Two-dimensional Hierarchical Dirichlet Process Mixture Model

Brian M. Brost

10 November 2015

Description

A hierarchical Dirichlet process mixture model for cluster estimation of 2-dimensional, normally distributed data.

Implementation

The file hdp.mixture.2d.sim.R simulates data according to the model statement presented below, and hdp.mixture.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 $\mathbf{s}_{jt} = (s_{1,jt}, s_{2,jt})'$ be observations at times $t = 1, \ldots, T$ for groups $j = 1, \ldots, J$. Also let $\boldsymbol{\mu}_h = (\mu_{1,h}, \mu_{2,h})'$, for $h = 1, \ldots, H$, be the locations of clusters, where the parameter H denotes the maximum number of clusters allowed unde the truncation approximation of Dirichlet process mixture (Gelman et al. 2014). Define h_{jt} to be an index variable that identifies the $\boldsymbol{\mu}_h$ associated with each $\mathbf{s}_{jt}$. Furthermore, denote the support of the Dirichlet process (i.e., all possible $\boldsymbol{\mu}_h$) as $\tilde{S}$.

$$
\begin{aligned}
\mathbf{s}_{jt} &\sim \mathcal{N}(\boldsymbol{\mu}_{h_{jt}}, \sigma^2 \mathbf{I}) \\
h_{jt} &\sim \text{Cat}\left(\pi_{j1}, \ldots, \pi_{jH}\right) \\
\pi_{jh} &= \eta_{jh} \prod_{l=1}^{h-1} (1 - \eta_{jl}) \\
\eta_{jh} &\sim \text{Beta}\left(\theta_j \theta_0, \theta_j \left(1 - \sum_{l=1}^{h} \pi_{0l}\right)\right) \\
\pi_{0h} &\sim \text{Stick}\left(\theta_0\right) \\
\boldsymbol{\mu}_h &\sim \text{Unif}(\tilde{\mathcal{S}}) \\
\theta_j &\sim \text{Gamma}\left(r_{\theta_j}, q_{\theta_j}\right) \\
\theta_0 &\sim \text{Gamma}\left(r_{\theta_0}, q_{\theta_0}\right) \\
\sigma &\sim \text{Unif}(l, u)
\end{aligned}
$$

The concentration parameter θ 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. Note that the lines in this model statement pertaining to h_{jt}, π_{jh}, η_{jt}, π_{0h}, and $\boldsymbol{\mu}_h$ comprise the stick-breaking representation of the hierarchical Dirichlet process mixture model, i.e.,

$$
\begin{aligned}
\boldsymbol{\mu}_{h_{jt}} &\sim \mathbf{P_j} \\
\mathbf{P}_j &\sim \text{DP}(\theta_j, \mathbf{P}_0) \\
\mathbf{P}_0 &\sim \text{DP}(\theta_0, \mathbf{P}_{00}) \\
\mathbf{P}_{00} &\sim \text{Unif}(\tilde{\mathcal{S}})
\end{aligned}
$$

Full conditional distributions

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

$$
\begin{aligned}
[\boldsymbol{\mu}_h|\cdot] &\propto \prod_{j=1}^{J}\prod_{t=1}^{T}[\mathbf{s}_{jt}|\boldsymbol{\mu}_{h_{jt}},\sigma^2]^{1_{\{h_{jt}=h\}}}[\boldsymbol{\mu}_h \mid \tilde{\mathcal{S}}] \\
&\propto \prod_{j=1}^{J}\prod_{\{t:h_{jt}=h\}} \mathcal{N}(\mathbf{s}_{jt}|\boldsymbol{\mu}_h,\sigma^2)1_{\{\boldsymbol{\mu}_h\in\tilde{\mathcal{S}}\}} \\
&\propto \prod_{j=1}^{J}\prod_{\{t:h_t=h\}} \exp\left\{-\frac{1}{2}\left((\mathbf{s}_{jt}-\boldsymbol{\mu}_h)'\left(\sigma^2\mathbf{I}\right)^{-1}(\mathbf{s}_{jt}-\boldsymbol{\mu}_h)\right)\right\}1_{\{\boldsymbol{\mu}_h\in\tilde{\mathcal{S}}\}} \\
&\propto \prod_{j=1}^{J}\prod_{\{t:h_t=h\}} \exp\left\{-\frac{1}{2}\left(\mathbf{s}_{jt}'\left(\sigma^2\mathbf{I}\right)^{-1}\mathbf{s}_{jt}-2\mathbf{s}_{jt}'\left(\sigma^2\mathbf{I}\right)^{-1}\boldsymbol{\mu}_h+\boldsymbol{\mu}_h'\left(\sigma^2\mathbf{I}\right)^{-1}\boldsymbol{\mu}_h\right)\right\}1_{\{\boldsymbol{\mu}_h\in\tilde{\mathcal{S}}\}} \\
&\propto \prod_{j=1}^{J}\prod_{\{t:h_t=h\}} \exp\left\{-\frac{1}{2}\left(-2\mathbf{s}_{jt}'\left(\sigma^2\mathbf{I}\right)^{-1}\boldsymbol{\mu}_h+\boldsymbol{\mu}_h'\left(\sigma^2\mathbf{I}\right)^{-1}\boldsymbol{\mu}_h\right)\right\}1_{\{\boldsymbol{\mu}_h\in\tilde{\mathcal{S}}\}} \\
&\propto \exp\left\{-\frac{1}{2}\left(-2\sum_{j=1}^{J}\sum_{\{t:h_t=h\}}\mathbf{s}_{jt}'\left(\sigma^2\mathbf{I}\right)^{-1}\boldsymbol{\mu}_h+\sum_{j=1}^{J}\boldsymbol{\mu}_h'\left(n_{jh}\left(\sigma^2\mathbf{I}\right)^{-1}\right)\boldsymbol{\mu}_h\right)\right\}1_{\{\boldsymbol{\mu}_h\in\tilde{\mathcal{S}}\}} \\
&\propto \exp\left\{-\frac{1}{2}\left(-2\sum_{j=1}^{J}\sum_{\{t:h_t=h\}}\mathbf{s}_{jt}'\left(\sigma^2\mathbf{I}\right)^{-1}\boldsymbol{\mu}_h+\boldsymbol{\mu}_h'\left(m_h\left(\sigma^2\mathbf{I}\right)^{-1}\right)\boldsymbol{\mu}_h\right)\right\}1_{\{\boldsymbol{\mu}_h\in\tilde{\mathcal{S}}\}} \\
&= \mathcal{N}(\mathbf{A}^{-1}\mathbf{b},\mathbf{A}^{-1})1_{\{\boldsymbol{\mu}_h\in\tilde{\mathcal{S}}\}}
\end{aligned}
$$

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

Group-specific probability mass for cluster location $\boldsymbol{\mu}_h$ (π_{jh}):

The stick-breaking construction of Dirichlet processes consistes of two components, namely a cluster weight and a cluster probability. Let η_{jh} denote the weight assigned to cluster $\boldsymbol{\mu}_h$ in group j. These group-specific weights are related to the global cluster probability (π_{0h}) such $\eta_{jh}\sim\text{Beta}\left(\theta_j\theta_0,\theta_j\left(1-\sum_{l=1}^{h}\pi_{0l}\right)\right)$. For $h=1,\ldots,H-1$, the associated full conditional is

$$
\begin{aligned}
[\eta_{jh}|\cdot] &\propto \prod_{t=1}^{T}[h_{jt}\mid\pi_{jh}]^{1_{\{h_{jt}=h\}}}\prod_{\tilde{h}=h+1}^{H}\prod_{t=1}^{T}\left[h_{jt}\mid\pi_{j\tilde{h}}\right]^{1_{\{h_{jt}=\tilde{h}\}}}[\eta_{jh}|\theta_j,\theta_0,\boldsymbol{\pi}_0] \\
&\propto \prod_{t=1}^{T}\pi_{jh}^{1_{\{h_{jt}=h\}}}\prod_{\tilde{h}=h+1}^{H}\prod_{t=1}^{T}\pi_{j\tilde{h}}^{1_{\{h_{jt}=\tilde{h}\}}}[\eta_{jh}|\theta_j,\theta_0,\boldsymbol{\pi}_0] \\
&\propto \pi_{jh}^{\sum_t 1_{\{h_{jt}=h\}}}\prod_{\tilde{h}=h+1}^{H}\pi_{j\tilde{h}}^{\sum_t 1_{\{h_{jt}=\tilde{h}\}}}[\eta_{jh}|\theta_j,\theta_0,\boldsymbol{\pi}_0] \\
&\propto \left(\eta_{jh}\prod_{l=1}^{h-1}(1-\eta_{jl})\right)^{\sum_t 1_{\{h_{jt}=h\}}}\prod_{\tilde{h}=h+1}^{H}\left(\eta_{j\tilde{h}}\prod_{l=1}^{\tilde{h}-1}(1-\eta_{jl})\right)^{\sum_t 1_{\{h_{jt}=\tilde{h}\}}}\eta_{jh}^{\theta_j\theta_0-1}(1-\eta_{jh})^{\theta_j\left(1-\sum_{l=1}^{h}\pi_{0l}\right)-1}
\end{aligned}
$$

$$
\begin{aligned}
&\propto \quad \eta_{jh}{}^{\sum_t 1_{\{h_{jt}=h\}}} \prod_{\tilde{h}=h+1}^{H} \left(\prod_{l=1}^{\tilde{h}-1} (1-\eta_{jl}) \right)^{\sum_t 1_{\{h_{jt}=\tilde{h}\}}} \eta_{jh}^{\theta_j\theta_0-1} (1-\eta_{jh})^{\theta_j\left(1-\sum_{l=1}^{h}\pi_{0l}\right)-1} \\
&\propto \quad \eta_{jh}{}^{\sum_t 1_{\{h_{jt}=h\}}} (1-\eta_{jh})^{\sum_{\tilde{h}=h+1}^{H}\sum_t 1_{\{h_{jt}=\tilde{h}\}}} \eta_{jh}^{\theta_j\theta_0-1} (1-\eta_{jh})^{\theta_j\left(1-\sum_{l=1}^{h}\pi_{0l}\right)-1} \\
&\propto \quad \eta_{jh}{}^{\sum_t 1_{\{h_{jt}=h\}}+\theta_j\theta_0-1} (1-\eta_{jh})^{\sum_{\tilde{h}=h+1}^{H}\sum_t 1_{\{h_{jt}=\tilde{h}\}}+\theta_j\left(1-\sum_{l=1}^{h}\pi_{0l}\right)-1} \\
&= \quad \text{Beta}\left(\sum_t 1_{\{h_{jt}=h\}} + \theta_j\theta_0, \sum_{\tilde{h}=h+1}^{H} \sum_t 1_{\{h_{jt}=\tilde{h}\}} + \theta_j \left(1-\sum_{l=1}^{h}\pi_{0l}\right)\right) \\
&= \quad \text{Beta}\left(n_{jh} + \theta_j\theta_0, \sum_{\tilde{h}=h+1}^{H} n_{j\tilde{h}} + \theta_j \left(1-\sum_{l=1}^{h}\pi_{0l}\right)\right)
\end{aligned}
$$

and $\eta_{jH} = 1$ to ensure $\sum_h \pi_{jh} = 1$. The varaible n_{jh} denotes the number of observations in group j allocated to cluster $\boldsymbol{\mu}_h$. Note that η_{jh} is sampled in order of decreasing n_{jh}, i.e., n_{jh} is sorted largest to smallest and η_{jh} is sampled in sequence. The group-specific cluster probabilities (π_{jh}) are deterministic and calculated as

$$
\pi_{jh} = \eta_{jh} \prod_{\tilde{h}=1}^{h-1} (1-\eta_{j\tilde{h}}).
$$

See page 553 in Gelman et al. (2014) and Section 5.2 in Ishwaran and James (2001) for the corresponding update in a Dirichlet process mixture model.

Global probability mass for cluster location $\boldsymbol{\mu}_h$ (π_{0h}):

Let η_{0h} denote the global weight assigned to cluster h, where $\eta_{0h} \sim \text{Beta}(1, \theta_0)$. For $h = 1, \ldots, H-1$, the associated full-conditional is

$$
\begin{aligned}
[\eta_{0h}|\cdot] &\propto [\eta_{0h}|1,\theta_0] \prod_{j=1}^{J} [\eta_{jh}|\theta_j, \theta_0, \boldsymbol{\pi}_0] \\
&\propto \eta_{0h}^{1-1} (1-\eta_{0h})^{\theta_0-1} \prod_{j=1}^{J} \eta_{jh}^{\theta_j\theta_0-1} (1-\eta_{jh})^{\theta_j\left(1-\sum_{l=1}^{h}\pi_{0l}\right)-1}
\end{aligned}
$$

The update proceeds via Metropolis-Hastings, and define $\eta_H = 1$ to ensure $\sum_h \pi_{0h} = 1$. Presumably, η_{0h} is sampled in order of decreasing m_h, the total number of observations allocated to $\boldsymbol{\mu}_h$ over all groups. In other words, m_h is sorted largest to smallest, and η_{0h} is sampled in sequence. The global cluster probabilities (π_{0h}) are deterministic and calculated as

$$
\pi_{0h} = \eta_{0h} \prod_{\tilde{h}=1}^{h-1} \left(1-\eta_{0\tilde{h}}\right).
$$

The probabilities π_{0h} are also calculated in order of decreasing m_h.

Dirichlet process concentration parameter (θ):

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

See page 553 in Gelman et al. (2014). Also see Escobar and West (1995) and West (1997?, white paper) for alternative full-conditionals for θ.

Latent cluster classification variable (h_{jt}):

$$\begin{aligned}
[h_{jt}|\cdot] &\sim \left[\mathbf{s}_{jt} \mid \boldsymbol{\mu}_{h_{jt}}, \sigma^2\right][h_{jt} \mid \pi_{jh}] \\
&\sim \text{Cat}\left(\frac{\pi_{jh}\left[\mathbf{s}_{jt} \mid \boldsymbol{\mu}_{h_{jt}}, \sigma^2\right]}{\sum_{\tilde{h}=1}^{H} \pi_{j\tilde{h}}\left[\mathbf{s}_{jt} \mid \boldsymbol{\mu}_{\tilde{h}}, \sigma^2\right]}\right) \\
&\sim \text{Cat}\left(\frac{\pi_{jh}\left(\mathcal{N}\left(\mathbf{s}_{jt} \mid \boldsymbol{\mu}_{h_{jt}}, \sigma^2\mathbf{I}\right)\right)}{\sum_{\tilde{h}=1}^{H} \pi_{j\tilde{h}}\left(\mathcal{N}\left(\mathbf{s}_{jt} \mid \boldsymbol{\mu}_{\tilde{h}}, \sigma^2\mathbf{I}\right)\right)}\right)
\end{aligned}$$

This update proceeds just as in multinomial sampling; see page 552 in Gelman et al. (2014).

Error in the observation process (σ):

$$[\sigma|\cdot] \propto \prod_{j=1}^{J}\prod_{t=1}^{T}[\mathbf{s}_{jt}|\boldsymbol{\mu}_{h_{jt}}, \sigma^2][\sigma]$$

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.