Two-dimensional Dirichlet Process Mixture Model Implemented using a Blocked Gibbs Sampler

Brian M. Brost

17 June 2015

Implementation

The file dp.mixture.blocked.2d.sim.R simulates data according to the model statement presented below, and dp.mixture.blocked.2d.mcmc.R contains the MCMC algorithm for model fitting. Model implementation follows the blocked Gibbs sampler approach of Ishwaran and James (2001) and Gelman et al. (2014).

Model statement

Let $\mathbf{s}_t = (s_{1,t}, s_{2,t})'$, for $t = 1, \ldots, T$, be observations and $\boldsymbol{\mu}_{0,h} = (\mu_{0,1,h}, \mu_{0,2,h})'$, for $h = 1, \ldots, H$, be the locations of clusters. Define $\tilde{S}$ to be the uniform support of the Dirichlet process (i.e., all possible $\boldsymbol{\mu}_{0,h}$).

$$
\begin{aligned}
\mathbf{s}_t &\sim \mathcal{N}(\boldsymbol{\mu}_{0,h}, \sigma^2\mathbf{I}) \\
\boldsymbol{\mu}_{0,h} &\sim \mathbf{G} \\
\mathbf{G} &\sim \text{DP}(\alpha_0, G_0) \\
G_0 &\sim \text{Unif}(\tilde{\mathcal{S}}) \\
\alpha_0 &\sim \text{Gamma}(r, q) \\
\sigma &\sim \text{Unif}(l, u)
\end{aligned}
$$

The concentration parameter α_0 affects the clustering in the Dirichlet process mixture: smaller values yield fewer clusters with more observations per cluster, whereas larger values yield more clusters with fewer observations per cluster.

Full conditional distributions

Cluster locations ($\boldsymbol{\mu}_{0,h}$):

$$
\begin{aligned}
[\boldsymbol{\mu}_{0,h}|\cdot] &\propto \prod_{\{t:h_t=h\}} [\mathbf{s}_t|\boldsymbol{\mu}_{0,h}, \sigma^2][\boldsymbol{\mu}_{0,h} \mid \mathbf{G}] \\
&\propto \prod_{\{t:h_t=h\}} \mathcal{N}(\mathbf{s}_t|\boldsymbol{\mu}_{0,h}, \sigma^2) 1_{\{\boldsymbol{\mu}_{0,h} \in \tilde{\mathcal{S}}\}} \\
&\propto \prod_{\{t:h_t=h\}} \exp\left\{-\frac{1}{2}\left((\mathbf{s}_t - \boldsymbol{\mu}_{0,h})' \left(\sigma^2\mathbf{I}\right)^{-1} (\mathbf{s}_t - \boldsymbol{\mu}_{0,h})\right)\right\} 1_{\{\boldsymbol{\mu}_{0,h} \in \tilde{\mathcal{S}}\}} \\
&\propto \prod_{\{t:h_t=h\}} \exp\left\{-\frac{1}{2}\left(\mathbf{s}_t' \left(\sigma^2\mathbf{I}\right)^{-1} \mathbf{s}_t - 2\mathbf{s}_t' \left(\sigma^2\mathbf{I}\right)^{-1} \boldsymbol{\mu}_{0,h} + \boldsymbol{\mu}_{0,h}' \left(\sigma^2\mathbf{I}\right)^{-1} \boldsymbol{\mu}_{0,h}\right)\right\} 1_{\{\boldsymbol{\mu}_{0,h} \in \tilde{\mathcal{S}}\}} \\
&\propto \prod_{\{t:h_t=h\}} \exp\left\{-\frac{1}{2}\left(-2\mathbf{s}_t' \left(\sigma^2\mathbf{I}\right)^{-1} \boldsymbol{\mu}_{0,h} + \boldsymbol{\mu}_{0,h}' \left(\sigma^2\mathbf{I}\right)^{-1} \boldsymbol{\mu}_{0,h}\right)\right\} 1_{\{\boldsymbol{\mu}_{0,h} \in \tilde{\mathcal{S}}\}} \\
&\propto \exp\left\{-\frac{1}{2}\left(-2\sum_{\{t:h_t=h\}} \mathbf{s}_t' \left(\sigma^2\mathbf{I}\right)^{-1} \boldsymbol{\mu}_{0,h} + \boldsymbol{\mu}_{0,h}' \left(n_h \left(\sigma^2\mathbf{I}\right)^{-1}\right) \boldsymbol{\mu}_{0,h}\right)\right\} 1_{\{\boldsymbol{\mu}_{0,h} \in \tilde{\mathcal{S}}\}} \\
&= \mathcal{N}(\mathbf{A}^{-1}\mathbf{b}, \mathbf{A}^{-1}) 1_{\{\boldsymbol{\mu}_{0,h} \in \tilde{\mathcal{S}}\}}
\end{aligned}
$$

where $\mathbf{A} = n_h \left(\sigma^2 \mathbf{I}\right)^{-1}$ and $\mathbf{b} = \sum_{\{t:h_t=h\}} \mathbf{s}_t' \left(\sigma^2 \mathbf{I}\right)^{-1}$; therefore, $[\boldsymbol{\mu}_{0,h}|\cdot] = \mathcal{N}(\mathbf{s}_t, \frac{\sigma^2}{n_h}\mathbf{I})$. Note that the product (or summation) is over all $\mathbf{s}_t$ that belong to cluster h (h_t is a latent class status that indicates membership of observation $\mathbf{s}_t$ to cluster h), and n_h is the number of observations allocated to h. Also note that 'proposed' values for $\boldsymbol{\mu}_{0,h}$ not in $\tilde{\mathcal{S}}$ are rejected, i.e., $[\boldsymbol{\mu}_{0,h}|\cdot] = \mathcal{TN}(\mathbf{s}_t, \frac{\sigma^2}{n_h}\mathrm{I})_{\tilde{\mathcal{S}}}$.

Latent class status (h_t):

$$[h_t|\cdot] \quad \sim \quad \frac{\pi_h[\mathbf{s}_t|\boldsymbol{\mu}_{0,h}, \sigma^2]}{\sum_{\tilde{h}}^{H} \pi_{\tilde{h}}[\mathbf{s}_t|\boldsymbol{\mu}_{0,\tilde{h}}, \sigma^2]},$$

just as in multinomial sampling. See page 552 in Gelman et al. (2014) for this full conditional.

Probability mass for cluster h (π_h):

$$\pi_h = v_h \prod_{\tilde{h}<h} (1 - v_{\tilde{h}}),$$

where

$$[v_h|\cdot] \quad \sim \quad \text{Beta}\left(1 + n_h, \alpha_0 + \sum_{\tilde{h}=h+1}^{H} n_{\tilde{h}}\right).$$

The parameter n_h denotes the number of observations allocated to cluster h. Note that v_h is sampled in order of decreasing n_h, i.e., n_h is sorted largest to smallest and v_h is sampled in sequence. The probabilities π_h are calculated in order of decreasing n_h as well. See page 553 in Gelman et al. (2014) and Section 5.2 in Ishwaran and James (2001) for this full conditional.

Dirichlet process concentration parameter (α_0):

$$[\alpha_0|\cdot] \quad \propto \quad \text{Gamma}(r + H - 1, q - \sum_{h=1}^{H-1} \log(1 - v_h)).$$

See page 553 in Gelman et al. (2014) for this full conditional.

Error in the observation process (σ):

$$[\sigma|\cdot] \quad \propto \quad \prod_{t=1}^{T} [\mathbf{s}_t|\boldsymbol{\mu}_{0,h}, \sigma^2][\sigma]$$

The update for σ proceeds using Metropolis-Hastings.

References

Gelman, A., J.B. Carlin, H.S. Stern, D.B. Dunson, A. Vehtari, and D.B. Rubin. 2014. Bayesian data analysis. CRC Press.

Ishwaran, H., and L.F. James. 2001. Gibbs sampling methods for stick-breaking priors. Journal of the American Statistical Association 96: 161–173.