Statistiek (WISB263)

Resit Exam

April 15, 2019

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 110 (10 extra BONUS points!!).
Grade= $min(100, \text{points})$.
Points distribution: 22–30–20–28(+10 extra BONUS points!!)

1. Let $\mathbf{X} = \{X_1, \dots, X_n\}$ be a random sample of i.i.d. random variables such that $X_i \sim \text{Unif}[\theta - 1/2, \theta + 1/2]$, for $i = 1, \dots, n$, where $\text{Unif}[\theta - 1/2, \theta + 1/2]$ denotes the uniform distribution in the interval $[\theta - 1/2, \theta + 1/2]$.
   (a) [7pt] Prove that $(X_{(1)}, X_{(n)})$ is a sufficient statistic for θ, where $X_{(1)} := \min_i X_i$ and $X_{(n)} := \max_i X_i$.
   (b) [7pt] Write the likelihood and find the maximum likelihood estimators $\hat{\theta}_{MLE}$ of θ. Show that the MLE is not unique.
   (c) [8pt] State conditions on the sample $\mathbf{X}$, in order to have that $\overline{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ is a MLE. How is it possible that the MLE does not depend *only* on the sufficient statistic?

2. Let us consider a die whose probabilities of getting 1, 2, 5 and 6 differ from 1/6. Given a discrete random variable Y, defined on the outcomes space $\Omega := \{1, 2, 3, 4, 5, 6\}$, for any $i \in \Omega$ we denote with p_i the probability mass function of Y (i.e. $p_i := \mathbb{P}(Y = i)$). We assume:
$$p_1 = p_2 = 1/6 - \theta, \quad p_3 = p_4 = 1/6, \quad p_5 = p_6 = 1/6 + \theta,$$
with $\theta \in \mathbb{R}$, and $|\theta| < 1/6$. Then, we perform an experiment by rolling *independently* the die n times and we collect the the number of times X_i that the outcome i appeared in the experiment. Hence, we have the random sample $\mathbf{X} = \{X_1, X_2, X_3, X_4, X_5, X_6\}$.
   (a) [7pt] Find a sufficient statistic for θ.
   (b) [7pt] Find the maximum likelihood estimator $\hat{\theta}_{MLE}$ of θ.
   (c) [7pt] What happens to the MLE of θ if in the sample there are only outcomes of type 3 and 4?
   (d) [9pt] We want to test now whether the die is actually biased. Thus, we consider the test:
$$\begin{cases} H_0: & \theta = 0, \\ H_1: & \theta \neq 0 \end{cases}$$
   If we collect the sample:
$$\mathbf{x} = \{9, 15, 19, 19, 14, 24\}$$
   test these hypotheses at $\alpha = 0.05$ level of significance.

3. Consider the two independent random variables U and V such that $U \sim \chi^2_m$ and $V \sim \chi^2_n$. Furthermore, let us consider the sample $\mathbf{X} := \{X_1, \dots, X_n\}$ of i.i.d. random variables such that $X_i \sim N(0, \sigma^2)$, where σ^2 is unknown.
   (a) [7pt] Prove that $U + V \sim \chi^2_{n+m}$
   (b) [6pt] Show that the maximum likelihood estimator of σ^2 is given by $\frac{1}{n} \sum_{i=1}^n X_i^2$.
   (c) [7pt] Give a $(1 - \alpha)$–confidence interval for σ^2.

4. The values of the real function:

$$x(t) = \beta_0 + \beta_1 t + \beta_2 t^2$$

have been measured at known time points t_i, $i = 1, 2, \ldots, n$. Due to the presence of measurement errors, any measurement has been modeled as:

$$X_i = \beta_0 + \beta_1 t_i + \beta_2 t_i^2 + \epsilon_i$$

where $i = 1, 2, \ldots, n$ and ϵ_i are i.i.d. random variables with zero expectation and variance σ^2.

(a) [7pt] Find the least squares estimators $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$ for the parameters $\beta_0, \beta_1, \beta_2$.

(b) [7pt] Calculate $\mathbb{E}(\hat{\beta}_i)$ for $i \in \{0, 1, 2\}$, and $\text{Cov}(\hat{\beta}_i, \hat{\beta}_j)$ for $i, j \in \{0, 1, 2\}$.

(c) [7pt] If we define $\hat{x}(t) := \hat{\beta}_0 + \hat{\beta}_1 t + \hat{\beta}_2 t^2$ and a, b are two real number such that $b > a > 0$, is the following statistic

$$\hat{I} := \int_a^b \hat{x}(t) dt$$

an unbiased estimator for the integral $I := \int_a^b x(t) dt$?

(d) [7pt] Find the $\text{Var}(\hat{I})$.

BONUS [10pt] Suppose that a *one-sided* statistical test is based on a statistic T such that, under H_0, it has distribution function F. Furthermore, suppose that the rejection region of the test consists of large values of T. If the test is applied to a *first* set of data, the statistic $T = t_1$ is observed, with an associated observed p-value of $p_1 := 1 - F(t_1) = 0.09$. On a second *independent* set of data the same test is applied, and this time a statistic $T = t_2$ is observed, with an associated observed p-value of $p_2 := 1 - F(t_2) = 0.07$.

(a) [10pt] Derive a way to *aggregate* the two p-values p_1, p_2 into a single *overall* p-value.
[**Hint:** if $U \sim \text{Unif}[0, 1]$, then $V := -2\ln(U) \sim \text{Exp}(1/2)(= \chi_2^2)$].