Probability and Statistics II
Final Exam

July 29, 2022

Course code: STA2002
Course name: Probability and Statistics II
Date: Monday, July 25, 2022
Time: 14:00-17:00 (three hours)

The questionnaire comes with five questions. You need to solve them all. You are entitled to bring along into the exam class a non-programmable calculator and two written double sided A4 sheets of paper prepared by yourself in advance. The total number of points is 100. The number of points each item carries is stated next to it. Note that in the case where a few items form a question, you are entitled to use an earlier item for proving a later one, even if you did not prove the former. Finally, you are asked to present a valid Student I.D. Card for inspection during the examination session.
Good luck!

1. (20) For a given series $X_1, \ldots, X_n$, both $\overline{X}$ and $\mathrm{Var}(X)$ were computed. A new observation X_{n+1} is added to the series. In terms only of $\overline{X}$, $\mathrm{Var}(X)$, n and X_{n+1} state a formula for

   (a) (4) the new arithmetic mean

   (b) (7) the new variance

   (c) (9) Assume now that $X_1 < X_2 < \cdots < X_n$ and that n is an odd number larger than 3. Suppose X_1 and X_n are removed from the series. For each of the three measures that are listed below, you need to select one out of the following four possibilities on the effect of this change.

   - no change
   - goes up
   - goes down
   - not enough information to tell

   Explain your answers. The measures are:

   i. (3) the arithmetic mean

   ii. (3) the median

   iii. (3) the variance

2. (20) A random sample of $X_1, \ldots, X_n$ is conducted where all $n \geq 2$ independent observations are identically distributed as a random variable X. Assume that n is an even number.

   (a) (4) Prove that

   $$\mathrm{E}\left(\frac{(X_2 - X_1)^2}{2}\right) = \mathrm{Var}(X).$$

   (b) (4) What is the distribution of $\frac{(X_2-X_1)^2}{2}$ in the case where $X \sim N(\mu, 1)$?

   (c) (12) Suppose now that $X \sim N(\mu, \sigma^2)$ and let $D_i = X_{2i} - X_{2i-1}$, $1 \leq i \leq n/2$.

   i. (4) Construct an UBE for σ^2 based only on D_i, $1 \leq i \leq n/2$.

   ii. (4) What is the MSE of the estimator you have just constructed? How is it compared with the MSE of the MLE for σ^2?

iii. (4) Based on the estimator you constructed in item (i), design a $1-\alpha$ confidence interval for σ^2 which is based on $X_1, \ldots, X_n$?

3. (20) Suppose $X \sim \chi^2_{(m)}$ for some even integer $m \geq 2$.

   (a) (3) Show how X can be sampled if your computer can draw series of independent standard normal random variables.

   (b) (2) Repeat the previous item but now where the series is of independent exponential random variables (the choice for the parameter is in your hands).

   (c) (2) Repeat the previous item but now where the series is of independent $[0, 1]$ uniform random variables.

   (d) (3) What is the density function of X, its expected value and its variance?

   (e) (10) Consider a random sample for n independent random variables, X_i, $1 \leq i \leq n$, all distributed as X.

      i. (3) What is the likelihood function of this sample?

      ii. (3) Suppose one wishes to test the null hypothesis that $m = m_0$ versus the alternative which says that $m = m_1$. Show that in order to execute the LRT all you need to know from the sample is its geometric mean.

      iii. (4) What are the LRTs for the case where $m_1 < m_0$?

4. (20) Let $X, Y \sim \exp(\lambda)$ be two independent random variables. Denote $\frac{X}{X+Y}$ by T.

   (a) (3) What is the support of the random variable T?

   (b) (7) Prove that $T \sim U[0, 1]$. Hint: Show that $P(T \geq t)$ for any t in the support equals $P(X \geq \frac{t}{1-t}Y)$. Then look for this probability for a specific value for $Y = y$ and conclude by using the complete probability law.

   (c) (5) Suppose a random sample of ratios of the type $T_i = \frac{X_i}{X_i+Y_i}$, $1 \leq i \leq n$, is conducted. How useful can these ratios be in estimating λ? Note: one can observe only the ratios and not the actual values of X_i and Y_i, $1 \leq i \leq n$.

(d) (5) Based on a random sample T_i, $1 \leq i \leq n$, suggest a goodness-of-fit test for testing the null hypothesis that says that the X and Y variables are independent and they follow an exponential distribution with a common parameter, against the alternative which says that this is not the case. Specifically, define ten possible cells (intervals) for the possible values for T which, under the null-hypothesis, are equally likely.

5. (20) Suppose m populations exist and we are interested in inspecting some numerical characteristics combining them. The model assumes that $X_{ij} \sim N(\mu_i, \sigma^2)$, $1 \leq j \leq n_i$, $1 \leq i \leq m$, for some parameters μ_i, $1 \leq i \leq m$ and $\sigma^2 > 0$, where all random variables involved are independent. Assume all sample sizes n_i, $1 \leq i \leq m$, are equal to some common value n. Suppose $m = 4$. Let $a = (1, -1/3, -1/3, -1/3)$ and $b = (0, 1, -1/2, -1/2)$.

   (a) (3) What are the MLEs for $\Sigma_{i=1}^4 a_i\mu_i$ and for $\Sigma_{i=1}^4 b_i\mu_i$?

   (b) (3) What are their MSE?

   (c) (3) Write a $1 - \alpha$ confidence level for $\Sigma_{i=1}^4 a_i\mu_i$. You can assume that the value of σ^2 is in your hands.

   (d) (4) Say in your own words what the null-hypothesis $\Sigma_{i=1}^4 a_i\mu_i = 0$ in fact says.

   (e) (3) Write an α test for the null-hypothesis $\Sigma_{i=1}^4 a_i\mu_i = 0$ against the alternative the this is not the case. You can assume that the value for σ^2 is in your hands.

   (f) (4) What is the covariance between the two estimators you stated in item (a)? Prove your claim.