Statistiek (WISB263)

Sketch of solutions: 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:

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

   Solution:
   Call $W_i := X_i - Y_i$. Then by the assumptions, W_i are i.i.d. random variables with common $\mathbb{E}(W_i) = 0$ and $\text{Var}(W_i) = 5$. Hence, by classical CLT:

$$p = \mathbb{P}\left(\frac{1}{\sqrt{500}\sqrt{5}} \sum_{i=1}^{500} W_i > \frac{50}{\sqrt{500}\sqrt{5}}\right) = \mathbb{P}\left(\frac{1}{50}\sum_{i=1}^{500} W_i > 1\right) \overset{CLT}{\approx} 1 - \Phi(1) \approx 0.16$$

   (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 $Y_n \xrightarrow{p} 0$?
   Solution:
   It is enough to prove that $Y_n \xrightarrow{p} 0$, since the convergence in probability implies the convergence in distribution. We need to show that for any $\epsilon > 0$, we have that $\lim_{n\to\infty} \mathbb{P}(|Y_n| > \epsilon) = 0$. We have that:

$$\mathbb{P}(|Y_n| > \epsilon) = 1 - \mathbb{P}(Y_n \le \epsilon) = 1 - F_{Y_n}(\epsilon).$$

   However, by independnece of Y_i, we have:

$$F_{Y_n}(\epsilon) = 1 - \mathbb{P}(Y_n > \epsilon) = 1 - (\mathbb{P}(Y_1 > \epsilon))^n = 1 - (1-\epsilon)^n.$$

   Thus

$$\mathbb{P}(|Y_n| > \epsilon) = (1-\epsilon)^n$$

   Therefore, we conclude:

$$\lim_{n\to\infty} \mathbb{P}(|Y_n| > \epsilon) = 0.$$

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 $t \in \mathbb{R}_{\ge 0}$ and $\tau > 0$. Five new independent components of this type have been tested, and during the experiment the <u>first</u> failure time was registered at 100 (days). No further observations were recorded.

(a) [4pt] Which is the the likelihood function of τ?
Solution:
Denoting with T_i, with $i \in \{1,\dots,5\}$ the failure times of each of the five components, the observed time $U := \min_{i\in\{1,\dots,5\}} T_i$. By the independence of T_i, we have that:

$$F_U(t) = 1 - (\mathbb{P}(T_1 > t))^n = 1 - (1 - F_{T_1}(t))^5 = 1 - e^{5t/\tau}$$

so that the probability density function is:

$$f_U(t) = \frac{5}{\tau} e^{5t/\tau}$$

Hence the likelihood function for τ is:

$$L(\tau; U) = \frac{5}{\tau} e^{5U/\tau}$$

(b) [8pt] Compute the maximum likelihood estimate of τ.
Solution:
Being the likelihood $L(\tau; u)$ the pdf of a random variable $U \sim \text{Exp}(\tau/5)$, then by the invariance principle $\hat{\tau}_{MLE} = 5U$. For our data an estimate of τ is then $100 * 5 = 500$.

(c) [8pt] Which is the distribution of the Maximum Likelihood Estimator (MLE) $\hat{\tau}_{MLE}$ of τ?
Solution:
Since $\hat{\tau}_{MLE} = 5U$ with $U \sim \text{Exp}(\tau/5)$, we have then that $\hat{\tau}_{MLE} \sim \text{Exp}(\tau)$.

(d) [4pt] Find the variance of $\hat{\tau}_{MLE}$.
Solution:
Since $\hat{\tau}_{MLE} \sim \text{Exp}(\tau)$, we have that $\text{Var}(\hat{\tau}_{MLE}) = \tau$.

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.
Solution:
By looking at the table we find:

$$\hat{\theta}_{MLE} = Y/10$$

(b) [6pt] Is $\hat{\theta}_{MLE}$ unbiased?
Solution:
If we calculate for $\theta = 1$ the expected value of $\hat{\theta}_{MLE}$, from the table we have:

$$\mathbb{E}_{\theta=1}(\hat{\theta}_{MLE}) = 1/10\mathbb{E}_{\theta=1}(Y) = 2.1 \neq 1$$

Thus, $\hat{\theta}_{MLE}$ is biased.

(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.
Solution:
We use the generalized likelihood–ratio test statistics:

$$\lambda = \frac{L(\theta_0)}{L(\hat{\theta}_{MLE})} = \frac{p(y|\theta = 1)}{p(y|\hat{\theta}_{MLE})}$$

The possible values of this test statistics are:

	$y = 10$	$y = 20$	$y = 30$	$y = 40$	$y = 50$	$y = 60$
λ	1	0.4	0.2	0.2	0.2	0

We reject H_0 for small values of λ. Since we have $\mathbb{P}(\lambda < 0.4|\theta = 1) = 0.3$, it follows that we reject H_0 at $\alpha = 0.03$ level of significance if $\lambda < 0.4$. Therefore, we reject H_0 for any y in the rejection region: $B = \{30, 40, 50, 60\}$.

(d) [8pt] In case we have $y = 20$, find an estimate of $\text{Var}(\hat{\theta}_{MLE})$.
Solution:
If $y = 20$, then $\hat{\theta}_{MLE} = 2$. Hence,

$$\mathbb{E}_{\theta=2}(\hat{\theta}_{MLE}) = 0.2 + 1 + 0.3 + 0.4 + 0.5 + 0 = 2.4$$

and

$$\mathbb{E}_{\theta=2}(\hat{\theta}^2_{MLE}) = 7.2$$

Therefore, an estimate for the variance is:

$$\widehat{\text{Var}}(\hat{\theta}_{MLE}) = 7.2 - 2.4^2 = 1.44$$

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.
Solution:
The sample is a realization of a multinomial random variable, so that the likelihood can be written as:

$$L(\theta; \mathbf{X}) = \frac{n!}{\prod_{i=1}^{6} X_i!} \left(\frac{1}{6} - \theta\right)^{X_1} \left(\frac{1}{6}\right)^{\sum_{i=2}^{5} X_i} \left(\frac{1}{6} + \theta\right)^{X_6} \tag{1}$$

(b) [6pt] Find a sufficient statistic for θ.
Solution:
The likelihood can be written as

$$L(\theta; \mathbf{X}) = h(\mathbf{X}) e^{X_1 \log(1/6 - \theta) + X_6 \log(1/6 + \theta)}$$

with $h(\mathbf{X}) := n!\,(1/6)^{\sum_{i=2}^{5} X_i} / \prod_{i=1}^{6} X_i!$. Hence, (X_1, X_6) is a sufficient statistics for θ by the factorization theorem.

(c) [6pt] Find the MLE $\hat{\theta}_{MLE}$ for θ. Is it always well defined?
Solution:
For $x_1 = x_6 = 0$, the likelihood does not depend on θ, so that the MLE is not defined. For all the other values we have:

$$\partial_\theta \ell(\theta) = -\frac{X_1}{1/6 - \theta} + \frac{X_6}{1/6 + \theta}$$

so that $\hat{\theta}_{MLE} = \frac{X_6 - X_1}{6(X_1 + X_6)}$. Notice in fact that:

$$\partial^2_{\theta\theta} \ell(\theta) = -\frac{X_1}{(1/6 - \theta)^2} - \frac{X_6}{(1/6 + \theta)^2} < 0$$

(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.
Solution:
We write the generalized -likelihood ratio:

$$\Lambda(\mathbf{X}) = \frac{L(0)}{L(\hat{\theta}_{MLE})} = \exp - \Big(X_1 \log(1 - 6\hat{\theta}_{MLE}) + X_6 \log(1 + 6\hat{\theta}_{MLE}) \Big)$$

From the data:

$$\hat{\theta}_{MLE} = \frac{2}{32} = 1/16, \quad -2 \log \Lambda(\mathbf{x}) = 4.62$$

Since $-2 \log \Lambda(\mathbf{X}) \xrightarrow{d} \chi^2_1$, we have:

$$4.62 > \chi^2_1(0.05) = 3.84$$

so that we reject H_0 at 0.05 level of significance.