

STATISTIQUE 1 (2A ENSAE)

Corrigé des TDs 1-9

Année 2020/2021

Professeur : Arnak Dalalyan, CREST-ENSAE.

Notes rédigées par : Martin Mugnier, CREST-ENSAE, martin.mugnier@ensae.fr¹.

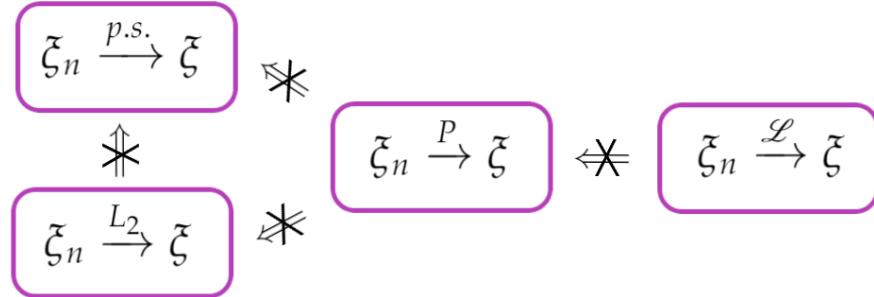
Table des matières

Fiche de TD 1	2
Fiche de TD 2	9
Fiche de TD 3	15
Fiche de TD 4	24
Fiche de TD 5	32
Fiche de TD 6	42
Fiche de TD 7	50
Fiche de TD 8	63
Fiche de TD 9	71

¹. Je remercie Arnak Dalalyan et Lucas Girard pour avoir mis à ma disposition leurs corrigés existants, ainsi que Lionel Riou-Durand pour sa participation à la correction de l'exercice 3 du TD2.

TD1 : DIFFÉRENTS TYPES DE CONVERGENCE

Le but de cette séance est de rappeler les différents types de convergence d'une suite de variables aléatoires réelles. En particulier, les différents exercices ci-dessous montrent que



Exercice 1. Montrer que si $\xi_n \xrightarrow{P} \xi$ et $\xi_n \xrightarrow{P} \eta$, alors les variables ξ et η sont presque sûrement égales, c'est-à-dire

$$\mathbf{P}(\xi = \eta) \triangleq \mathbf{P}(\{\omega : \xi(\omega) = \eta(\omega)\}) = 1.$$

Rappel sur les mesures positives : soit (E, \mathcal{A}) un espace mesurable et μ une mesure positive sur (E, \mathcal{A}) , i.e. une application $\mu : \mathcal{A} \longrightarrow [0, \infty]$ qui vérifie les propriétés suivantes :

(i) $\mu(\emptyset) = 0$;

(ii) Pour toute famille $(A_n)_{n \in \mathbb{N}}$ de parties mesurables disjointes ($A_n \in \mathcal{A}, \forall n \in \mathbb{N}$),

$$\mu(\bigcup_{n \in \mathbb{N}} A_n) = \sum_{n \in \mathbb{N}} \mu(A_n) \quad (\sigma\text{-additivité}).$$

On a les deux propriétés importantes suivantes :

P1 Si $A \subset B$, $\mu(A) \leq \mu(B)$.

P2 Si $A_n \in \mathcal{A}$, $\mu(\bigcup_{n \in \mathbb{N}} A_n) \leq \sum_{n \in \mathbb{N}} \mu(A_n)$.

On suppose ici que ξ_n, ξ, η sont des v.a.r. définies sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbf{P})$, i.e. un espace mesurable (Ω, \mathcal{A}) doté d'une mesure de probabilité \mathbf{P} , i.e. une mesure positive telle que $\mathbf{P}(\Omega) = 1$. Supposons $\xi_n \xrightarrow{P} \xi$ et $\xi_n \xrightarrow{P} \eta$. Soit l'ensemble

$$A := \{\omega : |\xi(\omega) - \eta(\omega)| > 0\}.$$

Il s'agit de montrer que $\mathbf{P}(A) = 0$. Pour tout $\epsilon > 0$, on pose

$$A_\epsilon := \{\omega : |\xi(\omega) - \eta(\omega)| \geq \epsilon\}.$$

Montrons d'abord

$$\forall \epsilon > 0, \mathbf{P}(A_\epsilon) = 0. \quad (1)$$

Soit $\epsilon > 0$. D'après l'inégalité triangulaire, pour tout $(\omega, n) \in \Omega \times \mathbb{N}^*$

$$|\xi(\omega) - \eta(\omega)| \leq |\xi_n(\omega) - \xi(\omega)| + |\xi_n(\omega) - \eta(\omega)|.$$

Ainsi,

$$A_\epsilon \subseteq \{\omega : |\xi_n(\omega) - \xi(\omega)| \geq \frac{\epsilon}{2} \text{ ou } |\xi_n(\omega) - \eta(\omega)| \geq \frac{\epsilon}{2}\}.$$

En utilisant **P1** puis **P2** (appelée aussi borne d'union), on a

$$\begin{aligned} \mathbf{P}(A_\epsilon) &\leq \mathbf{P}\left(\omega : |\xi_n(\omega) - \xi(\omega)| \geq \frac{\epsilon}{2} \text{ ou } |\xi_n(\omega) - \eta(\omega)| \geq \frac{\epsilon}{2}\right) \\ &\leq \mathbf{P}(\omega : |\xi_n(\omega) - \xi(\omega)| \geq \frac{\epsilon}{2}) + \mathbf{P}(\omega : |\xi_n(\omega) - \eta(\omega)| \geq \frac{\epsilon}{2}). \end{aligned}$$

Or, par hypothèse et définition de la convergence en probabilité,

$$\lim_{n \rightarrow +\infty} \mathbf{P}(\omega : |\xi_n(\omega) - \xi(\omega)| \geq \frac{\epsilon}{2}) = \lim_{n \rightarrow +\infty} \mathbf{P}(\omega : |\xi_n(\omega) - \eta(\omega)| \geq \frac{\epsilon}{2}) = 0.$$

On en déduit $\mathbf{P}(A_\epsilon) = 0$. Comme ϵ était quelconque, (1) est vraie. En remarquant à présent que

$$A = \bigcup_{n \in \mathbb{N}^*} A_{1/n},$$

une seconde application de **P2** donne

$$\mathbf{P}(A) = \mathbf{P}(\bigcup_{n \in \mathbb{N}^*} A_{1/n}) \leq \sum_{n \in \mathbb{N}^*} \mathbf{P}(A_{1/n}) = 0.$$

Dans les trois exercices suivants, on suppose que

$$(\Omega, \mathcal{A}, \mathbf{P}) = ([0, 1], \mathcal{B}_{[0,1]}, \lambda),$$

où λ est la mesure de Lebesgue sur $[0, 1]$.

Exercice 2. Pour tout $n \in \mathbb{N}$, on pose

$$\xi_n(\omega) = (-1)^n \mathbb{1}_{[0,1/2]}(\omega) + (-1)^{n+1} \mathbb{1}_{(1/2,1]}(\omega).$$

1. La suite $\{\xi_n\}$, converge-t-elle en loi ?

Soit $E := \{-1, 1\}$ et $\mathcal{B} := \{\emptyset, \{-1, 1\}, \{-1\}, \{1\}\}$. Pour tout $n \in \mathbb{N}$,

$$\xi_n : (\Omega, \mathcal{A}) \rightarrow (E, \mathcal{B})$$

et on peut vérifier que la loi de probabilité de ξ_n ne dépend pas de n . Soit $\xi : (\Omega, \mathcal{A}) \rightarrow (E, \mathcal{B})$ la variable aléatoire définie sur $(\Omega, \mathcal{A}, \mathbf{P})$ telle que

$$\xi(\omega) = 2 \times \mathbb{1}_{[0,1/2]}(\omega) - 1.$$

Pour tout $n \in \mathbb{N}$, pour tout $B \in \mathcal{B}$,

$$\mathbf{P}(\xi_n \in B) = \mathbf{P}(\xi \in B).$$

Par exemple,

$$\begin{aligned}\mathbf{P}(\xi_n = 1) &= \mathbf{P}(\omega : \xi_n(\omega) = 1) \\ &= \mathbf{P}(w : \omega \in [0, 1/2] \text{ si } n \text{ est pair}, \omega \in (1/2, 1] \text{ sinon}) \\ &= \lambda([0, 1/2]) = \lambda((1/2, 1]) = 1/2 \\ &= \mathbf{P}(\omega : \xi(\omega) = 1) \\ &= \mathbf{P}(\xi = 1).\end{aligned}$$

Autrement dit, $\xi_n \xrightarrow{D} \xi$ pour tout $n \in \mathbb{N}$ (égalité en distribution). On vérifie alors facilement que pour tout $n \in \mathbb{N}$ et tout point de continuité t de $F_\xi(\cdot) := \mathbf{P}(\xi \leq \cdot)$,

$$\mathbf{P}(\xi_n \leq t) = F_\xi(t).$$

Par définition de la convergence en loi,

$$\xi_n \xrightarrow{\mathcal{L}} \xi.$$

2. La suite $\{\xi_n\}$, converge-t-elle en probabilité ?

Intuition : non, car la fréquence des sauts de ξ_n est constante et égale à 1/2. Montrons-le par l'absurde. Soit ξ une variable aléatoire définie sur $(\Omega, \mathcal{A}, \mathbf{P})$ telle que

$$\xi_n \xrightarrow{P} \xi.$$

Alors $\xi_{2n} \xrightarrow{P} \xi$ et $\xi_{2n+1} \xrightarrow{P} \xi$. Or, pour tout $(\omega, n) \in \Omega \times \mathbb{N}$

$$\begin{aligned}|\xi_{2n}(\omega) - \xi_{2n+1}(\omega)| &= \left| \mathbb{1}_{[0,1/2]}(\omega) - \mathbb{1}_{(1/2,1]}(\omega) - (-\mathbb{1}_{[0,1/2]}(\omega) + \mathbb{1}_{(1/2,1]}(\omega)) \right| \\ &= 2 \left| \mathbb{1}_{[0,1/2]}(\omega) - \mathbb{1}_{(1/2,1]}(\omega) \right| \\ &= 2.\end{aligned}$$

Ainsi, pour tout $n \in \mathbb{N}$

$$\mathbf{P}(\omega : |\xi_{2n}(\omega) - \xi_{2n+1}(\omega)| > 1) = 1.$$

Notons que l'on a aussi

$$\begin{aligned}
\mathbf{P}(\omega : |\xi_{2n}(\omega) - \xi_{2n+1}(\omega)| > 1) &= \mathbf{P}(\omega : |\xi_{2n}(\omega) - \xi(\omega) + \xi(\omega) - \xi_{2n+1}(\omega)| > 1) \\
&\stackrel{\text{I.T. + P1}}{\leq} \mathbf{P}(\omega : |\xi_{2n}(\omega) - \xi(\omega)| + |\xi(\omega) - \xi_{2n+1}(\omega)| > 1) \\
&\stackrel{\mathbf{P1}}{\leq} \mathbf{P}(\omega : |\xi_{2n}(\omega) - \xi(\omega)| > 1/2 \text{ ou } |\xi(\omega) - \xi_{2n+1}(\omega)| > 1/2) \\
&\stackrel{\mathbf{P2}}{\leq} \mathbf{P}(\omega : |\xi_{2n}(\omega) - \xi(\omega)| > 1/2) + \mathbf{P}(\omega : |\xi(\omega) - \xi_{2n+1}(\omega)| > 1/2) \\
&\xrightarrow{n \rightarrow +\infty} 0,
\end{aligned}$$

où "I.T." signifie "inégalité triangulaire" et où la limite provient de l'hypothèse de convergence posée. On aboutit donc à une contradiction et on conclut que la suite $\{\xi_n\}$ ne converge pas en probabilité.

3. La suite $\{\xi_n\}$, converge-t-elle presque sûrement ?

La suite $\{\xi_n\}$ converge presque sûrement vers ξ ssi

$$\mathbf{P}(\{\omega : \xi_n(\omega) \xrightarrow{n \rightarrow +\infty} \xi(\omega)\}) = 1.$$

Or, pour tout $\omega \in [0, 1]$

$$\limsup \xi_n(\omega) = 1 \text{ et } \liminf \xi_n(\omega) = -1.$$

Ainsi, $\mathbf{P}(\{\omega : \xi_n(\omega) \xrightarrow{n \rightarrow +\infty} \xi(\omega)\}) = 0$. La suite $\{\xi_n\}$ ne converge pas presque sûrement.

Exercice 3. Pour tout $n \in \mathbb{N}$, on pose $\xi_n(\omega) = n \mathbb{1}_{[0,1]}(n\omega)$.

1. La suite $\{\xi_n\}$, converge-t-elle en probabilité ?

Pour tout $n \in \mathbb{N}^*$,

$$\begin{aligned}
\mathbf{P}(\xi_n = n) &= \mathbf{P}\left(\omega : 0 \leq \omega \leq \frac{1}{n}\right) = \lambda\left([0, \frac{1}{n}]\right) = \frac{1}{n}, \\
\mathbf{P}(\xi_n = 0) &= 1 - \mathbf{P}(\xi_n = n) = 1 - \frac{1}{n}.
\end{aligned}$$

Soit $\epsilon > 0$ et $n \in \mathbb{N}^*$,

$$\begin{aligned}
\mathbf{P}(\xi_n > \epsilon) &\leq \mathbf{P}(\xi_n \neq 0) \quad \text{d'après P1} \\
&= \mathbf{P}(\xi_n = n) \\
&= \frac{1}{n} \\
&\xrightarrow{n \rightarrow +\infty} 0.
\end{aligned}$$

La suite $\{\xi_n\}$ converge en probabilité vers 0.

2. La suite $\{\xi_n\}$, converge-t-elle presque sûrement ?

Soit $A :=]0, 1] \subset \Omega$. Pour tout $\omega \in A$

$$\exists N \in \mathbb{N}^*, \forall n \geq N : n > 1/\omega$$

car \mathbf{R} est archimédien. Ainsi, pour tout $\omega \in A$

$$\lim_{n \rightarrow +\infty} \xi_n(\omega) = \lim_{n \rightarrow +\infty} \mathbb{1}_{[0,1]}(n\omega) = 0.$$

Comme $\lambda(A) = 1$, on a

$$\mathbf{P}(\omega : \xi_n(\omega) \xrightarrow{n \rightarrow +\infty} 0) = 1.$$

La suite $\{\xi_n\}$ converge presque sûrement vers 0.

3. La suite $\{\xi_n\}$, converge-t-elle au sens L_2 ?

Supposons que $\{\xi_n\}$ converge au sens L_2 . Il existe une v.a.r. ξ définie sur $(\Omega, \mathcal{A}, \mathbf{P})$ telle que $\xi_n \xrightarrow{L_2} \xi$. Par conséquent,

$$\xi_n \xrightarrow{P} \xi.$$

Or, d'après 1), $\xi_n \xrightarrow{P} 0$. D'après, l'exercice 1, les variables ξ et 0 sont presque sûrement égales, i.e.

$$\xi \stackrel{p.s.}{=} 0.$$

On a donc le résultat suivant : si $\{\xi_n\}$ converge au sens L_2 , alors $\{\xi_n\}$ converge vers 0 (en tant que v.a.r. presque sûrement nulle). Il reste à regarder si $\xi_n \xrightarrow{L_2} 0$ ou non.

$$\begin{aligned} E(|\xi_n|^2) &= \int_{\Omega} |\xi_n(\omega)|^2 d\mathbf{P}(\omega) \quad \text{par définition de l'espérance} \\ &= \int_0^1 n^2 \left(\mathbb{1}_{[0,1]}(n\omega) \right)^2 d\lambda(\omega) \\ &= n^2 \int_0^1 \mathbb{1}_{[0, \frac{1}{n}]}(\omega) d\lambda(\omega) \\ &= n^2 \int_0^{1/n} 1 d\lambda(\omega) \\ &= n^2 \times \frac{1}{n} \\ &= n. \end{aligned}$$

$\lim_{n \rightarrow +\infty} E(|\xi_n|^2) \neq 0$. Donc la suite $\{\xi_n\}$ ne converge pas vers 0 au sens L_2 et, partant, elle ne converge pas au sens L_2 .

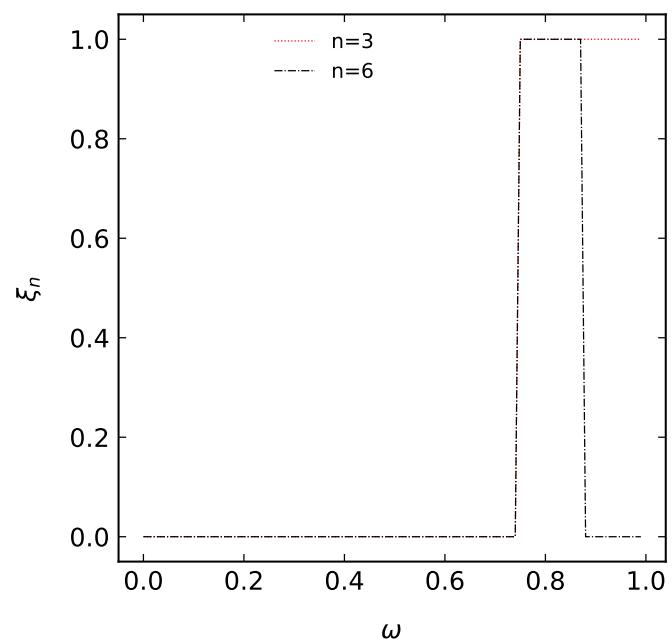
Exercice 4. Pour tout $n \in \mathbb{N}$, soit k_n l'unique nombre entier vérifiant $2^k \leq n < 2^{k+1}$. On définit la suite $\{\xi_n\}$ par

$$\xi_n(\omega) = \mathbb{1}_{[n, n+1]}(2^{k_n+1}\omega).$$

1. Tracer les graphiques de ξ_3 et ξ_6 .

On peut vérifier que, pour tout $\omega \in \Omega$

$$\begin{aligned}\xi_3(\omega) &= \mathbb{1}_{[3/4,1]}(\omega), \\ \xi_6(\omega) &= \mathbb{1}_{[3/4,7/8]}(\omega).\end{aligned}$$



2. Montrer que la suite $\{\xi_n\}$ converge en probabilité vers zéro.

Soit $\epsilon > 0, n \in \mathbb{N}^*$.

$$\begin{aligned}\mathbf{P}(\omega : \xi_n(\omega) > \epsilon) &\leq \mathbf{P}(\omega : \xi_n(\omega) \neq 0) \quad \text{d'après P1} \\ &= \mathbf{P}(\omega : \xi_n(\omega) = 1) \\ &= \lambda\left(\left[\frac{n}{2^{k_n+1}}, \frac{n+1}{2^{k_n+1}}\right]\right) \\ &= \frac{1}{2^{k_n+1}} \\ &= \exp\left((k_n + 1) \times \underbrace{\log(1/2)}_{<0}\right) \\ &\xrightarrow[n \rightarrow +\infty]{} 0,\end{aligned}$$

la limite se déduisant de l'encadrement suivant :

$$\begin{aligned}2^{k_n} \leq n < 2^{k_n+1} &\iff k_n \log(2) \leq \log(n) \leq (k_n + 1) \log(2) \\ &\implies k_n + 1 \geq \frac{\log(n)}{\log(2)} \xrightarrow[n \rightarrow +\infty]{} +\infty.\end{aligned}$$

La suite $\{\xi_n\}$ converge en probabilité vers 0.

3. Montrer que la suite $\{\xi_n\}$ ne converge pas presque sûrement.

Si $\xi_n \xrightarrow{p.s.} \xi$ alors $\xi_n \xrightarrow{P} \xi$ et donc, d'après l'exercice 1 et la question précédente, $\xi \stackrel{p.s.}{=} 0$ donc $\xi_n \xrightarrow{p.s.} 0$. Montrons que ce n'est pas le cas. Soit $A := [1/2, 1[$. Comme $\mathbf{P}(A) = 1/2 > 0$, il suffit de montrer :

$$\forall \omega \in A, \xi_n(\omega) \not\xrightarrow[n \rightarrow +\infty]{} 0.$$

Soit $\omega \in A$. On va montrer que $\xi_n(\omega) \not\xrightarrow[n \rightarrow +\infty]{} 0$ car cette suite contient une infinité d'éléments égaux à 1. En effet,

$$\forall k \in \mathbb{N}^*, \exists n \in \mathbb{N} : k_n = k \text{ et } \frac{n}{2^{k+1}} \leq \omega < \frac{n+1}{2^{k+1}}.$$

On en déduit que, pour ce n , $\xi_n(\omega) = 1$. On a donc,

$$\forall k \in \mathbb{N}^*, \exists n \in [2^k, 2^{k+1}[: \xi_n(\omega) = 1.$$

Comme les intervalles $([2^k, 2^{k+1}[)_{k \in \mathbb{N}^*}$ sont disjoints, il y a bien une infinité de termes égaux à 1 dans la suite $\{\xi_n(\omega)\}$. Donc $\xi_n \not\xrightarrow{} 0$ et, a fortiori, $\{\xi_n\}$ ne converge pas presque sûrement.

TD2 : MÉTHODES EMPIRIQUES

Exercice 1. (Inégalité de Hoeffding) Soient V_1, \dots, V_n des variables indépendantes (pas nécessairement de même loi) à valeurs dans $[-1, 1]$. On pose $\bar{V}_n = \frac{1}{n} \sum_{i=1}^n V_i$ et $\mu = \mathbf{E}[\bar{V}_n]$. Le but de l'exercice est de prouver que

$$\mathbf{P}(\bar{V}_n - \mu \geq t) \leq e^{-nt^2/2}, \quad \forall t > 0.$$

1. Montrer que pour tout $\lambda > 0$, on a

$$\mathbf{P}(\bar{V}_n - \mu \geq t) \leq e^{-n\lambda(t+\mu)} \prod_{i=1}^n \mathbf{E}[e^{\lambda V_i}].$$

Soit $\lambda > 0, t > 0$,

$$\begin{aligned} \mathbf{P}(\bar{V}_n - \mu \geq t) &= \mathbf{P}\left(e^{n\lambda\bar{V}_n} \geq e^{n\lambda(t+\mu)}\right) \quad \text{mêmes événements car } n, \lambda > 0 \\ &\leq \frac{\mathbf{E}\left[e^{n\lambda\bar{V}_n}\right]}{e^{n\lambda(t+\mu)}} \quad \text{par l'inégalité de Markov car } e^{n\lambda\bar{V}_n} \geq 0 \\ &= e^{-n\lambda(t+\mu)} \mathbf{E}\left[\prod_{i=1}^n e^{\lambda V_i}\right] \\ &= e^{-n\lambda(t+\mu)} \prod_{i=1}^n \mathbf{E}\left[e^{\lambda V_i}\right] \quad \text{car les } (V_i)_{i=1}^n \text{ sont indépendantes.} \end{aligned}$$

A noter que l'on a $\mathbf{E}[e^{\lambda V_i}] \leq e^\lambda < +\infty$ car $V_i \in [-1, 1]$ p.s. et $\lambda > 0$.

2. Posons $\varphi_i(\lambda) = \log \mathbf{E}[e^{\lambda V_i}]$. Calculer la dérivée et la dérivée seconde de φ_i .

Soit $i \in \llbracket 1, n \rrbracket$, $\lambda \in \mathbf{R}_+^*$,

$$\varphi'_i(\lambda) := \frac{d\varphi}{d\lambda}(\lambda) = \frac{\frac{d\mathbf{E}[e^{\lambda V_i}]}{d\lambda}(\lambda)}{\mathbf{E}[e^{\lambda V_i}]}.$$

Le calcul du numérateur se fait en appliquant le Théorème de dérivation d'une intégrale dépendant d'un paramètre (un corollaire du Théorème de convergence dominée). Posons $F_i(\lambda) := \exp(\varphi_i) = \mathbf{E}[e^{\lambda V_i}]$. On a

$$\begin{aligned} F_i(\lambda) &= \int e^{\lambda v} dP^{V_i}(v) \quad \text{par définition de l'espérance, où } P^{V_i} \equiv \text{la loi de la v.a. } V_i \\ &= \int_{[-1,1]} f(\lambda, v) dP^{V_i}(v), \end{aligned}$$

où $f(\lambda, v) := e^{\lambda v}$. On a

- (i) dérivabilité : $\forall v \in [-1, 1], \lambda \mapsto e^{\lambda v}$ est dérivable et $\frac{\partial f}{\partial \lambda}(\lambda, v) = ve^{\lambda v}$;
- (ii) domination : $\forall \lambda > 0, \forall v \in [-1, 1], \left| \frac{\partial f}{\partial v}(\lambda, v) \right| \leq |V_i e^{\lambda V_i}|$;
- (iii) intégrabilité : $\forall \lambda > 0, |V_i e^{\lambda V_i}| \leq e^\lambda \in L_1$.

Ainsi, on peut dériver sous le signe intégrale (somme) et on obtient

$$\frac{d\mathbf{E}[e^{\lambda V_i}]}{d\lambda}(\lambda) = F'_i(\lambda) = \mathbf{E}\left[\left(e^{\lambda V_i}\right)'\right] = \mathbf{E}\left[V_i e^{\lambda V_i}\right].$$

On conclut

$$\varphi'_i(\lambda) = \frac{\mathbf{E}[V_i e^{\lambda V_i}]}{\mathbf{E}[e^{\lambda V_i}]}.$$

Une seconde application du Théorème de dérivation sous le signe intégral donne

$$\begin{aligned} \varphi''_i(\lambda) &= \frac{(\mathbf{E}[V_i e^{\lambda V_i}])' \mathbf{E}[e^{\lambda V_i}] - \mathbf{E}[V_i e^{\lambda V_i}] (\mathbf{E}[e^{\lambda V_i}])'}{(\mathbf{E}[e^{\lambda V_i}])^2} \\ &= \frac{\mathbf{E}\left[(V_i e^{\lambda V_i})'\right]}{\mathbf{E}[e^{\lambda V_i}]} - \frac{(\mathbf{E}[V_i e^{\lambda V_i}])^2}{(\mathbf{E}[e^{\lambda V_i}])^2} \\ &= \underbrace{\frac{\mathbf{E}[V_i^2 e^{\lambda V_i}]}{\mathbf{E}[e^{\lambda V_i}]}}_{\leq 1 \text{ car } V_i^2 \leq 1 \text{ p.s.}} - \underbrace{\left(\frac{\mathbf{E}[V_i e^{\lambda V_i}]}{\mathbf{E}[e^{\lambda V_i}]}\right)^2}_{\geq 0} \leq 1. \end{aligned}$$

3. Montrer que $\varphi''_i(x) \leq 1$ pour tout $x \in \mathbb{R}$. En déduire que $\varphi_i(\lambda) \leq \lambda \mathbf{E}[V_i] + \lambda^2/2$.

Par Taylor, on en déduit que $\forall \lambda \in \mathbf{R} \exists \theta \in [0, 1]$:

$$\varphi_i(\lambda) = \lambda \underbrace{\varphi'_i(0)}_{=\mathbf{E}[V_i]} + \frac{\lambda^2}{2} \underbrace{\varphi''_i(\theta \lambda)}_{\leq 1} \leq \lambda \mathbf{E}[V_i] + \frac{\lambda^2}{2}.$$

4. Utiliser la question 1 et la question précédente, pour vérifier que

$$\mathbf{P}(\bar{V}_n - \mu \geq t) \leq e^{-n\lambda t + n\lambda^2/2}, \quad \forall \lambda, t > 0.$$

Conclure.

En utilisant les questions 1 et 3, on obtient $\forall \lambda > 0, \forall t > 0, \forall n \in \mathbb{N}^*$:

$$\mathbf{P}(\bar{V}_n - \mu \geq t) \leq e^{-n\lambda(t+\mu)} \prod_{i=1}^n \mathbb{E}[e^{\lambda V_i}] \quad (q.1)$$

$$= e^{-n\lambda(t+\mu)} \exp \left\{ \sum_{i=1}^n \varphi_i(\lambda) \right\}$$

$$\leq e^{-n\lambda(t+\mu)} \exp \left\{ \sum_{i=1}^n (\lambda \mathbb{E}[V_i] + \lambda^2/2) \right\} \quad (q.3)$$

$$= e^{-n\lambda(t+\mu)} \exp \left\{ n\mu + n\lambda^2/2 \right\}$$

$$= e^{-n\lambda t + n\lambda^2/2}.$$

Puisque l'inégalité obtenue est vraie pour tout $\lambda > 0$, on peut optimiser la borne par rapport à $\lambda > 0$. Puisque la fonction $\lambda \mapsto \lambda^2/2 - \lambda t$ est minimale pour $\lambda = t$ on obtient :

$$\mathbf{P}(\bar{V}_n - \mu \geq t) \leq \min_{\lambda > 0} e^{-n\lambda t + n\lambda^2/2} = \exp \left\{ n \min_{\lambda > 0} (\lambda^2/2 - \lambda t) \right\} = e^{-nt^2/2}.$$

Exercice 2. Soient X_1, \dots, X_n des variables aléatoires i.i.d. de fonction de répartition F^* . Rappelons que la fonction de répartition empirique est définie par

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}(X_i \leq x).$$

1. Montrer que

$$\mathbf{P}\left(2\hat{F}_n(x) - 1 \geq 2F^*(x) - 1 + t\right) \leq e^{-nt^2/2}, \quad \forall t \in \mathbb{R}_+.$$

Utilisons l'inégalité de Hoeffding pour $V_i = 2\mathbb{1}(X_i \leq x) - 1 \in [-1, 1]$, $i \in \mathbb{N}^*$. On a $\bar{V}_n = 2\hat{F}_n(x) - 1$, et $\mu = \mathbb{E}[\bar{V}_n] = \mathbb{E}[V_1] = 2F^*(x) - 1$, donc :

$$P\left(2\hat{F}_n(x) - 1 \geq 2F^*(x) - 1 + t\right) = P\left(\bar{V}_n - \mu \geq t\right) \leq e^{-nt^2/2}, \quad \forall t > 0. \quad (2)$$

2. En déduire que

$$\mathbf{P}\left(|\hat{F}_n(x) - F^*(x)| \geq y\right) \leq 2e^{-2ny^2}, \quad \forall y \in \mathbb{R}_+.$$

Remarque : En appliquant l'inégalité de Hoeffding sur les variables $-V_i \in [-1, 1]$, $i \in \mathbb{N}^*$, on obtient une seconde inégalité :

$$\mathbf{P}\left(1 - 2\hat{F}_n(x) \geq 1 - 2F^*(x) + t\right) \leq e^{-nt^2/2}, \quad \forall t > 0. \quad (3)$$

Les deux inégalités (2) et (3) se réécrivent pour $y = t/2 > 0$:

$$\left\{ \begin{array}{l} (2) \\ (3) \end{array} \right\} \Leftrightarrow \left\{ \begin{array}{l} \mathbf{P}(\hat{F}_n(x) - F^*(x) \geq y) \leq e^{-2ny^2} \\ \mathbf{P}(\hat{F}_n(x) - F^*(x) \leq -y) \leq e^{-2ny^2} \end{array} \right\}.$$

Enfin $\forall y > 0$ on obtient par sous-additivité :

$$\begin{aligned} \mathbf{P}\left(|\hat{F}_n(x) - F^*(x)| \geq y\right) &\leq 2e^{-2ny^2} = \mathbf{P}\left(\{\hat{F}_n(x) - F^*(x) \geq y\} \cup \{\hat{F}_n(x) - F^*(x) \leq -y\}\right) \\ &\leq \mathbf{P}(\hat{F}_n(x) - F^*(x) \geq y) + \mathbf{P}(\hat{F}_n(x) - F^*(x) \leq -y) \\ &\leq 2e^{-2ny^2}. \end{aligned}$$

3. Soit $\alpha \in]0, 1[$. Trouver un nombre réel $a > 0$ tel que

$$\mathbf{P}\left(F^*(x) \in [\hat{F}_n(x) - a, \hat{F}_n(x) + a]\right) \geq 1 - \alpha.$$

En utilisant la question 2, $\forall y > 0$ on obtient par inclusion :

$$\begin{aligned} \mathbf{P}\left(F^*(x) \in [\hat{F}_n(x) - y, \hat{F}_n(x) + y]\right) &\geq \mathbf{P}\left(F^*(x) \in [\hat{F}_n(x) - y, \hat{F}_n(x) + y]\right) \\ &= 1 - \mathbf{P}\left(|\hat{F}_n(x) - F^*(x)| \geq y\right) \\ &\geq 1 - 2e^{-2ny^2}. \end{aligned}$$

En posant $\alpha = 2e^{-2ny^2} \in]0, 1[$, on définit implicitement

$$y = \sqrt{-\frac{1}{2n} \log\left(\frac{\alpha}{2}\right)} = \sqrt{\frac{\log(2/\alpha)}{2n}} > 0.$$

Autrement dit, pour tout $\alpha \in]0, 1[$

$$\mathbf{P}\left(F^*(x) \in \left[\hat{F}_n(x) - \sqrt{\frac{\log(2/\alpha)}{2n}}, \hat{F}_n(x) + \sqrt{\frac{\log(2/\alpha)}{2n}}\right]\right) \geq 1 - \alpha.$$

Exercice 3. Soient X_1, \dots, X_n des variables indépendantes et de même loi de moment d'ordre 4 fini. Rappelons les notations

$$\mu = \mathbf{E}[X_1], \quad \sigma^2 = \mathbf{E}[X_1^2] - \mu^2, \quad S_n = \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}_n^2.$$

Supposons, pour simplifier les calculs, que $\mu = \mathbf{E}[X_1] = 0$.

On peut supposer $\mu = 0$ sans perte de généralité, car on peut toujours se ramener à l'étude des variables centrées $Y_i = X_i - \mu$, puisque $\sigma^2 = \mathbb{V}(X_1) = \mathbb{V}(Y_1)$ et $S_n = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 = \frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y}_n)^2$.

- Montrer que la suite $\{n^{1/4}\bar{X}_n : n \in \mathbb{N}^*\}$ converge en probabilité vers zéro.

On remarque que $\forall n \in \mathbb{N}^*$, $n^{1/4}\bar{X}_n = (n^{1/2}\bar{X}_n) \times n^{-1/4}$.

Pour $Z \sim \mathcal{N}(0, \sigma^2)$, on a par le Théorème Central Limite (TCL) : $n^{1/2}\bar{X}_n \xrightarrow{\mathcal{L}} Z$, car les v.a. X_i sont iid centrées et de carré intégrables. De plus la convergence $n^{-1/4} \rightarrow 0$ ne dépend pas de $\omega \in \Omega$, donc a fortiori $n^{-1/4} \xrightarrow{p.s.} 0 \Rightarrow n^{-1/4} \xrightarrow{P} 0$.

Le lemme de Slutsky nous dit que pour toute fonction $\varphi : \mathbf{R}^2 \mapsto \mathbf{R}$ continue on a donc $\varphi(n^{1/2}\bar{X}_n, n^{-1/4}) \xrightarrow{\mathcal{L}} \varphi(Z, 0)$. En posant $\varphi(x, y) = xy$, on obtient $n^{1/4}\bar{X}_n \xrightarrow{\mathcal{L}} \varphi(Z, 0) = 0$. Puisque la limite : 0 est une constante déterministe, la convergence en loi est équivalente à la convergence en probabilité.

- Trouver la loi limite de la suite $\{\sqrt{n}(S_n - \sigma^2) : n \in \mathbb{N}^*\}$.

— En utilisant seulement le TCL univarié. On a $\forall n \in \mathbb{N}^*$:

$$\sqrt{n}(S_n - \sigma^2) = \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n X_i^2 - \sigma^2 \right) - \left(n^{1/4}\bar{X}_n \right)^2.$$

Or par le TCL, puisque les X_i admettent un moment d'ordre 4, pour $W \sim \mathcal{N}(0, \mathbb{E}[X_1^4] - (\sigma^2)^2)$ on a

$$\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n X_i^2 - \sigma^2 \right) \xrightarrow{\mathcal{L}} W,$$

et on sait par la question 1 que

$$n^{1/4}\bar{X}_n \xrightarrow{P} 0.$$

Donc, par le lemme de Slutsky pour $\psi(x, y) = x - y^2$ continue, on obtient

$$\sqrt{n}(S_n - \sigma^2) = \psi \left(\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n X_i^2 - \sigma^2 \right), n^{1/4}\bar{X}_n \right) \xrightarrow{\mathcal{L}} \psi(W, 0) = W.$$

— En utilisant le TCL multivarié et la δ -méthode. Soit

$$\begin{aligned} \varphi &: \mathbf{R} \times \mathbf{R}_+^* \rightarrow \mathbf{R} \\ (x, y) &\mapsto y - x^2. \end{aligned}$$

On a $S_n = \varphi(\bar{X}_n, \bar{X}_n^2)$. Pour tout $k \in \{1, 2, 3, 4\}$, on note $\alpha_k := \mathbb{E}[X^k]$ et d'après l'énoncé $|\alpha_k| < +\infty$ pour tout k . Comme $(X_i)_{i=1}^n$ est un échantillon i.i.d., le Théorème de la Limite Centrale multivarié assure :

$$\sqrt{n} \left(\begin{pmatrix} \bar{X}_n \\ \bar{X}_n^2 \end{pmatrix} - \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} \right) \xrightarrow{\mathcal{L}} \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \alpha_2 - \alpha_1^2 & \alpha_3 - \alpha_1\alpha_2 \\ \alpha_3 - \alpha_1\alpha_2 & \alpha_4 - \alpha_2^2 \end{pmatrix} \right) \equiv \begin{pmatrix} T_1 \\ T_2 \end{pmatrix}.$$

La fonction φ est différentiable sur $\mathbf{R} \times \mathbf{R}_+^*$, donc au point $\theta = (\alpha_1, \alpha_2)'$, et a pour dérivée en ce point $\varphi'(\theta) = (-2\alpha_1, 1)$. En appliquant la Delta-Méthode, on obtient

$$\sqrt{n} \left(\varphi(\bar{X}_n, \bar{X}^2_n) - \varphi(\alpha_1, \alpha_2) \right) \xrightarrow{\mathcal{L}} -2\alpha_1 T_1 + T_2,$$

qui est une loi Gaussienne de moyenne zéro et dont la variance peut être exprimée en terme des $\alpha_1, \dots, \alpha_4$. Lorsque $\alpha_1 = 0$, cette variance vaut $\alpha_4 - \alpha_2^2$. En remarquant que S_n ne change pas lorsque l'on centre chaque observation X_i (i.e., on la remplace par $Y_i = X_i - \alpha_1$), on peut supposer sans perte de généralité $\alpha_1 = 0$. En notant $\mu_k = \mathbf{E}[Y_i^k]$ le moment centré d'ordre k de X_i , et en remarquant que $S_n = \varphi(\bar{Y}_n, \bar{Y}^2_n)$ et $\varphi(\mu_1, \mu_2) = \mu_2 = \sigma^2$, on a :

$$\sqrt{n} (S_n - \sigma^2) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \mu_4 - \mu_2^2).$$

TD3 : ESTIMATION DE PARAMÈTRES

Exercice 1. Soient X_1, \dots, X_n des variables indépendantes de loi $\mathcal{N}(0, \theta^*)$, où $\theta^* > 0$ est un paramètre inconnu. On considère les deux estimateurs suivants du paramètre θ^* :

$$\hat{\theta}_n = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2, \quad \tilde{\theta}_n = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2.$$

Le but de cet exercice est de comparer ces 2 estimateurs.

Le modèle statistique est

$$(\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n), \{\mathbf{P}_\theta = \mathcal{N}(0, \theta)^{\otimes n}, \theta \in \Theta := \mathbf{R}_+^*\}).$$

On est dans le cadre d'un modèle paramétrique identifié. \mathbf{P}_θ est la loi des observations (X_1, \dots, X_n) paramétrée par θ et $\Theta \subset \mathbf{R}$ est l'espace des paramètres.

1. Montrer que

$$\hat{\theta}_n = \frac{n-1}{n^2} \sum_{i=1}^n X_i^2 - \frac{2}{n^2} \sum_{k<\ell} X_k X_\ell,$$

où $\sum_{k<\ell}$ est un raccourci pour $\sum_{\ell=2}^n \sum_{k=1}^{\ell-1}$.

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

Or

$$\bar{X}_n^2 = \frac{1}{n^2} \left(\sum_{i=1}^n X_i \right)^2 = \underbrace{\frac{1}{n^2} \left(\sum_{i=1}^n X_i^2 + 2 \sum_{k<l} X_k X_l \right)}_{\text{le carré d'une somme}}.$$

On en déduit le résultat annoncé :

$$\begin{aligned} \hat{\theta}_n &= \left(\frac{1}{n} - \frac{1}{n^2} \right) \sum_{i=1}^n X_i^2 - \frac{2}{n^2} \sum_{k<\ell} X_k X_\ell \\ &= \frac{n-1}{n^2} \sum_{i=1}^n X_i^2 - \frac{2}{n^2} \sum_{k<\ell} X_k X_\ell. \end{aligned}$$

Remarque : on pouvait aussi utiliser le symbole de Kronecker, $\delta_{ij} = \mathbb{1}\{i = j\}$, et remarquer que

$$\begin{aligned}
\hat{\theta}_n &= \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \\
&= \frac{1}{n} \sum_{i=1}^n \left(\sum_{j=1}^n \frac{1}{n} X_j (n\delta_{ij} - 1) \right)^2 \\
&= \frac{1}{n^3} \sum_{i=1}^n \sum_{j=1}^n X_j^2 (n\delta_{ij} - 1)^2 + \frac{2}{n^3} \sum_{i=1}^n \sum_{j=2}^n \sum_{k=1}^{j-1} X_j X_k (n\delta_{ij} - 1) (n\delta_{ik} - 1) \\
&= \frac{1}{n^3} \sum_{j=1}^n X_j^2 \left(\sum_{i=1}^n (n\delta_{ij} - 1)^2 \right) + \frac{2}{n^3} \sum_{j=2}^n \sum_{k=1}^{j-1} X_j X_k \left(\sum_{i=1}^n (n\delta_{ij} - 1) (n\delta_{ik} - 1) \right).
\end{aligned}$$

Comme un seul élément seulement dans la suite $\{(n\delta_{ij} - 1)^2 : i = 1, \dots, n\}$ vaut $(n-1)^2$ et les autres valent 1, on a

$$\sum_{i=1}^n (n\delta_{ij} - 1)^2 = (n-1)^2 + (n-1) = n(n-1).$$

De même, si $j \neq k$

$$\sum_{i=1}^n (n\delta_{ij} - 1)(n\delta_{ik} - 1) = n - 2 - 2(n-1) = -n.$$

On obtient donc

$$\hat{\theta}_n = \frac{1}{n^3} \times n(n-1) \sum_{j=1}^n X_j^2 + \frac{2}{n^3} \times (-n) \sum_{k < l} X_k X_l.$$

2. Déduire de la question précédente que $\hat{\theta}_n$ est biaisé alors que $\tilde{\theta}_n$ est sans biais.

Soit $\theta \in \Theta$,

$$\mathbf{E}_\theta(\hat{\theta}_n) = \frac{n-1}{n^2} \sum_{i=1}^n \mathbf{E}_\theta(X_i^2) - \frac{2}{n^2} \sum_{k < l} \mathbf{E}_\theta(X_k X_l) \quad \text{par linéarité de l'espérance.}$$

Comme les $(X_i)_{i=1}^n$ sont i.i.d. et $\mathbf{E}_\theta(X_1^2) = \mathbf{E}_{N \sim \mathcal{N}(0, \theta)}(N^2) = \theta$, on a

$$\sum_{i=1}^n \mathbf{E}_\theta(X_i^2) = \sum_{i=1}^n \mathbf{E}_\theta(X_1^2) = n\mathbf{E}_\theta(X_1^2) = n\theta,$$

et $k \neq l \implies X_k \perp\!\!\!\perp X_l \implies \mathbf{E}_\theta(X_k X_l) = \mathbf{E}_\theta(X_k) \mathbf{E}_\theta(X_l) = 0$ (car $= \mathbf{E}_{N \sim \mathcal{N}(0, \theta)}(N)$). Ainsi,

$$\mathbf{E}_\theta(\hat{\theta}_n) = \frac{n-1}{n^2} \times n\theta = \frac{n-1}{n}\theta.$$

Calculons le bias de $\hat{\theta}_n$. Le bias de $\hat{\theta}_n$ pour une valeur θ du paramètre s'exprime

$$B_\theta(\hat{\theta}_n) := b_n(\theta, \hat{\theta}_n) := \mathbf{E}_\theta(\hat{\theta}_n) - \theta = -\frac{1}{n}\theta.$$

C'est une fonction de θ ! On en déduit de l'estimateur $\hat{\theta}_n$ est biaisé négativement (et le bias décroît lorsque $n \rightarrow \infty$). Calculons à présent le bias de $\tilde{\theta}_n$. On remarque d'abord que

$$\tilde{\theta}_n = \frac{n}{n-1} \hat{\theta}_n.$$

Soit $\theta \in \Theta$. Par linéarité de l'espérance,

$$\begin{aligned}\mathbf{E}_\theta(\tilde{\theta}_n) &= \frac{n}{n-1} \mathbf{E}(\hat{\theta}_n) - \theta^* \\ &= \frac{n}{n-1} \frac{n-1}{n} \theta \\ &= \theta.\end{aligned}$$

On en déduit

$$\forall \theta \in \Theta, B_\theta(\tilde{\theta}_n) = \theta - \theta = 0.$$

L'estimateur $\tilde{\theta}_n$ est sans biais.

3. On pose

$$Y = \sum_{i=1}^n X_i^2, \quad Z = \sum_{k<\ell} X_k X_\ell.$$

Montrer que

$$\mathbf{E}[YZ] = \mathbf{Cov}(Y, Z) = 0 \quad \text{et} \quad \mathbf{E}[Z^2] = \frac{n(n-1)}{2} \sigma^4.$$

Soit $\theta \in \Theta$. D'après les calculs précédents, on sait déjà que

$$\mathbf{E}_\theta(Z) = 0.$$

Donc

$$\mathbf{Cov}_\theta(Y, Z) := \mathbf{E}_\theta(YZ) - \mathbf{E}_\theta(Y)\mathbf{E}_\theta(Z) = \mathbf{E}_\theta(YZ).$$

Ensuite,

$$\mathbf{E}_\theta(YZ) = \sum_{i=1}^n \sum_{k=2}^n \sum_{l=1}^{k-1} \mathbf{E}_\theta(X_i^2 X_k X_l).$$

- Si $i = k$, alors $\mathbf{E}_\theta(X_i^2 X_l X_k) = \mathbf{E}_\theta(X_k^3 X_l) \stackrel{(\text{H})}{=} \mathbf{E}_\theta(X_k^3) \mathbf{E}_\theta(X_l) = 0$;
- Si $i = l$, alors de manière analogue $\mathbf{E}_\theta(X_i^2 X_k X_l) = \mathbf{E}_\theta(X_l^3) \mathbf{E}_\theta(X_k) = 0$;
- Si $i \neq k$ et $i \neq l$, alors

$$\mathbf{E}_\theta(X_i^2 X_k X_l) = \mathbf{E}_\theta(X_i^2) \mathbf{E}_\theta(X_k) \mathbf{E}_\theta(X_l) = 0.$$

Donc

$$\mathbf{E}_\theta(YZ) = 0.$$

On a

$$\begin{aligned}\mathbf{E}_\theta(Z^2) &= \sum_{k < l} \sum_{i < j} \mathbf{E}_\theta(X_k X_l X_i X_j) \quad (\text{tous les doubles produits}) \\ &= \sum_{k < l} \mathbf{E}_\theta(X_k X_l X_k X_l) + \sum_{k < l} \sum_{i < j} \mathbf{1}\{(i, j) \neq (k, l)\} \mathbf{E}_\theta(X_k X_l X_i X_j).\end{aligned}$$

Pour $k \neq l$, $\mathbf{E}_\theta(X_k^2 X_l^2) \stackrel{(\perp)}{=} \mathbf{E}_\theta(X_k^2) \mathbf{E}_\theta(X_l^2) = \theta \times \theta = \theta^2$. A présent, pour $(i, j) \neq (k, l)$ avec $i \neq j$ et $k \neq l$, on a soit

- $i \neq k$. Et dans ce cas, soit $i \neq l$ et

$$\mathbf{E}_\theta(X_k X_l X_i X_j) \stackrel{(\perp)}{=} \underbrace{\mathbf{E}_\theta(X_i)}_{=0} \mathbf{E}_\theta(X_k X_l X_j) = 0.$$

Soit $i = l$ avec $i \neq j$ et $l \neq k$ et

$$\mathbf{E}_\theta(X_k X_l X_i X_j) = \mathbf{E}_\theta(X_k X_l^2 X_j) \stackrel{(\perp)}{=} \mathbf{E}_\theta(X_k) \mathbf{E}(X_l^2 X_j) = 0.$$

- $j \neq l$, etc.
- $j \neq l$ et $i \neq k$, etc.

D'où, par linéarité de l'espérance,

$$\mathbf{E}_\theta(Z^2) = \sum_{k < l} \theta^2 = \frac{n(n-1)}{2} \theta^2.$$

En effet, pour calculer $\sum_{k < l} 1$, on peut soit

- Chercher à répondre à la question “combien de choix de deux éléments distincts parmi $\{1, \dots, n\}$ sont possibles?” (car l'ordre est imposé). On trouve

$$\frac{A_n^2}{2} = \frac{n(n-1)}{2} = C_n^2 = \binom{n}{2} = \frac{n(n-1)}{2}.$$

- Faire le calcul :

$$\begin{aligned}\sum_{k < l} &= \sum_{l=2}^n \underbrace{\sum_{k=1}^{l-1}}_{l-1 \text{ termes dans cette somme}} \\ &= \sum_{l=2}^n (l-1) \\ &= \sum_{l=1}^{n-1} l = \frac{(n-1)n}{2}.\end{aligned}$$

4. Déduire de la question précédente que

$$\mathbf{Var}(\hat{\theta}_n) = \frac{(n-1)^2}{n^3} \left(\mathbf{E}[X_1^4] - \frac{n-3}{n-1} \sigma^4 \right).$$

Est-ce qu'on a utilisé le fait que X_1 est de loi gaussienne ?

On a $\hat{\theta}_n = \frac{n-1}{n^2}Y - \frac{2}{n^2}Z$ et $\mathbf{Cov}_{\theta}(Y, Z) = 0$ donc, pour tout $\theta \in \Theta$,

$$\begin{aligned}\mathbf{Var}_{\theta}(\hat{\theta}_n) &= \mathbf{Var}_{\theta}\left(\frac{n-1}{n^2}Y - \frac{2}{n^2}Z\right) \\ &= \mathbf{Var}_{\theta}\left(\frac{n-1}{n^2}Y\right) + \mathbf{Var}_{\theta}\left(-\frac{2}{n^2}Z\right) \\ &= \frac{(n-1)^2}{n^4}\mathbf{Var}_{\theta}(Y) + \frac{4}{n^4}\mathbf{Var}_{\theta}(Z).\end{aligned}$$

Or

$$\begin{aligned}\mathbf{Var}_{\theta}(Y) &= n \times \mathbf{Var}_{\theta}(X_1^2) \quad \text{car } (X_i)_{i=1}^n \text{ i.i.d.} \\ &= n \times \left(\mathbf{E}_{\theta}(X_1^4) - \theta^2\right)\end{aligned}$$

et

$$\begin{aligned}\mathbf{Var}_{\theta}(Z) &= \mathbf{E}_{\theta}(Z^2) \quad \text{car } Z^2 \text{ est centré} \\ &= \frac{n(n-1)}{2}\theta^2.\end{aligned}$$

D'où,

$$\begin{aligned}\mathbf{Var}_{\theta}(\hat{\theta}_n) &= \frac{(n-1)^2}{n^4} \times n \times \left[\mathbf{E}_{\theta}(X_1^4) - \theta^2\right] + \frac{4}{n^4} \times \frac{n(n-1)}{2}\theta^2 \\ &= \frac{(n-1)^2}{n^3} \left[\mathbf{E}_{\theta}(X_1^4) - \theta^2 + \frac{2}{n-1}\theta^2\right] \\ &= \frac{(n-1)^2}{n^3} \left[\mathbf{E}_{\theta}(X_1^4) - \frac{n-3}{n-1} \underbrace{\theta^2}_{:=(\sigma^2)^2=\sigma^4}\right].\end{aligned}$$

On a utilisé que les X_i sont centrés mais pas l'hypothèse de gaussianité.

5. Calculer les risques des estimateurs $\hat{\theta}_n$ et $\tilde{\theta}_n$ dans le cas gaussien. Comparer les.

On s'intéresse ici au risque quadratique ("Mean Squared Error") :

$$\begin{aligned}R_n(\theta, \hat{\theta}_n) &= \mathbf{E}_{\theta}[(\hat{\theta}_n - \theta)^2] \\ &= \mathbf{Var}_{\theta}(\hat{\theta}_n) + [\mathbf{E}_{\theta}(\hat{\theta}_n - \theta)]^2 \\ &= \mathbf{Var}_{\theta}(\hat{\theta}_n) + b_n(\theta, \hat{\theta}_n)^2.\end{aligned}$$

Ici, on obtient

$$R_n(\theta, \hat{\theta}_n) = \frac{(n-1)^2}{n^3} \left[\mathbf{E}_{\theta}(X_1^4) - \frac{n-3}{n-1}\sigma^4\right] + \underbrace{\left(-\frac{1}{n}\sigma^2\right)^2}_{\frac{\sigma^4}{n^2}}.$$

Sous l'hypothèse de gaussianité (le modèle considéré), on a $X_1 \sim \mathcal{N}(0, \theta)$, autrement dit $X_1 \stackrel{d}{=} \theta^{1/2} N$ où $N \sim \mathcal{N}(0, 1)$ et on déduit

$$\mathbf{E}_\theta(X_1^4) = \mathbf{E}\left[\left(\theta^{1/2}N\right)^4\right] = \theta^2 \underbrace{\mathbf{E}\left(N^4\right)}_{=3} = 3\theta^2 = 3\sigma^4.$$

D'où

$$\begin{aligned} R_n(\theta, \hat{\theta}_n) &= \frac{\sigma^4}{n^2} \left[\frac{3(n-1)^2}{n} - \frac{(n-1)^2}{n} \times \frac{n-3}{n-1} + 1 \right] \\ &= \frac{\sigma^4}{n^2} \left[1 + \frac{3(n-1)^2}{n} - \frac{(n-1)(n+3)}{n} \right] \\ &= \frac{\sigma^4}{n^2} [1 + 2(n-1)] \\ &= \frac{\sigma^4}{n^2} (2n-1). \end{aligned}$$

Comme $\tilde{\theta}_n$ est sans bias, on a

$$\begin{aligned} R_n(\theta, \tilde{\theta}_n) &= \mathbf{Var}_\theta(\tilde{\theta}_n) \\ &= \frac{n^2}{(n-1)^2} \mathbf{Var}_\theta(\hat{\theta}_n) \\ &= \frac{n^2}{(n-1)^2} \times \frac{(n-1)^2}{n^3} \left[3\sigma^4 - \frac{n-3}{n-1}\sigma^4 \right] \\ &= \frac{\sigma^4}{n} \left[3 - \frac{n-3}{n-1} \right] \\ &= \frac{2\sigma^4}{n-1}. \end{aligned}$$

On peut vérifier que $\frac{2}{n-1} > \frac{2n-1}{n^2}$ pour tout $n > 1$. Ainsi $\hat{\theta}_n$ (non débiaisé) est préférable à $\tilde{\theta}_n$ (débiaisé) en termes de risque quadratique.

Exercice 2. Soient X_1, \dots, X_n des variables aléatoires i.i.d. de loi uniforme sur $[0, \theta^*]$ où $\theta^* > 0$ est un paramètre inconnu. On considère l'estimateur

$$\hat{\theta}_n = \max_{i=1, \dots, n} X_i.$$

1. Calculer la fonction de répartition $F_{n,\theta}$ de la variable $\hat{\theta}_n$ lorsque les X_i sont de loi $\mathcal{U}([0, 1])$.

Modèle statistique :

$$(\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n), \{\mathbf{P}_\theta = \mathcal{U}([0, \theta])^{\otimes n}, \theta \in \Theta := \mathbf{R}_+^*\}).$$

On a

$$\begin{aligned}
F_{n,\theta}(x) &= \mathbf{P}_\theta \left(\max_{1 \leq i \leq n} X_i \leq x \right) \\
&= \mathbf{P}_\theta \left(\bigcap_{i=1}^n \{X_i \leq x\} \right) \\
&\stackrel{(II)}{=} \prod_{i=1}^n \mathbf{P}_\theta(X_i \leq x) \\
&= \prod_{i=1}^n \left(\frac{x}{\theta} \mathbb{1}_{[0,\theta]}(x) + \mathbb{1}_{[\theta,+\infty[}(x) \right) \quad \text{car } X_i \stackrel{\mathbf{P}_\theta}{\sim} \mathcal{U}([0,\theta]) \text{ pour tout } i \in \{1, \dots, n\} \\
&= \left(\frac{x}{\theta} \right)^n \mathbb{1}_{[0,\theta]}(x) + \mathbb{1}_{[\theta,+\infty[}(x).
\end{aligned}$$

2. Montrer que pour tout $\delta > 0$, on a

$$\mathbf{P}_\theta(|\hat{\theta}_n - \theta| \geq \delta) = \mathbf{P}_\theta(\hat{\theta}_n \leq \theta - \delta) \xrightarrow[n \rightarrow \infty]{} 0.$$

En déduire que $\hat{\theta}_n$ est un estimateur faiblement consistant de θ^* .

Soit $\delta > 0$. Comme $\hat{\theta}_n \leq \theta$ presque-sûrement si $X_1, \dots, X_n \sim \mathcal{U}([0, \theta])$, on a

$$\begin{aligned}
\mathbf{P}_\theta(|\hat{\theta}_n - \theta| \geq \delta) &= \mathbf{P}_\theta(\hat{\theta}_n - \theta \leq -\delta) \\
&= \mathbf{P}_\theta(\hat{\theta}_n \leq \theta - \delta) \\
&= F_{n,\theta}(\theta - \delta) \\
&= \left(\frac{\theta - \delta}{\theta} \right)^n \\
&= \left(1 - \frac{\delta}{\theta} \right)^n,
\end{aligned}$$

où la dernière égalité est vraie pour $\delta \in]0, \theta[$, et implique pour ces δ

$$\mathbf{P}_\theta(|\hat{\theta}_n - \theta| \geq \delta) \xrightarrow[n \rightarrow +\infty]{} 0.$$

Donc $\hat{\theta}_n \xrightarrow{P} \theta^*$, i.e., $\hat{\theta}_n$ est faiblement consistant.

3. En utilisant un argument de monotonie, prouver que cet estimateur est également fortement consistant.

La suite $\{\hat{\theta}_n(\omega)\}_{n \in \mathbb{N}^*}$ est croissante pour tout $\omega \in \Omega$. De plus, elle est bornée par θ^* p.s. Donc, pour tous les $\omega \in \Omega_0$, avec $\mathbf{P}(\Omega_0) = 1$, on sait que la suite $\{\hat{\theta}_n(\omega)\}_{n \in \mathbb{N}^*}$ converge. Soit $\hat{\theta}_\infty(\omega)$ la limite de cette suite (on peut poser $\hat{\theta}_\infty(\omega) = 0$ si $\omega \notin \Omega_0$). On sait que $\hat{\theta}_n \xrightarrow{p.s.} \hat{\theta}_\infty$. Donc $\hat{\theta}_n \xrightarrow{P} \hat{\theta}_\infty$.

Il en découle (cf. Exercice 1 , TD 1) que $\hat{\theta}_\infty = \theta^*$ p.s. Par conséquent,

$$\hat{\theta}_n \xrightarrow{p.s.} \theta^*.$$

Donc $\hat{\theta}_n$ est fortement consistant.

4. Montrer que la suite $\{n(\theta^* - \hat{\theta}_n)\}$ converge en loi et déterminer sa limite.

Dans la question 2), en prenant $\delta = x/n$, il vient

$$\begin{aligned} \mathbf{P}_{\theta^*}(n(\theta^* - \hat{\theta}_n) \leq x) &= \mathbf{P}_{\theta^*}\left(\hat{\theta}_n \geq \theta^* - \frac{x}{n}\right) \\ &= \begin{cases} 1 & \text{si } x > n\theta^*, \\ 1 - (1 - x/n\theta^*) & \text{si } x \in (0, n\theta^*], \\ 0 & \text{si } x < 0. \end{cases} \end{aligned}$$

Lorsque θ^* et x sont fixes et $n \rightarrow +\infty$, on a

$$\left(1 - \frac{x}{n\theta^*}\right)^n = \exp\left(n \ln\left(1 - \frac{x}{n\theta^*}\right)\right) \sim \exp\left(-\frac{x}{\theta^*}\right).$$

On a donc

$$\mathbf{P}_{\theta^*}(n(\theta^* - \hat{\theta}_n) \leq x) \xrightarrow{n \rightarrow +\infty} \left(1 - e^{-\frac{x}{\theta^*}}\right) \mathbb{1}_{[0,+\infty]}(x).$$

Autrement dit, $n(\theta^* - \hat{\theta}_n) \xrightarrow{\mathcal{L}} \mathcal{E}(1/\theta^*)$.

Exercice 3. (facultatif) Soient $\{X_i : i \in \mathbb{N}^*\}$ une suite de variables iid de fonction de répartition F . On pose $\theta = F(0)$. Soit N une variable aléatoire de loi de Poisson de paramètre $\lambda > 0$ et indépendante de $\{X_i : i \in \mathbb{N}^*\}$. On pose

$$N_1 = \sum_{i=1}^N \mathbb{1}(X_i \leq 0), \quad N_2 = \sum_{i=1}^N \mathbb{1}(X_i > 0).$$

Montrer que N_1 et N_2 sont deux variables aléatoires indépendantes et que toutes les deux sont de loi de Poisson dont on déterminera les paramètres.

Soit $\theta \in \Theta := [0, 1]$. On a $\text{Supp}(N_1) = \text{Supp}(N_2) = \text{Supp}(N) = \mathbb{N}$. Ainsi, pour tout $k \in \mathbb{N}$

$$\begin{aligned}
\mathbf{P}_\theta(N_1 = k) &= \sum_{j=0}^{+\infty} \mathbf{P}_\theta \left(\{N_1 = k\} \cap \{N = j\} \right) \\
&= \sum_{j=k}^{+\infty} \mathbf{P}_\theta \left(\left\{ \sum_{i=1}^j \mathbb{1}(X_i \leq 0) = k \right\} \cap \{N = j\} \right) \\
&\stackrel{(1)}{=} \sum_{j=k}^{+\infty} \mathbf{P}_\theta \left(\underbrace{\sum_{i=1}^j \mathbb{1}(X_i \leq 0)}_{\text{v.a. de loi Binomiale}(j, \theta)} = k \right) \mathbf{P}(N = j) \\
&= \sum_{j=k}^{+\infty} \binom{j}{k} \theta^k (1 - \theta)^{j-k} \times \frac{\lambda^j}{j!} e^{-\lambda} \\
&= \frac{e^{-\lambda}}{k!} \sum_{j=k}^{+\infty} \frac{j!}{(j-k)!} \theta^k (1 - \theta)^{j-k} \frac{\lambda^j}{j!} \\
&= \frac{(\lambda\theta)^k}{k!} e^{-\lambda} \sum_{j=k}^{+\infty} \frac{(\lambda(1-\theta))^{j-k}}{(j-k)!} \\
&= \frac{(\lambda\theta)^k}{k!} e^{-\lambda} \sum_{i=0}^{+\infty} \frac{(\lambda(1-\theta))^i}{i!} \\
&= \frac{(\lambda\theta)^k}{k!} e^{-\lambda} e^{\lambda(1-\theta)} \\
&= \frac{(\lambda\theta)^k}{k!} e^{-\lambda\theta}.
\end{aligned}$$

On en déduit que $N_1 \sim \text{Poisson}(\lambda\theta)$. On prouve de la même manière que $N_2 \sim \text{Poisson}(\lambda(1-\theta))$. Pour prouver que $N_1 \perp\!\!\!\perp N_2$, on peut calculer les probabilités jointes en remarquant que, pour tout $(k, j) \in \mathbb{N}^2$,

$$\begin{aligned}
\mathbf{P}_\theta \left(\{N_1 = k\} \cap \{N_2 = j\} \right) &= \mathbf{P}_\theta \left(\{N_1 = k\} \cap \{N = k+j\} \right) \\
&= \binom{k+j}{k} \theta^k (1 - \theta)^j \frac{\lambda^{k+j}}{(k+j)!} e^{-\lambda} \\
&= \frac{(\theta\lambda)^k}{k!} e^{-(\theta\lambda)} \times \frac{((1-\theta)\lambda)^j}{j!} e^{-(1-\theta)\lambda} \\
&= \mathbf{P}_\theta(N_1 = k) \mathbf{P}_\theta(N_2 = j).
\end{aligned}$$

On a bien $N_1 \perp\!\!\!\perp N_2$.

TD4 : MÉTHODE DES MOMENTS

Exercice 1. Soient X_1, \dots, X_n des variables indépendantes de densité

$$f(x; \theta) = \frac{1}{a} e^{-(x-b)/a} \mathbb{1}(x \geq b), \quad \forall x \in \mathbb{R},$$

où $\theta = (a, b) \in]0, +\infty[\times \mathbb{R}$ est le paramètre inconnu.

1. Calculer $\mu_1(\theta) = \mathbf{E}_\theta[X_1]$ et $\mu_2(\theta) = \mathbf{E}[X_1^2]$.

Pour tout $x \in \mathbb{R}$,

$$\frac{dP^{X_1}}{d\lambda}(x) =: f(x; \theta) =: \frac{1}{a} e^{-(x-b)/a} \mathbb{1}(x \geq b),$$

où λ désigne la mesure de Lebesgue. Il s'agit de la loi exponentielle translatée (de paramètre de location b et de paramètre d'échelle a). On la note $\mathcal{E}(a, b)$ dans ce qu'il suit.

Rappel loi exponentielle $E \sim \mathcal{E}(u)$, $u > 0$.

$$\begin{aligned} \text{Si paramétrée par } u, \quad \frac{dP^E}{d\lambda} &= ue^{-ux} \mathbb{1}(x \geq 0), \\ \text{Si paramétrée par } a \text{ tel que } u = \frac{1}{a} &= \frac{1}{a} e^{-\frac{x}{a}} \mathbb{1}(x \geq 0). \end{aligned}$$

On a

$$\begin{cases} E_u(E^k) = \frac{k!}{u^k}, & E_a(E^k) = a^k k!, \\ E_u(E) = \frac{1}{u}, & E_a(E) = a, \\ \mathbf{Var}_u(E) = \frac{1}{u^2}, & \mathbf{Var}_a(E) = a^2, \\ F_u(x) := P_u(E \leq x) = 1 - e^{-ux}, & F_a := P_a(E \leq x) = 1 - e^{-\frac{1}{a}x}. \end{cases}$$

Le modèle statistique est ici

$$(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \{P_\theta = \mathcal{E}(a, b)^{\otimes n}, \theta := (a, b) \in \Theta := \mathbb{R}_+^* \times \mathbb{R}\}).$$

Ce modèle est

- paramétrique : $\Theta \subset \mathbb{R}^2$,
- identifiable,
- dominé par rapport à la mesure de Lebesgue.

On propose deux méthodes pour calculer $\mu_1(\theta)$ et $\mu_2(\theta)$.

Méthode 1 En remarquant que

$$X_1 = aY_1 + b, \quad Y_1 \sim \mathcal{E}(1).$$

En effet, pour tout $x \in \mathbf{R}_+$,

$$\mathbf{P}(aY_1 + b \leq x) = \mathbf{P}\left(Y_1 \leq \frac{x-b}{a}\right) = F_1\left(\frac{x-b}{a}\right).$$

En utilisant la règle de Leibniz pour l'intégration, $aY_1 + b$ admet donc pour densité par rapport à la mesure de Lebesgue

$$f_{aY_1+b}(x, \theta) = \frac{1}{a} F'_1\left(\frac{x-b}{a}\right) = f(x; \theta) \quad \lambda - p.p.$$

D'où, pour tout $\theta \in \Theta$,

$$\begin{aligned} \mu_1(\theta) &:= \mathbf{E}_\theta(X_1) = \mathbf{E}_\theta(aY_1 + b) \\ &= a \underbrace{\mathbf{E}_\theta(Y_1)}_{=1} + b \quad \text{par linéarité de } \mathbf{E}_\theta(\cdot) \\ &= a + b, \end{aligned}$$

et

$$\begin{aligned} \mu_2(\theta) &:= \mathbf{E}_\theta(X_1^2) = \mathbf{E}_\theta[(aY_1 + b)^2] \\ &= a^2 \underbrace{\mathbf{E}_\theta(Y_1^2)}_{=2} + 2ab\mathbf{E}_\theta(Y_1) + b^2 \quad \text{par linéarité de } \mathbf{E}_\theta(\cdot) \\ &= 2a^2 + 2ab + b^2. \end{aligned}$$

Méthode 2 Par un calcul d'intégrales et un changement de variables. Soit $\theta \in \Theta$,

$$\begin{aligned} \mu_1(\theta) &= \int_{\mathbf{R}} xf(x; \theta) \lambda(dx) \\ &= \int_{\mathbf{R}} \frac{1}{a} e^{-(x-b)/a} \mathbb{1}(x \geq b) dx \\ &= \frac{1}{a} \int_b^{+\infty} e^{-(x-b)/a} dx \\ &= \frac{1}{a} \int_0^{+\infty} (ay + b)e^{-y} a dy \quad \text{en posant } y = \frac{x-b}{a}, \quad x = ay + b, \quad \frac{dy}{dx} = \frac{1}{a} \\ &= b \underbrace{\int_0^{+\infty} e^{-y} dy}_{\Gamma(1)=1=[-e^{-y}]_0^{+\infty}} + a \underbrace{\int_0^{+\infty} ye^{-y} dy}_{\Gamma(2)=1!=1} \\ &= a + b. \end{aligned}$$

Rappel fonction Gamma : $\Gamma(z) = \int_0^{+\infty} x^{z-1} e^{-x} dx$, $z \in \mathbf{R}$, et on a

$$\Gamma(z+1) = z\Gamma(z) \text{ pour } z > 0,$$

$$\Gamma(1) = 1,$$

$$\Gamma(n+1) = n! \text{ pour tout entier } n \in \mathbb{N}.$$

De même, on obtient

$$\begin{aligned}
\mu_2(\theta) &= \int_0^\infty e^{-y} (ay + b)^2 dy \quad \text{avec le même changement de variable} \\
&= a^2 \underbrace{\int_0^{+\infty} y^2 e^{-y} dy}_{\Gamma(3)=2!=2} + b^2 \underbrace{\int_0^{+\infty} e^{-y} dy}_{\Gamma(1)=1} + 2ab \underbrace{\int_0^{+\infty} ye^{-y} dy}_{\Gamma(2)=1} \\
&= 2a^2 + 2ab + b^2.
\end{aligned}$$

2. Déterminer l'estimateur par la méthode des moments $\hat{\theta}_n^{\text{MM}} = (\hat{a}_n^{\text{MM}}, \hat{b}_n^{\text{MM}})$ du paramètre θ .

Notons les contreparties empiriques de $\mu_r(\theta)$, $r \in \{1, 2\}$,

$$m_{n,r} := \frac{1}{n} \sum_{i=1}^n (X_i)^r.$$

On pose donc

$$\begin{aligned}
m_{n,1} &:= \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i, \\
m_{n,2} &:= \bar{X}^2_n = \frac{1}{n} \sum_{i=1}^n (X_i)^2.
\end{aligned}$$

Méthode pour déterminer l'estimateur de la méthode des moments (EMM) : on cherche $\hat{\theta}_n^{\text{MM}} := (\hat{a}, \hat{b}) \in \Theta$ tel que

$$\mu_r(\hat{\theta}_n^{\text{MM}}) = m_{n,r}, \quad r = 1, 2.$$

C'est-à-dire, tel que

$$\begin{aligned}
\begin{cases} \mu_1(\hat{\theta}_n^{\text{MM}}) = \hat{a} + \hat{b} = \bar{X}_n \\ \mu_2(\hat{\theta}_n^{\text{MM}}) = 2\hat{a}^2 + 2\hat{a}\hat{b} + \hat{b}^2 = \bar{X}^2_n \end{cases} &\iff \begin{cases} \hat{a} + \hat{b} = \bar{X}_n \\ (\hat{a} + \hat{b})^2 + \hat{a}^2 = \bar{X}^2_n \end{cases} \\
&\iff \begin{cases} \hat{a}^2 = \bar{X}^2_n - (\bar{X}_n)^2 \\ \hat{b} = \bar{X}_n - \hat{a} \end{cases} \quad \hat{a} \in \mathbf{R}_+^* \\
&\iff \begin{cases} \hat{a} = \sqrt{\bar{X}^2_n - (\bar{X}_n)^2} \\ \hat{b} = \bar{X}_n - \sqrt{\bar{X}^2_n - (\bar{X}_n)^2} \end{cases}.
\end{aligned}$$

Or

$$\begin{aligned}
\bar{X}^2_n - (\bar{X}_n)^2 &= \frac{1}{n} \sum_{i=1}^n (X_i)^2 - (\bar{X}_n)^2 \\
&= \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \\
&=: S_n \quad \text{l'estimateur biaisé de la variance.}
\end{aligned}$$

On a donc une unique solution au système :

$$\hat{\theta}_n^{\text{MM}} = \left(\sqrt{S_n}, \bar{X}_n - \sqrt{S_n} \right).$$

3. Montrer que $\hat{\theta}_n^{\text{MM}}$ est asymptotiquement normal et déterminer sa variance limite.

On va utiliser le cas général du résultat du TD2, exercice 3, lorsque X_1 n'est pas centré. D'après le TD2, exercice 3 appliqué à $X_i - \mu_1$ pour avoir des variables centrées, on a pour tout $\theta \in \Theta$,

$$\sqrt{n}(S_n - \mathbf{Var}_\theta(X_1)) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \mathbf{Var}_\theta([X_1 - \mu_1]^2)\right).$$

Rq : la variance asymptotique était seulement égale à $\mathbf{Var}_\theta(X_1^2)$ dans le TD2, mais ici $\mathbf{Var}_\theta([X_1 - \mu_1]^2)$ car X_1 n'est pas forcément centrée.

$$\mathbf{Var}_\theta([X_1 - \mu_1]^2) = \mathbf{E}_\theta([X_1 - \mu_1]^4) - \underbrace{\left\{ \mathbf{E}_\theta([X_1 - \mu_1]^2) \right\}}_{=\mathbf{Var}_\theta(X_1)=a^2}^2.$$

En effet,

$$\begin{aligned} \mathbf{Var}_\theta(X_1) &= \mathbf{E}_\theta(X_1^2) - \mathbf{E}_\theta(X_1)^2 \\ &= \mu_2(\theta) - \mu_1(\theta)^2 \\ &= a^2 + (a+b)^2 - (a+b)^2 \\ &= a^2. \end{aligned}$$

D'où,

$$\begin{aligned} \mathbf{Var}_\theta([X_1 - \mu_1]^2) &= \mathbf{E}_\theta\left(\left[X_1 - \underbrace{\mu_1}_{\mu_1=a+b}\right]^4\right) - a^4 \\ &= \mathbf{E}_\theta([X_1 - b]^4) - a^4 \\ &= \mathbf{E}_\theta([X_1 - b]^4) - 4a\mathbf{E}_\theta([X_1 - b]^3) + 6a^2\mathbf{E}_\theta([X_1 - b]^2) \\ &\quad - 4a^3\mathbf{E}_\theta(X_1 - b) + a^4 - a^4 \quad \text{par linéarité de } \mathbf{E}_\theta(\cdot) \text{ et formule du binôme.} \end{aligned}$$

Or on a pour tout $z \in \mathbb{R}$,

$$\begin{aligned}
\mathbf{P}_\theta(X_1 - b \leq z) &= \mathbf{P}_\theta(X_1 \leq b + z) \\
&= \int_{-\infty}^{b+z} f(x; \theta) dx \\
&= \int_{-\infty}^z \frac{1}{a} e^{-\frac{(x-b)}{a}} \mathbb{1}(x \geq b) dx \\
&= \int_{-\infty}^z \frac{1}{a} e^{-\frac{z}{a}} \mathbb{1}(y \geq 0) dx \\
&= 1 - e^{-\frac{z}{a}}.
\end{aligned} \tag{4}$$

D'où $X_1 - b \stackrel{D}{=} \mathcal{E}(a)$ d'espérance a . Ainsi, en utilisant $\mathbf{E}_\theta([X_1 - b]^k) = k!a^k$, on a

$$\begin{aligned}
\mathbf{Var}_\theta([X_1 - \mu_1]^2) &= (4!a^4) - 4a(3!a^3) + 6a^2(2!a^2) - 4a^3(1!a) \\
&= a^4 [24 - 24 + 12 - 4] \\
&= 8a^4.
\end{aligned}$$

Donc, pour tout $\theta \in \Theta$

$$\sqrt{n}(S_n - a^2) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 8a^4).$$

Pour conclure, il faut utiliser le deuxième théorème de continuité (" δ -méthode") appliqué à la fonction $g : x \mapsto \sqrt{x}$ de \mathbb{R}_+ dans \mathbb{R} . g est C^1 sur \mathbb{R}_+^* (cela suffit car $a \in \mathbb{R}_+^*$ par définition de Θ), $g'(x) = 1/(2\sqrt{x})$ et donc $g'(a^2) = 1/2a \neq 0$ car $a \in \mathbb{R}_+^*$ par hypothèse. La δ -méthode donne

$$\sqrt{n}(g(S_n^2) - g(a^2)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 2a^2),$$

autrement dit, pour tout $\theta \in \Theta$

$$\sqrt{n}(\hat{a}_n^{\text{MM}} - a) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 2a^2).$$

On a montré la normalité asymptotique de \hat{a}_n^{MM} .

4. (Facultative) Quelle est la loi de la variable $Y_1 = (X_1 - b)/a$?

Déjà montré précédemment. Peut aussi se voir d'après l'égalité (4) en posant $z = za$. On conclut $Y_1 \sim \mathcal{E}(1)$.

Exercice 2. (*Modèle de mélange.*) Soient X_1, \dots, X_n des variables aléatoires i.i.d. dont la densité f est un mélange de deux densités gaussiennes $\mathcal{N}(0, 1)$ et $\mathcal{N}(0, 4)$:

$$f(x; \theta) = \frac{\theta}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) + \frac{(1-\theta)}{2\sqrt{2\pi}} \exp\left(-\frac{x^2}{8}\right),$$

où $\theta \in]0, 1[$ est un paramètre inconnu que l'on souhaite estimer. Expliciter $\hat{\theta}_n^{\text{MM}}$, l'estimateur de θ

obtenu à l'aide de la méthode des moments. Montrer que l'estimateur $\hat{\theta}_n$ est consistant et déterminer la loi limite de $\sqrt{n}(\hat{\theta}_n^{\text{MM}} - \theta)$ lorsque $n \rightarrow \infty$.

On peut noter " $X_1 \sim \theta\mathcal{N}(0,1) + (1-\theta)\mathcal{N}(0,4)$ ", ce qui signifie :

$$X_1 \stackrel{d}{=} DY + (1-D)Z,$$

où $D, Y, Z \perp\!\!\!\perp$ et

$$\begin{aligned} D &\sim \text{Ber}(\theta), \\ Y &\sim \mathcal{N}(0,1), \\ Z &\sim \mathcal{N}(0,4). \end{aligned}$$

Le modèle statistique est ici

$$\left(\mathbf{R}^n, \underbrace{\mathcal{B}(\mathbf{R}^n)}_{\text{loi continue}}, \left\{ P_\theta = \{\theta\mathcal{N}(0,1) + (1-\theta)\mathcal{N}(0,4)\}^{\otimes n}, \theta \in \Theta :=]0,1[\right\} \right).$$

Calcul de $\hat{\theta}_n^{\text{MM}}$, l'estimateur par la méthode des moments. On peut commencer par calculer $\mu_1(\theta) = \mathbf{E}_\theta(X_1)$. Pour tout $\theta \in \Theta$,

$$\begin{aligned} \mathbf{E}_\theta(X_1) &= \mathbf{E}_\theta(DY + (1-D)Z) & \text{ou } &= \theta \int_{\mathbf{R}} x\phi(x) dx + (1-\theta) \frac{1}{2} \int_{\mathbf{R}} x\phi\left(\frac{x}{2}\right) dx \\ &= \mathbf{E}_\theta \{ \mathbf{E}_\theta [DY + (1-D)Z | D] \} & &= \theta \times 0 + (1-\theta) \times 0 \\ &\stackrel{D \perp\!\!\!\perp (Y,Z)}{=} \mathbf{E}_\theta \left\{ D \underbrace{\mathbf{E}_\theta(Y)}_{=0} + (1-D) \underbrace{\mathbf{E}_\theta(Z)}_{=0} \right\} \\ &= 0 & &= 0. \end{aligned}$$

Donc on ne peut pas utiliser le moment d'ordre 1, $\mu_1(\theta)$, pour déterminer un EMM. Solution ?

- "Au lieu d'utiliser les k premiers moments pour construire l'EMM de $\theta \in \mathbf{R}^k$, on peut utiliser k moments quelconques $\mu_{r_1}, \dots, \mu_{r_k}$, pourvu qu'ils soient finis" : EMM avec un autre moment (cf. polycopié de Tsybakov, p.105).
- méthode des moments généralisés.

Ici, $\dim(\Theta) = 1$, donc on a besoin d'un seul moment. Regardons ce que donne le moment d'ordre 2. Soit $\theta \in \Theta$,

$$\begin{aligned} \mu_2(\theta) &:= \mathbf{E}_\theta(X_1^2) = \int_{\mathbf{R}} x^2 f(x; \theta) dx \quad \text{par la formule de transfert et par définition de } \frac{dP^{X_1}}{d\lambda} \\ &= \theta \underbrace{\int_{\mathbf{R}} \frac{x^2}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx}_{=\mathbf{E}(N^2) \text{ où } N \sim \mathcal{N}(0,1)} + (1-\theta) \underbrace{\int_{\mathbf{R}} \frac{x^2}{2\sqrt{2\pi}} e^{-\frac{x^2}{8}} dx}_{=\mathbf{E}(N^2) \text{ où } N \sim \mathcal{N}(0,4)} \\ &= \theta \times 1 + (1-\theta) \times 4 \\ &= 4 - 3\theta. \end{aligned}$$

On résout en $\hat{\theta}$:

$$\begin{aligned} m_{n,2} &:= \frac{1}{n} \sum_{i=1}^n X_i^2 = \mu_2(\hat{\theta}_n^{\text{MM}}) \\ \iff 4 - 3\hat{\theta}_n^{\text{MM}} &= \overline{X^2}_n \\ \iff \hat{\theta}_n^{\text{MM}} &= \frac{4}{3} - \frac{1}{3}\overline{X^2}_n = \frac{4}{3} - \frac{1}{3n} \sum_{i=1}^n X_i^2. \end{aligned}$$

Consistance de $\hat{\theta}_n^{\text{MM}}$. On utilise la Loi Forte des Grands Nombres (LGFN) et le Continuous Mapping Theorem (CMT). On a $(X_i)_{i=1}^n$ i.i.d. et $X_1 \in L_2$, donc la LGFN assure, pour tout $\theta \in \Theta$,

$$\frac{1}{n} \sum_{i=1}^n X_i^2 \xrightarrow{p.s.} \mathbf{E}_\theta(X_1^2) = 4 - 3\theta.$$

La fonction $\varphi : z \mapsto \frac{4}{3} - \frac{1}{3}z$ est continue sur \mathbf{R} et donc par le CMT

$$\hat{\theta}_n^{\text{MM}} = \varphi(\overline{X^2}_n) \xrightarrow{p.s.} \varphi(4 - 3\theta) = \theta, \quad \forall \theta \in \Theta.$$

Donc $\hat{\theta}_n^{\text{MM}}$ est fortement et donc faiblement consistant.

Normalité asymptotique de $\hat{\theta}_n^{\text{MM}}$. Méthodes : TCL, Slutsky et δ -méthode. On a $\hat{\theta}_n^{\text{MM}} = \varphi\left(\frac{1}{n} \sum_{i=1}^n X_i^2\right)$ et comme $(X_i)_{i=1}^n$ sont i.i.d. et appartiennent à L_4 , $(X_i^2)_{i=1}^n$ sont i.i.d. et appartiennent à L_2 . Ainsi, par le TCL, pour tout $\theta \in \Theta$

$$\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n X_i^2 - \underbrace{\mathbf{E}_\theta(X_1^2)}_{=\mu_2(\theta)=4-3\theta} \right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \mathbf{Var}_\theta(X_1^2)).$$

C'est exactement ce dont on a besoin ici puisque $\varphi(\mu_2(\theta)) = \theta$ le paramètre qu'on cherche à estimer (c'est de là que vient justement la définition de la fonction φ). On a

$$\begin{aligned} \mathbf{Var}_\theta(X_1^2) &= \underbrace{\mathbf{E}_\theta(X_1^4)}_{=\theta\mathbf{E}_{N \sim \mathcal{N}(0,1)}(N^4)+(1-\theta)\mathbf{E}_{N \sim \mathcal{N}(0,4)}(N^4)} - \underbrace{(\mathbf{E}_\theta(X_1^2))^2}_{=(4-3\theta)^2=16-24\theta+9\theta^2} \\ &= 3\theta + (1 - \theta) \times 3 \times 16 - 16 + 24\theta - 9\theta^2 \\ &= 32 - 9\theta^2 - 21\theta, \quad \text{bien } > 0 \text{ car } \theta \in]0, 1[, \end{aligned}$$

où l'on a utilisé le fait que $N \sim \mathcal{N}(0, \sigma^2) \implies \mathbf{E}(N^4) = 3\sigma^4$. À présent, on sait que φ est \mathcal{C}^1 sur \mathbf{R} et $\varphi'(z) = \frac{-1}{3} \neq 0$. Donc par la δ -méthode ("Deuxième Théorème de Continuité"), on a

$$\sqrt{n} \left(\varphi \left(\frac{1}{n} \sum_{i=1}^n X_i^2 \right) - \varphi(\mu_2(\theta)) \right) \xrightarrow{\mathcal{L}} \mathcal{N} \left(0, \left(\frac{-1}{3} \right)^2 \times (32 - 9\theta^2 - 21\theta) \right),$$

soit, pour tout $\theta \in \Theta$,

$$\sqrt{n} \left(\hat{\theta}_n^{\text{MM}} - \theta \right) \xrightarrow{\mathcal{L}} N \left(0, \frac{32}{9} - \theta^2 - \underbrace{\frac{21}{9}\theta}_{=\frac{7}{3}\theta} \right).$$

$\hat{\theta}_n^{\text{MM}}$ est asymptotiquement normal, de variance asymptotique $\frac{32}{9} - \theta^2 - \frac{7}{3}\theta$.

Exercice 3. (facultatif) Soient $\{X_i : i \in \mathbb{N}^*\}$ une suite de variables iid à valeurs réelles et soit N une variable aléatoire à valeurs dans \mathbb{N} . On pose

$$S = \sum_{n=1}^N X_n$$

Montrer que si N est indépendant de X_i pour tout $i \in \mathbb{N}$, alors

$$\mathbf{E}[S] = \mathbf{E}[N] \mathbf{E}[X_1].$$

On suppose $N \perp\!\!\!\perp X_i$ pour tout $i \in \mathbb{N}^*$. Il est clair que $\sum_{k=0}^{+\infty} \mathbf{1}(N = k) = 1$. Par conséquent,

$$\begin{aligned} \mathbf{E}(S) &= \mathbf{E}\left(S \sum_{k=0}^{+\infty} \mathbf{1}(N = k)\right) \\ &= \sum_{k=0}^{+\infty} \mathbf{E}(S \mathbf{1}(N = k)) \quad \text{par linéarité de } \mathbf{E}(\cdot) \\ &= \sum_{k=1}^{+\infty} \mathbf{E}\left[\left(\sum_{i=1}^N X_i\right) \mathbf{1}(N = k)\right] \\ &= \sum_{k=1}^{+\infty} \mathbf{E}\left[\left(\sum_{i=1}^k X_i\right) \mathbf{1}(N = k)\right] \\ &= \sum_{k=1}^{+\infty} \sum_{i=1}^k \mathbf{E}(X_i \mathbf{1}(N = k)) \quad \text{par linéarité de } \mathbf{E}(\cdot) \\ &\stackrel{(\perp\!\!\!\perp)}{=} \sum_{k=1}^{+\infty} \sum_{i=1}^k \mathbf{E}(X_i) \mathbf{P}(N = k) \\ &= \mathbf{E}(X_1) \sum_{k=1}^{+\infty} k \mathbf{P}(N = k) \quad \text{car les } X_i \text{ sont i.i.d.} \\ &= \mathbf{E}(X_1) \mathbf{E}(X_1). \end{aligned}$$

TD5 : ESTIMATEUR DU MAXIMUM DE VRAISEMBLANCE

Exercice 1. Soient X_1, \dots, X_n des variables aléatoires i.i.d. de loi de Paréto, dont la densité est donnée par

$$f(x; \theta) = \frac{\theta}{x^{\theta+1}} \mathbb{1}(x \geq 1), \quad x \in \mathbb{R}.$$

- On suppose d'abord que l'espace de paramètres est $\Theta =]1, +\infty[$. Estimer θ par la méthode des moments.

Le modèle statistique est :

$$([1, +\infty[^n, \mathcal{B}([1, +\infty[), \{\mathbf{P}_\theta = \text{Pareto}(\theta)^{\otimes n}, \theta \in \Theta\}).$$

Cherchons l'EMM. Soit $\theta \in \Theta$, la formule de transfert donne

$$\begin{aligned} \mathbf{E}_\theta[X_1] &= \int_{\mathbb{R}} x \frac{\theta}{x^{\theta+1}} \mathbb{1}\{x \geq 1\} dx \\ &= \theta \int_1^{+\infty} x^{-\theta} dx \\ &= \theta \times \left[\frac{x^{-\theta+1}}{-\theta+1} \right]_{x=1}^{x=+\infty} \\ &= \theta \times \left[0 - \frac{1}{-\theta+1} \right] \quad \text{si } \theta > 1 \text{ non intégrable sinon.} \\ &= \frac{\theta}{\theta-1}. \end{aligned}$$

Posons $m_{n,1} := \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}_n$: contrepartie empirique du moment d'ordre 1 $\mu_1(\theta)$. L'estimateur $\hat{\theta}_n^{\text{MM}}$ est défini comme solution (si elle existe) de

$$\mu_1(\hat{\theta}_n^{\text{MM}}) = m_{n,1}, \quad \hat{\theta}_n^{\text{MM}} \in \Theta.$$

Soit,

$$\begin{aligned} \frac{\hat{\theta}_n^{\text{MM}}}{\hat{\theta}_n^{\text{MM}} - 1} = \bar{X}_n &\iff \hat{\theta}_n^{\text{MM}} = \bar{X}_n(\hat{\theta}_n^{\text{MM}} - 1) \\ &\iff 1 = \bar{X}_n(1 - \frac{1}{\hat{\theta}_n^{\text{MM}}}) \\ &\iff \bar{X}_n \frac{1}{\hat{\theta}_n^{\text{MM}}} = \bar{X}_n - 1 \\ &\iff \hat{\theta}_n^{\text{MM}} = \frac{\bar{X}_n}{\bar{X}_n - 1}. \end{aligned}$$

Il s'agit d'une unique solution qui définit donc bien l'EMM. En effet, pour tout $i \in \llbracket 1, n \rrbracket$, $X_i \geq 1$ de loi continue donc $\bar{X}_n > 1$ avec proba 1 (p.s.) et donc $\hat{\theta}_n^{\text{MM}}$ est bien défini presque sûrement.

2. On considère maintenant le cas plus général où $\Theta =]0, +\infty[$. Expliquer pourquoi la méthode des moments est inapplicable dans ce cas. Proposer un estimateur par la méthode des moments généralisée.

En conduisant le même calcul que précédemment, on déduit que μ_2 n'est pas défini pour tout $\theta \in \Theta$. Le second moment n'existe pas et l'EMM n'est donc pas utilisable. On utilise la méthode des moments généralisés. On cherche μ_r tel que

$$\mu_r(\theta) = \mathbf{E}_\theta [\varphi_r(X_1)],$$

pour une fonction φ_r connue et $r \in \{1, \dots, k\}$. Ici on prend $k = \dim(\Theta) = 1$ et il faut de plus

$$\forall \theta \in \Theta, \mathbf{E}_\theta [|\varphi(X_1)|] < +\infty.$$

Soit $p > 1$. On peut par exemple prendre

$$\varphi_p : x \mapsto \mathbb{1}\{x \geq p\}$$

qui est intégrable car prend seulement deux valeurs. On peut également prendre

$$\varphi_p : x \mapsto 1/X^p$$

qui est intégrable car $\theta > 0$. Pour le premier cas, on obtient pour tout $\theta \in \Theta$

$$\mu_{\varphi_p}(\theta) := \mathbf{E}_\theta [\varphi_p(X_1)] \tag{5}$$

$$= \mathbf{E}_\theta [\mathbb{1}\{X_1 \geq p\}] \tag{6}$$

$$= \mathbf{P}_\theta [X_1 \geq p] \tag{7}$$

$$= \int_{\mathbf{R}} \mathbb{1}\{x \geq p\} \frac{\theta}{x^{1+\theta}} \mathbb{1}\{x \geq 1\} dx \quad ; (p > 1) \tag{8}$$

$$= \int_p^{+\infty} \theta x^{-\theta-1} dx \tag{9}$$

$$= \theta \left[-\frac{x^{-\theta}}{\theta} \right]_{x=p}^{x=+\infty} \tag{10}$$

$$= \theta \times \frac{p^{-\theta}}{\theta} \tag{11}$$

$$= \frac{1}{p^\theta}. \tag{12}$$

Posons la contrepartie empirique : $m_{n,\varphi_p} := \frac{1}{n} \sum_{i=1}^n \varphi_p(X_i) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}\{X_i \geq p\}$. Pour un p fixé (par exemple $p = 2$), on résout en $\hat{\theta}$ l'équation :

$$\begin{aligned} \mu_{\varphi_p}(\hat{\theta}) &= m_{n,\varphi_p} \\ \iff \frac{1}{p^{\hat{\theta}}} &= \frac{1}{n} \sum_{i=1}^n \mathbb{1}\{X_i \geq p\} \\ \iff p^{\hat{\theta}} &= \left(\frac{1}{n} \sum_{i=1}^n \mathbb{1}\{X_i \geq p\} \right)^{-1} \\ \iff \hat{\theta} &= -\frac{\ln \left(\frac{1}{n} \sum_{i=1}^n \mathbb{1}\{X_i \geq p\} \right)}{\ln(p)}. \end{aligned}$$

La solution est unique sous réserve d'avoir $\sum_{i=1}^n \mathbb{1}\{X_i \geq p\} \neq 0$. Ainsi, pour $p > 1$, on définit $\hat{\theta}_n^{\text{GMM}}$ (ou $\hat{\theta}_n^{\text{MMG}}$)

$$\hat{\theta}_{n,p}^{\text{MMG}} := \begin{cases} -\frac{\ln(\frac{1}{n} \sum_{i=1}^n \mathbb{1}\{X_i \geq p\})}{\ln(p)} & \text{si } \sum_{i=1}^n \mathbb{1}\{X_i \geq p\} > 0, \\ \text{non défini sinon.} & \end{cases}$$

3. Déterminer l'estimateur du maximum de vraisemblance (EMV) du paramètre θ . Montrer qu'il peut être vu comme un estimateur obtenu par la méthode des moments généralisée.

Le modèle est dominé : $\forall \theta \in \Theta, \mathbf{P}_\theta << \lambda^{\otimes n}$, mesure de Lebesgue sur $([1, +\infty[^n, \mathcal{B}([1, +\infty[^n))$. On peut donc définir la vraisemblance du modèle, i.e. la densité des observations par rapport à la mesure qui domine le modèle vue comme une fonction du paramètre indexant le modèle paramétrique, θ :

$$\begin{aligned} L_n(x_1, \dots, x_n; \theta) &= \frac{d\mathbf{P}_\theta}{d\lambda^{\otimes n}}(x_1, \dots, x_n) \\ &\stackrel{iid}{=} \prod_{i=1}^n \frac{d\text{Pareto}(\theta)}{d\lambda}(x_i) \\ &= \prod_{i=1}^n f(x_i; \theta) \\ &= \theta^n \frac{1}{(\prod_{i=1}^n x_i)^{\theta+1}} \prod_{i=1}^n \mathbb{1}\{x_i \geq 1\} \\ &= \frac{\theta^n}{(\prod_{i=1}^n x_i)^{\theta+1}} \mathbb{1}\{x_{(1)} \geq 1\} \end{aligned}$$

où $x_{(1)} = \min_{1 \leq i \leq n} x_i$. Si $X_{(1)} \geq 1$, la log-vraisemblance est

$$\begin{aligned} \ell_n(\theta) &:= -\frac{1}{n} L_n(X_1, \dots, X_n; \theta) \\ &= -\frac{1}{n} \log \left[\frac{\theta^n}{(\prod_{i=1}^n X_i)^{\theta+1}} \right] \\ &= \frac{(\theta+1)}{n} \log \left(\prod_{i=1}^n X_i \right) - \log(\theta) \\ &= \frac{(\theta+1)}{n} \sum_{i=1}^n \log(X_i) - \log(\theta) \end{aligned}$$

et $\ell_n(\theta) = +\infty$ si $X_{(1)} < 1$. Par définition, on a

$$\hat{\theta}_n^{\text{EMV}} := \arg \min_{\theta \in \Theta} \ell_n(\theta).$$

La fonction $\theta \mapsto \log(\theta)$ est concave en θ donc $\theta \mapsto -\log(\theta)$ est convexe en θ . La fonction $\theta \mapsto \frac{(\theta+1)}{n} \sum_{i=1}^n \log(X_i)$ est affine donc convexe en θ . Ainsi, $\theta \mapsto \ell_n(\theta)$ est convexe en θ et donc

tout point stationnaire (extremum local) est un minimum global pour $\ell_n(\cdot)$. De plus, $\ell_n(\cdot)$ est différentiable et on a

$$\ell'_n(\theta) = \frac{\partial \ell_n}{\partial \theta}(\theta) = \frac{1}{n} \sum_{i=1}^n \log(X_i) - \frac{1}{\theta}, \quad (13)$$

$$\ell''_n(\theta) = \frac{\partial \ell'_n}{\partial \theta}(\theta) = \frac{1}{\theta^2} > 0. \quad (14)$$

On déduit de (14) que ℓ_n est strictement convexe. Ainsi, la condition de premier ordre (CPO) est nécessaire et suffisante pour avoir un minimum global :

$$\ell'_n(\theta) = 0 \iff \theta = \left(\frac{1}{n} \sum_{i=1}^n \log(X_i) \right)^{-1}.$$

On en déduit, presque sûrement

$$\hat{\theta}_n^{\text{EMV}} = \left(\frac{1}{n} \sum_{i=1}^n \log(X_i) \right)^{-1}.$$

Pour montrer que l'EMV peut-être obtenu par les méthode des moments généralisés, on peut remarquer que $\hat{\theta}_n^{\text{EMV}}$ est une fonction continue (l'inverse) de la moyenne empirique des $\log(X_i)$. Posons

$$\begin{aligned} \varphi : [1, +\infty[&\rightarrow \mathbf{R} \\ x &\mapsto \log(x) \end{aligned}$$

et

$$\begin{aligned} \mu(\theta) &= \mathbf{E}_\theta [\varphi(X_1)] \\ &= \int_{\mathbf{R}} \log(x) f(x; \theta) dx \\ &= \theta \int_1^{+\infty} \frac{\log(x)}{x^{\theta+1}} dx \\ &= \theta \int_0^{+\infty} \frac{y}{e^{y(\theta+1)}} e^y dy \quad x = e^y \\ &= \int_0^{+\infty} \theta y e^{-\theta y} dy \\ &= \underbrace{\frac{1}{\theta} \int_0^{+\infty} z e^{-z} dz}_{\Gamma(2)=1} \quad z = \theta y = \frac{1}{\theta}. \end{aligned}$$

Note : on pouvait aussi procéder par I.P.P. Soit $m := \frac{1}{n} \sum_{i=1}^n \log(X_i)$. On résout en $\hat{\theta}$: $\mu(\hat{\theta}) = m \iff \hat{\theta} = (\frac{1}{n} \sum_{i=1}^n \log(X_i))^{-1}$, une unique solution. Donc

$$\hat{\theta}_n^{\text{EMV}} = \hat{\theta}_n^{\text{MMG}} \text{ avec } \varphi = \log = \left(\frac{1}{n} \sum_{i=1}^n \log(X_i) \right)^{-1}.$$

4. Déterminer la loi limite et la variance limite de l'EMV.

$\hat{\theta}_n^{\text{EMV}}$ est consistant par la LGN et le CMT. Posons $Y_i = \log(X_i)$. On a $\{Y_i\}_{i=1}^n$ i.i.d. et intégrables donc

$$\frac{1}{n} \sum_{i=1}^n Y_i \xrightarrow{P} \mathbf{E}_\theta[\log(X_1)] = \frac{1}{\theta}$$

par la LGN et question 3. La fonction $z \mapsto \frac{1}{z}$ est \mathcal{C}^0 sur \mathbf{R}_+^* donc par le CMT

$$\hat{\theta}_n^{\text{EMV}} \xrightarrow{P} \theta, \quad \forall \theta \in \Theta.$$

Ainsi, on en déduit

$$\forall \theta \in \Theta, \quad \hat{\theta}_n^{\text{EMV}} \xrightarrow{\mathcal{L}} \theta, \quad \text{i.e. } \delta_\theta \text{ Dirac en } \theta.$$

Il s'agit plutôt de montrer la normalité asymptotique ici. En vue d'appliquer le TCL puis la δ -méthode, vérifions dans un premier temps que $Y_1 \in L_2$. En effet, pour tout $\theta \in \Theta$,

$$\begin{aligned} \mathbf{E}_\theta[Y_1^2] &= \theta \int_1^{+\infty} \frac{(\log(x))^2}{x^{\theta+1}} dx \\ &= \theta \int_0^{+\infty} \frac{y^2}{e^{\theta y+y}} e^y dy \quad x = e^y \\ &= \frac{1}{\theta^2} \int_0^{+\infty} z^2 e^{-z} dz \quad z = \theta y \\ &= \frac{1}{\theta^2} \Gamma(3) = \frac{2}{\theta^2} < +\infty. \end{aligned}$$

On en déduit

$$\begin{aligned} \mathbf{Var}_\theta[Y_1] &= \mathbf{E}_\theta[Y_1^2] - \mathbf{E}_\theta[Y_1]^2 \\ &= \frac{2}{\theta^2} - \left(\frac{1}{\theta}\right)^2 = \frac{1}{\theta^2}. \end{aligned}$$

Le TCL assure

$$\sqrt{n} \left(\underbrace{\frac{1}{n} \sum_{i=1}^n Y_i}_{= \frac{1}{\hat{\theta}_n^{\text{EMV}}}} - \underbrace{\mathbf{E}_\theta[Y_1]}_{= \frac{1}{\theta}} \right) \xrightarrow{\mathcal{L}} \mathcal{N} \left(0, \underbrace{\mathbf{Var}_\theta[Y_1]}_{= \frac{1}{\theta^2}} \right).$$

On applique la δ -méthode avec $g : x \mapsto 1/x$ pour tout $x \in \mathbf{R}_+^*$ qui est \mathcal{C}^0 et telle que $g'(x) = -\frac{1}{x^2}$. On a

$$g' \left(\frac{1}{\theta} \right) = -\frac{1}{(1/\theta)^2} = -\theta^2 \neq 0$$

pour tout $\theta \in \Theta$. Ainsi, la δ -méthode donne, pour tout $\theta \in \Theta$,

$$\sqrt{n} \left(g \left(\frac{1}{\hat{\theta}_n^{\text{EMV}}} \right) - g \left(\frac{1}{\theta} \right) \right) \xrightarrow{\mathcal{L}} \mathcal{N} \left(0, \frac{1}{\theta^2} \times \underbrace{g' \left(\frac{1}{\theta} \right)^2}_{=(-\theta^2)^2 = \theta^4} \right),$$

c'est-à-dire

$$\sqrt{n} (\hat{\theta}_n^{\text{EMV}} - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \theta^2) [\mathbf{P}_\theta] \forall \theta \in \Theta.$$

On conclut que $\hat{\theta}_n^{\text{EMV}}$ est asymptotiquement normal et a pour variance asymptotique θ^2 (dont on montre ci-dessous qu'il s'agit de l'inverse de l'information de Fisher, $I_1(\theta)^{-1}$).

Autre méthode utilisant Slutsky : On a

$$\sqrt{n} (\hat{\theta}_n^{\text{EMV}} - \theta) = \underbrace{\hat{\theta}_n^{\text{EMV}} \times \theta \times (-1)}_{\xrightarrow{\text{p.s.}} -\theta^2} \times \underbrace{\sqrt{n} \left(\frac{1}{\hat{\theta}_n^{\text{EMV}}} - \frac{1}{\theta} \right)}_{\xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{1}{\theta^2}\right)}$$

Le théorème de Slutsky implique que

$$\sqrt{n} (\hat{\theta}_n^{\text{EMV}} - \theta) \xrightarrow{\mathcal{L}} (-\theta^2) \times \mathcal{N}\left(0, \frac{1}{\theta^2}\right) = \mathcal{N}(0, \theta^2).$$

5. Calculer l'information de Fisher du modèle de Paréto et la comparer avec l'inverse de la variance asymptotique.

D'après le cours, l'information de Fisher est la variance du score (lui-même défini comme la dérivée de la log-vraisemblance). On distingue l'information de Fisher pour une observation, $I_1(\theta)$ pour un n -échantillon (X_1, \dots, X_n) , $I_n(\theta)$. On a

$$I_1(\theta) := \mathbf{E}_\theta \left[\left(\frac{\partial}{\partial \theta} \log f(X_1; \theta) \right)^2 \right] \quad (15)$$

$$= -\mathbf{E}_\theta \left[\frac{\partial^2}{\partial \theta^2} \log f(X_1; \theta) \right]. \quad (16)$$

Si l'échantillon est i.i.d., on a $I_n(\theta) = n \times I_1(\theta)$. Ici, on a

$$\frac{\partial}{\partial \theta} \log f(x; \theta) = \frac{\partial}{\partial \theta} (\log \theta - (\theta + 1) \log x) = \frac{1}{\theta} - \log x.$$

D'où

$$\begin{aligned} I_1(\theta) &= \mathbf{E}_\theta \left[\left(\log X_1 - \frac{1}{\theta} \right)^2 \right] \\ &= \mathbf{Var}_\theta(Y_1) \\ &= \frac{1}{\theta^2}. \end{aligned}$$

On remarque que

$$\sqrt{n} (\hat{\theta}_n^{\text{EMV}} - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, I_1(\theta)^{-1}\right) [\mathbf{P}_\theta] \forall \theta \in \Theta.$$

Exercice 2. Une chaîne de montage produit des objets, dont on veut estimer la durée moyenne de fabrication. On suppose que les durées de fabrication T_i sont indépendantes et de loi exponentielle de paramètre θ . Le $n^{\text{ième}}$ objet est donc fabriqué à la date $T_1 + \dots + T_n$. On observe uniquement le nombre d'objets N_t fabriqués à la date t .

$(T_i)_{i \geq 1}$ durées de fabrication i.i.d. avec $T_1 \sim \mathcal{E}(\theta)$, i.e.

$$\forall x \in \mathbf{R}, \frac{dP^{T_1}}{d\lambda}(x) = \theta e^{-\theta x} \mathbb{1}\{x \geq 0\},$$

d'espérance $\mathbf{E}_\theta[T_1] = 1/\theta$. Observations : N_t = nombre d'objets fabriqués à la date t (fixé).

- Montrer que $\mathbf{P}(N_t \leq n) = \mathbf{P}(T_1 + \dots + T_{n+1} > t)$.

Soit $t > 0$ et $n \in \mathbb{N}$,

$$\begin{aligned} \{N_t \leq n\} &\iff \text{au plus } n \text{ objets ont été fabriqués à la date } t \\ &\iff \text{le } n+1 \text{ ème objet a été fabriqué à une date ultérieure à } t \\ &\iff T_1 + \dots + T_{n+1} > t, \end{aligned}$$

où $T_1 + \dots + T_{n+1}$ est la date de fabrication du $n+1$ ème objet. On a donc égalité des deux événements.

- Déterminer la loi de la variable aléatoire $S_n = T_1 + \dots + T_n$. Rappelons ici que la fonction gamma est définie par

$$\Gamma(a) = \int_0^\infty x^{a-1} e^{-x} dx, \quad \forall a > 0,$$

et elle vérifie $\Gamma(a+1) = a\Gamma(a)$.

Rappel loi Gamma : $X \sim \Gamma(p, \theta)$, $p > 0$, $\theta > 0$ si et seulement si

$$\frac{dP^X}{d\lambda}(x) = \frac{\theta^p}{\Gamma(p)} x^{p-1} e^{-\theta x} \mathbb{1}\{x \in \mathbf{R}_+\}, \forall x \in \mathbf{R}.$$

On a de plus $\mathbf{E}(X) = \frac{p}{\theta}$, $\mathbf{Var}(X) = \frac{p}{\theta^2}$ et d'autres propriétés de convolutions :

- Soit $a \in \mathbf{R}_+^*$, $X \sim \Gamma(p, \theta) \implies aX \sim \Gamma(p, \frac{\theta}{a})$,
- $X \sim \Gamma(p, \theta)$, $Y \sim \Gamma(q, \theta)$, $X \perp\!\!\!\perp Y \implies X * Y \sim \Gamma(p+q, \theta)$ avec $*$ = somme de v.a. indépendantes (convolution).

Rappel fonction gamma Γ : $\gamma(n) = (n+1)!$ pour $n \in \mathbb{N}$.

Rappel : la famille exponentielle est une sous-famille des lois Gamma ($\mathcal{E}(\theta) = \Gamma(1, \theta)$). On a en particulier $\mathcal{E}(\theta) * \mathcal{E}(\theta) \stackrel{d}{=} \Gamma(2, \theta)$. Montrons par récurrence $\underbrace{\mathcal{E}(\theta) * \dots * \mathcal{E}(\theta)}_{n \text{ fois}} \stackrel{d}{=} \Gamma(n, \theta)$, i.e.

$$S_n \sim \Gamma(n, \theta).$$

Pour $n = 1$, $S_1 = T_1 \sim \mathcal{E}(\theta) = \Gamma(1, \theta)$. On peut ensuite utiliser la caractérisation par l'espérance de fonctions mesurables bornées ou bien passer par les fonctions caractéristiques. Utilisons la

première méthode. Pour tout $n \in \mathbb{N}$, on note $f_n(x) = \frac{dP_{S_n}}{d\lambda}$ la densité de S_n . On a

$$f_1(x) = \theta e^{-\theta x} \mathbf{1}\{x \in \mathbf{R}_+\},$$

$S_n \perp\!\!\!\perp T_{n+1}$ et $S_{n+1} = S_n + T_{n+1}$. Soit $h : \mathbf{R} \rightarrow \mathbf{R}$ mesurable bornée. On a

$$\begin{aligned} \mathbf{E}_\theta[h(S_{n+1})] &= \mathbf{E}_\theta[h(S_n + T_{n+1})] \\ &= \int_0^{+\infty} \int_0^{+\infty} h(x+y) f_n(x) \theta e^{-\theta y} dx dy \quad \text{produit des densités par indépendance.} \end{aligned}$$

En posant $z = x + y \iff x = z - y$, on obtient

$$\begin{aligned} \int_0^{+\infty} h(x+y) f_n(x) dx &= \int_y^{+\infty} h(z) f_n(z-y) dz \\ &= \int_0^{+\infty} h(z) f_n(z-y) \mathbf{1}\{y \leq z\} dz. \end{aligned}$$

Une application du théorème de Fubini donne

$$\begin{aligned} \mathbf{E}_\theta[h(S_{n+1})] &= \int_0^{+\infty} \int_0^{+\infty} h(z) f_n(z-y) \mathbf{1}\{y \leq z\} \theta e^{-\theta y} dy dz \\ &= \int_0^{+\infty} h(z) \left\{ \int_0^{+\infty} f_n(z-y) \theta e^{-\theta y} \mathbf{1}\{y \leq z\} dy \right\} dz \\ &= \int_0^{+\infty} h(z) \underbrace{\left\{ \int_0^z f_n(z-y) \theta e^{-\theta y} dy \right\}}_{=f_{n+1}(z) \text{ car } h \text{ mesurabe bornée}} dz. \end{aligned}$$

On en déduit, pour tout $z \geq 0$,

$$\begin{aligned} f_{n+1}(z) &= \int_0^z f_n(z-y) \theta e^{-\theta y} dy \\ &= \int_0^z \frac{\theta^n}{\Gamma(n)} t^{n-1} e^{-\theta t} \mathbf{1}\{t \in \mathbf{R}_+\} \theta e^{-\theta z} e^{\theta t} dt \quad z-y=t \iff y=z-t \\ &= \frac{\theta^{n+1}}{\Gamma(n)} e^{-\theta z} \underbrace{\int_0^z t^{n-1} dt}_{=\frac{z^n}{n}} \\ &= \frac{\theta^{n+1}}{\Gamma(n) n} z^n e^{-\theta z}, \end{aligned}$$

qui est bien la densité d'une $\Gamma(n+1, \theta)$. Par récurrence, on a donc bien montré que, pour tout $n \in \mathbb{N}$, $x \geq 0$

$$f_n(x) = \frac{\theta^n x^{n-1}}{(n-1)!} e^{-x\theta},$$

soit $S_n \sim \Gamma(n, \theta)$.

3. Montrer que N_t suit une loi de Poisson dont on déterminera le paramètre.

Le support de N_t est \mathbb{N} . Soit $n \in \mathbb{N}$, $\theta \in \Theta := \mathbf{R}_+^*$. On a

$$\{N_t = n\} \iff \{N_t \leq n \text{ et } N_t > n - 1\}.$$

D'où,

$$\begin{aligned} \mathbf{P}_\theta(N_t = n) &= \mathbf{P}_\theta(N_t \leq n) - \mathbf{P}_\theta(N_t \leq n - 1) \\ &= \mathbf{P}_\theta(S_{n+1} > t) - \mathbf{P}_\theta(S_n > t) \quad \text{d'après Q1} \\ &= \mathbf{P}_\theta(S_n \leq t) - \mathbf{P}_\theta(S_{n+1} \leq t) \quad \text{par passage au complémentaire} \\ &= \frac{\theta^n}{\Gamma(n)} \int_0^t x^{n-1} e^{-\theta x} dx - \frac{\theta^{n+1}}{\Gamma(n+1)} \int_0^t x^n e^{-\theta x} dx \\ &= \frac{\theta^n}{n!} \left[n \int_0^t x^{n-1} e^{-\theta x} dx - \underbrace{\theta \int_0^t x^n e^{-\theta x} dx}_{= \int_0^{+\infty} x^n (-\theta) e^{-\theta x} dx} \right] \\ &= \frac{\theta^n}{n!} \left\{ n \int_0^t x^{n-1} e^{-\theta x} dx + [x^n e^{-\theta x}]_0^t - n \int_0^t x^{n-1} e^{-\theta x} dx \right\} \quad \text{par une I.P.P. de la seconde intégrale} \\ &= \frac{\theta^n}{n!} t^n e^{-\theta t} \\ &= \frac{(\theta t)^n}{n!} e^{-\theta t}. \end{aligned}$$

On a donc $N_t \sim \text{Poisson}(\theta t)$.

4. Quel est l'EMV de θ ?

On a une seule observation N_t à une date fixe t connue. Le modèle statistique est

$$(\mathbb{N}, \mathcal{P}(\mathbb{N}), \{\mathbf{P}_\theta = \text{Poisson}(\theta t), \theta \in \Theta = \mathbf{R}_+^*\}).$$

Il s'agit d'un modèle discret, dominé par la mesure de comptage. La vraisemblance du modèle est donnée par

$$\begin{aligned} L_t(N_t; \theta) &:= \frac{d\mathbf{P}_\theta}{d\mu}(N_t) \\ &= \mathbf{P}_\theta(N_t = N_t) \\ &= \mathbf{P}_{P \sim \text{Poisson}(\theta t)}(P = N_t) \\ &= \frac{(\theta t)^{N_t}}{N_t!} e^{-\theta t}. \end{aligned}$$

La log-vraisemblance est alors

$$\ell_t(\theta) := -\log L(N_t; \theta) \quad (17)$$

$$= \log(N_t!) + \theta t - N_t \log(\theta t) \quad (18)$$

$$= \theta t - N_t \log(\theta) + \underbrace{\log(N_t!) - N_t \log(t)}_{\text{constante indép. de } \theta}. \quad (19)$$

La fonction $\theta \mapsto \theta t$ est linéaire en θ donc convexe en θ . Comme $N_t \geq 0$, la fonction $\theta \mapsto -N_t \log(\theta)$ est convexe en θ donc ℓ_t est convexe en θ . En fait, on peut montrer que ℓ_t est strictement convexe en θ (calculer la dérivée seconde) donc tout extremum local est un minimum global et ce minimum global est unique, i.e. les CPO sont nécessaires et suffisantes pour déterminer uniquement l'EMV. On a

$$\ell'_t(\theta) = 0 \iff \theta = \frac{N_t}{t}.$$

On a donc

$$\hat{\theta}_t^{\text{EMV}} = \frac{N_t}{t}.$$

Rq : il s'agit de l'EMM (avec moment d'ordre 1) car $\mathbf{E}_\theta(N_t) = \theta t$ et $\hat{\theta}t = N_t \iff \hat{\theta} = N_t/t$. On a une seule observation ici donc $\bar{N}_t = \frac{1}{1} \sum_1 N_t = N_t$.

5. Calculer le risque de $\hat{\theta}_t^{\text{MV}}$ et étudier son comportement lorsque $t \rightarrow +\infty$.

On a

$$\begin{aligned} R_\theta(\hat{\theta}_t^{\text{EMV}}) &= \mathbf{Var}_\theta(\hat{\theta}_t^{\text{EMV}}) + \left[\underbrace{\mathbf{E}_\theta(\hat{\theta}_t^{\text{EMV}})}_{=\frac{1}{t}\mathbf{E}_\theta(N_t)=\frac{\theta t}{t}=\theta} - \theta \right]^2 \\ &= \mathbf{Var}_\theta\left(\frac{N_t}{t}\right) \\ &= \frac{1}{t^2} \underbrace{\mathbf{Var}_\theta(N_t)}_{=\theta t} \quad \text{car } N_t \stackrel{\mathbf{P}_\theta}{\sim} \text{Poisson}(\theta t) \\ &= \frac{\theta}{t}. \end{aligned}$$

Ainsi, on a

$$R_\theta(\hat{\theta}_t^{\text{EMV}}) = \frac{\theta}{t} \xrightarrow[t \rightarrow +\infty]{} 0.$$

On retrouve la vitesse habituelle en t pour le risque quadratique (équivalent $\sqrt{t} \approx \sqrt{n}$ pour racine du risque = distance L_2).

TD6 : ESTIMATION DANS UN PROBLÈME DE SONDAGE

Exercice 1. (Problème de sondages) Soit N le nombre d'habitants d'une commune. Il s'agit de faire un sondage de popularité de deux candidats (candidat A et candidat B) qui se présentent aux élections municipales. On choisit un échantillon de n habitants auxquels on pose la question : "Pour qui voteriez-vous aux élections ?" A l'issue de ce sondage, on obtient les données X_1, \dots, X_n , où

$$X_i = \begin{cases} 1, & \text{si le } i^{\text{ème}} \text{ habitant questionné préfère le candidat } A, \\ 0, & \text{si le } i^{\text{ème}} \text{ habitant questionné préfère le candidat } B, \end{cases} \quad i = 1, \dots, n.$$

Pour les raisons évidentes, il est impossible de questionner tous les habitants. Donc $n < N$ (dans la pratique, on a toujours $n \ll N$). Notons θ la part d'habitants de la commune qui préfèrent le candidat A . Le but du sondage est d'estimer θ et de donner un intervalle confiance pour θ .

Définissons les valeurs déterministes x_1, \dots, x_N par

$$x_j = \begin{cases} 1, & \text{si le } j^{\text{ème}} \text{ habitant préfère le candidat } A, \\ 0, & \text{si le } j^{\text{ème}} \text{ habitant préfère le candidat } B, \end{cases} \quad j = 1, \dots, N.$$

On a alors

$$\theta = \frac{1}{N} \sum_{j=1}^N x_j.$$

Définissons aussi

$$\sigma^2 = \frac{1}{N} \sum_{j=1}^N (x_j - \theta)^2.$$

On appelle θ *moyenne de population* et σ^2 *variance de population*.

1. Proposer un modèle statistique pour ce problème. Pour cela,

- (a) Calculer $\mathbf{P}_\theta((X_1, X_2, X_3, X_4) = (0, 1, 1, 0))$ en utilisant le fait qu'il s'agit d'un tirage au hasard sans remise d'une population de taille N , car chaque habitant peut apparaître au maximum une fois dans l'échantillon.

Le vecteur aléatoire observé (X_1, \dots, X_n) est à valeurs dans $\{0, 1\}^n$. Pour définir le modèle, il nous faut trouver sa loi. On commence par un exemple. Soit $n = 4$ et $(a_1, \dots, a_4) = (0, 1, 1, 0)$. Les tirages sont indépendants mais non identiquement distribués (la composition de l'urne change). On a

$$\begin{aligned} \mathbf{P}_\theta((X_1, \dots, X_4) = (a_1, \dots, a_4)) &= \mathbf{P}_\theta(X_1 = 0, X_2 = 1, X_3 = 1, X_4 = 0) \\ &= \frac{N_B}{N} \times \frac{N_A}{N-1} \times \frac{N_A-1}{N-2} \times \frac{N_B-1}{N-3} \\ &= \mathbf{P}_\theta((X_1, \dots, X_4) = (1, 1, 0, 0)), \end{aligned}$$

où $N_A = N\theta$ = nombre d'habitants qui préfèrent A et $N_B = N(1 - \theta)$ = nombre d'habitants qui préfèrent B . On remarque que seul le nombre de 1 (ou de 0) joue dans la distribution de la probabilité.

(b) Soit $(a_1, \dots, a_n) \in \{0, 1\}^n$ tel que $a_1 + \dots + a_n = a$. Deviner la forme de

$$\mathbf{P}_\theta((X_1, \dots, X_n) = (a_1, \dots, a_n))$$

et prouver la formule obtenue par récurrence sur n .

a est le nombre de personnes dans l'échantillon qui préfèrent A . On aura alors

$$\begin{aligned} & \mathbf{P}_\theta((X_1, \dots, X_n) = (a_1, \dots, a_n)) \\ &= \frac{\text{nombre de cas possibles pour avoir } a_1 + \dots + a_n = a}{\text{nombre de tirages possibles}} \\ &= \frac{\mathcal{A}_{N_A}^a \times \mathcal{A}_{N_B}^{n-a}}{\mathcal{A}_N^n} \quad (*) \\ &= \frac{N_A \times (N_A - 1) \times \dots \times (N_A - a + 1) \times N_B \times (N_B - 1) \times \dots \times (N_B - (n - a) + 1)}{N \times (N - 1) \times \dots \times (N - n + 1)}, \end{aligned}$$

où $\mathcal{A}_{|E|}^x = \frac{|E|!}{(|E|-x)!} = |E| \times (|E| - 1) \times \dots \times (|E| - x + 1)$ désigne le nombre d'arrangements de x éléments (x -liste) de E , i.e. le nombre de manière de choisir x éléments distincts parmi $|E|$ en tenant compte de l'ordre.

Récurrence sur n :

- $n = 1$. Si $a = 0$, on observe $a_1 = 0$ avec proba N_B/N . Si $a = 1$, on observe $a_1 = 1$ avec proba N_A/N . Cela correspond au tirage au hasard d'un individu du groupe A (ou B) qui est bien la formule avec la convention $\mathcal{A}_n^0 = 1$ (une seule façon de choisir 0 élément).
- Supposons $(*)$ vraie pour $n \in \mathbb{N}$. Soit $(a_1, \dots, a_n) \in \{0, 1\}^n$. On considère le cas $a_{n+1} = 1$ (le cas $a_{n+1} = 0$ se traite de façon similaire). En posant $a = a_1 + \dots + a_n + a_{n+1} = a_1 + \dots + a_n + 1$, on a par la formule de Bayes (ou celle des probabilités totales)

$$\begin{aligned} & \mathbf{P}_\theta((X_1, \dots, X_{n+1}) = (a_1, \dots, a_{n+1})) \\ &= \mathbf{P}_\theta(X_{n+1} = a_{n+1} | (X_1, \dots, X_n) = (a_1, \dots, a_n)) \times \mathbf{P}_\theta((X_1, \dots, X_n) = (a_1, \dots, a_n)) \\ &= \frac{\# \text{d'habitants préférant A après les } n \text{ premiers tirages}}{\# \text{habitants restants après } n \text{ tirages}} \times \mathbf{P}_\theta((X_1, \dots, X_n) = (a_1, \dots, a_n)) \\ &= \frac{N_A - (a_1 + \dots + a_n)}{N - n} \times \mathbf{P}_\theta((X_1, \dots, X_n) = (a_1, \dots, a_n)) \\ &= \frac{N_A - a + 1}{N - n} \times \frac{\mathcal{A}_{N_A}^{a-1} \mathcal{A}_{N_B}^{n-a+1}}{\mathcal{A}_N^n} \quad \text{par hypothèse de récurrence} \\ &= \frac{\mathcal{A}_{N_A}^a \mathcal{A}_{N_B}^{n+1-a}}{\mathcal{A}_N^{n+1}}. \end{aligned}$$

On retrouve bien $(*)$ au rang $n + 1$ avec $a_{n+1} = 1$. Le 3ème élément du modèle statistique est donc

$$\{\mathbf{P}_\theta \text{ où } \mathbf{P}_\theta \text{ est la distribution des observations caractérisée par } (*), \theta \in]0, 1[\}\}.$$

2. Montrer que $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ est un estimateur sans biais de θ .

3. Montrer que

$$\mathbf{Cov}(X_i, X_j) = -\frac{\sigma^2}{N-1} \quad \text{pour } i \neq j \quad \text{et} \quad \mathbf{Var}(\bar{X}_n) = \frac{\sigma^2}{n} \left(1 - \frac{n-1}{N-1}\right).$$

On peut remarquer que la loi du vecteur (X_1, \dots, X_n) est invariante par permutation. En particulier, on a

$$(X_1, X_2, \dots, X_{n-1}, X_n) \stackrel{d}{=} (X_n, X_2, \dots, X_{n-1}, X_1).$$

Ainsi X_n a exactement la même loi que X_1 . De même, le couple (X_i, X_j) avec $i < j$ a la même loi que le couple (X_1, X_2) . On obtient alors

$$\begin{aligned} \mathbf{E}_\theta(X_n) &= \mathbf{E}_\theta(X_1) = \mathbf{P}_\theta(X_1 = 1) = \frac{N_A}{N} = \frac{N\theta}{N} = \theta, \quad \forall n \in \mathbb{N}, \\ \mathbf{E}_\theta(X_i X_j) &= \mathbf{E}_\theta(X_1 X_2) = \mathbf{P}_\theta(X_1 = 1; X_2 = 1) = \frac{N_A \times (N_A - 1)}{N \times (N - 1)} = \theta \times \frac{N\theta - 1}{N - 1}. \end{aligned}$$

Ainsi,

$$\mathbf{E}_\theta(\bar{X}_n) = \frac{1}{n} \sum_{i=1}^n \theta = \theta,$$

et, en remarquant que $\sigma^2 = \theta(1 - \theta)$ (puisque $x_j^2 = x_j, \forall j$),

$$\begin{aligned} \mathbf{Cov}_\theta(X_i, X_j) &= \mathbf{E}_\theta(X_i X_j) - \mathbf{E}_\theta(X_i) \mathbf{E}_\theta(X_j) \\ &= \theta \left(\frac{N\theta - 1}{N - 1} - \theta \right) \\ &= \theta \times \frac{\theta - 1}{N - 1} \\ &= -\frac{\sigma^2}{N - 1}. \end{aligned}$$

Finalement, on obtient

$$\begin{aligned} \mathbf{Var}_\theta(\bar{X}_n) &= \frac{1}{n^2} \sum_{i=1}^n \mathbf{Var}_\theta(X_i) + \frac{2}{n^2} \sum_{i < j} \mathbf{Cov}_\theta(X_i, X_j) \quad (\text{voir TD2}) \\ &= \frac{1}{n} \mathbf{Var}_\theta(X_1) - \frac{2}{n^2} \times \frac{n(n-1)}{2} \times \frac{\sigma^2}{N-1} \\ &= \frac{\sigma^2}{n} - \frac{2}{n^2} \times \frac{n(n-1)}{2} \times \frac{\sigma^2}{N-1} \\ &= \frac{\sigma^2}{n} \left(1 - \frac{n-1}{N-1}\right). \end{aligned}$$

4. Calculer $\mathbf{E}[s_n^2]$, où $s_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2$, et proposer un estimateur sans biais $\hat{\sigma}_n^2$ de la variance σ^2 .

$$\begin{aligned}
s_n^2 &= \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \\
&= \frac{1}{n} \sum_{i=1}^n X_i^2 - (\bar{X}_n)^2 \\
&= \frac{1}{n} \sum_{i=1}^n X_i - (\bar{X}_n)^2 \quad \text{car } X_i \in \{0, 1\} \\
&= \bar{X}_n(1 - \bar{X}_n).
\end{aligned}$$

Donc

$$\begin{aligned}
\mathbf{E}_\theta[s_n^2] &= \mathbf{E}_\theta[\bar{X}_n] - \mathbf{E}_\theta[(\bar{X}_n)^2] \\
&= \theta - \mathbf{Var}_\theta[\bar{X}_n] - (\mathbf{E}_\theta[\bar{X}_n])^2 \\
&= \theta - \frac{\sigma^2}{n} \left(1 - \frac{n-1}{N-1}\right) - \theta^2 \\
&= \sigma^2 \left(1 - \frac{1}{n} \left(1 - \frac{n-1}{N-1}\right)\right) \\
&= \sigma^2 \times \frac{n \times (N-1) - (N-1) + n-1}{n \times (N-1)} \\
&= \sigma^2 \times \frac{(n-1) \times N}{n \times (N-1)}.
\end{aligned}$$

On en déduit que

$$\hat{V}_n^2 = \frac{n(N-1)}{N(n-1)} s_n^2$$

est un estimateur sans biais de σ^2 .

5. On se place maintenant dans le cadre asymptotique où $N \rightarrow \infty$, $n = n(N) \rightarrow \infty$ et $n/N \rightarrow 0$.

(a) Montrer que \bar{X}_n et \hat{v}_n^2 sont des estimateurs consistants de θ et σ^2 .

On a vu que \bar{X}_n est sans biais et que

$$\mathbf{Var}_\theta(\bar{X}_n) = \frac{\sigma^2}{n} \left(1 - \frac{n-1}{N-1}\right) \xrightarrow[N \rightarrow \infty]{} 0.$$

Donc $\bar{X}_n \xrightarrow{P} \theta$, ce qui veut dire que \bar{X}_n est consistant. De plus, comme

$$\hat{v}_n^2 = \frac{n(N-1)}{N(n-1)} \times \bar{X}_n(1 - \bar{X}_n) \xrightarrow{P} \theta(1 - \theta) = \sigma^2$$

on en déduit que \hat{v}_n^2 est un estimateur consistant de σ^2 .

(b) On admet que $\sqrt{n}(\bar{X} - \theta)$ converge en loi vers $\mathcal{N}(0, \sigma^2)$. Démontrer la normalité asymptotique de \hat{v}_n^2 .

tique

$$\sqrt{n} \frac{\bar{X}_n - \theta}{\hat{\sigma}_n} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1).$$

On cherche une statistique pivotale asymptotique, i.e. une statistique T_n qui dépend de θ (et bien-sûr des données) telle que, pour tout $\theta \in \Theta$, la distribution limite de T_n est indépendante de θ . Pour cela, on admet que

$$\sqrt{n}(\bar{X}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2).$$

Le Théorème de Slutsky, le CMT et la question 5.a) impliquent que

$$T_n := \frac{\sqrt{n}(\bar{X}_n - \theta)}{\hat{\sigma}_n} = \underbrace{\frac{\sqrt{n}(\bar{X}_n - \theta)}{\sigma}}_{\xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)} \times \underbrace{\frac{\sigma}{\hat{\sigma}_n}}_{\xrightarrow{P} 1} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1), \forall \theta \in \Theta.$$

(c) En utilisant le résultat ci-dessus, trouver un l'intervalle aléatoire $[A, B]$ tel que

$$\lim_{n \rightarrow \infty} \mathbf{P}([A, B] \text{ contient } \theta) = 95\%.$$

D'après la question 5.b), on a

$$\lim_{n \rightarrow \infty} \mathbf{P}_\theta(T_n \in [a, b]) \xrightarrow{n \rightarrow \infty} \mathbf{P}(Z \in [a, b]),$$

où $Z \sim \mathcal{N}(0, 1)$. Si l'on choisit $[a, b] = [-1.96, 1.96]$ on a alors

$$\mathbf{P}(Z \in [a, b]) = 0.95.$$

D'autre part,

$$\begin{aligned} T_n \in [a, b] &\iff a\hat{\sigma}_n \leq \sqrt{n}(\bar{X}_n - \theta) \leq b\hat{\sigma}_n \\ &\iff -\frac{b\hat{\sigma}_n}{\sqrt{n}} \leq \theta - \bar{X}_n \leq -\frac{a\hat{\sigma}_n}{\sqrt{n}} \\ &\iff \bar{X}_n - \frac{b\hat{\sigma}_n}{\sqrt{n}} \leq \theta \leq \bar{X}_n - \frac{a\hat{\sigma}_n}{\sqrt{n}}. \end{aligned}$$

Pour un tel couple de réels (a, b) , on a donc bien

$$\mathbf{P}_\theta \left(\left[\bar{X}_n - \frac{b\hat{\sigma}_n}{\sqrt{n}}, \bar{X}_n - \frac{a\hat{\sigma}_n}{\sqrt{n}} \right] \ni \theta \right) \xrightarrow{n \rightarrow \infty} 0.95.$$

L'intervalle aléatoire d'intérêt ("intervalle de confiance pour θ de niveau asymptotique 95%") est donc

$$\left[\bar{X}_n - \frac{1.96\hat{\sigma}_n}{\sqrt{n}}, \bar{X}_n + \frac{1.96\hat{\sigma}_n}{\sqrt{n}} \right].$$

6. Application numérique : donner l'intervalle de confiance de niveau asymptotique 95% pour θ lorsque $N = 8000$, $n = 100$, $n_1 = \sum_{i=1}^n \mathbb{1}\{X_i = 1\} = 65$.

On a

$$\begin{aligned}\bar{X}_n &= \frac{n_1}{n} = 0.65, \\ \hat{\sigma}_n^2 &= \frac{100 \times 7999}{99 \times 8000} \times 0.65 \times 0.35 = 0.2298, \\ \hat{\sigma}_n &= 0.48.\end{aligned}$$

L'intervalle $[a, b]$ est donc

$$\left[0.65 - \frac{1.96 \times 0.48}{10}; 0.65 + \frac{1.96 \times 0.48}{10} \right] = [0.56; 0.74].$$

Exercice 2. (facultatif) Soient f et g deux densités de probabilité définies sur $[0, 1]$. On suppose que

$$f(x) \leq M g(x), \quad \forall x \in [0, 1]. \quad (20)$$

Soient $\{(X_i, U_i) : i \in \mathbb{N}^*\}$ une suite de variables aléatoires iid telles que

- la loi de X_1 a pour densité g ,
- U_1 est de loi uniforme sur $[0, 1]$,
- X_1 et U_1 sont indépendantes.

On définit les variables aléatoires

$$N = \min \{n : f(X_n) \geq M U_n g(X_n)\}, \quad Y = X_N.$$

1. Soit $q = \mathbf{P}(f(X_n) \geq M U_n g(X_n))$. Montrer que $q = 1/M$.

D'après la loi des espérances itérées, on a

$$\begin{aligned}q &= \mathbf{P}(f(X_n) \geq M U_n g(X_n)) = \mathbf{E}(\mathbf{P}(f(X_n) \geq M U_n g(X_n) | X_n)) \\ &= \mathbf{E}\left(\mathbf{P}\left(U_n \leq \frac{f(X_n)}{M g(X_n)} | X_n\right)\right) \quad \text{avec pour convention } 0/0 = 0 \\ &= \mathbf{E}\left(\frac{f(X_n)}{M g(X_n)}\right) \quad \text{car } U_n | X_n \stackrel{d}{=} U_n \sim \text{Unif}([0, 1]) \text{ car } U_n \perp\!\!\!\perp X_n \\ &= \int_0^1 \frac{f(x)}{M g(x)} g(x) dx \\ &= \frac{1}{M} \int_0^1 f(x) dx \\ &= \frac{1}{M},\end{aligned}$$

car f est une densité sur $[0, 1]$.

2. Montrer que N suit la loi géométrique de paramètre q , c'est-à-dire

$$\mathbf{P}(N = k) = (1 - q)^{k-1} q, \quad k \in \mathbb{N}^*.$$

Soit $(Z_n)_{n \in \mathbb{N}^*}$ la suite de v.a. i.i.d. telle que $Z_n = \mathbb{1}\{f(X_n) \geq MU_ng(X_n)\}$ pour tout $n \in \mathbb{N}^*$. On a, pour tout $k \in \mathbb{N}^*$,

$$\begin{aligned} \mathbf{P}(N = k) &= \mathbf{E} \left(Z_k \prod_{j=1}^{k-1} (1 - Z_j) \right) \\ &\stackrel{\text{II}}{=} \mathbf{E}(Z_k) \prod_{j=1}^{k-1} \mathbf{E}(1 - Z_j) \\ &= \mathbf{P}(f(X_k) \geq MU_k g(X_k)) \prod_{j=1}^{k-1} \mathbf{P}(f(X_j) < MU_j g(X_j)) \\ &= q(1 - q)^{k-1}. \end{aligned}$$

3. Pour toute fonction mesurable bornée $h : [0, 1] \rightarrow \mathbb{R}$, montrer que

$$\mathbf{E}[h(Y)] = \int_0^1 h(y) f(y) dy. \quad (21)$$

En déduire que Y a pour densité f .

Soit $h : [0, 1] \rightarrow \mathbb{R}$ une fonction mesurable bornée. Par la loi des espérances itérées, on a

$$\begin{aligned} \mathbf{E}[h(Y)] &= \mathbf{E}[\mathbf{E}[h(X_N)|N]] \\ &= \sum_{n \in \mathbb{N}^*} \mathbf{E}[h(X_N)|N = n] \mathbf{P}(N = n) \\ &= \sum_{n \in \mathbb{N}^*} \mathbf{E}[h(X_n)|N = n] \mathbf{P}(N = n) \\ &= \sum_{n \in \mathbb{N}^*} \mathbf{E}[h(X_n)|f(X_n) \geq MU_ng(X_n)] \mathbf{P}(N = n) \quad \text{par hyp. i.i.d.} \\ &= \sum_{n \in \mathbb{N}^*} \frac{\int_0^1 \int_0^1 h(y) \mathbb{1}\left\{u \leq \frac{f(y)}{g(y)M}\right\} g(y) dy du}{\mathbf{P}(f(X_n) \geq MU_ng(X_n))} \mathbf{P}(N = n) \\ &= \sum_{n \in \mathbb{N}^*} \frac{\int_0^1 h(y) \frac{f(y)}{g(y)M} g(y) dy}{1/M} \mathbf{P}(N = n) \\ &= \int_0^1 h(y) f(y) dy \underbrace{\left(\sum_{n \in \mathbb{N}^*} \mathbf{P}(N = n) \right)}_{=1} \\ &= \int_0^1 h(y) f(y) dy. \end{aligned}$$

On en déduit que Y a pour densité f .

Remarque Cet exercice montre que si (23) est vrai et si on est capable de générer des variables aléatoires de densité g , alors on sera également capable de générer des variables aléatoires de densité f . Cette méthode porte le nom de méthode d'acceptation-rejet. Elle a toutefois un inconvénient : le nombre moyen de tirage de X_i nécessaire pour calculer Y est $E[N] = M$. Ce nombre peut être excessivement grand lorsque M est trop grand.

TD7 : ESTIMATEUR DU MAXIMUM DE VRAISEMBLANCE

Exercice 1. Soient X_1, \dots, X_n des variables aléatoires i.i.d. pouvant prendre les valeurs 0, 1, 2 avec probabilités $\theta/2, \theta/2, 1 - \theta$. Dans cet exercice, on note N_0, N_1 et N_2 le nombre de 0, de 1 et de 2 dans l'échantillon, c'est-à-dire,

$$N_a = \sum_{i=1}^n \mathbb{1}(X_i = a). \quad (22)$$

1. Dans quel intervalle de \mathbb{R} varie θ ?

On a $\theta \in \Theta \subset \mathbb{R}$ et $X_i \in \{0, 1, 2\}$ pour tout $i \in \{1, \dots, n\}$ presque sûrement. Pour une variable aléatoire réelle Z , on définit son support, noté $\text{supp}(Z)$, par

$$\text{supp}(Z) := \{x \in \mathbb{R} : \mathbf{P}(Z \in B(x, r)) > 0, \forall r > 0\}.$$

Ici X est discrète et prend un nombre fini de valeurs ; on a $\text{supp}(X) = \{0, 1, 2\}$. De plus,

$$\mathbf{P}_\theta(X_1 = 0) = \theta/2, \quad (23)$$

$$\mathbf{P}_\theta(X_1 = 1) = \theta/2, \quad (24)$$

$$\mathbf{P}_\theta(X_1 = 2) = 1 - \theta. \quad (25)$$

Pour que \mathbf{P}_θ soit une probabilité, il faut

$$\forall x \in \text{supp}(X), 0 \leq \mathbf{P}_\theta(X_1 = k) \leq 1, \quad (26)$$

$$\sum_{x \in \text{supp}(X)} \mathbf{P}_\theta(X_1 = x) = 1. \quad (27)$$

(27) est vraie pour tout $\theta \in \mathbb{R}$. (23), (24) et (26) donnent $0 \leq \theta \leq 2$. (25) et (26) donnent $0 \leq \theta \leq 1$. On en déduit

$$\Theta = [0, 1] \cap [0, 2] = [0, 1].$$

2. Proposer un estimateur $\hat{\theta}_n$ de θ par la méthode du maximum de vraisemblance.

Le modèle statistique est

$$(\{0, 1, 2\}^n, \mathcal{P}(\{0, 1, 2\}^n), \{\mathbf{P}_\theta^{\otimes n} : \theta \in \Theta = [0, 1]\}).$$

Le modèle est discret, dominé par la mesure de comptage μ (la densité de probabilité \mathbf{P}_θ par rapport à μ est simplement la fonction de probabilité : $p_\theta(\cdot) = \mathbf{P}_\theta(X = \cdot)$) et on a bien

$\forall \theta \in \Theta, p_\theta << \mu$). La vraisemblance s'écrit

$$\begin{aligned} L_n(X_1, \dots, X_n; \theta) &:= \frac{d\mathbf{P}_\theta^{\otimes n}}{d\mu}(X_1, \dots, X_n) \\ &= \prod_{i=1}^n \mathbf{P}_\theta(X_1 = X_i). \end{aligned}$$

Or, pour tout $x \in \text{supp}(X)$,

$$\begin{aligned} \mathbf{P}_\theta(X_1 = x) &= \left(\frac{\theta}{2}\right)^{\mathbb{1}\{x=0\}} \times \left(\frac{\theta}{2}\right)^{\mathbb{1}\{x=1\}} \times (1-\theta)^{\mathbb{1}\{x=2\}} \\ &= \left(\frac{\theta}{2}\right)^{\mathbb{1}\{x \leq 1\}} \times (1-\theta)^{\mathbb{1}\{x=2\}}. \end{aligned}$$

D'où

$$\begin{aligned} L_n(X_1, \dots, X_n; \theta) &= \prod_{i=1}^n \left(\frac{\theta}{2}\right)^{\mathbb{1}\{X_i \leq 1\}} \times (1-\theta)^{\mathbb{1}\{X_i=2\}} \\ &= \left(\frac{\theta}{2}\right)^{N_0+N_1} \times (1-\theta)^{N_2}. \end{aligned}$$

On en déduit la formule de l'opposée de la log-vraisemblance renormalisée par $1/n$:

$$\ell_n(\theta) = -\frac{1}{n} \log L_n(X_1, \dots, X_n; \theta) = -\frac{N_0 + N_1}{n} \log(\theta/2) - \frac{N_2}{n} \log(1-\theta), \quad (28)$$

avec $n = N_0 + N_1 + N_2$.

Existence et unicité de l'EMV : A partir de maintenant, on pose $\tilde{\Theta} = \text{int}(\Theta) =]0, 1[$. La fonction $\theta \mapsto \ell_n(\theta)$ est \mathcal{C}^2 sur $\tilde{\Theta}$ et on a

$$\forall \theta \in \tilde{\Theta}, \quad \begin{cases} \ell'_n(\theta) = -\frac{N_0+N_1}{\theta n} + \frac{N_2}{(1-\theta)n} = -\frac{N_0+N_1-\theta n}{n\theta(1-\theta)}, \\ \ell''_n(\theta) = \frac{N_0+N_1}{n\theta^2} + \frac{N_2}{n(1-\theta)^2} > 0 \text{ p.s..} \end{cases}$$

On ne déduit que ℓ_n est strictement convexe, i.e., si un minimum existe, il est global, unique, et entièrement caractérisé par les conditions du premier ordre (CNS ici) :

$$\begin{aligned} \ell'_n(\theta) &= 0 \\ \iff \theta &= \frac{N_0 + N_1}{n}. \end{aligned}$$

On en déduit $\hat{\theta}_n = \frac{N_0 + N_1}{n}$.

Remarque : On a

$$\hat{\theta}_n = \frac{1}{n} \sum_{i=1}^n \mathbb{1}\{X_i \leq 1\}.$$

On a également

$$\begin{aligned}
\mathbf{E}_\theta[\mathbb{1}\{X_1 \leq 1\}] &= \mathbf{P}_\theta(X_1 \leq 1) \\
&= \mathbf{P}_\theta(X_1 = 0 \cup X_1 = 1) \\
&= \theta/2 + \theta/2 \\
&= \theta.
\end{aligned}$$

On en conclut que $\hat{\theta}_n$ est l'EMM généralisé avec $\varphi(X) = \mathbb{1}\{X \leq 1\}$.

3. Calculer le risque quadratique de $\hat{\theta}_n$.

Soit $Z_i = \mathbb{1}\{X_i \leq 1\}$. Il est clair que $Z_i \stackrel{iid}{\sim} \mathbf{P}_\theta^Z = \text{Ber}(\theta)$. Comme $\hat{\theta}_n = \frac{1}{n} \sum_{i=1}^n Z_i$, on a

$$\begin{aligned}
\text{biais}(\hat{\theta}_n, \theta) &= \mathbf{E}_\theta(\hat{\theta}_n - \theta) \\
&= \mathbf{E}_\theta\left(\frac{1}{n} \sum_{i=1}^n Z_i\right) - \theta \\
&= \underbrace{\mathbf{E}_\theta(Z_1)}_{=\theta} - \theta \\
&= 0
\end{aligned}$$

pour tout $\theta \in \tilde{\Theta}$. $\hat{\theta}_n$ est un estimateur consistant. De plus, $\mathbf{Var}_\theta(\hat{\theta}_n) \stackrel{\text{H}}{=} \frac{1}{n^2} \sum_{i=1}^n \mathbf{Var}_\theta(Z_i) \stackrel{i.d.}{=} \frac{n}{n^2} \mathbf{Var}_\theta(Z_1) = \frac{\theta(1-\theta)}{n}$. D'où le risque quadratique de $\hat{\theta}_n$:

$$\begin{aligned}
R(\hat{\theta}_n, \theta) &:= \mathbf{E}_\theta [(\hat{\theta}_n - \theta)^2] \\
&= \text{biais}(\hat{\theta}_n, \theta)^2 + \mathbf{Var}_\theta(\hat{\theta}_n) \\
&= \frac{\theta(1-\theta)}{n}.
\end{aligned}$$

4. Calculer l'information de Fisher de ce modèle. Ce dernier, est-il régulier ?

Par définition de l'information de Fisher, on a

$$I_1(\theta) := \mathbf{Var}_\theta[s(X_1; \theta)]$$

avec $s(x; \theta)$ le score du modèle, i.e.

$$s(x; \theta) := \frac{\partial}{\partial \theta} \log \mathbf{P}_\theta(X_1 = x).$$

Il s'agit de la dérivée de la log-vraisemblance par rapport à θ , évaluée en x , pour une observa-

tion. Ici, pour tout $x \in \text{supp}(X)$,

$$\begin{aligned}
s(x; \theta) &= \frac{\partial}{\partial \theta} \log \mathbf{P}_\theta(X_1 = x) \\
&= \frac{\partial}{\partial \theta} [\mathbb{1}(x \leq 1) \log(\theta/2) + \mathbb{1}(x = 2) \log(1 - \theta)] \\
&= \frac{\mathbb{1}(x \leq 1)}{\theta} - \frac{\mathbb{1}(x = 2)}{1 - \theta} \\
&= \frac{\mathbb{1}(x \leq 1) - \theta(\mathbb{1}(x \leq 1) + \mathbb{1}(x = 2))}{\theta(1 - \theta)} \\
&= \frac{\mathbb{1}(x \leq 1) - \theta}{\theta(1 - \theta)},
\end{aligned}$$

car $\mathbb{1}(x \leq 1) + \mathbb{1}(x = 2) = 1$. D'où, $\mathbf{Var}_\theta[s(X_1; \theta)] = \frac{\mathbf{Var}_\theta(Z_1)}{\theta^2(1-\theta)^2} = \frac{1}{\theta(1-\theta)}$ et

$$I_1(\theta) = \frac{1}{1 - \theta}.$$

Remarque : pour n observations i.i.d., l'information de Fischer est donnée par $I_n(\theta) = nI_1(\theta)$.

Le Théorème 5.2 du poly d'A. Tsybakov (p. 118) assure que

$$\forall \theta \in \tilde{\Theta}, \quad \sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \theta(1 - \theta)).$$

sous les hypothèses

H1 : $\tilde{\Theta}$ est un ouvert de \mathbf{R} (ok ici par définition de $\tilde{\Theta}$) et $f(x; \theta) > 0 \iff f(x; \theta') > 0, \forall \theta, \theta' \in \tilde{\Theta}$ ("le support ne dépend pas de θ "). Ici, $f(x; \theta) = \mathbf{P}_\theta(X_1 = x) = (\theta/2)^{\mathbb{1}(x=1)} \times (1 - \theta)^{\mathbb{1}(x=2)} > 0$ pour tout $x \in \text{supp}(X), \theta \in \tilde{\Theta}$.

H2 : Pour μ presque tout x (ici pour tout $x \in \text{supp}(X)$), $\theta \mapsto f(x; \theta)$ et $\theta \mapsto \log f(x; \theta)$ sont \mathcal{C}^2 sur $\tilde{\Theta}$. Ici, $\theta \mapsto f(x; \theta) = \mathbf{P}_\theta(X_1 = x)$ est bien \mathcal{C}^2 et $\theta \mapsto \log f(x; \theta) = \log \mathbf{P}_\theta(X_1 = x) = \mathbb{1}(x \leq 1) \log(\theta/2) + \mathbb{1}(x = 2) \log(1 - \theta)$ est bien \mathcal{C}^2 (car \log est \mathcal{C}^2 et les polynômes sont \mathcal{C}^∞ donc \mathcal{C}^2).

H3 : Pour tout $\theta^* \in \tilde{\Theta}$, il existe un intervalle ouvert $U_{\theta^*} \subset \tilde{\Theta}$ contenant θ^* et une fonction borélienne $\Lambda(x)$ tels que pour tout $\theta \in U_{\theta^*}$ et μ presque tout $x \in \text{supp}(X)$ en notant $\ell(x; \theta) = \log f(x; \theta)$,

$$\begin{cases} |\ell'(x; \theta)| &\leq \Lambda(x) \\ |\ell'(x; \theta)|^2 &\leq \Lambda(x) \\ |\ell''(x; \theta)| &\leq \Lambda(x) \end{cases}$$

et

$$\int \Lambda(x) \sup_{\theta \in U_{\theta^*}} f(x; \theta) d\mu(x) < +\infty.$$

Soit $\theta^* \in \tilde{\Theta}, \varepsilon > 0$. Si $\theta \in U_{\theta^*} :=]\theta^* - \varepsilon, \theta^* + \varepsilon[$, alors pour tout $x \in \text{supp}(X)$

$$|\ell'(x; \theta)| \leq \frac{1}{\theta} + \frac{1}{1 - \theta} \leq \frac{1}{\theta^* - \varepsilon} + \frac{1}{1 - \theta^* - \varepsilon}.$$

On a aussi

$$\begin{aligned}
|\ell'(x; \theta)|^2 &\leq \frac{1}{\theta^2} + \frac{1}{(1-\theta)^2} + 2 \frac{1}{\theta(1-\theta)} \\
&\leq \frac{1}{\theta^2} + \frac{1}{(1-\theta)^2} + 2 \left(\frac{1}{\theta^2} + \frac{1}{(1-\theta)^2} \right) \quad \text{car } \theta \in \tilde{\Theta} \implies \frac{1}{\theta} \leq \frac{1}{\theta^2} \text{ et } \frac{1}{1-\theta} \leq \frac{1}{(1-\theta)^2} \\
&\leq 3 \left(\frac{1}{\theta^2} + \frac{1}{(1-\theta)^2} \right) \\
&\leq 3 \left(\frac{1}{(\theta^* - \varepsilon)^2} + \frac{1}{(1 - \theta^* - \varepsilon)^2} \right)
\end{aligned}$$

et

$$|\ell''(x; \theta)| \leq \frac{1}{\theta^2} + \frac{1}{(1-\theta)^2} \leq \frac{1}{(\theta^* - \varepsilon)^2} + \frac{1}{(1 - \theta^* - \varepsilon)^2}.$$

Ainsi, $\Lambda(x) := 3 \left(\frac{1}{(\theta^* - \varepsilon)^2} + \frac{1}{(1 - \theta^* - \varepsilon)^2} \right) \mathbb{1}\{x \in \text{supp}(X)\}$ convient et on a, comme $\theta \mapsto f(x; \theta)$ est continue sur $\overline{U_{\theta^*}}$ et donc atteint ses bornes (théorème des valeurs extrêmes),

$$\sup_{\theta \in U_{\theta^*}} f(x; \theta) \leq C < +\infty,$$

pour une certaine constante $C > 0$. On en conclut,

$$\begin{aligned}
\int \Lambda(x) \sup_{\theta \in U_{\theta^*}} f(x; \theta) d\mu(x) &\leq C \int \Lambda(x) d\mu(x) \\
&= C \sum_{x \in \text{supp}(X)} 3 \left(\frac{1}{(\theta^* - \varepsilon)^2} + \frac{1}{(1 - \theta^* - \varepsilon)^2} \right) \\
&= 9C \left(\frac{1}{(\theta^* - \varepsilon)^2} + \frac{1}{(1 - \theta^* - \varepsilon)^2} \right) < +\infty.
\end{aligned}$$

H 4 : L'information de Fischer vérifie bien

$$I_1(\theta) > 0, \forall \theta \in \tilde{\Theta}.$$

Exercice 2. Soient X_1, \dots, X_n des variables aléatoires i.i.d. de loi de Cauchy $\mathcal{C}(\theta^*)$ avec $\theta^* > 0$. C'est-à-dire, chaque X_i a pour densité par rapport à la mesure de Lebesgue la fonction

$$f(x, \theta^*) = \frac{\theta^*}{\pi((\theta^*)^2 + x^2)}, \quad \forall x \in \mathbb{R}. \tag{29}$$

1. Montrer que ce modèle est régulier.

On doit ici encore vérifier les hypothèses H1-H4 du poly d'A. Tsybakov.

H1 : $\Theta = \mathbb{R}_+$ et même support pour tout $x \in \mathbb{R}$ ($f(x; \theta) > 0, \forall (x, \theta) \in \mathbb{R} \times \Theta$).

H2 : $\theta \mapsto f(x; \theta)$ est $C^2(\mathbf{R}_+^*)$ car c'est un ratio de polynômes. Pour tout $x \in \mathbf{R}$,

$$\begin{aligned}\ell(x; \theta) &:= \log f(x; \theta) \\ &= \log \theta - \log(\theta^2 + x^2) - \log \pi\end{aligned}$$

comme fonction de θ est $C^2(\mathbf{R}_+^*)$ car \log est $C^2(\mathbf{R}_+^*)$.

H3 : On a

$$\ell'(x; \theta) := s(x, \theta) = \frac{\partial}{\partial \theta} \ell(x; \theta) = \frac{1}{\theta} - \frac{2\theta}{\theta^2 + x^2},$$

et

$$\begin{aligned}\ell''(x; \theta) &= -\frac{1}{\theta^2} - 2 \times \frac{(\theta^2 + x^2) - \theta \times 2 \times \theta}{(\theta^2 + x^2)^2} \\ &= -\frac{1}{\theta^2} + 2 \frac{\theta^2 - x^2}{(\theta^2 + x^2)^2}.\end{aligned}$$

Comme $(\theta^2 + x^2)^2 \geq \theta^2$, on a

$$\frac{2\theta^2}{(\theta^2 + x^2)^2} \leq \frac{2\theta^2}{(\theta^2)^2} = \frac{2}{\theta^2}.$$

On a également,

$$\begin{aligned}\frac{2x^2}{(\theta^2 + x^2)^2} &= \frac{2x^2}{\theta^4 + x^4 + 2\theta^2 x^2} \\ &= \frac{2}{\underbrace{x^2}_{>0} + 2\theta^2 + \underbrace{\frac{\theta^4}{x^2}}_{>0}} \\ &\leq \frac{2}{2\theta^2} = \frac{1}{\theta^2}.\end{aligned}$$

Soit $\varepsilon > 0$. Sur le voisinage $]\theta^* - \varepsilon, \theta^* + \varepsilon[:= U_{\theta^*}$, l'inégalité triangulaire donne

$$|\ell'(x; \theta)| \leq \frac{1}{\theta} + \frac{2\theta}{\theta^2} = \frac{3}{\theta} \leq \frac{3}{\theta^* - \varepsilon},$$

et

$$|\ell''(x; \theta)|^2 \leq \frac{9}{(\theta^* - \varepsilon)^2}$$

et

$$|\ell''(x; \theta)| \leq \frac{1}{\theta^2} + \frac{2}{\theta^2} + \frac{1}{\theta^2} \leq \frac{4}{(\theta^* - \varepsilon)^2}.$$

Donc en prenant $\Lambda(x) = \text{le maximum ou la somme des fonctions dominantes (indépendante de } x\text{), on a}$

$$\int \Lambda(x) \sup_{\theta \in U_{\theta^*}} f(x; \theta) d\lambda(x) \leq C \int \sup_{\theta \in U_{\theta^*}} f(x; \theta) d\lambda(x).$$

On peut borner $f > 0$ par

$$|f(x; \theta)| \leq \frac{\theta^* + \varepsilon}{\pi((\theta^* - \varepsilon)^2 + x^2)}$$

qui est bien intégrable par rapport à Lebesgue.

H4 : Le score est

$$s(x; \theta) = \frac{\partial}{\partial \theta} \ell(x; \theta) = \frac{1}{\theta} - \frac{2\theta}{\theta^2 + x^2}.$$

Supposons $I_1(\theta) = 0 = \mathbf{E}_\theta[s(X_1; \theta)^2]$ pour un certain $\theta \in \Theta$. Alors,

$$\frac{2\theta}{\theta^2 + X^2} = \frac{1}{\theta} \text{ p.s.}$$

Donc $X^2 = \theta^2$ p.s.. Or comme la loi de Cauchy admet une densité absolument continue par rapport à la mesure de Lebesgue, on a

$$\mathbf{P}_\theta(X^2 = \theta^2) = \underbrace{\mathbf{P}_\theta(X = -\theta)}_{=0} + \underbrace{\mathbf{P}_\theta(X = \theta)}_{=0} = 0.$$

On en conclut $I_1(\theta) > 0, \forall \theta \in \Theta$.

2. Calculer la dérivée de la log-vraisemblance $\ell_n(\theta)$ et montrer qu'elle s'annule en un seul point, si l'événement $\{X_1 \neq 0, \dots, X_n \neq 0\}$ est réalisé.

$$L_n(X_1, \dots, X_n; \theta) \stackrel{iid}{=} \prod_{i=1}^n f(X_i; \theta).$$

On en déduit

$$\begin{aligned} \ell_n(\theta) &:= -\frac{1}{n} \log L_n(X_1, \dots, X_n; \theta) \\ &= -\log(\theta) + \frac{1}{n} \sum_{i=1}^n \log(\theta^2 + X_i^2) + \log \pi. \end{aligned}$$

$\theta \mapsto \log(\theta)$ est \mathcal{C}^2 sur Θ et on a

$$\ell'_n(\theta) = -\frac{1}{\theta} + \frac{1}{n} \sum_{i=1}^n \frac{2\theta}{\theta^2 + X_i^2}.$$

D'où

$$\begin{aligned} \psi(\theta) &:= \theta \ell'_n(\theta) \\ &= -1 + \frac{2}{n} \sum_{i=1}^n \underbrace{\frac{\theta^2}{\theta^2 + X_i^2}}_{=1-\frac{X_i^2}{\theta^2+X_i^2}} \\ &= 1 - \frac{2}{n} \sum_{i=1}^n \frac{X_i^2}{\theta^2 + X_i^2}. \end{aligned}$$

La fonction ψ est strictement croissante (sauf si tous les X_i sont nuls, qui un événement de probabilité nulle) et \mathcal{C}^0 sur \mathbf{R}_+^* . De plus,

$$\psi(0) = -1 \text{ et } \lim_{t \rightarrow +\infty} \psi(t) = 1.$$

Il en découle (théorème de la bijection monotone) qu'il existe un unique point $\hat{\theta}_n$ tel que

$$\psi(\hat{\theta}_n) = 0.$$

3. En déduire qu'en dehors d'un événement de probabilité nulle, l'EMV existe, est unique et est asymptotiquement efficace.

$\hat{\theta}_n$ vérifie

- $\theta\ell'_n(\theta) < 0$ pour tout $0 < \theta < \hat{\theta}_n$, et donc $\ell'_n(\theta) < 0$ pour tout $0 < \theta < \hat{\theta}_n$.
- $\hat{\theta}_n \ln'(\hat{\theta}_n) = 0$ qui implique $\ell'_n(\hat{\theta}_n) = 0$ car $\hat{\theta}_n \in \Theta = \mathbf{R}_+^*$.
- $\theta\ell'_n(\theta) > 0$ pour tout $\theta > \hat{\theta}_n$, et donc $\ell'_n(\theta) > 0$ pour tout $\theta > \hat{\theta}_n$.

Ainsi, $\hat{\theta}_n$ est le seul point de minimum de ℓ_n , c'est donc l'EMV (il existe et est unique). On a montré que le modèle est régulier, un théorème du cours assure que l'EMV est asymptotiquement efficace.

Exercice 3. Soient X_1, \dots, X_n des variables aléatoires i.i.d. de loi \mathbf{P}_{θ^*} , où \mathbf{P}_θ est la loi sur \mathbb{R} admettant pour densité

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

1. Le modèle statistique $\{\mathbf{P}_\theta : \theta \in \mathbb{R}\}$ est-il identifiable ?

Oui, si $X \sim \mathbf{P}_\theta$, alors

$$\mathbf{E}_\theta(X) = \frac{\theta + \theta + 1}{2} = \theta + 1/2.$$

Donc

$$\begin{aligned} \mathbf{P}_\theta = \mathbf{P}_{\theta'} &\implies \mathbf{E}_\theta(X) = \mathbf{E}_{\theta'}(X) \\ &\implies \theta + \frac{1}{2} = \theta' + \frac{1}{2} \\ &\implies \theta = \theta'. \end{aligned}$$

Ok : \mathbf{P}_θ est injective.

2. Le modèle statistique $\{\mathbf{P}_\theta : \theta \in \mathbb{R}\}$ est-il régulier ?

Non, car le support dépend de θ . Si $\theta = 0$ et $\theta' = 2$, on a

- $f(x; \theta) > 0, \forall x \in]0, 1[$,
- mais $f(x; \theta') = 0, \forall x \in]0, 1[$. L'hypothèse H1 n'est pas respectée.

On remarque également que l'hypothèse H2 n'est pas respectée non plus.

3. L'estimateur du maximum de vraisemblance du paramètre θ^* est-il unique ? Déterminer l'ensemble de points de maximum de la vraisemblance de ce modèle.

$$\begin{aligned} L_n(X_1, \dots, X_n; \theta) &\stackrel{iid}{=} \prod_{i=1}^n \mathbb{1}\{X_i \in [\theta, \theta + 1]\} \\ &= \mathbb{1}\{X_{(1)} \geq \theta\} \mathbb{1}\{X_{(n)} \leq \theta + 1\} \end{aligned}$$

où $X_{(1)} = \min_{1 \leq i \leq n} X_i$ et $X_{(n)} = \max_{1 \leq i \leq n} X_i$ sont des statistiques d'ordre. On peut réécrire

$$L_n(X_1, \dots, X_n; \theta) = \mathbb{1}\{\theta \in [X_{(n)} - 1; X_{(1)}]\} \in \{0, 1\}.$$

Cette expression de $L_n(X_1, \dots, X_n; \theta)$ montre que toute valeur de θ incluse entre $X_{(n)} - 1$ et $X_{(1)}$ maximise la vraisemblance. On en déduit que l'EMV n'est pas unique.

4. On considère l'estimateur $\hat{\theta}_n = X_{(1)} = \min_{1 \leq i \leq n} X_i$. Sans calculer le biais, montrer que cet estimateur est biaisé.

Raisonnons par l'absurde. Supposons que $\hat{\theta}_n$ est sans biais, i.e.

$$\mathbf{E}_\theta(\hat{\theta}_n) = \theta \quad \forall \theta \in \mathbf{R}.$$

Sous \mathbf{P}_θ , on sait que $X_1, \dots, X_n \in [\theta, \theta + 1]$ p.s. Donc $X_{(1)} \geq \theta$ p.s.

$$\mathbf{E}_\theta(\hat{\theta}_n) = \theta \iff \mathbf{E}_\theta(\hat{\theta}_n - \theta) = 0.$$

Mais

$$\begin{aligned} \hat{\theta}_n - \theta \geq 0 \text{ et } \mathbf{E}_\theta(\hat{\theta}_n - \theta) = 0 &\implies \hat{\theta}_n - \theta = 0 \text{ p.s.} \\ &\implies \min_{1 \leq i \leq n} X_i = \theta \text{ p.s.} \\ &\implies \exists i \text{ t.q. } X_i = \theta. \end{aligned}$$

Or, une borne d'union donne

$$\begin{aligned} \mathbf{P}_\theta(\exists i : X_i = \theta) &= \mathbf{P}_\theta\left(\bigcup_{i=1}^n \{X_i = \theta\}\right) \\ &\leq \sum_{i=1}^n \mathbf{P}_\theta(X_i = \theta) \\ &= 0. \end{aligned}$$

5. Déterminer la densité $g_n(x; \theta^*)$ de la variable aléatoire $\hat{\theta}_n$.

On commence par calculer la fonction de répartition de $\hat{\theta}_n$. Pour tout $(x, \theta) \in \mathbf{R} \times \mathbf{R}$,

$$\begin{aligned}
G_n(x; \theta) &:= \mathbf{P}_\theta(\hat{\theta}_n \leq x) \\
&= \mathbf{P}_\theta(\min_{1 \leq i \leq n} X_i \leq x) \\
&= 1 - \mathbf{P}_\theta(\min_{1 \leq i \leq n} X_i > x) \\
&= 1 - \mathbf{P}_\theta\left(\bigcap_{i=1}^n \{X_i > x\}\right) \\
&\stackrel{iid}{=} 1 - \mathbf{P}_\theta(X_1 > x)^n \\
&= 1 - (1 - F_\theta(x))^n
\end{aligned}$$

avec $F_\theta(x) := \mathbf{P}_\theta(X_1 \leq x)$ la f.d.r. de X_1 . On en déduit,

$$g_n(x; \theta) = G'_n(x; \theta) \quad (30)$$

$$= -n(1 - F_\theta(x))^{n-1} \times (1 - F_\theta(x))' \quad (31)$$

$$= nf(x; \theta)(1 - F_\theta(x))^{n-1} \quad (32)$$

où

$$F_\theta(x) = \int_{-\infty}^x f(t; \theta) dt = (x - \theta)\mathbb{1}_{[\theta, \theta+1]}(x) + \mathbb{1}_{[\theta+1, +\infty[}(x).$$

Donc

$$g_n(x; \theta) = \begin{cases} 0 & \text{si } x < \theta, \\ n(1 - x + \theta)^{n-1} & \text{si } x \in [\theta; \theta + 1], \\ 0 & \text{si } x > \theta + 1. \end{cases}$$

, i.e.,

$$g_n(x; \theta) = \mathbb{1}_{[\theta, \theta+1]}(x) \times n(1 - x + \theta)^{n-1}.$$

6. Calculer le risque de $\hat{\theta}_n$. Proposer un estimateur $\bar{\theta}_n$ qui a exactement la même variance mais qui est sans biais. Comparer les deux risques.

$$\mathbf{E}_\theta(\hat{\theta}_n) = \int_{\mathbf{R}} x g_n(x; \theta) dx = n \int_{\theta}^{\theta+1} x(1 - x + \theta)^{n-1} dx.$$

En faisant le changement de variable $y = \theta + 1 - x \implies x = \theta + 1 - y$, on obtient

$$\begin{aligned}
\mathbf{E}_\theta(\hat{\theta}_n) &= n \int_0^1 (\theta + 1 - y)y^{n-1} dy \\
&= (\theta + 1) \times n \int_0^1 y^{n-1} dy - n \int_0^1 y^n dy \\
&= \theta + 1 - \frac{n}{n+1} = \theta + \frac{1}{n+1}.
\end{aligned}$$

On retrouve le résultat de la question 4. On a

$$\begin{aligned}
\mathbf{E}_\theta[\hat{\theta}_n^2] &= n \int_0^1 (\theta + 1 - y)^2 y^{n-1} dy \\
&= (\theta + 1)^2 - \frac{2n(\theta + 1)}{n+1} + \frac{n}{n+2} \\
&= \theta^2 + \frac{2\theta}{n+1} + 1 - \frac{2n}{n+1} + \frac{n}{n+2} \\
&= \theta^2 + \frac{2\theta}{n+1} + \frac{n^2 + 3n + 2 - 2n^2 - 4n + n^2 + n}{(n+1)(n+2)} \\
&= \theta^2 + \frac{2\theta}{n+1} + \frac{2}{(n+1)(n+2)}.
\end{aligned}$$

Ainsi,

$$\begin{aligned}
R(\hat{\theta}_n, \theta) &= \mathbf{E}_\theta[(\hat{\theta}_n - \theta)^2] \\
&= \mathbf{E}_\theta(\hat{\theta}_n^2) - 2\theta\mathbf{E}_\theta(\hat{\theta}_n) + \theta^2 \\
&= \frac{2\theta}{n+1} + \frac{2}{(n+1)(n+2)} - \frac{2\theta}{n+1} \\
&= \frac{2}{(n+1)(n+2)}.
\end{aligned}$$

Le risque ne dépend pas de la valeur de θ .

Remarque : le risque de $\hat{\theta}_n$ décroît en n^2 (n pour le RMSE), c'est plus rapide que la vitesse "habituelle" en \sqrt{n} du TCL. Cela est dû au fait que le modèle considéré n'est pas régulier.

L'estimateur $\bar{\theta}_n = \hat{\theta}_n - \frac{1}{n+1}$ a évidemment la même variance mais il est sans biais. Son risque est donc

$$\begin{aligned}
R(\bar{\theta}_n, \theta) &= R(\hat{\theta}_n, \theta) - \text{biais}(\hat{\theta}_n)^2 \\
&= \frac{2}{(n+1)(n+2)} - \frac{1}{(n+1)^2} \\
&= \frac{1}{n+1} \left(\frac{2}{n+2} - \frac{1}{n+1} \right) \\
&= \frac{n}{(n+1)^2(n+2)}.
\end{aligned}$$

Ce risque est presque deux fois plus petit que celui de $\hat{\theta}_n$.

7. Montrer que $\hat{\theta}_n$ est faiblement consistant. En déduire qu'il est également fortement consistant.

Soit $\theta^* \in \mathbf{R}$. On a $\hat{\theta}_n \xrightarrow{L_2(\mathbf{P}_{\theta^*})} \theta^*$ car le risque tend vers 0 quand $n \rightarrow +\infty$. Donc

$$\hat{\theta}_n \xrightarrow{P} \theta^* \quad \mathbf{P}_{\theta^*}.$$

De plus, la suite $\{\hat{\theta}_n\}$ est décroissante donc bornée inférieurement par θ^* . Donc $\hat{\theta}_n \xrightarrow{p.s.} \eta$. On en déduit que $\hat{\theta}_n \xrightarrow{P} \eta$, donc $\eta = \theta^*$ p.s. (cf. TD 1) et

$$\hat{\theta}_n \xrightarrow{p.s.} \theta^*, \quad \mathbf{P}_{\theta^*}.$$

8. En calculant la fonction de répartition de la variable aléatoire $Y_n = n(\hat{\theta}_n - \theta^*)$, montrer que cette suite converge en loi vers une loi exponentielle.

For all $\theta \in \mathbf{R}$,

$$\begin{aligned} \mathbf{P}_\theta(Y_n \leq t) &= \mathbf{P}_\theta\left(\hat{\theta}_n \leq \theta + \frac{t}{n}\right) = G_n\left(\theta + \frac{t}{n}; \theta\right) \\ &= 1 - \left(1 - F_\theta\left(\theta + \frac{t}{n}\right)\right)^n. \end{aligned}$$

Si $t \leq 0$, il est clair que $\mathbf{P}_\theta(Y_n \leq t) = 0$. Si $t > 0$ est fixé, pour n assez grand, on a $t/n \leq 1$ et, par conséquent,

$$\begin{aligned} \mathbf{P}_\theta(Y_n \leq t) &= 1 - \left(1 - \frac{t}{n}\right)^n \\ &= 1 - \exp(n \log(1 - \frac{t}{n})) \\ &\underset{n \rightarrow \infty}{\sim} 1 - \exp(-n \times \frac{t}{n}) = 1 - e^{-t}. \end{aligned}$$

On a donc

$$\mathbf{P}_\theta(Y_n \leq t) \xrightarrow{n \rightarrow +\infty} (1 - e^{-t}) \mathbf{1}(t \geq 0).$$

On retrouve à droite la f.d.r. d'une loi exponentielle de paramètre 1, notée $\mathcal{E}(1)$. Il en résulte que

$$Y_n \xrightarrow{\mathcal{L}} \mathcal{E}(1),$$

i.e.

$$n(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{E}(1).$$

9. (Facultatif) On peut également considérer l'estimateur $\tilde{\theta}_n = \max_i X_i - 1$. Dire, sans faire des calculs, lequel des deux estimateurs $\hat{\theta}_n$ et $\tilde{\theta}_n$ est préférable ?

Soit $\theta^* \in \mathbf{R}$. On pose $\tilde{\theta}_n = \max_{1 \leq i \leq n} X_i - 1$. Les deux estimateurs $\hat{\theta}_n, \tilde{\theta}_n$ sont équivalents en

termes de risque. En effet, les deux variables aléatoires

$$\hat{\theta}_n - \theta^* \text{ et } \theta^* - \tilde{\theta}_n$$

ont exactement la même loi. Pour s'en convaincre, on pose

$$\tilde{X}_i = 2\theta^* + 1 - X_i.$$

On vérifie facilement que $\tilde{X}_1, \dots, \tilde{X}_n \stackrel{iid}{\sim} \text{Unif}([\theta^*, \theta^* + 1])$. D'autre part,

$$\begin{aligned}\theta^* - \tilde{\theta}_n &= \theta^* - \max_{1 \leq i \leq n} X_i + 1 \\ &= \min_{1 \leq i \leq n} (\theta^* - X_i + 1) \\ &= \min_{1 \leq i \leq n} (2\theta^* + 1 - X_i) - \theta^* \\ &= \min_{1 \leq i \leq n} \tilde{X}_i - \theta^* \\ &\stackrel{d}{=} \underbrace{\min_{1 \leq i \leq n} X_i}_{=\hat{\theta}_n} - \theta^*.\end{aligned}$$

Il en découle que les variables $|\hat{\theta}_n - \theta^*|$ et $|\tilde{\theta}_n - \theta^*|$ ont la même loi, donc le même moment d'ordre 2. Donc les risques de $\hat{\theta}_n$ et $\tilde{\theta}_n$ sont égaux.

TD8 : ESTIMATEUR DU MAXIMUM DE VRAISEMBLANCE

Exercice 1. (données censurées) Soient Y_1, \dots, Y_n des variables aléatoires i.i.d. de loi exponentielle de moyenne $\theta^* > 0$. On considère le cadre des données censurées : pour une certaine constante connue $c > 0$, on observe les variables aléatoires

$$X_i = Y_i \wedge c = \min(Y_i, c).$$

1. Déterminer la fonction de répartition $F_{\theta^*}(x)$ de X_1 .

On a

$$\begin{aligned} \forall x \geq c, X_1 \leq x \text{ p.s.} &\implies \forall x \geq c, F_{\theta^*}(x) = 1, \\ \forall x \leq 0, X_1 \geq x \text{ p.s.} &\implies \forall x \leq 0, F_{\theta^*}(x) = 0. \end{aligned}$$

Pour tout $x \in]0, c[$,

$$\begin{aligned} F_{\theta^*}(x) &:= \mathbf{P}_\theta(X_1 \leq x) \\ &= \mathbf{P}_\theta(\min(Y_1, c) \leq x) \\ &= \mathbf{P}_\theta(0 \leq Y_1 \leq x) \quad \text{car on ne peut avoir } X_1 \leq x \text{ et } Y_1 > c \\ &= \int_0^x \frac{1}{\theta^*} e^{-\frac{t}{\theta^*}} dt \\ &= 1 - e^{-\frac{x}{\theta^*}}. \end{aligned}$$

On en déduit

$$F_{\theta^*}(x) = \left(1 - e^{-\frac{x}{\theta^*}}\right) \mathbb{1}_{[0,c]}(x) + \mathbb{1}_{[c,+\infty[}(x).$$

2. Montrer que l'estimateur par la méthode des moments s'écrit sous la forme $\hat{\theta}_n^{\text{MM}} = \psi(\bar{X}_n)$, où $\psi : [0, c[\rightarrow [0, \infty[$ est une fonction strictement croissante et infiniment différentiable.

Calculons $\mu(\theta) := \mathbf{E}_\theta(X_1)$. On a la décomposition suivante

$$X_1 = Y_1 \mathbb{1}\{Y_1 \leq c\} + c \mathbb{1}\{Y_1 > c\}.$$

Par linéarité de l'espérance, on a

$$\begin{aligned}
\mu(\theta) &:= \mathbf{E}_\theta(X_1) \\
&= \mathbf{E}_\theta[Y_1 \mathbb{1}\{Y_1 \leq c\}] + c \mathbf{E}_\theta[\mathbb{1}\{Y_1 > c\}] \\
&= \int_0^{+\infty} y \times \frac{1}{\theta} e^{-y/\theta} \mathbb{1}\{y \leq c\} dy + ce^{-c/\theta} \\
&= \theta \int_0^{c/\theta} e^{-z} dz + ce^{-c/\theta} \text{ avec le changement de variable } y = \theta z.
\end{aligned}$$

Une intégration par parties donne

$$\int_0^a z e^{-z} dz = -ae^{-a} + 1 - e^{-a}.$$

On en conclut

$$\begin{aligned}
\mu(\theta) &= \theta \times \left(-\frac{c}{\theta} e^{-c/\theta} + 1 - e^{-c/\theta} \right) + ce^{-c/\theta} \\
&= \theta(1 - e^{-c/\theta}).
\end{aligned}$$

L'estimateur par la méthode des moments, $\hat{\theta}_n^{\text{MM}}$, est solution de

$$\mu(\hat{\theta}_n^{\text{MM}}) = \bar{X}_n.$$

Montrons que l'on peut inverser μ . La fonction μ est strictement croissante car μ est dérivable comme produit de fonctions dérivables et on a

$$\mu'(\theta) = 1 - e^{-c/\theta} - \theta \frac{c}{\theta^2} e^{-c/\theta} = e^{-c/\theta} (e^{c/\theta} - 1 - \frac{c}{\theta}).$$

Or on a que $e^x \geq 1 + x$ pour tout $x \in \mathbf{R}$. De plus, $\mu(0) = 0$ et

$$\begin{aligned}
\lim_{\theta \rightarrow +\infty} \mu(\theta) &= \lim_{\theta \rightarrow +\infty} \theta \times (1 - e^{-c/\theta}) \\
&= \lim_{x \rightarrow 0} \frac{1 - e^{-cx}}{x} = c,
\end{aligned}$$

puisque le développement limité de e^x en 0 est $e^x \underset{x \rightarrow 0}{=} 1 + x + o(x)$. Donc $\mu : [0, +\infty[\longrightarrow [0, c[$ est strictement croissante (elle réalise une bijection) et ainsi $\psi = \mu^{-1}$ existe et est strictement croissante de $[0, c[$ à $[0, +\infty[$. Ainsi,

$$\hat{\theta}_n^{\text{MM}} = \psi(\bar{X}_n).$$

De plus, $\mu \in \mathcal{C}^\infty \implies \psi \in \mathcal{C}^\infty$.

3. Montrer que $\hat{\theta}_n^{\text{MM}}$ est fortement consistant et asymptotiquement normal.

Consistance: Comme ψ est continue et $\bar{X}_n \xrightarrow{p.s.} \mu(\theta^*)$ par la loi forte des grands nombres (les $(X_i)_i$ étant i.i.d. bornées elles appartiennent à L^∞ , a fortiori à L^1). Le premier théorème de

continuité (“CMT”) donne

$$\hat{\theta}_n^{\text{MM}} \xrightarrow{p.s.} \psi(\mu(\theta^*)) = \theta^*.$$

Normalité asymptotique : les $(X_i)_i$ étant i.i.d. de L^2 , le Théorème central limite donne

$$\sqrt{n}(\bar{X}_n - \mu(\theta^*)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \text{Var}_{\theta^*}(X_1)).$$

ψ étant C^∞ , la delta-méthode assure

$$\sqrt{n}(\psi(\bar{X}_n) - \underbrace{\theta^*}_{=\psi(\mu(\theta^*))}) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2) \quad \forall \theta^* \in \Theta,$$

où $\sigma^2 = \psi'(\theta^*)^2 \times \text{Var}_{\theta^*}(X_1)$.

4. Montrer que le modèle $\{F_\theta : \theta > 0\}$ est dominé par la mesure

$$\mu = \lambda + \delta_c$$

où λ désigne la mesure de Lebesgue et δ_c est la masse de Dirac en c .

Mesure qui domine le modèle : ici i.i.d., on cherche μ telle que

$$\forall \theta \in \Theta, P_\theta^{X_1} << \mu.$$

Soit $A \in \mathcal{B}(\mathbf{R})$ tel que $\mu(A) = 0$. Il faut montrer que $\mathbf{P}_\theta(X_1 \in A) = 0, \forall \theta \in \Theta$.

$$\begin{aligned} \mu(A) = 0 &\iff \lambda(A) = 0 \text{ et } \delta_c(A) = 0 \\ &\iff \lambda(A) = 0 \text{ et } c \notin A. \end{aligned}$$

D'où,

$$\begin{aligned} \mathbf{P}_\theta(X_1 \in A) &= \mathbf{P}_\theta(Y_1 \in A, Y_1 < c) + \underbrace{\mathbf{P}_\theta(c \in A, Y_1 \geq c)}_{\leq \mathbf{P}_\theta(c \in A)} \\ &= \mathbf{P}_\theta(Y_1 \in A, Y_1 < c) \\ &\leq \mathbf{P}_\theta(Y_1 \in A) \\ &= \int_A \frac{1}{\theta} e^{-y/\theta} dy = 0 \text{ car } \lambda(A) = 0. \end{aligned}$$

5. Déterminer la densité de P_θ par rapport à μ et en déduire que le fonction de vraisemblance peut s'écrire sous la forme

$$L_n(x_1, \dots, x_n; \theta) = \frac{1}{\theta^{N_c}} \exp \left\{ -\frac{1}{\theta} \sum_{i=1}^n x_i \right\} \mathbb{1}(x_{(n)} \leq c), \quad N_c = \sum_{i=1}^n \mathbb{1}(x_i < c)$$

pour tout $x_1, \dots, x_n \geq 0$.

Trouvons la densité par la méthode d'intégration contre des fonctions tests (cf. TD6, Exercice 2). Soit $h : \mathbf{R} \rightarrow \mathbf{R}$ une fonction mesurable bornée. On a

$$\begin{aligned}\mathbf{E}_\theta[h(X_1)] &= \mathbf{E}_\theta[h(Y_1)\mathbf{1}\{Y_1 \leq c\}] + h(c)\mathbf{P}_\theta(Y_1 \geq c) \\ &= \int_0^{+\infty} h(y) \frac{e^{-y/\theta}}{\theta} \mathbf{1}(y < c) dy + h(c)e^{-c/\theta} \\ &= \int_0^{+\infty} h(y) \frac{e^{-y/\theta}}{\theta} \mathbf{1}(y < c)\mu(dy) + \int_0^{+\infty} h(y)e^{-c/\theta} \mathbf{1}(y = c)\mu(dy) \\ &= \int_0^{+\infty} h(y) \left\{ \frac{e^{-y/\theta}}{\theta} \mathbf{1}(y < c) + e^{-c/\theta} \mathbf{1}(y = c) \right\} \mu(dy).\end{aligned}$$

Donc, la densité de P_θ par rapport à μ est

$$f(x; \theta) = \frac{1}{\theta} e^{-x/\theta} \mathbf{1}(x < c) + e^{-c/\theta} \mathbf{1}(x = c).$$

On peut aussi l'écrire

$$f(x; \theta) = \left(\frac{1}{\theta} \right)^{\mathbf{1}(x < c)} \times e^{-x/\theta} \mathbf{1}(x \leq c).$$

Finalement,

$$\begin{aligned}L_n(X_1, \dots, X_n; \theta) &\stackrel{i.i.d.}{=} \prod_{i=1}^n f(X_i; \theta) \\ &= \frac{1}{\theta^{N_c}} e^{-\frac{1}{\theta} \sum_{i=1}^n X_i} \mathbf{1}(\max_{1 \leq i \leq n} X_i \leq c).\end{aligned}$$

6. Trouver l'estimateur du maximum de vraisemblance $\hat{\theta}_n^{\text{MV}}$ et montrer qu'il est fortement consistant.

L'opposée de la log-vraisemblance re-normalisée est donnée par

$$\begin{aligned}\ell_n(\theta) &:= -\frac{1}{n} \log L_n(X_1, \dots, X_n; \theta) \\ &= \frac{N_c}{n} \log(\theta) + \frac{1}{\theta} \bar{X}_n.\end{aligned}$$

On fait abstraction de l'indicatrice car elle vaut 1 p.s.. Ainsi, ℓ_n est bien dérivable et on a

$$\ell'_n(\theta) = \frac{N_c}{\theta n} - \frac{\bar{X}_n}{\theta^2} = \frac{N_c}{\theta^2 n} \left(\theta - \frac{n \bar{X}_n}{N_c} \right).$$

La condition de premier ordre est vérifiée pour $\hat{\theta}_n^{\text{MV}} = \frac{n \bar{X}_n}{N_c} = \frac{\bar{X}_n}{\frac{N_c}{n}}$ et on peut vérifier que ℓ_n est strictement convexe. Donc $\hat{\theta}_n^{\text{MV}} = \frac{n \bar{X}_n}{N_c}$ est bien l'unique EMV de θ^* .

Consistance forte : notons $W_i = \binom{X_i}{\mathbf{1}\{X_i < c\}}$. Comme les $(W_i)_i$ sont i.i.d. $\in L^1$, la loi forte des

grands nombres s'applique et donne

$$\bar{W}_n := \frac{1}{n} \sum_{i=1}^n W_i = \left(\frac{\bar{X}_n}{\frac{1}{n} N_c} \right) \xrightarrow{p.s.} \mathbf{E}_\theta(W_1) = \begin{pmatrix} \mathbf{E}_\theta(X_1) \\ \mathbf{P}_\theta(X_1 < c) \end{pmatrix}.$$

Par le CMT appliqué à la fonction $g(a, b) = a/b$ qui est \mathcal{C}^0 , on obtient

$$\hat{\theta}_n^{\text{MV}} = g(\bar{W}_n) \xrightarrow{p.s.} g(\mathbf{E}_\theta(W_1)) = \frac{\theta^*(1 - e^{-c/\theta^*})}{1 - e^{-c/\theta^*}} = \theta^*.$$

7. Prouver que l'EMV $\hat{\theta}_n^{\text{MV}}$ est asymptotiquement normal.

On a

$$\begin{aligned} \sqrt{n}(\hat{\theta}_n^{\text{MV}} - \theta^*) &= \frac{1}{N_c/n} \times \frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - \theta^* \mathbb{1}(Y_i < c)) \\ &= \frac{1}{N_c/n} \sqrt{n} \bar{Z}_n, \end{aligned}$$

où les variables aléatoires $Z_i = X_i - \theta^* \mathbb{1}(Y_i < c)$ sont i.i.d., bornées donc L^∞ , et de moyenne nulle :

$$\mathbf{E}_{\theta^*}(Z_1) = \mu(\theta^*) - \theta^* \mathbf{P}_\theta(Y_1 < c) = \theta^*(1 - e^{-c/\theta^*}) - \theta^*(1 - e^{-c/\theta^*}) = 0.$$

Le Théorème central limite entraîne

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n Z_i \xrightarrow{\mathcal{L}} \mathcal{N}(0, \mathbf{Var}_\theta(Z_1)).$$

D'autre part, on a

$$N_c/n \xrightarrow{P} 1 - e^{-c/\theta^*}.$$

Par le lemme de Slutsky, on a

$$\sqrt{n}(\hat{\theta}_n^{\text{MV}} - \theta^*) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{\mathbf{Var}_\theta(Z_1)}{(1 - e^{-c/\theta^*})^2}\right).$$

Rq : on aurait également pu utiliser la delta-méthode et le TCL multivarié.

8. Quel est le comportement de la variance limite de l'EMV lorsque c tends vers $+\infty$? Expliquer intuitivement ce résultat.

Soit $v(c, \theta^*) := \frac{\mathbf{Var}_\theta(Z_1)}{(1-e^{-c/\theta^*})^2}$ la variance asymptotique de $\hat{\theta}_n^{\text{MV}}$. On a

$$\begin{aligned}\mathbf{Var}_\theta(Z_1) &= \mathbf{Var}_\theta((X_1 - \theta)\mathbb{1}(Y_1 < c) + c\mathbb{1}(Y_1 \geq c)) \\ &= \mathbf{E}_\theta[((Y_1 - \theta)\mathbb{1}(Y_1 < c) + c\mathbb{1}(Y_1 \geq c))^2] \quad \text{car } Z_1 \text{ centrée} \\ &= \mathbf{E}_\theta[(Y_1 - \theta)^2\mathbb{1}(Y_1 < c)] + c^2\mathbf{P}_\theta(Y_1 \geq c) \\ &= \mathbf{E}_\theta[Y_1^2\mathbb{1}(Y_1 < c)] - 2\theta\mathbf{E}_\theta[Y_1\mathbb{1}(Y_1 < c)] + \theta^2(1 - e^{-c/\theta}) + c^2e^{-c/\theta}.\end{aligned}$$

Une intégration par parties donne

$$\begin{aligned}\mathbf{E}_\theta[(Y_1 - \theta)^2\mathbb{1}(Y_1 < c)] &= \int_0^c y^2 \frac{e^{-y/\theta}}{\theta} dy \\ &= -c^2e^{-c/\theta} + 2 \int_0^c ye^{-y/\theta} dy \\ &= 2\theta^2 - c^2e^{-c/\theta} - 2\theta ce^{-c/\theta} - 2\theta^2e^{-c/\theta},\end{aligned}$$

et

$$\begin{aligned}\mathbf{E}_\theta[Y_1\mathbb{1}(Y_1 < c)] &= \int_0^c y \frac{e^{-y/\theta}}{\theta} dy \\ &= \theta - ce^{-c/\theta} - \theta e^{-c/\theta}.\end{aligned}$$

D'où

$$\begin{aligned}\mathbf{Var}_\theta(Z_1) &= \theta^2 - 2\theta ce^{-c/\theta} - \theta^2e^{-c/\theta} + 2\theta ce^{-c/\theta} \\ &= \theta^2(1 - e^{-c/\theta}).\end{aligned}$$

On en conclut

$$v(c, \theta) = \frac{\theta^2}{1 - e^{-c/\theta}}.$$

En particulier, on a $v(c, \theta) > \theta^2$. Logique : on est moins précis avec les observations censurées. Il est clair que $\lim_{c \rightarrow +\infty} v(c, \theta) = \theta^2$, c'est la variance limite de l'EMV dans un modèle exponentiel (d'après le TCL : $\sqrt{n}(\bar{Y}_n - \mathbf{E}_\theta(Y_1)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \mathbf{Var}_\theta(Y_1))$ avec $\mathbf{E}_\theta(Y_1) = \theta$ et $\mathbf{Var}_\theta(Y_1) = \frac{1}{(\frac{1}{\theta})^2} = \theta^2$). L'intuition est que $v(c, \theta^2)$ est décroissante de c et décroît vers θ^2 où la décroissance est la conséquence du fait que plus c est grand, plus les observations non-censurées sont nombreuses, et donc on arrive à mieux estimer θ .

9. (facultatif) Comparer les variances limites de $\hat{\theta}_n^{\text{MM}}$ et de $\hat{\theta}_n^{\text{MV}}$.

Revenons au calcul de la variance asymptotique $v^{\text{MM}}(\theta^*)$ de $\hat{\theta}_n^{\text{MM}}$. D'après la question 3, on a

$$v^{\text{MM}}(\theta^*) = \psi'(\mu(\theta^*))^2 \times \mathbf{Var}_{\theta^*}(X_1).$$

Or

$$\begin{aligned}
\text{Var}_\theta(X_1) &= \mathbf{E}_\theta(X_1^2) - \mu(\theta)^2 \\
&= \int_0^c \frac{y^2}{\theta} e^{-y/\theta} dy + c^2 e^{-c/\theta} - \theta^2(1 - e^{-c/\theta})^2 \\
&= 2\theta^2(1 - e^{-c/\theta}) - c^2 e^{-c/\theta} - 2\theta c e^{-c/\theta} + c^2 e^{-c/\theta} - \theta^2(1 - e^{-c/\theta})^2 \\
&= \theta^2(1 - e^{-2c/\theta}) - 2\theta c e^{-c/\theta}.
\end{aligned}$$

On sait que $\psi(\mu(\theta)) = \theta$. D'où,

$$\begin{aligned}
\psi'(\mu(\theta)) \times \mu'(\theta) = 1 &\implies \psi'(\mu(\theta)) = \frac{1}{\mu'(\theta)} \\
&= \frac{1}{1 - e^{-c/\theta} - (c/\theta)e^{-c/\theta}}.
\end{aligned}$$

Rq : on pouvait aussi utiliser directement la formule pour la dérivée de la bijection réciproque. Il en découle que

$$\begin{aligned}
v^{\text{MM}}(\theta) &= \frac{\theta^2(1 - e^{-2c/\theta}) - 2\theta c e^{-c/\theta}}{(1 - e^{-c/\theta} - (c/\theta)e^{-c/\theta})^2} \\
&= \theta^2 \times \frac{1 - e^{-2a} - 2ae^{-a}}{(1 - e^{-a} - ae^{-a})^2},
\end{aligned}$$

où $a = \frac{c}{\theta}$. D'où

$$\begin{aligned}
\frac{v^{\text{MM}}(\theta)}{v^{\text{MV}}(\theta)} &= \frac{1 - e^{-2a} - 2ae^{-a}}{(1 - e^{-a} - ae^{-a})^2} \times (1 - e^{-a}) \\
&= \frac{1 - e^{-2a} - 2ae^{-a} - e^{-a} + e^{-3a} + 2ae^{-2a}}{1 + e^{-2a} + a^2 e^{-2a} - 2e^{-a} - 2ae^{-a} + 2ae^{-2a}} \\
&= 1 + \frac{e^{-3a} + e^{-a} - 2e^{-2a} - a^2 e^{-2a}}{1 + e^{-2a} + a^2 e^{-2a} - 2e^{-a} - 2ae^{-a} + 2ae^{-2a}} \\
&= 1 + \frac{e^{-2a}(e^{a/2} - e^{-a/2} - a)(e^{a/2} - e^{-a/2} + a)}{1 + e^{-2a} + a^2 e^{-2a} - 2e^{-a} - 2ae^{-a} + 2ae^{-2a}}.
\end{aligned}$$

Comme

$$e^x > e^{-x} + 2x, \forall x > 0, \quad (33)$$

on en déduit

$$\frac{v^{\text{MM}}(\theta)}{v^{\text{MV}}(\theta)} > 1,$$

i.e. l'estimateur $\hat{\theta}_n^{\text{MM}}$ n'est pas asymptotiquement efficace.

Justification de (33) : pour $g(x) = e^x - e^{-x} - 2x$, on a

$$g(0) = 0 \text{ et } g'(x) = e^x + e^{-x} - 2 = (e^{x/2} - e^{-x/2})^2.$$

Donc $g'(x) > 0, \forall x > 0$. Il en découle que g est strictement croissante sur $]0, +\infty[$. Donc $g(x) > g(0) = 0, \forall x \in]0, +\infty[$.

Exercice 2. (facultatif) Soit $I(\theta)$ l'information de Fisher du modèle $\{P_\theta : \theta \in \Theta\}$, où Θ est un intervalle ouvert de \mathbb{R} . Soit $g : B \rightarrow \Theta$ une fonction continument différentiable et strictement croissante définie sur un intervalle ouvert B . Trouver une formule qui exprime l'information de Fisher $J(\beta)$ du modèle $\{Q_\beta = P_{g(\beta)} : \beta \in B\}$ en fonction de $I(\cdot)$ et de la dérivée de g .

On a $\Theta \subset \mathbb{R}$. On suppose que l'information de Fischer existe, et on note

$$\ell(x; \theta) := \log f(x; \theta) \quad \forall \theta \in \Theta,$$

où $f(x; \theta)$ est une densité de probabilité de P_θ par rapport à une mesure dominante. Pour tout $\beta \in B$, $g(\beta) \in \Theta$ et on note \mathbf{E}_θ l'espérance calculée pour $X \sim P_\theta$. En utilisant la formule de dérivation d'une composée de fonctions ("chain rule"), on obtient

$$\begin{aligned} J(\beta) &:= \mathbf{E}_{g(\beta)} \left[\left(\frac{\partial}{\partial t} \ell(X; g(t)) \Big|_{t=\beta} \right)^2 \right] \\ &= \mathbf{E}_{g(\beta)} \left[\left(g'(\beta) \times \frac{\partial}{\partial t} \ell(X; t) \Big|_{t=g(\beta)} \right)^2 \right] \\ &= [g'(\beta)]^2 \times \mathbf{E}_{g(\beta)} \left[\left(\frac{\partial}{\partial t} \ell(X; t) \Big|_{t=g(\beta)} \right)^2 \right] \\ &= [g'(\beta)]^2 \times I(g(\beta)). \end{aligned}$$

Comme g est strictement croissante, $J(\beta) = 0$ ssi $I(g(\beta)) = 0$.

TD9 : MÉTHODE BAYÉSIENNE : MOYENNE A POSTERIORI

Exercice 1. Soient Y_1, \dots, Y_n des variables aléatoires i.i.d. de loi de Poisson de moyenne $\theta^* > 0$, c'est-à-dire, $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \mathcal{P}(\theta^*)$.

1. On considère d'abord la mesure a priori π_0 égale à la mesure de Lebesgue sur $]0, +\infty[$. Trouver la densité de loi a posteriori, π_n , et vérifier que la loi a posteriori est une loi gamma.

Soit $S_n = X_1 + \dots + X_n = n\bar{X}_n$. Comme $f(k, \theta) = P_\theta(X_1 = k)$, la vraisemblance de ce modèle est donnée par

$$\begin{aligned} L_n(X_1, \dots, X_n; \theta) &= \prod_{i=1}^n f(X_i; \theta) = \prod_{i=1}^n \frac{\theta^{X_i}}{X_i!} e^{-\theta} \\ &= \frac{\theta^{S_n}}{X_1! \cdot \dots \cdot X_n!} e^{-n\theta} \\ &\propto \theta^{S_n} e^{-n\theta}. \end{aligned}$$

Par définition de π_0 , on a

$$d\pi_0(\theta) = \mathbb{1}_{]0, +\infty[}(\theta) d\lambda(\theta).$$

avec λ la mesure de Lebesgue. On en déduit que

$$\begin{aligned} \pi_n(\theta) &\propto L_n(X_1, \dots, X_n; \theta) \frac{d\pi_0}{d\lambda}(\theta) \\ &\propto \theta^{S_n} e^{-n\theta} \mathbb{1}_{]0, +\infty[}(\theta). \end{aligned}$$

On voit clairement que c'est une loi gamma de paramètres $(S_n + 1, 1/n)$.

2. On rappelle que l'espérance d'une v.a. ξ de loi gamma de paramètres (p, a) est égale à ap . Calculer l'estimateur bayésien

$$\hat{\theta}_n^B = \int_0^\infty \theta \pi_n(\theta) d\theta.$$

Il découle de la question précédente que

$$\hat{\theta}_n^B = \frac{S_n + 1}{n} = \bar{X}_n + \frac{1}{n}.$$

3. Vérifier que

$$\sqrt{n} (\hat{\theta}_n^B - \bar{X}_n) \xrightarrow[n \rightarrow \infty]{p.s.} 0.$$

C'est immédiat :

$$\begin{aligned}\sqrt{n}(\hat{\theta}_n^B - \bar{X}_n) &= \sqrt{n} \left(\bar{X}_n + \frac{1}{n} - \bar{X}_n \right) \\ &= \frac{1}{\sqrt{n}} \xrightarrow{p.s.} 0.\end{aligned}$$

4. Calculer l'information de Fisher dans ce modèle.

Le score est donné par

$$s(x; \theta) = \frac{\partial \log f(x; \theta)}{\partial \theta} = \frac{\partial}{\partial \theta} \left(x \log \theta - \theta \right) = \frac{x}{\theta} - 1.$$

Donc

$$I(\theta) = \text{Var}_{\theta} \left[\frac{X_1}{\theta} - 1 \right] = \frac{\text{Var}_{\theta}(X_1)}{\theta^2} = \frac{1}{\theta},$$

où on a utilisé que la variance d'une variable de loi de Poisson de paramètre θ est égale à θ .

5. On admet que le modèle de Poisson est régulier, montrer que $\hat{\theta}_n^B$ est asymptotiquement normal et déterminer la variance limite.

Comme (a) le modèle est régulier, (b) l'EMV est $\hat{\theta}_n^{\text{MV}} = \bar{X}_n$ et (c) ce dernier est fortement consistant d'après la loi forte des grands nombres, on peut utiliser le théorème vu en cours qui dit que

$$\sqrt{n}(\bar{X}_n - \theta^*) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, I(\theta^*)^{-1}) = N(0, \theta^*).$$

(On pouvait aussi, bien-sûr, utiliser directement le TCL).

6. Vérifier que $\{\text{Gamma}(p, a) : p > 0, a > 0\}$ est une famille conjuguée pour le modèle de Poisson.

Soit π_0 la densité de la loi a priori, choisie comme une loi gamma de paramètres (p, a) :

$$\pi_0(\theta) \propto \theta^{p-1} e^{-\theta/a} \mathbb{1}_{]0, \infty[}(\theta).$$

En répétant les arguments de la question 1, il vient

$$\begin{aligned}\pi_n(\theta) &\propto \left(\theta^{S_n} e^{-n\theta} \right) \left(\theta^{p-1} e^{-\theta/a} \right) \mathbb{1}_{]0, \infty[}(\theta) \\ &= \theta^{S_n+p-1} e^{-(n+1/a)\theta} \mathbb{1}_{]0, \infty[}(\theta).\end{aligned}$$

Par conséquent, π_n est la densité d'une loi gamma de paramètres (\hat{p}_n, \hat{a}_n) donnés par

$$\hat{p}_n = S_n + p; \quad \hat{a}_n = \frac{a}{an + 1}.$$

Il en découle que la famille des lois gamma est conjuguée pour le modèle de Poisson. On peut remarquer au passage que dans le cas limite $p \rightarrow 1$ et $a \rightarrow +\infty$, on obtient le résultat de la question 1.

Exercice 2. Nous avons observé des variables aléatoires iid de loi uniforme sur l'intervalle $[\theta^*, \theta^* + 1]$, mais la valeur de θ^* ne nous a pas été dévoilée. Nous décidons d'utiliser l'estimateur bayésien utilisant la mesure de Lebesgue comme mesure a priori.

1. Trouver la forme explicite de l'estimateur bayésien $\hat{\theta}_n^B$.

On suppose que la mesure a priori est la mesure de Lebesgue sur \mathbb{R} . En notant π_0 la densité a priori, on a $\pi_0(\theta) = 1, \forall \theta \in \mathbb{R}$. Nous avons déjà étudié le modèle iid uniforme (modèle paramétrique, identifiable et dominé par la mesure de Lebesgue). La vraisemblance s'écrit

$$\begin{aligned} L_n(X_1, \dots, X_n; \theta) &\stackrel{iid}{=} \prod_{i=1}^n \frac{d\mathcal{U}([\theta, \theta+1])}{d\lambda}(X_i) \\ &= \prod_{i=1}^n \mathbb{1}_{[\theta, \theta+1]}(X_i) \\ &= \mathbb{1}(\theta \leq X_{(1)} \leq X_{(n)} \leq \theta + 1) \\ &= \mathbb{1}_{[X_{(n)} - 1; X_{(1)}]}(\theta), \end{aligned}$$

où $X_{(n)} = \max_i X_i$, $X_{(1)} = \min_i X_i$ et la troisième égalité est une égalité p.s... On en déduit la forme de la densité a posteriori associée à la densité a priori $\pi_0(\theta) = 1$,

$$\begin{aligned} \pi_n(\theta) &\propto L_n(X_1, \dots, X_n; \theta) \pi_0(\theta) \\ &\propto \mathbb{1}_{[X_{(n)} - 1; X_{(1)}]}(\theta). \end{aligned}$$

D'après un théorème du cours, cela implique que

$$\begin{aligned} \hat{\theta}_n^B &= \frac{\int_{\mathbb{R}} \theta L_n(X_1, \dots, X_n; \theta) \pi_0(\theta) \lambda(d\theta)}{\int_{\mathbb{R}} L_n(X_1, \dots, X_n; \theta) \pi_0(\theta) \lambda(d\theta)} \\ &= \frac{\int_{X_{(n)} - 1}^{X_{(1)}} \theta d\theta}{\int_{X_{(n)} - 1}^{X_{(1)}} 1 d\theta} \\ &= \frac{X_{(1)} + X_{(n)} - 1}{2}. \end{aligned}$$

Remarque : la dernière égalité s'obtient par calcul intégral ou bien en remarquant que $\hat{\theta}_n^B = E[Z|X_1, \dots, X_n]$ avec $Z|X_1, \dots, X_n \sim \text{Unif}([X_{(n)} - 1, X_{(1)}])$.

2. L'estimateur $\hat{\theta}_n^B$ est-il sans biais ?

Méthode 1 (calcul) : On veut calculer, pour $\theta \in \mathbb{R}$:

$$E_\theta[\hat{\theta}_n^B - \theta] = \frac{1}{2} \left\{ E_\theta[X_{(1)}] + E_\theta[X_{(n)}] - 1 \right\} - \theta.$$

Il suffit donc de déterminer les lois marginales du vecteur $(X_{(1)}, X_{(n)})$, puis de calculer les intégrales ci-dessus.

- Densité et espérance de $X_{(1)}$. Par définition, $X_{(1)}$ prend valeurs p.s. dans $[\theta, \theta + 1]$. Soit $y \in]\theta, \theta + 1[$, on a

$$\begin{aligned} \mathbf{P}_\theta(X_{(1)} \leq y) &= \mathbf{P}_\theta(\min_{1 \leq i \leq n} X_i \leq y) \\ &= 1 - \mathbf{P}_\theta(\min_{1 \leq i \leq n} X_i > y) \\ &= 1 - \mathbf{P}_\theta(\bigcap_{1 \leq i \leq n} \{X_i > y\}) \\ &\stackrel{iid}{=} 1 - [\mathbf{P}_\theta(X_1 > y)]^n \\ &= 1 - (\theta + 1 - y)^n \quad \text{car } X_1 \sim \text{Unif}(]\theta, \theta + 1[) \end{aligned}$$

On en déduit la densité de $X_{(1)}$ par rapport à la mesure de Lebesgue :

$$f_{X_{(1)}}(y; \theta) = n(\theta + 1 - y)^{n-1} \mathbf{1}_{]\theta, \theta+1[}(y).$$

Ainsi, par le théorème de transfert,

$$\begin{aligned} E[X_{(1)}] &= \int_{\mathbb{R}} ny(\theta + 1 - y)^{n-1} \mathbf{1}_{]\theta, \theta+1[}(y) dy \\ &= n \int_{\theta}^{\theta+1} y(\theta + 1 - y)^{n-1} dy \\ &= n \int_0^1 (\theta + 1 - t)t^{n-1} dt \quad (\text{avec le changement de variable } t = \theta + 1 - y) \\ &= n(\theta + 1) \int_0^1 t^{n-1} dt - n \int_0^1 t^n dt \\ &= \theta + 1 - \frac{n}{n+1} \\ &= \theta + \frac{1}{n+1}. \end{aligned}$$

La dernière égalité utilise que pour $k \in \mathbb{N}^*$, $\int_0^1 t^k dt = 1/(k+1)$.

- Densité et espérance de $X_{(n)}$. Par définition, $X_{(n)}$ prend valeurs p.s. dans $[\theta, \theta + 1]$. Soit $x \in]\theta, \theta + 1[$,

$$\mathbf{P}_\theta(X_{(n)} \leq x) = \mathbf{P}_\theta(\bigcap\{X_i \leq x\}) \stackrel{iid}{=} [\mathbf{P}_\theta(X_1 \leq x)]^n = (x - \theta)^n.$$

On en déduit la densité de $X_{(n)}$ par rapport à la mesure de Lebesgue :

$$f_{X_{(n)}}(x; \theta) = n(x - \theta)^{n-1} \mathbf{1}_{]\theta, \theta+1[}(x).$$

Ainsi, on a d'après la théorème de transfert

$$\begin{aligned}
E_\theta[X_{(n)}] &= \int_{\mathbb{R}} xn(x - \theta)^{n-1} 1_{[\theta, \theta+1]}(x) dx \\
&= n \int_{\theta}^{\theta+1} x(x - \theta)^{n-1} dx \\
&= n \int_0^1 (t + \theta)t^{n-1} dy \quad \text{avec le changement de variable } t = x - \theta \\
&= n \int_0^1 y^n dy + n\theta \int_0^1 y^{n-1} dy \\
&= \frac{n}{n+1} + \frac{n\theta}{n} \\
&= \theta + \frac{n}{n+1}.
\end{aligned}$$

On a donc, pour tout $\theta \in \mathbb{R}$,

$$E_\theta[\hat{\theta}_n^B - \theta] = \frac{1}{2} \left\{ \underbrace{\theta + \frac{1}{n+1}}_{=E_\theta[X_{(1)}]} + \underbrace{\theta + \frac{n}{n+1}}_{E_\theta[X_{(n)}]} - 1 \right\} - \theta = \theta - \theta = 0.$$

On conclut que $\hat{\theta}_n^B$ est un estimateur sans biais de θ .

Méthode 2 : On a

$$\hat{\theta}_n^B - \theta = \frac{X_{(1)} - \theta + X_{(n)} - 1 - \theta}{2}. \quad (34)$$

Soit $\theta \in \Theta = \mathbb{R}$. On pose $U_i = X_i - \theta$ et $V_i = \theta + 1 - X_i$. On peut vérifier que les variables aléatoires $\{U_i\}$ sont iid de loi uniforme sur $[0, 1]$ et il en est de même pour les V_i . On peut réécrire

$$\begin{cases} X_i &= U_i + \theta, \\ X_i - 1 &= -V_i + \theta. \end{cases}$$

Cela montre que pour obtenir X_i (resp. $X_i - 1$), on applique seulement une translation $+\theta$ à U_i (resp. $-V_i$). De plus, comme $-V_i \leq 0$, on a

$$\begin{aligned}
X_{(1)} &= U_{(1)} + \theta, \\
X_{(n)} - 1 &= \theta - V_{(1)}.
\end{aligned}$$

L'équation (34) devient

$$\hat{\theta}_n^B - \theta = \frac{U_{(1)} - V_{(1)}}{2}.$$

Par linéarité de l'espérance et comme $U_{(1)}$ et $V_{(1)}$ sont deux variables aléatoires de même loi, cela implique que $E_\theta[\hat{\theta}_n^B - \theta] = 0$. L'estimateur bayésien est donc sans biais.

3. Peut-on affirmer que la famille $\{\mathcal{N}(\mu, 1) : \mu \in \mathbb{R}\}$ est conjuguée pour le modèle $\{\mathcal{U}([\theta, \theta+1]) : \theta \in \mathbb{R}\}$?

Il suffit de faire une simple vérification pour se convaincre que cette famille n'est pas conjuguée. C'est le support de la loi a posteriori qui va poser problème ici (pas égal à \mathbb{R} , le support d'une Gaussienne standard).

4. (facultatif) Prouver que la suite $\{n(\hat{\theta}_n^B - \theta^*)\}_{n \in \mathbb{N}}$ converge en loi.

Intuition : on a déjà vu que pour $U_i \stackrel{iid}{\sim} \text{Unif}([0, 1])$, on a $nU_{(1)} \xrightarrow{\mathcal{L}} \text{Exp}(1)$ et $n(1 - U_{(n)}) \xrightarrow{\mathcal{L}} \text{Exp}(1)$. On peut se ramener à ce cas ici en remarquant

$$\begin{aligned} n(\hat{\theta}_n^B - \theta^*) &= n \times \frac{X_{(1)} - \theta^* + X_{(n)} - \theta^* - 1}{2} \\ &= \frac{nU_{(1)} - n(1 - U_{(n)})}{2}. \end{aligned}$$

Attention : la convergence des lois marginales de $(nU_{(1)}, n(1 - U_{(n)}))$ ne suffit en général pas à déterminer la convergence en loi d'une fonctionnelle du **vecteur** $(nU_{(1)}, n(1 - U_{(n)}))$. On détermine d'abord la distribution limite de la suite des vecteurs deux-dimensionnels $(Z_n, Z'_n) = (nU_{(1)}, n(1 - U_{(n)}))$, puis on utilise le premier théorème de continuité.

Pour tout $x, y \in [0, \infty]$, on a

$$\begin{aligned} \mathbb{P}\left(nU_{(1)} > x, n(1 - U_{(n)}) > y\right) &= \mathbb{P}\left(\min_i U_i > \frac{x}{n}, \max_i U_i < 1 - \frac{y}{n}\right) \\ &= \mathbb{P}\left(\bigcap_{i=1}^n U_i \in \left] \frac{x}{n}; 1 - \frac{y}{n} \right[\right) \\ &= \left(1 - \frac{x+y}{n}\right)^n \mathbb{1}(x+y < n) \\ &\xrightarrow[n \rightarrow \infty]{} e^{-(x+y)}. \end{aligned}$$

On en déduit que la suite des vecteurs deux-dimensionnels (Z_n, Z'_n) converge en loi vers (Z, Z') où Z et Z' sont deux variables aléatoires indépendantes de loi exponentielle de paramètre 1. En outre,

$$n(\hat{\theta}_n^B - \theta^*) = \frac{nU_{(1)} - n(1 - U_{(n)})}{2} = \frac{Z_n - Z'_n}{2}.$$

En utilisant le premier théorème de continuité, on conclut que $n(\hat{\theta}_n^B - \theta^*)$ converge en loi vers $(Z - Z')/2$.