Department of Mathematics
University of Toledo

Master of Science Degree
Comprehensive Examination
Probability and Statistical Theory

May 13, 1995

Instructions:

You will do a total of five problems.

Three should be chosen from the *Detailed Knowledge* Questions, numbered 1-4.

Two should be chosen from the *General Knowledge* Questions, numberd 5-8.

List the problems to be graded on the cover of your Blue Book.

Show all of your computations in your Blue Book.

Prove all of your assertions or quote appropriate theorems.

Books, notes, and calculators *may be used.*

This is a three hour test.

DETAILED KNOWLEGDE QUESTIONS, Numbers 1-4 (Select *THREE*)--

1. Objects I and II have weights w_1 and w_2, respectively. First, object I is weighed, then object II is weighed, and finally the two objects are weighed together. Denote the three observed measurements by Y_1, Y_2, and Y_3, respectively. Assume that the three measurements are independent, normally distributed, unbiased, and that the variances are proportional to the number of objects being weighed. That is, $Var(Y_1) = Var(Y_2) = \sigma^2$ and $Var(Y_3) = 2\sigma^2$.

a. Find the maximum likelihood estimator for the vector (w_1, w_2).

b. Find the biases and variances of the two maximum likelihood estimators.

c. Find the biases and variances of the estimators Y_1 and Y_2.

d. Which estimators are preferable, Y_1 and Y_2 or the maximum likelihood estimators? State your reasoning.

7. A quality control plan for an assembly line involves sampling $n = 10$ finished items per day and counting Y, the number of defective items. If θ denotes the probability of observing a defective item, then Y has a binomial distribution, when the number of items produced by the line is large. In other words, we can assume that $Y \sim B(n, \theta)$ with $n = 10$. However, θ varies from day to day and is assumed to have a uniform distribution on the interval from 0 to 1/4.

(a) Find the expected value of Y for any given day.

(b) Find the standard deviation of Y.

(c) Find the correlation coefficient between Y and θ.

3. Let $X_1, \ldots, X_n$ be a random sample from a population with density function given by $f(x;\theta) = e^{-(x-\theta)}$, where $-\infty < \theta < x < \infty$.

(a) Show that $X_{(1)} = \min(X_1, X_2, \ldots, X_n)$ is a sufficient statistic.

(b) Is $X_{(1)}$ a complete sufficient statistic? Explain your reasoning.

(c) Let $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$. Are $X_{(1)}$ and S^2 independent? Explain your reasoning.

(d) Are $X_{(1)}$ and $n\bar{X} - X_{(1)}$ independent? Explain your reasoning.

(e) Find the maximum likelihood estimate $\hat{\theta}$ of θ.

(f) Find the distribution function of $n(\hat{\theta} - \theta)$. In addition, find the mean and variance of $n(\hat{\theta} - \theta)$.

4. Let $X_1, \ldots, X_n$ be a random sample from a $N(\mu, \sigma^2)$ population, where $-\infty < \mu < \infty$ and $\sigma^2 > 0$.

(a) Let $\bar{X} = \frac{1}{n}\sum_{i=1} X_i$ and $S_n^2 = \sum_{i=1}^n (X_i - \bar{X})^2$. Find $\mathrm{Cov}(\bar{X}, S_n^2)$.

(b) If σ^2 is known, find the UMVU estimator of μ^4. Does the UMVU estimator achieve the Cramér-Rao lower bound?

(c) If $\mu = 0$ is known, find the UMVU estimator of $\exp(\sigma^2/2)$. Does the UMVU estimator achieve the Cramér-Rao lower bound?

(d) Let $\mu = 0$ and $n = 20$. Suppose that we wish to test the null hypothesis $H_0 : \sigma^2 = 1$ versus the alternative hypothesis $H_a : \sigma^2 < 1$. Use the Neyman-Pearson lemma to find the most powerful critical region of size $\alpha = 0.05$.

GENERAL KNOWLEDGE QUESTIONS, Numbers 5-8 (Select *TWO*)--

5 . In the following we define a multinomial experiment and a random vector $\underset{\sim}{X}$ with the multinomial distribution.

Let $S = \{\omega_1, \omega_2, \ldots, \omega_d\}$. The experiment consists of n independent trials where each trial consists of selecting one of the elements of S. For notation, we assume that $p_i = P(\omega_i$ is selected on any one trial) for $i = 1,2,\ldots,d$. Clearly $\sum_{i=1}^{d} p_i = 1$. This sequence of selections is called a sequence of multinomial trials and constitutes a multinomial experiment.

Further, for each i, we define the random variable X_i to be the number of times ω_i is selected in the n trials. The random vector $\underset{\sim}{X} = (X_1, X_2, \ldots, X_d)'$ (' means transpose) is called a multinomial random vector and is said to have the multinomial distribution with parameters n and $\underset{\sim}{p} = (p_1, p_2, \ldots, p_d)'$, that is $\underset{\sim}{X} \sim MN(n,\underset{\sim}{p})$.

a. Write down the mean vector $\underset{\sim}{\mu} = E(\underset{\sim}{X})$ in terms of n and $\underset{\sim}{p}$.

b. The distribution of $\underset{\sim}{X}$ as defined is singular. Find a vector in $\underset{\sim}{a} \in R^d$ and a scalar c such that $\underset{\sim}{a}'\underset{\sim}{x} = c$ for all $\underset{\sim}{x} \in \text{Range}(\underset{\sim}{X})$.

c. In the case d=2 (a very special case), find the variance-covariance matrix Σ. Also find the eigenvalues and the associated eigenvectors of Σ.

d. In part c, you should have found one of the eigenvalues to be 0. Also, the associated eigenvector $\underset{\sim}{a_0}$ should satisfy $\underset{\sim}{a_0}'\underset{\sim}{x} = c$ (some constant) for all $\underset{\sim}{x} \in \text{Range}(\underset{\sim}{X})$. Explain why these things are true.

6 . For this problem, you should employ the following table for the Binomial(n=4,p) distribution.

x	p=.2	p=.4	p=.5	p=.6	p=.8
0	0.4096	0.1296	0.0625	0.0256	0.0016
1	0.4096	0.3456	0.2500	0.1536	0.0256
2	0.1536	0.3456	0.3750	0.3456	0.1536
3	0.0256	0.1536	0.2500	0.3456	0.4096
4	0.0016	0.0256	0.0625	0.1296	0.4096

Let X have a binomial distribution with n=4 and $p \in \{0,.2,.4,.5,.6,.8,1\}$ (i.e., assume a highly restricted parameter space).

a. Based upon one observation X, for each $x \in$ Range(X), find the maximum likelihood estimate of p.

b. Derive the likelihood ratio test of H_0: $p \in \{0,.2,.4,.5\}$ versus H_1: $p \in \{.6,.8,1\}$ based on one observation of X. Find a test so that the level of significance α is less than or equal to .10. Find α. [Hint: you can probably guess the test, but to confirm that it is the likelihood ratio test, you need to compute the likelihood ratio at all $x \in$ Range(X). Also recall that the official definition of α in this case is max{ P(reject H_0) : $p \in \{0,.2,.4,.5\}$ }].

c. Find the power function for your test. Use the entire parameter space as the domain of the power function.

7 . Say that n treatment observations are Bernoulli(p) and m control observations are Bernoulli(q). Recall that in the presence of ties, the Wilcoxon rank sum statistic, W_s, is the sum of the treatment midranks. Denote by X and Y the number of succeses (ones) for the treatment and control, respectively.

a. Show that, in this Bernoulli situation, $W_s = 1/2 \{ n(n+m+1-Y) + mX \}$.

In parts b -d, our aim is to test H_0: p=q versus H_1:p>q. We will explore two possibilities in the simple case where m=3 and n=2.

b. Say that there are a total of three treatment and control observations tied at one. In the spirit of the approach taken in our course, under H_0, find the exact conditional distribution of W_s given X+Y = 3 (and thus $d_0 = n+m-3 = 2$ = the number of observations tied at zero and $d_1 = 3$ = the number of observations tied at one).

c. It is also possible to find the exact unconditional null distribution of W_s. Find this distribution. Note that it will depend upon the unknown value of p (=q).

d. According to the formula in part "a", the largest W_s can be is 10.5. Say that we decide to reject H_0 only when $W_s = 10.5$. Find the maximum possible value of the level of significance α in this case.

8. Let T and U be nonnegative, independent random variables. The variable $X = \min(T, U)$ is a censored observation of the failure time variable T, and $\delta = I_{[T \leq U]}$ is the indicator variable for the event of an uncensored observation of T. Furthermore, let $F(t) = P(T \leq t)$, $S(t) = 1 - F(t)$, $G(t) = P(U \leq t)$, $C(t) = 1 - G(t)$, and $H(t) = P(X \leq t)$.

(a) Show that $H(t) = 1 - S(t)C(t)$.

(b) Show that $P(X \leq t, \delta = 1) = -\int_0^t C(u-)dS(u)$.

(c) Show that $P(X \leq t, \delta = 0) = -\int_0^t S(u)dC(u)$.

(d) Let $\tau_F = \inf\{t;\ F(t) = 1\}$, show that $\tau_H = \min(\tau_F, \tau_G)$.