MATHÉMATIQUES ET SCIENCES HUMAINES

H. ROUANET
B. LECLERC

Le rôle de la distribution normale en statistique

Mathématiques et sciences humaines, tome 32 (1970), p. 57-74
<http://www.numdam.org/item?id=MSH_1970__32__57_0>

M.S.H. (8e année, no 32, 1970, p. 57-74)

LE RÔLE DE LA DISTRIBUTION NORMALE EN STATISTIQUE

par
H. ROUANET * ET B. LECLERC **

I. — INTRODUCTION ***

Le présent article a été rédigé à l'intention des mathématiciens qui, chargés d'un enseignement de *statistique élémentaire* destiné à des étudiants en sciences humaines, cherchent à s'informer sur le bien-fondé des notions qu'ils enseignent, et se sentent parfois mal à l'aise devant des affirmations péremptoires de nature extra-mathématique et qui en réalité, reflètent seulement une tradition plus ou moins contestable.

Si nous avons choisi la distribution normale pour objet de cet article, c'est qu'il n'est sans doute pas de thème plus maltraité par de telles affirmations.

En 1967, on lit encore, dans un ouvrage rédigé à l'intention de « praticiens »: « c'est un fait d'expérience que de nombreux phénomènes concrets sont distribués au moins approximativement, selon une loi normale ».

Mais si d'autre part, nous ouvrons Kendall et Stuart [9] (qui passent auprès des gens éclairés pour des experts en statistique théorique et appliquée), nous lisons:

« La distribution normale a une histoire curieuse... La découverte que les erreurs d'observation doivent, sous certaines hypothèses plausibles, être distribuées normalement, conduisit à la croyance générale qu'elles l'étaient en fait. Cette croyance s'étendit à des distributions comme celle des tailles d'individus... Mais dans la deuxième moitié du 19e siècle, on s'est aperçu que les distributions observées réellement dans la pratique sont rarement du type normal et qu'apparemment il fallait rejeter la distribution normale comme représentation de phénomènes naturels. »

Pourtant, c'est un fait indiscutable que la distribution normale joue un rôle privilégié en statistique même si ce rôle n'est pas toujours celui qu'on croit. C'est sur la *nature de ce rôle* qu'il convient de s'entendre.

Dans cet article, nous tenterons de dégager de manière quelque peu détaillée, le rôle réel de la distribution normale, ce qui nous permettra de démystifier en les ramenant à leur juste niveau — celui d'une controverse assez stérile — certains arguments trop souvent avancés pour « justifier» ou au contraire pour « dénoncer» le caractère privilégié de la distribution normale.

---

* Laboratoire de Psychologie Expérimentale et Comparée de la Sorbonne, associé au C.N.R.S.
** Centre de Mathématique Sociale, EPHE, VIe section.
*** Nous tenons à remercier M. Reuchlin d'avoir bien voulu nous faire part de ses remarques.

Pour préparer la discussion, un rappel préalable de quelques *propriétés mathématiques* essentielles de la distribution normale nous est apparu indispensable.

En effet, le rôle de la distribution normale ne peut être compris qu'à partir de la connaissance d'au moins quelques propriétés essentielles de cette distribution, ne serait-ce que pour bien distinguer, au niveau de la discussion, propriétés mathématiques et arguments extra-mathématiques.

D'autre part, un tel rappel ne sera peut-être pas inutile pour des lecteurs n'ayant qu'une formation rudimentaire en statistique mathématique.

L'article comprendra donc:

— un exposé purement mathématique, divisé en deux parties pour des raisons essentiellement didactiques;

— une discussion sur le rôle de la distribution normale.

II. PROPRIÉTÉS ÉLÉMENTAIRES DE LA DISTRIBUTION NORMALE. [1]

Nous appellerons propriétés élémentaires de la distribution normale des propriétés:

— qui, à notre avis, peuvent être énoncées (et, au moins partiellement, démontrées par des méthodes élémentaires) à des étudiants du premier cycle en sciences humaines;

— dont la démonstration est immédiate au moyen des fonctions caractéristiques (dont par ailleurs on n'est pas obligé de faire usage au niveau de l'enseignement).

1. *Généralités*

La distribution normale est continue (elle a une densité), définie sur R tout entier, dépendant de deux paramètres μ et σ^2. μ est simultanément la moyenne, la médiane et le mode, σ^2 est la variance. Dire qu'une variable aléatoire X est normale (μ, σ^2) équivaut à dire que la variable aléatoire $Z = \dfrac{X - \mu}{\sigma}$ est normale (0, 1).

La fonction de répartition $F(z)$ de Z est tabulée dans tous les ouvrages de statistique. Très souvent, la fonction inverse l'est également et parfois la densité.

La forme analytique de la densité d'une distribution normale centrée est :

$$f(x) = K A^{x^2}$$

où:

$$|A| < 1$$

L'étudiant peut aisément construire une courbe de ce genre. Prenons par exemple: $K = 1$ et $A = 0{,}6$ (proche de $e^{-1/2} = 0{,}607$).

1. Ou « loi » normale.

On a:

x	0	1	2	3	4
$f(x)$	1	0,61	0,13	0,01	0,00

Pour que la courbe représente la densité de la distribution normale (0,1), on doit avoir :

$$K = \frac{1}{\sqrt{2\,\pi}} \qquad \text{et} \qquad A = \frac{1}{\sqrt{e}} = e^{-\frac{1}{2}}$$

$$y = \frac{1}{\sqrt{2\,\pi}}\, e^{-\frac{1}{2}x^2}$$

d'où pour la distribution normale (μ, σ^2) :

$$y = \frac{1}{\sqrt{2\,\pi}}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Les formes précises de la courbe de densité et surtout de la fonction de répartition sont importantes: il faut bien placer les branches infinies et les points d'inflexion de la densité: $\pm\,\sigma$.

La densité a une forme « en cloche », mais d'autres distributions de probabilité que la distribution normale ont des densités en cloche: il est instructif de superposer les courbes de densité (ainsi que les fonctions de répartition) de plusieurs de ces distributions (en particulier, distribution de Student avec 1 ou 2 degrés de liberté) à celle de la distribution normale (0,1).

2. *Le cortège normal*

A partir d'une suite X_1, ..., X_n de variables aléatoires normales (0,1) indépendantes, on définit les variables aléatoires suivantes, dont les fonctions de répartition sont tabulées.

1) χ^2 (centré) à n degrés de liberté (souvent noté χ^2_n):

$$\chi^2_n = X_1^2 + X_2^2 + \ldots + X_n^2$$

On a:

$$\mathrm{E}\,(\chi^2_n) = n \qquad \text{et} \qquad \mathrm{Var}\,(\chi^2_n) = 2\,n.$$

2) t de Student à n degrés de liberté:

$$t = \frac{X\,\sqrt{n}}{\sqrt{\chi^2_n}}$$

où X est normale (0,1) et indépendante de χ^2_n.

Pour $n = 1$, la distribution du t est également appelée distribution de Cauchy; c'est la distribution du rapport de deux variables normales (0,1) indépendantes.

3) $F(n_1, n_2)$ de Fisher-Snedecor :

$$F(n_1, n_2) = \frac{\chi^2_{n_1}/n_1}{\chi^2_{n_2}/n_2}$$

où χ^2_{n1} et χ^2_{n2} sont indépendantes.

3. *Normalité et linéarité: stabilité de la loi normale*

Si les variables aléatoires X et Y sont normales et indépendantes, les variables aléatoires suivantes, sont également normales, $\forall (a, b) \in \mathbf{R}^2$.

$$\begin{aligned} Z &= a X + b \\ T &= X + Y \end{aligned}$$

Un type de distribution $\mathcal{T}$ est une classe d'équivalence pour la relation sur les fonctions de répartition.

$$F_1 \sim F_2 \Leftrightarrow (\exists\ (a, b) \in \mathbf{R}^2)\ (\forall\ x \in \mathbf{R})\ F_1(x) = F_2(a x + b)$$

Ce type est dit stable si:

$$F_1 \in \mathcal{T} \quad \text{et} \quad F_2 \in \mathcal{T} \Rightarrow F_1 \star F_2 \in \mathcal{T}$$

où $\star$ désigne le produit de convolution (si X_1 et X_2 sont indépendantes, si F_1 est la f.r. de X_1 et F_2 de X_2, $F_1 \star F_2$ est celle de la variable aléatoire $X_1 + X_2$).

Les distributions normales constituent donc un type stable de distributions.

4. *Statistiques construites à partir d'un échantillon normal*

Soit $(X_1, \ldots, X_i, \ldots, X_n)$ une suite finie de n variables normales indépendantes équidistribuées, $X_i \sim N(\mu, \sigma^2)$. Une réalisation de cette suite sera appelée échantillon au hasard dans une distribution parente normale d'espérance μ et de variance σ^2. (Par extension de langage, on appellera également échantillon la suite de variables aléatoires elles-mêmes).

Alors:

— Les statistiques $\overline{X} = \frac{1}{n} \Sigma X_i$ (moyenne observée)

et $s^2 = \frac{1}{n} \Sigma (X_i - \overline{X})^2$ (variance observée)

sont indépendantes en probabilité.

— $\overline{X}$ a une distribution normale: $\left(\mu, \frac{\sigma^2}{n}\right)$

— $\frac{n\, s^2}{\sigma^2}$ suit la distribution du χ^2 à $n - 1$ degrés de liberté.

Ceci se démontre simplement, les X_i étant supposés centrés, en effectuant une transformation orthogonale T sur le vecteur:

$$\begin{bmatrix} X_1 \\ \cdot \\ \cdot \\ \cdot \\ X_n \end{bmatrix}$$

telle que:

$$T \begin{bmatrix} X_1 \\ \cdot \\ \cdot \\ \cdot \\ X_n \end{bmatrix} = \begin{bmatrix} Y_1 \\ \cdot \\ \cdot \\ \cdot \\ Y_n \end{bmatrix}$$

où:

$$Y_1 = \overline{X} \sqrt{n}.$$

Les Y_i ont même distribution que les X_i et sont indépendants, et:

$$\Sigma\, X_i^2 = \Sigma\, Y_i^2 = n\, \overline{X}^2 + \sum_{i=2}^{n} Y_i^2 = n\, \overline{X}^2 + \sum_{i=1}^{n} (X_i - \overline{X})^2.$$

— Enfin:

$$t = \frac{\dfrac{\sqrt{n}\,(\overline{X} - \mu)}{\sigma}}{\sqrt{\dfrac{n\, s^2}{(n-1)\,\sigma^2}}} = \frac{\sqrt{n-1}\,(\overline{X} - \mu)}{s}$$

suit la distribution du t de Student à $n - 1$ degrés de liberté, quels que soient μ et σ.

5. *Moyenne réduite centrée d'un échantillon de taille n : théorème de convergence normale*

Théorème

Soit $X_1, \ldots, X_n$ une suite de variables aléatoires indépendantes, équidistribuées d'espérance μ et de variance finie σ^2 (échantillon d'une distribution parente (μ, σ^2)).

Posons:

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i \qquad \text{(moyenne observée)},$$

la variable aléatoire réduite:

$$V_n = \frac{\overline{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}}$$

converge en distribution (ou « en loi »), lorsque $n \to \infty$, vers la variable aléatoire *normale* réduite (0,1). On dira que la variable V_n est *asymptotiquement* distribuée normalement (0,1); d'autre part, on dira que $\overline{X}_n$ est *approximativement* distribuée normalement $\left(\mu, \frac{\sigma^2}{n}\right)$ (nous disons: approximativement, et non: asymptotiquement, car il est clair que la distribution asymptotique de $\overline{X}_n$ est la distribution ponctuelle concentrée au point μ).

Exemple classique: variable aléatoire réduite associée à la somme de variables de Bernoulli équidistribuées (« convergence de la distribution binomiale vers la distribution normale »).

La démonstration de ce théorème se fait généralement par les fonctions caractéristiques. Elle demande alors que la fonction caractéristique de X_i admette un développement de Taylor d'ordre 3. Nous verrons que cette condition n'est pas en réalité nécessaire.

6. *Loi multinormale*

Le vecteur $(X_1, \ldots, X_n)$ est distribué suivant une distribution multinormale si et seulement si toute combinaison linéaire des X_i, $i = 1, \ldots, n$ est une variable aléatoire normale.

Si aucun des X_i n'est une combinaison linéaire d'autres, la densité correspondante est celle de la loi multinormale à n dimensions:

$$f(x_1, \ldots, x_n) = K \exp \left[-\tfrac{1}{2} \sum_{i=1}^{n} \sum_{i=1}^{n} a_{ij} (x_i - \mu_i)(x_j - \mu_j)\right]$$

où K est une constante convenable, et la matrice définie positive:

$$||\, a_{ij} \,|| = ||\, \sigma_{ij} \,||^{-1}, \qquad \sigma_{ij} = \mathrm{E}\,[(X_i - \mu_i)(X_j - \mu_j)].$$

Soit P_{mn} une application linéaire de rang $m \leqslant n$ de R^n sur R^m.

Le vecteur $(Y_1, \ldots, Y_m) = P_{mn} (X_1, \ldots, X_n)$ suit la distribution multinormale à m dimensions.

La distribution d'un sous-vecteur $(X_1, \ldots, X_p)$, $p < n$ conditionné par le sous-vecteur $(X_{p+1}, \ldots, X_q)$, $p < q \leqslant n$ est multinormale à p dimensions.

On a enfin le résultat suivant :

Si $(X_1, \ldots, X_n)$ suit la distribution normale à n dimensions, centrée, il existe une application linéaire Q de R_n sur lui-même telle que si $(Z_1, \ldots, Z_n) = Q\,(X_1, \ldots, Xn)$, les Z_i suivent indépendamment la distribution normale (0,1).

On a alors:

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}\, x_i\, x_j = \sum_{i=1}^{n} z_i^2 .$$

Conséquence: si $(X_1, \ldots, X_n)$ a une distribution normale à n dimensions, la variable aléatoire:

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} (X_i - \mu_i)(X_j - \mu_j)$$

a une distribution du χ^2 à n degrés de liberté.

Pour $n = 2$, la densité de la distribution est:

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left\{ -\frac{1}{2(1-\rho^2)} \left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) + \left(\frac{y-\mu_y}{\sigma_y}\right)^2 \right] \right\}$$

On a: $\rho = 0 \Leftrightarrow X$ et Y sont indépendantes.

La régression de Y en X est linéaire. On a:

$$E^x(Y) = \mu_y + \rho \frac{\sigma_y}{\sigma_x}(x - \mu_x)$$

et de même:

$$E^y(X) = \mu_x + \rho \frac{\sigma_x}{\sigma_y}(y - \mu_y).$$

La distribution multinormale joue un grand rîle dans les processus stochastiques (processus laplaciens et du second ordre. Voir par exemple, Girault [5], ch. 5, Loeve [10], Parzen [13], ch. 3).

III. AUTRES PROPRIÉTÉS

Nous rassemblons ici d'autres propriétés qui sont, soit moins élémentaires, soit moins souvent mentionnées que les propriétés précédentes et qui pourtant éclairent considérablement le rôle de la distribution normale. C'est pourquoi il est utile de les connaître, même si on n'est pas amené à les enseigner à un niveau élémentaire.

1. *Réciproques et généralisations des propriétés énoncées dans la partie II*

Référence bibliographique principale: *Renyi* [16], chap. 6, § 5, où se trouvent tous les résultats cités ci-dessous et quelques autres.

Théorème de Lévy-Cramer (Réciproque de II - 3).

Si X et Y sont des variables aléatoires indépendantes et si $X + Y$ est normalement distribuée, X et Y sont normalement distribuées (ou l'une est normale et l'autre certaine).

(Conséquence: une somme d'un nombre fini de variables aléatoires non toutes normales ou constantes n'est jamais *exactement* normale.)

Théorème de S. Bernstein

Si X et Y sont des variables aléatoires indépendantes de même distribution et de variance finie, et si $X + Y$ et $X - Y$ sont indépendantes, X et Y sont normalement distribuées.

Ce résultat est connu et assez souvent cité. En réalité, on peut démontrer que la condition « et de variance finie » n'est pas nécessaire et dans ce domaine, le résultat le plus fort est celui-ci:

Théorème de Darmois et Skitovitch

Si $X_1, \ldots, X_r$ sont des variables aléatoires indépendantes, si $a_1, \ldots, a_r, b_1, \ldots, b_r$ sont des réels non nuls et si les variables aléatoires:

$$Z_1 = \sum_{k=1}^{r} a_k X_k \qquad \text{et} \qquad Z_2 = \sum_{k=1}^{r} b_k X_k$$

sont indépendantes, les variables aléatoires $X_1, \ldots, X_r$ sont normales.

Théorèmes de Lukacs, Kawata, Sakamoto (Réciproques de II - 4).

Si $X_1, \ldots, X_n$ sont des v.a. indépendantes équidistribuées, leur distribution est normale si et seulement si, les variables aléatoires:

$$X = \sum_{i=1}^{n} X_i \qquad \text{et} \qquad Y = \sum_{i=1}^{n} \left(X_i - \frac{X}{n}\right)^2$$

sont indépendantes.

Un résultat plus faible est donné sous le nom de théorème de Geary avec sa démonstration dans Dugué [3], fasc. 2, ch. II, ainsi que le suivant:

Théorème de Daly

Toute statistique indépendante de l'origine est indépendante de la moyenne, si les variables parentes sont normales.

Les propriétés de la distribution normale unidimensionnelle se généralisent à la loi multinormale. Par exemple, pour le théorème de Lukacs, on a :

Si $X_1, \ldots, X_n$ sont des vecteurs aléatoires indépendants équidistribués (échantillon dans une distribution parente multidimensionnelle), et si le vecteur moyenne observée et la matrice des covariances observées sont indépendants, la distribution parente est multinormale.

Kendall et Stuart [9], vol. 1, p. 364, donnent cette caractérisation des échantillons multinormaux:

— Soit $X_1, \ldots, X_i, \ldots, X_n$ un échantillon dans une distribution parente p-dimensionnelle:

$$X_i = (X_i^1, \ldots, X_i^p) \qquad i = 1, \ldots, n \qquad , n > 1$$

— Soient les vecteurs:

$$X^j = (X_1^j, \ldots, X_n^j) \qquad j = 1, \ldots, p.$$

Si la distribution de l'échantillon a une densité qui est fonction uniquement des longueurs des vecteurs x^j et des angles qu'ils font entre eux (ou de leurs cosinus : coefficients de corrélation), la distribution parente est multinormale.

En particulier, pour le cas unidimensionnel, si la densité de l'échantillon $X_1, \ldots, X_n$ est fonction uniquement de $\sum_1^n x_i^2$, la distribution parente est normale.

2. *Stabilité et domaine d'attraction*

Le type de la distribution normale est stable. Il est donc (Girault [5]) indéfiniment divisible: une variable aléatoire de ce type est distribuée comme une somme d'un nombre n arbitraire de variables du même type. La famille des distributions indéfiniment divisibles est elle-même incluse dans la famille des distributions de processus numériques à accroissements aléatoires indépendants. Pour la distribution normale, on a le processus de Wiener-Lévy (Girault [5], Parzen [14]).

Les propriétés des distributions indéfiniment divisibles et des distributions stables se trouvent dans Gnedenko et Kolmogorov [7]. Citons (ch. 7, § 33 et 37) :

— pour que la fonction de répartition $F(x)$ soit celle de limites de sommes de termes indépendants et équidistribués, il est nécessaire et suffisant qu'elle soit stable.

Théorème.

Toute distribution indéfiniment divisible a un domaine d'attraction non vide.

La distribution normale a pour « domaine d'attraction » l'ensemble des distributions telles que, si toutes les variables (X_i) sont indépendantes équidistribuées suivant l'une d'entre elles, il existe deux suites de constantes (A_i) et (B_i) telles que la variable aléatoire:

$$Y_n = \frac{1}{B_n} \sum_{i=1}^{n} X_i - A_n$$

converge en distribution vers une variable normale.

Théorème

La fonction de répartition $F(x)$ appartient au domaine d'attraction de la distribution normale si et seulement si:

$$\lim_{y \to \infty} \frac{y^2 \int_{|x| > y} dF(x)}{\int_{|x| < y} x^2 \, dF(x)} = 0.$$

Par exemple, la distribution de densité:

$$f(x) = \begin{cases} \dfrac{1}{|x|^3} & \text{si} \quad |x| > 1 \\ 0 & \text{si} \quad -1 \leqslant x \leqslant +1 \end{cases}$$

appartient au domaine d'attraction de la distribution normale.

D'autre part, on vérifie immédiatement que les distributions de variance finie appartiennent au domaine d'attraction de la distribution normale.

3. *Théorèmes-limites centraux*

Théorème de Lindeberg et Feller (voir Cramer [2]).

Soit $X_1, \ldots, X_n, \ldots$ une suite de variables aléatoires indépendantes centrées, $\sigma_1, \ldots, \sigma_n, \ldots$ la suite de leurs écarts-types, tous finis, $F_1(x), \ldots, F_n(x), \ldots$ la suite de leurs fonctions de répartition.

Soit: $S_n^2 = \sum_{i=1}^{n} \sigma_i^2$, $\mathscr{F}_n(x)$ la fonction de répartition de la variable $\frac{1}{S_n} \sum_{1}^{n} X_i$.

La « condition de Lindeberg »:

$$(\forall \varepsilon > 0) \lim_{n \to \infty} \frac{1}{S_n^2} \sum_{i=1}^{n} \int_{|x| > \varepsilon S_n} x^2 \, dF_i(x) = 0 \tag{1}$$

est nécessaire et suffisante pour que l'on ait à la fois:

$$. \; S_n \xrightarrow[n \to \infty]{} \infty \tag{2}$$

$$. \; \frac{\sigma_n}{S_n} \xrightarrow[n \to \infty]{} 0 \tag{3}$$

[1]

$$. \; \lim_{n \to \infty} \mathscr{F}_n(x) = \Phi(x) \tag{4}$$

où $\Phi(x)$ est la fonction de répartition de la loi normale réduite centrée.

Dans le cas où $\lim_{n \to \infty} S_n = S$ fini, les variables X_1, et $\sum_{i=2}^{\infty} X_i$ sont indépendantes et de variance finie.

D'après le théorème de Lévy-Cramer, la CNS pour que leur somme soit normale est qu'elles soient toutes deux normales. Reprenant ceci pour les suites (X_i), $i > 1$, $i > 2$, etc., on en déduit que

---

1. Les conditions (2) et (3) sont équivalentes à la suivante (Rao, Loève) :

$$\lim_{n \to \infty} \max_{i \leqslant n} \frac{\sigma_i}{S_n} = 0.$$

On peut aussi comparer avec ces énoncés celui donné par Renyi, chap. 8.

la CNS pour que la somme de la série (X_i) soit exactement normale est que chacune des variables soit normale.

Dans les cas où $\frac{\sigma_n}{S_n} > \alpha > 0$ pour une infinité de valeurs de n, on conçoit, sans pousser la discussion, que la variable $\frac{1}{S_n} \sum_{i=1}^{n} X_i$ ne peut converger en distribution vers une variable normale que si la suite $\left(\frac{X_n}{S_n}\right)$ qui ne tend pas vers 0, tend elle-même vers une variable aléatoire normale. B. V. Gnedenko, dans [6], montre que:

$$(\forall\, \varepsilon > 0) \quad P\left[\max_{1 \leqslant i \leqslant n} |X_i| < S_n\right] \leqslant \frac{1}{\varepsilon^2 S_n^2} \sum_{i=1}^{n} \int_{|x| > \varepsilon S_n} x^2\, dF_i(x).$$

La réalisation de la condition de Lindeberg entraîne que le second membre tend vers 0 pour $n \to \infty$, donc que les variables $\frac{X_i}{S_n}$ tendent uniformément vers 0 en probabilité.

A partir du théorème de Lindeberg, on peut montrer aisément un certain nombre de théorèmes de convergence classiques:

1) Soit $X_1, X_2, \ldots, X_n \ldots$ une suite de variables aléatoires indépendantes, centrées, *uniformément bornées*.

La condition nécessaire et suffisante pour que la variable $\frac{1}{S_n} \sum_{i=1}^{n} X_i$ tende vers la loi normale est que $\sum_{i=1}^{n} \sigma_i^2 \xrightarrow[n \to \infty]{} \infty$.

En particulier, une somme réduite de variables de Bernoulli indépendantes, de paramètre p_i différents, tend vers la loi normale si et *seulement si* (Cramer [1], p. 218).

$$\sum_{i=1}^{n} p_i (1 - p_i) \xrightarrow[n \to \infty]{} \infty$$

(Cas particulier: p_i égaux: « convergence de la distribution binomiale vers la normale ».)

2) Si les variables de la suite (X_i) indépendantes, sont équidistribuées, de variance finie σ^2, la condition de Lindeberg devient:

$$\lim_{n \to \infty} \frac{1}{\sigma^2} \int_{|x| > \varepsilon \sqrt{n}\sigma} x^2\, dF = 0$$

et est toujours remplie. Le théorème II-5 se démontre ainsi sans faire appel à une autre condition que l'existence du second moment. On retrouve: toute distribution de variance finie appartient au domaine d'attraction de la distribution normale.

Théorème de Liapounov (Renyi [16], ch. 8, § 3).

Reprenant les notations du théorème de Lindeberg-Feller, la « condition de Liapounov »:

$$\exists\ \delta > 0 \quad \text{tel que} \ \frac{1}{S_n^{2+\delta}} \sum_1^n E(|X_i|^{2+\delta}) \xrightarrow[n \to \infty]{} 0 \tag{5}$$

entraîne la proposition (4):

$$\lim_{n \longrightarrow \infty} \mathscr{F}_n(x) = \Phi(x).$$

(Ce théorème est souvent présenté avec $\delta = 1$).

En fait, on a aussi (5) $\Rightarrow$ (2) (évident) et (5) $\Rightarrow$ (3) (car on a toujours:

$$[\mathrm{E}(|X|^{2+\delta})]^{\frac{1}{2+\delta}} > \sigma_X).$$

La réalisation de la condition de Liapounov entraine donc toujours celle de la condition de Lindeberg. (Gnedenko et Kolmogorov [7], p. 5 et 103.)

N.B. — Les théorèmes-limites centraux ont été généralisés au cas multi-dimensionnel (voir, par exemple, Rao [15]).

Ces théorèmes ont trait à la fonction de répartition (convergence en distribution ou en loi). Pour la densité, il existe des théorèmes « locaux » (Gnedenko [6], ch. VIII, § 43).

Application (Cramer [1], p. 218).

Soit une fonction $g(X_1, X_2, \ldots, X_n)$ de n variables aléatoires indépendantes, avec n grand. Si la fonction g a des dérivées continues du premier et du second ordre au voisinage du point $\mu = (\mu_1, \mu_2, \ldots, \mu_n)$, où μ_i désigne l'espérance de X_i, on peut écrire un développement de Taylor:

$$g(X_1, X_2, \ldots, X_n) = g(\mu_1, \mu_2, \ldots, \mu_n) + \sum_i^n c_i(X_i - \mu_i) + R$$

où c_i est la valeur de $\dfrac{\delta g}{\delta X_i}$ au point μ, le reste R contenant les dérivées du second ordre. Le premier terme du membre de droite est une constante, alors que le deuxième est une somme de n variables indépendantes centrées. D'après le théorème-limite central, la somme des deux premiers termes, sous des conditions générales, est approximativement normale, avec une espérance égale au premier terme.

Dans de nombreux cas, il est possible de montrer que lorsque $n \longrightarrow \infty$, la présence du terme R n'a pas d'influence sur la distribution de sorte que la fonction g a, pour de grandes valeurs de n, une distribution approximativement normale.

4. *Quantiles observés et distribution conjointe des moments* (voir Cramer [1], ch. 28)

Soit (X_i) une suite de n variables aléatoires indépendantes équidistribuées (échantillon) de fonction de répartition $F(x)$.

Soit $p \in]\,0, 1\,[$ et ξ tel que $F(\xi) = p$.

— Soit Z_n le quantile observé de l'échantillon (Xi) $1 \leqslant i \leqslant n$ correspondant à la proportion p. Si $F(x)$ est continue et deux fois continument dérivable en ξ, la variable aléatoire Z_n est approximativement normale lorsque n croît, de moyenne ξ et d'écart-type:

$$\frac{1}{f(\xi)}\sqrt{\frac{p(1-p)}{n}},$$

où $f(\xi)$ est la valeur de la dérivée (densité) de $F(x)$ au point ξ.

— Soit m'_{nk} le k-ième moment observé des (X_l), $l \leqslant i \leqslant n$ et m_k le moment théorique correspondant supposé existant.

La distribution conjointe d'un nombre quelconque de variables aléatoires $\sqrt{n}\,(m'_{nk} - m_k)$ tend, lorsque n croît, vers la distribution multinormale centrée, définie par les moments du second ordre:

$$\text{Var}\,[\sqrt{n}\,(m'_{nk} - m_k)] = m_{2k} - m_k^2$$

$$\text{Cov}\,[\sqrt{n}\,(m'_{nk} - m_k),\ \sqrt{n}\,(m'_{nl} - m_l)] = m_{k+l} - m_k\, m_l$$

— Soit μ'_{nk} le k-ième moment observé centré des (X_i), $1 \leqslant i \leqslant n$ et μ_k le moment théorique centré correspondant.

On a:

$$\text{E}\,(\mu'_{nk}) = \mu_k$$

$$\text{Var}\,(\mu'_{nk}) = \frac{1}{n}\,(\mu_{2k} - 2\,k\mu_{k+1}\,\mu_{k+1} - \mu_k^2 + k^2\,\mu_2\,\mu_{k-1}^2)$$

μ'_{nk} est, pour n croissant, approximativement normale.

La normalité asymptotique des quantiles empiriques se montre par le calcul. Celle des moments est une conséquence des théorèmes-limites centraux (linéarité).

5. *Tendance vers la normalité des estimateurs du maximum de vraisemblance* (Dugué [3], ch. VII, fasc. 1).

Soit $f(x, \Theta)$ la densité de la distribution de l'échantillon X de taille n, où $\Theta = (\theta_1, \ldots, \theta_p)$ est le vecteur des paramètres de la distribution parente.

Soit $T_n = (t_n^1, \ldots, t_n^p)$ l'estimateur du maximum de vraisemblance de Θ obtenu à partir de l'échantillon.

Si T_n tend en probabilité vers Θ lorsque n croît, si la fonction $f(x, \Theta)$ est dérivable deux fois par rapport à θ_i et θ_j, $\forall\ i, j = 1, \ldots, p$, sous le signe d'intégration, si la matrice:

$$\text{I} = \left\| \text{E}\,\frac{1}{f^2(X_i, \Theta)}\,\frac{\partial f}{\partial\theta_i}\,\frac{\partial f}{\partial\theta_j} \right\|_{\substack{i = 1, \ldots, p \\ j = 1, \ldots, p}}$$

existe et si les fonctions:

$$\frac{\partial^2 \operatorname{Log} f(x, y)}{\partial \theta_i \, \partial \theta_j}$$

où: $y = (y_1, \ldots, y_p)$ sont continues en $y_1, \ldots, y_p$ pour $y = \Theta$ uniformément par rapport à x, la distribution de probabilité du vecteur $\sqrt{n}\,(T_n - \Theta)$ tend vers une distribution multinormale à p dimensions de moyenne nulle et de matrice de covariance I^{-1}.

IV. RÔLE DE LA DISTRIBUTION NORMALE EN STATISTIQUE

Avant d'aborder le rôle de la distribution normale, nous proposerons sur le rôle en statistique des distributions en général, une distinction qui nous paraît importante.

Un premier rôle des distributions est de fournir des *modèles*, c'est-à-dire des représentations idéalisées de phénomènes concrets [1].

Un autre rôle des distributions est de fournir des *intermédiaires de calcul*, principalement sous forme de distributions d'échantillonnage de statistiques servant soit à estimer un paramètre, soit à tester une hypothèse.

Par exemple: *la distribution de Bernoulli* (distribution d'une variable à valeurs dans $\{0,1\}$ joue un rôle important comme modèle pour formaliser des phénomènes dont on ne retient qu'une propriété dichotomique.

Mais cette distribution ne joue aucun rôle comme intermédiaire de calcul. Quand on veut *calculer* à partir d'un modèle de Bernoulli, par exemple pour estimer la fréquence d'un événement, on travaille sur une distribution nouvelle, dérivée (à partir de l'opération de convolution) de la distribution de Bernoulli, mais qui n'est plus la distribution de Bernoulli : la distribution binomiale.

Si nous en venons maintenant au rôle de la distribution normale, on s'aperçoit que celle-ci joue un rôle à la fois comme modèle et comme intermédiaire de calcul.

C'est chacun de ces rôles qu'il convient d'examiner, afin d'apprécier à sa juste valeur l'importance de la distribution normale.

1) *Rôle de la distribution normale comme modèle*

Pour qu'une distribution normale représente une approximation raisonnable d'un phénomène, il est nécessaire de supposer que la distribution du phénomène est à peu près unimodale, symétrique et peu étalée aux extrémités. Dans ce cas on fera souvent au moins à titre de première approximation, la supposition d'un modèle normal.

Le choix d'une distribution normale comme modèle peut résulter de considérations plutôt théoriques ou plutôt pratiques:

1. L'objectif essentiel de cet article n'étant pas méthodologique, nous ne tenterons pas d'approfondir la notion de modèle en statistique et de préciser ce qu'on doit entendre par modèle plus ou moins réaliste, plus ou moins artificiel, etc. Disons simplement qu'un modèle est toujours une approximation de la réalité ou plutôt de la représentation que nous nous faisons de cette réalité. A cet égard, il n'y a pas, en statistique, de type de distribution privilégiée : par exemple, une distribution discrète reposant sur l'hypothèse de tirages indépendants peut ne pas être « moins approximative » qu'une distribution continue comme la distribution normale.

Choix de la distribution normale à partir de considérations théoriques

L'importance de la distribution normale comme modèle résultant de considérations théoriques a donné et donne encore lieu à de nombreuses discussions et beaucoup de malentendus. Cette importance tient certainement en grande partie à des raisons d'ordre *historique*. Le terme de « normal » est d'ailleurs le résultat (et non pas, bien entendu, la cause) de ces raisons. Sans entrer dans les détails, on a, dans le passé :

a) constaté, ou admis que la distribution des erreurs aléatoires (non systématiques) de mesure, en physique, est souvent normale ;

b) plus ou moins hâtivement assimilé la distribution de nombreux phénomènes variables à une telle distribution d'erreurs. Au début du siècle, suivant la célèbre boutade, les physiciens croyaient à la distribution normale parce qu'ils pensaient que c'était un théorème mathématique et les mathématiciens parce que c'était un fait expérimental ! Apparemment, dans les sciences dites exactes, on est revenu de ces conceptions simplistes ; on souhaiterait pouvoir en dire autant des sciences humaines.

Il est important pédagogiquement de dénoncer l'exagération manifeste consistant à « justifier » le modèle normal à partir d'un théorème limite central (surtout d'un théorème faisant intervenir des variables indépendantes) :

— Il n'y a aucune raison pour que les conditions que requiert un théorème-limite central doivent être considérées comme des contraintes satisfaites par un phénomène naturel « obéissant aux lois du hasard » ; la constatation qu'une distribution n'est pas normale ne saurait être signe d' « anormalité » devant éveiller les soupçons ;

— Inversement, la constatation qu'une distribution est normale ne saurait être nécessairement interprétée en supposant que la variable correspondante est la somme de variables *indépendantes*, vérifiant les conditions d'un théorème limite central restrictif : le fait qu'un phénomène est distribué normalement n'implique nullement l'intervention d'un mécanisme rappelant celui, réel ou supposé, des erreurs d'observation (la distribution normale peut, par exemple, résulter de l'addition de v.a. non indépendantes).

Choix de la distribution normale à partir de considérations pratiques

En réalité, l'importance de la distribution normale comme modèle tient essentiellement à des raisons de commodité pratique : facilité des calculs analytiques (encore qu'aujourd'hui la facilité de ces calculs soit moins importante que la facilité des calculs numériques), existence de tables détaillées. C'est pourquoi lorsque la distribution normale n'apparaît pas directement à partir de considérations théoriques on cherche souvent à s'y ramener. On peut distinguer à cet égard trois manières courantes de se ramener à une distribution normale :

— En premier lieu, la technique de *normalisation*, utilisée en particulier en psychométrie ; elle consiste, en partant d'une échelle de notes plus ou moins arbitraire à la transformer empiriquement de façon à obtenir une distribution de notes sensiblement normale.[1]

— En second lieu, la technique de *transformation des variables :* on suppose que le phénomène étudié à une distribution telle que par une transformation mathématique appropriée (par exemple, transformation logarithmique) la distribution devienne normale (exemple : la distribution log. normale — distribution d'une variable dont le logarithme est distribué normalement).

1. On doit également mentionner un deuxième type de normalisation consistant, à partir d'une distribution non normale, à modifier les questions du test ou à en introduire de nouvelles de manière à rapprocher la distribution des notes de la normale.

— En troisième lieu, on peut opérer un *codage de la variable* revenant à considérer une variable dérivée ayant une distribution plus proche de la distribution normale que la variable primitive. Considérons par exemple, une expérience de temps de réaction: on recueille, disons 100 TR au cours de 100 épreuves successives. Supposons qu'au lieu des TR primitifs on considère les moyennes (ou les médianes) des TR pris par blocs de 10, de 5 ou même de 2. Moyennant des conditions très générales, ces moyennes (ou médianes) de blocs auront une distribution plus proche de la normale que la distribution des TR primitifs. On pourra donc, pour ces variables dérivées, considérer le modèle normal comme admissible: en réalité, cela revient à utiliser les propriétés des distributions d'échantillonnage de la moyenne ou de la médiane: cf. § III).

Toutes ces techniques donnent au modèle une grande extensibilité, donc une grande importance — mais une importance qui ne repose pas sur le fait qu'il représente adéquatement des phénomènes naturels considérés comme « donnés » (cf. les commentaires de Kendall cités en tête de cet article) : le modèle normal est souvent réaliste non pas grâce à ses vertus « naturelles » mais à la suite d'une « construction ». Dans certains cas, la construction d'un modèle normal peut confiner à l'artificiel. Il ne faut pas à tout prix chercher à se ramener au modèle normal ni hésiter, le cas échéant, à en utiliser un autre plus directement approprié. Ainsi dans l'exemple d'un sondage d'opinion, on sait que les distributions asymétriques ou bimodales se rencontrent fréquemment : vouloir en pareil cas construire un modèle normal serait d'une grande maladresse. A cet égard, on peut considérer qu'avec l'introduction progressive de nombreux modèles autres que le modèle normal (liée à l'utilisation de moyens de calcul numérique rapides) l'*importance de la distribution normale comme modèle apparaît nettement comme allant en diminuant.*

Avant de clore ce paragraphe, mentionnons un avantage du modèle normal : les distributions de nombreuses statistiques peuvent facilement être obtenues sous forme analytique en supposant la distribution parente normale; lorsque la distribution parente n'est pas normale, ces distributions constituent souvent des approximations valables surtout lorsque l'effectif de l'échantillon est élevé. Il en résulte que nombre de méthodes d'inférence statistique valables rigoureusement sous le modèle normal restent valables en première approximation lorsque ce modèle n'est plus acceptable : on dit que le modèle normal est *robuste. Le modèle normal peut être donc utile alors même qu'il n'est pas réaliste* [1].

2) *Rôle de la distribution normale comme intermédiaire de calcul*

Tout d'abord, le rôle de la distribution normale est essentiel en statistique linéaire (par statistique linéaire, nous entendons la statistique fondée sur le calcul de sommes et de moyennes).

Dans le cadre des modèles d'échantillonnage habituels (échantillon considéré comme la réalisation d'une famille de v.a. indépendantes équidistribuées), le théorème-limite central sous ses diverses formes justifie effectivement l'importance de la distribution normale comme intermédiaire de calcul, par le biais des distributions d'échantillonnage.

En effet:

— Ou les v.a. qui interviennent sont elles-mêmes considérées comme distribuées normalement, et alors les sommes et moyennes seront exactement normalement distribuées;

— Ou les v.a. ne sont pas distribuées normalement, mais elles vérifient les conditions d'une forme du théorème-limite central, et dans ce cas leur somme ou leur moyenne sera lorsque n est grand approximativement distribuée normalement.

---

1. Dans certains milieux, l'importance attribuée au modèle normal revêt un caractère hautement fétichiste : le fait qu'une distribution *expérimentale* ne présente pas toutes les marques extérieures de la normalité est invoqué à l'occasion pour refuser le calcul d'un t, d'un écart-type, voire d'une moyenne !

Cela dit:

a) Les conditions de validité des théorèmes de convergence normale sont peu restrictives, mais elles peuvent ne pas être remplies, auquel cas on ne rencontrera pas la distribution normale comme distribution-limite (un contre-exemple classique est fourni par la distribution de Cauchy: une moyenne de variables indépendantes de Cauchy a une distribution de Cauchy; le calcul de sommes et de moyennes ne conduit pas nécessairement à la distribution normale !);

b) Même lorsque les conditions de convergence normale sont remplies, la convergence peut être très lente, auquel cas le théorème-limite est de peu d'utilité pratique (exemple: cas d'une distribution parente fortement asymétrique);

c) Ensuite, des statistiques définies à partir de ces sommes ou moyennes auront des distributions définies à partir de la distribution normale (cortège normal) [1];

d) Enfin, des classes importantes de statistiques: moments observés, fonctions de ces moments, quantiles, estimateurs obtenus par la méthode du maximum de vraisemblance ont sous des conditions très générales quand n est grand, une distribution d'échantillonnage approximativement normale.

C'est ce qui explique que l'on rencontre si souvent la distribution normale et ses dérivés comme intermédiaire de calcul dans de nombreux contextes où la distribution normale ne joue *aucun rôle* comme distribution parente, par exemple:

— Dans des problèmes de fréquence: cf. convergence de la distribution binomiale vers la normale;

— En statistique non-paramétrique, dans des problèmes de rangs (sommes ou moyennes de rangs);

— Dans les problèmes de sondage où les ensembles parents étant finis, les distributions parentes sont à support borné (donc à variance finie), ce qui entraîne l'apparition de la distribution normale à partir du théorème-limite central;

— Également dans des contextes non-linéaires: distribution normale comme distribution-limite de statistiques, telle que la médiane, etc.

En conclusion, on peut dire que l'*importance de la distribution normale comme intermédiaire de calcul à l'inverse de son rôle comme modèle ne cesse d'aller en augmentant.*

Terminologie

Au lieu de distribution normale, on dit aussi distribution de Gauss, ou de Laplace-Gauss, ou même (en France) de Laplace; « distribution normale de Gauss » est un pléonasme.

Il est important de dire aux étudiants que le terme traditionnel de « normal » n'implique aucun jugement de valeur sur la distribution correspondante, pas plus que le terme de « norme de vecteur » n'implique un jugement de valeur sur le vecteur. Une distribution non normale n'est pas une distribution exceptionnelle ou « anormale ». Dire qu' « il est normal de rencontrer la distribution normale » n'est (aujourd'hui) qu'un mauvais calembour. Cela dit, on peut penser que, les malentendus une fois levés, le terme en lui-même ne présente aucun caractère gênant. D'autre part, il présente les avantages d'éviter la contestation entre Laplace et Gauss et de se prêter à la formation de dérivés (normalité, normaliser, etc.).

---

1. Lesquelles, à leur tour, pourront souvent être approchées par une distribution normale : approximations normales du χ^2 ou du t lorsque le nombre de degrés de liberté est élevé.

BIBLIOGRAPHIE

[1] Cramér H., *Mathematical methods of statistics*, Princeton University Press, 1946.

[2] Cramér H., *Random variables and probability distributions*, Cambridge University Press, 1937.

[3] Dugue D., *Traité de statistique théorique et appliquée*, I et II, Masson, 1958.

[4] Feller W., *An introduction to probability theory and its applications*, I (1964) et II (1966), Wiley.

[5] Girault M., *Processus aléatoires*, Dunod, 1965.

[6] Gnedenko B. V., *The theory of probability*, Chelsea Publishing Company, 1963.

[7] Gnedenko B. V. et Kolmogorov A. N., *Limit distributions for sums of independant random variables*, Addison Wesley, 1954.

[8] Hennequin P. L. et Tortrat A., *Théorie des probabilités et quelques applications*, Masson, 1965.

[9] Kendall M. G. et Stuart A., *The advanced theory of statistics*, I : "Distribution theory", (1958) ; II : "Inference and relationship " (1961). III : "Design and analysis, and time-series" (1966).

[10] Lévy P., *Théorie de l'addition des variables aléatoires*, Gauthier-Villars, 1954.

[11] Loeve M., *Probability theory*, Van Nostrand Company, 1963.

[12] Lukacs E., *Fonctions caractéristiques*, Dunod, 1964.

[13] Miller K. S., *Multidimensional gaussian distributions*, Wiley, 1964.

[14] Parzen E., *Stochastic processes*, Holden Day, 1962.

[15] Rao R. C., *Linear statistical inference and its applications*, Wiley, 1965.

[16] Renyi A., *Calcul des probabilités*, Dunod, 1966.

[17] Wilks S. S., *Mathematical statistics*, Wiley, 1962.