The Gibbs Sampler
(Chapter 12)

Michael Tsiang

Stats 102C: Introduction to Monte Carlo Methods

UCLA

Acknowledgements: Qing Zhou

Outline

The Gibbs Sampler

- Another common Markov Chain Monte Carlo method is the **Gibbs sampler**, first proposed by Geman and Geman 1984[1] in an application to Gibbs distributions (from statistical physics).

- The Gibbs sampler (or Gibbs sampling) is useful when the target distribution is a multivariate distribution:

$$\pi(\boldsymbol{x}) = \pi(x_1, x_2, \ldots, x_d),$$

  for some $d > 1$.

- **Main Idea**: Break up the problem of sampling from a complicated high dimensional distribution into a sequence of easier problems by sampling from its univariate conditional distributions.

[1] https://doi.org/10.1109%2FTPAMI.1984.4767596

The Gibbs Sampler

In the Gibbs sampling scenario:

- We are unable to sample from the target distribution $\pi(\boldsymbol{x})$.
- All of the univariate conditional distributions

$$\pi(x_i|\boldsymbol{x}_{-i}) := \pi(x_i|x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_d),$$

  for $i = 1, 2, \ldots, d$, are known and we can sample from them.

The Gibbs sampler is a special case of the Metropolis-Hastings algorithm, where the proposal distributions are the conditional distributions of $\pi(\boldsymbol{x})$, and the proposals are always accepted.

We will describe two types of Gibbs samplers.

The Gibbs Sampler

Random-Scan Gibbs Sampler

Goal: Generate $\boldsymbol{X} \sim \pi(\boldsymbol{x}) = \pi(x_1, x_2, \ldots, x_d)$.

Let $\boldsymbol{x}^{(t)} = (x_1^{(t)}, x_2^{(t)}, \ldots, x_d^{(t)})$, and

$$\boldsymbol{x}_{-i}^{(t)} = (x_1^{(t)}, \ldots, x_{i-1}^{(t)}, x_{i+1}^{(t)}, \ldots, x_d^{(t)}).$$

At the $t+1$ iteration:

1. Randomly select a coordinate i from $\{1, 2, \ldots, d\}$.
2. Generate $x_i^{(t+1)} \sim \pi(x_i | \boldsymbol{x}_{-i}^{(t)})$ and leave the remaining components unchanged, i.e., $\boldsymbol{x}_{-i}^{(t+1)} = \boldsymbol{x}_{-i}^{(t)}$.

The random-scan Gibbs sampler updates a single (randomly chosen) coordinate of $\boldsymbol{x}^{(t)}$ by sampling from its conditional distribution, fixing all other coordinates in the vector.

The Gibbs Sampler

Systematic-Scan Gibbs Sampler

Goal: Generate $\boldsymbol{X} \sim \pi(\boldsymbol{x}) = \pi(x_1, x_2, \ldots, x_d)$.

Let $\boldsymbol{x}^{(t)} = (x_1^{(t)}, x_2^{(t)}, \ldots, x_d^{(t)})$. At the $t+1$ iteration:

For $i = 1, 2, \ldots, d$:

- Generate

$$x_i^{(t+1)} \sim \pi(x_i | x_1^{(t+1)}, x_2^{(t+1)}, \ldots, x_{i-1}^{(t+1)}, x_{i+1}^{(t)}, \ldots, x_d^{(t)}).$$

The systematic-scan Gibbs sampler cycles through the coordinates of $\boldsymbol{x}^{(t)}$ in order, updating each coordinate individually by sampling from its conditional distribution, fixing all other coordinates in the vector.

Outline

1. The Gibbs Sampler
2. Examples
   - Example 1: Bivariate Normal
   - Example 2: The Beta-Binomial Model

Example 1: Bivariate Normal

- Let $\boldsymbol{x} = (x_1, x_2)$, and let the target distribution $\pi(\boldsymbol{x})$ be the bivariate normal distribution
$$\mathcal{N}_2\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix}\right),$$
so
$$\pi(x_1, x_2) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{x_1^2 - 2\rho x_1 x_2 + x_2^2}{2(1-\rho^2)}\right].$$
- We want to use Gibbs sampling to sample from $\pi(\boldsymbol{x})$.

Example 1: Bivariate Normal

- To use Gibbs sampling, we need to find the conditional distributions $\pi(x_1|x_2)$ and $\pi(x_2|x_1)$.
- From the definition of conditional probability, notice that

$$\pi(x_1|x_2) \; = \; \frac{\pi(x_1, x_2)}{\pi(x_2)} \; \propto \; \pi(x_1, x_2).$$

- Since $\pi(x_1|x_2)$ is a function of only x_1 (all x_2 terms are considered fixed), this shows that the conditional distribution of $\pi(x_1|x_2)$ is proportional to the joint distribution $\pi(x_1, x_2)$.
- The marginal $\pi(x_2) = \displaystyle\int \pi(x_1, x_2)\, \mathrm{d}x_1$ is the normalizing constant for $\pi(x_1|x_2)$.

Example 1: Bivariate Normal

Using proportionality, we have

$$
\begin{aligned}
\pi(x_1|x_2) &\propto \pi(x_1, x_2) \\
&\propto \exp\left[-\frac{x_1^2 - 2\rho x_1 x_2 + x_2^2}{2(1-\rho^2)}\right] \\
&\propto \exp\left[-\frac{x_1^2 - 2\rho x_1 x_2}{2(1-\rho^2)}\right] \\
&= \exp\left[-\frac{x_1^2 - 2\rho x_1 x_2 + (\rho^2 x_2^2 - \rho^2 x_2^2)}{2(1-\rho^2)}\right] \\
&\propto \exp\left[-\frac{(x_1 - \rho x_2)^2}{2(1-\rho^2)}\right] \\
&= \exp\left[-\frac{1}{2}\left(\frac{x_1 - \rho x_2}{\sqrt{1-\rho^2}}\right)^2\right],
\end{aligned}
$$

which we recognize as a normal distribution with mean ρx_2 and variance $1-\rho^2$. So $\pi(x_1|x_2) \sim \mathcal{N}(\rho x_2, 1-\rho^2)$.

Example 1: Bivariate Normal

The systematic-scan Gibbs sampler for the bivariate normal:

- Let $\boldsymbol{x}^{(t)} = (x_1^{(t)}, x_2^{(t)})$ denote the Markov chain at time t.
- Generate $\boldsymbol{x}^{(t+1)} = (x_1^{(t+1)}, x_2^{(t+1)})$ by:

$$\begin{aligned}
x_1^{(t+1)} \mid x_2^{(t)} &\sim \mathcal{N}(\rho x_2^{(t)}, 1-\rho^2) \\
x_2^{(t+1)} \mid x_1^{(t+1)} &\sim \mathcal{N}(\rho x_1^{(t+1)}, 1-\rho^2).
\end{aligned}$$

Example 1: Bivariate Normal

By induction, one can show that

$$\begin{aligned}
\begin{pmatrix} x_1^{(t)} \\ x_2^{(t)} \end{pmatrix} \quad &\sim \quad \mathcal{N}_2\left(\begin{pmatrix} \rho^{2t-1} x_2^{(0)} \\ \rho^{2t} x_2^{(0)} \end{pmatrix}, \begin{pmatrix} 1-\rho^{4t-2} & \rho-\rho^{4t-1} \\ \rho-\rho^{4t-1} & 1-\rho^{4t} \end{pmatrix} \right) \\
&\xrightarrow{t\to\infty} \mathcal{N}_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right).
\end{aligned}$$

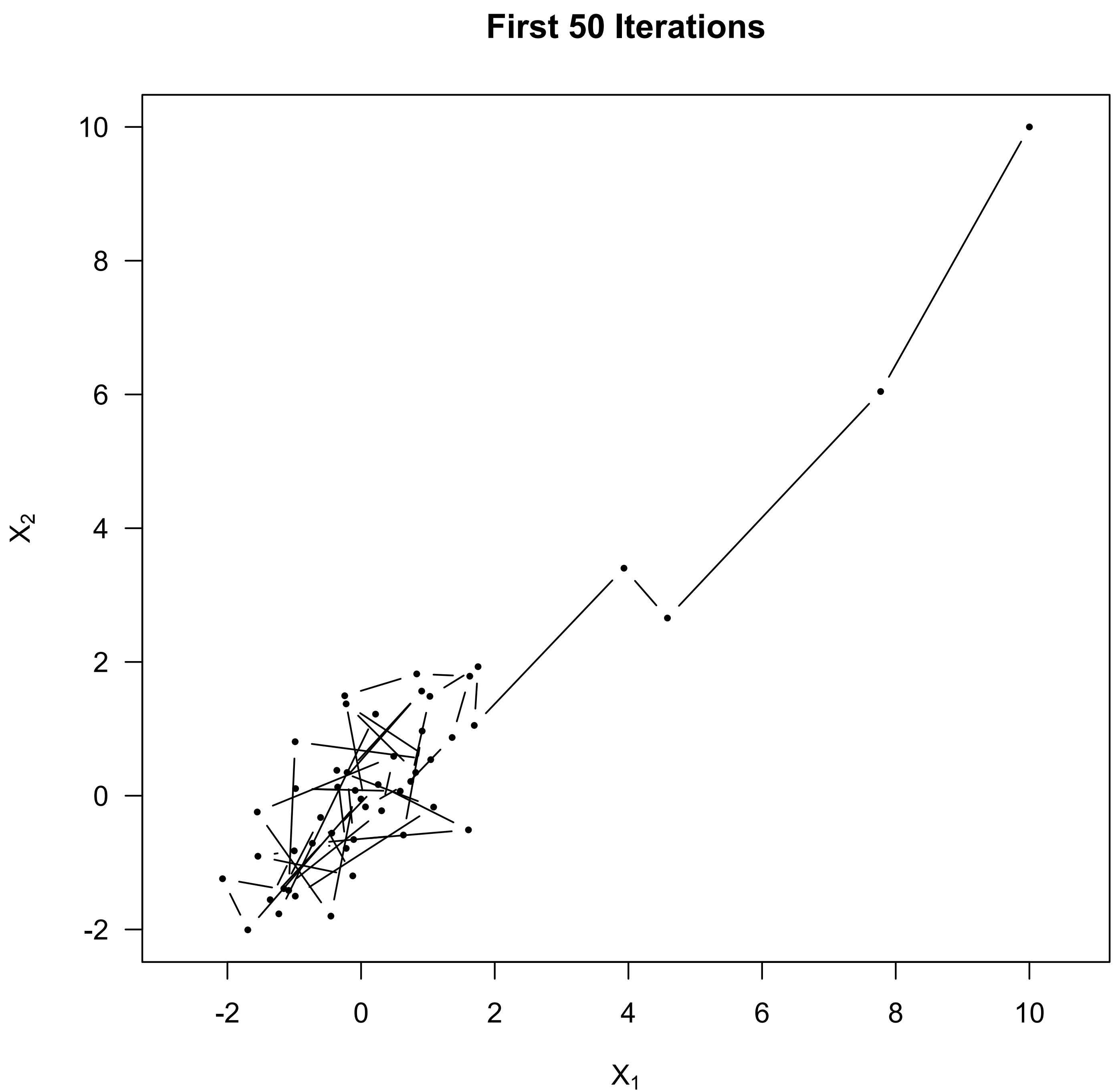
So, as $t \to \infty$, the joint distribution of $\boldsymbol{x}^{(t)}$ converges to $\pi(\boldsymbol{x})$.

Example 1: Bivariate Normal

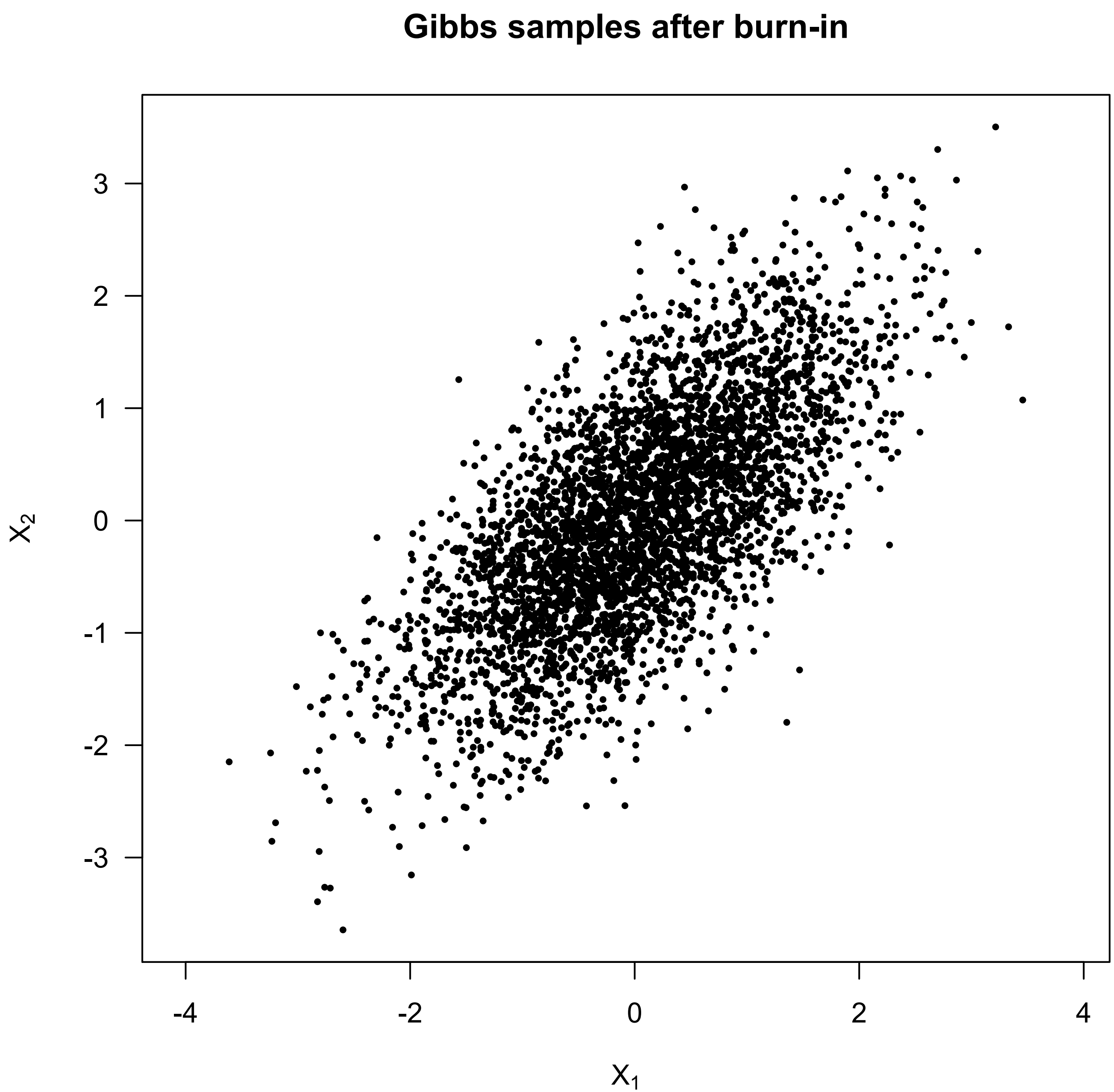
R Code for Gibbs sampler to sample from bivariate normal:

```
> # Set correlation
> rho <- 0.7

> set.seed(9999) # for reproduceability
> n <- 5000 # specify length of chain
> X <- matrix(0, nrow = n, ncol = 2) # create space for chain
> X[1, ] <- c(10, 10) # specify initial state

> # Systematic-scan Gibbs sampler
> for (t in 2:n) {
+     # Generate X1^(t) from X1 | X2^(t-1)
+     X[t, 1] <-
+     rnorm(1, mean = rho * X[t - 1, 2], sd = sqrt(1 - rho^2))
+
+     # Generate X2^(t) from X2 | X1^(t)
+     X[t, 2] <-
+     rnorm(1, mean = rho * X[t, 1], sd = sqrt(1 - rho^2))
+ }
```

Example 1: Bivariate Normal



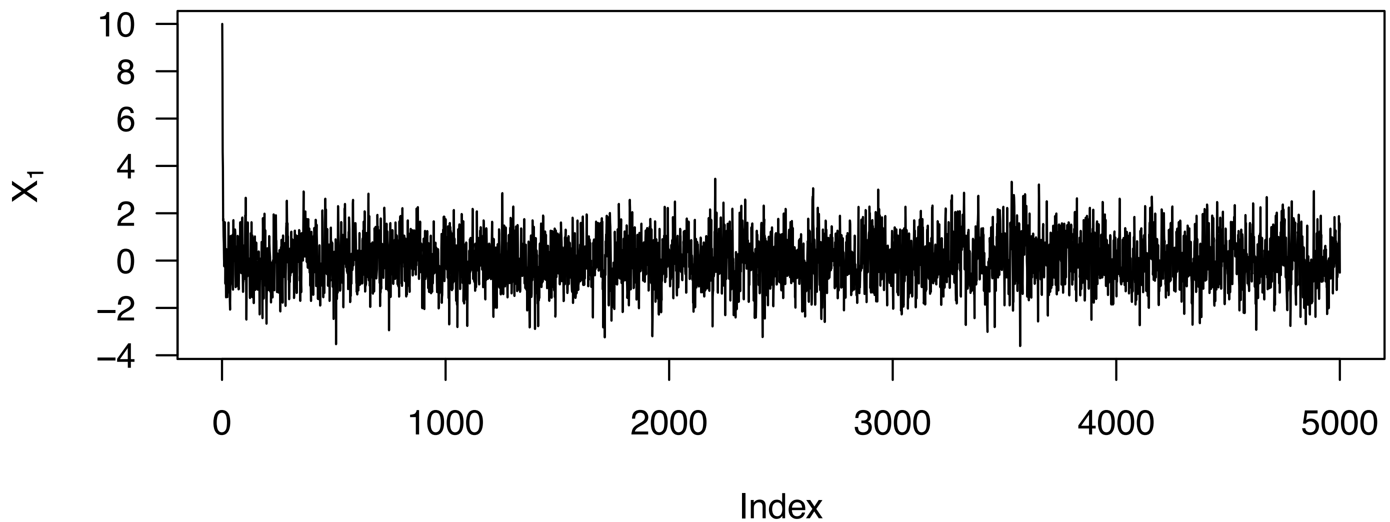
Example 1: Bivariate Normal

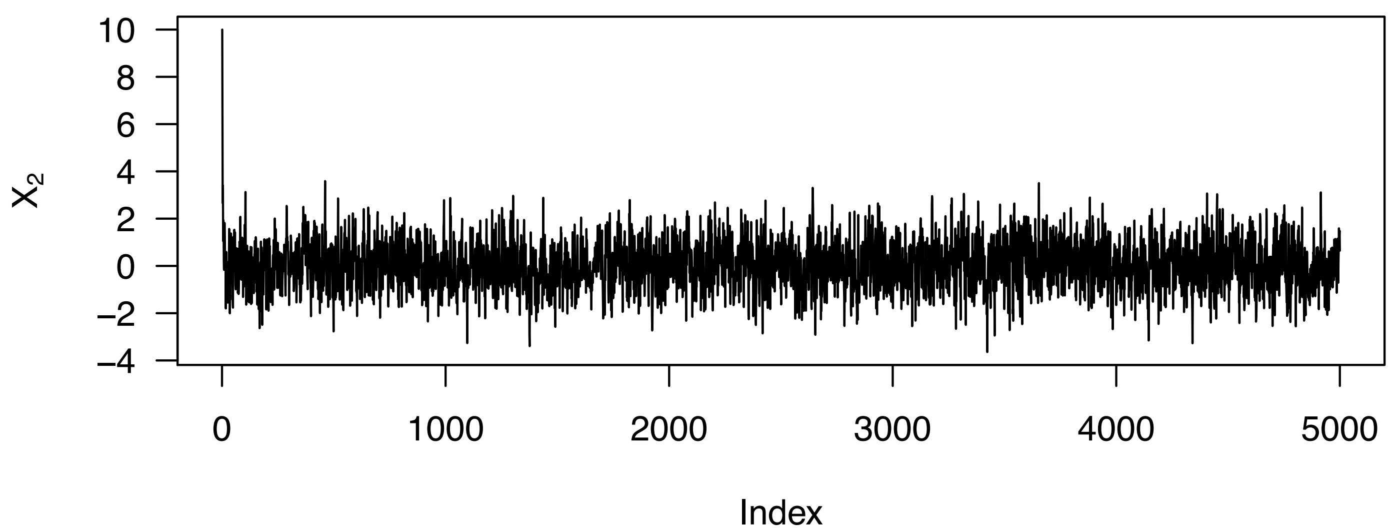


R Code for the plots:

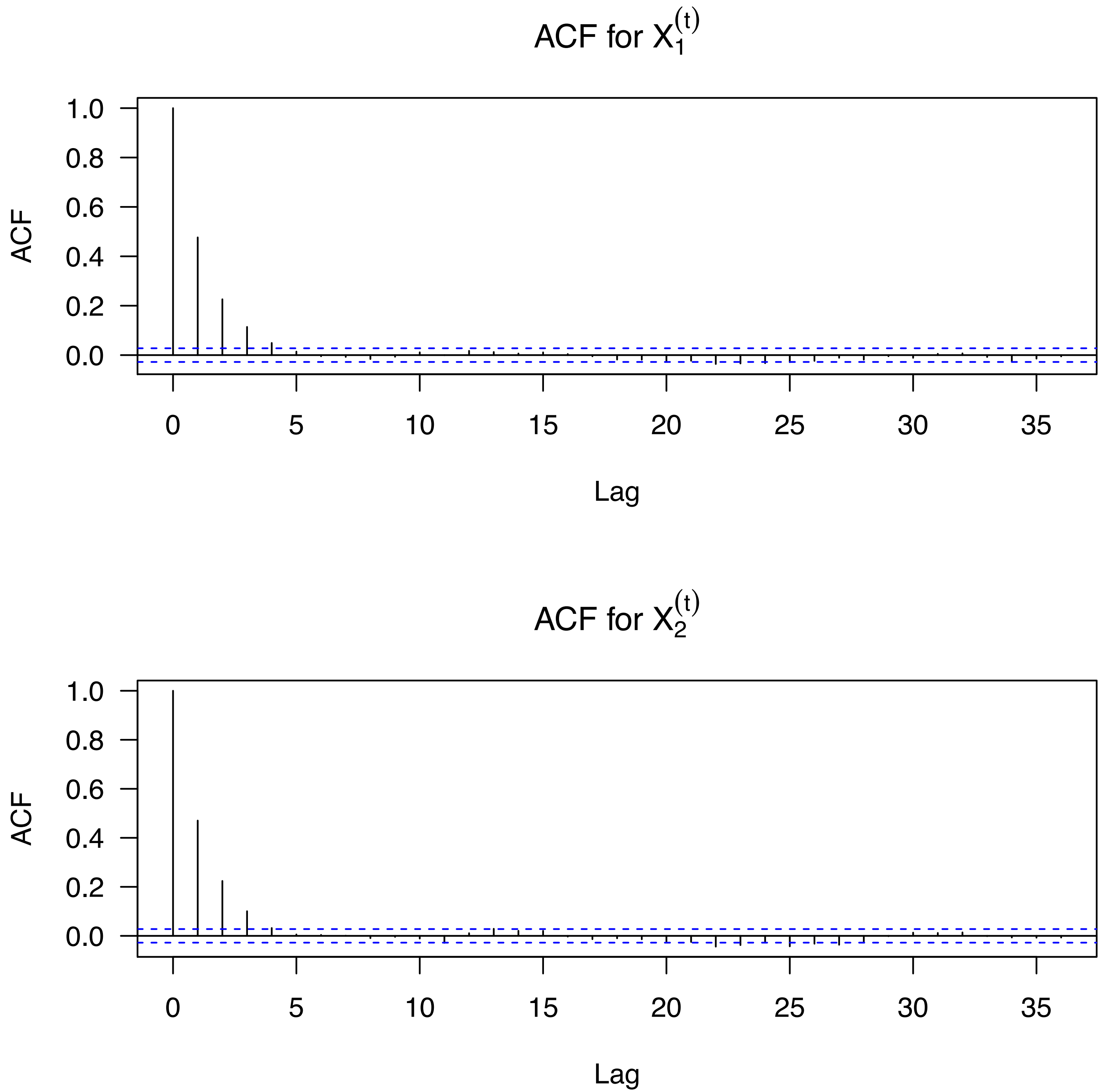
```
> # Plot the sample path for first 50 iterations
> plot(X[1:50, ],
+       type = "b", pch = 19, cex = 0.4, asp = 1, las = 1,
+       xlab = expression(X[1]), ylab = expression(X[2]),
+       main = "First 50 Iterations"
+ )

> # Scatterplot of samples after 1000 burn-in iterations
> plot(X[1001:n, ],
+       pch = 19, cex = 0.4, asp = 1, las = 1,
+       xlab = expression(X[1]), ylab = expression(X[2]),
+       main = "Gibbs samples after burn-in"
+ )
```

Example 1: Bivariate Normal





Example 1: Bivariate Normal



R Code for the plots:

```
> # Plot trace plot for each coordinate
> par(mfrow = c(2, 1))
> plot(X[, 1],type = "l", ylab = expression(X[1]), las = 1)
> plot(X[, 2], type = "l", ylab = expression(X[2]), las = 1)

> # Plot autocorrelation function for each coordinate
> par(mfrow = c(2, 1))
> acf(X[, 1], main = expression(paste("ACF for ", X[1]^(t))),
+     las = 1
+ )
> acf(X[, 2], main = expression(paste("ACF for ", X[2]^(t))),
+     las = 1
+ )
```

- Let the bivariate target distribution $\pi(x, y)$, for fixed n, α, β, be given by

$$\pi(x, y) \propto \binom{n}{x} y^{x+\alpha-1}(1-y)^{n-x+\beta-1},$$

for $x = 0, 1, \ldots, n$ and $y \in [0, 1]$.

- We want to use Gibbs sampling to sample from $\pi(x, y)$.

Example 2: The Beta-Binomial Model

- Using proportionality, we have

$$\begin{aligned}\pi(x|y) &\propto \pi(x,y)\\ &\propto \binom{n}{x} y^{x+\alpha-1}(1-y)^{n-x+\beta-1}\\ &\propto \binom{n}{x} y^{x}(1-y)^{n-x},\end{aligned}$$

which we recognize as a binomial distribution with n trials and success probability y. So $\pi(x|y) \sim \mathrm{Bin}(n,y)$.

- Similarly, since

$$\pi(y|x) \ \propto \ \pi(x,y) \ \propto \ y^{x+\alpha-1}(1-y)^{n-x+\beta-1},$$

then we recognize $\pi(y|x)$ as a beta distribution with parameters $x+\alpha$ and $n-x+\beta$. So

$$\pi(y|x) \sim \mathrm{Beta}(x+\alpha, n-x+\beta).$$

Example 2: The Beta-Binomial Model

Systematic-scan Gibbs sampler for $\pi(x, y)$:

- Let $(x^{(t)}, y^{(t)})$ denote the Markov chain at time t.
- Generate $(x^{(t+1)}, y^{(t+1)})$ by:

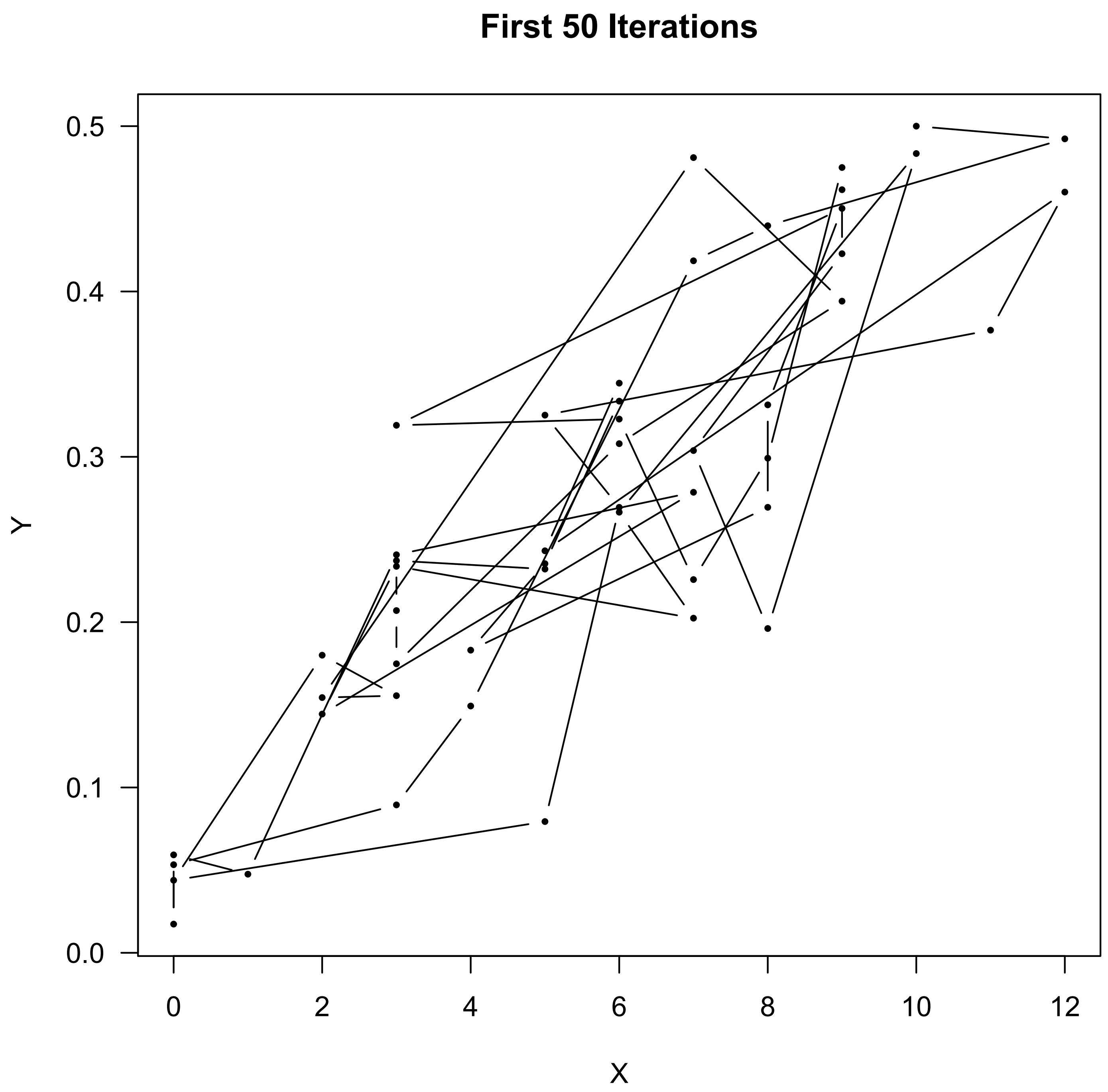
$$\begin{aligned} x^{(t+1)} \mid y^{(t)} &\sim \mathrm{Bin}(n, y^{(t)}) \\ y^{(t+1)} \mid x^{(t+1)} &\sim \mathrm{Beta}(x^{(t+1)} + \alpha, n - x^{(t+1)} + \beta). \end{aligned}$$

Example 2: The Beta-Binomial Model

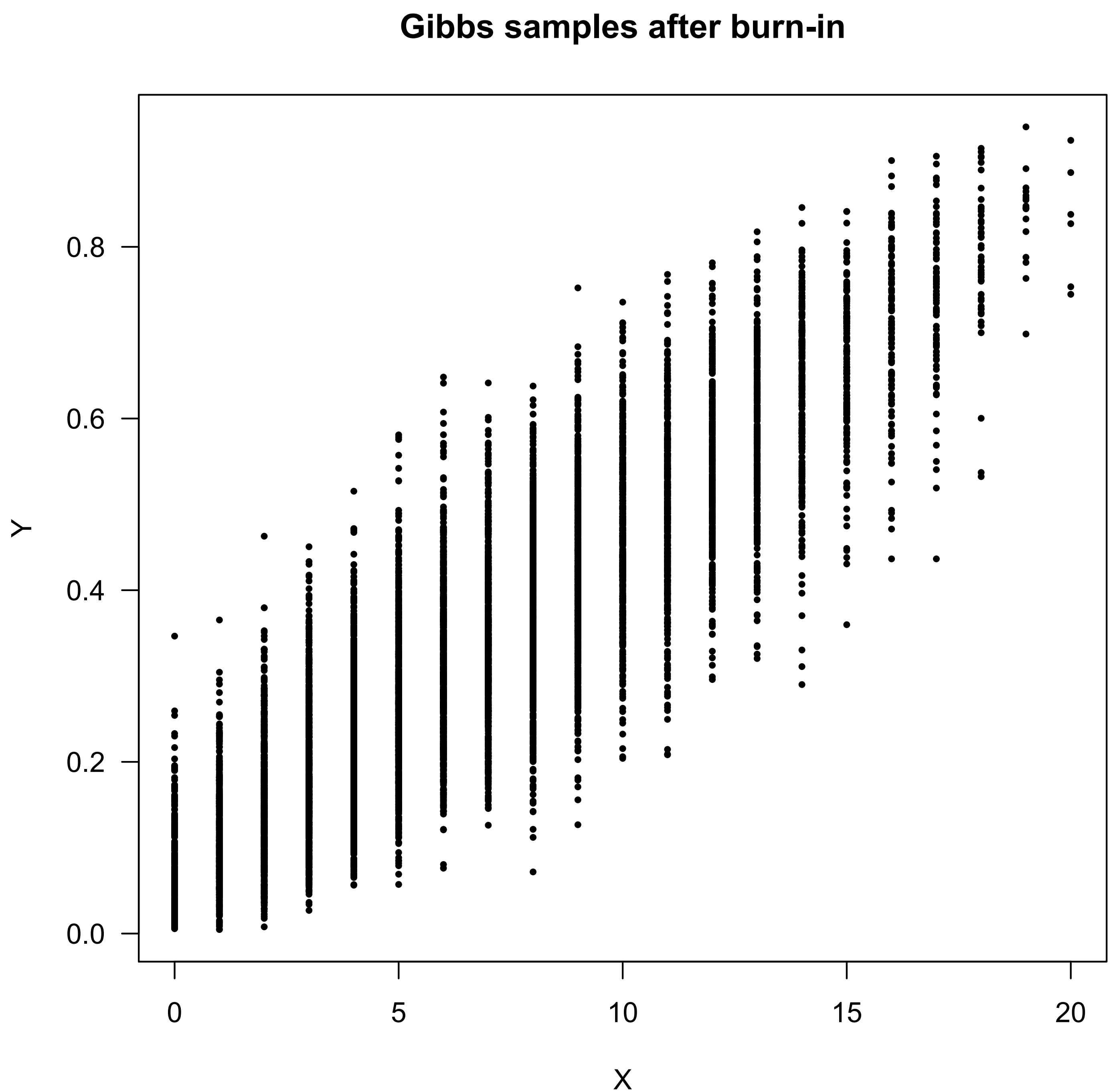
R Code for Gibbs sampler to sample from $\pi(x, y)$:

```
> # Set parameters
> n <- 20
> alpha <- 2
> beta <- 4

> set.seed(9999) # for reproduceability
> N <- 10000 # specify length of chain
> X <- matrix(0, nrow = N, ncol = 2) # create space for chain
> X[1, ] <- c(10, 0.5) # specify initial state

> # Systematic-scan Gibbs sampler
> for (t in 2:N) {
+     # Generate X^(t) from X | Y^(t-1)
+     X[t, 1] <- rbinom(1, size = n, prob = X[t - 1, 2])
+
+     # Generate Y^(t) from Y | X^(t)
+     X[t, 2] <- rbeta(1, X[t, 1] + alpha, n - X[t, 1] + beta)
+ }
```

Example 2: The Beta-Binomial Model



Example 2: The Beta-Binomial Model

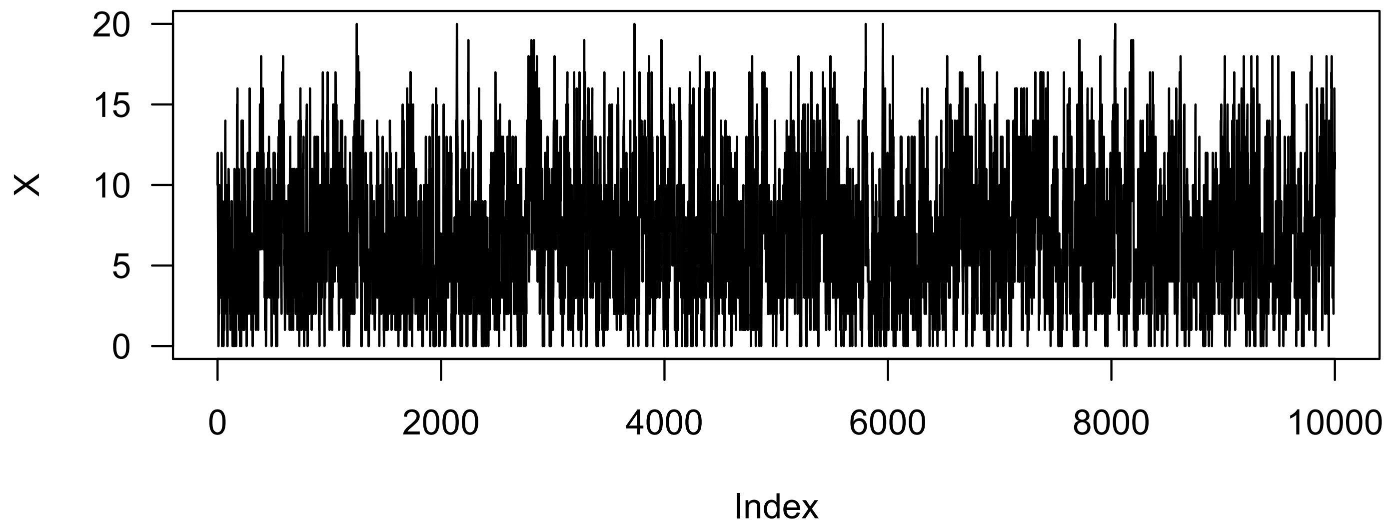


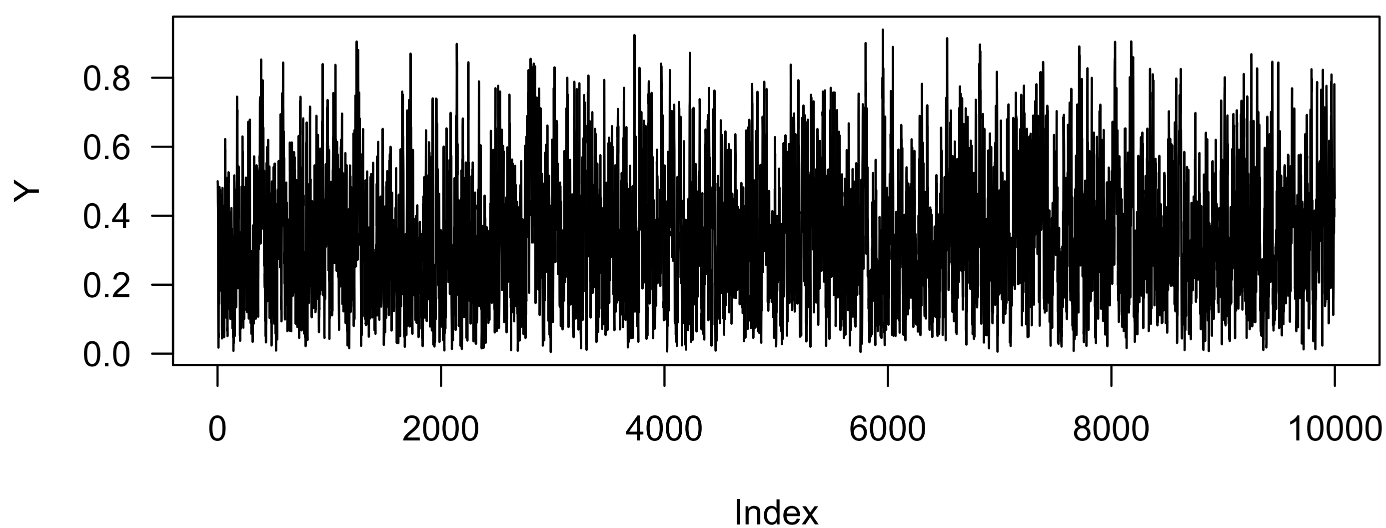
Example 2: The Beta-Binomial Model

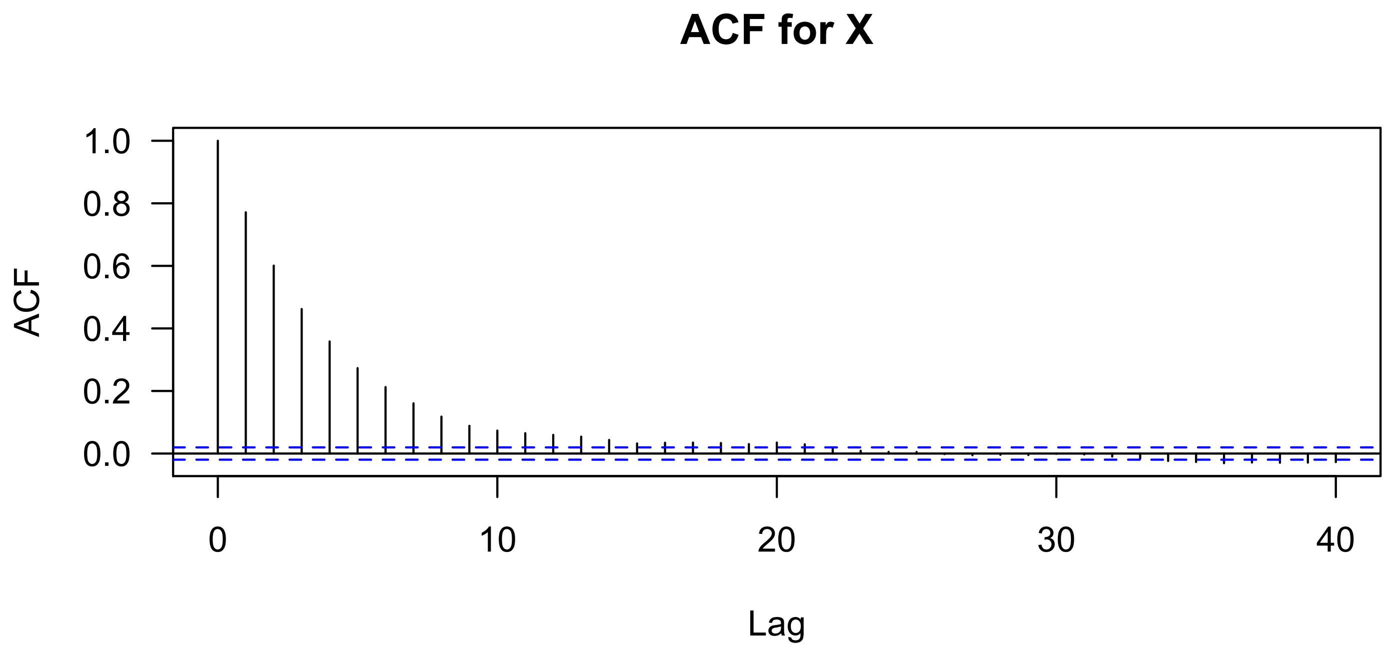
R Code for the plots:

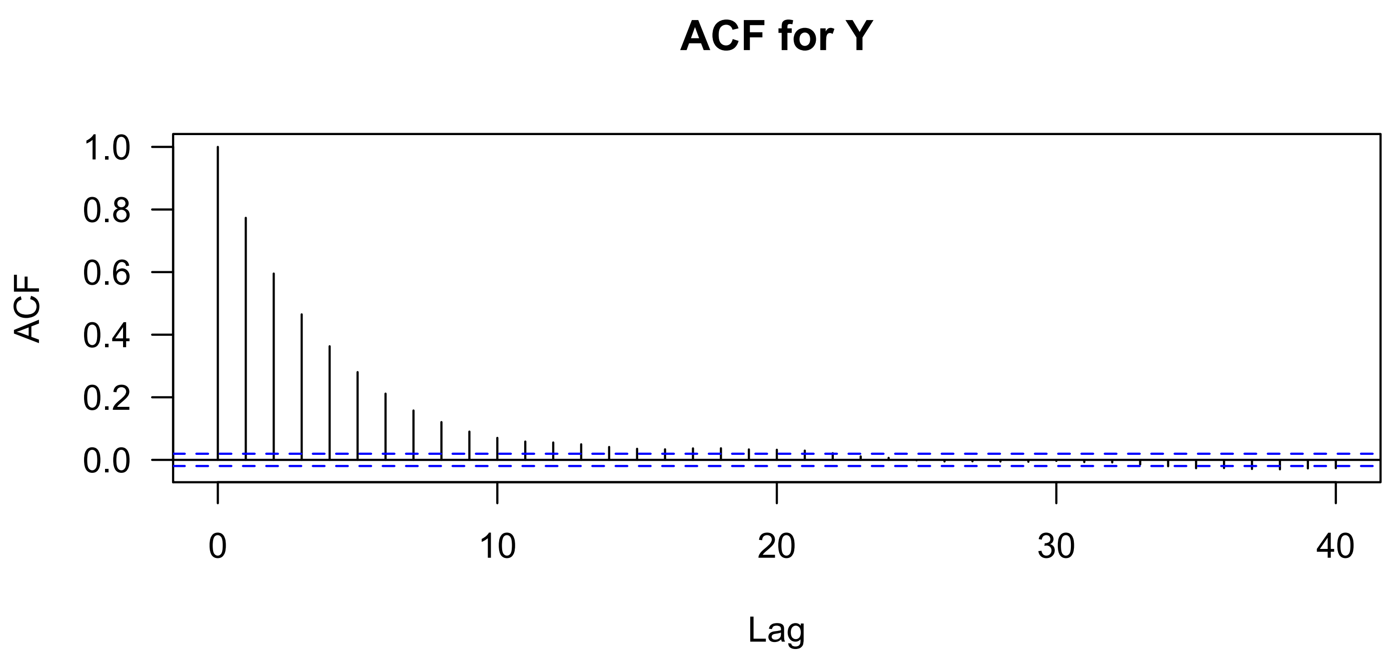
```
> # Plot the sample path for first 50 iterations
> plot(X[1:50, ], type = "b", pch = 19, cex = 0.4, las = 1,
+      main = "First 50 Iterations", xlab = "X", ylab = "Y"
+ )

> # Scatterplot of samples after 1000 burn-in iterations
> plot(X[1001:N,], pch = 19, cex = 0.4, las = 1,
+     main = "Gibbs samples after burn-in",
+     xlab = "X", ylab = "Y"
+ )
```

Example 2: The Beta-Binomial Model






ACF for X
1.0
0.8
0.6
0.4
0.2
0.0
ACF
0
10
20
30
40
Lag

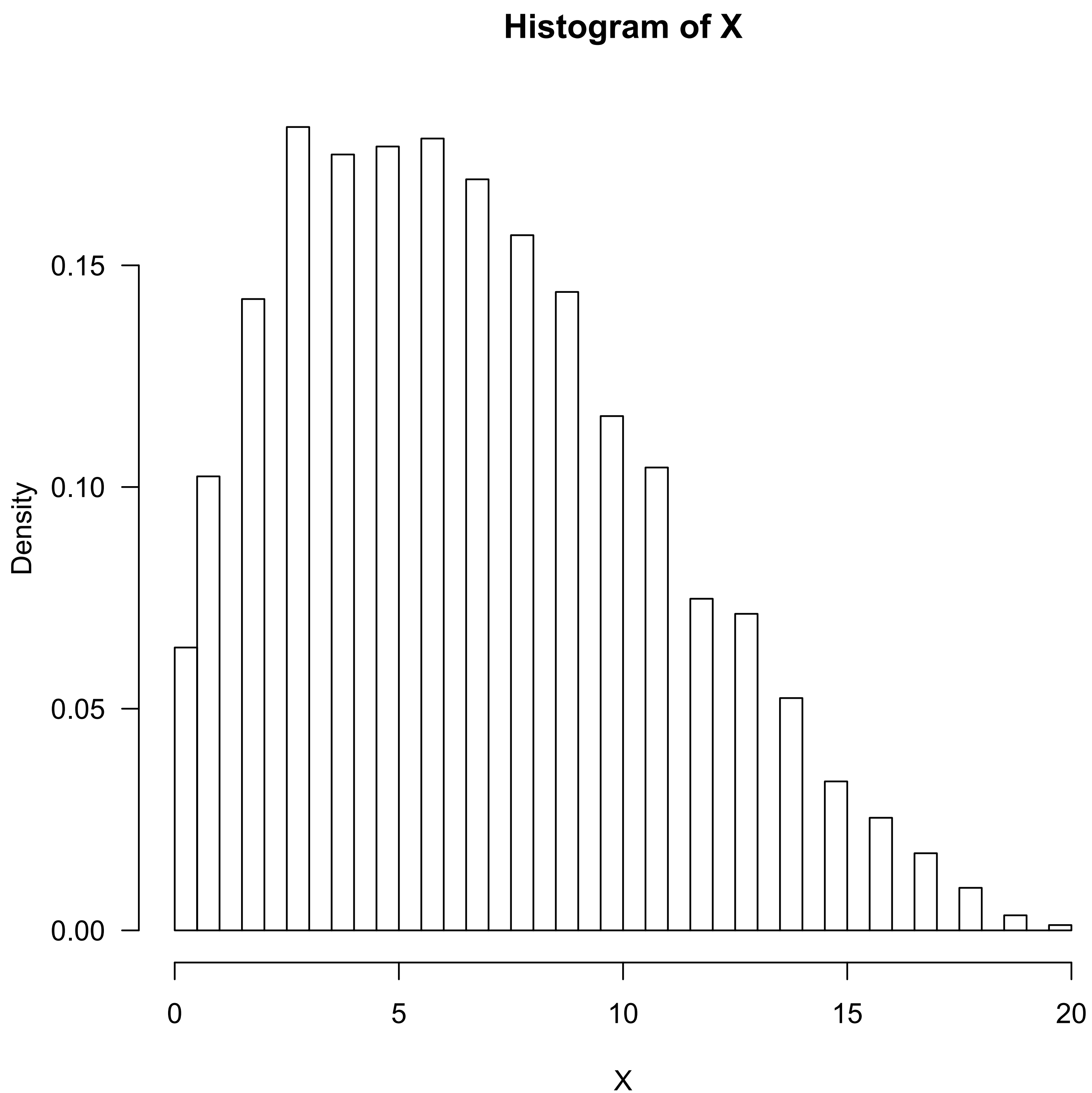

ACF for Y
1.0
0.8
0.6
0.4
0.2
0.0
ACF
0
10
20
30
40
Lag

Example 2: The Beta-Binomial Model

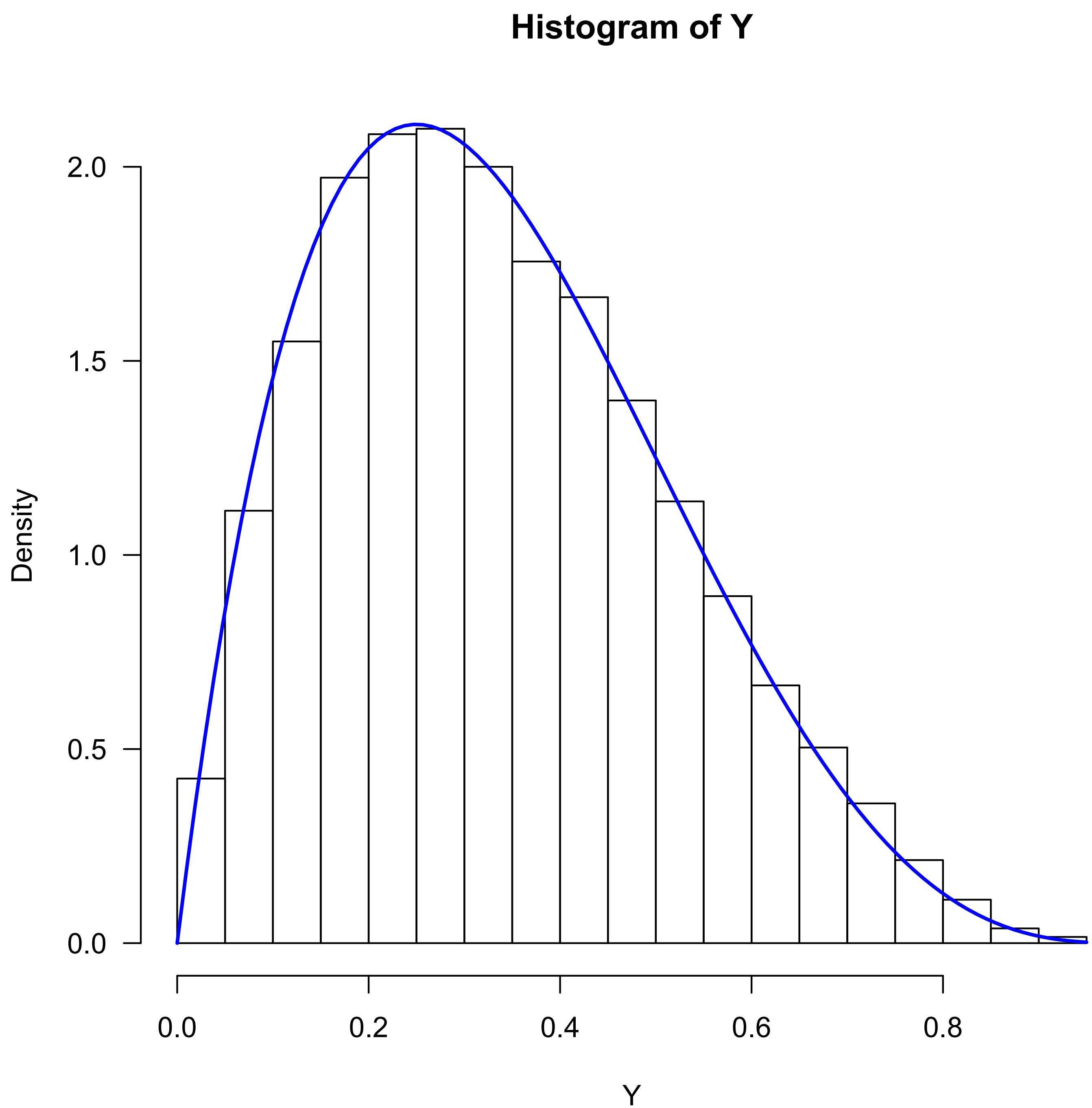
R Code for the plots:

```
> # Plot trace plot for each coordinate
> par(mfrow = c(2, 1))
> plot(X[, 1], type = "l", ylab = "X", las = 1)
> plot(X[, 2], type = "l", ylab = "Y", las = 1)

> # Plot autocorrelation function for each coordinate
> par(mfrow = c(2, 1))
> # Autocorrelation for X
> acf(X[, 1], main = "ACF for X", las = 1)
> # Autocorrelation for Y
> acf(X[, 2], main = "ACF for Y", las = 1)
```

Example 2: The Beta-Binomial Model



Example 2: The Beta-Binomial Model



Example 2: The Beta-Binomial Model

R Code for the plots:

```
> # Histogram of X: Beta-Binomial(n,alpha,beta)
> hist(X[, 1], breaks = 30, prob = TRUE,
+      las = 1, xlab = "X", main="Histogram of X"
+ )

> # Histogram of Y: Beta(alpha,beta)
> hist(X[, 2], prob = TRUE,
+      las = 1, xlab = "Y", main="Histogram of Y"
+ )
> # Superimpose theoretical Beta(alpha,beta)
> curve(dbeta(x, alpha, beta),
+       add = TRUE, col = "blue", lwd = 2
+ )
```

Example 2: The Beta-Binomial Model

- The marginal $\pi(x)$ is called the **beta-binomial** distribution.
- The beta-binomial distribution commonly arises in Bayesian statistics as the probability distribution for the number of successes in n Bernoulli trials (i.e., the binomial distribution), where the success probability is itself random and follows a $\mathrm{Beta}(\alpha, \beta)$ distribution.
- In Bayesian terminology, this example describes the **beta-binomial model**:
  - $\pi(y) \sim \mathrm{Beta}(\alpha, \beta)$ is the prior distribution
  - $\pi(x|y) \sim \mathrm{Bin}(n, y)$ is the likelihood
  - $\pi(y|x) \sim \mathrm{Beta}(x + \alpha, n - x + \beta)$ is the posterior distribution