

Volume 70, 2012

Editores

Cassio Machiaveli Oishi

Universidade Estadual Paulista - UNESP
Presidente Prudente, SP, Brasil

Fernando Rodrigo Rafaeli

Universidade Estadual Paulista - UNESP
São José do Rio Preto, SP, Brasil

Rosana Sueli da Motta Jafelice (Editor Chefe)

Universidade Federal de Uberlândia - UFU
Uberlândia, MG, Brasil

Rubens de Figueiredo Camargo

Universidade Estadual Paulista - UNESP
Bauru, SP, Brasil

Sezimária de Fátima P. Saramago

Universidade Federal de Uberlândia - UFU
Uberlândia, MG, Brasil

Vanessa Avansini Botta Pirani (Editor Adjunto)

Universidade Estadual Paulista - UNESP
Presidente Prudente, SP, Brasil



A Sociedade Brasileira de Matemática Aplicada e Computacional - SBMAC publica, desde as primeiras edições do evento, monografias dos cursos que são ministrados nos CNMAC.

Para a comemoração dos 25 anos da SBMAC, que ocorreu durante o XXVI CNMAC em 2003, foi criada a série **Notas em Matemática Aplicada** para publicar as monografias dos minicursos ministrados nos CNMAC, o que permaneceu até o XXXIII CNMAC em 2010.

A partir de 2011, a série passa a publicar, também, livros nas áreas de interesse da SBMAC. Os autores que submeterem textos à série Notas em Matemática Aplicada devem estar cientes de que poderão ser convidados a ministrarem minicursos nos eventos patrocinados pela SBMAC, em especial nos CNMAC, sobre assunto a que se refere o texto.

O livro deve ser preparado em **Latex (compatível com o Miktex versão 2.7)**, **as figuras em eps** e deve ter entre **80 e 150 páginas**. O texto deve ser redigido de forma clara, acompanhado de uma excelente revisão bibliográfica e de **exercícios de verificação de aprendizagem** ao final de cada capítulo.

Veja todos os títulos publicados nesta série na página
http://www.sbmac.org.br/p_notas.php



Sociedade Brasileira de Matemática Aplicada e Computacional

2012

MODELAGEM ESTOCÁSTICA E GERAÇÃO DE AMOSTRAS DE VARIÁVEIS E VETORES ALEATÓRIOS

Rubens Sampaio
rsampaio@puc-rio.br

Departamento de Engenharia Mecânica
Pontifícia Universidade Católica do Rio de Janeiro

Roberta de Queiroz Lima
roberta_10_lima@hotmail.com

Departamento de Engenharia Mecânica
Pontifícia Universidade Católica do Rio de Janeiro



Sociedade Brasileira de Matemática Aplicada e Computacional

São Carlos - SP, Brasil
2012

Coordenação Editorial: Elbert Einstein Nehrer Macau

Coordenação Editorial da Série: Rosana Sueli da Motta Jafelice

Editora: SBMAC

Capa: Matheus Botossi Trindade

Patrocínio: SBMAC

Copyright ©2012 by Rubens Sampaio e Roberta Lima. Direitos reservados, 2012 pela SBMAC. A publicação nesta série não impede o autor de publicar parte ou a totalidade da obra por outra editora, em qualquer meio, desde que faça citação à edição original.

**Catálogo elaborado pela Biblioteca do IBILCE/UNESP
Bibliotecária: Maria Luiza Fernandes Jardim Froner**

Sampaio, Rubens

Modelagem Estocástica e Geração de Amostras de Variáveis
e Vetores Aleatórios - São Carlos, SP :

SBMAC, 2012, 144 p., 21.5 cm - (Notas em Matemática
Aplicada; v. 70)

e-ISBN 978-85-8215-031-3

1. Modelagem Estocástica 2. Método de Monte Carlo
3. Geração

I. Sampaio, Rubens II. Lima, Roberta III. Título. IV. Série

CDD - 51

Agradecimentos

Os autores agradecem ao Projeto de Cooperação Internacional Brasil-França, *Capes-Cofecub*, número 672/10, ao *Projeto Pro-Engenharia*, número 086/2008, ao *CNPq*, e a *FAPERJ* pelo suporte que possibilitou a preparação dessas Notas.

Conteúdo

Prefácio	11
1 Variáveis Aleatórias, Vetores Aleatórios e Processos Estocásticos	13
1.1 Variáveis aleatórias	13
1.1.1 Variáveis aleatórias discretas	14
1.1.2 Variáveis aleatórias contínuas	17
1.1.3 Propriedades hilbertianas	20
1.1.4 Sequências de variáveis aleatórias	21
1.2 Vetores aleatórios	24
1.3 Processos estocásticos	29
1.4 Exercícios	30
2 Máximo de Entropia e Informação	33
2.1 Construção do modelo estocástico	33
2.2 Princípio da Entropia Máxima	35
2.2.1 Variáveis aleatórias discretas	35
2.2.2 Variáveis aleatórias contínuas	39
2.2.3 Vetores aleatórios	46
2.2.4 Matrizes aleatórias	49
2.3 Exercícios	52
3 Método de Monte Carlo e Geração de Amostras de Variáveis e Vetores Aleatórios	55
3.1 Método de Monte Carlo	55
3.1.1 Lei dos grandes números	57
3.1.2 Teorema do limite central	58
3.1.3 Integração com o método de Monte Carlo	61
3.2 Geração de amostras de variáveis e vetores aleatórios	69
3.2.1 Geradores baseados em congruência linear	71
3.2.2 Parâmetros do gerador baseado em congruência linear	72

3.2.3	Outros geradores de amostras de variáveis aleatórias	74
3.3	Método da Transformada Inversa	75
3.4	Geração de amostras de variáveis aleatórias gaussianas	78
3.5	Método da Rejeição	81
3.5.1	Método da Rejeição aplicado a Variáveis Aleatórias	81
3.5.2	Método da Rejeição aplicado a Vetores Aleatórios	86
3.6	Exercícios	89
4	Método de Monte Carlo com Cadeia de Markov	91
4.1	Processos estocásticos e cadeias de Markov	92
4.2	Algoritmo de Metropolis-Hastings aplicado a variáveis aleatórias	100
4.2.1	Aquecimento da cadeia de Markov	104
4.2.2	Densidade de probabilidade proposta	104
4.3	Algoritmo de Metropolis-Hastings aplicado a vetores aleatórios	108
4.4	Exercícios	113
5	Expansão de Karhunen-Loève	115
5.1	Expansão de Karhunen-Loève	115
5.1.1	Expansão de KL aplicada a processos estocásticos gaussianos	118
5.1.2	Exemplos numéricos da expansão de KL	118
5.1.3	O processo de discretização	118
5.1.4	Estimativas dos momentos de um processo estocástico	119
5.1.5	Estimativas dos autovalores e autofunções da expansão de KL	121
5.1.6	Estimativas das variáveis aleatórias da expansão de KL	124
5.2	Geração de realizações processos estocásticos a partir da expansão de KL	125
5.3	Exercícios	129
6	Propagação de Incertezas	131
6.1	Propagação de incertezas através do método de Monte Carlo	131
6.2	Exercícios	137
	Bibliografia	139

Prefácio

Modelagem estocástica começa com o estudo de jogos de azar (dados, baralho, roleta, etc). Os egípcios já jogavam dados. Porém nesse contexto modelagem estocástica se confunde com combinatória e não é nisso que estamos interessados. Nos interessa modelos contínuos.

Modelagem estocástica é mais difícil do que modelagem determinística, e a implementação dos modelos é muito mais custosa. Vejamos um pequeno exemplo: no caso mais simples de modelagem determinística os parâmetros do modelo são constantes. Seu equivalente em modelagem estocástica é transformar essas constantes em variáveis aleatórias. Para cada realização da variável aleatória o seu valor é uma constante, mas essa constante pode ser diferente para outras realizações e as realizações seguem uma distribuição de probabilidade. Ou seja, os objetos mais simples de um modelo estocástico, as variáveis aleatórias, são funções, portanto objetos de dimensão infinita. Em um modelo determinístico, se ele é coerente, temos uma solução uma vez fixadas condições suplementares. Em um modelo estocástico não é assim. Para cada realização dos parâmetros do modelo temos um modelo determinístico que tem uma solução, e temos que fazer estatísticas dessas soluções para sabermos como elas se distribuem; o que interessa em modelagem estocástica é a distribuição de probabilidade. Os valores obtidos para uma realização não representam o comportamento de um sistema estocástico. Para que verdades desse tipo fossem absorvidas pela indústria foram necessários anos de *controle de qualidade*, que não funcionava, até que a modelagem estocástica fosse promovida a ferramenta essencial de engenharia. Daí o grande interesse atual em fiabilidade, que representa uma nova maneira de projetar sistemas levando em conta incertezas.

Ou seja, podemos pensar que a solução de um modelo probabilístico pode ser decomposta em três etapas: geração dos objetos estocásticos segundo as distribuições de probabilidade, solução de um modelo determinístico associado a cada realização e, finalmente, fazer estatísticas das soluções encontradas (por exemplo, construindo histogramas). Os histogramas representam a solução de um modelo estocástico. Quando se deseja uma aproximação da solução, muitas vezes, se pode

fornecer, por exemplo, a média e a variância da distribuição de probabilidade (histograma). Em alguns casos se calcula apenas uma probabilidade, pois os momentos são muito complicados de calcular.

Essas Notas visam divulgar a Modelagem Estocástica e a Quantificação de Incertezas que, cada vez mais, vem crescendo em importância em todas as áreas. Os conceitos principais, desenvolvidos nas Notas, são a construção de um modelo probabilístico e a propagação de incertezas. Na construção de um modelo probabilístico se apresenta o Princípio da Máxima Entropia que ajuda quando se tem pouca informação. A propagação de incertezas é feita resolvendo vários problemas determinísticos para uma escolha randômica de parâmetros seguindo uma distribuição de probabilidade. Ou seja, um problema de geração de variáveis e vetores aleatórios e de processos estocásticos. Uma vez resolvido os problemas se faz o cálculo das estatísticas de interesse, em geral médias e variâncias.

Rio de Janeiro , 01 de agosto de 2012.

Rubens Sampaio
Roberta Lima

Capítulo 1

Variáveis Aleatórias, Vetores Aleatórios e Processos Estocásticos

Variáveis aleatórias, vetores aleatórios e processos estocásticos formam a base do estudo de incertezas. Por isso, antes de abordar especificamente os tópicos de modelagem estocástica e geração, é necessário fazer algumas definições básicas e estabelecer uma notação padrão para todo o texto. O objetivo é facilitar a compreensão dos próximos capítulos.

1.1 Variáveis aleatórias

Considere o espaço de probabilidade $(\Omega, \mathcal{F}, Pr)$ de um determinado experimento. Nesse espaço de probabilidade, Ω é um conjunto não vazio chamado de espaço amostral que contém todas as possibilidades de resultados desse experimento, \mathcal{F} é uma σ -álgebra sobre Ω e Pr é uma medida de probabilidade sobre (Ω, \mathcal{F}) .

Sendo \mathcal{F} uma σ -álgebra sobre Ω , tem-se que \mathcal{F} é uma coleção de subconjuntos do espaço amostral Ω . Cada um desses subconjuntos é chamado de evento e, além disso:

1. \mathcal{F} é não vazio;
2. se $A \in \mathcal{F}$, logo $\Omega \setminus A \in \mathcal{F}$;
3. se $A_1, A_2, \dots \in \mathcal{F}$, logo $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

Uma consequência dos axiomas 1 e 3 é que \mathbf{F} contém o conjunto vazio, \emptyset , e contém o todo o espaço amostral, Ω .

Comentários: No desenvolvimento que se segue se definirá variáveis aleatórias discretas e contínuas, passando ao largo, i. e. sem comentários, de variáveis aleatórias que não sejam nem discretas nem contínuas. O tratamento que será dado é insatisfatório e pode ser sistematizado usando a Teoria da Medida. Porém ela é perfeitamente dispensável para o entendimento das ideias. Na verdade a formalização da Teoria de Probabilidade através da Teoria da Medida só veio muito depois com o trabalho de Kolmogoroff (Cf. [21]).

Em Análise se define os conceitos de espaço topológico, conjunto aberto, e função contínua. O tratamento é semelhante ao que, em parte, fizemos acima com σ -álgebra, evento, função mensurável. Seja $X : \Omega \rightarrow \mathbb{R}$, se A for um aberto qualquer de \mathbb{R} e $X^{-1}(A) \in \mathbf{F}$ dizemos que X é mensurável. Veremos a seguir que variáveis aleatórias são funções mensuráveis.

1.1.1 Variáveis aleatórias discretas

DEFINIÇÃO 1.1.1 (VARIÁVEL ALEATÓRIA DISCRETA)

Uma variável aleatória discreta real X , no espaço de probabilidade (Ω, \mathbf{F}, Pr) , é uma função real que associa a cada ponto amostra w , pertencente ao espaço de amostras Ω , um número $X(w) \in \mathbb{R}$ (Cf.[1, 37, 13]), isto é:

$$\begin{aligned} X : \quad \Omega &\longrightarrow \mathbb{R} \\ w &\longmapsto X(w), \end{aligned} \tag{1.1.1}$$

tal que sua imagem, $X(\Omega)$, é um subconjunto contável de \mathbb{R} e, tal que $X^{-1}(x) = \{w \in \Omega : X(w) = x\} \in \mathbf{F}, \forall x \in \mathbb{R}$.

A notação utilizada para representar variáveis aleatórias é letras maiúsculas. A função de probabilidade, também chamada de função de massa, p_X , de uma variável aleatória discreta, X , é definida pela função:

$$\begin{aligned} p_X : \quad \mathbb{R} &\longrightarrow [0, 1] \\ x &\longmapsto p_X(x) = Pr(\{w \in \Omega : X(w) = x\}) . \end{aligned} \tag{1.1.2}$$

Dessa forma, $p_X(x)$ é a probabilidade de X ter o valor x . Por esse motivo, é usual utilizar a notação simplificada:

$$p_X(x) = Pr(X = x). \tag{1.1.3}$$

PROPOSIÇÃO 1.1.2 . Seja p_X a função de massa de uma variável aleatória discreta X . Então:

$$(i) p_X(x) = 0, \forall x \notin X(\Omega);$$

$$(ii) \sum_{x \in X(\Omega)} p_X(x) = Pr \left(\bigcup_{x \in X(\Omega)} \{w \in \Omega : X(w) = x\} \right) = Pr(\Omega) = 1.$$

Assim, sendo X uma variável aleatória real discreta que assume n valores distintos x_1, \dots, x_n com probabilidades p_1, \dots, p_n , tem-se que:

$$p_i \geq 0 \quad i = 1, \dots, n, \quad (1.1.4)$$

$$\sum_{i=1}^n p_i = 1. \quad (1.1.5)$$

Quando trabalha-se com uma variável aleatória discreta X , uma questão de interesse é determinar a probabilidade de X assumir um valor menor ou igual a $x \in \mathbb{R}$. Ou seja, a probabilidade de X assumir um valor na semi-reta $(-\infty, x]$ para algum $x \in \mathbb{R}$ fixo.

Lembrando da definição de variável aleatória, essa probabilidade existe se e somente se $X^{-1}((-\infty, x]) = \{w \in \Omega : X(w) \leq x\}$ pertence ao espaço de eventos \mathbf{F} . Porém, observa-se que:

$$\begin{aligned} X^{-1}((-\infty, x]) &= \{w \in \Omega : X(w) \leq x\} \\ &= \bigcup_{\substack{y \in X(\Omega) \\ y \leq x}} \{w \in \Omega : X(w) = y\}. \end{aligned} \quad (1.1.6)$$

Ou seja, $X^{-1}((-\infty, x])$ é uma união contável de eventos em \mathbf{F} e, dessa forma pertence a \mathbf{F} , $\forall x \in \mathbb{R}$.

A probabilidade de X assumir um valor menor ou igual a $x \in \mathbb{R}$ pode ser calculada através da função distribuição de probabilidade cumulativa de X , P_X :

$$\begin{aligned} P_X : \quad \mathbb{R} &\longrightarrow [0, 1] \\ x &\longmapsto P_X(x) = Pr(\{w \in \Omega : X(w) \leq x\}). \end{aligned} \quad (1.1.7)$$

É usual utilizar a notação simplificada:

$$P_X(x) = Pr(X \leq x). \quad (1.1.8)$$

PROPOSIÇÃO 1.1.3 *Seja P_X a distribuição de probabilidade cumulativa de X . Então:*

$$(i) 0 \leq P_X(x) \leq 1, \forall x \in \mathbb{R};$$

(ii) P_X é monotonicamente não-decrescente: $a \leq b \implies P(a) \leq P(b)$;

(iii) $Pr(X \geq x) = 1 - P_X(x)$ e $Pr(a \leq X < b) = P_X(b) - P_X(a)$;

(iv) $\lim_{x \rightarrow +\infty} P_X(x) = 1$ e $\lim_{x \rightarrow -\infty} P_X(x) = 0$;

(v) P_X é contínua à esquerda;

(vi) P_X é uma função de variação limitada e sua variação total é igual a 1.

A média, ou valor esperado, de uma variável aleatória real discreta, X , que assume n valores distintos x_1, \dots, x_n com probabilidades p_1, \dots, p_n , é definida por:

$$\mu_X = E[X] = \sum_{i=1}^n x_i p_i. \quad (1.1.9)$$

Dado uma função (mensurável) $h : \mathbb{R} \rightarrow \mathbb{R}$ e uma variável aleatória real discreta, X que assume os valores x_1, \dots, x_n com as respectivas probabilidades p_1, \dots, p_n , então $Y = h(X)$ é também uma variável aleatória real discreta. O valor esperado de Y pode ser calculado por:

$$E[h(X)] = \sum_{i=1}^n h(x_i) p_i. \quad (1.1.10)$$

Uma importante propriedade do operador média é a linearidade, como mostrado a seguir. Sejam X e Y duas variáveis aleatórias reais discretas e sejam α e β constantes reais. Tem-se que:

$$E[\alpha X + \beta Y] = \alpha E[X] + \beta E[Y]. \quad (1.1.11)$$

A variância de X , σ_X^2 , é definida por:

$$\text{var}(X) = \sigma_X^2 = E[(X - \mu_X)^2] = \sum_{i=1}^n (x_i - \mu)^2 p_i. \quad (1.1.12)$$

Verifica-se que a variância não depende do valor médio. Isto é, se a variável é "deslocada" por uma quantidade β , fazendo-se $X + \beta$, a variância da variável aleatória resultante permanece inalterada. Porém, se a variável for multiplicada por um fator de escala α , a variância é então multiplicada por α^2 . Assim, sendo α e β duas constantes reais e X uma variável aleatória discreta, cuja variância vale $\text{var}(X)$, então:

$$\text{var}(\alpha X + \beta) = \alpha^2 \text{var}(X). \quad (1.1.13)$$

O k -ésimo q -quantil de uma variável aleatória discreta X é um valor $x \in \mathbb{R}$ tal que, para $0 < k < q$, tem-se:

$$Pr(X \leq x) = P_X(x) = \frac{k}{q}. \quad (1.1.14)$$

Quando $k = 1$ e $q = 2$, o primeiro 2-quantil de X é um valor $x \in \mathbb{R}$ tal que

$$Pr(X \leq x) = P_X(x) = \frac{1}{2}, \quad (1.1.15)$$

também chamado de mediana de X .

1.1.2 Variáveis aleatórias contínuas

DEFINIÇÃO 1.1.4 (VARIÁVEL ALEATÓRIA CONTÍNUA)

Uma variável aleatória contínua real X , no espaço de probabilidade $(\Omega, \mathcal{F}, Pr)$, é uma função real que associa a cada ponto amostra w , pertencente ao espaço de amostras Ω , um número $X(w) \in \mathbb{R}$ (Cf.[1, 37, 24]), isto é:

$$\begin{aligned} X : \quad \Omega &\longrightarrow \mathbb{R} \\ w &\longmapsto X(w), \end{aligned} \quad (1.1.16)$$

tal que $X^{-1}(x) = \{w \in \Omega : X(w) \leq x\} \in \mathcal{F}, \forall x \in \mathbb{R}$.

O mapa (1.1.16) é sugerido na figura (1.1).

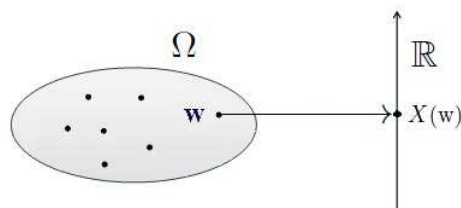


Figura 1.1: Variável aleatória (mapa de Ω para \mathbb{R}).

A distribuição de probabilidade cumulativa, P_X , de uma variável aleatória, X , é definida por:

$$\begin{aligned} P_X : \quad \mathbb{R} &\longrightarrow [0, 1] \\ x &\longmapsto P_X(x) = Pr(\{w \in \Omega : X(w) \leq x\}) \end{aligned} \quad (1.1.17)$$

onde $Pr(\{\omega \in \Omega : X(\omega) \leq x\})$ indica a probabilidade de X assumir um valor menor ou igual a x . Por esse motivo, é usual utilizar a notação simplificada:

$$P_X(x) = Pr(X \leq x). \quad (1.1.18)$$

PROPOSIÇÃO 1.1.5 . *Seja P_X a distribuição de probabilidade cumulativa de X . Então:*

- (i) $0 \leq P_X(x) \leq 1, \forall x \in \mathbb{R}$;
- (ii) P_X é monotonicamente não-decrescente: $a \leq b \implies P(a) \leq P(b)$;
- (iii) $Pr(X \geq x) = 1 - P_X(x)$ e $Pr(a \leq X < b) = P_X(b) - P_X(a)$;
- (iv) $\lim_{x \rightarrow +\infty} P_X(x) = 1$ e $\lim_{x \rightarrow -\infty} P_X(x) = 0$;
- (v) P_X é contínua à esquerda:

$$P_X(x^-) = P_X(x) \left(P_X(x^-) = \lim_{h \rightarrow 0^+} P_X(x - h) \right);$$
- (vi) P_X é uma função de variação limitada e sua variação total é igual a 1;
- (vii) $Pr(X = x) = P_X(x^+) - P_X(x^-) \left(P_X(x^+) = \lim_{h \rightarrow 0^+} P_X(x + h) \right)$.

Sendo P_X a distribuição de probabilidade cumulativa da variável aleatória contínua X , diz-se que X segue a lei P_X . Se P_X puder ser escrita na forma:

$$P_X(x) = Pr(X \leq x) = \int_{-\infty}^x p_X(u) du, \quad (1.1.19)$$

então, p_X é chamada de função densidade de probabilidade de X e, satisfaz:

$$p_X(x) \geq 0 \quad \forall x \in \mathbb{R}, \quad (1.1.20)$$

$$\int_{-\infty}^{\infty} p_X(x) dx = 1. \quad (1.1.21)$$

Essa representação nem sempre é possível, mas nessas notas só trataremos desse caso.

A média, ou valor esperado, de uma variável aleatória contínua real, X , é definida por:

$$\mu_X = E[X] = \int_{-\infty}^{\infty} x p_X(x) dx. \quad (1.1.22)$$

Dado uma função (mensurável) $h : \mathbb{R} \rightarrow \mathbb{R}$ e uma variável aleatória real contínua, X , com função densidade de probabilidade p_X , então o valor esperado de $h(X)$ pode ser calculado por:

$$E[h(X)] = \int_{-\infty}^{\infty} h(x)p(x) dx. \quad (1.1.23)$$

Uma importante propriedade do operador média é a linearidade. Sejam X e Y duas variáveis aleatórias reais contínuas e sejam α e β constantes reais, tem-se que:

$$E[\alpha X + \beta Y] = \alpha E[X] + \beta E[Y]. \quad (1.1.24)$$

A variância de X , σ_X^2 , é calculada por:

$$\sigma_X^2 = \text{var}(X) = E[(X - \mu_X)^2] = E[X^2] - (E[X])^2. \quad (1.1.25)$$

Sendo α e β duas constantes reais e X uma variável aleatória contínua, cuja variância vale $\text{var}(X)$, então:

$$\text{var}(\alpha X + \beta) = \alpha^2 \text{var}(X). \quad (1.1.26)$$

O k -ésimo q -quantil de uma variável aleatória contínua X é um valor $x \in \mathbb{R}$ tal que, para $0 < k < q$, tem-se:

$$Pr(X \leq x) = P_X(x) = \frac{k}{q}. \quad (1.1.27)$$

Quando $k = 1$ e $q = 2$, o primeiro 2-quantil de X é um valor $x \in \mathbb{R}$ tal que

$$Pr(X \leq x) = P_X(x) = \frac{1}{2}, \quad (1.1.28)$$

também chamado de mediana de X .

O k -ésimo q -quantil, x , associado a uma variável aleatória X estabelece um intervalo $(-\infty, x)$ a partir da função distribuição de probabilidade cumulativa, P_X . Assim, a medida de probabilidade de X assumir um valor nesse intervalo é k/q .

A noção de quantil induz ao conceito de histograma. Os histogramas representam uma forma de se visualizar a função densidade de probabilidade de X através de uma contagem de frequência de um número finito de amostras de X . Primeiramente faz-se a divisão do suporte de X em intervalos iguais e posteriormente conta-se quantas das amostras de X possuem um valor contido em cada um desses intervalos. Quando o número de intervalos é suficientemente grande, fixada uma certa precisão, o formato do histograma se aproxima da função densidade de probabilidade de X , p_X .

Porém para que histogramas possam ser comparados com funções densidade de probabilidade, é necessário que a área delimitada por eles seja igual a um. Nesse caso, são chamados de histogramas normalizados.

No programa MATLAB, histogramas construídos a partir de amostras de uma variável X podem ser traçados através do comando `hist`. Exemplos de histogramas normalizados aparecem nos próximos capítulos, como pode ser visto nas figuras (3.2) e (3.10).

EXEMPLO 1.1.6 *Seja X uma variável aleatória contínua com a seguinte função densidade de probabilidade:*

$$p_X(x) = \begin{cases} \frac{1}{2}, & \text{se } x \in [0, 1], \\ \frac{1}{2}, & \text{se } x \in [2, 3], \end{cases}$$

Dessa forma, $x \in [1, 2]$ é um primeiro 2-quantil de X com distribuição de probabilidade cumulativa, P_X , dada por:

$$P_X(x) = \begin{cases} \frac{1}{2}x, & \text{se } x \in [0, 1], \\ \frac{1}{2} + \frac{1}{2}(x-2), & \text{se } x \in [2, 3]. \end{cases}$$

1.1.3 Propriedades hilbertianas

DEFINIÇÃO 1.1.7 (RELAÇÃO DE EQUIVALÊNCIA)

*Sejam X e Y duas variáveis aleatórias reais. Diz-se que X é equivalente a Y , e nota-se $X \sim Y$, se e somente se $A = \{w \in \Omega : X(w) \neq Y(w)\}$ é *Pr*-quase impossível. Assim:*

$$X \sim Y \iff \text{Pr}(\{w \in \Omega : X(w) \neq Y(w)\}) = 0. \quad (1.1.29)$$

A partir dessa relação de equivalência, uma estrutura hilbertiana $L^2(\Omega)$ pode ser construída para variáveis aleatórias. Isso torna possível a utilização de toda teoria de Espaços de Hilbert e dos resultados associados (Cf. [47]).

Considere o conjunto das variáveis aleatórias em Ω :

$$\mathcal{C}(\Omega) = \{X : \Omega \longrightarrow \mathbb{R}\}, \quad (1.1.30)$$

e o subespaço vetorial, $\mathcal{V}(\Omega)$, das variáveis aleatórias simples em Ω :

$$\mathcal{V}(\Omega) = \{X \in \mathcal{C}(\Omega) : X(\Omega) \text{ é um conjunto finito}\}. \quad (1.1.31)$$

A relação \sim é uma relação de equivalência em $\mathcal{C}(\Omega)$:

$$\begin{aligned} X &\sim X \\ X \sim Y &\implies Y \sim X \\ X \sim Y \text{ e } Y \sim Z &\implies X \sim Z. \end{aligned} \tag{1.1.32}$$

A classe de equivalência de X é:

$$\tilde{X} = \{Y \in \mathcal{C}(\Omega) : Y \sim X\}. \tag{1.1.33}$$

Seja $\mathcal{C}(\Omega, Pr)$ o conjunto das classes de equivalência:

$$\mathcal{C}(\Omega, Pr) = \{\tilde{X} : X \in \mathcal{C}(\Omega)\}, \tag{1.1.34}$$

e $\mathcal{V}(\Omega, Pr)$ o conjunto das classes de equivalência das variáveis aleatórias simples:

$$\mathcal{V}(\Omega, Pr) = \{\tilde{X} \in \mathcal{C}(\Omega, Pr) : X \in \mathcal{V}(\Omega)\}. \tag{1.1.35}$$

Para $\tilde{X}, \tilde{Y} \in \mathcal{V}(\Omega, Pr)$ considere o produto escalar em $\mathcal{V}(\Omega, Pr)$:

$$(\tilde{X}, \tilde{Y}) = E[XY]. \tag{1.1.36}$$

Sendo $L^2(\Omega, Pr)$ o completamento de $\mathcal{V}(\Omega, Pr)$ para o produto escalar definido em (1.1.36), então $L^2(\Omega, Pr)$ é um espaço de Hilbert para o produto escalar (1.1.36). A norma de um elemento $X \in L^2(\Omega, Pr)$ é:

$$\|X\| = \sqrt{E[X^2]}. \tag{1.1.37}$$

1.1.4 Sequências de variáveis aleatórias

Seja $\{X_i\}_{i \in \mathbb{N}}$ uma sequência de variáveis aleatórias, ou seja $X_i \in L^2(\Omega, Pr)$, $\forall i \in \mathbb{N}$. Sequências deste tipo aparecem, por exemplo, em aplicações do Método de Monte Carlo, teorema do limite central e lei dos grandes números.

A estrutura hilbertiana do conjunto $L^2(\Omega, Pr)$, induz uma noção de convergência ligada à norma, ou ao produto escalar deste espaço. Porém, quando trabalha-se com sequências de variáveis aleatórias, o conceito de convergência de sequências de variáveis aleatórias não é único. A seguir são mostrados alguns tipos de convergência.

Convergência em média quadrática

A sequência de variáveis aleatórias $\{X_i\}_{i \in \mathbb{N}}$ converge para a variável aleatória X em média quadrática se e somente se

$$\|X_i - X\| = \sqrt{E[(X_i - X)^2]} \rightarrow 0 \text{ quando } i \rightarrow +\infty.$$

Essa é a convergência na norma 1.1.36.

Convergência em média

A sequência de variáveis aleatórias $\{X_i\}_{i \in \mathbb{N}}$ converge para a variável aleatória X em média se e somente se

$$E[|X_i - X|] \rightarrow 0 \text{ quando } i \rightarrow +\infty.$$

Convergência em probabilidade

A sequência de variáveis aleatórias $\{X_i\}_{i \in \mathbb{N}}$ converge para a variável aleatória X em probabilidade se e somente se

$$\forall \epsilon > 0 : Pr(|X_i - X| \geq \epsilon) \rightarrow 0 \text{ quando } i \rightarrow +\infty.$$

Convergência quase-certa

A sequência de variáveis aleatórias $\{X_i\}_{i \in \mathbb{N}}$ converge para a variável aleatória X quase certamente se e somente se

$$Pr(X_i \rightarrow X) = 1.$$

Convergência em distribuição, ou em Lei

A sequência de variáveis aleatórias $\{X_i\}_{i \in \mathbb{N}}$ converge em distribuição para a variável aleatória X se e somente se

$$P_{X_i}(x) \rightarrow P_X(x) \text{ quando } i \rightarrow +\infty, \forall x \text{ onde } P_X \text{ é contínua.}$$

EXEMPLO 1.1.8 Seja $\{X_i\}_{i \in \mathbb{N}}$ uma seqüência de variáveis aleatórias discretas que assumem os valores 1 e 2, cada uma delas com função densidade de probabilidade:

$$p_{X_i}(x) = \begin{cases} \frac{1}{i}, & \text{se } x = 1 \\ 1 - \frac{1}{i}, & \text{se } x = 2. \end{cases} \quad (1.1.38)$$

Verifica-se que $\{X_i\}_{i \in \mathbb{N}}$ converge para a variável aleatória X constante igual a 2 em média quadrática, pois:

$$(X_i - X)(w) = \begin{cases} -1, & \text{com probabilidade } \frac{1}{i} \\ 0, & \text{com probabilidade } 1 - \frac{1}{i}. \end{cases} \quad (1.1.39)$$

Assim:

$$\|X_i - X\| = \sqrt{E[(X_i - X)^2]} = \sqrt{(-1)^2 \frac{1}{i} + (0)^2 \left(1 - \frac{1}{i}\right)} = \sqrt{\frac{1}{i}}. \quad (1.1.40)$$

Logo, $\|X_i - X\| \rightarrow 0$ quando $i \rightarrow +\infty$.

EXEMPLO 1.1.9 Seja $\{X_i\}_{i \in \mathbb{N}}$ uma seqüência de variáveis aleatórias discretas definida por:

$$X_i(w) = \begin{cases} i, & \text{com probabilidade } 1/n \\ -1, & \text{com probabilidade } 1 - 1/n. \end{cases} \quad (1.1.41)$$

Verifica-se que $\{X_i\}_{i \in \mathbb{N}}$ converge para a variável aleatória X constante igual a -1 em probabilidade, pois:

$$(X_i - X)(w) = (X_i + 1)(w) \begin{cases} \frac{1}{n+1}, & \text{com probabilidade } \frac{1}{i} \\ 0, & \text{com probabilidade } 1 - \frac{1}{i}. \end{cases} \quad (1.1.42)$$

Assim:

$$\forall \epsilon > 0, \Pr(|X_i - X| \geq \epsilon) = \Pr(|X_i + 1| \geq \epsilon) = 1/i. \quad (1.1.43)$$

Logo, $\forall \epsilon > 0 : \Pr(|X_i - X| \geq \epsilon) \rightarrow 0$ quando $i \rightarrow +\infty$.

EXEMPLO 1.1.10 Seja $\{X_i\}_{i \in \mathbb{N}}$ uma sequência de variáveis aleatórias discretas definida por:

$$X_i(w) = \begin{cases} i, & \text{com probabilidade } 1/n^2 \\ 0, & \text{com probabilidade } 1 - 1/n^2, \end{cases} \quad (1.1.44)$$

Tem-se que $\{X_i\}_{i \in \mathbb{N}}$ converge para a variável aleatória X constante igual a 0 em probabilidade, porém não converge em média quadrática.

Verifique que $X_i - X = X_i - 0 = X_i$. Assim:

$$\forall \epsilon > 0, \Pr(|X_i - X| \geq \epsilon) = \Pr(|X_i| \geq \epsilon) = 1/i^2. \quad (1.1.45)$$

Logo, $\forall \epsilon > 0 : \Pr(|X_i - X| \geq \epsilon) \rightarrow 0$ quando $i \rightarrow +\infty$.

Entretanto:

$$\|X_i - X\| = \sqrt{E[(X_i - X)^2]} = \sqrt{(i)^2 \frac{1}{i^2} + (0)^2 \left(1 - \frac{1}{i^2}\right)} = \sqrt{1}, \quad (1.1.46)$$

de forma que $\|X_i - X\|$ não converge a zero quando $i \rightarrow +\infty$.

1.2 Vetores aleatórios

DEFINIÇÃO 1.2.1 (VETOR ALEATÓRIO)

Um vetor aleatório real $\{X\}$, de dimensão n , é composto por n variáveis aleatórias, $X_1, \dots, X_j, \dots, X_n$. Representa-se $\{X\}$ por:

$$\{X\} = \begin{bmatrix} X_1 \\ \vdots \\ X_j \\ \vdots \\ X_n \end{bmatrix} = (X_1 \ \dots \ X_j \ \dots \ X_n)^T. \quad (1.2.47)$$

O vetor $\{X\}$ é uma função real, que associa a cada ponto amostra w , pertencente ao espaço de amostras Ω , um ponto do espaço n -dimensional, $\{X(w)\} \in \mathbb{R}^n$ (Cf. [1]), isto é:

$$\{X\} : \begin{array}{l} \Omega \longrightarrow \mathbb{R}^n \\ w \longmapsto \{X(w)\} = \{x\} = (x_1, \dots, x_n)^T, \end{array} \quad (1.2.48)$$

tal que $\{X\}^{-1}(\{x\}) = \{w \in \Omega : X_1(w) \leq x_1, \dots, X_n(w) \leq x_n\}$ pertence ao espaço de eventos $F, \forall \{x\} \in \mathbb{R}^n$.

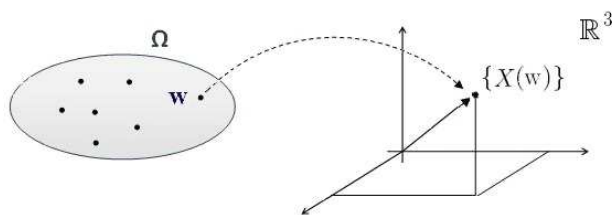


Figura 1.2: Vetor aleatório (mapa de Ω para \mathbb{R}^3).

O mapa representado em (1.2.48) é esquematizado na figura (1.2) para o caso em que $n = 3$.

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

$$P_{\{X\}} : \mathbb{R}^n \longrightarrow [0, 1] \\ \{x\} \longmapsto P_{\{X\}}(\{x\}) = \Pr(\{w \in \Omega : X_1(w) \leq x_1, \dots, X_n(w) \leq x_n\}). \quad (1.2.49)$$

onde $\{x\} \in \mathbb{R}^n$ é o vetor:

$$\{x\} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \quad (1.2.50)$$

cujo elemento de volume é definido por: $d\{x\} = dx_1 \cdots dx_n$. É usual utilizar a notação simplificada:

$$P_{\{X\}}(\{x\}) = \Pr(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n). \quad (1.2.51)$$

PROPOSIÇÃO 1.2.2 *Seja $P_{\{X\}}$ a distribuição probabilidade cumulativa de $\{X\}$. Então:*

(i) $0 \leq P_{\{X\}}(\{x\}) \leq 1, \forall \{x\} \in \mathbb{R}^n;$

(ii) $\lim_{\substack{x_1 \rightarrow +\infty \\ \dots \\ x_n \rightarrow +\infty}} P_{\{X\}}(\{x\}) = 1$ e $\lim_{\substack{x_i \rightarrow -\infty \\ (i=1, \dots, n)}} P_{\{X\}}(\{x\}) = 0;$

(iii) $P_{\{X\}}$ é uma função de variação limitada e sua variação total é igual a 1.

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

$$p_{\{X\}}(\{x\}) = \frac{\partial^n}{\partial x_1 \partial x_2 \cdots \partial x_n} P_{\{X\}}(\{x\}), \quad (1.2.52)$$

chamado também de densidade de probabilidade conjunta das variáveis aleatórias X_1, \dots, X_n . A partir de $p_{\{X\}}$, a função densidade de probabilidade marginal de cada variável aleatória X_j , $j = 1, \dots, n$, pode ser calculada por:

$$p_{X_j}(x_j) = \int_{\mathbb{R}^{n-1}} p_{\{X\}}(\{x\}) dx_1 \cdots dx_{j-1} dx_{j+1} \cdots dx_n. \quad (1.2.53)$$

Quando X_1, \dots, X_n formam um vetor de variáveis aleatórias independentes, a função densidade de probabilidade conjunta $p_{\{X\}}$ pode ser escrita como (Cf.[48]):

$$p_{\{X\}}(\{x\}) = p_{X_1}(x_1) \times \cdots \times p_{X_n}(x_n). \quad (1.2.54)$$

Vale observar que a partir da densidade de probabilidade conjunta do vetor $\{X\}$, é possível determinar a função densidade de probabilidade marginal da cada variável aleatória X_j , $j = 1, \dots, n$, através de (1.2.53). Porém, a partir das densidades de probabilidades marginais $p_{X_1}(x_1), \dots, p_{X_n}(x_n)$, é possível determinar a densidade de probabilidade conjunta do vetor, $p_{\{X\}}$, somente no caso em que X_1, \dots, X_n são variáveis aleatórias independentes.

EXEMPLO 1.2.3 *Seja $\{X\} \in \mathbb{R}^2$ um vetor aleatório composto pelas variáveis aleatórias X_1 e X_2 . Seu suporte é $k = [0, \infty) \times [0, \infty) \subset \mathbb{R}^2$ e sua função densidade de probabilidade conjunta é*

$$p_{\{X\}}(\{x\}) = \mathbb{1}_k(\{x\}) 2e^{-x_1-x_2}, \quad (1.2.55)$$

mostrada na figura (1.3).

A partir de (1.2.53), calcula-se as funções densidades de probabilidade marginais de X_1 e X_2 :

$$p_{X_1}(x_1) = \int_{\mathbb{R}} p_{\{X\}}(\{x\}) dx_2 = \mathbb{1}_{[0, \infty)}(x) 2e^{-x}, \quad (1.2.56)$$

$$p_{X_2}(x_2) = \int_{\mathbb{R}} p_{\{X\}}(\{x\}) dx_1 = \mathbb{1}_{[0, \infty)}(x) 2e^{-y}. \quad (1.2.57)$$

Verifica-se que o produto $p_{X_1}(x_1) \times p_{X_2}(x_2)$ não é igual a função densidade de probabilidade conjunta do vetor $p_{\{X\}}(\{x\})$. Dessa forma, pode-se afirmar que as variáveis aleatórias X_1 e X_2 não são independentes.

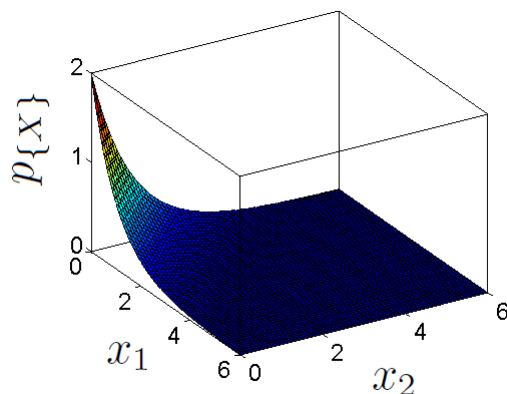


Figura 1.3: Gráfico da função densidade de probabilidade conjunta $p_{\{X\}}$ (1.2.55) na região $[0, 6] \times [0, 6]$.

A média de um vetor aleatório, $\{X\} \in \mathbb{R}^n$, é um vetor $\{\mu_X\} \in \mathbb{R}^n$:

$$E[\{X\}] = \{\mu_X\} = \begin{bmatrix} \mu_{X_1} \\ \vdots \\ \mu_{X_n} \end{bmatrix}, \quad (1.2.58)$$

onde cada componente j representa o valor esperado da variável aleatória X_j :

$$\mu_{X_j} = E[X_j] = \int_{-\infty}^{\infty} x p_{X_j}(x) dx \quad i = 1, \dots, n. \quad (1.2.59)$$

Dado uma função (mensurável) $h : \mathbb{R}^n \rightarrow \mathbb{R}$ e o vetor aleatório $\{X\} \in \mathbb{R}^n$, com função densidade de probabilidade conjunta $p_{\{X\}}$, então o valor esperado de $h(\{X\} \in \mathbb{R})$ pode ser calculado por:

$$E[h(\{X\})] = \int_{\mathbb{R}^n} h(\{x\}) p_{\{X\}}(\{x\}) d\{x\}. \quad (1.2.60)$$

A variância do vetor aleatório, $\{X\} \in \mathbb{R}^n$, é um vetor no \mathbb{R}^n dado por:

$$\{\sigma_X^2\} = \begin{bmatrix} \sigma_{X_1}^2 \\ \vdots \\ \sigma_{X_n}^2 \end{bmatrix}, \quad (1.2.61)$$

onde cada componente j representa a variância da variável aleatória X_j :

$$\sigma_{X_j}^2 = E[(X_j - \mu_{X_j})^2] = E[X_j^2] - (E[X_j])^2 \quad j = 1, \dots, n. \quad (1.2.62)$$

Caso, X_1, \dots, X_n formem um vetor de variáveis aleatórias independentes, tem-se que:

$$E[X_1 \cdots X_n] = \int_{\mathbb{R}^n} x_1 \cdots x_n p_{\{X\}}(\{x\}) d\{x\} = E[X_1] \cdots E[X_n]. \quad (1.2.63)$$

A matriz de covariância de $\{X\} \in \mathbb{R}^n$ é uma matriz $n \times n$ definida por:

$$[C] = \begin{bmatrix} E[(X_1 - \mu_{X_1})^2] & \cdots & E[(X_1 - \mu_{X_1})(X_n - \mu_{X_n})] \\ E[(X_2 - \mu_{X_2})(X_1 - \mu_{X_1})] & \cdots & E[(X_2 - \mu_{X_2})(X_n - \mu_{X_n})] \\ \vdots & \ddots & \vdots \\ E[(X_n - \mu_{X_n})(X_1 - \mu_{X_1})] & \cdots & E[(X_n - \mu_{X_n})^2] \end{bmatrix} \quad (1.2.64)$$

Os elementos na diagonal são:

$$C_{jj} = \sigma_{X_j}^2 \quad j = 1, \dots, n. \quad (1.2.65)$$

Quando as variáveis aleatórias X_1, \dots, X_n são independentes, $[C]$ é uma matriz diagonal, isto é, $C_{jk} = 0$, $j, k = 1, \dots, n$, $\forall j \neq k$.

EXEMPLO 1.2.4 *Seja $\{X\} \in \mathbb{R}^2$ um vetor aleatório composto pelas variáveis aleatórias gaussianas X_1 e X_2 . A função densidade de probabilidade conjunta de $\{X\}$ é*

$$p_{\{X\}}(\{x\}) = \frac{1}{\sqrt{(2\pi)^2 \det[C]}} \exp \left\{ -\frac{1}{2} \langle [C]^{-1}(\{x\} - \{\mu_X\}); (\{x\} - \{\mu_X\}) \rangle_{\mathbb{R}^2} \right\}, \quad (1.2.66)$$

onde $\{\mu_X\}$ é o vetor de médias de $\{X\}$ e $[C]$ é a matriz de covariância. Supondo:

$$\{\mu_X\} = \begin{bmatrix} 1 \\ 3 \end{bmatrix} \quad e \quad C = \begin{bmatrix} 1.0 & 0.5 \\ 0.5 & 1.0 \end{bmatrix}, \quad (1.2.67)$$

isto é, supondo $[C]$ uma matriz não diagonal, as variáveis aleatórias X_1 e X_2 não são independentes. O gráfico de $p_{\{X\}}$, com $\{\mu_X\}$ e $[C]$ definidos em (1.2.67), é mostrado na figura (1.4).

EXEMPLO 1.2.5 *Sejam X_1 e X_2 duas variáveis aleatórias uniformes independentes com suporte $[0, 1]$. Sorteiam-se dois valores x_1 e x_2 (amostras de X_1 e X_2 respectivamente). Deseja-se calcular a probabilidade de a soma dos quadrados desses dois valores ser menor ou igual a 1.*

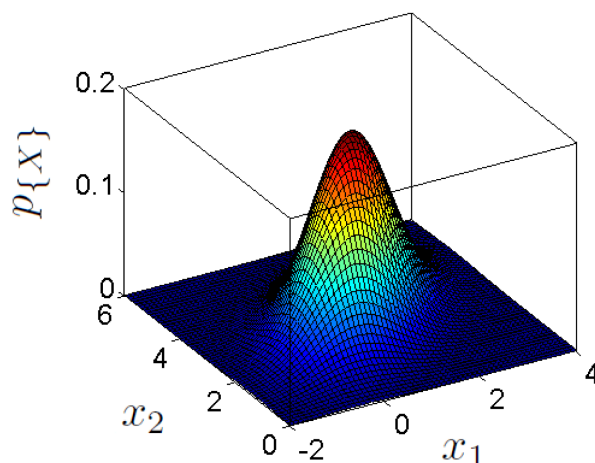


Figura 1.4: Gráfico da função densidade de probabilidade conjunta gaussiana $p_{\{X\}}$ na região $[-2, 4] \times [0, 6]$ para $\{\mu_X\} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$ e $[C] = \begin{bmatrix} 1.0 & 0.5 \\ 0.5 & 1.0 \end{bmatrix}$.

Dado que X_1 e X_2 são independentes, a função densidade de probabilidade conjunta do vetor aleatório $\{X\}$ formado por X_1 e X_2 é

$$p_{\{X\}}(\{x\}) = p_{X_1}(x_1) \times p_{X_2}(x_2) = \mathbf{1}_a(\{x\}), \quad (1.2.68)$$

onde $a = [0, 1] \times [0, 1] \subset \mathbb{R}^2$. Assim, a probabilidade de $X^2 + Y^2 \leq 1$ vale

$$Pr(X_1^2 + X_2^2 \leq 1) = \iint_R \mathbf{1}_a(\{x\}) \, dx_1 \, dx_2, \quad (1.2.69)$$

onde R é a região definida por $x^2 + y^2 \leq 1$, $x \geq 0$, $y \geq 0$, que é um quarto de círculo de raio 1. Assim, $Pr(X_1^2 + X_2^2 \leq 1) = \pi/4$.

1.3 Processos estocásticos

DEFINIÇÃO 1.3.1 (PROCESSO ESTOCÁSTICO)

Um processo estocástico, \mathcal{X} , é uma função que associa a cada ponto amostra w , pertencente ao espaço de amostras Ω , e a cada valor de um parâmetro t , pertencente a um conjunto T , um número $\mathcal{X}(w, t) \in \mathbb{R}$ (Cf. [1, 28]). Isto é:

$$\begin{aligned} \mathcal{X} : \quad \Omega \times T &\longrightarrow \mathbb{R} \\ (w, t) &\longmapsto \mathcal{X}(t, w). \end{aligned} \quad (1.3.70)$$

Na maioria dos processos estocásticos, o parâmetro t está associado ao tempo. Pode-se também definir \mathcal{X} como um mapeamento que associa a cada ponto amostra $w \in \Omega$ uma função real de um parâmetro t pertencente a um conjunto T , criando-se dessa forma uma família \mathcal{F} de funções de t ($t \in T$).

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

Quando T é um conjunto infinito, um processo estocástico pode ser interpretado como um número infinito de variáveis aleatórias indexadas pelo parâmetro t . Fixado um valor de $t = t_j$, a variável aleatória $\mathcal{X}(t_j, w)$ tem a função distribuição de probabilidade cumulativa:

$$P_{\mathcal{X}(t_j, w)}(x) = Pr(\mathcal{X}(t_j, w) < x) \quad (1.3.71)$$

e uma função densidade de probabilidade $p_{\mathcal{X}(t_j, w)}$.

A média de um processo estocástico, \mathcal{X} , é definida por (1.3.72), a variância por (1.3.73), a autocorrelação por (1.3.74) e a covariância por (1.3.75):

$$\mu_{\mathcal{X}}(t) = E[\mathcal{X}(t, w)] = \int_{-\infty}^{\infty} x p_{\mathcal{X}(t, w)}(x) dx, \quad (1.3.72)$$

$$\sigma_{\mathcal{X}}^2(t) = E[\{\mathcal{X}(t, w) - \mu_{\mathcal{X}}(t)\}^2] = E[\mathcal{X}^2(t, w)] - (E[\mathcal{X}(t, w)])^2, \quad (1.3.73)$$

$$R_{\mathcal{X}\mathcal{X}}(t_1, t_2) = E[\mathcal{X}(t_1, w) \mathcal{X}(t_2, w)], \quad (1.3.74)$$

$$\begin{aligned} C(t_1, t_2) &= E[\{\mathcal{X}(t_1, w) - \mu_{\mathcal{X}}(t_1)\} \{\mathcal{X}(t_2, w) - \mu_{\mathcal{X}}(t_2)\}] \\ &= E[\mathcal{X}(t_1, w)\mathcal{X}(t_2, w)] - \mu_{\mathcal{X}}(t_1)\mu_{\mathcal{X}}(t_2) \\ &= R_{\mathcal{X}\mathcal{X}}(t_1, t_2) - \mu_{\mathcal{X}}(t_1)\mu_{\mathcal{X}}(t_2) \end{aligned} \quad (1.3.75)$$

1.4 Exercícios

EXERCÍCIO 1.4.1 *Seja* (Ω, F, Pr) *um espaço de probabilidade no qual:*

$$\Omega = \{1, 2, 3, 4, 5, 6\}, \quad F = \{\emptyset, \{2, 4, 6\}, \{1, 3, 5\}, \Omega\},$$

e sejam U , V e W funções em Ω definidas por:

$$U(w) = w, \quad V(w) = \begin{cases} 1, & \text{se } w \text{ é par} \\ 0, & \text{se } w \text{ é ímpar} \end{cases}, \quad W(w) = w^2$$

para todo $w \in \Omega$. Determine quais dessas funções são variáveis aleatórias no espaço de probabilidade $(\Omega, \mathcal{F}, Pr)$.

EXERCÍCIO 1.4.2 Determine qual valor de constante c que torna a função p , definida por

$$p_X(x) = \begin{cases} \frac{c}{x(x+1)}, & \text{se } x = 1, 2, \dots, \\ 0, & \text{caso contrário,} \end{cases},$$

uma função de massa de uma variável aleatória discreta X .

EXERCÍCIO 1.4.3 Demonstre as propriedades da função distribuição de probabilidade cumulativa de uma variável aleatória contínua real enunciadas em (1.1.5).

EXERCÍCIO 1.4.4 Mostre que a variância de uma variável aleatória contínua real X pode ser escrita como $\sigma_X^2 = E[X^2] - E[X]^2$.

EXERCÍCIO 1.4.5 Mostre que se X e Y são variáveis aleatórias contínuas reais independentes, com médias $\mu_X = E[X]$ e $\mu_Y = E[Y]$ respectivamente, então $E[XY] = E[X]E[Y]$.

EXERCÍCIO 1.4.6 Seja X uma variável aleatória contínua com função densidade de probabilidade, p_X , bilateral exponencial:

$$p_X(x) = \frac{1}{2}ce^{-c|x|},$$

onde c é um parâmetro da densidade. Mostre que a média e variância de X valem 0 e $2c^{-2}$ respectivamente.

EXERCÍCIO 1.4.7 Seja X uma variável aleatória discreta, definida por:

$$X(w) = \begin{cases} 0 & \text{com probabilidade } 1/2 \\ 1 & \text{com probabilidade } 1/2, \end{cases}$$

e seja Y uma variável aleatória discreta definida por $Y = 1 - X$. Mostre que a função distribuição de probabilidade cumulativa de X é igual a função distribuição de probabilidade cumulativa de Y (convergência em distribuição), mas $|X - Y| = 1$ (não há convergência em média).

EXERCÍCIO 1.4.8 Seja $\{X\} \in \mathbb{R}^2$ um vetor aleatório composto pelas variáveis aleatórias X_1 e X_2 . Seu suporte é $k = [2, 6] \times [0, 5] \subset \mathbb{R}^2$ e sua função densidade de probabilidade conjunta é

$$p_{\{X\}}(\{x\}) = \mathbb{1}_k(\{x\}) c(2x_1 + x_2).$$

Determine:

- (a) o valor da constante c ;
- (b) as funções densidade de probabilidade marginais de X_1 e X_2 ;
- (c) as funções distribuição de probabilidade cumulativa marginais de X_1 e X_2 ;
- (d) $Pr(3 \leq X \leq 4, Y \geq 2)$;
- (e) $Pr(X \geq 3)$;
- (f) $Pr(X + Y \geq 4)$;
- (g) a função distribuição conjunta de $\{X\}$;
- (h) se X_1 e X_2 são variáveis aleatórias independentes.

Capítulo 2

Máximo de Entropia e Informação

Na análise estocástica de um sistema, duas diferentes abordagens podem ser destacadas: na primeira, considera-se que a incerteza está presente nos parâmetros do sistema (modelagem estocástica paramétrica) e na segunda, considera-se que a incerteza está presente na modelagem empregada para o sistema (modelagem estocástica não paramétrica).

No capítulo 1, foram apresentadas as definições básicas de objetos estocásticos, ou seja, variáveis aleatórias, vetores aleatórios e processos estocásticos. Nesse capítulo, uma metodologia para a construção de modelos probabilísticos para esses objetos é apresentada. Vale observar que essa metodologia pode ser empregada tanto em abordagens paramétricas (determina-se as funções densidades de probabilidades associadas a variáveis e vetores aleatórios) quanto em abordagens não-paramétricas (determina-se as funções densidades de probabilidades associadas a matrizes aleatórias).

2.1 Construção do modelo estocástico

O trabalho de simulações estocásticas de um sistema geralmente segue uma determinada organização:

1. constrói-se um **modelo determinístico** para o sistema;
2. constrói-se um **modelo estocástico** para o sistema: na abordagem paramétrica, seleciona-se os parâmetros escalares ou vetoriais do sistema considerados aleatórios e determina-se suas respectivas funções densidades de

probabilidades; na abordagem não-paramétrica, determina-se as funções densidades de probabilidades para as matrizes que estão associadas ao modelo do sistema.

3. aplica-se o método de Monte Carlo para obter-se inferências estatísticas sobre a resposta do sistema.

As etapas da construção dos modelos determinístico e estocástico do sistema são fundamentais para obter-se resultados realistas nas simulações estocásticas. Esses dois modelos são utilizados no Método de Monte Carlo e influenciam de forma direta as respostas obtidas.

No método de Monte Carlo, tipicamente gera-se amostras dos objetos que foram considerados aleatórios (usando às distribuições de probabilidade previamente fixadas - modelo estocástico) e, para cada amostra gerada calcula-se uma resposta determinística do sistema (através do modelo determinístico). Esses resultados determinísticos são agregados ao resultado final desejado (estatísticas e aproximações da densidade de probabilidade da resposta do sistema podem ser calculadas). A figura (2.1) mostra um esquema com as principais etapas do método.

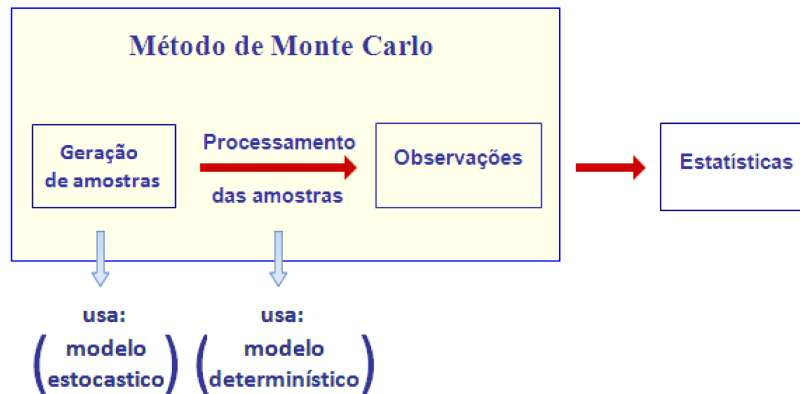


Figura 2.1: Método de Monte Carlo.

Assim, a escolha das funções densidades de probabilidade dos parâmetros considerados aleatórios deve ser feita de forma cautelosa. Essas funções podem ser determinadas:

- a partir da construção de histogramas com resultados de experimentos;
- através do Princípio da Máxima Entropia (PEM).

Muitas vezes, a realização de experimentos com os parâmetros pode ser muito custosa ou até inviável. Por isso, essa estratégia é pouco utilizada na determinação das funções densidades de probabilidade. O mais comum e frequente na literatura é fazer isso através do princípio da máxima entropia.

A seguir é mostrado como funciona esse princípio.

2.2 Princípio da Entropia Máxima

O Princípio da Entropia Máxima (PEM) é um método eficaz de construção de modelos probabilísticos para variáveis aleatórias escalares e vetoriais, contínuas e discretas. Ou seja, é empregado para a construção da função de massa ou densidade de probabilidade de objetos estocásticos, utilizando para isso somente as informações disponíveis sobre esses objetos.

Por exemplo, para uma variável aleatória vetorial, $\{X\} \in \mathbb{R}^n$, o método permite construir a função densidade de probabilidade quando se conhece somente:

- o suporte $k_n \in \mathbb{R}^n$ dessa variável ou;
- o suporte e o vetor média, ou;
- o suporte, o vetor média e a matriz de covariância.

A função densidade de probabilidade obtida pelo PEM depende diretamente da informação conhecida sobre a variável aleatória.

Dessa forma, o Princípio da Entropia Máxima consiste em maximizar a entropia (incerteza) sujeita a restrições definidas pelas informações disponíveis. O método baseia-se na teoria da informação. O conceito de entropia usado é o de Shannon (Cf. [43]) e, a ideia de maximizar a entropia foi proposta por Jaynes (Cf. [17, 18]) no contexto da Mecânica Estatística.

O Princípio estabelece que:

De todas as distribuições de probabilidade consistentes com as restrições impostas, escolhe-se aquela que maximiza a incerteza (entropia).

Para explicar o PEM, inicialmente serão tratadas as variáveis aleatórias discretas tanto nos casos em que assume um conjunto finito de valores, quanto nos casos em que assume um conjunto infinito. Posteriormente são tratadas as variáveis aleatórias contínuas e, por último, os vetores aleatórios e matrizes aleatórias.

2.2.1 Variáveis aleatórias discretas

Seja X uma variável aleatória discreta que assume n valores distintos x_1, \dots, x_n . Suponha que deseja-se determinar a função de massa p dessa variável aleatória X ,

ou seja, deseja-se determinar os valores p_1, \dots, p_n , tais que:

$$p_1 + \dots + p_n = 1, \quad p_i \geq 0, \quad \forall i \in \{1, \dots, n\}, \quad (2.2.1)$$

e que maximizam a medida da entropia de Shannon dada por:

$$S(p) = - \sum_{i=1}^n p_i \ln(p_i). \quad (2.2.2)$$

Seja \mathcal{C}_{ad} o espaço das funções $x \mapsto p(x)$ de $\{x_1, \dots, x_n\}$ em $[0, 1]$ e, que atendem ao vínculo definido em (2.2.1).

Para descobrir qual a expressão da função $p \in \mathcal{C}_{ad}$ que maximiza essa entropia S (2.2.2), um problema de otimização deve ser resolvido,

$$\max_{p \in \mathcal{C}_{ad}} S(p). \quad (2.2.3)$$

Para resolvê-lo, utiliza-se o método dos multiplicadores de Lagrange, de forma que o problema com vínculo é transformado em um problema sem vínculo através de um multiplicador de Lagrange. Nesse problema, a Lagrangiana é:

$$\begin{aligned} L(p, \lambda_0) &= S(p) - (\lambda_0 - 1) \left(\sum_{i=1}^n p_i - 1 \right) \\ &= S(p) - \lambda_0 \sum_{i=1}^n p_i + \lambda_0 + \sum_{i=1}^n p_i - 1, \end{aligned} \quad (2.2.4)$$

onde $(\lambda_0 - 1)$ é o multiplicador de Lagrange correspondente ao vínculo (2.2.1). Fazendo-se:

$$\begin{aligned} \frac{\partial L}{\partial p_i}(p, \lambda_0) &= -\ln(p_i) - p_i \frac{1}{p_i} - \lambda_0 + 1 = 0 \\ &= -\ln(p_i) - \lambda_0 = 0, \quad i = 1, \dots, n, \end{aligned} \quad (2.2.5)$$

verifica-se que:

$$\ln(p_i) = -\lambda_0 \implies p_i = e^{-\lambda_0}, \quad i = 1, \dots, n. \quad (2.2.6)$$

Assim, quando nenhuma informação sobre a variável aleatória X é conhecida, a função de massa p que maximiza a entropia de Shannon é a uniforme. Nesse caso, tem-se:

$$p_i = \frac{1}{n}, \quad i = 1, \dots, n. \quad (2.2.7)$$

Caso X assumisse um número infinito de valores o problema de otimização não teria solução. Isto é, não existe uma distribuição de probabilidade que maximize a entropia de Shannon. Logo o PEM pode não fornecer uma solução que possibilite a construção do modelo probabilístico.

Vejamos agora uma situação mais geral, quando mais vínculos são prescritos. Suponha que a função de massa esteja submetida a m vínculos escritos na forma:

$$\sum_{i=1}^m p_i g_{r i}(x_1, \dots, x_n) = a_r, \quad r = 1, \dots, m. \quad (2.2.8)$$

Nesse caso, o problema de maximização da entropia S sujeito aos $m+1$ vínculos (2.2.1 e 2.2.8) é transformado em um problema sem vínculos através de $m+1$ multiplicadores de Lagrange:

$$L(p, \lambda_0, \dots, \lambda_m) = S(p) - (\lambda_0 - 1) \left(\sum_{i=1}^n p_i - 1 \right) - \sum_{r=1}^m \lambda_r \left(\sum_{i=1}^n p_i g_{r i} \right) \quad (2.2.9)$$

onde $(\lambda_0 - 1), \lambda_1, \dots, \lambda_m$ são os $m+1$ multiplicadores de Lagrange. Fazendo-se:

$$\frac{\partial L}{\partial p_i} = 0 \implies -\ln(p_i) - \lambda_0 - \sum_{r=1}^m \lambda_r g_{r i} = 0, \quad (2.2.10)$$

ou seja:

$$p_i = \exp(-\lambda_0 - \lambda_1 g_{1i} - \lambda_2 g_{2i} - \dots - \lambda_m g_{mi}), \quad i = 1, \dots, n. \quad (2.2.11)$$

Os multiplicadores $\lambda_0 - 1, \lambda_1, \dots, \lambda_m$ são calculados substituindo-se p_i (2.2.11) em (2.2.1) e (2.2.8), tal que:

$$\sum_{i=1}^n \exp\left(-\lambda_0 - \sum_{j=1}^m \lambda_j g_{ji}\right) = 1 \quad (2.2.12)$$

e

$$\sum_{i=1}^n g_{r i} \exp\left(-\lambda_0 - \sum_{j=1}^m \lambda_j g_{ji}\right) = a_r, \quad r = 1, \dots, m. \quad (2.2.13)$$

EXEMPLO 2.2.1 *Seja X uma variável aleatória discreta que assume n finitos valores distintos $\{x_1, \dots, x_n\}$. Suponha que a média dessa variável seja conhecida (μ). Deseja-se determinar a função de massa p de X que maximiza a entropia S (2.2.2), tal que:*

$$\sum_{i=1}^n p_i = 1 \quad e \quad \sum_{i=1}^n x_i p_i = \mu. \quad (2.2.14)$$

Escrevendo esse problema de otimização através dos multiplicadores de Lagrange:

$$L(p, \lambda_0, \lambda_1) = - \sum_{i=1}^n p_i \ln(p_i) - (\lambda_0 - 1) \left(\sum_{i=1}^n p_i - 1 \right) - \lambda_1 \left(\sum_{i=1}^n x_i p_i - \mu \right) \quad (2.2.15)$$

$$\frac{\partial L}{\partial p_i} = 0 \implies -\ln(p_i) - \lambda_0 - \lambda_1 x_i = 0. \quad (2.2.16)$$

Dessa forma:

$$p_i = \exp(-\lambda_0 - \lambda_1 x_i), \quad i = 1, \dots, n. \quad (2.2.17)$$

Os valores de λ_0 e λ_1 são determinados a partir de (2.2.14):

$$\sum_{i=1}^n \exp(-\lambda_0 - \lambda_1 x_i) = 1 \quad e \quad \sum_{i=1}^n x_i \exp(-\lambda_0 - \lambda_1 x_i) = \mu. \quad (2.2.18)$$

EXEMPLO 2.2.2 Seja X uma variável aleatória discreta que assume infinitos valores distintos $\{0, 1, 2, \dots\}$. Suponha que a média dessa variável seja conhecida (μ). Deseja-se determinar a função de massa p de X que maximiza a entropia S (2.2.2), tal que:

$$\sum_{i=0}^{\infty} p_i = 1 \quad e \quad \sum_{i=0}^{\infty} i p_i = \mu. \quad (2.2.19)$$

Escrevendo esse problema de otimização através dos multiplicadores de Lagrange:

$$L(p, \lambda_0, \lambda_1) = - \sum_{i=0}^{\infty} p_i \ln(p_i) - (\lambda_0 - 1) \left(\sum_{i=0}^{\infty} p_i - 1 \right) - \lambda_1 \left(\sum_{i=0}^{\infty} i p_i - \mu \right) \quad (2.2.20)$$

$$\frac{\partial L}{\partial p_i} = 0 \implies -\ln(p_i) - \lambda_0 - \lambda_1 i = 0. \quad (2.2.21)$$

Dessa forma:

$$p_i = \exp(-\lambda_0 - \lambda_1 i), \quad i = 0, 1, 2, \dots. \quad (2.2.22)$$

Os multiplicadores $(\lambda_0 - 1)$ e λ_1 são determinados a partir dos vínculos prescritos em (2.2.19):

$$\sum_{i=0}^{\infty} \exp(-\lambda_0 - \lambda_1 i) = \exp(-\lambda_0) \sum_{i=0}^{\infty} \exp(-\lambda_1 i) = \frac{\exp(-\lambda_0)}{1 - \exp(-\lambda_1)} = 1, \quad (2.2.23)$$

$$\sum_{i=0}^{\infty} i \exp(-\lambda_0 - \lambda_1 i) = \exp(-\lambda_0) \frac{\exp(-\lambda_1)}{(1 - \exp(-\lambda_1))^2} = \mu. \quad (2.2.24)$$

Assim:

$$\exp(-\lambda_1) = \frac{\mu}{1 + \mu} \implies \lambda_1 = -\ln\left(\frac{\mu}{1 + \mu}\right), \quad (2.2.25)$$

$$\exp(-\lambda_0) = 1 - \frac{\mu}{1 + \mu} \implies \lambda_0 = -\ln\left(\frac{1}{1 + \mu}\right). \quad (2.2.26)$$

A expressão (2.2.22) pode ser reescrita como:

$$p_i = \exp\left[\ln\left(\frac{1}{1 + \mu}\right) + i \ln\left(\frac{\mu}{1 + \mu}\right)\right], \quad i = 0, 1, 2, \dots \quad (2.2.27)$$

$$p_i = \left(\frac{1}{1 + \mu}\right) \exp\left[i \ln\left(\frac{\mu}{1 + \mu}\right)\right], \quad i = 0, 1, 2, \dots \quad (2.2.28)$$

2.2.2 Variáveis aleatórias contínuas

Os casos de aplicação do Princípio da Entropia Máxima de maior interesse são relacionados com as variáveis aleatórias reais contínuas. Nesse caso, o suporte da variável aleatória é a primeira informação de interesse sobre a variável. Observe que esse suporte pode ser um intervalo finito, $k = [a, b]$ ($a, b \in \mathbb{R}$), pode ser um intervalo semi-infinito $[a, \infty]$, ($a \in \mathbb{R}$), ou pode ser todos os reais $(-\infty, \infty)$.

DEFINIÇÃO 2.2.3 (MEDIDA DA ENTROPIA DE SHANNON)

Caso tenha-se uma variável aleatória real contínua, X , com suporte $k = [a, b] \subset \mathbb{R}$ e com função densidade de probabilidade p , a medida da entropia de Shannon, S , é dada por:

$$S(p) = - \int_k p(x) \ln p(x) dx. \quad (2.2.29)$$

Como p é uma função densidade de probabilidade, sabe-se que:

$$\int_a^b p(x) dx = 1 \quad \text{e} \quad p(x) \geq 0, \quad \forall x \in [a, b]. \quad (2.2.30)$$

Seja C_{ad} o espaço das funções $\{x\} \mapsto p(x)$ do \mathbb{R} em $\mathbb{R}^{\geq 0}$, tendo todas o mesmo suporte conhecido $k \subset \mathbb{R}$ e, atendendo aos vínculos definidos em (2.2.30).

Para descobrir qual a expressão da função $p \in C_{ad}$ que maximiza essa entropia S (2.2.30), um problema de otimização deve ser resolvido,

$$\max_{p \in C_{ad}} S(p). \quad (2.2.31)$$

Assim como feito no caso das variáveis aleatórias discretas, esse problema de otimização com o vínculo (2.2.30) será transformado em um problema sem vínculo através de um multiplicador de Lagrange. A Lagrangiana é:

$$L(p, \lambda_0) = - \int_k p(x) \ln p(x) dx - (\lambda_0 - 1) \left[\int_k p(x) dx - 1 \right]. \quad (2.2.32)$$

Ela pode ser escrita como:

$$L(p, \lambda_0) = u(\lambda_0) - \int_a^b h(p, \lambda_0) dx, \quad (2.2.33)$$

com:

$$u(\lambda_0) = (\lambda_0 - 1), \quad (2.2.34)$$

$$h(p, \lambda_0) = p(x)[\ln p(x) + (\lambda_0 - 1)]. \quad (2.2.35)$$

Pelo cálculo variacional, os extremos de máximo e mínimo da Lagrangiana L verificam:

$$\frac{\partial}{\partial p(x)} h(p, \lambda_0) = 0. \quad (2.2.36)$$

Dessa forma, obtém-se como extremo a densidade:

$$p(x) = \mathbb{1}_{[a,b]}(x) \exp(-\lambda_0). \quad (2.2.37)$$

Porém, como:

$$\frac{\partial^2}{\partial p(x)^2} h(p, \lambda_0) > 0, \quad (2.2.38)$$

o extremo identificado é um máximo e assim, a densidade p que maximiza a entropia, S , é a densidade uniforme no intervalo $k = [a, b]$, ou seja:

$$p(x) = \mathbb{1}_{[a,b]}(x) \frac{1}{b-a}. \quad (2.2.39)$$

Frequentemente é comum trabalhar-se com problemas em que mais vínculos são prescritos, como por exemplo a média e momentos de ordem superior. Suponha que a variável aleatória X esteja submetida a m vínculos escritos na forma:

$$\int_k p(x) g_r(x) dx = a_r \quad r = 1, \dots, m. \quad (2.2.40)$$

Para essa variável aleatória, a Lagrangiana é escrita como:

$$\begin{aligned} L(p, \lambda_0, \dots, \lambda_m) &= S(p) - (\lambda_0 - 1) \left(\int_k p(x) dx - 1 \right) \\ &\quad - \sum_{r=1}^m \lambda_r \left(\int_k p(x) g_r(x) dx - a_r \right), \end{aligned} \quad (2.2.41)$$

onde $\lambda_0 - 1, \lambda_1, \dots, \lambda_m$ são os $m + 1$ multiplicadores de Lagrange. Reescreve-se (2.2.41) como:

$$L(p, \lambda_0) = u(\lambda_0, \dots, \lambda_m) - \int_k h(p, \lambda_0, \dots, \lambda_m) dx, \quad (2.2.42)$$

com:

$$u(\lambda_0, \dots, \lambda_m) = (\lambda_0 - 1) + \sum_{r=1}^m \lambda_r a_r, \quad (2.2.43)$$

$$h(p, \lambda_0, \dots, \lambda_m) = p(x) [\ln p(x) + (\lambda_0 - 1) + \sum_{r=1}^m \lambda_r g_r(x)]. \quad (2.2.44)$$

Pelo cálculo variacional, os extremos de máximo e mínimo da Lagrangiana L verificam:

$$\frac{\partial}{\partial p(x)} h(p, \lambda_0) = 0. \quad (2.2.45)$$

Dessa forma, obtém-se como extremo a densidade:

$$p(x) = \mathbb{1}_{[a,b]}(x) \exp(-\lambda_0 - \lambda_1 g_1(x) - \lambda_2 g_2(x) - \dots - \lambda_m g_m(x)). \quad (2.2.46)$$

Os multiplicadores $\lambda_0 - 1, \lambda_1, \dots, \lambda_m$ são calculados substituindo-se a expressão p (2.2.46) em (2.2.30) e (2.2.40).

EXEMPLO 2.2.4 *Seja X uma variável aleatória contínua com suporte compacto $[a, b]$ e com valores para média e segundo momento prescritos: $E[X] = \mu$ e $E[X^2] = \alpha^2$. Deseja-se determinar a densidade de probabilidade p de X que maximiza a entropia S (2.2.29), tal que:*

$$\int_a^b p(x) dx = 1, \quad \int_a^b p(x) x dx = \mu, \quad \int_a^b p(x) x^2 dx = \alpha^2. \quad (2.2.47)$$

Pela expressão (2.2.46), tem-se que:

$$p(x) = \mathbb{1}_{[a,b]}(x) \exp(-\lambda_0 - \lambda_1 x - \lambda_2 x^2). \quad (2.2.48)$$

EXEMPLO 2.2.5 *Seja X uma variável aleatória contínua com suporte $[0, \infty)$ e com valor da média prescrito: $E[X] = \mu$. Deseja-se determinar a densidade de probabilidade p de X que maximiza a entropia S (2.2.29), tal que:*

$$\int_0^\infty p(x) dx = 1, \quad \int_0^\infty p(x) x dx = \mu. \quad (2.2.49)$$

Pela expressão (2.2.46), tem-se que:

$$p(x) = \exp(-\lambda_0 - \lambda_1 x). \quad (2.2.50)$$

Calculando-se os multiplicadores λ_0 e λ_1 , chega-se a seguinte expressão para a densidade p :

$$p(x) = \mathbb{1}_{[0,\infty)}(x) \frac{1}{\mu} e^{-\frac{x}{\mu}}. \quad (2.2.51)$$

EXEMPLO 2.2.6 *Seja X uma variável aleatória contínua tal que:*

1. seu suporte seja $[0, \infty)$;
2. sua média seja $E[X] = \mu \in [0, \infty)$;
3. o valor esperado de $E[\log(X)] = q$.

Nesse caso, pelo PEM a densidade de probabilidade p de X que maximiza a entropia de Shannon é a densidade Gamma:

$$p(x) = \mathbb{1}_{[0,+\infty)}(x) \frac{1}{\mu} \left(\frac{1}{\delta^2}\right)^{\frac{1}{\delta^2}} \frac{1}{\Gamma(1/\delta^2)} \left(\frac{x}{\mu}\right)^{\frac{1}{\delta^2}-1} \exp\left(-\frac{x}{\delta^2 \mu}\right), \quad (2.2.52)$$

onde:

- Γ é a função Gamma: $\Gamma(a) = \int_0^\infty t^{a-1} \exp(-t) dt$;

- $\delta = \frac{\sigma}{\mu}$ é um fator de dispersão adimensional.

EXEMPLO 2.2.7 Seja X uma variável aleatória contínua com suporte $(-\infty, \infty)$ e com valores para média e segundo momento prescritos: $E[X] = \mu$ e $E[X^2] = \alpha^2$. Deseja-se determinar a densidade de probabilidade p de X que maximiza a entropia S (2.2.29), tal que:

$$\int_{-\infty}^{\infty} p(x) dx = 1, \quad \int_{-\infty}^{\infty} p(x) x dx = \mu, \quad \int_{-\infty}^{\infty} p(x) x^2 dx = \alpha^2. \quad (2.2.53)$$

Pela expressão (2.2.46), tem-se que:

$$p(x) = \exp(-\lambda_0 - \lambda_1 x - \lambda_2 x^2) = a \exp(-b(x-c)^2), \quad (2.2.54)$$

onde a , b e c devem ser determinados por:

$$\begin{aligned} a \int_{-\infty}^{\infty} \exp[-b(x-c)^2] dx &= 1, \\ a \int_{-\infty}^{\infty} x \exp[-b(x-c)^2] dx &= \mu, \\ a \int_{-\infty}^{\infty} x^2 \exp[-b(x-c)^2] dx &= \alpha^2. \end{aligned} \quad (2.2.55)$$

Usando o seguinte resultado (Cf. [15]):

$$\int_{-\infty}^{\infty} \exp[-b(x-c)^2] dx = \sqrt{\frac{\pi}{b}}, \quad \text{para } b > 0, \quad (2.2.56)$$

verifica-se que:

$$a = \frac{1}{\sqrt{2\pi\sigma}}, \quad b = \frac{1}{2\sigma^2}, \quad c = \mu, \quad (2.2.57)$$

onde a variância $\sigma^2 = \alpha^2 - \mu^2$.

Assim, a função densidade de probabilidade que maximiza a entropia de Shannon é a Gaussiana:

$$p(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left[-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2}\right]. \quad (2.2.58)$$

A seguir é mostrado um resumo da construção de modelos probabilísticos para variáveis aleatórias contínuas através do Princípio da Entropia Máxima.

Variáveis aleatórias contínuas com suporte compacto $[a, b]$:

Seja X uma variável aleatória contínua real com suporte compacto $[a, b]$. Deseja-se determinar a densidade de probabilidade p de X que maximiza a entropia S (2.2.29) tal que:

$$\int_a^b p(x) dx = 1 \quad \text{e} \quad p(x) \geq 0 \quad \forall x \in [a, b]. \quad (2.2.59)$$

1. Caso nenhuma informação adicional sobre X seja conhecida, a densidade de probabilidade que maximiza a entropia de Shannon é a uniforme no intervalo $[a, b]$:

$$p(x) = \mathbb{1}_{[a,b]}(x) \frac{1}{b-a}. \quad (2.2.60)$$

2. Caso a média de X seja conhecida, $E[X] = \mu$, a densidade de probabilidade que maximiza a entropia de Shannon é a exponencial truncada:

$$p(x) = \mathbb{1}_{[a,b]}(x) \exp(-\lambda_0 - \lambda_1 x), \quad (2.2.61)$$

onde:

$$e^{-\lambda_0} \int_a^b e^{-\lambda_1 x} dx = 1 \quad \text{e} \quad e^{-\lambda_0} \int_a^b x e^{-\lambda_1 x} dx = \mu. \quad (2.2.62)$$

3. Caso $a = 0$, $b = 1$ e, os valores de $E[\ln(X)]$ e $E[\ln(1-X)]$ sejam conhecidos e iguais a k_1 e k_2 respectivamente, a densidade de probabilidade que maximiza a entropia de Shannon é a densidade beta:

$$p(x) = \mathbb{1}_{[0,1]}(x) \frac{1}{B(m, n)} x^{(m-1)} (1-x)^{n-1}, \quad (2.2.63)$$

onde B é a função beta:

$$B(m, n) = \frac{\Gamma(m)\Gamma(n)}{\Gamma(m+n)}, \quad (2.2.64)$$

e onde $\Gamma(m) = \int_0^\infty t^{m-1} \exp(-t) dt$. Os valores de m e n são determinados a partir dos vínculos $E[\ln(X)]$ e $E[\ln(1-X)]$. De forma que:

$$\frac{1}{B(m, n)} \int_0^1 x^{(m-1)} (1-x)^{n-1} \ln(x) dx = E[\ln(X)] = k_1, \quad (2.2.65)$$

$$\frac{1}{\beta(m, n)} \int_0^1 x^{(m-1)} (1-x)^{n-1} \ln(1-x) dx = E[\ln(1-X)] = k_2. \quad (2.2.66)$$

Variáveis aleatórias contínuas com suporte $[0, \infty)$:

Seja X uma variável aleatória contínua real com suporte $[0, \infty)$. Deseja-se determinar a densidade de probabilidade p de X que maximiza a entropia S :

$$S(p) = - \int_0^{\infty} p(x) \ln p(x) dx. \quad (2.2.67)$$

tal que:

$$\int_0^{\infty} p(x) dx = 1 \text{ e } p(x) \geq 0, \quad \forall x \in [0, \infty). \quad (2.2.68)$$

1. Caso nenhuma informação adicional sobre X seja conhecida, o problema de otimização não possui solução.
2. Caso a média de X seja conhecida, $E[X] = \mu$, a densidade de probabilidade que maximiza a entropia de Shannon é a exponencial:

$$p(x) = \mathbf{1}_{[0, \infty)}(x) \frac{1}{\mu} e^{-\frac{x}{\mu}}. \quad (2.2.69)$$

3. Caso sejam conhecidos o valor esperado de X e o valor esperado de $\log(X)$, ou seja, $E[X] = \mu$ e $E[\log(X)] = q$, a densidade de probabilidade p que maximiza a entropia de Shannon é a densidade Gamma:

$$p(x) = \mathbf{1}_{[0, +\infty)}(x) \frac{1}{\mu} \left(\frac{1}{\delta^2}\right)^{\frac{1}{\delta^2}} \frac{1}{\Gamma(1/\delta^2)} \left(\frac{x}{\mu}\right)^{\frac{1}{\delta^2}-1} \exp\left(-\frac{x}{\delta^2 \mu}\right), \quad (2.2.70)$$

onde Γ é a função Gamma ($\Gamma(a) = \int_0^{\infty} t^{a-1} \exp(-t) dt$) e $\delta = \frac{\sigma}{\mu}$ um fator de dispersão adimensional.

4. Caso os valores de $E[\ln(X)]$ e $E[\ln(1-X)]$ sejam conhecidos e iguais a k_1 e k_2 respectivamente, a densidade de probabilidade que maximiza a entropia de Shannon é:

$$p(x) = \mathbf{1}_{]0, +\infty)}(x) \frac{1}{B(m, n)} x^{(m-1)} (1+x)^{-(n+m)}, \quad (2.2.71)$$

onde B é a função beta (2.2.64) e, os valores de m e n são determinados a partir dos vínculos $E[\ln(X)]$ e $E[\ln(1-X)]$. De forma que:

$$\frac{1}{B(m, n)} \int_0^{\infty} x^{(m-1)} (1+x)^{-(n+m)} \ln(x) dx = E[\ln(X)] = k_1, \quad (2.2.72)$$

$$\frac{1}{\beta(m, n)} \int_0^\infty x^{(m-1)} (1+x)^{-(n+m)} \ln(1-x) dx = E[\ln(1-X)] = k_2. \quad (2.2.73)$$

Variáveis aleatórias contínuas com suporte $(-\infty, \infty)$:

Seja X uma variável aleatória contínua real com suporte $[-\infty, \infty)$.

Deseja-se determinar a densidade de probabilidade p de X que maximiza a entropia S :

$$S(p) = - \int_{-\infty}^{\infty} p(x) \ln p(x) dx. \quad (2.2.74)$$

tal que:

$$\int_{-\infty}^{\infty} p(x) dx = 1 \quad \text{e} \quad p(x) \geq 0 \quad \forall x \in \mathbb{R}. \quad (2.2.75)$$

1. Caso nenhuma informação adicional sobre X seja conhecida, o problema de otimização não possui solução.
2. Caso somente a média de X seja conhecida, o problema de otimização não possui solução.
3. Caso sejam conhecidos $E[X] = \mu$ e $E[X^2] = \alpha^2$, a densidade de probabilidade que maximiza a entropia de Shannon é a Gaussiana:

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left[-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2} \right]. \quad (2.2.76)$$

2.2.3 Vetores aleatórios

De forma análoga ao caso das variáveis aleatórias contínuas, a medida de incerteza de um vetor aleatório $\{X\}$ no \mathbb{R}^n é definido pela entropia de sua função densidade de probabilidade p :

$$S(p) = - \int_{\mathbb{R}^n} p(\{x\}) \ln p(\{x\}) d\{x\}. \quad (2.2.77)$$

Como p é uma função densidade de probabilidade, sabe-se que:

$$p(\{x\}) \geq 0, \quad \forall \{x\} \in \mathbb{R}^n \quad \text{e} \quad \int_{\mathbb{R}^n} p(\{x\}) d\{x\} = 1. \quad (2.2.78)$$

Além de (2.2.78), outros vínculos podem ser prescritos para o vetor aleatório $\{X\}$. Suponha que m informações sobre $\{X\}$ sejam conhecidas e estejam escritas na forma:

$$\int_{\mathbb{R}^n} p(\{x\}) \{g_r(\{x\})\} d\{x\} = \{a_r\} \in \mathbb{R}^{\nu_r}, \quad r = 1, \dots, m. \quad (2.2.79)$$

onde os expoentes ν_1, \dots, ν_m são m inteiros iguais ou superiores a um.

Seja C_{ad} o espaço das funções $\{x\} \mapsto p(\{x\})$ do \mathbb{R}^n em $\mathbb{R}^{\geq 0}$, tendo todas o mesmo suporte conhecido $k_n \subset \mathbb{R}^n$ (com a possibilidade de que $k_n = \mathbb{R}^n$) e, atendendo aos vínculos definidos em (2.2.78) e (2.2.79).

Para descobrir qual a expressão da função $p \in C_{ad}$ que maximiza essa entropia S (2.2.77), um problema de otimização (2.2.80) deve ser resolvido.

$$\max_{p \in C_{ad}} S(p). \quad (2.2.80)$$

Para resolver este problema, $m+1$ multiplicadores de Lagrange são introduzidos $(\lambda_0 - 1) \in \mathbb{R}$, $\{\lambda_1\} \in \mathbb{R}^{\nu_1}, \dots, \{\lambda_m\} \in \mathbb{R}^{\nu_m}$ associados aos $m+1$ vínculos. A Lagrangiana, nesse caso, é:

$$\begin{aligned} L(p, \lambda_0, \{\lambda_1\}, \dots, \{\lambda_m\}) = & S(p) - (\lambda_0 - 1) \left(\int_{\mathbb{R}^n} p(\{x\}) d\{x\} - 1 \right) \\ & - \sum_{r=1}^m \langle \{\lambda_r\}, \left\{ \int_{\mathbb{R}^n} p(\{x\}) \{g_r(\{x\})\} d\{x\} - a_r \right\rangle_{\mathbb{R}^{\nu_r}} \end{aligned} \quad (2.2.81)$$

onde $\langle \{u\}, \{v\} \rangle_{\mathbb{R}^{\nu_r}} = u_1 v_1 + \dots + u_{\nu_r} v_{\nu_r}$ é o produto escalar euclidiano no \mathbb{R}^{ν_r} .

Reescreve-se (2.2.81) como:

$$\begin{aligned} L(p, \lambda_0, \{\lambda_1\}, \dots, \{\lambda_m\}) = & u(\lambda_0, \{\lambda_1\}, \dots, \{\lambda_m\}) \\ & - \int_{\mathbb{R}^n} h(p, \lambda_0, \{\lambda_1\}, \dots, \{\lambda_m\}) d\{x\}, \end{aligned} \quad (2.2.82)$$

com:

$$u(\lambda_0, \{\lambda_1\}, \dots, \{\lambda_m\}) = (\lambda_0 - 1) + \sum_{r=1}^m \langle \{\lambda_r\}, \{a_r\} \rangle, \quad (2.2.83)$$

e

$$\begin{aligned} h(p, \lambda_0, \{\lambda_1\}, \dots, \{\lambda_m\}) = & p(\{x\}) [\ln p(\{x\}) + (\lambda_0 - 1) + \\ & \sum_{r=1}^m \langle \{\lambda_r\}, \{g_r(\{x\})\} \rangle_{\mathbb{R}^{\nu_r}}] . \end{aligned} \quad (2.2.84)$$

Pelo cálculo variacional, os extremos de máximo e mínimo da Lagrangiana L verificam:

$$\frac{\partial}{\partial p(\{x\})} h(p, \lambda_0, \{\lambda_1\}, \dots, \{\lambda_m\}) = 0. \quad (2.2.85)$$

Resolvendo esse problema de otimização, chega-se ao seguinte resultado:

$$p(\{x\}) = \mathbb{1}_{k_n}(\{x\}) \exp\left(-\lambda_0 - \sum_{r=1}^m \langle \{\lambda_r\}, \{g_r(\{x\})\} \rangle_{\mathbb{R}^{\nu_r}}\right). \quad (2.2.86)$$

EXEMPLO 2.2.8 *Seja $\{X\}$ uma vetor aleatório contínuo, com dimensão n e com suporte compacto $k_n = [a_1, b_1] \times \dots \times [a_n, b_n]$. Deseja-se determinar a densidade de probabilidade p de $\{X\}$ que maximiza a entropia S (2.2.77), tal que:*

$$p(\{x\}) \geq 0, \quad \int_{\mathbb{R}^n} p(\{x\}) d\{x\} = 1. \quad (2.2.87)$$

Pela expressão (2.2.86), tem-se que:

$$p(\{x\}) = \mathbb{1}_{k_n}(\{x\}) \exp(-\lambda_0), \quad (2.2.88)$$

ou seja, p é uma função constante. Substituindo (2.2.88) no vínculo (2.2.87):

$$\int_{\mathbb{R}^n} \mathbb{1}_{k_n}(\{x\}) \exp(-\lambda_0) d\{x\} = 1 \quad (2.2.89)$$

$$\exp(-\lambda_0) \int_{\mathbb{R}^n} \mathbb{1}_{k_n}(\{x\}) d\{x\} = \exp(-\lambda_0) \int_{\mathbb{R}^n} \mathbb{1}_{k_n}(\{x\}) d\{x\} = 1 \quad (2.2.90)$$

$$\exp(-\lambda_0) |k_n| = 1 \implies \exp(-\lambda_0) = \frac{1}{|k_n|}, \quad (2.2.91)$$

com $|k_n| = \int_{\mathbb{R}^n} \mathbb{1}_{k_n}(\{x\}) d\{x\}$ sendo a medida de k_n . Substituindo (2.2.91) em (2.2.88), determina-se que:

$$p(\{x\}) = \mathbb{1}_{k_n}(\{x\}) \frac{1}{|k_n|}. \quad (2.2.92)$$

O valor de λ_0 é calculado igualando-se (2.2.88) a (2.2.92):

$$\lambda_0 = -\ln\left(\frac{1}{|k_n|}\right). \quad (2.2.93)$$

EXEMPLO 2.2.9 Seja $\{X\}$ uma vetor aleatório de dimensão n , com suporte $k_n = (0, +\infty) \times \cdots \times (0, +\infty) \subset \mathbb{R}^n$ e com vetor média conhecido $E[\{X\}] = \{\mu_X\} \in k_n$, ou seja:

$$E[\{X\}] = \begin{bmatrix} \mu_{X_1} \\ \vdots \\ \mu_{X_n} \end{bmatrix}. \quad (2.2.94)$$

A função densidade de probabilidade de $\{X\}$ construída pelo Princípio da Entropia Máxima é:

$$p(\{x\}) = p_{X_1}(x_1) \times \cdots \times p_{X_n}(x_n), \quad (2.2.95)$$

onde a densidade p_{X_j} da variável aleatória X_j é escrita como uma função exponencial (2.2.51):

$$p_{X_j}(x_j) = \mathbf{1}_{(0, \infty)}(x_j) \frac{1}{\mu_{X_j}} e^{-\frac{x_j}{\mu_{X_j}}}, \quad j = 1, \dots, n, \quad (2.2.96)$$

mostrando que, pelo Princípio da Entropia Máxima, as n variáveis aleatórias X_1, \dots, X_n são independentes.

EXEMPLO 2.2.10 Seja $\{X\}$ um vetor aleatório de dimensão n , com suporte $k_n = \mathbb{R}^n$, com vetor média conhecida $E[\{X\}] = \{\mu_X\} = (\mu_{X_1} \cdots \mu_{X_n})^T \in k_n$, onde $\mu_{X_j} = E[X_j]$ e com matriz de covariância $[C]$ real simétrica positiva definida conhecida.

A função densidade de probabilidade de $\{X\}$ construída pelo Princípio da Entropia Máxima é a densidade Gaussiana:

$$p_{\{X\}}(\{x\}) = \frac{1}{\sqrt{(2\pi)^n \det[C]}} \times \exp \left\{ -\frac{1}{2} \left\langle [C]^{-1}(\{x\} - \{\mu_X\}); (\{x\} - \{\mu_X\}) \right\rangle_{\mathbb{R}^n} \right\}. \quad (2.2.97)$$

2.2.4 Matrizes aleatórias

Nas seções anteriores do trabalho foi mostrado como o Princípio da Entropia Máxima (PEM) determina as funções densidades de probabilidade de variáveis e vetores aleatórios (casos em que apenas algumas informações sobre essas variáveis são conhecidas). Dessa forma, quando se faz uma abordagem estocástica paramétrica de um sistema, o PEM pode ser utilizado para construir as densidades de probabilidade dos parâmetros do sistema que são considerados aleatórios (e portanto representados por variáveis/vetores aleatórios).

Porém, essa abordagem paramétrica, em alguns problemas, revela-se não ser a mais adequada para modelar incertezas. Ao invés de se considerar que a incerteza está presente nos parâmetros do sistema, pode ser necessário considerá-la intrínseca ao modelo. Nesse caso, as incertezas estão nas matrizes que compõem o sistema de equações do modelo e, assim, o modelo é composto por matrizes aleatórias.

Igualmente ao feito nos casos das variáveis/vetores aleatórios, o PEM pode ser utilizado para construir as densidades de probabilidade de matrizes aleatórias. Esta abordagem foi introduzida recentemente por Soize (Cf. [46]) em 2000.

A notação utilizada para representar subconjunto de matrizes é:

- $\mathbb{M}_n(\mathbb{R})$: conjunto de todas as matrizes reais quadradas $n \times n$;
- $\mathbb{M}_n^S(\mathbb{R})$: conjunto de todas as matrizes reais simétricas $n \times n$;
- $\mathbb{M}_n^{+0}(\mathbb{R})$: conjunto de todas as matrizes reais simétricas positivo-semidefinidas $n \times n$;
- $\mathbb{M}_n^+(\mathbb{R})$: conjunto de todas as matrizes reais simétricas positivo-definidas $n \times n$;
- $\mathbb{M}_n^D(\mathbb{R})$: conjunto de todas as matrizes reais diagonais $n \times n$.

$$\mathbb{M}_n^D(\mathbb{R}) \subset \mathbb{M}_n^+(\mathbb{R}) \subset \mathbb{M}_n^{+0}(\mathbb{R}) \subset \mathbb{M}_n^S(\mathbb{R}) \subset \mathbb{M}_n(\mathbb{R}) \quad (2.2.98)$$

A norma de uma matriz:

$$\|A\| = \sup_{\|x\| \leq 1} \|[A]x\|, \quad x \in \mathbb{R}^n. \quad (2.2.99)$$

e a norma Frobenius (or Hilbert-Schmidt norm) é:

$$\|A\|_F^2 = \text{tr} \{ [A]^T [A] \} = \sum_{j=1}^n \sum_{k=1}^n [A]_{jk}^2 \quad (2.2.100)$$

tal que: $\|A\| \leq \|A\|_F \leq \sqrt{n} \|A\|$.

Seja $[A_n]$ uma matriz aleatória com valores em $\mathbb{M}_n^+(\mathbb{R})$ cuja distribuição cumulativa de probabilidade $P_{[A_n]} = p_{[A_n]}([A_n])$ $\tilde{d}A_n$ é definida através de uma função densidade de probabilidade desconhecida:

$$[A_n] \longmapsto p_{[A_n]}([A_n]), \quad \text{do } \mathbb{M}_n^+(\mathbb{R}) \quad \text{em } \mathbb{R}^+ = [0, +\infty) \quad (2.2.101)$$

com respeito ao elemento de volume:

$$\tilde{d}A_n = 2^{n(n-1)/4} \prod_{1 \leq i < j \leq n} d[A_n]_{ij}. \quad (2.2.102)$$

Esta função densidade de probabilidade desconhecida pode ser construída através do PEM utilizando somente as informações conhecidas.

Suponha que essas informações conhecidas sobre $[A_n]$ sejam:

1. $[A_n] \in \mathbb{M}_n^+(\mathbb{R})$ e $\int_{\mathbb{M}_n^+(\mathbb{R})} p_{[A_n]}([A_n]) \tilde{d}A_n = 1$;
2. o valor esperado de $[A_n]$ é conhecido, ou seja:

$$E([A_n]) = \int_{\mathbb{M}_n^+(\mathbb{R})} [A_n] p_{[A_n]}([A_n]) \tilde{d}A_n = [\underline{A}_n] .$$

3. $\int_{\mathbb{M}_n^+(\mathbb{R})} \ln \det [A_n] p_{[A_n]}([A_n]) \tilde{d}A_n = v$, com $|v| < +\infty$.

A entropia relacionada a $p_{[A_n]}$ é dada por:

$$S(p_{[A_n]}) = - \int_{\mathbb{M}_n^+(\mathbb{R})} p_{[A_n]}([A_n]) \ln p_{[A_n]}([A_n]) \tilde{d}A_n. \quad (2.2.103)$$

Seja C_{ad} o espaço das funções $[A_n] \mapsto p_{[A_n]}$ do $\mathbb{M}_n^+(\mathbb{R})$ em $\mathbb{R}^+ = [0, +\infty)$ atendendo aos vínculos definidos acima.

Para descobrir qual a expressão da função $p_{[A_n]} \in C_{ad}$ que maximiza essa entropia S (2.2.77), um problema de otimização (2.2.104) deve ser resolvido:

$$\max_{p \in C_{ad}} S(p). \quad (2.2.104)$$

Para simplificar o problema, são feitas algumas manipulações algébricas. A matriz positiva-definida $[\underline{A}_n]$ pode ser decomposta usando a decomposição de Cholesky:

$$[\underline{A}_n] = [\underline{L}_{A_n}]^T [\underline{L}_{A_n}], \quad (2.2.105)$$

sendo $[\underline{L}_{A_n}]$ uma matriz triangular superior. Consequentemente, a matriz aleatória $[A_n]$ pode ser escrita como:

$$[A_n] = [\underline{L}_{A_n}]^T [G_n] [\underline{L}_{A_n}], \quad (2.2.106)$$

onde $[G_n]$ é uma matriz aleatória com valores em $\mathbb{M}_n^+(\mathbb{R})$ tais que:

$$[\underline{G}_n] = E([G_n]) = [I_n]. \quad (2.2.107)$$

A distribuição cumulativa de probabilidade $P_{[G_n]} = p_{[G_n]}([G_n]) \tilde{d}G_n$ possui uma densidade de probabilidade com respeito ao elemento de volume: $\tilde{d}G_n =$

$2^{n(n-1)/4} \prod_{1 \leq i < j \leq n} d[A_n]_{ij}$ e, pode ser escrita como:

$$p_{[G_n]}([G_n]) = \mathbb{1}_{\mathbb{M}_n^+(\mathbb{R})}([G_n]) \times C_{G_n} \times (\det [G_n])^{(n+1)\frac{(1-\delta_A^2)}{2\delta_A^2}} \times \exp \left\{ -\frac{(n+1)}{2\delta_A^2} \text{tr}[G_n] \right\}, \quad (2.2.108)$$

onde a constante de normalização C_{G_n} vale:

$$C_{G_n} = \frac{(2\pi)^{-n(n-1)/4} \left(\frac{n+1}{2\delta_A^2} \right)^{n(n+1)(2\delta_A^2)^{-1}}}{\left\{ \prod_{j=1}^n \Gamma \left(\frac{n+1}{2\delta_A^2} + \frac{1-j}{2} \right) \right\}}. \quad (2.2.109)$$

O parâmetro de dispersão δ_A é dado por:

$$\delta_A = \left\{ \frac{E(\| [G_n] - [\underline{G}_n] \|_F^2)}{\| [\underline{G}_n] \|_F^2} \right\}^{\frac{1}{2}}, \quad (2.2.110)$$

com $0 < \delta < \sqrt{\frac{n+1}{n+5}}$.

Além disso, a matriz positiva - definida $[G_n]$ pode ser decomposta usando a decomposição de Cholesky:

$$[G_n] = [L_n]^T [L_n], \quad (2.2.111)$$

onde $[L_n]$ é uma matriz triangular superior com valores em $\mathbb{M}_n(\mathbb{R})$. Fazendo essa decomposição, as entradas da matriz triangular superior $[L_n]$ tornam-se variáveis aleatórias independentes, caracterizadas por:

- caso $j < j'$: $[L_n]_{jj'} = \sigma_n U_{jj'}$, onde $\sigma_n = \delta_A (n+1)^{-1/2}$ e $U_{jj'}$ é uma variável aleatória real Gaussiana com média zero e variância um.
- caso $j = j'$: $[L_n]_{jj} = \sigma_n \sqrt{2V_j}$, onde V_j é uma variável aleatória real positiva Gamma, com função densidade de probabilidade:

$$p_{V_j}(v) = \mathbb{1}_{\mathbb{R}^+}(v) \frac{1}{\Gamma \left(\frac{n+1}{2\sigma_A^2} + \frac{1-j}{2} \right)} v^{\left(\frac{n+1}{2\delta_A^2} - \frac{1+j}{2} \right)} e^{-v}. \quad (2.2.112)$$

2.3 Exercícios

EXERCÍCIO 2.3.1 *Seja X uma variável aleatória contínua real com suporte $(-\infty, \infty)$. Mostre que:*

- (a) caso somente o segundo momento de X seja conhecido e caso $E[X^2] = \text{var}(X) = \sigma^2$, a densidade de probabilidade p de X que maximiza a entropia de Shannon é Gaussiana com média zero e variância σ^2 ;
- (b) caso sejam conhecidos $E[X] = \mu$ e $E[(X - \mu)^2] = \sigma^2$, a densidade de probabilidade p de X que maximiza a entropia de Shannon é Gaussiana com média μ e variância σ^2 ;
- (c) caso somente a variância de X seja conhecida, a densidade de probabilidade p de X que maximiza a entropia de Shannon é Gaussiana com média μ arbitrária e variância σ^2 .

Mostre também que em todos esses casos, a máxima entropia, S , vale

$$\frac{1}{2} \ln(2\pi e \sigma^2)$$

e é independente da média μ .

EXERCÍCIO 2.3.2 Verifique que em todos os três casos do exercício anterior (2.3.1) a máxima entropia tende para infinito quando σ tende para o infinito. Verifique também que se somente a média de X for conhecida, o problema de otimização não possui solução.

EXERCÍCIO 2.3.3 Verifique que em todos os três casos do exercício (2.3.1) a máxima entropia tende para $-\infty$ quando $\sigma \rightarrow 0$.

EXERCÍCIO 2.3.4 Sejam X e Y duas variáveis aleatórias contínuas reais com suporte $[0, \infty)$. Suponha que X possui função densidade de probabilidade exponencial, p_X , e Y possui função densidade Gamma, p_Y , ambas com a mesma média μ . Mostre que a entropia de p_X é maior do que a entropia de p_Y .

Capítulo 3

Método de Monte Carlo e Geração de Amostras de Variáveis e Vetores Aleatórios

3.1 Método de Monte Carlo

Os Métodos de Monte Carlo são métodos numéricos baseados em amostragem randômica usados em aproximação. Aproximar significa fornecer um valor, a aproximação, uma cota do erro cometido e um algoritmo que forneça uma metodologia para redução do erro de aproximação conforme desejado. Evidentemente reduzir o erro pode implicar um custo de computação e tempo de cálculo maior.

O método foi criado na metade da década de 40 por Stanislaw Ulam ¹ e John Von Neumann ² para resolver problemas de difusão de nêutrons em Los Alamos (Cf.[44]) e, rapidamente se difundiu na engenharia, física e matemática. O desenvolvimento do método ocorre em paralelo com o do computador digital e uma descrição do desenvolvimento pode ser achada em Metropolis.

A capacidade do método de Monte Carlo em obter aproximações numéricas para problemas complexos o tornou uma importante ferramenta matemática. A maneira como é feita a implementação do método de Monte Carlo está intrinsecamente correlacionada com o problema no qual ele será aplicado. Porém, todas

¹Stanislaw Ulam (1909 – 1984). Renomado matemático húngaro.

²John Von Neumann (1903 – 1957). Renomado matemático húngaro de etnia judaica.

essas diferentes abordagens do método tendem a seguir um determinado padrão. Geralmente, deseja-se gerar realizações de um objeto estocástico (sujeito a alguma densidade de probabilidade) e posteriormente utilizar as amostras obtidas para aproximar alguma função de interesse. Assim, o método faz uso de amostragem aleatória como ferramenta para produzir observações sobre as quais se realizam inferências estatísticas para extrair informações sobre as observações.

O modelo estocástico utilizado nas simulações de Monte Carlo influencia de forma direta os seus resultados (2.1). Dessa forma, a construção desse modelo deve ser feita de forma cautelosa e sistemática. Uma opção para fazer isso é utilizar o Princípio da Máxima Entropia, como visto no segundo capítulo.

As etapas do método de Monte Carlo são tipicamente:

1. Gerar realizações de um objeto estocástico com uma determinada densidade de probabilidade.
2. Realizar cálculos determinísticos usando cada uma das amostras geradas.
3. Agregar os resultados dos cálculos determinísticos no resultado final desejado (estatísticas ou aproximações da densidade de probabilidade da resposta).
4. Fazer uma análise da convergência da resposta, ou seja, assegurar-se que os resultados são representativos e que alterando-se o número de realizações da variável ou vetor aleatório, as estatísticas da resposta permanecem dentro de uma margem de erro prescrita.

EXEMPLO 3.1.1 *Para melhor compreender essas etapas do método, considere, por exemplo, que deseja-se calcular uma aproximação para o número π através da relação de volume existente entre um cubo e uma esfera de raio r inscrita nesse cubo. Essa relação é mostrada na figura (3.1).*

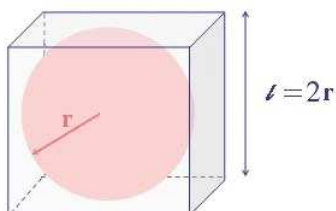
$$\frac{V_{\text{cubo}}}{V_{\text{esfera}}} = \frac{l^3}{\frac{4}{3}\pi r^3} = \frac{6}{\pi}$$


Figura 3.1: Relação de volume entre cubo e esfera inscrita no cubo.

Neste exemplo, as etapas do Método de Monte Carlo para aproximar π são:

1. Gerar m amostras distribuídas uniformemente dentro de um cubo de lado $l = 2r$ centrado na origem. Cada amostra i é um vetor composto por três coordenadas $\{x\}^{(i)} = (x_1^{(i)} \ x_2^{(i)} \ x_3^{(i)})^T$.

2. Contar quantas amostras estão dentro da esfera de raio r , ou seja quantas amostras atendem a condição: $x_1^2 + x_2^2 + x_3^2 < r^2$.
3. Calcular a razão R entre o número de amostras dentro da esfera e o número total de amostras m .
4. Aproximação: $\hat{\pi} = 6R$.
5. Verificar se o valor da aproximação obtida para π obtida está dentro de uma margem de erro prescrita.

O método de Monte Carlo é fundamentado em dois teoremas:

- Lei dos Grandes Números: garante a convergência das aproximações obtidas através do método;
- Teorema do Limite Central: especifica a forma dessa convergência.

Nesses dois fundamentos são utilizados os conceitos de convergência de sequências de variáveis aleatórias introduzidos no primeiro capítulo.

3.1.1 Lei dos grandes números

TEOREMA 3.1.2 (LEI DOS GRANDES NÚMEROS)

Seja X_1, X_2, \dots uma sequência de variáveis aleatórias independentes, cada uma delas com média μ e variância finita σ^2 . Seja a variável aleatória $S_n = X_1 + X_2 + \dots + X_n$. Tem-se que S_n/n converge em média quadrática para μ , ou seja:

$$\left\| \frac{S_n}{n} - \mu \right\| = \sqrt{E \left[\left(\frac{S_n}{n} - \mu \right)^2 \right]} \rightarrow 0, \text{ quando } n \rightarrow +\infty. \quad (3.1.1)$$

DEMONSTRAÇÃO.

Verifique que:

$$E \left[\frac{S_n}{n} \right] = \frac{1}{n} E[S_n] = \frac{1}{n} (E[X_1] + E[X_2] + \dots + E[X_n]) = \frac{1}{n} n\mu = \mu. \quad (3.1.2)$$

Assim:

$$E \left[\left(\frac{S_n}{n} - \mu \right)^2 \right] = \text{var} \left(\frac{1}{n} S_n \right). \quad (3.1.3)$$

Desenvolvendo a expressão da variância:

$$\begin{aligned}
E \left[\left(\frac{S_n}{n} - \mu \right)^2 \right] &= \text{var} \left(\frac{1}{n} S_n \right) = \frac{1}{n^2} \text{var}(X_1 + X_2 + \cdots + X_n) \\
&= \frac{1}{n^2} (\text{var}(X_1) + \cdots + \text{var}(X_n)) = \frac{1}{n^2} (n\sigma^2) = \frac{\sigma^2}{n}.
\end{aligned} \tag{3.1.4}$$

Dessa forma, tem-se que $E \left[\left(\frac{S_n}{n} - \mu \right)^2 \right] = \frac{\sigma^2}{n} \rightarrow 0$ quando $n \rightarrow +\infty$. ■

3.1.2 Teorema do limite central

TEOREMA 3.1.3 (TEOREMA DO LIMITE CENTRAL)

Seja X_1, X_2, \dots uma sequência de variáveis aleatórias independentes e identicamente distribuídas, cada uma delas com média μ e variância σ^2 . Seja $S_n = X_1 + X_2 + \cdots + X_n$. Pela Lei dos Grandes Números tem-se que: $E[S_n] = \mu n$ e $\text{var}(S_n) = \sigma^2 n$. Seja Z_n a variável aleatória:

$$Z_n = \frac{S_n - \mu n}{\sigma \sqrt{n}}, \tag{3.1.5}$$

com média nula e variância 1, ou seja: $E[Z_n] = 0$ e $\text{var}(Z_n) = 1$. Quando $n \rightarrow +\infty$, Z_n converge em probabilidade para uma variável aleatória Z com distribuição cumulativa de probabilidade gaussiana de média nula e variância um. Assim, sendo P_{Z_n} a função distribuição cumulativa de Z_n e P_Z a função distribuição cumulativa de Z :

$$P_{Z_n}(Z_n \leq x) \rightarrow P_Z(Z \leq x) \text{ quando } n \rightarrow +\infty, \forall x \text{ onde } P_Z \text{ é contínua.} \tag{3.1.6}$$

Pode-se dizer que:

$$\Pr(Z_n \leq x) \rightarrow \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{\alpha^2}{2}} d\alpha, \text{ quando } n \rightarrow +\infty. \tag{3.1.7}$$

EXEMPLO 3.1.4 Para ilustrar a lei dos grandes números e teorema do limite central, suponha que o método de Monte Carlo seja aplicado para obter uma aproximação para o valor de π , segundo o experimento descrito exemplo (3.1.1). Suponha também que esse experimento seja repetido $n = 10^3$ vezes, através de uma rotina MATLAB, utilizando-se em cada repetição $m = 10^4$ amostras distribuídas uniformemente dentro do cubo.

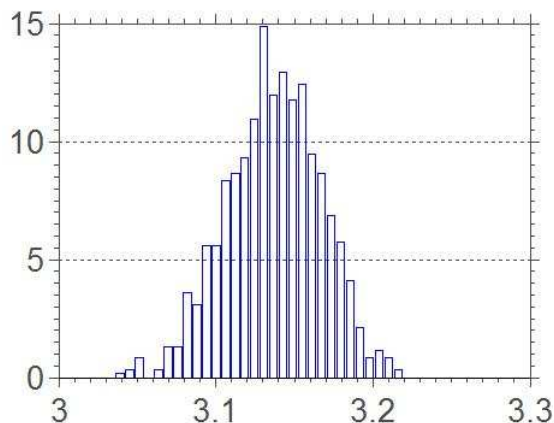


Figura 3.2: Histograma normalizado das $n = 10^3$ aproximações de π obtidas pelo método de Monte Carlo.

Ao final da simulação, obtém-se uma sequência de aproximações $\hat{\pi}_1, \dots, \hat{\pi}_n$. O gráfico de frequência, histograma normalizado, dessas aproximações é mostrado na figura (3.2).

Pode-se considerar que os experimentos formam uma sequência de variáveis aleatórias Π_1, Π_2, \dots independentes e identicamente distribuídas, com média μ e variância desconhecida σ^2 respectivamente. Aproximações para a média μ (que acredita-se ser π) e para a variância (que é desconhecida), podem ser obtidas através dos resultados dos experimentos $\hat{\pi}_1, \dots, \hat{\pi}_n$. Faz-se:

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n \hat{\pi}_i, \quad (3.1.8)$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (\hat{\pi}_i - \hat{\mu})^2. \quad (3.1.9)$$

Quando o número n de realizações é grande, como por exemplo, $n = 10^3$, supõe-se que $\hat{\mu}$ converge para π e que $\hat{\sigma}^2$ converge para o real valor de σ^2 . Na rotina MATLAB desenvolvida, com $n = 10^3$ repetições do experimento, os valores encontrados para $\hat{\mu}$ e $\hat{\sigma}^2$ foram respectivamente 3.1414 e 8.4325×10^{-4} .

Pela lei dos grandes números, sendo $S_n = \Pi_1 + \dots + \Pi_n$, tem-se que:

$$\mu_{\frac{S_n}{n}} = E \left[\frac{S_n}{n} \right] = \frac{1}{n} (E[\Pi_1] + E[\Pi_2] + \dots + E[\Pi_n]) = \mu = \pi, \quad (3.1.10)$$

$$\sigma^2_{\frac{S_n}{n}} = E \left[\left(\frac{S_n}{n} - \mu \right)^2 \right] = \frac{1}{n^2} (\text{var}(\Pi_1) + \dots + \text{var}(\Pi_n)) = \frac{\sigma^2}{n}. \quad (3.1.11)$$

Utilizando as aproximações $\hat{\mu}$ e $\hat{\sigma}^2$ calculadas através de (3.1.8) e (3.1.9), a média e variância de S_n/n podem ser aproximadas por:

$$\hat{\mu}_{\frac{S_n}{n}} = \hat{\mu}, \quad (3.1.12)$$

$$\hat{\sigma}^2_{\frac{S_n}{n}} = \frac{\hat{\sigma}^2}{n} = \frac{1}{n^2} \sum_{i=1}^n (\hat{\pi}_i - \hat{\mu})^2. \quad (3.1.13)$$

Assim, supondo que $\hat{\sigma}^2$ converge para o real valor de σ^2 quando $n = 10^3$, a média e variância de S_n/n valem:

$$\mu_{\frac{S_n}{n}} = \pi, \quad (3.1.14)$$

$$\sigma^2_{\frac{S_n}{n}} = \frac{\sigma^2}{n} = \frac{8.4325 \times 10^{-4}}{n}. \quad (3.1.15)$$

Na rotina MATLAB desenvolvida para esse exemplo, foram construídas as curvas $\hat{\sigma}^2_{\frac{S_n}{n}}$ (3.1.13) e $\sigma^2_{\frac{S_n}{n}}$ (3.1.15) em função do número de repetições n do experimento. O gráfico da figura (3.3) mostra os resultados obtidos. Verifica-se as duas curvas são de fato bastante próximas e, como esperado, com o aumento de n , $\hat{\sigma}^2_{\frac{S_n}{n}}$ decai com uma razão próxima a $1/n$.

Sendo Z_n a variável aleatória definida por:

$$Z_n = \frac{S_n - \mu n}{\sigma \sqrt{n}}, \quad (3.1.16)$$

pelo teorema do limite central, quando $n \rightarrow +\infty$, Z_n converge em probabilidade para uma variável aleatória Z com distribuição cumulativa de probabilidade gaussiana de média nula e variância um.

A partir de da tabela de probabilidades da distribuição cumulativa gaussiana $\mathcal{N}(0, 1)$, verifica-se que com probabilidade 0.954 uma amostra de $\mathcal{N}(0, 1)$ está no intervalo $(-2, 2)$. Assim:

$$\Pr \left(-2 < \frac{S_n - \mu n}{\sigma \sqrt{n}} < 2 \right) = 0.954. \quad (3.1.17)$$

Em outras palavras, com probabilidade 0.954, μ estará no intervalo:

$$\frac{S_n - 2\sigma\sqrt{n}}{n} < \mu < \frac{S_n + 2\sigma\sqrt{n}}{n}. \quad (3.1.18)$$

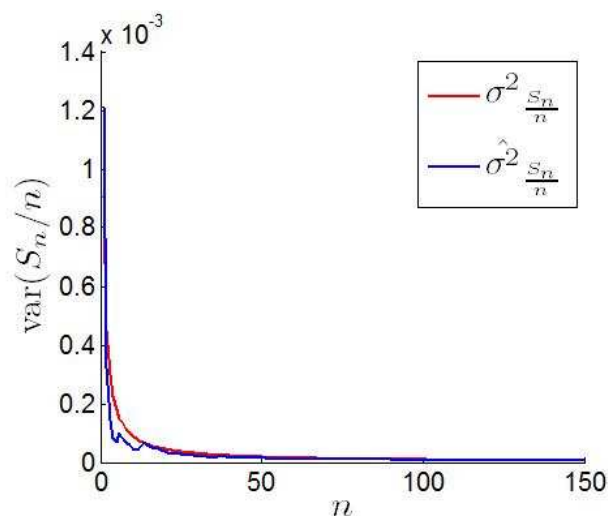


Figura 3.3: Variância de S_n/n em função do número de repetições n do experimento descrito em (3.1.).

É importante ressaltar que como neste exemplo a variância σ^2 era desconhecida e foi aproximada por $\hat{\sigma}^2$ (3.1.9), a variável aleatória Z_n não converge de fato em probabilidade para uma variável aleatória Z com distribuição cumulativa de probabilidade gaussiana de média nula e variância um.

3.1.3 Integração com o método de Monte Carlo

Como mencionado no início desse capítulo, um exemplo clássico do uso do método de Monte Carlo é a aproximação do cálculo de integrais multidimensionais em regiões com contornos complicados.

Suponha que deseja-se calcular a integral de uma dada função real multidimensional g em uma certa região $A \subset \mathbb{R}^d$,

$$I = \int_A g(\{x\}) \, d\{x\}. \quad (3.1.19)$$

Se g é uma função simples, sua integral, I , pode ser calculada facilmente. Por exemplo, no caso unidimensional, se $g(x) = x^3$ e $A = [0, 1]$ então, $I = \int_A g(x) \, dx = \left. \frac{x^4}{4} \right|_0^1 = \frac{1}{4}$.

Porém, se g é uma função mais difícil ou está definida em uma região com contorno complicado, pode não existir uma forma fechada para I . Algumas funções

unidimensionais em que isso ocorre são, por exemplo, $g(x) = \frac{\text{sen}(x)}{x}$, $g(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}$ e $g(x) = \sqrt{x}e^{-x}$, para $A = [a, b]$, $a < b < \infty$.

Nesses casos, métodos de integração numérica devem ser aplicados para obter-se aproximações para I , como por exemplo, método do trapézio, método de Simpson e método de Monte Carlo.

As vantagens do método de Monte Carlo sobre os outros métodos de integração numérica são:

- sua simplicidade de implementação;
- garantia de convergência da resposta (devido a Lei dos Grandes Números e teorema Central do Limite);
- eficiência no cálculo de aproximações de integrais multidimensionais.

Para mostrar como o método é aplicado, suponha que deseja-se calcular a integral multidimensional definida em (3.1.19). Suponhamos que ela seja escrita como:

$$I = \int_A g(\{x\}) d\{x\} = \int_A \frac{g(\{x\})}{p(\{x\})} p(\{x\}) d\{x\}, \quad (3.1.20)$$

onde p é a função densidade de probabilidade conjunta de um vetor aleatório $\{X\} \in \mathbb{R}^d$. Caso I seja uma integral unidimensional, p é a função densidade de probabilidade de uma variável aleatória, X .

Reescreve-se (3.1.20) como:

$$I = \int_A g(\{x\}) d\{x\} = \int_A h(\{x\}) p(\{x\}) d\{x\}, \quad (3.1.21)$$

onde $h(\{x\}) = g(\{x\})/p(\{x\})$, $\forall \{x\} \in A$. Como visto na equação (1.2.60) do primeiro capítulo, a integral I pode ser interpretada como o valor esperado de $h(\{X\})$, ou seja:

$$I = E[h(\{X\})] = \int_A h(\{x\}) p(\{x\}) d\{x\}. \quad (3.1.22)$$

Dessa forma, uma aproximação \hat{I} para a integral I é obtida através de:

$$\hat{I} = \sum_{i=1}^m h(\{x\}^{(i)}). \quad (3.1.23)$$

onde $\{x\}^{(1)}, \{x\}^{(2)}, \dots, \{x\}^{(m)}$ são amostras do vetor aleatório $\{X\}$ com função densidade de probabilidade conjunta p .

A seguir são mostrados alguns exemplos de integração com o método de Monte Carlo para funções unidimensionais e multidimensionais.

EXEMPLO 3.1.5 Para ilustrar o método de Monte Carlo aplicado em problemas de integração, suponha que deseja-se calcular a integral da função unidimensional:

$$g(x) = \frac{\text{sen}(x)}{x}, \quad (3.1.24)$$

no intervalo $A = [0, 3]$.

Uma primeira forma de fazer o cálculo, é escrever a integral na forma:

$$I = \int_A g(x) dx = \int_A \frac{g(x)}{p_{X_1}(x)} p_{X_1}(x) dx, \quad (3.1.25)$$

onde p_{X_1} é a função densidade de probabilidade de uma variável aleatória, X_1 , uniforme em $[0, 3]$, ou seja:

$$p_{X_1}(x) = \mathbb{1}_{[0,3]} \frac{1}{3}. \quad (3.1.26)$$

Assim, para obter-se uma aproximação \hat{I}_1 de I através do método de Monte Carlo, geram-se m amostras $x_1^{(1)}, x_1^{(2)}, \dots, x_1^{(m)}$ de X_1 e faz-se:

$$\hat{I}_1 = \sum_{i=1}^m \frac{g(x_1^{(i)})}{p_{X_1}(x_1^{(i)})} = \sum_{i=1}^m \frac{3 \text{sen}(x_1^{(i)})}{x_1^{(i)}}. \quad (3.1.27)$$

Uma outra opção para fazer o cálculo de uma aproximação, \hat{I}_2 , é escrever a integral como

$$I = \int_A g(x) dx = \int_A \frac{g(x)}{p_{X_2}(x)} p_{X_2}(x) dx, \quad (3.1.28)$$

onde p_{X_2} é a função densidade de probabilidade de outra variável aleatória, X_2 , definida por:

$$p_{X_2}(x) = \mathbb{1}_{[0,3]} \frac{e^{-x}}{1 - e^{-3}}. \quad (3.1.29)$$

Assim, para calcular-se \hat{I}_2 através do método de Monte Carlo, geram-se m amostras $x_2^{(1)}, x_2^{(2)}, \dots, x_2^{(m)}$ de X_2 e faz-se:

$$\hat{I}_2 = \sum_{i=1}^m \frac{g(x_2^{(i)})}{p_{X_2}(x_2^{(i)})} = \sum_{i=1}^m \frac{\text{sen}(x_2^{(i)})}{x_2^{(i)}} \frac{1 - e^{-15}}{5 e^{-5x_2^{(i)}}}. \quad (3.1.30)$$

Esse exemplo foi implementado em MATLAB e, aproximações da integral, \hat{I}_1 e \hat{I}_2 , foram calculadas utilizando-se diferentes número de amostras de X_1 e X_2 . A tabela (3.1) mostra os resultados obtidos.

Número de amostras	\hat{I}_1	\hat{I}_2
10^3	1.8221	1.8492
10^4	1.8504	1.8525
10^5	1.8552	1.8483
10^6	1.8490	1.8497
10^7	1.8490	1.8481

Tabela 3.1: Aproximações da integral I calculadas com diferentes número de amostras das variáveis aleatórias X_1 e X_2 .

EXEMPLO 3.1.6 Para ilustrar a integração de uma função multidimensional através do método de Monte Carlo, suponha que deseja-se calcular uma aproximação da integral da função:

$$g(\{x\}) = 1 + \frac{x_1^2 + x_2^2}{5}, \quad (3.1.31)$$

com $\{x\} = (x_1 \ x_2)^T \in \mathbb{R}^2$ na região $A = A_1 \times A_2 = [0, 2] \times [0, 2] \subset \mathbb{R}^2$.

Para isso, escreve-se a integral como:

$$I = \int_A g(\{x\}) \, d\{x\} = \int_A \frac{g(\{x\})}{p(\{x\})} p(\{x\}) \, d\{x\}, \quad (3.1.32)$$

onde p é a função densidade de probabilidade uniforme conjunta de um vetor aleatório $\{X\} \in \mathbb{R}^2$, ou seja:

$$p(\{x\}) = \mathbb{1}_{[0,2] \times [0,2]} \frac{1}{4}. \quad (3.1.33)$$

Assim, para obter-se uma aproximação \hat{I} de I através do método de Monte Carlo, geram-se m amostras $\{x\}^{(1)}, \{x\}^{(2)}, \dots, \{x\}^{(m)}$ de $\{X\}$ e faz-se:

$$\hat{I} = \sum_{i=1}^m \frac{g(\{x\}^{(i)})}{p(\{x\}^{(i)})} p(\{x\}^{(i)}). \quad (3.1.34)$$

Esse exemplo foi implementado em MATLAB e, aproximações da integral, \hat{I} , foram calculadas utilizando-se diferentes número de amostras de $\{X\}$. A tabela (3.2) mostra os resultados obtidos.

EXEMPLO 3.1.7 Suponha que deseja-se calcular aproximações do volume de uma esfera n -dimensional de raio $r = 1$ centrada na origem, ou seja deseja-se calcular:

$$I_{(n)} = \int_A 1 \, d\{x\}, \quad (3.1.35)$$

Número de amostras	\hat{I}
10^3	6.1655
10^4	6.1300
10^5	6.1315
10^6	6.1342
10^7	6.1333

Tabela 3.2: Aproximações da integral I calculadas com diferentes número de amostras do vetor aleatório $\{X\}$.

onde $A = \{ \{x\} \in \mathbb{R}^n \mid \|\{x\}\| \leq 1 \}$, através de dois métodos de integração numérica. Como a região A não é simples, para o cálculo de (3.1.35), utiliza-se a relação de volume entre o cubo n -dimensional e esfera n -dimensional inscrita nesse cubo. Considera-se um cubo de lado l . Assim, a razão de volume vale:

$$R_{(n)} = \frac{\text{Volume}_{\text{esfera}}}{\text{Volume}_{\text{cubo}}} = \frac{I_{(n)}}{l^n}, \quad (3.1.36)$$

podendo ser escrita como:

$$R_{(n)} = \frac{\int_A 1 \, d\{x\}}{l^n}. \quad (3.1.37)$$

O primeiro método para obter uma aproximação da integral, $\hat{I}_{(n)1}$, é determinístico. Para utilizá-lo, gera-se uma malha uniforme em todo o cubo n -dimensional. Cada dimensão do cubo é dividida em q intervalos, de forma que o cubo fica repartido em q^n cubinhos n -dimensionais. Após criada essa malha, verifica-se o centroide de cada cubinho i , representado pelo vetor $\{c\}^{(i)}$, atende a condição $\|\{c\}^{(i)}\| \leq 1$. Em caso afirmativo, considera-se que o cubinho está dentro da esfera n -dimensional de raio $r = 1$ e, em caso negativo, considera-se que o cubinho está fora. Somando os volumes de todos os cubinhos considerados dentro da esfera e dividindo pelo volume total do cubo n -dimensional, obtém-se uma aproximação para a razão de volumes $\hat{R}_{(n)1}$. Dessa forma:

$$\hat{R}_{(n)1} = \frac{\sum_{i=1}^{q^n} \mathbb{1}_{\|\{c\}^{(i)}\| \leq 1} (l/q)^n}{l^n}, \quad (3.1.38)$$

$$\hat{I}_{(n)1} = \hat{R}_{(n)1} l^n. \quad (3.1.39)$$

Na figura (3.4) é mostrado o cubo de dimensão $n = 3$ e lado $l = 2$ dividido em $q = 20^3$ cubinhos. Os cubinhos, cujo centroide atendem a condição $\|\{c\}^{(i)}\| \leq 1$, são representados em vermelho.

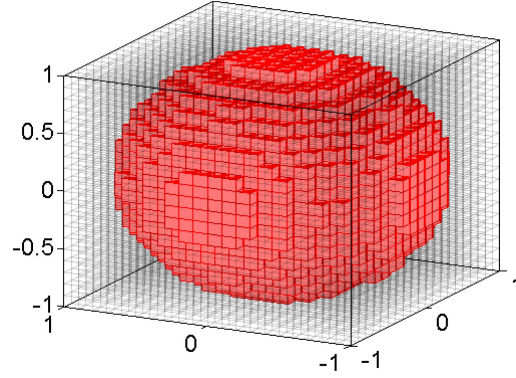


Figura 3.4: Malha uniforme em um cubo de dimensão $n = 3$ e lado $l = 2.0$.

Esse método fornece uma aproximação determinística e, quanto mais refinada for a malha, menor será o erro da aproximação. Porém, é importante observar que quando o número de dimensões, n , cresce, o número de vezes que a função $\mathbb{1}_{\|\{c^{(i)}\}\| \leq 1}$ tem que ser avaliada, q^n , cresce de forma exponencial. Dessa forma, o método em dimensões elevadas é apresenta um alto custo computacional.

Outra opção para obter-se uma aproximação, $\hat{I}_{(n)2}$, do volume da esfera n -dimensional, é usar-se o método de Monte Carlo. Para isso, são geradas m amostras de um vetor aleatório $\{X\} \in \mathbb{R}^n$ com função densidade de probabilidade conjunta uniforme na região $[-1, 1] \times \dots \times [-1, 1] \in \mathbb{R}^n$. São gerados, dessa forma, os vetores $\{x\}^{(1)}, \dots, \{x\}^{(m)}$.

Para cada vetor i gerado, verifica-se se $\|\{x\}^{(i)}\| \leq 1$. Em caso afirmativo, considera-se que amostra está dentro da esfera n -dimensional de raio $r = 1$ e, em caso negativo, considera-se que está fora. O número de amostras consideradas dentro da esfera dividido pelo número total de amostras, m , é uma aproximação para a razão de volumes $\hat{R}_{(n)2}$. Dessa forma:

$$\hat{R}_{(n)2} = \frac{\sum_{i=1}^m \mathbb{1}_{\|\{x\}^{(i)}\| \leq 1}}{m}, \quad (3.1.40)$$

$$\hat{I}_{(n)2} = \hat{R}_{(n)2} l^n. \quad (3.1.41)$$

O número de vezes que a função $\mathbb{1}_{\|\{x\}^{(i)}\| \leq 1}$ é avaliada no método de Monte Carlo não está associada com a dimensão n da esfera, diferentemente do método determinístico com a malha uniforme. Assim, em casos onde n é grande, a

integração através de Monte Carlo pode ser implementada de uma forma mais eficiente computacionalmente.

Para avaliar a qualidade das aproximações obtidas para o volume da esfera n -dimensional, faz-se a comparação com o volume exato, que é calculado através da seguinte fórmula:

$$I_{(n)} = \frac{\pi^{n/2}}{\Gamma(\frac{n}{2} + 1)} r^n, \quad (3.1.42)$$

onde o raio da esfera vale $r = 1$. Define-se erro da aproximação como:

$$\text{erro} = \frac{|I_{(n)} - \hat{I}_{(n)}|}{\hat{I}_{(n)}}. \quad (3.1.43)$$

Os dois métodos de aproximação foram implementados em MATLAB para os casos de cubo de lado $l = 2.0$ e esferas com raio $r = 1$. Para a construção da malha uniforme, cada lado do cubo foi dividido em $q = 20$ partes e, no método de Monte Carlo, foram utilizadas $m = 10^5$ amostras de $\{X\}$. Na tabela (3.3), são mostrados o erro da aproximação obtida e tempo de computação (em segundos) gasto em cada um dos dois métodos para diferentes números de dimensões n .

É importante observar que, quando n é pequeno, os erros e os tempos de computação gastos pelos dois métodos são próximos, de forma que o método de Monte Carlo não apresenta vantagens sobre o método da malha uniforme. Porém, quando o número de dimensões cresce, o tempo de computação do método da malha uniforme atinge valores muito elevados, enquanto o tempo do método de Monte Carlo varia muito pouco.

Dessa forma, quando trabalha-se em problemas com grande número de dimensões, usar Monte Carlo representa redução do custo computacional. Os resultados da tabela (3.3) são mostrados de forma gráfica nas figuras (3.5) e (3.6).

Observa-se também na tabela (3.3) que a razão entre o volume da esfera n -dimensional e do cubo n -dimensional reduz-se drasticamente com o aumento do número de dimensões, de forma que estimá-la não é uma tarefa fácil. Por esse motivo, quando o número de dimensões é muito grande o Método de Monte Carlo também encontra dificuldades para calcular uma aproximação de R . Sendo a razão de volumes muito pequena, a probabilidade de ter-se uma amostra localizada dentro da esfera é muito baixa. Assim, o que pode acontecer no método de Monte Carlo é ter-se todas as m amostras geradas fora da esfera.

n	Malha Uniforme		Monte Carlo		Vol. esfera	$R_{(n)}$
	Erro	Tempo	Erro	Tempo		
2	0.0059	0.0204	0.0013	0.0237	3.1416	0.7854
3	0.0084	0.0062	0.0020	0.0075	4.1888	0.5236
4	0.0061	0.0258	0.0054	0.0091	4.9348	0.3084
5	0.0076	0.7000	0.0102	0.0116	5.2638	0.1645
6	0.0158	17.550	0.0060	0.0144	5.1677	0.0807
7	0.0261	97.471	0.0054	0.0231	4.7248	0.0369
8	0.0005	6496.1	0.0369	0.0341	4.0587	0.0159
9	0.0334	48700	0.0609	0.0644	3.2985	0.0064
10	?	?	0.0042	0.0483	2.5502	0.0025
20	?	?	?	?	2.58×10^{-2}	2.46×10^{-8}
30	?	?	?	?	2.19×10^{-5}	2.04×10^{-14}
50	?	?	?	?	1.73×10^{-13}	1.53×10^{-28}
100	?	?	?	?	2.36×10^{-40}	1.86×10^{-70}

Tabela 3.3: Erro das aproximações e tempos de computação (em segundos) gastos nos métodos de Monte Carlo e malha uniforme para diferentes números de dimensões n .

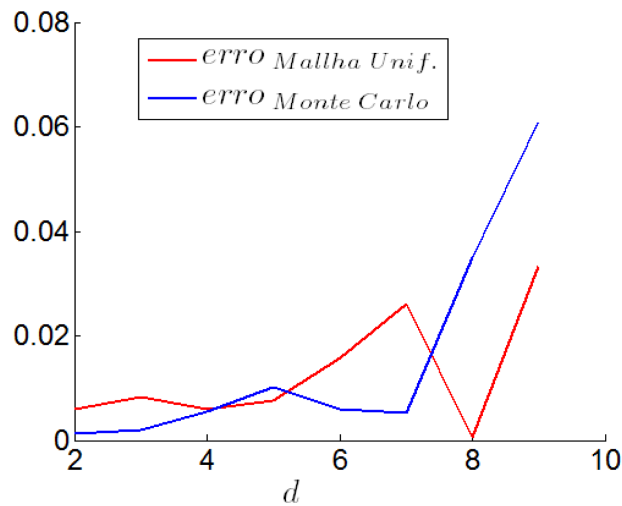


Figura 3.5: Erro da aproximação obtida para o volume da esfera em função do número de dimensões consideradas.

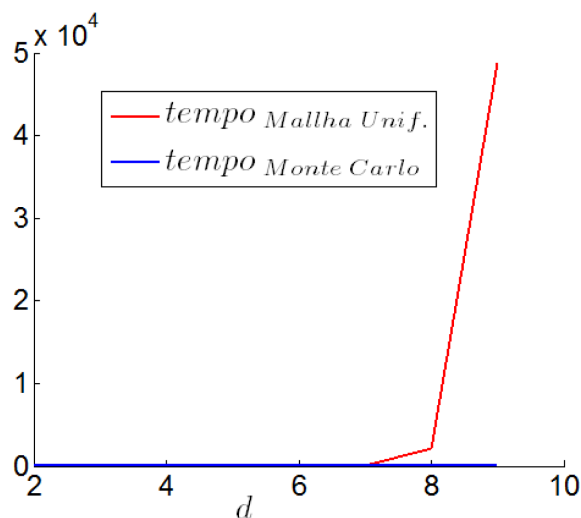


Figura 3.6: Tempo de computação gasto para o cálculo da aproximação do volume da esfera em função do número de dimensões consideradas.

3.2 Geração de amostras de variáveis e vetores aleatórios

Foi visto no início do capítulo, após serem construídos os modelos determinístico e estocástico (utilizando-se, por exemplo, o PEM) de um sistema, que o Método de Monte Carlo pode ser aplicado para obter-se inferências estatísticas sobre a resposta desse sistema (2.1).

Porém como o método tipicamente envolve a geração de observações de objetos estocásticos com alguma densidade de probabilidade, torna-se indispensável estudar-se geração de amostras:

- variáveis e vetores aleatórios;
- processos e campos estocásticos

para poder-se trabalhar com o método Monte Carlo.

Por exemplo, dado um sistema, suponha que o Princípio da Máxima Entropia tenha sido utilizado para determinar-se a densidade de probabilidade de parâmetros considerados aleatórios. Caso deseje-se aplicar o Método de Monte Carlo para esse problema, será necessário gerar amostras dos objetos estocásticos.

Quando deseja-se gerar amostras de uma variável aleatória real, X , com uma

densidade de probabilidade p , busca-se obter uma sequência de valores $\{x^{(i)}\}_{i \in \mathbb{N}}$, com $x_i \in \mathbb{R}, \forall i$ que representem realizações de X .

Muitos livros e artigos chamam os geradores de amostras de variáveis aleatórias frequentemente de geradores de números aleatórios (RNG-*Random Number Generator*). Abaixo segue uma lista com os principais requisitos de um RNG, espera-se que ele:

1. seja rápido, pois em suas aplicações podem ser necessário gerar milhões de amostras;
2. seja repetível, para tornar possível uma depuração do processo;
3. seja passível de análise, para poder-se estudar as propriedades dos gráficos de frequência construídos a partir das amostras geradas;
4. tenha um longo período;
5. seja aparentemente aleatório para a aplicação a qual é destinado.

No início do desenvolvimento de geradores, muitas técnicas híbridas de sistemas analógicos e digitais foram experimentadas. Como exemplo, tem-se os circuitos eletrônicos geradores de ruído branco. No entanto, estes métodos apresentavam muitas desvantagens: eram muito lentos, não repetíveis, tendiam a tornar-se tendenciosos e, exigiam equipamentos muito especializados. Por isso, essas técnicas foram sendo substituídas com o passar dos anos. Hoje, praticamente todos os geradores são baseados em algoritmos.

Como o computador é um dispositivo determinístico, pode parecer impossível que ele seja utilizado como um gerador. De fato, amostras geradas por computadores são obtidas através de algoritmos determinísticos. No entanto, elas parecem ser aleatórias e, para confirmar esse comportamento, deveriam ser submetidas a rigorosos testes. Muitas vezes, essas amostras são chamadas de '*pseudo-aleatórias*' (Cf. [44]).

Em (Cf. [20]), o autor Knuth descreve uma tentativa feita por ele de construir um gerador de amostras de variáveis aleatórias. Ele escreveu um algoritmo que trabalhava com números decimais de 10 dígitos, gerando uma sequência de valores $\{x^{(i)}\}_{i \in \mathbb{N}}$ a partir de um valor inicial.

Apesar de extenso, Knuth verificou que seu algoritmo podia convergir rapidamente para um ponto fixo. Além disso, testando-o, ele percebeu que, mesmo utilizando diferentes valores iniciais, a sequência de números começava rapidamente a se repetir.

Dessa forma, o algoritmo de Knuth é um excelente exemplo de que a uma grande complexidade em um algoritmo não garante que ele será um bom gerador de amostras de variáveis aleatórias. Muitas vezes, a complexidade esconde o comportamento de repetição dos números gerados. E assim, conclui-se que: *amostras*

de variáveis aleatórias não devem ser geradas com um método escolhido ao acaso (Cf. [20]).

Na sequência de amostras, $x^{(1)}, x^{(2)}, \dots, x^{(i)}, \dots$ com $x^{(i)} \in \mathbb{R}, \forall i \in \mathbb{N}$, produzida por geradores baseados em algoritmos, o termo $x^{(i)}$ é definido através de uma função f dependente de algumas variáveis e parâmetros:

$$x^{(i)} = f(\text{variáveis, parâmetros}) . \quad (3.2.44)$$

Cada vez que uma amostra é requisitada ao gerador, o algoritmo calcula a função f prescrita utilizando os parâmetros e valores atuais de suas variáveis. Geralmente, os parâmetros da função são valores fixos e, o que muda de amostra para amostra são as variáveis.

As variáveis podem ser quantidades externas ao algoritmo, ou valores armazenados internamente que mudam a cada iteração (nesse caso, as variáveis recebem o nome de sementes). Quando o valor da variável é atribuída por um processo externo, dificulta-se muito que a geração de uma sequência de números seja repetível. Por esta razão, a variável deve ser gerada internamente à função f . Uma exceção a essa regra é a atribuição do primeiro valor $x^{(0)}$.

Muitos algoritmos simplificam o cálculo dos números aleatórios, utilizando como semente para a geração o valor $x^{(i)}$ o próprio valor de $x^{(i-1)}$.

Vários métodos de geração de amostras de variáveis e vetores aleatórios foram desenvolvidos ao longo dos últimos anos. Como por exemplo, os geradores baseados em congruência linear, baseados no Método da Transformada Inversa, baseados em Cadeias de Markov (Monte Carlo Cadeia de Markov) entre outros.

3.2.1 Geradores baseados em congruência linear

Geradores baseados em Congruência linear são utilizados para gerar uma sequência de amostras de variáveis aleatórias com uma densidade de probabilidade uniforme.

Este gerador envolve três parâmetros fixos a , c e m e também um valor inicial $x^{(0)}$ (a semente). A sequência de números geradas pode ser definida como:

$$x^{(i+1)} = (a x^{(i)} + c) \bmod m, \quad \forall i \in \mathbb{N}. \quad (3.2.45)$$

Na operação $z = y \bmod m$, diz-se que z é congruente a y módulo m , se $z - y$ for divisível por m . Por exemplo, tomando-se: $a = 13$, $c = 0$, $m = 31$ e $x^{(0)} = 1$, os primeiros termos da sequência definida em (3.2.45) valem:

$$1, 13, 14, 27, 10, 6, 16, 22, 7, 29, 5, 3, \dots$$

Nesse exemplo, pode-se observar que os primeiros 30 termos da sequência são uma permutação dos inteiros de 1 a 30 e, a partir do trigésimo primeiro termo, a sequência começa a se repetir. Dessa forma, ela possui um período igual a $m - 1$.

Para obter-se uma sequência de valores distribuídos uniformemente no intervalo $[0, 1]$, basta dividir os valores obtidos por m . No exemplo acima, a sequência passa a ser:

$$0.032, 0.419, 0.451, 0.871, 0.322, 0.193, 0.516, \dots$$

Supondo que:

$$a \geq 2, \quad (3.2.46)$$

obtém-se uma generalização para (3.2.45) (Cf. [20]):

$$x^{(i+k)} = \left(a^k x^{(i)} + \frac{(a^k - 1)c}{a - 1} \right) \bmod m. \quad (3.2.47)$$

3.2.2 Parâmetros do gerador baseado em congruência linear

Nesta seção da dissertação, são apresentadas algumas importantes considerações na escolha dos valores dos parâmetros a , c e m do gerador de amostras de variáveis aleatórias baseado em congruência linear.

O maior número possível de diferentes valores gerados por um LCRNG é igual a m , o módulo. Por este motivo, em qualquer aplicação real, deseja-se que m seja o maior possível, por exemplo, 2^{32} . Aliás, uma escolha para m da forma 2^k , onde k é o tamanho da palavra do computador (*wordlength* - tamanho do registrador do processador), pode ser implementada muito eficientemente (Cf. [44]).

Essa eficiência está associada a forma como o computador, que é um dispositivo binário, calcula a divisão na operação do módulo. O cálculo do resto de uma divisão feita em notação binária (cujo divisor seja uma potência de 2) é muito simples. Basta alinhar pela direita o divisor e o dividendo representados em notação binária e, o resultado do resto da divisão é dado pelos algarismos do dividendo que correspondem aos zeros do divisor. Na tabela (3.4) é mostrado um exemplo.

	NOTAÇÃO DECIMAL	NOTAÇÃO BINÁRIA
Dividendo	13	1101
Divisor	4	100
Resto		01

Tabela 3.4: Cálculo do resto de uma divisão em representação binária para divisor sendo uma potência de 2.

Vale observar que as potências de 2 em representação binária são compostas apenas por um dígito 1 seguidos de zeros, como mostrado na tabela (3.5).

Dessa forma, observa-se que quando o módulo da expressão (3.2.45) é uma potência de 2, a conta do resto é bastante simplificada. Nesse caso, as operações necessárias ao cálculo são:

NOTAÇÃO DECIMAL	NOTAÇÃO BINÁRIA
2^0	1
2^1	10
2^2	100
2^3	1000
2^4	10000
\vdots	\vdots

Tabela 3.5: Potências de 2 em representação decimal binária.

1. copiar o dividendo ($a x^{(i)} + c$) da memória para o processador;
2. copiar o divisor (módulo m) da memória para o processador;
3. contar quantos zeros possui o divisor do bit menos significativo até encontrar o primeiro algarismo 1;
4. deslocar o dividendo para a direita de acordo com o tamanho do dividendo menos esse número de zeros e;
5. deslocar o dividendo de volta a posição original.

Este procedimento engloba cinco operações, sendo que esse número pode ainda crescer dependendo da arquitetura do processador do computador utilizado como gerador. Além disso, a cópia dos dados da memória para o processador é uma operação custosa, dado que o processador trabalha em frequências muito diferentes da frequência do barramento ao qual se conectam as memórias (limitações físicas das memórias *RAM*). Dessa forma, a economia da cópia de dados da memória para o processador reduziria significativamente o custo computacional do cálculo do resto de uma divisão.

Uma outra característica a ser observada é que operações no processador consomem ciclos de processamento e, portanto, tempo. Assim, quanto menos operações forem feitas com os dados no processador, em menos tempo um gerador de amostras de variáveis aleatórias será capaz de produzir uma sequência de valores.

A opção encontrada, mais simples e eficiente para minimizar a operação de cópia dos dados da memória para o processador, é utilizar como módulo uma potência de 2^k . Onde k é o tamanho da palavra do computador.

Tipicamente, as informações são armazenadas na memória do computador divididos em palavras de tamanho k . Dessa forma, por exemplo, um número de 8 bits 10110101 armazenado em uma memória com tamanho de palavra igual a 4 estaria na forma: 1011 0101. Assim, o resto da divisão desse número por 2^k ,

onde $k = 4$, é igual a palavra menos significativa de 1011 0101, ou seja 0101 (em representação binária).

Utilizando potências de 2^k para o módulo, a operação do cálculo do resto exige muito pouco custo computacional. Neste caso, o LCRNG precisa apenas o cálculo de uma multiplicação e uma adição. Vale observar se for desejado que os números gerados pelo LCRNG estejam uniformemente distribuídos no intervalo $[0, 1]$, deverá ser realizada uma divisão por m para cada valor calculado.

3.2.3 Outros geradores de amostras de variáveis aleatórias

Outros algoritmos, diferentes dos geradores de congruência linear, foram propostos nos últimos anos. Como exemplo, tem-se o algoritmo, criado na década de 50, baseado na sequência de *Fibonacci*. Nessa classe de geradores, o valor de um elemento $x^{(i+1)}$ depende não somente de $x^{(i)}$ e sim de mais de um valor anterior da sequência. Assim

$$x^{(i+1)} = (x^{(i)} + x^{(i-1)}) \bmod m. \quad (3.2.48)$$

Esse gerador, apesar de ter um período maior do que m , produz uma sequência de valores com uma forte relação entre termos consecutivos.

Geradores baseados na sequência de *Fibonacci* também foram desenvolvidos por Green Smith e Klem (Cf. [20]) no final da década de 50. A ideia era utilizar no cálculo de $x^{(i+1)}$ elementos anteriores, ou seja, $x^{(i-k)}$, sendo que k deve ser maior ou igual a 16. Assim:

$$x^{(i+1)} = (x^{(i)} + x^{(i-k)}) \bmod m, \quad k \geq 16. \quad (3.2.49)$$

Outros geradores de números foram propostos, como o algoritmo de G. J. Mitchell e D. P. Moore (Cf. [20]) de 1958 mostrado na equação (3.2.50):

$$x^{(i)} = (x^{(i-24)} + x^{(i-55)}) \bmod m, \quad i \geq 55, \quad (3.2.50)$$

onde m é par, e $x^{(0)} \dots, x^{(54)}$ são inteiros não todos pares. Mais um exemplo interessante é um gerador não linear que utiliza como semente para o termo $i + 1$, o inverso do termo $x^{(i-1)}$, isto é:

$$x^{(i+1)} = (a (x^{(i-1)})^{-1} + c) \bmod p, \quad \text{com } p \text{ primo.} \quad (3.2.51)$$

Knuth sugeriu em 1998 um outro tipo de gerador baseado em congruência não linear (Cf. [12]) no qual os termos anteriores ao elemento $x^{(i)}$ formam um polinômio de grau 2. Sendo $d \in \mathbb{R}$,

$$x^{(i+1)} = (d (x^{(i-2)})^2 + a x^{(i-1)} + c) \bmod m. \quad (3.2.52)$$

3.3 Método da Transformada Inversa

Até a presente seção, somente foram apresentados métodos para se gerar amostras de variáveis aleatórias com uma densidade de probabilidade uniforme. Porém, muitas vezes ao aplicar o método de Monte Carlo, é necessário simular o comportamento de variáveis aleatórias com outras densidades de probabilidade, como por exemplo, a densidade Gaussiana, Exponencial, Beta, etc (Cf. [12]). Por isso, nesta seção é mostrado como obter amostras dessas densidades diferentes da uniforme através do Método da Transformada Inversa.

Segundo esse método, uma amostra $x^{(i)}$ de uma variável aleatória X com distribuição cumulativa de probabilidade P contínua e estritamente monotônica, pode ser gerada através das seguintes etapas (Cf.[7, 38]):

1. gere uma amostra $u^{(i)}$ de uma variável aleatória U com densidade de probabilidade uniforme no intervalo unitário $[0, 1]$ (pode-se utilizar por exemplo algum gerador de congruência linear apresentado anteriormente);
2. faça $x^{(i)} = P^{-1}(u^{(i)})$.

A demonstração de que os valores gerados, $\{x^{(i)}\}_{i \in \mathbb{N}}$, são de fato amostras de X é mostrada a seguir.

Como já definido anteriormente, a distribuição cumulativa de probabilidade P de uma variável aleatória arbitrária X é não decrescente, mas não necessariamente estritamente monotônica e, não necessariamente contínua. Observe que P pode ter períodos de estabilidade ou descontinuidades. Contudo, sabe-se que:

$$\lim_{x \rightarrow -\infty} P(x) = 0 \quad \text{e} \quad \lim_{x \rightarrow \infty} P(x) = 1 . \quad (3.3.53)$$

Supondo que a função P é contínua e estritamente monotônica, tem-se que a função inversa P^{-1} existe e é contínua. Ela é definida como:

$$P^{-1}(y) = \inf \{x \mid P(x) \geq y, \quad 0 < y < 1\} . \quad (3.3.54)$$

Assim, para $-\infty < x < \infty$ e $0 < y < 1$:

$$P^{-1}(P(x)) = x \quad \text{e} \quad P(P^{-1}(y)) = y . \quad (3.3.55)$$

Sendo U uma variável aleatória uniforme no intervalo unitário $[0, 1]$, tem-se que:

$$\begin{aligned} Pr(P^{-1}(U) \leq x) &= Pr(P(P^{-1}(U)) \leq P(x)) \\ &= Pr(U \leq P(x)) . \end{aligned} \quad (3.3.56)$$

Como U é distribuída uniformemente no intervalo $[0, 1]$, a probabilidade $Pr(U \leq P(x))$ é igual ao tamanho do intervalo $[0, P(x)]$. Dessa forma:

$$Pr(P^{-1}(U) \leq x) = P(x), \quad (3.3.57)$$

o que justifica o método da Transformada Inversa.

A tabela (3.6) mostra algumas distribuições em que o método da Transformada Inversa pode ser aplicado.

Distribuição	pdf $p(x)$	Inversa $P^{-1}(u)$
Uniforme $\mathcal{U}(a, b)$ $a \leq x \leq b$	$\frac{1}{b-a}$	$a + (b - a)u$
Exponencial $\mathcal{E}(\lambda)$ $\lambda > 0; x \geq 0$	$\lambda e^{-\lambda x}$	$-\frac{1}{\lambda} \ln(1 - u)$
Beta $\mathcal{B}(\alpha, 1)$ $\alpha > 0; 0 \leq x \leq 1$	$\alpha x^{\alpha-1}$	$u^{\frac{1}{\alpha}}$
Beta $\mathcal{B}(1, \beta)$ $\beta > 0; 0 \leq x \leq 1$	$\beta (1 - x)^{\beta-1}$	$1 - (1 - u)^{\frac{1}{\beta}}$
Logística $\mathcal{L}(\alpha, \beta)$ $\beta > 0; -\infty < x; \alpha < \infty$	$\frac{e^{-(x-\alpha)/\beta}}{\beta[1+e^{-(x-\alpha)/\beta}]^2}$	$\alpha + \beta \ln[u/(1 - u)]$
Weibull $\mathcal{W}(\alpha, \beta)$ $\alpha, \beta > 0; x \geq 0$	$\frac{\alpha}{\beta} \left(\frac{x}{\beta}\right)^{\alpha-1} e^{-(x/\beta)^\alpha}$	$\beta[-\ln(1 - u)]^{1/\alpha}$
Cauchy $\mathcal{C}(\alpha, \beta)$ $\beta > 0; -\infty < x; \alpha < \infty$	$\frac{1}{\pi} \frac{\beta}{(x-\alpha)^2 + \beta^2}$	$\alpha + \beta \tan \pi[u - (1/2)]$

Tabela 3.6: Método da Transformada Inversa aplicado a algumas distribuições.

A função distribuição cumulativa de probabilidade de uma variável aleatória Gaussiana não possui uma expressão analítica fechada de sua inversa. Por isso, uma opção para gerar amostras desse tipo de variáveis é utilizar o método de Box-Muller proposto em 1958.

EXEMPLO 3.3.1 *Suponha que deseja-se gerar amostras de uma variável aleatória X com uma função densidade de probabilidade p dada por:*

$$p(x) = \begin{cases} 2x, & 0 \leq x \leq 1, \\ 0, & x < 0 \text{ ou } x > 1. \end{cases} \quad (3.3.58)$$

A função distribuição cumulativa de probabilidade P será:

$$P(x) = \begin{cases} 0, & x < 0, \\ \int_0^x 2y \, dy = x^2, & 0 \leq x \leq 1, \\ 1, & x > 1. \end{cases} \quad (3.3.59)$$

Aplicando o Método da Transformada Inversa, cada amostra $x^{(i)}$ de X é obtida por:

$$x^{(i)} = P^{-1}(u^{(i)}) = \sqrt{u^{(i)}}, \quad \forall i \in \mathbb{N}, \quad (3.3.60)$$

onde cada $u^{(i)}$ é uma amostra de uma variável aleatória uniforme, U , no intervalo unitário $[0, 1]$.

Para verificar a eficiência do método apresentado, foi implementada uma rotina MATLAB que compara o histograma normalizado das amostras de X (obtidas com o Método da Transformada Inversa) com a função densidade de probabilidade p . Nessa rotina, as amostras de U foram geradas através do comando `rand` do MATLAB. A figura (3.7) mostra os gráficos obtidos.

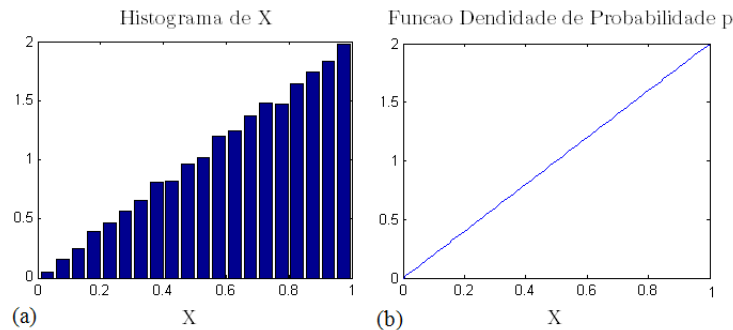


Figura 3.7: (a) Histograma normalizado construído a partir de 10^4 amostras de X . (b) Gráfico da função densidade de probabilidade p (3.3.58).

Na tabela (3.7) pode-se comparar as estatísticas (média, variância e desvio padrão) calculadas com diferentes números de amostras de X . Sendo a média de X igual a 0,6667, a variância 0.0556 e, o desvio padrão 0.2357, verifica-se que o Método da Transformada Inversa obtém boas aproximações para essas estatísticas.

Número de amostras	10^3	5×10^3	10^4
Média	0.6727	0.6634	0.6658
Variância	0.0565	0.0557	0.0554
Desvio padrão	0.2377	0.2360	0.2353

Tabela 3.7: Estatísticas das amostras de X obtidas pelo método da Transformada Inversa.

3.4 Geração de amostras de variáveis aleatórias gaussianas

Amostras de variáveis aleatórias Gaussianas podem ser geradas através do método de Box-Muller proposto em 1958 (Cf. [8]).

Nesse método, amostras de duas variáveis aleatórias Gaussianas, X_1 e X_2 , com média nula e variância um são obtidas através de amostras de duas variáveis aleatórias uniformes, U_1 e U_2 , no intervalo unitário $[0, 1]$.

Seja $\{X\} \in \mathbb{R}^2$ um vetor aleatório composto por duas variáveis aleatórias gaussianas independentes, X_1 e X_2 , de média nula e variância um. A função densidade de probabilidade conjunta de $\{X\}$ é o produto:

$$p(\{x\}) = \frac{1}{\sqrt{2\pi}} e^{-x_1^2/2} \times \frac{1}{\sqrt{2\pi}} e^{-x_2^2/2} = \frac{1}{2\pi} e^{-(x_1^2+x_2^2)/2}. \quad (3.4.61)$$

Como esta densidade conjunta é radialmente simétrica, X_1 e X_2 podem ser escritos em termos de duas variáveis aleatórias que representem as coordenadas polares: R e Θ , sendo o suporte de Θ o intervalo com $[0, 2\pi]$. Assim:

$$X_1 = R \cos(\Theta) \quad \text{e} \quad X_2 = R \sin(\Theta). \quad (3.4.62)$$

Sendo Θ uniformemente distribuído no intervalo $[0, 2\pi]$, amostras $\theta^{(i)}, i \in \mathbb{N}$, são obtidas por:

$$\theta^{(i)} = 2\pi u_1^{(i)}, \quad (3.4.63)$$

onde $u_1^{(i)}$ são amostras de uma variável aleatória uniforme, U_1 , no intervalo unitário $[0, 1]$. A probabilidade de R assumir um valor menor ou igual a $r \in \mathbb{R}$ pode ser escrito como:

$$Pr(R \leq r) = \int_{\Theta=0}^{2\pi} \int_{r'=0}^r \frac{1}{2\pi} e^{-r'^2/2} r' dr' d\theta = \int_{r'=0}^r e^{-r'^2/2} r' dr' . \quad (3.4.64)$$

Fazendo uma mudança de variáveis: $r'^2/2 = s$, $r' dr' = ds$, calcula-se:

$$Pr(R < r) = \int_{s=0}^{r^2/2} e^{-s} ds = 1 - e^{-r^2/2}. \quad (3.4.65)$$

A solução da distribuição de probabilidade (3.4.65) é assumir que:

$$R = \sqrt{-2 \ln(U_2)},$$

sendo U_2 uma variável aleatória uniforme no intervalo unitário $[0, 1]$.

Dessa forma, para gerar amostras $x_1^{(i)}$ e $x_2^{(i)}$, com $i \in \mathbb{N}$, de X_1 e X_2 :

1. geram-se as amostras $u_1^{(i)}$ de U_1 e $u_2^{(i)}$ de U_2 ;
2. calcula-se: $\theta^{(i)} = 2\pi u_1^{(i)}$ e $r^{(i)} = \sqrt{-2 \ln(u_2^{(i)})}$;
3. calcula-se: $x_1^{(i)} = r^{(i)} \cos(\theta^{(i)})$ e $x_2^{(i)} = r^{(i)} \sin(\theta^{(i)})$.

Vale observar que as amostras geradas de X_1 e X_2 (Gaussianas padrão, ou seja, média nula e variância um) podem ser transformadas em amostras de variáveis aleatórias Gaussianas Y_1 e Y_2 com médias μ_1 , μ_2 e variâncias σ_1^2 , σ_2^2 através de:

$$y_1^{(i)} = \mu_1 + \sigma_1 x_1^{(i)}, \quad (3.4.66)$$

$$y_2^{(i)} = \mu_2 + \sigma_2 x_2^{(i)}. \quad (3.4.67)$$

EXEMPLO 3.4.1 *Suponha que deseja-se gerar amostras de duas variáveis aleatórias Gaussianas Y_1 e Y_2 através do método de Box-Muller. A média de Y_1 vale $\mu_1 = 2,0$ e, sua variância $\sigma_1^2 = 2,0$. Para X_2 , $\mu_1 = 3,0$ e $\sigma_2^2 = 3,0$. Suas funções distribuições de probabilidade são:*

$$p_{Y_1}(y_1) = \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-(y_1-\mu_1)^2/2\sigma_1^2}, \quad (3.4.68)$$

$$p_{Y_2}(y_2) = \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-(y_2-\mu_2)^2/2\sigma_2^2}. \quad (3.4.69)$$

As amostras são geradas através de uma rotina implementada em MATLAB. Nessa rotina, as amostras das variáveis uniformes unitárias em $[0, 1]$, U_1 e U_2 , foram geradas com o comando `rand`. Os gráficos da figura (3.8) comparam os gráficos de frequência das amostras de Y_1 e Y_2 (obtidas com o Método de Box-Muller) com as função densidade de probabilidade p_{Y_1} e p_{Y_2} .

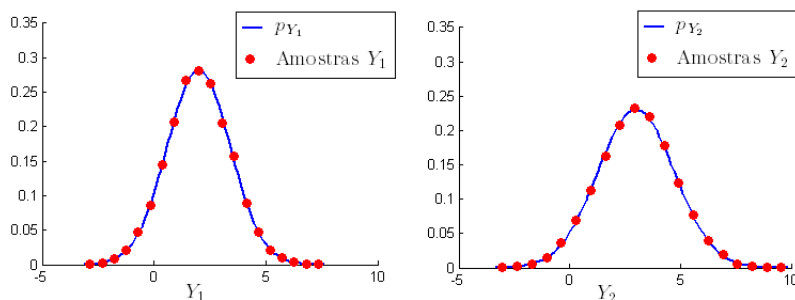


Figura 3.8: (a) Gráficos de frequência de 10^4 amostras de Y_1 . (b) Gráficos de frequência de 10000 amostras de Y_2 .

Na tabela (3.8) pode-se comparar as estatísticas (média, variância e desvio padrão) calculadas com diferentes números de amostras de Y_1 e Y_2 . Verifica-se que o Método de Box-Muller obtém boas aproximações para essas estatísticas.

	X_1			X_2		
N amostras	10^3	5×10^3	10^4	10^3	5×10^3	10^4
Média	2.0406	2.0109	1.9970	2.9658	3.0374	3.0030
Variância	1.9476	1.9709	1.9928	3.0270	2.9254	2.9461

Tabela 3.8: Estatísticas das amostras de X_1 e X_2 obtidas pelo método de Box-Muller.

Quando deseja-se simular o comportamento de uma variável aleatória através de uma sequência numérica, tem-se como principal objetivo a obtenção de amostras que de fato comportem-se de forma similar à variável aleatória.

Nas seções anteriores, discutiu-se os principais requisitos de um gerador, como por exemplo o tamanho do período do gerador. Porém, apesar do período ser um critério importante, não garante que os números obtidos com a sequência simulam se fato uma variável aleatória. Ou seja, se são 'suficientemente' aleatórios.

Existem testes que têm como objetivo verificar a não-aleatoriedade de uma sequência. Isto é, são capazes de mostrar que uma determinada sequência de

amostras não é aleatória. Vale observar porém, que esses testes não são capazes de revelar se a sequência é de fato aleatória. Exemplos de testes podem ser encontrados em (Cf. [20]).

A consideração de que uma sequência produzida por um gerador de números aleatórios é aleatória, está sempre associada a realização de um número finito n de testes T_1, T_2, \dots, T_n . Nada pode ser afirmado sobre o comportamento se a sequência quando ela for submetida a um teste T_{n+1} .

No entanto, cada avaliação feita permite que se tenha mais confiança sobre a real aleatoriedade da sequência.

3.5 Método da Rejeição

O método da Transformada Inversa, apresentado anteriormente, permite gerar amostras de uma variável aleatória X com densidade de probabilidade p diferente da uniforme. Entretanto, para aplicar esse método, é necessário:

- calcular a integral analítica da função p para obter a distribuição de probabilidade P e,
- inverter P , ou seja determinar P^{-1} .

Dessa forma, em problemas em que o cálculo da integral de p e da função inversa, P^{-1} , é difícil e em situações em que deseja-se gerar amostras de um vetor aleatório (cuja função densidade de probabilidade multidimensional é não inversível), o método da Transformada Inversa não pode ser aplicado. A geração deve ser feita assim, através de outros métodos e, uma opção é utilizar o método da rejeição proposto por von Neumann em 1951 (Cf. [22, 38, 39, 23]).

Esse método, explicado a seguir, é aplicável em problemas de geração de amostras de variáveis aleatórias e vetores aleatórios (com função densidade de probabilidade multidimensional).

3.5.1 Método da Rejeição aplicado a Variáveis Aleatórias

O método propõe gerar amostras de uma variável aleatória X , com densidade p , através de amostras de uma outra variável aleatória, Y , com densidade de probabilidade h . Vale observar que as duas variáveis, X e Y , devem estar definidas no mesmo suporte k .

A função densidade de probabilidade h é chamada de densidade proposta e definida pelo usuário do método. Cada amostra $y^{(j)}$ de Y é aceita ou rejeitada como uma amostra $x^{(i)}$ de X de acordo com um critério definido pelo método, tal que, no fim do processo de geração obtém-se uma sequência de amostras $x^{(i)}$ que respeitam a distribuição p .

No método da Rejeição, as amostras de X formam uma sequência de valores $x^{(i)}$, $i \in \mathbb{N}$ e, as amostras de Y formam uma sequência $u^{(j)}$, $j \in \mathbb{N}$.

Sejam p e h as funções densidades de X e Y respectivamente e, seja c uma constante real tal que:

$$\frac{p(x)}{c h(x)} \leq 1, \forall x \in k. \quad (3.5.70)$$

Uma amostra $x^{(i)}$ de X , pode ser gerada através das seguintes etapas:

1. gere uma amostra $y^{(j)}$ de Y com densidade de probabilidade h (pode-se utilizar por exemplo o método da Transformada Inversa apresentado anteriormente);
2. gere uma amostra $u^{(j)}$ de uma variável aleatória U com densidade de probabilidade uniforme no intervalo unitário $[0, 1]$ (pode-se utilizar por exemplo algum gerador de congruência linear apresentado anteriormente);
3. se $u^{(j)} \leq \frac{p(y^{(j)})}{c h(y^{(j)})}$, onde c é tal que a condição (3.5.70) seja respeitada, faz-se então $x^{(i)} = y^{(j)}$.

Para verificar que de fato a sequência gerada respeita a densidade p , basta mostrar que:

$$\begin{aligned} Pr\left(Y \leq x \mid U \leq \frac{p(y)}{c h(y)}\right) &= \frac{Pr\left(Y \leq x, U \leq \frac{p(y)}{c h(y)}\right)}{Pr\left(U \leq \frac{p(y)}{c h(y)}\right)} \\ \frac{\int_{-\infty}^x \int_0^{\frac{p(y)}{c h(y)}} 1 \cdot h(y) \, du \, dy}{\int_{-\infty}^{\infty} \frac{p(y)}{c h(y)} h(y) \, dy} &= \frac{\int_{-\infty}^x \frac{p(y)}{c h(y)} h(y) \, dy}{\frac{1}{c} \int_{-\infty}^{\infty} p(y) \, dy} \quad (3.5.71) \\ \frac{\frac{1}{c} \int_{-\infty}^x p(y) \, dy}{\frac{1}{c}} &= \int_{-\infty}^x p(y) \, dy = Pr(X \leq x), \end{aligned}$$

isto é, a probabilidade de $X \leq x$ é igual a probabilidade condicional de ter-se $Y \leq x$, dado que $U \leq p(y)/c h(y)$.

O número de iterações necessárias para se gerar uma amostra, $x^{(i)}$, de X é uma variável aleatória discreta, N , que assume os valores 1, 2, \dots com uma função de massa geométrica, isto é:

$$Pr(N = n) = (1 - q)^{n-1} q, \quad (3.5.72)$$

onde q é a probabilidade de sucesso, ou seja, a probabilidade de $y^{(j)}$ ser aceito como $x^{(i)}$. Assim:

$$q = \Pr \left(U \leq \frac{p(y)}{c h(y)} \right) = \int_{-\infty}^{\infty} \frac{p(y)}{c h(y)} h(y) dy = \frac{1}{c}. \quad (3.5.73)$$

Para aumentar a eficiência do método de geração, ou seja, ter-se o maior possível valor para q , c deve ser escolhido como o menor valor que respeita a condição (3.5.70).

EXEMPLO 3.5.1 *Suponha que deseja-se gerar amostras, $x^{(i)}$, de uma variável aleatória X com função densidade de probabilidade Beta:*

$$p(x) = \mathbb{1}_{[0,1]}(x) \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\int_0^1 u^{\alpha-1}(1-u)^{\beta-1} du}, \quad (3.5.74)$$

com parâmetros $\alpha = 2$ e $\beta = 2$. Para isso, utiliza-se o método da Rejeição, de forma que os $x^{(i)}$ são gerados através de amostras de uma outra variável aleatória, Y , com densidade de probabilidade, h , uniforme no intervalo $[0, 1]$.

O máximo da função p ocorre em $x = 0.5$ e vale 1.5. Assim, o menor valor possível para c que respeita a condição (3.5.70) é $c = 1.5$.

Traçando a curva de densidade p e da função $c \cdot h$, como mostrado em (3.9), observa-se que cada amostra $y^{(j)}$ de Y e cada amostra $u^{(j)}$ de U multiplicada por c forma um ponto de coordenadas cartesianas $(y^{(j)}, 1.5 \cdot u^{(j)})$ dentro do retângulo de base 1 e altura 1.5.

O valor $y^{(j)}$ é aceito como $x^{(i)}$ se:

$$c \cdot u^{(j)} \leq \frac{p(y^{(j)})}{h(y^{(j)})} = p(y^{(j)}), \quad (3.5.75)$$

isto é, se a ordenada, $1.5 \cdot u^{(j)}$, do ponto $(y^{(j)}, 1.5 \cdot u^{(j)})$ for menor do que $p(y^{(j)})$. Verifica-se assim, que $y^{(j)}$ é aceito como amostra de X se formar um ponto $(y^{(j)}, 1.5 \cdot u^{(j)})$ localizado abaixo da curva p .

Esse exemplo foi implementado em MATLAB para o caso de $m = 10^5$ amostras de X . O histograma normalizado construído a partir das amostras geradas e a densidade de probabilidade desejada p são mostrados na figura (3.10).

EXEMPLO 3.5.2 *Suponha que deseja-se gerar amostras de uma variável aleatória X com função densidade de probabilidade Gamma:*

$$p(x) = \mathbb{1}_{[0,+\infty)}(x) \frac{1}{\mu} \left(\frac{1}{\delta^2} \right)^{\frac{1}{\delta^2}} \frac{1}{\Gamma(1/\delta^2)} \left(\frac{x}{\mu} \right)^{\frac{1}{\delta^2}-1} \exp \left(-\frac{x}{\delta^2 \mu} \right), \quad (3.5.76)$$

onde:

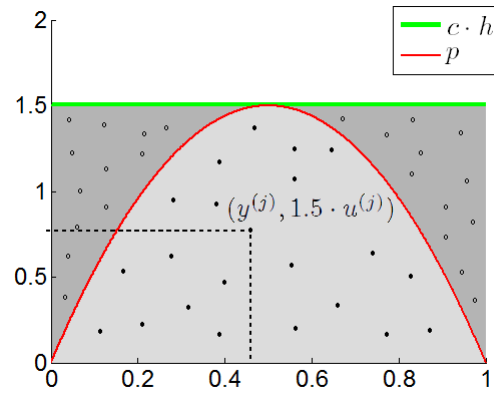


Figura 3.9: Função densidade de probabilidade Beta (com parâmetros $\alpha = 2$ e $\beta = 2$) e função densidade de probabilidade uniforme em $[0, 1]$ multiplicada pela constante $c = 1.5$.

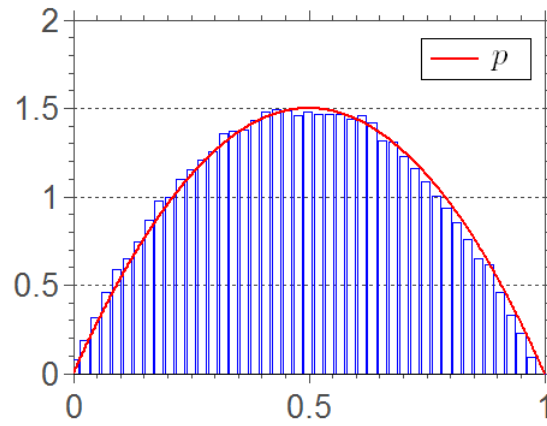


Figura 3.10: Histograma normalizado construído com 10^5 amostras de X , geradas através do método da Rejeição, e densidade de probabilidade Beta, p .

- Γ é a função Gamma: $\Gamma(a) = \int_0^{\infty} t^{a-1} \exp(-t) dt$;
- $\delta = \frac{\sigma}{\mu}$ é um fator de dispersão adimensional, com $\mu = E[X]$ e $\sigma = \sqrt{E[(X - \mu)^2]}$.

Utiliza-se o método da Rejeição e, como função densidade proposta utiliza-se a densidade exponencial, h , com $\lambda = 0.5$:

$$h(x) = \mathbb{1}_{[0,+\infty)}(x) \lambda e^{-\lambda x}. \quad (3.5.77)$$

O máximo da função p ocorre em $x = \delta^2 \mu \left(\frac{1}{\delta^2} - 1 \right) = 2.81$ e vale 0.55. Assim, o menor valor possível para c que respeita a condição (3.5.70) é $c = 4.45$.

Traçando a curva de densidade p e da função $c \cdot h$, como mostrado em (3.11), observa-se que cada amostra $y^{(j)}$ de Y e cada amostra $u^{(j)}$ de U multiplicada por $c \cdot h(y^{(j)})$ forma um ponto de coordenadas cartesianas $(y^{(j)}, 4.45 \cdot u^{(j)} \cdot h(y^{(j)}))$ localizado abaixo da curva $c \cdot h$.

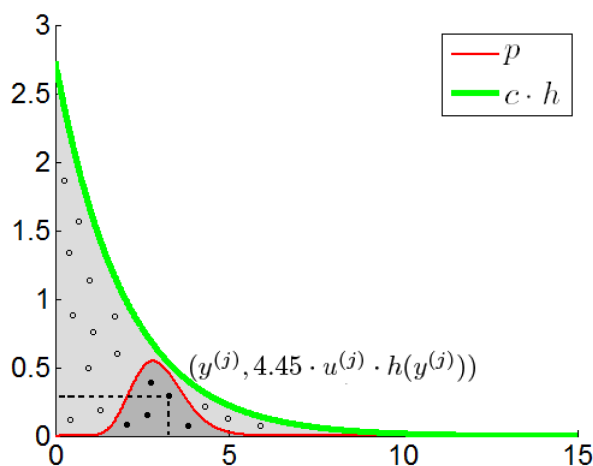


Figura 3.11: Função densidade de probabilidade Gamma (com média 3.0 e desvio padrão 0.25) e função densidade de probabilidade exponencial (com média 2.0) multiplicada pela constante $c = 4.45$.

O valor $y^{(j)}$ é aceito como $x^{(i)}$ se:

$$c \cdot u^{(j)} \leq \frac{p(y^{(j)})}{h(y^{(j)})} \implies c \cdot u^{(j)} \cdot h(y^{(j)}) \leq p(y^{(j)}), \quad (3.5.78)$$

isto é, se a ordenada, $4.45 \cdot u^{(j)} \cdot h(y^{(j)})$, do ponto $(y^{(j)}, 4.45 \cdot u^{(j)} \cdot h(y^{(j)}))$ for menor do que $p(y^{(j)})$. Verifica-se assim, que $y^{(j)}$ é aceito como amostra de X se formar um ponto $(y^{(j)}, 4.45 \cdot u^{(j)} \cdot h(y^{(j)}))$ localizado abaixo da curva p .

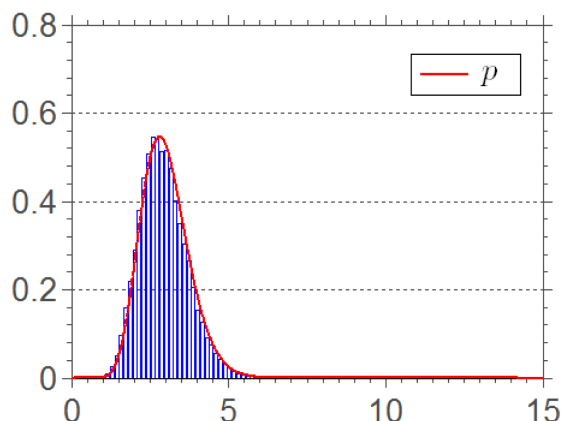


Figura 3.12: Histograma normalizado construído com 10^5 amostras de X , geradas através do método da Rejeição, e densidade de probabilidade Gamma, p .

Esse exemplo foi implementado em MATLAB para o caso de $m = 10^5$ amostras de X . O histograma normalizado construído a partir das amostras geradas e a densidade de probabilidade desejada p são mostrados na figura (3.10).

A probabilidade de $y^{(j)}$ ser aceito como $x^{(i)}$ vale:

$$q = \Pr\left(U \leq \frac{p(y)}{c h(y)}\right) = \int_{-\infty}^{\infty} \frac{p(y)}{c h(y)} h(y) dy = \frac{1}{c} = 0.22, \quad (3.5.79)$$

Assim, a variável aleatória discreta N , que representa o número de iterações necessárias para se gerar uma amostra, $x^{(i)}$, de X , possui a seguinte função de massa geométrica:

$$\Pr(N = n) = (1 - q)^{n-1} q = (0.78)^{n-1} 0.22. \quad (3.5.80)$$

O histograma normalizado construído a partir das amostras, $n^{(i)}$, de N geradas durante a simulação MATLAB e a função de massa geométrica (3.5.80) são mostrados na figura (3.13).

3.5.2 Método da Rejeição aplicado a Vetores Aleatórios

Como dito no início da seção, o método da Rejeição pode ser aplicado para a geração de amostras de variáveis e vetores aleatórios.

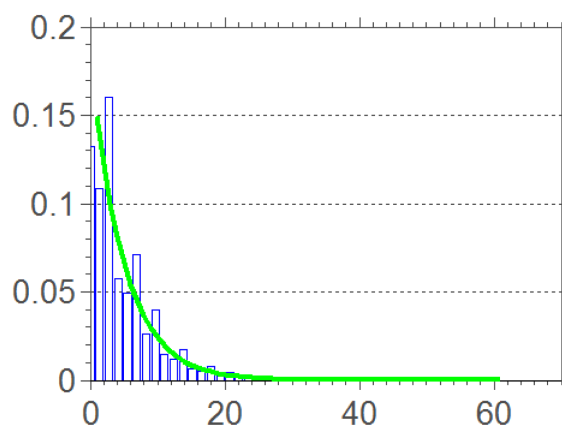


Figura 3.13: Histograma normalizado construído com 10^5 amostras de N .

EXEMPLO 3.5.3 *Suponha que deseja-se gerar amostras de um vetor aleatório $\{X\} = (X_1 \ X_2)^T$ com suporte $k = [0, 1] \times [0, 1] \subset \mathbb{R}^2$ e função densidade de probabilidade conjunta:*

$$p(\{x\}) = \mathbb{1}_k(\{x\}) 2.5e^{-x_1-x_2}. \quad (3.5.81)$$

Para isso, utiliza-se o método da Rejeição, de forma que as amostra $\{x\}^{(i)} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}^{(i)}$ são geradas através de amostras de um outro vetor aleatório, $\{Y\} = (Y_1 \ Y_2)^T$, composto pelas variáveis aleatórias independentes Y_1 e Y_2 . Esse vetor possui densidade de probabilidade conjunta uniforme:

$$h(\{x\}) = \mathbb{1}_{[0,1]}(x_1) \times \mathbb{1}_{[0,1]}(x_2) = \mathbb{1}_k(\{x\}), \quad (3.5.82)$$

isto é, h é o produto de duas funções densidades de probabilidade uniformes em $[0, 1]$.

O máximo da função p ocorre em $x = 0$, $y = 0$ e vale 2.5. Assim, o menor valor possível para c que respeita a condição (3.5.70) é $c = 2.5$. A figura (3.14) mostra as funções p e h .

Esse exemplo foi implementado em MATLAB para o caso de $m = 2 \times 10^5$ amostras de $\{X\}$. O histograma normalizado construído a partir das amostras geradas é mostrado na figura (3.15).

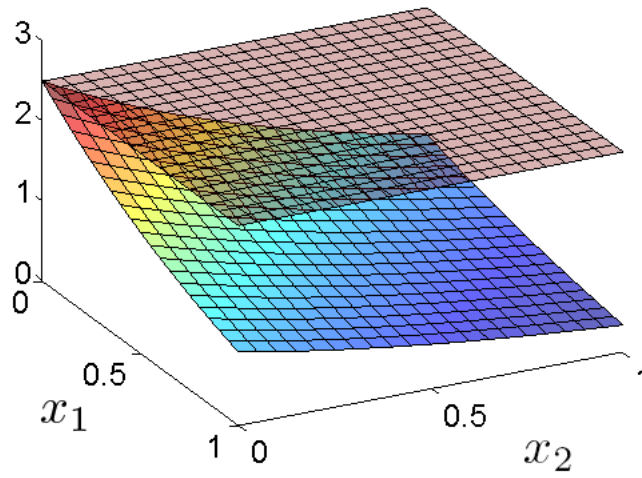


Figura 3.14: Função densidade de probabilidade conjunta p e função densidade de probabilidade proposta conjunta, h , multiplicada pela constante $c = 2.5$.

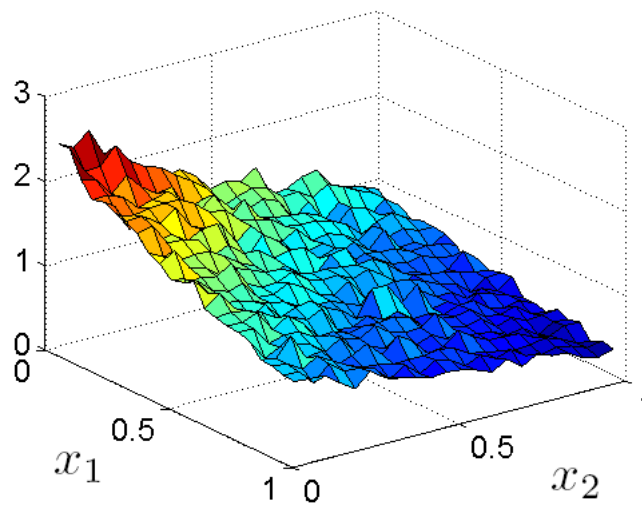


Figura 3.15: Histograma normalizado construído com 2×10^5 amostras de $\{X\}$, geradas através do método da Rejeição.

3.6 Exercícios

EXERCÍCIO 3.6.1 *Implemente uma rotina para calcular uma aproximação para o valor de π , segundo o experimento descrito exemplo (3.1.1).*

EXERCÍCIO 3.6.2 *Implemente uma rotina para calcular uma aproximação para a integral das funções,*

$$g(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}} \quad e \quad g(x) = \sqrt{x}e^{-x},$$

no intervalo $k = [1, 5]$, através do método de integração com Monte Carlo.

EXERCÍCIO 3.6.3 *Implemente uma rotina para calcular o volume de uma esfera n -dimensional através dos métodos de Monte Carlo e da malha uniforme, como descrito no exemplo (3.1.7). Faça a dimensão da esfera valer 4, 5, 6 e 9. Calcule em cada caso o erro da aproximação e o tempo computacional gasto para o cálculos.*

EXERCÍCIO 3.6.4 *Implemente uma rotina para gerar amostras de uma variável aleatória U com densidade de probabilidade uniforme no intervalo $[1, 3]$ através de um gerador baseado em congruência linear. Gere 10^2 amostras de U e posteriormente aumente esse número para 10^3 e 10^4 . Construa o histograma de cada uma das sequências de amostras.*

EXERCÍCIO 3.6.5 *Implemente uma rotina para gerar amostras de uma variável aleatória discreta X com a seguinte função densidade de probabilidade exponencial:*

$$p(x) = \mathbb{1}_{[0, \infty)} 5e^{-5x},$$

através de um gerador baseado no método da Transformada Inversa. Gere 10^2 , 10^3 e 10^4 amostras de X e construa o histograma de cada uma das sequências de amostras.

EXERCÍCIO 3.6.6 *Implemente uma rotina para gerar amostras de uma variável aleatória discreta X com a seguinte função densidade de probabilidade:*

$$p(x) = \mathbb{1}_{[0, 1)} 3x^{3-1},$$

através de um gerador baseado no método da Transformada Inversa. Gere 10^2 , 10^3 e 10^4 amostras de X e construa o histograma de cada uma das sequências de amostras.

EXERCÍCIO 3.6.7 *Implemente uma rotina para gerar amostras de uma variável aleatória X com densidade de probabilidade gaussiana (com média 5 e variância 0.25) através de um gerador baseado no método de Box-Muller. Gere 10^2 , 10^3 e 10^4 amostras de X e construa o histograma de cada uma das sequências de amostras.*

EXERCÍCIO 3.6.8 *Implemente uma rotina para gerar amostras de uma variável aleatória X com densidade de probabilidade gaussiana (com média 3 e variância 0.1) através de um gerador baseado no método da Rejeição. Utilize como densidade de probabilidade proposta a função densidade Cauchy. Gere 10^3 , 10^4 e 10^5 amostras de X e construa o histograma de cada uma das seqüências de amostras.*

Capítulo 4

Método de Monte Carlo com Cadeia de Markov

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 (Cf. [45]) (caso de distribuições com função distribuição de probabilidade inversíveis), método da Rejeição, 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 densidade de probabilidade difícil de ser normalizada. Ou seja, a função tem a forma de uma determinada densidade de probabilidade, porém a constante pela qual deve ser multiplicada para se tornar de fato uma distribuição (a integral da função em seu suporte, k , deve ser igual a um) é difícil de ser determinada.
- necessitam gerar amostras de variáveis aleatórias com densidade 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, será analisado o método

de geração de amostras de variáveis e vetores aleatórios: Monte Carlo com Cadeia de Markov (MCMC).

4.1 Processos estocásticos e cadeias de Markov

No primeiro capítulo foi apresentada a definição de um processo estocástico. A seguir é mostrada a definição de um processo estocástico de Markov.

DEFINIÇÃO 4.1.1 (PROCESSO DE MARKOV)

Um processo de Markov é um processo estocástico \mathcal{X} , com parâmetro $t \in T$, no qual tendo-se $t_1 \leq t_2 \leq \dots \leq t_{n-1} \leq t_n$, tem-se (Cf. [28]):

$$\begin{aligned} Pr(\mathcal{X}(t_n, w) \leq x_n | \mathcal{X}(t_{n-1}, w) = x_{n-1}, \dots, \mathcal{X}(t_1, w) = x_1) = \\ = Pr(\mathcal{X}(t_n) \leq x_n | \mathcal{X}(t_{n-1}) = x_{n-1}), \end{aligned} \quad (4.1.1)$$

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

$$\mathcal{X}_1, \mathcal{X}_2, \mathcal{X}_3, \dots, \mathcal{X}_n, \dots, \quad \forall n \in \mathbb{N}. \quad (4.1.2)$$

DEFINIÇÃO 4.1.2 (CADEIA DE MARKOV)

Seja $[T]$ uma matriz $k \times k$ com elementos: $T_{i,j} : i, j = 1, \dots, k$. Um processo estocástico com parâmetro discreto $(\mathcal{X}_0, \mathcal{X}_1, \dots)$ 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 (Cf. [14])

$$\begin{aligned} P(\mathcal{X}_{n+1} = s_j | \mathcal{X}_n = s_i, \mathcal{X}_{n-1} = s_{i_{n-1}}, \dots, \mathcal{X}_1 = s_{i_1}, \mathcal{X}_0 = s_{i_0}) \\ = P(\mathcal{X}_{n+1} = s_j | \mathcal{X}_n = s_i) \\ = T_{i,j}. \end{aligned} \quad (4.1.3)$$

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:

$$T_{ij} \geq 0, \quad \forall i, j \in \{1, \dots, k\}, \quad (4.1.4)$$

e

$$\sum_{j=1}^k T_{ij} = 1, \quad \forall i \in \{1, \dots, k\}, \quad (4.1.5)$$

ou seja, o somatório dos elementos de qualquer linha de T é sempre igual a 1. A propriedade (4.1.4) está associada ao fato de que probabilidades são sempre não negativas e, a propriedade (4.1.5) ao fato de que:

$$\begin{aligned} P(\mathcal{X}_{n+1} = s_1 | \mathcal{X}_n = s_i) &+ P(\mathcal{X}_{n+1} = s_2 | \mathcal{X}_n = s_i) + \dots \\ &+ P(\mathcal{X}_{n+1} = s_k | \mathcal{X}_n = s_i) = 1. \end{aligned} \quad (4.1.6)$$

Outra importante característica (além da matriz de transição) de uma cadeia de Markov $(\mathcal{X}_0, \mathcal{X}_1, \dots)$ é sua distribuição de probabilidade inicial. Essa distribuição é representada por um vetor $\{\pi(0)\}$ de k componentes:

$$\begin{aligned} \{\pi(0)\} &= [\pi_1(0) \ \pi_2(0) \ \dots \ \pi_k(0)] \\ &= [P(\mathcal{X}_0 = s_1) \ P(\mathcal{X}_0 = s_2) \ \dots \ P(\mathcal{X}_0 = s_k)]. \end{aligned} \quad (4.1.7)$$

Cada componente $\pi_i(0)$ de $\{\pi(0)\}$ está associada a probabilidade da cadeia ter no instante inicial (\mathcal{X}_0) o estado s_i . Como $\{\pi(0)\}$ é uma distribuição de probabilidade, tem-se:

$$\sum_{i=1}^k \pi_i(0) = 1. \quad (4.1.8)$$

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 :

$$\begin{aligned} \{\pi(n)\} &= [\pi_1(n) \ \pi_2(n) \ \dots \ \pi_k(n)] \\ &= [P(\mathcal{X}_n = s_1) \ P(\mathcal{X}_n = s_2) \ \dots \ P(\mathcal{X}_n = s_k)], \end{aligned} \quad (4.1.9)$$

com:

$$\sum_{i=1}^k \pi_i(n) = 1. \quad (4.1.10)$$

DEFINIÇÃO 4.1.3 (CHAPMAN-KOLOMOGROV)

A distribuição de probabilidade de uma cadeia de Markov $(\mathcal{X}_1, \mathcal{X}_2, \dots)$, com um número finito 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 (Cf. [50]):

$$\{\pi(n)\} = \{\pi(0)\} [T]^n. \quad (4.1.11)$$

DEMONSTRAÇÃO.

Para todo $j \in \{1, \dots, k\}$, tem-se que:

$$\begin{aligned} \pi_j(n+1) &= P(\mathcal{X}_{n+1} = x_j) \\ &= \sum_{i=1}^k P(\mathcal{X}_n = s_i, \mathcal{X}_{n+1} = x_j) \\ &= \sum_{i=1}^k P(\mathcal{X}_n = s_i) \cdot P(\mathcal{X}_{n+1} = s_j | \mathcal{X}_n = s_i) \\ &= \sum_{i=1}^k \pi_i(n) T_{i,j}. \end{aligned} \quad (4.1.12)$$

Escrevendo (4.1.12) em forma matricial: $\{\pi(n+1)\} = \{\pi(n)\} [T]$. Assim, recursivamente:

$$\begin{aligned} \{\pi(n)\} &= \{\pi(n-1)\} [T] = \{\pi(n-2)\} [T][T] = \{\pi(n-2)\} [T]^2 \\ &\vdots \\ &= \{\pi(0)\} [T]^n. \end{aligned} \quad (4.1.13)$$

■

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.

DEFINIÇÃO 4.1.4 (CADEIA DE MARKOV IRREDUTÍVEL)

Uma cadeia de Markov $(\mathcal{X}_0, \mathcal{X}_1, \dots)$, com um número finito de estados, $S = \{s_1, \dots, s_k\}$, e com matriz de transição $[T]$, é dita ser irredutível se, para todo $s_i, s_j \in S$, $\exists n \in \mathbb{N}$ tal que $(T_{i,j})^n > 0$. Onde, $(T_{i,i})^n$ representa o elemento da diagonal i, i da matriz $[T]^n$.

DEFINIÇÃO 4.1.5 (PERÍODO DE UM ESTADO DA CADEIA DE MARKOV)

O período de um estado s_i , $s_i \in S$, de uma cadeia de Markov, $d(s_i)$, é definido como:

$$d(s_i) = \text{mdc} \left\{ n \geq 1 : (T^n)_{i,i} > 0 \right\}, \quad (4.1.14)$$

onde $(T_{i,i})^n$ representa o elemento da diagonal i, i da matriz $[T]^n$ e, $\text{mdc} \{a_1, \dots, a_n\}$ indica o máximo divisor comum de a_1, \dots, a_n . Quando $d(s_i) = 1$, então o estado s_i é dito não-periódico.

DEFINIÇÃO 4.1.6 (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 4.1.7 (DISTRIBUIÇÃO DE PROBABILIDADE ESTACIONÁRIA DE UMA CADEIA DE MARKOV)

Seja uma cadeia de Markov $(\mathcal{X}_0, \mathcal{X}_1, \dots)$ 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:

$$\pi_i^* \geq 0, \text{ para } i = 1, \dots, k, \quad \sum_{i=1}^k \pi_i^* = 1, \quad (4.1.15)$$

e

$$\{\pi^*\} [T] = \{\pi^*\}. \quad (4.1.16)$$

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

TEOREMA 4.1.8 A distribuição de probabilidade $\{\pi(n)\}$ (quando $n \rightarrow \infty$) de uma cadeia de Markov irreduzível e não-periódica converge para distribuição de probabilidade estacionária (Cf. [14]), ou seja:

$$\lim_{n \rightarrow \infty} \{\pi(n)\} = \{\pi^*\}, \quad (4.1.17)$$

$$\lim_{n \rightarrow \infty} \sum_{i=1}^k |\pi_i^* - \pi_i(n)| = 0. \quad (4.1.18)$$

DEFINIÇÃO 4.1.9 (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:

$$\pi_i T_{i,j} = \pi_j T_{j,i}. \quad (4.1.19)$$

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 4.1.10 *Suponha que a matriz de transição $[T]$ de uma cadeia de Markov com 4 possíveis estados seja:*

$$[T] = \begin{bmatrix} 1/2 & 1/2 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (4.1.20)$$

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 autovalor 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 (4.1).

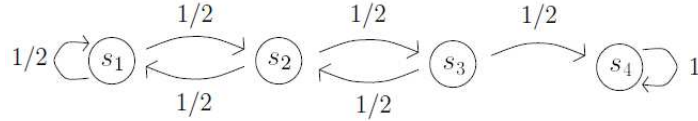


Figura 4.1: Grafo de Transição de uma cadeia de Markov com quatro estados (s_1, s_2, s_3 e s_4).

EXEMPLO 4.1.11 *Suponha que a matriz de transição $[T]$ de uma cadeia de Markov com 4 possíveis estados seja:*

$$[T] = \begin{bmatrix} 1/2 & 1/2 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 1/2 & 1/2 \end{bmatrix}. \quad (4.1.21)$$

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 (4.2).



Figura 4.2: Grafo de Transição de uma cadeia de Markov com duas distribuições estacionárias.

EXEMPLO 4.1.12 *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):

$$\begin{aligned} \{\pi(0)\} &= [\pi_1(0) \ \pi_2(0)] \\ &= [P(X_0 = s_1) \ P(X_0 = s_2)], \end{aligned} \quad (4.1.22)$$

com:

$$\pi_1(0) + \pi_2(0) = 1. \quad (4.1.23)$$

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:

$$[T] = \begin{bmatrix} 1 - p & p \\ q & 1 - q \end{bmatrix} \quad (4.1.24)$$

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:

$$\begin{aligned} P(\mathcal{X}_{n+1} = s_1) &= P(\mathcal{X}_n = s_1, \mathcal{X}_{n+1} = s_1) + P(\mathcal{X}_n = s_2, \mathcal{X}_{n+1} = s_1) \\ &= P(\mathcal{X}_n = s_1)P(\mathcal{X}_{n+1} = s_1 | \mathcal{X}_n = s_1) + \\ &\quad + P(\mathcal{X}_n = s_2)P(\mathcal{X}_{n+1} = s_1 | \mathcal{X}_n = s_2) \\ &= (1 - p) P(\mathcal{X}_n = s_1) + q P(\mathcal{X}_n = s_2), \end{aligned} \quad (4.1.25)$$

ou seja:

$$\pi_1(n+1) = T_{11} \pi_1(n) + T_{21} \pi_2(n). \quad (4.1.26)$$

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

$$\begin{aligned} P(\mathcal{X}_{n+1} = s_2) &= P(\mathcal{X}_n = s_1, \mathcal{X}_{n+1} = s_2) + P(\mathcal{X}_n = s_2, \mathcal{X}_{n+1} = s_2) \\ &= P(\mathcal{X}_n = s_1)P(\mathcal{X}_{n+1} = s_2 | \mathcal{X}_n = s_1) \\ &\quad + P(\mathcal{X}_n = s_2)P(\mathcal{X}_{n+1} = s_2 | \mathcal{X}_n = s_2) \\ &= p P(\mathcal{X}_n = s_1) + (1-q) P(\mathcal{X}_n = s_2), \end{aligned} \quad (4.1.27)$$

ou seja:

$$\pi_2(n+1) = T_{12} \pi_1(n) + T_{22} \pi_2(n). \quad (4.1.28)$$

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, $\{\pi(n)\}$:

$$\{\pi(n+1)\} = \{\pi(n)\} [T]. \quad (4.1.29)$$

De forma recursiva, pode-se determinar $\{\pi(n)\}$ em função de $\{\pi(0)\}$ (equação de Chapman-Kolomogrov (4.1.11)):

$$\{\pi(n)\} = \{\pi(0)\} [T]^n. \quad (4.1.30)$$

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

$$\begin{cases} \pi_1(n) = (1-p-q)^n \pi_1(0) + q \sum_{j=0}^{n-1} (1-p-q)^j \\ \pi_2(n) = (1-p-q)^n \pi_2(0) + p \sum_{j=0}^{n-1} (1-p-q)^j \end{cases} \quad (4.1.31)$$

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

- 1º caso: $p = q = 0$

$$\begin{cases} \pi_1(n) = \pi_1(0) \\ \pi_2(n) = \pi_2(0) \end{cases} \quad (4.1.32)$$

- 2º caso: $0 < p + q < 2$

$$\sum_{j=0}^{n-1} (1-p-q)^j = \frac{1 - (1-p-q)^n}{p+q}. \quad (4.1.33)$$

Assim, substituindo (4.1.33) em (4.1.31):

$$\begin{cases} \pi_1(n) = \frac{q}{p+q} + (1-p-q)^n \left[\pi_1(0) - \frac{q}{p+q} \right] \\ \pi_2(n) = \frac{p}{p+q} + (1-p-q)^n \left[\pi_2(0) - \frac{p}{p+q} \right] \end{cases} \quad (4.1.34)$$

Lembrando que:

$$\begin{cases} \pi_1(n) = T_{11}^n \pi_1(0) + T_{21}^n \pi_2(0) = T_{11}^n \pi_1(0) + T_{21}^n [1 - \pi_1(0)] \\ \pi_2(n) = T_{12}^n \pi_1(0) + T_{22}^n \pi_2(0) = T_{12}^n [1 - \pi_2(0)] + T_{22}^n \pi_2(0), \end{cases} \quad (4.1.35)$$

$$\begin{cases} \pi_1(n) = [T_{11}^n - T_{21}^n] \pi_1(0) + T_{21}^n \\ \pi_2(n) = [T_{22}^n - T_{12}^n] \pi_2(0) + T_{12}^n, \end{cases} \quad (4.1.36)$$

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

$$[T]^n = \frac{q}{p+q} \begin{bmatrix} q & p \\ q & p \end{bmatrix} + \frac{(1-p-q)^n}{p+q} \begin{bmatrix} p & -p \\ -q & q \end{bmatrix}. \quad (4.1.37)$$

Tomando-se o limite de (4.1.37) quando $n \rightarrow \infty$:

$$\lim_{n \rightarrow \infty} [T]^n = \frac{q}{p+q} \begin{bmatrix} q & p \\ q & p \end{bmatrix}. \quad (4.1.38)$$

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:

$$\{\pi^*\} = \lim_{n \rightarrow \infty} \{\pi(n)\} = \{\pi(0)\} \lim_{n \rightarrow \infty} [T]^n. \quad (4.1.39)$$

Assim:

$$\{\pi^*\} = [\pi_1(0) \quad \pi_2(0)] \cdot \frac{1}{p+q} \begin{bmatrix} q & p \\ q & p \end{bmatrix}. \quad (4.1.40)$$

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 (4.3).

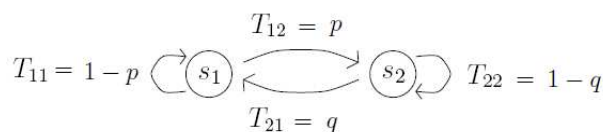


Figura 4.3: Grafo de Transição de uma cadeia de Markov com dois estados (s_1 e s_2).

4.2 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 , com suporte k 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 suporte k de X e, tem uma função densidade de probabilidade, h , chamada de densidade 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 (Cf. [8, 37]):

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):

$$R(x^{(i)}, u^{(i)}) = \frac{p(u^{(i)}) h(x^{(i)})}{p(x^{(i)}) h(u^{(i)})}. \quad (4.2.41)$$

3. Avaliar a função α dada por (4.2.42):

$$\alpha(x^{(i)}, u^{(i)}) = \min [1, R(x^{(i)}, u^{(i)})]. \quad (4.2.42)$$

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:

$$x^{(i+1)} = \begin{cases} u^{(i)}, & \text{se } \rho^{(i)} \leq \alpha \\ x^{(i)}, & \text{se } \rho^{(i)} > \alpha \end{cases} \quad (4.2.43)$$

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, $\alpha(x^{(i)}, u^{(i)})$).

$$T_{x^{(i)}, u^{(i)}} = h(u^{(i)}) \alpha(x^{(i)}, u^{(i)}). \quad (4.2.44)$$

O teorema da convergência da distribuição de probabilidade de uma cadeia de Markov, enunciado na seção anterior, 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 (4.1.19) é satisfeita, ou seja, que:

$$p(x^{(i)}) T_{x^{(i)}, u^{(i)}} = p(u^{(i)}) T_{u^{(i)}, x^{(i)}}. \quad (4.2.45)$$

Com efeito, combinando as equações (4.2.45), (4.2.42) e (4.2.41):

$$\begin{aligned}
 p(x^{(i)})T_{x^{(i)},u^{(i)}} &= p(x^{(i)})h(u^{(i)})\alpha(x^{(i)},u^{(i)}) \\
 &= f(x^{(i)})h(u^{(i)})\min\left[1, R(x^{(i)},u^{(i)})\right] \\
 &= \min\left[f(u^{(i)})h(u^{(i)}), f(x^{(i)})h(x^{(i)})\right] \\
 &= \min\left[f(u^{(i)})h(u^{(i)}), f(u^{(i)})h(u^{(i)})/R(x^{(i)},u^{(i)})\right] \\
 &= f(u^{(i)})h(u^{(i)})\min\left[1, 1/R(x^{(i)},u^{(i)})\right].
 \end{aligned} \tag{4.2.46}$$

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

$$p(x^{(i)})T_{x^{(i)},u^{(i)}} = p(u^{(i)})h(u^{(i)})\alpha(u^{(i)},x^{(i)}) = p(u^{(i)})T_{u^{(i)},x^{(i)}}. \tag{4.2.47}$$

EXEMPLO 4.2.1 Para exemplificar o uso do método de Monte Carlo com Cadeia de Markov, suponha que deseja-se calcular a seguinte integral:

$$I = \int_0^\infty (x - \alpha\beta)^2 p(x) = dx, \tag{4.2.48}$$

onde p é a função densidade de probabilidade Gamma

$$p(x) = \mathbb{1}_{[0,\infty)}(x) \frac{1}{\beta\Gamma(\alpha)} \left(\frac{x}{\alpha}\right)^{\alpha-1} \exp\left(-\frac{x}{\beta}\right), \tag{4.2.49}$$

e Γ a função Gamma.

Essa integral (4.2.48) representa a variância de uma variável aleatória X função densidade de probabilidade Gamma e vale $I = \alpha\beta^2$. Para calculá-la através do MCMC, são geradas amostras de uma variável aleatória X com função densidade de probabilidade Gamma, p e a partir dessas amostras faz-se uma estimativa para a média de X , ou seja, para a integral I . Como densidade de probabilidade proposta, h , utiliza-se a seguinte função:

$$h(u) = \mathbb{1}_{[0,\infty)}(x) 0,5 \exp(-0,5 u). \tag{4.2.50}$$

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 através do método da Transforma Inversa fazendo-se $u^{(i)} = -2 \ln(v^{(i)})$, onde cada $v^{(i)}$ é sorteado da distribuição uniforme no intervalo $[0, 1]$.

Nesse exemplo, a razão de Metropolis vale:

$$R = \frac{p(u^{(i)})h(x^{(i)})}{p(x^{(i)})h(u^{(i)})} = \left(\frac{u^{(i)}}{x^{(i)}}\right)^{\alpha-1} \exp\left(\frac{u^{(i)} - x^{(i)}}{\beta}\right) \exp(-0,5(x^{(i)} - u^{(i)})), \tag{4.2.51}$$

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

$$x^{(i+1)} = \begin{cases} u^{(i)}, & \text{se } \rho^{(i)} \leq \min[1, R] \\ x^{(i)}, & \text{se } \rho^{(i)} > \min[1, R] \end{cases} \quad (4.2.52)$$

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

A partir das m amostras geradas $x^{(1)}, x^{(2)}, \dots, x^{(m)}$, a integral (4.2.48) é estimada pela expressão (4.2.53) e o desvio padrão por (4.2.54).

$$\hat{I} = \frac{1}{m} \sum_{i=1}^m (x^{(i)} - \alpha\beta)^2 = \frac{1}{m} \sum_{i=1}^m (x^{(i)} - 2)^2 \quad (4.2.53)$$

$$\hat{\sigma}_I = \frac{1}{m} \sqrt{\sum_{i=1}^m \left((x^{(i)} - \alpha\beta)^2 - \hat{I} \right)^2} \quad (4.2.54)$$

Com o objetivo de avaliar a qualidade dos resultados fornecidos pelo MCMC, a integral (4.2.48) também foi estimada a partir de amostras geradas com a função `random('gam', α, β)` do MATLAB.

A tabela (4.1) mostra os resultados obtidos para diferentes números de amostras, m .

m	MCMC $\hat{I} \pm \hat{\sigma}_I$	Função <code>random('gam', α, β)</code> $\hat{I} \pm \hat{\sigma}_I$
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 4.1: Estimativas da média e variância da variável aleatória X com função densidade de probabilidade Gamma ($\alpha = 2$ e $\beta = 1$)

Tanto o método de MCMC, quanto a função do MATLAB

`random('gam', α, β)`

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 (4.4) mostra semelhança de formato entre o histograma normalizado das amostras obtidas por MCMC (10^7 amostras) e a função densidade de probabilidade Gamma (4.2.49).

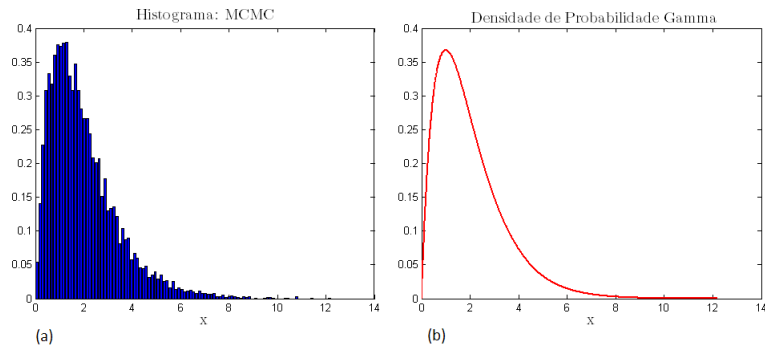


Figura 4.4: (a) Histograma normalizado das amostras obtidas por MCMC (10^4 amostras). (b) Função densidade de probabilidade Gamma com média $\mu = \alpha\beta$ e variância $\sigma^2 = \alpha\beta^2$.

4.2.1 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 geradas m é suficientemente grande, esse período de aquecimento exerce pouca influência sobre a distribuição das amostras obtidas.

4.2.2 Densidade 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 densidade é escolhida arbitrariamente, é costume utilizar-se uma densidade 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 terceiro capítulo.

Em princípio, a definição de h é totalmente livre, porém é usual escolher-se a densidade h com uma forma semelhante a densidade da qual deseja-se gerar amostras, p (Cf. [50]). Essa escolha faz com que a razão de Metrópolis, R , assumna na maioria das iterações do algoritmo um valor próximo a 1, e consequentemente

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') (Cf. [50]). 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 densidade proposta, h , na taxa de mistura da cadeia.

EXEMPLO 4.2.2 *Suponha que se deseja gerar amostras através do MCMC de uma variável aleatória X com a seguinte densidade de probabilidade Gamma p :*

$$p(x) = \mathbb{1}_{[0,+\infty)}(x) \frac{1}{\mu} \left(\frac{1}{\delta^2}\right)^{\frac{1}{\delta^2}} \frac{1}{\Gamma(1/\delta^2)} \left(\frac{x}{\mu}\right)^{\frac{1}{\delta^2}-1} \exp\left(-\frac{x}{\delta^2 \mu}\right), \quad (4.2.55)$$

com $\mu = 3.0$, $\delta = 0.25$ e $\sigma = 0.75$. Supõe-se que o valor inicial da cadeia é $x^{(1)} = 3$.

Como densidade proposta, duas funções, h_1 e h_2 , de suporte positivo $[0, +\infty)$ são analisadas:

$$h_1(u) = \mathbb{1}_{[0,+\infty)}(u) \lambda e^{-\lambda u}, \quad (4.2.56)$$

$$h_2(u) = \mathbb{1}_{[0,+\infty)}(u) \frac{\alpha}{\beta} \left(\frac{u}{\beta}\right)^{\alpha-1} e^{-(u/\beta)^\alpha}, \quad (4.2.57)$$

onde $\lambda = 0.08$, $\alpha = 3.0$ e $\beta = 3.5$. Observando a figura (4.5), verifica-se que primeira função, h_1 , possui uma forma bem diferente da densidade p , enquanto a segunda densidade, h_2 , possui um formato semelhante.

Uma rotina implementada em MATLAB testou o algoritmo de Metrópolis-Hastings para as duas densidades propostas. Na figura (4.2.2) é possível comparar as primeiras 500 amostras para as duas funções h . No caso da densidade h_1 , a cadeia pode ser considerada mal misturada ('poorly mixing') (há uma grande repetição sequencial de elementos), enquanto no caso da densidade h_2 (com forma mais semelhante a função p), é considerada bem misturada ('well mixing').

Construindo o gráfico de frequências das amostras geradas com h_1 e h_2 , verifica-se que com $m = 10^3$ amostras de X , nenhuma das distribuições propostas gera um

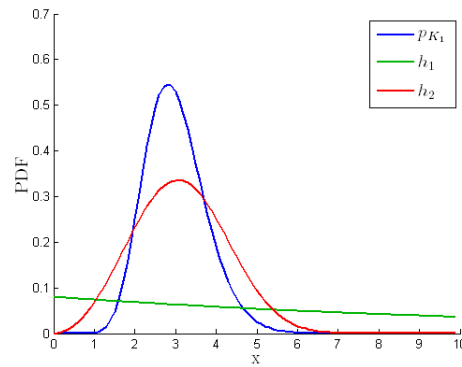


Figura 4.5: Função densidade de probabilidade p (com $\mu = 3.0$, $\delta = 0.25$ e $\sigma = 0.75$) e as duas distribuições propostas: h_1 (com $\lambda = 0.08$) e h_2 (com $\alpha = 3.0$ e $\beta = 3.5$).

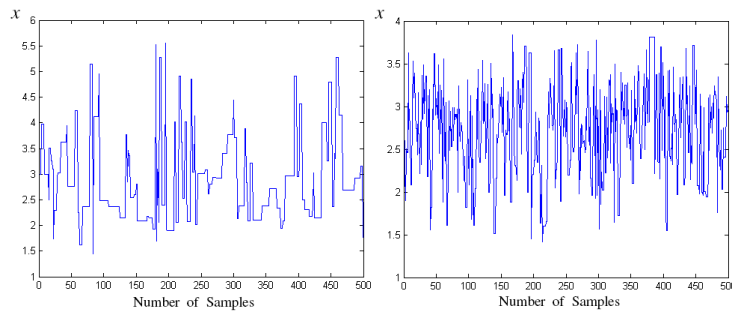


Figura 4.6: Primeiras 500 amostras da sequência $x^{(i)}$ geradas através de (a) h_1 com $\lambda = 0.08$ e (b) h_2 com $\alpha = 3.0$ e $\beta = 3.5$.

gráfico de frequência semelhante a densidade de probabilidade p , como é mostrado na figura (4.7).

Quando o número de amostras cresce para 8×10^3 , (4.8), o gráfico de frequência das amostras geradas com h_1 continua bem diferente da densidade de probabilidade p , enquanto o gráfico de frequência das amostras geradas com h_2 se aproxima mais de p .

Aumentando ainda mais o número de amostras, $m = 10^6$, ambos os gráficos de frequência são boas representações da função p .

Com este exemplo, verifica-se que a escolha de uma densidade proposta h que melhor aproxime a forma de p , reduz o número de amostras necessárias para o

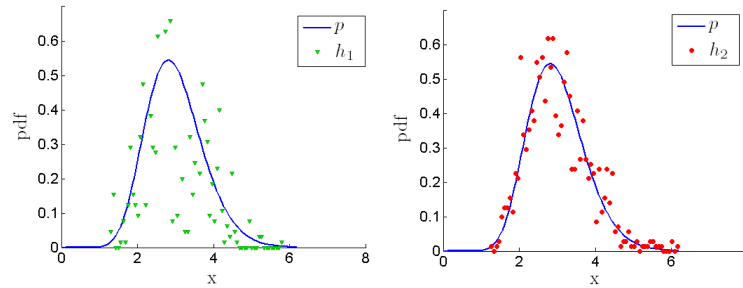


Figura 4.7: Gráfico de frequência construído com 10^3 amostras de X geradas através de h_1 e h_2 .

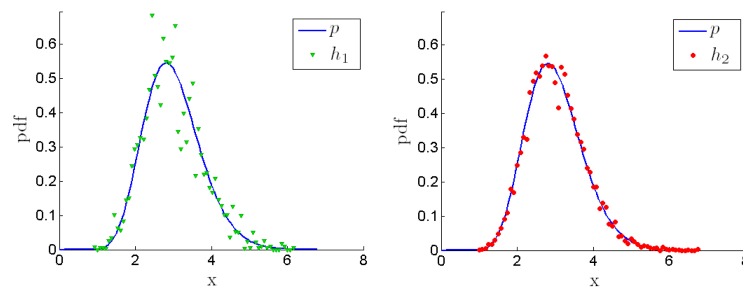


Figura 4.8: Gráfico de frequência construído com 8×10^3 amostras de X geradas através de h_1 e h_2 .

MCMC apresentar um bom desempenho. No entanto, é importante notar que se um número suficiente de amostras é gerado, essencialmente qualquer densidade 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 (4.2) 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.

EXEMPLO 4.2.3 Suponha que se deseja gerar amostras através do MCMC de uma variável aleatória X com a seguinte densidade de probabilidade p :

$$p(x) = \mathbb{1}_{[0, \infty)}(x) \cdot c \left(\frac{x}{a} \right)^{a-1} \exp\left(-\frac{x}{b}\right), \quad (4.2.58)$$

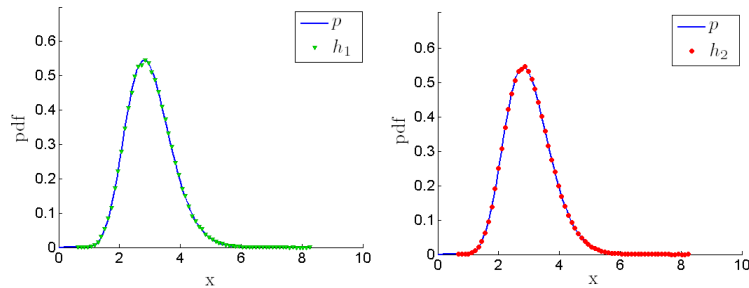


Figura 4.9: Gráfico de frequência construído com 10^6 amostras de X geradas através de h_1 e h_2 .

Tabela 4.2: Estimativas de média e variância de X calculadas a partir das amostras geradas com MCMC e com o comando `gamrnd` do MATLAB.

m	h_1 $\mu \pm \sigma$	h_2 $\mu \pm \sigma$	MATLAB $\mu \pm \sigma$
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

onde $a = 16.0$, $b = 0.18$ e c é uma constante de normalização com valor desconhecido.

Como distribuições propostas para o MCMC, uma função h , de suporte positivo $(0, +\infty)$, foi analisada:

$$h(u) = \mathbb{1}_{[0, \infty)}(x) \lambda e^{-\lambda u}, \quad (4.2.59)$$

com $\lambda = 1.0$. Supõe-se que o valor inicial da cadeia é $x^{(1)} = 3$. O histograma normalizado construído com 10^4 amostras de X obtidas através de h é mostrado na figura (4.10).

4.3 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 aleatórios.

Um vetor aleatório real de dimensão n foi definido no primeiro capítulo como uma função real que associa a cada ponto amostra w , pertencente ao espaço de amostras Ω , um ponto do espaço n -dimensional, $\{X(w)\} \in \mathbb{R}^n$. Assim, cada

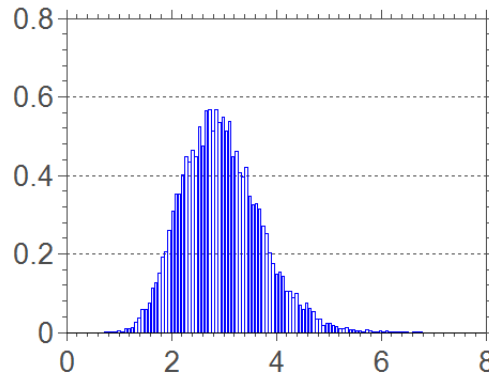


Figura 4.10: Histograma normalizado construído com 10^4 amostras de X .

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

$$\{x\}^{(i)} = \begin{bmatrix} x_1^{(i)} \\ \vdots \\ x_n^{(i)} \end{bmatrix}. \quad (4.3.60)$$

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 (4.3.61). A escolha da componente a ser gerada pode ser de forma sequencial ou aleatória.

$$\{u\}^{(i+1)} = [u_1^{(i)} \quad u_2^{(i)} \quad \dots \quad u_j^{(i+1)} \quad \dots \quad u_n^{(i)}] \quad (4.3.61)$$

EXEMPLO 4.3.1 *Suponha que deseja-se gerar amostras do vetor aleatório $\{X\} = (X_1 \ X_2)^T \in \mathbb{R}^2$ com a seguinte função densidade de probabilidade conjunta Gaussiana mostrada na figura (4.11) com média $\{\mu_X\} = \begin{bmatrix} 3 \\ 3 \end{bmatrix}$ e matriz de covariância*

$[C] = \begin{pmatrix} 1.0 & 0.1 \\ 0.1 & 1.0 \end{pmatrix}$. A função densidade de probabilidade conjunta do vetor é:

$$p(\{x\}) = \frac{1}{\sqrt{(2\pi)^2 \det[C]}} \times \exp \left\{ -\frac{1}{2} \left\langle [C]^{-1} (\{x\} - \{\mu_X\}); (\{x\} - \{\mu_X\}) \right\rangle_{\mathbb{R}^2} \right\}. \quad (4.3.62)$$

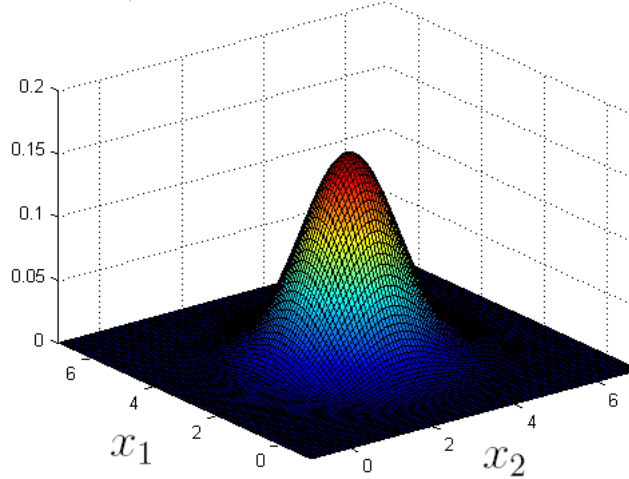


Figura 4.11: Gráfico da função densidade de probabilidade conjunta gaussiana $p_{\{X\}}$ na região $[0, 7] \times [0, 7]$ para $\{\mu_X\} = \begin{bmatrix} 3 \\ 3 \end{bmatrix}$ e $[C] = \begin{bmatrix} 1.0 & 0.1 \\ 0.1 & 1.0 \end{bmatrix}$.

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

$$h([u_1 \ u_2]) = h_{X_1}(u_1) \times h_{X_2}(u_2). \quad (4.3.63)$$

Cada uma das funções h_{X_1} e h_{X_2} está definida no domínio $(-\infty, \infty)$. Elas são dadas por:

$$h_{X_1}(u_1) = \mathcal{N}(0, 1) = \mathbf{1}_{(-\infty, \infty)}(x) \frac{e^{-u_1^2/2}}{\sqrt{2\pi}}, \quad (4.3.64)$$

$$h_{X_2}(u_2) = \mathbf{1}_{(-\infty, \infty)}(x) \frac{0.1}{2} e^{-0.1|u_2|}. \quad (4.3.65)$$

Para gerar amostras $u_2^{(i)}$ segundo h_{X_2} , um número $\beta^{(i)}$ é sorteado com distribuição de probabilidade uniforme unitária no intervalo $[0, 1]$ e faz-se:

$$u_2^{(i)} = \begin{cases} \ln(2\beta^{(i)})/0.1, & \text{se } \beta^{(i)} < 0.5 \\ 0, & \text{se } \beta^{(i)} = 0.5 \\ \ln(2 - 2\beta^{(i)})/0.1, & \text{se } \beta^{(i)} > 0.5 \end{cases} \quad (4.3.66)$$

O algoritmo de Metropolis-Hastings multidimensional foi implementado em MATLAB. As figuras (4.12), (4.13) e (4.14) mostram os histogramas normalizados construídos a partir de diferentes números de amostras. Verifica-se que 10^4 amostras são insuficientes para construir um histograma normalizado similar à densidade p . No entanto, quando este número cresce para 10^7 , o histograma normalizado torna-se representativo. Como no exemplo (4.2.2), 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 (4.3). Nas simulações, considerou-se $\{\mu_X\} = \begin{bmatrix} 3 \\ 3 \end{bmatrix}$ e matriz de covariância $[C] = \begin{bmatrix} 1.0 & 0.1 \\ 0.1 & 1.0 \end{bmatrix}$.

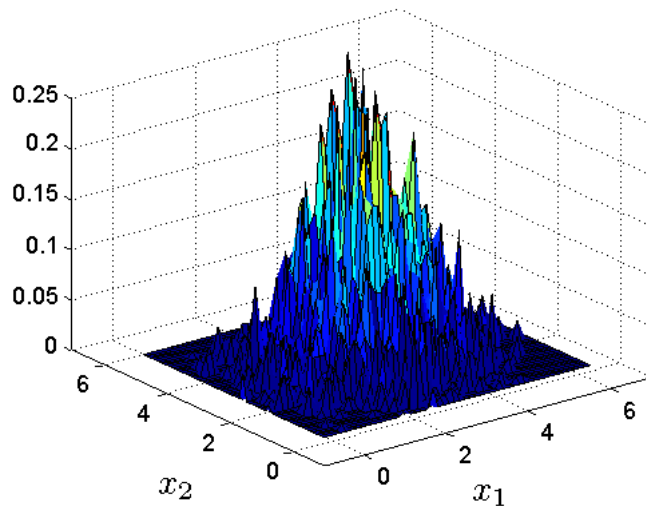


Figura 4.12: Histograma normalizado construído com 10^4 amostras de um vetor aleatório Gaussiano.

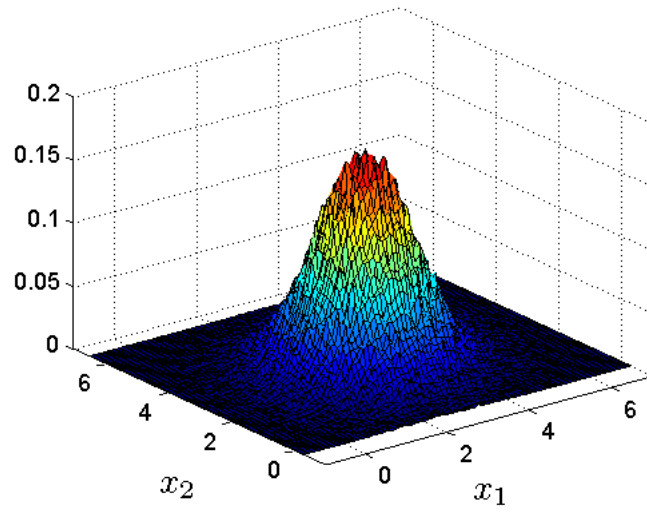


Figura 4.13: Histograma normalizado construído com 10^6 amostras de um vetor aleatório Gaussiano.

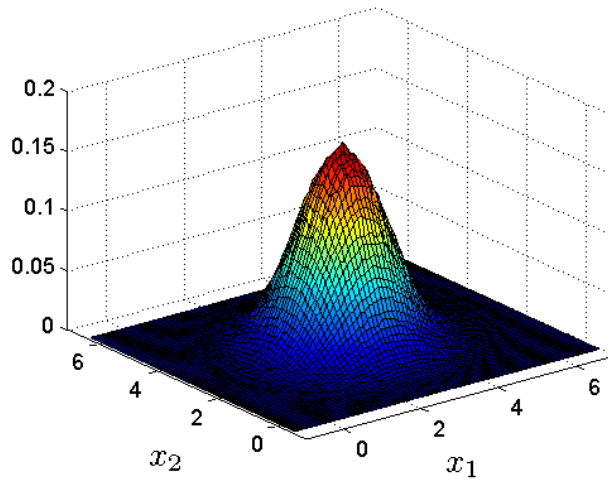


Figura 4.14: Histograma normalizado construído com 10^7 amostras de um vetor aleatório Gaussiano.

Tabela 4.3: Estimativas de $\{\mu_X\}$ e $[C]$.

N	MCMC		MATLAB	
	$\{\mu_X\}$	$[C]$	$\{\mu_X\}$	$[C]$
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	$\begin{pmatrix} 2.9985 \\ 3.0030 \end{pmatrix}$	$\begin{pmatrix} 0.9977 & 0.1019 \\ 0.1019 & 1.0027 \end{pmatrix}$	$\begin{pmatrix} 3.0004 \\ 2.9995 \end{pmatrix}$	$\begin{pmatrix} 1.0002 & 0.1003 \\ 0.1003 & 1.0015 \end{pmatrix}$

4.4 Exercícios

EXERCÍCIO 4.4.1 *Implemente uma rotina para gerar amostras de uma variável aleatória X com densidade de probabilidade Gamma:*

$$p(x) = \mathbf{1}_{[0, \infty)}(x) \frac{1}{\beta \Gamma(\alpha)} \left(\frac{x}{\alpha}\right)^{\alpha-1} \exp\left(-\frac{x}{\beta}\right), \quad (4.4.67)$$

com $\alpha = 0.5$ e $\beta = 0.8$ através de um algoritmo baseado em MCMC. Como densidade de probabilidade proposta, h , utilize a seguinte função:

$$h(u) = \mathbf{1}_{[0, \infty)}(x) 0,5 \exp(-0,5 u). \quad (4.4.68)$$

Gere 10^2 amostras de X e posteriormente aumente esse número para 10^3 e 10^4 . Construa o histograma de cada uma das seqüências de amostras.

EXERCÍCIO 4.4.2 *Implemente uma rotina para gerar amostras de um vetor aleatório $\{X\} = (X_1 \ X_2)^T \in \mathbb{R}^2$ com função densidade de probabilidade conjunta Gaussiana com média $\{\mu_X\} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ e matriz de covariância $[C] = \begin{pmatrix} 1.0 & 0.4 \\ 0.4 & 1.0 \end{pmatrix}$.*

Como distribuições de probabilidade propostas, duas distribuições (uma para componente de $\{X\}$) devem ser utilizadas. Cada uma delas está definida no domínio $(-\infty, \infty)$ e são elas:

$$h_{X_1}(u_1) = \mathcal{N}(0, 1) = \mathbf{1}_{(-\infty, \infty)}(x) \frac{e^{-u_1^2/2}}{\sqrt{2\pi}}, \quad (4.4.69)$$

$$h_{X_2}(u_2) = \mathbf{1}_{(-\infty, \infty)}(x) \frac{0.1}{2} e^{-0.1|u_2|}. \quad (4.4.70)$$

Gere 10^6 amostras de $\{X\}$ e posteriormente aumente esse número para 10^7 e 10^8 . Estime o vetor de média $\{\hat{\mu}_X\}$ e a matriz de covariância $[\hat{C}]$ através das amostras geradas.

Capítulo 5

Expansão de Karhunen-Loève

5.1 Expansão de Karhunen-Loève

DEFINIÇÃO 5.1.1 (EXPANSÃO DE KARHUNEN-LOÈVE)

Seja $\mu_{\mathcal{X}}$ a média em função do tempo de um processo estocástico \mathcal{X} de parâmetro $t \in T$ e, seja $C(t_1, t_2)$ a sua função de covariância. A expansão de Karhunen-Loève de \mathcal{X} é expressa por:

$$\mathcal{X}(t, w) = \mu_{\mathcal{X}}(t) + \sum_{k=1}^{\infty} \sqrt{\lambda_k} \psi_k(t) X_k(w), \quad (5.1.1)$$

onde cada ψ_k é uma autofunção e cada λ_k o correspondente autovalor do seguinte problema:

$$\int_T C(t_1, t_2) \psi_k(t_2) dt_2 = \lambda_k \psi_k(t_1), \quad t \in T, \quad (5.1.2)$$

e onde $\{X_k\}_{k \in \mathbb{N}_{\geq 1}}$ é uma sequência de variáveis aleatórias duas a duas decorrelacionadas, tal que:

$$E[X_k] = 0 \quad e \quad E[X_k X_m] = \delta_{km}, \quad \forall k, m \in \mathbb{N}, \quad (5.1.3)$$

e é definida por:

$$X_k(w) = \frac{1}{\lambda_k} \int_T [\mathcal{X}(t, w) - \mu_{\mathcal{X}}(t)] \psi_k(t) dt, \quad \forall k \in \mathbb{N}. \quad (5.1.4)$$

Dessa forma, a expansão de Karhunen-Loève representa uma maneira muito especial de decompor um processo estocástico em uma soma infinita de termos que dependem do tempo (as autofunções ψ_k) e de uma sequência de variáveis aleatórias, X_k .

Na prática, a expansão de KL é aproximada por uma soma finita com d termos:

$$\mathcal{X}(t, w) \approx \mu_{\mathcal{X}}(t) + \sum_{k=1}^d \sqrt{\lambda_k} \psi_k(t) X_k(w). \quad (5.1.5)$$

Surtem então duas questões a serem analisadas:

1. se uma boa aproximação para o processo estocástico \mathcal{X} pode ser obtida, fixado um grau de precisão, truncando-se a série em d termos e;
2. quantos termos são necessários para obter-se essa boa aproximação.

A resposta para essas questões está relacionada com uma importante característica da expansão de Karhunen-Loève: o decaimento do valor dos autovalores, λ_k , em função do crescimento do índice k .

A ordenação de forma decrescente da sequência de autovalores λ_k obtidas por (5.1.2) permite que se avalie a importância desses autovalores na expansão de KL. Quanto menor o autovalor, menor é a sua influência na série de KL.

Assim, para determinar o número de termos utilizados na expansão de KL truncada, é usual analisar o decaimento dos λ_k e considerar os d primeiros autovalores, fazendo com que a influência dos termos desprezados seja a menor possível.

É importante observar que esse número de termos d a serem considerados vai depender do processo estocástico que está sendo expandido e de sua função de covariância (Cf. [42, 3, 40, 49]).

Os próximos exemplos ilustram essa característica do decaimento dos autovalores para processos com diferentes funções de covariância.

EXEMPLO 5.1.2 *Considere um processo estocástico, \mathcal{X} , com parâmetro t definido para $t \in T = [-b, b]$ e com uma função de covariância C exponencial:*

$$C(t_1, t_2) = \exp\left(\frac{-|t_1 - t_2|}{a}\right), \quad a > 0. \quad (5.1.6)$$

Aplicando-se a expansão de Karhunen-Loève a esse processo, as autofunções ψ_k e seus correspondentes autovalores λ_k são calculados através da seguinte integral (5.1.2):

$$\int_T \exp(-c|t_1 - t_2|) \psi_k(t_2) dt_2 = \lambda_k \psi_k(t_1), \quad (5.1.7)$$

onde $c = 1/a$. A expressão (5.1.7) pode ser reescrita como:

$$\int_{-b}^{t_1} \exp(-c(t_1 - t_2))\psi_k(t_2) dt_2 + \int_{t_1}^b \exp(-c(t_2 - t_1))\psi_k(t_2) dt_2 = \lambda_k \psi_k(t_1). \quad (5.1.8)$$

Diferenciando (5.1.8) duas vezes em relação a t_1 e, fazendo algumas manipulações matemáticas, obtém-se uma equação diferencial em termos de ψ_k . Resolvendo essa equação (Cf. [52]), chega-se a seguinte expressão para os autovalores λ_k :

$$\lambda_k = \begin{cases} \frac{2a}{1 + a^2\omega_k^2}, & \text{para } k = 2, 4, 6, \dots \\ \frac{2a}{1 + a^2v_k^2}, & \text{para } k = 1, 3, 5, \dots \end{cases} \quad (5.1.9)$$

e para as correspondentes autofunções ψ_k :

$$\psi_k(t) = \begin{cases} \sin(\omega_k t) / \sqrt{b - \frac{\sin(2\omega_k b)}{2\omega_k}}, & \text{para } k = 2, 4, 6, \dots \\ \sin(v_k t) / \sqrt{b + \frac{\sin(2v_k b)}{2v_k}}, & \text{para } k = 1, 3, 5, \dots \end{cases} \quad (5.1.10)$$

onde ω_k e v_k são soluções da equação:

$$\begin{cases} a\omega + \tan(\omega b) = 0, & \text{para } k = 2, 4, 6, \dots \\ 1 - av \tan(vb) = 0, & \text{para } k = 1, 3, 5, \dots \end{cases} \quad (5.1.11)$$

Na figura (5.1) é apresentado o gráfico dos primeiros 20 autovalores λ_k calculados para a função de covariância exponencial (5.1.6) com diferentes valores de a .

EXEMPLO 5.1.3 Considere um processo estocástico \mathcal{X} , com parâmetro t definido para $t \in T = [-b, b]$ e com uma função de covariância C :

$$C(t_1, t_2) = \delta(t_1 - t_2), \quad \forall t_1, t_2 \in T. \quad (5.1.12)$$

Aplicando a expansão de KL a \mathcal{X} , verifica-se por (5.1.2) que as autofunções ψ_k podem ser quaisquer funções ortogonais e, os autovalores λ_k são todos constantes iguais a 1, isto é, $\lambda_k = 1, \forall k$. Nesse caso, quando $C(t_1, t_2) = \delta(t_1 - t_2)$ não há decaimento dos autovalores.

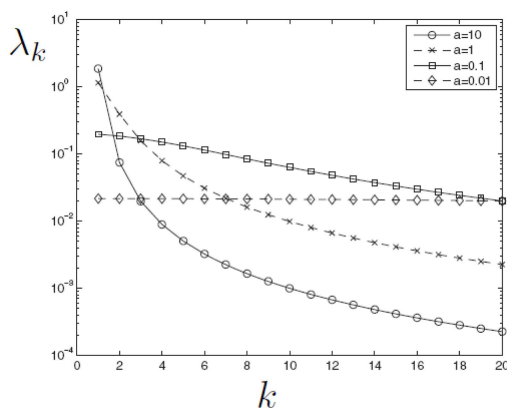


Figura 5.1: Primeiros 20 autovalores λ_k calculados para a função de covariância com diferentes valores de a .

5.1.1 Expansão de KL aplicada a processos estocásticos gaussianos

Os processos estocásticos gaussianos quando submetidos à expansão de KL apresentam um comportamento especial. Eles são representados pela soma (5.1.1) onde todos os X_k são variáveis aleatórias gaussianas independentes.

Isso se deve ao fato de que, quando trabalha-se com variáveis aleatórias gaussianas, decorrelação é equivalente a independência e, além disso, combinações lineares de variáveis gaussianas resultam em variáveis aleatórias gaussianas.

Dessa forma, a expansão truncada de KL permite parametrizar um processo estocástico gaussianos em um número finito de variáveis aleatórias gaussianas.

5.1.2 Exemplos numéricos da expansão de KL

Na seção anterior, mostrou-se como obter expressões analíticas da expansão de Karhunen-Loève para processos estocásticos a partir da expressão analítica da função de covariância. Nesta seção é mostrado como a expansão de KL pode ser estimada numericamente a partir das realizações do processo estocástico.

Antes de explicar como isso é feito, é necessário definir o que é considerado uma realização de um processo estocástico em uma simulação numérica.

5.1.3 O processo de discretização

O computador é um dispositivo capaz de fazer contas em um domínio discreto. Por isso, toda vez que um algoritmo é implementado em alguma linguagem com-

putacional, é escrito para fazer contas discretas. Quando se deseja trabalhar em um domínio contínuo, é necessário fazer uma discretização. No francês há uma palavra que define bem essa transformação, é a chamada "numérisation".

Por isso, quando se trabalha com realizações numéricas de um processo estocástico \mathcal{X} , primeiramente faz-se uma discretização do parâmetro t (geralmente associado ao tempo). Considera-se, por exemplo, que ele assumirá n valores $t_1, t_2, \dots, t_j, \dots, t_n, \forall t_j \in \mathbf{T}$ e $\forall j \in \mathbb{N}$.

Fixados esses n instantes, pode-se então associar a cada instante t_j uma variável aleatória $\mathcal{X}(t_j, \mathbf{w})$. No total, ficam definidas n variáveis:

$$\mathcal{X}(t_1, \mathbf{w}), \mathcal{X}(t_2, \mathbf{w}), \dots, \mathcal{X}(t_j, \mathbf{w}), \dots, \mathcal{X}(t_n, \mathbf{w}).$$

Cada realização de \mathcal{X} é representada por realizações dessas n variáveis aleatórias, ou seja: $\mathcal{X}(t_1, \mathbf{w}^{(i)}), \dots, \mathcal{X}(t_j, \mathbf{w}^{(i)}), \dots, \mathcal{X}(t_n, \mathbf{w}^{(i)})$.

É usual e prático representar cada realização através de um vetor com n componentes:

$$\{\mathcal{X}(t, \mathbf{w})\}^{(i)} = \begin{bmatrix} \mathcal{X}(t_1, \mathbf{w}^{(i)}) \\ \vdots \\ \mathcal{X}(t_j, \mathbf{w}^{(i)}) \\ \vdots \\ \mathcal{X}(t_n, \mathbf{w}^{(i)}) \end{bmatrix}. \quad (5.1.13)$$

Um processo estocástico é dito estar especificado até ordem n quando a função densidade de probabilidade conjunta $p_{\mathcal{X}(t_1, \mathbf{w})\mathcal{X}(t_2, \mathbf{w})\dots\mathcal{X}(t_n, \mathbf{w})}$ é conhecida para qualquer conjunto de valores $(t_1, t_2, \dots, t_i, \dots, t_n)$, tais que $t_i \in \mathbf{T}, \forall i$.

Assim, dado um processo especificado até uma ordem n , é possível gerar realizações numéricas desse processo. Ou seja, é possível obter os vetores:

$$\{\mathcal{X}(t, \mathbf{w})\}^{(1)}, \{\mathcal{X}(t, \mathbf{w})\}^{(2)}, \dots.$$

5.1.4 Estimativas dos momentos de um processo estocástico

A partir de um número grande, m , de realizações numéricas de um processo estocástico \mathcal{X} , estimativas para a média, variância e correlação desse processo podem ser calculadas (Cf. [51]).

Suponha que cada realização do processo, $\{\mathcal{X}(t, \mathbf{w})\}^{(i)}$, é representada por um vetor $(n \times 1)$ e que nessa realização, o parâmetro t assume os valores

$$t_1, t_2, \dots, t_j, \dots, t_n \in \mathbf{T}.$$

Dessa forma, a média, $\mu_{\mathcal{X}}(t)$, pode ser estimada através de um vetor com n componentes. Cada componente j representa a estimativa da média da variável aleatória $\mathcal{X}(t_j, \mathbf{w})$, sendo calculada por:

$$\hat{\mu}_{\mathcal{X}}(t_j) = \frac{1}{m} \sum_{i=1}^m \mathcal{X}(t_j, \mathbf{w}^{(i)}), \quad j = 1, \dots, n. \quad (5.1.14)$$

Assim, o vetor $\{\hat{\mu}_{\mathcal{X}}(t)\}$ que estima a média do processo é:

$$\{\hat{\mu}_{\mathcal{X}}(t)\} = \begin{bmatrix} \hat{\mu}_{\mathcal{X}}(t_1) \\ \vdots \\ \hat{\mu}_{\mathcal{X}}(t_j) \\ \vdots \\ \hat{\mu}_{\mathcal{X}}(t_n) \end{bmatrix}. \quad (5.1.15)$$

Tendo calculado $\{\hat{\mu}_{\mathcal{X}}(t)\}$, um procedimento semelhante pode ser feito para estimar a variância do processo, $\sigma_{\mathcal{X}}^2$ (1.3.73). Em cada instante t_j :

$$\hat{\sigma}_{\mathcal{X}}^2(t_j) = \frac{1}{m} \sum_{i=1}^m (\hat{\mu}_{\mathcal{X}}(t_j) - \mathcal{X}(t_j, \mathbf{w}^{(i)}))^2, \quad j = 1, \dots, n. \quad (5.1.16)$$

Assim, define-se o vetor:

$$\{\hat{\sigma}_{\mathcal{X}}^2(t)\} = \begin{bmatrix} \hat{\sigma}_{\mathcal{X}}^2(t_1) \\ \vdots \\ \hat{\sigma}_{\mathcal{X}}^2(t_j) \\ \vdots \\ \hat{\sigma}_{\mathcal{X}}^2(t_n) \end{bmatrix}. \quad (5.1.17)$$

Para estimar a função de covariância em dois instantes quaisquer $t_j, t_k \in \mathbb{T}$, faz-se:

$$\hat{C}(t_j, t_k) = \frac{1}{m} \sum_{i=1}^m [\mathcal{X}(t_j, \mathbf{w}^{(i)}) - \hat{\mu}_{\mathcal{X}}(t_j)] \cdot [\mathcal{X}(t_k, \mathbf{w}^{(i)}) - \hat{\mu}_{\mathcal{X}}(t_k)]. \quad (5.1.18)$$

Dessa forma, a função de covariância pode ser representada através de uma matriz $[\hat{C}]$ ($n \times n$):

$$[\hat{C}] = \begin{bmatrix} \hat{C}(t_1, t_1) & \cdots & \hat{C}(t_1, t_n) \\ \vdots & \ddots & \vdots \\ \hat{C}(t_n, t_1) & \cdots & \hat{C}(t_n, t_n) \end{bmatrix}. \quad (5.1.19)$$

Representando as m realizações de um processo estocástico \mathcal{X} em uma matriz $[\mathcal{X}](n \times m)$, de forma que cada coluna de $[\mathcal{X}]$ seja uma realização do processo:

$$[\mathcal{X}] = \begin{bmatrix} \left. \begin{array}{c} | \\ \{\mathcal{X}(t, w)\}^{(1)} \\ | \end{array} \right\} & \left. \begin{array}{c} | \\ \{\mathcal{X}(t, w)\}^{(2)} \\ | \end{array} \right\} & \cdots & \left. \begin{array}{c} | \\ \{\mathcal{X}(t, w)\}^{(m)} \\ | \end{array} \right\} \end{bmatrix}, \quad (5.1.20)$$

ou seja:

$$[\mathcal{X}] = \begin{bmatrix} \mathcal{X}(t_1, w^{(1)}) & \mathcal{X}(t_1, w^{(2)}) & \cdots & \mathcal{X}(t_1, w^{(m)}) \\ \vdots & \vdots & \vdots & \vdots \\ \mathcal{X}(t_n, w^{(1)}) & \mathcal{X}(t_n, w^{(2)}) & \cdots & \mathcal{X}(t_n, w^{(m)}) \end{bmatrix}. \quad (5.1.21)$$

e criando uma matriz $[\mathcal{X}_0](n \times m)$ em que cada coluna k seja igual a: $\{\mathcal{X}(t, w)\}^{(k)} - \{\hat{\mu}_{\mathcal{X}}(t)\}$,

$$[\mathcal{X}_0] = \begin{bmatrix} \{\mathcal{X}(t, w)\}^{(1)} - \{\hat{\mu}_{\mathcal{X}}(t)\} & \cdots & \{\mathcal{X}(t, w)\}^{(m)} - \{\hat{\mu}_{\mathcal{X}}(t)\} \end{bmatrix}, \quad (5.1.22)$$

a matriz de covariância $[\hat{C}]$ pode ser escrita como:

$$[\hat{C}] = \frac{1}{M} [\mathcal{X}_0] [\mathcal{X}_0]^T, \quad (5.1.23)$$

onde $(\cdot)^T$ indica a matriz transposta.

Assim, a partir de realizações numéricas de um processo estocástico, \mathcal{X} , é possível obter a matriz $[\hat{C}]$ que estima a sua função de covariância C .

5.1.5 Estimativas dos autovalores e autofunções da expansão de KL

Os termos da expansão de Karhunen-Loève em relação ao tempo são calculados através de um problema de autovalor envolvendo a função de covariância (5.1.2).

Quando trabalha-se em um domínio discreto e, calcula-se uma estimativa dessa função de covariância através da matriz $[\hat{C}]$, um problema de autovalor equivalente a (5.1.2) pode ser definido por:

$$[\hat{C}] \left\{ \hat{\psi} \right\} \Delta t = \hat{\lambda} \left\{ \hat{\psi} \right\}, \quad (5.1.24)$$

onde Δt representa o incremento de tempo utilizado para fazer-se a discretização do processo estocástico.

Resolver (5.1.24), significa obter n autovalores $\hat{\lambda}_k$ e seus n correspondentes autovetores $\left\{ \hat{\psi} \right\}_k$.

Esses $\hat{\lambda}_k$ calculados são considerados estimativas para os autovalores de (5.1.2) e, os $\{\hat{\psi}\}_k$ são considerados estimativas para as autofunções ψ_k .

EXEMPLO 5.1.4 *Considere o processo estocástico:*

$$\mathcal{X}(t, w) = A_1 t + A_2, \quad (5.1.25)$$

com parâmetro $t \in T$ e com A_1 e A_2 sendo variáveis aleatórias. Essas variáveis formam um vetor aleatório $\{A\} = (A_1 \ A_2)^T$ com vetor média e matriz de covariância dados por:

$$\{\mu_A\} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \quad [C_A] = \begin{bmatrix} 1 & 1/2 \\ 1/2 & 1 \end{bmatrix}. \quad (5.1.26)$$

Dado que A_1 e A_2 são variáveis aleatórias gaussianas, \mathcal{X} é uma combinação linear de gaussianas, ou seja, é um processo estocástico gaussiano.

Assim, para qualquer valor de $t_j \in T$, a função densidade de probabilidade correspondente $p_{\mathcal{X}(t_j, w)}$ é gaussiana.

Esse processo estocástico pode ser representado por:

$$\mathcal{X}(t, w) = [T] \{A\} = \begin{bmatrix} t & 1 \end{bmatrix} \begin{bmatrix} A_1 \\ A_2 \end{bmatrix}, \quad (5.1.27)$$

e terá média e variância:

$$\mu_{\mathcal{X}}(t) = [T] \{\mu_A\} = t + 2 \quad (5.1.28)$$

$$\sigma_{\mathcal{X}}^2(t) = [T][C_A][T]^T = t^2 + t + 1. \quad (5.1.29)$$

Dessa forma, a função densidade de probabilidade, $p_{\mathcal{X}(t_j, w)}$, da variável aleatória correspondente a $\mathcal{X}(t_j, w)$ é conhecida para qualquer valor de $t_j \in T$:

$$p_{\mathcal{X}(t_j, w)}(x) = \frac{1}{\sqrt{2\pi} \sqrt{t_j^2 + t_j + 1}} e^{-\frac{(x - (t_j + 2))^2}{2(t_j^2 + t_j + 1)}}, \quad (5.1.30)$$

e conseqüentemente, o processo é dito estar especificado até primeira ordem.

A partir da densidade de probabilidade (5.1.30), 10^4 realizações foram geradas para esse processo (no intervalo de tempo $[0, 10]$ e com um $\Delta t = 0.1$). Estimou-se a matriz de covariância $[\hat{C}]$ por (5.1.23) e, posteriormente resolveu-se o problema de autovalor (5.1.24).

Os autovalores estimados são mostrados na figura (5.2).

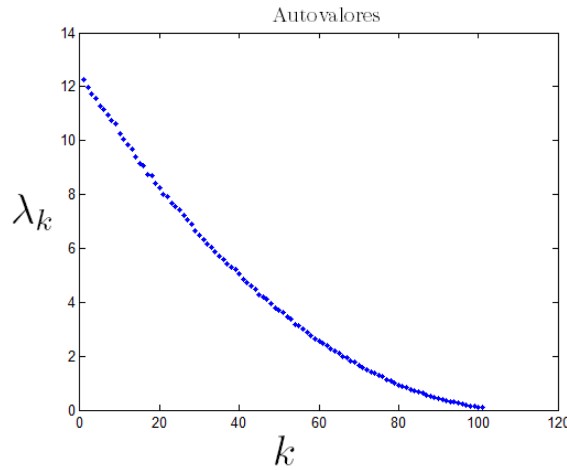


Figura 5.2: Autovalores, λ_k , estimados a partir da matriz de covariância $[\hat{C}]$ do processo (5.1.25).

EXEMPLO 5.1.5 Considere o seguinte processo estocástico:

$$\mathcal{X}(t, w) = A_1 \cos(\omega t) + A_2 \sin(\omega t), \quad (5.1.31)$$

com parâmetro $t \in T$ com A_1 e A_2 sendo variáveis aleatórias gaussianas. O vetor aleatório $\{A\} = (A_1 \ A_2)^T$ possui vetor média e matriz de covariância dados por (5.1.26). Da mesma forma que no exemplo anterior, 2.2.6 \mathcal{X} é gaussiano, o processo estocástico deste exemplo também é gaussiano. Ele pode ser representado por:

$$\mathcal{X}(t, w) = [T] \{A\} = \begin{bmatrix} \cos(\omega t) & \sin(\omega t) \end{bmatrix} \begin{bmatrix} A_1 \\ A_2 \end{bmatrix}, \quad (5.1.32)$$

e, possui média e variância:

$$\mu_{\mathcal{X}}(t) = [T] \{\mu_A\} = \cos(\omega t) + 2 \sin(\omega t), \quad (5.1.33)$$

$$\sigma_{\mathcal{X}}^2(t) = [T][C_A][T]^T = 1 + \frac{\cos(2\omega t)}{2}. \quad (5.1.34)$$

A função densidade de probabilidade $p_{\mathcal{X}(t_j, w)}$ é conhecida para qualquer valor de $t_j \in T$:

$$p_{\mathcal{X}(t_j, w)}(x) = \frac{1}{\sqrt{2\pi} \sqrt{\sigma_{\mathcal{X}}^2(t_j)}} e^{-\frac{(x - \mu_{\mathcal{X}}(t_j))^2}{2\sigma_{\mathcal{X}}^2(t_j)}} \quad (5.1.35)$$

e, o processo é dito estar especificado até primeira ordem.

A partir da densidade de probabilidade (5.1.30), 10^4 realizações foram geradas para esse processo (no intervalo de tempo $[0, 10]$ e com um $\Delta t = 0.1$). Estimou-se a matriz de covariância $[\hat{C}]$ por (5.1.23) e, posteriormente resolveu-se o problema de autovalor (5.1.24).

Os autovalores estimados são mostrados na figura (5.3).

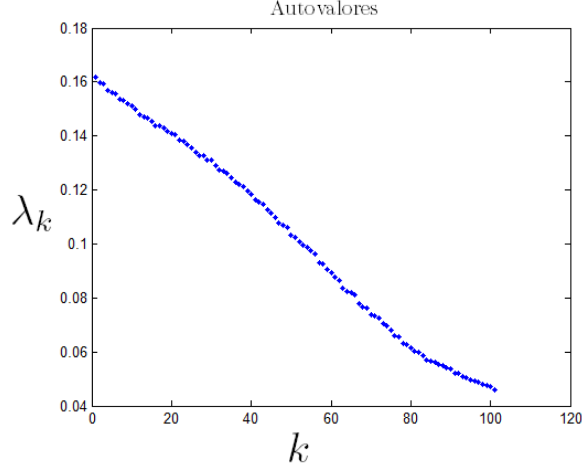


Figura 5.3: Autovalores, λ_k , estimados a partir da matriz de covariância $[\hat{C}]$ do processo (5.1.31).

5.1.6 Estimativas das variáveis aleatórias da expansão de KL

Quando trabalha-se com a expansão de KL de um processo estocástico \mathcal{X} em um domínio discreto, após ter-se calculado estimativas para os autovalores, $\hat{\lambda}_k$, e para as autofunções, $\{\hat{\psi}\}_k$, é necessário estimar-se a função densidade de probabilidade das variáveis aleatórias X_i da expansão.

A partir da expressão (5.1.4), cada realização $x_k^{(i)}$, $i \in \mathbb{N}$, de cada uma das variáveis aleatórias X_k da expansão de KL pode ser calculada por:

$$x_k^{(i)} = \frac{1}{\lambda_k} \langle (\{\mathcal{X}(t, w)\}^{(i)} - \{\hat{\mu}_{\mathcal{X}}(t)\}), \{\hat{\psi}\}_k \rangle_{\mathbb{R}^n}, \quad (5.1.36)$$

onde $\{\mathcal{X}(t, w)\}^{(i)}$ representa uma realização numérica do processo estocástico \mathcal{X} , ou seja é o vetor definido em (5.1.13) e pertence ao \mathbb{R}^n .

Para estimar a função densidade de probabilidade dessa variável aleatória X_k , basta calcular um número m grande de realizações de X_k e, construir o histograma normalizado dessas m realizações.

As m realizações de X_k podem ser expressas por um vetor $\{x_k\}$, $(m \times 1)$, (cada componente representa uma realização de X_k). Esse vetor pode ser calculado por:

$$\{x_k\} = \frac{1}{\lambda_k} [\mathcal{X}_0]^T \{\hat{\psi}\}_k, \quad (5.1.37)$$

onde a matriz $[\mathcal{X}_0]$ está definida em (5.1.22) e tem dimensão $(n \times m)$.

5.2 Geração de realizações processos estocásticos a partir da expansão de KL

A expansão de Karhunen-Loève é uma poderosa técnica de parametrização de um processo estocástico que possibilita sua aproximação através de um processo de dimensão finita, ou seja, um vetor aleatório. Ela pode ser utilizada, por exemplo, para facilitar e reduzir o custo computacional da geração de realizações desse processo.

Quando trabalha-se com o método de Monte Carlo, geralmente deseja-se gerar realizações de uma variável, vetor aleatório ou processo estocástico e posteriormente utilizar as amostras geradas para aproximar alguma função de interesse. Assim, o método utiliza a amostragem aleatória como ferramenta para produzir observações sobre as quais se realizam inferências estatísticas com objetivo de extrair informações sobre as observações (Cf. [44]).

Essa seção mostra como gerar amostras de processos estocásticos, que seriam utilizadas por exemplo no Método de Monte Carlo, a partir da expansão de KL.

Dado um processo estocástico, \mathcal{X} , suponha que o vetor de média $\{\hat{\mu}_{\mathcal{X}}(t)\}$ e a matriz de covariância $[\hat{C}]$ possam ter sido estimados por (5.1.24) e (5.1.23) e, que o problema de autovalor (5.1.24) tenha sido resolvido. Suponha também que a função densidade de probabilidade de cada variável aleatória X_i da expansão de Karhunen-Loève tenha sido estimada através do histograma normalizado de suas amostras (calculadas por (5.1.37)).

Considere, por exemplo, que n autovalores $\hat{\lambda}_k$ e n autovetores $\{\hat{\psi}_k\}$ ($n \times 1$) foram calculados.

Pela expansão truncada de KL, com d termos ($d < n$), para esse processo, verifica-se que cada realização numérica de \mathcal{X} pode ser escrita como:

$$\{\mathcal{X}^{(KL)}(t, w)\}^{(i)} = \{\hat{\mu}_{\mathcal{X}}(t)\} + \sum_{k=1}^d \sqrt{\hat{\lambda}_k} \{\hat{\psi}_k\} x_k^{(i)}. \quad (5.2.38)$$

onde $x_k^{(i)}$ representa uma realização de variável aleatória X_k .

Assim, vetores $\{\mathcal{X}(t, w)\}^{(i)}$ são gerados através de realizações de um número finito, d , das variáveis aleatórias X_k .

EXEMPLO 5.2.1 *Suponha que deseja-se gerar-se realizações através da expansão de KL de um processo estocástico, \mathcal{X} , com parâmetro $t \in T$, sendo $t \geq 0$, caracterizado por um passeio aleatório com incrementos com densidade de probabilidade Γ , tal que:*

1. $\mathcal{X}(0, w) = 0$;
2. $\mathcal{X}(t, w)$, $t \geq 0$, é um processo estocástico com incrementos independentes;
3. a densidade de probabilidade da variável aleatória $Y = \mathcal{X}(t_2, w) - \mathcal{X}(t_1, w)$, com $t_2 > t_1 \geq 0$, é Γ . Sua média é $\mu_Y = m(t_2 - t_1)$ e seu coeficiente de variação é $\delta_Y = \frac{\delta}{\sqrt{t_2 - t_1}}$, com $m > 0$ e $\delta > 0$ fixados.

Para gerar-se realizações através da expansão de KL de \mathcal{X} , primeiramente foram geradas 10^4 realizações desse processo (no intervalo de tempo $[0, 10]$ e com diferentes discretizações de tempo Δt). Através dessas realizações, estimou-se a matriz de covariância $[\hat{C}]$ por (5.1.23) e, posteriormente resolveu-se o problema de autovalor (5.1.24).

As aproximações obtidas para os vinte primeiros autovalores $\hat{\lambda}_k$ são mostradas na figura 5.4.

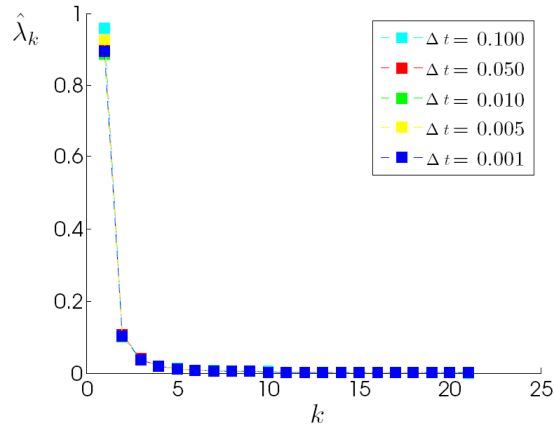


Figura 5.4: Estimativas dos autovalores $\hat{\lambda}_k$ da matriz $[\hat{C}]$ para diferentes discretizações de tempo Δt .

A partir dos histogramas normalizados das variáveis aleatórias da expansão de KL (5.1.37), foi possível estimar as funções densidades de probabilidade de cada variável X_k . Os gráficos das densidades estimadas para as quatro primeiras variáveis são mostradas em (5.5).

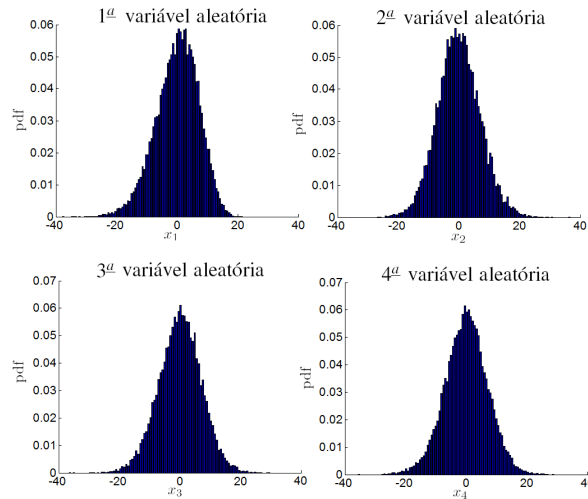


Figura 5.5: Estimativas para as densidade de probabilidade das quatro primeiras variáveis aleatórias X_k da expansão de KL.

Tendo as densidades de probabilidade das variáveis aleatórias X_k , amostras de cada uma dessas variáveis foram obtidas através do Método de Monte Carlo Cadeia de Markov (MCMC). Assim, pela expressão da expansão truncada de KL (5.2.38) foi possível calcular realizações de \mathcal{X} .

Para analisar a influência do número de termos d usados na expressão truncada de KL, duas estratégias foram utilizadas.

Primeira estratégia: a partir das amostras de \mathcal{X} geradas por KL, estimou-se a matriz de covariância do processo, $[\hat{C}_d]$, calculou-se a norma Frobenius dessa matriz e analisou-se o erro:

$$Erro = \frac{\left| [\hat{C}_d] \right|_{Fr} - \left| [\hat{C}] \right|_{Fr}}{\left| [\hat{C}] \right|_{Fr}}$$

para diferentes números de termos d usados na expressão truncada de KL. Os resultados obtidos são mostrados na figura (5.6) mostra. Verifica-se que quanto maior o número de termos usados na expansão, menor será o erro (5.2.39).

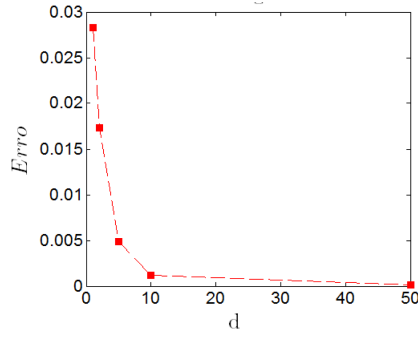


Figura 5.6: Erro (5.2.39) obtido para diferentes números de termos d .

Segunda estratégia: fez-se uma comparação entre:

- a função densidade de probabilidade da variável aleatória $\mathcal{X}(t_j, w)$ (para algum $t_j \in T$ fixo);
- o histograma normalizado da variável aleatória $\mathcal{X}^{(KL)}(t_j, w)$, calculado através das realizações do processo estocástico geradas pela expansão de KL.

Nas figuras (5.7) e (5.8) são mostrados as densidades $p_{\mathcal{X}(t=2, w)}$ (nesse caso, $t_j = 2$) e os histogramas normalizados da variável aleatória $\mathcal{X}^{(KL)}(t = 2, w)$ para os casos em que consideraram-se $d = 1$, $d = 2$, $d = 5$ e $d = 10$ termos na expansão de KL. Pelas figuras, verifica-se que quanto se aumenta o número de termos usados na expansão, o histograma de $\mathcal{X}^{(KL)}(t = 2, w)$ se aproxima da densidade $p_{\mathcal{X}(t=2, w)}$.

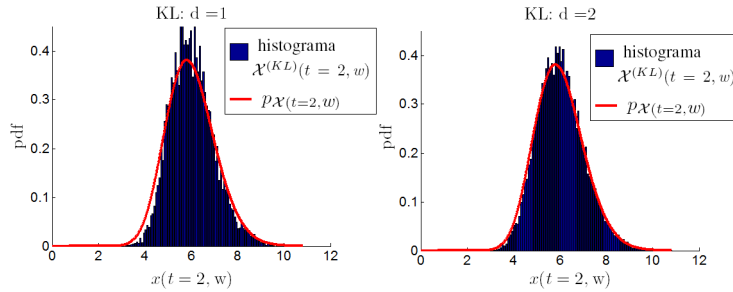


Figura 5.7: Densidades $p_{\mathcal{X}(t=2, w)}$ e histogramas normalizados de $\mathcal{X}^{(KL)}(t = 2, w)$ para $d = 1$ e $d = 2$ termos.

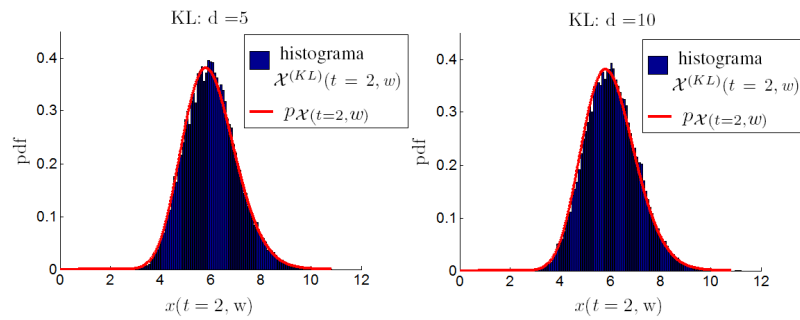


Figura 5.8: Densidades $p_{\mathcal{X}(t=2, w)}$ e histogramas normalizados de $X^{KL}(t=2, w)$ para $d=5$ e $d=10$ termos.

5.3 Exercícios

EXERCÍCIO 5.3.1 *Implemente uma rotina capaz de estimar os 100 primeiros autovalores do processo estocástico do exemplo (5.1.4) através da matriz de covariância do processo.*

EXERCÍCIO 5.3.2 *Implemente uma rotina capaz de estimar os 100 primeiros autovalores do processo estocástico do exemplo (5.1.5) através da matriz de covariância do processo.*

EXERCÍCIO 5.3.3 *Implemente uma rotina que construa os histogramas das amostras das cinco primeiras variáveis aleatórias da expansão de Karhunen-Loève do processo estocástico, \mathcal{X} , com parâmetro $t \in T$, sendo $t \geq 0$, do exemplo (5.2.1).*

Capítulo 6

Propagação de Incertezas

6.1 Propagação de incertezas através do método de Monte Carlo

Quando parâmetros de um sistema são considerados objetos aleatórios (como por exemplo, variáveis e vetores aleatórios), as incertezas associadas a esses parâmetros são propagadas para a resposta do sistema.

Assim, a resposta, antes considerada determinística, passa a ser aleatória e depende diretamente do modelo estocástico construído para os parâmetros. Uma forma de quantificar essa incerteza propagada é fazer-se inferências de estatísticas dessa resposta através do método de Monte Carlo.

Como visto no segundo capítulo, o trabalho de simulações estocásticas de um sistema geralmente envolve:

1. a construção de um modelo determinístico para o sistema;
2. a construção de um modelo estocástico paramétrico ou não paramétrico para o sistema;
3. a aplicação do método de Monte Carlo para obter-se inferências estatísticas sobre a resposta do sistema.

A seguir são mostrados dois exemplos de propagação de incertezas em sistemas dinâmicos.

EXEMPLO 6.1.1 *Seja x uma função real que associa a cada valor de tempo $t \in \mathbb{R}^{\geq 0}$ um valor $x(t) \in \mathbb{R}$ e representa a resposta de algum sistema dinâmico em função do*

tempo. Considere que a dinâmica desse sistema é regida pela equação diferencial:

$$\frac{dx(t)}{dt} = a x(t), \quad (6.1.1)$$

com a sendo uma constante real e com condição inicial $x(0) = x_0 = 1.0$. Essa equação possui solução analítica:

$$x(t) = e^{at}, \quad (6.1.2)$$

e representa o modelo determinístico do sistema dinâmico. Suponha agora que a constante que multiplica $x(t)$ em (6.1.1) é uma variável aleatória contínua, representada por A . Para construir seu modelo estocástico, é necessário determinar sua função densidade de probabilidade, ou seja, p_A . Uma opção para isso é utilizar o Princípio da Máxima Entropia.

Considere, por exemplo, que a única informação disponível sobre A seja o seu suporte $k = [1, 2]$. Assim, através do PEM, verifica-se que a função que maximiza a entropia de Shannon é a uniforme no intervalo $[1, 2]$, isto é:

$$p_A(x) = \mathbb{1}_{[1,2]}(x) \frac{1}{2}. \quad (6.1.3)$$

A incerteza associada ao parâmetro A é propagada para a resposta do sistema, de forma que essa resposta, antes considerada determinística, passa a ser aleatória. Ela é caracterizada por um processo estocástico de parâmetro $t \in \mathbb{R}^{\geq 0}$, representado por \mathcal{X} .

Para cada valor t_j de t , $\mathcal{X}(t_j, w)$ é uma variável aleatória com uma função densidade de probabilidade dependente de p_A . Uma forma de quantificar essa propagação incerteza é fazer-se inferências de estatísticas da resposta através do método de Monte Carlo.

Esse exemplo foi implementado em uma rotina MATLAB gerando-se m amostras da variável aleatória A . Para cada amostra $a^{(i)}$, uma realização numérica do processo estocástico, \mathcal{X} , foi calculada através do modelo determinístico do sistema (6.1.2).

Vale observar que no capítulo 5, apresentou-se a definição de uma realização numérica de um processo estocástico. Dado um número finito, n , de valores do parâmetro t , isto é, t_1, \dots, t_n , cada realização de \mathcal{X} é um vetor de n componentes $\{\mathcal{X}(t, w)\}^{(i)}$, em que cada componente j representa uma realização da variável aleatória $\mathcal{X}(t_j, w^{(i)})$.

A partir dessas m realizações de \mathcal{X} , é possível estimar um vetor de média e um vetor de variância do processo como mostrado em (5.1.15) e (5.1.16). Eles representam estimativas das funções $\mu_{\mathcal{X}}$ e $\sigma_{\mathcal{X}}^2$ avaliadas nos instantes selecionados t_1, \dots, t_n .

Na figura (6.1) são mostradas essas estimativas de média e variância do processo obtidas através do método de Monte Carlo. Essa figura recebe o nome especial de gráfico envelope.

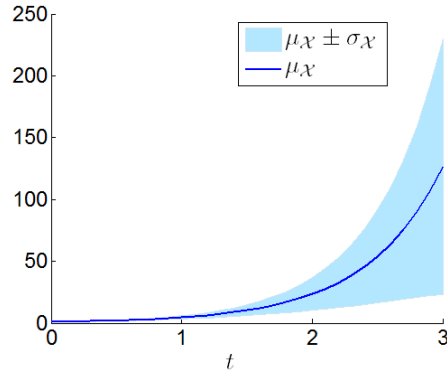


Figura 6.1: Estimativas de média e variância de \mathcal{X} ao longo do tempo.

Observando (6.1), verifica-se que o instante em que o processo apresenta maior variância é o último da simulação, no caso $t = 3.0$. Construindo o histograma normalizado com as m amostras da variável aleatória $\mathcal{X}(t = 3.0, w)$, obtém-se uma aproximação para a sua função densidade de probabilidade, $p_{\mathcal{X}(3.0, w)}$. Esse histograma é mostrado em (6.2).

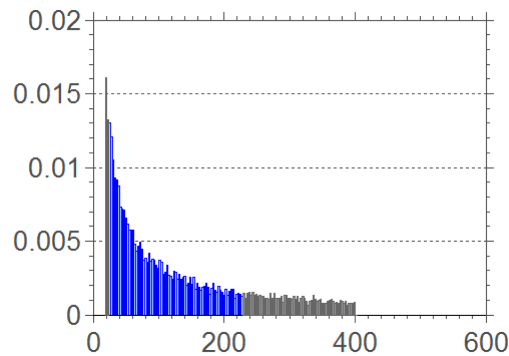


Figura 6.2: Histograma normalizado construído com 2×10^4 amostras de $\mathcal{X}(t = 3.0, w)$. A parte destacada em azul representa o intervalo $[\mu_{\mathcal{X}(t=3.0, w)} - \sigma_{\mathcal{X}(t=3.0, w)}, \mu_{\mathcal{X}(t=3.0, w)} + \sigma_{\mathcal{X}(t=3.0, w)}]$.

EXEMPLO 6.1.2 Considere a equação da dinâmica do sistema do exemplo anterior (6.1.1). Suponha que nessa equação a constante que multiplica $x(t)$ em (6.1.1) uma variável aleatória contínua, representada por A . Sabe-se que essa variável:

- só assume valores positivos, isto é seu suporte é $[0, \infty)$;
- tem valor esperado conhecido, $E[A] = \mu$;
- tem $E[\log(A)] < \infty$.

Uma opção para determinar a função densidade de probabilidade de A , p_A , e assim construir o modelo estocástico do sistema, é aplicar o Princípio da Máxima Entropia utilizando somente as informações disponíveis sobre A .

Através do PEM, verifica-se que a função que maximiza a entropia de Shannon é a densidade de probabilidade Gamma, isto é:

$$p_A(x) = \mathbb{1}_{[0, +\infty)}(x) \frac{1}{\mu} \left(\frac{1}{\delta^2}\right)^{\frac{1}{\delta^2}} \frac{1}{\Gamma(1/\delta^2)} \left(\frac{x}{\mu}\right)^{\frac{1}{\delta^2}-1} \exp\left(-\frac{x}{\delta^2\mu}\right), \quad (6.1.4)$$

onde:

- Γ é a função Gamma: $\Gamma(a) = \int_0^\infty t^{a-1} \exp(-t) dt$;
- $\delta = \frac{\sigma}{\mu}$ é um fator de dispersão adimensional.

A incerteza associada ao parâmetro A é propagada para a resposta do sistema, de forma que ela é caracterizada por um processo estocástico de parâmetro $t \in \mathbb{R}^{\geq 0}$, representado por \mathcal{X} .

Fazendo um procedimento semelhante ao feito no exemplo anterior (6.1.1), obtém-se o gráfico envelope para a resposta do sistema estocástico, como mostrado na figura (6.3) e, obtém-se uma aproximação para função densidade de probabilidade, $p_{\mathcal{X}(t=3.0, w)}$ através do histograma normalizado construído com as amostras da variável aleatória $\mathcal{X}(t = 3.0, w)$, (6.4).

Comparando os gráficos envelope e os histogramas normalizados dos dois últimos exemplos, verifica-se a influência que o modelo estocástico adotado para o parâmetro A do sistema exerce sobre a resposta estocástica do sistema \mathcal{X} .

EXEMPLO 6.1.3 Seja x uma função real que associa a cada valor de tempo $t \in \mathbb{R}^{\geq 0}$ um valor $x(t) \in \mathbb{R}$ e representa a resposta de algum sistema dinâmico em função do tempo. Considere que a dinâmica desse sistema é regida pela equação diferencial:

$$\frac{dx(t)}{dt} = a t x(t), \quad (6.1.5)$$

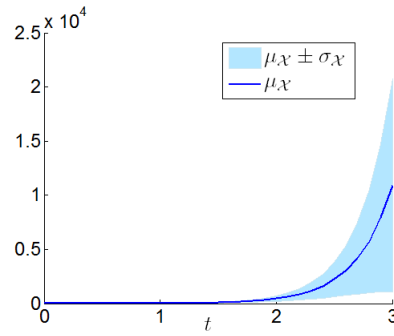


Figura 6.3: Estimativas de média e variância de \mathcal{X} ao longo do tempo.

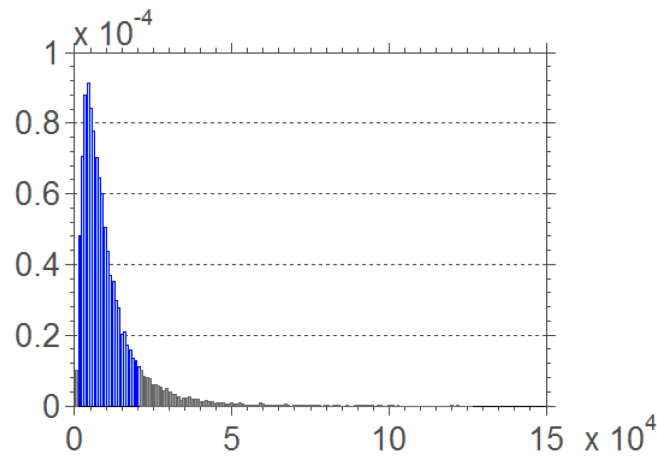


Figura 6.4: Histograma normalizado construído com 2×10^4 amostras de $\mathcal{X}(t = 3.0, w)$. A parte destacada em azul representa o intervalo $[\mu_{\mathcal{X}(t=3.0, w)} - \sigma_{\mathcal{X}(t=3.0, w)}, \mu_{\mathcal{X}(t=3.0, w)} + \sigma_{\mathcal{X}(t=3.0, w)}]$.

onde a é uma constante real e com condição inicial $x(0) = x_0 = 1.0$. Essa equação possui solução analítica:

$$\begin{aligned}
 \frac{dx(t)}{x(t)} &= a t dt \\
 \ln(x(t)) &= \frac{a t^2}{2} + c_1 \\
 x(t) &= e^{\left(\frac{a t^2}{2}\right)} + c_2.
 \end{aligned} \tag{6.1.6}$$

onde c_1 e c_2 são constantes reais. Como $x(0) = 1 + c_2 = 1.0$, tem-se que $c_2 = 0$.

Suponha agora que a constante que multiplica $x(t)$ em (6.1.5) é uma variável aleatória contínua, representada por A , com função densidade de probabilidade uniforme no intervalo $[1, 2]$. Dessa forma, o produto At é um processo estocástico de parâmetro $t \in \mathbb{R}^{\geq 0}$ e, a incerteza associada a esse processo é propagada para a resposta do sistema, de forma que ela é caracterizada por um processo estocástico de parâmetro $t \in \mathbb{R}^{\geq 0}$, representado por \mathcal{X} .

Fazendo um procedimento semelhante ao feito nos dois exemplos anteriores, (6.1.1) e (6.1.2), obtém-se o gráfico envelope para a resposta do sistema estocástico, como mostrado na figura (6.5) e, obtém-se uma aproximação para função densidade de probabilidade, $p_{\mathcal{X}(t=3.0, w)}$ através do histograma normalizado construído com as amostras da variável aleatória $\mathcal{X}(t = 3.0, w)$, mostrado na figura (6.6).

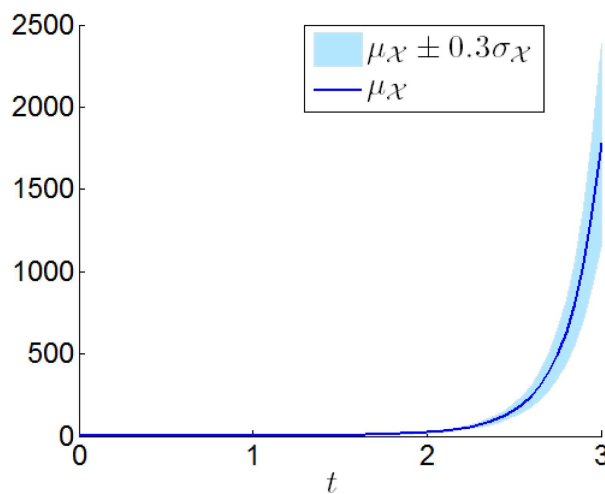


Figura 6.5: Estimativas de média e variância de \mathcal{X} ao longo do tempo.

Comentários: Os três exemplos apresentados para propagação de incertezas são acadêmicos e escolhidos exatamente devido a simplicidade. Na bibliografia citamos alguns artigos onde se encontram exemplos de modelagem estocástica e de propagação de incertezas nas áreas de produção de voz e colunas de perfuração de petróleo. Esse minicurso toca apenas em alguns problemas, os interessados deverão consultar a bibliografia aconselhada. Os livros de [11, 6, 26, 4, 19] devem ser examinados pois são obras-primas. Outros livros também foram citados que ampliam o escopo da teoria.

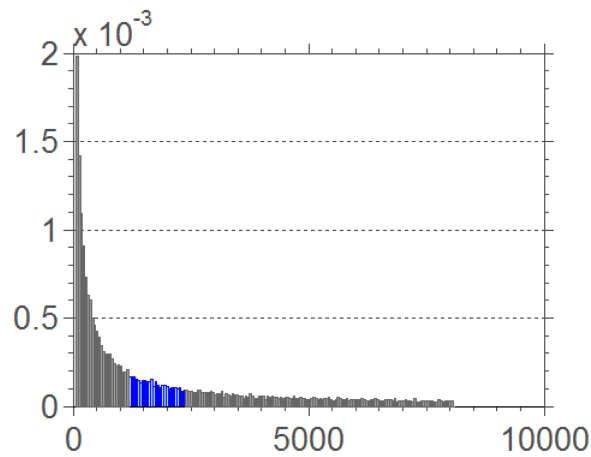


Figura 6.6: Histograma normalizado construído com 3×10^4 amostras de $\mathcal{X}(t = 3.0, w)$. A parte destacada em azul representa o intervalo $[\mu_{\mathcal{X}(t=3.0, w)} - 0.3 \sigma_{\mathcal{X}(t=3.0, w)}, \mu_{\mathcal{X}(t=3.0, w)} + 0.3 \sigma_{\mathcal{X}(t=3.0, w)}]$.

6.2 Exercícios

EXERCÍCIO 6.2.1 *Implemente uma rotina para refazer o exemplo (6.1.1).*

EXERCÍCIO 6.2.2 *Implemente uma rotina para refazer o exemplo (6.1.2).*

EXERCÍCIO 6.2.3 *Implemente uma rotina para refazer o exemplo (6.1.3).*

Bibliografia

- [1] J. P. Albuquerque e W. Finamore. *Probabilidade, Variáveis Aleatórias e Processos Estocásticos*. Editora PUC-Rio e Editora Interciência, Rio de Janeiro, Brasil, 2008.
- [2] L. Arnold. *Random Dynamical Systems*. Springer-Verlag, 1998.
- [3] S. Bellizzi e R. Sampaio. Smooth Karhunen-Loève decomposition to analyze randomly vibrating systems. *Journal of Sound and Vibration*, 325(3):491–498, 2009.
- [4] L. Breiman. *Probability*. Addison-Wesley Publishing Co., Reading, Massachusetts, 1968.
- [5] E. Cataldo, C. Soize, R. Sampaio e C. Descelliers. Probabilistic modelling of a nonlinear dynamical system used for producing voice. *Computational Mechanics*, 43:265–275, 2009.
- [6] K. L. Chung. *A Course in Probability Theory*. Academic Press, USA, third edition, 2001.
- [7] L. Devroye e D. P. Kroese. *Non-Uniform Random Variate Generation*. Springer-Verlag, New York, USA, 1986.
- [8] W. L. Dunn e J. K. Shultis. *Exploring Monte Carlo Methods*. Elsevier, USA, 2010.
- [9] M. Emery. *Stochastic Calculus in Manifolds*. Universitext Springer-Verlag, 1990.
- [10] A. T. Fabro, T. Ritto, R. Sampaio e J. R. Arruda. Stochastic analysis of a cracked rod modeled via the spectral element method. *Mechanics Research Communications*, 37: 326 - 331, 2010.

- [11] W. Feller. *An introduction to probability theory and its applications*, volume vol.1 (3rd ed.) and vol.2 (2nd ed.). John Wiley & Sons, Inc., New York, 1968 and 1971 [1st ed. of vol. 1,1950].
- [12] J.E. Gentle. *Random Number Generation and Monte Carlo Methods*. Springer, New York, USA, second edition, 2003.
- [13] G. Grimmett e Dominic. Welsh. *Probability an Introduction*. Oxford Science Publications, USA, New York, 1986.
- [14] O. Haggstrom. *Finite Markov Chains and Algorithmic Applications*. Cambridge University Press, USA, 2002.
- [15] Kapur J.N. e Kesavan H.K. *Entropy Optimization Principles with Applications*. Academic Press, Inc., USA, 1992.
- [16] E. P. Hsu. *Stochastic Analysis on Manifolds*. American Mathematical Society, Providence, RI, 2002.
- [17] E. Jaynes. Information theory and statistical mechanics. *The Physical Review*, 106(4):620–630, 1957.
- [18] E. Jaynes. Information theory and statistical mechanics ii. *The Physical Review*, 108:171–190, 1957.
- [19] M. Kac. *Statistical Independence in Probability, Analysis and Number Theory, Carus Mathematical Monograph 12*. John Wiley & Sons, Inc., New York, 1959.
- [20] D. Knuth. *Seminumerical Algorithms: The Art of Computer Programming*, volume 2. Addison Wesley, USA, third edition, 1998.
- [21] A. Kolmogoroff. *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Springer-Verlag, Berlin, 1933.
- [22] D. P. Kroese. *Monte Carlo Methods*. Lectures notes on Monte Carlo methods used at the 2011 Summer School of the Australian Mathematical Sciences Institute (AMSI), University of Queensland, 2011.
- [23] J. S. Liu. *Monte Carlo Strategies in Scientific Computing*. Springer, USA, 2001.
- [24] M. Loève. *Probability Theory I*. Springer-Verlag, USA, 1977.
- [25] J. Mauprivez, E. Cataldo e R. Sampaio. Estimation of random variables associated to a model for the vocal folds using anns. *Inverse Problems in Science & Engineering*, 20(2):209–225, 2012.

- [26] J. Neveu. *Mathematical foundations of the calculus of probability*. Holden-Day, Inc., San Francisco, 1965.
- [27] B. Oksendal. *Stochastic Differential Equations, an Introduction with Applications*. Universitext Springer-Verlag, sixth edition, 2003.
- [28] A. Papoulis. *Probability, Random Variables and Stochastic Processes*. McGraw-Hill Science, USA, third edition, 1991.
- [29] T. Ritto e R. Sampaio. Stochastic drill-string dynamics with uncertainty on the imposed speed and on the bit-rock parameters. *International Journal for Uncertainty Quantification*, 2(2):111–124, 2012.
- [30] T. Ritto, C. Soize e R. Sampaio. Nonlinear dynamics of a drill-string with uncertainty model of the bit-rock interaction. *International Journal of Non-Linear Mechanics*, 44:865–876, 2009.
- [31] T. Ritto, C. Soize e R. Sampaio. Robust optimization of the rate of penetration of a drill-string using a stochastic nonlinear dynamical model. *Computational Mechanics*, 45(5):415–427, 2010.
- [32] T. Ritto, C. Soize e R. Sampaio. Probabilistic model identification of the bit-rock-interaction-model uncertainties in nonlinear dynamics of a drill-string. *Mechanics Research Communications*, 37: 584-589, 2010.
- [33] T. Ritto, C. Soize e R. Sampaio. Uncertainties in drill-string dynamics: the role of a random weight-on-hook. *Journal of the Brazilian Soc. Mechanical Sciences*, 32(3):250–258, 2010.
- [34] T. Ritto, F. Buezas e R. Sampaio. A new measure of efficiency for model reduction: application to a vibroimpact system. *Journal of Sound and Vibration*, 330:1977–1984, 2011.
- [35] T. Ritto, R. Holdorf, R. Sampaio e E. S. de Cursi. Robust optimization of a flexible rotor-bearing system using the campbell diagram. *Engineering Optimization*, 43(1):77–96, 2011.
- [36] T. Ritto, R. Sampaio e F. Rochinha. Model uncertainties of flexible structures vibrations induced by internal flows. *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, 33(3):373–380, 2011.
- [37] C. P. Robert e G. Casella. *Monte Carlo Statistical Methods*. Springer, USA, second edition, 2004.
- [38] S. M. Ross. *Simulation*. Elsevier, USA, 2006.

- [39] R. Y. Rubinstein. *Simulation and the Monte Carlo Method*. John Wiley & Sons, USA, 2008.
- [40] R. Sampaio e S. Bellizzi. Poms analysis of randomly vibrating systems obtained from Karhunen-Loève expansion. *Journal of Sound and Vibration*, 279:774–793, 2006.
- [41] R. Sampaio e E. Cataldo. Two strategies to model uncertainties in structural dynamics. *Shock and Vibration*, 17(2):171–186, 2010.
- [42] R. Sampaio, C. Wolter e M. Trindade. Reduced-order model for an impacting beam using the Karhunen-Loève expansion. *Tendências em Matemática Aplicada e Computacional, TEMA*, 3(2):217–226, 2002.
- [43] C.E. Shannon. A mathematical theory of communication. *Bell System Tech. J.*, 27:379–423 and 623–659, 1948.
- [44] R. Shonkwiler e F. Mendivil. *Explorations in Monte Carlo Methods*. Springer, New York, USA, 2009.
- [45] I.M. Sobol. *A Primer for the Monte Carlo Method*. CRC Press, USA, 1994.
- [46] C. Soize. A nonparametric model of random uncertainties for reduced matrix models in structural dynamics. *Probabilistic Engineering Mechanics*, 15: 277–294, 2000.
- [47] J. E. Souza de Cursi e R. Sampaio. *Modelagem Estocástica e Quantificação de Incertezas*. Sociedade Brasileira de Matemática Aplicada e Computacional, São Carlos (SP), Brasil, 2012.
- [48] M. R. Spiegel. *Probabilidade e Estatística (Coleção Schaum)*. McGraw-Hill do Brasil, São Paulo, Brasil, 1978.
- [49] M. Trindade, C. Wolter e R. Sampaio. Karhunen-Loève decomposition of coupled axial-bending vibrations of beams subject to impacts. *Journal of Sound and Vibration*, 279(3–5):1015–1036, 2005.
- [50] B. Walsh. *Markov Chain Monte Carlo and Gibbs Sampling*. Lecture Notes, EEB 581, april 2004.
- [51] T. L. Ortiz, K. Wirsching e P. Paez. *Random Vibrations: Theory and Practice*. John Wiley & Sons, USA, 1995.
- [52] D. Xiu. *Numerical Methods for Stochastic Computations*. Princeton University Press, USA, 2010.

Índice

- cadeia de Markov, 93
 - Chapman-Kolmogorov, 94
 - distribuição de probabilidade estacionária, 95
 - distribuição de probabilidade inicial, 93
 - distribuição de probabilidade reversível, 96
 - grafo, 96
 - irredutibilidade, 94
 - matriz de transição, 93
 - período, 95
- convergência
 - distribuição, 22
 - média, 22
 - média quadrática, 22
 - probabilidade, 22
 - quase certa, 22
- expansão de Karhunen-Loève, 116
- gerador, 69
 - congruência linear, 71
 - método da rejeição, 81
 - método da Transformada Inversa, 75
 - método de Box-Muller, 78
- lei dos grandes números, 57
- método de Box-Muller, 78
- método de Monte Carlo, 34, 56
 - integração, 61
- Metropolis-Hastings
 - densidade de probabilidade proposta, 105
 - variáveis aleatórias, 101
 - vetores aleatórios, 109
- Princípio da Entropia Máxima, 35
 - matrizes aleatórias, 49
 - variáveis aleatórias contínuas, 39
 - variáveis aleatórias discretas, 35
 - vetores aleatórios, 46
- processo estocástico, 30
 - autocorrelação, 30
 - covariância, 30
 - média, 30
 - Markov, 92
 - variância, 30
- propagação de incertezas, 131
- teorema central do limite, 58
- variável aleatória contínua, 17
 - distribuição de probabilidade cumulativa, 17
 - densidade de probabilidade, 18
 - média, 19
 - mediana, 19
 - quantil, 19
 - variância, 19
- variável aleatória discreta, 14
 - distribuição de probabilidade cumulativa, 15
 - função de massa, 14

- média, 16
- mediana, 17
- quantil, 17
- variância, 16
- vetor aleatório, 25
 - distribuição cumulativa de probabilidade, 25
 - covariância, 28
 - densidade de probabilidade, 26
 - média, 27
 - variância, 27