PhD Qualifying Examination
Department of Statistics, University of Florida

August 23, 2002, 8:00 am - 12:00 noon

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.

The following abbreviations are used throughout:

- ANOVA = analysis of variance
- GLM = generalized linear model
- iid = independent and identically distributed
- ML = maximum likelihood
- UMP = uniformly most powerful

1. Let $\{X_n, n \geq 1\}$ be a sequence of random variables such that for some $\beta \in (0, \infty)$,

$$\sup_{n \geq 1} E|X_n|^\beta < \infty \; .$$

(a) Prove that for all $\alpha \in (0, \beta)$ the sequence of random variables $\{|X_n|^\alpha, n \geq 1\}$ is uniformly integrable.

(b) Demonstrate by example that the sequence of random variables $\{|X_n|^\beta, \; n \geq 1\}$ is not necessarily uniformly integrable.

2. Let $S_n = \sum_{i=1}^n X_i, \; n \geq 1$ where $\{X_n, n \geq 1\}$ is a sequence of independent random variables and let $\{b_n, n \geq 1\}$ be a sequence of positive constants with $b_n \uparrow \infty$.

(a) Prove that if

$$\frac{S_n}{b_n} \to 0 \quad \text{almost certainly,}$$

then

$$\frac{\max_{1 \leq j \leq n} |S_j|}{b_n} \to 0 \quad \text{almost certainly.}$$

(b) Prove that if

$$\frac{S_n}{b_n} \xrightarrow{P} 0,$$

then

$$\frac{\max_{1 \leq j \leq n} |S_j|}{b_n} \xrightarrow{P} 0.$$

3. Consider the linear model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

where $\mathbf{X}$ is $n \times p$ of rank p ($p < n$), $\boldsymbol{\beta}$ is a fixed unknown parameter vector, and $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$. Let SS_E be the residual sum of squares.

(a) Find the maximum likelihood estimate of $\boldsymbol{\beta}$ given that $\mathbf{A}\boldsymbol{\beta} = \mathbf{m}$, where $\mathbf{A}$ is a known matrix of order $s \times p$ and rank s ($s \leq p$), and $\mathbf{m}$ is a known vector.

(b) Find SS_E^c, the residual sum of squares under the condition $\mathbf{A}\boldsymbol{\beta} = \mathbf{m}$.

(c) Find the probability that $SS_E^c \geq SS_E + 5$, given that $\sigma^2 = 1$ and $\mathbf{A}\boldsymbol{\beta} = \mathbf{m}$.

(d) What distribution does $(SS_E^c - SS_E)/SS_E$ have given that $\mathbf{A}\boldsymbol{\beta} = \mathbf{m}$?

(e) Are $(SS_E^c - SS_E)$ and SS_E distributed independently? Why or why not?

4. Consider the balanced, two-fold nested model

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

where $i = 1, 2, \ldots, a$, $j = 1, 2, \ldots, b$, $k = 1, 2, \ldots, n$, the α_i are fixed, and β_{ij} and ϵ_{ijk} are distributed independently as $N\left(0, \sigma^2_\beta\right)$ and $N\left(0, \sigma^2_\epsilon\right)$, respectively.

(a) Show that $\phi = \sum_{i=1}^a \lambda_i \alpha_i$ is estimable if and only if ϕ is a contrast; that is, $\sum_{i=1}^a \lambda_i = 0$.

(b) Find $100(1-\alpha)$% simultaneous confidence intervals for all $\phi = \sum_{i=1}^a \lambda_i \alpha_i$ such that ϕ is a contrast.

(c) Develop an exact test statistic for testing the hypothesis $H_0 : \sigma^2_\beta = 2\,\sigma^2_\epsilon$ versus $H_a : \sigma^2_\beta \neq 2\,\sigma^2_\epsilon$. Also, draw the corresponding rejection region for an α-level of significance.

(d) Let h^2 be defined as

$$h^2 = \frac{\sigma^2_\beta}{\sigma^2_\beta + \sigma^2_\epsilon}\ .$$

Let $\hat{h}^2$ be its ANOVA estimator. Show how to obtain the expected value of $\hat{h}^2$ given that $\sigma^2_\beta = 5\,\sigma^2_\epsilon$.

5. (a) Write down a complete definition of a GLM for a set of responses, $Y_1, \ldots, Y_n$. Define all the components of the model, such as link function, linear predictor, variance function, dispersion parameter, etc.

(b) Derive the maximum likelihood estimating equations for the regression parameters of a GLM.

(c) Show that the Fisher scoring algorithm for finding the ML estimates of the regression parameters is equivalent to iteratively reweighted least squares (IRLS).

(d) Explain how you would calculate a GLM version of Cook's D based on the final iteration of IRLS.

6. Consider the gamma density

$$f(y; \lambda, \beta) = \begin{cases} \frac{\beta^\lambda}{\Gamma(\lambda)}\, y^{\lambda-1}\, e^{-\beta y} & y > 0 \\ 0 & \text{otherwise} \end{cases}$$

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.

7. (a) State the Neyman-Pearson Lemma. (Hint: Recall that there are three parts - existence, sufficiency, and necessity.)

(b) Let $X_1, \dots, X_n$ be iid N(μ, σ^2). Suppose that $\mu = \mu_0$ is known. Using **only** the Neyman-Pearson Lemma, show that there exists a UMP test for testing $H : \sigma \leq \sigma_0$ against $K : \sigma > \sigma_0$, which rejects when $\sum_{i=1}^n (X_i - \mu_0)^2$ is too large.

8. Let X be a discrete random variable with support $\mathcal{X}$ and mass function $P_\theta(X = x)$ which depends on the unknown parameter $\theta \in \Theta$. Consider using the observed value of X to estimate θ under the loss function $L(\cdot, \cdot)$. As usual, let the risk function of an estimator $\delta : \mathcal{X} \mapsto \Theta$ be the expected loss; that is,

$$R(\theta, \delta) = \sum_{x \in \mathcal{X}} L\left(\theta, \delta(x)\right) \, P_\theta(X = x) \; .$$

(a) Give the definition of *minimax estimator*.

(b) Let $\pi(\theta)$ be a proper prior density for θ. The *Bayes risk* of the estimator δ with respect to π is defined as $r(\pi, \delta) = \int_\Theta R(\theta, \delta) \, \pi(\theta) \, d\theta$. Let δ_π denote the *Bayes estimator with respect to* π. How is δ_π defined?

(c) Give the definition of *least favorable prior*.

(d) Suppose that, for a particular prior π^*,

$$r(\pi^*, \delta_{\pi^*}) = \sup_{\theta \in \Theta} R(\theta, \delta_{\pi^*}) \; .$$

Show that

1. The estimator δ_{π^*} is minimax.
2. If δ_{π^*} is the unique Bayes estimator with respect to π^*, then δ_{π^*} is the unique minimax estimator.
3. The prior π^* is least favorable.

(e) Suppose $X \sim$ Geometric(θ); that is,

$$P_\theta(X = x) = \theta(1 - \theta)^x$$

for $x \in \{0, 1, 2, \dots\}$ and $\theta \in (0, 1)$. Note that E$(X|\theta) = \frac{1}{\theta} - 1$ and Var$(X|\theta) = \frac{1-\theta}{\theta^2}$. Consider a prior for θ that puts positive probability on only two values of θ. Specifically, consider the prior given by $P(\theta = 1/4) = 2/3$ and $P(\theta = 1) = 1/3$. Show that the Bayes estimator of θ with respect to this prior under squared error loss is minimax.