Technische Universiteit Delft | Mekelweg 4
Faculteit Elektrotechniek, Wiskunde en Informatica | 2628 CD Delft

Exam Introduction to Statistics (AM2080)
November 4, 2024, 13.30–16.30

This written endterm exam contains 5 questions, each question counts for 20% of the final grade of the test. You are only allowed to use a personally made cheat-sheet, the sheet with information on probability distributions, and the tables for normal, binomial, chi-square and Student-t distributions. You are not allowed to use any books or (other) notes.

1. Let $X_1, \ldots, X_n$ be independent random variables with density
$$p_\theta(x) = \theta x^{-(\theta+1)}, \quad x \geq 1,$$
where $\theta > 0$.
   (a) Determine the maximum likelihood estimator for θ.
   (b) Determine the Bayes estimator for θ based on the prior distribution with density $\pi(\theta) = e^{-\theta}$ for $\theta > 0$ and zero elsewehere.

2. Let $X_1, \ldots, X_n \sim \mathcal{N}(\mu, 1)$ and $Y_1, \ldots, Y_m \sim \mathcal{N}(\mu + \delta, 4)$ be independent random variables, with unknown parameters $\mu, \delta \in \mathbb{R}$. We want to test the hypothesis $H_0 : \delta = 0$ against $H_1 : \delta \neq 0$ via the statistic
$$T_{n,m} = \frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{m}\sum_{j=1}^{m} Y_j.$$
   (a) Show that $T_{n,m} \sim \mathcal{N}(0, \frac{m+4n}{nm})$ under H_0.
   (b) Describe a test procedure to test H_0 against H_1 based on the statistic $T_{n,m}$ by determining the rejection region for the significance level $\alpha = 0.1$.
   (c) Suppose that a test rejects H_0 if $T_{n,m} > 1$. Determine the power of this test as a function of n, m, and δ. For $n = m = 16$, compute the size of this test, as well as the power against the alternative $\delta = -1$.

3. Let $X_1, X_2, \ldots, X_n$ be independent random variables, all with distribution function
$$F_\theta(x) = 1 - e^{-\theta\sqrt{x}} \text{ for } x \geq 0$$
and 0 elsewhere. The parameter $\theta > 0$ is unknown.
   (a) Show that $\theta^2 X_{(1)} = \theta^2 \min_{1 \leq i \leq n} X_i$ is a pivot, and that its distribution function under the distribution with parameter θ is given by
$$P_\theta\left(\theta^2 X_{(1)} \leq t\right) = \begin{cases} 1 - e^{-n\sqrt{t}} & \text{for } t \geq 0 \\ 0 & \text{for } t < 0. \end{cases}$$
   (b) Using this pivot from part (a) and for $0 < \alpha < 1$, find $0 < c < d$ such that $\left[cX_{(1)}^{-1/2}, dX_{(1)}^{-1/2}\right]$ is a confidence interval for θ with confidence level $1 - \alpha$

4. Let $X_1, \ldots, X_n$ be independent random variables with density

$$p_\theta(x) = \theta x^{\theta-1}, \quad x \in (0,1),$$

where $\theta > 0$.

(a) Compute the method of moments estimator for θ based on the first moment.

The maximum likelihood estimator for θ is given by

$$\hat{\theta}_{\text{mle}} = -\frac{n}{\sum \log X_i}.$$

(b) Use this to show that the Likelihood Ratio test statistic for testing $H_0 : \theta = 3$ against alternative $H_1 : \theta \neq 3$ is given by

$$\lambda_n(X_1, \ldots, X_n) = \left(\frac{\hat{\theta}_{\text{mle}}}{3}\right)^n \left(\prod_{i=1}^n X_i\right)^{\hat{\theta}_{\text{mle}}-3}$$

In the book, a theorem is stated on the asymptotic distribution of $2 \log \lambda_n$ under the null hypothesis. You may assume the conditions of this theorem are satisfied in the current setting.

(c) Use this asymptotic distribution of $2 \log \lambda_n$ to derive an approximate critical region for λ_n at significance level $\alpha = 0.05$.

5. Consider the simple linear regression model,

$$Y_i = \alpha + \beta x_i + e_i, \quad 1 \leq i \leq n,$$

where $e_1, \ldots, e_n$ are independent random variables with a $\mathcal{N}(0, \sigma^2)$ distribution. The x_i's are nonrandom. It is given that the maximum likelihood estimators for β and α are given by

$$\hat{\beta} = \frac{\sum_{i=1}^n (x_i - \bar{x}_n)(Y_i - \bar{Y}_n)}{\sum_{i=1}^n (x_i - \bar{x}_n)^2} \text{ and } \hat{\alpha} = \bar{Y}_n - \hat{\beta}\bar{x}_n$$

where, as usual, $\bar{x}_n = \frac{1}{n}\sum_{i=1}^n x_i$ and $\bar{Y}_n = \frac{1}{n}\sum_{i=1}^n Y_i$

(a) Write the estimator for β as linear combination of the observed Y_i's, so as

$$\hat{\beta} = \sum_{i=1}^n \lambda_i Y_i$$

and specify λ_i for $1 \leq i \leq n$.

(b) Show that $\hat{\alpha}$ and $\hat{\beta}$ are unbiased estimators for α and β respectively.

(c) Show that the Mean Squared Error of the estimator $\hat{\beta}$ is given by

$$\text{MSE}(\hat{\beta}) = \frac{\sigma^2}{(n-1)s_x^2}$$

where $s_x^2 = \frac{1}{n-1}\sum_{i=1}^n (x_i - \bar{x}_n)^2$.