Department of Statistics
University of Florida

PhD Qualifying Exam

January 8, 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. Consider the gamma density,

$$f(y;\lambda,\beta) = \frac{\beta^\lambda}{\Gamma(\lambda)} y^{\lambda-1} e^{-\beta y}\,, \quad y > 0\,,$$

where $\lambda > 0$ and $\beta > 0$.

(a) Show that this density can be written in exponential dispersion form.

(b) Identify the dispersion and canonical parameters, ϕ and θ respectively, in terms of λ and β.

(c) Identify the cumulant function, $b(\theta)$, and hence derive the canonical link and variance functions for a gamma GLM.

(d) Derive the deviance function for a gamma GLM.

(e) Show that the scaled deviance is approximately the square of the standardized gamma variable if ϕ is small.

2. Suppose that a population of individuals is partitioned into k sub-populations or groups, $G_1,\ldots,G_k$, with relative frequencies, $\pi_1,\ldots,\pi_k$. Multivariate measurements made on individuals in group j are distributed as $N_p(\boldsymbol{\mu}_j,\boldsymbol{\Sigma})$, for $j = 1,\ldots,k$. Let $\mathbf{x}$ be an observation made on an individual drawn at random from the combined population. The prior odds that the individual belongs to G_j rather than G_1 is π_j/π_1, $j = 2,\ldots,p$.

(a) Show that the corresponding (log) posterior odds are of the form

$$\log\frac{\pi_j(\mathbf{x})}{\pi_1(\mathbf{x})} = \log\frac{\pi_j}{\pi_1} + \alpha_j + \boldsymbol{\beta}_j^t\mathbf{x}\,,$$

and find expressions for α_j and $\boldsymbol{\beta}_j$ in terms of $\boldsymbol{\mu}_j$ and $\boldsymbol{\Sigma}$.

(b) What rule would you use to classify $\mathbf{x}$; i.e. how would decide which group the measurement $\mathbf{x}$ came from?

(c) What simplifications can be made if the k normal means, $\boldsymbol{\mu}_j$, lie in a straight line in R^p?

(d) Consider the case in which the measurements are univariate (i.e. $p = 1$). Suppose, as before, that conditional on being in group j the measurements are distributed as $N(\mu_j,\sigma_j^2)$. (So both means and variances are allowed to differ among groups.) Show that this assumption leads to a quadratic classification rule.

3. Let $\text{MS}_1, \text{MS}_2, \ldots, \text{MS}_k$ be independently distributed mean squares in an ANOVA table for a data vector $\boldsymbol{y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ such that $\nu_i \text{MS}_i/\delta_i \sim \chi^2_{\nu_i}, i = 1, 2, \ldots, k$. Consider the linear combination

$$\sum_{i=1}^{k} c_i \text{MS}_i = \boldsymbol{y}'\boldsymbol{A}\boldsymbol{y},$$

where $c_i \geq 0$ for $i = 1, 2, \ldots, k$, and $\boldsymbol{A}$ is a matrix of rank r.

(a) Find the values of θ and ν in terms of the δ_i such that $\boldsymbol{y}'\boldsymbol{A}\boldsymbol{y}$ is approximately distributed as $\theta\chi^2_\nu$.

(b) Let $\tau_1, \tau_2, \ldots, \tau_r$ be the nonzero eigenvalues of $\boldsymbol{A}\boldsymbol{\Sigma}$. Show that $\Delta \geq 1$, where

$$\Delta = \frac{r\sum_{i=1}^{r} \tau_i^2}{(\sum_{i=1}^{r} \tau_i)^2}.$$

(c) Show that the approximation in (a) is exact, i.e., $\boldsymbol{y}'\boldsymbol{A}\boldsymbol{y} \sim \theta\chi^2_\nu$ if and only if $\Delta = 1$.

(d) Define the function $\lambda(\boldsymbol{\tau})$ as

$$\lambda(\boldsymbol{\tau}) = \Delta,$$

where $\boldsymbol{\tau} = (\tau_1, \tau_2, \ldots, \tau_r)'$. Suppose that τ_1 and τ_r are the largest and smallest of the the τ_i, respectively. Write $\lambda(\boldsymbol{\tau})$ as

$$\lambda(\boldsymbol{\tau}) = \frac{r(1 + p^2 + \sum_{i=1}^{r-2} p_i^2)}{(1 + p + \sum_{i=1}^{r-2} p_i)^2},$$

where $p = \tau_r/\tau_1$, $p_i = \tau_{i+1}/\tau_1, i = 1, 2, \ldots, r-2$. For a fixed p, let $\lambda_{\max}(p)$ be the supremum of $\lambda(\boldsymbol{\tau})$ over a region R determined by $p \leq p_i \leq 1, i = 1, 2, \ldots, r-2$.

(i) Show that $\lambda_{\max}(p)$ is a monotone decreasing function of p $(0 < p \leq 1)$.

(ii) Show that the approximation in (a) is exact if and only if $\lim_{p \to 0} \lambda_{\max}(p) = 1$.

4. Consider the simple linear regression model

$$y(x) = \beta_0 + \beta_1 x + \epsilon$$

Let $\eta = \beta_0 + \beta_1 x$ be the mean response at x. The model is fitted to a data set consisting of n $(n > 4)$ observations. The usual assumptions of normality, independence, and equality of the error variances are assumed to be valid.

(a) Let x_0 be the value of x at which $\eta = 0$. Show how to obtain a $(1-\alpha)100\%$ confidence interval on x_0. Please give all the necessary details.

(b) Consider the ratio $\hat{\psi} = \bar{y}^2/\text{MS}_E$, where $\bar{y}$ is the average of all n observations, and MS_E is the error mean square. Show how to obtain the expected value of $\hat{\psi}$. Assume that the error variance is equal to one.

5. (a) Let $X_1, \cdots, X_n$ be iid with mean θ and finite second moment. Define $\bar{X} = n^{-1}\sum_{i=1}^{n} X_i$. Assuming squared error loss, show that $a\bar{X} + b$ (a and b being constants) is an inadmissible estimator of θ if (i) $a > 1$, (ii) $a < 0$, and (iii) $a = 1, b \neq 0$.

(b) Let $X_1, \cdots, X_n$ be iid $\mathrm{N}(\theta, 1)$, where $\theta \in (-\infty, \infty)$. Define $\bar{X} = n^{-1}\sum_{i=1}^{n} X_i$. Assume squared error loss.

(i) Show that a linear estimator $a\bar{X} + b$, where $a \in (0, 1)$ and b (real) are known constants, is a Bayes estimator of θ with respect to the $\mathrm{N}(b/(1-a), a/(n(1-a)))$ prior with finite Bayes risk a/n.

(ii) Show that every real valued constant b is an admissible estimator of θ.

(iii) Show that $\bar{X}$ is an admissible estimator of θ.

(iv) Characterize the class of all admissible linear estimators of the form $a\bar{X} + b$ of θ.

6. 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) Use delta method to find the asymptotic distribution of $\hat{\theta}_n$ after suitable normalization.

(c) Find a UMP size α test of $H_0 : \theta \leq \theta_0$ against $H_1 : \theta > \theta_0$.

(d) Find asymptotically (as $n \to \infty$) the explicit form of the critical region of the test in (c).

7. Let $X_1, X_2, \ldots$ be independent random variables. Show that $\sup_{n\geq 1} X_n < \infty$ a.s. if and only if there exists a finite constant M such that $\sum_{n=1}^{\infty} P(X_n > M) < \infty$. Be sure to specify where the hypothesis of independence is needed in your proof (or alternatively, state clearly that the independence assumption is not needed if your proof does not require it).

8. Let $\{X_n, n \geq 1\}$ be independent random variables. Show that for any $\lambda > 0$ and $0 < c < 1$ we have

$$P\Big(\max_{1\leq k\leq n} |S_k| \geq \lambda\Big) \times \min_{1\leq j\leq n} P\Big(|S_n - S_j| \leq (1-c)\lambda\Big) \leq P\Big(|S_n| \geq c\lambda\Big),$$

or equivalently

$$P\Big(\max_{1\leq k\leq n} |S_k| \geq \lambda\Big) \leq \frac{P\big(|S_n| \geq c\lambda\big)}{1 - \max_{1\leq j\leq n} P\big(|S_n - S_j| > (1-c)\lambda\big)}.$$

Hint: Consider events of the form $\big[|S_j| < \lambda, 1 \leq j \leq k-1, |S_k| \geq \lambda, |S_n - S_k| < (1-c)\lambda\big]$.