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.

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

3. **(*)** Soit une variable aléatoire X sur l'espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$. On suppose que pour tout $\omega \in \Omega$, $-\omega \in \Omega$ et également que la 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 $\mathbb{E}(X) = 0$.

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.

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.

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 $X + 1$)

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 $F_X^{-1}(U)$?

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?

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.

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 $\phi(u) = e^{-|u|}$.

11. **(***)** Soit X une variable aléatoire réelle intégrable telle que $\mathbb{E}[X] \geq 0$.

(a) Montrer que pour tout $\lambda > 0$, $X \leq \lambda\mathbb{E}[X] + X\mathbb{1}_{\{X>\lambda\mathbb{E}[X]\}}$.

(b) On suppose que, de plus, $0 < \mathbb{E}[X^2] < +\infty$. Montrer que

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

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

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

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?

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 $\mathbb{E}\big[|X_1 - X_2|\big]$?

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 $\mathbb{R}^2$?

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 $Z = \min\{L_1, L_2\}$.

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.

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 $Z = \min_{1 \leq i \leq n} X_i$. 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 $\mathbb{P}(N = k,\ Z > t) = \dfrac{e^{-nt}}{n}$ 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.

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 $\prod_{i=1}^n X_i$.

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.

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. $U = x^t Z$ a une densité sur $\mathbf{R}$.

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

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 $(X_1 - X_2, X_3 - X_1)$?

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.

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 $Z/\sqrt{W}$.

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.

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.

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

   (a) Démontrer que $\mathbb{E}\big[e^{<s,X>}\big] = e^{\frac{1}{2}\,{}^t s\,\Gamma\, s}$ pour tout $s \in \mathbf{R}^4$. En utilisant l'unicité du développement en série entière, en déduire que $\mathbb{E}\big[< s, X >^4 \big] = 3\left({}^t s\,\Gamma\, s\right)^2$. En déduire 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) Déduire également que $\mathbb{E}[X_1X_2X_3] = 0$.

(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 $\mathrm{cov}(X^2, Y^2)$.

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 $\mathbb{L}^p$?

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 $\theta \neq 1/2$.

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.

3. **(**)** 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 $\alpha_n = 1/n^2$.

4. **(**)** 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 $X_n \xrightarrow[n\to\infty]{p.s.} 0$ si et seulement si $X_n \xrightarrow[n\to+\infty]{\mathcal{P}} 0$. 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 $Y_n \xrightarrow[n\to\infty]{p.s.} 0$ si et seulement si $Y_n^* \xrightarrow[n\to+\infty]{\mathcal{P}} 0$.

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 $X_n \xrightarrow[n\to+\infty]{\mathcal{P}} -\infty$. La suite (X_n) converge-t-elle dans $\mathbb{L}^2$?

6. **(**)** On suppose $X_n \xrightarrow[n\to\infty]{\mathcal{L}} c$ 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 $X_n \xrightarrow[n\to+\infty]{\mathcal{P}} c$.

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.

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

9. **(***)** Soit $(X_n)_n$ une suite de v.a.i.i.d. centrées de variance commune 1. Soit $(a_{i,n})_{1\leq i\leq n,\, n\in\mathbf{N}}$ 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 $\max_{1\leq i\leq n}|a_{i,n}| \underset{n\to\infty}{\longrightarrow} 0$, 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 $|z_j| \leq 1$ et $|z'_j| \leq 1$ pour tout j, alors
   $$\Big|\prod_{j=1}^n z_j - \prod_{j=1}^n z'_j\Big| \leq \sum_{j=1}^n |z_j - z'_j|.$$

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

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).

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?

2. **(**)** Soit un n-échantillon $(X_1, \dots, X_n)$ de v.a.i.i.d. suivant une loi de densité $f_\theta(x) = (\theta+1)\, x^\theta\, \mathbb{I}_{0<x\leq 1}$ 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.

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
   $$R(\widehat{\theta}) \leq 2\, \frac{R(\widehat{\theta}_1)\, R(\widehat{\theta}_2)}{R(\widehat{\theta}_1) + R(\widehat{\theta}_2)}.$$

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.

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 $\widetilde{\theta}$?

6. **(**)** 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 $p \in]0, 1[$ 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.

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 $(X_1, \dots, X_n)$ du modèle statistique et on pose $\widehat{\theta}_n^{(1)} = \frac{1}{2} \max(X_1, \dots, X_n)$ et $\widehat{\theta}_n^{(2)} = \min(X_1, \dots,$ 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.

8. **(**)** Soit X une v.a. suivant une loi binomiale $\mathcal{B}(n, \frac{1}{\lambda})$ 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.

9. **(**)** Soit un n-échantillon de v.a.i.i.d. $(X_1, \dots, X_n)$ 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%.

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 $(X_1 \dots, X_n)$ un échantillon observé de ce modèle.

(a) Soit l'estimateur $\widehat{T}_n = (\overline{X}_n, \overline{\sigma}_n^2)$, où $\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$ et $\overline{\sigma}_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2$. En utilisant le théorème de Cochran (on considérera le sous-espace engendré par le vecteur de $\mathbf{R}^n$, $(1, \dots, 1)'$), 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.

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

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

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

12. **(***)** Soit un n-échantillon $(X_1, \ldots, X_n)$ 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?