Department of Statistics
University of Florida

PhD Qualifying Exam

August 2007

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 your chosen identifying number on every page.
5. Do not write your name anywhere on your exam.
6. Write only on one side each page of paper, and start each question on a new page.
7. Clearly label each part of each question with the question number and the part, e.g., **1(a)**.
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.
10. Do not write too near the upper left corner of the page where the pages will be stapled together.

1. Suppose $\mathbf{X}|\boldsymbol{\theta} \sim N(\boldsymbol{\theta}, \mathbf{I}_p)$, and $\boldsymbol{\theta}$ has the $N(\mathbf{0}, \tau^2\mathbf{I}_p)$ prior. Note that the posterior distribution of $\boldsymbol{\theta}$ given $\mathbf{X} = \mathbf{x}$ is $N((1-B)\mathbf{x}, (1-B)\mathbf{I}_p)$, where $B = (1+\tau^2)^{-1}$.

(a) Show that the Bayes estimator of $\boldsymbol{\theta}$ under the squared error loss $L(\boldsymbol{\theta}, \mathbf{a}) = (\boldsymbol{\theta} - \mathbf{a})^T(\boldsymbol{\theta} - \mathbf{a})$ is given by $\hat{\boldsymbol{\theta}}_B(\mathbf{X}) = (1-B)\mathbf{X}$.

(b) A general empirical Bayes estimator of $\boldsymbol{\theta}$ is given by $\hat{\boldsymbol{\theta}}_{EB}(\mathbf{X}) = (1 - \hat{B}(S))\mathbf{X}$, where $S = \sum_{i=1}^p X_i^2$, and $\hat{B}(S)$ (purported to estimate B) depends on $\mathbf{X}$ only through S. For any general estimator $\mathbf{e}$ of $\boldsymbol{\theta}$, let $r(\mathbf{e})$ denote its Bayes risk under the given likelihood, the prior and the given loss. Show that $r(\hat{\boldsymbol{\theta}}_{EB}) = r(\hat{\boldsymbol{\theta}}_B) + E[(\hat{B}(S) - B)^2 S]$.

(c) The relative savings loss (RSL) of $\hat{\boldsymbol{\theta}}_{EB}$ with respect to $\mathbf{X}$ is defined by

$$\text{RSL}(\hat{\boldsymbol{\theta}}_{EB}, \mathbf{X}) = [r(\hat{\boldsymbol{\theta}}_{EB}) - r(\hat{\boldsymbol{\theta}}_B)]/[r(\mathbf{X}) - r(\hat{\boldsymbol{\theta}}_B)].$$

Show that $\text{RSL}(\hat{\boldsymbol{\theta}}_{EB}, \mathbf{X}) = (pB)^{-1}E[(\hat{B}(S) - B)^2 S]$.

(d) The James-Stein empirical Bayes estimator of $\boldsymbol{\theta}$ has $\hat{B}(S) = (p-2)/S$, $(p \geq 3)$. Show that for this particular empirical Bayes estimator, the RSL expression given in (c) simplifies to $2/p$.

2. Let $X_1, \ldots, X_n$ be iid $\text{Bin}(1, \theta)$. Then the MLE of $\theta(1-\theta)$ is given by (you need not derive) $T_n = \bar{X}_n(1 - \bar{X}_n)$, where $\bar{X}_n = \sum_{i=1}^n X_i/n$.

(a) Show that for $\theta \neq \frac{1}{2}$, $n^{1/2}[T_n - \theta(1-\theta)] \xrightarrow{d} N(0, h(\theta))$, where $h(\theta) = (1-2\theta)^2\theta(1-\theta)$.

(b) Show that when $\theta = \frac{1}{2}$, $n(T_n - \frac{1}{4}) \xrightarrow{d} -\frac{1}{4}\chi_1^2$.

(c) For $n \geq 2$, the UMVUE of $\theta(1-\theta)$ is given by (you need not derive) $S_n = \frac{n}{n-1}T_n$. Show that $n^{1/2}(S_n - T_n) \to 0$ in probability as $n \to \infty$.

3. Consider the random one-way model,

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \quad i = 1, 2, \ldots, k; \; j = 1, 2, \ldots, n,$$

where $\alpha_i \sim N(0, \sigma_\alpha^2)$, $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$, and the α_i's and the ϵ_{ij}'s are mutually independent.

(a) Obtain a $(1-\alpha)$ 100% confidence interval on $n\,\sigma_\alpha^2 + \sigma_\epsilon^2$.

(b) Obtain a $(1-\alpha)$ 100% confidence interval on the ratio $\sigma_\alpha^2/\sigma_\epsilon^2$.

(c) Use parts (a) and (b) to obtain an exact confidence region on $(\sigma_\epsilon^2, \sigma_\alpha^2)$ with a confidence coefficient $\geq 1 - 2\,\alpha$.

(d) Use the result in (c) to obtain exact simultaneous confidence intervals on σ_α^2 and σ_ϵ^2 with a joint confidence coefficient $\geq 1 - 2\,\alpha$.

(e) Assume now that the above model is unbalanced, that is, $j = 1, 2, \ldots, n_i$, $i = 1, 2, \ldots, k$. All the assumptions made earlier regarding the random effects remain valid here. Let SS_A be the sum of squares associated with α_i in the model. What distribution does SS_A have? Please be specific giving all the necessary details.

4. Consider the general balanced model,

$$\mathbf{y} = \mathbf{X}\mathbf{g} + \mathbf{Z}\mathbf{h},$$

where $\mathbf{X}\mathbf{g} = \sum_{i=0}^{\nu-p} \mathbf{H}_i\boldsymbol{\beta}_i$ is fixed and $\mathbf{Z}\mathbf{h} = \sum_{i=\nu-p+1}^{\nu+1} \mathbf{H}_i\boldsymbol{\beta}_i$ is random, $0 \leq p \leq \nu$; the $\boldsymbol{\beta}_i$'s are mutually independent such that $\boldsymbol{\beta}_i \sim N(\mathbf{0}, \sigma_i^2\mathbf{I}_{c_i})$, $i = \nu-p+1, \nu-p+2, \ldots, \nu+1$. Let $\mathbf{P}_i$ be the matrix corresponding to the i^{th} sum of squares in the model whose rank in m_i, $(i = 0, 1, \ldots, \nu+1)$.

(a) Suppose that $\boldsymbol{\lambda}'\mathbf{g}$ is an estimable linear function of $\mathbf{g}$. What is its B.L.U.E? (just give the expression). Can you compute this B.L.U.E without knowing the model's variance components?

(b) Show that $[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}' - \sum_{i=0}^{\nu-p} \mathbf{P}_i]\mathbf{X} = \mathbf{0}$.

(c) Show that $m_i = \text{rank}(\mathbf{P}_i\mathbf{X})$ for $i = 0, 1, \ldots, \nu - p$.

(d) Show that $\text{rank}(\sum_{i=0}^{\nu-p} \mathbf{P}_i) = \sum_{i=0}^{\nu-p} m_i$.

(e) Show that $\text{rank}(\mathbf{X}) = \sum_{i=0}^{\nu-p} m_i$.

5. Let $\{X_n, n \geq 1\}$ be a sequence of independent nonnegative random variables and let $\{Y_n, n \geq 1\}$ be a sequence of random variables such that Y_n and X_n are identically distributed for each $n \geq 1$. Prove that if

$$\sum_{n=1}^{\infty} X_n < \infty \quad \text{a.c.},$$

then

$$\sum_{n=1}^{\infty} Y_n < \infty \quad \text{a.c.}$$

6. Let $\{X_n, n \geq 1\}$ be a sequence of independent $\mathcal{L}_2$ random variables with $EX_n = 0$, $n \geq 1$. Set $s_n^2 = \sum_{j=1}^{n} EX_j^2$, $n \geq 1$, and suppose that $s_n^2 \to \infty$. Prove that if

$$\lim_{n\to\infty} \frac{X_n}{s_n} = 0 \quad \text{a.c.} \tag{6.1}$$

and

$$\lim_{n\to\infty} \frac{1}{s_n^2} \sum_{j=1}^{n} E(X_j^2 I_{[|X_j| > \varepsilon_0 s_j]}) = 0 \quad \text{for some } \varepsilon_0 \in (0, \infty), \tag{6.2}$$

then

$$\frac{\sum_{j=1}^{n} X_j}{s_n} \xrightarrow{d} N(0, 1) \tag{6.3}$$

and

$$\frac{\max_{1\leq j\leq n} |X_j|}{s_n} \to 0 \quad \text{a.c.} \tag{6.4}$$

7. Suppose that $Y_1, \ldots, Y_n$ are independent, and satisfy $E\boldsymbol{\psi}(Y_i; \boldsymbol{\beta}, \boldsymbol{x}_i) = \mathbf{0}$, $i = 1, \ldots, n$, for some known function $\boldsymbol{\psi}$ and "true" but unknown value of $\boldsymbol{\beta}$. An M-estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ is a solution of the system of equations

$$\sum_{i=1}^{n} \boldsymbol{\psi}(y_i; \boldsymbol{\beta}, \boldsymbol{x}_i) = \mathbf{0}.$$

Under appropriate regularity conditions, $\hat{\boldsymbol{\beta}}$ is approximately normally distributed with mean $\boldsymbol{\beta}$ and covariance matrix $n^{-1}V_n$, where V_n is the so-called "sandwich matrix" $V_n = A_n^{-1} B_n A_n^{-T}$, with

$$A_n = -\frac{1}{n}\sum_{i=1}^{n} E\big[\boldsymbol{\psi}'(y_i; \boldsymbol{\beta}, \boldsymbol{x}_i)\big] \quad \text{and} \quad B_n = \frac{1}{n}\sum_{i=1}^{n} E\Big[\boldsymbol{\psi}(Y_i; \boldsymbol{\beta}, \boldsymbol{x}_i)\boldsymbol{\psi}(Y_i; \boldsymbol{\beta}, \boldsymbol{x}_i)^T\Big],$$

where $\boldsymbol{\psi}'(y; \boldsymbol{\beta}, \boldsymbol{x}) = \partial\boldsymbol{\psi}(y; \boldsymbol{\beta}, \boldsymbol{x})/\partial\boldsymbol{\beta}^T = \big(\partial\psi_j(y; \boldsymbol{\beta}, \boldsymbol{x})/\partial\beta_k\big)_{1\le j,k\le p}$. (You are free to use these results in the remainder of this problem if you need them.)

Now, suppose in particular that $\hat{\boldsymbol{\beta}}$ is the solution in $\boldsymbol{\beta}$ to the system of equations

$$\sum_{i=1}^{n} \frac{w_i(y_i - \mu_i)}{V(\mu_i)g'(\mu_i)}\boldsymbol{x}_i = \mathbf{0}, \qquad g(\mu_i) = \boldsymbol{x}_i^T\boldsymbol{\beta}, \quad i = 1, \ldots, n, \tag{7.1}$$

where $w_1, \ldots, w_n > 0$ are known weights, $g(\mu)$ is a smooth, strictly monotone link function, and $V(\mu)$ is a smooth, strictly positive "working variance function."

(a) Assuming that $\mu_i = E(Y_i)$ satisfies

$$g(\mu_i) = \boldsymbol{x}_i^T\boldsymbol{\beta}, \tag{7.2}$$

for all $i = 1, \ldots, n$, show that $\hat{\boldsymbol{\beta}}$ is approximately normally distributed (you are not required to specify any regularity conditions) with mean $\boldsymbol{\beta}$ and covariance matrix $n^{-1}V_n = n^{-1}A_n^{-1}B_nA_n^{-T}$, and show that the factors of the sandwich matrix V_n have the form $A_n = n^{-1}\, X^T\Omega X$ and $B_n = n^{-1}\, X^T\Omega^* X$. Give the form of Ω and Ω^*.

(b) Assuming that (7.2) holds, suggest a consistent estimator of V_n (in the sense that $\hat{V}_n - V_n \xrightarrow{P} 0$). Just give the estimator; you do not have to prove that it is consistent nor specify any regularity conditions.

(c) In a quasi-likelihood analysis, one assumes in addition to (7.2) that

$$\text{Var}(Y_i) = \frac{\phi}{w_i}V(\mu_i), \quad i = 1, \ldots, n. \tag{7.3}$$

Show that V_n takes a particularly simple form in this case.

(d) Assuming that both (7.2) and (7.3) hold, give a consistent estimator of ϕ. Just give the estimator; you do not have to prove that it is consistent nor specify any regularity conditions.

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. 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 $E(Y|\boldsymbol{U})$ (i.e., $g\big(E(Y|\boldsymbol{U})\big) = \eta$) 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, $U_i \sim$ 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 other words, show that $g^*\big(E(Y)\big) = \boldsymbol{x}^T\boldsymbol{\beta}^*$ for some link function g^* and vector of regression coefficients $\boldsymbol{\beta}^*$, specifying the function g^* and giving a formula for the elements of $\boldsymbol{\beta}^*$ in terms of $\boldsymbol{\beta}$, $\boldsymbol{z}$, and $\sigma_1, \ldots, \sigma_q$).