Correction TD 3 : EMV, EMM, Efficacité

Correction exercice 1. Calculs d'EMV

On rappelle que

- Si X est discrète, la vraisemblance d'un échantillon ponctuel $x_1, \cdots x_n$ est donné par

$$L(\theta, x_1, \cdots, x_n) = \prod_{i=1}^{n} \mathbb{P}_\theta(X_i = x_i)$$

- Si X est de densité f_θ, la vraisemblance d'un échantillon ponctuel $x_1, \cdots x_n$ est donné par

$$L(\theta, x_1, \cdots, x_n) = \prod_{i=1}^{n} f_\theta(x_i).$$

1. Soit $X \sim \mathcal{B}(p)$. On a donc $\mathbb{P}_p(X = 0) = 1 - p$ et $\mathbb{P}_p(X = 1) = p$, ce qui peut se réécrire $\mathbb{P}_p(X = x) = p^x(1-p)^{1-x}$ (remplacer x par 0 ou 1 pour s'en convaincre).

   La vraisemblance est alors donnée par

$$\begin{aligned} L(p, x_1, \cdots, x_n) &= \prod_{i=1}^{n} \mathbb{P}_p(X_i = x_i) \\ &= \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} \\ &= p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i} \end{aligned}$$

   Remarque : Cette fonction ne dépend en réalité que de p et du nombre de 1 dans l'échantillon $\sum_{i=1}^{n} x_i$. On peut la représenter en tant que fonction de deux variables. Lorsqu'on observe les réalisations x_i de X_i, il suffit alors de chercher le p qui la maximise, ce qui peut se faire en dérivant par rapport à p.

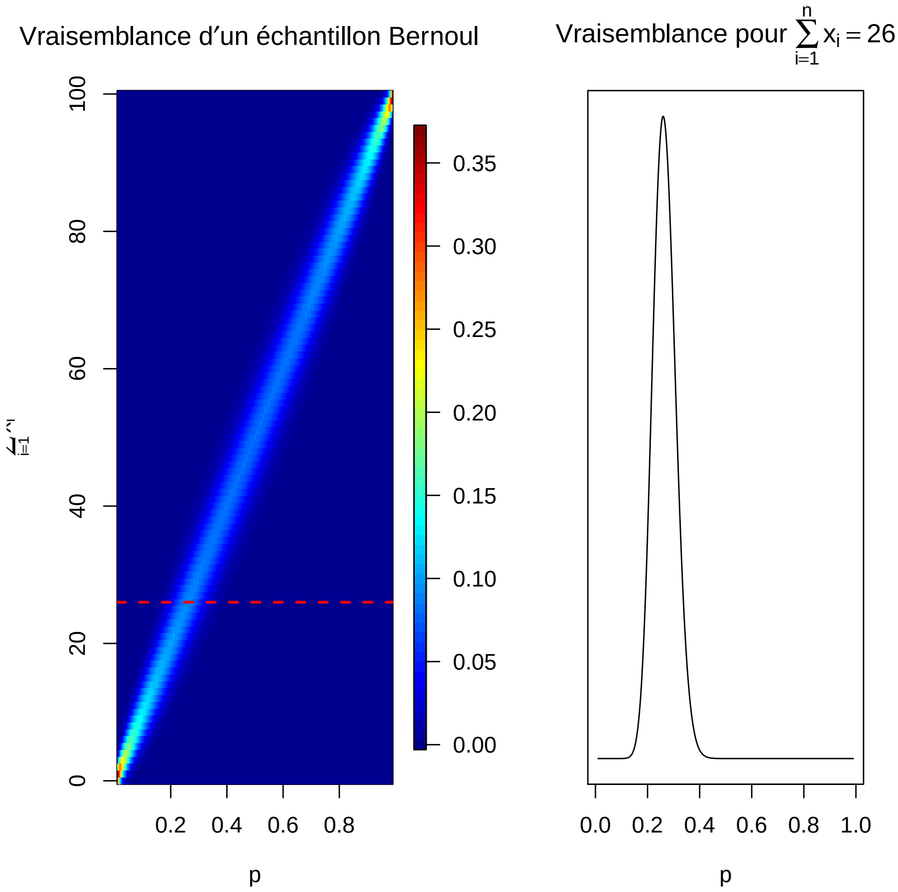

Figure 1: Représentation de la vraisemblance d'un échantillon Bernoulli, avec $n = 100$. À gauche, on représente $(p, \sum_{i=1}^{n} x_i) \mapsto p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i}$, et à droite, une coupe de cette fonction pour $\sum_{i=1}^{n} x_i = 26$, i.e. la fonction $p \mapsto p^{26}(1-p)^{100-26}$. Pour constuire l'estimateur par maximum de vraisemblance, on souhaite donc fixer les x_i et trouver, en fonction des x_i le p qui maximise cette vraisemblance.

La vraisemblance étant $\mathcal{C}_1$, on peut dès lors dériver pour trouver (en fonction des x_i) le p pour lequel le maximum est atteint, mais comme on n'aime pas trop dériver le produit, on va plutôt passer au log (qui est une fonction croissante), et chercher le maximum de la log-vraisemblance l, qui correspondra aussi au maximum de la vraisemblance L (car le log est croissant).

Remarque : On parle de log, et je noterais indifféremment log ou ln selon l'humeur, mais en réalité, il s'agira toujours du log népérien.

Cela donne

$$\begin{aligned} l(p, x_1, \cdots x_n) &= \ln\Big(L(p, x_1, \cdots, x_n)\Big) \\ &= \ln\left(p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i}\right) \\ &= \sum_{i=1}^{n} x_i \ln(p) + \left(n - \sum_{i=1}^{n} x_i\right) \ln(1-p) \end{aligned}$$

On peut donc résoudre

$$\begin{aligned}
\frac{\partial}{\partial p} l(\hat{p}, x_1, \cdots x_n) = 0 &\Leftrightarrow \frac{\sum_{i=1}^n x_i}{\hat{p}} - \frac{n - \sum_{i=1}^n x_i}{1-\hat{p}} = 0 \\
&\Leftrightarrow \frac{\sum_{i=1}^n x_i}{\hat{p}} = \frac{n - \sum_{i=1}^n x_i}{1-\hat{p}} \\
&\Leftrightarrow (1-\hat{p}) \sum_{i=1}^n x_i = \hat{p} n - \hat{p} \sum_{i=1}^n x_i \\
&\Leftrightarrow \sum_{i=1}^n x_i = \hat{p} n \\
&\Leftrightarrow \hat{p} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}
\end{aligned}$$

On peut vérifier qu'il s'agit bien d'un maximum (plusieurs façon, mais comme l est $\mathcal{C}_2$, par exemple en dérivant une seconde fois).

$$\frac{\partial^2}{\partial p^2} l(\hat{p}, x_1, \cdots x_n) = -\frac{\sum_{i=1}^n x_i}{\hat{p}^2} - \frac{n - \sum_{i=1}^n x_i}{(1-\hat{p})^2} < 0,$$

donc il s'agit bien d'un maximum.

On vient de calculer l'estimateur du maximum de vraisemblance pour un échantillon ponctuel $\hat{p} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}$. On en déduit l'estimateur de p par maximum de vraisemblance (en tant que statistique) pour un échantillon Bernoulli :

$$\hat{P}_{EMV} = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}.$$

2. On peut maintenant aller plus vite !

   Soit $X \sim \mathcal{G}(p)$, alors on peut écrire la vraisemblance, et la log-vraisemblance d'un échantillon ponctuel $x_1, \cdots, x_n$:

$$\begin{aligned}
L(p, x_1, \cdots x_n) &= \prod_{i=1}^n \mathbb{P}_p(X_i = x_i) \\
&= \prod_{i=1}^n (1-p)^{x_i - 1} p \\
l(p, x_1, \cdots x_n) &= \ln \Big(L(p, x_1, \cdots, x_n) \Big) \\
&= \ln \Big(\prod_{i=1}^n (1-p)^{x_i - 1} p \Big) \\
&= \sum_{i=1}^n \Big((x_i - 1) \ln(1-p) + \ln(p) \Big) \\
&= \left(\Big(\sum_{i=1}^n x_i \Big) - n \right) \ln(1-p) + n \ln(p)
\end{aligned}$$

Ce qui permet de dériver

$$\begin{aligned}
\frac{\partial}{\partial p} l(\hat{p}, x_1, \cdots x_n) = 0 &\Leftrightarrow -\frac{\left(\sum_{i=1}^n x_i\right) - n}{1-\hat{p}} + \frac{n}{\hat{p}} = 0 \\
&\Leftrightarrow \hat{p}\Big(\sum_{i=1}^n x_i\Big) - n\hat{p} = n(1-\hat{p}) \\
&\Leftrightarrow \hat{p} \sum_{i=1}^n x_i = n \\
&\Leftrightarrow \hat{p} = \frac{n}{\sum_{i=1}^n x_i} = \frac{1}{\frac{1}{n}\sum_{i=1}^n x_i} = \frac{1}{\bar{x}}
\end{aligned}$$

et c'est bien un maximum car l est $\mathcal{C}_2$ et

$$\frac{\partial^2}{\partial p^2} l(\hat{p}, x_1, \cdots x_n) = -\frac{\left(\sum_{i=1}^n x_i\right) - n}{(1-\hat{p})^2} - \frac{n}{\hat{p}^2} < 0.$$

L'estimateur par maximum de vraisemblance de p dans un modèle $\mathcal{G}(p)$ est donc

$$\hat{P}_{EMV} = \frac{1}{\bar{X}}.$$

3. Soit $X \sim \mathcal{B}in(N, p)$, alors on peut écrire la vraisemblance, et la log-vraisemblance d'un échantillon ponctuel $x_1, \cdots, x_n$:

$$\begin{aligned}
L(p, x_1, \cdots x_n) &= \prod_{i=1}^n \mathbb{P}_p(X_i = x_i) \\
&= \prod_{i=1}^n \binom{N}{x_i} p^{x_i}(1-p)^{N-x_i} \\
l(p, x_1, \cdots x_n) &= \ln\Big(L(p, x_1, \cdots, x_n)\Big) \\
&= \ln\left(\prod_{i=1}^n \binom{N}{x_i} p^{x_i}(1-p)^{N-x_i}\right) \\
&= \sum_{i=1}^n \left(\ln\binom{N}{x_i} + x_i \ln(p) + (N-x_i)\ln(1-p)\right) \\
&= \sum_{i=1}^n \ln\binom{N}{x_i} + \Big(\sum_{i=1}^n x_i\Big)\ln(p) + \Big(Nn - \sum_{i=1}^n x_i\Big)\ln(1-p)
\end{aligned}$$

Ce qui permet de dériver

$$\begin{aligned}
\frac{\partial}{\partial p} l(\hat{p}, x_1, \cdots x_n) = 0 &\Leftrightarrow 0 + \frac{\sum_{i=1}^n x_i}{\hat{p}} - \frac{Nn - \sum_{i=1}^n x_i}{1-\hat{p}} = 0 \\
&\Leftrightarrow \frac{\sum_{i=1}^n x_i}{\hat{p}} = \frac{Nn - \sum_{i=1}^n x_i}{1-\hat{p}} \\
&\Leftrightarrow (1-\hat{p})\sum_{i=1}^n x_i = \hat{p}\Big(Nn - \sum_{i=1}^n x_i\Big) \\
&\Leftrightarrow \sum_{i=1}^n x_i = \hat{p}Nn \\
&\Leftrightarrow \hat{p} = \frac{1}{nN}\sum_{i=1}^n x_i = \frac{\bar{x}}{N}
\end{aligned}$$

et c'est bien un maximum car l est $\mathcal{C}_2$ et

$$\frac{\partial^2}{\partial p^2} l(\hat{p}, x_1, \cdots x_n) = -\frac{\sum_{i=1}^n x_i}{\hat{p}^2} - \frac{Nn - \sum_{i=1}^n x_i}{(1-\hat{p})^2} < 0.$$

L'estimateur par maximum de vraisemblance de p dans un modèle $\mathcal{G}(p)$ est donc

$$\hat{P}_{EMV} = \frac{\bar{X}}{N}.$$

4. Dans cette question, on va considérer une variable continue, et donc la vraisemblance va s'écrire en fonction de la densité de cette variable. On peut noter qu'on supprime la partie indicatrice dans le calcul, qui a juste l'effet de rendre la vraisemblance nulle (et donc la log-vraisemblance égale à $-\infty$) dès qu'un des x_i n'est pas dans l'intervalle autorisé (ici $\mathbb{R}^+$).

   Soit $X \sim \mathcal{E}(\lambda)$, alors on peut écrire la vraisemblance, et la log-vraisemblance d'un échantillon ponctuel $x_1, \cdots, x_n$.

$$\begin{aligned} L(\lambda, x_1, \cdots x_n) &= \prod_{i=1}^n f_\lambda(x_i) \\ &= \prod_{i=1}^n \lambda e^{-\lambda x_i} \\ l(\lambda, x_1, \cdots x_n) &= \ln\Big(L(\lambda, x_1, \cdots, x_n)\Big) \\ &= \ln\left(\prod_{i=1}^n \lambda e^{-\lambda x_i}\right) \\ &= \sum_{i=1}^n \Bigg(\ln(\lambda) - \lambda x_i\Bigg) \\ &= n\ln(\lambda) - \lambda\sum_{i=1}^n x_i \end{aligned}$$

   Ce qui permet de dériver

$$\begin{aligned} \frac{\partial}{\partial \lambda} l(\hat{\lambda}, x_1, \cdots x_n) = 0 &\Leftrightarrow \frac{n}{\hat{\lambda}} - \sum_{i=1}^n x_i = 0 \\ &\Leftrightarrow \frac{n}{\hat{\lambda}} = \sum_{i=1}^n x_i \\ &\Leftrightarrow \hat{\lambda} = \frac{n}{\sum_{i=1}^n x_i} = \frac{1}{\bar{x}} \end{aligned}$$

   et c'est bien un maximum car l est $\mathcal{C}_2$ et

$$\frac{\partial^2}{\partial \lambda^2} l(\hat{\lambda}, x_1, \cdots x_n) = -\frac{1}{\hat{\lambda}^2} < 0.$$

   L'estimateur par maximum de vraisemblance de p dans un modèle $\mathcal{G}(p)$ est donc

$$\hat{\Lambda}_{EMV} = \frac{1}{\bar{X}}.$$

Correction exercice 2. Un exemple

1. Il serait naturel d'utiliser une loi Binomiale pour modéliser le nombre de pièces défectueuses par lots, en considérant que chaque pièce à une probabilité p d'être défectueuse, et qu'il y a indépendance

entre les pièces. Si p est petit (de l'ordre de $\frac{1}{n}$), une binomiale peut-être approchée par une Poisson. En effet, si $X_n \sim \mathcal{B}in(n, p_n)$, avec $p_n = \frac{\lambda}{n}$, on a alors, pour $k \geq 0$ fixé, et $n \geq k$,

$$\begin{aligned}
\mathbb{P}(X_n = k) &= \binom{n}{k}\Big(\frac{\lambda}{n}\Big)^k\Big(1 - \frac{\lambda}{n}\Big)^{n-k} \\
&= \frac{n(n-1)\cdots(n-k+1)}{k!n^k}\frac{\lambda^k}{\big(1-\frac{\lambda}{n}\big)^k}\big(1 - \frac{\lambda}{n}\big)^n
\end{aligned}$$

Pour k fixé, on peut donc faire un DL en $n \to +\infty$ de $\mathbb{P}(X_n = k)$, ce qui donne

$$\begin{aligned}
\mathbb{P}(X_n = k) &= \frac{1+o(1)}{k!}\frac{\lambda^k}{\big(1+o(1)\big)^k}e^{n\ln\left(1-\frac{\lambda}{n}\right)} \\
&= \big(1+o(1)\big)\frac{\lambda^k}{k!}e^{-n\left(\frac{\lambda}{n}+o(\frac{1}{n})\right)} \\
&= \big(1+o(1)\big)\frac{\lambda^k}{k!}e^{-\lambda+o(1)} \\
&\xrightarrow{n\to+\infty} \frac{\lambda^k}{k!}e^{-\lambda} \\
&= \mathbb{P}\Big(\mathcal{P}(\lambda) = k\Big).
\end{aligned}$$

Cela justifie l'utilisation de la loi de Poisson pour modéliser le nombre de pièce défectueuses.

2. Soit $X \sim \mathcal{P}(\lambda)$, alors on peut écrire la vraisemblance, et la log-vraisemblance d'un échantillon ponctuel $x_1, \cdots, x_n$.

$$\begin{aligned}
L(\lambda, x_1, \cdots x_n) &= \prod_{i=1}^n \mathbb{P}_\lambda(X_i = x_i) \\
&= \prod_{i=1}^n \frac{\lambda^{x_i}}{x_i!}e^{-\lambda} \\
l(\lambda, x_1, \cdots x_n) &= \ln\Big(L(\lambda, x_1, \cdots, x_n)\Big) \\
&= \ln\left(\prod_{i=1}^n \frac{\lambda^{x_i}}{x_i!}e^{-\lambda}\right) \\
&= \sum_{i=1}^n \left(x_i\ln(\lambda) - \ln(x_i!) - \lambda\right) \\
&= \ln(\lambda)\Big(\sum_{i=1}^n x_i\Big) - \left(\sum_{i=1}^n \ln(x_i!)\right) - n\lambda
\end{aligned}$$

Ce qui permet de dériver

$$\begin{aligned}
\frac{\partial}{\partial\lambda}l(\hat\lambda, x_1, \cdots x_n) = 0 &\Leftrightarrow \frac{\sum_{i=1}^n x_i}{\hat\lambda} - 0 - n = 0 \\
&\Leftrightarrow \frac{\sum_{i=1}^n x_i}{\hat\lambda} = n \\
&\Leftrightarrow \hat\lambda = \frac{\sum_{i=1}^n x_i}{n} = \bar{x}
\end{aligned}$$

et c'est bien un maximum car l est $\mathcal{C}_2$ et

$$\frac{\partial^2}{\partial\lambda^2}l(\hat\lambda, x_1, \cdots x_n) = -\frac{\sum_{i=1}^n x_i}{\hat\lambda^2} < 0.$$

L'estimateur par maximum de vraisemblance de λ dans un modèle $\mathcal{P}(\lambda)$ est donc

$$\hat\Lambda_{EMV} = \bar{X}.$$

3. On a $p = \mathbb{P}(X \geq k) = \sum_{j=k}^{+\infty} \frac{\lambda^j}{j!} e^{-\lambda}$, on peut donc utiliser cette formule pour une estimation plug-in de la probabilité :
$$\hat{P} = \sum_{j=k}^{+\infty} \frac{\hat{\Lambda}_{EMV}^j}{j!} e^{-\hat{\Lambda}_{EMV}}.$$

4. On peut calculer directement
$$\begin{aligned}
\mathbb{E}[\hat{\Lambda}_{EMV}] &= \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[X_i] \\
&= \mathbb{E}[X_1] \text{ car } X_i \text{ i.i.d.} \\
&= \lambda \text{ l'espérance d'une v.a. de Poisson a été calculée en TD de proba}
\end{aligned}$$

   Donc $\hat{\Lambda}_{EMV}$ est un estimateur sans biais de λ.

5. Pour calculer la borne de Cramer-Rao, il faut calculer l'information de Fisher, et pour cela, le plus simple est souvent d'utiliser la formule
$$I_n(\lambda) = -\mathbb{E}_\lambda\Big[\frac{\partial^2}{\partial \lambda^2} l(\hat{\lambda}, X_1, \cdots X_n)\Big].$$
   Or, on a
$$\frac{\partial^2}{\partial \lambda^2} l(\lambda, X_1, \cdots X_n) = -\frac{\sum_{i=1}^n X_i}{\lambda^2},$$
   d'où
$$\begin{aligned}
I_n(\lambda) &= \mathbb{E}_\lambda\Big[\frac{\sum_{i=1}^n X_i}{\lambda^2}\Big] \\
&= \frac{\sum_{i=1}^n \mathbb{E}_\lambda[X_i]}{\lambda^2} \\
&= \frac{n\lambda}{\lambda^2} \\
&= \frac{n}{\lambda}
\end{aligned}$$
   On trouve donc la Borne de Cramèr-Rao :
$$BCR = \frac{1}{I_n(\lambda)} = \frac{\lambda}{n}$$

6. Pour savoir si l'EMV est efficace il suffit de calculer sa variance, et de voir si elle atteint la borne de Cramèr-Rao. On peut donc calculer
$$\begin{aligned}
\text{var}\left(\hat{\Lambda}_{EMV}\right) &= \text{var}(\bar{X}) \\
&= \frac{1}{n^2} \sum_{i=1}^{n} \text{var}(X_i) \text{ par indépendance} \\
&= \frac{\text{var}(X_i)}{n} \\
&= \frac{\lambda}{n}
\end{aligned}$$
   Comme $\text{var}\left(\hat{\Lambda}_{EMV}\right) = BCR$, l'estimateur par maximum de vraisemblance de λ est un estimateur efficace, dans le modèle Poisson.

7. Pour l'estimation ponctuelle, il suffit de calculer
$$\hat{\lambda}_{EMV} = \bar{x} = \frac{15 + 20 + \cdots + 6}{10} = 11.5$$

8. On peut utiliser l'estimation plug-in de $p = \mathbb{P}(X \geq 20)$:

$$\hat{p} = \sum_{j=20}^{+\infty} \frac{\hat{\lambda}_{EMV}^j}{j!} e^{-\hat{\lambda}_{EMV}}.$$

En calculant les premières valeurs (autre possibilité : utiliser $\mathbb{P}(X \geq 20) = 1 - \mathbb{P}(X \leq 20)$), on trouve :

$$\begin{aligned} \hat{p} &= \frac{(11.5)^{20}}{20!} e^{-11.5} + \frac{(11.5)^{21}}{21!} e^{-11.5} + \frac{(11.5)^{22}}{22!} e^{-11.5} + \cdots \\ &\approx 0.0068 + 0.0037 + 0.002 + 0.001 + 0.0005 \text{ en utilisant les 5 premières valeurs} \\ &\approx 0.0142 \end{aligned}$$

Pour information, en utilisant un logiciel qui permet de faire le calcul complet, on trouve $\hat{p} \approx 0.0143178$, ce qui n'est pas loin du résultat qu'on a trouvé (au passage, les 3 premières valeurs fournissent déjà un résultat correct).

Cela signifie qu'avec cette estimation, moins de 2% des lots devront être jetés.

Remarque : En réalité, ce type d'estimation plug-in n'est pas la meilleur façon de faire, mais ça donne une idée.

Correction exercice 3. Bornes de Cramér-Rao

1. $\mathcal{B}in(N, p)$, N connu

   On rappelle qu'on a trouvé à l'exercice 1, $\hat{P}_{EMV} = \frac{\bar{X}}{N}$, et que

$$\frac{\partial^2}{\partial p^2} l(p, X_1, \cdots X_n) = -\frac{\sum_{i=1}^n X_i}{p^2} - \frac{Nn - \sum_{i=1}^n X_i}{(1-p)^2}$$

   On peut alors calculer l'information de Fisher :

$$\begin{aligned} I_n(p) &= -\mathbb{E}_p\Big[\frac{\partial^2}{\partial p^2} l(p, X_1, \cdots X_n)\Big] \\ &= \mathbb{E}_p\Big[\frac{\sum_{i=1}^n X_i}{p^2}\Big] + \mathbb{E}_p\Big[\frac{Nn - \sum_{i=1}^n X_i}{(1-p)^2}\Big] \\ &= \frac{n}{p^2}\mathbb{E}_p[X_i] + \frac{1}{(1-p)^2}\Big(Nn - n\mathbb{E}_p[X_i]\Big) \\ &= \frac{n}{p^2} Np + \frac{1}{(1-p)^2}\Big(Nn - nNp\Big) \\ &= nN\Big(\frac{1}{p} + \frac{1}{(1-p)}\Big) \\ &= \frac{Nn}{p(1-p)} \end{aligned}$$

   On trouve alors la borne de Cramèr-Rao:

$$BCR = \frac{1}{I_n(p)} = \frac{p(1-p)}{Nn}.$$

   Pour savoir si $\hat{P}_{EMV}$ est efficace, il reste à savoir si sa variance atteint la borne de Cramèr-Rao. On calcule donc

$$\text{var}\left(\hat{P}_{EMV}\right) = \text{var}\left(\frac{\sum_{i=1^n} X_i}{Nn}\right) = \frac{1}{n^2N^2}\,\text{var}\Big(\sum_{i=1^n} X_i\Big) = \frac{n}{n^2N^2}\,\text{var}\Big(X_i\Big) = \frac{1}{nN^2} Np(1-p) = \frac{p(1-p)}{Nn}.$$

   Donc l'estimateur par maximum de vraisemblance $\hat{P}_{EMV}$ est un estimateur sans biais et efficace de p.

2. $\mathcal{N}(m, \sigma^2)$, σ connu

   On va rapidement :

   - Calculer la log-vraisemblance et la dériver
   - La maximiser
   - Vérifier si l'EMV est sans biais
   - Calculer la BCR
   - Calculer la variance de l'estimateur et vérifier s'il est sans biais.

$$
\begin{aligned}
l(m, x_1, \cdots, x_n) &= \sum_{i=1}^{n} \log\Big(\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-m)^2}{2\sigma^2}}\Big) \\
&= \sum_{i=1}^{n} \Bigg(-\log(\sqrt{2\pi}\sigma) - \frac{(x_i-m)^2}{2\sigma^2} \Bigg) \\
&= -n\log(\sqrt{2\pi}\sigma) - \frac{\sum_{i=1}^{n}(x_i-m)^2}{2\sigma^2} \\
\frac{\partial}{\partial m} l(m, x_1, \cdots, x_n) &= \frac{\sum_{i=1}^{n} 2(x_i-m)}{2\sigma^2} \\
\frac{\partial}{\partial m} l(m, x_1, \cdots, x_n) &= \frac{\sum_{i=1}^{n}(x_i-m)}{\sigma^2} \\
\frac{\partial^2}{\partial m^2} l(m, x_1, \cdots, x_n) &= -\frac{n}{\sigma^2}
\end{aligned}
$$

   Comme la vraisemblance est $\mathcal{C}_2$, on peut résoudre $\frac{\partial}{\partial m} l(\hat{m}, x_1, \cdots, x_n) = 0$ pour trouver l'argmax, ce qui donne

$$
\begin{aligned}
\frac{\partial}{\partial m} l(\hat{m}, x_1, \cdots, x_n) = 0 &\Leftrightarrow \frac{\sum_{i=1}^{n}(x_i-\hat{m})}{\sigma^2} = 0 \\
&\Leftrightarrow \hat{m} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}.
\end{aligned}
$$

   De plus, comme $\frac{\partial^2}{\partial m^2} l(m, x_1, \cdots, x_n) < 0$, il s'agit bien d'un maximum. L'EMV de m dans le modèle gaussien à variance connue est donc

$$\hat{M}_{EMV} = \bar{X}.$$

   De plus, c'est bien un estimateur sans biais, car $\mathbb{E}[\hat{M}_{EMV}] = \mathbb{E}[\bar{X}] = m$. On peut dès lors se demander s'il est efficace, et pour cela, on calcule l'information de Fisher du modèle :

$$I_n(m) = -\mathbb{E}_m\Big[\frac{\partial^2}{\partial m^2} l(m, X_1, \cdots, X_n)\Big] = \frac{n}{\sigma^2}.$$

   On a donc $BCR = \frac{\sigma^2}{n}$, or $\mathrm{var}(\hat{M}_{EMV}) = \mathrm{var}(\bar{X}) = \frac{\sigma^2}{n}$. Donc l'estimateur par maximum de vraisemblance $\hat{M}_{EMV}$ est un estimateur sans biais et efficace de m.

3. $\mathcal{N}(m, v)$, m connu

   On procède comme dans l'exercice précédent : on calcule et dérive tout d'abord la log-vraisemblance :

$$
\begin{aligned}
l(v, x_1, \cdots, x_n) &= \sum_{i=1}^{n} \log\Big(\frac{1}{\sqrt{2\pi v}} e^{-\frac{(x_i-m)^2}{2v}}\Big) \\
&= \sum_{i=1}^{n} \Bigg(-\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(v) - \frac{(x_i-m)^2}{2v} \Bigg) \\
&= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(v) - \frac{\sum_{i=1}^{n}(x_i-m)^2}{2v} \\
\frac{\partial}{\partial v} l(v, x_1, \cdots, x_n) &= -\frac{n}{2v} + \frac{\sum_{i=1}^{n}(x_i-m)^2}{2v^2} \\
\frac{\partial^2}{\partial v^2} l(v, x_1, \cdots, x_n) &= \frac{n}{2v^2} - 2\frac{\sum_{i=1}^{n}(x_i-m)^2}{2v^3}
\end{aligned}
$$

Comme la vraisemblance est $\mathcal{C}_2$, on peut résoudre $\frac{\partial}{\partial v} l(\hat{v}, x_1, \cdots, x_n) = 0$ pour trouver l'argmax, ce qui donne

$$\begin{aligned}
\frac{\partial}{\partial v} l(\hat{v}, x_1, \cdots, x_n) = 0 &\Leftrightarrow -\frac{n}{2\hat{v}} + \frac{\sum_{i=1}^n (x_i - m)^2}{2\hat{v}^2} = 0 \\
&\Leftrightarrow \frac{n}{2\hat{v}} = \frac{\sum_{i=1}^n (x_i - m)^2}{2\hat{v}^2} \\
&\Leftrightarrow \hat{v} = \frac{\sum_{i=1}^n (x_i - m)^2}{n}
\end{aligned}$$

Pour vérifier que c'est bien un maximum, le signe de la dérivée seconde n'est pas évident, il faut donc le calculer en $\hat{v}$ On a

$$\frac{\partial^2}{\partial v^2} l(\hat{v}, x_1, \cdots, x_n) = \frac{n}{2\hat{v}^2} - 2\frac{\sum_{i=1}^n (x_i - m)^2}{2\hat{v}^3} = \frac{n}{2\hat{v}^2} - 2\frac{n\hat{v}}{2\hat{v}^3} = -\frac{n}{2\hat{v}^2} < 0.$$

Donc il s'agit bien d'un maximum. L'EMV de v dans le modèle gaussien à variance connue est donc

$$\hat{V}_{EMV} = \frac{\sum_{i=1}^n (X_i - m)^2}{n}.$$

De plus, c'est bien un estimateur sans biais, car $\mathbb{E}[\hat{v}_{EMV}] = \mathbb{E}[(X-m)^2] = \text{var}(X) = v$. On peut dès lors se demander s'il est efficace, et pour cela, on calcule l'information de Fisher du modèle :

$$I_n(v) = -\mathbb{E}_v\Big[\frac{\partial^2}{\partial v^2} l(v, X_1, \cdots, X_n)\Big] = -\frac{n}{2v^2} + 2\frac{\sum_{i=1}^n (X_i - m)^2}{2v^3} = -\frac{n}{2v^2} + 2\frac{nv}{2v^3} = \frac{n}{2v^2}.$$

On a donc $BCR = \frac{v^2}{n}$.

Il reste encore à calculer la variance de $\hat{V}_{EMV}$, et là, il y a plus de travail qu'à la question précédente. On remarque d'abord quel

$$\begin{aligned}
\text{var}(\hat{V}_{EMV}) &= \text{var}\left(\frac{\sum_{i=1}^n (X_i - m)^2}{n}\right) \\
&= \frac{1}{n^2}\sum_{i=1}^n \text{var}\Big((X_i - m)^2\Big) \text{ en utilisant l'indépendance} \\
&= \frac{\text{var}\Big((X-m)^2\Big)}{n}
\end{aligned}$$

Or, par stabilité des lois normales, $Y := X - m \sim \mathcal{N}(0, v)$. On veut donc calculer $\text{var}\Big((X-m)^2\Big) = \text{var}(Y^2) = \mathbb{E}[Y^4] - \mathbb{E}[Y^2]^2$.

On sait déjà que $\mathbb{E}[Y^2] = \mathbb{E}[(X-m)^2] = \text{var}(X) = v$. Pour le moment d'ordre 4 de Y, on ne peut ici qu'utiliser la densité, et appliquer la formule du transfert, puis utiliser une I.P.P (un peu maline)

$$\begin{aligned}
\mathbb{E}_v[Y^4] &= \int_{\mathbb{R}} y^4 \frac{1}{\sqrt{2\pi v e^{-\frac{y^2}{2v}}}} \mathrm{d}y \\
&= -\frac{\sqrt{v}}{\sqrt{2\pi}} \int_{\mathbb{R}} y^3 \left(-\frac{y}{v} e^{-\frac{y^2}{2v}}\right) \mathrm{d}y \\
&= -\frac{\sqrt{v}}{\sqrt{2\pi}} \int_{\mathbb{R}} y^3 \left(e^{-\frac{y^2}{2v}}\right)' \mathrm{d}y \\
&\overset{I.P.P}{=} -\frac{\sqrt{v}}{\sqrt{2\pi}} \left[e^{-\frac{y^2}{2v}} y^3\right]_{-\infty}^{+\infty} + \frac{\sqrt{v}}{\sqrt{2\pi}} \int_{\mathbb{R}} 3y^2 e^{-\frac{y^2}{2v}} \mathrm{d}y \\
&= 0 + 3v \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi v}} y^2 e^{-\frac{y^2}{2v}} \mathrm{d}y \\
&= 3v\mathbb{E}[Y^2] \\
&= 3v^2
\end{aligned}$$

On peut donc synthétiser :

$$
\begin{aligned}
\mathbb{E}[Y^4] &= 3v^2 \\
\operatorname{var}(Y^2) &= \mathbb{E}[Y^4] - \mathbb{E}[Y^2]^2 \\
&= 3v^2 - v^2 = 2v^2 \\
\operatorname{var}(\hat{V}_{EMV}) &= \frac{\operatorname{var}\Big((X-m)^2\Big)}{n} \\
&= \frac{\operatorname{var}\Big(Y^2\Big)}{n} \\
&= \frac{2v^2}{n} \\
&= BCR.
\end{aligned}
$$

Donc, la borne de Cramèr-Rao étant atteinte, l'estimateur par maximum de vraisemblance $\hat{V}_{EMV}$ est un estimateur sans biais et efficace de v.

Correction exercice 4. Maximum de vraisemblance vs méthode des moments

Soient $X_1, \cdots, X_n$, une suite de v.a. i.i.d. de densité $f_v(x) = C_v x^{v-1} \mathbb{1}_{0 \leq x \leq 1}$, avec $v \geq 1$.

1. On peut utiliser le fait que $\int_{\mathbb{R}} f_v(x)\mathrm{d}x = 1$, ce qui donne :

$$
\begin{aligned}
\frac{1}{C} &= \int_{x \in \mathbb{R}} x^{v-1} \mathbb{1}_{0 \leq x \leq 1} \mathrm{d}x \\
&= \int_0^1 x^{v-1} \mathrm{d}x \\
&= \Big[\frac{x^v}{v}\Big]_0^1 \\
&= \frac{1}{v}
\end{aligned}
$$

Donc $C = v$.

2. On calcule d'abord la log-vraisemblance, et ses dérivées :

$$
\begin{aligned}
l(v, x_1, \cdots, x_n) &= \sum_{i=1}^n \Big(\ln(C_v) + (v-1)\ln(x_i) \Big) \\
&= n\ln(v) + (v-1) \sum_{i=1}^n \ln(x_i) \frac{\partial}{\partial v} l(v, x_1, \cdots, x_n) \quad = \frac{n}{v} + \sum_{i=1}^n \ln(x_i) \frac{\partial}{\partial v^2} l(v, x_1, \cdots, x_n) = -
\end{aligned}
$$

On trouve donc

$$
\frac{\partial}{\partial v} l(\hat{v}, x_1, \cdots, x_n) = 0 \Leftrightarrow \frac{n}{\hat{v}} + \sum_{i=1}^n \ln(x_i) = 0 \Leftrightarrow \hat{v} = \frac{1}{\frac{\sum_{i=1}^n \ln(x_i)}{n}}.
$$

Et il s'agit bien d'un maximum, car $\frac{\partial}{\partial v^2} l(v, x_1, \cdots, x_n) < 0$.

On a donc

$$
\hat{V}_{EMV} = \frac{1}{\frac{\sum_{i=1}^n \ln(X_i)}{n}}
$$

.

3. On a

$$
\begin{aligned}
\mathbb{E}[X] &= \int_0^1 x v x^{v-1} \mathrm{d}x \\
&= v\Big[\frac{x^{v+1}}{v+1}\Big]_0^1 \\
&= \frac{v}{v+1}
\end{aligned}
$$

4. Il suffit d'inverser l'équation précédente :

$$\begin{aligned}
E[X] = \frac{v}{v+1} &\Leftrightarrow (v+1)\mathbb{E}[X] = v \\
&\Leftrightarrow v\Big(\mathbb{E}[X] - 1\Big) = \mathbb{E}[X] \\
&\Leftrightarrow v = \frac{\mathbb{E}[X]}{\mathbb{E}[X] - 1}
\end{aligned}$$

On peut donc construire l'estimateur de v par la méthode des moments, en posant

$$\hat{V}_{EMM} = \frac{\bar{X}}{\bar{X}-1} = \frac{\frac{1}{n}\sum_{i=1}^{n} X_i}{\frac{1}{n}\sum_{i=1}^{n} X_i - 1}.$$

Correction exercice 5. Calcul de l'information de Fisher

1. On utilise la définition, et on pose $s(\theta, x_1, \cdots, x_n) = \frac{\partial}{\partial\theta} l(\theta, x_1, \cdots, x_n)$. On peut alors écrire :

$$\begin{aligned}
\mathbb{E}\Big[s(\theta, X_1, \cdots, X_n)\Big] &= \mathbb{E}\left[\frac{\partial}{\partial\theta}\ln\Big(L(\theta, X_1, \cdots, X_n)\Big)\right] \\
&= \mathbb{E}\left[\frac{\frac{\partial}{\partial\theta}L(\theta, x_1, \cdots, x_n)}{L(\theta, x_1, \cdots, x_n)}\right] \text{ en utilisant la dérivée de } \ln(u) = \frac{u'}{u} \\
&= \int_{x_1, \cdots, x_n} \frac{\frac{\partial}{\partial\theta}L(\theta, x_1, \cdots, x_n)}{L(\theta, x_1, \cdots, x_n)} L(\theta, x_1, \cdots, x_n)\mathrm{d}x_1 \cdots \mathrm{d}x_n \text{ par transfert} \\
&= \int_{x_1, \cdots, x_n} \frac{\partial}{\partial\theta} L(\theta, x_1, \cdots, x_n)\mathrm{d}x_1 \cdots \mathrm{d}x_n \\
&= \frac{\partial}{\partial\theta}\int_{x_1, \cdots, x_n} L(\theta, x_1, \cdots, x_n)\mathrm{d}x_1 \cdots \mathrm{d}x_n \text{ en inversant } \frac{\partial}{\partial\theta} \text{ et } \int \\
&= \frac{\partial}{\partial\theta} 1 \text{ en utilisant que } (x_1, \cdots, x_n) \mapsto L(\theta, x_1, \cdots, x_n) \text{ est une densité} \\
&= 0
\end{aligned}$$

2. Dans le cas i.i.d., on a $l(\theta, x_1, \cdots, x_n) = \sum_{i=1}^{n} \ln\big(f_\theta(x_i)\big)$. On peut calculer

$$\begin{aligned}
I_n(\theta) &= \mathbb{E}_\theta\left[s(\theta, X_1, \cdots, X_n)^2\right] \\
&= \mathrm{var}\Big(s(\theta, X_1, \cdots, X_n)\Big) \text{ car } \mathbb{E}[s(\theta, X_1, \cdots, X_n)] = 0 \\
&= \mathrm{var}\Big(\frac{\partial}{\partial\theta}\sum_{i=1}^{n}\ln\big(f_\theta(X_i)\big)\Big) \\
&= \sum_{i=1}^{n}\mathrm{var}\Big(\frac{\partial}{\partial\theta}\ln\big(f_\theta(X_i)\big)\Big) \text{ en utilisant l'indépendance des } X_i \\
&= n\,\mathrm{var}\Big(\frac{\partial}{\partial\theta}l(\theta, X_1)\Big) \\
&= nI_1(\theta).
\end{aligned}$$

3. On peut alors calculer

$$\begin{aligned}
I_1(\theta) &= \mathbb{E}_\theta\left[\Big(\frac{\partial}{\partial\theta}\ln\big(f_\theta(X_1)\big)\Big)^2\right] \\
&= \mathbb{E}_\theta\left[\Big(\frac{\frac{\partial}{\partial\theta}f_\theta(X_1)}{f_\theta(X_1)}\Big)^2\right]
\end{aligned}$$

Or, comme $\ln(u)' = \frac{u'}{u}$, on a $\ln(u)'' = \frac{u''}{u} - \frac{(u')^2}{u^2}$, donc

$$\Big(\frac{\frac{\partial}{\partial\theta} f_\theta(X_1)}{f_\theta(X_1)}\Big)^2 = \frac{\frac{\partial^2}{\partial\theta^2} f_\theta(X_1))}{f_\theta(X_1))} - \frac{\partial^2}{\partial\theta^2} l(\theta, X_1).$$

Cela donne :

$$\begin{aligned}
I_1(\theta) &= \mathbb{E}_\theta\left[\frac{\frac{\partial^2}{\partial\theta^2} f_\theta(X_1)}{f_\theta(X_1)}\right] - \mathbb{E}_\theta\left[\frac{\partial^2}{\partial\theta^2} l(\theta, X_1)\right] \\
&= \int_{x_1} \frac{\frac{\partial^2}{\partial\theta^2} f_\theta(x_1)}{f_\theta(x_1)} f_\theta(x_1)\mathrm{d}x_1 - \mathbb{E}_\theta\left[\frac{\partial^2}{\partial\theta^2} l(\theta, X_1)\right] \\
&= \int_{x_1} \frac{\partial^2}{\partial\theta^2} f_\theta(x_1)\mathrm{d}x_1 - \mathbb{E}_\theta\left[\frac{\partial^2}{\partial\theta^2} l(\theta, X_1)\right] \\
&= \frac{\partial^2}{\partial\theta^2}\int_{x_1} f_\theta(x_1)\mathrm{d}x_1 - \mathbb{E}_\theta\left[\frac{\partial^2}{\partial\theta^2} l(\theta, X_1)\right] \text{ en inversant } \frac{\partial}{\partial\theta} \text{ et } \int \\
&= \frac{\partial^2}{\partial\theta^2} 1 - \mathbb{E}_\theta\left[\frac{\partial^2}{\partial\theta^2} l(\theta, X_1)\right] \text{ en inversant } \frac{\partial}{\partial\theta} \text{ et } \int \\
&= 0 - \mathbb{E}_\theta\left[\frac{\partial^2}{\partial\theta^2} l(\theta, X_1)\right] \text{ en inversant } \frac{\partial}{\partial\theta} \text{ et } \int
\end{aligned}$$

Cela donne finalement :

$$I_n(\theta) = -n\mathbb{E}_\theta\left[\frac{\partial^2}{\partial\theta^2}\ln(f_\theta(X))\right].$$

Correction exercice 6. Calcul de la Borne de Cramér-Rao

1. Comme T est sans biais, on a
$$\mathbb{E}[T] - \theta = 0,$$
On a alors, en inversant dérivée partielle par rapport à θ, et intégrale sur $x_1, \cdots, x_n$,
$$\begin{aligned}
I &:= \int_{x_1,\cdots,x_n} \Big(T(x_1,\cdots,x_n) - \theta\Big)\frac{\partial}{\partial\theta} L(\theta, x_1, \cdots, x_n)\mathrm{d}x_1\cdots\mathrm{d}x_n \\
&= \frac{\partial}{\partial\theta}\int_{x_1,\cdots,x_n} T(x_1,\cdots,x_n)L(\theta, x_1,\cdots,x_n) - \theta\frac{\partial}{\partial\theta}\int_{x_1,\cdots,x_n} L(\theta, x_1,\cdots,x_n) \\
&= \frac{\partial}{\partial\theta}\mathbb{E}[T] - \frac{\partial}{\partial\theta}1 \\
&= \frac{\partial}{\partial\theta}\theta - 0 \\
&= 1
\end{aligned}$$

2. On en déduit que
$$\begin{aligned}
\mathbb{E}[(T-\theta)\frac{\partial}{\partial\theta} l(\theta, X)] &= \int_{x_1,\cdots,x_n} \Big(T(x_1,\cdots,x_n) - \theta\Big)\Big(\frac{\partial}{\partial\theta} l(\theta, x_1,\cdots,x_n)\Big)L(\theta, x_1,\cdots,x_n)\mathrm{d}x_1\cdots\mathrm{d}x_n \\
&= \int_{x_1,\cdots,x_n} \Big(T(x_1,\cdots,x_n) - \theta\Big)\frac{\frac{\partial}{\partial\theta} L(\theta, x_1,\cdots,x_n)}{L(\theta, x_1,\cdots,x_n)} L(\theta, x_1,\cdots,x_n)\mathrm{d}x_1\cdots\mathrm{d}x_n \\
&= \int_{x_1,\cdots,x_n} \Big(T(x_1,\cdots,x_n) - \theta\Big)\frac{\partial}{\partial\theta} L(\theta, x_1,\cdots,x_n)\mathrm{d}x_1\cdots\mathrm{d}x_n \\
&= 1,
\end{aligned}$$
d'après la question précédente.

3. On pose alors $X = (T-\theta)$ et $Y = \frac{\partial}{\partial\theta} l(\theta, X)$, et on applique l'inégalité de Cauchy-Schwartz :

$$\begin{aligned}
\mathbb{E}\Big[(T-\theta)\frac{\partial}{\partial\theta} l(\theta, X)\Big]^2 &\leq \mathbb{E}\Big[(T-\theta)^2\Big]\mathbb{E}\Bigg[\Big(\frac{\partial}{\partial\theta} l(\theta, X)\Big)^2\Bigg] \\
&\leq \operatorname{var}(T) I_n(\theta)
\end{aligned}$$

Or $\mathbb{E}\Big[(T-\theta)\frac{\partial}{\partial\theta} l(\theta, X)\Big] = 1$, d'après la question précédente, on obtient alors

$$\operatorname{Var}(T) I_n(\theta) \geq 1.$$

En particulier, en définissant une borne $B(\theta) = I_n(\theta)$, on trouve que la variance de n'importe quel estimateur sans biais est toujours minorée par $B(\theta)$. C'est la Borne de Cramér-Rao.

Correction exercice 7. Maximum de vraisemblance et efficacité

Soient $X_1, \cdots, X_n$, une suite de v.a. i.i.d. de densité $f_\theta(x) = C_\theta e^{-\frac{1}{\theta}x} \mathbb{1}_{x \in \mathbb{R}^+}$, avec $\theta \geq 0$.

1. On utilise le fait que $\int_{x\in\mathbb{R}} f_\theta(x)\mathrm{d}x = 1$, et on obtient

$$\begin{aligned}
\frac{1}{C_\theta} &= \int_{\mathbb{R}^+} e^{-\frac{x}{\theta}} \mathrm{d}x \\
&= \Big[-\theta e^{-\frac{x}{\theta}}\Big]_0^{+\infty} \\
&= \theta
\end{aligned}$$

Ce qui donne $C_\theta = \frac{1}{\theta}$ (en réalité, on reconnait une loi exponentielle de paramètre $\frac{1}{\theta}$).

2. On peut écrire la log-vraisemblance et ses dérivées, puis annuler la dérivée pour maximiser la vraisemblance :

$$\begin{aligned}
l(\theta, x_1, \cdots, x_n) &= \sum_{i=1}^n \ln\Big(\frac{1}{\theta} e^{-\frac{x_i}{\theta}}\Big) \\
&= -n\ln(\theta) - \frac{1}{\theta}\sum_{i=1}^n x_i \\
\frac{\partial}{\partial\theta} l(\theta, x_1, \cdots, x_n) &= -\frac{n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^n x_i \\
\frac{\partial^2}{\partial\theta^2} l(\theta, x_1, \cdots, x_n) &= \frac{n}{\theta^2} - \frac{2}{\theta^3}\sum_{i=1}^n x_i
\end{aligned}$$

On peut alors chercher à annuler la dérivée :

$$\begin{aligned}
\frac{\partial}{\partial\theta} l(\hat\theta, x_1, \cdots, x_n) = 0 &\Leftrightarrow -\frac{n}{\hat\theta} + \frac{1}{\hat\theta^2}\sum_{i=1}^n x_i = 0 \\
&\Leftrightarrow \hat\theta = \frac{1}{n}\sum_{i=1}^n x_i
\end{aligned}$$

Pour vérifier que c'est bien un max, on peut évaluer la dérivée seconde en $\hat\theta$:

$$\begin{aligned}
\frac{\partial^2}{\partial\theta^2} l(\hat\theta, x_1, \cdots, x_n) &= \frac{n}{\hat\theta^2} - \frac{2}{\hat\theta^3}\sum_{i=1}^n x_i \\
&= \frac{n}{\hat\theta^2} - \frac{2}{\hat\theta^3} n\hat\theta \\
&= -\frac{n}{\hat\theta^2} < 0.
\end{aligned}$$

Donc l'EMV de θ s'écrit

$$\hat{T}_{EMV} = \frac{1}{n}\sum_{i=1}^n X_i = \bar{X}.$$

3. On a
$$\mathbb{E}[\hat{T}_{EMV}] = \mathbb{E}[\bar{X}] = \mathbb{E}[X] = \theta,$$
(on a déjà calculé en TD de proba que l'espérance d'une exponentielle de paramètre λ vallait $\frac{1}{\lambda}$, donc si $\lambda = \frac{1}{\theta}$, l'espérance vaut θ).

   Donc $\hat{T}_{EMV}$ est un estimateur sans biais de θ.

4. On peut reprendre l'expression de la dérivée seconde, on a alors
$$I_n(\theta) = -\mathbb{E}_\theta\left[\frac{n}{\theta^2} - \frac{2}{\theta^3}\sum_{i=1}^n X_i\right] = -\frac{n}{\theta^2} + \frac{2}{\theta^3}\sum_{i=1}^n \theta = \frac{n}{\theta^2}.$$
On trouve alors la borne de Cramér-Rao du modèle :
$$BCR = \frac{\theta^2}{n}.$$

5. On peut calculer sa variance :
$$\begin{aligned}
\mathrm{var}(\hat{T}_{EMV}) &= \mathrm{var}(\bar{X}) \\
&= \frac{1}{n^2}\sum_{i=1}^n \mathrm{var}(X_i) \\
&= \frac{1}{n^2} n\theta^2 \text{ on a déjà fait le calcul de la variance en TD de proba.} \\
&= \frac{\theta^2}{n} = BCR
\end{aligned}$$
Donc $\hat{T}_{EMV}$ est un estimateur efficace de θ, car il est sans biais et sa variance atteint la borne de Cramér-Rao.

Correction exercice 8. Methode des moments pour la loi Gamma

On définit la densité d'une loi Gamma de paramètres (k, β) (on note $X \sim \mathrm{Gamma}(k, \beta)$) par
$$f_{k,\beta}(x) = \frac{\beta^k x^{k-1}}{\Gamma(k)} e^{-\beta x} \mathbb{1}_{x \in \mathbb{R}^+},$$
où la fonction Γ désigne la fonction Gamma d'Euler définie par
$$\Gamma(z) = \int_{t \geq 0} t^{z-1} e^{-t} \mathrm{d}t.$$

1. On peut faire une I.P.P. :
$$\begin{aligned}
\Gamma(z+1) &= \int_{t \geq 0} t^z e^{-t} \mathrm{d}t \\
&\overset{I.P.P.}{=} \Big[-t^z e^{-t} \Big]_0^{+\infty} - \int_{t \geq 0} z t^{z-1}(-e^{-t}) \mathrm{d}t \\
&\overset{I.P.P.}{=} 0 + z\Gamma(z)
\end{aligned}$$

2. Soit $p \in \mathbb{N}$, on peut calculer :
$$\begin{aligned}
\mathbb{E}[X^p] &= \int_{x \geq 0} x^p \frac{\beta^k x^{k-1}}{\Gamma(k)} e^{-\beta x} \mathrm{d}x \\
&= \frac{\beta^k}{\Gamma(k)} \int x \geq 0 x^{p+k-1} e^{-\beta x} \mathrm{d}x \\
&= \frac{\beta^k}{\Gamma(k)} \int u \geq 0 \frac{u}{\beta}^{p+k-1} e^{-u} \frac{\mathrm{d}u}{\beta} \text{ en posant (chgt de var) } u = \beta x \\
&= \frac{\beta^k}{\beta^{k+p}\Gamma(k)} \int u \geq 0 u^{p+k-1} e^{-u} \mathrm{d}u \\
&= \frac{\Gamma(k+p)}{\beta^p \Gamma(k)} \text{ par définition de la fonction } \Gamma.
\end{aligned}$$

3. On peut alors calculer, en utilisant l'expression précédente, pour $p = 1, 2$

$$\begin{aligned}
\mathbb{E}[X] &= \frac{\Gamma(k+1)}{\beta\Gamma(k)} \\
&= \frac{k\Gamma(k)}{\beta\Gamma(k)} \text{ en utilisant la question 1} \\
&= \frac{k}{\beta} \\
\mathbb{E}[X^2] &= \frac{\Gamma(k+2)}{\beta^2\Gamma(k)} \\
&= \frac{(k+1)k\Gamma(k)}{\beta^2\Gamma(k)} \\
&= \frac{(k+1)k}{\beta^2} \\
\mathrm{var}(X) &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 \\
&= \frac{(k+1)k}{\beta^2} - \frac{k^2}{\beta^2} \\
&= \frac{k}{\beta^2}.
\end{aligned}$$

4. On a

$$\begin{aligned}
\mathbb{E}[X] &= \frac{k}{\beta} \\
\mathrm{var}(X) &= \frac{k}{\beta^2}
\end{aligned}$$

On peut alors inverser ses relations :

$$\begin{aligned}
k &= \frac{\mathbb{E}[X]^2}{\mathrm{var}(X)} \\
\beta &= \frac{\mathbb{E}[X]}{\mathrm{var}(X)}
\end{aligned}$$

On peut alors estimer k et β par la méthode des moments :

$$\begin{aligned}
\hat{K} &= \frac{\bar{X}^2}{\hat{S}_n^2} \\
\hat{B} &= \frac{\bar{X}}{\hat{S}_n^2}
\end{aligned}$$