Aula 2: Algoritmo de Metropolis-Hastings

Prof. Dr. Eder Angelo Milani

30/06/2023

Métodos Monte Carlo via Cadeia de Markov

Métodos de Monte Carlo via Cadeia de Markov (MCMC) abrangem uma estrutura geral de métodos introduzidos por Metropolis et. al (1953) e Hastings (1970) para integração de Monte Carlo.

A abordagem MCMC para amostrar da $f(.)$ é construir uma cadeia de Markov com distribuição estacionária $f(.)$ e executá-la por um tempo suficientemente longo, até que ela convirja (aproximadamente) para a distribuição estacionária.

Obs.: Seja $X_0, X_1, \ldots,$ um processo estocástico com espaço de estados finito ou infinito enumerável.

Propriedade de Markov: a probabilidade de que a cadeia assuma um certo valor futuro, quando o seu estado atual é conhecido, não se altera, caso seja conhecido o seu comportamento passado.

Em termos probabilístico, temos que

$$P(X_{t+1} = y | x_t, x_{t-1}, \ldots) = P(X_{t+1} = y | x_t)$$

O objetivo é gerar valores de uma distribuição $p(.)$ simulando uma cadeia de Markov.

O Algoritmo de Metropolis-Hastings

Os algoritmos de Metropolis-Hastings são uma classe dos métodos de Monte Carlo via cadeia de Markov, incluindo os casos especiais do amostrador de Metropolis, amostrador de Gibbs e o passeio aleatório.

A principal ideia é gerar uma cadeia de Markov $\{x_t, t = 0, 1, \ldots\}$ tal que sua distribuição estacionária é a distribuição desejada. O algoritmo deve especificar, para um determinado estado x_t, como gerar o próximo estado x_{t+1}. Em todos os algoritmos de amostragem Metropolis-Hastings (M-H) existe um valor candidato y gerado a partir de uma distribuição de posposta $g(.|x_t)$. Se o valor candidato é aceito, a cadeia se move para o estado y no tempo $t + 1$ e $x_{t+1} = y$; caso contrário, a cadeia permanece no estado x_t e $x_{t+1} = x_t$. *Note que a distribuição de proposta pode depender do estado anterior x_t.*

A escolha da distribuição de proposta é muito flexível, mas a cadeia gerada por essa escolha deve satisfazer certas condições de regularidade. A distribuição de proposta deve ser escolhida para que a cadeia gerada convirja para uma distribuição estacionária (a distribuição desejada f). Uma distribuição de proposta com o mesmo suporte da distribuição desejada usualmente satisfaz as condições de regularidade.

Amostrador de Metropolis-Hastings

O amostrdor de Metropolis-Hastings gera uma cadeia de markov $\{x_0, x_1, ...\}$ como a seguir

1- Escolha a distribuição de proposta $g(.|x_t)$.

2 - Gerar x_0 da distribuição g.

3 - Repetir

(a) gerar y da distribuição $g(.|x_t)$;

(b) gerar u da distribuição uniforme(0,1);

(c) se $u \leq \frac{f(y)g(x_t|y)}{f(x_t)g(y|x_t)}$ aceitar y e adotar $x_{t+1} = y$, caso contrário, $x_{t+1} = x_t$;

(d) incrementar t.

obs.:

(i) As repetições devem ocorrer até que a cadeia convirja para a distribuição estacionária de acordo com algum critério.

(ii) No passo 3(c) o valor y é aceito com probabilidade

$$\alpha(x_t, y) = min\left(1, \frac{f(y)g(x_t|y)}{f(x_t)g(y|x_t)}\right),$$

assim é necessário apenas conhecer a densidade da distribuição desejada a menos de uma constante.

Assumindo que a distribuição de proposta satisfaz as condições de regularidade, a cadeia M-H irá convergir para a única distribuição estacionária π. O algoritmo é projetado para que a distribuição estacionária da cadeia M-H seja realmente a distribuição desejada, f.

Exemplo 1 Faça o que se pede a seguir

(1) Gerar uma amostra de tamanho 200 da distribuição Bernoulli($\theta = 0.3$).

(2) Utilizando a abordagem Bayesiana, considere a ditribuição a priori como Beta($\alpha = 2, \beta = 2$), obtenha uma cadeia de Markov com 11 mil valores da distribuição a posteriori.

(3) Obtenha uma mostra aleatória de 1.000 valores da distribuição a posteriori.

(4) Compare gráficamente as distribuições teórica e amostrada.

(5) Obtenha a partir da amostra a estimativa pontual adotando a função perda quadrática, compare com o valor teórico.

Solução:

```r
set.seed(2023)
n <- 100
theta=0.3
z <- rbinom(n, 1, theta)
table(z)
```

```
## z
##  0  1
## 67 33
```

A distribuição a priori é $h(\theta) \propto \theta(1-\theta)$, combinando com a função de verossimilhança $L(\theta|\underline{z}) \propto \theta^{\sum z_i}(1-\theta)^{200-\sum z_i}$, tem se

$$h(\theta|\underline{z}) \propto h(\theta)L(\theta|\underline{z}) \propto \theta^{1+\sum z_i}(1-\theta)^{1+200-\sum z_i}.$$

Vamos adotar que a geração do candidato será realizada utilizando a distribuição Uniforme$(0,1)$. Logo o valor y é aceito com probabilidade

$$\alpha(x_t, y) = \alpha(x_t, y) = min\left(1, \frac{f(y)g(x_t|y)}{f(x_t)g(y|x_t)}\right),$$

note que a distribuição f é a densidade a posteriori, no entanto não temos a forma fechada, mas sabemos que existe uma constante k tal que

$$h(\theta|\underline{z}) = k\theta^{1+\sum z_i}(1-\theta)^{1+200-\sum z_i},$$

logo

$$\alpha(x_t, y) = min\left(1, \frac{h(y|\underline{z})g(x_t|y)}{h(x_t|\underline{z})g(y|x_t)}\right) = min\left(1, \frac{ky^{1+\sum z_i}(1-y)^{1+200-\sum z_i}}{kx_t^{1+\sum z_i}(1-x_t)^{1+200-\sum z_i}}\right),$$

note que podemos eliminar a constante k da última expressão, ou seja, não é necessário conhecer a constante que normaliza a distribuição a posteriori. Observe também, que na equação acima, foi substituído a expressão da geradora de candidato.

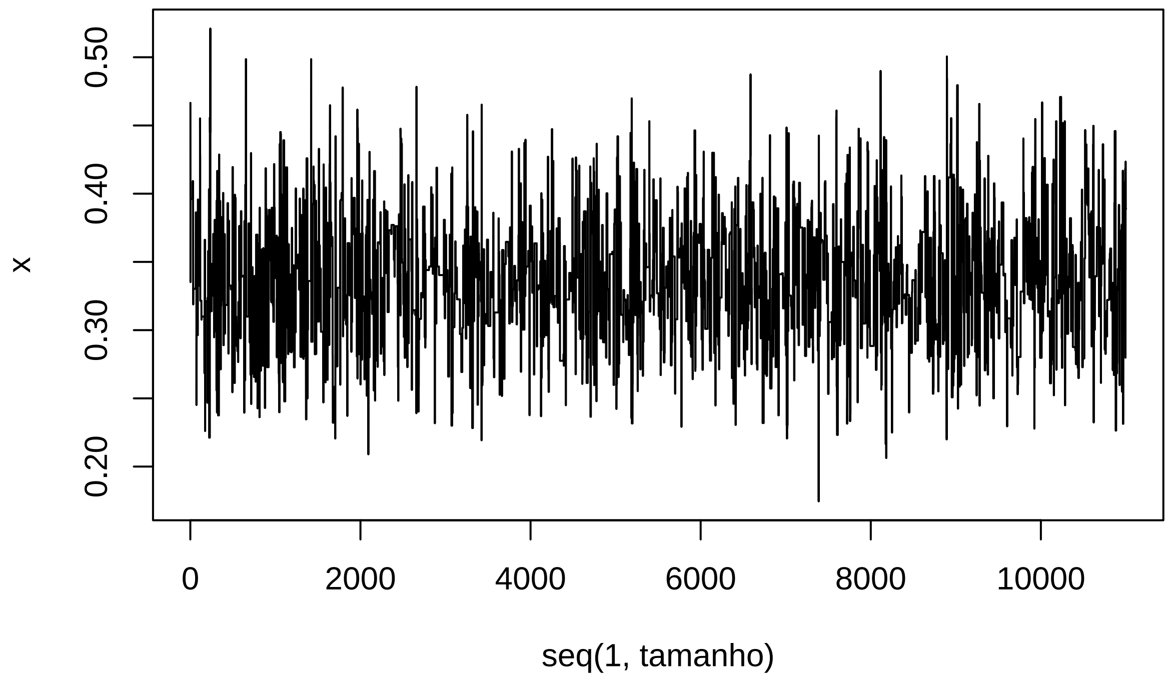
```r
set.seed(2023)

## definindo o calculo da razao
razao = function(amostra, x, y){
  n <- length(amostra)
  numerador <- y^(1+ sum(amostra))*(1-y)^(1+n- sum(amostra))
  denominador <- x^(1+ sum(amostra))*(1-x)^(1+n- sum(amostra))
  return(numerador/denominador)
}

tamanho = 11000
x = numeric(tamanho)
x[1] = runif(1)
x[1]
```

```
## [1] 0.4666139
```

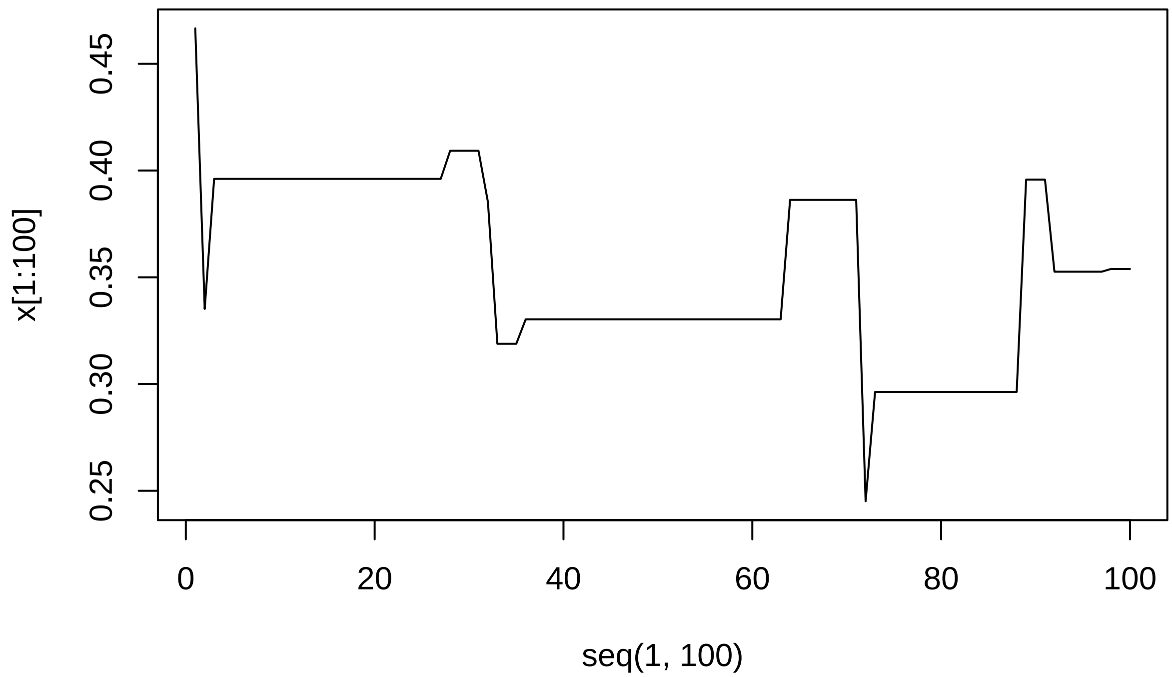
```r
for (t in 2:tamanho){
  y = runif(1)
  u = runif(1)
  x[t]=ifelse(u <= razao(z, x[t-1], y), y, x[t-1])
}

summary(x)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.1745  0.3048  0.3363  0.3364  0.3665  0.5211
```

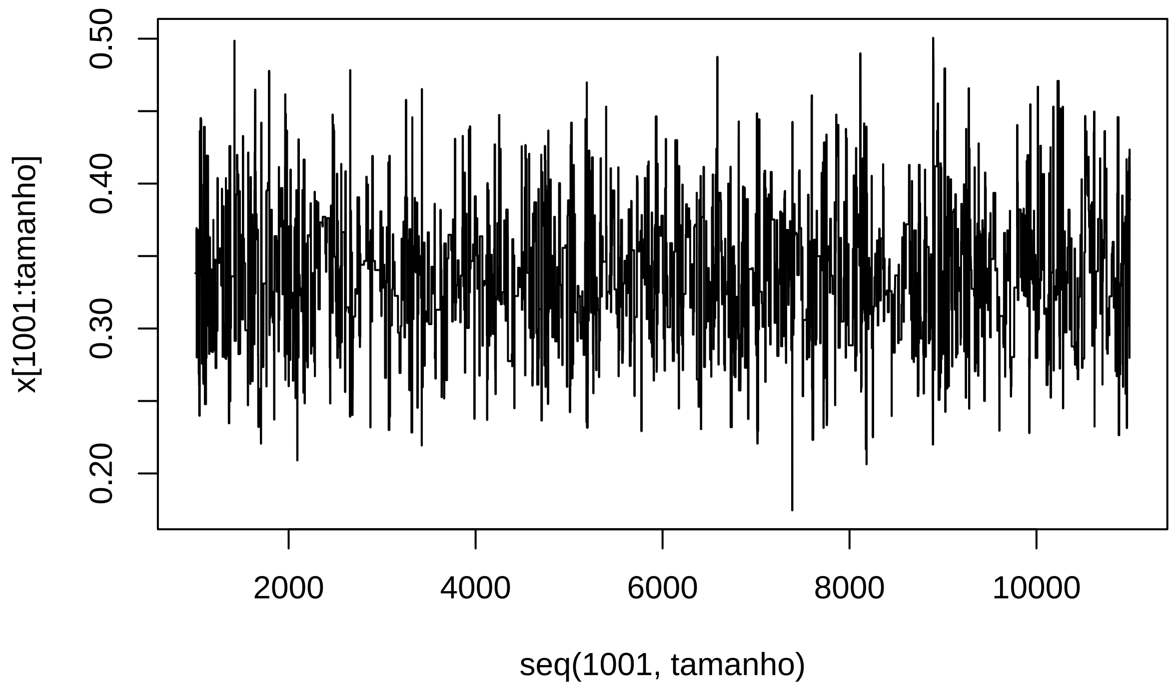
```r
# grafico dos valores gerados
plot(seq(1,tamanho), x, type = "l")
```




```r
plot(seq(1,100), x[1:100], type = "l")
```



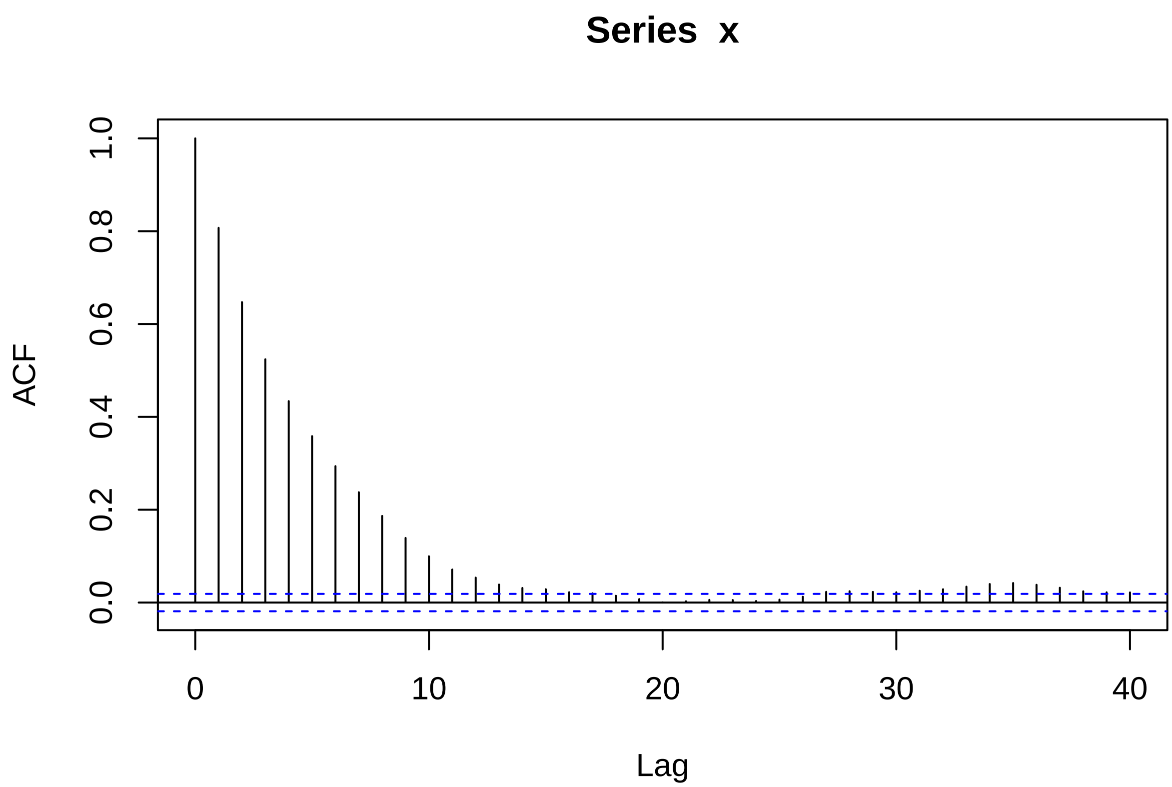

```
# desconsiderar as primeiras 1000 observações - eliminar o efeito do valor inicial
plot(seq(1001,tamanho), x[1001:tamanho], type = "l")
```



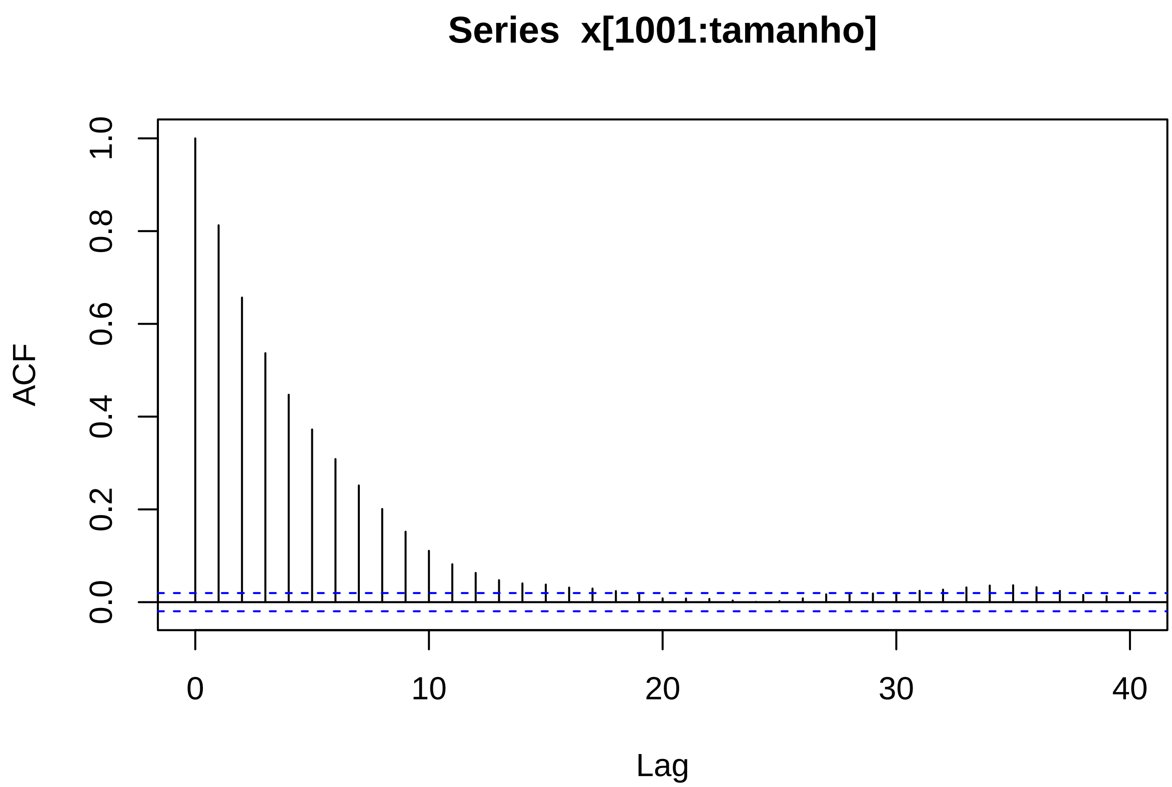

```r
# quantidade de rejeições
1-length(unique(x))/tamanho
```

```
## [1] 0.8547273
```

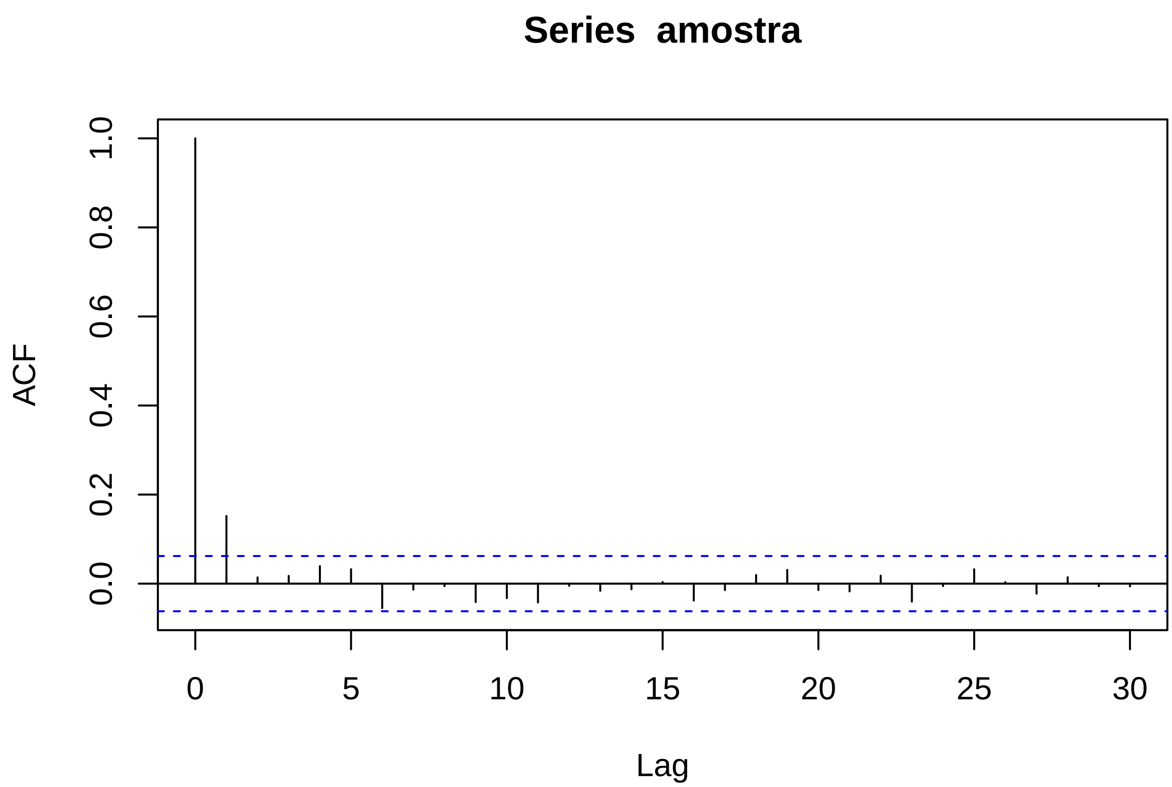
```r
# calcular a correlacao na cadeia
acf(x)
```



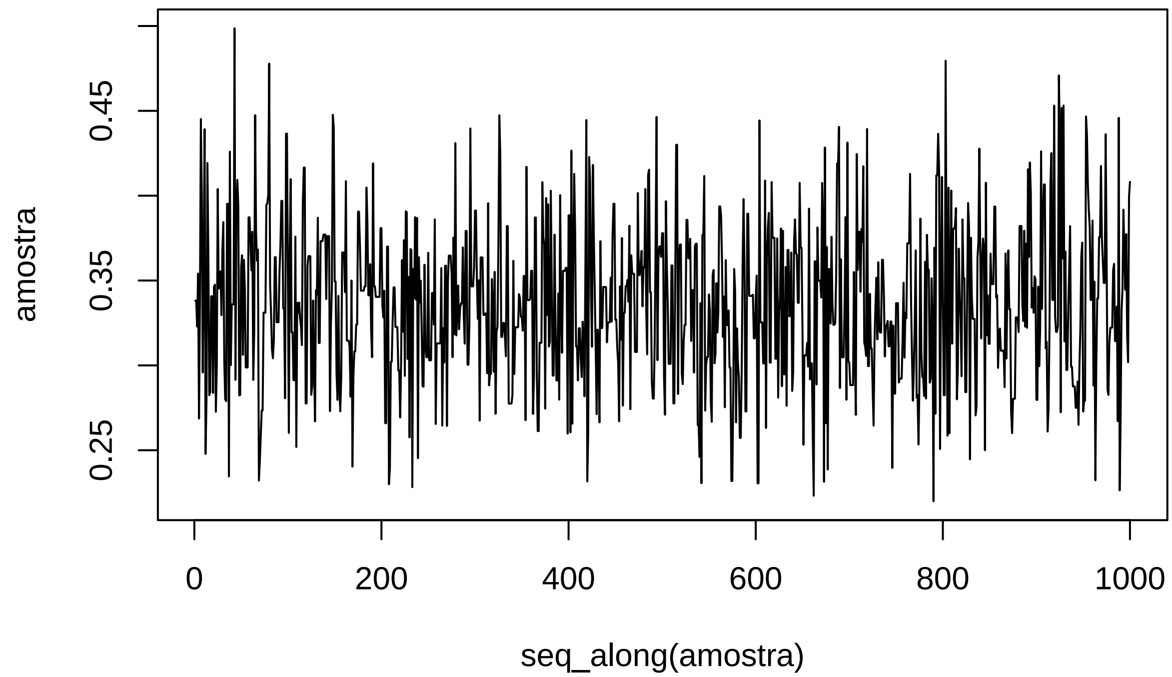

```
# calcular a correlacao na cadeia excluindo os primeiros 1000 valores
acf(x[1001:tamanho])
```




```
# amostra aletoria - pegando os valores 1001, 10011, 10021, ...
amostra=x[seq(1001,tamanho,by=10)]

# calculo da acf para a amostra
acf(amostra)
```




```r
# plot da amostra independente
plot(seq_along(amostra), amostra, type = "l")
```



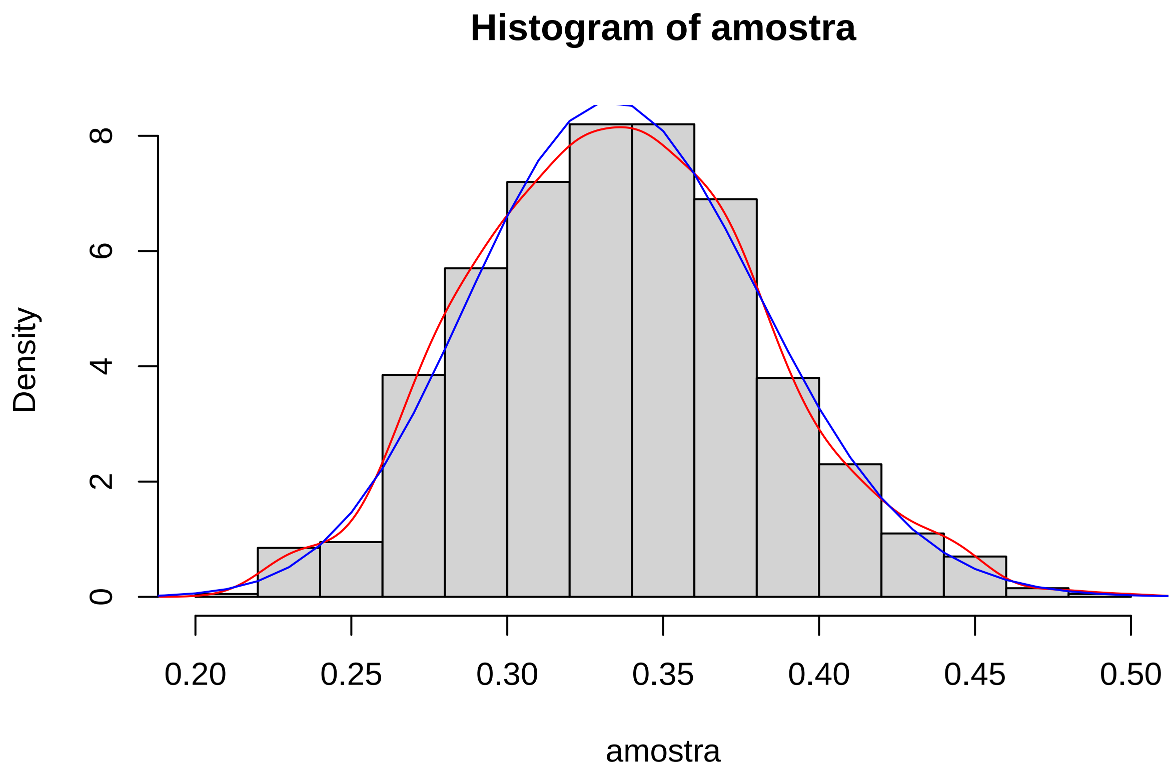

```
# resumo da amostra
summary(amostra)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.2199  0.3021  0.3360  0.3359  0.3665  0.4986
```

```
# comparação com o valor teórico
```

Sabemos que a distribuição a posteriori é uma Beta$(\alpha + \sum x_i, \beta + n - \sum x_i)$, ou seja, Beta$(35, 69)$. A comparação gráfica pode ser feita utilizando o código a seguir.

```
hist(amostra, freq = F)
lines(density(amostra), col="red")
theta_aux <- seq(0.01, 0.99, 0.01)
lines(theta_aux, dbeta(theta_aux, 35, 69), col="blue")
```



Sabemos que a média a posteriori é a estimativa pontual quando utilizamos a função perda quadrática. Para o caso da distribuição Beta(α, β), a média a posteriori é $\alpha/(\alpha+\beta)$, para este exemplo, tem-se $E(\theta|\underline{z}) = 35/(69+35) = 35/104 = 0,3365$. O resultado utilizando a amostra da distribuição a posteriori é

```
mean(amostra)
```

```
## [1] 0.3359213
```

Exercícios para a próxima aula prática

Exércicio 1 Simule uma amostra aleatória de tamanho 30 da distribuição Poisson$(\lambda = 4)$. Considerando a distribuição a priori como Gama$(\alpha = 2, \beta = 2)$, obtenha uma cadeia de Markov de 11 mil valores da distribuição a posteriori. Obtenha uma amostra aleatória de tamanho 1.000 da distribuição a posteriori. Compare as estimativas obtidas pela amostra com os resultados teóricos usando gráficos e medidas resumo.

Exércicio 2 Seja $X_1, ..., X_n$ uma amostra aleatória da distribuição Normal(μ, σ^2), sendo que $\sigma^2 = 1$. Considerando como distribuição a priori para o parâmetro μ a t-Student com $v = 4$ graus de liberdade, obtenha uma cadeia de Markov de 21 mil valores da distribuição a posteriori. Obtenha uma amostra aleatória de tamanho 1.000 da distribuição a posteriori. Compare as estimativas obtidas pela amostra com os resultados teóricos usando gráficos e medidas resumo.