EXAMEN STATISTIQUE - 1SN

Lundi 15 janvier 2025 (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

$$f(x_i; \theta) = \frac{2x_i}{\theta} \exp\left(-\frac{x_i^2}{\theta}\right) I_{\mathbb{R}^+}(x_i),$$

où $I_{\mathbb{R}^+}$ est la fonction indicatrice sur $\mathbb{R}^+$ ($I_{\mathbb{R}^+}(x) = 1$ si $x > 0$ et 0 sinon) et où θ est un paramètre inconnu que l'on cherche à estimer.

1. Montrer que l'estimateur du maximum de vraisemblance du paramètre θ est

$$\widehat{\theta}_{\text{MV}} = \frac{1}{n} \sum_{k=1}^{n} X_k^2.$$

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

2. Déterminer la loi de $Y_k = X_k^2$ et en déduire en utilisant les tables que $E[Y_k] = \theta$ et $\text{var}(Y_k) = \theta^2$.

3. L'estimateur $\widehat{\theta}_{\text{MV}}$ est-il un estimateur non biaisé et convergent du paramètre θ ?

4. Déterminer la borne de Cramér-Rao d'un estimateur non biaisé de θ. L'estimateur $\widehat{\theta}_{\text{MV}}$ est-il l'estimateur efficace du paramètre θ ?

5. Déterminer $E[X_k]$ et en déduire un estimateur des moments du paramètre θ.

6. On suppose désormais que le paramètre θ est muni d'une loi a priori de densité

$$f(\theta) = \frac{a}{\theta^2} e^{-\frac{a}{\theta}} I_{\mathbb{R}^+}(\theta).$$

   avec $a > 0$ et où $I_{\mathbb{R}^+}$ est la fonction indicatrice sur $\mathbb{R}^+$ ($I_{\mathbb{R}^+}(\theta) = 1$ si $\theta > 0$ et 0 sinon). Déterminer l'estimateur du maximum a posteriori du paramètre θ noté $\widehat{\theta}_{\text{MAP}}$. Montrer que la loi a posteriori du paramètre θ est une loi inverse gamma dont on déterminera les paramètres. En déduire l'estimateur MMSE de θ noté $\widehat{\theta}_{\text{MMSE}}$.

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

$$f(x_i;\theta) = \frac{2x_i}{\theta} \exp\left(-\frac{x_i^2}{\theta}\right) I_{\mathbb{R}^+}(x_i),$$

où $I_{\mathbb{R}^+}$ est la fonction indicatrice sur $\mathbb{R}^+$ ($I_{\mathbb{R}^+}(x) = 1$ si $x > 0$ et 0 sinon). On désire utiliser les observations $x_1, ..., x_n$ pour déterminer si $\theta = \theta_0 > 0$ ou si $\theta = \theta_1 < \theta_0$. On considère donc le test d'hypothèses

$$H_0 : \theta = \theta_0, \quad H_1 : \theta = \theta_1 \qquad \text{avec } \theta_1 < \theta_0.$$

1. 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. On admettra que la variable aléatoire $Y_k = X_k^2$ est de moyenne $E[Y_k] = \theta$ et de variance $\text{var}(Y_k) = \theta^2$. En déduire la loi asymptotique de T_n sous les deux hypothèses H_0 et H_1.

3. On note F 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 F, n et θ_0. En déduire la valeur du seuil S_α en fonction de $F^{-1}(\alpha)$, de n et de θ_0.

4. Déterminer les caractéristiques opérationnelles du récepteur (courbes COR) pour ce test et montrer qu'elles ne dépendent que de $\sqrt{n}$ et de $\frac{\theta_1}{\theta_0}$. Analyser les performances du test en fonction de $\frac{\theta_1}{\theta_0}$ et de n. Tracer l'allure des courbes COR pour différentes valeurs de n et de $\frac{\theta_1}{\theta_0}$.

5. On désire vérifier que l'ensemble des observations $(x_1, ..., x_n)$ suit une loi de densité $f(x_i;\theta)$ avec $\theta = 1$ à l'aide d'un test du χ^2. Déterminer la fonction de répartition de cette loi et en déduire que l'intervalle $]0, +\infty[$ est la réunion de trois intervalles équiprobables pour la loi de densité $f(x_i; 1)$ que l'on précisera. On compte le nombre d'observations x_i appartenant à ces trois intervalles et on trouve $K_1 = 13$, $K_2 = 8$ et $K_3 = 9$. Quelle est la valeur de la statistique du test du χ^2 ? Exprimer le seuil de ce test noté S_α en fonction du risque α et de l'inverse de la fonction de répartition d'une loi du χ^2 dont on précisera le nombre de degrés de liberté. On donne $S_{0.05} = 5.991$. Qu'en conclut-on ?

LOIS DE PROBABILITÉ DISCRÈTES

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

LOI	**Probabilités**	Moyenne	Variance	Fonction Caractéristique
Uniforme	$p_k = \frac{1}{n}$ $k \in \{1, ..., n\}$	$\frac{n+1}{2}$	$\frac{n^2-1}{12}$	$\dfrac{e^{it}(1-e^{itn})}{n(1-e^{it})}$
Bernoulli	$p_1 = P[X = 1] = p$ $p_0 = P[X = 0] = q$ $p \in [0,1] \quad q = 1-p$	p	pq	$pe^{it} + q$
Binomiale $B(n, p)$	$p_k = \binom{n}{k} p^k q^{n-k}$ $p \in [0,1] \quad q = 1-p$ $k \in \{0, 1, ..., n\}, \binom{n}{k} = \frac{n!}{k!(n-k)!}$	np	npq	$(pe^{it} + q)^n$
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	$p_{1,...,m} = \frac{n!}{k_1!...k_m!} p_1^{k_1} ... p_m^{k_m}$ $p_j \in [0,1] \quad q_j = 1 - p_j$ $k_j \in \{0, 1, \ldots, n\}$ $\sum_{j=1}^m k_j = n \quad \sum_{j=1}^m p_j = 1$	np_j	Variance : $np_j q_j$ Covariance : $-np_j p_k$	$\left(\sum_{j=1}^m p_j e^{it}\right)^n$
Poisson $P(\lambda)$	$p_k = e^{-\lambda} \dfrac{\lambda^k}{k!}$ $\lambda > 0 \quad k \in \mathbb{N}$	λ	λ	$\exp[\lambda(e^{it} - 1)]$
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	$f(x) = \frac{1}{b-a}$ $x \in]a, b[$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\dfrac{e^{itb} - e^{ita}}{it(b-a)}$
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 $\Gamma(n+1) = n!\ \forall n \in \mathbb{N}$	$\dfrac{\nu}{\theta}$	$\dfrac{\nu}{\theta^2}$	$\dfrac{1}{\left(1 - i\frac{t}{\theta}\right)^\nu}$
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	$\dfrac{1}{1+t^2}$
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 $\mathcal{N}_p(\boldsymbol{m}, \boldsymbol{\Sigma})$	$f(x) = K e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{m})^T \Sigma^{-1} (\boldsymbol{x}-\boldsymbol{m})}$ $K = \dfrac{1}{\sqrt{(2\pi)^p \det(\Sigma)}}$ $x \in \mathbb{R}^p$	$\boldsymbol{m}$	$\boldsymbol{\Sigma}$	$e^{i\boldsymbol{u}^T\boldsymbol{m} - \frac{1}{2}\boldsymbol{u}^T\boldsymbol{\Sigma}\boldsymbol{u}}$
Khi_2 χ^2_ν $\Gamma\left(\frac{\nu}{2}, \frac{1}{2}\right)$	$f(x) = k e^{-\frac{x}{2}} x^{\frac{\nu}{2}-1}$ $k = \dfrac{1}{2^{\frac{\nu}{2}} \Gamma\left(\frac{\nu}{2}\right)}$ $\nu \in \mathbb{N}^*,\ x \geq 0$	ν	2ν	$\dfrac{1}{(1-2it)^{\frac{\nu}{2}}}$
Cauchy $c_{\lambda,\alpha}$	$f(x) = \dfrac{1}{\pi\lambda\left(1 + \left(\frac{x-\alpha}{\lambda}\right)^2\right)}$ $\lambda > 0,\ \alpha \in \mathbb{R}$	(-)	(-)	$e^{i\alpha t - \lambda \lvert t \rvert}$
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 $\Gamma(n+1) = n!\ \forall n \in \mathbb{N}$	$\frac{a}{a+b}$	$\frac{ab}{(a+b)^2(a+b+1)}$	(*)