Tufts University
Department of Mathematics
Spring 2022

MA 166: Statistics

Midterm exam (v1.0) [1]

16 March 2022

PLEASE READ THESE INSTRUCTIONS BEFORE ANYTHING ELSE.

This exam is closed book and closed notes. While you are taking this exam, you may not communicate or otherwise exchange information regarding this exam or related content, with any human, either in person or by electronic means, and either to give or to receive help. The work you present must be your own, and yours only. Violations of the letter or even the spirit of this rule would be considered an extremely serious breach of ethics, honor, and conduct, and I would be obliged by Tufts University to report even so much as any suspicion I might have of such a violation to the Office of the Dean of Students.

Please sign and print your name below to indicate that you are aware of the above instructions, and that you will comply with them while you are taking this exam. You must turn in this entire exam booklet, signed where indicated, as the first four pages of your completed final examination.

Name: ____________________ Signature: ____________________

YOU MAY NOW TURN THE PAGE.

POTENTIALLY USEFUL INFORMATION

You may use the following information without proof or justification.

- You can use any of the quantities mentioned below without explaining what they are, whenever you need them.
  - * $f_Z(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$ (Normal distribution)
  - * $\int_{-\infty}^{z_{1-\alpha}} dx\ f_Z(x) = \int_{z_\alpha}^{+\infty} dx\ f_Z(x) = \alpha$
  - * $f_{\chi_n^2}(x) = \frac{1}{2^{n/2}\Gamma\left(\frac{n}{2}\right)} x^{(m/2)-1} e^{-x/2}$ (Chi squared distribution)
  - * $\int_0^{\chi^2_{1-\alpha,n}} dx\ f_{\chi_n^2}(x) = \int_{\chi^2_{\alpha,n}}^{+\infty} dx\ f_{\chi_n^2}(x) = \alpha$
  - * $f_{F_{m,n}}(x) = \frac{\Gamma\left(\frac{m+n}{2}\right) m^{m/2} n^{n/2} x^{(m/2)-1}}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)(n+mx)^{(m+n)/2}}$ (Fisher distribution)
  - * $\int_0^{f_{1-\alpha,m,n}} dx\ f_{F_{m,n}}(x) = \int_{f_{\alpha,m,n}}^{+\infty} dx\ f_{F_{m,n}}(x) = \alpha$
  - * $f_{T_n}(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-(n+1)/2}$ (Student T distribution)
  - * $\int_{-\infty}^{t_{1-\alpha,n}} dx\ f_{T_n}(x) = \int_{t_{\alpha,n}}^{+\infty} dx\ f_{T_n}(x) = \alpha$

- For data $X_1 = x_1, \ldots, X_n = x_n$ sampled from a pdf $f_X(x;\theta)$, the *likelihood function* is

$$L(\theta) = \prod_{j=1}^{n} f_X(x_j;\theta).$$

  This expression also works for discrete r.v.s if we replace $f_X(x_j;\theta)$ by $p_X(k_j;\theta)$.

- An *estimator* $\hat{\theta}$, applied to data $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$, results in an *estimate* $\theta_e = \hat{\theta}(x_1, \ldots, x_n)$.

- An *estimator* $\hat{\theta}$ is defined to be *sufficient* if the likelihood function obeys the *first factorization criterion for sufficiency*,

$$L(\theta) = f_{\hat{\theta}}(\theta_e)\ b(x_1, \ldots, x_n).$$

  for some function b. In other words, the likelihood function factors into the pdf for the estimator, times a function of the data alone.

- We proved that the *second factorization criterion for sufficiency*,

$$L(\theta) = g\left[\hat{\theta}(x_1, \ldots, x_n)\,;\,\theta\right]\ b(x_1, \ldots, x_n),$$

  is completely equivalent to the first factorization criterion for sufficiency. It states that the likelihood function factors into some function g of the estimator and the parameter, times another function b of the data alone. This is clearly a weaker requirement for sufficiency than that of the first factorization criterion, and hence often more useful.

THE EXAM QUESTIONS

Point values for each problem are given. You must show all your work and justify all your reasoning in order to receive credit, and also to receive partial credit. You do not have to provide formal two-column proofs for your arguments, but you do have to present them in such a way that a mathematically literate reader will understand and be convinced by your reasoning.

1. (25 points) Let $X_1 = k_1, X_2 = k_2, \ldots, X_n = k_n$ be n independent random samples from the geometric distribution $p_X(k;\theta) = (1-\theta)\theta^k$ where $k \in \{0, 1, \ldots, \infty\}$. Find the maximum likelihood estimator $\hat{\theta}$ for the parameter θ.

2. (40 points) Let $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$ be n independent random samples from a pdf of the general form

$$f_X(x;\theta) = \exp\left[K(x)p(\theta) + S(x) + q(\theta)\right],$$

where θ is a parameter. You may also assume that the support of f does not depend on θ.

   (a) (25 points) Show that the estimator

$$\hat{\theta}(\vec{x}) := \sum_{j=1}^{n} K(x_j),$$

   which may or may not be unbiased for θ, is a sufficient estimator.

   (b) (15 points) Show that the exponential distribution

$$f_X(x;\theta) = \begin{cases} \theta e^{-\theta x} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

   is of the general form described in this problem, and identify the corresponding sufficient estimator.

3. (35 points) You have reason to believe that X is a normal random variable, but you do not know either its mean or its variance, and you are able to take only three samples of X in your laboratory. Those samples turned out to be $X_1 = x_1 = 1$, $X_2 = x_2 = 2$ and $X_3 = x_3 = 3$.

   (a) (7 points) Find the sample mean, $\overline{x}$.

   (b) (7 points) Find the sample variance, s^2.

   (c) (7 points) Find the 90% confidence interval for the actual mean μ, based on the three measurements that you took. You may express your answer in terms of quantities given in the "Potentially Useful Information" section, if you wish.

   (d) (7 points) Suppose that your purpose in taking your three samples was to test $H_0 : \mu = 1$ against $H_1 : \mu > 1$ at the $100(1-\alpha)\%$ level of confidence. Give a criterion involving α for accepting or rejecting H_0.

   (e) (7 points) Repeat part (d) with the null hypothesis $H_0 : \mu = 2$, and the alternative hypothesis $H_1 : \mu \neq 2$. This time, you should be able to give and justify a definite answer as to whether or not H_0 should be rejected, given your three data points.