

SIMMA 2020

VI Simpósio de Matemática da
Região Fluminense

Introdução ao Método de Monte Carlo via
Cadeias de Markov (MCMC)

Felipe Leite Coelho da Silva
(UFRRJ)

felipeleiterural@gmail.com

Sumário

Introdução

Cadeias de Markov

Método de Monte Carlo Simples

Método de Monte Carlo via Cadeias de Markov

Algoritmo de Metropolis-Hastings

Amostrador de Gibbs

Introdução Inferência Bayesiana

História do MCMC

- ▶ Métodos de Monte Carlo tiveram grande impulso na metade da década de 40 para resolver problemas de difusão de nêutrons.

História do MCMC

- ▶ Métodos de Monte Carlo tiveram grande impulso na metade da década de 40 para resolver problemas de difusão de nêutrons.
- ▶ Algoritmo de Metropolis (Metropolis et al.,1953).

História do MCMC

- ▶ Métodos de Monte Carlo tiveram grande impulso na metade da década de 40 para resolver problemas de difusão de nêutrons.
- ▶ Algoritmo de Metropolis (Metropolis et al.,1953).
- ▶ Método de amostragem de Monte Carlo usando Cadeias de Markov e suas aplicações (Hastings, W. K., 1970).

Cadeias de Markov

- ▶ Um processo estocástico é uma família de variáveis aleatórias $\{X_{(t)}, t \in T\}$ definidas em um espaço de probabilidade, onde t varia no conjunto T .
 - ▶ o conjunto T é um conjunto de índices;
 - ▶ os valores assumidos por $X_{(t)}$ são os estados;

Cadeias de Markov

- ▶ Um processo estocástico é uma família de variáveis aleatórias $\{X_{(t)}, t \in T\}$ definidas em um espaço de probabilidade, onde t varia no conjunto T .
 - ▶ o conjunto T é um conjunto de índices;
 - ▶ os valores assumidos por $X_{(t)}$ são os estados;
- ▶ O conjunto S formado por todos os possíveis estados é chamado de espaço de estados.

Cadeias de Markov

- ▶ Um processo estocástico é uma família de variáveis aleatórias $\{X_{(t)}, t \in T\}$ definidas em um espaço de probabilidade, onde t varia no conjunto T .
 - ▶ o conjunto T é um conjunto de índices;
 - ▶ os valores assumidos por $X_{(t)}$ são os estados;
- ▶ O conjunto S formado por todos os possíveis estados é chamado de espaço de estados.
- ▶ Uma cadeia de Markov é um processo estocástico X_0, X_1, \dots tal que a distribuição de X_t dados todos os valores anteriores de $\{X_0, \dots, X_{t-1}\}$ depende apenas de X_{t-1} , isto é

$$P(X_t \in A | X_0, \dots, X_{t-1}) = P(X_t \in A | X_{t-1}) \quad (1)$$

para qualquer subconjunto $A \subset S$.

- ▶ A evolução da cadeia de Markov é dada pela probabilidade (ou núcleo) de transição:
 - ▶ Espaço de estados discretos:

$$P(x, y) = P(X_t = y | X_{t-1} = x), \quad x, y \in S. \quad (2)$$

- ▶ Espaço de estados contínuos:

$$P(x, y) = P(X_t \leq y | X_{t-1} = x), \quad x, y \in S. \quad (3)$$

- ▶ A evolução da cadeia de Markov é dada pela probabilidade (ou núcleo) de transição:

- ▶ Espaço de estados discretos:

$$P(x, y) = P(X_t = y | X_{t-1} = x), \quad x, y \in S. \quad (2)$$

- ▶ Espaço de estados contínuos:

$$P(x, y) = P(X_t \leq y | X_{t-1} = x), \quad x, y \in S. \quad (3)$$

A função de probabilidade de transição satisfaz as seguintes condições:

(i)

$$P(x, y) \geq 0, \forall x, y \in S; \quad (4)$$

(ii)

$$\sum_{y \in S} P(x, y) = 1, \forall x \in S. \quad (5)$$

- ▶ Cadeia de Markov homogênea:

$$P(X_n = y | X_{n-1} = x) = P(X_1 = y | X_0 = x) = P(x, y),$$

para todo $n \geq 1$ e para todo $x, y \in S$.

- ▶ Cadeia de Markov homogênea:

$$P(X_n = y | X_{n-1} = x) = P(X_1 = y | X_0 = x) = P(x, y),$$

para todo $n \geq 1$ e para todo $x, y \in S$.

- ▶ A matriz de transição P com (i, j) -ésimo elemento dado por $P(x_i, x_j)$.
- ▶ Supondo que o espaço de estados S é finito, com r elementos, então a matriz de transição P é dada por

$$P = \begin{bmatrix} P(x_1, x_1) & \cdots & P(x_1, x_r) \\ P(x_2, x_1) & P(x_2, x_2) & \cdots \\ \vdots & \cdots & \cdots \\ P(x_r, x_1) & \cdots & P(x_r, x_r) \end{bmatrix}$$

- ▶ Propriedades da matriz de transição.

A probabilidade de transição do estado x para o estado y em n passos é denotado por $P^n(x, y)$.

Assim,

$$\begin{aligned} P^n(x, y) &= P(X_n = y | X_0 = x) \\ &= \sum_{x_1} \dots \sum_{x_{n-1}} P(X_n = y, X_{n-1} = x_{n-1}, \dots, X_1 = x_1 | X_0 = x) \\ &= \sum_{x_1} \dots \sum_{x_{n-1}} P(X_n = y | X_{n-1} = x_{n-1}) \dots P(X_1 = x_1 | X_0 = x) \\ &= \sum_{x_1} \dots \sum_{x_{n-1}} P(x, x_1) P(x_1, x_2) \dots P(x_{n-1}, y) \end{aligned} \tag{6}$$

▶ Equações de Chapman-Kolmogorov

$$P^{n+m}(x, y) = P(X_{n+m} = y | X_0 = x) = \sum_z P^n(x, z) P^m(z, y) \quad (7)$$

- ▶ Assumindo que a cadeia é homogênea e as operações são sempre em função dos elementos do espaço de estados, temos que

$$P^{n+m} = P^n P^m, \text{ em particular } P^{n+1} = P^n P \quad (8)$$

- ▶ A distribuição de probabilidade da cadeia no n -ésimo passo pode ser representada pelo vetor π_n , com as componentes $\pi_n(x_i)$, $\forall x_i \in S$.

- ▶ A distribuição de probabilidade da cadeia no n -ésimo passo pode ser representada pelo vetor π_n , com as componentes $\pi_n(x_i)$, $\forall x_i \in S$.
- ▶ Espaço de estados finito: $S = \{x_1, x_2, \dots, x_r\}$.

$$\pi_n = (\pi_n(x_1), \pi_n(x_2), \dots, \pi_n(x_r))$$

- ▶ A distribuição de probabilidade da cadeia no n -ésimo passo pode ser representada pelo vetor π_n , com as componentes $\pi_n(x_i)$, $\forall x_i \in S$.
- ▶ Espaço de estados finito: $S = \{x_1, x_2, \dots, x_r\}$.

$$\pi_n = (\pi_n(x_1), \pi_n(x_2), \dots, \pi_n(x_r))$$

- ▶ Distribuição de probabilidade inicial ($n = 0$) da cadeia:

$$\pi_0 = (\pi_0(x_1), \pi_0(x_2), \dots, \pi_0(x_r))$$

- ▶ A distribuição de probabilidade da cadeia no n -ésimo passo pode ser representada pelo vetor π_n , com as componentes $\pi_n(x_i)$, $\forall x_i \in S$.
- ▶ Espaço de estados finito: $S = \{x_1, x_2, \dots, x_r\}$.

$$\pi_n = (\pi_n(x_1), \pi_n(x_2), \dots, \pi_n(x_r))$$

- ▶ Distribuição de probabilidade inicial ($n = 0$) da cadeia:

$$\pi_0 = (\pi_0(x_1), \pi_0(x_2), \dots, \pi_0(x_r))$$

- ▶ Assim,

$$\begin{aligned} \pi_n &= P(X_n = y) \\ &= \sum_{x \in S} P(X_n = y | X_0 = x) P(X_0 = x) \\ &= \sum_{x \in S} P^n(x, y) \pi_0(x) \end{aligned}$$

- ▶ Forma matricial:

$$\pi_n = \pi_0 P^n \quad (9)$$

Assim,

$$\pi_n = \pi_0 P^n = \pi_0 P^{n-1} P = \pi_{n-1} P \quad (10)$$

Exemplo 1:

Hippie viajante (Kleinrock, 1975). Um hippie viajante começa tomando carona em uma cidade 0. Por algum motivo, ele sempre pega uma nova carona ao final de cada dia, com o primeiro que se dispõe a levá-lo, sem se preocupar com o destino. Ele acaba viajando aleatoriamente pelas cidades próximas com as probabilidade ilustradas.

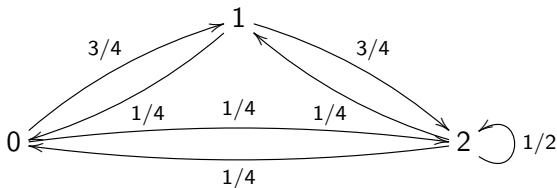


Figura: Grafo de transição de uma cadeia de Markov com três estados (0, 1, 2).

Exemplo 1:

- ▶ Depois do primeiro dia, ele estará na cidade 1 com probabilidade $3/4$ ou na cidade 2 com probabilidade $1/4$.

Exemplo 1:

- ▶ Depois do primeiro dia, ele estará na cidade 1 com probabilidade $3/4$ ou na cidade 2 com probabilidade $1/4$.
- ▶ A probabilidade de retornar à cidade 0 passando pela cidade 1 no segundo dia é de $3/16$ e de retornar à cidade 0 pela cidade 2 no segundo dia é de $1/16$.

Exemplo 1:

- ▶ Depois do primeiro dia, ele estará na cidade 1 com probabilidade $3/4$ ou na cidade 2 com probabilidade $1/4$.
- ▶ A probabilidade de retornar à cidade 0 passando pela cidade 1 no segundo dia é de $3/16$ e de retornar à cidade 0 pela cidade 2 no segundo dia é de $1/16$.
- ▶ Portanto, a probabilidade de retornar à cidade 0 no segundo dia é de $4/16 = 1/4$

Exemplo 1:

- ▶ Depois do primeiro dia, ele estará na cidade 1 com probabilidade $3/4$ ou na cidade 2 com probabilidade $1/4$.
- ▶ A probabilidade de retornar à cidade 0 passando pela cidade 1 no segundo dia é de $3/16$ e de retornar à cidade 0 pela cidade 2 no segundo dia é de $1/16$.
- ▶ Portanto, a probabilidade de retornar à cidade 0 no segundo dia é de $4/16 = 1/4$
- ▶ Sendo assim, podemos fazer o cálculo do viajante retornar à cidade 0 nos dias 3, 4 etc e também para as outras cidades.

Exemplo 1:

▶ $S = \{0, 1, 2\}$

Exemplo 1:

- ▶ $S = \{0, 1, 2\}$
- ▶ Matriz de transição:

$$P = \begin{bmatrix} 0 & 3/4 & 1/4 \\ 1/4 & 0 & 3/4 \\ 1/4 & 1/4 & 1/2 \end{bmatrix}$$

Exemplo 1:

- ▶ $S = \{0, 1, 2\}$
- ▶ Matriz de transição:

$$P = \begin{bmatrix} 0 & 3/4 & 1/4 \\ 1/4 & 0 & 3/4 \\ 1/4 & 1/4 & 1/2 \end{bmatrix}$$

- ▶ Se a próxima cidade visitada só depende da cidade atual e também não depende do dia, tem-se uma cadeia de Markov homogênea.

$$P(X_n = y | X_{n-1} = x) = P(X_1 = y | X_0 = x) = P(x, y), \quad (11)$$

para todo $n \geq 1$ e para todo $x, y \in S$.

Exemplo 1:

- ▶ Evolução do sistema
- ▶ Seja $\pi_n(x) = P(X_n = x)$ a probabilidade de encontrar o sistema no estado x no passo n .

Exemplo 1:

- ▶ Evolução do sistema
- ▶ Seja $\pi_n(x) = P(X_n = x)$ a probabilidade de encontrar o sistema no estado x no passo n .
- ▶ Supondo que $\pi_0 = (\pi_0(0), \pi_0(1), \pi_0(2)) = (1, 0, 0)$
- ▶ Estado inicial: π_0

Exemplo 1:

- ▶ Evolução do sistema
- ▶ Seja $\pi_n(x) = P(X_n = x)$ a probabilidade de encontrar o sistema no estado x no passo n .
- ▶ Supondo que $\pi_0 = (\pi_0(0), \pi_0(1), \pi_0(2)) = (1, 0, 0)$
- ▶ Estado inicial: π_0
- ▶ Estados seguintes:

$$\pi_1 = \pi_0 P$$

$$\pi_2 = \pi_1 P = (\pi_0 P) P = \pi_0 P^2$$

$$\vdots = \vdots$$

$$\pi_n = \pi_{n-1} P = \pi_0 P^n$$

Tabela: Viajante iniciou a sua viagem pela cidade 0.

| n | 0 | 1 | 2 | 3 | 4 | ... | ∞ |
|------------|---|-------|-------|-------|-------|-----|----------|
| $\pi_n(0)$ | 1 | 0.000 | 0.250 | 0.188 | 0.203 | ... | 0.200 |
| $\pi_n(1)$ | 0 | 0.750 | 0.062 | 0.359 | 0.254 | ... | 0.280 |
| $\pi_n(2)$ | 0 | 0.250 | 0.688 | 0.453 | 0.543 | ... | 0.520 |

Tabela: Viajante iniciou a sua viagem pela cidade 1.

| n | 0 | 1 | 2 | 3 | 4 | ... | ∞ |
|------------|---|-------|-------|-------|-------|-----|----------|
| $\pi_n(0)$ | 0 | 0.250 | 0.188 | 0.203 | 0.199 | ... | 0.200 |
| $\pi_n(1)$ | 1 | 0.000 | 0.375 | 0.250 | 0.289 | ... | 0.280 |
| $\pi_n(2)$ | 0 | 0.750 | 0.438 | 0.547 | 0.512 | ... | 0.520 |

Tabela: Viajante iniciou a sua viagem pela cidade 3.

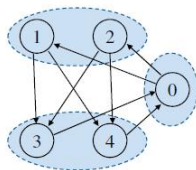
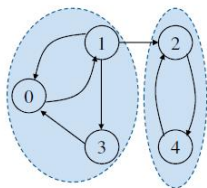
| n | 0 | 1 | 2 | 3 | 4 | ... | ∞ |
|------------|---|-------|-------|-------|-------|-----|----------|
| $\pi_n(0)$ | 0 | 0.250 | 0.188 | 0.203 | 0.199 | ... | 0.200 |
| $\pi_n(1)$ | 0 | 0.250 | 0.312 | 0.266 | 0.285 | ... | 0.280 |
| $\pi_n(2)$ | 1 | 0.500 | 0.500 | 0.531 | 0.516 | ... | 0.520 |

Tabela: Viajante iniciou a sua viagem pela cidade ?

| n | 0 | 1 | 2 | 3 | 4 | ... | ∞ |
|------------|-----|-------|-------|-------|-------|-----|----------|
| $\pi_n(0)$ | 0.5 | 0.125 | 0.219 | 0.195 | 0.201 | ... | 0.200 |
| $\pi_n(1)$ | 0.3 | 0.425 | 0.206 | 0.308 | 0.271 | ... | 0.280 |
| $\pi_n(2)$ | 0.2 | 0.450 | 0.575 | 0.497 | 0.528 | ... | 0.520 |

Definições

- Uma cadeia de Markov é **irredutível** se existem n, m tais que $P^n(x, y) > 0$ e $P^m(y, x) > 0 \forall x, y \in S$.



- ▶ O período de um estado $x \in S$ de uma cadeia de Markov é definido por $d(x) = \text{mdc}\{n \geq 1 | P^n(x, x) > 0\}$.

- ▶ O período de um estado $x \in S$ de uma cadeia de Markov é definido por $d(x) = \text{mdc}\{n \geq 1 | P^n(x, x) > 0\}$.
- ▶ Quando $d(x) = 1$, então o estado x é aperiódico.

- ▶ O período de um estado $x \in S$ de uma cadeia de Markov é definido por $d(x) = \text{mdc}\{n \geq 1 | P^n(x, x) > 0\}$.
- ▶ Quando $d(x) = 1$, então o estado x é aperiódico.
- ▶ Uma cadeia é **aperiódica** se todos os estados são aperiódicos ($d(x) = 1$) e a cadeia é periódica, se $d(x) > 1$.

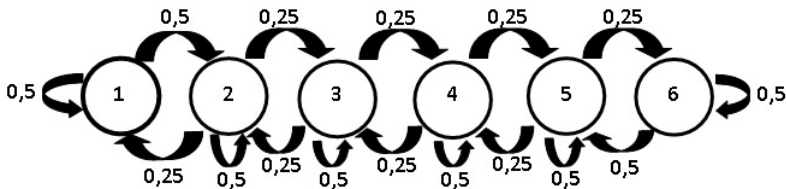


Figura: Grafo de transição.

- ▶ Uma cadeia de Markov com núcleo de transição $P(x, y)$ é dita ser **reversível** se existe uma distribuição de probabilidade π que satisfaz

$$\pi(x)P(x, y) = \pi(y)P(y, x) \quad (12)$$

para todo $x, y \in S$.

Distribuição estacionária e distribuição limite

- ▶ Uma distribuição estacionária $\pi(x)$, $x \in S$, satisfaz as seguintes condições:
 - ▶ $\pi(x) \geq 0$, $\forall x \in S$;
 - ▶ $\sum_{x \in S} \pi(x) = 1$;
 - ▶ $\sum_{x \in S} \pi(x)P(x, y) = \pi(y)$, $\forall y \in S$
- ▶ Forma matricial:

$$\pi P = \pi$$

Distribuição estacionária e distribuição limite

- ▶ Uma distribuição estacionária $\pi(x)$, $x \in S$, satisfaz as seguintes condições:

- ▶ $\pi(x) \geq 0$, $\forall x \in S$;

- ▶ $\sum_{x \in S} \pi(x) = 1$;

- ▶ $\sum_{x \in S} \pi(x)P(x, y) = \pi(y)$, $\forall y \in S$

- ▶ Forma matricial:

$$\pi P = \pi$$

- ▶ Distribuição limite:

$$\lim_{n \rightarrow \infty} P^n(x, y) = \pi(y), y \in S$$

Distribuição estacionária e distribuição limite

- ▶ Se uma cadeia de Markov é reversível, então a distribuição de probabilidade π da cadeia é estacionária (Gamerman e Lopes, 2006; Robert e Casella, 1999; Tierney, 1994).

Distribuição estacionária e distribuição limite

- ▶ Se uma cadeia de Markov é reversível, então a distribuição de probabilidade π da cadeia é estacionária (Gamerman e Lopes, 2006; Robert e Casella, 1999; Tierney, 1994).
- ▶ Se uma cadeia de Markov é irredutível, aperiódica e com distribuição estacionária π então
 - (i) a cadeia é recorrente positiva;
 - (ii) π é a única distribuição estacionária da cadeia;
 - (iii) $\pi(y)$ é a distribuição limite, isto é, $\lim_{n \rightarrow \infty} P^n(x, y) = \pi(y)$, $\forall x, y \in S$;
 - (iv) portanto, a distribuição de probabilidade π_n (quando $n \rightarrow \infty$) da cadeia converge para a distribuição de probabilidade estacionária π , independentemente da distribuição inicial da cadeia (Gamerman e Lopes, 2006; Tierney, 1994).

Problema do hippie viajante:

- ▶ Distribuição estacionária: $\pi^\infty = [0,200 \quad 0,280 \quad 0,520]$
- ▶ Distribuição limite:

$$\lim_{n \rightarrow \infty} P^n(x, y) = \begin{bmatrix} 0,200 & 0,280 & 0,520 \\ 0,200 & 0,280 & 0,520 \\ 0,200 & 0,280 & 0,520 \end{bmatrix}$$

- ▶ Portanto, a distribuição de probabilidade π_n da cadeia converge para distribuição estacionária π .

Método de Monte Carlo Simples

Método de Monte Carlo Simples

- ▶ É um método de simulação para obter integrais de funções complexas ou multidimensionais de forma aproximada.

Método de Monte Carlo Simples

- ▶ É um método de simulação para obter integrais de funções complexas ou multidimensionais de forma aproximada.

- ▶ Também é utilizado para resolver problemas de otimização.

Método de Monte Carlo Simples

- ▶ A ideia inicial do método é escrever a integral que se deseja calcular como um valor esperado. Então, vamos calcular a integral de $g(\theta)$ no intervalo (a, b) , isto é,

$$I = \int_a^b g(\theta) d\theta. \quad (13)$$

Método de Monte Carlo Simples

- ▶ A ideia inicial do método é escrever a integral que se deseja calcular como um valor esperado. Então, vamos calcular a integral de $g(\theta)$ no intervalo (a, b) , isto é,

$$I = \int_a^b g(\theta) d\theta. \quad (13)$$

- ▶ Esta integral (13) pode ser reescrita como

$$I = \int_a^b (b-a)g(\theta) \frac{1}{(b-a)} d\theta = (b-a)E[g(\theta)], \quad (14)$$

sendo θ uma v.a. com distribuição $U(a, b)$.

- ▶ Para estimar $E[g(\theta)]$, basta encontrarmos uma amostra aleatória de tamanho n , $\theta_1, \dots, \theta_n$, da distribuição $U(a, b)$. Assim, podemos obter uma amostra de valores $g(\theta_1), \dots, g(\theta_n)$ da função $g(\theta)$ e, a integral acima pode ser estimada pela média amostral, isto é

$$\hat{I} = (b - a) \frac{1}{n} \sum_{i=1}^n g(\theta_i). \quad (15)$$

- ▶ Para estimar $E[g(\theta)]$, basta encontrarmos uma amostra aleatória de tamanho n , $\theta_1, \dots, \theta_n$, da distribuição $U(a, b)$. Assim, podemos obter uma amostra de valores $g(\theta_1), \dots, g(\theta_n)$ da função $g(\theta)$ e, a integral acima pode ser estimada pela média amostral, isto é

$$\hat{I} = (b - a) \frac{1}{n} \sum_{i=1}^n g(\theta_i). \quad (15)$$

- ▶ Não é difícil verificar que esta estimativa é não viesada.

Algoritmo

- ▶ Passo 1: gerar $\theta_1, \dots, \theta_n$ da distribuição $U(a, b)$;
- ▶ Passo 2: calcule $g(\theta_1), \dots, g(\theta_n)$;
- ▶ Passo 3: calcule a média amostral $\bar{g} = \frac{1}{n} \sum_{i=1}^n g(\theta_i)$;
- ▶ Passo 4: calcule $\hat{I} = (b - a)\bar{g}$.

Exemplo:

- ▶ Suponha que queremos calcular $\int_2^5 e^{-x} dx$.
- ▶ A integral pode ser reescrita por

$$(5 - 2) \int_2^5 \frac{e^{-x}}{(5 - 2)} dx, \quad (16)$$

e será aproximada usando n valores simulados da distribuição Uniforme no intervalo $(2, 5)$ e calculando $y_i = e^{-x_i}, i = 1, \dots, n$.

Exemplo:

- ▶ O valor aproximado da integral é $\frac{3}{n} \sum_{i=1}^n y_i$.

Tabela: Resultado da integral pelo método de Monte Carlo.

| n | 50 | 200 | 500 | 1000 |
|------------------|--------|--------|--------|--------|
| Valor aproximado | 0.1302 | 0.1299 | 0.1288 | 0.1284 |

- ▶ A integral pode ser calculada de forma exata,

$$\int_2^5 e^{-x} dx = 0,1286. \quad (17)$$

Integral geral

- ▶ Para resolver uma integral geral

$$I = \int_a^b g(\theta)p(\theta)d\theta = E[g(\theta)],$$

utilizamos o mesmo algoritmo alterando a geração dos θ'_i s do passo 1 da $p(\theta)$ e calcula-se

$$\hat{I} = \frac{1}{n} \sum_{i=1}^n g(\theta_i).$$

Integral geral

- ▶ Para resolver uma integral geral

$$I = \int_a^b g(\theta)p(\theta)d\theta = E[g(\theta)],$$

utilizamos o mesmo algoritmo alterando a geração dos θ'_i s do passo 1 da $p(\theta)$ e calcula-se

$$\hat{I} = \frac{1}{n} \sum_{i=1}^n g(\theta_i).$$

- ▶ Como há independência das gerações, pela Lei Forte dos Grandes Números $\hat{I} \rightarrow I$, para $n \rightarrow \infty$.

Erro Padrão de Monte Carlo

- ▶ Além disso, $E[g(\theta_i)] = I$ e $Var(g(\theta_i)) = \frac{1}{n} \sum_{i=1}^n (g(\theta_i) - \bar{g})^2$.

Erro Padrão de Monte Carlo

- ▶ Além disso, $E[g(\theta_i)] = I$ e $Var(g(\theta_i)) = \frac{1}{n} \sum_{i=1}^n (g(\theta_i) - \bar{g})^2$.
- ▶ Portanto, a variância do estimador é

$$Var(\hat{I}) = \frac{1}{n^2} \sum_{i=1}^n (g(\theta_i) - \bar{g})^2.$$

Erro Padrão de Monte Carlo

- ▶ Além disso, $E[g(\theta_i)] = I$ e $Var(g(\theta_i)) = \frac{1}{n} \sum_{i=1}^n (g(\theta_i) - \bar{g})^2$.
- ▶ Portanto, a variância do estimador é

$$Var(\hat{I}) = \frac{1}{n^2} \sum_{i=1}^n (g(\theta_i) - \bar{g})^2.$$

- ▶ Erro padrão de Monte Carlo:

$$EP(\hat{I}) = \sqrt{\frac{1}{n^2} \sum_{i=1}^n (g(\theta_i) - \bar{g})^2}.$$

Monte Carlo via Cadeias de Markov

- ▶ O método de Monte Carlo via Cadeias de Markov (MCMC) permite gerar amostras de uma v.a. X que possui uma distribuição de interesse qualquer.

Monte Carlo via Cadeias de Markov

- ▶ O método de Monte Carlo via Cadeias de Markov (MCMC) permite gerar amostras de uma v.a. X que possui uma distribuição de interesse qualquer.
- ▶ No contexto de inferência clássica (frequentista), esta distribuição pode surgir de um modelo complexo.

Monte Carlo via Cadeias de Markov

- ▶ O método de Monte Carlo via Cadeias de Markov (MCMC) permite gerar amostras de uma v.a. X que possui uma distribuição de interesse qualquer.
- ▶ No contexto de inferência clássica (frequentista), esta distribuição pode surgir de um modelo complexo.
- ▶ No contexto de inferência bayesiana, a distribuição de interesse é a distribuição *a posteriori* do vetor paramétrico.

Monte Carlo via Cadeias de Markov

- ▶ Os métodos de MCMC são uma alternativa aos métodos não iterativos em problemas complexos.

Monte Carlo via Cadeias de Markov

- ▶ Os métodos de MCMC são uma alternativa aos métodos não iterativos em problemas complexos.
- ▶ Ele é um método de simulação iterativa, baseado em cadeias de Markov, e assim os valores gerados não serão mais independentes.

Monte Carlo via Cadeias de Markov

- ▶ Os métodos de MCMC são uma alternativa aos métodos não iterativos em problemas complexos.
- ▶ Ele é um método de simulação iterativa, baseado em cadeias de Markov, e assim os valores gerados não serão mais independentes.
- ▶ De acordo com Gamermam e Migon (1999), a ideia central do MCMC é construir uma cadeia de Markov que seja fácil de simular e que tenha a distribuição estacionária dada pela distribuição de interesse.

- ▶ Suponha uma cadeia de Markov $\{X_t, t = 0, 1, \dots\}$.

- ▶ Suponha uma cadeia de Markov $\{X_t, t = 0, 1, \dots\}$.
- ▶ Os métodos MCMC requerem que a cadeia seja:
 - ▶ homogênea, isto é, as probabilidades de transição de um estado para outro são invariantes;
 - ▶ irredutível, isto é, cada estado pode ser atingido a partir de qualquer outro em um número finito de iterações;
 - ▶ aperiódica, isto é, não atinge o mesmo ponto com regularidade fixa e não há estados absorventes.

- ▶ Suponha uma cadeia de Markov $\{X_t, t = 0, 1, \dots\}$.
- ▶ Os métodos MCMC requerem que a cadeia seja:
 - ▶ homogênea, isto é, as probabilidades de transição de um estado para outro são invariantes;
 - ▶ irredutível, isto é, cada estado pode ser atingido a partir de qualquer outro em um número finito de iterações;
 - ▶ aperiódica, isto é, não atinge o mesmo ponto com regularidade fixa e não há estados absorventes.
- ▶ Logo,

$$\lim_{i \rightarrow \infty} X_t^i = \pi,$$

e

$$\frac{1}{n} \sum_{i=1}^n g(X_t^i) \rightarrow E_{\pi}[g(X_t)],$$

quando $n \rightarrow \infty$ e i denota a i -ésima realização da cadeia.

- ▶ A ideia é que conforme o número de iterações aumenta, a cadeia gradualmente esquece os valores iniciais e eventualmente converge para uma distribuição de equilíbrio (Tierney, 1994).

- ▶ A ideia é que conforme o número de iterações aumenta, a cadeia gradualmente esquece os valores iniciais e eventualmente converge para uma distribuição de equilíbrio (Tierney, 1994).
- ▶ Na prática, as iterações iniciais são descartadas, como se formassem uma amostra de aquecimento.

- ▶ A ideia é que conforme o número de iterações aumenta, a cadeia gradualmente esquece os valores iniciais e eventualmente converge para uma distribuição de equilíbrio (Tierney, 1994).
- ▶ Na prática, as iterações iniciais são descartadas, como se formassem uma amostra de aquecimento.
- ▶ Na literatura, existem várias propostas para avaliar a **convergência** das cadeias. Uma forma amplamente utilizada, e de fácil aplicação, é a inspeção gráfica através da análise da trajetória de uma ou mais cadeias em períodos distintos de tempo. Outros critérios empíricos são os diagnósticos de Gelman e Rubin (1992) e Geweke (1992).

- ▶ A ideia é que conforme o número de iterações aumenta, a cadeia gradualmente esquece os valores iniciais e eventualmente converge para uma distribuição de equilíbrio (Tierney, 1994).
- ▶ Na prática, as iterações iniciais são descartadas, como se formassem uma amostra de aquecimento.
- ▶ Na literatura, existem várias propostas para avaliar a **convergência** das cadeias. Uma forma amplamente utilizada, e de fácil aplicação, é a inspeção gráfica através da análise da trajetória de uma ou mais cadeias em períodos distintos de tempo. Outros critérios empíricos são os diagnósticos de Gelman e Rubin (1992) e Geweke (1992).
- ▶ Os métodos MCMC mais utilizados são o amostrador de Gibbs e o algoritmo de Metropolis-Hastings.

Algoritmo de Metropolis-Hastings

- ▶ O algoritmo de Metropolis foi apresentado inicialmente por Metropolis et al. (1953) e generalizado por Hastings (1970) resultando no algoritmo de Metropolis-Hastings.

Algoritmo de Metropolis-Hastings

- ▶ O algoritmo de Metropolis foi apresentado inicialmente por Metropolis et al. (1953) e generalizado por Hastings (1970) resultando no algoritmo de Metropolis-Hastings.
- ▶ Utiliza a ideia dos métodos de rejeição, isto é, um valor é gerado de uma distribuição auxiliar e aceito com uma dada probabilidade.
 - ▶ Suponha que a cadeia esteja no estado θ e um valor θ^* é gerado de uma distribuição proposta $q(\cdot|\theta)$.
 - ▶ A distribuição proposta pode depender do estado atual da cadeia, por exemplo $q(\cdot|\theta)$ poderia ser uma distribuição normal centrada em θ .
 - ▶ Assim, o novo valor θ^* é aceito com probabilidade

$$\alpha(\theta, \theta^*) = \min \left(1, \frac{\pi(\theta^*)q(\theta|\theta^*)}{\pi(\theta)q(\theta^*|\theta)} \right),$$

em que π é a distribuição de interesse.

Algoritmo de Metropolis-Hastings

O algoritmo de Metropolis-Hastings pode ser especificado pelos seguintes passos:

- ▶ Passo 1: Inicialize o contador de iterações $i = 0$ e especifique um valor inicial $\theta^{(0)}$;

Algoritmo de Metropolis-Hastings

O algoritmo de Metropolis-Hastings pode ser especificado pelos seguintes passos:

- ▶ Passo 1: Inicialize o contador de iterações $i = 0$ e especifique um valor inicial $\theta^{(0)}$;
- ▶ Passo 2: Gere um novo valor θ^* da distribuição proposta $q(\cdot|\theta^{(i)})$;

Algoritmo de Metropolis-Hastings

O algoritmo de Metropolis-Hastings pode ser especificado pelos seguintes passos:

- ▶ Passo 1: Inicialize o contador de iterações $i = 0$ e especifique um valor inicial $\theta^{(0)}$;
- ▶ Passo 2: Gere um novo valor θ^* da distribuição proposta $q(\cdot|\theta^{(i)})$;
- ▶ Passo 3: Calcule a probabilidade de aceitação $\alpha(\theta^{(i)}, \theta^*)$ e gere $u \sim U(0, 1)$.

Algoritmo de Metropolis-Hastings

O algoritmo de Metropolis-Hastings pode ser especificado pelos seguintes passos:

- ▶ Passo 1: Inicialize o contador de iterações $i = 0$ e especifique um valor inicial $\theta^{(0)}$;
- ▶ Passo 2: Gere um novo valor θ^* da distribuição proposta $q(\cdot|\theta^{(i)})$;
- ▶ Passo 3: Calcule a probabilidade de aceitação $\alpha(\theta^{(i)}, \theta^*)$ e gere $u \sim U(0, 1)$.
- ▶ Passo 4: Se $u < \alpha$ então aceite o novo valor e faça $\theta^{(i+1)} = \theta^*$, caso contrário rejeite e faça $\theta^{(i+1)} = \theta^{(i)}$.

Algoritmo de Metropolis-Hastings

O algoritmo de Metropolis-Hastings pode ser especificado pelos seguintes passos:

- ▶ Passo 1: Inicialize o contador de iterações $i = 0$ e especifique um valor inicial $\theta^{(0)}$;
- ▶ Passo 2: Gere um novo valor θ^* da distribuição proposta $q(\cdot|\theta^{(i)})$;
- ▶ Passo 3: Calcule a probabilidade de aceitação $\alpha(\theta^{(i)}, \theta^*)$ e gere $u \sim U(0, 1)$.
- ▶ Passo 4: Se $u < \alpha$ então aceite o novo valor e faça $\theta^{(i+1)} = \theta^*$, caso contrário rejeite e faça $\theta^{(i+1)} = \theta^{(i)}$.
- ▶ Passo 5: Incremente o contador de i para $i + 1$ e volte ao passo 2.

A cadeia de Markov gerada pelo algoritmo de Metropolis-Hastings:

- (i) tem $\pi(\theta)$ como sua distribuição estacionária devido ao fato da cadeia ser reversível;
- (ii) é aperiódica, pois permite a ocorrência de eventos tais como $\{\theta_{t+1} = \theta_t\}$, isto é, a probabilidade desses eventos não é zero;
- (iii) é irreduzível, pois $q(\cdot|\cdot)$ satisfaz a seguinte condição

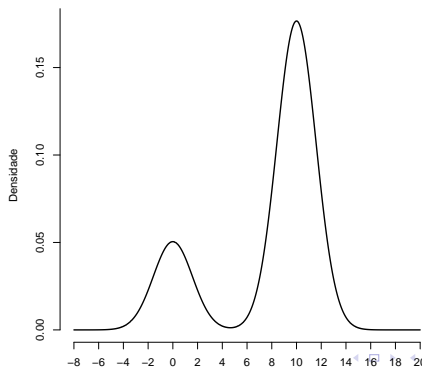
$$q(\theta^*|\theta) > 0, \quad \forall \theta, \theta^* \in S;$$

- (iv) portanto, a distribuição de probabilidade da cadeia converge para a distribuição de interesse $\pi(\theta)$.

Exemplo 2

- ▶ Como gerar uma amostra da distribuição bimodal dada por

$$p(x) = 0,3f_N(x, 0, 10/4) + 0,7f_N(x, 10, 10/4)?$$



Exemplo 1

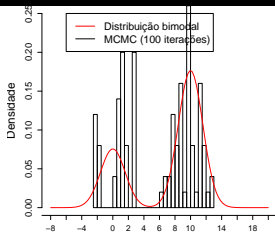
- ▶ Como gerar uma amostra da distribuição bimodal dada por

$$p(x) = 0,3f_N(x, 0, 10/4) + 0,7f_N(x, 10, 10/4) ?$$

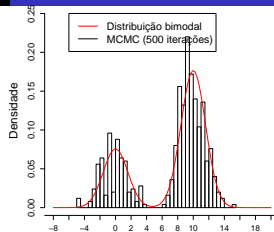
- ▶ Solução:

Distribuição proposta: $X^* \sim N(x^{(i)}, 100)$, então

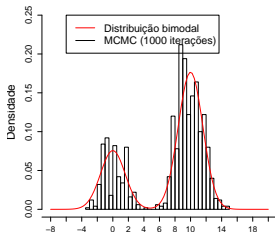
$$\begin{aligned}\alpha(x, x^*) &= \min \left(1, \frac{p(x^*)q(x|x^*)}{p(x)q(x^*|x)} \right) \\ &= \min \left(1, \frac{[0,3e^{-0,2x^{*2}} + 0,7e^{-0,2(x^{*2}-10)^2}]e^{-\frac{1}{2}\frac{(x-x^*)^2}{\sigma^2}}}{[0,3e^{-0,2x^2} + 0,7e^{-0,2(x-10)^2}]e^{-\frac{1}{2}\frac{(x^*-x)^2}{\sigma^2}}} \right) \\ &= \min \left(1, \frac{[0,3e^{-0,2x^{*2}} + 0,7e^{-0,2(x^{*2}-10)^2}]}{[0,3e^{-0,2x^2} + 0,7e^{-0,2(x-10)^2}]} \right) \quad (18)\end{aligned}$$



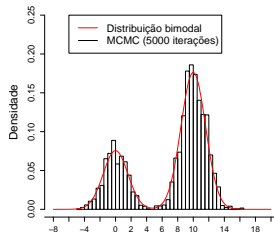
(a) 100 iterações.



(b) 500 iterações.



(c) 1000 iterações.



(d) 5000 iterações.

Amostrador de Gibbs

- ▶ O amostrador de Gibbs é um caso especial do algoritmo de Metropolis-Hastings, com aplicações em métodos Bayesianos.

Amostrador de Gibbs

- ▶ O amostrador de Gibbs é um caso especial do algoritmo de Metropolis-Hastings, com aplicações em métodos Bayesianos.
- ▶ No amostrador de Gibbs a cadeia irá sempre se mover para um novo valor, e as transições de um estado para outro são feitas de acordo com as distribuições condicionais completas.

Amostrador de Gibbs

- ▶ O amostrador de Gibbs é um caso especial do algoritmo de Metropolis-Hastings, com aplicações em métodos Bayesianos.
- ▶ No amostrador de Gibbs a cadeia irá sempre se mover para um novo valor, e as transições de um estado para outro são feitas de acordo com as distribuições condicionais completas.
- ▶ Seja θ um vetor paramétrico. A distribuição condicional completa de θ_i é a distribuição da i -ésima componente de θ condicionada em todas as outras componentes.

Amostrador de Gibbs

- ▶ O amostrador de Gibbs é um caso especial do algoritmo de Metropolis-Hastings, com aplicações em métodos Bayesianos.
- ▶ No amostrador de Gibbs a cadeia irá sempre se mover para um novo valor, e as transições de um estado para outro são feitas de acordo com as distribuições condicionais completas.
- ▶ Seja θ um vetor paramétrico. A distribuição condicional completa de θ_i é a distribuição da i -ésima componente de θ condicionada em todas as outras componentes.
- ▶ Se as distribuições condicionais completas são completamente conhecidas, então podemos utilizar o algoritmo amostrador de Gibbs.

Algoritmo do Amostrador de Gibbs

- ▶ Passo 1: inicialize o contador de iterações da cadeia $i = 0$;
- ▶ Passo 2: especifique os valores iniciais de $\theta^{(0)} = \theta_1^{(0)}, \dots, \theta_d^{(0)}$;
- ▶ Passo 3: obtenha um novo valor de $\theta^{(i)}$ a partir de $\theta^{(i-1)}$ através da geração sucessiva dos valores:

$$\theta_1^{(i)} \sim p(\theta_1 | \theta_2^{(i-1)}, \dots, \theta_d^{(i-1)})$$

$$\theta_2^{(i)} \sim p(\theta_2 | \theta_1^{(i)}, \theta_3^{(i-1)}, \dots, \theta_d^{(i-1)})$$

⋮

$$\theta_d^{(i)} \sim p(\theta_d | \theta_1^{(i)}, \theta_2^{(i)}, \dots, \theta_{d-1}^{(i)}).$$

- ▶ Passo 4: Incremente o contador de i para $i + 1$ e retorne ao passo 3 até obter convergência.

Problema 2:

- ▶ Suponha que queremos gerar amostras de uma distribuição Normal bivariada dada por

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right], \quad (19)$$

onde $\rho = 0,9$.

- ▶ **Algoritmo de Gibbs.** As distribuições condicionais completas podem ser obtidas usando resultados da Normal bivariada (Johnson e Wichern, 1998):

$$Y|X \sim N(\rho x, 1 - \rho^2)$$

$$X|Y \sim N(\rho y, 1 - \rho^2)$$

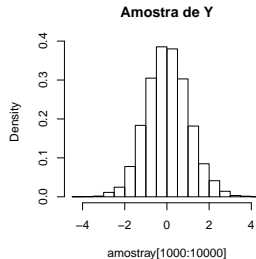
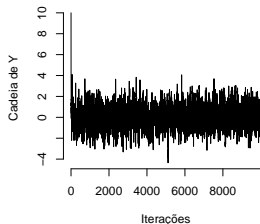
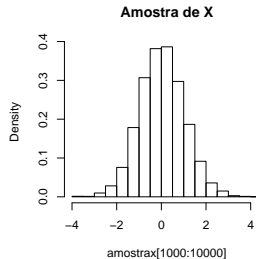
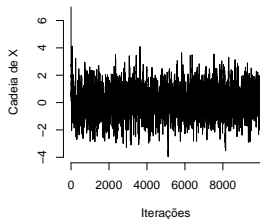
Solução: Amostrador de Gibbs.

- (1) Inicializar o contador de iterações da cadeia $i = 1$;
- (2) Especificar os valores iniciais da cadeia: $\theta_{(1)} = (x_{(1)}, y_{(1)})$;
- (3) Obtenha um novo valor de $\theta_{(i+1)}$ a partir de $\theta_{(i)}$ através da geração sucessiva dos valores

$$x_{(i+1)} \sim N(\rho y_{(i)}, 1 - \rho^2)$$

$$y_{(i+1)} \sim N(\rho x_{(i+1)}, 1 - \rho^2)$$

- (4) Incrementar o contador de $i + 1$ para $i + 2$ e retornar ao passo 3 até obter convergência.



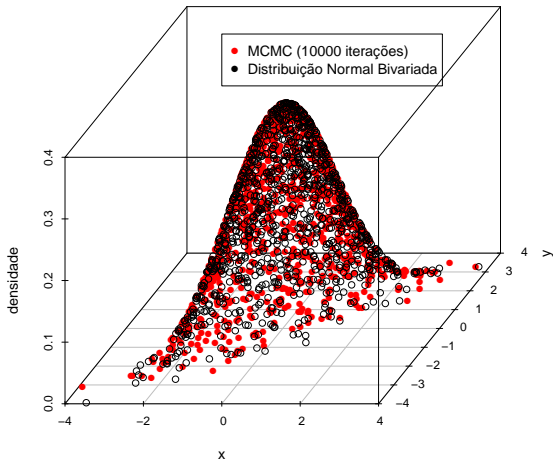


Figura: Distribuição Normal Bivariada

Inferência Bayesiana

- ▶ A informação sobre uma quantidade de interesse θ (não observável) é fundamental na Estatística.

Inferência Bayesiana

- ▶ A informação sobre uma quantidade de interesse θ (não observável) é fundamental na Estatística.
- ▶ Na inferência bayesiana, os parâmetros de um modelo são considerados quantidades aleatórias.

Inferência Bayesiana

- ▶ A informação sobre uma quantidade de interesse θ (não observável) é fundamental na Estatística.
- ▶ Na inferência bayesiana, os parâmetros de um modelo são considerados quantidades aleatórias.
- ▶ Seja y o vetor de observações e θ o vetor de parâmetros. Suponha que temos uma distribuição *a priori* $p(\theta)$, com uma incerteza inicial antes de y ser observado, e a função de verossimilhança do modelo $p(y|\theta)$. Assim, com a especificação de $p(y|\theta)$ e $p(\theta)$, temos um modelo probabilístico dado por

$$p(y, \theta) = p(y|\theta) p(\theta).$$

- ▶ Como os dados y possuem informação sobre θ , pode-se utilizar y para atualizar a informação de θ através da distribuição condicional de θ dado y . Pelo teorema de Bayes, a distribuição *a posteriori* $p(\theta|y)$ é dada por

$$p(\theta|y) = \frac{p(y, \theta)}{p(y)} = \frac{p(y|\theta) p(\theta)}{\int p(y, \theta) d\theta}. \quad (20)$$

- ▶ Como os dados y possuem informação sobre θ , pode-se utilizar y para atualizar a informação de θ através da distribuição condicional de θ dado y . Pelo teorema de Bayes, a distribuição *a posteriori* $p(\theta|y)$ é dada por

$$p(\theta|y) = \frac{p(y, \theta)}{p(y)} = \frac{p(y|\theta) p(\theta)}{\int p(y, \theta) d\theta}. \quad (20)$$

- ▶ Na maior parte das aplicações, $p(\theta|y)$ não possui forma analítica fechada (conhecida).

- ▶ Como os dados y possuem informação sobre θ , pode-se utilizar y para atualizar a informação de θ através da distribuição condicional de θ dado y . Pelo teorema de Bayes, a distribuição *a posteriori* $p(\theta|y)$ é dada por

$$p(\theta|y) = \frac{p(y, \theta)}{p(y)} = \frac{p(y|\theta) p(\theta)}{\int p(y, \theta) d\theta}. \quad (20)$$

- ▶ Na maior parte das aplicações, $p(\theta|y)$ não possui forma analítica fechada (conhecida).
- ▶ Para aproximar a distribuição *a posteriori* dada em (20) podemos utilizar a integração pelo método de Monte Carlo ou MCMC.

- ▶ Utilizando o fato que o denominador na equação (20) não depende de θ , podemos escrever

$$p(\theta|y) \propto p(y|\theta)p(\theta). \quad (21)$$

Exemplo:

- ▶ Seja Y_1, Y_2, \dots, Y_n uma amostra aleatória simples de uma população $N(\mu, \sigma^2)$, com μ e σ^2 desconhecidos. O interesse é estimar (μ, σ^2) . Considere que $\phi = (\sigma^2)^{-1}$ é a precisão.

Exemplo:

- ▶ Seja Y_1, Y_2, \dots, Y_n uma amostra aleatória simples de uma população $N(\mu, \sigma^2)$, com μ e σ^2 desconhecidos. O interesse é estimar (μ, σ^2) . Considere que $\phi = (\sigma^2)^{-1}$ é a precisão.
- ▶ A função de verossimilhança é dada por

$$L(\phi, \mu; y) = (2\pi)^{-n/2} \phi^{n/2} \exp \left\{ \frac{-\phi}{2} \sum_{i=1}^n (y_i - \mu)^2 \right\}. \quad (22)$$

Exemplo:

- ▶ Seja Y_1, Y_2, \dots, Y_n uma amostra aleatória simples de uma população $N(\mu, \sigma^2)$, com μ e σ^2 desconhecidos. O interesse é estimar (μ, σ^2) . Considere que $\phi = (\sigma^2)^{-1}$ é a precisão.
- ▶ A função de verossimilhança é dada por

$$L(\phi, \mu; y) = (2\pi)^{-n/2} \phi^{n/2} \exp \left\{ \frac{-\phi}{2} \sum_{i=1}^n (y_i - \mu)^2 \right\}. \quad (22)$$

- ▶ Objetivo:

$$p(\phi, \mu | y) = ?$$

Exemplo:

- ▶ Seja Y_1, Y_2, \dots, Y_n uma amostra aleatória simples de uma população $N(\mu, \sigma^2)$, com μ e σ^2 desconhecidos. O interesse é estimar (μ, σ^2) . Considere que $\phi = (\sigma^2)^{-1}$ é a precisão.
- ▶ A função de verossimilhança é dada por

$$L(\phi, \mu; y) = (2\pi)^{-n/2} \phi^{n/2} \exp \left\{ \frac{-\phi}{2} \sum_{i=1}^n (y_i - \mu)^2 \right\}. \quad (22)$$

- ▶ Objetivo:

$$p(\phi, \mu | y) = ?$$

- ▶ Sejam as distribuições *a priori* independentes $\mu \sim N(\mu_0, \tau^2)$ e $\phi \sim G(\alpha, \beta)$, com μ_0, τ^2, α e β conhecidos.

- ▶ Logo, a distribuição a posteriori é dada por

$$\begin{aligned} p(\phi, \mu|y) &\propto p(y|\phi, \mu)p(\phi, \mu) \\ &\propto \phi^{n/2} \exp\left\{-\frac{\phi}{2} \sum_{i=1}^n (y_i - \mu)^2\right\} \\ &\exp\left\{-\frac{1}{2} \sum_{i=1}^n \frac{(\mu - \mu_o)^2}{\tau^2}\right\} \phi^{\alpha-1} \exp\{-\beta\phi\}. \end{aligned}$$

- ▶ Logo, a distribuição a posteriori é dada por

$$\begin{aligned} p(\phi, \mu|y) &\propto p(y|\phi, \mu)p(\phi, \mu) \\ &\propto \phi^{n/2} \exp\left\{-\frac{\phi}{2} \sum_{i=1}^n (y_i - \mu)^2\right\} \\ &\exp\left\{-\frac{1}{2} \sum_{i=1}^n \frac{(\mu - \mu_o)^2}{\tau^2}\right\} \phi^{\alpha-1} \exp\{-\beta\phi\}. \end{aligned}$$

- ▶ A distribuição conjunta acima não tem forma conhecida. Entretanto, as condicionais completas a posteriori de cada parâmetro são fáceis de obter,

- $p(\mu|\phi, y) \propto \exp\left\{-\frac{\phi}{2} \sum_{i=1}^n (y_i - \mu)^2\right\} \exp\left\{-\frac{1}{2} \sum_{i=1}^n \frac{(\mu - \mu_o)^2}{\tau^2}\right\};$
- $p(\phi|\mu, y) \propto \phi^{\alpha+n/2-1} \exp\left\{-\phi\left[\beta + \frac{1}{2} \sum_{i=1}^n (y_i - \mu)^2\right]\right\}.$

Segue então que

- ▶ $(\mu|y, \phi) \sim N(m, v)$, em que

$$v = (n\phi + (\tau^2)^{-1})^{-1} \text{ e}$$

$$m = \left(\phi \sum_{i=1}^n y_i + \frac{\mu_0}{\tau^2}\right) v.$$

- ▶ $(\phi|y, \mu) \sim \text{Gama}(\alpha + \frac{n}{2}, \beta + \frac{1}{2} \sum_{i=1}^n (y_i - \mu)^2)$.

Portanto, podemos utilizar o amostrador de Gibbs que pode ser implementado facilmente gerando valores destas distribuições alternadamente.

Solução: Implementada no programa R.

Solução 1: Amostrador de Gibbs.

- (1) Inicializar o contador de iterações da cadeia $i = 1$;
- (2) Especificar os valores iniciais da cadeia: $\theta_{(1)} = (\mu_{(1)}, \phi_{(1)})$;
- (3) Obtenha um novo valor de $\theta_{(i+1)}$ a partir de $\theta_{(i)}$ através da geração sucessiva dos valores

$$\mu_{(i+1)} \sim N \left(\left(\phi_{(i)} \sum_{j=1}^n y_j + \frac{\mu_0}{\tau^2} \right) v, (n\phi_{(i)} + (\tau^2)^{-1})^{-1} \right)$$

$$\phi_{(i+1)} \sim G \left(\alpha + \frac{n}{2}, \beta + \frac{1}{2} \sum_{j=1}^n (y_j - \mu_{(i+1)})^2 \right)$$

- (4) Incrementar o contador de $i + 1$ para $i + 2$ e retornar ao passo 3 até obter convergência.

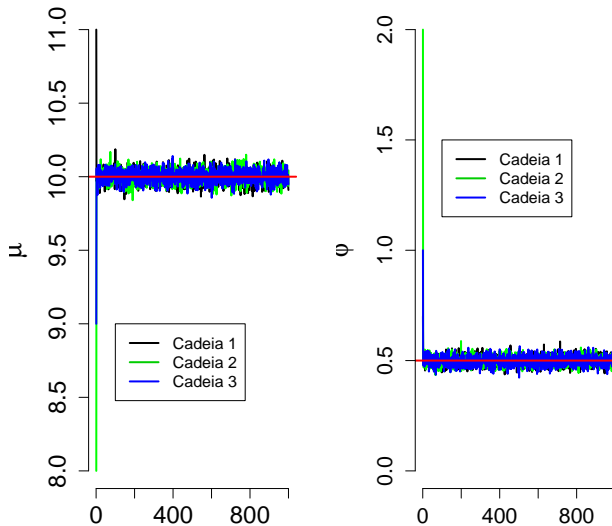


Figura: Via amostrador de Gibbs: Cadeias com 1000 iterações. ▶

Bibliografia

- ▶ Gamerman D., Lopes H.F. (2006). Markov Chain Monte Carlo - Stochastic Simulation for Bayesian Inference. Chapman and Hall, 2^a Ed.
- ▶ Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In J. M. Bernardo, J. O. Berger, A. P.
- ▶ Gelman A., Rubin D.B. (1992). Inference from iterative simulation using multiple sequences. Statistical Science, 7, 457-472.
- ▶ Geman S., Geman D. (1984). Stochastic relaxation, Gibbs distributions and the bayesian restoration of images. IEEE Transactions on Pattern Analysis and Machine Intelligence, 6, 721-741.

- ▶ Hastings W.K. (1970). Monte Carlo sampling methods using Markov Chains and their applications. *Biometrika*, 57, 97-109.
- ▶ Johnson, R. A., Wichern, D. W. Applied Multivariate Statistical Analysis. 4. ed.. New Jersey: Prentice Hall, 1998.
- ▶ Metropolis N., Rosenbulth A.W., Rosenbulth M.N., Teller A.H., Teller E. (1953). Equation of state calculations by fast computing machine. *Journal of Chemical Physics*, 21, 1089-1091.
- ▶ Migon, H. S.; Gamerman, D. (1999). Statistical inference: an integrated approach. London: Arnold.

- ▶ Robert, C. P. e Casella, G. (2004). Monte Carlo Statistical Methods. New York: Springer- Verlag, 2^a Ed.
- ▶ Ross, S. M. (2013). Simulation. Academic Press, 5^a Ed.
- ▶ Souza, R. C.; Brasil, G. H. (1988). Formulação Estrutural - Abordagens Clássica e Bayesiana: Semelhança e Dessemelhanças. Revista de Econometria.
- ▶ Tierney, L. (1994). Markov Chains for exploring posterior distributions. Annals of Statistics, 22, 1701-1728.

Trabalhos desenvolvidos

- ▶ A bottom-up Bayesian extension for long term electricity consumption forecasting. **Energy**, 2019, vol. 167, 198-210.
- ▶ Simulation of the energy efficiency auction prices via the Markov Chain Monte Carlo method. **Energies**, Accepted, 2020.

Anexo

Exemplo: Algoritmo de Metropolis-Hastings

- ▶ Em uma certa população de animais sabe-se que cada animal pode pertencer a uma das 4 linhagens genéticas com probabilidades

$$p_1 = \frac{1}{2} + \frac{\theta}{4}, \quad p_2 = \frac{1 - \theta}{4}, \quad p_3 = \frac{1 - \theta}{4}, \quad p_4 = \frac{\theta}{4}$$

sendo $0 < \theta < 1$ um parâmetro desconhecido.

- ▶ Para qualquer $\theta \in (0, 1)$ é fácil verificar que $p_i > 0, \forall i$ e $\sum_{i=1}^4 p_i = 1$.

Observando-se n animais dentre os quais y_i pertencem à linhagem i , então o vetor aleatório $Y = (y_1; y_2; y_3; y_4)$ tem distribuição multinomial com parâmetros n, p_1, p_2, p_3, p_4 e portanto,

$$p(y|\theta) = \frac{n!}{y_1! y_2! y_3! y_4!} p_1^{y_1} p_2^{y_2} p_3^{y_3} p_4^{y_4} \propto (2 + \theta)^{y_1} (1 - \theta)^{y_2 + y_3} \theta^{y_4}.$$

Assumindo que a distribuição *a priori* $\theta \sim U(0, 1)$, segue que a a distribuição *a posteriori* de θ é dada por

$$p(\theta|y) \propto (2 + \theta)^{y_1} (1 - \theta)^{y_2 + y_3} \theta^{y_4}.$$

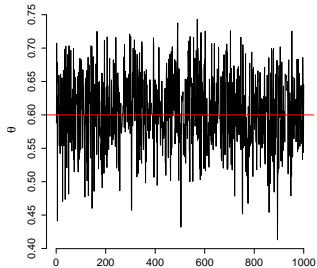
Tomando como distribuição proposta a $U(0, 1)$, então $q(\theta) = 1$, para todo θ .

Assim, a razão de Hastings é dada por

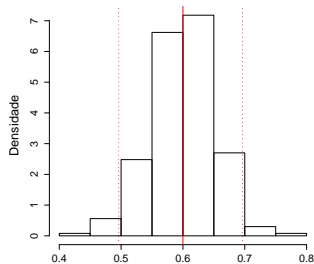
$$\frac{p(\theta^*|y)q(\theta|\theta^*)}{p(\theta|y)q(\theta^*|\theta)} = \frac{(2 + \theta^*)^{y_1} (1 - \theta^*)^{y_2 + y_3} (\theta^*)^{y_4}}{(2 + \theta)^{y_1} (1 - \theta)^{y_2 + y_3} \theta^{y_4}}.$$

Observação:

- ▶ Para verificar a eficiência do algoritmo foram geradas 200 observações de um modelo Multinomial com os parâmetros definidos no problema e $\theta = 0,6$.
- ▶ Foram realizadas 10000 iterações.



(a) Período de aquecimento de 1000 iterações, e com raleamento de 9 iterações.



(b) Histograma.