Université Hassiba Benbouali–Chlef
Faculté des Sciences Exactes et Informatique
Département de Mathématiques

1^{ère} **année Master MAS**
Statistique Inférentielle
Année : 2019/2020

Rattrapage

EXERCICE N° 1:

Soit $X_1, \ldots, X_n$ un échantillon que l'on modélise par une densité

$$f_\theta(x) = cx^{\theta-1}\mathbb{1}_{[0,1]}(x),$$

où $\theta > 0$ est un paramètre inconnu.

1. Déterminer la constante c pour que f_θ soit une densité.
2. Calculer l'espérance et la variance de X.
3. Vérifier que le modèle est de la famille exponentielle.
4. Déterminez une statistique exhaustive pour le paramètre θ.
5. Trouver un estimateur $\widetilde{\theta}_n$ de θ par la méthode des moments.
6. Trouver l'estimateur du maximum de vraisemblance $\widehat{\theta}_n$.
7. Calculer l'information de Fisher $I_n(\theta)$.

EXERCICE N° 2:

Dans un centre avicole, des études antérieures ont montré que la masse d'un oeuf choisi au hasard peut être considérée comme la réalisation d'une variable aléatoire normale X, de moyenne m et de variance σ^2.
On admet que les masses des oeufs sont indépendantes les unes des autres. On prend un échantillon de $n = 16$ oeufs que l'on pèse. Les mesures sont données dans le tableau suivant :

$$51.41;\ 52.22;\ 52.38;\ 52.62;\ 53.28;\ 53.30;\ 53.32;\ 53.39$$
$$54.78;\ 54.93;\ 54.99;\ 55.28;\ 55.95;\ 57.05;\ 57.31;\ 59.30$$

1. Donner une estimation des paramètres m et σ^2.
2. Donner un intervalle de confiance au niveau 95% de la masse moyenne m d'un oeuf.
3. Donner un intervalle de confiance au niveau 95%, pour σ^2.
4. Tester si la moyenne m est égale à 54 contre $m = 55$.

Correction du rattrapage

EXERCICE N° 1:

1. Pour que f_θ soit une densité, il faut que $c \geq 0$ et

$$\int_0^1 cx^{\theta-1}\,\mathrm{d}x = 1 \Rightarrow c\left(\frac{x^\theta}{\theta}\Big]_0^1\right) = 1 \Rightarrow c\left(\frac{1}{\theta}\right) = 1$$
$$\Rightarrow \boxed{c = \theta}$$

2. On a

$$\begin{aligned}\mathbb{E}[X] &= \int_{-\infty}^{+\infty} x f_\theta(x)\,\mathrm{d}x = \int_0^1 x\theta x^{\theta-1}\,\mathrm{d}x = \int_0^1 \theta x^\theta\,\mathrm{d}x \\ &= \frac{\theta}{\theta+1}x^{\theta+1}\Big]_0^1 = \frac{\theta}{\theta+1}.\end{aligned}$$

De plus,

$$\begin{aligned}\mathbb{E}[X^2] &= \int_{-\infty}^{+\infty} x^2 f_\theta(x)\,\mathrm{d}x = \int_0^1 x^2\theta x^{\theta-1}\,\mathrm{d}x = \int_0^1 \theta x^{\theta+1}\,\mathrm{d}x \\ &= \frac{\theta}{\theta+2}x^{\theta+2}\Big]_0^1 = \frac{\theta}{\theta+2}.\end{aligned}$$

$$\begin{aligned}\mathbb{V}[X] &= \mathbb{E}[X^2] - \mathbb{E}^2[X] = \frac{\theta}{\theta+2} - \frac{\theta^2}{(\theta+1)^2} \\ &= \frac{\theta}{(\theta+2)(\theta+1)^2}.\end{aligned}$$

3. On a

$$\mathcal{L}(\theta, x) = \theta x^{\theta-1}\mathbb{1}_{[0,1]}(x) = \theta\exp\{(\theta-1)\ln x\}\mathbb{1}_{[0,1]}(x)$$

La loi appartient à la famille exponentielle avec $\beta(\theta) = \theta$, $\alpha(\theta) = \theta - 1$, $T(x) = \ln x$ et $\xi(x) = \mathbb{1}_{[0,1]}(x)$.

4. Comme la loi appartient à la famille exponentielle et son support ne dépend pas de θ, une statistique exhaustive pour le paramètre θ est :

$$T(X_1, \ldots, X_n) = \sum_{i=1}^n \ln X_i.$$

5. On résout l'équation :

$$\begin{aligned}\mathbb{E}[X] = \bar{X}_n \quad &\Rightarrow \quad \frac{\theta}{\theta+1} = \bar{X}_n \\ &\Rightarrow \quad \theta = \frac{\bar{X}_n}{1-\bar{X}_n}\end{aligned}$$

L'estimateur des moments résultant de cette dernière égalité est

$$\widetilde{\theta}_n = \frac{\bar{X}_n}{1-\bar{X}_n}.$$

6. La vraisemblance de $(X_1, \ldots, X_n)$ est définie par

$$\mathcal{L}(X_1, \ldots, X_n; \theta) \quad := \quad \prod_{i=1}^{n} f_\theta(X_i) = \prod_{i=1}^{n} \theta X_i^{\theta-1} \mathbb{1}_{[0,1]}(X_i)$$

On a alors

$$\begin{aligned} \ln \mathcal{L}(X_1, \ldots, X_n; \theta) &= n \ln \theta + (\theta - 1) \sum_{i=1}^{n} \ln X_i \\ \frac{\partial}{\partial \theta} \ln \mathcal{L}(X_1, \ldots, X_n; \theta) &= \frac{n}{\theta} + \sum_{i=1}^{n} \ln X_i \end{aligned}$$

On obtient :

$$\frac{\partial}{\partial \theta} \ln L(X_1, \ldots, X_n; \theta) = 0 \iff \widehat{\theta}_n = -\frac{n}{\sum_{i=1}^{n} \ln X_i}.$$

7. On a

$$\frac{\partial^2}{\partial \theta^2} \ln \mathcal{L}(X_1, \ldots, X_n; \theta) \quad = \quad -\frac{n}{\theta^2}$$

L'information de Fisher $I_n(\theta)$:

$$I_n(\theta) \quad = \quad -\mathbb{E}\left[\frac{\partial^2}{\partial \theta^2} \ln \mathcal{L}(X_1, \ldots, X_n; \theta)\right] = \mathbb{E}\left[\frac{n}{\theta^2}\right] = \frac{n}{\theta^2}.$$

EXERCICE N° 2:

1. Une estimation des paramètres m et σ^2 : $\overline{x}_n = 54.47$ et ${s'_n}^2 = 4.6$.
2. On a

$$s'_n = \sqrt{4.6} = 2.14.$$

Dans la table de la loi de Student, pour 15 ddl, nous trouvons la valeur $t_{15,0.975} = 2.13$.
L'intervalle de confiance à 95% de la masse moyenne d'un oeuf :

$$\begin{aligned} \left[\overline{x}_n - \frac{s'_n}{\sqrt{n}} t_{n-1,1-\frac{\alpha}{2}}, \overline{x}_n + \frac{s'_n}{\sqrt{n}} t_{n-1,1-\frac{\alpha}{2}}\right] &= \left[54.47 - \frac{2.14}{\sqrt{16}} \times 2.13, 54.47 + \frac{2.14}{\sqrt{16}} \times 2.13\right] \\ &= [53.33, 55.61] \end{aligned}$$

3. Un intervalle de confiance au niveau 95%, pour σ^2 est :

$$\begin{aligned} \left[\frac{(n-1){s'_n}^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}}, \frac{(n-1){s'_n}^2}{\chi^2_{n-1,\frac{\alpha}{2}}}\right] &= \left[\frac{15 \times 4.6}{27.49}, \frac{15 \times 4.6}{6.26}\right] \\ &= [2.51, 11.01] \end{aligned}$$

4. Soit

$$\begin{cases} H_0 \ : \ m = m_0 \\ H_1 \ : \ m = m_1 \end{cases}$$

Par le théorème de Neyman-Pearson, on rejette l'hypothèse H_0 si

$$\frac{\mathcal{L}\left(m_0, X_1, \ldots, X_n\right)}{\mathcal{L}\left(m_1, X_1, \ldots, X_n\right)} \leq K$$

$$\begin{aligned}
&\Leftrightarrow \quad -\frac{1}{2}\sum_{i=1}^{n}(X_i - m_0)^2 + \frac{1}{2}\sum_{i=1}^{n}(X_i - m_1)^2 \leq \log(K) \\
&\Leftrightarrow \quad (m_0 - m_1)\sum_{i=1}^{n} X_i - \frac{nm_0^2}{2} + \frac{nm_1^2}{2} \leq \log(K) \\
&\Leftrightarrow \quad (m_0 - m_1)\sum_{i=1}^{n} X_i \leq \log(K) + \frac{nm_0^2}{2} - \frac{nm_1^2}{2}
\end{aligned}$$

Comme $m_0 < m_1$, l'inégalité devient

$$\sum_{i=1}^{n} X_i \geq \frac{1}{m_1 - m_0}\left(\log(K) + \frac{nm_1^2}{2} - \frac{nm_0^2}{2}\right)$$

La région critique sera de la forme $W = \left\{\bar{X}_n \geq K_\alpha\right\}$. Sous l'hypothèse H_0, $\dfrac{\sqrt{n}(\bar{X}_n - m_0)}{\sqrt{S_n'^2}}$ suit une loi de Student $\mathcal{T}_{n-1}$, il vient donc :

$$K_\alpha = m_0 + \frac{s'_n}{\sqrt{n}} t_{n-1,1-\alpha}$$

avec $s'_n = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x}_n)^2}$ et $t_{n-1,1-\alpha}$ est le quantile d'ordre $1 - \alpha$ de la loi de Student à $n - 1$ degrés de liberté.

On a

$$\begin{aligned}
K_\alpha &= m_0 + \frac{s'_n}{\sqrt{n}} t_{n-1,1-\alpha} \\
&= 54 + \frac{2.14}{4} \times 1.75 = 54.94.
\end{aligned}$$

On a $\overline{x}_{16} = 54.47 < K_\alpha = 54.94$, on n'est pas dans la région critique : alors on **ne rejette pas** l'hypothèse H_0.

Examen Final

EXERCICE N° 1:
Soit $(X_1, \ldots, X_n)$ un échantillon de v.a. i.i.d de même loi que X, où X suit la loi Géométrique de paramètre p définie

$$\forall k \in \mathbb{N}^*, \qquad \mathbb{P}(X = k) = p(1-p)^{k-1}.$$

1. Montrer que

$$\forall m \in \mathbb{N},\ \forall k \in \mathbb{N},\ \mathbb{P}(X > m + k | X > m) = \mathbb{P}(X > k).$$

2. Montrer que

$$\forall t \geq n, \qquad \mathbb{P}\left(\sum_{i=1}^{n} X_i = t\right) = \mathrm{C}_{t-1}^{n-1} p^n (1-p)^{t-n}.$$

3. En déduire que $T(X_1, \ldots, X_n) = \sum_{i=1}^{n} X_i$ est une statistique exhaustive pour le paramètre p.

EXERCICE N° 2:
Soit $(X_1, \ldots, X_n)$ un échantillon de v.a. i.i.d de même loi que X, où X suit la loi normale $\mathcal{N}(\theta, \theta^2)$ où $\theta > 0$ est un paramètre inconnu.

1. Vérifier que le modèle est de la famille exponentielle.
2. Déterminez une statistique exhaustive pour le paramètre θ.
3. Trouver l'estimateur du maximum de vraisemblance $\widehat{\theta}_n$.
4. Calculer l'information de Fisher $I_n(\theta)$.

EXERCICE N° 3:
Dans un centre avicole, des études antérieures ont montré que la masse d'un oeuf choisi au hasard peut être considérée comme la réalisation d'une variable aléatoire normale X, de moyenne m et de variance σ^2.
On admet que les masses des oeufs sont indépendantes les unes des autres. On prend un échantillon de $n = 16$ oeufs que l'on pèse. Les mesures sont données dans le tableau suivant :

$$51.51;\ 52.07;\ 52.22;\ 52.62;\ 53.13;\ 53.32;\ 53.39;\ 53.79$$
$$54.76;\ 54.78;\ 54.93;\ 54.99;\ 55.12;\ 55.28;\ 55.56;\ 55.95$$

1. Donner une estimation des paramètres m et σ^2.
2. Donner un intervalle de confiance au niveau 95% de la masse moyenne m d'un oeuf.
3. Donner un intervalle de confiance au niveau 95%, pour σ^2.
4. Tester si la moyenne m est égale à 54 contre $m = 53$.

Correction de l'examen Final

EXERCICE N° 1:

Soit $(X_1, \dots, X_n)$ un échantillon de v.a. i.i.d de même loi que X, où X suit la loi Géométrique de paramètre p définie

$$\forall k \in \mathbb{N}^*, \qquad \mathbb{P}(X = k) = p(1-p)^{k-1}.$$

1. Pour tous $m \in \mathbb{N}$ et $k \in \mathbb{N}$, comme $\{X > m+k\} \subset \{X > k\}$,

$$\mathbb{P}(X > m+k / X > m) = \frac{\mathbb{P}(X > m+k\} \cap \{X > n\})}{\mathbb{P}(X > m)} = \frac{\mathbb{P}(X > m+k)}{\mathbb{P}(X > m)}.$$

D'autre part, pour tout $l \in \mathbb{N}$

$$\mathbb{P}(X > l) = \sum_{k \geq l+1} \mathbb{P}(X = k) = \sum_{k \geq l+1} p(1-p)^{k-1} = (1-p)^l \sum_{k-(l+1) \geq 0} p(1-p)^{k-(l+1)} = (1-p)^l.$$

Finalement, pour tous $m \in \mathbb{N}$ et $k \in \mathbb{N}$,

$$\mathbb{P}(X > m+k / X > m) = \frac{(1-p)^{m+k}}{(1-p)^m} = (1-p)^k = \mathbb{P}(X > k).$$

2. On pose $S_n := X_1 + \dots + X_n$.
Par récurrence sur n on peut montrer que S_n suit la loi Binomiale négative de paramètres p et n, $\sum_{i=1}^{n} X_i \sim \mathcal{NB}(n,p)$, c'est-à-dire

$$\mathbb{P}(S_n = t) = \mathrm{C}_{t-1}^{n-1} p^n (1-p)^{t-n} \qquad \text{pour } t = n, n+1, \dots$$

Il est clair que ceci est vrai pour $n = 1$, et pour S_{n+1} on trouve que pour $t = n+1, n+2, \dots$:

$$\begin{aligned}
\mathbb{P}(S_{n+1} = t) &= \sum_{k=n}^{t-1} \mathbb{P}(S_n = k \wedge X_{n+1} = t-k) = \sum_{k=n}^{t-1} \mathbb{P}(S_n = k) \times \mathbb{P}(X_{n+1} = t-k) \\
&= \sum_{k=n}^{t-1} \mathrm{C}_{k-1}^{n-1} p^n (1-p)^{k-n} \times p(1-p)^{t-k-1} \\
&= p^{n+1}(1-p)^{t-n-1} \sum_{k=n}^{t-1} \mathrm{C}_{k-1}^{n-1} = p^{n+1}(1-p)^{t-n-1} \mathrm{C}_t^n
\end{aligned}$$

3. On veut montrer que $T(X_1, \dots, X_n) = \sum_{i=1}^{n} X_i$ est une statistique exhaustive pour p. On écrit :

$$\mathbb{P}(X_1 = x_1, \dots, X_n = x_n | T = t) = \frac{\mathbb{P}(X_1 = x_1, \dots, X_n = x_n, T = t)}{\mathbb{P}(T = t)}$$

$$\begin{aligned}
\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n, T = t) &= \mathbb{P}(X_1 = x_1, \ldots, X_{n-1} = x_{n-1}, X_n = t - \sum_{i=1}^{n-1} x_i) \\
&= \mathbb{P}(X_1 = x_1) \times \mathbb{P}(X_2 = x_2) \times \cdots \times \mathbb{P}(X_n = t - \sum_{i=1}^{n-1} x_i) \\
&= p(1-p)^{x_1 - 1} \times \cdots \times p(1-p)^{x_{n-1}-1} \times p(1-p)^{t - \sum_{i=1}^{n-1} x_i - 1} \\
&= p^n (1-p)^{\sum_{i=1}^{n-1} x_i - (n-1)} (1-p)^{t - \sum_{i=1}^{n-1} x_i - 1} \\
&= p^n (1-p)^{t-n}
\end{aligned}$$

On sait que $\sum_{i=1}^{n} X_i \sim \mathcal{NB}(n, p)$, alors :

$$\mathbb{P}(T = t) = \mathbb{P}\left(\sum_{i=1}^{n} X_i = t\right) = \mathrm{C}_{t-1}^{n-1} p^n (1-p)^{t-n}$$

Finalement,

$$\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n | T = t) = \frac{p^n (1-p)^{t-n}}{\mathrm{C}_{t-1}^{n-1} p^n (1-p)^{t-n}} = \frac{1}{\mathrm{C}_{t-1}^{n-1}}$$

qui ne dépend pas de p, alors $T = \sum_{i=1}^{n} X_i$ est une statistique exhaustive pour le paramètre p.

Exercice n° 2:

1. On a

$$\mathcal{L}(\theta, x) = \frac{\mathrm{e}^{-\frac{1}{2}}}{\theta\sqrt{2\pi}} \exp\left\{\frac{1}{\theta} x - \frac{1}{2\theta^2} x^2\right\}$$

La loi appartient à la famille exponentielle avec $\beta(\theta) = \frac{\mathrm{e}^{-\frac{\theta}{2}}}{\sqrt{2\pi}\theta}$, $\alpha(\theta) = \left(\frac{1}{\theta}, -\frac{1}{2\theta^2}\right)$, $T(x) = (x, x^2)$ et $\xi(x) = 1$.

2. Comme la loi appartient à la famille exponentielle, une statistique exhaustive pour le paramètre θ est :

$$T(X_1, \ldots, X_n) = \left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2\right).$$

La vraisemblance s'écrit :

$$\begin{aligned}
\mathcal{L}(\theta, x_1, \ldots, x_n) &= \prod_{i=1}^{n} \frac{1}{\theta\sqrt{2\pi}} \exp\left[-\frac{1}{2\theta^2}(x_i - \theta)^2\right] = \frac{1}{(\theta\sqrt{2\pi})^n} \exp\left[-\frac{1}{2\theta^2} \sum_{i=1}^{n} (x_i - \theta)^2\right] \\
&= \underbrace{\frac{1}{(\theta\sqrt{2\pi})^n} \exp\left[\frac{1}{\theta} \sum_{i=1}^{n} x_i - \frac{1}{2\theta^2} \sum_{i=1}^{n} x_i^2 - \frac{n}{2}\right]}_{g(\theta, T(x_1, \ldots, x_n))} \underbrace{1}_{h(x_1, \ldots, x_n)}.
\end{aligned}$$

Par le critère de factorisation Neyman-Fisher, on en déduit que $T(X_1, \ldots, X_n) = \left(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2\right)$ est une statistique exhaustive pour le paramètre θ.

3. Calculons à présent l'estimateur du maximum de vraisemblance sur $\Theta =]0, +\infty[$, on a

$$\begin{aligned}\mathcal{L}(x_1, \ldots, x_n; \theta) &= \prod_{i=1}^n f_\theta(x_i) = \prod_{i=1}^n \frac{1}{\theta\sqrt{2\pi}} \exp\left[-\frac{1}{2\theta^2}(x_i - \theta)^2\right] \\ &= \frac{1}{(\theta\sqrt{2\pi})^n} \exp\left[-\frac{1}{2\theta^2}\sum_{i=1}^n (x_i - \theta)^2\right]\end{aligned}$$

On écrit la log-vraisemblance :

$$\ln \mathcal{L}(x_1, \ldots, x_n; \theta) = -n \ln \sqrt{2\pi} - n \ln \theta - \frac{1}{2\theta^2}\sum_{i=1}^n (x_i - \theta)^2$$

La fonction de log-vraisemblance est dérivable sur $\Theta =]0, +\infty[$ et on

$$\frac{\partial}{\partial\theta} \ln \mathcal{L}(x_1, \ldots, x_n; \theta) = \frac{-n}{\theta} - \frac{n\bar{x}_n}{\theta^2} + \frac{\sum_{i=1}^n x_i^2}{\theta^3}.$$

$$\begin{aligned}\frac{\partial}{\partial\theta} \ln \mathcal{L}(x_1, \ldots, x_n; \theta) = 0 \quad &\Rightarrow \quad -n\theta^2 - n\bar{x}_n\theta + \sum_{i=1}^n x_i^2 = 0 \\ &\Rightarrow \quad \boxed{\widehat{\theta}_n = -\frac{\bar{X}_n}{2} + \sqrt{\frac{\sum_{i=1}^n X_i^2}{n} + \frac{\bar{X}_n^2}{4}}.}\end{aligned}$$

4. L'information de Fisher $I_n(\theta)$:

$$\frac{\partial^2 \ln \mathcal{L}(\theta, x_1, \ldots, x_n)}{\partial\theta^2} = \frac{n}{\theta^2} + \frac{2n\bar{x}_n}{\theta^3} - \frac{3}{\theta^4}\sum_{i=1}^n x_i^2.$$

Ainsi,

$$\begin{aligned}I_n(\theta) &= -\mathbb{E}\left[\frac{\partial^2 \ln \mathcal{L}(\theta, X_1, \ldots, X_n)}{\partial\theta^2}\right] = -\frac{n}{\theta^2} - \frac{2n}{\theta^3} \times \mathbb{E}\left[\bar{X}_n\right] + \frac{3}{\theta^4} \times \mathbb{E}\left[\sum_{i=1}^n X_i^2\right] \\ &= -\frac{n}{\theta^2} - \frac{2n}{\theta^3} \times \theta + \frac{3n}{\theta^4} \times (2\theta^2) = \frac{3n}{\theta^2}.\end{aligned}$$

<u>EXERCICE N° 3</u>:

1. Une estimation des paramètres m et σ^2 : $\overline{x}_n = 53.96$ et ${s'_n}^2 = 1.92$.
2. On a

$$s'_n = \sqrt{1.92} = 1.38.$$

Dans la table de la loi de Student, pour 15 ddl, nous trouvons la valeur $t_{15,0.975} = 2.13$. L'intervalle de confiance à 95% de la masse moyenne d'un oeuf :

$$\begin{aligned}\left[\overline{x}_n - \frac{s'_n}{\sqrt{n}} t_{n-1,1-\frac{\alpha}{2}}, \overline{x}_n + \frac{s'_n}{\sqrt{n}} t_{n-1,1-\frac{\alpha}{2}}\right] &= \left[53.96 - \frac{1.38}{\sqrt{16}} \times 2.13, 53.96 + \frac{1.38}{\sqrt{16}} \times 2.13\right] \\ &= [53.22, 54.69]\end{aligned}$$

3. Un intervalle de confiance au niveau 95%, pour σ^2 est :

$$\left[\frac{(n-1){s'_n}^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}}, \frac{(n-1){s'_n}^2}{\chi^2_{n-1,\frac{\alpha}{2}}}\right] = \left[\frac{15\times 1.92}{27.49}, \frac{15\times 1.92}{6.26}\right]$$
$$= [1.05, 4.59]$$

4. Soit

$$\begin{cases} H_0 \,:\, m = m_0 \\ H_1 \,:\, m = m_1 \end{cases}$$

Par le théorème de Neyman-Pearson, on rejette l'hypothèse H_0 si

$$\frac{\mathcal{L}(m_0, X_1, \ldots, X_n)}{\mathcal{L}(m_1, X_1, \ldots, X_n)} \le K$$

$$\begin{aligned}
\Leftrightarrow\quad & -\frac{1}{2}\sum_{i=1}^{n}(X_i - m_0)^2 + \frac{1}{2}\sum_{i=1}^{n}(X_i - m_1)^2 \le \log(K)\\
\Leftrightarrow\quad & (m_0 - m_1)\sum_{i=1}^{n} X_i - \frac{nm_0^2}{2} + \frac{nm_1^2}{2} \le \log(K)\\
\Leftrightarrow\quad & (m_0 - m_1)\sum_{i=1}^{n} X_i \le \log(K) + \frac{nm_0^2}{2} - \frac{nm_1^2}{2}
\end{aligned}$$

Comme $m_0 > m_1$, l'inégalité devient

$$\sum_{i=1}^{n} X_i \le \frac{1}{m_1 - m_0}\left(\log(K) + \frac{nm_1^2}{2} - \frac{nm_0^2}{2}\right)$$

La région critique sera de la forme $W = \left\{\bar{X}_n \le K_\alpha\right\}$. Sous l'hypothèse H_0, $\dfrac{\sqrt{n}(\bar{X}_n - m_0)}{\sqrt{S'^2_n}}$ suit une loi de Student $\mathcal{T}_{n-1}$, il vient donc :

$$K_\alpha = m_0 - \frac{s'_n}{\sqrt{n}} t_{n-1,1-\alpha}$$

avec $s'_n = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x}_n)^2}$ et $t_{n-1,1-\alpha}$ est le quantile d'ordre $1-\alpha$ de la loi de Student à $n-1$ degrés de liberté.

On a

$$K_\alpha = m_0 - \frac{s'_n}{\sqrt{n}} t_{n-1,1-\alpha}$$
$$= 54 - \frac{1.38}{4} \times 1.75 = 53.39.$$

On a $\overline{x}_{16} = 53.96 > K_\alpha = 53.39$, on n'est pas dans la région critique : alors on **ne rejette pas** l'hypothèse H_0.

Université Hassiba Benbouali–Chlef
Faculté des Sciences Exactes et Informatique
Département de Mathématiques

$1^{\text{ère}}$ **année Master MAS**
Séries Chronologiques
Année : 2018/2019

Examen Final

EXERCICE N° 1:
Soit $(X_1, \ldots, X_n)$ un échantillon de v.a. i.i.d de même loi que X, où X admet pour densité de probabilité :

$$f_X(x) = cx \exp\left(-\lambda x^2\right) \mathbb{1}_{\{x>0\}},$$

avec $\lambda \in]0, +\infty[$.

1. Déterminer la constante c pour que f_X soit une densité.
2. Calculez l'espérance $\mathbb{E}[X]$ et la variance $\mathbb{V}[X]$ de la variable aléatoire X.
3. Vérifier que le modèle est de la famille exponentielle.
4. Déterminez une statistique exhaustive pour le paramètre λ.
5. Trouver un estimateur de λ par la méthode des moments.
6. Donner l'estimateur du maximum de vraisemblance $\widehat{\lambda}_n$.
   (i) Cet estimateur est-il biaisé ? Calculez son risque quadratique.
   (ii) Cet estimateur est-il convergent ?
7. Calculer l'information de Fisher $I_n(\lambda)$.

EXERCICE N° 2:
Un fabricant de thermomètres prélève un échantillon de taille $n = 76$ dans un lot qu'il vient de produire. La valeur indiquée par ces thermomètres plongés dans un liquide maintenu à la température constante de 16 degrés est considérée comme une v.a. X de loi normale dont l'écart type σ est inconnu. Il obtient des observations de températures telles que

$$\sum_{i=1}^{76} x_i = 1140 \text{ et } \sum_{i=1}^{76} x_i^2 = 17195.$$

1. Donner une estimation de la précision d'un thermomètre, mesurée par la dispersion σ^2 des mesures.
2. Construire un intervalle de confiance de niveau 0.95 pour la précision d'un thermomètre.
3. Combien aurait-il fallu prélever de thermomètres pour que le résultat de la question précédente soit fourni avec un risque de 1% ?

Il désire effectuer le test entres les deux hypothèses suivantes :

$$H_0 \quad : \quad \sigma = 1$$

Contre

$$H_1 \quad : \quad \sigma = 2$$

4. Pour un risque de première espèce $\alpha = 0.05$, indiquer ce que sera sa conclusion.
5. Calculer la puissance du test.
6. Quelle devrait être la taille d'échantillon minimum pour que cette puissance soit supérieure à 0.95 ?
7. Que se produit-il si on intervertit les deux hypothèses ?

Corrigé de l'examen final

EXERCICE N° 1:

1. Pour que f_X soit une densité, il faut que

$$\int_{-\infty}^{+\infty} f_X(x)\,\mathrm{d}x = 1$$

$$\begin{aligned}
&\Longrightarrow \int_{-\infty}^{+\infty} cx \exp\left(-\lambda x^2\right) \mathbb{1}_{\{x>0\}}\,\mathrm{d}x = 1\\
&\Longrightarrow \frac{c}{2\lambda} \underbrace{\int_{0}^{+\infty} \exp(-u)\,\mathrm{d}u}_{=1} = 1\\
&\Longrightarrow \frac{c}{2\lambda} = 1\\
&\Longrightarrow c = 2\lambda.
\end{aligned}$$

2. Calculons l'espérance de X :

   Posons $u = \lambda x^2$, alors on a $x = \sqrt{\frac{u}{\lambda}}$ et $\mathrm{d}u = 2\lambda x \mathrm{d}x$. On a,

$$\begin{aligned}
\mathbb{E}[X] &= 2\lambda \int_0^\infty x^2 \mathrm{e}^{-\lambda x^2}\,\mathrm{d}x\\
&= \frac{1}{\sqrt{\lambda}} \underbrace{\int_0^\infty \sqrt{u}\mathrm{e}^{-u}\,\mathrm{d}u}_{=\frac{\sqrt{\pi}}{2}} = \frac{1}{2}\sqrt{\frac{\pi}{\lambda}}
\end{aligned}$$

   où

$$\Gamma\left(\frac{3}{2}\right) = \frac{\sqrt{\pi}}{2}.$$

   De plus,

$$\begin{aligned}
\mathbb{E}[X^2] &= 2\lambda \int_0^\infty x^3 \mathrm{e}^{-\lambda x^2}\,\mathrm{d}x\\
&= \frac{1}{\lambda} \underbrace{\int_0^\infty u\mathrm{e}^{-u}\,\mathrm{d}u}_{=1} = \frac{1}{\lambda}
\end{aligned}$$

   où

$$\Gamma(n+1) = \int_0^\infty v^n \mathrm{e}^{-v}\,\mathrm{d}v = n!, \quad \text{pour } n \text{ entier.}$$

   Finalement

$$\mathrm{Var}[X] = \mathbb{E}\left[X^2\right] - (\mathbb{E}[X])^2 = \frac{4-\pi}{4\lambda}.$$

3. Nous avons :

$$\begin{aligned} L(\lambda, x) &:= \lambda \times 2x\mathbb{1}_{\{x>0\}} \times \exp\left(-\lambda \times x^2\right) \\ &= \beta(\lambda) \times \xi(x) \times \exp\left(\alpha(\lambda) \times T(x)\right), \end{aligned}$$

La loi appartient donc à la famille exponentielle avec les notations usuelles $\beta(\lambda) = \lambda$, $\xi(x) = 2x\mathbb{1}_{\{x>0\}}$, $\alpha(\lambda) = -\lambda$ et $T(x) = x^2$.

4. D'après le résultat du cours sur les familles exponentielles, on obtient que $T(X_1, \ldots, X_n) = \sum_{i=1}^{n} X_i^2$ est une statistique exhaustive pour le paramètre λ.

5. L'espérance de X est

$$\mathbb{E}(X) = \frac{1}{2}\sqrt{\frac{\pi}{\lambda}} \text{ donc } \lambda = \frac{\pi}{4[\mathbb{E}(X)]^2}$$

L'estimateur des moments résultant de cette dernière égalité est

$$\widetilde{\lambda}_n = \frac{\pi}{4[\overline{X}_n]^2}$$

6. La vraisemblance de $(x_1, \ldots, x_n)$ est définie par

$$\begin{aligned} L(x_1, \ldots, x_n; \lambda) &:= \prod_{i=1}^{n} f_X(x_i) = \prod_{i=1}^{n} 2\lambda \times x_i\mathbb{1}_{\{x_i>0\}} \times \exp\left(-\lambda \times x_i^2\right) \\ &= 2^n\lambda^n \exp\left(-\lambda \sum_{i=1}^{n} x_i^2\right) \prod_{i=1}^{n} x_i\mathbb{1}_{\{x_i>0\}} \end{aligned}$$

On a alors

$$\begin{aligned} \ln L(x_1, \ldots, x_n; \lambda) &= n\ln 2 + n\ln\lambda - \lambda\sum_{i=1}^{n} x_i^2 + \sum_{i=1}^{n} \ln x_i \\ \frac{\partial}{\partial\lambda}\ln L(x_1, \ldots, x_n; \lambda) &= \frac{n}{\lambda} - \sum_{i=1}^{n} x_i^2 \end{aligned}$$

On obtient :

$$\frac{\partial}{\partial\lambda}\ln L(x_1, \ldots, x_n; \lambda) = 0 \iff \widehat{\lambda}_n = \frac{n}{\sum_{i=1}^{n} X_i^2}$$

(i) La moyenne de l'estimateur $\widehat{\lambda}_n$ est

$$\mathbb{E}[\widehat{\lambda}_n] = \mathbb{E}\left[\frac{n\lambda}{\sum_{i=1}^{n} \lambda X_i^2}\right]$$

On effectue tout d'abord le changement de variable $Y = \lambda X^2$, on écrit la fonction de répartition :

$$\begin{aligned} \mathbb{F}_Y(y) &= \mathbb{P}(Y \leq y) = \mathbb{P}(\lambda X^2 \leq y) \\ &= \mathbb{P}\left(X \leq \sqrt{\frac{y}{\lambda}}\right) = \mathbb{F}_X\left(\sqrt{\frac{y}{\lambda}}\right), \end{aligned}$$

ce qui implique

$$\begin{aligned} f_Y(y) &= \mathbb{F}'_Y(y) = \mathbb{F}'_X\left(\sqrt{\frac{y}{\lambda}}\right) \\ &= \frac{1}{2\sqrt{\lambda y}} f_X\left(\sqrt{\frac{y}{\lambda}}\right) = \frac{1}{2\sqrt{\lambda y}} 2\lambda\sqrt{\frac{y}{\lambda}}\exp\left(-\lambda\left(\sqrt{\frac{y}{\lambda}}\right)^2\right) \\ &= \mathrm{e}^{-y} \end{aligned}$$

On constate ainsi que la densité de Y est la loi exponentielle. Puisque les X_i sont indépendants, les Y_i le sont aussi et grâce aux propriétés de la loi gamma nous savons donc que

$$Z = \sum_{i=1}^{n} Y_i \sim \Gamma(n, 1)$$

Ceci nous permet de calculer le biais de $\widehat{\lambda}_n$:

$$\mathbb{E}\left[\widehat{\lambda}_n\right] = \mathbb{E}\left[\frac{n\lambda}{Z}\right] = n\lambda\mathbb{E}\left[\frac{1}{Z}\right].$$

De plus,

$$\mathbb{E}\left[\frac{1}{Z}\right] = \int_0^{+\infty} \frac{1}{z}\frac{1}{\Gamma(n)} z^{n-1}\exp(-z)\,\mathrm{d}z = \frac{1}{\Gamma(n)}\int_0^{+\infty} z^{n-2}\exp(-z)\,\mathrm{d}z.$$

On reconnaît la quantité $\Gamma(n-1)$ et la relation $\Gamma(n) = (n-1)\Gamma(n-1)$ permet de conclure pour le calcul de l'espérance :

$$\mathbb{E}\left[\widehat{\lambda}_n\right] = \frac{n\lambda}{(n-1)}.$$

L'estimateur est donc biaisé, mais asymptotiquement sans biais.

Pour calculer la variance de $\widehat{\lambda}_n$, on calcule d'abord $\mathbb{E}\left[\frac{1}{Z^2}\right]$. Avec le même changement de variable que précédemment, on obtient

$$\mathbb{E}\left[\frac{1}{Z^2}\right] = \frac{1}{(n-1)(n-2)}$$

d'où

$$\mathbb{V}\left[\widehat{\lambda}_n\right] = \frac{(n\lambda)^2}{(n-1)(n-2)} - \frac{(n\lambda)^2}{(n-1)^2} = (n\lambda)^2\left(\frac{1}{(n-1)(n-2)} - \frac{1}{(n-1)^2}\right).$$

Soit

$$\mathbb{V}\left[\widehat{\lambda}_n\right] = \frac{n^2\lambda^2}{(n-1)^2(n-2)}.$$

Son risque quadratique est égal a

$$\begin{aligned} \mathcal{R}(\widehat{\lambda}_n, \lambda) &= \text{Biais}^2\left(\widehat{\lambda}_n\right) + \mathbb{V}[\widehat{\lambda}_n] \\ &= \frac{\lambda^2}{(n-1)^2} + \frac{n^2\lambda^2}{(n-1)^2(n-2)} = \frac{(n^2+n-2)\lambda^2}{(n-1)^2(n-2)}. \end{aligned}$$

(ii) Cet estimateur est convergent car il est asymptotiquement sans biais et

$$\lim_{n\to+\infty} \mathbb{V}[\widehat{\lambda}_n] = \lim_{n\to+\infty} \frac{n^2\lambda^2}{(n-1)^2(n-2)} = 0.$$

7. L'information de Fisher $I_n(\lambda)$:

$$\begin{aligned} I_n(\lambda) &= -\mathbb{E}\left[\frac{\partial^2}{\partial\lambda^2}\ln L(X_1,\ldots,X_n;\lambda)\right] = -\mathbb{E}\left[-\frac{n}{\lambda^2}\right] \\ &= \frac{n}{\lambda^2} \end{aligned}$$

EXERCICE N° 2:

Un fabricant de thermomètres prélève un échantillon de taille $n = 76$ dans un lot qu'il vient de produire. La valeur indiquée par ces thermomètres plongés dans un liquide maintenu à la température constante de 16 degrés est considérée comme une v.a. X de loi normale dont l'écart type σ est inconnu. Il obtient des observations de températures telles que

$$\sum_{i=1}^{76} x_i = 1140 \text{ et } \sum_{i=1}^{76} x_i^2 = 17195.$$

1. On obtient comme estimation de la variance :

$$\widehat{\sigma^2}_{76} = \frac{1}{76}\sum_{i=1}^{76}(x_i-\mu)^2 = \frac{1}{76}\sum_{i=1}^{76}(x_i-16)^2 = \frac{171}{76} = 2.25$$

2. Un intervalle de confiance au niveau 95%, pour σ^2 : On a

$$\sum_{i=1}^{76}(x_i-\mu)^2 = 17195 - 2\times 16\times 1140 + 76\times(16)^2 = 171$$

Donc

$$\begin{aligned} \left[\frac{\sum_{i=1}^{76}(x_i-\mu)^2}{\chi^2_{n,1-\frac{\alpha}{2}}}, \frac{\sum_{i=1}^{76}(x_i-\mu)^2}{\chi^2_{n,\frac{\alpha}{2}}}\right] &= \left[\frac{171}{102.00}, \frac{171}{53.78}\right] \\ &= [1.68, 3.18] \end{aligned}$$

Il désire effectuer le test entres les deux hypothèses suivantes :

$$H_0 : \sigma = 1$$

Contre

$$H_1 : \sigma = 2$$

3. La région critique est donc définie par le théorème de Neyman et Pearson comme l'ensemble des points $(x_1,\ldots,x_n)$ tels que :

$$\sum_{i=1}^{76}(x_i-\mu)^2 \geq C$$

où C est le quantile d'ordre $1-\alpha$ de la loi $\chi^2_{(76)}$ suivie par $\sum_{i=1}^{76}(x_i-\mu)^2$ sous l'hypothèse nulle.

$$\begin{aligned} \alpha &= \mathbb{P}\left(\text{Rejeter } H_0 | H_0 \text{ vraie}\right), \\ &= \mathbb{P}_{H_0}\left(\sum_{i=1}^{n}(X_i-\mu)^2 > C\right) = 1 - \mathbb{P}_{H_0}\left(\sum_{i=1}^{n}(X_i-\mu)^2 \leq C\right) \end{aligned}$$

Pour $\alpha = 0.05$, on lit dans la table, la valeur $C = 97.35$.

Les observations donnent $\sum_{i=1}^{76}(x_i - \mu)^2 = 171$, ce qui conduit à refuser l'hypothèse nulle $\sigma = 1$.

4. La puissance du test :

$$\begin{aligned}
1-\beta &= \mathbb{P}\left(\text{Rejeter } H_0 | H_1 \text{ vraie}\right), \\
&= \mathbb{P}_{H_1}\left(\sum_{i=1}^{n}(X_i - \mu)^2 \geq C\right) \\
&= \mathbb{P}\left(\frac{1}{\sigma_1^2}\sum_{i=1}^{n}(X_i - \mu)^2 \geq \frac{C}{\sigma_1^2}\right) \\
&= \mathbb{P}\left(X \geq 24.25\right)
\end{aligned}$$

où X suit la loi $\chi^2_{(76)}$. On trouve

$$1-\beta \approx 1.$$

5. L'interversion des hypothèses change le sens de l'inégalité, la région critique étant définie par :

$$\sum_{i=1}^{76}(x_i - \mu)^2 \leq C$$

où $\frac{C}{4}$ est le quantile d'ordre α de la loi $\chi^2_{(76)}$ suivie par $\frac{1}{4}\sum_{i=1}^{76}(x_i - \mu)^2$ sous l'hypothèse nulle. Pour $\alpha = 0.05$, on lit dans la table, la valeur $\frac{C}{4} = 56.92$. Donc $C = 227.68$.

Les observations donnent $\sum_{i=1}^{76}(x_i - \mu)^2 = 171$, ce qui conduit à refuser l'hypothèse nulle $\sigma = 2$.

Examen Final

EXERCICE N° 1:

Soit $X_1, \ldots, X_n$ un échantillon que l'on modélise par une une loi uniforme $\mathcal{U}[0, \theta]$ où $\theta > 0$ est un paramètre inconnu.

1. Déterminer l'estimateur du maximum de vraisemblance.
   (i) Est-il biaisé ?
   (ii) Calculez son risque quadratique. Est-il convergent ?
2. Proposer une statistique libre pour le paramètre θ.
3. Donner un intervalle de confiance au niveau $1 - \alpha$, pour θ.

On cherche à tester

$$H_0 \quad : \quad \theta = \theta_0$$
$$\text{Contre}$$
$$H_1 \quad : \quad \theta = \theta_1$$

où $\theta_0 < \theta_1$ sont deux quantités connues.

4. Pourquoi ne peut-on pas utiliser ici le test de Neyman-Pearson ?

Soit $M = \max_{1 \leq i \leq n} X_i$. On propose le test suivant : On rejette H_0 lorsque $M > K$ (K constante donnée).

5. Si $\theta_0 = 1$, $\theta_1 = 2$, $n = 20$ et que la valeur observée de M est 0.96 :
   (i) Quelle valeur prendre pour K pour obtenir un niveau de 5% ?
   (ii) Quelle conclusion tirer sur H_0 ?
   (iii) Calculer la puissance du test.
6. Quelle conclusion tirer sur H_0 lorsque la valeur observée de M est 1.04 ?

EXERCICE N° 2:

Soit $(X_1, \ldots, X_n)$ un échantillon de v.a. i.i.d de même loi que X, où X suit la loi normale $\mathcal{N}(\theta, \theta)$ où $\theta > 0$ est un paramètre inconnu.

1. Vérifier que le modèle est de la famille exponentielle.
2. Déterminez une statistique exhaustive pour le paramètre θ.
3. Trouver l'estimateur du maximum de vraisemblance $\widehat{\theta}_n$.
4. Calculer l'information de Fisher $I_n(\theta)$.

Université Hassiba Benbouali–Chlef
Faculté des Sciences
Département de Mathématiques

1ère **année Master MAS**
Statistique Inférentielle
Année : 2017/2018

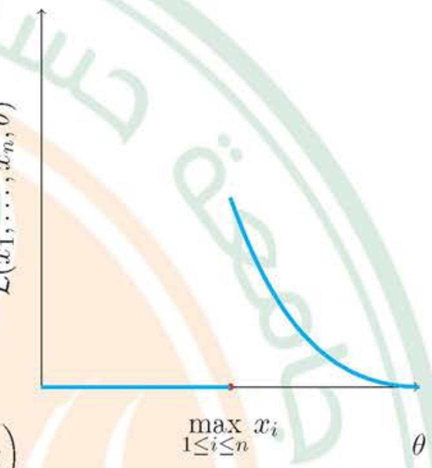
Corrigé de l'Examen Final

EXERCICE N° 1: 13 points

Soit $X_1, \dots, X_n$ un échantillon que l'on modélise par une loi uniforme $\mathcal{U}[0,\theta]$ où $\theta > 0$ est un paramètre inconnu.

1. On écrit la vraisemblance :

$$\mathcal{L}(x_1,\dots,x_n,\theta) = \prod_{i=1}^{n} \frac{1}{\theta} \mathbb{1}_{[0,\theta]}(x_i) = \left(\frac{1}{\theta}\right)^n \prod_{i=1}^{n} \mathbb{1}_{\{0 \le x_i \le \theta\}}$$
$$= \left(\frac{1}{\theta}\right)^n \mathbb{1}_{\left\{\max\limits_{1\le i\le n} x_i \le \theta\right\}} \mathbb{1}_{\left\{\min\limits_{1\le i\le n} x_i \ge 0\right\}}$$



Ceci implique que $\hat{\theta}_n = \max\limits_{1\le i\le n} X_i$ est l'estimateur du maximum de vraisemblance pour la loi uniforme $\mathcal{U}[0,\theta]$.

(i) D'abord calculons la loi de $\hat{\theta}_n$. On a :

$$\mathbb{F}_{\hat{\theta}_n}(x) = \mathbb{P}\left(\hat{\theta}_n \le x\right) = \mathbb{P}\left(\max_{1\le i\le n} X_i \le x\right)$$
$$= \mathbb{P}\left(X_1 \le x, X_2 \le x, \dots, X_n \le x\right)$$
$$= \mathbb{P}\left(X_1 \le x\right)\mathbb{P}\left(X_2 \le x\right)\dots\mathbb{P}\left(X_n \le x\right)$$
$$= \left(\frac{x}{\theta}\right)^n$$

Ainsi, on obtient la densité de $\hat{\theta}_n$:

$$f_{\hat{\theta}_n}(x) = n\frac{x^{n-1}}{\theta^n}$$

$$\mathbb{E}[\hat{\theta}_n] = \int_0^\theta nx\frac{x^{n-1}}{\theta^n}\,\mathrm{d}x = \frac{n}{\theta^n}\int_0^\theta x^n\,\mathrm{d}x$$
$$= \frac{n}{n+1}\theta.$$

On remarque que cet estimateur est asymptotiquement sans biais. En effet,

$$\lim_{n\longrightarrow+\infty} \mathbb{E}[\hat{\theta}_n] = \lim_{n\longrightarrow+\infty} \frac{n}{n+1}\theta = \theta.$$

(ii) Pour calculer la variance de $\hat{\theta}_n$, on calcule d'abord

$$\mathbb{E}[(\hat{\theta}_n)^2] = \int_0^\theta nx^2\frac{x^{n-1}}{\theta^n}\,\mathrm{d}x = \frac{n}{\theta^n}\int_0^\theta x^{n+1}\,\mathrm{d}x$$
$$= \frac{n}{n+2}\theta^2.$$

On en déduit que :

$$\mathbb{V}[\hat{\theta}_n] = \frac{n}{n+2}\theta^2 - \frac{n^2}{(n+1)^2}\theta^2 = \frac{n}{(n+2)(n+1)^2}\theta^2.$$

Son risque quadratique est égal a

$$\mathcal{R}(\hat{\theta}_n, \theta) = \mathbb{V}[\hat{\theta}_n] + \text{Biais}^2[\hat{\theta}_n] = \frac{n}{(n+2)(n+1)^2}\theta^2 + \frac{1}{(n+1)^2}\theta^2 = \frac{2\theta^2}{(n+1)(n+2)}.$$

Cet estimateur est convergent car il est asymptotiquement sans biais et

$$\lim_{n\to+\infty} \mathbb{V}[\hat{\theta}_n] = \lim_{n\to+\infty} \frac{n}{(n+2)(n+1)^2}\theta^2 = 0.$$

2. On pose $T = \dfrac{\hat{\theta}_n}{\theta}$. Pour $t \in \mathbb{R}$, calculons

$$\begin{aligned}\mathbb{F}_T(t) = \mathbb{P}(T \le t) = \mathbb{P}\left(\frac{\hat{\theta}_n}{\theta} \le t\right) = \mathbb{P}\left(\hat{\theta}_n \le t\theta\right) \\ = \mathbb{F}_{\hat{\theta}_n}(t\theta) = \left(\frac{t\theta}{\theta}\right)^n = t^n.\end{aligned}$$

On en déduit que la loi de T ne dépend pas de θ. Ainsi, T est bien une statistique libre pour θ.

3. On cherche le plus court intervalle $[a, b]$ tel que

$$\mathbb{P}(a \le T_n \le b) = \int_a^b f_{T_n}(t)\, \mathrm{d}t = 1 - \alpha$$

Comme f est une fonction strictement décroissante sur $]1, +\infty[$, nécessairement $a = 1$, alors

$$\int_1^b \frac{n}{t^{n+1}}\, \mathrm{d}t = 1 - \alpha$$

soit $b = \alpha^{-\frac{1}{n}}$. Alors, $\mathbb{P}(a \le T_n \le b) = 1 - \alpha$ s'écrit

$$\mathbb{P}\left(a \le \frac{\theta}{\hat{\theta}_n} \le b\right) = \mathbb{P}\left(\hat{\theta}_n \le \theta \le \hat{\theta}_n \alpha^{-\frac{1}{n}}\right) = 1 - \alpha$$

autrement, l'intervalle de confiance s'écrit

$$\left[\hat{\theta}_n, \alpha^{-\frac{1}{n}}\hat{\theta}_n\right]$$

4. Les densités n'ont pas même support. Le rapport de vraisemblance n'est donc pas défini.
5. Si $\theta_0 = 1$, $\theta_1 = 2$, $n = 20$ et que la valeur observée de M est 0.96 :
   (i) Pour avoir un niveau $\alpha \in]0, 1[$, il suffit de choisir K tel que $\mathbb{P}\left(\max_{1\le i\le n} X_i > K\right) = \alpha$ càd $K = (1-\alpha)^{\frac{1}{n}}$. Pour $\alpha = 0.05$ et $n = 20$, on prend $K = (0.95)^{0.05} = 0.997$.
   (ii) Comme $M = 0.96 < 0.997 = K$, on **ne rejette pas** l'hypothèse H_0.
   (iii) La puissance du test est :

$$1 - \beta = \mathbb{P}_{H_1}\left(\max_{1\le i\le n} X_i > K\right) = \mathbb{P}\left(\frac{\max_{1\le i\le n} X_i}{2} > \frac{K}{2}\right) = 1 - \left(\frac{K}{2}\right)^n = 0.999$$

6. Comme $M = 1.04 > 0.997 = K$, on **rejette** l'hypothèse H_0.

EXERCICE N° 2: 7 points

Soit $(X_1, \ldots, X_n)$ un échantillon de v.a. i.i.d de même loi que X, où X suit la loi normale $\mathcal{N}(\theta, \theta)$ où $\theta > 0$ est un paramètre inconnu.

1. On a
$$L(\theta, x) = \frac{e^{-\frac{\theta}{2}}}{\sqrt{2\pi\theta}} \exp\{-\frac{1}{2\theta}x^2\}e^x$$
La loi appartient à la famille exponentielle avec $\beta(\theta) = \frac{e^{-\frac{\theta}{2}}}{\sqrt{2\pi\theta}}$, $\alpha(\theta) = -\frac{1}{2\theta}$, $T(x) = x^2$ et $\xi(x) = e^x$.
2. Comme la loi appartient à la famille exponentielle, une statistique exhaustive pour le paramètre θ est :
$$T(X_1, \ldots, X_n) = \sum_{i=1}^{n} X_i^2.$$
3. La vraissemblance s'ecrit :
$$\begin{aligned}\mathcal{L}(\theta, x_1, \ldots, x_n) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\theta}} \exp\left\{\frac{-1}{2}\frac{(x_i-\theta)^2}{\theta}\right\} \\ &= \left(\frac{1}{\sqrt{2\pi\theta}}\right)^n \exp\left\{\frac{-1}{2\theta}\sum_{i=1}^{n}(x_i-\theta)^2\right\}\end{aligned}$$
la log-vraissemblance est :
$$\begin{aligned}\ln \mathcal{L}(\theta, x_1, \ldots, x_n) &= -n\ln(\sqrt{2\pi\theta}) - \frac{1}{2\theta}\sum_{i=1}^{n}(x_i-\theta)^2. \\ &= -\frac{n}{2}\ln(\theta) - n\ln(\sqrt{\pi}) - \frac{1}{2\theta}\sum_{i=1}^{n}x_i^2 + \sum_{i=1}^{n}x_i - \frac{n\theta}{2}. \\ \frac{\partial \ln \mathcal{L}(\theta, x_1, \ldots, x_n)}{\partial\theta} &= -\frac{n}{2\theta} - \frac{n}{2} + \frac{1}{2\theta^2}\sum_{i=1}^{n}x_i^2.\end{aligned}$$
d'où
$$\frac{\partial \ln \mathcal{L}(\theta, x_1, \ldots, x_n)}{\partial\theta} = 0 \quad \Leftrightarrow \quad -n\theta^2 - n\theta + \sum_{i=1}^{n}x_i^2 = 0$$
$$\Rightarrow \begin{cases} \theta_1 = -\frac{1}{2} + \sqrt{\frac{1}{4} + \frac{1}{n}\sum_{i=1}^{n}X_i^2}. \\ \theta_2 = -\frac{1}{2} - \sqrt{\frac{1}{4} + \frac{1}{n}\sum_{i=1}^{n}X_i^2}. \end{cases}$$
Comme $\theta > 0$, l'estimateur du maximum de vraisemblance est $\widehat{\theta}_n = -\frac{1}{2}+\sqrt{\frac{1}{4} + \frac{1}{n}\sum_{i=1}^{n}X_i^2}$.
4. L'information de Fisher $I_n(\theta)$:
$$\frac{\partial^2 \ln \mathcal{L}(\theta, x_1, \ldots, x_n)}{\partial\theta^2} = \frac{n}{2\theta^2} - \frac{1}{\theta^3}\sum_{i=1}^{n}x_i^2.$$
Ainsi,
$$\begin{aligned}I_n(\theta) &= -\mathbb{E}\left[\frac{\partial^2 \ln \mathcal{L}(\theta, X_1, \ldots, X_n)}{\partial\theta^2}\right] \\ &= -\frac{n}{2\theta^2} + \frac{n}{\theta^3}(\theta+\theta^2) \\ &= \frac{n}{2\theta^2} + \frac{n}{\theta}.\end{aligned}$$

Examen Final

Soit $(X_1, \dots, X_n)$ un échantillon de v.a. i.i.d de même loi que X, où X admet pour densité de probabilité :

$$f_X(x) = \frac{x}{\sigma^2} \exp\left(-\frac{x^2}{2\sigma^2}\right) \mathbb{1}_{\{x>0\}},$$

avec $\sigma \in]0, +\infty[$. Cette loi est appelée loi de Rayleigh et on utilisera la notation habituelle $X \sim \mathcal{R}(\sigma^2)$. On admettra que les premiers moments d'une loi de Rayleigh sont donnés par les relations suivantes :

$$\mathbb{E}[X^k] = \sigma^k 2^{\frac{k}{2}} \Gamma\left(1 + \frac{k}{2}\right), \qquad k \in \mathbb{N}^*.$$

Avec $\Gamma(a+1) = \int_0^{+\infty} t^a e^{-t}\, \mathrm{d}t$, $\Gamma(x+1) = x\Gamma(x)$ et $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$.

1. Vérifier que le modèle est de la famille exponentielle et donner sa paramétrisation canonique.
2. Déterminez une statistique exhaustive pour le paramètre σ^2.
3. Déterminer, sans calcul d'intégrale, l'espérance $\mathbb{E}[T(X)]$.
4. Trouver un estimateur de σ^2 par la méthode des moments.
   (i) Est-il biaisé ? Proposer un estimateur sans biais de σ^2 qu'on note $\widetilde{\sigma^2}_n$
5. Donner l'estimateur du maximum de vraisemblance $\widehat{\sigma^2}_n$.
   (i) Cet estimateur est-il biaisé ? Calculez son risque quadratique.
   (ii) Cet estimateur est-il convergent ?
6. Calculer l'information de Fisher $I_n(\sigma^2)$.

On désire effectuer le test d'hypothèses :

$$H_0 \;:\; \sigma^2 = \sigma_0^2$$

Contre

$$H_1 \;:\; \sigma^2 = \sigma_1^2$$

où σ_0^2 et σ_1^2 sont deux quantités connues.

7. Déterminer la forme de la région critique du test de Neyman-Pearson en fonction des valeurs de σ_0^2 et σ_1^2. On supposera dans la suite que $\sigma_0^2 < \sigma_1^2$.
8. Donner les expressions des risques α et β de ce test. On admet que $\sum_{i=1}^{n} X_i^2 \sim \Gamma(n, 2\sigma^2)$.

Corrigé de l'examen final

1. Nous avons :

$$\begin{aligned} L(\sigma^2, x) &:= \frac{1}{\sigma^2} \times x\mathbb{1}_{\{x>0\}} \times \exp\left(-\frac{1}{2\sigma^2} \times x^2\right) \\ &= \beta(\sigma^2) \times \xi(x) \times \exp\left(\alpha(\sigma^2) \times T(x)\right), \end{aligned}$$

La loi appartient donc à la famille exponentielle avec les notations usuelles $\beta(\sigma^2) = \frac{1}{\sigma^2}$, $\xi(x) = x\mathbb{1}_{\{x>0\}}$, $\alpha(\sigma^2) = -\frac{1}{2\sigma^2}$ et $T(x) = x^2$.
Pour la forme canonique, on utilise la mesure dominante $\nu = x\mathbb{1}_{\{x>0\}} \cdot \mu$ et le reparamétrage $\theta = -\frac{1}{2\sigma^2}$, $\Theta = \mathbb{R}_-^*$. On obtient

$$L(\theta, x) = \exp\left(\theta x^2 + \ln(-2\theta)\right)$$

et $\Psi_\nu(\theta) = -\ln(-2\theta)$, $T(x) = x^2$.

2. D'après le résultat du cours sur les familles exponentielles, on obtient que $T(X_1, \dots, X_n) = \sum_{i=1}^{n} X_i^2$ est une statistique exhaustive pour le paramètre σ^2.

3. On avait $T(x) = x^2$ et $\Psi_\nu(\theta) = -\ln(-2\theta)$ d'où

$$\mathbb{E}(X^2) = \Psi'_\nu(\theta) = -\frac{1}{\theta} = 2\sigma^2$$

4. La moyenne d'une loi de Rayleigh est

$$\mathbb{E}(X) = \sigma\sqrt{\frac{\pi}{2}} \text{ donc } \sigma^2 = \frac{2}{\pi}[\mathbb{E}(X)]^2$$

L'estimateur des moments résultant de cette dernière égalité est

$$\widetilde{\sigma^2}_M = \frac{2}{\pi}\overline{X}_n^2$$

(i) La moyenne de l'estimateur $\widetilde{\sigma^2}_M$ est

$$\begin{aligned} \mathbb{E}(\widetilde{\sigma^2}_M) &= \frac{2}{\pi}\mathbb{E}\left[\left(\overline{X}_n\right)^2\right] = \frac{2}{\pi}\mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2\right] \\ &= \frac{2}{n^2\pi}\mathbb{E}\left[\sum_{i=1}^{n} X_i \sum_{j=1}^{n} X_j\right] = \frac{2}{n^2\pi}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}\left[X_i X_j\right] \\ &= \frac{2}{n^2\pi}\left[n\mathbb{E}(X^2) + (n^2 - n)\mathbb{E}^2(X)\right] \\ &= \frac{2}{n^2\pi}\left[2n\sigma^2 + (n^2 - n)\frac{\pi}{2}\sigma^2\right] \\ &= \frac{\pi(n-1) + 4}{n\pi}\sigma^2 \end{aligned}$$

qu'on peut aussi calculer par une autre méthode

$$\begin{aligned}
\mathbb{E}(\widetilde{\sigma^2}_M) &= \frac{2}{\pi}\mathbb{E}[\overline{X}_n]^2 = \frac{2}{\pi}\left(\mathbb{V}(\overline{X}_n) + \mathbb{E}^2[\overline{X}_n]\right) \\
&= \frac{2}{\pi}\left(\frac{4-\pi}{2n}\sigma^2 + \sigma^2\frac{\pi}{2}\right) \\
&= \frac{\pi(n-1)+4}{n\pi}\sigma^2
\end{aligned}$$

On en déduit un estimateur non-biaisé du paramètre σ^2

$$\widetilde{\sigma^2}_n = \frac{n\pi}{\pi(n-1)+4}\widetilde{\sigma^2}_M = \frac{2n}{\pi(n-1)+4}[\overline{X}_n]^2$$

5. La vraisemblance de $(x_1, \ldots, x_n)$ est définie par

$$\begin{aligned}
L(x_1, \ldots, x_n; \sigma^2) &:= \prod_{i=1}^{n} f_X(x_i) = \prod_{i=1}^{n} \frac{1}{\sigma^2} \times x_i \mathbb{1}_{\{x_i>0\}} \times \exp\left(-\frac{1}{2\sigma^2} \times x_i^2\right) \\
&= \left(\frac{1}{\sigma^2}\right)^n \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n} x_i^2\right)\prod_{i=1}^{n} x_i \mathbb{1}_{\{x_i>0\}}
\end{aligned}$$

On a alors

$$\begin{aligned}
\ln L(x_1, \ldots, x_n; \sigma^2) &= -n\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n} x_i^2 + \sum_{i=1}^{n}\ln x_i \\
\frac{\partial}{\partial\sigma^2}\ln L(x_1, \ldots, x_n; \sigma^2) &= -n\frac{1}{\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n} x_i^2
\end{aligned}$$

On obtient :

$$\frac{\partial}{\partial\sigma^2}\ln L(x_1, \ldots, x_n; \sigma^2) = 0 \iff \widehat{\sigma^2}_n = \frac{1}{2n}\sum_{i=1}^{n} X_i^2$$

(i) La moyenne de l'estimateur $\widehat{\sigma^2}_n$ est

$$\begin{aligned}
\mathbb{E}[\widehat{\sigma^2}_n] &= \frac{1}{2n}\sum_{i=1}^{n}\mathbb{E}[X_i^2] \\
&= \frac{1}{2n}n(2\sigma^2) = \sigma^2
\end{aligned}$$

L'estimateur $\widehat{\sigma^2}_n$ est donc un estimateur sans biais de σ^2. La variance de l'estimateur $\widehat{\sigma^2}_n$ est

$$\begin{aligned}
\mathbb{V}[\widehat{\sigma^2}_n] &= \frac{1}{4n^2}\sum_{i=1}^{n}\mathbb{V}[X_i^2] \\
&= \frac{1}{4n^2}\sum_{i=1}^{n}\mathbb{V}[X_i^2] \\
&= \frac{1}{4n^2}n\mathbb{V}[X^2] = \frac{1}{4n}\left(\mathbb{E}[X^4] - \mathbb{E}^2[X^2]\right) \\
&= \frac{1}{4n}(8\sigma^4 - 4\sigma^4) = \frac{\sigma^4}{n}
\end{aligned}$$

Comme $\widehat{\sigma^2}_n$ est sans biais, son risque quadratique est égal a

$$\mathcal{R}(\widehat{\sigma^2}_n, \sigma^2) = \mathbb{V}[\widehat{\sigma^2}_n] = \frac{\sigma^4}{n}.$$

(ii) Cet estimateur est convergent car il est sans biais et

$$\lim_{n\to+\infty} \mathbb{V}[\widehat{\sigma^2}_n] = \lim_{n\to+\infty} \frac{\sigma^4}{n} = 0.$$

6. L'information de Fisher $I_n(\sigma^2)$:

$$\begin{aligned} I_n(\sigma^2) &= -\mathbb{E}\left[\frac{\partial^2}{\partial(\sigma^2)^2}\ln L(X_1,\dots,X_n;\sigma^2)\right] = \mathbb{E}\left[-\frac{n}{\sigma^4}+\frac{1}{\sigma^6}\sum_{i=1}^n X_i^2\right] \\ &= -\frac{n}{\sigma^4}+\frac{1}{\sigma^6}\sum_{i=1}^n \mathbb{E}\left[X_i^2\right] = -\frac{n}{\sigma^4}+\frac{2n\sigma^2}{\sigma^6} = \frac{n}{\sigma^4} \end{aligned}$$

7. Le test de Neyman-Pearson est défini par

$$\text{Rejet de } H_0 \text{ si} \qquad \frac{L(X_1,\dots,X_n;\sigma_0^2)}{L(X_1,\dots,X_n;\sigma_1^2)} < K$$

Mais

$$\begin{aligned} \frac{L(X_1,\dots,X_n;\sigma_0^2)}{L(X_1,\dots,X_n;\sigma_1^2)} < K &\iff \frac{\left(\frac{1}{\sigma_0^2}\right)^n \exp\left(-\frac{1}{2\sigma_0^2}\sum_{i=1}^n x_i^2\right)\prod_{i=1}^n x_i \mathbb{1}_{\{x_i>0\}}}{\left(\frac{1}{\sigma_1^2}\right)^n \exp\left(-\frac{1}{2\sigma_1^2}\sum_{i=1}^n x_i^2\right)\prod_{i=1}^n x_i \mathbb{1}_{\{x_i>0\}}} < K \\ &\iff \left(\frac{\sigma_1^2}{\sigma_0^2}\right)^n \exp\left(-\frac{1}{2}\left(\frac{1}{\sigma_0^2}-\frac{1}{\sigma_1^2}\right)\sum_{i=1}^n x_i^2\right) < K \\ &\iff \left(\frac{1}{\sigma_0^2}-\frac{1}{\sigma_1^2}\right)\sum_{i=1}^n x_i^2 > K' \end{aligned}$$

Pour $\sigma_1^2 > \sigma_0^2$, on a $\dfrac{1}{\sigma_0^2} - \dfrac{1}{\sigma_1^2} > 0$ et donc on en déduit le test équivalent

$$\text{Rejet de } H_0 \text{ si} \qquad \sum_{i=1}^n X_i^2 > C$$

Pour $\sigma_1^2 < \sigma_0^2$, on a $\dfrac{1}{\sigma_0^2} - \dfrac{1}{\sigma_1^2} < 0$ et donc le test s'écrit

$$\text{Rejet de } H_0 \text{ si} \qquad \sum_{i=1}^n X_i^2 < C$$

La statistique du test de Neyman-Pearson est donc

$$T_n = \sum_{i=1}^n X_i^2.$$

La région critique de ce test est définie par

— Si $\sigma_1^2 > \sigma_0^2$, $W = \left\{(X_1, \ldots, X_n) : \quad \sum_{i=1}^{n} X_i^2 > C\right\}$

— Si $\sigma_1^2 < \sigma_0^2$, $W = \left\{(X_1, \ldots, X_n) : \quad \sum_{i=1}^{n} X_i^2 < C\right\}$

8. On admet que $\sum_{i=1}^{n} X_i^2 \sim \Gamma(n, 2\sigma^2)$, de plus on a $\sigma_0^2 < \sigma_1^2$.

   Le risque de première espèce α est défini par

$$\begin{aligned}
\alpha &= \mathbb{P}\left(\text{Rejeter } H_0 | H_0 \text{ vraie}\right), \\
&= \mathbb{P}_{H_0}\left(\sum_{i=1}^{n} X_i^2 > C\right) = 1 - \mathbb{P}_{H_0}\left(\sum_{i=1}^{n} X_i^2 \leq C\right) \\
&= 1 - \Gamma_{(n, 2\sigma_0^2)}(C)
\end{aligned}$$

   De la même façon

$$\begin{aligned}
\beta &= \mathbb{P}\left(\text{Rejeter } H_1 | H_1 \text{ vraie}\right), \\
&= \mathbb{P}_{H_1}\left(\sum_{i=1}^{n} X_i^2 \leq C\right) \\
&= \Gamma_{(n, 2\sigma_1^2)}(C)
\end{aligned}$$

Examen Final

EXERCICE N° 1: 6 points

Soit $X_1, \ldots, X_n$ un échantillon que l'on modélise par une une loi uniforme $\mathcal{U}[0, \theta]$ où $\theta > 0$ est un paramètre inconnu.

1. Donner l'estimateur du maximum de vraisemblance $\hat{\theta}_n$.
2. Donnez la loi (fonctions de répartition et de densité) de $\hat{\theta}_n$.
3. Cet estimateur est-il biaisé ? Calculez son risque quadratique. Est-il convergent ?

EXERCICE N° 2: 6 points

Soit $(X_1, \ldots, X_n)$ un échantillon de v.a. i.i.d de même loi que X, où X admet pour densité de probabilité :

$$f_\theta(x) = k\sqrt{\frac{2}{\pi}} \exp\left(-\frac{x^2}{2\theta}\right) \mathbb{1}_{\{x>0\}},$$

avec $\theta \in]0, +\infty[$ un paramètre réel inconnu et k une constante à déterminer.

1. Montrer que la constante k est égale à $\frac{1}{\sqrt{\theta}}$. On rappelle que $\int_{-\infty}^{+\infty} \exp\left(-\frac{x^2}{2}\right) \, \mathrm{d}x = \sqrt{2\pi}$.
2. La loi de la variable aléatoire appartient-elle à la famille exponentielle
3. Déterminer une statistique exhaustive pour le paramètre θ.

EXERCICE N° 3: 8 points

On considère X une variable aléatoire continue, de densité :

$$f_\theta(x) = \frac{\theta}{1-\theta} x^{\frac{2\theta-1}{1-\theta}} \mathbb{1}_{]0,1[}(x)$$

où $\theta \in \left]\frac{1}{2}, 1\right[$ est un paramètre inconnu.

Soit $X_1, \ldots, X_n$ un échantillon pour cette variable aléatoire.

1. Calculez l'espérance $\mathbb{E}[X]$ et la variance $\mathbb{V}[X]$ de la variable aléatoire X.
2. Trouver un estimateur $\hat{\theta}_{n1}$ de θ par la méthode des moments.
   (i) Cet estimateur est-il biaisé ? Calculez son risque quadratique.
   (ii) Cet estimateur est-il convergent ?
3. Soit la variable aléatoire $Y = \log X$. Donnez sa fonction de répartition et sa densité.
4. Déterminer l'estimateur du maximum de vraisemblance $\hat{\theta}_{n2}$. Écrivez l'estimateur obtenu en fonction des variables $Y_1, \ldots, Y_n$.

Corrigé de l'Examen Final

EXERCICE N° 1: 6 points
Soit $X_1, \ldots, X_n$ un échantillon que l'on modélise par une loi uniforme $\mathcal{U}[0, \theta]$ où $\theta > 0$ est un paramètre inconnu.

1. On écrit la vraisemblance :

$$\begin{aligned}\mathcal{L}(x_1, \ldots, x_n, \theta) &= \prod_{i=1}^{n} \frac{1}{\theta} \mathbb{1}_{[0,\theta]}(x_i) \\ &= \left(\frac{1}{\theta}\right)^n \prod_{i=1}^{n} \mathbb{1}_{\{0 \le x_i \le \theta\}} \\ &= \left(\frac{1}{\theta}\right)^n \mathbb{1}_{\left\{\max\limits_{1 \le i \le n} x_i \le \theta\right\}} \mathbb{1}_{\left\{\min\limits_{1 \le i \le n} x_i \ge 0\right\}}\end{aligned}$$

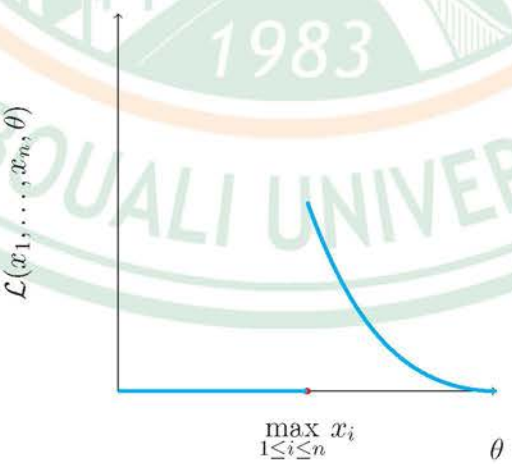
Pour trouver l'estimateur du maximum de vraisemblance, il nous faut maximiser la fonction de vraisemblance c'est à dire trouver la valeur de θ qui maximise

$$\left(\frac{1}{\theta}\right)^n \mathbb{1}_{\left\{\max\limits_{1 \le i \le n} x_i \le \theta\right\}}$$

$$\theta^{-n} \mathbb{1}_{\left\{\max\limits_{1 \le i \le n} x_i \le \theta\right\}}$$

Or la fonction $g(\theta) = \theta^{-n}$ définie pour $\theta \ge 0$ a pour dérivée la fonction $g'(\theta) = -n\,\theta^{-n-1}$. Comme $g'(\theta) < 0$ sur $[0, +\infty[$ cela implique que $g(x)$ est décroissante sur $[0, +\infty[$. Revenons à la fonction de vraisemblance. On a vu que :

$$\begin{aligned}\mathcal{L}(x_1, \ldots, x_n, \theta) &= 0 \quad \text{si } \theta < \max_{1 \le i \le n} x_i \\ &= \theta^{-n} \quad \text{si } \theta \ge \max_{1 \le i \le n} x_i\end{aligned}$$



Comme nous avons montré que θ^{-n} est décroissante sur $[0, +\infty[$. Nous pouvons conclure que la fonction de vraisemblance $\mathcal{L}(x_1, \dots, x_n, \theta)$ atteint son maximum lorsque θ vaut $\max\limits_{1\le i\le n} x_i$.

Ceci implique que $\hat{\theta}_n = \max\limits_{1\le i\le n} X_i$ est l'estimateur du maximum de vraisemblance pour la loi uniforme $\mathcal{U}[0, \theta]$.

2. Calculons la loi de $\hat{\theta}_n$. On a :

$$\begin{aligned}
\mathbb{F}_{\hat{\theta}_n}(x) &= \mathbb{P}\left(\hat{\theta}_n \le x\right) \\
&= \mathbb{P}\left(\max_{1\le i\le n} X_i \le x\right) \\
&= \mathbb{P}\left(X_1 \le x, X_2 \le x, \dots, X_n \le x\right) \\
&= \mathbb{P}\left(X_1 \le x\right)\mathbb{P}\left(X_2 \le x\right)\dots\mathbb{P}\left(X_n \le x\right) \\
&= \left(\frac{x}{\theta}\right)^n
\end{aligned}$$

Ainsi, on obtient la densité de $\hat{\theta}_n$:

$$f_{\hat{\theta}_n}(x) = n\frac{x^{n-1}}{\theta^n}$$

3.

$$\begin{aligned}
\mathbb{E}[\hat{\theta}_n] &= \int_0^\theta nx\frac{x^{n-1}}{\theta^n}\,\mathrm{d}x = \frac{n}{\theta^n}\int_0^\theta x^n\,\mathrm{d}x \\
&= \frac{n}{n+1}\theta.
\end{aligned}$$

On remarque que cet estimateur est asymptotiquement sans biais. En effet,

$$\lim_{n\longrightarrow+\infty} \mathbb{E}[\hat{\theta}_n] = \lim_{n\longrightarrow+\infty} \frac{n}{n+1}\theta = \theta.$$

Pour calculer la variance de $\hat{\theta}_n$, on calcule d'abord

$$\begin{aligned}
\mathbb{E}[(\hat{\theta}_n)^2] &= \int_0^\theta nx^2\frac{x^{n-1}}{\theta^n}\,\mathrm{d}x = \frac{n}{\theta^n}\int_0^\theta x^{n+1}\,\mathrm{d}x \\
&= \frac{n}{n+2}\theta^2.
\end{aligned}$$

On en déduit que :

$$\mathbb{V}[\hat{\theta}_n] = \frac{n}{n+2}\theta^2 - \frac{n^2}{(n+1)^2}\theta^2 = \frac{n}{(n+2)(n+1)^2}\theta^2.$$

Son risque quadratique est égal a

$$\mathcal{R}(\hat{\theta}_n, \theta) = \mathbb{V}[\hat{\theta}_n] + \text{Biais}^2[\hat{\theta}_n] = \frac{n}{(n+2)(n+1)^2}\theta^2 + \frac{1}{(n+1)^2}\theta^2 = \frac{2\theta^2}{(n+1)(n+2)}.$$

Cet estimateur est convergent car il est asymptotiquement sans biais et

$$\lim_{n\to+\infty} \mathbb{V}[\hat{\theta}_n] = \lim_{n\to+\infty} \frac{n}{(n+2)(n+1)^2}\theta^2 = 0.$$

EXERCICE N° 2: 6 points

Soit $(X_1, \ldots, X_n)$ un échantillon de v.a. i.i.d de même loi que X, où X admet pour densité de probabilité :

$$f_\theta(x) = k\sqrt{\frac{2}{\pi}} \exp\left(-\frac{x^2}{2\theta}\right) \mathbb{1}_{\{x>0\}},$$

avec $\theta \in]0, +\infty[$ un paramètre réel inconnu et k une constante à déterminer.

1. Pour que f_θ soit une densité, il faut que

$$\begin{aligned}
&\int_{-\infty}^{+\infty} f_\theta(x)\,\mathrm{d}x = 1 \\
&\qquad \Longrightarrow \int_{-\infty}^{+\infty} k\sqrt{\frac{2}{\pi}} \exp\left(-\frac{x^2}{2\theta}\right) \mathbb{1}_{\{x>0\}}\,\mathrm{d}x = 1 \\
&\qquad \Longrightarrow k\sqrt{\frac{2}{\pi}}\sqrt{\theta} \underbrace{\int_{0}^{+\infty} \exp\left(-\frac{u^2}{2}\right)\mathrm{d}u}_{=\sqrt{\frac{\pi}{2}}} = 1 \\
&\qquad \Longrightarrow k\sqrt{\frac{2}{\pi}}\sqrt{\theta}\sqrt{\frac{\pi}{2}} = 1 \\
&\qquad \Longrightarrow k = \frac{1}{\sqrt{\theta}}.
\end{aligned}$$

2. La loi de la variable aléatoire appartient à la famille exponentielle si on peut écrire la densité sous la forme :

$$f_\theta(x) = \exp\left\{a(x)\alpha(\theta) + b(x) + \beta(\theta)\right\}.$$

On a :

$$\begin{aligned}
f_\theta(x) &= \sqrt{\frac{2}{\pi}}\frac{1}{\sqrt{\theta}} \exp\left(-\frac{x^2}{2\theta}\right) \mathbb{1}_{\{x>0\}} \\
&= \exp\left(-\frac{x^2}{2\theta} + \ln\sqrt{\frac{2}{\pi}} - \frac{1}{2}\ln\theta\right) \mathbb{1}_{\{x>0\}}
\end{aligned}$$

La loi de X appartient à la famille exponentielle avec

$$a(x) = x^2,\ \alpha(\theta) = -\frac{1}{2\theta},\ b(x) = \ln\sqrt{\frac{2}{\pi}} \text{ et } \beta(\theta) = -\frac{1}{2}\ln\theta.$$

3. On écrit la vraisemblance :

$$\begin{aligned}
\mathcal{L}(x_1, \ldots, x_n, \theta) &= \prod_{i=1}^{n} f_\theta(x_i) \\
&= \left(\frac{1}{\sqrt{\theta}}\right)^n \left(\sqrt{\frac{2}{\pi}}\right)^n \exp\left(-\frac{1}{2\theta}\sum_{i=1}^{n} x_i^2\right) \prod_{i=1}^{n} \mathbb{1}_{\{x_i>0\}} \\
&= \underbrace{\left(\frac{1}{\sqrt{\theta}}\right)^n \left(\sqrt{\frac{2}{\pi}}\right)^n \exp\left(-\frac{1}{2\theta}\sum_{i=1}^{n} x_i^2\right)}_{g(\theta, T(x_1,\ldots,x_n))} \underbrace{\prod_{i=1}^{n} \mathbb{1}_{\{x_i>0\}}}_{h(x_1,\ldots,x_n)}.
\end{aligned}$$

Par le critère de factorisation Neyman-Fisher, on en déduit que $T(X_1, \dots, X_n) = \sum_{i=1}^{n} X_i^2$ est une statistique exhaustive pour le paramètre θ.

EXERCICE N° 3: 8 points

On considère X une variable aléatoire continue, de densité :

$$f_\theta(x) = \frac{\theta}{1-\theta} x^{\frac{2\theta-1}{1-\theta}} \mathbb{1}_{]0,1[}(x)$$

où $\theta \in \left]\frac{1}{2}, 1\right[$ est un paramètre inconnu.

Soit $X_1, \dots, X_n$ un échantillon pour cette variable aléatoire.

1. L'espérance :

$$\begin{aligned}
\mathbb{E}[X] &= \int_{-\infty}^{+\infty} x f_\theta(x)\, \mathrm{d}x \\
&= \int_{-\infty}^{+\infty} x \frac{\theta}{1-\theta} x^{\frac{2\theta-1}{1-\theta}} \mathbb{1}_{]0,1[}(x)\, \mathrm{d}x = \int_0^1 \frac{\theta}{1-\theta} x^{\frac{\theta}{1-\theta}}\, \mathrm{d}x \\
&= \frac{\theta}{1-\theta}(1-\theta) x^{\frac{1}{1-\theta}} \Big]_0^1 = \theta.
\end{aligned}$$

Pour calculer la variance, on calcule d'abord

$$\begin{aligned}
\mathbb{E}[X^2] &= \int_{-\infty}^{+\infty} x^2 f_\theta(x)\, \mathrm{d}x \\
&= \int_{-\infty}^{+\infty} x^2 \frac{\theta}{1-\theta} x^{\frac{2\theta-1}{1-\theta}} \mathbb{1}_{]0,1[}(x)\, \mathrm{d}x = \int_0^1 \frac{\theta}{1-\theta} x^{\frac{1}{1-\theta}}\, \mathrm{d}x \\
&= \frac{\theta}{1-\theta} \frac{1-\theta}{2-\theta} x^{\frac{2-\theta}{1-\theta}} \Big]_0^1 = \frac{\theta}{2-\theta}.
\end{aligned}$$

On en déduit que :

$$\mathbb{V}[X] = \frac{\theta}{2-\theta} - \theta^2.$$

2. Un estimateur de θ par la méthode des moments est obtenu en résolvant

$$\mathbb{E}[X] = \overline{X}_n \Longrightarrow \theta = \overline{X}_n.$$

L'estimateur obtenu par la méthode des moments est alors $\hat{\theta}_{n1} = \overline{X}_n$.

(i) Cet estimateur est sans biais, en effet

$$\mathbb{E}[\hat{\theta}_{n1}] = \mathbb{E}[\overline{X}_n] = \mathbb{E}[X] = \theta.$$

La variance :

$$\mathbb{V}[\hat{\theta}_{n1}] = \mathbb{V}[\overline{X}_n] = \frac{1}{n}\mathbb{V}[X] = \frac{1}{n}\left(\frac{\theta}{2-\theta} - \theta^2\right).$$

Comme $\hat{\theta}_{n1}$ est sans biais, son risque quadratique est égal a

$$\mathcal{R}(\hat{\theta}_{n1}, \theta) = \mathbb{V}[\hat{\theta}_{n1}] = \frac{1}{n}\left(\frac{\theta}{2-\theta} - \theta^2\right).$$

(ii) Cet estimateur est convergent car il est sans biais et

$$\lim_{n\to+\infty} \mathbb{V}[\hat{\theta}_{n1}] = \lim_{n\to+\infty} \frac{1}{n}\left(\frac{\theta}{2-\theta} - \theta^2\right) = 0.$$

3. Soit la variable aléatoire $Y = \log X$. La fonction de répartition de Y :

$$\begin{aligned}\mathbb{F}_Y(y) &:= \mathbb{P}(Y \leq y) = \mathbb{P}(\ln X \leq y)\\ &= \mathbb{P}(X \leq \mathrm{e}^y) = \mathbb{F}_X(\mathrm{e}^y)\end{aligned}$$

De plus

$$\mathbb{F}_X(x) = \int_{-\infty}^{x} f_\theta(t)\,\mathrm{d}t = \int_{0}^{x} \frac{\theta}{1-\theta} t^{\frac{2\theta-1}{1-\theta}}\,\mathrm{d}t = \frac{\theta}{1-\theta}\frac{1-\theta}{\theta} t^{\frac{\theta}{1-\theta}}\Big]_0^x$$

Alors

$$\mathbb{F}_X(x) = \begin{cases} 0 & \text{si } x < 0 \\ x^{\frac{\theta}{1-\theta}} & \text{si } 0 \leq x \leq 1 \\ 1 & \text{si } x > 1 \end{cases}$$

Donc la fonction de répartition de Y est donnée par :

$$\mathbb{F}_Y(y) = \begin{cases} e^{\frac{\theta}{1-\theta}y} & \text{si } y \leq 0 \\ 1 & \text{si } y > 0 \end{cases}$$

La densité de Y :

$$f_Y(y) = \mathbb{F}'_Y(y) = \frac{\theta}{1-\theta} e^{\frac{\theta}{1-\theta}y}\mathbb{1}_{\{y<0\}}.$$

4. On écrit la vraisemblance :

$$\begin{aligned}\mathcal{L}(x_1,\ldots,x_n,\theta) &= \prod_{i=1}^{n} f_\theta(x_i)\\ &= \prod_{i=1}^{n} \frac{\theta}{1-\theta} x_i^{\frac{2\theta-1}{1-\theta}} \mathbb{1}_{]0,1[}(x_i)\\ &= \left(\frac{\theta}{1-\theta}\right)^n \prod_{i=1}^{n} x_i^{\frac{2\theta-1}{1-\theta}} \mathbb{1}_{]0,1[}(x_i)\end{aligned}$$

La log-vraisemblance :

$$\ln \mathcal{L}(x_1,\ldots,x_n,\theta) = n\ln\left(\frac{\theta}{1-\theta}\right) + \frac{2\theta-1}{1-\theta}\sum_{i=1}^{n} \ln x_i + \ln\prod_{i=1}^{n} \mathbb{1}_{]0,1[}(x_i)$$

La dérivée par rapport à θ est :

$$\frac{\partial \ln \mathcal{L}(x_1,\ldots,x_n,\theta)}{\partial\theta} = \frac{n}{\theta(1-\theta)} + \frac{1}{(1-\theta)^2}\sum_{i=1}^{n} \ln x_i$$

On pose :

$$\begin{aligned}\frac{\partial \ln \mathcal{L}(x_1,\ldots,x_n,\theta)}{\partial\theta} = 0 &\Longrightarrow \frac{\theta-1}{\theta} = \frac{1}{n}\sum_{i=1}^{n} \ln x_i = \overline{y}_n\\ &\Longrightarrow \theta(1-\overline{y}_n) = 1\\ &\Longrightarrow \theta = \frac{1}{1-\overline{y}_n}\end{aligned}$$

Donc, l'estimateur du maximum de vraisemblance de θ est :

$$\hat{\theta}_{n2} = \frac{1}{1-\overline{Y}_n}.$$