STATISTICS PH.D. COMPREHENSIVE EXAM

August 11, 2006 1-4 PM

Directions: The exam consists of six questions of equal point value. Make sure to write your ID number (last 4 digits of your Social Security number) on your answer sheets. Do not put your name on any of your answer sheets. All solutions should be complete and rigorous. If you cannot solve a problem, try to outline an approach that may lead to a solution.

1) Consider the balanced two-way ANOVA model with interaction:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \eta_{ij} + \epsilon_{ijk} \text{ for } i = 1, ..., I \;\; j = 1, ..., J, \;\text{ and } k = 1, ..., K$$

where the ϵ_{ijk} are iid $N(0, \sigma^2)$. Which of the following are estimable? If estimable, provide the Least Squares (LS) estimate, and how you would compute a 95% confidence interval for the estimable function.

(a) μ

(b) $\alpha_i - \alpha_j$ where $i \neq j$

(c) $\mu + \alpha_i + \beta_j + \eta_{ij}$

(d) $\alpha_i - \alpha_k + \eta_{ij} - \eta_{kj}$

2. Let U_k have a chi-squared distribution with $k > 0$ degrees of freedom (a chi-squared distribution with non-integer degrees of freedom is defined by a relationship with the gamma distribution). Letting n be a positive integer, determine in (a) through (c) the limiting distributions as $n \to \infty$ of

(a)

$$\frac{U_n}{n}$$

(b)

$$\frac{U_n - n}{\sqrt{2n}}$$

(c) $U_{1/n}$

(d) Let $X_1, X_2, ..., X_n$ be independent $N(\mu, \sigma^2)$ random variables. Define the sample variance by $s^2 = \sum_i (X_i - \bar{X})^2/(n-1)$, where $\bar{X} = \sum_i X_i/n$ is the sample mean. Find the large sample distribution of

$$\sqrt{n}\{log(s^2) - log(\sigma^2)\}$$

as $n \to \infty$.

3. Suppose (X_1, X_2) has a joint distribution and that $\mathrm{E}(X_i^2)$ is finite for $i = 1, 2$. Assume $\mathrm{E}(X_1|X_2) = X_2$ and $\mathrm{E}(X_2|X_1) = X_1$. Show that $X_1 = X_2$ with probability 1.

4. Let $F(x)$ be the cumulative distribution function (cdf) for a known continuous density function $f(x)$ on R^1.

(a) Show that $F(x)^\theta$ (i.e. $F(x)$ raised to the θ power) is a cdf on R^1 for each $\theta > 0$. Find the density function associated with $F(x)^\theta$.

(b) Let $X_1, X_2, ..., X_n$ be iid random variables with cdf $F(x)^\theta$ for $\theta > 0$. Find a complete sufficient statistic for θ.

(c) Find a minimum variance unbiased estimator (MVUE) for $1/\theta$. Does this MVUE achieve the Cramer-Rao lower bound?

5. (Continuation of 4). Let $X_1, X_2, ..., X_n$ be iid random variables with cdf $F(x)^\theta$ for $\theta > 0$, where $F(x)$ is a known cdf with a continuous density function $f(x)$ on R^1.

(a) Find the maximum likelihood estimator of $1/\theta$. What is the exact distribution of the MLE? What is the large sample distribution of the MLE (suitably centered and scaled)?

(b) Derive a general form of the rejection region for a most powerful size-α test of $H_0 : \theta = 1$ against $H_1 : \theta > 1$. Also, derive a suitable large sample approximation to the rejection region.

6. Consider the linear model $Y = X\beta + e$ where X is a fixed n by p matrix, β is a p by 1 parameter vector and $e \sim N_n(0, \sigma^2 I)$, where σ^2 is unknown. We are interested in testing $H_0 : A\beta = 0$, where A is a k by p matrix of constants, with $k < p$. For simplicity, assume that $A\beta$ is estimable.

(a) Using linear model theory, propose a test of H_0. Discuss properties of the test, and provide a careful description of how the test is carried out. Justify all steps in any derivations.

(b) Repeat part (a), assuming that $e \sim N_n(0, V)$, where V is a known positive definite matrix.