Yazhou

Department of Statistics
University of Wisconsin, Madison
PhD Qualifying Exam Part I
Tuesday, August 30, 2011
12:30-4:30pm, Room 133 SMI

- There are a total of FOUR (4) problems in this exam. Please do a total of THREE (3) problems.
- Each problem must be done in a separate exam book.
- Please turn in THREE (3) exam books.
- Please write your code name and NOT your real name on each exam book.

1. Consider random variables Y and Δ, where $\Delta \sim \text{Bernoulli}(p)$, for an unknown parameter $p \in (0,1)$, and

$$
\begin{aligned}
Pr(Y=0 \mid \Delta=1) &= 1, \\
Pr(Y=k \mid \Delta=0) &= \frac{\Gamma(k+\lambda/\theta)}{k!\Gamma(\lambda/\theta)}(1+\theta)^{-\lambda/\theta}(1+1/\theta)^{-k}, \quad k=0,1,2,\ldots,
\end{aligned}
$$

where $\theta > 0$ and $\lambda > 0$ are unknown parameters.

(a) Derive the marginal distribution of Y. Find $E(Y)$ and $\text{var}(Y)$.

(b) Derive the marginal distribution of Y when $\theta \to 0$.

For the rest of the question we extend the statistical model to $\theta = 0$ by associating the derived limiting marginal distribution of Y in (b) with $\theta = 0$.

(c) Let $Y_1, \cdots, Y_n$ be independent and identically distributed copies of Y, and T_1 and T_2 be the sample mean and sample variance of Y_i's, respectively. Take $\theta = 0$ and derive a consistent estimator of (λ, p) as an explicit function of T_1 and T_2.

For parts (d) and (e) below assume that we observe n independent observations (Y_i, Δ_i) along with vector valued covariates $\mathbf{x}_i$ and $\mathbf{z}_i$. Consider modeling p and λ based on $\mathbf{z}_i$ and $\mathbf{x}_i$ with the following link functions:

$$
\begin{aligned}
\log(\lambda_i) &= \mathbf{x}_i^t \beta, \\
\log\left(\frac{p_i}{1-p_i}\right) &= \mathbf{z}_i^t \gamma,
\end{aligned}
$$

where β and γ are vectors of dimensions q_1 and q_2, respectively.

(d) Consider the Fisher information matrix of (θ, β, γ) given by

$$
I(\theta, \beta, \gamma) = \begin{bmatrix} I_{\theta\theta} & I_{\theta\beta} & I_{\theta\gamma} \\ I_{\theta\beta}^t & I_{\beta\beta} & I_{\beta\gamma} \\ I_{\theta\gamma}^t & I_{\beta\gamma}^t & I_{\gamma\gamma} \end{bmatrix},
$$

where superscript t denotes matrix transpose (e.g. $I_{\theta\beta}^t$ is the transpose of $I_{\theta\beta}$), $I_{\theta\theta}$ is scalar, and others are matrices of appropriate dimensions. Derive the components $I_{\theta\theta}$, $I_{\theta\beta_j}$, $I_{\theta\gamma_j}$, $I_{\beta_j\beta_k}$, $I_{\beta_j\gamma_k}$, and $I_{\gamma_j\gamma_k}$ of this information matrix.

(e) Derive the score test for testing the null hypothesis $H_0 : \theta = 0$ against the alternative $H_1 : \theta > 0$.

Hint: You may find the following useful result. If X has Negative Binomial distribution with parameters r and p, $r > 0$, $p \in (0,1)$, then $E(X) = \frac{pr}{1-p}$ and $\text{var}(X) = \frac{pr}{(1-p)^2}$. The p.d.f. of X at $X = x$ is given by

$$
\frac{\Gamma(x+r)}{x!\Gamma(r)}(1-p)^r p^x.
$$

2. Let $X_1, ..., X_n$, $n > 1$, be real-valued observations such that $E(X_j) = \mu$, $\text{Var}(X_j) = \sigma^2 > 0$, $j = 1, ..., n$, and the correlation coefficient between X_i and X_j is r_{ij}, $1 \le i < j \le n$. Assume that μ and σ^2 are unknown parameters.

   (a) Assume that r_{ij}'s are known. Derive the best linear unbiased estimator (BLUE) $\hat{\mu}$ of μ.

   (b) Suppose that $r_{ij} = \rho$ for all (i, j), $i \neq j$, but ρ is unknown. Show that $\hat{\mu}$ derived in part (a) is still the BLUE and that $\hat{\mu} = \bar{X}$, the sample mean.

   (c) In part (b), show that $\hat{\mu}$ is actually the uniformly minimum variance unbiased estimator (UMVUE) when $(X_1, ..., X_n)$ is normally distributed.

   (d) Suppose that $(X_1, ..., X_n)$ is normally distributed and that r_{ij}'s are known. Consider testing the following hypotheses:

   $$H_0 : \mu \le 0 \quad \text{versus} \quad H_1 : \mu > 0.$$

   Show that the uniformly most powerful unbiased (UMPU) test of size $\alpha \in (0, 1)$ is

   $$\phi = \begin{cases} 1 & \hat{\mu} > c(U) \\ 0 & \hat{\mu} \le c(U) \end{cases}$$

   where $c(U)$ is determined by

   $$E(\phi|U) = \alpha$$

   for every U and the expectation is taken under $\mu = 0$. Obtain the explicit form of U.

   (e) Assume that $(X_1, ..., X_n)$ is normally distributed and $r_{ij} = \rho$ for all (i, j), $i \neq j$, with a known ρ. Show that the UMPU test of size α in part (d) reduces to

   $$\phi = \begin{cases} 1 & T > c_0 \\ 0 & T \le c_0 \end{cases}$$

   where c_0 is a constant and

   $$T = \sqrt{n}\bar{X}/S, \qquad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

   Find the value of c_0.

3. A system of interest involves three random variables, X, Y, and S, where S has a Poisson distribution with mean 2λ, for a parameter $\lambda > 0$, and where X and Y are conditionally independent Bernoulli variables, given $S = s$, with

$$P(X = 1|S = s) = \frac{1}{2^{s+1}} \quad \text{and} \quad P(Y = 1|S = s) = \frac{\theta}{2^s}$$

for a second parameter $\theta \in (0, 1)$. The random variable S is unobservable.

(a) Calculate the joint distribution for X and Y, and confirm that these two variables are positively correlated.

(b) We have n independent and identically distributed copies (X_i, Y_i) of (X, Y), and with these we aim to test the null hypothesis $H_0 : \theta = 1/2$. A pair (X_i, Y_i) is *discordant* if $X_i \neq Y_i$. Calculate $P(X_i = 1|X_i \neq Y_i)$ generally and under the null hypothesis, and describe a hypothesis test that uses only the discordant pairs when the number M of such pairs is at least one. As θ moves away from $1/2$, what are two sources of increasing power?

(c) Again consider n independent and identically distributed copies (X_i, Y_i) of (X, Y). Here we aim to estimate θ. Show that $\hat{\theta}_n = \bar{Y}_n/(2\bar{X}_n)$ is consistent and asymptotically normal, and compute its asymptotic variance, where $\bar{Y}_n = \frac{1}{n}\sum_{i=1}^n Y_i$ and $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$.

(d) Examine the finite sample bias of estimator $\hat{\lambda}_n$ of λ, where

$$\hat{\lambda}_n = \begin{cases} -\log(2\bar{X}_n) & \text{if } \bar{X}_n > 0 \\ -\log(2/n) & \text{if } \bar{X}_n = 0 \end{cases}$$

Hint: Recall that

$$E[\hat{\lambda}_n] = E[\hat{\lambda}_n|\bar{X}_n > 0]P[\bar{X}_n > 0] + E[\hat{\lambda}_n|\bar{X}_n = 0]P[\bar{X}_n = 0].$$

4. Suppose that we have n subjects. For the ith subject we have two independent continuous random variables T_i and C_i, but they are not completely observable. Rather, we can observe $Y_i = \min\{T_i, C_i\}$ and $\Delta_i = 1(T_i \le C_i)$. We also obtain a fixed vector $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ip})'$ of covariates for subject i. In summary, the observed data are n triplets $(Y_i, \Delta_i, \boldsymbol{x}_i)$, $i = 1, \ldots, n$. Denote the observed values of (Y_i, Δ_i) by (y_i, δ_i), $i = 1, \ldots, n$.

   We assume that C_i's are independent and identically distributed (iid) with a known probability density function (pdf) $g(\cdot)$ which does not depend on $\boldsymbol{x}_i$. Let $f(t|\boldsymbol{x}_i)$ and $F(t|\boldsymbol{x}_i)$ be unknown probability density function (pdf) and cumulative distribution function (CDF) of T_i for given $\boldsymbol{x}_i$, respectively, and set $h(t|\boldsymbol{x}_i) = f(t|\boldsymbol{x}_i)/[1 - F(t|\boldsymbol{x}_i)]$. The distribution of T_i depends on $\boldsymbol{x}_i$ and is specified through $h(t|\boldsymbol{x}_i)$ as follows,

$$h(t|\boldsymbol{x}_i) = h_0(t) \exp\{\boldsymbol{x}_i'\boldsymbol{\beta}\}, \tag{1}$$

   where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ is a vector of p unknown parameters, and $h_0(t)$ is an unknown function that does not depend on i.

   (a) Write down the joint density function for the random variables $\{(Y_i, \Delta_i), i = 1, 2, \ldots, n\}$.

   (b) Assume that T_i has pdf $f(t|\boldsymbol{x}_i) = \lambda_i \exp\{-\lambda_i t\}$ for $t \ge 0$, where $\lambda_i > 0$ is an unknown parameter. Show that the function $h_0(t)$ is constant, and find estimators $\tilde{h}_0(t)$ of $h_0(t)$ and $\tilde{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ by maximizing the joint density function in (a) in those variables, with the data fixed at their observed values.

   For the following questions, both $h_0(t)$ and $\boldsymbol{\beta}$ in (1) are again assumed to be unknown but we no longer assume that the distribution of T_i specified by (1) is necessarily an exponential distribution.

   (c) For any given $\boldsymbol{\beta}$, let $\hat{h}_0(t)$ be a function which maximizes the joint density function in (a) with the data fixed at their observed values. The $\hat{h}_0(t)$ may depend on $\boldsymbol{\beta}$. Let $\mathcal{D}=\{y_i : \delta_i = 1, i = 1, \cdots, n\}$. Show that $\hat{h}_0(t) = 0$ for $t \notin \mathcal{D}$. Also derive $\hat{h}_0(t)$ for $t \in \mathcal{D}$.

   (d) Plugging $\hat{h}_0(t)$ you obtained in (c) into the joint density function in (a) to get a function of $\boldsymbol{\beta}$, find estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ by maximizing the resulting function of $\boldsymbol{\beta}$ with the data fixed at their observed values.