Université Paris I, Panthéon - Sorbonne

Licence M.I.A.S.H.S. troisième année 2023 – 2024

Feuilles de TD, cours de L3 Statistique 2

Jean-Marc Bardet (Université Paris 1, SAMM)

Email: `bardet@univ-paris1.fr`
Page oueb: `http://samm.univ-paris1.fr/-Jean-Marc-Bardet-`

U.F.R. 27 et Equipe SAMM (Statistique, Analyse et Modélisation Multidisiplinaire)
Université Panthéon-Sorbonne, 90 rue de Tolbiac, 75013 Paris.

Feuille n^o 1:
Variables aléatoires

1. **(*)** Soit l'espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ où $\Omega = [0,1]$, $\mathcal{A}$ la tribu borélienne sur Ω et $\mathbb{P}$ la probabilité uniforme sur $[0,1]$.

   (a) On pose X la variable aléatoire telle que $X(\omega) = 1-\omega$ pour tout $\omega \in \Omega$. Déterminer la loi de probabilité de X, son espérance et sa variance.

   (b) Répondre aux mêmes questions pour $Y(\omega) = -\ln(\omega)$.

   (c) On pose $Z(\omega) = \omega$ pour $\omega \in [0.5, 1]$ et $Z(\omega) = 0$ pour $\omega \in [0, 0.5[$. Déterminer la fonction de répartition de Z, son espérance et sa variance.

Proof. (a) La variable X est à valeurs dans $[0,1]$ car $1-\omega \in [0,1]$ pour tout $\omega \in [0,1]$. Donc pour $x \leq 0$, $F_X(x) = 0$ et pour $x \geq$, $F_X(x) = 1$. Pour $x \in [0,1]$, on a $\{X \leq x\} = \{\omega \in [0,1],\ 1-\omega \leq x\} = \{\omega \in [0,1],\ 1-x \leq \omega\} = [1-x, 1]$. Or $\mathbb{P}([1-x,1]) = x$ car $\mathbb{P}$ mesure la longueur de l'intervalle, d'où $F_X(x) = x$. La loi de X est donc celle d'une variable uniforme sur $[0,1]$, d'où $\mathbb{E}[X] = 1/2$ et $\mathrm{var}(X) = 1/12$.

(b) La variable Y est à valeurs dans $[0, +\infty[$ car $\ln(\omega) \leq 0$ pour tout $\omega \in [0,1]$. Donc pour $y \leq 0$, $F_Y(y) = 0$. Pour $y \geq 0$, on a $\{Y \leq y\} = \{\omega \in [0,1],\ -\ln(\omega) \leq y\} = \{\omega \in [0,1],\ e^{-y} \leq \omega\} = [e^{-y}, 1]$. Or $\mathbb{P}([e^{-y}, 1]) = 1 - e^{-y}$ car $\mathbb{P}$ mesure la longueur de l'intervalle, d'où $F_Y(y) = 1 - e^{-y}$: Y suit la loi exponentielle $\mathcal{E}(1)$, donc $\mathbb{E}[Y] = 1$ et $\mathrm{var}(Y) = 1$.

(c) Les valeurs prises par Z sont $\{0\} \cup [0.5, 1]$. Ainsi, pour $z < 0$ alors $F_Z(z) = 0$, et pour $z \geq 1$, $F_Z(z) = 1$. Si $z \in [0, 0.5[$, $F_Z(z) = \mathbb{P}(Z = 0) = \mathbb{P}([0, 0.5[) = 0.5$. Si $z \in [0.5, 1]$, $F_Z(z) = 0.5 + \mathbb{P}(0.5 \leq Z \leq z) = 0.5 + \mathbb{P}([0.5, z]) = 0.5 + (z - 0.5) = z$.
$\mathbb{E}[Z] = \int_{[0,1/2]} 0\, d\omega + \int_{[1/2,1]} \omega\, d\omega = 0 + \left[\omega^2/2\right]_{1/2}^1 = 3/8$.
Et $\mathbb{E}[Z^2] = \int_{[0,1/2]} 0\, d\omega + \int_{[1/2,1]} \omega^2\, d\omega = 0 + \left[\omega^3/3\right]_{1/2}^1 = 7/24$, d'où $\mathrm{var}(Z) = 7/24 - 9/64 = 29/192 \simeq 0.151$.

□

2. **(*)** Sur $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé, on considère une v.a. réelle positive X de fonction de répartition F_X. Déterminer dans les 2 cas suivants l'espérance et la variance de X:

$$F_X(t) = \frac{1}{2}\left(e^t \mathbb{I}_{]-\infty,0[}(t) + (2 - e^{-t})\mathbb{I}_{[0,\infty[}(t)\right);$$
$$F_X(t) = \frac{1}{4}(t+2)\mathbb{I}_{[-1,0[\cup[1,2[}(t) + \frac{3}{4}\mathbb{I}_{[0,1]}(t) + \mathbb{I}_{[2,\infty[}(t).$$

Proof. Dans le premier cas, la fonction de répartition est continue et dérivable sur $\mathbf{R}^*$: la v.a. est donc continue et sa densité est $f_X(x) = \frac{1}{2} e^{-|x|}$ pour $x \in \mathbf{R}$: loi de Laplace. On alors $\mathbb{E}[X] = 0$ et $\mathrm{var}(X) = 2$.
Dans le second cas, il y a 2 sauts: en -1 avec un saut de hauteur de $1/4$ et en 0 avec un saut de hauteur $1/4$. La mesure de probabilité de X peut donc s'écrire:

$$\mathbb{P}(X \in B) = \frac{1}{4}\left(\delta_{\{-1\}}(B) + \delta_{\{0\}}(B)\right) + \frac{1}{4}\int_B \mathbb{I}_{[-1,0[\cup[1,2[}(t)\, dt \quad \text{pour } B \in \mathcal{B}(\mathbf{R}).$$

D'où $\mathbb{E}[X] = \frac{1}{4}\big(-1 + 0 - 1/2 + 3/2\big) = 0$ et $\mathrm{var}(X) = \mathbb{E}[X^2] = \frac{1}{4}\big(1 + 0 + \int_{-1}^0 t^2 dt + \int_1^2 t^2 dt\big) = \frac{1}{4}\big(1 + \frac{8}{3}\big) = \frac{11}{12}$. □

3. **(*)** Soit une variable aléatoire X sur l'espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$. On suppose que loi de X est symétrique, c'est-à-dire que la loi de X est la même que celle de $-X$.

   (a) Montrer que $\mathbb{P}(X \leq 0) \geq 1/2$ et $\mathbb{P}(X < 0) \leq 1/2$. Conclusion?

   (b) Montrer que si $\mathbb{E}(|X|) < \infty$ alors $\mathbb{E}(X) = 0$.

Proof. (a) On a d'après la formule des probabilités totales $\mathbb{P}(X > 0) + \mathbb{P}(X < 0) + \mathbb{P}(X = 0) = 1$. De plus $\mathbb{P}(X > 0) = \mathbb{P}(-X < 0) = \mathbb{P}(X < 0)$ puisque X et $-X$ ont même loi, donc même fonction de répartition. D'où $\mathbb{P}(X > 0) + \mathbb{P}(X < 0) = 2\mathbb{P}(X < 0)$. Par suite, $2\mathbb{P}(X > 0) = 1 - \mathbb{P}(X = 0)$, d'où $\mathbb{P}(X > 0) \leq 1/2$. Or, la formule des probabilités totales donne également $\mathbb{P}(X \leq 0) = 1 - \mathbb{P}(X > 0)$ et ainsi $\mathbb{P}(X \leq 0) \geq 1/2$. $\mathbb{P}(X < 0) \leq 1/2$ en découle également.

(b) On va traiter les 2 cas de v.a. Si X est une v.a. discrète à valeurs dans $I=(x_j)_{j\in J}$. Comme X et $-X$ ont même loi, forcément quand $x_j \in I$, alors $-x_j \in I$ et $\mathbb{P}(X=x_j)=\mathbb{P}(X=-x_j)$. Or $\mathbb{E}(X)=\sum_{j\in J} x_j\,\mathbb{P}(X=x_j)=\sum_{j\in J,\,x_j>0} x_j\,\mathbb{P}(X=x_j)+\sum_{j\in J,\,x_j<0} x_j\,\mathbb{P}(X=x_j)+0*\mathbb{P}(X=0)$. En conséquence

$$\mathbb{E}[X]=\sum_{j\in J,\,x_j>0}\big(x_j\,\mathbb{P}(X=x_j)-x_j\,\mathbb{P}(X=-x_j)\big)=0.$$

Pour une v.a. continue, $F_X(x)=\mathbb{P}(X\le x)=1-\mathbb{P}(X>x)=1-\mathbb{P}(-X<-x)=1-F_X(-x)$ car la variable est continue. Donc en tout x où F_X est dérivable, $F_X'(x)=0-(F_X(-x))'=F_X'(-x)$, d'où $f_X(x)=f_X(-x)$: la densité est une fonction paire. Donc $\mathbb{E}[X]=\int_{-\infty}^{\infty} x\,f_X(x)dx=0$ car la fonction $x\to x\,f_X(x)$ est impaire.

□

4. **(**)** Sur $(\Omega,\mathcal{A},\mu)$ un espace probabilisé, on considère une v.a. réelle positive X de fonction de répartition F_X. Montrer, en utilisant Fubini, que pour $n\in\mathbb{N}^*$:

$$\mathbb{E}[X^n]=\int_0^\infty n\,t^{n-1}(1-F_X(t))\,dt=\int_0^\infty n\,t^{n-1}\mathbb{P}(X>t)\,dt.$$

Montrer que l'hypothèse X positive est nécessaire.

Proof. Du fait que les fonctions intervenant dans l'intégrale sont mesurables positives, on peut écrire avec Fubini, quitte à obtenir $+\infty$,

$$\begin{aligned}\int_0^\infty n\,t^{n-1}\mathbb{P}(X>t)\,dt &= \int_0^\infty n\,t^{n-1}\int_{]t,\infty[} d\mathbb{P}_X(x)\,dt\\ &= \int_0^\infty\int_{]t,\infty[} n\,t^{n-1}\,d\mathbb{P}_X(x)\,dt\\ &= \int_0^\infty\int_{[0,x[} n\,t^{n-1}\,dt\,d\mathbb{P}_X(x)\quad\text{en réécrivant le domaine d'intégration}\\ &= \int_0^\infty x^n\,d\mathbb{P}_X(x)=\mathbb{E}|X^n].\end{aligned}$$

Si $n=1$ et X peut être négative, alors on peut avoir $\mathbb{E}[X]<0$ ce qui n'est pas possible avec une telle formule. □

5. **(**)** Soit X une v.a. réelle normale centrée réduite. Soit la v.a. $Y=e^X$. On dit que Y suit une loi log-normale.

(a) Montrer que Y à une mesure de probabilité absolument continue par rapport à la mesure de Lebesgue de densité $f_Y(y)=\dfrac{1}{\sqrt{2\pi}}\dfrac{1}{y}\,e^{-\ln^2(y)/2}$ si $z>0$ et 0 sinon.

(b) Pour $a\in[-1,1]$, soit Y_a la v.a. de densité $f_a(y)=f_Y(y)(1+a\sin(2\pi\ln(y))$. Montrer que Y et Y_a ont mêmes moments, et en déduire que les moments ne caractérisent pas une loi de probabilité.

Proof. (a) Il est clair que $Y:\Omega\to]0,\infty[$ et Y v.a. car $x\in\mathbf{R}\mapsto e^x$ est une fonction continue donc mesurable (borélienne). Donc pour $y\le 0$, $F_Y(y)=0$. Et pour $y>0$,

$$F_Y(y)=\mathbb{P}(Y\le y)=\mathbb{P}(X\le\ln(y))=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\ln(y)} e^{-t^2/2}dt.$$

Il est clair que pour tout $y>0$ cette fonction est dérivable (donc continue) et sa limite en 0^+ est 0: F_Y est continue sur $\mathbf{R}$ et dérivable partout sauf en 0, donc Y est une v.a. continue.
Sa dérivée, donc sa densité, sur $]-\infty,0[$ est 0 et pour $y>0$,

$$f_Y(y)=\frac{\partial}{\partial}\big(F_X(\ln(y))-F_X(-\infty)\big)=\frac{1}{y}\,f_X(\ln(y))=\frac{1}{\sqrt{2\pi}\,y}\,e^{-\ln^2(y)/2}.$$

(b) Pour tout $a\in[-1,1]$, il est clair que $f_a(y)$ est mesurable positive, et son intégrale existe car $f_a\le(1+|a|)f_Y$. De plus,

$$\int_0^\infty f_a(y)\,dy=\int_0^\infty f_Y(y)\,dy+a\int_0^\infty f_Y(y)\,\sin(2\pi\,\ln(y))\,dy=1+a\int_\infty^\infty e^x\,f_Y(e^x)\,\sin(2\pi\,x)\,dx.$$

Mais $e^x\,f_Y(e^x)=\frac{1}{\sqrt{2\pi}}\,e^{-x^2/2}$ fonction paire sur $\mathbf{R}$, donc $e^x\,f_Y(e^x)\,\sin(2\pi\,x)$ est une fonction impaire intégrable, donc son intégrale sur $\mathbf{R}$ est nulle. On en déduit que $\int_0^\infty f_a(y)\,dy=1$ pour tout $a\in[-1,1]$.
Si on calcule $\mathbb{E}[Y_a^n]$ (qui existe car $\mathbb{E}[Y^n]$ existe) alors:

$$\int_0^\infty y^n\,f_a(y)\,dy=\int_0^\infty y^n\,f_Y(y)\,dy+a\int_0^\infty y^n f_Y(y)\,\sin(2\pi\,\ln(y))\,dy=\mathbb{E}[Y^n]+a\,\frac{1}{\sqrt{2\pi}}\int_\infty^\infty e^{(n-1)x}\,e^{-x^2/2}\,\sin(2\pi\,x)\,dx.$$

Mais $e^{(n-1)x-x^2/2}=e^{-(n-1)^2/2}\,e^{-(x-(n-1))^2/2}$. Par changement de variable, on obtient

$$\begin{aligned}\int_{\infty}^{\infty} e^{(n-1)x}\,e^{-x^2/2}\,\sin(2\pi\,x)\,dx &= e^{-(n-1)^2/2}\int_{\infty}^{\infty} e^{-(x-(n-1))^2/2}\,\sin(2\pi\,x)\,dx\\ &= e^{-(n-1)^2/2}\int_{\infty}^{\infty} e^{-z^2/2}\,\sin(2\pi\,(z+(n-1)))\,dz = e^{-(n-1)^2/2}\int_{\infty}^{\infty} e^{-z^2/2}\,\sin(2\pi\,z)\,dz=0,\end{aligned}$$

par parité. Donc $\mathbb{E}[Y_a^n]=\mathbb{E}[Y^n]$ pour tout $n\geq 0$ et tout $a\in[-1,1]$: les moments ne caractérise pas la loi, puisque clairement Y_a et Y n'ont pas la même loi si $a\neq 0$.

□

6. **(*)** Soit $X \overset{\mathcal{L}}{\sim} \mathcal{E}(\lambda)$ loi exponentielle de paramètre $\lambda>0$. Quelle est la loi de $Y=[X+1]$? (partie entière de $X+1$)

Proof. Y prend ses valeurs dans $\mathbf{N}^*$ et $\mathbb{P}(Y=k)=\mathbb{P}(k\leq X+1<k+1)=F_X(k)-F_X(k-1)$, donc $\mathbb{P}(Y=k)=(1-e^{-\lambda\,k})-(1-e^{-\lambda(k-1)})=\big(1-e^{-\lambda}\big)e^{-\lambda(k-1)}$. Ainsi $\mathbb{P}(Y=k)=\big(1-e^{-\lambda}\big)\big(e^{-\lambda}\big)^{k-1}$ pour $k\geq 1$: $Y \overset{\mathcal{L}}{\sim} \mathcal{G}\big(1-e^{-\lambda}\big)$: loi géométrique. □

7. **(**)** Soit U une variable aléatoire uniforme sur $[0,1]$. Soit X une variable de fonction de répartition F_X que l'on supposera strictement croissante et dérivable sur $\mathbf{R}$.
   (a) Montrer F_X est une fonction admettant une application réciproque sur $]0,1[$, notée F_X^{-1}.
   (b) Démontrer que la loi de la variable $F_X^{-1}(U)$ est la même que celle de X.
   (c) A partir de la touche `RAND` d'une calculatrice, comment obtenir une réalisation d'une variable aléatoire de loi exponentielle de paramètre 3?
   (d) Même question si $F_X(x)=\arctan(x)/\pi+1/2$. Quelle est alors l'espérance de $F_X^{-1}(U)$?

Proof. (a) Si F_X est strictement croissante et dérivable, donc continue, alors comme $\lim_{x\to-\infty}F_X(x)=0$ et $\lim_{x\to\infty}F_X(x)=1$, on en déduit que F_X : $\mathbf{R}ø]0,1[$. De plus, pour tout $y\in]0,1[$, s'il existe $x<x'$ tel que $F_X(x)=F_X(x')=y$ alors F_X ne serait pas strictement croissante: F_X est bien bijective, et admet une fonction réciproque F_X^{-1} sur $]0,1[$.

(b) Comme F_X est dérivable et strictement croissante, sa dérivée ne s'annule pas, donc F_X^{-1} est dérivable et strictement croissante sur $]0,1[$, donc continue: $F_X^{-1}(U)$ est bien une variable aléatoire sur $(\Omega,\mathcal{A},\mathbb{P})$ qui prend ses valeurs dans $\mathbf{R}$. On a pour tout $x\in\mathbf{R}$, en utiliant le fait que $F_X(F_X^{-1}(U))=U$ et F_X strictement croissante,

$$\mathbb{P}(F_X^{-1}(U)\leq x)=\mathbb{P}(U\leq F_X(x))=F_U(F_X(x))=F_X(x)\qquad\text{car } F_U(u)=u \text{ pour tout } u\in[0,1].$$

La v.a. $F_X^{-1}(U)$ a donc même fonction de répartition que X, ces deux v.a. ont donc même loi.

(c) On sait que la touche `RAND` fournit une réalisation d'une v.a. uniforme sur $[0,1]$. Pour obtenir une réalisation d'une v.a. exponentielle de paramètre, il faudra donc calculer $V=F_X^{-1}(U)$. Or $F_X(x)=1-e^{-3\,x}$, d'où $x=-\ln(1-F_X(x))/3$ et on en déduit que $V=-\ln(1-U)/3$.

(d) Si $F_X(x)=\arctan(x)/\pi+1/2$ alors $x=\pi\,\tan(F_X(x)-1/2)$ soit $W=F_X^{-1}(U)=\pi\,\tan(U-1/2)$.
La densité de X est $\frac{1}{\pi(1+x^2)}$, c'est une v.a. qui suit une loi de Cauchy. On a alors $\mathbb{E}[W]=\mathbb{E}[X]=\int_{-\infty}^{\infty}\frac{x}{\pi(1+x^2)}\,dx$. Mais cette intégrale n'existe pas car:

$$\mathbb{E}[|X|]=\int_{-\infty}^{\infty}\frac{|x|}{\pi(1+x^2)}\,dx=2\int_0^{\infty}\frac{x}{\pi(1+x^2)}\,dx=\frac{1}{\pi}\left[\ln(1+x^2)\right]_0^{\infty}=+\infty.$$

□

8. **(*)** Calculer la fonction génératrice d'une variable aléatoire suivant une loi géométrique de paramètre p. De même pour celle d'une loi de Poisson de paramètre λ. En déduire que la somme de 2 v.a. indépendantes de lois de Poisson de paramètres λ_1 et λ_2 est une loi de Poisson. En est-il de même pour la loi géométrique?

Proof. Si X v.a. de loi géométrique de paramètre p alors pour $z \in [-1,1]$,

$$g(z) = \mathbb{E}[z^X] = \sum_{k=1}^{\infty} z^k (1-p)^{k-1} p = p\, z \sum_{k=1}^{\infty} (z\,(1-p))^{k-1} = p\, z \sum_{k=0}^{\infty} (z\,(1-p))^{k} = \frac{p\, z}{1-(1-p)\,z}.$$

Si X v.a. de loi de Poisson de paramètre λ alors pour $z \in [-1,1]$,

$$g(z) = \mathbb{E}[z^X] = e^{-\lambda} \sum_{k=0}^{\infty} z^k\, \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(z\,\lambda)^k}{k!} = e^{-\lambda} e^{z\,\lambda} = e^{(z-1)\,\lambda}.$$

Si X_1 et X_2 sont deux v.a. indépendantes de lois de Poisson de paramètres λ_1 et λ_2, alors par indépendance

$$\mathbb{E}[z^{X_1+X_2}] = \mathbb{E}[z^{X_1}]\,\mathbb{E}[z^{X_2}] = e^{(z-1)\,\lambda_1}\, e^{(z-1)\,\lambda_2} = e^{(z-1)\,(\lambda_1+\lambda_2)}$$

qui caractérise la loi de Poisson de paramètre $\lambda_1+\lambda_2$.
Par le même raisonnement, si X_1 et X_2 sont deux v.a. indépendantes de lois géométriques de paramètres p_1 et p_2,

$$\mathbb{E}[z^{X_1+X_2}] = \frac{p_1\, z}{1-(1-p_1)\,z}\, \frac{p_2\, z}{1-(1-p_2)\,z} = \frac{p_1\, p_2\, z^2}{(1-(1-p_1)\,z)(1-(1-p_2)\,z)}$$

qui ne peut clairement pas être simplifié pour faire apparaître la fonction génératrice d'une loi géométrique. □

9. **(*)** Calculer la fonction caractéristique d'une variable aléatoire : a/ gaussienne, b/ de Poisson, c/ exponentielle, d/ uniforme, e/ gamma, f/ binomiale. En déduire que la somme de 2 v.a. gaussiennes indépendantes est gaussienne.

Proof. a/ Pour $X \overset{\mathcal{L}}{\sim} \mathcal{N}(m\,,\sigma^2)$, on peut toujours écrire que $X \overset{\mathcal{L}}{\sim} m+\sigma\, Z$, avec $Z \overset{\mathcal{L}}{\sim} \mathcal{N}(0\,,1)$. On a alors $\phi_X(u) = \mathbb{E}\big[e^{i\,u\,(m+\sigma\,Z)}\big] = e^{i\,u\,m}\phi_Z(\sigma\, u)$. Mais:

$$\phi_Z(u) = \frac{1}{\sqrt{2\,\pi}} \int_{-\infty}^{\infty} e^{i\,u\,x - x^2/2}\, dx = \frac{1}{\sqrt{2\,\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left((x-iu)+u^2\right)}\, dx = e^{-u^2/2},$$

après changement de variable $y = x - iu$. D'où $\phi_X(u) = e^{-\frac{1}{2}\sigma^2\, u^2 + i\,u\,m}$.
b/ Si $X \overset{\mathcal{L}}{\sim} \mathcal{P}(\lambda)$, alors pour $u \in \mathbf{R}$,

$$\phi_X(u) = e^{-\lambda} \sum_{k=0}^{\infty} \frac{1}{k}\, \lambda^k e^{iuk} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{1}{k} \left(\lambda\, e^{iu}\right)^k = e^{-\lambda} e^{\lambda\, e^{iu}} = e^{\lambda\,(e^{iu}-1)}.$$

c/ Si $X \overset{\mathcal{L}}{\sim} \mathcal{E}(\lambda)$, alors pour $u \in \mathbf{R}$,

$$\phi_X(u) = \int_0^{\infty} \lambda\, e^{-\lambda\,x + i\,u\,x} dx = \Big[\frac{\lambda}{i\,u-\lambda}\, e^{(-\lambda+i\,u)x}\Big]_0^{\infty} = \frac{\lambda}{-i\,u+\lambda}.$$

d/ Si $X \overset{\mathcal{L}}{\sim} \mathcal{U}([a,b])$, alors pour $u \in \mathbf{R}^*$,

$$\phi_X(u) = \frac{1}{b-a} \int_a^b e^{i\,u\,x} dx = \frac{1}{b-a} \Big[\frac{\lambda}{i\,u}\, e^{i\,ux}\Big]_a^b = \frac{1}{i\,(b-a)\,u} \big(\cos(ub) - \cos(ua) + i(\sin(ub) - \sin(ua))\big).$$

e/ Si $X \overset{\mathcal{L}}{\sim} \Gamma(\alpha,\beta)$, alors pour $u \in \mathbf{R}$,

$$\phi_X(u) = \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^{\infty} x^{\alpha-1}\, e^{-\beta\,x + i\,u\,x} dx = \frac{\beta^\alpha}{(\beta - i\,u)^\alpha \Gamma(\alpha)} \int_0^{\infty} y^{\alpha-1}\, e^{-y} dy = \big(1 - iu/\beta\big)^{-\alpha},$$

avec le changement de variable $y = (\beta - i\,u)x$, soit $dy = (\beta - i\,u)dx$.
f/ Si $X \overset{\mathcal{L}}{\sim} \mathcal{B}(n,p)$, alors pour $u \in \mathbf{R}$,

$$\phi_X(u) = \sum_{k=0}^{n} \binom{k}{n} p^k (1-p)^{n-k} e^{iuk} = (1-p)^n \sum_{k=0}^{n} \binom{k}{n} \Big(\frac{p\,e^iu}{1-p}\Big)^k = (1-p)^n \Big(1 + \frac{p\,e^iu}{1-p}\Big)^n = \big(1 + p\,(e^{iu}-1)\big)^n,$$

en utilisant la formule du binome.

Si deux v.a. X et X' sont gaussiennes de lois $\mathcal{N}(m\,,\sigma^2)$ et $\mathcal{N}(m'\,,\sigma'^2)$ sont indépendantes, alors $\phi_{X+X'}(u) = \phi_X(u)\,\phi_{X'}(u)$ par l'indépendance, soit $\phi_{X+X'}(u) = e^{-\frac{1}{2}\sigma^2\,u^2 + i\,u\,m - \frac{1}{2}\sigma'^2\,u^2 + i\,u\,m'} = e^{-\frac{1}{2}(\sigma^2+\sigma'^2)\,u^2 + i\,u\,(m+m')}$, soit la loi $\mathcal{N}(m+m'\,,\sigma^2+\sigma'^2)$. □

10. **(***)** En utilisant la formule d'inversion de la fonction caractéristique pour les v.a. continues, démontrer que la fonction de caractéristique d'une v.a. de Cauchy de densité $f(x) = \pi^{-1}(1+x^2)^{-1}$ sur $\mathbf{R}$ est $\phi(u) = e^{-|u|}$.

Proof. On part de la formule de la densité caractéristique $\phi(u) = e^{-|u|}$. Comme on sait que X est une variable "continue" et que cette fonction caractéristique est intégrable, on utilise la formule d'inversion:

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \phi_X(u)\, e^{-i\,u\,x} du \quad \text{pour tout } x \in \mathbf{R}.$$

Donc pour $x \in \mathbf{R}$, $f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-|u|-i\,u\,x} du$. On en déduit que:

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{0} e^{u-i\,u\,x} du + \frac{1}{2\pi} \int_{0}^{\infty} e^{-u-i\,u\,x} du = \frac{1}{2\pi} \Big(\Big[\frac{e^{u-i\,u\,x}}{1-i\,x} \Big]_{-\infty}^{0} + \Big[\frac{e^{-u-i\,u\,x}}{-1-i\,x} \Big]_{0}^{\infty} \Big) = \frac{1}{\pi}\, \frac{1}{1+x^2}.$$

Par unicité de la fonction caractéristique, on en déduit que celle-ci est bien celle d'une loi de Cauchy. □

11. **(**)** Soit X une variable aléatoire réelle intégrable telle que $\mathbb{E}[X] \geq 0$.
    (a) Montrer que pour tout $\lambda > 0$, $X \leq \lambda\, \mathbb{E}[X] + X\, \mathbb{I}_{\{X > \lambda\, \mathbb{E}[X]\}}$.
    (b) On suppose que, de plus, $0 < \mathbb{E}[X^2] < +\infty$. Montrer que

$$\big(\mathbb{E}\big[X\, \mathbb{I}_{\{X > \lambda \mathbb{E}[X]\}}\big]\big)^2 \leq \mathbb{E}[X^2]\, \mathbb{P}\big(X > \lambda \mathbb{E}[X]\big).$$

    (c) Montrer que pour tout $\lambda \in]0,1[$ on a l'Inégalité de Paley-Zygmund:

$$\mathbb{P}\big(X > \lambda\, \mathbb{E}[X]\big) \geq (1-\lambda)^2 \frac{\mathbb{E}[X]^2}{\mathbb{E}[X^2]}.$$

Proof. (a) Si $X > \lambda\, \mathbb{E}[X]$ alors le terme de droite vaut X, donc l'inégalité est vérifiée. Si $X \leq \lambda\, \mathbb{E}[X]$, alors $\lambda\, \mathbb{E}[X] + X\, \mathbb{I}_{\{X > \lambda\, \mathbb{E}[X]\}} = \lambda\, \mathbb{E}[X]$. Mais comme cela a lieu pour $X \leq \lambda\, \mathbb{E}[X]$, l'inégalité est bien vérifiée. Elle l'est donc dans tous les cas.

(b) On utilise l'inégalité de Cauchy-Schwarz et le fait que

$$\mathbb{E}\big[\mathbb{I}^2_{\{X > \lambda\, \mathbb{E}[X]\}}\big] = \mathbb{E}\big[\mathbb{I}_{\{X > \lambda \mathbb{E}[X]\}}\big] = \mathbb{P}\big(X > \lambda\, \mathbb{E}[X]\big).$$

(c) Grâce à la première question, $X - \lambda\, \mathbb{E}[X] \leq X\, \mathbb{I}_{\{X > \lambda\, \mathbb{E}[X]\}}$. En prenant l'espérance on obtient donc que $(1-\lambda)\, \mathbb{E}[X] \leq \mathbb{E}\big[X\, \mathbb{I}_{\{X > \lambda\, \mathbb{E}[X]\}}\big]$. Comme $\mathbb{E}[X] \geq 0$ et $\lambda \in [0,1]$, alors $(1-\lambda)\, \mathbb{E}[X] \geq 0$. Donc

$$(1-\lambda)^2\, \big(\mathbb{E}[X]\big)^2 \leq \big(\mathbb{E}\big[X\, \mathbb{I}_{\{X > \lambda\, \mathbb{E}[X]\}}\big]\big)^2.$$

Le résultat final est alors obtenu grâce à celui de la deuxième question.

□

Feuille n^o 2:
Vecteurs aléatoires

1. **(*)** Soit (X, Y) un couple de variables aléatoires à valeurs dans $\mathbf{R}^2$ dont la loi a pour densité par rapport à la mesure de Lebesgue sur $\mathbf{R}^2$,

$$f_{(X,Y)}(x, y) = \frac{2}{\pi}\, e^{-x(1+y^2)} \mathbb{I}_{\{x,y \geq 0\}}.$$

   (a) Vérifier que $f_{(X,Y)}$ est bien une densité.

   (b) Déterminer les lois de X et de Y. Les variables X et Y sont-elles indépendantes?

Proof. (a) En premier lieu, $f_{(X,Y)}$ est borélienne positive. Ensuite, en utilisant Fubini (les fonctions sont positives):

$$\int_{\mathbf{R}^2} f_{(X,Y)}(x, y)\, dx\, dy = \frac{2}{\pi} \int_0^\infty \Big[- \frac{e^{-x(1+y^2)}}{1+y^2} \Big]_0^\infty dy = \frac{2}{\pi} \int_0^\infty \frac{1}{1+y^2}\, dy = 1.$$

(b) On a

$$\begin{aligned} f_X(x) &= \int_0^\infty f_{(X,Y)}(x, y)\, dy = \frac{2\, e^{-x}}{\pi} \int_0^\infty e^{-x\, y^2}\, dy = \frac{2\, e^{-x}}{\pi\, \sqrt{2\, x}} \int_0^\infty e^{-z^2/2}\, dz = \frac{e^{-x}}{\pi\, \sqrt{2\, x}} \sqrt{2\pi} = \frac{e^{-x}}{\sqrt{\pi\, x}} \quad \text{si } x > 0 \\ &= 0 \quad \text{si } x \leq 0 \\ f_Y(y) &= \int_0^\infty f_{(X,Y)}(x, y)\, dx = \frac{2}{\pi}\, \frac{1}{1+y^2} \quad \text{si } y \geq 0 \\ &= 0 \quad \text{si } y < 0. \end{aligned}$$

Il est clair que les 2 variables ne sont pas indépendantes car $f_X(x)\, f_Y(y) \neq f_{(X,Y)}(x, y)$.

□

2. **(*)** Soient X_1 et X_2 deux v.a. indépendantes de même loi uniforme sur $[0, 1]$.

   (a) Déterminer les fonctions de répartition des v.a. $U = \min\{X_1, X_2\}$ et $V = \max\{X_1, X_2\}$, et en déduire les densités de probabilité de U et V.

   (b) Calculer $\mathrm{cov}(U, V)$. Les variables U et V sont-elles indépendantes?

   (c) Que vaut $\mathbb{E}\big[|X_1 - X_2|\big]$?

Proof. (a) On a $F_V(v) = 0$ pour $v \notin [0, 1]$. Si $v \in [0, 1]$, alors $F_V(v) = \mathbb{P}(X_1 \leq v \cap X_2 \leq v) = v^2$ par indépendance de X_1 et X_2. Comme c'est une fonction continue sur $\mathbf{R}$ et dérivable par morceaux, alors V admet une densité et $f_V(v) = 2\, v\, \mathbb{I}_{v \in [0,1]}$.
De même, $\mathbb{P}(U \leq u) = 1 - \mathbb{P}(U > u) = \mathbb{P}(X_1 > u \cap X_2 > u) = 1 - (1 - u)^2$ pour $u \in [0, 1]$. D'où $f_U(u) = 2\,(1 - u)\, \mathbb{I}_{u \in [0,1]}$.

(b) On a $\mathbb{E}[U] = 2 \int_0^1 u(1 - u) du = [u^2 - \frac{2}{3} u^3]_0^1 = \frac{1}{3}$ et $\mathbb{E}[V] = 2 \int_0^1 v^2 dv = [\frac{2}{3} v^3]_0^1 = \frac{2}{3}$. Et $\mathbb{E}[UV] = \mathbb{E}[X_1 X_2] = \mathbb{E}[X_1]\, \mathbb{E}[X_2] = \frac{1}{4}$ par indépendance. D'où $\mathrm{cov}(U, V) = \frac{1}{4} - \frac{2}{9} = \frac{1}{36}$.
Les variables ne sont pas indépendantes car $\mathrm{cov}(U, V) \neq 0$.

(c) Il est clair que $\mathbb{E}\big[|X_1 - X_2|\big] = \mathbb{E}[V - U] = \mathbb{E}[V] - \mathbb{E}[U] = \frac{1}{3}$.

□

3. **(**)** On considère $X = (X_1, X_2)$ un vecteur aléatoire à valeurs dans $\mathbb{R}^2$. On suppose que X est absolument continue, c'est-à-dire que la mesure de probabilité de X est absolument continue par rapport à la mesure de Lebesgue λ_2 sur $\mathbb{R}^2$.

   (a) Montrer alors que la loi de X_1 admet une densité f_1 par rapport à la mesure de Lebegue λ_1 sur $\mathbb{R}$, que l'on exprimera en fonction de f.

(b) Calculer f_1 et f_2 pour f telle que :

$$f(x_1, x_2) = \begin{cases} e^{-x_1} & \text{si } x_1 \geq x_2 \geq 0 \\ 0 & \text{sinon.} \end{cases}$$

A-t-on $f(x_1, x_2) = f_1(x_1) f_2(x_2)$ pour λ_2-presque tout $(x_1, x_2) \in \mathbb{R}^2$? Quelle conclusion en tirer sur X_1 et X_2?

(c) On suppose maintenant que $X = (X_1, X_1)$ où X_1 est absolument continue par rapport à λ_1. Le vecteur aléatoire X est-il absolument continue par rapport à la mesure de Lebesgue λ_2 sur $\mathbb{R}^2$?

Proof. (a) La fonction de répartition de X_1 est, pour $x \in \mathbf{R}$,

$$F_{X_1}(x) = \mathbb{P}(X_1 \leq x) = \int_{x_1 \leq x} \int_{\mathbf{R}} f(x_1, x_2)\, dx_2\, dx_1 = \int_{-\infty}^{x} \Big(\int_{\mathbf{R}} f(x_1, x_2)\, dx_2 \Big)\, dx_1,$$

par Fubini. Donc $F_{X_1}(x)$ s'écrit sous la forme $\int_{-\infty}^{x} f_1(x_1)\, dx_1$ avec $f_1(x_1) = \int_{\mathbf{R}} f(x_1, x_2)\, dx_2$. Comme f est borélienne positive, f_1 l'est également. Donc la loi de X_1 est absolument continue par rapport à la mesure de Lebesgue sur $\mathbf{R}$, et sa densité est f_1.

(b) On a:

$$\begin{aligned} f_1(x_1) &= \int_0^{x_1} e^{-x_1}\, dx_2\, \mathbb{I}_{x_1 \geq 0} = x_1\, e^{-x_1} \mathbb{I}_{x_1 \geq 0} \quad \text{loi } \Gamma(2, 1) \\ f_2(x_2) &= \int_{x_2}^{\infty} e^{-x_1}\, dx_1\, \mathbb{I}_{x_2 \geq 0} = e^{-x_2} \mathbb{I}_{x_2 \geq 0} \quad \text{loi } \mathcal{E}(1) \end{aligned}$$

Il est clair que $f(x_1, x_2) \neq f_1(x_1)\, f_2(x_2)$, donc les variables X_1 et X_2 ne sont pas indépendantes.

(c) On a $X \in D = \{(x, y) \in \mathbf{R}^2,\ y = x\}$ bissectrice du plan. Donc l'ensemble des valeurs prises par X est un ensemble D de $\mathbf{R}^2$ de mesure de Lebesgue $\lambda_2(D) = 0$: le vecteur aléatoire X n'est pas absolument continue par rapport à la mesure de Lebesgue λ_2 sur $\mathbb{R}^2$, mais il l'est par rapport à la mesure de Lebesgue sur D.

□

4. **(**)** Soit L une v.a. positive admettant une densité de probabilité f et X une v.a. de loi uniforme sur $[0, 1]$ indépendante de L. On définit deux v.a. L_1 et L_2 par $L_1 = X\, L$ et $L_2 = (1 - X)\, L$ (cela représente par exemple la rupture aléatoire en 2 morceaux de longueurs L_1 et L_2 d'une certaine molécule de longueur initiale (aléatoire) L).

   (a) Déterminer la loi du couple (L_1, L_2), ainsi que les lois marginales de L_1 et L_2.

   (b) Que peut-on dire du couple (L_1, L_2) lorsque $f(y) = \mathbb{I}_{[0,+\infty[}(y) \lambda^2 y e^{-\lambda y}$ $(\lambda > 0)$?

   (c) Déterminer la loi de $Z = \min\{L_1, L_2\}$.

Proof. (a) Soit $g : \mathbf{R}^2 \to \mathbf{R}_+$ mesurable. Alors

$$\mathbb{E}[g(L_1, L_2)] = \mathbb{E}[g(X\, L, (1 - X)\, L)] = \mathbb{E}[h(X, L)],$$

où $h(x, \ell) = g(xl, (1 - x)\ell)$.
Le but est de trouver une formule du type

$$\mathbb{E}[g(L_1, L_2)] = \int_{\mathbf{R}^2} g(x_1, x_2) f_{(L_1, L_2)}(x_1, x_2) dx_1 dx_2.$$

Si on prend $g = \mathbb{I}_C$ avec $C \in \mathcal{B}(\mathbf{R}^2)$, cela nous fournira une densité du couple (L_1, L_2). Ou, si par exemple, $g(l_1, l_2) = \mathbb{I}_{]-\infty, x]}(l_1) 1_{]-\infty, y]}(l_2)$, on trouvera que

$$\mathbb{E}[g(L_1, L_2)] = \mathbb{P}(L_1 \leq x,\ L_2 \leq y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{(L_1, L_2)}(x_1, x_2) dx_1 dx_2.$$

Le but de ce qui suit est trouver une formule explicite pour $f_{(L_1, L_2)}$. On travaille avec une fonction g arbitraire (c'est juste plus simple à écrire.)
Par théorème de transfert appliqué au couple (X, L),

$$\mathbb{E}[g(L_1, L_2)] = \mathbb{E}[h(X, L)] = \int_{\mathbf{R}^2} h(x, \ell) \mathbb{P}_{(X, L)}(dx, d\ell).$$

Par indépendance de X et L,

$$\mathbb{P}_{(X, L)}(dx, d\ell) = \mathbb{P}_X(dx) \otimes \mathbb{P}_L(d\ell) = \mathbb{I}_{[0,1]}(x) f(\ell) dx d\ell.$$

Donc,

$$\mathbb{E}[g(L_1,L_2)] = \int_{\mathbf{R}_+} f(\ell)(\int_0^1 h(x,\ell)dx)d\ell = \int_{\mathbf{R}_+} f(\ell)\int_0^1 g(xl,(1-x)\,\ell)dxd\ell.$$

Soit le changement de variables $x_1 = x\,\ell, x_2 = (1-x)\,\ell$. Alors

$$\frac{\partial(x_1,x_2)}{\partial(x,\ell)} = \begin{pmatrix} \ell & x \\ -\ell & 1-x \end{pmatrix},$$

avec det $\frac{\partial(x_1,x_2)}{\partial(x,\ell)} = \ell = x_1 + x_2$. Par conséquent,

$$\mathbb{E}[g(L_1,L_2)] = \int_{\mathbf{R}_+^2} g(x_1,x_2)\frac{f(x_1+x_2)}{x_1+x_2}dx_1dx_2.$$

La densité commune de (L_1,L_2) est donc

$$f_{(L_1,L_2)}(x_1,x_2) = \frac{f(x_1+x_2)}{x_1+x_2}\mathbb{I}_{x_1,x_2\geq 0}.$$

Lois marginales : puisque $X \sim 1-X$, clairement, $L_1 \sim L_2$: les deux coordonnées suivent la même loi. Soit maintenant $g : \mathbf{R}_+ \to \mathbf{R}_+$ une fonction test mesurable.

$$\mathbb{E}[g(L_1)] = \int_{\mathbf{R}_+}\left(\int_0^1 g(\ell\,x)dx\right)f(\ell)d\ell.$$

Changement de variables : $\ell\,x = y$, avec ℓ fixé, donc $\ell\,dx = dy,\ dx = \frac{1}{\ell}dy$. Cela donne

$$\mathbb{E}[g(L_1)] = \int_{\mathbf{R}_+} f(\ell)(\frac{1}{l}\int_0^\ell g(y)dy)dl = \int_0^\infty g(y)\left(\int_y^\infty \frac{f(\ell)}{\ell}d\ell\right)dy,$$

où on a utilisé Fubini. L_1 possède donc la densité

$$f_{L_1}(y) = \int_y^\infty \frac{f(\ell)}{\ell}d\ell, y > 0.$$

(b) Si $f(y) = \mathbb{I}_{[0,+\infty[}(y)\lambda^2 ye^{-\lambda y}$, nous avons

$$f(\ell)/\ell = \lambda^2 e^{-\lambda\,\ell},$$

et

$$\int_y^\infty \frac{f(\ell)}{\ell}d\ell = \lambda e^{-\lambda y}:$$

L_1 et L_2 suivent donc une loi exponentielle de paramètre λ.

(c) $\min(L_1,L_2) = \min(X,1-X)\,L$. Donc, avec $g : \mathbf{R}_+ \to \mathbf{R}_+$ une fonction test mesurable,

$$\begin{aligned}\mathbb{E}[g(Z)] &= \int_0^\infty f(\ell)\left(\int_0^1 g(\min(u,1-u)\,\ell)du\right)d\ell \\ &= \int_0^\infty f(\ell)\left(\int_0^{\frac12} g(\ell\,u)du + \int_{\frac12}^1 g(\ell\,(1-u))du\right)d\ell.\end{aligned}$$

On fait un changement de variables dans la deuxième expression: $1-u = v$, donc $v \in [0,\frac12]$ et

$$\int_{\frac12}^1 g(\ell\,(1-u))du = \int_0^{\frac12} g(\ell\,v)dv.$$

On reconnait la première expression... Donc, en posant $y = \ell\,u$, avec ℓ fixé, $dy = \ell\,du$,

$$\begin{aligned}\mathbb{E}[g(Z)] = 2\int_0^\infty f(\ell)\left(\int_0^{\frac12} g(\ell\,u)\,du\right)d\ell &= 2\int_0^\infty \frac{f(\ell)}{\ell}\left(\int_0^{l/2} g(y)dy\right)d\ell \\ &= 2\int_0^\infty g(y)\left(\int_{2y}^\infty \frac{f(\ell)}{\ell}d\ell\right)dy.\end{aligned}$$

On conclut que pour $y > 0$,

$$f_Z(y) = 2\int_{2y}^\infty \frac{f(\ell)}{\ell}\,d\ell.$$

□

5. **(**)** On considère un couple indépendant de v.a. (X,Y). On suppose que X admet une densité f et que Y est une variable discrète qui prend ses valeurs dans $\{y_n\,,n\in I\}$, $I\subseteq \mathbf{N}$ où $(y_n)_{n\in I}\subset\mathbb{R}$. Montrer que $Z = X+Y$ possède une densité f_Z et donner sa formule.

Proof. Puisque $Z = X + y_n$ sur $\{Y = y_n\}$, on a

$$\begin{aligned}
\mathbb{P}(Z \le z) &= \sum_{n\in I} \mathbb{P}(Y = y_n, X \le z - y_n) = \sum_{n\in I} \mathbb{P}(Y = y_n)\mathbb{P}(X \le z - y_n) = \sum_{n\in I} \mathbb{P}(Y = y_n) F_X(z - y_n) \\
&= \sum_{n\in I} \mathbb{P}(Y = y_n) \int_{-\infty}^{z-y_n} f(x)\,dx = \sum_{n\in I} \mathbb{P}(Y = y_n) \int_{-\infty}^{z} f(u - y_n) du \\
&= \int_{-\infty}^{z} \Big(\sum_{n\in I} \mathbb{P}(Y = y_n) f(u - y_n) \Big)\, du,
\end{aligned}$$

avec le changement de variables $u = x + y_n$ et puis Fubini. Donc, la densité de Z est donnée par

$$f_Z(z) = \sum_{n\in I} P(Y = y_n)\, f(z - y_n).$$

□

6. **(***)** Soit $(X_1, \cdots, X_n)$ un échantillon de n v.a.i.i.d. de loi exponentielle de paramètre 1.

   (a) Montrer que $\mathbb{P}(\exists (i,j) \in \{1, \cdots, n\}^2, i \neq j, X_i = X_j) = 0$.

   (b) On pose $Z = \min_{1\le i\le n} X_i$. Déterminer la loi de Z.

   (c) Soit $N = \min\{1 \le i \le n, X_i = Z\}$. Montrer que N est une v.a. et établir que $\mathbb{P}(N = k,\ Z > t) = \dfrac{e^{-nt}}{n}$ pour $k = 1, \cdots, n$ et $t > 0$. En déduire que Z et N sont des v.a. indépendantes et préciser la loi de N.

Proof. (a) $\mathbb{P}\big(\exists (i,j) \in \{1, \cdots, n\}^2, i \neq j, X_i = X_j\big) = \int_D f(x)f(y)\, d\lambda_2(x,y)$, où D est la première bissectrice, soit $D = \{(x,y) \in \mathbf{R}^2,\ x = y\}$. Comme $\lambda_2(D) = 0$ alors $\int_D f(x)f(y)\, d\lambda_2(x,y) = 0$.

(b) Z prend ses valeurs dans $[0, \infty[$. Pour $z < 0$, $F_Z(z = \mathbb{P}(Z \le z) = 0$. Pour $z \ge 0$,

$$F_Z(z) = 1 - \mathbb{P}\big(X_1 > z \cap \ldots \cap X_n > z\big) = 1 - \Big(\int_z^\infty e^{-x}\, dx \Big)^n = 1 - e^{-n\, z}.$$

En conséquence, Z suit une loi exponentielle de paramètre n.

(c) Si $N = \min\{1 \le i \le n, X_i = Z\}$, cela signifie que N est le plus petit indice pour lequel X_i atteint son minimum. Mais les applications X_i et Z sont mesurables, donc les applications $Y_i = i\, \mathbb{I}_{X_i = Z} + n\, \mathbb{I}_{X_i \neq Z}$ également, d'où l'application $\min_{1\le i\le n}(Y_i)$ également. Donc N est une variable aléatoire.
Deux preuves possibles:

- Par la formule des probabilités totales: $\sum_{j=1}^n \mathbb{P}(N = j, Z > t) = \mathbb{P}(Z > t)$. Mais par le fait que les v.a. (X_i) sont i.i.d., alors $\mathbb{P}(N = j, Z > t) = \mathbb{P}(N = k, Z > t)$ pour tout j. D'où $n\, \mathbb{P}(N = k, Z > t) = \mathbb{P}(Z > t)$, d'où le résultat.
- $\mathbb{P}(N = k, Z > t) = \mathbb{P}(\min_{i\neq k} X_i \ge X_k, X_k > t)$. Comme X_k et $\min_{i\neq k} X_i$ sont indépendants, et comme $\mathbb{P}(\min_{i\neq k} X_i > x_k) = e^{-(n-1)x_k}$ pour $x_k > 0$, on en déduit que:

$$\mathbb{P}(N = k, Z > t) = \int_t^\infty e^{-x_k}\, e^{-(n-1)x_k}\, dx\, dx_k = \frac{e^{-nt}}{n}.$$

De ceci, on en déduit que $\mathbb{P}(N = k \mid Z > t) = \mathbb{P}(N = k\ \cap\ Z > t)/\mathbb{P}(Z > t) = \frac{1}{n}$ et ceci pour tout $k = 1, \ldots, n$ et tout $t > 0$: les deux variables sont indépendantes.
Et $\mathbb{P}(N = k, Z > t) = \frac{1}{n}$ pour tout k: N suit une loi uniforme sur $\{1, \ldots, n\}$.

□

7. **(**)** Soient $X_1, \ldots, X_n$ des v.a.i.i.d., uniformes sur $[0, 1]$,.

   (a) On pose $W_i = -\log(X_i)$. Montrer que W_i suit une loi exponentielle de paramètre 1.

   (b) On rappelle qu'une loi Gamma $\Gamma(\alpha, \beta)$ de paramètres (p, α) avec $\alpha, \beta > 0$ est une loi continue de densité sur $\mathbf{R}$:

   $$f_{(\alpha,\beta)}(x) = \frac{\beta^\alpha}{\Gamma(\alpha)}\, x^{\alpha-1} e^{-\beta\, x}\, \mathbb{I}_{x>0}$$

   Soient U, V indépendants telles que $U \overset{\mathcal{L}}{\sim} \Gamma(\alpha_1, \beta)$ et $V \overset{\mathcal{L}}{\sim} \Gamma(\alpha_2, \beta)$. Quelle est la loi de $U + V$?

   (c) En déduire la loi de $W_1 + \cdots + W_n$.

   (d) Utiliser le résultat précédent pour trouver la loi de $\prod_{i=1}^n X_i$.

Proof. (a) $\mathbb{P}(W_i > x) = \mathbb{P}(-\log(X_i) > x) = \mathbb{P}(\log(X_i) < -x) = \mathbb{P}(X_i < e^{-x}) = \mathbb{P}(X_i \leq e^{-x})$, car X_i possède une densité. Enfin, $\mathbb{P}(X_i \leq e^{-x}) = e^{-x}$, par définition de la loi uniforme. Donc $\mathbb{P}(W_i \leq x) = 1 - e^{-x}$: on reconnait la fonction de répartition d'une loi exponentielle de paramètre 1.

(b) On montre en utilisant la fonction caractéristique que la somme de deux v.a. indépendantes Z_1 et Z_2 de lois $\Gamma(\alpha_1, \beta)$ et $\Gamma(\alpha_2, \beta)$, respectives, suit encore une loi Gamma: $\Gamma(\alpha_1 + \alpha_2, \beta)$. En effet, la fonction caractéristique d'une loi $\Gamma(\alpha, \beta)$ est $\phi(u) = \big(1 - i\,\frac{u}{\beta}\big)^{\alpha}$. D'où $\phi_{Z_1+Z_2}(u) = \phi_{Z_1}(u)\,\phi_{Z_2}(u) = \big(1 - i\,\frac{u}{\beta}\big)^{\alpha_1}\big(1 - i\,\frac{u}{\beta}\big)^{\alpha_2} = \big(1 - i\,\frac{u}{\beta}\big)^{\alpha_1+\alpha_2}$ en utilisant l'indépendance entre Z_1 et Z_2.

(c) Puisque la loi exponentielle de paramètre β est une $\Gamma(1, \beta)$, nous avons donc que $W_1 \overset{\mathcal{L}}{\sim} \Gamma(1, 1)$ et donc $W_1 + W_2 + \cdots + W_n \overset{\mathcal{L}}{\sim} \Gamma(n, 1)$.

(d) Soit $Y = W_1 + W_2 + \cdots + W_n \overset{\mathcal{L}}{\sim} \Gamma(n, 1)$. Donc, pour $x \in (0, 1)$,

$$\mathbb{P}(X_1 \cdot X_2 \cdot \ldots \cdot X_n \leq x) = \mathbb{P}(e^{-Y} \leq x) = \mathbb{P}(-Y \leq \log x) = \mathbb{P}(Y \geq -\log x) = 1 - F_Y(-\log x),$$

avec F_Y la fonction de répartition de Y. La densité de $X_1 \cdot X_2 \cdot \ldots \cdot X_n$ est donc donnée par

$$\frac{1}{x}\, f_{(n,1)}(-\log x)\, \mathbb{I}_{x\in(0,1)},$$

avec $f_{(n,1)}$ la densité de la loi $\Gamma(n, 1)$.

$\square$

8. **(**)** Soient X et Y deux variables aléatoires exponentielles indépendantes de paramètres $\alpha > 0$ et $\beta > 0$. On pose $S = \min(X, Y)$ et $T = |X - Y|$.

   (a) Calculer $\mathbb{P}(S > a, T > b, X > Y)$ et $\mathbb{P}(S > a, T > b, X < Y)$.

   (b) En déduire $\mathbb{P}(X < Y)$, la loi de S, et la loi de T.

Proof. (a) Choisissons a et b des réels positifs (les autres cas ne sont pas informatifs). Alors

$$\begin{aligned}\mathbb{P}(S > a, T > b, X > Y) = \mathbb{P}(Y > a, X - Y > b) &= \int_a^{\infty} \beta\, e^{-\beta\, y} \int_{b+y} \alpha\, e^{-\alpha\, x} dx\, dy \\ &= \int_a^{\infty} \beta\, e^{-\alpha\,(b+y) - \beta\, y}\, dy = \frac{\beta\, e^{-\alpha\, b - (\alpha+\beta)\, a}}{\alpha + \beta}.\end{aligned}$$

Par symmétrie, $\mathbb{P}(S > a, T > b, X < Y) = \dfrac{\alpha\, e^{-\beta\, b - (\alpha+\beta)\, a}}{\alpha + \beta}$.

(b) Il suffit de choisir $a = b = 0$ pour en déduire $\mathbb{P}(X < Y) = \dfrac{\alpha\, e^{-\beta\, b - (\alpha+\beta)\, a}}{\alpha + \beta}$.
Par ailleurs, on choisissant $b = 0$, on a

$$\mathbb{P}(S > a) = \mathbb{P}(S > a, X < Y) + \mathbb{P}(S > a, X < Y) = \frac{\beta\, e^{-(\alpha+\beta)\, a}}{\alpha + \beta} + \frac{\alpha\, e^{-(\alpha+\beta)\, a}}{\alpha + \beta} = e^{-(\alpha+\beta)\, a}$$

donc S suit une loi exponentielle de paramètre $\alpha + \beta$.
Pour la loi de T, on fixe $a = 0$ et

$$\mathbb{P}(T > b) = \mathbb{P}(T > b, X < Y) + \mathbb{P}(T > b, X < Y) = \frac{\beta\, e^{-\alpha\, b}}{\alpha + \beta} + \frac{\alpha\, e^{-\beta\, b}}{\alpha + \beta} = \frac{\beta\, e^{-\alpha\, b} + \alpha\, e^{-\beta\, b}}{\alpha + \beta}.$$

On en déduit que la densité de T est $\dfrac{\alpha\,\beta}{\alpha + \beta}\left(e^{-\alpha\, x} + e^{-\beta\, x}\right) \mathbb{I}_{x\geq 0}$.

$\square$

9. **(**)** Soit (X_1, X_2, X_3) vecteur aléatoire centré de matrice de covariance

$$A = \begin{bmatrix} 2 & 1 & 3 \\ 1 & 5 & 6 \\ 3 & 6 & 9 \end{bmatrix}$$

   (a) Calculer la variance de $X_3 - \alpha_1 X_1 - \alpha_2 X_2$.

   (b) En déduire que X_3 est une combinaison linéaire de X_1 et X_2 p.s.

(c) Plus généralement, pour un vecteur aléatoire Y de matrice de covariance Γ, donner une condition nécessaire et suffisante sur Γ pour que l'une des composantes de Y soit une fonction affine des autres composantes de Y p.s.

(d) Soit maintenant Z un vecteur aléatoire à valeurs dans $\mathbf{R}^d$, $d \geq 1$. Supposons que Z admette une densité f par rapport à la mesure de Lebesgue sur $\mathbf{R}^d$. Soit $x \in \mathbf{R}^d$ un vecteur non-nul. Montrer qu'alors la v.a. $U = {}^t x\, Z$ a une densité sur $\mathbf{R}$.

(e) Si Y est un vecteur aléatoire de matrice de covariance non-inversible, peut-il avoir une densité?

Proof. (a) On a $\mathrm{var}(X_3 - \alpha_1 X_1 - \alpha_2 X_2) = \mathrm{cov}\big((-\alpha_1, -\alpha_2, 1)\, X\big) = (-\alpha_1, -\alpha_2, 1)\, A\, {}^t(-\alpha_1, -\alpha_2, 1)$, donc $\mathrm{var}(X_3 - \alpha_1 X_1 - \alpha_2 X_2) = 2\,\alpha_1^2 + 5\,\alpha_2^2 + 2\,\alpha_1\,\alpha_2 - 6\,\alpha_1 - 12\,\alpha_2 + 9$.

(b) On peut calculer le déterminant de la matrice A, et on montre que $\det(A) = 0$. Donc 0 est valeur propre. On peut alors déterminer le sous-espace propre associé à 0. Cela revient à résoudre le système:

$$\left\{ \begin{array}{rcl} 2x + y + 3z & = & 0 \\ x + 5y + 6z & = & 0 \end{array} \right. \Longleftrightarrow \left\{ \begin{array}{rcl} x & = & y \\ z & = & -y \end{array} \right. .$$

Le sous-espace, qui est de dimension 1 est donc généré par le vecteur $(1, 1, -1)$. On en déduit qu'en choisissant $\alpha_1 = \alpha_2 = 1$ alors $\mathrm{var}(X_3 - \alpha_1 X_1 - \alpha_2 X_2) = 0$, soit $X_3 = X_1 + X_2$ p.s.

(c) Il est clair que cette CNS est "la matrice de covariance admet 0 comme valeur propre" (ou bien son déterminant est nul).

(d) Comme x est un vecteur non nul, on peut alors considérer $(f_2, \ldots, f_d)$ famille orthonormée de $d-1$ vecteurs de $\mathbf{R}^d$ telle que $(\frac{x}{\|x\|}, f_2, \ldots, f_d)$ soit une base orthonormale de R^d (on a $\|x\| > 0$ car x non nul). Ainsi, pour $u \in \mathbf{R}$, on a

$$\begin{aligned} \mathbb{P}(U \leq u) &= \int_{{}^t xz \leq u} f\Big(({}^t xz)\, \frac{x}{\|x\|^2} + \sum_{j=2}^{d} < f_j, z > f_j\Big)\, d\lambda_d(z) = \int_{\|x\| z_1' \leq u} f\Big(z_1'\, \frac{x}{\|x\|} + \sum_{j=2}^{d} z_j' f_j\Big)\, dz_1' \ldots dz_d' \\ &= \int_{z_1' \leq \frac{u}{\|x\|}} \Big(\int_{\mathbf{R}^{d-1}} f\Big(z_1'\, \frac{x}{\|x\|} + \sum_{j=2}^{d} z_j' f_j\Big)\, dz_2' \ldots dz' d\Big) dz_1' \end{aligned}$$

après un changement de variable de déterminant $= 1$ (changement d'une base orthonormale à une autre base orthonormale) et en utilisant Fubini. si l'on note

$$f_x(z_1') = \int_{\mathbf{R}^{d-1}} f\Big(z_1'\, \frac{x}{\|x\|} + \sum_{j=2}^{d} z_j' f_j\Big)\, dz_2' \ldots dz' d$$

on a pu écrire F_U sous la forme $\int_{-\infty}^{\frac{u}{\|x\|}} f_x(z_1') dz_1' = \int_{-\infty}^{u} \|x\|\, f_x(t/\|x\|)\, dt$, où f_x est une fonction positive mesurable, donc U est une variable continue par rapport à la mesure de Lebesgue.

(e) On montre que Y a une densité sur le sous-espace vectoriel de $\mathbf{R}^d$ constitué par la somme directe des sous-espaces propres des valeurs propres non nulles de la matrice de coviance.

□

Feuille n^o 3:
Vecteurs gaussiens

1. **(*)** Soit $X = (X_1, X_2, X_3)$ un vecteur gaussien centré de matrice de covariance

$$\Gamma = \begin{pmatrix} 3 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

   (a) Quelle est la loi de X_3 et celle de (X_1, X_2) et que peut-on dire de ces 2 vecteurs aléatoires?
   (b) Déterminer la densité de la loi de (X_1, X_2, X_3).
   (c) Quelle est la loi de $(X_1 - X_2, X_3 - X_1)$?

Proof. (a) $X_3 = A\,X$ avec $A = {}^t(0,0,1)$ est donc une variable gaussienne comme combinaison linéaire issue de X et $\mathbb{E}[X_3] = A\,\mathbb{E}[X] = 0$, $\mathrm{var}(X_3) = A\,\mathrm{cov}(X)\,{}^tA = 1$. Donc $X_3 \overset{\mathcal{L}}{\sim} \mathcal{N}(0,1)$.

$(X_1, X_2) = B\,X$ avec $B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$. Donc le vecteur (X_1, X_2) est gaussien de loi $\mathcal{N}\Big(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 3 & 1 \\ 1 & 2 \end{pmatrix}\Big)$.

(X_1, X_2) et X_3 sont deux vecteurs issus du même vecteur gaussien. De plus, pour tous $(a_1, a_2) \in \mathbf{R}^2$, $\mathrm{cov}(X_3, a_1X_1 + a_2X_2) = a_1\mathrm{cov}(X_3, X_1) + a_2\mathrm{cov}(X_3, X_2) = 0$ d'après la matrice de covariance. Donc (X_1, X_2) et X_3 sont indépendants.

(b) La densité de la loi de (X_1, X_2, X_3) est directement donnée par le cours:

$$\begin{aligned} f(x_1, x_2, x_3) &= \frac{1}{(2\pi)^{3/2}\sqrt{\det(\Gamma)}} \exp\Big(-\frac{1}{2}\,(x_1, x_2, x_3)\Gamma^{-1}\,{}^t(x_1, x_2, x_3)\Big) \\ &= \frac{1}{(2\pi)^{3/2}\sqrt{5}} \exp\left(-\frac{1}{10}\,(x_1, x_2, x_3) \begin{pmatrix} 2 & -1 & 0 \\ -1 & 3 & 0 \\ 0 & 0 & 5 \end{pmatrix} {}^t(x_1, x_2, x_3)\right) \end{aligned}$$

(c) $Z = (X_1 - X_2, X_3 - X_1) = B\,X$ avec $B = \begin{pmatrix} 1 & -1 & 0 \\ -1 & 0 & 1 \end{pmatrix}$. Donc Z est gaussien. Son espérance est 0 et sa matrice de variance covariance est:

$$\mathrm{cov}(Z) = B\,\mathrm{cov}(X)\,{}^tB = \begin{pmatrix} 1 & -1 & 0 \\ -1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 3 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ -1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 3 & -2 \\ -2 & 4 \end{pmatrix}$$

□

2. **(**)** Soit X une v.a. réelle normale centrée réduite et soit Y une v.a. indépendante de X, à valeurs dans $\{-1, 1\}$ telle que $\mathbb{P}(Y = 1) = 0.5$. On considère la v.a. $Z = X\,Y$.

   (a) Déterminer la mesure de probabilité de Z.
   (b) Déterminer $\mathrm{cov}(X, Z)$. Les variables X et Z sont-elles indépendantes?
   (c) Déterminer la mesure de probabilité de $X + Z$. En déduire que la somme de 2 variables gaussiennes non-corrélées peut ne pas être gaussienne.

Proof. (a) Pour tout $z \in \mathbf{R}$, on a $\mathbb{P}(Z \le z) = \mathbb{P}(Z \le z \cap Y = 1) + \mathbb{P}(Z \le z \cap Y = -1)$ par la formule des probabilités totales. D'où $\mathbb{P}(Z \le z) = \mathbb{P}(Z \le z \mid Y = 1)\mathbb{P}(Y = 1) + \mathbb{P}(Z \le z \mid Y = -1)\mathbb{P}(Y = -1) = \frac{1}{2}\big(\mathbb{P}(X \le z) + \mathbb{P}(-X \le z)\big)$ car X et Y sont indépendantes. Or $\mathbb{P}(-X \le z) = \mathbb{P}(X \le z)$ car X a une loi symétrique, donc $\mathbb{P}(Z \le z) = \mathbb{P}(X \le z)$: $Z \overset{\mathcal{L}}{\sim} \mathcal{N}(0,1)$.

(b) $\mathrm{cov}(X, Z) = \mathbb{E}[X\,Z] = \mathbb{E}[Y\,X^2] = \mathbb{E}[Y]\,\mathbb{E}[X^2] = 0$ car Y est indépendante de X et du fait que $\mathbb{E}[Y] = 0$.
X et Y ne sont pas indépendantes car $\mathbb{P}(|X| < 1) > 0$, $\mathbb{P}(|Z| > 1) > 0$ et $\mathbb{P}(|X| < 1 \cap |Z| > 1) = 0$, donc $\mathbb{P}(|X| < 1 \cap |Z| > 1) \neq \mathbb{P}(|X| < 1)\,\mathbb{P}(|Z| > 1)$.

(c) On a $X + Z = X + X\,Y = X(1+Y) = 2\,X\,U$, où $U \overset{\mathcal{L}}{\sim} \mathcal{B}(1/2)$ et U indépendante de X. Donc pour tout $B \in \mathcal{B}(\mathbf{R})$, alors $\mathbb{P}(X + Z \in B) = \frac{1}{2}\,\delta_{\{0\}\in B} + \frac{1}{2}\,\mathbb{P}_{2X}(B)$, où $\mathbb{P}_{2X}$ est une loi $\mathcal{N}(0,4)$.
On a donc X et Z qui sont deux v.a. gaussiennes centrées réduites, non corrélées, mais telles que X et Z ne sont pas indépendantes et une combinaison linéaire de X et Z, telle que $X + Z$, ne sont pas gaussienne.

□

3. **(*)** Soit X et Y deux v.a. indépendantes de loi commune $\mathcal{N}(0,1)$. Déterminer la loi de $Z = \frac{1}{\sqrt{2}}\big(X+Y\big)$, celle de $W = \frac{1}{2}(X-Y)^2$ et enfin celle de $Z/\sqrt{W}$.

Proof. Comme X et Y sont deux v.a. indépendantes de loi $\mathcal{N}(0,1)$, alors le vecteur (X,Y) est gaussien (centré réduit). Comme Z et $Z' = \frac{1}{\sqrt{2}}\big(X-Y\big)$ sont des combinaisons linéaires de (X,Y), alors le vecteur $\big(Z,Z'\big)$ est gaussien.
Par ailleurs, $E[Z] = \mathbb{E}[Z'] = 0$, $\mathrm{var}(Z) = \mathrm{var}(Z') = 1$ et $\mathrm{cov}(Z,Z') = \frac{1}{2}\big(\mathrm{var}(X^2) - \mathrm{var}(Y^2)\big) = 0$. Comme le vecteur $\big(Z,Z'\big)$ est gaussien, on en déduit que Z et Z' sont indépendantes et de loi $\mathcal{N}(0,1)$.
Comme $Z' \overset{\mathcal{L}}{\sim} \mathcal{N}(0,1)$ et $W = (Z')^2$ alors $W \overset{\mathcal{L}}{\sim} \chi^2(1)$.
Enfin, comme Z et W sont indépendantes, $Z/\sqrt{W} \overset{\mathcal{L}}{\sim} t(1)$, loi de student à 1 degré de liberté. □

4. **(**)** Soient X et Y des v.a. indépendantes de loi $\mathcal{N}(0,1)$.
   (a) Déterminer la loi du couple de $(X+Y, X-Y)$. Que remarque-t-on?
   (b) Déterminer également la loi du couple $(X/Y, Y)$ puis celle de la v.a. X/Y. Les v.a. X/Y et Y sont elles indépendantes?
   (c) En déduire la densité de la loi de Student de degré 1.

Proof. (a) Voir l'exercice précédent $(X+Y, X-Y) \overset{\mathcal{L}}{\sim} \mathcal{N}\Big(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}\Big)$ et $(X+Y)$ et $X-Y$ indépendantes.

(b) Le couple $(X/Y, Y)$ prend ses valeurs dans $\mathbf{R}^2$. De plus, les lois de X/Y et Y sont symétrique. Pour déterminer la densité du couple, considérons $u \in \mathbf{R}$ et $y < 0$ (le cas $y \geq 0$ sera obtenu par symmétrie. Alors:

$$\begin{aligned} F_{(X/Y,Y)}(u,y) &= \mathbb{P}(X/Y \leq u \cap Y \leq y) \\ &= \mathbb{P}(X \geq uY \cap Y \leq y) \\ &= \int_{-\infty}^{y} f_X(y') \int_{uy'}^{\infty} f_X(x')\, dx'\, dy' \quad \text{(Fubini)} \\ &= \int_{-\infty}^{y} f_X(y')\,(1 - F_X(uy'))\, dy'. \end{aligned}$$

En considérant $\frac{\partial^2}{\partial y\, \partial u} F_{(X/Y,Y)}(u,y)$ on obtient donc que pour $y < 0$ et $u \in \mathbf{R}$,

$$f_{(X/Y,Y)}(u,y) = \frac{\partial^2}{\partial y\, \partial u} F_{(X/Y,Y)}(u,y) = -y\, f_X(y)\, f_X(uy).$$

On en déduit que pour tout $(u,y) \in \mathbf{R}^2$,

$$f_{(X/Y,Y)}(u,y) = \frac{|y|}{2\,\pi} \exp\Big(-\frac{1}{2}\, y^2(1+u^2)\Big).$$

On peut alors facilement déduire la densité de X/Y:

$$f_{X/Y}(u) = \int_{-\infty}^{\infty} \frac{|y|}{2\,\pi} \exp\Big(-\frac{1}{2}\, y^2(1+u^2)\Big)\, dy = \frac{1}{\pi}\, \frac{1}{1+u^2},$$

c'est-à-dire la densité d'une loi de Cauchy.
Les v.a. X/Y et Y ne sont donc pas indépendantes car le produit des densités est différent de la densité du couple.

(c) La loi de Student de degré 1 est celle de $X/\sqrt{Y^2}$, puisque X et Y sont indépendantes et $Y^2 \overset{\mathcal{L}}{\sim} \chi^2(1)$. Mais $\sqrt{Y^2} = |Y|$, et par symétrie $X/|Y| \overset{\mathcal{L}}{\sim} X/Y$. La loi de Student de degré 1 est donc la loi de Cauchy.

□

5. **(**)** Soit $X = (X_1, X_2)$ un vecteur gaussien centré de matrice de covariance

$$\Gamma = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}.$$

   (a) Montrer que Γ est bien une matrice de variance-covariance et déterminer ses valeurs propres et leurs sous-espaces propres associés.
   (b) Démontrer que $\mathbb{E}[(2X_1 - X_2)^2] = 0$. En déduire la densité de la loi de (X_1, X_2) par rapport à une mesure que l'on précisera.
   (c) Généraliser à un vecteur gaussien quelconque dont la matrice de covariance est singulière.

Proof. (a) On a Γ symétrique, Trace$(\Gamma) = 5 > 0$ et $\det(\Gamma) = 0 \geq 0$: la matrice Γ est bien celle d'une variance-covariance.
On déduit que 0 et 5 sont les 2 valeurs propres. Le sous-espace propre associé à 0 est $\{(x,y) \in \mathbf{R}^2,\ x + 2y = 0\}$ donc engendré par le vecteur $(2,-1)$. Le sous-espace propre associé à 5 est $\{(x,y) \in \mathbf{R}^2,\ -4x + 2y = 0\}$ donc engendré par le vecteur $(1,2)$.

(b) On a $\mathbb{E}[(2X_1 - X_2)^2] = \mathrm{var}(2X_1 - X_2) = 4\mathrm{var}(X_1) + \mathrm{var}(X_2) - 4\mathrm{cov}(X_1, X_2) = 4 + 4 - 8 = 0$.
On a donc $2X_1 = X_2$ p.s.
De ce qui précède on en déduit que le vecteur X prend uniquement ses valeurs sur la droite $2x = y$. Et sa mesure de probabilité sur cette droite est gaussienne et centrée:

$$\mathbb{P}\big((X_1, X_2) \in A\big) = \int_A \frac{1}{\sqrt{2\pi}}\, e^{-t^2/2} d\lambda_{\{y=2x\}}(t)$$

La densité de (X_1, X_2) sur la droite $y = 2x$ est la densité gaussienne centrée réduite.

(c) Si la matrice de covariance est singulière, alors 0 est valeur propre de multiplicité $m_0 \geq 1$. Or si $X \overset{\mathcal{L}}{\sim} \mathcal{N}_n\big(0, \Gamma\big)$, alors comme Γ est symétrique, il existe une matrice Q orthogonale (telle que $Q^t Q = I_n$), $\Gamma = Q\, D\, {}^tQ$ avec D une matrice diagonale contenant les valeurs propres, et on supposera que les m_0 premières valeurs sont des 0. On peut alors écrire que $X = \Gamma^{1/2}\, Z$, où $Z \overset{\mathcal{L}}{\sim} \mathcal{N}(0, I_n)$, donc $X = Q\, D^{1/2}\, {}^tQ\, Z$. Mais on a $Z' = {}^t(Z'_1, \ldots, Z'_n) = {}^tQ\, Z \overset{\mathcal{L}}{\sim} \mathcal{N}(0, I_n)$, d'où $D^{1/2}\, {}^tQ\, Z = D^{1/2}\, Z' = {}^t(0, \ldots, 0, \lambda_1 Z'_{m_0}, \ldots, \lambda_p Z'_n)$, où les $\lambda_1, \ldots, \lambda_p$ sont les autres valeurs propres que 0. Donc X ne dépend que des $Z'_{m_0+1}, \ldots, Z'_n$, X est donc un vecteur gaussien appartenant uniquement à $E_0^\perp$ (orthogonal du sev propre associé à 0) de dimension $n - m_0$.

□

6. **(***)** Soit $X = (X_1, X_2, X_3, X_4)$ un vecteur gaussien centré de matrice de covariance Γ.

(a) Démontrer que $\mathbb{E}\big[e^{<s,X>}\big] = e^{\frac{1}{2}\, {}^ts\, \Gamma\, s}$ pour tout $s \in \mathbf{R}^4$. En utilisant l'unicité du développement en série entière, en déduire que $\mathbb{E}\big[< s, X >^4 \big] = 3\, \big({}^ts\, \Gamma\, s\big)^2$. En déduire que

$$\mathbb{E}\big[X_1 X_2 X_3 X_4\big] = \mathbb{E}[X_1 X_2]\, \mathbb{E}[X_3 X_4] + \mathbb{E}[X_1 X_3]\, \mathbb{E}[X_2 X_4] + \mathbb{E}[X_1 X_4]\, \mathbb{E}[X_2 X_3].$$

(b) Déduire également que $\mathbb{E}[X_1 X_2 X_3] = 0$.

(c) Si (X, Y) est un vecteur gaussien d'espérance m et de matrice de variance-covariance $\Gamma = \begin{pmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_X^2 \end{pmatrix}$, en déduire $\mathrm{var}(X^2)$ et $\mathrm{cov}(X^2, Y^2)$.

Proof. (a) On va utiliser la densité de X. Ainsi pour tout $s \in \mathbf{R}^4$,

$$\begin{aligned}
\mathbb{E}\big[e^{<s,X>}\big] &= \int_{\mathbf{R}^4} \frac{1}{(2\pi)^2 \sqrt{\det(\Gamma)}} \exp\Big({}^ts\, x - \frac{1}{2}\, {}^tx\, \Gamma^{-1}\, x\Big)\, dx \\
&= \int_{\mathbf{R}^4} \frac{1}{(2\pi)^2} \exp\Big({}^ts\, \Gamma^{1/2}\, z - \frac{1}{2}\, {}^tz\, z\Big)\, dz \\
&= \int_{\mathbf{R}^4} \frac{1}{(2\pi)^2} \exp\Big(-\frac{1}{2}\Big({}^t(z - \Gamma^{1/2}\, s)(z - \Gamma^{1/2}\, s) - {}^ts\, \Gamma\, s\Big)\Big)\, dz \\
&= e^{\frac{1}{2}\, {}^ts\, \Gamma\, s} \int_{\mathbf{R}^4} \frac{1}{(2\pi)^2} \exp\Big(-\frac{1}{2}\, {}^tz'\, z'\Big)\, dz' \\
&= e^{\frac{1}{2}\, {}^ts\, \Gamma\, s}.
\end{aligned}$$

On sait que $e^x = \sum_{k=0}^\infty \frac{x^k}{k!}$ pour tout $x \in \mathbf{R}$. Donc $\mathbb{E}\big[e^{<s,X>}\big] = \mathbb{E}\Big[\sum_{k=0}^\infty \frac{<s,X>^k}{k!}\Big]$ pour tout $s \in \mathbf{R}^4$. Par ailleurs, la série étant convergente et les moments $\mathbb{E}\big[|<s,X>|^k\big] < \infty$ étant tous finis (car $< s, X >$ est une variable gaussienne) alors pour tout $s \in \mathbf{R}^4$,

$$\mathbb{E}\big[e^{<s,X>}\big] = 1 + \sum_{k=1}^\infty \frac{\mathbb{E}\big[< s, X >^k \big]}{k!}.$$

Mais on a également pour tout $s \in \mathbf{R}^4$,

$$e^{\frac{1}{2}\, {}^ts\, \Gamma\, s} = 1 + \sum_{j=1}^\infty \frac{\big({}^ts\, \Gamma\, s\big)^j}{2^j\, j!}.$$

Par unicité du développement en série entière, il y a donc égalité des 2 développements et donc égalité des coefficients devant les différents moments en s. Pour le moment d'ordre 4, on en déduit donc que:

$$\frac{\mathbb{E}\big[< s, X >^4 \big]}{4!} = \frac{\big({}^ts\, \Gamma\, s\big)^2}{2^2\, 2!} \quad \Longrightarrow \quad \mathbb{E}\big[< s, X >^4 \big] = 3\, \big({}^ts\, \Gamma\, s\big)^2 \quad \text{pour tout } s \in \mathbf{R}^4.$$

Si $s = {}^t(s_1, s_2, s_3, s_4)$, en développant $< s, X >^4 = (s_1X_1 + s_2X_2 + s_3X_3 + s_4X_4)^4$, on obtient que le terme en $s_1s_2s_3s_4$ est $6\, X_1X_2X_3X_4$, donc celui de $\mathbb{E}\big[< s, X >^4 \big]$ est $6\,\mathbb{E}[X_1X_2X_3X_4]$.
D'autre part, ${}^t s\,\Gamma\, s = \sum_{i=1}^4 s_i^2 \mathbb{E}[X_i^2] + \sum_{1\leq i<j\leq 4} s_i s_j \mathbb{E}[X_iX_j]$, donc si l'on regarde le terme en $s_1s_2s_3s_4$ de $\big({}^t s\,\Gamma\, s\big)^2$, on obtient:

$$2\,\big(\mathbb{E}[X_1X_2]\,\mathbb{E}[X_3X_4] + \mathbb{E}[X_1X_3]\,\mathbb{E}[X_2X_4] + \mathbb{E}[X_1X_4]\,\mathbb{E}[X_2X_3]\big)$$

Par conséquent, en égalisant les 2 termes en $s_1s_2s_3s_4$, on obtient bien que:

$$\mathbb{E}\big[X_1X_2X_3X_4\big] = \mathbb{E}[X_1X_2]\,\mathbb{E}[X_3X_4] + \mathbb{E}[X_1X_3]\,\mathbb{E}[X_2X_4] + \mathbb{E}[X_1X_4]\,\mathbb{E}[X_2X_3].$$

(b) Si l'on prend dans le développement précédent $\mathbb{E}\big[< s, X >^3 \big]$, l'égalité des développements en série entière montre que nécessairement $\mathbb{E}\big[< s, X >^3 \big] = 0$, donc par exemple $\mathbb{E}[X_1X_2X_3] = 0$.

(c) Commençons par le cas où $m = {}^t(0,0)$. Alors $\mathrm{var}(X^2) = \mathbb{E}[X^4] - \mathbb{E}[X^2]^2 = 3\,\sigma_X^4 - \sigma_X^4 = 2\,\sigma_X^4$ d'après l'égalité précédente (en prenant $X_1 = X_2 = X_3 = X_4 = X$) et $\mathrm{cov}(X^2, Y^2) = \mathbb{E}[X^2Y^2] - \sigma_X^2\sigma_Y^2 = 2\,\mathbb{E}[X\,Y]\,\mathbb{E}[X\,Y] = 2\,\sigma_{XY}^2$ (en prenant $X = X_1 = X_2$ et $Y = X_3 = X_4$).
Si (X, Y) est maintenant un vecteur gaussien d'espérance $m = {}^t(\mathbb{E}[X], \mathbb{E}[Y])$, on en déduit que

$$\mathrm{var}(X^2) = 4\,\mathbb{E}[X]^2\,\sigma_X^2 + 2\,\sigma_X^4 \quad \text{et} \quad \mathrm{cov}(X^2, Y^2) = 4\,\mathbb{E}[X]\,\mathbb{E}[Y]\,\sigma_{XY} + 2\,\sigma_{XY}^2.$$

□

Feuille n^o 4:
Convergence et théorèmes limites

0. **(*)** Soit X_0 une v.a. définie sur $(\Omega, \mathcal{A}, \mathbb{P})$ et pour $n \in \mathbf{N}^*$, on définit $X_n = X_0/n$. Démontrer que (X_n) converge en loi, en probabilité et presque-sûrement vers une limite que l'on précisera. Qu'en est-il pour la convergence dans $\mathbb{L}^p$?

Proof.

- Pour la convergence en loi, on commence par la caractérisation avec les fonctions de répartition. Ainsi pour $x \in \mathbf{R}$, on a:
$$F_{X_n}(x) = \mathbb{P}(X_n \leq x) = \mathbb{P}(X_0 \leq x\, n).$$
Quand $x > 0$ alors $x\, n \underset{n\to\infty}{\longrightarrow} \infty$, et $\lim_{n\to\infty} \mathbb{P}(X_0 \leq x\, n) = 1$.
Quand $x < 0$ alors $x\, n \underset{n\to\infty}{\longrightarrow} -\infty$, et $\lim_{n\to\infty} \mathbb{P}(X_0 \leq x\, n) = 0$. Or il n'existe qu'une seule v.a. dont la fonction caractéristique vaut 0 sur $]|\infty, 0[$ et 1 sur $]0, \infty[$, c'est celle de masse de Dirac en 0, donc presque sûrement la constante 0. Donc $X_n \overset{\mathcal{L}}{\underset{n\to\infty}{\longrightarrow}} 0$ (on notera que la convergence n'a pas lieu pour $x = 0$, mais ce n'est pas grave puisque la fonction de répartition de 0 n'est pas continue en 0).
- On peut également obtenir la convergence en loi avec les fonctions caractéristiques. Ainsi on a pour tout $u \in \mathbf{R}$:
$$\phi_{X_n}(u) = \mathbb{E}\big[e^{i\,u X_n}\big] = \mathbb{E}\big[e^{i\,u X_0/n}\big] = \int_\Omega e^{i\,u X_0(\omega)/n}\, d\mathbb{P}(\omega).$$
Or pour tout ω et tout u, on a $e^{i\,u X_0(\omega)/n} \underset{n\to\infty}{\longrightarrow} 1$. De plus pour tout $n \in \mathbf{N}^*$, $\omega \in \Omega$ et $u \in \mathbf{R}$, $|e^{i\,u X_0(\omega)/n}| \leq 1$, et $\int_\Omega 1\, d\mathbb{P} = 1 < \infty$. On peut donc appliquer le Théorème de convergence dominée de Lebesgue et pour tout $u \in \mathbf{R}$, $\phi_{X_n}(u) \underset{n\to\infty}{\longrightarrow} 1$. Et comme 1 est la fonction caractéristique de la constante 0, par le théorème d'inversion on en déduit que $X_n \overset{\mathcal{L}}{\underset{n\to\infty}{\longrightarrow}} 0$.
- On montre également la convergence en loi avec la caractérisation par les espérances. Soit $g : \mathbf{R} \to \mathbf{R}$, bornée et continue. Alors:
$$\mathbb{E}\big[g(X_n)\big] = \mathbb{E}\big[g(X_0/n)\big] = \int_\Omega g(X_0(\omega)/n)\, d\mathbb{P}(\omega).$$
Or g est continue sur $\mathbf{R}$, donc comme pour tout $\omega \in \Omega$, $X_0(\omega)/n \underset{n\to\infty}{\longrightarrow} 0$, $gX_0(\omega)/n) \underset{n\to\infty}{\longrightarrow} g(0)$ par continuité en 0. Par ailleurs, g étant bornée, pour tout $n \in \mathbf{N}^*$ et $\omega \in \Omega$, $|gX_0(\omega)/n)| \leq \sup|g|$ et $\int_\Omega \sup|g|\, d\mathbb{P} = \sup|g| < \infty$. On peut donc appliquer le Théorème de convergence dominée de Lebesgue et pour tout g continue bornée, $\mathbb{E}\big[g(X_n)\big] \underset{n\to\infty}{\longrightarrow} \mathbb{E}\big[g(0)\big] = g(0)$. On en déduit que $X_n \overset{\mathcal{L}}{\underset{n\to\infty}{\longrightarrow}} 0$.
- On va considérer la convergence en probabilité. On sait que X_n converge en loi vers 0. Or si (X_n) convergeait en probabilité vers une autre limite que 0, comme la convergence en probabilités entraîne la convergence en loi, il y aurait une contradiction. Donc si (X_n) converge en probabilité ce ne peut être que vers 0. Pour le démontrer, on pose $\varepsilon > 0$ et on a
$$\mathbb{P}\big(|X_n - 0| \geq \varepsilon\big) = \mathbb{P}\big(|X_0| \geq \varepsilon\, n\big) = \mathbb{P}\big(X_0 \geq \varepsilon\, n\big) + \mathbb{P}\big(X_0 \leq -\varepsilon\, n\big).$$
On a $\mathbb{P}\big(X_0 \leq -\varepsilon\, n\big) = F_{X_0}(-\varepsilon\, n) \underset{n\to\infty}{\longrightarrow} 0$ car $-\varepsilon\, n \underset{n\to\infty}{\longrightarrow} -\infty$. De plus $\mathbb{P}\big(X_0 \geq \varepsilon\, n\big) = 1 - \mathbb{P}\big(X_0 < \varepsilon\, n\big)$ et $\mathbb{P}\big(X_0 < \varepsilon\, n\big) \underset{n\to\infty}{\longrightarrow} 1$. Donc on a bien $\mathbb{P}\big(|X_n - 0| \geq \varepsilon\big) \underset{n\to\infty}{\longrightarrow} 0$ et ainsi $X_n \overset{\mathcal{P}}{\underset{n\to+\infty}{\longrightarrow}} 0$.
Cependant, on aurait pu faire beaucoup plus simple puisque la convergence en loi vers une constante est équivalent à la convergence en probabilité vers une constante, donc ayant montré $X_n \overset{\mathcal{L}}{\underset{n\to\infty}{\longrightarrow}} 0$ on avait directement $X_n \overset{\mathcal{P}}{\underset{n\to+\infty}{\longrightarrow}} 0$.
- La convergence presque sûre est souvent la plus difficile à démontrer. Dans cet exercice c'est la plus simple! En effet, pour tout $\omega \in \Omega$, $X_n(\omega) = X_0(\omega)/n \underset{n\to\infty}{\longrightarrow} 0$: il y a convergence "sûre" sur Ω donc $X_n \overset{p.s.}{\underset{n\to\infty}{\longrightarrow}} 0$.
Si on avait initialement commencé par cette convergence, on avait directement la convergence en loi et probabilité, puisque la convergence presque sûre entraîne les 2 autres.
- Pour la convergence dans $\mathbb{L}^p$ avec $p > 0$, celle-ci n'est possible que si $X_0 \in \mathbb{L}^p$, donc si $\|X_0\|_p = \big(\mathbb{E}[|X_0|^p]\big)^{1/p} < \infty$. Dans ce cas, $\mathbb{E}[|X_n - 0|^p] = \frac{\mathbb{E}[|X_0|^p]}{n^p} \underset{n\to\infty}{\longrightarrow} 0$, donc $X_n \overset{\mathbb{L}^p}{\underset{n\to\infty}{\longrightarrow}} 0$.
Si $X_0 \notin \mathbb{L}^p$, alors (X_n) ne converge pas dans $\mathbb{L}^p$. Question? Peut-il exister une v.a. X_0 qui n'appartiennent à aucun $\mathbb{L}^p$ pour tout $p > 0$. Ceci est possible par exemple en considérant X_0 une variable positive continue de densité $f(x) = C\,(x\, \ln^2(x))^{-1}$ pour $x \geq 2$. On pourra trouver C car l'intégrale de f existe sur $[2, \infty[$ (intégrale de Bertrand), mais en revanche $\mathbb{E}[X_0^p] = C \int_2^\infty x^{p-1} \ln^{-2}(x)\, dx = \infty$ pour tout $p > 0$.

$\square$

1. **(*)** Soit $(X_n)_n$ une suite i.i.d. de variables de Bernoulli de paramètre $\theta \in]0,1[$.

   (a) Montrer que $X_n = n^{-1}\sum_{i=1}^n X_i \xrightarrow[n\to+\infty]{\mathcal{P}} \theta$. En déduire que $\overline{X}_n(1-\overline{X}_n) \xrightarrow[n\to+\infty]{\mathcal{P}} \theta(1-\theta)$.

   (b) Montrer que $\sqrt{n}(\overline{X}_n - \theta) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0, \theta(1-\theta))$.

   (c) Montrer que $\sqrt{n}(\overline{X}_n - \theta)^2 \xrightarrow[n\to+\infty]{\mathcal{P}} 0$.

   (d) Déterminer la loi limite de $\sqrt{n}\left(\overline{X}_n(1-\overline{X}_n) - \theta(1-\theta)\right)$ pour $\theta \neq 1/2$.

Proof. (a) En appliquant la loi des grands nombres dont les hypothèses sont respectées, comme $\mathbb{E}[X_1] = \theta$, alors $X_n = n^{-1}\sum_{i=1}^n X_i \xrightarrow[n\to+\infty]{\mathcal{P}} \theta$.

Soit la fonction $g(x) = x(1-x)$ pour $x \in \mathbf{R}$, fonction continue sur $\mathbf{R}$: donc $g(\overline{X}_n) \xrightarrow[n\to+\infty]{\mathcal{P}} g(\theta)$, soit le résultat.

(b) On applique le théorème de la limite centrale dont les hypothèses sont respectées et avec $\mathrm{var}(X_1)\theta(1-\theta)$, on obtient bien $\sqrt{n}(\overline{X}_n - \theta) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0, \theta(1-\theta))$.

(c) D'après le résultat précédent, $n^{1/4}(\overline{X}_n - \theta) = n^{-1/4}\big(\sqrt{n}(\overline{X}_n - \theta)\big) \xrightarrow[n\to\infty]{\mathcal{L}} 0$. Comme cette limite en loi est vers une constante, cette limite est aussi en probabilité. On applique alors la fonction carré, qui est continue, et on obtient bien $\sqrt{n}(\overline{X}_n - \theta)^2 \xrightarrow[n\to+\infty]{\mathcal{P}} 0$ en probabilité.

(d) On applique la delta-méthode, avec la fonction $g(x) = x(1-x)$ pour $x \in \mathbf{R}$, et on obtient pour $\theta \neq 1/2$,

$$\sqrt{n}\big(g(\overline{X}_n) - g(\theta)\big) = \sqrt{n}\left(\overline{X}_n(1-\overline{X}_n) - \theta(1-\theta)\right) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}\big(0\,,\,\theta(1-\theta)(1-2\theta)^2\big)$$

car $g'(x) = 1-2x$ et $g'(\theta) = 1-2\theta$.
Pour $\theta = 1/2$, en allant plus loin dans le développement de Taylor

$$n\left(g(\overline{X}_n) - g(\theta)\right) \simeq \frac{1}{2}\left(\sqrt{n}(\overline{X}_n - \theta)\right)^2 g''(\theta) \quad\Longrightarrow\quad n\left(\overline{X}_n(1-\overline{X}_n) - \frac{1}{4}\right) \xrightarrow[n\to\infty]{\mathcal{L}} -\frac{1}{4}\chi^2(1).$$

□

2. **(*)** Soit $(X_k)_{k\in\mathbb{N}}$ une suite de v.a. telle que $\mathbb{E}[X_k^2] < \infty$ pour tout $k \in \mathbf{N}$ et $\mathrm{cov}(X_i, X_j) = 0$ pour $i \neq j$. On suppose qu'il existe des réels m et C tels que pour tout k, $\mathbb{E}[X_k] = m$ et $\mathrm{var}X_k \leq C$. Montrer que la suite des $\overline{X}_n$ converge vers m dans $\mathbb{L}^2$ et en probabilité.

Proof. On a $\mathbb{E}\big[(\overline{X}_n - m)^2\big] = \frac{1}{n^2}\sum_{1\leq i\leq n}\mathrm{var}(X_i) \leq \frac{C}{n} \xrightarrow[n\to\infty]{} 0$. D'où la convergence de $\overline{X}_n$ vers m dans $\mathbb{L}^2$. Et comme la convergence dans $\mathbb{L}^2$ entraîne celle en probabilité, on a aussi $\overline{X}_n \xrightarrow[n\to+\infty]{\mathcal{P}} m$.

□

3. **(**) [Hors Progamme]** Soit $(X_n)_{n\in\mathbb{N}}$ une suite de v.a. indépendantes à valeurs dans $\mathbb{N}$. Montrer que :

   (a) $((X_n)$ converge en probabilité vers $0) \Longleftrightarrow$ (la suite $(\mathbb{P}(X_n > 0))$ tend vers 0).

   (b) $((X_n)$ converge presque-sûrement vers $0) \Longleftrightarrow$ (la série $\displaystyle\sum_{n\geq 1}^{\infty}\mathbb{P}(X_n > 0) < \infty$).

   (c) On suppose que (X_n) suit une loi de Poisson de paramètre α_n. Etudier la convergence de la suite (X_n) dans $\mathbb{L}^1$ puis presque-sûrement dans les cas où $\alpha_n = 1/n$ et $\alpha_n = 1/n^2$.

Proof. (a) On va utiliser le fait que les X_n sont à valeurs dans $\mathbf{N}$ et ainsi pour tout $0 < \varepsilon < 1$, $\mathbb{P}(|Xn - 0| \geq \varepsilon) = \mathbb{P}(X_n > 0)$. D'où le résultat (si $\varepsilon \geq 1$ alors $\mathbb{P}(X_n \geq \varepsilon) \leq \mathbb{P}(X_n > 0)$).

(b) On sait que puisque les v.a. sont indépendantes, alors $(X_n \xrightarrow[n\to\infty]{p.s.} 0) \Longleftrightarrow (\forall\varepsilon > 0, \displaystyle\sum_{n\geq 1}^{\infty}\mathbb{P}(|X_n - 0| \geq \varepsilon) < \infty$).

Mais comme précédemment cela revient à remplacer par $\mathbb{P}(X_n > 0)$.

(c) Comme (X_n) suit une loi de Poisson de paramètre α_n avec $\alpha_n \underset{n\to\infty}{\longrightarrow} 0$ alors comme $\mathbb{E}[X_n] = \alpha_n$, on a aussi $\mathbb{E}[|X_n - 0|] \underset{n\to\infty}{\longrightarrow} 0$, donc la suite (X_n) conerge vers 0 dans $\mathbb{L}^1$.
Si on utilise l'équivalence précédente, on commence par calculer $\mathbb{P}(X_n > 0) = 1 - \mathbb{P}(X_n = 0) = 1 - e^{-\alpha_n} \sim \alpha_n$. On en déduit que si $\alpha_n = 1/n$ alors $\sum_{n\geq 1} \mathbb{P}(X_n > 0) = +\infty$ donc il n'y a pas convergence presque-sûre, mais si $\alpha_n = 1/n^2$ alors $\sum_{n\geq 1} \mathbb{P}(X_n > 0) < \infty$ donc $X_n \underset{n\to\infty}{\overset{p.s.}{\longrightarrow}} 0$.

□

4. **(**) [Hors Progamme]** Soit (X_n) une suite de v.a. positives telle que $X_{n+1} \leq X_n$ p.s. pour tout $n \in \mathbf{N}$. Montrer que $X_n \underset{n\to\infty}{\overset{p.s.}{\longrightarrow}} 0$ si et seulement si $X_n \underset{n\to+\infty}{\overset{\mathcal{P}}{\longrightarrow}} 0$. En déduire que si (Y_n) est une suite de v.a. non nécessairement positives, et si on note $Y_n^* = \sup_{k\geq n} |Y_k|$ alors $Y_n \underset{n\to\infty}{\overset{p.s.}{\longrightarrow}} 0$ si et seulement si $Y_n^* \underset{n\to+\infty}{\overset{\mathcal{P}}{\longrightarrow}} 0$.

Proof. La convergence p.s. implique celle en probabilité. Il suffit donc de démontrer la réciproque. Supposons donc que pour $\varepsilon > 0$ fixé, $\mathbb{P}(X_n > \varepsilon) \to 0$. Posons $A_n := \{X_n \leq \varepsilon\}$. Nous avons donc que $P(A_n^c) \to 0$. Donc il existe une sous-suite $(n_k)_k$ telle que $\sum_{k=1}^{\infty} P(A_{n_k}^c) < \infty$. Ici on utilise : si $x_n \geq 0$ et $x_n \to 0$, alors il existe toujours une sous-suite telle que $\sum_{k=1}^{\infty} x_{n_k} < \infty$. Et on applique avec $x_n = \mathbb{P}(A_n^c)$.
Par le lemme de Borel-Cantelli, nous savons que $\mathbb{P}(\limsup A_{n_k}^c) = 0$ ou encore $\mathbb{P}(\liminf A_{n_k}) = 1$. Pour tout $\omega \in \liminf A_{n_k}$ (qui est un ensemble de proba 1) il existe donc $k_0 = k_0(\omega)$ tel que $A_{n_{k_0}}$ est réalisé, ce qui veut dire que $X_{n_{k_0}}(\omega) \leq \varepsilon$. Puisque la suite est décroissante, on en déduit que $\limsup_{n\to\infty} X_n(\omega) \leq \varepsilon$.b Ou encore

$$\limsup_{n\to\infty} X_n \leq \varepsilon \text{ presque surement.}$$

Puisque ε est arbitraire, cela implique que $\limsup_{n\to\infty} X_n = 0$ p.s. donc le résultat.

(Y_n^*) est une suite de v.a. positive et décroissante. De plus $Y_n^* \underset{n\to\infty}{\overset{p.s.}{\longrightarrow}} 0$ est équivalent à $Y_n \underset{n\to\infty}{\overset{p.s.}{\longrightarrow}} 0$. D'où le résultat.

□

5. **(**)** Soit $\Omega = [0,1]$ et soit la suite (X_n) de v.a. définies sur $(\Omega, \mathcal{B}([0,1]), \mathbb{P})$ où $\mathbb{P}$ est la loi uniforme sur $[0,1]$ et telle que pour tout $n \in \mathbf{N}$:

$$X_n(\omega) = (n+1)^2\,\omega^n - (n+1) \quad \text{pour tout } \omega \in [0,1].$$

Que vaut $\mathbb{E}[X_n]$? Démontrer pourtant que $X_n \underset{n\to+\infty}{\overset{\mathcal{P}}{\longrightarrow}} -\infty$. La suite (X_n) converge-t-elle dans $\mathbb{L}^2$?

Proof. On a $\mathbb{E}[X_n] = \int_0^1 \big((n+1)^2\,\omega^n - (n+1)\big)\,d\omega = \Big[(n+1)\,\omega^{n+1}\Big]_0^1 - (n+1) = 0$.
Pour tout $\omega \in [0,1[$, $(n+1)^2\,\omega^n \underset{n\to\infty}{\longrightarrow} 0$ donc $X_n(\omega) = -\infty$. Ainsi pour tout $\varepsilon > 0$, $\mathbb{P}\big(|X_n + (n+1)| \geq \varepsilon\big) = \mathbb{P}\big(\{1\}\big) = 0$.
On en déduit donc que la variable $X_n + (n+1)$ tend en probabilité vers 0 donc $X_n \underset{n\to+\infty}{\overset{\mathcal{P}}{\longrightarrow}} -\infty$.
Si X_n converge en probabilité vers $-\infty$, alors la seule limite possible dans $\mathbb{L}^2$ pour (X_n) est $-\infty$. Mais pour toute suite réelle (a_n),

$$\begin{aligned}\mathbb{E}\big[(X_n - a_n)^2\big] &= \mathbb{E}[X_n^2] - 2\,a_n\,\mathbb{E}[X_n] + a_n^2 \\ &= (n+1)^2\Big(\frac{(n+1)^2}{2n+1} - 1\Big) + a_n^2 \\ &\underset{n\to\infty}{\longrightarrow} +\infty.\end{aligned}$$

Donc X_n ne tend pas vers $-\infty$ dans $\mathbb{L}^2$ et on s'aperçoit même que c'est pour $a_n = 0$ que la distance $\mathbb{L}^2$ entre X_n et a_n est la plus petite (mais pour autant X_n ne tend pas du tout vers 0 dans $\mathbb{L}^2$).

□

6. **(**)** On suppose $X_n \underset{n\to\infty}{\overset{\mathcal{L}}{\longrightarrow}} c$ pour une suite de v.a. $(X_n)_n$ à valeurs réelles et $c \in \mathbf{R}$. Soit $\phi : \mathbf{R}_+ \to \mathbf{R}$ définie par $\phi(x) = \min(x, 1)$.

(a) Soit $\varepsilon > 0$. En utilisant une caractérisation de la convergence en loi, quelle est la limite de $\mathbb{E}[\phi(|X_n - c|/\varepsilon)]$ quand $n \to \infty$?

(b) En déduire que $X_n \underset{n\to+\infty}{\overset{\mathcal{P}}{\longrightarrow}} c$.

Proof. (a) La fonction ϕ est continue et bornée. Comme $X_n \xrightarrow[n\to\infty]{\mathcal{L}} c$ alors $|X_n - c|/\varepsilon \xrightarrow[n\to\infty]{\mathcal{L}} 0$, alors $\mathbb{E}[\phi(|X_n - c|/\varepsilon)] \xrightarrow[n\to\infty]{} \mathbb{E}[\phi(0)] = 0$.

(b) Pour $\varepsilon > 0$, $\phi(|X_n - c|/\varepsilon) = 1$ pour $|X_n - c| \geq \varepsilon$ et $\phi(|X_n - c|/\varepsilon) = |X_n - c|/\varepsilon$ sinon. On en déduit que

$$\begin{aligned}
\mathbb{E}[\phi(|X_n - c|/\varepsilon)] &= \mathbb{E}\big[\phi(|X_n - c|/\varepsilon)\, \mathbb{I}_{|X_n - c|\geq \varepsilon}\big] + \mathbb{E}\big[\phi(|X_n - c|/\varepsilon)\, \mathbb{I}_{|X_n - c|<\varepsilon}\big] \\
&= \mathbb{E}\big[\mathbb{I}_{|X_n - c|\geq \varepsilon}\big] + \int_{c-\varepsilon}^{c+\varepsilon} \frac{|x-c|}{\varepsilon}\, dF_{X_n}(x) \\
&= \mathbb{P}\big(|X_n - c| \geq \varepsilon\big) + \int_{c-\varepsilon}^{c+\varepsilon} \frac{|x-c|}{\varepsilon}\, dF_{X_n}(x).
\end{aligned}$$

Mais on peut écrire que

$$\begin{aligned}
\int_{c-\varepsilon}^{c+\varepsilon} \frac{|x-c|}{\varepsilon}\, dF_{X_n}(x) &= \frac{1}{\varepsilon} \int_{c}^{c+\varepsilon} (x-c)\, dF_{X_n}(x) + \frac{1}{\varepsilon} \int_{c+\varepsilon}^{c} (c-x)\, dF_{X_n}(x) \\
&= \frac{1}{\varepsilon} \Big(\Big[(x-c)F_{X_n}(x)\Big]_c^{c+\varepsilon} - \int_c^{c+\varepsilon} F_{X_n}(x)\, dx + \Big[(c-x)F_{X_n}(x)\Big]_{c-\varepsilon}^{c} + \int_{c-\varepsilon}^{c} F_{X_n}(x)\, dx \Big) \\
&\xrightarrow[n\to\infty]{} \frac{1}{\varepsilon} \Big(\varepsilon - (c+\varepsilon - c) + 0 + 0 \Big) = 0,
\end{aligned}$$

car pour tout $x > c$, $F_{X_n}(x) \xrightarrow[n\to\infty]{} 1$ et pour tout $x < c$, $F_{X_n}(x) \xrightarrow[n\to\infty]{} 0$. En conséquence $\mathbb{E}[\phi(|X_n - c|/\varepsilon)]$ et $\mathbb{P}\big(|X_n - c| \geq \varepsilon\big)$ on même limite, donc $\mathbb{P}\big(|X_n - c| \geq \varepsilon\big) \xrightarrow[n\to\infty]{} 0$, soit $X_n \xrightarrow[n\to+\infty]{\mathcal{P}} c$.

□

7. **(***)** Soit $(X_n)_{n\in\mathbb{N}}$ une suite de v.a.i.i.d. telle que $\mathbb{E}[X_1^2] < \infty$. Soit $Y_k = \overline{X}_k$ pour $k \geq 1$. Peut-on obtenir la loi faible des grands nombres pour la suite (Y_k)? La loi forte des grands nombres? Dans le cas particulier où les X_k suivent une loi normale centrée réduite déterminer un théorème de limite centrale.

Proof. On va supposer, sans perte de généralité $\mathbb{E}[X_1] = 0$ puisque (Y_n) a la même espérance que (X_n) et on notera $\sigma^2 = \mathrm{var}(X_1)$. On sait que $\mathrm{var}(Y_k) = \frac{\sigma^2}{k}$. Par ailleurs, $\mathrm{cov}(Y_k, Y_\ell) = \frac{\sigma^2}{k}$ pour $k \leq \ell$. On en déduit que:

$$\begin{aligned}
\mathrm{var}\Big(\frac{1}{n}\sum_{k=1}^n Y_k\Big) &= \frac{1}{n^2}\Big(\sum_{k=1}^n \mathrm{var}(Y_k) + 2\sum_{1\leq k<\ell\leq n} \mathrm{cov}(Y_k, Y_\ell)\Big) \\
&= \frac{\sigma^2}{n^2}\Big(1 + 2\sum_{1\leq k<\ell\leq n} \frac{1}{k}\Big) \\
&= \frac{\sigma^2}{n^2}\Big(1 + 2\sum_{1\leq k<n} \frac{n-k}{k}\Big) \\
&\simeq \frac{2\,\sigma^2\, \ln(n)}{n} \xrightarrow[n\to\infty]{} 0.
\end{aligned}$$

En utilisant l'Inégalité de Bienaymé-Tchebychev, on en déduit que $Y_n \xrightarrow[n\to+\infty]{\mathcal{P}} 0$.

Pour une suite numérique (u_n) convergeant vers une limite ℓ, on sait que la somme de Cesaro $\frac{1}{n}\sum_{k=1}^n u_k$ converge vers ℓ. Dans notre cas, considérons $\Omega \setminus N$ l'ensemble des $\omega \in \Omega$ tels que $Y_n(\omega) \xrightarrow[n\to\infty]{} 0$. Et comme $Y_n \xrightarrow[n\to\infty]{p.s.} 0$, on sait que $\mathbb{P}(N) = 0$ (ensemble négligeable). Comme somme de Césaro, on a donc pour tout $\omega \in \Omega \setminus N$, $\frac{1}{n}\sum_{k=1}^n Y_k(\omega) \xrightarrow[n\to\infty]{} 0$, ce qui implique:

$$\frac{1}{n}\sum_{k=1}^n Y_k \xrightarrow[n\to\infty]{p.s.} 0.$$

Si $X_1 \overset{\mathcal{L}}{\sim} \mathcal{N}(0,1)$, alors $(Y_1, \dots, Y_n)$ est un vecteur gaussien et $\Big(\frac{1}{n}\sum_{k=1}^n Y_k\Big)_n$ est une suite de v.a. gaussienne. On en déduit donc du calcul effectué plus haut que

$$\frac{1}{n}\sum_{k=1}^n Y_k \overset{\mathcal{L}}{\sim} \mathcal{N}\Big(0\,,\, \mathrm{var}\Big(\frac{1}{n}\sum_{k=1}^n Y_k\Big)\Big) \implies \sqrt{\frac{n}{2\,\ln(n)}}\,\frac{1}{n}\sum_{k=1}^n Y_k \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1).$$

□

8. **(**)** Appliquer le théorème limite central à une suite de v.a.i.i.d. de loi de Poisson de paramètre 1 pour trouver la limite de la suite

$$u_n = e^{-n}\sum_{k=0}^n \frac{n^k}{k!}.$$

Proof. On sait que si les X_k sont des v.a.i.i.d. de loi de Poisson de paramètre 1, alors $\sum_{k=1}^n X_k$ suit une loi de Poisson de paramètre n. Donc

$$u_n = e^{-n}\sum_{k=0}^n \frac{n^k}{k!} = \mathbb{P}\Big(\sum_{k=1}^n X_k \le n\Big) = \mathbb{P}\Big(\frac{1}{n}\sum_{k=1}^n X_k - 1 \le 0\Big).$$

Mais si on applique le TLC à (X_k) ce qui est possible puisque les X_k sont des v.a.i.i.d. de variance 1, alors:

$$\sqrt{n}\Big(\frac{1}{n}\sum_{k=1}^n X_k - 1\Big) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1).$$

De ceci on en déduit que $u_n \underset{n\to\infty}{\longrightarrow} \mathbb{P}\big(\mathcal{N}(0,1)\le 0\big) = 1/2$.

□

9. **(***)** Soit $(X_n)_n$ une suite de v.a.i.i.d. centrées de variance commune 1. Soit $(a_{i,n})_{1\le i\le n,\, n\in\mathbf{N}}$ une famille de réels telle que $\sum_{i=1}^n a_{i,n}^2 = 1$ pour tout $n\in\mathbf{N}^*$. On va montrer que si $\max_{1\le i\le n}|a_{i,n}| \underset{n\to\infty}{\longrightarrow} 0$, alors la suite des (S_n) telle que $S_n = \sum_{i=1}^n a_{i,n}X_i$ converge en loi vers la loi $\mathcal{N}(0,1)$.

   (a) Montrer que si (z_j) et (z'_j) sont deux familles de nombres complexes tels que $|z_j|\le 1$ et $|z'_j|\le 1$ pour tout j, alors
   $$\Big|\prod_{j=1}^n z_j - \prod_{j=1}^n z'_j\Big| \le \sum_{j=1}^n |z_j - z'_j|.$$

   (b) Déterminer la fonction caractéristique de S_n. En déduire sa limite en utilisant l'inégalité précédente.

Proof. (a) Par récurrence sur n, montrons que

$$\Big|\prod_{j=1}^n z_j - \prod_{j=1}^n z'_j\Big| \le \sum_{j=1}^n |z_j - z'_j|.$$

En effet, la relation est clairement vraie pour $n=1$. Si elle est vraie au rang n alors par inégalité triangulaire,

$$\begin{aligned}
\Big|\prod_{j=1}^{n+1} z_j - \prod_{j=1}^{n+1} z'_j\Big| &= \Big|z_{n+1}\prod_{j=1}^n z_j - z'_{n+1}\prod_{j=1}^n z'_j\Big| \\
&\le \Big|(z_{n+1}-z'_{n+1})\prod_{j=1}^n z_j\Big| + \Big|z'_{n+1}\Big(\prod_{j=1}^n z_j - \prod_{j=1}^n z'_j\Big)\Big| \\
&\le \big|z_{n+1}-z'_{n+1}\big| + \Big|\prod_{j=1}^n z_j - \prod_{j=1}^n z'_j\Big| \le \sum_{j=1}^{n+1}|z_j - z'_j|.
\end{aligned}$$

La relation est donc vraie au rang $n+1$.

(b) Considérons la fonction caractéristique de S_n et montrons qu'elle converge vers celle d'une loi $\mathcal{N}(0,1)$.
Soit $u\in\mathbf{R}$. Alors $\phi_{S_n}(u) = \mathbb{E}\big[e^{i\,u\,S_n}\big] = \prod_{k=1}^n \mathbb{E}\big[e^{i\,u\,a_{k,n}X_k}\big]$ en utilisant l'indépendance des X_k. Comme par hypothèse $\max_{1\le i\le n}|a_{i,n}| \underset{n\to\infty}{\longrightarrow} 0$, alors pour tout $k\in\{1,\dots,n\}$, on peut utiliser un développement de Taylor d'ordre 2 en 0 de chaque fonction caractéristique $\phi_{X_k}(u\,a_{k,n}) = \mathbb{E}\big[e^{i\,u\,a_{k,n}X_k}\big]$ et on obtient avec les $\varepsilon_{k,n}$ tels que $\max_{1\le k\le n}|\varepsilon_{k,n}| \underset{n\to\infty}{\longrightarrow} 0$,

$$\phi_{X_k}(u\,a_{k,n}) = \phi_{X_k}(0) + \phi'_{X_k}(0)\,u\,a_{k,n} + \frac{1}{2}\phi''_{X_k}(0)\,u^2\,a_{k,n}^2 + o\big(u^2\,a_{k,n}^2\big) = 1 - \frac{1}{2}u^2\,a_{k,n}^2 + \varepsilon_{k,n}\,\big(u^2\,a_{k,n}^2\big),$$

car $\phi'_{X_k}(0) = i\,\mathbb{E}[X_k] = 0$ et $\phi''_{X_k}(0) = -\mathbb{E}[X_k^2] = -1$. En utilisant l'inégalité, comme pour n suffisamment grand $\Big|1-\frac{1}{2}u^2\,a_{k,n}^2\Big| \le 1$ pour tout $k\in\{1,\dots,n\}$ du fait que $\max_{1\le k\le n}|a_{k,n}| \underset{n\to\infty}{\longrightarrow} 0$, ceci entraîne que

$$\begin{aligned}
\Big|\phi_{S_n}(u) - \prod_{k=1}^n\Big(1-\frac{1}{2}u^2\,a_{k,n}^2\Big)\Big| &\le \sum_{k=1}^n |\varepsilon_{k,n}|\,u^2\,a_{k,n}^2 \\
&\le u^2 \max_{1\le k\le n}|\varepsilon_{k,n}| \sum_{k=1}^n a_{k,n}^2 \\
&\le u^2 \max_{1\le k\le n}|\varepsilon_{k,n}| \underset{n\to\infty}{\longrightarrow} 0.
\end{aligned}$$

Pour n suffisamment grand, on a avec le développement de Taylor de la fonction log et $\max_{1\le k\le n}|\varepsilon'_{k,n}| \underset{n\to\infty}{\longrightarrow} 0$,

$$\log\Big(\prod_{k=1}^n\Big(1-\frac{1}{2}u^2\,a_{k,n}^2\Big)\Big) = \sum_{k=1}^n \log\Big(1-\frac{1}{2}u^2\,a_{k,n}^2\Big) = \sum_{k=1}^n\Big(-\frac{1}{2}u^2\,a_{k,n}^2(1+\varepsilon'_{k,n})\Big) \underset{n\to\infty}{\longrightarrow} -\frac{1}{2}u^2.$$

De ceci on en déduit que $\prod_{k=1}^{n}\left(1-\frac{1}{2}\,u^2\,a_{k,n}^2\right) \underset{n\to\infty}{\longrightarrow} e^{-u^2/2}$ et de même $\phi_{S_n}(u) \underset{n\to\infty}{\longrightarrow} e^{-u^2/2}$ qui est la fonction caractéristique d'une loi $\mathcal{N}(0,1)$.

□

10. **(**)** Soit $(X_n)_n$ une suite de v.a.i.i.d. centrées de variance commune $\sigma^2 > 0$.

(a) Rappeler la limite en loi de S_n telle que $S_n = \dfrac{1}{\sigma\sqrt{n}}\displaystyle\sum_{j=1}^{n} X_j$.

(b) Décomposer la variable S_{2n} en fonction de S_n et d'une variable aléatoire S'_n indépendante de S_n et de même loi.

(c) En raisonnant par l'absurde, montrer que S_n ne converge pas en probabilité (on pourra montrer que si c'était le cas, S'_n convergerait aussi en probabilité et étudier sa limite).

Proof. (a) Le TLC peut pleinement s'appliquer et on a $S_n = \dfrac{1}{\sigma\sqrt{n}}\displaystyle\sum_{j=1}^{n} X_j = \sqrt{n}\,\frac{1}{\sigma}\left(\overline{X}_n - 0big\right) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1)$.

(b) On a $\frac{1}{\sigma}\sum_{j=1}^{n} X_j = \sqrt{n}\,S_n$, donc $\frac{1}{\sigma}\sum_{j=1}^{2n} X_j = \sqrt{n}\,S_n + \frac{1}{\sigma}\sum_{j=n+1}^{2n} X_j = \sqrt{2n}\,S_{2n}$. On en déduit donc que $S_{2n} = \frac{1}{\sqrt{2}}\,S_n + S'_n$ avec $S'_n = \frac{1}{\sigma\sqrt{2n}}\sum_{j=n+1}^{2n} X_j$. Il est clair que comme les X_i sont indépendantes, alors $\frac{1}{\sqrt{2}}\,S_n$ et S'_n sont indépendantes. Et comme les X_i ont toutes même loi, $\sum_{j=1}^{n} X_j \overset{\mathcal{L}}{\sim} \sum_{j=n+1}^{2n} X_j$, donc $\frac{1}{\sqrt{2}}\,S_n \overset{\mathcal{L}}{\sim} S'_n$.

(c) Comme $S_n \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1)$, il est clair que si S_n converge en probabilité, ce ne peut être que vers une variable aléatoire S_∞ qui suit une loi $\mathcal{N}(0,1)$. Supposons que $S_n \xrightarrow[n\to+\infty]{\mathcal{P}} S_\infty$. Alors $\frac{1}{\sqrt{2}}\,S_n \xrightarrow[n\to+\infty]{\mathcal{P}} \frac{1}{\sqrt{2}}\,S_\infty$ et $S_{2n} \xrightarrow[n\to+\infty]{\mathcal{P}} S_\infty$. Or $S'_n = \frac{1}{\sqrt{2}}\,S_n - S_{2n}$ donc $S'_n \xrightarrow[n\to+\infty]{\mathcal{P}} \left(\frac{1}{\sqrt{2}}-1\right)S_\infty$. Mais comme on a supposé que S_n convergeait en probabilités, on a aussi $S'_n \xrightarrow[n\to+\infty]{\mathcal{P}} \frac{1}{\sqrt{2}}\,S'_\infty$ où $S'_\infty \overset{\mathcal{L}}{\sim} S_\infty$ et S'_∞ indépendante de S'_∞. Donc à moins que $S_\infty = 0$, ce qui n'est pas possible puisque la loi de S_∞ est $\mathcal{N}(0,1)$, on ne peut pas avoir $(\frac{1}{\sqrt{2}}-1\Big)\,S_\infty = \frac{1}{\sqrt{2}}\,S'_\infty$, d'où l'impossibilité de converger en probabilité.

□

Feuille n^o 5:
Estimation paramétrique

1. **(*)** Soit un n-échantillon $(X_1,\dots,X_n)$ de v.a.i.i.d. suivant une loi de Poisson $\mathcal{P}(\theta)$, où $\theta>0$ est inconnu. Déterminer un estimateur du maximum de vraisemblance de θ. Est-il biaisé? Convergent? Efficace?

Proof. Les variables sont à valeurs dans $\mathbf{N}$. La vraisemblance s'écrit:

$$L_{(X_1,\dots,X_n)}(x_1,\dots,x_n)=\mathbb{P}\big(X_1=x_1\cap\dots\cap X_n=x_n\big)=\prod_{k=1}^n e^{-\lambda}\,\frac{\lambda^{x_k}}{x_k!},$$

en utilisant l'indépendance. On en déduit que la log-vraisemblance en $X_1,\dots,X_n$ vaut:

$$\ell_n(\lambda)=-n\,\lambda+\log(\lambda)\sum_{k=1}^n X_k-\sum_{k=1}^n\log(X_k!).$$

Pour majorer, on peut dériver et on arrive à un point critique vérifiant $-n+\frac{1}{\lambda}\sum_{k=1}^n X_k$, soit $\lambda=\frac{1}{n}\sum_{k=1}^n X_k$. Si on dérive 2 fois on obtient $-\frac{1}{\lambda^2}\sum_{k=1}^n X_k$, donc une fonction toujours négative. Ainsi

$$\widehat{\lambda}=\frac{1}{n}\sum_{k=1}^n X_k.$$

Il est clair que $\mathbb{E}[\widehat{\lambda}]=\mathbb{E}[X_1]=\lambda$: l'estimateur est sans biais.
Commençons par calculer le risque quadratique:

$$R(\widehat{\lambda})=\mathrm{var}(\widehat{\lambda})=\frac{1}{n}\,\mathrm{var}(X_1)=\frac{\lambda}{n}.$$

Le modèle est régulier. On peut calculer l'information de Fisher pour une variable X_1. Pour cela on calcule $\frac{\partial}{\partial\lambda}\log\Big(e^{-\lambda}\,\frac{\lambda^{X_1}}{X_1!}\Big)=\frac{\partial}{\partial\lambda}\big(-\lambda+X_1\,\log(\lambda)-\log(X_1!)\big)=-1+\frac{X_1}{\lambda}$ et on en déduit $\frac{\partial^2}{\partial^2\lambda}\log\Big(e^{-\lambda}\,\frac{\lambda^{X_1}}{X_1!}\Big)=-\frac{X_1}{\lambda^2}$. D'où l'information de Fisher:

$$I(\lambda)=-\mathbb{E}\Big[-\frac{X_1}{\lambda^2}\Big]=\frac{1}{\lambda}.$$

Donc on retrouve $I^{-1}(\lambda)=\lambda=\mathrm{var}(X_1)$: l'estimateur est efficace. □

2. **(**)** Soit un n-échantillon $(X_1,\dots,X_n)$ de v.a.i.i.d. suivant une loi de densité $f_\theta(x)=(\theta+1)\,x^\theta\,\mathbb{I}_{0<x\le 1}$ par rapport à la mesure de Lebesgue sur $\mathbf{R}$, où $\theta>0$ est inconnu. Déterminer l'estimateur du maximum de vraisemblance $\widehat{\theta}$ de θ. Après avoir calculé $\mathbb{E}[\log(X_1)]$, montrer que $\widehat{\theta}$ est convergent.

Proof. Les valeurs prises par X_1 sont dans $]0,1]$. La vraisemblance de $(X_1,\dots,X_n)$ revient au calcul de la densité par rapport à la mesure de Lebesgue sur $]0,1]^n$ soit:

$$f_{(X_1,\dots,X_n)}(x_1,\dots,x_n)=\prod_{k=1}^n(\theta+1)\,x_k^\theta=(\theta+1)^n\,\Big(\prod_{k=1}^n x_k\Big)^\theta$$

du fait que les X_i sont indépendants. On en déduit que la log-vraisemblance calculée en $(X_1,\dots,X_n)$ vaut:

$$\ell_n(\theta)=n\,\ln(\theta+1)+\theta\sum_{k=1}^n\ln(X_k).$$

On résout l'équation $\frac{\partial}{\partial\theta}\ell_n(\theta)=0$ et on obtient $\theta=-1-n\,\Big(\sum_{k=1}^n\ln(X_k)\Big)^{-1}$. Il suffit de vérifier $\frac{\partial^2}{\partial\theta^2}\ell_n(\theta)<0$ ce qui est le cas puisque $\frac{\partial^2}{\partial\theta^2}\ell_n(\theta)=-n\,(\theta+1)^2$. Ainsi la fonction est concave et le maximum est bien atteint en

$$\widehat{\theta}=-1-n\,\Big(\sum_{k=1}^n\ln(X_k)\Big)^{-1}.$$

On veut appliquer la loi forte des grands nombres aux $(\ln(1-X_i))_i$, dont les hypothèses sont vérifiées si $\mathbb{E}[|\ln(1-X_1)|]<\infty$. Tout d'abord $\ln(1-X_1)=-|\ln(1-X_1)|$ et donc $\mathbb{E}[|\ln(1-X_1)|]$ existe si $\mathbb{E}[\ln(1-X_1)]>-\infty$. Mais

$\mathbb{E}[\ln(1-X_1)] = (1+\theta)\int_0^1 \ln(1-x)\,(1-x)^\theta\,dx$. Après le changement de variable $y = 1-x$, on obtient par intégration par parties,

$$\mathbb{E}[\ln(1-X)] = (1+\theta)\int_0^1 \ln(y)\,y^\theta\,dy = \Big[y^{\theta+1}\,\ln(y)\Big]_0^1 - \int_0^1 y^\theta dy = 0 - \frac{1}{1+\theta}.$$

Par conséquent, $\displaystyle\frac{1}{n}\sum_{k=1}^n \ln(X_k) \xrightarrow[n\to\infty]{p.s.} \mathbb{E}[\ln(X_1)] = -\frac{1}{1+\theta}$. De ceci, en considérant la fonction $g(x) = -1 - 1/x$ qui est une fonction continue sur $]-\infty, 0[$, on peut écrire que $\widehat{\theta} = g(\overline{X}_n) \xrightarrow[n\to\infty]{p.s.} g(\mathbb{E}[\ln(X_1)]) = \theta$: l'estimateur est bien convergent. □

3. **(*)** On considère un n-échantillon $(X_1, \dots, X_n)$ dont la loi dépend d'un paramètre $\theta \in \mathbf{R}$ inconnu. Soit $\widehat{\theta}_1$ et $\widehat{\theta}_2$ deux estimateurs non biaisés de θ, tels que $\mathbb{E}[\widehat{\theta}_1^2 + \widehat{\theta}_2^2] < \infty$. Pour $\alpha \in [0,1]$ on considère $\widehat{\theta} = \alpha\,\widehat{\theta}_1 + (1-\alpha)\,\widehat{\theta}_2$ et on notera $R(\widehat{\theta}_1)$, $R(\widehat{\theta}_2)$ et $R(\widehat{\theta})$ les risques quadratiques de $\widehat{\theta}_1$, $\widehat{\theta}_2$ et $\widehat{\theta}$.
   (a) On suppose que $\widehat{\theta}_1$ et $\widehat{\theta}_2$ sont indépendants. Déterminer α en fonction de $R(\widehat{\theta}_1)$ et $R(\widehat{\theta}_2)$, de telle manière que le $R(\widehat{\theta})$ soit minimum. Que vaut alors $R(\widehat{\theta})$?
   (b) Si $\widehat{\theta}_1$ et $\widehat{\theta}_2$ ne sont pas indépendants, montrer que l'on a tout de même
   $$R(\widehat{\theta}) \le 2\,\frac{R(\widehat{\theta}_1)\,R(\widehat{\theta}_2)}{R(\widehat{\theta}_1) + R(\widehat{\theta}_2)}.$$

Proof. (a) On a $R(\widehat{\theta}) = \mathbb{E}\big[\big(\alpha\,(\widehat{\theta}_1-\theta) + (1-\alpha)\,(\widehat{\theta}_2-\theta)\big)^2\big] = \alpha^2\,\mathrm{var}(\widehat{\theta}_1) + (1-\alpha)^2\,\mathrm{var}(\widehat{\theta}_2) = \alpha^2\,R(\widehat{\theta}_1) + (1-\alpha)^2\,R(\widehat{\theta}_2)$ car les estimateurs sont tous les 3 sans biais. Il suffit donc de chercher α de telle manière que cette expression soit minimale. On dérive par rapport à α et on obtient que le minimum est obtenu pour $\alpha = R(\widehat{\theta}_2)\,\big(R(\widehat{\theta}_1) + R(\widehat{\theta}_2)\big)^{-1}$. En remplaçant, on obtient que
$$R(\widehat{\theta}) = \frac{R(\widehat{\theta}_1)\,R(\widehat{\theta}_2)}{R(\widehat{\theta}_1) + R(\widehat{\theta}_2)}.$$

(b) Si $\widehat{\theta}_1$ et $\widehat{\theta}_2$ ne sont pas indépendants, on utilise l'inégalité $(a+b)^2 \le 2\,(a^2+b^2)$. En conséquence,
$$\mathbb{E}\big[\big(\alpha\,(\widehat{\theta}_1-\theta) + (1-\alpha)\,(\widehat{\theta}_2-\theta)\big)^2\big] \le 2\,\big(\mathbb{E}\big[\alpha^2(\widehat{\theta}_1-\theta)^2\big] + \mathbb{E}\big[(1-\alpha)^2(\widehat{\theta}_2-\theta)^2\big]\big) = 2\big(\alpha^2 R(\widehat{\theta}_1) + (1-\alpha)^2 R(\widehat{\theta}_2)\big).$$
Il suffit alors de reprendre la valeur de α minimisant l'expression précédente et on obtient le résultat demandé. □

4. **(**)** On considère un n-échantillon $(X_1, \dots, X_n)$ de v.a.i.i.d. suivant une loi $\mathcal{E}(\lambda)$, où $\lambda > 0$ est inconnu.
   (a) Déterminer l'estimateur $\widehat{\lambda}$ par maximum de vraisemblance de λ.
   (b) Déterminer la loi de $\sum_{j=1}^n X_j$. En déduire que $\widehat{\lambda}$ est biaisé.
   (c) L'estimateur $\widehat{\lambda}$ est-il convergent?
   (d) Etablir un TLC vérifié par $1/\widehat{\lambda}$. En déduire un intervalle de confiance asymptotique de niveau 95% pour λ.
   (e) En utilisant la Delta-Méthode déterminer un TLC vérifié par $\widehat{\lambda}$. En déduire que $\widehat{\lambda}$ est asymptotiquement efficace.

Proof. (a) L'ensemble des valeurs prises par X_1 est $[0,\infty[$. La vraisemblance de $(X_1, \dots, X_n)$ est donc la densité par rapport à la mesure de Lebesgue sur $[0,\infty[^n$ et elle vaut pour $(x_1, \dots, x_n) \in [0,\infty[^n$,
$$f_{(X_1,\dots,X_n)}(x_1, \dots, x_n) = \prod_{k=1}^n \lambda\,e^{-\lambda\,x_k} = \lambda^n\,e^{-\lambda\,\sum_{k=1}^n x_k}$$
du fait que les X_i sont indépendants. On en déduit que la log-vraisemblance calculée en $(X_1, \dots, X_n)$ vaut:
$$\ell_n(\lambda) = n\,\ln(\lambda) - \lambda\sum_{k=1}^n X_k.$$
On résout l'équation $\frac{\partial}{\partial\lambda}\ell_n(\lambda) = 0$ et on obtient $\lambda = n\Big(\displaystyle\sum_{k=1}^n X_k\Big)^{-1} = \big(\overline{X}_n\big)^{-1}$. Il suffit de vérifier $\frac{\partial^2}{\partial\lambda^2}\ell_n(\lambda) < 0$ ce qui est le cas puisque $\frac{\partial^2}{\partial\lambda^2}\ell_n(\theta) = -n\,\lambda^2$. Ainsi la fonction est concave et le maximum est bien atteint en
$$\widehat{\lambda} = \big(\overline{X}_n\big)^{-1}.$$

(b) On sait que $X_1 \overset{\mathcal{L}}{\sim} \Gamma(1,\lambda)$ et comme les X_i sont indépendantes, alors $\sum_{j=1}^n X_j \overset{\mathcal{L}}{\sim} \Gamma(n,\lambda)$.
On en déduit que

$$\mathbb{E}[\widehat{\lambda}] = n\int_0^\infty \frac{1}{x}\, f_{\Gamma(n,\lambda)}(x)\, dx = n\int_0^\infty \lambda^{n-1}\,\frac{x^{n-2}}{(n-1)!}\, e^{-\lambda\, x} dx = \frac{n}{n-1}\,\lambda \int_0^\infty f_{\Gamma(n-1,\lambda)}(x)\, dx = \frac{n}{n-1}\,\lambda.$$

L'estimateur $\widehat{\lambda}$ est donc biaisé.

(c) Par la loi des grands nombres dont les hypothèses sont bien vérifiées car les (X_i) sont des v.a.i.i.i.d. et $\mathbb{E}[|X_1|] < \infty$, alors $\overline{X}_n \xrightarrow[n\to\infty]{p.s.} \mathbb{E}[X_1] = 1/\lambda$. On peut alors définir $g(x) = 1/x$ qui est une fonction continue sur $]0,\infty[$, d'où $g(\overline{X}_n) \xrightarrow[n\to\infty]{p.s.} g(\mathbb{E}[X_1]) = \lambda$: l'estimateur est bien convergent.

(d) Comme $\mathbb{E}[X_1^2] < \infty$, les (X_i) vérifient bien un TLC et on a

$$\sqrt{n}\,\frac{1}{\lambda}\,\big(\overline{X}_n - 1/\lambda\big) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1) \Longrightarrow \sqrt{n}\,\Big(\frac{\overline{X}_n}{\lambda} - 1\Big) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1).$$

Pour déterminer un intervalle de confiance asymptotique on peut d'abord écrire avec $q \simeq 1.96$ le quantile à 97.5% d'une loi normale centrée réduite,

$$\mathbb{P}\Big(\Big|\sqrt{n}\,\Big(\frac{\overline{X}_n}{\lambda} - 1\Big)\Big| \leq q\Big) = 0.95 \quad\Longrightarrow\quad \mathbb{P}\Big(\frac{\overline{X}_n}{1+q/\sqrt{n}} \leq \lambda \leq \frac{\overline{X}_n}{1-q/\sqrt{n}}\Big) = 0.95.$$

On en déduit donc que $\Big[\frac{\overline{X}_n}{1+q/\sqrt{n}}\,,\,\frac{\overline{X}_n}{1-q/\sqrt{n}}\Big]$ est un intervalle asymptotique de λ de niveau 95%

(e) On applique la Delta-Méthode au TLC obtenu précédemment, avec la fonction g dont la dérivée $g'(x) - 1/x^2$ ne s'annule pas. Et $g'(1/\lambda) = -\lambda^2$. On en déduit donc que:

$$\sqrt{n}\,\big(\widehat{\lambda} - \lambda\big) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}\big(0\,,\,\lambda^2\big).$$

Le modèle étant régulier et les variables indépendantes, il suffit de calculer l'information de Fisher de X_1. Comme $\frac{\partial}{\partial\theta}\ell_1(\theta) = \frac{1}{\lambda} - X_1$ et $-I_1(\lambda) = \frac{\partial^2}{\partial\theta^2}\ell_1(\theta) = -\frac{1}{\lambda^2}$. Donc $I_1(\lambda)^{-1} = \lambda^2$, qui est bien la variance asymptotique du TLC précédent: l'estimateur est asymptotiquement efficace.

□

5. **(**)** On considère un n-échantillon $(X_1,\dots,X_n)$ de v.a.i.i.d. suivant la loi binomiale $\mathcal{B}(k,p)$, où $p \in]0,1[$ est inconnu alors que $k \in \mathbb{N}^*$ est connu. On voudrait estimer la probabilité $\theta = \mathbb{P}(X_1 = 1)$.

(a) Déterminer l'estimateur $\widehat{p}$ par maximum de vraisemblance de p. Est-il biaisé? Efficace? Etablir un TLC vérifié par $\widehat{p}$.

(b) En déduire un estimateur $\widehat{\theta}$ de θ. Est-il convergent? Etablir un TLC vérifié par $\widehat{\theta}$ pour $p \neq 1/k$. En utilisant le lemme de Slutsky, déterminer un intervalle de confiance asymptotique de niveau 95% pour θ.

(c) Soit $\widetilde{\theta} = \frac{1}{n}\sum_{j=1}^n \mathbb{I}_{X_j=1}$. Montrer que $\widetilde{\theta}$ est non biaisé. Etablir un TLC vérifié par $\widetilde{\theta}$. Pour p proche de 0, de 1 et de $1/k$, quel estimateur préférer entre $\widehat{\theta}$ et $\widetilde{\theta}$?

Proof. (a) Les valeurs prises par X_1 sont dans $\{0,1,\dots,k\}$. La vraisemblance de $(X_1,\dots,X_n)$ est donc pour $(x_1,\dots,x_n) \in \{0,1,\dots,k\}^n$

$$\mathbb{P}\big(X_1 = x_1,\dots,X_n = x_n\big) = \prod_{j=1}^n \tbinom{k}{x_j}\, p^{x_j}\,(1-p)^{k-x_j} = \Big(\prod_{j=1}^n \tbinom{k}{x_j}\Big)\, p^{\sum_{j=1}^n x_j}\,(1-p)^{nk-\sum_{j=1}^n x_j}$$

grâce à l'indépendance des X_i. De ceci on en en déduit que la log-vraisemblance du modèle en $(X_1,\dots,X_n)$ vaut:

$$\ell_n(p) = \ln\Big(\prod_{j=1}^n \tbinom{k}{X_j}\Big) + \ln(p)\sum_{j=1}^n X_j + \ln(1-p)\,\Big(nk - \sum_{j=1}^n X_j\Big).$$

On résout l'équation $\frac{\partial}{\partial p}\ell_n(p) = 0$ et on obtient $(1-p)\displaystyle\sum_{j=1}^n X_j = p\Big(nk - \sum_{j=1}^n X_j\Big)$ soit $p = \frac{nk}{}n\sum_{j=1}^n X_j = \frac{1}{k}\,\overline{X}_n$.
Il suffit de vérifier $\frac{\partial^2}{\partial p^2}\ell_n(p) < 0$ ce qui est le cas puisque $\frac{\partial^2}{\partial p^2}\ell_n(p) = -\frac{1}{p^2}\sum_{j=1}^n X_j - \frac{1}{(1-p)^2}\Big(nk - \sum_{j=1}^n X_j\Big)$.
Ainsi la fonction est concave et le maximum est bien atteint en

$$\widehat{p} = \frac{1}{k}\,\overline{X}_n.$$

We have $\mathbb{E}[\widehat{p}] = \frac{1}{k}\,\mathbb{E}[X_1] = \frac{1}{k}\,k\,pp$: l'estimateur n'est pas biaisé.
Le risque quadratique de l'estimateur vaut $\mathrm{var}(\widehat{p}) = \frac{1}{n\,k^2}\,var(X_1) = \frac{p(1-p)}{n\,k}$.
Par ailleurs, l'information de Fisher $I_1(p)$ pour X_1 est obtenue à partir de

$$I_1(p) = -\mathbb{E}\Big[\frac{\partial^2}{\partial p^2}\ell_1(p)\Big] = -\mathbb{E}\Big[-\frac{X_1}{p^2} - \frac{k-X_1}{(1-p)^2}\Big] = \frac{k}{p} + \frac{k}{1-p} = \frac{k}{p(1-p)}.$$

On en déduit donc que $I_1^{-1}(p) = \frac{k}{p(1-p)} = n\,\mathrm{var}(\widehat{p})$: l'estimateur est efficace.
Les hypothèses du TLC sont bien vérifiée pour les (X_i) car $\mathrm{var}(X_1) < \infty$. On obtient donc:

$$\sqrt{n}\,\big(k\,\widehat{p} - k\,p\big) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}\big(0\,,\,k\,p\,(1-p)\big) \quad\Longrightarrow\quad \sqrt{n}\,\big(\widehat{p} - p\big) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}\big(0\,,\,\frac{1}{k}\,p\,(1-p)\big).$$

(b) On sait que $\theta = k\,p\,(1-p)^{k-1}$. On considère g tel que $g(x) = k\,x\,(1-x)^{k-1}$. On considérera donc $\widehat{\theta} = g(\widehat{p})$.
Comme $\widehat{p} \xrightarrow[n\to\infty]{p.s.} p$ par la loi forte des grands nombres, comme g est continue sur $\mathbf{R}$, alors $\widehat{\theta} = g(\widehat{p}) \xrightarrow[n\to\infty]{p.s.} g(p) = \theta$: l'estimateur est convergent.
On peut également utiliser la Delta-Méthode pour obtenir un TLC sur $\widehat{\theta}$ à partir du TLC sur $\widehat{p}$. En effet, $g'(x) = g(x) = k\,(1-x)^{k-2}(1-k\,x)$ et $g'(p) \neq 0$ pour $k\,p \neq 1$. Ainsi:

$$\sqrt{n}\,\big(g(\widehat{p}) - g(p)\big) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}\big(0\,,\,(g'(p))^2\,\frac{1}{k}\,p\,(1-p)\big) \Longrightarrow \sqrt{n}\,\big(\widehat{\theta} - \theta\big) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}\big(0\,,\,k\,p\,(1-p)^{2k-3}(1-k\,p)^2\big).$$

Comme $\widehat{p} \xrightarrow[n\to\infty]{p.s.} p$ et la fonction $p \to k\,p\,(1-p)^{2k-3}(1-k\,p)^2$, on en déduit que $k\,\widehat{p}\,(1-\widehat{p})^{2k-3}(1-k\,\widehat{p})^2 \xrightarrow[n\to\infty]{p.s.} k\,p\,(1-p)^{2k-3}(1-k\,p)^2$. En utilisant le Lemme de Stlusky, on en déduit le TLC:

$$\sqrt{n}\,\frac{1}{\sqrt{k\,\widehat{p}\,(1-\widehat{p})^{2k-3}(1-k\,\widehat{p})^2}}\,\big(\widehat{\theta} - \theta\big) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}\big(0\,,\,1\big).$$

Avec $q \simeq 1.96$ le quantile à 97.5% d'une loi normale centrée réduite, on en déduit l'intervalle de confiance asymptotique à 95% de θ:

$$\Big[\widehat{\theta} - q\,\frac{\sqrt{k\,\widehat{p}\,(1-\widehat{p})^{2k-3}(1-k\,\widehat{p})^2}}{\sqrt{n}}\,,\,\widehat{\theta} + q\,\frac{\sqrt{k\,\widehat{p}\,(1-\widehat{p})^{2k-3}(1-k\,\widehat{p})^2}}{\sqrt{n}}\Big].$$

(c) On a $\mathbb{E}[\widetilde{\theta}] = \mathbb{E}[\mathbb{I}_{X_j=1}] = \mathbb{P}(X_1 = 1) = \theta$: l'estimateur n'est pas biaisé.
La suite $(\mathbb{I}_{X_j=1})_j$ est une suite de v.a.i.i.d. et $\mathrm{var}(\mathbb{I}_{X_1=1}) = \theta(1-\theta) < \infty$, donc on peut appliquer le TLC. On obtient:

$$\sqrt{n}\,\big(\widetilde{\theta} - \theta\big) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}\big(0\,,\,\theta(1-\theta)\big).$$

Si on écrit le ratio entre les 2 variances asymptotiques, avec $\theta = k\,p\,(1-p)^{k-1}$, on obtient:

$$\frac{k\,p\,(1-p)^{2k-3}(1-k\,p)^2}{\theta(1-\theta)} = (1-p)^{k-2}\,\frac{(1-k\,p)^2}{1-k\,p\,(1-p)^{k-1}}$$

Si $p \to 0$, ce ratio est équivalent à $(1-(2k-2)p$, donc < 1 et c'est $\widehat{\theta}$ qui est le plus intéressant. Si $p \to 1$, ce ratio est équivalent à $(k-1)^2(1-p)^{k-2}$, donc < 1 et c'est toujours $\widehat{\theta}$ qui est le plus intéressant. De même pour $p \to 1/p$. En revanche, pour $p = 1/2$, on peut montrer que pour $k < 11$, $\widetilde{\theta}$ est le plus intéressant, mais pour $k \geq 11$ cela devient $\widehat{\theta}$.

□

6. **(**)** On considère un n-échantillon $(X_1, \ldots, X_n)$ de v.a.i.i.d. suivant la loi binomiale $\mathcal{B}(k,p)$, où $p \in]0,1[$ et $k \in \mathbb{N}^*$ sont inconnus. Obtenir explicitement l'estimateur par maximum de vraisemblance de (p,k) est très difficile, dans cet exercice on essaie autre chose.

   (a) Rappeler l'espérance m et la variance σ^2 de X_1.

   (b) Déterminer des estimateurs naturels de m et σ^2. Sont-ils convergents?

   (c) En déduire des estimateurs convergents de p et de k.

Proof. (a) On a $m = \mathbb{E}[X_1] = k\,p$ et $\sigma^2 = \mathrm{var}(X_1) = k\,p(1-p)$.

(b) On sait que $\overline{X}_n = \frac{1}{n}\sum_{j=1}^n X_j$ et $\overline{\sigma}^2 = \frac{1}{n}\sum_{j=1}^n \big(X_j - \overline{X}_n\big)^2$, moyenne et variance empiriques, sont des estimateurs naturels de m et σ^2.
Comme les (X_i) sont des v.a.i.i.d., que $\sigma^2 < \infty$, on peut appliquer la loi forte des grands nombres et on obtient $\overline{X}_n \xrightarrow[n\to\infty]{p.s.} m$. Par ailleurs, $\overline{\sigma}^2 = \frac{1}{n}\sum_{j=1}^n X_j^2 - \big(\overline{X}_n\big)^2$. On peut appliquer la loi forte des grands nombres à la suite (X_i^2) qui sont des v.a.i.i.d. d'espérance finie, donc $\frac{1}{n}\sum_{j=1}^n X_j^2 \xrightarrow[n\to\infty]{p.s.} \mathbb{E}[X_1^2]$. Comme $\big(\overline{X}_n\big)^2 \xrightarrow[n\to\infty]{p.s.} m^2$, on en déduit que $\overline{\sigma}^2 \xrightarrow[n\to\infty]{p.s.} \sigma^2$.

(c) Il est clair que $\sigma^2/m = 1-p$, donc $p = 1-\sigma^2/m$ et $k = m/p$ donc $k = \frac{m^2}{m-\sigma^2}$. La fonction $h : (x,y) \in \{(x,y) \in]0,\infty[^2, x > y\} \mapsto \Big(1 - \frac{y}{x}\,,\,\frac{x^2}{x-y}\Big)$ étant continue, alors $h\big(\overline{X}_n, \overline{\sigma}^2\big) \xrightarrow[n\to\infty]{p.s.} h(m,\sigma^2)$ et on en déduit que

$$\widehat{p}_n = 1 - \frac{\overline{\sigma}^2}{\overline{X}_n} \xrightarrow[n\to\infty]{p.s.} p \quad \text{et} \quad \widehat{k}_n = \frac{\overline{X}_n^2}{\overline{X}_n - \overline{\sigma}^2} \xrightarrow[n\to\infty]{p.s.} k.$$

□

7. **(**)** Soit $((\Omega')^n, \mathcal{A}'_n, (\mathbb{P}_\theta)^{\otimes n}, \theta \in]0,\infty[)$ un modèle paramétrique tel que $\mathbb{P}_\theta$ admette pour densité par rapport à la mesure de Lebesgue sur $\mathbf{R}$:

$$f(x) = \frac{1}{\log 2}\,\frac{1}{x}\,\mathbb{I}_{x\in]\theta,2\theta[}.$$

(a) Montrer que l'estimateur du maximum de vraisemblance pour ce modèle n'est pas unique.

(b) On observe $(X_1,\ldots,X_n)$ du modèle statistique et on pose $\widehat{\theta}_n^{(1)} = \dfrac{1}{2}\max(X_1,\ldots,X_n)$ et $\widehat{\theta}_n^{(2)} = \min(X_1,\ldots,$ Déterminer la loi de ces estimateurs et montrer qu'ils convergent. Sans rentrer dans les calculs, donner un court raisonnement montrant qu'ils sont biaisés.

Proof. (a) Le modèle statistique implique que la vraisemblance pour $(x_1,\ldots,x_n) \in]0,\infty[^n$ vaut

$$L(x_1,\ldots,x_n) = \prod_{j=1}^n \frac{1}{\log 2}\,\frac{1}{x_j}\,\mathbb{I}_{x_j\in]\theta,2\theta[} = \frac{1}{(\log 2)^n}\,\frac{1}{\prod_{j=1}^n x_j}\,\mathbb{I}_{\bigcap_{j=1}^n\{x_j\in]\theta,2\theta[\}}.$$

On en déduit que cette vraisemblance est non nulle seulement si θ est tel que $\min_j(x_j) \geq \theta$ et si $\max_j(x_j) \leq 2\theta$, donc si $\frac{1}{2}\max_j(x_j) \leq \theta \leq \min_j(x_j)$. De plus sur l'intervalle $\big]\frac{1}{2}\max_j(x_j)\,,\,\min_j(x_j)\big[$, la fonction vraisemblance est positive strictement et ne dépend pas de θ: elle vaut $\frac{1}{(\log 2)^n}\,\frac{1}{\prod_{j=1}^n x_j}$. Donc le maximum de la vraisemblance est atteint pour toute valeur de θ dans l'intervalle $\big]\frac{1}{2}\max_j(x_j)\,,\,\min_j(x_j)\big[$: l'estimateur n'est pas unique.

(b) On a $X_i \in]\theta,2\theta[$. On en déduit que $\widehat{\theta}_n^{(1)} \in]\theta/2,\theta[$ et $\widehat{\theta}_n^{(2)} \in]\theta,2\theta[$.
Pour $x \in]\theta/2,\theta[$,

$$\begin{aligned}
\mathbb{P}\big(\widehat{\theta}_n^{(1)} \leq x\big) &= \mathbb{P}\big(\max(X_1,\ldots,X_n) \leq 2\,x\big) \\
&= \prod_{j=1}^n \mathbb{P}(X_i \leq 2\,x) \quad \text{(Indépendance)} \\
&= \Big(\frac{1}{\log 2}\int_\theta^{2x}\frac{dt}{t}\Big)^n \quad \text{(Identiquement distribuées)} \\
&= \Big(\frac{1}{\log 2}\,\log\big(\frac{2x}{\theta}\big)\Big)^n \\
\Longrightarrow \quad f_{\widehat{\theta}_n^{(1)}}(x) &= \frac{n}{x\,(\log 2)^n}\Big(\log\big(\frac{2x}{\theta}\big)\Big)^{n-1}\,\mathbb{I}_{x\in]\theta/2,\theta[}.
\end{aligned}$$

De même, pour $x \in]\theta, 2\,\theta[$,

$$\begin{aligned}
\mathbb{P}\big(\widehat{\theta}_n^{(2)} \leq x\big) &= 1 - \mathbb{P}\big(\min(X_1,\ldots,X_n) > x\big) \\
&= 1 - \prod_{j=1}^n \mathbb{P}(X_i \geq x) \quad \text{(Indépendance)} \\
&= 1 - \Big(\frac{1}{\log 2}\int_x^{2\theta}\frac{dt}{t}\Big)^n \quad \text{(Identiquement distribuées)} \\
&= 1 - \Big(\frac{1}{\log 2}\,\log\big(\frac{2\theta}{x}\big)\Big)^n \\
\Longrightarrow \quad f_{\widehat{\theta}_n^{(2)}}(x) &= \frac{n}{x\,(\log 2)^n}\Big(\log\big(\frac{2\theta}{x}\big)\Big)^{n-1}\,\mathbb{I}_{x\in]\theta,2\theta[}.
\end{aligned}$$

Pour $\varepsilon > 0$, on a

$$\mathbb{P}\big(|\widehat{\theta}_n^{(1)} - \theta| \geq \varepsilon\big) = \mathbb{P}\big(\widehat{\theta}_n^{(1)} \leq \theta - \varepsilon\big) = \Big(1 + \frac{1}{\log 2}\,\log\big(1 - \frac{\varepsilon}{\theta}\big)\Big)^n.$$

Comme $\log\big(1 - \frac{\varepsilon}{\theta}\big) < 0$, il est clair que $\mathbb{P}\big(|\widehat{\theta}_n^{(1)} - \theta| \geq \varepsilon\big) \xrightarrow[n\to\infty]{} 0$: on a bien $\widehat{\theta}_n^{(1)} \xrightarrow[n\to+\infty]{\mathcal{P}} \theta$.
On raisonne de la même manière pour $\widehat{\theta}_n^{(1)}$.
Enfin, comme $\widehat{\theta}_n^{(1)} \in]\theta/2,\theta[$ pour tout n, on ne peut que avoir $\mathbb{E}[\widehat{\theta}_n^{(1)}] < \theta$: l'estimateur est biaisé.

□

8. **(**)** Soit X une v.a. suivant une loi binomiale $\mathcal{B}\big(n, \frac{1}{\lambda}\big)$ où $n \in \mathbf{N}^*$ est connu et $\lambda \geq 1$ est inconnue. On observe une réalisation de X est on estime λ par un estimateur $\widehat{\lambda} = \widehat{\lambda}(X)$.

   (a) Montrer que si $\widehat{\lambda}$ est un estimateur sans biais de λ, alors pour tout $\lambda \in [1, \infty[$,

$$\lambda^{n+1} - \sum_{k=0}^{n} \binom{n}{k} \widehat{\lambda}(k)\,(\lambda - 1)^{n-k} = 0.$$

   (b) En déduire qu'il n'existe pas d'estimateur $\widehat{\lambda}$ sans biais de λ.

Proof. (a) Si $\widehat{\lambda}(X)$ est un estimateur sans biais de λ, alors pour tout $\lambda \in [1, \infty[$, $\mathbb{E}[\widehat{\lambda}(X)] = \lambda$. Mais $X \to \widehat{\lambda}(X)$ étant une fonction mesurable, on a

$$\mathbb{E}[\widehat{\lambda}(X)] = \sum_{k=0}^{n} \widehat{\lambda}(k) \tbinom{n}{k}\, \lambda^{-k}\,(1 - \lambda^{-1})^{n-k} = \sum_{k=0}^{n} \widehat{\lambda}(k) \tbinom{n}{k}\, \lambda^{-n}\,(\lambda - 1)^{n-k} \lambda.$$

Il ne reste plus qu'à passer le λ^{-n} de l'autre côté pour obtenir la formule demandée.

(b) On rappelle que $\widehat{\lambda}$ dépend de X mais ne peut pas dépendre de λ. Donc pour l'estimateur soit sans biais, il faut que le polynôme en λ $\lambda^{n+1} - \sum_{k=0}^{n} \binom{n}{k} \widehat{\lambda}(k)\,(\lambda - 1)^{n-k}$ soit nul pour tout $\lambda > 1$. Or c'est un polynôme de degré $n + 1$, qui a au plus $n + 1$ racine. Ce polynôme ne pouvant pas être nul, il est donc impossible que l'estimateur soit sans biais.

□

9. **(**)** Soit un n-échantillon de v.a.i.i.d. $(X_1, \ldots, X_n)$ de loi admettant la densité f par rapport à la mesure de Lebesgue sur $\mathbb{R}$ avec:

$$f(x) = (1 - \theta) \cdot \mathbb{I}_{]-1/2,0[}(x) + (1 + \theta) \cdot \mathbb{I}_{]0,1/2[}(x),$$

où $\theta \in]-1, 1[$ est un paramètre inconnu.

   (a) Déterminer l'estimateur $\widehat{\theta}_n$ du maximum de vraisemblance.

   (b) Est-il sans biais? Converge-t-il? $\widehat{\theta}_n$ est-il un estimateur efficace?

   (c) Quelle est la loi limite de $\sqrt{n}(\widehat{\theta}_n - \theta)$? En déduire un intervalle de confiance à 95%.

Proof. (a) Les X_i prennent leurs valeurs dans $]-1/2, 0[\cup]0, 1/2[$ et pour $(x_1, \ldots, x_n) \in \big(]-1/2, 0[\cup]0, 1/2[\big)^n$, la vraisemblance de $(X_1, \ldots, X_n)$ est:

$$f_{(X_1,\ldots,X_n)}(x_1, \ldots, x_n) = \prod_{j=1}^{n} f(x_j) = (1-\theta)^{\sum_{j=1}^n \mathbb{I}_{x_j<0}}(1+\theta)^{\sum_{j=1}^n \mathbb{I}_{x_j>0}}.$$

Par conséquent, la log-vraisemblance de $(X_1, \ldots, X_n)$ comme fonction de θ vaut:

$$\ell_n(\theta) = \sum_{j=1}^{n} \mathbb{I}_{X_j<0}\, \ln(1-\theta) + \Big(n - \sum_{j=1}^{n} \mathbb{I}_{X_j<0}\Big)\, \ln(1+\theta).$$

On en déduit que $\ell_n'(\theta) = -\dfrac{1}{1-\theta} \sum_{j=1}^{n} \mathbb{I}_{X_j<0} + \dfrac{1}{1+\theta}\Big(n - \sum_{j=1}^{n} \mathbb{I}_{X_j<0}\Big)$. En résolvant l'équation $\ell_n'(\theta) = 0$, on obtient $\theta = 1 - \dfrac{2}{n} \sum_{j=1}^{n} \mathbb{I}_{X_j<0}$. il reste à calculer $\ell_n''(\theta) = -\dfrac{1}{(1-\theta)^2} \sum_{j=1}^{n} \mathbb{I}_{X_j<0} - \dfrac{1}{(1+\theta)^2}\Big(n - \sum_{j=1}^{n} \mathbb{I}_{X_j<0}\Big) < 0$: la fonction est bien concave et le point critique est un maximum global. Donc

$$\widehat{\theta} = 1 - \frac{2}{n} \sum_{j=1}^{n} \mathbb{I}_{X_j<0}.$$

(b) On a $\mathbb{E}[\mathbb{I}_{X_j<0}] = (1-\theta) \int_{-1/2}^{0} dt = \frac{1}{2}\,(1-\theta)$. On en déduit que $\mathbb{E}[\widehat{\theta}] = 1 - 2\,\frac{1}{2}\,(1-\theta) = \theta$: l'estimateur est sans biais.

La suite $(\mathbb{I}_{X_j<0})_j$ est une suite de v.a.i.i.d. telle que $\mathbb{E}[\mathbb{I}_{X_j<0}] < \infty$: on peut appliquer la loi forte des grands nombres et on obtient:

$$\frac{1}{n} \sum_{j=1}^{n} \mathbb{I}_{X_j<0} \xrightarrow[n\to\infty]{p.s.} \mathbb{E}[\mathbb{I}_{X_j<0}] = \frac{1}{2}\,(1-\theta) \quad\Longrightarrow\quad \widehat{\theta} \xrightarrow[n\to\infty]{p.s.} \theta.$$

On a $\mathrm{var}(\mathbb{I}_{X_j<0}) = \frac{1}{2}(1-\theta) - \frac{1}{4}(1-\theta)^2 = \frac{1}{4}(1-\theta^2)$ donc $\mathrm{var}(\widehat{\theta}) = \frac{1}{n}(1-\theta^2)$. Par ailleurs, l'information de Fisher de X_1 sur θ est:

$$I_1(\theta) = -\mathbb{E}\big[\ell_1''(\theta)\big] = -\mathbb{E}\Big[-\frac{1}{(1-\theta)^2}\mathbb{I}_{X_1<0} - \frac{1}{(1+\theta)^2}\Big(1-\mathbb{I}_{X_1<0}\Big)\Big] = \frac{1}{2(1-\theta)} + \frac{1}{2(1+\theta)} = \frac{1}{1-\theta^2}.$$

On a bien $I_1^{-1}(\theta) = n\,\mathrm{var}(\widehat{\theta})$: l'estimateur sans biais est efficace.

(c) Les hypothèses pour le TLC sont aussi vérifiées et comme $\mathrm{var}(\mathbb{I}_{X_j<0}) = \frac{1}{4}(1-\theta^2) < \infty$, on a

$$\sqrt{n}\Big(\frac{1}{n}\sum_{j=1}^n \mathbb{I}_{X_j<0} - \frac{1}{2}(1-\theta)\Big) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}\big(0\,,\,\frac{1}{4}(1-\theta^2)\big) \quad\Longrightarrow\quad \sqrt{n}\,\big(\widehat{\theta}-\theta\big) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}\big(0\,,\,(1-\theta^2)\big).$$

(d) Le TLC précédent peut être étendu par le Lemme de Slutsky et on obtient:

$$\sqrt{n}\,\frac{1}{\sqrt{1-\theta^2}}\,\big(\widehat{\theta}-\theta\big) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}\big(0\,,\,1\big).$$

Si on note $q \simeq 1.96$ le quantile à 97.5% de la loi normale centrée réduite, alors le TLC précédent permet d'obtenir l'intervalle de confiance asymptotique à 95% sur θ:

$$\Big[\widehat{\theta} - q\,\frac{\sqrt{1-\theta^2}}{\sqrt{n}}\,,\,\widehat{\theta} + q\,\frac{\sqrt{1-\theta^2}}{\sqrt{n}}\Big].$$

□

10. **(**)** On considère le modèle paramétrique gaussien $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mathcal{N}(m,\sigma^2)^{\otimes n}, (m,\sigma^2) \in \mathbb{R}\times]0,+\infty[)$ et $(X_1\ldots,X_n)$ un échantillon observé de ce modèle.

(a) Soit l'estimateur $\widehat{T}_n = (\overline{X}_n, \overline{\sigma}_n^2)$, où $\overline{X}_n = \dfrac{1}{n}\displaystyle\sum_{i=1}^n X_i$ et $\overline{\sigma}_n^2 = \dfrac{1}{n-1}\displaystyle\sum_{i=1}^n (X_i - \overline{X}_n)^2$. En utilisant le théorème de Cochran (on considérera le sous-espace engendré par le vecteur de $\mathbf{R}^n$, $(1,\ldots,1)'$), montrer que cet estimateur est sans biais et que ses 2 composantes sont indépendantes. Est-il efficace?

(b) Déterminer l'estimateur du maximum de vraisemblance de (m,σ^2). Est-il biaisé? Efficace? Comparer avec l'estimateur précédent.

Proof. (a) Si on note le vecteur $Z = (Z_1,\ldots,Z_n)$ tel que $X_i = m + Z_i$ alors Z est un vecteur gaussien centré de matrice de variance-covariance $\sigma^2 I_n$. De plus $\overline{X}_n = \overline{Z}_n + m$. Soit le vecteur de $\mathbf{R}^n$, $\mathbb{I} = \frac{1}{\sqrt{n}}(1,\ldots,1)'$, qui est tel que $\|\mathbb{I}\| = 1$ (norme euclidienne). Pour tout vecteur $U \in \mathbf{R}^n$, la projection orthogonale de U sur la droite vectorielle engendrée par $\mathbb{I}$, qui est normé, est le vecteur $< \mathbb{I}, U > \mathbb{I}$. Ainsi la projection orthogonale de $Z = (Z_1,\ldots,X_n)'$ est le vecteur $\big(\overline{Z}_n,\ldots,\overline{Z}_n\big)'$. Ainsi $\overline{Z}_n = (1,0,\ldots,0)\,P_{\mathbb{I}}(Z)$, où $P_{\mathbb{I}}(Z)$ désigne la projection orthogonale sur le sous-espace vectorielle engrendré par $\mathbb{I}$.
Par ailleurs, $\sum_{i=1}^n (X_i - \overline{X}_n)^2 = \big\|X - P_{\mathbb{I}}(X)\big\|^2 = \big\|Z - P_{\mathbb{I}}(Z)\big\|^2$. Donc $\overline{\sigma}_n^2 = \frac{1}{n-1}\big\|P_{\mathbb{I}^\perp}(Z)\big\|^2$.
Il est bien connu que $\overline{X}_n$ est un estimateur sans biais de $\mathbb{E}[X_1] = m$. Par ailleurs, d'après le Théorème de Cochran, $\big\|P_{\mathbb{I}^\perp}(Z)\big\|^2 \overset{\mathcal{L}}{\sim} \sigma^2\,\chi^2(n-1)$ car $\dim(\mathbb{I}^\perp) = n-1$. Or $\mathbb{E}[\chi^2(n-1)] = n-1$, donc $\mathbb{E}[\overline{\sigma}_n^2] = \sigma^2$, l'estimateur est sans biais, et il en est de même pour $\widehat{T}_n$.
D'après le Théorème de Cochran, $P_{\mathbb{I}}(Z)$ et $P_{\mathbb{I}^\perp}(Z)$ sont deux variables gaussiennes indépendantes, donc $\overline{X}_n = m + (1,0,\ldots,0)\,P_{\mathbb{I}}(Z)$ et $\overline{\sigma}_n^2 = \frac{1}{n-1}\big\|P_{\mathbb{I}^\perp}(Z)\big\|^2$ sont deux variables aléatoires indépendantes.
Pour chercher si l'estimateur est efficace, on va déterminer la matrice d'information de Fisher de X_1 pour (m,σ^2). Pour cela, on écrit la log-vraisemblance de X_1:

$$\ell_1(m,\sigma^2) = \ln\big(f_{X_1}(X_1)\big) = -\frac{1}{2}\Big(\ln(2\pi) + \ln(\sigma^2) + \frac{(X_1-m)^2}{\sigma^2}\Big).$$

On en déduit $\nabla\ell_1(m,\sigma^2) = \begin{pmatrix} (X_1-m)/\sigma^2 \\ \big((X_1-m)^2-\sigma^2\big)/2\sigma^4 \end{pmatrix}$, d'où

$$I_1(m,\sigma^2) = -\mathbb{E}\Big[\begin{pmatrix} -\frac{1}{\sigma^2} & -\frac{X_1-m}{\sigma^4} \\ -\frac{X_1-m}{\sigma^4} & \frac{1}{2\sigma^4} - \frac{(X_1-m)^2}{\sigma^6} \end{pmatrix}\Big] = \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{pmatrix} \Longrightarrow I_1^{-1}(m,\sigma^2) = \begin{pmatrix} \sigma^2 & 0 \\ 0 & 2\sigma^4 \end{pmatrix}$$

Il nous reste à déterminer la matrice de variance-covariance de $\widehat{T}_n$. Il est bien connu que $\mathrm{var}(\overline{X}_n) = \frac{1}{n}\sigma^2$ et d'après la question précédente, $\mathrm{cov}(\overline{X}_n, \overline{\sigma}_n^2) = 0$ (indépendance). Il nous reste à calculer $\mathrm{var}(\overline{\sigma}_n^2)$. On utilise le Théorème de Cochran pour calculer $\mathrm{var}\Big(\frac{1}{n-1}\,\big\|P_{\mathbb{I}^\perp}(Z)\big\|^2\Big)=\mathrm{var}\Big(\frac{1}{n-1}\,\sigma^2\,\chi^2(n-1)\Big)$ car la dimension de $\mathbb{I}^\perp$ est $n-1$. On en déduit que $\mathrm{var}(\overline{\sigma}^2_n)=\frac{2}{n-1}\,\sigma^4$. Et ainsi:

$$n\,\mathrm{cov}\big(\widehat{T}_n\big)=\begin{pmatrix}\sigma^2 & 0\\ 0 & \frac{2n}{n-1}\,\sigma^4\end{pmatrix}>I_1^{-1}(m,\sigma^2).$$

L'estimateur n'est pas efficace mais il est asymptotiquement efficace.

(b) En maximisant la log-vraisemblance, on trouve comme unique estimateur $\widetilde{T}_n=(\overline{X}_n,\widetilde{\sigma}^2_n)$, où $\widetilde{\sigma}^2_n=\dfrac{1}{n}\displaystyle\sum_{i=1}^n(X_i-\overline{X}_n)^2$.
Il est clair que cet estimateur est biaisé (voir précédemment), mais il est asymptotiquement sans biais.
La notion d'efficacité s'applique aux estimateurs sans biais, donc pas approprié pour $\widetilde{T}_n$.
Le risque quadratique de $\widehat{T}_n$ est $R(\widehat{T}_n)=\mathrm{var}(\overline{X}_n)+\mathrm{var}(\overline{\sigma}^2_n)=\frac{1}{n}\,\sigma^2+\frac{2}{n-1}\,\sigma^4$.
Le risque quadratique de $\widetilde{T}_n$ est $R(\widetilde{T}_n)=\mathrm{var}(\overline{X}_n)+\mathrm{var}(\widetilde{\sigma}^2_n)+B^2(\widetilde{\sigma}^2_n)=\frac{1}{n}\,\sigma^2+\frac{2}{n}\,\sigma^4+\big(\frac{1}{n}\,\sigma^2\big)^2=\frac{1}{n}\,\sigma^2+\frac{2n+1}{n^2}\,\sigma^4$.
On remarque que $R(\widetilde{T}_n)<R(\widehat{T}_n)$: l'estimateur $\widetilde{T}_n$ converge plus vite (au sens quadratique) que $\widehat{T}_n$.

$\square$

11. **(***)** Soit $((\Omega')^n,\mathcal{A}'_n,(\mathbb{P}_\theta)^{\otimes n},\theta\in]0,\infty[)$ un modèle paramétrique tel que $\mathbb{P}_\theta$ admette pour densité par rapport à la mesure de Lebesgue sur $\mathbf{R}$:

$$f(x)=\mathbb{I}_{x\in]\theta,\theta+1[}.$$

(a) Montrer que l'estimateur du maximum de vraisemblance pour ce modèle n'est pas unique.

(b) On pose $\widehat{\theta}^{(1)}_n=\max(X_1,\ldots,X_n)-1$ et $\widehat{\theta}^{(2)}_n=\min(X_1,\ldots,X_n)$. Pour chacun de ces estimateurs, déterminer la loi, l'espérance et la variance. Sont-ils convergents? Efficaces?

(c) Déterminer un estimateur de la forme $\widehat{\theta}_n=\alpha\,\widehat{\theta}^{(1)}_n+(1-\alpha)\,\widehat{\theta}^{(2)}_n$, avec $\alpha\in[0,1]$, qui soit sans biais.

(d) Pour $n=2$, déterminer $\mathrm{cov}\big(\widehat{\theta}^{(1)}_2,\widehat{\theta}^{(2)}_2\big)$. Le risque quadratique de $\widehat{\theta}_2$ est-il inférieur à ceux de $\widehat{\theta}^{(1)}_2$ et $\widehat{\theta}^{(1)}_2$? $\widehat{\theta}_2$ est-il efficace?

Proof. (a) Les (X_i) prennent leurs valeurs dans $]0,\infty[$. Donc pour $(x_1,\ldots,x_n)\in]0,\infty[^n$, la vraisemblance du modèle en $(x_1,\ldots,x_n)$ vaut:

$$f_{(X_1,\ldots,X_n)}(x_1,\ldots,x_n)=\prod_{j=1}^n\mathbb{I}_{x_j\in]\theta,\theta+1[}=\mathbb{I}_{\max_j(x_j)-1<\theta<\min_j(x_j)}$$

du fait de l'indépendance. Comme fonction de θ, cette vraisemblance prise en $(X_1,\ldots,X_n)$ est donc maximale (en valant 1) pour tout réel compris strictement entre $\max_j(X_j)-1$ et $\min_j(X_j)$: il n'y a pas unicité de l'estimateur.

(b) Les deux estimateurs $\widehat{\theta}^{(1)}_n$ et $\widehat{\theta}^{(2)}_n$ prennent respectivement leurs valeurs dans $]\theta-1,\theta[$ et $]\theta,\theta+1[$. Pour $x\in]\theta-1,\theta[$, on a:

$$\mathbb{P}\big(\widehat{\theta}^{(1)}_n\le x\big)=\mathbb{P}\le\big(\max_j(X_j)\le x+1\big)=\big(\mathbb{P}\big(X_j\le x+1\big)\big)^n=(x+1-\theta)^n.$$

De ceci, on en déduit que $f_{\widehat{\theta}^{(1)}_n}(x)=n\,(x+1-\theta)^{n-1}\mathbb{I}_{x\in]\theta-1,\theta[}$.
On a également

$$\mathbb{E}[\widehat{\theta}^{(1)}_n]=n\int_{\theta-1}^{\theta}x\,(x+1-\theta)^{n-1}dx=n\int_0^1(y+\theta-1)\,y^{n-1}dy=n\Big[\frac{y^{n+1}}{n+1}+(\theta-1)\,\frac{y^n}{n}\Big]_0^1=\frac{n}{n+1}+(\theta-1)=\theta-\frac{1}{n+1}.$$

Donc $\widehat{\theta}^{(1)}_n$ est biaisé mais asymptotiquement non biaisé.

$$\mathrm{var}(\widehat{\theta}^{(1)}_n)=n\int_{\theta-1}^{\theta}x^2\,(x+1-\theta)^{n-1}dx-\Big(\theta-\frac{1}{n+1}\Big)^2=n\int_0^1(y+\theta-1)^2\,y^{n-1}dy-\Big(\theta-\frac{1}{n+1}\Big)^2=\frac{n}{(n+2)(n+1)^2}.$$

Par symétrie, $f_{\widehat{\theta}^{(1)}_n}(x)=n\,(x-\theta)^{n-1}\mathbb{I}_{x\in]\theta,\theta+1[}$, $\mathbb{E}[\widehat{\theta}^{(2)}_n]=\theta+\frac{1}{n+1}$ et $\mathrm{var}(\widehat{\theta}^{(2)}_n)=\frac{n}{(n+2)(n+1)^2}$.
Comme $\widehat{\theta}^{(1)}_n$ et $\widehat{\theta}^{(2)}_n$ sont asymptotiquement sans biais, comme leurs variances tendent vers 0 avec n, alors ces deux estimateurs sont convergents (d'après l'Inégalité de Bienaymé-Tchebychev).

(c) On a $\mathbb{E}[\widehat{\theta}_n]=\alpha\,\mathbb{E}[\widehat{\theta}^{(1)}_n]+(1-\alpha)\,\mathbb{E}[\widehat{\theta}^{(2)}_n]=\theta+(1-2\alpha)\,\frac{1}{n+1}$. Donc pour que $\widehat{\theta}_n$ soit un estimateur sans biais il faut que $\alpha=1/2$.

(d) Comme pour $n=2$, $\max_j(X_j)\min_j(X_j) = X_1X_2$, alors $\widehat{\theta}_2^{(1)}\widehat{\theta}_2^{(2)} = X_1X_2 - \widehat{\theta}_2^{(2)}$. En conséquence,

$$\operatorname{cov}\big(\widehat{\theta}_2^{(1)},\widehat{\theta}_2^{(2)}\big) = \mathbb{E}\big[X_1X_2\big] - \mathbb{E}\big[\widehat{\theta}_2^{(2)}\big] - \mathbb{E}[\widehat{\theta}_n^{(1)}]\,\mathbb{E}[\widehat{\theta}_n^{(2)}] = \big(\theta+\frac{1}{2}\big)^2 - \big(\theta+\frac{1}{3}\big) - \big(\theta^2-\frac{1}{9}\big) = \frac{1}{36}.$$

Comme on a $\operatorname{var}\big(\widehat{\theta}_2^{(1)}\big) = \operatorname{var}\big(\widehat{\theta}_2^{(1)}\big) = \frac{2}{4\times 9} = \frac{1}{18}$, alors

$$\operatorname{var}(\widehat{\theta}_2) = \frac{1}{4}\left(\operatorname{var}\big(\widehat{\theta}_2^{(1)}\big) + \operatorname{var}\big(\widehat{\theta}_2^{(1)}\big) + 2\operatorname{cov}\big(\widehat{\theta}_2^{(1)},\widehat{\theta}_2^{(2)}\big)\right) = \frac{1}{4}\,\frac{1}{6} = \frac{1}{24}.$$

L'estimateur $\widehat{\theta}_2$ est clairement plus intéressant que $\widehat{\theta}_2^{(1)}$ ou $\widehat{\theta}_2^{(2)}$.

□

12. **(***)** Soit un n-échantillon $(X_1,\dots,X_n)$ de v.a. telles que X_1 suit une loi de Bernoulli de paramètre $1/2$, et $\mathbb{P}(X_{n+1}=1 \mid X_n=1) = \mathbb{P}(X_{n+1}=0 \mid X_n=0) = p$ et $\mathbb{P}(X_{n+1}=1 \mid X_n=0) = \mathbb{P}(X_{n+1}=0 \mid X_n=1) = 1-p$ pour $n\in\mathbb{N}^*$, avec $p\in]0,1[$ un paramètre inconnu.

    (a) Montrer que pour tout $k\in\mathbf{N}^*$, $X_k \overset{\mathcal{L}}{\sim} \mathcal{B}(1/2)$.

    (b) Déterminer l'estimateur de maximum de vraisemblance de p. Est-il biaisé?

Proof. (a) Il est clair que les (X_i) prennent leurs valeurs dans $\{0,1\}$. Par ailleurs,

$$\mathbb{P}(X_2=1) = \mathbb{P}(X_2=1\,|\,X_1=1)\,\mathbb{P}(X_1=1) + \mathbb{P}(X_2=1\,|\,X_1=0)\,\mathbb{P}(X_1=0) = \frac{1}{2}\,\big(p+(1-p)\big) = \frac{1}{2}.$$

De ceci, on en déduit que $\mathbb{P}(X_2=0) = 1-\mathbb{P}(X_2=1) = \frac{1}{2}$, et par itération on a bien $X_k \overset{\mathcal{L}}{\sim} \mathcal{B}(1/2)$ pour tout k.

(b) La vraisemblance de $(X_1,\dots,X_n)$ est pour $(x_1,\dots,x_n)\in\{0,1\}^n$:

$$\begin{aligned}
\mathbb{P}\big(X_1=x_1\cap\ldots\cap X_n=x_n\big) &= \mathbb{P}\big(X_n=x_n\,|\,X_{n-1}=x_{n-1}\cap\ldots\cap X_1=x_1\big)\,\mathbb{P}\big(X_{n-1}=x_{n-1}\cap\ldots\cap X_1=x_1\big)\\
&= \mathbb{P}\big(X_n=x_n\,|\,X_{n-1}=x_{n-1}\big)\times\mathbb{P}\big(X_{n-1}=x_{n-1}\,|\,X_{n-2}=x_{n-2}\big)\times\\
&\qquad\times\cdots\times\mathbb{P}\big(X_2=x_2\,|\,X_1=x_1\big)\times\mathbb{P}\big(X_1=x_1\big)\\
&= p^{\mathbb{I}_{x_n=x_{n-1}}}(1-p)^{1-\mathbb{I}_{x_n=x_{n-1}}}\,p^{\mathbb{I}_{x_{n-1}=x_{n-2}}}(1-p)^{1-\mathbb{I}_{x_{n-1}=x_{n-2}}}\times\\
&\qquad\times\cdots\times p^{\mathbb{I}_{x_2=x_1}}(1-p)^{1-\mathbb{I}_{x_2=x_1}}\,\frac{1}{2}\\
&= \frac{1}{2}\,p^{\sum_{j=2}^n\mathbb{I}_{x_j=x_{j-1}}}(1-p)^{(n-1)-\sum_{j=2}^n\mathbb{I}_{x_j=x_{j-1}}}.
\end{aligned}$$

En utilisant ce que l'on sait déjà sur le maximum de vraisemblance pour un échantillon de v.a.i.i.d. de loi $\mathcal{B}(p)$, on en déduit que l'estimateur par maximum de vraisemblance est unique et vaut:

$$\widehat{p} = \frac{1}{n-1}\sum_{j=2}^n \mathbb{I}_{X_j=X_{j-1}}.$$

On a

$$\begin{aligned}
\mathbb{E}\big[\mathbb{I}_{X_j=X_{j-1}}\big] = \mathbb{P}\big(X_j=X_{j-1}\big) &= \mathbb{P}\big(X_j=X_{j-1}\cap X_{j-1}=1\big) + \mathbb{P}\big(X_j=X_{j-1}\cap X_{j-1}=0\big)\\
&= \frac{1}{2}\,\Big(\mathbb{P}\big(X_j=1\,|\,X_{j-1}=1\big) + \mathbb{P}\big(X_j=0\,|\,X_{j-1}=0\big)\Big) = \frac{1}{2}\,(p+p) = p.
\end{aligned}$$

Ainsi $\mathbb{E}\big[\widehat{p}\big] = p$: l'estimateur est non biaisé.

□

Feuille n^o 6:
Tests paramétriques

1. **(*)** Dix coureurs de 1500 m testent si la prise d'un complément alimentaire non dopant améliore leurs performances. Ils courrent un 1500 m sans le produit (ou plutôt avec un placébo) et 2 jours après un autre 1500 m avec le produit. On note $D_1, \ldots, D_{10}$ la différence entre leur temps de cours sans et avec le complément alimentaire. On supposera que les D_i sont des v.a.i.i.d. gaussiennes. On veut tester entre H_0: le produit n'a pas d'effet, et H_1: le produit a un effet positif. Sur les 10 coureurs, on a obtenu en moyenne 3s de moins entre les 2 courses avec un écart-type de 4s. Déterminer le résultat du test avec un niveau de risque de 5%.

Proof. Si les D_i ont pour loi commune une loi $\mathcal{N}(m, \sigma^2)$. Les hypothèse de test deviennent alors H_0: $m = 0$ et H_1: $m < 0$. On va utiliser naturellement comme statistique de test la moyenne empirique $\overline{D}_{10}$ en tant qu'estimateur efficace de m. Les règles de décision à associer sont pour H_0: $\overline{D}_{10} \geq s_\alpha$ et pour H_1: $\overline{D}_{10} < s_\alpha$, où s_α est un réel dépendant de $\alpha = 0.05$. Pour déterminer ce seuil, on utilise la définition de α, à savoir $\mathbb{P}_{H_0}(\text{Choisir } H_1) = \alpha$, soit $\mathbb{P}_{m=0}(\overline{D}_{10} < s_\alpha) = \alpha$.
Or on sait que sous H_0, donc si $m = 0$, $\sqrt{10}\,\overline{D}_{10}/\overline{\sigma} \overset{\mathcal{L}}{\sim} t(9)$, loi de Student. Donc si on note q le quantile d'ordre 0.05 pour la loi $t(9)$, et $q \simeq -1.83$, on a $s_\alpha = q\,\overline{\sigma}/\sqrt{10}$, soit $s_\alpha \simeq -2.31$. Comme pour cette expérience $\overline{D}_{10} = -3 < s_\alpha$, on choisit H_1: le complément améliore les performances. □

2. **(**)** On a lancé 100 fois une pièce de monnaie et obtenu 58 fois pile. Avec un niveau $\alpha = 0.05$, tester si la pièce est équilibrée (on pourra utiliser d'abord une inégalité de Bienaymé-Tchebychev puis une approximation gaussienne). Déterminer une formule permettant de calculer la fonction puissance du test en fonction de p. Tester également si la pièce est truquée toujours avec une erreur de première espèce $\alpha = 0.05$. Aboutit-on à la même conclusion du test?

Proof. Le modèle statistique sous-jacent est $(\{0,1\}^n, \mathcal{P}(\{0,1\}^n), \mathcal{B}(p)^{\otimes n}, p \in [0,1])$, avec $n = 100$. Le problème de test s'écrit $H_0 :\ p = 1/2$ contre $H_1 : p \neq 1/2$.
On doit commencer par choisir une statistique de test. Naturellement comme la question porte sur p, on utilise $\widehat{p}_n = \overline{X}_n$, les X_i étant donc des v.a.i.i.d. de loi $\mathcal{B}(p)$. Les régles de décision par rapport au problème de test sont: H_0 est choisi lorsque $|\widehat{p}_n - 1/2| \leq s_\alpha$ et H_1 si $|\widehat{p}_n - 1/2| > s_\alpha$ où s_α est un seuil dépendant de α, niveau du test.
Pour déterminer s_α, on sait que l'on veut $\mathbb{P}_{H_0}(\text{Choisir } H_1) \leq \alpha$ par définition de l'erreur de première espèce. Deux possibilités:

- Avec l'Inégalité de Bienaymé-Tchebychev, sous H_0, $\mathbb{P}(|\widehat{p}_n - 1/2| \geq s_\alpha) \leq \frac{1}{4\,n}\,\frac{1}{s_\alpha^2}$. On choisit donc s_α tel que $\frac{1}{4\,n}\,\frac{1}{s_\alpha^2} = \alpha$, soit $s_\alpha = \sqrt{1/20} \simeq 0.22$. Concrètement on choisira H_0 car $\widehat{p}_n = 0.58$ pour cet échantillon, donc $|\widehat{p}_n - 1/2| = 0.08 < s_\alpha$.
- On peut utiliser le TLC, et on a alors sour H_0, $\sqrt{n}(\widehat{p}_n - 1/2) \overset{\mathcal{L}}{\sim} \mathcal{N}(0, 1/4)$ en considérant que $n = 100$ est suffisamment grand pour accepter l'approximation gaussienne. On sait donc que $\mathbb{P}(20\,|\widehat{p}_n - 1/2| \geq q) \simeq \mathbb{P}(|Z| \geq q) = 0.05$ avec Z une v.a. gaussienne centrée réduite et $q \simeq 1.96$ le quantile à 97.5% de cette loi. D'où $s_\alpha \simeq 1.96/20 \simeq 0.098$. Concrètement on choisira H_0 puisque $|\widehat{p}_n - 1/2| = 0.08 < 0.098$

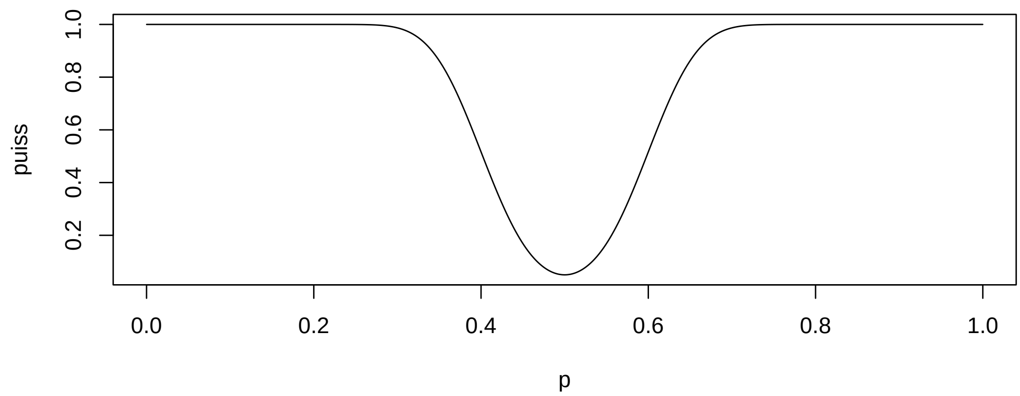
Pour déterminer la fonction puissance du test, on en revient à sa définition:

$$P_\alpha = 1 - \mathbb{P}_{H_1}(\text{Choisir } H_0) = 1 - \mathbb{P}_{\left(X_1 \overset{\mathcal{L}}{\sim} \mathcal{B}(p)\right)}\big(|\widehat{p}_n - 1/2| \leq s_\alpha\big).$$

En utilisant l'approximation gaussienne, sous H_1, c'est-à-dire si $X_1 \overset{\mathcal{L}}{\sim} \mathcal{B}(p)$ alors $10\,(\widehat{p}_n - p) \overset{\mathcal{L}}{\sim} \mathcal{N}(0, p(1-p))$ donc $\widehat{p}_n - 1/2 \overset{\mathcal{L}}{\sim} \mathcal{N}\big(p - 1/2\,,\, \frac{p(1-p)}{100}\big)$. Aussi, avec $Z \overset{\mathcal{L}}{\sim} \mathcal{N}(0,1)$, on a:

$$\mathbb{P}_{\left(X_1 \overset{\mathcal{L}}{\sim} \mathcal{B}(p)\right)}\big(|\widehat{p}_n - 1/2| \leq s_\alpha\big) = \mathbb{P}\Big(Z \leq \frac{10(1/2 - p + s_\alpha)}{\sqrt{p(1-p)}}\Big) - \mathbb{P}\Big(Z \leq \frac{10(1/2 - p - s_\alpha)}{\sqrt{p(1-p)}}\Big)$$

D'où le graphe suivant (obtenu avec `R`) de la puissance en fonction de p:



Si on veut tester maintenant que la pièce est truquée, on échange les hypothèses du test $H_0: \; p \neq 1/2$ contre $H_1: p = 1/2$. Donc la règle décision associée à H_0 devient $|\widehat{p}_n - 1/2| \geq s'_\alpha$ et le seuil s'_α est obtenu par:

$$\sup_{p\in[0,1]\setminus\{1/2\}} \mathbb{P}_{\left(X_1 \overset{\mathcal{L}}{\sim} \mathcal{B}(p)\right)}\left(|\widehat{p}_n - 1/2| \leq s'_\alpha\right) = 0.05.$$

Il est intuitif qu'à s'_α constant, ce sup est atteint en $p = 1/2$, ce qui revient à chosir s'_α tel que $\mathbb{P}_{\left(X_1 \overset{\mathcal{L}}{\sim} \mathcal{B}(1/2)\right)}\left(|\widehat{p}_n - 1/2| \leq s'_\alpha\right) = 0.05$. On retrouve avec l'approximation gaussienne $20\left(\widehat{p}_n - 1/2\right) \overset{\mathcal{L}}{\sim} \mathcal{N}(0,1)$ donc on utilise le quantile à 52.5% de la loi $\mathcal{N}(0,1)$ qui vaut $q' \simeq 0.063$. On a ainsi $20\, s'_\alpha = q'$ d'où $s'_\alpha \simeq 0.0031$. Donc dans ce cas, avec $\widehat{p}_n = 0.58$, $|\widehat{p}_n - 1/2| = 0.08 > s'_\alpha$: on choisit H_0, la pièce est truquée. □

3. **(*)** Une chaîne des magasins décide d'adopter une nouvelle politique de gestion des stocks pour l'ensemble de ses succursales. Auparavant, le bénéfice mensuel d'une succursale était en moyenne égal à 300000 euros. Après la mise en place de la nouvelle politique pour 100 succursales "pilotes", on observe un bénéfice moyen de 305000 euros avec un écart-type de 20000 euros. En utilisant une approximation gaussienne, peut-on conclure à l'efficacité de cette politique au niveau 5%? 1%? Et que se passe-t-il si l'hypothèse testée est que la politique n'a pas d'effet?

Proof. Le modèle statistique sous-jacent est $([0,\infty[^n, \mathcal{B}([0,\infty[^n), \mathbb{P}_m^{\otimes n}, m \in [0,\infty[)$, avec $n = 100$. La loi $\mathbb{P}_m$ n'est pas connue, mais on sait que si $X \overset{\mathcal{L}}{\sim} \mathbb{P}_m$ alors $\mathbb{E}[X] = m$ (on est dans une situation semi-paramétrique).

On veut tester: $H_0: \; m > 300000$ contre $H_1: \; m \leq 300000$. Sans connaissance sur $\mathbb{P}_m$, on va utiliser $\overline{X}_n$ comme estimateur de m pour établir les règles de décision: H_0 choisie si $\overline{X}_n - 300000 \geq s_\alpha$ et H_1 choisie si $\overline{X}_n - 300000 < s_\alpha$, où s_α est un seuil dépendant de α erreur de première espèce.
Sa détermination s'obtient par l'équation $\sup_{m>300000} \mathbb{P}_{H_0}\left(\overline{X}_n - 300000 < s_\alpha\right) = \alpha$, ce qui revient à $\mathbb{P}_{m=300000}\left(\overline{X}_n - 300000 < s_\alpha\right) = \alpha$. Pour obtenir cette probabilité, on va utiliser le TLC2, dont les hypothèses sont supposées satisfaites (v.a.i.i.d. et $\mathbb{E}[X_1^2] < \infty$) ce qui permet d'écrire:

$$\sqrt{100}\,\frac{1}{20000}\left(\overline{X}_n - 300000\right) = \frac{1}{2000}\left(\overline{X}_n - 300000\right) \simeq \overset{\mathcal{L}}{\sim} \mathcal{N}(0,1). \qquad (1)$$

On a doit donc avoir $\dfrac{s_\alpha}{2000} \simeq q_\alpha$ où q_α est le quantile d'ordre α de la loi $\mathcal{N}(0,1)$. Donc si $\alpha = 0.05$, alors $q_\alpha \simeq -1.96$, soit $s_\alpha \simeq -3920$ et il est clair que $\overline{X}_n - 300000 > s_\alpha$, donc H_0 est choisie. Et ce sera clairement la même chose pour $\alpha = 0.01$.

Echangeons les hypothèses et désormais on veut tester H_0: $m = 300000$ contre H_1: $m > 300000$. La statistique de test est la même et les règles de décision sont désormais: H_0 choisie si $\overline{X}_n - 300000 \leq s'_\alpha$ et H_1 choisie si $\overline{X}_n - 300000 > s_\alpha$, où s'_α est un seuil dépendant de α erreur de première espèce. Pour trouver s'_α, on utilise l'équation $\mathbb{P}_{H_0}\left(\overline{X}_n - 300000 > s'_\alpha\right) = \alpha$. Comme sous H_0 on a encore (**??**), on en déduit que $\dfrac{s'_\alpha}{2000} \simeq q_{1-\alpha}$, d'où pour $\alpha = 0.05$, $s'_\alpha \simeq 3920$ et ainsi H_0 est rejetée: la politique a eu un effet positif. Et si $\alpha = 0.01$, comme $q_{0.99} \simeq 2.33$, on a $s'_\alpha \simeq 4660$, H_0 est encore rejetée (mais de peu). □

4. **(**)** Dans le cas d'un échantillon $(X_1, \cdots, X_n)$ de v.a.i.i.d. de loi $\mathcal{E}(\lambda)$, où λ est inconnu. On considère les différents problèmes de test suivants:

   (1) $H_0: \; \lambda = 1$ contre $H_1: \; \lambda = 2$;

(2) $H_0:\ \lambda = 1$ contre $H_1:\ \lambda > 1$;

(3) $H_0:\ \lambda < 1$ contre $H_1:\ \lambda > 1$;

(4) $H_0:\ \lambda = 1$ contre $H_1:\ \lambda \neq 1$.

(a) En utilisant le TLC donné exercice 4 du TD5, déterminer le test de Wald de niveau 5% (statistique de test, région critique en fonction du niveau $\alpha = 0.05$) pour ces différents tests.

(b) Faire ensuite la même chose avec le test du rapport de vraisemblance.

Proof. (a) On considère l'estimateur $\widehat{\lambda} = 1/\overline{X}_n$, qui sera la statistique de test. Dans l'exercice 4 du TD5, on a montré que:

$$\sqrt{n}\,\big(\widehat{\lambda} - \lambda\big) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}\big(0\,,\,\lambda^2\big).$$

On peut donc encore écrire que:

$$\sqrt{n}\,\Big(\frac{\widehat{\lambda}}{\lambda} - 1\Big) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}\big(0\,,\,1\big).$$

Pour le test (1), les règles de décision sont H_0 choisie si $\widehat{\lambda} \leq s_\alpha$, et H_1 choisie si $\widehat{\lambda} > s_\alpha$. Pour déterminer cette zone critique on résout $\mathbb{P}_{H_0}(\text{Choisir } H_1) = \mathbb{P}_{\lambda=1}(\widehat{\lambda} > s_\alpha) = \alpha$. Or, sous H_0, $\sqrt{n}\,\big(\widehat{\lambda} - 1\big) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}\big(0\,,\,1\big)$. Donc comme $\mathbb{P}_{\lambda=1}(\sqrt{n}\,(\widehat{\lambda} - 1) > \sqrt{n}\,(s_\alpha - 1)) = \alpha$, on en déduit que $\sqrt{n}\,(s_\alpha - 1) = q_{1-\alpha}$ quantile d'ordre 0.95 pour la loi $\mathcal{N}(0,1)$, soit $q_{1-\alpha} \simeq 1.645$. Par conséquent, la zone critique est

$$\widehat{\lambda}_n = \frac{1}{\overline{X}_n} \in \Big]1 + 1.645\,\frac{1}{\sqrt{n}}\,,\infty\Big[.$$

Pour le test (2), les règles de décision sont H_0 choisie si $\widehat{\lambda} \leq s_\alpha$, et H_1 choisie si $\widehat{\lambda} > s_\alpha$: c'est exactement le même test que le précédent.

Pour le test (3), les règles de décision sont H_0 choisie si $\widehat{\lambda} \leq s_\alpha$, et H_1 choisie si $\widehat{\lambda} > s_\alpha$. Pour déterminer cette zone critique on résout $\mathbb{P}_{H_0}(\text{Choisir } H_1) = \sup_{\lambda<1} \mathbb{P}_{\lambda}(\widehat{\lambda} > s_\alpha) = \alpha$. Mais ce sup est atteint en $\lambda = 1$. On retombe donc exactement sur la même zone critique que précédemment.

Pour le test (4), les règles de décision sont H_0 choisie si $|\widehat{\lambda} - 1| \leq s_\alpha$, et H_1 choisie si $|\widehat{\lambda} - 1| > s_\alpha$. Pour déterminer cette zone critique on résout $\mathbb{P}_{H_0}(\text{Choisir } H_1) = \mathbb{P}_{\lambda=1}(|\widehat{\lambda} - 1| > s_\alpha) = \alpha$. Or, sous H_0, $\sqrt{n}\,\big(\widehat{\lambda} - 1\big) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}\big(0\,,\,1\big)$. Donc comme $\mathbb{P}_{\lambda=1}(\sqrt{n}\,|\widehat{\lambda} - 1| > \sqrt{n}\,s_\alpha) = \alpha$, on en déduit que $\sqrt{n}\,s_\alpha = q_{1-\alpha/2}$ quantile d'ordre 0.975 pour la loi $\mathcal{N}(0,1)$, soit $q_{1-\alpha/2} \simeq 1.96$. Par conséquent, la zone critique est

$$\widehat{\lambda}_n = \frac{1}{\overline{X}_n} \in \Big] - \infty, 1 - 1.96\,\frac{1}{\sqrt{n}}\Big[\bigcup \Big]1 + 1.96\,\frac{1}{\sqrt{n}}\Big[.$$

(b) On va maintenant considérer le test du rapport de vraisemblance qui se définit par:

$$\widehat{T} = \frac{\sup_{\lambda\in\Lambda_0} L_\lambda(X_1,\ldots,X_n)}{\sup_{\lambda\in\Lambda_1} L_\lambda(X_1,\ldots,X_n)},$$

les ensembles Λ_0 et Λ_1 étant relatifs respectivement aux hypothèses H_0 et H_1, avec, pour $\lambda > 0$,

$$L_\lambda(X_1,\ldots,X_n) = \prod_{j=1}^n \lambda\, e^{-\lambda\, X_j} = \lambda^n\ e^{-\lambda\,\sum_{j=1}^n X_j},$$

du fait que les (X_i) sont des v.a.i.i.d. On peut alors décliner la statistique $\widehat{T}$ suivant les différents problèmes de test:

Pour le test (1), $\widehat{T} = \dfrac{L_1(X_1,\ldots,X_n)}{L_2(X_1,\ldots,X_n)} = 2^{-n}\,e^{\sum_{j=1}^n X_j} = 2^{-n}\,e^{n\,\overline{X}_n}$. Comme la zone critique est donnée par $\widehat{T} < s$, cela revient donc à avoir $\overline{X}_n < s'$. Or d'après le TLC classique sous H_0, $\sqrt{n}\big(\overline{X}_n - 1\big) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0\,,\,1)$. Donc comme $\mathbb{P}_{\lambda=1}\big(\overline{X}_n < s'\big) = \alpha$, qui revient à $\mathbb{P}_{\lambda=1}\big(\sqrt{n}\,(\overline{X}_n - 1) < \sqrt{n}\,(s' - 1)\big) = \alpha$, on a donc $\sqrt{n}\,(s' - 1) \simeq -1.645$ et région critique est donc:

$$\overline{X}_n \in \Big] - \infty\,,\,1 - \frac{1.645}{\sqrt{n}}\Big[.$$

Pour le test (2), on a

$$\widehat{T} = \frac{L_1(X_1,\ldots,X_n)}{\sup_{\lambda>1} L_\lambda(X_1,\ldots,X_n)} = \frac{e^{-\sum_{j=1}^n X_j}}{L_1(X_1,\ldots,X_n)\,\mathbb{I}_{\widehat{\lambda}_n<1} + L_{\widehat{\lambda}_n}(X_1,\ldots,X_n)\,\mathbb{I}_{\widehat{\lambda}_n\geq 1}}$$
$$= \frac{e^{-n\,\overline{X}_n}}{e^{-n\,\overline{X}_n}\,\mathbb{I}_{\overline{X}_n>1} + (\overline{X}_n)^{-n}\,e^{-n}\,\mathbb{I}_{\overline{X}_n\leq 1}}.$$

La zone critique correspond à $\widehat{T} < s$. Donc

- Si $\overline{X}_n > 1$, cela revient à $1 < s$, toujours vrai;
- Si $\overline{X}_n \leq 1$, cela revient à $\exp\Big(n\left(\ln(\overline{X}_n) - \overline{X}_n + 1\right)\Big) < s$ ou encore $\ln(\overline{X}_n) - \overline{X}_n < s'$. La fonction $x \in]0,\infty[\mapsto \ln(x) - x$ est une fonction croissante sur $]0,1[$ puis décroissante sur $]1,\infty[$, s'annulant en 1. Comme on est dans le cas $\overline{X}_n \leq 1$, cela revient donc à $\overline{X}_n < s''$.

De ces deux conditions, on en arrive à une seule qui est $\overline{X}_n < s''$. Pour trouver la valeur de s'', on utilise le TLC sous H_0, soit $\sqrt{n}\big(\overline{X}_n - 1\big) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0\,,1)$. Par conséquent, on peut utiliser la même méthode qu'au-dessus et obtenir pour zone critique:

$$\overline{X}_n \in \Big] -\infty\,, 1 - \frac{1.645}{\sqrt{n}}\Big[.$$

Pour le test (3) on retombe après calculs sur la même zone critique qu'avec le test de Wald.

Pour le test (4), la statistique de test $\widehat{T}$ se simplifie en:

$$\widehat{T} = \frac{L_1(X_1,\dots,X_n)}{\sup_{\lambda\neq 1} L_\lambda(X_1,\dots,X_n)} = \frac{e^{-\sum_{j=1}^n X_j}}{L_{\widehat{\lambda}_n}(X_1,\dots,X_n)} = \frac{e^{-n\,\overline{X}_n}}{(\overline{X}_n)^{-n}\,e^{-n}}.$$

La zone critique correspond à $\widehat{T} < s$, ce qui revient à $\exp\Big(n\left(\ln(\overline{X}_n) - \overline{X}_n + 1\right)\Big) < s$ ou encore $\ln(\overline{X}_n) - \overline{X}_n < s'$. Avec l'étude de la fonction faite précédemment, on en arrive à une zone critique de la forme $|\overline{X}_n - 1| > s''$. En utilisant les résultats précédents, on en arrive à une zone critique de la forme:

$$\overline{X}_n \in \Big] -\infty\,, 1 - \frac{1.96}{\sqrt{n}}\Big[\bigcup\Big]1 + \frac{1.96}{\sqrt{n}}, +\infty\Big[$$

□

5. **(***)** On suppose que le nombre de clients N attendant à la caisse d'un grand magasin à 11h00 du matin peut être modélisé par une loi de Poisson de paramètre $\theta > 0$. On sait que pour les clients la limite du supportable en attente est de voir au maximum 5 clients devant une caisse. La question se pose de savoir si l'on doit augmenter ou nom le nombre de personnes travaillant aux caisses: ceci conduit à un problème de test sur l'opportunité d'embaucher.

   (a) Pour la direction, peu désireuse d'embaucher, le problème se pose de la manière suivante: $H_0:\ \theta \leq 5$ contre $H_1:\ \theta > 5$ à tester avec le niveau 5%; ainsi, la direction veut que la probabilité d'embaucher alors qu'il n'y en a pas besoin est contrôlée (moins de 5%) sans s'intéresser à l'autre erreur possible (qui est de ne pas embaucher alors qu'il le fallait).

   (b) Pour les syndicats, la position est inverse: on veut surtout éviter de ne pas embaucher alors qu'il le fallait. Ils vont tabler sur le fait l'erreur de seconde espèce du test soit de 5%.

   Dans les deux cas, sachant que l'on dispose 1/ d'une seule observation de valeur 6; 2/ de 100 observations de moyenne 5.4, déterminer les résultats des tests du rapport de vraisemblance. On pourra utiliser les quantiles d'ordre 0.95 d'une loi de Poisson de paramètre 5 (9) et d'une loi de Poisson de paramètre 500 (537).

Proof. Le modèle statistique paramétrique est $(\mathbf{N}^n, \mathcal{P}(\mathbf{N}^n), \mathcal{P}(\theta)^{\otimes n}, \theta > 0)$, avec $n = 1$, puis $n = 100$. La vraisemblance est donc:

$$L_\theta(X_1,\dots,X_n) = \prod_{j=1}^n e^{-\theta}\frac{\theta^{X_j}}{X_j!} = e^{-n\,\theta}\,\frac{\theta^{\sum_{j=1}^n X_j}}{\prod_{j=1}^n X_j!}.$$

Dans le premier cas, le problème de test est $H_0:\ \theta \leq 5$ contre $H_1:\ \theta > 5$, c'est-à-dire $H_0:\ \theta \in \Theta_0 =]-\infty, 5]$ contre $H_1:\ \theta \in \Theta_1 =]5,\infty[$. Le test du rapport de vraisemblance est donc

$$\widehat{T} = \frac{\sup_{\theta\in\Theta_0} L_\theta(X_1,\dots,X_n)}{\sup_{\theta\in\Theta_1} L_\theta(X_1,\dots,X_n)} = \frac{\sup_{\theta\leq 5} e^{-n\,\theta}\,\theta^{\sum_{j=1}^n X_j}}{\sup_{\theta>5} e^{-n\,\theta}\,\theta^{\sum_{j=1}^n X_j}}.$$

Si on considère la fonction $f(x) = e^{-nx}x^p$ avec $p \geq 0$, alors $f'(x) = e^{-nx}x^{p-1}(p - nx)$, donc $f'(x) = 0$ pour $x = p/n$, et $f'(x) > 0$ si $x < p/n$, $f'(x) < 0$ si $x > p/n$: la fonction atteint un maximum local et global en $x = p/n$. En conséquence, $L_\theta(X_1,\dots,X_n)$ est maximisée sur $\mathbf{R}$ en $\overline{X}_n$. Donc:

- Si $\overline{X}_n \leq 5$, alors $\sup_{\theta\leq 5} e^{-n\,\theta}\,\theta^{\sum_{j=1}^n X_j} = e^{-n\,\overline{X}_n}\,\overline{X}_n^{\sum_{j=1}^n X_j}$ et $\sup_{\theta>5} e^{-n\,\theta}\,\theta^{\sum_{j=1}^n X_j} = e^{-5\,n}\,5^{\sum_{j=1}^n X_j}$. D'où:

$$\widehat{T} = \frac{e^{-n\,\overline{X}_n}\,\overline{X}_n^{\sum_{j=1}^n X_j}}{e^{-5\,n}\,5^{\sum_{j=1}^n X_j}} = e^{5n}\Big(\frac{\overline{X}_n}{5\,e}\Big)^{n\,\overline{X}_n}.$$

- Si $\overline{X}_n > 5$, on doit inverser la statistique précédente, et
$$\widehat{T} = e^{-5n}\Big(\frac{\overline{X}_n}{5\,e}\Big)^{n\,\overline{X}_n}.$$
Or la fonction $x \in [0,5] \mapsto \left(\frac{x}{5e}\right)^{-nx}$ est décroissante pour $x \leq 5$. Donc la région critique $\widehat{T} < s$ est obtenue pour $\overline{X}_n > s'$. Et l'inverse, la fonction $x \in [5,\infty[\mapsto \left(\frac{x}{5e}\right)^{-nx}$ est aussi décroissante pour $x > 5$, donc la région critique est aussi obtenue pour $\overline{X}_n > s'$.

Dans les 2 cas, on obtient pour règle de décision: H_0 choisie si $\overline{X}_n \leq s'$ et H_1 choisie si $\overline{X}_n > s'$.

(a) Dans la position de la direction, on veut $\sup \mathbb{P}_{\theta\in\Theta_0}(\text{Choisir } H_1) = \sup \mathbb{P}_{\theta\leq 5}(\overline{X}_n > s'_\alpha) = \alpha$. Le sup étant atteint en $\theta = 5$, et comme si $X_1 \overset{\mathcal{L}}{\sim} \mathcal{P}(5)$ alors $\sum_{k=1}^n X_k \overset{\mathcal{L}}{\sim} \mathcal{P}(5n)$, cela revient à trouver s'_α vérifiant $n\,s'_\alpha = q_n$ où q_n est le quantile d'ordre $1-\alpha$ d'une loi de Poisson de paramètre $5\,n$.
Si $\alpha = 0.05$ et $n = 1$, $s'_\alpha = q_1 = 9$. Donc H_0 est choisie.
Si $\alpha = 0.05$ et $n = 100$, $s'_\alpha = q_{100}/100 = 5.37$. Donc H_1 est choisie.

(b) Dans la position des syndicats, on veut que $\sup \mathbb{P}_{\theta\in\Theta_1}(\text{Choisir } H_0) = \sup \mathbb{P}_{\theta>5}(\overline{X}_n < s'_\beta) = \beta$. Le sup étant encore atteint en $\theta = 5$, $n\,s'_\beta = (q'_n)^-$, où q'_n est le quantile d'ordre $\beta = 0.05$ pour une loi de Poisson de paramètre $5\,n$.
Si $\beta = 0.05$ et $n = 1$, $s'_\beta = (q'_1)^- = 2^-$. Donc H_1 est choisie.
Si $\beta = 0.05$ et $n = 100$, $s'_\alpha = (q'_{100})^-/100 = 4.64$. Donc H_1 est choisie.

□