

La courbe de Gauss : Pas si cloche que ça!

Daniel Justens

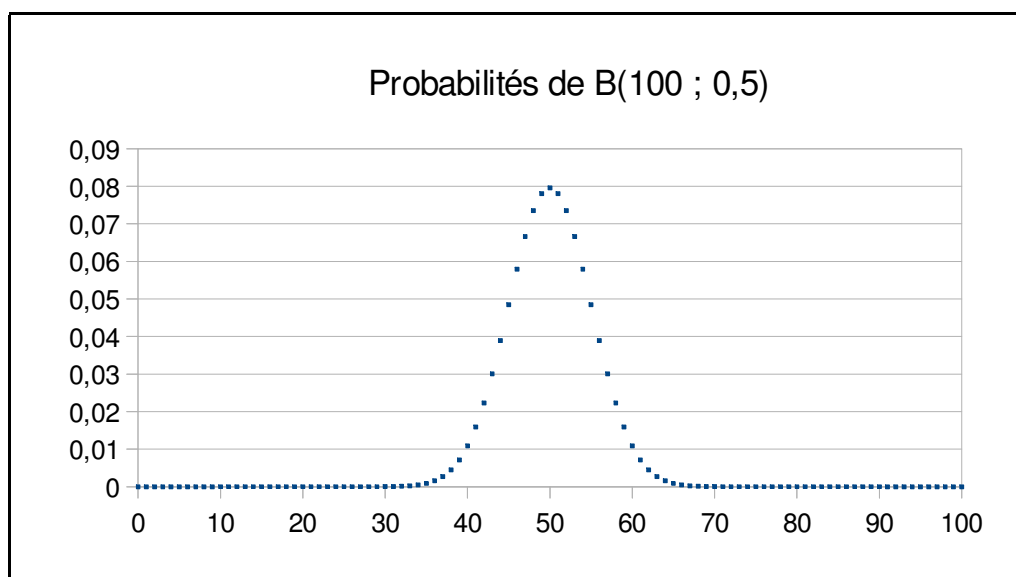
La célèbre courbe en chapeau de gendarme est probablement l'une des plus utilisées en statistique inférentielle et peut-être aussi l'une des plus mal connues. Quelle est son histoire?



Ses origines remonteraient au XVII^e siècle lorsque Jacques Bernoulli s'intéressa au jeu de pile ou face (vers 1677). Il construisit ce que l'on appelle aujourd'hui un « schéma de Bernoulli » donnant naissance à une distribution de probabilité à valeurs dans les naturels, bien connue des lycéens : la distribution binomiale. Il s'agit de réaliser une expérience aléatoire à deux issues possibles, que l'on nomme arbitrairement *réussite* ou *succès* (R ou S) et *échec* (E) et de répéter n fois cette même expérience, dans des conditions identiques et de manière indépendante. Si on peut probabiliser l'événement « réussite », en lui attribuant la probabilité p , on peut également probabiliser la variable X : « nombre de réussites lors de la réalisation de n expériences identiques et indépendantes ». Cette variable se note $B(n,p)$. La probabilité d'observer exactement k réussites sur n est donnée par :

$$p_k = P[X = k] = p^k (1-p)^{n-k} \frac{n!}{(n-k)! k!}$$

La représentation graphique de ces probabilités, vues comme fonction de X , fait apparaître immédiatement un nuage de points à l'allure bien régulière. Voici le cas particulier de 100 jets d'une pièce de monnaie équilibrée avec $p = 0,5$:



Attention : nous avons affaire ici à une fonction à valeurs dans les naturels et décrivant des probabilités. Nous sommes encore loin de la courbe de Gauss, à valeur dans les réels et décrivant une *densité de probabilité*.

Il faut prendre conscience des problèmes liés au calcul de produits de type « exponentielle x factorielle » lorsque le nombre d'expériences devient très grand. Imaginons un sondage effectué entre deux tours de présidentielle auprès de 2000 personnes. Soit un candidat D (de droite) opposé à un candidat G (de gauche - nous laissons au lecteur le soin de définir ce qu'il appelle réussite ou échec lors du vote). Dans une situation équilibrée, les deux valeurs p et $(1-p)$ sont proches de 0,5, ce qui est souvent le cas au second tour. Le produit des deux exponentielles donne environ $(0,5)^{2000}$ soit, en première approximation à 10^{-20} près : 0. Amusant ! Peut-être ! mais totalement inutilisable ! La factorielle de 2000, énorme quant à elle, est également difficile à approcher : les tableurs affichent sans complexe : « ≠ valeur ». En fait le produit intervenant dans le calcul de chaque probabilité a tout de l'indétermination : $\infty \cdot 0$.

C'est dans ce genre de situation que les théorèmes limites trouvent toute leur utilité. Pour y arriver, il fallait trouver une formule approchée de la factorielle. Une première formule fut avancée conjointement par un Français Abraham de Moivre (1667 - 1754) et un écossais James Stirling (1692-1770) qui proposèrent :

$$1 \cdot 2 \cdot 3 \cdot \dots \cdot n = n! \approx \sqrt{2\pi n} \left(\frac{n}{e}\right)^n$$

Formule pleine de mystère introduisant les deux nombres transcendants les plus « connus », π et e . Elle permet de donner une bonne approximation des probabilités intervenant dans la distribution binomiale. Ainsi, la probabilité d'obtenir 50 piles ou 50 faces exactement, lors de 100 jets devient :

$$\left(\frac{1}{2}\right)^{100} \cdot \frac{100!}{50!50!} \approx \left(\frac{1}{2}\right)^{100} \frac{\sqrt{2\pi \cdot 100} \left(\frac{100}{e}\right)^{100}}{\sqrt{2\pi \cdot 50} \left(\frac{50}{e}\right)^{50} \sqrt{2\pi \cdot 50} \left(\frac{50}{e}\right)^{50}} = \frac{1}{\sqrt{50\pi}} = 0,07979$$

C'est également Abraham de Moivre qui a nous fourni la première version d'un théorème limite dans le cas particulier du jeu de pile ou face, en calculant une forme asymptotique des probabilités binomiales. Ce résultat daterait de 1712, mais n'aurait été publié qu'en 1733. Que dit exactement ce théorème de *de Moivre - Laplace* ? Il exprime une convergence en loi de probabilité ou plus simplement *convergence en loi* et fait partie d'un ensemble plus vaste connu sous l'appellation « lois des grands nombres ». La première opération consiste à réduire la variable binomiale. On calcule ses moyenne et écart-type :

$$m = np \quad \text{et} \quad \sigma = \sqrt{np(1-p)}$$

et on travaille avec la variable

$$X_{\text{réduite}} = Z = \frac{X - m}{\sigma} = \frac{X - np}{\sqrt{np(1-p)}}$$

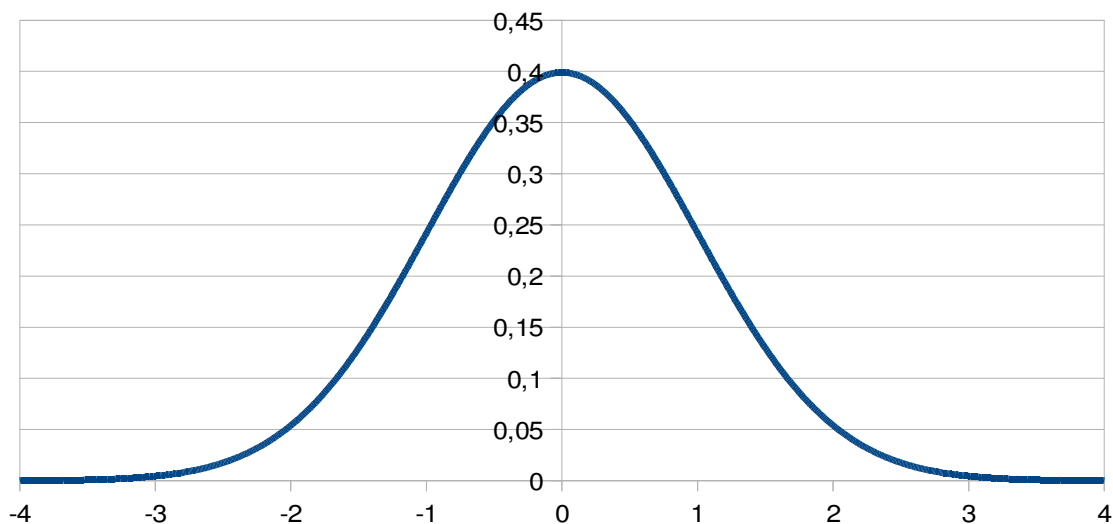
ce qui assure l'extension du domaine, de \mathbb{N} dans \mathbb{R} et le passage des probabilités aux densités de probabilité qui expriment des *probabilités par unité de mesure*. On définit enfin la notion de « fonction de répartition », $F(x)$, égale à la probabilité d'observer une valeur inférieure à x . Le résultat de *de Moivre* établit que la fonction de répartition de cette variable centrée réduite converge,

au sens de la convergence des fonctions, vers la fonction de répartition

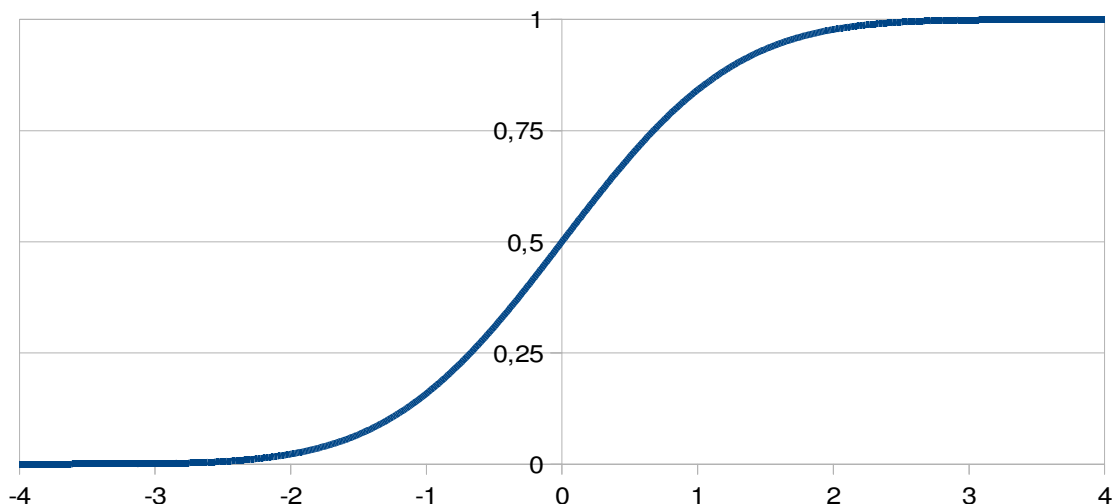
$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{t^2}{2}} dt$$



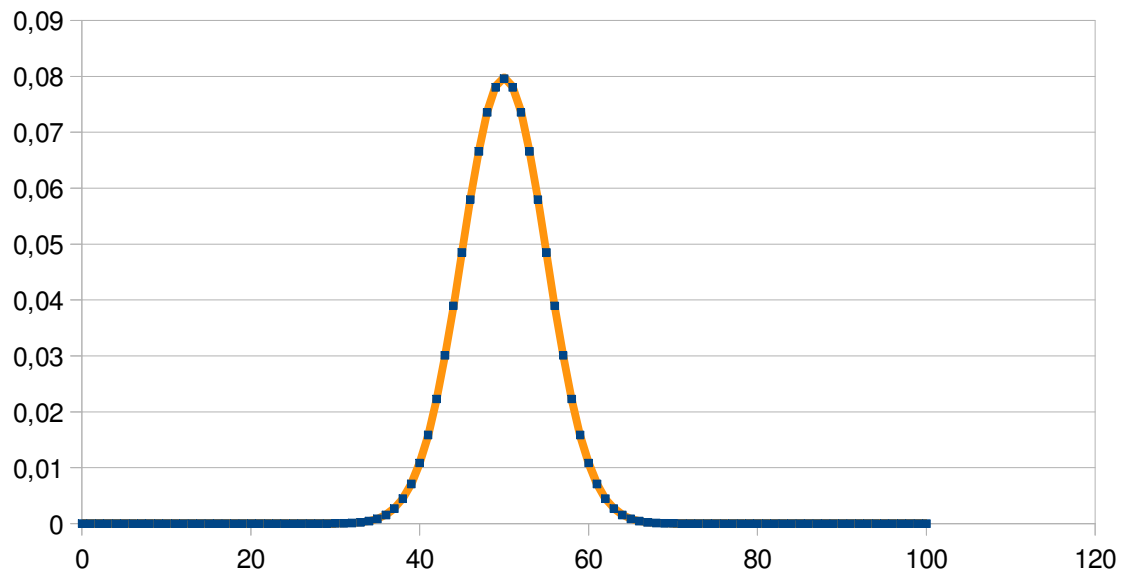
C'est cette fonction de répartition qui définit ce que l'on appelle la « distribution normale centrée réduite ». Remarquons que la fonction à intégrer ne possède pas de primitive élémentaire mais on peut aussi considérer que cette fonction définit la primitive de $f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$. C'est cette dernière fonction $f(z)$ qui a la forme bien connue dite en « chapeau de gendarme » (et non en cloche comme on a coutume de le dire).



Une première table de cette loi fut publiée par Laplace en 1781. Le statisticien australien William Fleetwood Sheppard (1863 - 1936) s'attela à améliorer cette première table entre 1903 et 1905, publiant les valeurs de la fonction densité et celles de sa primitive. Cette dernière représentation de la courbe de Gauss est moins usuelle :



Comparons les probabilités de la binomiale $B(100 ; 0,5)$ avec les densités associées aux intervalles de longueur unité donnés par l'approximation normale établie par de Moivre :



On constate que l'approximation est excellente.

Mais une question reste en suspens : pourquoi cette courbe porte-t-elle le nom de « courbe de Gauss » ? Notons tout d'abord que cette appellation est surtout anglo-saxonne. En France, on parle de « loi de Laplace » ou de « loi de Laplace-Gauss ». Le statisticien belge Adolphe Quetelet (1796 - 1874), l'appelait « courbe des possibilités » ou « loi des possibilités ». Henri Poincaré est probablement à l'origine de l'appellation « distribution normale » parce que *cette loi apparaît souvent dans la nature*. Et Gauss dans tout cela ?



Il est sans doute à l'origine de la méthode dite des moindres sommes de carrés (1809) (LSS : least sum of squares, souvent appelée simplement « moindres carrés » comme si le seul agrégat de plusieurs nombres était leur somme!), bien que cette attribution de cette découverte lui soit contestée, Legendre ayant publié ces résultats avant lui (1805).

Considérons les faits : la découverte le 1^{er} janvier 1801 d'un nouvel astéroïde par l'astronome italien Giuseppe Piazzi fut déterminante. Cet astéroïde fut suivi pendant 6 semaines, permettant une accumulation de mesures. Plusieurs modèles calibrés furent proposées alors ayant pour objet de prédire sa position future. La plupart des prédictions furent erronées. Seul l'astronome Allemand Franz Xaver von Zach (1754 - 1832) fut à même de localiser l'astéroïde un an plus tard, son estimation étant ajustée sur base de la méthode de Gauss.

L'idée de Gauss est simple : dès que l'on dispose d'un modèle pour un phénomène, à toute série d'observations de ce phénomène on peut associer une statistique d'erreurs ou de résidus, définis par la différence entre les observations et les valeurs

fournies par le modèle. Cette méthode permet de calibrer facilement un modèle paramétrique. En effet, Gauss propose de rendre toutes les erreurs positives par mise au carré et de les agréger par sommation. Il est bien conscient que cette méthode est arbitraire mais lui même écrit que toute autre façon de procéder conduit à des calculs d'une extrême complexité. C'est la définition de l'erreur totale sous forme de *somme des carrés des erreurs*. Reste ensuite à calibrer les paramètres du modèle en minimisant cette erreur totale. Un exemple simple ? Comment modéliser une variable au moyen d'une constante ? La méthode de Gauss fournit la réponse : le « meilleur résumé » (celui minimisant la somme des carrés des erreurs) est la *moyenne arithmétique*, aujourd'hui devenue usuelle. Ainsi, une série d'observations x_1, x_2, \dots, x_n sera-t-elle « au mieux » représentée par la constante :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

L'erreur totale minimale répartie sur le nombre d'observations définit la variance du phénomène. L'usage le plus courant de cette méthode est sans doute la mesure de quantités physiques à partir de données expérimentales. En effet, Gauss fait aussi l'hypothèse que la distribution des résidus est normale de moyenne nulle ce qui autorise également le calcul du niveau de précision de l'estimateur « moyenne ». On comprend dès lors les raisons du succès de l'astronome Zach. Et aussi celles qui ont conduit à baptiser la célèbre distribution du nom de celui qui l'a rendue indispensable.

Encadré : Le théorème central limite

L'intitulé « théorème central limite » vient probablement de la traduction de l'expression allemande « zentraler Grenzwertsatz » utilisée par le mathématicien hongrois George Pólya, (1887 - 1985). Il s'agit effectivement d'un résultat *central* en statistique inférentielle et son objet porte bien sur une *limite*. Il n'est pas sûr que le contenu de cette expression soit toujours clair.

En fait, le théorème central limite établit la convergence en loi de toute somme de variables aléatoires indépendantes de variances finies, vers une loi normale. Et cela sans faire d'hypothèses sur les distributions de probabilité des variables sommées. C'est un résultat impressionnant. Précisons-le. En calculant moyenne et variance de la variable « somme », on peut réduire celle-ci. Les hypothèses garantissent le calcul de la variance. Le théorème dit que la fonction de répartition de la variable *somme réduite* converge, lorsque le nombre de termes de la somme tend vers l'infini, vers la fonction de répartition normale centrée réduite que nous avons présentée plus haut.

Intuitivement, ce résultat affirme que tout phénomène aléatoire qui est la résultante d'un grand nombre de causes indépendantes dont les effets se traduisent par de petites modifications (à la hausse ou à la baisse) se distribue selon la même distribution *normale*. Ceci est un gage de régularité. Pour une variable distribuée normalement, la majorité des observations se situent au voisinage de sa moyenne. Les observations très éloignées sont de moins en moins probables. Et la décroissance de cette probabilité suit une fonction densité exponentielle négative tendant très rapidement vers 0. Les phénomènes suivant cette distribution sont donc faciles à gérer et leur comportement généralement très prévisible.

La conclusion à en tirer est sans appel : la régularité d'un phénomène n'est pas un effet de l'intervention d'actes réfléchis. On observe exactement le contraire. Les financiers le savent bien puisque les distributions de taux de rendements qu'ils étudient et qui sont la résultante de l'action d'un grand nombre d'acteurs économiques pensants, ne se distribuent jamais normalement. La forme particulière des distributions de taux, incluant trop de valeurs extrêmes et - ce n'est pas paradoxal - trop de valeurs centrales, porte le nom barbare de leptokurticité. En physique, en finance ou en biologie, tout comme en mathématique, l'ordre est fondé sur le désordre.