ANNALES DE L'I. H. P.

J.F. STEFFENSEN

Deux problèmes du Calcul des Probabilités

Annales de l'I. H. P., tome 3, nº 3 (1933), p. 319-344

<http://www.numdam.org/item?id=AIHP_1933__3_3_319_0>

Deux problèmes du Calcul des Probabilités

PAR

J. F. STEFFENSEN

I. Mesures de dépendance entre événements fortuits.

1. — On dit généralement qu'on a affaire à une loi de probabilité, lorsqu'une même expérience peut fournir un certain nombre de résultats chacun avec sa probabilité, la somme des probabilités étant égale à l'unité. Les cas possibles seront souvent exprimés par des nombres ; sinon, on peut toujours leur associer des nombres différents et parler ainsi dans tous les cas d'une variable aléatoire prenant certaines valeurs avec certaines probabilités.

Dans ce qui suit nous nous occuperons d'abord d'une variable aléatoire x prenant un nombre fini de valeurs distinctes

$$x_1, x_2, \cdots x_k \qquad (k > 1)$$

avec certaines probabilités, et d'une autre variable aléatoire y prenant les valeurs distinctes

$$y_1, y_2, \cdots y_l \qquad (l > 1)$$

chacune avec sa probabilité, la somme des probabilités étant égale à l'unité dans chacun des deux cas.

Il s'agit d'examiner les expressions qu'on a proposées pour mesurer en quelque sorte la dépendance qui peut exister entre les deux variables aléatoires x et y. On peut supposer $k > 1$, $l > 1$, parce que si $k = 1$ ou $l = 1$, l'une des variables n'est plus aléatoire.

2. — Avant d'aborder ces questions il faut dire un mot de la notation employée. La probabilité pour que x prenne la valeur x_i lorsqne y prend la valeur y_j, sera désignée par le symbole ordinaire p_{ij}. On

obtiendra la probabilité pour que x prenne la valeur x_i (quelle que soit la valeur de y) en prenant la somme de p_{ij} par rapport à toutes les valeurs de l'indice j. Nous écrirons le résultat sous la forme

$$p_{i*} = \sum_j p_{ij}, \tag{1}$$

en désignant par un astérisque la sommation par rapport à toutes les valeurs de l'indice que l'astérisque remplace. On a, de même

$$p_{*j} = \sum_i p_{ij}, \tag{2}$$

et

$$p_{**} = \sum_{ij} p_{ij} = 1. \tag{3}$$

La probabilité p_{ij} peut s'annuler pour certaines valeurs de i et de j. Au contraire, p_{i*} et p_{*j} sont toujours positives, puisqu'il ne faut tenir compte que, des valeurs possibles de x_i et de y_j. On a, de plus, $p_{i*} < 1$, $p_{*j} < 1$, parce que $k > 1$, $l > 1$.

Une autre probabilité que nous aurons à utiliser est la probabilité pour que y prenne la valeur y_j, lorsque x prend la valeur x_i. Nous désignerons cette probabilité par $p_{(i)j}$; de même $p_{i(j)}$ signifiera la probabilité pour que x prenne la valeur x_i, lorsque y prend la valeur y_j.

Nous croyons que la notation proposée présente quelque avantage sur la notation un peu plus compliquée employée par TSCHUPROW [1], et ses disciples, à savoir

$$p_{i|j} = p_{ij}, \qquad p_{i|} = p_{i*}, \qquad p_{|j} = p_{*j},$$
$$p^{(i)}{}_{|j} = p_{(i)j}, \qquad p^{(j)}{}_{i|} = p_{i(j)}.$$

3. — Après ces préliminaires, nous pouvons écrire, d'après le théorème de la multiplication des probabilités

$$p_{ij} = p_{i*}p_{(i)j}, \tag{4}$$

et

$$p_{ij} = p_{*j}p_{i(j)}, \tag{5}$$

(1) A. A. TSCHUPROW : *Grundbegriffe und Grundprobleme der Korrelationstheorie.* Leipzig, 1925.

et la *condition d'indépendance* des deux variables aléatoires x et y sera donnée par

$$p_{ij} = p_{i*}p_{*j}, \tag{6}$$

pour toutes les valeurs de i et j.

L'autre cas extrême se présente si l'une des variables aléatoires est fonction de l'autre. Dans ce cas, l'une des valeurs de x est toujours associée à la même valeur de y, et on a nécessairement $k = l$. On peut alors, sans restreindre la généralité, associer x_i à y_i, et la *condition de dépendance complète* s'écrit :

$$p_{ii} = p_{i*} = p_{*i}; \qquad p_{ij} = 0 \text{ pour } i \neq j. \tag{7}$$

Entre ces deux extrêmes se trouvent les cas où il y a une dépendance aléatoire ou « stochastique », comme dit TSCHUPROW, entre x et y, et le problème important se présente de trouver une mesure du *degré* de cette dépendance. Évidemment, ce problème ne comporte pas une solution unique ; plusieurs mesures sont possibles et, en effet, on en a proposé un certain nombre (1) Nous allons d'abord rappeler brièvement quelques unes des plus importantes en commençant par le *coefficient de corrélation.*

Écrivons (2)

$$\sigma_{rs} = \sum_{j} p_{ij}x_i^r y_j^s, \tag{8}$$

de sorte que σ_{rs} désigne l'espérance mathématique du produit $x^r y^s$, ou

$$\sigma_{rs} = \mathcal{E}x^r y^s, \tag{9}$$

et posons ensuite

$$m_{rs} = \sum_{ij} p_{ij}(x_i - \sigma_{10})^r(y_j - \sigma_{01})^s, \tag{10}$$

ou

$$m_{rs} = \mathcal{E}(x - \sigma_{10})^r(y - \sigma_{01})^s,$$

(1) Voir, par exemple : Charles JORDAN : *Statistique Mathématique*, chap. XII. J'ai en vue seulement le cas où les probabilités p_{ij} sont données. Ainsi, la corrélation des rangs (critères de SPEARMAN, ESSCHER, LINDEBERG, etc.) ne rentre pas dans les cadres de la présente analyse.

(2) J'emploie ordinairement les symboles $\bar{\sigma}$ et $\bar{m}$ pour les moments théoriques, et les symboles σ et m pour les moments déduits des observations. Ici, où nous ne nous occuperons guère des observations, il n'y a pas lieu de retenir cette distinction.

avec ces notations le coefficient de corrélation ρ sera défini par la relation

$$\rho^2 = \frac{m_{11}^2}{m_{20}m_{02}}. \tag{11}$$

On sait qu'on a toujours $0 \leq \rho^2 \leq 1$, et que ρ^2 s'annule dans le cas d'indépendance complète. Mais le fait que $\rho^2 = 0$ ne suffit pas pour prouver l'indépendance ; d'autre part on a dépendance complète pour une valeur de ρ^2 inférieure à 1, à l'exception du cas où la relation entre x et y est linéaire. Pour ces raisons le coefficient de corrélation n'est pas satisfaisant comme mesure de la dépendance aléatoire (1).

4. Une mesure plus adaptée à ce but est le *quotient de corrélation* introduit par KARL PEARSON. Si l'on pose, d'une manière générale,

$$\sigma_{0s}^{x_i} = \sum_j p_{(i)j} y_j^s = \mathcal{E}^{x_i} y^s, \tag{12}$$

de sorte que $\sigma_{0s}^{x_i}$ signifie l'espérance mathématique de y^s, lorsque x prend la valeur x_i, et si l'on pose ensuite

$$m_{0s}^x = \sum_j p_{(i)j}(y_j - \sigma_{01}^{x_i})^s = \mathcal{E}^{x_i}(y - \sigma_{01}^{x_i})^s, \tag{13}$$

le quotient de corrélation de y par rapport à x, qu'on désigne par η_{yx}, est défini par la relation

$$\eta_{yx}^2 = 1 - \frac{1}{m_{02}} \sum_i p_{i*} m_{02}^x, \tag{14}$$

qu'on peut aussi écrire sous la forme

$$\eta_{yx}^2 = \frac{1}{m_{02}} \sum_i p_{i*}(\sigma_{01}^{x_i} - \sigma_{01})^2. \tag{15}$$

Naturellement on peut définir d'une manière analogue, le quotient de corrélation de x par rapport à y qu'on désigne par η_{yx}.

On a toujours $0 \leq \eta^2 \leq 1$. Si $\eta^2 = 1$, la dépendance est complète, et réciproquement, si la dépendance est complète, on a $\eta^2 = 1$. S'il y a indépendance complète, η^2 s'annule ; mais le fait que $\eta^2 = 0$ ne suffit pas pour prouver l'indépendance. Je renvoie, pour la démons-

(1) TSCHUPROW : *l. c.*, pp. 55-56. Si, toutefois, $k = l = 2$, ρ se confond avec la mesure φ mentionnée ci-dessous ; voir TSCHUPROW *l. c.*, p. 63.

tration de ces faits, aux travaux qui ont été publiés sur cette question (1).

Il convient d'ajouter que, si $\eta_{yx}^2 = 0$, on dit selon KARL PEARSON, que « y n'est pas corrélée à x »(2). Si l'on adopte cette terminologie, les expressions « corrélation » et « dépendance aléatoire » n'ont pas le même sens. Pour cette raison nous avons évité de faire usage du mot « corrélation » pour désigner une dépendance aléatoire.

On voit que le quotient de corrélation, quoique préférable au coefficient de corrélation pour mesurer une dépendance aléatoire, n'est pas une mesure parfaite ; nous allons examiner maintenant une autre mesure, proposée également par KARL PEARSON.

5. — On voit que le coefficient de corrélation et le quotient de corrélation ont ceci de commun qu'ils dépendent des valeurs x_i et y_i que prennent les variables aléatoires ; par conséquent ce n'est qu'au moyen d'un artifice qu'ils peuvent être employés dans les cas où les événements possibles ne sont pas représentés par des nombres. Au contraire, dans les mesures dont nous nous occuperons à présent, les valeurs des variables aléatoires n'interviennent pas ; seules entrent en en jeu les probabilités des divers cas possibles, cas que nous pouvons d'ailleurs continuer à appeler x_i et y_j selon notre convention, initiale.

La *contingence moyenne* de KARL PEARSON est le nombre φ défini par la relation (3)

$$\varphi^2 = \sum_{ij} \frac{(p_{ij} - p_{i*}p_{*j})^2}{p_{i*}p_{*j}}, \tag{16}$$

qu'on peut aussi écrire

$$\varphi^2 = \sum_{ij} \frac{p_{ij}^2}{p_{i*}p_{*j}} - 1. \tag{17}$$

Il résulte de (16) que φ^2 s'annule si $p_{ij} = p_{i*}p_{*j}$ pour toutes les valeurs de i et j, c'est-à-dire s'il y a indépendance complète, et seulement dans ce cas. D'autre part, s'il y a dépendance complète, on a $k = l$, $p_{*i} = p_{i*} = p_{ii}$ et $p_{ij} = 0$ pour $i \neq j$, de sorte que, d'après (17), la valeur de φ^2 est égale, à une unité près, au nombre des cas possibles pour x (ou pour y), c'est-à-dire $\varphi^2 = k - 1 = l - 1$.

(1) Voir, par exemple : TSCHUPROW *l. c.*, pp. 52-54.
(2) TSCHUPROW : *l. c.*, p. 33 et p. 53 ; Charles JORDAN : *l. c.*, p. 305.
(3) φ^2 est le *mean square contingency* de PEARSON.

Afin d'obtenir que la mesure φ^2 proposée par KARL PEARSON prenne la valeur 1 s'il y a dépendance complète, TSCHUPROW (1) l'a légèrement modifiée en divisant φ^2 par $\sqrt{(k-1)(l-1)}$. On obtient ainsi la mesure τ définie par la relation

$$\tau^2 = \frac{\varphi^2}{\sqrt{(k-1)(l-1)}}. \tag{18}$$

Il reste à démontrer que τ^2 ne peut pas dépasser l'unité, et que la valeur 1 est atteinte *seulement* dans le cas de dépendance complète. La démonstration qui suit est due à M. G. HÖSSJER et m'a été communiquée au commencement de 1930 par M. H. CRAMÉR (2).

Écrivons (17) sous la forme

$$\varphi^2 + 1 = \sum_{i=1}^{k}\sum_{j=1}^{l}\frac{p_{ij}}{p_{i*}}\cdot\frac{p_{ij}}{p_{*j}}.$$

S'il n'y a pas dépendance complète, on a, au moins pour certaines valeurs de i et j,

$$0 < p_{ij} < p_{*j}$$

et, par conséquent,

$$\varphi^2 + 1 < \sum_{i=1}^{k}\sum_{j=1}^{l}\frac{p_{ij}}{p_{i*}} = \sum_{i=1}^{k}\frac{p_{i*}}{p_{i*}} = k,$$

donc $\varphi < \sqrt{k-1}$. On obtient de la même manière $\varphi < \sqrt{l-1}$ et, en multipliant les deux inégalités, $\varphi^2 < \sqrt{(k-1)(l-1)}$ ou $\tau^2 < 1$.

6. — On a objecté (3) à la contingence moyenne dans la forme proposée par TSCHUPROW, que l'extension aux cas où $k \neq l$ est « évidemment erronée » (« offenbarirrtümlich »). Je dois pourtant avouer que je ne puis découvrir aucune erreur dans les raisonnements qui précèdent.

Une autre objection s'attache aux « poids » $\frac{1}{p_{i*}p_{*j}}$ figurant dans (16). Ces poids, dit-on (4), ont été introduits uniquement pour obtenir

(1) *l. c.*, p. 41. Dans le cas $k = l = 2$ on a $\tau^2 = \varphi^2$.

(2) Voir aussi R. v. MISES *Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und theoretischen Physik*, 1931, p. 360, pour le cas $k = l$.

(3) V. MISES : *l. c.*, p. 359.

(4) H. CRAMÉR : *Remarks on Correlation. Skandinavisk Aktuarietidskrift*, 1924, p. 231. Voir aussi v. MISES : *l. c.*, p. 359.

des résultats simples, à savoir la formule (17), qui est certainement le plus simple possible.

C'est pourquoi je crois utile de montrer qu'on peut introduire la contingence moyenne d'une manière parfaitement naturelle, sans faire usage des poids, en écrivant

$$\varphi^2 = \sum_{ij} (p_{(i)j} - p_{*j})(p_{i(j)} - p_{i*}). \tag{19}$$

Puisque la condition d'indépendance $p_{ij} = p_{i*}\, p_{*j}$ s'écrit aussi au moyen des équations (4) et (5)

$$p_{i*} = p_{i(j)}, \quad p_{*j} = p_{(i)j}, \tag{20}$$

on peut s'assurer directement que l'expression (19) s'annule dans le cas d'indépendance, et seulement dans ce cas ; il n'y a pas besoin de donner la démonstration en détail. L'identité de (19) et de (16) résulte immédiatement si l'on écrit

$$\frac{(p_{ij} - p_{i*}p_{*j})^2}{p_{i*}p_{*j}} = \frac{p_{ij} - p_{i*}p_{*j}}{p_{i*}} \cdot \frac{(p_{ij} - p_{i*}p_{*j})}{p_{*j}}$$

et si l'on emploie (4) et (5).

7. — Nous avons supposé jusqu'ici que k et l sont des nombres finis. Si ces nombres tendent vers l'infini, on a pour τ^2 l'expression peu commode

$$\tau^2 = \lim_{kl \to \infty} \frac{1}{\sqrt{(k-1)(l-1)}} \sum_{i=1}^{k} \sum_{j=1}^{l} \frac{(p_{ij} - p_{i*}p_{*j})^2}{p_{i*}p_{*j}}$$

et il faudrait encore démontrer l'existence de la limite.

Il y a donc lieu de s'occuper de la question de trouver une mesure applicable également pour les valeurs infinies de k et l, question qui n'est pas sans importance pour les applications. Nous devons d'abord mentionner une mesure proposée par M. H. CRAMÉR dans le mémoire cité ci-dessus.

M. CRAMÉR propose comme mesure de la dépendance aléatoire *la valeur minimum* de l'expression

$$\sum_{ij} (p_{ij} - u_i v_j)^2 \tag{21}$$

pour toutes les valeurs réelles des variables u_i et v_j. Une mesure analogue est donnée dans le cas continu.

La mesure (21) s'annule, comme la contingence moyenne, dans le cas où les deux variables aléatoires sont complètement indépendantes l'une de l'autre et seulement dans ce cas. Dans tous les autres cas la mesure est positive, et M. CRAMÉR démontre que sa valeur ne peut pas excéder $1/4$. Mais cette mesure présente des difficultés sur lesquelles M. CRAMÉR lui-même a attiré l'attention. Non seulement son application est difficile, mais la dépendance complète des variables aléatoires n'est pas caractérisée par une même valeur de la mesure dans tous les cas. Pour ces raisons nous ne nous en occuperons pas en détail.

8. — Je vais montrer maintenant comment, par des considérations assez naturelles, on peut être conduit à une nouvelle définition de la mesure, qui ne donne pas lieu aux objections mentionnées jusqu'ici.

Nous prenons comme point de départ la contingence moyenne pour une table *tétradique* (1), c'est-à-dire pour une table comportant seulement deux possibilités pour x et deux possibilités pour y. On sait, (voir nº 5) que pour une telle table, la contingence moyenne possède, les qualités suivantes : φ^2 étant toujours comprise entre 0 et 1, prend la valeur 0 dans le cas d'indépendance complète et seulement dans ce cas, et la valeur 1 dans le cas de dépendance complète et seulement dans ce cas.

Considérons une table tétradique comportant les quatre possibilités

x_i se présente, $\quad$ x_i ne se présente pas,
y_j se présente, $\quad$ y_j ne se présente pas.

Soit p_{aj} la probabilité pour que x_i ne se présente pas (2) tandis que y_j se présente ; on définira de même p_{ia} et p_{aa}. On aura

$$(22) \qquad \begin{cases} p_{ia} = p_{i*} - p_{ij}, \qquad p_{aj} = p_{*j} - p_{ij}, \\ p_{aa} = 1 - p_{i*} - p_{*j} + p_{ij}. \end{cases}$$

On en déduit les relations, d'ailleurs évidentes,

$$(23) \qquad p_{a*} = 1 - p_{i*}, \qquad p_{*a} = 1 - p_{*j}.$$

(1) Selon Karl PEARSON : *tetrachoric.*
(2) L'indice « a » est une abréviation du mot « absent ».

Nous désignerons par φ_{ij} la contingence moyenne de la table tétradique. On a, d'après la définition de φ^2,

$$\left\{\begin{aligned}\varphi^2_{ij} &= \frac{(p_{ij} - p_{i*}p_{*j})^2}{p_{i*}p_{*j}} + \frac{(p_{ia} - p_{i*}p_{*a})^2}{p_{i*}p_{*a}} \\ &+ \frac{(p_{aj} - p_{a*}p_{*j})^2}{p_{a*}p_{*j}} + \frac{(p_{aa} - p_{a*}p_{*a})^2}{p_{a*}p_{*a}}\end{aligned}\right.$$

qui se réduit facilement, au moyen des équations (22) et (23), à

$$\varphi^2_{ij} = \frac{(p_{ij} - p_{i*}p_{*j})^2}{p_{i*}(1 - p_{i*})p_{*j}(1 - p_{*j})}. \tag{24}$$

On peut aussi écrire, au moyen de (4) et (5),

$$\varphi^2_{ij} = \frac{(p_{i(j)} - p_{i*})(p_{(i)j} - p_{*j})}{(1 - p_{i*})(1 - p_{*j})}. \tag{25}$$

Nous désignerons la nouvelle mesure par ψ, et nous la définirons par la relation

$$\psi^2 = \sum_{ij} p_{i*}p_{*j}\varphi^2_{ij}. \tag{26}$$

On aura aussi, au moyen de (24),

$$\psi^2 = \sum_{ij} \frac{(p_{ij} - p_{i*}p_{*j})^2}{(1 - p_{i*})(1 - p_{*j})}. \tag{27}$$

De (26) on conclut d'abord que ψ^2 est une sorte de *moyenne des grandeurs* φ^2_{ij}. Puisque $\sum_{ij} p_{i*}p_{*j} = 1$, et que φ^2_{ij} est toujours comprise entre 0 et 1, il s'ensuit que la somme (26) est convergente et que ψ^2 est aussi comprise entre ces limites.

Puisque la somme (27) contient seulement des termes positifs, elle s'annule seulement si tous les termes s'annulent. La condition nécessaire et suffisante pour cela est évidemment $p_{ij} = p_{i*}p_{*j}$ pour toutes les valeurs de i et j, c'est-à-dire qu'il y ait indépendance complète entre x et y.

Reste à démontrer que $\psi^2 = 1$ est la condition nécessaire et suffisante pour que la dépendance soit complète. Or, la condition nécessaire et suffisante pour qu'il y ait dépendance complète entre x_i et y_j (c'est-à-dire que, ou bien y_j se présente toujours si x_i se présente, ou bien y_j ne se présente jamais si x_i se présente) est $\varphi^2_{ij} = 1$. Par conséquent, la condition nécessaire et suffisante pour qu'il y ait dépendance complète entre x et y est que $\varphi_{ij}^2 = 1$ pour *toutes* les valeurs de i et j. Mais ceci veut dire, selon (26), que $\psi^2 = 1$, puisque ni p_{i*} ni p_{*j} ne peuvent s'annuler, et que

$$\sum_{ij} p_{i*}p_{*j} = 1.$$

9. — Il faut observer que dans le cas particulier assez important où i et j prennent seulement les valeurs 1 et 2, la nouvelle mesure est identique à la contingence moyenne de PEARSON. En effet dans ce cas on a

$$\varphi_{11}^2 = \varphi_{12}^2 = \varphi_{21}^2 = \varphi_{22}^2 = \varphi^2,$$

et puisque $\sum_{ij} p_{i*}p_{*j}$, il résulte de (26) que $\psi^2 = \varphi^2$.

10. — La mesure ψ a le désavantage de ne pas s'appliquer directement aux lois de probabilité continues. On peut toujours du reste, en partant d'une telle loi $f(x,y)$, construire une loi discontinue p_{ij} en posant, par exemple,

$$p_{ij} = \int_{j-\frac{1}{2}}^{j+\frac{1}{2}} \int_{i-\frac{1}{2}}^{i+\frac{1}{2}} f(x,y)dxdy, \tag{28}$$

et calculer ψ par ces probabilités. Mais si, au lieu de l'unité, on choisit un autre intervalle d'intégration, la valeur de ψ changera également.

11. — Pour obtenir une approximation de ψ au moyen d'une expérience donnée, on remplacera les probabilités par les fréquences relatives correspondantes. Nous désignerons par H_{ij} la fréquence absolue de la combinaison (x_i, y_j) ; de plus, nous écrirons

$$H_{i*} = \sum_j H_{ij}, \quad H_{*j} = \sum_i H_{ij} \qquad \text{et} \qquad N = H_{**},$$

de sorte que N désigne le nombre total des observations. Si l'on représente par Ψ l'approximation de ψ obtenue de cette manière, on trouvera au moyen de (27)

$$\Psi^2 = \frac{1}{N^2} \sum_{ij} \frac{(NH_{ij} - H_{i*}H_{*j})^2}{(N - H_{i*})(N - H_{*j})}, \tag{29}$$

qui est la forme la plus commode pour le calcul, si les fréquences absolues sont données.

Pour la contingence moyenne on obtient au moyen de (17) l'expression suivante, en désignant par Φ l'approximation de φ,

$$\Phi^2 = \sum_{ij} \frac{H^2_{ij}}{H_{i*}H_{*j}} - 1. \tag{30}$$

En calculant Ψ ou Φ il faudra omettre les termes dont les dénominateurs s'annulent. Ce cas ne se présente d'ailleurs presque jamais pour Ψ, puisque, si $H_{i*} = N$ ou $H_{*j} = N$, il n'y a pas de table à double entrée. Au contraire, on a assez souvent $H_{i*} = 0$ ou $H_{*j} = 0$ vers les extrémités de la table, et les termes correspondants de Φ tombent en défaut.

Pour les calculs numériques il est inutile sans doute d'appeler l'attention sur ces faits. Au contraire, des erreurs dans le traitement théorique de cette sorte de questions sont assez fréquentes dans la littérature. Ainsi, par exemple, TSCHUPROW — d'ailleurs l'un des savants les plus scrupuleux dans le domaine de la statistique mathématique — essaye de calculer l'espérance mathématique de $\frac{H_{ij}}{H_{i*}}$ de la manière suivante [1].

Le nombre H_{i*} peut prendre l'une ou l'autre des valeurs 0, 1, 2... N. La probabilité pour que H_{i*} prenne la valeur ν est la probabilité pour que x_i se présente ν fois dans N épreuves, c'est-à-dire

$$C_N^\nu p_{i*}^\nu (1 - p_{i*})^{N-\nu}.$$

Or, l'espérance mathématique de H_{ij} dans le cas où H_{i*} prend la valeur ν est

$$\nu p_{(i)j} = \nu \frac{p_{ij}}{p_{i*}}.$$

Donc l'espérance mathématique de $\frac{H_{ij}}{H_{i*}}$ est obtenue sous la forme

$$\mathcal{E}\frac{H_{ij}}{H_{i*}} = \sum_{\nu=0}^{N} C_N^\nu p_{i*}^\nu (1 - p_{i*})^{N-\nu} \cdot \frac{1}{\nu} \cdot \nu \frac{p_{ij}}{p_{i*}}$$
$$= \frac{p_{ij}}{p_{i*}} \sum_{\nu=0}^{N} C_N^\nu p_{i*}^\nu (1 - p_{i*})^{N-\nu} = \frac{p_{ij}}{p_{i*}}.$$

1) *L. c.*, p. 82. Nous avons remplacé la notation de TSCHUPROW par la nôtre.

Il est évident que ce résultat a été obtenu seulement en assignant au produit $\frac{1}{\nu}\cdot\nu$, indéterminé pour $\nu = 0$, la valeur parfaitement arbitraire 1. Mais il y a d'autres cas où le résultat auquel conduit cette manière de procéder est non seulement indéterminé mais infini. Ainsi, sur la même page, TSCHUPROW obtient pour l'espérance mathématique de $\frac{H_{ij}^2}{H_{i*}^2}$ l'expression suivante

$$\mathcal{E}\,\frac{H_{ij}^2}{H_{i*}^2} = \frac{p_{ij}^2}{p_{i*}^2} + \frac{p_{ij}}{p_{i*}}\left(1 - \frac{p_{ij}}{p_{i*}}\right)\sum_{\nu=0}^{N} C_N^{\nu}\, p_{i*}^{\nu}(1-p_{i*})^{N-\nu}\cdot\frac{1}{\nu}.$$

Ici, le premier terme de la somme est infini, objection bien plus sérieuse que celle indiquée par TSCHUPROW lui-même, à savoir que la sommation ne conduit pas à un résultat simple.

Malheureusement, l'espérance mathématique d'un quotient de valeurs observées se présente souvent, si l'on veut calculer l'erreur moyenne de certaines mesures de dépendance déduites d'une expérience donnée. Il semble donc qu'une révision de quelques-uns des résultats obtenus dans ce domaine par TSCHUPROW et d'autres soit fort désirable.

Naturellement, il faut *bien distinguer* entre les valeurs théoriques des mesures de dépendance et les valeurs déduites de l'expérience par les formules (29) et (30). Dans ce dernier cas, le résultat est affecté d'erreurs accidentelles, le plus grand inconvénient étant qu'on peut obtenir une valeur positive de la mesure, quand elle devrait être nulle parce qu'il n'y a aucune dépendance. D'ordinaire, la valeur positive est très petite dans ce cas, mais il y a toujours des cas douteux qui sont quelquefois les plus intéressants. Dans ces circonstances il serait très important de posséder une mesure de l'incertitude du résultat obtenu, c'est-à-dire de connaître l'erreur moyenne. Malheureusement le calcul exact de l'erreur moyenne d'une mesure de dépendance dérivée des observations est non seulement assez compliqué, mais le résultat doit contenir des fonctions dont on connaît seulement les valeurs observées. De plus, le calcul approché a été praticable jusqu'ici, seulement dans le cas où les observations possèdent déjà une assez grande précision, de sorte que le besoin d'une mesure de l'incertitude du résultat tiré des observations, ne s'est pas fait sentir. Ce cas se présente par exemple lorsque le nombre N est si grand qu'on peut avec avantage développer suivant les puissances de $\frac{1}{N}$; il se présente encore lorsqu'on néglige le carré des différences entre les valeurs théoriques et empiriques (1). Peut-être y a-t-il dans la manière même de poser le problème de l'exactitude, un cercle vicieux ; en effet, dans le cas où la mesure de dépendance déduite des observations est douteuse, son erreur moyenne déduite des mêmes observations le sera nécessairement, et à plus forte raison si l'on considère les approximations grossières employées. Dans ces circonstances, nous recommandons d'appliquer la mesure empirique Ψ seulement dans les cas où le nombre total des observations est assez grand, et de se méfier toujours des petites valeurs trouvées pour Ψ.

Sur la probabilité pour que la descendance d'une personne s'éteigne.

1. — Le mathématicien danois A. K. ERLANG s'est occupé du problème suivant dont il ne connaissait pas la solution complète :

Soit a_n la probabilité pour qu'un nouveau-né ait n enfants ; on a évidemment $a_0 + a_1 + a_2 + \ldots = 1$. Quelle est la probabilité pour que sa descendance s'éteigne ?

La question avait été soumise aux lecteurs du *Matematisk Tidsskrift* (2) après la mort de M. ERLANG, et j'avais déjà publié une petite note en danois sur ce sujet (3), quand le statisticien anglais M. W. PALIN ELDERTON appela mon attention sur le fait que le même problème avait été posé déjà en 1875 par F. GALTON et une solution donnée par le révérend H. W. WATSON dans un mémoire intitulé *On the Probability of the Extinction of Families* (4).

Ni la solution d'ERLANG, ni celle de WATSON n'étaient cependant complètes ; il restait ensuite à démontrer l'identité des deux solutions et l'unicité de la solution, ce qui donne lieu à une analyse assez inté-

(1) Felix BERNSTEIN : *Über den mittleren Fehler der Potenzmomente* (Zeitschrift für die gesammte Versicherungswissenschaft, vol. XXX, pp. 365-377).
(2) *Matematisk Tidsskrift*, B. 1929, p. 36, problème 15.
(3) *Ibid.*, B. 1930, p. 19.
(4) *Journal of the Anthropological Institute*, vol. IV, 1875, p. 138. Voir aussi Francis GALTON : *Natural Inheritance*, 1889, p. 242.

ressante. C'est pourquoi j'ai cru utile de m'occuper encore une fois de ces questions, en ajoutant quelques résultats nouveaux.

Le problème de GALTON était réellement celui de l'extinction des *noms de famille*; mais, en ce qui concerne le traitement mathématique, peu importe si a_n signifie la probabilité pour qu'une personne ait n *enfants* ou pour qu'un mâle ait n *fils*, la seule différence étant l'interprétation du résultat. On peut même aller plus loin et parler de la descendance des êtres quelconques au lieu de personnes ; j'ai pourtant préféré garder la terminologie d'ERLANG, la généralisation étant évidente.

Or, en considérant la manière dont le problème a été posé, on voit tout de suite que plusieurs hypothèses ont été admises tacitement, et qu'il faut faire certaines réserves.

D'abord, l'énoncé admet l'*existence* de la probabilité cherchée ; mais cette existence n'est pas évidente, puisque l'infini entre en considération, ainsi que nous le verrons. Pour l'instant nous nous contenterons d'examiner ce que doit être la probabilité, lorsqu'elle existe.

Ensuite, il a été admis que les probabilités a_n restent *constantes dans les générations successives*. Cette hypthèse est nécessaire, tant qu'on ne sait pas par quoi la remplacer ; mais elle ne peut être vraie qu'en moyenne puisque le croisement avec des familles d'une plus grande ou plus faible fertilité doit influencer la fertilité future. Toutefois, il est naturel de commencer par chercher la solution qui serait correcte pour une population homogène avec une fertilité assez stable.

Finalement, nous aurons à appliquer les règles ordinaires pour la multiplication des probabilités, quoique les événements dont il s'agit ne sont peut-être pas toujours entièrement *indépendants*. C'est d'ailleurs le cas dans un grand nombre d'applications du calcul des probabilités aux problèmes de biologie, et l'objection perd en force s'il s'agit de la descendance d'une population au lieu de celle d'une personne.

Négligeons donc ces scrupules en première approximation ; on voit alors immédiatement que, si l'on désigne la probabilité inconnue par x, on a la relation suivante

$$x = a_0 + a_1 x + a_2 x^2 + \cdots \qquad (\Sigma a_n = 1). \tag{1}$$

En effet la probabilité cherchée se compose des probabilités : pour qu'il n'y ait aucun enfant ; pour qu'il y ait un enfant dont la descendance s'éteingne ; pour qu'il y aient deux enfants dont les descendances s'éteignent ; et ainsi de suite.

La relation (1), qui a été proposée par M. ERLANG lui-même, ne suffit pas toujours pour répondre à la question posée. Si par exemple, $a_1 = 1$, de sorte que tous les autres a_n s'annulent, l'équation (1) se réduit à l'identité $x = x$ qui ne contribue pas à déterminer l'inconnue, laquelle est nulle évidemment, dans ce cas particulier. Il est vrai que ce cas est un cas extrême ; mais une difficulté plus sérieuse se présente : en effet, il y a toujours la racine $x = 1$ (à cause de $\Sigma a_n = 1$) et souvent une autre racine comprise entre 0 et 1 qu'on peut, par conséquent, interpréter comme une probabilité. Comment faut-il alors choisir entre ces deux racines ?

Avant de montrer que l'on peut poser la question de telle manière qu'il n'y ait qu'une seule réponse possible, nous allons examiner les circonstances dans lesquelles l'équation (1) possède deux racines appartenant à l'intervalle de 0 à 1. Puisque seulement les racines de cet intervalle nous intéressent, la convergence dela série $\Sigma a_n x^n$ est assurée à cause de $\Sigma a_n = 1$, les coefficients a_n étant positifs ou nuls. Il n'y a donc pas besoin de discuter la question délicate s'il y a une limite du nombre d'enfants qu'une personne puisse avoir ; les nombres très grands auront en tout cas une probabilité très faible, comme dans la théorie des erreurs.

Si on introduit dans (1) $a_0 = 1 - a_1 - a_2 - \ldots$ on trouve

$$1 - x = a_1(1 - x) + a_2(1 - x^2) + a_3(1 - x^3) + \cdots$$

Pour examiner quelles sont les racines possibles en dehors de $x = 1$, divisons les deux membres par $1 - x$; on obtient

$$1 = a_1 + a_2(1 + x) + a_3(1 + x + x^2) + \cdots \tag{2}$$

et en faisant varier x d'une manière continue de 0 à 1, on voit que si

$$1 \leqslant a_1 + 2a_2 + 3a_3 + \cdots \tag{3}$$

il y a une seule racine appartenant à l'intervalle mentionné, et que cette racine est unique. On conclut de (1) que cette racine est $\geqslant a_0$.

La condition (3) peut être exprimée d'une manière qui met en évidence son sens biologique. En effet, en posant

$$m = a_1 + 2a_2 + 3a_3 + \cdots \tag{4}$$

on voit que m est *le nombre moyen d'enfants*, ou l'espérance mathématique du nombre d'enfants, et il est probable que cette constante est reliée d'une certaine façon à la durée probable de la famille. Il y a lieu de croire que M. ERLANG était convaincu que si $m > 1$ on doit préférer la plus petite des racines de l'équation (1) ; si, au contraire, $m < 1$ il y a seulement la solution $x = 1$ qui, dans le cas $m = 1$, est une racine double.

Il faut pourtant admettre que les réflexions précédentes ne constituent pas une démonstration mathématique ; c'est pourquoi nous allons attaquer le problème d'un autre côté, en le généralisant un peu, d'une manière qui nous donne la certitude de l'existence de la probabilité cherchée.

2. — Commençons par chercher *la fonction de fréquence de la* s^e *génération*, c'est-à-dire la probabilité que la s^e génération contienne n membres. Dans notre terminologie, les enfants seront la première génération, les petits-enfants la seconde génération, et ainsi de suite.

Soit $a_n^{(s)}$ la probabilité pour que la s^e génération contienne n membres, de sorte que $a_n^{(1)} = a_n$. On a évidemment

$$0 \leqslant a_n^{(s)} \leqslant 1, \qquad \sum_{n=0}^{\infty} a_n^{(s)} = 1. \tag{5}$$

Nous employerons les fonctions $f_s(x)$ définies par

$$f_s(x) = a_0^{(s)} + a_1^{(s)}x + a_2^{(s)}x^2 + \cdots \tag{6}$$

et nous écrirons $f_1(x) = f(x)$, de sorte que

$$f(x) = a_0 + a_1 x + a_2 x^2 + \cdots \tag{7}$$

La convergence de (6) dans l'intervalle de 0 à 1 résulte de (5).

Or, on peut exprimer les fonctions de fréquence d'une certaine génération, par exemple a_r^{s+1}, au moyen des fonctions de fréquence de la génération précédente, c'est-à-dire les $a_n^{(s)}$. On a, en effet, d'après un théorème bien connu du calcul des probabilités

$$a_r^{(s+1)} = \sum_{n=0}^{\infty} a_n^{(s)} \sum_{(\nu)} \frac{n!}{\nu_0!\,\nu_1! \cdots \nu_r!}, \tag{8}$$

la somme $\sum\limits_{(\nu)}$ portant sur toutes les valeurs de ν satisfaisant aux deux conditions suivantes

$$
(9) \qquad \left\{ \begin{aligned} \nu_0 + \nu_1 + \nu_2 + \cdots + \nu_r &= n, \\ \nu_1 + 2\nu_2 + \cdots + r\nu_r &= r. \end{aligned} \right.
$$

En effet, la somme $\sum\limits_{(\nu)}$ représente la probabilité pour que n personnes aient un nombre total de r enfants, en désignant par ν_k le nombre de personnes parmi les n qui auront k enfants.

Pour écrire (8) d'une manière plus commode, développons d'abord $f_s(f(x))$ en série entière, ce qui est évidemment toujours possible.

De (6) et (7), il résulte que

$$
f_s(f(x)) = \sum_{n=0}^{\infty} a_n^{(s)}(a_0 + a_1 x + a_2 x^2 + \cdots)^n.
$$

Pour trouver le coefficient de x^r dans le développement de $f_s(f(x))$, considérons le terme général du développement de l'expression

$$
(a_0 + a_1 x + \cdots + a_r x^r)^n
$$

où l'on a négligé dans la parenthèse, les puissances supérieures à r. Ce terme est

$$
\frac{n!}{\nu_0! \nu_1! \cdots \nu_r!}(a_0)^{\nu_0}(a_1 x)^{\nu_1} \cdots (a_r x^r)^{\nu_r}
$$

où

$$
\nu_0 + \nu_1 + \cdots + \nu_r = n.
$$

Parmi ces termes, il faut retenir seulement ceux pour lesquels on a

$$
\nu_1 + 2\nu_2 + \cdots + r\nu_r = r,
$$

de sorte que l'exposant de x soit égal à r.

On voit ainsi que le coefficient cherché est identique au second membre de (8) avec les conditions (9). On a, par conséquent,

$$
(10) \qquad a_r^{(s+1)} = \frac{1}{r!} D_{x=0}^r f_s(f(x))
$$

où D est le symbole de différentiation.

Nous pouvons maintenant écrire

$$f_s(f(x)) = \sum_{r=0}^{\infty} a_r^{(s+1)} x^r$$

ou

$$f_s(f(x)) = f_{s+1}(x). \tag{11}$$

On peut dire qu'on obtient $f_s(x)$ en *itérant* $f(x)$ s fois, $f(x)$ elle-même étant la première itérée.

La comparaison de (11) et de (10) montre qu'on a

$$a_n^{(s)} = \frac{1}{n!} D_{x=0}^n f_s(x). \tag{12}$$

On peut exprimer ce résultat ainsi :

La probabilité que la s^e génération contienne n membres est égale au coefficient de x^n dans le développement en série de puissances de la s^e itérée de $f(x)$.

Ceci constitue le résultat trouvé antérieurement par H. W. WATSON dans le cas où $f(x)$ est un *polynôme*. Il reste à montrer l'avantage qu'on peut tirer de ce résultat pour l'équation (1) d'ERLANG.

3. — Avant de chercher la probabilité pour que la descendance s'éteigne, considérons la probabilité que la s^e génération n'existe pas. On obtient cette probabilité en posant $n = 0$ dans (12) ; elle est donc

$$a_0^{(s)} = f_s(0). \tag{13}$$

Si l'on écrit, pour abréger

$$x_s = a_0^{(s)}, \tag{14}$$

on obtient au moyen de (13)

$$x_s = f_s(0) = f(f_{s-1}(0)) = f(a_0^{(s-1)}) = f(x_{s-1}),$$

de sorte que

$$x_{s+1} = f(x_s). \tag{15}$$

Au moyen de cette relation, on peut calculer successivement x_2, x_3, x_4,... en commençant par $x_1 = a_0$. Si cette suite converge, elle donnera évidemment une racine de l'équation d'ERLANG qui s'écrit

$$x = f(x). \tag{16}$$

La question de la convergence de l'itération des fonctions analytiques a été étudiée particulièrement en France par M. KOENIGS, puis par MM. LÉAU, FATOU et JULIA, et en Norvège par M. TAMBS LYCHE. On peut prévoir que cette branche intéressante de l'analyse pourra devenir un instrument important pour l'analyse de certaines questions du calcul des probabilités et de statistique mathématique. Cependant quelques réflexions élémentaires suffiront pour notre but particulier.

Pour démontrer la convergence de la suite x_1, x_2, x_3,..., observons d'abord que x_s étant une probabilité, est toujours comprise entre 0 et 1. Par conséquent, la suite x_1, x_2, x_3,... sera, convergente si elle est monotone. Or, $f(x)$ est une fonction croissante dans l'intervalle de 0 à 1. Par conséquent, si $x_s < x_{s+1}$ on a $f(x_s) < f(x_{s+1})$, c'est-à-dire $x_{s+1} < x_{s+2}$. Il reste seulement à démontrer que $x_1 < x_2$. Puisque $x_1 = a_0$ on a

$$x_2 = f(a_0) = a_0 + a_1 a_0 + a_2 a_0^2 + \cdots$$

de sorte que $x_2 > x_1$ (à l'exception du cas banal où $a_0 = 1$, $x_s = 1$). La suite x_s va donc en croissant vers sa limite, à l'exception du cas $x_s = 1$.

Nous allons encore démontrer que x_s tend vers la *plus petite* racine de (16) appartenant à l'intervalle de a_0 à 1. Pour le prouver, il suffit de montrer qu'il n'y a aucune racine entre x_s et x_{s+1}. Soit donc x une racine, et retranchons (15) de (16) ; il vient

$$x - x_{s+1} = f(x) - f(x_s),$$

et puisque $f(x)$ est une fonction croissante, il en résulte que si $x_s < x$, on aura $x_{s+1} < x$.

D'autre part, il est évident que l'équation (1) ne peut pas avoir une racine plus petite que a_0, comme nous l'avons déjà fait observer.

Pour résumer, nous avons démontré que la probabilité

$$\lim_{s \to \infty} a_0^{(s)} \tag{17}$$

existe, et qu'elle est la *plus petite* racine, dans l'intervalle de a_0 à 1, de l'équation (1) d'ERLANG. Il semble naturel de donner à la probabilité (17) le nom de « la probabilité pour que la descendance d'une personne s'éteigne » ; du moins, il semble difficile de donner un autre sens précis à ce nom. Nous avons ainsi complètement résolu le problème de Watson-Erlang.

4. — Les coefficients $a_n^{(s)}$ sont, comme nous l'avons déjà dit, les *fonctions de fréquence* des générations successives, et nous allons les examiner d'un peu plus près à ce point de vue. Par la méthode usuelle on peut calculer les paramètres caractéristiques de ces distributions, tels que les moments, les semi-invariants, etc. La classe de paramètres la plus accessible dans le cas des fonctions de fréquence discontinues est constituée en général, par les *moments factoriels* (¹). En posant $n^{(r)} = n\,(n - 1)\ldots(n - r + 1)$, on a pour les moments factoriels de la première génération

$$\sigma_{(r)} = \sum_{n=r}^{\infty} n^{(r)} a_n, \tag{18}$$

et pour ceux de la s^e génération

$$\sigma_{(r)}^{(s)} = \sum_{n=r}^{\infty} n^{(r)} a_n^{(s)}, \tag{19}$$

de sorte que

$$\sigma_{(r)}^{(1)} = \sigma_{(r)}.$$

On obtient alors en différentiant (6) r fois

$$f_s^{(r)}(x) = \sum_{n=r}^{\infty} n^{(r)} a_n^{(s)} x^{n-r} \; ; \tag{20}$$

il faut observer qu'en écrivant $f_s^{(r)}(x)$, les r différentiations *suivent* les s itérations, et que les opérations de différentation et d'itération ne sont pas commutatives.

Si l'on pose $x = 1$, dans (20), on obtient

$$\sigma_{(r)}^{(s)} = f_s^{(r)}(1), \tag{21}$$

d'où, en particulier,

$$\sigma_{(r)} = f^{(r)}(1). \tag{22}$$

Par des différentiations successives de (11), on peut exprimer les moments de la s^e génération au moyen des moments de la première génération. Ainsi, en prenant la première dérivée, on obtient d'abord

$$f_s'(f(x)) f_1'(x) = f_{s+1}'(x). \tag{23}$$

(1) Voir, par exemple, J. F. Steffensen : *Some Recent Researches in the Theory of Statistics and Actuarial Science*, pp. 41-42. J'écris d'ordinaire, $\bar{\sigma}_{(r)}$ pour les valeurs théoriques, et $\sigma_{(r)}$ pour les valeurs observées ; ici il n'y a pas lieu de faire cette distinction.

Si l'on pose $x = 1$, on trouve, puisque $f(1) = 1$, $f'(1) = m$,

$$f'_{s+1}(1) = mf'_s(1),$$

et par des applications successives de cette relation

$$f'_{s+1}(1) = m^s f'_1(1) = m^{s+1},$$

de sorte que, d'après (21),

$$\sigma^{(s)}_{(1)} = m^s. \tag{24}$$

Ceci revient à dire que *l'espérance mathématique du nombre des membres de la s^e génération est m^s*, résultat qu'on pourrait d'ailleurs prévoir. En effet, en partant de N individus et en observant que m est le nombre moyen d'enfants *par individu*, on voit que l'espérance mathématique totale des enfants est Nm, des petits-enfants Nm^2, et ainsi de suite.

Soit, en particulier, $m = 1$; dans ce cas l'espérance mathématique du nombre des membres de la s^e génération est indépendante de s et égale à l'unité. Pourtant, nous avons déjà vu que si $m = 1$, la probabilité pour que la descendance s'éteigne est égale à l'unité. On voit combien on doit se méfier de l'intuition dans ces questions.

On obtient ensuite, en différentiant (23),

$$f''_s(f(x))[f'(x)]^2 + f'_s(f(x))f''(x) = f''_{s+1}(x), \tag{25}$$

d'où, en posant $x = 1$,

$$f''_{s+1}(1) = f''_s(1)[f'(1)]^2 + f'_s(1)f''(1),$$

de sorte que, par (21) et (24),

$$\sigma^{(s+1)}_{(2)} = \sigma^{(s)}_{(2)}m^2 + \sigma_{(2)}m^s.$$

On démontre par des applications successives de cette formule ou par induction, que

$$\sigma^{(s)}_{(2)} = \frac{m^s(m^s - 1)}{m(m-1)}\sigma_{(2)}. \tag{26}$$

On trouve ensuite, par les formules liant les moments factoriels aux semi-invariants [1], en désignant le carré de l'erreur moyenne par $\mu_2^{(s)}$,

$$\mu_2^{(s)} = \frac{m^s(m^s - 1)}{m(m - 1)} \mu_2. \tag{27}$$

L'espérance mathématique du nombre des membres de la s^e génération avec son erreur moyenne est ainsi, avec la valeur (27) de $\mu_2^{(s)}$,

$$m^s \pm \sqrt{\mu_2^{(s)}}. \tag{28}$$

Cette formule est valable pour la descendance d'un seul individu. Si nous pàrtons de N individus nouveau-nés, la formule correspondante sera

$$\mathrm{N}m \pm^s \sqrt{\mathrm{N}\mu_2^{(s)}}. \tag{29}$$

5. — Il semble désirable d'appliquer la théorie précédente à un cas particulier où il soit possible de former effectivement les itérées $f_s(x)$. Le choix des fonctions $f(x)$ sera ainsi très limité, d'autant plus qu'il faudra choisir une fonction $f(x)$ qui pourrait au besoin représenter une distribution réelle. M. ELDERTON m'a suggéré le cas où les probabilités a_n pourraient être, dans certaines conditions idéales, en progression géométrique. Élargissant cette hypothèse je vais faire une exception pour a_0 qui n'entrera pas dans la progression géométrique. Si insignifiante que paraisse cette modification, elle permet pourtant de représenter les faits d'une façon beaucoup plus satisfaisante, en tenant compte des cas de stérilité extraordinaire.

Posons donc,

$$a_0 = \frac{c}{\lambda}, \qquad a_n = \frac{(\lambda - c)(\lambda - 1)}{\lambda^{n+1}}, \qquad (n > 0). \tag{30}$$

c et λ désignant des constantes.

Nous avons supposé ici

$$\lambda > 1, \qquad \lambda \geqslant c \geqslant 0, \tag{31}$$

et la condition $\sum_0^\infty a_n = 1$ est satisfaite.

(1) Voir, par exemple, STEFFENSEN : *l. c.*, p. 45. La formule à employer est

$$\mu_2 = \sigma_{(2)} + \sigma_{(1)} - \sigma_{(2)}^2.$$

Au moyen des valeurs (30) on obtient à partir de (7)

$$f(x) = 1 + (\lambda - c)\left(\frac{\lambda - 1}{\lambda - x} - 1\right). \tag{32}$$

On tire aussitôt de cette équation

$$f'(1) = m = \frac{\lambda - c}{\lambda - 1}. \tag{33}$$

Pour obtenir les itérées de $f(x)$, il convient d'abord d'éliminer λ de (32), au moyen de (33). Puisque, par la dernière équation,

$$\lambda = 1 + \frac{1 - c}{m - 1}, \tag{34}$$

(32) devient

$$f(x) = 1 + \frac{m(1 - c)}{m - 1}\left(\frac{1 - c}{m - c - (m - 1)x} - 1\right). \tag{35}$$

On démontre alors que

$$f_s(x) = 1 + \frac{m^s(1 - c)}{m^s - 1}\left(\frac{1 - c}{m^s - c - (m^s - 1)x} - 1\right). \tag{36}$$

On peut réduire la démonstration à la solution d'une équation aux différences finies d'une type connu ([1]), mais on peut aussi procéder par induction, de la manière suivante. Remarquons d'abord que (36) est satisfaite pour $s = 1$, parce qu'elle se réduit alors à (35). Nous devons donc prouver uniquement que si (36) est satisfaite pour une valeur particulière de s, elle est aussi satisfaite pour la valeur suivante. Or, substituons dans (36), $f(x)$ à x ; le premier membre devient $f_s(f(x)) = f_{s+1}(x)$. Quant au second membre, on prouve d'abord, au moyen de (35), que

$$\frac{1 - c}{m^s - c - (m^s - 1)f(x)} = \frac{(m^s - 1)m}{m^{s+1} - 1} \cdot \frac{1 - c}{m^{s+1} - c - (m^{s+1} - 1)x} + \frac{m - 1}{m^{s+1} - 1} ;$$

puis, un calcul facile dont je puis omettre les détails montre qu'on obtient au second membre de (36), une expression de la même forme où s a été remplacée par $s + 1$. La démonstration est donc complète.

1) BOOLE : *Finite Differences*, third edition, p. 295.

En dérivant (36), on obtient

$$(37) \qquad f_s^{(r)}(x) = \frac{r!(m^s - 1)^{r-1} m^s (1-c)^2}{[m^s - c - (m^s - 1)x]^{r+1}} \qquad (r > 0)$$

d'où, pour $x = 1$,

$$(38) \qquad f_s^{(r)}(1) = r! m^s \left(\frac{m^s - 1}{1 - c}\right)^{r-1} \qquad (r > 0).$$

Pour calculer l'erreur moyenne, remarquons d'abord qu'on a

$$\mu_2 = \sigma_{(2)} + \sigma_{(1)} - \sigma_{(1)}^2$$

ou, par (22) et (38),

$$\begin{aligned} \mu_2 &= f''(1) + f'(1) - [f'(1)]^2 \\ &= 2m\frac{m-1}{1-c} + m - m^2, \end{aligned}$$

c'est-à-dire

$$(39) \qquad \mu_2 = m(m-1)\,\frac{1+c}{1-c}.$$

Pour la s^e génération on a donc, d'après (27),

$$(40) \qquad \mu_2^{(s)} = m^s(m^s - 1)\,\frac{1+c}{1-c}$$

avec la valeur (33) de m.

La probabilité que la s^e génération contienne n membres est égale au coefficient de x^n dans le développement de (36), soit

$$(41) \qquad a_n^{(s)} = m^s \left(\frac{1-c}{m^s - c}\right)^2 \left(\frac{m^s - 1}{m^s - c}\right)^{n-1} \qquad (n > 0),$$

$$(42) \qquad a_o^{(s)} = c\,\frac{m^s - 1}{m^s - c}.$$

On vérifie facilement que

$$\sum_{n=0}^{\infty} a_n^{(s)} = 1.$$

La probabilité pourque la descendance s'éteigne est donnée par (17). Or, si $m < 1$, on tire de (42)

$$(43) \qquad \lim_{s \to \infty} a_0^{(s)} = 1 \qquad (m < 1),$$

comme on devait le prévoir.

D'autre part, si $m = 1$ on a, d'après (33), $c = 1$, et (42) devient indéterminée. Dans ce cas, il est préférable d'éliminer c par (33), donc

$$c = 1 - (m - 1)(\lambda - 1)\ ;$$

on peut mettre ainsi (42) sous la forme

$$a_0^{(s)} = \frac{1 - (m - 1)(\lambda - 1)}{1 + \dfrac{m - 1}{m^s - 1}(\lambda - 1)},$$

et on voit que, pour $m \to 1$,

(44)
$$a_0^{(s)} = \frac{1}{1 + \dfrac{\lambda - 1}{s}}, \qquad (m = 1).$$

On tire de cette relation

(45)
$$\lim_{s \to \infty} a_0^{(s)} = 1, \qquad (m = 1)$$

en accord avec la théorie générale.

Si, enfin, $m > 1$, on tire de (42)

(46)
$$\lim_{s \to \infty} a_0^{(s)} = c. \qquad (m > 1).$$

Ce dernier résultat montre, en particulier, que si *toutes* les probabilités a_n (a_0 incluse) sont en progression géométrique, la probabilité pour que la descendance s'éteigne est égale à $\frac{1}{m}$. En effet, dans ce cas, on aura, d'après (30),

$$\lambda = \frac{a_0}{a_1} = \frac{c}{\lambda} \cdot \frac{\lambda^2}{(\lambda - c)(\lambda - 1)},$$

d'où $\lambda = 1 + c$, de sorte que (34) donne $c = \frac{1}{m}$.

Observons encore que la solution (46) est obtenue en accord avec la théorie générale, comme la plus petite racine de l'équation $x = f(x)$, c'est-à-dire, d'après (32),

(47)
$$x = 1 + (\lambda - c)\left(\frac{\lambda - 1}{\lambda - x} - 1\right).$$

En effet les racines de cette équation sont égales à $x = 1$ et $x = c$; mais puisque nous avons supposé $m > 1$, il résulte de (33) que $c < 1$.

6. — Avant de terminer, je dirai encore un mot sur l'hypothèse que nous avons faite et d'après laquelle toute les probabilités a_n, à l'exception de a_0, sont en progression géométrique. J'ai eu la chance de recevoir, quelques jours avant cette conférence, deux notes très récentes de M. A. J. LOTKA (1) donnant les valeurs des probabilités pour qu'un mâle nouveau-né ait n fils, déduites du recensement de la population des États-Unis en 1920. Ces valeurs s'accordent très bien avec l'hypothèse mentionnée ci-dessus. Je renvoie pour les détails, aux articles de M. LOTKA, en ajoutant pourtant que la table qu'il dresse des probabilités $a_n^{(s)}$ au moyen d'itérations numériques s'obtient plus facilement par les formules directes (41) et (42) qu'il ne connaissait pas.

Il ne faut pourtant pas s'attendre à ce que l'harmonie soit toujours aussi parfaite, surtout s'il s'agit d'une statistique, non pas d'un pays entier mais d'un groupe particulier de personnes possédant une fertilité spéciale. Un exemple m'a été fourni par M. H. C. J. NYBÖLLE, traitant d'une famille danoise où l'hypothèse (30) ne s'applique pas bien. Ces résultats sont pourtant d'une nature trop particulière pour trouver place ici.

(Conférences faites à l'Institut Henri-Poincaré en décembre 1931).

(Manuscrit reçu le 12 février 1932).

(1) Alfred J. LOTKA : *The Extinction of Families*, I-II. *Journal of the Washington Academy o Sciences*, vol. XXI, oct. nov. 1931.