

PLAIDOYER POUR LA LOI NORMALE

AIMÉ FUCHS

Parmi les nombreuses lois de probabilité qui constituent l'arsenal de la statistique, la loi normale occupe assurément une place de choix, tant dans la théorie que dans les applications. C'est aussi la seule loi, nous semble-t-il, qui soit connue du grand public. Qui ne connaît pas en effet la célèbre "courbe en cloche" ou "courbe de Gauss" qui en est le graphe? C'est que la loi normale s'est imposée depuis fort longtemps à l'attention des observateurs dans les disciplines les plus diverses : physique, physiologie, génétique, biologie, ... On y rencontre partout des diagrammes de fréquence ayant approximativement l'allure d'une courbe en cloche. La quasi-universalité de ces observations, jointe à l'attrait exercé par la forme harmonieuse de la courbe, n'a pas manqué de frapper les esprits. C'est ainsi que F. Galton (1822–1911), le célèbre physiologiste anglais, qui fut l'un des premiers à introduire la méthodologie statistique dans sa discipline, a fait part de son enthousiasme dans les termes suivants :

"I know of scarcely anything so apt to impress the imagination as the wonderful form of cosmic order expressed by the Law of Frequency of Error... It reigns with serenity and in complete self-effacement amidst the wildest confusion."

Natural Inheritance, 1889, p. 62.

Le lecteur aura compris que par "Law of Frequency of Error," il fallait entendre "loi normale" (voir Fig. 1).

C'est en statistique, et principalement dans la théorie des tests d'hypothèses, que la loi normale a trouvé son domaine d'application privilégié. Il suffit d'ouvrir un traité pour s'apercevoir que la plupart des lois usuelles (Student, Fisher, Snédécour, Chi-Deux, ...) relèvent de près ou de loin de la loi normale.

Observons néanmoins qu'il y a des cas où la loi normale est contre-indiquée pour décrire un phénomène. Un exemple suffira pour le montrer. Considérons la loi d'Ohm $I = V/R$ et supposons que la tension V soit non aléatoire et connue, mais que la résistance R soit aléatoire et suive une loi normale, ce qui, à première vue, paraît raisonnable. Or cette hypothèse aurait pour conséquence que l'intensité du courant I n'admet pas d'espérance mathématique, ce qui est difficilement acceptable pour l'ingénieur. Ainsi l'hypothèse que R suit une loi normale est à rejeter comme irréaliste.

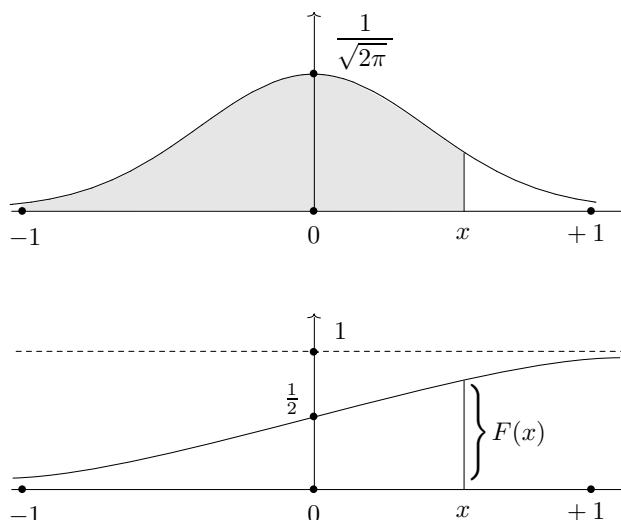


Figure 1. Loi normale

Cependant en dépit d'exemples de ce type qu'il serait loisible de multiplier la loi normale garde toute son importance, comme la suite de cet exposé s'emploiera à montrer. Nous verrons que les propriétés les plus profondes de la loi normale sont de nature mathématique. Nous les avons regroupées dans des paragraphes selon qu'elles concernent la théorie de l'information, la géométrie, la statistique théorique. Nous faisons précéder leur présentation d'un bref historique.

1. Historique de la loi normale. — La loi normale est associée aux noms prestigieux de Gauss et de Laplace; dans la littérature allemande et anglo-saxonne elle est connue sous le nom de *Loi de Gauss*; certains mathématiciens français y ont fait référence sous le nom de *Loi de Laplace*; pour mettre tout le monde d'accord, Maurice Fréchet a proposé de l'appeler *Loi de Laplace-Gauss*. C'est cet usage qui a prévalu en France.

Ces mathématiciens ont tout deux introduit cette loi, mais leurs approches étaient fondamentalement différentes. Celle de Gauss, d'inspiration statistique, est la plus originale et a été à l'origine de la *théorie de l'estimation par le maximum de vraisemblance*. Celle de Laplace, d'inspiration probabiliste, prolonge les idées de de Moivre et a donné naissance à la problématique liée au "théorème central limit." nous allons résumer ci-dessous ces deux approches.

L'approche de Gauss [4; surtout les paragraphes 175-178]. — La loi normale a été introduite par Gauss à propos d'un problème d'estimation de paramètre, dans un contexte totalement étranger au Calcul des Probabilités, celui du mouvement des corps célestes. Désignons par θ une

quantité (inconnue) dont n observations indépendantes ont fourni les n approximations x_1, \dots, x_n . Gauss se propose d'estimer θ à partir des seules valeurs observées x_1, \dots, x_n .

a) Une première estimation de θ est fournie par la *méthode des moindres carrés*; elle consiste à prendre pour estimation de θ la valeur $\tilde{\theta}$ qui réalise le minimum de la fonction $\theta \mapsto \sum_{k=1}^n (x_k - \theta)^2$; un calcul immédiat donne $\tilde{\theta} = (x_1 + \dots + x_n)/n = \bar{x}$ (la moyenne arithmétique de (x_1, \dots, x_n)).

b) Gauss propose une autre méthode, qui, dans le langage actuel, est connue sous le nom de *méthode du maximum de vraisemblance*. Il introduit une fonction $f(x)$ qui représente la (densité de) probabilité de l'erreur x commise lors d'une observation; il fait sur f les hypothèses (H) suivantes :

$$f > 0, \int_{-\infty}^{+\infty} f(x) dx = 1;$$

f est paire (i.e., on commet aussi souvent une erreur positive qu'une erreur négative);

$f(x)$ est décroissante lorsque $|x| \rightarrow +\infty$ (i.e., on commet moins souvent de grandes erreurs que des petites).

Les erreurs afférentes aux valeurs observées x_1, \dots, x_n de θ sont $x_1 - \theta, \dots, x_n - \theta$, de sorte que la (densité de) probabilité pour que ces erreurs aient été commises simultanément est, compte tenu de l'indépendance des observations :

$$(1) \quad L(\theta) = f(x_1 - \theta) \dots f(x_n - \theta).$$

(En langage moderne, c'est la *fonction de vraisemblance*.)

Il est à présent naturel de chercher les densités f vérifiant les hypothèses (H) telles que la fonction $L(\theta)$ définie par (1) prenne sa valeur maximale au point $\theta = \tilde{\theta} = \bar{x}$. En langage moderne, *on cherche les densités f vérifiant (H) telles que l'estimation de θ par la méthode des moindres carrés coïncide avec son estimation par le maximum de vraisemblance*. Gauss montre que les seules densités qui satisfont ces conditions sont les densités *normales* (centrées). Son raisonnement, élémentaire et digne du grand Gauss, mérite d'être reproduit.

c) Ecrivons (1) sous la forme :

$$\text{Log } L(\theta) = \text{Log } f(x_1 - \theta) + \dots + \text{Log } f(x_n - \theta).$$

On cherche f vérifiant (H) telle que, pour tout n , pour tous x_1, \dots, x_n , on ait

$$\left[\frac{\partial}{\partial \theta} \text{Log } L(\theta) \right]_{\theta=\bar{x}} = 0,$$

c'est-à-dire, en posant $g = \text{Log } f$,

$$g'(x_1 - \bar{x}) + \cdots + g'(x_n - \bar{x}) = 0.$$

Pour résoudre cette équation, Gauss prend $x_2 = \cdots = x_n = x_1 - ny$ avec y réel. Alors $\bar{x} = x_1 - (n-1)y$; d'où $x_1 - \bar{x} = (n-1)y$ et $x_2 - \bar{x} = \cdots = x_n - \bar{x} = -y$. On a alors, pour tout n et tout y

$$g'[(n-1)y] + (n-1)g'(-y) = 0.$$

Comme, par hypothèse, f est paire, on a $g'(-y) = -g'(y)$, d'où

$$g'[(n-1)y] = (n-1)g'(y),$$

ou encore

$$\frac{g'[(n-1)y]}{(n-1)y} = \frac{g'(y)}{y} = k;$$

soit encore $g'(x) = kx$ et $\text{Log } f(x) = g(x) = k(x^2/2) + C$. En adaptant les constantes de telle façon que f vérifie les hypothèses (H), on voit que f est la *densité d'une loi normale* (centrée).

L'approche de Laplace [7]. — Elle prolonge les recherches du mathématicien anglais de Moivre [10] sur l'approximation de la loi binomiale par la loi normale. Cette approche peut être décrite de la façon suivante.

Procédons à une suite de parties indépendantes de "Pile" ou "Face." Jouons avec une pièce de monnaie *parfaite* (ce qui signifie que la probabilité d'amener "Pile" en une partie est $\frac{1}{2}$, quelle que soit la partie). Désignons par X_n ($n \geq 1$) le nombre de "Pile" amenées en les n premières parties; il est aisé de voir que la probabilité pour que X_n prenne la valeur x , où x est un entier compris entre 0 et n , est égale à $p_x = (1/2^n) \binom{n}{x}$, le symbole $\binom{n}{x}$ désignant le coefficient binomial égal à $n!/(x!(n-x)!)$. La variable aléatoire X_n est pour cette raison appelée *variable aléatoire binomiale* et sa loi, c'est-à-dire, l'ensemble des valeurs de p_x pour x variant dans l'ensemble $\{0, 1, \dots, n\}$, est notée $B(n, \frac{1}{2})$. Cette loi peut d'ailleurs être représentée par un diagramme à bâtons, qui, à toute valeur de x associe la valeur correspondante de p_x .

Or la moyenne $\mu = n/2$ et l'écart-type $\sigma = \frac{1}{2}\sqrt{n}$ de X_n tendent tous deux vers l'infini lorsque le nombre n de parties tend vers l'infini. Pour pallier cet inconvénient, on effectue sur X_n un changement d'origine et un changement d'échelle de façon à obtenir une variable aléatoire centrée (i.e., de moyenne nulle) et réduite (i.e., d'écart-type unité). La variable aléatoire $Y_n = \frac{X_n - \mu}{\sigma} = \frac{X_n - \frac{n}{2}}{\frac{1}{2}\sqrt{n}}$ réalise ces deux impératifs; sa loi est donnée par les valeurs : $p_y = p_x$, $y = \frac{x - \mu}{\sigma}$ ($x \in \{0, 1, \dots, n\}$).

Le diagramme à bâtons des p_y se déduit de celui des p_x en effectuant sur les abscisses le changement d'origine et le changement d'échelle dont il a été question ci-dessus. Dans les Figures 1_a et 1_b, nous donnons les diagrammes des p_x et des p_y pour les valeurs $n = 2, 4, 8, 16$. On constatera que pour les valeurs $n = 8, n = 16$, le diagramme de p_y ressemble fortement à une courbe en cloche. Cette tendance ne fera que s'accroître lorsque n augmente, au point que, lorsque n tend vers l'infini, le diagramme des p_y "converge" vers la célèbre courbe en cloche. On exprime ceci en langage moderne comme suit :

Pour tout x réel,

$$P\{Y_n \leq x\} \rightarrow \frac{1}{\sqrt{2\pi}} \int_0^x e^{-t^2/2} dt \quad (n \rightarrow +\infty);$$

et l'on dit que la suite des variables aléatoires Y_n converge (en loi) vers une variable aléatoire de loi normale $\mathcal{N}(0, 1)$ centrée, réduite. Ce résultat avait été démontré par A. de Moivre, dès 1733, par des calculs laborieux d'estimation des coefficients binomiaux; sa démonstration fut incluse dans son ouvrage fondamental *The doctrine of chances* [10]. Ce fut à l'époque une des premières performances à mettre à l'actif du Calcul des Probabilités.

Au siècle dernier, F. Galton a inventé un dispositif fort ingénieux pour simuler la tendance de la loi binomiale $B(n, \frac{1}{2})$ vers la loi normale; il fit tomber des billes (du sable) sur un réseau de clous plantés en quinconce sur une planche inclinée (planche de Galton). La figure 2 (page suivante) est suffisamment parlante pour nous dispenser de tout commentaire.

Laplace reprit le problème de de Moivre, mais traita le cas général où la pièce de monnaie n'est pas nécessairement parfaite, c'est-à-dire, où la probabilité d'amener "Pile" en une partie est égale à un nombre quelconque p compris strictement entre 0 et 1. Le nombre X_n ($n \geq 1$) de "Pile" amenées en les n premières parties admet encore une loi binomiale, mais cette dernière est la loi $B(n, p)$ donnée par

$$p_x = P\{X_n = x\} = \binom{n}{x} p^x q^{n-x}, \quad x \in \{0, 1, \dots, n\}, \quad q = 1 - p.$$

La moyenne $\mu = np$ et l'écart-type $\sigma = \sqrt{npq}$ tendent encore tous deux vers l'infini lorsque n tend vers l'infini. On remplace encore X_n par la variable aléatoire centrée réduite $Y_n = \frac{X_n - \mu}{\sigma} = \frac{X_n - np}{\sqrt{npq}}$.

Laplace essaie ensuite de démontrer, pour cette nouvelle suite (Y_n), un résultat analogue à celui auquel avait abouti de Moivre dans le cas

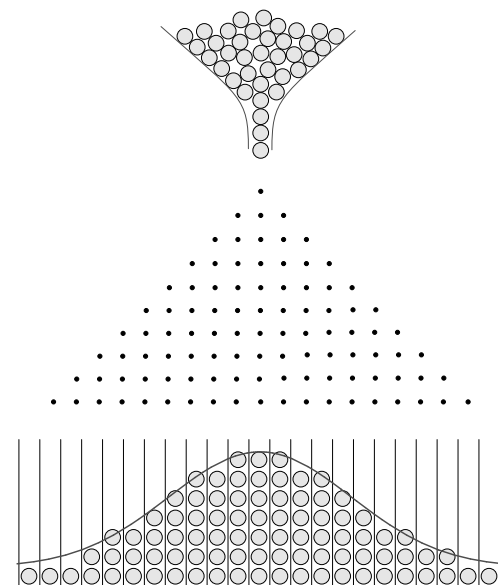


Figure 2. Planche de Galton

symétrique $p = \frac{1}{2}$. Il y parvient effectivement et démontre (avec des notations modernes) que pour tout x réel

$$P\left\{\frac{X_n - np}{\sqrt{npq}} \leq x\right\} \rightarrow \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt \quad (n \rightarrow \infty),$$

c'est-à-dire, que la suite des $Y_n = \frac{X_n - np}{\sqrt{npq}}$ converge (en loi) vers une variable aléatoire de loi $\mathcal{N}(0, 1)$.

Pour établir cette convergence Laplace invente une méthode nouvelle où pointe déjà la théorie des fonctions caractéristiques qui permet à l'heure actuelle d'arriver au résultat sans trop de difficultés.

L'étude de l'approximation de la loi binomiale par la loi normale n'a été pour Laplace qu'un prélude à des développements plus profonds. Il a en effet mis la loi normale en rapport avec la théorie des erreurs d'observation et c'est finalement cette approche qui s'est par la suite avérée la plus féconde. Donnons de cette approche un aperçu succinct.

Procédons à une suite d'observations et désignons par X_n ($n \geq 1$) l'erreur afférente à la $n^{\text{ième}}$ observation. On interprète les X_n comme des variables aléatoires indépendantes centrées et du second ordre, c'est-à-dire, si l'on désigne par σ_n l'écart-type de X_n , on suppose $0 < \sigma_n < \infty$ pour tout $n \geq 1$. Introduisons ensuite la somme $S_n = X_1 + \dots + X_n$ des n premières erreurs. On s'intéresse au comportement de cette somme pour de grandes valeurs de n et à cet effet on commence par la normer pour obtenir une

variable aléatoire d'écart-type unité. La variable aléatoire $Y_n = \frac{S_n}{\sigma(S_n)}$, où $\sigma(S_n)$ désigne l'écart-type de S_n , satisfait cette condition.

Laplace se pose alors la question de savoir à *quelles conditions la loi de Y_n est approchée par la loi normale $\mathcal{N}(0, 1)$ pour de grandes valeurs de n* . Une analyse très fine, dont il n'est pas question ici de donner les détails, lui a montré qu'une condition essentielle pour qu'il en soit ainsi est que :

“Le plus grand des écarts-type individuels est négligeable devant l'écart-type de la somme.”

ou, en langage mathématique,

$$\frac{1}{\sigma(S_n)} \sup_{1 \leq k \leq n} \sigma_k \rightarrow 0 \quad (n \rightarrow \infty). \quad (1.1)$$

Cette condition est par exemple réalisée si toutes les erreurs X_n sont indépendantes et ont même loi.

2. Les continuateurs de Laplace. — Nous avons déjà mentionné que c'était l'approche de Laplace qui s'était avérée la plus féconde pour le développement du Calcul des Probabilités. Elle contenait en germe ce que est devenu par la suite le problème “central limit.” Il s'agit d'établir des conditions (suffisantes) pour qu'une suite de variables aléatoires converge (en loi) vers une variable aléatoire normale. Tout théorème qui fournit des conditions de ce type est appelé “théorème central limit.” C'est Polyà [11] qui, en 1920, a désigné un tel théorème par le nom de “Zentraler Grenzwertsatz,” qui fut traduit en anglais par “central limit theorem” pour être adapté en français sous la dénomination de “théorème central limite.” Ce n'est pas le lieu ici d'aborder le détail de cette problématique. Nous voudrions néanmoins attirer l'attention du lecteur sur le fait que, parmi les hypothèses de tout théorème central limite figure une condition d'une nature tout à fait intuitive et qui, pour la suite de Laplace étudiée au paragraphe précédent, de terme général $Y_n = \frac{S_n}{\sigma(S_n)}$, $S_n = X_1 + \dots + X_n$, n'est autre que la condition (1.1). Cette condition est en effet essentielle pour la convergence (en loi) de la suite (Y_n) vers une variable aléatoire normale et Laplace lui-même s'était rendu compte que, s'il y avait dans la somme S_n une variable aléatoire disproportionnée par rapport aux autres, elle ferait prévaloir sa loi. Cependant il n'avait pas vu que cette condition n'était pas tout à fait suffisante et qu'il fallait lui adjoindre une condition de régularité supplémentaire. Les continuateurs de Laplace, en premier lieu J.L. Lindeberg [8] ont bien vu le problème et ont énoncé des théorèmes en conséquence, même dans des cas de suites plus générales que celles de Laplace.

L'exploitation la plus fructueuse de l'idée sous-jacente à la condition (1.1) a été faite par Gnedenko et Kolmogorov [6], qui ont montré qu'*une variable aléatoire normale est la somme d'un grand nombre de variables aléatoires indépendantes petites dont aucune n'est prépondérante*. Cette propriété rend bien compte de la quasi-universalité de la loi normale dans les disciplines les plus diverses.

La proposition écrite en italique peut s'énoncer de façon précise comme suit.

THÉORÈME 2.1. — *Considérons une suite de variables aléatoires X_n dont chacune est la somme d'un nombre fini de variables aléatoires $X_{n,1}, \dots, X_{n,k_n}$ avec $k_n \rightarrow +\infty$ ($n \rightarrow +\infty$). Pour tout $\varepsilon > 0$ introduisons la*

variable aléatoire tronquée $X^\varepsilon = \begin{cases} X, & \text{si } |X| \leq \varepsilon; \\ 0, & \text{sinon;} \end{cases}$ et supposons

- a) $\sum_{1 \leq k \leq n} |X_{nk}| \rightarrow 0$ ($n \rightarrow \infty$) (en probabilité);
 b) *Pour tout $\varepsilon > 0$ $\sum_{1 \leq k \leq n} \mathbb{E}[X_{nk}^\varepsilon] \rightarrow \mu$ et $\sum_{1 \leq k \leq n} \text{Var}(X_{nk}^\varepsilon) \rightarrow \sigma^2$ quand n tend vers l'infini.*

Alors la loi de X_n converge vers la loi normale $\mathcal{N}(\mu, \sigma^2)$, de moyenne μ et de variance σ^2 .

Le lecteur aura remarqué que c'est la condition a) de ce théorème qui assure que S_n est la somme des quantités aléatoires "petites."

3. La loi normale et la théorie de l'information. — Depuis les travaux d'avant-garde de C.E. Shannon et de N. Wiener la théorie de l'information et la notion d'*entropie* qui en est un des outils principaux a connu une large diffusion. La loi normale n'a pas manqué, une fois de plus, de se tailler une place de choix dans cette nouvelle théorie. Nous allons en effet voir que la notion d'entropie permet de distinguer la loi normale parmi d'autres lois au moyen d'une propriété d'extremum. Commençons par rappeler brièvement la définition de l'*entropie de Shannon* :

Soit X une variable aléatoire admettant une densité de probabilité f de support \mathbb{R} et supposons en outre X centrée (de moyenne nulle) et réduite (d'écart-type unité). Alors la quantité

$$H(f) = - \int_{-\infty}^{+\infty} f \text{Log } f \, dx$$

est appelée l'entropie de (Shannon) de la densité f (ou de la variable aléatoire X).

La propriété d'extremum à laquelle nous avons fait allusion s'énonce comme suit.

THÉORÈME 3.1. — Parmi toutes les densités de probabilité centrées, réduites et de support \mathbb{R} , il y en a une et une seule qui maximise l'entropie et c'est la densité de la loi normale $\mathcal{N}(0, 1)$.

La démonstration de ce théorème ne fait appel qu'à des notions élémentaires de convexité et mérite d'être reproduite.

a) Commençons par établir l'inégalité de convexité élémentaire

$$y \operatorname{Log} x - y \operatorname{Log} y \leq x - y \quad (3.1)$$

valable pour tout $x > 0$ et tout $y > 0$, l'égalité ayant lieu si et seulement si $x = y$.

En effet, la fonction Log étant concave, on a l'inégalité $\operatorname{Log} x \leq x - 1$, avec égalité si et seulement si $x = 1$. En remplaçant dans cette inégalité x par $\frac{x}{y}$ où $x, y > 0$, il vient $\operatorname{Log} \frac{x}{y} \leq \frac{x}{y} - 1$, avec égalité si et seulement si $\frac{x}{y} = 1$, c'est-à-dire, $x = y$. Ceci établit (3.1).

b) Désignons par g la densité de la loi normale $\mathcal{N}(0, 1)$, c'est-à-dire, $g(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ pour tout x réel. Elle est centrée et réduite. On a, en vertu de (2.1), pour toute densité f centrée, réduite et de support \mathbb{R}

$$f \operatorname{Log} g \leq f \operatorname{Log} f + g - f, \quad (3.2)$$

avec égalité si et seulement si $f = g$.

c) Une simple vérification montre que

$$\int_{-\infty}^{+\infty} f \operatorname{Log} g \, dx = \int_{-\infty}^{\infty} g \operatorname{Log} g \, dx.$$

d) En intégrant (3.2) et en utilisant c) il vient

$$\int_{-\infty}^{+\infty} g \operatorname{Log} g \, dx = \int_{-\infty}^{\infty} f \operatorname{Log} f \, dx,$$

c'est-à-dire, $H(f) \leq H(g)$, avec égalité si et seulement si $f = g$. Ceci achève la démonstration.

Une autre notion d'information, moins connue que celle de Shannon, mais très utile en théorie de l'estimation, est la *quantité d'information de Fisher*. Elle permet également de distinguer la loi normale parmi d'autres lois au moyen d'une propriété d'extremum. Rappelons sa définition.

Soit X une variable aléatoire admettant une densité de probabilité suffisamment régulière f et supposons en outre X centrée et réduite. Alors la quantité (positive)

$$I(f) = \int_{-\infty}^{+\infty} \left(\frac{f'}{f}\right)^2 f dx$$

est appelée la *quantité d'information de Fisher* de la densité f (ou de la variable aléatoire X).

THÉORÈME 3.2. — *Pour toute densité de probabilité f centrée, réduite et suffisamment régulière, on a $I(f) - 1 \geq 0$, avec égalité si et seulement si f est la densité de la loi normale $\mathcal{N}(0, 1)$.*

La démonstration de ce théorème est encore élémentaire.

Posons $f_1(x) = x\sqrt{f(x)}$, $f_2(x) = \frac{f'(x)}{f(x)}\sqrt{f(x)}$. Alors $\int_{-\infty}^{+\infty} f_1(x)^2 dx = 1$, $\int_{-\infty}^{+\infty} f_2^2 dx = I(f)$ et $\int_{-\infty}^{+\infty} f_1 f_2 dx = \int_{-\infty}^{+\infty} x f'(x) dx = -1$ (grâce à une intégration par parties). Il en résulte que, pour tout λ réel, on a $\int (\lambda f_1 + f_2)^2 dx = (\lambda - 1)^2 + I(f) - 1$, d'où, pour $\lambda = 1$, $I(f) - 1 = \int (f_1 + f_2)^2 dx$. On en déduit que $I(f) - 1 \geq 0$ et que l'on a l'égalité si et seulement si $f_1 + f_2 = 0$, c'est-à-dire, $x + \frac{f'}{f} = 0$, soit, compte tenu du fait que f est une densité de probabilité, $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$. Ceci achève la démonstration.

Les deux théorèmes que nous venons d'énoncer et qui présentent la loi normale comme un objet mathématique extrémal, ont certes un intérêt propre. Il convient cependant de signaler qu'il ont été largement mis à profit par Y.L. Linnik dans ses travaux sur la connexion entre la notion d'entropie et le théorème central limite. Dans un travail très dense [9], il a fourni de ce dernier une démonstration toute entière reposant sur la théorie de l'information. Il a en outre établi un lien très étroit entre la *convergence d'une suite de lois de probabilité vers la loi normale* et la *croissance de l'entropie des termes de cette suite*.

Il ne peut être question ici d'entrer dans les détails techniques du travail de Linnik qui reste, sans doute, une des contributions majeures à la connaissance de la nature profonde de la loi normale.

4. Une propriété de conservation de la loi normale. — Nous nous proposons, dans ce paragraphe, de commenter brièvement une propriété de la loi normale connue sous le nom de *propriété de conservation*. Elle affirme que la somme de deux variables aléatoires normales et *indépendantes* est

encore une variable aléatoire normale. En d'autres termes, le caractère de normalité de deux variables aléatoires se transmet à la somme de ces variables, *si ces dernières sont indépendantes*, ou encore, le *caractère de normalité est conservé par addition de variables aléatoires indépendantes*. L'hypothèse de l'indépendance des termes de la somme est essentielle, car si elle n'est pas vérifiée, il peut arriver que la somme de deux variables normales ne soit pas normale.

Cette propriété de conservation est bien connue et facile à démontrer, malheureusement elle n'est pas caractéristique de la loi normale. D'autres lois, par exemple, la loi de Cauchy la possèdent également. En revanche, il est une propriété de cohérence, bien plus profonde que la propriété de conservation dont nous venons de parler, et qui, elle, est propre à la loi normale. Elle fait l'objet du théorème suivant, connu sous le nom de théorème de Cramer-Lévy.

THÉORÈME 4.1. — *Si une variable aléatoire normale peut être décomposée en la somme de deux variables aléatoires indépendantes, alors chacun des termes de la somme suit une loi normale.*

Ce théorème, conjecturé par Paul Lévy en 1935 et démontré en 1936 par Cramer, est un des bijoux du Calcul des Probabilités. Son énoncé d'une clarté et d'une simplicité rares, affirme une étonnante cohérence de la loi normale. Malheureusement on n'en connaît pas de démonstration élémentaire. Celles qu'on rencontre dans la littérature font toutes appel à un profond théorème d'analyse (le théorème de factorisation des fonctions entières de Hadamard) et c'est à juste titre que Paul Lévy a qualifié cet état de choses de "scandaleux."

Ce théorème fournit, en outre, un argument sérieux qui justifie le remplacement de la première loi de Laplace (de densité $f(x) = \frac{1}{2}e^{-|x|}$ pour x réel) par la seconde (la loi normale) dans la théorie des erreurs. En effet, la première loi de Laplace ne jouit pas d'une propriété de cohérence analogue à celle que possède la loi normale. Pour le voir, il suffit d'observer qu'une variable aléatoire X qui admet pour loi la première loi de Laplace peut être représentée comme la somme de deux variables aléatoires *indépendantes* X_1, X_2 , où X_1 suit la loi exponentielle de paramètre 1 et X_2 la loi de $(-X_1)$. (On dit que X est la "symétrisée" de X_1 .)

5. La loi normale dans l'espace. — La loi normale s'est également imposée dans l'espace, principalement à la suite des travaux de Maxwell sur la répartition des vitesses des particules dans un gaz parfait. Dans ses tentatives pour modéliser le mouvement désordonné observé dans un tel gaz, Maxwell a été amené à représenter la vitesse d'une particule par un vecteur aléatoire $X = (X_1, X_2, X_3)$, dont les composantes X_1, X_2, X_3 sont des variables aléatoires indépendantes admettant toutes la même loi

normale $\mathcal{N}(0, 1)$. La loi du vecteur X est encore appelée *loi normale (à trois dimensions)*. La longueur $|X| = \sqrt{X_1^2 + X_2^2 + X_3^2}$, c'est-à-dire, le module de la vitesse d'une particule, est une variable aléatoire à valeurs positives dont la loi est connue sous le nom de *loi de Maxwell*; elle admet pour densité $f(x) = \sqrt{\frac{2}{\pi}} x^2 e^{-x^2/2}$ ($x \geq 0$).

La projection du vecteur X sur le plan de coordonnées $0x_1x_2$ est un vecteur dont la longueur $\sqrt{X_1^2 + X_2^2}$ est une variable aléatoire à valeurs positives dont la loi est appelée *loi de Rayleigh*; elle admet pour densité $f(x) = xe^{-x^2/2}$ ($x \geq 0$). Il n'est pas besoin de souligner que les deux lois dont nous venons de faire état et qui sont dans la filiation de la loi normale jouent un rôle essentiel dans la théorie statistique des gaz parfaits.

Revenons à la loi du vecteur aléatoire $X = (X_1, X_2, X_3)$. Son principal intérêt aux yeux des physiciens est qu'elle est *isotrope*, c'est-à-dire, invariante par rapport à toute rotation autour de l'origine. Cette dernière propriété s'est finalement avérée caractéristique des vecteurs aléatoires à composantes normales centrées indépendantes identiquement distribuées. La propriété est en outre vraie quelle que soit la dimension de l'espace sous-jacent, comme énoncée dans le théorème suivant.

THÉORÈME 5.1. — *Soit $X = (X_1, \dots, X_n)$ un vecteur aléatoire dans \mathbb{R}^n ($n \geq 2$), dont les composantes X_1, \dots, X_n sont des variables aléatoires indépendantes. Alors les deux propriétés suivantes sont équivalentes :*

a) *les composantes X_1, \dots, X_n suivent toutes une même loi normale centrée ;*

b) *la loi de X est isotrope, c'est-à-dire, invariante par rapport à toute rotation autour de l'origine.*

L'implication a) \Rightarrow b) est banale, c'est sa réciproque b) \Rightarrow a) qui fait problème. On en trouvera une démonstration dans [3]. Dans le cas $n = 2$, elle peut se démontrer élémentairement si l'on suppose en outre que le vecteur aléatoire $X = (X_1, X_2)$ admet une densité $f(x_1, x_2)$ dérivable.

En effet, si b) est vérifiée, la densité $f(x_1, x_2)$ au point (x_1, x_2) ne dépend que de la distance de ce point à l'origine; on a donc $f(x_1, x_2) = f(x_1)f(x_2) = \varphi(x_1^2 + x_2^2)$, d'où, en dérivant par rapport à x_1 et à x_2 ,

$$f'(x_1)f(x_2) = 2x_1 \varphi'(x_1^2 + x_2^2), \quad f(x_1)f'(x_2) = 2x_2 \varphi'(x_1^2 + x_2^2),$$

d'où

$$\frac{1}{x_1} \frac{f'(x_1)}{f(x_1)} = \frac{1}{x_2} \frac{f'(x_2)}{f(x_2)} = c,$$

c'est-à-dire, $f'(x) - cx f(x) = 0$, d'où $f(x) = ke^{cx^2/2}$.

Toujours dans le cas $n = 2$, signalons une autre propriété curieuse de la loi normale, elle aussi liée à une rotation autour de l'origine. Elle affirme

PLAIDOYER POUR LA LOI NORMALE

que si X_1, X_2 sont deux variables aléatoires *indépendantes* de même loi et si $X_1 - X_2$ et $X_1 + X_2$ sont également *indépendantes*, alors la loi commune de X_1, X_2 est normale. Cette propriété, qui est associée aux noms de Bernstein [1] et Darmois [2] a été considérablement généralisée. Nous n'y insisterons pas puisqu'elle intéresse surtout les mathématiciens.

En revanche, nous voudrions, pour clore ce paragraphe, retenir l'attention du lecteur sur une remarquable connexion entre le volume de la boule dans \mathbb{R}^n et la loi normale. Voici de quoi il s'agit.

Désignons par $V_n(R)$ le volume de la boule de l'espace euclidien \mathbb{R}^n , à n dimensions, centrée à l'origine et de rayon R . Ses premières valeurs sont données dans le tableau suivant.

n	1	2	3
$V_n(R)$	$2R$	πR^2	$\frac{4}{3}\pi R^3$

Projetons la masse constituant ce volume sur l'un des axes de coordonnées que nous prendrons comme axe de x . On obtient ainsi une distribution de masses sur la droite qui admet une densité notée $g_n(x, R)$. Pour x en dehors de l'intervalle $[-R, +R]$, cette densité est évidemment nulle et pour x à l'intérieur de cet intervalle, un calcul aisé en fournit la valeur. Nous la consignons dans le tableau suivant.

n	1	2	3
$g_n(x, R)$	1	$2\sqrt{R^2 - x^2}$	$\pi\sqrt{R^2 - x^2}$

On norme cette densité en la divisant par $V_n(R)$ et l'on obtient ainsi une *densité de probabilité* sur la droite que nous noterons $f_n(x, R) = \frac{g_n(x, R)}{V_n(R)}$. Elle est encore nulle pour x en dehors de l'intervalle $[-R, +R]$. Pour x à l'intérieur de cet intervalle, sa valeur pour $n = 1, 2, 3$ est donnée dans le tableau suivant

n	1	2	3
$f_n(x, R)$	$\frac{1}{2R}$	$\frac{2}{\pi} \frac{1}{R} \sqrt{1 - \frac{x^2}{R^2}}$	$\frac{3}{4} \frac{1}{R} \left(1 - \frac{x^2}{R^2}\right)$

Finalement nous prenons dans $f_n(x, R)$ le rayon R égal à \sqrt{n} et nous obtenons en définitive une suite de densités de probabilité sur la droite, de terme général $f_n(x, \sqrt{n})$. Le comportement asymptotique de cette suite fait apparaître un phénomène tout à fait surprenant qui fait l'objet du théorème suivant.

THÉORÈME 5.2. — Pour tout x réel, la suite de terme général $f_n(x, \sqrt{n})$ converge, lorsque n tend vers l'infini, vers la densité de la loi normale $\mathcal{N}(0, 1)$. Autrement dit, pour tout x réel, on a $f_n(x, \sqrt{n}) \rightarrow \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$, lorsque n tend vers l'infini.

Ce phénomène, que le calcul constate, n'a encore reçu, à notre connaissance, aucune explication qui satisfasse l'esprit. Connaître sa démonstration n'apporterait rien au lecteur. Qu'il lui suffise, comme à nous-même, de s'étonner de l'apparition de la loi normale dans un domaine qui, à première vue, lui est totalement étranger.

6. Conclusion. — Nous n'avons pas la prétention d'avoir fait un inventaire exhaustif des remarquables propriétés de la loi normale. Pour revenir à la statistique, signalons encore que dans la théorie des tests d'hypothèse pour petits échantillons, il convient de supposer que la variable parente dont est issu l'échantillon suit une loi normale, car *alors la moyenne et la variance expérimentale sont des variables aléatoires indépendantes*; en d'autres termes, si (X_1, \dots, X_n) est un tel échantillon, alors la moyenne $\bar{X} = \frac{1}{n} \sum_{k=1}^n X_k$ et la variance expérimentale $S_n^2 = \frac{1}{n} \sum_{k=1}^n (X_k - \bar{X})^2$ sont des variables aléatoires *indépendantes*. Cette dernière propriété s'est d'ailleurs avérée *caractéristique de la loi normale*, comme l'a montré Geary [5] et ceci fournit une raison supplémentaire pour opter pour la loi normale comme loi parente.

Une autre propriété de la loi normale centrée réduite consiste en le fait que sa densité est un vecteur propre de la transformation de Fourier; on montre, en effet, que la transformée de Fourier de $(1/\sqrt{2\pi})e^{-x^2/2}$ est $e^{-t^2/2}$.

Nous voudrions enfin terminer en mentionnant que la loi normale s'est avérée un outil puissant dans les parties de la mathématique où on l'attendait le moins : en théorie des nombres (distribution des diviseurs d'entiers), en algèbre (distribution des racines d'une équation algébrique dont les coefficients sont des variables aléatoires normales, etc.).

Bigliographie

- [1] Bernstein (S.N.). — On a property characterizing the Gaussian law, *Tr. Leningrad Politechn. Inst.*, t. **3**, 1954.
- [2] Darmois (G.). — Sur une propriété caractéristique de la loi de probabilité de Laplace, *Compte Rendus Acad. Sc. Paris*, t. **232**, 1951.

- [3] Fuchs (A.) et Letta (G.). — Sur certaines équations fonctionnelles liées à la loi normale, *L'Enseignement Mathématique*, 2^{ième} série, t. **25**, 1979.
- [4] Gauss (G.F.). — *Theoria motus corporum coelestium*, Liber II, Section III, 1809.
- [5] Geary (R.C.). — Distribution of Student's ratio for non-normal samples, *J. Roy. Stat. Soc.*, Suppl. ser. B 3, 1936.
- [6] Gnedenko (V.) et Kolmogorov (A.N.). — *Limit distribution for sums of independent random variables* (English translation by K.L. Chung). — Addison-Wesley, Cambridge, Mass., 1954.
- [7] Laplace (P.-A.). — *Théorie analytique des probabilités*. — Paris, 1812.
- [8] Lindeberg (J.N.). — Eine neue Herleitung des Exponentialgesetzes der Wahrscheinlichkeitsrechnung, *Math. Zeitschrift*, t. **15**, 1922.
- [9] Linnik (Y.V.). — An information-theoretic proof of the central limit theorem with the Lindeberg condition, *Theory of Probability and its Applications*, t. **4**, 1959.
- [10] de Moivre (A.). — *The doctrine of chances or a method of calculating the probability of events in play*. — London, Millar, 1756; réimpression par Chelsea, New York, 1967.
- [11] Polyà (G.). — Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung und das Momentenproblem, *Math. Zeitschrift*, t. **8**, 1920.

Quelques notions du Calcul des probabilités

Si l'on prélève au hasard un élève dans une classe donnée et que l'on mesure sa taille, on obtient un nombre qui dépend de l'issue de ce choix; on dit que c'est une *variable aléatoire* et on a l'habitude de la désigner par une capitale X, Y, \dots . Une *réalisation* particulière d'une telle variable pour une issue donnée s'écrit par une lettre minuscule : x, y, \dots .

Une caractéristique importante d'une variable aléatoire X est sa *loi*, c'est-à-dire, la donnée, pour tout intervalle $[a, b]$, $a < b$, de la probabilité $P\{X \in [a, b]\}$ pour que X prenne une valeur dans cet intervalle lors du prélèvement au hasard. La loi est parfaitement déterminée par la fonction $X \mapsto F(x) = P\{X \leq x\}$, appelée *fonction de répartition* de X .

S'il existe une fonction f telle que $F(x) = \int_{-\infty}^x f(t) dt$, on dit que f est la *densité de probabilité* de X . On voit que $F(x)$ représente l'aire délimitée par le graphe de f , l'axe des x et la droite parallèle à l'axe des y d'abscisse x .

Une variable aléatoire X peut être sommairement décrite par une *caractéristique de position* (par exemple, la moyenne ou espérance mathématique) et une *caractéristique de dispersion* (par exemple, la variance, l'écart-type). Dans le cas où il existe une densité f , on définit :

$$\begin{aligned} \mathbb{E}[X] &= \int_{-\infty}^{+\infty} x f(x) dx && : \text{espérance mathématique de } X; \\ \text{Var } X &= \mathbb{E}[(X - \mathbb{E}[X])^2] && : \text{variance de } X; \\ \sigma(X) &= \sqrt{\text{Var } X} && : \text{écart-type de } X. \end{aligned}$$

Si $\mathbb{E}[X] = 0$, la variable X est dite *centrée*; si $\text{Var } X = 1$, elle est dite *réduite*. Pour des raisons de commodité, il est souvent utile de remplacer une variable aléatoire donnée X par $(X - \mathbb{E}[X])/\sigma(X)$, qui est alors *centrée et réduite*.

Comme exemples de lois à densité, nous citerons les deux lois que Laplace avait proposées dans sa théorie des erreurs. La plus importante, et la seule qui ait été retenue, est la seconde loi de Laplace, connue sous le nom de loi normale. Les raisons pour lesquelles la première loi de Laplace a été abandonnée sont multiples. L'une d'elles est donnée dans le texte.

Exemple 1. — La loi normale centrée réduite, notée $\mathcal{N}(0, 1)$, ou seconde loi de Laplace, a pour densité f et pour fonction de répartition F les fonctions données par :

$$\begin{aligned} f(x) &= \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, && x \in \mathbb{R}; \\ F(x) &= \int_{-\infty}^x f(t) dt. \end{aligned}$$

PLAIDOYER POUR LA LOI NORMALE

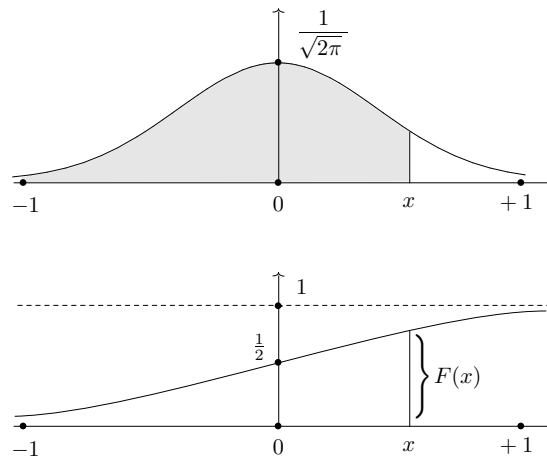


Figure 3. Loi normale

La valeur $F(x)$ de la Figure 1_b est égale à l'aire hachurée dans la Figure 1_a.

Exemple 2. — La première loi de Laplace a pour densité f et pour fonction de répartition F données par

$$f(x) = \frac{1}{2}e^{-|x|}, \quad x \in \mathbb{R};$$

$$F(x) = \int_{-\infty}^x f(t) dt.$$

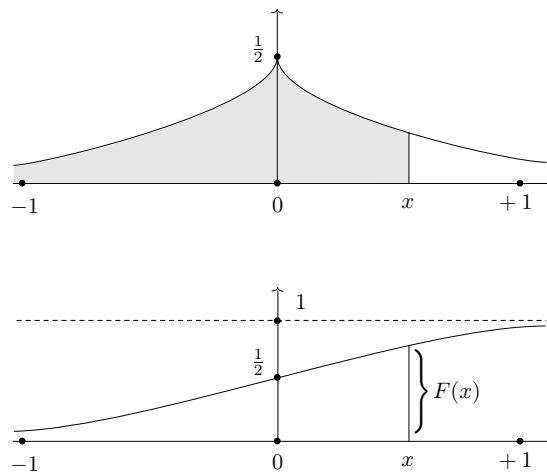


Figure 4. Première loi de Laplace