

SAGOMBAYE NODJIRAM

Inadéquation du modèle normal et description d'un histogramme par moyenne et écart type

Les cahiers de l'analyse des données, tome 16, n° 4 (1991),
p. 437-438

http://www.numdam.org/item?id=CAD_1991__16_4_437_0

© Les cahiers de l'analyse des données, Dunod, 1991, tous droits réservés.

L'accès aux archives de la revue « Les cahiers de l'analyse des données » 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/>

INADÉQUATION DU MODÈLE NORMAL ET DESCRIPTION D'UN HISTOGRAMME PAR MOYENNE ET ÉCART TYPE

[MODÈ. HISTOG.]

SAGOMBAYE NODJIRAM*

1 Valeurs négatives, valeurs fractionnaires et inadéquation du modèle normal

Dans une publication récente analysant les "Facteurs de risque de fracture de la hanche" nous relevons les données suivantes:

Facteur de risque	malades (176 cas)	témoins (672 cas)
consommation d'alcool	1,8 ± 3,3 once/sem	1,7 ± 2,9 once/sem
cigarettes fumées	2,8 ± 7,3 cig/jour	1,8 ± 5,7 cig/jour

Ces indications numériques, exprimées en moyenne ± écart type, sont absurdes: $1,8 - 3,3 = - 1,5$; $2,8 - 7,3 = - 4,5$: on ne peut consommer une quantité négative d'alcool! ni fumer un nombre négatif de cigarettes.

Il est bien connu que le calcul du couple {moyenne, écart-type} suggère une courbe en cloche symétrique unimodale, analogue à une distribution normale.

Ce modèle, très fréquemment utilisé, ne convient, en toute rigueur, que pour une distribution s'étendant de $-\infty$ à $+\infty$; en particulier, il est inadéquat pour toute grandeur essentiellement positive. C'est pourquoi, dans certain cas, on préfère postuler que c'est le logarithme de la grandeur mesurée qui suit une loi normale. Cependant, si la variable normale ajustée n'a qu'une faible probabilité d'être négative, le modèle n'est pas absurde.

Tel n'est pas le cas dans les exemples que nous avons relevés: non seulement la distribution ajustée s'étend nettement en deçà de zéro; mais, à un écart type à gauche, dans le noyau dense de la distribution, on trouve déjà des valeurs négatives.

(*) Étudiant en doctorat à l'Université Pierre et Marie Curie, Paris.

Voici un deuxième exemple issu d'une étude consacrée à la "Variété des pratiques transfusionnelles lors d'un pontage coronarien":

Hôpital	moyenne écart-type	valeurs extrêmes
H1	0,4 ± 0,2	0 ... 4
H2	0,9 ± 0,2	0 ... 4

Nombres d'Unités transfusées

S'il est admissible qu'une cigarette soit fumée à moitié, on ne peut considérer comme la norme le fractionnement des unités transfusées.

2 L'alternative d'une distribution concentrée en deux points

Sans connaître les distributions exactes, il est facile d'en deviner la forme. Certains sujets ne fument pas du tout, d'autres fument environ un paquet par jour, les cas intermédiaires sont rares. De même, certains s'abstiennent de boire et l'on pourra assimiler à une valeur typique ce que boivent les autres. Quant à l'Hôpital H1, la moitié environ des sujets n'y recevraient pas de transfusion, les autres recevant d'ordinaire une seule unité, à moins qu'on n'use de 1/2 unités?

Il est facile de préciser ces conjectures par le calcul. Soit ξ une variable aléatoire de moyenne μ et d'écart-type σ ; supposons que ξ ne prend que les deux valeurs 0 et M, la probabilité de M étant π : on a les relations:

d'où:

$$\mu = \pi M \quad ; \quad \mu^2 + \sigma^2 = \pi M^2 \quad ;$$

$$M = \mu + (\sigma^2 / \mu) \quad ; \quad \pi = \mu / M$$

Pour les cigarettes, avec $2,8 \pm 7,3$, on trouve : $M = 21.8 (\approx 20!)$ et $\pi = 0.13$: un sujet sur 8, environ, fume un paquet par jour. Et avec $1,8 \pm 5,7$ il vient: $M = 19.9 (\approx 20)$ et $\pi = 0.09$: un sujet sur 11 fume un paquet par jour.

Le cas des transfusions dans l'hôpital H1 est plus curieux. Notons {a, b, c, d} les probabilités respectives qu'on transfuse 1, 2, 3, 4 unités; il vient:

$$\mu = a + 2b + 3c + 4d = 0,4 \quad ;$$

$$\mu^2 + \sigma^2 = a + 4b + 9c + 16d = 0,2 \quad ;$$

le système est incompatible (à moins de supposer que les transfusions sont administrées par demi unités...).

Références

D. T. Felson & coll. : "Thiazide diuretics and the risk of hip fracture, Results from the Framingham Study"; *JAMA* 1991; 265; pp. 370-373; traduit dans *JAMA* fr., 15 mars 1991.
 L.T. Goodnough & coll. : "The variability of transfusion practice in coronary artery bypass surgery"; *JAMA*, 1991; 265; pp. 86-90; traduit dans *JAMA* fr. cardio., 30 Juin 1991.