Slice sampler for DPM using constructive definition of Dirichlet process [1–3]

Let $\mathbf{V} = V_1, V_2, \ldots,$ has distribution λ^∞ where $\lambda = \text{Beta}(1, \alpha)$. Let $\boldsymbol{\theta}^* = (\theta_1^*, \theta_2^*, \ldots,)$ be independent of $V_1, V_2, \ldots$ and has distribution G_0^∞. Let $\pi_1 = V_1, \pi_2 = (1 - V_1)V_2, \ldots$. Then $(\pi_1, \pi_2, \ldots)$ constitutes a random discrete probability distribution. The Dirichlet process $\text{DP}(\alpha G_0)$ can be understood as the distribution of the random probability measure P given by

$$P = \sum_{h=1}^{\infty} \pi_h \delta_{\theta_h^*}$$

in view of its constructive definition (also called the stick-breaking construction). The DPM model stated that, given such a $P, (y_1, \ldots, y_n)$ are i.i.d. with common pdf $\int N(y; \theta) dP(\theta)$ and this can be expressed as

$$f(y) = \sum_{h=1}^{\infty} \pi_h N(y; \theta_h^*)$$

The problem is to obtain the posterior distribution of P given $\mathbf{y}$ or alternatively, to obtain the conditional distribution of $\mathbf{V}$ and $\boldsymbol{\theta}^*$ given $\mathbf{y}$ and then convert it to the posterior distribution of P. It may seem that we have substituted a problem involving a finite number of variables with a problem involving an infinite number of random variables and made the problem more complicated. Some clever argumentation of random variables makes this problem simpler and amenable to computation.

Given P, the pdf of y_1 is $f(y) = \sum_{h=1}^{\infty} \pi_h N(y; \theta_h^*)$. Let $u_1 \in [0, 1], S_1 \in \{1, 2, \ldots\}$ be (augmented) random variables such that, given P or equivalently given $\mathbf{V}, \boldsymbol{\theta}^*$, the joint pdf-pmf of (y_1, u_1, S_1) is

$$I(u_1 \leq \pi_{S_1}) N(y_1, \theta_{S_1}^*).$$

Then, again conditional on P, the marginal density of y_1 is $f(y) = \sum_{h=1}^{\infty} \pi_h N(y; \theta_h^*)$ which agrees with what we started in the beginning. Conditional on P, the marginal distribution of S_1 is the discrete distribution $(\pi_1, \pi_2, \ldots)$, and the marginal distribution of u_1 has pdf $\sum_{h=1}^{\infty} I(u \leq \pi_h)$. We also introduce $(u_i, S_i), i = 2, \ldots, n$ to go with $y_2, \ldots, y_n$ in a similar way. Thus we have the data $\mathbf{y}$, the augmented (unobserved) variables $\mathbf{u} = (u_1, \ldots, u_n)$, $\mathbf{S} = (S_1, \ldots, S_n)$ and the parameters $\mathbf{V}, \boldsymbol{\theta}^*$. Their joint distribution may be written as

$$\prod_{i=1}^{n} I(u_i \leq \pi_{S_i}) N(y_i, \theta_{S_i}^*) \times \lambda^\infty \times G_0^\infty. \tag{1}$$

Let $n_h = \sum_{i=1}^{n} I(S_i = h), h = 1, 2, \ldots$ be the multiplicities among $\mathbf{S}$. Note that $n_h = 0$ if

$h > \max\{S_i\}$.

From (1), it is easy to see that the conditional distributions of $\mathbf{u}$ given $\mathbf{y}, \mathbf{S}, \mathbf{V}, \boldsymbol{\theta}^*$ are independent and

$$u_i \sim \mathrm{U}[0, \pi_{S_i}], \quad i = 1, \ldots, n.$$

If we drop the $\mathbf{u}$, the joint distribution of $(\mathbf{y}, \mathbf{S}, \mathbf{V}, \boldsymbol{\theta}^*)$ can be written as

$$\prod_{i=1}^{n} \pi_{S_i} N(y_i, \theta^*_{S_i}) \times \lambda^\infty \times G_0^\infty. \tag{2}$$

From the form of the joint distribution in (2), the conditional distribution of $\mathbf{V}$ given $(\mathbf{y}, \mathbf{S}, \boldsymbol{\theta}^*)$ are independent and the marginal conditional distributions are

$$V_h \sim \mathrm{Beta}(1 + n_h, \alpha + \sum_{k>h} n_k), h = 1, 2, \ldots, \max\{S_i\}$$

Note that the conditional distributions of V_h remain unchanged from $\mathrm{Beta}(1, \alpha)$ for $h > \max\{S_i\}$. Again, using (1), the conditional distributions of $\boldsymbol{\theta}^*$ given $(\mathbf{y}, \mathbf{S}, \mathbf{u}, \mathbf{V})$ (notice that we brought back $\mathbf{u}$ here) are independent and the marginal conditional distributions are

$$\theta^*_h \sim G_0(dz) \prod_{i:S_i=h} N(y_i; \theta^*_h)$$

which is exactly same as that in the finite mixture model case. Again the conditional distributions of θ^*_h remain unchanged from G_0 when $h > \max\{S_i\}$ and when $n_h = 0, h < \max S_i$. However $\mathbf{u}$ and $\mathbf{S}$ are only augmented variables and not part of the data $\mathbf{y}$ and therefore we have to provide their (full) conditional distributions.

Let $H_i(u_i) = \{h : u_i \leq \pi_h\}, i = 1, 2, \ldots, n$. Again, from (1) it is easy to see that the conditional distributions of $\mathbf{S}$ given $\mathbf{y}, \mathbf{u}, \mathbf{V}, \boldsymbol{\theta}^*$ are independent and

$$P(S_i = h \mid \mathbf{y}, \mathbf{u}, \mathbf{V}, \boldsymbol{\theta}^*) = \frac{I(u_i \leq \pi_h) N(y_i; \theta^*_h)}{\sum_{h \in H_i(u_i)} I(u_i \leq \pi_h) N(y_i; \theta^*_h)}$$

Note that the above is zero if $h \notin H_i(u_i)$ and the denominator is a finite sum. All these make computations feasible.

To sample the V_hs and θ^*_hs, [3] recommends drawing until the minimum h^* such that $\sum_{h=1}^{h^*} \pi_h > 1 - \min(u_1, \ldots, u_n)$ which provides a tight upper bound to the maximum element in $H_i(u_i)$.

References

[1] Omiros Papaspiliopoulos. A note on posterior sampling from dirichlet mixture models. 2008.

[2] Omiros Papaspiliopoulos and Gareth O Roberts. Retrospective markov chain monte carlo methods for dirichlet process hierarchical models. *Biometrika*, 95(1):169–186, 2008.

[3] Stephen G Walker. Sampling the dirichlet mixture model with slices. *Communications in StatisticsSimulation and Computation®*, 36(1):45–54, 2007.