

REVUE DE STATISTIQUE APPLIQUÉE

P. THIONET

Note de P. Thionet sur : Comparaison entre la distribution binomiale symétrique et la distribution « normale discontinue »

Revue de statistique appliquée, tome 20, n° 3 (1972), p. 87-88

http://www.numdam.org/item?id=RSA_1972__20_3_87_0

© Société française de statistique, 1972, tous droits réservés.

L'accès aux archives de la revue « Revue de statistique appliquée » (<http://www.sfds.asso.fr/publicat/rsa.htm>) 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/>

Note de P. THIONET sur :

COMPARAISON

ENTRE LA DISTRIBUTION BINOMIALE SYMÉTRIQUE ET LA DISTRIBUTION « NORMALE DISCONTINUE »

Le petit papier proposé à la R.S.A, nous a paru fort intéressant.

Point 1 : A toute loi d'une variable X , à densité $f(x)$, $\int_{-\infty}^{+\infty} f(x) dx = 1$, on peut faire correspondre la loi discrète :

$$pr(Y = j) = C \cdot f(j), j = \dots -6, -5, \dots, 0, 1, 2, \dots$$

j pouvant prendre toute valeur entière, —ou un sous-ensemble de celles-ci, selon la valeur de la constante C . La convergence de l'intégrale de $f(x)$ assure celle de la série de terme général $u_n = f(n)$, ou encore $f(-n)$.

C'est d'ailleurs un fait banal pour la loi exponentielle (continue sur $0, +\infty$) et la loi "géométrique" (discrète, sur N) : $u_n = q p^n$, $q = 1 - p$.

Ainsi la loi "normale discontinue" de LISMAN et Van ZUYLEN n'a rien de bien mystérieux.

Point 2 : Il est bien vrai que le maximum de $H = - \int p(x) \log p(x) dx$ sous des contraintes du type $\int p(x) g_j(x) dx = c_j$, $j = 1, 2, \dots, m$, comme le maximum de $H^* = - \sum p(i) \log p(i)$, sous les contraintes : $\sum p(i) g_j(i) = c_j^*$, correspondent à : $\log p(x) = -1 + \sum_j \lambda_j g_j(x)$, $\log p(i) = -1 + \sum_j \lambda_j g_j(i)$.

Cependant, il paraît abusif de définir une loi binomiale par l'espérance μ et la variance σ^2 , car la loi est définie par un paramètre p , pris sur $(0, 1)$ et un paramètre n essentiellement entier positif. Posant $np = c_1$, $np(1 - p) = c_2$, on en tire : $n = c_1^2 / (c_1 - c_2)$, $p = (c_1 - c_2) / c_1$. Néanmoins on doit avoir : $0 < c_2 < c_1$ et surtout : c_1^2 divisible par $(c_1 - c_2)$.

Il paraît excessif d'avoir l'air de penser qu'on pourra ne pas spécifier, dans la définition de la distribution binomiale, sur quel sous-ensemble de valeurs de i la probabilité $p(i)$ est irrémédiablement nulle. (je veux dire : $i < 0$ et $i > n$). Quoi d'étonnant à ce qu'on trouve alors un certain $p(i)$ non nul pour toutes valeurs entières de la variable pour l'homologue de la loi normale ? Inversement, partant d'une loi continue définie sur $(0, +\infty)$, qui se soucie des i entiers négatifs ? Nous écrivons p_i pour $p(i)$.

Donnons-nous le support $(0, 1, 2, 3)$ de la distribution, ce qui (compte tenu de la condition $p_0 + p_1 + \dots = 1$) permet déjà de maximiser $-H^*$ pour EX et VX imposés, On trouve justement la distribution binomiale !

En effet, $p_0 = 1 - p_1 - p_2 - p_3$; EX et EX^2 ne renferment pas p_0 ; dans H^* on porte l'expression de p_0 et en dérive en p_1 , p_2 et p_3 ; il vient :

$$\log p_0 - \log p_i = -\lambda_i - \mu i^2 \quad \text{Identifions avec la loi binomiale :}$$

ou :

$$\text{Log } p_1/p_0 = (\lambda + \mu) = K \dots = \text{Log } p/q + \text{Log } 3$$

$$\text{Log } p_2/p_0 = 2\lambda + 4\mu = 2K + 2M = 2\text{Log } p/q + \text{Log } 3$$

$$\text{Log } p_3/p_0 = 3\lambda + 9\mu = 3K + 6M = 3\text{Log } p/q$$

L'identification se révèle possible : $\text{Log } p/q = K + 2M$, $2M = -\text{Log } 3$,
 $K = \text{Log } \frac{3p}{q}$, $2K + 2M = 2\text{Log } 3 + 2\text{Log } p/q - \text{Log } 3 = 2\text{Log } p/q + \text{Log } 3$.

Conclusion : la distribution binomiale maximise $-H$ (pour EX et VX donnés) sur le support $(0, 1, 2, 3)$.

Malheureusement, ce n'est déjà plus exact sur $(0, 1, 2, 3, 4)$.

On peut cependant déterminer la nouvelle loi ainsi définie sur $(0, 1, \dots, n)$:

$$p_i/p_0 = K^i M^{i(i-1)}$$

p_0 étant choisi pour normer la distribution sur $(0, 1, \dots, n)$.

Si n tend vers l'infini, on ne trouve pas la loi de Poisson.

Point 3 : Pour trouver les lois binomiales et de Poisson, on peut songer : soit à modifier l'expression de H^* (en fait à substituer à Log un autre opérateur car l'information de Shannon n'est qu'un cas particulier d'information au sens de Schützenberger, fonctionnelle linéaire très générale) ; - soit à modifier la notion d'écart sur un axe. Ce qu'on va faire ici :

Avec une loi de Poisson, on a :

$$\text{Log } p_i/p_{i-1} = K - \text{Log } i,$$

au lieu de :

$$\text{Log } p_i/p_{i-1} = K + M [i(i-1) - (i-1)(i-2)]$$

ci-dessus :

autrement dit, la seconde contrainte ne consiste pas à imposer $VX (= EX)$, mais à imposer la valeur de la moyenne géométrique de X , soit GX :

$$\text{Log } GX = p_1 \text{Log } 1 + p_2 \text{Log } 2 + p_3 \text{Log } 3 + \dots$$

moyenne calculée (bien entendu) sur la distribution tronquée à l'origine.

Le cas binomial est un peu moins simple : la seconde contrainte est

$$GX/G(n-X) = C^{\text{re}}$$

ce qui postule que la distribution n'existe que sur $(0, 1, 2, \dots, n)$; la contrainte est $GX/F(n-X) = 1$ pour la loi binomiale symétrique.

P. Thionet