Stat 471: Lecture 25
Bayesian Inference II.

Moo K. Chung
mchung@stat.wisc.edu

November 7, 2003

1. Consider a simple linear model

$$Y_i = \beta_0 + \epsilon_i,\ i = 1, \cdots, n$$

where $\epsilon_i \sim N(0,1)$. Suppose 3, 2, 4, 2, 1 are observations. We know somehow that the prior to be $\pi(\beta_0) \sim N(3,1)$. Estimate β_0 using MCMC.

Solution. This is a dumb example but serves the purpose of illustrating the use of MCMC for parameter estimation. Note that $Y_i|\beta_0 \sim N(\beta_0, 1)$. The posterior is

$$\begin{aligned} \pi(\beta_0|y) &= cf(y_1, \cdots, y_n|\beta_0)\pi(\beta_0) && (1)\\ &= cf(y_1|\beta_0)\cdots f(y_n|\beta_0)\pi(\beta_0) && (2)\end{aligned}$$

where c is a normalizing constant, i.e.

$$c^{-1} = \int_{-\infty}^{\infty} f(y_1, \cdots, y_n|\beta_0)\pi(\beta_0)\, d\beta_0.$$

Then the posterior is

$$\pi(\beta_0|y) \propto \exp\Big(-\frac{(\beta_0 - 3)^2}{2}\Big)\exp\Big(-\frac{\sum_{i=1}^n(\beta_0 - y_i)^2}{2}\Big).$$

The Bayes estimate $\hat{\beta}_0 = \mathbb{E}(\beta_0|y)$ can be computed analytically in this particular example but requires some algebraic manipulation and if the prior is something like $\pi(\beta_0) \propto \frac{1}{1+\beta_0^2}$, it may not even possible to compute the Bayes estimate analytically.

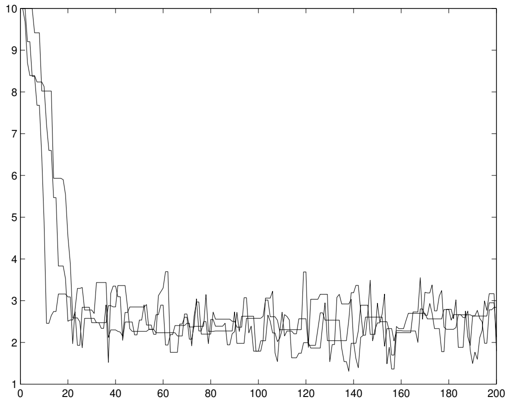
This gives a motivation for computing $\hat{\beta}_0$ via MCMC. We use the random-walk Metropolis algorithm with symmetric propositional density $q(y|x) \sim N(x,1)$. For this particular distribution

$$q(y|x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{(y-x)^2}{2}} = q(x|y).$$

So function α in the MCMC algorithm is trivially

$$\alpha(a,b) = \min\Big[1, \frac{\pi(a|y)}{\pi(b|y)}\Big].$$

Then we generate Markov chain $X_0 \rightarrow X_1 \rightarrow X_2 \cdots$ with invariant distribution $\pi(\beta_0|y)$. For large i, $X_i \sim \pi(\beta_0|y)$.



```
n=200;
y=[3 2 4 2 1];
X(1)=10;
for i=2:n
  candidate=normrnd(X(i-1),1);
  pi_n=normpdf(candidate,3,1)*
       prod(normpdf(candidate,y,1));
  pi_d=normpdf(X(i-1),3,1)*
       prod(normpdf(X(i-1),y,1));
  alpha=min([1,pi_n/pi_d]);
  u=rand;
  if u <= alpha
    X(i)=candidate;
  else
    X(i)=X(i-1);
  end;
end
>> mean(X(50:200))
ans =
    2.4498
>> mean(y)
ans =
    2.4000
```

2. *Hierarchical model.* Assume $Y_i \sim N(\beta_0, 1)$ in the previous model. Give prior $\beta_0 \sim \pi(\beta_0|\gamma)$, $\gamma \sim \pi(\gamma)$. In this case, we construct the Gibbs sampler

$$\beta_0 \sim \pi(\beta_0|y, \gamma),$$

$$\gamma \sim \pi(\gamma|y, \beta_0).$$