EXAMEN STATISTIQUE - 1SN

Lundi 15 janvier 2024 (10h-11h30)

Partiel sans document (Une feuille A4 recto-verso autorisée)

Exercice 1 : Estimation (10 points)

On considère n observations $x_1, ..., x_n$ issues d'un vecteur $(X_1, ..., X_n)$ de n variables aléatoires indépendantes de mêmes lois de densités

$$f_r(x_i; \lambda) = \frac{\lambda^r}{\Gamma(r)} \frac{1}{x_i^{r+1}} \exp\left(-\frac{\lambda}{x_i}\right) I_{\mathbb{R}^+}(x_i),$$

où $I_{\mathbb{R}^+}$ est la fonction indicatrice sur $\mathbb{R}^+$ ($I_{\mathbb{R}^+}(x) = 1$ si $x > 0$ et 0 sinon) et où r est un paramètre supposé connu vérifiant $r > 2$. La moyenne et la variance d'une telle loi appelée loi inverse-gamma et notée $\mathcal{IG}(r, \lambda)$ (voir tables) sont

$$E[X_i] = \frac{\lambda}{r-1} \text{ et } \mathrm{var}[X_i] = \frac{\lambda^2}{(r-1)^2(r-2)}.$$

On admettra que si $X_i \sim \mathcal{IG}(r, \lambda)$, alors $\frac{1}{X_i}$ suit une loi gamma de paramètres r et λ, c'est-à-dire $\frac{1}{X_i} \sim \mathcal{G}(r, \lambda)$, et qu'inversement si une variable Y_i est de loi gamma, i.e., $Y_i \sim \mathcal{G}(r, \lambda)$, alors $\frac{1}{Y_i} \sim \mathcal{IG}(r, \lambda)$.

1. (2pts) Montrer que l'estimateur du maximum de vraisemblance du paramètre λ est

$$\widehat{\lambda}_{\text{MV}} = \frac{nr}{\sum_{k=1}^{n} \frac{1}{X_k}}.$$

   On justifiera que la vraisemblance admet un maximum en ce point.

2. (2pt) En utilisant les propriétés des lois gamma et inverse gamma données en début d'énoncé, déterminer la loi de $Y_k = \frac{1}{X_k}$. Montrer que $Y = \sum_{k=1}^{n} \frac{1}{X_k}$ suit une loi gamma dont on précisera les paramètres (on pourra déterminer la fonction caractéristique de Y). En déduire la loi de $\frac{1}{Y}$.

3. (2pts) Calculer l'espérance de $\widehat{\lambda}_{\text{MV}}$. En déduire un estimateur non biaisé de λ noté $\widetilde{\lambda}_{\text{MV}}$. L'estimateur $\widetilde{\lambda}_{\text{MV}}$ est-il convergent?

4. (2pts) Déterminer la borne de Cramér-Rao d'un estimateur non biaisé de λ. L'estimateur $\widetilde{\lambda}_{\text{MV}}$ est-il l'estimateur efficace du paramètre λ ?

5. (2pts) On suppose désormais que le paramètre λ est muni d'une loi a priori de densité

$$f(\lambda) = \begin{cases} e^{1-\lambda} & \text{si } \lambda > 1 \\ 0 \text{ sinon} \end{cases}$$

   Déterminer l'estimateur du maximum a posteriori du paramètre λ noté $\widehat{\lambda}_{\text{MAP}}$ et le comparer à $\widehat{\lambda}_{\text{MV}}$ pour $n \to \infty$. Commenter ce résultat.

Exercice 2 : Tests Statistiques (10 points)

On considère n observations $x_1, ..., x_n$ issues d'un vecteur $(X_1, ..., X_n)$ de n variables aléatoires indépendantes de mêmes lois de densités

$$f_r(x_i; \lambda) = \frac{\lambda^r}{\Gamma(r)} \frac{1}{x_i^{r+1}} \exp\left(-\frac{\lambda}{x_i}\right) I_{\mathbb{R}^+}(x_i),$$

où $I_{\mathbb{R}^+}$ est la fonction indicatrice sur $\mathbb{R}^+$ ($I_{\mathbb{R}^+}(x) = 1$ si $x > 0$ et 0 sinon) et où r est un paramètre supposé connu. Pour $r > 2$, la moyenne et la variance d'une telle loi appelée loi inverse-gamma et notée $\mathcal{IG}(r, \lambda)$ sont (voir table)

$$E[X_i] = \frac{\lambda}{r-1} \text{ et } \mathrm{var}[X_i] = \frac{\lambda^2}{(r-1)^2(r-2)}.$$

On désire utiliser les observations $x_1, ..., x_n$ pour déterminer si $\lambda = \lambda_0 > 0$ ou si $\lambda = \lambda_1 > \lambda_0$. On considère donc le test d'hypothèses

$$H_0 : \lambda = \lambda_0, \quad H_1 : \lambda = \lambda_1 \qquad \text{avec } \lambda_1 > \lambda_0.$$

1. (2pt) Déterminer la statistique T_n du test de Neyman Pearson et la région critique associée (sans chercher pour l'instant à déterminer le seuil associé à cette région noté S_α).

2. (1pt) En admettant que $Y_i = \frac{1}{X_i}$ suit une loi gamma de paramètres r et λ, i.e., $Y_i \sim \mathcal{G}(r, \lambda)$, déterminer la loi asymptotique de T_n sous les deux hypothèses H_0 et H_1.

3. (2pts) On note G la fonction de répartition d'une loi du normale $\mathcal{N}(0, 1)$. En utilisant la loi asymptotique trouvée à la question précédente, exprimer le risque de première espèce α en fonction du seuil du test de Neyman Pearson noté S_α, de $G(\alpha)$, n, r et λ_0. En déduire la valeur du seuil S_α en fonction de $G^{-1}(\alpha)$ et de n, r et λ_0.

4. (2pts) Déterminer les caractéristiques opérationnelles du récepteur (courbes COR) pour ce test et montrer qu'elles ne dépendent que de nr et de $\frac{\lambda_1}{\lambda_0}$. Analyser les performances du test en fonction de $\frac{\lambda_1}{\lambda_0}$ et tracer l'allure des courbes COR pour différentes valeurs de $\frac{\lambda_1}{\lambda_0}$.

5. (3pts) On désire vérifier que les observations x_i suivent la loi de densité $f_r(x_i; \lambda)$ avec $r = \lambda = 1$ à l'aide d'un test de Kolmogorov.

   - Déterminer la fonction de répartition de la loi de densité $f_r(x_i; \lambda)$ notée F lorsque $r = \lambda = 1$.
   - On observe l'échantillon de taille $n = 4$ suivant : $x_1 = 2$, $x_2 = 3$, $x_3 = 4$ et $x_4 = \frac{3}{2}$. Le tableau suivant résume les quantités nécessaires pour effectuer le test

$x_{(i)}$	1.5	2	3	4
$F\left(x_{(i)}\right)$	0.5134	0.6065	0.7165	0.7788
$\hat{F}\left(x_{(i)}^-\right)$	0	0.25	0.50	0.75
$\hat{F}\left(x_{(i)}^+\right)$	0.25	0.50	0.75	1
$E_i^- = \lvert F\left(x_{(i)}\right) - \hat{F}\left(x_{(i)}^-\right)\rvert$	0.5134	0.3565	0.2165	0.0288
$E_i^+ = \lvert F\left(x_{(i)}\right) - \hat{F}\left(x_{(i)}^+\right)\rvert$	0.2634	0.1065	0.0335	0.2212

   où $(x_{(1)}, x_{(2)}, x_{(3)}, x_{(4)})$ est l'échantillon ordonné.
   Expliquer ce que représentent $\hat{F}\left(x_{(i)}^-\right)$ et $\hat{F}\left(x_{(i)}^+\right)$.

   - Rappeler la région critique du test de Kolmogorov. Pour $\alpha = 0.01$ et $n = 4$, on a $S_{0.01} = 0.7342$. Que peut-on en conclure ?

LOIS DE PROBABILITÉ DISCRÈTES

$p_k = P\left[X = k\right] \qquad p_{1,\ldots,m} = P\left[X_1 = k_1, \ldots, X_m = k_m\right]$

LOI	**Probabilités**	Moyenne	Variance	Fonction Caractéristique
Uniforme	$p_k = \frac{1}{n}$ $k \in \{1, \ldots, n\}$	$\frac{n+1}{2}$	$\frac{n^2-1}{12}$	$\dfrac{e^{it}\left(1 - e^{itn}\right)}{n\left(1 - e^{it}\right)}$
Bernoulli	$p_1 = P\left[X = 1\right] = p$ $p_0 = P\left[X = 0\right] = q$ $p \in [0, 1] \quad q = 1 - p$	p	pq	$pe^{it} + q$
Binomiale $B\left(n, p\right)$	$p_k = \binom{n}{k} p^k q^{n-k}$ $p \in [0, 1] \quad q = 1 - p$ $k \in \{0, 1, \ldots, n\}, \binom{n}{k} = \frac{n!}{k!(n-k)!}$	np	npq	$\left(pe^{it} + q\right)^n$
Binomiale négative	$p_k = \binom{n+k-1}{n-1} p^n q^k$ $p \in [0, 1] \quad q = 1 - p$ $k \in \mathbb{N}, \binom{n}{k} = \frac{n!}{k!(n-k)!}$	$n\dfrac{q}{p}$	$n\dfrac{q}{p^2}$	$\left(\dfrac{p}{1 - qe^{it}}\right)^n$
Multinomiale	$p_{1,\ldots,m} = \frac{n!}{k_1! \ldots k_m!} p_1^{k_1} \ldots p_m^{k_m}$ $p_j \in [0, 1] \quad q_j = 1 - p_j$ $k_j \in \{0, 1, \ldots, n\}$ $\sum_{j=1}^m k_j = n \quad \sum_{j=1}^m p_j = 1$	np_j	Variance : $np_j q_j$ Covariance : $-np_j p_k$	$\left(\sum_{j=1}^m p_j e^{it}\right)^n$
Poisson $P\left(\lambda\right)$	$p_k = e^{-\lambda}\dfrac{\lambda^k}{k!}$ $\lambda > 0 \quad k \in \mathbb{N}$	λ	λ	$\exp\left[\lambda\left(e^{it} - 1\right)\right]$
Géométrique	$p_k = pq^{k-1}$ $p \in [0, 1] \quad q = 1 - p$ $k \in \mathbb{N}^*$	$\dfrac{1}{p}$	$\dfrac{q}{p^2}$	$\dfrac{pe^{it}}{1 - qe^{it}}$

LOIS DE PROBABILITÉ CONTINUES

LOI	**Densité de probabilité**	Moyenne	Variance	Fonction Caractéristique
Uniforme	$f(x) = \frac{1}{b-a}$ $x \in]a, b[$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\frac{e^{itb} - e^{ita}}{it(b-a)}$
Gamma $\mathcal{G}(\nu, \theta)$	$f(x) = \frac{\theta^\nu}{\Gamma(\nu)} e^{-\theta x} x^{\nu-1}$ $\theta > 0,\ \nu > 0$ $x \geq 0$ avec $\Gamma(n+1) = n!\ \forall n \in \mathbb{N}$	$\frac{\nu}{\theta}$	$\frac{\nu}{\theta^2}$	$\frac{1}{\left(1 - i\frac{t}{\theta}\right)^\nu}$
Inverse gamma $\mathcal{IG}(\nu, \theta)$	$f(x) = \frac{\theta^\nu}{\Gamma(\nu)} e^{-\frac{\theta}{x}} \frac{1}{x^{\nu+1}}$ $\theta > 0,\ \nu > 0$ $x \geq 0$ avec $\Gamma(n+1) = n!\ \forall n \in \mathbb{N}$	$\frac{\theta}{\nu-1}$ si $\nu > 1$	$\frac{\theta^2}{(\nu-1)^2(\nu-2)}$ si $\nu > 2$	(*)
Première loi de Laplace	$f(x) = \frac{1}{2} e^{-\lvert x \rvert}, \quad x \in \mathbb{R}$	0	2	$\frac{1}{1+t^2}$
Normale univariée $\mathcal{N}(m, \sigma^2)$	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}, \quad x \in \mathbb{R}$	m	σ^2	$e^{imt - \frac{\sigma^2 t^2}{2}}$
Normale multivariée $\mathcal{N}_p(\boldsymbol{m}, \boldsymbol{\Sigma})$	$f(x) = K e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{m})^T \Sigma^{-1} (\boldsymbol{x}-\boldsymbol{m})}$ $K = \frac{1}{\sqrt{(2\pi)^p \det(\Sigma)}}$ $x \in \mathbb{R}^p$	$\boldsymbol{m}$	$\boldsymbol{\Sigma}$	$e^{i\boldsymbol{u}^T\boldsymbol{m} - \frac{1}{2}\boldsymbol{u}^T\boldsymbol{\Sigma}\boldsymbol{u}}$
Khi$_2$ χ^2_ν $\Gamma\left(\frac{1}{2}, \frac{\nu}{2}\right)$	$f(x) = k e^{-\frac{x}{2}} x^{\frac{\nu}{2}-1}$ $k = \frac{1}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)}$ $\nu \in \mathbb{N}^*,\ x \geq 0$	ν	2ν	$\frac{1}{(1-2it)^{\frac{\nu}{2}}}$
Cauchy $c_{\lambda,\alpha}$	$f(x) = \frac{1}{\pi\lambda\left(1 + \left(\frac{x-\alpha}{\lambda}\right)^2\right)}$ $\lambda > 0,\ \alpha \in \mathbb{R}$	(-)	(-)	$e^{i\alpha t - \lambda \lvert t \rvert}$
Beta $B(a, b)$	$f(x) = k x^{a-1} (1-x)^{b-1}$ $k = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}$ $a > 0,\ b > 0$ $x \in]0, 1[$ avec $\Gamma(n+1) = n!\ \forall n \in \mathbb{N}$	$\frac{a}{a+b}$	$\frac{ab}{(a+b)^2(a+b+1)}$	(*)