NRES 779: Bayesian Hierarchical Modeling in Natural Resources

Lab 05: Markov Chain Monte Carlo

Contents

I Motivation

This problem challenges you to understand how the wickedly clever Markov chain Monte Carlo algorithm takes a multivariate joint distribution and breaks it into a series of univariate, marginal distributions that can be approximated one at a time. There are only two unknowns in this problem, but the same principles and approach would apply if there were two hundred. The accompanying document, MCMCMath1.pdf, describes the math that stands behind the coding that you will do here. You should study this thoroughly before proceeding.

II Problem

You will write code using conjugate relationships, also known as Gibbs updates, to draw samples from marginal posterior distributions of a mean (μ) and variance (σ^2). Specifically, you are coding your own MCMC sampler. This is what JAGS does behind the scenes.

The model statement is

$$
\begin{aligned}
y_i &\sim \text{Normal}(\mu, \sigma^2) \\
\mu &\sim \text{Normal}(\mu_0, \sigma_0^2) \\
\sigma^2 &\sim \text{IG}(\alpha, \beta)
\end{aligned}
$$

NOTE: We are using the shape and scale parameterization for α and β, respectively, in the inverse gamma distribution.

III Preliminaries

1. Set the seed for random numbers in R with `set.seed(1991)`.
2. Simulate 100 data points from a normal distribution with mean $\mu = 100$ and variance $\sigma^2 = 25$. You will use these data to test the conjugate functions that you will write in section IV below. Call the data set y. Be careful here. R requires the standard deviation, not the variance, as a parameter.

   Simulating data is always a good way to test methods. Your method should be able to *recover* the generating parameters. *Recover* is usually taken to mean that if you simulated 100 data sets, 95% of the time, the 95% credible intervals would cover the true parameters used to simulate the data. If your intuition is telling you that this seems like a frequentist way of doing things, you are right! It's often referred to as *Calibrated Bayes*, which combines Bayesian and Frequentist ideas (see Little, 2006, for a very interesting article on a Bayes/Frequentist roadmap). **However, there is no theory that dictates a Bayesian credible interval must cover a parameter that was used to simulate the data** Why? It has to do with priors.
3. Write a function called `draw.mean` that:
   - calculates the parameters of the posterior distributions for μ using a normal-normal conjugate relationship where the variance is assumed to be known.
   - Make a random draw of from that distribution.
4. Write a function called `draw.var` that:
   - calculates the parameters of the posterior distributions for σ^2 using a inverse gamma–normal conjugate relationship where the mean is assumed to be known.
   - Makes a random draw from that distribution.

IV Writing a Sampler

Now execute these steps:

5. Set up a matrix for storing samples from the posterior distribution of the mean. The number of columns should equal the number of chains (3) and number of rows should equal the number of iterations (10,000). Do the same thing for storing samples from the posterior distribution of the variance.
6. Assign initial values to the first row of each matrix, a different value for each of the chains. These can be virtually any value within the support of the random variable.
7. Set up nested `for` loops to iterate from **two** to the total number of iterations for each of the three chains for each parameter (recall we already selected values for the first row–these were our initial values). Use the conjugate functions `draw.mean` and `draw.var` to draw a sample from the distribution of the mean using the value of

the variance at the previous current iteration. Then make a draw from the variance using the value of the mean from the current iteration. Repeat. Assume vague priors for the mean and variance:

$$\begin{aligned} \mu &\sim \text{Normal}(\mu_0 = 0, \sigma_0^2 = 10000) \\ \sigma^2 &\sim \text{IG}(\alpha = 0.01, \beta = 0.01) \end{aligned} \tag{1}$$

V Trace Plots & Plots of Marginal Posterior Distributions

8. Discard the first 1000 iterations as *burn-in*. On the same figure, make three plots (one for each chain) with the value of the mean as a function of iteration number. This is called a trace plot.
9. For all chains combined, make a histogram of the samples of the mean retained after burn-in. Put a vertical line on the plot showing the generating value.
10. Repeat steps 8-9 for the variance.
11. For both μ and σ^2, calculate the mean of all the chains combined and its standard deviation. Interpret these quantities.
12. Compare the standard deviation of the posterior distribution of μ with an approximation using the standard deviation of the data divided by the square root of the sample size. What is this approximation called in the frequentist world?
13. Vary the number of values in the simulated data set, e.g., $n = 10, 100, 1000$. What happens to the mean and variance of the posterior distributions as n gets large? Explain why you think this happens.
14. Make the burnin = 1 instead of 1000. Does this change your results? Why or why not?
15. Reverse the order of the conjugate functions in step 7 so that the variance is drawn first followed by the mean. Be careful, this involves a bit more than simply reversing the order of the functions in the loop. Does this reordering have an effect on the posteriors? Why or why not?

References

Little, R. J. 2006. Calibrated Bayes: a Bayes/frequentist roadmap. The American Statistician **60**:213–223.