UNIVERSITÉ TOULOUSE 1 CAPITOLE

TSE
Toulouse School of Economics
École d'économie de Toulouse

Année Universitaire 2012-2013
SESSION 1

MASTER 1

MATHEMATICAL STATISTICS
(durée 1h30)

Jeudi 10 janvier 2013 ~ 09h00 -10h30

C. THOMAS

****

Without documents.

Please justify all your answers. No calculator.

Q 1

We consider a sample from the gaussian model where $X \sim \mathcal{N}(\mu, \sigma^2)$ and $\sigma \in]0, \infty[$. We consider the following estimator of σ^2 : $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$.

a) What is the theoretical bias B of this estimator $\hat{\sigma}^2$?

b) Propose an estimator $\hat{B}$ for the bias of $\hat{\sigma}^2$ based on N independent samples of size n, denoted by X_i^m for $i = 1, \cdots, n$, and $m = 1, \cdots, N$. Indicate at least two arguments in favor of the choice of this estimator $\hat{B}$?

Q 2

For a sample of stores in a given region, one considers the characteristic X corresponding to their annual sales in euros. A point has coordinates $(0.2, 0.8)$ on the Lorenz curve of this variable X. How do you interpret these coordinates ?

Q 3

Explain how you measure the dependence (correlation) between two qualitative variables.

Q 4

With a Bernoulli sample of size $n = 3$ and parameter θ, one would like to test $H_0 : \theta = \frac{1}{3}$ against the alternative $\theta = \frac{2}{3}$. Find a constant K so that the level of the test based on the following critical region $\{\sum_{i=1}^{3} X_i \geq K\}$ is less than or equal to 0.05.

Problem

Let $X_1, \cdots, X_n$ be a sample of size n from the exponential distribution with parameter θ and density given by $f_\theta(x) = \frac{1}{\theta}\exp(-\frac{1}{\theta}x)$, for $x \geq 0$ and 0 otherwise. The mean and variance of this distribution are given by $\mathbb{E}(X) = \theta$ and $\mathbb{V}ar(X) = \theta^2$.

Part 1.

1) find a sufficient statistic for θ in this model.

2) compute the Fisher information contained in this sample about the parameter θ.

3) find the maximum likelihood estimator $\hat{\theta}$ of θ based on the sample $X_1, \cdots, X_n$.

4) prove that $\hat{\theta}$ is the optimal unbiased estimator of θ.

5) the objective now is to estimate the parameter $\lambda = \frac{1}{\theta}$. Find the maximum likelihood estimator $\hat{\lambda}$ of λ.

Part 2. We consider the following four estimators of θ

$$\hat{\theta}_1 = \frac{X_1 + X_2}{2}, \hat{\theta}_2 = \bar{X}, \hat{\theta}_3 = \min(X_1, X_2, X_3), \hat{\theta}_4 = \hat{q}_{0.5},$$

where $\hat{q}_{0.5}$ is the empirical median of the sample.

1) find the mean, variance and mean square error of the three estimators $\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3$ and compare them at finite distance.

2) is there an efficient estimator among these three?

3) what is the relative efficiency of $\hat{\theta}_1$ with respect to $\hat{\theta}_2$? and of $\hat{\theta}_3$ with respect to $\hat{\theta}_2$? explain the results with information arguments.

4) what is the asymptotic distribution of $\hat{\theta}_2$ and $\hat{\theta}_4$ when the sample size increases to infinity?

5) what is the asymptotic relative efficiency of $\hat{\theta}_4$ with respect to $\hat{\theta}_2$?

Part 3. We have a first sample $X_1, \cdots, X_n$ from the exponential distribution with mean $\frac{1}{\lambda}$ and a second sample $Y_1, \cdots, Y_n$, from the exponential distribution with mean $\frac{1}{\lambda+\mu}$. We assume that $Y_1, \cdots, Y_n$ are independent from $X_1, \cdots, X_n$.

1) find the maximum likelihood estimator $\hat{\mu}$ of μ based on the pooled data $X_1, \cdots, X_n, Y_1, \cdots, Y_n$

2) derive the asymptotic distribution of $\hat{\mu}$.

En français Justifiez toutes vos réponses. Pas de calculatrice.

Q 1

On considère un échantillon du modèle gaussien où $X \sim \mathcal{N}(\mu, \sigma^2)$ et $\sigma \in]0, \infty[$. On considère l'estimateur suivant de σ^2 : $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$.

a) Quel est la biais théorique B de cet estimateur $\hat{\sigma}^2$?

b) Proposer un estimateur $\hat{B}$ pour le biais de $\hat{\sigma}^2$ basé sur N échantillons indépendents de taille n, notés X_i^m pour $i = 1, \cdots, n$, et $m = 1, \cdots, N$. Indiquer au moins deux arguments en faveur du choix de cet estimateur $\hat{B}$?

Q 2

Pour un échantillon de magasins dans une région donnée, on considère une caractéristique X correspondant au chiffre d'affaires annuel en euros. Un point a pour coordonnées $(0.2, 0.8)$ sur la courbe de Lorenz de cette variable X. Comment interprétez-vous ces coordonnées ?

Q 3

Expliquer comment on mesure la dépendence (corrélation) entre variables qualitatives.

Q 4

Avec un échantillon de Bernoulli de taille $n = 3$ et de paramètre θ, on voudrait tester $H_0 : \theta = \frac{1}{3}$ contre l' alternative $\theta = \frac{2}{3}$. Trouver une constante K de façon que le niveau du test basé sur la région critique suivante $\{\sum_{i=1}^{3} X_i \geq K\}$ soit inférieur ou égal à 0.05.

Problème

Soit $X_1, \cdots, X_n$ un échantillon de taille n de la loi exponentielle de paramètre θ et de densité donnée par $f_\theta(x) = \frac{1}{\theta}\exp(-\frac{1}{\theta}x)$, pour $x \geq 0$ et 0 sinon. La moyenne et la variance de cette distribution sont données par $\mathbb{E}(X) = \theta$ et $\mathbb{V}ar(X) = \theta^2$.

Partie 1.

1) trouver une statistique exhaustive pour le paramètre θ dans ce modèle.

2) calculer l' information de Fisher contenue dans l'échantillon sur le paramètre θ.

3) trouver l'estimateur du maximum de vraisemblance $\hat{\theta}$ de θ basée sur l'échantillon $X_1, \cdots, X_n$.

4) montrer que $\hat{\theta}$ est l'estimateur sans biais optimal de θ.

5) l'objectif maintenant est d'estimer un paramètre $\lambda = \frac{1}{\theta}$. Trouver l'estimateur du maximum de vraisemblance $\hat{\lambda}$ de λ.

Partie 2. On considère les quatre estimateurs suivants de θ

$$\hat{\theta}_1 = \frac{X_1 + X_2}{2}, \hat{\theta}_2 = \bar{X}, \hat{\theta}_3 = \min(X_1, X_2, X_3), \hat{\theta}_4 = \hat{q}_{0.5},$$

où $\hat{q}_{0.5}$ est la médiane empirique de l'échantillon.

1) trouver la moyenne, la variance et l'erreur quadratique moyenne des trois estimateurs $\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3$ et les comparer à distance finie.

2) y a-t-il un estimateur efficace parmi ces trois estimateurs ?

3) quelle est l'efficacité relative de $\hat{\theta}_1$ par rapport à $\hat{\theta}_2$? et de $\hat{\theta}_3$ par rapport à $\hat{\theta}_2$? expliquer les résultats avec des arguments relatifs à l'information.

4) quelle est la loi asymptotique de $\hat{\theta}_2$ et de $\hat{\theta}_4$ quand la taille de l'échantillon croit vers l'infini ?

5) quelle est l'efficacité relative asymptotique de $\hat{\theta}_4$ par rapport à $\hat{\theta}_2$?

Partie 3. On a un premier échantillon $X_1, \cdots, X_n$ de la loi exponentielle de moyenne $\frac{1}{\lambda}$ et un deuxième échantillon $Y_1, \cdots, Y_n$, de la loi exponentielle de moyenne $\frac{1}{\lambda+\mu}$. On suppose que $Y_1, \cdots, Y_n$ sont indépendants de $X_1, \cdots, X_n$.

1) trouver l'estimateur du maximum de vraisemblance $\hat{\mu}$ de μ basé sur l'échantillon combiné $X_1, \cdots, X_n, Y_1, \cdots, Y_n$

2) trouver la loi asymptotique de $\hat{\mu}$.