Statistiek (WISB263)

Final Exam

January 30, 2017

Schrijf uw naam op elk in te leveren vel. Schrijf ook uw studentnummer op blad 1.

(The exam is an *open–book* exam: notes and book are allowed. The scientific calculator is allowed as well).

The maximum number of points is 100.

Points distribution: 25-20-30-25

1. Given two parameters $a > 0$ and $k > 0$, let $\mathbf{X} = \{X_1, \dots, X_n\}$ be a random sample of n i.i.d. observations sampled from the random variable X with density function:

$$f_X(x; a, k) := \begin{cases} k\, e^{-k\,(x-a)} & x \geq a, \\ 0 & x < a \end{cases}$$

   (a) (8pt) Find sufficient statistics for a, k and for the couple (a, k).

   (b) (5pt) Determine, in case it exists, the maximum likelihood estimator of a in case k is known.

   (c) (5pt) Determine, in case it exists, the maximum likelihood estimator of k in case a is known.

   (d) (7pt) Determine, in case it exists, the maximum likelihood estimator of the couple (a, k).

2. We consider the following three random samples of size 100:

$$\mathbb{X}_i := \{X_{i,1}, X_{i,2}, \dots X_{i,100}\},$$

with $i \in \{1, 2, 3\}$. Each sample $\mathbb{X}_i$ consists of i.i.d. normal random variables, such that $X_{i,j} \sim N(50, \sigma_i^2)$ for any $j \in \{1, \dots, 100\}$. Moreover the samples are independent (i.e. $X_{i,j} \perp X_{\ell m}$, for any $i \neq \ell$). We want to test:

$$\begin{cases} H_0: & \sigma_1^2 = \sigma_2^2 = \sigma_3^2, \\ H_1: & \text{the variances are not equal.} \end{cases}$$

   (a) [10pt] Show that the Generalized Likelihood Ratio Test (GLRT) statistic Λ is such that:

$$-2 \log \Lambda = 300 \log \left(\frac{1}{3} \sum_{i=1}^{3} s_i^2 \right) - 100 \sum_{i=1}^{3} \log s_i^2$$

   where $s_i^2 := 1/100 \sum_{j=1}^{100} (X_{i,j} - 50)^2$, with $i \in \{1, 2, 3\}$.

   (b) [10pt] If the collected data $\mathbf{x}_i = \{x_{i,1}, \dots, x_{i,100}\}$, with $i \in \{1, 2, 3\}$, are such that:

$$\textstyle\sum_{j=1}^{100} x_{1,j} = 5040, \qquad \sum_{j=1}^{100} x_{2,j} = 4890, \qquad \sum_{j=1}^{100} x_{3,j} = 4920,$$

$$\textstyle\sum_{j=1}^{100} x_{1,j}^2 = 264200, \quad \sum_{j=1}^{100} x_{2,j}^2 = 250000, \quad \sum_{j=1}^{100} x_{2,j}^2 = 251700$$

   perform a GLRT at $\alpha = 0.05$ level of significance (you can consider the sample size $n = 100$ large enough for applying large sample results).

3. The life times (in hours) of $n = 30$ batteries have been measured from a company interested in the performances of a new product. In this way, a sample $\mathbb{X} = \{X_1, \dots X_{30}\}$ of i.i.d. random variable X_j, representing the life time of the j−th battery, has been collected. In the following table the empirical cumulative distribution function $\hat{F}_{30}(x)$ (i.e. $\hat{F}_n(x) = 1/n \sum_{j=1}^{n} \mathbf{1}(X_j \leq x)$) is reported:

x (in hours)	1	2	4	6	8	11	13	27	29	42
$\hat{F}_{30}(x)$	7/30	12/30	16/30	20/30	23/30	26/30	27/30	28/30	29/30	1

(a) [6pt] Determine an estimator of the probability that the battery produced lasts more than 9 hours (i.e. $\mathbb{P}(X > 9)$).

(b) [8pt] Derive an approximated 95% confidence interval for the probability that the battery produced lasts more than 9 hours.

Due to previous statistical analyses performed on similar batteries, we can assume now that the sample is a collection of 30 i.i.d. exponential random variable with expected value θ (i.e. $X_i \sim \mathrm{Exp}(1/\theta)$).

(c) [8pt] Under these parametric assumptions, calculate the maximum likelihood estimator of the probability that the battery produced lasts more than 9 hours.

(d) [8pt] If we denote with $p(\theta)$ the probability that the battery produced lasts more than 9 hours, propose a test for testing the hypotheses:

$$\begin{cases} H_0: & p = 0.32 \\ H_1: & p = 0.16. \end{cases}$$

at the α level of significance.

4. Let the independent random variables $Y_1, Y_2, \dots, Y_n$ be such that we have the following linear model:

$$Y_i = \alpha + \beta x_i + \epsilon_i$$

for $i = 1, \dots, n$, where ϵ_i are i.i.d. normal random variables such that $\epsilon_i \sim N(0, \sigma^2)$. Let $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ be the model in the matrix formalism. After we collected a sample of size $n = 42$, we have that:

$$(\mathbf{X}^\top \mathbf{X})^{-1} = \begin{pmatrix} 0.03 & -0.015 \\ -0.015 & 0.04 \end{pmatrix}$$

Furthermore, we know that the least squares estimate is $\hat{\boldsymbol{\beta}}^\top = (\hat{\beta}_0, \hat{\beta}_1) = (1.90, 0.65)$ and that the residual sum of squares $\|\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\|^2 = 160$.

(a) [8pt] Compute the 95% confidence intervals for β_0 and β_1

(b) [10pt] Consider the test:

$$\begin{cases} H_0: & \beta_0 = 2, \\ H_1: & \beta_0 \neq 2. \end{cases}$$

Will H_0 be rejected at a significance level of 5%? And at a significance level of 1%?

(c) [7pt] Under the previous H_0, it holds that $\mathbb{P}(\hat{\beta}_0 > 1.90) = 0.61$ and that $\mathbb{P}(\hat{\beta}_0 < 1.90) = 0.39$. For which values of the significance level α, the null hypothesis H_0 will be rejected with the given data?