ANNALES DE L'I. H. P.

R.A. FISHER

Conclusions fiduciaires

Annales de l'I. H. P., tome 10, n° 3 (1948), p. 191-213
<http://www.numdam.org/item?id=AIHP_1948__10_3_191_0>

Conclusions fiduciaires

par

R. A. FISHER.

Introduction. — Le thème général de mes remarques est le problème compliqué du raisonnement inductif, dont le but est d'établir un code de processus mental pour raisonner du particulier au général, analogue au code plus ou moins acceptable auquel nous nous référons lorsque nous parlons de raisonnement rigoureux du type déductif le plus familier.

Les conclusions inductives sont des conclusions typiquement incertaines, mais ce n'est pas, je pense, une raison pour qu'elles ne soient établies rigoureusement, pourvu que les conclusions tirées comportent une spécification adéquate et une mesure de l'incertitude entraînée. *Un* type d'incertitude est spécifié de façon convenable par le concept de probabilité mathématique. Dans les applications aux jeux de hasard concrets, par exemple, on sait que si le matériel n'est pas faussé et s'il est utilisé correctement, la probabilité d'un résultat ou d'une suite de résultats peut être calculée rigoureusement, bien que le résultat d'un coup particulier soit incertain.

Les premiers auteurs étudiant la logique des probabilités ont mis en évidence un point de la plus grande importance, qui est très facilement négligé, à savoir : que la probabilité est une qualité, non d'un événement particulier en soi, mais d'un événement regardé comme membre typique d'une *population* bien définie d'événements. La probabilité qu'il pleuve dans le Comté de Cambridge est conçue en rapport avec la fréquence de pluie sur le Comté à toutes heures du jour et époques de l'année. Cette fréquence n'est naturellement pas nécessairement la même que celle de la pluie à 9 heures du matin, sur le bourg de Cambridge, bien que l'événement particulier auquel le concept de probabilité est appliqué

doive toujours en fait, avoir quelque spécification particulière dans le temps et l'espace. Dans les applications pratiques, nous sommes guidés par le fait que les probabilités arrivent à être bien déterminées et soient apparemment applicables au cas où le guide est nécessaire. Mais dans les applications mathématiques, nous sommes dans l'obligation de tenir compte de toutes les circonstances particulières qui importent dans notre problème.

Ceci m'amène à une considération plus générale. Dans un raisonnement déductif, il est juste de choisir à volonté quelques-uns de nos axiomes valables et de tirer les conclusions que nous pouvons en déduire sans considérer d'autres axiomes. L'existence d'autres axiomes dans notre système ne doit pas exclure une telle solution. Mais dans un raisonnement fondé sur des données expérimentales, il serait tout à fait illégitime de faire une sélection des renseignements que nous désirons utiliser et d'en tirer des conclusions, comme s'ils existaient seuls. C'est naturellement ce que font invariablement les politiciens, et c'est ce qui fait dire que n'importe quoi peut être prouvé par des statistiques. En fait, pour un raisonnement inductif une restriction supplémentaire s'ajoute à celles exigées par un processus purement déductif, à savoir que la totalité des observations particulières doit être prise en considération tant que ces observations importent pour un point de la question.

En statistique mathématique nous cherchons à tirer des conclusions rigoureuses au sujet de certaines quantités hypothétiques inconnues, habituellement appelées paramètres, ou à tirer des conclusions indépendantes de ces paramètres inconnus. Dans certains cas, je pense que de telles conclusions peuvent être rigoureusement exprimées en termes de probabilité, mais la base du raisonnement qui y conduit est différente de celle qui fut mise en avant originairement par Bayes au XVIII^{e} siècle et qui imprégna les manuels, au moins en ce siècle. Pour distinguer les probabilités actuellement utilisées de celles de Bayes, dont elles diffèrent par le contenu logique mais auxquelles elles ressemblent assez étroitement au point de vue de l'exposé formel, elles ont été appelées probabilités fiduciaires.

J'ai choisi quelques exemples pour illustrer les méthodes de raisonnement, mais je pense que le temps est passé où un raisonnement pouvait être décrit comme s'imposant au sens absolu. Il me semble, que, en réalité, nous sondons les possibilités que l'esprit humain a de raisonner

rigoureusement dans un champ nouveau, et ce raisonnement ne peut s'imposer que pour ceux qui ont admis auparavant d'accepter certaines formes typiques ou canoniques de raisonnement auxquelles n'importe quel cas particulier peut être comparé. En extrême logique je peux me tromper, mais au moins là où toutes les questions fondamentales sont en discussion, la seule direction pratique qui nous est ouverte, est de concevoir clairement le processus intellectuel exact d'une méthode et ensuite de peser, considérer, critiquer et finalement décider si la méthode est ou non acceptable, si j'ose dire, pour notre conscience scientifique.

Le type de relation mathématique qui semble rendre possible des énoncés probabilistes déduits des données et absolument indépendants de tous les paramètres inconnus peut être illustré dans ce qui suit où j'ai employé le mot « statistique » pour toute quantité calculable à partir des données sans information extrinsèque, que cette quantité doive ou non être utilisée pour une estimation, et où j'ai supposé que le nombre des paramètres inconnus était supérieur à un.

Tout ensemble de statistiques dont la distribution simultanée est indépendante des paramètres, est appelé un ensemble « ancillaire » de statistiques.

Le plus petit ensemble de statistiques tel que, lorsqu'il est donné, la distribution de toute statistique fonctionnellement indépendante soit indépendante des paramètres est appelé un ensemble exhaustif de statistiques.

Tout ensemble de fonctions faisant intervenir tous les paramètres en même temps que les statistiques de l'ensemble exhaustif et possédant une distribution simultanée indépendante des paramètres, est appelé un ensemble de quantités pivotales.

Ce sont de telles relations mathématiques qui semblent ouvrir la porte à des inférences de la nature recherchée, sans recours à aucune probabilité à priori, comme Bayes a dû le faire. J'espère que mes exemples illustreront ces idées en pratique.

1. **Perte de précision due à une pondération inexacte.** — En vue d'introduire et de simplifier l'outil mathématique de la théorie des probabilités fiduciaires ou, pourrait-on dire, de la théorie inverse des

variables aléatoires, considérons le problème classique d'une pondération inexacte.

Supposons que nous possédions N observations $x_1, \ldots, x_N$, chacune de précision connue; ainsi x_i a une variance σ_i^2 ou, si nous préférons, une invariance $w_i = \frac{1}{\sigma_i^2}$. C'est un résultat connu que la moyenne est définie avec la plus grande précision possible par :

$$\hat{x} = \frac{\sum wx}{\sum w}.$$

On a facilement la variance de $\hat{x}$

$$V(\hat{x}) = \frac{\sum [w^2 V(x)]}{\left[\sum (w)\right]^2}$$

et en remplaçant $V(x)$ par $\frac{1}{W}$

$$V(\hat{x}) = \frac{1}{\sum (w)}.$$

Si cependant, les poids réels nous sont inconnus ou si les connaissant, nous les avons négligés et utilisé la moyenne arithmétique

$$\bar{x} = \frac{\sum x}{N},$$

nous trouverons :

$$V(\bar{x}) = \frac{\sum V(x)}{N^2} = \frac{\sum \frac{1}{w}}{N^2},$$

et en comparant ces deux variances on voit que :

$$\frac{V(\hat{x})}{V(x)} = \frac{N^2}{\sum \frac{1}{w} \sum w}$$

et inférieur à 1, à moins que toutes les valeurs de w ne soient égales, auquel cas les deux moyennes sont équivalentes.

Un exemple concret éclairera ce résultat général. Supposons que la variance vraie $\frac{1}{w}$ ait même répartition qu'une variable aléatoire donnée

$$w = C\chi^2,$$

où χ^2 est une variable aléatoire ayant la probabilité élémentaire

$$df = \frac{1}{\frac{n-2}{2}!}\left(\frac{1}{2}\chi^2\right)^{\frac{n-2}{2}} e^{-\frac{1}{2}\chi^2} d\left(\frac{1}{2}\chi^2\right).$$

On voit aisément que :

$$E\left(\frac{1}{2}\chi^2\right) = \frac{n}{2}$$

et

$$E\left(\frac{2}{\chi^2}\right) = \frac{2}{n-2} \qquad \text{si} \quad n > 2$$

ou infinie dans les autres cas,
et que

$$\frac{1}{E\left(\frac{1}{2}\chi^2\right) E\left(\frac{2}{\chi^2}\right)} = \frac{n-2}{n} \qquad \text{si} \quad n > 2$$

et nul dans les autres cas.

Cette quantité est le rapport entre la variance correspondant à des poids corrects et celle correspondant à des poids égaux dans le cas considéré.

Le point que je désire établir maintenant est que ce rapport, égal à $\frac{n-2}{n}$, est en réalité le produit de deux facteurs attribuables à des causes différentes. En effet, en prenant la moyenne arithmétique et en donnant aux différentes observations des poids égaux, nous n'avons pas simplement négligé les poids réels, ce qui, du fait du manque d'observations appropriées, pourrait être inévitable, mais nous avons aussi fondé notre estimation sur l'emploi de la moyenne arithmétique; or, ce procédé pourrait être amélioré par la connaissance de la distribution réelle des poids même si l'observation de poids différents est impossible.

En fait, un ensemble d'observations ayant chacune une loi d'erreur

*

normale, de variance σ^2 telle que $\frac{1}{\sigma^2} = c\chi^2$ aurait la répartition

$$df = \int_{\chi^2=0}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}C\chi^2 x^2} k\sqrt{C}\, dx \frac{1}{\frac{n-2}{2}!} \left(\frac{\chi^2}{2}\right)^{\frac{n-2}{2}} e^{-\frac{\chi^2}{2}} d\left(\frac{\chi^2}{2}\right),$$

$$df = \frac{\frac{n-1}{2}!}{\frac{n-2}{2}!} (1 + Cx^2)^{-\frac{n+1}{2}} \sqrt{\frac{C}{\pi}}\, dx.$$

Ce type de distribution est communément appelé distribution de « Student » bien que le raisonnement logique développé ici soit tout à fait différent de celui qui conduit à ce genre de distribution. Nous voyons que dans le problème présent, la simple ignorance des poids réels a remplacé un certain nombre de distributions normales par un type différent pour lequel la moyenne arithmétique n'est pas une estimation jouissant d'un mérite particulier.

Si nous déterminons la quantité d'information fournie par la nouvelle distribution par la formule très utile suivante :

$$i_\mu = \int_{-\infty}^{+\infty} \left(\frac{d}{d\mu} \log f\right)^2 df,$$

où μ représente la valeur vraie que nous cherchons et où $x - \mu$ remplace x dans l'expression de df, nous trouvons .

$$\begin{aligned} \int_{-\infty}^{+\infty} \left(\frac{d}{dx} \log f\right)^2 df &= \int \left[\frac{(n+1)Cx}{1+Cx}\right]^2 df \\ &= (n+1)^2 C \int \left[\frac{1}{1+Cx^2} - \frac{1}{(1+Cx^2)^2}\right] df \\ &= (n+1)^2 C \left[\frac{n}{n+1} - \frac{n(n+2)}{(n+1)(n+3)}\right] = nC\frac{n+1}{n+3}. \end{aligned}$$

La valeur moyenne W des poids étant nC, on voit que l'absence de distinction entre les observations de précision variable réduit l'information originairement utilisable dans le rapport $\frac{n+1}{n+3}$. En outre, la variance de la distribution de student est $C\frac{1}{(n-2)}$, de sorte que l'efficacité de la moyenne dans l'estimation de la vraie valeur μ, centre d'une

distribution de student est $\frac{(n-2)(n+3)}{n(n+1)}$; c'est le second facteur qui intervient dans la décomposition de $\frac{n-2}{n}$ en deux facteurs.

Le seul manque de la connaissance des poids réels réduit donc la précision possible dans le rapport $\frac{n+1}{n+3}$ $\left(\text{égal à } \frac{3}{5} \text{ pour } n=2\right)$, ce qui constitue une réduction beaucoup plus satisfaisante que le rapport $\frac{n-2}{n}$.

2. **Erreur d'échantillonnage aléatoire dans l'estimation d'un coefficient de régression.** — Si nous avons deux variables x et y et si nous savons que y a une loi de répartition normale de variance connue σ^2 autour d'une valeur moyenne donnée en fonction de x par $\alpha+\beta x$, nous pouvons estimer la valeur de β à partir de N groupes (x, y), choisi au hasard, en calculant :

$$A = \Sigma(x-\bar{x})^2,$$
$$B = \Sigma(x-\bar{x})y$$

et prenant pour β

$$b = \frac{B}{A}.$$

En discutant la précision d'une telle estimation b, j'ai suggéré il y a plus de vingt ans, qu'une bonne méthode, en même temps très simple, est de considérer notre échantillon comme appartenant à une population d'échantillons aléatoires, ayant tous le même ensemble de valeurs $x_1 \ldots x_N$, de la variable indépendante.

Dans une telle population d'échantillons, il est aisé de voir que :

$$V(B) = V\left[\Sigma(x-\bar{x})y\right] = \sigma^2 \Sigma(x-\bar{x})^2 = A\sigma^2,$$

d'où :

$$V(b) = \frac{1}{A^2} V(B) = \frac{\sigma^2}{A}.$$

Avec cette restriction, b est réparti normalement autour de sa vraie valeur β, avec la variance connue $\frac{\sigma^2}{A}$. Nous pouvons donc tout de

suite étendre la population d'échantillons auxquels la même distribution s'applique, en disant que c'est la distribution de tous les échantillons ayant une valeur constante A, même si les valeurs de x_1 à x_N sont différentes et même indépendamment du nombre N de groupes constituant un échantillon, mais nous n'avons pas ainsi la distribution de notre estimation pour des échantillons correspondant à N donné et à A quelconque.

Si nous imaginons que la population d'échantillons à laquelle notre exposé s'applique (avec ses tests appropriés) est spécifiée uniquement par le nombre constant N, la distribution sera différente. En particulier elle dépendra de la variation de la quantité observée A, d'un échantillon à un autre et, par conséquent, de la distribution de la variable x, dont nous ne nous sommes pas occupés dans la méthode que je propose comme correcte.

En second lieu, si en fait la distribution de x était donnée comme partie des données du problème, cette information serait inutile dans l'estimation de b et en particulier dans la détermination des erreurs d'estimation auxquelles b est exposé.

Supposons, par exemple, que x a une loi de répartition normale de variance connue α; il est alors bien connu que la distribution de A, lorsque N est donné, est donnée par

$$A = \alpha\chi^2$$

avec $n = N - 1$ degrés de liberté.

Dans ce cas, la distribution de notre erreur $b - \beta$ pour tous les échantillons de grandeur N sera définie par la probabilité élémentaire :

$$df = \int_{\chi^2=0}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\alpha\chi^2 \frac{(b-\beta)^2}{2\sigma^2}} \frac{\chi\sqrt{\alpha}}{\sigma} db \frac{\left(\frac{1}{2}\chi^2\right)^{\frac{N-3}{2}}}{\frac{N-3}{2}!} e^{-\frac{1}{2}\chi^2} d\left(\frac{1}{2}\chi^2\right),$$

$$df = \frac{\frac{N-2}{2}!}{\frac{N-3}{2}!} \left[1 + \frac{(b-\beta)^2}{\sigma^2}\right]^{-\frac{N}{2}} \frac{1}{\sigma\sqrt{\pi}} \sqrt{\alpha}\, db.$$

Il est clair que si α était inconnu sauf par l'information que fournit la valeur A observée dans notre échantillon, ce pas serait rétrograde, car il remplacerait une distribution dépendant d'un paramètre connu A par

une autre dépendant d'un paramètre inconnu α. Même si α est connu, la nouvelle distribution donne au sujet de la valeur β ou de la signification de l'écart entre b et une valeur hypothétique, des informations plus réduites que celles fournies sur la moyenne par les distributions normales distinctes, dont elle est composée.

Comme dans l'exemple qui précède en négligeant la valeur de A véritablement observée dans notre échantillon et en nous appuyant par contre, sur la grandeur N de l'échantillon et notre connaissance de α, nous réduisons l'information dans le rapport

$$\frac{n+1}{n+3} = \frac{N}{N+2}.$$

La morale que je désire tirer de cet exemple est que la grandeur N de l'échantillon peut être sans importance pour certains tests demandés ; que d'autres caractères de l'échantillon doivent éventuellement être spécifiés ou que la spécification de la population d'échantillons à laquelle nos tests de signification sont appliqués, ne doit pas être écartée à la légère, par des phrases telles que « échantillonnage répété à partir de la même population », avec la signification sous-entendue que les échantillons doivent avoir la même grandeur. Au contraire, l'intelligence exacte de ce qui constitue la population des possibilités auxquelles nos propositions probabilistes s'appliquent, constitue, à mon sens, le noyau logique de ce processus de raisonnement inductif que doit offrir la théorie de l'estimation.

3. **Élimination d'une variance inconnue.** — Nous pouvons varier l'exemple précédent en supposant que y est distribué normalement autour de la valeur moyenne $\alpha + \beta x$, avec une variance σ^2 indépendante de x mais inconnue. Nous pouvons, selon le procédé habituel, utiliser les données pour obtenir une évaluation S^2 de la variance inconnue, soit :

$$s^2 = \frac{1}{N-2}\left(C - \frac{B^2}{A}\right),$$

où

$$C = \Sigma(y - \bar{y})^2;$$

l'estimation b, de β, est inchangée. Nous écrirons encore

$$b = \frac{B}{A},$$

mais la précision de b est fondamentalement affectée : l'erreur $b - \beta$ est en effet distribuée normalement autour de O, mais avec une variance inconnue $\frac{\sigma^2}{A}$. En suivant la méthode de « student » nous pouvons cependant trouver la répartition du rapport $\frac{b-\beta}{s}$.

Si nous posons

$$\frac{b-\beta}{s} = \frac{t}{\sqrt{A}},$$

la loi de répartition

$$df = \frac{1}{\sqrt{2\pi}} e^{-\frac{A(b-\beta)^2}{2\sigma^2}} \frac{\sqrt{A}}{\sigma} db$$

s'écrit

$$df = \frac{1}{\sqrt{2\pi}} e^{-\frac{s^2+2}{2\sigma^2}} \frac{s}{\sigma} dt,$$

multipliée par la probabilité élémentaire

$$df = \frac{1}{\frac{N-4}{2}!} \left[\frac{1}{2}(N-2)\frac{s^2}{\sigma^2}\right]^{\frac{N-4}{2}} e^{-\frac{(N-2)s^2}{2\sigma^2}} d\left[\frac{1}{2}(N-2)\frac{s^2}{\sigma^2}\right]$$

elle donne la probabilité composée t et $\frac{S}{\sigma}$.

Intégrons par rapport à $\frac{S}{\sigma}$, l'intégrale étant étendue à toutes les valeurs possibles de ce rapport; la distribution de t devient

$$\frac{\left[\frac{1}{2}(N-3)\right]!}{\left[\frac{1}{2}(N-4)\right]!} \left[1 + \frac{t^2}{N-2}\right]^{-\frac{N-1}{2}} \frac{dt}{\sqrt{\pi(N-2)}},$$

c'est la distribution de student pour $N-2$ degrés de liberté; cette distribution ne dépend que de N et a été naturellement mise en tables. Pour saisir la signification d'une déviation observée $b - \beta$ nous pouvons multiplier par $\frac{\sqrt{A}}{s}$ et comparer la valeur de t observée aux valeurs en

pour cent, données par la table. Ceci est naturellement mathématiquement équivalent à attribuer à $b - \beta$ la distribution obtenue en substituant $\frac{\sqrt{A}(b-\beta)}{s}$ à t dans la distribution de t donnée plus haut. Notons que maintenant s ainsi que A doit être traité comme un paramètre fixe de la loi d'erreurs.

J'ai dit mathématiquement équivalent, car une distinction logique doit être établie ici. Dans le cas du paramètre A, qui intervient comme une caractéristique de notre échantillon, nous pourrions imaginer que notre probabilité se rapporte à une population d'échantillons obtenue en répétant l'échantillonnage de la même population. Elle pourrait, il est vrai, ne pas être la population de tous ces échantillons, mais seulement de ceux qui correspondent à une valeur constante A et, avec A fixé, nous avons vu qu'il est inutile de stipuler que les échantillons ont la même grandeur N ou qu'ils sont tirés de populations ayant même distribution de x. Mais dans le cas de s il n'est même plus possible de supposer que nous avons affaire à une telle population choisie de façon à correspondre à une valeur s donnée, car une telle sélection nous ramènerait simplement à une distribution normale de variance inconnue $\frac{\sigma^2}{A}$. La population d'échantillons à laquelle nos résultats s'appliquent doit être tirée de populations ayant des valeurs différentes de σ^2, cette diversité étant conforme à la distribulion que nous avons imposée au rapport $\frac{s^2}{\sigma^2}$ et à la valeur connue de s^2.

Ceci vu, la question qui se pose est de savoir jusqu'à quel point cela vaut la peine d'essayer de relier la théorie des tests de signification à la conception d'échantillonnages répétés à partir d'une même population.

Pour commencer, la distribution de la quantité t qui est une fonction des statistiques observables A, b, s et du paramètre inconnu β, peut être reproduite par échantillonnages répétés, au hasard d'une population du type spécifié; mais également la distribution est valable si les différents échantillons sont tirés au hasard d'un certain nombre de populations différant par les valeurs de β, σ^2, etc., qui les caractérisent. Il n'y a ici aucune indication permettant de préciser que l'échantillonnage est répété à partir de la même population. Le seul paramètre de la distribution est N et nous devons exiger uniquement que tous les

seuls échantillons de la population à laquelle notre probabilité se rapporte aient la même grandeur N.

Nous pouvons cependant facilement lever cette restriction, car la distribution de la fonction

$$P = \int_{-\infty}^{t} \frac{\frac{N-3}{2}!}{\frac{N-4}{2}!} \left[1 + \frac{u^2}{N-2}\right]^{\frac{N-1}{2}} \frac{du}{\sqrt{\pi(N-2)}}$$

est la même pour des échantillons tirés de toutes populations et la même pour des échantillons de toutes grandeurs. Une simple transformation fonctionnelle suffit, en fait, pour lever toute limitation due aux caractéristiques particulières de notre population et la grandeur de l'échantillon. J'appellerai des quantités telles que t et P des quantités « pivotales ». Ce sont typiquement des fonctions à la fois des observations et des paramètres inconnus, ayant la propriété que leur distribution est indépendante de tous les paramètres et par conséquent du fait que les échantillons sont tirés de la même population ou de populations différentes.

La logique des méthodes que je discute consiste à reconnaître, dans la validité absolue de telles distributions, les moyens d'éliminer les paramètres inconnus ou, alternativement, de faire des affirmations probabilistes à leur sujet. De telles propositions étant tout à fait distinctes dans leur contenu logique des propositions de probabilités inverses ou probabilités de Bayes, nous les appellerons des énoncés de probabilités fiduciaires.

4. **Information ancillaire.** — Même lorsque la loi d'erreur n'est pas normale, on peut donner une illustration de la perte de précision, due à la non-utilisation de la totalité de l'information fournie par notre échantillon, relative aux erreurs auxquelles est sujette une estimation déduite d'un tel échantillon.

Considérons N groupes d'observations (x, y) et supposons que la distribution simultanée de x et y soit donnée en fonction d'un paramètre θ par la fréquence élémentaire

$$df = e^{-\left(\theta x + \frac{y}{\theta}\right)} dx\, dy.$$

I. *Estimation de* 0. — La somme des logarithmes des fréquences élémentaires est

$$-0\,X-\frac{Y}{0},$$

où X et Y représentent respectivement les sommes de x et y dans notre échantillon

$$X=\Sigma x_i, \qquad Y=\Sigma y_i.$$

Nous pouvons majorer cette expression en posant $0^2=T^2\equiv\frac{Y}{X}$ et cette solution de l'équation du maximum de vraisemblance sera nécessairement une estimation d'un bon rendement « efficient » bien que, nous le verrons, elle ne soit pas une estimation suffisante, « sufficient ».

II. *Quantité moyenne d'information dans le cas d'un seul couple d'observations.* — Celle-ci, que nous désignerons par i_0, est la valeur moyenne de

$$\left[\frac{\partial}{\partial 0}(\log df)\right]^2$$

ou de

$$\left[x-\frac{y}{0^2}\right]^2;$$

or

$$E(x^2)=\frac{2}{0^2}, \qquad E(xy)=1, \qquad E(y^2)=2\,0^2,$$

d'où

$$i_0=\frac{2}{0^2},$$

ou si l'on désire l'information relative à $\log 0$

$$i_{\log 0}=2.$$

III. *Distribution des estimations déduites d'échantillons correspondant à* N *donné.* — On voit facilement que la répartition simultanée de x et y est

$$df=\frac{X^{N-1}}{(N-1)!}\,\frac{Y^{N-1}}{(N-1)!}\,e^{-0X-\frac{Y}{0}}\,dX\,dY. \tag{1}$$

Posons

$$\frac{Y}{X} = T^2, \qquad X = \frac{U}{T}, \qquad dX\,dY = \frac{2U}{T}\,dU\,dT,$$
$$XY = U^2, \qquad Y = UT,$$

la distribution de T et U est alors pour N donné,

$$(2) \qquad df = 2\,e^{-U\left(\frac{T}{\theta}+\frac{\theta}{T}\right)}\frac{U^{2N-1}\,dU}{[(N-1)!]^2}\,\frac{dT}{T}.$$

Intégrons par rapport à U de o à $+\infty$; la distribution de T pour N donné quelconque est alors

$$(A) \qquad df = 2\,\frac{(2N-1)!}{[(N-1)!]^2}\left(\frac{\theta}{T}+\frac{T}{\theta}\right)^{-2N}\frac{dT}{T}.$$

Si nous calculons comme précédemment la valeur moyenne de l'information relative à θ fourni par une seule observation à partir de cette distribution, nous trouvons :

$$\frac{4N^2}{(2N+1)\theta^2} = \frac{2N}{2N+1}\,\frac{2N}{\theta^2}$$

qui montre que de la quantité d'information contenue dans les observations à l'origine, une fraction $\frac{1}{2N+1}$ a été perdue quand nous avons substitué aux données les deux seules valeurs T et N.

IV. *Distribution des estimations pour des échantillons correspondant à une valeur* U *donnée.* — Si nous intégrons l'équation (2) par rapport à T de θ à $+\infty$ nous trouvons la distribution de U pour N donné

$$df = 4K_0(2U)\frac{U^{2N-1}\,dU}{[(N-1)!]^2},$$

où K_θ désigne selon la notation classique une fonction de Bessel de seconde espèce, telle que

$$K_0(x) = \int_0^{+\infty} e^{-x\,\mathrm{ch}\,z}\,dz.$$

Il est important de noter que la distribution de U est indépendante de θ.

La distribution de T, U donné, est simplement

$$(B) \qquad \frac{1}{2K_0(2U)}\, e^{-U\left(\frac{T}{\theta}+\frac{\theta}{T}\right)}\frac{dT}{T},$$

et nous remarquerons que cette distribution est indépendante de N. La somme d'information relative à θ fourni par une seule observation, à partir de la distribution (B), est

$$\frac{2U}{\theta^2}\,\frac{K_1(2U)}{K_0(2U)},$$

et de ceci résulte que la valeur moyenne pour des échantillons de N groupes est $\frac{2N}{\theta^2}$, ce qui montre qu'en utilisant la distribution (B) comme loi d'erreur nous ne perdons aucune information. J'ai appelé cette méthode la Recherche d'informations à l'aide de statistiques « ancillaires » telles que U dans le problème actuel. Une telle valeur étant une statistique nous est connue lorsque nous possédons un échantillon, de même que N. Elle peut, comme U le fait dans le problème précédent, remplacer N dans la spécification de la classe d'échantillons à laquelle notre courbe d'erreur s'applique; on peut simplement ajouter une autre restriction. Sa distribution doit être indépendante du paramètre à estimer.

A propos du problème précédent, remarquons que lorsque nous disons que la statistique T est insuffisante « insufficient » nous voulons dire seulement qu'elle est « insufficient » si elle est complétée uniquement par la valeur ancillaire N. Si, au contraire, elle est complétée par la valeur ancillaire U, l'ensemble T, U fournit une méthode exhaustive d'estimation avec la courbe d'erreur (B), de telle sorte que « l'insufficiency » en réalité tient plutôt à la valeur ancillaire N qu'à l'estimation T, qui n'est pas modifiée elle-même bien que sa courbe d'erreur le soit lorsque la totalité de l'information utilisable est employée.

Le problème de trouver des valeurs qui, comme U dans cet exemple, ne sont fonction que des observations et ont des distributions indépendante des paramètres, méritent beaucoup d'autres recherches. Il y a dix ans, aux réunions du Tri-centenaire qui eurent lieu à Harvard, je le mis en lumière sous la forme dite « Problème du Nil ».

« La terre cultivable d'un village égyptien est inégalement fertile. La fertilité de chaque partie est connue avec exactitude en fonction de la hauteur atteinte par le Nil. Mais la hauteur de la crue a une influence

inégale sur les différentes parties du territoire. On demande de diviser le terrain entre les familles du village de sorte que les récoltes des lots assignés à chacun soient dans une proportion déterminée par avance, quelle que soit la hauteur atteinte par le fleuve. »

Si la distribution (1) de l'exemple ci-dessus représente la fertilité du pays en fonction du paramètre inconnu 0 correspondant à la hauteur de la crue du Nil, il est clair que le problème est résolu en utilisant la famille d'hyperboles équilatères U = const. comme limites des terrains distribués.

V. *Origine et échelle d'une Loi élémentaire de forme connue.* — L'emploi d'une information ancillaire de nature plus compliquée peut être illustré par un problème où nous avons un échantillon de valeurs $x_1, \ldots x_N$, tirés d'une distribution caractérisée par la probabilité élémentaire par exemple.

$$\Sigma q \frac{1}{2} e^{-\left|\frac{x-\alpha}{\beta}\right|} \frac{dx}{\beta}, \qquad df = \Phi\left(\frac{x-\alpha}{\beta}\right)\frac{dx}{\beta},$$

où Φ' est une fonction donnée et α, β deux paramètres inconnus.

Les équations du maximum de vraisemblance sont :

$$\sum\left[\Phi'\left(\frac{x-\alpha}{\beta}\right)\right] = 0,$$
$$\sum\left[\frac{x-\alpha}{\beta}\Phi'\left(\frac{x-\alpha}{\beta}\right)\right] + N = 0,$$

où Φ est la première dérivée de la fonction Φ et où Σ désigne une sommation étendue aux N valeurs de x observées. Supposons que A et B soient des valeurs de α et β satisfaisant à ces équations. Alors si $x = A + uB$, l'échantillon fournira un ensemble de N valeurs de u telles que

$$\Sigma\Phi'(u) = 0,$$
$$\Sigma u\Phi'(u) + N = 0;$$

l'ensemble des valeurs de u satisfaisant à ces conditions, caractéristique de notre échantillon, peut être appelé le « complexe » de l'échantillon. Évidemment, il ne dépend que des rapports des (N — 1) intervalles séparant nos observations lorsqu'elles sont rangées par ordre de grandeur, de sorte que la probabilité d'obtenir un échantillon de complexe donné est indépendante de α et β.

Considérons seulement les échantillons de même complexe que celui observé; les 2 fonctions

$$t_1 = \frac{A - \alpha}{\beta}, \qquad t_2 = \frac{B}{\beta}$$

qui dépendent des paramètres et des observations ont une probabilité composée indépendante des paramètres. En effet,

$$df \propto \prod_{i=1}^{N} \Phi(t_1 + u_i t_2)\, t_2^{N-2}\, dt_1\, dt_2.$$

Celle-ci, naturellement, est compliquée en pratique car elle fait intervenir le complexe particulier de l'échantillon observé et ses propriétés numériques en dépendent aussi bien que de la fonction Φ.

En théorie cependant, elle est simple en ce sens que la distribution est connue indépendamment de notre ignorance de α et β.

Nous pouvons en déduire : 1° la probabilité composée de A et B lorsque α et β sont donnés

$$df \propto \prod_{i=1}^{N} \Phi\left[\frac{A - \alpha}{\beta} + u_i \frac{B}{\beta}\right] \frac{B^{N-2}}{\beta^N}\, dA\, dB.$$

2° La probabilité fiduciaire composée de α et β lorsque A et B sont donnés

$$df \propto \prod_{i=1}^{N} \Phi\left[\frac{A - \alpha}{\beta} + u_i \frac{B}{\beta}\right] \frac{B^{N-1}}{\beta^{N+1}}\, d\alpha\, d\beta.$$

En général, dans cet exemple, rien de ce qui touche notre échantillon n'est superflu ou hors du sujet. Tout ce que nous avons fait c'est de partager les données en une partie ancillaire à N — 2 degrés de liberté (le complexe), suceptible d'être déterminée par N — 2 fonctions ne dépendant que des observations, qui sont fonctionnellement indépendantes et dont la probabilité composée est indépendante des paramètres; et en une partie de deux autres fonctions dépendant non seulement de l'échantillon, mais aussi des paramètres et dont la probabilité composée lorsque le complexe est donné, en est également indépendante des paramètres, et enfin qui peuvent être utilisées pour fournir leur distribution fiduciaire.

VI. *Conclusions tirées d'une population normale.* — Un exemple de distribution fiduciaire simultanée sans la nécessité d'information ancillaire, puisqu'il suffit d'utiliser des statistiques « sufficient », est fourni par des échantillons issus d'une distribution normale.

Considérons un échantillon de valeurs $x_1, \ldots x_N$ et nous calculons

$$\bar{x} = \frac{1}{N} \Sigma x,$$

$$s^2 = \frac{1}{N-1} \Sigma (x - \bar{x})^2;$$

si maintenant nous imaginons un second échantillon issu de la même population, nous aurons

$$\bar{x}' = \frac{1}{N'} \Sigma x',$$

$$s'^2 = \frac{1}{N'-1} \Sigma (\bar{x} - \bar{x}')^2$$

et à partir des deux échantillons nous pourrons calculer deux fonctions dont la distribution simultanée sera indépendante des paramètres de la population échantillonnée

$$z = \log s - \log s'$$

et

$$t^2 = \frac{(\bar{x} - \bar{x}')^2 (N + N' - 2)}{\left(\frac{1}{N} + \frac{1}{N'}\right) [(N-1)s^2 + (N'-1)s'^2]}.$$

Supposons qu'on connaisse la distribution simultanée de z et t, et qu'on ait observé un échantillon; on pourra alors en déduire la distribution de l'ensemble $\bar{x}'$ et s' caractérisant un autre échantillon. Nous pouvons ainsi établir une distribution simultanée des éléments d'un second échantillon non encore observé ayant n'importe quelle grandeur donnée. Si le second échantillon croît indéfiniment, les valeurs $\bar{x}'$ et s' viennent à différer en probabilité aussi peu que l'on veut des valeurs μ et σ de la population hypothétique d'où l'exemple est tiré.

Alors,

$$e^{2z} = \frac{s^2}{\sigma^2}$$

est distribué comme $\frac{X^2}{N-1}$ pour $N-1$ degrés de liberté et

$$t = (x-\mu)\frac{\sqrt{N}}{\sigma}$$

est distribué normalement avec la variance unité.

La distribution fiduciaire simultanée de μ et σ sera

$$\frac{1}{\sqrt{2\pi}} e^{-\frac{N(x-\mu)^2}{2\sigma^2}} \frac{\sqrt{N}\, d\mu}{\sigma} \frac{1}{\frac{N-1}{2}!} \left[\frac{(N-1)s^2}{2\sigma^2}\right]^{\frac{N}{2}} e^{-\frac{(N-1)s^2}{2\sigma^2}} \frac{d\sigma}{\sigma}$$

dont la distribution marginale trouvée en intégrant par rapport à μ est la distribution fiduciaire ordinaire de σ pour s donné, tandis que celle trouvée en intégrant par rapport à σ est la distribution de Student pour

$$t = (x-\mu)\frac{\sqrt{N}}{s}.$$

VII. *Distribution fiduciaire simultanée des pourcentages.* — Supposons qu'une variable x ait une distribution continue de sorte que la probabilité $x < x'$ soit une fonction non décroissante et continue de x', d'ailleurs de forme inconnue.

Si nous considérons la médiane de la distribution $\xi_{.5}$, la probabilité pour que exactement r parmi les N observations excèdent cette médiane et $N-r$ lui soient inférieurs est

$$2^{-N}\frac{N!}{r!\,(N-r)!};$$

et ceci peut être considéré comme la probabilité fiduciaire pour que $\xi_{.5}$, ou peut-être tout un intervalle de valeurs médianes soit compris entre les $r^{\text{ième}}$ et $(r+1)^{\text{ième}}$ valeurs observées, ordonnées selon les valeurs croissantes.

Plus généralement, si ξ_p représente le paramètre qui excède la fraction p de la population échantillonnée, la probabilité pour que dans un échantillon de N observations r exactement excèdent ξ_p est

$$p^{N-r}(1-p)^r\frac{N!}{r!\,(N-r)!};$$

et cette expression peut être considérée comme la probabilité fiduciaire

que le paramètre ξ_p tombe dans l'intervalle défini par les $r^{\text{ième}}$ et $(r+1)^{\text{ième}}$ observations. Nous avons ainsi N points donnés de la fonction de fréquence fiduciaire de ξ_p pour toutes les valeurs de p.

En outre, la distribution fiduciaire simultanée de ξ_p et de $\xi_{p'}$, avec $p > p'$, est donnée par les termes du développement trinome

$$p'^{N-r'}(p-p')^{r'-r}(1-p)^{r}\frac{N!}{r!\,(r'-r)!\,(N-r')!}$$

déterminant la probabilité fiduciaire pour que ξ_p soit trouvé entre les $r^{\text{ième}}$ et les $(r+1)^{\text{ième}}$ observations et que simultanément $\xi_{p'}$ soit trouvé entre les $r'^{\text{ième}}$ et $(r'+1)^{\text{ième}}$ lorsque $r' \geqq r$.

Les séries complètes de conclusions de cette sorte peuvent être déduites en prenant comme quantités pivotales les fractions de la population dépassées par la plus grande, la suivante en grandeur, etc., de nos valeurs observées. Désignons-les par $p_1, p_2 \ldots p_N$ est déterminée par la probabilité élémentaire

$$N!\,dp_1\,dp_2 \ldots dp_N, \qquad p_1 \geqq p_2 \geqq \ldots \geqq p_N.$$

Leur distribution est simultanée, quelle que soit la nature de la population échantillonnée, pourvu que sa fonction de répartition soit continue.

On peut en déduire par intégration les probabilités de toutes les inégalités établies plus haut. Ici le nombre de paramètres dans la distribution fiduciaire est égal au nombre d'observations. Puisque rien n'est donné au sujet de la distribution, sauf sa continuité, la «métrique» de x n'est pas spécifiée et nous ne devons pas être influencés par le fait que pour une «métrique» arbitraire, certaines observations, rangées par ordre de valeurs croissantes, sont groupées tandis que d'autres peuvent sembler éloignées.

VIII. *La table quadruple.* — Supposons que de $a+b$ animaux soumis à l'expérience a meurent et b survivent et que des $c+d$ animaux de contrôle c meurent et d survivent. L'emploi d'animaux de contrôle dans de telles expériences signifie que nous désirons juger l'effet du traitement sur le taux de mortalité par une comparaison contenue dans l'expérience elle-même et non par rapport à des espérances fondées sur d'autres expériences. Le problème statistique est le suivant : le témoignage de cette expérience contenue en elle-même, justifie-t-il la

conclusion que le traitement testé a élevé ou diminué le taux de mortalité par comparaison avec le comportement des « contrôles » ?

Il est facile de calculer que si la probabilité vraie de la mort des animaux traités est p et si la mort et la survie sont indépendantes pour tous les animaux, la probabilité du résultat observé pour ceux-ci sera :

$$\frac{(a+b)!}{a!\,b!}p^a(1-p)^b;$$

de même si p' est la probabilité vraie de la mort des animaux de contrôle, la probabilité du résultat observé sera :

$$\frac{(c+d)!}{c!\,d!}p'^c(1-p')^d,$$

Si nous choisissons comme hypothèse à tester que $p=p'$, la probabilité composée de ce qui a été observé dans les deux groupes est le produit :

$$\frac{(a+b)!\,(c+d)!}{a!\,b!\,c!\,d!}p^{a+c}(1-p)^{b+d},$$

qui contient seulement le facteur inconnu $p^{a+c}\,(1-p)^{\,b+d}$, de sorte qu'avec l'information ancillaire que $a+c$ en tout sont morts, la probabilité de chaque résultat possible sera proportionnelle à

$$\frac{1}{a!\,b!\,c!\,d!},$$

et l'on peut montrer qu'elle est égale à

$$\frac{(a+b)!\,(c+d)!\,(a+c)!\,(b+d)!}{a!\,b!\,c!\,d!\,(a+b+c+d)!}.$$

Les résultats admissibles soumis à l'information ancillaire utilisée, forment une suite linéaire discontinue ayant à une extrémité un résultat d'observation où soit a soit d est nul, et à l'autre extrémité un autre, où b ou c est nul. Les fréquences relatives des membres d'une suite donnée étant connues indépendamment du paramètre inconnu p, nous avons un test objectif en calculant la probabilité des valeurs particulières observées en même temps que celle de tous les ensembles possibles d'observations ayant les mêmes sommes marginales, disproportionnées dans le même sens, et pour une plus grande étendue que celles observées.

La convention qui consiste à prendre pour degré de signification d'un test, la fréquence parmi des échantillons répétés sans restriction à partir de la même population, ne concorde pas avec cette méthode de raisonnement. Par exemple, si de trois animaux traités, tous meurent, et de trois animaux de contrôle tous survivent, le raisonnement que j'ai développé conclut que quelles que puissent être les véritables fréquences de la mort, les quatre résultats symbolisés par

$$\begin{array}{c|c} 3 & 0 \\ \hline 0 & 3 \end{array} \qquad \begin{array}{c|c} 2 & 1 \\ \hline 1 & 2 \end{array} \qquad \begin{array}{c|c} 1 & 2 \\ \hline 2 & 1 \end{array} \qquad \begin{array}{c|c} 0 & 3 \\ \hline 3 & 0 \end{array}$$

doivent arriver avec des fréquences dans les rapports 1 : 9, 9 : 1; par conséquent, ce que nous avons observé est favorable au point de vue que le traitement imposé augmente la fréquence de mort, le degré de signification de cette conclusion étant $1/20^e$, puisque ce cas est le résultat le plus favorable parmi les 20 résultats également fréquents.

Cependant, si nous considérons les essais répétés à partir de la même expérience, la probabilité du résultat observé est $p^3\,q^3$ et son maximum a lieu lorsque $p = q$, la fréquence étant alors $1/64^e$. Par conséquent, le degré de signification de nos observations devrait être au moins de $1/64^e$.

Considérons les 44 cas restants que nous avons négligés; dans 30 cas 4 ou 2 animaux sont morts, dans 12 cas 5 ou 1, dans 2 cas 6 ou 0. Si 4 ou 2 animaux sont morts, nous avons une suite telle que :

$$\begin{array}{c|c} 2 & 1 \\ \hline 0 & 3 \end{array} \qquad \begin{array}{c|c} 1 & 2 \\ \hline 1 & 2 \end{array} \qquad \begin{array}{c|c} 0 & 3 \\ \hline 2 & 1 \end{array}$$

avec des fréquences dans le rapport 1 : 3 : 1, de sorte que même le plus probable arrivera aussi souvent que 1 sur 5 essais ayant un résultat de cette série.

Si 5 ou 1 animaux sont morts, la série sera :

$$\begin{array}{c|c} 1 & 2 \\ \hline 0 & 3 \end{array} \qquad \begin{array}{c|c} 0 & 3 \\ \hline 1 & 2 \end{array}$$

comportant seulement deux possibilités de fréquences égales, tandis que si 6 ou 0 animaux sont morts, les animaux traités ou de contrôle doivent s'être comportés exactement de la même façon. De ces 44 cas on

peut donc dire que l'expérience n'a pas fourni assez d'informations de la nature cherchée pour que n'importe quelle opinion solide soit formée, favorable ou défavorable à la théorie à tester. Notre expérience comporte assez fréquemment des résultats non concluants, mais à mon point de vue, ceci n'affecte pas notre jugement de signification lorsqu'il arrive que le nombre réel des morts rende une décision possible.

Peut-être une analogie éclaircira-t-elle cette distinction logique. Nous pouvons tester si un animal est homozygote ou hétérozygote en le croisant avec un récessif; s'il est hétérozygote, nous attendons, disons un nombre égal de souris noires et brunes, s'il est homozygote toutes seront noires. Une partie de six souris est née et toutes sont noires; nous pouvons alors conclure, je pense, que si le parent avait été hétérozygote, une telle portée ne se serait produite qu'une fois sur 64 essais; mais ceci suppose que la portée est nécessairement de six souris. En fait, il y a une probabilité finie bien qu'indéterminée, pour que la portée soit de moins de cinq, auquel cas je dirai que le test a fait défaut. Je n'aimerais pas rehausser la signification du résultat qui a été obtenue, en raison du fait qu'une répétition du test pourrait donner une évidence moins impérative que celle effectivement obtenue. La distribution marginale dans le premier problème m'apparaît ainsi analogue au nombre de souris classées dans le second, et devoir être acceptée comme partie des données du problème statistique correspondant, indépendamment de sa fréquence de réalisation comme résultat d'une répétition physique.