Department of Statistics
University of Florida

PhD Qualifying Exam

August 20, 2004

Instructions:

1. You have exactly four hours to answer questions in this examination.
2. There are 8 problems of which you must answer 6.
3. Only your first 6 problems will be graded.
4. Write only on one side of the paper, and start each question on a new page.
5. Clearly label each part of each question with the question number and the part, e.g., **1(a)**.
6. Write your **number** on every page.
7. Do not write your name anywhere on your exam.
8. You must show your work to receive credit.
9. While the eight questions are equally weighted, within a given question, the parts may have different weights.

1. Let $X_1, \cdots, X_n$ be iid with common pdf (or pf) $f_\theta(x) = \exp[\theta x - \psi(\theta)]h(x)$, where $\theta \in \Theta$, some open interval in the real line.

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

(b) Find $E_\theta(X_1)$ and $V_\theta(X_1)$.

(c) Find the asymptotic distribution of some suitably normalized $\bar{X}_n$, where $\bar{X}_n = n^{-1}\sum_{i=1}^n X_i$.

(d) Using the delta method, find the asymptotic distribution of some suitably normalized $\hat{\theta}_n$.

(e) Find the UMP test for $H_0 : \theta \le \theta_0$ against the alternatives $H_1 : \theta > \theta_0$.

2. (a) Suppose $X \sim \text{Bin}(1, p)$, where $p \in [0, 1]$. Assume squared error loss.

(i) Show that $\frac{1}{2}X + \frac{1}{4}$ is a Bayes estimator of p with constant risk $\frac{1}{16}$.

(ii) Show that $\frac{1}{2}X + \frac{1}{4}$ is an admissible estimator of p.

(b) Let $X_{11}, X_{12}, X_{21}, X_{22}$ have a joint multinomial$(1; \theta/3, (1-\theta)/3, (1-\theta)/3, (1+\theta)/3)$ distribution, where $\theta \in [0, 1]$. Note that exactly one of the X_{ij} is 1, and the rest are zeroes. Assume squared error loss.

(i) Show that conditional on $X_{11} + X_{21} = 1$, $\frac{3}{4}X_{11} + \frac{1}{4}X_{21}$ is an admissible estimator of θ.

(ii) Show that conditional on $X_{12} + X_{22} = 1$, $\frac{1}{2}(X_{22} - X_{12})$ is an admissible estimator of θ.

(iii) Show that $T = \frac{3}{4}X_{11} + \frac{1}{4}X_{21} + \frac{1}{2}(X_{22} - X_{12})$ is an admissible estimator of θ conditional on either $X_{11} + X_{21} = 1$ or $X_{12} + X_{22} = 1$.

(iv) Show that T has constant risk $3/16$.

(v) Show that the estimator $S = \frac{3}{4}X_{11} - \frac{1}{8}(X_{12} + X_{21}) + \frac{1}{2}X_{22}$ of θ has risk $\frac{3}{32}(1+\theta)$, and hence, dominates T as an estimator of θ.

3. Let $\{Y_n, n \ge 1\}$ be a sequence of independent, nonnegative random variables and suppose that $\sum_{n=1}^\infty Y_n$ converges almost certainly. Let $\{X_n, n \ge 1\}$ be a sequence of random variables with X_n and Y_n having the same distribution for each $n \ge 1$.

(a) Prove that $\sum_{n=1}^\infty X_n$ converges almost certainly irrespective of the joint distributions of the $\{X_n, n \ge 1\}$.

(b) Demonstrate, by providing an example of sequences of random variables $\{Y_n, n \ge 1\}$ and $\{X_n, n \ge 1\}$, that part (a) can fail if the assumption that the $\{Y_n, n \ge 1\}$ are nonnegative is dispensed with.

4. Let $\{X_n, n \ge 1\}$ be a sequence of independent, mean 0, $\mathcal{L}_2$ random variables and let $S_n = \sum_{j=1}^n X_j$ and $s_n^2 = \sum_{j=1}^n EX_j^2$, $n \ge 1$. Prove that if $s_n^2 \to \infty$, $EX_n^2 = o(s_n^2)$, and $\frac{S_n}{s_n} \xrightarrow{d} N(0,1)$, then

$$\frac{\max_{1 \le j \le n} |X_j|}{s_n} \xrightarrow{P} 0.$$

5. Suppose that $\boldsymbol{u} \sim N(\boldsymbol{\mu}, \boldsymbol{I}_k)$ with $\boldsymbol{\mu} \neq \boldsymbol{0}$. Let $W = \boldsymbol{u}'\boldsymbol{u}$. It is known that W is distributed as ${\chi'_k}^2(\lambda)$, where $\lambda = \frac{1}{2}\boldsymbol{\mu}'\boldsymbol{\mu}$, and ${\chi'_k}^2(\lambda)$ denotes the noncentral chi-squared distribution with k degrees of freedom and a noncentrality parameter λ.

(a) Show that W can be written as

$$W = W_1 + W_2$$

where W_1 and W_2 are independently distributed such that $W_1 \sim {\chi'_1}^2(\lambda)$, $W_2 \sim \chi^2_{k-1}$.

(b) Let $Y_1 = e^{W_1}$. Give an approximate expression for the variance of Y_1 using the *Delta Method*.

6. Consider the one-way model,

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij},$$

$i = 1, 2, \ldots, k$; $j = 1, 2, \ldots, n_i$, where α_i is fixed and the ϵ_{ij} are independently distributed as normal variates with zero means and variances σ_i^2 $(i = 1, 2, \ldots, k)$.

It is assumed that the σ_i^2 are known, but are not equal. Let $\mu_i = \mu + \alpha_i$ be the mean of the ith treatment, and let $\psi_c = \sum_{i=1}^k c_i\mu_i$ be a linear function of the μ_i, where $c_1, c_2, \ldots, c_k$ are constants.

(a) Show that the simultaneous $(1-\alpha)100\%$ confidence intervals on all linear functions of the μ_i of the form ψ_c are given by

$$\hat{\psi}_c \pm \left[\chi^2_{\alpha,k} \sum_{i=1}^{k} \frac{c_i^2\sigma_i^2}{n_i}\right]^{1/2},$$

where $\hat{\psi}_c = \sum_{i=1}^k c_i\bar{y}_i$, and $\bar{y}_i$ is the sample mean of treatment i $(i = 1, 2, \ldots, k)$, and $\chi^2_{\alpha,k}$ is the upper α-quantile of the chi-squared distribution with k degrees of freedom.

(b) Suppose now that the σ_i^2 are unknown. Show how you can obtain an exact $(1-\alpha)100\%$ confidence region on the vector $(\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2)$.

7. Recall that distributions in an exponential dispersion family have densities of the form

$$f(y;\theta,\phi) = \exp\left\{\frac{1}{\phi}\big[\theta y - b(\theta)\big] + c(y;\phi)\right\}, \qquad (*)$$

for $\theta \in \Theta \subset \mathbb{R}$ and $1/\phi \in \Lambda \subset \mathbb{R}$. In a generalized linear model (GLM), we assume that $Y_1, \ldots, Y_n$ are independent and that Y_i has density $f(y;\theta_i,\phi/w_i)$, where f has the form $(*)$ and $w_1, \ldots, w_n$ are known weights. This defines the random, or stochastic part of the GLM.

(a) Describe the systematic part of the GLM relating the mean $\mu_i = E(Y_i)$ to the vector $\boldsymbol{x}_i$ of predictor values for the ith observation.

(b) Derive the maximum likelihood estimating equations (score equations) for the vector of regression coefficients $\boldsymbol{\beta}$.

(c) Derive the form of the elements of the "observed" and "expected" Fisher information matrices for $\boldsymbol{\beta}$. Express the expected Fisher information matrix in matrix form.

(d) What is the canonical link function for the density given in $(*)$? Show that if ϕ is known and the canonical link is used, then there is a simple sufficient statistic for the vector of regression parameters, $\boldsymbol{\beta}$, and give its form. Show also that the score equations and the observed Fisher information for β can be simplified when the canonical link is used and give the simplified forms.

8. The standard Cauchy (or Cauchy$(0, 1)$) distribution has density $f(u) = \frac{1}{\pi}\frac{1}{1+u^2}$, $-\infty < u < \infty$, cumulative distribution function $F(u) = \frac{1}{2} + \frac{1}{\pi}\arctan(u)$, and characteristic function $\varphi(t) = e^{-|t|}$. If U has a standard Cauchy distribution, then the distribution of $\sigma U + \mu$, where $-\infty < \mu < \infty$, and $\sigma > 0$, is known as a Cauchy distribution with location parameter (median) μ and scale parameter σ. We will denote this distribution by the notation Cauchy(μ, σ).

(a) If U_1 and U_2 are independent random variables, $U_i \sim$ Cauchy(μ_i, σ_i), $i = 1, 2$, and a_1 and a_2 are real numbers, what is the distribution of $a_1U_1 + a_2U_2$? Justify (prove) your answer.

(b) Suppose that Y is a binary response following a generalized linear mixed model (GLMM) for $E(Y|\boldsymbol{U})$, with link function $g = F^{-1}$ (where F is the cdf of the Cauchy$(0, 1)$ distribution as given above) and linear predictor $\eta = \boldsymbol{x}^T\boldsymbol{\beta} + \boldsymbol{z}^T\boldsymbol{U}$, where $\boldsymbol{x} \in \mathbb{R}^p$ and $\boldsymbol{z} \in \mathbb{R}^q$ are known covariates, $\boldsymbol{\beta} \in \mathbb{R}^p$ is a vector of regression parameters, and $\boldsymbol{U} = (U_1, \ldots, U_q)$ is a vector of independent random variables with $U_i \sim$ Cauchy$(0, \sigma_i)$, $i = 1, \ldots, q$. Show that the marginal model for Y is a GLM, identify its link function, and express its coefficient vector in terms of the elements of the GLMM. In this context, what general statement can you make about the effect of a change in x_i on the marginal mean $E(Y)$ versus its effect on the conditional mean $E(Y|\boldsymbol{U})$?