STATISTICS PH.D. COMPREHENSIVE EXAM
August 15, 2008 1-3 PM

Directions: The exam consists of five questions. Problems 1-4 are worth 20 points each while problem 5 is worth 40 points. 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 model:

$$y_{ij} = \mu + \alpha_i + \beta x_j + \epsilon_{ij} \text{ for } i = 1, ..., I \;\; j = 1, ..., J$$

where the x_j are fixed with $\sum_j x_j = 0$ and $\sum_j x_j^2 > 0$. The ϵ_{ij} are iid with mean zero. Which of the following are estimable? Give justifications for your answers.

   (a) μ

   (b) $\mu + \alpha_i$

   (c) β

   (d) $\alpha_i - \alpha_j$

   (e) Clearly describe, both in words, and with a graphical (or pictorial) representation, the mean structure in this model, and provide an interpretation for each of the estimable functions in (a)-(d).

2. 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 $e \sim N_n(0, \sigma^2 V)$, where σ^2 is an unknown scalar and where V is a known positive definite matrix. We are interested in testing $H_0 : a'\beta = 0$, where a is a p by 1 vector of constants and $a'\beta$ is estimable. 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.

3. Let $X_1, X_2, ...$ be independent random variables such that $\Pr(X_n = n^\alpha) = 1/n$ and $\Pr(X_n = 0) = 1 - 1/n$ for $n = 1, 2, ...$ where α is a constant. For what values of α, $-\infty < \alpha < \infty$ is it true that:

   (a) $X_n \to 0$ in probability as $n \to \infty$?

   (b) $X_n \to 0$ almost surely as $n \to \infty$?

(c) $X_n \to 0$ in r^{th} mean as $n \to \infty$?

4. Let $X_1,\ X_2, ...,\ X_n$ be iid $N(\mu, \sigma^2)$ random variables with $-\infty < \mu < \infty$ and $\sigma^2 > 0$. Find the minimum variance unbiased estimator of μ/σ. To obtain the final answer, it may help to recall that the density function for a χ^2_r distribution (where $r > 0$ is the degrees of freedom) is given by

$$f(y) = \frac{y^{.5r-1} exp(-.5y)}{\Gamma(.5r)2^{.5r}}$$

for $y > 0$ and zero otherwise.

5. Assume that $y_i = f(x_i) + \varepsilon_i,\ i = 1, \ldots, n$ for some unknown function f defined on $x \in [0, 1]$ and ε_i are independent errors with mean 0 and variance σ^2. The x_is are fixed by design. The regressogram estimator is a flexible regression estimator defined by splitting the domain into λ (non-overlapping) bins and taking averages of the y_i within these bins. Let

$$R_k = \left\{ x : \frac{k-1}{\lambda} \leq x < \frac{k}{\lambda} \right\}, \quad k = 1, \ldots, \lambda$$

and let

$$\bar{y}_k = \frac{\sum_{i=1}^n y_i I_{R_k}(x_i)}{\sum_{i=1}^n I_{R_k}(x_i)}$$

where $I_A(x)$ is the indicator function $I_A(x) = 1$ if $x \in A$ and 0 otherwise. Finally the regressogram estimator is given as

$$\hat{f}(x) = \sum_{k=1}^{\lambda} \bar{y}_k I_{R_k}(x).$$

Now consider estimation of f at a particular x_0 and assume that f is continuous and differentiable on $[0, 1]$. Also assume that for a given n, the design points are given by $x_i = (2i - 1)/2n,\ i = 1, \ldots, n$. **Note: you may use any of the results in previous steps to solve subsequent parts, even if you could not show they are true.**

(a) Give an expression for the variance of $\hat{f}(x_0)$ in terms of n and λ.

(b) What is the mean of $\hat{f}(x_0)$ in terms of $f(x_i),\ i = 1, \ldots, n$? Show that

$$\text{Bias}(\hat{f}(x_0)) = O(\lambda^{-1})$$

(i.e. $|\text{Bias}(\hat{f}(x_0))|/\lambda$ is bounded above for large λ) by using Taylor approximation of f around x_0. That is, use $f(x) = f(x_0) + r(x)$ where $r(x) = f'(\xi)(x - x_0)$ for some ξ between x and x_0.

(c) Suppose that $\lambda = \lfloor n^{1/2} \rfloor$ as $n \to \infty$. Show that

$$\frac{\sqrt{n/\lambda}}{\sigma} \left(\hat{f}(x_0) - f(x_0) \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1).$$

(d) Use (c) to provide an asymptotic confidence interval for $f(x_0)$ assuming σ is known.