Statistics 582, Problem Set 6 Solutions

Wellner; 2/15/2006

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.

Solution: A. When $X \sim$ Bernoulli(θ), the Information for θ is $I(\theta) = \{\theta(1-\theta)\}^{-1}$, so the Jeffrey's prior for θ is has density

$$\lambda_J(\theta) = \frac{1}{\sqrt{\theta(1-\theta)}}\,.$$

This density is proportional to the Beta$(1/2, 1/2)$ density

$$\lambda(\theta) = \frac{\Gamma(1)}{\Gamma(1/2)\Gamma(1/2)}\theta^{1/2-1}(1-\theta)^{1/2-1}1_{(0,1)}(\theta)$$

(which is also known as the "arcsin" distribution because the corresponding distribution function is $\Lambda(\theta) = (2/\pi)\arcsin(\sqrt{\theta})$; it arises naturally as the limiting distribution of the proportion of time a random walk stays positive). Thus

$\lambda_J((0,1)) = \Gamma(1/2)^2 = \pi$. The corresponding posterior distribution is proportional to

$$\theta^{x-1/2}(1-\theta)^{1-x-1/2} = \theta^{(x+1/2)-1}(1-\theta)^{(3/2-x)-1};$$

i.e. the posterior density is Beta$(x+1/2, 3/2-x)$ if $X = x$ is observed.

B. When $X \sim$ Geometric(θ) with $P_\theta(X = x) = (1-\theta)^{x-1}\theta$ for $x = 1, 2, \ldots$, the score function is

$$\dot{l}_\theta(x) = \frac{1}{1-\theta}\left(\frac{1}{\theta} - X\right)$$

and the information for θ is

$$I(\theta) = E_\theta(\dot{l}_\theta(X))^2 = -E_\theta \ddot{l}_{\theta\theta}(X) = \frac{1}{\theta^2(1-\theta)}.$$

Hence the Jeffrey's prior for θ has density

$$\lambda_J(\theta) = \frac{1}{\theta\sqrt{(1-\theta)}}.$$

Here we have

$$\int_0^1 \lambda_J(\theta)d\theta = \int_0^1 \frac{1}{\theta\sqrt{(1-\theta)}}d\theta = \infty$$

because the integral diverges at 0. Hence this density does not correspond to a finite measure. It corresponds in some sense to a Beta$(0, 1/2)$ density; but recall that the Beta(α, β) densities are defined for $\alpha, \beta > 0$. In spite of this the posterior density is proportional to

$$\theta^0(1-\theta)^{x-1-1/2} = (1-\theta)^{x-1/2-1}$$

which corresponds to Beta$(1, x-1/2)$ if $X = x$ is observed. This is completely proper for any $x \in \{1, 2, \ldots\}$ since $x - 1/2 > 0$.

When $X_1, \ldots, X_n$ are i.i.d. Geometric(θ) so that $\sum_1^n X_i \sim$ Negative Binomial(n, θ), then the density of the data is proportional to $\theta^n(1-\theta)^{y-n}$ with $y \geq n$, and since $I_n(\theta) = n/[\theta^2(1-\theta)]$, the posterior is proportional to

$$\theta^{n-1}(1-\theta)^{y-n-1/2} = \theta^{n-1}(1-\theta)^{y-n+1/2-1}$$

which has finite mass for all $y \geq n$. Hence the posterior is proper for all $n \geq 1$.

C. From Example 3.2.7 the information matrix for the Weibull density is

$$I(\theta) = \begin{pmatrix} \frac{\beta^2}{\alpha^2} & \frac{a}{\alpha} \\ \frac{a}{\alpha} & \frac{b^2}{\beta^2} \end{pmatrix}$$

where $a = -(1-\gamma)$ and $b^2 = \pi^2/6 + (1-\gamma)^2$. Therefore the Jeffrey's prior is given by

$$\lambda_J(\theta) = \det(I(\theta))^{1/2} = \frac{\pi/\sqrt{6}}{\alpha}\,.$$

This is not the density of a finite measure:

$$\int_0^\infty \int_0^\infty \lambda_J(\alpha,\beta) d\alpha d\beta = \frac{\pi}{\sqrt{6}} \int_0^\infty \int_0^\infty \frac{1}{\alpha} d\alpha d\beta = \infty\,.$$

The posterior density is proportional to

$$\begin{aligned} p_\theta(x)\lambda_J(\theta) &= \frac{\beta}{\alpha}\left(\frac{x}{\alpha}\right)^{\beta-1} \exp(-(x/\alpha)^\beta)\frac{\pi/\sqrt{6}}{\alpha} 1_{(0,\infty)}(\alpha)1_{(0,\infty)}(\beta) \\ &= \frac{\pi}{\sqrt{6}}\frac{\beta}{\alpha^2}\left(\frac{x}{\alpha}\right)^{\beta-1} \exp(-(x/\alpha)^\beta)1_{(0,\infty)}(\alpha)1_{(0,\infty)}(\beta)\,. \end{aligned}$$

When we try to compute the normalizing constant to form a density (namely the marginal density $p(x)$), we find, however, that (using the change of variables $u = \alpha^{-\beta}$ in the second line so that $du = -\beta\alpha^{-\beta-1}d\alpha$)

$$\begin{aligned} p(x) &= \frac{\pi}{\sqrt{6}} \int_0^\infty \int_0^\infty \frac{\beta}{\alpha^2}\left(\frac{x}{\alpha}\right)^{\beta-1} \exp(-(x/\alpha)^\beta) d\alpha d\beta \\ &= \frac{\pi}{\sqrt{6}} \int_0^\infty \frac{x^{\beta-1}}{x^\beta} \int_0^\infty x^\beta \exp(-x^\beta u) du d\beta \\ &= \frac{\pi}{\sqrt{6}} \int_0^\infty \frac{1}{x} d\beta = \infty\,. \end{aligned}$$

Hence the posterior (for one observation) is not proper for any x.

For n observations we compute

$$p_\theta(\underline{x})\lambda_J(\theta) = \frac{\beta^n}{\alpha^n}\frac{(\prod x_i)^{\beta-1}}{\alpha^{n(\beta-1)}} \exp(-(\sum x_i^\beta/\alpha^\beta)\frac{\pi/\sqrt{6}}{\alpha} 1_{(0,\infty)}(\alpha)1_{(0,\infty)}(\beta)\,.$$

Therefore, using the same change of variables as before, $u = \alpha^{-\beta}$,

$$\begin{aligned} p(\underline{x}) &= \frac{\pi}{\sqrt{6}} \int_0^\infty \beta^n (\prod x_i)^{\beta-1} \int_0^\infty \frac{1}{\alpha^{n+1}}\frac{1}{\alpha^{n(\beta-1)}} \exp(-\sum x_i^\beta/\alpha^\beta) d\alpha d\beta \\ &= \frac{\pi}{\sqrt{6}} \int_0^\infty \beta^{n-1} (\prod x_i)^{\beta-1} \int_0^\infty u^{n-1} \exp(-\sum x_i^\beta u) du d\beta \\ &= \frac{\pi}{\sqrt{6}}\Gamma(n) \int_0^\infty \beta^{n-1} \frac{(\prod x_i)^{\beta-1}}{(\sum x_i^\beta)^n} d\beta = \frac{\pi}{\sqrt{6}}\frac{\Gamma(n)}{\prod x_i} \int_0^\infty \beta^{n-1} \frac{(\prod x_i)^\beta}{(\sum x_i^\beta)^n} d\beta \end{aligned}$$

This converges if the x_i's are not all equal. This can be seen by writing the last integral as follows:

$$\begin{aligned}\int_0^\infty \beta^{n-1}\frac{(\prod x_i)^\beta}{(\sum x_i^\beta)^n}d\beta &= \int_0^\infty \beta^{n-1}\frac{1}{\left(\frac{\sum_1^n x_i^\beta}{(\prod x_i^\beta)^{1/n}}\right)^n}d\beta \\ &= \int_0^\infty \beta^{n-1}\frac{1}{(\sum_1^n y_i^\beta)^n}d\beta \\ &= \int_0^1 \beta^{n-1}\frac{1}{(\sum_1^n y_i^\beta)^n}d\beta + \int_1^\infty \beta^{n-1}\frac{1}{(\sum_1^n y_i^\beta)^n}d\beta.\end{aligned}$$

where $y_i \equiv x_i/(\prod x_i)^{1/n}$, $i = 1, \ldots, n$. But for $\beta \ge 1$ we have

$$\left(\frac{1}{n}\sum_1^n y_i^\beta\right)^{1/\beta} \ge \frac{1}{n}\sum_1^n y_i,$$

by Liapunov's inequality, so

$$\sum_1^n y_i^\beta \ge n\left(\frac{1}{n}\sum_1^n y_i\right)^\beta$$

where $1 = (\prod_1^n y_i)^{1/n} \le n^{-1}\sum_1^n y_i$ with strict inequality if the y_i's are not all equal to 1, or, equivalently, if the x_i's are not all equal. Thus the integrals in the last display are bounded above by

$$\begin{aligned}&\int_0^1 \beta^{n-1}\frac{1}{(ny_{(1)}^\beta)^n}d\beta + \int_1^\infty \beta^{n-1}\frac{1}{n^n\overline{y}^{n\beta}}d\beta \\ &= \int_0^1 \beta^{n-1}\frac{1}{(ny_{(1)}^\beta)^n}d\beta + \int_1^\infty \frac{\beta^{n-1}}{n^n}\exp(-n\beta\log(\overline{y}))d\beta \\ &< \infty\end{aligned}$$

where $\log(\overline{y}) > 0$ since $\overline{y} > 1$ if the x_i's are not all equal. Thus the resulting posterior distribution is proper for $n \ge 2$ as long as all the observations are not all equal.

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

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

(iii) If $R_i = 1$, $Y_i \sim \text{Bernoulli}(\theta_{X_i})$; 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

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

and hence the likelihood for θ is

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

Thus

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

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 $1 \le j \le B$.

Solution: A. Note that by conditioning first on X_1, R_1 and then just on X_1 we get

$$\begin{aligned}
E(\hat{\psi}_n) &= E\{\frac{R_1}{\xi_{X_1}}Y_1\} = EE\{\frac{R_1}{\xi_{X_1}}Y_1|X_1, R_1\} = E\left\{\frac{R_1}{\xi_{X_1}}E\{Y_1|X_1, R_1\}\right\} \\
&= E\left\{\frac{R_1}{\xi_{X_1}}\theta_{X_1}\right\} = EE\left\{\frac{R_1}{\xi_{X_1}}\theta_{X_1}|X_1\right\} = E\left\{\frac{\theta_{X_1}}{\xi_{X_1}}E\{R_1|X_1\}\right\} \\
&= E\left\{\frac{\theta_{X_1}}{\xi_{X_1}}\xi_{X_1}\right\} = E\{\theta_{X_1}\} \\
&= B^{-1}\sum_{j=1}^{B}\theta_j = \psi(\theta)\,.
\end{aligned}$$

B. Since the (X_i, R_i, Y_i)'s are i.i.d.,

$$\begin{aligned}
Var(\hat{\psi}_n) &= n^{-1} Var\left(\frac{R_1 Y_1}{\xi_{X_1}}\right) \\
&= n^{-1}\left\{EVar\left(\frac{R_1 Y_1}{\xi_{X_1}}|R_1, X_1\right) + Var\left(E\left(\frac{R_1 Y_1}{\xi_{X_1}}|R_1, X_1\right)\right)\right\} \\
&= n^{-1}\left\{E\left(\frac{R_1^2}{\xi_{X_1}^2}\theta_{X_1}(1-\theta_{X_1})\right) + Var\left(\frac{R_1}{\xi_{X_1}}\theta_{X_1}\right)\right\} \\
&= n^{-1}\left\{EE\left(\frac{R_1^2}{\xi_{X_1}^2}\theta_{X_1}(1-\theta_{X_1})|X_1\right)\right. \\
&\qquad \left. + EVar\left(\frac{R_1}{\xi_{X_1}}\theta_{X_1}|X_1\right) + Var\left(E\left(\frac{R_1}{\xi_{X_1}}\theta_{X_1}|X_1\right)\right)\right\} \\
&= n^{-1}\left\{E\left(\frac{\theta_{X_1}(1-\theta_{X_1})}{\xi_{X_1}}\right)\right. \\
&\qquad \left. + E\left(\frac{\theta_{X_1}^2}{\xi_{X_1}^2}\xi_{X_1}(1-\xi_{X_1})\right) + Var(\theta_{X_1})\right\} \\
&= n^{-1}\left\{\frac{1}{B}\sum_{j=1}^{B}\frac{\theta_j(1-\theta_j)}{\xi_j} + \frac{1}{B}\sum_{j=1}^{B}\theta_j^2\frac{1-\xi_j}{\xi_j} + \frac{1}{B}\sum_{j=1}^{B}(\theta_j - \overline{\theta})^2\right\} \\
&= n^{-1}\left\{\frac{1}{B}\sum_{j=1}^{B}\frac{\theta_j}{\xi_j} - \psi(\theta)^2\right\}.
\end{aligned}$$

C. Since $\xi_j \geq \delta$ and $\theta_j \leq 1$ for all j, it follows that

$$Var(\hat{\psi}_n) \leq n^{-1}\frac{1}{B}\sum_{j=1}^{B}\frac{1}{\delta} = \frac{1}{n\delta}.$$