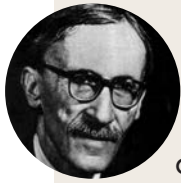


LES LOIS STABLES DE LÉVY

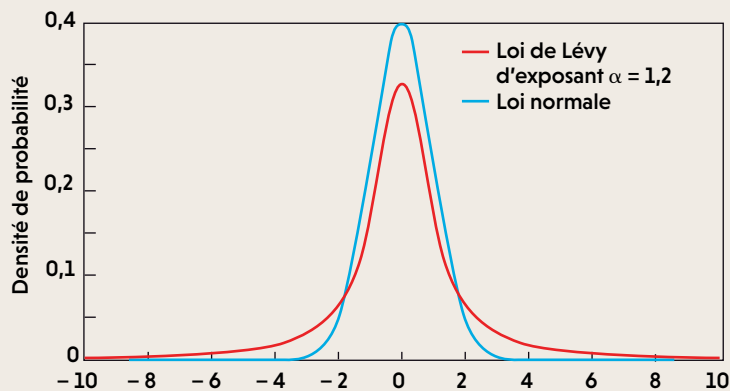


Paul Lévy était un pionnier de la théorie moderne des processus aléatoires. Il découvrit dans les années 1920, en même temps que Boris Gnedenko en Union soviétique, les distributions « stables » qui portent aujourd'hui son nom et étendit le théorème de la limite centrale au cas du « hasard sauvage ». 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 α , 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,

indépendamment du fait que l'aléa sous-jacent soit sage ou sauvage, la moyenne ou l'écart-type d'un échantillon ne sont tout simplement pas les indicateurs pertinents, même si on peut les estimer avec précision.

Il en est ainsi des études de fiabilité, où, pour déterminer la probabilité de défaillance d'un système, on cherche à estimer celle de son « maillon le plus faible ». Par exemple, pour choisir la dimension et les caractéristiques d'un barrage, il faut avoir une idée précise de la pression maximale que l'ouvrage pourra être amené à supporter, et non de la pression moyenne qu'il subira en conditions normales. Pour cela, il faut caractériser directement les valeurs extrêmes dans l'échantillon d'observations. Les notions d'écart-type, de moyenne et le théorème de la limite centrale ne sont alors pas d'un grand secours...

Parallèlement aux travaux de Lévy sur le théorème de la limite centrale, un groupe de statisticiens, dont sir Ronald Fisher et Leonard

Tippett en Grande-Bretagne, Boris Gnedenko en Union soviétique, Maurice Fréchet en France et Ludwig von Mises en Autriche, ainsi qu'Emil Julius Gumbel, mathématicien allemand réfugié en France, développaient une nouvelle branche des statistiques dont le but est d'étudier non plus les valeurs moyennes des échantillons, mais leurs valeurs extrêmes, c'est-à-dire maximales ou minimales.

De même que le théorème de la limite centrale décrit le comportement de la moyenne d'un grand échantillon, cette théorie des valeurs extrêmes décrit le comportement du maximum d'un échantillon. Ces statisticiens ont montré que, selon la nature de la variable aléatoire étudiée, la valeur maximale de l'échantillon ne peut obéir qu'à l'une parmi trois distributions différentes : les distributions de Fréchet, de Gumbel ou de Weibull (voir l'encadré ci-contre).

Comme la loi normale et la loi de Lévy, qui sont des candidats naturels pour modéliser les sommes de variables indépendantes, les lois de valeurs extrêmes sont des candidats privilégiés pour la modélisation des valeurs maximales d'un échantillon. Gumbel présenta en 1941 une application de ces concepts à l'analyse statistique des crues fluviales. Cette théorie allait être mise en application dans un contexte dramatique.

APPRIVOISER LA RARETÉ... ET LA MER

Les Pays-Bas ont toujours vécu sous la menace de la montée des eaux. Les nombreuses digues qui parsemaient le pays ont longtemps suffi à le protéger. Mais dans la nuit du 31 janvier au 1^{er} février 1953, une montée des eaux due à une tempête eut raison de leur résistance et submergea le Sud-Ouest du pays, tuant plus de 1800 personnes et des milliers de têtes de bétail.

Sous le choc, le gouvernement néerlandais lança peu après un vaste projet de construction de barrages. La commission chargée de revoir les normes de sécurité exigea de fixer la hauteur des digues de façon à ce que la probabilité que la marée les dépasse soit inférieure à 1/10000. Une équipe de scientifiques, armée de la récente théorie des valeurs extrêmes, se mit alors à étudier marées et crues passées afin d'estimer la distribution de probabilité de la hauteur maximale des inondations. Ils montrèrent qu'elle suit une loi de Gumbel, et que la hauteur requise dépassait 5 mètres, une nette augmentation par rapport aux digues de l'époque. Sur la base de ces estimations, d'énormes barrages furent construits (voir la photo page 26) ; ces édifices protègent encore le pays aujourd'hui.

Dans une ambiance de foi grandissante dans les méthodes quantitatives, la théorie des valeurs extrêmes fut par la suite appliquée aux

LES LOIS DE VALEURS EXTRÊMES



Considérons un échantillon $x_1, x_2, \dots, x_n$ de tirages indépendants d'une distribution de probabilité G . La théorie des valeurs extrêmes s'intéresse au comportement de la valeur maximale $U_n = \max(x_1, x_2, \dots, x_n)$ d'un échantillon de grande taille (n grand). Un des résultats fondamentaux de la théorie, le théorème de Fisher-Tippett-Gnedenko, montre que s'il existe des constantes de normalisation a_n et b_n telles que la distribution de $(U_n - a_n)/b_n$ ait une limite quand n tend vers l'infini, alors la distribution limite de la suite normalisée de maxima des variables aléatoires est nécessairement de la forme :

$$F(x; \mu, \sigma, \xi) = \exp \left\{ - \left[1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} \right\}$$

définie pour $1 + \xi(x - \mu)/\sigma > 0$, et $\sigma > 0$.

On distingue alors trois cas.

Le premier, quand $\xi < 0$, nommé distribution de Weibull, décrit le cas où les variables x_i sont bornées.

Le second, lorsque $\xi = 0$, correspond à la distribution de Gumbel. Elle est au maximum d'un échantillon ce que la loi normale est à la moyenne : elle représente le cas où le hasard décrit par la distribution G est sage.

Enfin, si les variables x_i suivent une loi de Pareto d'exposant α , alors $\xi = 1/\alpha > 0$. La distribution, dite de Fréchet, décrit le maximum d'un échantillon qui obéit au hasard sauvage. Plus ξ est grand, plus la queue de distribution est importante, c'est-à-dire plus les événements extrêmes sont fréquents. Ces trois lois de probabilité sont adaptées aux distributions de valeurs extrêmes.

