Department of Statistics
University of Florida

PhD Qualifying Exam

August 22, 2003

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. Write your **number** on every page.
6. Do not write your name anywhere on your exam.
7. You must show your work to receive credit.
8. While the eight questions are equally weighted, within a given question, the parts may have different weights.

1. Suppose that X has a binomial distribution with m trials and probability μ.

(a) Express the binomial mass function in exponential form in terms of the canonical parameter $\theta = \text{logit}(\mu)$.

(b) Derive the deviance measure of fit $D(y, \mu)$ for the binomial model, where $Y = X/m$.

(c) Show that the deviance can be approximated by the Pearson chisquared statistic, χ^2, if m is large, where

$$\chi^2 = \frac{m(Y-\mu)^2}{\mu(1-\mu)}\,.$$

(d) Suppose that $X_i \sim B(m_i, \mu_i)$, $i = 1, \ldots, n$, independently, and that the success probabilities satisfy a linear logistic model. Let $H_0 \subset H_1$ be nested hypotheses. Show that the corresponding model deviances satisfy the Pythagorean relationship

$$D(\mathbf{y}, \hat{\boldsymbol{\mu}}_0) = D(\mathbf{y}, \hat{\boldsymbol{\mu}}_1) + D(\hat{\boldsymbol{\mu}}_1, \hat{\boldsymbol{\mu}}_0)\,,$$

where $\mathbf{y} = (y_1, \ldots, y_n)$ is the vector of observed proportions, and $\hat{\boldsymbol{\mu}}_0$ and $\hat{\boldsymbol{\mu}}_1$ are the vectors of fitted proportions under H_0 and H_1 respectively.

2. Suppose that $X \sim B(m, \mu)$ where $\mu = e^\theta/(1+e^\theta)$, and let Z be defined by the equation $X = m\mu + \sqrt{m\mu(1-\mu)}Z$.

(a) Show that for $c > 0$

$$E\left\{\log\frac{X+c}{m\mu}\right\} = \frac{c}{m\mu} - \frac{1-\mu}{2m\mu} + O(m^{-3/2})\,.$$

(b) Find the corresponding expansion for

$$E\left\{\log\frac{m-X+c}{m(1-\mu)}\right\}\,.$$

(c) Consider the estimate

$$\tilde{\theta} = \log\frac{X+c}{m-X+c}\,.$$

Show that

$$E(\tilde{\theta}) = \theta + \frac{(1-2\mu)(c-\frac{1}{2})}{m\mu(1-\mu)} + O(m^{-3/2})\,.$$

(d) Show that the expected value of the corresponding ML estimate, $\hat{\theta}$, does not converge to θ as m increases.

3. Consider the unbalanced random one-way model

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

where α_i and ϵ_{ij} are independently distributed such that $\alpha_i \sim N(0, \sigma^2_\alpha)$, and $\epsilon_{ij} \sim N(0, \sigma^2_\epsilon)$, $i = 1, 2, \cdots, a$; $j = 1, 2, \cdots, n_i$. Let SS_α be the sum of squares,

$$\text{SS}_\alpha = \sum_{i=1}^{a} \frac{y_{i.}^2}{n_i} - \frac{y_{..}^2}{n_.},$$

where $n_. = \sum_{i=1}^{a} n_i$.

(a) Write SS_α as $\mathbf{y}'\mathbf{A}\mathbf{y}$, indicating what the matrix $\mathbf{A}$ is, where $\mathbf{y}$ is the vector of all observations.

(b) Make use of part (a) to show that

$$E(\text{SS}_\alpha) = (n_. - \frac{1}{n_.}\sum_{i=1}^{a} n_i^2)\sigma^2_\alpha + (a-1)\sigma^2_\epsilon$$

(c) What distribution does SS_α have if

(i) $\sigma^2_\alpha = 0$

(ii) $\sigma^2_\alpha \neq 0$

(d) Show that SS_α and SS_E, the residual sum of squares, are independent.

(e) Show how to obtain the expected vatue of $\frac{\text{SS}_\alpha}{\text{SS}_E}$.

4. Consider the balanced mixed model

$$y_{ijk} = \mu + \alpha_{(i)} + \beta_{(j)} + \gamma_{i(k)} + (\alpha\beta)_{(ij)} + \epsilon_{i(jk)},$$

$i = 1, 2, \cdots, a$; $j = 1, 2, \cdots, b$; $k = 1, 2, \cdots, c$, where $\alpha_{(i)}$ and $\beta_{(j)}$ are fixed unknown parameters representing the *ith* level and *jth* level of factors A and B, respectively, and $\gamma_{i(k)}$ and $\epsilon_{i(jk)}$ are distributed independently as $N(0, \sigma^2_{\gamma(\alpha)})$ and $N(0, \sigma^2_\epsilon)$, respectively.

(a) Derive a test statistic for testing each of the following hypotheses:

(i) $H_0 : \alpha_{(i)} = 0$ for all i.

(ii) $H_0 : \beta_{(j)} = 0$ for all j.

Please specify the distribution and degrees fo freedom of the test statistic under H_0 in each case.

(b) Show how to obtain an exact confidence interval on $\theta = \sigma^2_{\gamma(\alpha)} + \sigma^2_\epsilon$, with a confidence coefficient greater than or equal to 0.95.

(c) Suppose that the interaction A*B is significant and therefore it is necessary to test each factor at fixed levels of the other factor.

(i) Give a test statistic that can be used to compare the means of two levels of A (say i and i') at the fixed level j of B.

(ii) Give a test statistic that can be used to compare the means of two levels of B (say j and j') at the fixed level i of A.

In each case, please describe the null distribution of the test statistic

(d) What is the *best linear unbiased estimator* of the least-squares mean for level i of A?

5. Let X_i $(i = 1, \cdots, m)$ and Y_j $(j = 1, \cdots, n)$ be mutually independent with the X_i iid $\mathrm{N}(\theta_1, \sigma_1^2)$, and the Y_j iid $\mathrm{N}(\theta_2, \sigma_2^2)$, where $\theta_k \in (-\infty, \infty)$ $(k = 1, 2)$ are unknown. Consider estimation of $\Delta = \theta_2 - \theta_1$ under squared error loss. Define $\bar{X} = m^{-1} \sum_{i=1}^{m} X_i$ and $\bar{Y} = n^{-1} \sum_{j=1}^{n} Y_j$.

(a) Assume $\sigma_k^2 (> 0)$ $(k = 1, 2)$ to be known. Consider independent $\mathrm{N}(0, \tau_1^2)$ and $\mathrm{N}(0, \tau_2^2)$ priors for θ_1 and θ_2 respectively. Show that the Bayes estimator of Δ is given by $(1 - B_2)\bar{Y} - (1 - B_1)\bar{X}$, where $B_1 = \sigma_1^2/(\sigma_1^2 + m\tau_1^2)$ and $B_2 = \sigma_2^2/(\sigma_2^2 + n\tau_2^2)$.

(b) If σ_k^2 $(k = 1, 2)$ are known, show that $\bar{Y} - \bar{X}$ is a minimax estimator of Δ.

(c) Suppose now σ_k^2 $(k = 1, 2)$ are unknown, but $\sigma_k^2 \le A_k$ $(k = 1, 2)$, where $A_k (> 0)$ $(k = 1, 2)$ are known. Show that $\bar{Y} - \bar{X}$ continues to be a minimax estimator of Δ.

6. (a) Let $X_1, \cdots, X_n$ be iid from the distribution with pdf given by $f_\mu(x) = (2\pi x^3)^{-1/2} \exp[-(x - \mu)^2/(2\mu^2 x)]$, where $x > 0$ and $\mu > 0$. Show that the family of pdf's has monotone likelihood ratio in $\sum_{i=1}^{n} X_i$.

(b) Consider a discrete random variable X assuming values 1, 2 and 3 with pf's p_0 and p_1 under H_0 and H_1 given by

x	1	2	3
$p_0(x)$	.0001	.0500	.9499
$p_1(x)$	.0100	.1000	.8900

(i) Find the MP size .05 test for testing H_0 against H_1.

(ii) The three possible nonrandomized tests at level (NOT size) .05 are given by (1) never reject, (2) reject iff $x = 1$ and (3) reject iff $x = 2$. Find the optimal among these three tests.

(iii) Comment on the result found in (ii).

7. Suppose that $\{X_n, n \ge 1\}$ is a sequence of identically distributed random variables with finite mean.

(a) Prove that

$$\frac{1}{n} \max_{1 \le j \le n} |X_j| \to 0 \text{ a.s.}$$

(b) Prove that

$$\lim_{n \to \infty} E\left[\frac{1}{n} \max_{1 \le j \le n} |X_j|\right] = 0.$$

8. Let X_n, $n \ge 1$, be random variables with respective characteristic functions ϕ_n, $n \ge 1$. Suppose that $\sup_{n \ge 1} E[g(X_n)] < \infty$ for some nonnegative function g satisfying $g(x) \to \infty$ as $x \to \infty$ and $g(x) \to \infty$ as $x \to -\infty$. Suppose further that $\phi_n(t) \to h(t)$ for some complex-valued function h. Prove that the sequence of random variables $\{X_n, n \ge 1\}$ converges in distribution.