Gibbs Samplers

Jonathan Navarrete

July 2, 2017

Introduction

Now that we've familiarized ourselves with MCMC and the Metropolis-Hastings algorithms, we begin to analyze a now-common MCMC algorithm called Gibbs sampler. The Gibbs sampler is in fact a special case of the Metropolis-Hastings algorithm for high dimensional target distributions.

We introduce the Gibbs sampler with a two-stage example. The **two-state Gibbs sampler algorithm** as described Robert & Casella goes as follows

`Take` $X_0 = x_0$

For $t = 1, 2, ...,$ `generate`

1. $Y_t \sim f_{Y|X}(\cdot|x_{t-1})$
2. $X_t \sim f_{X|Y}(\cdot|y_t)$

The *two-stage* Gibbs sampler creates a Markov chain from a joint distribution in the following way. If two random variables X and Y have joint density $f(x, y)$, with corresponding conditional densities $f_{Y|X}$ and $f_{X|Y}$, the two stage Gibbs sampler generates a Markov chain (X_t, Y_t) by generating Y_t from conditional density $f_{Y|X}$ and then generating X_t from conditional density $f_{X|Y}$.

We illustrate the implementation of the Gibbs sampler with a simple example. Consider a bivariate Normal distribution where

$$X, Y \sim N_2(\begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \begin{pmatrix} \sigma_X^2 & \sigma_{XY}^2 \\ \sigma_{YX}^2 & \sigma_Y^2 \end{pmatrix})$$

The marginal distributions of X and Y are $N(\mu_X, \sigma_X)$ and $N(\mu_X, \sigma_X)$. The conditional distributions of Y and X are

$$Y|X = x \sim N(\mu_Y + \frac{\rho\sigma_Y}{\sigma_X}(x - \mu_X), (1 - \rho^2)\sigma_Y^2)$$

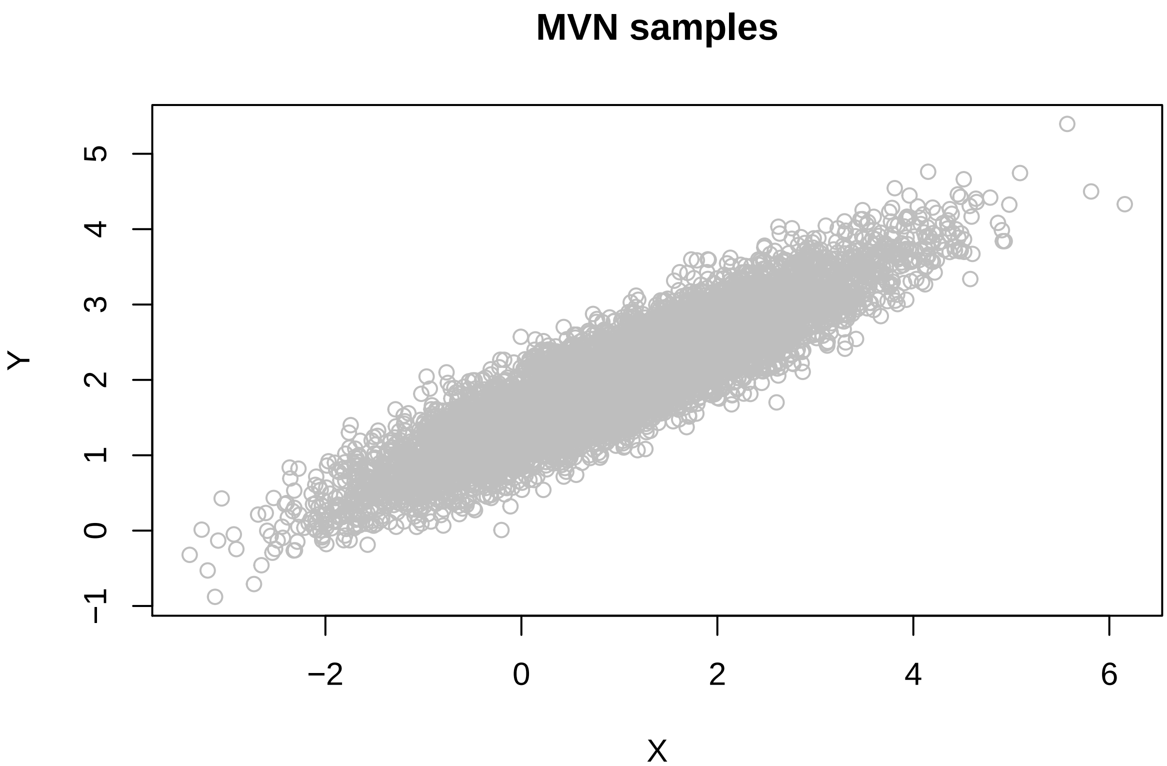
and

$$X|Y = y \sim N(\mu_X + \frac{\rho\sigma_X}{\sigma_Y}(y - \mu_Y), (1 - \rho^2)\sigma_X^2)$$

ρ is the correlation between X and Y, and $(1 - \rho^2)\sigma_X^2$ is the variance.

```r
N = 10000

rho = 0.9
mu_x = 1
mu_y = 2
sd_x = 1.2
sd_y = 0.75

s1 = sqrt(1 - rho^2) * sd_x
s2 = sqrt(1 - rho^2) * sd_y

MVN = matrix(data = NA, nrow = N, ncol = 2,
           dimnames = list(NULL, c("X", "Y")))

MVN[1, ] = c(mu_x, mu_y)
```

```r
Y = MVN[1, 2] ## get Y vals
for(i in 1:(N-1)){
  mx = mu_x + rho * (Y - mu_y) * sd_x/sd_y
  X = rnorm(n = 1, mx, s1)
  MVN[i+1, 1] = X
  my = mu_y + rho * (X - mu_x) * sd_y/sd_x
  Y = rnorm(n = 1, mean = my, sd = s2)
  MVN[i+1, 2] = Y
}


plot(MVN, type = "p", col = 8,
     main = "MVN samples")
```




```r
## means
colMeans(MVN)
```

```
##        X        Y
## 1.005803 2.003126
```

```r
## correlation
cor(MVN)
```

```
##           X         Y
## X 1.0000000 0.9002251
## Y 0.9002251 1.0000000
```

Beta-Binomial revisited

In the introduction to these notes, we saw a Bayesian example of the Beta-Binomial distribution. From Casella's paper *Explaining the Gibbs Sampler*, we revisit this example.

$$X|\theta \sim Bin(n,\theta), \text{ and } \theta \sim Beta(a,b)$$

have joint density

$$f(x,\theta) = \binom{n}{x}\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{x+a-1}(1-\theta)^{n-x+b-1}$$

is the a $Beta(x+a, n-x+b)$ distribution.

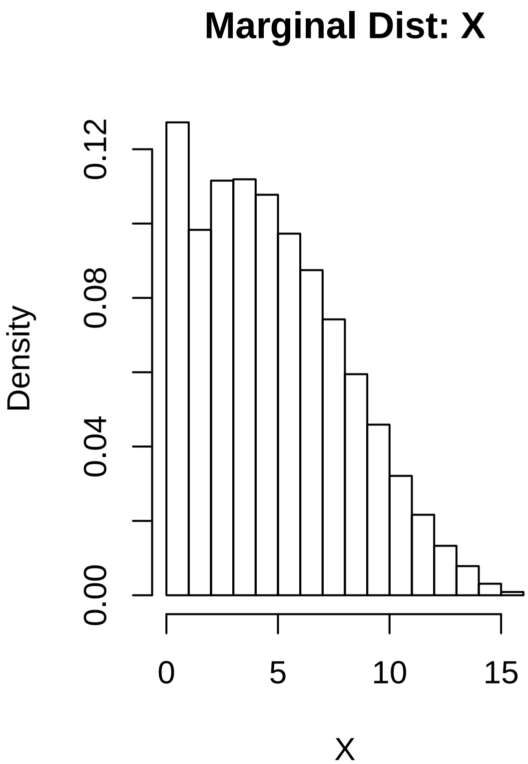
Suppose we are interested in calculating some characteristics of the marginal distributions of $X|\theta$ and $\theta|a,b,x,n$.

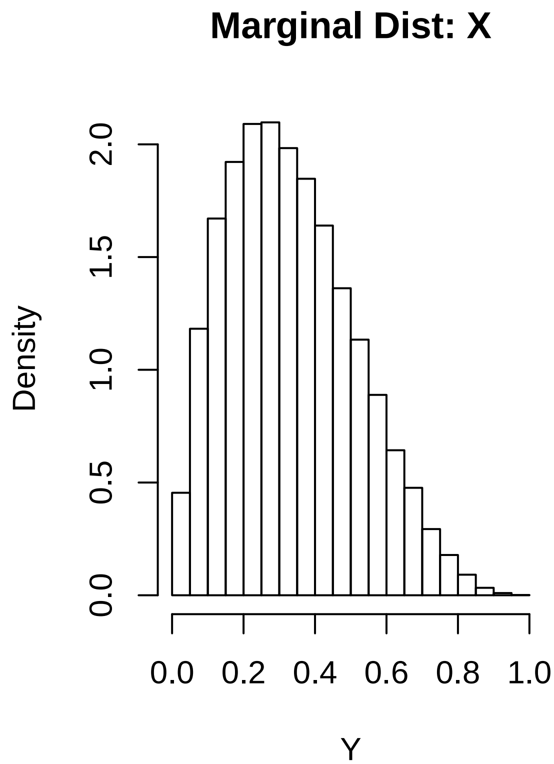
$$f(x|\theta) \text{ is } Bin(n,\theta) f(\theta|x) \text{ is } Beta(x+a, n-x+b)$$

Therefore, we follow an iterative algorithm of

$$X_i \sim f(x|\theta) Y_{i+1} \sim f(y|X_i = x_i)$$

```
N = 10^5
n = 16
a = 2
b = 4
X = numeric(N)
Y = numeric(N)

## initial values
X[1] = 0.2
Y[1] = 0.34 ## theta values

for(i in 1:N){
  X[i+1] = rbinom(n = 1, size = n, prob = Y[i])
  Y[i+1] = rbeta(1, a + X[i+1], n - X[i+1]+b)
}

par(mfrow = c(1,2))
hist(X, main = "Marginal Dist: X", probability = TRUE)
hist(Y, main = "Marginal Dist: X", probability = TRUE)
```


Marginal Dist: X
Density
0.00
0.04
0.08
0.12
0
5
10
15
X


Marginal Dist: X
Density
0.0
0.5
1.0
1.5
2.0
0.0
0.2
0.4
0.6
0.8
1.0
Y