TD 1 Statistique - 1HY

Exercice 1. Durée d'attente à un feu rouge.
On considère la variable aléatoire "durée d'attente à un feu rouge". La durée maximale d'attente à ce feu rouge est notée θ, paramètre inconnu strictement positif. On observe un échantillon $t_1, ..., t_n$ de taille n, où t_i désigne la durée d'attente observée pour le $i^{\text{ème}}$ individu. On fait l'hypothèse que les variables aléatoires T_i associées aux observations t_i sont indépendantes et de même loi uniforme sur $[0, \theta]$, i.e., $T_i \sim U[0, \theta]$.

1. Représenter le graphe de la densité de la loi $U[0, \theta]$ et préciser ses paramètres moyenne et variance.
2. On désire estimer le paramètre θ. Déterminer le biais et la variance de la statistique $\overline{T} = \frac{1}{n}\sum_{i=1}^{n} T_i$.
   Montrer que $\widehat{\theta}_1 = 2\overline{T}$ est un estimateur sans biais et convergent en probabilité de θ.
3. Montrer que l'estimateur du maximum de vraisemblance de θ est $Y_n = \sup_i T_i$.
   - En utilisant l'équivalence des événements suivante $Y_n < y \iff T_i < y, \forall i = 1, ..., n$, calculer la fonction de répartition de Y_n. En déduire sa densité et calculer $E[Y_n]$ et $\operatorname{var}(Y_n)$.
   - Montrer que la statistique $\widehat{\theta}_2 = \frac{n+1}{n} Y_n$ est un estimateur sans biais et convergent en probabilité de θ.
4. Lequel des deux estimateurs $\widehat{\theta}_1$ et $\widehat{\theta}_2$ choisiriez-vous pour estimer θ ?

Exercice 2. La durée de fonctionnement d'un matériel électrique est représentée par une variable aléatoire réelle X suivant une loi de Weibull de densité :

$$f(x; \theta, \lambda) = \frac{\lambda}{\theta} x^{\lambda - 1} \exp\left\{-\frac{x^\lambda}{\theta}\right\} \qquad x > 0$$

avec $\theta > 0$ et $\lambda > 0$. On suppose que λ est connu.

1. Déterminer la loi de $U = X^\lambda$ puis calculer $E\left(X^\lambda\right)$ et $\operatorname{Var}\left(X^\lambda\right)$.
2. On considère un échantillon $(X_1, ..., X_n)$ de même loi que X. Calculer l'estimateur du maximum de vraisemblance $\widehat{\theta}_n$ de θ. Cet estimateur est-il sans biais ? convergent ? efficace ? Calculer son erreur quadratique moyenne.

Exercice 3. Estimation de l'écart type
Soient $x_1, ..., x_n$ les réalisations respectives de n variables aléatoires $X_1, ..., X_n$ indépendantes et de même loi gaussienne $\mathcal{N}(0, \sigma^2)$.

1. Déterminer un minorant de la variance de tout estimateur sans biais de σ.
2. Déterminer a pour que
   $$\widehat{\sigma} = \frac{a}{n}\sum_{i=1}^{n} |X_i|$$
   soit un estimateur sans biais de σ. L'estimateur obtenu $\widehat{\sigma}$ est-il convergent ? efficace ?

Exercice 4. Soient $X_1, ..., X_n$, n variables aléatoires indépendantes de même loi de densité :

$$f(x) = \beta e^{\beta(\alpha - x)} 1_{[\alpha, +\infty[}(x)$$

1. α étant connu, déterminer l'estimateur du maximum de vraisemblance de $\omega = 1/\beta$ noté $\widehat{\omega}$. Vérifier qu'il est sans biais et convergent. Montrer enfin que $\widehat{\omega}$ est l'estimateur efficace de ω.

2. β étant connu, étudier l'estimateur du maximum de vraisemblance de α noté $\widehat{\alpha}$. On admet que la densité de probabilité de $\widehat{\alpha}$ est :

   $$\pi(u) = n\beta e^{n\beta(\alpha - u)} 1_{[\alpha, +\infty[}(u)$$

   En s'aidant de ce qui a été fait à la première question, déterminer le biais et la variance de $\widehat{\alpha}$. En déduire un estimateur sans biais et convergent de α noté $\widetilde{\alpha}$. Que dire de l'efficacité de $\widetilde{\alpha}$?

3. β étant connu, déterminer l'estimateur de α obtenu à l'aide de la méthode des moments appliquée à $E[X_i]$ noté $\overline{\alpha}$. Comparer les deux estimateurs $\overline{\alpha}$ et $\widehat{\alpha}$.

Exercice 1

1) La densité de T_i est

$$f(t_i;\theta)=\begin{cases}\frac{1}{\theta}\text{ si } t_i\in[0,\theta]\\ 0\text{ sinon}\end{cases}$$

On a de plus (voir tables ou calculs élémentaires)

$$E[T_i]=\frac{\theta}{2}\text{ et var }[T_i]=\frac{\theta^2}{12}$$

2)

$$\begin{aligned}E\left[\bar{T}\right] &= E[T_1]=\frac{\theta}{2}\\ var\left[\bar{T}\right] &= \frac{var[T_1]}{n}=\frac{\theta^2}{12n}\end{aligned}$$

Donc

$$\begin{aligned}E\left[\widehat{\theta}_1\right] &= 2E\left[\bar{T}\right]=\theta\\ \text{var}\left[\widehat{\theta}_1\right] &= 4\text{var}\left[\bar{T}\right]=\frac{\theta^2}{3n}\underset{n\to\infty}{\to}0\end{aligned}$$

$\widehat{\theta}_1$ est donc un estimateur non biaisé et convergent de θ.

3) La vraisemblance de $t_1,...,t_n$ est

$$f(t_1,...,t_n;\theta)=\begin{cases}\frac{1}{\theta^n}\text{ si } t_i\in[0,\theta],\forall i=1,...,n\\ 0\text{ sinon}\end{cases}$$

Maximiser la vraisemblance de $t_1,...,t_n$ revient à maximiser $\frac{1}{\theta^n}$ sous les contraintes $t_i\in[0,\theta],\forall i=1,...,n$. Puisque $\frac{1}{\theta^n}$ est une fonction décroissante de θ et que les contraintes impose $t_i\leq\theta,\forall i=1,...,n$, on en déduit

$$Y_n=\sup_i T_i$$

a) On a

$$\begin{aligned}P[Y_n<y] &= P[T_1<y\text{ et }T_2<y\text{ ... et }T_n<y]\\ &= \prod_{k=1}^{n}P[T_k<y]\\ &= P[T_1<y]^n\\ &= \begin{cases}\left(\frac{y}{\theta}\right)^n\text{ si } y\in[0,\theta]\\ 1\text{ si } y\geq\theta\\ 0\text{ si } y\leq 0\end{cases}\end{aligned}$$

La densité de Y_n est donc

$$f(y)=\begin{cases}n\frac{y^{n-1}}{\theta^n}\text{ si } y\in[0,\theta]\\ 0\text{ sinon}\end{cases}$$

On en déduit

$$\begin{aligned}E[Y_n] &= \int_0^\theta n\frac{y^n}{\theta^n}dy=\frac{n\theta}{n+1}\\ E\left[Y_n^2\right] &= \int_0^\theta n\frac{y^{n+1}}{\theta^n}dy=\frac{n\theta^2}{n+2}\end{aligned}$$

et donc

$$\text{var}[Y_n] = \frac{n\theta^2}{(n+2)(n+1)^2}.$$

$\widehat{\theta}_2 = \frac{n+1}{n} Y_n$ est donc un estimateur non biaisé de θ. Puisque

$$\text{var}\left[\widehat{\theta}_2\right] = \left(\frac{n+1}{n}\right)^2 \text{var}[Y_n] = \frac{\theta^2}{n(n+2)} \underset{n\to\infty}{\to} 0$$

$\widehat{\theta}_2 = \frac{n+1}{n} Y_n$ est un estimateur convergent de θ.

4) Puisque la variance de $\widehat{\theta}_2$ est plus faible que la variance de $\widehat{\theta}_1$ (au moins pour n grand car $\text{var}\left[\widehat{\theta}_2\right] \simeq \frac{\theta^2}{n^2}$ et $\text{var}\left[\widehat{\theta}_1\right] = \frac{\theta^2}{3n}$), on choisira l'estimateur $\widehat{\theta}_2$.

Exercice 2

1) un changement de variables élémentaire conduit à

$$f(u) = \begin{cases} \frac{1}{\theta}\exp\left(-\frac{u}{\theta}\right) \text{ pour } u > 0 \\ 0 \text{ si } u \leq 0 \end{cases}$$

Des calculs simples ou les tables de la loi gamma permettent d'obtenir

$$E[U] = \theta \text{ et } \text{var}[U] = \theta^2$$

2) La vraisemblance de $t_1, ..., t_n$ est

$$\begin{aligned} f(x_1, ..., x_n; \theta) &= \prod_{k=1}^{n} \left[\frac{\lambda}{\theta} x_k^{\lambda-1} \exp\left\{-\frac{x_k^\lambda}{\theta}\right\}\right] \qquad x_i > 0 \\ &= \left(\frac{\lambda}{\theta}\right)^n \left[\prod_{k=1}^{n} x_k\right]^{\lambda-1} \exp\left\{-\frac{1}{\theta}\sum_{k=1}^{n} x_k^\lambda\right\} \end{aligned}$$

et son logarithme

$$\ln f(x_1, ..., x_n; \theta) = n\ln\lambda - n\ln\theta + (\lambda-1)\ln\left[\prod_{k=1}^{n} x_k\right] - \frac{1}{\theta}\sum_{k=1}^{n} x_k^\lambda$$

Puisque le domaine de définition de la vraisemblance est indépendant de θ, la recherche du maximum de vraisemablance de θ se fait comme suit

$$\begin{aligned} \frac{\partial \ln f(x_1, ..., x_n; \theta)}{\partial\theta} &\geq 0 \Leftrightarrow -\frac{n}{\theta} + \frac{1}{\theta^2}\sum_{k=1}^{n} x_k^\lambda \geq 0 \\ &\Leftrightarrow \theta \leq \frac{1}{n}\sum_{k=1}^{n} x_k^\lambda \end{aligned}$$

d'où

$$\hat{\theta}_n = \frac{1}{n}\sum_{k=1}^{n} X_k^\lambda$$

En remarquant que $\hat{\theta}_n = \frac{1}{n}\sum_{k=1}^{n} U_k$, on a

$$\begin{aligned} E\left[\hat{\theta}_n\right] &= E[U_1] = \theta \\ \text{var}\left[\hat{\theta}_n\right] &= \frac{\text{var}(U_1)}{n} = \frac{\theta^2}{n} \end{aligned}$$

Donc $\hat{\theta}_n$ est un estimateur sans biais et convergent de θ. La borne de Cramér-Rao d'un estimateur non biaisé de θ est

$$\begin{aligned}
\text{BCR}\,(\theta) &= \frac{-1}{E\left[\frac{\partial^2 \ln f(x_1,...,x_n;\theta)}{\partial\theta^2}\right]} \\
&= \frac{-1}{\frac{n}{\theta^2} - \frac{2}{\theta^3}E\left[\sum_{k=1}^n X_k^\lambda\right]} \\
&= \frac{-1}{\frac{n}{\theta^2} - \frac{2}{\theta^3}n\theta} = \frac{\theta^2}{n}
\end{aligned}$$

Puisque $\text{var}(\hat{\theta}_n) = \text{BCR}\,(\theta)$, $\hat{\theta}_n$ est l'estimateur efficace de θ. Son erreur quadratique moyenne est

$$E\left[\left(\hat{\theta}_n - \theta\right)^2\right] = \text{var}\left[\hat{\theta}_n\right] + \text{biais}^2\left[\hat{\theta}_n\right] = \frac{\theta^2}{n}$$

Exercice 3

1) Un minorant de la variance de tout estimateur non biaisé de σ est donné par la borne de Cramér-Rao

$$\text{BCR}\,(\sigma) = \frac{-1}{E\left[\frac{\partial^2 \ln f(x_1,...,x_n;\sigma)}{\partial\sigma^2}\right]}.$$

La vraisemblance de l'échantillon gaussien est

$$f(x_1,...,x_n;\sigma) = \frac{1}{\sigma^n\left(\sqrt{2\pi}\right)^n}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n x_i^2\right)$$

d'où la log-vraisemblance

$$\ln f(x_1,...,x_n;\sigma) = -n\ln(\sigma) - \frac{n}{2}\ln(2\pi) - \frac{1}{2\sigma^2}\sum_{i=1}^n x_i^2.$$

Les dérivées de cette log-vraisemblance se calculent facilement

$$\begin{aligned}
\frac{\partial \ln f(x_1,...,x_n;\sigma)}{\partial\sigma} &= -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^n x_i^2 \\
\frac{\partial^2 \ln f(x_1,...,x_n;\sigma)}{\partial\sigma^2} &= \frac{n}{\sigma^2} - \frac{3}{\sigma^4}\sum_{i=1}^n x_i^2
\end{aligned}$$

ce qui permet d'obtenir

$$E\left[\frac{\partial^2 \ln f(x_1,...,x_n;\sigma)}{\partial\sigma^2}\right] = \frac{n}{\sigma^2} - \frac{3}{\sigma^4}\times(n\sigma^2) = -\frac{2n}{\sigma^2}.$$

La borne de Cramér-Rao d'un estimateur non biaisé de σ est donc

$$\text{BCR}\,(\sigma) = \frac{\sigma^2}{2n}.$$

2) Pour que

$$\widehat{\sigma} = \frac{a}{n}\sum_{i=1}^n |X_i|$$

soit un estimateur sans biais de σ, il faut

$$E[\widehat{\sigma}] = \sigma$$

c'est-à-dire

$$E[\widehat{\sigma}] = \frac{a}{n}\sum_{i=1}^{n} E[|X_i|] = aE[|X_1|] = \sigma.$$

Quelques calculs élémentaires permettent de calculer $E[|X_1|]$

$$\begin{aligned}
E[|X_1|] &= \int_{-\infty}^{\infty} \frac{|u|}{\sigma\sqrt{2\pi}} \exp\left(-\frac{u^2}{2\sigma^2}\right) du \\
&= \frac{2}{\sigma\sqrt{2\pi}} \int_{0}^{\infty} u \exp\left(-\frac{u^2}{2\sigma^2}\right) du \\
&= \sqrt{\frac{2}{\pi}}\sigma
\end{aligned}$$

d'où

$$a = \sqrt{\frac{\pi}{2}}$$

et

$$\widehat{\sigma} = \frac{1}{n}\sqrt{\frac{\pi}{2}}\sum_{i=1}^{n} |X_i|.$$

La variance de cet estimateur est

$$\text{var}[\widehat{\sigma}] = \frac{a^2}{n}\text{var}[|X_1|].$$

La variance de $|X_1|$ se calcule comme suit

$$\text{var}[|X_1|] = E[X_1^2] - E^2[|X_1|] = \sigma^2 - \frac{2}{\pi}\sigma^2 = \left(1 - \frac{2}{\pi}\right)\sigma^2$$

d'où

$$\text{var}[\widehat{\sigma}] = \left(\frac{\pi}{2} - 1\right)\frac{\sigma^2}{n}.$$

L'estimateur $\widehat{\sigma}$ est donc convergent mais il n'est pas efficace. On montrerait que

$$\widetilde{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} X_i^2}$$

est l'estimateur efficace du paramètre σ.

Exercice 4

1) La vraisemblance de $x_1, ..., x_n$ est

$$\begin{aligned}
f(x_1, ..., x_n; \omega) &= \prod_{k=1}^{n}\left[\frac{1}{\omega}\exp\left\{\frac{\alpha - x_k}{\omega}\right\}\right] \qquad x_k > \alpha \\
&= \left(\frac{1}{\omega}\right)^n \exp\left\{-\frac{1}{\omega}\left(\sum_{k=1}^{n} x_k - n\alpha\right)\right\}
\end{aligned}$$

et son logarithme

$$\ln f(x_1, ..., x_n; \omega) = -n\ln\omega + -\frac{1}{\omega}\left(\sum_{k=1}^{n} x_k - n\alpha\right)$$

Puisque le domaine de définition de la vraisemblance est indépendant de ω, la recherche du maximum de vraisemablance de ω se fait comme suit

$$\begin{aligned}
\frac{\partial \ln f(x_1,...,x_n;\omega)}{\partial \omega} &\geq 0 \Leftrightarrow -\frac{n}{\omega}+\frac{1}{\omega^2}\left(\sum_{k=1}^{n} x_k - n\alpha\right) \geq 0 \\
&\Leftrightarrow \omega \leq \frac{1}{n}\sum_{k=1}^{n} x_k - \alpha
\end{aligned}$$

d'où

$$\hat{\omega} = \frac{1}{n}\sum_{k=1}^{n} X_k - \alpha$$

On montre aisément que $U_k = X_k - \alpha$ suit une loi exponentielle de densité

$$f(u) = \begin{cases} \frac{1}{\omega}\exp\left(-\frac{u}{w}\right) \text{ pour } u>0 \\ 0 \text{ si } u \leq 0 \end{cases}$$

Donc comme dans l'exercice précédent

$$\begin{aligned}
E\left[\hat{\omega}\right] &= E\left[U_1\right] = \omega \\
\text{var}\left[\hat{\omega}\right] &= \frac{\text{var}\left(U_1\right)}{n} = \frac{\omega^2}{n}
\end{aligned}$$

Donc $\hat{\omega}$ est un estimateur sans biais et convergent de ω. La borne de Cramer-Rao d'un estimateur non biaisé de ω est

$$\text{BCR}\left(\omega\right) = \frac{-1}{E\left[\frac{\partial^2 \ln f(x_1,...,x_n;\omega)}{\partial \omega^2}\right]} = \frac{\omega^2}{n}$$

Puisque $\text{var}\left[\hat{\omega}\right] = \text{BCR}\left(\omega\right)$, $\hat{\omega}$ est l'estimateur efficace de ω.

2) La vraisemblance de $x_1,...,x_n$ est la même que précédemment mais elle est maintenant paramétrée par α

$$\begin{aligned}
f(x_1,...,x_n;\alpha) &= \prod_{k=1}^{n}\left[\frac{1}{\omega}\exp\left\{\frac{\alpha - x_k}{\omega}\right\}\right] 1_{[\alpha,+\infty[}(x_i) \\
&= \left(\frac{1}{\omega}\right)^n \exp\left\{-\frac{1}{\omega}\left(\sum_{k=1}^{n} x_k - n\alpha\right)\right\}\prod_{k=1}^{n} 1_{[\alpha,+\infty[}(x_i) \\
&= \left(\frac{1}{\omega}\right)^n \exp\left\{-\frac{1}{\omega}\sum_{k=1}^{n} x_k + \frac{n\alpha}{\omega}\right\}\prod_{k=1}^{n} 1_{[\alpha,+\infty[}(x_i).
\end{aligned}$$

La vraisemblance est une fonction croissante de α mais α doit vérifier les contraintes

$$x_i \geq \alpha, \; \forall i = 1,...,n.$$

On en déduit

$$\widehat{\alpha}_{\text{MV}} = \min_{i=1,...,n} X_i.$$

La densité de $\min_{i=1,...,n} X_i$ était donnée dans l'énoncé

$$\pi\left(u\right) = n\beta e^{n\beta(\alpha-u)} 1_{[\alpha,+\infty[}(u).$$

On remarque qu'elle est similaire à celle de X_i mais que β devient $n\beta$. On en déduit que la moyenne de $\widehat{\alpha}_{\text{MV}}$ est égale à l'expression de $E[X_i]$ dans laquelle on a remplacé β par $n\beta$, soit

$$E[\widehat{\alpha}_{\text{MV}}] = \alpha + \frac{1}{n\beta}.$$

De même

$$\mathrm{var}[\widehat{\alpha}_{\mathrm{MV}}] = \frac{1}{n^2\beta^2}.$$

On en déduit que

$$\widetilde{\alpha} = \widehat{\alpha}_{\mathrm{MV}} - \frac{1}{n\beta} = \min_{i=1,\dots,n} X_i - \frac{1}{n\beta}$$

est un estimateur non biaisé de α. La variance de cet estimateur est

$$\mathrm{var}[\widetilde{\alpha}] = \frac{1}{n^2\beta^2}.$$

Comme cette variance tend vers 0 lorsque $n \to \infty$ et que l'estimateur $\widetilde{\alpha}$ est non biaisé, il est convergent. Comme les bornes du domaine de définition de la densité de X_i dépendent de α, il n'existe pas de borne de Cramér-Rao et donc l'efficacité de $\widetilde{\alpha}$ n'a pas de sens.

3) La moyenne de X_i est

$$E[X_i] = \alpha + \frac{1}{\beta}.$$

L'estimateur de α obtenu à l'aide de la méthode des moments appliquée à $E[X_i]$ est

$$\overline{\alpha} = \frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{\beta}.$$

Cet estimateur est non biaisé puisque

$$E[\overline{\alpha}] = -\frac{1}{\beta} + \frac{1}{n}\sum_{i=1}^{n} E[X_i] = -\frac{1}{\beta} + \frac{1}{n}\sum_{i=1}^{n}\left(\alpha + \frac{1}{\beta}\right) = \alpha.$$

La variance de l'estimateur $\overline{\alpha}$ est

$$\mathrm{var}\left[\overline{\alpha}\right] = \frac{\mathrm{var}(X_1)}{n} = \frac{1}{n\beta^2}.$$

L'estimateur $\overline{\alpha}$ est donc également convergent mais la vitesse de convergence (de l'ordre de $\frac{1}{n}$) est plus faible que celle de $\widetilde{\alpha}$ (qui est de l'ordre de $\frac{1}{n^2}$). On préfèrera donc $\widetilde{\alpha}$ à $\overline{\alpha}$.