Partiel de Statistique

Mercredi 4 décembre 2013, 8h00-10h00

Partiel sans documents (une feuille manuscrite A4 est autorisée)
Barème indicatif. Exo 1 : 10 pts, Exo 2 : 10 pts

Exercice 1 : Estimation

On considère n variables aléatoires $X_1, ..., X_n$ indépendantes suivant la même loi discrète à valeurs dans l'ensemble des entiers non nuls $\mathbb{N}^* = \mathbb{N}\setminus\{0\}$ définie par

$$P\left[X_i = x_i; \theta\right] = \frac{\theta^{x_i - 1}}{(1+\theta)^{x_i}}, \qquad x_i = 1, 2, ...$$

avec $\theta > 0$. On admettra que la moyenne et la variance d'une telle loi sont définies par $E\left[X_i\right] = \theta + 1$ et $\mathrm{var}(X_i) = \theta\left(\theta + 1\right)$.
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 $\widehat{\theta}_{\mathrm{MV}}$.
2) L'estimateur $\widehat{\theta}_{\mathrm{MV}}$ est-il un estimateur sans biais et convergent du paramètre θ ?
3) Déterminer la borne de Cramer-Rao pour un estimateur non biaisé du paramètre θ. L'estimateur $\widehat{\theta}_{\mathrm{MV}}$ est-il l'estimateur efficace du paramètre θ ?
5) On désire désormais estimer le paramètre $a = \frac{1}{1+\theta} \in \left]0, 1\right[$.

- Quel est l'estimateur du maximum de vraisemblance du paramètre a noté $\widehat{a}_{\mathrm{MV}}$?
- On suppose que le paramètre a est muni d'une loi a priori (appelée loi beta de paramètres α et β) de densité
$$p(a) = \frac{a^{\alpha-1}(1-a)^{\beta-1}}{B\left(\alpha, \beta\right)} I_{]0,1[}(a)$$
où $I_{]0,1[}(a)$ est la fonction indicatrice sur l'intervalle $]0, 1[$ ($I_{]0,1[}(a) = 1$ si $a \in]0, 1[$ et $I_{]0,1[}(a) = 1$ si $a \notin]0, 1[$) et où $B(\alpha, \beta)$ est la fonction beta dont l'expression n'est pas importante dans cet exercice. Montrer que la loi a posteriori de $a|\, x_1, ..., x_n$ est aussi une loi beta dont on précisera les paramètres.
- Déterminer l'estimateur du maximum a posteriori du paramètre a noté $\widehat{a}_{\mathrm{MAP}}$ et étudier son comportement lorsque $n \to \infty$.

$\boxed{\textbf{Exercice 2 : Modèle linéaire}}$

On considère n vecteurs aléatoires $X_1, ..., X_n$ appartenant à $\mathbb{R}^2$ définis par

$$X_i = M\theta + e_i$$

où M est une matrice connue de taille 2×2 possèdant toutes les propriétés nécessaires à la résolution de cet exercice, θ est un vecteur inconnu de $\mathbb{R}^2$ et

$$e_i \sim \mathcal{N}_2 \left(\left(\begin{array}{c} 0 \\ 0 \end{array} \right), \sigma^2 \left(\begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array} \right) \right).$$

Pour simplifier, on notera $e_i \sim \mathcal{N}\left(0, \sigma^2 I_2\right)$ (où 0 est le vecteur nul de $\mathbb{R}^2$ et I_2 est la matrice identité de taille 2×2) et on suppose que les vecteurs $e_1, ..., e_n$ sont indépendants. On suppose également (au moins dans un premier temps) que le paramètre σ^2 est connu. On désire effectuer le test d'hypothèses

$$\begin{array}{lll} H_0 & : & \theta = \theta_0 = (0,0)^T \\ H_1 & : & \theta = \theta_1 \neq \theta_0 \end{array}$$

avec $M\theta_1 \neq 0$.

1) En utilisant l'expression de la densité d'un vecteur Gaussien rappelée à la fin de cet exercice, déterminer la vraisemblance des vecteurs $x_1, ..., x_n$ (attention chaque observation x_i est un vecteur de $\mathbb{R}^2$). Montrer que la statistique du test de Neyman-Pearson peut s'écrire

$$T = \theta_1^T M^T \sum_{i=1}^{n} X_i$$

et donner la région critique de ce test.

2) On rappelle que si X est un vecteur gaussien de $\mathbb{R}^2$ de moyenne μ et de matrice de covariance Σ et si a est un vecteur non nul de $\mathbb{R}^2$, alors $a^T X$ est une variable aléatoire gaussienne de moyenne $E\left[a^T X\right] = a^T \mu$ et de variance $a^T \Sigma a$. En déduire la loi de T sous les deux hypothèses H_0 et H_1.

3) En utilisant les résultats de la question précédente, exprimer les risques de première et de seconde espèces α et β en fonction du seuil du test de Neyman Pearson noté S_α, du paramètre θ_1 et de la fonction $\phi(x) = 1 - F(x)$ où F est la fonction de répartition de la loi normale $\mathcal{N}(0,1)$. En déduire l'expression analytique des courbes COR de ce détecteur et montrer qu'elles dépendent de n, σ et de $\nu = \|M\theta_1\| = \sqrt{\theta_1^T M^T M \theta_1}$. Représenter graphiquement la forme approximative de ces courbes COR pour différentes valeurs des paramètres n et de σ.

4) On suppose désormais que le paramètre σ^2 est inconnu. Déterminer l'estimateur du maximum de vraisemblance de ce paramètre. En déduire la règle de décision liée au test du rapport des vraisemblances maximales.

Rappel : on rappelle que la densité de probabilité d'un vecteur gaussien à deux dimensions $X \sim \mathcal{N}_2(\mu, \Sigma)$ de vecteur moyenne μ et de matrice de covariance Σ est

$$\frac{1}{2\pi\sqrt{\det \Sigma}} \exp\left[-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right]$$