STAT 135 Concepts of Statistics **Name:** _____

Summer 2021 Lec 001

Practice Midterm

In class, 07/15/2021

Time Limit: 120 Minutes

Instructions: This exam is open book. It contains 3 pages (including this cover page) and 5 questions. Total of points is 71.

1. Start with the problems that you find the easiest. If you are stuck, move on to the next question.
2. Write clearly and show you work. Ambiguous or insufficient reasoning will be marked incorrect.
3. The bCourses page will be closed at 11:02 AM. Make sure you leave ample time to scan and submit you answers.
4. The abbreviations used in this exam include: probability density function (**pdf**), cumulative distribution function (**cdf**), Method of Moments (**MM**) estimator, Maximum likelihood estimator (**MLE**), Confidence interval (**CI**), Likelihood ratio test (**LRT**)

---

1. (34 points) Let $X_1, \ldots, X_n$ be i.i.d from a population with pdf

$$f(x|\theta) = \begin{cases} |x|/\theta^2, & \text{if } -\theta \le x \le \theta \\ 0, & \text{otherwise.} \end{cases}$$

in which $\theta > 0$ is an unknown parameter.

   (a) (3 points) Find a MM estimator of θ;

   (b) (4 points) Find the MLE of θ. Is it a sufficient statistic for θ? Explain why;

   (c) (5 points) *(Bonus question) Denote the MLE by $\hat{\theta}_{\text{MLE}}$. Prove that $\hat{\theta}_{\text{MLE}}$ can only take values in $[0, \theta]$ and its the cdf satisfies

$$P(\hat{\theta}_{\text{MLE}} \le t) = \left(\frac{t}{\theta}\right)^{2n}, \quad t \in [0, \theta].$$

   (d) (8 points) Show that $Y = \frac{2n+1}{2n}\hat{\theta}_{\text{MLE}}$ is an unbiased estimator of θ. Calculate $\text{Var}(Y)$.

   (e) (8 points) Use the cdf in (c) to explain why $\hat{\theta}_{\text{MLE}}$ can not be approximated by a Normal distribution as $n \to \infty$. Is it contradictory to Theorem D of Lecture 4?

(f) (6 points) One can estimate $\text{Var}(\hat{\theta}_{\text{MM}}) \approx \frac{\theta^2}{12n}$ for sufficiently large n. With this result, we can calculate the relative efficiency

$$\text{eff}(Y, \hat{\theta}_{\text{MM}}) = \frac{\text{Var}(Y)}{\text{Var}(\hat{\theta}_{\text{MM}})}.$$

Show that $\text{eff}(Y, \hat{\theta}_{\text{MM}}) \to 0$ for large n, which shows the unbiased estimator Y is a much more efficient estimator than $\hat{\theta}_{\text{MM}}$.

2. (12 points) Suppose $X_1, \ldots, X_n$ are independently sampled from a Kumaraswamy distribution with parameter β. The probability density function of this distribution is

$$f(x|\beta) = \begin{cases} \beta(1-x)^{\beta-1}, & \text{if } x \in (0,1), \\ 0, & \text{otherwise.} \end{cases}$$

(a) (6 points) Find the MM estimator $\hat{\beta}_{MM}$ for β;
(*Hint:* $\int_0^1 x^{\alpha-1}(1-x)^{\beta-1} = \text{Beta}(\alpha, \beta)$.)

(b) (6 points) Suppose a researcher collected 100 observations and the sample mean is 0.45. Calculate the MM estimate for β, and use the Delta Method to approximate the SE of your estimate $\hat{\beta}_{MM}$.

3. (16 points) Suppose $X_1, \ldots, X_n$ are i.i.d observations from a population with pmf

$$P(X = x|\theta) = \theta^x(1-\theta)^{1-x}, \quad , x = 0 \text{ or } 1, \quad 0 \le \theta \le \frac{1}{2}$$

.

(a) (6 points) Find the MM estimator and MLE of θ;

(b) (4 points) Try to find the MSE of each of the estimators (*See Page 29 of Lecture 1 slides for the definition of MSE*);

(c) (6 points) Which estimator is preferred? Justify your choice.

4. (25 points) Let $X_1, \ldots, X_n$ be i.i.d with beta$(\mu, 1)$ pdf and $Y_1, \ldots, Y_m$ be i.i.d with beta$(\theta, 1)$. Also assume that X's and Y's are independent of each other.

(a) (8 points) Find an LRT of $H_0 : \theta = \mu$ versus $H_1 : \theta \neq \mu$, and show that the LRT statistic is

$$T(\mathbf{X}_n, \mathbf{Y}_m) = \frac{\sum_{i=1}^n \log X_i}{\sum_{i=1}^n \log X_i + \sum_{i=1}^m \log Y_i}.$$

(b) (5 points) Find the distribution of T when H_0 is true, and show how to get a test of size $\alpha = 0.05$.

(*Hint*: $-\log X_i \sim \text{Gamma}(1, 1/\mu)$ and $-\log Y_i \sim \text{Gamma}(1, 1/\theta)$. Also, $W/(W+V) \sim \text{Beta}(n, m)$ if $W \sim \text{Gamma}(n, 1/\mu)$, $V \sim \text{Gamma}(m, 1/\mu)$ and W is independent of V.)

(c) (4 points) Determine the threshold in the LRT rejection region so that the significance level $\alpha = 0.05$. You might need `R` to help compute the upper α quantile of a beta(a, b) distribution: `qbeta(alpha, a, b, lower.tail=FALSE)`

(d) (8 points) Suppose $n = 13$ and $m = 17$. The samples collected are

$$\begin{aligned}\mathbf{X}_{13} = \{&0.610,\ 0.344,\ 0.289,\ 0.700,\ 0.710,\ 0.266,\ 0.244,\ 0.919,\\ &0.022,\ 0.006,\ 0.073,\ 0.849,\ 0.773\},\\ \mathbf{Y}_{17} = \{&0.781,\ 0.479,\ 0.821,\ 0.766,\ 0.444,\ 0.443,\ 0.290,\ 0.862,\\ &0.684,\ 0.151,\ 0.931,\ 0.753,\ 0.694,\ 0.121,\ 0.264,\ 0.731,\ 0.575\}.\end{aligned}$$

Will you reject the null hypothesis? Report the p-value. You might need `R` to help compute the area under the curve of a beta(a, b) density: `pbeta(alpha, a, b, lower.tail=TRUE)`

5. (18 points) In a study of the effect of cigarette smoking on the carbon monoxide diffusing capacity (DL) of the lung, researchers found that current smokers had DL readings significantly lower than those of either ex-smokers or non-smokers. The carbon monoxide diffusing capacities for a random sample of n = 20 current smokers are listed here:

$$\begin{aligned}\{&103.768,\ 92.295,\ 100.615,\ 102.754,\ 88.602,\\ &61.675,\ 88.017,\ 108.579,\ 73.003,\ 90.677,\\ &71.210,\ 73.154,\ 123.086,\ 84.023,\ 82.115,\\ &106.755,\ 91.052,\ 76.014,\ 89.222,\ 90.479\}\end{aligned}$$

(a) (6 points) Compute the sample mean and sample standard deviation of the above data;

(b) (6 points) Do these data indicate that the mean DL reading for current smokers is significantly lower than 100, which is the average for nonsmokers? Use a one-sided hypothesis test, with $\alpha = 0.01$. Since $n < 30$, you will need to use exact Student's t distribution to find the rejection regions of the test.

(c) (6 points) Calculate the 95% exact CI for population variance σ^2. Use this result and the duality between the CIs and hypothesis tests to make a decision about

$$H_0 : \sigma = 14.9 \text{ versus } H_1 : \sigma \neq 14.9.$$