Théorèmes Limites

Pour finir ce cours nous allons donner deux exemples de théorèmes limites pour des suites de v.a. réelles indépendantes : la loi faible des grands nombres et le théorème de la limite centrale ou théorème « central limit », TCL en abrégé dans la suite.

1. Loi des grands nombres.

Nous commençons par deux inégalités très classiques du calcul des probabilités ; elles sont d'usage fréquent.

1.1. Inégalités de Markov et Bienaymé–Tchebychev.

Tout d'abord l'inégalité de Markov dont la démonstration est d'un simplicité enfantine.

Proposition 1. *Soit X une v.a.r. positive et intégrable. Alors,*

$$\forall a > 0, \qquad a\,\mathbb{P}(X \geq a) \leq \mathbb{E}[X].$$

La démonstration de cette inégalité est élémentaire. Il suffit de remarquer que, comme X est une v.a.r. positive,

$$X = X\,\mathbf{1}_{X<a} + X\,\mathbf{1}_{X\geq a} \geq X\,\mathbf{1}_{X\geq a} \geq a\,\mathbf{1}_{X\geq a},$$

et d'utiliser la croissance de l'espérance, pour obtenir

$$\mathbb{E}[X] \geq a\,\mathbb{E}\left[\mathbf{1}_{X\geq a}\right] = a\,\mathbb{P}(X \geq a).$$

L'inégalité ci-dessus est vraie pour $a = 0$ mais ne présente dans ce cas aucun intérêt.

Remarque(s). Notons que si X est une v.a.r. positive et $r > 0$, alors $X(\omega) \geq a$ si et seulement si $X^r(\omega) \geq a^r$, et ce pour tout $a > 0$. En particulier, si X est une v.a.r. positive qui posséde un moment d'ordre r, on a

$$\forall a > 0, \qquad a^r\,\mathbb{P}(X \geq a) = a^r\,\mathbb{P}\left(X^r \geq a^r\right) \leq \mathbb{E}\left[X^r\right].$$

Cette remarque très simple conduit à une seconde inégalité connue sous le nom d'inégalité de Bienaymé–Tchebychev.

Proposition 2. *Soient $a > 0$ et X une v.a.r. de carré intégrable. Alors,*

$$\mathbb{P}\left(|X - \mathbb{E}[X]| \geq a\right) \leq \frac{\mathbb{V}[X]}{a^2}.$$

Il suffit d'appliquer la remarque précédente à la variable $Y = |X - \mathbb{E}[X]|$ avec $r = 2$. En effet, pour tout $a > 0$,

$$a^2\,\mathbb{P}(Y \geq a) \leq \mathbb{E}\left[Y^2\right] = \mathbb{V}[X].$$

Remarque(s). Ces inégalités sont valables indépendamment de la loi de la v.a.r. X. Il n'est donc pas très surprenant qu'elles ne soient pas extrèmement précises comme on peut s'en convaincre sur l'exemple suivant. Soient X de loi uniforme sur $[0, 1]$ et $a = 1$; $\mathbb{E}[X] = 1/2$ et donc l'inégalité de Markov donne $0 = \mathbb{P}(X \geq 1) \leq 1/2$, qui n'est pas optimal !

1.2. Convergence en moyenne quadratique.

Nous allons nous intéresser à la convergence d'une suite de v.a.r. $(X_n)_{n\in\mathbf{N}}$. Nous devons préciser en quel sens on doit comprendre cette convergence.

Définition. Soient $(X_n)_{n\in\mathbf{N}}$ une suite de v.a.r. de carré intégrable et X une v.a.r., toutes définies sur le même espace de probabilité. $(X_n)_{n\in\mathbf{N}}$ converge vers X *en moyenne quadratique* si :

$$\lim_{n\to+\infty} \mathbb{E}\left[|X_n - X|^2\right] = 0.$$

Il existe différentes notions de convergence pour les suite variables aléatoires réelles. Définissons un autre mode de convergence : la convergence en probabilité.

Définition. Soient $(X_n)_{n\in\mathbf{N}}$ une suite de v.a.r., X une v.a.r., toutes définies sur le même espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$. On dit que $(X_n)_{n\in\mathbf{N}}$ converge vers X *en probabilité* si :

$$\forall\varepsilon > 0, \qquad \lim_{n\to+\infty} \mathbb{P}\left(|X_n - X| \geq \varepsilon\right) = 0.$$

Remarque(s). Un moyen d'établir la convergence en probabilité de $(X_n)_{n\in\mathbf{N}}$ vers X consiste à montrer que, pour un $r > 0$, $\mathbb{E}\left[|X_n - X|^r\right]$ converge vers 0 et d'utiliser l'inégalité de Markov puisque

$$\forall\varepsilon > 0, \qquad \mathbb{P}\left(|X_n - X| \geq \varepsilon\right) \leq \mathbb{E}\left[|X_n - X|^r\right]/\varepsilon^r.$$

En particulier, la convergence en moyenne quadratique entraîne la convergence en probabilité.

Exemple. Pour illustrer la notion de convergence en probabilité, considérons X une v.a.r. uniforme sur $[0, 1]$ et, pour $n \in \mathbf{N}^*$, $X_n = X + n^2\mathbf{1}_{X\leq 1/n}$. Alors, la suite $(X_n)_{n\in\mathbf{N}^*}$ converge vers X en probabilité. En effet, si $0 < \varepsilon < 1$, alors

$$\mathbb{P}\left(|X_n - X| \geq \varepsilon\right) = \mathbb{P}(X \leq 1/n) = 1/n.$$

Par contre, la suite $(X_n)_{n\in\mathbf{N}^*}$ ne converge pas vers X en moyenne quadratique puisque

$$\mathbb{E}\left[|X_n - X|^2\right] = n^4\,\mathbb{P}(X \leq 1/n) = n^3.$$

1.3. Loi faible des grands nombres.

Imaginons un instant que l'on lance une pièce non truquée un grand nombre de fois, disons n fois. On note P_n le nombre de fois où « pile » est apparu. Intuitivement, si n est grand la fréquence d'apparition de pile, P_n/n, va être proche de $1/2$. Le théorème suivant, la loi faible des grands nombres, confirme cette intuition.

Théorème 3. *Soit $(X_n)_{n\in\mathbf{N}^*}$ une suite de v.a.r. indépendantes, de même loi, de carré intégrable et définies sur le même espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$. On note, pour tout $n \geq 1$, $S_n = X_1 + \ldots + X_n$. Alors, avec $\mu = \mathbb{E}[X_1]$,*

$$\frac{S_n}{n} = \frac{1}{n}(X_1 + \ldots + X_n) \longrightarrow \mu, \qquad \textit{en moyenne quadratique,} \quad \textit{quand } n \to \infty.$$

Remarque(s). Le résultat reste vrai si on suppose seulement que les v.a.r. sont 2 à 2 non-corrélées et possèdent la même moyenne et la même variance au lieu de les supposer indépendantes et de même loi.

La démonstration est relativement facile : comme les v.a.r. sont indépendantes et de même loi – c'est pareil sous les hypothèses de la remarque précédente – on a, notant σ^2 la variance de X_1,

$$\mathbb{E}\left[n^{-1}S_n\right] = \mu, \qquad \mathbb{E}\left[\left|n^{-1}S_n - \mu\right|^2\right] = \mathbb{V}\left[n^{-1}S_n\right] = \frac{\sigma^2}{n}\ ;$$

le résultat s'en suit immédiatement.

2. Théorème de la limite centrale.

Nous avons vu au paragraphe précédent deux types de convergence pour les suites de v.a.r. : la convergence en moyenne quadratique et la convergence en probabilité. Toutefois pour énoncer le théorème de la limite centrale nous aurons besoin d'une notion plus faible de convergence.

2.1. Convergence en loi.

Pour ce type de convergence, les variables aléatoires peuvent être définies sur des espaces de probabilité différents.

Définition. Une suite de v.a.r. $(X_n)_{n\in\mathbf{N}}$ *converge en loi* vers une v.a.r. X si, pour toute fonction $f : \mathbf{R} \longrightarrow \mathbf{R}$, continue et bornée,

$$\mathbb{E}[f(X_n)] \longrightarrow \mathbb{E}[f(X)], \quad \text{quand } n \to +\infty.$$

Remarque(s). Comme déjà dit, les v.a.r. X_n peuvent être définies sur des espaces de probabilité différents puisque les v.a. n'interviennent qu'au travers d'espérances.

Dans la définition, la continuité de la fonction f est importante. En effet, si X_n est une v.a. prenant les valeurs $1/n$ et 1 avec probabilité 1/2 alors X_n converge en loi vers X où X prend les valeurs 0 et 1 avec probabilité 1/2 puisque, si f est continue bornée,

$$\mathbb{E}[f(X_n)] = (f(1/n) + f(1))/2 \longrightarrow (f(0) + f(1))/2 = \mathbb{E}[f(X)].$$

Pourtant, si on prend $f = \mathbf{1}_{\{0\}}$ qui n'est pas continue au point 0,

$$\mathbb{E}[f(X_n)] = \mathbb{P}(X_n = 0) = 0, \qquad \mathbb{E}[f(X)] = \mathbb{P}(X = 0) = 1/2.$$

En général, si X_n converge en loi vers X, on ne peut pas dire que $\mathbb{P}(X_n \in A)$ converge vers $\mathbb{P}(X \in A)$ car $x \longmapsto \mathbf{1}_A(x)$ n'est pas une fonction continue.

Exemple. Soit X_n une v.a. de loi uniforme sur $\{i/n,\ i = 0, \ldots, n-1\}$. Alors X_n converge en loi vers X de loi uniforme sur $[0, 1]$. En effet, si f est une fonction continue et bornée, on a

$$\mathbb{E}[f(X_n)] = \frac{1}{n}\sum_{i=0}^{n-1} f(i/n) \longrightarrow \int_0^1 f(x)dx = \mathbb{E}[f(X)],$$

puisque $\mathbb{E}[f(X_n)]$ est la somme de Riemann de f associée à la subdivision de pas $1/n$ sur $[0, 1]$.

Dans cet exemple, les v.a.r. X_n sont discrètes et la v.a.r. X est absolument continue. Si toutes les v.a.r. sont discrètes et à valeurs dans le même ensemble discret on peut obtenir une caractérisation simple de la convergence en loi.

Proposition 4. *Soient X_n, $n \in \mathbf{N}$, et X des v.a.r. à valeurs dans $\mathbf{N}$. Alors X_n converge en loi vers X si et seulement si, pour tout $k \in \mathbf{N}$, $\mathbb{P}(X_n = k) \longrightarrow \mathbb{P}(X = k)$ quand $n \to +\infty$.*

Rappelons que, dans ce cas, si f est continue bornée,

$$\mathbb{E}[f(X_n)] = \sum_{i\geq 0} f(i)\mathbb{P}(X_n = i), \qquad \mathbb{E}[f(X)] = \sum_{i\geq 0} f(i)\mathbb{P}(X = i)\ ;$$

on montre que la condition est nécessaire en considérant, pour tout entier k, la fonction continue $f(x) = (1 - 2|x-k|)^+$ qui vérifie $f(p) = 0$ pour tout $p \in \mathbf{N}$ tel que $p \neq k$. La condition est suffisante par passage à la limite dans $\mathbb{E}[f(X_n)]$, f étant bornée. En effet, si on fixe $0 < \varepsilon < 1/2$, il existe un entier m tel que $\sum_{i=0}^m \mathbb{P}(X = i) \geq 1 - \varepsilon$, puisque $\sum_{i\geq 0} \mathbb{P}(X = i) = 1$. Comme $\mathbb{P}(X_n = i) \longrightarrow \mathbb{P}(X = i)$ pour tout i on en déduit que, si n est suffisamment grand disons $n \geq p$, $\sum_{i=0}^m \mathbb{P}(X_n = i) \geq 1 - 2\varepsilon$. On a donc, pour tout entier n,

$$\mathbb{E}[f(X_n)] - \mathbb{E}[f(X)] = \sum_{i>m} f(i)\left\{\mathbb{P}(X_n = i) - \mathbb{P}(X = i)\right\} + \sum_{i=0}^{m} f(i)\left\{\mathbb{P}(X_n = i) - \mathbb{P}(X = i)\right\},$$

et, si $n \geq p$, l'inégalité triangulaire donne

$$\left|\mathbb{E}[f(X_n)] - \mathbb{E}[f(X)]\right| \leq 3\varepsilon\|f\|_\infty + \|f\|_\infty \max_{0\leq i\leq m} |\mathbb{P}(X_n = i) - \mathbb{P}(X = i)|,$$

où $\|f\|_\infty = \sup_{i\in\mathbf{N}} |f(i)|$. Cette dernière inégalité permet de conclure puisque m est fini.

Exemple. On suppose que X_n suit la loi $\mathcal{B}(n, p_n)$. Si $np_n \longrightarrow \lambda > 0$, alors X_n converge en loi vers une v.a. X suivant une loi de Poisson $\mathcal{P}(\lambda)$. Appliquons le critère précédent. On a, si $k \leq n$, $\mathbb{P}(X_n = k) = C_n^k (p_n)^k (1-p_n)^{n-k}$. Tout d'abord $\mathbb{P}(X_n = 0) = (1-p_n)^n = \exp\left\{n\ln(1-p_n)\right\}$. Comme $\lim_{x\longrightarrow 0} \ln(1-x)/x = -1$, on a comme $p_n \longrightarrow 0$, $\mathbb{P}(X_n = 0) \longrightarrow e^{-\lambda} = \mathbb{P}(X = 0)$. Un calcul facile conduit à, pour $n \geq k+1$,

$$\frac{\mathbb{P}(X_n = k+1)}{\mathbb{P}(X_n = k)} = \frac{(n-k)p_n}{(k+1)(1-p_n)} = \frac{1 - kn^{-1}}{1 - p_n}\,\frac{np_n}{k+1} \longrightarrow \frac{\lambda}{k+1}.$$

Il vient alors, pour tout $k \in \mathbf{N}$, $\mathbb{P}(X_n = k) \longrightarrow e^{-\lambda}\lambda^k/k!$.

Nous admettrons le résultat suivant qui donne des caractérisations de la convergences en loi.

Théorème 5. *Soient $(X_n)_{n\in\mathbf{N}}$ une suite de v.a.r. et X une v.a.r. Alors les propositions suivantes sont équivalentes :*

- *$(X_n)_{n\in\mathbf{N}}$ converge en loi vers X ;*
- *pour tout $t \in \mathbf{R}$, $\varphi_{X_n}(t) \longrightarrow \varphi_X(t)$;*
- *pour tout $x \in \mathbf{R}$ où F_X est continue, $F_{X_n}(x) \longrightarrow F_X(x)$.*

De plus, si $(X_n)_{n\in\mathbf{N}}$ converge vers X en probabilité alors $(X_n)_{n\in\mathbf{N}}$ converge vers X en loi.

Exemple. Soit Y_n une v.a. de loi géométrique λ/n, $\lambda > 0$. Montrons que $X_n = Y_n/n$ converge en loi vers une v.a. X de loi $\mathcal{E}xp(\lambda)$. Soient φ_n la fonctions caractéristique de X_n et φ celle de X. On a

$$\forall t \in \mathbf{R}, \qquad \varphi_n(t) = \frac{(\lambda/n)\exp(it/n)}{1-(1-\lambda/n)\exp(it/n)}, \qquad \varphi(t) = \frac{\lambda}{\lambda - it}.$$

Comme $\lim_{z\to 0}\left(e^z - 1\right)/z = 1$, on a, pour tout $t \in \mathbf{R}$,

$$\varphi_n(t) = \frac{\lambda}{n\left(\exp(-it/n)-1\right)+\lambda} \longrightarrow \frac{\lambda}{\lambda - it} = \varphi(t).$$

Exercice. Reprendre l'exemple de la page 4 à l'aide du critère sur les fonctions caratéristiques.

2.2. Le TCL.

Nous démontrons maintenant un résultat important de la théorie des probabilités : le théorème de la limite centrale.

Théorème 6. *Soit $(X_n)_{n\in\mathbf{N}}$ une suite de v.a.r. indépendantes, de même loi, de carré intégrable. On note $\mu = \mathbb{E}[X_1]$ et $\sigma^2 = \mathbb{V}[X_1]$; on suppose que $\sigma^2 > 0$. Alors,*

$$\frac{\sqrt{n}}{\sigma}\left(\frac{S_n}{n} - \mu\right) \longrightarrow X, \qquad \textit{en loi,} \quad \textit{quand } n \to +\infty,$$

où X est une v.a.r. de loi $\mathcal{N}(0,1)$ et $S_n = X_1 + \ldots + X_n$.

Remarque(s). Comme la loi limite possède une fonction de répartition continue, on a, pour tout intervalle $]a,b[$,

$$\mathbb{P}\left(a < \sqrt{n}\left(n^{-1}S_n - \mu\right)/\sigma < b)\right) \longrightarrow \frac{1}{\sqrt{2\pi}}\int_a^b \exp(-u^2/2)\,du.$$

La loi des grands nombres nous dit que $n^{-1}S_n$ converge vers μ en moyenne quadratique. Le TCL montre que la vitesse de convergence est en $1/\sqrt{n}$. De plus, si n est suffisamment grand, $\sqrt{n}(n^{-1}S_n - \mu)/\sigma$ se comporte comme une v.a.r. normale centrée réduite.

On a en particulier, pour n assez grand,

$$\mathbb{P}\left(\left|n^{-1}S_n - \mu\right| \geq \alpha\right) = \mathbb{P}\left(\left|\sqrt{n}\left(n^{-1}S_n - \mu\right)/\sigma\right| \geq \alpha\sqrt{n}/\sigma\right) \simeq 2\left(1 - \Phi\left(\alpha\sqrt{n}/\sigma\right)\right),$$

où Φ est la fonction de répartition de la loi normale centrée réduite.

Connaissant deux des nombres n, α et ε, il est possible au moyen de cette approximation de déterminer le troisième pour que

$$\mathbb{P}\left(\left|n^{-1}S_n - \mu\right| \geq \alpha\right) \leq \varepsilon.$$

Le TCL se démontre en utilisant les fonctions caractéristiques. On note Y_k la variable aléatoire $(X_k - \mu)/\sigma$. On a donc $\mathbb{E}[Y_k] = 0$ et $\mathbb{V}[Y_k] = 1$ et les Y_k demeurent indépendantes. Avec ces notations, on doit montrer que

$$R_n = \frac{1}{\sqrt{n}}\sum_{k=1}^{n} Y_k \longrightarrow X, \qquad \text{en loi quand } n \to +\infty$$

où X suit la loi normale centrée réduite. Déterminons la fonction caractéristique de la v.a. R_n, φ_n : par indépendance, on a, notant φ la fonction caractéristique de Y_1,

$$\forall t \in \mathbf{R}, \qquad \varphi_n(t) = \varphi\left(t/\sqrt{n}\right)^n \ ;$$

comme Y_1 est de carré intégrable, centrée réduite, sa fonction caractéristique est de classe C^2 et on a $\varphi'(0) = i\,\mathbb{E}[Y_1] = 0$ et $\varphi''(0) = i^2\,\mathbb{E}[Y_1^2] = -1$. φ possède donc un DL à l'ordre 2 en 0 de la forme

$$\varphi(t) = 1 - t^2/2 + t^2\varepsilon(t), \qquad \text{avec } \lim_{t\to 0}\varepsilon(t) = 0.$$

Soit $t \in \mathbf{R}$ fixé. Puisque $|z^n - w^n| \leq n|z-w|$ si $|z| \leq 1$ et $|w| \leq 1$, pour n assez grand, $1 - t^2/(2n)$ appartient à $[0,1]$, et on a comme $|\varphi(t/\sqrt{n})| \leq 1$,

$$\left|\varphi_n(t) - \left(1 - \frac{t^2}{2n}\right)^n\right| = \left|\left(1 - \frac{t^2}{2n} + \frac{t^2}{n}\varepsilon\left(\frac{t}{\sqrt{n}}\right)\right)^n - \left(1 - \frac{t^2}{2n}\right)^n\right| \leq n\frac{t^2}{n}\left|\varepsilon\left(\frac{t}{\sqrt{n}}\right)\right|.$$

Comme $\ln(1+z)/z \longrightarrow 1$ si $z \longrightarrow 0$, on obtient

$$\lim_{n\to+\infty}\left(1 - \frac{t^2}{2n}\right)^n = \lim_{n\to+\infty}\exp\left\{n\ln\left(1 - \frac{t^2}{2n}\right)\right\} = \exp(-t^2/2),$$

ce qui entraîne, puisque $\varepsilon(t/\sqrt{n})$ tend vers 0 si $n \to +\infty$,

$$\forall t \in \mathbf{R}, \qquad \lim_{n\to+\infty}\varphi_n(t) = \exp(-t^2/2),$$

ce qui démontre le TCL via le Théorème 5.