erreur d'un petit nombre de dix-millièmes entre $x = 2$ et $x = 4$.

Par exemple,

$$\frac{1}{2} y_0 + \sum_{1}^{2,5} y_x = \frac{1}{2} \Theta\left(2{,}5k + \frac{k}{2}\right) + 0{,}5 \times 0{,}0079 = 0{,}3630.$$

Ce résultat donne lieu à l'énoncé suivant :

La probabilité d'un écart allant de 0 à 2,5 (0 et 2,5 compris) est

0,3630 (à quelques dix-millièmes près).

on a de même, toujours pour $m = 100$, $p = 0{,}1$, $q = 0{,}9$

$$\begin{array}{ll} y_0 = 0{,}1319 & \frac{1}{2}y_0 = 0{,}0659 \\ y_{-1} = 0{,}1199 & \frac{1}{2}y_0 + y_{-1} = 0{,}1858 \\ y_{-2} = 0{,}0988 & \frac{1}{2}y_0 + y_{-1} + y_{-2} = 0{,}2846 \\ y_{-3} = 0{,}0743 & \frac{1}{2}y_0 + y_{-1} + y_{-2} + y_{-3} = 0{,}3589 \\ \cdots\cdots & \cdots\cdots\cdots\cdots \end{array}$$

La formule d'interpolation est ici, de $x = -1$ à $x = -2$:

$$(2) \qquad \frac{1}{2}y_0 + \sum_{-1}^{-x} y_x = \frac{1}{2}\Theta\left(kx + \frac{k}{2}\right) - (x-1) \times 0{,}0058,$$

avec

$$k = 0{,}2281 \qquad \frac{k}{2} = 0{,}1140 ;$$

le nombre 0,0058 a été choisi de manière que la formule donne, 0,1858 pour $x = -1$ et 0,3589 pour $x = -3$.

Pour $x = -2$ (critérium), la formule donne

0,2842 au lieu de 0,2846 (valeur exacte).

La formule (2) donne donc, à quelques dix-millièmes près, la probabilité d'un écart donné compris entre -1 à -3, les extrémités -1 et -3 de l'intervalle pouvant être atteintes.

IV. — ÉCART PROBABLE

43. *L'écart probable est l'écart qui a chances égales d'être dépassé ou de ne pas être dépassé.*

Cette définition usuelle est à préciser.

Nous étudions, *ne l'oublions pas*, la fonction

$$y_x = \frac{m!}{(mp-x)!\,(mq+x)!}\,p^{mp-x}q^{mq+x}.$$

44. *Cas de* $p = q = 0{,}5$.

I. *m est pair.*

Soit un entier x défini par

$$\begin{cases} Y_x = y_{-x} + y_{-x+1} + \cdots + y_{-1} + y_0 + y_1 + \cdots + y_x < 0{,}5 \\ \qquad\qquad\qquad Y_{x+1} = Y_{-x} + y_{-x-1} + y_{x+1} > 0{,}5 \end{cases}$$

on dira que x est l'écart *probable* (ou mieux l'écart *le plus probable*) à moins de 1 unité près.

Il est clair que l'*interpolation*, telle que nous l'avons faite, permet de définir l'écart probable, à 0,1 près, à 0,01 près, etc.

Si l'on interpole par la fonction Θ, et si α est l'écart probable, la formule (4) du nº 35

$$y_{-x} + y_{-x+1} + \cdots + y_0 + y_1 + \cdots + y_x = \Theta\left(kx + \frac{k}{2}\right)$$

donnera

$$0{,}5 = \Theta\left(k\alpha + \frac{k}{2}\right).$$

Cette relation donne une valeur approchée de l'écart probable α : *approchée* puisque la fonction Θ est une représentation approchée de la somme des y; les Tables de la fonction Θ donnent

$$k\alpha + \frac{k}{2} = 0{,}477$$

d'où l'on tire

$$\alpha = \frac{0{,}477}{k} - 0{,}5.$$

Si l'on se reporte à la valeur de k (nº 35 au début) qui est $\frac{1}{\sqrt{m:2}}$, on a

$$(1) \qquad \text{écart probable} \quad \alpha = 0{,}477\sqrt{m:2} - 0{,}5.$$

Cette formule peut se déduire de la formule d'Eggenberger (nº 38) comme l'a remarqué M. Mirimanoff.

Observations. — A. C'est *à tort* que, dans la formule usuelle, 0,5 ne figure pas.

B. On écrit souvent, 0,477 avec plus de 3 décimales : cela est *inutile*, l'approximation obtenue n'est pas meilleure.

C. Il faut distinguer *l'écart probable à gauche* et *l'écart probable à droite*.

$$\alpha' = \text{éc. prob. à gauche} = -(0{,}477\sqrt{m : 2} - 0{,}50),$$
$$\alpha'' = \text{éc. prob. à droite} = 0{,}477\sqrt{m : 2} - 0{,}50;$$

où les valeurs de m diffèrent quand $p \neq q$.
l'écart probable *à gauche* est tel que

$$\frac{y_0}{2} + \sum_{-1}^{-\alpha'} y_x = 0{,}25$$

l'écart probable *à droite* est tel que

$$\frac{y_0}{2} + \sum_{+1}^{+\alpha''} y_x = 0{,}25$$

Pour $p = q = 0{,}5$,

$$\alpha' = \alpha''.$$

Exemple : $m = 970$ et $m = 968$ $(p = q = 0{,}5)$.
Pour

$$m = 970, \qquad k = \frac{1}{\sqrt{970 : 2}} = 0{,}04545$$

d'où (1)

$$\alpha = 10{,}005 = 10 + \varepsilon;$$

pour

$$m = 968, \qquad k = \frac{1}{\sqrt{968 : 2}} = 0{,}04543$$

d'où (1)

$$\alpha = 9{,}994 = 10 - \varepsilon'.$$

Dans ces deux cas, l'écart probable est donc très sensiblement **10**. En effet, le calcul *direct* (*exact*) des y donne

pour $m = 970$ $\quad y_0 + 2y_1 + 2y_2 + \cdots + 2y_{10} = 0{,}4998$ (val. exacte),
pour $m = 968$ $\quad y_0 + 2y_1 + 2y_2 + \cdots + 2y_{10} = 0{,}5002$ (val. exacte);

on a remplacé

$$y_{-10} + y_{-9} + \cdots + y_{-1} + y_0 + y_1 + \cdots + y_{10}$$

par la somme *identique*

$$y_0 + 2y_1 + 2y_2 + \cdots + 2y_{10}.$$

Les écarts probables à gauche et à droite sont, pour $m = 970$,

$$\alpha' = -(10 + \varepsilon), \qquad \alpha'' = 10 + \varepsilon,$$

et on a

$$\frac{y_0}{2} + y_1 + y_2 + \cdots + y_{10} = \frac{y_0}{2} + y_{-1} + y_{-2} + \cdots + y_{-10} = 0{,}25 + \eta.$$

II. m est *impair*. On écrira ici (comparer avec le nº **37**) :

$$\mathrm{Y}_{x+0,5} = y_{-x-1,5} + y_{-x-0,5-1} + \cdots + y_{0,5} + y_{+0,5} + \cdots + y_{x+0,5} < 0{,}5$$
$$\mathrm{Y}_{x+0,5+1} = \mathrm{Y}_{x+0,5} + y_{-x-0,5-1} + y_{x+0,5+1} > 0{,}5$$

x sera l'écart probable à 1 unité près.

Puisque

$$y_{-x-0,5} = y_{x+0,5},$$

on a

$$\mathrm{Y}_{x+0,5} = 2y_{0,5} + 2y_{1,5} + \cdots + 2y_{x+0,5};$$

cette somme a pour valeur *approchée* (nº **37**).

$$\Theta[(x + 1)k].$$

Soit $s_1 + 0{,}5$ l'écart probable à une unité près. On a

$$\mathrm{Y}_{s_{h+0,5}} = \Theta[(s_1 + 1)k] < 0{,}5,$$
$$\mathrm{Y}_{s_{h+0,5+1}} = \Theta[(s_1 + 2)k] > 0{,}5;$$

On a bien, l'écart probable à gauche étant $-(10 + \varepsilon)$

$$\frac{1}{2} y_0 + y_{-1} + \cdots + y_{-10} = 0{,}25 + \eta$$

et l'écart probable à droite étant encore $10 + \varepsilon$,

$$\frac{1}{2} y_0 + y_1 + \cdots y_{10} = 0{,}25 + \eta$$

Soit $s_n + 0{,}5$ l'écart probable à $\frac{1}{n}$ près. On a semblablement

$$\Theta[(s_n + 1)k] < 0{,}5 \qquad \Theta\left[\left(s_n + 1 + \frac{1}{n}\right)k\right] > 0{,}5;$$

enfin si

$$\text{écart probable } \alpha = s + 0{,}5,$$

c'est-à-dire si $\frac{1}{n}$ est négligeable, compte tenu de l'approximation que la fonction Θ peut donner, on a

$$\Theta[(s+1)k] = 0{,}5\,;$$

d'où il résulte

$$(s+1)k = 0{,}477$$

$$s+1 = \frac{0{,}477}{k} \qquad s = \frac{0{,}477}{k} - 1$$

puis

$$\alpha = s + 0{,}5 = \frac{0{,}477}{k} - 0{,}5,$$

$$\alpha = 0{,}477\sqrt{m} : 2 - 0{,}5 :$$

c'est la même formule que précédemment (m pair).

Soit par exemple $m = 431$. On a

$$\alpha = 0{,}477\sqrt{431} : 2 - 0{,}5 = 6{,}502 = 6{,}5 + \varepsilon$$

Par ailleurs, en prenant les valeurs exactes des y :

$$y_{-6,5} + y_{-5,5} + \cdots + y_{-0,5} + y_{0,5} + y_{1,5} + \cdots + y_{6,5} =$$
$$2y_{0,5} + 2y_{1,5} + \cdots + 2y_{6,5} = 0{,}5007 :$$

il y a concordance, *à quelques millièmes près* ; en effet, puisque

$$2y_{0,5} + 2y_{1,5} + \cdots + 2y_{6,5} > 0{,}5$$

la formule devrait donner

$$\alpha = 6{,}5 - \varepsilon$$

au lieu de $6{,}5 + \varepsilon$. L'emploi de la fonction Θ, qui ne donne que des valeurs *approchées*, comporte des erreurs de cet ordre.

Ici encore, nous avons l'écart probable à gauche

$$-(6{,}5 + \eta),$$

et l'écart probable à droite

$$6{,}5 - \eta :$$

1/4	des nomb. y	ont leurs	absc. compr. entre	-0 et $-(6{,}5 + \eta)$
1/4	»	»	»	0 et $6{,}5 + \eta$
1/4	»	»	plus petites que	$-(6{,}5 + \eta)$
1/4	»	»	plus grandes que	$6{,}5 + \eta$

45. III. *Cas de $p \neq q$.* — Quand $p \neq q$, y_x et y_{-x} *diffèrent l'un de l'autre*, sauf pour une seule valeur particulière de x, qui n'est pas à considérer ici.

L'écart probable qui se rapporte aux valeurs négatives de x (*écart probable à gauche*) diffère donc de l'écart probable qui se rapporte aux valeurs positives de x (*écart probable à droite*).

Traitons le cas de $m = 100$; $p = 0{,}1$; $q = 0{,}9$.

Nous avons trouvé (n° 42)

$$y_0 = 0{,}1319$$
$$y_0 + 2y_1 = 0{,}1858,$$
$$y_0 + 2y_1 + 2y_2 = 0{,}6223\,;$$

l'écart probable *à droite* est donc compris entre 1 et 2, puisque

$$y_0 + 2y_1 < 0{,}50 \qquad y_0 + 2y_1 + 3y_2 > 0{,}50.$$

Cela revient à

$$\frac{1}{2}\,y_0 + y_1 < 0{,}25, \qquad \frac{1}{2}\,y_0 + y_1 + y_2 > 0{,}25.$$

Nous avons *interpolé* par la formule

$$\frac{1}{2}\,y_0 + \sum_1^x y_x = \frac{1}{2}\,\Theta\left(kx + \frac{1}{2}\right) + (x - 2) \times 0{,}0079 \qquad k = 0{,}2495\,;$$

nous écrivons en conséquence

$$\frac{1}{2}\,\Theta\left(kx + \frac{1}{2}\right) + (x - 2) \times 0{,}0079 = 0{,}25 \qquad k = 0{,}2495\,;$$

la valeur de x qui vérifie cette équation :

$$x = 1{,}45$$

est l'écart probable à droite.

De même

$$y_0 = 0{,}1319,$$
$$y_0 + 2y_{-1} = 0{,}3717,$$
$$y_0 + 2y_{-1} + 2y_{-2} = 0{,}5693,$$

l'écart probable *à gauche* est donc compris entre -1 et -2, puisque

$$y_0 + 2y_{-1} < 0{,}50$$
$$y_0 + 2y_{-1} - 2y_{-2} > 0{,}50,$$

ce qui revient à

$$\frac{1}{2}y_0 + y_{-1} < 0{,}25 \qquad \frac{1}{2}y_0 + y_{-2} + y_{-2} > 0{,}25\,;$$

Nous avons interpolé par la formule

$$\frac{1}{2}y_0 + \sum_{-1}^{-x} y_x = \frac{1}{2}\Theta\left(kx + \frac{k}{2}\right) - (x-1) \times 0{,}0058 \qquad k = 0{,}2281\,;$$

nous écrirons en conséquence

$$\frac{1}{2}\Theta\left(kx + \frac{k}{2}\right) - (x-1) \times 0{,}0058 = 0{,}25\,;$$

la valeur de x qui vérifie cette équation,

$$x = 1{,}66$$

est l'*écart probable à gauche*.

Les équations qui interviennent se résolvent sans peine par tatonnement. Les écarts probables obtenus sont exacts, comme les interpolations, à un petit nombre de millièmes près.

Ici,

1/4 des val. de y ont leurs absc. comprises entre.	.	0 et		1,45
1/4 »	»	» . .	0 et	— 1,66
1/4 »	»	plus grandes que.	.	1,45
1/4 »	»	plus petites que.	.	— 1,65

46. — L'écart probable est fort long à calculer quand $p \neq q$. On peut en calculer rapidement une valeur approchée, que nous appellerons *écart probable moyen*.

Il s'agit d'une courbe de probabilité simple

$$(12) \qquad y = \frac{m\,!}{(mp-x)\,!\,(mq+x)\,!}\,p^{mp-x}\,q^{mq+x} \qquad p \neq q.$$

Nous avons montré (n° 39) que cette fonction admet la représentation *approchée*

$$(13) \qquad y = \frac{m'\,!}{\left(\frac{m'}{2}-x\right)!\left(\frac{m'}{2}+x\right)}\,0{,}5^{\,m'} \qquad p = q = 0{,}5$$

où l'on a $m' = 4mpq$.

On a posé aussi

$$k' = \sqrt{2 : m'} = \sqrt{2:4mpq} = \frac{1}{\sqrt{2mpq}}$$

et la fonction (13) a pour écarts probables :

$$\text{écart probl. à gauche} = -\left[0{,}477\sqrt{m' : 2} - 0{,}5\right],$$
$$\text{écart probl. à droite} = +\left[0{,}477\sqrt{m' : 2} - 0{,}5\right].$$

L'expression

$$(14) \qquad 0{,}477\sqrt{m' : 2} - 0{,}5$$

est ce que nous appellerons *l'écart probable moyen* de la fonction (12).

Remplaçons m' par sa valeur.

L'écart probable moyen de (12) a pour valeur

$$(15) \qquad 0{,}477\sqrt{2mpq} - 0{,}5.$$

Les écarts probables moyens à gauche et à droite de (12) seront

$$-(0{,}477\sqrt{2mpq} - 0{,}5), \qquad +(0{,}477\sqrt{2mpq} - 0{,}5).$$

Dans le cas de $m = 100$, $p = 0{,}1$, $q = 0{,}9$ (n° 45), ils sont

$$-1{,}52 \quad \text{et} \quad +1{,}52$$

les écarts probables *vrais* (ou du moins approchés à quelques millièmes) étant, comme on l'a vu,

$$-1{,}66 \quad \text{et} \quad +1{,}45.$$

Dans les cas usuels, l'écart probable à gauche vrai diffère de moins d'une demi-unité de l'écart probable à gauche moyen ; de même *l'écart probable à droite vrai diffère de moins d'une demi-unité de l'écart probable à droite vrai.*

La formule (14) peut être déduite d'une formule donnée par Eggenberger, que nous avons indiquée au n° 38.

47. Si $p \neq q$ la probabilité moyenne π d'un écart compris entre $-x$ et x est semblablement (se reporter au n° 42)

$$\pi = \Theta\left(kx + \frac{k}{2}\right)$$
$$k = \sqrt{2mpq}.$$

IV. — RÔLE DE LA FONCTION Θ

48. On admet généralement que l'intégrale (nº 35) dite de LAPLACE,

$$\frac{2k}{\sqrt{\pi}}\int_0^x e^{-k^2x^2}\,dx$$

ou que la fonction Θ, cela revient au même, représente EXACTEMENT la *loi du hasard*, par exemple la loi des erreurs accidentelles d'observation.

A ce sujet, on lira avec intérêt un Mémoire de M. Maurice FRÉCHET [1].

Nous verrons que la *fonction de probabilité simple* ou *fonction binomiale* (nº 25) *représente convenablement* un grand nombre de phénomènes naturels dont la cause fondamentale est assimilable au hasard (Chapitre XI). La fonction Θ, qui donne une bonne représentation approchée de la *fonction de probabilité simple*, interviendra pour les calculs pratiques d'écarts et d'interpolation de cette fonction.

[1] Sur l'hypothèse de l'additivité des erreurs partielles, *Bull. des Sciences Math.* 2e série, t. LII, mai 1928. M. Fréchet pense qu'il serait nécessaire de reprendre et de développer l'étude expérimentale des lois de probabilité et de confronter ces résultats avec les lois qu'on a proposées.

CHAPITRE IV

LA MODE

49. La *Mode* est l'abscisse de l'ordonnée maximum de la fonction

(1) $$y = \frac{m\,!}{(mp - x)\,!\,(mq + x)\,!}\,p^{mp-x}q^{mq+x},$$

et plus généralement de toute fonction représentative d'une statistique.

Il est *très important* de connaître la mode.

La Mode en première approximation.

La fonction (1) est approximativement représentée par la fonction (n° 28).

$$y = \frac{1}{\sqrt{2\pi mpq}}\,e^{\frac{q-p}{2mpq}x - \frac{x^2}{2mpq} - \frac{q-p}{6(mpq)^2}x^3};$$

le maximum, ou mode, est donc la racine voisine de zéro de l'équation

$$\frac{dy}{dx} = 0$$

ou

(2) $$\frac{q-p}{3mpq}\,x^2 + x - \frac{q-p}{2} = 0.$$

A l'équation (2) on peut pratiquement substituer celle-ci :

(3) $$x = \frac{q-p}{2}$$

car le coefficient de x^2, qui est $\frac{q-p}{3mpq}$, est ordinairement petit.

La formule (3) *donne la mode en première approximation* et cette approximation suffit presque toujours.

50. *La mode en deuxième approximation.*

Reportons-nous à la formule (5) du n° 27 :

$$\log y = -\log\sqrt{2\pi} + \frac{1}{2}\log\frac{m}{(mp-x)(mq+x)} + (mp-x)\log\frac{mp}{mp-x}$$
$$+ (mq+x)\log\frac{mq}{mq+x} + \frac{M}{12m} - \frac{M}{12}\frac{m}{(mp-x)(mq+x)}$$

où les deux derniers termes ont été négligés.

On a, *en utilisant cette formule*,

$$\frac{d\log y}{dx} = \frac{1}{2}\left(\frac{1}{mp-x} - \frac{1}{mq+x}\right)\left[1 - \frac{1}{6}\left(\frac{1}{mp-x} + \frac{1}{mq+x}\right)\right]$$
$$+ \frac{1}{M}\log\frac{q}{p} + \frac{1}{M}\log\frac{mp-x}{mq+x},$$

Il s'agit donc de calculer la racine voisine de zéro de l'équation

$$(4)\quad \frac{1}{2}\left(\frac{1}{mp-x} - \frac{1}{mq+x}\right)\left[1 - \frac{1}{6}\left(\frac{1}{mp-x} + \frac{1}{mq+x}\right)\right]$$
$$+ \frac{1}{M}\log\frac{q}{p} + \frac{1}{M}\log\frac{mp-x}{mq+x} = 0;$$

cela revient à calculer l'abscisse commune aux deux courbes.

$$(5)\quad \begin{cases} y_1 = \left(\dfrac{1}{mp-x} - \dfrac{1}{mq+x}\right)\left[1 - \dfrac{1}{6}\left(\dfrac{1}{mp-x} + \dfrac{1}{mq+x}\right)\right], \\ y_2 = -\dfrac{2}{M}\log\dfrac{q}{p} - \dfrac{2}{M}\log\dfrac{mp-x}{mq+x}. \end{cases}$$

Ce calcul est très facile : 1° parce que nous connaissons une valeur approchée $q-p$ de la racine cherchée (form. 3) ; 2° parce que les deux courbes (4) diffèrent peu de lignes droites aux environs de leur point commun.

On observera que l'ordonnée maximum de la fonction (1) quand on envisage SEULEMENT *les valeurs entières* de

$$mp - x, \qquad mq + x$$

correspond à la valeur *entière* de x vérifiant les inégalités

$$mp - q < x < mp + p.$$

Le calcul que nous avons fait se rapporte au cas où l'on fait varier x de façon continue.

EXEMPLES. — Soit $m = 10\,000$, $p = 0{,}001$, $q = 0{,}999$

On a

$$y_{-1} = 0{,}11379 \qquad y_0 = 0{,}12517 \qquad y_1 = 0{,}12516$$

A priori, le maximum est très voisin de $\frac{1}{2}$ puisque y_0 et y_1 diffèrent très peu l'un de l'autre.

La formule (3) donne

$$\frac{q-p}{2} = \frac{0{,}998}{2} = 0{,}499$$

Soit encore

$$m = 100 \qquad p = 0{,}1 \qquad q = 0{,}9.$$

La formule (3) donne pour valeur approchée de l'abscisse correspondant au maximum

$$\frac{q-p}{2} = 0{,}4$$

Appliquons les formules (5) à **0,35** et à **0,45**.

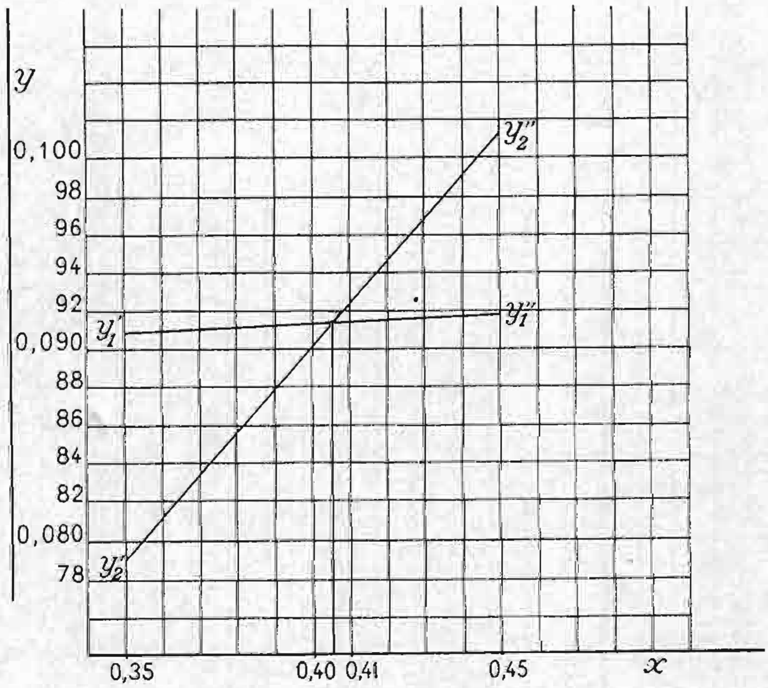


Fig. 2. — Résolution graphique d'une équation transcendante.

On a

$$mp = 10 \qquad mq = 90$$

$$x = 0{,}35 \qquad\qquad x = 0{,}45$$

$$mp - x = 9{,}65 \qquad mq + x = 90{,}35 \qquad mp - x = 9{,}55 \qquad mq - x = 90{,}45$$

$$\frac{1}{mp - x} = 0{,}10363 \qquad \frac{1}{mq + x} = 0{,}01107 \qquad \frac{1}{mp - x} = 0{,}10471 \qquad \frac{1}{mp + x} = 0{,}01109$$

$$y_1' = 0{,}09078 \qquad\qquad y_1'' = 0{,}09184$$

$$\log(mp - x) = 0,98453$$
$$\log(mq - x) = 1,95593$$
$$\log\frac{mp - x}{mq + x} = \bar{1},02860 = -0,97140$$
$$\log\frac{2}{M} = 0,66325$$
$$y_2' = 0,0790$$

$$\log(mp - x) = 0,98000$$
$$\log(mq + x) = 1,95641$$
$$\log\frac{mp - x}{mq + x} = \bar{1},02359 = -0,97641$$
$$\frac{2}{M}\log\frac{q}{p} = 4,3945$$
$$y_2'' = 0,1010$$

La figure 2, qui équivaut à une interpolation par parties proportionnelles, montre que la racine commune aux équations (5) est un peu plus grande que 0,4.

Appliquons les formules (5) à $x = 0,40$ et $x = 0,41$.

On trouve

pour $x = 0,40$	pour $x = 0,41$
$y_1' = 0,09132$	$y_1'' = 0,09143$
$y_2' = 0,09050$	$y_2'' = 0,09280$

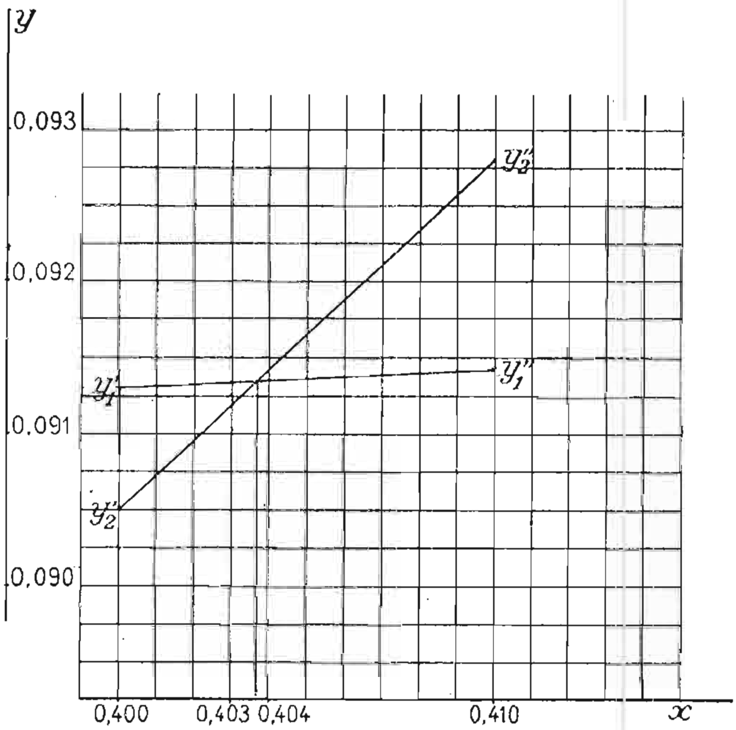


Fig. 3.. — Complément de la figure 2.

L'interpolation graphique, par parties proportionnelles (fig. 3) montre que x est très voisin de 0,403 à 0,404.

Cas de h différent de zéro

51. Pour la fonction

$$y = \frac{m!}{(mp - x)!\,(mq + x)!}\, p^{mp-x} q^{mq+x},$$

la mode est donnée par la formule approchée

$$x = \frac{q - p}{2}.$$

S'il s'agit de la fonction plus générale qui interviendra bientôt dans nos calculs

$$y_{x+h} = \frac{m!}{(mp - x - h)!\,(mq + x + h)!}\, p^{mp-x-h} q^{mq+x+h},$$

la mode sera donnée par

$$x + h = \frac{q - p}{2}.$$

Soit la statistique

....			
3	$y_{-2+h} = 10{,}1$		**où**
5	$y_{-1+h} = 13{,}2$	$h = -0{,}089$	$p = 0{,}289$ $\quad q = 0{,}711$
7	$y_h = 15{,}1$		
9	$y_{1+h} = 15{,}5$		$q - p = 0{,}422$
11	$y_{2+h} = 13{,}0$		
....			

La mode est donnée par

$$x + h = \frac{q - p}{2} = \left(\frac{q - p}{2} - h\right) + h = (0{,}211 - h) + h = 0{,}300 + h.$$

Ecrivons

(a)		
7	$7 + 0$	y_{0+h}
$7 + z$	$7 + z$	$y_{0,300+h}$
9	$7 + 9$	y_{1+h}
....		

on voit que

$$z = 2 \times 0{,}300 = 0{,}600$$

la mode, *dans l'échelle* (a) est donc **7,600.**

Soit encore la statistique

$$
\begin{array}{ll}
\ldots\ldots\ldots\ldots\ldots & \\
-\,0^\circ,5 & y_{-1+h} \\
-\,1^\circ,5 & y_h \\
-\,2^\circ,5 & y_{1+h} \\
\ldots\ldots\ldots\ldots &
\end{array}
$$

où

$$h = 0,5228 \qquad p = 0,2915 \qquad q = 0,7085$$

$$\frac{q - p}{2} = 0,2085$$

Ici

$$x + h = \frac{q - p}{2} = (0,2085 - h) + h = -\,0,3143 + h.$$

Ecrivons

$$
\begin{array}{ll}
-\,1^\circ,5 + 1 & y_{-1+h} \\
-\,1^\circ,5 + x & y_{-0,3143+h} \\
-\,1^\circ,5 + 0 & y_{-2+h}
\end{array}
$$

on voit que la mode est

$$-\,1^\circ,5 + 0,3143 = -\,1^\circ,1857.$$

CHAPITRE V

ROLE DISTRIBUTIF DES CONSTANTES m, p, q [1]

---

52. La *probabilité* de tirer $10 - x$ boules rouges et $90 + x$ boules noires quand on fait 100 tirages d'une urne contenant 1 boule rouge et 10 boules noires est (n° 25)

$$(1) \qquad y_x = \frac{100\,!}{(10 - x)\,!\,(90 + x)\,!}\, 0{,}1^{10-x} \times 0{,}9^{90+x}.$$

Le *nombre probable* de sorties de $10 - x$ boules rouges et de $90 + x$ boules noires quand on fait 1000 séries de 100 tirages chacune est

$$(2) \qquad Y_x = 1000 y_x$$

où les Y_x sont *arrondis à l'unité*, puisque ce sont évidemment des nombres entiers.

Cette notion de « **nombre probable** » est tout aussi importante que la notion de « **probabilité** ».

Dans le cas présent (1,2), les valeurs de y_x et Y_x sont les suivantes [2] :

[1] *Revue générale des Sciences* 31 mai et 15 juin 1928.
[2] Voir le Tableau en Note à la fin du volume.

		Nombres probables des sorties			
$y_{-13} = 0{,}000\,075$	$Y_{-13} =$	0 fois	23 rouges	et	77 noires
$y_{-12} = 0{,}000\,198$	$Y_{-12} =$	0 »	22 »		78 »
$y_{-11} = 0{,}000\,496$	$Y_{-11} =$	0 »	21 »		79 »
$y_{-10} = 0{,}001\,171$	$Y_{-10} =$	1 »	20 »		80 »
$y_{-9} = 0{,}002\,602$	$Y_{-9} =$	3 »	19 »		81 »
$y_{-8} = 0{,}005\,426$	$Y_{-8} =$	5 »	18 »		82 »
$y_{-7} = 0{,}010\,592$	$Y_{-7} =$	11 »	17 »		83 »
$y_{-6} = 0{,}019\,292$	$Y_{-6} =$	19 »	16 »		84 »
$y_{-5} = 0{,}032\,682$	$Y_{-5} =$	33 »	15 »		85 »
$y_{-4} = 0{,}051\,304$	$Y_{-4} =$	51 »	14 »		86 »
$y_{-3} = 0{,}074\,302$	$Y_{-3} =$	74 »	13 »		87 »
$y_{-2} = 0{,}098\,788$	$Y_{-2} =$	99 »	12 »		88 »
$y_{-1} = 0{,}119\,877$	$Y_{-1} =$	120 »	11 »		89 »
$y_{0} = 0{,}131\,865$	$Y_{0} =$	132 »	10 »		90 »
$y_{1} = 0{,}130\,416$	$Y_{1} =$	130 »	9 »		91 »
$y_{2} = 0{,}114\,823$	$Y_{2} =$	115 »	8 »		92 »
$y_{3} = 0{,}088\,895$	$Y_{3} =$	89 »	7 »		93 »
$y_{4} = 0{,}059\,579$	$Y_{4} =$	60 »	6 »		94 »
$y_{5} = 0{,}033\,866$	$Y_{5} =$	34 »	5 »		95 »
$y_{6} = 0{,}015\,911$	$Y_{6} =$	16 »	4 »		96 »
$y_{7} = 0{,}005\,892$	$Y_{7} =$	6 »	3 »		97 »
$y_{8} = 0{,}001\,623$	$Y_{8} =$	2 »	2 »		98 »
$y_{9} = 0{,}000\,295$	$Y_{9} =$	0 »	1 »		99 »
$y_{10} = 0{,}000\,027$	$Y_{10} =$	0 »	0 »		100 »
	Total :	1 000			

D'après ce tableau, la série composée de 12 rouges et 88 noires sortira probablement 99 fois ;

la série composée de 11 rouges et 89 noires sortira probablement 120 fois, etc.

On peut préciser.

Considérons la série composée de 12 rouges et de 88 noires.

Il y a 1000 séries, donc 1000 cas possibles.

La série envisagée doit sortir 99 fois : il y a 99 cas favorables à la sortie de 12 rouges et 88 noires. Il y a donc

$$1\,000 - 99 = 901$$

cas défavorables.

La probabilité π de sortie de 12 rouges et 88 noires est $\frac{99}{1\,000}$; la probabilité de non sortie est $1 - \frac{99}{1\,000} = \frac{901}{1\,000}$.

L'écart probable moyen concernant le nombre de sorties 99 de cette série (n° 46) est, à moins d'une demi-unité près :

$$\pm \left(0{,}477 \sqrt{2 \times 1\,000 \times \frac{99}{1\,000} \times \frac{901}{1\,000}} - 0{,}5\right) = \pm 6.$$

L'écart probable a chances égales d'être dépassé ou de ne pas être dépassé. Donc si l'on envisage par exemple 40 groupes de 1000 séries chacun, sur ces 40 groupes :

dans 20 groupes, 12 rouges et 88 noires sortiront plus de $99 - 6 = 93$ fois et moins de $99 + 6 = 105$ fois,

dans 10 groupes, 12 rouges et 88 noires sortiront 93 fois ou moins de 93 fois,

dans 10 groupes enfin, 12 rouges et 88 noires sortiront 105 fois ou plus de 105 fois.

Les nombres effectifs de sorties s'écarteront d'ailleurs peu, en moins, de 93 et peu, en plus, de 105; on pourrait préciser dans une certaine mesure ce mot : peu.

Importante remarque concernant les ajustements de statistiques

53. Il y a lieu de signaler aux statisticiens une particularité *intéressante.*

Considérons les nombres suivants où les Y sont arrondis à l'unité

$$y_x = \frac{50!}{(5-x)!\,(45+x)!}\, 0{,}1^{5-x} \times 0{,}9^{45+x} \qquad \text{et} \qquad Y_x = 500\, y_x.$$

On a

		500 y_x
$y_{-9} = 0{,}000\,21$	$Y_{-9} = 0$	
$y_{-8} = 0{,}000\,72$	$Y_{-8} = 0$	0,360
$y_{-7} = 0{,}002\,22$	$Y_{-7} = 1$	1,110
$y_{-6} = 0{,}006\,13$	$Y_{-6} = 3$	3,065
$y_{-5} = 0{,}015\,18$	$Y_{-5} = 8$	7,590
$y_{-4} = 0{,}033\,33$	$Y_{-4} = 17$	16,665
$y_{-3} = 0{,}064\,28$	$Y_{-3} = 32$	32,140
$y_{-2} = 0{,}107\,63$	$Y_{-2} = 54$	53,815
$y_{-1} = 0{,}154\,10$	$Y_{-1} = 77$	77,050
$y_0 = 0{,}184\,92$	$Y_0 = 92$	$500 = y_0$ 92,460
$y_1 = 0{,}180\,91$	$Y_1 = 90$	90,455
$y_2 = 0{,}138\,57$	$Y_2 = 69$	69,285
$y_3 = 0{,}077\,94$	$Y_3 = 39$	38,970
$y_4 = 0{,}028\,63$	$Y_4 = 14$	14,315
$y_5 = 0{,}005\,15$	$Y_5 = 3$	2,575
	Total : 499	

Le total n'étant pas 500 mais 499, pour obtenir 500, il faut remplacer

$$500\, y_x \qquad \text{par} \qquad (500 + \varepsilon) y_x.$$

où ε est *juste assez grand* pour augmenter l'un des Y, *et un seul*, de 1 unité.

En consultant la 3e colonne des produits *exacts* par 500, et $500\, y_0$, on voit que le multiplicateur doit être un peu plus grand que

$$500 \times \frac{92,5}{92,460}$$

Ce qui remplacera 92 par 93, mais n'augmente aucun des autres nombres Y.

La répartition de 500 séries se fait donc comme il suit :

$$1 \quad 3 \quad 8 \quad 17 \cdots \quad 77 \quad \mathbf{93} \quad 90 \cdots \quad 3.$$

Cas d'exception. Soit

$$y_x = \frac{60\,!}{(30 - x)\,!\,(30 + x)\,!}\, 0,5^{60} \qquad \text{et} \qquad Y_x = 300\, y_x,$$

où les nombres Y_x sont arrondis à l'unité ; ici

Valeurs de x	$300\,y$	Y_x	
0	30,7734	31	31 à compter 1 fois
± 1	29,7807	30	30, 27, ... à compter 2 fois
± 2	26,9268	27	
± 3	22,8996	23	
± 4	18,2151	18	
± 5	13,5087	14	
± 6	9,3813	9	
± 7	6,0852	6	
± 8	3,6831	4	
± 9	2,0775	2	
± 10	1,0908	1	
± 11	0,5322	1	
± 12	0,2409	0	
		Total : 301	

Il faudrait un multiplicateur qui transforme 13,5087 en un nombre un peu inférieur à 13,5 ; mais 14 (pour 13,5087) étant compté deux fois, si on le transforme en 13, le total sera $301 - 2 = 299$; *il n'est donc pas possible d'obtenir le total* 300.

Cette particularité et la manière de résoudre le problème posé se rencontrent dans TOUS LES AJUSTEMENTS DE STATISTIQUES. Si par un procédé de calcul quelconque on remplace des nombres entiers donnés a, b, c..., par des nombres a', b', c',... si $\Sigma a'$ diffère de Σa de quelques unités, on doit procéder semblablement pour amener $\Sigma a'$ à la valeur Σa.

Ecarts et fréquences

54. Il est temps d'introduire la notion de *fréquence* : *les nombres x sont des* ÉCARTS (n° 42), *les nombres* Y *sont des* FRÉQUENCES.

Dans le Tableau qui précède, il y a par exemple 31 fréquences dont l'écart est zéro ; 30 dont l'écart est 1 ; 30 aussi dont l'écart est -1 ;... ; 1 dont l'écart est 1 ; 1 dont l'écart est -1.

Nous n'envisageons ici que la loi de distribution (3), *infra* ; mais ces dénominations, **Ecarts** et **Fréquences**, s'appliquent à une loi de distribution *quelconque*.

Le rôle des constantes m, p, q, apparaît quand on distribue un *même* nombre N de fréquences par des lois

$$(3) \qquad Y = N \frac{m!}{(mp - x)!\,(mq + x)!} p^{mp-x} q^{mq+x}$$

où m, p, q, ont des valeurs différentes.

Exemple. On peut dresser le Tableau de la page 73.

Colonne A : 100 fréquences sont réparties en 23 groupes, définis par leurs écarts :

écarts -11, -10, -9,..., -1, 0, 1, ..., $+11$;

Colonne B : 100 fréquences sont réparties en 19 groupes, définies par leurs écarts : -9, -8, ..., $+9$;

Colonne C : 100 fréquences sont réparties en 16 groupes, définis par leurs écarts : -8, -7, ..., $+6$, $+7$.

On aperçoit ainsi pourquoi des figures correspondantes (fig. 4, 5, 6) présentent des aspects *plus ou moins étalés*, ou

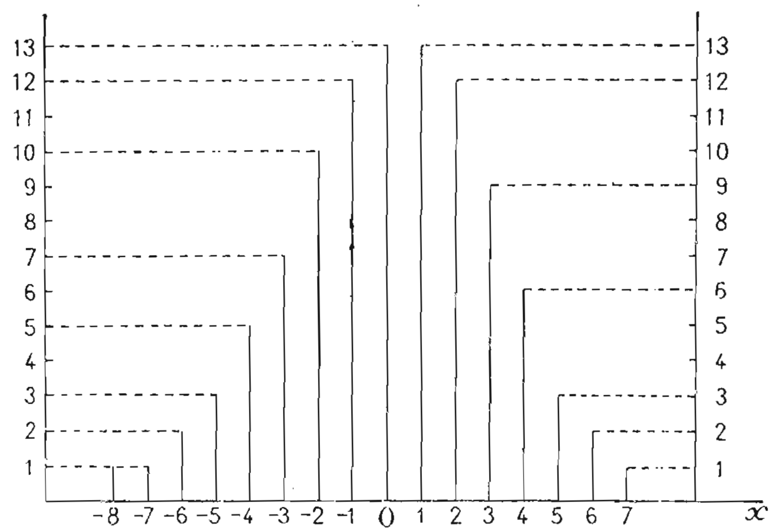


Fig. 4. — A. Répartition de 100 fréquences en 16 groupes, — 8 à + 7.

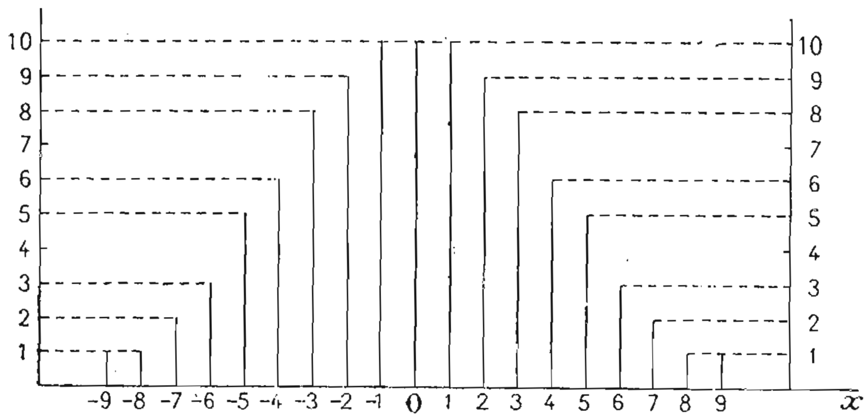


Fig. 5. — B. Répartition de 100 fréquences en 19 groupes – 9 à + 9.

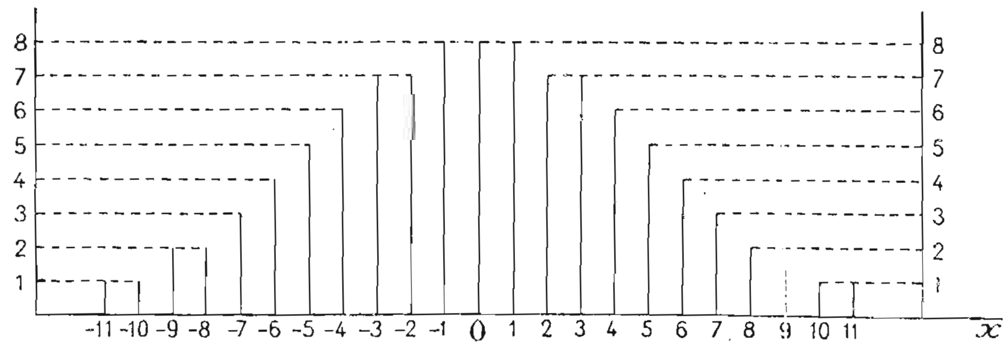


Fig. 6. — C. Répartition de 100 fréquences en 23 groupes, — 11 à + 11.

x	A $m = 100$ $p = q = 0,5$ $100\,y$	B $m = 60$ $p = q = 0,5$ $100\,y$	C $m = 100$ $p = 0,1 \quad q = 0,9$ $100\,y$
— 12	0		
— 11	1		
— 10	1	0	
— 9	2	1	0
— 8	2	1	1
— 7	3	2	1
— 6	4	3	2
— 5	5	5	3
— 4	6	6	5
— 3	7	8	7
— 2	7	9	10
— 1	8	10	12
0	8	10	13
1	8	10	13
2	7	9	12
3	7	8	9
4	6	6	6
5	5	5	3
6	4	3	2
7	3	2	1
8	2	1	0
9	2	1	100
10	1	0	
11	1	100	
12	0		
	100		

aplatis, surtout si l'on relie les extrémités des ordonnées par un trait *continu,* comme on le fait souvent.

Le rôle des constantes m, p, q est de modifier la distribution d'un nombre donné de fréquences.

CHAPITRE VI

INTRODUCTION DE LA CONSTANTE DE DEPLACEMENT [1]

55. Les formules (4) du n° 25 et (3) du n° 54

$$y_x = \frac{m!}{(mp - x)!\,(mq + x)!}\, p^{mp-x} q^{mq+x}$$

$$Y_x = N \frac{m!}{(mp - x)!\,(mq + x)!}\, p^{mp-x} q^{mq+x}$$

sont *insuffisantes* pour l'étude des problèmes qui se présenteront dans la suite. Nous devons les remplacer par les formules plus générales

$$(1) \qquad y_{x+h} = \frac{m!}{(mp - h - x)!\,(mq + h + x)!}\, p^{mp-h-x} q^{mq+h+x} \qquad x = 0, \pm 1, \pm 2, \cdots$$

$$(2) \qquad Y_{x+h} = N \frac{m!}{(mp - h - x)!\,(mq + h + x)!}\, p^{mp-h-x} q^{mq+h+x} \qquad x = 0, \pm 1, \pm 2, \cdots$$

où h *ordinairement fractionnaire*, mais pouvant être entier, est en général, *mais pas forcément*, compris entre $-$ 1 et $+$ 1.

Quand mp, mq sont *entiers*, quand h et x sont *entiers*, on a (n° 17)

$$(3) \qquad \Sigma y_{x+h} = 1,$$

[1] *Revue Gén. des Sciences*, 31 maî et 15 juin 1928.

où le signe Σ indique qu'on donne à x toutes les valeurs possibles, c'est-à-dire les valeurs ne rendant pas

$$mp - h - x, \qquad mq + h + x$$

négatifs.

Quand mp, mq ne sont pas entiers, la relation (3) n'est peut-être pas exactement vérifiée. On n'aurait donc pas

$$\Sigma N y_{x+h} = N \Sigma y_{x+h} = N\,;$$

mais la différence

$$\Sigma N y_{x+h} - N$$

apparaît comme si faible qu'on peut la négliger dans les applications.

Formules de récurrence

Les formules (8) et (9) du n° 32 deviennent ici

$$(4)\quad \left\{\begin{array}{l}
y_{-1+h} = \dfrac{mq+h}{mp-h+1} \times \dfrac{p}{q} y_h\,; \quad y_{-2+h} = \dfrac{mq+h-1}{mp-h+2} \times \dfrac{p}{q} y_{-1+h}\,; \\[2ex]
\qquad y_{-3+h} = \dfrac{mq+h-2}{mp-h+3} \times \dfrac{p}{q} y_{-2+h}\,; \;\cdots \\[2ex]
y_{1+h} = \dfrac{mp-h}{mq+h+1} \times \dfrac{q}{p} y_h\,; \quad y_{2+h} = \dfrac{mp-h-1}{mq+h+2} \times \dfrac{q}{p} y_{1+h}\,; \\[2ex]
\qquad y_{3+h} = \dfrac{mp-h-2}{mq+h+3} \times \dfrac{q}{p} y_{2+h}\,; \;\cdots
\end{array}\right.$$

On a par conséquent, formule (2)

$$(5)\quad \left\{\begin{array}{l}
Y_{-1+h} = \dfrac{mq+h}{mp-h+1} \times \dfrac{p}{q} Y_h\,; \quad Y_{-2+h} = \dfrac{mq+h-1}{mp-h+2} \times \dfrac{p}{q} Y_{-1+h}\,; \\[2ex]
\qquad Y_{-3+h} = \dfrac{mq+h-2}{mp-h+3} \times \dfrac{p}{q} Y_{-2+h}\,; \;\cdots \\[2ex]
Y_{1+h} = \dfrac{mp-h}{mq+h+1} \times \dfrac{q}{p} Y_h\,; \quad Y_{2+h} = \dfrac{mp-h-1}{mq+h+1} \times \dfrac{q}{p} Y_{1+h}\,; \\[2ex]
\qquad Y_{3+h} = \dfrac{mp-h-2}{mq+h-3} \times \dfrac{q}{p} Y_{2+h}\,; \;\cdots
\end{array}\right.$$

Quant à y_h et Y_h on pourra les calculer, *le premier* par la for-

mule (5) du n° 27 où l'on remplacera mp par $mp - h$ et mq par $mq + h$, et où l'on fera $x = 0$,
ce qui donne

$$(6)\quad \left\{\begin{aligned} \log y_h = &-\log\sqrt{2\pi} + \frac{1}{2}\log\frac{m}{(mp-h)(mq+h)} \\ &+ (mp-h)\log\frac{mq}{mp-h} + (mq+h)\log\frac{mq}{mq+h} \\ &+ \frac{M}{12m} - \frac{M}{12} \times \frac{m}{(mp-h)(mq+h)}, \end{aligned}\right.$$

c'est un cas particulier de la formule

$$(7)\quad \left\{\begin{aligned} \log y_{x+h} = &-\log\sqrt{2\pi} + \frac{1}{2}\log\frac{m}{(mp-h-x)(mq+h+x)} \\ &+ (mp-h-x)\log\frac{mp}{mp-h-x} + (mq+h+x)\log\frac{mq}{mq+h+x} \\ &+ \frac{M}{12m} - \frac{M}{12} \times \frac{m}{(mp-h-x)(mq+h+x)}, \end{aligned}\right.$$

obtenue en remplaçant x par $x + h$ dans la formule (5) du n° 27.
On aura ensuite

$$(8)\qquad Y_h = Ny_h.$$

56. Importante remarque. — *Il est* EXTRÊMEMENT PRÉCIEUX *de faire la remarque suivante,* qui dispense d'employer la formule (6).

Quand on a à calculer

$$Y_{-i+h},\quad Y_{-i+h+1},\ \cdots,\quad Y_{-1+h},\quad Y_h,\quad Y_{1+h},\ \cdots,\quad Y_{j+h}$$

où

$$\begin{aligned} &Y_{-i+h-1},\quad Y_{-i+h-2},\ \cdots \\ &Y_{j+h+1},\quad Y_{j+h+2},\ \cdots \end{aligned}$$

sont *négligeables,* on peut se borner à calculer des valeurs *provisoires*

$$\begin{aligned} &Y'_{-1+h},\quad Y'_{-2+h},\quad Y'_{-3+h},\ \cdots,\quad Y'_{i+h} \\ &Y'_{1+h},\quad Y'_{2+h},\ \cdots,\quad Y'_{j+h} \end{aligned}$$

des Y en faisant $Y = 1$ dans les formules (5). On fera ensuite la somme

$$\begin{aligned} \Sigma = &Y'_{-1+h} + Y'_{-2+h} + \cdots + Y'_{-i+h+1} \\ &+ Y'_{1+h} + Y'_{2+h} + \cdots + Y'_{j+h} \end{aligned}$$

et on obtiendra les Y en multipliant les Y′ par $\frac{N}{\Sigma}$, de telle sorte que

$$(9) \qquad Y_{i+h} = Y'_{i+h} \times \frac{N}{\Sigma}.$$

Comme on déduit de (9) :

$$\Sigma Y_{i+h} = \frac{N}{\Sigma} \Sigma Y'_{i+h} = \frac{N}{\Sigma} \Sigma = N,$$

$$\Sigma \frac{Y_{x+h}}{N} = 1, \qquad \Sigma y_{x+h} = 1$$

la formule (9) suppose donc que, comme au n° **18**,

$$\Sigma y_{x+h} = 1 :$$

or il n'en est peut-être pas ainsi. Mais certainement, dans les cas usuels,

$$1 - \Sigma y_{x+h}.$$

est assez petit pour pouvoir être négligé.

D'ailleurs, si on emploie la formule (6) ce qui donne y_h, puis

$$Y_h = N y_h,$$

si on emploie ensuite les formules de récurrence (5), la somme Σ' des Y obtenus ainsi différera *un peu* de N et on devra multiplier tous ces Y par $\frac{N}{\Sigma}$, ce qui revient au calcul à faire à partir de $Y_h' = 1$.

CHAPITRE VII

FORMULES DE SOMMATION [1]
MOMENTS

I. — CAS DE $h = 0$

57. Nous considérons les nombres y_x définis par la formule

$$y_x = \frac{m!}{(mp - x)!\,(mq + x)!}\, p^{mp-x} q^{mq+x}$$

où mp, mq peuvent être fractionnaires. Les valeurs de x sont

$$0, \quad \pm 1, \quad \pm 2, \quad \pm 3, \cdots$$

avec

$$mp - x > 0, \qquad mq + x > 0.$$

On a

$$y_{x-1} = \frac{m!}{(mp - x + 1)!\,(mq + x - 1)!}\, p^{mp-x+1} \times q^{mq+x-1};$$

donc

$$\frac{y_x}{y_{x-1}} = \frac{mp - x + 1}{mq + x} \times \frac{q}{p} = \frac{mpq - qx + q}{mpq + px}$$

et

$$(1) \qquad mpqy_x + xpy_x - mpqy_{x-1} + (x - 1)qy_{x-1} = 0.$$

[1] *Annales Soc. Scient. de Bruxelles*, 1927 et RAGNAR FRISCH, *Skandinavisk Aktuarietidskrift*, 1924, p. 161, *Biometrika*, 1925 p. 170; Cf. aussi D. MIRIMANOFF, sur une formule de M. de Montessus de Ballore, *Enseignement Mathématique*, 1929, p. 144.

Faisons $x = 1, 2, 3, \ldots, n$:

$$\begin{array}{llll}
mpqy_1 + py_1 & - mpqy_0 & & = 0 \\
mpqy_2 + 2py_2 & - mpqy_1 & + qy_1 & = 0 \\
mpqy_3 + 3py_3 & - mpqy_2 & + 2qy_2 & = 0 \\
\ldots\ldots\ldots & & & \\
mpqy_n + npy_n & - mpqy_{n-1} & + (n-1)qy_{n-1} & = 0.
\end{array}$$

Ajoutons ces identités en écrivant leur somme comme il suit :

$$(2)\qquad mpq \left|\begin{array}{l} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_{n-1} \\ y_n \end{array}\right. + p \left|\begin{array}{l} y_1 \\ 2y_2 \\ 3y_3 \\ \\ (n-1)y_{n-1} \\ ny_n \end{array}\right. - mq \left|\begin{array}{l} y_0 \\ y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_{n-1} \end{array}\right. + q \left|\begin{array}{l} \\ \\ y_1 \\ 2y_2 \\ \vdots \\ (n-1)y_{n-1} \end{array}\right. = 0.$$

Posons

$$(3)\qquad \left\{\begin{array}{l}
y_1 + y_2 + y_3 + \cdots + y_n = s'_{0,n} \\
y_1 + 2y_2 + 3y_3 + \cdots + ny_n = s'_{1,n} \\
y_1 + 2^2y_2 + 3^2y_3 + \cdots + n^2y_n = s'_{2,n} \\
y_1 + 2^3y_2 + 3^3y_3 + \cdots + n^3y_n = s'_{3,n} \\
y_1 + 2^4y_2 + 3^4y_3 + \cdots + n^4y_n = s'_{4,n} \\
\ldots\ldots\ldots\ldots\ldots\ldots \\
y_{-1} + y_{-2} + y_{-3} + \cdots + y_{-n'} = s''_{0,n'} \\
y_{-1} + 2y_{-2} + 3y_{-3} + \cdots + n'y_{-n'} = s''_{1,n'} \\
y_{-1} + 2^2y_{-2} + 3^2y_{-3} + \cdots + n'^2y_{-n'} = s''_{2,n'} \\
y_{-1} + 2^3y_{-2} + 3^3y_{-3} + \cdots + n'^3y_{-n'} = s''_{3,n'} \\
y_{-1} + 2^4y_{-2} + 3^4y_{-3} + \cdots + n'^4y_{-n'} = s''_{4,n'} \\
\ldots\ldots\ldots\ldots\ldots\ldots
\end{array}\right.$$

Les sommes s sont connues sous le nom de MOMENTS.

L'identité (2) s'écrit

$$mpq\, s'_{0,n} + ps'_{1,n} - mpq(s'_{0,n} + y_0 - y_n) + q(s'_{1,n} - ny_n) = 0;$$

après réductions :

$$(4)\qquad s'_{1,n} - mpqy_0 + (mpq - qn)y_n = 0.$$

De même

$$(4')\qquad s''_{1,n} - mpqy_0 + (mpq - pn')y_{n'} = 0.$$

Première série de formules de sommation

58. Multiplions maintenant la formule (1) par

$$x = (x-1)+1$$

en *associant* x à y_x et $x-1$ à y_{x-1} :

$$mpqxy_x + px^2y_x - mpq(x-1)y_{x-1} - mpqy_{x-1}$$
$$+ q(x-1)^2y_{x-1} + q(x-1) = 0.$$

Faisons

$$x = 1, 2, 3, \cdots, n$$

où n est un nombre entier quelconque, et ajoutons les équations obtenues :

$$\begin{array}{l|l|l|l|l|l}
 & & & & y_0 & \\
mqp & y_1 \quad +p & y_1 \quad -(mpq-q) & y_1 \quad -mpq & y_1 \quad +q & y_1 \quad =0; \\
 & 2y_2 & 2^2y_2 & 2y_2 & y_2 & 2^2y_2 \\
 & 3y_3 & 3^2y_3 & 3y_3 & y_3 & 3^2y_3 \\
 & \vdots & \vdots & \vdots & \vdots & \vdots \\
 & ny_n & n^2y_n & (n-1)y_{n-1} & y_{n-1} & (n-1)^2y_{n-1}
\end{array}$$

cela revient à

$$\begin{array}{l|l|l|l|l|l}
mpq & s'_{1,n} \quad +p & s'_{2,n} \quad -(mpq-q) & s'_{1,n} \quad -mpq & y_0 \quad +q & s'_{2,n} \quad =0; \\
 & & & -ny_n & +s'_{0,n} & -ny_n \\
 & & & & -y_n &
\end{array}$$

ou, puisque $p+q=1$,

$$s'_{2,n} + qs'_{1,n} - mpq(s'_{0,n} + y_0) + (n+1)(mpq - qn)y_n = 0.$$

De même

$$s''_{2,n'} + ps''_{1,n'}\, mpq(s''_{0,n'} + y_0) + (n'+1)(mpq - pn)y_{n'} = 0.$$

59. Multiplions maintenant la formule (1) par

$$x^2 = (x-1)^2 + 2(x-1) + 1,$$

toujours en associant x et y_x, $x-1$ et y_{x-1} ; on a

$$mpqx^2y_x + px^3y_x - mpq[(x-1)^2 + 2(x-1) + 1]y_{x+1}$$
$$+ q[(x-1)^3 + 2(x-1)^2 + (x-1)]y_{x-1} = 0,$$

ou

$$px^3y_x + q(x-1)^3y_{x-1} + mpqx^2y_x - (mpq-2q)(x-1)^2y_{x-1}$$
$$-(2mpq-q)(x-1)y_{x-1} - mpqy_{x-1} = 0,$$

Faisons

$$x = 1, 2, 3, \cdots, n$$

et ajoutons, en modifiant l'ordre des termes pour mieux mettre les réductions en évidence :

$$p\left|\begin{array}{l} y_1 \\ 2^3y_2 \\ \vdots \\ \vdots \\ n^3y_n \end{array}\right. + q\left|\begin{array}{l} y_1 \\ 2^3y_2 \\ \vdots \\ (n-1)^3y_{n-1} \\ \end{array}\right. + mpq\left|\begin{array}{l} y_1 \\ 2^2y_2 \\ \vdots \\ \vdots \\ n^2y_n \end{array}\right. -(mpq-2q)\left|\begin{array}{l} y_1 \\ 2^2y_2 \\ \vdots \\ (n-1)^2y_{n-1} \\ \end{array}\right. -(2mpq-q)\left|\begin{array}{l} y_1 \\ 2y_2 \\ \vdots \\ (n-1)y_{n-1} \\ \end{array}\right. -mpq\left|\begin{array}{l} y_0 \\ y_1 \\ y_2 \\ \vdots \\ (n-1)y_n \end{array}\right. = 0;$$

ceci revient à

$$p\left|\begin{array}{l} s'_{3,n} \\ \\ \end{array}\right. + q\left|\begin{array}{l} s'_{3,n} \\ -ny_n \\ \end{array}\right. + mpq\left|\begin{array}{l} s'_{2,n} \\ \\ \end{array}\right. -(mpq-2q)\left|\begin{array}{l} s'_{2,n} \\ -n^2y_n \\ \end{array}\right. -(2mpq-q)\left|\begin{array}{l} s'_{1,n} \\ -ny_n \\ \end{array}\right. -mpq\left|\begin{array}{l} y_0 \\ +s'_{0,n} \\ -y_n \end{array}\right. = 0,$$

ou

$$s'_{3,n} + 2qs'_{2,n} - (2mpq-q)s'_{1,n} - mpq(s'_{0,n} + y_0)$$
$$+ (n+1)^2(mpq - qn)y_n = 0;$$

de même

$$s''_{3,n'} + 2ps''_{2,n'} - (2mpq-p)s''_{1,n'} - mpq(s''_{0,n'} + y_0)$$
$$+ (n'+1)^2(mpq - pn')y_{n'} = 0.$$

Nous exprimons ainsi les sommes

$$s_1, \quad s_2, \quad s_3$$

en fonction des sommes de degrés moindres.

60. En multipliant (1) par

$$x^3 = (x-1)^3 + 3(x-1)^2 + 3(x-1) + 1$$
$$x^4 = (x-1)^4 + 4(x-1)^3 + 6(x-1)^2 + 4(x-1) + 1$$

et en associant toujours x à y_x et $x-1$ à y_{x-1}, on obtient semblablement les sommes s_4, s_5 qui sont écrites un peu plus loin.

On a ainsi les formules :

$$
(4)\quad \left\{
\begin{aligned}
& s'_{1,n} - mpqy_0 + (mpq - qn)y_n = 0 \\
& s'_{2,n} + qs'_{1,n} - mpq(s'_{0,n} + y_0) + (n+1)(mpq - qn)y_n = 0 \\
& s'_{3,n} + 2qs'_{2,n} - (2mpq - q)s'_{1,n} - mpq)s'_{0,n} + y_0) \\
& \qquad\qquad + (n+1)^2(mpq - qn)y_n = 0 \\
& s'_{4,n} + 3qs'_{3,n} - (3mpq - 3q)s'_{2,n} \\
& \qquad\qquad - (3mpq - q)s'_{1,n} - mpq(s'_{0,n} + y_0) \\
& \qquad\qquad + (n+1)^3(mpq - qn)y_n = 0 \\
& s'_{5,n} + 4qs'_{4,n} - (4mpq - 6q)s'_{3,n} - (6mpq - 4q)s'_{2,n} \\
& \qquad\qquad - (4mpq - q)s'_{1,n} - mpq(s'_{0,n} + y_0) \\
& \qquad\qquad + (n+1)^4(mpq - qn)y_n = 0. \\
& \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots
\end{aligned}
\right.
$$

On obtient les formules en s'' en remplaçant p par q :

$$
(5)\quad \left\{
\begin{aligned}
& s''_{1,n} - mpqy_0 + (mpq - pn)y_n = 0 \\
& s''_{2,n'} + ps''_{1,n'} - mpq(s''_{0,n'} + y_0) + (n'+1)(mpq - pn')y_{n'} = 0. \\
& \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots
\end{aligned}
\right.
$$

61. Les formules (4) et (5) sont EXACTES quel que soient les nombres entiers n.

Deux cas se présentent si l'on donne à n ses valeurs extrêmes.

1° Si mp est entier, on peut donner à n la valeur mp et

$$mpq - qn \quad \text{devient nul,}$$

les termes complémentaires en y_n *disparaissent*,

2° Si mp n'est pas entier, si l'on donne à n la valeur *entière*

$$mp + k$$

où k est la plus petite fraction possible, les termes complémentaires en y_n peuvent être négligés (de même pour les termes en n'.

On a ainsi les formules suivantes, EXACTES si mp entier, TRÈS-APPROCHÉES si mp n'est pas entier, *sous condition qu'on épuise les valeurs possibles de n,* ce que nous indiquons par la notation

$$
\begin{matrix}
s_1', & s_2', & s_3', & \cdots \\
s_1'', & s_2'', & s_3'', & \cdots :
\end{matrix}
$$

$$
(6)\quad \left\{ \begin{array}{l}
s_1' - mpqy_0 = 0 \\
s_2' + qs_1' - mpq(s_0' + y_0) = 0 \\
s_3' + 2qs_2' - (2mpq - q)s_1' - mpq(s_0' + y_0) = 0 \\
s_4' + 3qs_3' - (3mpq - 3q)s_2' - (3mpq - q)s_1' - mpq(s_0' + y_0) = 0 \\
s_5' + 4qs_4' - (4mpq\ 6q)s_3' - (6mpq - 4q)s_2' \\
\qquad\qquad - (4mpq - q)s_1' - mpq(s_0' + y_0 = 0 \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
s_1'' - mpqy_0 = 0 \\
s_2'' + ps_1'' - mpq(s_0'' + y_0) = 0 \\
s_3'' + 2ps_2'' - (2mpq - p)s_1'' - mpq(s_0'' + y_0) = 0 \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots
\end{array} \right.
$$

Deuxième série de formules de sommation

62. Reprenons la formule (1) du n° 57. Multiplions-la successivement par

$$
\begin{array}{l}
x - 1, \\
(x - 1)^2 = x^2 - 2x + 1, \\
(x - 1)^3 = x^3 - 3x^2 + 3x - 1, \\
(x - 1)^4 = x^4 - 4x^3 + 6x^2 - 6x + 1,
\end{array}
$$

en *associant*, comme précédemment x à y et $x - 1$ à y_{x-1}.

En procédant exactement comme tout à l'heure, on arrive aux formules suivantes, qui, au fond ne diffèrent pas des précédentes.

$$
(4^1)\quad \left\{ \begin{array}{l}
s'_{2,n} - ps'_{1,n} - mpqs'_{0,n} + n(mpq - qn)y_n = 0 \\
s'_{3,n} - 2ps'_{2,n} - (2mpq - p)s'_{1,n} + mpqs'_{0,n} + n^2(mpq - qn)y_n = 0 \\
s'_{4,n} - 3ps'_{3,n} - (3mpq - 3p)s'_{2,n} + (3mpq - p)s'_{1,n} \\
\qquad\qquad - mpqs'_{0,n} + n^3(mpq - qn)y_n = 0 \\
s'_{5,n} - 4ps'_{4,n} - (4mpq - 6p)s'_{3,n} + (6mpq - 4p)s'_{2,n} \\
\qquad\qquad - (4mpq - p)s'_{1,n} + mpqs'_{0,n} + n^4(mpq - qn)y_n = 0 \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots
\end{array} \right.
$$

Les formules analogues aux formules (5) et (6) sont

$$(5^1)\quad \begin{cases} s''_{2,n'} - qs''_{1,n'} - mpqs''_{0,n'} + n'(mpq - pn')y_{n'} = 0 \\ \dots\dots\dots\dots\dots\dots\dots\dots \end{cases}$$

$$(6^1)\quad \begin{cases} s_2' - ps_1' \quad - mpqs_0' = 0 \\ s_3' - 2ps_2' - (2mpq - p)s_1' + mpqs_0' = 0 \\ s_4' - 3ps_3' - (3mpq - 3p)s_2' + (3mpq - p)s_1' - mpqs_0' = 0 \\ s_5' - 4ps_4' - (4mpq - 6p)s_3' \\ \qquad + (6mpq - 4p)s_2' - (4mpq - p)s_1' + mpqs_0' = 0 \\ \dots\dots\dots\dots\dots\dots\dots\dots \\ s_2' - \quad qs_1'' - mpqs_0'' = 0 \\ s_3'' - 2qs_2'' - (2mpq - q)s_1'' + mpqs_0'' = 0 \\ \dots\dots\dots\dots\dots\dots\dots\dots \end{cases}$$

Les formules (4^1) et (5^1) sont EXACTES, quel que soit le nombre entier n positif, comme le sont les formules (4) et (5).

Quand aux formules (6^1) elles sont *exactes* si l'on peut donner à n la valeur mp (formules s') ou mq (formules s''). Si cela n'est pas possible, c'est-à-dire si mp, mq ne sont pas entiers, les formules (5^1), (6^1) sont seulement *très-approchées* quand on donne à n la valeur entière se rapprochant le plus possible de np pour les s', de nq pour les s''.

63. En écrivant (formules 6 et 6^1) :

$$(7)\quad \begin{cases} s_5' - mpq(4s_3' + 6s_2' + 4s_1' + s_0') \\ \qquad + q(4s_4' + 6s_3' + 4s_2' + s_1') - mpqy_0 = 0 \\ s_5' - mpq(4s_3' - 6s_2' + 4s_1' - s_0') \\ \qquad - p(4s_4' - 6s_3' + 4s_2' - s_1') = 0, \end{cases}$$

on aperçoit immédiatement les formes de s_n', qu'il est facile de vérifier.

La combinaison des formules (7) donne de nouvelles formules intéressantes.

II. CAS DE $h \neq 0$.

64. Considérant la fonction (n° 55)

$$y_{x+h} = \frac{m!}{(mp - h - x)!\,(mq + h + x)!}\, p^{mp-h-x} \times q^{mq+h+x},$$

nous avons

$$y_{x+h+1} = \frac{m\,!}{(mp-h-x-1)\,!\,(mq+h+x+1)\,!}\, p^{mp-h-x-1} \times q^{mq+h+x+1}$$

et

$$\frac{y_{x+h}}{y_{x+h+1}} = \frac{mq+h+x+1}{mp-h-x} \times \frac{p}{q} = \frac{mpq+p(x+h)+p}{mpq-q(x+h)}.$$

Posons

$$(8) \qquad mpq + ph = \lambda \qquad mpq - qh = \mu.$$

d'où

$$(9) \qquad h = \lambda - \mu\,;$$

la formule qui précède s'écrit

$$\frac{y_{x+h}}{y_{x+h+1}} = \frac{\lambda + px + p}{\mu - qx}.$$

On a ainsi

$$(10) \qquad \mu y_{x+h} - qxy_{x+h} - \lambda y_{x+h+1} - p(x+1)y_{x+h+1} = 0.$$

Posons (comparer avec les formules 3)

$$(11) \qquad \left\{ \begin{array}{lllll} y_{1+h} & + y_{2+h} & + \cdots + & y_{n+h} & = s'_{0,n} \\ y_{1+h} & + 2y_{2+h} & + \cdots + & ny_{n+h} & = s'_{1,n} \\ y_{1+h} & + 2^2y_{2+h} & + \cdots + & n^2y_{n+h} & = s'_{2,n} \\ & & & & \\ y_{-1+h} & + y_{-2+h} & + \cdots + & y_{-n'+h} & = s''_{0,n'} \\ y_{-1+h} & + 2y_{-2+h} & + \cdots + & n'y_{-n'+h} & = s''_{1,n'} \\ y_{-1+h} & + 2^2y_{-2+h} & + \cdots + & n'^2y_{-n'+h} & = s''_{2,n'}. \end{array} \right.$$

A. Dans (10), remplaçons x par 0, 1, 2, 3, ..., $n-1$ et faisons la somme des égalités obtenues, ce que nous écrivons

$$\mu \left| \begin{array}{l} y_h \\ y_{1+h} \\ y_{2+h} \\ \vdots \\ y_{n-1+h} \end{array} \right. - q \left| \begin{array}{l} \\ y_{1+h} \\ 2y_{2+h} \\ \vdots \\ (n-1)y_{n-1+h} \end{array} \right. - \lambda \left| \begin{array}{l} \\ y_{1+h} \\ y_{2+h} \\ \vdots \\ y_{n-1+h} \\ y_{n+h} \end{array} \right. - p \left| \begin{array}{l} \\ y_{1+h} \\ 2y_{2+h} \\ \\ (n-1)y_{n-1+h} \\ ny_{n+h} \end{array} \right. = 0;$$

on a (11)

$$\mu(y_h + s'_{0,n-1}) - qs'_{1,n-1} - \lambda s'_{0,n} - ps'_{1,n} = 0.$$

ou encore

$$(12) \qquad \mu(s'_{0,n} + y_h - y_{n+h}) - q(s'_{1,n} - ny_{n+h}) - \lambda s'_{0,n} - ps'_{1,n} = 0.$$

B. Multiplions (10) par x et écrivons comme il suit le résultat de la multiplication :

$$\begin{aligned} \mu x y_{x+h} - qx^2 y_{x+h} - \lambda(x+1)y_{x+h+1} + \lambda y_{x+h+1} - p(x+1)^2 y_{x+h+1} \\ + p(x+1)y_{x+h+1} = 0\,; \end{aligned}$$

faisons $x = 0, 1, 2, 3, \ldots, n-1$ et ajoutons les égalités obtenues; on a

$$\begin{array}{l|l l|l l|l l|l l|l l|l} \mu & y_{1+h} & -q & y_{1+h} & -\lambda & y_{1+h} & +\lambda & y_{1+h} & -p & y_{1+h} & +p & y_{1+h} = 0 \\ & 2y_{2+h} & & 2^2 y_{2+h} & & 2y_{2+h} & & y_{2+h} & & 2^2 y_{2+h} & & 2y_{2+h} \\ & 3y_{3+h} & & 3^3 y_{3+h} & & 3y_{3+h} & & y_{3+h} & & 3^2 y_{3-h} & & 3y_{3+h} \\ & \vdots & & \vdots & & \vdots & & \vdots & & \vdots & & \vdots \\ & (n-1)_{n-1+h} & & (n-1)^2 y_{n-1+h} & & & & & & & & \\ & & & & & ny_{n+h} & & y_{n+h} & & n^2 y_{n+h} & & ny_{n+h} \end{array}$$

ou bien

$$(13) \qquad \left\{ \begin{aligned} &\mu(s'_{1,n} - ny_{n+h}) - q(s'_{2,n} - n^2 y_{n+h}) - \lambda s'_{1,n} \\ &\qquad + \lambda s'_{0,n} - ps'_{2,n} + ps'_{1,n} = 0. \end{aligned} \right.$$

C. Dans (12) changeons n en $-n'$; il faudra changer p en q, remplacer λ par μ et μ par λ ; on a ainsi

$$(14) \qquad \lambda(s''_{0,n'} + y_h - y_{-n'+h}) - p(s''_{1,n'} - n'y_{-n'+h}) - \mu s''_{0,n'} - qs''_{1,n'} = 0.$$

D. Dans (13) changeons de même n en n', p en q, échangeons λ et μ ; il vient

$$(15) \qquad \left\{ \begin{aligned} &\lambda(s''_{0,n'} - n'y_{-n'+h}) - p(s''_{2,n'} - n'^2 y_{-n'+h}) - \mu s''_{1,n'} \\ &\qquad + \mu s''_{0,n'} - qs''_{2,n'} + qs''_{1,n'} = 0. \end{aligned} \right.$$

65. *Formules approchées.* — On peut négliger les termes en y_{n+h}, $y_{-n'+h}$, si l'on prend n et n' assez grand, ce que nous ferons toujours.

Posons donc (comparer avec les formules 11)

$$(16)\quad \begin{cases} s'_0 = y_{1+h} + y_{2+h} + y_{3+h} + \cdots \\ s''_0 = y_{-1+h} + y_{-2+h} + y_{-3+h} + \cdots \\ s'_1 = y_{1+h} + 2y_{2+h} + 3y_{3+h} + \cdots \\ s''_1 = y_{-1+h} + 2y_{-2+h} + 3y_{-3+h} + \cdots \\ s'_2 = y_{1+h} + 2^2y_{2+h} + 3^2y_{3+h} + \cdots \\ s''_2 = y_{-1+h} + 2^2y_{-2+h} + 3^2y_{-3+h} + \cdots \end{cases}$$

en excluant seulement les y_{n+h}, $y_{-n'+h}$ négligeables.

Les identités (12, 14) deviennent

$$(17)\quad \begin{cases} -\lambda s'_0 + \mu(s_0' + y_h) = s_1' \\ \lambda(s''_0 + y_h) - \mu s''_0 = s''_1 \end{cases}$$

et l'on en tire

$$\lambda = \frac{s'_1 s''_0 + s''_1 s'_0 + s''_1 y_h}{y_h(s_0' + s''_0 + y_h)} \qquad \mu = \frac{s''_1 s'_0 + s_1' s''_0 + s_1' y_h}{y_h(s'_0 + s''_0 + y_h)}.$$

On a d'autre part

$$(18)\quad s'_0 + s''_0 + y_h = 1$$

dans le cas où mp, mq sont entiers et quand $h = 0$ (nº 18).

Dans le cas général, l'égalité (18) n'est, vraisemblablement qu'*approchée*. Mais son degré d'approximation est tel que nous pouvons l'employer.

Elle permet d'écrire

$$(19)\quad \lambda = \frac{a}{y_h} + s''_1 \qquad \mu = \frac{a}{y_h} + s'_1$$

en posant

$$(20)\quad a = s'_1 s_0 + s''_1 s_0';$$

On en déduit (9)

$$(21)\quad \lambda - \mu = s''_1 - s_1',$$

$$(22)\quad h = s''_1 - s_1'.$$

Nous allons exprimer y_h en fonction des sommes (16). A cet effet, nous écrivons les équations (13, 15), en négligeant les termes en y_{n+h}, $y_{-n'+h}$:

$$(23)\quad \begin{cases} \mu s'_1 - \lambda(s_1' - s_0') = s_2' - p s_1' \\ \lambda s''_1 - \mu(s''_1 - s''_0) = s''_2 - q s''_1; \end{cases}$$

multiplions la première par s_1'', la deuxième par s_1' et ajoutons :

$$\lambda s_0' s_1'' + \mu s_0'' s_1' = s_2' s_1'' + s_2'' s_1' - s_1' s_1'';$$

posons

$$b = s_2' s_1'' + s_2'' s_1' - s_1' s_1''; \tag{24}$$

remplaçons λ et μ par leurs valeurs (19) ; on a

$$\left(\frac{a}{y_h} + s_1''\right) s_0' s_1'' + \left(\frac{a}{y_h} + s_1''\right) s_0'' s_1' = s_2' s_1'' + s_2'' s_1' - s_1' s_1''$$

d'où

$$\frac{a}{y_h} = \frac{b - (s_1')^2 s_0'' - (s_1'')^2 s_0'}{a}, \tag{25}$$

et

$$y_h = \frac{a^2}{b - (s_1')^2 s_0'' - (s_1'')^2 s_0'}. \tag{26}$$

La relation (25) permet d'écrire (19) et (20) comme il suit :

$$\lambda = \frac{b - (s_1')^2 s_0'' - (s_1'')^2 s_0'}{a} + s_1'' = \frac{b - (s_1')^2 s_0'' + s_0'' s_1' s_1''}{a}$$
$$= \frac{b}{a} + s_1' s_0'' \frac{s_1'' - s_1'}{a}$$

ou (22)

$$\left\{\begin{aligned} \lambda &= \frac{b}{a} + h \frac{s_1' s_0''}{a}; \qquad \text{de même} \\ \mu &= \frac{b}{a} - h \frac{s_1' s_0''}{a}. \end{aligned}\right. \tag{27}$$

Les équations (23) donnent ensuite

$$ps_1' = s_2' + (\lambda - \mu) s_1' - \lambda s_0'; \qquad qs_1'' = s_2'' - (\lambda - \mu) s_1'' - \mu s_0''; \tag{28}$$

remplaçons $\lambda - \mu$ par sa valeur (9), λ et μ par leurs valeurs (27), on trouve

$$\left\{\begin{aligned} p &= \frac{s_2'}{s_1'} - \frac{b}{a} \frac{s_0'}{s_1'} + \left(h - h \frac{s_0' s_0''}{a}\right) \\ q &= \frac{s_2''}{s_1''} - \frac{b}{a} \frac{s_0''}{s_1''} - \left(h - h \frac{s_0' s_0''}{a}\right). \end{aligned}\right. \tag{29}$$

Il est facile de voir que $p + q = 1$, quels que soient s_0', s_0'', s', s'', s_2', s_2''. *Les formules* (29) *comportent donc un élément de vérification des calculs effectués.*

66. Formules de sommation concernant les quantités Y_{x+h}. Nous considérons ici les quantités Y_{x+h} du n° 55.

Nous avons

$$Y_{x+h} = N \frac{m!}{(mp - x - h)!(mq + x + h)!} p^{mp-x-h} q^{mq+x+h}$$

$$(30) \qquad = N y_{x+h}$$

Nous posons

$$(31) \quad \left\{ \begin{aligned} S'_0 &= Y_{1+h} + Y_{2+h} + Y_{3+h} + \cdots \\ S''_0 &= Y_{-1+h} + Y_{-2+h} + Y_{-3+h} + \cdots \\ S'_1 &= Y_{1+h} + 2Y_{2+h} + 3Y_{3+h} + \cdots \\ S''_1 &= Y_{-1+h} + 2Y_{-2+h} + 3Y_{-3+h} + \cdots \\ S'_2 &= Y_{1+h} + 2^2 Y_{2+h} + 3^2 Y_{3+h} + \cdots \\ S''_2 &= Y_{-1+h} + 2^2 Y_{-2+h} + 3^2 Y_{-3+h} + \cdots \end{aligned} \right.$$

et de plus

$$(32) \qquad S = S'_0 + S''_0 + Y_h ;$$

nous nous arrêtons aux Y_{n+h}, $Y_{-n'+h}$ quand Y_{n+h+1}, $Y_{n'+h+1}$,... sont négligeables.

Nous avons (16)

$$(33) \quad S'_0 = Ns'_0,\ S''_0 = Ns''_0,\ S'_1 = Ns'_1, S''_1 = Ns''_1,\ S'_2 = Ns'_2,\ S''_2 = Ns''_2.$$

Par conséquent (32)

$$S = N(s'_0 + s''_0 + y_h)$$

ou (18)

$$(34) \qquad S = N$$

inégalité vraisemblablement approchée seulement (texte concernant la formule (18) ; ensuite (22, 34)

$$h = s''_1 - s'_1 = \frac{S''_1 - S'_1}{N},$$

$$(35) \qquad h = \frac{S''_1 - S'_1}{S}.$$

Remplaçons les s par les S dans a et dans b (20, 24) ; posons

$$(36) \qquad A = S'_1 S''_0 + S''_1 S'_0,$$

$$(37) \qquad B = S'_2 S''_1 + S''_2 S'_1 - S'_1 S''_1 :$$

on a

$$A = N^2 a, \ B = N^2 b \ ;$$

donc (**27**)

$$(38) \quad \begin{cases} \lambda = \dfrac{B}{A} + h \dfrac{S'_1 S''_0}{A} \\ \mu = \dfrac{B}{A} - h \dfrac{S''_1 S'_0}{A} \end{cases}$$

et (**29**)

$$(39) \quad \begin{cases} p = \dfrac{S'_2}{S'_1} - \dfrac{B}{A}\dfrac{S'_0}{S'_1} + \left(h - h \dfrac{S'_0 S''_0}{A}\right) \\ q = \dfrac{S''_2}{S''_1} - \dfrac{B}{A}\dfrac{S''_0}{S''_1} - \left(h - h \dfrac{S'_0 S''_0}{A}\right) \end{cases}$$

où h est défini par (**35**). On notera les formules suivantes de *vérification*, déduites de (23) :

$$(40) \qquad p = \frac{S_2'}{S_1'} + \lambda - \mu - \lambda \frac{S_0'}{S_1'}, \qquad q = \frac{S_2''}{S_1''} + \mu - \lambda - \mu \frac{S_0''}{S_1''}.$$

67. Connaissant les sommes S_0, S_1, S_2, S (**31, 32**), les formules (**39, 38**) permettent de calculer p, q, λ, μ ; les formules (**8**)

$$mpq = \lambda - ph = \mu + qh$$

permettent ensuite, λ, μ, p, q étant connus, de calculer m.

La formule (26) permet de calculer aussi Y_h; on a

$$Y_h = Sy_h = \frac{Sa^2}{b - (s_1')^2 s_0'' - (s_1'')s_0'} = \frac{S \dfrac{A^2}{S^4}}{\dfrac{B^2}{S^2} - \dfrac{(S_1')^2 S_0''}{S^3} - \dfrac{(S_1'')^2 S_0'}{S^3}}$$

$$(41) \qquad Y_h = \frac{A^2}{BS - (S_1')^2 S_0'' - (S_1'')^2 S_0'}$$

Si m = **100** ; p = **0,1** ; q = **0,9** ; h = **0,3**, on a

$$BS = 9{,}07849, \quad (S_1')^2 S_0'' = 0{,}48984, \quad (S_1'')^2 S_0' = 0{,}73505,$$

d'où (**41**)

$$BS - (S_1')^2 S_0'' - (S_1'')^2 S_0' = 7{,}85360$$

$$Y_h = 0{,}13304 \ ;$$

Le calcul direct à 4 décimales, donne 0,1330 (formule 5 du n° 27, où l'on remplace x par 0,3).

En remplaçant S par $S_0' + S_0'' + Y_h$, on tire de cette formule

$$(42) \qquad BY_h^2 + [B(S_0' + S_0'') - (S_1')^2S_0'')^2S_0']Y_h - A^2 = 0,$$

formule qui permet de calculer Y_h en partant des sommes S_0', S_0'' $\cdots$, S_2', S_2''.

Quand Y_h (observé) *a une valeur très douteuse, cette relation permet de le remplacer par un* Y_h *calculé à partir de* S_0', S_0'', $\cdots$, S_2', S_2'' *et d'aborder les calculs des* Y_{x+h} *avec l'*Y_h *calculé au lieu de l'*Y_h *observé.*

Cette remarque peut être *précieuse, car il n'est pas nécessaire de prendre pour* Y_h *la plus grande valeur des* Y *observés* ; il suffit pratiquement de prendre Y_h voisin de la plus grande valeur des Y.

On pourra donc prendre l'Y douteux, s'il existe, pour Y_h et le remplacer par la racine positive de (42), si cet Y douteux n'est pas trop loin de l'Y maximum.

68. Nous allons donner une idée de la puissance de cette méthode de calcul en traitant le cas suivant.

Reportons-nous aux nombres, 1, 3, 6, 15, ..., 10, 2 du n° , qui sont déduites de la loi de probabilité simple

$$y_{x+h} = \frac{m\,!}{(mp - x - h)\,!\,(mq + x + h)\,!}\, p^{mp-x-h} \times q^{mq+x+h}$$

où $m = 100$; $p = 0,1$; $q = 0,9$; $h = 0,3$

Calculons *directement* le nombre y_h par la formule (5) du n° 27 où l'on remplace m, p, q par leurs valeurs

100 ; 0,1 ; 0,9

et x par 0,3 ; calculons ensuite les y_{x+h} par les formules de récurrence (4) du n° 55. On trouve les nombres désignés dans le Tableau qui suit par le terme « Données ».

Partons de ces *données*. On trouve les nombres désignés par le terme « Calculé », en effectuant les calculs indiqués sur le Tableau et pages 94, 95 :

	Données					Calculé	
	$y_{-13+h} =$ 0,0001	× 13 =	0,0013	× 13 =	0,0169	0,0001 = y'_{-13+h}	
	0003	× 12 =	0036	× 12 =	0432	0003	
	0006	× 11 =	0066	× 11 =	0726	0006	
	0015	× 10 =	0150	× 10 =	1500	0015	
	0033	× 9 =	0297	× 9 =	2673	0032	
	0067	× 8 =	0536	× 8 =	4288	0067	
	0128	× 7 =	0896	× 7 =	6272	0127	
	0228	× 6 =	1368	× 6 =	8208	0227	
	0377	× 5 =	1885	× 5 =	9425	0377	
	0578	× 4 =	2312	× 4 =	9248	0578	
	0817	× 3 =	2451	× 3 =	7353	0816	
	1057	× 2 =	2114	× 2 =	4228	1057	
	$y_{-1+h} =$ 0,1246	× 1 =	0,1246	× 1 =	1246	1247 = y'_{-1+h}	
0,4556	0,4556 = S_0''		1,3370 = S_1''		5,5768 = S_2''	0,4553	0,4553 = Σ''
0,1329 = y_h						0,1330	0,1330 = y_h'
	$y_{1+h} =$ 0,1271	× 1 =	0,1271	× 1 =	0,1271	0,1272 = y'_{1+h}	
	1078	× 2 =	2156	× 2 =	4302	1079	
	0801	× 3 =	2403	× 3 =	7209	0801	
	0512	× 4 =	2048	× 4 =	8192	0512	
	0276	× 5 =	1380	× 5 =	6900	0276	
	0121	× 6 =	0726	× 6 =	4356	0121	
	0041	× 7 =	0287	× 7 =	2009	0042	
	0010	× 8 =	0080	× 8 =	0640	0010	
	$y_{9+h} =$ 0,0002	× 9 =	0,0018	× 9 =	0,0162	0002 = y'_{9+h}	
0,4112	0,4112 = S_0'		1,0369 = S_1'		3,5041	0,4115	0,4115 = Σ'
0,9997 = S							0,9998 = Σ

On a

$$
\begin{array}{rlrl}
S_0''S_1' &= 0{,}47241 & S_2'S_1'' &= 4{,}68498 \\
S_0'S_0'' &= 0{,}54977 & S_2''S_1' &= 5{,}78258 \\
\hline
A &= 1{,}02218 & S_1'S_1'' &= 1{,}38634 \\
 & & C &= 9{,}08122
\end{array}
\qquad h = \frac{0{,}3001}{0{,}9997} = 0{,}3001
$$

$$
\begin{array}{rlrlrlrl}
\dfrac{S_2'}{S_1'} &= 3{,}3794 & \dfrac{S_2''}{S_1''} &= 4{,}1711 & \dfrac{C}{A} &= 8{,}8842 & \dfrac{C}{A} &= 8{,}8842 \\
\dfrac{C}{A}\dfrac{S_0'}{S_1'} &= 3{,}5232 & \dfrac{C}{A}\dfrac{S_0''}{S_1''} &= 3{,}0274 & h\dfrac{S_0''S_1'}{A} &= 0{,}1387 & h\dfrac{S_0'S_1''}{A} &= 0{,}1615 \\
\hline
 & -0{,}1438 & & 1{,}1437 & & 9{,}0229 = \lambda & & 8{,}7227 = \mu
\end{array}
$$

$$-0{,}1438 + 1{,}1437 = 1{,}0001 \qquad \lambda - \mu = 0{,}3002$$

$$(h = 0{,}3001)$$

$$
\begin{array}{rlrl}
h - h\dfrac{S_0'S_0''}{A} &= 0{,}2451 & h - h\dfrac{S_0'S_0''}{A} &= 0{,}2451 \\
\hline
 & 0{,}1013 = p & & 0{,}8986 = q
\end{array}
\qquad p + q = 0{,}9999
$$

$$mpq = 8{,}9924$$

$$mp = 10{,}0071 \qquad mq = 88{,}7700 \qquad m = mp + mq = 98{,}7771$$

Calcul des y'.

Prenons

$$m = 98{,}78 \qquad mq = 88{,}77 \qquad mp = 10{,}01 \qquad h = 0{,}30$$

on a

$$mq + h = 89{,}07 \qquad mp - h = 9{,}71\,;$$

calculons y_h' pour la formule (5) du n° 27 où $h = x$; on a

$$\log \frac{m}{(mp - h)(mq + h)} = \bar{1}{,}05772,$$

$$^1/_2 \log \frac{m}{(mp - h)(mq + h)} = \bar{1}{,}52886\,;$$

avec 7 décimales, pour les multiplications,

$$\log \frac{mp}{mp - h} = 0{,}0132149$$

$$\log \frac{mq}{mq + h} = -\,0{,}0014653$$

$$(mp - h) \times \frac{mp}{mp - h} = 0{,}128316679$$

$$(mq + h) \times \frac{mq}{mq + h} = -\,0{,}130514271$$

$$(mp - h) \log \frac{mp}{mp - h} + (mp + h) \log \frac{mq}{mq + h} = -\,0{,}00220$$

(réduit à 5 décimales)

puis

$$\frac{\mathrm{M}}{12m} = 0{,}00037, \qquad \frac{\mathrm{M}}{12} \times \frac{m}{(mp - h)(mq + h)} = 0{,}00413,$$

$$-\log \sqrt{2\pi} = \bar{1}{,}60091\,;$$

finalement

$$\log y_h' = \bar{1}{,}12381\,; \qquad y_h' = 0{,}1330.$$

On calcule ensuite

$$y'_{-1+h}, y'_{-2+h}, \ldots, y'_{-13+h}$$
$$y'_{1+h}, \ y'_{2+h}, \ \ldots, y'_{9+h},$$

par les formules de récurrence (4) du numéro 55, qui donnent la colonne « calculé » du Tableau précédent.

La comparaison des y' et des y, des Σ'', Σ', Σ avec S_0'', S_0', S justifie complètement la méthode de calcul. ***Elle satisfait à tous les besoins de la statistique.***

Quant à p, q, mp, mq, m qui étaient primitivement 0,1 ; 0,9 ; 10 ; 90 ; 100, les nombres trouvés **0,1013 ; 0,8986 ; 10,0071; 88,7700—; 98,7771** les restituent avec une erreur relative d'environ un centième.

Mais ce ne sont pas ces nombres qui nous intéressent le plus ; *ce sont les nombres y' dont la coïncidence avec les y importait.*

Au Chapitre IX, nous verrons un exemple avec vérifications au cours des calculs.

Cas où les données sont incomplètes

69. *Il arrive que les termes en Y_{n+h}, $Y_{-n'+h}$ ne soient pas négligeables* dans les équations **13** à **15** du n° **64**.

Ce cas se présente si les valeurs extrêmes des Y ne sont pas connues ; par exemple si les *données* du Tableau du n° 68, étant limitées à **0,0067** dans un sens et à **0,0276** dans l'autre sens, les nombres

0,0033 à 0,0001 et 0,0121 à 0,0002

étaient inconnus.

Dans ce cas, on écrit les équations **23** à **26** complètes, avec les S et les Y :

$$(52)\quad \begin{cases} \mu(S_{0,n} + Y_h - Y_{n+h}) - \lambda S'_{0,n} + S'_{1,n} + qnY_{n+h} = 0, \\ \lambda(S''_{0,n'} + Y_h - Y_{-n'+h}) - \mu S''_{0,n'} + S''_{1,n'} + pn'Y_{-n'+h} = 0, \end{cases}$$

$$(53)\quad \begin{cases} \mu(S'_{1,n} - nY_{n+h}) - \lambda(S'_{1,n} - S'_{0,n}) + pS'_{1,n} + qn^2Y_{n+h} - S'_{2,n} = 0, \\ \lambda(S''_{1,n'} - n'Y_{-n'+h}) - \mu(S''_{1,n'} - S''_{0,n'}) \\ \qquad\qquad + qS''_{1,n'} + pn^2Y_{-n'+h} - S''_{2,n'} = 0. \end{cases}$$

On aurait ici

$$n = 5, \quad n' = 8.$$

Puis on résout les quatre équations par rapport à p, q, λ, μ, ce qui se fait aisément.

Trois cas peuvent se présenter.

1° On trouve exactement.

$$p + q = 1 :$$

on poursuit les calculs.

2° On a, à peu près seulement,

$$p + q = 1 ;$$

on modifie alors légèrement p, q en les remplaçant par des nombres très voisins p', q' tels que

$$p' + q' = 1$$

et on poursuit les calculs avec p', q' et les valeurs qu'on a trouvées pour λ, μ ; on a

$$h = \lambda - \mu, \qquad mp'q' = \lambda - p'h = \mu + q'h$$

relations qui donnent h, mp', mq', m.

3° La relation

$$p + q = 1$$

n'est pas vérifiée et, pour la vérifier, les corrections à faire subir à p, q sont inadmissibles ; ou bien λ ou μ sont *négatifs* ou bien p ou q sont *négatifs*.

Dans ce cas, les données Y n'appartiennent pas à une courbe de probabilité simple.

Note. — Un cas spécial se présente assez souvent : les Y manquant sont *petits* sans pouvoir être négligeables. On peut alors les obtenir graphiquement, ce qui permet de se servir des équations (27), (29) du n° 65.

On en trouvera un exemple au n° 92, statistique B.

CHAPITRE VIII

LES MOYENNES

I. — MOYENNE ARITHMÉTIQUE (1)

70. Considérons des nombres quelconques

$$a_1, a_2, \cdots, a_l ;$$

la *moyenne arithmétique* M_a de ces nombres est définie par

$$M_a = \frac{a_1 + a_2 + \cdots + a_l}{l}.$$

Il arrive que plusieurs de ces nombres a soient égaux ; si l'on envisage

$$\begin{array}{l} \alpha_1 \text{ nombres } b_1, \\ \alpha_2 \cdots\cdots\cdots b_2, \\ \quad\cdots\cdots \\ \alpha_i \cdots\cdots\cdots b_i, \end{array}$$

la moyenne arithmétique des b est

$$M_a = \frac{b_1 + b_1 + \cdots + b_2 + b_2 + \cdots + b_i + b_i + \cdots}{\alpha_1 + \alpha_2 + \cdots + \alpha_i}$$

ou

$$M_a = \frac{\alpha_1 b_1 + \alpha_2 b_2 + \cdots + \alpha_i b_i}{\alpha_1 + \alpha_2 + \cdots + \alpha_i}.$$

(1) *Revue Générale des Sciences*, 31 janvier 1930. — CH. JORDAN, *Statistique Mathématique*, Gauthier-Villars, Paris, 1927.

71. I. Soit la fonction de probabilité simple

$$y_x = \frac{m\,!}{(mp - x)\,!\,(mq + x)\,!}\,p^{mp-x}q^{qm+x}\,;$$

Si l'on a une urne où des boules rouges et noires sont mélangées dans la proportion $\frac{p}{q}$, si l'on fait m tirages, si mp, mq sont des nombres entiers :

*y_x est la probabilité de sortie de $mp - x$ boules rouges ; y_x est aussi **la** probabilité de l'écart x* ; ainsi

La probabilité de l'écart *zéro* est y_0 ;
. . . . $+ 1 \cdots y_1$;
. . . . $+ 2 \cdots y_2$;
. . . . .
. . . . $- 1 \cdots y_{-1}$;
. . . . $- 2 \cdots y_{-2}$;
. . . . $- 3 \cdots y_{-3}$;

la moyenne M_a de ces écarts sera par conséquent, en probabilité,

$$M_a = \frac{0 \times y_0 + 1 \times y_1 + 2 \times y_2 + \cdots - 1 \times y_{-1} - 2y_{-2} - \cdots}{y_0 + y_1 + y_2 + \cdots + y_{-1} + y_{-2} + \cdots}$$

Si l'on pose, en épuisant les valeurs possibles de x :

$$\begin{cases} s_1'' = y_{-1} + 2y_{-2} + \cdots + (mq)y_{-mq}, \\ s_1' = y_1 + 2y_2 + \cdots + (mp)y_{mp}, \end{cases}$$

on a

$$M_a = \frac{s_1' - s_1''}{y_0 + y_1 + y_2 + \cdots + y_{mp} + y_{-1} + y_{-2} + \cdots + y_{-mq}}\,;$$

et puisque mp, mq sont entiers (n° 61, 1^re^ et 6^e^ formule 6)

$$s_1' - s_1'' = 0$$

avec (n° 18)

$$y_0 + y_1 + \cdots + y_{mp} + y_{-1} + y_{-2} + \cdots + y_{-mq} = 1$$

donc

(1) $$M_a = 0.$$

Si mp, mq ne sont pas entiers, la formule

$$s_1'' = s_1' = mpq$$

est seulement approchée (premières formules 4 et 5 du n° 60) ;

par conséquent la formule (1) est elle-même *approchée*, en fait très approchée.

II. Considérons maintenant la fonction

$$y_{x+h} = \frac{m\,!}{(mp - x - h)\,!\,(mq + x + h)\,!}\, p^{mp-x-h} q^{mq+x+h}$$

et reportons-nous aux formules (approchées, mais très approchées) du n° 65 :

$$\left\{\begin{array}{l} -\lambda s_0' + \mu(s_0' + y_h) = s_1' \\ \lambda(s_0'' + y_h) - \mu s_0'' = s_1'' \\ s_0' = y_{1+h} + y_{2+h} + y_{3+h} + \cdots \\ s_0'' = y_{1+h} + y_{-2+h} + y_{-3+h} + \cdots \\ s_1' = y_{1+h} + 2y_{2+h} + 3y_{3+h} + \cdots \\ s_1'' = y_{-1+h} + 2y_{2+h} + 3y_{-3+h} + \cdots ; \end{array}\right.$$

les deux premières donnent par soustraction

$$\lambda - \mu = h = s_1'' - s_1'$$

puisque l'on a très sensiblement

$$s_0'' + y_h + s_0' = 1.$$

Comme

$$\mathrm{M}_a = \frac{y_{1+h} + 2y_{2+h} + 3y_{3+h} + \cdots - (y_{-1+h} + 2y_{-2+h} + 3y_{-3+h} + \cdots)}{\Sigma y_{x+h}},$$

on a

$$\mathrm{M}_a = \frac{s_1' - s_1''}{\Sigma y_{x+h}} = -h.$$

Il ne sera pas inutile d'approfondir cette question.

La *moyenne arithmétique* des nombres

$$y_{x+h} = \mathrm{A}\frac{m\,!}{(mp - x - h)\,!\,(mq + x + h)\,!}\, p^{mp-x-h} q^{mq+x+h}$$

où l'on donne à x toutes les valeurs entières possibles, c'est-à-dire toutes les valeurs entières telles que

$$mp - x - h \geqslant 0, \qquad mq + x + h \geqslant 0$$

est un nombre k tel que l'on ait

$$y_{-1+k} + 2y_{-2+k} + 3y_{-3+k} + \cdots = y_{1+k} + 2y_{2+k} + 3y_{3+k} + \cdots$$

1° m et mp sont entiers et h est nul.

Ici,

$$\begin{aligned} y_{-1} &+ 2y_{-2} + 3y_{-3} + \cdots = mpq \\ y_1 &+ 2y_2 \ \ + 3y_3 \ \ + \cdots = mpq \end{aligned}$$

donc la moyenne arithmétique M_a ou — h est nulle.

Soit $m = 100$; $p = 0{,}1$; $9 = 0{,}9$; $h = 0$, $A = 100$; la moyenne arithmétique est zéro ; c'est aussi la valeur de h. On a ici (page 68)

		Multiplicateurs	Produits
y_{-13}	= 0,0075	13	0,0975
	= 0,0198	12	0,2376
y_{-11}	= 0,0496	11	0,5456
	= 0,1171	10	1,1710
y_{-9}	= 0,2602	9	2,3418
	= 0,5426	8	4,3408
y_{-7}	= 1,0592	7	7,4144
	= 1,9292	6	11,5752
y_{-5}	= 3,2682	5	16,3410
	= 5,1304	4	20,5216
y_{-3}	= 7,4302	3	22,2906
	= 9,8788	2	19,7576
y_{-1}	= 11,9877	1	11,9877
			118,6124 = S_1''
y_0	= 13,1865		
y_1	= 13,0416	1	13,0416
y_2	= 11,4823	2	22,9646
y_3	= 8,8895	3	26,6685
y_4	= 6,9579	4	23,8316
y_5	= 3,3866	5	16,9330
y_6	= 1,5911	6	9,5466
y_7	= 0,5892	7	4,1244
y_8	= 0,1623	8	1,2984
y_9	= 0,0295	9	0,2655
S	= 99,9697		118,6742 = S_1'

On voit que $S_1'' = S_1'$; la petite différence vient de ce qu'on n'a pas épuisé les valeurs possibles de x. On a aussi

$$h = \frac{S_1'' - S_1'}{S} = -\frac{0{,}0618}{99{,}9697}$$

et on aurait $h = 0$ si l'on avait épuisé les valeurs possibles de x.

2° m et mp sont entiers ; h est entier et n'est pas nul.

Soit

$$m = 100\,; \qquad p = 0{,}1\,; \qquad q = 0{,}9\,; \qquad h = 2\,; \qquad A = 100.$$

Nous réalisons ce cas en écrivant la suite qui précède à partir de

$$y_h = 9{,}8788 = \mathrm{A}\ \frac{m\,!}{(mp - h)\,!\,(mq + h)\,!}\ p^{mp-h}q^{mq+h}\ ;$$

on a

		Multiplicateurs	Produits
y_{-11+h} =	0,0075	11	0,0925
y_{-10+h} =	0,0198	10	0,1980
y_{-9+h} =	0,0496	9	0,4464
y_{-8+h} =	0,1171	8	0,9368
y_{-7+h} =	0,2602	7	1,5214
y_{-6+h} =	0,5426	6	3,1556
y_{-5+h} =	1,0592	5	5,2960
y_{-4+h} =	1,9292	4	7,7168
y_{-3+h} =	3,2682	3	9,8046
y_{-2+h} =	5,1304	2	10,2608
y_{-1+h} =	7,4302	1	7,4302
			46,8631 = S_1''
y_h =	9,8788		
y_{1+h} =	11,9877	1	11,9877
y_{2+h} =	13,1865	2	26,3730
y_{3+h} =	13,0416	3	39,1248
y_{4+h} =	11,4823	4	45,9292
y_{5+h} =	8,8895	5	44,4475
y_{6+h} =	5,9579	6	34,8474
y_{7+h} =	3,3866	7	23,7062
y_{8+h} =	1,5911	8	12,7288
y_{9+h} =	0,5892	9	5,3028
y_{10+h} =	0,1623	10	1,6230
y_{11+h} =	0,0295	11	0,3245
S =	99,9970		246,3949 = S_1'

Ici

$$h = \frac{S_1'' - S_1'}{S} = -\frac{199{,}5318}{99{,}9970}\ ;$$

Nous trouvons un nombre un peu différent de — 2, parce que nous n'avons pas épuisé toutes les valeurs possibles de x, nous arrêtant à — 11 et à — 11 : sinon nous eussions trouvé exactement — 2.

Comme au 1°, la moyenne arithmétique correspond à la valeur 13,1865 de y ; elle correspond à y_0 ; donc à

$$y_{2+h} = y_{2-2} = y_0\,;$$

la moyenne arithmétique est donc 2 ; elle est égale à — h.

3° m et mp sont entiers, h est fractionnaire.

Dans la formule

$$y_{x+h} = \frac{m\,!}{(mp - x - h)\,!\,(mq + x + h)\,!}\, p^{mp-x-h} q^{mq+x+h},$$

la moyenne correspond, comme au 1°, à

$$\frac{m\,!}{mp\,!\;mq\,!}\, p^{mp} q^{mq}\,;$$

elle correspond donc à $x + h = 0$, donc à

$$x = -h\,;$$

par exemple pour

$$m = 100\,; \quad p = 0{,}1\,; \quad q = 0{,}9\,; \quad h = 0{,}3\,; \quad A = 100,$$

la moyenne arithmétique est (n° 68)

$$x = M_a = -0{,}3\,;$$

la valeur correspondante de y est y_0, tout comme précédemment.

4° m, mp, h, sont quelconques.

Nous sommes toujours en présence de la formule

$$y_{x+h} = \frac{m\,!}{(mp - x - h)\,!\,(mq + x + h)\,!}\, p^{mp-x-h} q^{mq+x+h}\,;$$

la moyenne arithmétique correspond encore à y_h; elle est donc encore

$$x = M_a = -h;$$

Cependant, cela suppose que

$$y_1 + 2y_2 + 3y_3 + \cdots = y_{-1} + 2y_{-2} + 3y_{-3} + \cdots$$

ce qui n'est rigoureusement exact que si m et mp sont entiers ; mais cette égalité est suffisamment approchée pour qu'on puisse l'employer comme si elle était exacte.

Il peut y avoir lieu de *placer* la moyenne obtenue.

Exemple (n° 88), la moyenne étant — **0,5055**,

pour 17/30 à 19/30 = 18/30 = 20/30 — 2/30		on a	y_{-1+h}
(Moyenne)	pour 20/30 — n/30	»	y_{-h+h}
pour 19/30 à 21/30 = 20/30		»	y_{-0+h}

on voit que

$$n = \frac{2}{30} \times 0{,}5055 \frac{1{,}0110}{30}$$

et que

$$M_a = \frac{20}{30} - \frac{1{,}0110}{30} = \frac{18{,}9890}{30}.$$

72. La mode est très sensiblement définie par (nº 49)

$$x + h = \frac{q - p}{2} \qquad \text{ou} \qquad x = \frac{q - p}{2} - h\,;$$

la mode et la moyenne arithmétique ne sont donc confondues que si

$$-h = \frac{q - p}{2} - h$$

ou si

$$q - p = 0,$$

donc si

$$p = q = 0{,}5.$$

Cependant, *dans le cas particulier de tirage de boules d'une urne avec mp entier*, les fréquences, ou les ordonnées à considérer ont pour abscisses

$$0, \quad \pm 1, \quad \pm 2, \quad \cdots;$$

l'ordonnée maximum, *parmi celles-ci*, a pour abscisse 0, et, d'autre part, la moyenne a pour abscisse 0 aussi : il y a identité entre l'une et l'autre ; mais l'abscisse de la plus grande des ordonnées à considérer n'est pas, absolument parlant, la mode.

La moyenne arithmétique diffère d'autant plus de la mode, abscisse de l'ordonnée maximum, que la courbe de probabilité diffère plus d'une courbe de probabilité simple, sauf cas tout à fait exceptionnels.

C'est une *erreur* de prendre la moyenne arithmétique pour valeur la plus probable ou mode, sauf dans le cas que nous venons d'envisager.

Mais on sait que la somme des écarts quadratiques

$$(x_1 - x_m)^2 + (x_2 - x_m)^2 + \cdots + (x_i - x_m)^2$$

est minimum si

$$x_m = M_a.$$

En effet, la somme en question, fonction de x_m est minimum si sa demi-dérivée par rapport à x_m est nulle. si

$$x_1 - x_m + x_2 - x_m + \cdots + x_i - x_m = 0$$

$$x_m = \frac{x_1 + x_2 + \cdots + x_i}{i} = \textit{moyenne arith. des } x_i.$$

73. Le calcul numérique de la moyenne arithmétique est grandement facilité par cette remarque : que retrancher un nombre constant p à chacun des termes dont on prend la moyenne, retranche aussi p de la moyenne ; par exemple

$$\frac{3+5+7+11}{4} = 6{,}5$$

$$\frac{(3-p)+(5-p)+(7-p)+(11-p)}{4} = 6{,}5-p.$$

Soit le tableau

$x =$	$y =$	$x' =$	$x'y =$
8	1	-4	-4
12	3	-3	-9
16	5	-2	-10
20	34	-1	-34
24	67	0	
28	21	1	21
32	23	2	46
36	8	3	24
40	7	4	28
	Total : 168		Total : -62

La moyenne des x' est

$$M_{x'} = -\frac{62}{168};$$

mais

$$x = 4x' + 24$$

donc

$$M_x = 4M_{x'} + 24 = -4 \times \frac{62}{168} + 24 = 25{,}476.$$

Cela est plus simple que de calculer

$$M_x = \frac{1 \times 8 + 3 \times 12 + 5 \times 16 + \cdots + 40 \times 7}{168}.$$

II. — MÉDIANE

74. A propos de la moyenne, s'introduit la notion de *Médiane*
Soit le tableau

Écarts	Fréquences
$x = 7$	$y = 28$ observations
12	36 »
17	42 »
22	26 »
27	20 »

Ce tableau est à interpréter comme il suit : il y a

28 observations comprises entre $x = 7 - \frac{12-7}{2}$ et $x = 7 + \frac{12-7}{2}$

donc 28	»	»	»	$x = 4,5$	et	$x = 9,5$
36	»	»	»	9,5	»	14,5
42	»	»	»	14,5	»	19,5
26	»	»	»	19,5	»	24,5
20	»	»	»	24,5	»	29,5

par conséquent

de $x = 4,5$	à	$x = 9,5$	il y a	$y_1 =$	28 observations
4,5	»	14,5	»	$y_2 = y_1 + 36 = 64$	»
4,5	»	19,5	»	$y_3 = y_2 + 42 = 106$	»
4,5	»	24,5	»	$y_4 = y_3 + 26 = 132$	»
4,5	»	29,5	»	$y_5 = y_4 + 20 = 152$	»

Quelle est l'abscisse x qui correspond à $\frac{152}{2} = 76$ observations ?
Cette abscisse x est la *médiane.*
A défaut de formules donnant

$$\begin{cases} y_1 + y_2 + \cdots \\ y_{-1} + y_{-2} + \cdots \end{cases}$$

en fonction de m, p, q ou

$$\begin{cases} y_{1+h} + y_{2+h} + \cdots \\ y_{-1+h} + y_{-2+h} + \cdots \end{cases}$$

en fonction de m, p, q, h, on calcule approximativement la médiane en faisant une interpolation par parties proportionnelles :

76 étant compris entre 64 et 106, la médiane x_m est comprise entre

$$14{,}5 \quad \text{et} \quad 19{,}5 \,;$$

interpolons par parties proportionnelles dans cet intervalle ;

$$\frac{x_m - 14{,}5}{19{,}5 - 14{,}5} = \frac{76 - 64}{106 - 64} = \frac{12}{42} \quad \text{d'où} \quad x_m = 14{,}5 + 5 \times \frac{12}{42} = 15{,}9 \,;$$

il serait *vain* de calculer beaucoup de décimales, d'écrire p. e.

$$x_m = 15{,}929,$$

procéder par parties proportionnelles étant une approximation.

Même dans les cas les plus simples, on ne sait pas calculer la *médiane*. On se sert souvent de la relation suivante :

$$M_0 + 2M_a = 3M$$

où M_0 est la mode, M_a la moyenne arithmétique, M la médiane ; cette relation *empirique* est *illusoire* (1).

III. — MOYENNES GÉOMÉTRIQUES ET MOYENNE HARMONIQUE

75. A la moyenne arithmétique se rattachent la moyenne géométrique et la moyenne harmonique.

Moyenne géométrique. — Soient des quantités $x_1, x_2, \cdots, x_n$; leur moyenne géométrique M_g est

$$M_g = \sqrt[n]{x_1 x_2 \cdots x_n} \,;$$

par conséquent

$$\log M_g = \frac{\log x_1 + \log x_2 + \cdots + \log x_n}{n} \,;$$

la moyenne géométrique peut être définie ainsi : *son logarithme est la moyenne arithmétique des logarithmes des éléments.*

Moyenne harmonique. — La moyenne harmonique de

$$x_1, \quad x_2, \quad \cdots, \quad x_n$$

est définie par

$$\frac{1}{M_h} = \frac{\dfrac{1}{x_1} + \dfrac{1}{x_2} + \cdots + \dfrac{1}{x_n}}{n}.$$

(1) Le lecteur pourra consulter à ce propos un Mémoire de M. L. BENDERSKY, *Bulletin des sciences mathématiques*, 1930 (non encore paru).

On ne connaît aucun moyen de déterminer exactement, ou approximativement, la moyenne géométrique ou la moyenne arithmétique.

La moyenne arithmétique est plus grande que la moyenne géométrique ; la moyenne géométrique est plus grande que la moyenne harmonique : il est facile de le démontrer.

IV. — DISPERSION

76. Soit M_a la moyenne arithmétique et $\xi_i = x_i - M_a$ les écarts comptés à partir de la moyenne arithmétique. Soient encore $f(x_i)$ les nombres de fréquences d'abscisses x_i.

L'écart quadratique, ou DISPERSION, est défini par

$$\sqrt{\frac{1}{N}\Sigma\xi_i^2 f(x_i)}, \qquad N = \Sigma f(x_i).$$

On *admet*, en statistique, que la *répartition* ou DISPERSION *des fréquences est mesurée par l'écart quadratique.* On notera que la *dimension* de l'écart quadratique est la même, *et cela est nécessaire*, que la dimension des grandeurs x.

V. — ÉCART MOYEN

77. L'ÉCART MOYEN est la moyenne des valeurs absolues des écarts entre les grandeurs x_i et leur médiane **M** :

$$\frac{1}{N}\Sigma \mid x_i - M \mid f(x_i).$$

I. Dans le cas de la *fonction de probabilité simple* (n° 57), on a

$$\text{Ecart moyen} = E_m = \frac{\cdots 3y_{-3} + 2y_{-2} + y_{-1} + y_1 + 2y_2 + 3y_3 + \cdots}{\cdots y_{-3} + y_{-2} + y_{-1} + y_0 + y_1 + y_2 + y_3 + \cdots}.$$

Le dénominateur a pour valeur *un*; le numérateur a pour valeur (nº 61, formule 6) :

$$s_1' + s_1'' = 2mpqy_0.$$

Or (nº 28, formule 1), *approximativement* :

$$2mpqy_0 = 2mpq \frac{1}{\sqrt{2\pi mpq}} = \frac{1}{\sqrt{\pi}} \sqrt{2mpq}.$$

Comme (nº 48, formule 15)

$$\mathrm{E}_p = \text{Ecart probable (moyen)} = (\text{appx}^{\text{t}})\ 0{,}447 \sqrt{2mpq} - 0{,}5,$$

on a

(1) $$\mathrm{E}_p = 0{,}845\ \mathrm{E}_m - 0{,}5 \qquad \text{(formule approchée).}$$

II. Toujours dans le cas de la *fonction de probabilité simple* (nº 57), on a (nº 76)

$$\text{Ecart quadr.} = \mathrm{E}_q = \frac{\cdots 3^2y_{-3}+2^2y_{-2}+y_{-1}+y_1+2^2y_2+3^2y_3+\cdots}{\cdots y_{-3}+y_{-2}+y_{-1}+y_0+y_1+y_2+y_3+\cdots},$$

$$\mathrm{E}_q = s_2' + s_2''.$$

Or (nº 61, formule 6)

$$s_2' + s_2'' = mpqy_0 + mpq(s_0' + s_0'') = mpq(y_0 + s_0' + s_0'');$$

donc

$$\mathrm{E}_q = mpq;$$

comme E_m a pour valeur (ci-dessus)

$$2mpqy_0 = \frac{1}{\sqrt{\pi}} \sqrt{2mpq},$$

il en résulte

$$\mathrm{E}_m = \mathrm{E}_q \sqrt{\frac{2}{\pi}},$$

$$\mathrm{E}_m = 0{,}798\ \mathrm{E}q \qquad \text{(formule approchée).}$$

CHAPITRE IX

CALCUL « A POSTERIORI » DES CONSTANTES m, p, q, h D'UNE LOI DE PROBABILITE SIMPLE ; CONDUITE DES CALCULS. ECARTS PROBABLES ABSOLUS ET RELATIFS

Notation. — Nous écrirons désormais y_{x+h} : la distinction entre y_{x+h} et Y_{x+h} n'étant plus nécessaire dans les applications.

78. *Nous poursuivons l'étude du problème fondamental que voici :*

On donne un jeu complet de nombres y_x vérifiant une loi de probabilité simple inconnue

$$(1) \qquad y_{x+h} = \mathrm{N} \frac{m\,!}{(mp - x - h)\,!\,(mq + x + h)}\, p^{mp-x-h} q^{mq+x+h}$$
$$x = 0 \pm 1, \pm 2, \pm 3, \cdots$$

on propose de calculer les constantes m, p, q, h.

Nous avons déjà étudié un problème analogue n° 68. Nous allons donner ici un type de calcul, concernant des nombres y_{x+h}, *qui ne suivent pas tout à fait exactement une loi de probabilité* (1). Nous introduisons des vérifications utiles aux calculateurs. Nous obtiendrons une loi (1), *à peu près suivie* par les nombres donnés.

On remarquera que la valeur de m est fractionnaire, tandis qu'elle est entière dans les nombres y_{x+h} dérivés de tirages de boules d'une urne. Cela n'a rien de surprenant. La soustraction, la division, opérations inverses de l'addition, de la multiplication, introduisent les nombres négatifs et fractionnaires, qui généralisent les nombres entiers positifs. De même nous faisons une opération *inverse* de celle qui consiste à calculer des probabilités de tirage de boules d'une urne ; cette opération inverse introduit des valeurs fractionnaires de m [1]. Nous reviendrons sur ce point à la fin du chapitre XI.

Nous partons de la suite O (Données ou Observations), page 112. Nous désignons la plus grande valeur de y, dans la suite O, par y_{x+h} ; exceptionnellement (n° 93, G) la plus grande valeur de y calculée (C) ne correspondra pas à y_{x+h} de O : cela n'a pas d'importance.

Nous utilisons toutes les données (à l'exemple de ce qui se fait quand on applique la Méthode des moindres carrés) : cela est *nécessaire* car les données, expérimentales, ne suivent jamais exactement une loi de probabilité simple, mais s'en écartent parfois très peu, parfois beaucoup, comme on le verra dans les Applications, et IL FAUT COMPENSER LES ÉCARTS EXPÉRIMENTAUX PAR L'UTILISATION DE TOUTES LES OBSERVATIONS.

79. Reportons-nous au Tableau de la page 113, à la colonne O qui sont les données (O = observé). On a (n° 66)

$$h = \frac{S_1'' - S_1'}{9\,998} = 0{,}15323 \qquad \log h = \bar{1}{,}185\,3457.$$

Ensuite

par logarithmes	$S_0''S_1' = 34\ 626\ 700$	directement	$S_0''S_1' = 34\ 626\ 696$
»	$S_0'S_1'' = 34\ 363\ 870$	»	$S_0'S_1'' = 34\ 363\ 875$

d'où

$A = S_0''S_1' + S_0'S_1''$ et ici $A = 68\,990\,571$ [2]

par logarithmes	$S_2'S_1'' = 205\ 042\ 400$	directement	$S_2'S_1'' = 205\,042\,425$
»	$S_2''S_1' = 194\ 078\ 300$	»	$S_2''S_1' = 194\,078\,295$
»	$S_1'S_1'' = 71\ 816\ 330$	»	$S_1'S_1'' = 71\,816\,325$

d'où

$B = S_2'S_1'' + S_2''S_1' - S_1'S_1''$ et ici $B = 327\,304\,395$ [2]

(1) *Revue Générale des Sciences*, 31 mai et 15 juin 1928.
(2) Il eut suffi de prendre les nombres calculés par logarithmes.

		O.			Vérification
	$y_{-7+h} =$	3	× 7 = 21	× 7 = 147	3 × 7² = 147
		29	× 6 = 174	× 6 = 1 044	29 × 6² = 1 044
		140	× 5 = 700	× 5 = 3 500	140 × 5² = 3 500
		372	× 4 = 1 488	× 4 = 5 952	372 × 4² = 5 952
		801	× 3 = 2 403	× 3 = 7 209	801 × 3² = 7 209
		1 362	× 2 = 2 724	× 3 = 5 448	1 362 × 2² = 5 448
	$y_{-1+h} =$	1 765	× 1 = 1 765	× 1 = 1 765	1 765 × 1² = 1 765
4 472		4 472 = S_0''	9 275 = S_1''	25 065 = S_2''	
1 821	$y_h =$	1 821			
	$y_{1+h} =$	1 554	× 1 = 1 554	× 1 = 1 554	1 554 × 1² = 1 554
		1 058	× 2 = 2 116	× 2 = 4 232	1 058 × 2² = 4 232
		592	× 3 = 1 776	× 3 = 5 328	592 × 3² = 5 328
		316	× 4 = 1 264	× 4 = 5 056	316 × 4² = 5 056
		122	× 5 = 610	× 5 = 3 050	122 × 5² = 3 050
		32	× 6 = 192	× 6 = 1 152	32 × 6² = 1 152
		20	× 7 = 140	× 7 = 980	20 × 7² = 980
		8	× 8 = 64	× 8 = 512	8 × 8² = 512
	$y_{9+h} =$	3	× 9 = 27	× 9 = 243	3 × 9² = 243
3 705		3 705 = S_0'	7 743 = S_1'	22 107 = S_2'	
9 998 = S					

80. Calcul de p, q.

On a (n° 66, formules 39).

$$p - \left(h - h\frac{S_0'S_0''}{A}\right) = \frac{S_2'}{S_1'} - \frac{C}{A}\frac{S_0'}{S_1'}$$

$$q + \left(h - h\frac{S_0'S_0''}{A}\right) = \frac{S_2''}{S_1''} - \frac{C}{A}\frac{S_0''}{S_1''}.$$

Ici,

$$\frac{S_2'}{S_1'} = 2{,}85509 \qquad \frac{S_2''}{S_1''} = 2{,}70243$$

$$\frac{C}{A}\frac{S_0'}{S_1'} = 2{,}27008 \qquad \frac{C}{A}\frac{S_0''}{S_1''} = 2{,}28744$$

donc

$$p - \left(h - h\frac{S_0'S_0''}{A}\right) = 0{,}58501,$$

$$q + \left(h - h\frac{S_0'S_0''}{A}\right) = 0{,}41499,$$

Vérification : la somme des deux premiers membres est (et doit être) 1.

De

$$h = 0{,}15323$$

$$h\frac{S_0'S_0''}{A} = 0{,}03680, \qquad h - h\frac{S_0'S_0''}{A} = 0{,}11643,$$

il résulte

$$p = 0{,}58501 + 0{,}11643 = 0{,}70144$$
$$q = 0{,}41499 - 0{,}11643 = 0{,}29856$$

et on a bien

$$p + q = 1.$$

81. Calcul de λ, μ.

On a (n° 66 formule 38) :

$$\lambda = \frac{B}{A} + h\frac{S_0''S_1'}{A}, \qquad \mu = \frac{B}{A} - h\frac{S_0'S_1''}{A}$$

Ici,

$$\frac{B}{A} = 4{,}74419, \qquad h\frac{S_0''S_1'}{A} = 0{,}07691, \qquad h\frac{S_0'S_1''}{A} = 0{,}07632$$

d'où

$$\lambda = 4{,}82110 \quad \mu = 4{,}66787\ ; \quad \textit{vérification}\ (\text{n}^\text{o}\ 66,\ \text{formule } 20)$$
$$\lambda - \mu = 0{,}15323 = h.$$

Vérification : appliquons les formules (40) du n° 66 :

ici

$$\lambda - \mu = h = 0{,}15323$$
$$\frac{S_2''}{S_1''} = 2{,}70243, \qquad \mu\frac{S_0''}{S_1''} = 2{,}25064\ ;$$

il en résulte, comme tout à l'heure :

$$q = 0{,}29856\ ;$$

il suffit de vérifier l'un des deux nombres p ou q.

82. Calcul de mp, mq, avec vérification :

On a (n° 66, formule 19)

		Vérification :
$mpq = \lambda - ph$	$mpq = \mu + qh$	
$ph = 0{,}10748$	$qh = 0{,}04575$	$ph + qh = h$
$mpq = 4{,}71362$	$mpq = 4{,}71362$	
$mp = 15{,}78838$	$mq = 6{,}71982$	
$m = 22{,}50820$	$m = 22{,}50820$	$mp + mq = m$

83. **Calcul des y' en négligeant la 5e décimale :**

$$mp - h = 15{,}6352 \qquad mq + h = 6{,}8730\ ;$$

1° Les formules (5) du n° 55 donnent

$$y'_{-1+h} = \frac{mq+h}{mp-h+1}\frac{p}{q}\,y'_h\ ; \qquad y'_{-2+h} = \frac{mq+h-1}{mp-h+2}\frac{p}{q}\,y'_{-1+h}\ ;$$
$$y'_{-3+h} = \frac{mq+h-2}{mp-h+3}\frac{p}{q}\,y'_{-2+h}\ ; \ \ldots$$

on prend $y_h' = 1$;

2° Les mêmes formules donnent

$$y'_{1+h} = \frac{mp - h}{mq + h + 1}\,\frac{q}{p}\,y_h'\,; \qquad y'_{2+h} = \frac{mp - h - 1}{mq + h + 2}\,\frac{q}{p}\,y'_{1+h}\,;$$

$$y'_{3+h} = \frac{mp - h - 2}{mq + h + 3}\,\frac{q}{p}\,y'_{2+h}\,; \ldots$$

on prend $y_h' = 1$ (n° 56) ;
on trouve ainsi les nombres suivants :

$y'_{-1+h} = 0{,}97073$	$y'_{1+h} = 0{,}84525$	
$y'_{-2+h} = 0{,}75955$	$y'_{2+h} = 0{,}59338$	
$y'_{-3+h} = 0{,}46666$	$y'_{3+h} = 0{,}34879$	
$y'_{-4+h} = 0{,}21627$	$y'_{4+h} = 0{,}17251$	
$y'_{-5+h} = 0{,}07075$	$y'_{5+h} = 0{,}07195$	
$y'_{-6+h} = 0{,}01439$	$y'_{6+h} = 0{,}02530$	
$y'_{-7+h} = 0{,}00130$	$y'_{7+h} = 0{,}00748$	$(S_0'')' = 2{,}49965$
	$y'_{8+h} = 0{,}0\ 185$	$y_h' = 1$
$(S_0'')' = 2{,}49965$	$y'_{9+h} = 0{,}00038$	$(S_0') = 2{,}06697$
	$(S_0')' = 2{,}06689$	$S' = 5{,}56654$

3° Pour avoir les y, il suffit de multiplier les y' par

$$\frac{S}{S'} = \frac{9998}{5{,}56662} = 1796 = y_h$$

on trouve ainsi les nombres C_1 du tableau qui suit qui, arrondis à l'unité donnent les nombres C_2. Ensuite, la somme des C_2 étant 9996 et la somme des données étant 9998, on procède comme il a étê dit au n° 53 : *on multlplie les nombres* C_2 *par* $1 + \varepsilon$, *où* ε *est choisi de manière telle que deux des nombres* C_2 *et deux seulement,* 1743 et 627, *soient augmentés chacun d'une unité.* On obtient ainsi les nombres *définitifs* C, dont la somme est 9998.

84. Le calcul des y' et des y a été fait par logarithmes à 7 décimales. En faisant le calcul par logarithmes à 5 décimales, à partir de

$$mp = 15{,}788 \quad mq = 6{,}720 \quad h = 0{,}153 \quad p = 0{,}701 \quad q = 0{,}299,$$

on trouve

$$S' = 5{,}5650 \qquad \frac{S}{S'} = \frac{9998}{5{,}5650} = 1797 :$$

à une unité près pour quelques-uns, les y' ainsi calculés sont identiques aux y calculés par logarithmes à 7 décimales.

	Observé O	Calculé			O — C		Écarts prob. des C
		C_1	C_2	C	+	—	±
1.........	$y_{-7+h} =$ 3	2,3	2	2	1		
2.........	29	25,9	26	26	3		2,9
3.........	140	127,1	127	127	13		7,1 *
4.........	372	388,4	388	388		16	12,8 *
5.........	801	838,1	838	838		37	18,2 *
6.........	1 362	1 364,2	1 364	1 364		2	22,7
7.........	1 765	1 743,49	1 743	1 744	21		25,1
8.........	$y_h =$ 1 821	1 796,1	1 796	1 796	25		25,4
9.........	1 554	1 518,1	1 518	1 518	36		23,7 *
10.........	1 058	1 065,7	1 066	1 066		8	20,3
11.........	592	626,45	626	627		35	15,9 *
12.........	316	309,8	310	310	6		11,4
13.........	122	129,2	129	129		7	7,1
14.........	32	45,54	46	46		14	4,1 *
15.........	20	13,46	13	13	7		1,5
16.........	8	3,3	3	3	5		
17.........	$y_{9+h} =$ 3	0,7	1	1	2		
	9 998		9 996	9 998	119	119	

85. Cette statistique se rapporte aux *nombres premiers*. Nous l'avons désignée par F dans le classement général : nos 90 à 92.

A gauche de la colonne O, on a inscrit les nombres 1, 2, 3, 4.

En voici l'explication : les nombres premiers inférieurs à 10^6 se répartissent comme il suit :

3 centaines	contiennent chacune....	1 nombre premier		
29 »	»	2 nombres premiers		
140 »	»	3 »		
372 »	»	4 »		etc.

En tout

$$3 \times 1 + 29 \times 2 \times 140 \times 3 \times 372 \times 4 + \cdots = 78498$$

nombres premiers.

Ne figurent pas :

1 centaine contenant 21 nombres premiers, 1 centaine contenant 25 nombres premiers. Il n'existe pas de centaines contenant 18 ou 19 ou 20 ou 22, 23, 24 nombres premiers [1].

[1] M. Kraitchik, *Recherches sur la Théorie des Nombres*, p. 24, Gauthier-Villars, 1924.

La fonction de probabilité simple définie par les nombres 9998 ;

$$h = 0{,}15323\,; \quad p = 0{,}70144\,; \quad q = 0{,}29856\,; \quad m = 22{,}50820\,;$$

qui donne les nombres de la colonne C, est une forme approchée de la fonction qui représente la distribution des nombres premiers par centaines.

Les écarts probables des nombres C doivent être calculés et doivent être mis en regard des différences O — C ; (nous avons marqué d'un * les écarts probables placés en face des O — C les plus grands) on se sert pour cela de la formule (n° 52)

$$\text{Ec. prob.} = \pm\, 0{,}477 \sqrt{2 \times \mathrm{S} \times \frac{y}{\mathrm{S}} \times \frac{\mathrm{S} - y}{\mathrm{S}}},$$

$$= \pm\, 0{,}477 \sqrt{2y \times \frac{\mathrm{S} - y}{\mathrm{S}}}\,;$$

ainsi, pour **1066**, l'écart probable est

$$\pm\, 0{,}477 \sqrt{2 \times 1066 \times \frac{9\,998 - 1\,066}{9\,998}} - 0{,}5 \doteq \pm\, 20{,}3.$$

Nous verrons au n° **90** l'usage à faire des écarts probables.

Ecart probable relatif

86. Dans certaines statistiques on considère non pas les nombres absolus d'événements, mais leurs pourcentages ; il arrivera, par exemple, qu'on divisera par 5 les nombres O de tout à l'heure.

Dans ce cas l'écart probable est divisé par $\sqrt{5}$ à 0,5 près.

Si l'on divise par n, l'écart probable, à 0,5 près, est divisé par $\sqrt{n}$, à peu près.

En effet, au lieu de

$$1\,066$$

de tout à l'heure, on a

$$1\,066 : 5 = 213{,}2\,;$$

au lieu du total 9998, on a

$$9\,998 : 5 = 1\,999{,}6\,;$$

donc l'écart probable est

$$\pm 0{,}477 \sqrt{2 \times \frac{1066}{5} \times \frac{9998 : 5 - 1066 : 5}{9998}} - 0{,}5,$$

ou

$$\frac{1}{\sqrt{5}} \sqrt{2 \times 1066 \times \frac{9998 - 1066}{9998}} - 0{,}5.$$

Il faut bien prendre garde que ces écarts probables sont *relatifs*, tandis que tout à l'heure, ils étaient *absolus*.

Nous en trouverons plus loin des statistiques où les écarts sont tantôt absolus, tantôt relatifs.

Voici ce qu'on obtient pour la statistique précédente, en divisant les nombres O et C par 5.

O'	C'	O' — C' +	O' — C' —	Écarts prob. des C' ±
0,6	0,4	0,2		
5,8	5,2	0,6		1,0
28,0	25,4	2,6		2,9
74,4	77,6		3,2	5,4
160,2	167,6		7,4	7,9
272,4	272,8		0,4	9,8
353,0	348,8	4,2		10,9
364,2	359,2	5,0		11,1
310,8	303,6	7,2		10,5
211,6	213,2		1,6	8,8
118,4	125,4		7,0	6,8
63,2	62,0	1,2		4,8
24,4	25,8		1,4	2,9
6,4	9,2		2,8	1,5 *
4,0	2,6	1,4		0,6 *
1,6	0,6	1,0		
0,6	0,2	0,4		

On voit que les écarts probables ne sont plus dépassés que dans deux cas : mais ici, les écarts probables sont *relatifs*.

Du non-dépassement de ces écarts probables *relatifs, on ne peut pas conclure que la statistique suit une loi de probabilité* *simple dans toute son étendue, exception faite pour les valeurs de x qui correspondent aux astérisques* : c'est la statistique *réduite* qui suit une loi de probabilité simple, sauf aux deux points précités, ce n'est pas la statistique vraie.

Toutefois l'emploi des statistiques réduites est très utile quand les différences O — C dépassent souvent, bien que d'assez peu, les écarts probables ; ces dépassements sont moindres dans les statistiques réduites correspondantes et les anomalies importantes sont ainsi mieux mises en évidence dans les statistiques réduites que dans les statistiques vraies (*Exemple* : Statistique E, n° 91).

CHAPITRE X

EXISTENCE DE STATISTIQUES SUIVANT LA LOI DE PROBABILITE SIMPLE

---

87. Il existe des statistiques qui suivent la Loi de probabilité simple, compte tenu des écarts probables.

Il est très important que de telles statistiques existent : cette existence est utile pour justifier ce que nous dirons au n° 90.

Nous allons étudier trois statistiques de ce genre.

La première est une statistique réduite, où l'on considérera les écarts probables relatifs. La seconde et la troisième sont des statistiques vraies, c'est-à-dire non réduites, où l'on considérera les écarts probables absolus.

Les observations de la température de l'air faites à Montsouris pendant les 50 années 1873-1922 ont donné lieu aux résultats T que voici [1] (Cette statistique est désignée par A_1 dans le classement du n° 90).

[1] L. Besson, *La température à Paris depuis* 50 *années d'observations.* Annales de l'Observatoire de Montsouris, 1926, p. 221.

Mois	Température maximum T la plus fréquemment observée dans le mois	T — 7 observé O	T — 7 calculé C	O — C +	O — C —	Ecarts prob. des C ±
Janvier........	8°	$y_{-6+h} = 1$	1	0	0	
Février........	9	2	2	0	0	
Mars..........	11	$y_{-4+h} = 4$	3	1		0,7
Avril..........	13	$y_{-3+h} = 6$	7		1	1,3
Mai...........	18	11	11	0	0	
Juin...........	22	15	15	0	0	
Juillet.........	24	$y_h = 17$	17	0	0	
Août..........	23	16	16	0	0	
Septembre.....	18	11	11	0	0	
Octobre.......	13	6	6	0	0	
Novembre.....	9	2	2	0	0	
Décembre.....	7	$y_{5+h} = 0$	0	0	0	
		91	91	1	1	

On a retranché 7 de T pour avoir une suite 0 dont les termes extrêmes soient nuls, car il n'est absolument rien qui impose 0° comme point de départ : on peut aussi bien prendre 7° que 0°.

Ici,

$$S = 91 \qquad S_0'' = 39 \qquad S_1'' = 87 \qquad S_2'' = 263$$
$$S_0' = 35 \qquad S_1' = 64 \qquad S_2' = 146$$
$$h = 0{,}253 \qquad A = 5\,541 \qquad B = 23\,966$$
$$p = 0{,}107 \qquad q = 0{,}893 \qquad \lambda = 4{,}439 \qquad \mu = 4{,}186$$
$$mp = 4{,}941 \qquad mq = 41{,}235 \qquad m = 46{,}176\,;$$

l'écart probable qui concerne y_{-4+h}

$$\pm\left(0{,}477\sqrt{2 \times 91 \times \frac{3}{91} \times \frac{88}{91}} - 0{,}5\right) = \pm\, 0{,}7$$

n'est dépassé que d'une manière insignifiante ; l'écart probable qui concerne y_{-3+h} n'est pas dépassé.

Ici, les écarts probables sont *relatifs*.

La figure est tout à fait analogue à la figure 4 : les écarts 0 — C y sont imperceptibles.

88. Voici un exemple de statistique *vraie*, suivant la loi de probabilité simple *sans que les écarts probables soient dépassés* (fig. 7).

Cette statistique est désignée par A_2 dans le classement du n° 90.

La colonne *observé* est une statistique de la fécondité de juments sous certaines conditions [1]...

	O observé	C calculé	O — C +	O — C —	Écarts probl. des B ±
1/30 à 3/30	$y'_{-9+h} =$ 2	$y_{-9+h} =$ 1,3			
3/30 » 5/30	7,5	4,0			
5/30 » 7/30	11,5	10,9	0,6		1,7
7/30 » 9/30	21,5	26,7		5,2	3,0
9/30 » 11/30	55	57,4		2,4	4,5
11/30 » 13/30	104,5	108,8		4,3	6,4
13/30 » 15/30	182	180,2	1,8		8,1
15/30 » 17/30	271,5	259,0	12,5		9,6
17/30 » 19/30	$y'_{-1+h} =$ 315	$y_{-1+h} =$ 320,0		5,0	10,6
19/30 » 21/30	$y_h =$ 337	$y_h =$ 335,8	1,2		10,8
21/30 » 23/30	$y'_{1+h} =$ 293,5	$y_{1+h} =$ 294,6		1,1	10,2
23/30 » 25/30	204	211,1		7,1	8,8
25/30 » 27/30	127	119,8	7,2		6,7
27/30 » 29/30	49	51,3		2,3	4,3
29/30 » 30/30	$y'_{5+h} =$ 19	19,2		0,2	2,5
	Total : 2 000	Total : 2 000,1	27,5	27,6	

On a ici

$$S_0'' = 970{,}5 \qquad S_1'' = 2\,384{,}5 \qquad S_2'' = 8\,068{,}5$$
$$S_0' = 692{,}5 \qquad S_1' = 1\,373{,}5 \qquad S_2' = 3\,511{,}5$$

$h = 0{,}5055 \quad A = 2\,984\,248 \quad B = 16\,173\,278$

$p = 0{,}2158 \quad q = 0{,}7842$

$\lambda = 5{,}6453 \quad \mu = 5{,}1398, \quad mp = 7{,}0597, \quad mq = 25{,}6543.$

Le calcul donne les nombres de la colonne B.

[1] Statistique donnée par G. UDNY YULE, *An Introduction to the Theory of Statistics*, 6e édition, London, 1922, p. 96.

Pour ces nombres B, on a

$$\begin{array}{llll} S_0'' = 968{,}3 & \text{au lieu de} & 970{,}5 & \text{observé} \\ S_0' = 696{,}0 & \text{»} & 692{,}5 & \text{»} \end{array}$$

La probabilité de l'écart h, quand on fait **2000,1** épreuves, est $\dfrac{\mathbf{335{,}8}}{\mathbf{2000{,}1}} = p'$; la probabilité complémentaire est

$$q' = 1 - \frac{335{,}8}{2\,000{,}1} = \frac{1\,664{,}3}{2\,000{,}1};$$

l'écart probable qui se rapporte à **335,8** est

$$\pm \left(0{,}477 \sqrt{2 \times 2\,000{,}1 \times \frac{335{,}8}{2\,000{,}1} \times \frac{1\,664{,}3}{2\,000{,}1}} - 0{,}5\right) = \pm\, 10{,}8.$$

L'écart probable n'est dépassé que de peu quand il l'est et les O — C ne suivent pas de loi définie : *les données* A *suivent donc une loi de probabilité simple.*

Pour la fonction

$$y = \frac{m\,!}{(mp - x - h)\,!\,(mq + x + h)\,!}\, p^{mp-x-h} q^{mq+x+h},$$

la mode est donnée par la relation (n° 49)

$$x' + h = \frac{q - p}{2} = \frac{0{,}7842 - 0{,}2158}{2} = 0{,}2842$$

et

$$\begin{aligned} x' &= 0{,}2842 - h = 0{,}2842 - 0{,}5055 \\ &= -\,0{,}2213 = (-\,0{,}2213 - h) + h = -\,0{,}7268 + h\,; \end{aligned}$$

y_{-1+h} se rapporte à l'interv. 17/30 à 19/30, donc se rapporte à 18/30;
y_h » 19/30 à 21/30, » 20/30;

$$\begin{array}{lll} 18/30 \text{ correspond à} & y_{-1+h} = y_{-1+0{,}5055} = y_{-0{,}4945}\,; \\ n/30 \quad\text{»} & y_{-0{,}7268+h} = y_{-0{,}7268+0{,}5055} = y_{-0{,}2213} \\ 20/30 \quad\text{»} & y_{0+h} = y_{0+0{,}5055} = y_{0{,}5055}\,; \end{array}$$

on voit que, dans l'échelle

$$-\,1 \qquad -\,0{,}7268 \qquad 0,$$

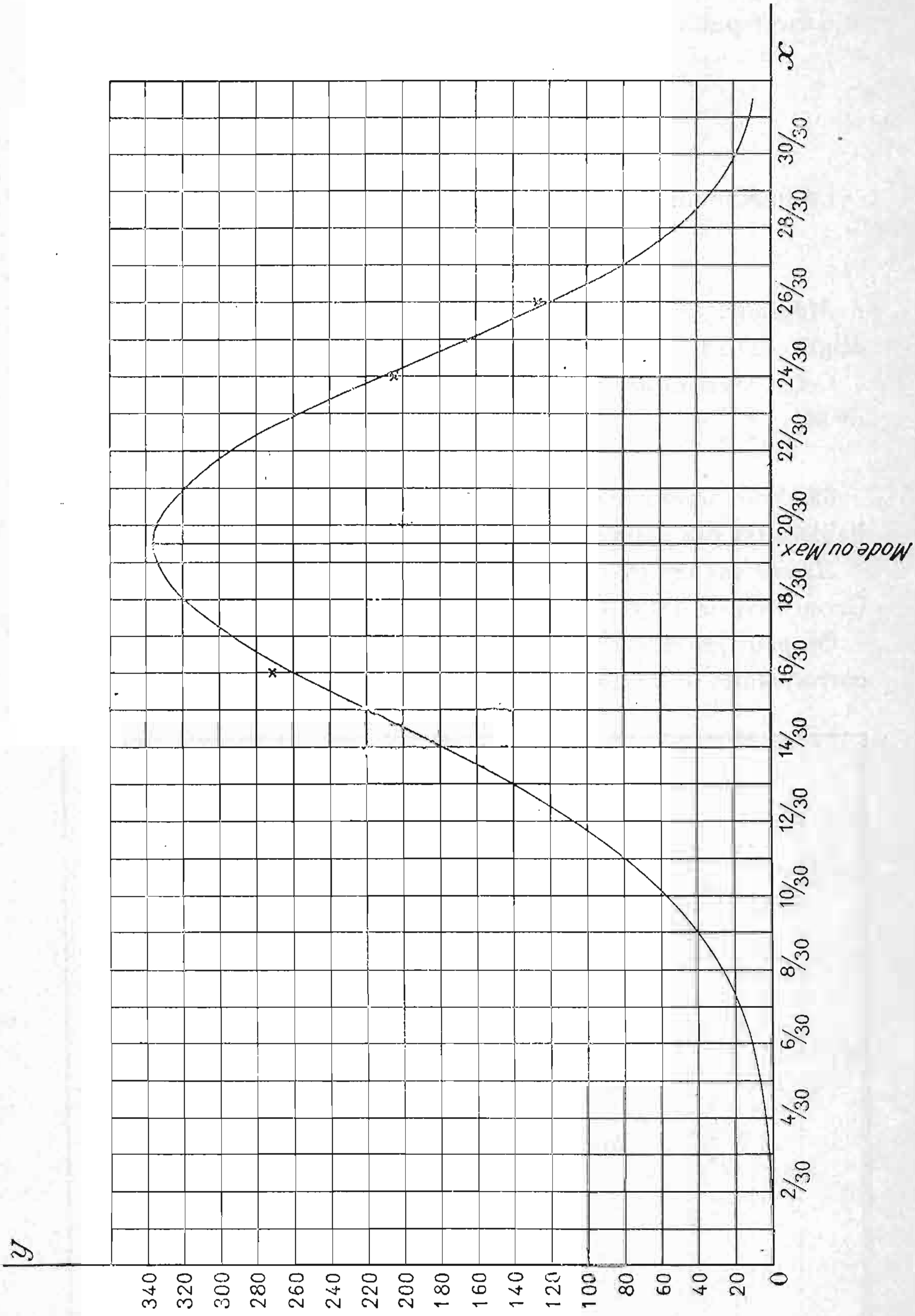


Fig. 7. — Exemple de statistique suivant presqu'exactement la Loi de probabilité simple. Sauf les trois +, qui sont presque sur la courbe, obtenue par le calcul, toutes les autres données sont sur la courbe.

n correspond à — 0,7268 ; donc

$$\frac{n}{20-18} = \frac{-0{,}7268}{-1} = 0{,}7268$$
$$n = 18 + 2 \times 0{,}7268 = 19{,}4536$$

Ce maximum a donc lieu pour

$$19{,}4536/30.$$

Moyenne. — La moyenne relative à cette statistique a été étudiée nº **73**.

Cette statistique est désignée par A_3 dans le classement du nº 90.

89. Voici un autre exemple de statistique vraie, où l'écart probable n'est pas dépassé.

Ascensions droites de l'Etoile polaire : 487 observations faites à Greenwich de 1836 à 1839 inclusivement.

On peut les répartir comme il suit par $0^s,5$ de temps; on a fait correspondre 0^s au nombre le plus élevé d'observations :

	Observé O	Calculé C	O — C +	O — C —	Écarts prob. ±
— 3^s,5	$y_{-7+h} =$ 1	2		1	
— 3	6	4	2		
— 2,5	12	11	1		1,8
— 2	21	21	0	0	
— 1,5	36	38		2	3,2
— 1	61	58	3		4,3
— 0,5	$y_{-1+h} =$ 73	75		2	4,9
0^s	$y_h =$ 82	84		2	5,2
+ 0,5	$y_{1+h} =$ 72	77		5	4,9
1	63	59	4		4,3
1,5	38	35	3		3,2
2	16	16	0	0	
2,5	5	6		1	0,7
3	1	1	0	0	
	487	487	13	13	

On a ici

$$h = 0{,}1704 \qquad A = 181\,020 \qquad B = 958\,834$$
$$p = 0{,}2550 \qquad q = 0{,}7450 \qquad \lambda = 5{,}3773 \qquad \mu = 5{,}2068$$
$$mp = 7{,}1595 \qquad mq = 20{,}9170 \qquad m = 28{,}0765$$

On ne peut pas dire que l'écart probable soit dépassé pour $y_{-6+h} = 4$, car au lieu de 6 observations, on aurait pu, raisonnablement, en trouver seulement 5 ; cette statistique suit donc la loi de probabilité simple.

Moyenne. — Elle correspond à $-h$;

$$-h = -0{,}1704 \;;$$

puisque $-1 + h$ correspond à $-0^s{,}5$,

$$-h + h = -0{,}1704 + h$$

correspond à

$$-0^s{,}5 \times 0{,}1704 = -0^s{,}0852.$$

Mode ou valeur la plus probable :

$$\frac{q - p}{2} = 0{,}255 \;;$$

la mode correspond à

$$(0{,}255 - h) + h = (0{,}255 - 0{,}170) + h = 0{,}085 + h \;;$$

puisque $1 + h$ correspond à $0^s{,}5$, la mode correspond à

$$0^s{,}5 \times 0{,}085 = 0^s{,}0425.$$

La figure représentative est tout à fait analogue à la figure 4. Les écarts 0 — C sont imperceptibles.

CHAPITRE XI

APPLICATION A L'ÉTUDE DE STATISTIQUES DIVERSES

90. Les statistiques que nous étudions présentent *un seul* maximum ; nous les appelons *statistiques du premier genre.*

Les nombres calculés suivent soit de très près, soit de près, soit de loin les nombres observés : les statistiques étudiées sont dans l'ordre qu'on vient ainsi de définir.

Pour *toutes*, l'*étendue de la statistique calculée est identique à l'étendue de la statistique observée,* ce qui est un fait des plus remarquables.

Parfois, nous en donnons des exemples à la fin, le calcul est impossible.

Nous admettons ceci :

Quand une statistique du premier genre est réductible par le calcul à la loi de probabilité simple, un seul phénomène est en jeu.

Quand une statistique du premier genre est réductible à la loi de probabilité simple, avec exceptions en quelques points, un phénomène principal est en jeu, mais des phénomènes secondaires sont aussi en jeu. Les points exceptionnels sont indiqués par la comparaison entre les écarts probables des nombres calculés et les écarts entre le calcul et l'observation.

Mais il n'y a réellement exception et indice de phénomènes secondaires que daus les deux cas suivants :

1° des écarts consécutifs $O - C$, qui *dépassent* les écarts probables, croissent puis décroissent :

$$O_1 - C_1 < O_2 - C_2 < \cdots < O_n - C_n$$
$$O_n - C_n > O_{n+1} - C_{n+1} > \cdots > O_{n+k} - C_{n+k},$$

tels sont les écarts disposés autour des nombres 5, 9, 11. 14 de la 1re colonne du Tableau concernant la statistique étudiée au n° 82 : il y a des *anomalies* autour de chacun des points, 5, 9, 11, 14.

Par contre, dans la statistique du n° 89, où les écarts probables ne sont pas dépassés, on doit admettre qu'il n'y a pas d'anomalies,

2° il existe un écart *isolé* (ou plusieurs écarts isolés) très grand, p. e. dont la probabilité est 1 : 100 ; une probabilité de 1 : 100, même de 1 : 50 ou 1 : 25 est si faible qu'on peut la regarder comme l'indice d'une anomalie.

Reportons-nous à la statistique B du n° 92, qu'elle est la probabilité de l'écart 8, qui y figure ?

Il y a 649 cas. Les cas « favorables » (colonne C du tableau) qui sont ici les cas envisagés, sont au nombre de 25 ; les cas défavorables sont donc au nombre de $649 - 25 = 624$; les probabilités correspondantes sont, par suite,

$$\frac{25}{649}, \qquad \frac{624}{649}.$$

La probabilité *moyenne* d'un écart 8 est (n° 47)

$$\Theta\left(8h + \frac{h}{2}\right) = \Theta(7{,}5\,h) ;$$

Ici

$$\frac{1}{h} = \sqrt{2mpq} = \sqrt{2 \times 649 \times \frac{25}{649} \times \frac{649}{624}},$$
$$h = 0{,}144,$$
$$\Theta(8{,}5h) = \Theta(1{,}224) = 0{,}917 ;$$

en examinant la probabilité complémentaire, tout-à-fait acceptable,

$$1 - 0{,}917 = 0{,}093,$$

on se rend compte que la probabilité 0,917 est absolument normale, que l'écart isolé 8 ne décèle pas d'anomalie.

Un écart 20 au lieu de 8, de probabilité moyenne

$$\Theta\left(20h + \frac{h}{2}\right) = \Theta(20{,}5h) = \Theta(1{,}471) = 0{,}962,$$

pourrait indiquer au contraire une anomalie.

Il y a ici élément d'appréciation.

Quand une statistique du premier genre n'est pas réductible à une loi de probabilité simple, des phénomènes du même ordre d'intensité se superposent.

L'analyse mathématique nous donne ainsi des renseignements qu'il n'est pas possible d'obtenir par une autre voie.

De plus, elle *réalise l'ajustage* quand la statistique est réductible, ou à peu près réductible, à une loi de probabilité simple.

La statistique F (n^{os} 84 et 92) concerne la Théorie des nombres.

Les statistiques A (n^{os} 84 et 91), C (n^o 92), Q (n^o 94) se rapportent à l'Astronomie.

Les statistiques B (n^o 92), H et K (n^o 93), S, T et U (n^o 95) appartiennent à la Démographie.

Les statistiques A_2 (n^{os} 88 et 91), D (n^o 92), R (n^o 95) ont l'Histoire Naturelle pour objet.

Les statistiques A_1 (n^{os} 87 et 91), E (n^o 92), G, I, J, L, M, N, O et P (n^o 93), V, X, Y et Z (n^o 95) concernent la Météorologie : leur nombre élevé vient de ce que certaines circonstances nous ont fait envisager souvent des questions de météorologie.

I. — STATISTIQUES DU PREMIER GENRE RÉDUCTIBLES A LA LOI DE PROBABILITÉ SIMPLE

91. *Statistiques entièrement réductibles à la loi de probabilité simple* : nous avons donné des exemples de telles statistiques aux numéros 87, 88, 89, statistiques A_1, A_2, A_3.

La statistique du n^o 87 montre que la température moyenne mensuelle à Monstouris est fonction d'un seul phénomène, que nous savons être l'obliquité de l'équateur sur l'écliptique.

92. *Statistiques où le calcul met en évidence un phénomène principal et un petit nombre de phénomènes secondaires.*

Voici de telles statistiques, B à F.

Ici, l'écart probable des nombres calculés joue un rôle important. Quand les conditions qu'on vient d'indiquer sont remplies, il y a indice d'un phénomène secondaire, qui ajoute son action à celle du phénomène principal.

Il faut distinguer avec soin les écarts probables *absolus*, qui indiquent tous des anomalies, des écarts probables *relatifs*, qui indiquent seulement les principales anomalies (n° 86).

B. — *Statistique concernant l'étude du paupérisme en Angleterre* :

Nombre des bureaux de bienfaisance en fonction du pourcentage de la population à laquelle la Loi sur la bienfaisance est applicable [1]

NOMBRE DE BUREAUX

Pourcentage de la population	Observé	Calculé	O — C		Écarts prob.
			+	—	±
	1 [2]	0	1		
	2 [2]	1	1		
	5 [2]	4	1		
	10 [2]	12		2	
0,75 à 1,25..........	18	26		8	2,9
1,25 » 1,75..........	48	46	2		
1,75 » 2,25..........	72	69	3		
2,25 » 2,75..........	89	88	1		
2,75 » 3,25..........	100	97	3		
3,25 » 3,75..........	90	93		3	
3,75 » 4,25..........	75	78		3	
4,25 » 4,75..........	60	57	3		
4,75 » 5,25..........	40	37	3		
5,25 » 5,75..........	21	22		1	
5,75 » 6,25..........	11	11	0	0	
6,25 » 6,75..........	5	5	0	0	
6,75 » 7,25..........	1	2		1	
7,25 » 7,75..........	1	1	0	0	
	649	649	18		

(1) G. UDNY YULE, *An Introduction to the Theory of Statistics*, 6e édition, Griffin à Londres, 1922, p. 93.

(2) Ces nombres ont été obtenus par extrapolation graphique : cela était nécessaire, car le calcul suppose que la statistique est *complète*. On peut utiliser les statistiques *incomplètes* (n° 69) mais cela complique beaucoup les calculs (Cf, n° 69, NOTE).

Ici

$h = 0{,}38829$ $\quad$ $A = 362239$ $\quad$ $B = 2528795$

$p = 0{,}71281$ $\quad$ $q = 0{,}28719$ $\quad$ $\lambda = 6{,}73296$ $\quad$ $\mu = 7{,}12124$

$mp = 24{,}40784$ $\quad$ $mq = 9{,}83396$ $\quad$ $m = 34{,}24181$

Les nombres observés suivent la loi de probabilité simple, l'écart O — C = 8 étant certainement possible, bien que plus grand que l'écart probable 2,9 : nulle part ailleurs, l'écart probable n'est dépassé.

La figure est encore tout à fait analogue à la figure 4 ; les données sont sur la courbe correspondant aux nombres calculés C. Les écarts O — C sont imperceptibles.

C. — *Distribution des étoiles variables* (fig. 8).

Les nombres d'étoiles variables, classées d'après les durées de leurs périodes sont indiqués ci-après [1] colonne O :

Périodes en jours	Nombres d'étoiles O	C	O — C +	O — C —	Ecarts prob. ∓
50 à 100	$y_{-5+h} = 5$	4	1		
100 » 150	15	13	2		1,9
150 » 200	24	32		8	3,2
200 » 250	71	60	11		4,3
250 » 300	78	83		5	5,0
300 » 350	$y_h = 85$	87		2	
350 » 400	64	65		1	
400 » 450	37	34	3		3,2
450 » 500	10	11		1	
500 » 550	$y_{4+h} = 2$ [2]	2			
	391	391	17	17	

Ici,

$h = 0{,}5117$ $\quad$ $A = 76{,}567$ $\quad$ $B = 225534$

$p = 0{,}3735$ $\quad$ $q = 0{,}6$[illegible]75 $\quad$ $\lambda = 3{,}1724$ $\quad$ $\mu = 2{,}6607$

$mp = 4{,}7519$ $\quad$ $mq = 8{,}0048$ $\quad$ $m = 12{,}5767.$

Il n'y a pas d'anomalies,

(1) BIGOURDAN, *Etoiles variables.* Ann. Beau des Longitudes 1909, p. 24. Comparer avec *Ibid.*, p. 55.

(2) Nombre obtenu par extrapolation graphique.

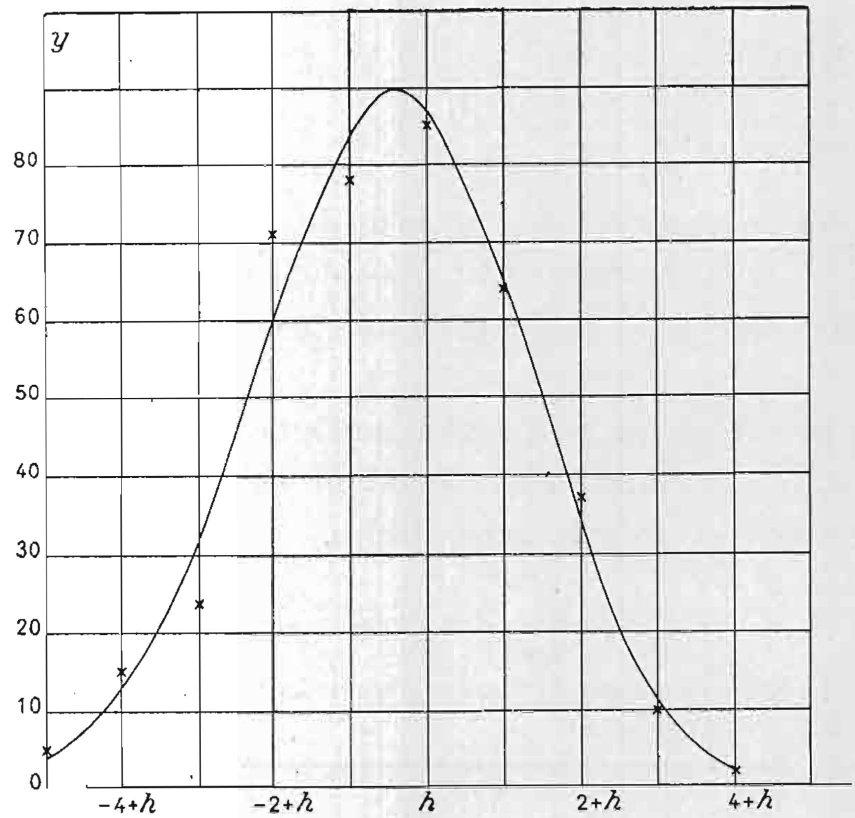


Fig. 8. — Distribution des étoiles variables. La courbe est obtenue par le calcul.

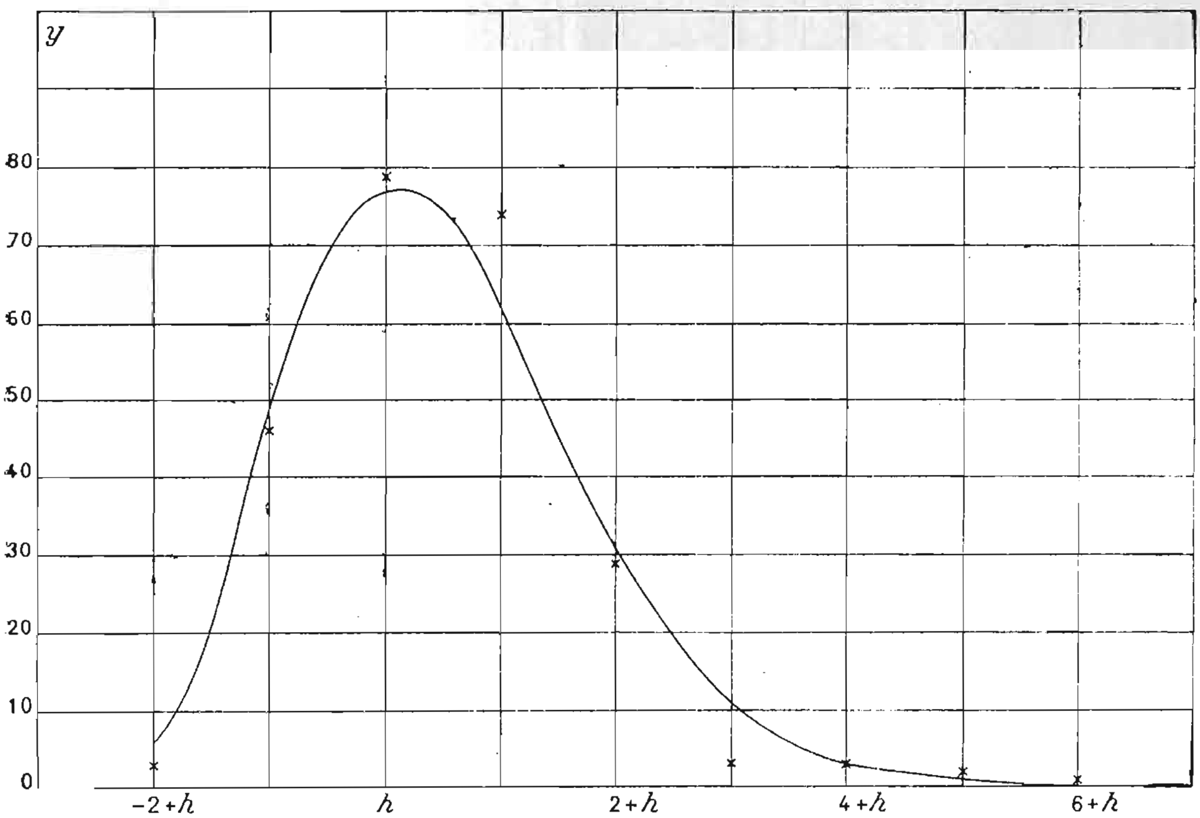


Fig. 9. — Etude d'un lot d'épinoches. La courbe est obtenue par le calcul. Ce lot n'est pas homogène, car les données correspondant à $1 + h$ et $3 + h$ sont trop en dehors des limites de l'écart probable.

D. — *Epinoches classées d'après les nombres de leurs plaques osseuses latérales* (1) (fig. 9).

Nombres de plaques	Nombres de sujets observés O	C	O — C +	O — C —	Écarts prob. ±
3	$y_{-2+h} =$ 3	6		3	
4	46	49		3	4,0
5	$y_h =$ 79	77	2		
6	$y_{1+h} =$ 74	62	12		4,5
7	29	31		2	
8	3	11		8	1,1
9	2	3	0	0	
10	2	1	1		
11	1	0	1		
	240	240	16	16	

Ici,

$$h = -0{,}488 \qquad A = 14105 \qquad B = 19266$$
$$p = 0{,}874 \qquad q = 0{,}126 \qquad \lambda = 1{,}080 \qquad \mu = 1{,}567$$
$$mp = 1{,}723 \qquad mq = 11{,}940 \qquad m = 13{,}663$$

Les anomalies pour 6 plaques et pour 8 plaques indiquent que ces épinoches appartenaient en grande majorité à une espèce dominante, ou race, ou sexe dominant.

Une ou deux autres espèces, ou races, ou sexe, ayant un peu plus de plaques que la classe dominante, étaient représentée par un petit nombre d'individus.

La courbe représentative est une des plus dissymétriques que nous connaissions.

Pour 6 plaques, la probabilité de l'écart 12 défini par (nos 44, 46)

$$\frac{1}{k} = \sqrt{2 \times 240 \times \frac{62}{240} \times \frac{178}{240}}$$

$$\log k = 1{,}01818$$

(1) L. Bertin, *Recherches bionomiques, biométriques et systématiques sur les Epinoches.* Thèse présentée à la Faculté des Sciences de Paris, 1925, in-4° 204 pages et 71 figures. Blondel la Rougery à Paris.

a pour valeur

$$\Theta\left(12k + \frac{k}{2}\right) = \Theta(12{,}5k) = \Theta(1{,}303)$$
$$= 0{,}935\ ;$$

la probabilité complémentaire **1 — 0,935 = 0,065 =** ou environ $\frac{1}{17}$ est assez faible pour qu'on puisse voir ici l'indice d'une anomalie, de même pour 8 plaques (n° 90).

La Mode

$$x + h = \frac{q - p}{2} = -0{,}374 = (-0{,}374 - h) + h = 0{,}114 + h$$

correspond à

$$6 \text{ plaques} + 0{,}114 = 6{,}114 \text{ plaques},$$

soit un plus de 6 plaques.

E. — *Statistique des températures observées à Paris, au Parc Saint-Maur, de* 1890 *à* 1899 [1] (fig. 10)

en tout 87 648 observations horaires réduites par M. Baldit au total 999,2 (pour 1 000). Les nombres O indiquent les nombres d'heures sur 1 000 (999,2) où les températures ont eu les valeurs x.

Températures en degrés centigrades	Observé O	Calculé C	O — C +	O — C —	Écarts prob. ±
— 15	$y_{-25+h} =$ 0,1	0,2			
14	0,3	0,3			
13	0,5	0,6			
12	0,6	0,8			
11	1,0	1,1			
10	1,5	1,6			
9	2,1	2,2			
8	2,8	3,0			
7	3,7	4,1			
6	5,3	5,4			
5	7,3	7,1			
4	7,7	9,1			
3	10,7	11,5			
2	13,8	14,3			
— 1	19,6	17,5			
0	28,0	21,1	6,9		2,6
+ 1	29,3	25,0			
2	30,8	29,1			
3	33,5	33,2			
4	33,6	37,4		3,8	3,1
5	39,5	41,3			
6	44.0	44,9			
7	47,0	48,0			
8	47,8	50,4			
9	5[illegible],8	52,0			
10	$y_h =$ 52,9	52,8			
11	48,7	52,6		3,9	4,2
12	48,2	5[illegible],5			
13	47,0	49,6			
14	45,2	46,9			
15	48,5	43,5	5,0		3,9
16	42,2	39,7			
17	36,5	35,6			
18	3[illegible],7	31,3			
19	27,2	27,0			
20	24,1	22,9			
21	19,8	19,1			
22	16,5	15,6			
23	12,5	12,5			
24	10,2	9,9			
25	7,9	7,6			
26	5,6	5,8			
27	4,4	4,3			
28	2,8	3,1			
29	2,2	2,2			
30	1,8	1,6			
31	0,9	1,1			
32	0,5	0,7			
33	0,3	0,5			
34	0,2	0,3			
35	$y_{25+h} =$ 0,1	0,2			
	999,2	999,2			

[1] A. Baldit, *Sur la fréquence des températures à Saint-Maur et à Zi-ka-weï*, 10 années, 1890-1899, *Ann. Soc. Météor. de France*, 1906, p. 195, Paris, Gauthier-Villars.

Ici :

$$h = -0{,}1858 \qquad A = 2\,875\,450 \qquad B = 163\,650\,820$$
$$p = 0{,}3743 \qquad q = 0{,}6257 \qquad \lambda = 56{,}8197 \qquad \mu = 57{,}0055$$
$$mp = 90{,}9147 \qquad mq = 152{,}0058 \qquad m = 242{,}9205$$

Un phénomène principal est en jeu : la variation de l'inclinaison de l'équateur sur l'écliptique.

Il y a anomalie certaine pour 0°, ce qui s'explique par l'action régulatrice de la glace fondant sur la température.

Il y a peut-être anomalie pour 15° et pour 4°.

Si l'on considère que l'on devrait multiplier les nombres 0 et C par

$$\frac{87\,648}{999{,}2} = 87{,}72$$

et les écarts probables par environ

$$\sqrt{\frac{87\,648}{999{,}2}} = 9{,}36$$

pour passer de cette statistique à la statistique *vraie*, et pour passer des écarts probables relatifs aux écarts probables absolus (n° 86), les anomalies pour 15°, 4° deviennent certaines ; il s'en ajoute même d'autres, en particulier une anomalie pour 11°.

Mode.

$$\frac{q-p}{2} = 0{,}1257 = 0{,}1257 - h + h = 0{,}1257 + 0{,}1858 + h$$
$$= 0{,}3115 + h ;$$

la mode est donc

$$10^\circ + 0^\circ{,}3115 = 10^\circ{,}3115 :$$

c'est la température la plus probable.

Moyenne. — Elle correspond à

$$-h = 0{,}1858 ;$$

soit

$$\begin{array}{rl} 10^\circ & y_{0+h} = y_{0-0,1858} \\ x^\circ & y_{0,1858+h} \\ 11^\circ & y_{1+h} = y_{1-0,1858} ; \end{array}$$

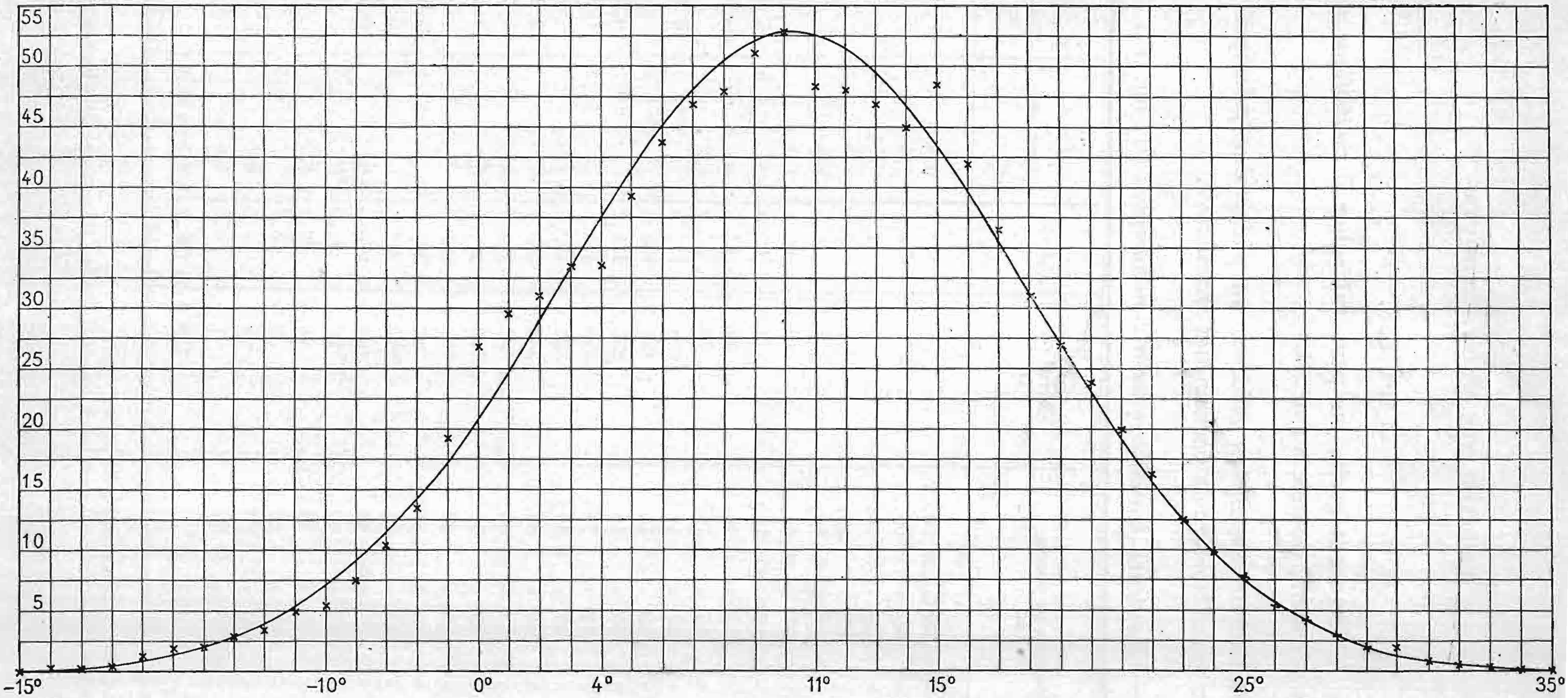


Fig. 10. — Températures au parc Saint-Maur : 87.648 observations horaires, réduites au total 1000 par M. Baldit, faites de 1890 à 1899. La courbe est le résultat du calcul. Les + sont les nombres de la statistique. Il y a des anomalies, indices de phénomènes secondaires, pour les températures 0°, 4°, 11°, 15°.

on a

$$x = 10^{o},1858.$$

F. — A cette catégorie, on peut rattacher *la statistique des nombres premiers*, donnée au n° 84.

93. *Statistique mettant en évidence un phénomène principal et un assez grand nombre de phénomènes secondaires.*

G. — *Hauteurs barométriques à Southampton* [1] (fig. 11)

Hauteurs en pouces L	Observé [2] O	Calculé C	O — C +	O — C —	Écarts prob. ±
28,45 à 28,55 ou 28,50....	1	0,2	0,8		
55 » 65 » 60....	2	0,5	1,5		
65 » 75 » 70....	2	1,3	0,7		
75 » 85 » 80....	4	2,8	1,2		
85 » 95 » 90....	8,5	6,0	2,5		
95 » 29,05 » 29,00....	13,5	12,1	1,4		
29,05 » 15 » 10....	21,5	23,3		1,8	
15 » 25 » 20....	37	42,6		5,6	
25 » 35 » 30....	79	73,9	5,1		
35 » 45 » 40....	108	121,0		13,0	6,8
45 » 55 » 50....	181,5	186,6		5,1	
55 » 65 » 60....	254,5	270,2		15,7	10,5
65 » 75 » 70....	348,5	365,7		17,2	11,9
75 » 85 » 80....	463,5	460,5	3,0		
85 » 95 » 90....	548,5	536,8	11,7		14,2
95 » 3 ,05 » 30,00....	$y_{-1+h} =$ 602,5	575,9	26,6		14,7
30,05 » 15 » 10....	$y_h =$ 619,5	564,4	55,1		14,5
15 » 25 » 20....	$y_{1+h} =$ 500	501,3		1,3	
25 » 35 » 30....	382	399,4		17,4	12,4
35 » 45 » 40....	237,5	281,8		44,3	10,6
45 » 55 » 50....	189,5	173,3	16,2		8,2
55 » 65 » 60....	88,5	91,0		2,5	
65 » 75 » 70....	43,5	39,6	3,9		
75 » 85 » 80....	7	13,7		6,7	
85 » 95 » 90....	4	3,5	0,5		
30,95 » 31,05 » 31,00....	1	0,6	0,4		
	4 748,0	4 748,0	130,6	130,6	

(1) YULE, *Loc. cit.* Statistique B, p. 97.
(2) Nombres de jours où la hauteur barométrique observée a été comprise dans les limites L.

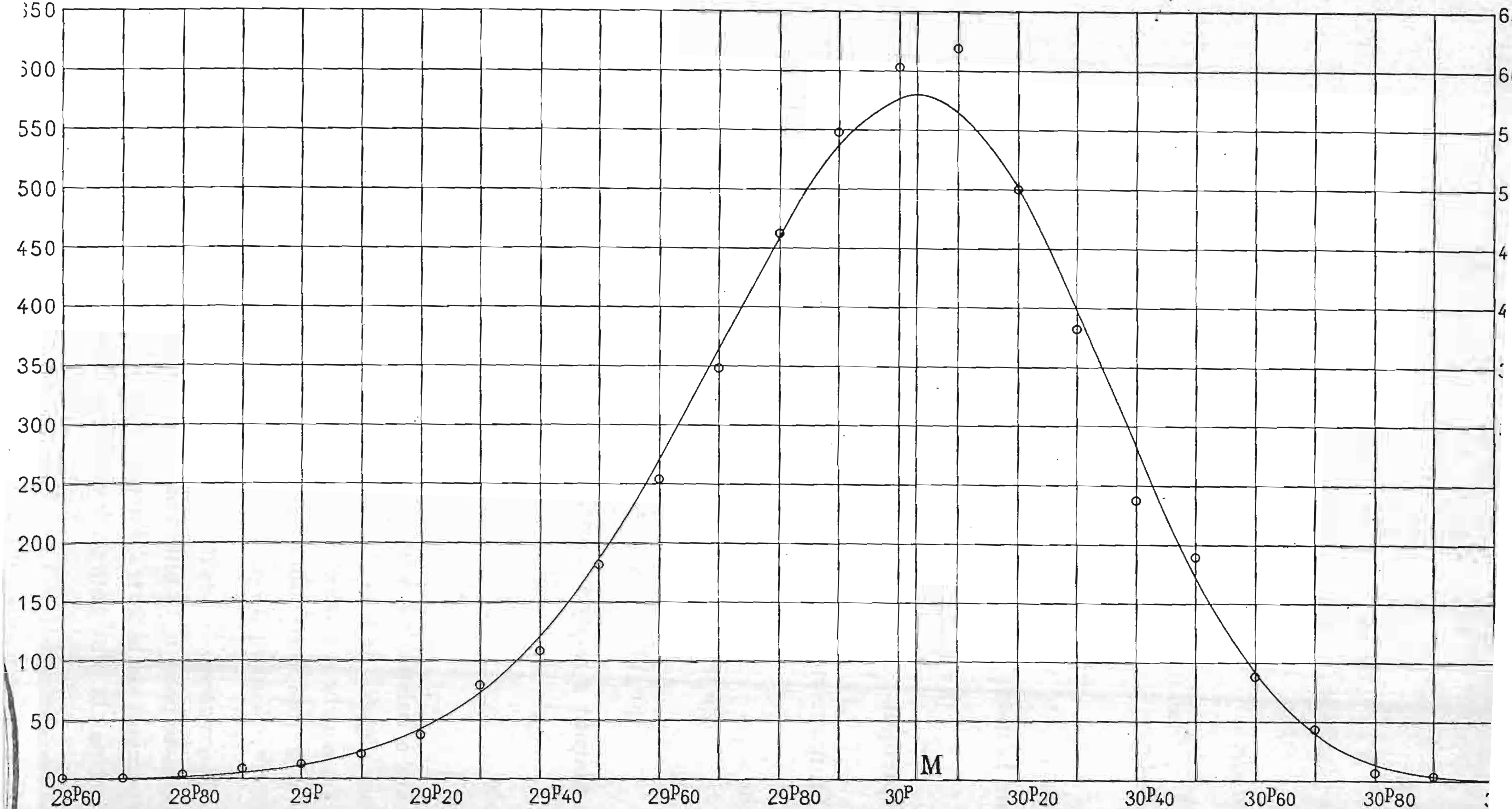


Fig. 11. — Hauteurs barométriques à Southampton pour 4.748 journées. La courbe est le résultat du calcul. Les observations sont indiquées par des o. Il y a une forte anomalie pour 30 p. 05 et une autre presque aussi forte pour 30 p. 40 ; ce sont les indices de phénomènes secondaires.

Ici,

$$h = 1{,}18576 \qquad A = 22\,745\,738 \qquad B = 235\,197\,600$$
$$p = 0{,}02835 \qquad q = 0{,}97165 \qquad \lambda = 10{,}83232 \qquad \mu = 9{,}64656$$
$$mp = 11{,}1137 \qquad mq = 380{,}9070 \qquad m = 392{,}0207$$

Mode.

$$\frac{q-p}{2} = 0{,}47165 = (0{,}47165 - h) + h = -0{,}71411 + h\,;$$

$$\begin{array}{rcl} -1 + h & \text{correspond à} & 30^{p}{,}00\,; \\ -0{,}71411 + h & \text{»} & 30^{p}{,}00 + x\,; \\ h & \text{»} & 30^{p}{,}10\,; \end{array}$$

donc la mode est

$$30^{p}{,}10 - \frac{0{,}71411}{10} = 30^{p}{,}028589 = 762^{mm}{,}71,$$

en prenant

$$1 \text{ pouce} = 0^{mm}{,}254.$$

La moyenne est

$$-1{,}18576\,;$$

donc

$$\begin{array}{lcl} 29^{p}{,}90 \text{ correspond à} & -2 + h = -2 + 1{,}18576 \\ \quad x \qquad\quad \text{»} & -1{,}18576 + 1{,}18576 \\ 30^{p}{,}00 \qquad \text{»} & -1 + 1{,}18576 \end{array}$$

en plaçant x relativement à

$$-1 \qquad -1{,}18576 \qquad -2,$$

on a

$$x = 29^{p}{,}918576 = 759^{mm}{,}93.$$

On a calculé les seuls écarts probables correspondant aux valeurs élevées de 0 — C.

La courbe des observations oscille autour de la courbe calculée. Il y a une forte anomalie pour $30^{p}{,}00$ à $30^{p}{,}20$, soit 762^{mm} à 767^{mm}. (Voir n° 101).

Observation. — Les formules de calcul ne supposent pas que l'on choisisse pour y_h la plus grande des valeurs observées. Il convient cependant de le faire. Il arrive parfois, comme ici, que y_h calculé n'est pas la plus grande des valeurs calculées.

H. — *Habitants du Pays de Galles classés d'après leurs poids* (1)

Poids en livres anglaises	Nombres d'individus Observé (2) O	C	O — C +	O — C —	Écarts prob. ±
90	$y_{-6+h} =$ 1	0	1		
100	2	0	2		
110	10	6	4		
120	23	37		14	3,5
130	68	90		22	5,4
140	153	139	14		6,7
150	$y_h =$ 178	154	24		6,8
160	134	131	3		
170	102	90	12		5,4
180	34	51		17	4,1
190	14	24		10	1,8
200	7	10		3	
210	8	4	4		
220	1	1	0	0	
230	$y_{8+h} =$ 2	0	2		
	737	737	66	66	

Ici

$h = 0{,}2551 \qquad A = 279\,742 \qquad B = 988\,096$

$p = 0{,}2124 \qquad q = 0{,}8876 \qquad \lambda = 3{,}6462 \qquad \mu = 3{,}3911$

$mp = 4{,}0756 \qquad mq = 32{,}1840 \qquad m = 36{,}2593$

L'écart probable est souvent dépassé. Puisque, dans l'ensemble, la statistique suit une loi de probabilité simple, une race prédomine, mais elle est accompagnée de races secondaires dont les représentants sont en nombres élevés.

(1) G. Udny Yule, *loc. cit.* Statistique B, p. 95.

(2) Le nombre 1 a été obtenu graphiquement (n° 69, Note).

I. — *Températures moyennes mensuelles à Paris ramenées au niveau de la mer* [1]

Ces températures ont été toutes diminuées de 2°,5 pour que les données extrêmes soient voisines de zéro; il n'y a aucune raison de prendre 0° pour ordonnée nulle.

	Temp.	O	C	O — C +	O — C −	Écarts prob. ±
Janvier	2°,5	0,0	0,4		0,4	
Février	3°,9	1,4	1,2	0,2		
Mars	6°,2	3,7	3,1	0,6		
Avril	10°,3	7,8	6,4	1,4		1,1
Mai	13°,4	10,9	11,0		0,1	
Juin	16°,9	$y_{-1+h} = 14{,}4$	15,4		1,0	
Juillet	18°,6	$y_h = 16{,}1$	17,7		1,6	2,0
Août	18°,0	$y_{1+h} = 15{,}5$	16,3		0,8	
Septembre	15°,0	12,5	12,0	0,5		
Octobre	10°,3	7,8	6,8	1,0		1,2
Novembre	6°,0	3,5	2,9	0,6		
Décembre	2°,9	0,4	0,8		0,4	
		94,0	94,0	4,7	4,7	

Ici

$$h = 0{,}016 \qquad A = 6283{,}76 \qquad B = 27684{,}28$$
$$p = 0{,}347 \qquad q = 0{,}653 \qquad \lambda = 4{,}413 \qquad \mu = 4{,}397$$
$$mp = 6{,}749 \qquad mq = 12{,}700 \qquad m = 19{,}449$$

Mode :

$$\frac{q-p}{2} = 0{,}153 = (0{,}153 - h) + h = 0{,}137 + h.$$

h correspond au 15 juillet
0,137 » 15 + x juillet
$h + 1$ » 15 août ;

donc

$$\frac{x}{30} = \frac{0{,}153}{1}, \qquad x = 4{,}59 :$$

[1] *Annuaire du Bureau des Longitudes*, 1929, p. 193.

le maximum correspond à 15 juillet + 5 jours = 20 juillet ; l'observation donne

$$23 \text{ jt} + 2^{j}5\ ;$$

le calcul conduit à une erreur de 6 jours sur 365.

Les écarts probables ne sont pas dépassés ou de bien peu. Mais ce sont des écarts probables *relatifs*. Si on exprimait les températures en 10es de degrés, on trouverait par exemple

	O	C	Ec. prob.
Jt	161	177	$\pm 2{,}5\sqrt{10} = \pm 8{,}0$

Il y a donc anomalie pour avril, octobre, juillet, cette dernière s'étendant sur juin et août.

Le phénomène principal en jeu est la variation de l'inclinaison de l'équateur sur l'ecliptique.

J. — *Températures mensuelles à Bordeaux* (1)

	Degrés centig. (2) O	C	O — C +	O — C —	Écarts probl. ±
Janvier	$y_{-6+h} =$ 0	0,4		0,4	
Février	1,2	1,1	0,1		
Mars	3,8	2,8	1,0		0,6
Avril	6,9	5,8	1,1		1,1
Mai	10,1	10,0	0,1		
Juin	13,2	14,4		1,2	1,8
Juillet	$y_h =$ 15,1	16,9		1,8	2,0
Août	15,5	16,2		0,7	
Septembre	13,0	12,4	0,6		
Octobre	8,4	7,2	1,2		1,2
Novembre	3,3	3,1	0,2		
Décembre	$y_{5+h} =$ 0,7	0,9		0,2	
	91,2	91,2	4,3	4,3	

(1) 1875-1928, *Communciation* de M. MÉMERY, directeur de l'observatoire de Talence, Gironde

(2) Les températures sont

5°,3 ;	6°,5 ;	9°,1 ;	12e,2 ;	15°,4 ;	18°,5 ;	20°,4 ;	20°,8 ;	18°,3 ;	13°,7 ;	8°,6 ;	6°,0
j	*f*	*m*	*a*	*m*	*j*	*j*	*a*	*s*	*o*	*n*	*d*

elles ont toutes été diminuées de 5°3, pour rendre le calcul possible.

Ici,

$h = -0,089$ $A = 6015,45$ $B = 27011,10$

$h = 0,289$ $q = 0,711$ $\lambda = 4,447$ $\mu = 4,472$

$mp = 6,291$ $mq = 15,478$ $m = 21,769$

Mêmes observations que pour la statistique I.

Mode : $\frac{q - p}{2} = 0,211$; la température moyenne maxima *calculée* se rapporte à

$$15 \text{ juillet} + 31^j \times 0,211 = 15 \text{ juillet} + 6^j541 = 22 \text{ juillet}$$

K. — *Hauteurs de taille aux États-Unis* (1) (fig. 12)

Observé O		Calculé C	O — C +	O — C —	Écarts prob. ±
$y_{-13+h} =$	4	0	4		
	1	0	1		
	3	0	3		
	7	0	7		
	6	2	4		
	10	13		3	
	15	58		43	4,6
	50	195		145	7,9
	526	512	14		14,6
	1 237	1 092	145		21,3
	947	1 929	18		28,0
	3 019	2 868	151		33,6
	3 475	3 633		158	37,2
$y_h =$	4 050	3 958	92		39,1
	3 631	3 731		100	37,6
	3 133	3 058	75		34,6
	2 075	2 189		114	28,7
	1 485	1 368	117		23,8
	680	748		68	17,7
	343	357		14	12,2
	181	149	32		7,7
	42	54		12	
	9	17		8	
	6	5	1		
$y_{11+h} =$	2	1	1		
Nombres d'individus . .	25 937	25 937	565	565	

(1) D'après Quételet. Ch. Jordan, *Statistique mathématique*, p. 209 ; Gauthier-Villars Paris, 1927.

Ici,

$h = -0{,}217758 \quad A = 575\,539\,306 \quad B = 3\,850\,966\,524$

$p = 0{,}63101 \quad q = 0{,}36899 \quad \lambda = 6{,}57693 \quad \mu = 6{,}79419$

$mp = 18{,}19703 \quad mq = 10{,}64045 \quad m = 28{,}83749$

Les écarts probables sont largement dépassés dans un sens ou dans l'autre, mais les nombres calculés suivent d'assez près, relativement parlant, les nombres observés. Les discordances peuvent être mises sur le compte d'un mélange de races peu différentes les unes des autres, comme il était en effet lorsque cette statistique a été établie.

Divisons les nombres de la statistique par 10 *; elle devient*

O	C	O — C		Écarts prob.
		+	—	±
0,4	0	0,4		
0,1	0	0,1		
0,3	0	0,3		
0,7	0	0,7		
0,6	0,2	0,4		
1,0	1,3		0,3	
1,5	5,8		4,3	1,1
5,0	19,5		14,5	2,1 *
52,6	51,2	1,4		4,3
123,7	109,2	14,5		6,4 *
194,7	192,9	1,8		8,5
301,9	286,8	15,1		10,3 *
347,5	363,3		15,8	11,4 *
405,0	395,8	9,2		12,4
363,1	373,1		10,0	11,6
313,3	305,8	7,5		10,6
207,5	218,9		11,4	9,1 *
148,5	136,8	11,7		7,2 *
68,0	74,8		6,8	5,2 *
34,3	35,7		1,4	3,5
18,1	14,9	3,2		2,1 *
4,2	5,4		1,3	
0,9	1,7		0,8	
0,6	0,5	0,1		
0,2	0,1	0,1		
2 593,7	2 593,7	56,5	56,5	

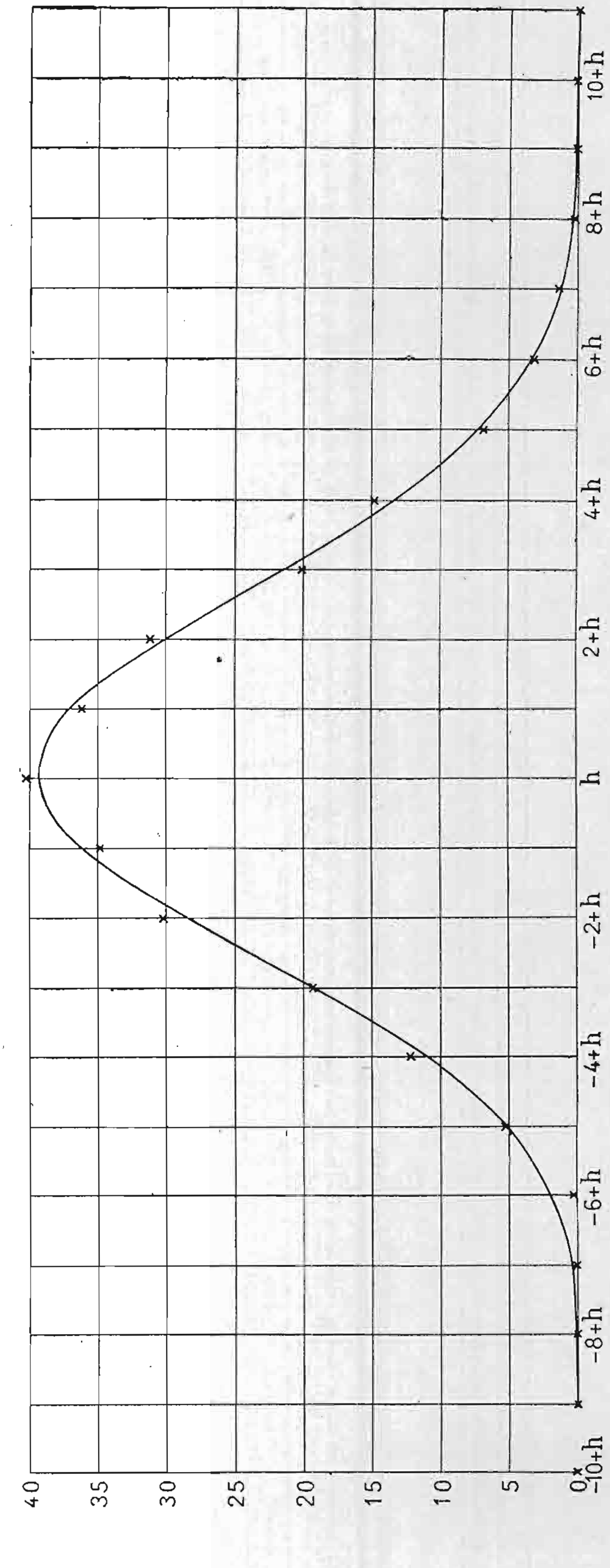


Fig. 12. — Statistique démographique. Les données (×) sont à peu près confondues avec la courbe calculée, parce que l'échelle du dessin est trop petite ; en fait un grand nombre de données sont en dehors des limites de l'écart probable. Les nombres de la statistique ont été divisée par 100.

Sous cette forme, il apparaît mieux que la statistique suit la même loi que les nombres calculés C, et les écarts principaux O — C désignés par un * sont plus visibles.

L. M. N. — *Températures à Greenharbour, Spitzberg* (1)

Températures en degrés centigrades	I Juillet Année 1912		II Juillet Année 1914		III Juin Année 1916	
	O	C	O	C	O	C
9,0 à 8,1.....						
8,0 » 7,1.....			2	0	1	5
7,0 » 6,1.....			6	2	22	15
6,0 » 5,1.....			1	5	46	35
5,0 » 4,1.....	4	1	12	14	59	68
4,0 » 3,1.....	6	4	20	32	102	106
3,0 » 2,1.....	14	18	39	56	135	134
2,0 » 1,1.....	34	53	114	93	124	137
1,0 » 0,1.....	140	113	127	127	121	109
0,0 » — 0,9.....	171	173	168	142	68	68
— 1,0 » — 1,9.....	178	186	101	127	31	31
— 2,0 » — 2,9.....	123	132	94	88	10	10
— 3,0 » — 3,9.....	71	55	51	44	1	2
— 4,0 » — 4,9.....	3	9	5	14		
— 5,0 » — 5,9.....						
— 6,0 » — 6,9.....						
— 7,0 » — 7,9.....						
	744	744	744	744	720	720

On a ici

I. $h = 0{,}52285$ $\quad A = 231717$ $\quad B = 532150$

$p = 0{,}2915$ $\quad q = 0{,}7085$

$\lambda = 2{,}5247$ $\quad \mu = 2{,}0019$

$mp = 3{,}3487$ $\quad mq = 8{,}1350$ $\quad m = 11{,}4837$

II. $h = 0{,}26613$ $\quad A = 325818$ $\quad B = 1368688$

$p = 0{,}2177$ $\quad q = 0{,}7823$

$\lambda = 4{,}3261$ $\quad \mu = 4{,}0600$

$mp = 5{,}4559$ $\quad mq = 19{,}6056$ $\quad m = 25{,}0615$

(1) *Loc. cit.*, à Statistiques à V — Z, p. 157.

III. $h = -0{,}41528$ $\quad A = 332605$ $\quad B = 1411402$

$p = 0{,}34742$ $\quad q = 0{,}65258$

$\lambda = 4{,}0281$ $\quad \mu = 4{,}4434.$

$$mp = 6{,}6114$$
$$mq = 12{,}4197$$
$$m = 19{,}0311$$

Ici, la comparaison des 0 et des C ne nécessite pas le calcul des écarts probables.

Pour I :

obs. $34 + 140 = 174$ calc. $53 + 113 = 166$

obs. $123 + 71 = 194$ calc. $132 + 55 = 187$,

il y a anomalies pour **2°,0 à 0°,1 et — 2°,0 à — 3°,9.**

Pour II :

obs. $39 + 114 = 153$ calc. $56 + 93 = 149$

obs. $168 + 101 = 269$ calc. $142 + 127 = 269$

il y a anomalies pour **3°,0 à 1°,1 et 0°,0 à — 1°,9.**

Pour III :

obs. $46 + 59 = 105$ calc. $35 + 68 = 103$

obs. $124 + 121 = 245$ calc. $137 + 109 = 246$

il y a anomalies pour **6°0 à 4°1 et 2°,0 à 0°1.**

O. — *Orages à grêle en France* (1)

	O	C	O — C +	O — C —
Janvier	2	$y_{-6+h} =$ 0	2	
Février	4	1	3	
Mars	27	80		53
Avril	390	344	46	
Mai	1 088	877	211	
Juin	1 283	1 424		141
Juillet	1 338	$y_h =$ 1 532		194
Août	1 109	1 093	16	
Septembre	622	500	122	
Octobre	131	133		2
Novembre	6	16		10
Décembre	0	$y_{5+h} =$ 0	0	0
	6 000	6 900	400	400

(1) 1868-1928. Observaitons communiquées par M. PIRON, chef du service de la branche *Grêle* à la compagnie d'assurances la Nationale.

Ici,

$h = -0{,}3333 \qquad A = 166479 \times 10^2 \qquad B = 37030 \times 10^3$

$p = 0{,}488 \qquad q = 0{,}512 \qquad \lambda = 2{,}3793 \qquad \mu = 2{,}0459$

$mp = 4{,}329 \qquad mq = 4{,}542 \qquad m = 8{,}871.$

Le calcul ne suit l'observation que de loin. Le calcul des écarts probables est inutile.

Les fortes anomalies de mai et septembre ont de l'intérêt pour les compagnies d'assurances, *qui ont reconnu la première, mais non pas la seconde.*

P. — *Proportion des orages en France* (1)

	O	C	O — C +	O — C −	Écarts prob. ±
Janvier	1,04	0,66	0,38		
Février	1,42	1,75		0,33	
Mars	3,47	4,00		0,53	
Avril	7,37	7,84		0,47	
Mai	16,00	12,97	3,03		1,8
Juin	19,02	17,80	1,22		
Juillet	18,50	19,84		1,34	
Août	15,60	17,37		1,77	
Septembre	9,45	11,34		1,89	
Octobre	5,07	5,06	0,01		
Novembre	1,89	1,28	0,61		
Décembre	1,17	0,09	0,08		
	100,00	100,00	5,33	5,33	

Ici

$h = -0{,}6277 \qquad A = 6052{,}5 \qquad B = 25931$

$p = 0{,}159 \qquad q = 0{,}841 \qquad \lambda = 3{,}935 \qquad \mu = 4{,}563$

$mp = 4{,}798 \qquad mq = 25{,}377 \qquad m = 30{,}175$

Le calcul des écarts probables est inutile.

Forte anomalie en mai, où la proportion des orages dépasse beaucoup le chiffre indiqué par le calcul. Cette anomalie a, on l'a dit, de l'intérêt pour les compagnies d'assurances-grêle.

(1) 1881-1913, proportion calculée par FRON, AUGER, DONGIER, d'après CH. MAURAIN, *L'Onde électrique*, avril 1929, Chiron à Paris.

Le phénomène principal est la variation de l'inclinaison de l'équateur sur l'écliptique.

Comparaison des orages et des orages à grêle en France

Orages à grêle ramenés du total 6 000 au total 100, pour permettre la comparaison qui suivra

	O	C	O — C +	O — C —
Janvier	0,03	0	0,03	
Février	0,07	0,02	0,05	
Mars	0,45	1,33		0,88
Avril	6,50	5,73	0,77	
Mai	18,13	14,62	3,51	
Juin	21,38	23,73		2,35
Juillet	22,60	25,53		2,93
Août	18,48	18,22	0,26	
Septembre	10,37	8,33	2,04	
Octobre	2,18	2,21		0,03
Novembre	0,10	0,27		0,17
Décembre	0,00	0,00	0	0
	99,99	99,99		

Comparaison de ces nombres avec ceux de la statistique P :

	O — C			
	+		—	
	Stat. O (ci-dessus)	Stat. P	Stat. O (ci-dessus)	Stat. P
Janvier	0,03	0,38		
Février	0,05		0,88	0,33
Mars				0,53
Avril	0,77			0,47
Mai	3,51	3,03		
Juin		1,22	2,35	
Juillet			2,93	1,34
Août	0,26			1,77
Septembre	2,04			1,89
Octobre		0,01	0,03	
Novembre		0,61	0,17	
Décembre	0	0,08	0	

Ce Tableau montre qu'en mai un phénomène d'une assez grande importance accroît le nombre des orages et le nombre des orages à grêle.

II. — STATISTIQUES NON RÉDUCTIBLES A LA LOI DE PROBABILITÉ SIMPLE

94. *Statistiques réductibles en apparence seulement à la loi de probabilité simple.*

Ces statistiques ont l'allure des statistiques que nous venons d'étudier. Mais quand on leur applique le calcul, on trouve des nombres C qui diffèrent trop des nombres O pour qu'on puisse les leur substituer.

On ne peut établir dans ce cas des formules relevant de la loi de probabilité simple.

Il est à présumer que la statistique est alors la résultante de statistiques composantes, bien que le graphique ne l'indique pas.

Q. — *Étoile variable U Céphée* (1)

Heures	Grandeur G	9,25 — G O	C	O — C +	O — C —
0,30	9,25	$y_{-7+h} =$ 0,00			
1	9,20	0,05	0,17		0,12
1,30	8,80	0,45	0,32	0,13	
2	8,40	0,85	0,64	0,23	
2,30	8,00	1,25	1,08	0,17	
3	7,65	1,60	1,56	0,04	
3,30	7,50	1,75	1,93		0,18
4	7,48	$y_h =$ 1,77	2,05		0,28
4,30	7,60	1,65	1,88		0,23
5	7,75	1,50	1,49	0,01	
5,30	8,25	1,00	1,02		0,02
6	8,45	0,80	0,60	0,20	
6,30	8,80	0,45	0,30	0,15	
7	9,20	$y_{6+h} =$ 0,05	0,13		0,08
		13,17	13,17	0,92	0,92

(1) BIGOURDAN, *Ann. du Bureau du Long.* 1909 ; Etoiles variables, p. 35.
(2) *Ibid.*, p. 38.

Ici

$$h = 0{,}0380 \qquad A = 159{,}5725 \qquad B = 1043{,}385$$
$$p = 0{,}5425 \qquad q = 0{,}4575 \qquad \lambda = 6{,}5578 \qquad \mu = 6{,}5198$$
$$mp = 14{,}304 \qquad mq = 12{,}039 \qquad m = 26{,}343.$$

Les nombres 0 — C montrent que la statistique n'est pas réductible à la loi de probabilité simple : ils sont trop grands vis-à-vis des C.

Cela n'a rien de surprenant car les nombres 0 suivent une loi mathématique bien définie. En effet, U Céphée est considérée comme formée de deux étoiles, l'une brillante, l'autre obscure et plus petite, dont les diamètres sont dans le rapport 5 à 3 ; le compagnon aurait une orbite relative ; en supposant celle-ci circulaire, le rayon de l'orbite serait double de celui de l'étoile principale.

On remarquera l'analogie des nombres 0 — C avec les nombres 0 — C de la statistique E.

95. *Statistiques formellement irréductibles à la loi de probabilité simple.*

On reconnaît qu'il en est ainsi quand p et q ne sont pas compris, chacun, entre 0 et 1.

Ici, des phénomènes de même ordre d'intensité sont superposés, bien que le graphique ne présente qu'un seul maximum.

R. — *Épis de blés classés d'après leurs longueurs* (1)

Longueurs en pouces	Nombres d'épis
3,0	1
3,5	0
4,0	1
4,5	0
5,0	2
5,5	3
6,0	9
6,5	8
7,0	12
7,5	19
8,0	32
8,5	40
9,0	$y_h = 67$
9,5	63
10,0	38
10,5	21
11,0	8
11,5	2
12,0	1

$S = 327 \quad S_0'' = 127 \quad S_1'' = 362 \quad S_2'' = 1\,574$

$S_0' = 133 \quad S_1' = 250 \quad S_2' = 618$

$h = -0,343 \quad A = 79896 \quad B = 526\,716$

$p = 1,7773 \quad q = -0,7773$

Bien que la courbe représentative soit en forme de cloche, la fonction de probabilité simple ne représente pas la statistique.

Les épis n'appartenaient pas à une même espèce ou race, mais à plusieurs espèces mélangées.

Le calcul par la formule (42) du n° 67 donne

$$y_h = 44.$$

(1) Horace Secrist, *An Introduction to statistical Methods* Macmillan, New-York, 1925, p. 228.

S. T. U. — *Individus de divers pays classés d'après leurs poids* (1)

Poids en livres anglaises	Nombres d'individus		
	Angleterre S	Écosse T	Irlande U
90	2		
100	26	1	5
110	133	8	1
120	388	22	7
130	694	63	42
140	1 240	173	57
150	1 075	255	51
160	881	275	36
170	492	168	25
180	304	125	13
190	174	67	8
200	75	24	1
210	62	14	1
220	33	7	
230	10	4	
240	9	2	
250	3	4	
260	1		

On a ici :

S. $h = -1{,}13132$, $\quad A = 15\,145\,073$, $\quad B = 67\,495\,136$

$$p = 1{,}0508, \qquad q = -0{,}0508.$$

T. $h = 0{,}013707$, $\quad A = 856\,914$, $\quad B = 3\,749\,894$

$$p = 1{,}2075, \qquad q = -0{,}2075.$$

U. $h = 0{,}9876$, $\quad A = 23\,958$, $\quad B = 67\,215$

$$p = 1{,}0595, \qquad q = -0{,}0595.$$

Aucune de ces trois statistiques n'est réductible à une loi de probabilité simple.

Dans chacune d'elles, des races différentes, où les individus sont en nombres comparables, sont mélangées.

On peut observer qu'en groupant les données, dans les cas où p, q, ne sont pas compris entre 0 et 1, on peut obtenir parfois une suite de fréquences donnant lieu à des valeurs de p, q, comprises entre 0 et 1.

Par exemple, en groupant 2 à 2 les fréquences de la statistique T, on obtient les nombres

$$9 \quad 85 \quad 428 \quad 443 \quad 192 \quad 38 \quad 11 \quad 6$$

(1) G. Vdny Yule. *loc. cit.*, statistique B, p. 95.

et le calcul appliqué à cette suite donne

$$p = 0{,}848 ; \quad q = 0{,}152 ; \quad h = 0{,}248 ; \quad m = 9{,}186.$$

Les nombres qu'on déduit à ces constantes sont

153 408 380 194 62 13 2 :

ils ne représentent pas les données.

Le groupement peut donc permettre d'établir une courbe de probabilité simple, mais cet artifice, souvent employé par les statisticiens, ne nous donne pas ici de résultat intéressant.

V. X. Y. Z. — *Températures du mois de juillet à Greenharbour, Spitzberg* (1)

Températures en degrés centigrades	I Année 1913	II Année 1915	III Année 1916	IV Années 1912 1913, 1914, 1915 1916 réunies
9,0 » 9,1	1		1	2
8,0 » 8,1	6		0	8
7,0 » 7,1	14	3	6	29
6,0 » 6,1	14	13	15	43
5,0 » 4,1	13	10	32	71
4,0 » 3,1	23	9	38	96
3,0 » 2,1	41	23	64	181
2,0 » 2,1	66	39	77	330
1,0 » 0,1	90	76	98	531
0,0 » — 0,9	158	163	124	784
— 1,0 » — 1,9	141	144	135	699
— 2,0 » — 2,9	98	147	95	557
— 3,0 » — 3,9	60	81	40	303
— 4,0 » — 4,9	15	30	13	70
— 5,0 » — 5,9	0	6	6	12
— 6,0 » — 6,9	2			2
— 7,0 » — 7,9	2			2

I : $h = 0{,}18548$ A $= 397\,242$ B $= 2\,503\,893$ $p = -0{,}71901$ $q = 1{,}71901$
II : $h = -0{,}56855$ A $= 310\,227$ B $= 1\,717\,368$ $p = -0{,}6409$ $q = 1{,}6409$
III : $h = -1{,}49731$ A $= 325\,818$ B $= 1\,368\,688$ $p = -0{,}20596$ $q = 1{,}20597$
IV : $h = -0{,}01935$ A $= 8\,947\,928$ B $= 44\,754\,688$ $p = -0{,}33334$ $q = 1{,}33334$

Ces températures ne suivent pas la loi de probabilité simple.

(1) Données communiquées par M. G. Rempp, professeur à l'Institut de Physique du globe à l'Université de Strasbourg.

Voir aussi : G. Rempp, *La variabilité de la température au Spitzberg*, Annuaire de l'Institut de Physique du globe de Strasbourg pour 1926.

Observation relative aux valeurs de la constante m

96. Considérons encore la statistique formée par les *deux premières* colonnes du Tableau qui suit :

	O	C_1	O — C_1 +	O — C_1 −	C_2	O — C_2 +	O — C_2 −
18	1	1	0	0	1	0	0
19	1	2		1	3		2
20	7	4	3		6	1	
21	9	9	0	0	10		1
22	13	16		3	18		5
23	20	28		8	30		10
24	46	44	2		46	0	0
25	73	66	7		66	7	
26	105	91	14		91	14	
27	130	119	11		119	11	
28	135	152		17	147		12
29	171	174		3	170	1	
30	177	190		13	190		13
31	210	195	15		194	16	
32	194	189	5		190	4	
33	171	172		1	174		3
34	143	149		6	151		8
35	111	121		10	123		12
36	89	93		4	94		5
37	72	68	4		68	4	
38	54	47	7		46	8	
39	27	30		3	29		2
40	20	19	1		17	3	
41	11	11	0	0	9	2	
42	4	6		2	5		1
43	4	3	1		2	2	
44	2	1	1		1	1	
	2 000	2 000	71	71	2 000	74	74

En voici l'explication :

On a partagé les nombres compris entre

$$9\,000\,000 \quad \text{et} \quad 10\,000\,000$$

en 2.000 tranches (1) (fig. 13).

9 000 001 à 9 000 500 ; 9 000 501 à 9 001 000 ;
9 001 001 à 9 001 500 ; 9 001 501 à 9 002 000 ; etc.

et on a compté combien il y a de nombres *premiers* dans chacune de ces 2.000 tranches.

On a trouvé — 1re et 2e colonne du Tableau :

1 tranche	contenant	18	nombres	premiers	;
1	»	19	»	»	;
7 tranches	»	20	»	»	;
9	»	21	»	»	;
.....	.....	.....	.....	.....	.....
2	»	44	»	»	.

Le graphique des nombres O affecte sensiblement la forme en cloche.

Si l'on soumet les nombres O au calcul, on trouve (2).

$h = -0,045$; $A = 5\,804\,430$; $B = 97\,442\,604$;
$p = 0,6021$; $q = 0,3979$; $\lambda = 16,7652$; $\mu = 16,8102$
$mp = 42,202$; $mq = 27,890$; $m = 70,092$;

les nombres C_1 ont été calculés à partir de ces constantes.

Donnons à m la valeur *entière* 70 voisine de 70,092 et prenons pour p et q les valeurs

$$0,6 \quad \text{et} \quad 0,4,$$

voisines de 0,6021 et 0,3979»; puis donnons à h la valeur O, voisine de — 0,045.

Calculons ensuite les nombres y définis par la formule

$$y = 2\,000 \frac{m\,!}{(mp\ x)\,!\,(mq + x)\,!} p^{mp-x} q^{mq+x}$$

$m = 70$; $p = 0,6$; $q = 0,4$; $x = 0, \ \pm 1, \ \pm 2, \ \pm 3, \cdots$:

Nous trouverons les nombres C_2. qui diffèrent *fort peu* des nombres C_1.

Les nombres C_2, étant ceux qu'on trouverait — théoriquement

(1) Ce travail a été fait par M. KRAITCHIK : *Étude de statistique des nombres premiers, Bulletin de la classe des Sciences de l'Académie Royale de Belgique*, 1er février 1930 .

(2) *Annales Soc. scient. de Bruxelles*, tome L, série A, 1930, et *Revue Gén. des Sciences*, 31 juillet 1930.

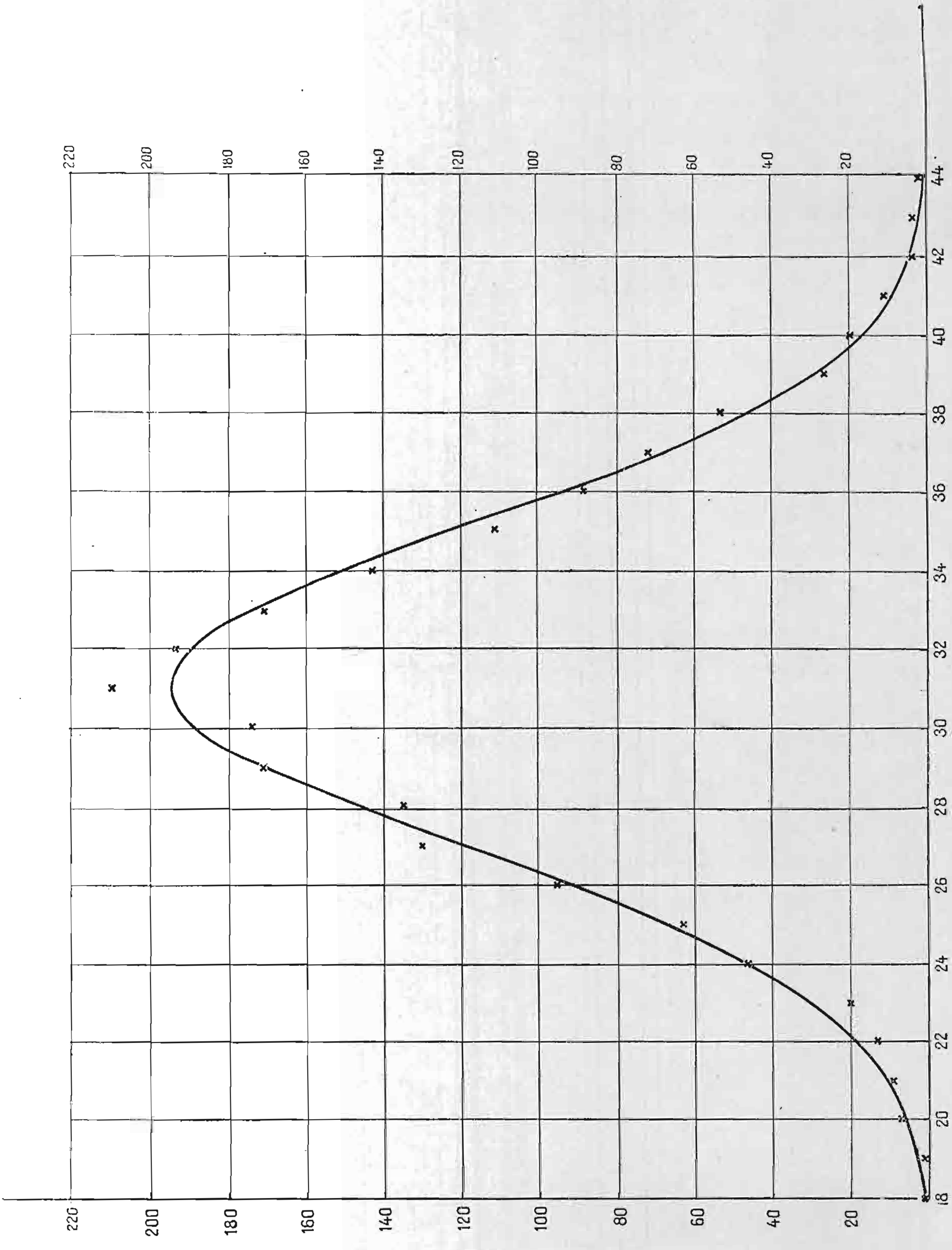


Fig. 13. — Dans les tranches de 500 nombres de la suite des nombres entiers, de 9 000 001 à 10 000 000, la répartition des nombres premiers est soumise aux lois du hasard.

— en faisant 2000 séries de tirages de 70 tirages chaque, d'une urne renfermant des boules rouges et noires dans la proportion 6 : 4, nous concluons :

Si les nombres O *peuvent être regardés comme représentant convenablement* 2000 *séries effectives de* 70 *tirages chaque, on pourra dire que les nombres premiers envisagés sont répartis suivant la loi du hasard,* en donnant au mot hasard, son sens le plus étroit.

Calculons les probabilités des écarts les plus élevés, 10 et 14.

Si

$$A = \Theta\left[\frac{10 + 0,5}{\sqrt{2 \times 2000 \times \frac{30}{2000} \times \frac{2000 - 30}{2000}}}\right]$$

$$= \Theta\left(\frac{10,5}{\sqrt{60}}\right) = \Theta(1,356) = 0,945,$$

la différence

$$1 - 0,945 = 0,055 = \text{environ } 1 : 20$$

est la probabilité d'un écart 10 ou plus grand que 10. Cette probabilité est certainement possible, car cet écart sera atteint — théoriquement — 1 fois si l'on fait 20 séries de 2000 épreuves.

La probabilité de l'écart 14 pour 91 tirages est, semblablement, 0,12 ; cet écart 14 sera atteint — théoriquement — une fois si l'on fait 9 séries de 2000 tirages chacun, puisque

$$0,12 = \text{environ } 1 : 9.$$

Par conséquent, le mot hasard étant pris dans son sens le plus restreint : *dans les tranches de* 500 *nombres de la suite des nombres entiers de* 9000.001 *à* 10.000 000, *la répartition des nombres premiers est soumise aux lois du hasard.*

97. Ce qui intéressera ici le statisticien, *c'est la possibilité, sans modifier sensiblement les résultats, de se borner à la considération de valeurs entières de la constante m* : on peut en effet remplacer les valeurs fractionnaires de m par la valeur entière la plus voisine, comme on vient de le voir.

A dire vrai, nous ne voyons nullement d'ailleurs la nécessité de rendre entières les valeurs de m.

CHAPITRE XII

CAS DE DONNÉES UNILATÉRALES

98. Il arrive que les *données* sont toutes croissantes, ou toutes décroissantes. Le graphique affecte alors non plus la forme de cloche, mais simplement d'un arc de courbe.

Appelons-les *unilatérales*, comme pouvant se rapporter toutes soit à des y_{-x+h} soit à des y_{x+h} d'une courbe de probabilité simple.

Les équations (52) et (53) du n° 69 déterminent *en principe* (ce mot sera expliqué un peu plus loin) la courbe de probabilité simple, *si elle existe.*

Soit par exemple le cas *théorique* obtenu en choisissant comme données les nombres

$$0{,}0015 \text{ à } 0{,}1057$$

de l'exemple numérique traité au n° 68.

Choisissons 0,0228 pour terme en y_h et écrivons, *comme dans le cas où les données sont complètes*,

	y_{-4+h}	$= 0,0015$	$\times 4 = 0,0060$	$\times 4 = 0,0240$
	y_{-3+h}	$= 0,0033$	$\times 3 = 0,0099$	$\times 3 = 0,0297$
	y_{-2+h}	$= 0,0067$	$\times 2 = 0,0134$	$\times 2 = 0,0268$
	y_{-1+h}	$= 0,0128$	$\times 1 = 0,0128$	$\times 1 = 0,0128$
$0,0243$		$0,0243 = S''_{0,4}$	$0,0421 = S''_{1,4}$	$0,0933 = S''_{2,4}$
$0,0228$	y_h	$= 0,0228$		
	y_{1+h}	$= 0,0377$	$\times 1 = 0,0377$	$\times 1 = 0,0377$
	y_{2+h}	$= 0,0578$	$\times 2 = 0,1156$	$\times 2 = 0,2312$
	y_{3+h}	$= 0,0817$	$\times 3 = 0,2451$	$\times 3 = 0,7353$
	y_{4+h}	$= 0,1057$	$\times 4 = 0,4228$	$\times 4 = 1,6912$
$0,2829$		$0,2829 = S'_{0,4}$	$0,8212 = S'_{1,4}$	$2,6954 = S'_{2,4}$
$0,3300 = S$				

On a ici

$$(a) \qquad m = 100 \qquad p = 0{,}1 \qquad q = 0{,}9 \qquad n = 4.$$

Pour h, il est *visible* que (tableau de la page 93).

$$(b) \qquad h = 0{,}3 - 6 = -5{,}7.$$

Remplaçons y_i par **10.000** y_i; on aura

$$S'_{0,n} + y_h - y_{n+h} = 2\,829 + 228 - 1\,057 = 1\,998$$
$$qny_{n+h} = 3\,805{,}4$$

et la première équation (52) du n° 69 devient

$$1\,998\mu - 8\,212 - 2\,829\lambda + 3\,805{,}4 = 0\,;$$

en remplaçant

$$\mu \text{ par } mpq - qh = 9 - 0{,}9h,$$
$$\lambda \text{ par } mpq + ph = 9 + 0{,}1h,$$

on a

$$2\,081{,}1h = -11\,885{,}6$$

et

$$h = -5{,}7112$$

pour $-$ **5,700** (b)

La première équation (52) est donc vérifiée par (a, b) ; il en est de même des autres équations (52, 53) : les équations (52, 53), permettent pratiquement de remonter à la courbe, comme déterminant m, h, p.

Mais il ne faut pas se dissimuler qu'un calcul de ce genre, *à propos d'uns statistique où les y ne sont pas exactement connus, risque d'être gravement faussé pour ce motif que les données sont trop peu nombreuses.*

99. Application aux erreurs accidentelles d'observation. — Certains statisticiens ont l'habitude *regrettable* et condamnable de donner dans leurs écrits, non pas les nombres qu'ils ont obtenus directement, mais des nombres qu'ils ont déduits de ceux-ci.

Il arrive souvent par exemple que des erreurs accidentelles d'observation classées comme il suit

	Nombres d'erreurs accid. observées
....	....
de $-0^s,4$ à $-0^s,3$	δ'
» $-0^s,3$ » $-0^s,2$	γ'
» $-0^s,2$ » $-0^s,1$	β'
» $-0^s,1$ » $0^s,0$	α'
» $0^s,0$ » $+0^s,1$	α''
» $+0^s,1$ » $+0^s,2$	β''
» $+0^s,2$ » $+0^s,3$	γ''
» $+0^s,3$ » $+0^s,4$	δ''
....	....

soient groupées de la manière suivante :

de $0^s,0$ à $0^s,1$	$\alpha' + \alpha'' = \alpha$ erreurs
» $0^s,1$ » $0^s,2$	$\beta' + \beta'' = \beta$ »
» $0^s,2$ » $0^s,3$	$\gamma' + \gamma'' = \gamma$ »
» $0^s,3$ » $0^s,4$	$\delta' + \delta'' = \delta$ »
....	....

et que ce ne soit pas les nombres

$$\cdots, \delta', \gamma', \cdots, \gamma'', \delta'', \cdots$$

qui soient donnés, mais les nombres

$$\alpha, \beta, \gamma, \delta, \cdots$$

On sera alors *réduit* à essayer si une courbe de probabilité simple où $h = 0$ et $p = q = 0,5$ convient.

Exemple (1). :

O	C	O — C +	O — C —
94	92	2	
88	89		1
78	79		1
58	66		8
51	51	0	0
36	37		1
26	25	1	
14	15		1
10	9	1	
7	5	2	
8	2	6	
470	470	12	12

On écrira ici

$94 = y_0$				1
	$y_1 = 88$	$\times\ 1 =$	88	0,965
	78	$\times\ 2 =$	156	0,862
	58	$\times\ 3 =$	174	0,719
	51	$\times\ 4 =$	204	0,558
	36	$\times\ 5 =$	180	0,402
	26	$\times\ 6 =$	156	0,269
	14	$\times\ 7 =$	98	0,167
	10	$\times\ 8 =$	80	0,096
	7	$\times\ 9 =$	63	0,051
	8	$\times\ 10 =$	80	0,025
376	$376 = S_0'$		$1\,279 = S_1'$	5,114
$470 = S_0$				

(1) C. S. Peirce, *On the Theory of errors of Observations, U. S. Coast Survey for the year ending*, november 1, 1870, Washington govt. Printing Office, 1873, Appendix, n° 21, pp. 200-204 and plate n° 27.

On pourra consulter à propos duMémoire de C. S. Pierce : E. B. Wilson, *Proceedings of the National Academy of Science*, vol. 15, p. 120-125 et la première formule **6**

$$S_1' = mpqY_0 \qquad (Y_0 \text{ pour } y_0)$$

du numéro **61** donnera

$$m = \frac{4S_1'}{Y_0} = 54{,}4$$

d'où

$$mp = mq = \frac{54{,}4}{2} = 27{,}2.$$

On emploiera ensuite les formules de récurrence (**4**) et (**5**) du n° **53**, où $h = 0$, en prenant $y_h = 1$, ce qui donne les nombres **1** ; **0,965** ;... qui ont pour somme **5,114**.

En multipliant ces nombres par

$$\frac{470}{5{,}114},$$

on obtient les nombres C

$$92,\ 89,\ \cdots$$

du tableau.

Mais il eut été beaucoup plus intéressant de connaître *séparément* les nombres d'erreurs d'observation positives et les nombres d'erreurs d'observation négatives et de leur appliquer es méthodes générales de calcul.

CHAPITRE XIII

APERÇU SUR LES STATISTIQUES DE GENRES SUPÉRIEURS

---

100. L'étude mathématique des statistiques dont les graphiques présentent plusieurs maximum, et des statistiques dont les graphiques en cloches non réductibles à des formules de probabilité simple laissent entrevoir la présence de plusieurs courbes de probabilité simple voisines et superposées, fera faire à la statistique mathématique des progrès d'une haute importance.

Nous nous bornerons à indiquer comment on peut faire correspondre des tirages de boules d'urnes à des statistiques formées de courbes de probabilité simple superposées.

Considérons les formules

$$\left\{\begin{aligned} &\mathrm{Y}_x = 500\,\frac{50\,!}{(5-x)\,!\,(45+x)\,!}\,0{,}1^{5-x}\times 0{,}9^{45+x},\ m=50\,;\ p=0{,}9\,;\ q=0{,}9 \\ &\mathrm{Z}_x = 1000\,\frac{100\,!}{(50-x)\,!\,(50+x)\,!}\,0{,}5^{100}\,;\ m=100\,;\ p=q=0{,}5 \end{aligned}\right.$$

$$\left\{\begin{aligned} &y_x = \frac{50\,!}{(5-x)\,!\,(45+x)\,!}\,0{,}1^{5-x}\times 0{,}9^{45+x} \\ &z_x = \frac{100\,!}{(50-x)\,!\,50+x)\,!}\,0{,}5^{100}. \end{aligned}\right.$$

On a

		A	Rouges	Noires
$y_{-9} = 0{,}000\,21$	$Y_{-9} =$	0	14	36
$y_{-8} = 0{,}000\,72$	$Y_{-8} =$	0	13	37
$y_{-7} = 0{,}002\,22$	$Y_{-7} =$	1	12	38
$y_{-6} = 0{,}006\,13$	$Y_{-6} =$	3	11	39
$y_{-5} = 0{,}015\,18$	$Y_{-5} =$	8	10	40
$y_{-4} = 0{,}033\,33$	$Y_{-4} =$	17	9	41
$y_{-3} = 0{,}064\,28$	$Y_{-3} =$	32	8	42
$y_{-2} = 0{,}107\,63$	$Y_{-2} =$	54	7	43
$y_{-1} = 0{,}154\,10$	$Y_{-1} =$	77	6	44
$y_{0} = 0{,}184\,92$	$Y_{0} =$	93	5	45
$y_{1} = 0{,}180\,91$	$Y_{1} =$	90	4	46
$y_{2} = 0{,}138\,57$	$Y_{2} =$	69	3	47
$y_{3} = 0{,}077\,94$	$Y_{3} =$	39	2	48
$y_{4} = 0{,}028\,63$	$Y_{4} =$	14	1	49
$y_{5} = 0{,}005\,15$	$Y_{5} =$	3	0	50
		500		

Cela veut dire que sur **500** séries de **50** tirages chacune, il est probable que

1 série sera composée de 12 rouges et de 38 noires,
3 » » 11 » 39 »
8 » » 10 » 40 »
. . . .

On a semblablement

		B	Rouges	Noires
$z_{-16} = 0{,}000\,458$	$Z_{-16} =$	0	66	34
$z_{-15} = 0{,}000\,864$	$Z_{-15} =$	1	65	35
$z_{-14} = 0{,}001\,560$	$Z_{-14} =$	2	64	36
$z_{-13} = 0{,}002\,698$	$Z_{-13} =$	3	63	37
$z_{-12} = 0{,}004\,474$	$Z_{-12} =$	4	62	38
$z_{-11} = 0{,}007\,111$	$Z_{-11} =$	7	61	39
$z_{-10} = 0{,}010\,844$	$Z_{-10} =$	11	60	40
$z_{-9} = 0{,}015\,869$	$Z_{-9} =$	16	59	41
$z_{-8} = 0{,}022\,292$	$Z_{-8} =$	22	58	42
$z_{-7} = 0{,}030\,069$	$Z_{-7} =$	30	57	43
$z_{-6} = 0{,}038\,953$	$Z_{-6} =$	39	56	44
$z_{-5} = 0{,}048\,474$	$Z_{-5} =$	48	55	45
$z_{-4} = 0{,}057\,958$	$Z_{-4} =$	58	54	46
$z_{-3} = 0{,}066\,591$	$Z_{-3} =$	67	53	47
$z_{-2} = 0{,}073\,527$	$Z_{-2} =$	74	52	48
$z_{-1} = 0{,}078\,028$	$Z_{-1} =$	78	51	49
$z_{0} = 0{,}079\,589$	$Z_{0} =$	80	50	50
$z_{1} = 0{,}078\,028$	$Z_{1} =$	78	49	51
$z_{2} = 0{,}073\,527$	$Z_{2} =$	74	48	52
$z_{3} = 0{,}066\,591$	$Z_{3} =$	67	47	53
$z_{4} = 0{,}057\,958$	$Z_{4} =$	58	46	54
$z_{5} = 5{,}048\,474$	$Z_{5} =$	48	45	55
$z_{6} = 0{,}038\,953$	$Z_{6} =$	39	44	56
$z_{7} = 0{,}030\,069$	$Z_{7} =$	30	43	57
$z_{8} = 0{,}022\,292$	$Z_{8} =$	22	42	58
$z_{9} = 0{,}015\,869$	$Z_{9} =$	16	41	59
$z_{10} = 0{,}010\,844$	$Z_{10} =$	11	40	60
$z_{11} = 0{,}007\,111$	$Z_{11} =$	7	39	61
$z_{12} = 0{,}004\,474$	$Z_{12} =$	4	38	62
$z_{13} = 0{,}002\,698$	$Z_{13} =$	3	37	63
$z_{14} = 0{,}001\,560$	$Z_{14} =$	2	36	64
$z_{15} = 0{,}000\,864$	$Z_{15} =$	1	35	65
$z_{16} = 0{,}000\,458$	$Z_{16} =$	0	34	66
		1 000		

Cela veut dire que sur 1 000 séries de 100 tirages chacune, il est probable que

1 série	sera	formée de	65	rouges	et de	35	noires
2	»	»	64	»	»	36	»
3	»	»	63	»	»	37	»

etc...

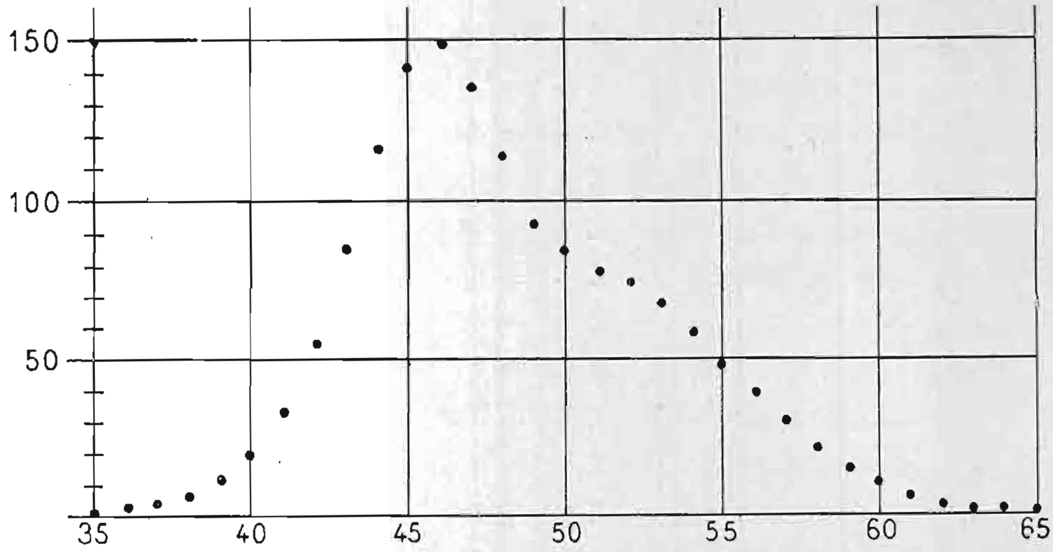


Fig. 14. — Exemple de deux suites de probabilité simple superposées. La figure laisse apercevoir un second maximum.

Réunissant les résultats obtenus *en tenant compte seulement des nombres de noires* :

Nombres de sorties	A	B	A + B
De 35 noires	0	1	1
» 36 »	0	2	2
» 37 »	0	3	3
» 38 »	1	4	5
» 39 »	3	7	10
» 40 »	8	11	19
» 41 »	17	16	33
» 42 »	32	22	54
» 43 »	54	30	84
» 44 »	77	39	116
» 45 »	93	48	141
» 46 »	90	58	148
» 47 »	69	67	136
» 48 »	39	74	113
» 49 »	14	78	92
» 50 »	3	80	83
» 51 »	0	78	78
» 52 »		74	74
» 53 »		67	67
» 54 »		58	58
» 55 »		48	48
» 56 »		39	39
» 57 »		30	30
» 58 »		22	22
» 59 »		16	16
» 60 »		11	11
» 61 »		7	7
» 62 »		4	4
» 63 »		3	3
» 64 »		2	2
» 65 »		1	1

A la colonne A + B correspond la figure 14 : elle présente un maximum pour $x = 46$ et une forte déviation vers $x = 52$, qui indique qu'elle n'est pas du premier genre (fig. 14).

101. Voici un exemple de superposition de deux courbes de probabilité simple, *qu'il a été possible d'obtenir par le calcul* (fig. 15).

Revenons à la statistique des hauteurs barométriques de Southampton (G, n° 93).

Multiplions les nombres C par

$$\frac{254{,}5}{270{,}24},$$

fraction dont le numérateur et le dénominateur sont les nombres

O (observé) et C (calculé)

pour 29 pouces 60.

Nous obtiendrons ainsi une nouvelle courbe C_2, *non tracée sur les figures* 11, 14, analogue à la courbe de la figure 11, avec les mêmes extrémités que celle-ci, mais un peu au-dessous.

Tous les points observés O, de 29 p. 60 à 30 p. 30 seront au-dessus de C_2 (non tracée, on l'a dit) ; $\frac{254{,}5}{270{,}24}$ *a été choisi à cet effet.*

Soit donc

$$C_2 = C \times \frac{254{,}5}{270{,}5}.$$

Nous obtenons, en décalant de gauche à droite de **29 p. 60** à **30 p. 30**, ce qui donne les quatrième et cinquième colonnes ci-dessous :

	O	C_2	O	C_2	$O - C_2$ +	$O - C_2$ —
28p,50	1	0,21			0,79	
60	2	0,51			1,49	
70	2	1,19			0,81	
80	4	2,65			1,35	
90	8,5	5,61			2,89	
29p, 0	13,5	11,37			2,13	
10	21,5	21,91				0,41
20	37	40,12				3,12
30	79	69,57			5,43	
40	108	113,96				5,96
50	181,5	175,79			5,71	
60			254,5	254,50		9,49
70			348,5	344,39	4,11	
80			463,5	433,66	29,84	
90			548,5	505,53	42,97	
30p, 0			602,5	542,30	60,20	
10			619,5	531,58	87,92	
20			500,0	472,16	27,84	
30			382,0	376,16	5,84	
40	237,5	265,42				27,92
50	189,5	163,23			26,27	
60	88,5	85,67			2,83	
70	43,5	37,28			6,22	
80	7	12,89				5,89
90	4	3,31			0,69	
31p, 0	1	0,55			0,45	
	$\Sigma O = 1\,029,0$	$1\,011,24 = \Sigma C_2$				

Les nombres **4,11** à **5,84** de l'avant-dernière colonne paraissent constituer les éléments d'une courbe de probabilité C_3 : nous allons nous en assurer.

Auparavant, ramenons le total **1011,24** à **1029,0** en multipliant les nombres C_2 par

$$\frac{1029,0}{1011,24},$$

ce qui donne

	O	$C_3 = C_2 \times \frac{1029,0}{1011,24}$	O	$C_3 = C_2 \times \frac{1029,0}{1011,24}$	O − C_3 +	−
28^p,50	1	0,21			0,79	
60	2	0,52			1,48	
70	2	1,21			0,79	
80	4	2,69			1,31	
90	8,5	5,71			2,79	
29^p, 0	13,5	11,57			1,93	
10	21,5	22,30				0,80
20	37	40,82				3,82
30	79	70,79			8,21	
40	108	115,97				7,97
50	181,5	178,89			2,11	
60			254,5	258,98		4,48
70			348,5	350,45		1,95
80			463,5	441,29	22,21	
90			548,5	514,42	34,08	
30^p, 0			602,5	551,84	50,66	
10			619,5	540,93	78,57	
20			500,0	480,46	19,54	
30			382,5	382,77		0,77
40	237,5	270,08				32,58
50	189,5	166,10			23,40	
60	88,5	87,10			1,32	
70	43,5	37,94			5,56	
80	7	13,12				5,12
90	4	3,37			0,63	
31^p, 0	1	0,57			0,43	
	ΣO = 1 029,0	1 029,04 = ΣC_3				

Les nombres suivants (données) sont extraits sauf les deux premiers et le dernier, de la colonne $O - C_3$:

Données		
zéro	pour	— 4,48
zéro	»	— 1,95
22,21	»	22,21
34,08	»	34,08
50,66	»	50,66
40 + 38,57	»	38,57
19,54	»	19,54
zéro	»	— 0,77
165,06		157,86

soumettons-les au calcul ordinaire avec S = 157,86 ; ils conduisent à

$$h = 0{,}0054\ ;\quad p = 0{,}449\ ;\quad q = 0{,}551\ ;\quad \lambda = 1{,}415\ ;\quad \mu = 1{,}409$$
$$mp = 2{,}564\ ;\quad mq = 3{,}147\ ;\quad m = 5{,}711.$$

On a mis à part, par interpolation graphique, le nombre 40 : on en fera l'usage qui sera indiqué plus loin. Partant des données précédentes, on trouve par le calcul :

Données	Calculé
zéro	0,84
zéro	7,53
22,21	17,15
34,08	32,17
50,66	44,58
78,57	33,72
19,54	16,55
zéro	5,32
	0
	157,86

Ce sont les éléments d'une courbe de probabilité simple C_4 *qui vient se superposer* à C_3, comme le montre le Tableau suivant, résumé des calculs effectués :

	O	C_3	C_4	anomalie	$C = C_3 + C_4$	O — C +	O — C —	Éc. prob. ±
28ᵖ,50	1	0,21			0,21	0,79		
60	2	0,52			0,51	1,48		
70	2	1,21			1,21	0,79		
80	4	2,69			2,69	1,31		
90	8,5	5,71			5,71	2,79		
29ᵖ, 0	13,5	11,57			11,57	1,93		
10	21,5	22,30			22,30		0,80	
20	37	40,82			40,82		3,82	
30	79	70,79			70,79	8,21		
40	108	115,97			115,97		7,97	
50	181,5	178,89	0		178,89	2,61		
60	254,5	258,98	0,84		259,82		5,32	
70	348,5	350,45	7,53		357,98		9,48	
80	463,5	441,29	17,15		458,44	5,06		
90	548,5	514,42	32,17		546,59	1,91		
30ᵖ, 0	602,5	551,84	44,58		596,42	6,08		
10	619,5	540,93	33,72	40,00	574,65	4,85		19,0 (pour 40)
20	500	480,46	16,55		497,01	2,99		
30	382	382,77	5,32		388,09		6,09	
40	237,5	270,08	0		270,08		32,58	10,2
50	189,5	166,10			166,10	23,40		8,0
60	88,5	87,18			87,18	1,32		
70	43,5	37,94			37,94	5,56		
80	7	13,12			13,12		6,12	
90	4	3,37			3,37	0,63		
31ᵖ, 0	1	0,57			0,57	0,43		
	4 748	4 550,18	157,86	40		72,14	72,18	
		157,86						
		40						
		4 748,04						

Subsistent *seules* les anomalies *a'*, *b*, *c* de la figure 15. On peut même se demander si ce sont des anomalies (nº 90).

En effet, pour *a'*

l'écart probable relatif à C_3 est

$$0{,}477\sqrt{2 \times 4550{,}18 \times \frac{540{,}93}{4550{,}18} \times \frac{4550{,}18 - 540{,}93}{4550{,}18}} - 0{,}5 = 16{,}0,$$

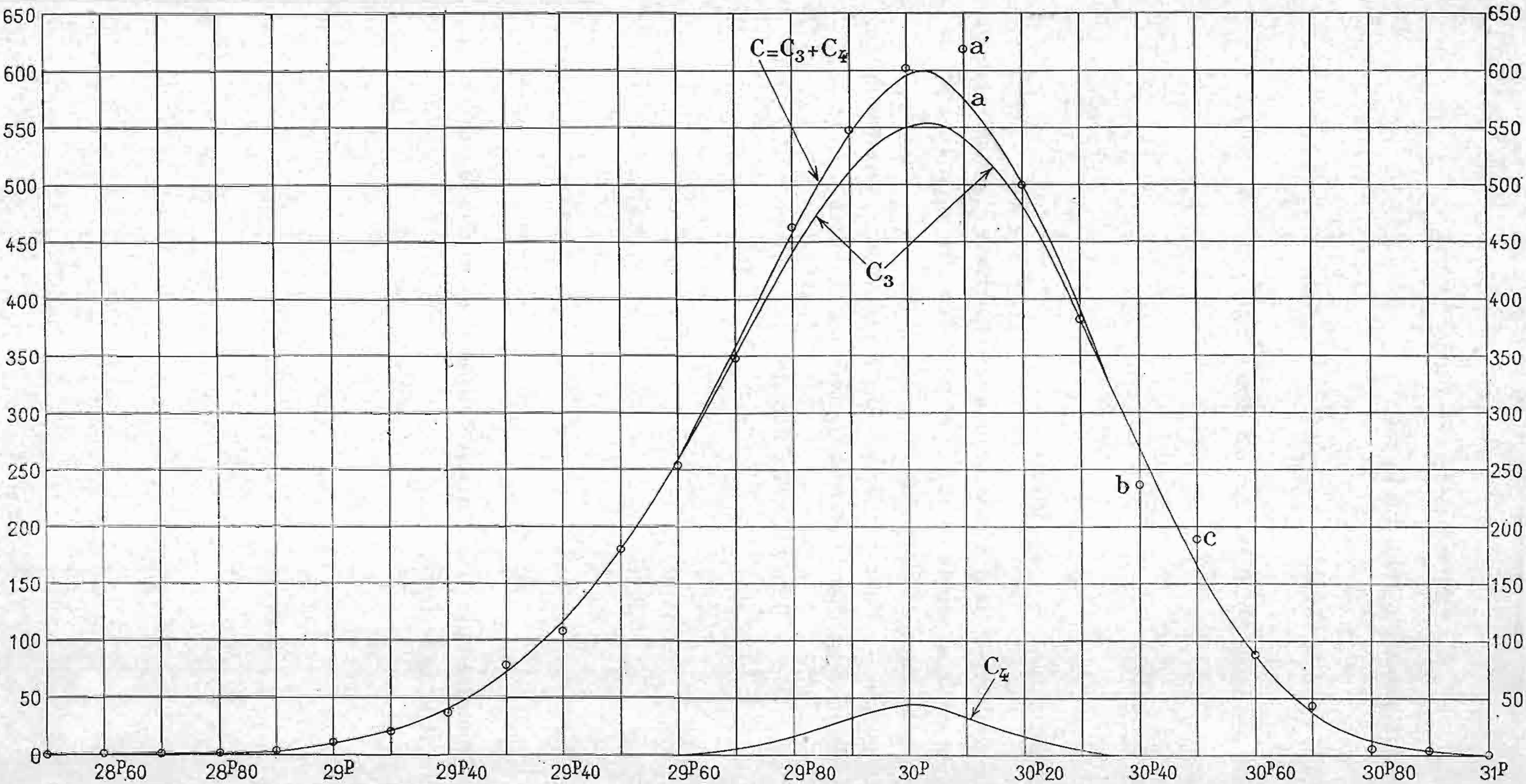


Fig. 15. — Hauteurs barométriques à Southampton. Il a été possible de mettre en évidence la superposition de deux phénomènes secondaires (30 p. et 30 p. 10) au phénomène principal. (Comparer avec la fig. 11).

et l'écart probable relatif à C_4 est

$$0,477\sqrt{2.\times 157,86 \times \frac{33,72}{157,86} \times \frac{157,86-33,72}{157,86}} - 0,5 = 3,0,$$

l'écart probable total pour a' est donc $16,0 + 3,0 = 19,0$;

pour b,

écart probable $= 10,2$;

pour c,

écart probable $= 8,0$.

En définitive, *il y a une anomalie importante pour* **30** *pouces, mise en évidence par la courbe* C_4 et peut-être des anomalies secondaires pour **30 p. 10, 30 p. 40, 30 p. 50.**

Remarque. — Pour rendre le calcul *formellement* irréprochable, il faudrait recalculer la courbe C_4 avec les nombres

0,84	au lieu de	zéro
7,53	»	zéro
22,81	»	22,81
34,08	»	34,08
50,66	»	50,66
38,57	»	38,57
19,54	»	19,54
5,32	»	zéro :

on obtiendrait des résultats si voisins des précédents qu'il n'est guère utile de le faire.

CHAPITRE XIV

STATISTIQUES CONTINUES ET STATISTIQUES DISCONTINUES

I. — CLASSIFICATION DES STATISTIQUES EN STATISTIQUES CONTINUES ET EN STATISTIQUES DISCONTINUES

102. Quand on tire m fois de suite des boules d'une urne, quand on note le nombre r, de boules rouges sorties et le nombre n_1 de boules noires sorties, quand on recommence ces m tirages un certain nombre de fois, les nombres

$$r_1, r_2, r_3, \cdots$$

constituent une statistique ; les nombres

$$n_1, n_2, n_3, \cdots$$

constituent aussi une statistique.

Ces statistiques sont *discontinues* par opposition aux statistiques *continues* dont nous allons parler.

103. Soit un lot d'épis de blé. Je les classe d'après leurs longueurs : 1° je puis les classer par millimètres (arrondis) de longueur, en comptant pour 6 millimètres par exemple les longueurs allant de 5 mm. 5 à 6 mm. 5 ; je puis aussi les classer par pouces, en comptant par exemple 3 pouces pour les longueurs allant de 2 p. 5 à 3 p. 5 : la statistique sera dite continue; les nombres qui figurent dans les statistiques continues dépendent de l'unité de mesure choisie. *On transforme les statistiques continues en statistiques discontinues, accessibles au calcul,* en choisissant une unité de mesure et en arrondissant les mesures aux multiples entiers de l'unité.

La plupart des statistiques sont continues. Il existe cependant des statistiques discontinues. Nous en avons vu un exemple (statistique D, ci-dessus). Un autre exemple serait les nombres de pétales de différentes roses portées par un même rosier.

104. Admettons qu'une statistique continue, faite en partant d'une certaine unité de longueur, de temps, de température, ou autre, suive la loi de probabilité simple

$$(1) \qquad y_x = \frac{m\,!}{(mp - x - h)\,!\,(mq + x + h)\,!}\, p^{mp-x-h} q^{mq+x+h}$$

et changeons l'unité de mesure.

La nouvelle statistique suivra-t-elle encore une loi de probabilité simple telle que (1), définie par de nouvelles valeurs (m', p', q', h') des constantes ?

Nous allons montrer que, sous certaines réserves, il en est ainsi, *malgré que la fonction* (1) *n'ait pas de théorème d'addition.*

Nous considérerons à cet effet une suite de nombres vérifiant la loi (1) ; nous les grouperons 3 à 3, puis 5 à 5, ce qui revient à changer l'unité de mesure, à la prendre d'abord 3 fois plus grande, à la prendre ensuite 5 fois plus grande, et nous verrons que les nombres obtenus par ces groupements peuvent encore, sous certaines réserves avons-nous dit, en fait compte tenu de l'écart probable ou d'écarts raisonnablement possibles, être représentés par de nouvelles lois (1).

II. — CHANGEMENT DE L'UNITÉ DE MESURE DANS LES STATISTIQUES CONTINUES

105. On a (voir NOTE, page 204), en posant

$$y_x = \frac{m!}{(mp - x)!\,(mq + x)!} p^{mp-x} q^{mq+x}$$

$$m = 1\,000 \quad p = 0{,}1 \quad q = 0{,}9:$$

$$\begin{aligned}
y_{-73} &= 0{,}000\ 000\ 0000\ 00523 \\
y_{-72} &= 0{,}000\ 000\ 0000\ 00992 \\
&\dots\dots\dots\dots \\
y_{-1} &= 0{,}041\ 600\ 783 \\
y_0 &= 0{,}042\ 016\ 791 \\
y_1 &= 0{,}041\ 970\ 157 \\
&\dots\dots\dots\dots \\
y_{59} &= 0{,}000\ 000\ 0000\ 0171\ 04 \\
y_{60} &= 0{,}000\ 000\ 0000\ 0052\ 223
\end{aligned}$$

Ces nombres ont été calculés avec suffisamment de décimales pour assurer 6 décimales exactes à *toutes les sommes et à tous les nombres* qu'on va en déduire.

106. Groupons les nombres y comme il suit ; posons

$$\begin{aligned}
&\dots\dots\dots\dots \\
z'_{-2} &= y_{-5} + y_{-6} + y_{-7} \\
z'_{-1} &= y_{-2} + y_{-3} + y_{-4} \\
z_0' &= y_1 + y_0 + y_{-1} \\
z_1' &= y_4 + y_3 + y_2 \\
z_2' &= y_7 + y_6 + y_5 \\
&\dots\dots\dots\dots
\end{aligned}$$

on trouve les résultats numériques suivants :

$z'_{-23} = 0,000\ 000\ 0000\ 21325$
0001 26516
0007 0348
0036 6295
0178 444
0812 44
3453 4
$z'_{-16} = 0,000\ 001\ 3689$
005 0549
017 3331
055 4432
164 2628
451 0239
$z'_{-10} = 0,001\ 146\ 0735$
2 691 093
5 830 038
11 633 973
21 347 746
35 954 872
55 477 724
078 264 662
100 733 432
$z'_{-1} = 0,118\ 025\ 925$
$z'_{0} = 0,125\ 587\ 731$

$z'_{1} = 0,121\ 058\ 055$
105 429 058
082 721 246
058 297 783
036 785 864
020 709 399
010 364 347
004 592 675
$z'_{9} = 0,001\ 794\ 2313$
615 1237
184 1369
047 8668
010 7411 8
$z'_{14} = 0,000\ 002\ 0670\ 0$
3386 66
0468 686
0054 2989
0005 2142 9
$z'_{19} = 0,000\ 000\ 0000\ 4104\ 35$

Dans quelle mesure les nombres z' qui résultent du groupement des y pris 3 à 3, suivent-ils une loi de probabilité simple, en raison du fait que les y suivent eux-mêmes une telle loi ?

Les procédés de calcul que nous avons développés donnent pour les z' :

$$S = 1,000000 \quad S_0'' = 0,431800 \quad S_1'' = 1,255819 \quad S_2'' = 5,148766$$
$$S_0' = 0,442613 \quad S_1' = 1,255822 \quad S_2'' = 4,925808$$
$$h = -0,00003 \quad A = 1,098105 \quad B = 10,074768$$
$$p' = 0,367811 \quad q' = 0,632189 \quad m = 43,37295 :$$

h est négligeable.

En calculant z_0' par la formule (5) du n° 27 où $x = 0$:

$$\log z_0' = -\log\sqrt{2\pi} + \frac{1}{2}\log\frac{1}{mpq} + \frac{M}{12m} - \frac{M}{12mpq},$$

on trouve

$$\log z_0' = -\bar{1},096316 \qquad z_0' = 0,124838\,;$$

les formules de récurrence (8) et (9) du n° 33 donnent ensuite les z' de la troisième colonne du Tableau que voici :

x	z_x' obtenus par groupement des y 3 à 3 O	z_x' calculés comme on l'a dit C
— 16	0,000 001	0,000 001
— 15	5	3
	17	12
	55	42
	164	137
	451	400
— 10	0,001 146	1 063
	2 691	2 573
	5 830	5 682
	11 634	11 455
	21 348	21 593
	35 955	36 339
	55 478	55 878
	78 265	78 473
	100 733	100 564
— 1	118 026	117 455
0	125 587	124 838
+ 1	121 058	120 437
	105 429	105 215
	82 721	82 955
	58 328	58 782
	36 786	37 250
	20 709	20 985
	10 364	10 430
	4 593	4 531
	1 794	1 701
10	0,000 615	543
	184	145
	48	31
	11	5
	2	1
15	0,000 000	0

On a ensuite, les nombres 100 000 O et 100 000 C étant arrondis à l'unité, et les écarts probables concernant les nombres 100 000 C (fig. 16) :

x	100 000 O O_1	100 000 C C_1	$O_1 - C_1$ +	$O_1 - C_1$ —	Écarts prob. ±
— 15	1	1	0	0	0,5
	2	1	1		0,5
	6	4	2		1,0
	16	14	2		2,4
	45	40	5		4,3
— 10	115	106	9		6,9 **
	269	257	12		10,8 *
	583	568	15		16,0
	1 163	1 146	17		23,0
	2 135	2 159		24	30,3
— 5	3 595	3 634		39	39,9
	5 548	5 588		40	49,0
	7 826	7 847		21	57,4
	10 073	10 056	17		64,1
— 1	11 803	11 745	58		68,7
0	12 559	12 482	77		70,5 *
+ 1	12 106	12 043	63		69,4
	10 543	10 521	22		65,4
	8 272	8 295		23	58,8
	5 833	5 878		45	50,2
5	3 679	3 725		46	40,4 *
	2 071	2 098		27	30,6
	1 036	1 043		7	21,7
	459	453	6		14,3
	179	170	9		8,8
10	62	54	8		4,9 **
	18	15	3		2,6
	5	3	2		1,2
13	1	1	0		0,5

L'*écart probable* n'est vraiment dépassé que pour les nombres où figure un * ou surtout deux **.

Multiplions les nombres O_1 et C_1 par x. Ils deviendront par exemple pour $x = 5$,

$$3\,679\,x \qquad \text{et} \qquad 3\,725\,x\,;$$

l'écart $O_1 - C_1$ deviendra

$$46\,x\,;$$

l'écart probable

$$E_5 = 0{,}477\sqrt{2 \times 100\,000 \times \frac{3\,725}{100\,000} \times \frac{100\,000 - 3\,725}{100\,000}} = 40{,}4\,;$$

deviendra

$$0{,}477\sqrt{2 \times 100\,000\,x \times \frac{3\,725\,x}{100\,000\,x} \times \frac{100\,000\,x - 3\,725\,x}{100\,000\,x}}$$
$$= E_5\sqrt{x} = 40{,}4\sqrt{x}.$$

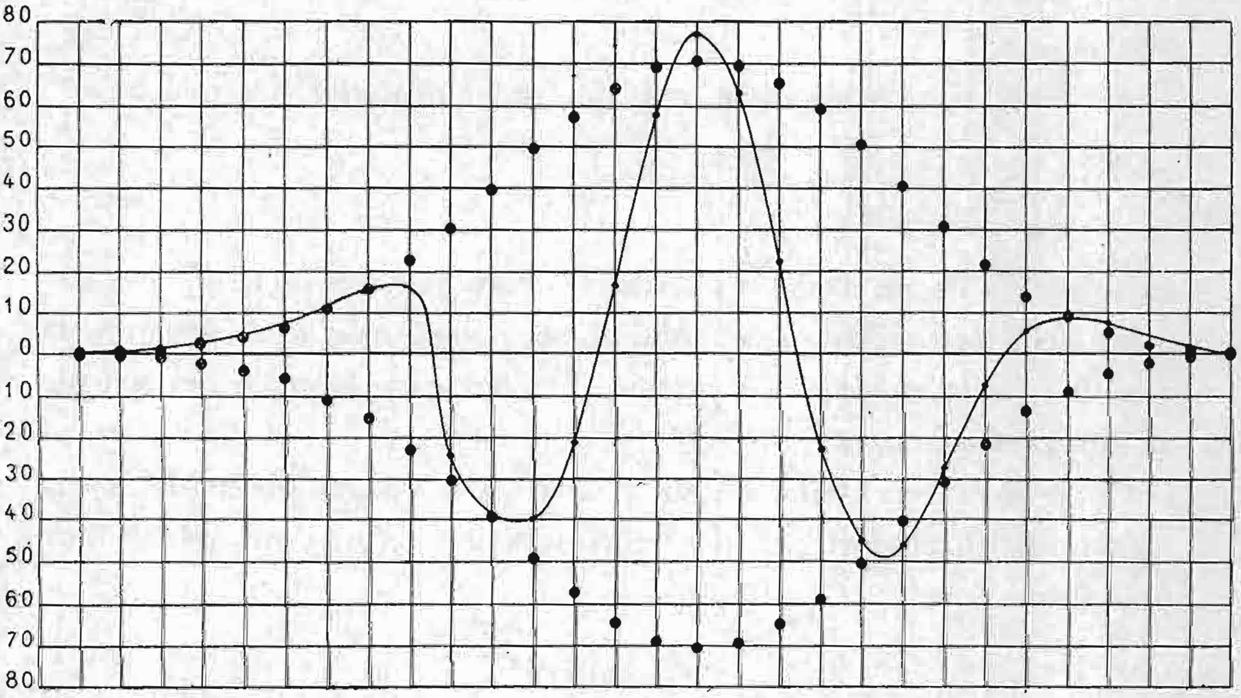


Fig. 16. — Comparaison entre une suite de nombres qui suivent approximativement une loi de probabilité simple et les écarts probables. La courbe se rapporte aux différences entre la suite de nombres et la loi de probabilité simple correspondante. Les ● figurent les écarts probables.

Donc si nous multiplions les nombres O_1 par $\left(\frac{40{,}4}{46}\right)^2 = 0{,}77\ 118$, ou, ce qui revient au même, si nous multiplions les nombres O non plus par 100 000 mais par 77118, l'écart probable n'est plus dépassé pour $x = 5$ dans les nombres 77118 z' ; comme

$$\left(\frac{10{,}8}{12}\right)^2 = 0{,}81,$$

il n'est pas dépassé non plus pour $x = -$ **11.**

Si l'on veut que l'écart probable ne soit pas dépassé non plus pour

$$x = 10 \qquad x = -11,$$

ce qui au point de vue *pratique* est bien inutile, il faudra multiplier par

$$100\,000\left(\frac{4,9}{8}\right)^2 = 37\,502,$$

au lieu **77118.**

On s'arrêterait à **77 118** si les nombres O provenaient d'une statistique.

En toute rigueur, compte tenu de l'écart probable, les nombres

$$37\,502\, z_x'$$

suivent une loi de probabilité simple, bien que résultant du groupement **3** *à* **3** *de nombres y qui suivent eux-mêmes une loi de probabilité simple ; ils la suivent tout comme s'ils se rapportaient à des tirages de boules d'une urne.*

On pourrait substituer l'*écart possible* à l'écart probable, mais cela nous entraînerait à des considérations qui n'ont guère leur place ici.

107. La comparaison des nombres O′ — C_1 du Tableau précédent et des nombres O — C relatifs à certaines statistiques, par exemple « Etoile polaire », n° 89, et B, n° 92, donne lieu de penser que le choix du nombre d'observations ou le choix de l'unité (millimètre, centimètre, par exemple), que le choix en fait du multiplicateur (**100 000, 77 118,** etc.) peut conduire à une représentation plus ou moins serrée, par la loi de probabilité simple, d'une statistique donnée.

En effet, nous venons de voir que les y_x sont *mieux* représentés que les

$$100\,000\, z_x \qquad \text{ou les} \qquad 77\,118\, z_x \,;$$

or l'unité de mesure des y_x est trois fois plus grande que celle des z_x.

Au point de vue pratique, dans l'étude des statistiques ordinaires, ces considérations n'ont peut-être actuellement qu'un intérêt médiocre : quelle que soit la manière de grouper, par exemple, les nombres de la statistique E, n° 95, les anomalies signalées dans cette statistique, ainsi que les conséquences de ces anomalies, resteront en évidence ; mais la manière de grouper pourrait intervenir dans des études particulièrement serrées de statistiques. Il n'est pas douteux en effet que telle unité de mesure donnera une meilleure répartition suivant la loi de probabilité simple que telle autre unité de mesure et que, parmi les unités de mesure, il en sera une donnant la répartition la meilleure. Nous dirons qu'une répartition z' (ou z'') est meilleure qu'une répartition z''' quand le multiplicateur faisant descendre les O — C des z' au-dessous des écarts probables est plus grand que le multiplicateur correspondant des z''' ; et parmi les répartitions z''', z'', z', y, il se peut qu'il y en ait une y qui soit optimum au point de vue de la loi de probabilité simple. Les groupements 5 à 5 vont préciser ce point de vue.

Nous devions d'ailleurs les mettre en évidence, pour bien établir la base de nos calculs.

108. On peut grouper aussi les y comme il suit :

$$
\begin{array}{l}
\cdots\cdots\cdots\cdots\cdots\cdots \\
z''_{-3} = y_{-7} + y_{-8} + y_{-9} \\
z''_{-2} = y_{-4} + y_{-5} + y_{-6} \\
z''_{-1} = y_{-1} + y_{-2} + y_{-3} \\
z_0'' = y_2 \;\; + y_1 \;\; + y_0 \\
z_1'' = y_5 \;\; + y_4 \;\; + y_3 \\
z_2'' = y_8 \;\; + y_6 \;\; + y_7 \\
\cdots\cdots\cdots\cdots\cdots\cdots
\end{array}
$$

$z''_{-23} = 0{,}000\ 000\ 0000\ 38882$	$z''_1 = 0{,}116\ 933\ 647$
0002 25163	98 388 272
0012 2829	74 509 645
0062 559	50 628 357
0298 004	30 767 696
1326 21	16 661 805
5508 3	8 010 836
$z''_{-16} = 0{,}000\ 002\ 1328$	$z''_8 = 0{,}003\ 405\ 4783$
007 6557	274 4252
025 7828	417 8439
080 2993	119 4093
231 9846	029 5759 7
260 8316	006 3101 8
$z''_{-10} = 0{,}001\ 536\ 8380$	$z''_{14} = 0{,}000\ 001\ 1518\ 9$
3 513 6818	1785 68
7 407 836	0233 150
14 377 679	0025 403
25 644 598	$z''_{18} = 0{,}000\ 000\ 0002\ 285\ 9$
41 957 923	
62 849 897	
86 014 816	
107 322 750	
$z''_{-1} = 0{,}121\ 805\ 821$	
$z''_0 = 0{,}125\ 445\ 274$	

Ici, pour les z'',

$$S = 1{,}000001 \quad S_0'' = 0{,}473401 \quad S_1'' = 1{,}427630 \quad S_2'' = 6{,}049517$$
$$S_0' = 0{,}401155 \quad S_1' = 1{,}094329 \quad S_2' = 4{,}135686$$

Ces nombres sont tous exacts à 0,000 001 près.

On en déduit

$$h = +0{,}333301 \quad A = 1{,}0913275 \quad B = 10{,}9620944$$
$$p = 0{,}368611 \quad q = 0{,}631389 \quad m = 43{,}31108.$$

et, par la formule 50 du n° 66 : $y_h = 0{,}124\ 363$. Les formules de récurrence du n° 55 donneront ensuite z''_{-x+h} et z''_{x+h}.

109. Groupons maintenant les y 5 à 5 de la manière suivante :

$$\begin{array}{l}
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
z'''_{-1} = y_{-3} + y_{-4} + y_{-5} + y_{-6} + y_{-7} \\
z_0''' = y_2 + y_1 + y_0 + y_{-1} + y_{-2} \\
z_1''' = y_7 + y_6 + y_5 + y_4 + y_3 \\
z_2''' = y_{12} + y_{11} + y_{10} + y_9 + y_8 \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots
\end{array}$$

Yoici les valeurs des nombres z''' :

$z'''_{-14} = 0,000\ 000\ 0000\ 24170$	$z'''_1 = 0,185\ 028\ 787$
4 58796	124 196 334
72 7215	62 002 496
958 504	22 679 807
$z'''_{-10} = 0,000\ 001\ 0456\ 11$	5 977 1122
69 1396	$z'''_6 = 0,001\ 113\ 4617$
414 5155	143 4523
$z'''_{-6} = 0,002\ 012\ 0812$	$z'''_8 = 0,000\ 012\ 4643\ 4$
7 855 8948	7092 78
24 497 509	0255 365
60 548 324	$z'''_{11} = 0,000\ 000\ 0005\ 581$
$z'''_{-2} = 0,117\ 633\ 144$	
$z'''_1 = 0,178\ 019\ 593$	
$z'''_0 = 0,207\ 785\ 821$	

Avec 6 décimales exactes :

$$\begin{array}{llll}
S = 1,000001 & S_0'' = 0,391061 & S_1'' = 0,747823 & S_2'' = 1,879891 \\
 & S_0' = 0,401155 & S_1' = 0,747825 & S_2' = 1,799113 \\
 & A = 0,592437 & B = 2,192005 & \\
h = -0,000002 & p = 0,421018 & q = 0,578982 & m = 15,17867
\end{array}$$

En négligeant h :

$$\log z_0''' = \bar{1},3094132 \qquad z_0''' = 0,203898.$$

On a ensuite

x	z''' par addition des y 5 à 5	z''' calculés	$10\,000\,z'''$ par groupement des y 5 à 5 O	$10\,000\,z'''$ par calcul C	O — C +	O — C —	Écarts prob. ±
— 9	0,000 009	0,000 001	0	0	0	0	0
— 8	69	24	1	0	1		0
— 7	415	262	4	3	1		1,2
— 6	2 012	1 702	20	17	3		2,8
— 5	7 856	7 746	79	77	2		5,9
— 4	24 498	25 340	245	253		8	10,6
— 3	60 548	62 555	606	626		20	16,3 *
— 2	0,117 633	0,119 003	1 176	1 190		14	21,8
— 1	0,178 019	0,176 309	1 780	1 763	17		25,7
0	0,207 786	0,203 898	2 078	2 039	39		27,2 *
+ 1	0,185 029	0,183 067	1 850	1 831	19		26,1
2	0,124 196	0,125 792	1 242	1 258		16	22,4
3	62 002	64 578	620	645		25	16,6 *
4	22 680	23 546	227	235		8	10,2
5	5 977	5 602	60	56	4		4,8
6	1 113	725	11	7	4		1,9 *
7	143	0,000 025	1	0	1		0
			10 000	10 000	91	91	

Comme on l'a expliqué à propos des z', il suffirait de substituer à 10 000 le plus petit des 3 nombres

$$10\,000\left(\frac{16,3}{20}\right)^2=6642,\quad 10\,000\left(\frac{27,2}{39}\right)^2=4\,864,\quad 10\,000\left(\frac{16,6}{25}\right)^2=4\,409,$$

donc 4 409, pour que les nombres (O — C) soient tous en deçà des écarts probables ; le statisticien négligera en effet 1,9.

Ainsi, *les nombres*

$$Az''' \qquad A \leqslant 4\,409,$$

où les z''' résultent du groupement 5 à 5 des y, suivent une loi de probabilité simple, compte tenu de l'écart probable, tout comme si les nombres Az''' résultaient du tirage de boules d'une urne.

III. — JUSTIFICATION DE LA CONSTANTE DE DÉPLACEMENT h

110. Dans les tirages de boules d'une urne, il n'y a pas lieu d'envisager de constante de déplacement.

Pourquoi celle-ci s'introduit-elle dans l'étude des statistiques ? Nous pouvons actuellement répondre à cette question.

Pour les z' de tout à l'heure, la constante de déplacement est nulle. Pour les z'', la constante de déplacement est $\frac{1}{3}$.

C'est donc la manière de grouper les éléments qui introduit la constante h.

Si nous apprécions une suite de données thermométriques en ramenant celles-ci aux degrés arrondis

$$\cdots 0^{\circ},\quad 1^{\circ},\quad 2^{\circ},\quad 3^{\circ},\ \cdots,$$

si nous trouvons $h = \frac{1}{3}$, nous trouverons $h = 0$ en appréciant les températures suivant l'échelle ...

$$\cdots\left(0+\frac{1}{3}\right)^{\circ},\quad \left(1+\frac{1}{3}\right)^{\circ},\quad \left(2+\frac{2}{3}\right)^{\circ},\ \cdots$$

de même qu'en passant de z' à z''.

En *refaisant* une statistique quelconque sur une échelle nouvelle, indiquée par la valeur de h, on pourrait donc s'arranger de manière à ce que h soit nul : non pas en modifiant l'unité de mesure, mais en décalant les mesures d'une certaine fraction, $\frac{1}{3}$ de degré dans le cas précédent.

Il n'y a aucune utilité pratique à procéder ainsi. — Il y aurait même souvent impossibilité, parce que la plupart des statistiques à étudier ne comportent pas la possibilité de changer d'échelle.

Mais il était nécessaire de montrer pourquoi la constante de déplacement s'introduit dans les calculs et pourquoi les nombres y_{x+h} se substituent ici aux nombres y_x qui représentent les probabilités de tirage de boules d'une urne.

IV. — LES CONSTANTES m, p, q DANS LES GROUPEMENTS

Etude de mpq

111. Quand nous groupons $2n + 1$ à $2n + 1$ des nombres qui suivent une loi de probabilité simple, quand nous écrivons

$$\begin{aligned}
&\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\\
z_{-1} &= y_{-3n-1} + y_{-3n} + \cdots + y_{-n-1}\\
z_0 &= y_{-n} + y_{-n+1} + \cdots + y_{-1} + y_0 + y_1 + \cdots + y_n\\
z_1 &= y_{n+1} + y_{n+2} + \cdots + y_{3n+1}\\
&\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots
\end{aligned}$$

nous obtenons, sous certaines conditions qui ont été indiquées, une suite

$$\dots,\quad z_2,\quad z_1,\quad z_0,\quad z_1,\quad z_2, \dots$$

telle que les nombres

$$\dots,\quad Az_{-2},\quad Az_{-1},\quad Az_0,\quad Az_1,\quad Az_2, \dots$$

où A est convenablement choisi, suit une loi de probabilité simple, compte tenu de l'écart probable.

Il en est de même des suites

$$\dots,\quad Bz_{-1},\quad Bz_{-2},\quad Bz_0,\quad Bz_1,\quad Bz_2, \dots$$

pourvu que

$$B < A.$$

On a sensiblement (n° 28)

$$y_0 = \frac{1}{\sqrt{2\pi mpq}}. \tag{1}$$

Soient m', p', q' les constantes relatives à la suite z ; on a, de même,

$$z_0 = \frac{1}{\sqrt{2\pi m'p'q'}};$$

or

$$z_0 = (2n+1)y_0 - \varepsilon_n = \frac{2n+1}{\sqrt{2\pi mpq}} - \varepsilon_n ;$$

donc

$$\frac{1}{\sqrt{2\pi m'p'q'}} = \frac{1}{\sqrt{2\pi mpq : (2n+1)^2}} - \varepsilon_n,$$

$$\sqrt{2\pi m'p'q'} = \sqrt{2\pi mpq : (2n+1)^2} + \eta_n$$

(2) $$m'p'q' = mpq : (2n+1)^2 + \omega_n,$$

Soit

$$m = 1000, \quad p = 0{,}1, \quad q = 0{,}9, \quad mpq = 90.$$

1° Groupons les y 3 à 3 ; pour les z' ainsi obtenus :

$$m'p'q' = mpq : (2n+1)^2_{+\omega_3} = mpq : 3^2_{+\omega_3} = 90 : 9_{+\omega_3} = 10 + \omega_3.$$

Nous avons trouvé (n° 102)

$$p = 0{,}367811, \quad q = 0{,}632189, \quad m = 43{,}37295,$$

d'où

$$m'p'q' = 10{,}09 ;$$

on a donc

$$\omega_3 = 0{,}09.$$

On peut trouver une valeur approchée de ω_3 ; en effet (n° 55)

$$y_{-1} = \frac{mq}{mp+1}\frac{p}{q}y_0, \quad y_1 = \frac{mp}{mq+1}\frac{q}{p}y_0 ;$$

donc

$$z_0 = y_0\left(1 + \frac{mpq}{mpq+q} + \frac{mpq}{mpq+p}\right)$$

$$= y_0 \frac{3(mpq)^2 + 2mpq + pq}{(mpq)^2 + 2mpq + pq} ;$$

$$\frac{1}{z_0^2} = \frac{1}{y_0^2}\frac{[(mpq)^2 + mpq + pq]^2}{[3(mpq)^2 + 2mpq + pq]^2}$$

ou formule (1)

$$m'p'q' = mpq \frac{[(mpq)^2 + mpq + pq]^2}{[3(mpq)^2 + 2mpq + pq]^2}$$

en remplaçant mpq par sa valeur 90, on trouve

$$m'p'q' = 10{,}074\ ;$$

la valeur *vraie* est 10,085.

Dans le groupement z'',

$$m''p''q'' = 10{,}045\ :$$

d'un groupement n à n à un autre groupement n à n, par exemple z' et z_4', $m'p'q'$ ne diffère pas beaucoup de $m''p''q''$ si n est petit ($n = 2$ ou $n = 3$).

2° Pour le groupement z''', la formule (2) donne

$$m'''p'''q''' = 3{,}6000 + \omega_n$$

et on a

$$m'''p'''q''' = 3{,}6990$$

d'où

$$\omega_n = 0{,}09999.$$

Étude de p et de q

112. Nous avons écrit

$$\begin{aligned}
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
z^1_{-1} &= y_{-4} + y_{-3} + y_{-2}\\
z_0' &= y_{-1} + y_0 \;\; + y_1\\
z_1' &= y_2 + y_3 + y_4\\
&\ldots\ldots\ldots\ldots\ldots
\end{aligned}$$

Cela revient à construire la fonction (fig. 17)

$$z' = y_{n-1} + y_n + y_{n+1}. \tag{3}$$

mais à envisager *seulement* les valeurs entières de x ; à la fonction z' correspond la courbe Γ_ξ, de même qu'à y correspond la courbe Γ_x ; nous spécifions : ξ, parce que en écrivant

$$z_n = y_{3n-1} + y_{3n} + y_{3n+1}$$

nous faisons correspondre l'indice n de z à l'indice $3n$ de y, ce qui introduit deux unités d'abscisses, les x et les ξ qui sont triples des x.

113. Pour déterminer la mode ou abscisse du maximum, nous remplaçons (nº 28, formule III où le terme en x^3 est négligé) la fonction

$$y = \frac{m\,!}{(mp - x)\,!\,(mq + x)\,!}\, p^{mp-x} q^{mq+x}$$

par la fonction

$$(y) = \frac{1}{\sqrt{2\pi mpq}}\, e^{\frac{q-p}{2mpq}x - \frac{x^2}{2mpq}}$$

et c'est l'abscisse $\frac{q-p}{2}$ du maximum de (y) que nous prenons pour valeur (valeur approchée) du maximum de y (nº 49).

A z' substituons de même

$$(z') = \mathrm{A}e^{\alpha(x-1)+\beta(x-1)^2} + \mathrm{A}e^{\alpha x+\beta x^2} + \mathrm{A}e^{\alpha(x+1)^2} + \mathrm{A}e^{\alpha(x+1)+\beta(x+1)^2}$$

où

$$\mathrm{A} = \frac{1}{\sqrt{2\pi mpq}} \qquad \alpha = \frac{q-p}{2mpq} \qquad \beta = -\frac{1}{2mpq}\,;$$

la valeur de x qui rendra (z') maximum s'obtiendra en écrivant

$$[\alpha + 2\beta(x-1)]\, e^{\alpha(x-1)+\beta(x-1)^2} + (\alpha + 2\beta x)e^{\alpha x+\beta x^2}$$
$$+ [\alpha + 2\beta x + 2\beta]^{\alpha(x+1)+\beta(x+1)^2} = 0.$$

Cette équation peut s'écrire

$$(\alpha + 2\beta x - 2\beta)e^{\alpha(x-1)+\beta(x-1)^2-\alpha x-\beta x^2}$$
$$+ (\alpha + 2\beta x) + (\alpha + 2\beta x + 2\beta)e^{\alpha(x+1)+\beta(x+1)^2-\alpha x-\beta x^2} = 0$$
$$(\alpha + 2\beta x)(1 + e^{-\alpha-2\beta x+\beta} + e^{\alpha+2\beta x+\beta})$$
$$-2\beta(e^{-\alpha-2\beta x+\beta} - e^{\alpha+2\beta x+\beta}) = 0\,;$$

on voit qu'elle admet la solution

$$\alpha + 2\beta x = 0,$$
$$\frac{q-p}{2mpq} - \frac{2x}{2mpq} = 0,$$
$$x = \frac{q-p}{2}\,:$$

au degré d'approximation où nous sommes, l'abscisse OH *du maximum de* Γ_ξ *ou* z *est donc la même que l'abscisse* OK *du maximum de* Γ_x *ou* y (fig. 17).

Mais, sous certaines conditions qui ont été précisées à l'aide de l'écart probable, Γ_ξ ou z' peut être représentée par

$$z = \mathrm{A}\,\frac{m'\,!}{(m'p' - \xi)\,!\,(m'q' + \xi)\,!}\,p^{m'p'-\xi}q^{m'q'+\xi}$$

et l'abscisse du maximum est ici

$$\omega\mathrm{H} = \xi = \frac{q'-p'}{2}\,;$$

on a par ailleurs

$$\omega\mathrm{H}_{\text{mesuré en }\xi} = \frac{1}{3}\,\mathrm{OK}_{\text{mesuré en }x}\,;$$

donc

$$\frac{\mathrm{OK}}{3} = \frac{q'-p'}{2}\,;$$

et, comme

$$\mathrm{OK} = \frac{q-p}{2},$$

il en résulte

$$\frac{1}{3}\,\frac{q-p}{2} = \frac{q'-p'}{2}$$

ou

$$q'-p' = \frac{q-p}{3}.$$

114. Le raisonnement valable pour le groupement 5 à 5, car ici nous avons à considérer

$$\begin{aligned}&[\alpha+2\beta(x-2)]e^{\alpha(x-2)+\beta(x-2)^2}+[\alpha+2\beta(x-1)]e^{\alpha(x-1)+\beta(x-1)^2}\\&\quad+[\alpha+2\beta x]e^{\alpha x+\beta x^2}+(\alpha+2\beta(x+1))]e^{\alpha(x+1)+\beta(x+1)^2}\\&\quad+[\alpha+2\beta(x+2)]e^{\alpha(x+2)+\beta(x+2)^2}=0,\end{aligned}$$

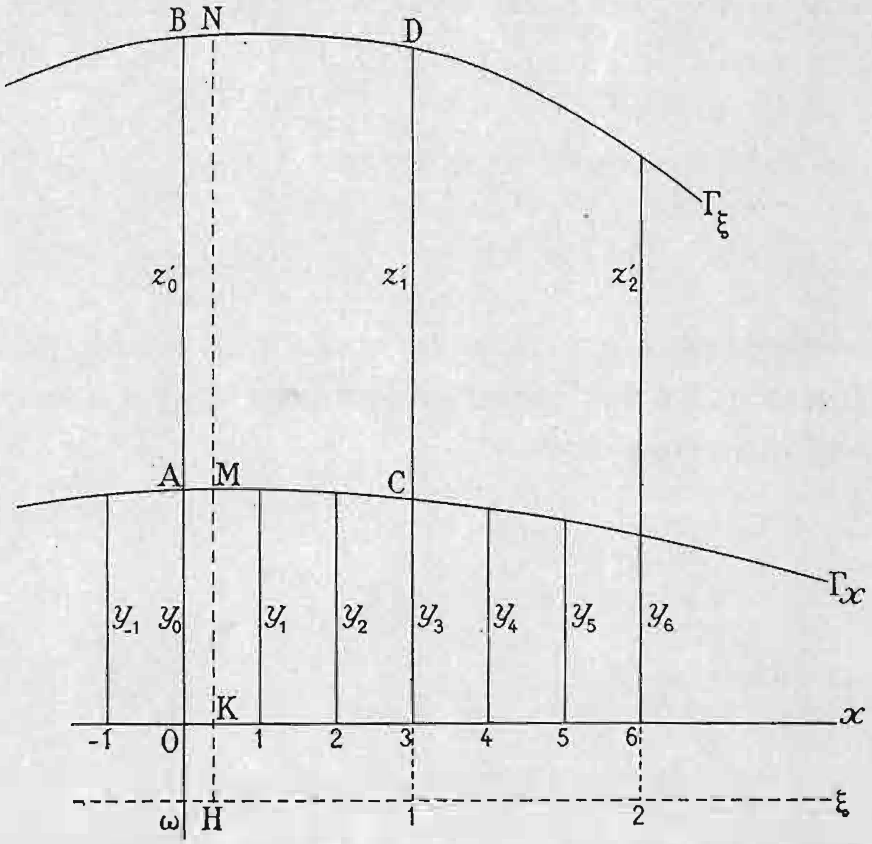
que nous écrivons

$$\begin{aligned}&[\alpha+2\beta(x-2)]e^{-2\alpha-4\beta x+4\beta}+[\alpha+2\beta(x-1)]e^{-\alpha-2\beta x+\beta}\\&\quad+[\alpha+2\beta x]+[\alpha+2\beta(x+1)]e^{\alpha+2\beta x+\beta}\\&\quad+[\alpha+2\beta(x+2)]e^{2\alpha+4\beta x+4\beta}=0\,;\end{aligned}$$

quand $\alpha + 2\beta x$ est nul, les trois termes du milieu disparaissent, on l'a vu, et il reste

$$-4\beta e^{4\beta} + 4\beta e^{4\beta}$$

qui est nul.



En réalité $\omega\,\xi$ est placé sur $0x$; on l'a placé sur une droite distincte de $0x$ pour faciliter la compréhension du texte.

Fig. 17. — Quand on groupe 2 à 2, 3 à 3, n à n les termes d'une suite (m, p, q) soumise à la loi de probabilité simple, on obtient, *sous certaines conditions*, une nouvelle suite soumise également à la loi de probabilité simple, *mais* les *constantes* m', p', q' de *la nouvelle suite diffèrent de* m, p, q : c'est un fait *capital*.

Le raisonnement est valable de même pour tout groupement symétrique $2n + 1$ à $2n + 1$:

$$\begin{array}{c} \cdots\cdots \\ y_{-n} + y_{-n+1} + \cdots + y_{-1} + y_0 + y_1 + \cdots + y_n \,; \\ \cdots\cdots \end{array}$$

Ici

$$q'_{2n+1} - p'_{2n+1} = \frac{q-p}{2n+1}.$$

115. Passons aux groupements dissymétriques, tel que

$$\begin{array}{c} \cdot\cdot\cdot\cdot\cdot \\ y_0 + y_1 + y_2 \\ \cdot\cdot\cdot\cdot\cdot \end{array}$$

Les courbes Γ_ξ et Γ_x sont *les mêmes* que tout à l'heure ; mais au lieu des points N et D, nous considérons les points N_1 et D_1 ; le maximum a encore pour abscisses ω H et OK ; on a donc encore

$$q'' - p'' = \frac{q' - p'}{3}$$

et, pour un groupement $2n + 1$ à $2n + 1$:

$$q''_{2n+1} - p''_{2n+1} = \frac{q' - p'}{2n + 1}.$$

116. *Groupements* 2 *à* 2, $2n$ *à* $2n$. — Le plus simple de ceux-ci est le groupement 2 à 2 et parmi les groupements 2 à 2 nous considérons celui qui résulte de :

$$\frac{mp\,!}{\left(mp + \frac{1}{2} - x\right)!\left(mq + \frac{1}{2} + x\right)!}\, p^{mp + \frac{1}{2} - x}\, q^{mq + \frac{1}{2} + x},$$

qui a pour valeur approchée,
où l'on a

$$\mathrm{A}e^{\frac{q-p}{2mpq}\left(x - \frac{1}{2}\right) + \frac{1}{2mpq}\left(x - \frac{1}{2}\right)^2}$$

Le groupement 2 à 2 s'obtiendra en donnant à x deux valeurs consécutives

$$x \quad \text{et} \quad x + 1$$

et nous aurons ainsi le terme général du groupement :

$$\mathrm{A}e^{\frac{q-p}{2mpq}\left(x - \frac{1}{2}\right) - \frac{1}{2mpq}\left(x - \frac{1}{2}\right)^2} + \mathrm{A}e^{\frac{q-p}{2mpq}\left(x + \frac{1}{2}\right) - \frac{1}{2mpq}\left(x + \frac{1}{2}\right)^2}.$$

Nous avons donc à chercher une racine de l'équation

$$\left[\alpha + 2\beta\left(x - \frac{1}{2}\right)\right]e^{\alpha\left(x - \frac{1}{2}\right) + \beta\left(x - \frac{1}{2}\right)^2}$$
$$+ \left[\alpha + 2\beta\left(x + \frac{1}{2}\right)\right]e^{\alpha\left(x + \frac{1}{2}\right) + \beta\left(x + \frac{1}{2}\right)^2} = 0$$

où l'on a

$$\alpha = \frac{q - p}{2mpq}, \qquad \beta = -\frac{1}{2mpq};$$

cette équation est encore vérifiée par

$$\alpha + 2\beta x = 0\,;$$

en effet,

$$\alpha\left(x - \frac{1}{2}\right) + \beta\left(x - \frac{1}{2}\right)^2 = \alpha x - \frac{1}{2}\alpha + \beta x^2 - \beta x + \frac{\beta}{4},$$

$$\alpha\left(x + \frac{1}{2}\right) + \beta\left(x + \frac{1}{2}\right)^2 = \alpha x + \frac{1}{2}\alpha + \beta x^2 + \beta x + \frac{\beta}{4}\,;$$

nous pouvons diviser l'équation par

$$e^{\alpha x + \beta x^2 + \frac{\beta}{4}}$$

et il reste

$$(\alpha + 2\beta x - \beta)e^{-\frac{1}{2}\alpha - \beta x} + (\alpha + 2\beta x + \beta)e^{\frac{1}{2}\alpha + \beta x} = 0\,;$$

si

$$\alpha + 2\beta x = 0,$$

on a bien

$$-\beta e^0 + \beta e^0 = 0.$$

De là on passera aux groupements $2n$ à $2n$ symétriques, et des groupements $2n$ à $2n$ symétriques, on passera aux groupements $2n$ à $2n$ dissymétriques.

117. D'une manière générale, mais *approximativement*, quand on groupe n à n, si p', q' sont les valeurs de p, q applicables au groupement, on a

$$q' - p' = \frac{q - p}{n}\,;$$

puisque

$$q' + p' = 1,$$

il en résulte

$$p' = \frac{1}{2} - \frac{q - p}{n}$$

$$q' = \frac{1}{2} + \frac{q - p}{n}.$$

Applications. — Pour z' et z'', on a par ces formules

$$p' = p'' = 0{,}3667 \qquad q' = q'' = 0{,}6333\,;$$

les valeurs vraies sont, avons-nous vu,

$$\begin{aligned} p' &= 0{,}3678 \qquad & q' &= 0{,}6362 \\ p'' &= 0{,}3686 \qquad & q'' &= 0{,}6314\,; \end{aligned}$$

nous les obtenons donc, pour la formule

$$\frac{q-p}{n},$$

à 1 ou 2 millièmes près.

Pour z''', la même formule donne

$$p''' = 0{,}4200 \qquad q''' = 0{,}5800$$

au lieu de

$$p''' = 0{,}4210 \qquad q''' = 0{,}5790 :$$

nous les obtenons encore à 0,001 près.

Il est très important de faire ressortir que p et q ne sont pas conservés, loin de là, quand on groupe les termes d'une suite qui procède suivant la loi de probabilité simple.

118. Il y a encore lieu de remarquer que *la formule*

$$q' - p' = \frac{q-p}{n}$$

n'est réversible que dans des limites étroites. Ecrivons en effet

$$q - p = n(q' - p') \qquad \text{avec} \qquad q + p = 1\,;$$

on a

$$p = \frac{1}{2} - n\,\frac{q'-p'}{2} \qquad q = \frac{1}{2} + n\,\frac{q'-p'}{2}.$$

et on aurait

$$q > 1$$

si

$$n\,\frac{q'-p'}{2} > \frac{1}{2}, \qquad n(q'-p') > 1, \qquad n > \frac{1}{q'-p'}\,;$$

par exemple si

$$q' = 0{,}9 \qquad p' = 0{,}1,$$

on a

$$q' - p' = 0,8 \qquad \frac{1}{q' - p'} = \frac{1}{0,8} = 1,25$$

et la formule *renversée* ne s'applique plus si

$$n > 1,25\,;$$

par exemple si $q' = p'$, on aurait

$$q = \frac{1}{2} + 2 \times \frac{0,8}{2} = 1,3$$

$$p = \frac{1}{2} - 2 + \frac{0,8}{2} = -0,3.$$

NOTE

TABLE DES VALEURS DE LA FONCTION (1)

$$y_x = \frac{m\,!}{(mp - x)\,!\,(mq + x)\,!}\, p^{mp-x} q^{mq+x}$$

calculée par M. F. J. Duarte

pour $m = 1\,000$ $p = 0{,}1$ $q = 0{,}9$

x	y	x	y	x	y	x	y
0	0,042 016 791			— 40	0,000 012 0864	+ 40	0,000 001 9202 8
— 1	0,041 600 783	+ 1	0,041 970 157	41	0,000 008 1909	41	0,000 001 1019 7
2	0,040 739 764	2	0,041 458 326	42	0,000 005 5055	42	0,000 000 6211 8
3	0,039 46[illegible] 274	3	0,040 494 178	43	0,000 003 6367	43	0,000 000 3438 5
4	0,037 820 887	4	0,039 105 551	44	0,000 002 4270	44	0,000 000 1868 6
5	0,035 859 804	5	0,037 333 918	45	0,000 001 5920	45	0,000 000 0996 59
6	0,033 642 059	6	0,035 232 340	46	0,000 001 0359	46	0,000 000 0521 47
7	0,031 231 569	7	0,032 862 800	47	0,000 000 6686 6	47	0,000 000 0267 62
8	0,028 693 201	8	0,030 293 132	48	0,000 000 4282 1	48	0,000 000 0134 66
9	0,026 090 046	9	0,027 593 744	49	0,000 000 2720 6	49	0,000 000 0066 406
— 10	0,023 481 415	+ 10	0,024 834 370	— 50	0,000 000 1715 0	+ 50	0,000 000 0032 084
11	0,020 919 046	11	0,022 081 531	51	0,000 000 1072 7	51	0,000 000 0015 182
12	0,018 449 436	12	0,019 393 557	52	0,000 000 0655 71	52	0,000 000 0007 0329
13	0,016 109 242	13	0,016 852 695	53	0,000 000 0409 97	53	0,000 000 0003 1880
14	0,013 926 801	14	0,014 412 105	54	0,000 000 0250 53	54	0,000 000 0001 4135

15	0,011 921 880	15	0,012 191 224	55	0,000 000 0151 94	55	0,000 000 0000 0127 9
16	0,010 106 191	16	0,010 182 535	56	0,000 000 0091 444	56	0,000 000 0000 2596 0
17	0,008 484 210	17	0,008 393 937	57	0,000 000 0054 620	57	0,000 000 0000 1074 4
18	0,007 054 197	18	0,006 830 361	58	0,000 000 0032 380	58	0,000 000 0000 0433 95
19	0,005 809 339	19	0,005 485 099	59	0,000 000 0019 052	59	0,000 000 0000 0171 04
— 20	0,004 738 914	+ 20	0,004 346 345	— 60	0,000 000 0011 127	+ 60	0,000 000 0000 0052 223
21	0,003 829 426	21	0,003 397 794	61	0,000 000 0006 4505		
22	0,003 065 633	22	0,002 620 208	62	0,000 000 0003 7120		
23	0,002 431 459	23	0,001 992 834	63	0,000 000 0002 1204		
24	0,001 910 744	24	0,001 494 626	64	0,000 000 0001 2024		
25	0,001 487 835	25	0,001 105 215	65	0,000 000 0000 6769 1		
26	0,001 148 019	26	0,000 805 6373	66	0,000 000 0000 3783 2		
27	0,000 877 8378	27	0,000 578 8074	67	0,000 000 0000 2099 3		
28	0,090 655 2365	28	0,000 409 7866	68	0,000 000 0000 1156 3		
29	0,000 499 6446	29	0,000 285 8312	69	0,000 000 0000 0632 6		
— 30	0,000 371 9569	+ 30	0,000 196 3942	— 70	0,000 000 0000 0343 6		
31	0,000 274 4720	31	0,000 132 8983	71	0,000 000 0000 0185 3		
32	0,000 200 7712	32	0,000 088 5514	72	0,000 000 0000 0099 2		
33	0,000 145 5884	33	0,000 058 0851	73	0,000 000 0000 0052 3		
34	0,000 104 6643	34	0,000 037 5004				
35	0,000 074 6003	35	0,000 023 8238				
36	0,000 052 7200	36	0,000 014 8898				
37	0,000 036 9425	37	0,000 009 1532 0				
38	0,000 025 6694	38	0,000 005 5329 7				
39	0,000 017 6874	35	0,000 003 2879 3				

(1) Le lecteur trouvera des Tables analogues pour $m = 1.000$ et

$p = 0{,}2$; $q = 0{,}8$; $p = 0{,}25$; $q = 0{,}75$; $p = 0{,}3$; $q = 0{,}7$; $p = 0{,}4$; $q = 0{,}6$; $p = 0{,}5$; $q = 0{,}5$

dans *Mémorial de l'Office National Météorologique de France*, n° 10, 1926. Chiron à Paris et Office Nat. Mét. de France, 176 rue de l'Université, Paris 7°

TABLE DES MATIÈRES

Pages

NOTE

TABLE DES FIGURES

Saint-Amand (Cher). — Imprimerie R. BUSSIÈRE — 7-11-1930