Partiel de Statistique

Mardi 8 décembre 2009, 8h00-10h00

Partiel sans documents (une feuille manuscrite A4 est autorisée)

Exercice 1 : Estimation

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

$$f(x; \mu, \lambda) = \sqrt{\frac{\lambda}{2\pi x^3}} \exp\left[-\frac{\lambda}{2\mu^2 x}(x-\mu)^2\right] I_{\mathbb{R}^+}(x)$$

avec $\lambda > 0$, $\mu > 0$ et où $I_{\mathbb{R}^+}(x)$ est la fonction indicatrice sur $\mathbb{R}^+$ ($I_{\mathbb{R}^+}(x) = 1$ si $x \in \mathbb{R}^+$ et $I_{\mathbb{R}^+}(x) = 0$ si $x \notin \mathbb{R}^+$). Cette loi est appelée loi de Wald et on utilisera la notation habituelle $X_k \sim W(\mu, \lambda)$. On admettra que la moyenne, la variance et la fonction caractéristique d'une loi de Wald sont respectivement

$$E[X_k] = \mu, \mathrm{Var}[X_k] = \frac{\mu^3}{\lambda} \text{ et } \phi_{\mu,\lambda}(t) = E\left[e^{itX_k}\right] = \exp\left[\frac{\lambda}{\mu}\left(1 - \sqrt{1 - \frac{2i\mu^2 t}{\lambda}}\right)\right].$$

Partie 1 : μ inconnu et λ connu

1) Déterminer l'estimateur du maximum de vraisemblance du paramètre μ noté $\widehat{\mu}_{\mathrm{MV}}$.
2) Montrer que l'estimateur $\widehat{\mu}_{\mathrm{MV}}$ est un estimateur sans biais et convergent du paramètre μ.
3) Déterminer la borne de Cramer-Rao pour les estimateurs non biaisés du paramètre μ. L'estimateur $\widehat{\mu}_{\mathrm{MV}}$ est-il l'estimateur efficace du paramètre μ ?

Partie 2 : μ connu et λ inconnu

1) Déterminer l'estimateur du maximum de vraisemblance du paramètre λ noté $\widehat{\lambda}_{\mathrm{MV}}$.
2) On suppose désormais qu'on dispose d'une information a priori sur le paramètre λ résumée dans la loi gamma $\Gamma(a, b)$ de densité

$$f(\lambda | a, b) = \frac{b^a}{\Gamma(a)} \lambda^{a-1} e^{-\lambda b} I_{\mathbb{R}^+}(\lambda)$$

On rappelle que la moyenne d'une loi gamma est donnée par $E[\lambda | a, b] = \frac{a}{b}$.

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

Partie 3 : μ inconnu et λ inconnu

1) Déterminer l'estimateur du maximum de vraisemblance du vecteur $\theta = (\mu, \lambda)$ noté $\widehat{\theta}_{\mathrm{MV}} = \left(\widehat{\mu}_{\mathrm{MV}}, \widehat{\lambda}_{\mathrm{MV}}\right)$.
2) Déterminer la matrice d'information de Fisher associée au vecteur θ et montrer qu'elle est diagonale. En déduire les variances minimales d'estimateurs non biaisés de μ et λ.

Exercice 2 : Test de Neyman-Pearson

On considère n variables aléatoires $X_1, ..., X_n$ indépendantes telles que $X_k \sim W(\mu, \lambda)$ (pour $k = 1, ..., n$) et on suppose dans cette partie que le paramètre λ est connu. On désire effectuer le test d'hypothèses

$$\begin{aligned} H_0 &: \mu = \mu_0, \\ H_1 &: \mu = \mu_1 \text{ avec } \mu_1 > \mu_0, \end{aligned}$$

1) Déterminer la statistique du test de Neyman-Pearson et la région critique de ce test.
2) On admet que si $(X_1, ..., X_n)$ est un échantillon de loi de Wald, alors

$$Y = \frac{\lambda}{\mu^2} \sum_{k=1}^{n} X_k \sim W\left(\mu', \lambda'\right)$$

avec $\mu' = \frac{n\lambda}{\mu}$ et $\lambda' = \frac{n^2\lambda^2}{\mu^2}$. En déduire une expression des risques α et β en fonction du seuil du test noté S_α, des paramètres μ_0, μ_1, λ et de la fonction G_a définie par

$$G_a(x) = \int_x^{\infty} \frac{a}{\sqrt{2\pi u^3}} \exp\left[-\frac{(u-a)^2}{2u}\right] du$$

Exercice 3 : Test d'ajustement

On dispose de n observations $x_1, ..., x_n$ générées suivant la loi de Wald et on se propose de tester si on peut approcher la loi de Wald par une loi normale. Pour ce, on centre et on réduit les observations x_k en formant

$$y_k = \frac{x_k - \overline{x}}{s_x} \text{ avec } \overline{x} = \frac{1}{n}\sum_{k=1}^{n} x_k \text{ et } s_x = \left[\frac{1}{n-1}\sum_{k=1}^{n}(x_k - \overline{x})^2\right]^{1/2}$$

et on désire effectuer un test du χ^2 (avec 4 classes) pour tester si la loi normale $\mathcal{N}(0,1)$ est adaptée aux observations y_k. Répondre aux question suivantes

- Comment doit-on choisir les 4 classes ?
- Quelle est la statistique ϕ utilisée pour le test du χ^2 ?
- Donner la règle de décision.et préciser comment on détermine le seuil associé à cette décision en fonction du risque de première espèce α.
- Que pensez vous du risque β associé à la décision ci-dessus ?