

Probability and Statistics

Y-DATA School of Data Science

P&P 5

Due: 07.12.2022

PROBLEM 1. Suppose that you received a free subscription to the lottery and you want to know whether you are in the regular track (probability 0.1 to win) or the premium track (probability 0.25 to win). To this end, you start counting the weeks in which you participated up to (and including) the first win. You set a decision rule: If you win before week 6, you will conclude that you are on the premium track.

- (1) Define an appropriate random variable X and determine its distribution.
- (2) Formulate the hypotheses H_0 and H_1 .
- (3) Write the rejection region of H_0 in terms of X .
- (4) Compute type I and type II errors.

PROBLEM 2. The advertisement of the fast food chain of restaurants “FastBurger” claims that the average waiting time for food in its branches is 30 seconds unlike the 50 seconds of their competitors. Mr. Skeptic does not believe much in advertising and decided to test its truth by the following test: he will go to one of the “FastBurger” branches, measure the waiting time, and if it is less than 40 seconds (the critical waiting time he fixed) he would believe in its advertisement. Otherwise, he will conclude that the service in “FastBurger” is no faster than in other fast food companies. Mr. Skeptic also assumes that waiting time is exponentially distributed.

- (1) What are the hypotheses Mr. Skeptic tests? Calculate the probabilities of errors of both types for his test.
- (2) Can you suggest a better test to Mr. Skeptic with the same significance level?

PROBLEM 3. Let X_1, \dots, X_n be an i.i.d. random sample from a distribution with the density function

$$f_{\theta}(x) = \frac{x}{\theta} e^{-\frac{x^2}{2\theta}}, x \geq 0, \theta > 0$$

Use the Neyman-Pearson lemma to find the most powerful test at level α for testing two simple hypotheses $H_0 : \theta = \theta_0$ vs. $H_1 : \theta = \theta_1$, where $\theta_1 > \theta_0$.

Hint: You may use without proof the fact that $\sum_{i=1}^n X_i^2 / \theta \sim \chi_{2n}^2$

PROBLEM 4. The lifetime of an automatic gear has normal distribution with known standard deviation of 30,000 km. The manufacturer claims that the expected lifetime is more than 120,000 km. To test the claim of the manufacturer, a sample of 15 cars was drawn. The average lifetime of the cars in the sample is 135,320 km.

- (1) Formulate the hypotheses H_0 and H_1 .
- (2) Would you reject the null hypothesis with significance level of 5%?
- (3) What is the minimal significance level for which you would reject the null?

PROBLEM 5. In the lecture, we saw that for $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} N(\theta, \sigma^2)$ (σ^2 known), the two-sided hypothesis test for

$$H_0 : \theta = \theta_0 \text{ vs. } H_0 : \theta \neq \theta_0$$

is given by the rejection region

$$R = \{|\bar{X}_n - \theta_0| \geq c\}$$

Show that for significance level α , $c = z_{1-\alpha/2} \sigma / \sqrt{n}$.

PROBLEM 6. This problem is a guided proof of the well-known one-sample t-test.

Let $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} N(\mu, \sigma^2)$ where σ^2 is unknown, and we want to test $H_0 : \mu = \mu_0$ against the two-sided alternative $\mu \neq \mu_0$. Denote for simplicity $\theta = (\mu, \sigma^2)$.

- (1) Write the likelihood function of X_1, \dots, X_n .
- (2) Plug-in the MLE estimators for μ and σ^2 in the likelihood function.
Explain why the obtained expression is $\sup_{\theta \in \Theta} L(\theta; X)$.
- (3) Plug-in μ_0 and $\hat{\sigma}_0^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu_0)^2$ in the likelihood function.
Explain why the obtained expression is $\sup_{\theta \in \Theta_0} L(\theta; X)$.
- (4) Show that the generalized likelihood ratio is

$$\lambda^*(X) = \left(1 + \frac{1}{n-1} \left(\frac{\bar{X}_n - \mu_0}{s/\sqrt{n}} \right)^2 \right)^{n/2}$$

where $s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \mu_0)^2$.

- (5) Define

$$T(X) = \frac{\bar{X}_n - \mu_0}{s/\sqrt{n}}$$

Explain why the rejection region $\{\lambda^*(X) \geq c\}$ is equivalent to the rejection region $\{|T(X)| \geq c^*\}$.

- (6) What is the distribution of $T(X)$?
Hint: The answer is in the slides of lecture 4.
- (7) Find the critical value c^* .