

# MAT02034 - Métodos bayesianos para análise de dados

Métodos de Monte Carlo via Cadeias de Markov

Rodrigo Citton P. dos Reis  
citton.padilha@ufrgs.br

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA  
DEPARTAMENTO DE ESTATÍSTICA

Porto Alegre, 2022

# Introdução

# Introdução

- ▶ Uma séria desvantagem da amostragem padrão de Monte Carlo ou amostragem por importância de Monte Carlo é que a determinação completa da forma funcional da densidade *a posteriori* é necessária para sua implementação.
- ▶ Situações em que as distribuições *a posteriori* são especificadas de forma incompleta ou são especificadas indiretamente não podem ser tratadas.
- ▶ Uma dessas instâncias é onde a distribuição *a posteriori* conjunta do vetor de parâmetros é especificada em termos de várias distribuições **condicionais** e **marginais**, mas não diretamente.
  - ▶ Isso realmente cobre uma gama muito grande de análises bayesianas.

# Introdução

- ▶ Acontece que é de fato possível em tais casos adotar um esquema de amostragem de **Monte Carlo iterativo**, que no **ponto de convergência** garantirá uma geração aleatória da **distribuição (alvo) *a posteriori* conjunta**.
- ▶ Esses procedimentos iterativos de Monte Carlo normalmente geram uma sequência aleatória com a **propriedade de Markov**, tal que essa **cadeia de Markov** é **ergódica** com a **distribuição limite** sendo a distribuição *a posteriori* alvo.
- ▶ Na verdade, existe toda uma classe de tais procedimentos iterativos coletivamente chamados de procedimentos de **Monte Carlo via Cadeias de Markov (MCMC)**.
  - ▶ Diferentes procedimentos desta classe são adequados para diferentes situações.

# Cadeias de Markov para MCMC

# Cadeias de Markov para MCMC

- Uma sequência de variáveis aleatórias  $\{X_n\}_{n \geq 0}$  é uma **cadeia de Markov** se para qualquer  $n$ , dado o valor atual,  $X_n$ , o **passado**  $\{X_j : j \leq n-1\}$  e o **futuro**  $\{X_j : j \geq n+1\}$  são **independentes**. Em outras palavras,

$$\Pr(A \cap B | X_n) = \Pr(A | X_n) \Pr(B | X_n),$$

em que  $A$  e  $B$  são eventos definidos respectivamente em termos do passado e do futuro.

# Cadeias de Markov para MCMC

- ▶ Entre as cadeias de Markov existe uma subclasse que possui ampla aplicabilidade.
- ▶ São as cadeias de Markov com **probabilidades de transição estacionárias** ou homogêneas no tempo, o que significa que a distribuição de probabilidade de  $X_{n+1}$  dado  $X_n = x$ , e o passado,  $\{X_j : j \leq n-1\}$  depende apenas de  $x$  e não depende dos valores de  $\{X_j : j \leq n-1\}$  ou  $n$ .

# Cadeias de Markov para MCMC

- ▶ Se o conjunto  $S$  de valores que  $\{X_n\}$  pode assumir, conhecido como **espaço de estados**, é contável, isso se reduz a especificar a **matriz de probabilidades de transição**  $P \equiv ((p_{ij}))$  em que para quaisquer dois valores  $i, j$  em  $S$ ,

$$p_{ij} = \Pr(X_{n+1} = j | X_n = i)$$

é a probabilidade de passar do estado  $i$  para o estado  $j$  em uma unidade de tempo.

- ▶ Para o espaço de estados  $S$  que não é contável, deve-se especificar um **kernel de transição** ou função de transição  $P(x, \cdot)$  em que

$$P(x, A) = \Pr(X_{n+1} \in A | X_n = x)$$

é a probabilidade de passar de  $x$  para  $A$  em uma etapa (passo).



# Cadeias de Markov para MCMC

- ▶ Dada a probabilidade de transição e a distribuição de probabilidade do valor inicial  $X_0$ , pode-se construir a distribuição de probabilidade conjunta de  $\{X_j : 0 \leq j \leq n\}$  para qualquer  $n$  finito.
  - ▶ Estas são conhecidas como as **distribuições finito-dimensionais**.
- ▶ Por exemplo, no caso do espaço de estados contável

$$\begin{aligned} & \Pr(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n) \\ &= \Pr(X_n = i_n | X_0 = i_0, \dots, X_{n-1} = i_{n-1}) \times \Pr(X_0 = i_0, \dots, X_{n-1} = i_{n-1}) \\ &= p_{i_{n-1}i_n} \Pr(X_0 = i_0, \dots, X_{n-1} = i_{n-1}) \\ &= \Pr(X_0 = i_0) p_{i_0i_1} p_{i_1i_2} \dots p_{i_{n-1}i_n}. \end{aligned}$$

# Cadeias de Markov para MCMC

- ▶ Uma distribuição de probabilidade  $\pi$  é chamada **estacionária** ou **invariante** para uma probabilidade de transição  $P$  ou a cadeia de Markov associada  $\{X_n\}$  se for o caso de quando a distribuição de probabilidade de  $X_0$  é  $\pi$  então o mesmo é verdadeiro para  $X_n$  para todo  $n > 1$ .
- ▶ No caso do espaço de estados contável uma distribuição de probabilidade  $\pi = \{\pi_i : i \in S\}$  é **estacionária** para uma matriz de probabilidade de transição  $P$  se para cada  $j$  em  $S$ ,

$$\begin{aligned}\Pr(X_1 = j) &= \sum_i \Pr(X_1 = j | X_0 = i) \Pr(X_0 = i) \\ &= \sum_i \pi_i p_{ij} = \Pr(X_0 = j) = \pi_j.\end{aligned}$$

# Cadeias de Markov para MCMC

Da mesma forma, se  $S$  é contínuo, uma distribuição de probabilidade  $\pi$  com densidade  $p(x)$  é estacionária para o kernel de transição  $P(\cdot, \cdot)$  se

$$\pi(A) = \int_A p(x) dx = \int_S P(x, A) p(x) dx,$$

para  $A \subset S$ .

# Cadeias de Markov para MCMC

- ▶ Uma cadeia de Markov  $\{X_n\}$  com um espaço de estados contável  $S$  e matriz de probabilidade de transição  $P \equiv ((p_{ij}))$  é dita **irredutível** se para quaisquer dois estados  $i$  e  $j$  a probabilidade da cadeia de Markov visitar  $j$  começando em  $i$  for positiva, ou seja, para algum  $n \geq 1$ ,

$$p_{ij}^{(n)} \equiv \Pr(X_n = j | X_0 = i) > 0.$$

- ▶ Uma noção semelhante de irredutibilidade, conhecida como irredutibilidade de Harris ou Doeblin, também existe para o caso geral do espaço de estados.

# Cadeias de Markov para MCMC

## Teorema (Lei dos grandes números para cadeias de Markov)

Seja  $\{X_n\}_{n \geq 0}$  uma cadeia de Markov com um espaço de estado contável  $S$  e uma matriz de probabilidade de transição  $P$ . Além disso, suponha que seja irredutível e tenha uma distribuição de probabilidade estacionária  $\pi = (\pi_i; i \in S)$ . Então, para qualquer função limitada  $h : S \rightarrow \mathbb{R}$  e para qualquer distribuição inicial de  $X_0$

$$\frac{1}{n} \sum_{i=0}^{n-1} h(X_i) \rightarrow \sum_j h(j) \pi_j = \mathbb{E}_\pi[h(X_j)] \quad (1)$$

em probabilidade conforme  $n \rightarrow \infty$ . (Às vezes é chamado de **Teorema Ergódico**.)

# Cadeias de Markov para MCMC

- ▶ Um resultado semelhante é válido quando o espaço de estados  $S$  não é contável.
- ▶ O valor limite será a integral de  $h$  em relação à distribuição estacionária  $\pi$ .
  - ▶ Uma condição suficiente para a validade deste é que a cadeia de Markov  $\{X_n\}$  seja Harris irreduzível e possua uma distribuição estacionária  $\pi$ .

## Cadeias de Markov para MCMC

- ▶ Para ver como isso é útil para nós, considere o seguinte: dada uma distribuição de probabilidade  $\pi$  em um conjunto  $S$  e uma função  $h$  em  $S$ , suponha que se deseja calcular a “integral de  $h$  em relação a  $\pi$ ”, que se reduz a  $\sum_j h(j)\pi_j$  no caso de  $S$  contável.
- ▶ Procure uma cadeia de Markov irreduzível  $\{X_n\}$  com espaço de estados  $S$  e distribuição estacionária  $\pi$ . Então, a partir de algum valor inicial  $X_0$ , “corra” (uma trajetória d) a cadeia de Markov  $\{X_j\}$  por um período de tempo, digamos  $0, 1, 2, \dots, n-1$  e considere como uma estimativa

$$\mu_n = \frac{1}{n} \sum_{j=0}^{n-1} h(X_j).$$

- ▶ Pela Lei dos grandes números para cadeias de Markov (LGNM), esta estimativa  $\mu_n$  será próxima de  $\sum_j h(j)\pi_j$  para  $n$  grande.

# Cadeias de Markov para MCMC

- ▶ Esta técnica é chamada de **Monte Carlo via cadeia de Markov (MCMC)**.
- ▶ Por exemplo, se alguém está interessado em  $\pi(A) \equiv \sum_{j \in A} \pi_j$  algum  $A \subset S$  então pela LGNCM isso se reduz a

$$\pi_n(A) \equiv \frac{1}{n} \sum_{j=0}^{n-1} I_A(X_j) \rightarrow \pi(A)$$

em probabilidade conforme  $n \rightarrow \infty$ , em que  $I_A(X_j) = 1$  se  $X_j \in A$  e 0 caso contrário.



## Cadeias de Markov para MCMC

- ▶ Uma cadeia de Markov irreduzível  $\{X_n\}$  com um espaço de estados contável  $S$  é chamada **aperiódica** se para algum  $i \in S$  o máximo divisor comum,  $\{n : p_{ii}^{(n)} > 0\} = 1$  ( $p_{ii}^{(n)}$  é a probabilidade de retornar ao estado  $i$  em  $n$  passos da cadeia).
- ▶ Então, além do LGNCM, vale o seguinte resultado sobre a convergência de  $\Pr(X_n = j)$

$$\sum_j |\Pr(X_n = j) - \pi_j| \rightarrow 0. \quad (2)$$

conforme  $n \rightarrow \infty$ , para qualquer distribuição inicial de  $X_0$ .

- ▶ Em outras palavras, para  $n$  grande a distribuição de probabilidade de  $X_n$  estará próxima de  $\pi$ .
- ▶ Existe um resultado semelhante para o caso do espaço de estado geral.

# Cadeias de Markov para MCMC

- ▶ Isso sugere que, em vez de fazer uma trajetória de comprimento  $n$ , pode-se fazer  $N$  trajetórias independentes, cada uma de comprimento  $m$ , de modo que  $n = Nm$  e, em seguida, da  $i$ -ésima trajetória, use apenas a  $m$ -ésima observação, digamos,  $X_{m,i}$  e considere a estimativa

$$\tilde{\mu}_{N,m} \equiv \frac{1}{N} \sum_{i=1}^N h(X_m, i).$$

- ▶ Outras variações também existem. Algumas das cadeias de Markov especiais usadas nos MCMC são discutidas nas próximas duas seções.

# Algoritmo Metropolis-Hastings

# Algoritmo Metropolis-Hastings

- ▶ Nesta seção, discutimos um método MCMC muito geral com amplas aplicações.
- ▶ Logo ficará claro por que essa importante descoberta levou a um progresso considerável na inferência baseada em simulação, particularmente na análise bayesiana.
- ▶ A ideia aqui não é simular diretamente a partir de uma determinada densidade alvo (que pode ser computacionalmente muito difícil), mas sim simular uma cadeia de Markov fácil **que tenha essa densidade alvo como a densidade de sua distribuição estacionária**.

# Algoritmo Metropolis-Hastings

- ▶ Seja  $S$  um conjunto finito ou contável.
- ▶ Seja  $\pi$  uma distribuição de probabilidade em  $S$ .
  - ▶ Chamaremos  $\pi$  de **distribuição alvo**.
- ▶ Seja  $Q \equiv ((q_{ij}))$  uma matriz de probabilidade de transição tal que para cada  $i$ , seja computacionalmente fácil gerar uma amostra da distribuição  $\{q_{ij} : j \in S\}$ .

# Algoritmo Metropolis-Hastings

Vamos gerar uma cadeia de Markov  $\{X_n\}$  como segue.

1. Se  $X_n = i$ , primeiro amostre da distribuição  $\{q_{ij} : j \in S\}$  e denote essa observação  $Y_n$ .
2. Em seguida, escolha  $X_{n+1}$  dos dois valores  $X_n$  e  $Y_n$  de acordo com

$$\begin{aligned}\Pr(X_{n+1} = Y_n | X_n, Y_n) &= \rho(X_n, Y_n) \\ \Pr(X_{n+1} = X_n | X_n, Y_n) &= 1 - \rho(X_n, Y_n),\end{aligned}$$

em que a “**probabilidade de aceitação**”  $\rho(\cdot, \cdot)$  é dada por

$$\rho(i, j) = \min \left\{ \frac{\pi_j}{\pi_i} \frac{q_{ji}}{q_{ij}}, 1 \right\}$$

para todo  $(i, j)$  tal que  $\pi_i q_{ij} > 0$ .

# Algoritmo Metropolis-Hastings

- Observe que  $\{X_n\}$  é uma cadeia de Markov com matriz de probabilidade de transição  $P = ((p_{ij}))$  dada por

$$p_{ij} = \begin{cases} q_{ij}\rho_{ij}, & j \neq i, \\ 1 - \sum_{k \neq i} p_{ik}, & j = i. \end{cases}$$

- $Q$  é chamada de “**probabilidade de transição proposta**” e  $\rho$  de “probabilidade de aceitação” .

# Algoritmo Metropolis-Hastings

- Uma característica significativa deste mecanismo de transição  $P$  é que  $P$  e  $\pi$  satisfazem

$$\pi_i p_{ij} = \pi_j p_{ji} \quad \text{para todo } i, j.$$

- Isso implica que para qualquer  $j$

$$\sum_i \pi_i p_{ij} = \pi_j \sum_i p_{ji} = \pi_j,$$

ou,  $\pi$  (a distribuição alvo) é uma distribuição de probabilidade estacionária para  $P$ .



# Algoritmo Metropolis-Hastings

- ▶ Agora suponha que  $S$  é irredutível em relação a  $Q$  e  $\pi_i > 0$  para todo  $i$  em  $S$ .
- ▶ Pode-se então mostrar que  $P$  é irredutível e, como tem uma distribuição estacionária  $\pi$ , a LGNCM está disponível.
- ▶ Este algoritmo é, portanto, muito flexível e útil.
- ▶ A escolha de  $Q$  está sujeita apenas à condição de que  $S$  seja irredutível em relação a  $Q$ .
- ▶ Claramente, não há perda de generalidade supor que  $\pi_i > 0$  para todo  $i$  em  $S$ .

# Algoritmo Metropolis-Hastings

- Uma condição suficiente para a aperiodicidade de  $P$  é que  $p_{ii} > 0$  para algum  $i$  ou equivalente

$$\sum_{j \neq i} q_{ij} \rho_{ij} < 1.$$

- Uma condição suficiente para isso é que exista um par  $(i, j)$  tal que  $\pi_i q_{ij} > 0$  e  $\pi_j q_{ji} < \pi_i q_{ij}$ .
- Lembre-se de que, se  $P$  for aperiódica, tanto o LGNCM (1) quanto o resultado de (2) serão válidos.

# Algoritmo Metropolis-Hastings

Se  $S$  não é finito ou contável, mas é um contínuo e a distribuição alvo  $\pi(\cdot)$  tem uma densidade  $g(\cdot)$ , então procede-se da seguinte forma:

1. Seja  $Q$  uma função de transição tal que para cada  $x$ ,  $Q(x, \cdot)$  tem uma densidade  $q(x, y)$ .
2. Em seguida, proceda como no caso discreto, mas defina a **“probabilidade de aceitação”**  $\rho(x, y)$  como

$$\rho(x, y) = \min \left\{ \frac{g(y)q(y, x)}{g(x)q(x, y)}, 1 \right\}$$

para todo  $(x, y)$  tal que  $g(x)q(x, y) > 0$ .

# Algoritmo Metropolis-Hastings

- ▶ Uma característica particularmente útil do algoritmo acima é que é suficiente conhecer  $g(\cdot)$  a menos de uma constante multiplicativa, pois na definição da “probabilidade de aceitação”  $\rho(\cdot, \cdot)$  apenas as razões  $g(y)/g(x)$  precisam ser calculados (note que a constante de normalização é cancelada na razão  $g(y)/g(x)$ ).
- ▶ Isso nos assegura que em aplicações bayesianas não é necessário ter a constante de normalização da densidade *a posteriori* disponível para cálculo das quantidades *a posteriori* de interesse.

## Algoritmo M-H: exemplo

- ▶ Suponha que  $X$  é o número de defeituosos na produção diária de um produto.
- ▶ Considere  $(X|Y, \theta) \sim \text{binomial}(Y, \theta)$ , em que  $Y$ , a produção de um dia, é uma **variável aleatória** com uma distribuição de Poisson com média conhecida  $\lambda$ , e  $\theta$  é a probabilidade de que qualquer produto seja defeituoso.
- ▶ A distribuição *a priori* é tal que  $(\theta|Y = y) \sim \text{Beta}(\alpha, \gamma)$ , com  $\alpha$  e  $\gamma$  conhecidos independentes de  $Y$ .

## Algoritmo M-H: exemplo

- Observe que  $X|\theta \sim \text{Poisson}(\lambda\theta)$ . Em seguida,  $\theta \sim \text{Beta}(\alpha, \gamma)$ . Portanto,

$$g(\theta|X = x) \propto \exp(-\lambda\theta)\theta^{x+\alpha-1}(1-\theta)^{\gamma-1}, \quad 0 \leq \theta \leq 1.$$

- A única dificuldade é que esta não é uma **distribuição padrão** e, portanto, as quantidades *a posteriori* não podem ser obtidas de forma fechada.

## Algoritmo M-H: exemplo

- ▶ Mesmo que o método da amostragem por rejeição (SIR também) possa ser empregado aqui, nós gostaríamos de usar este exemplo para ilustrar o algoritmo M-H.
- ▶ A cadeia de Markov necessária é gerada tomando a densidade de transição  $q(\theta, \theta^c) = q(\theta^c | \theta) = q(\theta^c)$ , independente de  $\theta$ . Uma boa escolha para  $q(\cdot)$  é a densidade de  $\text{Beta}(x + \alpha, \gamma)$ .

## Algoritmo M-H: exemplo

- ▶ Em nosso exemplo, suponha  $X = 1$ ,  $\alpha = 1$ ,  $\gamma = 49$  e  $\lambda = 100$ .
- ▶ Ainda, a probabilidade de aceitação é

$$\rho(\theta, \theta^c) = \min \left\{ \frac{g(\theta^c)q(\theta)}{g(\theta)q(\theta^c)}, 1 \right\} = \min \{ \exp[-\lambda(\theta^c - \theta)], 1 \}.$$



## Algoritmo M-H: exemplo

Assim, os passos envolvidos neste **algoritmo M-H “independente”** são os seguintes. Comece em  $t = 0$  com um valor  $\theta_0$  no suporte da distribuição alvo; neste caso,  $0 < \theta_0 < 1$ . Dado  $\theta_t$ , gere o próximo valor na cadeia conforme indicado abaixo.

- a. Gere  $\theta_t^c$  de  $Beta(x + \alpha, \gamma)$ .
- b. Seja

$$\theta_{t+1} = \begin{cases} \theta_t^c & \text{com probabilidade } \rho_t, \\ \theta_t & \text{caso contrário.} \end{cases}$$

em que  $\rho_t = \min \{ \exp[-\lambda(\theta_t^c - \theta_t)], 1 \}$ .

- c. Faça  $t = t + 1$  e volte ao passo (a).

Rode esta cadeia até  $t = n$ , um inteiro grande adequadamente escolhido.

# Algoritmo M-H: exemplo

```
# Configuração
```

```
x <- 1
```

```
alpha <- 1
```

```
gamma <- 49
```

```
lambda <- 100
```

```
n <- 1000
```

```
theta <- vector(length = n)
```

## Algoritmo M-H: exemplo

```
# Algoritmo M-H

theta[1] <- 0.3 # theta_0 = 0.3

for (t in 1:(n-1)){

  theta_cand <- rbeta(
    n = 1, shape1 = x + alpha, shape2 = gamma)

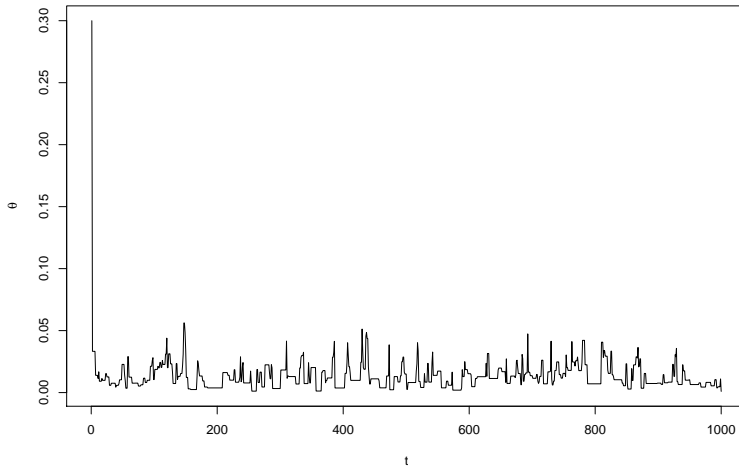
  rho <- min(exp(-lambda * (theta_cand - theta[t])),
             1)

  theta[(t+1)] <- sample(c(theta_cand, theta[t]),
                        size = 1,
                        prob = c(rho, 1 - rho))

}
```

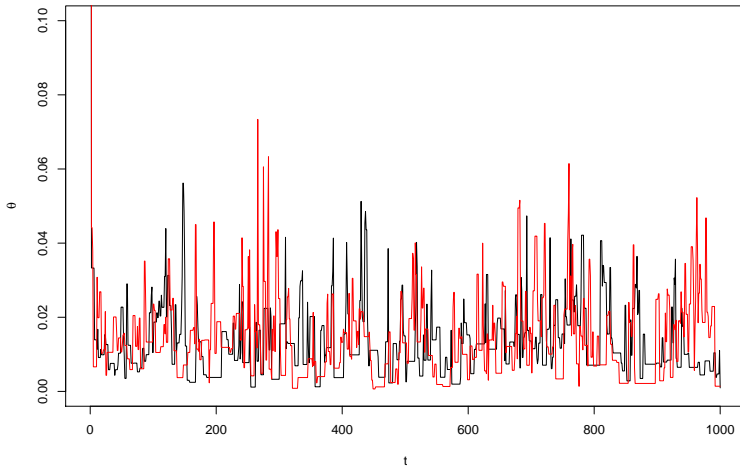
# Algoritmo M-H: exemplo

Trajetória da cadeia (1.000 passos)



# Algoritmo M-H: exemplo

Duas trajetórias da cadeia (1.000 passos)

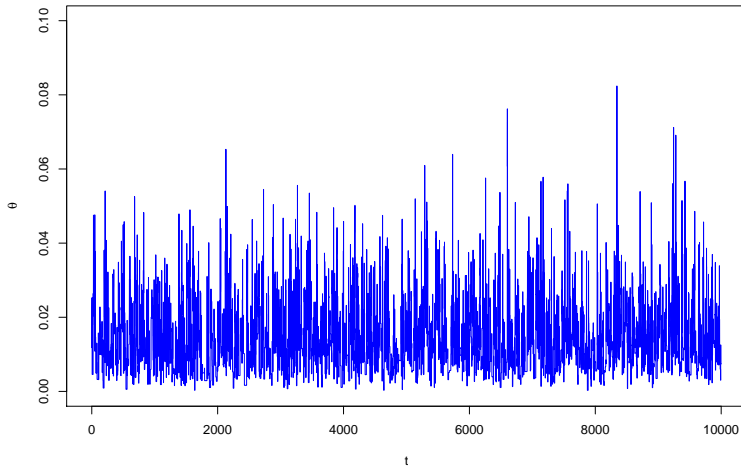


## Algoritmo M-H: exemplo

- ▶ Para que a escolha do valor inicial não influencie na amostra da distribuição alvo, e para “garantirmos” que estamos amostrando da distribuição alvo (dist. estacionária/limite), uma estratégia de amostragem consiste em deixar a cadeia “correr” por um período inicial (**burn-in period**, ou **warm-up**), e só após este período iniciar a amostragem das realizações de  $\theta$ .
- ▶ Escolha de valores para o **burn-in** podem variar (1.000, 10.000, etc.).
  - ▶ Esta escolha está relacionada com a convergência da cadeia. Logo, está proximalmente relacionada com a escolha da distribuição proposta  $q(\cdot, \cdot)$ .

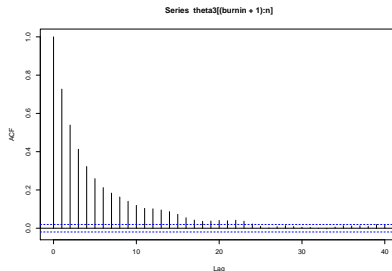
# Algoritmo M-H: exemplo

Trajetória da cadeia (descartando 10.000 primeiras observações)



# Algoritmo M-H: exemplo

- Note ainda, que estas observações são correlacionadas.

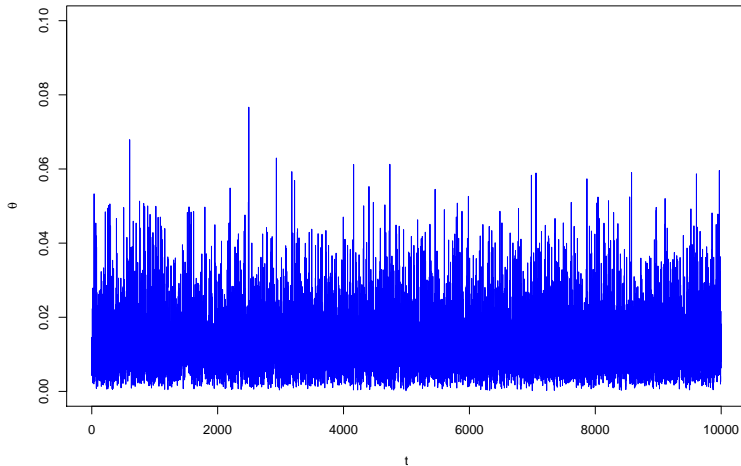


- Uma estratégia consiste em amostrar da “distribuição limite” a cada  $k$  observações.
  - Isto também permite “correr” uma trajetória maior da cadeia, aumentando a eficiência do método em explorar o espaço paramétrico de  $\theta$ .

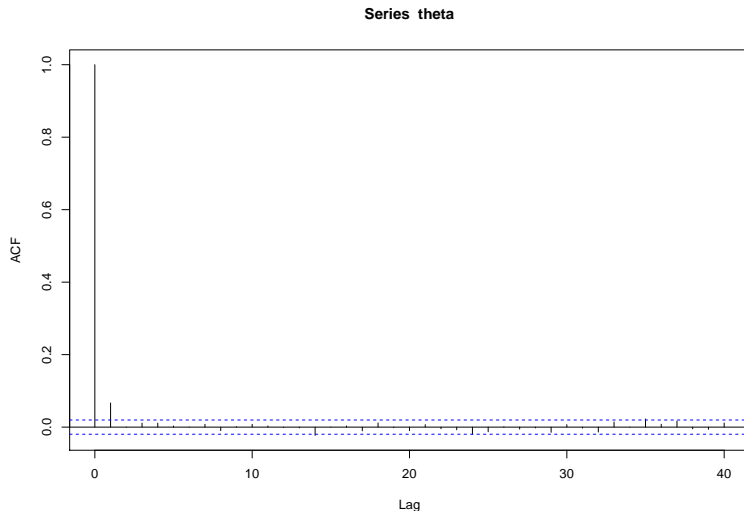


# Algoritmo M-H: exemplo

Trajetória da cadeia (descartando 10.000 primeiras observações, saltando de 10 em 10)

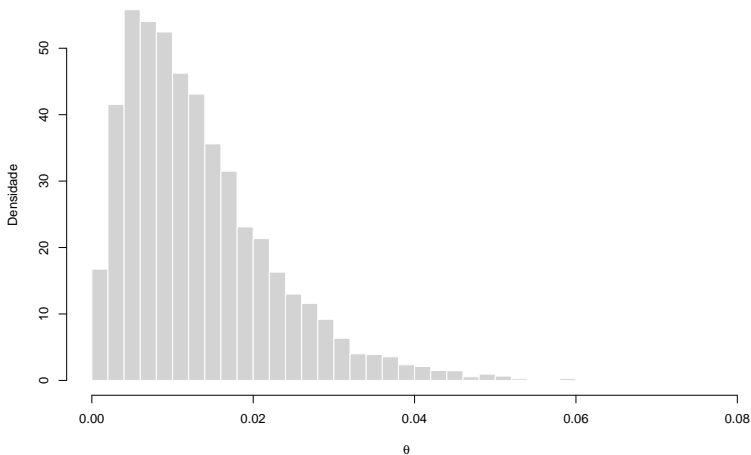


# Algoritmo M-H: exemplo



# Algoritmo M-H: exemplo

Amostra da dist. a posteriori



# Amostrador de Gibbs

# Amostrador de Gibbs

- ▶ A maioria dos novos problemas que a inferência bayesiana é solicitada a resolver são de alta dimensão.
  - ▶ Aplicações em áreas como **microarrays (DNA)** e **processamento de imagens** são alguns exemplos.
- ▶ A análise bayesiana de tais problemas invariavelmente envolve **distribuições alvo** (*a posteriori*) que são distribuições multivariadas de alta dimensão.

# Amostrador de Gibbs

- ▶ O **amostrador de Gibbs** é uma técnica especialmente adequada para gerar uma cadeia de Markov aperiódica irreduzível que tem como distribuição estacionária uma distribuição alvo em um espaço de alta dimensão, mas com alguma estrutura especial.
- ▶ O aspecto mais interessante desta técnica é que para “correr” esta cadeia de Markov, **basta gerar observações a partir de distribuições univariadas**.

# Amostrador de Gibbs

O amostrador de Gibbs no contexto de uma distribuição de probabilidade bivariada pode ser descrito como segue.

- ▶ Seja  $\pi$  uma distribuição de probabilidade alvo de um vetor aleatório bivariado  $(X, Y)$ .
- ▶ Para cada  $x$ , seja  $P(x, \cdot)$  a distribuição de probabilidade condicional de  $Y$  dado  $X = x$ .
  - ▶ Da mesma forma, seja  $Q(y, \cdot)$  a distribuição de probabilidade condicional de  $X$  dado  $Y = y$ .
- ▶ Observe que para cada  $x$ ,  $P(x, \cdot)$  é uma distribuição univariada, e para cada  $y$ ,  $Q(y, \cdot)$  também é uma distribuição univariada.

# Amostrador de Gibbs

Agora gere uma **cadeia de Markov bivariada**  $Z_n = (X_n, Y_n)$  como segue:

1. Comece com algum  $X_0 = x_0$ .
2. Gere uma observação  $Y_0$  da distribuição  $P(x_0, \cdot)$
3. Então gere uma observação  $X_1$  de  $Q(Y_0, \cdot)$ .
4. Em seguida, gere uma observação  $Y_1$  de  $P(X_1, \cdot)$  e assim por diante.

No passo  $n$  da cadeia, se  $Z_n = (X_n, Y_n)$  for conhecido, então gere  $X_{n+1}$  de  $Q(Y_n, \cdot)$  e  $Y_{n+1}$  de  $P(X_{n+1}, \cdot)$ .



# Amostrador de Gibbs

- ▶ Se  $\pi$  é uma distribuição discreta concentrada em  $\{(x_i, y_j) : 1 \leq i \leq K, 1 \leq j \leq L\}$  e se  $\pi_{ij} = \pi(x_i, y_j)$  então  $P(x_i, y_j) = \pi_{ij}/\pi_{i\cdot}$  e  $Q(y_j, x_i) = \pi_{ij}/\pi_{\cdot j}$ , em que  $\pi_{i\cdot} = \sum_j \pi_{ij}$  e  $\pi_{\cdot j} = \sum_i \pi_{ij}$ .
- ▶ Assim, a matriz de probabilidades de transição  $R = ((r_{(ij),(k\ell)}))$  da cadeia  $\{Z_n\}$  é dada por

$$r_{(ij),(k\ell)} = Q(y_j, x_k)P(x_k, y_\ell) = \frac{\pi_{kj}}{\pi_{\cdot j}} \frac{\pi_{k\ell}}{\pi_{k\cdot}}.$$

# Amostrador de Gibbs

- ▶ Pode-se verificar que esta cadeia é irredutível, aperiódica e tem  $\pi$  como sua distribuição estacionária.
- ▶ Assim, LGNCM (1) e (2) valem neste caso.
- ▶ Assim, para  $n$  grande,  $Z_n$  pode ser visto como uma amostra de uma distribuição próxima a  $\pi$  e pode-se aproximar  $\sum_{i,j} h(i,j)\pi_{ij}$  por  $\sum_{i=1}^n h(X_i, Y_i)/n$ .
  - ▶ Por exemplo, se  $h(X_i, Y_i) = X_i$ , então  $\sum_{i=1}^n h(X_i, Y_i)/n = \sum_{i=1}^n X_i/n$  é uma estimativa de  $\sum_{i,j} h(i,j)\pi_{ij} = \sum_i i \sum_j \pi_{ij} = \sum_i i\pi_{i\cdot} = E_{\pi}(X)$ .

# Amostrador de Gibbs

A extensão multivariada do caso bivariado mencionado acima é muito direta.

- ▶ Suponha que  $\pi$  seja uma distribuição de probabilidade de um vetor aleatório  $k$ -dimensional  $(X_1, X_2, \dots, X_k)$ .
- ▶ Se  $\mathbf{u} = (u_1, u_2, \dots, u_k)$  é qualquer vetor de comprimento  $k$ , seja  $\mathbf{u}_{-i} = (u_1, u_2, \dots, u_{i-1}, u_{i+1}, \dots, u_k)$  seja o vetor de comprimento  $k - 1$  resultante da eliminação do componente  $u_i$ .
- ▶ Seja  $\pi(\cdot | \mathbf{x}_{-i})$  a distribuição condicional univariada de  $X_i$  dado que  $\mathbf{X}_{-i} \equiv (X_1, X_2, \dots, X_{i-1}, X_{i+1}, \dots, X_k) = \mathbf{x}_{-i}$ .
- ▶ Agora começando com algum valor inicial para  $\mathbf{X}_0 = (x_{01}, x_{02}, \dots, x_{0k})$  gere  $\mathbf{X}_1 = (X_{11}, X_{12}, \dots, X_{1k})$  sequencialmente gerando  $X_{11}$  de acordo com a distribuição univariada  $\pi(\cdot | \mathbf{x}_{0_{-1}})$  e, em seguida, gerando  $X_{12}$  de acordo com  $\pi(\cdot | X_{11}, x_{03}, \dots, x_{0k})$  e assim por diante.

# Amostrador de Gibbs

- ▶ A característica mais importante a reconhecer aqui é que todas as distribuições condicionais univariadas,  $X_i | \mathbf{X}_{-i} = \mathbf{x}_{-i}$ , conhecidas como **condicionais completas** devem permitir facilmente a amostragem delas, o que acaba sendo o caso na maioria dos problemas **hierárquicos de Bayes**.
- ▶ Assim, o amostrador de Gibbs é particularmente bem adaptado para cálculos bayesianos com **distribuição *a priori* hierárquicas**.

## Amostrador de Gibbs: exemplo

- Novamente considere o problema de estimar  $\theta$ , a probabilidade de um item produzido ser defeituoso. Relembrando:

$$(X|Y, \theta) \sim \text{bin}(Y, \theta)$$

$$(Y|\lambda) \sim \text{Poi}(\lambda)$$

$$(\theta|Y = y) \sim \text{Beta}(\alpha, \gamma)$$

## Amostrador de Gibbs: exemplo

- ▶ Embora outros métodos já tenham sido utilizados para obter uma amostra da distribuição *a posteriori* com sucesso, o amostrador de Gibbs oferece uma excelente alternativa.
- ▶ Em vez de focar em  $\theta|X$  diretamente, veja-o como um **componente marginal** de  $(Y, \theta|X)$ .
- ▶ A partir da distribuição *a posteriori* conjunta,  $g(Y, \theta|X = x)$ , vamos obter as **distribuições condicionais completas** para  $Y$  e  $\theta$ .
- ▶ Note que,

$$g(Y, \theta|X = x) \propto \frac{1}{(y - x)!} \lambda^y e^{-\lambda} \theta^{\alpha+x-1} (1 - \theta)^{y-x} (1 - \theta)^{\gamma-1}.$$

## Amostrador de Gibbs: exemplo

► Portanto, temos:

$$\begin{aligned} p(y|\theta, X = x) &\propto \frac{1}{(y-x)!} \lambda^y e^{-\lambda} (1-\theta)^{y-x} \\ &= \frac{1}{(y-x)!} \lambda^y \left( \frac{\lambda^{-x}}{\lambda^{-x}} \right) e^{-\lambda} \left( \frac{e^{-\lambda\theta}}{e^{-\lambda\theta}} \right) (1-\theta)^{y-x} \\ &\propto \frac{1}{(y-x)!} e^{-\lambda(1-\theta)} [\lambda(1-\theta)]^{y-x}, \end{aligned}$$

e

$$p(\theta|y, X = x) \propto \theta^{\alpha+x-1} (1-\theta)^{\gamma+y-x-1}.$$

## Amostrador de Gibbs: exemplo

- ▶ Assim, temos que as distribuições condicionais completas são dadas por

$$\begin{aligned}(Y|\theta, X = x) &\sim x + Poi(\lambda(1 - \theta)) \\ (\theta|Y = y, X = x) &\sim Beta(\alpha + x, \gamma + y - x)\end{aligned}$$

- ▶ Note que estas são distribuições padrões e fáceis de simular.



# Amostrador de Gibbs: exemplo

```
# Configuração
```

```
x <- 1
```

```
alpha <- 1
```

```
gamma <- 49
```

```
lambda <- 100
```

```
n <- 1000
```

```
theta <- vector(length = n)
```

```
Y <- vector(length = n)
```

## Amostrador de Gibbs: exemplo

```
# Amostrador de Gibbs

theta[1] <- 0.3 # theta_0 = 0.3

for (t in 1:(n-1)){

  Y[t] <- x + rpois(n = 1,
                    lambda = lambda * (1 - theta[t]))

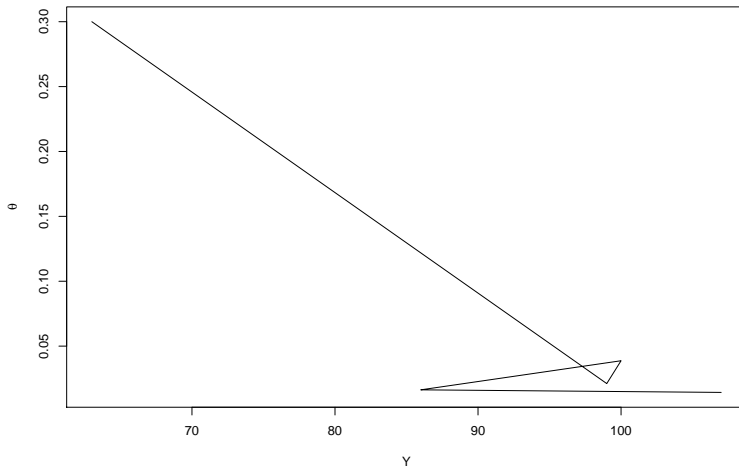
  theta[(t+1)] <- rbeta(n = 1,
                       shape1 = alpha + x,
                       shape2 = gamma + Y[t] - x)

}

Y[n] <- x + rpois(n = 1,
                  lambda = lambda * (1 - theta[n]))
```

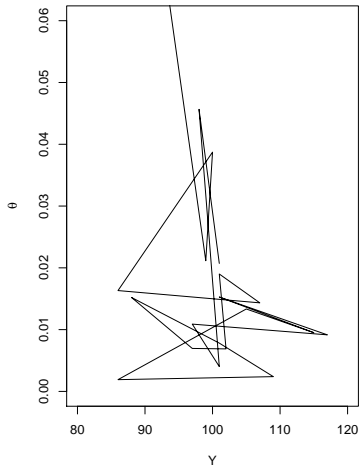
# Amostrador de Gibbs: exemplo

Trajetória da cadeia (5 primeiros passos)

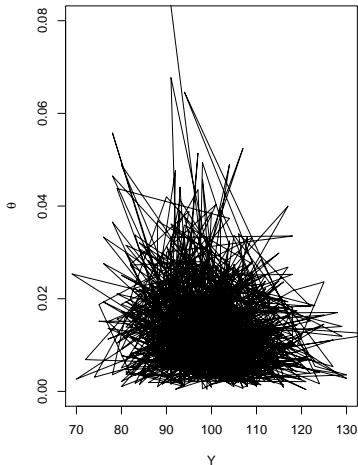


# Amostrador de Gibbs: exemplo

Trajetória da cadeia (20 primeiros passos)

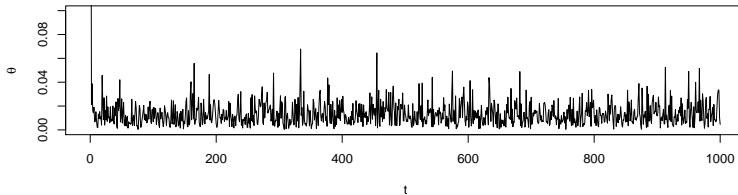


Trajetória da cadeia (1.000 passos)

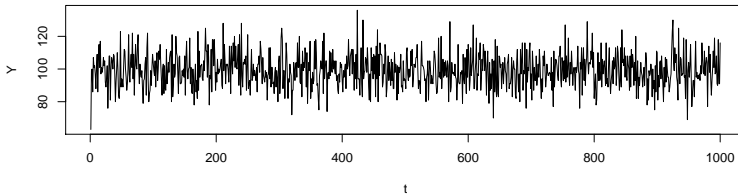


# Amostrador de Gibbs: exemplo

Trajectoria da cadeia (1.000 passos)



Trajectoria da cadeia (1.000 passos)



## Amostrador de Gibbs: exemplo

- Mais uma vez podemos descartar observações gerados no período de *burn-in* e armazenar apenas observações a cada  $k$  passos da cadeia (*thinning*).

```
# Configuração
```

```
n <- 10000
```

```
burnin <- 10000
```

```
k <- 10
```

```
theta <- vector(length = n)
```

```
Y <- vector(length = n)
```

## Amostrador de Gibbs: exemplo

```
# Amostrador de Gibbs

# Período de burn-in

theta_b <- 0.3 # theta_0 = 0.3

Y_b <- x + rpois(n = 1,
               lambda = lambda * (1 - theta_b))

for (t in 1:(burnin-1)){

  theta_b <- rbeta(n = 1,
                 shape1 = alpha + x,
                 shape2 = gamma + Y_b - x)

  Y_b <- x + rpois(n = 1,
                 lambda = lambda * (1 - theta_b))

}
```

## Amostrador de Gibbs: exemplo

```
# Amostrador de Gibbs

# Período pós burn-in

theta_n <- theta_b
Y_n <- Y_b

for (t in 1:(k*n) ){

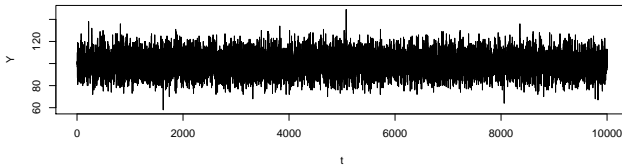
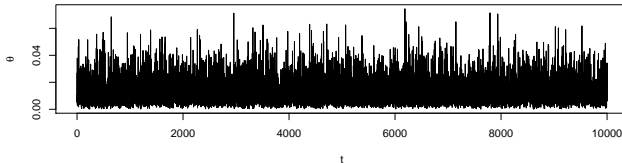
  theta_n <- rbeta(n = 1,
                  shape1 = alpha + x,
                  shape2 = gamma + Y_n - x)

  Y_n <- x + rpois(n = 1,
                  lambda = lambda * (1 - theta_n))

  if (t %% k == 0) {
    theta[(t/k)] <- theta_n
    Y[(t/k)] <- Y_n
  }
}
```

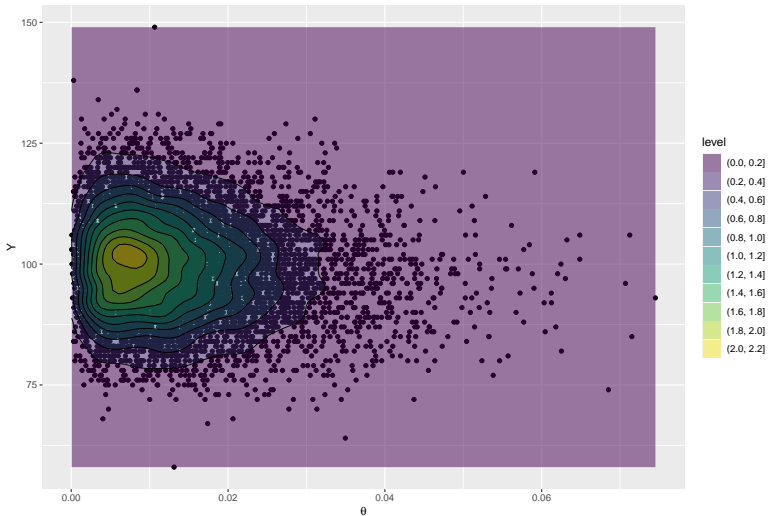


# Amostrador de Gibbs: exemplo

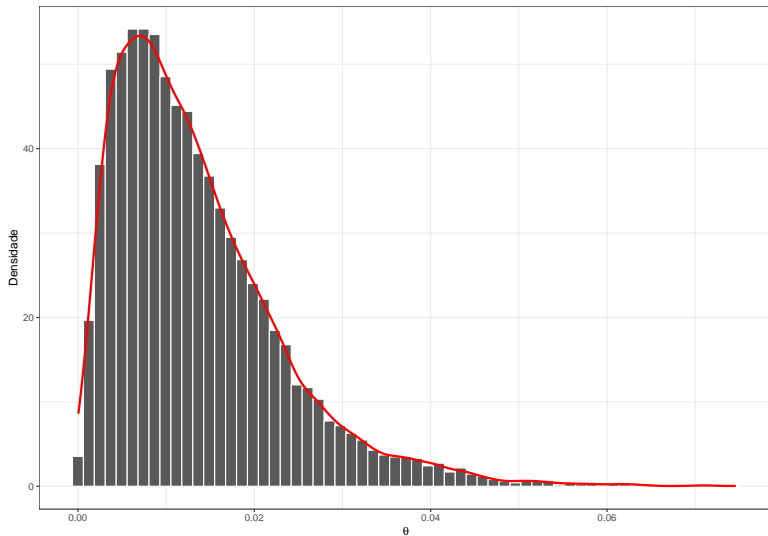


# Amostrador de Gibbs: exemplo

Amostra e densidade (estimada) conjunta da distribuição a posteriori



## Amostrador de Gibbs: exemplo



## Amostrador de Gibbs: exemplo

- ▶ Assim como nos outros métodos de Monte Carlo, uma vez que temos uma amostra da distribuição *a posteriori* podemos realizar inferências para o parâmetro de interesse com base em resumos da distribuição:
  - ▶ média *a posteriori*;
  - ▶ mediana *a posteriori*;
  - ▶ probabilidade *a posteriori* da ocorrência de um certo evento de interesse ( $\theta \geq 0.07$ , por exemplo);
  - ▶ intervalos de credibilidade.

```
# média_a_posteriori_  
mean(theta)
```

```
## [1] 0.01347836
```

```
# mediana_a_posteriori_  
median(theta)
```

```
## [1] 0.011396
```

```
# probabilidade_a_posteriori_ theta >= 0.07  
mean(theta >= 0.07)
```

```
## [1] 4e-04
```

## Amostrador de Gibbs: exemplo

```
# intervalos de credibilidade (quantílico)  
quantile(theta, c(0.025, 0.975))
```

```
##           2.5%           97.5%  
## 0.001723109 0.038419274
```

```
# intervalos de credibilidade (HPD)  
library(HDInterval)  
hdi(theta, credMass = 0.95)
```

```
##           lower           upper  
## 0.0004380295 0.0322941573  
## attr("credMass")  
## [1] 0.95
```

# Avaliando a convergência das saídas do MCMC

# Avaliando a convergência das saídas do MCMC

- ▶ Como já vimos, abordagens baseadas em amostragem de Monte Carlo para inferência fazem uso de **teoremas limite**, como a lei dos grandes números, para justificar sua validade.
- ▶ Quando adicionamos mais uma dimensão a esta amostragem e adotamos esquemas MCMC, são necessários teoremas limite mais fortes.
  - ▶ Teoremas ergódicos para cadeias de Markov, como os dados nas equações (1) e (2) são esses resultados úteis.
- ▶ Este procedimento necessariamente depende de esperar até que a cadeia de Markov **convirja** para a distribuição invariante alvo, e então amostrar desta distribuição.

# Avaliando a convergência das saídas do MCMC

- ▶ No entanto, como saber que a cadeia de fato convergiu para a distribuição invariante?
- ▶ Nesta aula, vamos considerar algumas estratégias para avaliação da convergência das cadeias geradas pelos métodos MCMC.
- ▶ Também veremos um exemplo de aparente convergência, mas que a distribuição *a posteriori* é imprópria.

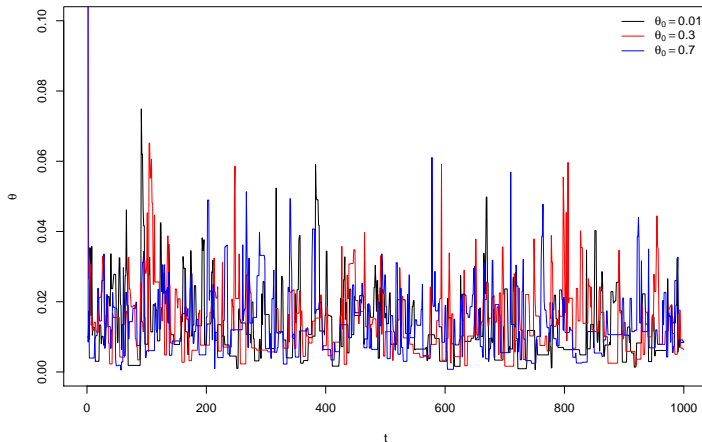


# Gráficos de traço ou histórico

- ▶ Os **gráficos de histórico (traço da cadeia)** são um dos métodos mais antigos de avaliar qualitativamente o desempenho da amostragem MCMC.
- ▶ O **número de iteração** MCMC está no **eixo x** e o **valor do parâmetro** gerado em cada iteração está no **eixo y**.
- ▶ Os valores sucessivos são unidos por uma linha.
- ▶ Quando mais de uma cadeia foi rodada, as linhas de todas as cadeias são apresentadas em diferentes cores no mesmo gráfico.

## Gráficos de traço ou histórico

- ▶ De volta ao exemplo ( $\theta$ , prob. de defeituoso).



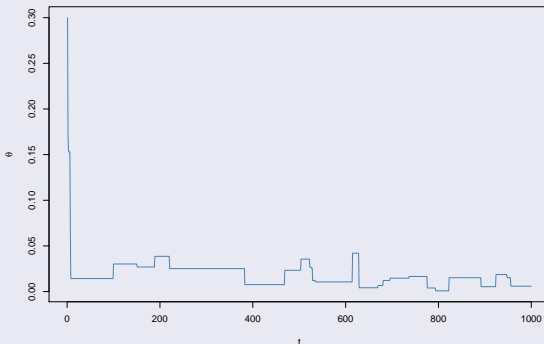
## Gráficos de traço ou histórico

- ▶ Apesar do fato de termos iniciado as três cadeias a partir de valores iniciais (bastante) diferentes, elas imediatamente passam a gerar valores da mesma faixa do espaço paramétricos.
- ▶ Além disso, elas se parecem com ruído branco - apenas rabiscos aleatórios nesse intervalo, sem nenhum padrão consistente.
  - ▶ Estas são características da rápida convergência MCMC (*good mixing*).

# Gráficos de traço ou histórico

## Exemplo de poor mixing

- Considere mais uma vez o exemplo de estimar a probabilidade de defeituosos,  $\theta$ , utilizando o mesmo modelo, com o algoritmo M-H, porém com a distribuição proposta  $Beta(1, 1)$ .



# Gráficos de autocorrelação

- ▶ Mencionamos que geralmente há dependência nas amostras produzidas por uma cadeia de Markov.
- ▶ A **autocorrelação** é uma medida quantitativa dessa dependência.
- ▶ Os valores de autocorrelação devem estar entre -1 (significando correlação negativa perfeita) e 1.

# Gráficos de autocorrelação

- ▶ A **autocorrelação de defasagem (*lag*) 1** na saída do MCMC é a correlação entre amostras da mesma cadeia separadas por 1 iteração.
  - ▶ Em nosso exemplo, podemos chamar os valores extraídos de iterações sucessivas de  $\theta^{(1)}$ ,  $\theta^{(2)}$ ,  $\theta^{(3)}$  e assim por diante.
  - ▶ Então, para calcular a autocorrelação de *lag* 1, emparelhamos  $\theta^{(1)}$  com  $\theta^{(2)}$ ,  $\theta^{(2)}$  com  $\theta^{(3)}$ ,  $\theta^{(3)}$  com  $\theta^{(4)}$ , etc.
- ▶ Da mesma forma, a autocorrelação de *lag*  $k$  é a correlação entre amostras extraídas com  $k$  iterações de separação.

# Gráficos de autocorrelação

- ▶ Em um gráfico de autocorrelação, *as defasagens estão no eixo x* e a altura de cada barra representa a magnitude da autocorrelação nessa defasagem.
- ▶ A primeira barra (para *lag 0*) sempre tem uma altura de 1; ela fornece uma escala visual com a qual comparar as alturas das barras restantes.
- ▶ Normalmente, na saída do MCMC, a autocorrelação de *lag 1* é positiva e a autocorrelação diminui à medida que a defasagem aumenta até atingir uma defasagem limite além do qual é essencialmente 0.

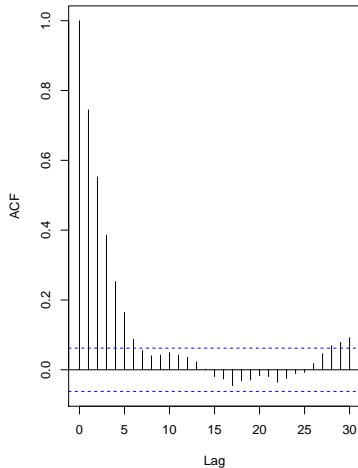
# Gráficos de autocorrelação

Uma cadeia de Markov na qual a **autocorrelação é grande na defasagem 1** e **decai lentamente à medida que a defasagem aumenta** é dita ser *mix slowly*, e convergirá lentamente em todos os três sentidos:

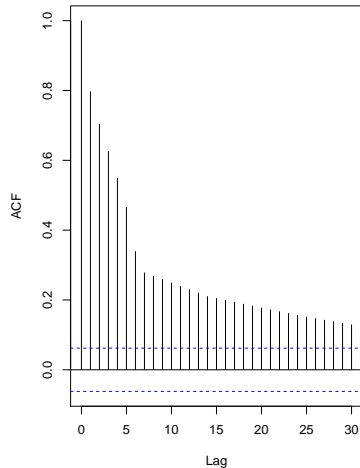
1. levará muito tempo para encontrar sua distribuição estacionária;
2. uma vez na distribuição estacionária, serão necessárias muitas iterações para explorar todo o suporte da distribuição;
3. e um número muito grande de iterações será necessário para obter uma estimativa útil e precisa das características da distribuição *a posteriori*.



# Gráficos de autocorrelação

Proposta Beta( $x + \alpha$ ,  $\gamma$ )

Proposta Beta(1, 1)

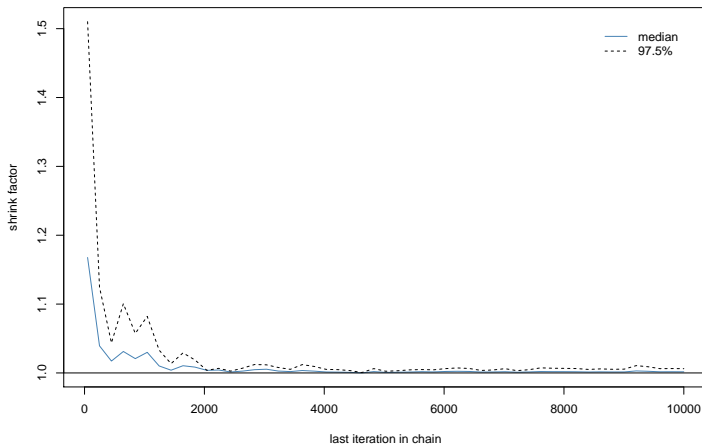


## Diagnóstico Brooks, Gelman e Rubin

- ▶ Além das verificações visuais de convergência, existem também verificações numéricas.
- ▶ Uma verificação numérica popular é uma medida de quanta variância existe entre as cadeias em relação a quanta variância existe dentro das cadeias.
- ▶ A ideia é que, se todas as cadeias se estabeleceram em uma amostragem representativa da distribuição invariante, a diferença média entre as cadeias deve ser a mesma que a diferença média (entre etapas) dentro das cadeias.
- ▶ Mas, se uma ou mais cadeias ficarem “presas” (em um subconjunto do espaço paramétrico), aumentará a variação entre as cadeias em relação à variação dentro da cadeia.

# Diagnóstico Brooks, Gelman e Rubin

- ▶ A figura a seguir mostra o gráfico dessa medida.



# Diagnóstico Brooks, Gelman e Rubin

- ▶ Podemos ver que durante o período de inicial, a medida excede muito 1.0.
- ▶ Após o período de inicial, a medida rapidamente se aproxima muito de 1.0.
  - ▶ Poderia ser utilizado para especificar o período de *burn-in*?

# Diagnóstico Brooks, Gelman e Rubin

- ▶ A medida numérica específica é chamada de estatística **Gelman-Rubin**<sup>1</sup>, ou estatística **Brooks-Gelman-Rubin**<sup>2</sup>, ou o “**fator de redução de escala potencial**”, ou simplesmente o “**fator de encolhimento**” (*shrink factor*).

---

<sup>1</sup>Andrew Gelman. Donald B. Rubin. Inference from Iterative Simulation Using Multiple Sequences. *Statist. Sci.* 7 (4) 457 - 472, November, 1992.  
<https://doi.org/10.1214/ss/1177011136>

<sup>2</sup>Stephen P. Brooks & Andrew Gelman (1998) General Methods for Monitoring Convergence of Iterative Simulations, *Journal of Computational and Graphical Statistics*, 7:4, 434-455, DOI: 10.1080/10618600.1998.10474787

## Diagnóstico Brooks, Gelman e Rubin

- ▶ Intuitivamente, seu valor é 1.0 se as cadeias forem totalmente convergentes, mas seu valor é maior que 1.0 se houver cadeias órfãs ou presas.
- ▶ Como uma heurística, se a estatística Gelman-Rubin for maior que 1.1, devemos nos preocupar que talvez as cadeias não tenham convergido adequadamente.
- ▶ A definição exata da estatística Gelman-Rubin envolve muitos detalhes que não são essenciais para os propósitos do nosso curso.

## Um exemplo (para alertar)

- ▶ Há uma situação importante, no entanto, onde a amostragem MCMC pode levar a inferências absurdas.
- ▶ É aqui que se recorre à amostragem MCMC sem perceber que a distribuição *a posteriori* alvo **não é uma distribuição de probabilidade**, mas imprópria.

# Um exemplo (para alertar)

## Probabilidade de defeituosos ( $\theta$ )

- ▶ Lembre-se que, neste problema,  $(X|Y, \theta) \sim \text{bin}(Y, \theta)$ , onde  $(Y|\lambda) \sim \text{Poi}(\lambda)$ . Anteriormente, trabalhamos com uma média conhecida  $\lambda$ , mas vamos ver agora se é possível lidar com esse problema com  $\lambda$  desconhecida. Como  $Y$  não é observado e apenas  $X$  é observado, existe um problema de “identificação” aqui, como pode ser visto observando que  $(X|\theta) \sim \text{Poi}(\lambda\theta)$ . Já temos a distribuição *a priori*  $\text{Beta}(\alpha, \gamma)$  em  $\theta$ . Suponha que  $0 < \alpha \leq 1$ . Considere uma distribuição *a priori* independente em  $\lambda$  de acordo com  $g(\lambda) \propto I(\lambda > 0)$ . Então,

$$g(\lambda, \theta|x) \propto \exp(-\lambda\theta)\lambda^x\theta^{x+\alpha-1}(1-\theta)^{\gamma-1}, 0 < \theta < 1, \lambda > 0.$$



## Um exemplo (para alertar)

Esta densidade conjunta é imprópria porque

$$\begin{aligned} & \int_0^\infty \int_0^1 \exp(-\lambda\theta) \lambda^x \theta^{x+\alpha-1} (1-\theta)^{\gamma-1} d\theta d\lambda \\ &= \int_0^1 \left( \int_0^\infty \exp(-\lambda\theta) \lambda^x d\lambda \right) \theta^{x+\alpha-1} (1-\theta)^{\gamma-1} d\theta \\ &= \int_0^1 \frac{\Gamma(x+1)}{\theta^{x+1}} \theta^{x+\alpha-1} (1-\theta)^{\gamma-1} d\theta \\ &= \Gamma(x+1) \int_0^1 \theta^{\alpha-2} (1-\theta)^{\gamma-1} d\theta = \infty. \end{aligned}$$

## Um exemplo (para alertar)

- ▶ De fato, as distribuições marginais (*a posteriori*) também são impróprias.
- ▶ No entanto, ela possui distribuições condicionais completas que são próprias (**CUIDADO!**):

$$\begin{aligned}(\lambda|\theta, x) &\sim \text{Gama}(x+1, \theta) \\ g(\theta|\lambda, x) &\propto \exp(-\lambda\theta)\theta^{x+\alpha-1}(1-\theta)^{\gamma-1}.\end{aligned}$$

## Um exemplo (para alertar)

- ▶ Assim, por exemplo, o amostrador de Gibbs pode ser empregado com sucesso com essas condicionais completas próprias.
- ▶ Para gerar  $\theta$  de  $g(\theta|\lambda, x)$ , pode-se usar o algoritmo M-H independente descrito anteriormente.
  - ▶ Este algoritmo é conhecido como Metropolis *in* Gibbs<sup>3</sup>
- ▶ O algoritmo é válido como um método MCMC, desde que a distribuição *a posteriori* seja própria. Neste exemplo, ilustramos o método, mas salientamos que qualquer inferência sobre as distribuições marginais *a posteriori* derivadas desta amostra **será totalmente errônea**.

---

<sup>3</sup>Também chamando de Metropolis *within* Gibbs, ou, Gibbs com passo de Metropolis.

## Um exemplo (para alertar)

```
# Configuração  
  
x <- 1  
alpha <- 1  
gamma <- 49  
# lambda <- 100  
  
n <- 10000  
theta <- vector(length = n)  
lambda <- vector(length = n)
```

## Um exemplo (para alertar)

```
# Amostrador de Gibbs
theta[1] <- 0.3 # theta_0 = 0.3

for (t in 1:(n-1)){
  lambda[t] <- rgamma(n = 1, shape = x + 1, rate = theta[t])

  # Passo de M-H
  theta_cand <- rbeta(
    n = 1, shape1 = x + alpha, shape2 = gamma)

  rho <- min(exp(-lambda[t] * (theta_cand - theta[t])), 1)

  theta[(t+1)] <- sample(c(theta_cand, theta[t]),
                        size = 1,
                        prob = c(rho, 1 - rho))
}
lambda[n] <- rgamma(n = 1, shape = x + 1, rate = theta[n])
```

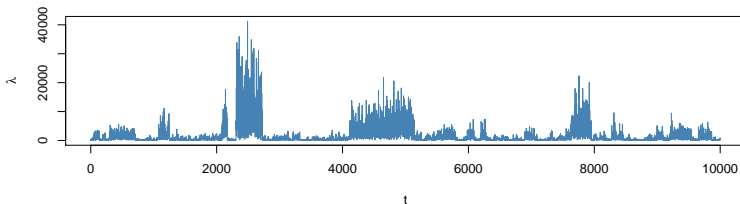
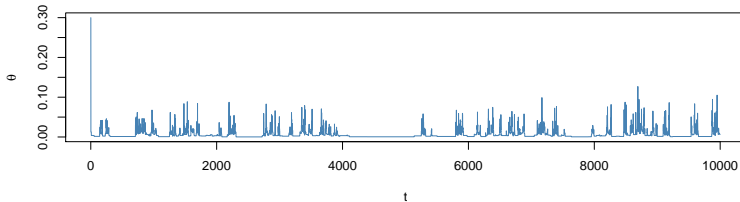
## Um exemplo (para alertar)

```
par(mfrow = c(2,1))

plot(theta, type = "l", col = "steelblue",
      ylab = expression(theta), xlab = "t",
      main = "")

plot(lambda, type = "l", col = "steelblue",
      ylab = expression(lambda), xlab = "t",
      main = "")
```

## Um exemplo (para alertar)



## Um exemplo (para alertar)

- ▶ De fato, a não convergência da cadeia encontrada no último exemplo está longe de ser incomum.
- ▶ Muitas vezes, quando temos uma distribuição *a priori* hierárquica, a distribuição no estágio final da hierarquia é uma “priori objetiva”<sup>4</sup> imprópria.
- ▶ Então não é fácil verificar se a *posteriori* conjunta é própria.
- ▶ Então, nenhum dos teoremas sobre a convergência das cadeias pode se aplicar, **mas a cadeia ainda pode parecer convergir**.
- ▶ Nesses casos, a inferência baseada no MCMC pode ser enganosa no sentido do que foi visto no exemplo acima.

---

<sup>4</sup>A priori de Jeffreys e a priori de Berger-Bernardo são regras de especificação da distribuição *a priori* conhecidas como “objetivas” (em contraste com a elicitação da *priori* a partir do conhecimento [subjetivo] do pesquisador).



## Para casa

- ▶ Rodar os códigos dos exemplos de aula.
  - ▶ Trazer as dúvidas para o Fórum Geral do Moodle e para a próxima aula.
- ▶ Atividade de avaliação (Moodle).

## Próxima aula

- ▶ Modelos bayesianos hierárquicos (introdução).

# Por hoje é só!

Bons estudos!

