Zeitschrift: Bulletin de la Société Vaudoise des Sciences Naturelles

Herausgeber: Société Vaudoise des Sciences Naturelles

Band: 54 (1921-1922)

Heft: 204

Artikel: Une application du calcul des probabilités aux statistiques des sciences biologiques expérimentales

Autor: Amann, J.

DOI: https://doi.org/10.5169/seals-270916

Nutzungsbedingungen

Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften auf E-Periodica. Sie besitzt keine Urheberrechte an den Zeitschriften und ist nicht verantwortlich für deren Inhalte. Die Rechte liegen in der Regel bei den Herausgebern beziehungsweise den externen Rechteinhabern. Das Veröffentlichen von Bildern in Print- und Online-Publikationen sowie auf Social Media-Kanälen oder Webseiten ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. Mehr erfahren

Conditions d'utilisation

L'ETH Library est le fournisseur des revues numérisées. Elle ne détient aucun droit d'auteur sur les revues et n'est pas responsable de leur contenu. En règle générale, les droits sont détenus par les éditeurs ou les détenteurs de droits externes. La reproduction d'images dans des publications imprimées ou en ligne ainsi que sur des canaux de médias sociaux ou des sites web n'est autorisée qu'avec l'accord préalable des détenteurs des droits. En savoir plus

Terms of use

The ETH Library is the provider of the digitised journals. It does not own any copyrights to the journals and is not responsible for their content. The rights usually lie with the publishers or the external rights holders. Publishing images in print and online publications, as well as on social media channels or websites, is only permitted with the prior consent of the rights holders. Find out more

Download PDF: 02.07.2025

ETH-Bibliothek Zürich, E-Periodica, https://www.e-periodica.ch

Une application du calcul des probabilités aux statistiques des sciences biologiques expérimentales

PAR

J. AMANN

(Présenté à la séance du 11 janvier 1922.)

On a souvent à faire, dans les sciences biologiques, à des données statistiques provenant d'un nombre restreint ou même très restreint d'observations. Les lois générales du calcul des probabilités n'étant rigoureusement valables que pour des nombres d'observations très grands (théoriquement indéfiniment grands), on est en droit de se demander quelle est la valeur relative des conclusions que l'on peut tirer des statistiques qui ne considèrent que des observations peu nombreuses.

Ce problème, qui s'est présenté, il y a longtemps déjà, pour les statistiques médicales par exemple, a été traité par divers auteurs [1].

Il m'a paru intéressant d'examiner quelques cas ressortissant des sciences biologiques expérimentales où il se présente sous une forme tout à fait analogue, sinon identique, parce que, dans la règle, les expériences en question ne peuvent porter — pour des raisons d'ordre pratique — que sur des individus en nombre restreint. Prenons un exemple applicable à beaucoup de ces cas.

Nous voulons comparer deux séries d'individus sous le rapport de la présence (ou de l'absence) d'un caractère particulier. L'une des séries comprend, par exemple, des individus qui ont été soumis à certaines conditions expérimentales et l'autre série, des individus soumis à des conditions différentes des premières sous un certain rapport bien déterminé.

La question que nous cherchons à résoudre est celle-ci : y a-t-il

[1] On trouvera des indications bibliographiques dans le travail du Professeur C. Liebermeister : « Ueber Wahrscheinlichkeitsrechnung..... » (R. von Volkmann : Sammlung klinischer Vorträge für innere Medizin, II, p. 935.)

une relation entre la présence (ou l'absence) d'un caractère et la différence des conditions auxquelles les deux séries ont été soumises ? Ou, autrement dit : la fréquence différente de ce caractère chez les individus des deux séries, dépend-elle de la différence des conditions expérimentales auxquelles elles ont été soumises ?

Pour cela, nous faisons la statistique des individus qui présentent (ou ne présentent pas) le caractère, en déterminant, pour chacune des séries, la proportion de ces individus par rapport au nombre total.

Si, pour l'une des séries, nous avons a individus présentant le caractère et b ne le présentant pas, la fraction $\frac{a}{a+b}$ est la proportion pour cette série. Si pour l'autre série, nous avons semblablement p individus présentant le caractère et q ne le présentant pas, le rapport en question est $\frac{p}{p+q}$.

Si $a + b$ et $p + q$ étaient de grands nombres (comme c'est parfois le cas), nous serions en droit de tirer, de la comparaison des deux rapports ci-dessus, une conclusion touchant la dépendance du caractère considéré, du facteur expérimental étudié. Dans le cas où ces rapports seraient suffisamment différents, on pourrait en effet admettre cette dépendance comme très probable ou même comme certaine si l'expérience a été faite dans des conditions telles qu'elles excluent toute autre interprétation.

Mais si $a + b$ et $p + q$ sont des nombres réduits, cette conclusion n'est plus admissible *a priori*, même au cas où les rapports sont fort différents.

La question qui se pose alors est celle-ci : quelle est la valeur de la probabilité que la différence observée entre les deux rapports $\frac{a}{a+b}$ et $\frac{p}{p+q}$ corresponde à une différence de même signe pour des observations réitérées un nombre de fois assez grand pour que l'on puisse conclure avec une certitude suffisante que cette différence dans les proportions est due aux conditions expérimentales ?

Ou, autrement dit, la fréquence plus grande des individus présentant (ou ne présentant pas) le caractère en question, constatée pour l'une des séries, a-t-elle une signification quelconque au point de vue statistique, ou bien peut-elle être attribuée au hasard ?

Cette question relève du calcul des probabilités qui permet de la résoudre, en nous donnant le moyen de calculer d'une manière

suffisante la valeur de cette probabilité ; ce qui nous permettra, le cas échéant, de répondre à la question posée ci-dessus.

Le problème énoncé plus haut, qui, comme on le voit immédiatement, comporte des applications nombreuses aux sciences naturelles en général, revient à celui-ci :

Etant données deux urnes dont nous savons que chacune contient un nombre considérable de boules, dont les unes sont blanches et les autres noires, les proportions des blanches aux noires, dans chaque urne, étant du reste inconnues, nous voudrions savoir si les boules blanches sont en proportion plus ou moins forte dans une urne que dans l'autre.

Pour ce faire, nous extrayons, d'une des urnes, un certain nombre de boules (l'une après l'autre et en les remettant chaque fois et mélangeant bien) : il s'en trouve a blanches et b noires (nombre des boules extraites $a + b$). De l'autre urne, nous tirons, de la même manière, un nombre quelconque $p + q$ de boules, dont p sont blanches et q sont noires.

Si nous avons constaté ainsi que la proportion des boules blanches au nombre des boules extraites est plus ou moins forte dans la première urne que dans la seconde, nous devons nous demander si les résultats obtenus par ces tirages, qui n'ont porté que sur des nombres restreints ou très restreints de boules, suffisent à nous permettre de conclure que la proportion des blanches est plus ou moins forte dans la première urne que dans la seconde.

Ou, en d'autres termes : étant donnés les résultats obtenus par ces deux tirages, quelle est la valeur de la probabilité que la proportion des boules blanches est plus ou moins forte dans la première urne que dans la seconde ?

L'application des formules générales auxquelles on arrive par la discussion mathématique (du reste assez compliquée) de ce problème, nous permettra d'obtenir assez facilement les valeurs des deux probabilités :

P que la proportion des boules blanches est plus forte dans la première urne que dans la 2me, et

1 - P qu'il n'en est pas ainsi, c'est-à-dire que les blanches sont en proportion moins forte ou égale à celle des noires.

Pour prendre l'exemple numérique très simple donné *loc. cit.*, supposons que le tirage de 4 boules de la première urne comprenne 3 blanches et 1 noire, et le tirage de 3 boules de la 2me urne, com-

prenne 1 blanche et 2 noires, c'est-à-dire $\frac{a}{a+b} = \frac{3}{4}$ et $\frac{p}{p+q} = \frac{1}{3}$. Ces rapports ne permettent pas de conclure que les boules blanches sont en proportion plus forte dans la première urne ; étant donné le petit nombre des boules extraites et le grand nombre de celles contenues dans les urnes, on doit se demander si la différence constatée $\frac{3}{4} - \frac{1}{2}$ n'est pas due au hasard, et si des tirages répétés un grand nombre de fois n'amèneraient pas à un résultat très différent de celui obtenu par cette première opération. La comparaison des rapports ci-dessus peut établir tout au plus une présomption : quel en est le degré de probabilité ?

En appliquant aux données ci-dessus la formule en question [1] nous obtenons :

$$P = 0{,}833333... \quad 1 - P = 0{,}166666... \quad \frac{P}{1-P} = 5.$$

La probabilité 1 — P n'étant pas d'ordre négligeable, on ne peut donc conclure avec certitude à une proportion plus forte des boules blanches dans la première urne ; mais la probabilité qu'il en est ainsi est cependant cinq fois plus grande que celle qu'il en est autrement ; ou autrement dit, on pourrait parier 5 contre 1 qu'il en est ainsi.

Lorsque la différence entre les résultats des tirages est moindre que dans l'exemple ci-dessus, par exemple 2 blanches et 2 noires au premier tirage et 1 blanche et 2 noires au deuxième :

$$\frac{a}{a+b} = \frac{2}{4} \quad \frac{p}{p+q} = \frac{1}{3}, \text{ les probabilités deviennent :}$$

$$P = 0.573 \quad 1 - P = 0.427 \qquad \frac{P}{1-P} = 1.34.$$

Dans ce cas, on ne peut donc parier que 1,34 contre 1 que les blanches sont en proportion plus forte dans la 1re urne.

En résumé, si les données fournies par des observations portant sur un nombre réduit d'individus, ne suffisent pas, dans la règle, à conclure avec une certitude suffisante, l'application des formules générales du calcul des probabilités fournit, dans chaque

[1] Liebermeister, *loc. cit.*, p. 946, formules I et II, et p. 958. Avec une table des logarithmes des factorielles de 0 ! à 1200 !

cas particulier, une évaluation exacte de la probabilité que présentent l'une et l'autre des alternatives possibles ; et cette évaluation est nécessaire pour qu'on puisse se rendre compte de la signification réelle des résultats statistiques obtenus et de leur valeur relative.

La supposition que nous avons faite ci-dessus, que le nombre des boules contenues dans les deux urnes était très grand, est réalisée même dans les cas où nous faisons l'application de la formule à la statistique fournie pour un petit nombre d'individus : elle correspond, en effet, à la possibilité théorique de répéter l'expérience et les constatations un nombre de fois indéfini. (Liebermeister, *loc. cit.*)

Quant aux critiques que l'on peut faire à ces statistiques, relativement au fait que, pour une raison ou pour une autre, les individus considérés ne seraient pas rigoureusement comparables, ou bien que les conditions de l'expérience n'exclûraient pas la possibilité que la présence ou l'absence du caractère dépende d'autres facteurs que celui auquel on veut l'attribuer, leur examen est du domaine de la critique expérimentale, et non pas de celui du calcul des probabilités : les formules que nous appliquons ayant été établies en admettant que les individus des deux séries ne diffèrent que par la présence ou l'absence du caractère considéré. Il en est de même du second point concernant la nature du facteur auquel est due la différence des proportions constatées : le calcul des probabilités ne s'en occupe pas : il se borne à évaluer la probabilité que cette différence existe ou n'existe pas.

Il est évident que la valeur de cette probabilité dépend, d'une part, du nombre des individus observés pour chaque série, et d'autre part, du rapport existant entre les proportions constatées des catégories considérées dans les deux séries.

Comme exemples de la valeur des probabilités dans certains cas particuliers, j'ai calculé, au moyen de la formule de Liebermeister, les probabilités qui correspondent aux données suivantes où les deux séries observées comprennent chacune 10 individus ($a + b = p + q = 10$).

$\frac{a}{a+b}$	$\frac{p}{p+q}$	P	1 — P	$\frac{P}{1-P}$ (environ)
$\frac{1}{10}$	$\frac{0}{10}$	0.76185....	0.23815....	3.2
$\frac{5}{10}$	$\frac{4}{10}$	0.6655.....	0.3345.....	2.0
$\frac{6}{10}$	$\frac{4}{10}$	0.8027.....	0.1973.....	4.0
$\frac{7}{10}$	$\frac{3}{10}$	0.9595.....	0.0405.....	24
$\frac{8}{10}$	$\frac{2}{10}$	0.9957.....	0.0043.....	248
$\frac{9}{10}$	$\frac{1}{10}$	0.9998.....	0.0002.....	5781
$\frac{9}{10}$	$\frac{0}{10}$	0.99998....	0.00002....	50000