JOURNAL DE LA SOCIÉTÉ STATISTIQUE DE PARIS

FRED MILHLAUD

Les modalités de la synergie de plusieurs facteurs. Un procédé pour leur étude en médecine, en hygiène, en sociologie

Journal de la société statistique de Paris, tome 105 (1964), p. 114-119

<http://www.numdam.org/item?id=JSFS_1964__105__114_0>

VI

VARIÉTÉ

Les modalités de la synergie de plusieurs facteurs
Un procédé pour leur étude en Médecine, en Hygiène, en Sociologie

L'un des problèmes que le médecin ou l'hygiéniste peuvent poser au statisticien est de déterminer comment agissent différents facteurs lorsque plusieurs conditions sont réunies qui peuvent influer sur le même phénomène.

Il s'agit par exemple d'expériences systématiques où plusieurs poisons ou plusieurs inoculations microbiennes sont administrés aux mêmes animaux, ou plusieurs médicaments aux mêmes malades.

Il peut aussi s'agir d'enquêtes statistiques pour déterminer les importances respectives de deux facteurs dans l'étiologie d'une maladie déterminée.

Cette recherche des étiologies intéresse au premier chef l'hygiéniste.

On peut rechercher des étiologies purement biologiques : âge, présence de tel ou tel antécédent morbide, activité particulière, influence de telle ou telle intoxication chronique, d'ordre professionnel ou alimentaire, etc. On peut aussi rechercher des étiologies sociales : situation familiale, transplantation, appartenance socio-professionnelle, etc. D'ailleurs ces étiologies se recouvrent, certaines activités ou certaines intoxications étant liées à certaines appartenances. On peut aussi bien étudier l'importance de facteurs sociaux dans d'autres phénomènes que des maladies, par exemple dans la délinquance ou dans le divorce. Le problème est le même, en ce qui concerne la méthodologie statistique, que la question posée soit d'ordre médical, sociologique, ou mixte.

Nous envisagerons le cas où on étudie un événement I non mesurable, par exemple la mort. On cherche comment se conjuguent deux facteurs A et B susceptibles de le provoquer. Ou bien, on cherche comment se conjuguent 2 facteurs A et B susceptibles de provoquer un même symptôme ou d'être à l'étiologie d'une même maladie.

Mais ici une réflexion s'impose. Sous la dénomination de I nous réunissons des phénomènes que nous ne voulons pas discerner. Il est possible que la mort par A s'accompagne d'autres symptômes que la mort par B. Il est possible, si A et B provoquent un même symptôme, qu'ils en provoquent chacun d'autres en même temps. Les événements dus à A et ceux dus à B sont peut-être identiques, mais rien n'est moins certain. Simplement nous avons défini I de façon à ce que ces différences éventuelles ne soient pas considérées.

On peut retenir un ensemble E_a des sujets soumis à A, et un ensemble E_b des sujets soumis à B. Ces deux ensembles ont une intersection qui est le sous-ensemble des sujets soumis à la fois à A et B.

$$\text{On a un sur-ensemble } = E_a - (E_a \cap E_b) + (E_a \cap E_b) + E_a - (E_a \cap E_b)$$
$$= E_a + E_b - (E_a \cap E_b).$$

On peut y concevoir que les événements I sont des événements I_a dus à A et I_b dus à B, non discernés.

La densité des I_a dans E_a est la probabilité r_a d'observer I quand on a réalisé A.

La densité des I_b dans E_b est la probabilité r_b d'observer I quand on a réalisé B.

Il y a *indépendance* de A et B si la probabilité r_a d'observer I quand on a réalisé A n'est pas influencée par la réalisation de B et la probabilité d'observer I quand on a réalisé B n'est pas influencée par la présence de A.

Il y a *synergie* si les probabilités se renforcent.

Il y a *contresynergie* si elles s'atténuent réciproquement.

On peut supposer que l'action de A sur I_b ne soit pas la même que l'action de B sur I_a mais cela, nous nous interdisons de le chercher puisque nous ne distinguons pas I_a de I_b.

Si nous administrons simultanément 2 poisons α et β il est possible que l'action de α sur certains éléments du syndrome dû à β ne soit pas la même que celle de β sur certains éléments du syndrome dû à α, mais nous ne considérons qu'un élément, commun aux 2 syndromes.

Soit r_a la probabilité de I quand A est présent.

Soit r_b la probabilité de I quand B est présent.

Il y a n cas où A et B sont présents.

Nous avons $E_a = n\,r_a$
$E_b = n\,r_b$

Soit p la probabilité de I dans ce cas

$$np = \text{sur-ensemble } E_{a,b} = E_a + E_b - (E_a \cap E_b)$$
$$= n\,r_a + n\,r_b - n\,r_a\,r_b$$
$$p = r_a + r_b - r_a\,r_b$$

L'écart-type de fréquence np est $\sqrt{npq}$.

Tel est le cas si A et B agissent de façon indépendante.

Il devient facile de vérifier cette indépendance et de définir la potentialisation ou la contresynergie.

Soit F la fréquence des cas où on observe I alors que A et B sont présents.

Il y a *potentialisation* si $F > np$
contresynergie si $F < np$.

On peut poser un *coefficient de synergie* $\boxed{c = \dfrac{F - np}{F + np}}$

Si $\sqrt{npq}$ est voisin de O, c est évalué sans risque d'erreur.
Sinon, on pose $(\alpha = F - np)$ évalué en $\sqrt{npq}$,
$(\beta = F - np)$ évalué de même.

En même temps que $c = \dfrac{\alpha}{\beta}$, on calcule $c = \dfrac{\alpha + 1}{\beta + 1}$, $c = \dfrac{\alpha + 2}{\beta + 2}$, $c = \dfrac{\alpha + 3}{\beta + 3}$. Devant une valeur de c, nous pouvons savoir la marge des erreurs possibles qui correspondent à des différences inférieures à 2 écarts-types entre les valeurs correspondantes de F. On en fera un tableau.

* * *

$c = +1$ signifie $F \neq 0$, $r_a = r_b = p = np = 0$, c'est-à-dire qu'aucun des facteurs A et B n'agit isolément, alors qu'ils agissent réunis. Il y a *potentialisation absolue.* Comme $p = 0$, $\sqrt{npq} = 0$, il n'y a aucune marge d'erreur.

$0 < c < +1$ signifie *potentialisation partielle.*

$c = 0$, signifie *indépendance*, sous réserve de la marge d'erreur.

$-1 < c < 0$ signifie *contresynergie partielle.*

$c = -1$ signifie $F = 0$, $r_a = r_b = p \neq 0$. Il y a *contresynergie absolue.*

On a donc pour s des valeurs comprises entre $+1$ qui signifierait synergie absolue, et -1 qui signifierait contresynergie absolue, toutes les valeurs positives signifiant synergie et les valeurs négatives contresynergie.

* * *

Le phénomène se complique de ce que A et B peuvent ne pas être des phénomènes dont on relève seulement la présence ou l'absence. Ils peuvent aussi être des quantités mesurables, par exemple des doses de médicaments, de poisons, de radiations, des durées de séjour dans un milieu d'une certaine nocivité, des durées de certaines habitudes ou d'exercice d'un certain travail, etc.

Soit y la probabilité de I, x_a la mesure de A et x_b la mesure de B, il nous semble impossible de trouver un procédé général pour établir algébriquement la forme de la fonction, continue ou non, qui lie y à la fois à x_a et x_b, compte tenu des potentialisations ou contresynergies possibles.

Il est cependant possible d'étudier le phénomène.

Nous construisons des tableaux à 4 lignes et 4 colonnes.

La 1re colonne correspond aux cas où $x_a = 0$
La 2e colonne correspond aux cas où $0 < x_a < s_a$
La 3e colonne correspond aux cas où $s_a < x_a < 2\, s_a$
La 4e colonne correspond aux cas où $2\, s_a < x_a < 3\, s_a$
La 1re ligne correspond aux cas où $x_b = 0$
La 2e ligne correspond aux cas où $0 < x_b < s_b$
La 3e ligne correspond aux cas où $s_b < x_b < 2\, s_b$
La 4e ligne correspond aux cas où $2\, s_b < x_b < 3\, s_b$

Si nous soupçonnons que A et B n'agissent qu'à partir d'un certain seuil, nous prenons des valeurs de s_a et s_b qui soient les valeurs minimales vraisemblables pour ces seuils.

Dans un tableau I, chaque case contient le nombre de sujets pour lesquels x_a et x_b sont dans les zones correspondantes.

Dans un tableau II, chaque case contient le nombre qui exprime la fréquence F de l'événement I dans les conditions correspondantes.

Pour savoir la valeur de np qui correspond à chaque valeur de F, nous considérons ces deux tableaux. Nous trouvons n dans la case correspondante du tableau I, nous avons r_a en divisant le total de la colonne du tableau II par le total de la même colonne du tableau I; nous avons r_b en divisant le total de la ligne du tableau II par celui de la même ligne du tableau I, et nous utilisons la formule $p = r_a + r_b - r_a\, r_b$.

Nous pouvons construire un tableau III où chaque case contient une valeur de np.

A partir des tableaux II et III, nous construisons un tableau IV où chaque case contient une valeur du coefficient de synergie.

Deux éventualités peuvent se présenter :

1° Ce coefficient est sensiblement le même dans toutes les cases du tableau.

2° Il n'est pas le même.

La première se divise en 3 sous-éventualités. Le coefficient trouvé montre potentialisation, synergie simple, ou contresynergie.

La deuxième se divise en 16 sous-éventualités principales :

1° Le coefficient est le même sur toutes les lignes de la même colonne et il s'élève au fur et à mesure que l'on se déplace vers la droite.

Cela signifie qu'il s'élève en fonction directe de x_a et qu'il est indépendant de x_b.

2° Il est le même sur toutes les lignes de la même colonne et il s'abaisse au fur et à mesure que l'on se déplace vers la droite.

Il est fonction inverse de x_a, et indépendant de x_b.

3° et 4° Il est le même sur toutes les lignes de la même colonne, et passe par un maximum, ou par un minimum, sur une colonne moyenne.

Il est fonction de x_a, passant par un maximum ou par un minimum pour une certaine valeur de x_a, et il est indépendant de x_b.

5° Il est le même sur toutes les colonnes de la même ligne, et croît au fur et à mesure que l'on se déplace de la ligne du haut vers celle du bas.

Cela signifie qu'il est fonction directe de x_b, et indépendant de x_a.

6° Il est le même sur toutes les colonnes de la même ligne, et décroît au fur et à mesure que l'on se déplace de la ligne du haut vers celle du bas.

Il est fonction inverse de x_b, et indépendant de x_a.

7° et 8° Il est le même sur toutes les colonnes de la même ligne et passe par un maximum ou par un minimum sur une ligne moyenne.

Il est fonction de x_b, passant par un maximum ou par un minimum pour une certaine valeur de x_b, et il est indépendant de x_a.

9° Il n'est pas le même sur toutes les colonnes de la même ligne ni sur toutes les lignes de la même colonne, mais varie le long des diagonales.

Il augmente au fur et à mesure que l'on va de l'angle supérieur gauche à l'angle inférieur droit.

Il diminue au fur et à mesure que l'on va de l'angle supérieur droit à l'angle inférieur gauche.

Cela signifie qu'il augmente comme augmentent x_a et x_b.

10° Il n'est pas le même sur toutes les colonnes de la même ligne ni sur toutes les lignes de la même colonne, mais varie le long des diagonales.

Il augmente au fur et à mesure que l'on va de l'angle supérieur droit à l'angle inférieur gauche.

Il diminue au fur et à mesure que l'on va de l'angle supérieur gauche à l'angle inférieur droit.

Cela signifie qu'il varie en fonction inverse de x_a et x_b.

11° Il augmente, sur tous les trajets possibles, au fur et à mesure que l'on se rapproche de l'angle supérieur droit.

Cela signifie qu'il varie en fonction directe de x_a et en fonction inverse de x_b.

12° Il augmente au fur et à mesure que l'on se rapproche de l'angle inférieur droit.

Cela signifie qu'il varie en fonction directe de x_b et en fonction inverse de x_a.

13° Il varie le long des deux diagonales en passant par un maximum.

Cela signifie qu'il varie en fonction de x_a et x_b selon une fonction qui passe par un maximum.

14° Il varie le long des deux diagonales en passant par un minimum.

Cela signifie qu'il varie en fonction de x_a et x_b selon une fonction qui passe par un minimum.

15° Les valeurs maximales sont à l'angle supérieur droit et à l'angle inférieur gauche, les valeurs minimales aux deux autres angles.

Au croisement des deux diagonales, on notera des valeurs qui seront les plus élevées de celles portées sur la diagonale qui joint les valeurs minimales et les moins élevées de celle qui joint les valeurs maximales.

Cela signifie que le coefficient varie en fonction de x_a et de x_b selon une courbe qui passe par un minimum si x_a et x_b varient dans le même sens, par un maximum dans le cas opposé.

16° Les valeurs minimales sont à l'angle supérieur droit et à l'angle inférieur gauche, les valeurs maximales aux deux autres.

Au croisement des deux diagonales, on notera encore des valeurs qui seront les plus élevées de celles portées sur la diagonale qui joint les valeurs minimales, et les moins élevées de celles portées sur la diagonale qui joint les valeurs maximales.

Cela signifie que le coefficient varie en fonction de x_a et x_b selon une courbe qui passe par un maximum si x_a et x_b varient dans le même sens, par un minimum dans le cas opposé.

Il y a donc de nombreux types de relations principales que le seul examen du tableau permettra de reconnaître.

*
* *

Il serait facile de mettre en évidence les modalités de synergie de plus de 2 facteurs.
Avec 3, A, B, C, on aurait :

$$p = (r_a + r_b + r_c) - (r_a r_b + r_a r_c + r_b r_c) + r_a r_b r_c$$

On aurait comme précédemment à considérer $\frac{\mathrm{F} - np}{np}$, et à poser $q = 1 - p$, $\sigma_b = \sqrt{npq}$.

Mais ce serait sans intérêt.

Il faut considérer $r_{a,b}$ = probabilité de I quand A et B sont présents,
$r_{a,c}$ = probabilité de I quand A et C sont présents,
$r_{b,c}$ = probabilité de I quand B et C sont présents.

On cherchera l'action de A sur la conjoncture de B et C présents,
B sur la conjoncture de A et C présents,
C sur la conjoncture de A et B présents.

Et on partira des formules :

$$p_a = r_{b,c} + r_a - r_{b,c}\, r_a$$
$$p_b = r_{a,c} + r_b - r_{a,c}\, r_b$$
$$p_c = r_{a,b} + r_c - r_{a,b}\, r_c$$

On se trouve ainsi ramené au cas précédent.

D^r^ Fred MILHAUD