AMS 241, Fall 2010, Homework 2

Vadim von Brzeski
vvonbrze@ucsc.edu

November 8, 2010

1 Gibbs Sampling Equations

The model we are studying is:

$$
\begin{aligned}
y_i|\theta_i, \phi &\sim k_N(y_i; \theta_i, \phi), \;\; i = 1, ..., n \\
\theta_i|G &\sim G, \;\; i = 1, ..., n \\
G|\alpha, \mu, \tau^2 &\sim DP(\alpha, G_0 = N(\mu, \tau^2)) \\
\alpha &\sim Gamma(a_\alpha, b_\alpha) \\
\mu &\sim N(a_\mu, b_\mu) \\
\tau^2 &\sim InvGamma(a_{\tau^2}, b_{\tau^2}) \\
\phi &\sim InvGamma(a_\phi, b_\phi)
\end{aligned}
$$

1.1 Posterior θ_i

$$
p(\theta_i|y_i, \{\theta_k, k \neq i\}, \alpha, \phi, \mu, \tau^2, data) = \frac{q_0 h(\theta_i|\phi, \mu\, \tau^2, y_i) + \sum_{j=1}^{n^{*-}} n_j^- q_j \delta_{\theta_j^{*-}}(\theta_i)}{q_0 + \sum_{j=1}^{n^{*-}} n_j^- q_j} \tag{1}
$$

where

$$
\begin{aligned}
q_j &= N(y_i; \theta_j^{*-}, \phi) \\
q_0 &= \alpha \int_{-\infty}^{\infty} k(y_i; \theta_i, \phi) g_0(\theta_i|\mu, \tau^2) d\theta_i \\
&= \alpha \int_{-\infty}^{\infty} N(y_i; \theta_i, \phi) N(\theta_i|\mu, \tau^2) d\theta_i \\
&= \alpha N(y_i; \mu, \phi + \tau^2)
\end{aligned}
$$

and

$$
\begin{aligned}
h(\theta_i|\phi, \mu\,\tau^2, y_i) &= C \cdot k_N(y_i|\theta_i, \phi) g_0(\theta_i|\mu, \tau^2) \\
&= C \cdot N(y_i; \theta_i, \phi) N(\theta_i|\mu, \tau^2) \\
&= N(\theta_i; m, v)
\end{aligned}
$$

where

$$
m = \frac{\mu/\tau^2 + y_i/\phi}{1/\tau^2 + 1/\phi} \qquad v = \frac{1}{1/\tau^2 + 1/\phi}
$$

So θ_i will either be a new distinct θ value drawn from $N(\theta_i; m, v)$ with probability proportional to q_0, or equal to an existing θ_j value with probability proportional to $n_j^- q_j$.

1.2 Posterior θ_j^*

$$
\begin{aligned}
p(\theta_j^*|w, n^*, \mu, \tau^2, \phi, data) &= C \cdot \left[\prod_{i:w_i=j} k(y_i|\theta_j^*, \phi)\right] g_0(\theta_j^*|\mu, \tau^2) \\
&= C \cdot \left[\prod_{i:w_i=j} N(y_i|\theta_j^*, \phi)\right] N(\theta_j^*|\mu, \tau^2) \\
&= N(\theta_j^*; m^*, v^*)
\end{aligned} \tag{2}
$$

where

$$
m^* = \frac{\mu/\tau^2 + n_j\bar{y}_j/\phi}{1/\tau^2 + n_j/\phi} \qquad v = \frac{1}{1/\tau^2 + n_j/\phi}
$$

So for each individual cluster (component) j, we indepedently draw a value for θ_j from the distribution defined in equation [2].

1.3 Posterior ϕ

$$
\begin{aligned}
p(\phi|\boldsymbol{\theta}, data) &= C \cdot p(\phi) \prod_{i=1}^{n} k(y_i|\theta_i, \phi) \\
&= C \cdot InvGamma(a_\phi, b_\phi) \prod_{i=1}^{n} N(y_i|\theta_i, \phi) \\
&= C \cdot \phi^{-(a_\phi+1)} exp(-b_\phi/\phi) \prod_{i=1}^{n} \phi^{-1/2} exp\left[\frac{-1}{2\phi}(y_i - \theta i)^2\right] \\
&= C \cdot \phi^{-(a_\phi+n/2+1)} exp(-b_\phi/\phi) exp\left[\sum_{i=1}^{n} \frac{-1}{2\phi}(y_i - \theta i)^2\right] \\
&= C \cdot \phi^{-(a_\phi+n/2+1)} exp\left[\frac{-1}{\phi}\left(b_\phi + \frac{nv}{2}\right)\right] \\
&= InvGamma(a_\phi + n/2, b_\phi + nv/2)
\end{aligned} \tag{3}
$$

where

$$nv = \sum_{i=1}^{n} (y_i - \theta_i)^2$$

1.4 Posterior μ, τ^2

$$\begin{aligned}
p(\mu, \tau^2 | \boldsymbol{\theta}^*) &= C \cdot p(\mu) p(\tau^2) \prod_{i=1}^{n^*} g_0(\theta_j^* | \mu, \tau^2) \\
&= C \cdot N(\mu | a_\mu, b_\mu) InvGamma(\tau^2 | a_{\tau^2}, b_{\tau^2}) \prod_{i=1}^{n^*} N(\theta_j^* | \mu, \tau^2)
\end{aligned}$$

We are not in a conjugate prior setting in this case for $p(\mu, \tau^2 | \boldsymbol{\theta}^*)$, but we can easily compute the conditional distributions needed for Gibbs sampling:

$$\begin{aligned}
p(\mu | \tau^2, \boldsymbol{\theta}^*) &= C \cdot N(\mu | a_\mu, b_\mu) \prod_{i=1}^{n^*} N(\theta_j^* | \mu, \tau^2) \\
&= N(m_\mu, s_\mu^2) \qquad (4)
\end{aligned}$$

where

$$m_\mu = \frac{a_\mu / b_\mu + n^* \bar{\boldsymbol{\theta}}^* / \tau^2}{1/b_\mu + n^*/\tau^2} \qquad s_\mu^2 = \frac{1}{1/b_\mu + n^*/\tau^2}$$

where n^* is the number of distinct θ^* values and $\bar{\boldsymbol{\theta}}^*$ is the mean of the distinct θ^* values.

$$p(\tau^2 | \mu, \boldsymbol{\theta}^*) = C \cdot InvGamma(\tau^2 | a_{\tau^2}, b_{\tau^2}) \prod_{i=1}^{n^*} N(\theta_j^* | \mu, \tau^2)$$

Using the same algebraic manipulation as was done for the posterior of ϕ above, we have:

$$p(\tau^2 | \mu, \boldsymbol{\theta}^*) = InvGamma(a_{\tau^2} + n^*/2, b_{\tau^2} + n^* \nu / 2) \qquad (5)$$

where

$$n^* \nu = \sum_{i=1}^{n^*} (\theta_i^* - \mu)^2$$

1.5 Posterior α

Here we just re-iterate the sampling scheme for α using the auxiliary variable scheme outlined by Escobar and West, 1995. First we draw η from :

$$p(\eta | \alpha, data) = Beta(\alpha + 1, n) \qquad (6)$$

Define probability $p = (a_\alpha + n^* - 1)/(n(b_\alpha - \log(\eta)) + a_\alpha + n^* - 1)$, and draw:

$$\begin{aligned}
&with\, probability\, p \;: && \alpha | \eta, n^*, data \sim Gamma(a_\alpha + n^*, b_\alpha - \log(\eta)) \\
&with\, probability\, 1 - p \;: && \alpha | \eta, n^*, data \sim Gamma(a_\alpha + n^* - 1, b_\alpha - \log(\eta)) \qquad (7)
\end{aligned}$$

1.6 Overview of Gibbs Sampling Scheme

A brief overview of the Gibbs sampling scheme is shown below. In our sampler, we maintain a *state* structure which consists of the θ_i values, the cluster (component) membership index for each y_i, and the values of all the parameters $\alpha, \mu, \tau^2, \phi$.

1.6.1 Initialization

At initialization, we assign each of the N y_i observations to belong to a different cluster, thus starting out with N separate clusters (components). Each cluster is initialized with a θ_j^* drawn from $g_0 = N(\mu, \tau^2)$ distribution, given the starting values of μ, τ^2.

1.6.2 MCMC Simulation

We typically run for 1000 burn-in iterations, followed by 4000 monitoring iterations. We arrived at the 1000/4000 values after preliminary experiments that showed that increasing iterations (e.g. 2000/10000) did not appreciably change the results, indicating that the chain had reached convergence.

The sampling scheme consists of :

- For $i = 1..n$, sample θ_i from the mixed distribution defined by (1). Adjust cluster (component) membership indeces of the data if an existing component was dropped or if a new component was added.
- For $j = 1..n^*$, sample θ_j^* from (2).
- Sample ϕ from (3).
- Sample $\mu|\tau^2$ from (4) and $\tau^2|\mu$ from (5).
- Sample α using (6) and (7).

1.7 Predictive Distribution Calculations

The posterior predictive distribution for a new observation θ_0 is given by the Polya-Urn scheme:

$$p(\theta_0|n^*, \boldsymbol{w}, \boldsymbol{\theta}^*, \alpha, \phi) = \frac{\alpha}{\alpha + n} G_0(\theta_0|\phi) + \frac{1}{\alpha + n} \sum_{j=1}^{n^*} n_j \delta_{\theta_j^*}(\theta_0) \tag{8}$$

The posterior predictive distribution for a new y_0 is given by :

$$p(y_0|data) = \int\int k(y_0; \theta_0, \phi) p(\theta_0|n*, \boldsymbol{w}, \boldsymbol{\theta}^*, \alpha, \phi) p(n^*, \boldsymbol{w}, \boldsymbol{\theta}^*, \alpha, \phi|data) \tag{9}$$

Thus given B sets of samples from our MCMC output, we can obtain samples from $p(y_0|data)$ as follows: for each set b of posterior parameter values in the MCMC output, first draw $\theta_{0,b}$ from $p(\theta_0|n*, \boldsymbol{w}, \boldsymbol{\theta}^*, \alpha, \phi)$, and then draw $y_{0,b}$ from $p(y_0|\theta_{0,b}, \phi_b) = N(\cdot; \theta_{0,b}, \phi_b)$

Samples from the prior predictive distribution $p(y_0)$ are obtained as follows:

- Draw (μ, τ^2) from $p(\mu)p(\tau^2)$. Since $p(\mu)$ and $p(\tau^2)$ are independent of each other, $p(\mu, \tau^2)$ does not form a conjugate prior for $G_0 = N(\mu, \tau^2)$. So we independently sample b times:
- $\tau_b^2 \sim InvGamma(a_{\tau^2}, b_{\tau^2})$
- $\mu_b \ \sim N(a_\mu, b_\mu)$
- $\theta_{0,b} \sim G_0 = N(\mu, \tau^2)$
- $y_{0,b} \sim \int N(y_{0,b}; \theta_{0,b}|\phi)p(\phi)d\phi = t_\nu(\theta_{0,b}, s^2)$, a t distribution with $\nu = 2a_\phi$ degrees of freedom, mean $\theta_{0,b}$ and scale $s^2 = b_\phi/a_\phi$

  The derivation of $y_{0,b} \sim t_\nu(\theta_{0,b}, s^2)$ is shown below:

$$
\begin{aligned}
\int N(y;\theta|\phi)IG(\phi;a,b)d\phi &= \int \frac{1}{\sqrt{2\pi\phi}} exp\left[\frac{-1}{2\phi}(y-\theta)^2\right]\frac{b^a}{\Gamma(a)}\phi^{-(a+1)}exp\left[\frac{-b}{\phi}\right]d\phi \\
&= C \cdot \int \phi^{-(a+1+1/2)} exp\left[\frac{-1}{2\phi}((y-\theta)^2+2b)\right]d\phi \\
Letting \quad z &= \frac{A}{2\phi} \\
\phi &= \frac{A}{2z} \\
|d\phi| &= \frac{A}{2z^2}dz \\
where \quad A &= ((y-\theta)^2+2b) \quad we \quad get \\
&= C \cdot \int \left(\frac{A}{2z}\right)^{-(a+1+1/2)} exp[-z]\frac{A}{2z^2}dz \\
&= C \cdot \left(\frac{A}{2}\right)^{-(a+1/2)} \int z^{(a+1/2-1)}exp(-z)dz \\
&= C \cdot \left(\frac{A}{2}\right)^{-(a+1/2)} \Gamma(a+1/2) \\
&= C \cdot \Gamma(a+1/2)\left[\frac{1}{2}(y-\theta)^2+b\right]^{-(a+1/2)} \\
&= C \cdot \Gamma(a+1/2)b^{-(a+1/2)}\left[1+\frac{(y-\theta)^2}{2b}\right]^{-(a+1/2)}
\end{aligned}
$$

  which we recognize as the kernel of a t-distribution. Letting $a = a_\phi = \nu/2$ and $b = b_\phi = \nu s^2/2$ and filling in the constants, we get the familiar form of the t-distribution, namely:

$$
y_{0,b} \sim \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})}\frac{1}{\sqrt{\pi\nu}s}\left[1+\frac{(y-\theta)^2}{\nu s^2}\right]^{-\frac{\nu+1}{2}}
$$