A Stochastic Sampling Algorithms for Bayesian Computation

We use a Markov chain Monte Carlo algorithm called Gibbs sampling (Gelfand and Smith 1990, Gilks et al. 1996) to draw samples from joint posterior distributions of model parameters. The Gibbs sampler is well suited to the model described in Section 2.2 because conditional posterior distributions of its parameters are relatively easy to sample. For example, when inferences are based on only 1 year of data (as in Sections 2.51 and 2.52), the joint posterior density is formed by taking the product of the likelihood function

\[ f(y_1 | X_1, \beta, \tau, \rho) = \left( \frac{\tau(1 - \rho^2)}{2\pi} \right)^{n/2} \exp \left[ -\frac{\tau d_1(1 - \rho^2)}{2} \right], \quad (A.1) \]

where \( \tau = 1/\sigma^2 \) and \( d_1 = \sum_{i=1}^{n}(y_{i1} - x_{i1}^T \beta)^2 \), and the prior. Assuming mutually independent, Uniform(0,1) priors for each component of \( \beta \), a Uniform(-1,1) prior for \( \rho \), and a conjugate Gamma(\( \varepsilon_1, \varepsilon_2 \)) prior for \( \tau \), the prior density function of model parameters is

\[ \pi(\beta, \tau, \rho) = \left( \frac{1}{1 - 0} \right)^2 \left( \frac{1}{1 + 1} \right) \frac{\varepsilon_1^{\varepsilon_1}}{\Gamma(\varepsilon_1)} \tau^{\varepsilon_1 - 1} \exp(-\tau\varepsilon_2). \quad (A.2) \]

Forming the product of (A.1) and (A.2) and excluding terms without parameters yields the joint posterior density (modulo the normalizing constant)

\[ p(\beta, \tau, \rho | y_1, X_1) \propto \tau^{n/2 + \varepsilon_1 - 1}(1 - \rho^2)^{n/2} \exp \left[ -\tau \left( \varepsilon_2 + \frac{d_1(1 - \rho^2)}{2} \right) \right]. \quad (A.3) \]

Random draws of \( \beta, \tau, \) and \( \rho \) are difficult to compute by sampling (A.3) directly. Gibbs sampling provides a sample from (A.3) by computing random draws from each parameter’s full conditional posterior distribution, which holds the values of other parameters constant. By alternating among the parameters, the Gibbs sampler yields a stochastic sequence (actually a Markov chain) whose stationary distribution is the joint posterior; thus, a sample from the stationary Markov chain is also a sample from (A.3).

The conditional posterior densities needed for Gibbs sampling are readily derived from the joint posterior density. For example, ignoring terms in (A.3) that don’t include \( \tau \) yields the full conditional density for \( \tau \) (modulo its normalizing constant)

\[ p(\tau | \beta, \rho, y_1, X_1) \propto \tau^{n/2 + \varepsilon_1 - 1} \exp \left[ -\tau \left( \varepsilon_2 + \frac{d_1(1 - \rho^2)}{2} \right) \right]. \quad (A.4) \]
This function is proportional to the density function of a Gamma distribution, so random draws from (A.4) are relatively easy to compute as follows:

$$
\tau \mid \beta, \rho, y_1, X_1 \sim \text{Gamma} \left( \varepsilon_1 + n/2, \varepsilon_2 + \frac{d_1(1 - \rho^2)}{2} \right).
$$

The full conditional densities of $\beta$ and $\rho$ are derived from (A.3) in the same way as that of $\tau$:

$$
p(\beta_j \mid \beta_{k(\neq j)}, \tau, \rho, y_1, X_1) \propto \exp \left[ -\frac{\tau d_1(1 - \rho^2)}{2} \right], \quad (A.5)
$$

$$
p(\rho \mid \beta, \tau, y_1, X_1) \propto (1 - \rho^2)^{n/2} \exp \left[ -\frac{\tau d_1(1 - \rho^2)}{2} \right]. \quad (A.6)
$$

These conditional densities do not have a familiar form, but still may be sampled using adaptive-rejection sampling (Gilks 1992, Gilks and Wild 1992) or other algorithms for computing random draws from univariate densities (Carlin and Louis 2000, p. 131–137).

When inferences are based on 2 years of data (as in Section 2.5.3), we form the joint posterior density as before and obtain

$$
p(\beta, \tau, \rho \mid y_1, y_2, X_1, X_2) \propto \tau^{n+\varepsilon_1-1}(1 - \rho^2)^{n/2} \exp \left[ -\tau \left( \varepsilon_2 + \frac{d_1(1 - \rho^2)}{2} + \frac{d_2}{2} \right) \right], \quad (A.7)
$$

where $d_2 = \sum_{i=1}^{n} (y_{i2} - x_{i2}^T \beta - \rho(y_{i1} - x_{i1}^T \beta))^2$. Gibbs sampling may be used to sample (A.7) by computing random draws from the following full-conditional distributions (modulo their normalizing constants):

$$
p(\tau \mid \beta, \rho, y_1, y_2, X_1, X_2) \propto \tau^{n+\varepsilon_1-1} \exp \left[ -\tau \left( \varepsilon_2 + \frac{d_1(1 - \rho^2)}{2} + \frac{d_2}{2} \right) \right], \quad (A.8)
$$

$$
p(\beta_j \mid \beta_{k(\neq j)}, \tau, \rho, y_1, y_2, X_1, X_2) \propto \exp \left[ -\tau \left( \frac{d_1(1 - \rho^2)}{2} + \frac{d_2}{2} \right) \right], \quad (A.9)
$$

$$
p(\rho \mid \beta, \tau, y_1, y_2, X_1, X_2) \propto (1 - \rho^2)^{n/2} \exp \left[ -\tau \left( \frac{d_1(1 - \rho^2)}{2} + \frac{d_2}{2} \right) \right]. \quad (A.10)
$$

Given a sample from the joint posterior distribution of model parameters, the method of composition (Tanner 1996) may be used to compute a sample from the posterior predictive distribution (7) associated with a particular set of management actions; then, Monte Carlo integration may be used to estimate the expected loss (8) associated with this set of management actions. We demonstrate these calculations, which are rather trivial for the autoregressive model, using the example in Section 2.5.3. Suppose Gibbs sampling has been used to compute an arbitrarily large sample from the joint posterior distribution (5), and let $\boldsymbol{\theta}^{(r)} = (\beta^{(r)}, \sigma^{2(r)}, \rho^{(r)})$ denote the $r$th element in this sample. We require a sample of the posterior predictive distribution of vegetation responses ($\tilde{y}_3 \mid \tilde{X}_3, y_1, y_2, X_1, X_2$) associated with the proposed management actions specified in $X_3$. By applying the method of composition to (7), the $r$th element $\tilde{y}_3^{(r)}$ is easily obtained by computing a random draw from the following, $n$-variate normal distribution: $\mathcal{N}(\tilde{X}_3\beta^{(r)} + \rho^{(r)}(y_2 - X_2\beta^{(r)}), \sigma^{2(r)}I)$,
where $I$ is the $n \times n$ identity matrix. The absolute-error loss function used in the example of Section 2.5.3 is $l(\hat{y}_3, 0.5) = \sum_{i=1}^{n} |\hat{y}_{i3} - 0.5|$. To estimate the expected loss associated with the proposed management actions $\hat{X}_3$, we use Monte Carlo integration to average over the posterior uncertainty expressed in the predictions of $\hat{y}_3$:

$$l(\hat{X}_3 | y_1, y_2, X_1, X_2) = \frac{1}{R} \sum_{r=1}^{R} l(\hat{y}_3^{(r)}, 0.5)$$

where $R$ denotes the number of draws computed from the posterior predictive distribution of $\hat{y}_3$.

**Literature Cited**


