Institut National des Télécommunications
Télécom INT, EI2, Module ST 21
Coordonnateur: Wojciech Pieczynski

Introduction aux Statistiques

Sujet d'Examen
6 novembre 2006
Durée : 1h30
Documents autorisés

N. B.
- chaque question est notée sur 2 points ;
- les exercices sont indépendants ;
- penser à admettre le résultat d'une question et passer à la question suivante.

Exercice 1

On considère une suite de variables aléatoires indépendantes $X_1, \ldots, X_n$ à valeurs dans $[0,1]$, de loi admettant pour densité $f_\theta(x) = \theta x + 1 - \frac{\theta}{2}$, où θ est un paramètre dans $[0,2]$. On souhaite estimer θ à partir de $X_1, \ldots, X_n$.

1. Montrer que pour tout $\theta \in [0,2]$, f_θ est bien une densité de probabilité sur $[0,1]$;
2. Montrer que l'espérance de chaque X_i est égale à $\frac{1}{2} + \frac{\theta}{12}$;
3. Montrer que la variance de chaque X_i est égale à $\frac{1}{12} - \frac{\theta^2}{144}$;
4. Montrer que l'estimateur

$$\hat{\theta}_1(X_1,...,X_n) = 12\left(\frac{X_1 + ... + X_n}{n} - \frac{1}{2}\right)$$

est sans biais et qu'il converge en moyenne quadratique.

On dispose d'une connaissance a priori sur le paramètre selon laquelle « il est plutôt proche de 2 », ce que l'on modélise par la densité de probabilité sur l'ensemble des paramètres donnée par $f(\theta) = \frac{\theta}{2}$. On se donne la fonction de perte $L(\theta', \theta) = (\theta' - \theta)^2$.

5. Montrer que l'estimateur bayésien $\hat{\theta}_2$ associé à L est donné par

$$\hat{\theta}_2(X_1,...,X_n) = \frac{\int_0^2 \theta[\prod_{i=1}^n (\theta X_i + 1 - \theta/2)]d\theta}{\int_0^2 [\prod_{i=1}^n (\theta X_i + 1 - \theta/2)]d\theta} \; ;$$

6. On considère maintenant que surestimer le paramètre est deux fois plus grave que de le sous-estimer, ce que l'on modélise par la fonction de perte $L^*(\theta',\theta) = \begin{cases} 2(\theta' - \theta) & \text{si } \theta' > \theta \\ \theta - \theta' & \text{si } \theta' < \theta \end{cases}$
Ecrire la fonction dont la minimisation donne l'estimateur bayésien associé à L^*.

Exercice 2

On souhaite trier automatiquement, à partir du poids, un résultat de pêche de poissons contenant dix mille spécimens, chacun pouvant appartenir à l'une des trois espèces a, b, et c. On suppose que les poids suivent les lois gaussiennes de densités f_a, f_b, f_c données par les moyennes m_a, m_b, m_c et les variances σ_a^2, σ_b^2, σ_c^2. Par ailleurs, on notera π_a, π_b, et π_c les probabilités de pêcher un spécimen de l'espèce, respectivement, a, b, et c.

1. Afin d'estimer les paramètres π_a, π_b, π_c, m_a, m_b, m_c, σ_a^2, σ_b^2, et σ_c^2 on prélève 1000 spécimens dont on note x_1, …, x_{1000} l'espèce (chaque x_i est dans $\Omega = \{a,b,c\}$) et y_1, …, y_{1000} le poids. Proposer des estimateurs pour π_a, m_a, et σ_a^2 ;

2. Donner la méthode bayésienne de classification des 9000 poissons restant minimisant la probabilité d'erreur ;

3. Ecrire la formule, sous forme d'une intégrale, donnant la probabilité de mal classer un poisson de l'espèce a ;

4. On suppose que l'espèce a est précieuse et l'on souhaite limiter la probabilité de l'erreur consistant à classer un poisson de cette espèce dans les espèces b ou c à ε. Décrire rapidement comment doit-on modifier la méthode précédente pour atteindre cet objectif.