Markov Chain Monte Carlo

February 26, 2019

Metropolis updates require lots of computation time because of the several steps required to calculate likelihoods, compute ratios and choose proposals vs current values. Gibbs sampling allows us to streamline that process by making proposals that are so smart that we retain *all* of them

We can usefully illustrate Gibbs updates by showing how we would use them to estimate the posterior distribution of the mean of a normally distributed random variable. We will call this mean θ. Recall that draws of the random variable y_i from the normal distribution with mean θ arise as

$$y_i \sim \text{normal}(\theta, \varsigma^2). \tag{1}$$

We can think of y_i, of course, as an observation on some ecological process. For this example, we begin by assuming that the variance of the observations, ς^2 is *known.* It is important to understand the "knowing" ς^2 is not the same as calculating it as the variance of a sample dataset. Rather we are treating it here as a fully observed quantity, as if we had calculated it from *all* of the potential observations. In the following discussion, it is particularly important to keep in mind that ς^2 is the variance of the distribution of the *observations* (y_i), not the variance of the distribution of the mean of the observations (θ). We have prior information about θ,

$$\theta \sim \text{normal}\left(\mu_0, \sigma_0^2\right). \tag{2}$$

This information might be informative or vague. We have a data set $\mathbf{y}$ with n observations. This allow us to formulate the posterior and joint distribution, our universal starting point, as

$$[\theta, \varsigma^2 \mid \mathbf{y}] \propto \prod_{i=1}^{n} \text{normal}(y_i|\theta, \varsigma^2)\text{normal}(\theta \mid \mu_0, \sigma_0^2)\text{inverse gamma}(\alpha_0, \beta_0) \tag{3}$$

where $\mu_0, \sigma^2$and α_0, β_0 are numeric arguments. They are known.

We use equation 3 to obtain the full-conditional distribution of θ,

$$[\theta|\cdot] \propto \prod_{i=1}^{n} \text{normal}\left(y_i|\theta, \varsigma^2\right) \text{normal}\left(\theta|\mu_0, \sigma_0^2\right). \tag{4}$$

For notational convenience, we define μ_1 and σ_1^2 as the parameters of the conditional posterior distribution of θ, that is

$$[\theta|\cdot] \propto \text{normal}(\mu_1, \sigma_1^2). \tag{5}$$

Note that σ_1^2 is the updated variance of the distribution of the *mean* not the variance of the distribution of the *observations*, which of course is ς^2. Note that we temporarily treating ς^2 as known. Equation 4 shows that we have a normal likelihood for the mean with known variance and a normal prior on the mean, which are conjugates. When this is the case, we can calculate the parameters of the conditional posterior distribution of θ directly using the formulas

$$\mu_1 = \frac{\left(\frac{\mu_0}{\sigma_0^2} + \frac{\sum_{i=1}^{n} y_i}{\varsigma^2}\right)}{\left(\frac{1}{\sigma_0^2} + \frac{n}{\varsigma^2}\right)} \tag{6}$$

$$\sigma_1^2 = \left(\frac{1}{\sigma_0^2} + \frac{n}{\varsigma^2}\right)^{-1}. \tag{7}$$

Notice that every quantity on the right hand side of these equations is known; μ_0 and σ_0^2 are known as priors. The y_i are observations in hand, and we are assuming (for now) that ς^2 is known. So, we have all we need to know to make a draw from the distribution of θ using equation 5 because we know μ_1 and σ_1^2. So, why wouldn't we just use equations 6 and 7 to estimate the parameters of the posterior of θ and be done with it? Because we must assume that ς^2 is known, which is virtually never the case. Somehow we must learn about ς^2 to find the marginal posterior of θ.

So, what about ς^2? Again the observations arise from $y_i \sim \text{normal}\left(\theta, \varsigma^2\right)$ and we seek to understand the conditional[1] posterior distribution of ς^2. If we assume that θ is known, then

$$\left[\varsigma^2|\cdot\right] \propto \prod_{i=1}^{n} \text{normal}\left(y_i|\theta,\varsigma^2\right) \text{inverse gamma}\left(\varsigma^2|\alpha_0,\beta_0\right). \tag{8}$$

We define the parameters of the full-conditional distribution of ς^2 as α_1 and β_1 for notational convenience so that

$$\left[\varsigma^2|\cdot\right] \propto \text{inverse gamma}\left(\alpha_1,\beta_1\right). \tag{9}$$

We have a normal likelihood with a known mean and unknown variance and an inverse gamma prior on the variance. When this is true we can compute the parameters of the full-conditional distribution of ς^2 using

$$\alpha_1 = \alpha_0 + \frac{n}{2} \tag{10}$$

$$\beta_1 = \beta_0 + \frac{\sum_{i=1}^{n}\left(y_i-\theta\right)^2}{2}. \tag{11}$$

Again, remember that β_0 and α_0 are known arguments to priors, so in practice they would be numeric. It follows that all quantities on the right hand side of equations 10 and 11 are known.

It might seem that we have tied ourselves in a knot. We need to know ς^2 to estimate θ and we need to know θ to estimate ς^2. This is just the kind of problem that MCMC can solve because at each step in the chain we pretend all of the parameters save one are *known*. Equations 6 - 11 give us all we need to construct a very fast sampler for θ and ς^2. Define k as the iteration in the chain. So, element 100 in the chain is indexed by $k = 100$. Be sure you understand that k is a superscript not an exponent. The algorithm is:

1. Use the current value of $\varsigma^{2(k)}$ to calculate $\mu_1^{(k+1)}$ and $\sigma_1^{2(k+1)}$ from equations 6 and 7. Make a draw from $\theta^{(k+1)} \sim \text{normal}\left(\mu_1^{(k+1)},\sigma_1^{2(k+1)}\right)$ and store it in the chain.

2. Use the updated value of $\theta^{(k+1)}$ to calculate $\alpha_1^{(k+1)}$ and $\beta_1^{(k+1)}$ using equations 10 and 11. Make a draw from $\varsigma^{2(k+1)} \sim \text{inverse gamma}\left(\alpha_1^{(k+1)},\beta_1^{(k+1)}\right)$ and store it in the chain.

3. Repeat 1-2.

[1] The distribution is conditional because we must know θ.

A sufficient number of repetitions usually converges on the posterior distributions of θ and ς^2 much more quickly than if we used Metropolis-Hastings. However, the estimates would be the same.