

LOI DES GRANDS NOMBRES ET THÉORÈME DE LIMITE CENTRALE

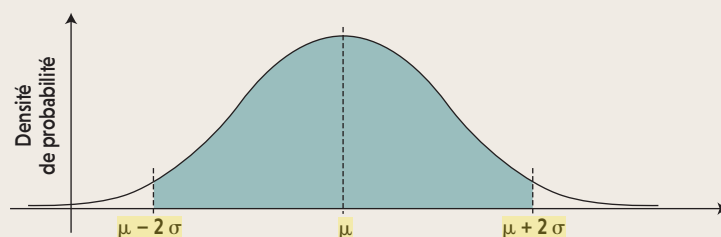
La loi des grands nombres énonce que lorsqu'on échantillonne une variable aléatoire, sous certaines conditions, lorsque la taille de l'échantillon est assez grande, la moyenne empirique de cet échantillon tend vers la moyenne théorique de la variable aléatoire échantillonnée.

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 μ et d'écart-type σ 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 μ , 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 μ et d'écart-type σ). 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 σ quand n tend vers l'infini. La densité de probabilité d'une variable gaussienne de moyenne m et écart-type σ 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 μ 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 μ 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 μ soit inférieur à 2σ .

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 xx^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 α de cette loi de puissance est petit, plus >