Partiel de Statistique

Mardi 19 Décembre 2000, 10h00-12h00

INP ENSEEIHT

On considère les observations $x_i, i = 1, ..., n$ (avec $n = 10$) définies par

$x_1 = 1$	$x_2 = 0$	$x_3 = 1$	$x_4 = 1$	$x_5 = 1$	$x_6 = 1$	$x_7 = 1$	$x_8 = 2$	$x_9 = 0$	$x_{10} = 0$

et on suppose que les variables aléatoires associées à ces observations sont indépendantes et issues de la même loi de Poisson $P(\lambda)$.

Rappels : on rappelle que si X suit une une loi de Poisson de paramètre λ, on a
$E[X] = Var[X] = \lambda$ *et* $\varphi_X(t) = E\left[e^{itX}\right] = \exp\left[\lambda\left(e^{it} - 1\right)\right]$

$1^{ère}$ Partie : Estimation

1) Déterminer l'estimateur du maximum de vraisemblance du paramètre λ noté $\widehat{\lambda}_{MV}$ (on prendra soin d'établir un tableau de variation associé à la fonction à maximiser). Dire si $\widehat{\lambda}_{MV}$ est un estimateur sans biais, convergent et efficace de λ.

2) En supposant que n est suffisamment grand pour utiliser les résultats du théorème de la limite centrale, donner une approximation de la loi de $\widehat{\lambda}_{MV}$. En déduire un intervalle de confiance pour le paramètre λ avec un seuil de confiance $\alpha = 0.95$.

3) De nombreuses études en astrophysique ont montré que le paramètre λ pouvait être muni de la loi a priori

$$g(\lambda) = ce^{-\lambda c} I_{\mathbb{R}^+}(\lambda)$$

où c est un paramètre connu et $I_{\mathbb{R}^+}(\lambda)$ est la fonction indicatrice de $\mathbb{R}^+$. Déterminer l'estimateur du Maximum A Posteriori du paramètre λ noté $\widehat{\lambda}_{MAP}$. Comparer $\widehat{\lambda}_{MV}$ et $\widehat{\lambda}_{MAP}$ pour n "grand" et commenter.

$2^{ème}$ Partie : Test d'ajustement

On désire vérifier si les données x_i sont issues d'un échantillon de variables aléatoires X_i de loi de Poisson (de paramètre inconnu). Effectuer un test du chi2 avec un risque $\alpha = 0.05$ en répartissant les données dans trois classes judicieusement choisies et conclure.

$3^{ème}$ Partie : Tests d'Hypothèses

On désire tester les deux hypothèses

$$\begin{cases} H_0 : \lambda = \lambda_0 \\ H_1 : \lambda = \lambda_1 \text{ (avec } \lambda_1 < \lambda_0) \end{cases}$$

1) Vérifier que la statistique du test de Neyman-Pearson peut s'écrire $T = \sum_{i=1}^{n} X_i$ et déterminer la région critique associée.

2) Déterminer la fonction caractéristique de T et en déduire que T suit une loi de Poisson que l'on précisera sous chaque hypothèse.

Remarque. Pour ceux qui n'arriveraient pas à traiter cette question, on supposera dans la suite que T suit une loi de Poisson de paramètre 10 sous H_0 et de paramètre 1 sous H_1.

3) Préciser le test de puissance maximale tel que le risque de première espèce α vérifie $\alpha \leq 0.05$. On précisera le risque maximal α, la décision prise au vu des données $x_i, i = 1, ..., 10$ et la puissance de ce test. Pour les applications numériques, on prendra $\lambda_0 = 1$ et $\lambda_1 = 0.1$.

4) On suppose que n est suffisamment grand pour pouvoir utiliser les résultats du théorème de la limite centrale.

a) Donner la loi approchée de T issue de ce théorème.

b) Quelle est la valeur du seuil obtenue lorsqu'on confond la loi de T avec son approximation. En comparant avec la valeur obtenue précédemment, dire ce que vous pensez de cette approximation pour $n = 10$.

c) Déterminer les courbes COR (caractéristiques opérationnelles du récepteur) découlant de cette loi approchée. On posera

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du$$

et on notera $\Phi^{-1}(x)$ son inverse. En supposant que n est suffisamment grand pour faire les approximations nécessaires, déterminer les paramètres qui influent sur la performance asymptotique $(n \to \infty)$ du test. De ces deux cas

Premier Cas : $n = 100, \lambda_0 = 1, \lambda_1 = 0.1$

Deuxième Cas : $n = 100, \lambda_0 = 2, \lambda_1 = 1.1$

indiquer celui qui engendre la meilleure performance.