Probability and Statistical Theory

MS Comprehensive Examination

April 13, 2019

Instructions:

Show all of your computations.
Prove all of your assertions or quote the appropriate theorems.
Books, notes, and calculators *may be used.*

You have three hours.

1. (20 points, 5 points each) The joint distribution of (X_1, X_2, X_3, X_4) is multinomial$(n, \boldsymbol{\pi})$ with $\boldsymbol{\pi} = (\pi_1, \cdots, \pi_4)$ and

$$P(X_1 = x_1, X_2 = x_2, X_3 = x_3, X_4 = x_4) = \frac{n!}{x_1!x_2!x_3!x_4!}\pi_1^{x_1}\pi_2^{x_2}\pi_3^{x_3}\pi_4^{x_4},$$

where $\sum_{i=1}^{4} \pi_i = 1$ and $\sum_{i=1}^{4} x_i = n$.

a. What is the marginal distribution of X_i $(i = 1, \cdots, 4)$? Find $\mathsf{E}(X_i)$ and $\mathsf{Var}(X_i)$ $(i = 1, \cdots, 4)$. (You could use the marginal distribution)

b. Show that $\mathsf{Cov}(X_i, X_j) = -n\pi_i\pi_j$ $(i \neq j)$.

c. Find $\mathsf{Var}(X_1 + X_2 + X_3)$ using the results above.

d. Under the assumption that $\pi_1 = \pi_2 = \pi_3 = \pi$, find the maximum likelihood estimator (MLE) $\hat{\pi}_{\text{MLE}}$ for π and the MLE $\hat{\theta}_{\text{MLE}}$ for the odds $\theta = \frac{\pi}{1-\pi}$.

2. (30 points, 5 points each) Suppose $\boldsymbol{X} = \{X_1 \cdots, X_n\}$ is a simple random sample (i.e. X_i's are iid) from

$$f(x|\theta) = \theta e^{-\theta x}, x \geq 0, \theta > 0.$$

a. Show that $T = \sum_{i=1}^{n} X_i$ is a complete sufficient statistic for θ.

b. Find $\mathsf{E}\left(\frac{X_n}{T}\right)$ using Basu's Theorem.

c. Find the uniformly minimum variance unbiased estimator (UMVUE) $\tilde{\theta}$ for θ.

d. Calculate the variance of $\tilde{\theta}$.

e. Does $\tilde{\theta}$ attain the C-R Lower Bound?

f. Find a uniformly most powerful (UMP) level α test $\phi(\boldsymbol{x})$ for $H_0 : \theta = \theta_0$ vs $H_1 : \theta > \theta_0$.

3. (20 points)

Consider a situation where $X \sim B(n=6, p)$ and we wish to test H_0: $p = 0.5$ versus H_A: $p \neq 0.5$ at level of significance $\alpha \leq .05$.

a. Give at least one reason why the set {0,6} is a reasonable critical region for this test.
b. If that is the critical region, what is the actual, exact, level of significance α?
c. Argue whether or not the normal approximation should be valid for the calculation in part b.
d. Whether you argue that it is valid or not, use the normal approximation to approximate α. Use the half-integer correction for continuity. Comment on the accuracy of this approximation.
e. Define the power function. In this problem, what is its domain? In this problem, what is its range?
f. Sketch the power function by calculating the power at each point in the set {k/10 : k=0,1,2,...,9,10}. Hint: there is a symmetry argument, plus some easy cases, so that there are only 4 (out of a possible 11) calculations to perform. Calculate exact probabilities.
g. Use the normal approximation to approximate Power(0.1). Argue whether this should, or should not, be valid. Use the half-integer correction for continuity.

4. (30 points)
Consider a continuous random variable X with density function $f(x) = kx$ for $0 \le x \le b$, and 0 elsewhere. Regard b as the unknown parameter of interest.

a. Show that $k = 2 / b^2$.
b. Find $\mu = E(X)$.
c. Show that the CDF (Cumulative Distribution Function) F(x) for this distribution is
$$F(x) = \begin{cases} 0 \ if \ x < 0 \\ x^2/b^2 \quad if \ 0 \le x \le b \\ 1 \ if \ x > b \end{cases}.$$
d. Find M = Median of this distribution.

For the remaining parts, let $X_1,\ldots,X_n$ denote a random sample from this distribution. And here is some notation: $x_{(n)} = \max\{x_i : 1 \le i \le n\}$ and $1_A(y) = 1$ if $y \in A$ and 0 if $y \in A^c$. This is like an indicator variable where the argument is not random. Also bold face **x** indicates the n-dimensional vector of the data.

e. Find the Method of Moments estimator for b.
f. Show that the Likelihood function can be written $L(b;\mathbf{x}) = 2^n/b^{2n} \left(\prod_{i=1}^{n} x_i\right) 1_{[0,b]}(x_{(n)})$.
g. Show that the MLE (Maximum Likelihood Estimator) for b is $\hat{b}_{ML} = x_{(n)}$.
h. Make a case that the maximum $X_{(n)}$ is a sufficient statistic for the parameter b.
i. Show that the density function of the maximum $X_{(n)}$ is $f_{(n)}(x) = \frac{2n}{b^{2n}} x^{2n-1}$ for $0 \le x \le b$.
j. Show that $X_{(n)}$ is biased, calculate the bias, and show what happens as n tends to infinity.

Now we will use this data to find the likelihood ratio test for H_0: b=1 versus H_A: $b \neq 1$ at level of significance $\alpha = .10$.

k. Let $\widehat{b_{ML,0}}$ denote the MLE under H_0. Find this MLE. Hint: This *is* as easy as it seems.
l. Recall that the likelihood ratio test statistic is $\lambda(\mathbf{x}) = L(\widehat{b_{ML,0}};\mathbf{x}) / L(\widehat{b_{ML}};\mathbf{x})$, i.e. the ratio of the maximum likelihoods – under H_0 for the numerator and over the entire parameter space for the denominator. Show that in this problem, $\lambda(\mathbf{x}) = x_{(n)}^{2n} 1_{[0,1]}(x_{(n)})$. Sketch $\lambda(\mathbf{x})$ as a function of $x_{(n)}$. Note that $\lambda(\mathbf{x})$ depends on the data only through the sufficient statistic $X_{(n)}$.
m. Recall that for the likelihood ratio test we reject H_0 if , $\lambda(\mathbf{x}) \le c$ for the constant c such that $P_{H_0}(\lambda(\mathbf{x}) \le c) = \alpha$. Find the critical region where we reject H_0 in terms of the sufficient statistic $X_{(n)}$ as a function of α and n.
n. Finally for $\alpha = .10$ as given, and for n=10, find the critical region and make a decision regarding H_0 if the maximum $x_{(n)}$ is 0.8. Repeat this if the maximum $x_{(n)}$ is 1.2.