Petite Classe 5 - MAP 311 - Groupes 1 et 12

Josselin Garnier

22 mai 2017 - salle PC n^o 16

Exercice 1. Soit $\alpha > 0$, et soit $(Z_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes de loi donnée par

$$\mathbb{P}(Z_n = 1) = \frac{1}{n^\alpha} \text{ et } \mathbb{P}(Z_n = 0) = 1 - \frac{1}{n^\alpha}.$$

Montrer que $Z_n \to 0$ dans L^1. Est-ce que la suite $(Z_n)_{n\geq 1}$ converge en probabilité ? Est-ce que la suite (Z_n) converge presque sûrement ?

Solution. On a,

$$\mathbb{E}(|Z_n|) = \mathbb{E}(Z_n) = \frac{1}{n^\alpha} \to 0,$$

ce qui signifie que $Z_n \to 0$ dans L^1. Donc $Z_n \to 0$ en probabilité car la convergence en moyenne implique la convergence en probabilité.

Pour tout $n \geq 1$, on note $A_n = \{Z_n = 1\}$. Si $\alpha > 1$ alors $\sum_{n\geq 1} \mathbb{P}(A_n) < \infty$. D'après le lemme de Borel-Cantelli, on a donc $\mathbb{P}(\limsup A_n) = 0$ c'est-à-dire p.s., il existe $n_0 \geq 1$ tel que pour tout $n \geq n_0$, $Z_n = 0$ et ainsi $Z_n \to 0$ presque sûrement.

Si $\alpha \leq 1$ alors $\sum_{n\geq 1} \mathbb{P}(A_n) = \infty$. Les A_n étant indépendants, d'après le lemme de Borel-Cantelli, on a donc $\mathbb{P}(\limsup A_n) = 1$ c'est-à-dire p.s $Z_n = 1$ une infinité de fois. On a aussi $\sum_{n\geq 1} \mathbb{P}(Z_n = 0) = \infty$, donc p.s. $Z_n = 0$ une infinité de fois. Donc Z_n ne converge pas presque sûrement.

Exercice 2. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes de loi exponentielle de paramètre $\lambda > 0$. Montrer que la convergence

$$\frac{1}{\ln(n)} \max_{1\leq k\leq n} X_k \quad \underset{n\to\infty}{\longrightarrow} \quad \frac{1}{\lambda}$$

a lieu en probabilité.

Solution. Par définition de la convergence en probabilité, il s'agit de montrer que pour tout $\epsilon > 0$ on a

$$\mathbb{P}\Big(\Big| \frac{1}{\ln(n)} \max_{1\leq k\leq n} X_k - \frac{1}{\lambda} \Big| > \epsilon \Big) \quad \underset{n\to\infty}{\longrightarrow} \quad 0.$$

Montrons d'abord que

$$\mathbb{P}\Big(\frac{1}{\ln(n)} \max_{1\leq k\leq n} X_k < \frac{1}{\lambda} - \epsilon \Big) \quad \underset{n\to\infty}{\longrightarrow} \quad 0$$

pour $\epsilon < 1/\lambda$. Par indépendance,

$$\begin{aligned}
\mathbb{P}\Big(\frac{1}{\ln(n)} \max_{1\leq k\leq n} X_k < \frac{1}{\lambda} - \epsilon \Big) &= \mathbb{P}\Big(\frac{1}{\ln(n)} X_1 < \frac{1}{\lambda} - \epsilon \Big)^n \\
&= \big(1 - \exp(-(1-\lambda\epsilon)\ln n) \big)^n = \big(1 - n^{-(1-\lambda\epsilon)} \big)^n \\
&= \exp\left(n \ln\left(1 - \frac{n^{\epsilon\lambda}}{n} \right) \right) \leq e^{-n^{\epsilon\lambda}} \quad \underset{n\to\infty}{\longrightarrow} \quad 0,
\end{aligned}$$

où on a utilisé $\ln(1+x) \leq x \ \forall x > -1$.
Ensuite, en passant au complémentaire,

$$\mathbb{P}\Big(\frac{1}{\ln(n)} \max_{1 \leq k \leq n} X_k > \frac{1}{\lambda} + \epsilon\Big) = 1 - \mathbb{P}\big(X_1 < (1/\lambda + \epsilon)\ln(n)\big)^n = 1 - \left(1 - \frac{1}{n^{1+\lambda\epsilon}}\right)^n.$$

Or

$$1 - \left(1 - \frac{1}{n^{1+\lambda\epsilon}}\right)^n = 1 - \exp\left(n \ln\left(1 - \frac{1}{n^{1+\lambda\epsilon}}\right)\right) \underset{n\to\infty}{\sim} \frac{1}{n^{\epsilon\lambda}} \underset{n\to\infty}{\longrightarrow} 0.$$

Exercice 3. Soit $f : \mathbb{R} \to \mathbb{R}$ une fonction continue bornée et $\alpha > 0$. Calculer les limites suivantes :

$$\lim_{n\to\infty} \int_{[0,1]^n} f\left(\frac{x_1 + \cdots + x_n}{n}\right) dx_1 \cdots dx_n \qquad \text{et} \qquad \lim_{n\to\infty} \sum_{k\geq 0} e^{-\alpha n} \frac{(\alpha n)^k}{k!} f\left(\frac{k}{n}\right).$$

Solution.

– Soient $(U_1, \ldots, U_n)$ des variables aléatoires indépendantes et de même loi uniforme sur $[0,1]$. On a

$$\int_{[0,1]^n} f\left(\frac{x_1 + \cdots + x_n}{n}\right) dx_1 \cdots dx_n = \mathbb{E}\Big(f\left(\frac{U_1 + \cdots + U_n}{n}\right)\Big).$$

D'après la loi forte des grands nombres, $\frac{U_1+\cdots+U_n}{n}$ converge presque sûrement vers $\mathbb{E}(U_1) = 1/2$. Donc $f\left(\frac{U_1+\cdots+U_n}{n}\right)$ converge presque sûrement vers $f(1/2)$. De plus, $|f\left(\frac{U_1+\cdots+U_n}{n}\right)|$ est majorée par $\sup|f|$. Donc

$$\int_{[0,1]^n} f\left(\frac{x_1 + \cdots + x_n}{n}\right) dx_1 \cdots dx_n \underset{n\to\infty}{\longrightarrow} f(1/2)$$

d'après le théorème de convergence dominée.

– Soit X une variable aléatoire de Poisson de paramètre αn. Alors

$$\sum_{k\geq 0} e^{-\alpha n} \frac{(\alpha n)^k}{k!} f\left(\frac{k}{n}\right) = \mathbb{E}(f(X)).$$

Or X a la même loi que $V_1 + V_2 + \cdots + V_n$, où $(V_i)_{1\leq i\leq n}$ sont des variables aléatoires indépendantes de même loi de Poisson de paramètre α. Donc

$$\sum_{k\geq 0} e^{-\alpha n} \frac{(\alpha n)^k}{k!} f\left(\frac{k}{n}\right) = \mathbb{E}\Big(f\left(\frac{V_1 + \cdots + V_n}{n}\right)\Big).$$

D'après la loi forte des grands nombres, $\frac{V_1+\cdots+V_n}{n}$ converge presque sûrement vers $\mathbb{E}(V_1) = \alpha$. Donc $f\left(\frac{V_1+\cdots+V_n}{n}\right)$ converge presque sûrement vers $f(\alpha)$. De plus, $|f\left(\frac{V_1+\cdots+V_n}{n}\right)|$ est majorée par $\sup|f|$. Donc

$$\sum_{k\geq 0} e^{-\alpha n} \frac{(\alpha n)^k}{k!} f\left(\frac{k}{n}\right) \underset{n\to\infty}{\longrightarrow} f(\alpha).$$

d'après le théorème de convergence dominée.

Exercice 4. (Lemme de Scheffé) Soient $(X_n)_{n\geq 1}$ des variables aléatoires **positives** qui convergent presque sûrement vers X. On suppose que $\mathbb{E}(X) < \infty$ et que $\mathbb{E}(X_n) \to \mathbb{E}(X)$ lorsque $n \to \infty$. Le but de cet exercice est de démontrer que $X_n \to X$ dans L^1.

On pose $Y_n = \min(X_n, X)$ et $Z_n = \max(X_n, X)$.

1. Démontrer que $\mathbb{E}(Y_n) \to \mathbb{E}(X)$ lorsque $n \to \infty$.
2. Démontrer que $\mathbb{E}(Z_n) \to \mathbb{E}(X)$ lorsque $n \to \infty$.

   Indication. On pourra écrire que $Z_n = X + X_n - Y_n$.
3. Conclure que $\mathbb{E}(|X_n - X|) \to 0$ lorsque $n \to \infty$.

   Indication. On pourra écrire que $|X_n - X| = Z_n - Y_n$.

Solution.

1. Y_n converge presque sûrement vers X, en étant majorée par X. Donc $\mathbb{E}(Y_n) \to \mathbb{E}(X)$ d'après le théorème de convergence dominée.
2. On a $\mathbb{E}(Z_n) = \mathbb{E}(X) + \mathbb{E}(X_n) - \mathbb{E}(Y_n) \to \mathbb{E}(X) + \mathbb{E}(X) - \mathbb{E}(X) = \mathbb{E}(X)$.
3. On a $\mathbb{E}(|X_n - X|) = \mathbb{E}(Z_n) - \mathbb{E}(Y_n) \to \mathbb{E}(X) - \mathbb{E}(X) = 0$.

Exercice 5. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes et de même loi donnée par $\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = -1) = \frac{1}{2}$. Soit $k \geq 1$ un entier. On pose $S_0 = k$ et pour tout $n \geq 1$

$$S_n = k + X_1 + \cdots + X_n.$$

Soit $T = \inf\{i \geq 1 : S_i = 0 \text{ ou } S_i = 2k\}$ (avec la convention $\inf \emptyset = \infty$).

1. Montrer que $T < \infty$ presque sûrement.
2. On pose $Z_n = S_{\min(n,T)}$. Montrer que Z_n converge presque sûrement vers une variable aléatoire dont on identifiera la loi. Est-ce que Z_n converge en probabilité ? En moyenne ?

Solution.

1. S'il y a $2k$ montées successives, alors $T < \infty$. Une application du lemme de Borel-Cantelli (semblable à celle de la séquence de 100 piles successifs) montre que dans la suite $(X_i)_{i\geq 1}$ l'événement "1 apparaît $2k$ fois de suite" arrive une infinité de fois presque sûrement. Donc $\mathbb{P}(T < \infty) = 1$.
2. Comme $T < \infty$ presque sûrement, Z_n converge presque sûrement vers S_T. Donc Z_n converge également en probabilité vers S_T (la convergence p.s. implique la convergence en probabilité). Comme $|Z_n| \leq 2k$, la suite converge également en moyenne.

Par symétrie,

$$\mathbb{P}(S_T = 0) = \mathbb{P}(S_T = 2k) = \frac{1}{2}.$$

Rédigeons cet argument formellement. Soit Φ l'application qui transforme une marche sur $\mathbb{Z}$ qui fait des sauts ± 1 en la marche obtenue en changeant les sauts $+1$ en sauts -1 et les sauts -1 en sauts $+1$. Notons $(S'_n)_{n\geq 0} = \Phi((S_n)_{n\geq 0})$ et $T' = \inf\{i \geq 1 : S'_i = 0 \text{ ou } S'_i = 2k\}$. Alors, par construction, $S_T = 0$ si et seulement si $S'_{T'} = 2k$. De plus, $(S_n)_{n\geq 0}$ et $(S'_n)_{n\geq 0}$ ont la même loi. Donc $\mathbb{P}(S_T = 0) = \mathbb{P}(S'_{T'} = 2k) = \mathbb{P}(S_T = 2k)$. Comme $\mathbb{P}(S_T = 0) + \mathbb{P}(S_T = 2k) = 1$, le résultat voulu en découle.

Exercice 6. (**Problème du collectionneur**) Soit $(X_k)_{k\geq 1}$ une suite de variables aléatoires indépendantes uniformément distribuées sur $\{1, 2, \ldots, n\}$. Soit

$$T_n = \inf\{m \geq 1 : \{X_1, \ldots, X_m\} = \{1, 2, \ldots, n\}\}$$

le premier temps où toutes les valeurs ont été observées.

Motivation : Votre petit frère collectionne les images des joueurs de la coupe du monde que l'on trouve dans les tablettes de chocolat. On suppose qu'il existe n images différentes, et qu'elles sont équitablement distribuées dans l'ensemble des tablettes (que l'on supposera infini). T_n est le nombre de tablettes qu'il faut acheter pour avoir toutes les images de la collection.

1. Soit $\tau_k^n = \inf\{m \geq 1 : |\{X_1, \ldots, X_m\}| = k\}$ pour tout $k \geq 1$. Montrer que les variables $(\tau_k^n - \tau_{k-1}^n)_{2\leq k\leq n}$ sont indépendantes, et déterminer leurs lois respectives.

2. En déduire que la convergence $\frac{T_n}{n \log n} \to 1$ a lieu en probabilité.

   Indication. On pourra utiliser l'inégalité de Bienaymé-Tchébychev.

Solution.

1. La quantité $\tau_k^n - \tau_{k-1}^n$ représente le temps mis pour obtenir un nouvel élément une fois qu'on en a obtenu $k-1$. Intuitivement, ces variables sont indépendantes et suivent une loi géométrique de paramètre $\frac{n-k+1}{n}$ (il reste $n-(k-1)$ éléments).

   Pour le démontrer formellement, on peut procéder comme suit. On a $\tau_1^n = 1$. Soit $(t_2, \ldots, t_n) \in (\mathbb{N}^*)^{n-1}$. On veut calculer $\mathbb{P}(\tau_2^n - \tau_1^n = t_2, \ldots, \tau_n^n - \tau_{n-1}^n = t_n)$. En posant $t_1 = 1$ et en notant $\mathcal{S}_n$ l'ensemble des permutations de $\{1, 2, \ldots, n\}$ ($\sigma(k)$ va désigner le numéro de la nouvelle image obtenue à l'instant τ_k^n), on a

$$\begin{aligned}
&\mathbb{P}(\tau_2^n - \tau_1^n = t_2, \ldots, \tau_n^n - \tau_{n-1}^n = t_n) \\
&= \sum_{\sigma \in \mathcal{S}_n} \mathbb{P}\Bigg(\bigcap_{k=1}^{n-1} \Big\{ X_{t_1+\cdots+t_k} = \sigma(k), X_{t_1+\cdots+t_k+1} \in \{\sigma(1), \ldots, \sigma(k)\}, \ldots, \\
&\qquad\qquad X_{t_1+\cdots+t_k+t_{k+1}-1} \in \{\sigma(1), \ldots, \sigma(k)\} \Big\} \bigcap \{X_{t_1+\cdots+t_n} = \sigma(n)\} \Bigg) \\
&= \sum_{\sigma \in \mathcal{S}_n} \frac{1}{n^n} \prod_{k=2}^{n} \left(\frac{k-1}{n}\right)^{t_k - 1} \\
&= \frac{n!}{n^n} \prod_{k=2}^{n} \left(\frac{k-1}{n}\right)^{t_k - 1} \\
&= \prod_{k=2}^{n} \left(\frac{n+1-k}{n}\right) \left(\frac{k-1}{n}\right)^{t_k - 1}.
\end{aligned}$$

Donc les variables aléatoires $(\tau_k^n - \tau_{k-1}^n)_{2 \le k \le n}$ sont indépendantes, et ont respectivement pour loi

$$\mathbb{P}(\tau_k^n - \tau_{k-1}^n = i) = \left(\frac{n+1-k}{n}\right)\left(\frac{k-1}{n}\right)^{i-1} = \left(\frac{n-k+1}{n}\right)\left(1 - \frac{n-k+1}{n}\right)^{i-1}$$

pour $i \ge 1$. Cette loi est bien la loi géométrique de paramètre $\frac{n-k+1}{n}$.

2. On a $T_n = \tau_1 + \sum_{k=2}^{n} (\tau_k - \tau_{k-1})$ et donc

$$\mathbb{E}(T_n) = 1 + \sum_{k=2}^{n} \frac{n}{n+1-k} = 1 + nH_{n-1},$$

où $H_n = \sum_{k=1}^{n} \frac{1}{k}$ est la série harmonique et

$$\mathrm{Var}(T_n) = \sum_{k=2}^{n} \mathrm{Var}(\tau_k - \tau_{k-1}) = \sum_{k=2}^{n} \frac{(k-1)/n}{((n+1-k)/n)^2} = n \sum_{k=1}^{n-1} \frac{n-k}{k^2} \le Cn^2$$

avec $C = \sum_{k \ge 1} \frac{1}{k^2}$. Donc pour tout $\epsilon > 0$,

$$\mathbb{P}\big(|T_n - \mathbb{E}(T_n)| \ge \epsilon n \log(n)\big) \le \frac{\mathrm{Var}(T_n)}{\epsilon^2 n^2 \log(n)^2} \le \frac{C}{\epsilon^2 \log(n)^2}.$$

Donc $(T_n - \mathbb{E}(T_n))/(n \log(n)) \to 0$ en probabilité. Or $\mathbb{E}(T_n) \sim n \log(n)$ quand $n \to \infty$. Donc pour tout $\epsilon > 0$ on a $\{|T_n - n\log(n)| \ge 2\epsilon n \log(n)\} \subset \{|T_n - \mathbb{E}(T_n)| \ge \epsilon n \log(n)\}$ pour n assez grand. On obtient ainsi le résultat.

Exercice 7. (**Biais par la taille**) On considère une population comportant un grand nombre n de foyers. On modélise la taille de ces foyers par une suite de variables aléatoires $(X_i)_{1\leq i\leq n}$ indépendantes de même loi sur $\mathbb{N}^*$, de moyenne $m = \mathbb{E}(X_1) = \sum_{k\geq 1} kp_k < \infty$ avec $p_k = \mathbb{P}(X_1 = k)$. Soit T la taille du foyer d'un individu pris au hasard dans la population. Montrer que pour tout entier $k \geq 1$, $\mathbb{P}(T = k)$ converge vers $\frac{k}{m}p_k$ lorsque $n \to \infty$.
Indication. On pourra commencer par calculer $\mathbb{P}(T = k|X_1, \ldots, X_n)$.

Solution. La population compte au total $X_1 + \cdots + X_n$ individus. On modélise le choix d'un individu au hasard par le tirage, conditionnellement à $X_1, \ldots, X_n$ (c'est-à-dire sachant $X_1, \ldots, X_n$), d'un entier selon la loi uniforme sur l'ensemble $\{1, 2, \ldots, X_1 + \cdots + X_n\}$. Pour tout $k \geq 1$, il y a $N_k = \mathbf{1}_{\{X_1=k\}} + \cdots + \mathbf{1}_{\{X_n=k\}}$ foyers de taille k qui comptent au total kN_k individus. Par conséquent, par définition de la loi uniforme (formule "cas favorables sur cas totaux"):

$$\mathbb{P}(T = k|X_1, \ldots, X_n) = \frac{k\sum_{i=1}^n \mathbf{1}_{\{X_i=k\}}}{\sum_{i=1}^n X_i} = \frac{k\frac{1}{n}\sum_{i=1}^n \mathbf{1}_{\{X_i=k\}}}{\frac{1}{n}\sum_{i=1}^n X_i}.$$

Donc, d'après la loi forte des grands nombres,

$$\mathbb{P}(T = k|X_1, \ldots, X_n) \quad \underset{n\to\infty}{\overset{p.s.}{\longrightarrow}} \quad \frac{k}{m}p_k.$$

Posons $F(X_1, \ldots, X_n) = \mathbb{P}(T = k|X_1, \ldots, X_n)$, de sorte que $\mathbb{P}(T = k) = \mathbb{E}(F(X_1, \ldots, X_n))$ (plus généralement, $\mathbb{E}(U) = \mathbb{E}(\mathbb{E}(U|V))$). D'après ce qu'on vient de voir, $F(X_1, \ldots, X_n)$ converge presque sûrement vers $\frac{k}{m}p_k$ en étant dominé par 1 (car c'est une probabilité conditionnelle), intégrable. D'après le théorème de convergence dominée,

$$\mathbb{P}(T = k) = \mathbb{E}(F(X_1, \ldots, X_n)) \quad \underset{n\to\infty}{\longrightarrow} \quad \frac{k}{m}p_k.$$

Interprétation. La loi de T n'est pas la loi de départ, car les grands foyers sont sur-représentés tandis que les petits foyers sont sous-représentés. Ce phénomène est appelé *biais par la taille*. Il s'agit sans doute du biais d'échantillonnage le plus connu. Le biais est d'autant plus important que la taille du foyer diffère de la taille moyenne m.

Pour aller plus loin

Exercice 8. (Extension du théorème de convergence dominée) Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires réelles telle que X_n converge en probabilité vers X. On suppose qu'il existe une variable aléatoire positive et intégrable Z, indépendante de n, telle que $|X_n| \leq Z$ pour tout $n \geq 1$. Montrer que X_n converge vers X en moyenne.
Indication. On pourra utiliser que si Y_n converge en probabilité vers Y alors il existe une sous-suite de Y_n qui converge presque sûrement vers Y.

Solution. Tout d'abord, montrons que $|X| \leq Z$ presque sûrement. Comme X_n converge en probabilité vers X, il existe une sous-suite $X_{\phi(n)}$ qui converge presque sûrement vers X. Comme $|X_{\phi(n)}| \leq Z$ pour tout $n \geq 1$, on en déduit que $|X| \leq Z$ presque sûrement.

On en déduit que la suite $\mathbb{E}(|X_n - X|)$ est bornée, car $\mathbb{E}(|X_n - X|) \leq \mathbb{E}(|X_n|) + \mathbb{E}(|X|) \leq 2\mathbb{E}(Z)$. Pour montrer que $\mathbb{E}(|X_n - X|) \to 0$, il suffit donc de montrer que 0 est la seule valeur d'adhérence de la suite $\mathbb{E}(|X_n - X|)$.

Pour cela, soit ϕ une extraction telle que $\mathbb{E}(|X_{\phi(n)} - X|) \to \ell$. Comme $X_{\phi(n)}$ converge aussi en probabilité vers X lorsque $n \to \infty$, il existe une extraction ψ telle que $X_{\phi(\psi(n))}$ converge presque sûrement vers X lorsque $n \to \infty$. On peut alors appliquer le théorème de convergence dominé usuel pour en déduire que $\mathbb{E}(|X_{\phi(\psi(n))} - X|) \to 0$. Donc $\ell = 0$, ce qui conclut.

Exercice 9. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires qui converge en probabilité vers X et soit $(Y_n)_{n\geq 1}$ une suite de variables aléatoires qui converge en probabilité vers Y. Montrer que (X_n, Y_n) converge en probabilité vers (X, Y).

Solution. Pour tout $\epsilon > 0$, on remarque que si

$$\|(X_n, Y_n) - (X, Y)\| \geq 2\epsilon.$$

alors $|X_n - X| \geq \epsilon$ ou $|Y_n - Y| \geq \epsilon$. En effet, par l'absurde, si $|X_n - X| < \epsilon$ et $|Y_n - Y| < \epsilon$, alors

$$\|(X_n, Y_n) - (X, Y)\| \leq \sqrt{\epsilon^2 + \epsilon^2} = \sqrt{2}\epsilon < 2\epsilon.$$

Ainsi, pour tout $\epsilon > 0$,

$$\mathbb{P}(\|(X_n, Y_n) - (X, Y)\| \geq 2\epsilon) \leq \mathbb{P}(|X_n - X| \geq \epsilon) + \mathbb{P}(|Y_n - Y| \geq \epsilon) \quad \underset{n\to\infty}{\longrightarrow} \quad 0,$$

d'où le résultat.

Exercice 10. Soient X et Y deux variables aléatoires réelles bornées. On suppose que $\mathbb{E}(X^n) = \mathbb{E}(Y^n)$ pour tout $n \geq 1$. Montrer que X et Y ont la même loi.
Indication. Utiliser le théorème de Weierstrass sur la densité des polynômes.

Solution. Soit $f : \mathbb{R} \to \mathbb{R}$ une fonction continue bornée. Il s'agit de montrer que $\mathbb{E}(f(X)) = \mathbb{E}(f(Y))$. Soient $a < b$ tels que $\mathbb{P}(a < X < b) = \mathbb{P}(a < Y < b) = 1$. On peut supposer que f est à support compact dans $[a, b]$. Il existe alors une suite $(P_n)_{n\geq 1}$ de polynômes à coefficients réels tels que P_n converge uniformément vers f sur $[a, b]$. En particulier, il existe une constante $C > 0$ telle que $\sup_{[a,b]} |P_n| \leq C$ pour tout $n \geq 1$.

Posons $X_n = P_n(X)$. Alors X_n converge presque sûrement vers $f(X)$ (car la convergence uniforme implique la convergence simple), et $|X_n| \leq C$ pour tout $n \geq 1$. D'après le théorème de convergence dominée, on en déduit que $\mathbb{E}(X_n) = \mathbb{E}(P_n(X)) \to \mathbb{E}(f(X))$. On montre de même que $\mathbb{E}(P_n(Y)) \to \mathbb{E}(f(Y))$.

Or, par hypothèse, $\mathbb{E}(P_n(X)) = \mathbb{E}(P_n(Y))$. Ceci conclut.

Exercice 11. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires réelles définies sur le même espace de probabilité. Montrer qu'il existe une suite $(c_n)_{n\geq 1}$ telle que X_n/c_n converge presque sûrement vers 0.

Solution. Soit $n \geq 1$, comme la fonction de répartition de $|X_n|$ tend vers 1 en ∞, il existe $c_n > 0$ suffisamment grand tel que $\mathbb{P}\big(|X_n| > \frac{c_n}{n}\big) \leq 2^{-n}$. Montrons que $(c_n)_{n\geq 1}$ convient.

D'après le (premier) lemme de Borel-Cantelli, il suffit de montrer que pour tout $\epsilon > 0$, la série $\sum_{n\geq 1} \mathbb{P}\Big(\Big|\frac{X_n}{c_n}\Big| > \epsilon\Big)$ converge. Alors, pour $n > 1/\epsilon$ (de sorte que $\epsilon > 1/n$),

$$\mathbb{P}\Big(\Big|\frac{X_n}{c_n}\Big| > \epsilon\Big) \leq \mathbb{P}\Big(\Big|\frac{X_n}{c_n}\Big| > \frac{1}{n}\Big) \leq \frac{1}{2^n}$$

par définition de c_n. D'où le résultat.

Exercice 12. Soit $(X_n)_{n\geq 0}$ une suite de variables aléatoires réelles positives, indépendantes et de même loi.

1. Montrer que p.s. $\sum_{n=0}^{\infty} X_n = \infty$, sauf dans un cas à préciser.
2. Montrer que $\alpha \sum_{n\geq 1} \mathbb{P}(X \geq \alpha n) \leq \mathbb{E}(X) < \alpha \sum_{n\geq 1} \mathbb{P}(X \geq \alpha n) + \alpha$.

   Indication. On pourra montrer que

$$\sum_{n\geq 0} \alpha n \mathbb{P}(\alpha n \leq X < \alpha(n+1)) \leq \mathbb{E}(X) \leq \sum_{n\geq 0} \alpha(n+1)\mathbb{P}(\alpha n \leq X < \alpha(n+1)).$$

3. Montrer que pour tout $\alpha > 0$ on a l'équivalence suivante :

$$\mathbb{E}(X) < \infty \iff \sum_{n\geq 0} \mathbb{P}(X \geq \alpha n) < \infty.$$

4. En déduire la dichotomie suivante : presque sûrement,

$$\text{la plus grande valeur d'adhérence de } \frac{X_n}{n} \text{ vaut } \quad \begin{cases} 0 & \text{si } \mathbb{E}(X_1) < \infty \\ \infty & \text{si } \mathbb{E}(X_1) = \infty \end{cases}.$$

Solution.

1. Tout d'abord, si X_1 est presque sûrement nulle, alors $\sum_{n\geq 1} X_n = 0$ p.s. Supposons que X_1 n'est pas nulle, alors on peut trouver $\epsilon > 0$ tel que $\mathbb{P}(X_1 > \epsilon) > 0$. On a alors que $\sum_{n\geq 1} \mathbb{P}(X_n > \epsilon) = \infty$, et les X_n sont indépendants, donc par le (deuxième) lemme Borel-Cantelli, p.s. les X_n sont supérieurs à ϵ une infinité de fois, donc $\sum_{n\geq 1} X_n = \infty$.
2. Soit X une v.a. positive et $\alpha > 0$. On vérifie aisément les encadrements de X suivants :

$$\sum_{n\geq 0} \alpha n \mathbf{1}_{\alpha n \leq X < \alpha(n+1)} \leq X \leq \sum_{n\geq 0} \alpha(n+1)\mathbf{1}_{\alpha n \leq X < \alpha(n+1)}.$$

   En prenant l'espérance de la ligne précédente on obtient

$$\begin{aligned} \sum_{n\geq 0} \alpha n \left(\mathbb{P}(X \geq \alpha n) - \mathbb{P}(X \geq \alpha(n+1))\right) &\leq \mathbb{E}(X) \\ &\leq \sum_{n\geq 0} \alpha(n+1)\left(\mathbb{P}(X \geq \alpha n) - \mathbb{P}(X \geq \alpha(n+1))\right) \end{aligned}$$

En reconnaissant des sommes téléscopiques, on en déduit que

$$\alpha \sum_{n\geq 1} \mathbb{P}(X \geq \alpha n) \leq \mathbb{E}(X) < \alpha \sum_{n\geq 1} \mathbb{P}(X \geq \alpha n) + \alpha.$$

3. C'est une conséquence immédiate de la question précédente.
4. Soit $\alpha > 0$. La plus grande valeur d'adhérence de $\frac{X_n}{n}$ est supérieure ou égale à α si et seulement si il existe une infinité de n tels que $X_n \geq \alpha n$. D'après la question précédente,

$$\sum_{n\geq 0} \mathbb{P}(X \geq \alpha n) \quad \left\{ \begin{array}{ll} < \infty & \text{si } \mathbb{E}(X_1) < \infty \\ = \infty & \text{si } \mathbb{E}(X_1) = \infty \end{array} \right. ,$$

donc, d'après les deux lemmes de Borel-Cantelli,

$$\mathbb{P}\left(\text{la plus grande valeur d'adhérence de } \frac{X_n}{n} \geq \alpha\right) = \left\{ \begin{array}{ll} 0 & \text{si } \mathbb{E}(X_1) < \infty \\ 1 & \text{si } \mathbb{E}(X_1) = \infty \end{array} \right.$$

et par conséquent

$$\text{la plus grande valeur d'adhérence de } \frac{X_n}{n} = \left\{ \begin{array}{ll} 0 & \text{si } \mathbb{E}(X_1) < \infty \\ \infty & \text{si } \mathbb{E}(X_1) = \infty \end{array} \right. \text{ p.s.}$$

Exercice 13. La convergence de l'exercice 2 a-t-elle lieu presque sûrement ? Dans L^1 ?

Solution. Montrons que la convergence a lieu p.s. et dans L^1. On pose

$$Z_n = \frac{1}{\ln(n)} \max_{1\leq k\leq n} X_k$$

(1) Soit $\epsilon \in (0,1)$ et posons $A_n = \{Z_n \leq (1-\epsilon)\lambda^{-1}\}$. Montrons que $\sum \mathbb{P}(A_n)$ converge. D'après les calculs de la solution de l'exercice 2, on a $\mathbb{P}(A_n) \leq e^{-n^{\epsilon\lambda}}$. Donc $\sum \mathbb{P}(A_n)$ converge. D'après le lemme de Borel-Cantelli, p.s., pour tout n suffisamment grand on a $Z_n \geq 1-\epsilon$. Donc p.s. pour tout $k \geq 1$, $Z_n \geq 1 - 1/k$ pour tout n suffisamment grand.

Posons $B_n = \{Z_n \geq (1+\epsilon)\lambda^{-1}\}$.Les calculs de la solution de l'exercice 2 donnent

$$\mathbb{P}(B_n) = 1 - \left(1 - \frac{1}{n^{1+\lambda\epsilon}}\right)^n \quad \underset{n\to\infty}{\sim} \quad \frac{1}{n^{\epsilon\lambda}}.$$

La série de terme général $1/n^{\epsilon\lambda}$ ne converge pas, il faut donc ruser un peu comme dans l'exercice 7 de la PC 1. Fixons $\eta > 0$ et posons $n_k = (1+\eta)^k$. Alors $\sum_k P(B_{n_k})$ converge et d'après le lemme de Borel-Cantelli, p.s., pour tout k suffisamment grand on a $Z_{n_k} \leq (1+\epsilon)\lambda^{-1}$. On encadre ensuite $n \geq 1$: $(1+\eta)^k \leq n \leq (1+\eta)^{k+1}$ et on écrit:

$$Z_n = \frac{\max(X_1,\ldots,X_n)}{\ln(n)} \leq \frac{\max(X_1,\ldots,X_{n_{k+1}})}{\ln(n)} \leq \frac{\max(X_1,\ldots,X_{n_{k+1}})}{\ln(n_{k+1})} \frac{\ln(n_{k+1})}{\ln(n_k)} = Z_{n_{k+1}} \frac{k+1}{k}.$$

Il s'ensuit que p.s. $Z_n \leq (1+2\epsilon)\lambda^{-1}$ pour tout n assez grand. Donc p.s. pour tout $k \geq 1$, $Z_n \leq \lambda + 1/k$ pour tout n suffisamment grand.

On conclut que $\lim_{n\to\infty} Z_n = 1$ p.s.

(2) Pour la convergence dans L^1, nous allons utiliser le lemme de Scheffé (exercice 4), qui nous garantit qu'il suffit de vérifier que $\mathbb{E}(Z_n) \to 1/\lambda$.

Montrons d'abord que

$$\int_0^1 \ln(1-v)v^{n-1}dv = -\frac{1+\frac{1}{2}+\cdots+\frac{1}{n}}{n}. \tag{1}$$

Pour cela, on écrit $\ln(1-v) = -\sum_{k\geq 1}\frac{v^k}{k}$. Comme

$$\sum_{k\geq 1}\int_0^1 \frac{v^{k+n-1}}{k}dv = \sum_{k\geq 1}\frac{1}{(k+n)k} < \infty,$$

on a

$$\int_0^1 \ln(1-v)v^{n-1}dv = -\sum_{k\geq 1}\frac{1}{(k+n)k}.$$

Pour calculer cette somme, remarquons que $\frac{1}{k(k+n)} = \frac{1}{n}\left(\frac{1}{k}-\frac{1}{k+n}\right)$, de sorte que

$$n\sum_{k\geq 1}\frac{1}{(k+n)k} = \sum_{k\geq 1}\left(\frac{1}{k}-\frac{1}{k+n}\right) = \lim_{N\to\infty}\sum_{k=1}^{N}\left(\frac{1}{k}-\frac{1}{k+n}\right) = \lim_{N\to\infty} H_N - H_{n+N} + \sum_{k=1}^{n}\frac{1}{k}$$

en notant $H_n = 1+\frac{1}{2}+\cdots+\frac{1}{n}$ la série harmonique. Comme $H_n = \ln(n)+O(1)$ quand $n\to\infty$, on en déduit que

$$n\sum_{k\geq 1}\frac{1}{(k+n)k} = \sum_{k=1}^{n}\frac{1}{k},$$

et (1) en découle.

Revenons maintenant au problème du comportement asymptotique de $\mathbb{E}(Z_n)$. Posons $Y_n = \max(X_1,\ldots,X_n)$, de sorte que $Z_n = Y_n/\log(n)$, et déterminons une densité de Y_n. Pour $u \geq 0$, $\mathbb{P}(Y_n \leq u) = (1-e^{-\lambda u})^n$. On en déduit que $\lambda n e^{-\lambda u}(1-e^{-\lambda u})^{n-1}\mathbf{1}_{u\geq 0}$ est une densité de Y_n. Ainsi,

$$\mathbb{E}(Z_n) = \int_0^\infty \frac{\lambda n}{\ln(n)} u e^{-\lambda u}(1-e^{-\lambda u})^{n-1}du = \int_0^\infty \frac{\lambda}{\ln(n)} u e^{-\lambda\frac{u}{n}}(1-e^{-\lambda\frac{u}{n}})^{n-1}du.$$

En faisant le changement de variable $v = 1-e^{-\lambda u}$, on en déduit que

$$\mathbb{E}(Z_n) = -\frac{n}{\lambda}\int_0^1 \ln(1-v)v^{n-1}dv = \frac{1}{\lambda}\left(1+\frac{1}{2}+\cdots+\frac{1}{n}\right)$$

compte tenu de (1). On en déduit que $\mathbb{E}(Z_n)\to 1/\lambda$ car $H_n \sim \log(n)$ lorsque $n\to\infty$, ce qui conclut.

Remarque. En utilisant la formule $\mathbb{E}(Y_n) = \int_0^\infty \mathbb{P}(Y_n \geq t)dt$ que nous verrons un peu plus tard (suivi du changement de variable $u = 1-e^{-t}$), le calcul de $\mathbb{E}(Y_n)$ est un peu plus facile.

Remarque. Voici un argument probabiliste qui permet de montrer que $\mathbb{E}(Y_n) = \frac{1}{\lambda}(1+\frac{1}{2}+\cdots+\frac{1}{n})$. Imaginons que nous disposons de n horloges et que X_k représente le temps auquel sonne la k-ième horloge. La variable aléatoire Y_n représente alors le temps au bout duquel sonne la dernière horloge. La première horloge sonne alors au temps $\min(X_1, \ldots, X_n)$, qui est distribué suivant une variable exponentielle de paramètre λn (d'espérance $\frac{1}{\lambda n}$ donc), et d'après la propriété d'absence de mémoire, en remettant le temps à 0, les temps au bout desquels sonnent les $n-1$ horloges restantes sont indépendantes et de même loi exponentielle de paramètre λ. On en déduit que $\mathbb{E}(Y_n) = \frac{1}{\lambda n} + \mathbb{E}(Y_{n-1})$, et le résultat voulu en découle.

Pour rédiger ceci de manière formelle, notons $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ le réarrangement croissant des variables $(X_1, \ldots, X_n)$. On montre alors que la densité conditionnelle de $(X_{(2)} - X_{(1)}, \ldots, X_{(n)} - X_{(1)})$ sachant $X_{(1)}$ est la même que celle du réarrangement croissant de $n-1$ variables aléatoires i.i.d. exponentielles de paramètre λ (ce calcul est proche de celui de l'exercice 13 de la PC4). On écrit alors :

$$\mathbb{E}(Y_n) = \mathbb{E}(X_{(n)}) = \mathbb{E}(X_{(n)} - X_{(1)}) + \mathbb{E}(X_{(1)}) = \mathbb{E}(\mathbb{E}[X_{(n)} - X_{(1)} | X_{(1)}]) + \frac{1}{\lambda n} = \mathbb{E}(Y_{n-1}) + \frac{1}{\lambda n}.$$