J. H. C. LISMAN

Comparaison entre la distribution binomiale symétrique et la distribution « normale discontinue »

Revue de statistique appliquée, tome 20, n° 3 (1972), p. 85-86
<http://www.numdam.org/item?id=RSA_1972__20_3_85_0>

COMPARAISON ENTRE LA DISTRIBUTION BINOMIALE SYMÉTRIQUE ET LA DISTRIBUTION "NORMALE DISCONTINUE"

J.H.C. LISMAN

Bureau Central du Plan - La Haye

On pourrait se trouver incliné à considérer la distribution binomiale symétrique comme une analogie discontinue de la distribution normale. Cette opinion est vraisemblable, mais elle n'est pas juste, ce que nous allons démontrer brièvement ici.

En nous référant à Shannon [1], l'entropie d'une distribution avec densité p(x) est définie par $H = -\int p(x)\ \log(x)\ dx$. Quand H est maximum, nous trouvons la situation la plus aléatoire. En ce cas, nous désignons la distribution comme la plus probable [2].

La distribution de fréquence la plus probable d'une variable stochastique $\underline{x}$ avec une densité de probabilité p(x) peut être trouvée en maximisant son entropie $H = -\int p(x) \log p(x)\ dx$ sous les contraintes $E\ f_j(\underline{x}) = C_j$ $(j = 1, \ldots, n)$, où C_j sont des constantes données. Ce procédé a été réalisé par Shannon [1]. En appliquant la méthode multiplicatrice de Lagrange [3], il en résulte :

$$\frac{d\left\{-p(x)\ \log\ p(x)\ dx + \sum_{j=1}^{n} \lambda_j\ p(x)\ f_j(x)\right\}}{d\ p(x)} = 0$$

ou

$$-\log\ p(x) - 1 + \sum_{j=1}^{n} \lambda_j f_j(x) = 0\ ,$$

de sorte que

$$p(x) = e^{-1}\ e^{\sum_{j=1}^{n} \lambda_j f_j(x)},$$

ce qui est valable aussi pour des distributions discontinues.

La distribution normale est obtenue à l'aide de deux contraintes

$$\int_{-\infty}^{+\infty} p(x)\ dx = 1 \quad \text{et} \quad E(x-a)^2 = \sigma^2,$$

a et σ^2 étant donnés.

Il en résulte :

$$p(x) = e^{-1}\, e^{\lambda_1+\lambda_2(x-a)^2} = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-a)^2}{2\sigma^2}}.$$

En vertu de la procédure susmentionnée et sous les mêmes contraintes, l'analogie discontinue de la distribution normale est définie comme

$$p_i = e^{-1}\, e^{\lambda_1+\lambda_2(i-a)^2}.$$

Les multiplicateurs λ_1 et λ_2 ne se laissent pas représenter en termes algébriques, à défaut d'une procédure de sommation appropriée.

Il est clair que ce type de distribution montre une différence fondamentale, en comparaison avec la distribution binomiale. Nous présentons une illustration numérique et nous opposons

$$p_i = e^{-1}\, e^{\lambda_1+\lambda_2(i-a)^2} \qquad (i = 0, \ldots, 4 \; ; \; a = 2)$$

à

$$p_i = \left(\frac{1}{2}\right)^n \frac{n!}{i!\,(n-i)!} \qquad (i = 0, \ldots, 4 \; ; \; n = 4)$$

en prenant $\sigma^2 = 1$.

Dans ce cas simple, les multiplicateurs λ_1 et λ_2 se laissent déterminer par un calcul algébrique ordinaire. On trouve les probabilités suivantes :

discontinue normale :	0,0638	0,2447	0,3830	0,2447	0,0638
binomiale symétrique :	0,0625	0,2500	0,3750	0,2500	0,0625

Les différences sont petites.

[1] C.E. SHANNON - A Mathematical Theory of Communication, Bell System Technical Journal 27, 1948, p. 379 et 623.

[2] J.H.C. LISMAN et M.C.A. VAN ZUYLEN - Note on the generation of most probable frequency distributions, Statistica Neerlandica, 26, 1972, n° 1.

[3] R. WEINSTOCK - Calculus of Variations, with Applications to Physics and Engineering, New-York, 1952, 48-51.