

Stat 8931, Fall 2005  
Homework 2  
Due Oct 5, 2005

**Q1** Do a Gibbs sampler for the problem described in Section 10.6 (or whatever the “Examples” subsection of “The Gibbs Update” section has turned into) of the lecture notes.

The likelihood is a function of the sufficient statistics

$$\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^n x_i$$
$$\hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{\mu}_n)^2$$

Suppose we have observed

$$n = 20$$
$$\hat{\mu}_n = 14.731$$
$$\hat{\sigma}_n^2 = 4.814$$

Suppose that we have independent priors for  $\mu$  and  $\lambda = 1/\sigma^2$  (the “precision”) which are

$$\mu \sim \text{Normal}(10, 2^2)$$
$$\lambda \sim \text{Gamma}(3, 1)$$

Although this is the kind of problem that WinBUGS is designed for, write your own Gibbs sampler in R.

Run the Gibbs sampler and get posterior expectation and MCSE for the functions of the state

- $\mu$  (the first component)
- $\sigma = \lambda^{-1/2}$  where  $\lambda$  is the second component.

As in the first assignment, run the Markov chain long enough to obtain MCSE smaller than 0.001 for each.

**S1** The first task is to derive the unnormalized posterior (almost given in the notes) and the “full conditionals” of each variable given the other. The unnormalized joint posterior density is

$$\begin{aligned} & -\frac{\lambda}{2} \sum_{i=1}^n (x_i - \hat{\mu}_n)^2 - \frac{n\lambda}{2} (\hat{\mu}_n - \mu)^2 + \frac{n}{2} \log(\lambda) - \frac{l}{2} (\mu - m)^2 + (a-1) \log(\lambda) - b\lambda \\ & = -\frac{n\lambda}{2} \hat{\sigma}_n^2 - \frac{n\lambda}{2} (\hat{\mu}_n - \mu)^2 + \frac{n}{2} \log(\lambda) - \frac{l}{2} (\mu - m)^2 + (a-1) \log(\lambda) - b\lambda \end{aligned}$$

The posterior precision of  $\mu$  given  $\lambda$  is the coefficient of  $-\mu^2/2$  in the above, which is  $l + n\lambda$ . The posterior mean of  $\mu$  given  $\lambda$  is the coefficient of  $\mu$  divided by the posterior precision, which is  $(lm + n\lambda\hat{\mu}_n)/(l + n\lambda)$ . Thus we have discovered one “full conditional”

$$\mu \mid \lambda \sim \text{Normal} \left( \frac{lm + n\lambda\hat{\mu}_n}{l + n\lambda}, \frac{1}{l + n\lambda} \right)$$

The shape parameter of  $\lambda$  given  $\mu$  is one plus the coefficient of  $\log(\lambda)$  in the above, which is  $n/2 + a$ . The inverse scale parameter of  $\lambda$  given  $\mu$  is the coefficient of  $-\lambda$  in the above, which is  $b + n\hat{\sigma}_n^2/2 + n(\hat{\mu}_n - \mu)^2/2$ . Thus we have discovered the other “full conditional”

$$\lambda \mid \mu \sim \text{Gamma} \left( a + \frac{n}{2}, b + \frac{n\hat{\sigma}_n^2 + n(\hat{\mu}_n - \mu)^2}{2} \right)$$

First we assign the data.

```
> n <- 20
> mu.hat <- 14.731
> sig2.hat <- 4.814
> m <- 10
> l <- 1/(2^2)
> a <- 3
> b <- 1
```

Then devise the Gibbs updates.

```
> mu.up <- function(lambda) {
+   muprec <- l + n * lambda
+   mumu <- (l * m + n * lambda * mu.hat)/muprec
+   rnorm(1, mean = mumu, sd = sqrt(1/muprec))
+ }
```

```

> lambda.up <- function(mu) {
+   lambdarate <- b + n * (sig2.hat + (mu.hat - mu)^2)/2
+   rgamma(1, shape = a + n/2, rate = lambdarate)
+ }

```

Now we are ready to run a simple Gibbs sampler. We start at the MLE.

```

> niter <- 1000

> mu <- mu.hat
> lambda <- 1/sig2.hat
> mupath <- double(niter)
> lambdapath <- double(niter)
> for (i in 1:niter) {
+   mu <- mu.up(lambda)
+   lambda <- lambda.up(mu)
+   mupath[i] <- mu
+   lambdapath[i] <- lambda
+ }

```

Figure 1 is produced by the following code

```

> acf(mupath)

```

and appears on p. 4.

Figure 2 is produced by the following code

```

> acf(lambdapath)

```

and appears on p. 5.

Almost unbelievable, there is no detectable autocorrelation. This MCMC is almost IIDMC. Thus there is no need for batch means in calculating MCSE.

```

> mu.pe <- mean(mupath)
> mu.mcse <- sd(mupath)/sqrt(length(mupath))
> lambda.pe <- mean(lambdapath)
> lambda.mcse <- sd(lambdapath)/sqrt(length(lambdapath))
> sigma.pe <- 1/sqrt(lambda.pe)
> sigma.mcse <- (1/2) * lambda.pe^(-3/2) * lambda.mcse

> mu.pe

```

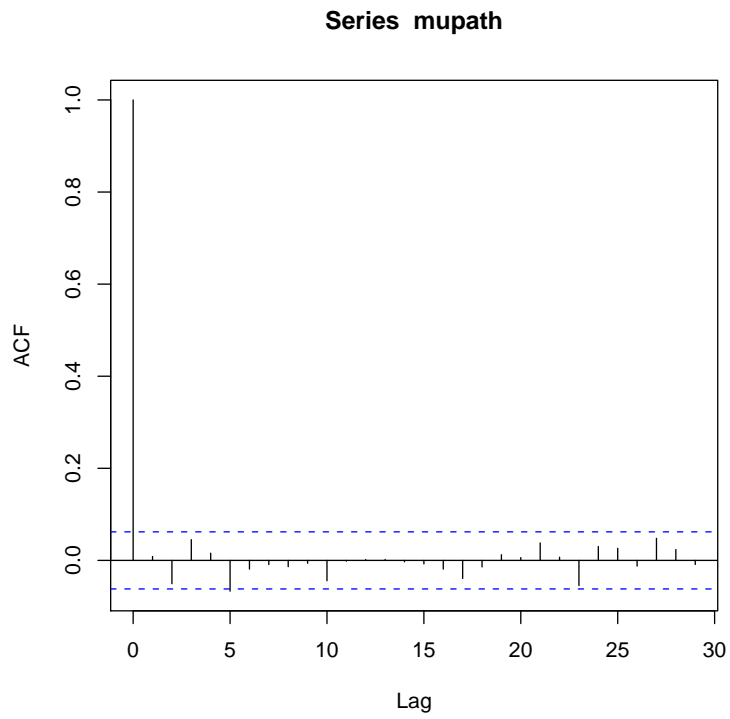


Figure 1: Autocorrelation Plot for  $\mu$  (Monte Carlo sample size 1000).

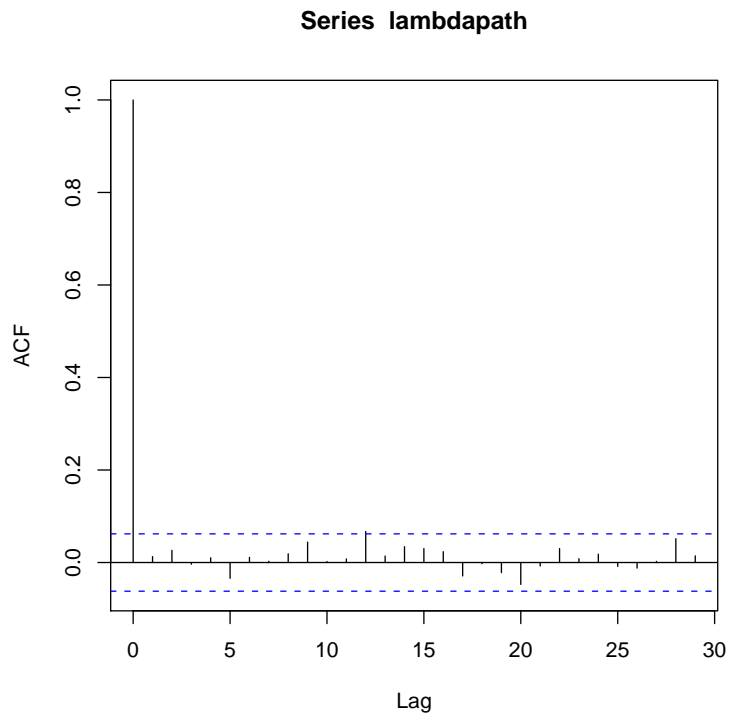


Figure 2: Autocorrelation Plot for  $\lambda$  (Monte Carlo sample size 1000).

```

[1] 14.47044

> mu.mcse

[1] 0.01434573

> sigma.pe

[1] 1.998533

> sigma.mcse

[1] 0.008994898

```

Looks like to satisfy the required accuracy for  $\mu$  we need at least  $14^2 = 196$  times the Monte Carlo sample size. So we redo.

```

> niter <- 200 * niter
> mu <- mu.hat
> lambda <- 1/sig2.hat
> mupath <- double(niter)
> lambdapath <- double(niter)
> for (i in 1:niter) {
+   mu <- mu.up(lambda)
+   lambda <- lambda.up(mu)
+   mupath[i] <- mu
+   lambdapath[i] <- lambda
+ }
> mu.pe <- mean(mupath)
> mu.mcse <- sd(mupath)/sqrt(length(mupath))
> lambda.pe <- mean(lambdapath)
> lambda.mcse <- sd(lambdapath)/sqrt(length(lambdapath))
> sigma.pe <- 1/sqrt(lambda.pe)
> sigma.mcse <- (1/2) * lambda.pe^(-3/2) * lambda.mcse

> mu.pe

[1] 14.49209

> mu.mcse

[1] 0.001026836

```

```
> sigma.pe
```

```
[1] 1.991998
```

```
> sigma.mcse
```

```
[1] 0.0006365476
```

Time for a sanity check. The posterior mean for  $\mu$ , which is 14.4921, is between the prior mean 10 and the observed data mean 14.731. How about  $\lambda$  and  $\sigma$ ? The posterior mean for  $\lambda$ , which is 0.252, is between the prior mean 3 and the observed data point estimate (not a mean)  $1/\hat{\sigma}_n^2 = 0.2077$ .

The prior mean for  $\sigma$  is an exercise in master's level theory.

$$\begin{aligned} E(\sigma) &= E(\lambda^{-1/2}) \\ &= \int_0^\infty \lambda^{-1/2} f(\lambda) d\lambda \\ &= \int_0^\infty \lambda^{-1/2} \cdot \frac{b^a}{\Gamma(a)} \lambda^{a-1} e^{-b\lambda} d\lambda \\ &= \frac{b^a}{\Gamma(a)} \cdot \frac{\Gamma(a - 1/2)}{b^{a-1/2}} \end{aligned}$$

```
> sigma.prior.mean <- sqrt(b) * gamma(a - 1/2)/gamma(a)
```

```
> sigma.prior.mean
```

```
[1] 0.6646702
```

So now we see that the posterior mean for  $\sigma$ , which is 1.992, is between the prior mean 0.6647 and the observed data point estimate (not a mean)  $\hat{\sigma}_n = 2.1941$ .

In all cases the posterior mean is a lot closer to the data point estimate than to the prior mean, as is to be expected with the fairly diffuse priors we used.

On a different issue, note that the lack of autocorrelation in this Gibbs sampler does not arise from lack of correlation between  $\mu$  and  $\lambda$

```
> cor(mupath, lambdapath)
```

```
[1] 0.1434924
```

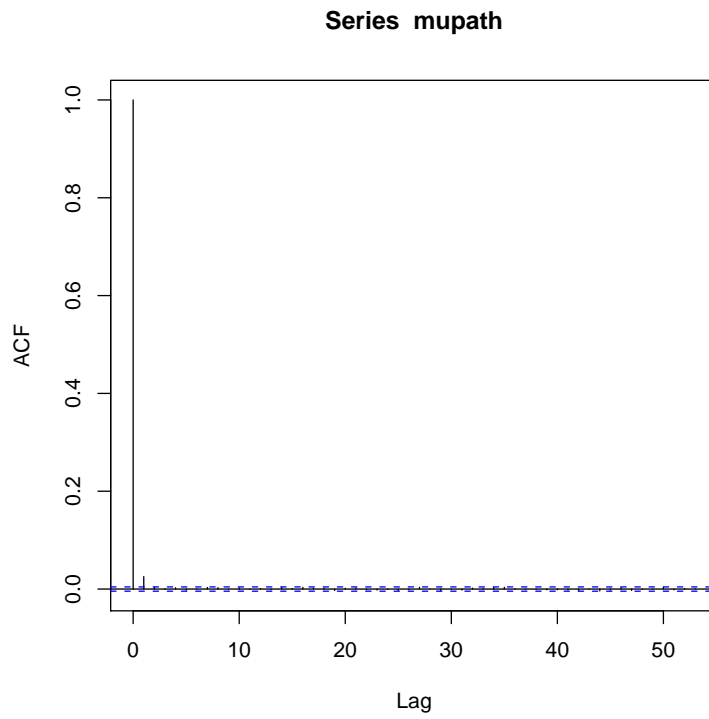


Figure 3: Autocorrelation Plot for  $\mu$  (Monte Carlo sample size  $2e+05$ ).



Apparently, the correlation just isn't enough to give rise to much autocorrelation in the Gibbs sampler.

Figure 3 is produced by the same code as Figure 1 but with our new larger sample as the "path" and appears on p. 8. In it we now see statistically significant lag-one autocorrelation, though it is of no practical significance. And this means we should have used the method of batch means with batch length at least two, and that our MCSE are a little bit low. We know from the formula for asymptotic variance being a sum of autocovariance terms that something like  $1+2\rho_1 = 1.0509$  (where  $\rho_1$  is the lag-one autocovariance) is the correct variance inflation factor, hence the square root of this, which is 1.0251, is the correct inflation factor for MCSE. But we won't worry about that 3% difference.