STATISTIQUES ET PROBABILITÉS

Autor(en): **de Montessus de Ballore, R.**

Objekttyp: **Article**

Zeitschrift: **L'Enseignement Mathématique**

Band (Jahr): **27 (1928)**

Heft 1: **L'ENSEIGNEMENT MATHÉMATIQUE**

PDF erstellt am: **21.07.2024**

Persistenter Link: https://doi.org/10.5169/seals-21869

On le montre par le procédé utilisé par M. Montel en s'appuyant sur ce que les fonctions limites de la suite sont toutes égales. De la même façon on établit la proposition suivante, qui complète dans un autre sens le théorème de Stieltjes:

VI. *Si les fonctions* $f(x; n)$ *sont également quasi-analytiques d'une classe* A_p *dans un intervalle et si* x_0 *appartenant à cet intervalle, la suite* $f^{(q)}(x_0; n)$ *converge pour chaque valeur de* q, *les fonctions* $f(x; n)$ *convergent uniformément à l'intérieur de l'intervalle.*

STATISTIQUES ET PROBABILITÉS

PAR

R. DE MONTESSUS DE BALLORE (Paris).

J'appellerai « Fonction de probabilité simple » la fonction

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

Elle se présente dès le début du Calcul des Probabilités.

On la retrouve dans la Théorie des Erreurs accidentelles d'observation, dans l'étude des statistiques simples de toute nature: démographie, météorologie, histoire naturelle, applications de la chimie, certains phénomènes physiques, etc.

La représentation *approchée* de la fonction de probabilité simple (1) par l'exponentielle

$$e^{-h^2x^2}$$

et par les variétés d'exponentielles, a donné beaucoup de mécomptes, il faut avoir le courage de le reconnaître, en dehors de l'étude des erreurs accidentelles d'observation.

Je vais montrer qu'on peut étudier *directement* la fonction (1) sans recourir à aucune représentation approchée, exponentielle ou autre, et que l'étude directe donne des renseignements d'un grand intérêt sur les statistiques [1].

I

1. — Tout d'abord, la formule de Stirling permet d'écrire, en logarithmes décimaux:

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

$$+ (mq+x)\log\frac{mq}{mq+x} + \frac{M}{12m} - \frac{M}{12}\times\frac{m}{(mp-x)(mq+x)} ;$$

$$-\log\sqrt{2\pi} = \bar{1},6009101 ; \qquad M = \log e = 0,4342945 . \qquad (2)$$

Cette formule donne y_x avec 5 décimales exactes, quand on prend les logarithmes avec 7 décimales.

On peut donc calculer, avec une approximation dépassant les besoins de toutes les statistiques, les valeurs numériques de y_x pour toutes les valeurs possibles de m ($m>0$, entier ou non entier), de p ($0<p<1$), de q ($p+q=1$), de x ($mp-x$, $mq+x$ entiers ou fractionnaires, mais $mp-x \geq 0$, $mq+x \geq 0$).

Quand m, $mp-x$, $mq+x$ sont entiers, on peut aussi calculer y_x par les Tables de factorielles [2]: les deux procédés donnent les mêmes résultats.

2. — Soit

$$y_x = \frac{1000\,!}{(100-x)\,!\,(900+x)\,!}\,0,1^{100-x} \times 0,9^{900+x} ,$$

c'est la formule (1) pour $m = 1000$; $p = 0,1$; $q = 0,9$. Dans le problème de 1000 tirages de boules d'une urne contenant 9 fois

---

[1] Le lecteur voudra bien se reporter pour les démonstrations des formules qui vont être données aux *Annales de la Société scientifique de Bruxelles*, 1927 et 1928.

[2] F. J. Duarte, Tables de log $n!$ à 33 décimales, pour toutes les valeurs entières de n, de $n = 1$ à $n = 3000$. Imprimerie A. Kundig à Genève, 1927, 30 fr. fr.

plus de boules rouges que de boules noires, on sera amené à calculer[1]

$$y_0 = 0{,}042\,017 \ ; \quad y_1 = 0{,}041\,970 \ ; \quad y_2 = 0{,}041\,458 \ ; \quad y_3 = 0{,}040\,494 \ldots$$
$$y_{-1} = 0{,}041\,601 \ ; \quad y_{-2} = 0{,}040\,740 \ ; \quad y_{-3} = 0{,}039\,466 \ldots$$

Mais, *perdant de vue ce problème de boules tirées d'une urne*, on peut de façon semblable, calculer (form. 1) les valeurs de y pour

$$x = h \ , \quad x = 1 + h \ ; \quad x = 2 + h \ , \ldots , \quad x = -1 + h \ ,$$
$$x = -2 + h \ , \ldots$$

h étant une constante numérique quelconque: la formule (2) le permet.

Aussi bien, mettant la constante h en évidence, *ce n'est plus la fonction* (1) *que nous étudierons, c'est la fonction plus générale.*

$$y_{x+h} = \frac{m\,!}{(mp - x - h)\,!\,(mq + x + h)\,!}\, p^{mp-x-h}\, q^{mq+x+h} \qquad (3)$$

où:

m est un nombre positif quelconque, entier ou fractionnaire,

p, q sont des nombres compris entre 0 et 1, dont la somme est 1;

h est un nombre positif ou négatif, ordinairement fractionnaire, voisin dans les applications, de 0, 1, —1;

x est un nombre *entier*, positif ou négatif: il n'y a pas lieu, dans l'étude des statistiques, de considérer des valeurs fractionnaires de x; les valeurs de x doivent vérifier les inégalités

$$mp - x - h \geqq 0 \ , \qquad mq + x + h \geqq 0 \ .$$

3. — On calcule les valeurs de y_{x+h} par la formule suivante, qui n'est autre que la formule (2) légèrement modifiée

$$\log y_{x+h} = -\log\sqrt{2\pi} + \frac{1}{2}\log\frac{m}{(mp - x - h)(mq + x + h)}$$
$$+ (mp - x - h)\log\frac{mp}{mp - x - h} + (mq + x + h)\log\frac{mq}{mq + x + h}$$
$$+ \frac{\mathrm{M}}{12m} - \frac{\mathrm{M}}{12} \times \frac{m}{(mp - x - h)(mq + x + h)} \ ; \qquad (4)$$

$$-\log\sqrt{2\pi} = \bar{1}{,}6009101 \ ; \qquad \mathrm{M} = \log e = 0{,}434\,2945 \quad \text{(log. décimaux)}$$

[1] Le lecteur trouvera quantités de tableaux de valeurs numériques de y_x dans: *Annales Soc. Sc. Brux.*, 1926; *Mém.* N° 10 de *l'Office nat. météor. de France; Calcul des probabilités et statistiques*, par R. de Montessus de Ballore, Chiron, Paris, 1926.

Mais quand on a plusieurs valeurs de y_{x+h} à calculer, il est plus simple, plus rapide, de calculer seulement y_h par la formule

$$\log y_h = -\log\sqrt{2\pi} + \frac{1}{2}\log\frac{m}{(mp-h)(mq+h)}$$

$$+ (mp-h)\log\frac{mp}{mp-h} + (mq+h)\log\frac{mq}{mq+h}$$

$$+ \frac{\mathrm{M}}{12m} - \frac{\mathrm{M}}{12}\times\frac{m}{(mp-h)(mq+h)}\ ; \qquad -\log\sqrt{2\pi} = \bar{1},600\,9101\ ;$$

$$\mathrm{M} = \log e = 0,434\,2945 \quad (\text{log. décimaux})\ , \tag{5}$$

et de calculer *ensuite* les autres valeurs de y_h par les formules de récurrence

$$y_{-1+h} = \frac{mq+h}{mp-h+1}\frac{p}{q}y_h\ , \qquad y_{-2+h} = \frac{mq+h-1}{mp-h+2}\frac{p}{q}y_{-1+h}\ ,$$

$$y_{-3+h} = \frac{mq+h-2}{mp-h+3}\frac{p}{q}y_{-2+h}\ , \ \ldots \tag{6}$$

$$y_{1+h} = \frac{mp-h}{mq+h+1}\frac{q}{p}y_h\ , \qquad y_{2+h} = \frac{mp-h-1}{mq+h+2}\frac{q}{p}y_{1+h}\ ,$$

$$y_{3+h} = \frac{mp+h-2}{mq+h+3}\frac{q}{p}y_{2+h}\ , \ \ldots \tag{7}$$

Les formules générales sont

$$y_{-x+h} = \frac{mq+h-(x-1)}{mp-h+x}\frac{p}{q}y_{-x+1+h}\ , \tag{6'}$$

$$y_{x+h} = \frac{mp-h-(x-1)}{mq+h+x}\frac{p}{q}y_{x-1+h}\ . \tag{7'}$$

4. — Ce ne sont pas précisément les nombres y_x de la formule (3) que nous allons considérer, mais les nombres

$$\mathrm{Y}_{x+h} = \mathrm{A}\frac{m\,!}{(mp-x-h)\,!\,(mq+x+h)}\,p^{mp-x-h}\,q^{mq+x+h} = \mathrm{A}y_{x+h}$$

$$(\text{A constante positive quelconque}) \tag{8}$$

les nombres Y_{x+h} obéissent, comme les nombres y_x, aux lois de récurrence (6) et (7).

II

5. — Nous poserons

$$\left\{\begin{aligned} s_0' &= y_{1+h} + y_{2+h} + y_{3+h} + \dots + y_{n+h} \\ s_0'' &= y_{-1+h} + y_{-2+h} + y_{-3+h} + \dots + y_{-n'+h} \end{aligned}\right. \qquad (9)$$

$$\left\{\begin{aligned} s_1' &= y_{1+h} + 2y_{2+h} + 3y_{3+h} + \dots + ny_{n+h} \\ s_1'' &= y_{-1+h} + 2y_{-2+h} + 3y_{-3+h} + \dots + n'y_{-n'+h} \end{aligned}\right. \qquad (10)$$

$$\left\{\begin{aligned} s_2' &= y_{1+h} + 2^2y_{2+h} + 3^2y_{3+h} + \dots + n^2y_{n+h} \\ s_2'' &= y_{-1+h} + 2^2y_{-2+h} + 3^2y_{-3+h} + \dots + n'^2y_{-n'+h} \end{aligned}\right. \qquad (11)$$

$$\left\{\begin{aligned} S_0' &= Y_{1+h} + Y_{2+h} + Y_{3+h} + \dots + Y_{n+h} \\ S_0'' &= Y_{-1+h} + Y_{-2+h} + Y_{-3+h} + \dots + Y_{-n'+h} \end{aligned}\right. \qquad (12)$$

$$\left\{\begin{aligned} S_1' &= Y_{1+h} + 2Y_{2+h} + 3Y_{3+h} + \dots + nY_{n+h} \\ S_1'' &= Y_{-1+h} + 2Y_{-2+h} + 3Y_{-3+h} + \dots + n'Y_{-n'+h} \end{aligned}\right. \qquad (13)$$

$$\left\{\begin{aligned} S_2' &= Y_{1+h} + 2^2Y_{2+h} + 3^2Y_{3+h} + \dots + n^2Y_{n+h} \\ S_2'' &= Y_{-1+h} + 2^2Y_{-2+h} + 3^2Y_{-3+h} + \dots + n^2Y_{-n'+h} \end{aligned}\right. \qquad (14)$$

$$\lambda = mpq + ph \quad (15) \qquad \mu = mpq - qh \quad (16)$$

Si l'on donne à n, n' des valeurs soumises aux conditions déjà énoncées au N° 2:

$$mp - n - h \geqq 0 \,, \qquad mq + n + h \geqq 0 \,, \qquad mp - n' + h \geqq 0 \,,$$
$$mq + n' + h \geqq 0 \,,$$

on a les formules EXACTES suivantes, *exactes quels que soient* n, n':

$$h = \lambda - \mu \qquad (17)$$

$$\lambda s_0' - \mu(s_0' + y_h - y_{n+h}) + s_1' - qny_{n+h} = 0 \qquad (18)$$

$$-\lambda(s_0'' + y_h - y_{-n'+h}) + \mu s_0'' + s_1'' - pn'y_{-n'+h} = 0 \qquad (19)$$

$$\lambda(s_0' - s_1') + \mu(s_1' - ny_{n+h}) + ps_1' - s_2' + qn^2y_{n+h} = 0 \qquad (20)$$

$$\lambda(s_1'' - n'y_{-n'+h}) + \mu(s_0'' - s_1'') + qs_1'' - s_2'' + pn'^2y_{-n'+h} = 0 \,. \qquad (21)$$

Les formules (18) à (21) sont nouvelles. Voici comment on les obtient.

Ecrivons comme il suit les formules (6'), (7'):

$$(\mu + qx) y_{-x+h} + [(x-1)p - \lambda] y_{-x+1-h} = 0 \ .$$

$$(\lambda + px) y_{x+h} + [(x-1)q - \mu] y_{x-1+h} = 0 \ ; \quad (6''), (7'')$$

faisons $x = 1, 2, 3, \dots n$ dans (7″) et ajoutons les résultats obtenus: après réductions, on trouve la formule (18).

On déduit la formule (19) de la formule (18) en comparant (6″) et (7″): il suffit de changer dans (18), les s' en s'', p en q, n en n', λ en μ.

Pour obtenir (20), on multiplie (7″) par $x - 1$ et on écrit le résultat

$$[px^2 + (\lambda - p)x - \lambda] y_{x+h} + [(x-1)^2 q - \mu(x-1)] y_{x-1+h} = 0 \ .$$

Dans cette formule, on fait encore $x = 1, 2, 3, \dots n$, on ajoute les résultats et cela donne la formule (20). De (20), on déduit (21), comme de (18) on a déduit (19).

Les formules (18) à (21) *généralisent* d'autres formules plus simples, relatives au cas de $h = 0$, que j'ai données peu après M. Ragnar FRISCH, dont j'ignorais les travaux [1].

Observons que les formules relatives au cas de $h = 0$ ne nous seraient ici d'aucune utilité.

Les formules (18) *à* (21) *se simplifient beaucoup si* n, n' *sont assez grands pour qu'on puisse négliger les termes en* y_{n+h}, y_{-n+h}, *ce qui est toujours le cas dans les statistiques que nous visons.*

Elles deviennent, en effet, dans cette hypothèse

$$-\lambda s'_0 + \mu(s'_0 + y_h) = s'_1 \qquad (22)$$

$$\lambda(s''_0 + y_h) - \mu s''_0 = s''_1 \qquad (23)$$

$$\lambda(s'_0 - s'_1) + \mu s'_1 = s'_2 - p s'_1 \qquad (24)$$

$$\lambda s''_1 + \mu(s''_0 - s''_1) = s''_2 - q s''_1 \ . \qquad (25)$$

---

[1] D. MIRIMANOFF, « A propos d'une formule de M. de Montessus de Ballore », *Enseignement Mathématique*, XXVIe année, 1927. M. Mirimanoff propose d'appeler les formules relatives au cas de $h = 0$: formules de Frisch-Montessus de Ballore. J'ai donné les formules relatives au cas de $h = 0$, dans: *Ann. Soc. Scient. de Bruxelles*, 1927: « La Formule fondamentale de la Statistique ».

Nous leur adjoindrons la formule

$$y_h + s_0' + s_0'' = 1 \tag{26}$$

qui, elle aussi, est approchée, et dont l'approximation est de même ordre que celle des formules (22) à (25).

6. — Les formules (15), (16), (17), (22)-(26) vont nous fournir l'instrument de calcul *nécessaire et suffisant* pour l'étude des questions qui suivront.

En les transformant algébriquement, en abandonnant les sommes s pour les sommes S, qui se présentent (et non pas les s) dans les calculs — les calculs de statistique — en posant

$$\left\{\begin{aligned} A &= S_0' S_1'' + S_0'' S_1' \\ C &= S_2' S_1'' + S_2'' S_1' - S_1' S_1'' \\ S &= S_0'' + S_0' + Y_h \end{aligned}\right. \tag{27}$$

on déduit des formules (17), (22) à (26):

$$h = \frac{S_1'' - S_1'}{S} \tag{28}$$

$$p = \frac{S_2'}{S_1'} - \frac{C}{A}\frac{S_0'}{S_1'} + h\left(1 - \frac{S_0' S_0''}{A}\right) \tag{29}$$

$$q = \frac{S_2''}{S_1''} - \frac{C}{A}\frac{S_0''}{S_1''} - h\left(1 - \frac{S_0' S_0''}{A}\right) \tag{30}$$

$$\lambda = \frac{C}{A} + h\frac{S_0'' S_1''}{A} \tag{31}$$

$$\mu = \frac{C}{A} - h\frac{S_0' S_1''}{A}. \tag{32}$$

III

7. — Le problème le plus simple posé par les statisticiens, *et que nous allons résoudre*, est celui-ci:

Des nombres donnés par une statistique

$$\mathrm{Y}_{-n'+h} < \mathrm{Y}_{-n'+1+h} < \mathrm{Y}_{-n'+2+h} < \cdots$$
$$< \mathrm{Y}_{-1+h} < \mathrm{Y}_h > \mathrm{Y}_{1+h} > \mathrm{Y}_{2+h} > \cdots > \mathrm{Y}_{n+h} \,.$$

où $\mathrm{Y}_{-n'+h}$, Y_{n+h} sont très voisins de zéro, suivent *à peu près*, avec de petits accidents locaux, la loi (1) ou (3), ce que l'on constate par un graphique.

On propose de calculer, SI CELA EST POSSIBLE, *les éléments* m, p, q, h *d'une courbe* (3) *dont les ordonnées restituent*, *dans l'ensemble*, les éléments Y.

Les formules (dans l'ordre de leur application):

(12) à (14), (27) à (32), (15) ou (16) qui donnent m (5) à (7) résolvent le problème.

Il n'est même pas nécessaire de faire appel à la formule (5): en effet, on pose *provisoirement* dans les formules (6) et (7)

$$y_h = 1 \,,$$

on calcule

$$y_{-1+h}\,, \quad y_{-2+h}\,, \quad y_{-3+h}\,, \quad y_{-n'+h}\,, \cdots,$$
$$y_{1+h}\,, \quad y_{2+h}\,, \quad y_{3+h}\,, \cdots, y_{n+h} \tag{33}$$

par les formules (6), (7) en partant de $y_h = 1$, on fait la somme Σ de ces nombres (33) et de y_h, autrement dit, on ajoute *un* à la somme des nombres (33); les Y *calculés* s'obtiennent en multipliant les y par S et en divisant les produits par Σ.

Puisque le côté pratique a tant d'importance pour les statisticiens, disons que le calcul de bout en bout ne demande pas plus de 2 à 4 heures.

IV. Exemples de calcul.

7. — Hauteurs de taille aux Etats-Unis (d'après Quételet)[1].

	Observé	Calculé		Observé	Calculé
y_{-13+h} =	4		y_{1+h} =	3631	3731
y_{-12+h} =	1		y_{2+h} =	3133	3058
y_{-11+h} =	3	0	y_{3+h} =	2075	2189
y_{-10+h} =	7	0	y_{4+h} =	1485	1368
y_{-9+h} =	6	2	y_{5+h} =	680	748
y_{-8+h} =	10	13	y_{6+h} =	343	357
y_{-7+h} =	15	58	y_{7+h} =	181	149
y_{-6+h} =	50	195	y_{8+h} =	42	54
y_{-5+h} =	526	513	y_{9+h} =	9	17
y_{-4+h} =	1237	1093	y_{10+h} =	6	5
y_{-3+h} =	1947	1929	y_{11+h} =	2	1
y_{-2+h} =	3019	3268			
y_{-1+h} =	3475	3634			
y_h =	4050	3958			

Total des nombres observés: 25937,
Total des nombres calculés : 25937, identique au précédent.

Dans le cas actuel, on a trouvé

$$S_0' = 3475 + 3019 + 1947 + \dots + 4 = 10\,300$$

$$S_0'' = 3631 + 3133 + 2075 + \dots + 2 = 11\,587$$

$$S_1' = 3475 \times 1 + 3019 \times 2 + 1947 \times 3 + \dots + 4 \times 13 = 23\,638$$

$$S_1'' = 3631 \times 1 + 3133 \times 2 + 2075 \times 3 + \dots + 2 \times 11 = 29\,286$$

$$S_1' = 3475 \times 1^2 + 3019 \times 2^2 + 1947 \times 3^2 + \dots + 4 \times 13^2 = 73\,560$$

$$S_2'' = 3631 \times 1^2 + 3133 \times 2^2 + 2075 \times 3^2 + \dots + 2 \times 11^2 = 101\,064$$

$$h = \frac{5648}{25937} = 0{,}217758$$

$$A = 575\,539\,306 \qquad C = 385\,0966524 \qquad S = 25\,937$$

$$p = 0{,}36899 \qquad q = 0{,}63101$$

$$\lambda = 6{,}79\,469 \qquad \mu = 6{,}57\,693$$

$$m = 28{,}837\,49 \; .$$

---

[1] Ch. Jordan, *Statistique Mathématique*, Paris, Gauthier-Villars, 1927; p. 209.

Les différences entre les nombres calculés et les nombres observés sont, au point de vue des statisticiens, insignifiantes.

Elles indiquent toutefois qu'il ne s'agit pas d'une race d'hommes parfaitement homogène: l'introduction de l'erreur probable le montre.

8. — Voici un autre exemple de calcul. Il s'agit de la statistique des températures observées à Paris, au Parc Saint-Maur de 1890 à 1899, en tout 87.648 observations horaires, réduites au total 1000 par M. Baldit, ou plutôt au total 999,2.

Les températures, notées au dixième de degré, ont été ramenées aux degrés ronds par M. Baldit.

Pour cette statistique, calculée par M. Duarte [1], à Genève, nous donnerons quelques détails de calcul, qui ne figurent pas dans la statistique précédente.

On a ici

$$S_0' = 50{,}8 + 47{,}8 + 47{,}0 + \dots + 0{,}1 = 461{,}3$$

$$S_0'' = 48{,}7 + 48{,}2 + 47{,}0 + \dots + 0{,}1 = 485{,}0$$

$$S_1' = 50{,}8 \times 1 + 47{,}8 \times 2 + 47{,}0 \times 3 + \dots + 0{,}1 \times 25 = 2948{,}1$$

$$S_1'' = 48{,}7 \times 1 + 48{,}2 \times 2 + 47{,}0 \times 3 + \dots + 0{,}1 \times 25 = 3133{,}8$$

$$S_2' = 50{,}8 \times 1^2 + 47{,}8 \times 2^2 + 47{,}0 \times 3^2 + \dots + 0{,}1 \times 25^2 = 27593{,}5$$

$$S_2'' = 48{,}7 \times 1^2 + 48{,}2 \times 2^2 + 47{,}0 \times 3^2 + \dots + 0{,}1 \times 25^2 = 29312{,}8$$

$$h = \frac{185{,}7}{999{,}2} = 0{,}185849$$

$$A = 2\,875\,450{,}44 \qquad C = 163\,650\,820{,}20 \qquad S = 999{,}2$$

(Noter que S diffère légèrement de 1000, mais cela n'a aucune importance pour nos calculs, qui sont basés sur cette donnée: 999,2.)

$$p = 0{,}625{,}743 \qquad q = 0{,}374\,257$$

$$\lambda = 57{,}005\,522 \qquad \mu = 56{,}819\,673$$

$$m = 242{,}9205\ .$$

[1] Je dois ici remercier M. Duarte, qui depuis plusieurs années a consacré une partie de son temps à effectuer de nombreux calculs numériques et autres, qui m'ont beaucoup aidé à mettre au jour ces nouvelles théories.

A	B	Calculé N_1	Observé N_2	t
0,004 3047	0,227	$Y_{25+h} = 0,2$	$Y_{25+h} = 0,1$	— 15°
0,006 5373	0,345	0,3	0,3	— 14°
0,009 7661	0,516	0,6	0,5	— 13°
0,014 3514	0,758	0,8	0,6	— 12°
0,020 7449	1,095	1,1	1,0	— 11°
0,029 4960	1,557	1,6	1,5	— 10°
0,041 2517	2,178	2,2	2,1	— 9°
0,056 7462	2,996	3,0	3,8	— 8°
0,076 7778	4,053	4,1	3,7	— 7°
0,102 1706	5,394	5,4	5,3	— 6°
0,133 7194	7,059	7,1	7,3	— 5°
0,172 1179	9,086	9,1	7,7	— 4°
0,217 8739	11,502	11,5	10,7	— 3°
0,271 2159	14,317	14,3	13,8	— 2°
0,331 9999	17,526	17,5	19,6	— 1°
0,399 6269	21,096	21,1	anom. 28,0	0°
0,472 9829	24,969	25,0	29,3	+ 1°
0,550 4132	29,056	29,1	30,8	+ 2°
0,629 7418	33,244	33,2	33,5	+ 3°
0,708 3421	37,393	37,4	anom. 33,6	+ 4°
0,783 2596	41,348	41,3	39,5	+ 5°
0,851 3823	44,914	44,9	44,0	+ 6°
0,909 6463	48,020	48,0	47,0	+ 7°
0,955 2589	50,428	50,4	47,8	+ 8°
0,985 9175	52,047	52,0	50,8	+ 9°
1	52,790	$Y_h = 52,8$	$Y_h = 52,9$	+ 10°
0,996 7058	52,616	52,6	anom. 48,7	+ 11°
0,976 1303	51,530	51,5	48,2	+ 12°
0,939 2632	49,584	49,6	47,0	+ 13°
0,887 9099	46,873	46,9	45,2	+ 14°
0,824 5453	43,528	43,5	anom. 48,5	+ 15°
0,752 1156	39,704	39,7	42,2	+ 16°
0,673 8108	35,570	35,6	36,5	+ 17°
0,592 8323	31,296	31,3	31,7	+ 18°
0,512 1791	27,038	27,0	27,2	+ 19°
0,434 4721	22,936	22,9	24,1	+ 20°
0,361 8296	19,101	19,1	19,8	+ 21°
0,295 8004	15,615	15,6	16,5	+ 22°
0,237 3529	12,530	12,5	12,5	+ 23°
0,186 9123	9,867	9,9	10,2	+ 24°
0,144 4353	7,625	7,6	7,9	+ 25°
0,109 5074	5,781	5,8	5,6	+ 26°
0,081 4496	4,300	4,3	4,4	+ 27°
0,059 4220	3,137	3,1	2,8	+ 28°
0,042 5163	2,244	2,2	2,2	+ 29°
0,029 8296	1,575	1,6	1,8	+ 30°
0,020 5188	1,083	1,1	0,9	+ 31°
0,013 8356	0,730	0,7	0,5	+ 32°
0,009 1436	0,483	0,5	0,3	+ 33°
0,005 9214	0,313	0,3	0,2	+ 34°
0,003 7570	0,198	$Y_{-2+h} = 0,2$	$Y_{-25+h} = 0,1$	+ 35°
Totaux Σ = 18,927 8414	999,201		S = 999,2	

anom. = anomalie

Dans la colonne A, figurent les nombres calculés par les formules (6), (7) en partant de $h = 1$. Les nombres B sont les nombres

$$A \times \frac{S}{\Sigma} = A \times \frac{999{,}2}{18{,}9278414} \cdot$$

Les nombres N_1 ont été obtenus en arrondissant les nombres B aux dixièmes.

Les nombres N_2 sont les nombres observés, les *données:* p. e. sur 999,2 observations (au total 87.648 observations effectives, réduites à 999,2) la température 10° a été observée 52,9 fois.

Dans les calculs pratiques, il n'est pas nécessaire de prendre autant de décimales: il suffit ordinairement de calculer p, q, λ, μ, avec trois décimales.

Les nombres N_1 sont discutés plus loin (N° 10-II).

V

9. — Les considérations développées aux paragraphes I-III se justifient d'elles-mêmes, puisque les calculs effectués retrouveraient évidemment h, m, p, q ($h = 0$) si l'on partait de probabilités calculées par la formule (1).

On notera que nous avons tenu compte non seulement des nombres isolés de la statistique étudiée, mais de l'ensemble de ces nombres, et cela est nécessaire, il n'est pas besoin d'insister.

Nous avons d'importantes remarques à faire: les voici.

I. — Pourquoi avons-nous introduit h ?

Parce que les calculs qu'on tenterait de faire en prenant $h = 0$, conduiraient à des équations incompatibles.

Cela s'explique facilement. Reportons-nous au N° 2, où nous avons introduit h, à dessein, dès le début.

Actuellement, nous avons ajusté (N° 8) la statistique des températures au Parc Saint-Maur: *non pas ajusté par une courbe analytique quelconque, mais nous avons ajusté par une fonction de probabilité simple, de la forme* (3), CE QUI EST CAPITAL.

Ces températures, relevées chacune au dixième de degré, sont groupées par degrés *ronds*

$$0°\ ,\quad 1°\ ,\quad 2°\ ,\quad 3°\ ,\ \ldots\quad -1°\ ,\quad -2°\ ,\ \ldots$$

Nous aurions eu, d'après les calculs précédents,

$$h = 0$$

si les températures avaient été groupées suivant l'échelle

$$\begin{array}{lll} -\ 0^\circ,18585\ , & -1^\circ - 0^\circ,18585\ , & -2^\circ - 0^\circ,18585\ , \ldots \\ & +1^\circ - 0^\circ,18585\ , & +2^\circ - 0^\circ,18585\ . \ldots \end{array}$$

soit

$$-\ 0^\circ,18585\ , \quad -\ 1^\circ,18585\ , \quad -\ 2^\circ,18585\ , \ldots\ 0^\circ,81415\ , \ldots$$

Mais nous ne connaissons pas *a priori* l'échelle précédente, qui correspond à $h = 0$: nous groupons donc suivant une échelle quelconque, quitte à introduire h dans les calculs, et à calculer h (form. 3).

II. — Dans les tirages de boules d'une urne, $mp - x$, $mq + x$ sont des *écarts*, *forcément entiers*; mais quand nous étudions une statistique, rien ne nous oblige à prendre des écarts entiers: nous n'avons pas à supposer que les écarts sont entiers, puisqu'à un écart fractionnaire, p. e. correspond une température exprimée en dixièmes, centièmes, même millièmes de degrés......

III. — Les calculs donnent pour m des valeurs FRACTIONNAIRES Dans un tirage de boules d'une urne, le nombre m d'épreuves, de tirages, est *entier*.

La seule manière d'interpréter ce *fait:* que l'on trouve m fractionnaire — ce qui n'empêche nullement les calculs numériques — est de se reporter aux cas classiques d'inversions de problèmes: les plus simples sont la soustraction et la division, qui *introduisent* en arithmétique de nouveaux algorithmes: les nombres négatifs et les nombres fractionnaires.

Quand nous calculons la probabilité d'un écart donné, à propos d'un jeu de hasard quelconque, nous étudions un problème *direct*; quand nous ramenons une statistique donnée à une courbe ou fonction de probabilité simple, nous étudions le problème *inverse* du précédent. L'inversion agrandit le champ des valeurs de m, oblige de considérer des valeurs de m non entières.

Par contre, les valeurs de m négatives, les valeurs de p, q négatives ou plus grandes que *un* ne sauraient être prises en considération, parce que les calculs à faire par les formules (2, 4) ne peuvent plus être faits.

VI

10. — Quand on calcule les éléments h, p, q, m d'une statistique dont le graphique paraît suivre la loi (1) ou (3), les cas suivants se présentent [1]:

I. — On trouve

$$0 < p < 1 \qquad 0 < q < 1 \qquad \text{(les calculs donnent } \textit{toujours } p + q = 1)$$

et

$$m > 0 \ .$$

De plus, les nombres calculés reproduisent *convenablement* les données (N° 7):

On doit conclure que la *statistique traduit un phénomène naturel bien défini:* dans le cas étudié, une race d'hommes à peu près homogène.

N'oublions pas que cette statistique se rapporte à des temps anciens, où il n'y avait guère de mélanges de races aux Etats-Unis.

II. — On trouve encore

$$0 < p < 1 \qquad 0 < q < 1 \qquad m > 0 \ ;$$

et les nombres calculés représentent les données, mais avec des divergences locales (N° 8) *la statistique traduit l'effet d'un phénomène principal et des effets de phénomènes secondaires*; la divergence pour 0° dans la statistique des températures de Montsouris s'explique facilement. Le calcul donne moins de températures *zéro* que l'observation n'en indique ; c'est parce que la glace qui se forme ou la glace qui fond est un régulateur de températures. *C'est au météorologiste à expliquer les anomalies qui se présentent pour 4°, 11°, 15°.* Il est d'un grand intérêt de les déceler, comme nous l'avons fait.

[1] Le lecteur trouvera de nombreux exemples de tous ces cas, tirés de la démographie de la météorologie, des sciences naturelles, dans: *Annales Soc. scient. de Bruxelles*, t. 48 (1928), série des sciences mathématiques, Mémoires, p. 1 (fasc. 3).

III. — On trouve

$$0 < p < 1 \qquad 0 < q < 1 \qquad m > 0 ,$$

mais la courbe calculée diffère notablement de la courbe observée.

IV. — On trouve

$$p \text{ ou } q > 1 \qquad \text{ou bien :} \qquad p \text{ ou } q < 0$$
$$m > 0 .$$

V. — On trouve

$$0 < p < 1 \qquad 0 < q < 1 , \qquad m < 0 .$$

Dans les cas III, IV, V, la représentation de la Statistique donnée par une fonction de probabilité simple (1) ou (3) est impossible.

Voici un exemple [1]: grains de blé d'espèces différentes, mélangés, classés par longueurs, courbe *graphique* offrant l'apparence de la fonction de probabilité simple.

Dans les cas III, IV, V, on doit présupposer la superposition de courbes (3), donc la superposition de phénomènes.

VII

11. — Nous venons de résoudre les problèmes que voici: la statistique est-elle due à un phénoméne unique ou à plusieurs phénomènes ?

Quand il y a un phénomène prépondérant et des phénomènes accessoires, nous avons mis ceux-ci en évidence (N° 8).

Nous avons ajusté la statistique quand elle est le fait d'un phénomène unique (N° 7) ou de plusieurs phénomènes, l'un de ceux-ci étant prépondérant (N° 8).

Nous pouvons résoudre avec facilité les problèmes suivants: interpolation; calcul de la mode: son abscisse est une fonction simple de $q-p$ [2]. *Jusqu'ici, on ne savait pas calculer la mode.*

---

[1] Ce cas est traité dans *Ann. Soc. scient. de Bruxelles*. Voir référence à la note du début du N° 10.

[2] *Ann. Soc. scient. de Bruxelles*, t. 48, 1928, sciences mathématiques, fasc. 1.

Enfin, nous avons mis en évidence un fait analytique important et inattendu:

Quand on varie le groupement des données élémentaires, p. e. quand dans une même statistique portant sur des longueurs, on prend comme unités le centimètre et le pouce, il en résulte ceci: h, p, q, m *n'ont pas les mêmes valeurs pour la statistique en centimètres et pour la statistique en pouces*, ce qui est surprenant à première vue pour p, q.

Il existe d'ailleurs un invariant, au sens large du mot, une fonction de m, p, q, qui conserve la même valeur quand on change d'unité [1].

Le problème qui se pose maintenant est celui-ci: *étude des statistiques qui relèvent de courbes* (1) *ou* (3) *superposées*, problème qui n'a encore été abordé que par des moyens de fortune et qui présente de très grandes difficultés. Il faut chercher à déterminer h, p, q, m pour chacune des courbes composantes.

Qu'il me soit permis en terminant, de remercier MM. Alliaume (Louvain), Fehr et Mirimanoff (Genève), Potin, le général Perrier et Fichot (Paris) qui ont bien voulu s'intéresser à mes travaux dans l'ordre d'idées que je viens d'exposer [2].

Paris, mai 1928.

---

[1] La question est traitée dans *Ann. Soc. scient. de Bruxelles*. Voir référence à la note du début du N° 10, et aussi t. 47, première partie, *Comptes rendus des séances*, p. 87.

[2] Le lecteur pourra approfondir ces questions pour la lecture d'un Mémoire publié dans la *Revue générale des Sciences*, 30 mai et 15 juin 1928, où nos Mémoires des *Ann. de la Soc. scient. de Bruxelles* sont résumés et commentés.