Convergence en loi. Théorème de la limite centrale.

Convergence en loi

1. Soient $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires réelles définies sur un espace $(\Omega, \mathscr{F}, \mathbb{P})$ et f une application continue de $\mathbb{R}$ dans $\mathbb{R}$. On suppose que $(X_n)_{n\in\mathbb{N}}$ converge en loi vers une variable aléatoire X.

Montrer que la suite $(f(X_n))_{n\in\mathbb{N}}$ converge en loi vers $f(X)$.

Solution de l'exercice 1. Tout d'abord, $f(X_n)$ est bien une variable aléatoire comme composée de la variable aléatoire X_n et de l'application f qui est continue. Si φ est une application de $\mathbb{R}$ dans $\mathbb{R}$ continue bornée, remarquons que $\varphi \circ f$ est également continue bornée. L'hypothèse de convergence en loi de la suite $(X_n)_{n\in\mathbb{N}}$ vers X implique $\lim \mathbb{E}[\varphi \circ f(X_n)] = \mathbb{E}[\varphi \circ f(X)]$ que l'on peut écrire $\lim \mathbb{E}[\varphi(f(X_n))] = \mathbb{E}[\varphi(f(X))]$. Ceci étant valable pour toute φ continue bornée, $(f(X_n))_{n\in\mathbb{N}}$ converge en loi vers $f(X)$.

2. a. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires réelles qui converge en loi vers une variable aléatoire réelle constante a. Montrer que la convergence a lieu aussi en probabilité.

b. Soit $(X_n)_{n\geq 1}$ une suite indépendante de variables aléatoires réelles de même loi de Cauchy de paramètre 1. Soit $S_n = \sum_{k=1}^n X_k$. Etudier les convergences en probabilité et en loi des suites $(\frac{1}{\sqrt{n}} S_n)_{n\geq 1}$, $(\frac{1}{n} S_n)_{n\geq 1}$ et $(\frac{1}{n^2} S_n)_{n\geq 1}$.

Indication : la fonction caractéristique de la loi de Cauchy est donnée par $\phi(t) = e^{-|t|}$.

Solution de l'exercice 2. a. Soit $\varepsilon > 0$ et f_ε définie par $f_\varepsilon(x) = \mathbb{1}_{]a-\varepsilon, a+\varepsilon[^c}(x) + \frac{1}{\varepsilon}|x - a|\mathbb{1}_{]a-\varepsilon,a+\varepsilon[}$. Cette fonction est continue bornée, donc la suite $(\mathbb{E}[f_\varepsilon(X_n)])_{n\geq 1}$ converge vers $\mathbb{E}[f_\varepsilon(a)] = 0$. Comme

$$\mathbb{P}(|X_n - a \geq \varepsilon|) = \mathbb{E}[\mathbb{1}_{]a-\varepsilon,a+\varepsilon[^c}(X_n)] \leq \mathbb{E}[f_\varepsilon(X_n)],$$

$(X_n)_{n\geq 1}$ converge en probabilité vers a.

b. La fonction caractéristique de $\frac{1}{\sqrt{n}} S_n$ est donnée par

$$\phi(t) = \mathbb{E}[e^{\frac{it}{\sqrt{n}} S_n}] = \left(\mathbb{E}[e^{\frac{it}{\sqrt{n}} X_1}]\right)^n = e^{-|t|\sqrt{n}},$$

car les X_n sont indépendantes. Comme la suite $\left(e^{-|t|\sqrt{n}}\right)_{n\geq 1}$ tend vers $1_{\{0\}}(t)$ qui définit une application non continue en 0. Donc la limite des fonctions caractéristiques n'est pas une fonction caractéristique, et $(\frac{1}{\sqrt{n}} S_n)_{n\geq 1}$ ne converge ni en loi, ni en probabilité.

La fonction caractéristique de $\frac{1}{n}S_n$ est $\phi(t) = e^{-|t|}$. La suite $(\frac{1}{n}S_n)_{n\geq 1}$ converge donc en loi vers une variable aléatoire de loi de Cauchy. Mais elle ne converge pas en probabilité. En effet, sinon, la suite $(\frac{S_n}{n} - \frac{S_{2n}}{2n})_{n\geq 1}$ convergerait en probabilité, donc aussi en loi, vers 0. Or, la fonction caractéristique de $\frac{S_n}{n} - \frac{S_{2n}}{2n}$ est donnée par $\phi(t) = \mathbb{E}[e^{\frac{it}{2n}(X_1+\cdots+X_n-(X_{n+1+\cdots+X_{2n}}))}] = e^{-|t|}$, qui ne converge pas vers 1, fonction caractéristique de la variable aléatoire 0.

La fonction caractéristique de $\frac{1}{n^2}S_n$ est $\phi(t) = e^{-\frac{|t|}{n}}$ qui converge vers 1. Ainsi, la suite $(\frac{1}{n^2}S_n)_{n\geq 1}$ converge en loi et donc en probabilité vers 0.

3. Soient X une variable aléatoire réelle, $(X_n)_{n\in\mathbb{N}}$ et $(Y_n)_{n\in\mathbb{N}}$ deux suites de v.a.r.

a. Montrer que pour tout $t \in \mathbb{R}$, $a > 0$ et $n \in \mathbb{N}$,

$$|\phi_{X_n+Y_n}(t) - \phi_{X_n}(t)| \leq 2\mathbb{P}(|Y_n| > a) + \mathbb{E}[\mathbb{1}_{]-\infty,a]}(|Y_n|)|e^{itY_n} - 1|].$$

b. Montrer que si $(X_n)_{n\in\mathbb{N}}$ converge en loi vers X et $(Y_n)_{n\in\mathbb{N}}$ converge en loi vers 0, alors la suite $(X_n + Y_n)_{n\in\mathbb{N}}$ converge en loi vers X.

c. Montrer que la convergence en loi de $(X_n)_{n\in\mathbb{N}}$ vers X n'implique pas la convergence en loi de $(X_n - X)_{n\in\mathbb{N}}$ vers 0.

Solution de l'exercice 3.

a) Pour tous $t \in \mathbb{R}$, $a > 0$ et $n \in \mathbb{N}$,

$$\begin{aligned}
|\phi_{X_n+Y_n}(t) - \phi_{X_n}(t)| &= |\mathbb{E}[e^{itX_n}(e^{itY_n} - 1)]| \leq \mathbb{E}[|e^{itY_n} - 1|] \\
&= \int_{\{|Y_n|>a\}} |e^{itY_n} - 1| d\mathbb{P} + \int_{\{|Y_n|\leq a\}} |e^{itY_n} - 1| d\mathbb{P} \\
&\leq 2\mathbb{P}(|Y_n| > a) + \int_{\{|Y_n|\leq a\}} |e^{itY_n} - 1| d\mathbb{P}
\end{aligned}$$

b) En fait, Y_n converge en probabilité vers 0. Soit $\varepsilon > 0$. Il existe $a_0 > 0$ tel que pour $y \leq a_0$, $|e^{it} - 1| \leq \varepsilon$. De plus, il existe $n_0 \in \mathbb{N}$ tel que pour tout $n \geq n_0$, $\mathbb{P}(|Y_n| > a_0) \leq \varepsilon$. Donc, $|\phi_{X_n+Y_n}(t) - \phi_{X_n}(t)| \leq 3\varepsilon$. La convergence en loi de $(X_n)_{n\in\mathbb{N}}$ vers X implique qu'il existe n_1 (que l'on peut prendre plus grand que n_0), tel que, pour tout $n \geq n_1$, $|\phi_{X_n}(t) - \phi_X(t)| \leq \varepsilon$. On a montré que pour tout $n \geq n_1$, $|\phi_{X_n+Y_n}(t) - \phi_X(t)| \leq 4\varepsilon$ d'où la convergence en loi demandée.

Pour éviter les ϵ on peut utiliser la notion de limite supérieure. Pour tout a positif on a :

$$\begin{aligned}
| \phi_{X_n+Y_n}(t) - \phi_X(t) | &\leq | \phi_{X_n+Y_n}(t) - \phi_{X_n}(t) | + | \phi_{X_n}(t) - \phi_X(t) | \\
&\leq 2\mathbb{P}[| Y_n | \geq a] + \mathbb{E}[\mathbb{1}_{[0,a]}(| Y_n |) | e^{itY_n} - 1 |] + | \phi_{X_n}(t) - \phi_X(t) | \\
&\leq 2\mathbb{P}[| Y_n | \geq a] + \sup_{x\in[0,a]} | e^{itx} - 1 | + | \phi_{X_n}(t) - \phi_X(t) |
\end{aligned}$$

or la limite supérieur d'une somme est plus petite que la somme des limites supérieures et si une suite converge sa limite supérieure est égale à sa limite donc pour tout réel a positif :

$$\begin{aligned}\overline{\lim}_n \mid \phi_{X_n+Y_n}(t) - \phi_X(t) \mid &\leq 2\overline{\lim}_n \mathbb{P}[\mid Y_n \mid \geq a] + \sup_{x\in[0,a]} \mid e^{itx} - 1 \mid + \overline{\lim}_n \mid \phi_{X_n}(t) - \phi_X(t) \mid \\ &= \sup_{x\in[0,a]} \mid e^{itx} - 1 \mid,\end{aligned}$$

la dernière égalité provenant de la convergence en loi de X_n vers X et de la convergence en probabilité de Y_n vers 0. Ainsi pour tout a réel positif :

$$\overline{\lim}_n \mid \phi_{X_n+Y_n}(t) - \phi_X(t) \mid \leq \sup_{x\in[0,a]} \mid e^{itx} - 1 \mid,$$

en faisant tendre a vers 0 on obtient donc $\overline{\lim}_n \mid \phi_{X_n+Y_n}(t) - \phi_X(t) \mid = 0$.
Or $(\mid \phi_{X_n+Y_n}(t) - \phi_X(t) \mid)_n$ est une suite de termes positifs donc elle est convergente et converge vers 0.

c) On prend X de loi symétrique (X a la même loi que $-X$), et on pose $X_n = -X$. Alors X_n converge en loi vers X, mais $X_n - X$ converge en loi vers $-2X$.

Applications du TCL

4. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes de même loi de Poisson de paramètre 1. Soit $S_n = \sum_{k=1}^n X_k$. Rappeler la loi de S_n et calculer la limite de la suite $\left(e^{-n}\sum_{k=0}^n \frac{n^k}{k!}\right)_{n\geq 1}$.

Solution de l'exercice 4. Une somme de v.a. de Poisson indépendantes est une v.a. de Poisson dont le paramètre (qui est aussi l'espérance et la variance) est la somme de ceux des v.a. qu'on a ajoutées. Ainsi S_n suit une loi de Poisson de paramètre n. En particulier, si $k \in \mathbb{N}$, alors $e^{-n}\frac{n^k}{k!} = \mathbb{P}(S_n = k)$. En sommant, on obtient

$$\left(e^{-n}\sum_{k=0}^n \frac{n^k}{k!}\right) = \mathbb{P}(S_n \leq n) = \mathbb{P}(S_n/n \leq 1).$$

La loi des grands nombres nous dit que $S_n/n \to \mathbb{E}[X_1] = 1$ p.s. Ainsi S_n/n est proche de 1, mais ce résultat ne nous dit pas s'il est un peu plus grand ou un peu plus petit. Pour avoir des informations sur $S_n/n - 1$, (et en particulier son signe), on applique le théorème de la limite centrale :

$$\frac{S_n - n}{\sqrt{n}} \xrightarrow[n\to\infty]{\text{loi}} \mathcal{N}(0,1).$$

D'après le corrolaire 4.2.1 du cours, on en déduit que, si Y suit une loi normale standard, alors, lorsque $n \to \infty$,

$$\mathbb{P}(S_n/n \leq 1) = \mathbb{P}(\frac{S_n - n}{\sqrt{n}} \leq 0) \to \mathbb{P}(Y \leq 0) = 1/2.$$

L'égalité $\mathbb{P}(Y \leq 0) = 1/2$ découle du fait que Y est symétrique.

Ainsi, on a montré que

$$\left(e^{-n}\sum_{k=0}^{n}\frac{n^k}{k!}\right) \to 1/2.$$

5. a. Soit $(p_n)_{n\geq 0}$ une suite de réels dans $]0,1[$ telle que $\lim_{n\to+\infty} np_n = \lambda > 0$. Soit $(X_n)_{n\geq 0}$ une suite de v.a. telles que pour tout n : $X_n \sim \mathcal{B}(n,p_n)$ et X une v.a. de loi de Poisson paramètre λ. Montrer que $(X_n)_{n\geq 0}$ converge en loi vers X.

b. Soit $(X_n)_{n\geq 0}$ une suite de v.a. réelles indépendantes de même loi $\mathcal{N}(0,1)$. Etudier le comportement asymptotique en loi de la suite $Y_n = \frac{1}{n}\sum_{k=1}^{n}\sqrt{k}X_k$.

Solution de l'exercice 5. a. Afin de montrer la convergence en loi de $(X_n)_{n\geq 0}$ vers X. Montrons que $\phi_{X_n}(t) \xrightarrow[n\to+\infty]{} \phi_X(t)$ pour tout $t \in \mathbb{R}$, où ϕ_X designe la fonction caractéristique d'une variable aléatoire X.

$$\begin{aligned}
\phi_{X_n}(t) &= \left(1 - p_n + p_n e^i t\right)^n \\
&= \exp\left(n\ln(1 - p_n + p_n e^i t)\right) \\
&= \exp\left(-np_n(1 - e^i t) + np_n\epsilon_n\right) \quad \text{où } \epsilon_n \xrightarrow[n\to+\infty]{} 0 \\
&\sim_n \exp\left(-np_n(1 - e^i t)\right) \\
&\xrightarrow[n\to+\infty]{} e^{\lambda(e^{it}-1)} = \phi_X(t).
\end{aligned}$$

b. Vu la forme de la variable de Y_n, on est tenté d'appliquer la loi des grands nombres. Cependant on ne peut pas l'appliquer car les variables $\sqrt{k}X_k$ ne sont pas identiquement distribuées. On va donc à nouveau utiliser les fonctions caractéristiques afin de démontrer un résultat de convergence en loi. Les variables aléatoires X_n sont indépendantes, donc Y_n est une variable gaussienne de moyenne 0 et de variance $\frac{n(n+1)}{2n^2}$ donc :

$$\phi_{Y_n}(t) = e^{-\frac{n(n+1)}{2n^2}t^2} \xrightarrow[n\to+\infty]{} e^{-t^2/2}.$$

On a donc montré que $(Y_n)_{n\geq 1}$ converge en loi vers une loi normale centrée réduite $\mathcal{N}(0,\frac{1}{2})$.

6. On suppose que l'intervalle de temps entre deux voitures successives à un passage à niveau (peu fréquenté) suit une loi exponentielle de moyenne 30 minutes. On suppose de plus qu'il y a indépendance entre les intervalles de temps séparant les instants de passage de voitures. Calculer (une valeur approchée) de la probabilité qu'il y ait plus de 50 voitures qui empruntent le passage à niveau une journée donnée.

Solution de l'exercice 6. Soit $(X_n)_{n\geq 0}$ une suite de v.a. i.i.d suivant la loi exponentielle de paramètre $1/30$. On a alors $\mathbb{E}[X_1] = 30$ et $Var(X_1) = 900$. X_n représente le temps (en minutes) qui sépare le passage de la $(n-1)$ième voiture de la $n-$ième voiture. On considère ensuite $S_n = X_1 + \ldots X_n$. S_n donne l'instant de passage de la $n-$ième voiture. On veut déterminer

$$\mathbb{P}[S_{50} \leq 24 \times 60].$$

Or on a

$$\mathbb{P}[S_{50} \leq 24 \times 60] = \mathbb{P}\left[\frac{S_{50} - 50 \times 30}{30 \times \sqrt{50}} \leq \frac{24 \times 60 - 50 \times 30}{30 \times \sqrt{50}}\right] \approx F\left(\frac{-2}{\sqrt{50}}\right),$$

où F est la fonction de répartition de la loi normale centrée réduite, l'approximation étant donnée par le TCL. De plus $F\left(\frac{-2}{\sqrt{50}}\right) \approx F(-0.283) = 1 - F(0.283)$. En utilisant une table ou un logiciel on obtient $F(0.283) \approx 0.61$. On a donc $\mathbb{P}[S_{50} \leq 24 \times 60] \approx 0.39$.

7. La somme des résultats de 10000 lancers d'un même dé est 35487. Pensez-vous que ce dé soit truqué ? *On pourra utiliser l'égalité approchée* $\int_{3,29}^{+\infty} e^{-\frac{x^2}{2}}\, dx \simeq 0,0005\sqrt{2\pi}$.

Solution de l'exercice 7. Une suite de lancers de dés peut être modélisée par une suite $(X_n)_{n\geq 1}$de variables aléatoires indépendantes de même loi uniforme sur l'ensemble fini $\{1, \ldots, 6\}$. On a

$$\begin{aligned}
\mathbb{E}[X_1] &= \frac{1}{6}(1+2+3+4+5+6) = \frac{7}{2}, \\
\mathbb{E}[X_1^2] &= \frac{1}{6}(1^2+2^2+3^2+4^2+5^2+6^2) = \frac{91}{6}, \\
\mathrm{Var}(X_1) &= \mathbb{E}[X_1^2] - \mathbb{E}[X_1]^2 = \frac{35}{12}.
\end{aligned}$$

Notons $S_n = X_1 + \ldots + X_n$ la somme des n premiers lancers. Le théorème central limite affirme que l'on a la convergence en loi

$$\frac{S_n - \frac{7n}{2}}{\sqrt{n\mathrm{Var}(X_1)}} \xrightarrow[n\to\infty]{(d)} \mathcal{N}(0,1).$$

Faisons l'approximation consistant à dire que pour $n = 10000$, la loi de $\frac{S_n - \frac{7n}{2}}{\sqrt{n\mathrm{Var}(X_1)}}$ est égale à la loi normale $\mathcal{N}(0,1)$. On a alors, pour tout $a > 0$, et avec $n = 10000$,

$$\mathbb{P}\left(\frac{7n}{2} - \sqrt{n\mathrm{Var}(X_1)}a \leq S_n \leq \frac{7n}{2} + \sqrt{n\mathrm{Var}(X_1)}a\right) = \frac{1}{\sqrt{2\pi}}\int_{-a}^{a} e^{-\frac{x^2}{2}}\, dx.$$

Prenons $a = 3,29$. Nous avons

$$\frac{1}{\sqrt{2\pi}}\int_{-3,29}^{3,29} e^{-\frac{x^2}{2}}\, dx = 1 - 2\frac{1}{\sqrt{2\pi}}\int_{3,29}^{+\infty} e^{-\frac{x^2}{2}}\, dx = 0,999.$$

Ainsi,

$$\mathbb{P}(35000 - 562 \leq S_{10000} \leq 35000 + 562) \simeq 0,999.$$

Nous observons une somme qui appartient à cet intervalle. Nous décidons donc de dire que la somme obtenue est compatible avec l'hypothèse que ce dé est équilibré.