Partiel de Statistique

Mardi 8 Janvier 2002, 8h00-10h00

INP TOULOUSE ENSEEIHT

Exercice 1

Dans la fabrication en grande série, on réalise l'expérience suivante : on prélève au hasard des pièces que l'on contrôle et on note le nombre de pièces bonnes tirées jusqu'à l'obtention de la première pièce défectueuse qui est une variable aléatoire X de loi géométrique de paramètre p :

$$P[X = x] = p(1-p)^x \qquad x \in \mathbb{N}$$

Pour une telle loi géométrique, on montre que $E[X] = \frac{1-p}{p}$ et $Var[X] = \frac{1-p}{p^2}$. On désire estimer le paramètre $\theta = \frac{1-p}{p}$ à partir de n observations de la variable X notées $x_1, ..., x_n$. On suppose que $(x_1, ..., x_n)$ est issu d'un échantillon $(X_1, ..., X_n)$ de même loi que X.

1) Ecrire la vraisemblance de $(X_1, ..., X_n)$ au point $(x_1, ..., x_n)$ en fonction de θ. En déduire l'estimateur du maximum de vraisemblance $\widehat{\theta}_{MV}$ du paramètre θ (on prendra soin d'étudier les variations de la vraisemblance en fonction de θ).

2) L'estimateur $\widehat{\theta}_{MV}$ est-il un estimateur sans biais et convergent de θ ?

3) L'estimateur $\widehat{\theta}_{MV}$ est-il l'estimateur efficace du paramètre θ ?

4) On désire construire à partir de $\widehat{\theta}_{MV}$ un intervalle de confiance pour le paramètre θ avec un coefficient de confiance $1 - \alpha = 0.95$. Quelle est la loi asymptotique de l'estimateur $\widehat{\theta}_{MV}$? En déduire une équation du second degré vérifiée par θ permettant de déterminer les bornes θ_1 et θ_2 de l'intervalle de confiance pour θ (on ne cherchera pas à résoudre cette équation).

Exercice 2

Deux informations binaires $\theta_i \in \{0, 1\}$ avec $i = 1, 2$ sont transmises vers un récepteur à travers un canal de transmission. Ces informations sont perturbées par un bruit supposé Gaussien centré de variance σ^2 noté $e_i \sim \mathcal{N}(0, \sigma^2)$. Le message reçu s'écrit alors $z = (x, y)$ avec $x = \theta_1 + e_1$ et $y = \theta_2 + e_2$, où les erreurs e_1 et e_2 sont supposées indépendantes (et de même loi). Le problème consiste à retrouver le vecteur $\theta = (\theta_1, \theta_2)$ à partir des observations x et y, i.e. à partir de z. On suppose qu'on dispose d'une information a priori sur les bits "0" et "1" qui se traduit par $P(\theta_i = 0) = P(\theta_i = 1) = \frac{1}{2}$.

1) Déterminer les probabilitiés $p_{00} = P[\theta = (0,0)|\, z]$, $p_{11} = P[\theta = (1,1)|\, z]$, $p_{01} = P[\theta = (0,1)|\, z]$ et $p_{10} = P[\theta = (1,0)|\, z]$.

2) Déterminer l'estimateur de θ noté $\widehat{\theta}_{MMSE}$ qui minimise $E\left[\left(\widehat{\theta} - \theta\right)^2\right]$. Cet estimateur est-il adapté au problème posé ?

3) Déterminer l'estimateur du maximum a Posteriori du paramètre $\theta = (\theta_1, \theta_2)$ noté $\widehat{\theta}_{MAP}$. Représenter dans le plan (x, y) les domaines correspondant à $\widehat{\theta}_{MAP} = (0,0)$, $\widehat{\theta}_{MAP} = (0,1)$, $\widehat{\theta}_{MAP} = (1,0)$ et $\widehat{\theta}_{MAP} = (1,1)$. Commenter la règle de décision obtenue.

4) Comment les résultats de la question 3) se modifient-ils lorsque $P(\theta_i = 0) = p < 1/2$?

Exercice 3

On dispose de n observations $x_1, ..., x_n$ issues d'un échantillon $(X_1, ..., X_n)$ de loi normale $\mathcal{N}(0, \sigma^2)$.

1) Déterminer suivant les valeurs de σ_0^2 et de σ_1^2, la statistique $T(X_1, ..., X_n)$ et la région critique du test de Neyman-Pearson associé au problème suivant :

$$\begin{cases} H_0 : \sigma^2 = \sigma_0^2 \\ H_1 : \sigma^2 = \sigma_1^2 \end{cases}$$

Commenter la forme de la statistique de test. Dans la suite de ce problème, on supposera $\sigma_1 > \sigma_0$.

2) Déterminer la valeur du risque de seconde espèce β lorsque $\alpha = 0.01, \sigma_0^2 = 1, \sigma_1^2 = 2 > 0$ et $n = 60$.

3) Déterminer les courbes COR (caractéristiques opérationnelles du récepteur) associées au problème précédent. On posera

$$\Phi_n(x) = \int_x^{+\infty} f_n(u)\, du$$

où $f_n(u)$ est la densité d'une loi du χ_n^2 et on notera $\Phi_n^{-1}(x)$ son inverse. Comment la puissance du test de Neyman Pearson dépend-elle de σ_0 et de σ_1 ? Représenter la forme approximative des courbes COR pour différentes valeurs de ces paramètres.

4) Quelle est la loi asymptotique de la statistique $T(X_1, ..., X_n)$ sous les deux hypothèses H_0 et H_1 ? En utilisant cette loi asymptotique, déterminer la valeur du risque de seconde espèce β lorsque $\alpha = 0.01, \sigma_0^2 = 1, \sigma_1^2 = 2 > 0$ et $n = 100$.