

## LOI DES GRANDS NOMBRES ET THÉORÈME DE LIMITE CENTRALE

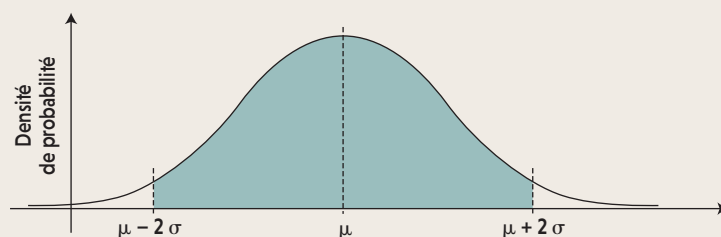
**L**a loi des grands nombres énonce que lorsqu'on échantillonne une variable aléatoire, sous certaines conditions, lorsque la taille de l'échantillon est assez grande, la moyenne empirique de cet échantillon tend vers la moyenne théorique de la variable aléatoire échantillonnée.

C'est sur cette loi que reposent notamment les sondages. Considérons un échantillon d'observations  $x_1, x_2, \dots, x_n$  d'une variable aléatoire  $X$ , d'espérance  $\mu$  et d'écart-type  $\sigma$  finis. La loi forte des grands nombres affirme que quand  $n$  tend vers l'infini, la moyenne empirique  $M_n = (x_1 + x_2 + \dots + x_n)/n$  converge presque sûrement vers  $\mu$ , c'est-à-dire que la probabilité que la moyenne empirique converge vers la moyenne théorique est égale à 1.

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} M_n = \mu\right) = 1$$

En d'autres termes, la moyenne d'un grand nombre d'observations aléatoires indépendantes n'est plus aléatoire ! La loi des grands nombres assure ainsi que la moyenne empirique est un estimateur convergent (ou « consistant ») de l'espérance mathématique. Cela reste vrai tant que la contribution de chaque terme  $x_i$  à la moyenne reste négligeable : aucune observation n'est assez grande pour dominer la somme. Cette situation est qualifiée de hasard « sage ».

Le théorème de la limite centrale est l'autre pilier de l'échantillonnage. Considérons des variables aléatoires  $X_i$  indépendantes et



de même distribution (d'espérance  $\mu$  et d'écart-type  $\sigma$ ). Alors la somme centrée  $(X_1 + X_2 + \dots + X_n - n\mu)/\sqrt{n}$  tend vers une variable aléatoire normale (ou gaussienne) d'espérance nulle et d'écart-type  $\sigma$  quand  $n$  tend vers l'infini. La densité de probabilité d'une variable gaussienne de moyenne  $m$  et écart-type  $\sigma$  est définie par la fonction

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

et sa représentation graphique est la célèbre courbe en cloche.

Le théorème de la limite centrale reste valable même si les distributions individuelles sont différentes, pour autant que l'écart-type de chacun des termes soit négligeable vis-à-vis de l'écart-type de la somme.

Ce théorème implique que la différence entre la moyenne empirique  $M_n$  de l'échantillon et l'espérance  $\mu$  de la variable aléatoire  $X$  est représentée par une distribution gaussienne dont l'écart-type est proportionnel à  $1/\sqrt{n}$ . Cela permet de contrôler l'erreur faite en approximant la moyenne théorique  $\mu$  par la moyenne empirique  $M_n$ . La probabilité que  $M_n$  se trouve dans l'intervalle  $[\mu - t\sigma, \mu + t\sigma]$  est représentée, pour  $n$  assez grand, par l'aire sous la courbe en cloche comprise entre les abscisses  $\mu - t\sigma$  et  $\mu + t\sigma$ . Il y a par exemple 95 % de chances que l'écart entre la moyenne empirique  $M_n$  et l'espérance  $\mu$  soit inférieur à  $2\sigma$ .

Le théorème de la limite centrale permet de quantifier la probabilité que la moyenne empirique approche la moyenne réelle avec une précision donnée.

Par exemple, si un sondage électoral porte sur 10 000 personnes et que 46 % d'entre elles se déclarent favorables au candidat A, le théorème de la limite centrale indique qu'il y a 95 % de chance que lors du vote, il recueille entre 45 et 47 % des voix. Concrètement, il n'est pas nécessaire de sonder un très grand échantillon de la population pour avoir une estimation fiable.

Le théorème de la limite centrale est invoqué pour justifier la représentation par une distribution gaussienne dans des domaines allant des sciences sociales à la physique, en passant par l'économie, la biologie ou la finance. La courbe en cloche est mise à toutes les sauces. On l'utilise même pour « normaliser » la distribution des notes dans les concours...

Cependant, l'omniprésence de la distribution gaussienne en statistique (d'où son qualificatif de « loi normale ») est peut-être due à la trop grande confiance des statisticiens plutôt

qu'à la réalité : tout phénomène ne peut pas être ramené à une somme de petites contributions indépendantes. De fait, en sciences économiques et sociales, la loi normale paraît plutôt être l'exception que la règle.

Dès le début du <sup>xx</sup>e siècle, l'économiste Vilfredo Pareto s'est par exemple intéressé à la richesse nationale italienne, et a constaté que les 20 % les plus riches de la population détenaient 80 % de la richesse totale du pays. Cette situation d'inégalité n'est manifestement pas bien représentée par une distribution gaussienne des richesses, où la richesse individuelle serait concentrée autour de la richesse moyenne et où les individus très riches ou très pauvres seraient très rares.

Pareto a proposé une distribution plus réaliste qui porte aujourd'hui son nom : la part d'individus ayant une richesse supérieure à un niveau  $u$  est proportionnelle à  $1/u^\alpha$ . Plus l'exposant  $\alpha$  de cette loi de puissance est petit, plus

## LES LOIS STABLES DE LÉVY

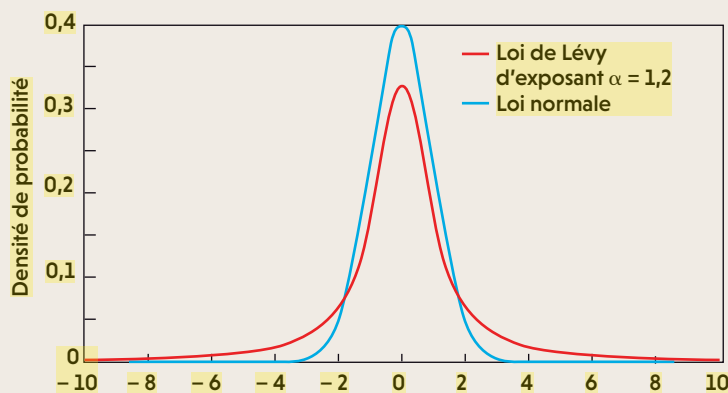


**P**aul Lévy était un pionnier de la théorie moderne des processus aléatoires. Il découvrit dans les années 1920, en même temps que Boris Gnedenko en Union soviétique, les distributions « stables » qui portent aujourd'hui son nom et étendit le théorème de la limite centrale au cas du « hasard sauvage ». Soit une suite  $X_1, X_2, \dots, X_n$  de variables aléatoires indépendantes centrées, dont la queue de distribution se comporte comme une loi de Pareto  $1/x^\alpha$  avec  $0 < \alpha < 2$  (l'écart-type est

alors infini) ; alors la somme  $X_1 + X_2 + \dots + X_n$  n'obéit plus au théorème de la limite centrale.

La distribution de  $(X_1 + X_2 + \dots + X_n)/n^{1/\alpha}$  se comporte, pour  $n$  assez grand, non comme une loi gaussienne, mais comme une « loi stable de Lévy » d'exposant. À la différence du théorème de la limite centrale, le facteur de normalisation est ici  $n^{1/\alpha}$  au lieu de  $n$ .

Les distributions stables n'ont pas de formules analytiques explicites, sauf dans le cas  $\alpha = 1$  (loi de Cauchy) et  $\alpha = 1/2$ . Leur queue de distribution se comporte comme  $1/|x|^\alpha$  pour  $|x|$  assez grand : la somme a la même queue de distribution que chacun de ses termes.



➤ cette proportion est grande et plus la distribution des richesses est inégalitaire. Face à de telles distributions, les notions de moyenne et d'écart-type se révèlent inadaptées et offrent une description trompeuse.

### LES MATHS DES INÉGALITÉS

En effet, la richesse moyenne et l'écart-type peuvent par exemple rester identiques alors que la distribution des richesses se modifie (ce qui est impossible dans le cadre d'une distribution gaussienne).

Pire, si l'exposant de la loi de Pareto est inférieur à deux, alors l'écart-type de la distribution est infini. S'il est inférieur à un, c'est aussi le cas pour la moyenne. Ces indicateurs sont donc dénués de sens pour les lois de Pareto.

Ce fait a conduit les économistes à définir d'autres indicateurs pour mieux caractériser les distributions inégalitaires. Le coefficient de Gini, proposé par le statisticien italien Corrado Gini en 1912, est ainsi un nombre compris entre 0 et 1 qui mesure le degré d'inégalité de la distribution des richesses. Plus il est proche de 1, plus la répartition est inégalitaire. Dans

le cas d'une loi de Pareto d'exposant  $\alpha$ , le coefficient de Gini vaut  $1/(2\alpha - 1)$ . Ces indicateurs se focalisent sur les plus grands événements dans un échantillon, événements d'occurrence faible, mais d'importance prépondérante. Ces « événements rares » forment des queues de distributions.

Dès lors que la moyenne et l'écart-type n'ont plus forcément de sens pour les distributions inégalitaires, qu'advient-il de la loi des grands nombres et du théorème de la limite centrale ? Dans les années 1920, Lévy fut l'un des premiers à remarquer que les hypothèses qui les sous-tendent peuvent cesser d'être valables dans le cas où des événements rares influent de façon prépondérante sur la moyenne de l'échantillon.

Lévy étendit le théorème de la limite centrale à ces situations : il montra que pour un échantillon de variables centrées dont la queue de distribution se comporte comme une loi de Pareto avec un exposant compris entre 0 et 2, la moyenne empirique se comporte non plus comme une variable gaussienne, mais selon une distribution que Lévy appela loi stable, dont la queue se comporte comme une loi de Pareto de même exposant. En d'autres termes, la loi limite est une loi de Lévy, et non une loi normale.

Pour ces distributions de Lévy, variance et écart-type n'ont plus de sens : ils sont infinis. Benoît Mandelbrot, qui fut élève de Lévy à l'École polytechnique, montra en 1963 que, loin d'être une curiosité mathématique, les situations décrites par les lois de Lévy étaient omniprésentes en économie, en finance et en assurance : les rendements des matières premières, la richesse des individus ou la taille des entreprises suivent des lois de Pareto.

Mandelbrot qualifia ces situations, où le comportement de l'échantillon n'est plus correctement caractérisé par la moyenne et l'écart-type, de hasard sauvage, par opposition au hasard sage de la loi gaussienne.

### DOMPTER LE HASARD SAUVAGE

Ce comportement sauvage est manifeste dans les rendements d'indices boursiers (voir l'encadré page 27). Alors que l'amplitude des fluctuations d'un échantillon gaussien reste de l'ordre de son écart-type, certains indices boursiers présentent des variations d'amplitude erratiques atteignant plusieurs dizaines de fois l'écart-type ! De nombreuses autres situations se conformant aux lois de Lévy ont été rapidement mises en évidence en physique statistique.

L'épithète « sauvage » suggère l'incontrôlabilité du hasard, mais une bonne compréhension de l'aléa aide parfois à l'appivoiser. C'est le cas dans certains systèmes physiques, ainsi que dans des situations, où,