Statistiek (WISB263)

Final Exam

June 27, 2022

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 use of a laptop is allowed as well, under the restriction that the invigilator can look at the screen at all times, and that students are not allowed to type on the computer and wifi is off. The scientific calculator is also allowed).`

The maximum number of points is 110 (10 extra BONUS points!!).
Grade= $min(100, \text{points})$.
Points distribution: 25–20–20-25-10 (+10 extra BONUS points!!)

1. Suppose that n integers are drawn uniformly at random with replacement from the set $\mathcal{I} \equiv \{1, 2, \dots, N\}$.
   (a) [5pt] Find the method of moments estimator $\hat{N}_1$ of N.
   (b) [5pt] Calculate $\mathbb{E}(\hat{N}_1)$ and $\text{Var}(\hat{N}_1)$.
   (c) [7pt] Find the Maximum Likelihood Estimator (MLE) $\hat{N}_2$ of N.
   (d) [4pt] Show that $\mathbb{E}(\hat{N}_2)$ is approximately $\frac{n}{n+1}N$ (use a continuous approximation of the discrete distribution of the MLE). With the same approximation show that $\text{Var}(\hat{N}_2)$ is approximately $N^2\left(\frac{n}{n+2} - \frac{n^2}{(n+1)^2}\right)$.
   (e) [4pt] Suppose that you have collected the sample $\mathbf{x} = \{28, 6, 22, 15\}$. Among $\hat{N}_1, \hat{N}_2$, which estimator would you prefer for estimating N?

2. Let $\mathbf{X} = \{X_1, \dots X_n\}$ a random sample of i.i.d. random variables with probability density function (pdf):

$$f_X(x) = \frac{1-\beta}{x^\beta},$$

   with $x \in (0,1)$ and $0 < \beta < 1$.
   (a) [6pt] Find a sufficient statistic for β.
   (b) [8pt] Determine the MLEs for β and for $n/(\beta - 1)$. Can you write these MLEs in terms of the sufficient statistic? Explain clearly the reason.
   (c) [6pt] Which is the asymptotic variance of the MLE?

3. The time Y (in years) served in prison for a certain crime was believed to follow a distribution with the following probability density function (pdf):

$$f_Y^{(1)}(y) = \frac{1}{9}y^2, \;\; 0 \le y \le 3.$$

   In order to check this law, a study was conducted in one prison on 100 people. It was reported that 16 convicted served less than one year in jail, 32 served between one and two years, and 52 served between two and three years.
   (a) [8pt] Are these data consistent with the pdf $f_Y^{(1)}(y)$? Perform an appropriate hypothesis test at 0.05 level of significance.
   (b) [7pt] Another theory was proposed for the distribution of Y. According to this new theory, we have a linear pdf:

$$f_Y^{(2)}(y) = \frac{2}{9}y, \;\; 0 \le y \le 3.$$

   Perform the same analysis as in point (a).

(c) [5pt] Imagine now we have only one realization y of the random variable Y. Perform the most powerful test for testing:

$$\begin{cases} H_0: & f_Y = f_Y^{(1)}, \\ H_1: & f_Y = f_Y^{(2)}. \end{cases}$$

If $\alpha = 0.05$, find the rejection region of the test and its power. In case $y = 0.3$ do we reject H_0 at the 0.05 level of significance?

4. Consider the two samples $\mathbf{X} = \{X_1, X_2, \dots, X_n\}$, with $X_i \overset{i.i.d}{\sim} N(\theta_1, \theta_2^2)$ and $\mathbf{Y} = \{Y_1, Y_2, \dots, Y_m\}$, with $Y_i \overset{i.i.d}{\sim} N(\theta_3, \theta_4^2)$. The two samples are independent, i.e., $X_i \perp\!\!\!\perp Y_j, \forall i, j$.

   (a) [4pt] Find the Maximum Likelihood Estimators (MLE) of $\theta_1, \theta_2, \theta_3$ and θ_4.

   (b) [7pt] if we want to test:

$$\begin{cases} H_0: & \theta_1 = \theta_3 \\ H_1: & \theta_1 \neq \theta_3. \end{cases}$$

   with θ_2 and θ_4 unknown, find the null parameter space Θ_0 and the Generalized Likelihood Ratio Test (GLRT) statistics. Which is its asymptotic distribution?

   (c) [7pt] If we assume now that $\theta_2 = \theta_4$ (still unknown), find also in this case the GLRT statistics and its asymptotic distribution for:

$$\begin{cases} H_0: & \theta_1 = \theta_3 \\ H_1: & \theta_1 \neq \theta_3. \end{cases}$$

   (d) [7pt] We collected the samples $\mathbf{x} = \{9.00, 10.87, 11.32\}$, $\mathbf{y} = \{9.54, 7.82, 7.74, 7.14\}$. Perform the test of point (c) at 0.05 level of significance by using a GLRT.

5. [10pt] We have performed an experiment and we collected 48 measurements. However, each of these 48 real numbers is rounded to the nearest integer. The sum of the original 48 numbers is approximated by the sum of these integers. If we assume that the errors made by rounding off are *i.i.d.* and have a uniform distribution over the interval $(-1/2, 1/2)$, compute *approximately* the probability that the sum of the integers is within two units of the *true* sum.

BONUS [10pt] Imagine that you are a fraudulent scientist and that you have decided to repeat the experiment if its p-value is smaller than a priori fixed value β, with $0 < \beta < 1$, and then to report only the largest one. Show that under H_0, the cumulative distribution function of the p-values of the fraudulent experiment is:

$$F(x; \beta) = \begin{cases} x^2, & \text{if } 0 \le x \le \beta, \\ (1+\beta)x - \beta, & \text{if } \beta < x \le 1. \end{cases}$$