Gibbs sampling for HDP-LDA

1 Chinese Restaurant Franchise

1.1 Definitions

$$G_0 \sim DP(\gamma, H) \qquad G_j \sim DP(\alpha_0, G_0) \qquad \theta_{ji} \sim G_j \qquad x_{ji} \sim F(\theta_{ji}) \tag{1}$$

$j = 1, \ldots, J$ restaurants

θ_{ji} is a customer in restaurant j

x_{ji} are observed words

t_{ji} is the table sat on by customer i in restaurant j

k_{jt} is the dish index of table t in restaurant j

n_{jtk} is the number of customers in restaurant j at table t served with dish k

m_{jk} is the number of tables in restaurant j served with dish k

1.2 HDP-LDA

H is the topic distribution over the vacabulary $H \sim Dirichlet(\beta)$

$\phi_1, \ldots, \phi_K$ are distinct dishes that restaurants serve $\phi_k \sim H$

F is the multinomial distribution over the vocabulary $x_{ji} \sim Mult(\theta_{ji})$

We have:

$$h(\phi_k) = \frac{\prod_v [\phi_k]_v^{\beta-1}}{C} \tag{2}$$

C is a constant.

The derivation of the conditional distribution of word x_{ji} with topic k given all other observations $f_k^{-ji}(x_{ji})$ (eq.30 on [Teh+2006]) is as follows: Let n_{kv}^{-ji} be number of words v served with dish k excluding the current observation. and $n_{k.}^{-ji}$ is the number of words served with dish k excluding the current observation.

When a new customer arrives at the restaurant j, if he sits on an existing table:

$$f_k^{-ji}(x_{ji}) = \frac{\beta + n_{kv}^{-ji}}{V\beta + n_{k.}^{-ji}} \tag{3}$$

If he takes a new table:

$$f_{k^{new}}^{-ji}(x_{ji}) = \frac{1}{V} \tag{4}$$

1.3 Posterior Sampling

The likelihood due to x_{ji} given $t_{ji} = t$ for some previously used t is $f_k^{-ji}(x_{ji})$, The likelihood for $t_{ji} = t^{new}$ can be calculated by integrating out the possible values for $k_{jt^{new}}$:

$$p(x_{ji}|\mathbf{t}^{-ji}, t_{ji} = t^{new}, \mathbf{k}) = \sum_{k=1}^{K}\left(\frac{m_{.k}}{m_{..} + \gamma} \cdot \frac{\beta + n_{kv}^{-ji}}{V\beta + n_{k.}^{-ji}}\right) + \frac{\gamma}{m_{..} + \gamma} \cdot \frac{1}{V} \tag{5}$$

The conditional distribution of the table index t_{ji} given the remainder of the variables:

$$P(t_{ji}|\mathbf{t^{-ji}}, \mathbf{k}) \propto \begin{cases} n_{jt.}^{-ji} \cdot \frac{\beta + n_{kv}^{-ji}}{V\beta + n_{k.}^{-ji}}, & \text{if t is previously used} \\ \alpha_0 \cdot p(x_{ji}|\mathbf{t}^{-ji}, t_{ji} = t^{new}, \mathbf{k}), & \text{if } t = t^{new} \end{cases} \tag{6}$$

If the sampled value of t_{ji} is t^{new}, we obtain a sample of $k_{jt^{new}}$ by sampling from the conditional distribution:

$$p(k_{jt^{new}} = k|\mathbf{t}, \mathbf{k}^{-jt^{new}}) \propto \begin{cases} m_{.k} \cdot \frac{\beta + n_{kv}^{-ji}}{V\beta + n_{k.}^{-ji}}, & \text{if k is previously used} \\ \frac{\gamma}{V}, & \text{if } k = k^{new} \end{cases} \tag{7}$$