Spring 1997 John Rust
Economics 551b 37 Hillhouse, Rm. 27

PROBLEM SET 2: SOLUTIONS

Bayesian Basics (II)

QUESTION 1

Define Kullback-Leibler distance by

$$K(\theta,\theta^*) = E_{\theta^*}\left[\log f(x|\theta) - \log f(x|\theta^*)\right] \leq 0.$$

We know equality holds only when $\theta = \theta^*$.

By definition,

$$\overline{\theta} = \underset{\theta \in \text{supp}(p(\theta))}{\arg\min} \left[-K(\theta,\theta^*)\right].$$

This implies $-K(\overline{\theta},\theta^*) < -K(\theta,\theta^*)$.(Assume uniqueness.)

We want to show convergence of following posterior odds ratio to zero,

$$\frac{p(\theta|x_1,\cdots,x_N)}{p(\overline{\theta}|x_1,\cdots,x_N)} = \frac{f(x_1,\cdots,x_N|\theta)p(\theta)}{f(x_1,\cdots,x_N|\overline{\theta})p(\overline{\theta})} \to 0 \qquad \text{if } \theta \neq \overline{\theta}.$$

Since priors $p(\theta)(\neq 0)$ and $p(\overline{\theta})(\neq 0)$ do not depend on sample size, only we need to show is the convergence of likelihood ratio to zero,

$$\frac{f(x_1,\cdots,x_N|\theta)}{f(x_1,\cdots,x_N|\overline{\theta})} = \frac{\Pi_{i=1}^N f(x_i|\theta)}{\Pi_{i=1}^N f(x_i|\overline{\theta})} \to 0 \qquad \text{if } \theta \neq \overline{\theta}.$$

or the convergence of the log likelihood ratio to $-\infty$,

$$\log\left(\frac{\Pi_{i=1}^N f(x_i|\theta)}{\Pi_{i=1}^N f(x_i|\overline{\theta})}\right) \to -\infty \qquad \text{if } \theta \neq \overline{\theta}.$$

Now,

$$\log\left(\frac{\Pi_{i=1}^N f(x_i|\theta)}{\Pi_{i=1}^N f(x_i|\overline{\theta})}\right) = N\left[\frac{1}{N}\sum_{i=1}^N f(x_i|\theta) - \frac{1}{N}\sum_{i=1}^N f(x_i|\overline{\theta})\right],$$

Since

$$\begin{aligned}
& \frac{1}{N}\sum_{i=1}^N \log f(x_i|\theta) - \frac{1}{N}\sum_{i=1}^N \log f(x_i|\overline{\theta}) \\
= \;& \left[\frac{1}{N}\sum_{i=1}^N \log f(x_i|\theta) - \frac{1}{N}\sum_{i=1}^N \log f(x_i|\theta^*)\right] - \left[\frac{1}{N}\sum_{i=1}^N \log f(x_i|\overline{\theta}) - \frac{1}{N}\sum_{i=1}^N \log f(x_i|\theta^*)\right] \\
\to \;& E_{\theta^*}\left[\log f(x|\theta) - \log f(x|\theta^*)\right] - E_{\theta^*}\left[\log f(x|\overline{\theta}) - \log f(x|\theta^*)\right] \qquad \text{(by LLN)} \\
= \;& \left[-K(\overline{\theta},\theta^*)\right] - \left[-K(\theta,\theta^*)\right] = c < 0.
\end{aligned}$$

Therefore,

$$\log\left(\frac{\Pi_{i=1}^{N} f(x_i|\theta)}{\Pi_{i=1}^{N} f(x_i|\overline{\theta})}\right) \to -\infty.$$

QUESTION 2

By the same argument,

$$\frac{p(\theta|x_1,\cdots,x_N)}{p(\theta^*|x_1,\cdots,x_N)} = \frac{f(x_1,\cdots,x_N|\theta)p(\theta)}{f(x_1,\cdots,x_N|\theta^*)p(\theta^*)} \to 0 \qquad \text{if } \theta \neq \theta^*, \text{ and } \theta \neq \theta^{**}.$$

and

$$\frac{p(\theta|x_1,\cdots,x_N)}{p(\theta^{**}|x_1,\cdots,x_N)} = \frac{f(x_1,\cdots,x_N|\theta)p(\theta)}{f(x_1,\cdots,x_N|\theta^{**})p(\theta^{**})} \to 0 \qquad \text{if } \theta \neq \theta^*, \text{ and } \theta \neq \theta^{**}.$$

Therefore,

$$\begin{aligned} \int p(\theta|x_1,\cdots,x_N)d\theta - [p(\theta^*|x_1,\cdots,x_N) + p(\theta^{**}|x_1,\cdots,x_N)] &\to 0 \\ [p(\theta^*|x_1,\cdots,x_N) + p(\theta^{**}|x_1,\cdots,x_N)] &\to 1 \end{aligned}$$

Now, observational equivalence implies

$$f(x|\theta^*) = f(x|\theta^{**}) \qquad \text{for all } x.$$

In this case, the following posterior odd ratio is identical to the ratio of priors,

$$\begin{aligned} & \frac{p(\theta^*|x_1,\cdots,x_N)}{p(\theta^{**}|x_1,\cdots,x_N)} \\ = & \frac{f(x_1,\cdots,x_N|\theta^*)p(\theta^*)}{f(x_1,\cdots,x_N|\theta^{**})p(\theta^{**})} \\ = & \frac{\Pi_{i=1}^{N} f(x_i|\theta^*)p(\theta^*)}{\Pi_{i=1}^{N} f(x_i|\theta^{**})p(\theta^{**})} \\ = & \frac{p(\theta^*)}{p(\theta^{**})}. \end{aligned}$$

Therefore convergence of posterior distribution can be written as

$$p(\theta|x_1,\cdots,x_N) \to \begin{cases} \delta_{\theta^*} & \text{with probability } \frac{p(\theta^*)}{p(\theta^*)+p(\theta^{**})} \\ \delta_{\theta^{**}} & \text{with probability } \frac{p(\theta^{**})}{p(\theta^*)+p(\theta^{**})} \end{cases}$$

QUESTION 3

See the solution for extra credit problem.

QUESTION 4

Sample output:(See also GAUSS program)
(1) Direct method

$$E(\widehat{\theta}) = \begin{bmatrix} 2.0127 \\ 0.9081 \end{bmatrix} \qquad Cov(\widehat{\theta}) = \begin{bmatrix} 1.1094 & \\ -0.3370 & 0.9797 \end{bmatrix}.$$

(2) Gibbs sampler (500 initial values, 50 iterations)

$$E(\widehat{\theta}) = \begin{bmatrix} 2.0324 \\ 0.9756 \end{bmatrix} \qquad Cov(\widehat{\theta}) = \begin{bmatrix} 0.9716 & \\ -0.2389 & 1.0639 \end{bmatrix}.$$

(3) Gibbs sampler (500 initial values, 1000 iterations)

$$E(\widehat{\theta}) = \begin{bmatrix} 2.0794 \\ 1.0093 \end{bmatrix} \qquad Cov(\widehat{\theta}) = \begin{bmatrix} 0.9842 & \\ -0.2834 & 0.9734 \end{bmatrix}.$$

(4) Gibbs sampler (1 initial value, 2500 iterations)

$$E(\widehat{\theta}) = \begin{bmatrix} 1.9370 \\ 1.0330 \end{bmatrix} \qquad Cov(\widehat{\theta}) = \begin{bmatrix} 0.9466 & \\ -0.3307 & 0.9822 \end{bmatrix}.$$