Statistiek (WISB263)

Resit Exam

April 19, 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: 32-20-26-22

1. Let $\mathbf{X} = \{X_1, \dots, X_n\}$ be a random sample of n i.i.d. Poisson random variables with parameter λ.

   (a) (8pt) Find the maximum likelihood for λ and its asymptotic sampling distribution.

   (b) (8pt) Find the maximum likelihood estimator for the parameter $\mu = e^{-\lambda}$.

   Suppose now that, rather than observing the actual values of the random variables X_i, we are just able to register whether they are null or positive. More precisely, only the events $X_i = 0$ or $X_i > 0$ for $i = 1, \dots, n$ are observed.

   (c) (8pt) Find the maximum likelihood for λ for these new observations.

   (d) (8pt) When does the maximum likelihood estimator not exist? Assuming that the true value of λ is λ_0, compute the probability that the maximum likelihood estimator does not exist.

2. Let $\mathbf{X} = \{X_1, \dots, X_n\}$ be a random sample of n i.i.d. random variables with densities:

$$f_X(x;\theta) = \begin{cases} \frac{\theta^3}{2} x^2 e^{-\theta x} & \text{if } x > 0, \\ 0 & \text{otherwise} \end{cases}$$

   with $\theta > 0$ is an unknown parameter. Moreover, consider another random sample $\mathbf{Y} = \{Y_1, \dots, Y_n\}$ of n i.i.d. random variables with densities:

$$f_Y(y;\mu) = \begin{cases} \frac{\mu^3}{2} y^2 e^{-\mu y} & \text{if } y > 0, \\ 0 & \text{otherwise} \end{cases}$$

   with $\mu > 0$ is another unknown parameter. We further assume that the two sample are independent (i.e. $X_i \perp Y_j$, for all i, j).

   (a) [10pt] Find the Generalized Likelihood Ratio Test (GLRT) statistic for testing:

$$\begin{cases} H_0: & \theta = \mu, \\ H_1: & \theta \neq \mu. \end{cases}$$

   Let us define now the following statistic:

$$T := \frac{\sum_{i=1}^n X_i}{\sum_{i=1}^n X_i + \sum_{j=1}^n Y_j}$$

   (b) [10pt] Show that the GLRT rejects H_0 if $T(1-T) < k$, for a suitable constant k.

3. A company wants to monitor the efficiency of two employees in completing an assigned task. For this reason, the performances of two employees (denoted by **A** and **B**) were measured by recording the times needed to complete the assigned tasks. Hence, the following two samples have been collected:

$$\mathbf{x_A} = \{5.18, 13.43, 6.31, 3.18, 4.91, 11.07\},$$

$$\mathbf{x_B} = \{5.50, 18.16, 8.14, 9.14, 14.24, 10.72\}$$

where the duration of each task is measured in hours.

(a) [10pt] Perform a test at 10% of significance for testing the hypothesis that *employee* **A** *is faster than* **B**. Discuss critically the choice of the test used.

Suppose now that the time T needed by an employee for completing a task can be modeled by a continuous random variable with the following probability density function:

$$f_T(t;\theta) = \begin{cases} \frac{1}{2\theta\sqrt{t}} e^{-\frac{\sqrt{t}}{\theta}} & \text{if } t > 0, \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

with $\theta > 0$ an unknown parameter.

(b) [8pt] Given a sample $\mathbb{T} = \{T_1, \ldots, T_n\}$ of i.i.d random variables sampled from $f_T(t;\theta)$, determine the maximum likelihood estimator of the probability $\mathbb{P}_\theta(T > 7)$.

(c) [8pt] Under the parametric model (1) for the random variable T and given the samples $\mathbf{x_A}$, $\mathbf{x_B}$, estimate the probability that the time needed by an employee for completing a task is larger than 7 hours, under the further assumption that 55% of the employees are similar to employee **A** and 45% to employee **B**.

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

$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 (x_i - 3.5)_+ + \epsilon_i$$

for $i = 1, \ldots, n$, where ϵ_i are i.i.d. normal random variables such that $\epsilon_i \sim N(0, \sigma^2)$ and with $(y)_+$ we denoted the positive part of the real number y (i.e. $(y)_+ := \max(0, y)$). We collect the following sample of observations

$$\mathbf{y} = \{1, 2, 4, 5, 4, 3, 1\}$$

corresponding to the predictors:

$$\mathbf{x} = \{0, 1, 2, 3, 4, 5, 6\}$$

(a) [8pt] If we rewrite the linear model using the usual matrix formalism

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

write down the design matrix $\mathbf{X}$ of the linear model.

(b) [6pt] Given that

$$(\mathbf{X}^\top \mathbf{X})^{-1} = \begin{pmatrix} 0.65 & -0.24 & 0.35 \\ -0.24 & 0.14 & -0.26 \\ 0.35 & -0.26 & 0.65 \end{pmatrix}$$

estimate the model coefficients and write down the fitted model.

(b) [8pt] Calculate the prediction of the fitted model at $x = 4.5$. Assuming that the sum of squared residuals equals 7.8, calculate a 95% confidence interval for this prediction.