STAT 578: Advanced Bayesian Modeling

Week 7 – Lesson 3

Metropolis and Metropolis-Hastings

Fall 2019

Metropolis Example

Normal Sample

Simple Bayesian model for a normal sample:

$$\begin{aligned} y_1, \ldots, y_n \mid \mu, \sigma^2 &\sim \text{iid N}(\mu, \sigma^2) \\ \mu &\sim \text{N}(\mu_0, \tau_0^2) \\ \sigma^2 &\sim \text{Inv-}\chi^2(\nu_0, \sigma_0^2) \end{aligned}$$

with μ and σ^2 independent under the prior.

Parameter space of $\theta = (\mu, \sigma^2)$ is embedded in two-dimensional real space.

Proposal Choice

For illustration, use bivariate normal proposal $J(\theta' \mid \theta)$:

$$\theta' \quad \sim \quad \mathrm{N}\big(\,\theta,\ \rho I\,\big)$$

- I is the 2×2 identity matrix.
- $\rho > 0$ adjusts width of the proposal.

Note: Can propose values outside of parameter space ($\sigma^2 \leq 0$), but they will never be accepted

(Alternative: Use transformed parameter $(\mu, \log \sigma^2)$.)

Ratio

Since the proposal is symmetric, the algorithm will be Metropolis.

The ratio is

$$r \quad = \quad \frac{p\big(\mu^*, (\sigma^2)^* \mid y\big)}{p\big(\mu^{t-1}, (\sigma^2)^{t-1} \mid y\big)} \quad = \quad \frac{p(\mu^*)\, p\big((\sigma^2)^*\big)\, p\big(y \mid \mu^*, (\sigma^2)^*\big)}{p(\mu^{t-1})\, p\big((\sigma^2)^{t-1}\big)\, p\big(y \mid \mu^{t-1}, (\sigma^2)^{t-1}\big)}$$

since the normalizing factor cancels.

Recall the likelihood:

$$p(y \mid \mu, \sigma^2) \quad \propto \quad \frac{1}{(\sigma^2)^{n/2}} \exp\left(-\frac{(n-1)s^2}{2\sigma^2}\right) \exp\left(-\frac{n}{2\sigma^2}(\mu - \bar{y})^2\right)$$

Data Example

Recall the Flint water data:

$$y_i \quad = \quad \text{logarithm of first-draw lead level (ppb), for observation } i$$

Sample statistics:

$$n = 271 \qquad \bar{y} \approx 1.40 \qquad s^2 \approx 1.684$$

For illustration, let:

$$\mu_0 = 1.10 \qquad \tau_0^2 = \sigma_0^2 = 1.17 \qquad \nu_0 = 1$$

For implementation in R:

```
> n <- 271
> ybar <- 1.40
> s.2 <- 1.684

> mu0 <- 1.10
> tau.2.0 <- sigma.2.0 <- 1.17
> nu0 <- 1
```

Define some convenience functions:

```
> dinvchisq <- function(x, df, scalesq){  # inverse chi-square density
+   if(x>0)
+     ((df/2)^(df/2) / gamma(df/2)) * sqrt(scalesq)^df * x^(-(df/2+1)) *
+       exp(-df*scalesq/(2*x))
+   else 0
+ }

> likelihood <- function(mu, sigma.2){
+   sigma.2^(-n/2) * exp(-(n-1)*s.2/(2*sigma.2)) *
+     exp(-n*(mu-ybar)^2/(2*sigma.2))
+ }
```

Note: Likelihood needed only up to proportionality in the parameters.

Define function for evaluating ratio r:

```
> ratio <- function(mu.prop, sigma.2.prop, mu.old, sigma.2.old){
+   dnorm(mu.prop,mu0,sqrt(tau.2.0)) * dinvchisq(sigma.2.prop,nu0,sigma.2.0) *
+     likelihood(mu.prop,sigma.2.prop)  /
+   (dnorm(mu.old,mu0,sqrt(tau.2.0)) * dinvchisq(sigma.2.old,nu0,sigma.2.0) *
+     likelihood(mu.old,sigma.2.old))
+ }
```

Note: `dnorm` is density of a normal distribution.

(In practice, should implement more carefully to avoid numerical problems.)

Set up storage space, choose ρ, and initialize chain:

```
> n.sim <- 1000
> mu.sim <- numeric(n.sim)
> sigma.2.sim <- numeric(n.sim)
> accept.prob <- numeric(n.sim-1)

> rho <- 0.01

> mu.sim[1] <- 2          # starting value
> sigma.2.sim[1] <- 2.5   # starting value
```

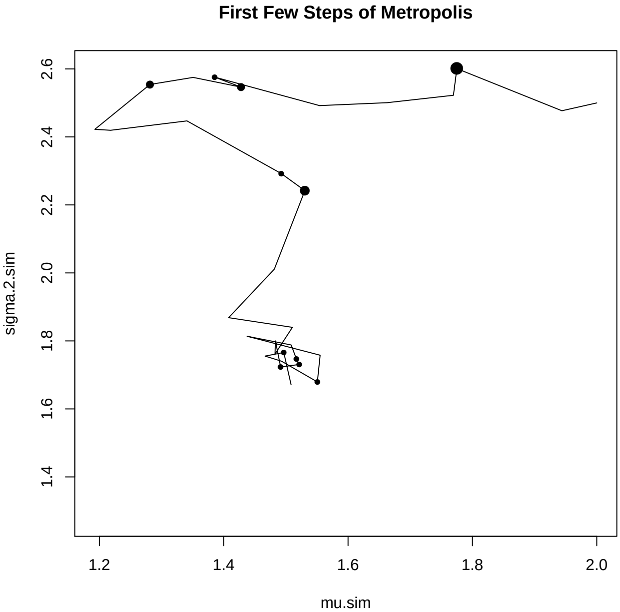
Run Metropolis algorithm:

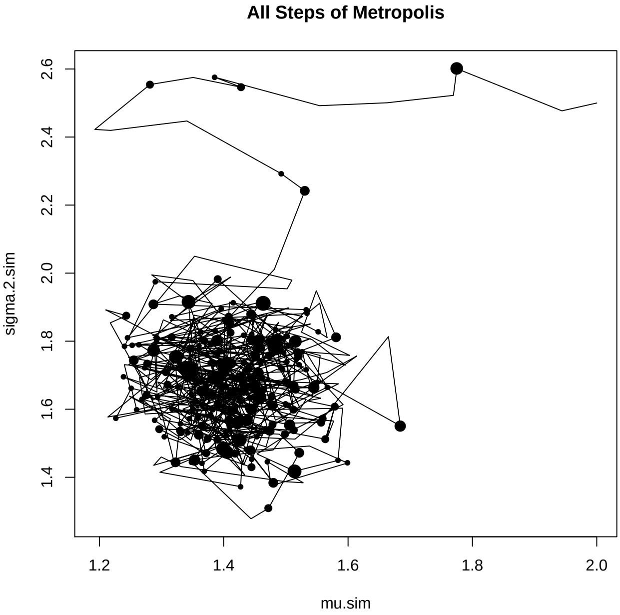
```
> for(t in 2:n.sim){
+   mu.prop <- rnorm(1, mu.sim[t-1], sqrt(rho))
+   sigma.2.prop <- rnorm(1, sigma.2.sim[t-1], sqrt(rho))
+
+   accept.prob[t-1] <-
+     min(ratio(mu.prop,sigma.2.prop,mu.sim[t-1],sigma.2.sim[t-1]), 1)
+   if(runif(1) < accept.prob[t-1]){
+     mu.sim[t] <- mu.prop
+     sigma.2.sim[t] <- sigma.2.prop
+   }else{
+     mu.sim[t] <- mu.sim[t-1]
+     sigma.2.sim[t] <- sigma.2.sim[t-1]
+   }
+ }
```

Note: `runif(1)` samples once from $\mathrm{U}(0,1)$

```
> mean(accept.prob)
[1] 0.5296723
```

Acceptance rate is higher than optimal – could improve by increasing ρ.


First Few Steps of Metropolis
sigma.2.sim
mu.sim
1.4
1.6
1.8
2.0
2.2
2.4
2.6
1.2
1.4
1.6
1.8
2.0



All Steps of Metropolis
sigma.2.sim
mu.sim
2.6
2.4
2.2
2.0
1.8
1.6
1.4
1.2
1.4
1.6
1.8
2.0

Warning: Improper Priors

Being able to construct a Markov chain with Gibbs, Metropolis, or Metropolis-Hastings does not automatically imply the posterior is proper!

When using an improper prior, propriety of the posterior must be checked separately.

Slice Sampling

A **slice sampler** is another MCMC technique.

It essentially uses special auxiliary variables (later).

Not covered in BDA3, but sometimes used by JAGS.