Probability and Statistical Theory

MS Comprehensive Examination

April 15, 2006

Instructions:

Please answer all three questions.
Point Values: 50, 30, 20

Record your answers in your blue books.

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. Suppose X_i's $(1 \le i \le n)$ are iid $N(\mu, \sigma^2)$ with

$$f(x|\sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

a. Let $Y_1 = X_1 + X_2$ and $Y_2 = X_1 - X_2$. Find $\mathrm{Cov}(Y_1, Y_2)$ and $\mathrm{P}(Y_1 < Y_2)$.

b. Let $\bar{X}$ be the sample mean. Find $\mathrm{E}|X_1 - \bar{X}|$.

Suppose BOTH μ and σ^2 UNKNOWN. Let $\boldsymbol{\theta}' = (\mu, \sigma^2)$. Answer the questions c-f.

c. Find a complete and sufficient statistic $T_0(\vec{X})$ for $\boldsymbol{\theta}$.

d. Show that

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \text{ and } S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

are independent.

e. Find the MLE $\hat{\boldsymbol{\theta}}$.

f. Consider the LRT for $H_0 : \mu \le 0$ vs. $H_1 : \mu > 0$. Show that the rejection region for the LRT test is

$$R = \left\{\vec{x} : \frac{\sqrt{n}\bar{X}}{S} > t_\alpha(n-1)\right\},$$

where $\mathrm{P}(T > t_\alpha(n-1)) = \alpha$ (T is student t distributed with degrees of freedom $n-1$).

2. Let $X_1, \ldots, X_n$ denote a random sample of size n from the distribution over the interval $[a, \infty)$ with density function $f(x) = c \times \exp[(a-x)/\beta]$ for $x \geq a$ with parameters $a \geq 0$ and $\beta > 0$, and c is a fixed constant to be determined.

a. Find c.
b. Find the mean μ and the variance σ^2. Since this is not a test on integration, you may use the facts that $\int_0^\infty y e^{-y} dy = 1$ and $\int_0^\infty y^2 e^{-y} dy = 2$.
c. Find the method of moments estimators of a and β.
d. Let θ denote the vector (a,β) of unknown parameters. Find a sufficient statistic S for θ.
e. Find the maximum likelihood estimator of θ.
f. Find the likelihood ratio test statistic for testing H_0: a=0 versus the alternative H_A: a>0. Is this a function of the sufficient statistic S?
g. Using the typical asymptotic approximation, give the result of this test of the dataset below. Use level of significance $\alpha = .10$. What is your conclusion?

Data: 4.15 2.31 2.44 3.80 2.27 0.94 3.85 0.76 0.88 2.66 4.25 1.85

Statistics:

N	Mean	StDev	Minimum	Q1	Median	Q3	Maximum
12	2.513	1.277	0.761	1.164	2.373	3.840	4.251

3.Consider the following experiment: In a sequence of independent trials, roll two fair, six-sided, dice (outcomes in {1,2,3,4,5,6} all equally likely on each trial). One die is green, the other, the one that I will call the "Stop Die", is red. Roll only until the first time a "6" is rolled on the Stop Die. So we roll until the red die shows a six. Let S denote the sum of all of the results on both dice, up to and including that last roll. Our aim is to find the expectation and standard deviation of S. The problem consists of some steps along the way, as detailed below.

a. Denote by Z the number of times the pair of dice is rolled. Name and write down the distribution of Z.
b. Derive E(Z) and Var(Z). Use any method you like, but show the derivation.
c. Let X_i denote the number rolled on the green die on trial 'i', and let Y_i denote the number rolled on the green die on trial 'i'. Note that $I \leq Z$. Write down the conditional distribution, expectation, and variance of X_i and Y_i given i<Z and given I=Z.
d. Use the formula E(S) = E(E(S|Z)) to find E(S).
e. Use the formula Var(S) = Var (E(S|Z)) + E(Var(S|Z)) to finish the problem by finding the standard deviation of S.