TD16

Convergence de variables aléatoires

Inégalités de concentration

Exercice 16.1 (★★)

En appliquant l'inégalité de Markov à $|X|$, où X est une variable aléatoire suivant la loi normale centrée réduite, montrer que pour tout réel x strictement positif, on a :

$$\int_0^x e^{-\frac{t^2}{2}} \,\mathrm{d}t \geq \sqrt{\frac{\pi}{2}} - \frac{1}{x}.$$

Soit $X \hookrightarrow \mathcal{N}(0,1)$. Notons $f : t \mapsto \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$ une densité de X. $|X|$ est positive et admet une espérance (car X en admet une), et on a par l'inégalité de Markov :

$$\forall x > 0, \quad P(|X| \geq x) \leq \frac{E(|X|)}{x}.$$

On a :

$$\begin{aligned} P(|X| \geq x) = P(X \geq x) + P(X \leq -x) &= \int_x^{+\infty} f(t) \,\mathrm{d}t + \int_{-\infty}^{-x} f(t) \,\mathrm{d}t \\ &= \int_{-\infty}^{+\infty} f(t) \,\mathrm{d}t - \int_{-x}^{x} f(t) \,\mathrm{d}t \\ &= 1 - 2\int_0^x f(t) \,\mathrm{d}t \end{aligned}$$

car f est paire et est une densité de probabilité. D'autre part on a par le théorème de transfert :

$$E(|X|) = \int_{-\infty}^{+\infty} |t| f(t) \,\mathrm{d}t = 2\int_0^{+\infty} t f(t) \,\mathrm{d}t$$

car $t \mapsto |t| f(t)$ est paire. Or on a pour tout $A > 0$:

$$\int_0^A t f(t) \,\mathrm{d}t = \frac{1}{\sqrt{2\pi}} \left[-e^{-t^2/2}\right]_0^A \underset{A \to +\infty}{\longrightarrow} \frac{1}{\sqrt{2\pi}}$$

Finalement on obtient que pour tout $x > 0$:

$$1 - \frac{2}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} \,\mathrm{d}t \leq \frac{2}{x\sqrt{2\pi}},$$

soit encore en multipliant par $\frac{\sqrt{2\pi}}{2} = \sqrt{\frac{\pi}{2}}$:

$$\sqrt{\frac{\pi}{2}} - \frac{1}{x} \mathrm{d}t \leq \int_0^x e^{-\frac{t^2}{2}} \,\mathrm{d}t.$$

Exercice 16.2 (★★)

On note Φ la fonction de répartition d'une variable aléatoire suivant une loi $\mathcal{N}(0,1)$.

1. Montrer à l'aide de l'inégalité de Bienaymé-Tchebychev que : $\forall x \in \mathbb{R}_+^*,\ 1 - \Phi(x) \leq \dfrac{1}{2x^2}$.
2. En déduire que $\displaystyle\int_0^{+\infty}(1-\Phi(t))\mathrm{d}t$ converge et calculer sa valeur à l'aide d'une intégration par parties.

1. On applique l'inégalité de Bienaymé-Tchebychev à $X \hookrightarrow \mathcal{N}(0,1)$ qui admet une variance. Pour tout $x > 0$, on a :
$$P(|X - E(X)|) \geq x) \leq \frac{V(X)}{x^2}.$$
D'où avec $E(X) = 0$ et $V(X) = 1$:
$$P(|X| \geq x) \leq \frac{1}{x^2}.$$
Or on a :
$$P(|X| \geq x) = P(X \geq x) + P(X \leq -x) = 1 - \Phi(x) + \Phi(-x) = 2 - 2\Phi(x)$$
car X est une variable à densité, et par propriété de la fonction de répartition d'une loi normale centrée réduite. On obtient donc que pour tout $x > 0$:
$$2 - 2\Phi(x) \leq \frac{1}{x^2} \quad \text{soit} \quad 1 - \Phi(x) \leq \frac{1}{2x^2}.$$
2. La fonction $t \mapsto 1 - \Phi(t)$ est continue sur $\mathbb{R}_+$ (car X à densité), donc l'intégrale est généralisée en $+\infty$. On a :
   - $1 - \Phi(x) \leq \dfrac{1}{2x^2}$ pour tout $x > 0$;
   - $0 \leq 1 - \Phi(x)$ pour tout $x \in \mathbb{R}_+$ car $\Phi(x) \in]0,1[$;
   - $\displaystyle\int_1^{+\infty} \frac{1}{2t^2}\,\mathrm{d}t$ converge en tant qu'intégrale de Riemann avec $\alpha = 2 > 1$.

   Par théorème de comparaison, on peut donc en déduire que l'intégrale considérée converge.

   Pour calculer cette intégrale, on effectue une intégration par partie sur le **segment** $[0, A]$ avec $A > 0$. Notons pour cela $f : t \mapsto \dfrac{1}{\sqrt{2\pi}}e^{-t^2/2}$ une densité de X.
$$\begin{array}{c|ccc} + & 1-\Phi(t) & & 1 \\ & & \searrow & \\ & & \int & \\ - & -f(t) & \longleftarrow & t \end{array}$$
   Les fonctions $t \mapsto t$ et $t \mapsto 1 - \Phi(t)$ sont bien de classe $\mathscr{C}^1$ sur $[0, A]$. On a donc :
$$\begin{aligned}\int_0^A (1-\Phi(t))\mathrm{d}t &= [-tf(t)]_0^A + \int_0^A tf(t)\,\mathrm{d}t = -Af(A) + [-f(t)]_0^A \\ &= -Af(A) - f(A) + \frac{1}{\sqrt{2\pi}} \underset{A\to+\infty}{\longrightarrow} \frac{1}{\sqrt{2\pi}}\end{aligned}$$
   par croissances comparées. On retrouve ainsi que $\displaystyle\int_0^{+\infty}(1-\Phi(t))\mathrm{d}t$ converge et vaut $\dfrac{1}{\sqrt{2\pi}}$.

Exercice 16.3 (★★★ - QSP ESCP 2022)

Soit $t \in \mathbb{R}$. Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires indépendantes définies sur un même espace probabilisé $(\Omega, \mathscr{A}, P)$, qui suivent toutes la même loi telle que $E(X_n) = V(X_n) = 1$.
Pour tout $n \in \mathbb{N}^*$, on pose $T_n = X_1 + \cdots + X_n$.

1. Pour tout entier $n > t$, comparer les événements $(T_n < t)$ et $(|T_n - n| \geq n - t)$.
2. Calculer $P\left(\bigcap_{n=1}^{+\infty} (T_n < t)\right)$.

Convergence en probabilité

Exercice 16.4 (★)

Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires, définies sur un même espace probabilisé, indépendantes, et qui suivent toutes la loi de Poisson de paramètre $\lambda > 0$.
Pour tout $n \in \mathbb{N}^*$, on pose $S_n = \dfrac{1}{n}\sum_{i=1}^{n} X_i$. Étudier la convergence en probabilité de la suite $(S_n)_{n\in\mathbb{N}^*}$.

Exercice 16.5 (★)

Soit $(\varepsilon_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires et X une variable aléatoire, toutes définies sur le même espace probabilisé. On suppose que pour tout $n \in \mathbb{N}^*$, $\varepsilon_n \hookrightarrow \mathscr{U}(\{1 - \frac{1}{n}, 1 + \frac{1}{n}\})$, et que X admet un moment d'ordre 1.
Pour tout $n \in \mathbb{N}^*$, on pose $X_n = \varepsilon_n X$. Montrer que la suite $(X_n)_{n\in\mathbb{N}^*}$ converge en probabilité vers X.

Exercice 16.6 (★★)

1. Soit (X_n) une suite de variables aléatoires à valeurs positives ayant toutes une espérance et une variance. On suppose que $\lim\limits_{n\to+\infty} E(X_n) = m$ et $\lim\limits_{n\to+\infty} V(X_n) = 0$.

   En appliquant l'inégalité de Markov à $(X_n - m)^2$, montrer que $X_n \xrightarrow{P} m$.
2. Soit $(Y_k)_{k\geq 1}$ une suite de variables i.i.d. suivant toutes une loi exponentielle de paramètre $\lambda > 1$. Pour tout $n \geq 1$, on pose $X_n = \prod_{k=1}^{n} Y_k$. À l'aide de l'inégalité de Markov, montrer la convergence en probabilité de la suite (X_n) vers la variable aléatoire certaine égale à 0.

Exercice 16.7 (★★)

On considère une suite $(X_n)_{n\in\mathbb{N}^*}$ de variables aléatoires mutuellement indépendantes suivant toutes la loi de Bernoulli de paramètre p. Pour tout $n \in \mathbb{N}^*$, on pose $Y_n = X_n + X_{n+1}$ et $\overline{Y_n} = \dfrac{1}{n}\sum_{k=1}^{n} Y_k$.

1. Soit $n < n'$. Donner une condition nécessaire et suffisante sur n et n' pour que Y_n et $Y_{n'}$ soient indépendantes.
2. Montrer que la suite $\overline{Y_n}$ converge en probabilité vers la variable certaine égale à $2p$.

Exercice 16.8 (★★★ - Convergence en probabilité d'une somme - ✍)

Soit $(X_n)_{n\in\mathbb{N}}$ et $(Y_n)_{n\in\mathbb{N}}$ deux suites de variables aléatoires et X et Y deux variables aléatoires, toutes définies sur un même espace probabilisé. On suppose que $X_n \xrightarrow{P} X$ et $Y_n \xrightarrow{P} Y$.

1. Soit $\varepsilon > 0$. Montrer que $[|X_n + Y_n - (X+Y)| \geq \varepsilon] \subset [|X_n - X| \geq \frac{\varepsilon}{2}] \cup [|Y_n - Y| \geq \frac{\varepsilon}{2}]$.
2. Montrer que $X_n + Y_n \xrightarrow{P} X + Y$.
3. On suppose de plus que toutes les variables aléatoires considérées sont à valeurs strictement positives. Montrer que $X_n \cdot Y_n \xrightarrow{P} X \cdot Y$.

1. Soit $\omega \in \Omega$ tel que :
$$|X_n(\omega) + Y_n(\omega) - (X(\omega) + Y(\omega))| \geq \varepsilon. \qquad (*)$$
Par l'inégalité triangulaire, on a :
$$\varepsilon \leq |X_n(\omega) - X(\omega)| + |Y_n(\omega) - Y(\omega)|.$$
Supposons que l'on ait $|X_n(\omega) - X(\omega)| < \dfrac{\varepsilon}{2}$ et $|Y_n(\omega) - Y(\omega)| < \dfrac{\varepsilon}{2}$. On aurait alors par somme :
$$|X_n(\omega) - X(\omega)| + |Y_n(\omega) - Y(\omega)| < \varepsilon$$
ce qui est contradictoire avec l'inégalité $(*)$. Ainsi on a $|X_n(\omega) - X(\omega)| \geq \dfrac{\varepsilon}{2}$ ou $|Y_n(\omega) - Y(\omega)| \geq \dfrac{\varepsilon}{2}$. D'où l'inclusion d'évènements :
$$[|X_n + Y_n - (X+Y)| \geq \varepsilon] \subset \left[|X_n - X| \geq \frac{\varepsilon}{2}\right] \cup \left[|Y_n - Y| \geq \frac{\varepsilon}{2}\right].$$
2. Par croissance d'une probabilité, on a :
$$\begin{aligned} P\left(|X_n + Y_n - (X+Y)| \geq \varepsilon\right) &\leq P\left(\left[|X_n - X| \geq \frac{\varepsilon}{2}\right] \cup \left[|Y_n - Y| \geq \frac{\varepsilon}{2}\right]\right) \\ &\leq P\left(|X_n - X| \geq \frac{\varepsilon}{2}\right) + P\left(|Y_n - Y| \geq \frac{\varepsilon}{2}\right). \end{aligned}$$
Puisque $X_n \xrightarrow{P} X$ et $Y_n \xrightarrow{P} Y$, on a :
$$\lim_{n\to+\infty} P\left(|X_n - X| \geq \frac{\varepsilon}{2}\right) = 0 = \lim_{n\to+\infty} P\left(|Y_n - Y| \geq \frac{\varepsilon}{2}\right).$$
Par théorème des gendarmes (une probabilité étant positive), on obtient que $\lim\limits_{n\to+\infty} P(|X_n + Y_n - (X+Y)| \geq \varepsilon)$ existe et vaut 0. Ceci étant vrai pour tout $\varepsilon > 0$, on a donc montré que $X_n + Y_n \xrightarrow{P} X + Y$.
3. Prenons le logarithme (possible car toutes les variables aléatoires considérées sont positives). On a :
$$\ln(X_n Y_n) = \ln(X_n) + \ln(Y_n).$$
Or $X_n \xrightarrow{P} X$ et $Y_n \xrightarrow{P} Y$ et le logarithme est continue sur $\mathbb{R}_+^*$, donc on a aussi $\ln(X_n) \xrightarrow{P} \ln(X)$ et $\ln(Y_n) \xrightarrow{P} \ln(Y)$. Par la question précédente, on obtient donc que :
$$\ln(X_n Y_n) = \ln(X_n) + \ln(Y_n) \xrightarrow{P} \ln(X) + \ln(Y) = \ln(XY).$$
Par composition par l'exponentielle qui est continue sur $\mathbb{R}$, on obtient bien que :
$$X_n \cdot Y_n \xrightarrow{P} X \cdot Y.$$

Exercice 16.9 (★★★★ - Convergence en probabilité d'un produit (ESCP 2017) - ✍)
Dans cet exercice, toutes les variables aléatoires sont définies sur le même espace probabilisé $(\Omega, \mathscr{A}, P)$.
Soit $(X_n)_n$, $(Y_n)_n$ deux suites de variables aléatoires et X, $Y)$ deux variables aléatoires.
On suppose que $(X_n)_n$ converge en probabilité vers X et que $(Y_n)_n$ converge en probabilité vers Y.

1. Vérifier que $X_nY_n - XY = (X_n - X)\,Y + (Y_n - Y)\,X + (X_n - X)\,(Y_n - Y)$.

Soit $\varepsilon > 0$. On pose :

$$A_n = [|X_nY_n - XY| > \varepsilon]\,, B_n = [|X_n - X|\,|Y| > \varepsilon/3]\,,$$
$$C_n = [|Y_n - Y|\,|X| > \varepsilon/3]\,, D_n = [|X_n - X|\,|Y_n - Y| > \varepsilon/3]\,.$$

2. Montrer que $P\,(A_n) \leqslant P\,(B_n) + P\,(C_n) + P\,(D_n)$.
3. Soit $\delta > 0$.
   (a) Montrer qu'il existe un réel $A > 0$ tel que pour tout $t > A, P(|Y| > t) < \frac{\delta}{2}$.
   (b) Montrer que pour tout $t > 0, P\,(B_n) \leqslant P\left(|X_n - X| > \frac{\varepsilon}{3t}\right) + P(|Y| > t)$.
   (c) Soit $t > A$. Montrer qu'il existe un entier N_0 tel que pour tout $n \geq N_0$,
   $$P\left(|X_n - X| > \frac{\varepsilon}{3t}\right) < \frac{\delta}{2}.$$
   (d) En déduire que $\lim\limits_{n\to+\infty} P\,(B_n) = 0$.

On montrerait de même et on admet ici que $\lim\limits_{n\to+\infty} P\,(C_n) = 0$.

4. Montrer que $P\,(D_n) \leqslant P\left(|X_n - X| > \sqrt{\frac{\varepsilon}{3}}\right) + P\left(|Y_n - Y| > \sqrt{\frac{\varepsilon}{3}}\right)$.
   En déduire que $\lim\limits_{n\to+\infty} P\,(D_n) = 0$.
5. Montrer que la suite (X_nY_n) converge en probabilité vers XY.

1. C'est un simple calcul...
2. Si on a à la fois $|X_n - X\|Y| \leqslant \frac{\varepsilon}{3}, |Y_n - Y|\,|X| \leqslant \frac{\varepsilon}{3}$ et $|X_n - X| \cdot |Y_n - Y| \leqslant \frac{\varepsilon}{3}$.
   Alors, par l'inégalité triangulaire, on a
   $$\begin{aligned} |X_nY_n - XY| &= |(X_n - X)\,Y + (Y_n - Y)\,X + (X_n - X)\,(Y_n - Y)| \\ &\leqslant |X_n - X|\,|Y| + |Y_n - Y|\,|X| + |X_n - X| \cdot |Y_n - Y| \\ &\leqslant \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} \leqslant \varepsilon. \end{aligned}$$
   Autrement dit, nous venons de prouver que $\overline{B_n} \cap \overline{C_n} \cap \overline{D_n} \subset [|X_nY_n - XY| \leqslant \varepsilon]$.
   En passant aux événements contraires, il vient
   $$|X_nY_n - XY| > \varepsilon] \subset B_n \cup C_n \cup D_n.$$
   Et donc
   $$P\,(|X_nY_n - XY| > \varepsilon) \leqslant P\,(B_n \cup C_n \cup D_n) \leqslant P\,(B_n) + P\,(C_n) + P\,(D_n)\,.$$
3. (a) On a $P(|Y| > t) = 1 - P(|Y| \leqslant t)$.

Or, $|Y|$ est une variable aléatoire sur $(\Omega, \mathscr{A}, P)$, donc sa fonction de répartition tend vers 1 en $+\infty$ et donc $\lim_{t \to +\infty} P(|Y| > t) = 0$. Et donc [1] il existe $A > 0$ tel que pour $t \geqslant A, g(x) < \frac{\delta}{2}$.
Et donc pour $t \geqslant A, 0 \leqslant P(|Y| > t) \leqslant g(t) < \frac{\delta}{2}$.

(b) Comme à la question 2, on a

$$[|Y| \leqslant t] \cap \left[|X_n - X| \leqslant \frac{\varepsilon}{3t}\right] \subset \underbrace{\left[|X_n - X|\,|Y| \leqslant \frac{\varepsilon}{3}\right]}_{=\overline{B_n}}.$$

Et donc

$$P(B_n) \leqslant P\left(|X_n - X| > \frac{\varepsilon}{3t}\right) + P(|Y| > t).$$

(c) C'est la définition de $X_n \xrightarrow{P} X$: puisque $\lim_{n \to +\infty} P\left(|X_n - X| > \frac{\varepsilon}{3t}\right) = 0$, il existe donc $N_0 \in \mathbf{N}$ tel pour $n \geqslant N_0, P\left(|X_n - X| > \frac{\varepsilon}{3t}\right) < \frac{\delta}{2}$.

(d) Soit $\delta > 0$ fixé. Et soit alors A comme à la question 3.a, et $t > A$.
On a alors, pour tout $n \in \mathbf{N}, P(B_n) \leqslant P\left(|X_n - X| > \frac{\varepsilon}{3t}\right) + \frac{\delta}{2}$.
Et en particulier, si N_0 est comme à la question 3.c, pour $n \geqslant N_0$, on a $0 \leqslant P(B_n) < \delta$.
Autrement dit, nous venons de prouver que pour tout $\delta > 0$, il existe $N_0 \in \mathbf{N}$ tel que pour
$n \geqslant N_0, |P(B_n)| \leqslant \delta$.
Nous reconnaissons là la définition de $\lim_{n \to +\infty} P(B_n) = 0$.

4. Toujours comme à la question 2, on a, par passage aux événements contraires,

$$\left[|X_n - X| \leqslant \sqrt{\frac{\varepsilon}{3}}\right] \cap \left[|Y_n - Y| \leqslant \sqrt{\frac{\varepsilon}{3}}\right] \subset \underbrace{\left[|X_n - X|| | Y_n - Y \Big| \leqslant \frac{\varepsilon}{3}\right]}_{=\overline{D_n}}$$

Et donc
$P(D_n) \leqslant P\left(\left[|X_n - X| > \sqrt{\frac{\varepsilon}{3}}\right] \cup \left[|Y_n - Y| > \sqrt{\frac{\varepsilon}{3}}\right]\right) \leqslant P\left(\left[|X_n - X| > \sqrt{\frac{\varepsilon}{3}}\right]\right) + P\left(\left[|Y_n - Y| > \sqrt{\frac{\varepsilon}{3}}\right]\right)$.
Puisque $X_n \xrightarrow{P} X$, on a $\lim_{n \to +\infty} P\left(\left[|X_n - X| > \sqrt{\frac{\varepsilon}{3}}\right]\right) = 0$.
Et de même, $Y_n \xrightarrow{P} Y$ et donc $\lim_{n \to +\infty} P\left(\left[|Y_n - Y| > \sqrt{\frac{\varepsilon}{3}}\right]\right)$.
On en déduit que $\lim_{n \to +\infty} P(D_n) = 0$.

5. D'après la question 2 , et les résultats de 3.(d) et 4 , on a

$$0 \leqslant P(A_n) \leqslant \underbrace{P(B_n) + P(C_n) + P(D_n)}_{n \to +\infty}$$

Et donc par le théorème des gendarmes, $\lim_{n \to +\infty} P(A_n) = 0$.
Autrement dit, $X_n Y_n \xrightarrow{P} XY$.

Convergence en loi

Exercice 16.10 (★)

Soit (X_n) une suite de variables aléatoires discrètes telles que $X_n \hookrightarrow \mathscr{U}(\llbracket 1, n \rrbracket)$. Étudier la convergence en loi de la suite (X_n).

Exercice 16.11 (★★)

Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires telle que pour tout $n \in \mathbb{N}^*$, $X_n \hookrightarrow \mathscr{P}\left(\frac{1}{n}\right)$.
Montrer que la suite $(X_n)_{n\in\mathbb{N}^*}$ converge en loi vers une variable aléatoire X dont on précisera la loi.
Montrer qu'il s'agit également d'une convergence en probabilité.

Exercice 16.12 (★★)

Une urne contient des boules numérotées de 1 à n ($n \geq 2$). On tire avec remise une boule de l'urne jusqu'à obtenir un numéro supérieur ou égal au premier numéro tiré. Soit X_n la variable égale au nombre de tirages effectués.

1. Montrer que pour tout $(n, k) \in (\mathbb{N} \setminus \{0, 1\})^2$, on a $P(X_n = k) = \dfrac{1}{n}\displaystyle\sum_{j=0}^{n-1}\left(\left(\frac{j}{n}\right)^{k-2} - \left(\frac{j}{n}\right)^{k-1}\right)$.

2. À l'aide des sommes de Riemann, déterminer $\displaystyle\lim_{n\to+\infty} P(X_n = k)$ pour tout $k \in \mathbb{N}$, $k \geq 2$.

3. En déduire que $(X_n)_{n\geq 2}$ converge en loi vers une variable aléatoire dont on précisera la loi.

Exercice 16.13 (★★★ - QSP HEC 2017)

Soit $(X_i)_{i\in\mathbb{N}^*}$ une suite de variables aléatoires réelles, indépendantes, définies sur un même espace probabilisé $(\Omega, \mathscr{A}, P)$ et soit $(p_i)_{i\in\mathbb{N}^*}$ une suite de réels de $]0, 1[$ telle que pour tout $i \in \mathbb{N}^*$, la variable aléatoire X_i suit une loi géométrique de paramètre p_i. On pose pour tout $i \in \mathbb{N}^*$: $q_i = 1 - p_i$.
On pose pour tout $n \in \mathbb{N}^*$: $Z_n = \min(X_1, X_2, \ldots, X_n)$.

1. Calculer pour tout $k \in \mathbb{N}^*$, la probabilité $P(Z_n \geq k)$. Quelle est la loi de Z_n ?

2. On suppose que pour tout $i \in \mathbb{N}^*$, on a $p_i = \dfrac{1}{(i+1)^2}$.

   Montrer que la suite de variables aléatoires $(Z_n)_{n\in\mathbb{N}^*}$ converge en loi vers une variable aléatoire dont on précisera la loi.

> 1. Notons tout d'abord que $Z_n(\Omega) = \mathbb{N}^*$. Pour tout $k \in \mathbb{N}^*$, on a :
>
> $$\begin{aligned} P(X_i \geq k) &= \sum_{j=k}^{+\infty} P(X_i = j) = \sum_{j=k}^{+\infty}(1-p_i)^{j-1}p_i \\ &= p_i \underbrace{(1-p_i)^{k-1}}_{\text{1er terme de la série géom.}} \frac{1}{1-(1-p_i)} = (1-p_i)^{k-1} \end{aligned}$$
>
> On obtient alors :
>
> $$\begin{aligned} P(Z_n \geq k) &= P(X_1 \geq k, \ldots, X_n \geq k) \\ &= P(X_1 \geq k) \ldots P(X_n \geq k) \quad \text{par indép. des } X_i \\ &= \left(\prod_{i=1}^{n}(1-p_i)\right)^{k-1} \end{aligned}$$

On en déduit que pour tout $k \in \mathbb{N}^*$ (en posant $Q_n = \prod_{i=1}^{n}(1-p_i)$) :

$$P(Z_n = k) = P(Z_n \geq k) - P(Z_n \geq k+1) = Q_n^{k-1} - Q_n^k = Q_n^{k-1}(1-Q_n).$$

Ainsi Z_n suit une loi géométrique de paramètre $1 - Q_n = 1 - \prod_{i=1}^{n}(1-p_i)$.

2.

> **Rappel.**
> Il s'agit ici d'une suite de variables aléatoires à valeurs dans $\mathbb{N}$. Pour montrer la convergence en loi de (Z_n), on étudie donc la convergence ponctuelle, c'est-à-dire $\lim_{n\to+\infty} P(Z_n = k)$ pour k fixé dans $\mathbb{N}$.

Prenons $k \in \mathbb{N}^*$ fixé. Pour tout $n \in \mathbb{N}$, on a :

$$P(Z_n = k) = Q_n^{k-1} - Q_n^k$$

avec $Q_n = \prod_{i=1}^{n}(1-p_i)$. En substituant par les valeurs de l'énoncé, on a :

$$\begin{aligned} Q_n &= \prod_{i=1}^{n}\left(1 - \frac{1}{(i+1)^2}\right) = \prod_{i=1}^{n}\left(\frac{(i+1)^2-1}{(i+1)^2}\right) = \prod_{i=1}^{n}\left(\frac{i(i+2)}{(i+1)(i+1)}\right) \\ &= \frac{1\times 3}{2\times 2}\frac{2\times 4}{3\times 3}\frac{3\times 5}{4\times 4}\cdots\frac{(n-1)(n+1)}{n\times n}\frac{n(n+2)}{(n+1)(n+1)} = \frac{n+2}{2(n+1)} \end{aligned}$$

après télescopage. On en déduit que $\lim_{n\to+\infty} Q_n = \frac{1}{2}$. On a ainsi que pour tout $k \in \mathbb{N}^*$:

$$\lim_{n\to+\infty} P(Z_n = k) = \lim_{n\to+\infty} (1-Q_n)^{k-1}Q_n = \left(1 - \frac{1}{2}\right)^{k-1}\frac{1}{2}.$$

(Z_n) converge donc en loi vers une loi géométrique $\mathscr{G}(1/2)$.

Exercice 16.14 (★)

Soit (X_n) une suite de variables aléatoires telles que $X_n \hookrightarrow \mathscr{N}(m, \frac{\sigma^2}{n})$.

1. Déterminer la fonction de répartition de X_n.
2. Étudier la convergence en loi de la suite (X_n).
3. Étudier la convergence en probabilité de la suite (X_n).

Exercice 16.15 (★★)

Pour $n \in \mathbb{N}^*$, soit f_n la fonction définie sur $\mathbb{R}$ par : $\quad \forall t \in \mathbb{R}, \quad f_n(t) = \frac{1}{\pi}\frac{n}{1+n^2t^2}$.

1. Montrer que f_n est une densité de probabilité.
2. Soit (X_n) une suite de variables aléatoires telles que X_n admet f_n pour densité.
   (a) Déterminer la fonction de répartition de X_n.
   (b) Montrer que (X_n) converge en loi vers la variable certaine égale à 0.
   (c) Montrer qu'il s'agit également d'une convergence en probabilité.

Exercice 16.16 (★★)
Soient (X_n) une suite de variables aléatoires indépendantes suivant la loi uniforme sur $[0,1]$. On pose pour tout $n \geq 1$, $Y_n = \min(X_1, \cdots, X_n)$.

1. Déterminer la fonction de répartition F_n de Y_n.
2. Montrer que (Y_n) converge en probabilité vers la variable constante égale à 0.
3. Montrer que la suite (nY_n) converge en loi vers une variable suivant une loi exponentielle.

1. On vérifie que :
$$F_n : x \in \mathbb{R} \mapsto 1 - (1 - F_{X_1}(x))^n = \begin{cases} 0 & \text{si } x \leq 0, \\ 1 - (1-x)^n & \text{si } x \in]0,1[, \\ 1 & \text{si } x \geq 1. \end{cases}$$
On remarquera en passant que F_n est continue sur $\mathbb{R}$ et $\mathscr{C}^1$ sauf éventuellement en 0 et 1 puisque F_{X_1} l'est, ce qui montre que Y_n est à densité.

2. Pour tout $\varepsilon > 0$, on a :
$$P(|Y_n - 0| \geq \varepsilon) \underbrace{=}_{Y_n \geq 0} P(Y_n \geq \varepsilon) = 1 - P(Y_n < \varepsilon) = 1 - P(Y_n \leq \varepsilon)$$
puisque Y_n est à densité. Ainsi, on a :
$$P(|Y_n - 0| \geq \varepsilon) = 1 - F_n(\varepsilon).$$
On a deux cas à traiter :
   - Si $\varepsilon \geq 1$, alors :
   $$P(|Y_n - 0| \geq \varepsilon) = 1 - F_n(\varepsilon) = 1 - 1 = 0 \to 0.$$
   - Si $\varepsilon \in]0,1[$, alors :
   $$P(|Y_n - 0| \geq \varepsilon) = 1 - F_n(\varepsilon) = (1-\varepsilon)^n \underset{n \to +\infty}{\to} 0$$
   car $1 - \varepsilon \in]0,1[$.

   Dans tous les cas, on a bien montré que $P(|Y_n - 0| \geq \varepsilon) \underset{n \to +\infty}{\to} 0$. Ainsi Y_n converge en probabilité vers 0.

3. Pour tout $n \geq 1$, posons $Z_n = nY_n$.

   Étape 1 : loi de Z_n.

Pour tout $x \in \mathbb{R}$, on a :

$$F_{Z_n}(x) = P(Z_n \leq x) = P(Y_n \leq \frac{x}{n}) = F_{Y_n}(x/n) = \begin{cases} 0 & \text{si } x/n \leq 0, \\ 1 - (1 - \frac{x}{n})^n & \text{si } x/n \in]0, 1[, \\ 1 & \text{si } x/n \geq 1. \end{cases}$$

$$= \begin{cases} 0 & \text{si } x \leq 0, \\ 1 - (1 - \frac{x}{n})^n & \text{si } x \in]0, n[, \\ 1 & \text{si } x \geq n. \end{cases}$$

Étape 2 : limite ponctuelle.

Soit $x \in \mathbb{R}$ **fixé**. On a deux cas à considérer :

- Si $x \leq 0$, $F_{Z_n}(x) = 0 \underset{n\to+\infty}{\longrightarrow} 0$.
- Supposons $x > 0$. Pour tout $n \in \mathbb{N}$ tel que $n > x$, on a :

$$F_{Z_n}(x) = 1 - (1 - \frac{x}{n})^n.$$

On cherche la limite de $(1 - \frac{x}{n})^n = \exp\left(n \ln(1 - \frac{x}{n})\right)$. En utilisant les équivalents usuels :

$$n \ln(1 - \frac{x}{n}) \underset{n\to+\infty}{\sim} n\left(-\frac{x}{n}\right) = -x \underset{n\to+\infty}{\longrightarrow} -x.$$

Par composition des **limites** par l'exponentielle qui est continue, on obtient :

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

On a donc montré que :

$$F_{Z_n}(x) \underset{n\to+\infty}{\longrightarrow} \begin{cases} 0 & \text{si } x \leq 0, \\ 1 - e^{-x} & \text{si } x > 0. \end{cases}$$

On reconnait ici la fonction de répartition de la loi exponentielle $\mathscr{E}(1)$. Ainsi (Z_n) converge en loi vers une variable Z suivant une loi $\mathscr{E}(1)$.

Exercice 16.17 (★★)

Considérons la fonction f définie sur $\mathbb{R}$ par : $\forall x \in \mathbb{R},\ f(x) = \dfrac{e^x}{(1+e^x)^2}.$

1. Montrer que f est une densité de probabilité.
2. Soit à présent $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires i.i.d. de densité f.
   (a) Déterminer la fonction de répartition des X_i.
   (b) En déduire la fonction de répartition de $M_n = \max(X_1, \dots, X_n)$.
   (c) Étudier la convergence en loi de la suite $(M_n - \ln(n))_{n\in\mathbb{N}}$.

Exercice 16.18 (★★★)

Soit (X_n) une suite de variables aléatoires discrètes telles que $n \times X_n$ suit une loi uniforme sur $[\![0, n]\!]$.

1. Rappelons que si `x` et `y` sont des vecteurs de même dimension, la commande `plt.step(x,y)` trace la fonction en escalier qui vaut `y[k]` sur l'intervalle $[\mathtt{x}[\mathtt{k}-1], \mathtt{x}[\mathtt{k}][$.

On considère le code `Python` suivant ainsi que la figure qu'il renvoie :

```python
n = np.array([4,10,30])
for i in range(3):
    x = np.linspace(0,1,n[i]+1)
    y = np.arange(0,n[i]+1)/(n[i]+1)
    plt.subplot(1,3,i+1)
    plt.step(x,y)
plt.show()
```

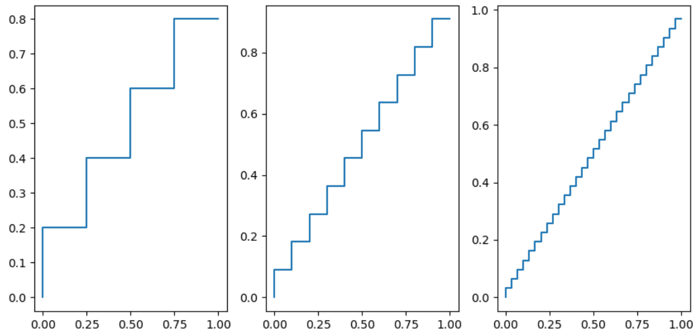

FIGURE 1 - *Fonctions en escaliers.*

Que peut-on conjecturer ?

2. Étudier la convergence en loi de (X_n).

Exercice 16.19 (★★)

On considère une suite (X_n) de variables aléatoires indépendantes suivant la loi de Poisson de paramètre 1, et on pose $S_n = X_1 + \cdots + X_n$.

1. Exprimer $P(S_n \leq n)$ sous forme d'une somme.

2. En utilisant le théorème central limite, montrer que : $\displaystyle\sum_{k=0}^{n} \frac{n^k}{k!} \underset{n\to+\infty}{\sim} \frac{e^n}{2}$.

Exercice 16.20 (★★★ - QSP HEC 2012)

Soit $(U_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires indépendantes, définies sur le même espace probabilisé $(\Omega, \mathscr{A}, P)$, de même loi uniforme sur $]0,1]$.

On pose, pour tout $n \in \mathbb{N}^*$:

$$X_n = \prod_{i=1}^{n} U_i^{\frac{1}{n}} \qquad \text{et} \qquad Y_n = (eX_n)^{\sqrt{n}}.$$

Montrer que la suite de variables aléatoires $(\ln(Y_n))_{n\in\mathbb{N}^*}$ converge en loi vers une variable aléatoire dont on précisera la loi.

Exercice 16.21 (★★★★ - Convergence en probabilité et convergence en loi (Oral HEC) - ✍)

Soient $(X_n)_{n\geq 1}$ et X des variables aléatoires définies sur un espace probabilisé $(\Omega, \mathscr{A}, P)$ telles que $X_n \xrightarrow{P} X$. On note F_n la fonction de répartition de X_n et F la fonction de répartition de X.

Soit x un point de continuité de F, et soit $\delta > 0$ fixé.

1. Montrer qu'on peut choisir $\varepsilon > 0$ tel que $F(x-\varepsilon) > F(x) - \delta$ et $F(x+\varepsilon) < F(x) + \delta$.

2. Montrer que : $[X_n \leq x] \subset [X \leq x+\varepsilon] \cup [|X_n - X| > \varepsilon]$.

   Montrer de même que : $[X \leq x-\varepsilon] \subset [X_n \leq x] \cup [|X_n - X| > \varepsilon]$.

3. En déduire que :

$$F_n(x) \leq F(x) + \delta + P(|X_n - X| > \varepsilon) \qquad \text{et} \qquad F(x) - \delta \leq F_n(x) + P(|X_n - X| > \varepsilon).$$

4. Montrer que $X_n \xrightarrow{\mathscr{L}} X$.

1. Puisque x est un point de continuité, on a par définition de la continuité d'une fonction en un point l'existence de $\varepsilon > 0$ tel que pour tout $y \in \mathbb{R}$:
$$|y - x| \leq \varepsilon \quad \Rightarrow \quad |F(y) - F(x)| < \delta.$$
Notons que l'on peut dans la définition de la limite d'une fonction en un point mettre des inégalités strictes ou larges comme on le souhaite.
D'où ici pour tout $y \in [x - \varepsilon, x + \varepsilon]$, on a :
$$F(x) - \delta < F(y) < F(x) + \delta.$$
En particulier pour $y = x \pm \varepsilon$, on obtient donc :
$$F(x - \varepsilon) > F(x) - \delta \quad \text{et} \quad F(x + \varepsilon) < F(x) + \delta.$$

2. Soit $\omega \in \Omega$ tel que $X_n(\omega) \leq x$. Supposons que :
$$X(\omega) > x + \varepsilon \quad \text{et } |X_n - X| \leq \varepsilon.$$
Alors on aurait :
$$X(\omega) = X_n(\omega) - (X_n(\omega) - X(\omega)) \geq X_n(\omega) - |X_n(\omega) - X(\omega)| > x + \varepsilon - \varepsilon = x$$
ce qui est contradictoire. Donc on a :
$$X(\omega) \leq x + \varepsilon \quad \text{ou } |X_n - X| > \varepsilon.$$
D'où l'inclusion d'évènements $[X_n \leq x] \subset [X \leq x + \varepsilon] \cup [|X_n - X| > \varepsilon]$. La deuxième inclusion découle de la précédente en permutant les rôles de X et X_n et en substituant $x - \varepsilon$ à x.

3. Par croissance et sous-additivité d'une probabilité, on a :
$$P(X_n \leq x) \leq P(X \leq x + \varepsilon) + P(|X_n - X| > \varepsilon)$$
et
$$P(X \leq x - \varepsilon) \leq P(X_n \leq x) + P(|X_n - X| > \varepsilon).$$
D'où avec la question 1. :
$$F_n(x) \leq F(x + \varepsilon) + P(|X_n - X| > \varepsilon) \leq F(x) + \delta + P(|X_n - X| > \varepsilon)$$
et
$$F(x) - \delta < F(x - \varepsilon) \leq F_n(x) + P(|X_n - X| > \varepsilon).$$

4. Puisque (X_n) converge en probabilité vers X, il existe un rang $N \in \mathbb{N}$ tel que pour tout $n \in \mathbb{N}$, on a :
$$n \geq N \quad \Rightarrow \quad 0 \leq P(|X_n - X| > \varepsilon) \leq \delta.$$
Avec les inégalités de la question précédente, on obtient que pour tout $n \geq N$, on a :
$$F(x) - 2\delta \leq F_n(x) \leq F(x) + 2\delta.$$

> Ainsi on a montré que :
>
> $$\forall \delta > 0,\ \exists N \in \mathbb{N},\ \forall n \in \mathbb{N},\ n \geq N \ \Rightarrow\ |F_n(x) - F(x)| \leq 2\delta.$$
>
> On reconnait ici la définition de limite (avec 2δ à la place de δ, ce qui n'a pas d'importance[a]). Ainsi on a bien que $\lim\limits_{n\to} F_n(x) = F(x)$. Ceci étant vrai pour tout x point de continuité de F, on a donc bien la convergence en loi de (X_n) vers X.
>
> [a] on pourrait effectuer un changement de variables pour se ramener à δ'

Approximation

Exercice 16.22 (★)

Soit S une variable aléatoire de loi $\mathscr{B}(45\,;0.7)$. Donner une approximation de la probabilité $P(28 < S \leq 39)$ sans correction de continuité puis avec correction de continuité.
Un calcul détaillé à l'aide de loi binomiale donne $P(28 < S \leq 39) \approx 0,8332$.

Exercice 16.23 (★★)

On effectue des lancers successifs d'un dé équilibré. On cherche à déterminer le nombre n de lancers nécessaires pour garantir avec moins de 5% d'erreur que la fréquence d'apparition du 1 sera $\dfrac{1}{6} \pm 0.01$.

1. **Méthode 1.** À l'aide de l'inégalité de Bienaymé Tchebychev.
   (a) Soit (Y_n) une suite de variables aléatoires mutuellement indépendantes, de même loi et admettant un moment d'ordre 2. On pose $S_n = \sum\limits_{k=1}^{n} Y_k$. Prouver que :
   $$\forall a > 0, \quad P\left(\left|\frac{S_n}{n} - E(Y_1)\right| \geq a\right) \leq \frac{V(Y_1)}{na^2}.$$
   (b) En déduire le nombre n de lancers nécessaires.
2. **Méthode 2.** À l'aide du théorème central limite, déterminer approximativement ce nombre n.
3. Un calcul détaillé à l'aide de la loi binomiale prouve qu'il faut $n \geq 5395$. Que dire des résultats obtenus précédemment ?

Exercice 16.24 (★★)

Dans le désert, une Renault 4L crève en moyenne tous les 4000 km. On considère donc qu'à chaque kilomètre, la probabilité de crever est de $\frac{1}{4000}$. Un équipage s'inscrit au 4L Trophy, rallye de 6000 km dans le désert. Il aimerait savoir combien de roues de secours emporter pour avoir moins de 10% de chances de manquer de roues de secours. Deux roues de secours sont-elles suffisantes ? Trois ? On donne $e^{-3/2} \approx 0.22$.

Exercice 16.25 (★★)

Une entreprise compte 300 employés. Chacun d'eux téléphone en moyenne 6 minutes par heure. Combien de lignes téléphonique doit faire installer l'entreprise afin que la probabilité que toutes les lignes soient occupées en même temps soit inférieure ou égale à 0.025 ?

Exercice 16.26 (★★★★ - QSP HEC 2018)

On considère la fonction `Python` suivante :

```
def bn(N,n,p):
    X = rd.geometric(p,[N,n])
    k = 0
    for i in range(N):
        Y = np.sum(X[i,:])
        if Y >=n/p :
            k = k+1
    f = k/N
    return f
```

1. De quel nombre réel `bn(N,1,0.4)` fournit-il une valeur approchée lorsque N est grand ?
2. Donner une valeur approximative de `bn(10000,10000,0.4)`.

1. Supposons que $n = 1$. Les variables Y_i obtenues à chaque passage dans la boucle `for` pour $i = 1, \dots, N$ sont donc i.i.d. et suivent toutes une loi $\mathscr{G}(p)$. On calcule alors, à l'aide d'une boucle `if`, la fréquence de réalisation de l'évènement $[Y_i >= 1/p]$. Par la loi faible des grands nombres (les Y_i admettant bien une même espérance $\frac{1}{p}$ et une même variance $\frac{q}{p^2}$), cette fréquence tend lorsque $N \to +\infty$ vers la probabilité théorique $P(Y \geq \frac{1}{p})$ où $Y \hookrightarrow \mathscr{G}(p)$.

   D'après la discussion précédente, le réel `bn(N,1,0.4)` est donc une valeur approchée de $P(Y \geq \frac{1}{0.4}) = P(Y \geq 2.5) = P(Y \geq 3)$ lorsque N est grand. On peut calculer cette valeur (avec $p = 0.4$) :

$$P(Y \geq 3) = \sum_{k=3}^{+\infty} P(Y = k) = \sum_{k=3}^{+\infty} (1-p)^{k-1} p = p \sum_{k=2}^{+\infty} (1-p)^k = p(1-p)^2 \frac{1}{1-(1-p)} = (1-p)^2 = 0.36.$$

2. Supposons cette fois n quelconque. Dans ce programme, on créée de façon indépendante et identiquement distribuées des variables aléatoires $X_{i,j}$ (avec $1 \leq i \leq N$ et $1 \leq j \leq n$) suivant toutes une loi $\mathscr{G}(p)$. Pour tout $1 \leq i \leq N$, on définit alors $Y_i = \sum_{j=1}^{n} X_{i,j}$. Les variables Y_i suivent toutes la même loi de probabilité et sont indépendantes par le lemme de coalition. On calcule alors, à l'aide d'une boucle `if`, la fréquence de réalisation de l'évènement $[Y_i >= n/p]$. Par la loi faible des grands nombres (les Y_i admettant bien une même espérance $\frac{n}{p}$ et une même variance $\frac{nq}{p^2}$), cette fréquence tend lorsque $N \to +\infty$ vers la probabilité théorique $P(S_n \geq \frac{n}{p})$, où l'on note $S_n = X_1 + \dots + X_n$ avec X_i des variables i.i.d. suivant toutes une loi $\mathscr{G}(p)$.

   Or on a :

$$P(S_n \geq \frac{n}{p}) = P(S_n - E(S_n) \geq 0) = P\left(\frac{S_n - E(S_n)}{\sqrt{V(S_n)}} \geq 0\right) \xrightarrow[n \to +\infty]{} P(S \geq 0)$$

   où $S \hookrightarrow \mathscr{N}(0,1)$, par le théorème limite central. Enfin on a $P(S \geq 0) = 1 - \Phi(0) = \frac{1}{2}$.

   On peut donc conclure, étant donné que N et n sont grands (égaux à 10000) que la valeur approximative de `bn(10000,10000,0.4)` est $\frac{1}{2}$.