Chapitre 6

Loi des grands nombres

Avec ce chapitre nous abordons le point essentiel du paradigme de l'application du calcul des probabilité qui a donné son nom à ce calcul : la relation qu'il y a entre l'observation du nombre des succès lors de "répétitions indépendantes" d'une expérience à l'issue incertaine, tel l'obtention d'une "face" dans un tirage à pile ou face, et la "chance de succès", dite "probabilité de réalisation de l'évènement". Nous commençons par deux inégalités qui nous serviront dans la preuve de ce point essentiel

6.1 Les inégalités de Markov et Bienaymé-Tchébicheff

Au chapitre 4 nous avons dit que la variance est une mesure de la dispertion de la loi d'un v.a. X loin de son espérance. L'inégalité de Bienaymé-Tchebicheff[1] précise ce point. La preuve de cette inégalité est "diaboliquement simple", l'idée se généralisant directement à une situation un peu plus générale dite "inégalité de Markov".

Proposition 6.1 (inégalité de Markov) *Soit Y une v.a. quelconque ; pour toute fonction $h : \mathbb{R}^+ \longrightarrow \mathbb{R}^+$, $a \mapsto h(a)$, croissante strictement positive pour $a > 0$ telle que $\mathbb{E}(h(Y)) < +\infty$, et pour tout $a > 0$ on a*

$$\mathbb{P}(\{|Y| \geq a\}) \leq \frac{1}{h(a)} \mathbb{E}(h(|Y|)). \tag{6.1}$$

Preuve : L'idée tient dans le fait de choisir l'écriture linéaire de la probabilité, à savoir $\mathbb{P}(\{|Y| \geq a\}) = \mathbb{E}(\mathbb{I}_{\{|Y| \geq a\}})$ et de remarquer que, comme h est croissante positive, on a $h(a)\mathbb{I}_{\{|Y| \geq a\}}) \leq h(|Y|)\mathbb{I}_{\{|Y| \geq a\}})(\leq h(|Y|))$. La suite découle de la linéarité et de la positivité de l'espérance :

$$\mathbb{P}(\{|Y| \geq a\}) = \mathbb{E}(\mathbb{I}_{\{|Y| \geq a\}}) = \frac{1}{h(a)} \mathbb{E}(h(a)\mathbb{I}_{\{|Y| \geq a\}}) \leq \frac{1}{h(a)} \mathbb{E}(h(|Y|)\mathbb{I}_{\{|Y| \geq a\}}) \leq \frac{1}{h(a)} \mathbb{E}(h(|Y|))$$

$\square$

En appliquant ce résultat à $Y := X - \mathbb{E}(X)$, en posant $h(x) := x^2$ et $a := \lambda$ nous obtenons l'inégalité de Bienaymé-Tchebicheff :

Proposition 6.2 *Pour toute v.a. $X \in L^2(\Omega)$ on a*

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq \lambda) \leq \frac{1}{\lambda^2} \mathit{Var}(X) \tag{6.2}$$

6.2 La Loi des Grands Nombres (LGN)

Théorème 6.3 *Soit $(X_i)_{i \geq 0}$ une suite de v.a. i.i.d., ayant une espérance μ et un écart-type. Pour chaque $n \geq 1$, soit $M_n := \frac{1}{n}(X_1 + \ldots + X_n)$ la moyenne des n premières. Alors la suite de ces moyennes M_n tend vers le nombre μ dans le sens suivant :*

$$\textit{Pour tout } \lambda > 0, \lim_{n \to +\infty} \mathbb{P}(\{|M_n - \mu| \geq \lambda\}) = 0. \tag{6.3}$$

[1] le français Bienaymé était ami du russe Tchebicheff (Chebyshev en transcription anglo-saxonne), tout comme du belge Quetelet- un des fondateurs de la "statistique" au sens étymologique : sciences de l'Etat (sociologie quantitative). Il est d'usage d'accoler le nom de Tchebicheff à l'inégalité de Bienaymé car c'est Tchebichev qui l'a utilisée le premier à la généralisation de la loi des Grands Nombres de Bernoulli

Preuve : Notons σ l'écart-type commun des X_i ; observons que l'espérance les M_n est également μ, et comme les v.a. X_i sont indépendantes, la variance des M_n est égale à $\frac{\sigma^2}{n}$; en effet

$$\mathbb{E}(M_n) = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(X_i) = \frac{1}{n}\sum_{i=1}^{n}\mu = \mu \text{ , et} \tag{6.4}$$

$$\mathrm{Var}\,(M_n) = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}\,(X_i) = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{\sigma^2}{n}. \tag{6.5}$$

A présent, il suffit d'écrire l'inégalité de Tchébicheff pour M_n :

$$0 \le \mathbb{P}(|M_n - \mu| \ge \lambda) = \mathbb{P}(|M_n - \mathbb{E}(M_n)| \ge \lambda) \le \frac{1}{\lambda^2}\mathrm{Var}\,(M_n) \le \frac{\sigma^2}{n\lambda^2}.$$

On conclut on observant que $\lim_{n\to+\infty}\frac{\sigma^2}{n\lambda^2} = 0$. □

Exemple : Considérons un phénomènes pouvant se produire ou non à chaque expérience, tel la sortie d'une face dans un tirage à pile ou face, ou la sortie d'un 6 dans le tirage d'un dé. On répète l'expérience de manière identique de manière à ce que le résultat d'une expérience n'influe pas sur les expériences suivantes. On compte le nombre $S(n)$ de fois où le phénomène s'est produit au cours des n premières expériences, et on forme le rapport $M(n) = S(n)/n$, égal au nombre moyen de "succès". On dira que le phénomène se produit de manière aléatoire avec la probabilité p dans les conditions de l'expérience choisies si l'on peut lui appliquer le modèle probabiliste suivant : l'apparition du phénomène lors de la i-ème expérience est une v.a. de Bernouilli $X_i \rightsquigarrow \mathcal{B}(1,p)$, les diverses v.a. X_i étant supposées indépendantes. Dans ce cas, le rapport $M(n) = S(n)/n$ observé est modélisé par la v.a. $M_n = \frac{1}{n}\sum_{i=1}^{n} X_i$.

Que nous dit la Loi des Grands Nombres pour ce modèle ? On a $\mu = \mathbb{E}(X_i) = p\cdot 1 + (1-p)\cdot 0 = p$; donnons nous un $\lambda > 0$, par exemple $\lambda = 0.01$. Plus n est grand, plus il est improbable que $M_n \notin]p-\lambda, p+\lambda[$. Si le phénomène considéré est effectivement aléatoire (dans le conditions de l'expérience), les valeurs observées du rapport $M(n)$ doivent donc, au fur et à mesure que n augmente, "se grouper autour" d'une valeur $\hat{p}$, sans que cela n'exclue de "petite excursions" hors de $]\hat{p}-\lambda, \hat{p}+\lambda[$, celles-ci "devenant de plus en plus rares".

Si tel est le cas, on considérera que "le phénomène se produit avec la probabilité $\hat{p}$".

L'exemple qui précède correspond à la Loi des Grands Nombre, telle qu'elle a été découverte par Jacques Bernouilli, pour le cas particulier de v.a. de Bernoulli, précisément, et qui constitue le *théorème de Bernoulli*. Ce cas à l'avantage de la simplicité, la loi commune des v.a. se réduisant au choix d'un unique paramètre p, que la Loi de Grands Nombre révèle par la *limite en probabilité* des $M_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. C'est cette application du théorème de Bernoulli qui fonde l'utilisation du calcul (abstrait) des probabilités comme cadre mathématique de la Statistique, littéralement "science des Etats", ou "Physique sociale" pour reprendre le nom d'un livre d'Adolphe Quételet (Bruxelles, 1869).

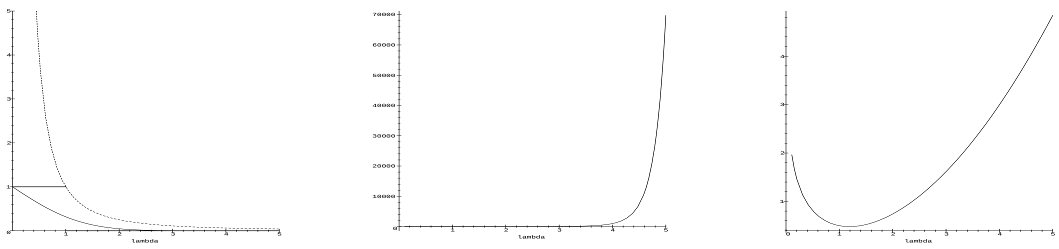
6.3 Portée réelle de ces résultats

6.3.1 Que nous apprennent vraiment l'inégalités de Bienaymé-Tchébicheff ?

Sur le premier graphique de la figure 6.1, nous avons représenté simultanément la fonction $\frac{1}{\lambda^2}\mathrm{Var}\,(X)$ pour une loi de variance 1, et la probabilité $\mathbb{P}\{|X - \mathbb{E}(X)| \ge \lambda\}$ pour deux v.a. centrées réduites : $X \rightsquigarrow \mathcal{N}(0,1)$ et $X = 2Y - 1$, avec $Y \rightsquigarrow \mathcal{B}(1, 0.5)$. Nous voyons que pour λ grand, la majoration est très grossière : pour $X = 2Y - 1$, nous voyons que, sauf en $\lambda = 1^-$ où loi et majorant se confondent, nous voyons que nous majorons 0 par $\frac{1}{\lambda^2}$. Sur le deuxième graphique, pour $X \rightsquigarrow \mathcal{N}(0,1)$, nous avons représenté le rapport du majorant divisé par la fonction qu'il est "chargé de majorer" ; nous λ grand, nous voyons que ce rapport devient considérable. Pour améliorer la lisibilité, nous avons représenté dans le troisième graphique le $\log_{10}$ du rapport précédent (les valeurs représentent donc le rapport comme un exposant de 10). Nous voyons que cette inégalité est à la fois médiocre, mais ne peut être améliorée en toute généralité à cause du cas de v.a. de Bernoulli pour lesquelles la majorations se révèle la pire.

6.3.2 Loi faible et loi forte

L'énoncé 6.3 de la LGN que nous avons donné est encore appelé "loi *faible* des grands nombres". La "loi *forte*" traite de l'ensemble $\Omega_0 = \{\omega \in \Omega \mid \lim_{n\to+\infty} M_n(\omega) = \mu\}$, c'est-à-dire de l'ensemble des



(a) **(b)** **(c)**

FIG. 6.1 – L'inégalité de Bienaymé-Tchebicheff : **(a)** en pointillé : la fonction $\frac{1}{\lambda^2}\mathrm{Var}\,(X)$ pour une loi de variance 1, en ligne continue la probabilité $\mathbb{P}\{|X - \mathbb{E}(X)| \geq \lambda\}$ pour deux v.a. centrées réduites : $X \rightsquigarrow \mathcal{N}(0,1)$ et $X = 2Y - 1$, avec $Y \rightsquigarrow \mathcal{B}(1, 0.5)$. Nous voyons que pour λ grand, la majoration est très grossière. **(b)** Pour $X \rightsquigarrow \mathcal{N}(0,1)$, quotient du majorant $\frac{1}{\lambda^2}\mathrm{Var}\,(X)$ par la valeur exacte $\mathbb{P}\{|X - \mathbb{E}(X)| \geq \lambda\}$. **(c)** $\log_{10}$ du quotient précédent.

états du monde pour lesquels la moyenne $M_n(\omega)$ tend effectivement vers l'espérance commune μ des v.a. X_i. Notons que pour avoir une suite infinie de v.a. indépendantes, Ω doit nécessairement être infini, et le plus simple pour avoir une simple suite de v.a. de Bernouilli, $\Omega = \{0,1\}^{\mathbb{N}^*}$, n'est pas dénombrable. Nous quittons donc le cadre élémentaire que nous nous sommes fixé pour ce cours. Indiquons néanmoins le résultat de la loi forte des grands nombres : le sous-ensemble Ω_0 ci-dessus est de probabilité égale à 1 ; en d'autres termes, les ω pour lesquels on n'a pas la convergence souhaitée forment un *ensemble négligeable*, c'est-à-dire de probabilité nulle. On dit aussi que la convergence $\lim_{n\to+\infty} M_n = \mu$ est *presque-sûre*, et on note $Y_n \xrightarrow{ps} \overline{Y}$ si la convergence $\lim_{n\to+\infty}(Y_n - \overline{Y}) = 0$ est presque-sûre.

6.3.3 Convergence en probabilité

L'énoncé 6.3 de la loi des grands nombres que nous avons donné s'exprime encore par la locution “la suite des M_n *tend en probabilité* vers le nombre μ”. De façon générale, on dit que la suite de v.a. $(Y_n)_{n\geq 1}$ *tend en probabilité vers la v.a.* $\overline{Y}$, et on note $Y_n \xrightarrow{\mathcal{P}} \overline{Y}$ si et seulement si

$$\text{pour pour } \varepsilon > 0, \ \lim_{n\to+\infty} \mathbb{P}(\{|Y_n - \overline{Y}| \geq \varepsilon\}) = 0 \ ;$$

en paraphrasant : quand n devient grand, il est de plus en plus improbable que $Y_n(\omega)$ s'écarte de $\overline{Y}(\omega)$ de plus de ε. On montre que la convergence presque-sûre implique toujours la convergence en probabilité.

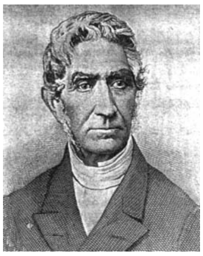
6.3.4 Et notre norme L^2 ?

Nous avons vu que $X \mapsto \|X\| := (\mathbb{E}(X^2))^{\frac{1}{2}}$ est une norme sur $L^2(\Omega)$. Nous avons donc la notion de convergence usuelle dans un espace normé, celle qui assure que $\lim_{n\to+\infty} \|Y_n - \overline{Y}\| = 0$: on dit dans ce cas que Y_n *tends vers* $\overline{Y}$ *dans* $L^2(\Omega)$, et on note $Y_n \xrightarrow{L^2} \overline{Y}$.

On montre que $Y_n \xrightarrow{L^2} \overline{Y}$ implique que $Y_n \xrightarrow{\mathcal{P}} \overline{Y}$. Par ailleurs, on montre que $Y_n \xrightarrow{ps} \overline{Y}$ implique que $Y_n \xrightarrow{L^2} \overline{Y}$. Nous voyons donc que la *convergence en probabilité* “$\xrightarrow{\mathcal{P}}$” est la plus faible de ces notions de convergence ; toutefois on montre aussi que si $Y_n \xrightarrow{\mathcal{P}} \overline{Y}$, alors il existe une sous-suite $k \mapsto n_k$ telle que $Y_{n_k} \xrightarrow{ps} \overline{Y}$.

Jacques Bernoulli (1654-1705) : 

Irénée-Jules Bienaymé (1796-1878) : 

Lambert Adolphe Jacques Quetelet (1796-1874) 

Pafnuty Lvovich Tchebicheff [Chebyshev] (1821-1894) : 

Andrei Andreyevich Markov (1856-1922) :