6 Théorèmes limites

Les théorèmes limites constituent les résultats théoriques les plus importants, des probabilités. Parmi eux, les principaux sont répertoires sous deux dénominations : lois des grands nombres d'une part et théorèmes centraux limites d'autre part. On s'accord généralement à les considérer comme des lois des grands nombres s'ils énoncent des conditions sous lesquelles la moyenne d'une suite de variables aléatoires converge vers leur espérance commune. Les théorèmes centraux limites par contre déterminent sous quelles hypothèses la somme d'un grand nombre de variables aléatoires est de distribution approximativement normale.

6.1 Inégalité de Markov

Soit ξ une variable aléatoire à valeurs non-négatives. Pour tout réel $a > 0$

$$P(\xi \geq a) \leq \frac{E(\xi)}{a}.$$

Démonstration: La démonstration qui suit s'applique à une variable ξ continue de densité f.

$$E(\xi) = \int_0^\infty x f(x)\,\mathrm{d}x = \int_0^a x f(x)\,\mathrm{d}x + \int_a^\infty x f(x)\,\mathrm{d}x \geq \int_a^\infty x f(x)\,\mathrm{d}x \geq$$
$$\geq \int_a^\infty a f(x)\,\mathrm{d}x = a \int_a^\infty f(x)\,\mathrm{d}x = aP(\xi \geq a).$$

6.2 Inégalité de Tchebyshev

Soit ξ une variable aléatoire d'espérance μ et de variance σ^2 finies. Pour tout réel $k > 0$

$$P(|\xi - \mu| \geq k) \leq \frac{\sigma^2}{k^2}.$$

Démonstration: On peut appliquer l'inégalité de Markov, avec $a = k^2$, à la variable $(\xi - \mu)^2$ puisque celle-ci est à valeurs non-négatives. On obtient

$$P((\xi - \mu)^2 \geq k^2) \leq \frac{E[(\xi - \mu)^2]}{k^2}$$

mais comme $(\xi - \mu)^2 \geq k^2$ équivaut à $|\xi - \mu| \geq k$ la formule peut être reécrite

$$P(|\xi - \mu| \geq k) \leq \frac{E[(\xi - \mu)^2]}{k^2} = \frac{\sigma^2}{k^2}.$$

L'importance des inégalités de Markov et de Tchebychev réside en ce qu'elles permettent de borner la valeur de certaines probabilités là où seule l'espérance de la distribution est connue, plus éventuellement sa variance. Il est évident que, si la distribution elle-même est connue, on ne recourra pas à des bornes, puisque la valeur exacte de ces probabilités est calculable.

Exemple: *On suppose que le nombre de pièces sortant d'une usine donnée en l'espace d'une semaine est une variable aléatoire d'espérance* 50*.*

a. Peut-on estimer la probabilité que la production de la semaine prochaine dépasse 75 pièces ?
b. On sait de plus que la variance de la production hebdomadaire est de 25. Peut-on estimer la probabilité que la production de la semaine à venir soit comprise entre 40 et 60 ?

Réponse: Soit ξ le nombre de pièces produites en une semaine.

a. L'inégalité de Markov donne

$$P(\xi > 75) \leq \frac{E(\xi)}{75} = \frac{2}{3}.$$

b. L'inégalité de Tchebychev donne

$$P(|\xi - 50| \geq 10) \leq \frac{\sigma^2}{10^2} = \frac{1}{4}$$

et donc

$$P(|\xi - 50| < 10) \geq 1 - \frac{1}{4} = \frac{3}{4}.$$

L'inégalité de Tchebychev étant valable pour n'importe quelle distribution de la variable ξ, il ne faut pas s'attendre à ce que la borne qu'elle fournit soit très proche de la probabilité exacte dans la majorité des cas.

6.3 La loi faible des grands nombres

Soient $\xi_1, \xi_2, \ldots$ une suite de variables aléatoires indépendantes et identiquement distribuées. On suppose que toutes admettent la même espérance finie $E(\xi) = \mu$. Pour tout $\varepsilon > 0$

$$P\left(\left|\frac{\xi_1 + \xi_2 + \cdots + \xi_n}{n} - \mu\right| > \varepsilon\right) \to 0 \quad \text{lorsque } n \to \infty.$$

6.4 Théorème central limite

Soient $\xi_1, \xi_2, \ldots$ une suite de variables aléatoires indépendantes et identiquement distribuées, d'espérance μ, de variance σ^2. Alors la distribution de

$$\xi_n^* = \frac{\xi_1 + \cdots + \xi_n - n\mu}{\sigma\sqrt{n}}$$

tend vers la distribution normale réduite lorsque $n \to \infty$, ce qui veut dire que pour tout a, b

$$P(a < \xi_n^* < b) \to \phi(b) - \phi(a).$$

En termes moins rigoureux, le théorème central limite affirme donc que la somme d'un grand nombre de fluctuations aléatoires indépendantes obéit, sous des conditions assez générales, à une distribution proche d'une distribution normale.

Notons que la distribution des sommes non réduites ξ_n ne converge pas vers une loi normale ; en effet la densité de probabilité de ξ_n devient de plus en plus plate et converge partout vers zéro si n augmente indéfiniment.

Remarquons d'ailleurs qu'il existe d'autres versions du théorème central limite.

Exemple: *Le nombre d'inscriptions à un cours de psychologie est une variable aléatoire de Poisson d'espérance 100. Le professeur donnant ce cours a décidé que, si le nombre des inscriptions est au-delà de 120, il créera deux sections et donnera donc deux cours, tandis qu'en deçà une seule classe sera formée. Quelle est la probabilité que ce professeur ait à donner deux fois le cours ?*

Réponse: La solution exacte est $e^{-100} \cdot \sum_{i=120}^{\infty} \frac{100^i}{i!}$, et son évaluation numérique n'est pas aisée. On se souvient cependant qu'une variable poissonienne de paramètre 100 peut être considérée comme la somme de 100 variables poissoniennes indépendantes de paramètre 1 chacune, ce qui donne l'occasion d'utiliser le théorème central limite pour obtenir une approximation de la réponse exacte. Soit ξ le nombre d'inscriptions. On a

$$P(\xi \geq 120) = P\left(\frac{\xi - 100}{\sqrt{100}} \geq \frac{120 - 100}{\sqrt{100}}\right) \approx 1 - \phi(2) = 0.0228$$

où l'on utilisé le fait qu'espérance et variance des variables poissoniennes sont toujours égales.