Department of Statistics
University of Florida

PhD Qualifying Exam

January 2009

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 in the form SN-x, where x is your number.
5. Do not write your name anywhere on your exam.
6. Write only on one side each sheet of paper. For each problem you do, start the problem on a new page. At the end of the exam, for each problem, staple together all pages for that problem in order.
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. Consider a standard linear model

$$Y_{n\times 1} = X_{n\times p}\beta_{p\times 1} + \epsilon_{n\times 1}, \quad \epsilon \sim N_n(0, \sigma^2 I),$$

where X is of full rank. Let $\hat{\beta}$ be the usual least squares estimate of β.

(a) Suppose that σ is known. Show that the likelihood function is proportional to

$$\exp\left(-\frac{1}{2\sigma^2}\|X(\hat{\beta} - \beta)\|^2\right).$$

(b) Suppose now that σ is unknown. Show that the maximum likelihood estimate of the variance is given by

$$\hat{\sigma}^2_{\text{mle}} = \frac{\|Y - \hat{Y}\|^2}{n},$$

where $\hat{Y}$ is the vector of fitted values, and deduce that

$$E(\hat{\sigma}^2_{\text{mle}}) < \sigma^2.$$

(c) Construct a confidence interval for σ.

2. Suppose the pairs (Y_i, X_i), $i = 1, \dots, n$ are iid realization of a random vector (Y, X) with a distribution π on $\mathbb{R}^2$, and that Y and X both have a finite second moment. We wish to estimate $\mu = E(Y)$. Suppose we know $E(X)$, and without loss of generality, we can assume that $E(X) = 0$. For any $\beta \in \mathbb{R}$, the random variable $Y - \beta X$ has expectation μ, and hence

$$\bar{Y} - \beta \bar{X} \tag{1}$$

is an unbiased estimate of μ.

(a) Find the value of β, call it β_{opt}, for which the estimate (1) has smallest variance.

(b) Show that, if the correlation between Y and X is not 0, then if we use β_{opt}, the variance of estimate (1) is strictly smaller than the variance of $\bar{Y}$.

(c) In general β_{opt} is not known, and must be estimated. Let $\hat{\alpha}$ and $\hat{\beta}$ be the usual estimates of the coefficients when we do simple linear regression of Y on X. Note that we do not assume that a linear regression model of the form

$$Y_i = \alpha + \beta X_i + \epsilon_i, \quad i = 1, \dots, n \qquad \epsilon_i \overset{\text{iid}}{\sim} N(0, \sigma^2)$$

is true: we do not assume that the errors are homoscedastic, that they are normally distributed, or even that they have mean 0. In fact we do not assume anything except that the pairs (Y_i, X_i) are iid from π. The values $\hat{\alpha}$ and $\hat{\beta}$ are just the usual expressions that arise when we do simple linear regression; they are simply functions of the data (Y_i, X_i), $i = 1 \dots, n$.

Show that $\hat{\alpha}$ is asymptotically equivalent to (1) with the optimal β, i.e. show that as $n \to \infty$, the two quantities $n^{1/2}(\hat{\alpha} - \mu)$ and $n^{1/2}(\bar{Y} - \beta_{\text{opt}} \bar{X} - \mu)$ have the same limiting distribution.

3. Prove the following: if $\{X_n : n \geq 1\}$ is uniformly integrable, then so is $\{S_n/n : n \geq 1\}$, where $S_n = \sum_{j=1}^{n} X_j$.

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

$$P(X_n = 1) = P(X_n = -1) = \frac{1}{4}, \qquad P(X_n = n) = P(X_n = -n) = \frac{1}{4n^2},$$

and

$$P(X_n = 0) = \frac{1}{2}\Big(1 - \frac{1}{n^2}\Big).$$

Note that $E(X_n) = 0$ and $\text{Var}(X_n) = 1$.

(a) Show that the triangular array $\{X_{nj} : 1 \leq j \leq n, n \geq 1\}$, with $X_{nj} = X_j/\sqrt{n}$, $1 \leq j \leq n$, $n \geq 1$, does not satisfy the Lindeberg condition.

(b) Show that nevertheless there exists a $\sigma^2 > 0$ such that

$$\frac{\sum_{j=1}^{n} X_j}{\sqrt{n}} \rightsquigarrow N(0, \sigma^2),$$

where $\rightsquigarrow$ denotes convergence in distribution.

5. Recall that distributions in an exponential dispersion family have densities of the form

$$f(y;\theta,\phi) = \exp\left\{\frac{1}{\phi}\big[\theta y - b(\theta)\big] + c(y;\phi)\right\}, \qquad (*)$$

for $\theta \in \Theta \subset \mathbb{R}$ and $1/\phi \in \Lambda \subset \mathbb{R}$. In a generalized linear model (GLM), we assume that $Y_1, \ldots, Y_n$ are independent and that Y_i has density $f(y;\theta_i,\phi_i)$, where f has the form $(*)$ and $\phi_i = \phi/w_i$, where $w_1, \ldots, w_n$ are known postive weights. This defines the random, or stochastic part of the GLM.

(a) Suppose that Y has density given by $(*)$. Show that $E(Y) = b'(\theta)$ and $\mathrm{Var}(Y) = \phi\, b''(\theta)$.

(b) Describe the systematic part of the GLM relating the mean $\mu_i = E(Y_i)$ to the vector $\boldsymbol{x}_i$ of predictor values for the ith observation.

(c) Derive the maximum likelihood estimating equations (score equations) for the vector of regression coefficients $\boldsymbol{\beta}$. Show that their solution does not depend on the value of ϕ.

(d) Derive the "expected" Fisher information matrix for $\boldsymbol{\beta}$, assuming that ϕ is known and express your result matrix form.

$$\frac{\partial^2 l_i}{\partial\beta_j\partial\beta_k} = \frac{\partial}{\partial\beta_k}\frac{\partial l_i}{\partial\beta_j} = \frac{\partial}{\partial\beta_k}\left(\frac{1}{\phi}\frac{w_i(y_i-\mu_i)}{V(\mu_i)g'(\mu_i)}\,x_{ij}\right) = \frac{\partial}{\partial\mu_i}\left(\frac{1}{\phi}\frac{w_i(y_i-\mu_i)}{V(\mu_i)g'(\mu_i)}\,x_{ij}\right)\cdot\frac{\partial\mu_i}{\partial\beta_k}$$

$$\frac{\partial}{\partial\mu_i}\left(\frac{1}{\phi}\frac{w_i(y_i-\mu_i)}{V(\mu_i)g'(\mu_i)}\,x_{ij}\right) = \frac{-1}{\phi}\frac{w_i}{V(\mu_i)g'(\mu_i)}\,x_{ij} + \frac{1}{\phi}\,w_i(y_i-\mu_i)x_{ij}\frac{\partial}{\partial\mu_i}\left(\frac{1}{V(\mu_i)g'(\mu_i)}\right)$$

$$\frac{\partial\mu_i}{\partial\beta_k} = \frac{\partial\mu_i}{\partial\eta_i}\frac{\partial\eta_i}{\partial\beta_k} = \frac{1}{\partial\eta_i/\partial\mu_i}\frac{\partial\eta_i}{\partial\beta_k} = \frac{1}{g'(\mu_i)}\,x_{ik}$$

Because $E(Y_i - \mu_i) = 0$, the second term drops out upon taking expectations and we have

$$E\left\{\frac{-\partial^2 l_i}{\partial\beta_j\partial\beta_k}\right\} = \frac{1}{\phi}\frac{w_i}{V(\mu_i)g'(\mu_i)}\,x_{ij}\cdot\frac{1}{g'(\mu_i)}\,x_{ik} = \frac{1}{\phi}\frac{w_i}{V(\mu_i)g'(\mu_i)^2}\,x_{ij}x_{ik}.$$

Thus the expected Fisher information matrix has (j,k)th element

$$I(\boldsymbol{\beta})_{j,k} = E\left\{\frac{-\partial^2 l}{\partial\beta_j\partial\beta_k}\right\} = \frac{1}{\phi}\sum_{i=1}^{n}\frac{w_i}{V(\mu_i)g'(\mu_i)^2}\,x_{ij}x_{ik}.$$

Letting

$$X = (x_{ij})_{\substack{1\le i\le n\\ 1\le j\le n}} = (\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n)^T \in \mathbb{R}^{n\times p}$$

and

$$D = D(\boldsymbol{\beta}) = \mathrm{diag}\left\{\frac{w_i}{V(\mu_i)g'(\mu_i)^2} : i = 1,\ldots,n\right\}$$

we have

$$I(\boldsymbol{\beta}) = \frac{1}{\phi}\,X^T D X.$$

(e) What is the canonical link function for the density given in $(*)$? Show that if ϕ is known and the canonical link is used, then there is a simple sufficient statistic for the vector of regression parameters, $\boldsymbol{\beta}$, and give its form. Show also that the likelihood equations and the observed Fisher information for β can be simplified when the canonical link is used and give the simplified forms.

6. Suppose that Y is a binary response following a probit-normal model, i.e., a generalized linear mixed model (GLMM) with a probit link and linear predictor $\eta = \boldsymbol{x}^T\boldsymbol{\beta} + \boldsymbol{z}^T\boldsymbol{U}$, where $\boldsymbol{x} \in \mathbb{R}^p$ and $\boldsymbol{z} \in \mathbb{R}^q$ are known covariates, $\boldsymbol{\beta} \in \mathbb{R}^p$ is a vector of unknown regression parameters, and $\boldsymbol{U} \sim N_q(0, \Sigma_u)$. Show that the marginal model for Y is a probit GLM and find an expression for the coefficient vector of the marginal GLM in terms of the parameters of the GLMM.

7. Let $X \sim N(\mu, \sigma^2)$, σ^2 known. For each $c \geq 0$ define an *interval estimator* for μ by $C(x) = [x - c\sigma, x + c\sigma]$, and consider the loss function $L(\mu, C) = b \times Length(C) - I(\mu \in C)$.

(a) Show that the risk function of C is given by

$$R(\mu, C) = 2bc\sigma - P(-c \leq Z \leq c),$$

where Z is a standard normal random variable.

(b) Show that for $b\sigma > 1/\sqrt{2\pi}$, the risk function is minimized at $c = 0$, so the best "interval estimator" for this case is the point x.

(c) Show that for $b\sigma \leq 1/\sqrt{2\pi}$, the risk function is minimized at $c = \sqrt{-2\log(\sqrt{2\pi}b\sigma)}$

(d) How does the interval in part (c) compare with the usual $1 - \alpha$ interval?

8. Suppose that we observe $X_{ij} \sim N(\theta_i, \sigma^2)$, $i = 1, \ldots, p$, $j = 1, \ldots, n_i$, where $p \geq 3$, σ^2 is known and the n_i are not necessarily equal. Consider two versions of the Stein estimator

1. $\delta^u = \left(1 - \frac{c\sigma^2}{\sum_i \bar{X}_i}\right)\bar{\mathbf{X}}$,
2. $\delta^w = \left(1 - \frac{c\sigma^2}{\sum_i n_i \bar{X}_i}\right)\bar{\mathbf{X}}$,

where $\bar{X}_i = (1/n_i)\sum_j X_{ij}$ and $\bar{\mathbf{X}}$ is the vector of these means. We use "u" for unweighted and "w" for weighted

(a) Show that δ^u is minimax if $c \leq 2(p-2)$.

(b) Show that δ^w is minimax and, in fact, has smaller risk than δ^u.

(c) If we assume that $\theta_i \sim N(0, \tau^2)$, independent, what is the Bayes estimator of θ_i? What is the marginal distribution of $\bar{X}_i$?

(d) Based on part (c), can you justify either δ^u or δ^w as an empirical Bayes estimator?