Leçon 11

Inégalités de Markov et de Tchebychev

Les inégalités sur des probabilités sont d'usage constant ; elles permettent notamment de quantifier le comportement (à l'infini) de la fonction de répartition d'une variable aléatoire. La plus simple et la plus fondamentale est l'*inégalité de Markov*[1].

1 Inégalité de Markov

Proposition 1 (Inégalité de Markov). *Soit X une variable aléatoire positive ; pour tout $t > 0$,*

$$\mathbb{P}(X \geq t) \leq \frac{1}{t}\,\mathbb{E}(X).$$

La démonstration est immédiate (et repose sur la propriété de conservation de l'ordre de l'intégrale déjà mise en œuvre de façon analogue dans les rappels de la théorie de l'intégration) : pour chaque $\omega \in \Omega$,

$$\mathbb{1}_{\{X \geq t\}}(\omega) = \mathbb{1}_{[t,\infty]}\big(X(\omega)\big) \leq \frac{1}{t}\,X(\omega)$$

puisque si $X(\omega) \geq t$, alors $1 \leq \frac{1}{t}\,X(\omega)$ et si $X(\omega) < t$, $0 \leq \frac{1}{t}\,X(\omega)$. L'intégrale conservant le sens des inégalités,

$$\mathbb{P}(X \geq t) = \int_\Omega \mathbb{1}_{\{X \geq t\}}\,d\,\mathbb{P} \leq \frac{1}{t}\int_\Omega X\,d\,\mathbb{P} = \frac{1}{t}\,\mathbb{E}(X)$$

qui est le résultat.

Comme à l'habitude, il suffit que X soit presque sûrement positive, l'argument se déployant pour l'ensemble des $\omega \in \{X \geq 0\}$ de probabilité 1.

L'inégalité de Markov n'a bien entendu d'intérêt que si X est intégrable, et $t > \mathbb{E}(X)$. En fait, sous l'hypothèse d'intégrabilité, elle fournit une décroissance

1. Andreï Markov, mathématicien russe (1856–1922).

de $1-F_X(t)$ vers 0 lorsque $t \to \infty$. Par continuité du majorant $\frac{1}{t}\,\mathbb{E}(X)$ en $t > 0$, il est équivalent d'avoir

$$\mathbb{P}(X > t) \;\leq\; \frac{1}{t}\,\mathbb{E}(X)$$

pour tout $t > 0$. En effet, si tel est le cas, pour tout $\varepsilon > 0$ tel que $t - \varepsilon > 0$,

$$\mathbb{P}(X \geq t) \;\leq\; \mathbb{P}(X > t - \varepsilon) \;\leq\; \frac{1}{t - \varepsilon}\,\mathbb{E}(X)$$

et l'affirmation s'ensuit quand $\varepsilon \to 0$. Prendre soin toutefois que ce raisonnement ne s'applique pas t par t.

La souplesse de l'inégalité de Markov, ainsi qu'il sera développé dans les paragraphes suivants, est illustré par le fait que si ϕ est une fonction croissante sur $\mathbb{R}$ à valeurs strictement positives, pour toute variable aléatoire réelle X et tout $t \in \mathbb{R}$,

$$\mathbb{P}(X \geq t) \;\leq\; \mathbb{P}\big(\phi(X) \geq \phi(t)\big) \;\leq\; \frac{1}{\phi(t)}\,\mathbb{E}\big(\phi(X)\big).$$

Une nouvelle fois, cette inégalité n'a d'intérêt que si $\phi(X)$ est intégrable, et le rapport de droite est inférieur à 1. Ce principe est appliqué ci-dessous pour différents exemples de fonctions ϕ, comme les fonctions puissances $\phi(x) = |x|^r$, $r > 0$, ou la fonction exponentielle $\phi(x) = e^{\theta x}$, $\theta > 0$.

2 Inégalité de Tchebychev

Sous des hypothèses de moments plus fortes, la décroissance est plus rapide. C'est notamment le cas pour des moments d'ordre 2 donnant lieu à l'inégalité de Tchebychev [2]. En fait, cette dernière peut être formulée sous la forme d'une inégalité de concentration autour de la valeur moyenne.

2. Pafnouti Tchebychev, mathématicien russe (1821–1894).

Proposition 2 (Inégalité de Tchebychev)**.** *Soit X une variable aléatoire de carré intégrable ; pour tout $t > 0$,*

$$\mathbb{P}\big(\big|X - \mathbb{E}(X)\big| \geq t\big) \ \leq \ \frac{1}{t^2}\,\mathrm{Var}(X).$$

Elle découle de la remarque de la fin du paragraphe précédent : il suffit d'appliquer l'inégalité de Markov à $Z^2 = [X - \mathbb{E}(X)]^2$ et t^2 puisque $|X - \mathbb{E}(X)| \geq t$ si et seulement si $Z^2 \geq t^2$ et

$$\mathbb{E}(Z^2) \ = \ \mathbb{E}\big(\big[X - \mathbb{E}(X)\big]^2\big) \ = \ \mathrm{Var}(X).$$

L'inégalité de Tchebychev peut être couplée avec l'identité de Bienaymé pour fournir *l'inégalité de Bienaymé-Tchebychev*. La démonstration est contenue dans l'énoncé.

Proposition 3 (Inégalité de Bienaymé-Tchebychev)**.** *Si $X_1, \ldots, X_n$ sont des variables aléatoires de carré intégrable deux à deux non corrélées, et si $S_n = X_1 + \cdots + X_n$, pour tout $t > 0$,*

$$\mathbb{P}\big(\big|S_n - \mathbb{E}(S_n)\big| \geq t\big) \ \leq \ \frac{1}{t^2}\,\mathrm{Var}(S_n) \ = \ \frac{1}{t^2}\sum_{k=1}^{n}\mathrm{Var}(X_k).$$

3 Inégalité exponentielle

Il est facile d'imaginer que la puissance 2 dans l'inégalité de Tchebychev pourrait être remplacée par n'importe quelle puissance $r > 0$, conduisant à une décroissance polynomiale sous une hypothèse de moment.

Une variation sur ce thème utilise la fonction exponentielle, et est liée à la transformée de Laplace (Leçon 10). Par exemple, si X est une variable aléatoire réelle et $t > 0$, pour tout $u > 0$, par la croissance de la fonction exponentielle et l'inégalité de Markov,

$$\mathbb{P}(X \geq t) \, = \, \mathbb{P}\big(e^{uX} \geq e^{ut}\big) \, \leq \, e^{-ut} \, \mathbb{E}(e^{uX}).$$

L'inégalité reste vraie pour tous $t, u \geq 0$. Une nouvelle fois, cette borne n'a d'intérêt que si $\mathbb{E}(e^{uX}) < \infty$, mais son profit est ailleurs : en effet, sous réserve de ces hypothèses d'intégrabilité, il est avantageux suivant les cas d'optimiser, à $t \geq 0$ fixé, sur les réels positifs u pertinents afin de dégager la meilleure borne possible.

Un exemple simple est fourni par le cas d'une variable X de loi binomiale $\mathcal{B}(n, \frac{1}{2})$. Il est aisé de vérifier que, pour tout réel u,

$$\mathbb{E}(e^{uX}) \, = \, \frac{1}{2^n} \sum_{k=0}^{n} \binom{n}{k} e^{uk} \, = \, \frac{1}{2^n} \, (e^u + 1)^n$$

d'après la formule du binôme. Ainsi, pour tous $t \geq 0$ et $u \geq 0$,

$$\mathbb{P}(X \geq t) \, \leq \, e^{-ut} \, \frac{1}{2^n} \, (e^u + 1)^n.$$

L'optimisation exacte en $u \geq 0$ n'est pas nécessairement facile, mais si $t = \frac{n}{2} + s$, $s \geq 0$, l'inégalité précédente se réécrit comme

$$\mathbb{P}\big(X \geq \tfrac{n}{2} + s\big) \, \leq \, e^{-us} \left(\frac{e^{\frac{u}{2}} + e^{-\frac{u}{2}}}{2} \right)^n.$$

Un développement en série montre que $\frac{e^{\frac{u}{2}} + e^{-\frac{u}{2}}}{2} \, (= \mathrm{ch}(\frac{u}{2})) \leq e^{\frac{u^2}{8}}$. Ainsi

$$\mathbb{P}\big(X \geq \tfrac{n}{2} + s\big) \, \leq \, e^{-us + \frac{n}{8} u^2}$$

et l'optimisation en u dans l'exponentielle ($u = \frac{4s}{n}$) implique au final que, pour tout $s \geq 0$,

$$\mathbb{P}\big(X \geq \tfrac{n}{2} + s\big) \, \leq \, e^{-\frac{2}{n} s^2}.$$

La valeur $\frac{n}{2}$ n'est pas une surprise, c'est simplement l'espérance de X, de sorte l'inégalité précédente exprime une inégalité de déviation au dessus de la valeur moyenne.

Cette borne exponentielle participe également du principe de grandes déviations qui sera évoqué en Leçon 19.

4 Inégalité de concentration

L'inégalité de Tchebychev et, pour des sommes de variables aléatoires non corrélées, l'inégalité de Bienaymé-Tchebychev, sont des exemples d'inégalités dites de *concentration*, au sens où la variable étudiée se concentre autour de sa valeur moyenne au taux fourni par la variance. Plus précisément, si X est une variable aléatoire réelle de carré intégrable, l'inégalité de Tchebychev exprime que pour tout $t > 0$,

$$\mathbb{P}\big(\big|X - \mathbb{E}(X)\big| \geq t\big) \,\leq\, \frac{1}{t^2}\,\mathrm{Var}(X).$$

Si pour $\varepsilon > 0$, $t > 0$ est choisi de sorte que $\frac{1}{t^2}\,\mathrm{Var}(X) = \varepsilon$, par passage au complémentaire, avec probabilité plus grande que $1 - \varepsilon$,

$$\big|X - \mathbb{E}(X)\big| \,\leq\, t \,=\, \sqrt{\frac{\mathrm{Var}(X)}{\varepsilon}}\,.$$

Si la variance de X est très faible, X sera donc proche de sa valeur moyenne (constante) $\mathbb{E}(X)$ avec forte probabilité. Cette discussion sera reprise dans la Leçon 21 sur les intervalles de confiance.

Ce type de raisonnement est d'autant plus puissant que l'inégalité est forte, ce qui dépend des variables considérées. Il est à ce titre instructif de reprendre l'exemple de la variable aléatoire X de loi binomiale $\mathcal{B}(n, \frac{1}{2})$ du paragraphe précédent pour laquelle il a été établi que, pour tout $s \geq 0$,

$$\mathbb{P}\big(X \geq \tfrac{n}{2} + s\big) \,\leq\, e^{-\frac{2}{n}s^2}.$$

Le raisonnement développé montre en fait qu'il en va de même pour $-X$ sous la forme

$$\mathbb{P}\big(-X \geq -\tfrac{n}{2} + s\big) \;\leq\; e^{-\frac{2}{n}s^2}$$

pour tout $s \geq 0$. Comme $\mathbb{E}(X) = \frac{n}{2}$, la somme de ces deux inégalités (et le fait que $\mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$) entraîne que

$$\mathbb{P}\big(\big|X - \mathbb{E}(X)\big| \geq s\big) \;\leq\; 2\,e^{-\frac{2}{n}s^2}$$

pour tout $s \geq 0$. En particulier, si s est de l'ordre de $10\sqrt{n}$ (par exemple !), alors, avec probabilité plus grande que $1 - 2e^{-200}$ (très proche de 1 donc), $|X - \mathbb{E}(X)| \leq s$. Il convient de rappeler ici que l'éventail des valeurs de X va jusqu'à n, et recentré autour de la valeur moyenne $\frac{n}{2}$, il n'est plus que $\sqrt{n}$. Ce taux en $\sqrt{n}$ dans le choix de u est lié au théorème central limite (Leçon 20).

Il s'agit donc là d'une propriété de concentration de la variable binomiale X autour de sa valeur moyenne $\mathbb{E}(X) = \frac{n}{2}$ à un taux exponentiel, plus fort que celui fourni par l'inégalité de Tchebychev, à savoir

$$\mathbb{P}\big(\big|X - \mathbb{E}(X)\big| \geq s\big) \;\leq\; \frac{n}{4s^2}$$

pour tout $s > 0$ puisque $\mathrm{Var}(X) = \frac{n}{4}$ (dans ce cas, si $s \geq 10\sqrt{n}$, la probabilité considérée n'est que de $1 - \frac{1}{400}$).

Ces exemples d'inégalités de concentration seront repris dans la Leçon 13 sur les sommes de variables aléatoires indépendantes.

Exercices

(*Une étoile * désignera une question de difficulté supérieure.*)

Exercice 1. Soit X une variable aléatoire positive et intégrable sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$; démontrer que pour tout $\varepsilon > 0$, il existe $\omega \in \Omega$ tel que $X(\omega) \leq (1+\varepsilon)\,\mathbb{E}(X)$. (*Indication* : utiliser l'inégalité de Markov pour $t = (1+\varepsilon)\,\mathbb{E}(X)$ et le fait que si $\mathbb{P}(A) > 0$, alors A est non vide.)

Exercice 2.

a) Soit X une variable aléatoire réelle de carré intégrable telle que $\mathbb{E}(X) = 0$ et $\mathbb{E}(X^2) > 0$; démontrer que $\mathbb{P}(X < 0) > 0$ et $\mathbb{P}(X > 0) > 0$. (*Indication* : procéder par contradiction.)

b*) Soit X une variable aléatoire réelle telle que $\mathbb{E}(X^2) \geq a > 0$ et $\mathbb{E}(X^4) \leq b < \infty$; démontrer que $\mathbb{P}(|X| > 0) \geq \frac{a^2}{b}$. (*Indication* : utiliser l'inégalité de Cauchy-Schwarz sur $\mathbb{E}(X^2) = \mathbb{E}(X^2 \mathbb{1}_{\{|X|>0\}})$.)

Exercice 3* (*Inégalité de Cantelli*). Pour X une variable aléatoire réelle sur $(\Omega, \mathcal{A}, \mathbb{P})$ de carré intégrable de variance $\sigma^2 = \mathrm{Var}(X)$, démontrer que pour tout $t > 0$,

$$\mathbb{P}\big(X - \mathbb{E}(X) \geq t\big) \;\leq\; \frac{\sigma^2}{\sigma^2 + t^2}\,.$$

Comparer avec l'inégalité de Tchebychev. (*Indication* : supposer sans perte de la généralité que $\mathbb{E}(X) = 0$ et $\sigma > 0$, appliquer l'inégalité de Markov avec un moment d'ordre 2 à

$$\mathbb{P}(X \geq t) \;=\; \mathbb{P}(X + a \geq t + a)$$

pour tout $a \geq 0$, et optimiser en a.)

Exercice 4* (*Inégalité de Paley*[3]*-Zygmund*[4]). Soit X une variable aléatoire positive sur $(\Omega, \mathcal{A}, \mathbb{P})$ telle que $0 < \mathbb{E}(X^2) < \infty$, et soit $t \in]0,1[$.

a) Comparer $\mathbb{E}(X \mathbb{1}_{\{X < t\,\mathbb{E}(X)\}})$ et $t\,\mathbb{E}(X)$. En déduire que

$$(1-t)\,\mathbb{E}(X) \;\leq\; \mathbb{E}\big(X \mathbb{1}_{\{X \geq t\,\mathbb{E}(X)\}}\big).$$

b) À partir de la question précédente, établir l'inégalité de Paley-Zygmund

$$\mathbb{P}\big(X \geq t\,\mathbb{E}(X)\big) \;\geq\; (1-t)^2\,\frac{[\mathbb{E}(X)]^2}{\mathbb{E}(X^2)}\,.$$

c) Si X suit la loi binomiale $\mathcal{B}(n, \frac{1}{2})$, proposer une minoration de $\mathbb{P}(X \geq \frac{tn}{2})$ pour $t \in]0,1[$.

Exercice 5. Vérifier l'inégalité

$$\mathbb{P}\big(-X \geq -\tfrac{n}{2} + s\big) \;\leq\; e^{-\frac{2}{n}s^2}$$

pour tout $s \geq 0$ du Paragraphe 4, où donc X est une variable aléatoire de loi binomiale $\mathcal{B}(n, \frac{1}{2})$.

Exercice 6. Une variable aléatoire réelle X sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ suit la loi de Poisson $\mathcal{P}(\theta)$ de paramètre $\theta > 0$.

a) Déterminer le domaine de définition de la transformée de Laplace L_X de (la loi de) X, et la calculer.

b) Soit $t \geq 0$ fixé ; démontrer que pour tout $u \geq 0$,

$$\mathbb{P}(X \geq t) \;\leq\; e^{-ut}\,L_X(u).$$

c) Déduire de la question précédente que pour tout $t > \theta$,

$$\mathbb{P}(X \geq t) \;\leq\; e^{-t\ln(t) + t(\ln(\theta)+1) - \theta}.$$

3. Raymond Paley, mathématicien anglais (1907–1933).
4. Antony Zygmund, mathématicien polonais et américain (1900–1992).