STAT 135 CONCEPTS OF STATISTICS
HOMEWORK 3

Assigned July 8, 2021, due July 15, 2021

This homework pertains to materials covered in Lecture 6, 7 and 8. The assignment can be typed or handwritten, with your name on the document, and **with properly labeled input code and computer output for those problems that require it**. To obtain full credit, please write clearly and show your reasoning. If you choose to collaborate, the write-up should be your own. Please show your work! Upload the file to the Week 3 Assignment on bCourses.

Note in this homework, we use the following abbreviations: Cramer-Rao bound (CR bound), Uniformly minimum-variance unbiased estimator (UMVUE), probability density/mass function (pdf/pmf), Maximum likelihood estimators (MLE) and likelihood ratio test (LRT).

Problem 1. Consider the i.i.d random variables $X_1, \dots, X_n$ from Poission(λ).

(1) Show that both sample mean $\bar{X}_n$ and sample variance $S^2 = (n-1)^{-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$ are both unbiased estimators of λ.
(2) Calculate the CR lower bound for any unbiased estimator of λ.
(3) Is either $\bar{X}_n$ or S^2 a UMVUE of λ?
(*Hint*: For i.i.d random variables, $\mathrm{Var}(S^2) = \frac{E(X-\mu)^4}{n} - \frac{\sigma^4(n-3)}{n(n-1)}$. For Poission($\lambda$), the 4th central moment is $E(X-\lambda)^4 = \lambda(1+3\lambda)$.)

Problem 2. Consider the i.i.d random variables $X_1, \dots, X_n$ from some population $f(x|\theta)$ with the true parameter being θ_0, which is sufficiently smooth to satisfy

$$\int_{\mathcal{X}} \frac{\partial}{\partial\theta} f(x|\theta)dx = \frac{\partial}{\partial\theta}\int_{\mathcal{X}} f(x|\theta)dx,$$
$$\int_{\mathcal{X}} \frac{\partial^2}{\partial^2\theta} f(x|\theta)dx = \frac{\partial^2}{\partial\theta^2}\int_{\mathcal{X}} f(x|\theta)dx,$$

in which $\mathcal{X}$ is the set of all possible values of the population.

(1) Suppose $X \sim f(x|\theta_0)$. Show that

$$E_{\theta_0}\left[\frac{\partial}{\partial\theta}\log f(X|\theta_0)\right] = 0.$$

(2) Denote $f(x_1, \dots, x_n|\theta) = \prod_{i=1}^n f(x_i|\theta)$ as the joint density. Prove

$$E_{\theta_0}\left[\frac{\partial}{\partial\theta}\log f(X_1, \dots, X_n|\theta_0)\right] = 0.$$

(3) Prove

$$E_{\theta_0}\left[\frac{\partial}{\partial\theta}\log f(X_1,\ldots,X_n|\theta_0)\right]^2 = nI(\theta_0).$$

Problem 3. Suppose $X_1,\ldots,X_n$ are independently sampled from a distribution with pdf

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

(1) Find the MLE, $\hat{\theta}_{MLE}$, of θ. Is it a sufficient statistic for θ?
(2) Calculate the Fisher information $I(\theta)$.
(3) Can we apply Theorem D of Lecture 5 to obtain the asymptotic normality $\sqrt{n}(\hat{\theta}_{MLE}-\theta) \xrightarrow{d} N(0,1/I(\theta))$? Explain why.
(4) If you answered no to the previous question, derive the exact sampling distribution of $\hat{\theta}_{MLE}$ and see why it can not be approximated by a Normal distribution.
(*Hint*: The density of the sample maximum $X_{(n)}$ is $f_{X_{(n)}}(x) = nf(x)F^{n-1}(x)$.)
(5) Show that $Y = \frac{n+1}{n}X_{(n)}$ is an unbiased estimator of θ. Calculate $\mathrm{Var}(Y)$, and show that it is smaller than the CR bound $\frac{1}{nI(\theta)}$. Is this result contradictory to the CR inequality in Theorem E?

Problem 4. Let's say we observed 14 i.i.d $N(\mu,\sigma^2)$ samples with both μ and σ^2 being unknown:

$$\{-1.398,\ 8.061,\ 13.609,\ 4.325,\ 12.140,\ -4.611,\ 6.669,$$
$$4.340,\ 1.776,\ 7.355,\ -3.100,\ -3.784,\ 9.962,\ -4.150\}.$$

(1) Find 99% exact confidence intervals for μ and σ^2.
(2) Find 99% bootstrap confidence intervals for μ and σ^2.
(3) Compare the two sets of CIs. Which one has more accuracy?
(4) How much larger a sample do you think you would need to halve the length of the confidence interval for μ?

Problem 5. Let $X_1,\ldots,X_n$ be i.i.d samples from a Bernoulli(p) population. It is desired to test

$$H_0: p = 0.59, \qquad \text{versus} \qquad H_1: p = 0.45.$$

(1) Derive the rejection region of the LRT.
(2) Use the central limit theorem to determine, approximately, the sample size needed so that type I error and type II error are both at most 0.01.

Problem 6. Suppose $X_1,\ldots,X_n$ are independently sampled from Pareto(θ,ν) with pdf

$$f(x|\theta) = \begin{cases} \frac{\theta\nu^\theta}{x^{\theta+1}}, & \text{if } x \geq \nu, \\ 0, & \text{otherwise.} \end{cases}$$

in which $\theta > 0$ and $\nu > 0$.

(1) Find the MLEs of θ and ν.

(2) Show that the LRT of

$$H_0 : \theta = 1,\ \nu \text{ unknown}, \qquad \text{versus} \qquad H_1 : \theta \neq 1,\ \nu \text{ unknown}$$

has rejection region of the form $\{T(\mathbf{X}_n) \leq c_1 \text{ or } T(\mathbf{X}_n) \geq c_2\}$ where $0 < c_1 < c_2$ and

$$T(\mathbf{X}_n) = \log \left[\frac{\prod_{i=1}^n X_i}{X_{(1)}^n} \right],$$

in which $X_{(1)}$ is the sample minimum.

Problem 7. 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.

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

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

(2) 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.)

(3) 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.

Problem 8. 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:

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

(1) Compute the sample mean and sample standard deviation of the above data.

(2) 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.

Problem 9. Suppose X is a random variable whose pmf under H_0 and H_1 is given by

x	1	2	3	4	5	6	7
$f(x\|H_0)$	0.01	0.01	0.01	0.01	0.02	0.01	0.93
$f(x\|H_1)$	0.06	0.05	0.04	0.03	0.02	0.01	0.79

Use the Neyman-Pearson Lemma to find the most powerful test for H_0 versus H_1 with $\alpha = 0.04$. Compute the probability of Type II error for this test.

Problem 10. Let X be a single observation from $f(x|\theta) = \theta x^{\theta-1}$, $0 < x < 1$.

(1) Find the most powerful test using significance level $\alpha = 0.05$ for testing the hypotheses $H_0 : \theta = 1$ versus $H_1 : \theta = 2$ (sketch the densities $f(x|H_0)$ and $f(x|H_1)$ for the two hypotheses).
(2) What is the power of the test?
(3) What is the p-value of $X = .8$?
(4) For fixed $\alpha = 0.05$, is the test uniformly most powerful against the alternative hypothesis $H_1 : \alpha > 1$?