PC 8 : Estimation

Exercice 1 (Méthode de Monte Carlo). Soit $g : [0, 1] \rightarrow [0, 1]$ une fonction mesurable et bornée. On souhaite calculer $m = \int_0^1 g(x) dx$. On pose $\sigma^2 := \int_0^1 g^2(x) dx - m^2$. Soient X et Y des v.a.r. i.i.d. de loi uniforme sur $[0, 1]$. On pose

$$U = \mathbf{1}_{Y \leq g(X)}, \quad V = g(X), \quad \text{et} \quad W = \frac{g(X) + g(1 - X)}{2}.$$

1. Calculer l'espérance et la variance de U, V, W , et comparer les variances ;
2. Proposer trois méthodes de type Monte-Carlo pour calculer m ;
3. On suppose dans la suite que g est monotone. Vérifier que $\mathbb{E}(g(X)g(1 - X)) \leq m^2$. Indication : on pourra d'abord montrer que $(g(x) - g(y))(g(1 - x) - g(1 - y)) \leq 0$ pour tous x, y .
4. Soit $(X_i)_{i \geq 1}$ une suite de v.a.r. i.i.d. de loi uniforme sur $[0, 1]$, et les estimateurs de m

$$A_n = \frac{1}{2n} \sum_{i=1}^{2n} g(X_i) \quad \text{et} \quad B_n = \frac{1}{2n} \sum_{i=1}^n (g(X_i) + g(1 - X_i)).$$

Lequel possède la plus petite erreur quadratique moyenne ?

5. Dans le cas où $g(x) = x^2$, déterminer le nombre n de simulations nécessaires garantissant une précision relative de 1% sur le calcul de m avec A_n et B_n .

Solution. 1. Comme $U^2 = U$ (Bernoulli) on a $\text{Var}(U) = \mathbb{E}(U)(1 - \mathbb{E}(U))$. à présent,

$$\mathbb{E}(U) = \mathbb{P}(Y \leq g(X)) = \iint_{[0,1]^2} \mathbf{1}_{y \leq g(x)} dx dy = \int_0^1 g(x) dx = m \quad \text{et} \quad \text{Var}(U) = m(1 - m).$$

Par le théorème du transfert

$$\mathbb{E}(V) = \int_0^1 g(x) dx = m \quad \text{et} \quad \mathbb{E}(V^2) = \int_0^1 g^2(x) dx \quad \text{donc} \quad \text{Var}(V) = \sigma^2.$$

Comme X et $1 - X$ ont même loi on a par linéarité de l'espérance

$$\mathbb{E}(W) = m \quad \text{et} \quad \mathbb{E}(W^2) = \frac{(\sigma^2 + m^2) + \int_0^1 g(x)g(1 - x) dx}{2}$$

et donc

$$\text{Var}(W) = \frac{(\sigma^2 - m^2) + \int_0^1 g(x)g(1 - x) dx}{2}$$

Comme g prend ses valeurs dans $[0, 1]$, il vient que V a une plus petite variance que U car

$$\text{Var}(U) - \text{Var}(V) = \int_0^1 g(x)(1 - g(x)) dx \geq 0.$$

D'autre part, la comparaison des variances de V et W donne :

$$\text{Var}(V) - \text{Var}(W) = \frac{\int_0^1 g(x)(g(x) - g(1 - x)) dx}{2}.$$

2. Par la LGN forte on a p.s.

$$m = \lim_{n \rightarrow \infty} \frac{\#\{1 \leq i \leq n : Y_i \leq g(X_i)\}}{n} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n g(X_i) = \lim_{n \rightarrow \infty} \frac{1}{2n} \sum_{i=1}^n (g(X_i) + g(1-X_i)).$$

3. Comme g est monotone, on a $(g(x) - g(y))(g(1-x) - g(1-y)) \leq 0$ pour tous x, y et donc $\mathbb{E}(g(X)g(1-X)) + \mathbb{E}(g(Y)g(1-Y)) - 2\mathbb{E}(g(X))\mathbb{E}(g(Y)) \leq 0$, c'est-à-dire $\mathbb{E}(g(X)g(1-X)) \leq m^2$. Il en découle que $\text{Var}(W) \leq \frac{\sigma^2}{2} = \frac{\text{Var}(V)}{2}$.

4. $EQM(\hat{m}) := \mathbb{E}((\hat{m} - m)^2) = \text{Var}(\hat{m}) + (\mathbb{E}(\hat{m}) - m)^2$. ($EQM = \text{Variance} + \text{Carré Du Biais}$). Les estimateurs A_n et B_n sont sans biais. On a $EQM(A_n) = \frac{\sigma^2}{2n}$ et $EQM(B_n) = \frac{\text{Var}(W)}{n} \leq \frac{\sigma^2}{2n}$. Mais A_n nécessite un échantillon deux fois plus gros que B_n .

5. $\frac{EQM(A_n)}{m^2} = \frac{1}{100}$ donne $n = \frac{50\sigma^2}{m^2}$. Pour $g(x) = x^2$ il vient $m = \frac{1}{3}$ et $\sigma^2 = \frac{4}{45}$ et donc $n \approx 40$ avec A_n , et en fait plus de 16 fois moins avec B_n car il s'avère que $\text{Var}(W) = \frac{1}{180}$.

Exercice 2 (Maximum de vraisemblance). On observe un échantillon $(X_1, \dots, X_n)$.

1. Donnez l'estimateur du maximum de vraisemblance dans le cas :

i) $(X_i)_{1 \leq i \leq n} \stackrel{i.i.d.}{\sim} \mathcal{B}(p), p \in]0, 1[$.

ii) $(X_i)_{1 \leq i \leq n} \stackrel{i.i.d.}{\sim} \mathcal{P}(\lambda), \lambda > 0$.

iii) $(X_i)_{1 \leq i \leq n} \stackrel{i.i.d.}{\sim} \mathcal{E}(\theta), \theta > 0$.

2. Donnez une convergence en loi non dégénérée de l'estimateur dans les 3 modèles précédents.

Exercice 3 (Comparaison d'estimateurs). Soit $(X_n)_{n \geq 1}$ des v.a.r. i.i.d. de loi uniforme sur $[0, \theta]$.

1. étudier la convergence, le biais, la variance, le risque quadratique moyen, et la fluctuation asymptotique de l'estimateur $m_n = 2(X_1 + \dots + X_n)/n$ de θ .

2. Montrer que l'estimateur de maximum de vraisemblance de θ est $M_n = \max(X_1, \dots, X_n)$.

3. Montrer que l'estimateur M_n a pour densité $(n/\theta)(x/\theta)^{n-1}\mathbf{1}_{[0, \theta]}(x)$ puis calculer $\mathbb{E}(M_n)$ et $\text{Var}(M_n)$. Comparer avec l'estimateur moyenne empirique m_n .

4. Montrer que $M_n \xrightarrow{\mathbb{P}} \theta$, puis que la convergence est en fait presque sûre.

5. Montrer que $W_n = n(M_n/\theta - 1)$ converge en loi quand $n \rightarrow \infty$, et déterminer la limite.

Solution. Pour un estimateur $\hat{\theta}_n$ de θ , l'écart quadratique moyen possède toujours une décomposition (carré-du-)biais-variance : $\mathbb{E}((\hat{\theta}_n - \theta)^2) = (\mathbb{E}(\hat{\theta}_n) - \theta)^2 + \text{Var}(\hat{\theta}_n)$. Cela découle du théorème de Pythagore dans L^2 : $\mathbb{E}((Z - \theta)^2) = \text{Var}(Z) + (\mathbb{E}(Z) - \theta)^2$. Note : $\text{Var}(Z) = \arg \min_{c \in \mathbb{R}} \mathbb{E}((Z - c)^2)$.

1. La LGN entraîne que $m_n \rightarrow \theta$ p.s. L'estimateur n'est pas biaisé : $\mathbb{E}(m_n) = \theta$. La variance est égale à l'écart quadratique moyen, et vaut $\frac{4}{n} \text{Var}(X_1) = \frac{\theta^2}{3n} = \mathcal{O}(\frac{1}{n})$. La fluctuation asymptotique est gaussienne, de vitesse $\sqrt{n}$, car par le TCL, $\sqrt{n}(m_n - \theta) \xrightarrow{\text{loi}} \mathcal{N}(0, \frac{\theta^2}{3})$. L'intervalle de confiance asymptotique s'obtient avec le résultat de fluctuation. Nul besoin du lemme de Slutsky car l'inversion en θ est facile ici : si $Z \sim \mathcal{N}(0, 1)$ et $J \subset \mathbb{R}$,

$$\mathbb{P}(Z \in J) \approx \mathbb{P}\left(\frac{\sqrt{3n}}{\theta}(m_n - \theta) \in J\right) = \mathbb{P}(\theta \in m_n/(1 + (3n)^{-1/2}J))$$

Tout intervalle J tel que $\mathbb{P}(Z \in J) = 1 - \alpha$ fournit l'intervalle de confiance $I = m_n/(1 + (3n)^{-1/2}J)$ de niveau $1 - \alpha$ pour θ . On peut chercher à choisir J de sorte que I soit petit ;

2. $F_{M_n}(x) = (x/\theta)^n \mathbf{1}_{[0,\theta]} + \mathbf{1}_{]\theta,\infty[}$ et $f_{M_n}(x) = F'_{M_n}(x) = (n/\theta)(x/\theta)^{n-1} \mathbf{1}_{[0,\theta]}(x)$. Par conséquent, $\mathbb{E}(M_n) = \frac{n}{n+1}\theta$ et $\text{Var}(M_n) = \frac{n}{(n+1)^2(n+2)}\theta^2 = \mathcal{O}(\frac{1}{n^2})$. L'estimateur M_n est biaisé (et asymptotiquement sans biais), mais il est plus rapide que m_n . L'écart quadratique moyen de M_n vaut $(\mathbb{E}(M_n) - \theta)^2 + \text{Var}(M_n) = \frac{\theta^2}{(n+1)^2} + \frac{n\theta^2}{(n+1)^2(n+2)} = \frac{2}{(n+1)(n+2)}\theta^2 = \mathcal{O}(\frac{1}{n^2})$;
3. Pour tout $\varepsilon > 0$ assez petit, $\mathbb{P}(|M_n - \theta| > \varepsilon) = \mathbb{P}(M_n < \theta - \varepsilon) = ((\theta - \varepsilon)/\theta)^n$. Cela donne la convergence en probabilité de M_n vers θ , et même la convergence p.s. en utilisant le lemme de Borel-Cantelli pour ε fixé (il faut ensuite poser $\varepsilon = 1/k$ et faire $\cap_k$). Alternativement, comme p.s. $(M_n)_{n \geq 1}$ est positive croissante et majorée par θ , elle converge p.s. et dans L^1 vers une v.a.r. $\leq \theta$ de moyenne $\lim_{n \rightarrow \infty} \mathbb{E}(M_n) = \theta$, qui est forcément égale p.s. à θ ;
4. $F_{W_n}(x) = \mathbb{P}(M_n \leq \theta(1 + x/n))$. Ainsi, $F_{W_n}(x) \rightarrow 1$ si $x > 0$, tandis que si $x \leq 0$ alors, pour $n \gg 1$, $F_{W_n}(x) = (1 + x/n)^n \rightarrow e^x$. Ainsi, $W_n \rightarrow -W$ en loi quand $n \rightarrow \infty$, où $W \sim \text{Exp}(1)$. L'intervalle de confiance asymptotique s'obtient avec la fluctuation : Pour tout $J \subset \mathbb{R}$, on a

$$\mathbb{P}(-W \in J) \approx \mathbb{P}(W_n \in J) = \mathbb{P}(\theta \in M_n/(1 + J/n)).$$

N'importe quel intervalle J tel que $\mathbb{P}(-W \in J) = 1 - \alpha$ fournit l'intervalle de confiance asymptotique $I = M_n/(1 + J/n)$ pour θ . On peut choisir J tel que I soit petit.

5. Pour tout réel $\theta \geq 0$, on a $L(\theta; X_1, \dots, X_n) = \theta^{-n} \prod_{k=1}^n \mathbf{1}_{[0,\theta]}(X_k)$, qui vaut 0 si $\theta < M_n$ et qui décroît comme θ^{-n} si $\theta \geq M_n$. Donc $M_n = \arg \max_{\theta \geq 0} L(\theta; X_1, \dots, X_n)$.

Exercice 4 (Méthode de capture-recapture). On souhaite estimer le nombre N de poissons vivant dans un bassin. Pour cela, on en pêche k que l'on marque. Ensuite, on pêche et on relâche successivement n poissons et l'on compte le nombre X de poissons marqués parmi les n . Donner l'estimateur du maximum de vraisemblance de N en fonction de X, n, k .

Exercice 5 (Régression linéaire). Soit $X \in M_{n,p}(\mathbb{R})$, $\beta \in \mathbb{R}^p$ avec $p < n$. On définit

$$Y = X\beta + \varepsilon,$$

où $\varepsilon \sim \mathcal{N}_n(0, \sigma^2 I_n)$. La matrice X est connue et on observe Y . L'objectif est d'estimer β . On supposera que X est de rang maximal, ce qui implique en particulier que $X'X$ est inversible. La vraisemblance d'un vecteur non indépendant est sa densité (qui généralise le cas où les composantes sont indépendantes).

1. Montrez $\hat{\beta}$, l'estimateur du maximum de vraisemblance, est caractérisé par

$$\hat{\beta} \in \underset{\beta \in \mathbb{R}^p}{\text{argmin}} \|Y - X\beta\|_2^2.$$

2. En déduire que

$$\hat{\beta} = (X'X)^{-1} X'Y.$$

3. Donnez la loi de $\hat{\beta}$.

4. On admet le Théorème de Cochran : Si $(V_i)_{1 \leq i \leq n}$ est une décomposition orthogonale en somme directe de $\mathbb{R}^n$ et que $(\Pi_{V_i})_{1 \leq i \leq n}$ sont les matrices de projection associées, alors, si $Z \sim \mathcal{N}_n(0, \sigma^2 I_n)$,

- $\Pi_{V_1} Z, \dots, \Pi_{V_k} Z$ sont des vecteurs aléatoires indépendants de loi $\mathcal{N}_n(0, \sigma^2 \Pi_{V_i})$.
- Pour tout $1 \leq i \leq n$, $\|\Pi_{V_i} Z\|_2^2$ suit la loi $\sigma^2 \chi_{p_i}^2$ où p_i est la dimension de V_i .

Donnez la loi de :

$$\hat{\sigma}^2 := \frac{\|Y - X\hat{\beta}\|_2^2}{n - p},$$

et en déduire qu'il s'agit d'un estimateur sans biais et convergent (lorsque $n \rightarrow +\infty$) de σ^2 .

Solution. 1. La log-vraisemblance associée à $(Y_i)_{1 \leq i \leq n}$ s'écrit

$$-\log(\sigma) - p \log(2\pi) - \frac{\|Y - X\beta\|_2^2}{2\sigma}.$$

Maximiser la log-vraisemblance revient à minimiser $\|Y - X\beta\|_2^2$.

2. On remarque que minimiser $\|Y - X\beta\|_2^2$ revient à projeter linéairement Y sur l'espace vectoriel engendré par les colonnes de X . $\hat{\beta}$ vérifie

$$\langle Y - X\hat{\beta}, X\beta \rangle_{\mathbb{R}^n} = 0, \quad \forall \beta \in \mathbb{R}^p \iff \langle X'(Y - X\hat{\beta}), \beta \rangle_{\mathbb{R}^p} = 0, \quad \forall \beta \in \mathbb{R}^p.$$

On en déduit que $\hat{\beta}$ vérifie :

$$X'Y = X'X\hat{\beta},$$

et comme $X'X$ est supposé inversible,

$$\hat{\beta} = (X'X)^{-1}X'Y.$$

3. On a

$$\hat{\beta} = (X'X)^{-1}X'(X\beta + \varepsilon) = \beta + (X'X)^{-1}X'\varepsilon.$$

Étant donné que $\varepsilon \sim \mathcal{N}_n(0, \sigma^2 I_n)$, on en déduit que

$$\hat{\beta} \sim \mathcal{N}_p(\beta, \sigma^2(X'X)^{-1}).$$

4. Par construction, $X\hat{\beta}$ est la projection orthogonale de Y sur l'espace vectoriel engendré par les colonnes de X . On a la décomposition orthogonale :

$$\varepsilon = Y - X\beta = (X\hat{\beta} - X\beta) + (Y - X\hat{\beta}),$$

où $X\hat{\beta} - X\beta$ est la projection orthogonale de $Y - X\beta$ sur l'espace vectoriel engendré par X et $Y - X\hat{\beta}$ sur son supplémentaire orthogonal de dimension $n - p$. On en déduit, par le théorème de Cochran, que

$$\frac{\|Y - X\hat{\beta}\|_2^2}{n - p} \sim \sigma^2 \frac{\mathcal{X}_{n-p}^2}{n - p}.$$

Puisque $\mathbb{E}(\mathcal{X}_{n-p}^2) = n - p$ et $\text{Var}(\mathcal{X}_{n-p}^2) = 2(n - p)$, on en déduit :

$$\mathbb{E}(\hat{\sigma}^2) = \sigma^2$$

et

$$\text{Var}(\hat{\sigma}^2) = \frac{2\sigma^2}{n - p}.$$

L'erreur quadratique moyenne tend vers 0 ce qui implique la convergence L^2 de l'estimateur et donc en probabilité.

Exercice 6 (Données censurées). Soit $(X_i, C_i)_{1 \leq i \leq n}$ une i.i.d. de variables aléatoires indépendantes. On suppose de plus que X_i est indépendante de C_i , pour tout i . Enfin, on suppose que X_1 admet la densité $x \mapsto f(x; \theta)$ par rapport à une mesure μ (discrète ou Lebesgue) pour un $\theta \in \Theta$ inconnu. On suppose que C_1 admet également une densité par rapport à μ , qu'on notera g et qui est inconnue.

On s'intéresse à l'estimation de $\theta \in \Theta$, mais on n'observe pas les X_i directement. On observe seulement :

$$(Y_i, Z_i)_{1 \leq i \leq n} = (\min(X_i, C_i), \mathbf{1}_{\{X_i \leq C_i\}})_{1 \leq i \leq n}.$$

Ici, C_i est appelée une variable de censure, c'est à dire que si elle est plus petite que X_i , on n'observera pas X_i (elle sera censurée), mais on aura l'information partielle que $X_i > C_i$ car on sait si Y_i vaut X_i ou C_i . L'objectif est d'estimer $\theta \in \Theta$ grâce aux seules observations de $(Y_i, Z_i)_{1 \leq i \leq n}$.

1. Soit $A \in \mathcal{B}(\mathbb{R})$ et $B \in \mathcal{P}(\{0, 1\})$. Montrez que

$$\mathbb{P}(Y_1 \in A, Z_1 \in B) = \int_{A \times B} [\bar{F}(y; \theta)g(y)]^{1-z} [\hat{G}(y)f(y; \theta)]^z d\mu(y)d(\delta_0 + \delta_1)(z),$$

où $\bar{F}$ est la fonction de survie associée à X_1 et $\hat{G}(y) = \mathbb{P}(C < y)$.

2. En déduire que maximiser la vraisemblance revient à résoudre le problème :

$$\hat{\theta} \in \operatorname{argmax}_{\theta \in \Theta} \prod_{i=1}^n [\bar{F}(y_i; \theta)]^{1-z_i} [f(y_i; \theta)]^{z_i}.$$

Remarque : La loi de C_1 n'intervient pas dans l'expression ci-dessus. En fait, ce résultat est aussi valide sans l'hypothèse que la loi de C_1 soit dominée par la même mesure μ , en revanche, l'hypothèse d'indépendance entre les X_i et les C_i est fondamentale.

3. On suppose que $(X_i)_{1 \leq i \leq n} \stackrel{i.i.d.}{\sim} \mathcal{E}(\theta)$, $\theta > 0$ est la date de mariage après 18 ans d'individus pris au hasard et C_i leur date de décès et indépendantes des X_i . On ne fait pas d'hypothèse supplémentaire sur la loi de $(C_i)_{1 \leq i \leq n}$. On observe les réalisations $(y_i, z_i)_{1 \leq i \leq n}$ des variables aléatoires $(\min(X_i, C_i), \mathbf{1}_{\{X_i \leq C_i\}})_{1 \leq i \leq n}$.

Donnez l'estimateur du maximum de vraisemblance $\hat{\theta}$.

4. Montrez que cet estimateur est fortement convergent. Indication : on pourra montrer que tout $c \geq 0$, $\mathbb{E}(\min(X, c)) = \int_0^c (1 - F(x; \theta)) dx$.
5. On introduit l'estimateur naïf qui consiste à supprimer les données censurées (on ne conserve que les données où $X_i \leq C_i$), c'est à dire :

$$\tilde{\theta} = \frac{\sum_{i=1}^n z_i}{\sum_{i=1}^n x_i \mathbf{1}_{\{x_i \leq c_i\}}}.$$

Montrez que cet estimateur n'est pas convergent.

Solution. 1. Soit $A \in \mathcal{B}(\mathbb{R})$. D'une part :

$$\begin{aligned} \mathbb{P}(Y_1 \in A, Z_1 = 1) &= \int \int \mathbf{1}_A(\min(x, c)) \mathbf{1}_{\{x \leq c\}} f(x; \theta) g(c) d\mu(x) d\mu(c) \\ &= \int \mathbf{1}_A(x) f(x; \theta) \int \mathbf{1}_{\{x \leq c\}} g(c) d\mu(c) d\mu(x) \\ &= \int_A f(x; \theta) \hat{G}(x) d\mu(x). \end{aligned}$$

D'autre part :

$$\begin{aligned}\mathbb{P}(Y_1 \in A, Z_1 = 0) &= \int \int \mathbf{1}_A(\min(x, c)) \mathbf{1}_{\{x > c\}} f(x; \theta) g(c) d\mu(x) d\mu(c) \\ &= \int \mathbf{1}_A(c) g(c) \int \mathbf{1}_{\{x > c\}} f(x; \theta) d\mu(x) d\mu(c) \\ &= \int_A g(c) \bar{F}(x; \theta) d\mu(c).\end{aligned}$$

Soit $B \in \mathcal{P}(\{0, 1\})$, on en déduit :

$$\mathbb{P}(Y_1 \in A, Z_1 \in B) = \int_{A \times B} [\bar{F}(y; \theta) g(y)]^{1-z} [\hat{G}(y) f(y; \theta)]^z d\mu(y) d(\delta_0 + \delta_1)(z),$$

2. Comme la suite observée est i.i.d., par construction la vraisemblance est :

$$\prod_{i=1}^n [\bar{F}(y_i; \theta) g(y_i)]^{1-z_i} [f(y_i; \theta) \hat{G}(y_i)]^{z_i} = \left(\prod_{i=1}^n [\bar{F}(y_i; \theta)]^{1-z_i} [f(y_i; \theta)]^{z_i} \right) \left(\prod_{i=1}^n [g(y_i)]^{1-z_i} [\hat{G}(y_i)]^{z_i} \right)$$

Le second produit ne dépend pas de $\theta \in \Theta$, on en déduit le résultat.

3. On a $f(y_i, \theta) = \theta \exp(-\theta y_i)$ et $\bar{F}(y_i, \theta) = \exp(-\theta y_i)$. En passant au logarithme on cherche à maximiser :

$$\begin{aligned}\sum_{i=1}^n (1 - z_i) \log [\bar{F}(y_i; \theta)] + \sum_{i=1}^n z_i \log [f(y_i; \theta)] &= -\theta \sum_{i=1}^n (1 - z_i) y_i + \sum_{i=1}^n z_i \log(\theta) - \theta \sum_{i=1}^n z_i y_i \\ &= \sum_{i=1}^n z_i \log(\theta) - \theta \sum_{i=1}^n y_i.\end{aligned}$$

La condition du premier ordre donne :

$$\hat{\theta} := \frac{\sum_{i=1}^n z_i}{\sum_{i=1}^n y_i}.$$

Il s'agit du nombre d'observation sur la durée totale observée. Si on n'observe aucune censure ($z_i = 1$ pour tout i), alors on retrouve l'estimateur du maximum de vraisemblance classique de l'observation directe de $(X_i)_{1 \leq i \leq n}$.

4. On a (indication)

$$\begin{aligned}\mathbb{E}(\min(X, c)) &= \int_0^{+\infty} \min(x, c) d\mathbb{P}_\theta(x) \\ &= \int_0^{+\infty} \int_0^c \mathbf{1}_{\{x > t\}} dt d\mathbb{P}_\theta(x) \\ &= \int_0^c \int_0^{+\infty} \mathbf{1}_{\{x > t\}} d\mathbb{P}_\theta(x) dt \\ &= \int_0^c (1 - F(t; \theta)) dt.\end{aligned}$$

Par la loi des grands nombres, on a

$$\hat{\theta} \xrightarrow{p.s.} \frac{\mathbb{P}(X \leq C)}{\mathbb{E}(\min(X, C))}.$$

Le numérateur, se réécrit :

$$\int \int \mathbf{1}_{\{x \leq c\}} d\mathbb{P}_\theta(x) d\mathbb{P}_C(c) = \int F(\theta; c) d\mathbb{P}_C(c) = \int (1 - e^{-\theta c}) d\mathbb{P}_C(c).$$

Le dénominateur se réécrit :

$$\begin{aligned} \mathbb{E}(\min(X, C)) &= \int \int \min(x, c) d\mathbb{P}_\theta(x) d\mathbb{P}_C(c) \\ &= \int \int_0^c (1 - F(x; \theta)) dx d\mathbb{P}_C(c) \\ &= \int \int_0^c e^{-\theta x} dx d\mathbb{P}_C(c) \\ &= \frac{1}{\theta} \int (1 - e^{-\theta c}) d\mathbb{P}_C(c). \end{aligned}$$

On en déduit

$$\hat{\theta} \xrightarrow{p.s.} \theta.$$

5. Par la loi forte des grands nombres, le dénominateur converge vers :

$$\mathbb{E}(X \mathbf{1}_{\{X \leq C\}}) = \mathbb{E}(X | X \leq C) \mathbb{P}(X \leq C).$$

Ainsi, on en déduit :

$$\tilde{\theta} \xrightarrow{p.s.} \frac{1}{\mathbb{E}(X | X \leq C)}.$$

Or, par construction,

$$\mathbb{E}(X | X \leq C) < \mathbb{E}(X),$$

ainsi, on surestime θ .