MAST30027: Modern Applied Statistics

Week 11 Lab Sheet

Suppose that $\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, with $\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$ and $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}$.

1. Show that the conditional distribution of $X_1|X_2 = x_2$ is normal with mean $\mu_1 + (x_2 - \mu_2)\sigma_{12}/\sigma_2^2$ and variance $\sigma_1^2 - \sigma_{12}/\sigma_2^2$.

   Solution: Let $\boldsymbol{\Sigma}^{-1} = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$, then the condional marginal distribution of X_1 given $X_2 = x_2$ is

$$\begin{aligned}
\frac{f(x_1, x_2)}{f(x_2)} &\propto f(x_1, x_2) \\
&\propto \exp\{-\tfrac{1}{2}[(x_1 - \mu_1)^2 a + 2(x_1 - \mu_1)(x_2 - \mu_2)b + (x_2 - \mu_2)^2 c]\} \\
&\propto \exp\{-\tfrac{1}{2}[x_1^2 a - 2x_1(\mu_1 a - (x_2 - \mu_2)b)]\} \\
&\propto \exp\{-\tfrac{1}{2}[x_1 - (\mu_1 - (x_2 - \mu_2)b/a)]^2 a\}
\end{aligned}$$

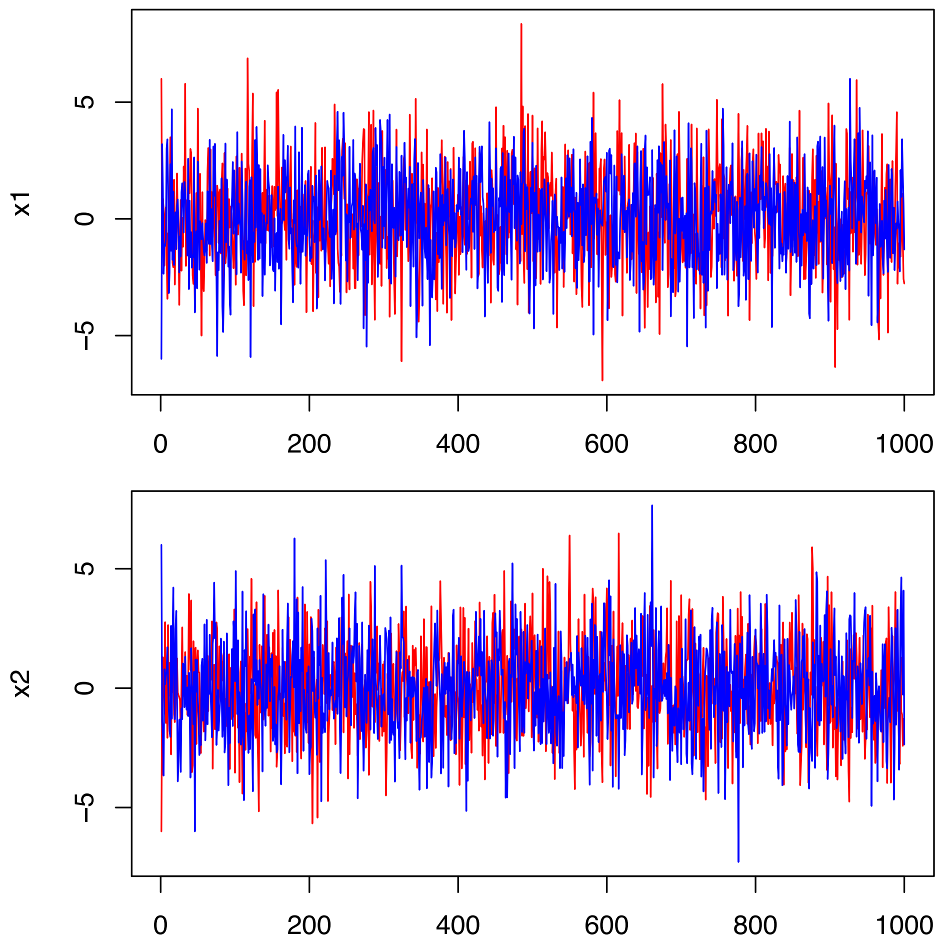
   Thus $X_1|X_2 = x_2 \sim N(\mu_1 - (x_2 - \mu_2)b/a, 1/a)$, where $a = \sigma_2^2/(\sigma_1^2\sigma_2^2 - \sigma_{12}^2)$ and $b = -\sigma_{12}/(\sigma_1^2\sigma_2^2 - \sigma_{12}^2)$, and thus $b/a = -\sigma_{12}/\sigma_2^2$ and $1/a = \sigma_1^2 - \sigma_{12}^2/\sigma_2^2$.

2. Write an R function that uses the Gibbs sampler to generate a sample of size $n = 1000$ from the $N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 4 & 1 \\ 1 & 4 \end{pmatrix}\right)$ distribution. Run at least two Gibbs sampling chains with different initial values. Make trace plots for X_1 and X_2 and see if samples from different chains are mixed well and behave similarly.

   Solution:

```
> set.seed(200)
> # params
> mu1 <- 0
> mu2 <- 0
> s11 <- 4
> s12 <- 1
> s22 <- 4
> # sample size
> nreps <- 1000
> # 1st initial values
> x1 <- 6
> x2 <- -6
> Gsamples <- matrix(nrow=nreps, ncol=2)
> Gsamples[1,] <- c(x1, x2)
> # main loop
> for (i in 2:nreps) {
+   x1 <- rnorm(1, mu1 + (x2 - mu2)*s12/s22, sqrt(s11 - s12/s22))
+   x2 <- rnorm(1, mu2 + (x1 - mu1)*s12/s11, sqrt(s22 - s12/s11))
+   Gsamples[i,] <- c(x1, x2)
+ }
> Gsamples1 = Gsamples
> # 2nd initial values
> x1 <- -6
> x2 <- 6
> Gsamples <- matrix(nrow=nreps, ncol=2)
> Gsamples[1,] <- c(x1, x2)
> # main loop
> for (i in 2:nreps) {
```

```
+   x1 <- rnorm(1, mu1 + (x2 - mu2)*s12/s22, sqrt(s11 - s12/s22))
+   x2 <- rnorm(1, mu2 + (x1 - mu1)*s12/s11, sqrt(s22 - s12/s11))
+   Gsamples[i,] <- c(x1, x2)
+ }
> Gsamples2 = Gsamples
> # trace plot
> par(mfrow=c(2,1), mar=c(2,4,1,1))
> plot(Gsamples1[,1], type="l", xlab="iteration", ylab="x1", col="red", ylim =
+        c(min(Gsamples1[,1], Gsamples2[,1]), max(Gsamples1[,1], Gsamples2[,1])))
> points(1:nreps, Gsamples2[,1], type="l", col="blue")
> plot(Gsamples1[,2], type="l", xlab="iteration", ylab="x2", col="red", ylim =
+        c(min(Gsamples1[,2], Gsamples2[,2]), max(Gsamples1[,2], Gsamples2[,2])))
> points(1:nreps, Gsamples2[,2], type="l", col="blue")
```



The trace plots show that samples from different chains are mixed well and behave similarly.

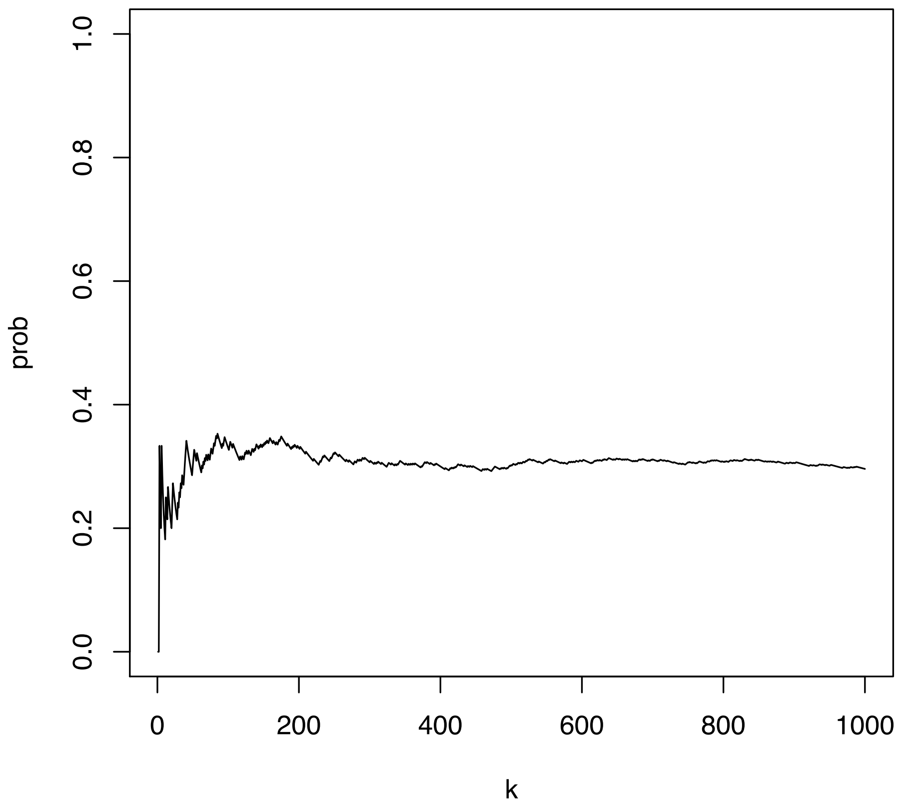
3. Use your simulator to estimate $\mathbb{P}(X_1 \geq 0, X_2 \geq 0)$. To get a feel for the convergence rate, calculate the estimate using samples $\{1, \ldots, k\}$, for $k = 1, \ldots, n$, and then plot the estimates against k.

   Solution: Here, I will use the samples from the first Gibbs sampling chain, but we can either use samples from the second chain or combine samples from two chains.

```
> par(mfrow=c(1,1))
> success <- apply(Gsamples1, 1, function(x) (x[1] > 0)&(x[2] > 0))
> mean(success)

[1] 0.296

> plot(1:nreps, cumsum(success)/(1:nreps), type="l", xlab="k", ylab="prob", ylim=c(0,1))
```



4. Now change $\mathbf{\Sigma}$ to $\begin{pmatrix} 4 & 2.8 \\ 2.8 & 4 \end{pmatrix}$ and generate another sample of size 1000. What do the traces/estimates look like now?

   Solution: We put `s12 <- 2.8` then re-run the code above, getting a different `Gsamples`. We plot the cumlative estimates on top of the previous graph using `lines`. The cumulative estimates are more volatile in the second case, reflecting the stronger autocorrelation in the Markov chain, caused by the stronger correlation between X_1 and X_2.

```
> success2 <- success
> mu1 <- 0
> mu2 <- 0
> s11 <- 4
> s12 <- 2.8
> s22 <- 4
> x1 <- 6
> x2 <- -6
> nreps <- 1000
> Gsamples <- matrix(nrow=nreps, ncol=2)
> Gsamples[1,] <- c(x1, x2)
> for (i in 2:nreps) {
+   x1 <- rnorm(1, mu1 + (x2 - mu2)*s12/s22, sqrt(s11 - s12/s22))
+   x2 <- rnorm(1, mu2 + (x1 - mu1)*s12/s11, sqrt(s22 - s12/s11))
+   Gsamples[i,] <- c(x1, x2)
+ }
> success <- apply(Gsamples, 1, function(x) (x[1] > 0)&(x[2] > 0))
> mean(success)
```

```
[1] 0.37

> plot(1:nreps, cumsum(success2)/(1:nreps), type="l", xlab="k", ylab="prob", ylim=c(0,1))
> lines(1:nreps, cumsum(success)/(1:nreps), col="red")
```