BDA 15: Multiparameter Models

Often we'll have to deal with models containing multiple unknowns. It's usually the case that within these models we're only interested in the posterior distribution of one parameter, which we call the marginal posterior distribution of that parameter.

To obtain the marginal posterior it's first necessary to model the joint posterior distribution over all unknowns. We can then integrate over the other unknowns to get the desired marginal posterior.

The computational equivalent of this procedure is to draw many samples from the joint posterior, and record the parameter of interest each time. In doing this, it's possible to build up a frequency distribution that approximates the marginal posterior of the parameter.

Consider the case where we seek to model the joint distribution of two parameters. We can get hold of the unnormalized posterior by taking the product of the joint prior and the sampling distribution:

$$p(\theta_1, \theta_2 | y) \; \propto \; p(y | \theta_1, \theta_2) \cdot p(\theta_1, \theta_2) \tag{1}$$

Let's say we want the marginal posterior of θ_1, $p(\theta_1 | y)$. We can get this by integrating the joint posterior over θ_2

$$p(\theta_1 | y) \; = \; \int p(\theta_1, \theta_2 | y) \cdot d\theta_2 \tag{2}$$

We can re-express the right hand side in terms of a conditional posterior and another marginal posterior

$$p(\theta_1 | y) \; = \; \int p(\theta_1 | \theta_2, y) \cdot p(\theta_2 | y) \cdot d\theta_2 \tag{3}$$

I think this form makes it clearer what we're doing - we're taking the weighted sum of θ_1's posterior densities condtional on θ_2, where the weights correspond to the posterior probabilities of the relevant value of θ_2.

Factoring the marginal posterior in this way also reveals a possible way to simulate it. We can first model the posterior distribution of θ_2, draw a sample θ_2, then use this to evaluate $p(\theta_1 | \theta_2, y)$ from which we can draw a value of θ_1. If we repeat this process many times, the resulting distribution will be similar to $p(\theta_1 | y)$

Now for an example. Say we're trying to model the mean μ of a Normally distributed random variable Y for which we don't know the variance. σ^2. We can formulate the marginal posterior as

$$p(\mu | y) = \int p(\mu | \sigma^2, y) \cdot p(\sigma^2 | y) \cdot d\sigma^2 \tag{4}$$

So if we're to follow the procedure outlined above in which we sample from $p(\sigma^2 | y)$, then from $p(\mu | \sigma^2, y)$, we'll need expressions for both of these terms. $p(\mu | \sigma^2, y)$ is quite straightforward:

$$p(\mu | \sigma^2, y) \propto p(y | \mu, \sigma^2) \cdot p(\mu | \sigma^2) \tag{5}$$

$$p(\mu | \sigma^2, y) \propto \prod_{i=1}^{n} \frac{1}{\sigma} \exp\left(-\frac{(y^{(i)} - \mu)^2}{2\sigma^2}\right) \tag{6}$$

We assume that the location and scale are independent so that $p(\mu | \sigma^2) = p(\mu) = \text{Unif}(-\infty, \infty) = \texttt{constant}$. We also assume that the observations are independent.

$p(\sigma^2|y)$ is a little more tricky to derive. Again we use the fact that

$$p(\sigma^2|y) \propto p(y|\sigma^2) \cdot p(\sigma^2) \tag{7}$$

For $p(y|\sigma^2)$ we can integrate $p(y|\sigma^2, \mu)$ over μ

$$p(y|\sigma^2) = \int \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y^{(i)} - \mu)^2}{2\sigma^2}\right) d\mu \tag{8}$$

$$= \frac{1}{(2\pi\sigma^2)^{n/2}} \cdot \exp\left(-\frac{(n-1)s^2}{2\sigma^2}\right) \cdot \int \exp\left(-\frac{n(\bar{y} - \mu)^2}{2\sigma^2}\right) d\mu \tag{9}$$

$$= \frac{1}{(2\pi\sigma^2)^{n/2}} \cdot \exp\left(-\frac{(n-1)s^2}{2\sigma^2}\right) \cdot \sqrt{2\pi} \frac{\sigma}{\sqrt{n}} \tag{10}$$

$$= \frac{1}{\sqrt{n} \cdot (2\pi\sigma^2)^{\frac{n-1}{2}}} \cdot \exp\left(-\frac{(n-1)s^2}{2\sigma^2}\right) \tag{11}$$

To derive this we made use of the fact that the integral in (9) is Normal and that the sample variance of our observations is $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y^{(i)} - \bar{y})$. We can take a noninformative prior on σ^2 so that

$$p(\sigma^2) \propto (\sigma^2)^{-1} \tag{12}$$

This prior is appropriate when the d.o.f. of the data dwarfs the d.o.f. of the usual conjugate prior. Using (12) and (11) we can now express (7):

$$p(\sigma^2|y) \propto (\sigma^2)^{-\frac{n+1}{2}} \cdot \exp\left(-\frac{(n-1)s^2}{2\sigma^2}\right) \tag{13}$$

This has the form of an inverse gamma function (i.e. $\frac{1}{\sigma^2} \sim \text{Gamma}(a, b)$).

Finally we're in a position where can do some modelling. To reiterate, the process we're going to follow is:

1. Sample σ^2 from $p(\sigma^2|y)$.

2. Use these values of σ^2 to draw samples of μ from $p(\mu|\sigma^2, y)$.

3. Chart the frequency of the μ values in a histogram. This distribution will be representative of the marginal posterior $p(\mu|y)$.

Before we do anything however, we'll need some data to work with. We'll use ten data points drawn from a Normal distribution with a true mean of 50 and a true variance of 10.

```
obs_y <- rnorm(n = 10, mean = 50, sd = sqrt(10))
```