# 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

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

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

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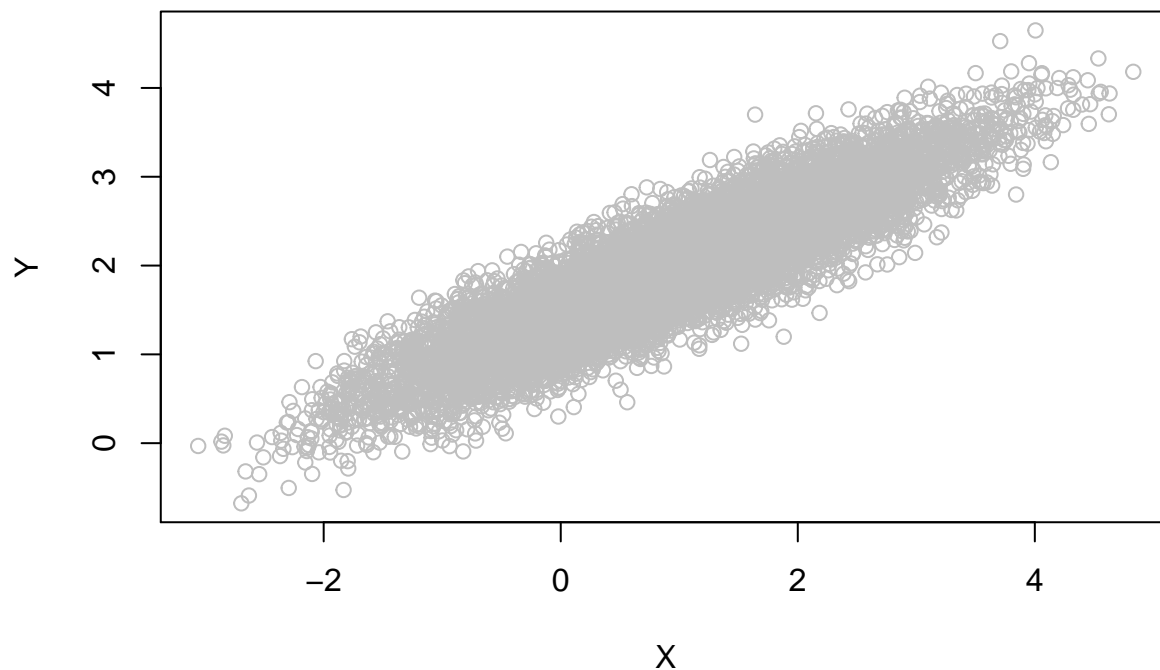
```
Y = MVN[1, 2] ## get Y vals
for(i in 1:(N-1)){
  mx = mu_x + rho * (Y - mu_y) * sd_x/sd_y
  X = rnorm(n = 1, mx, s1)
  MVN[i+1, 1] = X
  my = mu_y + rho * (X - mu_x) * sd_y/sd_x
  Y = rnorm(n = 1, mean = my, sd = s2)
  MVN[i+1, 2] = Y
}


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

**MVN samples**



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

```
##         X         Y
## 0.9236048 1.9486663
```

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

```
##           X         Y
## X 1.0000000 0.8969959
## Y 0.8969959 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

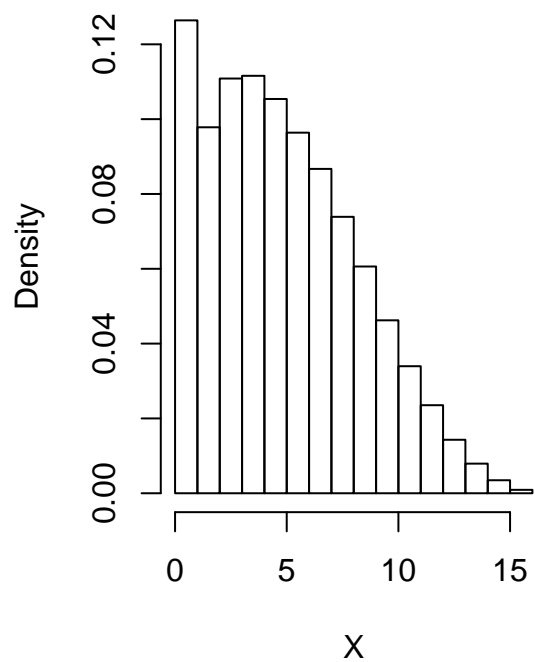$$X_i \sim f(x|\theta) Y_{i+1} \sim f(y|X_i = x_i)$$

```
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)
```
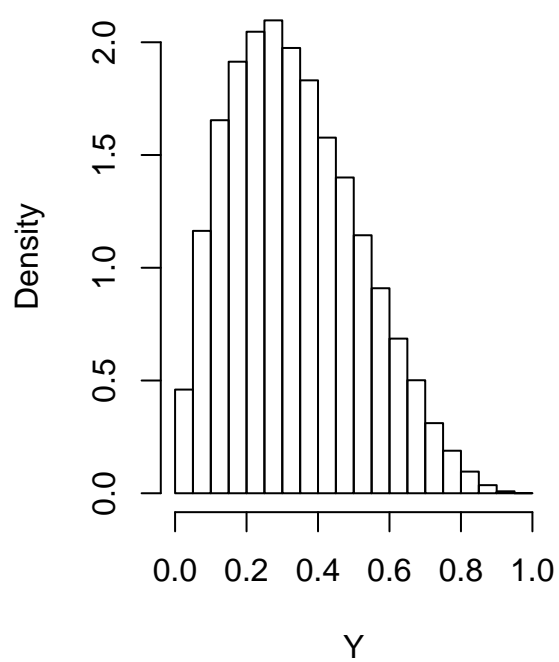
**Marginal Dist: X**



**Marginal Dist: X**

## Bayesian Analysis with Normal Data

Suppose we have a random sample $y_1, ..., y_n | \mu, \tau \sim N(\mu, 1/\tau)$. Here we use a precision parameter $\tau$ instead of $\sigma$ for standard devation (or $\sigma^2$ for variance) . Place two independent priors on the unkown parameters $\mu$ and $\tau$.

$$\mu \sim N(a, 1/b) \perp \tau \sim Gamma(c, d)$$

The posterior is given by

$$p(\mu, \tau | \mathbf{y}) = \frac{L(\mu, \tau | \mathbf{y}) \times p(\mu, \tau)}{m(\mathbf{y})}$$

where $p(\mu, \tau) = p(\mu) \times p(\tau)$ and $m(\mathbf{y}) = \int L(\mu, \tau | \mathbf{y}) \times p(\mu, \tau)$

The likelihood is

$$
\begin{aligned}
L(\mu, \tau | \mathbf{y}) &= \prod_{i=1}^{n} f(y_i | \mu, \tau) \\
&= \frac{\tau}{2\pi}^{n/2} exp(\frac{\tau}{2} \sum (y_i - \mu)^2) \\
&= \frac{\tau}{2\pi}^{n/2} exp \frac{\tau}{2}(n-1)s^2 - \frac{\tau}{2}(\bar{y} - \mu)^2 \text{ ( obtained after some algebra )} \\
&\propto \tau^{n/2} exp \frac{\tau}{2}(n-1)s^2 - \frac{\tau}{2}(\bar{y} - \mu)^2 \text{ ( drop unimportant constants )}
\end{aligned}
$$

The posterior is then given by

$$
\begin{aligned}
p(\mu, \tau) &\propto L(\mu, \tau | \mathbf{y}) \times p(\mu, \tau) \\
&= \tau^{n/2} exp(\frac{\tau}{2}(n-1)s^2 - \frac{\tau}{2}(\bar{y} - \mu)^2) \times \frac{d^c}{\Gamma(c)} \tau^{c-1} exp(-b\tau) \times \sqrt{\frac{b}{2\pi}} exp(-\frac{b}{2}(\mu - a)^2) \\
&\text{( next, drop unimportant constants )} \\
&\propto \tau^{n/2} exp(\frac{\tau}{2}(n-1)s^2 - \frac{\tau}{2}(\bar{y} - \mu)^2) \times \tau^{c-1} exp(-b\tau) \times exp(-\frac{b}{2}(\mu - a)^2) \\
&= \tau^{n/2+c-1} exp(\frac{\tau}{2}(n-1)s^2 - \frac{\tau}{2}(\bar{y} - \mu)^2) \times exp(-b\tau) \times exp(-\frac{b}{2}(\mu - a)^2)
\end{aligned}
$$

And the final result is some function that is not recognizable as any commonly known distribution. Hence, we're stuck if we are to try to sample from this distribution directly!

However, if we are too look at the conditional posteriors, $\mu | \tau, \mathbf{y}$

$$p(\mu | \tau, \mathbf{y}) \propto exp(1\frac{1}{2}(n\tau + b)(\mu - \hat{\mu}_\tau)^2)$$

and $\tau | \mu, \mathbf{y}$

$$p(\tau | \mu, \mathbf{y}) \propto \tau^{(n/2+c)-1} exp(-\tau(d + \frac{n(\bar{y} - \mu)^2 + (n-1)s^2}{2}))$$

we find that there are recognizable distributions. They are $N(\hat{\mu}, \frac{1}{n\tau+b})$ and $Gamma(\frac{n}{2} + c, d + \frac{n(\bar{y}-\mu)^2 + (n-1)s^2}{2})$

Given that we can sample from the conditional distributions, we can utilize a Gibbs sampler to sample from the posterior distribution $p(\mu, \tau | \mathbf{y})$.