PC 1 - Correction

Exercice 1 *Démonstrations de cours*

1. Considérons l'estimateur empirique de la moyenne $\overline{X}_n = \frac{1}{n}\sum\limits_{i=1}^{n} X_i$. Calcul du biais

$$E[\overline{X}_n] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \frac{1}{n}\sum_{i=1}^{n} E[X_1] = E[X_1] = m \tag{1}$$

   Calcul de la variance :

$$Var(\overline{X}_n) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_1) = \frac{1}{n}Var(X_1) = \frac{\sigma^2}{n} \tag{2}$$

   Estimateur sans biais de m à variance asymptotiquement nul => consistant.

2. Considérons l'estimateur empirique de la variance $\hat{V}_n = \frac{1}{n}\sum\limits_{i=1}^{n}(X_i - \overline{X})^2$. Calcul du biais :

$$\begin{aligned} E[\hat{V}_n] &= \tfrac{1}{n}\sum_{i=1}^{n} E[(X_i - \overline{X})^2] \\ &= \tfrac{1}{n}\sum_{i=1}^{n}(E[X_i^2] - 2E[X_i\overline{X}] + E[\overline{X}^2]) \end{aligned} \tag{3}$$

   Or, nous avons :
   — For any $i \in \{1, ..., n\}$:

$$E[X_i^2] = \sigma^2 + m^2 \tag{4}$$

   — For any $i \in \{1, ..., n\}$:

$$\begin{aligned} E[X_i\overline{X}] &= E[X_i \tfrac{1}{n}\sum_{j=1}^{n} X_j] \\ &= \tfrac{1}{n}\sum_{j=1}^{n} E[X_iX_j] \\ &= \tfrac{1}{n}\sum_{j=1, j\neq i}^{n} E[X_iX_j] + \tfrac{1}{n}E[X_i^2] \\ &= \tfrac{1}{n}\sum_{j=1, j\neq i}^{n} E[X_i]E[X_j] + \tfrac{1}{n}(Var(X_i) + m^2) \\ &= \tfrac{n-1}{n}m^2 + \tfrac{1}{n}(\sigma^2 + m^2) \\ &= m^2 + \tfrac{1}{n}\sigma^2 \end{aligned} \tag{5}$$

   —

$$E[\overline{X}^2] = Var(\overline{X}) + E[\overline{X}]^2 = \frac{\sigma^2}{n} + m^2 \tag{6}$$

   Ainsi, finalement :

$$E[\hat{V}_n] = \tfrac{1}{n}\sum_{i=1}^{n}(\sigma^2 + m^2 - 2m^2 - \tfrac{2}{n}\sigma^2 + \tfrac{1}{n}\sigma^2 + m^2) = \tfrac{n-1}{n}\sigma^2 \tag{7}$$

Exercice 2 *Estimateur des moments d'une loi uniforme*

1. D'une part :

$$m_1(\theta) = E_\theta[X_i] = \int_{\mathbb{R}} x\frac{1}{\theta}\mathbb{1}_{[0;1]}(x)dx = \int_0^\theta x\frac{1}{\theta}dx = \frac{\theta}{2} \tag{8}$$

D'autre part :

$$\hat{m_1}(\theta) = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{9}$$

Ainsi :

$$\frac{\hat{\theta}_n}{2} = \frac{1}{n}\sum_{i=1}^{n} X_i, \tag{10}$$

et donc :

$$\hat{\theta}_n = \frac{2}{n}\sum_{i=1}^{n} X_i, \tag{11}$$

2. L'espérance de $\hat{\theta}_n$ vaut :

$$E[\hat{\theta}_n] = \frac{2}{n}\sum_{i=1}^{n} E[X_i] = \frac{2}{n}\sum_{i=1}^{n}\frac{\theta}{2} = \theta. \tag{12}$$

Sa variance vaut :

$$Var(\hat{\theta}_n) = \frac{4}{n^2}Var(\sum_{i=1}^{n} X_i) = \frac{4}{n^2}\sum_{i=1}^{n} var(X_i) = \frac{4}{n^2}\sum_{i=1}^{n}\frac{\theta^2}{12} = \frac{\theta^2}{3n} \tag{13}$$

3. $\hat{\theta}_n$ est donc non biaisé. De plus, sa variance tend est asymptotiquement nulle. Ainsi, il s'agit d'un estimateur consistant.

Exercice 3 *Comparaison d'estimateurs d'un paramètre d'une loi uniforme*

1. *Pour rappel, la fonction de densité d'une variable aléatoire uniforme sur* $[0;\theta]$ *vaut* $f_\theta(x) = \frac{1}{\theta}\mathbf{1}_{[0;\theta]}$.

Si un tel estimateur sans biais existe, on a :

$$E[S_n^{(1)}] = g(\theta). \tag{14}$$

Or, on a :

$$E[S_n^{(1)}] = \frac{1}{n}\sum_{i=1}^{n} E[T(X_i)] = E[T(X_1)] = \int_{-\infty}^{\infty} T(x)f_\theta(x)dx = \frac{1}{\theta}\int_0^\theta T(x)dx. \tag{15}$$

Ainsi, il vient :

$$\int_0^\theta T(x)dx = \theta g(\theta) \tag{16}$$

D'où :

$$T(x) = [xg(x)]' = xg'(x) + g(x). \tag{17}$$

Réciproquement, $T(x) = xg'(x) + g(x)$ convient.

2. (a) La fonction de répartition s'écrit (cas de tirages indépendants) :

$$\begin{aligned} F_X(x;\theta,n) &= P(X_{n:n} \le x;\theta,n) = P(X_1 \le x, ..., X_n \le x;\theta,n) \\ &= P(X_1 \le x;\theta,n)P(X_2 \le x;\theta,n)...P(X_n \le x;\theta,n) = \begin{cases} 0 \quad si \quad x \le 0 \\ (\frac{x}{\theta})^n \quad si \quad 0 \le x \le \theta \\ 1 \quad si \quad x \ge \theta \end{cases} \end{aligned} \tag{18}$$

Ainsi, en dérivant la fonction de répartition, on trouve la fonction densité :

$$f_{\theta,n}(x) = \begin{cases} 0 \quad si \quad x \le 0 \\ n\frac{1}{\theta}(\frac{x}{\theta})^{n-1} \quad si \quad 0 \le x \le \theta \\ 0 \quad si \quad x \ge \theta \end{cases}, \tag{19}$$

ce qui s'écrit encore :

$$f_{\theta,n}(x) = \frac{n}{\theta^n} x^{n-1} \mathbf{1}_{[0,\theta]} \tag{20}$$

(b) Si un tel estimateur sans biais existe, on a :

$$E[S_n^{(2)}] = g(\theta). \tag{21}$$

Or, on a :

$$E[S_n^{(2)}] = E[\tilde{T}(X_{n:n})] = \int_{-\infty}^{\infty} \tilde{T}(x) f_{\theta,n}(x)dx = \int_{-\infty}^{\infty} \tilde{T}(x) \frac{n}{\theta^n} x^{n-1} \mathbf{1}_{[0,\theta]} dx = \frac{n}{\theta^n} \int_0^{\theta} \tilde{T}(x) x^{n-1} dx. \tag{22}$$

Ainsi, il vient :

$$\int_0^{\theta} \tilde{T}(x) x^{n-1} dx = \frac{\theta^n}{n} g(\theta) \tag{23}$$

D'où :

$$\tilde{T}(x) = \frac{[x^n g(x)]'}{n x^{n-1}} = \frac{x g'(x)}{n} + g(x). \tag{24}$$

Réciproquement, $\frac{xg'(x)}{n} + g(x)$ convient.

3. Ici $g(x) = x$. Ainsi $T(x) = 2x$ et $\tilde{T}(x) = \frac{n+1}{n} x$. Le calcul du risque quadratique moyen de $S_n^{(1)}$ donne :

$$\begin{aligned}
&EQM(S_n^{(1)}, \theta) = E[(S_n^{(1)} - \theta)^2] = E[{S_n^{(1)}}^2] - 2E[S_1^{(n)}]\theta + \theta^2] = E[(\tfrac{1}{n} \sum_{i=1}^n T(X_i))^2] - 2\theta^2 + \theta^2 \\
&= \tfrac{1}{n^2} E[\sum_{i=1}^n T(X_i)^2 + \sum_{i=1}^n \sum_{j=1, j\neq i}^n T(X_i)T(X_j)] - \theta^2 \\
&= \tfrac{1}{n^2} \sum_{i=1}^n E[T(X_i)]^2 + \sum_{i=1}^n \sum_{j=1, j\neq i}^n E[T(X_i)T(X_j)] - \theta^2 \\
&= \tfrac{1}{n^2} \sum_{i=1}^n E[T(X_i)^2] + \tfrac{1}{n^2} \sum_{i=1}^n \sum_{j=1, j\neq i}^n E[T(X_i)]E[T(X_j)] - \theta^2, \qquad independance \quad des \quad X_i \\
&= \tfrac{1}{n^2} n E[T(X_1)^2] + \tfrac{1}{n^2} \sum_{i=1}^n \sum_{j=1, j\neq i}^n E[T(X_1)]E[T(X_1)] - \theta^2, \qquad X_i \quad identiquement \quad distribues \\
&= \tfrac{1}{n} \int_{-\infty}^{\infty} T(x)^2 f_\theta(x)dx + \tfrac{1}{n^2} n(n-1)\theta^2 - \theta^2 \\
&= \tfrac{1}{n\theta} \int_0^{\theta} (2x)^2 dx - \tfrac{\theta^2}{n} \\
&= \tfrac{1}{3n}\theta^2
\end{aligned} \tag{25}$$

Le calcul du risque quadratique moyen de $S_n^{(1)}$ donne :

$$\begin{aligned}
&EQM(S_n^{(2)}, \theta) = E[(S_n^{(2)} - \theta)^2] = E[{S_n^{(2)}}^2] - 2E[S_2^{(n)}]\theta + \theta^2] = E[\tilde{T}(X_{n:n})^2] - 2\theta^2 + \theta^2 \\
&= \tfrac{1}{n} \int_{-\infty}^{\infty} \tilde{T}(x)^2 f_\theta(x)dx - \theta^2 \\
&= \int_0^{\theta} (\tfrac{n+1}{n})^2 x^2 \tfrac{n}{\theta^n} x^{n-1} dx - \theta^2 \\
&= \tfrac{1}{n(n+2)}\theta^2
\end{aligned} \tag{26}$$

Finalement, la différence des erreurs quadratique s'écrit

$$EQM(S_n^{(2)}, \theta) - EQM(S_n^{(1)}, \theta) = \frac{1}{n(n+2)}\theta^2 - \frac{1}{3n}\theta^2 = \frac{1-n}{3n(n+2)}\theta^2, \tag{27}$$

ce qui est négatif, quel que soit $\theta \in \mathbb{R}_+^*$ et quel que soit $n \geq 1$. Ainsi, $EQM(S_n^{(2)}, \theta) \leq EQM(S_n^{(1)}, \theta)$ et $S_n^{(2)}$ est préférable à $S_n^{(1)}$, qui est donc inadmissible pour tout $n \geq 1$.

Exercice 4 *Matrice d'Information de Fisher et Borne de Cramer-Rao*

1. La densité de probabilité s'écrit :

$$p(Z;(\beta_1,\beta_2)) = \frac{1}{(2\pi\sigma^2)^{n/2}} exp\left(\tfrac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i-\beta_1-\beta_2 t_i)^2\right) \tag{28}$$

Ainsi :

$$ln(p(Z;(\beta_1,\beta_2))) = -\frac{n}{2}ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i-\beta_1-\beta_2 t_i)^2 \tag{29}$$

Puis, en dérivant par rapport au couple (β_1,β_2), on trouve :

$$\frac{\partial ln(p(Z;(\beta_1,\beta_2)))}{\partial(\beta_1,\beta_2)} = \begin{bmatrix} \frac{\partial ln(p(Z;(\beta_1,\beta_2)))}{\partial\beta_1} \\ \frac{\partial ln(p(Z;(\beta_1,\beta_2)))}{\partial\beta_2} \end{bmatrix} = \frac{1}{\sigma^2}\begin{bmatrix} \sum_{i=1}^{n}(X_i-\beta_1-\beta_2 t_i) \\ \sum_{i=1}^{n} t_i(X_i-\beta_1-\beta_2 t_i) \end{bmatrix} \tag{30}$$

En notant $Y_i = X_i-\beta_1-\beta_2 t_i$ et en remarquant que $Y_i \sim \mathcal{N}(0,\sigma^2)$, et que pour tout $i \neq j$, $Cov(Y_i,Y_j)=0$, la matrice d'information de Fisher peut se calculer par :

$$F(\beta_1,\beta_2) = Cov(\tfrac{\partial ln(p(Z;(\beta_1,\beta_2)))}{\partial(\beta_1,\beta_2)}) = \tfrac{1}{\sigma^4}\begin{bmatrix} Var(\sum_{i=1}^{n} Y_i) & Cov(\sum_{i=1}^{n} Y_i, \sum_{i=1}^{n} t_iY_i) \\ Cov(\sum_{i=1}^{n} Y_i, \sum_{i=1}^{n} t_iY_i) & Var(\sum_{i=1}^{n} t_iY_i) \end{bmatrix}$$
$$= \tfrac{1}{\sigma^4}\begin{bmatrix} n\sigma^2 & \sum_{i=1}^{n} t_i\sigma^2 \\ \sum_{i=1}^{n} t_i\sigma^2 & \sum_{i=1}^{n} t_i^2\sigma^2 \end{bmatrix} = \tfrac{1}{\sigma^2}\begin{bmatrix} n & \bar{t} \\ \bar{t} & ||t||^2 \end{bmatrix} \tag{31}$$

avec $\bar{t} = \sum_{i=1}^{n} t_i$.

2. L'inverse de la matrice de Fisher s'écrit :

$$F^{-1}(\beta_1,\beta_2) = \frac{\sigma^2}{n||t||^2-\bar{t}^2}\begin{bmatrix} ||t||^2 & -\bar{t} \\ -\bar{t} & n \end{bmatrix} \tag{32}$$

En appliquant la propriété proposé dans l'énoncé de l'exercice avec $\theta = (\beta_1,\beta_2)$ et $g(\beta_1,\beta_2) = \beta_1$, la borne de Cramer-Rao demandé s'écrit :

$$\frac{\partial g(\beta_1,\beta_2)}{\partial(\beta_1,\beta_2)}^T F^{-1}(\beta_1,\beta_2)\frac{\partial g(\beta_1,\beta_2)}{\partial(\beta_1,\beta_2)} = \begin{bmatrix} 1 & 0 \end{bmatrix}\frac{\sigma^2}{n||t||^2-\bar{t}^2}\begin{bmatrix} ||t||^2 & -\bar{t} \\ -\bar{t} & n \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \frac{\sigma^2||t||^2}{n||t||^2-\bar{t}^2} \tag{33}$$

3. On suppose à présent β_2 connu. Ainsi, la dérivation de la log-vraisemblance se fait uniquement par rapport à β_1. Il vient :

$$\frac{\partial ln(p(Z;\beta_1)))}{\partial\beta_1} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\beta_1-\beta_2 t_i), \tag{34}$$

et la matrice d'information de Fisher s'écrit ensuite :

$$F(\beta_1) = \frac{1}{\sigma^4}Var(\sum_{i=1}^{n} Y_i) = \frac{1}{\sigma^4}n\sigma^2 = \frac{n}{\sigma^2} \tag{35}$$

Finalement, une borne inférieure de la variance d'un estimateur sans biais de β_1 est cette fois $\frac{\sigma^2}{n}$.

Cette borne est donc inférieure à la précédente. En fait, toute l'information est utilisée pour estimer β_1 dans ce cas, alors que dans le premier cas, il y a une dispersion de l'information sur l'estimation de β_1 et de β_2.

Exercice 5 Estimateur du maximum de vraisemblance du paramètre d'une loi uniforme

1. Notons pour commencer, que par définition du tirage de la loi uniforme proposée, on a pour tout $i \in \{1, ..., n\}$; $0 \leq x_i \leq \theta$. Rappelons également, que la vraisemblance est une fonction de θ pour des réalisations $x_1, ..., x_n$ données. Ainsi, la vraisemblance s'écrit :

$$L(\theta; x_1, ..., x_n) = \prod_{i=1}^{n} f_\theta(x_i) = \begin{cases} \frac{1}{\theta^n} & si \quad \theta \geq max(x_1, ...x_n), \\ 0 & sinon. \end{cases} \tag{36}$$

Ainsi, le θ maximisant la vraisemblance est solution de :

$$\max_{\theta \geq max(x_1, ..., x_n)} \frac{1}{\theta^n}. \tag{37}$$

dont la solution vaut $max(x_1, ..., x_n)$. Ainsi, l'estimateur du maximmum de vraisembance s'écrit :

$$\hat{\theta}_n = max(X_1, ..., X_n). \tag{38}$$

2. Sa fonction de répartition s'écrit :

$$\begin{aligned} F_{\hat{\theta}_n} &= \mathbb{P}(\hat{\theta}_n \leq x) = \mathbb{P}(max(X_1, ..., X_n \leq x) = \mathbb{P}(X_1 \leq x, ..., X_n \leq x) = \prod_{i=1}^{n} \mathbb{P}(X_i \leq x) \\ &= \mathbb{P}(X_1 \leq x)^n = \mathbb{P}(X_1 \leq x)^n = \begin{cases} (\frac{x}{\theta})^n & x \leq \theta, \\ 1 & sinon. \end{cases} \end{aligned} \tag{39}$$

Ainsi, sa densité s'écrit :

$$f_{\hat{\theta}_n}(x) = \frac{n}{\theta^n} x^{n-1} \mathbb{1}_{[0;\theta]}(x) \tag{40}$$

3. Son espérance s'écrit donc :

$$E[\hat{\theta}_n] = \int_{-\infty}^{+\infty} x f_{\hat{\theta}_n}(x) dx = \int_{0}^{\theta} x \frac{n}{\theta^n} x^{n-1} dx = \frac{n}{\theta^n} \int_{0}^{\theta} x^n dx = \frac{n}{n+1} \theta. \tag{41}$$

D'où un biais :

$$B(\hat{\theta}_n, \theta) = \frac{n}{n+1}\theta - \theta = \frac{-1}{n+1}\theta \tag{42}$$

D'autre part, on trouve :

$$E[\hat{\theta}_n^2] = \int_{-\infty}^{+\infty} x^2 f_{\hat{\theta}_n}(x) dx = \frac{n}{\theta^n} \int_{0}^{\theta} x^{n+1} dx = \frac{n}{n+2} \theta^2. \tag{43}$$

Ainsi :

$$Var(\hat{\theta}_n) = E[\hat{\theta}_n^2] - E[\hat{\theta}_n]^2 = \frac{n}{n+2}\theta^2 - \frac{n^2}{(n+1)^2}\theta^2 = \frac{n}{(n+2)(n+1)^2}\theta^2 \tag{44}$$

Enfin, le risque quadratique moyen s'écrit :

$$RQM(\hat{\theta}_n, \theta) = Biais(\hat{\theta}_n, \theta)^2 + Var(\hat{\theta}_n) = \frac{1}{(n+1)^2}\theta^2 + \frac{n}{(n+2)(n+1)^2}\theta^2 = \frac{2}{(n+1)(n+2)}\theta^2 \tag{45}$$

Exercice 6 Estimateur du maximum de vraisemblance

1. Par le fait que les variables sont indépendantes et identiquement distribuées, la vraisemblance s'écrit :

$$L(x_1, ..., x_n; \theta) = f(X_1 = x_1, ..., X_n = x_n; \theta) = \prod_{i=1}^{n} f(X_i = x_i; \theta) = \prod_{i=1}^{n} \frac{x_i}{\theta} exp(\frac{-x_i}{2\theta}) I_{\mathbb{R}_+}(x_i) \tag{46}$$

Ainsi, la Log-vraisemblance s'écrit :

$$\ln L(x_1, ..., x_n; \theta) = \sum_{i=1}^{n} \left(\ln(x_i) - \ln(\theta) - \frac{x_i^2}{2\theta}\right) = -n\ln(\theta) + \sum_{i=1}^{n} \left(\ln(x_i)\right) - \frac{1}{2\theta} \sum_{i=1}^{n} \left(x_i^2\right) \tag{47}$$

Calculons les dérivées d'ordre 1 et 2 par rapport à θ de cette expression. On trouve :

$$\begin{cases} \frac{\partial \ln L(x_1,...,x_n;\theta)}{\partial\theta} & = -\frac{n}{\theta} + \frac{1}{2\theta^2} \sum_{i=1}^{n} x_i^2 \\ \frac{\partial^2 \ln L(x_1,...,x_n;\theta)}{\partial\theta^2} & = \frac{n}{\theta^2} - \frac{1}{\theta^3} \sum_{i=1}^{n} x_i^2 \end{cases} \tag{48}$$

Ainsi, $\frac{\partial \ln L(x_1,...,x_n;\theta)}{\partial\theta} = 0$ équivaut à :

$$\theta = \frac{1}{2n} \sum_{i=1}^{n} x_i^2, \tag{49}$$

et pour cette valeur de θ, on trouve :

$$\frac{\partial^2 \ln L(x_1, ..., x_n; \theta)}{\partial\theta^2} = \frac{-4n^3}{[\sum_{i=1}^{n} x_i^2]^2} \tag{50}$$

Ainsi, la fonction de vraisemblance admet un unique extremum, et il s'agit d'un maximum global. D'où l'estimateur du maximum de vraisemblance suivant :

$$\hat{\theta} = \frac{1}{2n} \sum_{i=1}^{n} X_i^2. \tag{51}$$

2. Moyenne :

$$E[\hat{\theta}] = E[\frac{1}{2n} \sum_{i=1}^{n} x_i^2] = \frac{1}{2n} \sum_{i=1}^{n} E[x_i^2] = \frac{1}{2n} \sum_{i=1}^{n} 2\theta = \theta \tag{52}$$

Variance (la troisième égalité provient de l'indépendance des x_i) :

$$\begin{gathered} Var(\theta) = \frac{1}{4n^2} Var(\sum_{i=1}^{n} x_i^2) = \frac{1}{4n^2} \sum_{i=1}^{n} Var(x_i^2) + \frac{1}{4n^2} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} Covar(x_i^2, x_j^2) = \frac{1}{4n} Var(x_1^2) + 0 \\ = \frac{1}{4n}\left(E[x_1^4] - E[x_1^2]^2\right) = \frac{1}{4n}\left(8\theta^2 - 4\theta^2\right) = \frac{\theta^2}{n} \underset{n\to\infty}{\longrightarrow} 0 \end{gathered} \tag{53}$$

Ainsi, l'estimateur est non biaisé et converge vers θ en moyenne quadratique.