DERIVATION OF FULL CONDITIONAL POSTERIOR FOR MARGINAL GIBBS SAMPLER

MATTIAS VILLANI

ABSTRACT. Just a little derivation of the full conditional posterior for the marginal Gibbs sampler for DP mixtures.

1. DERIVATION

The aim is to derive the posterior for θ_i conditional on all other $\theta_{-i} = (\theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_n)$. The prior is given by the Polya scheme

$$p(\theta_i|\theta_1, \ldots, \theta_{i-1}) = \frac{\alpha}{\alpha + i - 1} P_0(\theta_i) + \frac{1}{\alpha + i - 1} \sum_{j=1}^{i-1} \delta_{\theta_j}.$$

So for the last observation $i = n$ we have the full conditional prior

$$p(\theta_i|\theta_{-i}) = \frac{\alpha}{\alpha + n - 1} P_0(\theta_i) + \frac{1}{\alpha + n - 1} \sum_{j=1}^{n-1} \delta_{\theta_j}.$$

Since the order of the observations y_i does matter in this model (the y_i are exchangeable), this can be taken as the full conditional prior for any θ_i. Let $\theta_1^*, \ldots, \theta_{k^{(-i)}}^*$ denote the $k^{(-i)}$ unique values among the elements in θ_{-i}, and let $n_j^{(-i)}$ be the number of values that are exactly θ_j^*. The prior can be written

$$p(\theta_i|\theta_{-i}) = \frac{\alpha}{\alpha + n - 1} P_0(\theta_i) + \frac{1}{\alpha + n - 1} \sum_{j=1}^{k^{(-i)}} n_j^{(-i)} \delta_{\theta_j^*}.$$

The full conditional posterior is

$$p(\theta_i|\theta_{-i}, \boldsymbol{y}) \propto p(\boldsymbol{y}|\theta_i, \theta_{-i}) p(\theta_i|\theta_{-i}) \propto p(y_i|\theta_i) p(\theta_i|\theta_{-i}) = \mathcal{K}(y_i|\theta_i) p(\theta_i|\theta_{-i}),$$

since $p(\boldsymbol{y}|\theta_i, \theta_{-i}) = \prod_{i=1}^n p(y_i|\theta_i)$. Now, substituting the conditional prior gives

$$\begin{aligned} p(\theta_i|\theta_{-i}, \boldsymbol{y}) &= \frac{\alpha}{\alpha + n - 1} \mathcal{K}(y_i|\theta_i) P_0(\theta_i) + \frac{1}{\alpha + n - 1} \sum_{j=1}^{k^{(-i)}} n_j^{(-i)} \mathcal{K}(y_i|\theta_i) \delta_{\theta_j^*} \\ &= \frac{\alpha}{\alpha + n - 1} p_0(y_i) p_0(\theta_i|y_i) + \frac{1}{\alpha + n - 1} \sum_{j=1}^{k^{(-i)}} n_j^{(-i)} \mathcal{K}(y_i|\theta_j^*) \delta_{\theta_j^*}, \end{aligned}$$

where $p_0(\theta_i|y_i) = \mathcal{K}(y_i|\theta_i) P_0(\theta_i)/p_0(y_i)$ is the posterior based on y_i and $p_0(y_i) = \int \mathcal{K}(y_i|\theta_i) dP_0(\theta_i)$ is the marginal likelihood of y_i under the base measure $P_0(\theta_i)$. The full conditional posterior

Villani: *Department of Statistics, Stockholm University* and Dept of Computer and Information Science, Linkoping University. *E-mail: mattias.villani@gmail*.com.

of θ_i is therefore a mixture distribution with one component being $p_0(\theta_i|y_i)$ and the other components being point masses $\delta_{\theta_j^*}$ for $j = 1, ..., k^{(-i)}$. The unnormalized weights of the mixture components are proportional to $\alpha p_0(y_i)$ and $n_j^{(-i)}\mathcal{K}(y_i|\theta_j^*)$. Hence by formulating a mixture in terms of the indicators I_i for component membership we obtain

$$\Pr(I_i = j|I_{-i}, \boldsymbol{y}) \propto \begin{cases} n_j^{(-i)}\mathcal{K}(y_i|\theta_j^*) & \text{for } j = 1, \ldots, k^{(-i)} \\ \alpha p_0(y_i) = \int \mathcal{K}(y_i|\theta_i)dP_0(\theta_i) & \text{for } j = k^{(-i)} + 1 \end{cases}.$$

Of course, when we simulate the indicators, we need normalize so that $\sum_{j=1}^{k^{(-i)}+1} \Pr(I_i = j|I_{-i}, \boldsymbol{y}) = 1$, by dividing by the sum of the unnormalized weights.

After having simulated $I_1, \ldots, I_n$, we sample the unique θ_j^* from

$$\begin{aligned} p(\theta_j^*|I_1, \ldots, I_n, \theta_{-j}^*, \boldsymbol{y}) &\propto p(\boldsymbol{y}|I_1, \ldots, I_n, \theta_j^*, \theta_{-j}^*)p(\theta_j^*|I_1, \ldots, I_n, \theta_{-j}^*). \\ &= \left(\prod_{i:I_i=1} \mathcal{K}(y_i|\theta_1^*) \cdots \prod_{i:I_i=k} \mathcal{K}(y_i|\theta_k^*)\right) P_0(\theta_j^*) \propto P_0(\theta_j^*) \prod_{i:I_i=j} \mathcal{K}(y_i|\theta_j^*), \end{aligned}$$

where k is the number of unique θ_j^* (as dictated by $I_1, \ldots, I_n$). Note that simulating from $\Pr(I_i = j|I_{-i}, \boldsymbol{y})$ may have opened up new components and for those components, θ_j^* is drawn from a posterior based on a single observation.