Department of Statistics
University of Florida

PhD Qualifying Exam

August 2006

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_i $(i = 1, \cdots, m)$ and Y_j $(j = 1, \cdots, n)$ be mutually independent with the X_i iid $N(\xi, \sigma^2)$, and the Y_j iid $N(\eta, \tau^2)$, where ξ and η are unknown. Consider the problem of estimating $\Delta = \eta - \xi$ under squared error loss. Assuming independent $\mathrm{N}(\mu_1, \zeta_1)$ and $\mathrm{N}(\mu_2, \zeta_2)$ priors for ξ and η, find the Bayes estimator of Δ.

(b) Consider the set up of part (a) where the problem is once again estimation of $\Delta = \eta - \xi$ under squared error loss.

(i) If σ^2 and τ^2 are known, show that $\bar{Y} - \bar{X}$ is a minimax estimator of Δ.

(ii) If σ^2 and τ^2 are unknown, but $\sigma^2 \leq A$ and $\tau^2 \leq B$, where A and B are known, show that $\bar{Y} - \bar{X}$ continues to be a minimax estimator of Δ.

2. (a) Let $X_1, \cdots, X_n$ and $Y_1, \cdots, Y_n$ constitute random samples independently drawn from the two normal populations $N(\mu_1, 1)$ and $N(\mu_2, 1)$. The Fieller-Creasy problem involves inference about the ratio $\theta = \mu_2/\mu_1$, where we assume that $\mu_2 \in (-\infty, \infty)$ and $\mu_1 \in (-\infty, \infty) - \{0\}$.

(i) Find the MLE of θ.

(ii) Find the asymptotic distribution of the MLE of θ.

(b) A random sample $X_1, \cdots, X_n$ $(n \geq 2)$ is drawn from a Power distribution with pdf

$$f_{\theta,\nu}(x) = (\theta x^{\theta-1}/\nu^{\theta}) I_{[0<x\leq\nu]},$$

where $\theta (> 0)$ is unknown, and $\nu > 0$. Consider testing $H_0 : \theta = 1$ against $H_1 : \theta \neq 1$.

(i) Show that when ν is known, the critical region of the GLRT can be determined from the percentiles of a chi-square distribution.

(ii) Find the form of the GLRT for unknown ν.

3. Let $\{Y_n : n \geq 1\}$, $\{Z_n : n \geq 1\}$, and $\{W_n : n \geq 1\}$ be independent sequences of random variables where:

$Y_1, Y_2, \ldots$ are independent and identically distributed with $E(Y_1) = 0$ and $\mathrm{Var}(Y_1) = 1$;

$Z_1, Z_2, \ldots$ are independent with $E(|Z_n|) = \infty$ for all $n \geq 1$; and

$W_1, W_2, \ldots$ are independent Bernoullis with $P(W_n = 1) = n^{-2} = 1 - P(W_n = 0)$.

Let $X_n = (1 - W_n)Y_n + W_n Z_n$, $n \geq 1$.

(a) Prove that $E(|X_n|) = \infty$. (1 point)

(b) Prove that $n^{-1/2} \sum_{k=1}^{n} X_k \rightsquigarrow N(0, 1)$, where $\rightsquigarrow$ denotes convergence in distribution. (9 points)

The moral of the story: the central limit theorem can hold even when the summands have no finite moments.

4. Let $(\Omega, \mathscr{F}, P)$ be a probability space, and suppose that $\mathscr{G}_1$, $\mathscr{G}_2$, and $\mathscr{G}_3$ are sub-σ-fields of $\mathscr{F}$. Let $\mathscr{G}_i \vee \mathscr{G}_j := \sigma(\mathscr{G}_i, \mathscr{G}_j)$ represent the smallest σ-field containing both $\mathscr{G}_i$ and $\mathscr{G}_j$. Show that if the random variable Y is integrable and $\mathscr{G}_1$ measurable and the σ-fields $\mathscr{G}_1 \vee \mathscr{G}_2$ and $\mathscr{G}_3$ are independent, then $E(Y|\mathscr{G}_2 \vee \mathscr{G}_3) = E(Y|\mathscr{G}_2)$.

Hint: Show that $E(Y|\mathscr{G}_2)$ is a version of $E(Y|\mathscr{G}_2 \vee \mathscr{G}_3)$. In doing so, consider sets of the form $A = B \cap C$, $B \in \mathscr{G}_2$, $C \in \mathscr{G}_3$, and note that this class of sets forms a π-system.

5. Consider $Y_{ji} \sim \text{Ber}(\pi_{ji})$ (indep) for $j = 1, 2$ and $i = 1, \ldots, n$, where $\text{logit}\,\pi_{1i} = \lambda_i + \Delta$ and $\text{logit}\,\pi_{2i} = \lambda_i$. Suppose the parameter of interest is the odds ratio, Δ',

$$\Delta' = \frac{\pi_{1i}}{1-\pi_{1i}} \Big/ \frac{\pi_{2i}}{1-\pi_{2i}} = \exp(\Delta)$$

.

(a) Find sufficient statistics for the nuisance parameters, λ_i, $i = 1, \ldots, n$.

(b) Derive the conditional likelihood for Δ' and the mle for Δ' based on this conditional likelihood (available in closed form).

(c) Describe how to construct an exact (i.e., non-large sample) confidence interval for Δ'.

(d) Based on the conditional likelihood, derive the score test of $H_0 : \Delta' = 1$. (This will take a simple form).

6. Suppose $Y_i|\pi_i \sim \text{Bin}(m_i, \pi_i)$ and $\pi_i \sim \text{Beta}(\alpha, \beta)$ for $i = 1, \ldots, n$.

(a) Derive the marginal distribution of Y_i (this is the beta-binomial distribution).

(b) Derive the mean and variance. In particular, show that the mean can be written in the form $E[Y_i] = m_i\pi$ and that the variance can be written in the following form, $m_i\pi(1-\pi)h(m_i, \alpha, \beta)$. (Hint: π will be a function of the beta parameters, α and β).

(c) Comment on the form of the overdispersion induced by this beta mixture of binomials.

(d) Given this structure of the marginal mean and variance of the beta-binomial distribution, propose a way to enter covariates into this model (in the spirit of glm's). Clearly this will now make the parameters of the beta distribution indexed by i, but in what way?

(e) Using quasi-likelihood as the basis for estimation, overdispersion in a binomial is modelled as $\text{Var}(Y_i) = \sigma^2 m_i\pi_i(1-\pi_i)$. Suppose we have available an estimator of π_i, $\hat{\pi}_i$. Given this estimator for π_i, suggest a simple moment-type estimator for σ^2.

(f) Suppose an investigator wants to choose between the parametric beta-binomial distribution and the quasi-likelihood approach (which also accounts for overdispersion) for his/her data. Assume the same mean model is used for both approaches. Suggest a simple graphical check to determine which approach is more appropriate for his/her data.

7. Consider the full-rank linear model,

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

where $\boldsymbol{X}$ is $n \times p$ of rank p $(n > p)$, $E(\boldsymbol{\epsilon}) = \mathbf{0}$, and $\mathrm{Var}(\boldsymbol{\epsilon}) = \boldsymbol{\Sigma}$.

(a) What is *the best linear unbiased estimator* (BLUE) of $\boldsymbol{\beta}$?

(b) Show that the BLUE of $\boldsymbol{\beta}$ is equal to $(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$ if and only if there exits a nonsingular matrix $\boldsymbol{F}$ such that $\boldsymbol{\Sigma}\boldsymbol{X} = \boldsymbol{X}\boldsymbol{F}$.

(c) Show that $(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$ is the BLUE of $\boldsymbol{\beta}$ if and only if

$$\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}') = \mathbf{0}.$$

(d) Let s^2 be defined as

$$s^2 = \frac{1}{n-p}\boldsymbol{y}'[\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']\boldsymbol{y}$$

Show that

$$E(s^2) \le \frac{1}{n-p}\sum_{i=1}^{n}\sigma_{ii}$$

where σ_{ii} is the i^{th} diagonal element of $\boldsymbol{\Sigma}$ $(i = 1, \ldots, n)$. Can this upper bound on $E(s^2)$ be attained?

8. The weights at birth of male lambs were recorded by an animal scientist. Each lamb was the progeny (offspring) of one of several rams (male sheep) that came from five distinct population lines, and each lamb had a different dam (a female sheep). The age of the dam was recorded as belonging to one of three categories, numbered 1 (1-2 years), 2 (between 2 and 3 years), and 3 (over 3 years). Let y_{ijkl} denote the weight at birth of the l^{th} lamb that is the offspring of the k^{th} sire (ram) in the j^{th} population line and of a dam belonging to the i^{th} age category. A possible model for y_{ijkl} is the balanced model

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

where $\alpha_{(i)}$ denotes the i^{th} age effect ($i = 1, 2, 3$), $\beta_{(j)}$, denotes the j^{th} line effect ($j = 1, 2, \ldots, 5$), $\delta_{j(k)}$ denotes the effect of the k^{th} sire ($k = 1, 2, 3, 4$) within the j^{th} line, $(\alpha\beta)_{(ij)}$ is the interaction effect associated with the age and line effects, and $\epsilon_{ijk(l)}$ is a random error ($l = 1, 2, 3$). The age and line effects are fixed, and the sire effect is random. It is assumed that the $\delta_{j(k)}$'s are independently distributed as $N(0, \sigma^2_{\delta(\beta)})$ and the $\epsilon_{ijk(l)}$'s are distributed as $N(0, \sigma^2_\epsilon)$ independently of each other and of the $\delta_{j(k)}$'s.

(a) Give appropriate test statistics for testing the significance of the

  (i) age effect.

  (ii) line efect.

  (iii) sire effect

  Give the degrees of freedom for each test statistic.

(b) Suppose that it is desired to compare the means of lines 1 and 2 for a fixed age, say age category 1, where such means are obtained by averaging over subscripts k and l. Give a t test statistic that can be used to compare these two means at the α level of significance. Also, specify the number of degrees of freedom for this test statistic.

(c) Consider the *heritability* function

$$h^2 = \frac{4\sigma^2_{\delta(\beta)}}{\sigma^2_{\delta(\beta)} + \sigma^2_\epsilon}$$

  Obtain an exact $(1 - \alpha)100\%$ confidence interval on h^2.

(d) Show how you can obtain an exact, but conservative, confidence interval with a confidence coefficient greater than or equal to $1 - \alpha$ for the function

$$g^2 = \exp(-\sigma^2_{\delta(\beta)} - \sigma^2_\epsilon) + \sigma^2_\epsilon$$

  Please give the necessary details.