Partiel de Statistique

Lundi 5 décembre 2011, 8h00-10h00

Partiel sans documents (une feuille manuscrite A4 est autorisée)
Barème indicatif. Exo 1 : 12pts, Exo 2 : 5 pts, Exo 3 : 3pts

Exercice 1 : Estimation

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

$$f(x;\theta) = \theta x^{\theta-1} I_{[0,1]}(x)$$

avec $\theta > 0$ et où $I_{[0,1]}(x)$ est la fonction indicatrice sur $[0,1]$ ($I_{[0,1]}(x) = 1$ si $x \in [0,1]$ et $I_{[0,1]}(x) = 0$ si $x \notin [0,1]$).

1) Montrer que la vraisemblance de $(x_1, ..., x_n)$ 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 $\widehat{\theta}_{\text{MV}}$.

2) Déterminer la loi de $Y_i = -\ln X_i$. En déduire la fonction caractéristique de Y_i et montrer que la variable aléatoire

$$Z = \sum_{i=1}^{n} Y_i = -\sum_{i=1}^{n} \ln X_i$$

suit une loi gamma de paramètres θ et n, ce que l'on notera $Z \sim \Gamma(\theta, n)$.

3) Déterminer l'estimateur du maximum du paramètre $a = \frac{1}{\theta}$ construit à partir des variables aléatoires X_i noté $\widehat{a}_{\text{MV}}$ (on pourra utiliser la propriété d'invariance fonctionnelle). Déterminer la moyenne et la variance de $\widehat{a}_{\text{MV}}$. En déduire que l'estimateur $\widehat{a}_{\text{MV}}$ est un estimateur sans biais et convergent du paramètre a.

4) Déterminer la borne de Cramer-Rao pour un estimateur non biaisé du paramètre a. L'estimateur $\widehat{a}_{\text{MV}}$ est-il l'estimateur efficace du paramètre a ?

5) Déterminer $E[X_i]$ en fonction de θ. En déduire un estimateur de θ noté $\widehat{\theta}_{\text{M}}$ en utilisant la méthode des moments.

6) On suppose désormais qu'on dispose d'une information a priori sur le paramètre θ résumée dans la loi gamma $\Gamma(\alpha, \beta)$ de densité

$$f(\theta | \alpha, \beta) = \frac{\alpha^{\beta}}{\Gamma(\beta)} \theta^{\beta-1} \exp(-\alpha\theta) I_{\mathbb{R}^+}(\theta)$$

- Montrer que la loi a posteriori de $\theta | x_1, ..., x_n$ est aussi une loi inverse-gamma dont on précisera les paramètres.
- Déterminer l'estimateur du maximum a posteriori du paramètre θ noté $\widehat{\theta}_{\text{MAP}}$.
- Expliquer le comportement de $\widehat{\theta}_{\text{MAP}}$ lorsque $n \to \infty$.

Exercice 2 : Test de Neyman-Pearson

Comme dans l'exercice précédent, on considère n variables aléatoires $X_1, ..., X_n$ indépendantes suivant la même loi de densité

$$f(x;\theta) = \theta x^{\theta-1} I_{[0,1]}(x)$$

avec $\theta > 0$ et on désire effectuer le test d'hypothèses

$$\begin{aligned} H_0 &: \theta = \theta_0 \\ H_1 &: \theta = \theta_1 \end{aligned}$$

avec $\theta_1 > \theta_0$.

1) Déterminer la statistique du test de Neyman-Pearson notée T_n (qui ne dépend que des variables aléatoires $X_1, ..., X_n$) et donner la région critique de ce test. Ce résultat est-il cohérent avec le fait que la moyenne de $Y_i = -\ln X_i$ est

$$E[Y_i] = \frac{1}{\theta}$$

2) On rappelle que

$$\sum_{i=1}^{n} Y_i = -\sum_{i=1}^{n} \ln X_i$$

suit une loi gamma de paramètres θ et n.

- Déterminer la densité de la variable aléatoire $U_n = \frac{1}{\theta T_n}$ et montrer que cette densité dépend de n mais est indépendante de θ. On notera $H_n(u)$ la fonction de répartition associée à la variable aléatoire U_n.
- Exprimer les risques de première et de seconde espèces α et β en fonction du seuil du test de Neyman Pearson noté S_α, des paramètres θ_0 et θ_1 et de la fonction H_n.
- En déduire l'expression analytique des courbes COR de ce détecteur. Représenter la forme approximative de ces courbes COR pour différentes valeurs des paramètres θ_0 et θ_1.

3) Avant de tester la valeur du paramètre θ, on se propose de vérifier si les données x_i suivant la loi de densité $f(x;\theta)$ (donnée ci-dessus) à l'aide du test de Kolmogorov.

- Déterminer la fonction de répartition associée à la loi de densité $f(x;\theta)$ et représenter la graphiquement.
- Donner le principe du test de Kolmogorov en précisant a) la statistique de test (fonction des observations x_i), b) la règle de décision, c) comment calculer le seuil du test à l'aide de la loi (asymptotique) de Kolmogorov.

LOIS DE PROBABILITÉ CONTINUES

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

LOI	**Densité de probabilité**	**m**	$\boldsymbol{\sigma}^2$	**F. C.**
Uniforme	$f(x) = \frac{1}{b-a}$ $x \in \left]a, b\right[$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\dfrac{e^{itb} - e^{ita}}{it(b-a)}$
Gamma $\Gamma(\theta, \nu)$	$f(x) = \frac{\theta^\nu}{\Gamma(\nu)} e^{-\theta x} x^{\nu-1}$ $\theta > 0,\ \nu > 0$ $x \geq 0$	$\dfrac{\nu}{\theta}$	$\dfrac{\nu}{\theta^2}$	$\dfrac{1}{\left(1 - i\frac{t}{\theta}\right)^\nu}$
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 $\Gamma\left(\frac{1}{2}, \frac{\nu}{2}\right)$	$f(x) = k e^{-\frac{x}{2}} x^{\frac{\nu}{2}-1}$ $k = \dfrac{1}{2^{\frac{\nu}{2}} \Gamma\left(\frac{\nu}{2}\right)}$ $\nu \in \mathbb{N}^*,\ \curvearrowleft \geq \nvdash$	ν	2ν	$\dfrac{1}{(1-2it)^{\frac{\nu}{2}}}$
Cauchy $c_{\lambda,\alpha}$	$f(x) = \dfrac{1}{\pi\lambda\left(1 + \left(\frac{x-\alpha}{\lambda}\right)^2\right)}$ $\lambda > 0,\ \alpha \in \mathbb{R}$	(-)	(-)	$e^{i\alpha t - \lambda\lvert t \rvert}$
Bêta	$f(x) = k x^{a-1} (1-x)^{b-1}$ $k = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}$ $a > 0,\ b > 0,\ x \in \left]0, 1\right[$	$\frac{a}{a+b}$	$\frac{ab}{(a+b)^2(a+b+1)}$	(*)