S3 Appendix: Posterior computation

for

Estimation of COVID-19 spread curves integrating global data and borrowing information

By SE YOON LEE, BOWEN LEI, and BANI K. MALLICK

Department of Statistics, Texas A&M University, College Station, Texas, 77843, U.S.A.
seyoonlee@stat.tamu.edu bowenlei@stat.tamu.edu bmallick@stat.tamu.edu

S.1 Posterior computation

We illustrate a full description of a posterior computation for the BHRRM (2) – (6) by using a Markov chain Monte Carlo (MCMC) simulation (Robert and Casella, 2013). To start with, for illustrative purpose, we shall use vectorized notations for the likelihood part (2), regression part (3), and its coefficients part (5):

\[
y_i|\theta_1, \theta_2, \theta_3, \xi, \sigma^2 \sim N_T(f(\theta_1, \theta_2, \theta_3, \xi_i), \sigma^2 I), \quad (i = 1, \cdots, N),
\]

\[
\theta_l|\alpha_l, \beta_l, \sigma^2_l \sim N_N(1\alpha_l + X\beta_l, \sigma^2_l I), \quad (l = 1, 2, 3),
\]

\[
\beta_l|\tau_l, \lambda_l, \sigma^2_l \sim N_p(0, \sigma^2_l \tau_l^2 \Lambda_l), \quad (l = 1, 2, 3).
\]

The \( T \)-dimensional vector \( y_i = (y_{i1}, \cdots, y_{it}, \cdots, y_{iT})^\top \) \( (i = 1, \cdots, N) \) is the observed infection trajectory for the country \( i \) across the times. The notation \( f(\theta_1, \theta_2, \theta_3, \xi_i) \) \( (i = 1, \cdots, N) \) is \( T \)-dimensional vector that describes the Richard curves across the times:

\[
f(\theta_1, \theta_2, \theta_3, \xi_i) = (f(1; \theta_1, \theta_2, \theta_3, \xi_i), \cdots, f(T; \theta_1, \theta_2, \theta_3, \xi_i))^\top, \quad (i = 1, \cdots, N).
\]

The vectors \( \theta_l = (\theta_{1l}, \cdots, \theta_{Nl})^\top \) \( (l = 1, 2, 3) \) and \( \xi = (\xi_1, \cdots, \xi_N)^\top \) are \( N \)-dimensional vectors for the four parameters of the Richards curve (1) across the \( N \) countries.

The matrix \( X \) is \( N \)-by-\( p \) design matrix whose \( i \)-th row vector is given by the \( p \) predictors \( x_i = (x_{i1}, \cdots, x_{ip})^\top \in \mathbb{R}^p, \quad (i = 1, \cdots, N) \). The notation \( I \) stands for an identity matrix. Before implementing, it is recommended that each of column vectors of the design matrix \( X \) is standardized.
that is, each column vector has been centered, and then columnwisely scaled so that each column vector has mean zero and unit Euclidean norm ($l_2$-norm).

The $p$-dimensional vector $\beta_l = (\beta_{l1}, \ldots, \beta_{lp})^\top$ ($l = 1, 2, 3$) denotes $p$ coefficients from the $l$-th regression. The vector $\lambda_l = (\lambda_{l1}, \ldots, \lambda_{lp})^\top$ ($l = 1, 2, 3$) is $p$-dimensional vector for the local-scale parameters, and the matrix $\Pi_l$ is $p$-by-$p$ diagonal matrix $\Pi_l = \text{diag}(\lambda_{l1}^2, \ldots, \lambda_{lp}^2)$ ($l = 1, 2, 3$). The $\tau_l$ ($l = 1, 2, 3$) is referred to as the global-scale parameter (Carvalho et al., 2010).

Under the formulation of BHARM (2) – (6), our goal is to sample from the full joint posterior distribution $\pi(\theta_1, \theta_2, \theta_3, \xi, \sigma^2, \Omega_1, \Omega_2, \Omega_3|y_{1:N})$ where $\Omega_l = \{\alpha_l, \beta_l, \lambda_l, \tau_l, \sigma^2_l\}$ ($l = 1, 2, 3$), whose proportional part is given by

$$\prod_{i=1}^{N} N_T(y_i|f_i(\theta_{1i}, \theta_{2i}, \theta_{3i}, \xi_i), \sigma^2 I) \left\{ \prod_{i=1}^{3} N_N(\theta_l|1\alpha_l + X\beta_l, \sigma^2_l I)N_p(\beta_l|0, \sigma^2_l \tau_l^2 \Pi_l)\pi(\lambda_l)\pi(\tau_l)\pi(\sigma_l^2) \right\} \cdot \left\{ \prod_{i=1}^{N} \log N(\xi_i|0, 1) \right\} \pi(\sigma).$$

To sample from the full joint density, we use a Gibbs sampler (Casella and George, 1992) to exploit conditional independences among the latent variables induced by the hierarchy. The following algorithm describes a straightforward Gibbs sampler

**Step 1.** Sample $\theta_1$ from its full conditional distribution

$$\pi(\theta_1|\cdot) \sim N_N(\Sigma_{\theta_1}, \{(1/\sigma^2)r + (1/\sigma^2)^2(1\alpha_1 + X\beta_1), \Sigma_{\theta_1}\}),$$

where $\Sigma_{\theta_1} = \{(1/\sigma^2)H + (1/\sigma^2)^2 I\}^{-1} \in \mathbb{R}^{N \times N}$. The matrix $H$ is $N$-by-$N$ diagonal matrix $H = \text{diag}(||h(\theta_{21}, \theta_{31}, \xi_1)||^2_2, \ldots, ||h(\theta_{2N}, \theta_{3N}, \xi_N)||^2_2)$, and the vector $r$ is a $N$-dimensional vector which is given by $r = (y_1^T h(\theta_{21}, \theta_{31}, \xi_1), \ldots, y_N^T h(\theta_{2N}, \theta_{3N}, \xi_N))^T$, where each of the $T$-dimensional vector $h(\theta_{2i}, \theta_{3i}, \xi_i)$ ($i = 1, \ldots, N$) is obtained by

$$h(\theta_{2i}, \theta_{3i}, \xi_i) = (h(1; \theta_{2i}, \theta_{3i}, \xi_i), \ldots, h(T; \theta_{2i}, \theta_{3i}, \xi_i))^T,$$

where $h(t; \theta_2, \theta_3, \xi) = [1 + \xi \cdot \exp\{-\theta_2 \cdot (t - \theta_3)\}]^{-1/\xi}$.

**Step 2.** Sample $\theta_{2i}$ and $\theta_{3i}, i = 1, \ldots, N$, independently from their full conditional distributions. Proportional parts of the distributions are given by

$$\pi(\theta_{2i}|\cdot) \propto \exp\left( -\frac{1}{2\sigma^2}||y_i - f(\theta_{1i}, \theta_{2i}, \theta_{3i}, \xi_i)||^2_2 - \frac{1}{2\sigma^2_2}(\theta_{2i} - \alpha_2 - x_i^T \beta_2)^2 \right),$$

$$\pi(\theta_{3i}|\cdot) \propto \exp\left( -\frac{1}{2\sigma^2}||y_i - f(\theta_{1i}, \theta_{2i}, \theta_{3i}, \xi_i)||^2_2 - \frac{1}{2\sigma^2_3}(\theta_{3i} - \alpha_3 - x_i^T \beta_3)^2 \right).$$

Here, $\| \cdot \|_2$ indicates the $l_2$-norm. Note that the two conditional densities are not known in...
closed forms because two parameters, $\theta_{2i}$ and $\theta_{3i}$, participate to the function $f(\theta_{1i}, \theta_{2i}, \theta_{3i}, \xi_i)$ in a nonlinear way. We use the Metropolis algorithm (Andrieu et al., 2003) with Gaussian proposal densities within this Gibbs sampler algorithm.

**Step 3.** Sample $\xi_i$, $i = 1, \cdots, N$, independently from its full conditional distribution. Proportional parts of the distributions are given by

$$
\pi(\xi_i|\cdot) \propto \exp \left( -\frac{1}{2\sigma^2} \| y_i - f(\theta_{1i}, \theta_{2i}, \theta_{3i}, \xi_i) \|^2 \right) \cdot \log \mathcal{N}(\xi_i|0, 1). \quad (S.1)
$$

Note that the density (S.1) is not expressed in a closed form distribution. Because the shape parameter $\xi_i$ is supported on $(0, \infty)$ and participates in the Richards curve (1) as an exponent, sampling from the density needs a delicate care, where by we employed the elliptical slice sampler (Murray et al., 2010).

**Step 4.** Sample $\sigma^2$ from its full conditional distribution

$$
\pi(\sigma^2|\cdot) \sim \mathcal{IG} \left( \frac{NT}{2}, \frac{1}{2} \sum_{i=1}^{N} \| y_i - f(\theta_{1i}, \theta_{2i}, \theta_{3i}, \xi_i) \|^2 \right).
$$

**Step 5.** Sample $\alpha_l$, $l = 1, 2, 3$, independently from their full conditional distributions

$$
\pi(\alpha_l|\cdot) \sim \mathcal{N}_1(1^\top (\theta_l - X\beta_l)/N, \sigma^2_l/N).
$$

**Step 6.** Sample $\beta_l$, $l = 1, 2, 3$, independently from conditionally independent posteriors

$$
\pi(\beta_l|\cdot) \sim \mathcal{N}_p(\Sigma_{\beta_l} X^\top (\theta_l - 1\alpha_l), \sigma^2_l \Sigma_{\beta_l}),
$$

where $\Sigma_{\beta_l} = [X^\top X + \Lambda^{-1}_{\alpha_l}]^{-1} \in \mathbb{R}^{p \times p}$, $\Lambda_l = \text{diag}(\lambda^2_{l1}, \cdots, \lambda^2_{lp}) \in \mathbb{R}^{p \times p}$, and $\Lambda_{\alpha_l} = \tau^2_{\alpha_l} \Lambda_l$.

**Step 7.** Sample $\lambda_{lj}$, $l = 1, 2, 3$, $j = 1, \cdots, p$, independently from conditionally independent posteriors

$$
\pi(\lambda_{lj}|\cdot) \sim \mathcal{N}(\beta_{lj}|0, \sigma^2_l \tau^2_{\alpha_l} \lambda^2_{lj}) \cdot \{1/(1 + \lambda^2_{lj})\}.
$$

Note that the densities $\pi(\lambda_{lj}|\cdot)$ ($l = 1, 2, 3$, $j = 1, \cdots, p$) are not expressed in closed forms: we use the slice sampler (Neal, 2003).

**Step 8.** Sample $\tau_l$, $l = 1, 2, 3$, independently from conditionally independent posteriors

$$
\pi(\tau_l|\cdot) \sim \mathcal{N}_p(\beta_l|0, \sigma^2_l \tau^2_{\alpha_l} \Lambda_l) \cdot \{1/(1 + \tau^2_l)\}.
$$
Note that the densities $\pi(\tau_l \mid -)$ ($l = 1, 2, 3$) are not expressed in closed forms: we use the slice sampler (Neal, 2003).

**Step 9.** Sample $\sigma^2_l$, $l = 1, 2, 3$, independently from their full conditionally distributions

$$
\pi(\sigma^2_l \mid -) \sim IG\left(\frac{N + p}{2}, \frac{\|\theta_l - 1 \alpha_l - X \beta_l\|^2}{2} + \beta_l^\top \Lambda_{i_l}^{-1} \beta_l\right).
$$

**Elliptical slice sampler for Step 3**

To start with we shall use the variable change ($\eta = \log \xi$) to the right hand side of (S.1):

$$
\pi(\eta_i \mid -) \propto L(\eta_i) \cdot N(\eta_i \mid 0, 1), \quad (i = 1, \cdots, N),
$$

such that $L(\eta_i) = \exp\{-\|y_i - f(\theta_{1i}, \theta_{2i}, \theta_{3i}, e^{\eta_i})\|^2/(2\sigma^2)\}$ corresponds to a likelihood part.

Now, we use the elliptical slice sampler (ESS) (Murray et al., 2010; Nishihara et al., 2014) to sample from $\eta_i^{(s+1)} \sim \pi(\eta_i \mid -)$ (S.2) ($i = 1, \cdots, N$) that exploits the Gaussian prior measure. Conceptually, ESS and the Metropolis-Hastings (MH) algorithm (Chib and Greenberg, 1995) are similar: both methods are comprised of two steps: *proposal step* and *criterion step*. A difference between the two algorithms arises in the criterion step. If the new candidate does not pass the criterion, then MH takes the current state as the next state: whereas, ESS re-proposes a new candidate until rejection does not take place, rendering the algorithm rejection-free. Further information for ESS is referred to the original paper (Murray et al., 2010). After ESS has been employed, the realized $\eta_i$ needs to be transformed back through $\xi = e^\eta$. Algorithm 1 illustrates the full description of algorithms: to avoid notation clutter, the index $i$ is omitted.
Algorithm 1: Elliptical slice sampler to sample from $\pi(\xi)$

**Circumstance:** At the Step 3 of the $s$-th iteration of the Gibbs sampler.

**Input:** Current state $\xi^{(s)}$.

**Output:** A new state $\xi^{(s+1)}$.

1. Variable change ($\eta = \log \xi$): $\eta^{(s)} = \log \xi^{(s)}$.
2. Implement elliptical slice sampler;
   a. Choose an ellipse: $\nu \sim \mathcal{N}_1(0, 1)$.
   b. Define a criterion function:
      $$\alpha(\eta, \eta^{(s)}) = \min \{ \mathcal{L}(\eta)/\mathcal{L}(\eta^{(s)}), 1 \} : \mathbb{R} \to [0, 1],$$
      where $\mathcal{L}(\eta) = \exp\{-||y - f(\theta_1, \theta_2, \theta_3, e^\eta)||_2^2/(2\sigma^2)\}$.
   c. Choose a threshold and fix: $u \sim \mathcal{U}[0, 1]$.
   d. Draw an initial proposal $\eta^*$:
      $$\phi \sim \mathcal{U}(-\pi, \pi)$$
      $$\eta^* = \eta^{(s)} \cos \phi + \nu \sin \phi$$
   e. if ($u < \alpha(\eta^*, \eta^{(s)})$) { $\eta^{(s+1)} = \eta^*$ } else {
      Define a bracket: ($\phi_{\min}, \phi_{\max}$) = ($-\pi, \pi$).
      while ($u \geq \alpha(\eta^*, \eta^{(s)})$) {
        Shrink the bracket and try a new point:
        if ($\phi > 0$) $\phi_{\max} = \phi$ else $\phi_{\min} = \phi$
        $\phi \sim \mathcal{U}(\phi_{\min}, \phi_{\max})$
        $$\eta^* = \eta^{(s)} \cos \phi + \nu \sin \phi$$
      }
      $\eta^{(s+1)} = \eta^*$
  }
4. Variable change ($\xi = e^\eta$): $\xi^{(s+1)} = \exp \eta^{(s+1)}$. 
References


