Journal d'imagerie diagnostique et interventionnelle 2017;1:S4-S8

Loi normale de Laplace-Gauss

The normal or Gaussian distribution; what you have to know!

P. Ingrand [a,b]

[a]*Épidémiologie & biostatistique, Inserm CIC 1402, 6 rue de la Milétrie, TSA 51115, 86073 Poitiers cedex, France*
[b]*Université de Poitiers, 6 rue de la Milétrie, TSA 51115, 86073 Poitiers cedex, France*

RÉSUMÉ

La loi normale occupe une place centrale en statistique, autant d'un point de vue théorique que dans ses applications. Elle renvoie vers des utilisations très fréquentes et diverses. C'est la raison pour laquelle quelques brefs rappels de notions élémentaires sur cette loi de probabilité importante sont proposés ici.

MOTS-CLÉS
Loi de probabilité
Statistique
Loi normale

KEYWORDS
Normal distribution
Gaussian distribution
Statistics

SUMMARY

In probability theory, the normal (or Gaussian) distribution is a common continuous probability distribution. Normal distributions are important in statistics and are often used in medicine to represent real-valued random variables, which distributions are not known. The normal distribution is associated with the central limit theorem. In its most general form, under some conditions (which include finite variance), this theorem states that averages of samples of observations of random variables independently drawn from independent distributions converge in distribution to the normal when the number of observations is sufficiently large. This article presents what a radiologist has to know on this distribution.

HISTORIQUE

Alors que l'intuition d'une loi générale gouvernant les chances de gains au jeu avait été formulée par J. Bernoulli dès 1677, la paternité de la loi normale est à partager entre le français P.-S. de Laplace et l'allemand F. Gauss à partir de 1777, à la suite de travaux menés parallèlement à propos de l'analyse des erreurs lors des observations astronomiques, d'où le nom de « loi normale de Laplace-Gauss ».
Ces premiers résultats ont trouvé tout leur sens en 1812 avec la formulation par Laplace du théorème central limite (voir ci-dessous), principe fondateur de la théorie des fluctuations d'échantillonnage, et donc de la science statistique. Le théorème central limite apporte la preuve du caractère universel de la loi normale pour toute caractéristique résultant de la répétition d'événements similaires à caractères aléatoire et indépendant. C'est par cette propriété fondamentale que cette loi sort du champ théorique des probabilités pour s'inscrire dans celui des statistiques et notamment du raisonnement inductif basé sur les observations tirées d'échantillons (l'inférence statistique).

PROPRIÉTÉS

Sa forme de courbe en cloche est bien connue (*Fig. 1*), et Bru nous en offre une description plus poétique : « C'est une cloche plate ou un

Auteur correspondant :
P. Ingrand,
Université de Poitiers, 6 rue de la Milétrie, TSA 51115, 86073 Poitiers cedex, France
Adresse e-mail :
pierre.ingrand@univ-poitiers.fr

chapeau de gendarme vu de face, lorsque les gendarmes en portaient un de cette forme-là (éventuellement, un boa qui a avalé un dromadaire). »

La loi normale comporte deux paramètres :

- la moyenne, notée μ (« mu » la lettre *m* grecque) qui correspond au centre de la distribution ;
- l'écart type, noté σ (« sigma » le *s* grec) qui représente la dispersion de part et d'autre de la moyenne.

Une variable *x* qui suit une loi normale de moyenne μ et d'écart type σ sera notée N(μ, σ).

On peut ainsi définir en théorie autant de lois normales différentes qu'il existe de couples moyens, écart type.

Par la transformation centrée réduite : toute loi normale de moyenne μ et d'écart type σ est ramenée à une loi unique, la loi normale centrée réduite (*z*) de moyenne nulle (c'est-à-dire une loi centrée sur 0) et d'écart type égal à 1 (c'est-à-dire une loi dont la dispersion est réduite à l'unité). On la nomme également « loi normale standard ».

$$z = \frac{x - \mu}{\sigma} \quad N(0, 1)$$

De ce fait, la loi centrée réduite est la seule loi normale qui mérite un intérêt particulier.

La courbe en cloche représente la fonction de densité de probabilité (*Fig. 1*) de la loi normale. Il s'agit d'une fonction continue, définie (en théorie) de $-\infty$ à $+\infty$. À partir de cette loi, une probabilité correspond à une surface comprise sous la courbe en cloche, calculée sur un intervalle (et non pas à une valeur ponctuelle de la fonction comme ce serait le cas avec une loi discrète).

Comme toute loi de probabilité, sa somme (sa surface totale) est égale à 1.

Ainsi : Pr($Z < +\infty$) = 1 ou Pr($Z > -\infty$) = 1 (avec la notation Pr : « probabilité de »).

La loi normale est parfaitement symétrique : Pr($Z > z$) = Pr($Z < -z$).

Des niveaux de probabilité remarquables sont associés à des écarts à la moyenne :

- 95 % dans l'intervalle moyenne ± 2 fois l'écart type (exactement 1,96 fois, *Fig. 1*) ;
- 68 % (approximativement deux tiers) dans l'intervalle moyen ± 1 fois l'écart type ;
- 99,7 % (la quasi-totalité) dans l'intervalle moyen ± 3 fois l'écart type.

Les valeurs de probabilité sont obtenues à partir de la fonction de répartition de la loi normale qui est définie comme l'intégrale depuis $-\infty$ jusqu'à *z* de la loi normale centrée réduite (*Fig. 2*).

Théorème central limite : la loi de probabilité de la somme (et par extension, de la moyenne) de variables aléatoires indépendantes et identiquement distribuées tend vers la loi normale quand le nombre de répétitions *n* tend vers l'infini. Ce théorème est vrai quelle que soit la distribution des variables considérées (non normales). La convergence s'effectue d'autant plus rapidement que la forme de cette distribution est plus proche de celle de la loi normale, et en particulier plutôt symétrique. Par consensus, il est admis pour toute loi de variable quantitative qu'une convergence satisfaisante est obtenue à partir de $n \geq 30$. C'est ce qui explique la dichotomie dans les applications statistiques entre les grands échantillons et les petits échantillons ($n < 30$).

La loi normale constitue donc la loi des moyennes. Par extension, elle est aussi très souvent utilisée comme loi des proportions (par convergence de la loi binomiale), et de beaucoup de paramètres observés sur échantillons. Mais contrairement à une idée reçue, peu de variables biométriques,

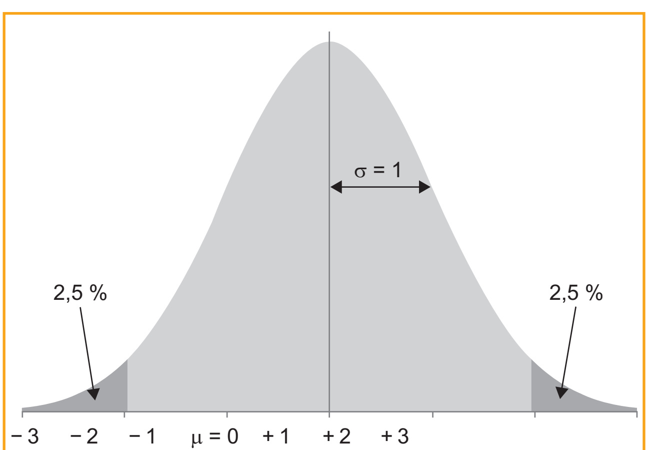

Figure 1. Fonction de densité de probabilité de la loi normale centrée réduite (loi *z*) de moyenne $\mu = 0$ et d'écart type $\sigma = 1$. La zone gris foncé représente 5 % de la distribution et la zone gris clair 95 %.

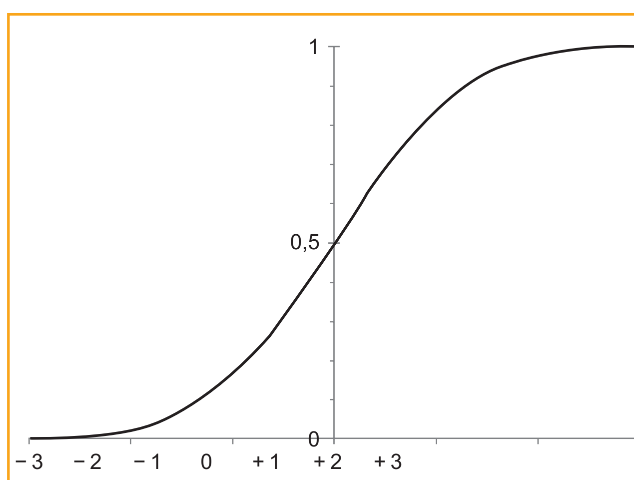

Figure 2. Fonction de répartition de la loi normale centrée réduite (loi *z*). L'ordonnée représente la probabilité d'observer une valeur de *Z* inférieure à une valeur *z* donnée en abscisse Pr($Z < z$). Par exemple Pr($Z < 0$) = 0,5.