Statistiek (WISB263)

Resit exam

July 12, 2023

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

`(The exam is a CLOSED-book exam: students can bring only two A4-sheets with personal notes. The use of the statistical tables is allowed. The scientific calculator is also allowed).`

The maximum number of points is 100.
Points distribution: 20–24–28–28.

1. (a) [12pt] Consider the sample $\mathbf{X} = \{X_1, \dots X_{500}\}$ of i.i.d. random variables such that $\mathbb{E}(X_i) = 2$ and $\text{Var}(X_i) = 3$. Moreover consider another sample $\mathbf{Y} = \{Y_1, \dots Y_{500}\}$ of i.i.d. random variables such that $\mathbb{E}(Y_i) = 2$ and $\text{Var}(Y_i) = 2$. Moreover, the two samples are independent (i.e., $X_i \perp Y_j, \forall i, j$).
   Find an approximated value of the probability p, defined by:

$$p := \mathbb{P}\left(\sum_{i=1}^{500} X_i > \sum_{i=1}^{500} Y_i + 50 \right).$$

   (b) [8pt] Let $(X_1, \dots, X_n)$ be a sequence of i.i.d. random variables such that $X_i \sim \text{Unif}[0,1]$. We consider the random variables Y_n, defined by:

$$Y_n := \min(X_1, \dots, X_n)$$

   Prove that $Y_n \xrightarrow{d} 0$. Is it also true that $Y_n \xrightarrow{p} 0$?

2. Suppose that a type of electronic component has lifetime T (measured in days) that is exponentially distributed, i.e., T has probability density function $f_T(t;\tau) = \frac{1}{\tau} e^{-t/\tau}$, with $t \in \mathbb{R}_{\geq 0}$ and $\tau > 0$. Five new independent components of this type have been tested, and during the experiment the first failure time was registered at 100 (days). No further observations were recorded.
   (a) [4pt] Which is the the likelihood function of τ?
   (b) [8pt] Compute the maximum likelihood estimate of τ.
   (c) [8pt] Which is the distribution of the Maximum Likelihood Estimator (MLE) $\hat{\tau}_{MLE}$ of τ?
   (d) [4pt] Find the variance of $\hat{\tau}_{MLE}$.

3. Consider one realization y of the discrete random variable Y, attaining values in $\Omega := \{10, 20, 30, 40, 50, 60\}$. Its probability mass function (pmf) $p(y;\theta) := \mathbb{P}_\theta(Y = y)$ depends on the unknown parameter θ, belonging to the discrete parameter space $\Theta := \{1, 2, 3, 4, 5, 6\}$. The pmf $p(y;\theta)$ is given by the following table:

y	10	20	30	40	50	60
$p(y;\theta=1)$	0.5	0.2	0.1	0.1	0.1	0
$p(y;\theta=2)$	0.2	0.5	0.1	0.1	0.1	0
$p(y;\theta=3)$	0.1	0.2	0.5	0.1	0.1	0
$p(y;\theta=4)$	0.1	0.1	0.2	0.5	0.1	0
$p(y;\theta=5)$	0.1	0.1	0.1	0.2	0.5	0
$p(y;\theta=6)$	0	0.1	0.1	0.1	0.2	0.5

   (a) [8pt] Find the maximum likelihood estimator $\hat{\theta}_{MLE}$ of θ.
   (b) [6pt] Is $\hat{\theta}_{MLE}$ unbiased?

(c) [6pt] Suppose we want to test:

$$\begin{cases} H_0: & \theta = 1, \\ H_1: & \theta \neq 1 \end{cases}$$

at $\alpha = 0.03$ level of significance. Propose a test statistic and find the rejection region of the test.

(d) [8pt] In case we have $y = 20$, find an estimate of $\mathrm{Var}(\hat{\theta}_{MLE})$.

4. We suspect that a gambler is cheating, in particular we believe that the gambler is using a biased die, in the sense that the probabilities of getting 1 and 6 differ from 1/6. We then consider a discrete random variable X, attaining values on $\Omega := \{1, 2, 3, 4, 5, 6\}$. For any $i \in \Omega$, we denote with p_i the probability mass function of X (i.e., $p_i := \mathbb{P}(X = i)$). We consider:

$$p_1 = 1/6 - \theta, \quad p_2 = p_3 = p_4 = p_5 = 1/6, \quad p_6 = 1/6 + \theta,$$

with $\theta \in \mathbb{R}$, and $|\theta| < 1/6$. We perform an experiment, by rolling, independently, the die n times. Therefore, 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) [6pt] Write the likelihood function for θ.

(b) [6pt] Find a sufficient statistic for θ.

(c) [6pt] Find the MLE $\hat{\theta}_{MLE}$ for θ. Is it always well defined?

(d) [10pt] In order to prove that the die is manipulated, we want to test:

$$\begin{cases} H_0: & \theta = 0, \\ H_1: & \theta \neq 0. \end{cases}$$

If we collect the sample:

$$\mathbf{x} = \{10, 14, 17, 21, 16, 22\}$$

test these hypotheses at $\alpha = 0.05$ level of significance.