Statistics 582, Problem Set 6

Wellner; 2/8/2006

Due: Wednesday, February 15, 2006.
Reminder: midterm exam, Friday February 10.
Reading: Chapter 5, section 8; start reading Chapter 6

1. Suppose that $X \sim P_\theta$ for $\theta \in \Theta \subset R^k$ has well-defined Fisher information matrix $I(\theta)$ for θ. The *Jeffreys prior* distribution Λ_J has density $\lambda_J(\theta) = \det(I(\theta))^{1/2}$ with respect to Lebesgue measure on Θ. Note that Λ_J may not be a finite measure, and even if Λ_J is a finite measure, it may not have total mass 1. If a prior distribution is a finite measure, then call it a *proper prior distribution*, and correspondingly if it is not a finite measure, call it an *improper prior distribution*. If the resulting posterior distribution is a finite measure, call it a *proper posterior distribution*, and (by convention) normalize it to have total mass 1.
   A. Suppose that $X \sim$ Bernoulli(θ). Find the Jeffrey's prior density λ_J for θ. Is Λ_J a finite measure? If it is finite, what is $\Lambda_J((0,1))$? Find the corresponding posterior distribution of Θ starting with the Jeffrey's prior.
   B. Suppose that $X \sim$ Geometric(θ), i.e. the number of trials until the first success in i.i.d. Bernoulli trials with probability θ of success for each trial – recall Chapter 1, section 1. Find the Jeffrey's prior density λ_J for θ. Is Λ_J a finite measure? If it is finite, what is $\Lambda_J((0,1))$? Find the corresponding posterior distribution of Θ starting with the Jeffrey's prior. If we observe $X_1, \ldots, X_n$ i.i.d. Geometric(θ), so that $\sum X_i \sim$ Negative Binomial(n, θ) is the posterior distribution "proper" for some n?
   C. Suppose that $X \sim$ Weibull(θ) with $\theta = (\alpha, \beta) \in (0, \infty) \times (0, \infty)$ as in chapters 3 and 4. Find the Jeffrey's prior density λ_J for θ. Is Λ_J a finite measure? If it is finite, what is $\Lambda_J((0,\infty)^2)$? Find the corresponding posterior distribution of Θ starting with the Jeffrey's prior.

2. Suppose that $(X_1, R_1, Y_1), \ldots, (X_n, R_n, Y_n)$ are i.i.d. with a distribution described as follows. Let $\theta = (\theta_1, \ldots, \theta_B) \in [0,1]^B \equiv \Theta$ where B is large, e.g. 10^{10}. Let $\xi = (\xi_1, \ldots, \xi_B)$ be a vector of *known* numbers with $0 < \delta \le \xi_j \le 1 - \delta < 1$ for $j = 1, \ldots, B$. Furthermore, suppose that:

   (i) $X_i \sim$ Uniform$\{1, \ldots, B\}$.

   (ii) $R_i \sim$ Bernoulli(ξ_{X_i}).

   (iii) If $R_i = 1$, $Y_i \sim$ Bernoulli(θ_{X_i}); if $R_i = 0$, Y_i is missing (i.e. not observed).

Our goal is to estimate

$$\psi = \psi(\theta) = P_\theta(Y_1 = 1) = \sum_{j=1}^{B} P(Y_1 = 1|X_1 = j)P(X_1 = j) = \frac{1}{B}\sum_{j=1}^{B}\theta_j.$$

Now the likelihood contribution of (X_i, R_i, Y_i) is

$$f(X_i)f(R_i|X_i)f(Y_i|X_i, R_i) = \frac{1}{B}\xi_{X_i}^{R_i}(1-\xi_{X_i})^{1-R_i}\theta_{X_i}^{Y_iR_i}(1-\theta_{X_i})^{(1-Y_i)R_i},$$

and hence the likelihood for θ is

$$\begin{aligned} L_n(\theta) &= \prod_{i=1}^{n}\frac{1}{B}\xi_{X_i}^{R_i}(1-\xi_{X_i})^{1-R_i}\theta_{X_i}^{Y_iR_i}(1-\theta_{X_i})^{(1-Y_i)R_i} \\ &\propto \prod_{i=1}^{n}\theta_{X_i}^{Y_iR_i}(1-\theta_{X_i})^{(1-Y_i)R_i}. \end{aligned}$$

Thus

$$\begin{aligned} l_n(\theta) &= \sum_{i=1}^{n}\{Y_iR_i\log\theta_{X_i} + (1-Y_i)R_i\log(1-\theta_{X_i})\} \\ &= \sum_{j=1}^{B} n_j\log\theta_j + \sum_{j=1}^{B} m_j\log(1-\theta_j) \end{aligned}$$

where

$$n_j = \#\{i:\ Y_i = 1,\ R_i = 1, X_i = j\}, \qquad m_j = \#\{i:\ Y_i = 0,\ R_i = 1, X_i = j\}.$$

Note that $n_j = m_j = 0$ for most j since $B >> n$. Thus the MLE for most θ_j is not defined. Furthermore, for most θ_j the posterior distribution is the prior distribution (especially if the prior is a product distribution on $\Theta = [0,1]^B$!). Thus both MLE and Bayes estimation fail.

Here is a purely frequentist solution: the Horovitz- Thompson estimator of ψ is

$$\hat{\psi}_n = \frac{1}{n}\sum_{i=1}^{n}\frac{R_i}{\xi_{X_i}}Y_i.$$

A. Show that

$$E(\hat{\psi}_n) = B^{-1}\sum_{j=1}^{B}\theta_j = \psi(\theta).$$

Thus $\hat{\psi}_n$ is an unbiased estimator of $\psi(\theta)$.
B. Show that the variance of $\hat{\psi}$ is given by

$$Var(\hat{\psi}_n) = \frac{1}{n}\left\{\frac{1}{B}\sum_{j=1}^{B}\frac{\theta_j}{\xi_j} - \psi(\theta)^2\right\}.$$

[Hint: use the formula $Var(Y) = EVar(Y|X) + Var[E(Y|X)]$ twice.]
C. Use B to show that

$$Var(\hat{\psi}_n) \le \frac{1}{n\delta}$$

under the assumption that $\xi_j \ge \delta > 0$ for all $1 \le j \le B$.

3. **Optional bonus problem:** A. Suppose that $X \sim F$, and let $m = F^{-1}(1/2)$, $\mu = E(X)$, $\sigma^2 = Var(X)$, and we assume that $E(X^2) < \infty$. Show that

$$|m - \mu| \le \sqrt{2\sigma^2}\,.$$

Hint: use Chebychev's inequality.
B. Let m and μ be the median and mean of F as in A, and let M be the *mode* of F, assuming that it is well-defined: i.e. we suppose that F has density f which is strictly increasing to the left of M and strictly decreasing to the right of M. Show that $\mu \le m \le M$ if there is an x_0 such that

$$f(m+x) - f(m-x)\begin{cases} \ge 0 & \text{for} \quad 0 \le x < x_0 \\ \le 0 & \text{for} \quad x_0 < x < \infty \end{cases}. \tag{1}$$

If the inequalities in (1) are reversed, then $M \le m \le \mu$. Hint: show that

$$m - \mu = \int_0^\infty \{F(m-x) + F(m+x) - 1\}dx\,.$$

C. Examine (1) and the conclusion for the distributions Gamma$(r, 1)$ with $r = 1, 2$.