Master of Applied Statistics

Theoretical Statistics Comprehensive Exam

May 2018

Directions: This is a closed book exam with a 4-hour time limit. Attached you will find four pages of formulas and tables for the normal, \( t \), chi-square, and \( F \) distributions. You may use a non-programmable, non-graphing calculator.

Answer Only Five of the Six Questions.

1. A transmitter is sending a message in binary code (“+” and “−” signals) that must pass through two independent relay stations before being sent on to the receiver. Schematically, the message is sent as follows

\[
\text{Transmitter} \implies \text{Relay 1} \implies \text{Relay 2} \implies \text{Receiver.}
\]

At each relay station, there is a 25 percent chance that a signal will be reversed; that is, for \( i = 1, 2, \)

\[
P(\text{“+” is sent by relay } i | \text{“−” is received by relay } i) = 0.25
\]

\[
P(\text{“−” is sent by relay } i | \text{“+” is received by relay } i) = 0.25.
\]

Suppose that “+” symbols make up 60 percent of the messages being sent by the transmitter.

(a) What is the probability a “+” is received from Relay 2?

(b) If a “+” is received from Relay 2, what is the probability that a “+” was sent by the transmitter?
2. The length of time to failure (in 100s of hours) for a transistor is a random variable with the following cumulative distribution function (cdf):

\[ F_X(x) = \begin{cases} 0, & x \leq 0 \\ 1 - \exp(-x^2), & x > 0. \end{cases} \]

(a) Find \( f_X(x) \), the probability density function (pdf) of \( X \), and graph it.

(b) Calculate the mean time to failure.

(c) Calculate the median time to failure.

(d) Calculate \( P(X > r + s | X > r) \), for \( r, s > 0 \).

3. A committee of 2 people is to be randomly selected from group containing 3 Republicans, 2 Democrats, and 1 Libertarian. Let \( Y_1 \) denote the number of Republicans on the committee and let \( Y_2 \) denote the number of Democrats on the committee.

(a) Find the joint probability mass function (pmf) of \( Y_1 \) and \( Y_2 \) and depict it in a two-way table. Make sure your joint pmf (and marginal pmfs) is (are) valid.

(b) Calculate the joint moment-generating function \( M_{Y_1,Y_2}(t_1,t_2) = E(e^{t_1Y_1+t_2Y_2}) \) and the marginal moment-generating functions of \( Y_1 \) and \( Y_2 \).

(c) Calculate the correlation of \( Y_1 \) and \( Y_2 \).
4. Monte Carlo methods can be used to evaluate an integral. Suppose that we wish to evaluate

\[ I(f) = \int_a^b f(x) \, dx. \]

Let \( g \) be a density function on \([a, b]\). If we can generate a large i.i.d. sample \( X_1, \ldots, X_n \) from \( g \), then we can approximate \( I(f) \) by

\[ \hat{I}(f) = \frac{1}{n} \sum_{i=1}^{n} \frac{f(X_i)}{g(X_i)}. \]

(a) Find \( E\left( \frac{f(X)}{g(X)} \right) \) where \( X \) has density function \( g \).

(b) Show that \( E(\hat{I}(f)) = I(f) \).

(c) Find an expression for \( \text{Var}(\hat{I}(f)) \) in terms of \( E\left( \frac{f(X)}{g(X)} \right) \) and/or \( \text{Var}\left( \frac{f(X)}{g(X)} \right) \).

(d) Propose a suitable \( g \) and describe a Monte Carlo method to evaluate

\[ I(f) = \frac{1}{\sqrt{2\pi}} \int_0^1 e^{-x^2/2} \, dx. \]

5. The Pareto distribution has been used in economics as a model for a density function with a slowly decaying tail:

\[ f(x|x_0, \theta) = \theta x_0^\theta x^{-\theta - 1}, \quad x > x_0, \quad \theta > 1. \]

Assume that \( x_0(>0) \) is known and that \( X_1, \ldots, X_n \) is an i.i.d. sample.

(a) Find the method of moments estimate of \( \theta \).

(b) Find the maximum likelihood estimate (mle) of \( \theta \).

(c) Find a sufficient statistic for \( \theta \).

(d) Find the asymptotic variance of the mle.
6. A survey of voter sentiment was conducted in four midcity political wards to compare the fraction of voters favoring candidate A. Random samples of 200 voters polled in each of the four wards, with the results as shown in the table below. The numbers of voters favoring A in the four samples can be regarded as four independent binomial random variables. Construct a likelihood ratio test of the hypothesis that the fractions of voters favoring candidate A are the same in all four wards.

<table>
<thead>
<tr>
<th>Opinion</th>
<th>Ward</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Favor A</td>
<td></td>
<td>76</td>
<td>53</td>
<td>59</td>
<td>48</td>
<td>236</td>
</tr>
<tr>
<td>Do not favor A</td>
<td></td>
<td>124</td>
<td>147</td>
<td>141</td>
<td>152</td>
<td>564</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>200</td>
<td>200</td>
<td>200</td>
<td>200</td>
<td>800</td>
</tr>
</tbody>
</table>

(a) Let \( p_i \) denote the proportion of voters who favor candidate A of ward \( i \), for \( i = 1, 2, 3, \) and 4. Find the maximum likelihood for this data under \( H_0 : p_1 = p_2 = p_3 = p_4 \).

(b) Find the maximum likelihood for this data under the whole parameter space \( \Omega = \{ 0 \leq p_1 \leq 1, 0 \leq p_2 \leq 1, 0 \leq p_3 \leq 1, 0 \leq p_4 \leq 1 \} \). (Hint: Find the maximum likelihood for each of the four samples, and then multiply them together.)

(c) Calculate the likelihood ratio and conduct the approximate \( \chi^2 \) test. Use \( \alpha = .05 \).