Department of Statistics
University of Florida

PhD Qualifying Exam

January 2008

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 near the upper left corner of the page where the pages will be stapled together.

1. (a) Suppose X has pdf $f(x|\theta)$. Consider the mixture prior $\pi(\theta) = \sum_{i=1}^{k} w_i \pi_i(\theta)$, where $\pi_i(\theta)$ are themselves pdf's and $w_i \geq 0$, $\sum_{i=1}^{k} w_i = 1$.

(i) Find the posterior $\pi(\theta|x)$ explicitly as a weighted average of the component posteriors $\pi_i(\theta|x)$.

(ii) Find also $E_\pi(\theta|x)$ and $V_\pi(\theta|x)$ in terms of $E_{\pi_i}(\theta|x)$ and $V_{\pi_i}(\theta|x)$, $i = 1, \cdots, k$.

(b) Prove or give a counterexample to the following statements:

(i) A minimax decision rule is always Bayes with respect to some proper prior.

(ii) An admissible decision rule with constant risk is minimax.

(iii) If $C_1, \cdots, C_k$ are all complete, then $C_1 \bigcap \cdots \bigcap C_k$ is essentially complete.

(iv) An admissible decision rule is always Bayes with respect to some proper prior.

2. Let $X_1, \cdots, X_n, Y_1, \cdots, Y_n$ be mutually independent where X_i is exponential with mean σ/θ_i and Y_i is exponential with mean $\sigma\theta_i$, $i = 1, \cdots, n$. In the above, $\theta_1, \cdots, \theta_n, \sigma$ are all unknown.

(a) Write down the likelihood function $L(\theta_1, \cdots, \theta_n, \sigma)$.

(b) Show that the MLE $\hat{\sigma}_n$ of σ is given by $\hat{\sigma}_n = n^{-1} \sum_{i=1}^{n} (X_i Y_i)^{1/2}$.

(c) Show that $\hat{\sigma}_n \to (\pi/4)\sigma$ in probability as $n \to \infty$.

(d) Give an intuitive explanation of the result in (c).

3. Consider the balanced fixed-effects one-way model,

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

where $\alpha_{(i)}$ is a fixed unknown parameter ($i = 1, 2, \ldots, k$), $\epsilon_{i(j)} \sim N(0, \sigma_\epsilon^2)$, and the $\epsilon_{i(j)}$'s are mutually independent. Let $\text{SS}_{\text{treat}} = n \sum_{i=1}^{k} (\bar{y}_{i.} - \bar{y}_{..})^2$ be the treatment sum of squares.

(a) Express SS_{treat} as a quadratic form in $\bar{\mathbf{y}}$, the vector of treatment sample means for the k treatments.

(b) Partition SS_{treat} into $k-1$ independent sums of squares each with one degree of freedom. What distribution does each sum of squares have? Please be specific.

(c) Deduce that the one-degree-of-freedom sums of squares in part (b) represent sums of squares of orthogonal contrasts among the true means of the treatments.

(d) Obtain $(1-\alpha)100\%$ simultaneous confidence intervals on the $k-1$ orthogonal contrasts in part (c) using Scheffé's procedure. What can you say about the actual joint coverage probability for these $k-1$ confidence intervals?

(e) If the F-test concerning the treatment effect is significant at the α-level, does it necessarily follow that every single confidence interval in part (d) must not contain zero? Why or why not?

4. Consider the linear model

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

$i = 1, 2, \ldots, a;\ j = 1, 2, \ldots, b;\ k = 1, 2, \ldots, c;\ l = 1, 2, \ldots, n.$

The effect $\alpha_{(i)}$ is fixed, but the remaining effects are independently distributed as normal random variables with zero means and variances given by σ^2_β, $\sigma^2_{\gamma(\beta)}$, $\sigma^2_{\alpha\beta}$, $\sigma^2_{\alpha\gamma(\beta)}$, σ^2_ϵ, respectively.

(a) Give the corresponding population structure. Then, indicate what the subscripts are for the $(\alpha\gamma)$ effect in the model, and point out its rightmost-bracket subscripts.

(b) Write down the expected mean squares for all the effects in the corresponding ANOVA table. What distributions do the sums of squares have in this ANOVA table?

(c) Give an expression for the power function of the F-test concerning the hypothesis $H_0 : \alpha_{(i)} = 0$ for all i.

(d) Let $\hat\sigma^2_{\gamma(\beta)}$ denote the ANOVA estimator of $\sigma^2_{\gamma(\beta)}$. Give an expression that can be used to compute the probability $P(\hat\sigma^2_{\gamma(\beta)} < 0)$. What parameter values must be specified in order to compute this probability?

(e) Let $\mathbf{g} = (\mu, \alpha_{(1)}, \alpha_{(2)}, \ldots, \alpha_{(a)})'$, and let $\boldsymbol{\lambda}'\mathbf{g}$ be an estimable linear function of $\mathbf{g}$. What is the B.L.U.E. of $\boldsymbol{\lambda}'\mathbf{g}$? Give also an expression for its variance.

5. Let $\{X_n, n \geq 1\}$ be a sequence of i.i.d. mean 0 random variables and set $S_n = \sum_{j=1}^n X_j,\ n \geq 1$. Prove that $E|S_n| = o(n)$.

6. Let$\{X_n, n \geq 1\}$ be a sequence of independent random variables with

$$EX_n = 0,\ 0 < EX_n^2 = \sigma_n^2 < \infty,\ n \geq 1.$$

Set

$$S_n = \sum_{j=1}^n X_j \quad \text{and} \quad s_n^2 = \sum_{j=1}^n \sigma_j^2,\ n \geq 1.$$

Prove that if

$$s_n^2 \to \infty,\ \sigma_n^2 = o(s_n^2),$$

and

$$\frac{S_n}{s_n} \xrightarrow{d} N(0, 1),$$

then

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

7. This problem concerns the inverse Gaussian distribution, which has cumulative distribution function (CDF)

$$F(y) = \begin{cases} 0, & y \le 0, \\ \Phi\Big(\sqrt{\frac{\lambda}{y}}\Big(-1+\frac{y}{\mu}\Big)\Big) + e^{2\lambda/\mu}\,\Phi\Big(-\sqrt{\frac{\lambda}{y}}\Big(1+\frac{y}{\mu}\Big)\Big), & y > 0, \end{cases}$$

where Φ denotes the standard normal CDF.

(a) Show that F has density f given by

$$f(y) = \begin{cases} 0, & y \le 0, \\ \Big(\frac{\lambda}{2\pi y^3}\Big)^{1/2} \exp\Big\{-\frac{\lambda(y-\mu)^2}{2\mu^2 y}\Big\}, & y > 0. \end{cases}$$

(b) Show that the density f can be written in exponential dispersion form. Identify: the canonical parameter θ and the dispersion parameter ϕ (in terms of λ and μ); the cumulant function, $b(\theta)$; the variance function, $V(\mu)$; and the canonical link for this distribution.

(c) Find the form of the deviance $D(\boldsymbol{y}, \hat{\boldsymbol{\mu}})$ for a GLM when the random component (i.e., the distribution of the responses) is inverse Gaussian.

8. (a) Suppose that $\boldsymbol{Y} = (Y_1, \ldots, Y_k)^T$ is a multinomial vector of counts based on m trials with probability vector $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_k)^T$, i.e., $\boldsymbol{Y} \sim \mathrm{MN}_k(m, \boldsymbol{\pi})$. Suppose further that π depends on some parameter vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)^T$. Show that the likelihood equations for $\boldsymbol{\theta}$ have the form

$$\sum_{j=1}^{k} \frac{y_j - m\pi_j}{\pi_j} \frac{\partial \pi_j}{\theta_l} = 0, \quad l = 1, \ldots, p.$$

(b) Now let $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n \sim \text{indep}\,\mathrm{MN}_k(m_i, \boldsymbol{\pi}_i)$, with

$$\pi_{ij} = \frac{\exp(\boldsymbol{x}_{ij}^T\boldsymbol{\beta})}{\sum_{r=1}^{k} \exp(\boldsymbol{x}_{ir}^T\boldsymbol{\beta})},$$

where $\boldsymbol{x}_{ij}$ is a vector of known covariates associated with each count. Write down the log-likelihood function for the parameter $\boldsymbol{\beta}$. Show that there exists a Poisson loglinear GLM for which (frequentist) likelihood inference concerning $\boldsymbol{\beta}$ is identical to that based on this multinomial model.