Department of Statistics
University of Florida

PhD Qualifying Exam

January 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. (a) Let $X_1, \cdots, X_n$ be iid Bin$(1, \theta)$, where $\theta \in [0, 1]$, and let $\bar{X} = n^{-1} \sum_{i=1}^{n} X_i$. Assume squared error loss. Show that $\delta(X_1, \cdots, X_n) = \frac{n\bar{X}+\alpha}{n+\alpha+\beta}$ is the Bayes estimator of θ under the Beta(α, β) prior.

(b) Show that the Bayes estimator of θ found in (a) is an admissible estimator under squared error loss.

(c) Let $X_1, \cdots, X_p$ be independently distributed with pdf's $f_{\theta_i}(x_i)$. Let $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_p)^T$. Writing $\mathbf{a} = (a_1, \cdots, a_p)^T$, suppose the loss in estimating $\boldsymbol{\theta}$ by $\mathbf{a}$ is $L(\boldsymbol{\theta}, \mathbf{a}) = \sum_{i=1}^{p} L(\theta_i, a_i)$. Consider the prior $\pi(\boldsymbol{\theta}) = \prod_{i=1}^{p} \pi_i(\theta_i)$ for $\boldsymbol{\theta}$, where the π_i are themselves pdf's. Show that the Bayes estimator of $\boldsymbol{\theta}$ is given by $\boldsymbol{\delta}^{\pi}(X) = (\delta_1^{\pi}(X_1), \cdots, \delta_p^{\pi}(X_p))^T$, where $\delta_i^{\pi}(X_i)$ is the Bayes estimator of θ_i under the prior π_i.

(d) Let $X_1, \cdots, X_p$ be independently distributed Binomial (n_i, θ_i), $i = 1, \cdots, p$ random variables. Assume the loss $L(\boldsymbol{\theta}, \mathbf{a}) = \sum_{i=1}^{p} (\theta_i - a_i)^2$. Find the prior under which the estimator $((X_1 + \alpha_1)/(n_1 + 2\alpha_1), \cdots, (X_p + \alpha_p)/(n_p + 2\alpha_p))^T$ is the Bayes estimator of $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_p)^T$.

2. (a) Let $X_1, \cdots, X_n$ be iid with common pdf $f_\theta(x) = (2x/\theta^3) I_{[0 \le x \le \theta]}$, where $\theta (> 0)$ is unknown, and I is the usual indicator function. Find the generalized likelihood ratio test of $H_0 : \theta = \theta_0$ against the alternatives $H_0 : \theta \neq \theta_0$. Show also that the critical region of this test can be determined exactly using percentiles of a chisquare distribution.

(b) If gene frequencies are in equilibrium, the genotypes AA, Aa and aa occur in a population with relative frequencies $(1 - \theta)^2$, $2\theta(1 - \theta)$ and θ^2, where $0 \le \theta \le 1$. Let X_1, X_2 and X_3 denote the corresponding observed counts in a random sample of size n. Find the MLE of θ, and find also its asymptotic distribution.

3. Suppose that $\mathbf{y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

(a) Show that the quadratic form $\mathbf{y}'\mathbf{A}\mathbf{y}$ can be represented as

$$\mathbf{y}'\mathbf{A}\mathbf{y} = \sum_{i=1}^{k} \lambda_i W_i,$$

where the W_i's are independently distributed as noncentral chi-squared variates with m_i degrees of freedom and noncentrality parameter θ_i, that is, $W_i \sim \chi'^2_{m_i}(\theta_i)$, $i = 1, 2, \ldots, k$. Indicate what λ_i, m_i, θ_i are equal to.

(b) Use part (a) to derive the moment generating function of $\mathbf{y}'\mathbf{A}\mathbf{y}$.

(c) Let $\phi(t)$ denote the moment generating function of $\mathbf{y}'\mathbf{A}\mathbf{y}$ found in part (b). Show that $\phi(t)$ exists in a small neighborhood of $t = 0$, that is, for $|t| \le t_0$ for some positive constant t_0.

(d) Use part (a) to show that if $\mathbf{A}\boldsymbol{\Sigma}$ is idempotent, then $\mathbf{y}'\mathbf{A}\mathbf{y}$ has the noncentral chi-squared distribution. Determine its degrees of freedom and noncentrality parameter.

(e) Show that if $\mathrm{tr}[(\mathbf{A}\boldsymbol{\Sigma})^2] = \mathrm{tr}(\mathbf{A}\boldsymbol{\Sigma}) = r$, where r is the rank of $\mathbf{A}$, then $\mathbf{y}'\mathbf{A}\mathbf{y}$ has the chi-squared distribution.

4. Consider the quadratic forms, $Q_1 = \mathbf{y}'\mathbf{A}\mathbf{y}$, $Q_2 = \mathbf{y}'\mathbf{B}\mathbf{y}$, where $\mathbf{y} \sim N(\mathbf{0}, \mathbf{\Sigma})$ and $\mathbf{A}$ and $\mathbf{B}$ are nonnegative definite matrices of order $n \times n$.

(a) Find the covariance of $\mathbf{y}'\mathbf{A}\mathbf{y}$ and $\mathbf{y}'\mathbf{B}\mathbf{y}$.

(b) Show that if Q_1 and Q_2 are uncorrelated, then they are also independent.

(c) Show that $E(\mathbf{y}'\mathbf{A}\mathbf{y}) \leq \lambda_{\max} \sum_{i=1}^{n} \sigma_{ii}$, where $\lambda_{\max}$ is the largest eigenvalue of $\mathbf{A}$ and σ_{ii} is the i^{th} diagonal element of $\mathbf{\Sigma}$.

(d) If $\mathbf{\Sigma}$ is known, can you compute the exact probability $P(\mathbf{y}'\mathbf{A}\mathbf{y} > \mathbf{y}'\mathbf{B}\mathbf{y})$? Please explain.

5. Consider $Y_i \sim \text{Ber}(\pi_i)$ (indep), $i = 1, \ldots, n$, where

$$\operatorname{logit} \pi_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1} X_{i2}$$

where X_{i1} and X_{i2} are binary covariates.

(a) Derive the mle for $\beta = (\beta_0, \beta_1, \beta_2, \beta_3)$ in closed form. *Hint: you may need to introduce some additional notation to do this.*

(b) Interpret the coefficient β_3.

(c) Suppose we replace the logit link with a probit link,

$$\Phi^{-1}(\pi_i) = \beta_0^\star + \beta_1^\star X_{i1} + \beta_2^\star X_{i2} + \beta_3^\star X_{i1} X_{i2}$$

where Φ^{-1} is the inverse of the cdf of a normal random variable. Derive the mle for $\beta^\star$ in closed form.

(d) Compare the mle's under the probit and the logit link. Will any relationship you see between the mle's hold in general? Explain.

(e) Suppose that we replace the scalar random variables Y_i with T-dimensional vectors $Y_i^\star = (Y_{i1}, \ldots, Y_{iT})^T$ (e.g., repeated measurements on unit i). To account for correlation among the components on Y_i, we introduce a random effect into the probit formulation as follows:

$$\Phi^{-1}(\pi_{ij}(b_i)) = \beta_0^\star + b_i + \beta_1^\star X_{i1} + \beta_2^\star X_{i2} + \beta_3^\star X_{i1} X_{i2} \tag{1}$$

$$b_i \sim N(0, \tau^2) \tag{2}$$

where $\pi_{ij}(b_i) = E[Y_{ij} \mid b_i]$. So, the conditional (on the random effect) mean follows a probit link. Derive the marginal mean, $E[Y_{ij}]$ and identify the link function. Is there a relationship between the β coefficients in (1) and the corresponding coefficients in the model for $E[Y_{ij}]$?

(f) Suppose we want to test for a missing covariate X_{i3}, but we don't want to fit the more complex model involving X_{i3},

$$\operatorname{logit} \pi_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1} X_{i2} + \beta_4 X_{i3}. \tag{3}$$

Derive a test for $H_0 : \beta_4 = 0$ which does not require fitting the full model in (3).

6. Suppose $Y_i|\pi_i \sim \text{Poisson}(e_i\lambda_i)$ and $\lambda_i \sim \text{Gamma}(\alpha, \beta)$ for $i = 1, \ldots, n$, where e_i is a constant. Note: The Gamma density is given by

$$f(\lambda_i) = \frac{1}{\Gamma(\alpha)\beta^\alpha}\lambda_i^{\alpha-1}\exp(-\lambda_i/\beta) \tag{4}$$

where $E[\lambda_i] = \alpha\beta$ and $\text{Var}[\lambda_i] = \alpha\beta^2$.

(a) Derive the marginal distribution of Y_i.

(b) Derive the mean and variance. In particular, show that the mean can be written in the form $E[Y_i] = e_i\lambda$ and that the variance can be written in the following form, $\text{Var}(Y_i) = e_i\lambda h(e_i, \alpha, \beta)$. *Hint: λ will be a function of the Gamma distribution parameters, α and β.*

(c) Comment on the form of the overdispersion induced by this gamma mixture of Poissons.

(d) Suppose we now index (α, β) by i and re-parameterize them as $\alpha_i = e_i\delta m$ and $\beta_i = \frac{1}{\delta e_i}$. Derive the variance under this parameterization and compare it to the one in (b).

(e) Given the parameterization in (d), now replace m with m_i and model it as $\log m_i = X_i\gamma$. Suppose we have a function that fits regular Poisson regression models and from it we obtain an estimate of γ, $\hat{\gamma}$. Given that we have an estimate for γ, propose a simple moment based estimator of δ. *Hint: This will be a function of $\hat{\gamma}$.*

7. Let $X_1, X_2, \ldots$ be independent random variables with

$$X_n = \begin{cases} \pm n^2, & \text{with probability } \dfrac{1}{12n^2} \text{ each,} \\ \pm n, & \text{with probability } \dfrac{1}{12} \text{ each,} \\ 0, & \text{with probability } 1 - \dfrac{1}{6} - \dfrac{1}{6n^2}, \end{cases}$$

and let $S_n = \sum_{k=1}^n X_k$. Prove that

$$\frac{S_n}{\sqrt{n^3/18}} \rightsquigarrow N(0, 1),$$

where $\rightsquigarrow$ denotes convergence in distribution. *Hint: Consider $Y_n = X_n I_{\{|X_n| \le n\}}$, $n \ge 1$. Recall that $\sum_{k=1}^n k^2 = n(n+1)(2n+1)/6$.*

8. Suppose that $\{X_n, n \ge 1\}$ is a sequence of nonnegative random variables, and let

$$M_n = \max_{1 \le k \le n} X_k.$$

(a) Prove that

$$E\big(M_n I_{\{M_n > \alpha\}}\big) \le \sum_{k=1}^n E\big(X_k I_{\{X_k > \alpha\}}\big).$$

(b) Prove that if $\{X_n, n \ge 1\}$ is uniformly integrable, then

$$\frac{E(M_n)}{n} \to 0.$$

(c) Give an example to show that the result of part (b) may fail without the hypothesis of uniform integrability.