

Examen Final – Janvier 2019

Durée 2h00 – Sans documents

Important. Suivant les réglements en vigueur,

1. les enseignants présents lors de l'épreuve ne peuvent communiquer que sur les fautes d'énoncé potentielles. Toute autre question durant la composition ne sera pas acceptée.
2. les étudiants sont tenus de se lever au moment de l'annonce de fin de la composition. En cas de refus, le responsable de l'UE sera fondé à ne pas prendre en compte la copie incriminée.
3. l'identification des copies et intercalaires doit se faire au moment de la remise de chaque copie par les enseignants et surveillants. Il ne sera pas accordé de délai pour cette raison en fin d'épreuve.

Formulaire

Loi	Notation	Densité
Beta	$\mathcal{B}eta(a,b)$	$f_{a,b}(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}\mathbb{1}_{x\in]0,1[}$
Gamma	$\gamma(a,b)$	$f_{a,b}(x) = \frac{1}{\Gamma(a)} x^{a-1}e^{-bx}b^a\mathbb{1}_{x>0}$
Exponentielle	$\mathcal{E}(\lambda)$	$f_\lambda(x) = \lambda e^{-\lambda x}\mathbb{1}_{x>0}$
Normale	$\mathcal{N}(\mu,\sigma^2)$	$f_{\mu,\sigma^2}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)}$
Poisson	$\mathcal{P}oi(\lambda)$	$f_\lambda(x) = \frac{e^{-\lambda}}{x!}\lambda^x\mathbb{1}_{x\in\mathbb{N}}$

French – English Lexicon

- échantillon – sample
- famille exponentielle – exponential family
- statistique complète – complete statistic
- statistique exhaustive – sufficient statistic
- statistique libre – ancilary statistic
- vraisemblance – likelihood

Exercice 1 (QCM – 4 pts). **Donner la bonne réponse en la justifiant. Seules les réponses justifiées seront validées. Il n'y a pas de points négatifs.**

1. *(1 pt)* Soit $X_1, \ldots, X_n$ un échantillon *i.i.d.* de loi $\mathcal{P}oi(\lambda)$. La forme générique de la loi conjuguée *a posteriori* est

(a) $\lambda \sim \gamma(a+n, b+\sum_{i=1}^n x_i)$
(b) $\lambda \sim \mathcal{P}oi(a+\sum_{i=1}^n x_i)$
(c) $\lambda \sim \mathcal{N}((a+\sum_{i=1}^n x_i)/(b+n), bn/b+n)$
(d) $\lambda \sim \gamma(a+\sum_{i=1}^n x_i, b+n)$

2. *(1 pt)* Soit $X_1, \ldots, X_n$ un échantillon *i.i.d.* de loi $\mathcal{E}(\lambda)$. On observe $(Z_1, \ldots, Z_n) = (\mathbb{1}(X_1 \leq \delta), \ldots, \mathbb{1}(X_n \leq \delta))$, avec δ connu. Si $z_1 = 1$ et $(z_2, \ldots, z_n) = (0, \ldots, 0)$, l'estimateur du maximum de vraisemblance de λ est donné par $\hat{\lambda} = \ldots$

(a) $-\log\left(\frac{n-1}{n}\right)/\delta$ **(b)** $\frac{(n-1)\delta}{n}$ **(c)** $\frac{\log(\delta)}{n-1}$ **(d)** $\frac{n+1}{\delta}$ **(e)** 0

3. *(1 pt)* Soit $X_1, \ldots, X_n$ un échantillon *i.i.d.* tel que $\mathbb{P}(X_i = 0) = p$ et $\mathbb{P}(X_i > x) = (1-p)e^{-\lambda x}$, $(x > 0)$. Une statistique exhaustive pour ce modèle est

(a) $\bar{X}_n$

(b) $\sum_{i=1}^n X_i / \sum_{i=1}^n \mathbb{1}_{X_i>0}$

(c) $(\sum_{i=1}^n \mathbb{1}_0(X_i), \sum_{i=1}^n \mathbb{1}_{X_i>0} X_i)$

(d) $(\sum_{i=1}^n \mathbb{1}_0(X_i), \sum_{i=1}^n \mathbb{1}_{X_i>0})$

4. *(1 pt)* Soit X de loi normale $\mathcal{N}(\mu, 1/v)$ avec μ inconnu et v connu. On considère sur μ une loi *a priori* $\mathcal{N}(m, 1/p)$. Lequels des codes ci-dessous permet d'évaluer la densité *a posteriori* $\pi(\mu \mid x)$ au point $\mu = 0.3$?

(a)
```
pnorm(0.3, (m*p^2 + x*v^2)/(x + v), sqrt((p*v)^2/(p^2+v^2)))
```

(b)
```
dnorm(0.3, (m*p + x*v)/(p+v), sqrt(1/(p+v)))
```

(c)
```
pnorm(0.3, (m*p + x*v)/(p+v), sqrt(1/(p+v)))
```

(d)
```
dnorm(0.3, (m*v + x*p)/(p+v), sqrt(1/(p+v)))
```

Solution Exercice 1.

1. La forme générique de la loi conjuguée *a posteriori* est

$$\textbf{(d)} \quad \lambda \sim \gamma(a + \sum_{i=1}^n x_i, b + n),$$

car la loi de Poisson a une densité proportionnelle à $\lambda^{x_i} \exp\{-\lambda\}$, qui se conjugue avec $\lambda^a \exp\{-b\lambda\}$.

2. Sachant que $\mathbb{P}(Z_1 = 1) = (1 - \exp\{-\lambda\delta\})$, la vraisemblance de l'échantillon est

$$L(\lambda \,|, 1, 0, \ldots, 0, \delta) = \exp\{-(n-1)\lambda\delta\}(1 - \exp\{-\lambda\delta\}).$$

L'estimateur du maximum de vraisemblance de λ est donc

$$\textbf{(a)} \quad -\log\left(\frac{n-1}{n}\right)/\delta.$$

3. La vraisemblance s'écrit

$$\begin{aligned} L(p \mid x_1, \ldots, x_n) &= \prod_{i=1}^n p^{\mathbb{1}_0(x_i)} \{(1-p)e^{-\lambda x_i}\}^{\mathbb{1}_{x_i>0}} \\ &= p^{\sum_{i=1}^n \mathbb{1}_0(x_i)} \times (1-p)^{\sum_{i=1}^n (1-\mathbb{1}_0(x_i))} \times e^{-\lambda \sum_{i=1}^n \mathbb{1}_{x_i>0} x_i}. \end{aligned}$$

Une statistique exhaustive pour ce modèle est donc

$$\textbf{(c)} \quad \left(\sum_{i=1}^n \mathbb{1}_0(X_i), \sum_{i=1}^n \mathbb{1}_{X_i>0} X_i\right)$$

4. Il s'agit de loi *a priori* conjuguée pour la densité normale. La loi *a posteriori* est donc une loi normale dont la moyenne est une combinaison convexe de l'observation et de la moyenne *a priori* (chacune étant multipliée par sa précision) et dont la précision et la somme des précisions du modèle et de la loi *a priori*, *i.e.*,

$$\pi(\mu \mid x) \sim \mathcal{N}\left(\frac{xv + mp}{v+p}, \frac{1}{v+p}\right).$$

Par ailleurs, la commande pour évaluer la densité d'une loi normale est `dnorm()`. La bonne réponse est donc la réponse **(b)**.

Exercice 2 (5 pts). Soit $X_1, \dots, X_n$, $n > 1$, un échantillon *i.i.d.* de loi uniforme sur $[0, \theta]$. Dans cet exercice, on cherche à construire un estimateur de $\lambda = 1/\theta$. Les statistiques d'ordre notées $X_{(1)}, \dots, X_{(n)}$ sont obtenues en faisant un tri croissant de l'échantillon, en particulier $X_{(1)} = \min\{X_1, \dots, X_n\}$ et $X_{(n)} = \max\{X_1, \dots, X_n\}$. Si l'on note F la fonction de répartition de X et f sa densité, on rappelle que la densité de la k-ème statistique d'ordre, $k \in \{1, \dots, n\}$, est

$$f_k(x) = \frac{n!}{(k-1)!(n-k)!} F(x)^{k-1}\{1 - F(x)\}^{n-k} f(x).$$

1. (a) *(1 pt)* Montrer qu'à l'exception de $X_{(1)}$, toutes les statistiques d'ordre admettent un moment inverse d'ordre 1 défini par $\mathbb{E}\left[1/X_{(k)}\right]$ et les calculer.

(b) *(0.5 pt)* En déduire que pour tout $k \geq 2$, $\hat{\lambda}_{n,k} = (k-1)/(nX_{(k)})$ est un estimateur sans biais de λ.

2. *(0.5 pt)* Donner une statistique exhaustive pour l'échantillon $X_1, \dots, X_n$.

3. *(1.5 pt)* Montrer que $\hat{\lambda}_{n,n}$ est l'estimateur de variance minimale uniformément parmi tous les estimateurs sans biais.

4. *(1.5 pt)* Justifier que pour tout $k \geq 2$, l'amélioration de l'estimateur $\hat{\lambda}_{k,n}$ par le théorème de Rao-Blackwell conduit à l'estimateur $\hat{\lambda}_{n,n}$.

Solution Exercice 2. La densité de la k-ème statistique d'ordre, $k \in \{1, \dots, n\}$, est

$$f_k(x) = \frac{n!}{(k-1)!(n-k)!} \frac{x^{k-1}(\theta - x)^{n-k}}{\theta^n} \mathbb{1}_{x \in [0,\theta]}.$$

1. (a) La fonction $x \mapsto x^{k-2}(\theta - x)^{n-k}$, $k \in \{1, \dots, n\}$, est intégrable sur en 0 si et seulement si $2 - k < 1$, *i.e.*, $k > 1$. Donc toutes les statistiques d'ordre $k > 1$ admettent un moment inverse d'ordre 1. Pour $k > 1$, en considérant le changement de variable $\phi : u \mapsto u/\theta$ qui réalise un $\mathcal{C}^1$-difféomorphisme de $[0, \theta]$ dans $[0, 1]$, on obtient

$$\begin{aligned}
\mathbb{E}\left[1/X_{(k)}\right] &= \frac{n!}{(k-1)!(n-k)!\theta} \int_0^1 u^{k-2}(1-u)^{n-k} \mathrm{d}u \\
&= \frac{n!}{(k-1)!(n-k)!\theta} \frac{\Gamma(k-1)\Gamma(n-k+1)}{\Gamma(n)} \\
&= \frac{n!}{(k-1)!(n-k)!\theta} \frac{(k-2)!(n-k)!}{(n-1)!} = \frac{n}{(k-1)\theta}.
\end{aligned}$$

(b) Pour tout $k \geq 2$, on déduit de la question précédente que

$$\mathbb{E}\left[\hat{\lambda}_{n,k}\right] = \frac{k-1}{n} \times \mathbb{E}\left[1/X_{(k)}\right] = \frac{1}{\theta} = \lambda.$$

L'estimateur est donc sans biais.

2. Pour un échantillon *i.i.d.* suivant la loi uniforme sur $[0,\theta]$, la vraisemblance s'écrit

$$L(\theta \mid X_1, \ldots, X_n) = \frac{1}{\theta^n} \prod_{i=1}^{n} \mathbb{1}_{X_i \in [0,\theta]} = \frac{1}{\theta^n} \mathbb{1}_{X_{(n)} \leq \theta}.$$

Le théorème de factorisation permet de conclure que la statistique d'ordre $X_{(n)}$ est une statistique exhaustive.

3. Ce résultat est une application du théorème de Lehmann-Scheffé. $\hat{\lambda}_{n,n}$ est un estimateur sans biais (Question 1.b). Par ailleurs la statistique $X_{(n)}$ est exhausitve (Question 2). Il reste à montrer que cette statistique est complète. Soit h une fonction mesurable, tel que pour tout $\theta > 0$, $\mathbb{E}_\theta\left[h\left(X_{(n)}\right)\right] = 0$, ce qui équivaut à :

$$\int_0^\theta h(x)x^{n-1}\mathrm{d}x = 0, \quad \text{pour tout } \theta > 0.$$

Supposons h continue. En dérivant l'égalité précédente par rapport à θ, on obtient

$$h(\theta)\theta^{n-1} = 0, \quad \text{pour tout } \theta > 0,$$

et donc h est la fonction nulle. Si la fonction h n'est pas continue, le résultat précédent reste vrai presque-sûrement. Il suffit d'exclure les points de discontinuité de h qui est $\mathbb{P}_\theta$ négligable. La statistique $X_{(n)}$ est donc complète et le théorème de Lehmann-Scheffé permet de conclure que $\mathbb{E}\left[\hat{\lambda}_{n,n} \mid X_{(n)}\right]$ est l'estimateur de variance minimale uniformément parmi tous les estimateurs sans biais. Or comme $\hat{\lambda}_{n,n}$ s'exprime comme une fonction mesurable de $X_{(n)}$, on a $\mathbb{E}\left[\hat{\lambda}_{n,n} \mid X_{(n)}\right] = \hat{\lambda}_{n,n}$.

4. Pour tout $k \geq 2$, $\hat{\lambda}_{k,n}$ est un estimateur sans biais de λ et l'amélioration cet estimateur par le théorème de Rao-Blackwell pour la statistique exhaustive $X_{(n)}$ est $\mathbb{E}\left[\hat{\lambda}_{k,n} \mid X_{(n)}\right]$. Or d'après la question précédente, cette statistique est également complète et donc $\mathbb{E}\left[\hat{\lambda}_{k,n} \mid X_{(n)}\right]$ est l'unique estimateur de variance minimale uniformément parmi tous les estimateurs sans biais, c'est à dire $\hat{\lambda}_{n,n}$. Ce qui constitue le résultat souhaité.

Exercice 3 (6.5 pts). Dans une classe, il y a $n = X_1$ étudiants. On note X_t le nombre d'étudiants présents la semaine t, sachant que chaque semaine un étudiant décide, indépendament des autres et du passé, de revenir la semaine suivante avec probabilité p. Lorsqu'un étudiant a décidé de ne pas venir, il ne reviendra plus jamais. $X_t \mid X_{t-1} = x_{t-1}$ suit une loi binomiale $\mathcal{B}(x_{t-1}, p)$, dont la densité est notée par la suite $f(x_t \mid x_{t-1}, p)$. Notons que les X_t ne sont pas indépendants.

Pour chaque étudiant $i \in \{1, \ldots, n\}$, le nombre T_i de cours auquel il aura assisté avant de ne plus venir suit une loi géométrique de paramètre $1-p$, c'est à dire que $P(T_i = k) = (1-p)p^{k-1}$, pour

$k > 0$. On a alors $E(T_1) = 1/(1-p)$ et $Var(T_1) = p/(1-p)^2$.

1. *(0.5 pt)* Montrer que la loi géométrique est dans la famille exponentielle. En déduire S une statistique exhaustive de dimension 1 quelque soit la taille de l'échantillon.

2. *(1 pt)* On admet que la vraisemblance s'écrit $L(p \mid x_1, \ldots, x_{s_0}) = \prod_{t=2}^{s_0} f(x_t \mid x_{t-1}, p)$, avec $s_0 = \inf\{t \mid X_t = 0\}$. Exprimer cette vraisemblance en fonction des T_i et montrer que l'estimateur du maximum de vraisemblance est donné par

$$\hat{p}_n = 1 - \frac{n}{\sum_{i=1}^n T_i}.$$

3. **(a)** *(0.5 pt)* Montrer que cet estimateur est biaisé.

(b) *(1 pt)* Écrire une procédure R qui estime le biais de cet estimateur.

4. *(1.5 pt)* Donner un intervalle de confiance asymptotique pour p au niveau de confiance $1 - \alpha$.

5. On mène une inférence bayésienne sur le paramètre p.

(a) *(1 pt)* Montrer que pour n'importe quel choix de loi *a priori* sur p, la loi *a posteriori* de p sachant $x_1, \ldots, x_{s_0}$ est égale à la loi *a posteriori* de p sachant S.

(b) *(0.5 pt)* On met une loi *a priori* $\mathcal{B}eta(\alpha, \beta)$ sur p. Quelle est la loi *a posteriori* de p ?

(c) *(0.5 pt)* Montrer que la moyenne de la loi *a posteriori* $\tilde{p}_n$ et $\hat{p}_n$ sont équivalents lorsque $n \to \infty$.

Solution Exercice 3.

1. La densité de la loi géométrique par rapport à la mesure de comptage est

$$f(y \mid p) = (1-p)p^{y-1}\mathbb{1}_{y\in\mathbb{N}} = \frac{1-p}{p} \times \mathbb{1}_{y\in\mathbb{N}} \times \exp(y \times \ln p).$$

La loi géométrique appartient donc à la famille exponentielle avec pour paramètre naturel $\eta(p) = \ln p$ et pour statistique naturelle $S(y) = y$. On en déduit que pour un n-échantillon $T_1, \ldots, T_n$ suivant la loi géométrique, $S(T_1, \ldots, T_n) = \sum_{i=1}^n T_i$ est une statistique exhaustive.

2. La vraisemblance s'écrit

$$\begin{aligned} L(p \mid x_1, \ldots, x_{s_0}) &= \prod_{t=2}^{s_0} \binom{x_{t-1}}{x_t} p^{x_t}(1-p)^{x_{t-1}-x_t} \\ &= p^{\sum_{t=2}^{s_0} x_t}(1-p)^n \prod_{t=2}^{s_0} \binom{x_{t-1}}{x_t}. \end{aligned}$$

Par ailleurs, $\sum_{t=1}^{s_0} x_t$ représente le total cumulé d'étudiants venus en cours, autrement dit

$$\sum_{t=1}^{s_0} x_t = \sum_{i=1}^{n} T_i.$$

Ainsi, sachant que $x_1 = n$, la vraisemblance peut s'écrire

$$L(p \mid x_1, \ldots, x_{s_0}) = p^{\sum_{i=1}^n T_i - n}(1-p)^n \prod_{t=2}^{s_0} \binom{x_{t-1}}{x_t}.$$

La log-vraisemblance est dérivable sur $]0,1[$ et

$$\frac{\mathrm{d}}{\mathrm{d}p}\ln L(p \mid x_1,\dots,x_{s_0}) = \frac{1}{p}\left(\sum_{i=1}^n T_i - n\right) - \frac{n}{1-p} = \frac{1}{p(1-p)}\left[(1-p)\sum_{i=1}^n T_i - n\right].$$

Elle admet donc un unique point critique donné par

$$\hat{p}_n = 1 - \frac{n}{\sum_{i=1}^n T_i}.$$

De plus, pour tout $p \in]0,1[$, ${\mathrm{d}^2}/{\mathrm{d}p^2} \ln L(p \mid x_1,\dots,x_{s_0}) < 0$, la log-vraisemblance est donc concave et le point critique est un maximum. Donc $\hat{p}_n$ est bien l'estimateur du maximum de vraisemblance pour p.

3. **(a)** La fonction $t \mapsto 1/t$ étant **strictement convexe** sur $\mathbb{R}_+^*$, l'inégalité de Jensen donne

$$\mathbb{E}\left[\hat{p}_n\right] = 1 - \mathbb{E}\left[\frac{n}{\sum_{i=1}^n T_i}\right] < 1 - \frac{n}{\mathbb{E}\left[\sum_{i=1}^n T_i\right]} = 1 - \frac{1}{\mathbb{E}\left[T_1\right]} = p.$$

Donc l'estimateur $\hat{p}_n$ est biaisé.

(b) On considère que le n-échantillon $T_1,\dots T_n$ est contenu dans un vecteur `y` et que la taille de l'échantillon bootstrap est `N`.

```
p.hat <- mean(y)
p.boot <- numeric(N)
for (k in 1: N) {
  p.boot[i] <- 1 - 1/mean(sample(y, replace = T))
}
mean(p.boot) - p.hat
```

4. $T_1,\dots,T_n$ étant une suite de variables aléatoires *i.i.d.* de carré intégrable, le théorème central-limite donne

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n T_i - \frac{1}{1-p}\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(0, \frac{p}{(1-p)^2}\right)$$

La fonction $g : x \mapsto 1 - 1/x$ est dérivable et pour $p \in]0,1[$, $g'(1/(1-p)) = (1-p)^2 \neq 0$. La méthode delta conduit à

$$\sqrt{n}\left(\hat{p}_n - p\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(0, p(1-p)^2\right).$$

Par ailleurs, en utilisant la loi de forte des grands nombre et la continuité de la fonction $x \mapsto 1/x$ sur $]0,1[$, on a

$$\hat{p}_n(1-\hat{p}_n)^2 \xrightarrow[n\to+\infty]{\mathbb{P}} p(1-p)^2.$$

Le lemme de Slutsky permet alors de conclure que

$$\frac{\sqrt{n}}{\sqrt{\hat{p}_n}(1-\hat{p}_n)}\left(\hat{p}_n - p\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(0,1\right).$$

On en déduit alors l'intervalle de confiance au niveau $1-\alpha$

$$\left[\hat{p}_n - q_{1-\alpha/2}(1-\hat{p}_n)\sqrt{\frac{\hat{p}_n}{n}}\,,\, \hat{p}_n + q_{1-\alpha/2}(1-\hat{p}_n)\sqrt{\frac{\hat{p}_n}{n}}\right].$$

5. **(a)** Soit π un loi *a priori* sur p. La statistique S étant exhaustive, il existe deux fonctions mesurables h et g telles que $L(p \mid x_1, \ldots, x_{s_0}) = h(x_1, \ldots, x_{s_0})g(p, S(T_1, \ldots, T_n))$. La loi *a posteriori* s'écrit

$$\begin{aligned}\pi(p \mid x_1, \ldots, x_{s_0}) &= \frac{L(p \mid x_1, \ldots, x_{s_0})\pi(p)}{\int L(p \mid x_1, \ldots, x_{s_0})\pi(p)\mathrm{d}p} \\ &= \frac{\cancel{h(x_1, \ldots, x_{s_0})}}{\cancel{h(x_1, \ldots, x_{s_0})}} \frac{g(p, S(T_1, \ldots, T_n)\pi(p)}{\int g(p, S(T_1, \ldots, T_n)\pi(p)\mathrm{d}p}.\end{aligned}$$

La loi *a posteriori* ne dépend donc que de la statistique $S(T_1, \ldots, T_n)$.

(b) Pour une loi *a priori* $\mathcal{B}eta(\alpha, \beta)$, on a

$$\pi(p \mid S(T_1, \ldots, T_n)) \propto p^{S(T_1, \ldots, T_n) - n + \alpha - 1}(1-p)^{n+\beta-1}.$$

La loi *a posteriori* est donc une loi $\mathcal{B}eta(\alpha + S(T_1, \ldots, T_n) - n, \beta + n)$.

(c) La moyenne de la loi *a posteriori* est

$$\tilde{p}_n = \frac{\alpha + S(T_1, \ldots, T_n) - n}{\alpha + S(T_1, \ldots, T_n) + \beta} = \frac{\alpha(1 - \hat{p}_n)/n + \hat{p}_n}{(\alpha + \beta)(1 - \hat{p}_n)/n + 1}.$$

On obtient alors directement

$$\tilde{p}_n \underset{+\infty}{\sim} \hat{p}_n \quad \text{et} \quad \tilde{p}_n \underset{n \to +\infty}{\longrightarrow} p.$$

Exercice 4 (4.5 pts). On souhaite calibrer un appareil de voltage connu pour être d'autant moins précis que la mesure à prendre est grande. On suppose que n mesures de voltage $X_1, \ldots, X_n$ forment un échantillon *i.i.d.* de loi $\mathcal{N}(m, \theta m^2)$, avec $m \in \mathbb{R}$ connu et $\theta > 0$ inconnu.

On pourra utiliser sans démonstration les résultats suivants :

- Si Y suit une loi normale $\mathcal{N}(0, \sigma^2)$, alors Y^2 suit une loi gamma $\gamma(\frac{1}{2}, \frac{1}{2\sigma^2})$.
- Si Y et Z sont des variables aléatoires indépendantes de loi respectives $\gamma(\alpha, \lambda)$ et $\gamma(\beta, \lambda)$, alors $Y + Z$ est de loi $\gamma(\alpha + \beta, \lambda)$.
- Pour tout $a > 0$, $\Gamma(a + 1) = a\Gamma(a)$.

1. *(1 pt)* Donner la vraisemblance $L(\theta \mid x_1, \ldots, x_n, m)$. En déduire $\hat{\theta}_n$ l'estimateur du maximum de vraisemblance.

2. *(1 pt)* Donner la loi de $\hat{\theta}_n$. En déduire que $\hat{\theta}_n$ est sans biais.

3. **(a)** *(1 pt)* Calculer l'information de Fisher $I_X(\theta)$. En déduire la borne de Cramer-Rao.

(b) *(0.5 pt)* Conclure que $\hat{\theta}_n$ est l'estimateur de variance minimale uniformément parmi les estimateurs sans biais.

4. On considère une loi a priori sur θ dont la densité est définie par :

$$\pi(\theta) = \frac{b^a}{\Gamma(a)} \left(\frac{1}{\theta}\right)^{a+1} e^{-\frac{b}{\theta}} \mathbb{1}_{\theta > 0}, \quad \text{avec } a > 0 \text{ et } b > 0 \text{ fixés.}$$

Cette loi est appelée loi inverse-gamma de paramètre (a, b).

(a) *(0.5 pt)* Donner la loi *a posteriori* de θ sachant $x_1, \ldots, x_n$.

(b) *(0.5 pt)* Calculer l'estimateur de $\theta_{\text{MAP}} = \arg\max_\theta \pi(\theta \mid x_1, \ldots, x_n)$.

Solution Exercice 4.

1. La vraisemblance s'écrit

$$L(\theta \mid x_1, \ldots, x_n, m) = \frac{1}{(2\pi m^2\theta)^{n/2}} \prod_{i=1}^n \exp\left(-\frac{\{x_i - m\}^2}{2\theta m^2}\right)$$
$$= \frac{1}{(2\pi m^2\theta)^{n/2}} \exp\left(-\frac{1}{2\theta m^2}\sum_{i=1}^n \{x_i - m\}^2\right).$$

La log-vraisemblance est dérivable sur $\mathbb{R}_+^*$ et

$$\frac{\mathrm{d}}{\mathrm{d}\theta} \ln L(\theta \mid x_1, \ldots, x_n, m) = -\frac{n}{2\theta} + \frac{1}{2\theta^2 m^2}\sum_{i=1}^n (x_i - m)^2$$
$$= \frac{1}{2\theta^2}\left(-n\theta + \frac{1}{m^2}\sum_{i=1}^n (x_i - m)^2\right).$$

Elle admet un unique point critique donné par

$$\hat{\theta}_n = \frac{1}{nm^2}\sum_{i=1}^n (x_i - m)^2.$$

De plus, la dérivée de la log-vraisemblance est positive sur $]0, \hat{\theta}_n]$ et négative sur $]\hat{\theta}_n, +\infty[$. Ce point critique est donc un maximum. On en conlcut que $\hat{\theta}_n$ est l'estimateur du maximum de vraisemblance de θ.

2. D'après l'indication de l'énoncé, pour tout $i \in \{1, \ldots, n\}$,

$$\frac{X_i - m}{m\sqrt{n}} \sim \mathcal{N}\left(0, \frac{\theta}{n}\right) \quad \Rightarrow \quad \frac{(X_i - m)^2}{m^2 n} \sim \gamma\left(\frac{1}{2}, \frac{n}{2\theta}\right).$$

On en déduit

$$\hat{\theta}_n \sim \gamma\left(\frac{n}{2}, \frac{n}{2\theta}\right).$$

Comme l'espérance d'une loi $\gamma(\alpha, \beta)$ est α/β, on en déduit que l'estimateur est sans biais :

$$\mathbb{E}\left[\hat{\theta}_n\right] = \frac{n/2}{n/2\theta} = \theta.$$

3. (a) Si l'on note $f(\cdot \mid m, \theta)$ la densité de la loi $\mathcal{N}(m, \theta m^2)$, alors la dérivé seconde de $\ln f(\cdot \mid m, \theta)$ est intégrable sur $]0, +\infty[$ et

$$I_X(\theta) = \mathbb{E}_\theta\left[-\frac{\mathrm{d}^2}{\mathrm{d}\theta^2} \ln f(x \mid m, \theta)\right] = \mathbb{E}_\theta\left[-\frac{1}{2\theta^2} + \frac{1}{\theta^3 m^2}(X - m)^2\right]$$
$$= -\frac{1}{2\theta^2} + \frac{1}{\theta^3 m^2}\mathbb{V}\mathrm{ar}\left[X - m\right] = \frac{1}{2\theta^2}.$$

La borne de Cramer-Rao est donc $2\theta^2/n$.

(b) La variance d'une loi gamma $\gamma(\alpha, \beta)$ est α/β^2. On en déduit que la variance de l'estimateur $\hat{\theta}_n$ est

$$\mathbb{V}\mathrm{ar}\left[\hat{\theta}_n\right] = \frac{2\theta^2}{n}.$$

Par suite, $\hat{\theta}_n$ est un estimateur sans biais dont la variance atteint la borne de Cramer-Rao. C'est l'estimateur de variance minimale uniformément parmi les estimateurs sans biais.

4. (a) Pour la loi *a posteriori* de θ sachant $x_1, \dots, x_n$, on a

$$\begin{aligned}\pi(\theta \mid x_1, \dots, x_n) &\propto \pi(\theta) L(\theta \mid x_1, \dots, x_n, m) \\ &\propto \left(\frac{1}{\theta}\right)^{a+1+n/2} \exp\left(-\frac{b}{\theta} - \frac{1}{2\theta m^2}\sum_{i=1}^n \{x_i - m\}^2\right) \mathbb{1}_{\theta>0}.\end{aligned}$$

On en déduit que la loi *a posteriori* est une loi inverse-gamma de paramètre

$$\left(a + \frac{n}{2}, b + \frac{1}{2m^2}\sum_{i=1}^n \{x_i - m\}^2\right).$$

(b) Le logarithme de la loi *a posteriori* de θ sachant $x_1, \dots, x_n$ est dérivable sur $\mathbb{R}_+^*$ et

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\theta} \ln \pi(\theta \mid x_1, \dots, x_n) &= -\frac{a+1+n/2}{\theta} + \left(b + \frac{1}{2m^2}\sum_{i=1}^n \{x_i - m\}^2\right)\frac{1}{\theta^2} \\ &= \frac{1}{\theta^2}\left(-\left\{a + 1 + \frac{n}{2}\right\}\theta + b + \frac{1}{2m^2}\sum_{i=1}^n \{x_i - m\}^2\right).\end{aligned}$$

On en déduit qu'il existe un unique point critique donné par

$$\theta_{\mathrm{MAP}} = \frac{b + \frac{1}{2m^2}\sum_{i=1}^n \{x_i - m\}^2}{a + 1 + n/2}.$$

La fonction étant concave (signe de la dérivée seconde ou tableau de variations), on en déduit que ce point critique est un maximum.

Here are the most glaring mistakes in the exam sheets :

- When trying to show that $\hat{p}$ is biased, using the tempting equality $\mathbb{E}[1/\sum_i T_i] = 1/\mathbb{E}[\sum_i T_i]$ leads to unbiasedness
- for unbiasedness questions, being asymptotically unbiased is *not the same as* being unbiased.
- in Exercise 1, for the mixed random variable with density $p\mathbb{I}_0(x) + (1-p)\exp(-\lambda x)\mathbb{I}(x > 0)$, the likelihood is not

$$\prod_I \{p\mathbb{I}_0(x_i) + (1-p)\exp(-\lambda x)\mathbb{I}(x_i > 0)\}$$

(with an unspecified x).

- when establishing completeness, $\int_0^\theta f(x)x^k \mathrm{d}x = 0$ does not obviously imply $f = 0$ (but requires a derivation in θ).
- Rao-Blackwell does not imply that $\mathbb{E}[\hat\lambda_{kn}|\hat\lambda_{nn}] = \hat\lambda_{nn}$, the identify is a consequence of the completeness of $\hat\lambda_{nn}$.
- when computing the maximum likelihood estimator, not using the logarithm makes life complicated, and so does looking at the derivative of $\log(\sqrt{2\pi\theta m}^n)$. Developing $\sum_i (x_i - m)^2$ does not help in the least.
- in Exercise 1, if one observes $Z_i = \mathbb{I}(X_i > \delta)$ writing the likelihood for the X_i's does not help.
- reverse engineering (constructing the argument from the provided estimator or unbiasedness result) is tempting but rather obvious to detect.
- it is not because $X_{(1)}$ is the smallest observation that its inverse first moment does not exist or that it is the only one that can be equal to zero.