EXAMEN STATISTIQUE - 1SN

Mardi 27 Novembre 2019 (14h-15h30)

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 échantillon $(X_1, ..., X_n)$ distribué suivant la même loi de densité

$$f(x_i; \boldsymbol{\theta}) = \frac{\beta}{\alpha} \exp\left(\beta x_i - \frac{1}{\alpha} e^{\beta x_i}\right), \quad x_i \in \mathbb{R}$$

avec $\alpha > 0$, $\beta > 0$ et $\boldsymbol{\theta} = (\alpha, \beta)^T$. On cherche tout d'abord à estimer le paramètre α à partir des observations $x_1, ..., x_n$ (β est connu dans les 5 premières questions de cet exercice), puis on s'intéresse à l'estimation du vecteur $\boldsymbol{\theta} = (\alpha, \beta)^T$ dans la dernière question de cet exercice.

1. Déterminer l'estimateur du maximum de vraisemblance du paramètre α noté $\widehat{\alpha}_{\text{MV}}$.
2. Montrer que $Y_i = e^{\beta X_i}$ suit une loi gamma dont on déterminera les paramètres. En s'aidant des tables, déterminer la moyenne et la variance de la variable aléatoire Y_i.
3. L'estimateur $\widehat{\alpha}_{\text{MV}}$ est-il un estimateur sans biais et convergent du paramètre α ?
4. Déterminer la borne de Cramer-Rao pour un estimateur non biaisé du paramètre α. L'estimateur $\widehat{\alpha}_{\text{MV}}$ est-il l'estimateur efficace du paramètre α ?
5. On suppose que α est muni d'une loi a priori inverse gamma IG(b, a) définie par

$$p(\alpha) \propto \frac{1}{\alpha^{a+1}} \exp\left(-\frac{b}{\alpha}\right) \mathbb{I}_{\mathbb{R}^+}(\alpha)$$

   où a et b sont deux paramètres connus, $\propto$ signifie "proportionnel à" et $\mathbb{I}_{\mathbb{R}^+}$ est la fonction indicatrice sur $\mathbb{R}^+$ (telle que $\mathbb{I}_{\mathbb{R}^+}(\alpha) = 1$ si $\alpha > 0$ et $\mathbb{I}_{\mathbb{R}^+}(\alpha) = 0$ sinon).

   - Montrer que l'estimateur du maximum a posteriori du paramètre α s'écrit

   $$\widehat{\alpha}_{\text{MAP}} = c_1(n)\widehat{\alpha}_{\text{MV}} + c_2(n)$$

     où $c_1(n)$ et $c_2(n)$ sont deux fonctions de n, a, b que l'on déterminera. Déterminer les limites de $c_1(n)$ et de $c_2(n)$ lorsque $n \to \infty$ et commenter le résultat obtenu.
   - Montrer que la loi de $\alpha | x_1, ..., x_n$ est une loi inverse gamma dont on déterminera les paramètres. En déduire l'estimateur MMSE de α.

   On suppose désormais que α et β sont deux paramètres inconnus que l'on cherche à estimer à partir de l'échantillon $(X_1, ..., X_n)$ et on pose $\boldsymbol{\theta} = (\alpha, \beta)^T$.

6. Montrer que l'estimateur du maximum de vraisemblance de $\boldsymbol{\theta} = (\alpha, \beta)^T$ construit à partir de l'observation des variables aléatoires X_i s'obtient comme la solution du système suivant

$$\alpha = \frac{1}{n} \sum_{i=1}^{n} \exp(\beta x_i), \quad g(\beta; x_1, ..., x_n) = \frac{1}{\beta} + \frac{1}{n} \sum_{i=1}^{n} x_i - \frac{\sum_{i=1}^{n} x_i \exp(\beta x_i)}{\sum_{i=1}^{n} \exp(\beta x_i)} = 0.$$

   Expliquer succinctement comment on peut trouver la solution de ce système.

Exercice 2 : Tests Statistiques (10 **points)**

On considère n observations $x_1, ..., x_n$ issues d'un échantillon $(X_1, ..., X_n)$ distribué suivant la même loi de Weibull de densité

$$f(x_i; \alpha) = \frac{\beta x_i^{\beta-1}}{\alpha} \exp\left(-\frac{x_i^\beta}{\alpha}\right) \mathbb{I}_{\mathbb{R}^+}(x_i),$$

où $\beta > 0$ est un paramètre connu, $\alpha > 0$, et où $\mathbb{I}_{\mathbb{R}^+}(x)$ est la fonction indicatrice sur $\mathbb{R}^+$. L'objectif de cet exercice est d'étudier un test statistique basé sur les observations $x_1, ..., x_n$ qui permet de déterminer si $\alpha = \alpha_0$ ou si $\alpha = \alpha_1 < \alpha_0$. On considère donc le test d'hypothèses

$$H_0 : \alpha = \alpha_0, \quad H_1 : \alpha = \alpha_1 \qquad \text{avec } \alpha_1 < \alpha_0.$$

1. Montrer que la statistique de test du théorème de Neyman-Pearson est

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

   et indiquer la région critique de ce test. Représenter cette région critique pour $\beta = 2$ et $n = 2$.

2. Montrer que $Y_i = \frac{2}{\alpha} X_i^\beta$ suit une loi du χ^2 à deux degrés de liberté, i.e., $Y_i \sim \chi_2^2$. En déduire la loi de $U_n = \frac{2}{\alpha} T_n$ sous les deux hypothèses H_0 et H_1.

3. On note $F_{2n}(x)$ la fonction de répartition d'une loi du χ_{2n}^2. Exprimer les risques de première et seconde espèce notés PFA et PND en fonction du seuil du test de Neyman-Pearson noté S, de F_{2n} et de α_0 et α_1. En déduire les caractéristiques opérationnelles du récepteur (courbes COR) pour ce test en fonction de $\alpha_0, \alpha_1, F_{2n}$ et F_{2n}^{-1}. Analyser les performances du test en fonction des valeurs de α_0 et α_1 et donner l'allure des courbes COR pour plusieurs valeurs du couple (α_0, α_1).

4. On désire vérifier que l'hypothèse d'une loi de Weibull pour les variables aléatoires $X_1, ..., X_n$ est correcte. Pour cela, on construit les variables $Y_i = \frac{2}{\alpha} X_i^\beta$ et on cherche à vérifier si ces variables suivent une loi du χ_2^2 à l'aide d'un test de Kolmogorov. Expliquer le principe de ce test (on précisera notamment comment on calcule la statistique de ce test et comment le seuil est calculé à partir d'un risque de première espèce donné).

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	$p_k = pq^{k-1}$ $p \in [0, 1] \quad q = 1 - p$ $k \in \mathbb{N}^*$	$\dfrac{1}{p}$	$\dfrac{q}{p^2}$	$\dfrac{pe^{it}}{1 - qe^{it}}$