Posterior Sampling for α

David Pfau

30 October 2009

To perform Metropolis sampling from α given the class assignment variables for each data point, we first sample a point from a symmetric distribution and then calculate the ratio of posterior probabilities to find the probability of accepting the new sample. We should note, this is not an ideal sampler near 0, since a symmetric proposal will often propose negative α and thus will often be rejected. Nonetheless, it's just plain quicker (and trivial to extend to asymmetric proposals). In our case we choose a diffuse Gamma prior for α with parameters $(1, 1)$. For $\alpha > 0$ the posterior is given by

$$p(\alpha|\mathbf{z}) \propto p(\mathbf{z}|\alpha)p(\alpha) = p(\mathbf{z}|\alpha)e^{-\alpha} = \prod_{i=1}^{N} p(z_i|\mathbf{z}_{1:i-1}, \alpha)e^{-\alpha} \tag{1}$$

Now, by the exchangeability of samples from a Chinese Restaurant Process, we can order the elements of $\mathbf{z}$ any way we like. In particular, suppose there are K occupied tables (in the argot of the CRP) with m_k customers sitting at table k, and that the first K customers all sit one-by-one at new tables. Since the probability of sitting at a new table is simply $\frac{\alpha}{j+\alpha}$, where j is the total number of customers already seated, the probability of K customers all sitting at new tables is given by

$$\frac{\alpha^K}{\prod_{j=0}^{K-1}(j+\alpha)} \tag{2}$$

If the next $m_1 - 1$ customers all sit at table 1, the first customer does so with probability $\frac{1}{K+\alpha}$, the next with probability $\frac{2}{K+1+\alpha}$ and so on. The same holds true for all other tables, with the appropriate number of customers-already-seated in the denominator. So the total probability of $\mathbf{z}$ given α is

$$\frac{\alpha^K}{\prod_{j=0}^{K-1}(j+\alpha)} \frac{(m_1-1)!}{\prod_{j=K}^{K+m_1-1}(j+\alpha)} \frac{(m_2-1)!}{\prod_{j=K+m_1}^{K+m_1+m_2-1}(j+\alpha)} \cdots \frac{(m_K-1)!}{\prod_{j=N-m_k}^{N-1}(j+\alpha)} \tag{3}$$

Which simplifies to

$$\frac{\alpha^K \prod_{l=1}^{K}(m_l-1)!}{\prod_{j=0}^{N-1}(j+\alpha)} = \alpha^K \prod_{l=1}^{K}(m_l-1)! \frac{\Gamma(\alpha)}{\Gamma(N+\alpha)} \tag{4}$$

And the propability of accepting the proposed α^* is simply

$$a(\alpha^*, \alpha) = min\left(1, \frac{\alpha^{*K}\Gamma(\alpha)\Gamma(N+\alpha^*)}{\alpha^K\Gamma(\alpha^*)\Gamma(N+\alpha)} e^{\alpha-\alpha^*}\right) \tag{5}$$

Which, remarkably, only depends on the assigned labels by the number of active components K.

Should we want to use an asymmetric proposal, for example one with variance proportional to the current α, we can easily modify the acceptance probability. Say we draw the proposal from a normal distribution with mean α and variance $\kappa\alpha$, then the acceptance probability becomes

$$a(\alpha^*, \alpha) = min\left(1, \frac{\alpha^{*K-1}\Gamma(\alpha)\Gamma(N+\alpha^*)}{\alpha^{K-1}\Gamma(\alpha^*)\Gamma(N+\alpha)} e^{\alpha-\alpha^*} e^{-\frac{1}{2}\frac{(\alpha-\alpha^*)^2}{\kappa^2}\left(\frac{1}{\alpha^{*2}}-\frac{1}{\alpha^2}\right)}\right) \tag{6}$$