Markov Chain Mote Carlo Methods

Chapter 6

STATS 102C: Introduction to Monte Carlo Methods

Bayesian Inference

The underlying difference between the Frequentist and Bayesian perspectives is what probability represents.

The Frequentist Perspective: $\binom{n}{y}\theta^{y}(1-\theta)^{n-y}$

- Probability represents the long-run relative frequency of random events.
- Parameters are considered (often unknown) fixed constants.

The Bayesian Perspective: $\binom{n}{y}\underline{\theta}^{y}(1-\theta)^{n-y} \quad \theta \sim f(\theta)$

- Probability represents one's subjective belief about random events.
- Parameters are considered random variables.

Frequentist Perspective

- Consider the scenario of flipping a coin n times, where the probability of heads on any given flip is θ. If Y is the number of heads in n flips, then $Y \sim Bin(n, \theta)$, with PMF/PDF

$$P_\theta(Y = y) = f(y|\theta) = \binom{n}{y} \theta^y (1-\theta)^{n-y}.$$

- The density of Y, considered as a function of θ, is called the **likelihood function** (or just **likelihood**): $L(\theta|y) = f(y|\theta)$. Suppose we observed $Y = y$ heads from n flips. Based on this data, we want to estimate θ. $\max_\theta L(\theta|y)$
- A standard (frequentist) way to estimate θ would be the maximum likelihood estimator, $\underline{\hat{\theta}_{MLE}} = \dfrac{y}{n}$.

- A $100(1-\alpha)$**% confidence interval for** θ is a random interval $[\ell(Y), u(Y)]$ such that, *before the data is gathered*,

$$P[\ell(Y) < \theta < u(Y)|\theta] = 1 - \alpha.$$

Once we observe $Y = y$, then the interval $[\ell(y), u(y)]$ is no longer random, so

$$P[\ell(y) < \theta < u(y)|\theta] = \begin{cases} 0 & \text{if } \theta \notin [\ell(y), u(y)] \\ 1 & \text{if } \theta \in [\ell(y), u(y)]. \end{cases}$$

- If we were to take many random samples and form a $100(1-\alpha)\%$ confidence interval from each one, about $100(1-\alpha)\%$ of these intervals would contain θ. The probability $1-\alpha$ is called the (frequentist) **coverage probability**.

Bayesian Perspective

In the Bayesian perspective, we are able to take our prior beliefs into account. We represent our beliefs about θ prior to observing data by a **prior distribution** $P(\theta)$.

- The prior distribution can represent past information, such as past experiments or literature, or subjective beliefs from a knowledgeable person.
- If no prior information is available (or we do not want to take it into account), we can use an **uninformative** (or **flat**) **prior**, which assigns equal density to all possibilities of the parameter.
- When using an uninformative prior, Bayesian estimators tends to be similar (sometimes identical) to frequentist estimators: The data easily outweighs a prior with no information.

- The **posterior distribution** $f(\theta|y)$ represents our updated beliefs about θ *after* observing the data. If the posterior is in the same parametric family as the prior, the prior and posterior are called **conjugate distributions**, and the prior is called a **conjugate prior** for the likelihood.

$\theta | Y \sim f(\;)$

$$P(A|B) = \frac{P(A,B)}{P(B)} = \frac{P(B|A)\cdot P(A)}{P(B)}$$

$$\underbrace{f(\theta|Y)}_{\text{posterior}} = \frac{f(Y|\theta)\cdot f(\theta)}{\int f(Y|\theta)\cdot f(\theta)\,d\theta}$$

hyper-parameters

$$\propto \underbrace{f(Y|\theta)}_{\text{likelihood}} \cdot \underbrace{f(\theta|\boxed{a_1, a_2, \ldots, a_k})}_{\text{prior}}$$

$f(Y|\theta_1\; \theta_2\; \theta_3 \ldots \theta_k) \cdot P(\theta_1)\, P(\theta_2) \ldots P(\theta_k)$

Example: Beta-Binomial Model

For our coin flipping example, suppose our prior is $P(\theta) \sim Beta(\alpha, \beta)$. Find the posterior mean.

$$y \overset{iid}{\sim} Bin(n, \theta) \qquad P(\theta) = \frac{1}{B(\alpha,\beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1}$$

$$f(\theta|y) \propto \binom{n}{y} \theta^{y} (1-\theta)^{n-y} \frac{1}{B(\alpha,\beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1}$$

$$= \binom{n}{y} \frac{1}{B(\alpha,\beta)} \theta^{y+\alpha-1} (1-\theta)^{n-y+\beta-1}$$

$$\propto \theta^{y+\alpha-1} (1-\theta)^{n-y+\beta-1}$$

set $\alpha' = y+\alpha$

$\beta' = n-y+\beta$

$$= \theta^{\alpha'-1} (1-\theta)^{\beta'-1}$$

$$f(\theta|y) = Beta(\alpha', \beta')$$

$$E[\theta|Y] = \frac{\alpha'}{\alpha'+\beta'}$$

$$= \frac{y+\alpha}{\cancel{y}+\alpha+n-\cancel{y}+\beta}$$

$$= \underbrace{\boxed{\frac{n}{n+\alpha+\beta}}}_{w}\frac{y}{n} + \underbrace{\boxed{\frac{\alpha+\beta}{n+\alpha+\beta}}}_{1-w}\frac{\alpha}{\alpha+\beta}$$

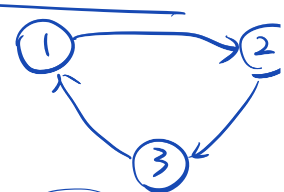
$$= w\,\underbrace{\frac{y}{n}}_{\hat\theta_{MLE}} + (1-w)\underbrace{\frac{\alpha}{\alpha+\beta}}_{E(\theta)}$$

Recall: $\theta \sim Beta(\alpha, \beta)$

$$E(\theta) = \frac{\alpha}{\alpha+\beta}$$

$$Var(\theta) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

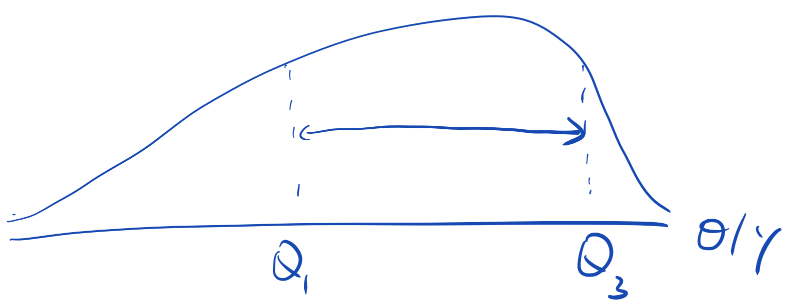
$$Mode = \frac{\alpha-1}{\alpha+\beta-2}$$



If y, n are large, $\hat\theta_{MLE} \approx E(\theta|Y)$

If y, n are small, $E(\theta) \approx E(\theta|Y)$

$$f(x) = \frac{g(x)}{Z_g}$$



The uninformative prior $\theta \sim Unif(0,1)$ $\quad E(\theta|Y) = \frac{\alpha'}{\alpha'+\beta'}$

$$f(\theta|y) \propto \binom{n}{y}\theta^y(1-\theta)^{n-y}(1) \qquad = \frac{y+1}{n+2}$$

$$\propto \theta^y(1-\theta)^{n-y}$$

set $\alpha' = y+1$

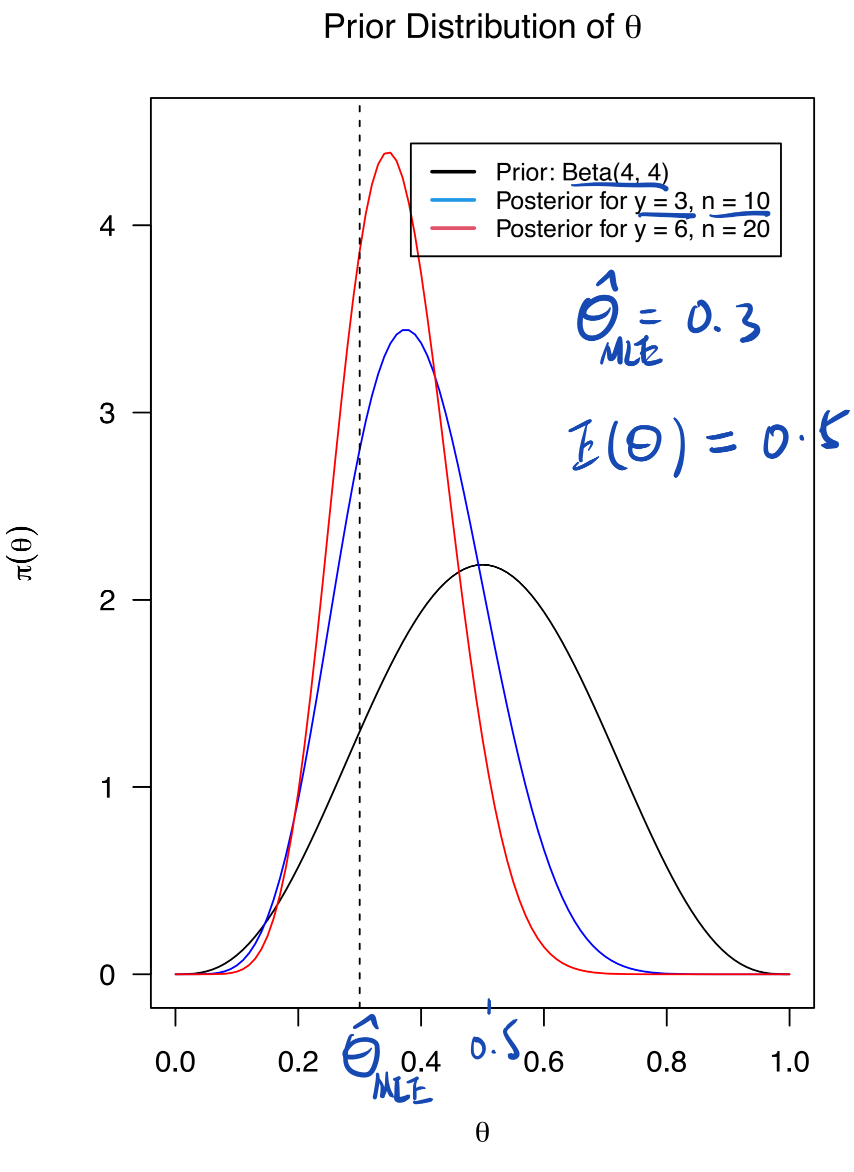
$\beta' = n-y+1$

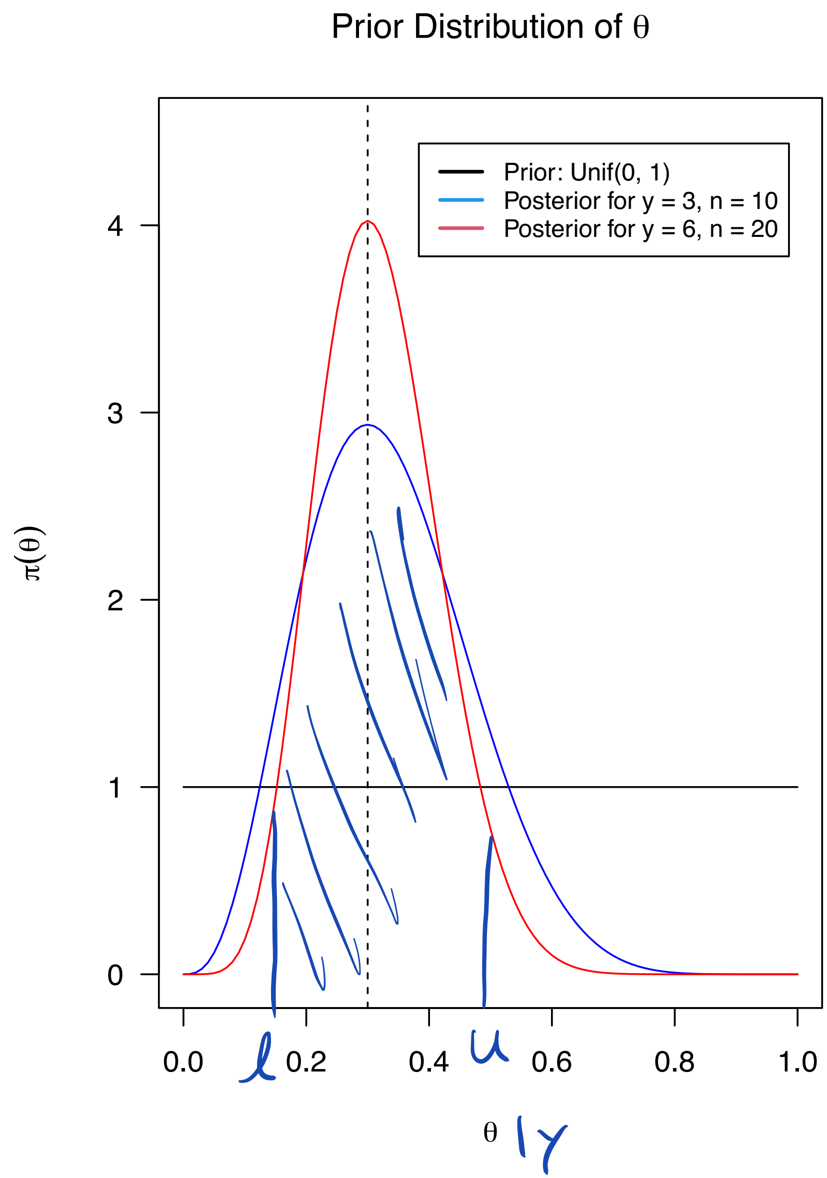
$$f(\theta|y) = Beta(\alpha', \beta')$$

$$Mode: \frac{\alpha'-1}{\alpha'+\beta'-2} = \frac{y}{n}$$

MAP

Example: Beta-Binomial Model with Different Priors





Credible Intervals

An interval $[\ell(y), u(y)]$, based on the observed data $Y = y$, is a $100(1-\alpha)\%$ credible interval for θ if

$$P[\ell(y) < \theta < u(y) | Y = y] = 1 - \alpha.$$

The probability $1 - \alpha$ is called the **(Bayesian) coverage probability**. The interpretation of a credible interval is that it describes the information about the location of the true value of θ *after* you have observed $Y = y$.

This is different from the frequentist interpretation of coverage probability, which describes the probability that the interval will cover the true value *before* the data is observed.

Quantile-based Credible Intervals

A Bayesian analogue to a frequentist confidence interval is to use posterior quantiles. If $\theta_{\alpha/2}$ and $\theta_{1-\alpha/2}$ are the $\alpha/2$ and $1-\alpha/2$ posterior quantiles of θ, then

$$P(\theta_{\alpha/2} < \theta < \theta_{1-\alpha/2} | Y = y) = 1 - \alpha,$$

so $[\theta_{\alpha/2}, \theta_{1-\alpha/2}]$ is a $100(1-\alpha)\%$ quantile-based credible interval for θ.

Here the quantile function for the Beta distribution in R is `qbeta()`.

Markov Chain Mote Carlo

The idea of Markov Chain Mote Carlo (MCMC) is to generate a Markov Chain such that its stationary distribution is the target distribution.

Main Goals:

- To generate a sequence of correlated samples from $\pi(x)$.
- To estimate $E_\pi[g(X)] = \int g(x)\pi(x)dx$.

X_1 X_2 . . . X_k

MCMC and Markov Chain

MCMC methods use Markov chain theory in reverse:

- The stationary distribution of the Markov chain is the target distribution $\pi(x)$ from which we want to sample.
- We want to find a transition kernel $Q(x, y)$ such that the corresponding Markov chain has $\pi(y)$ as its stationary distribution.

  $Q = \left(g(x,y) \right) = \left(P_{xy} \right)$

- Once we find a suitable transition kernel, we can simulate the Markov chain for a large number of steps (the **burn-in period**) until (approximate) convergence. The simulated observations after convergence are approximately from $\pi(y)$.

  proposal dist function

- The **transition kernel** for a Markov chain $\{X_t : t = 0, 1, 2, \ldots\}$ is the conditional density $Q(x, y)$ of $Y|X = x$.

$$g(x,y) = P(X_{t+1} = y \mid X_t = x) = P(Y = y \mid X = x)$$

$$\sum_y \underline{P(Y = y \mid X = x)} = 1 \qquad \int_y \underline{f(y \mid x = x)\, dy} = 1$$

Example: Uniform Transition Kernel

```
set.seed(9999) # for reproduceability

X <- 0 # specify initial state
n <- 10000 # specify length of chain
delta <- 1 # specify width parameter

for (t in 2:n) {
        # Take next step of Markov chain
        X[t] <- runif(1, X[t - 1] - delta, X[t - 1] + delta)
}
```

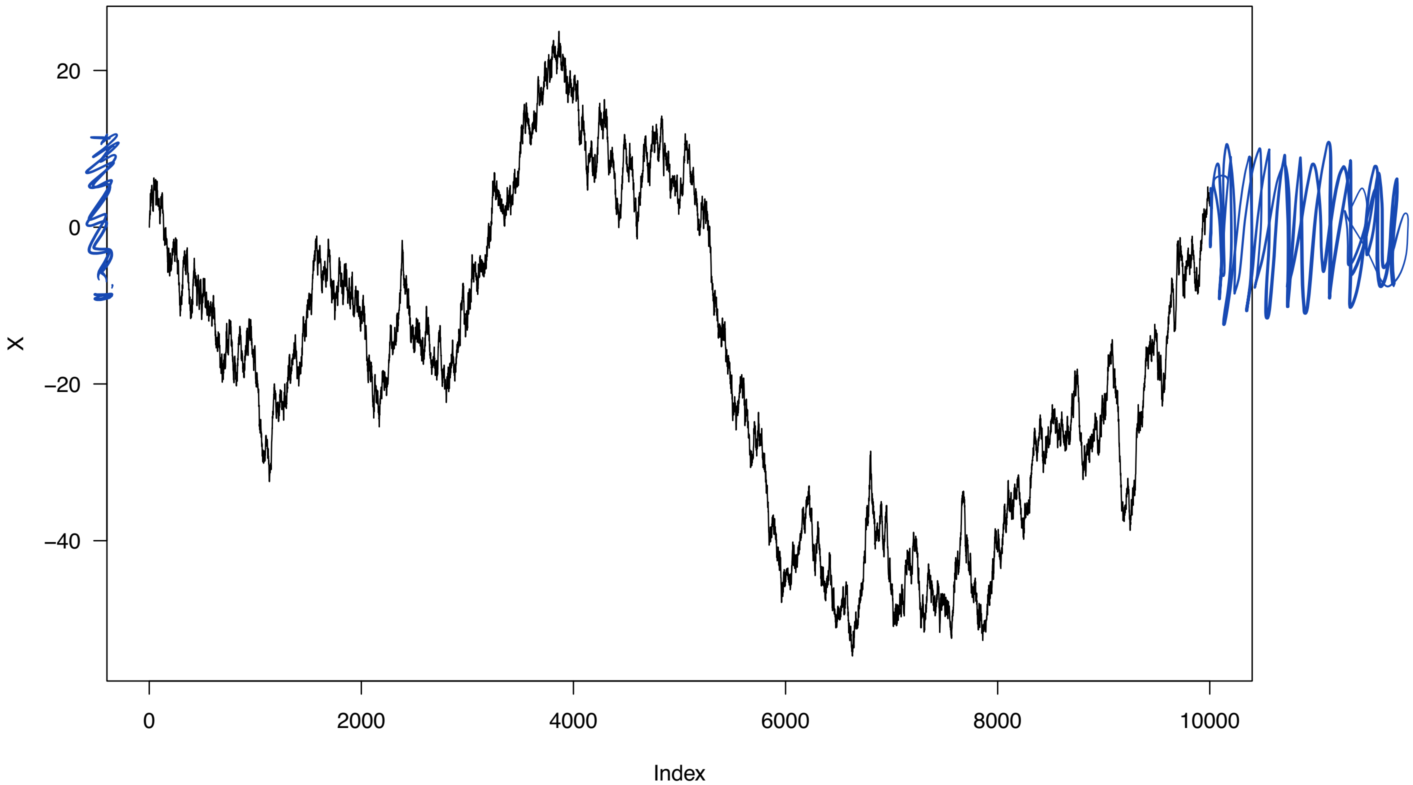
$f(y \mid x) = \text{Unif}(x-c, \; x+c)$

Say $c=1$ and $x_0 = 0$

$f(y \mid x_0 = 0) = \text{Unif}(-1, \; +1) \hookrightarrow x_1 = \frac{1}{2}$

$f(y \mid x_1 = \frac{1}{2}) = \text{Unif}(-0.5, \; 1.5)$

```r
# Plot the Markov chain over time
plot(X, type = "l", las = 1)
```



Global Balance

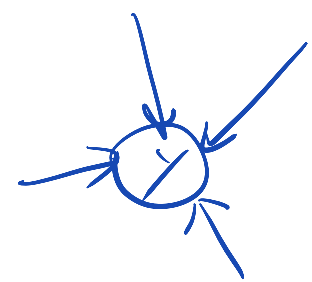
Let $\{X_0, X_1, X_2, \ldots\}$ be a Markov chain with state space $\mathbb{Z}$ and transition kernel $Q(x, y)$. A distribution $\pi(x)$ on $\mathbb{Z}$ is a stationary distribution of the Markov chain if it satisfies the **global balance** equations

$$\pi(y) = \int \pi(x) Q(x, y) dx, \quad \text{for all } y \in \mathbb{Z}.$$

If $\mathbb{Z}$ is discrete, the global balance equations can be written as

$$\pi(y) = \boxed{\sum_{x \in \mathbb{Z}} \pi(x) Q(x, y),} \quad \text{for all } y \in \mathbb{Z}.$$

$\pi(i) = \sum \pi(j) P_{ji}$



Detailed Balance

Let $\{X_0, X_1, X_2, \ldots\}$ be a Markov chain with state space $\mathbb{Z}$ and transition kernel $Q(x, y)$. The Markov chain is **time reversible** if there is a distribution $\pi(x)$ on $\mathbb{Z}$ such that $Q(x, y)$ and $\pi(x)$ satisfy the **detailed balance** (or **time reversibility**) condition

$$\underbrace{\pi(x)\overset{q(y|x)}{Q(x,y)}}_{x \to y} = \underbrace{\pi(y)\overset{q(x|y)}{Q(y,x)}}_{y \to x}, \quad \text{for all } x, y \in \mathbb{Z}. \tag{1}$$

Theorem (Detailed Balance Implies Global Balance): If $Q(x, y)$ and $\pi(x)$ satisfy the detailed balance condition, then $\pi(x)$ is a stationary distribution of the Markov chain with $Q(x, y)$ as the one-step transition kernel.

Let G be any specified irreducible Markov transition matrix (kernel)

$$G = \begin{array}{c c} & j \\ i & \left(\; g(j \mid i) \; \right) \end{array}$$

and $g(\cdot \mid i)$ is a proposal dist. funct. given i

when $X_t = i$, generate $Y \sim g(\cdot \mid i)$

If $Y = j$, then set $X_{t+1} = j$ w/ prob. $\alpha(i, j)$

If $Y \neq j$, then set $X_{t+1} = i$ w/ prob. $1 - \alpha(i, j)$

$$0 \leq \alpha(i, j) \leq 1$$

$$P(X_{t+1} = j \mid X_t = i) = \underline{g(j \mid i) \cdot \alpha(i, j)}$$

Recall: $\pi(x) \underline{Q(x, y)} = \pi(y) Q(y, x)$

$$\pi(i) g(j \mid i) \alpha(i, j) = \pi(j) g(i \mid j) \cdot \alpha(j, i)$$

If set $\alpha(i, j) = \boxed{\dfrac{\pi(j) g(i \mid j)}{\pi(i) g(j \mid i)}}$, then $\alpha(j, i) = 1$

If set $\alpha(j, i) = \dfrac{\pi(i) g(j \mid i)}{\pi(j) g(i \mid j)}$, then $\alpha(i, j) = \boxed{1}$

Set $\alpha(i, j) = \min \left\{ \dfrac{\pi(j) g(i \mid j)}{\pi(i) g(j \mid i)}, \; 1 \right\}$

$$r = mh = \frac{\pi(j) g(i \mid j)}{\pi(i) g(j \mid i)}$$

Metropolis-Hastings (M-H) sampling algorithm

There is a candidate point Y generated from a proposal distribution $g(.|X_t)$. If this candidate point is accepted, the chain moves to state Y at time $t+1$ and $X_{t+1} = Y$; otherwise the chain stays in state X_t and $X_{t+1} = X_t$. (Rizzo's book)

$$\alpha(i,j) = \min\left\{\frac{\pi(j)\, g(i|j)}{\pi(i)\, g(j|i)},\ 1\right\}$$

if $\dfrac{\pi(j)\, g(i|j)}{\pi(i)\, g(j|i)}$ $\uparrow$, most likely accept $X_{t+1} = j$

$\downarrow$, most likely accept $X_{t+1} = i$

$u \sim \text{Unif}(0,1)$

If (u < mh) accept $X_{t+1} = j$

ow let $X_{t+1} = i$

M-H Algorithm $\pi(x)$

1. Choose a proposal distribution $g(.|X_t)$.
2. Generate X_0 from the distribution g.
3. Repeat until the chain has converged to stationary distribution according to some criterion:
4. Generate Y from $g(.|X_t)$
5. Generate U from Uniform(0, 1).
6. Compute $\overset{mh}{r} = \boxed{\frac{\pi(Y)g(X_t|Y)}{\pi(X_t)g(Y|X_t)}}$
7. If (U < r), accept Y and set $X_{t+1} = Y$; otherwise set $X_{t+1} = X_t$
8. Increment t.

$$\begin{pmatrix} \pi_1 \\ \pi_1 \end{pmatrix} = \underset{N \times N}{\begin{pmatrix} \; - \; \end{pmatrix}} \underset{N \times 1}{\begin{pmatrix} | \end{pmatrix}}$$

$$\pi^T P = \pi^T$$

set $c=1$ $\quad g(Y|\underline{x}) = \text{Unif}(x-c, \; x+c) \;\; // \;\; \frac{1}{2}$

e.g. $\quad g(Y|x=0) = \text{Unif}(0-1, \; 0+1) \;\; G \;\; i,$

$g(y \mid x=1.1) = \text{Unif}(1.1-1,\ 1.1+1) \to \frac{1}{2}$

Example: Standard Normal Distribution $N(0, 1)$

Use the Metropolis-Hastings sampler to generate a sample from Standard Normal Distribution $N(0, 1)$.

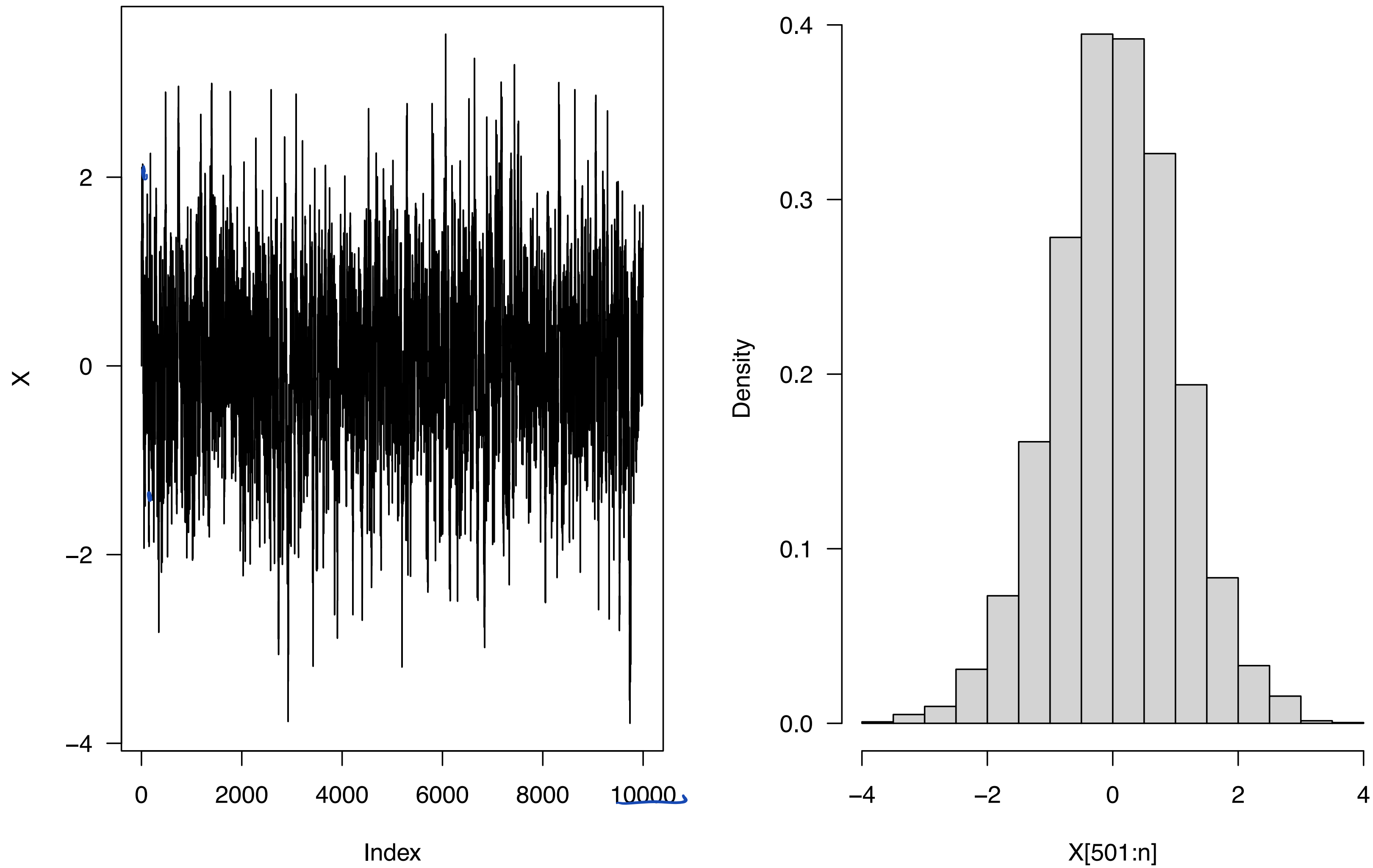
```
set.seed(9999) # for reproduceability
n <- 10000 # specifiy length of chain
X <- 0 # initialize chain
C <- 1
for (t in 2:n) {
     # Generate Y from proposal
     Y <- runif(1, X[t - 1] - C, X[t - 1] + C)
     # Compute MH ratio
     r <- min(1, exp(-0.5 * (Y^2 - X[t - 1]^2)))

     U <- runif(1, 0, 1) # Generate U from Unif(0,1)
     if (U <= r) {
          X[t] <- Y # Move to Y if U <= r
     } else{
          X[t] <- X[t - 1] # Stay at X[t - 1] if U > r
     }
}
```

$$\frac{\pi(Y)\,g(X_t \mid Y)}{\pi(X_t)\,g(Y \mid X_t)} = \frac{\frac{1}{\sqrt{2\pi}} e^{-\frac{Y^2}{2}}\, g(X_t \mid Y)}{\frac{1}{\sqrt{2\pi}} e^{-X_t^2/2}\, g(Y \mid X_t)} = e^{-\frac{1}{2}(Y^2 - X_t^2)}$$

```r
par(mfrow = c(1, 2))
# Plot the Markov chain over time
plot(X, type = "l", las = 1)

# Plot histogram after 500 burn-in iterations
hist(X[501:n], prob = TRUE, las = 1, main = "")
```



$$\theta \sim f(\theta \mid y)$$

$$\frac{\pi(y) \cdot g(x_t \mid y)}{\pi(x_t)\, g(y \mid x_t)}$$

① Assign $\theta^{(0)}$

② set $t=1$

③ sample $\theta^* \sim g(\theta^* \mid \theta^{t-1})$

④ Compute $mh = \dfrac{f(\theta^* \mid y)\, g(\theta^{t-1} \mid \theta^*)}{f(\theta^{t-1} \mid y)\, g(\theta^* \mid \theta^{t-1})}$

⑤ $u \sim \text{Unif}(0,1)$

if $(u < mh)$ set $\theta^t = \theta^*$

o.w. set $\theta^t = \theta^{t-1}$

⑥ $t = t+1$

$$y \sim \text{Ber}(\theta)$$

$$f(y \mid \theta) = \theta^y (1-\theta)^{1-y} \qquad y = \{0, 1\}$$

$$f(\theta) = 1$$

$$f(\theta \mid y) \propto \theta^y (1-\theta)^{1-y}$$

$$f(\theta \mid y=0) = \int_0^1 \theta^y (1-\theta)^{1-y}\, d\theta$$

$$g(\cdot \mid \theta^{t-1}) = \text{Unif}(0,1)$$

$$mh = \frac{f(\theta^* \mid y) \cdot g(\theta^{t-1} \mid \theta^*)}{f(\theta^{t-1} \mid y)\, g(\theta^* \mid \theta^{t-1})} = \frac{f(\theta^* \mid y)}{f(\theta^{t-1} \mid y)}$$

Two main challenges to MCMC simulation inference

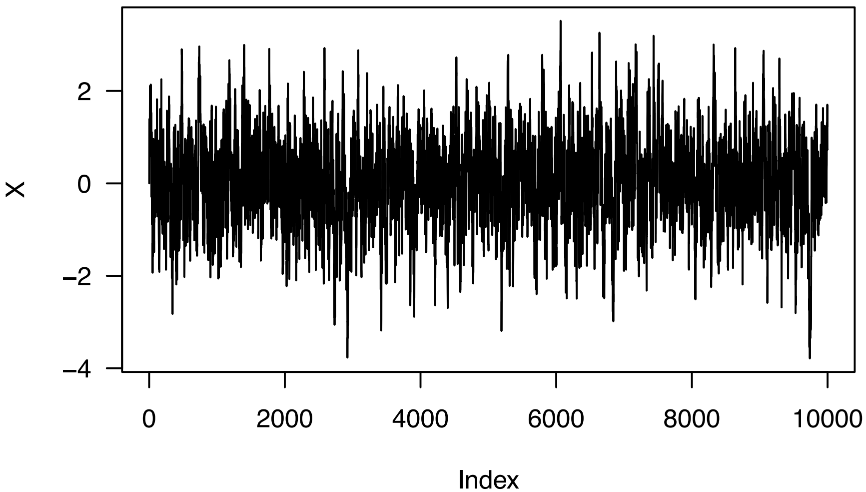
1. If the iterations have not proceeded long enough, the simulations may be grossly un-representative of the target distribution.

2. Simulation inference from correlated draws is generally less precise than from the same number of independent draws. Such correlation can cause inefficiencies in simulations.

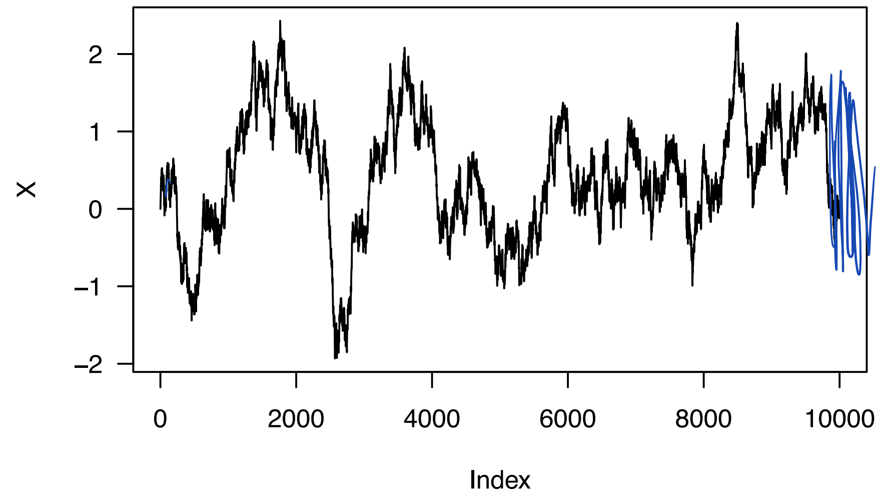
   Reference: Bayesian Data Analysis 3rd edition by Andrew Gelman etc.

Visualization of the Problem

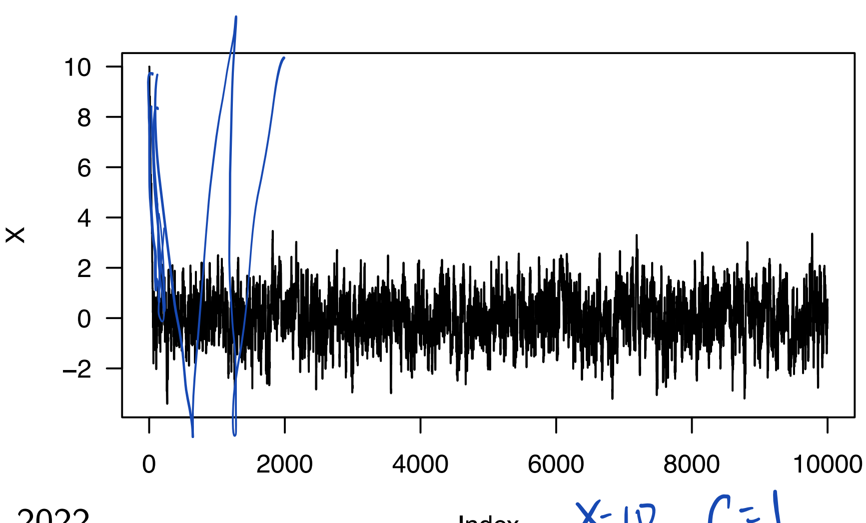
One way to visualize the problem is to plot the Markov chain over time. We return to the example 6.3.1 and consider different X_0 and C.



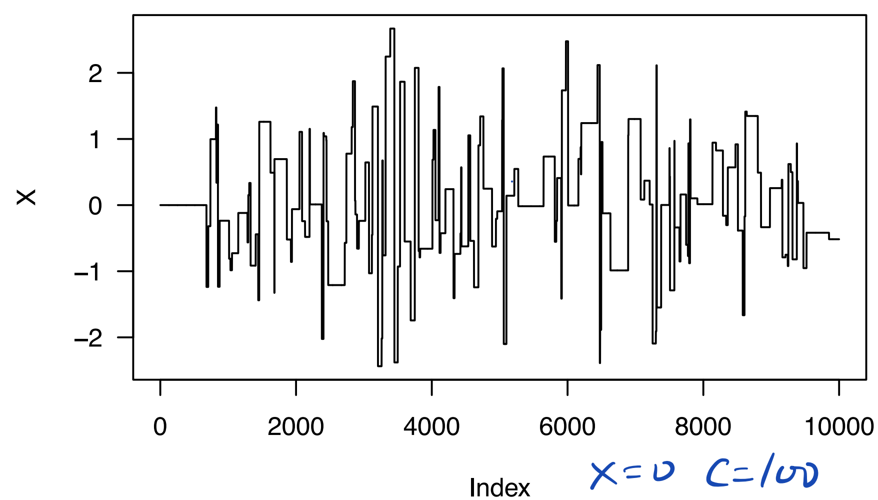
X = 0 and C = 1



X = 0 and C = 0.1



X=10 C=1



X=0 C=100

The variance for correlated samples

When we consider direct Monte Carlo method, the estimator for the mean was computed by summing up samples since they were independent. However, in MCMC the samples are not independent, so we cannot apply the same way to compute the variance of our estimator. If the Markov chain is irreducible, aperiodic, and has a stationary distribution, then the distribution of X_n will converge to the stationary distribution.

$X_0 \; X_1 \; \cdots \quad X_k \; \cdots \quad X_n \; \cdots$

If k and n are large, $f(X_k) = f(X_n) = f_{stationary}(x)$

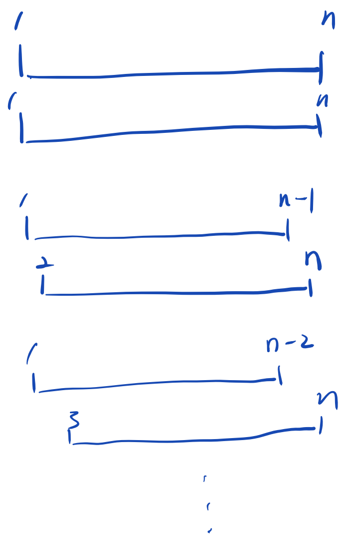
$$Var(X_k) \cong Var(X_n) \cong \sigma^2_{stationary}$$

$$\underbrace{Cov(X_k, X_n)}_{\text{function of } n} = \left[Var(X_k)\, Var(X_n)\right]^{\frac{1}{2}} \cdot corr(X_k, X_n)$$

$$= \sigma^2_{stationary} \cdot corr(X_k, X_n) \simeq 0 \Rightarrow X_k \perp X_n$$

k fixed

AC - : Auto correlation of function



$\Rightarrow \quad \mathrm{corr}(X_{1:n}, X_{1:n}) = 1$

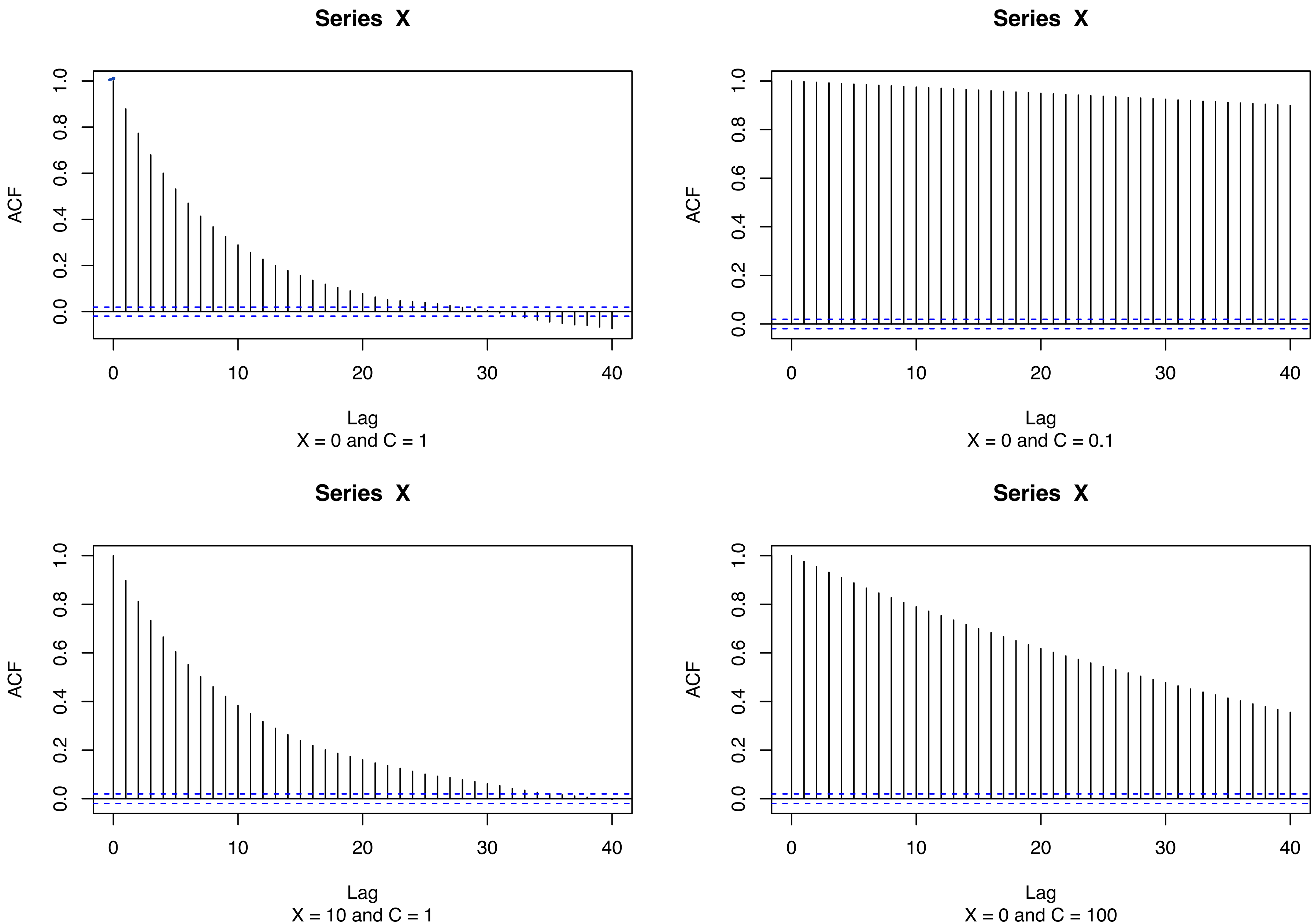
$\Rightarrow \quad \mathrm{corr}(X_{1:n-1}, X_{2:n}) \qquad \mathrm{lag} = 1$

$\Rightarrow \quad \mathrm{corr}(X_{1:n-2}, X_{3:n}) \qquad \mathrm{lag} = 2$

$\vdots$

$\mathrm{corr}(X_{1:n-t}, X_{1+t:n}) \qquad \mathrm{lag} = t$

Example 6.3.1, (Cont.)



Choosing the Proposal Distribution

The Metropolis-Hastings algorithm induces a Markov chain that converges to the stationary (limiting) distribution $\pi(x)$ if the Markov chain satisfies certain regularity conditions:

- Irreducibility
  (All states communicate with each other.)
- Aperiodicity
  (Every state has a period of 1.)
- Positive recurrence
  (Expected time until the chain returns to any state is finite.)

Positive recurrence is trivially satisfied for irreducible and aperiodic Markov chains with finite state spaces.

Sufficient (but not necessary) Conditions

- The induced Markov chain is irreducible if
  $q(x, y) > 0 \quad$ for all $x, y \in Supp(\pi)$. In other words, every state can be reached in a single transition.

- The induced Markov chain is aperiodic if

$$P(x^{(t)} = x^{(t-1)}) > 0$$

  In other words, there is positive probability that the chain remains in the current state.

Spread of the proposal distribution

The spread (variance) of the proposal distribution affects the behavior of the Markov chain in two main ways:

- Acceptance rate
  (How often the proposal is accepted.)
- Mixing rate
  (How long it takes to move through the state space.)

A Markov chain that moves quickly through the state space is said to have **good mixing behavior**.

Good Mixing Behavior

- If the spread of the proposal is too large:
  - The proposal moves through the state space quickly (good mixing behavior).
  - The proposed state may be far from the current state, and the probability of acceptance will be low.
- If the spread of the proposal is too small:
  - The probability of acceptance will be high.
  - The chain will take a long time to move through the state space, and low density regions will be undersampled.
- Both situations above will exhibit high **autocorrelation**: Correlation between subsequent values in the chain.
- An ideal proposal distribution will balance between good mixing behavior and a high acceptance rate.

Expected lifetime beta-binomial model (Example 11.3 in Statistical Computing with R — 2nd edition)

A machine produces parts that are either defective or not. Let p denote the proportion of working among all part that might be produced by this machine. If we measure time discretely, the probability that an individual survives k period is

$$P(T = k) = p^k(1 - p),\ k = 0, 1, 2, \ldots$$

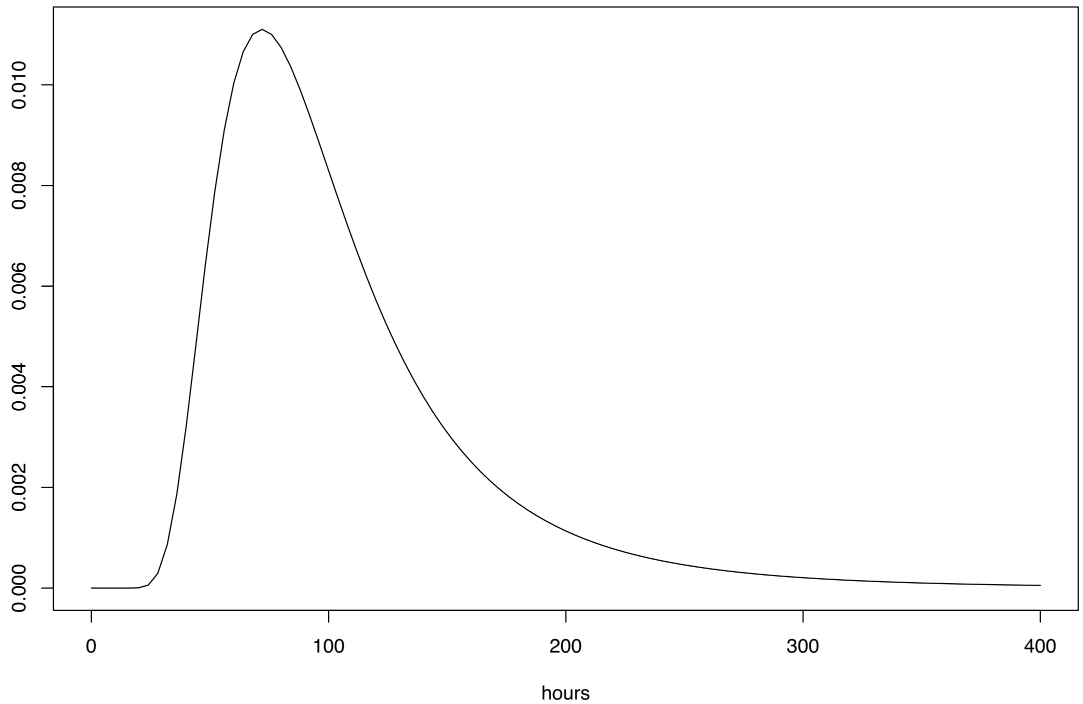
Suppose that 20 parts are tested for 24 hours and four failure times are recorded as follows:

5.1, 14.0, 14.6, 14.7, while all other parts survived with unobserved failure times. Suppose we regard each hour of testing as a Bernoulli trial.

Develop a model for predicting the expected future lifetime of a randomly selected item from this machine.

The density $f_\mu(\mu)$ should be evaluated using logarithms for numerical stability.

```
f.mu <- function(x) {
    exp(- lbeta(432, 5) + 431 * log(x) -
    431 * log(1 + x) - 6 * log(1 + x))
}
curve(f.mu(x), from=0, to=400, xlab="hours", ylab="")
```

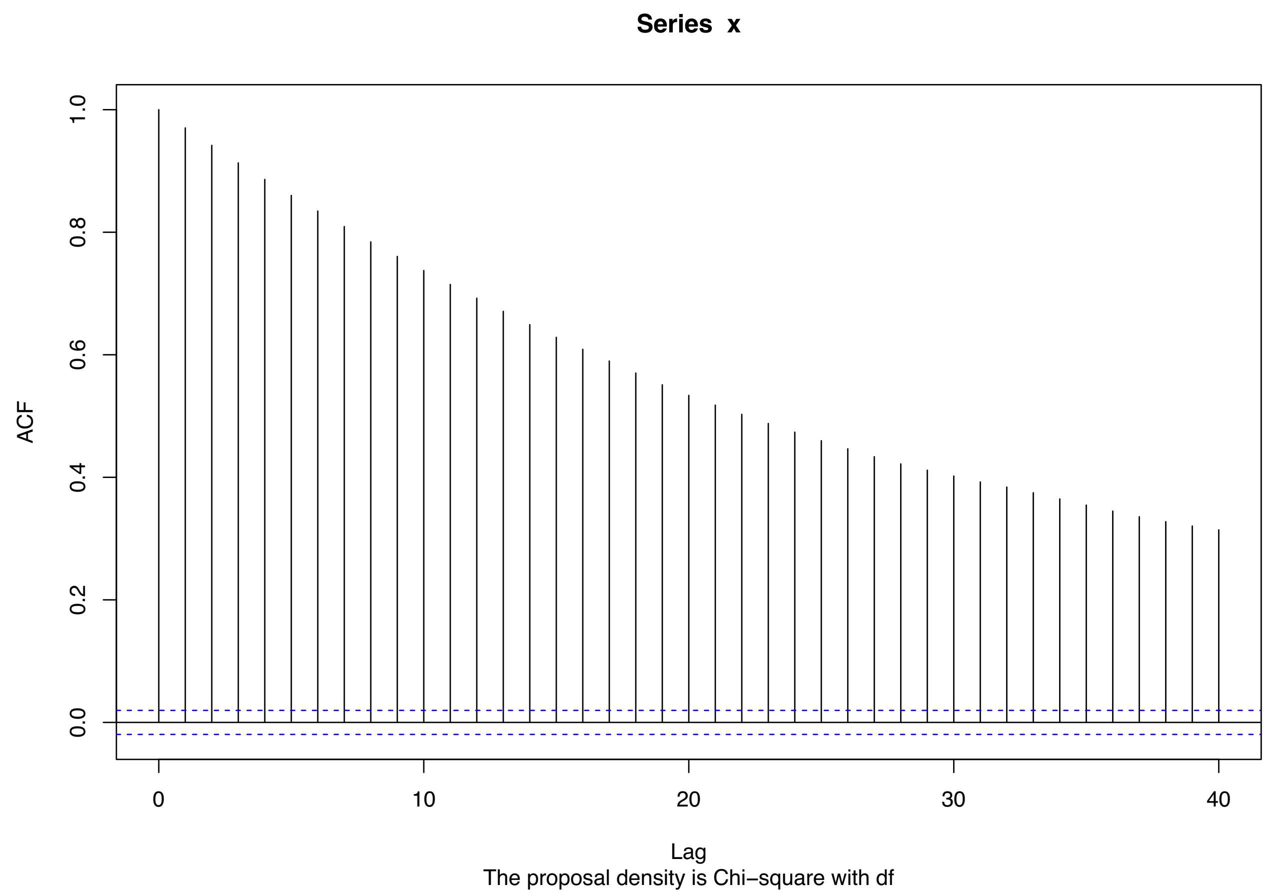


The next step is to generate replicates from $f_\mu(\mu)$ using M-H algorithm. The ratio $f_\mu(Y)/f_\mu(X_t)$ can be simplified and computed with logarithms using **fr**:

```
fr <- function(x, y) {
     a <- 431 * (log(y) - log(x))
     b <- 437 * (log(1 + x) - log(1+y))
     return(exp(a + b))
}
```

Implementation of the algorithm

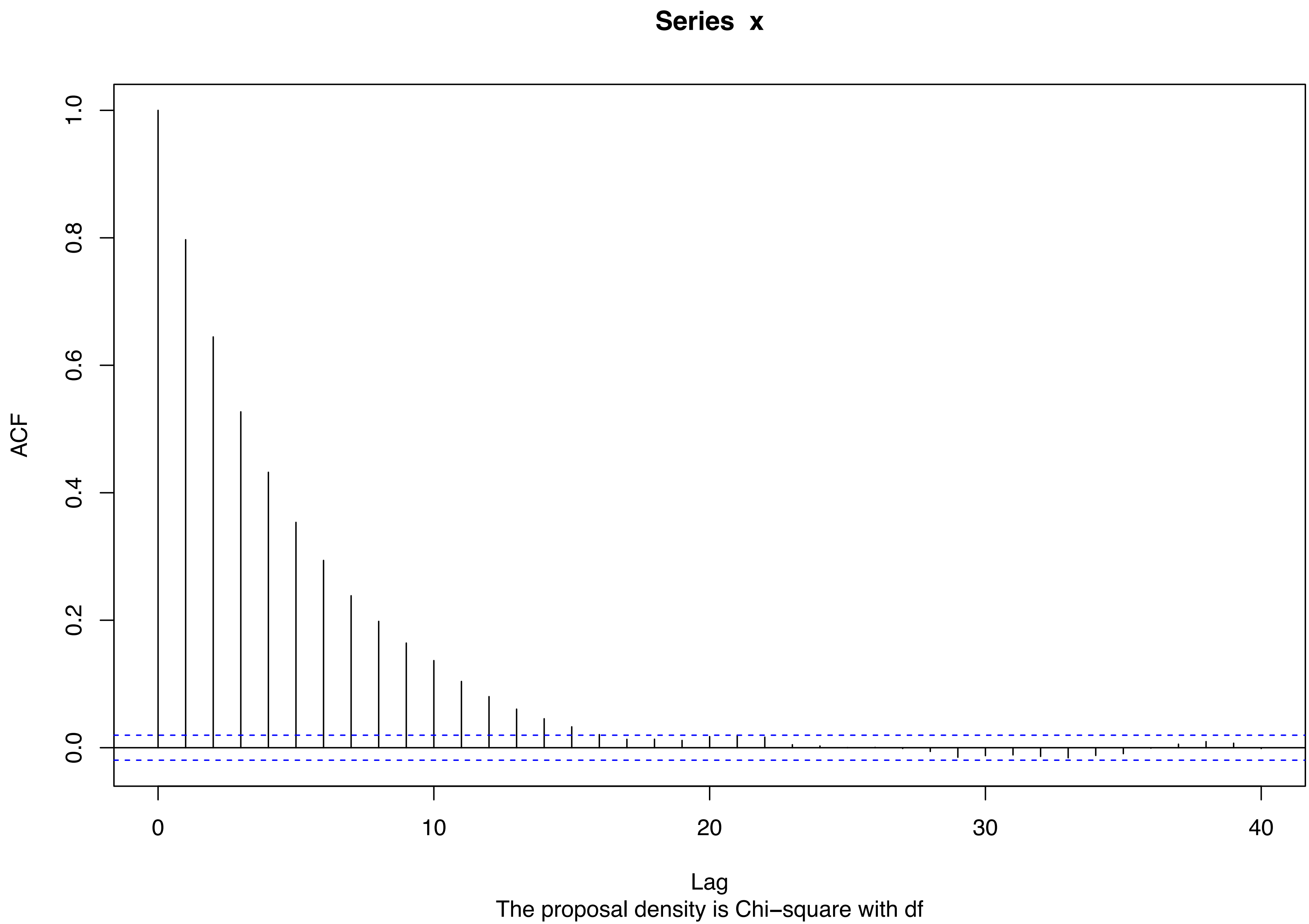
```
m <- 10000; x <- numeric(m);  k <- 0
x[1] <- rchisq(1, df = 1)
u <- runif(m)
for (i in 2:m) {
    xt <- x[i-1]
    y <- rchisq(1, df = xt)
    r <- fr(xt, y) * dchisq(xt, df = y) /  dchisq(y, df = xt)
  if (u[i] <= r) {
    x[i] <- y
  } else {
    x[i] <- xt
    k <- k+1
  }
}
```


Series x
ACF
1.0
0.8
0.6
0.4
0.2
0.0
0
10
20
30
40
Lag
The proposal density is Chi−square with df

Considering a Different Proposal Distribution

We can consider a different proposal distribution to improve the generated output. Here we use the gamma distribution with two parameters, shape and rate.

```
m <- 10000; x <- numeric(m); a <- 5
x[1] <- rlnorm(1)  #initialize chain
k <- 0
u <- runif(m)
for (i in 2:m) {
 xt <- x[i-1]
 y <- rgamma(1, shape=a, rate=a/xt)
 r <- fr(xt, y) *
   dgamma(xt, shape=a, rate=a/y) / dgamma(y, shape=a, rate=a/xt)
 if (u[i] <= r) {
   x[i] <- y
 } else {
   x[i] <- xt
   k <- k+1     #y is rejected
 }
}
```


Series x
ACF
1.0
0.8
0.6
0.4
0.2
0.0
0
10
20
30
40
Lag
The proposal density is Chi−square with df