Dept Copy

Department of Statistics
University of Wisconsin, Madison
PhD Qualifying Exam Part I
August 31, 2010
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. Let $X_1, ..., X_n$ be independent and identically distributed with the uniform distribution on the interval (θ_1, θ_2), where $n > 2$ and $-\infty < \theta_1 < \theta_2 < \infty$. Let $X_{(1)}$ and $X_{(n)}$ be the smallest and largest order statistics, respectively.

   (a) Derive the conditional distribution of X_1 given $X_{(n)} = x$.

   (b) Derive the conditional distribution of $X_{(1)}$ given $X_{(n)} = x$.

   (c) Let $\alpha \in (0, 1)$. Derive a uniformly most accurate unbiased (UMAU) upper confidence bound for θ_1 with confidence coefficient $1 - \alpha$.

2. Suppose that $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim}$ a cumulative distribution function F. Consider the use of the chi-square goodness-of-fit test. Let $A_1 = (-\infty, 0]$ and $A_2 = (0, +\infty)$. Define $n_j = \sum_{i=1}^n \mathrm{I}(X_i \in A_j)$. Define $p_j = P(X \in A_j)$, $j = 1, 2$. Define $\chi^2(\mathbf{p}) = \sum_{j=1}^2 \frac{(n_j - np_j)^2}{np_j}$.

(a) Consider the null hypothesis $H_0^{(1)} : \mathbf{p} = \mathbf{p}_0$, where $\mathbf{p}_0 = (p_{0;1}, p_{0;2})^T$ is fully specified. Find the limit distribution of $\chi^2(\mathbf{p})$ under $H_0^{(1)}$.

(b) Consider the null hypothesis $H_0^{(2)} : F = N(\mu, 1)$, where μ is unknown. Let $p_j(\mu)$ denote p_j as a function of μ. Suppose we estimate μ by $\overline{X}$. Define $\widehat{\mathbf{p}} = (\widehat{p}_1, \widehat{p}_2)$, where $\widehat{p}_j = p_j(\overline{X})$. Find the limit distribution of $\chi^2(\widehat{\mathbf{p}})$ under $H_0^{(2)}$.

(c) Find the minimum chi-square test statistic, $\inf_{\mathbf{p} \in \mathcal{P}_0} \chi^2(\mathbf{p})$, where $\mathcal{P}_0 = \{\mathbf{p} = (p_1, p_2) : p_1 \geq 0,\ p_2 \geq 0,\ p_1 + p_2 = 1\}$.

3. Suppose that y_j, $j = 1, \cdots, n$, are independent observations with y_j following binomial distribution $\text{Bin}(m, \pi_j)$, where

$$\log\left(\frac{\pi_j}{1-\pi_j}\right) = \beta_0 + \beta_1 x_j, \tag{1}$$

$x_1 < \cdots < x_n$ are covariates, $\sum x_j/n = 0$, $\sum x_j^2/n = 1$, and β_0 and β_1 are unknown parameters.

(a) Find the MLE of β_0 and β_1.

(b) Compute the Fisher information for β_0 and β_1.

(c) Establish the limiting distribution for the MLE of β_0 and β_1 as $n \to \infty$ assuming appropriate conditions.

(d) Construct an asymptotic level α test with asymptotic power converging to one for testing the following hypothesis under model (1):

$$H_0 \ : \ \pi_1 = \pi_2 = \cdots = \pi_n.$$

4. Suppose we have n i.i.d. observations: $\boldsymbol{y} = (y_1, \ldots, y_n)$ with the density function $g()$ with respect to a reference measure μ.

State all the regularity conditions you need when answering the following questions.

(a) Consider a parametric family $\mathcal{F} = \{f(y|\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta\}$ to approximate $g(\cdot)$. Assume $g(\cdot) \notin \mathcal{F}$. Let $\ell(\boldsymbol{y}|\boldsymbol{\theta})$ be the log likelihood function based on the parametric family $\mathcal{F}$ and $\widehat{\boldsymbol{\theta}}_{\boldsymbol{y}}$ be the MLE obtained as the solution of $\partial\ell(\boldsymbol{y}|\boldsymbol{\theta})/\partial\boldsymbol{\theta} = 0$. When $n \to \infty$, does $\widehat{\boldsymbol{\theta}}_{\boldsymbol{y}}$ converge or diverge? If you think it diverges, prove it. If you think it converges to a $\boldsymbol{\theta}^*$, give the formula for $\boldsymbol{\theta}^*$ and find the asymptotic distribution of $\sqrt{n}(\widehat{\boldsymbol{\theta}}_{\boldsymbol{y}} - \boldsymbol{\theta}^*)$.

(b) Consider K candidate parametric families $\mathcal{F}_k = \{f_k(y|\boldsymbol{\theta}_k), \boldsymbol{\theta}_k \in \Theta_k\}$, $k = 1, \ldots, K$, to approximate $g(\cdot)$. Assume $g(\cdot) \notin \mathcal{F}_k$, $k = 1, \ldots, K$. Prior to observing data, the investigator's belief in the kth family is indexed by π_k. The investigator's prior beliefs about the parameter vector $\boldsymbol{\theta}_k$ are summarized by the prior density $m_k(\boldsymbol{\theta}_k)$. The posterior probability of the kth family is

$$Pr(\mathcal{F}_k|\boldsymbol{y}) = \frac{f_k(\boldsymbol{y})\pi_k}{\sum_{k=1}^{K} f_k(\boldsymbol{y})\pi_k},$$

where $f_k(\boldsymbol{y}) = \int f_k(\boldsymbol{y}|\theta_k)m_k(\boldsymbol{\theta}_k)d\boldsymbol{\theta}_k$.

Let $\ell_k(\boldsymbol{y}|\boldsymbol{\theta}_k)$ denote the log likelihood function based on the parametric family $\mathcal{F}_k$ and $\widehat{\boldsymbol{\theta}}_k$ be the MLE obtained as the solution of $\partial\ell_k(\boldsymbol{y}|\boldsymbol{\theta}_k)/\partial\boldsymbol{\theta}_k = 0$.

Prove that

$$\log[\pi_k f_k(\boldsymbol{y})] = BIC_k + O_p(1),$$

where $BIC_k = \log[f_k(\boldsymbol{y}|\widehat{\boldsymbol{\theta}}_k)] - p_k \log(n)/2$ with $p_k = \dim(\boldsymbol{\theta}_k)$.