

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->

Feuille n° 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 $x \geq 1$, $F_X(x) = 1$. 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 $\text{var}(X) = 1/12$.

(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 $\text{var}(Y) = 1$.

(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 $z \in [0.5, 1]$, $F_Z(z) = 0.5 + \mathbb{P}(0.5 \leq Z \leq z) = 0.5 + \mathbb{P}([0.5, z]) = 0.5 + (z - 0.5) = z$.

$$\mathbb{E}[Z] = \int_{[0, 1/2]} 0 \, d\omega + \int_{[1/2, 1]} \omega \, d\omega = 0 + [\omega^2/2]_{1/2}^1 = 3/8.$$

$$\text{Et } \mathbb{E}[Z^2] = \int_{[0, 1/2]} 0 \, d\omega + \int_{[1/2, 1]} \omega^2 \, d\omega = 0 + [\omega^3/3]_{1/2}^1 = 7/24, \text{ d'où } \text{var}(Z) = 7/24 - 9/64 = 29/192 \simeq 0.151.$$

□

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 :

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

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

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 $\text{var}(X) = 2$.

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:

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

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

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

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

Pour une v.a. continue, $F_X(x) = \mathbb{P}(X \leq x) = 1 - \mathbb{P}(X > x) = 1 - \mathbb{P}(-X < -x) = 1 - F_X(-x)$ 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 $x \rightarrow x f_X(x)$ est impaire. \square

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. \square

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 $f_Y(y) = \frac{1}{\sqrt{2\pi}} \frac{1}{y} e^{-\ln^2(y)/2}$ 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 $Y : \Omega \rightarrow]0, \infty[$ et Y v.a. car $x \in \mathbf{R} \mapsto e^x$ est une fonction continue donc mesurable (borélienne). Donc pour $y \leq 0$, $F_Y(y) = 0$. Et pour $y > 0$,

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

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 $y > 0$,

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

- (b) Pour tout $a \in [-1, 1]$, il est clair que $f_a(y)$ est mesurable positive, et son intégrale existe car $f_a \leq (1 + |a|)f_Y$. 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 caractérisent pas la loi, puisque clairement Y_a et Y n'ont pas la même loi si $a \neq 0$. □

6. (*) Soit $X \stackrel{\mathcal{L}}{\sim} \mathcal{E}(\lambda)$ 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 $\mathbb{P}(Y = k) = (1 - e^{-\lambda k}) - (1 - e^{-\lambda(k-1)}) = (1 - e^{-\lambda})e^{-\lambda(k-1)}$. Ainsi $\mathbb{P}(Y = k) = (1 - e^{-\lambda})(e^{-\lambda})^{k-1}$ pour $k \geq 1$: $Y \stackrel{\mathcal{L}}{\sim} \mathcal{G}(1 - e^{-\lambda})$: 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} .

- Montrer F_X est une fonction admettant une application réciproque sur $]0, 1[$, notée F_X^{-1} .
- Démontrer que la loi de la variable $F_X^{-1}(U)$ est la même que celle de X .
- A partir de la touche **RAND** d'une calculatrice, comment obtenir une réalisation d'une variable aléatoire de loi exponentielle de paramètre 3?
- 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 $\lim_{x \rightarrow -\infty} F_X(x) = 0$ et $\lim_{x \rightarrow \infty} F_X(x) = 1$, on en déduit que $F_X : \mathbf{R} \rightarrow]0, 1[$. 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 utilisant le fait que $F_X(F_X^{-1}(U)) = U$ et F_X strictement croissante,

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

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 $F_X(x) = 1 - e^{-3x}$, 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:

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

□

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 $z \in [-1, 1]$,

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

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. \square

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

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

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 $X \stackrel{\mathcal{L}}{\sim} \mathcal{P}(\lambda)$, alors pour $u \in \mathbf{R}$,

$$\phi_X(u) = e^{-\lambda} \sum_{k=0}^{\infty} \frac{1}{k!} \lambda^k e^{i u k} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{1}{k!} (\lambda e^{i u})^k = e^{-\lambda} e^{\lambda e^{i u}} = e^{\lambda(e^{i u} - 1)}.$$

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

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

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

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

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

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

avec le changement de variable $y = (\beta - i u)x$, soit $dy = (\beta - i u)dx$.

f/ Si $X \stackrel{\mathcal{L}}{\sim} \mathcal{B}(n, p)$, alors pour $u \in \mathbf{R}$,

$$\phi_X(u) = \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} e^{i u k} = (1-p)^n \sum_{k=0}^n \binom{n}{k} \left(\frac{p e^{i u}}{1-p} \right)^k = (1-p)^n \left(1 + \frac{p e^{i u}}{1-p} \right)^n = (1 + p(e^{i u} - 1))^n,$$

en utilisant la formule du binôme.

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 $\mathcal{N}(m+m', \sigma^2 + \sigma'^2)$. \square

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:

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

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

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^0 e^{u-iux} du + \frac{1}{2\pi} \int_0^{\infty} e^{-u-iux} du = \frac{1}{2\pi} \left(\left[\frac{e^{u-iux}}{1-ix} \right]_{-\infty}^0 + \left[\frac{e^{-u-iux}}{-1-ix} \right]_0^{\infty} \right) = \frac{1}{\pi} \frac{1}{1+x^2}.$$

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

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

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

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

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

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

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

- (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 $(1-\lambda) \mathbb{E}[X] \leq \mathbb{E}[X \mathbb{I}_{\{X > \lambda \mathbb{E}[X]\}}]$. Comme $\mathbb{E}[X] \geq 0$ et $\lambda \in [0, 1]$, alors $(1-\lambda) \mathbb{E}[X] \geq 0$. Donc

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

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

Feuille n° 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):

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

(b) On a

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

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 $\text{cov}(U, V)$. Les variables U et V sont-elles indépendantes?
 (c) Que vaut $\mathbb{E}[|X_1 - X_2|]$?

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 $f_V(v) = 2v \mathbb{I}_{v \in [0, 1]}$.

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 $\text{cov}(U, V) = \frac{1}{4} - \frac{2}{9} = \frac{1}{36}$.

Les variables ne sont pas indépendantes car $\text{cov}(U, V) \neq 0$.

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

□

3. (***) On considère $X = (X_1, X_2)$ un vecteur aléatoire à valeurs dans \mathbf{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 \mathbf{R}^2 .

- (a) Montrer alors que la loi de X_1 admet une densité f_1 par rapport à la mesure de Lebesgue λ_1 sur \mathbf{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 $x \in \mathbf{R}$,

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

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 $L_1 = XL$ 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 $g : \mathbf{R}^2 \rightarrow \mathbf{R}_+$ mesurable. Alors

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

où $h(x, \ell) = g(x\ell, (1 - x)\ell)$.

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, $g(l_1, l_2) = \mathbb{I}_{] - \infty, x]}(l_1) \mathbb{I}_{] - \infty, y]}(l_2)$, 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 de 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,

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

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,

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

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 $g : \mathbf{R}_+ \rightarrow \mathbf{R}_+$ 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

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

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 $f(y) = \mathbb{I}_{[0, +\infty[}(y) \lambda^2 y e^{-\lambda y}$, 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 $g : \mathbf{R}_+ \rightarrow \mathbf{R}_+$ une fonction test mesurable,

$$\begin{aligned} \mathbb{E}[g(Z)] &= \int_0^\infty f(\ell) \left(\int_0^1 g(\min(u, 1-u)\ell) du \right) d\ell \\ &= \int_0^\infty f(\ell) \left(\int_0^{\frac{1}{2}} g(\ell u) du + \int_{\frac{1}{2}}^1 g(\ell(1-u)) du \right) d\ell. \end{aligned}$$

On fait un changement de variables dans la deuxième expression: $1-u = v$, donc $v \in [0, \frac{1}{2}]$ et

$$\int_{\frac{1}{2}}^1 g(\ell(1-u)) du = \int_0^{\frac{1}{2}} g(\ell v) dv.$$

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

$$\begin{aligned} \mathbb{E}[g(Z)] &= 2 \int_0^\infty f(\ell) \left(\int_0^{\frac{1}{2}} g(\ell u) du \right) d\ell = 2 \int_0^\infty \frac{f(\ell)}{\ell} \left(\int_0^{\ell/2} g(y) dy \right) d\ell \\ &= 2 \int_0^\infty g(y) \left(\int_{2y}^\infty \frac{f(\ell)}{\ell} d\ell \right) dy. \end{aligned}$$

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

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

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, \dots, X_n) un échantillon de n v.a.i.i.d. de loi exponentielle de paramètre 1.

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

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

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

Proof. (a) $\mathbb{P}(\exists(i, j) \in \{1, \dots, n\}^2, i \neq j, X_i = X_j) = \int_D f(x)f(y) d\lambda_2(x, y)$, 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 $z < 0$, $F_Z(z) = \mathbb{P}(Z \leq z) = 0$. Pour $z \geq 0$,

$$F_Z(z) = 1 - \mathbb{P}(X_1 > z \cap \dots \cap X_n > z) = 1 - \left(\int_z^\infty e^{-x} dx \right)^n = 1 - e^{-nz}.$$

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

(c) Si $N = \min\{1 \leq i \leq n, X_i = Z\}$, 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 $\min_{1 \leq i \leq n} (Y_i)$ 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.
- $\mathbb{P}(N = k, Z > t) = \mathbb{P}(\min_{i \neq k} X_i \geq X_k, X_k > t)$. 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:

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

De ceci, on en déduit que $\mathbb{P}(N = k | Z > t) = \mathbb{P}(N = k \cap Z > t) / \mathbb{P}(Z > t) = \frac{1}{n}$ et ceci pour tout $k = 1, \dots, n$ et tout $t > 0$: les deux variables sont indépendantes.

Et $\mathbb{P}(N = k, Z > t) = \frac{1}{n} e^{-nt}$ pour tout k : N suit une loi uniforme sur $\{1, \dots, n\}$.

□

7. (***) Soient X_1, \dots, 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 $U \stackrel{\mathcal{L}}{\sim} \Gamma(\alpha_1, \beta)$ et $V \stackrel{\mathcal{L}}{\sim} \Gamma(\alpha_2, \beta)$. Quelle est la loi de $U + V$?

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

(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 reconnaît 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 $\phi(u) = (1 - i \frac{u}{\beta})^{-\alpha}$. D'où $\phi_{Z_1+Z_2}(u) = \phi_{Z_1}(u) \phi_{Z_2}(u) = (1 - i \frac{u}{\beta})^{-\alpha_1} (1 - i \frac{u}{\beta})^{-\alpha_2} = (1 - i \frac{u}{\beta})^{-\alpha_1 - \alpha_2}$ 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 $W_1 \stackrel{\mathcal{L}}{\sim} \Gamma(1, 1)$ et donc $W_1 + W_2 + \dots + W_n \stackrel{\mathcal{L}}{\sim} \Gamma(n, 1)$.

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

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

avec F_Y la fonction de répartition de Y . La densité de $X_1 \cdot X_2 \cdot \dots \cdot X_n$ 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 $\Gamma(n, 1)$. □

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

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

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

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

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 $\frac{\alpha \beta}{\alpha + \beta} (e^{-\alpha x} + e^{-\beta x}) \mathbb{I}_{x \geq 0}$. □

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

- (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:

$$\begin{cases} 2x + y + 3z = 0 \\ x + 5y + 6z = 0 \end{cases} \iff \begin{cases} x = y \\ z = -y \end{cases}.$$

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 $\text{var}(X_3 - \alpha_1 X_1 - \alpha_2 X_2) = 0$, 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, \dots, f_d) famille orthonormée de $d - 1$ vecteurs de \mathbf{R}^d telle que $(\frac{x}{\|x\|}, f_2, \dots, f_d)$ soit une base orthonormale de \mathbf{R}^d (on a $\|x\| > 0$ car x non nul). Ainsi, pour $u \in \mathbf{R}$, on a

$$\begin{aligned} \mathbb{P}(U \leq u) &= \int_{{}^t x z \leq u} f\left({}^t x z \frac{x}{\|x\|^2} + \sum_{j=2}^d \langle f_j, z \rangle f_j\right) d\lambda_d(z) = \int_{\|x\|z'_1 \leq u} f\left(z'_1 \frac{x}{\|x\|} + \sum_{j=2}^d z'_j f_j\right) dz'_1 \dots dz'_d \\ &= \int_{z'_1 \leq \frac{u}{\|x\|}} \left(\int_{\mathbf{R}^{d-1}} f\left(z'_1 \frac{x}{\|x\|} + \sum_{j=2}^d z'_j f_j\right) dz'_2 \dots dz'_d \right) dz'_1 \end{aligned}$$

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

$$f_x(z'_1) = \int_{\mathbf{R}^{d-1}} f\left(z'_1 \frac{x}{\|x\|} + \sum_{j=2}^d z'_j f_j\right) dz'_2 \dots dz'_d$$

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 covariance.

□