

Le goulot d'étranglement

Mathématiques et sciences humaines, tome 5 (1964), p. 59-60

http://www.numdam.org/item?id=MSH_1964__5_59_0

© Centre d'analyse et de mathématiques sociales de l'EHESS, 1964, tous droits réservés.

L'accès aux archives de la revue « Mathématiques et sciences humaines » (<http://msh.revues.org/>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

LE GOULOT D'ETRANGLEMENT

Le problème du choix des supports de publicité conduit à analyser le phénomène de la "duplication de lecture": si l'on place des annonces dans un certain nombre de magazines, quelle est la probabilité pour qu'une personne soit touchée par la même annonce dans plusieurs magazines ?

La duplication entre les magazines A et B peut être mesurée par le nombre de personnes qui lisent à la fois A et B: $N(AB)$, ou par la proportion de ces personnes dans la population: $P(AB)$.

Plus généralement on pourra représenter les duplications entre η supports par un ensemble de 2^n proportions:

$$\begin{aligned} &P(A), P(B), P(C) \dots \\ &P(AB), P(AC), P(BC) \dots \\ &P(ABC), P(ABD) \dots \end{aligned}$$

Ces nombres ne sont pas directement utilisables: d'une part en général une partie d'entre eux seulement sont connus; d'autre part, il suffit que le nombre de lecteurs d'un support se modifie pour que toutes les duplications où il figure se modifient.

On désire donc représenter la duplication par des coefficients qui présentent une certaine stabilité lorsque les proportions varient dans des limites à préciser et qui permettent par suite d'évaluer les valeurs probables des duplications inconnues à partir de celles qui sont connues.

Le problème semble présenter un intérêt très général puisqu'il se pose chaque fois que l'on doit manipuler des tableaux, dichotomiques à η dimensions (soit pour évaluer la valeur de certaines cases à partir des marges, soit pour apprécier la dépendance entre les diverses variables).

M.J. Ville (Principes d'analyse matricielle, page 191) définit un coefficient:

$$\Delta_{AB} = \frac{P(A/B)}{P(A)} = \frac{P(B/A)}{P(B)} = \frac{P(AB)}{P(A) P(B)}$$

Ces coefficients peuvent être généralisés pour un nombre quelconque de variables:

$$P(AB) = \Delta_{AB} P(A) P(B)$$

$$P(ABC) = \Delta_{ABC} \Delta_{AB} \Delta_{AC} \Delta_{BC} P(A) P(B) P(C)$$

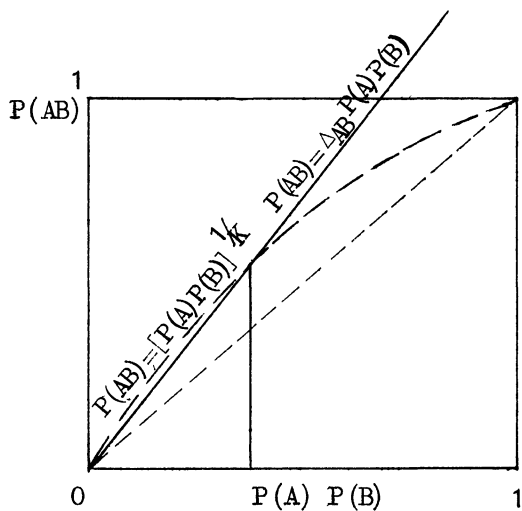
$$P(ABCD) = \Delta_{ABCD} \Delta_{ABC} \Delta_{ABD} \Delta_{ACD} \Delta_{BCD} \Delta_{AB} \Delta_{AC} \Delta_{AD} \Delta_{BC} \Delta_{BD} \Delta_{CD} P(A) P(B) P(C) P(D)$$

Ces coefficients ont l'avantage d'être simples et symétriques: ils utilisent la totalité de l'information disponible (il y a également 2^n coefficients Δ).

En outre, ils permettent de formuler de façon simple les diverses hypothèses d'indépendance (indépendance complète, indépendance marginale, indépendance partielle, etc...): chacune consiste à évaluer à 1 certains coefficients ou certains produits de coefficients. En fait, on constate d'ailleurs que lorsque le nombre de variables s'accroît les coefficients tendent vers 1.

Si ces coefficients étaient sensiblement constants, ils permettraient par exemple de calculer $P(AB)$ connaissant $P(A)$ et $P(B)$, etc...

Mais ils offrent l'inconvénient de ne pas être stables: par exemple, si $\Delta_{AB} > 1$, $P(AB)$ donné par la formule sera supérieur à 1 pour des valeurs élevées de $P(A)$ ou de $P(B)$ (cf. figure ci-jointe).



On est donc conduit à utiliser une relation non linéaire entre

$$P(AB) \text{ et } P(A) \cdot P(B):$$

On a obtenu par exemple de bons résultats expérimentaux avec:

$$P(AB) = \left[P(A), P(B) \right] \frac{1}{K}$$

Mais dans cette formule, le coefficient K n'a pas de signification intuitive et elle cesse d'être symétrique au-delà de 2 variables.

J. Durand