Université Paris I, Panthéon - Sorbonne

Master M.A.E.F. 2017 – 2018

Feuilles de TD, cours de Statistique 1

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:

Révisions sur les variables et vecteurs aléatoires

1. **(*)** On considère deux v.a. X et Y définies sur le même espace de probabilité $(\Omega, \mathcal{A}, P)$. Supposons que X et Y ont la même loi. Est-t-il nécessairement vrai que $\forall \omega \in \Omega, \quad X(\omega) = Y(\omega)$?

2. **(*)** L'éclairage d'une pièce se compose de deux néons montés en série (ce qui signifie que chaque ampoule ne fonctionne que lorsque les deux fonctionnent). La durée de vie de chaque néon suit une loi de exponentielle de moyenne un an. On note T la durée de vie du système d'éclairage. Donner la loi de T, son espérance et sa variance.

3. **(*)** Un point est choisi au hasard sur un segment de longueur L. Interpréter cet énoncé et trouver la probabilité que le rapport entre le plus petit et le plus grand segment soit inférieure à $1/4$.

4. **(*)** La taille moyenne de 500 élèves de collège est 151 cm et l'écart-type 15 cm. En supposant que la taille soit distribuée suivant une loi normale, trouver combien d'élèves ont leur taille comprise entre (a) 120 et 155 cm (b) 185 cm et plus.

5. **(*)** La durée de vie d'un composant électrique peut être représentée par une variable aléatoire réelle X de densité de probabilité f définie par

$$f(x) = \begin{cases} kx(5-x) & \text{si } 0 \leq x \leq 5 \\ 0 & \text{sinon} \end{cases}$$

   (a) Calculer la valeur de k. Représenter graphiquement la fonction de répartition F_X de X.

   (b) Calculer $\mathbb{E}(X)$ et $\text{var}(X)$. Déterminer la médiane de X, c'est-à-dire le réel m tel que $P(X \leq m) = 0.5$.

6. **(**)** Sur $(\Omega, \mathcal{A}, P)$ un espace de probabilité, 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 :

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

7. **(*)** Soit $X \sim \text{U}([0,1])$ (loi uniforme sur $[0,1]$). Trouver la loi de $Y = \exp(X)$ et de $Z = X^{1/n}$, $n \in N$.

8. **(***)** 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 $f_Y(y) = \frac{1}{\sqrt{2\pi}}\frac{1}{z}e^{-\ln^2(z)/2}$ 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 les mêmes moments, et en déduire que les moments ne caractérisent pas une loi de probabilité.

9. **(*)** Soit X de loi de Poisson $\mathcal{P}(\lambda)$ et Y de loi de Poisson $\mathcal{P}(\mu)$. En supposant que X et Y sont indépendantes, donner la loi de la somme $X + Y$, en justifiant la réponse par les calculs.

10. **(**)** 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 $P(Y = 1) = 0.5$. On considère la v.a. $Z = X \cdot Y$. Déterminer la mesure de probabilité de Z, puis celle de $X + Z$. En déduire que la somme de 2 variables gaussiennes peut ne pas être gaussienne.

11. **(**)** Soit X et Y deux v.a. indépendantes définies sur le même espace de probabilité (Ω, P), X suivant une loi de Bernoulli de paramètre p, et Y suivant une loi de Poisson de paramètre λ. On pose pour tout $\omega \in \Omega$,

$$Z(\omega) = \begin{cases} 0 & \text{si} \quad X(\omega) = 0 \\ Y(\omega) & \text{si} \quad X(\omega) = 1 \end{cases}$$

   (a) Montrer que pour tout $\omega \in \Omega \quad Z(\omega) = X(\omega)Y(\omega)$.

   (b) Calculer l'espérance et la variance de Z.

12. **(*)** Soit X et Y deux variables définies sur le même espace de probabilité $(\Omega, \mathcal{A}, P)$ et absolument continues par rapport à la mesure de Lebesgue sur $\mathbb{R}$ (de densité f_X et f_Y). Déterminer la densité de $X + Y$ lorsque X et Y sont indépendantes. La calculer explicitement dans les cas où X et Y sont des v.a. uniformes sur [0,1], puis des lois exponentielles de paramètres respectifs λ et μ.

13. **(**)** Soit X et Y deux variables aléatoires telles que le vecteur $Z = (X, Y)$ soit absolument continu par rapport à la mesure de Lebesgue sur $\mathbb{R}^2$, de densité f_Z telle que :

$$f_Z(x, y) = k \cdot \frac{|x|}{(x^2 + |y| + 1)^4} \quad \text{pour} \quad (x, y) \in \mathbb{R}^2.$$

   Déterminer la valeur de k. Déterminer la loi de X et celle de Y. Déterminer leurs espérances, leurs variances et $\text{cov}(X, Y)$. Les variables X et Y sont-elles indépendantes ?

14. **(**)** Soit $(\varepsilon_n)_{n\in\mathbb{N}}$ une suite de v.a.i.i.d. gaussiennes centrées réduites. Soit $a \in]-1, 1[$. On considère la suite de v.a. $(X_n)_{n\in\mathbb{N}}$ telle que :

   (a) $X_0 = 0$ et $X_{n+1} = a \cdot X_n + \varepsilon_{n+1}$ pour $n \in \mathbb{N}$. Déterminer la loi de X_n, puis celle de $(X_1, \cdots, X_n)$. Les variables X_i sont-elles indépendantes ?

   (b) $X_0 \sim \mathcal{N}(0, (1 - a^2)^{-1})$ et $X_{n+1} = a \cdot X_n + \varepsilon_{n+1}$ pour $n \in \mathbb{N}$. Déterminer la loi de X_n, puis celle de $(X_1, \cdots, X_n)$. Les variables (X_i) sont-elles indépendantes ?

   (c) $X_0 = 0$ et $X_{n+1} = \varepsilon_{n+1} - \varepsilon_n$ pour $n \in \mathbb{N}$. Déterminer la loi de X_n, puis celle de $(X_1, \cdots, X_n)$. Les variables (X_i) sont-elles indépendantes ?

15. **(***)** 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 Lebesgue 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 $\mathbb{R}^2$?

16. **(**)** Soit $(X)_n$ une suite des v.a.i.i.d. de loi de Bernoulli $\mathcal{B}(p)$ avec $0 < p < 1$. Pour $n \geq 1$, on pose $Y_n = X_n \cdot X_{n+1}$.

   (a) Donner la loi de Y_n. Calculer $E(Y_n)$ et $\mathrm{Var}(Y_n)$.

   (b) Calculer $\mathrm{Cov}(Y_n, Y_{n+1})$ et $\mathrm{Cov}(Y_n, Y_m)$ pour $m \neq n+1$.

   (c) Pour $n \geq 1$, soit $S_n = \frac{1}{n} \sum_{k=1}^{n} Y_k$. Calculer $E(S_n)$ et $\mathrm{Var}(S_n)$.

   (d) Soit $\varepsilon > 0$. En utilisant l'inégalité de Tchébychev, montrer que si $n \to \infty$,

$$P(|S_n - p^2| \geq \varepsilon) \to 0.$$

   Que peut-on en conclure quant à S_n ?

17. **(**)** Soit $(X_1, \cdots, X_n)$ un échantillon de loi exponentielle de paramètre 1.

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

   (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 $P(N = K, Z > t) = \dfrac{e^{-nt}}{n}$ 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.

18. **(***)** Soit $(X_n)_{n \in \mathbb{N}}$ une suite de v.a. indépendantes identiquement distribuées suivant une loi de Bernoulli de paramètre p. Pour tout $n \geq 1$ on définit par récurrence $T_n = \inf\{k > T_{n-1}, X_k = 1\}$ et $T_0 = 0$.

   (a) Montrer que pour $n \in \mathbb{N}$, T_n est une v.a.

   (b) Montrer que $(T_{i+1} - T_i)_{i \in \mathbb{N}}$ forme une suite de v. a. indépendantes et identiquement distribuées.

   (c) Calculer la loi de T_1 et sa fonction caractéristique.

   (d) En déduire la loi de T_n.

Feuille n^o 2 : **Révisions sur l'espérance conditionnelle et les théorèmes limites**

1. **(**)** Sur $(\Omega, \mathcal{A}, \mu)$ un espace probabilisé, on considère une v.a. réelle positive X de fonction de répartition F_X. Montrer que pour $n \in \mathbb{N}^*$:
$$\mathbb{E}X^n = \int_0^\infty nt^{n-1}(1 - F_X(t))dt = \int_0^\infty nt^{n-1}P(X > t)dt.$$
Montrer que l'hypothèse X positive est nécessaire.

2. **(*)** Soit X et Y des variables aléatoires indépendantes appartenant à $\mathbb{L}^1(\Omega, \mathcal{A}, P)$. Comparer les lois des couples $(X, X+Y)$ et $(Y, X+Y)$. En déduire que $\mathbb{E}(X \mid X+Y) = \mathbb{E}(Y \mid X+Y) = \frac{1}{2} \cdot (X+Y)$.

3. **(*)** On jette indépendamment deux dés, dont les résultats sont X_1 et X_2. Quelle est la loi de X_1 sachant que $X_1 + X_2$ est paire ?

4. **(*)** Soit X une variable aléatoire réelle et soit $c \in \mathbb{R}$ une constante. Déterminer la loi de X sachant $\min(X, c)$, puis, en supposant que X est $\mathbb{L}^1(\Omega, \mathcal{A}, P)$, déterminer $\mathbb{E}(X \mid \min(X, c))$.

5. **(**)** Soit X une variable aléatoire suivant une loi exponentielle de paramètre $\beta > 0$. Démontrer que
$$P(X > t + s \mid X > t) = P(X > s), \quad \text{pour tout } (s,t) \in \mathbb{R}_+^2.$$
Prouver que pour tout $t \in \mathbb{R}_+$, $\lim_{h \to 0} \frac{1}{h} \cdot P(t < X < t+h \mid X > t) = \beta$.

6. **(**)** Soient X et Y deux variables telles que $(X+Y)$ soit un vecteur gaussien non dégénéré. Déterminer $\mathbb{E}(X \mid Y)$. Généraliser au cas où Y est un vecteur gaussien de $\mathbb{R}^n$ et de matrice de covariance Σ_Y.

7. **(**)** Soit la chaîne de Markov $(X_n)_{n \in \mathbb{N}}$ à valeur dans $\{x_1, x_2\}$ et de matrice de transition $Q = (q_{ij})_{1 \leq i,j \leq 2}$. Rappeler la définition de (X_n). Déterminer la prédiction optimale au sens des moindres carrés de X_{n+1} et X_{n+2} connaissant $(X_0, \cdots, X_n)$, soit $\mathbb{E}(X_{n+1} \mid (X_0, \cdots, X_n))$ et $\mathbb{E}(X_{n+2} \mid (X_0, \cdots, X_n))$.

8. **(**)** Soit $(X_n)_{n \in \mathbb{Z}}$ une suite de variables aléatoires définie par
$$X_n = \varepsilon_n - \theta \varepsilon_{n-1}, \quad \text{pour } n \in \mathbb{Z},$$
avec $|\theta| < 1$ et (ε_n) une suite de v.a.i.i.d. centrées et de variance σ^2. Montrer que le prédicteur linéaire optimal $\widehat{X}_{n+1}$ au sens des moindres carrés de X_{n+1} à partir de $\{X_{n-j}, j \in \mathbb{N}\}$ est :
$$\widehat{X}_{n+1} = \mathbb{E}(X_{n+1} \mid (X_{n-j}, j \in \mathbb{N})) = -\sum_{j=1}^{\infty} \theta^j X_{n+1-j}.$$

9. **(**)** Dans le modèle de l'exercice précédent, on suppose que X_1, X_2, X_4 et X_5 sont connus, mais pas X_3. Déterminer le prédicteur linéaire optimal $\widehat{X}_3$ au sens des moindres carrés de X_3 à partir de (X_1, X_2), puis à partir de (X_4, X_5) et enfin à partir de (X_1, X_2, X_4, X_5).

10. **(***)** Répondre aux mêmes questions dans le cas du modèle
$$X_n = \phi X_{n-1} + \varepsilon_n, \quad \text{pour } n \in \mathbb{Z},$$
où $\phi \in \mathbb{R}$ et (ε_n) une suite de v.a.i.i.d. centrées de variance σ^2.

11. **(**)** 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 $0) \iff$ (la suite $(P(X_n > 0))$ tend vers 0).

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

(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 $\alpha_n = 1/n^2$.

12. **(**)** Soit $(X_n)_{n\in\mathbb{N}}$ une suite de v.a. non corrélées et dans $\mathbb{L}^2$ à valeurs dans $\mathbb{R}$. On suppose qu'il existe des réels m et C tels que pour tout n, $\mathbb{E}(X_n) = m$ et $\mathrm{var} X_n \leq C$. Montrer que la suite des $\bar{X}_n$ converge vers m dans $\mathbb{L}^2$ et en probabilité.

13. **(***)** Soit $(X_n)_{n\in\mathbb{N}}$ une suite de v.a.i.i.d., dans $\mathbb{L}^2$ et à valeurs dans $\mathbb{R}$. Soit $Y_k = \bar{X}_k$ pour $k \geq 1$. Peut-on obtenir la loi faible des grands nombres pour la suite (Y_k) ? La loi forte des grands nombres? Un théorème de limite centrale ? Traiter ensuite le cas particulier où les X_k suivent une loi normale centrée réduite.

14. **(*)** Trouver la probabilité d'obtenir 200 fois 1 en lançant 1000 fois un dé honnête.

15. **(*)** On sait que 3% des réservations de places d'avions ne sont pas honorées le jour du départ. Pour un avion de 400 places, combien de réservations au maximum peuvent être acceptées par une compagnie aérienne pour que la probabilité que des personnes ayant réservé n'aient pas de place le jour du départ soit inférieure à 5% ?

16. **(*)** Soit $(X_1, \cdots, X_n)$ un échantillon de v.a.i.i.d. de même loi ayant pour densité par rapport à la mesure de Lebesgue :
$$f(x) = \exp(-x + \theta) \cdot \mathbb{I}_{x \geq \theta},$$
où θ est un nombre réel donné. On définit $m_n = \min\{X_1, \cdots, X_n\}$. Déterminer la fonction de répartition, la densité ainsi que l'espérance de m_n. En utilisant l'inégalité de Markov, vérifier que m_n converge en probabilité vers θ.

17. **(*)** Soit $(X_1, \cdots, X_n)$ un échantillon de v.a.i.i.d. de même loi exponentielle de paramètre β. On pose
$$T_n = \frac{n}{\sum_{i=1}^{n} X_i}.$$
Montrer que $\sqrt{n}(T_n - \beta)$ converge en loi vers une loi normale dont précisera l'espérance et la variance.

18. **(**)** Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé et soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires indépendantes sur $(\Omega, \mathcal{A}, P)$ et de même loi P_X (à laquelle est associée la fonction de répartition F_X). Pour tout $n \in \mathbb{N}^*$, on définit la fonction de répartition empirique de $(X_1, \cdots, X_n)$ par :
$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}_{X_i \leq x} \quad \text{pour} \quad x \in \mathbb{R}.$$

(a) Rappeler l'expression de F_X lorsque P_X est une loi binomiale de paramètre $(3, 1/3)$ et lorsque P_X est la loi gaussienne centrée réduite.

(b) Montrer que pour tout $x \in \mathbb{R}$, $F_n(x)$ est une variable aléatoire sur $(\Omega, \mathcal{A}, P)$ à valeurs dans $[0, 1]$. Pour tout $n \in \mathbb{N}^*$ et tout $x \in \mathbb{R}$, calculer $\mathbb{E} F_n(x)$ en fonction de F_X.

(c) Démontrer que pour tout $x \in \mathbb{R}$, $F_n(x) \xrightarrow[n\to+\infty]{\mathcal{P}} F(x)$.

(d) Montrer que pour tout $x \in \mathbb{R}$, $F_n(x)$ vérifie un théorème de limite centrale que l'on établira.

(e) On suppose que les $(X_n)_{n\in\mathbb{N}^*}$ ne sont plus indépendants. A partir d'un contre-exemple, montrer que les convergences précédentes n'ont plus lieu en général.

Feuille n^o 3 : **Statistiques exhaustives**

1. **(*)** Soit $((\Omega')^n, \mathcal{A}'_n, (P_\theta)^{\otimes n}, \theta \in \Theta)$ un modèle paramétrique. Dans tous les cas suivants, préciser Ω', $\mathcal{A}'_n$ et la mesure dominante, et vérifier (sans utiliser les résultats sur les modèles exponentiels) si la statistique $\widehat{S}_n = X_1 + \ldots + X_n$ est une statistique exhaustive pour ce modèle :
   (a) P_θ est la loi de Bernoulli de paramètre $\theta = p \in]0,1[$;
   (b) P_θ est la loi exponentielle de paramètre $\theta = \lambda \in]0,+\infty[$;
   (c) P_θ est la loi normale de moyenne $\theta \in \mathbb{R}$ et de variance 1;
   (d) P_θ est la loi normale de paramètre $\theta = (m, \sigma^2) \in \mathbb{R} \times]0,+\infty[$;
   (e) P_θ est la loi de Poisson de paramètre $\theta = (\lambda) \in \mathbb{R} \times]0,+\infty[$;
   (f) P_θ est la loi uniforme sur $[0,\theta]$ où $\theta \in]0,+\infty[$;

2. **(*)** On considère le modèle paramétrique gaussien $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mathcal{N}(0,\sigma^2)^{\otimes n}, \sigma^2 \in]0,+\infty[)$. Montrer (sans utiliser les résultats sur les modèles exponentiels) que la statistique $\widehat{T}_n = X_1^2 + \ldots + X_n^2$ est exhaustive pour ce modèle.

3. **(**)** Soit $((\Omega')^n, \mathcal{A}'_n, (P_\theta)^{\otimes n}, \theta \in]0,\infty[)$ un modèle paramétrique tel que P_θ admette une densité par rapport à la mesure de Lebesgue de densité :
$$f(x) = \frac{1}{\log 2} \cdot \frac{1}{x} \cdot \mathbb{I}_{]\theta,2\theta[}(x).$$
   Préciser Ω' et $\mathcal{A}'_n$. Montrer que la statistique $\widehat{T} = (\min(X_1,\ldots,X_n), \max(X_1,\ldots,X_n))$ est une statistique exhaustive minimale pour ce modèle.

4. **(**)** On considère un n-échantillon de v.a.i.i.d. de loi gamma de paramètres $(a,b) \in]0,\infty[^2$. Préciser le modèle statistique paramétrique induit. Déterminer une statistique exhaustive pour ce modèle. Est-elle minimale ? complète ?

5. **(***)** 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 $P(X_{n+1} = 1 \mid X_n = 1) = P(X_{n+1} = 0 \mid X_n = 0) = p$ et $P(X_{n+1} = 1 \mid X_n = 0) = P(X_{n+1} = 0 \mid X_n = 1) = 1-p$ pour $n \in \mathbb{N}^*$, avec $p \in]0,1[$ un paramètre inconnu. Montrer que la statistique d'ordre $\widehat{T}_n = (X_{(1)}, \cdots, X_{(n)})$ n'est pas une famille exhaustive pour ce modèle paramétrique.

6. **(**)** On considère un n-échantillon de v.a.i.i.d. de loi uniforme sur $]\theta, \theta+1[$, où $\theta \in \mathbb{R}$.
   (a) Préciser le modèle statistique induit.
   (b) Montrer que la statistique $\widehat{T} = (\min(X_1,\ldots,X_n), \max(X_1,\ldots,X_n))$ est une statistique exhaustive minimale pour ce modèle, mais n'est pas complète.
   (c) Soit $X_{(1)} \leq \cdots \leq X_{(n)}$ l'échantillon ordonné. Soit $\widehat{R}_n = X_{(n)} - X_{(1)}$. Montrer que $\widehat{R}_n$ est une statistique libre pour le modèle.

7. **(*)** En reprenant les différents modèles de l'exercice 1., déterminer l'information de Fisher du modèle et comparer la à celle induite par la statistique $\widehat{S}_n = X_1 + \ldots + X_n$.

8. **(**)** Soit un n-échantillon de v.a.i.i.d. de loi admettant la densité f par rapport à la mesure de Lebesgue sur $\mathbb{R}$ telle que :
$$f(x) = \frac{\exp(x-\theta)}{(1+\exp(x-\theta))^2},$$
où $\theta \in \mathbb{R}$ est un paramètre inconnu. Déterminer une statistique exhaustive minimale et complète pour ce modèle et déterminer son information de Fisher.

9. **(***)** Soit un n-échantillon de v.a.i.i.d. de loi admettant la densité f par rapport à la mesure de Lebesgue telle que :
$$f(x) = \frac{1}{\theta} \cdot x^{-1+1/\theta} \cdot \exp(-x^{1/\theta}) \cdot \mathbb{I}_{]0,\infty[}(x),$$
où $\theta \in]0,\infty[$ est un paramètre inconnu. Déterminer une statistique exhaustive minimale et complète pour ce modèle et déterminer son information de Fisher. Montrer que la statistique $\widehat{T}_2 = \dfrac{\log(X_2)}{\log(X_1)}$ est libre. En déduire que $\widehat{T}_n = \dfrac{\prod_{i=2}^n \log(X_i)}{\log(X_1)}$ l'est également.

10. **(***)** Soit $(\varepsilon_i)_{n\in\mathbb{N}}$ une suite de v.a.i.i.d. de loi gaussienne $\mathcal{N}(0,\sigma^2)$, avec $\sigma^2 \in]0,\infty[$ inconnue. On suppose que $(X_1,\cdots,X_n)$ est un n-échantillon tel que :
$$\forall k \in \{1,\cdots,n\}, \quad X_k = \varepsilon_{k+1} - a\cdot\varepsilon_k,$$
où $a \in \mathbb{R}$ est inconnu.
    (a) Montrer que le modèle paramétrique lié à l'échantillon $(X_1,\cdots,X_n)$ (on notera $\theta = (\sigma^2, a)$) appartient à la famille exponentielle.
    (b) Déterminer une statistique exhaustive minimale et complète pour ce modèle.
    (c) Déterminer l'expression de l'information de Fisher pour ce modèle (ne pas faire les calculs).

11. **(***)** Soit $(\varepsilon_i)_{n\in\mathbb{N}}$ une suite de v.a.i.i.d. de loi de Bernoulli de paramètre $p \in]0,1[$ inconnu. Soit $(X_1,\cdots,X_n)$ un n-échantillon tel que :
$$\forall k \in \{1,\cdots,n\}, \quad X_k = \varepsilon_{k+1}\cdot\varepsilon_k.$$
    (a) Montrer que le modèle paramétrique lié à l'échantillon $(X_1,\cdots,X_n)$ appartient à la famille exponentielle.
    (b) Déterminer une statistique exhaustive minimale et complète pour ce modèle.
    (c) Soit la statistique d'ordre $\widehat{R}_n = (X_{(1)},\cdots,X_{(n)})$. Est-ce une statistique exhaustive ?

Feuille n^o 4 : **Estimation paramétrique**

1. **(*)** Soit $((\Omega')^n, \mathcal{A}'_n, (P_\theta)^{\otimes n}, \theta \in \Theta)$ un modèle paramétrique. Dans tous les cas suivants, préciser un estimateur exhaustif complet et sans biais du paramètre θ, déterminer s'il est efficace, et s'il ne l'est pas, donner la fonction du paramètre pouvant être estimé de façon efficace et sans biais :

   (a) P_θ est la loi de Bernoulli de paramètre $\theta = p \in]0,1[$;

   (b) P_θ est la loi exponentielle de moyenne $\theta = \lambda \in]0,+\infty[$;

   (c) P_θ est la loi normale de moyenne $\theta \in \mathbb{R}$ et de variance 1;

   (d) P_θ est la loi normale de moyenne 0 et de variance $\theta = \sigma^2 \in]0,+\infty[$;

   (e) P_θ est la loi normale de moyenne 0 et d'écart-type $\theta = \sigma \in]0,+\infty[$;

   (f) P_θ est la loi de Poisson de paramètre $\theta = \lambda \in]0,+\infty[$;

   (g) P_θ est la loi dont la densité par rapport à la mesure de Lebesgue est $f_\theta(x) = \dfrac{\log\theta}{\theta-1} \cdot \theta^x \cdot \mathbb{I}_{]0,1[}(x)$.

2. **(**)** On considère le modèle paramétrique gaussien $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mathcal{N}(m,\sigma^2)^{\otimes n}, (m,\sigma^2) \in \mathbb{R}\times]0,+\infty[)$.

   (a) En utilisant les résultats sur les modèles exponentiels, montrer que 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 $\overline{\sigma}_n^2 = \dfrac{1}{n-1}\displaystyle\sum_{i=1}^n (X_i - \overline{X}_n)^2$ est exhaustif et complet pour ce modèle.

   (b) En utilisant le théorème de Cochran, montrer que cet estimateur est sans biais. Est-il efficace ? Déterminer une région de confiance à 95% pour (m,σ^2).

   (c) Déterminer l'estimateur du maximum de vraisemblance de (m,σ^2). Est-il biaisé ? Efficace ? Déterminer une région de confiance à 95% pour (m,σ^2).

3. **(**)** Soit $((\Omega')^n, \mathcal{A}'_n, (P_\theta)^{\otimes n}, \theta \in]0,\infty[)$ un modèle paramétrique tel que P_θ admette une densité par rapport à la mesure de Lebesgue de densité :

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

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

   (b) On pose $\widehat{\theta}_n^{(1)} = \dfrac{1}{2}\max(X_1,\dots,X_n)$ et $\widehat{\theta}_n^{(2)} = \min(X_1,\dots,X_n)$. Pour chacun de ces estimateurs, déterminer leur loi et montrer qu'ils convergent. Sans rentrer dans les calculs, donner un court raisonnement montrant qu'ils sont biaisés. Déterminer une région de confiance à 95% pour θ.

4. **(***)** Soit $((\Omega')^n, \mathcal{A}'_n, (P_\theta)^{\otimes n}, \theta \in]0,\infty[)$ un modèle paramétrique tel que P_θ admette une densité par rapport à la mesure de Lebesgue de densité :

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

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

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

   (c) Déterminer un estimateur de la forme $\widehat{\theta}_n = \alpha \cdot \widehat{\theta}_n^{(1)} + (1-\alpha)\cdot\widehat{\theta}_n^{(2)}$, avec $\alpha \in]0,1[$, qui soit sans biais. Cet estimateur a-t-il une variance inférieure à celle des précédents ?

5. **(**)** 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 $P(X_{n+1} = 1 \mid X_n = 1) = P(X_{n+1} = 0 \mid X_n = 0) = p$ et $P(X_{n+1} = 1 \mid X_n = 0) = P(X_{n+1} = 0 \mid X_n = 1) = 1 - p$ pour $n \in \mathbb{N}^*$, avec $p \in]0,1[$ un paramètre inconnu. Déterminer l'estimateur de maximum de vraisemblance de p. Est-il biaisé ? Convergent ?

6. **(*)** On considère un n-échantillon $(X_1, \ldots, X_n)$ de v.a.i.i.d. suivant la loi binomiale $\mathcal{B}(p,k)$, où $p \in]0,1[$ est inconnu alors que $k \in \mathbb{N}^*$ est connu. On voudrait estimer la probabilité $P(X_1 = 1)$. Montrer que c'est une fonction de p, soit $g(p)$. Déterminer un estimateur sans biais de variance uniformément minimum pour $g(\theta)$. Est-il efficace ?

7. **(**)** Soit un n-échantillon de v.a.i.i.d. de loi admettant la densité f par rapport à la mesure de Lebesgue sur $\mathbb{R}$ telle que :

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

où $\theta \in]-1,1[$ est un paramètre inconnu. Est-ce un modèle exponentiel ? Est-il régulier ? Déterminer l'estimateur $\widehat{\theta}_n$ du maximum de vraisemblance. Est-il sans biais ? Converge-t-il ? Quelle est la loi limite de $\sqrt{n}(\widehat{\theta}_n - \theta_0)$, où θ_0 désigne le "vrai" paramètre ? En déduire un intervalle de confiance à 95%.

8. **(***)** Soit $(\varepsilon_i)_{n\in\mathbb{N}}$ une suite de v.a.i.i.d. de loi de Bernoulli de paramètre $p \in]0,1[$ inconnu. Soit $(X_1, \cdots, X_n)$ un n-échantillon tel que :

$$\forall k \in \{1, \cdots, n\}, \quad X_k = \prod_{i=1}^{k} \varepsilon_i$$

Déterminer l'estimateur du maximum de vraisemblance de p. Est-il convergent ? Sans biais ? Efficace ? Comparer ces résultats avec ceux obtenus par l'appartenance du modèle à la famille exponentielle.

9. **(**)** Soit (X_i) est une suite de v.a.i.i.d. de même loi que X, où X admet pour densité de probabilité $f(x,\theta)$ par rapport à la mesure Lebesgue, définie par :

$$f(x,\theta) = k \cdot x^{-(p+1)} \cdot \exp(-\frac{\theta}{x}) \cdot \mathbb{I}_{x>0},$$

avec $\theta \in]0,+\infty[$ un paramètre réel inconnu et $p > 0$ un nombre connu.

   (a) Montrer que $k = \dfrac{\theta^p}{\Gamma(p)}$. Déterminer une statistique exhaustive complète minimale. Déterminer le paramètre pouvant être estimer efficacement à partir de cette statistique.

   (b) On pose $U = \dfrac{1}{X}$ et $Z = 2\cdot\theta\cdot U$. Déterminer $\mathbb{E}(U)$ et $\mathrm{var}(U)$ ainsi que $\mathbb{E}(Z)$ et $\mathrm{var}(Z)$. Déterminer $\widehat{\theta}_n$ l'estimateur du maximum de vraisemblance de θ. Montrer que :

$$\mathbb{E}(\widehat{\theta}_n) = \frac{n \cdot p \cdot \theta}{n \cdot p - 1}, \quad \mathrm{var}(\widehat{\theta}_n) = \frac{(n \cdot p \cdot \theta)^2}{(n \cdot p - 1)^2 (n \cdot p - 2)}.$$

   L'estimateur $\widehat{\theta}_n$ est-il sans biais? convergent? efficace?

10. **(**)** Soit $(X_1, \cdots, X_n)$ un échantillon de v.a.i.i.d. de loi de Pareto de densité :

$$f(x;\theta) = \frac{3\theta^3}{x^4} \cdot \mathbb{I}_{x\geq\theta}, \quad \text{où } \theta \text{ est un paramètre positif inconnu.}$$

   (a) Montrer que cet échantillon n'appartient pas à la famille exponentielle.

(b) Calculer $\widehat{\theta}_n$ l'estimateur de maximum de vraisemblance de θ.

(c) Déterminer la fonction de répartition puis la densité de l'estimateur $\widehat{\theta}_n$. En déduire que $\widehat{\theta}_n$ est un estimateur convergent de θ. Est-il asymptotiquement efficace ?

(d) Calculer $\mathbb{E}(\widehat{\theta}_n)$ et en déduire un estimateur sans biais de θ, puis un intervalle de confiance à 95% de θ.

Feuille n^o 5 : **Tests paramétriques**

1. **(*)** 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 ? Déterminer la fonction puissance du test. Tester si la pièce est truquée. Aboutit-on au même résultat ?

2. **(*)** 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 320000 euros avec un écart-type de 20000 euros. Peut-on conclure à l'efficacité de cette politique au niveau 5% ? 1% ?

3. **(*)** 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 et de Wald.

4. **(*)** Dans le cas d'un échantillon $(X_1, \cdots, X_n)$ de v.a.i.i.d. de loi $\mathcal{N}(m, 1)$, où m est inconnu, déterminer le test de Wald de niveau 5% (statistique de test, région critique) pour les hypothèses :

   (a) $H_0 \ : m = 0$ contre $H_1 \ : m = 2$;

   (b) $H_0 \ : m = 0$ contre $H_1 \ : m > 0$;

   (c) $H_0 \ : m = 0$ contre $H_1 \ : m \neq 0$.

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

5. **(**)** Dans le cas d'un échantillon $(X_1, \cdots, X_n)$ de v.a.i.i.d. déterminer le test du rapport de vraisemblance de niveau 5% (statistique de test, région critique), pour les hypothèses :

   (a) $H_0 \ : \theta = \theta_0$ contre $H_1 \ : \theta = \theta_1 > \theta_0$;

   (b) $H_0 \ : \theta = \theta_0$ contre $H_1 \ : \theta > \theta_0$;

   (c) $H_0 \ : \theta = \theta_0$ contre $H_1 \ : \theta \neq \theta_0$;

   (d) $H_0 \ : \theta < \theta_0$ contre $H_1 \ : \theta > \theta_1 > \theta_0$;

   (e) $H_0 \ : \theta \neq \theta_0$ contre $H_1 \ : \theta = \theta_0$,

   où $(\theta_0, \theta_1) \in \Theta^2$ sachant que la loi de X_1 est 1/ une loi de Bernoulli de paramètre $\theta \in]0, 1[$, 2/ une loi uniforme sur $[0, \theta]$ avec $\theta > 0$, 3/ une loi exponentielle de paramètre $\theta > 0$.

6. **(**)** On considère une seule variable aléatoire X observée, de loi binomiale $\mathcal{B}(k,p)$ où $k \in \mathbb{N}^*$ est connu, mais $p \in]0,1[$ est inconnu.

   (a) Montrer que $p \in]0,1[\mapsto L_p(x)$, où $x \in \{0,1,\ldots,k\}$ est une fonction croissante sur $]0,x/k]$, puis décroissante sur $[x/k,1[$.

   (b) On veut tester $H_0 \ : p > p_0$ contre $H_1 \ : p \leq p_0$ au niveau α. Vérifier que le rapport de vraisemblance de ce problème est une fonction décroissante de l'observation. En déduire la forme de la région de rejet du test.

7. **(**)** Soit $X_1, \cdots, X_n$ des v.a.i.i.d. de loi exponentielle de paramètre $1/\theta$ avec $\theta > 0$, inconnu. On souhaite faire un choix entre les deux hypothèses suivantes :

$$H_0 : \ \theta = 1, \qquad H_1 : \theta = \theta_1 > 1$$

   (a) Appliquer le test du rapport de vraisemblance (ici Neyman-Pearson) de niveau α. Vérifier que la région de rejet de H_0 est de la forme

$$\{\widehat{X}_n > \mu_\alpha, \}$$

   où $\mu_\alpha > 0$. Calculer l'erreur de seconde espèce β en fonction de μ_α, θ_1 et de la fonction de répartition G_{2n} de la loi du $\mathcal{X}^2_{2n}$.

   (b) Soit N_n le nombre de X_j supérieurs à $\dfrac{1}{2}(1+\theta_1)$. On accepte l'hypothèse H_1 lorsque $N_n \geq \mu'_\alpha$. Pour tout $\alpha \in]0,1[$, ce test est-il moins puissant que celui de Neyman-Pearson ?

8. **(**)** Soit $(X_1, \cdots, X_n)$ un échantillon de v.a.i.i.d. de densité par rapport à la mesure de Lebesgue sur $\mathbb{R}_+$:

$$f(x) = e^{-(x-\theta)} \, \mathbb{I}_{x \geq \theta},$$

   où $\theta \in \mathbb{R}$ est inconnu.

   (a) Montrer que la statistique $\widehat{\theta}_n = \min(X_1, \cdots, X_n)$ est exhaustive pour θ. Quelle est le biais et la variance de cet estimateur ? En déduire un test des hypothèses $H_0 \ : \theta \leq \theta_0$ contre $H_1 \ : \theta > \theta_0$ de niveau α utilisant comme statistique de test $\widehat{\theta}_n$.

   (b) Déterminer le test $\widehat{T}$ du rapport de vraisemblance de niveau α pour le même problème de test. Conclusion ? Pour α fixé, quelle est explicitement la région critique ? Déterminer la fonction puissance de $\widehat{T}$.

9. **(**)** Soit $(X_1, \cdots, X_n)$ un échantillon de v.a.i.i.d. de densité par rapport à la mesure de Lebesgue sur $\mathbb{R}_+$:

$$f(x) = \frac{\alpha \cdot r^\alpha}{x^{\alpha+1}} \, \mathbb{I}_{x \geq r},$$

   où $\alpha \in \mathbb{R}$ et $r > 0$ sont inconnus. Un tel modèle s'appelle le modèle de Pareto, et pourrait être appliqué en économie pour modéliser des revenus, sachant que r serait une sorte de revenu minimal.

   (a) A quelle condition une variable de Pareto admet une espérance ? une variance ?

   (b) Donner l'estimateur du maximum de vraisemblance $(\widehat{a}_n, \widehat{r}_n)$ de ces deux paramètres.

   (c) On désirer tester $H_0 \ : a = 1$ contre $H_1 \ : a \neq 1$. Montrer que le test $\widehat{T}$ du rapport de vraisemblance de niveau α dans ce cas admet une région critique de la forme $[\widehat{a}_n > s_1] \cup [\widehat{a}_n < s_2]$.

10. **(***)** Soit $(X_1, \cdots, X_n)$ et $(X'_1, \cdots, X'_n)$ deux échantillons de v.a.i.i.d. de loi exponentielle de moyennes respectives θ et θ' où $\theta > 0$ et $\theta' > 0$ sont inconnus. On désirer tester $H_0 \ : \theta = \theta'$ contre $H_1 \ : \theta \neq \theta'$. Déterminer le test du rapport de vraisemblance de niveau α. Montrer que la région critique de ce test peut s'exprimer en fonction de la statistique $\widehat{T} = \dfrac{\sum_{i=1}^n X_i}{\sum_{i=1}^n X_i + \sum_{i=1}^n Y_i}$. Déterminer la loi de $\widehat{T}$ sous l'hypothèse H_0.

11. **(***)** Soit $(X_1, \cdots, X_n)$ un échantillon de v.a.i.i.d. de loi gaussiennes $\mathcal{N}(m, \sigma^2)$ où $m \in \mathbb{R}$ et $\sigma^2 > 0$ sont deux paramètres inconnus. Montrer que pour le problème de test $H_0 \ : m = m_0$ contre $H_1 \ : m \neq m_0$ de niveau α, le test du rapport de vraisemblance se ramène à utiliser une statistique de Student. Quel est également le test de Wald pour ce problème ?