EXAMEN STATISTIQUE - 1MF2E

Vendredi 6 mars 2020 (8h-9h45)

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

Exercice 1 : Estimation (10 points)

On considère n variables aléatoires indépendantes $X_1, ..., X_n$ suivant la même loi de densité

$$f(x_i; \theta) = \theta^2 x_i \exp(-\theta x_i) \mathbb{I}_{\mathbb{R}^+}(x_i)$$

où $\theta \in]0, \infty[$ est un paramètre inconnu et $\mathbb{I}_{\mathbb{R}^+}$ est la fonction indicatrice sur $\mathbb{R}^+$.

1. Déterminer la vraisemblance de $X_1, ..., X_n$ et montrer qu'elle admet un unique maximum global pour une valeur de θ que l'on déterminera. En déduire l'estimateur du maximum de vraisemblance du paramètre θ noté $\widehat{\theta}_{\text{MV}}$ en fonction des variables X_i.

2. Déterminer la fonction caractéristique puis la loi de

$$U = \sum_{i=1}^{n} X_i$$

   et montrer que $T = \frac{1}{U}$ suit une loi inverse gamma IG$(\theta, 2n)$. Exprimer $\widehat{\theta}_{\text{MV}}$ en fonction de T et en déduire un estimateur non biaisé de θ noté $\tilde{\theta}_{\text{MV}}$. Déterminer la variance de $\tilde{\theta}_{\text{MV}}$.

3. Déterminer la borne de Cramér-Rao pour un estimateur non biaisé du paramètre θ à partir de l'observation des variables $X_1, ..., X_n$. L'estimateur $\tilde{\theta}_{\text{MV}}$ est-il l'estimateur efficace de θ ?

4. On désire maintenant construire un estimateur Bayésien du paramètre θ à partir d'une loi a priori définie par (loi gamma $\Gamma(a, b)$)

$$p(\theta) = \frac{a^b}{\Gamma(b)} e^{-a\theta} \theta^{b-1} \mathbb{I}_{\mathbb{R}^+}(\theta)$$

   où Γ est la fonction gamma dont l'expression n'est pas importante dans cet exercice.

   - Montrer que la densité de $\theta | x_1, ..., x_n$ est proportionnelle à la densité d'une loi gamma dont on précisera les paramètres.
   - Déterminer l'estimateur du maximum a posteriori du paramètre θ noté $\widehat{\theta}_{\text{MAP}}$ et étudier son comportement par rapport à l'estimateur du maximum de vraisemblance lorsque $n \to \infty$.

Exercice 2 : Test Statistique (5 points)

On considère n variables aléatoires $X_1, ..., X_n$ indépendantes de densités

$$f(x_i; \theta) = \theta^2 x_i \exp(-\theta x_i) \mathbb{I}_{\mathbb{R}^+}(x_i)$$

avec $\theta > 0$ et où $\mathbb{I}_{\mathbb{R}^+}$ indique la fonction indicatrice sur $\mathbb{R}^+$. On veut à l'aide d'observations $x_1, ..., x_n$ de ces variables aléatoires tester les deux hypothèses suivantes

$$H_0 : \theta = \theta_0 \text{ contre } H_1 : \theta = \theta_1 \qquad \text{avec } \theta_1 > \theta_0.$$

1. À l'aide du théorème de Neyman-Pearson, calculer la statistique T_n du test le plus puissant et indiquer la région critique de ce test que l'on représentera graphiquement pour $n = 2$ (comme d'habitude, on retiendra pour T_n la fonction seule des observations).

2. On admettra que θT_n suit une loi gamma $\Gamma(1, 2n)$ (qui est donc indépendante du paramètre θ). En déduire une expression des risques de première et seconde espèce α et β du test en fonction de θ_0 et θ_1 et de la fonction de répartition d'une loi $\Gamma(1, 2n)$ notée F_{2n}.

3. Déterminer les courbes COR associées à ce test et expliquer comment les performances de ce test dépendent des valeurs de θ_0 et θ_1. Illustrer ce résultat à l'aide de quelques courbes COR associées à différentes valeurs de θ_0 et θ_1.

Exercice 3 : Test d'adéquation (5 points)

Un résultat classique de la littérature est que si X est une variable aléatoire de la loi gamma $\Gamma(a, b)$, alors on peut approcher la loi de $X^{1/3}$ par une loi normale de moyenne $\mu = (ab)^{1/3}$ et de variance $\sigma^2 = \frac{b}{9\mu}$. Afin de vérifier cette approximation, on génère des données distribuées suivant la loi $\Gamma(a, b)$ notées $x_1, ..., x_n$ et on considère le test d'adéquation suivant :

$$\begin{aligned} H_0 &: Y_i \sim \mathcal{N}(0, 1) \\ H_1 &: \text{non } H_0 \end{aligned}$$

où Y_i est obtenu en centrant et en réduisant les variables $X_i^{1/3}$.

1. Rappeler comment on peut centrer et réduire les variables $x_i^{1/3}$ pour obtenir les variables y_i.

2. On désire effectuer un test du χ^2 avec $K = 4$ classes. Expliquer comment déterminer ces 4 classes.

3. Rappeler l'expression de la statistique du test du χ^2 et la région critique de ce test.

4. Exprimer le seuil du test du χ^2 en fonction du risque de première espèce α.

5. Peut-on calculer la puissance du test ? Pourquoi ?

LOIS DE PROBABILITÉ CONTINUES

$\mathbf{m}$: moyenne σ^2 : variance **F. C. :** fonction caractéristique

LOI	**Densité de probabilité**	$\mathbf{m}$	σ^2	**F. C.**
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 $\Gamma(\theta, \nu)$	$f(x) = \frac{\theta^\nu}{\Gamma(\nu)} e^{-\theta x} x^{\nu-1}$ $\theta > 0,\ \nu > 0$ $x \geq 0$	$\dfrac{\nu}{\theta}$	$\dfrac{\nu}{\theta^2}$	$\dfrac{1}{\left(1 - i\frac{t}{\theta}\right)^\nu}$
Inverse gamma IG(θ, ν)	$f(x) = \frac{\theta^\nu}{\Gamma(\nu)} e^{-\frac{\theta}{x}} \frac{1}{x^{\nu+1}}$ $\theta > 0,\ \nu > 0$ $x \geq 0$	$\frac{\theta}{\nu-1}$ si $\nu > 1$	$\frac{\theta^2}{(\nu-1)^2(\nu-2)}$ si $\nu > 2$	(*)
Première loi de Laplace	$f(x) = \frac{1}{2} e^{-\lvert x \rvert}$	0	2	$\dfrac{1}{1+t^2}$
Normale $\mathcal{N}\left(m, \sigma^2\right)$	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}$	m	σ^2	$e^{imt - \frac{\sigma^2 t^2}{2}}$
Khi$_2$ χ^2_ν $\Gamma\left(\frac{1}{2}, \frac{\nu}{2}\right)$	$f(x) = k e^{-\frac{x}{2}} x^{\frac{\nu}{2}-1}$ $k = \dfrac{1}{2^{\frac{\nu}{2}} \Gamma\left(\frac{\nu}{2}\right)}$ $\nu \in \mathbb{N}^*,\ x \geq 0$	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	$f(x) = k x^{a-1} (1-x)^{b-1}$ $k = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}$ $a > 0,\ b > 0,\ x \in]0, 1[$	$\frac{a}{a+b}$	$\frac{ab}{(a+b)^2(a+b+1)}$	(*)

LOIS DE PROBABILITÉ DISCRÈTES

$\mathbf{m}$: moyenne σ^2 : variance **F. C. :** fonction caractéristique

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

LOI	**Probabilités**	**m**	σ^2	**F. C.**
Uniforme	$p_k = \frac{1}{n}$ $k \in \{1, ..., n\}$	$\frac{n+1}{2}$	$\frac{n^2-1}{12}$	$\dfrac{e^{it}\left(1 - e^{itn}\right)}{n\left(1 - e^{it}\right)}$
Bernoulli	$p_1 = P[X = 1] = p$ $p_0 = P[X = 0] = q$ $p \in [0, 1] \quad q = 1 - p$	p	pq	$pe^{it} + q$
Binomiale $B(n, p)$	$p_k = C_n^k p^k q^{n-k}$ $p \in [0, 1] \quad q = 1 - p$ $k \in \{0, 1, ..., n\}$	np	npq	$\left(pe^{it} + q\right)^n$
Binomiale négative	$p_k = C_{n+k-1}^{n-1} p^n q^k$ $p \in [0, 1] \quad q = 1 - p$ $k \in \mathbb{N}$	$n\dfrac{q}{p}$	$n\dfrac{q}{p^2}$	$\left(\dfrac{p}{1 - qe^{it}}\right)^n$
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	$\exp\left[\lambda\left(e^{it} - 1\right)\right]$
Géométrique	$p_k = pq^{k-1}$ $p \in [0, 1] \quad q = 1 - p$ $k \in \mathbb{N}^*$	$\dfrac{1}{p}$	$\dfrac{q}{p^2}$	$\dfrac{pe^{it}}{1 - qe^{it}}$