

3

Método de Monte Carlo com Cadeia de Markov

3.1

Introdução

No capítulo anterior, foram apresentados algoritmos de geração de amostras de variáveis aleatórias com função densidade de probabilidade uniforme e não uniforme. Para variáveis aleatórias não uniformes, mostrou-se que amostras podiam ser obtidas através do Método da Transformada Inversa [7] (caso de distribuições com função distribuição de probabilidade inversíveis), ou Método de Box-Muller (em caso de variáveis aleatórias Gaussianas).

Entretanto, esses métodos podem não serem eficazes em problemas mais complexos. Eles são difíceis de serem aplicados em problemas que por exemplo:

- envolvem variáveis aleatórias com função de distribuição de probabilidade difícil de ser invertida;
- envolvem variáveis aleatórias com função de distribuição de probabilidade difícil de ser normalizada. Ou seja, a função tem a forma de uma determinada distribuição de probabilidade, porém a constante pela qual deve ser multiplicada para se tornar de fato uma distribuição (a integral da função em seu domínio, \mathcal{D} , deve ser igual a um) é difícil de ser determinada.
- necessitam gerar amostras de distribuições de probabilidade condicional.
- necessitam gerar amostras de vetores aleatórios (com uma função densidade de probabilidade conjunta);

Dessa forma, os métodos apresentados tem utilidade limitada e novos métodos mais abrangentes foram desenvolvidos. Nesse capítulo da dissertação, será analisado o método de geração de amostras de variáveis e vetores aleatórios: Monte Carlo com Cadeia de Markov (MCMC).

3.2

Processos Estocásticos e Cadeias de Markov

Para tratar do método de Monte Carlo com Cadeia de Markov, é necessário previamente estudar processos estocásticos e cadeia de Markov.

No capítulo anterior, uma variável aleatória real foi definida como uma função real X que associa a cada ponto-amostra w , pertencente ao espaço de amostras Ω , um número $X(w) \in \mathbb{R}$.

Um processo estocástico é definido como um mapeamento que associa a cada ponto amostra $w \in \Omega$ uma função real de um parâmetro t pertencente a um conjunto \mathcal{T} (na maioria dos processos estocásticos, o parâmetro t está associado ao tempo) [27]. Dessa forma, cria-se uma família \mathcal{F} de funções de t ($t \in \mathcal{T}$).

$$\begin{aligned} X : \quad \Omega &\longmapsto \mathcal{F} \\ w &\longmapsto X(t, w), \quad t \in \mathcal{T} \end{aligned} \quad (3-1)$$

Assim, um processo estocástico é uma função de duas variáveis (w e t) cujos domínios são Ω e $\mathcal{T} \subset \mathbb{R}$, respectivamente. Ao se fixar, por exemplo, o valor w_i para w , o processo estocástico passa a representar uma única função $X(t, w_i)$ de t . Ao se fixar o valor t_i para t , o processo estocástico passa a representar uma variável aleatória que associa a cada ponto amostra um número real $X(t_i, w)$.

Para simplificar a notação, $X(t, w)$, usada para representar um processo estocástico, a variável w será omitida. Dessa forma, será utilizada a representação $X(t)$.

Um processo estocástico pode ser interpretado como um número infinito de variáveis aleatórias indexadas pelo parâmetro t . Fixado um valor de $t = t_i$, a variável aleatória $X(t_i)$ possui uma função distribuição de probabilidade:

$$P_{X(t_i)}(x) = Pr(X(t_i) \leq x) \quad (3-2)$$

e uma função densidade de probabilidade $p_{X(t_i)}$.

A média de um processo estocástico, $X(t)$, é definida pela equação (3-3), a variância por (3-4), a autocorrelação por (3-5) e a autocovariância por (3-6):

$$\mu_X(t) = E[X(t)] = \int_{-\infty}^{\infty} x p_{X(t)}(x) dx \quad (3-3)$$

$$\sigma_X^2(t) = E[\{X(t) - \mu_X(t)\}^2] = E[X^2(t)] - (E[X(t)])^2 \quad (3-4)$$

$$R_{XX}(t_1, t_2) = E[X(t_1) X(t_2)] \quad (3-5)$$

$$\begin{aligned}
\sigma_{XX}(t_1, t_2) &= E[\{X(t_1) - \mu_X(t_1)\} \{X(t_2) - \mu_X(t_2)\}] \\
&= E[X(t_1)X(t_2)] - \mu_X(t_1)\mu_X(t_2) \\
&= R_{XX}(t_1, t_2) - \mu_X(t_1)\mu_X(t_2)
\end{aligned} \tag{3-6}$$

O coeficiente de autocorrelação é:

$$\rho_{XX}(t_1, t_2) = \frac{\sigma_{XX}(t_1, t_2)}{\sigma_X(t_1)\sigma_X(t_2)} \tag{3-7}$$

Um processo de Markov é um processo estocástico no qual tendo-se $t_1 \leq t_2 \leq \dots \leq t_{n-1} \leq t_n$, tem-se [5]:

$$\begin{aligned}
Pr(X(t_n) \leq x_n | X(t_{n-1}) = x_{n-1}, \dots, X(t_1) = x_1) &= \\
&= Pr(X(t_n) \leq x_n | X(t_{n-1}) = x_{n-1})
\end{aligned} \tag{3-8}$$

ou seja, a probabilidade do processo estocástico $X(t)$ assumir um valor x_n em $t = t_n$ depende apenas de $X(t_{n-1})$. Essa definição também vale para processos estocásticos de parâmetro discreto (caso em que $X(t_n)$ pode ser substituído por X_n).

O conceito de cadeia de Markov foi criado por um matemático russo (Andrei. A. Markov) em 1906. Mas somente na década de 20 o termo "cadeia de Markov" começou a ser empregado. Foi Bernstein (outro matemático russo) que o utilizou pela primeira vez em um artigo no ano de 1926.

A cadeia de Markov é formada por um processo de Markov com parâmetro discreto, ou seja por uma sequência de variáveis aleatórias:

$$X_1 \ X_2 \ X_3 \ \dots \ X_n \ \dots \quad \forall n \in \mathbb{N}. \tag{3-9}$$

Definição de Cadeia de Markov [14]: Seja $[T]$ uma matriz $k \times k$ com elementos: $T_{i,j} : i, j = 1, \dots, k$. Um processo estocástico com parâmetro discreto (X_0, X_1, \dots) com um número finito de estados $S = \{s_1, \dots, s_k\}$ é dito ser uma cadeia de Markov com matriz de transição $[T]$, se para todo n , todo $i, j \in \{1, \dots, k\}$ e todo $i_0, \dots, i_{n-1} \in \{1, \dots, k\}$, tem-se:

$$\begin{aligned}
P(X_{n+1} = s_j | X_n = s_i, X_{n-1} = s_{i_{n-1}}, \dots, X_1 = s_{i_1}, X_0 = s_{i_0}) \\
&= P(X_{n+1} = s_j | X_n = s_i) \\
&= T_{i,j}.
\end{aligned} \tag{3-10}$$

Os elementos da matriz $[T]$ são chamados de probabilidades de transição. O elemento $T_{i,j}$ representa a probabilidade da cadeia assumir no instante $n + 1$ o estado s_j dado que no instante n ela assume o estado i . Quando a matriz $[T]$ é invariante no tempo, a cadeia de Markov é dita ser homogênea.

Toda matriz de transição satisfaz:

$$T_{ij} \geq 0 \quad \forall i, j \in \{1, \dots, k\}, \quad (3-11)$$

e

$$\sum_{j=1}^k T_{ij} = 1 \quad \forall i \in \{1, \dots, k\} \quad (3-12)$$

ou seja, o somatório dos elementos de qualquer linha de T é sempre igual a 1. A propriedade (3-11) está associada ao fato de que probabilidades são sempre não negativas e, a propriedade (3-12) ao fato de que:

$$\begin{aligned} P(X_{n+1} = s_1 | X_n = s_i) + P(X_{n+1} = s_2 | X_n = s_i) + \dots \\ + P(X_{n+1} = s_k | X_n = s_i) = 1. \end{aligned} \quad (3-13)$$

Outra importante característica (além da matriz de transição) de uma cadeia de Markov (X_0, X_1, \dots) é sua distribuição de probabilidade inicial. Essa distribuição é representada por um vetor $\{\pi(0)\}$ de k componentes:

$$\begin{aligned} \{\pi(0)\} &= [\pi_1(0) \pi_2(0) \dots \pi_k(0)] \\ &= [P(X_0 = s_1) P(X_0 = s_2) \dots P(X_0 = s_k)]. \end{aligned} \quad (3-14)$$

Cada componente $\pi_i(0)$ de $\{\pi(0)\}$ está associada a probabilidade da cadeia ter no instante inicial (X_0) o estado s_i . Como $\{\pi(0)\}$ é uma distribuição de probabilidade, tem-se:

$$\sum_{i=1}^k \pi_i(0) = 1. \quad (3-15)$$

De forma similar, os vetores $\{\pi(1)\}, \{\pi(2)\}, \dots$ denotam as distribuições da cadeia de Markov nos instantes $1, 2, \dots$. Assim, no instante n :

$$\begin{aligned} \{\pi(n)\} &= [\pi_1(n) \pi_2(n) \dots \pi_k(n)] \\ &= [P(X_n = s_1) P(X_n = s_2) \dots P(X_n = s_k)] \end{aligned} \quad (3-16)$$

com:

$$\sum_{i=1}^k \pi_i(n) = 1. \quad (3-17)$$

A distribuição de probabilidade de uma cadeia de Markov (X_1, X_2, \dots) , com um número finitos de estados $\{s_1, \dots, s_k\}$ e com distribuição inicial de probabilidade $\{\pi(0)\}$, em qualquer instante n satisfaz a equação de Chapman-Kolomogrov [18]:

$$\{\pi(n)\} = \{\pi(0)\} [T]^n. \quad (3-18)$$

Demonstração de (3-18): para todo $j \in \{1, \dots, k\}$, tem-se que:

$$\begin{aligned} \pi_j(n+1) &= P(X_{n+1} = x_j) \\ &= \sum_{i=1}^k P(X_n = s_i, X_{n+1} = x_j) \\ &= \sum_{i=1}^k P(X_n = s_i) \cdot P(X_{n+1} = s_j | X_n = s_i) \\ &= \sum_{i=1}^k \pi_i(n) T_{i,j}. \end{aligned} \quad (3-19)$$

Escrevendo (3-19) em forma matricial: $\{\pi(n+1)\} = \{\pi(n)\} [T]$. Assim, recursivamente:

$$\begin{aligned} \{\pi(n)\} &= \{\pi(n-1)\} [T] = \{\pi(n-2)\} [T][T] = \{\pi(n-2)\} [T]^2 \\ &\vdots \\ &= \{\pi(0)\} [T]^n. \end{aligned} \quad (3-20)$$

□

A seguir são enunciadas algumas definições que serão importantes para a apresentação do Método de Monte Carlo com Cadeia de Markov feita na próxima seção da dissertação.

Definição de cadeia de Markov irreduzível: uma cadeia de Markov (X_0, X_1, \dots) , com um número finito de estados $S = \{s_1, \dots, s_k\}$ e com matriz de transição $[T]$, é dita ser irreduzível se, para todo $s_i, s_j \in S$, $\exists n \in \mathbb{N}$ tal que $T_{i,j}^n > 0$.

Definição do período de um estado da cadeia de Markov: o período de um estado $s_i, s_j \in S$ de uma cadeia de Markov, $d(s_i)$, é definido como:

$$d(s_i) = mdc \{n \geq 1 : T_{i,i}^n > 0\} \quad (3-21)$$

onde $mdc \{a_1, a_2, \dots\}$ indica o máximo divisor comum de a_1, a_2, \dots . Quando $d(s_i) = 1$, então o estado s_i é dito não-periódico.

Definição de cadeia de Markov não-periódica: uma cadeia de Markov, com um número finito de estados $S = \{s_1, \dots, s_k\}$ é dita não-periódica se todos os seus estados são não-periódicos.

Definição da distribuição de probabilidade estacionária de uma cadeia de Markov:

seja uma cadeia de Markov (X_0, X_1, \dots) com um número finito de estados $S = \{s_1, \dots, s_k\}$ e com matriz de transição $[T]$. O vetor $\{\pi^*\}$ é dito ser uma distribuição de probabilidade estacionária, ou invariante, da cadeia de Markov se:

$$\pi_i^* \geq 0 \quad \text{para } i = 1, \dots, k \quad \text{e} \quad \sum_{i=1}^k \pi_i^* = 1 \quad (3-22)$$

e

$$\{\pi^*\} [T] = \{\pi^*\}. \quad (3-23)$$

Para qualquer cadeia de Markov irredutível e não-periódica, existe uma única distribuição de probabilidade estacionária $\{\pi^*\}$ [14]. Para calculá-la, basta resolver o problema de auto-vetor associado ao auto-valor 1: $\{\pi^*\} [T] = \{\pi^*\}$.

Existe um teorema que afirma que a distribuição de probabilidade $\{\pi(n)\}$ (quando $n \rightarrow \infty$) de uma cadeia de Markov irredutível e não-periódica converge para distribuição de probabilidade estacionária, ou seja:

$$\{\pi(n)\} \xrightarrow{n \rightarrow \infty} \{\pi^*\} \quad (3-24)$$

$$\lim_{n \rightarrow \infty} \frac{1}{2} \sum_{i=1}^k |\pi_i^* - \pi_i(n)| = 0 \quad (3-25)$$

Detalhes da prova desse teorema podem ser encontrados em [14].

Definição da distribuição de probabilidade reversível de uma cadeia de Markov:

Uma distribuição de probabilidade $\{\pi\}$ em S é dita ser reversível se, para todo $i, j \in \{1, \dots, k\}$ tem-se:

$$\pi_i T_{i,j} = \pi_j T_{j,i}. \quad (3-26)$$

Quando uma cadeia de Markov possui uma distribuição de probabilidade reversível é considerada uma cadeia de Markov reversível. Além disso, se $\{\pi\}$ é uma distribuição de probabilidade reversível da cadeia, será também distribuição estacionária.

Exemplo 3.2.1 (Cadeia de Markov). Suponha que a matriz de transição $[T]$ de uma cadeia de Markov com 4 possíveis estados seja:

$$[T] = \begin{bmatrix} 1/2 & 1/2 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (3-27)$$

Essa cadeia é não irredutível, não-periódica e possui uma distribuição estacionária $\{\pi^*\} = [0 \ 0 \ 0 \ 1]$. Verifique que $\{\pi^*\}$ é o auto-vetor associado ao auto-valor 1: $\{\pi^*\} = [T] \{\pi^*\}$.

Uma forma usual de representar uma cadeia de Markov é através de um gráfico chamado grafo de transição. O grafo consiste de nós representando os possíveis estados da cadeia e setas ligando os nós representando as probabilidades de transição entre os estados. Quando a probabilidade de transição de um estado i para o outro j é nula, ou seja, $T_{ij} = 0$, ela não deve ser mostrada no grafo.

O grafo da cadeia de Markov para esse exemplo (com os estados s_1, s_2, s_3 e s_4) é mostrado na figura (3.1).

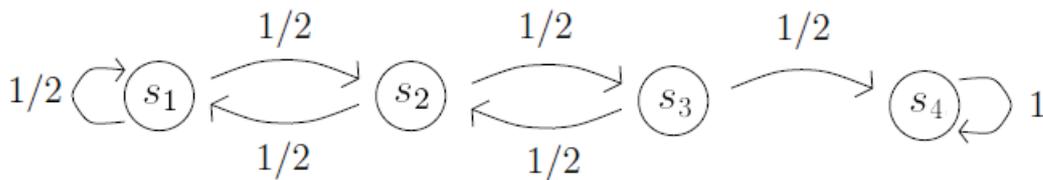


Figura 3.1: Grafo de Transição de uma cadeia de Markov com quatro estados (s_1, s_2, s_3 e s_4).

Exemplo 3.2.2 (Cadeia de Markov). Suponha que a matriz de transição $[T]$ de uma cadeia de Markov com 4 possíveis estados seja:

$$[T] = \begin{bmatrix} 1/2 & 1/2 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 1/2 & 1/2 \end{bmatrix}. \quad (3-28)$$

Essa cadeia é não irredutível, não-periódica e possui duas distribuições estacionárias: $\{\pi^*\} = [1/2 \ 1/2 \ 0 \ 0]$ e $\{\pi^*\} = [0 \ 0 \ 1/2 \ 1/2]$.

O grafo da cadeia de Markov para esse exemplo é mostrado na figura (3.2).



Figura 3.2: Grafo de Transição de uma cadeia de Markov com duas distribuições estacionárias.

Exemplo 3.2.3 (Cadeia de Markov). Suponha que uma máquina de uma fábrica possui dois possíveis estados de operação: quebrada (estado s_1) e funcionando (estado s_2). No primeiro dia de operação da fábrica, a probabilidade da máquina estar quebrada é $\pi_1(0)$ e, a probabilidade de estar funcionando é $\pi_2(0)$. Além disso, a probabilidade do estado da máquina (funcionando ou quebrada) em um dia n depende somente de seu estado no dia anterior, ou seja, de $n - 1$.

De acordo com essas suposições, o estado de operação da máquina no decorrer dos dias forma uma cadeia de Markov X_0, X_1, X_2, \dots . A distribuição de probabilidade inicial dessa cadeia é um vetor $\{\pi(0)\}$ de 2 componentes (os dois possíveis estados):

$$\begin{aligned} \{\pi(0)\} &= [\pi_1(0) \ \pi_2(0)] \\ &= [P(X_0 = s_1) \ P(X_0 = s_2)] \end{aligned} \quad (3-29)$$

com:

$$\pi_1(0) + \pi_2(0) = 1 \quad (3-30)$$

Considere a que as probabilidades da máquina estar funcionando /ou quebrada em um dia n , dado que dia anterior, $n - 1$, estava funcionando /ou quebrada são:

$$\begin{aligned} P(X_{n+1} = s_1 | X_n = s_1) &= 1 - p & P(X_{n+1} = s_2 | X_n = s_1) &= p \\ P(X_{n+1} = s_1 | X_n = s_2) &= q & P(X_{n+1} = s_2 | X_n = s_2) &= 1 - q. \end{aligned}$$

com $p, q \geq 0$. Assim, a matriz de transição $[T]$ 2×2 dessa cadeia será:

$$[T] = \begin{bmatrix} 1 - p & p \\ q & 1 - q \end{bmatrix} \quad (3-31)$$

Pergunta-se: Qual a distribuição de probabilidade estacionária $\{\pi^*\}$ dessa cadeia?

A probabilidade da máquina estar quebrada no dia $n + 1$ em função do seu estado em n é:

$$\begin{aligned} P(X_{n+1} = s_1) &= P(X_n = s_1, X_{n+1} = s_1) + P(X_n = s_2, X_{n+1} = s_1) \\ &= P(X_n = s_1)P(X_{n+1} = s_1 | X_n = s_1) \\ &\quad + P(X_n = s_2)P(X_{n+1} = s_1 | X_n = s_2) \\ &= (1 - p) P(X_n = s_1) + q P(X_n = s_2) \end{aligned} \quad (3-32)$$

ou seja:

$$\pi_1(n + 1) = T_{11} \pi_1(n) + T_{21} \pi_2(n). \quad (3-33)$$

A probabilidade da máquina estar funcionando no dia $n + 1$ em função do seu estado em n é:

$$\begin{aligned} P(X_{n+1} = s_2) &= P(X_n = s_1, X_{n+1} = s_2) + P(X_n = s_2, X_{n+1} = s_2) \\ &= P(X_n = s_1)P(X_{n+1} = s_2 | X_n = s_1) \\ &\quad + P(X_n = s_2)P(X_{n+1} = s_2 | X_n = s_2) \\ &= p P(X_n = s_1) + (1 - q) P(X_n = s_2) \end{aligned} \quad (3-34)$$

ou seja:

$$\pi_2(n + 1) = T_{12} \pi_1(n) + T_{22} \pi_2(n). \quad (3-35)$$

Dessa forma, a distribuição de probabilidade da cadeia no dia $n + 1$, $\{\pi(n + 1)\}$, pode ser determinada a partir da distribuição do dia anterior, $\{\pi(n)\}$:

$$\{\pi(n + 1)\} = \{\pi(n)\} [T]. \quad (3-36)$$

De forma recursiva, pode-se determinar $\{\pi(n)\}$ em função de $\{\pi(0)\}$ (equação de Chapman-Kolomogrov (3-18)):

$$\{\pi(n)\} = \{\pi(0)\} [T]^n \quad (3-37)$$

Desenvolvendo as contas para essa cadeia, cada elemento do vetor $\{\pi(n)\}$ é representado em termos de $\{\pi(0)\}$ como:

$$\begin{cases} \pi_1(n) = (1 - p - q)^n \pi_1(0) + q \sum_{j=0}^{n-1} (1 - p - q)^j \\ \pi_2(n) = (1 - p - q)^n \pi_1(0) + p \sum_{j=0}^{n-1} (1 - p - q)^j. \end{cases} \quad (3-38)$$

A partir desse ponto, as contas podem ser separadas em 2 casos:

- 1º caso: $p = q = 0$:

$$\begin{cases} \pi_1(n) = \pi_1(0) \\ \pi_2(n) = \pi_2(0) \end{cases} \quad (3-39)$$

- 2º caso: $0 < p + q < 2$:

$$\sum_{j=0}^{n-1} (1 - p - q)^j = \frac{1 - (1 - p - q)^n}{p + q}. \quad (3-40)$$

Assim, substituindo (3-40) em (3-38):

$$\begin{cases} \pi_1(n) = \frac{q}{p + q} + (1 - p - q)^n \left[\pi_1(0) - \frac{q}{p + q} \right] \\ \pi_2(n) = \frac{p}{p + q} + (1 - p - q)^n \left[\pi_2(0) - \frac{p}{p + q} \right] \end{cases}. \quad (3-41)$$

Lembrando que:

$$\begin{cases} \pi_1(n) = T_{11}^n \pi_1(0) + T_{21}^n \pi_2(0) = T_{11}^n \pi_1(0) + T_{21}^n [1 - \pi_1(0)] \\ \pi_2(n) = T_{12}^n \pi_1(0) + T_{22}^n \pi_2(0) = T_{12}^n [1 - \pi_2(0)] + T_{22}^n \pi_2(0) \end{cases} \quad (3-42)$$

$$\begin{cases} \pi_1(n) = [T_{11}^n - T_{21}^n] \pi_1(0) + T_{21}^n \\ \pi_2(n) = [T_{22}^n - T_{12}^n] \pi_2(0) + T_{12}^n \end{cases} \quad (3-43)$$

e supondo que $0 < p + q < 2$, pode-se determinar cada um dos elementos de $[T]^n$:

$$[T]^n = \frac{q}{p+q} \begin{bmatrix} q & p \\ q & p \end{bmatrix} + \frac{(1-p-q)^n}{p+q} \begin{bmatrix} p & -p \\ -q & q \end{bmatrix}. \quad (3-44)$$

Tomando-se o limite de (3-44) quando $n \rightarrow \infty$:

$$\lim_{n \rightarrow \infty} [T]^n = \frac{q}{p+q} \begin{bmatrix} q & p \\ q & p \end{bmatrix}. \quad (3-45)$$

Observe que quando $n \rightarrow \infty$, todas as linhas da matriz $[T]^n$ são iguais, ou seja, a probabilidade da cadeia ter um determinado estado no instante $n+1$ não depende de seu estado do instante n .

Visto que a cadeia desse exemplo é irredutível e não-periódica, a distribuição de probabilidade estacionária $\{\pi^*\}$ existe e pode ser calculada por:

$$\{\pi^*\} = \lim_{n \rightarrow \infty} \{\pi(n)\} = \{\pi(0)\} \lim_{n \rightarrow \infty} [T]^n. \quad (3-46)$$

Assim:

$$\{\pi^*\} = [\pi_1(0) \quad \pi_2(0)] \cdot \frac{1}{p+q} \begin{bmatrix} q & p \\ q & p \end{bmatrix}. \quad (3-47)$$

O grafo da cadeia de Markov para esse exemplo da máquina da fábrica (com os estados s_1 e s_2) é mostrado na figura (3.3).

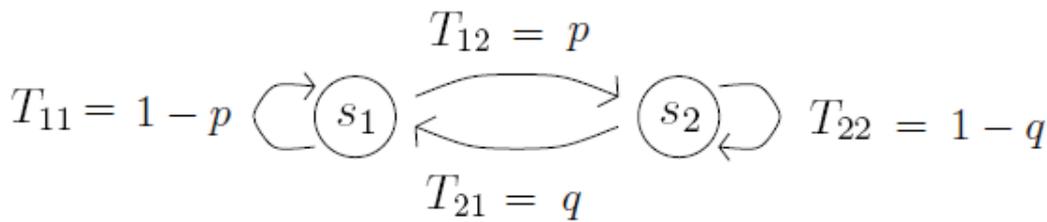


Figura 3.3: Grafo de Transição de uma cadeia de Markov com dois estados (s_1 e s_2).

3.3

O Algoritmo de Metropolis-Hastings Aplicado a Variáveis Aleatórias

A história do algoritmo de Metropolis-Hastings começou no ano de 1953, quando Metropolis percebeu que poderia simular o movimento molecular utilizando uma cadeia de Markov. Ao invés de fazer simulações baseadas em estatísticas (da energia potencial mínima de uma coleção de partículas),

ele optou por considerar que a distribuição espacial das moléculas em um determinado instante era função apenas da distribuição no instante anterior. Ou seja, um processo de Markov.

Quase duas décadas depois, Hastings expandiu o trabalho de Metropolis e formalizou o método de Monte Carlo com Cadeia de Markov (MCMC) baseado no algoritmo de Metropolis-Hastings. O algoritmo permite gerar amostras $x^{(1)}, x^{(2)}, \dots, x^{(i)}$ de uma determinada variável aleatória X , no domínio \mathcal{D} e com uma complicada função densidade de probabilidade p . Vale observar que p pode ser uma densidade conjunta ou até mesmo não normalizada.

A simplicidade do método está associada ao fato de que ao invés de gerar amostras de X a partir de p , gera-se a partir de uma outra variável aleatória Y . Essa variável deve estar definida no mesmo domínio \mathcal{D} e X e, tem uma função densidade de probabilidade, h , chamada de distribuição proposta.

Essa função h é definida pelo usuário do método e, cada amostra $u^{(i)}$ de Y é aceita ou rejeitada como uma amostra $x^{(i)}$ de X de acordo com um critério definido pelo algoritmo de Metropolis-Hastings.

As amostras de X formam uma sequência de valores $x^{(i)}$, $i \in \mathbb{N}$ e, as amostras de Y formam uma sequência $u^{(i)}$, $i \in \mathbb{N}$. O valor inicial $x^{(1)}$ deve ser especificado pelo usuário do algoritmo e, tendo o termo $x^{(i)}$, as etapas para obter-se o próximo termo $x^{(i+1)}$ são [32] [16]:

1. Gerar uma amostra $u^{(i)}$ de Y segundo um procedimento desejado (pode-se utilizar por exemplo o método da Transformada Inversa para a função densidade h).
2. Calcular a função R (geralmente chamada de razão de Metropolis ou razão de Hastings):

$$R(x^{(i)}, u^{(i)}) = \frac{p(u^{(i)}) h(x^{(i)})}{p(x^{(i)}) h(u^{(i)})}. \quad (3-48)$$

3. Avaliar a função α dada por (3-49):

$$\alpha(x^{(i)}, u^{(i)}) = \min [1, R(x^{(i)}, u^{(i)})]. \quad (3-49)$$

4. Sortear um valor $\rho^{(i)}$ da distribuição uniforme unitária no intervalo $[0, 1]$.
5. Aceitar $u^{(i)}$ como a nova amostra $x^{(i+1)}$ com probabilidade α ou rejeita-lá com probabilidade $1 - \alpha$, isto é:

$$x^{(i+1)} = \begin{cases} u^{(i)} & \text{se } \rho^{(i)} \leq \alpha \\ x^{(i)} & \text{se } \rho^{(i)} > \alpha \end{cases} \quad (3-50)$$

A sequência de amostras de X geradas através do algoritmo de Metropolis-Hastings formam uma cadeia de Markov. Cada amostra $u^{(i)}$ é aceita como uma nova amostra $x^{(i+1)}$ dependendo apenas do valor de $x^{(i)}$ (e não das amostras anteriores).

Quando a razão de Metropolis é maior ou igual a um, a função $\alpha(x^{(i)}, u^{(i)}) = 1$ e, assim o valor $u^{(i)}$ é sempre aceito. Entretanto, se $R < 1$, a função $\alpha(x^{(i)}, u^{(i)}) = R(x^{(i)}, u^{(i)})$ e, o valor proposto $u^{(i)}$ é aceito com probabilidade $\alpha = R$. Dessa forma, a probabilidade de transição do estado $x^{(i)}$ para $u^{(i)}$ é calculado pela multiplicação da probabilidade de gerar $u^{(i)}$ (isto é, $h(u^{(i)})$) com a probabilidade de aceitá-lo como $x^{(i+1)}$ (isto é, $\alpha(x^{(i)}, u^{(i)})$).

$$T_{x^{(i)}, u^{(i)}} = h(u^{(i)}) \alpha(x^{(i)}, u^{(i)}) \quad (3-51)$$

O teorema da convergência da distribuição de probabilidade de uma cadeia de Markov, enunciado na seção anterior da dissertação, garante que se um número suficiente de amostras de uma cadeia irredutível e aperiódica for gerada, a distribuição de probabilidade da cadeia π converge para a distribuição estacionária π^* .

Assim para demonstrar que as amostras geradas através do algoritmo de Metropolis-Hastings estão de acordo com a distribuição de probabilidade p , é necessário mostrar que elas convergem para a distribuição estacionária π^* da cadeia.

Porém, como uma distribuição de probabilidade estacionária atende sempre a condição de reversibilidade, basta provar que a equação (3-26) é satisfeita, ou seja, que:

$$p(x^{(i)}) T_{x^{(i)}, u^{(i)}} = p(u^{(i)}) T_{u^{(i)}, x^{(i)}} \quad (3-52)$$

Com efeito, combinando as equações (3-52), (3-49) e (3-48):

$$\begin{aligned} p(x^{(i)}) T_{x^{(i)}, u^{(i)}} &= p(x^{(i)}) h(u^{(i)}) \alpha(x^{(i)}, u^{(i)}) \\ &= f(x^{(i)}) h(u^{(i)}) \min [1, R(x^{(i)}, u^{(i)})] \\ &= \min [f(u^{(i)}) h(u^{(i)}), f(x^{(i)}) h(x^{(i)})] \\ &= \min [f(u^{(i)}) h(u^{(i)}), f(u^{(i)}) h(u^{(i)}) / R(x^{(i)}, u^{(i)})] \\ &= f(u^{(i)}) h(u^{(i)}) \min [1, 1/R(x^{(i)}, u^{(i)})] . \end{aligned} \quad (3-53)$$

Como $1/R(x^{(i)}, u^{(i)}) = R(u^{(i)}, x^{(i)})$, tem-se que:

$$p(x^{(i)}) T_{x^{(i)}, u^{(i)}} = p(u^{(i)}) h(u^{(i)}) \alpha(u^{(i)}, x^{(i)}) = p(u^{(i)}) T_{u^{(i)}, x^{(i)}} . \quad (3-54)$$

Exemplo 3.3.1 (MCMC). Para exemplificar o uso do método de Monte Carlo com Cadeia de Markov, suponha que deseja-se calcular a seguinte integral:

$$I = \int_0^{\infty} (x - \alpha\beta)^2 p(x) dx \quad (3-55)$$

onde p é a função densidade de probabilidade Gamma (3-56) e Γ a função Gamma.

$$p(x) = \frac{1}{\beta\Gamma(\alpha)} \left(\frac{x}{\alpha}\right)^{\alpha-1} \exp -\frac{x}{\beta} \quad (3-56)$$

Essa integral (3-55) representa a variância da função densidade de probabilidade Gamma e vale $I = \alpha\beta^2$. Para calculá-la através do MCMC, utilizou-se a seguinte distribuição proposta h :

$$h(u) = 0,5 \exp -0,5 u \quad u \geq 0. \quad (3-57)$$

Supôs-se como valor inicial da cadeia $x^{(1)} = 1$. Como parâmetros da distribuição Gamma supôs-se: $\alpha = 2$ e $\beta = 1$. As amostras $u^{(1)}, u^{(2)}, \dots$ com distribuição de probabilidade h são geradas por $u^{(i)} = -2 \ln(v^{(i)})$, onde cada $v^{(i)}$ é sorteado da distribuição uniforme no intervalo $[0, 1]$.

Nesse exemplo, a razão de Metropolis vale:

$$R = \frac{f(u^{(i)}) h(x^{(i)})}{f(x^{(i)}) h(u^{(i)})} = \left(\frac{u^{(i)}}{x^{(i)}}\right)^{\alpha-1} \exp \frac{u^{(i)} - x^{(i)}}{\beta} \exp -0,5(x^{(i)} - u^{(i)}) \quad (3-58)$$

e a atualização das amostras, segundo o algoritmo de Metropolis-Hastings, é feita da seguinte forma:

$$x^{(i+1)} = \begin{cases} u^{(i)} & \text{se } \rho^{(i)} \leq \min[1, R] \\ x^{(i)} & \text{se } \rho^{(i)} > \min[1, R]. \end{cases} \quad (3-59)$$

onde $\rho^{(i)}$ é sorteado da distribuição uniforme no intervalo $[0, 1]$.

A partir das N amostras geradas $x^{(1)}, x^{(2)}, \dots, x^{(N)}$, a integral (3-55) é estimada pela expressão (3-60) e o desvio padrão por (3-61).

$$I_N = \frac{1}{N} \sum_{i=1}^N (x^{(i)} - \alpha\beta)^2 = \frac{1}{N} \sum_{i=1}^N (x^{(i)} - 2)^2 \quad (3-60)$$

$$\sigma(I_N) = \frac{1}{N} \sqrt{\sum_{i=1}^N \left((x^{(i)} - \alpha\beta)^2 - I_N \right)^2} \quad (3-61)$$

Duas rotinas MATLAB foram desenvolvidas na implementação do algoritmo de Metropolis-Hastings nesse exemplo: `MCMC_VARIAVELALEATORIA` e `MCMC_INTEGRALGAMMA`.

Com o objetivo de avaliar a qualidade dos resultados fornecidos pelo MCMC, a integral (3-55) também foi estimada a partir de amostras geradas com a função $random('gam', \alpha, \beta)$ do MATLAB.

A tabela (3.1) mostra os resultados obtidos para diferentes números de amostras (N).

N	MCMC $I_N \pm \sigma(I_N)$	Função $random('gam', \alpha, \beta)$ $I_N \pm \sigma(I_N)$
10^4	1.9635 ± 0.0440	2.0672 ± 0.0477
10^5	1.9871 ± 0.0140	2.0063 ± 0.0144
10^6	1.9885 ± 0.0044	1.9982 ± 0.0044
10^7	1.9979 ± 0.0014	2.0001 ± 0.0014

Tabela 3.1: Estimativas da variância da função densidade de probabilidade gamma ($\alpha = 2$ e $\beta = 1$)

Tanto o método de MCMC quanto a função do MATLAB $random('gam', \alpha, \beta)$ fornecem resultados que convergem para a resposta correta $I = 2$. Em ambos os casos, quanto maior o número de amostras menor é a variância. A figura (3.4) mostra semelhança de formato entre o histograma das amostras obtidas por MCMC (10^7 amostras) e a função densidade de probabilidade gamma (3-56).

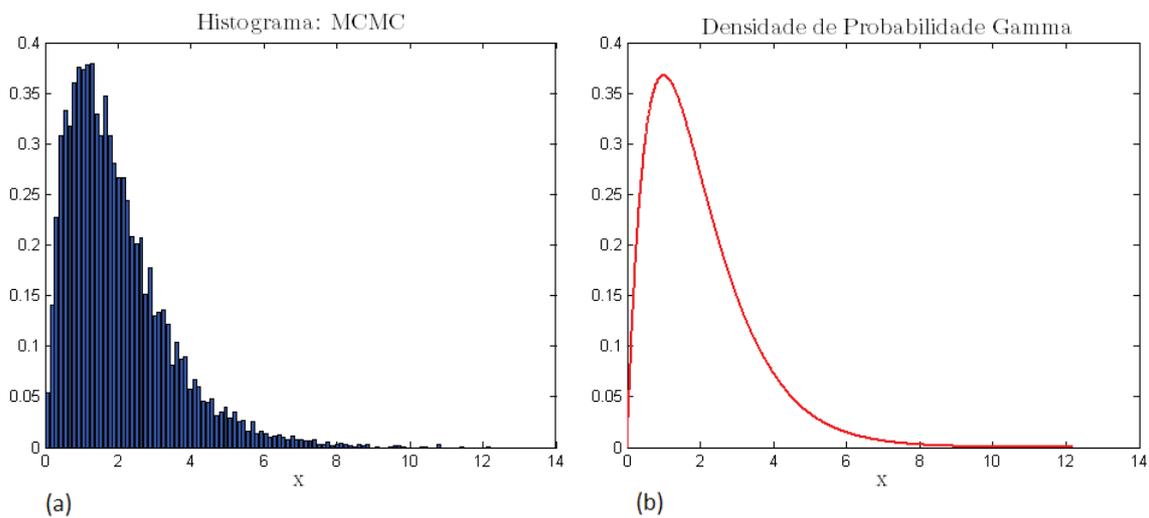


Figura 3.4: (a) Histograma das amostras obtidas por MCMC (10^4 amostras). (b) Função densidade de probabilidade gamma ($\sigma^2 = \alpha\beta^2$).

3.3.1

Aquecimento da Cadeia de Markov

Muitos livros recomendam descartar os primeiros elementos da sequência gerada pelo algoritmo de Metropolis-Hastings. Esses elementos constituem o período de aquecimento ("burn-in") da cadeia de Markov e, representam o período em que a cadeia ainda não convergiu para a distribuição de probabilidade estacionária.

Desprezá-los é uma opção de anular o efeito que eles geram sobre a distribuição das amostras obtidas com o MCMC, fazendo com que essa distribuição fique mais próxima da distribuição desejada.

Entretanto, quando o número de amostras N é suficientemente grande, esse período de aquecimento pode ter pouca influência sobre a distribuição das amostras obtidas. Podendo assim ser desprezado.

3.3.2

Distribuição de Probabilidade Proposta

No algoritmo de Metropolis-Hastings é necessário definir uma variável aleatória Y com função densidade de probabilidade h e gerar amostras u_i de Y . Como essa distribuição é escolhida arbitrariamente, é costume utilizar-se uma distribuição simples e de fácil amostragem (como por exemplo a uniforme ou a exponencial). Para esta tarefa pode-se por exemplo utilizar o Método da Transformada Inversa apresentado no primeiro capítulo.

Em princípio, a definição de h é totalmente livre, porém é usual escolher-se a distribuição h com uma forma semelhante a distribuição da qual deseja-se gerar amostras, p [18]. Essa escolha faz com que a razão de Metrópolis, R , assumam na maioria das iterações do algoritmo um valor próximo a 1, e conseqüentemente mais rapidamente as amostras geradas abrangeriam todo o espaço amostral.

Dessa forma, uma escolha apropriada de h pode reduzir o número de iterações necessárias para obter-se uma distribuição de amostras próximas a p .

O termo mistura ("mixing") é utilizado para representar a taxa em que o espaço amostral é explorado. Uma cadeia é dita ser mal misturada ("poorly mixing") se com um elevado número de amostras apenas uma pequena região do espaço amostral é ocupada. Caso contrário, é dita ser bem misturada ("well mixing") [18]. Escolhendo-se a função $h = f$, a razão R será sempre igual a 1 e, dessa forma, a cadeia teria uma mistura perfeita. Porém, para essa escolha ser possível, gerar amostras de p deveria ser uma tarefa simples (o que conseqüentemente não justificaria o uso do algoritmo de Metropolis-Hastings).

No próximo exemplo é mostrado a influência da escolha da distribuição proposta, h , na taxa de mistura da cadeia.

Exemplo 3.3.2 (MCMC). Suponha que se deseja gerar amostras de uma variável aleatória X com a seguinte distribuição de probabilidade Qui-Quadrado Inversa Escalada p :

$$p(x) = C x^{-n/2} \exp \frac{-a}{2x}. \tag{3-62}$$

O parâmetro a vale 4, $n = 5$ e C é uma constante de normalização. Supõe-se que o valor inicial da cadeia é $x^{(1)} = 1$.

Como distribuição proposta h , escolhe-se inicialmente a distribuição uniforme no intervalo $[0, 100]$. Posteriormente, escolhe-se a distribuição qui-quadrado \mathcal{X}_1^2 .

$$h(u) = \frac{1}{2^{1/2}\Gamma(1/2)} u^{-1/2} e^{-u/2} \tag{3-63}$$

A rotina MCMC_MISTURACADEIA implementada em MATLAB testou o algoritmo de Metrópolis-Hastings para as duas distribuições propostas. Na figura (3.5) é possível comparar as primeiras 500 amostras para as duas funções h . No caso da distribuição uniforme, a cadeia pode ser considerada mal misturada ("poorly mixing") (há uma grande repetição sequencial de elementos), enquanto no caso da distribuição chi-quadrado (com forma mais semelhante a função p), é considerada bem misturada ("well mixing").

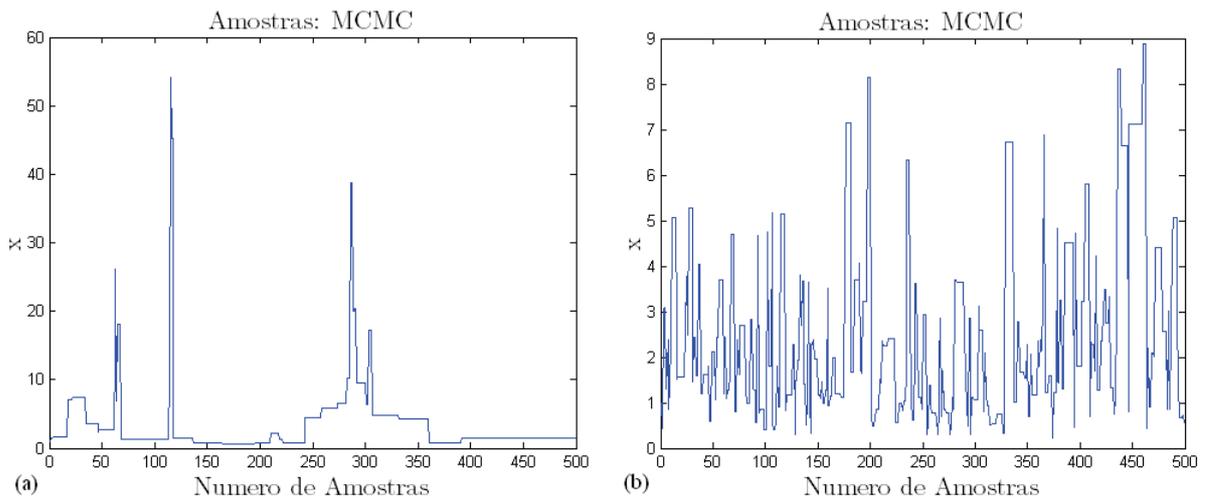


Figura 3.5: Amostras geradas de X obtidas utilizando-se (a) h uniforme e (b) h qui-quadrado.

Exemplo 3.3.3 (MCMC). Suponha que se deseja gerar amostras de uma variável aleatória X com a seguinte distribuição de probabilidade Gamma p :

$$p(x) = \frac{1}{\underline{x}} \left(\frac{1}{\delta_X^2} \right)^{\frac{1}{\delta_X^2}} \frac{1}{\Gamma(1/\delta_X^2)} \left(\frac{x}{\underline{x}} \right)^{\frac{1}{\delta_X^2} - 1} \exp \left(-\frac{x}{\delta_X^2 \underline{x}} \right). \quad (3-64)$$

Observe que (3-64) é uma forma alternativa de se escrever a distribuição Gamma, onde \underline{x} é a média de X , δ_X o coeficiente de variação ($\delta_X = \sigma_X/\underline{x}$) e σ_X o desvio padrão.

Como distribuições propostas para o MCMC, duas funções, h_1 e h_2 , de suporte positivo $(0, +\infty)$ foram analisadas. A primeira, (3-65), tem uma forma totalmente diferente da distribuição de probabilidade p , enquanto a segunda, (3-66), possui uma forma similar. Os gráficos de p , h_1 e h_2 são mostrados na figura (3.6).

$$h_1(u) = \lambda e^{-\lambda u} \quad (3-65)$$

$$h_2(u) = \frac{\alpha}{\beta} \left(\frac{u}{\beta} \right)^{\alpha-1} e^{-(u/\beta)^\alpha} \quad (3-66)$$

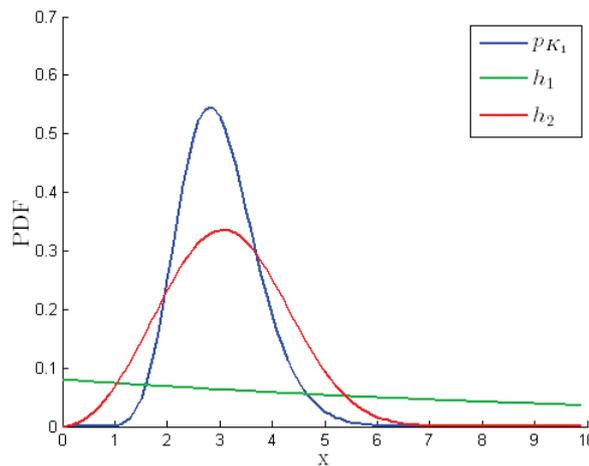


Figura 3.6: Função densidade de probabilidade p (com $\underline{x} = 3.0$, $\delta_X = 0.25$ e $\sigma_X = 0.75$) e as duas distribuições propostas: h_1 (com $\lambda = 0.08$) e h_2 (com $\alpha = 3.0$ e $\beta = 3.5$).

O algoritmo de Metropolis-Hastings foi aplicado para gerar amostras de X usando cada uma das distribuições propostas, h_1 e h_2 , e um valor inicial da cadeia de $x^{(1)} = \underline{x} = 3.0$. Os resultados obtidos são mostrados na figura (3.7). Para as duas funções propostas, é possível notar que 5×10^3 amostras são insuficientes para reproduzir p . Porém, quando esse número cresce para 5×10^4 , ambas as funções, h_1 e h_2 , apresentam bons resultados, embora o histograma gerado com h_2 se assemelha mais com a função de densidade de probabilidade desejada p .

Com este exemplo, verifica-se que a escolha de uma distribuição proposta de h que melhor aproxima a forma de p , reduz o número de amostras necessárias para o MCMC apresentar um bom desempenho. No entanto, é importante

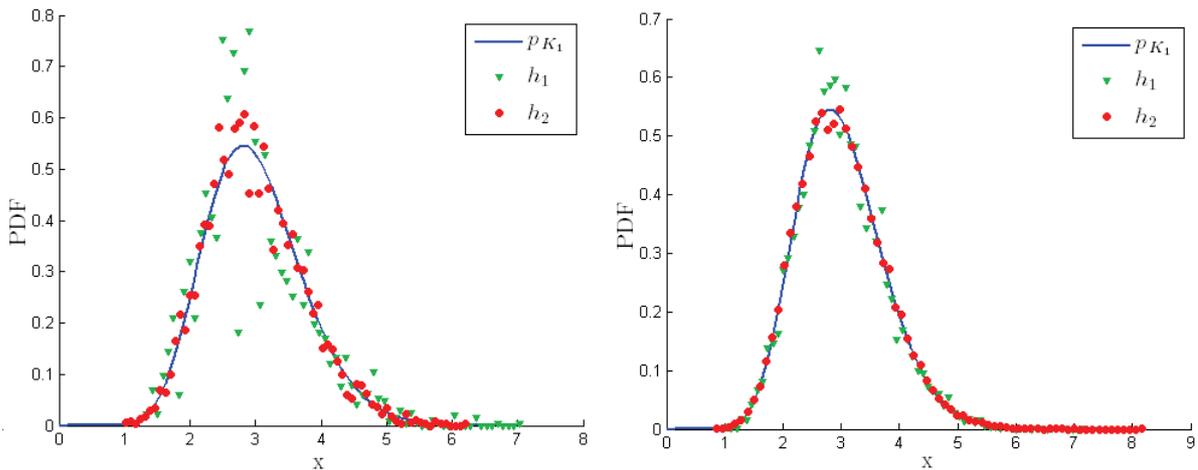


Figura 3.7: Gráfico de frequência de (a) 5×10^3 e (b) 5×10^4 amostras de X .

notar que se um número suficiente de amostras é gerado, essencialmente qualquer distribuição de proposta h pode ser eficaz.

Como a gama função densidade de probabilidade é muito usual, muitos softwares já desenvolvidos incluem em suas bibliotecas comandos capazes de gerar amostras segundo essa distribuição. No MATLAB, esta função é chamada de *gamrnd*. Comparando na tabela (3.2) as estimativas de média e variância de X calculadas a partir de amostras do MCMC e *gamrnd*, observamos que ambos os métodos apresentam resultados semelhantes, mesmo quando o número de amostras é pequeno.

Tabela 3.2: estimativas de média e variância de X calculadas a partir das amostras geradas com MCMC e com o comando *gamrnd* do MATLAB.

N	h_1	h_2	MATLAB
	$\bar{x} \pm \sigma_X$	$\bar{x} \pm \sigma_X$	$\bar{x} \pm \sigma_X$
10^3	3.0583 ± 0.7013	3.0252 ± 0.7550	3.0080 ± 0.7548
10^4	2.9714 ± 0.7519	2.9957 ± 0.7439	3.0067 ± 0.7487
10^5	2.9983 ± 0.7454	2.9998 ± 0.7523	3.0008 ± 0.7495

A influência de h_1 e h_2 na sequência de amostras de MCMC pode ser observado na figura (3.8). Note-se que com h_2 , a sequência de valores de $x^{(i)}$ apresenta poucos valores repetidos, formando Cadeia de Markov bem misturada. Com h_1 , o oposto ocorre (tem muitos períodos de valores repetidos), formando uma Cadeia de Markov mau misturada.

3.4

O Algoritmo de Metropolis-Hastings Aplicado a Vetores Aleatórios

Nesta seção do trabalho é mostrado uma extensão do algoritmo de Metropolis-Hastings para casos em que se deseja gerar amostras de vetores

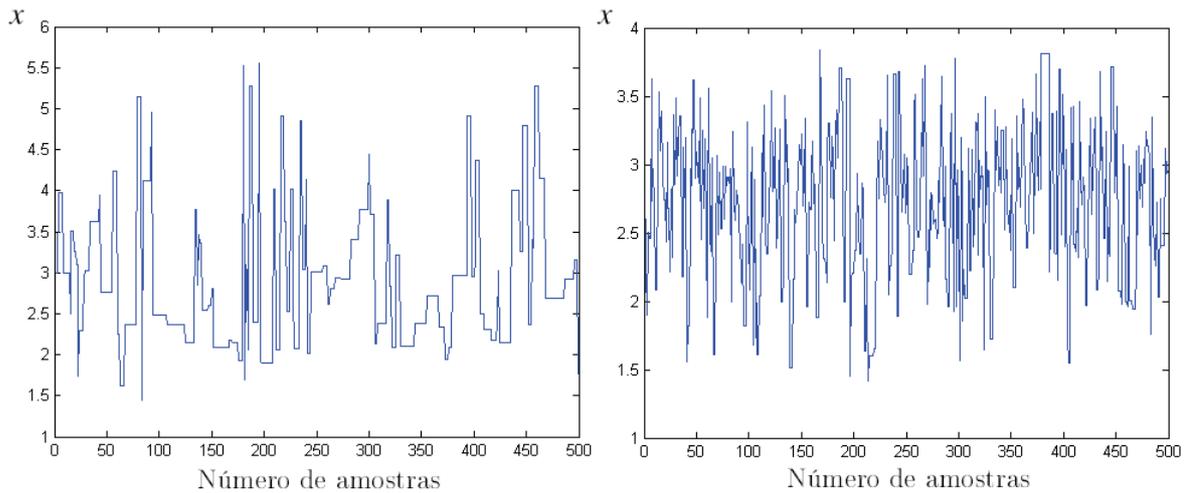


Figura 3.8: Primeiras 500 amostras da sequência $x^{(i)}$ obtidas pela distribuição proposta (a) h_1 e (b) h_2 .

aleatórios.

Um vetor aleatório real de dimensão n é definido como uma função real, $\{X\}$, que associa a cada ponto amostra w , pertencente ao espaço de amostras Ω , um ponto do espaço n -dimensional, $\{x(w)\} \in \mathbb{R}^n$ [27]. Pode-se escrever:

$$\begin{aligned} \{X\} : \quad \Omega &\longmapsto \mathbb{R}^n \\ w &\longmapsto \{x(w)\}. \end{aligned} \tag{3-67}$$

O mapa representado em (3-67) é esquematizado na figura (3.9) para o caso em que $n = 3$.

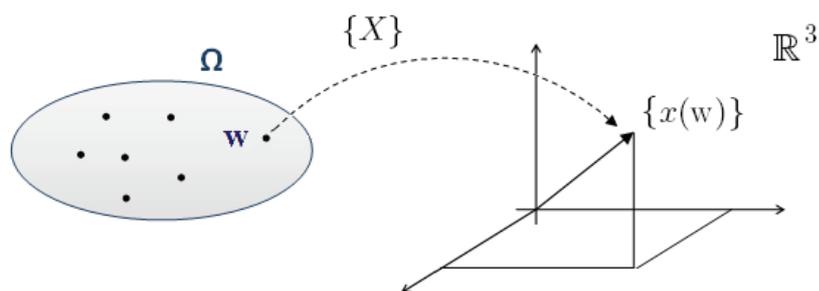


Figura 3.9: Vetor aleatório (mapa de Ω para \mathbb{R}^3).

Assim, um vetor aleatório $\{X\}$ com dimensão n é um vetor formado por n variáveis aleatórias:

$$\{X\} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}. \quad (3-68)$$

A função distribuição de probabilidade associada a um vetor aleatório $\{X\}$ é a função:

$$\begin{aligned} P : \quad \mathbb{R}^n &\longmapsto [0, 1] \\ \{x\} &\longmapsto P(\{x\}) \end{aligned} \quad (3-69)$$

onde

$$P(\{x\}) = Pr(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n). \quad (3-70)$$

A função densidade de probabilidade, p , do vetor aleatório $\{X\}$ é:

$$p(\{x\}) = \frac{\partial^n}{\partial X_1 \partial X_2 \dots \partial X_n} P(\{x\}), \quad (3-71)$$

chamado também de densidade de probabilidade conjunta das variáveis aleatórias X_1, \dots, X_n .

Cada amostra de $\{X\}$ é um vetor de dimensão n :

$$\{x\}^{(i)} = \begin{pmatrix} x_1^{(i)} \\ \vdots \\ x_n^{(i)} \end{pmatrix} \quad (3-72)$$

Assim como no caso de geração de amostras de variáveis aleatórias, o Algoritmo de Metropolis-Hastings para gerar amostras de vetores aleatórios baseia-se no mesmo critério de rejeição ou aceitação de amostras geradas por uma distribuição de probabilidade proposta h . Porém, essa distribuição pode ser:

- conjunta, de forma que, a cada iteração do algoritmo, é gerada um vetor $u^{(i)}$ de dimensão n ;
- não conjunta. Nesse caso, n funções $h_{x_1}, h_{x_2}, \dots, h_{x_n}$ são utilizadas (uma para cada componente de $\{X\}$). A cada iteração, é escolhida uma componente j para ser gerada segundo h_{x_j} . Assim, o vetor $\{u\}^{(u+1)}$ é dado por (3-73). A escolha da componente a ser gerada pode ser de forma sequencial ou aleatória.

$$\{u\}^{(i+1)} = [u_1^{(i)} \quad u_2^{(i)} \quad \dots \quad u_j^{(i+1)} \quad \dots \quad u_n^{(i)}] \quad (3-73)$$

Exemplo 3.4.1 (MCMC aplicado a vetores aleatórios). Suponha que deseje-se gerar amostras do vetor aleatório $\{X\} = [X_1 \ X_2]^T$ com a seguinte função densidade de probabilidade conjunta Gaussiana mostrada na figura (3.10) com média $m_X = \begin{pmatrix} 3 \\ 3 \end{pmatrix}$ e matriz de covariância $C = \begin{pmatrix} 1.0 & 0.1 \\ 0.1 & 1.0 \end{pmatrix}$:

$$p(x_1, x_2) = \frac{1}{\sqrt{(2\pi)^n \det[C]}} \exp \left\{ -\frac{1}{2} \langle [C]^{-1}(x - m_X); (x - m_X) \rangle_{\mathbb{R}^2} \right\} \quad (3-74)$$

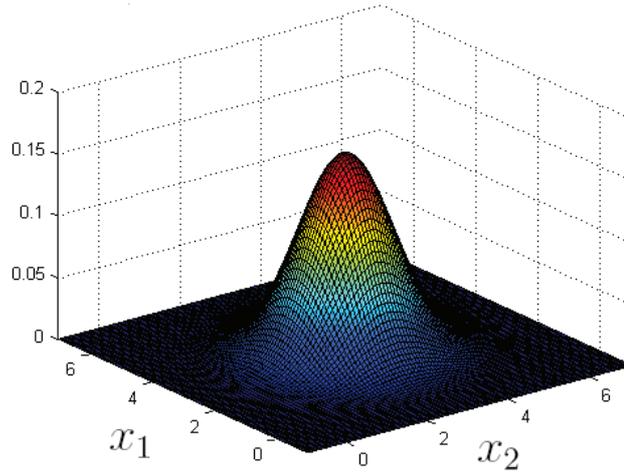


Figura 3.10: Densidade de probabilidade Gaussiana.

Como distribuições de probabilidade propostas, duas distribuições (uma para componente de $\{X\}$) são utilizadas. Assim:

$$h([u_1 \ u_2]) = h_1(u_1) \times h_2(u_2) \quad (3-75)$$

Cada função, h_1 e h_2 , está definida no domínio $(-\infty, \infty)$. São dadas por:

$$h_1(u_1) = \mathcal{N}(0, 1) = \frac{e^{-u_1^2/2}}{\sqrt{2\pi}} \quad (3-76)$$

$$h_2(u_2) = \frac{0.1}{2} e^{-0.1|u_2|} \quad (3-77)$$

Para gerar amostras $u_2^{(i)}$ segundo h_2 , um número $\beta^{(i)}$ é sorteado com distribuição de probabilidade uniforme unitária no intervalo $[0, 1]$ e faz-se:

$$u_2^{(i)} = \begin{cases} \ln(2\beta^{(i)})/0.1 & \text{se } \beta^{(i)} < 0.5 \\ 0 & \text{se } \beta^{(i)} = 0.5 \\ \ln(2 - 2\beta^{(i)})/0.1 & \text{se } \beta^{(i)} > 0.5 \end{cases} \quad (3-78)$$

O algoritmo de Metropolis-Hastings multidimensional foi implementado em MATLAB através da rotina MCMC_VETORALEATORIO. A Figura (3.11) mostra os histogramas obtidos com dois número de amostras. É possível notar que 10^4 amostras não são insuficientes para construir um histograma similar

à densidade p . No entanto, quando este número cresce para 10^6 , o histograma torna-se representativo. Como no exemplo anterior, o método MCMC foi comparado com a função do MATLAB `mvrnd` capaz de gerar amostras de um vetor Gaussiano. Estimativas do vetor de médias e matriz de covariância para diferentes números de amostras são apresentados na tabela (3.3). Nas simulações, considerou-se a média $m_X = \begin{pmatrix} 3 \\ 3 \end{pmatrix}$ e a matriz de covariância $C = \begin{pmatrix} 1.0 & 0.1 \\ 0.1 & 1.0 \end{pmatrix}$.

Tabela 3.3: Estimativas de m_X e C .

N	MCMC		MATLAB	
	m_X	C	m_X	C
10^5	$\begin{pmatrix} 3.0017 \\ 2.9916 \end{pmatrix}$	$\begin{pmatrix} 1.0022 & 0.0961 \\ 0.0961 & 1.0175 \end{pmatrix}$	$\begin{pmatrix} 3.0019 \\ 3.0054 \end{pmatrix}$	$\begin{pmatrix} 1.0054 & 0.1043 \\ 0.1043 & 0.9946 \end{pmatrix}$
10^6	$\begin{pmatrix} 2.9985 \\ 3.0030 \end{pmatrix}$	$\begin{pmatrix} 0.9977 & 0.1019 \\ 0.1019 & 1.0027 \end{pmatrix}$	$\begin{pmatrix} 3.0004 \\ 2.9995 \end{pmatrix}$	$\begin{pmatrix} 1.0002 & 0.1003 \\ 0.1003 & 1.0015 \end{pmatrix}$

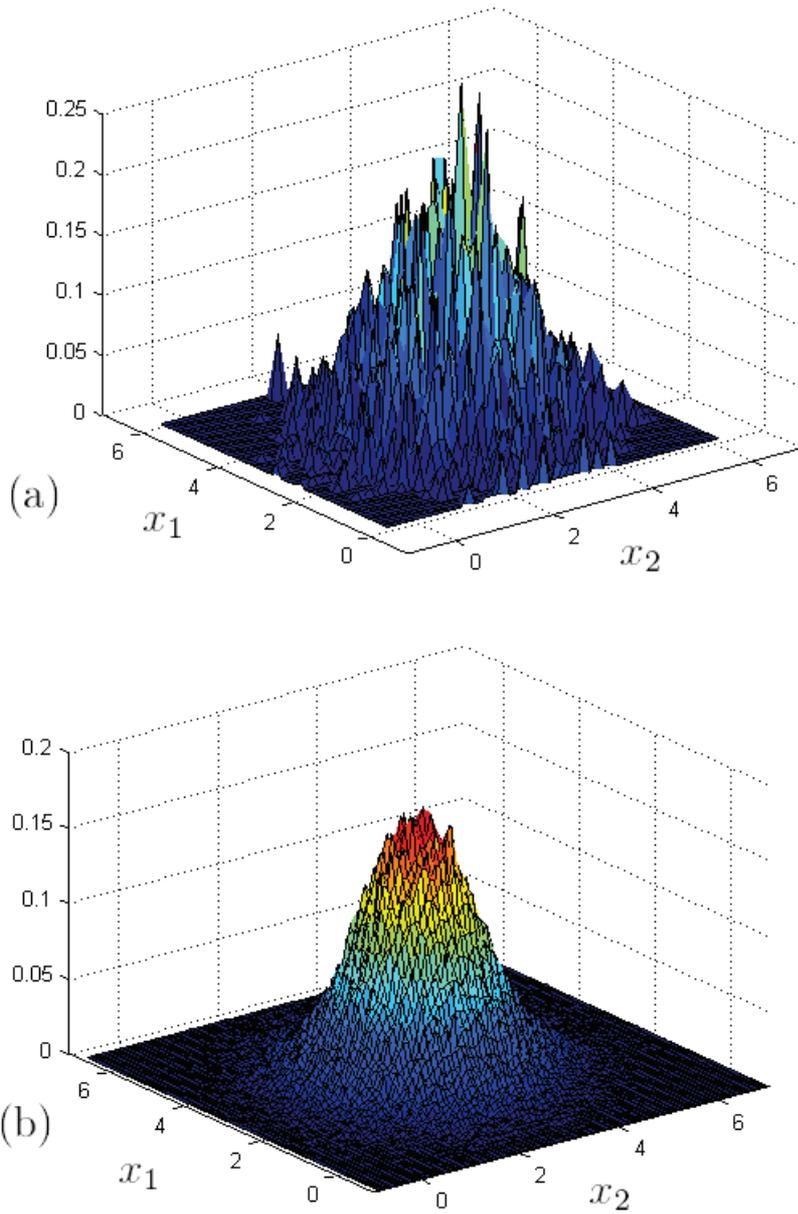


Figura 3.11: Histograma para uma distribuição de probabilidade de um vetor aleatório de duas dimensões (a) 10^4 e (b) 10^6 amostras.