STATISTICS Ph. D. COMPREHENSIVE EXAM

January 13, 2003

Instructions: Write your ID number on your answer sheets. Do not put your name on any of your answer sheets. All solutions should be rigorously explained. Show all work.

1. Let $Z_1, Z_2 \overset{iid}{\sim} N(0, \theta)$, where $\theta = Var(Z_i) > 0$. Let

$$\begin{aligned} X &= (Z_1 - Z_2)^2, \text{and} \\ Y &= Z_1 Z_2. \end{aligned}$$

The general problem is to compare the four estimators

$$\begin{aligned} \hat{\theta}_1 &= X/2 \\ \hat{\theta}_2 &= |Y| \\ \hat{\theta}_3 &= \text{the moment estimator based on} \\ & \quad\ \text{the conditional distribution of Y given X} \\ \hat{\theta}_4 &= E\left[\frac{Z_1^2 + Z_2^2}{2}|(X,Y)\right] \end{aligned}$$

in terms of their mean squared errors $MSE_i(\theta) = E[(\hat{\theta}_i - \theta)^2]$, $i = 1, 2, 3, 4$.

(a) Show that $MSE_2(\theta) < MSE_1(\theta)$.

(b) A student suggests a quick way of computing $E_\theta[Y|X]$. Given $X = d^2$, $Z_2 = Z_1 \pm d$, so that $Y = Z_1(Z_1 \pm d)$ has conditional expectation $E_\theta[Z_1^2 \pm dZ_1] = E_\theta[Z_1^2] = \theta$. Is this calculation correct? Justify your answer.

(c) Find an explicit expression for $\hat{\theta}_3$ in terms of X and Y and compute $MSE_3(\theta)$. *Hint:* Transform (Z_1, Z_2) into $(U = Z_1 - Z_2, V = Z_1 + Z_2)$.

(d) Find an explicit expression for $\hat{\theta}_4$ in terms of X and Y and compute $MSE_4(\theta)$.

(e) Find an estimator that has smaller MSE than any of the above four. *Ignorable hint:* Try combining $\hat{\theta}_2$ and $\hat{\theta}_4$.

2. Give necessary definitions and then formulate and prove Basu's theorem. Hint: This is a theorem about very specific relationship between sufficient and ancillary statistics.

3. Prove the following theorem that is the main tool in finding a minimal sufficient statistic.

   Theorem *Let $f(x^n|\theta)$ be the pmf or pdf of a sample $X^n = (X_1, \ldots, X_n)$. Suppose there exists a function $T(x^n)$ such that, for every two sample points x^n and y^n, the ratio $f(x^n|\theta)/f(y^n|\theta)$ is a constant as a function of θ if and only if $T(x^n) = T(y^n)$. Then $T(X^n)$ is a minimal sufficient statistic for θ.*

4. Let $X_1, \ldots, X_n \overset{iid}{\sim} Poisson(\lambda)$. Let $\theta = 1/\lambda$.

   (a) Find the MLE of θ, say $\hat{\theta}$.

   (b) Find the Fisher information for θ, $I(\theta)$.

   (c) What is the large sample distribution of $\hat{\theta}$?

   (d) Prove $\hat{\theta} \xrightarrow{P} \theta$, that $\hat{\theta}$ converges in probability to θ.

5. Suppose X and Y are independent Poisson random variables with means λ and μ, respectively.

   (a) Derive the form for the uniformly most powerful unbiased (UMPU) test of size α for testing $H_0 : \lambda \leq \mu$ against $H_a : \lambda > \mu$. Clearly state the reference distribution for this test and how the rejection region is found for a given value of α.

   (b) Assume that the observed value of $X + Y$ is large. Show how to approximate the critical region for the test derived above.

6. Consider the simple linear regression model

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \text{ for } i = 1, \ldots, n,$$

   where $\bar{x} = 0$. As usual, define Y to be the $n \times 1$ vector of responses, X to be the $n \times 2$ design matrix, β to be the 2×1 vector of regression coefficients, and ϵ to be the $n \times 1$ vector of errors:

$$Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \text{ and } \epsilon = \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix}.$$

Assume X is full rank and

$$cov(\epsilon_i, \epsilon_j) = \left\{ \begin{array}{ll} \sigma^2 & \text{if } i = j \\ \sigma^2\rho & \text{if } i \neq j \end{array} \right\},$$

where ρ is a known, positive constant such that $0 < \rho < 1$.

(a) Write out the covariance matrix of the vector ϵ as $cov(\epsilon) = \sigma^2 V$, where V is positive definite.

(b) The ordinary least squares (OLS) estimate of β is best linear unbiased (BLUE) if and only if $C(VX) \subset C(X)$, where $C(X)$ is the column space of X. Show that the OLS estimate of β for the regression model is BLUE.

(c) Derive the BLUE of β, say $\hat{\beta}$. Specify the distribution of $\hat{\beta}$.

(d) Describe in detail how to test $H_0 : \beta_1 = 0$ against $H_a : \beta_1 \neq 0$ assuming σ^2 is unknown.