Loi normale et information

par Guy CHERBIT et Jean BARRAUD
Université de Paris VII

INTRODUCTION

La probabilité d'un événement E mesurant le degré de vraisemblance a priori de cet événement, il est légitime de poser :

$$I(E) = \text{Log} \frac{1}{P(E)}$$

comme étant la mesure de l'incertitude a priori de l'événement E ; le logarithme permettant, pour tout couple d'événements indépendants, d'écrire :

$$I(E \cap F) = I(E) + I(F)$$

On appelle alors incertitude (ou entropie) d'un aléa X la valeur moyenne des incertitudes $I(X=x)$, soit, dans le cas d'un aléa continu de densité $f(x)$:

$$H(X) = \int_{\mathbb{R}} f(x) \text{Log} \frac{1}{f(x)} dx$$

On appelle information (au sens de Shannon) l'opposé de l'entropie, c'est-à-dire :

$$I(X) = \int_{\mathbb{R}} f(x) \text{Log} f(x) dx$$

La loi normale se prête facilement à ces calculs et est particulièrement bien adapté à l'étude des sommes de variables indépendantes.

Nous allons montrer qu'il s'agit de beaucoup plus qu'une simple adéquation strictement formelle.

LOI NORMALE ET INFORMATION

Lorsque d'une distribution théorique quelconque on extrait des échantillons, ceux-ci n'ont en général pas tous la même moyenne ni la même variance.

Les fluctuations d'échantillonnage observées sont d'autant plus faibles que la variance de la distribution théorique est petite. Si en effet la dispersion autour de la moyenne est petite, toutes les valeurs sont groupées autour de la moyenne et les échantillons diffèrent peu entre eux.

L'information d'un échantillon quelconque est donc de la forme :

$$I = I_0 + \varphi(v)$$

où φ est une fonction décroissante et I_0 l'information intrinsèque. Nous signifions par là que tout échantillon porte en lui "l'empreinte" de la distribution théorique.

Si l'on choisit $\varphi(v) = \alpha v + \beta$ avec $\alpha < 0$

On a alors :

$$I_0 = I - \alpha v - \beta$$

Soit encore, dans le cas d'une distribution continue

$$I_0 = \int_{\mathbb{R}} \left[f(x) \log f(x) - \alpha (x-\bar{x})^2 f(x) - \beta f(x) \right] dx$$

Posons alors le problème suivant :

"Parmi toutes les distributions continues théoriques, quelles sont celles qui véhiculent une information intrinsèque minimale ?".

Les conditions limites sont imposées par le fait que $f(x)$ est une densité, c'est-à-dire

$$f(-\infty) = f(+\infty) = 0$$

La densité cherchée est solution de l'équation de Lagrange qui s'écrit ici :

$$1 + \log f(x) - \alpha (x-\bar{x})^2 - \beta = 0$$

d'où

$$\log f(x) = \alpha (x-\bar{x})^2 + \beta - 1$$

$$f(x) = e^{\alpha (x-\bar{x})^2} e^{\beta - 1}$$

Compte tenu de :

$$\int_{\mathbb{R}} f(x) dx = 1 \quad \text{et} \quad \int_{\mathbb{R}} f(x) \cdot (x-\bar{x})^2 dx = \sigma^2$$

On trouve les valeurs des constantes α et β :

$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\bar{x})^2}{2\sigma^2}}$$

La loi normale rend donc I_0 extrémale.

En choisissant $g(x)$ pour que $f(x) + g(x)$ soit une densité, il est immédiat que

$$I_0(f + g) \geq I_0(f)$$

La distribution cherchée est donc bien la loi normale.

CONCLUSION

Le résultat précédent explique l'importance de la loi de Gauss et la raison pour laquelle elle est une loi limite. La convergence vers la loi normale traduit donc une minimalisation de l'information d'une distribution.

Cela *était* prévisible puisque lorsque l'effectif augmente et que toutes les valeurs sont distribuées au hasard, l'information totale diminue (et l'on tend vers la loi normale).

En terme d'information, la loi de Gauss est donc une loi économique.

N.B. : Le résultat démontré nous ayant paru fondamental, nous avons entrepris des recherches bibliographiques une fois notre travail terminé. Nous n'avons rien trouvé d'analogue si ce n'est dans l'ouvrage de J.M. FAVERGE, "Méthodes Statistiques en psychologie appliquée", tome 3^{ème}, page 882, où l'auteur rapproche la loi normale de la notion d'information, mais en se plaçant dans un contexte sensiblement différent.