L2 MASS Université Lyon 1, Probabilités-statistique 2
Année Universitaire 2020 − 2021

Correction TD 2

Exercice 1 :

1. On peut écrire $X = Y_1 + Y_2$. En effet, vérifions que $\forall \omega \in \Omega, X(\omega) = Y_1(\omega) + Y_2(\omega)$. Soit $\omega \in \Omega$,
   - Si $X(\omega) > a$. Alors $\mathbb{1}_{X(\omega)>A} = 1$ et $\mathbb{1}_{X(\omega)\leq a} = 0$. Donc $Y_1(\omega) = X(\omega)\mathbb{1}_{X(\omega)>a} = X(\omega)$ et $Y_2(\omega) = X(\omega)\mathbb{1}_{X(\omega)\leq a} = 0$. Dans ce cas là, on a bien $X(\omega) = Y_1(\omega) + Y_2(\omega)$.
   - Si $X(\omega) \leq a$. Alors $\mathbb{1}_{X(\omega)>a} = 0$ et $\mathbb{1}_{X(\omega)\leq a} = 1$. Donc $Y_1(\omega) = X(\omega)\mathbb{1}_{X(\omega)>a} = 0$ et $Y_2(\omega) = X(\omega)\mathbb{1}_{X(\omega)\leq a} = X(\omega)$. Dans ce cas là, on a bien $X(\omega) = Y_1(\omega) + Y_2(\omega)$.

2. Remarquons d'abord que :

$$\begin{aligned}
&Y_1 = X\mathbb{1}_{X>a} \geq a\mathbb{1}_{X>a}\\
&Y_2 = X\mathbb{1}_{X\leq a} \geq 0 \text{ car } X \geq 0 \text{ p.s.}
\end{aligned}$$

   On peut donc écrire

$$\begin{aligned}
\mathbb{E}[X] &= \mathbb{E}[Y_1 + Y_2]\\
&= \mathbb{E}[Y_1] + \mathbb{E}[Y_2] \text{ par linéarité de l'espérance}\\
&= \mathbb{E}\Big[X\mathbb{1}_{X>a}\Big] + \mathbb{E}\Big[X\mathbb{1}_{X\leq a}\Big]\\
&\geq a\mathbb{E}[\mathbb{1}_{X>a}] + 0 \text{ par croissance de l'espérance}\\
&\geq a\mathbb{P}(X > a) \text{ (Espérance d'une indicatrice d'un évenement)}
\end{aligned}$$

   On obtient alors

$$\mathbb{P}(X > a) \leq \frac{\mathbb{E}[X]}{a}$$

Exercice 2 :
On peut utiliser l'inégalité de Markov avec $Y = (X - \mathbb{E}[X])^2$, qui vérifie $Y \geq 0$ p.s. et $\mathbb{E}[Y] = \mathrm{var}(X) < +\infty$, on trouve

$$\forall b > 0, \mathbb{P}(Y > b) \leq \frac{\mathbb{E}[Y]}{b},$$

en remplaçant Y par $(X - \mathbb{E}[X])^2$, et en prenant $b = a^2$, on obtient

$$\forall a > 0, \mathbb{P}\Big((X - \mathbb{E}[X])^2 > a^2\Big) \leq \frac{\mathbb{E}\Big[(X - \mathbb{E}[X])^2\Big]}{a^2},$$

ce qui donne

$$\forall a > 0, \mathbb{P}\Big(|X - \mathbb{E}[X]| > a\Big) \leq \frac{\mathrm{var}(X)}{a^2}.$$

Exercice 3 :
Soit $\epsilon > 0$. On veut montrer que $\mathbb{P}\Big(|\bar{X}_n - m| > \epsilon\Big) \to +\infty$. L'inégalité précédente donne que :

$$\mathbb{P}\Big(|\bar{X}_n - \mathbb{E}[\bar{X}_n]| > \epsilon\Big) \leq \frac{\mathrm{var}(\bar{X}_n)}{\epsilon^2}.$$

Or,

$$\begin{aligned}
\mathbb{E}[\bar{X}_n] &= \mathbb{E}\Big[\frac{1}{n}\sum_{i=1}^{n} X_i\Big] \\
&= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\Big[X_i\Big] \\
&= \frac{1}{n}\sum_{i=1}^{n} m = m \\
\mathrm{var}(\bar{X}_n) &= \mathrm{var}\Big(\frac{1}{n}\sum_{i=1}^{n} X_i\Big) \\
&= \frac{1}{n^2}\,\mathrm{var}\Big(\sum_{i=1}^{n} X_i\Big) \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{var}(X_i) \text{ par indépendance} \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{\sigma^2}{n}.
\end{aligned}$$

Ce qui donne finalement

$$\mathbb{P}\Big(|\bar{X}_n - m| > \epsilon\Big) \leq \frac{\sigma^2}{n\epsilon^2} \xrightarrow{n\to+\infty} 0$$

Cela achève la preuve, et on a bien $X_n \xrightarrow{\mathbb{P}} m$.

Exercice 4 :
On peut écrire :

$$\begin{aligned}
\mathbb{E}\Big[(\hat{T}_n - \theta)^2\Big] &= \mathbb{E}\left[\Big(\hat{T}_n - \mathbb{E}[\hat{T}_n] + \mathbb{E}[\hat{T}_n] - \theta\Big)^2\right] \\
&= \mathbb{E}\left[\Big(\hat{T}_n - \mathbb{E}[\hat{T}_n]\Big)^2 + \Big(\mathbb{E}[\hat{T}_n] - \theta\Big)^2 + 2\Big(\hat{T}_n - \mathbb{E}[\hat{T}_n]\Big)\Big(\mathbb{E}[\hat{T}_n] - \theta\Big)\right] \\
&= \mathbb{E}\left[\Big(\hat{T}_n - \mathbb{E}[\hat{T}_n]\Big)^2\right] + \mathbb{E}\left[\Big(\mathbb{E}[\hat{T}_n] - \theta\Big)^2\right] + 2\mathbb{E}\left[\hat{T}_n - \mathbb{E}[\hat{T}_n]\right]\Big(\mathbb{E}[\hat{T}_n] - \theta\Big) \\
&= \mathrm{var}(\hat{T}_n) + b_n^2 + 2\bigg(\mathbb{E}[\hat{T}_n] - \mathbb{E}[\hat{T}_n]\bigg)\Big(\mathbb{E}[\hat{T}_n] - \theta\Big) \\
&= \mathrm{var}(\hat{T}_n) + b_n^2
\end{aligned}$$

ce qui achève la preuve de la décomposition du risque quadratique. On en déduit qu'un estimateur asymptotiquement sans biais (i.e. tel que $b_n \xrightarrow{n\to+\infty} 0$) qui vérifirait $\mathrm{var}(\hat{T}_n) \xrightarrow{n\to+\infty} 0$ vérifierait alors $\mathbb{E}\Big[(\hat{T}_n - \theta)^2\Big] \xrightarrow{n\to+\infty} 0$.
On a alors, par inégalité de Markov,

$$\forall\epsilon > 0, \mathbb{P}\Big(|\hat{T}_n - \theta| > a\Big) = \mathbb{P}\Big((\hat{T}_n - \theta)^2 > a^2\Big) \leq \frac{\mathbb{E}\Big[(\hat{T}_n - \theta)^2\Big]}{a^2} \xrightarrow{n\to+\infty} 0.$$

Par définition de la convergence en probabilité, cela signifie que $\hat{T}_n \xrightarrow{\mathbb{P}} \theta$. On en déduit que $\hat{T}_n$ est un estimateur consistant de θ.

Exercice 5 :
Soient $U_1, U_2 \sim_{i.i.d.} \mathcal{U}([0,1])$.

1. On peut calculer la fonction de répartition de $\min(U_1, U_2)$: soit $t \in \mathbb{R}$, on a

$$\begin{aligned} F_{\min(U_1,U_2)}(t) &= \mathbb{P}\Big(\min(U_1, U_2) \leq t\Big) \\ &= 1 - \mathbb{P}\Big(\min(U_1, U_2) > t\Big) \end{aligned}$$

or, pour que le minimum de deux valeurs soit plus grand que t, il faut et il suffit que chacune des valeurs soit plus grande que t. On a donc égalité des évènements $\{\min(U_1, U_2) > t\}$ et $\{U_1 > t\} \cap \{U_2 > t\}$, et par conséquent

$$\begin{aligned} F_{\min(U_1,U_2)}(t) &= 1 - \mathbb{P}\Big(U_1 > t \text{ et } U_2 > t\Big) \\ &= 1 - \mathbb{P}\Big(U_1 > t\Big)\mathbb{P}\Big(U_2 > t\Big) \text{ par indépendance} \\ &= 1 - \Big(1 - \mathbb{P}\Big(U_1 \leq t\Big)\Big)\Big(1 - \mathbb{P}\Big(U_2 \leq t\Big)\Big) \\ &= 1 - \Big(1 - F_{U_1}(t)\Big)\Big(1 - F_{U_2}(t)\Big) \end{aligned}$$

il ne reste qu'à calculer la fonction de répartition d'une variable aléatoire uniforme sur $[0, 1]$:

$$\begin{aligned} F_U(t) &= \int_0^t \mathrm{d}x = t \text{ si } t \in [0, 1] \\ &= 0 \text{ si } t \leq 0 \\ &= 1 \text{ si } t \geq 1 \end{aligned}$$

On trouve alors

$$\begin{aligned} F_{\min(U_1,U_2)}(t) &= 1 - (1 - t)^2 \text{ si } t \in [0, 1] \\ &= 0 \text{ si } t \leq 0 \\ &= 1 \text{ si } t \geq 1 \end{aligned}$$

Remarque : on aurait pu directement calculer $\mathbb{P}\Big(U_1 > t\Big) = \mathbb{P}\Big(U_2 > t\Big) = \int_t^1 \mathrm{d}x = 1 - t$ pour $t \in [0, 1]$.

2. On peut calculer la fonction de répartition de U_1^2 : soit $t \in \mathbb{R}$, on a

$$\begin{aligned} F_{U_1^2}(t) &= \mathbb{P}\Big(U_1^2 \leq t\Big) \\ &= \mathbb{P}\Big(-\sqrt{t} \leq U_1 \leq \sqrt{t}\Big) \text{ sit} \geq 0 \\ &= \mathbb{P}\Big(U_1 \leq \sqrt{t}\Big) \text{ car } U_1 \geq 0 \text{ p.s.} \\ &= F_{U_1}(\sqrt{t}) \\ &= \sqrt{t} \text{ si } t \in [0, 1] \\ &= 1 \text{ si } t \geq 1 \end{aligned}$$

Il reste à calculer $F_{U_1^2}(t)$ pour $t \leq 0$, mais comme $U_1^2 \geq 0$ p.s., on a $\forall t \leq 0, F_{U_1^2}(t) = \mathbb{P}\Big(U_1^2 \leq t\Big) = 0$.

Exercice 6 :
Soit $X_1 \sim \mathcal{N}(7, 4)$ et $X_2 \sim \mathcal{N}(1, 4)$ indépendantes.

1. Par le théorème de stabilité des lois normales, $X_1 + X_2$ est de loi normale, il ne reste plus qu'à calculer son espérance et sa variance :

$$\begin{aligned} \mathbb{E}[X_1 + X_2] &= \mathbb{E}[X_1] + \mathbb{E}[X_2] = 7 + 1 = 8 \\ \mathrm{var}(X_1 + X_2) &= \mathrm{var}(X_1) + \mathrm{var}(X_2) \text{ par indépendance} \\ &= 4 + 4 = 8 \end{aligned}$$

donc

$$X_1 + X_2 \sim \mathcal{N}(8, 8).$$

2. Par le théorème de stabilité des lois normales, $\frac{X_1-2X_2}{5}$ est de loi normale, il ne reste plus qu'à calculer son espérance et sa variance :

$$\begin{aligned}
\mathbb{E}\Big[\frac{X_1-2X_2}{5}\Big] &= \frac{\mathbb{E}[X_1]-2\mathbb{E}[X_2]}{5} = \frac{7-2}{5} = 1\\
\text{var}\left(\frac{X_1-2X_2}{5}\right) &= \text{var}\left(\frac{1}{5}X_1\right) + \text{var}\left(-\frac{2}{5}X_2\right) \text{ par indépendance}\\
&= \left(\frac{1}{5}\right)^2 \text{var}\Big(X_1\Big) + \left(\frac{-2}{5}\right)^2 \text{var}\Big(X_2\Big)\\
&= \frac{1}{25}4 + \frac{4}{25}4 = \frac{20}{25} = \frac{4}{5}
\end{aligned}$$

donc

$$\frac{X_1-2X_2}{5} \sim \mathcal{N}\Big(1, \frac{4}{5}\Big).$$

3. Pour centrer et réduire X_2, il suffit de définir

$$X_2' = \frac{X_2-1}{2} \sim \mathcal{N}(0,1).$$

Exercice 7 :

1. On utilise l'additivité de $\mathbb{P}$:

$$\mathbb{P}\big(\{1,2,3\}\big) = \mathbb{P}\big(\{1\}\big) + \mathbb{P}\big(\{2\}\big) + \mathbb{P}\big(\{3\}\big) = \frac{3}{4}.$$

2. On pose $Z = \frac{X_1+1}{2}$. Comme X_1 est à valeurs dans $\{-1,1\}$, la variable Z est à valeurs dans $\{\frac{-1+1}{2}, \frac{1+1}{2}\} = \{0,1\}$. Il s'agit donc d'une variable aléatoire de Bernoulli. Son paramètre est donné par

$$\mathbb{P}(Z=1) = \mathbb{P}(X_1=1) = \mathbb{P}\big(\{3,4\}\big) = \frac{1}{2}$$

. Donc $Z \sim \mathcal{B}(\frac{1}{2})$.

3. On peut calculer

$$\begin{aligned}
\mathbb{P}(X_1=-1) = \mathbb{P}\big(\{1,2\}\big) &= \frac{1}{2}\\
\mathbb{P}(X_2=2) = \mathbb{P}\big(\{2\}\big) &= \frac{1}{4}\\
\mathbb{P}(|X_2|=2) = \mathbb{P}\big(\{1,2\}\big) &= \frac{1}{2}\\
\mathbb{P}(X_1=-1 \text{ et } |X_2|=2) = \mathbb{P}\big(\{1,2\}\big) &= \frac{1}{2}
\end{aligned}$$

4. On a donc

$$\mathbb{P}(X_1=-1 \text{ et } |X_2|=2) = \frac{1}{2} \neq \frac{1}{4} = \mathbb{P}(X_1=-1)\mathbb{P}(|X_2|=2),$$

donc X_1 et X_2 ne sont pas indépendantes.

5. On peut calculer

$$\begin{aligned}
\mathbb{E}[X_1] &= \frac{1}{4}(-1) + \frac{1}{4}(-1) + \frac{1}{4}(1) + \frac{1}{4}(1) = 0\\
\mathbb{E}[X_2] &= \frac{1}{4}(-2) + \frac{1}{4}(2) + \frac{1}{4}(-1) + \frac{1}{4}(1) = 0\\
\mathbb{E}[X_1X_2] &= \frac{1}{4}(-1)(-2) + \frac{1}{4}(-1)(2) + \frac{1}{4}(1)(-1) + \frac{1}{4}(1)(1) = 0\\
\text{Cov}(X_1,X_2) &= \mathbb{E}[X_1X_2] - \mathbb{E}[X_1]\mathbb{E}[X_2] = 0
\end{aligned}$$

On a donc un exemple de variables qui sont de covariance nulle **mais pas indépendantes.**

Exercice 8 :

1. Soit X_i une suite de v.a. i.i.d. d'espérance m et de variance σ^2. On pose $\hat{T}_n = \frac{1}{n}\sum_{i=1}^n \Big(X_i - \bar{X}_n\Big)^2$. On peut tout d'abord calculer

$$\begin{aligned}
\hat{T}_n^2 &= \frac{1}{n}\sum_{i=1}^n \Big(X_i - \bar{X}_n\Big)^2 \\
&= \frac{1}{n}\sum_{i=1}^n \left(X_i^2 - 2X_i\bar{X}_n + \bar{X}_n^2\right) \\
&= \frac{1}{n}\sum_{i=1}^n X_i^2 - 2\frac{1}{n}\sum_{i=1}^n X_i\bar{X}_n + \frac{1}{n}\sum_{i=1}^n \bar{X}_n^2 \\
&= \frac{1}{n}\sum_{i=1}^n X_i^2 - 2\bar{X}_n\left(\frac{1}{n}\sum_{i=1}^n X_i\right) + \bar{X}_n^2\frac{1}{n}\sum_{i=1}^n 1 \\
&= \frac{1}{n}\sum_{i=1}^n X_i^2 - 2\bar{X}_n\bar{X}_n + \bar{X}_n^2 \\
&= \frac{1}{n}\sum_{i=1}^n X_i^2 - \bar{X}_n^2
\end{aligned}$$

Pour savoir si $\hat{T}_n$ est un estimateur sans biais de σ^2, il suffit de calculer l'espérance :

$$\begin{aligned}
\mathbb{E}\Big[\hat{T}_n^2\Big] &= \mathbb{E}\left[\Big(\frac{1}{n}\sum_{i=1}^n X_i^2\Big) - \bar{X}_n^2\right] \\
&= \frac{1}{n}\sum_{i=1}^n \mathbb{E}\Big[X_i^2\Big] - \mathbb{E}\Big[\bar{X}_n^2\Big] \\
&= \frac{1}{n}\sum_{i=1}^n \mathbb{E}\Big[X_1^2\Big] - \mathbb{E}\Big[\bar{X}_n^2\Big] \text{ car les } X_i \text{ ont même loi} \\
&= \mathbb{E}\Big[X_1^2\Big] - \mathbb{E}\Big[\bar{X}_n^2\Big]
\end{aligned}$$

On peut ensuite utiliser que $\mathrm{var}(Y) = \mathbb{E}[Y^2] - \mathbb{E}[Y]^2$, ce qui donne $\mathbb{E}[Y^2] = \mathrm{var}(Y) + \mathbb{E}[Y]^2$, en l'appliquant à X_1 et à $\bar{X}_n$. On obtient alors

$$\begin{aligned}
\mathbb{E}\Big[\hat{T}_n^2\Big] &= \mathrm{var}(X_1) + \mathbb{E}[X_1]^2 - \mathrm{var}(\bar{X}_n) - \mathbb{E}[\bar{X}_n]^2 \\
&= \sigma^2 + m^2 - \mathrm{var}(\bar{X}_n) - \mathbb{E}[\bar{X}_n]^2
\end{aligned}$$

or

$$\begin{aligned}
\mathbb{E}[\bar{X}_n] &= \frac{1}{n}\sum_{i=1}^n \mathbb{E}[X_i] \\
&= m \\
\mathrm{var}(\bar{X}_n) &= \frac{1}{n^2}\sum_{i=1}^n \mathrm{var}(X_i) \text{ par indépendance} \\
&= \frac{\sigma^2}{n}
\end{aligned}$$

On obtient finalement

$$\begin{aligned}
\mathbb{E}\Big[\hat{T}_n^2\Big] &= \sigma^2 + m^2 - \frac{\sigma^2}{n} - m^2 \\
&= \frac{(n-1)\sigma^2}{n}
\end{aligned}$$

Comme on n'a pas $\mathbb{E}\left[\hat{T}_n^2\right] = \sigma^2$, l'estimateur $\hat{T}_n^2$ est un estimateur biaisé de σ^2. Par contre, on a bien $\frac{n-1}{n} \xrightarrow{n\to+\infty} 1$, d'où $\mathbb{E}\left[\hat{T}_n^2\right] \xrightarrow{n\to+\infty} \sigma^2$. Donc $\hat{T}_n^2$ est un estimateur asymptotiquement sans biais de σ^2.

2. On cherche a tel que $\hat{S}_n^2 = a\hat{T}_n^2$ soit un estimateur sans biais de σ^2, i.e. tel que $\mathbb{E}\left[\hat{S}_n^2\right] = \sigma^2$. Il suffit de résoudre

$$\sigma^2 = \mathbb{E}\left[\hat{S}_n^2\right] = \mathbb{E}\left[a\hat{T}_n^2\right] = a\mathbb{E}\left[\hat{T}_n^2\right] = a\frac{(n-1)\sigma^2}{n}.$$

Cela donne $a = \frac{n}{n-1}$, et donc

$$\hat{S}_n^2 = \frac{n}{n-1}\hat{T}_n^2 = \frac{n}{n-1}\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2,$$

qui est un estimateur sans biais de la variance.