ANNALES DE L'I. H. P.

ALF GULDBERG

Sur les lois de probabilités et la corrélation

Annales de l'I. H. P., tome 5, nº 2 (1935), p. 159-176
<http://www.numdam.org/item?id=AIHP_1935__5_2_159_0>

Sur les lois de probabilités et la corrélation

PAR

ALF. GULDBERG

J'ai eu l'honneur, il y a quelque temps, de publier dans ce recueil une série de conférences sur les fonctions discontinues d'une variable (1). Le présent article constitue l'application des résultats énoncés dans ces conférences au problème de la dépendance aléatoire, ou « stochastique » suivant la terminologie de TSCHUPROW, de deux variables aléatoires x et y.

TSCHUPROW a eu le grand mérite de donner, le premier, une définition claire et précise du concept de dépendance stochastique entre deux variables aléatoires (2) : « Das Wesen der stochastichen Verbundenheit zwischen zwei zufälligen Variablen besteht darin, dass die möglichen Werte der einen Variablen in Verbindung mit verschiedenen möglichen Werten der anderen Variablen auftreten, und dass jeder solchen Kombination eine bestimmte Wahrscheinlichkeit zukommt. Die Gesamtheit der verschiedenen Kombinationen der möglichen Werte der beiden Variablen und der diesen Kombinationen zukommenden Wahrscheinlichkeiten wollen wir als *das Abhängigkeitsgesetz der Variablen* bezeichnen. »

« Ist das Abhängigkeitsgesetz gegeben, so kennt man alles, was über die stochastische Verbundenheit zwischen den Variablen ausgesagt werden kann. Alles übrige lässt sich aus dem Abhängigkeitsgesetze deduzieren. Man darf mithin die Bestimmung des *Abhängigkeitsgesetzes als die eigentliche Hauptaufgabe der Forschung betrachten.* »

(1) *Annales de l'Institut Henri-Poincaré*, vol. III, 1933, p. 229.
(2) TSCHUPROW : « *Grundbegriffe und Grundprobleme der Korrelationstheorie* », p. 30.

Dans cet article, je traiterai ce problème en ne considérant, pour fixer les idées, que des lois de probabilités discontinues simples.

Soit $f(x, y)$ la loi de probabilité des variables aléatoires x et y. Nous appellerons moment, d'ordres f et g, ou d'ordre total $f+g$, l'expression

$$s_{fg} = \sum\sum x^f y^g \cdot f(x, y) \; ;$$

s_{fg} est donc l'espérance mathématique du produit $x^f y^g$, que nous désignerons pour abréger par :

$$s_{fg} = \mathrm{E} x^f y^g.$$

En particulier, on a

$$s_{10} = \mathrm{E}x, \qquad s_{01} = \mathrm{E}y,$$

s_{10} est l'espérance mathématique de x, s_{01} est l'espérance mathématique de y.

Nous désignerons les moments calculés par rapport à s_{10} et s_{01} par

$$m_{fg} = \sum\sum (x - s_{10})^f \cdot (y - s_{01})^g \cdot f(x, y).$$

Posons ensuite

$$r_{fg} = \frac{m_{fg}}{m_{20}^{\frac{f}{2}} \cdot m_{02}^{\frac{g}{2}}}.$$

En particulier, on a

$$r_{11} = \frac{m_{11}}{\sqrt{m_{20} \cdot m_{02}}}$$

r_{11} est appelé *coefficient de corrélation.*

Outre ces paramètres nous introduisons les semi-invariants définis par l'identité en u et v :

$$e^{\frac{1}{1}(\lambda_{10}u + \lambda_{01}v) + \frac{1}{1 \cdot 2}(\lambda_{20}u^2 + 2\lambda_{11}uv + \lambda_{02}v^2) + \cdots}$$
$$= \sum\sum f(x, y) \cdot e^{xu + yv} = 1 + \sum\sum \frac{u^f \cdot v^g}{f!\, g!} \cdot s_{fg}.$$

Cette identité est toute formelle, elle remplace le système d'équations linéaires auquel elle conduit. Les premiers semi-invariants sont :

$$\lambda_{10} = s_{10}, \quad \lambda_{01} = s_{01}, \quad \lambda_{20} = s_{20} - s_{10}^2, \; \lambda_{11} = s_{11} - s_{10}s_{01} \quad \lambda_{02} = s_{02} - s_{01}^2.$$

Une fois la loi de dépendance donnée, tous ces paramètres sont connus et inversement la loi de dépendance est complètement caractérisée par un nombre suffisant de ces paramètres ; cependant un seul de ces paramètres, par exemple le coefficient de corrélation, ne caractérise pas complètement la loi de dépendance.

Deux problèmes importants peuvent se poser. Le premier est de trouver les moments d'une loi donnée, le second d'établir des critères permettant de dire si une certaine table de corrélation peut être représenté par une loi déterminée. Je traiterai ces problèmes pour la loi de Bernoulli, pour la loi de Poisson qui est son cas limite, pour la loi de Pascal, et enfin pour la loi hypergéométrique. Nous verrons que la solution s'obtient facilement dans tous les cas au moyen des équations aux différences finies caractérisant la loi donnée.

1. — *Loi de Bernoulli pour deux variables.* —

$$f(x, y) = \frac{k!}{x!\, y!\, (k-x-y)!} \cdot p^x q^y (l - p - q)^{k-y-x}$$

est la probabilité pour que l'événement E, de probabilité constante p, se présente x fois, et que l'événement F de probabilité constante q, se présente y fois en k épreuves, les événements E, F, et non (EF), devant se présenter nécessairement à chaque épreuve.

La fonction $f(x,y)$ satisfait aux deux équations aux différences finies suivantes :

$$(1) \qquad (1 - p - q)\,(x + 1) \cdot f(x, y) = p(k - x - y) \cdot f(x, y)$$

$$(2) \qquad (1 - p - q)\,(y + 1) \cdot f(x, y) = q(k - x - y) \cdot f(x, y).$$

De ces équations on déduit des formules de récurrence pour les moments s_{fg} d'ordre $(f + g)$. Multiplions l'équation (1) par

$$(1 + x)^{n-1} \cdot y^r$$

et l'équation (2) par

$$x^r \cdot (1 + y)^{n-1} \;;$$

on a :

$$\begin{aligned}(1 - p - q)\,(x + 1)^n \cdot y^r \cdot f(x + 1, y) &= kpx^{n-1} \cdot y^r \cdot f(x, y) + \\ &+ kp(n - 1)\,x^{n-2} \cdot y^r \cdot f(x, y) + \cdots + kpy^r \cdot f(x, y) - \\ - px^n \cdot y^r \cdot f(x,y) - p(n-1)x^{n-1} \cdot y^r \cdot f(x,y) &- \cdots - pxy^r \cdot f(x,y) - \\ - px^{n-1} \cdot y^{r+1} \cdot f(x,y) - p(n-1)x^{n-2} \cdot y^{r+1} \cdot f(x.y) &- \cdots - py^{r+1} \cdot f(x,y).\end{aligned}$$

et

$$(1 - p - q)(y + 1)^n \cdot y^r \cdot f(x, y + 1) = kqy^{n-1} \cdot x^r \cdot f(x, y) +$$
$$+ kq(n - 1)y^{n-2} \cdot x^r \cdot f(x, y) + \cdots + kqx^r \cdot f(x, y) -$$
$$- qy^n \cdot x^r \cdot f(x, y) - q(n - 1)y^{n-1} \cdot x^r \cdot f(x, y) - \cdots - qyx^r \cdot f(x, y) -$$
$$- qy^{n-1} \cdot x^{r+1} \cdot f(x, y) - q(n - 1)y^{n-2} \cdot x^{r+1} \cdot f(x, y) - \cdots - qx^{r+1} \cdot f(x, y).$$

Faisons la somme pour toutes les valeurs de x et y, en remarquant que

$$\sum\sum x^r \cdot y^n \cdot f(x, y) = \sum\sum (x + 1)^r \cdot y^n \cdot f(x + 1, y)$$
$$= \sum\sum x^r \cdot (y + 1)^n \cdot f(x, y + 1) = s_{rn},$$

et

$$\sum\sum f(x, y) = 1.$$

On trouve les deux formules de récurrence :

$$(3) \quad \begin{cases} (1 - p - q)s_{nr} = kps_{n-1r} + kp(n - 1)s_{n-2r} + \cdots + kps_{0r} - \\ - ps_{nr} - p(n - 1)s_{n-1r} - p\dfrac{(n-1)(n-2)}{1 \cdot 2}s_{n-2r} - \cdots - ps_{1r} - \\ - ps_{n-1r+1} - p(n - 1)s_{n-2r+1} - \cdots - ps_{0r+1}, \end{cases}$$

et

$$(4) \quad \begin{cases} (1 - p - q)s_{rn} = kqs_{rn-1} + kq(n - 1)s_{rn-2} + \cdots + kqs_{r0} - \\ - qs_{rn} - q(n - 1)s_{rn-1} - q\dfrac{(n-1)(n-2)}{1 \cdot 2}s_{rn-2} - \cdots - qs_{r1} - \\ - qs_{r+1n-1} - q(n - 1)s_{r+1n-2} - \cdots - qs_{r+10}. \end{cases}$$

Prenons dans ces formules $n = 1$, $r = 0$; on a :

$$(1 - q)s_{10} + ps_{01} = kp$$
$$qs_{10} + (1 - p)s_{01} = kq$$

d'où

$$s_{10} = kp, \qquad s_{01} = kq.$$

Posons dans nos formules (3) et (4) $n = 2$, $r = 0$, et dans la formule (3), $n = 1$, $r = 1$; on trouve 3 équations pour les moments s_{20}, s_{11}, s_{02} d'où l'on déduit les valeurs

$$s_{20} = (kp)^2 + kp(1-p), \qquad s_{11} = k(k-1)pq, \qquad s_{02} = (kq)^2 + kq(1-q).$$

On a, d'ailleurs :

$$m_{20} = kp(1-p), \qquad m_{11} = -kpq, \qquad m_{02} = kq(1-q)$$

et

$$\lambda_{10} = kp, \quad \lambda_{01} = kq, \quad \lambda_{20} = kp(1-p), \quad \lambda_{11} = -kpq, \quad \lambda_{02} = kq(1-q).$$

Le coefficient de corrélation est

$$r_{11} = \frac{m_{11}}{\sqrt{m_{20}\cdot m_{02}}} = -\sqrt{\frac{pq}{(1-p)(1-q)}}.$$

on trouve donc la même valeur que dans le cas limite $k \to \infty$, c'est-à-dire la corrélation normale (1).

Divisons nombre à nombre les équations (1) et (2) ; on aura :

$$\text{(A)} \qquad \frac{(x+1)\cdot f(x+1, y)}{(y+1)\cdot f(x, y+1)} = \frac{p}{q}$$

ou

$$\text{(5)} \qquad q(x+1)\cdot f(x+1, y) = p(y+1)\cdot f(x, y+1)$$

Faisons la somme pour toutes les valeurs de x et y ; on trouve :

$$qs_{10} = ps_{01}.$$

En multipliant l'équation (5) par $(1 + x)$ et par $(1 + y)$ et en prenant la somme pour toutes les valeurs de x et y, on aura :

$$qs_{20} = ps_{11} + ps_{01}$$
$$ps_{02} = qs_{11} + qs_{10}.$$

Eliminons p et q ; il vient :

$$\frac{s_{10}}{s_{01}} = \frac{s_{20}}{s_{11} + s_{01}} = \frac{s_{11} + s_{10}}{s_{02}}.$$

On obtient également ces relations en éliminant k, p, q entre les expressions des moments de la loi de Bernouilli. D'après M. Frisch (2),

(1) Voir Darmois : *Statistique Mathématique*, p. 240.
(2) Metron, vol. X, nº 3, p. 51.

on appelle ces relations les critères complets auxquels doivent satisfaire les moments empiriques d'une table de corrélation, pour pouvoir être représentés par une loi de Bernoulli.

Écrivons les équations aux différences finies (1), (2), (A) de la loi de Bernoulli, de la manière suivante :

$$\frac{(x+1)\cdot f(x+1,y)}{f(x,y)} + \frac{p}{1-p-q}\cdot(x+y) = \frac{pk}{1-p-q},$$

$$\frac{(y+1)\cdot f(x,y+1)}{f(x,y)} + \frac{q}{1-p-q}\cdot(x+y) = \frac{qk}{1-p-q},$$

$$\frac{(x+1)\cdot f(x+1,y)}{(x+1)\cdot f(x,y+1)} = \frac{p}{q},$$

Exprimons les constantes p, q, k de la loi de Bernoulli en fonction des semi-invariants ; en posant

$$\lambda_{20}\lambda_{01} + \lambda_{02}\lambda_{10} - \lambda_{10}\lambda_{01} = \Delta,$$

on trouve

$$\frac{p}{1-p-q} = \frac{\lambda_{10}\lambda_{01} - \lambda_{01}\lambda_{20}}{\Delta},$$

$$\frac{q}{1-p-q} = \frac{\lambda_{10}\lambda_{01} - \lambda_{10}\lambda_{02}}{\Delta},$$

$$k = \frac{\lambda_{01}^2}{\lambda_{01} - \lambda_{02}} = \frac{\lambda_{10}^2}{\lambda_{10} - \lambda_{20}},$$

$$\frac{kp}{1-p-q} = \frac{\lambda_{10}^2\cdot\lambda_{01}}{\Delta},$$

$$\frac{kq}{1-p-q} = \frac{\lambda_{01}^2\cdot\lambda_{10}}{\Delta},$$

Les équations aux différences finies peuvent donc s'écrire

$$P(x,y) \equiv \frac{(x+1)\cdot f(x+1,y)}{f(x,y)} + \frac{\lambda_{10}\lambda_{01} - \lambda_{01}\lambda_{20}}{\Delta}(x+y) = \frac{\lambda_{10}^2\lambda_{01}}{\Delta}$$

$$Q(x,y) \equiv \frac{(y+1)\cdot f(x,y+1)}{f(x,y)} + \frac{\lambda_{10}\lambda_{01} - \lambda_{10}\lambda_{02}}{\Delta}(x+y) = \frac{\lambda_{01}^2\lambda_{10}}{\Delta}$$

$$R(x,y) \equiv \frac{(x+1)\cdot f(x+1,y)}{(y+1)\cdot f(x,y+1)} = \frac{\lambda_{10}}{\lambda_{01}}.$$

$P(x, y)$, $Q(x, y)$, $R(x, y)$ désignant les premiers membres de ces équations, on voit que les expressions

$$\frac{P(x, y) \cdot \Delta}{\lambda_{10}^2 \lambda_{01}} \equiv \alpha_1(x, y)$$

$$\frac{Q(x, y) \cdot \Delta}{\lambda_{01}^2 \lambda_{10}} \equiv \alpha_2(x, y)$$

$$\frac{R(x, y) \cdot \lambda_{01}}{\lambda_{10}} \equiv \alpha_3(x, y),$$

sont égales à l'unité pour toutes les valeurs de x, y dans le cas d'une loi de Bernoulli. Nous appellerons ces relations, avec M. FRISCH les « *critères locaux* » d'une loi de Bernoulli.

Donnons-nous une table de corrélation $[H(x_i, y_j) ; x_i, y_{2j}]$.

Formons ses semi-invariants λ_{10}, λ_{01}, λ_{20}, λ_{20}. Si les conditions exprimées par les critères locaux correspondants $\alpha(x, y)$ sont approximativement satisfaites pour toutes les valeurs de x et y, la table de corrélation doit pouvoir être representée d'une manière approchée par une loi de Bernoulli.

2. — *Loi de Poisson.* — Si dans la loi de Bernoulli

$$f(x, y) = \frac{k!}{x!\, y!\, (k - x - y)!} \cdot p^x q^y (1 - p - q)^{k-x-y},$$

k tend vers l'infini, et p, q vers zéro, de façon que le produit kp tende vers une constante a et kq vers une constante b, nous aurons la *loi de Poisson*

$$f(x, y) = \frac{a^x b^y}{x!\, y!} e^{-a-b}.$$

La loi de Poisson est caractérisée par les équations aux différences finies suivantes :

$$(x + 1) \cdot f(x + 1, y) = a \cdot f(x, y)$$
$$(y + 1) \cdot f(x, y + 1) = b \cdot f(x, y).$$

Comme pour la loi de Bernoulli, nous pouvons en déduire les deux formules de récurrence pour les moments de la loi de Poisson, à savoir

$$s_{nr} = as_{n-1\,r} + a(n - 1)s_{n-2\,r} + \cdots \qquad + as_{0r}$$
$$s_{rn} = bs_{rn-1} + b(n - 1)s_{rn-2} + \cdots \qquad + bs_{r0}.$$

d'où en particulier

$$s_{10} = a, \qquad s_{01} = b, \qquad s_{20} = a^2 + a, \qquad a_{11} = ab, \qquad s_{02} = b^2 + b.$$

On trouve facilement que les semi-invariants marginaux sont constants

$$\lambda_{10} = \lambda_{20} = \lambda_{30} = \cdots = a, \quad \lambda_{01} = \lambda_{02} = \lambda_{03} = \cdots = b, \quad \text{et} \quad \lambda_{ik} = 0, i, k \neq 0.$$

Si les semi-invariants marginaux d'une table de corrélation sont constants et les autres semi-invariants λ_{ik} $(i,\ k \neq 0)$ égaux à zéro, la table peut être représentée d'une manière approximative par une loi de Poisson.

3. — *Loi de Pascal.* — La loi de Pascal pour deux variables

$$f(x, y) = \frac{(k + x + y)\,!}{k\,!\,x\,!\,y\,!}\, p^x q^y (1 - p - q)^{k+1},$$

donne la probabilité pour que dans une série de $(k + x + y + 1)$ épreuves : l'événement A, de probabilité p, se produise x fois, l'événement B de probabilité q, se produise y fois, et l'événement défavorable *non* (A, B), de probabilité $(1 - p - q)$, se produise $k + 1$ fois.

La fonction de Pascal pour deux variables satisfait aux équations aux différences finies suivantes :

$$(x + 1)\cdot f(x + 1, y) = p(k + 1 + x + y)\cdot f(x, y)$$
$$(y + 1)\cdot f(x, y + 1) = q(k + 1 + x + y)\cdot f(x, y),$$

En multipliant la première équation par $(1 + x)^{n-1}\cdot y^r$ et la seconde par $(1 + y)^{n-1}\cdot x^r$, et en faisant la somme de 0 à ∞, on trouve comme plus haut les formules de récurrences pour les moments

$$\begin{aligned} s_{nr} = {} & p(k + 1)s_{n-1r} + \\ & + p(k + 1)(n - 1)s_{n-2r} + \cdots + p(k + 1)s_{0r} \\ & + ps_{nr} + p(n - 1)s_{n-1r} + \cdots + ps_{1r} \\ & + ps_{n-1\ r+1} + p(n - 1)s_{n-2\ r+1} + \cdots + ps_{0r+1} \\ s_{rn} = {} & q(k + 1)s_{rn-1} + \\ & + q(k + 1)(n - 1)s_{rn-2} + \cdots + q(k + 1)s_{r0} \\ & + qs_{r+1\ n-1} + q(n - 1)s_{r+1\ n-2} + \cdots + qs_{r+10} \\ & + qs_{rn} + q(n - 1)s_{rn-1} + \cdots + qs_{r1} \end{aligned}$$

De ces formules on tire

$$\lambda_{10} = s_{10} = \frac{p(k+1)}{1-p-q} \qquad \lambda_{01} = s_{01} = \frac{q(k+1)}{1-p-q}$$

et

$$s_{20} = s_{10}^2 + s_{10} \cdot \frac{1-q}{1-p-q}, \qquad s_{02} = s_{01}^2 + s_{01} \cdot \frac{1-p}{1-p-q},$$

$$s_{11} = s_{10}s_{01} + s_{10} \frac{q}{1-p-q},$$

d'où

$$\lambda_{20} = m_{20} = \frac{(k+1)p(1-p)}{(1-p-q)^2}, \qquad \lambda_{11} = m_{11} = \frac{(k+1)pq}{(1-p-q)^2},$$

$$\lambda_{02} = m_{02} = \frac{(k+1)q(1-p)}{(1-p-q)^2}$$

et par conséquent :

$$\lambda_{10} < \lambda_{20}, \qquad \lambda_{10} < \lambda_{02}.$$

Le coefficient de corrélation est

$$r_{11} = \sqrt{\frac{p \cdot q}{(1-p)(1-q)}}.$$

En introduisant dans les équations aux différences finies de la loi de Pascal, à la place de p, q, k les semi-invariants λ_{10}, λ_{01}, λ_{20}, λ_{02} on voit facilement que la forme des équations est la même que dans le cas de la loi de Bernoulli (et Poisson), à cela près cependant que pour la loi de Pascal on a $\lambda_{10}, < \lambda_{20}$, $\lambda_{01} < \lambda_{02}$, pour la loi de Bernoulli $\lambda_{10} > \lambda_{20}$, $\lambda_{01}, > \lambda_{02}$, et pour la loi de Poisson $\lambda_{10} = \lambda_{20}$, $\lambda_{01} = \lambda_{02}$.

Les critères locaux auxquels doit satisfaire une table de corrélation pour pouvoir être représentée par une loi de Bernoulli, ou par une loi de Poisson ou par une loi de Pascal, ont donc la même forme dans les trois cas (voir les critères locaux pour la loi de Bernoulli), mais avec les conditions $\lambda_{10}, > \lambda_{20}$, $\lambda_{01}, > \lambda_{02}$ pour la loi de Bernoulli, les conditions $\lambda_{10} = \lambda_{20}$, $\lambda_{01} = \lambda_{02}$ pour la loi de Poisson et enfin $\lambda_{10}, < \lambda_{20}$, $\lambda_{01} < \lambda_{02}$ pour la loi de Pascal.

Exemple. — 1. Un dé est jeté 12 fois. On a noté le nombre de fois que sont sortis, d'une part le 6 et d'autre part le 1. On répète la même épreuve 1.100 fois.

Les résultats obtenus ont été consignés dans le tableau de corrélation suivant ; x désigne le nombre de fois que le 6 est sorti et y le nombre correspondant pour le 1.

y \ x	0	1	2	3	4	5	6	7	8	9	Somme
0......	16	23	28	24	13	8	2	3			117
1......	26	71	83	54	35	4	2	1			276
2......	36	84	99	62	33	7	3				324
3......	24	45	66	54	22	7	1				219
4......	16	31	31	20	6	2	1				107
5......	10	15	7	4	1						37
6......	1	10	4	1							16
7......	2		2								4
8......											
9......											
Somme	131	279	320	219	110	28	9	4			1.100

Les moments, les semi-invariants et le coefficient de corrélation de ce tableau sont les suivants :

$$s_{10} = 2{,}03, \quad s_{01} = 2{,}11, \quad s_{02} = 5{,}92, \quad s_{02} = 6{,}30, \quad s_{11} = 4{,}03$$

$$\lambda_{20} = m_{20} = 1{,}18, \quad \lambda_{02} = m_{02} = 1{,}88, \quad \lambda_{11} = u_{11} = -0{,}25, \quad r_{11} = -0{,}14.$$

D'après la nature des épreuves, on doit s'attendre à ce qu'on puisse approcher ce tableau au moyen de la loi de Bernoulli

$$\left(p = q = \frac{1}{6}, \qquad k = 12\right).$$

Dans ce cas, le premier critère total est l'égalité des rapports

$$\frac{s_{10}}{s_{01}} = \frac{s_{20}}{s_{11} + s_{01}} = \frac{s_{11} + s_{10}}{s_{02}}.$$

Or, nous trouvons

$$\frac{s_{10}}{s_{01}} = 0{,}97 \qquad \frac{s_{20}}{s_{11} + s_{01}} = 0{,}96 \qquad \frac{s_{11} + s_{10}}{s_{02}} = 0{,}96$$

ce qui constitue un bon accord. Donnons aussi quelques valeurs pour les critères locaux, on a :

$$\alpha_1 (0,0) = 0{,}54,\ \alpha_1 (1,0) = 1{,}02,\ \alpha_1 (2,0) = 0{,}73,\ \alpha_1 (3,0) = 1{,}11,\ \alpha_1 (0,1) = 1{,}09,\ \alpha_1 (1,1) = 1{,}00$$

$$\alpha_1 (2,1) = 0{,}92,\ \alpha_1 (3,1) = 1{,}22,\ \alpha_1 (4,1) = 0{,}52,\ \alpha_1 (0,2) = 1{,}00,\ \alpha_1 (1,2) = 1{,}08,\ \alpha_1 (2,2) = 0{,}95$$

$$\alpha_1 (3,2) = 1{,}11,\ \alpha_1 (4,1) = 0{,}76,\ \alpha_1 (0,3) = 0{,}95,\ \alpha_1 (1,3) = 1{,}35,\ \alpha_1 (2,3) = 1{,}23,\ \alpha_1 (3,3) = 0{,}98$$

$$\alpha_1 (4,3) = 1{,}02,\ \alpha_1 (0,4) = 0{,}97,\ \alpha_1 (1,4) = 1{,}06,\ \alpha_1 (2,4) = 1{,}09,\ \alpha_1 (3,4) = 0{,}88,\ \alpha_1 (4,4) = 1{,}03$$

$$\alpha_3 (0,0) = 0{,}88,\ \alpha_3 (1,0) = 0{,}99,\ \alpha_3 (2,0) = 1{,}11,\ \alpha_3 (1,1) = 0{,}99,\ \alpha_3 (2,2) = 0{,}94,\ \alpha^3 (3,3) = 1{,}10.$$

Comme on le voit l'accord est assez satisfaisant.

Comme comparaison considérons le tableau de la loi de Bernoulli multiplié par 1.100. Les moments théoriques sont :

$$s_{10} = 2, \quad s_{01} = 2, \quad s_{20} = 5{,}66, \quad s_{11} = 3{,}66, \quad s_{02} = 5{,}66,$$
$$r_{11} = -\,0{,}2.$$

L'accord entre les moments théoriques et les moments expérimentaux est bon.

$y =$ \ $x =$	0	1	2	3	4	5	6	7	8
0.........	9	25	35	29	16	7	2		
1.........	25	71	87	67	33	12	3	1	
2.........	35	87	98	67	29	9	2		
3.........	29	67	67	38	14	4	1		
4.........	16	33	29	14	5	1			
5.........	7	12	9	4	1				
6.........	2	3	2	1					
7.........		1							
8.........									

Exemple. 2. — Le livre *Thesauraus logarithmorum* de VEGA contient, dans sa première partie, les logarithmes à dix décimales des nombres de 1.000 à 11.000. Chacune des pages du livre contient cinq colonnes, et chaque colonne 60 lignes. BRUNS a examiné combien de fois dans les premières 1.000 colonnes une ligne finit par un zéro ; il a comparé ensuite la série trouvé à celle que fournirait la loi de Bernoulli à une variable (pour $p = 0{,}1$, $k = 60$). L'accord constaté est satisfaisant. Il faut noter qu'ici tout hasard est exclu. J'ai généralisé ces recherches. Les tables de VEGA *Logarithmisches-Trigonometrisches Handbuch*, 39e édition, contient des logarithmes à 7 décimales rangés dans chaque page sur 10 colonnes, chaque colonne contenant 50 lignes. On a noté combien de fois, dans 1.000 colonnes, les lignes finissent par un nombre pair (2, 4, 6 ou 8), désigné par x, et combien de fois ces lignes finissent par un nombre impair (1, 3, 5, 7 ou 9), désigné par y. Les résultats sont inscrits dans le tableau de la page suivante.

On trouve comme moments :

$$s_{10} = 20{,}12, \quad s_{01} = 24{,}86, \quad s_{20} = 416{,}47, \quad s_{02} = 630{,}47$$
$$s_{11} = 487{,}26, \quad m_{20} = 11{,}81, \quad m_{02} = 12{,}35, \quad m_{11} = -12{,}86.$$

$y =$ \ $x =$	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	Somme
14.																						1			1
15.																	1				1	1		1	4
16.																			2	1		1			4
17.											1				1	1	1	3	2	3	3				15
18.												1	1	2	1	2	4	1	3		2				17
19.									1		1		2	2	4	4	3	4	1						22
20.						1					2	1	2	7	4	7	6	4	2	1					37
21.								1	1	1	7	6	9	11	14	6	5	2							63
22.							1	3	3	5	2	12	13	16	15	8	5	1							84
23.							1	1	5	7	9	14	26	18	10	4									95
24.					1		3	8	4	15	19	31	27	9	5		1								123
25.				1			8	6	7	25	19	25	15	6	5	2									119
26.							4	8	22	22	21	13	7	3	4										104
27.	1		1		1	4	8	8	20	18	12	3	2	1											79
28.	1				3	7	6	17	10	14	9	1	2												70
29.				1	4	5	14	16	13	4	4														61
30.			1	3	3	8	12	7	6	3	1														44
31.		1	3	5	5	11	8	3																	36
32.	1	1	1	1	1	5	1																		11
33.			1	1	3	1																			6
34.		2	1	2																					5
Somme.	3	4	8	14	21	42	66	78	92	114	107	107	106	75	63	34	26	15	10	5	6	3		1	1 000

En supposant que notre tableau puisse être approché par une loi de Bernoulli, on en déduit

$$p = 0{,}41, \quad q = 0{,}50, \quad k = 49{,}41, \quad r_{11} = -0{,}84.$$

Les valeurs des moments de la loi de Bernoulli (pour $p = 0{,}4$, $q = 0{,}5$, $k = 50$) seraient

$$s_{10} = 20, \quad s_{01} = 25, \quad s_{20} = 412, \quad s_{02} = 637{,}5, \quad s_{11} = 490,$$

$$m_{20} = 12, \quad m_{02} = 12{,}5, \quad m_{11} = -10, \quad r_{11} = -0{,}81.$$

Il existe donc un bon accord avec la théorie.

Donnons aussi quelques valeurs des critères locaux pour des valeurs voisines des termes les plus élevés :

$$\alpha_3(19{,}24) = 0{,}75, \quad \alpha_3(19{,}25) = 0{,}82, \quad \alpha_3(19{,}26) = 1{,}08, \quad \alpha_3(19{,}27) = 0{,}67$$
$$\alpha_3(19{,}28) = 1{,}92, \quad \alpha_3(18{,}23) = 1{,}71, \quad \alpha_3(18{,}24) = 1{,}08, \quad \alpha_3(18{,}25) = 1{,}49$$
$$\alpha_3(18{,}26) = 0{,}96, \quad \alpha_3(18{,}27) = 1{,}5, \quad \alpha_3(18{,}28) = 0{,}83, \quad \alpha_3(18{,}29) = 0{,}52$$
$$\alpha_3(21{,}20) = 0{,}43, \quad \alpha_3(21{,}21) = 1{,}64, \quad \alpha_3(21{,}22) = 1{,}39, \quad \alpha_3(21{,}23) = 0{,}81$$
$$\alpha_1(18{,}26) = 0{,}99, \quad \alpha_1(19{,}26) = 1{,}01, \quad \alpha_1(20{,}26) = 1{,}00, \quad \alpha_1(22{,}26) = 1{,}03.$$

Il y a certainement quelques irrégularités, mais en gros notre tableau se laisse approcher d'une manière assez satisfaisante par la loi de Bernoulli.

4. *La loi hypergéométrique.* — La loi hypergéométrique de deux variables a la forme :

$$f(x, y) = \frac{a!\qquad b!\qquad c!}{x!\,(a-x)!\,y!\,(b-y)!\,(k-x-y)!\,(c-k+x+y)!}\;\frac{k!\,(N-k)!}{N!}$$

$f(x, y)$ est la probabilité de tirer x boules blanches et y boules noires d'une urne dont la composition initiale est (p, q, r, $p + q + r = 1$) ce qui signifie que l'urne renferme $a = Np$ boules blanches, $b = Nq$ boules noires et $c = Nr$ boules rouges ; on suppose en outre que l'on effectue k tirages sans remettre dans l'urne les boules tirées.

La fonction $f(x, y)$ satisfait aux équations aux différences finies suivantes

$$(x + 1)\,(c - k + x + y + 1) \cdot f(x + 1), y) = (a - x)\,(k - x - y) \cdot f(x, y)$$
$$(y + 1)\,(c - k + x + y + 1) \cdot f(x, y + 1) = (b - y)\,(k - x - y) \cdot f(x, y).$$

En multipliant la première équation par $(1 + x)^{n-1} \cdot y^r$ et la seconde par $(1 + y)^{n-1} \cdot x^r$ et faisant la somme pour toutes les valeurs de x et de y, on aura, comme plus haut, les formules de récurrence des moments

$$\begin{aligned}(c - k)s_{nr} + s_{n+1r} + s_{nr+1} = ak(s_{n-1r} &+ (n-1)s_{n-2r} + \cdots + s_{0r}) \\ - (a + k)(s_{nr} &+ (n-1)s_{n-1r} + \cdots + s_{1r}) \\ - a(s_{n-1r+1} &+ (n-1)s_{n-2r+1} + \cdots + s_{0r+1}) \\ + (s_{nr+1} &+ (n-1)s_{n-1r+1} + \cdots + s_{1r+1}) \\ + (s_{n+1r} &+ (n-1)s_{nr} + \cdots + s_{2r}).\end{aligned}$$

$$\begin{aligned}&(c - k)s_{rn} + s_{rn+1} + s_{r+1n} = bk(s_{rn-1} + (n-1)s_{rn-2} + \cdots + s_{r0}) \\ &\quad - (b + k)(s_{rn} + (n-1)s_{rn-1} + \cdots + s_{r1}) \\ &\quad - b(s_{r+1n-1} + (n-1)s_{r+1n-2} + \cdots + s_{r+10}) \\ &\quad + (s_{r+1n} + (n-1)s_{r+1n-1} + \cdots + s_{r+11}) \\ &\quad + (s_{rn+1} + (n-1)s_{rn} + \cdots + s_{r2}) \\ &\qquad (n \geqq 1).\end{aligned}$$

On déduit de ces formules

$$s_{10} = k\frac{a}{N}, \quad s_{01} = k\frac{b}{N}, \quad s_{20} = \frac{N-k}{N-1}k\frac{a}{N}\cdot\left(1 - \frac{s}{N}\right) + \left(k\cdot\frac{a}{N}\right)^2$$

$$s_{11} = \frac{N}{N-1}\cdot k(k-1)\cdot\frac{a}{N}\cdot\frac{b}{N}, \quad s_{02} = \frac{N-k}{N-1}k\frac{b}{N}\left(1 - \frac{b}{N}\right) + \left(\frac{b}{N}k\right)^2$$

d'où l'on tire

$$\lambda_{10} = k\frac{a}{N}, \quad \lambda_{01} = k\frac{b}{N}, \quad \lambda_{20} = \frac{N-k}{N-1}k\frac{a}{N}\left(1 - \frac{a}{N}\right), \quad r_{11} = -\frac{N-k}{N-1}k\frac{a}{N}\frac{b}{N}$$

$$\lambda_{02} = \frac{N-k}{N-1}k\frac{b}{N}\left(1 - \frac{b}{N}\right), \quad \lambda_{11} = -\sqrt{\frac{p\cdot q}{(1-p)(1-q)}}.$$

On trouve donc la même expression pour le coefficient de corrélation, qu'on parte de la loi hypergéométrique ou de la loi de Bernouilli.

En éliminant les constantes a, b, c, k entre les expressions trouvées des moments on aura les critères totaux.

Les constantes a, b, c, k s'expriment en fonction des semi-invariants en posant

$$\frac{\lambda_{10}\lambda_{11}-\lambda_{20}\lambda_{01}}{\lambda_{11}}=\Delta, \qquad \Delta\,\frac{\lambda_{20}-\lambda_{10}(\Delta-\lambda_{10})}{\Delta(\lambda_{20}-\lambda_{10})+\lambda_{10}^2}=\nabla.$$

On a :

$$a=\frac{\nabla\cdot\lambda_{10}}{\Delta}\equiv\nabla_1,\quad b=\frac{\nabla\cdot\lambda_{01}}{\Delta}\equiv\nabla_2,\quad c=\nabla\left(1-\frac{\lambda_{10}+\lambda_{01}}{\Delta}\right)\equiv\nabla_3,\quad k=\Delta.$$

Les équations aux différences finies de la loi hypergéométrique s'écrivent alors

$$P(x,y)\equiv(\nabla_3-\Delta)\frac{(x+1)f(x+1,y)}{f(x,y)}+\frac{(x+1)^2f(x+1,y)}{f(x,y)}+\frac{(x+1)yf(x+1,y)}{f(x,y)}$$
$$+(\nabla_1+\Delta)x+\nabla_1x-xy-x^2=\nabla\cdot\lambda_{10},$$

$$Q(x,y)\equiv(\nabla_3-\Delta)\frac{(y+1)f(x,y+1)}{f(x,y)}+\frac{(y+1)^2f(x,y+1)}{f(x,y)}+\frac{(y+1)xf(x,y+1)}{f(x,y)}$$
$$+(\nabla_2+\Delta)y+\nabla_2y-xy-y^2=\nabla\cdot\lambda_{01},$$

$$R(x,y)\equiv\frac{(x+1)f(x+1,y)}{(y+1)f(x,y+1)}=\frac{\nabla_1-x}{\nabla_2-y},$$

et les critères locaux :

$$\bar{\alpha}_1(x,y)\equiv\frac{P(x,y)}{\nabla\cdot\lambda_{10}},\quad \bar{\alpha}_2(x,y)\equiv\frac{Q(x,y)}{\nabla\cdot\lambda_{01}},\quad \bar{\alpha}_3(x,y)\equiv\frac{R(x,y)\,(\nabla_2-y)}{\nabla_1-x}.$$

Pour toutes les valeurs de x et y les $\bar{\alpha}(x, y)$ sont égaux à l'unité dans le cas de la loi hypergéométrique. Considérons un tableau de corrélation donné et formons les semi-invariants λ_{01}, λ_{10}, λ_{20}, λ_{02}. Si les critères locaux correspondants $\bar{\alpha}(x, y)$ sont approximativement satisfaits, le tableau peut être approché par une loi hypergéométrique.

Exemple. — On note dans le tableau, reproduit ci-après nombre de fois que dans une partie de whist la main a reçu x piques et y trèfles, dans 2.000 données.

Ce tableau doit être approximé par une loi hypergéométrique (avec $a = 13$, $b = 13$, $c = 26$, $k = 13$).

Les distributions théoriques marginales sont

$$\frac{\mathrm{H}(r): \quad 26, \quad 160, \quad 412, \quad 573, \quad 477, \quad 249, \quad 83, \quad 18, \quad 2, \quad 0}{r: \quad 0, \quad 1, \quad 2, \quad 3, \quad 4, \quad 5, \quad 6, \quad 7, \quad 8, \quad 9}.$$

Le tableau donne pour les moments et le coefficient de corrélation les valeurs suivantes :

$$s_{10} = 3{,}16, \quad s_{01} = 3{,}26, \quad s_{20} = 12{,}04, \quad s_{02} = 12{,}70; \quad s_{11} = 9{,}56,$$
$$r_{11} = -\,0{,}35.$$

$y =$ \ $x =$	0	1	2	3	4	5	6	7	8	9	Somme
0.		1	4	9	11	7	5	5			42
1.		6	12	42	49	31	14	9	1		164
2.	2	18	69	87	116	59	32	13	1		397
3.	7	50	110	191	137	57	20				572
4.	10	49	112	135	95	13	2				448
5.	14	37	83	60	48	13	2				257
6.	7	17	31	21	5	1	3				85
7.	2	6	11	6	2	1					28
8.	2		2	1	1						6
9.	1										1
Somme . . :	45	134	434	552	464	203	86	30	2		2 000

Les moments et le coefficient de corrélation de la loi hypergéométrique sont :

$$s_{01} = s_{10} = 3{,}25, \quad s_{20} = s_{02} = 12{,}43, \quad s_{11} = 9{,}94 \cdots$$
$$r_{11} = -\,0{,}33.$$

On a donc un bon accord entre les moments empiriques et les moments théoriques.

Je donne encore quelques critères locaux voisins du terme le plus élevé :

$\overline{\alpha}_1(2,1) = 1,44,$	$\overline{\alpha}_1(3,1) = 1,23,$	$\overline{\alpha}_1(4,1) = 0,92,$	$\overline{\alpha}_1(5,1) = 1,06,$
$\overline{\alpha}_1(1,2) = 1,13,$	$\overline{\alpha}_1(1,3) = 0,74,$	$\overline{\alpha}_1(1,4) = 0,78,$	$\overline{\alpha}_1(1,5) = 0,70,$
$\overline{\alpha}_1(2,2) = 0,75,$	$\overline{\alpha}_1(3,2) = 1,72,$	$\overline{\alpha}_1(4,2) = 0,83,$	$\overline{\alpha}_1(5,2) = 0,98,$
$\overline{\alpha}_1(2,3) = 0,99,$	$\overline{\alpha}_1(3,3) = 0,81,$	$\overline{\alpha}_1(4,3) = 0,67,$	$\overline{\alpha}_1(5,3) = 0,93,$
$\overline{\alpha}_1(2,4) = 0,85,$	$\overline{\alpha}_1(3,4) = 0,51,$	$\overline{\alpha}_1(4,4) = 0,70,$	$\overline{\alpha}_1(5,4) = 0,80.$

On voit que la loi hypergéométrique donne une approximation assez satisfaisante du tableau précédent.

En terminant je ne puis mieux faire que de citer les sages paroles de M. E. Borel :

« La théorie des probabilités, comme toute théorie mathématique d'ailleurs, ne saurait prétendre résoudre *à priori* des questions concrètes, qui sont du domaine de l'expérience ; son seul rôle — déjà très beau si elle sait le remplir — est de guider l'expérience et l'observation par l'interprétation qu'elle fournit de leurs résultats. »

Le Gérant : E. Schneider.

Imprimerie des Presses Universitaires de France,
Paris-Saint-Amand (France). — 12-6-1935.