BAYESIAN ECONOMETRICS
FINAL EXAM
SOLUTION

HEDIBERT FREITAS LOPES
APRIL 7TH 2014

Question 1. (1.0)

Suppose that $x|\theta \sim N(\theta, 1)$ (for example, x is a measurement of a physical constant θ made with an instrument with variance 1). The prior distribution for θ elicited by the scientist A corresponds to a $N(5, 1)$ distribution and the scientist B elicits a $N(15, 1)$ distribution. The value $x = 6$ was observed. What prior fits the data better?

Solution: We should compare the prior predictives, ie. $p(x|H_A)$ and $p(x|H_B)$, where H_A stands for all knowledge gathered by scientist A before x is observed. Similarly for H_B. Since both priors can be written as $N(\mu_s, 1)$, where $s \in \{A, B\}$, then

$$p(x|H_s) \int p_N(x|\theta, 1)p(\theta|\mu_s, 1)d\theta = p_N(x|\mu_s, 2) = \begin{cases} p_N(x|5, 2) & \text{Scientist A} \\ p_N(x|15, 2) & \text{Scientist B} \end{cases}.$$

Hence the Bayes factor, for $x = 6$, is

$$BF_{A,B} = \frac{p_N(6|5, 2)}{p_N(6|15, 2)} = \frac{0.176032700000}{0.000007991871} = 22026.47.$$

The data ($x = 6$) supports Scientist A's prior views substantially more than Scientist B's.

Question 2. (1.0)

Consider the following longitudinal data model:

$$\begin{aligned} y_{ti}|x_{ti}, \alpha_i, \sigma^2 &\sim N(\alpha_i x_{ti}, \sigma^2) \\ \alpha_i|\alpha, \tau^2 &\sim N(\alpha, \tau^2), \end{aligned}$$

where y_{it} refers to the outcomes for individual (or more generally, group) i at time i and α_i is a person-specific random (slope) effect. We assume $i = 1, 2, \ldots, N$ and $t = 1, 2, \ldots, T$ (i.e., a balanced panel). Derive the conditional posterior distribution $p(\alpha_i|\alpha, \sigma^2, \tau^2, y, x)$.

Proof. Let $\gamma = (\alpha, \sigma^2, \tau^2)$, $\texttt{data} = \{y, x\}$, $y_i = (y_{1i}, \ldots, y_{Ti})'$ and $x_i = (x_{1i}, \ldots, x_{Ti})'$. Then,

$$\begin{aligned} p(\alpha_i|\gamma, \texttt{data}) &\propto \exp\{-0.5\sigma^{-2}(y_i - x_i\alpha_i)'(y_i - x_i\alpha_i)\} \\ &\times \exp\{-0.5\tau^{-2}(\alpha_i - \alpha)^2\} \\ &\propto \exp\{-0.5[(\alpha_i^2(x_i'x_i/\sigma^2 + 1/\tau^2) - 2\alpha_i(x_i'y_i/\sigma^2 + \alpha/\tau^2)]\}. \end{aligned}$$

Therefore,

$$(\alpha_i|\gamma, \texttt{data}) \sim N(\mu_i, V_i),$$

for $V_i = 1/(x_i'x_i/\sigma^2 + 1/\tau^2)$ and $\mu_i = V_i(x_i'y_i/\sigma^2 + \alpha/\tau^2)$.

□

Question 3. (1.0)

Let $x|\theta, \mu \sim N(\theta, \sigma^2)$, σ^2 known, $\theta|\mu \sim N(\mu, \tau^2)$, τ^2 known and $\mu \sim N(0, 1)$. Obtain the following distributions:

(a) $(\theta|x, \mu)$

(b) $(\mu|x)$

(c) $(\theta|x)$

Question 4. (1.0)

Count data is a typical kind of discrete data, where the observations y_i are equal to zero or a positive integer; that is, $y_i \in \{0, 1, \ldots, \}$. Such data are often modeled by the Poisson distribution,

$$p(y_i|\theta_i) = \frac{e^{-\theta_i}\theta_i^{y_i}}{y_i!} \qquad \text{where } \theta_i = \exp\{-x_i'\beta\}.$$

Find an expression for the posterior distribution, $p(\beta|y)$, on the assumption that $p(\beta) \equiv N(\beta_0, B_0)$, and discuss possible ways to simulate from this distribution.

Solution: $p(\beta|y) \propto p(\beta)p(y|\beta)$, or

$$\begin{aligned} p(\beta|y) &\propto \exp\{-0.5(\beta' B_0^{-1}\beta - 2\beta' B_0^{-1} b_0)\} \\ &\times \prod_{i=1}^{n} \exp\{-\exp\{-x_i'\beta\}\} \exp\{-y_i x_i'\beta\}. \end{aligned}$$

If dim(β) is small, say up to 3, a simple SIR might be good enough to generate draws from the posterior distribution. In this case, a possible proposal would obtained by using a normal approximation to the Poisson distribution. The approximation will be better when θ_i is large. When dim(β) is moderate or large an MCMC scheme (Metropolis-Hastings algorithm, for example) might be the natural alternative. For more details you can read Gamerman's (1997) *Statistics and Computing* paper entitled "Sampling from the posterior distribution in generalized linear mixed models" (volume 7, pages 57-68).

Question 5. (2.0)

The random variable x has double exponential distribution with parameters μ and σ, denoted by $DE(\mu, \sigma)$, if its density is

$$f(x|\mu, \sigma) = \frac{1}{2\sigma} \exp\left\{-\frac{|x-\mu|}{\sigma}\right\}, \quad \text{for } x \in \mathbb{R} \,.$$

Show that if $x|y \sim N(0, y)$ and $y \sim Exp(1/2)$ then $x \sim DE(0, 1)$.

Proof. The marginal density of x is

$$\begin{aligned} p(x) &= \int_0^\infty p(x|y)p(y)dy \\ &= \int_0^\infty (2\pi y)^{-1/2} \exp\{-0.5x^2/y\} 0.5 \exp\{-0.5y\} dy \\ &= (2\pi)^{-1/2}(0.5) \int_0^\infty y^{-1/2} \exp\{-0.5(y + x^2 y^{-1})\} dy. \end{aligned}$$

Now, let $z = y^{1/2}$ so that $dz = 0.5y^{-1/2}dy$ and

$$p(x) = (2\pi)^{-1/2} \int_0^\infty \exp\{-0.5(z^2 + x^2 z^{-2})\} dz.$$

Andrews and Mallows (1974) show that

$$\int_0^\infty \exp\{-0.5(a^2u^2 + b^2u^{-2})\} du = \left(\frac{\pi}{2a^2}\right)^{1/2} \exp\{-|ab|\}.$$

Therefore,

$$p(x) = \frac{1}{2} \exp\{-|x|\}, \quad \text{for } x \in \mathbb{R}.$$

□

Question 6. (2.0)

Suppose that the density for a time series y_t, $t = 1, 2, \ldots, T$, conditioned on its lags, the model parameters, and other covariates, can be expressed as

$$y_t|\beta_1, \beta_2, \sigma_1^2, \sigma_2^2, \tau, x_t \sim \begin{cases} N(\beta_1 x_t, \sigma_1^2) & t \le \tau \\ N(\beta_2 x_t, \sigma_2^2) & t > \tau \end{cases}$$

In this model, τ is a changepoint: For periods until τ, ie. $t \le \tau$, one regression is assumed to generate y, and following τ, ie. $t > \tau$, a new regression is assumed to generate y. Suppose you employ priors of the form

$$\begin{aligned} \beta_1 &\sim N(b_{01}, B_{01}) \qquad \beta_2 \sim N(b_{02}, B_{02}) \\ \sigma_1^2 &\sim IG(c_{01}, d_{01}) \qquad \sigma_2^2 \sim IG(c_{02}, d_{02}) \\ \tau &\sim U\{1, 2, \ldots, T-1\}. \end{aligned}$$

Note that τ is treated as a parameter of the model, and by placing a uniform prior over the elements $1, 2, \ldots, T-1$ a changepoint is assumed to exist. Describe how the Gibbs sampler can be employed to estimate the parameters of this model.

Solution: Conditioning on τ, the above problem much simpler and two standard Gibbs samplers are combined (for data before and after τ). More precisely, $(\beta_i|\sigma_i^2, \tau, \tilde{y}_i, \tilde{x}_i) \sim N(b_{1i}, B_{1i})$ and $(\sigma_i^2|\beta_i, \tau, \tilde{y}_i, \tilde{x}_i) \sim IG(c_{1i}, d_{1i})$, for $i = 1, 2$, $\tilde{y}_1 = (y_1, \ldots, y_\tau)'$, $\tilde{x}_1 = (x_1, \ldots, x_\tau)'$, $\tilde{y}_2 = (y_{\tau+1}, \ldots, y_T)'$ and $\tilde{x}_2 = (x_{\tau+1}, \ldots, x_T)'$. It is easy to see that $B_{1i}^{-1} = B_{0i}^{-1} + \sigma_i^{-2}\tilde{x}_i'\tilde{x}_i$, $B_{1i}^{-1}b_{1i} = B_{0i}^{-1}b_{0i} + \sigma_i^{-2}\tilde{x}_i'\tilde{y}_i$, $c_{1i} = c_{0i} + n_i/2$ and $d_{1i} = d_{0i} + (\tilde{y}_i - \beta_i\tilde{x}_i)'(\tilde{y}_i - \beta_i\tilde{x}_i)/2$, for $n_1 = \tau$ and $n_2 = T - \tau$. Finally, the full conditional distribution of τ is

$$Pr(\tau|\beta_1, \beta_2, \sigma_1^2, \sigma_2^2, y, x) \propto \prod_{t=1}^{\tau} p_N(y_t; \beta_1 x_t, \sigma_1^2) \prod_{t=\tau+1}^{T} p_N(y_t; \beta_2 x_t, \sigma_2^2),$$

for $\tau \in \{1, 2, \ldots, T-1\}$.

Question 7. (2.0)

Consider the simple regression model

$$y_i^*|\beta, \sigma^2 \sim N(\beta x_i, \sigma^2).$$

Assume that x_i is observed for all $i = 1, 2, \ldots, n$, and that y_i^* is observed (and equal to y_i) for $i = 1, 2, \ldots, m$, where $m < n$. For $i = m+1, m+2, \ldots, n$, the dependent variable is missing, and is presumed to be generated by the above model. The missing data are missing completely at random. Let $N(b_0, B_0)$ and $IG(c, d)$ be the prior distributions of β and σ^2, respectively. Describe a data augmentation approach for ***filling in*** the missing data and fitting the regression model.

Solution: Similar to the previous example, let us assume that we know the true values of $y_{m+1}, y_{m+2}, \ldots, y_n$. In this case, the full conditional distributions of β and σ^2 are normal and inverse gamma, respectively. More precisely, $(\beta|\sigma^2, x, y) \sim N(b_1, B_1)$ and $(\sigma^2|\beta, x, y) \sim IG(c_1, d_1)$, where $B_1^{-1} = B_0^{-1} + \sigma^{-2}x'x$, $B_1^{-1}b_1 = B_0^{-1}b_0 + \sigma^{-2}x'y$, $c_1 = c + n/2$ and $d_1 = d + (y - \beta x)'(y - \beta x)/2$. Sampling the missing becomes straightforward, ie. sample y_i from $N(\beta x_i, \sigma^2)$, for $i = m+1, \ldots, n$. A natural set of initial values for the MCMC are $\hat{\beta}_{OLS}$ and $\hat{\sigma}^2_{OLS}$ based on observations $\{(y_1, x_1), \ldots, (y_m, x_m)\}$. For the missing, the starting values could then be $\hat{y}_i = x_i\hat{\beta}$, for $i = m+1, \ldots, m$.

Useful results

- $x \sim N(\mu, \sigma^2)$ when

$$p(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}, \qquad \text{for } x, \mu \in \mathbb{R}, \sigma^2 > 0.$$

- $x \sim IG(a, b)$ when

$$p(x|a, b) = \frac{b^a}{\Gamma(a)} x^{-(a+1)} e^{-b/x}, \qquad \text{for } x, a, b > 0.$$

- Andrews and Mallows (1974)[1] show that

$$\int_0^\infty \exp\{-0.5(a^2u^2 + b^2u^{-2})\} du = \left(\frac{\pi}{2a^2}\right)^{1/2} \exp\{-|ab|\}.$$

[1] Andrews and Mallows (1974) Scale mixtures of normal distributions. *JRSS-B*, **36**, 99-102. See also, West (1987) On scale mixtures of normal distributions. *Biometrika*, **74**, 646-648.