

Probabilidade, Estatística e Processos Estocásticos

Carlos Alberto Ynoguti

25 de janeiro de 2011

Agradecimentos

Ao Prof. Dr. Dayan Adionel Guimarães pela criteriosa revisão do texto.

Sumário

Lista de Figuras	vii
1 Probabilidade	1
1.1 Introdução.	1
1.2 Teoria de Conjuntos.	2
1.2.1 Lei de De Morgan.	4
1.2.2 Princípio da Dualidade.	5
1.3 Definições de Probabilidade.	5
1.3.1 Frequência Relativa.	5
1.3.2 Axiomática.	6
1.3.3 Clássica.	7
1.4 Cálculo de probabilidades usando métodos de contagem.	7
1.4.1 Amostragem com reposição e ordenação.	8
1.4.2 Amostragem sem reposição e com ordenação.	8
1.4.3 Permutação de n objetos distintos.	9
1.4.4 Amostragem sem reposição e sem ordenação.	10
1.4.5 Amostragem com reposição e sem ordenação.	11
1.5 Probabilidade Conjunta.	11
1.5.1 Probabilidades Marginais.	12
1.6 Probabilidade Condicional.	12
1.6.1 Regra de Bayes.	13
1.7 Eventos independentes.	14
1.8 Experimentos sequenciais e diagramas em árvore	16
1.9 Exercícios	19
2 Variáveis Aleatórias	25
2.1 Definição.	25
2.2 Função distribuição cumulativa.	27
2.3 Tipos de Variáveis Aleatórias	30
2.3.1 Discretas	30
2.3.2 Contínuas	31
2.3.3 Mistas	32
2.4 Função Densidade de Probabilidade	33
2.4.1 Definição	33
2.4.2 Propriedades	34
2.4.3 Caso Discreto	35
2.5 Algumas variáveis aleatórias discretas importantes	36
2.5.1 Bernoulli	36

2.5.2	Binomial	37
2.5.3	Poisson	37
2.5.4	Geométrica	38
2.6	Algumas variáveis aleatórias contínuas importantes	38
2.6.1	Uniforme	39
2.6.2	Exponencial	40
2.6.3	Rayleigh	40
2.6.4	Gaussiana	41
2.6.5	Gama	45
2.6.6	m-Erlang	46
2.6.7	Chi-Quadrado (χ^2)	47
2.6.8	Cauchy	47
2.6.9	Laplace	48
2.7	Densidades Condicionais	49
2.8	Variáveis Aleatórias Múltiplas	51
2.8.1	Função Distribuição de Probabilidade Conjunta	51
2.8.2	Densidades marginais	52
2.8.3	Caso multidimensional	53
2.8.4	Função distribuição de probabilidade condicional	54
2.8.5	Independência Estatística de Variáveis Aleatórias	56
2.9	Funções de Variáveis Aleatórias	56
2.9.1	Caso Unidimensional	56
2.9.2	Caso Multidimensional	61
2.10	Exercícios	64
3	Médias Estatísticas de Variáveis Aleatórias	72
3.1	Médias	72
3.1.1	Média de uma Variável Aleatória	72
3.1.2	Média de uma Função de uma Variável Aleatória	74
3.1.3	Médias para Variáveis Múltiplas	75
3.1.4	Média da Soma de Funções	76
3.1.5	Média do Produto de Duas Variáveis Aleatórias Independentes	77
3.1.6	Média Quadrática da Soma de Duas Variáveis Aleatórias	77
3.1.7	Média condicional	78
3.2	Momentos	78
3.2.1	N -ésimo momento	78
3.2.2	Momentos Centrais	79
3.2.3	Variância	79
3.2.4	Caso Multidimensional	80
3.2.5	Variáveis Aleatórias Descorrelacionadas e Ortogonais	82
3.3	Funções Características	83
3.3.1	Caso multidimensional	85
3.4	Exercícios	86
4	Métodos computacionais para geração de números aleatórios	90
4.1	Introdução	90
4.2	Método do resíduo da potência	90
4.3	Método da transformada	92

4.4	O método da rejeição	94
4.5	Geração de funções de uma variável aleatória	97
4.6	Geração de misturas de variáveis aleatórias	98
4.7	Exercícios	98
5	Somas de Variáveis Aleatórias e o Teorema do Limite Central	100
5.1	Introdução	100
5.2	Médias de somas	100
5.3	Fdp da soma de duas v.a.'s	103
5.4	Função geratriz de momentos	105
5.5	FGM da soma de v.a.'s independentes	109
5.6	Somas de v.a.'s gaussianas independentes	111
5.7	Somas aleatórias de v.a.'s independentes	112
5.8	Teorema do limite central	116
5.9	Aplicações do Teorema do Limite Central	118
5.10	Exercícios	120
6	Limitantes Superiores para a Probabilidade de Cauda	125
6.1	Desigualdade de Markov	125
6.2	Desigualdade de Chebyshev	126
6.3	Limitante de Chernoff	128
6.4	Exercícios	129
7	A média amostral	132
7.1	Introdução	132
7.2	Valor esperado e variância	132
7.3	Média amostral de números grandes	134
7.4	Leis de Números Grandes	135
7.4.1	Lei Fraca de Números Grandes	136
7.4.2	Lei Forte de Números Grandes	137
7.5	Exercícios	138
8	Processos Estocásticos	140
8.1	Definição	140
8.2	Tipos de procesos estocásticos	142
8.3	Variáveis aleatórias a partir de processos estocásticos	143
8.4	Sequências aleatórias independentes e identicamente distribuídas	145
8.5	Processo de Contagem	147
8.6	Processo de Poisson	148
8.7	Processo sinal telegráfico aleatório	152
8.8	Processo movimento Browniano	154
8.9	Médias estatísticas de processos aleatórios	155
8.9.1	Momentos	159
8.9.2	Função de autocovariância	159
8.10	Classificação dos processos estocásticos	160
8.10.1	Processos estocásticos estacionários e não estacionários	160
8.10.2	Processos estacionários no sentido amplo	161
8.10.3	Processos ergódicos	164
8.11	Exercícios	166

9	Processamento de Sinais Aleatórios	173
9.1	Sistemas lineares e invariantes no tempo	173
9.2	Filtragem linear de um processo estocástico	174
9.3	Espectro densidade de potência	177
9.4	Correlações cruzadas	181
9.4.1	Função de correlação cruzada	181
9.4.2	Densidade espectral cruzada	184
9.4.3	Filtragem de processos estocásticos	186
9.5	Processos gaussianos	188
9.6	Processo ruído branco gaussiano	191
9.7	Exercícios	193
10	Cadeias de Markov	199
10.1	Processos de Markov	199
10.2	Cadeias de Markov de Tempo discreto	202
10.2.1	Probabilidade de transição para n passos	203
10.2.2	Probabilidades dos estados	204
10.2.3	Probabilidades em regime	206
10.3	Cadeias de Markov em tempo contínuo	207
10.3.1	Tempos de ocupação de estados	209
10.3.2	Taxas de transição e probabilidades de estados dependentes de tempo	210
10.4	Probabilidades de Estados em Regime e Equações de Balanço Globais	214
10.5	Classes de estados, propriedades de recorrência e probabilidades limite	218
10.5.1	Classes de estados	218
10.5.2	Propriedades de recorrência	220
10.5.3	Probabilidades limite	223
10.5.4	Probabilidades limite para as cadeias de Markov de tempo contínuo	226
10.6	Exercícios	228
A	Tabelas Matemáticas	234
A.1	Identidades trigonométricas	234
A.2	Coeficientes Binomiais	235
A.3	Derivadas	235
A.4	Integrais indefinidas	236
A.5	Integrais definidas	237
B	Tabelas de transformadas de Fourier	238
B.1	Definição	238
B.2	Propriedades	238
B.3	Pares de transformadas	239
C	Séries de Taylor	240
C.1	Série de Taylor para funções de uma variável	240
C.2	Expansões mais utilizadas	240

D	Variáveis aleatórias discretas	242
D.1	Bernoulli	242
D.2	Binomial	242
D.3	Geométrica	242
D.4	Binomial negativa	243
D.5	Poisson	243
E	Variáveis aleatórias contínuas	244
E.1	Uniforme	244
E.2	Exponencial	244
E.3	Gaussiana (Normal)	244
E.4	Gama	245
E.5	m-Erlang	245
E.6	Chi-Quadrado (χ^2)	245
E.7	Rayleigh	245
E.8	Cauchy	246
E.9	Laplace	246
F	Valores da distribuição normal	247
	Bibliografia	250

Lista de Figuras

1.1	Espaço amostral para o arremesso de um dado.	2
1.2	Representação do a) complemento, b) união, c) interseção de eventos, e d) eventos disjuntos.	4
1.3	Demonstração da lei de De Morgan.	4
1.4	Espaço amostral para a derivação da regra de Bayes.	13
2.1	Uma v.a. associa um número $x = X(\zeta)$ a cada resultado ζ no espaço amostral S de um experimento aleatório.	25
2.2	Eventos equivalentes.	27
2.3	$P[a < X \leq b] = F_X(b) - F_X(a)$	28
2.4	Exemplo de uma fdc de uma v.a. discreta.	31
2.5	Gráfico da fdc de v.a. contínua X	31
2.6	Gráfico de $F'_X(x)$	32
2.7	Um exemplo de v.a. mista.	33
2.8	A função densidade de probabilidade especifica a probabilidade de intervalos de largura infinitesimal.	34
2.9	A probabilidade de um intervalo $[a, b]$ é a área sob a fdp naquele intervalo.	34
2.10	Fdc's condicional e incondicional de X	51
2.11	a) Dependência entre X e Y , b) $f_X(x)$, e c) $f_Y(y)$	57
2.12	Uma transformação da v.a. X e um exemplo das fdp's correspondentes de X e Y	59
2.13	Função de uma v.a. com duas raízes.	59
2.14	Uma transformação quadrática da v.a. X	60
2.15	Função densidade de probabilidade de Rayleigh.	64
3.1	Função densidade de probabilidade gaussiana com média m e variância σ^2	73
3.2	$Y = g(X)$	74
4.1	Método da transformada para gerar uma variável aleatória com fdc $F_X(x)$	92
4.2	Gerando uma variável aleatória com distribuição de Bernoulli.	93
4.3	Gerando uma variável aleatória com distribuição Binomial.	94
4.4	Método da rejeição para gerar uma variável aleatória com fdp $f_X(x)$	95
4.5	Método da rejeição para gerar uma variável aleatória com distribuição gama ($0 < \alpha < 1$).	96
5.1	Região de integração para a obtenção de $F_W(w)$	103
5.2	Região de integração para a obtenção de $F_W(w)$	104

5.3	O número de caras em 50 arremessos de uma moeda ideal: 400 repetições experimentais versus a fmp binomial.	117
6.1	Região A (sombreada).	127
6.2	Um limitante superior exponencial usado para obter a probabilidade de cauda (limitante de Chernoff).	128
7.1	Convergência de uma sequência de médias amostrais obtidas a partir de uma sequência de v.a.'s com distribuição Gaussiana de média 4 e variância 10.	137
8.1	Um processo estocástico que representa a temperatura de uma cidade. .	141
8.2	Um conjunto com um número finito de funções amostra.	141
8.3	Funções amostra de quatro tipos de processos estocásticos: $X(t)$ é um processo contínuo no tempo e na amplitude; $X(n)$, obtido a partir da amostragem de $X(t)$ em instantes de tempo inteiros n , é discreto no tempo e contínuo na amplitude; $Y(t)$ é obtida a partir da quantização de $X(t)$ nos instantes de amostragem, e é um processo discreto na amplitude e contínuo no tempo; finalmente, $Y(n)$, um processo discreto no tempo e na amplitude, é obtido a partir da amostragem de $Y(t)$	143
8.4	Função amostra de um processo de contagem	148
8.5	Função amostra de um processo telegráfico aleatório	152
8.6	Forma de onda do pulso $p(t)$	155
8.7	Erro de detecção devido ao ruído.	156
8.8	Processo estocástico comprimido no tempo.	157
8.9	Fdp dos processos x e y	157
8.10	Funções de autocorrelação para os processos $X(t)$ e $Y(t)$	158
8.11	Processo aleatório $X(t) = A \cos(\omega_c t + \theta)$	162
8.12	Classificação dos processos estocásticos.	165
9.1	Filtro passa faixa ideal $H(f)$ com frequência central f_0 e largura de banda B Hz.	179
9.2	A correlação cruzada entre a entrada e a saída de um filtro linear invariante no tempo é a convolução da resposta a impulso do filtro com a função de autocorrelação da entrada. A densidade espectral cruzada entre a entrada e a saída é o produto do espectro densidade de potência da entrada com a função de transferência do filtro. A densidade espectral de potência da saída é o produto da densidade espectral cruzada da entrada e da saída e o complexo conjugado da função de transferência do filtro. .	188
10.1	Transições para o estado j	211
10.2	Balanço global de fluxo de probabilidade.	215
10.3	Diagrama de transição de estados para o sistema M/M/1.	216
10.4	Diagrama de taxa de transição para um processo de nascimento e morte geral.	217
10.5	Instantes de recorrência para o estado i	224

Capítulo 1

Probabilidade

1.1 Introdução.

Em muitos problemas físicos de interesse, existe um elemento de incerteza, ou aleatoriedade. Independente de quanto possamos conhecer da história passada de um dado fenômeno, somos essencialmente incapacitados de prever seu comportamento futuro de forma precisa. Ex. cara ou coroa.

Foi observado que nestes casos certas médias tendem a um valor constante à medida em que o número de observações cresce. (No exemplo da cara e coroa, quais seriam estas médias?) Desde que as médias geralmente exibem tal regularidade, e são portanto razoavelmente previsíveis, parece ser desejável desenvolver um estudo sobre o cálculo destas médias. Este é o domínio da teoria matemática da probabilidade e estatística. O propósito desta é descrever e prever tais médias em termos de probabilidades de eventos.

Algumas definições importantes:

Definição 1.1. *Experimento aleatório:* um experimento é chamado aleatório se seu resultado não pode ser predito precisamente porque as condições em que é realizado não podem ser predeterminadas com precisão suficiente.

Exemplo: arremesso de dados ou moedas.

Definição 1.2. *Resultados:* são os resultados particulares da execução de um experimento.

Exemplo: cara, coroa.

Definição 1.3. Eventos: são conjuntos de resultados que atendem a algumas especificações.

Exemplo: no caso de jogar dados, o evento “número ímpar em um arremesso” pode resultar de qualquer um de 3 resultados 1,3,5. Desta forma este é um evento de 3 resultados. Portanto, eventos são agrupamentos de resultados em classes.

1.2 Teoria de Conjuntos.

Definição 1.4. Espaço amostral: o espaço amostral S é definido como uma coleção de todos os resultados possíveis de um experimento aleatório. Cada resultado é um elemento ou amostra deste espaço e pode ser convenientemente representado por um ponto no espaço amostral.

Exemplo 1.1. No caso do dado, o espaço amostral consiste de 6 elementos $\zeta_1, \zeta_2, \zeta_3, \zeta_4, \zeta_5, \zeta_6$, onde os ζ_i representam o resultado “ i pontos”. O evento, por outro lado, é um subconjunto de S .

O evento “número ímpar em um arremesso”, denotado por A_o , é um subconjunto de S (ou um conjunto com os elementos ζ_1, ζ_3 e ζ_5). Similarmente, o evento “número par em um arremesso”, denotado por A_e é outro subconjunto de S (ou um conjunto com os elementos ζ_2, ζ_4 e ζ_6). O evento “número menor ou igual a 4 em uma jogada”, denotado por B é formado pelos elementos $\zeta_1, \zeta_2, \zeta_3$ e ζ_4 .

Na Figura 1.1 abaixo, tem-se uma representação gráfica destes eventos em um diagrama de Venn.

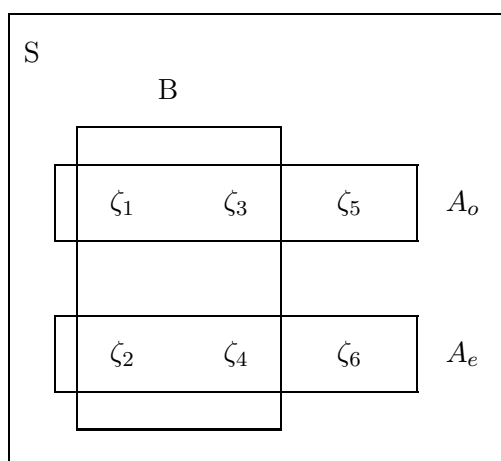


Figura 1.1: Espaço amostral para o arremesso de um dado.

Definição 1.5. O **complemento** de um evento A , denotado por A^c , é o evento que contém todos os pontos de S que não estão em A .

No exemplo acima, quais são os eventos complementares de A_o , A_e , e B ?.

Definição 1.6. Um evento que não contém elementos é chamado de **evento nulo**, e é denotado por ϕ .

Observe que o evento nulo é o complemento do espaço amostral S : $\phi = S^c$.

Importante!!!!

Definição 1.7. A **união** de eventos A e B , denotada por $A \cup B$, é aquele que contém todos os pontos em A e B .

Verifique no exemplo anterior quais são os eventos $A_o \cup A_e$, $A_o \cup B$, $A_e \cup B$). Observe que $A \cup B = B \cup A$.

Importante!!!!

Definição 1.8. A **interseção** dos eventos A e B , denotada por $A \cap B$ ou simplesmente AB , é o evento que contém pontos comuns a A e a B . Este evento também é conhecido como **evento conjunto** AB .

Observe que $AB = BA$. ~~Na figura 1.2 abaixo, tem-se estes conceitos mostrados graficamente em diagramas de Venn.~~

Importante!!!!

Definição 1.9. Se os eventos A e B são tais que $AB = \phi$ então A e B são ditos eventos **disjuntos** ou **mutuamente exclusivos**.

Isto quer dizer que A e B não podem ocorrer simultaneamente. (A e A^c são mutuamente exclusivos).

Estes conceitos são mostrados de forma gráfica na Figura 1.2.

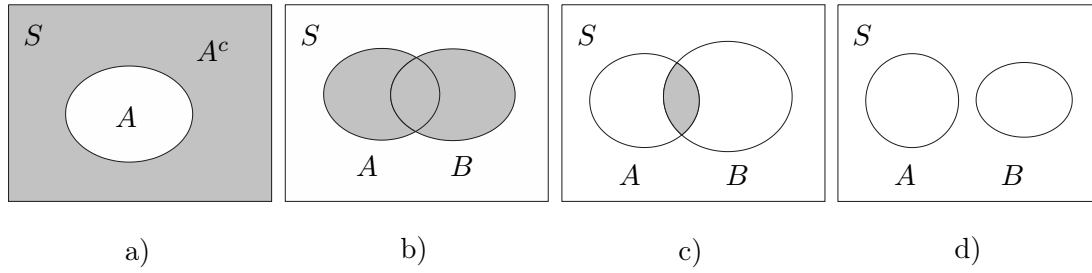


Figura 1.2: Representação do a) complemento, b) união, c) interseção de eventos, e d) eventos disjuntos.

1.2.1 Lei de De Morgan.

Teorema 1.1. *Se A e B são eventos em um espaço amostral então:*

$$\overline{A + B} = \overline{A} \overline{B} \quad (1.1)$$

Equivalentemente, podemos escrever:

$$\overline{AB} = \overline{A} + \overline{B} \quad (1.2)$$

Demonstração. A lei de De Morgan pode ser facilmente demonstrada por meio de diagramas de Venn:

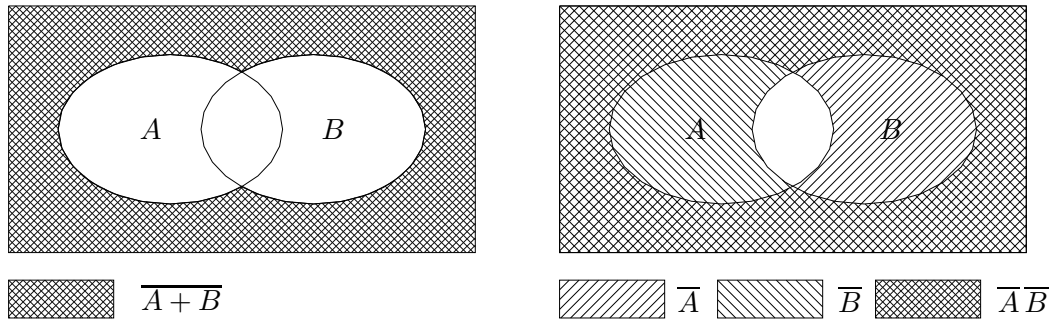


Figura 1.3: Demonstração da lei de De Morgan.

□

Observação

A aplicação repetida da equação (1.1) leva ao seguinte: se em uma identidade de conjuntos substituímos todos os conjuntos pelos seus complementos, todas as uniões por intersecções, e todas as intersecções por uniões, a identidade é preservada.

Exemplo 1.2. *Seja a identidade*

$$A(B + C) = AB + AC \quad (1.3)$$

Usando (1.1) segue que

$$\overline{A(B + C)} = \overline{A} + \overline{B + C} = \overline{A} + \overline{B} \overline{C}$$

Similarmente

$$\overline{AB + AC} = \overline{AB} \overline{AC} = (\overline{A} + \overline{B})(\overline{A} + \overline{C})$$

e desde que os dois lados de (1.3) são iguais, seus complementos também o são. Portanto

$$\overline{A} + \overline{B + C} = (\overline{A} + \overline{B})(\overline{A} + \overline{C}) \quad (1.4)$$

Estas identidades podem ser facilmente conferidas por meio de diagramas de Venn.

1.2.2 Princípio da Dualidade.

Sabemos que $\overline{S} = \phi$ e $\overline{\phi} = S$. Além disso, se em uma identidade como (1.3) todas as barras forem removidas, a identidade é preservada. Isto leva à seguinte versão da lei de De Morgan:

Proposição 1.1. *Se em uma identidade de conjuntos substituirmos todas as uniões por intersecções, todas as intersecções por uniões, e os conjuntos S e ϕ pelos conjuntos ϕ e S respectivamente, a identidade é preservada.*

Aplicando o teorema acima às identidades

$$A(B + C) = AB + AC \quad S = A + S$$

obtemos as identidades

$$A + BC = (A + B)(A + C) \quad \phi = \phi A$$

1.3 Definições de Probabilidade.

1.3.1 Frequência Relativa.

Embora o resultado de um experimento aleatório seja imprevisível, existe uma regularidade estatística sobre este, e a definição por frequência relativa baseia-se nesta regularidade.

Importante!!!!**Definição 1.10.** A probabilidade $P(A)$ de um evento A é dada pelo limite

$$P(A) = \lim_{n \rightarrow \infty} \frac{n_A}{n} \quad (1.5)$$

onde n_A é o número de ocorrências de A e n é o número de tentativas.**Observações importantes**

1. Segue da definição que $0 \leq P(A) \leq 1$.
2. Se A e B são dois eventos mutuamente exclusivos

$$P(A + B) = P(A) + P(B) = \lim_{n \rightarrow \infty} \frac{n_A + n_B}{n} \quad (1.6)$$

3. Se A_1, A_2, \dots, A_N não forem mutuamente exclusivos então:

$$P(A_1 + A_2 + \dots + A_N) < P(A_1) + P(A_2) + \dots + P(A_N) \quad (1.7)$$

1.3.2 Axiomática.*Genericamente:*

$$P(A_1 + A_2 + \dots + A_N) \leq P(A_1) + P(A_2) + \dots + P(A_N)$$

Definição 1.11. A aproximação axiomática para a probabilidade é baseada nos três postulados seguintes e nada mais:

1. A probabilidade $P(A)$ de um evento A é um número positivo associado a este evento

$$P(A) \geq 0 \quad (1.8)$$

2. A probabilidade do espaço amostral é igual a 1

Importante!!!!

$$P(S) = 1 \quad (1.9)$$

3. Se os eventos A e B são mutuamente exclusivos, então

$$P(A + B) = P(A) + P(B) \quad (1.10)$$

$$P(A+B) = P(A) + P(B) - P(AB)$$

Propriedades:

$$P(\phi) = 0$$

$$P(A^c) = 1 - P(A)$$

$$P(A + B) = P(A) + P(B) - P(AB) \leq P(A) + P(B)$$

evento impossível

 A^c complemento de A

probabilidade da união

Exemplo 1.3. Determinar a probabilidade de obtenção de uma cara e duas coroas em 3 arremessos de uma moeda ideal.

Solução. Neste caso, os resultados possíveis são:

- | | | |
|---------------|---------------|---|
| 1) ca, ca, ca | 5) co, ca, ca | $P_A = \binom{N}{K} (1-p)^{N-K} p^K$ $P_{1co2ca} = \binom{3}{2} (1-p_{ca})^{3-2} p_{ca}^2$ $= \binom{3}{1} (1-p_{co})^{3-1} p_{co}^1$ |
| 2) ca, ca, co | 6) co, ca, co | |
| 3) ca, co, ca | 7) co, co, ca | |
| 4) ca, co, co | 8) co, co, co | |

São possíveis 8 resultados mutuamente exclusivos $\Rightarrow P(A_i) = 1/8$

$\therefore P(1ca, 2co) = P(A_4) + P(A_6) + P(A_7) = 3/8$.

1.3.3 Clássica.

Importante!!!!

Definição 1.12. A probabilidade $P(A)$ de um evento A é determinada a priori sem experimentação real, e é dada pela expressão

$$P(A) = \frac{n_A}{n} \quad (1.11)$$

onde:

n : número de resultados possíveis,

n_A : número de resultados favoráveis ao evento A .

Versão melhorada da definição clássica

Definição 1.13. A probabilidade de um evento é igual à razão entre seus resultados favoráveis e o número total de resultados, desde que todos os resultados sejam equiprováveis.

Exemplo 1.4. Arremesso de um dado $P(\text{ímpar}) = 3/6 = 1/2$.

Fim da Aula 1

1.4 Cálculo de probabilidades usando métodos de contagem.

Em muitos experimentos com espaços amostrais finitos, os resultados podem ser assumidos como sendo equiprováveis. A probabilidade de um evento é então a razão entre o número de resultados no evento de interesse e o número total de resultados no espaço amostral. O cálculo das probabilidades se reduz a contar o número de resultados de um evento.

Este é o caso de um dado ou de um baralho

Suponha que um teste de múltipla escolha tem k questões e para a questão i o estudante precisa selecionar uma entre n_i respostas possíveis. Qual é o número total de modos de responder a todo o teste?

A resposta à questão i pode ser vista como a especificação da i -ésima componente de uma k -upla, de modo que a questão acima é equivalente a: quantas k -uplas ordenadas distintas (x_1, \dots, x_k) são possíveis se x_i é um elemento de um conjunto com n_i elementos distintos?

O número de k -uplas ordenadas distintas (x_1, \dots, x_k) com componentes x_i , de um conjunto com n_i elementos distintos é dado por

$$\text{número de } k\text{-uplas ordenadas distintas} = n_1 n_2 \dots n_k \quad (1.12)$$

Muitos problemas de contagem podem ser colocados como problemas de amostragem onde selecionamos bolas em urnas ou objetos em populações. Iremos agora usar a Equação 1.12 para desenvolver fórmulas combinatoriais para vários tipos de amostragem.

1.4.1 Amostragem com reposição e ordenação.

Suponha que escolhemos k objetos de um conjunto A que tem n objetos distintos, com reposição. Iremos nos referir ao conjunto A como a *população*. O experimento produz uma k -upla ordenada (x_1, \dots, x_k) , onde $x_i \in A, i = 1, 2, \dots, k$. A Equação 1.12, com $n_1 = n_2 = \dots = n_k = n$ implica que

$$\text{número de } k\text{-uplas ordenadas distintas} = n^k \quad (1.13)$$

Seja um LED que pode assumir 4 cores diferentes. Quantas sequencias diferentes de cores pode existir com 3 LEDs?

Importante!!!!

Exemplo 1.5. Uma urna contém cinco bolas numeradas. Suponha que selecionamos duas bolas da urna com reposição. Quantos pares ordenados distintos são possíveis? Qual é a probabilidade de retirar duas vezes a mesma bola?

Solução. A Equação 1.13 diz que o número de pares ordenados é $5^2 = 25$. Na Tabela abaixo temos os pares possíveis. Cinco dos resultados possíveis são de bolas com o mesmo número. Se supomos que todos os resultados possíveis são equiprováveis, então a probabilidade de retirar a mesma bola duas vezes é $5/25 = 0,2$.

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)

Note que a ordem da sequencia importa, ou seja, a diferenca na ordem faz com que a sequencia seja diferente!

1.4.2 Amostragem sem reposição e com ordenação.

O problema agora consiste em escolher k objetos em sucessão, sem reposição, de uma população A de n objetos distintos. Claramente, $k \leq n$. O número de resultados possíveis na primeira retirada é $n_1 = n$, e o número de resultados possíveis na segunda

retirada é $n_2 = n - 1$, e assim por diante, até $n_k = n - (k - 1)$ na retirada final. Desta forma, a Equação 1.12 fornece

$$\text{número de } k\text{-uplas ordenadas distintas} = n(n-1)\dots(n-k+1) \quad (1.14)$$

Exemplo 1.6. *Uma urna contém cinco bolas numeradas. Suponha que selecionamos duas bolas da urna em sucessão, e sem reposição. Quantos pares ordenados distintos são possíveis? Qual é a probabilidade de que a primeira bola tenha um número maior que a segunda?*

Solução. A Equação 1.14 mostra que o número de pares ordenados possíveis é $5(4) = 20$. Estes são mostrados na Tabela abaixo. Dez pares ordenados nesta tabela têm o primeiro número maior que o segundo, de forma que a probabilidade deste evento é $10/20 = 0,5$.

	(1,2)	(1,3)	(1,4)	(1,5)
(2,1)		(2,3)	(2,4)	(2,5)
(3,1)	(3,2)		(3,4)	(3,5)
(4,1)	(4,2)	(4,3)		(4,5)
(5,1)	(5,2)	(5,3)	(5,4)	

1.4.3 Permutação de n objetos distintos.

Considere uma amostragem sem reposição com $k = n$. Isto equivale a retirar objetos de uma urna até que ela esteja vazia. Então o número de seqüências possíveis de n objetos distintos é igual ao número de n -uplas da amostragem sem reposição com $k = n$. Da Equação 1.14, temos

$$\text{número de permutações de } n \text{ objetos} = n(n-1)\dots(2)(1) = n! \quad (1.15)$$

Para n grande, a fórmula de Stirling é bastante útil:

$$n! \approx \sqrt{2\pi} n^{n+\frac{1}{2}} e^{-n} \quad (1.16)$$

Exemplo 1.7. *Encontre o número de permutações de três objetos distintos 1,2,3.*

Solução. A Equação 1.15 fornece $3! = 6$. As seis permutações são

123 312 231 132 213 321

1.4.4 Amostragem sem reposição e sem ordenação.

Suponha que pegamos k objetos de um conjunto de n objetos distintos sem reposição e armazenamos o resultado sem nos importarmos com a ordem. Chamamos o subconjunto resultante de k objetos selecionados de uma “combinação de tamanho k ”.

Da Equação 1.15, existem $k!$ seqüências nas quais os objetos ~~selecionados~~ podem ter sido selecionados. Então se C_k^n denota o número de combinações de tamanho k de um conjunto de tamanho n , então $C_k^n k!$ é o número total de amostras ordenadas distintas de k objetos, a qual é dada pela Equação 1.14. Então

$$C_k^n k! = n(n-1) \dots (n-k+1) = \frac{n!}{(n-k)!} \quad (1.17)$$

e o número de combinações diferentes de tamanho k de um conjunto de tamanho n , $k \leq n$, é

$$C_k^n = \frac{n(n-1) \dots (n-k+1)}{k!} = \frac{n!}{k!(n-k)!} \equiv \binom{n}{k} \quad (1.18)$$

A expressão $\binom{n}{k}$ é chamada de *coeficiente binomial*.

Importante!!!!

Note que escolher k objetos de um conjunto de n é equivalente a escolher os $(n-k)$ objetos que não foram selecionados. Segue então que

$$\binom{n}{k} = \binom{n}{n-k} \quad (1.19)$$

Exemplo 1.8. Encontre o número de modos de selecionar dois objetos de $A = \{1, 2, 3, 4, 5\}$ sem se importar com a ordem.

Solução. A Equação 1.18 fornece

$$\binom{5}{2} = \frac{5!}{2!3!} = 10 \quad (1.20)$$

Abaixo temos a listagem destes 10 pares.

$$\begin{array}{cccc} (1,2) & (1,3) & (1,4) & (1,5) \\ & (2,3) & (2,4) & (2,5) \\ & & (3,4) & (3,5) \end{array}$$

Note que (i,k) e (k,i) são o mesmo evento! (4,5)

Exemplo 1.9. Encontre o número de permutações distintas de k bolas brancas e $(n-k)$ bolas pretas.

Solução. Este problema é equivalente ao seguinte problema de amostragem: coloque n etiquetas numeradas de 1 a n em uma urna, onde cada etiqueta representa uma posição no arranjo das bolas; pegue uma combinação de k etiquetas e coloque as k bolas brancas nas posições correspondentes.

Cada combinação de tamanho k leva a um arranjo diferente (permutação) de k bolas brancas e $(n-k)$ bolas pretas.

Então o número de permutações distintas de k bolas brancas e $(n-k)$ bolas pretas é C_k^n .

Este exemplo mostra que a amostragem sem reposição e sem ordenação é equivalente a particionar o conjunto de n objetos distintos em dois conjuntos: B , contendo os k itens que foram retirados da urna, e B^c , contendo os $(n - k)$ deixados na urna.

Suponha que particionemos um conjunto de n objetos distintos em F subconjuntos B_1, B_2, \dots, B_F , onde ao subconjunto B_j são associados k_j elementos e $k_1 + k_2 + \dots + k_F = n$.

Neste caso, o número de combinações distintas é dado por

$$\frac{n!}{k_1! k_2! \dots k_F!} \quad (1.21)$$

A Equação 1.21 é chamada de coeficiente multinomial. O coeficiente binomial é o caso $F = 2$ dos coeficientes multinomiais.

1.4.5 Amostragem com reposição e sem ordenação.

Suponha que tomemos k objetos de um conjunto de n objetos distintos com reposição e armazenamos os resultados sem nos importarmos com a ordem. Isto pode ser feito preenchendo-se um formulário com n colunas, uma para cada objeto distinto. Cada vez que um objeto é selecionado, um “x” é colocado na coluna correspondente. Por exemplo, se selecionamos 5 objetos de 4 objetos distintos, um formulário destes poderia ter a seguinte forma:

Objeto 1	Objeto 2	Objeto 3	Objeto 4
xx		x	xx

Note que este formulário pode ser resumido pela sequência $xx / / x / xx$, onde o símbolo “/” é usado para separar as entradas para as diferentes colunas. Desta forma os $(n - 1)$ /’s indicam as linhas entre as colunas, e onde nada aparece entre /’s consecutivos se o objeto correspondente não foi selecionado.

Cada arranjo diferente de 5 x’s e 3 /’s leva a um formulário distinto.

Se identificarmos os x’s com bolas brancas e os /’s com bolas pretas, então este problema foi considerado no Exemplo 1.9, e o número de arranjos diferentes é dado por $\binom{8}{3}$.

No caso geral o formulário irá envolver k x’s e $(n - 1)$ /’s. Então o número de modos diferentes de escolher k objetos de um conjunto de n objetos distintos com reposição e sem ordenação é dado por

$$\binom{n - 1 + k}{k} = \binom{n - 1 + k}{n - 1} \quad (1.22)$$

1.5 Probabilidade Conjunta.

Ao invés de lidar com um experimento, consideremos agora dois experimentos e seus respectivos resultados. Por exemplo, os dois experimentos podem ser dois arremessos consecutivos de um único dado ou um único arremesso de dois dados. Em ambos os casos, o espaço amostral consiste de 36 duplas (i, j) , onde $i, j = 1, 2, 3, 4, 5, 6$. Se os dados são ideais, a cada ponto do espaço amostral é associada uma probabilidade $1/36$. Podemos agora considerar eventos conjuntos tais como $\{i \text{ é par, } j = 3\}$, e determinar

as probabilidades associadas a tais eventos a partir do conhecimento das probabilidades dos pontos amostrais.

Definição 1.14. Se os resultados possíveis de um experimento são $A_i, i = 1, 2, \dots, n$, e os resultados possíveis de um segundo experimento são $B_j, j = 1, 2, \dots, m$, então os resultados possíveis do experimento combinado são dados pelo conjunto $(A_i, B_j), i = 1, 2, \dots, n, j = 1, 2, \dots, m$. A cada resultado conjunto (A_i, B_j) associa-se uma probabilidade conjunta $P(A_i, B_j)$ que satisfaz a condição

$$0 \leq P(A_i, B_j) \leq 1 \quad (1.23)$$

Exemplo 1.10. Retirar duas cartas em sucessão (com ou sem reposição) de um baralho.

Solução. Vamos considerar os seguintes eventos

Evento A : retirar um ás na primeira tentativa

Evento B : retirar um ás na segunda tentativa

AB é o evento de retirar dois ases.

Calcule esta probabilidade, considerando com reposicao e sem reposicao.

1.5.1 Probabilidades Marginais.

Assumindo que os resultados $B_j, j = 1, 2, \dots, m$ são mutuamente exclusivos, segue que

$$\text{Importante!!!!} \sum_{j=1}^m P(A_i, B_j) = P(A_i) \quad (1.24)$$

Demonstre isso usando o diagrama de Venn!

Similarmente, se os resultados $A_i, i = 1, 2, \dots, n$ são mutuamente exclusivos então

$$\sum_{i=1}^n P(A_i, B_j) = P(B_j) \quad (1.25)$$

Além disso, se todos os resultados dos dois experimentos são mutuamente exclusivos temos

$$\sum_{i=1}^n \sum_{j=1}^m P(A_i, B_j) = 1 \quad (1.26)$$

$P(A_i)$ e $P(B_j)$ são chamadas de **probabilidades marginais**. É fácil ver que a generalização do tratamento acima para mais de dois experimentos é direta.

1.6 Probabilidade Condicional.

Considere um experimento combinado no qual um evento conjunto ocorre com probabilidade $P(A, B)$. Suponha que o evento B ocorreu e queremos determinar a probabilidade de ocorrência do evento A . Esta probabilidade é chamada de probabilidade condicional e denota-se por $P(A|B)$. Probabilidade de A ocorrer dado que B ocorreu.

Exemplo: encontre a probabilidade de receber o bit "0" dado que o bit "1" foi transmitido no BSC abaixo

Exemplo 1.11. No exemplo anterior, se a primeira carta não é recolocada no baralho, fica evidente que a retirada de um ás na segunda tentativa é influenciada pelo resultado da primeira.

1.6.1 Regra de Bayes.

Importante!!!!

Teorema 1.2. Teorema de Bayes. Seja um experimento fornecendo dois resultados A e B . Então,

$$P(AB) = P(A|B)P(B) = P(B|A)P(A) \quad (1.27)$$

Demonstração. Sejam as seguintes grandezas:

- N : número total de tentativas;
- n_B : número de resultados favoráveis ao evento B ;
- n_{AB} : número de resultados favoráveis ao evento A dentro das n_B tentativas.

Estas grandezas são mostradas em um diagrama de Venn na Figura 1.4.

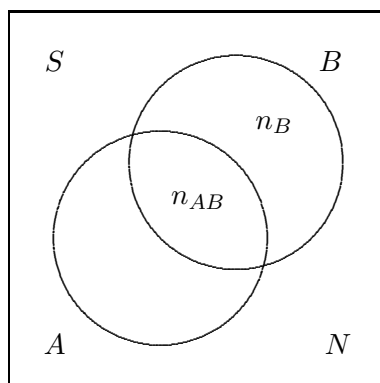


Figura 1.4: Espaço amostral para a derivação da regra de Bayes.

Observe que n_{AB} é o número de tentativas que são favoráveis ao evento AB . Assim

$$P(AB) = \lim_{N \rightarrow \infty} \frac{n_{AB}}{N} = \lim_{N \rightarrow \infty} \left(\frac{n_B}{N} \right) \left(\frac{n_{AB}}{n_B} \right) \quad (1.28)$$

Do diagrama acima, podemos extrair as seguintes expressões:

$$P(B) = \lim_{N \rightarrow \infty} \frac{n_B}{N} \quad (1.29)$$

$$P(A|B) = \lim_{N \rightarrow \infty} \frac{n_{AB}}{n_B} \quad (1.30)$$

Aqui estamos implicitamente usando o fato que $n_B \rightarrow \infty$ à medida que $N \rightarrow \infty$. Observe que n_{AB} é o número de tentativas favoráveis ao evento A dentro das n_B tentativas favoráveis ao evento B . Isto representa a probabilidade condicional $P(A|B)$. Combinando (1.28), (1.29) e (1.30), temos:

$$P(A|B) = \frac{P(AB)}{P(B)} \quad (1.31)$$

E por um desenvolvimento similar, pode-se demonstrar que

$$P(B|A) = \frac{P(AB)}{P(A)} \quad (1.32)$$

Combinando 1.31 e 1.32, chegamos à Regra de Bayes

$$P(AB) = P(A|B)P(B) = P(B|A)P(A) \quad (1.33)$$

- 1) Qual é a probabilidade de A ocorrer dado que B ocorreu, sendo A e B eventos disjuntos?
 2) Qual é a condição para que $P(A/B)=P(B/A)$? □

Extensão para mais eventos

Uma generalização bastante útil da regra de Bayes é a seguinte: considere os eventos $A_i, i = 1, 2, \dots, n$, mutuamente exclusivos tais que

$$\bigcup_{i=1}^n A_i = S \quad (1.34)$$

e um evento arbitrário B com probabilidade não nula. Então, a regra de Bayes pode ser reescrita como

$$P(A_i|B) = \frac{P(A_i, B)}{P(B)} = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^n P(B|A_j)P(A_j)} \quad (1.35)$$

No canal BSC apresentado acima, qual é a probabilidade de se ter o bit "1" na saída, assumindo que a entrada é equiprovável?

1.7 Eventos independentes.

Importante!!!!

Definição 1.15. Um evento A é dito independente de B se

$$P(A|B) = P(A) \quad (1.36)$$

Importante!!!!

Teorema 1.3. Se A e B são eventos independentes então

$$P(AB) = P(A)P(B) \quad (1.37)$$

Dois eventos são ditos independentes quando a ocorrência de um não afeta a probabilidade de ocorrência do outro.

Demonstração. Pela Regra de Bayes, temos que

$$P(AB) = P(A|B)P(B)$$

Mas como A e B são independentes,

$$P(A|B) = P(A)$$

Substituindo este resultado na Equação acima, chegamos a

$$P(AB) = P(A)P(B)$$

□

Exemplo 1.12. *Suponha que uma moeda é jogada três vezes. Se assumimos que as jogadas são independentes e a probabilidade de caras é p , encontre a probabilidade dos eventos nenhuma coroa, uma coroa, duas coroas e três coroas.*

Solução. A probabilidade para as sequências de caras e coroas é dada por

$$\begin{aligned} P[\{CCC\}] &= P[\{C\}]P[\{C\}]P[\{C\}] &= p^3 \\ P[\{CCK\}] &= P[\{C\}]P[\{C\}]P[\{K\}] &= p^2(1-p) \\ P[\{CKC\}] &= P[\{C\}]P[\{K\}]P[\{C\}] &= p^2(1-p) \\ P[\{KCC\}] &= P[\{K\}]P[\{C\}]P[\{C\}] &= p^2(1-p) \\ P[\{KKC\}] &= P[\{K\}]P[\{K\}]P[\{C\}] &= p(1-p)^2 \\ P[\{KCK\}] &= P[\{K\}]P[\{C\}]P[\{K\}] &= p(1-p)^2 \\ P[\{CKK\}] &= P[\{C\}]P[\{K\}]P[\{K\}] &= p(1-p)^2 \\ P[\{KKK\}] &= P[\{K\}]P[\{K\}]P[\{K\}] &= (1-p)^3 \end{aligned}$$

onde usamos o fato de que as jogadas são independentes. Seja k o número de caras em três tentativas. Então

$$\begin{aligned} P[k=0] &= P[KKK] = (1-p)^3 \\ P[k=1] &= P[KKC, KCK, CKK] = 3p(1-p)^2 \\ P[k=2] &= P[CCK, CKC, KCC] = 3p^2(1-p) \\ P[k=3] &= P[CCC] = p^3 \end{aligned}$$

Observações

A definição de independência estatística pode ser estendida a três ou mais eventos. Para que três eventos A_1 , A_2 e A_3 sejam estatisticamente independentes, precisam satisfazer as seguintes condições

$$\begin{aligned} P(A_1, A_2) &= P(A_1)P(A_2) \\ P(A_1, A_3) &= P(A_1)P(A_3) \\ P(A_2, A_3) &= P(A_2)P(A_3) \\ P(A_1, A_2, A_3) &= P(A_1)P(A_2)P(A_3) \end{aligned} \tag{1.38}$$

Para o caso geral, os eventos $A_i, i = 1, 2, \dots, n$ são estatisticamente independentes se as probabilidades dos eventos conjuntos tomados 2, 3, \dots, n eventos de cada vez possam ser fatoradas no produto das probabilidades dos eventos individuais.

1.8 Experimentos sequenciais e diagramas em árvore

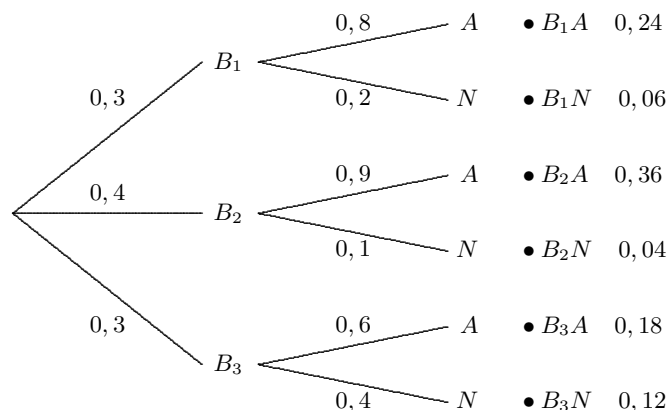
Muitos experimentos consistem de uma sequência de *subexperimentos*. O procedimento adotado para cada subexperimento pode depender dos resultados dos subexperimentos anteriores. Podemos usar um diagrama em árvore para representar a natureza sequencial dos subexperimentos. Seguir o procedimento e anotar as observações do experimento é equivalente a seguir a sequência de ramificações da raiz para as folhas da árvore. Cada folha corresponde a um resultado do experimento.

É natural modelar probabilidades condicionais em termos de experimentos sequenciais e ilustrá-las através de diagramas em árvores. Na raiz da árvore, a probabilidade de um evento particular é descrito pelo nosso conhecimento a priori. Se os resultados possíveis do primeiro resultado são descritos pelos eventos B_1, \dots, B_m , então $\{B_1, \dots, B_m\}$ é um espaço de eventos. A partir da raiz, desenhemos ramos para cada evento B_i . Seguir um ramo a partir da raiz corresponde a observar os resultados do primeiro subexperimento. Associamos a cada ramo as probabilidades a priori $P[B_1], \dots, P[B_m]$. Para cada evento B_i , temos probabilidades condicionais descrevendo o resultado do segundo subexperimento. Então para cada um dos ramos do primeiro conjunto, desenhemos um novo ramo e associamos a ele esta probabilidade condicional. Se seguirmos uma sequência de ramos da raiz a uma determinada folha, especificamos o resultado de um dado subexperimento. Desta forma, as folhas representam os resultados do experimento completo. A probabilidade de cada resultado é o produto das probabilidades dos ramos entre a raiz da árvore e a folha que correspondente ao resultado. Em geral, associamos às folhas os resultados e as probabilidades correspondentes.

Isto é uma descrição complicada para um procedimento extremamente simples, como veremos nos exemplos a seguir.

Exemplo 1.13. *Uma companhia tem três máquinas B_1 , B_2 e B_3 que fabricam resistores de $1k\Omega$. Observou-se que 80% dos resistores produzidos por B_1 têm tolerância de 50Ω do valor nominal. A máquina B_2 produz 90% dos resistores com tolerância de 50Ω do valor nominal. A porcentagem para a máquina B_3 é de 60%. A cada hora, a máquina B_1 produz 3000 resistores, B_2 produz 4000 resistores, e B_3 produz 3000 resistores. Todos os resistores são misturados em um recipiente comum e empacotados para envio. Desenhe um diagrama em árvore para este experimento. Qual a probabilidade de escolher um resistor da máquina B_2 com tolerância maior que 50Ω ?*

Solução. Seja A o evento “o resistor selecionado é aceitável” (tem tolerância de 50Ω), e N o complemento de A : “o resistor selecionado não é aceitável”. O procedimento de testar um resistor pode ser decomposto em dois passos: primeiro, identificamos qual máquina (B_1 , B_2 ou B_3) produziu o resistor; depois, verificamos se o resistor é aceitável ou não. Estes dois passos correspondem à seguinte árvore:

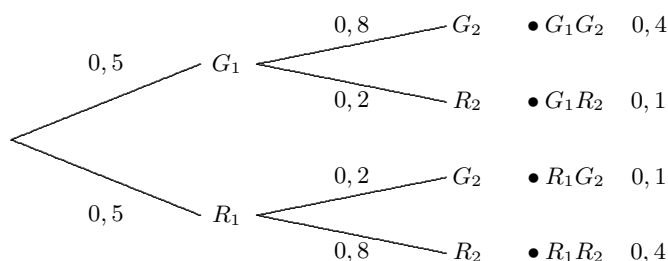


Para usar a árvore para encontrar a probabilidade do evento B_2N , um resistor não aceitável da máquina B_2 , começamos da esquerda e verificamos que a probabilidade de alcançar B_2 é $P[B_2] = 0,4$. Andamos então para a direita em direção ao nó B_2N e multiplicamos $P[B_2]$ por $P[N|B_2] = 0,1$, e obtemos $P[B_2N] = 0,4 \times 0,1 = 0,04$.

Podemos observar neste exemplo uma propriedade geral de todos os diagramas em árvore que representam experimentos sequenciais: a soma das probabilidades dos ramos que deixam um determinado nó é sempre 1. Isto é uma consequência da lei da probabilidade total e da propriedade da probabilidade condicional, vistas anteriormente.

Exemplo 1.14. Suponha que os engenheiros de tráfego tenham coordenado a temporização de dois faróis para encorajar uma sequência de faróis verdes. Em particular, a temporização foi projetada de modo que, com probabilidade 0,8 um motorista encontre o segundo farol com a mesma cor do primeiro. Assumindo que o primeiro farol seja verde ou vermelho com a mesma probabilidade, qual é a probabilidade $P[G_2]$ de que o segundo farol seja verde? Calcule $P[G_1|R_2]$, a probabilidade condicional de que o primeiro farol seja verde, dado que o segundo é vermelho.

Solução. Neste caso, a árvore que descreve o problema é:



A probabilidade do segundo farol ser verde é

$$P[G_2] = P[G_1G_2] + P[R_1G_2] = 0,4 + 0,1 = 0,5$$

O evento W de ter que esperar por pelo menos um farol é dado por

$$W = \{R_1 G_2 \cup G_1 R_2 \cup R_1 R_2\}$$

e desta forma, a probabilidade de esperar por pelo menos um farol é dada por

$$P[W] = P[R_1 G_2] + P[G_1 R_2] + P[R_1 R_2] = 0,1 + 0,1 + 0,4 = 0,6$$

Para encontrar $P[G_1|R_2]$, precisamos de $P[R_2]$. Notando que $R_2 = \{G_1 R_2 \cup R_1 R_2\}$, temos:

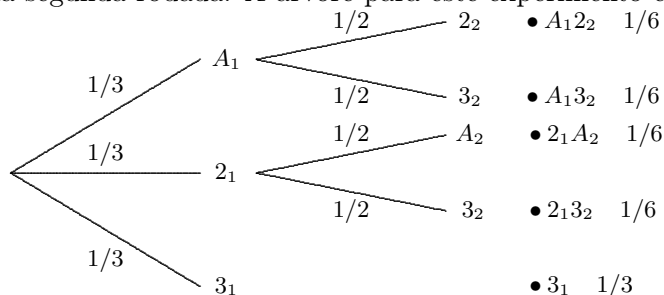
$$P[R_2] = P[G_1 R_2] + P[R_1 R_2] = 0,1 + 0,4 = 0,5$$

Desde que $P[G_1 R_2] = 0,1$, a probabilidade condicional de observar o primeiro farol verde dado que o segundo é vermelho é dada por:

$$P[G_1|R_2] = \frac{P[G_1 R_2]}{P[R_2]} = \frac{0,1}{0,5} = 0,2 \quad (1.39)$$

Exemplo 1.15. Considere o jogo do Três. Você embaralha um baralho de três cartas: às, 2 e 3. Se o às vale um ponto, você retira cartas do baralho até que a soma seja 3 ou mais. Você ganha se o total for 3. Calcule $P[W]$, a probabilidade de vencer o jogo.

Solução. Seja C_i o evento “ C é a i -ésima carta retirada. Por exemplo, 3_2 é o evento de tirar um 3 na segunda rodada. A árvore para este experimento é então:

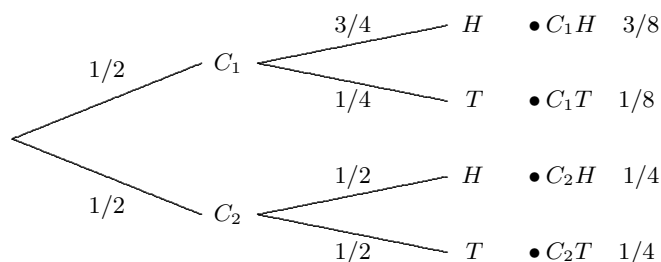


Você vence se $A_1 2_2$, $2_1 A_2$ ou 3_1 ocorrerem. Desta forma, a probabilidade de vencer é dada por

$$P[W] = P[A_1 2_2] + P[2_1 A_2] + P[3_1] = \frac{1}{3} \frac{1}{2} + \frac{1}{3} \frac{1}{2} + \frac{1}{3} = \frac{2}{3}$$

Exemplo 1.16. Suponha que você tem duas moedas, uma viciada e outra não, mas você não sabe qual é qual. A moeda 1 é viciada (tem probabilidade $3/4$ de dar cara). Suponha que você pegue uma moeda de forma aleatória e a arremesse. Seja C_i o evento “a moeda i foi selecionada”. Vamos denotar por H (cara) e T (coroa) os possíveis resultados de um arremesso. Dado que o resultado de um arremesso é uma cara, calcule $P[C_1|H]$, a probabilidade de você ter selecionado a moeda viciada. Dado que o resultado é uma coroa, calcule $P[C_1|T]$, a probabilidade de ter selecionado a moeda viciada.

Solução. Primeiro, contruímos a árvore que descreve o problema:



Para encontrar as probabilidades condicionais, temos:

$$P[C_1|H] = \frac{P[C_1H]}{P[H]} = \frac{P[C_1H]}{P[C_1H] + P[C_2H]} = \frac{3/8}{3/8 + 1/4} = \frac{3}{5}$$

Similarmente,

$$P[C_1|T] = \frac{P[C_1T]}{P[T]} = \frac{P[C_1T]}{P[C_1T] + P[C_2T]} = \frac{1/8}{1/8 + 1/4} = \frac{1}{3}$$

Como esperávamos, é mais provável termos selecionado a moeda 1 quando o primeiro arremesso resultou em cara, e é mais provável termos selecionado a moeda 2 quando o primeiro arremesso resultou em coroa.

Fim da Aula 2

1.9 Exercícios

1. Quatro moedas ideais são arremessadas simultaneamente.
 - (a) Quantos resultados são possíveis?
 - (b) Associe probabilidades adequadas para a obtenção de quatro coroas, uma cara, duas caras, três caras e quatro caras neste experimento.

Resp:

- (a) 16
 - (b) $P[4 \text{ coroas}] = 1/16$
 $P[1 \text{ cara}] = 1/4$
 $P[2 \text{ caras}] = 3/8$
 $P[3 \text{ caras}] = 1/4$
 $P[4 \text{ caras}] = 1/16$
2. Três dados não viciados são jogados. Calcule as probabilidades dos eventos de se obter uma soma de 8, 9 e 10 pontos.

Resp: $P[8] = 21/216$ $P[9] = 25/216$ $P[10] = 27/216$

3. Uma certa cidade tem 8 faróis aleatoriamente localizados, quatro dos quais ficam verdes por meio minuto na direção leste-oeste e meio minuto na direção norte-sul, três permanecem verdes por $1/4$ de minuto na direção leste-oeste e $3/4$ de minuto na direção norte-sul, e o último permanece verde $3/4$ de minuto na direção leste-oeste e $1/4$ de minuto na direção norte-sul.

Assuma que todos os faróis são independentes, isto é, não existe nenhum tipo de sincronização entre eles.

Um automóvel está viajando de forma aleatória através da cidade. Encontre a probabilidade de o automóvel encontrar um sinal verde na direção leste-oeste. Faça o mesmo para a direção norte-sul.

Qual é a probabilidade de um automóvel viajando aleatoriamente pela cidade encontre um sinal verde?

Resp:

$$P[\text{verde na direção L-O}] = 7/16$$

$$P[\text{verde na direção N-S}] = 9/16$$

$$P[\text{verde}] = 1/2$$

4. Uma urna contém 3 bolas vermelhas e 2 brancas. Duas bolas são retiradas em sucessão, a primeira bola sendo recolocada antes da retirada da segunda.

(a) Quantos resultados são possíveis?

(b) Associe probabilidades a cada um destes resultados.

Resp:

(a) 4

(b) $P[1a.V, 2a.V] = 9/25$

$$P[1a.V, 2a.B] = 6/25$$

$$P[1a.B, 2a.V] = 6/25$$

$$P[1a.B, 2a.B] = 4/25$$

5. Repita o problema anterior se a primeira bola não for recolocada antes da segunda retirada.

(a) 4

(b) $P[1a.V, 2a.V] = 3/10$

$$P[1a.V, 2a.B] = 3/10$$

$$P[1a.B, 2a.V] = 3/10$$

$$P[1a.B, 2a.B] = 1/10$$

6. No problema anterior, se sabemos que a primeira retirada foi de uma bola branca, qual é a probabilidade de a segunda retirada ser também de uma bola branca ?

Resp: $1/4$

7. No problema 5), se sabemos que a segunda bola é vermelha, qual a probabilidade de a primeira também ter sido vermelha? Qual a probabilidade da primeira bola ter sido branca?

Resp: a) $1/2$ b) $1/2$

8. Uma urna contém 3 bolas vermelhas, 5 bolas brancas e 8 bolas pretas. Outra urna contém 6 bolas vermelhas, 7 bolas brancas e 4 bolas pretas. Uma bola é retirada de cada urna. Encontre a probabilidade de obter duas bolas da mesma cor.

Resp: $85/272$

9. A caixa I contém 3 bolas vermelhas e 5 bolas brancas, e a caixa II, 4 vermelhas e 2 brancas. Extraí-se ao acaso uma bola da primeira caixa e coloca-se na segunda, sem observar a cor. Extraí-se então uma bola da segunda caixa. Qual a probabilidade da mesma ser branca?

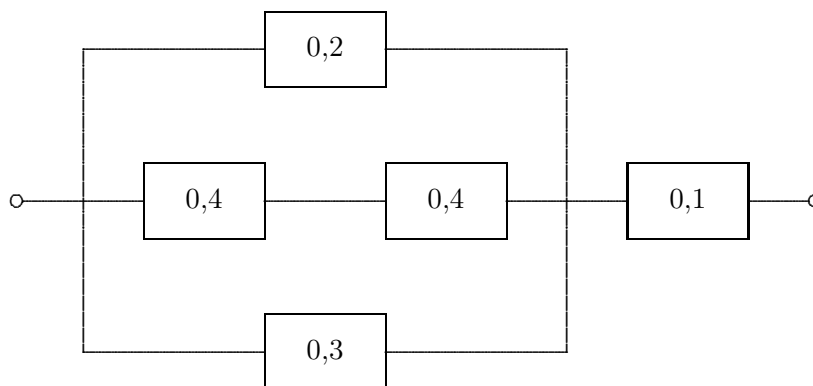
Resp: $21/56$

10. Em certo colégio, 25 % dos estudantes foram reprovados em matemática, 15 % em química e 10 % em matemática e química ao mesmo tempo. Um estudante é selecionado aleatoriamente.

- Se ele foi reprovado em química, qual é a probabilidade de ele ter sido reprovado em matemática?
- Se ele foi reprovado em matemática, qual é a probabilidade de ele ter sido reprovado em química?
- Qual é a probabilidade de ele ter sido reprovado em matemática ou química?

Resp: a) $2/3$ b) $2/5$ c) $0,30$

11. A rede comutada mostrada na figura abaixo opera se e somente se existe pelo menos um caminho fechado de comutadores entre a entrada e a saída. Assumindo que os comutadores falhem de forma independente e que a probabilidade de falha de cada comutador são aquelas dadas na figura, calcule a probabilidade de esta rede funcionar.



Resp: $0,865$

12. Uma urna contém duas bolas pretas e três bolas brancas. Duas bolas são selecionadas aleatoriamente da urna sem reposição, e a sequência de cores é anotada. Encontre a probabilidade de retirar duas bolas pretas.

Resp: $1/10$

13. Lança-se uma moeda viciada de modo que $P[\text{cara}] = 2/3$ e $P[\text{coroa}] = 1/3$. Se aparecer cara, então seleciona-se aleatoriamente um número dentre os de 1 a 9; se aparecer coroa, seleciona-se aleatoriamente um número dentre os de 1 a 5. Encontre a probabilidade p de um número par ser selecionado.

Resp: $p = 58/135$

14. Dois dígitos são selecionados aleatoriamente de 1 a 9, sem reposição. Se a soma é par, encontre a probabilidade p de ambos os números serem ímpares.

Resp: $p = 5/8$

15. Telefones celulares realizam *handoffs* à medida em que se movem de uma célula para outra. Suponha que durante uma chamada, os telefones realizam zero *handoffs* (H_0), um *handoff* (H_1), ou dois *handoffs* (H_2). Adicionalmente, cada chamada pode ser longa (L) ou breve (B).

Sabendo que $P[L, H_0] = 0.1$, $P[B, H_1] = 0.1$, $P[H_2] = 0.3$, $P[B] = 0.6$ e $P[H_0] = 0.5$, calcule:

- (a) A probabilidade de não ocorrer nenhum *handoff* durante uma chamada.
- (b) A probabilidade de uma chamada ser breve.
- (c) A probabilidade de uma chamada ser longa ou existirem pelo menos dois *handoffs*.

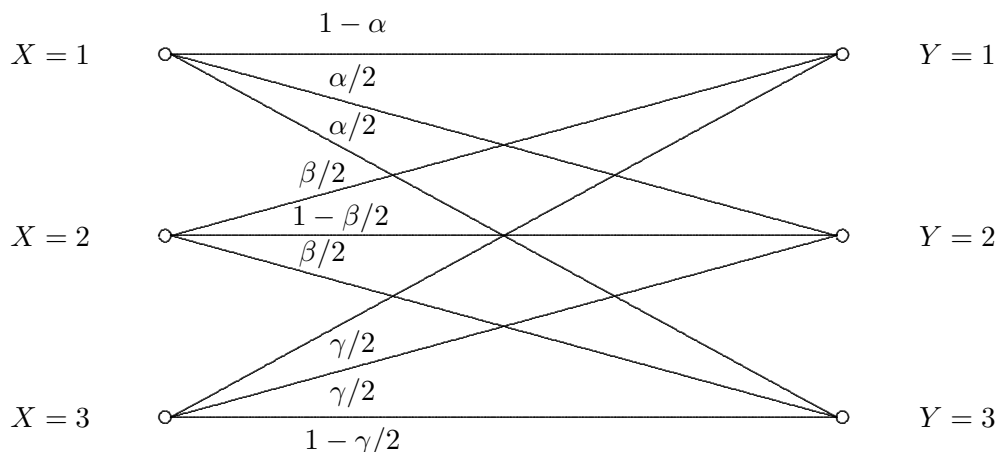
Resp: a) 0.5 b) 0.6 c) 0.5

16. Três máquinas A , B e C produzem 50%, 30% e 20% respectivamente, do total de peças de uma fábrica. As porcentagens de produção de peças defeituosas destas máquinas são 3%, 4% e 5%, respectivamente.

- (a) Se uma peça é selecionada aleatoriamente, ache a probabilidade dela ser defeituosa.
- (b) Suponha que uma peça, selecionada aleatoriamente, seja considerada defeituosa. Encontre a probabilidade dela ter sido produzida pela máquina A .

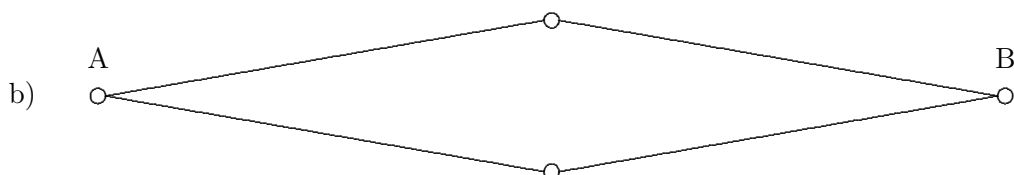
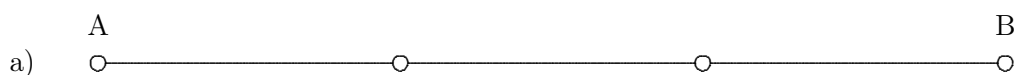
Resp: a) 0,037 b) $15/37$

17. No sistema de comunicação ternário mostrado na figura abaixo, um 3 é enviado três vezes mais frequentemente que um 1, e um 2 é enviado duas vezes mais frequentemente que um 1. Um 1 é observado. Qual a probabilidade de um 1 ter sido enviado?



Resp: $\frac{1 - \alpha}{1 - \alpha + \beta + 1,5\gamma}$

18. Para a comunicação entre os terminais A e B são necessários enlaces que são representados nas figuras abaixo por arcos. Sendo p a probabilidade de que um enlace esteja ocupado, determine a probabilidade de que não exista caminho livre para comunicação em cada uma das seguintes configurações:



Resp: a) $3(1-p)p^2 + 3(1-p)^2p + p^3$ b) $(2p(1-p) + p^2)^2$

19. Durante a recepção de mensagens codificadas, consistindo de pulsos de formas A e B, estabeleceu-se que de cada 10 combinações equiprováveis, três são do tipo AAB, cinco são do tipo AB, e duas são do tipo ABB. Qual é a probabilidade de que um pulso escolhido aleatoriamente seja da forma A?

Resp: $31/60$

20. Sabendo que a probabilidade de um homem viver mais de dez anos é $1/4$, a probabilidade de sua esposa viver mais de dez anos é $1/3$, encontre a probabilidade dos seguintes eventos

- (a) ambos estarem vivos depois de dez anos,
- (b) ao menos um estar vivo depois de dez anos,
- (c) nenhum deles estar vivo depois de dez anos,
- (d) somente a esposa estar viva depois de dez anos.

Dica: considere os eventos

A : o homem está vivo daqui a 10 anos.

B : sua esposa está viva daqui a 10 anos.

Resp: a) $1/12$ b) $1/2$ c) $1/2$ d) $1/4$

21. A urna 1 contém 5 bolas brancas e 7 bolas pretas. A urna 2 contém 3 bolas brancas e 12 bolas pretas. Uma moeda ideal é arremessada. Se o resultado é cara, então seleciona-se uma bola da urna 1, enquanto que se o resultado é coroa, seleciona-se uma bola da urna 2. Suponha que uma bola branca tenha sido selecionada. Qual a probabilidade do resultado do arremesso da moeda ter sido coroa?

Resp: $P[co|B] = 12/37$

22. Sejam os seguintes eventos:

- A : uma família tem crianças de ambos os sexos.
- B : uma família tem no máximo um menino.

- (a) Mostre que A e B são independentes, se uma família tem 3 crianças.
(b) Mostre que A e B são dependentes, se uma família tem 2 crianças.

Capítulo 2

Variáveis Aleatórias

2.1 Definição.

O resultado de um experimento aleatório pode ser um número real (como no caso do arremesso de dados) ou pode ser não numérico, mas descrito por palavras (por exemplo “cara” e “coroa”).

Entretanto estamos geralmente interessados não no resultado, mas em alguma medida ou atributo numérico deste. Por exemplo, se jogamos uma moeda n vezes, podemos estar interessados no número total de caras e não na ordem específica na qual ocorreram as caras e as coroas.

Assim, podemos definir uma função que associa um valor numérico ao resultado do experimento aleatório. Desde que os resultados são aleatórios, os resultados das medidas também o serão. Desta forma faz sentido falar em probabilidades dos valores numéricos resultantes.

O conceito de variável aleatória formaliza esta noção:

Definição 2.1. Uma variável aleatória X é uma função que associa um número real $X(\zeta)$ a cada resultado ζ no espaço amostral de um experimento aleatório.

Lembre-se que uma função é simplesmente uma regra que associa um valor numérico a cada elemento de um conjunto, como mostrado graficamente na Figura 2.1.

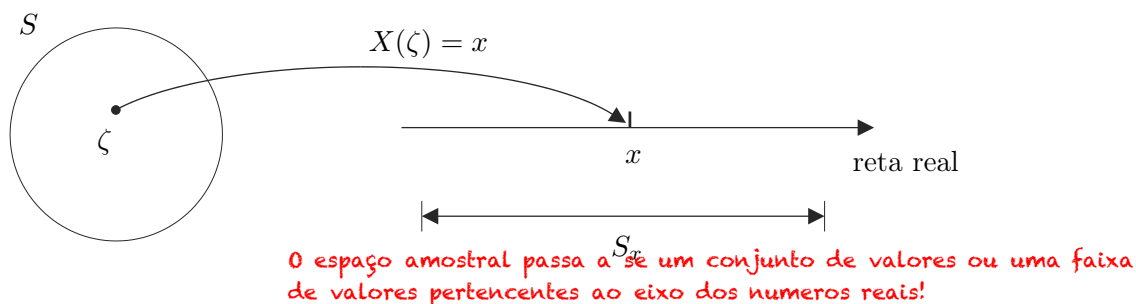


Figura 2.1: Uma v.a. associa um número $x = X(\zeta)$ a cada resultado ζ no espaço amostral S de um experimento aleatório.

A especificação de uma medida de um experimento aleatório define uma função no espaço amostral, e portanto uma v.a. O espaço amostral S é o domínio da v.a., e o conjunto S_X de todos os valores tomados por X é a faixa da v.a. Então S_X é um subconjunto do conjunto de todos os números reais.

Podemos ver $X(\cdot)$ como uma função que mapeia os pontos amostrais $\zeta_1, \zeta_2, \dots, \zeta_m$ em números reais x_1, x_2, \dots, x_n . Assim, X é uma **variável aleatória** que assume valores x_1, x_2, \dots, x_n . Observe que m não é necessariamente igual a n . Mais de um ponto amostral pode ser mapeado em um mesmo valor de x .

Exemplo 2.1. Especifique o espaço amostral de um experimento que consiste em jogar uma moeda 3 vezes.

Solução. O espaço amostral para este experimento é

$$S = \{CCC, CCK, CKC, CKK, KCC, KCK, KKC, KKK\},$$

onde C corresponde a “cara” e K corresponde a “coroa”.

Seja X o número de caras em três jogadas da moeda. X associa a cada resultado ζ em S um número do conjunto $S_X = 0, 1, 2, 3$. A tabela abaixo lista os oito resultados de S e os valores de X correspondentes.

ζ	CCC	CCK	CKC	KCC	CKK	KCK	KKC	KKK
$X(\zeta)$	3	2	2	2	1	1	1	0

X é então uma v.a. que toma valores no conjunto $S_X = 0, 1, 2, 3$.

A função ou regra que associa valores a cada resultado é fixa ou determinística, como, por exemplo, na regra “número de caras em 3 jogadas de uma moeda”. A aleatoriedade nos valores observados deve-se à aleatoriedade dos argumentos da função X , ou seja os resultados ζ_i do experimento.

Em outras palavras, a aleatoriedade dos valores observados de X é induzida pelo experimento aleatório, e devemos portanto ser capazes de calcular as probabilidades dos valores observados em termos das probabilidades dos resultados do experimento.

Exemplo 2.2. O evento $\{X = k\} = \{k \text{ caras em 3 jogadas de uma moeda}\}$ ocorre quando o resultado do experimento contém k caras. Calcule as probabilidades dos eventos $\{X = k\}$, $k = 0, 1, 2, 3$.

Solução. A probabilidade do evento $\{X = k\}$ é dada pela soma das probabilidades dos resultados correspondentes ou eventos elementares. Seja p a probabilidades de caras e $(1 - p)$ a probabilidade de coroas. Desta forma, temos

$$\begin{aligned} p_0 &= P[X = 0] = P[\{KKK\}] = (1 - p)^3 \\ p_1 &= P[X = 1] = P[\{CKK\}]P[\{KCK\}]P[\{KKC\}] = 3(1 - p)^2p \\ p_2 &= P[X = 2] = P[\{CCK\}]P[\{CKC\}]P[\{KCC\}] = 3(1 - p)p^2 \\ p_3 &= P[X = 3] = P[\{CCC\}] = p^3 \end{aligned}$$

Note que as jogadas das moedas são independentes, ou seja $P(AB) = P(A)P(B)$

O exemplo acima ilustra a seguinte técnica geral para encontrar as probabilidades de eventos envolvendo a v.a. X : seja S_X o conjunto de valores que podem ser assumidos por X , e B algum subconjunto de S_X .

S_X pode ser visto como um novo espaço amostral, e B como um evento neste espaço.

Seja A o conjunto de resultados ζ em S que levam a valores $X(\zeta)$ em B , como mostrado na Figura 2.2, isto é

$$A = \{\zeta : X(\zeta) \text{ em } B\}$$

então o evento B em S_X ocorre sempre que o evento A em S ocorre. Desta forma, a probabilidade do evento B é dada por

$$P[A] = P[B] = P[\zeta : X(\zeta) \text{ em } B]$$

Referimo-nos aos eventos A e B como eventos equivalentes.

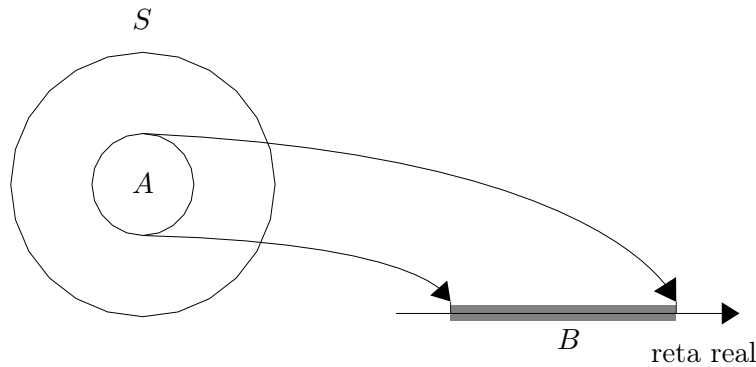


Figura 2.2: Eventos equivalentes.

2.2 Função distribuição cumulativa.

Importante!!!!

Definição 2.2. A função distribuição cumulativa (fdc) de uma v.a. X é definida como a probabilidade do evento $\{X \leq x\}$:

$$F_X(x) \triangleq P[X \leq x], \quad -\infty < x < \infty \quad (2.1)$$

isto é, a probabilidade da v.a. X tomar um valor no intervalo $(-\infty, x]$.

Em termos do espaço amostral, a fdc é a probabilidade do evento $\{\zeta : X(\zeta) \leq x\}$. O evento $\{X \leq x\}$ e sua probabilidade variam à medida que x varia; em outras palavras, $F_X(x)$ é uma função da variável x .

A fdc é simplesmente uma maneira conveniente de especificar a probabilidade de todos os intervalos semi-infinitos da reta real, e seus complementos, uniões e interseções.

Propriedades

Os axiomas de probabilidade e seus corolários implicam que a fdc tem as seguintes propriedades:

1. $0 \leq F_X(x) \leq 1$
2. $\lim_{x \rightarrow \infty} F_X(x) = 1$
3. $\lim_{x \rightarrow -\infty} F_X(x) = 0$
4. $F_X(x)$ é uma função não decrescente de x , isto é, se $a < b$, então $F_X(a) \leq F_X(b)$.
5. A probabilidade de eventos que correspondem a intervalos da forma $(a < X \leq b)$ podem ser expressas em termos da fdc

$$P[a < X \leq b] = F_X(b) - F_X(a) \quad \text{Importante!!!!} \quad (2.2)$$

Demonstração.

$$P[a < X \leq b] = P[X \leq b] - P[X \leq a] = F_X(b) - F_X(a)$$

Isto pode ser facilmente visto na Figura abaixo

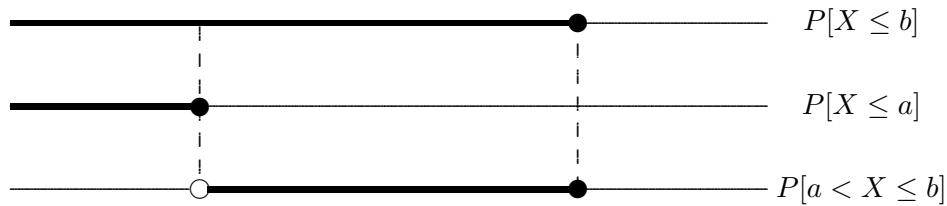


Figura 2.3: $P[a < X \leq b] = F_X(b) - F_X(a)$

□

6. A probabilidade que uma v.a. X toma em um ponto específico, digamos b , é dada pela magnitude do salto da fdc no ponto b . Segue que se a fdc é contínua em um ponto b , então o evento tem probabilidade zero.

Demonstração. Desejamos calcular $P[X = b]$. Seja $a = b - \varepsilon$, $\varepsilon > 0$. Usando (2.2), podemos escrever

$$P[a < X \leq b] = P[b - \varepsilon < X \leq b] = F_X(b) - F_X(b - \varepsilon) \quad (2.3)$$

À medida que $\varepsilon \rightarrow 0$, o lado esquerdo de (2.3) aproxima $P[X = b]$, e então

$$P[X = b] = F_X(b) - F_X(b^-) \quad (2.4)$$

□

7. Seja o intervalo $\{a \leq X \leq b\} = \{X = a\} \cup \{a < X \leq b\}$. Então

$$\begin{aligned} P[a \leq X \leq b] &= P[X = a] + P[a < X \leq b] \\ &= F_X(a) - F_X(a^-) + F_X(b) - F_X(a) \\ &= F_X(b) - F_X(a^-) \end{aligned} \quad (2.5)$$

8. Se a fdc é contínua nos limites de um intervalo, então os limites têm probabilidade zero, e portanto podem ser incluídos ou excluídos do intervalo sem afetar a probabilidade. Em outras palavras, se a fdc é contínua nos pontos $x = a$ e $x = b$, então

$$P[a < X < b] = P[a \leq X < b] = P[a < X \leq b] = P[a \leq X \leq b] \quad (2.6)$$

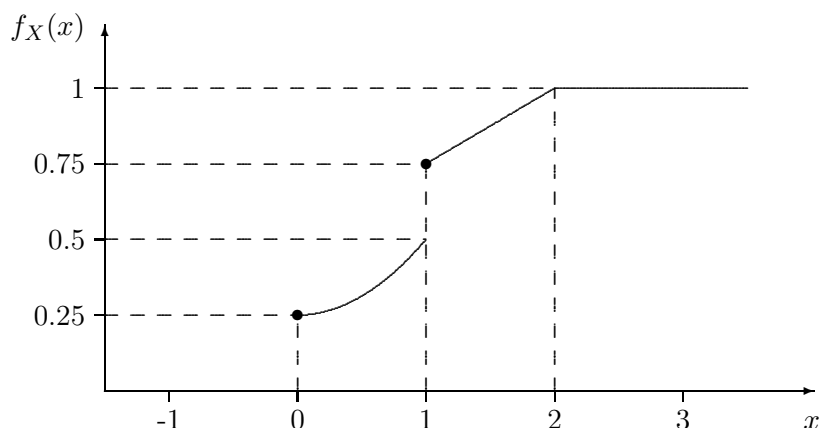
Exemplo 2.3. A fdc de uma variável aleatória X é dada por

$$F_X(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{4}(x^2 + 1) & 0 \leq x < 1 \\ \frac{1}{4}x + \frac{1}{2} & 1 \leq x < 2 \\ 1 & x \geq 2 \end{cases}$$

Encontre a probabilidade dos eventos:

$$a) \{X < 1\} \quad b) \{X = 1\} \quad c) \{X = 0\} \quad d) \{|x - 1| > 1/2\} \quad e) \{x \leq 0\}$$

Solução. A primeira coisa a fazer é analisar como esta função se comporta: das equações acima, podemos ver que esta é uma função nula para $x < 0$; para $0 \leq x < 1$ assume a forma de uma parábola, e no intervalo $1 \leq x < 2$ o de uma reta; finalmente, assume um valor constante igual a 1 para $x > 2$. Abaixo temos um gráfico desta função.



A partir da análise do gráfico, fica fácil resolver o problema:

a) A probabilidade do evento $\{X < 1\}$ é dado pelo valor da fdc no ponto imediatamente anterior a $X = 1$. Portanto, $P[X < 1] = 1/2$.

- b) A probabilidade do evento $\{X = 1\}$ é dada pelo valor do salto da fdc em $X = 1$. Portanto, $P[X = 1] = 1/4$.
- c) Pelas mesmas razões do item b), $P[X = 0] = 1/4$.
- d) O evento $\{|x - 1| > 1/2\}$ pode ser visto como um círculo de raio $1/2$ com centro em $X = 1$.
Desta forma, $P[|x - 1| > 1/2] = 1 - P[1/2 < X \leq 3/2] = 1 - [F_X(3/2) - F_X(1/2)] = 7/16$
- e) $P[X \leq 0] = F_X(0) = 1/4$

2.3 Tipos de Variáveis Aleatórias

2.3.1 Discretas

Variáveis aleatórias discretas tomam valores de um conjunto finito $S_X = \{x_0, x_1, \dots, x_n\}$. Aparecem geralmente em aplicações que envolvem contagem, de modo que geralmente temos $S_X = \{0, 1, \dots\}$.

Importante!!!!

Definição 2.3. A *função massa de probabilidade (fmp)* de X é o conjunto de probabilidades $p_X(x_k) = P[X = x_k]$ dos elementos em S_X .

Alguns livros consideram a fmp como uma fdp.

Definição 2.4. A fdc de uma v.a. discreta pode ser escrita como uma soma ponderada de funções degrau unitário

$$F_X(x) = \sum_k p_X(x_k) u(x - x_k) \quad (2.7)$$

onde $p_X(x_k) = P[X = x_k]$ fornece a magnitude dos saltos na fdc.

Exemplo 2.4. Seja a v.a. X definida como “número de caras em três arremessos de uma moeda ideal”. Determine a fdc de X .

Solução. Do Exemplo 2.1 sabemos que X toma apenas os valores 0, 1, 2 e 3. Do Exemplo 2.2, se fizermos $p = 0.5$ as probabilidades para cada um destes resultados são $1/8$, $3/8$, $3/8$ e $1/8$, respectivamente, de modo que $F_X(x)$ é simplesmente a soma das probabilidades dos resultados de 0,1,2,3 que são menores ou iguais a x . A fdc resultante tem portanto descontinuidades nos pontos 0,1,2 e 3. A fdc de X definida desta maneira pode ser vista na Figura 2.4.

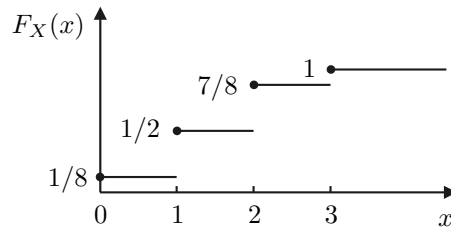


Figura 2.4: Exemplo de uma fdc de uma v.a. discreta.

2.3.2 Contínuas

São as v.a.'s cujas fdc's $F_X(x)$ são contínuas em todos os pontos e, as quais, adicionalmente, são suficientemente suaves de modo que podem ser escritas como uma integral de alguma função $f(x)$ não negativa.

$$F_X(x) = \int_{-\infty}^x f(t) dt \quad (2.8)$$

Para v.a.'s contínuas, a fdc é contínua em todos os pontos, de modo que a propriedade 6 implica que $P[X = x] = 0, \forall x$.

Exemplo 2.5. O tempo de transmissão X de mensagens em um sistema de comunicação obedece a lei de probabilidade exponencial com parâmetro λ , isto é $P[X > x] = e^{-\lambda x}, x > 0$. Encontre a fdc de X . Calcule $P[T < X \leq 2T], T = 1/\lambda$.

Solução. Por definição, a fdc de X é dada por $F_X(x) = P[X \leq x] = 1 - P[X > x]$. Desta forma, temos

$$F_X(x) = \begin{cases} 0, & x \leq 0 \\ 1 - e^{-\lambda x}, & x > 0 \end{cases}$$

Na Figura 2.5 tem-se um desenho da fdc de X .

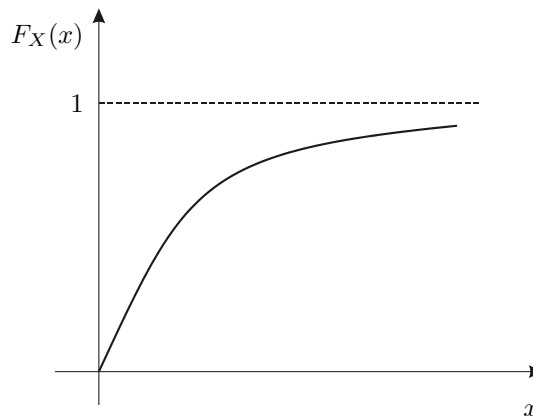


Figura 2.5: Gráfico da fdc de v.a. contínua X .

Da propriedade 5 temos que

$$P[T < X \leq 2T] = F_X(2T) - F_X(T) = 1 - e^{-2} - (1 - e^{-1}) = e^{-1} - e^{-2} \approx 0.233$$

Note que $F_X(x)$ é contínua para todo x . Note também que sua derivada existe para todos os pontos, exceto em $x = 0$.

Na Figura 2.6 tem-se o gráfico de $F'_X(x)$.

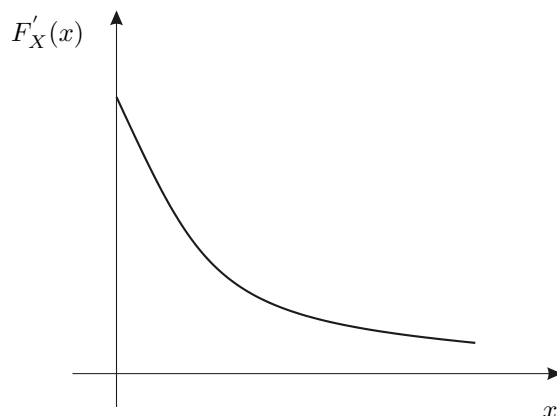


Figura 2.6: Gráfico de $F'_X(x)$.

2.3.3 Mistas

São v.a.'s cujas fdc's têm saltos em um número finito de pontos x_0, x_1, \dots, x_n mas que também aumentam de forma contínua por pelo menos um intervalo de valores de x . A fdc destas variáveis tem a forma

$$F_X(x) = pF_1(x) + (1 - p)F_2(x) \quad (2.9)$$

onde

- $0 < p < 1$
- $F_1(x)$ é a fdc de uma v.a. discreta.
- $F_2(x)$ é a fdc de uma v.a. contínua.

Exemplo 2.6. O tempo de espera X de um usuário em um sistema de filas é zero se ele encontra o sistema livre, e com um tempo de espera exponencialmente distribuído se encontra o sistema ocupado. As probabilidades de ele encontrar o sistema livre ou ocupado são p e $(1 - p)$, respectivamente. Encontre a fdc de X .

Solução.

$$F_X(x) = P[X \leq x] = P[X \leq x | \text{livre}]p + P[X \leq x | \text{ocupado}](1 - p)$$

Note que $P[X \leq x | \text{livre}] = 1$ quando $x \geq 0$ e 0 caso contrário. Desta forma

$$F_X(x) = \begin{cases} 0, & x < 0 \\ p + (1-p)(1 - e^{-\lambda x}), & x \geq 0 \end{cases}$$

O gráfico da fdc é mostrado na Figura 2.7. Note que $F_X(x)$ pode ser expressa como a soma de uma função degrau com amplitude p e uma função contínua de x .

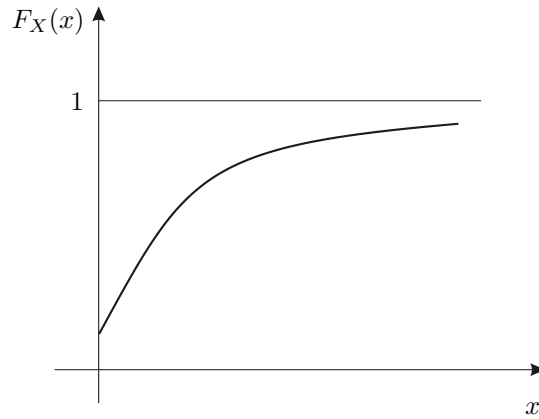


Figura 2.7: Um exemplo de v.a. mista.

2.4 Função Densidade de Probabilidade

2.4.1 Definição

Importante!!!!

Definição 2.5. A função densidade de probabilidade (fdp) de uma v.a. X , se existir, é definida como a derivada de $F_X(x)$:

$$f_X(x) = \frac{dF_X(x)}{dx} \quad (2.10)$$

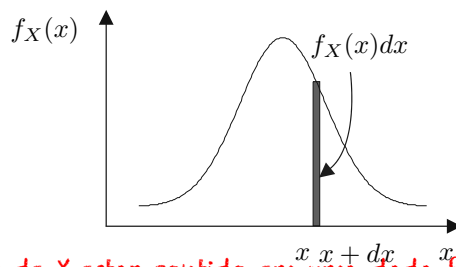
A fdp representa a “densidade” de probabilidade no ponto x no seguinte sentido: a probabilidade de que X esteja em um intervalo pequeno na vizinhança de x , isto é $\{x < X \leq x + h\}$, é

$$P[\{x < X \leq x + h\}] = F_X(x + h) - F_X(x) = \frac{F_X(x + h) - F_X(x)}{h} h \quad (2.11)$$

Se a fdc tem uma derivada em x , então à medida que $h \rightarrow 0$

$$P[\{x < X \leq x + h\}] \approx f_X(x)h \quad (2.12)$$

Então $f_X(x)$ representa a “densidade” de probabilidade no ponto x no sentido de que a probabilidade de que X esteja em um pequeno intervalo na vizinhança de x é aproximadamente $f_X(x)h$, conforme mostrado na Figura 2.8.



A probabilidade de X estar contido em uma dada faixa de valores é igual a área sobre a curva de $f_X(x)$, para a faixa de valores de interesse.

Figura 2.8: A função densidade de probabilidade especifica a probabilidade de intervalos de largura infinitesimal.

2.4.2 Propriedades **Importante!!!!**

1. A derivada da fdc, quando existir, é positiva desde que a fdc é uma função não decrescente de x , então

$$f_X(x) \geq 0 \quad (2.13)$$

2. Seja $f_X(x)$ uma função não negativa, a qual chamaremos de função densidade de probabilidade, e que especifica as probabilidades de eventos da forma “ X cai em um pequeno intervalo de largura dx ao redor do ponto x ”. As probabilidades de eventos envolvendo X são então expressas em termos da fdp adicionando probabilidades de intervalos de largura dx . À medida que as larguras dos intervalos se aproximam de zero, obtemos uma integral em termos da fdp. Por exemplo, a probabilidade de um intervalo $[a, b]$ é dada por

$$P[a \leq x \leq b] = \int_a^b f_X(x) dx \quad (2.14)$$

A probabilidade de um intervalo é portanto a área sob $f_X(x)$ naquele intervalo (ver Figura 2.9). A probabilidade de qualquer evento que consiste na união de intervalos disjuntos pode ser encontrada adicionando-se as integrais da fdp sobre cada um dos intervalos.

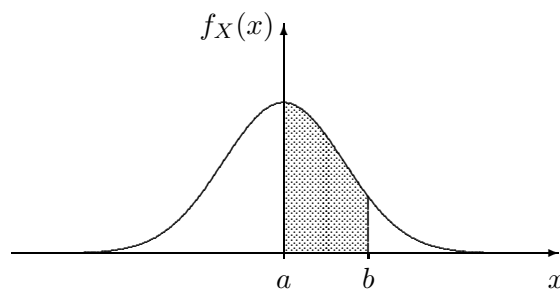


Figura 2.9: A probabilidade de um intervalo $[a, b]$ é a área sob a fdp naquele intervalo.

3. A fdc de X pode ser obtida integrando-se a fdp

$$F_X(x) = \int_{-\infty}^x f_X(t) dt \quad (2.15)$$

4. Fazendo $x \rightarrow +\infty$ na equação (2.15), obtemos a condição de normalização para as fdp's

$$\int_{-\infty}^{+\infty} f_X(t)dt = 1 \quad (2.16)$$

5. Uma fdp válida pode ser formada a partir de qualquer função $g(x)$ não negativa e contínua por partes que tenha uma integral finita

$$\int_{-\infty}^{+\infty} g(x)dx = c < \infty \quad (2.17)$$

Fazendo $f_X(x) = g(x)/c$ obtemos uma função que satisfaz a condição de normalização. Note que a fdp precisa ser definida para todos os valores reais de x ; se X não toma valores em alguma região da reta real, simplesmente fazemos $f_X(x) = 0$ na região.

2.4.3 Caso Discreto

A derivada da fdc não existe em pontos onde ela não é contínua. Então a noção de fdp definida na equação (2.10) não se aplica a v.a.'s discretas nos pontos em que a fdc não é contínua. Podemos generalizar a definição da função densidade de probabilidade notando a relação entre as funções degrau unitário e delta de Dirac.

Definição 2.6. A função *degrau unitário* $u(x)$ é definida como

$$u(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases} \quad (2.18)$$

Definição 2.7. A função *delta de Dirac* $\delta(x)$ é definida em termos da função degrau unitário pela seguinte equação

$$u(x) = \int_{-\infty}^{+\infty} \delta(t)dt \quad (2.19)$$

Na seção 2.3.1 vimos que a fdc de uma v.a. discreta pode ser escrita como uma soma ponderada de funções degrau unitário

$$F_X(x) = \sum_k p_X(x_k)u(x - x_k) \quad (2.20)$$

onde a função massa de probabilidade é dada por $p_X(x) = P[X = x]$.

Para generalizar a definição da fdp de modo que a Equação (2.15) valha também para v.a.'s discretas, podemos notar o seguinte: a integral de uma função delta localizada

em $x = b$, isto é $\delta(x - b)$, irá gerar uma função degrau que começa em $x = b$, isto é, $u(x - b)$.

Importante!!!!

Definição 2.8. Usando a equação (2.15), podemos definir a fdp de uma v.a. discreta como

$$p_X(x) = \sum_k P[X = x_k] \delta(x - x_k) \quad (2.21)$$

Desta forma, a definição generalizada da função densidade de probabilidade coloca uma função delta de peso $P[X = x_k]$ nos pontos x_k onde a fdc não é contínua.

2.5 Algumas variáveis aleatórias discretas importantes

As variáveis aleatórias discretas aparecem em geral em aplicações que envolvem contagens. As distribuições discretas mais comuns são:

2.5.1 Bernoulli

Usos mais frequentes

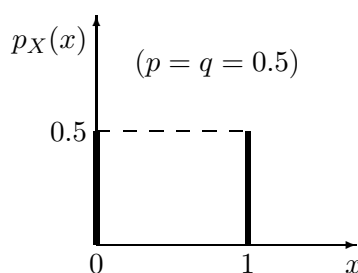
A distribuição de Bernoulli é o valor da função indicadora I_A para algum evento A ; $X = 1$ se A ocorre, e $X = 0$ caso contrário. Para estes testes, assume-se que a probabilidade de A ocorrer é p .

Domínio: $S_X = \{0, 1\}$

Função massa de probabilidade

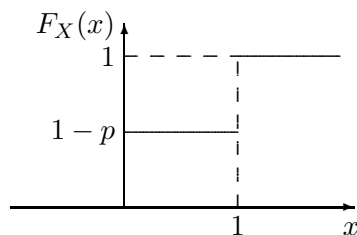
$$p_X(x) = \begin{cases} 1 - p = q, & X = 0 \\ p, & X = 1 \end{cases}$$

$$0 \leq p \leq 1$$



Função distribuição cumulativa

$$F_X(x) = \begin{cases} 0, & x < 0 \\ 1 - p, & 0 \leq x < 1 \\ 1, & x \geq 1 \end{cases}$$



2.5.2 Binomial

Usos mais frequentes

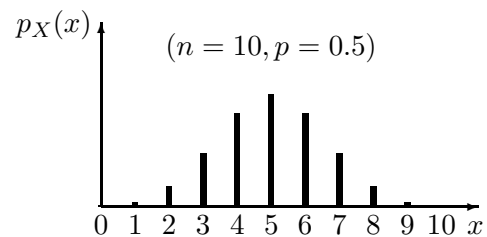
X é o número de sucessos em n experimentos de Bernoulli e, portanto, a soma de n variáveis aleatórias independentes e identicamente distribuídas, com distribuição de Bernoulli com probabilidade de sucesso igual a p .

Domínio: $S_X = \{0, 1, \dots, n\}$

Função massa de probabilidade

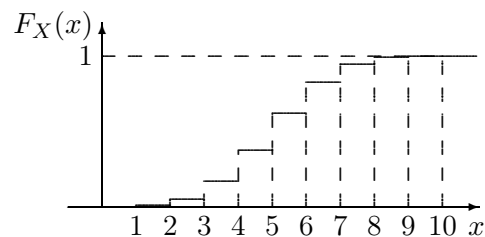
$$p_X(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

$$x = 0, 1, \dots, n$$



Função distribuição cumulativa

$$F_X(x) = \sum_k \binom{n}{k} p^k (1-p)^{n-k} u(x - x_k)$$



2.5.3 Poisson

Usos mais frequentes

Em muitas aplicações, estamos interessados em contar o número de ocorrências de um evento em um certo intervalo de tempo ou em uma determinada região do espaço. A variável aleatória de Poisson conta o número de eventos que ocorrem em uma unidade de tempo quando o tempo entre os eventos é exponencialmente distribuído com média $1/\alpha$.

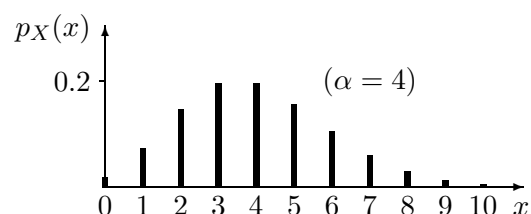
A distribuição de Poisson pode ser derivada da distribuição binomial fazendo-se $n \rightarrow \infty$ e $p \rightarrow 0$.

Domínio: $S_X = \{0, 1, 2, \dots\}$

Função massa de probabilidade

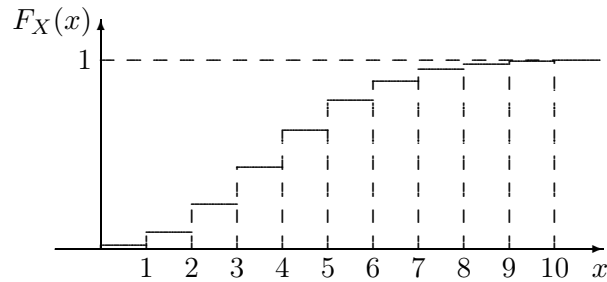
$$p_X(x) = \frac{\alpha^x}{x!} e^{-\alpha}$$

$$x = 0, 1, \dots \quad \text{e} \quad \alpha > 0$$



Função distribuição cumulativa

$$F_X(x) = \sum_{k=0}^{\infty} \frac{\alpha^k e^{-\alpha}}{k!} u(x - k)$$



2.5.4 Geométrica

Usos mais frequentes

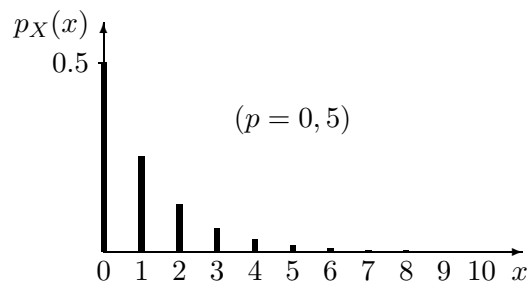
X é o número de falhas antes do primeiro sucesso em uma sequência de testes de Bernoulli independentes, cada uma com probabilidade de sucesso igual a p . É a única variável aleatória discreta sem memória.

Domínio $S_X = \{0, 1, 2, \dots\}$

Função massa de probabilidade

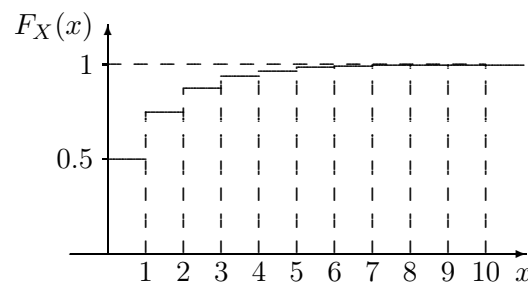
$$p_X(x) = p(1-p)^x$$

$$x = 0, 1, 2, \dots$$



Função distribuição cumulativa

$$F_X(x) = \sum_{k=0}^{\infty} p(1-p)^k u(x - k)$$



2.6 Algumas variáveis aleatórias contínuas importantes

Estamos sempre limitados a medidas de precisão finita, de modo que toda variável aleatória encontrada na prática é uma variável aleatória discreta. Entretanto, existem várias razões pelas quais é interessante utilizar modelos que utilizem variáveis aleatórias contínuas. Primeiro, em geral, variáveis aleatórias contínuas são em geral mais fáceis de lidar analiticamente. Segundo, as formas limite de muitas variáveis aleatórias discretas

geram variáveis aleatórias contínuas. Finalmente, existem algumas “famílias” de variáveis aleatórias contínuas que podem ser utilizadas para modelar uma grande variedade de situações pelo ajuste de alguns poucos parâmetros.

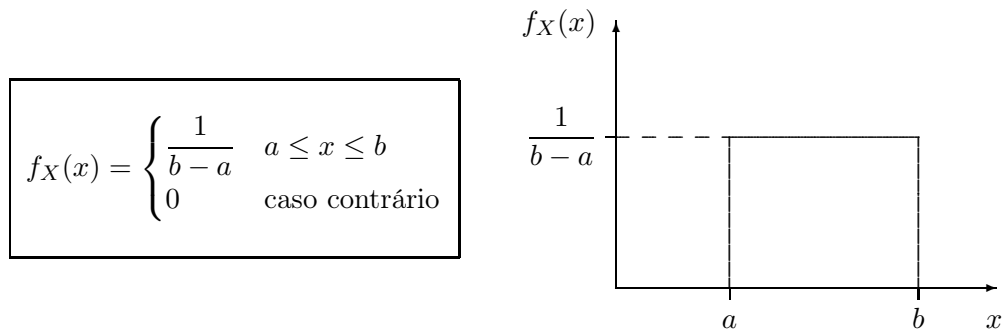
2.6.1 Uniforme

Usos mais frequentes

A variável aleatória uniforme aparece em situações onde todos os valores em um intervalo da reta real são equiprováveis. Esta distribuição é bastante usada em modelamentos de ruído de quantização.

Domínio: $S_X = [a, b]$

Função densidade de probabilidade

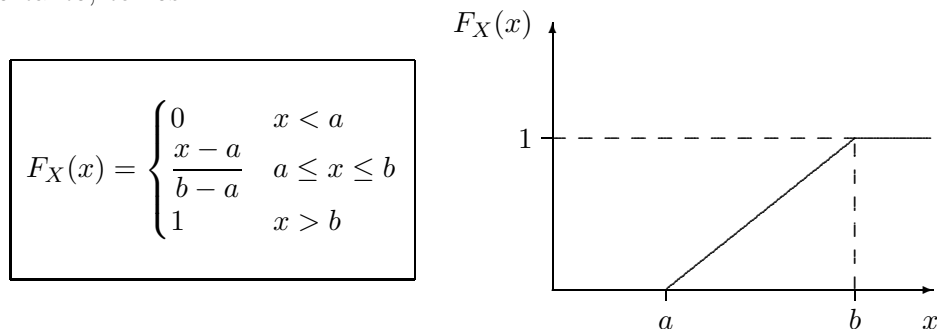


Função distribuição cumulativa

Neste caso, temos 3 situações possíveis:

1. $x < a$ $F_X(x) = \int_{-\infty}^x 0 \, dy = 0$
2. $a \leq x \leq b$ $F_X(x) = \int_a^x \frac{1}{b-a} \, dy = \frac{x-a}{b-a}$
3. $x > b$ $F_X(x) = \int_a^b \frac{1}{b-a} \, dy = \frac{b-a}{b-a} = 1$

Portanto, temos:



2.6.2 Exponencial

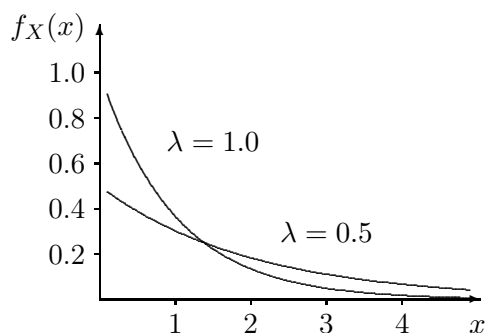
Usos mais frequentes

A variável aleatória exponencial modela o tempo de duração de eventos que ocorrem segundo a distribuição de Poisson. É a única variável aleatória contínua sem memória.

Domínio: $S_X = [0, \infty)$

Função densidade de probabilidade

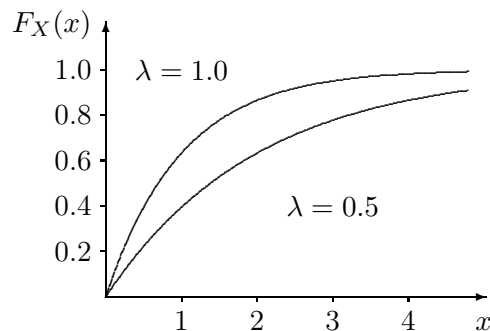
$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \text{ e } \lambda > 0 \\ 0 & \text{caso contrário} \end{cases}$$



Função distribuição cumulativa

$$F_X(x) = \int_0^x \lambda e^{-\lambda y} dy = \lambda \frac{e^{-\lambda y}}{-\lambda} \Big|_0^x = -1(e^{-\lambda x} - e^0) \Rightarrow$$

$$F_X(x) = \begin{cases} 1 - e^{-\lambda x} & x \geq 0, \lambda > 0 \\ 0 & \text{caso contrário} \end{cases}$$



2.6.3 Rayleigh

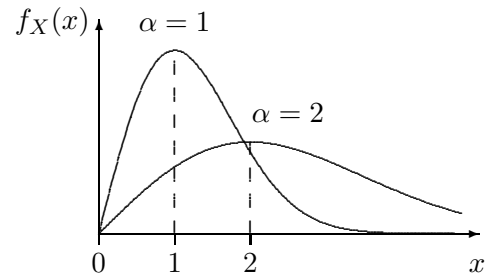
Usos mais frequentes

Modelamento de desvanecimento.

Domínio: $S_X = [0, \infty)$

Função densidade de probabilidade

$$f_X(x) = \begin{cases} \frac{x}{\alpha^2} e^{-\frac{x^2}{2\alpha^2}} & x > 0, \alpha > 0 \\ 0 & \text{caso contrário} \end{cases}$$

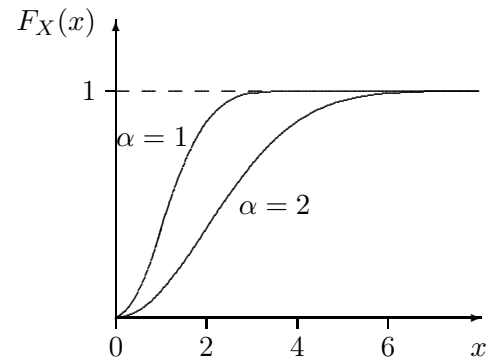


Função distribuição cumulativa

$$F_X(x) = \int_0^x \frac{y}{\alpha^2} e^{-\frac{y^2}{2\alpha^2}} dy \quad \text{Fazendo } u = y^2/2, \text{ temos que } du = y dy.$$

$$F_X(x) = \int_0^{x^2/2} \frac{1}{\alpha^2} e^{-\frac{u}{\alpha^2}} du = \frac{1}{\alpha^2} \frac{e^{-\frac{u}{\alpha^2}}}{-\frac{1}{\alpha^2}} \Big|_0^{x^2/2} = 1 - e^{-\frac{x^2}{2\alpha^2}}$$

$$F_X(x) = \begin{cases} 1 - e^{-\frac{x^2}{2\alpha^2}} & x \geq 0, \alpha > 0 \\ 0 & \text{caso contrário} \end{cases}$$



2.6.4 Gaussianiana

Usos mais frequentes

Curvas em forma de sino aparecem em várias aplicações de teoria de probabilidade. Os modelos de probabilidade nestas aplicações são membros da família de v.a.'s Gaussianas. De fato, sob uma grande faixa de condições X pode ser usada para aproximar a soma de um grande número de variáveis aleatórias independentes. Pelo fato de ocorrerem tão frequentemente na prática, as v.a.'s Gaussianas são também chamadas de v.a.'s normais.

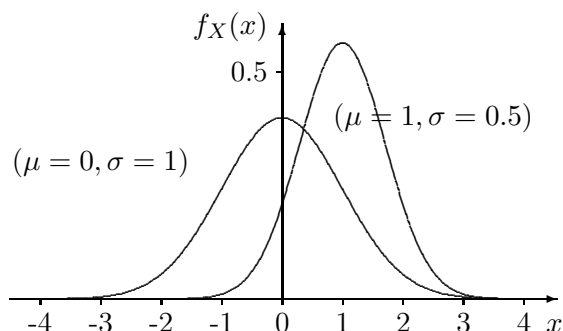
Também, sob uma grande variedade de condições, a variável aleatória gaussiana pode ser utilizada para aproximar a soma de um grande número de variáveis aleatórias independentes. (Veja o Teorema do Limite Central no Capítulo 5)

Seguindo a convenção de vários textos na área, usaremos a notação X é $N(\mu, \sigma^2)$ para nos referirmos a uma v.a. X com distribuição Gaussianiana de média μ e variância σ^2 . Nesta notação, o N quer dizer (obviamente) *normal*.

Domínio: $S_X = (-\infty, \infty)$

Função densidade de probabilidade

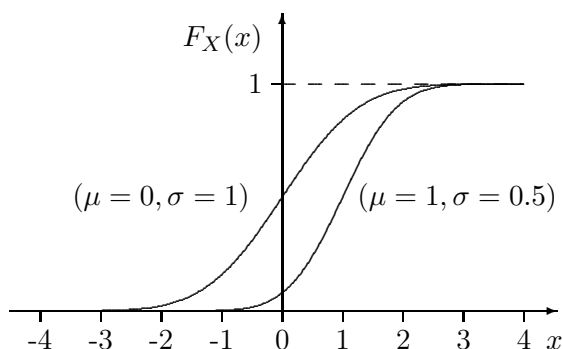
$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



O gráfico de $f_X(x)$ tem formato de sino, com centro em $x = \mu$. σ reflete a largura do sino: se σ é pequeno, o sino é estreito, com um pico agudo e alto. Por outro lado, se σ é grande, o sino é largo, e o pico é baixo e menos pontudo. A altura do pico é dada por $1/(\sigma\sqrt{2\pi})$.

Função distribuição cumulativa

$$F_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x e^{-\frac{(y-\mu)^2}{2\sigma^2}} dy$$



Observações

- É impossível expressar a integral de uma fdp Gaussiana entre limites finitos de forma analítica. Desta forma, a única solução é calcular estes valores de forma numérica. Nas Tabelas do Apêndice F tem-se os valores da fdc de uma variável aleatória Gaussiana $N(0, 1)$ para valores de -4 a 0.
- Observe que como a variável aleatória gaussiana $N(0, 1)$ é simétrica em relação à origem, estas tabelas também fornecem os valores da fdc no intervalo 0 a 4.
- Para valores fora deste intervalo, as probabilidades são muito baixas.

Para aprender como usar esta tabela, vamos introduzir a seguinte propriedade das variáveis aleatórias Gaussianas:

Teorema 2.1. Se X é uma variável aleatória Gaussiana com parâmetros μ e σ , então $Y = aX + b$ é uma variável aleatória Gaussiana com parâmetros $a\mu + b$ e $a\sigma$.

Este teorema diz que qualquer transformação linear de uma variável aleatória Gaussiana produz outra variável aleatória Gaussiana. Este teorema nos permite relacionar

as propriedades de uma variável aleatória Gaussiana arbitrária com as propriedades de uma variável aleatória Gaussiana específica.

Definição 2.9. *Variável aleatória normal padrão.* A variável aleatória normal padrão Z é a variável aleatória Gaussiana com parâmetros $\mu = 0$ e $\sigma = 1$.

As tabelas definidas contêm valores de $F_Z(z)$. Introduzimos a notação especial $\Phi(z)$ para esta função.

Definição 2.10. *Fdc normal padrão.* A fdc da variável normal padrão Z é

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-u^2/2} du$$

Dada a tabela de valores de $\Phi(z)$, usamos o seguinte teorema para encontrar as probabilidades de uma variável aleatória Gaussiana com parâmetros μ e σ

Teorema 2.2. *Se X é uma variável aleatória Gaussiana com parâmetros μ e σ , a fdc de X é*

$$F_X(x) = \Phi\left(\frac{x - \mu}{\sigma}\right)$$

E a probabilidade de X estar no intervalo $(a, b]$ é

$$P[a < X \leq b] = \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right)$$

Usando este teorema, transformamos os valores de uma variável aleatória Gaussiana, X , para valores equivalentes da variável aleatória normal padrão, Z . Para um valor particular x da variável aleatória X , o valor correspondente para a variável aleatória Z é

$$z = \frac{x - \mu}{\sigma} \quad (2.22)$$

Note que z é adimensional. Ele representa x como um número de desvios padrões em relação ao valor esperado de X .

Exemplo 2.7. *Suponha que a sua pontuação em um teste seja $x = 46$, uma amostra de uma variável aleatória Gaussiana com valor esperado 61 e desvio padrão 10. Expresse este resultado como uma amostra da variável aleatória normal padrão Z .*

Solução. Pela Equação (2.22),

$$z = \frac{46 - 61}{10} = -1.5$$

Assim, esta pontuação corresponde a 1.5 desvios padrões menos que o valor esperado.

Para encontrar as probabilidades das variáveis aleatórias Gaussianas, usamos os valores de $\Phi(z)$ apresentados nas tabelas. Note que estas foram calculadas apenas para valores negativos de x . Para valores positivos, devemos usar a seguinte propriedade:

Teorema 2.3. *Para a variável aleatória normal padrão,*

$$\Phi(-z) = 1 - \Phi(z)$$

Exemplo 2.8. *Se X é uma variável aleatória Gaussiana com $\mu = 61$ e $\sigma = 10$, calcule $P[X \leq 46]$*

Solução. Aplicando o Teorema 2.2 e o resultado do Exemplo 2.7, temos

$$P[X \leq 46] = F_X(46) = \Phi(-1.5) = 0,067$$

Isto sugere que, se seu resultado está 1,5 desvios padrões abaixo da média, você está na região dos 6,7% piores, dentro da população das pessoas que fizeram o teste.

A função distribuição cumulativa complementar $Q(x)$.

Uma outra maneira de se calcular as probabilidades de eventos de variáveis aleatórias envolvendo distribuições gaussianas é através do uso da função distribuição cumulativa complementar, definida como

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-y^2/2} dy \quad (2.23)$$

Observe que a função $Q(x)$ corresponde ao valor da probabilidade do evento $P[X > x]$, sendo portanto o complemento da fdc $F_X(x)$, de modo que vale a identidade

$$Q(x) + F_X(x) = 1 \quad (2.24)$$

Desta simetria, pode-se concluir facilmente que a tabela de valores da função $Q(x)$ pode ser obtida diretamente da tabela de valores de $\Phi(x)$. Surge então a pergunta: por que estudar a função $Q(x)$ se já temos a função $\Phi(x)$? Para responder a esta questão, vamos dar uma olhada em outra função, denominada *função erro complementar*, definida como

$$\operatorname{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^{\infty} e^{-y^2} dy \quad (2.25)$$

Esta função tem uma expansão em séries da forma

$$\operatorname{erfc}(x) = 1 - \frac{2}{\sqrt{\pi}} \left[\sum_{i=0}^{\infty} \frac{(-1)^i x^{(2i+1)}}{(2i+1)i!} \right] \quad (2.26)$$

Comparando as Equações (2.23) e (2.25), podemos estabelecer as seguintes relações

$$\operatorname{erfc}(x) = 2Q(x\sqrt{2}) \quad Q(x) = \frac{1}{2} \operatorname{erfc}\left(\frac{x}{\sqrt{2}}\right) \quad (2.27)$$

Para x grande o suficiente (assintoticamente), podemos usar a seguinte representação da função $Q(x)$:

$$Q(x) = \frac{e^{-x^2/2}}{x\sqrt{2\pi}} \left(1 - \frac{1}{x^2} + \frac{1 \times 3}{x^4} - \frac{1 \times 3 \times 5}{x^6} + \dots \right) \quad (2.28)$$

Na prática, as seguintes aproximações são utilizadas

$$Q(x) \approx \frac{1}{x\sqrt{2\pi}} e^{-x^2/2}, x \gg 1 \quad (2.29)$$

$$Q(x) \approx \frac{1}{x\sqrt{2\pi}} \left(1 - \frac{0.7}{x^2} \right) e^{-x^2/2}, x > 2 \quad (2.30)$$

2.6.5 Gama

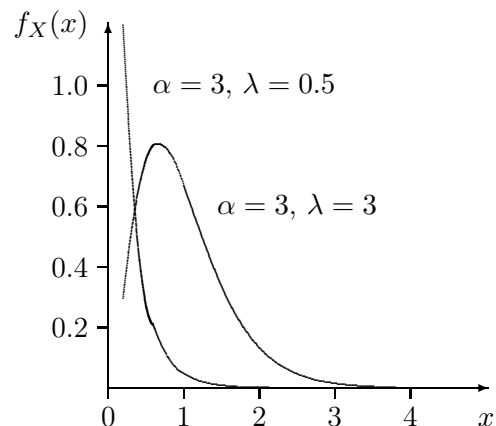
Usos mais frequentes

A distribuição gama não tem muitas aplicações práticas, mas tem um interesse teórico bastante grande, pois serve de base para a derivação de outras distribuições, estas sim de grande interesse prático.

Domínio: $S_X = [0, \infty)$

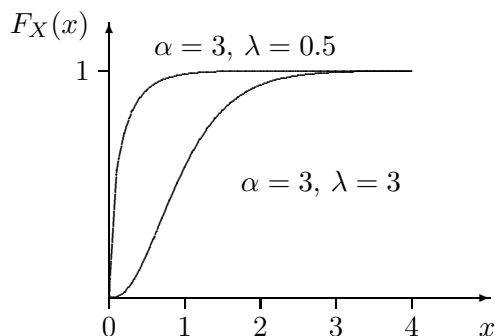
Função densidade de probabilidade

$$f_X(x) = \frac{\lambda(\lambda x)^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)}$$



Função distribuição cumulativa

$$F_X(x) = \int_0^x \frac{\lambda(\lambda y)^{\alpha-1} e^{-\lambda y}}{\Gamma(\alpha)} dy$$



2.6.6 m-Erlang

Usos mais frequentes

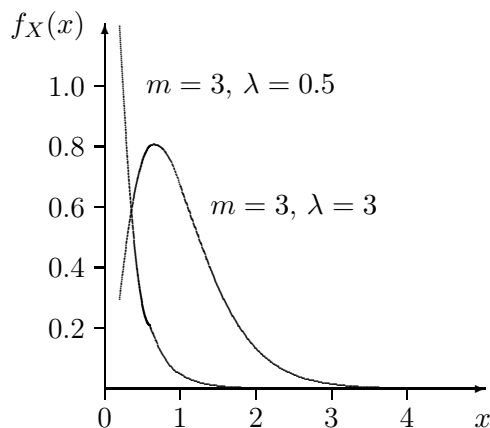
A variável aleatória m-Erlang é obtida pela soma de m variáveis aleatórias independentes com distribuição exponencial de parâmetro λ .

Observação: é um caso especial da distribuição Gama, fazendo-se com que o parâmetro $\alpha = m$ seja um número inteiro positivo.

Domínio: $S_X = [0, \infty)$

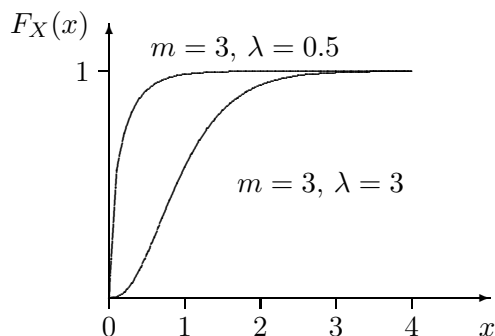
Função densidade de probabilidade

$$f_X(x) = \frac{\lambda e^{-\lambda x} (\lambda x)^{m-1}}{(m-1)!}, x > 0$$



Função distribuição cumulativa

$$F_X(x) = \int_0^x \frac{\lambda e^{-\lambda y} (\lambda y)^{m-1}}{(m-1)!} dy$$



2.6.7 Chi-Quadrado (χ^2)

Usos mais frequentes

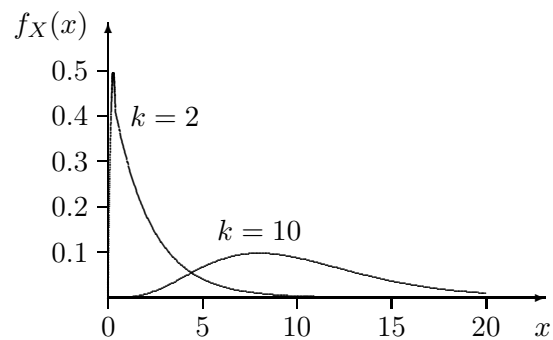
A soma de k variáveis aleatórias gaussianas independentes de média zero e variância unitária, ao quadrado, é uma variável aleatória com distribuição χ^2 com k graus de liberdade.

Observação: é um caso especial da distribuição Gama, fazendo-se $\alpha = k/2$, k inteiro positivo, e $\lambda = 1/2$.

Domínio: $S_X = [0, \infty)$

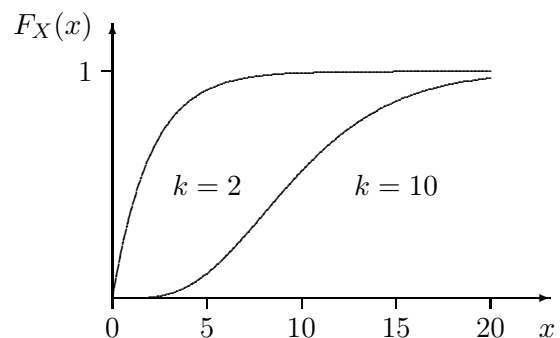
Função densidade de probabilidade

$$f_X(x) = \frac{x^{(k-2)/2} e^{-x/2}}{2^{k/2} \Gamma(k/2)}$$



Função distribuição cumulativa

$$F_X(x) = \int_0^x \frac{y^{(k-2)/2} e^{-y/2}}{2^{k/2} \Gamma(k/2)} dy$$



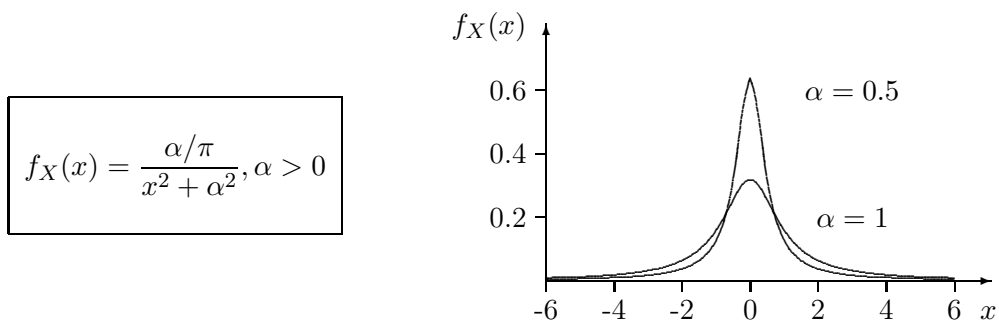
2.6.8 Cauchy

Usos mais frequentes

A distribuição de Cauchy não tem aplicação prática, mas tem um grande interesse teórico pelas suas peculiaridades.

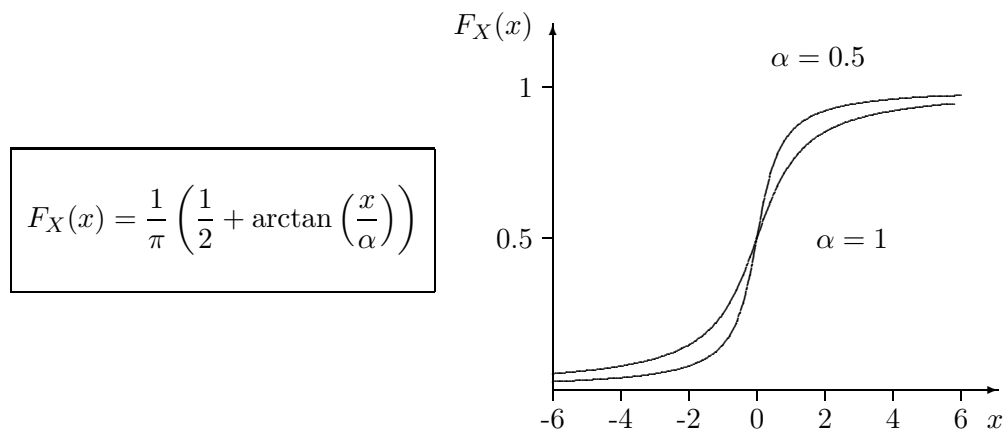
Domínio: $S_X = [-\infty, \infty)$

Função densidade de probabilidade



Função distribuição cumulativa

$$\begin{aligned} F_X(x) &= \int_{-\infty}^x \frac{\alpha/\pi}{\alpha^2 + u^2} du \\ &= \frac{a}{\pi} \left[\frac{1}{a} \arctan\left(\frac{u}{a}\right) \right]_{-\infty}^x \end{aligned}$$



2.6.9 Laplace

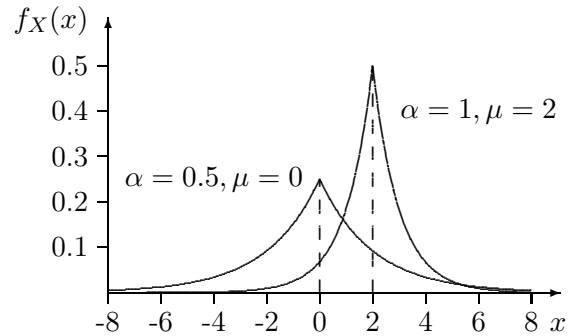
Usos mais frequentes

A distribuição de Laplace é também conhecida como distribuição exponencial dupla. É a distribuição das diferenças entre duas variáveis aleatórias iid com distribuição exponencial.

Domínio: $S_X = [-\infty, \infty)$

Função densidade de probabilidade

$$f_X(x) = \frac{\alpha}{2} e^{-\alpha|x-\mu|}, \alpha > 0$$



Função distribuição cumulativa

Por causa da presença do módulo na expressão da fdp, precisamos derivar a fdc em duas etapas:

Primeira etapa: $x \leq \mu$

$$F_X(x) = \int_{-\infty}^x \frac{\alpha}{2} e^{-\alpha|y-\mu|} dy \quad \text{Fazendo } z = y - \mu, \text{ temos que } dz = dy, \text{ e então}$$

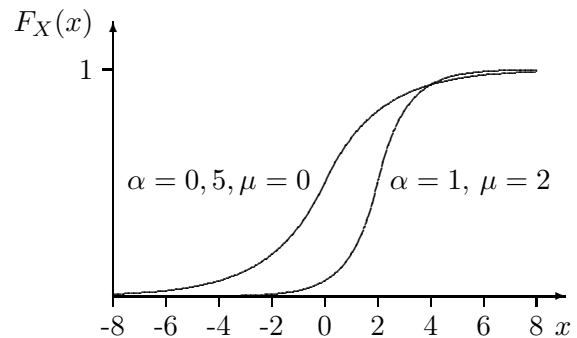
$$F_X(x) = \int_{-\infty}^{x-\mu} \frac{\alpha}{2} e^{-\alpha z} dz = \frac{\alpha}{2} \frac{e^{-\alpha y}}{-\alpha} \Big|_{-\infty}^{x-\mu} = \frac{1}{2} e^{-\alpha(x-\mu)}$$

Segunda etapa: $x > \mu$

$$F_X(x) = \int_{-\infty}^{\mu} \frac{\alpha}{2} e^{-\alpha|y-\mu|} dy + \int_{\mu}^x \frac{\alpha}{2} e^{-\alpha|y-\mu|} dy \quad \text{Fazendo } z = y - \mu, dz = dy, \text{ e então:}$$

$$F_X(x) = \int_{-\infty}^{\mu} \frac{\alpha}{2} e^{-\alpha z} dz + \int_{\mu}^x \frac{\alpha}{2} e^{-\alpha z} dz = \frac{\alpha}{2} \frac{e^{-\alpha y}}{-\alpha} \Big|_{-\infty}^{\mu} + \frac{\alpha}{2} \frac{e^{-\alpha y}}{-\alpha} \Big|_{\mu}^x = \frac{1}{2} - \frac{1}{2} e^{-\alpha(x-\mu)}$$

$$F_X(x) = \begin{cases} \frac{1}{2} e^{\alpha(x-\mu)}, & x \leq \mu \\ \frac{1}{2} - \frac{1}{2} e^{-\alpha(x-\mu)}, & x > \mu \end{cases}$$



2.7 Densidades Condicionais

Se temos informação adicional sobre o experimento sob análise, então nossas expectativas podem (ou não) ser alteradas. Por exemplo, ao fazermos apostas em um hipódromo, se sabemos que um cavalo está machucado ou doente, mesmo que seja um campeão, diminuímos nossa confiança nele.

Nesta seção iremos mostrar como determinar a influência de uma informação adicional na fdc de uma variável aleatória. Isto é bastante fácil se lembrarmos que a fdc é na verdade uma probabilidade:

Definição 2.11. Função de distribuição condicional. A função de distribuição condicional $F_X(x|B)$ de uma variável aleatória X dado o evento B é definida como

$$F_X(x|B) = P[X \leq x|B] = \frac{P[X \leq x, B]}{P[B]}$$

Propriedades

A função distribuição condicional $F_X(x|B)$ tem as mesmas propriedades de uma fdc comum. Dentre elas, podemos destacar:

1. $F_X(-\infty|B) = 0$
2. $F_X(\infty|B) = 1$
3. $P[a < X \leq b|B] = F_X(b|B) - F_X(a|B)$

Definição 2.12. Se X é uma variável aleatória discreta, então a função massa de probabilidade condicional é dada por

$$p_X(x_k|B) = P[X = x_k|B] = \frac{P[X = x_k, B]}{P[B]}$$

Se X é uma variável aleatória contínua, então a função densidade de probabilidade condicional é dada por

$$f_X(x|B) = \frac{dF_X(x|B)}{dx}$$

Exemplo 2.9. Seja $B \triangleq \{X \leq 10\}$. Determine $F_X(x|B)$.

Solução. Para resolver este problema, vamos analisá-lo em duas partes:

1. para $x \geq 10$, o evento $\{X \leq 10\}$ é um subconjunto do evento $\{X \leq x\}$. Desta forma, $P[X \leq 10, X \leq x] = P[X \leq 10]$, e então podemos escrever

$$F_X(x|B) = \frac{P[X \leq 10, X \leq x]}{P[X \leq 10]} = \frac{P[X \leq 10]}{P[X \leq 10]} = 1$$

2. para $x \leq 10$, o evento $\{X \leq x\}$ é um subconjunto do evento $\{X \leq 10\}$. Desta forma, $P[X \leq 10, X \leq x] = P[X \leq x]$, e então podemos escrever

$$F_X(x|B) = \frac{P[X \leq 10, X \leq x]}{P[X \leq 10]} = \frac{P[X \leq x]}{P[X \leq 10]}$$

Na Figura abaixo temos uma versão gráfica deste resultado.

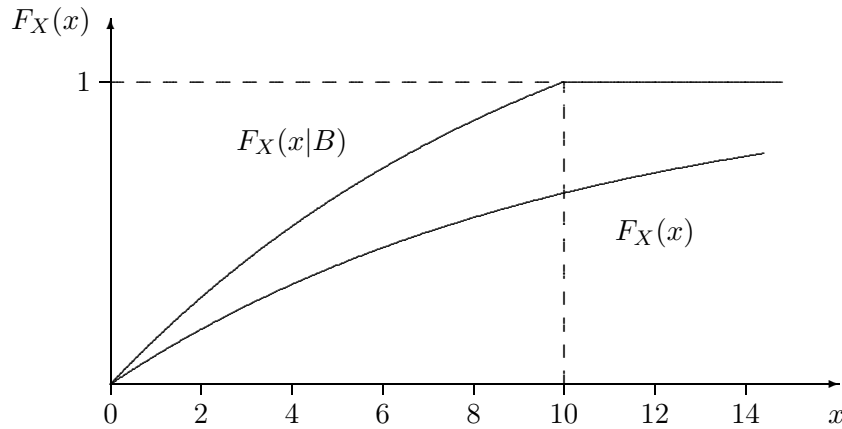


Figura 2.10: Fdc's condicional e incondicional de X .

2.8 Variáveis Aleatórias Múltiplas

Quando lidamos com experimentos combinados ou tentativas repetidas de um mesmo experimento, encontramos v.a.'s múltiplas e suas fdc's e fdp's. Variáveis aleatórias múltiplas são basicamente funções multidimensionais definidas em um espaço amostral de um experimento combinado.

2.8.1 Função Distribuição de Probabilidade Conjunta

Sejam duas v.a.'s X_1 e X_2 , cada uma delas podendo ser contínua, discreta ou mista.

Definição 2.13. A função distribuição cumulativa conjunta (fdc conjunta) para as duas v.a.'s pode ser definida como

$$F_{X_1 X_2}(x_1, x_2) = P[X_1 \leq x_1, X_2 \leq x_2] = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} f_{X_1 X_2}(u_1, u_2) du_1 du_2 \quad (2.31)$$

onde $f_{X_1 X_2}(x_1, x_2)$ é a função densidade de probabilidade conjunta (fdp conjunta). Esta última pode ser expressa na forma

$$f_{X_1 X_2}(x_1, x_2) = \frac{\partial^2}{\partial x_1 \partial x_2} F_{X_1 X_2}(x_1, x_2) \quad (2.32)$$

2.8.2 Densidades marginais

Teorema 2.4. Quando a fdp conjunta $f_{X_1X_2}(x_1, x_2)$ é integrada sobre uma das variáveis, obtemos a fdp da outra variável, isto é

$$\begin{aligned}\int_{-\infty}^{+\infty} f_{X_1X_2}(x_1, x_2) dx_1 &= f_{X_2}(x_2) \\ \int_{-\infty}^{+\infty} f_{X_1X_2}(x_1, x_2) dx_2 &= f_{X_1}(x_1)\end{aligned}$$

As fdp's $f_{X_1}(x_1)$ e $f_{X_2}(x_2)$ obtidas a partir da integração de uma das variáveis são chamadas de **fdp's marginais**.

Corolário 2.5. Se $f_{X_1X_2}(x_1, x_2)$ é integrada sobre ambas as variáveis, obtemos

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{X_1X_2}(x_1, x_2) dx_1 dx_2 = F(\infty, \infty) = 1 \quad (2.33)$$

Corolário 2.6. $F(-\infty, -\infty) = F(-\infty, x_2) = F(x_1, -\infty) = 0$

No caso de v.a.'s discretas, substituímos as integrais por somatórios.

Teorema 2.7. Para as v.a.'s discretas X e Y , temos:

$$\begin{aligned}p_X(x_i) &= P[X = x_i] = P[X = x_i, Y = y_1 \text{ ou } X = x_i, Y = y_2 \text{ ou } \dots] \\ &= \sum_{j=-\infty}^{\infty} p_{XY}(x_i, y_j)\end{aligned}$$

$$\begin{aligned}p_Y(y_j) &= P[Y = y_j] = P[Y = y_j, X = x_1 \text{ ou } Y = y_j, X = x_2 \text{ ou } \dots] \\ &= \sum_{i=-\infty}^{\infty} p_{XY}(x_i, y_j)\end{aligned}$$

E a expressão correspondente à Equação 2.33 para o caso discreto é

Teorema 2.8.

$$\sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} p_{XY}(x_i, y_j) = F(\infty, \infty) = 1 \quad (2.34)$$

Exemplo 2.10. Duas linhas de produção fabricam um certo tipo de peça. Suponha que a capacidade (em qualquer dia) seja 5 peças na linha I e 3 peças na linha II. Admita que o número de peças realmente produzidas em qualquer linha seja uma v.a. e que (X, Y) represente a v.a. bidimensional que fornece o número de peças produzidas pela linha I e a linha II, respectivamente. A Tabela 2.1 fornece a distribuição de probabilidade conjunta de (X, Y) . Calcule as probabilidades marginais.

Tabela 2.1: Exemplo de probabilidades conjunta e marginal.

$\downarrow Y \ X \rightarrow$	0	1	2	3	4	5	Soma
0	0	0,01	0,03	0,05	0,07	0,09	0,25
1	0,01	0,02	0,04	0,05	0,06	0,08	0,26
2	0,01	0,03	0,05	0,05	0,05	0,06	0,25
3	0,01	0,02	0,04	0,06	0,06	0,05	0,24
Soma	0,03	0,08	0,16	0,21	0,24	0,28	1

Solução. Na Tabela 2.1, cada casa representa

$$p_{XY}(x_i, y_j) = P[X = x_i, Y = y_j]$$

A última linha e a última coluna fornecem os totais marginais, isto é, a soma das 6 colunas e 4 linhas da tabela. As probabilidades que aparecem nas margens, linha e coluna, representam a distribuição de probabilidade de Y e de X , respectivamente. Por exemplo, $P[Y = 1] = 0.26$, $P[X = 3] = 0.21$, etc.

Em virtude da forma de apresentação da Tabela 2.1 aludiremos, de modo muito usual à distribuição marginal de X ou à distribuição marginal de Y , sempre que tivermos uma v.a. bidimensional (X, Y) , quer discreta, quer contínua.

2.8.3 Caso multidimensional

A generalização das expressões acima para v.a.'s multidimensionais é direta. Suponha que X_i , $i = 1, 2, \dots, n$ são v.a.'s com uma fdc conjunta definida por

$$\begin{aligned} F_{X_1 X_2 \dots X_n}(x_1, x_2, \dots, x_n) &= P[X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n] \\ &= \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \dots \int_{-\infty}^{x_n} f_{X_1 X_2 \dots X_n}(u_1, u_2, \dots, u_n) du_1 du_2 \dots du_n \end{aligned} \quad (2.35)$$

onde $f_{X_1 X_2 \dots X_n}(x_1, x_2, \dots, x_n)$ é a fdp conjunta.

Tomando as derivadas parciais de $F_{X_1 X_2 \dots X_n}(x_1, x_2, \dots, x_n)$ dadas por (2.35), obtemos

$$f_{X_1 X_2 \dots X_n}(x_1, x_2, \dots, x_n) = \frac{\partial^n}{\partial x_1 \partial x_2 \dots \partial x_n} F_{X_1 X_2 \dots X_n}(x_1, x_2, \dots, x_n) \quad (2.36)$$

Um número qualquer de variáveis de $f_{X_1 X_2 \dots X_n}(x_1, x_2, \dots, x_n)$ pode ser eliminado integrando-se sobre estas variáveis. Por exemplo, integrando-se sobre x_2 e x_3 leva a

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{X_1 X_2 X_3 X_4 \dots X_n}(x_1, x_2, x_3, x_4, \dots, x_n) dx_2 dx_3 = f_{X_1 X_4 \dots X_n}(x_1, x_4, \dots, x_n) \quad (2.37)$$

Segue também que

$$F_{X_1 X_2 \dots X_n}(x_1, \infty, \infty, x_4, \dots, x_n) = F_{X_1 X_4 \dots X_n}(x_1, x_4, \dots, x_n)$$

e

$$F_{X_1 X_2 \dots X_n}(x_1, -\infty, -\infty, x_4, \dots, x_n) = 0.$$

2.8.4 Função distribuição de probabilidade condicional

Teorema 2.9. *Sejam duas v.a.'s X_1 e X_2 com fdp conjunta $f_{X_1 X_2}(x_1, x_2)$. A fdc $F_{X_1}(x_1)$ condicionada por*

$$x_2 - \Delta x_2 < X_2 \leq x_2$$

onde Δx_2 é algum incremento positivo, é dada por

$$F_{X_1}(x_1|x_2) = \frac{\int_{-\infty}^{x_1} f_{X_1 X_2}(u_1, x_2) du_1}{f_{X_2}(x_2)}$$

Demonstração. Sejam X_1 e X_2 duas v.a.'s com fdp conjunta $f_{X_1 X_2}(x_1, x_2)$. Queremos determinar $P[X_1 \leq x_1]$ condicionada por

$$x_2 - \Delta x_2 < X_2 \leq x_2$$

onde Δx_2 é algum incremento positivo. Em outras palavras, desejamos calcular a probabilidade do evento $(X_1 \leq x_1 | x_2 - \Delta x_2 < X_2 \leq x_2)$. Usando as relações estabelecidas anteriormente para a probabilidade condicional de um evento, a probabilidade do evento $(X_1 \leq x_1 | x_2 - \Delta x_2 < X_2 \leq x_2)$ pode ser expressa como

$$\begin{aligned} P[X_1 \leq x_1 | x_2 - \Delta x_2 < X_2 \leq x_2] &= \frac{P[X_1 \leq x_1, x_2 - \Delta x_2 < X_2 \leq x_2]}{P[x_2 - \Delta x_2 < X_2 \leq x_2]} \\ &= \frac{\int_{-\infty}^{x_1} \int_{x_2 - \Delta x_2}^{x_2} f_{X_1 X_2}(u_1, u_2) du_1 du_2}{\int_{x_2 - \Delta x_2}^{x_2} f_{X_2}(u_2) du_2} \end{aligned} \quad (2.38)$$

Vamos agora utilizar um resultado da teoria do cálculo diferencial e integral para continuarmos com a nossa prova:

Teorema 2.10. Teorema do Valor Médio: se f for uma função contínua em $[a, b]$ e diferenciável em (a, b) , então existe $c \in (a, b)$ tal que $f(b) - f(a) = f'(c)(b - a)$.

De acordo com o Teorema do Valor Médio enunciado acima, existem pontos c e $c' \in (x_2 - \Delta x_2, x_2)$ tais que

$$\frac{\int_{-\infty}^{x_1} \int_{x_2 - \Delta x_2}^{x_2} f_{X_1 X_2}(u_1, u_2) du_1 du_2}{\int_{x_2 - \Delta x_2}^{x_2} f_{X_2}(u_2) du_2} = \frac{\int_{-\infty}^{x_1} f_{X_1 X_2}(u_1, c) \Delta x_2 du_1}{f_{X_2}(c') \Delta x_2} \quad (2.39)$$

Fazendo agora $\Delta x_2 \rightarrow 0$, temos que c e c' aproximam-se de x_2 , e desta forma, podemos reescrever (2.39) como

$$\frac{\int_{-\infty}^{x_1} f_{X_1 X_2}(u_1, c) \Delta x_2 du_1}{f_{X_2}(c') \Delta x_2} = \frac{\int_{-\infty}^{x_1} f_{X_1 X_2}(u_1, x_2) du_1}{f_{X_2}(x_2)} \quad (2.40)$$

que é a fdc condicional da v.a. X_1 dada a v.a. X_2 , ou seja

$$F_{X_1}(x_1|x_2) = \frac{\int_{-\infty}^{x_1} f_{X_1 X_2}(u_1, x_2) du_1}{f_{X_2}(x_2)} \quad (2.41)$$

□

Corolário 2.11. $F_{X_1}(-\infty|x_2) = 0$ e $F_{X_1}(+\infty|x_2) = 1$.

Teorema 2.12.

$$f_{X_1}(x_1|x_2) = \frac{f_{X_1 X_2}(x_1, x_2)}{f_{X_2}(x_2)} \quad (2.42)$$

Demonstração. Este corolário é demonstrado diretamente derivando (2.40) em relação a x_1 , obtemos a fdp $f_{X_1 X_2}(x_1|x_2)$ correspondente na forma □

Alternativamente, podemos expressar a fdp conjunta $f_{X_1 X_2}(x_1, x_2)$ em termos das fdp's condicionais

$$f_{X_1 X_2}(x_1, x_2) = f_{X_1}(x_1|x_2) f_{X_2}(x_2) = f_{X_2}(x_2|x_1) f_{X_1}(x_1) \quad (2.43)$$

A extensão das relações dadas acima para o caso multidimensional é direta:

$$f_{X_1 \dots X_n}(x_1, \dots, x_n) = f_{X_1 \dots X_k}(x_1, \dots, x_k | x_{k+1}, \dots, x_n) f_{X_{k+1} \dots X_n}(x_{k+1}, \dots, x_n) \quad (2.44)$$

onde k é qualquer inteiro na faixa $1 < k < n$. A fdc condicional conjunta correspondente à fdp $f_{X_1 \dots X_k}(x_1, \dots, x_k | x_{k+1}, \dots, x_n)$ é dada por

$$F_{X_1 \dots X_k}(x_1, \dots, x_k | x_{k+1}, \dots, x_n) = \frac{\int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_k} f_{X_1 \dots X_k}(u_1, \dots, u_k | x_{k+1}, \dots, x_n) du_1 \dots du_k}{f_{X_{k+1} \dots X_n}(x_{k+1}, \dots, x_n)} \quad (2.45)$$

Esta fdc condicional satisfaz as propriedades previamente estabelecidas para estas funções tais como

$$\begin{aligned} F_{X_1 X_2 \dots X_k}(\infty, x_2, \dots, x_k | x_{k+1}, \dots, x_n) &= F_{X_2 \dots X_k}(x_2, \dots, x_k | x_{k+1}, \dots, x_n) \\ F_{X_1 X_2 \dots X_k}(-\infty, x_2, \dots, x_k | x_{k+1}, \dots, x_n) &= 0 \end{aligned}$$

2.8.5 Independência Estatística de Variáveis Aleatórias

Já definimos a independência estatística para dois ou mais eventos de uma espaço amostral S . Este conceito pode ser estendido para variáveis aleatórias definidas em um espaço amostral gerado por um experimento combinado ou por várias tentativas de um único experimento. Se os experimentos gerarem resultados mutuamente exclusivos, a probabilidade de um resultado em um experimento é independente de um resultado em qualquer outro experimento. Isto é, a probabilidade conjunta dos resultados pode ser fatorada no produto das probabilidades correspondentes a cada resultado. Consequentemente, as variáveis aleatórias correspondentes aos resultados nestes experimentos são independentes no sentido de que sua fdp conjunta pode ser fatorada no produto das fdp's marginais.

Definição 2.14. *As v.a.'s multidimensionais são estatisticamente independentes se e somente se*

$$F_{X_1 X_2 \dots X_n}(x_1, x_2, \dots, x_n) = F_{X_1}(x_1) F_{X_2}(x_2) \dots F_{X_n}(x_n) \quad (2.46)$$

ou alternativamente

$$f_{X_1 X_2 \dots X_n}(x_1, x_2, \dots, x_n) = f_{X_1}(x_1) f_{X_2}(x_2) \dots f_{X_n}(x_n) \quad (2.47)$$

2.9 Funções de Variáveis Aleatórias

2.9.1 Caso Unidimensional

Um problema que surge frequentemente em aplicações de probabilidade é o seguinte: dada uma v.a. X , caracterizada por sua fdp $f_X(x)$, calcular a fdp da v.a. $Y = g(X)$,

onde $g(X)$ é alguma função de X . Chamemos a fdp desejada de $f_Y(y)$.

Teorema 2.13. *Sejam duas v.a.'s X e Y , com $Y = g(X)$. Nestas condições, a fdp de Y é dada por*

$$f_Y(y) = \frac{f_X(x)}{|g'(X)|} \Big|_{x=g^{-1}(y)}$$

Demonstração. Inicialmente, vamos analisar os gráficos da Figura 2.11.

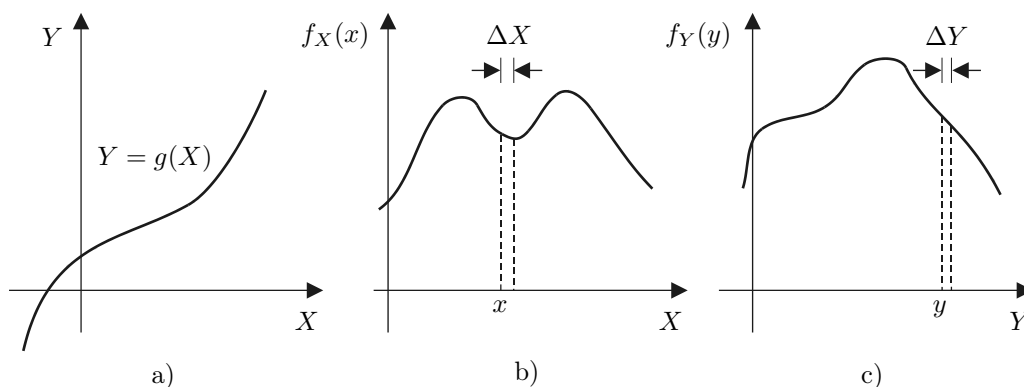


Figura 2.11: a) Dependência entre X e Y , b) $f_X(x)$, e c) $f_Y(y)$.

Se X sofre uma variação $\Delta X \rightarrow 0$, e a variação correspondente em Y é dada por ΔY , então é óbvio que a probabilidade de observar X no intervalo $[x, x + \Delta x]$ é a mesma de observar Y no intervalo $[y, y + \Delta y]$. Mas estas probabilidades são dadas por $f_X(x)\Delta x$ e $f_Y(y)\Delta y$, respectivamente. Portanto

$$\lim_{\Delta x \rightarrow 0} f_X(x)\Delta x = f_Y(y)\Delta y \quad (2.48)$$

Propriamente falando a equação acima deveria ser expressa como

$$\lim_{\Delta x \rightarrow 0} f_X(x)|\Delta x| = f_Y(y)|\Delta y| \quad (2.49)$$

pois as probabilidades são iguais às magnitudes das áreas sob ΔX e ΔY , respectivamente. Desta forma

$$f_Y(y) = \frac{f_X(x)}{\left| \frac{dY}{dX} \right|} = \frac{f_X(x)}{|g'(X)|} \quad (2.50)$$

Observe que $f_Y(y)$ é uma função de y . Desta forma, no lado direito da equação acima, a variável x deve ser expressa em termos de y . Assumindo que $y = g(x)$ tem uma inversa $x = g^{-1}(y)$, temos

$$f_Y(y) = \frac{f_X(x)}{|g'(X)|} \Big|_{x=g^{-1}(y)} \quad (2.51)$$

□

Exemplo 2.11. A função densidade de probabilidade de uma variável aleatória X é dada por

$$f_X(x) = \begin{cases} \frac{x^2}{81}, & -3 < x < 6 \\ 0, & \text{caso contrário} \end{cases}$$

Calcule a função densidade de probabilidade da variável aleatória $U = \frac{1}{3}(12 - x)$.

Solução. Neste caso, $g(x) = 1/3(12 - x)$. Assim, a derivada e a inversa de $g(x)$ são dadas por:

$$g'(x) = \frac{1}{3}(0 - 1) = -\frac{1}{3} \qquad g^{-1}(U) = 12 - 3U$$

Aplicando o Teorema 2.13, temos:

$$f_U(u) = \left. \frac{\frac{X^2}{81}}{\left| -\frac{1}{3} \right|} \right|_{g^{-1}(U)} = \left. \frac{x^2}{27} \right|_{g^{-1}(U)} = \frac{(4 - U)^2}{3}$$

Ainda, para X variando no intervalo $(-3, 6)$, U varia no intervalo $(2, 5)$, e a solução final é então dada por:

$$f_U(u) = \begin{cases} \frac{(4 - u)^2}{3}, & 2 < u < 5 \\ 0, & \text{caso contrário} \end{cases}$$

Exemplo 2.12. Considere a v.a. Y definida como $Y = aX + b$, $a > 0$. Se X tem fdp dada por $f_X(x)$, encontre a fdp de Y em termos da fdp de X .

Solução. Na Figura 2.12a) tem-se o mapeamento de X contra Y . Notamos que este mapeamento é linear e monotônico. Sejam $F_X(x)$ e $F_Y(y)$ as fdc's para X e Y , respectivamente. Então

$$F_Y(y) = P[Y \leq y] = P[aX + b \leq y] = P\left[X \leq \frac{y - b}{a}\right] = \int_{-\infty}^{\frac{y-b}{a}} f_X(x) dx = F_X\left(\frac{y - b}{a}\right)$$

Derivando a equação acima em relação a y , obtemos a relação entre as respectivas fdp's

$$f_Y(y) = \frac{1}{a} f_X\left(\frac{y - b}{a}\right)$$

Ou seja, se a fdp de X é da forma da Figura 2.12b), a fdp de Y será aquela mostrada na Figura 2.12c).

Uma outra forma de resolver este problema é aplicando diretamente o Teorema 2.13. Neste caso, temos:

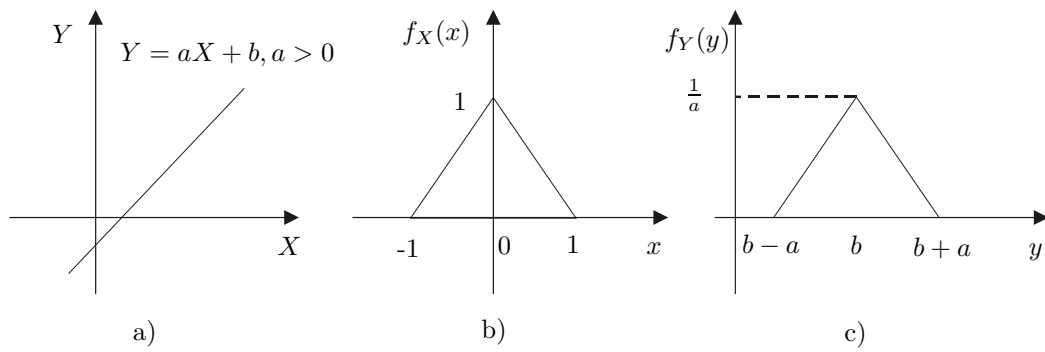


Figura 2.12: Uma transformação da v.a. X e um exemplo das fdp's correspondentes de X e Y .

$$f_Y(y) = \frac{f_X(x)}{g'(x)} \Big|_{x=g^{-1}(y)} = \frac{f_X(x)}{a} \Big|_{x=(y-b)/a} = \frac{1}{a} f_X\left(\frac{y-b}{a}\right)$$

Até agora assumiu-se implicitamente que existe uma correspondência biunívoca entre X e Y ou seja, existe apenas um valor de X para um dado Y , e vice-versa. Se, por outro lado, para um dado valor de Y existir mais de um valor de X , as equações acima devem ser modificadas. O seguinte corolário trata deste caso:

Corolário 2.14. Quando a equação $Y = g(X)$ tem duas raízes, x_1 e x_2 , a fdp $f_Y(y)$ é dada por

$$f_Y(y) = \frac{f_X(x_1)}{|g'_1(x_1)|} \Big|_{x_1=g_1^{-1}(y)} + \frac{f_X(x_2)}{|g'_2(x_2)|} \Big|_{x_2=g_2^{-1}(y)}$$

Demonstração. Considere a relação $Y = g(X)$ mostrada na Figura 2.13.

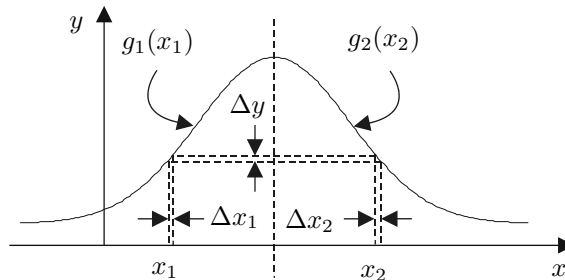


Figura 2.13: Função de uma v.a. com duas raízes.

Nesta Figura, para um dado valor de Y existem dois valores correspondentes para X . Então a equação $Y = g(X)$ tem duas raízes, x_1 e x_2 . Vamos quebrar esta função em duas outras, cada qual com uma única raiz: $Y = g_1(X_1)$ e $Y = g_2(X_2)$.

Note que agora temos uma correspondência unívoca entre X e Y em cada uma destas funções. Então x_1 e x_2 são funções de y com uma única raiz. Chamemos as

relações inversas de $x_1 = g_1^{-1}(y)$ e $x_2 = g_2^{-1}(y)$.

Da Figura 2.13 temos que Y está no intervalo $(y, y + \Delta y)$ quando x_1 está no intervalo $(x_1, x_1 + \Delta x_1)$ ou quando x_2 está no intervalo $(x_2, x_2 + \Delta x_2)$. Os dois últimos eventos são mutuamente exclusivos, pois X pode assumir o valor x_1 ou o valor x_2 mas não ambos. Desta forma, temos

$$f_Y(y)|\Delta y| = \lim_{\substack{\Delta x_1 \rightarrow 0 \\ \Delta x_2 \rightarrow 0}} (f_X(x_1)|\Delta x_1| + f_X(x_2)|\Delta x_2|) \quad (2.52)$$

$$f_Y(y) = \left. \frac{f_X(x_1)}{|g'_1(x_1)|} \right|_{x_1=g^{-1}(y)} + \left. \frac{f_X(x_2)}{|g'_2(x_2)|} \right|_{x_2=g^{-1}(y)} \quad (2.53)$$

□

Se existirem n valores de X para um dado valor de Y , podemos estender este resultado para o seguinte corolário:

Corolário 2.15. Quando a equação $Y = g(X)$ tem n raízes, x_1, \dots, x_n , a fdp $f_Y(y)$ é dada por

$$f_Y(y) = \left. \frac{f_X(x_1)}{|g'_1(x_1)|} \right|_{x_1=g_1^{-1}(y)} + \dots + \left. \frac{f_X(x_n)}{|g'_n(x_n)|} \right|_{x_n=g_n^{-1}(y)}$$

onde x_1, x_2, \dots, x_n são os valores de X quando $Y = y$.

Exemplo 2.13. Considere a v.a. Y definida como $Y = aX^2 + b$, $a > 0$. Se X tem fdp dada por $f_X(x)$, encontre a fdp de Y em termos da fdp de X .

Solução. Na Figura 2.14 temos o mapeamento de Y em relação a X .

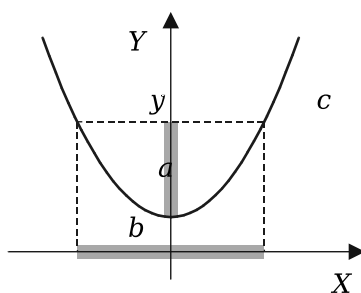


Figura 2.14: Uma transformação quadrática da v.a. X .

Para determinar a fdc de Y , observamos que

$$F_Y(y) = P[Y \leq y] = P[aX^2 + b \leq y] = P[|X| \leq \sqrt{\frac{y-b}{a}}]$$

e então

$$F_Y(y) = F_X\left(\sqrt{\frac{y-b}{a}}\right) - F_X\left(-\sqrt{\frac{y-b}{a}}\right)$$

Derivando a equação acima em relação a y , obtemos a fdp de Y em termos da fdp de X

$$f_Y(y) = \frac{f_X\left[\sqrt{\frac{y-b}{a}}\right]}{2a\sqrt{\frac{y-b}{a}}} + \frac{f_X\left[-\sqrt{\frac{y-b}{a}}\right]}{2a\sqrt{\frac{y-b}{a}}}$$

Utilizando agora o Corolário 2.9.1, temos: a equação $g(X) = aX^2 + b = y$ tem duas soluções reais $x_1 = \sqrt{\frac{y-b}{a}}$ e $x_2 = -\sqrt{\frac{y-b}{a}}$, e portanto, $f_Y(y)$ consiste de dois termos correspondentes a estas duas soluções

$$f_Y(y) = \frac{f_X\left[x_1 = \sqrt{\frac{y-b}{a}}\right]}{\left|g'_X\left[x_1 = \sqrt{\frac{y-b}{a}}\right]\right|} + \frac{f_X\left[x_2 = -\sqrt{\frac{y-b}{a}}\right]}{\left|g'_X\left[x_2 = -\sqrt{\frac{y-b}{a}}\right]\right|} = \frac{f_X\left[\sqrt{\frac{y-b}{a}}\right]}{2a\sqrt{\frac{y-b}{a}}} + \frac{f_X\left[-\sqrt{\frac{y-b}{a}}\right]}{2a\sqrt{\frac{y-b}{a}}}$$

2.9.2 Caso Multidimensional

Teorema 2.16. Considere duas v.a.'s X e Y e sua fdp conjunta $f_{XY}(x, y)$. Sejam U e V outras duas v.a.'s relacionadas a X e Y por $U = U(X, Y)$ e $V = V(X, Y)$. Suponha que tanto U como V assumem valores únicos para valores particulares de X e Y , e vice-versa. Então

$$f_{UV}(u, v) = \frac{f_{XY}(x, y)}{J\left(\frac{u, v}{x, y}\right)}$$

Demonstração. Considere duas v.a.'s X e Y e sua fdp conjunta $f_{XY}(x, y)$. Sejam U e V outras duas v.a.'s relacionadas a X e Y por $U = U(X, Y)$ e $V = V(X, Y)$. Suponha que tanto U como V assumem valores únicos para valores particulares de X e Y , e vice-versa. Similarmente ao caso unidimensional, para obter $f_{UV}(u, v)$ a partir de $f_{XY}(x, y)$, observe que

$$f_{UV}(u, v)|dudv| = f_{XY}(x, y)|dxdy| \quad (2.54)$$

Portanto

$$f_{UV}(u, v) = \frac{f_{XY}(x, y)}{\left|\frac{dudv}{dxdy}\right|} \quad (2.55)$$

A relação entre os dois elementos de área nos dois sistemas de coordenadas pode ser expressa em termos do Jacobiano como

$$dudv = J\left(\frac{u,v}{x,y}\right) dx dy \quad (2.56)$$

onde J é o Jacobiano da transformação, dado pelo determinante

$$J\left(\frac{u,v}{x,y}\right) = \begin{vmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{vmatrix} \quad (2.57)$$

Portanto

$$f_{UV}(u,v) = \frac{f_{XY}(x,y)}{J\left(\frac{u,v}{x,y}\right)} \quad (2.58)$$

Note que para que o Jacobiano exista as derivadas parciais de u e v em relação a x e a y devem também existir. □

Teorema 2.17. *Se X e Y são funções de múltiplos valores, isto é, se $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ são as soluções das equações $U = U(X, Y)$ e $V = V(X, Y)$ então*

$$f_{UV}(u,v) = \frac{f_{XY}(x_1, y_1)}{\left|J\left(\frac{u,v}{x_1, y_1}\right)\right|} + \frac{f_{XY}(x_2, y_2)}{\left|J\left(\frac{u,v}{x_2, y_2}\right)\right|} + \dots + \frac{f_{XY}(x_n, y_n)}{\left|J\left(\frac{u,v}{x_n, y_n}\right)\right|} \quad (2.59)$$

O resultado acima pode ser estendido a qualquer número de variáveis. Suponha que temos n v.a.'s X_1, X_2, \dots, X_n com uma fdp conjunta $f_{X_1 X_2 \dots X_n}(x_1, x_2, \dots, x_n)$. Desejamos encontrar a fdp conjunta $f_{Y_1 Y_2 \dots Y_n}(y_1, y_2, \dots, y_n)$ de n v.a.'s relacionadas com X_1, X_2, \dots, X_n por

$$Y_i = Y_i(X_1, X_2, \dots, X_n), \quad i = 1, 2, \dots, n$$

$$X_j = X_j(Y_1, Y_2, \dots, Y_n), \quad j = 1, 2, \dots, n$$

Assume-se que todas essas funções sejam de valor único e com derivadas parciais contínuas em todos os pontos. Assim, temos

$$f_{Y_1 Y_2 \dots Y_n}(y_1, y_2, \dots, y_n) |dy_1, dy_2, \dots, dy_n| = f_{X_1 X_2 \dots X_n}(x_1, x_2, \dots, x_n) |dx_1, dx_2, \dots, dx_n| \quad (2.60)$$

Portanto

$$f_{Y_1 Y_2 \dots Y_n}(y_1, y_2, \dots, y_n) = \frac{f_{X_1 X_2 \dots X_n}(x_1, x_2, \dots, x_n)}{\left| \frac{dy_1, dy_2, \dots, dy_n}{dx_1, dx_2, \dots, dx_n} \right|} \quad (2.61)$$

A razão dos elementos de área é dada pelo Jacobiano da transformação $J\left(\frac{y_1, \dots, y_n}{x_1, \dots, x_n}\right)$

$$dy_1, dy_2, \dots, dy_n = J\left(\frac{y_1, y_2, \dots, y_n}{x_1, x_2, \dots, x_n}\right) dx_1, dx_2, \dots, dx_n \quad (2.62)$$

onde

$$J\left(\frac{y_1, y_2, \dots, y_n}{x_1, x_2, \dots, x_n}\right) = \begin{vmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \dots & \frac{\partial y_1}{\partial x_n} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \dots & \frac{\partial y_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_n}{\partial x_1} & \frac{\partial y_n}{\partial x_2} & \dots & \frac{\partial y_n}{\partial x_n} \end{vmatrix} \quad (2.63)$$

Portanto

$$f_{Y_1 Y_2 \dots Y_n}(y_1, y_2, \dots, y_n) = \frac{f_{X_1 X_2 \dots X_n}(x_1, x_2, \dots, x_n)}{J\left(\frac{y_1, y_2, \dots, y_n}{x_1, x_2, \dots, x_n}\right)} \quad (2.64)$$

Pode-se mostrar que

$$J\left(\frac{y_1, y_2, \dots, y_n}{x_1, x_2, \dots, x_n}\right) = \frac{1}{J\left(\frac{x_1, x_2, \dots, x_n}{y_1, y_2, \dots, y_n}\right)} \quad (2.65)$$

Se Y_1, Y_2, \dots, Y_n são funções de múltiplos valores de X_1, X_2, \dots, X_n , uma equação similar a (2.59) deve ser utilizada. (Qual ?)

Exemplo 2.14. Para ilustrar o exemplo de transformação de uma fdp de segunda ordem, considere o caso do arremesso de um dardo. Assuma que ambas as variáveis X e Y que descrevem as coordenadas de um ponto onde o dardo atinge o alvo são independentes e tem fdp's normais (gaussianas)

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} \quad e \quad f_Y(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{y^2}{2\sigma^2}}$$

Encontre a fdp $f_{R\Theta}(r, \theta)$ onde R é a distância do ponto à origem e Θ o ângulo do ponto em relação ao eixo x . As relações entre as variáveis são as seguintes:

$$R = \sqrt{X^2 + Y^2} \quad e \quad \Theta = \arctg\left(\frac{Y}{X}\right)$$

Solução.

$$J\left(\frac{R, \Theta}{X, Y}\right) = \begin{vmatrix} \frac{\partial R}{\partial X} & \frac{\partial R}{\partial Y} \\ \frac{\partial \Theta}{\partial X} & \frac{\partial \Theta}{\partial Y} \end{vmatrix}$$

Assim, a fdp $f_{R\Theta}(r, \theta)$ pode ser escrita como

$$f_{R\Theta}(r, \theta) = \frac{f_{XY}(x, y)}{\left| J \left(\frac{r, \theta}{x, y} \right) \right|} = \frac{r}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} = \frac{r}{2\pi\sigma^2} e^{-\frac{r^2}{2\sigma^2}}$$

A variável Θ não aparece na equação acima. Isto quer dizer que as variáveis R e Θ são independentes e $f_{\Theta}(\theta)$ precisa ser uma constante. Desde que Θ varia no intervalo $[0, 2\pi]$, é evidente que $f_{\Theta}(\theta)$ é uma constante de modo a termos $\int_0^{2\pi} f_{\Theta}(\theta) d\Theta = 1$.

Portanto

$$f_{R\Theta}(r, \theta) = \left(\frac{1}{2\pi} \right) \left(\frac{r}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}} \right) = f_R(r) f_{\Theta}(\theta)$$

onde

$$f_{\Theta}(\theta) = \begin{cases} \frac{1}{2\pi}, & 0 < \Theta < 2\pi \\ 0, & \text{caso contrário} \end{cases}$$

$$f_R(r) = \frac{r}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}}$$

$f_R(r)$ é conhecida como função densidade de Rayleigh.

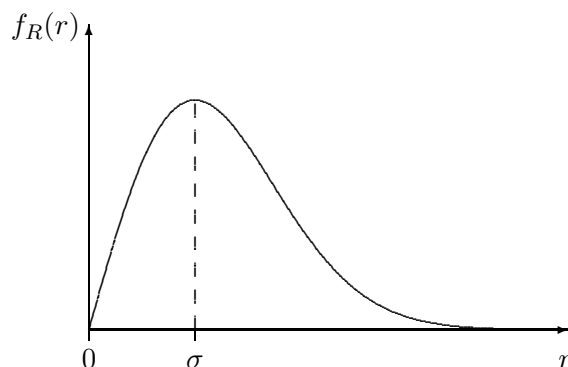


Figura 2.15: Função densidade de probabilidade de Rayleigh.

2.10 Exercícios

1. A função densidade de probabilidade da amplitude de um certo sinal (em volts) é dada por

$$f_X(x) = x e^{-x} u(x)$$

- (a) Qual a probabilidade da amplitude do sinal ser maior que 1 volt?
- (b) Qual é a probabilidade de observar a amplitude do sinal na faixa de 1 a 2 volts?

Resp: a) $2e^{-1}$ b) $2e^{-1} - 3e^{-2}$

2. A função densidade de probabilidade conjunta $f_{XY}(x, y)$ de duas v.a.'s contínuas X e Y é dada por

$$f_{XY}(x, y) = xye^{-\frac{x^2+y^2}{2}}u(x)u(y)$$

- (a) Encontre as seguintes funções densidade de probabilidade: $f_X(x)$, $f_Y(y)$, $f_{XY}(x|Y = y)$, $f_{XY}(y|X = x)$.
 (b) As v.a.'s X e Y são independentes?

Resp:

- (a) $f_X(x) = xe^{-\frac{x^2}{2}}$
 $f_Y(y) = ye^{-\frac{y^2}{2}}$
 $f_{XY}(x|Y = y) = xe^{-\frac{x^2}{2}}$
 $f_{XY}(y|X = x) = ye^{-\frac{y^2}{2}}$
 (b) sim

3. A função densidade de probabilidade conjunta $f_{XY}(x, y)$ de duas v.a.'s contínuas X e Y é dada por

$$f_{XY}(x, y) = ke^{-(x^2+2xy+2y^2)}$$

- (a) Determine o valor da constante k .
 (b) Determine as funções densidade de probabilidade $f_X(x)$, $f_Y(y)$, $f_{XY}(x|Y = y)$, e $f_{XY}(y|X = x)$.
 (c) Estas duas v.a.'s são independentes?

- (a) $k = 1/\pi$
 (b) $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$
 $f_Y(y) = \frac{1}{\sqrt{\pi}} e^{-y^2}$
 $f_{XY}(x|Y = y) = \frac{1}{\sqrt{\pi}} e^{-(x^2+2xy+y^2)}$
 $f_{XY}(y|X = x) = \sqrt{\frac{2}{\pi}} e^{-\left(\frac{x^2}{2}+2xy+2y^2\right)}$
 (c) não

4. O sinal de entrada X e o sinal de saída Y de um retificador de meia onda são relacionados por

$$Y = \begin{cases} X^2, & X > 0 \\ 0, & X \leq 0 \end{cases}$$

A função densidade de probabilidade do sinal de entrada é dada por

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}$$

Encontre $f_Y(y)$.

$$\text{Resp: } f_Y(y) = \begin{cases} \frac{1}{2\sigma\sqrt{2\pi}y} e^{-\frac{y}{2\sigma^2}}, & y > 0 \\ 0, & \text{caso contrário} \end{cases}$$

5. Repita o problema anterior para um retificador de onda completa.

Dica: para um retificador de onda completa, os sinais de entrada e saída estão relacionados por $Y = X^2$.

$$\text{Resp: } f_Y(y) = \begin{cases} \frac{1}{\sigma\sqrt{2\pi}y} e^{-\frac{y}{2\sigma^2}}, & y > 0 \\ 0, & \text{caso contrário} \end{cases}$$

6. Suponha que três usuários de telefone tenham uma linha em comum. Qual a probabilidade de mais de um deles utilizar a linha ao mesmo tempo? Admita que, em média, um usuário utilize o aparelho durante 5 minutos por hora.

Resp: $425/21600 \approx 0,0197$

7. Se 20% dos bits transmitidos por um transmissor acusam defeito, determine a probabilidade de que, em 4 bits transmitidos ao acaso:

- (a) Um seja errado
- (b) Nenhum esteja errado
- (c) Ao menos dois estejam errados

Resp: a) 0,4096 b) 0,4096 c) 0,1808

8. Se os defeitos de um tecido seguem uma lei de Poisson com média de defeito a cada 500 m, qual a probabilidade de que o intervalo entre dois defeitos consecutivos seja:

- (a) no mínimo 1250 m
- (b) entre 1000 m e 1250 m
- (c) menor que 1000 m

Resp: a) $e^{-5/2} \approx 0,082$ b) $e^{-2} - e^{-5/2} \approx 0,053$ c) $1 - e^{-2} \approx 0,865$

9. Sabe-se que a média de carros com um pneu furado durante a travessia de um determinado túnel é de 0,06 casos/mês. Calcular a probabilidade de pelo menos 2 carros terem um pneu furado ao passar pelo túnel durante um mês de tráfego normal, sabendo-se que a distribuição é de Poisson.

Resp: 0,0017

10. Suponha que a variável aleatória X tem uma distribuição de chi-quadrado, com 10 graus de liberdade. Se pedirmos para determinar dois números a e b , tais que $P(a < x < b) = 0,85$, por exemplo, deveremos compreender que existem muitos

pares dessa espécie. Determine dois diferentes conjuntos de valores (a, b) que satisfaçam à condição acima. Suponha que, em aditamento ao acima, se exija que $P(X < a) = P(X > b)$.

Resp: $a = 4,45$ e $b = 16,97$

11. A *fdp* de uma variável aleatória X é $f_X(x)$. Uma variável aleatória Y é definida como $Y = aX + b$, $a < 0$. Determine a *fdp* de Y em termos da *fdp* de X .

Resp: $f_Y(y) = -\frac{1}{a}f_X\left(\frac{y-b}{a}\right)$, $a < 0$

12. Verifique quais das funções abaixo podem ser consideradas *fdc*'s. Justifique sua resposta.

$$\text{a) } y = \begin{cases} 0 & x < 0 \\ x^2 & 0 \leq x < 1 \\ 1 & x \geq 1 \end{cases} \quad \text{b) } y = \begin{cases} 1 - e^{-2x} & x \geq 0 \\ 0 & x < 0 \end{cases} \quad \text{c) } y = \begin{cases} -2x & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Resp: Apenas o item c) não pode ser *fdc*.

13. A *fdc* conjunta de duas variáveis aleatórias X e Y é dada por

$$F_{XY}(x, y) = \begin{cases} (1 - e^{-\alpha x})(1 - e^{-\beta y}) & x \geq 0, y \geq 0 \\ 0 & \text{caso contrário} \end{cases}$$

- (a) Encontre as *fdc*'s marginais.
 (b) Encontre as probabilidades dos eventos
 i) $A = \{X \leq 1, Y \leq 1\}$
 ii) $B = \{X > x, Y > y\}$, $x > 0, y > 0$

Dica: use a lei de De Morgan

Resp:

$$\begin{aligned} \text{(a) } F_X(x) &= \begin{cases} 1 - e^{-\alpha x}, & x \geq 0 \\ 0, & \text{caso contrário} \end{cases} \\ F_Y(y) &= \begin{cases} 1 - e^{-\beta y}, & y \geq 0 \\ 0, & \text{caso contrário} \end{cases} \\ \text{(b) i. } P[X \leq 1, Y \leq 1] &= (1 - e^{-\alpha})(1 - e^{-\beta}) \\ \text{ii. } P[X > x, Y > y] &= e^{-\alpha x}e^{-\beta y} \end{aligned}$$

14. Uma variável aleatória X tem função densidade de probabilidade dada por

$$f_X(x) = \frac{c}{x^2 + 1}, \quad -\infty < X < \infty$$

- (a) Determine o valor da constante c .
 (b) Calcule a probabilidade do evento $[1/3 \leq X^2 \leq 1]$.
 (c) Determine a função distribuição de probabilidade de X .

Resp:

- (a) $c = 1/\pi$
- (b) $P[1/3 \leq X^2 \leq 1] = 1/6$
- (c) $F_X(x) = \frac{1}{2} + \frac{1}{\pi} \arctg(x)$

15. Seja a variável aleatória X com função densidade de probabilidade dada por

$$f_X(x) = \begin{cases} 6x(1-x), & 0 < x < 1 \\ 0, & \text{caso contrário} \end{cases}$$

Determine uma função $Y = h(X)$ que tenha a função densidade de probabilidade

$$f_Y(y) = \begin{cases} 12y^3(1-y^2), & 0 < y < 1 \\ 0, & \text{caso contrário} \end{cases}$$

Dica: Admita que a função incógnita h seja tal que os intervalos $X \leq x$ e $Y \leq y$ se correspondam biunívoca e continuamente, de forma que $P[X \leq x] = P[Y \leq y]$, ou seja $F_X(x) = F_Y(y)$.

Resp: $Y = \sqrt{X}$

16. Assuma que duas variáveis aleatórias X e Y têm função densidade de probabilidade conjunta dada por

$$f_{XY}(x, y) = \frac{1}{2\pi} \exp \left[-\frac{1}{2}(x^2 + y^2) \right]$$

Sejam duas outras variáveis aleatórias U e W definidas da seguinte maneira:

$$U \triangleq 3X + 5Y \qquad W \triangleq X + 2Y$$

Determine a função densidade de probabilidade conjunta de U e W .

Resp: $f_{UW}(u, w) = \frac{1}{2\pi} e^{-\frac{1}{2}(5U^2 - 26UW + 34W^2)}$

17. Seja uma v.a. com fdp dada por

$$f_X(x) = ke^{-\lambda|x|}, \quad \lambda > 0, \quad -\infty < x < \infty$$

onde k é uma constante.

- (a) Calcule o valor de k .
- (b) Encontre a função distribuição cumulativa de X .
- (c) Calcule $P[1 \leq X \leq 2]$ usando a fdp, para $\lambda = 1$.
- (d) Calcule $P[1 \leq X \leq 2]$ usando a fdc, para $\lambda = 1$.

Resp:

- (a) $k = \frac{\lambda}{2}$
- (b) $F_X(x) = \begin{cases} \frac{1}{2} e^{\lambda x}, & x < 0 \\ 1 - \frac{1}{2} e^{-\lambda x}, & x \geq 0 \end{cases}$
- (c) $E[X] = 0, \text{Var}[X] = \frac{2}{\lambda^2}$
- (d) $\frac{1}{2}(e^{-1} - e^{-2}) \approx 0,1163$
- (e) $\frac{1}{2}(e^{-1} - e^{-2}) \approx 0,1163$

18. A probabilidade de uma chamada telefônica não durar mais do que t minutos é geralmente descrita por uma *fdc* exponencial

$$F_T(t) = \begin{cases} 1 - e^{-t/3} & t \geq 0 \\ 0 & \text{caso contrário} \end{cases}$$

Qual é a *fdp* da duração em minutos de uma conversa telefônica? Qual é a probabilidade de uma conversação durar entre 2 e 4 minutos?

Resp:

- (a) $f_T(t) = \begin{cases} \frac{1}{3} e^{-t/3}, & t \geq 0 \\ 0, & \text{caso contrário} \end{cases}$
- (b) $P[2 \leq t \leq 4] = e^{-2/3} - e^{-4/3} \approx 0,25$

19. Expresse os valores extremos das *fdc*'s conjuntas $F_{XY}(x, y)$ por números ou em termos das *fdc*'s $F_X(x)$ e $F_Y(y)$.

- (a) $F_{XY}(-\infty, 2)$ (b) $F_{XY}(\infty, \infty)$
 (c) $F_{XY}(\infty, y)$ (d) $F_{XY}(\infty, -\infty)$

Resp: a) 0 b) 1 c) $F_Y(y)$ d) 0

20. Considere as variáveis aleatórias X e Y com *fdp* conjunta

$$f_{XY}(x, y) = \begin{cases} 4xy & 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0 & \text{caso contrário} \end{cases}$$

X e Y são independentes?

Resp: sim

21. Sejam X e Y duas v.a.'s com *fdp* conjunta dada por

$$f_{XY}(x, y) = \begin{cases} A(x + y), & 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0, & \text{caso contrário} \end{cases}$$

- (a) Calcule o valor de A .
- (b) Calcule as fdp's marginais.
- (c) X e Y são independentes?

Resp:

- (a) $A = 1$
- (b) $f_X(x) = x + 1/2$
 $f_Y(y) = y + 1/2$
- (c) não

22. Que distribuição de probabilidade você pode utilizar para modelar as seguintes situações?

- (a) Número de toques entre erros de digitação, dado que cada toque tem uma certa probabilidade de estar com erro;
- (b) Número de toques com erro dentre m toques, dado que cada toque tem uma certa probabilidade de estar com erro;
- (c) Tempo entre chegadas sucessivas, dado que as chegadas são sem memória;
- (d) Tempo de serviço de um dispositivo que consiste de m servidores sem memória, em série.

Resp: (a) Geométrica (b) Binomial (c) Exponencial (d) m-Erlang

23. Uma fonte binária gera dígitos 0 e 1 de forma aleatória com probabilidades 0,6 e 0,4, respectivamente.

- (a) Qual é a probabilidade de que ocorram dois 1s e três 0s em uma sequência de cinco dígitos?
- (b) Qual a probabilidade de ocorrerem pelo menos três 1s em uma sequência de cinco dígitos?

Resp: (a) 0,3456 (b) 0,31744

24. Seja $Y = e^X$. Encontre a fdp de Y se $X = N(\mu, \sigma^2)$.

$$\text{Resp: } f_Y(y) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma y} e^{-\frac{(\ln(y)-\mu)^2}{2\sigma^2}} & y > 0 \\ 0 & \text{caso contrário} \end{cases}$$

25. A função densidade de probabilidade conjunta de duas variáveis aleatórias X e Y é dada por

$$f_{XY}(x, y) = A \sin(x + y), \quad 0 \leq x \leq \pi/2, \quad 0 \leq y \leq \pi/2$$

Determine:

- (a) A constante A .

- (b) A função distribuição de probabilidade conjunta $F_{XY}(x, y)$
- (c) As funções distribuições de probabilidade marginais $F_X(x)$ e $F_Y(y)$.
- (d) A probabilidade do evento $\frac{\pi}{6} \leq X \leq \frac{\pi}{4}$.

Resp:

- (a) $A = 0,5$
- (b) $F_{XY}(x, y) = 0,5 [\sin(x) + \sin(y) - \sin(x + y)]$
- (c) $F_X(x) = 0,5 [1 - \cos(x) + \sin(x)]$ $F_Y(y) = 0,5 [1 - \cos(y) + \sin(y)]$
- (d) 0,18

26. Uma companhia aérea sabe que 5% das pessoas que fazem reservas em um determinado voo não comparecem para o embarque. Consequentemente, sua política é vender 52 passagens para um voo que pode transportar até 50 passageiros. Qual a probabilidade de haver assentos disponíveis para todos os passageiros que comparecerem ao embarque?

Resp: 0,7405

27. Suponha que um sinal $x(t)$ alimenta um dispositivo cuja saída seja $y(t)$. Se X tem distribuição uniforme no intervalo $(0, 2)$ e $y = \ln(x)$, calcule $f_Y(y)$ e $F_Y(y)$. Faça os esboços das curvas pertinentes, indicando valores em alguns pontos notáveis (onde a função corta os eixos ou muda abruptamente).

Resp:

$$f_Y(y) = \begin{cases} \frac{e^y}{2}, & -\infty < y < \ln(2) \\ 0, & y \geq \ln(2) \end{cases}$$

$$F_Y(y) = \begin{cases} \frac{e^y}{2}, & -\infty < y < \ln(2) \\ 1, & y \geq \ln(2) \end{cases}$$

28. Sabe-se que a distância (em metros) do ponto de aterrissagem de um paraquedista em relação ao centro da área alvo pode ser modelada como uma variável aleatória contínua X com distribuição de Rayleigh de parâmetro $\sigma^2 = 100$.

- (a) Encontre a probabilidade de o paraquedista aterrisar dentro de um raio de 10m do centro da área alvo.
- (b) Encontre o raio r tal que a probabilidade do evento $\{X > r\}$ seja e^{-1} .

Resp:

- (a)
- (b)

29. Uma fonte gera um sinal de ruído com distribuição gaussiana de média zero e potência 2 W. Encontre a probabilidade de a amplitude do sinal exceder 5 volts.

Resp: $Q(5/\sqrt{2}) \approx 2,0563 \cdot 10^{-4}$

30. Repita o problema anterior, se a potência for de 1 W.

Resp: $Q(5) \approx 2,89 \cdot 10^{-7}$

Capítulo 3

Médias Estatísticas de Variáveis Aleatórias

3.1 Médias

O conceito de médias assume uma posição extremamente importante em processos aleatórios. Como mencionado anteriormente, os processos aleatórios são caracterizados pela **regularidade estatística**. Usando o termo regularidade estatística indicamos que o processo não pode ser predito especificamente, mas pode ser predito em uma base média. Por exemplo, no experimento de jogar moedas não é possível prever o resultado de uma jogada particular, mas em média, podemos confiar que metade das jogadas irão ser caras, e a outra metade, coroas, dado que esta média seja feita sobre um número suficientemente grande de jogadas.

3.1.1 Média de uma Variável Aleatória

Considere uma v.a. X que pode assumir n valores x_1, x_2, \dots, x_n . Suponha que o experimento (representado por X) foi repetido N vezes ($N \rightarrow \infty$) e sejam m_1, m_2, \dots, m_n o número de tentativas favoráveis aos resultados x_1, x_2, \dots, x_n , respectivamente. Então o valor médio de X é dado por

$$E[X] = \frac{1}{N} (m_1 x_1 + m_2 x_2 + \dots + m_n x_n) = \frac{m_1}{N} x_1 + \frac{m_2}{N} x_2 + \dots + \frac{m_n}{N} x_n \quad (3.1)$$

No limite quando $N \rightarrow \infty$, a razão m_i/N tende a $f_X(x_i)$ de acordo com a definição por frequência relativa de probabilidade. Portanto

$$E[X] = \sum_{i=1}^n x_i p_X(x_i) \quad (3.2)$$

O valor médio é também chamado de valor esperado da v.a. X .

Definição 3.1. A média ou valor esperado de uma v.a. discreta é dado por

$$E[X] = \sum_{i=1}^n x_i p_X(x_i) \quad (3.3)$$

Se X é uma v.a. contínua, o valor médio é dado por

$$E[X] = \int_{-\infty}^{+\infty} x f_X(x) dx \quad (3.4)$$

Exemplo 3.1. Uma fdp gaussiana geral é dada por

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}$$

Encontre o valor médio de X .

Solução. Na Figura 3.1 tem-se um esboço de $f_X(x)$. Para esta distribuição, temos

$$E[X] = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} x e^{-\frac{(x-m)^2}{2\sigma^2}} dx$$

Fazendo $X = Y + m$, podemos reescrever a equação acima como

$$E[X] = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} (y + m) e^{-\frac{y^2}{2\sigma^2}} dy = \frac{1}{\sqrt{2\pi}\sigma} \left[\int_{-\infty}^{+\infty} y e^{-\frac{y^2}{2\sigma^2}} dy + m \int_{-\infty}^{+\infty} e^{-\frac{y^2}{2\sigma^2}} dy \right]$$

O integrando da primeira integral é uma função ímpar de y , e por isso, o resultado é zero. O da segunda integral é uma função par de y , de modo que podemos reescrever a equação acima como

$$E[X] = \frac{1}{\sqrt{2\pi}\sigma} 2m \int_0^{+\infty} e^{-\frac{y^2}{2\sigma^2}} dy = \frac{1}{\sqrt{2\pi}\sigma} 2m \frac{1}{2} \sqrt{2\pi\sigma^2} = m$$

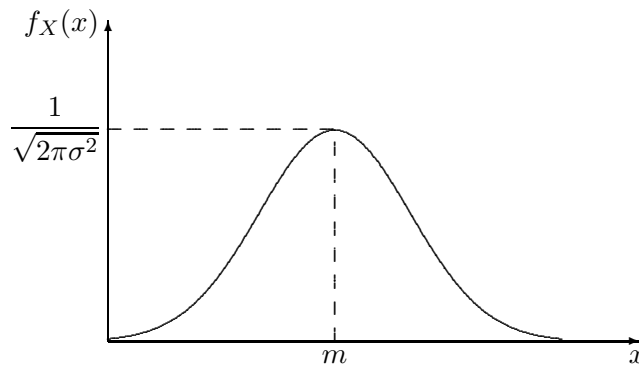


Figura 3.1: Função densidade de probabilidade gaussiana com média m e variância σ^2 .

3.1.2 Média de uma Função de uma Variável Aleatória

Frequentemente desejamos encontrar a média de uma função de uma v.a. ao invés da média da própria v.a. Como um exemplo simples disso, analisemos o caso de um sinal de ruído cuja amplitude é representada por uma v.a. X . Na prática estamos mais interessados no valor quadrático médio do sinal do que no valor médio deste.

De forma geral, desejamos obter a expressão do valor médio de uma v.a. Y a qual por sua vez é uma função da v.a. X

$$Y = g(X) \quad (3.5)$$

Teorema 3.1. *Sejam duas v.a.'s X e Y relacionadas por $Y = g(X)$. Então o valor médio de Y é dado por*

$$E[Y] = \int_{-\infty}^{+\infty} y f_Y(y) dy = \int_{-\infty}^{+\infty} g(X) f_X(x) dx \quad (3.6)$$

Demonstração. Considere o gráfico da Figura 3.2, onde aparece um esboço da curva de X contra $Y = g(X)$

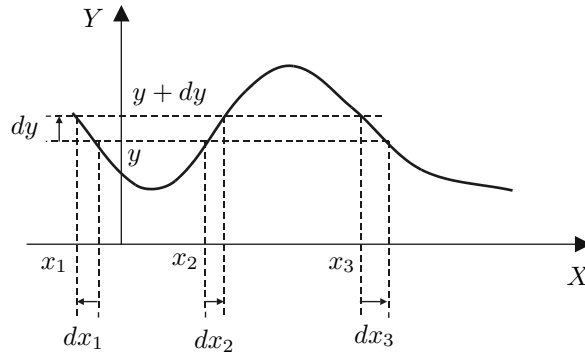


Figura 3.2: $Y = g(X)$.

Da figura, podemos ver que $y = g(x_1) = g(x_2) = g(x_3)$, então podemos escrever

$$f_Y(y)dy = f_X(x_1)dx_1 + f_X(x_2)dx_2 + f_X(x_3)dx_3 \quad (3.7)$$

Multiplicando ambos os lados da equação por y , obtemos

$$y f_Y(y)dy = g(x_1) f_X(x_1)dx_1 + g(x_2) f_X(x_2)dx_2 + g(x_3) f_X(x_3)dx_3 \quad (3.8)$$

Então para cada diferencial em (3.8) correspondem um ou mais diferenciais em (3.6). À medida que dy cobre o eixo y , os dx 's correspondentes são não sobrepostos e cobrem todo o eixo x . Desta forma, as integrais em (3.8) e (3.6) são iguais, e a prova está completa.

□

Teorema 3.2. Se X é uma v.a. discreta, (3.6) pode ser reescrita como

$$E[Y] = \sum_i g(x_i)P[X = x_i] = \sum_i g(x_i)p_X(x_i) \quad (3.9)$$

Exemplo 3.2. Encontrar o valor quadrático médio da distribuição gaussiana do Exemplo 3.1.

Solução. Temos que $Y = g(X) = X^2$. Então

$$E[Y] = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} x^2 e^{-\frac{(x-m)^2}{2\sigma^2}} dx$$

Fazendo $u = x - m$, reescrevemos a equação acima como

$$E[Y] = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} (u+m)^2 e^{-\frac{u^2}{2\sigma^2}} du$$

Resolvendo as três integrais acima, chega-se a (resolver)

$$E[Y] = E[X^2] = \sigma^2 + m^2$$

3.1.3 Médias para Variáveis Múltiplas

Seja Z uma v.a. que é função de duas v.a.'s X e Y

$$Z = g(X, Y) \quad (3.10)$$

Então

$$E[Z] = \int_{-\infty}^{+\infty} z f_Z(z) dz \quad (3.11)$$

Podemos calcular $E[Z]$ a partir de (3.11) e do conhecimento da densidade conjunta $f_{XY}(x, y)$. Entretanto, podemos determinar $E[Z]$ diretamente a partir da densidade conjunta $f_{XY}(x, y)$ usando o seguinte teorema

Teorema 3.3. Sejam duas v.a.'s X e Y com fdp conjunta $f_{XY}(x, y)$, e a v.a. Z definida por $Z = g(X, Y)$. Então o valor médio de Z é dado por

$$E[Z] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(X, Y) f_{XY}(x, y) dx dy \quad (3.12)$$

Demonstração. A prova desta relação é similar à da equação (3.9). Se a variável Z está no intervalo $[z, z + \Delta z]$, então as variáveis X e Y estão na região limitada por $[x, x + \Delta x]$ e $[y, y + \Delta y]$. A área desta região é obviamente $\Delta x \Delta y$. Segue também que

$$f_Z(z)\Delta z = f_{XY}(x, y)\Delta x\Delta y \quad (3.13)$$

Integrando ambos os lados de (3.13), chegamos à equação (3.12).

□

Teorema 3.4. *Para v.a.'s discretas, (3.12) pode ser reescrita como*

$$E[Z] = \sum_i \sum_j g(x_i, y_j) p_{XY}(x_i, y_j) \quad (3.14)$$

Podemos estender facilmente a equação (3.12) para o caso de uma função de n v.a.'s:

Corolário 3.5. *Seja Z uma v.a. que é função de n v.a.'s X_1, \dots, X_n :*

$$Z = g(X_1, \dots, X_n)$$

Então a média de Z é dada por

$$E[Z] = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} g(X_1, \dots, X_n) f_{X_1, \dots, X_n}(x_1, \dots, x_n) dx_1 \dots dx_n \quad (3.15)$$

Se algumas das v.a.'s são discretas, a equação (3.15) ainda é válida desde que a distribuição discreta é considerada um caso limite da distribuição contínua através do uso da função impulso.

3.1.4 Média da Soma de Funções

Teorema 3.6. *Se $g_1(X, Y), \dots, g_n(X, Y)$ são funções de X e Y então*

$$E[g_1(X, Y) + \cdots + g_n(X, Y)] = E[g_1(X, Y)] + \cdots + E[g_n(X, Y)] \quad (3.16)$$

A prova é trivial e segue diretamente da definição das médias. Então a média da soma é igual à soma das médias. Alguns exemplos simples disto são

$$E[X + Y] = E[X] + E[Y] \quad (3.17)$$

$$E[X^2 + Y^2] = E[X^2] + E[Y^2] \quad (3.18)$$

Estes resultados podem ser estendidos a funções de qualquer número de v.a.'s.

3.1.5 Média do Produto de Duas Variáveis Aleatórias Independentes

Teorema 3.7. *Para v.a.'s independentes a média do produto é igual ao produto das médias individuais.*

Demonstração. Se $Z = XY$

$$E[Z] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xy f_{XY}(x, y) dx dy \quad (3.19)$$

Se X e Y são independentes $f_{XY}(x, y) = f_X(x)f_Y(y)$, e desta forma podemos escrever

$$E[Z] = \int_{-\infty}^{+\infty} x f_X(x) dx \int_{-\infty}^{+\infty} y f_Y(y) dy = E[X]E[Y] \quad (3.20)$$

Então, se X e Y são v.a.'s independentes

$$E[XY] = E[X]E[Y] \quad (3.21)$$

□

Este resultado pode ser estendido a qualquer número de variáveis. Na verdade, a equação (3.21) é um caso especial de um resultado mais geral:

Teorema 3.8. *Se X e Y são independentes, então para $Z = g_1(X)g_2(Y)$ temos que*

$$E[Z] = E[g_1(X)]E[g_2(Y)] \quad (3.22)$$

Em outras palavras

$$E[g_1(X)g_2(Y)] = E[g_1(X)]E[g_2(Y)] \quad (3.23)$$

3.1.6 Média Quadrática da Soma de Duas Variáveis Aleatórias

O valor quadrático médio de $Z = X + Y$ é dado por

$$\begin{aligned} E[Z^2] &= E[(X + Y)^2] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x + y)^2 f_{XY}(x, y) dx dy \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x^2 f_{XY}(x, y) dx dy + \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} y^2 f_{XY}(x, y) dx dy \\ &\quad + 2 \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xy f_{XY}(x, y) dx dy \end{aligned} \quad (3.24)$$

Se as v.a.'s X e Y são independentes, então

$$\begin{aligned} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x^2 f_{XY}(x, y) dx dy &= \int_{-\infty}^{+\infty} x^2 f_X(x) dx \int_{-\infty}^{+\infty} f_Y(y) dy = \\ &= \int_{-\infty}^{+\infty} x^2 f_X(x) dx = E[X^2] \end{aligned}$$

Similarmente

$$\begin{aligned} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} y^2 f_{XY}(x, y) dx dy &= \int_{-\infty}^{+\infty} f_X(x) dx \int_{-\infty}^{+\infty} y^2 f_Y(y) dy = \\ &= \int_{-\infty}^{+\infty} y^2 f_Y(y) dy = E[Y^2] \end{aligned}$$

E usando (3.21), podemos escrever

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xy f_{XY}(x, y) dx dy = E[XY] = E[X]E[Y]$$

Portanto, para as v.a.'s independentes X e Y temos

$$E[(X + Y)^2] = E[X^2] + E[Y^2] + 2E[X]E[Y] \quad (3.25)$$

Se $E[X]$ ou $E[Y]$ ou ambos forem zero, então temos

$$E[(X + Y)^2] = E[X^2] + E[Y^2] \quad (3.26)$$

3.1.7 Média condicional

Definição 3.2. A *média condicional* (ou *valor esperado condicional*) de uma v.a. X dado que outra v.a. $Y = y$ é denotada por $E[X|Y = y]$ e é definida como

$$E[X|Y = y] = \int_{-\infty}^{+\infty} x f_X(x|Y = y) dx \quad (3.27)$$

Isto segue da definição básica da média.

3.2 Momentos

3.2.1 N -ésimo momento

Definição 3.3. O *n -ésimo momento* de uma v.a. X é definido como o valor esperado da n -ésima potência de X .

$$E[X^n] = \int_{-\infty}^{+\infty} x^n f_X(x) dx \quad (3.28)$$

3.2.2 Momentos Centrais

Definição 3.4. O *n*-ésimo momento central da v.a. X é seu momento ao redor de seu valor médio $E[X]$, e é dado por

$$E[(X - E[X])^n] = \int_{-\infty}^{+\infty} (x - E[X])^n f_X(x) dx \quad (3.29)$$

3.2.3 Variância

Definição 3.5. O segundo momento central sobre a média é chamado de **variância** e é denotado por σ_X^2 .

$$\sigma_X^2 = E[(X - E[X])^2] \quad (3.30)$$

Das expressões da seção 3.1.4, segue que

$$\sigma_X^2 = E[X^2] - 2E[X]E[X] + E[E^2[X]] = E[X^2] - 2E^2[X] + E^2[X] = E[X^2] - E^2[X] \quad (3.31)$$

Então a variância de uma v.a. é igual à sua média quadrática menos o quadrado de sua média.

Exemplo 3.3. Encontre a variância σ_X^2 de uma variável aleatória X com distribuição gaussiana.

Solução. No Exemplo 3.1 vimos que $E[X] = m$, e no Exemplo 3.2, foi mostrado que $E[X^2] = \sigma^2 + m^2$.

Desta forma, pela equação (3.31) temos que $\sigma_X^2 = \sigma^2 + m^2 - m^2 = \sigma^2$.

A variância tem uma grande importância principalmente na análise de sinais, pois está intimamente ligada à potência dos mesmos (na verdade, ela corresponde à potência de um sinal de média nula). Nos teoremas a seguir são derivadas algumas propriedades importantes da variância.

Teorema 3.9. Se X sempre toma o valor a , então $\text{Var}[X] = 0$.

Demonstração. Desde que X sempre toma o valor a , $P[X = a] = 1$. Neste caso, $E[X] = a$, e $\text{Var}[X] = (a - a)^2 P[X = a] = 0$. \square

Este teorema diz que a variância de X é zero quando X é determinística.

Teorema 3.10. *Se $Y = X + b$, então $\text{Var}[Y] = \text{Var}[X]$.*

Demonstração. Dada a variável aleatória X , $E[Y] = E[X] + b$, e desta forma, a variância de Y é dada por

$$\begin{aligned}\text{Var}[Y] &= E[(X + b) - (E[X] + b)]^2 \\ &= E[(X - E[X])^2] = \text{Var}[X]\end{aligned}$$

□

Ou seja, o deslocamento da variável aleatória X de uma constante não muda a sua variância.

Teorema 3.11. *Se $Y = aX$, então $\text{Var}[Y] = a^2 \text{Var}[X]$.*

Demonstração. Desde que $E[Y] = aE[X]$, temos

$$\text{Var}[Y] = E[(aX - aE[X])^2] = E[a^2(X - E[X])^2] = a^2 \text{Var}[X].$$

□

3.2.4 Caso Multidimensional

Definição 3.6. *Sejam duas v.a.'s X_1 e X_2 com fdp conjunta $f_{X_1 X_2}(x_1, x_2)$. O **momento conjunto** é definido como*

$$E[X_1^k X_2^n] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x_1^k x_2^n f_{X_1 X_2}(x_1, x_2) dx_1 dx_2 \quad (3.32)$$

Definição 3.7. *Sejam duas v.a.'s X_1 e X_2 com fdp conjunta $f_{X_1 X_2}(x_1, x_2)$. O **momento central conjunto** é definido como*

$$E[(X_1 - m_1)^k (X_2 - m_2)^n] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x_1 - m_1)^k (x_2 - m_2)^n f_{X_1 X_2}(x_1, x_2) dx_1 dx_2 \quad (3.33)$$

onde $m_i = E[X_i]$.

Uma propriedade bastante útil quando lidamos com distribuições bidimensionais é a desigualdade de Cauchy-Schwarz, que é apresentada a seguir

Teorema 3.12. Desigualdade de Cauchy-Schwarz. *Sejam X e Y duas variáveis aleatórias. Então*

$$[E(XY)]^2 \leq E[X^2] E[Y^2] \quad (3.34)$$

Demonstração. Considere a expressão $E[(X - \alpha Y)^2]$ para duas variáveis aleatórias X e Y quaisquer, e uma variável real α . Esta expressão, quando vista como um quadrado em α , é sempre não negativa, isto é:

$$E[(X - \alpha Y)^2] \geq 0$$

Expandindo o quadrado, temos

$$E[X^2] - 2\alpha E[XY] + \alpha^2 E[Y^2] \geq 0$$

Vamos escolher agora o valor de α de modo que o lado esquerdo da equação acima seja mínimo:

$$\alpha = \frac{E[XY]}{E[Y^2]}$$

o que resulta na desigualdade

$$E[X^2] - \frac{[E(XY)]^2}{E[Y^2]} \geq 0 \Rightarrow [E(XY)]^2 \leq E[X^2] E[Y^2]$$

□

De particular importância para nós são os momentos conjuntos e momentos centrais conjuntos correspondentes a $k = n = 1$. Estes momentos conjuntos são chamados de **correlação** e **covariância** das v.a.'s X_1 e X_2 , respectivamente, e serão estudados com mais detalhes adiante.

Para v.a.'s multidimensionais podemos definir momentos conjuntos de qualquer ordem. Entretanto, os momentos que são mais úteis em aplicações práticas são as correlações e covariâncias entre pares de v.a.'s. Suponha que $X_i, i = 1, 2, \dots, n$ são v.a.'s com fdp conjunta $f_{X_1 X_2 \dots X_n}(x_1, x_2, \dots, x_n)$. Seja $f_{X_i X_j}(x_i, x_j)$ a fdp conjunta das v.a.'s X_i e X_j .

Definição 3.8. *A correlação entre duas variáveis aleatórias X_i e X_j é dada pelo momento conjunto*

$$\rho_{ij} = E[X_i X_j] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x_i x_j f_{X_i X_j}(x_i, x_j) dx_i dx_j \quad (3.35)$$

Definição 3.9. A **covariância** de duas variáveis aleatórias X_i e X_j , cujas médias são, respectivamente, m_i e m_j , é dada por

$$\begin{aligned}
 K_{ij} &= E[(X_i - m_i)(X_j - m_j)] \\
 &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x_i - m_i)(x_j - m_j) f_{X_i X_j}(x_i, x_j) dx_i dx_j \\
 &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x_i x_j f_{X_i X_j}(x_i, x_j) dx_i dx_j - m_i m_j \\
 &= E[X_i X_j] - m_i m_j
 \end{aligned} \tag{3.36}$$

As matrizes $n \times n$ com elementos ρ_{ij} e μ_{ij} são chamadas respectivamente de **matriz de correlação** e **matriz de covariância** das v.a.'s $X_i, i = 1, 2, \dots, n$.

3.2.5 Variáveis Aleatórias Descorrelacionadas e Ortogonais

Definição 3.10. Duas v.a.'s X_i e X_j são ditas **descorrelacionadas** se

$$E[X_i X_j] = E[X_i]E[X_j] = m_i m_j$$

Neste caso, a covariância $K_{ij} = 0$. Note que quando X_i e X_j são estatisticamente independentes, também são descorrelacionadas. Entretanto, se X_i e X_j são descorrelacionadas, não são necessariamente estatisticamente independentes.

Definição 3.11. Duas v.a.'s X_i e X_j são ditas **ortogonais** se $E[X_i X_j] = 0$.

Esta condição acontece quando X_i e X_j são descorrelacionadas e uma ou ambas as v.a.'s tem média zero.

3.3 Funções Características

Definição 3.12. A *função característica* de uma v.a. X é definida como a média estatística

$$\Psi(j\omega) \equiv E[e^{j\omega X}] = \int_{-\infty}^{+\infty} e^{j\omega x} f_X(x) dx \quad (3.37)$$

onde a variável ω é real e $j = \sqrt{-1}$ é a constante imaginária.

$\Psi(j\omega)$ pode ser vista como a transformada de Fourier da fdp $f_X(x)$. Assim, a transformada inversa de Fourier é dada por

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \Psi(j\omega) e^{-j\omega x} d\omega \quad (3.38)$$

Uma propriedade útil da função característica é sua relação com os momentos da v.a. O seguinte teorema relaciona estas duas grandezas:

Teorema 3.13. Sejam uma variável aleatória X e sua correspondente função característica $\Psi(j\omega)$. Então

$$E[X^n] = (-j)^n \frac{d^n \Psi(j\omega)}{d\omega^n} \Big|_{\omega=0} \quad (3.39)$$

Demonstração. A derivada primeira de $\Psi(j\omega)$ em relação a ω leva a

$$\frac{d\Psi(j\omega)}{d\omega} = j \int_{-\infty}^{+\infty} x e^{j\omega x} f_X(x) dx \quad (3.40)$$

Avaliando a expressão acima em $\omega = 0$, obtemos o primeiro momento (média)

$$E[X] = m_X = -j \frac{d\Psi(j\omega)}{d\omega} \Big|_{\omega=0} \quad (3.41)$$

O processo de diferenciação pode ser repetido, de modo que a n -ésima derivada de $\Psi(j\omega)$ avaliada em $\omega = 0$ leva ao n -ésimo momento

$$E[X^n] = (-j)^n \frac{d^n \Psi(j\omega)}{d\omega^n} \Big|_{\omega=0} \quad (3.42)$$

□

Então os momentos de uma v.a. podem ser determinados a partir da função característica. Por outro lado suponha que a função característica possa ser expandida em uma série de Taylor sobre o ponto $\omega = 0$, isto é

$$\Psi(j\omega) = \sum_{n=0}^{\infty} \left[\frac{d^n \Psi(j\omega)}{d\omega^n} \right]_{\omega=0} \frac{\omega^n}{n!} \quad (3.43)$$

Usando a relação em (3.42) para eliminar a derivada em (3.43), obtemos uma expressão para a função característica em termos de seus momentos na forma

$$\Psi(j\omega) = \sum_{n=0}^{\infty} E[X^n] \frac{(j\omega)^n}{n!} \quad (3.44)$$

A função característica fornece um meio simples de determinar a fdp da soma de v.a.'s estatisticamente independentes:

Teorema 3.14. *Seja $X_i, i = 1, 2, \dots, x_n$ um conjunto de n v.a.'s estatisticamente independentes e seja*

$$Y = \sum_{i=1}^n X_i$$

Então a função característica de Y é dada por

$$\Psi_Y(j\omega) = \prod_{i=1}^n \Psi_{X_i}(j\omega)$$

Demonstração. O problema consiste em determinar a fdp de Y . Iremos fazer isto encontrando primeiro a sua função característica a então calculando a transformada inversa de Fourier.

$$\begin{aligned} \Psi_Y(j\omega) &= E[e^{j\omega Y}] \\ &= E \left[\exp \left(j\omega \sum_{i=1}^n X_i \right) \right] \\ &= E \left[\prod_{i=1}^n (e^{j\omega X_i}) \right] \end{aligned} \quad (3.45)$$

Desde que as v.a.'s são estatisticamente independentes,

$$f_{X_1 X_2 \dots X_n}(x_1, x_2, \dots, x_n) = f_{X_1}(x_1) f_{X_2}(x_2) \dots f_{X_n}(x_n)$$

e desta forma a integral múltipla da equação acima pode ser fatorada em n integrais simples, cada uma correspondendo à função característica de um dos X_i . Portanto

$$\Psi_Y(j\omega) = \prod_{i=1}^n \Psi_{X_i}(j\omega) \quad (3.46)$$

□

Corolário 3.15. *Se além de independentes, as v.a.'s X_i forem identicamente distribuídas, as $\Psi_{X_i}(j\omega)$ são idênticas, e a expressão acima reduz-se a*

$$\Psi_Y(j\omega) = [\Psi_{X_i}(j\omega)]^n \quad (3.47)$$

Observações:

- A fdp de Y pode ser determinada a partir da transformada inversa de Fourier de $\Psi_Y(j\omega)$, dada pela equação (3.38).
- Desde que a função característica da soma de n v.a.'s estatisticamente independentes é igual ao produto das funções características das v.a.'s individuais $X_i, i = 1, 2, \dots, n$, segue que no domínio da transformada, a fdp de Y é a convolução das fdp's de X_i . Geralmente a convolução é mais difícil de calcular do que o método da função característica descrito acima para determinar a fdp de Y .

3.3.1 Caso multidimensional

Para lidar com v.a.'s n -dimensionais, é conveniente definir uma transformada de Fourier n -dimensional da fdp conjunta.

Definição 3.13. *Se $X_i, i = 1, 2, \dots, n$ são v.a.'s com fdp $f_{X_1 X_2 \dots X_n}(x_1, x_2, \dots, x_n)$, a função característica n -dimensional é definida como*

$$\begin{aligned} \Psi(j\omega_1, j\omega_2, \dots, j\omega_n) &= E \left[\exp \left(j \sum_{i=1}^n \omega_i X_i \right) \right] \\ &= \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} \exp \left(j \sum_{i=1}^n \omega_i X_i \right) f_{X_1 X_2 \dots X_n}(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n \end{aligned} \quad (3.48)$$

De especial interesse é a função característica bi-dimensional

$$\Psi(j\omega_1, j\omega_2) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{j(\omega_1 X_1 + \omega_2 X_2)} f_{X_1 X_2}(x_1, x_2) dx_1 dx_2 \quad (3.49)$$

Observe que as derivadas parciais de $\Psi(j\omega_1, j\omega_2)$ em relação a ω_1 e a ω_2 podem ser utilizadas para gerar os momentos conjuntos. Por exemplo, é fácil mostrar que

$$E[X_1, X_2] = - \left. \frac{\partial^2 \Psi(j\omega_1, j\omega_2)}{\partial \omega_1 \partial \omega_2} \right|_{\omega_1 = \omega_2 = 0} \quad (3.50)$$

3.4 Exercícios

1. Se $F_X(\omega)$ é a transformada de Fourier de uma função densidade de probabilidade $f_X(x)$ e m_n representa o n -ésimo momento da v.a. X ,

$$m_n = \int_{-\infty}^{+\infty} x^n f_X(x) dx$$

Então mostre que

(a)

$$m_n = (-j)^n \left. \frac{d^n F_X(\omega)}{d\omega^n} \right|_{\omega=0}$$

(b) se $F_X(\omega)$ é expandida em série de Taylor, então

$$F_X(\omega) = m_0 - jm_1\omega - \frac{m_2\omega^2}{2!} + j\frac{m_3\omega^3}{3!} + \cdots = \sum_{n=0}^{\infty} (-j)^n m_n \left(\frac{\omega^n}{n!} \right)$$

2. Use os resultados do problema anterior para determinar o valor médio e o valor quadrático médio de

(a) um sinal gaussiano;

(b) um sinal com $f_X(x) = xe^{-x}u(x)$

Dica: encontre $F_X(\omega)$ e expanda em séries de potências como no problema anterior. O segundo e terceiro coeficientes representam os valores médios e quadrático médio, respectivamente.

Resp:

(a) $E[X] = m \quad E[X^2] = \sigma^2 + m^2$

(b) $E[X] = 2 \quad E[X^2] = 6$

3. Seja X uma v.a. com média μ e desvio padrão $\sigma > 0$, e seja X^* a v.a. padronizada correspondente de X , isto é $X^* = (X - \mu)/\sigma$. Mostre que $E[X^*] = 0$ e $\text{Var}[X^*] = 1$. (Logo $\sigma_{X^*} = 1$).

4. Encontre o n -ésimo momento de X , se X é uma v.a. uniformemente distribuída no intervalo $[a, b]$.

Resp: $E[X^n] = \frac{b^{n+1} - a^{n+1}}{(b-a)(n+1)}$

5. Encontre a média e a variância de uma v.a. gaussiana aplicando o teorema dos momentos sobre a função característica.

6. Dado que

$$m = \int_{-\infty}^{\infty} x f_x(x) dx \quad E[X^2] = \int_{-\infty}^{\infty} x^2 f_x(x) dx \quad \sigma_X^2 = \int_{-\infty}^{\infty} (x - m)^2 f_x(x) dx$$

Mostre que $\sigma_X^2 = E[X^2] - m^2$

7. Encontre a função característica de uma variável aleatória X com distribuição de Cauchy, cuja função densidade de probabilidade é dada por

$$f_X(x) = \frac{a}{\pi(x^2 + a^2)}, \quad -\infty < x < \infty$$

Resp: $\Psi(j\omega) = e^{-a\omega}$

8. Seja $Y = a \cos(\omega t + \Theta)$ onde a , ω , e t são constantes, e Θ é uma variável aleatória com distribuição uniforme no intervalo $(0, 2\pi)$. A variável aleatória Y resulta na amostragem da amplitude de uma senóide com fase aleatória Θ . Encontre $E[Y]$ e $E[Y^2]$.

Resp: $E[Y] = 0 \quad E[Y^2] = \frac{a^2}{2}$

9. Mostre que o primeiro e segundo momentos de uma variável aleatória com distribuição χ_n^2 são respectivamente n e $(n^2 + 2n)$, aplicando o teorema dos momentos sobre a função característica.

A fdp de uma distribuição χ_n^2 é dada pela expressão

$$f_Y(y) = \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} y^{\frac{(n-1)}{2}} e^{-\frac{y}{2}}, \quad y \geq 0$$

onde $\Gamma(p)$ é a função gama, definida por

$$\Gamma(p) = \int_0^\infty t^{p-1} e^{-t} dt, \quad p > 0$$

$$\Gamma(p+1) = p \Gamma(p)$$

Dica: faça $u = y/2$

10. Determine os momentos de uma variável aleatória X com distribuição $N(0, 1)$.

Resp: $E[X^n] = \begin{cases} 0 & n = 1, 3, 5, 7, \dots \\ 1 \cdot 3 \cdot 5 \cdots (n-1) & n = 2, 4, 6, 8, \dots \end{cases}$

11. Dada uma variável aleatória discreta ξ que assume dois valores 0 e 1 com probabilidades p e q , respectivamente, prove que $\sigma_\xi^2 \leq 0,25$. Encontre o valor para o qual $\sigma_\xi^2 = 0,25$.

Resp: $q = 0,5$

12. Sabe-se que para uma variável aleatória X positiva, o segundo e o quarto momentos são dados por $2\sigma^2$ e $8\sigma^4$, respectivamente. Se $Y = X^2$, determine a média e a variância de Y .

Resp: $E[Y] = 2\sigma^2 \quad \text{Var}[Y] = 4\sigma^4$.

13. Se uma variável aleatória X tem fmp dada por

$$p_X(x_k) = \begin{cases} 0,5 & x = -1 \\ 0,5 & x = +1 \\ 0 & \text{caso contrário} \end{cases}$$

mostre que a função característica de X é dada por $\cos(\omega)$.

14. Demonstre a consistência da definição da função característica. Faça as suposições necessárias para a demonstração.

Dica: unicidade das transformadas, função impulso, e propriedade de deslocamento no domínio do tempo

15. Seja η a média de uma variável aleatória X . Mostre que se

$$\int_{-\infty}^{\eta} f_X(x) dx = \int_{\eta}^{\infty} f_X(x) dx$$

então $F_X(\eta) = 1/2$.

16. Seja (X, Y) uma v.a. bidimensional com fdp conjunta

$$f_{XY}(x, y) = \frac{x^2 + y^2}{4\pi} e^{-(x^2 + y^2)/2}, -\infty < x < \infty, -\infty < y < \infty$$

Mostre que X e Y são descorrelacionadas mas não independentes.

17. Seja X uma variável aleatória $N(0, \sigma^2)$.

- (a) Calcule $f_X(x|X > 0)$
- (b) Calcule $E[X|X > 0]$
- (c) Calcule $Var[X|X > 0]$

Resp:

$$(a) f_X(x|X > 0) = \begin{cases} 0 & x < 0 \\ 2 \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} & x \geq 0 \end{cases}$$

$$(b) \sqrt{\frac{2}{\pi}} \sigma$$

$$(c) \sigma^2 \left(1 - \frac{2}{\pi}\right) \approx 0,363\sigma^2$$

18. Suponha que a fmp conjunta de uma variável aleatória bidimensional (X, Y) seja dada por

$$p_{XY}(x, y) = \begin{cases} 1/3 & (0, 1), (1, 0), (2, 1) \\ 0 & \text{caso contrário} \end{cases}$$

- (a) Encontre as fmps marginais.
- (b) X e Y são independentes?
- (c) X e Y são descorrelacionadas?

Resp:

$$(a) \ p_X(x) = \begin{cases} 1/3 & x = 0, 1, 2 \\ 0 & \text{caso contrário} \end{cases}$$

$$p_Y(y) = \begin{cases} 1/3 & x = 0 \\ 2/3 & x = 1 \\ 0 & \text{caso contrário} \end{cases}$$

- (b) não
- (c) sim

Capítulo 4

Métodos computacionais para geração de números aleatórios

4.1 Introdução

Em simulações de sistemas reais às vezes nos deparamos com a necessidade de gerar números aleatórios segundo alguma distribuição para testar nossas idéias. Por exemplo, se queremos simular um canal de comunicação ruidoso, devemos gerar números aleatórios segundo uma distribuição gaussiana de média zero e variância igual à potência do ruído de canal. Por outro lado, se queremos simular o tráfego de dados em um determinado serviço, devemos gerar números com distribuição exponencial para o tempo entre chegadas de clientes.

Neste capítulo serão apresentados alguns algoritmos computacionais para a geração de números de forma aleatória, segundo uma dada distribuição. Inicialmente será apresentado o algoritmo para a geração de números com distribuição uniforme entre 0 e 1, que irá servir de base para os demais algoritmos.

4.2 Método do resíduo da potência

O primeiro problema a ser abordado quando queremos gerar números aleatórios no intervalo $[0, 1]$ é que existem infinitos pontos dentro deste intervalo, mas o computador só pode representar números com precisão finita. Precisamos nos contentar então em gerar números de forma equiprovável dentro de um conjunto limitado, por exemplo $\{0, 1, \dots, M-1\}$ ou $\{1, 2, \dots, M\}$. Dividindo estes números por M , obtemos números no intervalo unitário. Podemos gerar distribuições bastante densas se fizermos M bastante grande.

O próximo passo consiste em encontrar um mecanismo para gerar números de forma aleatória. A forma preferida para gerar números aleatória através do computador é através de fórmulas recursivas que possam ser implementadas de forma fácil e rápida. No método do resíduo da potência utiliza-se a seguinte fórmula:

$$Z_k = \alpha Z_{k-1} \mod M \quad (4.1)$$

onde α é um inteiro entre 0 e M , e M é um número primo (p) ou uma potência inteira de um número primo (p^m).

Exemplo 4.1. *Encontre as sequências geradas pela Equação (4.1) para:*

1. $M = 11, \alpha = 7, Z_0 = 1$
2. $M = 11, \alpha = 3, Z_0 = 1$
3. $M = 2^2, \alpha = 7, Z_0 = 1$

Solução. Usando (4.1), temos:

1. Para $M = 11, \alpha = 7$ e $Z_0 = 1$, temos:

$$Z_1 = \text{resto de } \frac{7 \times 1}{11} = 7$$

$$Z_2 = \text{resto de } \frac{7 \times Z_1}{11} = \text{resto de } \frac{7 \times 7}{11} = \text{resto de } \frac{49}{11} = 5$$

e assim por diante. A sequência resultante é:

$$1, 7, 5, 2, 3, 10, 4, 6, 9, 8, 1, 7, 5, 2, 3, 10, 4, 6, 9, 8, 1, 7, 5, 2, 3, 10, 4, 6, 9, 8, \dots$$

Note que a sequência passa por todos os inteiros de 1 a 10, e então passa a se repetir indefinidamente.

2. Para este caso, a sequência gerada é:

$$1, 3, 9, 5, 4, 1, 3, 9, 5, 4, 1, 3, 9, 5, 4, \dots$$

Esta sequência não passa por todos os inteiros de 1 a 10 antes de começar a se repetir.

3. Para o último caso, a sequência gerada é:

$$1, 2, 0, 0, 0, \dots$$

Do Exemplo acima, podemos notar que a escolha de α influi diretamente na sequência gerada: se α é divisor de M , então a sequência gerada pela Equação (4.1) irá eventualmente ser toda nula; caso contrário, a sequência será periódica com período máximo $M - 1$. Para que a sequência tenha o máximo comprimento possível, α deve ser uma raiz primitiva de M , um conceito cujo estudo está fora do escopo deste texto.

Uma coisa a ser notada sobre este algoritmo é que as sequências produzidas pela Equação (4.1) não são realmente aleatórias, mas sim periódicas. Por esta razão, as sequências produzidas por (4.1) são chamadas de **pseudo-aleatórias**.

Se fizermos M grande o suficiente, então os números gerados não irão se repetir durante uma dada simulação, e a sequência gerada tem a aparência de uma sequência aleatória.

Vários estudos foram feitos para determinar bons valores para M e α . Uma combinação que é bastante usada é:

$$Z_i = 7^5 Z_{i-1} \mod (2^{31} - 1) \quad (4.2)$$

ou seja, $\alpha = 7^5 = 16807$ e $M = 2^{31} - 1$. Esta combinação gera sequências pseudo-aleatórias de comprimento $M - 1 = 2^{31} - 1 - 1 = 2147483646$ elementos, o que é mais que suficiente para a maioria das aplicações.

A escolha de Z_0 determina o ponto em que a sequência irá se iniciar, e por isso, este parâmetro é conhecido como a “semente” do gerador aleatório.

Nas seções a seguir, iremos descrever algoritmos para gerar sequências de números aleatórios com outras distribuições de probabilidade a partir das sequências geradas nesta seção.

4.3 Método da transformada

Suponha que U seja uniformemente distribuída no intervalo $[0, 1]$. Seja $F_X(x)$ a fdc de uma variável aleatória que estamos interessados em gerar. Vamos definir a variável aleatória $Z = F_X^{-1}(U)$; isto é, primeiro selecionamos U e depois encontramos Z , como indicado na Figura 4.1. A fdc da variável Z encontrada desta maneira é dada por:

$$F_Z(z) = P[Z \leq x] = P[F_X^{-1}(U) \leq x] = P[U \leq F_X(x)]$$

Mas se U é uniformemente distribuída em $[0, 1]$ e $0 \leq h \leq 1$, então $P[U \leq h] = h$. Então:

$$P[Z \leq x] = F_X(x)$$

e $Z = F_X^{-1}(U)$ tem a fdc desejada.

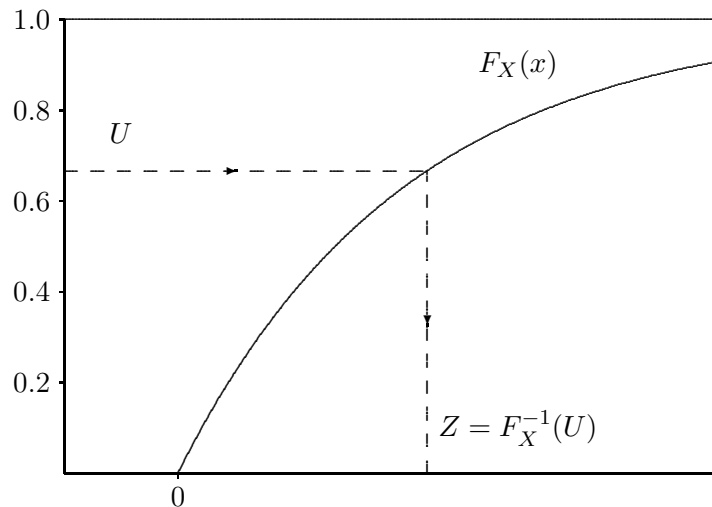


Figura 4.1: Método da transformada para gerar uma variável aleatória com fdc $F_X(x)$.

Método da transformada para gerar X

1. Gere U com distribuição uniforme no intervalo $[0, 1]$.
2. Faça $X = F_X^{-1}(U)$

Exemplo 4.2. *Determine X para gerar uma sequência de números aleatórios com distribuição exponencial de parâmetro λ a partir de uma sequência de números aleatórios uniformemente distribuídos no intervalo $[0, 1]$.*

Solução. Precisamos inverter a expressão $u = F_X(x) = 1 - e^{-\lambda x}$. Com isto, obtemos

$$X = -\frac{1}{\lambda} \ln(1 - U)$$

Note que podemos usar a expressão mais simples $X = -\ln(U)/\lambda$, desde que $(1 - U)$ também é uniformemente distribuída no intervalo $[0, 1]$.

Exemplo 4.3. *Para gerar uma variável aleatória com distribuição de Bernoulli de probabilidade de sucesso p , notamos da Figura 4.2 que*

$$X = \begin{cases} 0, & U \leq p \\ 1, & U > p \end{cases}$$

Em outras palavras, particionamos o intervalo $[0, 1]$ em dois segmentos de comprimentos p e $1 - p$, respectivamente. A saída X é determinada pelo intervalo em que U cair.

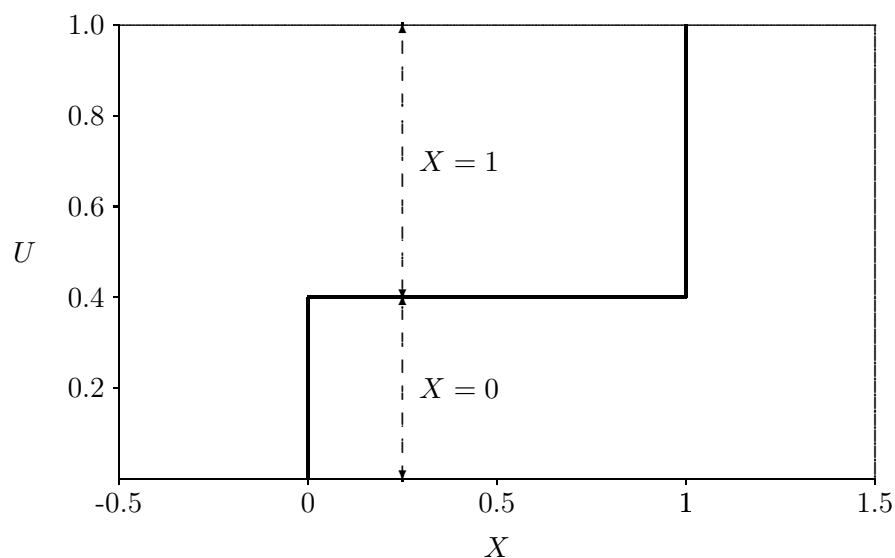


Figura 4.2: Gerando uma variável aleatória com distribuição de Bernoulli.

Exemplo 4.4. Gere uma variável aleatória com distribuição binomial de parâmetros $n = 5$ e $p = 1/2$.

Solução. Para gerar uma variável aleatória com distribuição binomial de parâmetros $n = 5$ e $p = 1/2$, poderíamos simplesmente gerar cinco variáveis aleatórias com distribuição de Bernoulli e assumir Y como sendo o número total de sucessos.

Alternativamente, podemos usar diretamente o método da transformada, como mostrado na Figura 4.3. O intervalo unitário é agora particionado em seis elementos. A eficiência do algoritmo de partição depende da ordem na qual fazemos a busca. Por exemplo, se fazemos a busca nos segmentos em ordem (de 0 a 5), serão necessárias em média 3.5 comparações para cada número gerado. Se fizermos a busca nos segmentos em ordem decrescente de probabilidade, então o número médio de comparações cai para 2.38.

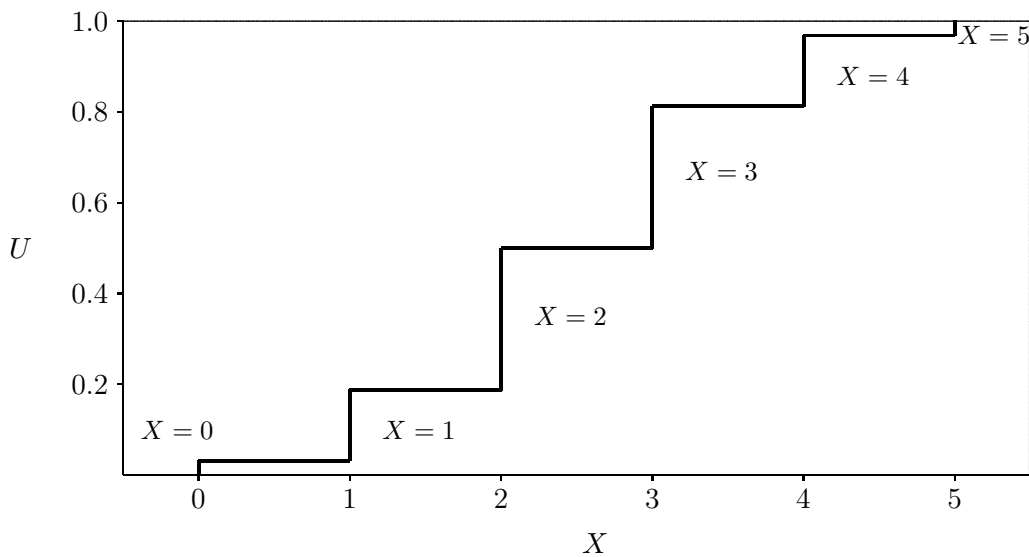


Figura 4.3: Gerando uma variável aleatória com distribuição Binomial.

Claramente qualquer variável aleatória finita discreta pode ser gerada dividindo-se o intervalo unitário em subintervalos com comprimentos determinados pela fmp. O próximo método é baseado na fdp ao invés da fdc de Z .

4.4 O método da rejeição

Iremos considerar uma versão simplificada deste algoritmo para explicar porque ele funciona. Depois o algoritmo será rerepresentado em sua forma geral.

Suponha que estamos interessados em gerar uma variável aleatória Z com fdp $f_X(x)$, como mostrado na Figura 4.4. Em particular, assumimos que:

- a fdp é não nula somente no intervalo $[0, a]$;
- a fdp assume valores no intervalo $[0, b]$.

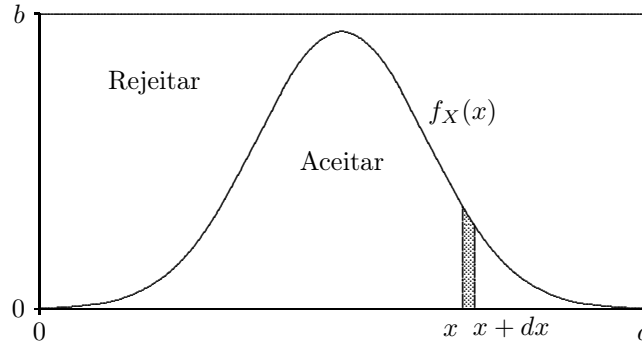


Figura 4.4: Método da rejeição para gerar uma variável aleatória com fdp $f_X(x)$.

O método da rejeição neste caso funciona da seguinte maneira:

1. Gere X_1 com distribuição uniforme no intervalo $[0, a]$.
2. Gere Y com distribuição uniforme no intervalo $[0, b]$.
3. Se $Y \leq f_X(X_1)$, então $Z = X_1$; senão, rejeite X_1 e retorne ao passo 1.

Note que este algoritmo irá realizar um número aleatório de passos antes de produzir a saída Z .

Iremos mostrar agora que a saída Z tem a fdp desejada: os passos 1 e 2 selecionam aleatoriamente um ponto em um retângulo de largura a e altura b . A probabilidade de selecionar um ponto em qualquer região é simplesmente a área da região dividida pela área total do retângulo, ab . Então a probabilidade de aceitar X_1 é a área da região abaixo de $f_X(x)$ dividida por ab . Mas a área sob qualquer fdp é 1, de modo que concluímos que a probabilidade de sucesso é $1/(ab)$. Considere agora a seguinte probabilidade:

$$\begin{aligned}
 P[\{x < X_1 \leq x + dx | X_1 \text{ ser aceito}\}] &= \frac{P[\{x < X_1 \leq x + dx\}, \{X_1 \text{ ser aceito}\}]}{P[X_1 \text{ ser aceito}]} \\
 &= \frac{\text{área sombreada}/(ab)}{1/(ab)} = \frac{f_X(x) dx/(ab)}{1/(ab)} \\
 &= f_X(x) dx
 \end{aligned}$$

Então, X_1 , quando aceito, tem a fdp desejada, e portanto Z tem a fdp desejada.

O algoritmo acima pode apresentar dois problemas: primeiro, se a diferença entre o retângulo e a fdp a ser gerada for muito grande, então o número de X_1 's que devem ser gerados antes da aceitação pode ser excessivamente alto; segundo, este método não pode ser utilizado se $f_X(x)$ não é limitada, ou se seu contradomínio não é limitado.

A versão geral deste algoritmo resolve estes dois problemas: suponha que queremos gerar X com fdp $f_X(x)$. Seja W uma variável aleatória com fdp $F_W(x)$ que é *fácil* de gerar, e tal que para alguma constante $K > 1$,

$$K f_W(x) \geq f_X(x), \forall x$$

ou seja, a região sob $K f_W(x)$ contém $f_X(x)$, como mostrado na Figura 4.5.

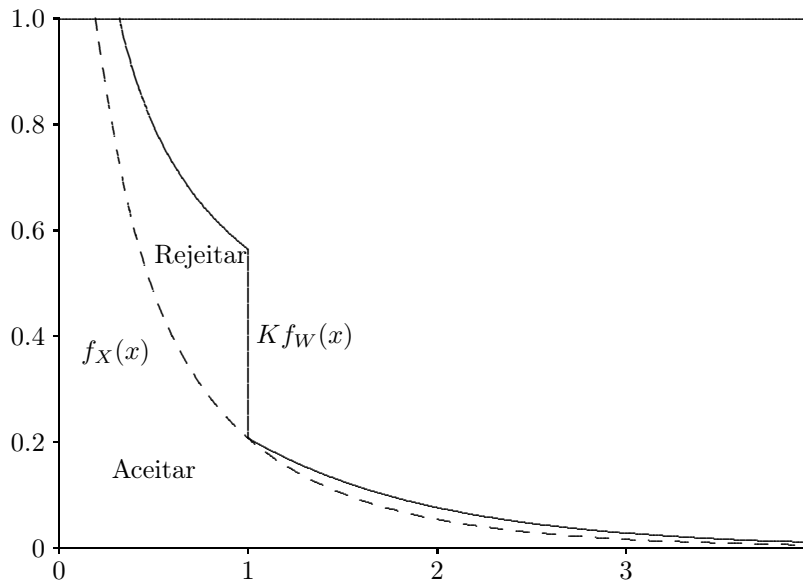


Figura 4.5: Método da rejeição para gerar uma variável aleatória com distribuição gama ($0 < \alpha < 1$).

Método da rejeição para gerar X

1. Gere X_1 com fdp $f_W(x)$. Defina $B(X_1) = K f_W(X_1)$.
2. Gere Y com distribuição uniforme no intervalo $[0, B(X_1)]$.
3. Se $Y \leq f_X(X_1)$, então $X = X_1$; senão, rejeite X_1 e retorne ao passo 1.

Exemplo 4.5. Mostre uma maneira de gerar uma variável aleatória com distribuição gama de parâmetros $0 < \alpha < 1$ e $\lambda = 1$, usando o método da rejeição.

Solução. Uma função $K f_W(x)$ que “cobre” $f_X(x)$ é

$$f_X(x) = \frac{x^{\alpha-1} e^{-x}}{\Gamma(\alpha)} \leq K f_W(x) = \begin{cases} \frac{x^{\alpha-1}}{\Gamma(\alpha)}, & 0 \leq x \leq 1 \\ \frac{e^{-x}}{\Gamma(\alpha)}, & x > 1 \end{cases}$$

A fdp $f_W(x)$ que corresponde à função no lado direito é

$$f_W(x) = \begin{cases} \frac{\alpha e x^{\alpha-1}}{\alpha + e}, & 0 \leq x \leq 1 \\ \alpha e \frac{e^{-x}}{\alpha + e}, & x > 1 \end{cases}$$

A fdc de W é

$$F_W(x) = \begin{cases} \frac{ex^\alpha}{\alpha + e}, & 0 \leq x \leq 1 \\ 1 - \alpha e \frac{e^{-x}}{\alpha + e}, & x > 1 \end{cases}$$

W pode ser gerada facilmente usando o método da transformação com

$$F_W^{-1}(u) = \begin{cases} \left[\frac{(\alpha + e)u}{e} \right]^{1/\alpha}, & u \leq e/(\alpha + e) \\ -\ln \left[(\alpha + e) \frac{(1-u)}{\alpha e} \right], & u > e/(\alpha + e) \end{cases}$$

Podemos usar o método da transformada para gerar esta $f_W(x)$, e então o método da rejeição para gerar qualquer variável aleatória com distribuição gama de parâmetros $0 < \alpha < 1$ e $\lambda = 1$. Finalmente, note que se fizermos $W = \lambda X$, então W terá distribuição gama com parâmetros α e λ .

4.5 Geração de funções de uma variável aleatória

Se tivermos um método simples para gerar uma variável aleatória X , podemos gerar facilmente qualquer variável aleatória que seja definida por $Y = g(x)$ ou mesmo $Z = h(X_1, X_2, \dots, X_n)$, onde X_1, X_2, \dots, X_n são n saídas do gerador de números aleatórios.

Exemplo 4.6. Método Box & Muller. *Pode-se mostrar que se U_1 e U_2 são variáveis aleatórias independentes e uniformemente distribuídas no intervalo unitário, então*

$$X = \cos(2\pi U_2) \sqrt{-2 \ln(U_1) \sigma_X^2} + \mu_X$$

e

$$Y = \sin(2\pi U_2) \sqrt{-2 \ln(U_1) \sigma_Y^2} + \mu_Y$$

são variáveis aleatórias gaussianas de médias μ_X e μ_Y variâncias σ_X^2 e σ_Y^2 , respectivamente. Este resultado pode então ser utilizado para produzir duas variáveis aleatórias gaussianas a partir de duas variáveis aleatórias com distribuição uniforme.

Exemplo 4.7. *Seja X_1, X_2, \dots, X_m uma sequência de variáveis aleatórias iid com distribuição exponencial de parâmetro λ . Iremos mostrar no Capítulo 5 que a variável aleatória*

$$Y = X_1 + X_2 + \dots + X_m$$

tem uma distribuição m-Erlang com parâmetro λ . Podemos então gerar uma variável aleatória m-Erlang gerando primeiro m variáveis aleatórias com distribuição exponencial de parâmetro λ através do método da transformada, e tomando a soma destas.

Desde que a variável aleatória m -Erlang é um caso especial da variável aleatória gama, para m grande pode ser preferível usar o método da rejeição descrito anteriormente.

4.6 Geração de misturas de variáveis aleatórias

Às vezes uma variável aleatória consiste de uma mistura de várias variáveis aleatórias. Para gerar este tipo de variável aleatória podemos primeiramente selecionar uma distribuição de acordo com alguma fmp, e então gerar uma amostra da variável aleatória selecionada. Este procedimento pode ser facilmente simulado, como mostrado da seguir:

Exemplo 4.8. *Uma variável aleatória exponencial de dois estágios tem fdp*

$$f_X(x) = pae^{-ax} + (1-p)be^{-bx}$$

Fica claro da expressão acima que X consiste da mistura de duas variáveis aleatórias exponenciais com parâmetros a e b , respectivamente.

X pode ser gerada da seguinte maneira:

- *Realize um teste de Bernoulli com probabilidade de sucesso p .*
 - *Se o resultado for um sucesso, use o método da transformada para gerar uma variável aleatória exponencial de parâmetro a .*
 - *Se o resultado for um fracasso, use o método da transformada para gerar uma variável aleatória exponencial de parâmetro b .*
-

4.7 Exercícios

1. Escreva um programa de computador para implementar um gerador de números aleatórios segundo a Equação (4.2).
 - (a) Para checar seu programa, encontre Z_{1000} ; com semente $Z_0 = 1$, ele deve ser 522329230.
 - (b) Gere 10000 números aleatórios no intervalo unitário e plote o histograma. O resultado é o esperado? Justifique sua resposta.
2. Suponha que estamos interessados em utilizar arremessos de uma moeda ideal para simular um experimento aleatório no qual existem seis resultados equiprováveis, $S = \{0, 1, 2, 3, 4, 5\}$. O seguinte algoritmo é proposto:
 - 1) Jogue uma moeda ideal três vezes e obtenha um número binário, associando cara com o zero e coroa com o 1.
 - 2) Se o resultado dos arremessos do passo 1) for a representação binária de um número em S , gere o número; caso contrário, retorne ao passo 1).

Este algoritmo é uma versão simplificada do método da rejeição.

- (a) Encontre a probabilidade de um número ser gerado no passo 2).
 - (b) Mostre que os números gerados no passo 2) são equiprováveis.
 - (c) Generalize o algoritmo acima para mostrar como o arremesso de moedas pode ser usado para simular qualquer experimento aleatório com urnas.
3. Encontre a transformação necessária para gerar uma variável aleatória com distribuição de Laplace.
 4. Uma variável aleatória mista Y tem fdp dada por

$$f_Y(x) = p\delta(x) + (1 - p)f_X(x)$$

onde X é uma variável aleatória com distribuição de Laplace, e p é um número entre 0 e 1. Encontre a transformação necessária para gerar Y .

5. Especifique o método de transformação necessário para gerar uma variável aleatória com distribuição de parâmetro α (α pequeno). Calcule o número médio de comparações necessário na busca.

Capítulo 5

Somas de Variáveis Aleatórias e o Teorema do Limite Central

5.1 Introdução

Uma grande variedade de questões pode ser respondida estudando-se uma v.a. W_n , definida como a soma de n v.a.'s

$$W_n = X_1 + X_2 + \cdots + X_n \quad (5.1)$$

Pelo fato de W_n ser uma função de n v.a.'s, poderíamos utilizar as distribuições conjuntas de X_1, X_2, \dots, X_n para derivar o modelo de probabilidade completo de W_n na forma de uma fdp ou de uma fmp. Entretanto, em muitas aplicações práticas, a natureza da análise das propriedades das v.a.'s nos permite aplicar técnicas que são mais simples do que analisar um modelo de probabilidade n -dimensional.

5.2 Médias de somas

Teorema 5.1. *Para qualquer conjunto de v.a.'s X_1, X_2, \dots, X_n , o valor esperado de $W_n = X_1 + X_2 + \cdots + X_n$ é*

$$E[W_n] = E[X_1] + E[X_2] + \cdots + E[X_n]$$

Demonstração. Vamos mostrar inicialmente que $E[W_2] = E[X_1] + E[X_2]$.

Sejam $g_1(X_1, X_2) = X_1$, $g_2(X_1, X_2) = X_2$ e $g(X_1, X_2) = g_1(X_1, X_2) + g_2(X_1, X_2)$.

Usando a propriedade da média de uma função de duas variáveis aleatórias, podemos escrever (para o caso contínuo)

$$\begin{aligned} E[g(X_1, X_2)] &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(X_1, X_2) f_{X_1 X_2}(X_1, X_2) dx_1 dx_2 \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} [g_1(X_1, X_2) + g_2(X_1, X_2)] f_{X_1 X_2}(X_1, X_2) dx_1 dx_2 \end{aligned}$$

$$\begin{aligned}
&= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g_1(X_1, X_2) f_{X_1 X_2}(X_1, X_2) dx_1 dx_2 \\
&+ \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g_2(X_1, X_2) f_{X_1 X_2}(X_1, X_2) dx_1 dx_2 \\
&= E[g_1(X_1, X_2)] + E[g_2(X_1, X_2)]
\end{aligned}$$

Portanto mostramos que $E[W_2] = E[X_1 + X_2] = E[X_1] + E[X_2]$.

Assumimos agora

$$E[W_{n-1}] = E[X_1] + E[X_2] + \cdots + E[X_{n-1}]$$

Note que $W_n = W_{n-1} + X_n$. Desde que W_n é uma soma de duas v.a.'s W_{n-1} e X_n ,

$$E[W_n] = E[W_{n-1}] + E[X_n] = E[X_1] + E[X_2] + \cdots + E[X_n]$$

□

Ou seja, a esperança da soma é igual à soma das esperanças quer as v.a.'s sejam independentes ou não. Para a variância de W_n , temos

Teorema 5.2. *A variância de $W_n = X_1 + X_2 + \cdots + X_n$ é*

$$\text{Var}[W_n] = \sum_{i=1}^n \text{Var}[X_i] + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{Cov}[X_i, X_j]$$

Demonstração. Da definição de variância, podemos escrever

$$\text{Var}[W_n] = E \left[(W_n - E[W_n])^2 \right]$$

Por conveniência, chamemos $\mu_i = E[X_i]$. Desde que $W_n = \sum_{i=1}^n X_i$ e $E[W_n] = \sum_{i=1}^n \mu_i$,

$$\text{Var}[W_n] = E \left[\left(\sum_{i=1}^n (X_i - \mu_i) \right)^2 \right] = E \left[\sum_{i=1}^n (X_i - \mu_i) \sum_{j=1}^n (X_j - \mu_j) \right]$$

Separando os termos para os quais $i = j$, temos

$$\begin{aligned}
\text{Var}[W_n] &= \sum_{i=1}^n E \left[(X_i - \mu_i)^2 + \sum_{j \neq i} (X_i - \mu_i)(X_j - \mu_j) \right] \\
&= \sum_{i=1}^n \text{Var}[X_i] + \sum_{i=1}^n \sum_{j \neq i} \text{Cov}[X_i, X_j]
\end{aligned}$$

Por último, notamos que $\text{Cov}[X_i, X_j] = \text{Cov}[X_j, X_i]$, e desta forma

$$\sum_{i=1}^n \sum_{j \neq i} \text{Cov}[X_i, X_j] = 2 \sum_{i=1}^n \sum_{j=i+1}^n \text{Cov}[X_i, X_j]$$

□

Quando X_1, X_2, \dots, X_n são mutuamente independentes, os termos $\text{Cov}[X_i, X_j] = 0$ se $j \neq i$ (veja Definição 3.9), e temos o seguinte resultado

Teorema 5.3. Quando X_1, X_2, \dots, X_n são mutuamente independentes, a variância de $W_n = X_1 + X_2 + \dots + X_n$ é a soma das variâncias

$$\text{Var}[W_n] = \text{Var}[X_1] + \text{Var}[X_2] + \dots + \text{Var}[X_n]$$

Exemplo 5.1. A entrada de um filtro digital é uma sequência aleatória $X_n = X_0, X_1, X_2, \dots$

O valor esperado de X_n é a função $\mu_X(n) = 0, \forall n$. A função de covariância de X_n é $C_X[X_n, X_k] = C_X[n - k] = 0, 8^{|n-k|}$. A saída do filtro é uma sequência aleatória Y_n , relacionada a X_n por

$$Y_n = X_n + X_{n-1} + X_{n-2}, \text{ para todo } n \text{ inteiro}$$

Qual é a variância de Y_n ?

Solução. Aplicando o teorema 5.2 obtemos para cada i ,

$$\begin{aligned} \text{Var}[Y_i] &= \text{Var}[X_i] + \text{Var}[X_{i-1}] + \text{Var}[X_{i-2}] + 2 \text{Cov}[X_i, X_{i-1}] \\ &\quad + 2 \text{Cov}[X_i, X_{i-2}] + 2 \text{Cov}[X_{i-1}, X_{i-2}] \end{aligned}$$

Desde que $\text{Var}[X_j] = C_X[0]$ e $\text{Cov}[X_i, X_j] = C_X[i - j]$,

$$\text{Var}[Y_i] = 3C_X[0] + 4C_X[1] + 2C_X[2] = 3 \times 0,8^0 + 4 \times 0,8^1 + 2 \times 0,8^2 = 7,48$$

A mesma estratégia pode ser utilizada para encontrar as propriedades de filtros digitais mais complexos com relação entre entrada e saída dada pela forma geral

$$Y_n = \sum_{i=0}^{N-1} a_i X_{n-i}$$

5.3 Fdp da soma de duas v.a.'s

Antes de analisar o modelo de probabilidade da soma de n v.a.'s, é instrutivo analisar a soma $W = X + Y$ de duas v.a.'s contínuas.

Teorema 5.4. A fdp de $W = X + Y$ é

$$f_W(w) = \int_{-\infty}^{+\infty} f_{XY}(x, w - x) dx = \int_{-\infty}^{+\infty} f_{XY}(w - y, y) dy$$

Demonstração. Para a prova deste teorema, vamos encontrar a fdp de W usando um procedimento em dois passos: primeiro encontramos a fdc $F_W(w)$ integrando a fdp conjunta $f_{XY}(x, y)$ sobre a região $X + Y \leq w$ mostrada na Figura 5.1, e depois encontramos a fdp $f_W(w)$ derivando a expressão de $F_W(w)$.

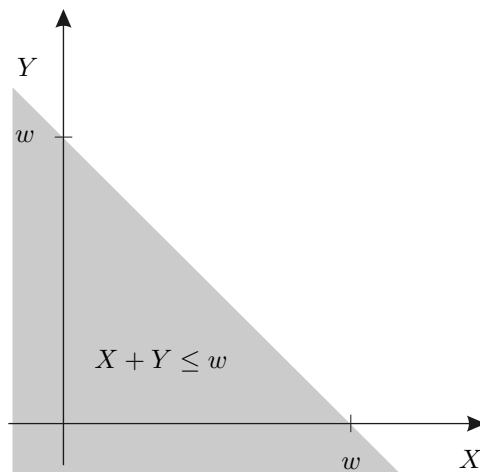


Figura 5.1: Região de integração para a obtenção de $F_W(w)$.

$$F_W(w) = P[X + Y \leq w] = \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{w-x} f_{XY}(x, y) dy \right) dx$$

Tomando as derivadas da fdc para encontrar a fdp, temos

$$f_W(w) = \frac{dF_W(w)}{dw} = \int_{-\infty}^{+\infty} \left(\frac{d}{dw} \left(\int_{-\infty}^{w-x} f_{XY}(x, y) dy \right) \right) dx = \int_{-\infty}^{+\infty} f_{XY}(x, w - x) dx$$

Através de um desenvolvimento similar, podemos mostrar também que

$$f_W(w) = \int_{-\infty}^{+\infty} f_{XY}(w - y, y) dy$$

□

Exemplo 5.2. Encontre a fdp de $W = X + Y$ se X e Y têm fdp conjunta

$$f_{XY}(x, y) = \begin{cases} 2 & 0 \leq y \leq 1, 0 \leq x \leq 1, x + y \leq 1 \\ 0 & \text{caso contrário} \end{cases}$$

Solução. A fdp de $W = X + Y$ pode ser encontrada usando-se o teorema 5.4. Note que X e Y são dependentes e que os valores possíveis de X, Y ocorrem na região triangular sombreada da Figura 5.2.

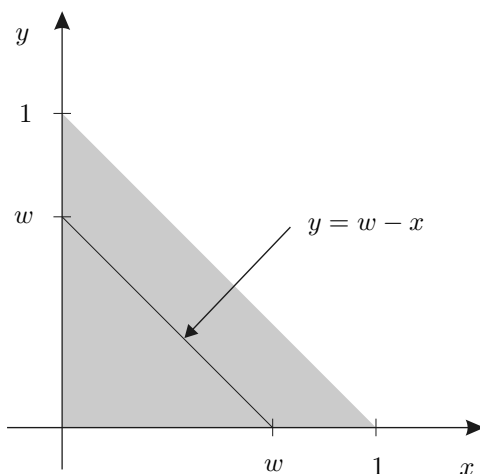


Figura 5.2: Região de integração para a obtenção de $F_W(w)$.

Portanto $0 \leq X + Y \leq 1$. Assim, $f_W(w) = 0$ para $w < 0$ ou $w > 1$. Para $0 \leq w \leq 1$, aplicando o teorema 5.4, chega-se a

$$f_W(w) = \int_0^w 2 \, dx = 2w \quad (0 \leq w \leq 1)$$

A expressão completa para a fdp de W é então dada por

$$f_W(w) = \begin{cases} 2w & 0 \leq w \leq 1 \\ 0 & \text{caso contrário} \end{cases}$$

Quando X e Y são independentes, a fdp conjunta de X e Y pode ser escrita como o produto das fdp's marginais $f_{XY}(x, y) = f_X(x)f_Y(y)$. Neste caso, podemos reescrever o teorema 5.4 como

Teorema 5.5. Quando X e Y são v.a.'s independentes, a fdp de $W = X + Y$ é

$$f_W(w) = \int_{-\infty}^{+\infty} f_X(w - y)f_Y(y) \, dy = \int_{-\infty}^{+\infty} f_X(x)f_Y(w - x) \, dx$$

Neste teorema combinamos duas funções de uma variável $f_X(\cdot)$ e $f_Y(\cdot)$ para produzir uma terceira função $f_W(\cdot)$. A combinação no teorema 5.5, é chamada de **convolução**, e é denotada por $f_W(\cdot) = f_X(\cdot) * f_Y(\cdot)$. De maneira geral, é melhor usar métodos de transformação para calcular a convolução de duas funções. Na linguagem de teoria de probabilidade, a transformada de uma fdp ou de uma fmp é uma **função geratriz de momentos**.

5.4 Função geratriz de momentos

A fdp da soma das v.a.'s independentes X_1, X_2, \dots, X_n é uma sequência de convoluções envolvendo as fdp's $f_{X_1}(x)$, $f_{X_2}(x)$, e assim por diante. Na teoria de sistemas lineares, uma convolução no domínio do tempo corresponde a uma multiplicação no domínio da frequência com as funções no tempo e na frequência relacionadas pela transformada de Fourier. Na teoria de probabilidade podemos, de forma similar, usar métodos de transformadas para substituir a convolução de fdp's por multiplicações de transformadas.

Definição 5.1. Função geratriz de momentos (FGM): Para uma v.a. X , a função geratriz de momentos (FGM) é dada por

$$\phi_X(s) = E[e^{sX}]$$

Esta definição se aplica tanto a v.a.'s contínuas como discretas. O que muda de um caso para outro é a forma de cálculo da esperança. Quando X é uma v.a. contínua

$$\phi_X(s) = \int_{-\infty}^{+\infty} e^{sx} f_X(x) dx \quad (5.2)$$

Esta equação indica que a FGM de uma v.a. contínua é similar à transformada de Laplace de uma função temporal. Para uma v.a. discreta Y a FGM torna-se

$$\phi_Y(s) = \sum_{y_i \in S_Y} e^{sy_i} p_Y(y_i) \quad (5.3)$$

Na forma integral da Equação (5.2), a FGM lembra a transformada de Laplace que é geralmente utilizada na teoria de sistemas lineares. A principal diferença é que a FGM é definida para valores reais de s .

Para uma dada v.a. X , existe uma faixa de valores possíveis de s para os quais $\phi_X(s)$ existe. O conjunto de valores de s para os quais $\phi_X(s)$ existe é chamada de **região de convergência**. Por exemplo, se X é uma v.a. não negativa, a região de convergência inclui todo $s \leq 0$. Para qualquer v.a. X , $\phi_X(s)$ sempre existe para $s = 0$.

Iremos usar a FGM avaliando suas derivadas em $s = 0$. À medida que a região de convergência inclui um intervalo não vazio $(-\varepsilon, \varepsilon)$ em torno da origem $s = 0$, podemos avaliar as derivadas da FGM em $s = 0$.

A exemplo da fmp de uma v.a. discreta e da fdp de uma v.a. contínua, a FGM é um modelo de probabilidade completo para uma v.a. Usando métodos de transformada inversa, é possível calcular a fmp ou a fdp a partir da FGM.

Exemplo 5.3. Se $X = a$, uma constante, então $f_X(x) = \delta(x - a)$, e

$$\phi_X(s) = \int_{-\infty}^{+\infty} e^{sx} \delta(x - a) dx = e^{sa}$$

Exemplo 5.4. Quando X tem uma fdp uniforme,

$$f_X(x) = \begin{cases} 1 & 0 \leq x \leq 1 \\ 0 & \text{caso contrário} \end{cases}$$

a FGM de X é

$$\phi_X(s) = \int_0^1 e^{sx} dx = \frac{e^s - 1}{s}$$

Exemplo 5.5. Seja a v.a. X com fdp exponencial

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \text{caso contrário} \end{cases}$$

a FGM de X é

$$\phi_X(s) = \int_0^{\infty} e^{sx} \lambda e^{-\lambda x} dx = \frac{\lambda}{\lambda - s}$$

Exemplo 5.6. Seja X uma v.a. de Bernoulli com

$$f_X(x) = \begin{cases} 1 - p & x = 0 \\ p & x = 1 \\ 0 & \text{caso contrário} \end{cases}$$

a FGM de X é

$$\phi_X(s) = E[e^{sx}] = (1 - p)e^0 + pe^s = 1 - p + pe^s$$

Exemplo 5.7. Seja X com fmp geométrica

$$f_X(x) = \begin{cases} (1 - p)^{x-1} p & x = 1, 2, \dots \\ 0 & \text{caso contrário} \end{cases}$$

a FGM de X é

$$\phi_X(s) = \sum_{x=1}^{\infty} e^{sx} (1 - p)^{x-1} p = pe^s \sum_{x=1}^{\infty} ((1 - p)e^s)^{x-1} = \frac{pe^s}{1 - (1 - p)e^s}$$

Exemplo 5.8. *Seja X com fmp de Poisson*

$$p_X(x) = \begin{cases} \alpha^x e^{-\alpha} / x! & x = 0, 1, 2, \dots \\ 0 & \text{caso contrário} \end{cases}$$

a FGM de X é

$$\phi_X(s) = \sum_{x=0}^{\infty} e^{sx} \alpha^x e^{-\alpha} / x! = e^{-\alpha} \sum_{x=0}^{\infty} (\alpha e^s)^x / x! = e^{\alpha(e^s - 1)}$$

A função geratriz de momentos tem algumas propriedades:

Teorema 5.6. *Para qualquer v.a. X , a FGM satisfaz*

$$\phi_X(s)|_{s=0} = 1$$

Demonstração.

$$\phi_X(s)|_{s=0} = E[e^{sX}]|_{s=0} = E[e^0] = 1$$

□

Este teorema é bastante útil para verificar se uma função pode ser uma FGM válida.

Teorema 5.7. *A FGM de $Y = aX + b$ satisfaz*

$$\phi_Y(s) = e^{sb} \phi_X(as)$$

Demonstração.

$$\phi_Y(s) = E[e^{s(aX+b)}] = e^{sb} E[e^{(as)X}] = e^{sb} \phi_X(as)$$

□

Como seu nome sugere, a função $\phi_X(s)$ é especialmente útil para encontrar os momentos de X .

Teorema 5.8. *Uma v.a. com FGM $\phi_X(s)$ tem n -ésimo momento*

$$E[X^n] = \left. \frac{d^n \phi_X(s)}{ds^n} \right|_{s=0}$$

Demonstração. A derivada primeira de $\phi_X(s)$ é

$$\frac{d\phi_X(s)}{ds} = \frac{d}{ds} \left(\int_{-\infty}^{+\infty} e^{sx} f_X(x) dx \right) = \int_{-\infty}^{+\infty} x e^{sx} f_X(x) dx$$

Avaliando esta derivada em $s = 0$, concluímos a prova para $n = 1$

$$\left. \frac{d\phi_X(s)}{ds} \right|_{s=0} = \int_{-\infty}^{+\infty} x f_X(x) dx = E[X]$$

Similarmente, a n -ésima derivada de $\phi_X(s)$ é dada por

$$\frac{d^n \phi_X(s)}{ds^n} = \int_{-\infty}^{+\infty} x^n e^{sx} f_X(x) dx$$

Avaliando a expressão acima em $s = 0$ completamos a prova do teorema. \square

Uma vantagem da FGM é que geralmente é mais fácil encontrar a FGM de X e tomar as derivadas para encontrar os momentos de X do que encontrá-los diretamente.

Exemplo 5.9. Encontre o n -ésimo momento de uma v.a. com fdp exponencial

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \text{caso contrário} \end{cases}$$

Solução. Podemos escrever o primeiro momento como

$$E[X] = \left. \frac{d\phi_X(s)}{ds} \right|_{s=0} = \left. \frac{\lambda}{(\lambda - s)^2} \right|_{s=0} = \frac{1}{\lambda}$$

o segundo momento como

$$E[X^2] = \left. \frac{d^2 \phi_X(s)}{ds^2} \right|_{s=0} = \left. \frac{2\lambda}{(\lambda - s)^3} \right|_{s=0} = \frac{2}{\lambda^2}$$

e o terceiro momento como

$$E[X^3] = \left. \frac{d^3 \phi_X(s)}{ds^3} \right|_{s=0} = \left. \frac{6\lambda}{(\lambda - s)^4} \right|_{s=0} = \frac{6}{\lambda^3}$$

Por indução, podemos afirmar que o n -ésimo momento de X é dado por

$$E[X^n] = \left. \frac{d^n \phi_X(s)}{ds^n} \right|_{s=0} = \left. \frac{n! \lambda}{(\lambda - s)^{n+1}} \right|_{s=0} = \frac{n!}{\lambda^n}$$

5.5 FGM da soma de v.a.'s independentes

FGM's são particularmente úteis para analisar a soma de v.a.'s independentes. Se $W = X + Y$ onde X e Y são v.a.'s com transformadas $\phi_X(s)$ e $\phi_Y(s)$ respectivamente, a transformada de W é

$$\phi_W(s) = E[e^{sW}] = E[e^{s(X+Y)}] = E[e^{sX}e^{sY}] \quad (5.4)$$

Geralmente a expressão acima é difícil de calcular. Entretanto, quando X e Y são independentes, podemos escrever a esperança do produto $e^{sX}e^{sY}$ como o produto das esperanças $E[e^{sX}]E[e^{sY}]$. Neste caso, encontrar $\phi_W(s)$ fica fácil se conhecermos $\phi_X(s)$ e $\phi_Y(s)$.

$$\phi_W(s) = E[e^{sX}e^{sY}] = E[e^{sX}]E[e^{sY}] = \phi_X(s)\phi_Y(s) \quad (5.5)$$

Quando as n v.a.'s X_1, X_2, \dots, X_n são independentes, a esperança do produto $g_1(X_1)g_2(X_2)\cdots g_n(X_n)$ pode ser escrita como o produto das esperanças

$$E[g_1(X_1)g_2(X_2)\cdots g_n(X_n)] = E[g_1(X_1)]E[g_2(X_2)]\cdots E[g_n(X_n)] \quad (5.6)$$

Esta expressão leva ao seguinte teorema

Teorema 5.9. *Para uma sequência X_1, X_2, \dots, X_n de n v.a.'s independentes, a FGM de $W = X_1 + X_2 + \cdots + X_n$ é*

$$\phi_W(s) = \phi_{X_1}(s)\phi_{X_2}(s)\cdots\phi_{X_n}(s)$$

Demonstração. Da definição de FGM

$$\phi_W(s) = E[e^{s(X_1+X_2+\cdots+X_n)}] = E[e^{sX_1}e^{sX_2}\cdots e^{sX_n}]$$

Usando a Equação (5.5) com $g_i(X_i) = e^{sX_i}$, a esperança do produto é

$$E[W] = E[e^{sX_1}]E[e^{sX_2}]\cdots E[e^{sX_n}] = \phi_{X_1}(s)\phi_{X_2}(s)\cdots\phi_{X_n}(s)$$

□

Quando X_1, X_2, \dots, X_n são independentes e identicamente distribuídas, $\phi_{X_i}(s) = \phi_X(s)$ para todo i , e o teorema 5.9 tem um corolário simples

Corolário 5.10. *Para as v.a.'s X_1, X_2, \dots, X_n independentes e identicamente distribuídas cada qual com FGM $\phi_X(s)$, a FGM de $W = X_1 + X_2 + \cdots + X_n$ é*

$$\phi_W(s) = [\phi_X(s)]^n$$

Vimos anteriormente que a fdp $f_W(w)$ é obtida através da convolução das fdp's individuais $f_{X_i}(x_i)$. A FGM $\phi_W(s)$ é simplesmente a multiplicação das FGM's individuais $\phi_{X_i}(s)$. Geralmente, o cálculo destas convoluções é um processo complexo e tedioso, e a alternativa seria transformar $f_X(x)$ em $\phi_X(s)$, e então usar o Corolário 5.10 para obter $\phi_W(s)$, e finalmente calcular a transformada inversa, obtendo-se assim $f_W(w)$.

Exemplo 5.10. *Seja K_1, K_2, \dots, K_n um conjunto de n v.a.'s independentes com distribuição de Poisson, tais que $E[K_i] = \lambda_i$. Encontre a FGM de $W = K_1 + K_2 + \dots + K_n$.*

Solução. Do Exemplo 5.8 sabemos que K_i tem FGM $\phi_{K_i}(s) = e^{\lambda_i(e^s - 1)}$. Pelo Corolário 5.10,

$$\phi_W(s) = e^{\lambda_1(e^s - 1)} e^{\lambda_2(e^s - 1)} \dots e^{\lambda_n(e^s - 1)} = e^{(\lambda_1 + \lambda_2 + \dots + \lambda_n)(e^s - 1)} = e^{(\lambda_T)(e^s - 1)}$$

onde $\lambda_T = \lambda_1 + \lambda_2 + \dots + \lambda_n$. Examinando o Exemplo 5.8, observamos que $\phi_W(s)$ é a FGM de uma v.a. com distribuição de Poisson com média λ_T . Portanto

$$f_W(w) = \begin{cases} \lambda_T^w e^{-\lambda_T} / w! & w = 0, 1, \dots, n \\ 0 & \text{caso contrário} \end{cases}$$

O modelo de probabilidade da soma de n v.a.'s identicamente distribuídas com distribuição de Poisson tem a mesma forma do modelo de probabilidade de cada v.a. individual. Esta propriedade é válida também para v.a.'s identicamente distribuídas com distribuição gaussiana. Para v.a.'s com outras distribuições esta propriedade não é mais válida.

Exemplo 5.11. *Encontre a FGM de uma v.a. Binomial K com fmp*

$$p_K(k) = \begin{cases} \binom{n}{k} p^k (1-p)^{(n-k)} & k = 0, 1, \dots, n \\ 0 & \text{caso contrário} \end{cases}$$

Solução. Calcular a FGM de K diretamente como $E[e^{sK}]$ é bastante complicado. Ao invés disso, lembremos que podemos representar K como $K = X_1 + X_2 + \dots + X_n$ onde cada X_i é uma v.a. de Bernoulli independente. Desta forma, do Exemplo 5.6

$$\phi_K(s) = (\phi_X(s))^n = (1 - p + pe^s)^n$$

Exemplo 5.12. *Uma v.a. Erlang- n T_n tem fdp*

$$f_{T_n}(t) = \begin{cases} \frac{\lambda^n t^{n-1} e^{-\lambda t}}{(n-1)!} & t \geq 0 \\ 0 & \text{caso contrário} \end{cases}$$

Encontre a FGM de T_n

Solução. A FGM de T_n pode ser calculada diretamente como

$$\phi_{T_n}(s) = \int_0^\infty e^{st} \frac{\lambda^n t^{n-1} e^{-\lambda t}}{(n-1)!} dt = \left(\frac{\lambda}{\lambda-s} \right)^n \underbrace{\int_0^\infty \frac{(\lambda-s)^n t^{n-1} e^{-(\lambda-s)t}}{(n-1)!} dt}_1$$

A integral (1) é igual a 1 pois é a integral de um fdp Erlang sobre todos os valores possíveis. Então

$$\phi_{T_n}(s) = \left(\frac{\lambda}{\lambda-s} \right)^n$$

No Exemplo 5.5 observamos que $\phi_X(s) = \lambda/(\lambda-s)$ é a FGM de uma v.a. exponencial X com média $1/\lambda$. Portanto, a soma de n v.a.'s exponenciais identicamente distribuídas, cada uma com média $1/\lambda$ tem FGM $(\lambda/(\lambda-s))^n$, que é exatamente a FGM de uma v.a. Erlang de ordem n .

Isto mostra que uma v.a. Erlang é a soma de v.a.'s exponenciais identicamente distribuídas.

5.6 Somas de v.a.'s gaussianas independentes

Seja X_1, X_2, \dots, X_n um conjunto de v.a.'s gaussianas independentes. Podemos usar a FGM de cada v.a. na soma para derivar algumas propriedades interessantes de $W_n = X_1 + X_2 + \dots + X_n$. Quando $n = 1$, $W = X_1$ é apenas uma v.a. gaussiana. Para encontrar sua FGM, encontramos inicialmente a FGM de uma v.a. $N(0, 1)$.

Teorema 5.11. *A FGM de uma v.a. gaussiana Z com média nula e variância unitária é*

$$\phi_Z(s) = e^{s^2/2}$$

Demonstração. A FGM de Z pode ser escrita como

$$\phi_Z(s) = \int_{-\infty}^{+\infty} e^{sz} f_Z(z) dz = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{sz} e^{-z^2/2} dz$$

Esta integral pode ser resolvida completando-se o quadrado no expoente

$$\phi_Z(s) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}(z^2 - 2sz + s^2)} e^{s^2/2} dz = e^{s^2/2} \underbrace{\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}(z-s)^2} dz}_1$$

O teorema se sustenta pois no lado direito temos uma integral de uma fdp gaussiana com média s e variância 1. \square

Teorema 5.12. *A FGM de uma v.a. gaussiana com média μ e variância σ^2 é*

$$\phi_X(s) = e^{s\mu + \sigma^2 s^2/2}$$

Demonstração. Uma v.a. gaussiana X com média μ e variância σ^2 pode ser expressa em termos da v.a. $Z \sim N(0,1)$ como

$$X = \sigma Z + \mu$$

Consequentemente, do Teorema 5.7, a FGM de X é

$$\phi_X(s) = e^{s\mu} \phi_Z(\sigma s) = e^{s\mu + \sigma^2 s^2/2}$$

□

Agora podemos apresentar o resultado principal desta seção.

Teorema 5.13. *A soma de n v.a.'s gaussianas independentes $W_n = X_1 + X_2 + \cdots + X_n$ tem uma distribuição gaussiana com média e variância dadas por*

$$E[W_n] = E[X_1] + E[X_2] + \cdots + E[X_n]$$

$$\text{Var}[W_n] = \text{Var}[X_1] + \text{Var}[X_2] + \cdots + \text{Var}[X_n]$$

Demonstração. Por conveniência, seja $\mu_i = E[X_i]$ e $\sigma_i^2 = \text{Var}[X_i]$. Desde que os X_i são independentes, sabemos que

$$\begin{aligned} \phi_W(s) &= \phi_{X_1}(s) \phi_{X_2}(s) \cdots \phi_{X_n}(s) \\ &= e^{s\mu_1 + \sigma_1^2 s^2/2} e^{s\mu_2 + \sigma_2^2 s^2/2} \cdots e^{s\mu_n + \sigma_n^2 s^2/2} \\ &= e^{s(\mu_1 + \mu_2 + \cdots + \mu_n) + (\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2) s^2/2} \end{aligned}$$

Da equação acima, pode-se ver que $\phi_W(s)$ é a FGM de uma v.a. gaussiana com média $\mu_1 + \mu_2 + \cdots + \mu_n$ e variância $\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2$. □

5.7 Somas aleatórias de v.a.'s independentes

Muitos problemas práticos podem ser analisados pela soma de v.a.'s identicamente distribuídas, mas cujo número de termos na soma é também uma v.a. Referimo-nos à v.a. resultante R como uma **soma aleatória** de v.a.'s independentes e identicamente distribuídas. Então, dada uma v.a. N e uma sequência de v.a.'s X_1, X_2, \dots, X_N identicamente distribuídas, seja

$$R = X_1 + X_2 + \cdots + X_N \tag{5.7}$$

Os dois exemplos a seguir descrevem processos estocásticos nos quais as observações são somas aleatórias de v.a.'s.

Exemplo 5.13. *Em um terminal de ônibus, conte o número de pessoas que chegam nos ônibus durante uma hora.*

Solução. Se o número de pessoas no i -ésimo ônibus é K_i e o número de ônibus que chegam é N , então o número de pessoas chegando durante uma hora é

$$R = K_1 + K_2 + \cdots + K_N$$

Em geral, o número N de ônibus que chegam irá ser uma v.a., e desta forma, R é uma soma aleatória de v.a.'s.

Exemplo 5.14. *Conte o número N de pacotes de dados transmitidos através de um link de comunicações em um minuto.*

Solução. Suponha que cada pacote é corretamente decodificado com probabilidade p , independentemente do resultado da decodificação de qualquer outro pacote. O número de pacotes decodificados corretamente em um minuto de transmissão é

$$R = X_1 + X_2 + \cdots + X_N$$

onde X_i é **1** se o i -ésimo pacote é decodificado corretamente, e **0**, caso contrário. Pelo fato de o número de pacotes transmitido N ser aleatório, R não é a v.a. binomial usual.

No exemplo acima, podemos utilizar os métodos utilizados para v.a.'s múltiplas para encontrar a fmp conjunta $f_{NR}(n, r)$. Entretanto não somos capazes de encontrar uma expressão simples em forma fechada para a fmp $f_R(r)$. Por outro lado, vamos demonstrar nesta seção que é possível expressar o modelo de probabilidade de R como uma fórmula para a FGM $\phi_R(s)$.

Embora nos exemplos acima tenhamos considerado apenas casos nos quais os X_i são v.a.'s discretas, será mais instrutivo enfatizar o caso em que os X_i são v.a.'s contínuas. Sejam as v.a.'s

$$W_n = X_1 + X_2 + \cdots + X_n \tag{5.8}$$

$$R = X_1 + X_2 + \cdots + X_N \tag{5.9}$$

É importante sabermos distinguir a v.a. W_n da v.a. R . Especificamente, W_n é a soma de um número determinístico particular n dos X_i e **não é** uma soma aleatória de v.a.'s. Portanto, a fdp de W_n é a fdp condicional de R dado que $N = n$. Em geral, encontrar a fdp ou a fmp de R é bastante difícil. Entretanto, encontrar a FGM de R é surpreendentemente fácil, como podemos ver no teorema a seguir

Teorema 5.14. *A soma aleatória de v.a.'s independentes e identicamente distribuídas $R = X_1 + X_2 + \cdots + X_N$ tem FGM dada por*

$$\phi_R(s) = \phi_N(\ln(\phi_X(s)))$$

Demonstração. Para encontrar $\phi_R(s) = E[e^{sR}]$, iremos usar iterações de esperanças, encontrando primeiro a esperança condicional $E[e^{sR}|N = n]$, e então tomando a esperança sobre N

$$\phi_R(s) = \sum_{n=0}^{\infty} E[e^{sR}|N = n] p_N(n) = \sum_{n=0}^{\infty} E[e^{s(X_1+X_2+\cdots+X_N)}|N = n] p_N(n)$$

Pelo fato de os X_i serem independentes de N ,

$$E[e^{s(X_1+X_2+\cdots+X_N)}|N = n] = E[e^{s(X_1+X_2+\cdots+X_n)}] = E[e^{sW_n}] = \phi_{W_n}(s)$$

Do teorema 5.10, sabemos que $\phi_{W_n}(s) = [\phi_X(s)]^n$, o que implica em

$$\phi_R(s) = \sum_{n=0}^{\infty} [\phi_X(s)]^n p_N(n)$$

Observamos que podemos escrever $[\phi_X(s)]^n = [e^{\ln(\phi_X(s))}]^n = e^{[\ln(\phi_X(s))]n}$. Isto implica

$$\phi_R(s) = \sum_{n=0}^{\infty} e^{[\ln(\phi_X(s))]n} p_N(n)$$

Reconhecendo que esta soma tem a mesma forma daquela da Equação (5.3), obtemos

$$\phi_R(s) = \phi_N(\ln(\phi_X(s)))$$

□

Exemplo 5.15. *O número N de páginas em uma transmissão de fax tem fmp geométrica com média $1/q = 4$. O número K de bits em uma página de fax também tem distribuição geométrica com média $1/p = 10^5$ bits, independentemente de qualquer outra página e do número de páginas. Encontre a FGM de B , o número total de bits em uma transmissão de fax.*

Solução. Quando a i -ésima página tem K_i bits, o número total de bits é a soma aleatória

$$B = K_1 + K_2 + \cdots + K_N$$

Então

$$\phi_B(s) = \phi_N(\ln(\phi_K(s)))$$

Do exemplo 5.7

$$\phi_N(s) = \frac{qe^s}{1 - (1-q)e^s} \quad \phi_K(s) = \frac{pe^s}{1 - (1-p)e^s}$$

Para calcular $\phi_B(s)$, substituímos $\ln(\phi_K(s))$ para toda ocorrência de s em $\phi_N(s)$. Equivalentemente, podemos substituir $\phi_K(s)$ para toda ocorrência de e^s em $\phi_N(s)$. Esta substituição leva a

$$\phi_B(s) = \frac{q \left(\frac{pe^s}{1 - (1-p)e^s} \right)}{1 - (1-q) \left(\frac{pe^s}{1 - (1-p)e^s} \right)} = \frac{pqe^s}{1 - (1-pq)e^s}$$

Podemos ver que B tem FGM de uma v.a. geométrica com média $1/(pq) = 400000$ bits.

Usando o teorema 5.14, podemos tomar as derivadas de $\phi_N(\ln(\phi_X(s)))$ para encontrar expressões simples para a média e variância de R

Teorema 5.15. *A soma aleatória das v.a.'s independentes e identicamente distribuídas $R = X_1 + X_2 + \cdots + X_N$ tem média e variância dadas por*

$$E[R] = E[N]E[X] \quad \text{Var}[R] = E[N] \text{Var}[X] + \text{Var}[N](E[X])^2$$

Demonstração. Pela regra da cadeia das derivadas,

$$\phi'_R(s) = \phi'_N(\ln(\phi_X(s))) \frac{\phi'_X(s)}{\phi_X(s)}$$

Desde que $\phi_X(0) = 1$, avaliando em $s = 0$, temos

$$E[R] = \phi'_R(0) = \phi'_N(0) \frac{\phi'_X(0)}{\phi_X(0)} = E[N]E[X]$$

Para a derivada segunda de $\phi_X(s)$ temos

$$\phi''_R(s) = \phi''_N(\ln(\phi_X(s))) \left(\frac{\phi'_X(s)}{\phi_X(s)} \right)^2 + \phi'_N(\ln(\phi_X(s))) \frac{\phi_X(s)\phi''_X(s) - [\phi'_X(s)]^2}{[\phi_X(s)]^2}$$

Novamente, avaliando em $s = 0$, temos

$$E[R^2] = E[N^2]\mu_X^2 + E[N] (E[X^2] - \mu_X^2)$$

Subtraindo $(E[R])^2 = (\mu_N\mu_X)^2$ de ambos os lados da equação acima completamos a prova. \square

Observe que $\text{Var}[R]$ contém dois termos: o primeiro termo $\mu_N \text{Var}[X]$ resulta da aleatoriedade de X , enquanto que o segundo termo $\text{Var}[N]\mu_X^2$ é uma consequência da aleatoriedade de N . Para visualizar isto, considere estes dois casos

- Suponha que N é determinístico, de modo que $N = n$ todas as vezes. Neste caso, $\mu_N = n$ e $\text{Var}[N] = 0$. A soma aleatória R é uma soma determinística ordinária $R = X_1 + X_2 + \cdots + X_n$ e $\text{Var}[R] = n \text{Var}[X]$.
- Suponha que N é aleatória, mas cada X_i é uma constante determinística x . Neste exemplo, $\mu_X = x$ e $\text{Var}[X] = 0$. além disso, a soma aleatória torna-se $R = Nx$ e $\text{Var}[R] = x^2 \text{Var}[N]$.

É importante enfatizar que os teoremas 5.14 e 5.15 exigem que N seja independente da sequência aleatória X_1, X_2, \dots, X_n , isto é, o número de termos na soma aleatória não pode depender dos valores dos termos da soma.

Exemplo 5.16. *Seja X_1, X_2, \dots uma sequência de v.a.'s gaussianas independentes e identicamente distribuídas com média 100 e variância 100. Se K é uma v.a. de Poisson com média 1, encontre a média e a variância de $R = X_1 + X_2 + \cdots + X_K$.*

Solução. A distribuição de R ou mesmo a FGM de R são difíceis de se obter. Entretanto, o Teorema 5.15 torna o cálculo dos momentos bastante fácil. Sabemos que uma v.a. de Poisson com média 1 também tem variância 1. Então

$$E[R] = E[X]E[K] = 100$$

$$\text{Var}[R] = E[K] \text{Var}[X] + \text{Var}[K](E[X])^2 = 100 + (100)^2 = 10100$$

Pode-se ver que a maior parte da variância é devida à aleatoriedade de K . Isto acontece porque é muito provável que K assuma os valores 0 e 1, e estas duas escolhas mudam de forma dramática a soma.

5.8 Teorema do limite central

Em um grande número de situações práticas, histogramas de medidas seguem aproximadamente uma curva em forma de sino. Um histograma é um gráfico de barras que divide o conjunto de medidas possíveis em intervalos iguais e mostra o número de medidas em cada intervalo.

Quando o tamanho de cada intervalo é pequeno e o número de medidas é grande, a forma da histograma assemelha-se bastante à forma da fdp da v.a. que descreve as medidas. Por exemplo, o primeiro gráfico da Figura 5.3 é um histograma derivado a partir de 400 repetições de um experimento. Em cada experimento alguém joga uma moeda 50 vezes e observa o número de coroas. O histograma segue aproximadamente uma curva em forma de sino. O segundo gráfico na Figura 5.3 mostra as probabilidades binomiais exatas do número de caras em 50 jogadas.

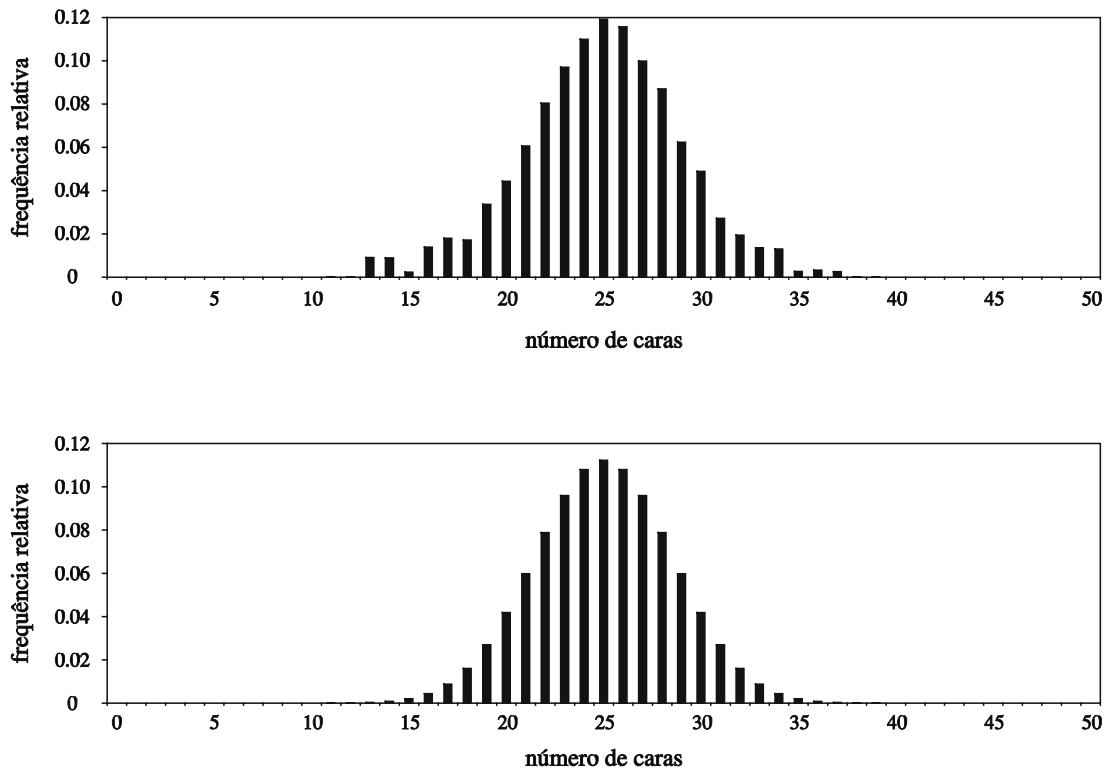


Figura 5.3: O número de caras em 50 arremessos de uma moeda ideal: 400 repetições experimentais versus a fmp binomial.

Lembremos que a fdp em forma de sino corresponde à de uma v.a. gaussiana. O teorema do limite central explica porque muitos fenômenos produzem dados que podem ser modelados como v.a.'s gaussianas na prática.

Iremos usar o teorema do limite central para estimar as probabilidades associadas com a soma de v.a.'s independentes e identicamente distribuídas $W_n = X_1 + X_2 + \cdots + X_n$. Entretanto, à medida que $n \rightarrow \infty$, $E[W_n] = n\mu_X$ e $\text{Var}[W_n] = n \text{Var}[X]$ tendem a infinito, o que faz com que seja muito difícil fazer uma afirmação matemática sobre a convergência da fdc $F_{W_n}(w)$. Portanto o teorema do limite central será escrito em termos da v.a. normalizada

$$Z_n = \frac{\sum_{i=1}^n X_i - n\mu_X}{\sqrt{n\sigma_X^2}}$$

Dizemos que a soma Z_n está normalizada pois para todo n

$$E[Z_n] = 0$$

$$\text{Var}[Z_n] = 1$$

Teorema 5.16. Teorema do limite central. Dada uma sequência X_1, X_2, \dots de v.a.'s independentes e identicamente distribuídas com valor esperado μ_X e variância σ_X^2 , a fdc de $Z_n = (\sum_{i=1}^n X_i - n\mu_X) / \sqrt{n\sigma_X^2}$ satisfaz

$$\lim_{n \rightarrow \infty} F_{Z_n}(z) = \Phi(z)$$

onde $\Phi(z)$ é a fdc de uma v.a. $N(0, 1)$.

A prova deste teorema é bastante complexa, e está fora do escopo deste texto. Além do Teorema 5.16 existem outros teoremas do limite central, cada um deles com sua própria restrição sobre a natureza da sequência W_n de v.a.'s.

Um aspecto singular do Teorema do Limite Central é o fato de não haver restrições sobre a natureza das v.a.'s X_i na soma. Elas podem ser contínuas, discretas ou mistas. Em todos os casos a fdc de sua soma assemelha-se mais e mais da fdc Gaussiana à medida que o número de termos na soma cresce. Algumas versões do Teorema do Limite Central aplicam-se a somas de sequências X_i que não são nem independentes e identicamente distribuídas.

Para usar o teorema do limite central, observe que podemos expressar a soma de v.a.'s identicamente distribuídas $W_n = X_1 + X_2 + \dots + X_n$ como

$$W_n = \sqrt{n\sigma_X^2} Z_n + n\mu_X \quad (5.10)$$

A fdc de W_n pode ser expressa em termos da fdc de Z_n como

$$F_{W_n}(w) \approx \Phi\left(\frac{w - n\mu_X}{\sqrt{n\sigma_X^2}}\right) \quad (5.11)$$

Para n grande, o teorema do limite central diz que $F_{Z_n}(z) \approx \Phi(z)$. Esta aproximação é a base para a maneira prática de se utilizar o teorema do limite central.

Corolário 5.17. Aproximação do teorema do limite central: Seja $W_n = X_1 + X_2 + \dots + X_n$ uma soma de v.a.'s independentes e identicamente distribuídas com $E[X] = \mu_X$ e $\text{Var}[X] = \sigma_X^2$. A fdc de W_n pode ser aproximada por

$$F_{W_n}(w) \approx \Phi\left(\frac{w - n\mu_X}{\sqrt{n\sigma_X^2}}\right)$$

Frequentemente chamamos a Definição 5.17 uma aproximação Gaussiana para W_n .

5.9 Aplicações do Teorema do Limite Central

O Teorema do Limite Central torna possível fazer cálculos rápidos e precisos que de outra maneira seriam bastante complexos e demorados. Nestes, a v.a. de interesse é

uma soma de outras v.a.'s, e calculamos as probabilidades dos eventos referindo-nos à v.a. Gaussiana correspondente.

Exemplo 5.17. *Um disco digital compacto (CD) contém amostras digitalizadas de uma forma de onda acústica.*

Em um CD player com um conversor D/A de 1 bit, cada amostra digital é representada com uma precisão de $\pm 0,5$ mV.

Para minimizar o erro de reprodução, a forma de onda é sobreamostrada tomando-se oito medidas independentes para cada amostra. O valor final da amostra da forma de onda é obtido calculando a média (média amostral) de oito medidas.

Qual é a probabilidade de o erro na amostra da forma de onda ser maior que 0.05 mV?

Solução. As medidas X_1, \dots, X_8 têm distribuição uniforme na faixa $(V - 0,5 \text{ mV}) < X < (V + 0,5 \text{ mV})$, onde V é o valor exato da amostra da forma de onda. O CD player produz a saída $U = W_8/8$ onde

$$W_8 = \sum_{i=1}^8 X_i$$

Para encontrar $P[|U - V| > 0.05]$ exatamente, precisaríamos encontrar o modelo de probabilidade exato para W_8 , ou calculando oito convoluções da fdp uniforme de X_i ou ainda, usando a função geratriz de momentos. De qualquer forma, o processo é extremamente complexo.

Alternativamente, podemos usar o Teorema do Limite Central para modelar W_8 como uma v.a. Gaussiana com $\mu = 8\mu_X = 8\text{mV}$ e variância $\text{Var}[W_8] = 8/12$.

Portanto, U é uma v.a. Gaussiana com $E[U] = E[W_8]/8 = V$ e $\text{Var}[W_8]/64 = 1/96$.

Finalmente, o erro $U - V$ na amostra da forma de onda de saída é Gaussiano com valor esperado zero e variância $1/96$. Segue então que

$$P[|U - V| > 0,05] = 2 \left[1 - \Phi \left(0,05/\sqrt{1/96} \right) \right] = 0,62$$

Exemplo 5.18. *Um modem transmite um milhão de bits. Cada bit é 0 ou 1 com probabilidades iguais. Estime a probabilidade de pelo menos 502000 uns.*

Solução. Seja W o número de uns em um milhão de bits. Note que $E[W] = 500000$ e $\text{Var}[W] = 10^6/4 = 250000$, de modo que $\sigma_W = 500$. Pela aproximação do Teorema do Limite Central,

$$P[W \geq 502000] = 1 - P[W < 502000] \approx 1 - \Phi \left(\frac{502000 - 500000}{500} \right) = 1 - \Phi(4)$$

Verificando os valores da função $\Phi(\cdot)$ em tabelas matemáticas, temos que

$$1 - \Phi(4) = Q(4) = 3,17 \times 10^{-5}$$

5.10 Exercícios

1. Seja W_n a soma de n arremessos independentes de um dado de quatro faces. Encontre a média e a variância de W_n .

$$\text{Resp: } E[W_n] = 2,5n \quad \text{Var}[W_n] = 1,25n$$

2. Sejam X e Y duas v.a.'s exponenciais independentes com médias $E[X] = 1/3$ e $E[Y] = 1/2$. Encontre a fdp de $W = X + Y$.

$$\text{Resp: } f_W(w) = 6(e^{-2w} - e^{-3w})$$

3. A v.a. K tem fmp dada por

$$f_K(k) = \begin{cases} 0,2, & k = 0, 1, 2, 3, 4 \\ 0, & \text{caso contrário} \end{cases}$$

Encontre a FGM $\phi_K(s)$ de K . Use-a para encontrar os quatro primeiros momentos de K .

$$\text{Resp: } \phi_K(s) = 0,2(1 + e^s + e^{2s} + e^{3s} + e^{4s})$$

$$E[K] = 2 \quad E[K^2] = 6 \quad E[K^3] = 20 \quad E[K^4] = 70,8$$

4. Seja K_1, K_2, \dots uma sequência de v.a.'s independentes e identicamente distribuídas, cada uma delas com distribuição dada por

$$f_K(k) = \begin{cases} 1/n, & k = 1, 2, \dots, n \\ 0, & \text{caso contrário} \end{cases}$$

Encontre a FGM de $J = K_1 + K_2 + \dots + K_m$

$$\text{Resp: } \phi_J(s) = \frac{e^{ms}(1 - e^{ns})^m}{n^m(1 - e^s)^m}$$

5. Seja X_1, X_2, \dots, X_n uma sequência de v.a.'s gaussianas independentes de média zero e variância tal que $\text{Var}[X_i] = i$. Encontre a fdp de

$$W = \alpha X_1 + \alpha^2 X_2 + \dots + \alpha^n X_n$$

$$\text{Resp: } f_W(w) = \frac{1}{\sqrt{2\pi\sigma_W^2}} e^{-w^2/2\sigma_W^2}$$

6. Seja X_1, X_2, \dots uma sequência de v.a.'s independentes e identicamente distribuídas com fdp exponencial

$$f_X(x) = \begin{cases} e^{-x}, & x \geq 0 \\ 0, & \text{caso contrário} \end{cases}$$

Seja N uma v.a. geométrica com média $1/p$. Qual é a FGM de $R = X_1 + X_2 + \dots + X_N$? Adicionalmente, encontre a fdp de R .

Resp:

$$\phi_R(s) = \frac{p}{p-s} \quad f_R(r) = \begin{cases} p e^{-pr}, & r \geq 0 \\ 0, & \text{caso contrário} \end{cases}$$

7. A v.a. X milissegundos é o tempo total de acesso (tempo de espera + tempo de leitura) para obter um bloco de informação de um disco de computador. X é uniformemente distribuída no intervalo de 0 a 12 milissegundos. Antes de realizar uma determinada tarefa, o computador precisa acessar 12 blocos de informação diferentes do disco. (Os tempos de acesso para blocos diferentes são independentes um do outro). O tempo total de acesso para todas as informações é uma v.a. A milissegundos.

- Calcule $E[X]$, o valor esperado para o tempo de acesso.
- Calcule $\text{Var}[X]$, a variância do tempo de acesso.
- Calcule $E[A]$, o valor esperado do tempo total de acesso.
- Calcule σ_A , o desvio padrão do tempo total de acesso.
- Use o Teorema do Limite Central para estimar $P[A > 75 \text{ ms}]$, a probabilidade do tempo total de acesso exceder 75 ms.
- Use o Teorema do Limite Central para estimar $P[A < 48 \text{ ms}]$, a probabilidade do tempo total de acesso ser menor que 48 ms.

Resp: a) $E[X] = 6 \text{ ms}$ b) $\text{Var}[X] = 12$ c) $E[A] = 72 \text{ ms}$ d) $\text{Var}[A] = 144$
 e) $P[A > 75] \approx 0,4013$ f) $P[A < 48] \approx 0,0227$

8. Seja X_1, X_2, \dots uma sequência de v.a.'s independentes e identicamente distribuídas com fdp uniforme entre 0 e 1, e seja N uma v.a. geométrica com média $1/p$.

- Qual é a FGM de $R = X_1 + X_2 + \dots + X_N$?
- Calcule a média e a variância de R .

Resp:

$$\begin{aligned} \text{(a)} \quad \phi_R(s) &= \frac{p(e^s - 1)}{s - (1-p)(e^s - 1)} \\ \text{(b)} \quad E[R] &= \frac{1}{2p} \quad \text{Var}[R] = \frac{3-2p}{12p^2} \end{aligned}$$

9. Seja X uma v.a. $N(0, 1)$. Encontre a média e a variância de $Y = 2X + 1$ usando a função geratriz de momentos.

Resp: $E[Y] = 1$ e $\text{Var}[Y] = 4$

10. Seja a função geratriz de momentos de uma v.a. discreta dada por

$$\phi_X(s) = 0.25e^s + 0.35e^{3s} + 0.40e^{5s}$$

Encontre $P[X = 0]$, $P[X = 1]$, $P[X = 2]$, $P[X = 3]$, $P[X = 4]$ e $P[X = 5]$.

Dica: lembre que, para o caso discreto, $\phi_X(s) = \sum_i e^{sx_i} f_X(x_i)$, e que $f_X(x_0) = P[X = x_0]$.

Resp: $P[X = 0] = 0$ $P[X = 1] = 0,25$ $P[X = 2] = 0$ $P[X = 3] = 0,35$
 $P[X = 4] = 0$ $P[X = 5] = 0,40$

11. Seja K_1, K_2, \dots uma sequência de v.a.'s iid com distribuição de Bernoulli, com fmp dada por

$$p_K(k) = \begin{cases} 1-p & k=0 \\ p & k=1 \\ 0 & \text{caso contrário} \end{cases}$$

Seja $M = K_1 + K_2 + \dots + K_n$.

- (a) Encontre a FGM $\phi_K(s)$
- (b) Encontre a FGM $\phi_M(s)$
- (c) Use $\phi_M(s)$ para calcular $E[M]$ e $Var[M]$.

Resp:

- (a) $\phi_K(s) = 1 - p + pe^s$
 - (b) $\phi_M(s) = (1 - p + pe^s)^n$
 - (c) $E[M] = np$ e $Var[M] = np(1 - p)$.
12. Suponha que durante o i -ésimo dia de dezembro, a energia X_i armazenada por um coletor solar é bem modelada por uma v.a. gaussiana com média $(32 - i)/4$ kWh e desvio padrão de 10 kWh. Assumindo que a energia armazenada a cada dia é independente de qualquer outro dia, qual é a fdp de Y , a energia total armazenada nos 31 dias de dezembro?

Resp: Gaussiana de média 124 e variância 3100

13. O k -ésimo momento de uma v.a. discreta é dado por

$$E[X^k] = 0,8, \quad k = 1, 2, \dots$$

- (a) Encontre a função geratriz de momentos de X .
- (b) Encontre $P[X = 0]$ e $P[X = 1]$.

Resp:

- (a) $\phi_X(s) = 0,2 + 0,8e^s$
 - (b) $P[X = 0] = 0,2$ e $P[X = 1] = 0,8$.
14. Seja X uma variável aleatória com distribuição $N(0, 1)$. Usando a função geratriz de momentos, determine $E[X^n]$ para $n = 1, 2, 3$.
- Resp:* $E[X] = 0$, $E[X^2] = 1$ e $E[X^3] = 0$.

15. As chamadas telefônicas podem ser classificadas como sendo de voz (V), se alguém está falando, ou de dados (D), se corresponder a uma transmissão de modem ou fax. Baseado em uma grande quantidade de observações realizadas por uma companhia telefônica, temos o seguinte modelo de probabilidade: $P[V] = 0.8$ e $P[D] = 0.2$. As chamadas de voz e de dados ocorrem independentemente umas das outras. Seja a variável aleatória K_n definida como o número de chamadas de dados em uma coleção de n chamadas telefônicas.

- Calcule $E[K_{100}]$, o número esperado de chamadas de dados em um conjunto de 100 chamadas.
- Calcule $\sigma_{K_{100}}$, o desvio padrão do número de chamadas de dados em um conjunto de 100 chamadas.
- Use o Teorema do Limite Central para estimar $P[K_{100} \geq 18]$, ou seja, a probabilidade de pelo menos 18 chamadas de dados em um conjunto de 100 chamadas telefônicas.
- Use o Teorema do Limite Central para estimar $P[16 \leq K_{100} \leq 24]$, ou seja, a probabilidade de existirem entre 16 e 24 chamadas de dados em um conjunto de 100 chamadas telefônicas.

Dica: $Q(-x) = 1 - Q(x)$.

Resp: (a) 20 (b) 4 (c) 0,6915 (d) 0,6826

16. Sejam X_1, X_2, \dots, X_n n variáveis aleatórias *iid* com distribuição de Cauchy

$$f_X(x) = \frac{a}{\pi(x\frac{1}{2} + a\frac{1}{2})}, -\infty < x < \infty$$

Seja a variável aleatória Y_n dada por

$$Y_n = \frac{1}{n}(X_1 + \dots + X_n) = \frac{1}{n} \sum_{i=1}^n X_i$$

- Encontre a função característica de Y_n .
- Encontre a fdp de Y_n .
- O Teorema do Limite Central se aplica neste caso? Justifique sua resposta.

Resp: (a) $\Psi_X(j\omega) = e^{-a|\omega|}$ (b) $F_{Y_n}(y_n) = \frac{a}{\pi(y_n^2 + a^2)}$ (c) não

17. Sejam X e Y duas variáveis aleatórias independentes com distribuição uniforme no intervalo $(0, 1)$. Encontre e esboce a fdp de $Z = X + Y$.

Dica: faça a análise para o intervalo $(0 < z < 1)$ e depois para o intervalo $(1 < z < 2)$.

$$\text{Resp: } f_Z(z) = \begin{cases} z & 0 < z < 1 \\ 2 - z & 1 < z < 2 \\ 0 & \text{caso contrário} \end{cases}$$

18. Seja K a soma de 20 variáveis aleatórias iid com distribuição de Bernoulli com probabilidade $p = 0,4$ de produzir um resultado igual a 1. Usando o Teorema do Limite Central, estime $P[K = 8]$, e compare com o valor exato para esta probabilidade.

Dica: Considere $P[7,5 < Z_n < 8,5]$ como aproximação para $P[K = 8]$. (Por quê?)

Resp: $P[K = 8] \approx 0,1811$.

19. O número N de serviços submetidos a um computador em uma hora é uma variável aleatória geométrica com parâmetro p , e os tempos de execução destes trabalhos são variáveis aleatórias independentes com distribuição exponencial de média $1/\alpha$. Encontre a fdp da soma dos tempos de execução dos trabalhos submetidos em uma hora.

$$\text{Resp: } f_R(r) = \begin{cases} \lambda_p e^{-\lambda_p} & r \geq 0 \\ 0 & \text{caso contrário} \end{cases}$$

20. As resistências dos resistores r_1, r_2, r_3 e r_4 são variáveis aleatórias independentes, cada uma delas uniformemente distribuída no intervalo $(450, 550)$. Usando o Teorema do Limite Central, calcule $P[1900 \leq r_1 + r_2 + r_3 + r_4 \leq 2100]$.

Resp: 0,9164

Capítulo 6

Limitantes Superiores para a Probabilidade de Cauda

Neste capítulo, iremos desenvolver desigualdades para probabilidades que podem ser muito difíceis de calcular exatamente. Geralmente, o desempenho de um sistema é determinado pela probabilidade de um evento indesejável. Por exemplo, a medida principal de um sistema de comunicação digital é a probabilidade de um erro de bit. Para um alarme de incêndio, a probabilidade de um falso alarme não pode ser muito grande; caso contrário o alarme pode ser ignorado quando houver um incêndio real. Quando o cálculo exato é muito difícil de realizar, um limitante superior oferece um meio de garantir que a probabilidade do evento indesejável não será muito alta.

6.1 Desigualdade de Markov

Teorema 6.1. Desigualdade de Markov. Para uma variável aleatória X não negativa e uma constante $c > 0$,

$$P[X \geq c] \leq \frac{E[X]}{c}$$

Demonstração. Desde que X é não negativo, $f_X(x) = 0$ para $x < 0$ e

$$\begin{aligned} E[X] &= \int_0^c x f_X(x) dx + \int_c^\infty x f_X(x) dx \\ &\geq \int_c^\infty x f_X(x) dx \geq c \int_c^\infty f_X(x) dx = cP[X \geq c] \end{aligned}$$

□

É importante lembrar que a desigualdade de Markov é válida somente para variáveis aleatórias não negativas. Como veremos no exemplo a seguir, geralmente o limitante fornecido pela desigualdade de Markov é bastante fraco.

Exemplo 6.1. *Seja X a altura (em pés) de um adulto selecionado aleatoriamente. Se o valor médio da altura é $E[X] = 5,5$, estime a probabilidade de um adulto ter pelo menos 11 pés usando a desigualdade de Markov.*

Solução. A desigualdade de Markov afirma que a probabilidade de um adulto ter pelo menos 11 pés satisfaz

$$P[X \geq 11] \leq \frac{5,5}{11} = 0,5$$

Dizemos que a desigualdade de Markov é folgada porque a probabilidade de uma pessoa ter uma altura maior que 11 pés é praticamente zero, enquanto que a desigualdade afirma meramente que ela é menor ou igual a 0,5. Embora esta desigualdade seja extremamente folgada para muitas variáveis aleatórias, ela é apertada (de fato, uma equação) com relação a algumas variáveis aleatórias.

Exemplo 6.2. *Suponha que uma v.a. Y tome o valor $c > 0$ com probabilidade p e o valor 0 caso contrário. Neste caso, $E[Y] = pc$ e utilizando a desigualdade de Markov, temos*

$$P[Y \geq c] \leq E[Y]/c = p$$

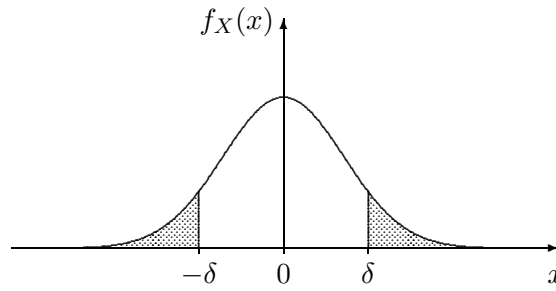
Desde que $P[Y \geq c] = p$, observamos que a desigualdade de Markov é de fato uma igualdade neste caso.

6.2 Desigualdade de Chebyshev

Teorema 6.2. Desigualdade de Chebyshev. *Seja X uma v.a. com média m_x e variância σ_X^2 finitas. Para todo número positivo δ*

$$P[|X - m_x| \geq \delta] \leq \frac{\sigma_X^2}{\delta^2}$$

Demonstração. $P[|X - m_x| \geq \delta]$ é a probabilidade da v.a. X ter um valor na região $A \equiv \{x : |X - m_x| \geq \delta\}$, mostrada na Figura 6.1.

Figura 6.1: Região A (sombreada).

Usando a expressão da variância, podemos escrever

$$\sigma_X^2 = \int_{-\infty}^{+\infty} (x - m_x)^2 f_X(x) dx \geq \int_A (x - m_x)^2 f_X(x) dx$$

Mas da definição da região A , temos $|X - m_x| \geq \delta \Rightarrow |X - m_x|^2 \geq \delta^2, \forall x \in A$. Assim

$$\int_A (x - m_x)^2 f_X(x) dx \geq \delta^2 \int_A f_X(x) dx$$

Mas $\int_A f_X(x) dx = P[|X - m_x| \geq \delta]$ e então podemos escrever

$$\sigma_X^2 \geq \delta^2 P[|X - m_x| \geq \delta] \Rightarrow P[|X - m_x| \geq \delta] \leq \frac{\sigma_X^2}{\delta^2}$$

□

Diferentemente da desigualdade de Markov, a desigualdade de Chebyshev é válida para todas as v.a.'s. Enquanto a desigualdade de Markov necessita apenas do valor esperado de uma v.a., a desigualdade de Chebyshev necessita também da variância. Por usar mais informações sobre a v.a., a desigualdade de Chebyshev geralmente fornece um limitante mais apertado do que a desigualdade de Markov.

Exemplo 6.3. Se a altura X de um adulto escolhido aleatoriamente tem valor esperado $E[X] = 5,5$ pés, e desvio padrão $\sigma_X = 1$ pés, use a desigualdade de Chebyshev para encontrar um limitante superior para $P[X \geq 11]$.

Solução. Desde que a altura X é não negativa, a probabilidade do evento $X \geq 11$ pode ser escrita como

$$P[X \geq 11] = P[X - \mu_X \geq 11 - \mu_X] = P[|X - \mu_X| \geq 5,5]$$

Usamos agora a desigualdade de Chebyshev para obter

$$P[X \geq 11] = P[|X - \mu_X| \geq 5,5] \leq \frac{\text{Var}[X]}{(5,5)^2} = \frac{1}{(5,5)^2} = 0,033$$

Embora este limitante seja melhor que o obtido pela desigualdade de Markov, é também bastante folgado. De fato, $P[X \geq 11]$ é na prática muitas ordens de magnitude menor que 0,033.

6.3 Limitante de Chernoff

O limitante de Chebyshev dado acima envolve a área das duas caudas da fdp. Em algumas aplicações estamos interessados somente na área de uma das caudas $(-\infty, -\delta)$ ou (δ, ∞) . Neste caso, podemos obter uma estimativa bastante justa, utilizando um limitante exponencial.

Teorema 6.3. Limitante de Chernoff. Para uma v.a. X e uma constante c arbitrárias,

$$P[X \geq c] \leq \min_{s \geq 0} e^{-sc} \phi_X(s)$$

Demonstração. Em termos da função degrau unitário, observamos que

$$P[X \geq c] = \int_c^{+\infty} f_X(x) dx = \int_{-\infty}^{+\infty} u(x - c) f_X(x) dx$$

Para todo $s \geq 0$, $u(x - c) \leq e^{s(x-c)}$, pois $e^{s(x-c)}$ representa uma família de curvas que passa pelo ponto c , como mostrado na Figura 6.2. Isto implica em

$$P[X \geq c] \leq \int_{-\infty}^{+\infty} e^{s(x-c)} f_X(x) dx = e^{-sc} \int_{-\infty}^{+\infty} e^{sx} f_X(x) dx = e^{-sc} \phi_X(s)$$

Este limitante é válido para qualquer $s \geq 0$. O limitante superior mais apertado é obtido selecionando-se o valor de s que minimiza $e^{-sc} \phi_X(s)$.

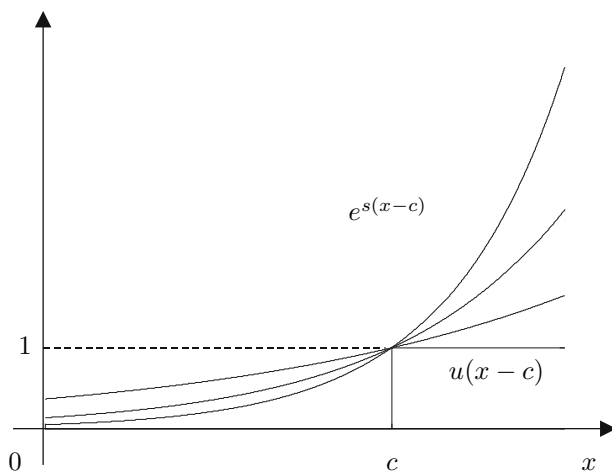


Figura 6.2: Um limitante superior exponencial usado para obter a probabilidade de cauda (limitante de Chernoff).

□

O limitante de Chernoff pode ser aplicado a qualquer v.a. Entretanto, para valores pequenos de c , $e^{-sc} \phi_X(s)$ irá ser minimizada por um valor negativo de s . Neste caso, o

valor de s não negativo que minimiza esta expressão é $s = 0$, o que fornece a resposta trivial: $P[X \geq c] \leq 1$

Exemplo 6.4. Se a altura X de um adulto escolhido aleatoriamente é uma v.a. gaussiana com valor esperado $E[X] = 5,5$ pés e desvio padrão $\sigma_X = 1$ pés, use o limitante de Chernoff para encontrar um limitante superior para $P[X \geq 11]$

Solução. Desde que X é $N(5,5,1)$, a FGM de X é

$$\phi_X(s) = e^{(11s+s^2)/2}$$

Então o limitante de Chernoff é

$$P[X \geq 11] \leq \min_{s \geq 0} e^{-11s} e^{(11s+s^2)/2} = \min_{s \geq 0} e^{(s^2-11s)/2}$$

Para encontrar s que minimiza a expressão acima, é suficiente encontrar s que minimize $h(s) = s^2 - 11s$. Tomando a derivada de $h(s)$ em relação a s e igualando a zero

$$\frac{dh(s)}{ds} = 2s - 11 = 0 \Rightarrow s = 5,5$$

Substituindo este valor de s ao limitante de Chernoff, chegamos a

$$P[X \geq 11] \leq e^{(s^2-11s)/2} \Big|_{s=5,5} = e^{-(5,5)^2/2} = 2,7 \times 10^{-7}$$

6.4 Exercícios

1. Em uma estação de metrô, existem usuários suficientes para completar exatamente três trens. Os trens chegam à estação segundo um processo de Poisson de taxa $\lambda = 0.5$ trens/minuto.

Seja X igual ao tempo em minutos requerido para servir os passageiros em espera. Encontre limitantes superiores para $P[X \geq 30 \text{ minutos}]$ usando as desigualdades de Markov, Chebyshev e Chernoff.

Dicas: i) o tempo entre chegadas pode ser modelado por uma variável aleatória com distribuição exponencial; ii) a soma de m variáveis aleatórias com distribuição exponencial é uma variável aleatória com distribuição m -Erlang.

Resp:

$$\text{Markov: } P[X \geq 30] = \frac{1}{5} \quad \text{Chebyshev: } P[X \geq 30] = \frac{1}{48}$$

$$\text{Chernoff: } P[X \geq 30] = 7,68 \cdot 10^{-4}$$

2. A média e a variância do tempo de resposta de um sistema de computador multiusuário são 15 segundos e 4 segundos, respectivamente. Estime a probabilidade de o tempo de resposta ser superior a 4 segundos da média, usando a desigualdade de Chebyshev.

Resp: 0,25

3. Dada uma v.a. X com fdp gaussiana de média zero e variância σ^2 , estime a probabilidade dos eventos $(-2\sigma \leq X \leq +2\sigma)$, $(-3\sigma \leq X \leq +3\sigma)$ e $(-4\sigma \leq X \leq +4\sigma)$ usando:

- (a) a função $Q(x)$;
- (b) a desigualdade de Chebyshev;
- (c) a desigualdade de Chernoff.

Resp:

- (a) Usando a função $Q(x)$:
 $P[-2\sigma \leq X \leq 2\sigma] = 0,9545$
 $P[-3\sigma \leq X \leq 3\sigma] = 0,9973$
 $P[-4\sigma \leq X \leq 4\sigma] = 0,9999$
- (b) Usando a desigualdade de Chebyshev:
 $P[-2\sigma \leq X \leq 2\sigma] = 0,75$
 $P[-3\sigma \leq X \leq 3\sigma] = 0,89$
 $P[-4\sigma \leq X \leq 4\sigma] = 0,9375$
- (c) Usando a desigualdade de Chernoff:
 $P[-2\sigma \leq X \leq 2\sigma] = 0,7293$
 $P[-3\sigma \leq X \leq 3\sigma] = 0,9778$
 $P[-4\sigma \leq X \leq 4\sigma] = 0,9993$

4. Use o limitante de Chernoff para mostrar que uma v.a. Z com distribuição $N(0,1)$ satisfaz

$$P[Z \geq c] \leq e^{-c^2/2}$$

Para $c = 1, 2, 3, 4$, verifique a diferença entre o limitante e o valor real da probabilidade.

Resp: Na tabela abaixo tem-se os valores aproximados pelo limitante de Chernoff, e os valores exatos, dados pela função $Q(x)$.

	Chernoff	$Q(x)$
$P[Z \geq 1]$	$\leq 0,6065$	0,1587
$P[Z \geq 2]$	$\leq 0,1353$	0,0228
$P[Z \geq 3]$	$\leq 0,0111$	$1,35 \cdot 10^{-3}$
$P[Z \geq 4]$	$\leq 0,0003$	$3,17 \cdot 10^{-5}$
$P[Z \geq 5]$	$\leq 3,7267 \cdot 10^{-6}$	$3,0 \cdot 10^{-7}$

5. Para uma variável aleatória arbitrária X , use a desigualdade de Chebyshev para estimar a probabilidade de X assumir um valor maior que k desvios padrões da média.

Resp: $\leq 1/k^2$.

6. Use o limitante de Chernoff para encontrar um limitante superior para $P[X \geq c]$ quando X é uma variável aleatória $N(\mu, \sigma^2)$.

Resp: $P[X \geq c] \leq e^{-\frac{(c-\mu)^2}{2\sigma^2}}$

7. Para uma variável aleatória arbitrária X , use a desigualdade de Chebyshev, calcule a probabilidade de X assumir valores maiores que k desvios padrões de seu valor esperado $E[x]$. Compare os valores obtidos com os valores exatos quando X é uma variável aleatória com distribuição $N(0, 1)$. Faça para 1, 2, 3 e 4 desvios padrões.

Resp: $1/k^2$.

8. Seja X uma variável aleatória com média 10 e variância 15. O que podemos dizer sobre $P[5 < X < 15]$?

Resp: $P[|X - 10| \leq 5] \geq 2/5$

Capítulo 7

A média amostral

7.1 Introdução

Vimos no Capítulo 1 que a frequência relativa é a razão entre o número de vezes que um evento ocorre e o número de vezes que um experimento é realizado. Se realizamos um experimento repetidas vezes, esperamos que a frequência relativa de cada evento convirja para uma constante à medida em que o número de repetições cresce.

Neste capítulo vamos definir a **média amostral** de uma v.a. e mostrar que muitas quantidades interessantes, incluindo a frequência relativa, podem ser expressas em termos da média amostral. Em seções posteriores, iremos mostrar matematicamente como a média amostral converge para uma constante à medida que o número de repetições de um experimento cresce.

Este capítulo, portanto, fornece a base matemática para a afirmativa de que embora o resultado de um único experimento aleatório seja imprevisível, padrões de comportamento previsíveis emergem quando coletamos mais e mais dados.

7.2 Valor esperado e variância

Para definir a média amostral, consideremos tentativas repetidas e independentes de um experimento aleatório. Cada tentativa resulta em uma observação da v.a. X . Depois de n tentativas, temos valores amostrais de n v.a.'s X_1, X_2, \dots, X_n todas com a mesma fdp de X . A média amostral é a média aritmética das observações.

Definição 7.1. Média Amostral: Para as v.a.'s X_1, X_2, \dots, X_n independentes e identicamente distribuídas com fdp $f_X(x)$, a média amostral de X é a variável aleatória

$$M_n(X) = \frac{X_1 + X_2 + \dots + X_n}{n} = \frac{1}{n} \sum_{i=1}^n X_i$$

A primeira coisa a ser notada é que $M_n(X)$ é uma função das v.a.'s X_1, X_2, \dots, X_n e portanto também é uma v.a. É importante distinguir a média amostral $M_n(X)$ do valor esperado $E[X]$ da v.a. X . Enquanto $M_n(X)$ é uma v.a., $E[X]$ é um número.

A média amostral e o valor esperado de X estão intimamente relacionados. Desta forma o propósito maior deste capítulo é explorar o fato de que à medida que n cresce sem limite, $M_n(X)$ aproxima $E[X]$.

Dependendo da definição da v.a. X , podemos usar a média amostral para descrever vários aspectos de um experimento. Por exemplo, se queremos explorar $P[A]$, a probabilidade de um evento arbitrário A , podemos definir uma v.a. indicadora X_A tal que $X_A = 1$ se o evento A ocorre, e $X_A = 0$ caso contrário. Neste caso, X_A é uma v.a. de Bernoulli com probabilidade de sucesso $P[A]$ de modo que $E[X_A] = P[A]$. Desde que as propriedades gerais do valor esperado de uma v.a. aplicam-se a $E[X_A]$, podemos ver que as técnicas para estimar valores esperados irão também nos permitir estimar as probabilidades de eventos arbitrários.

O valor esperado e a variância de $M_n(X)$ revelam as propriedades mais importantes da média amostral:

Teorema 7.1. *A média amostral $M_n(X)$ tem valor esperado e variância dados por*

$$E[M_n(X)] = E[X]$$

$$\text{Var}[M_n(X)] = \frac{\text{Var}[X]}{n}$$

Demonstração. Usando a Definição 7.1, o Teorema 5.1 e o fato de que $E[X_i] = E[X]$ para todo i ,

$$E[M_n(X)] = \frac{1}{n} (E[X_1] + E[X_2] + \cdots + E[X_n]) = \frac{1}{n} \underbrace{(E[X] + \cdots + E[X])}_{n \text{ vezes}} = E[X]$$

Para a variância, temos o seguinte:

$$\begin{aligned} \text{Var}[M_n(X)] &= E[M_n^2(X)] - E^2[M_n(X)] = E[M_n^2(X)] - E^2[X] \\ &= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n E[X_i X_j] - E^2[X] \\ &= \frac{1}{n^2} \sum_{\substack{i=1 \\ i=j}}^n E[X_i^2] + \frac{1}{n^2} \sum_{\substack{i=1 \\ i \neq j}}^n \sum_{j=1}^n E[X_i X_j] - E^2[X] \\ &= \frac{1}{n} (\sigma_X^2 + E^2[X]) + \frac{1}{n^2} n(n-1) E^2[X] - E^2[X] \\ &= \frac{\sigma_X^2}{n} \end{aligned}$$

□

Quando $M_n(X)$ é vista como uma estimativa para a média m_x , nota-se que seu valor esperado é igual a m_x e sua variância decresce inversamente com o número n de

amostras. À medida que $n \rightarrow \infty$, a variância σ_X^2 tende a zero. Uma estimativa de um parâmetro (neste caso a média m_x) que satisfaz a condição de que seu valor esperado converge para o valor real do parâmetro e a variância converge para zero à medida que $n \rightarrow \infty$ é dita uma **estimativa consistente**.

7.3 Média amostral de números grandes

Quando aplicamos a desigualdade de Chebyshev a $Y = M_n(X)$, obtemos informações importantes sobre as amostras independentes de uma v.a.

Teorema 7.2. *Para qualquer constante c , a média amostral $M_n(X)$ satisfaz*

$$(a) \ P[|M_n(X) - \mu_X| \geq c] \leq \frac{\text{Var}[X]}{nc^2} = \alpha$$

$$(b) \ P[|M_n(X) - \mu_X| < c] \geq 1 - \frac{\text{Var}[X]}{nc^2} = 1 - \alpha$$

Demonstração. Seja $Y = M_n(X)$. O Teorema 7.1 diz que

$$E[Y] = E[M_n(X)] = \mu_X \quad \text{Var}[Y] = \text{Var}[M_n(X)] = \text{Var}[X]/n$$

Aplicando a desigualdade de Chebyshev a $Y = M_n(X)$, conseguimos provar o item a). O item b) é apenas uma reafirmação do item a), desde que

$$P[|M_n(X) - \mu_X| \geq c] = 1 - P[|M_n(X) - \mu_X| < c]$$

□

Observações

O Teorema 7.2(b) contém duas desigualdades. Uma desigualdade,

$$|M_n(X) - \mu_X| < c$$

define um evento. Este evento diz que a média amostral está $\pm c$ unidades do valor esperado. O comprimento do intervalo que define este evento, $2c$ unidades, é chamado de **intervalo de confiança**.

A outra desigualdade afirma que a probabilidade da média amostral estar no intervalo de confiança é pelo menos $1 - \alpha$. Chamaremos a quantidade $1 - \alpha = 1 - \text{Var}[X]/(nc^2)$ de **coeficiente de segurança**. Se α é pequeno, podemos ter grande confiança de que $M_n(X)$ está no intervalo $(\mu_X - c, \mu_X + c)$.

No Teorema 7.2(a) observamos que para qualquer número c positivo, independente de quão pequeno seja, podemos fazer α tão pequeno quanto desejarmos escolhendo n grande o suficiente.

Em uma aplicação prática, c indica a precisão desejada de uma estimativa para μ_X , $1 - \alpha$ indica a confiança que temos de ter alcançado esta precisão, e n nos diz quantas amostras foram necessárias para alcançar o valor desejado de $1 - \alpha$.

Alternativamente, dados $\text{Var}[X]$, n e α , o Teorema 7.2(b) nos diz o tamanho c do intervalo de confiança.

Exemplo 7.1. O Teorema 7.2(b) dá origem a declarações que ouvimos no noticiário, tais como:

“Baseada em uma amostra de 1103 eleitores, a porcentagem de pessoas que apoiam o candidato José da Silva é de 58 % com precisão de mais ou menos 3 pontos percentuais.”

Comente estes fatos.

Solução. O experimento consiste em observar um eleitor escolhido aleatoriamente e determinar se o mesmo apóia ou não o candidato José da Silva.

Vamos associar o valor $X = 1$ se o eleitor apoiar o candidato José da Silva, e $X = 0$ caso contrário.

Portanto, X é uma v.a. com distribuição de Bernoulli com valor esperado $E[X] = p$ e variância $p(1 - p)$, onde $p = p_X(1)$. Para $c = 0,03$, o Teorema 7.2(b) diz

$$P[|M_n(X) - p| < 0,03] \geq 1 - \frac{p(1-p)}{n(0,03)^2} = 1 - \alpha$$

Desta forma o coeficiente de confiança para a estimativa de p é dado por

$$1 - \alpha = 1 - \frac{p(1-p)}{n(0,03)^2}$$

Devemos sempre ter em mente que temos grande confiança em nosso resultado quando α é pequeno. Entretanto, dede que não sabemos o valor real de p , gostaríamos de ter confiança em nossos resultados independentemente do valor de p .

Analisando a função $x(1 - x)$ para x entre 0 e 1, verifica-se que a mesma tem um máximo igual a $1/4$ em $x = 1/2$. Então para todos os valores de p entre 0 e 1, $\text{Var}[X] = p(1 - p) \leq 0,25$. Desta forma, podemos concluir que

$$1 - \alpha \geq 1 - \frac{0,25}{n(0,03)^2} = 1 - \frac{277,778}{n}$$

Então para $n = 1103$ amostras, $1 - \alpha \geq 0,75$. Isto nos diz que nossa estimativa de p está dentro de 3 pontos percentuais de p com probabilidade de pelo menos $1 - \alpha = 0,75$.

7.4 Leis de Números Grandes

Além da Desigualdade de Chebyshev e do Teorema 7.2(a), os quais descrevem as propriedades estatísticas de coleções de dados, temos as leis de números grandes, que se referem a limites quando estas coleções crescem sem limite.

7.4.1 Lei Fraca de Números Grandes

Teorema 7.3. Lei Fraca de Números Grandes. Se $\text{Var}[X] < \infty$, então para qualquer constante $c > 0$, a média amostral $M_n(X)$ satisfaz

$$(a) \lim_{n \rightarrow \infty} P[|M_n(X) - \mu_X| \geq c] = 0$$

$$(b) \lim_{n \rightarrow \infty} P[|M_n(X) - \mu_X| < c] = 1$$

Demonstração. A prova deste teorema segue diretamente dos resultados do Teorema 7.2 desde que

$$P[|M_n(X) - \mu_X| \geq c] = 1 - P[|M_n(X) - \mu_X| < c]$$

□

A lei fraca de números grandes afirma que, para um valor suficientemente grande e *fixo* de n , a probabilidade da média amostral usando n amostras estar perto da média real é alta.

Como podemos ver no exemplo seguinte, a lei fraca de números grandes também valida a interpretação de frequência relativa de probabilidades.

Exemplo 7.2. Suponha que realizemos n tentativas independentes de um experimento. Vamos definir a v.a. indicadora para o evento A como

$$X_i = \begin{cases} 1, & \text{se } A \text{ ocorre na tentativa } i \\ 0, & \text{caso contrário} \end{cases}$$

X_1, X_2, \dots é uma sequência aleatória de Bernoulli com probabilidade de sucesso $P[A]$. Então $E[X_i] = P[A]$ e $\text{Var}[X_i] = P[A](1 - P[A])$.

A frequência relativa de A em n tentativas é

$$R_n = M_n(X) = \frac{X_1 + X_2 + \dots + X_n}{n}$$

Desde que $E[R_n] = E[X_i] = P[A]$, o Teorema 7.3(a) diz que

$$\lim_{n \rightarrow \infty} P[|R_n - P[A]| \geq c] = 0$$

Portanto, à medida que $n \rightarrow \infty$, $R_n \rightarrow P[A]$, que é a versão matemática da afirmação de que à medida que o número de observações cresce sem limite, a frequência relativa de qualquer evento aproxima a probabilidade do evento.

7.4.2 Lei Forte de Números Grandes

Suponha que realizemos uma série de medidas independentes da mesma v.a. Seja X_1, X_2, \dots a sequência resultante de v.a.'s identicamente distribuídas com média μ . Considere agora uma **sequência de médias amostrais** que resulta das medidas acima: M_1, M_2, \dots , onde M_j é a média amostral usando as amostras X_1 até X_j . Por causa da regularidade estatística do experimento, espera-se que esta sequência de médias amostrais convirja para μ , isto é, esperamos que com probabilidade alta, cada sequência particular de médias amostrais aproxime-se de μ e permaneça lá, como mostrado na Figura 7.1. Formalmente, podemos escrever este resultado da seguinte maneira:

Teorema 7.4. *Seja X_1, X_2, \dots uma sequência de v.a.'s independentes e identicamente distribuídas com média $E[X] = \mu$ e variância finitas. Então*

$$P \left[\lim_{n \rightarrow \infty} M_n(X) = \mu \right] = 1$$

Este resultado é similar àquele obtido no Teorema 7.3, mas na verdade faz uma afirmação dramaticamente diferente: afirma que com probabilidade 1, toda sequência de cálculos de médias amostrais irá eventualmente aproximar-se e permanecer perto de $E[X] = \mu$. Este é o tipo de convergência que esperamos observar em situações reais onde haja regularidade estatística.

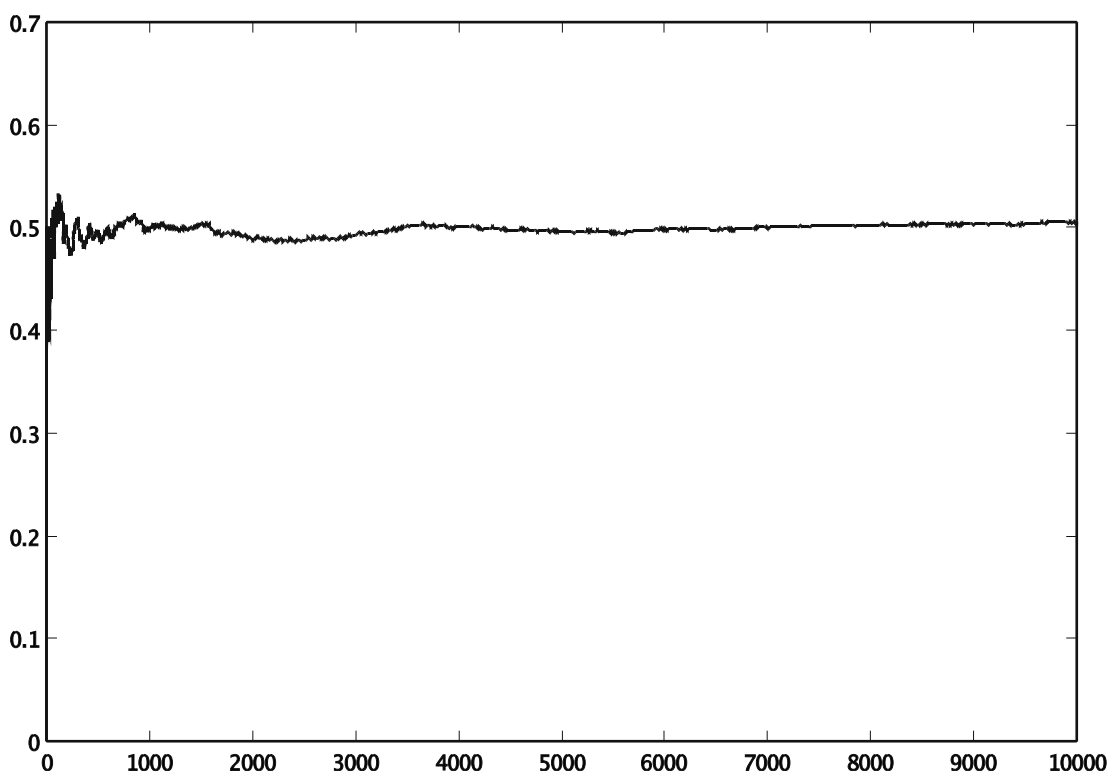


Figura 7.1: Convergência de uma sequência de médias amostrais obtidas a partir de uma sequência de v.a.'s com distribuição Gaussiana de média 4 e variância 10.

7.5 Exercícios

1. Suponha que o número de emissões de partículas de uma massa radioativa em t segundos é uma v.a. com distribuição de Poisson com média λt . Use a desigualdade de Chebyshev para obter um limitante para $P[|N(t)/t - \lambda| > \varepsilon]$.

Resp: $P[|N(t)/t - \lambda| \geq \varepsilon] \leq \lambda/\varepsilon^2 t$

2. Suponha que 10 % dos eleitores estão a favor de certa lei. Um grande número n de eleitores é consultado e obtém-se uma estimativa por frequência relativa $f_A(n)$ da proporção acima. Use o Teorema 7.2 para determinar quantos eleitores devem ser consultados de modo a termos uma probabilidade de pelo menos 0,95 de $f_A(n)$ diferir de 0,10 em menos de 0,02.

Resp: $n = 4500$

3. Um dado ideal é arremessado 100 vezes. Use o Teorema 7.2 e encontre um limitante para a probabilidade de o número total de pontos estar entre 300 e 400.

Resp: $P[|M_n(x) - 350| \leq 50] = 0,9994$

4. Seja X_i uma sequência de v.a.'s Gaussianas independentes de média zero e variância unitária. Compare o limitante dado pela Teorema 7.2 com o valor exato para $n = 10$ e $n = 100$.

Resp: Para $c = 4$, temos:

(a) Valor exato: $P[|M_n(x)| < 4] = 1 - 2Q(4) \approx 0,9999367$

(b) $n = 10$: $P[|M_n(x)| < 4] \geq 0,99375$

(c) $n = 100$: $P[|M_n(x)| < 4] \geq 0,999375$

5. (Para ser feito no MATLAB) Gere sequências de números aleatórios com diversas distribuições, variando a média (e a variância, quando for o caso) e calcule as sequências de médias amostrais. Com isto podemos comprovar na prática a lei forte de números grandes.
6. Deseja-se medir uma tensão constante mas desconhecida. Cada medida X_j é na verdade a soma da tensão desejada v com a tensão do ruído N_j de média zero e desvio padrão de $1 \mu\text{V}$

$$X_j = v + N_j$$

Assuma que as tensões do ruído são v.a.'s independentes. Quantas medidas serão necessárias de modo que a probabilidade de $M_n(X)$ esteja a $\varepsilon = 1 \mu\text{V}$ da média verdadeira seja pelo menos 0,99?

Resp: $n \geq 100$

7. Seja X uma variável aleatória com função densidade de probabilidade $f_X(x)$, e seja X_1, X_2, \dots, X_n um conjunto de variáveis aleatórias independentes, cada qual com função densidade de probabilidade $f_X(x)$. O conjunto de variáveis aleatórias X_1, X_2, \dots, X_n é chamado de uma *amostra aleatória de tamanho n de X* . A média amostral é definida como

$$\overline{X}_n = \frac{1}{n} (X_1 + \cdots + X_n) = \frac{1}{n} \sum_{i=1}^n X_i$$

Seja X_1, X_2, \dots, X_n uma amostra aleatória de X com média μ e variância σ^2 . Quantas amostras de X devemos tomar para que a probabilidade da média amostral desviar da média real μ por mais que $\sigma/10$ seja de pelo menos 0,95?

Resp: $n \geq 2000$

Capítulo 8

Processos Estocásticos

8.1 Definição

A noção de processo estocástico é uma extensão do conceito de v.a. Considere, por exemplo, a temperatura X de uma certa cidade ao meio dia. A temperatura X é uma v.a. e toma valores diferentes a cada dia. Para obter as estatísticas completas de X , precisamos armazenar valores de temperatura durante vários dias (um grande número de tentativas). A partir destes dados podemos determinar $f_X(x)$, a fdp da v.a. X .

Mas a temperatura é também função do tempo. À uma da tarde, por exemplo, a temperatura pode ter uma distribuição totalmente diferente daquela obtida para o meio dia. Então a v.a. X é uma função do tempo, e pode ser expressa como $X(t)$.

Definição 8.1. *Uma v.a. que é uma função do tempo é chamada de um **processo estocástico** (ou **processo aleatório**).*

Para especificar uma v.a. X , repetimos um experimento várias vezes e a partir dos resultados, determinamos a sua fdp $f_X(x)$. Similarmente, para especificar um processo estocástico $X(t)$, fazemos a mesma coisa para cada valor de t .

Continuando com nosso exemplo, precisamos armazenar temperaturas diárias para cada valor de t (cada hora do dia). Isto pode ser feito armazenando-se temperaturas a cada instante do dia. Este procedimento fornece uma forma de onda $X(t; \zeta_i)$ onde ζ_i indica o dia em que foi feita a medida. Precisamos repetir este procedimento todos os dias por um grande número de dias.

A coleção de todas as formas de onda possíveis é conhecida como o **conjunto** do processo estocástico $X(t)$, e uma forma de onda nesta coleção é uma **função amostra** (ao invés de um ponto amostral) do processo estocástico. As amplitudes das funções amostra em algum instante $t = t_1$ são os valores que a v.a. $X(t_1)$ assume em várias tentativas. Na Figura 8.1 tem-se o conceito que acabamos de definir em forma gráfica.

Podemos ver um processo estocástico de outra forma. No caso de uma v.a., o resultado de cada tentativa de um experimento aleatório é um número. Para um processo estocástico o resultado de cada tentativa é uma forma de onda (uma função amostra) que é uma função de t . O número de formas de onda em um conjunto pode ser finito ou infinito. No caso do processo estocástico $X(t)$ (a temperatura de uma cidade), o

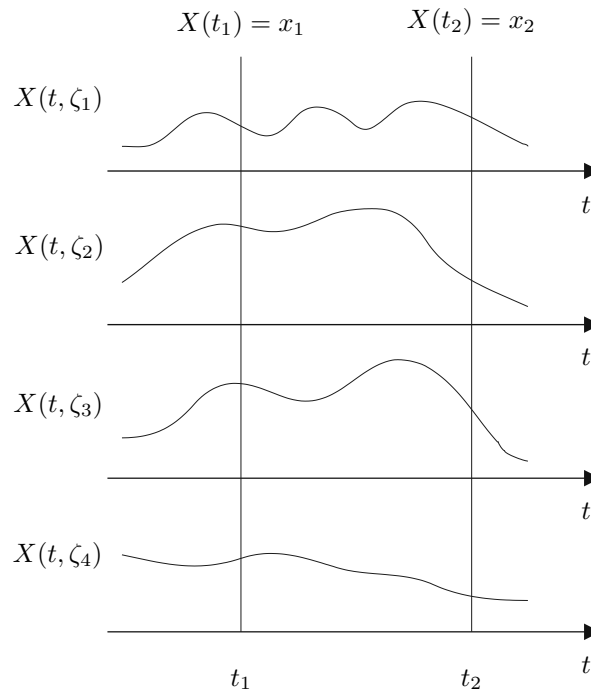


Figura 8.1: Um processo estocástico que representa a temperatura de uma cidade.

conjunto tem infinitas formas de onda. Por outro lado, se considerarmos a saída de um gerador de sinais binários (sobre um período de 0 a $10T$) existem no máximo 2^{10} formas de onda neste conjunto (Figura 8.2).

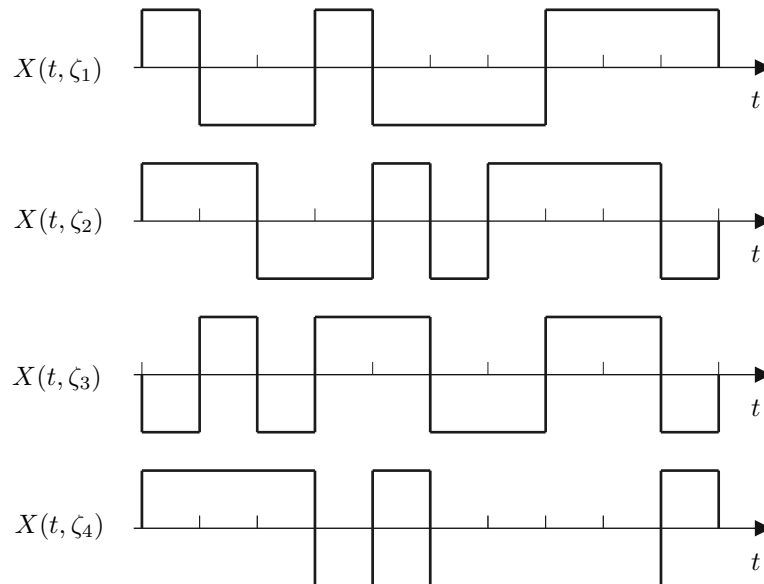


Figura 8.2: Um conjunto com um número finito de funções amostra.

Um ponto que precisa ser esclarecido é que as formas de onda (funções amostra) não são aleatórias, mas determinísticas. A aleatoriedade neste caso é associada não com a forma de onda mas com a incerteza de qual delas vai ocorrer em uma dada tentativa.

Isto é completamente análogo ao caso de uma v.a. Por exemplo, no experimento de jogar uma moeda quatro vezes em sucessão, existem 16 resultados possíveis, todos conhecidos. A aleatoriedade nesta situação está associada não aos resultados mas com a incerteza de qual deles irá ocorrer em uma dada tentativa.

8.2 Tipos de procesos estocásticos

Os processos estocásticos podem ser classificados em termos dos valores que podem assumir assim como dos instantes de tempo em que podem sofrer mudanças. Segundo esta ótica, podem ser classificados em processos de valor discreto e valor contínuo, e processos de tempo discreto e tempo contínuo.

Definição 8.2. Processos de valor contínuo e de valor discreto: $X(t)$ é um processo de valores discretos se o conjunto de todos os valores possíveis de $X(t)$ para todos os instantes de tempo t é um conjunto contável S_X ; caso contrário, $X(t)$ é um processo de valores contínuos.

Definição 8.3. Processos de tempo contínuo e tempo discreto. O processo estocástico $X(t)$ é de tempo discreto se $X(t)$ é definido apenas para um conjunto de instantes de tempo $t_n = nT$, onde T é uma constante e n é um inteiro; caso contrário, $X(t)$ é um processo de tempo contínuo

Estes conceitos são ilustrados na Figura 8.3. Nesta, podemos identificar que para o processo $X(t)$, existem quatro possibilidades básicas:

- amplitude discreta, tempo discreto
- amplitude discreta, tempo contínuo
- amplitude contínua, tempo discreto
- amplitude contínua, tempo contínuo

Para um processo de tempo discreto, a função amostra é completamente descrita pela sequência ordenada de variáveis aleatórias $X_n = X(nT)$.

Definição 8.4. Sequência aleatória. Uma sequência aleatória é uma sequência ordenada de variáveis aleatórias X_0, X_1, \dots

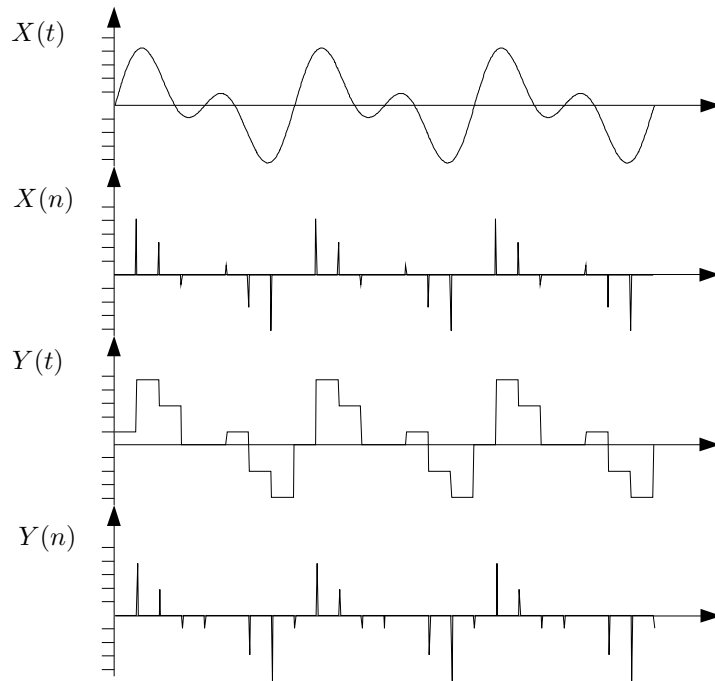


Figura 8.3: Funções amostra de quatro tipos de processos estocásticos: $X(t)$ é um processo contínuo no tempo e na amplitude; $X(n)$, obtido a partir da amostragem de $X(t)$ em instantes de tempo inteiros n , é discreto no tempo e contínuo na amplitude; $Y(t)$ é obtida a partir da quantização de $X(t)$ nos instantes de amostragem, e é um processo discreto na amplitude e contínuo no tempo; finalmente, $Y(n)$, um processo discreto no tempo e na amplitude, é obtido a partir da amostragem de $Y(t)$.

Além de caracterizar os processos estocásticos em relação à sua natureza temporal e das amplitudes, podemos classificá-los quanto ao seu tempo de duração:

Definição 8.5. *Processos de duração finita, semi-infinita e infinita.*

- a) $X(t)$ é um processo de duração finita $t_2 - t_1$ se para todo s , $x(t, s) = 0$ para $t < t_1$ e $t > t_2 > t_1$.
- b) $X(t)$ é um processo de duração semi-infinita se para todo s , $x(t, s) = 0$, para $t < t_1$.
- c) caso contrário, $X(t)$ é um processo de duração infinita.

8.3 Variáveis aleatórias a partir de processos estocásticos

Suponha que estamos observando um processo estocástico em um instante de tempo particular t_1 . Neste caso, cada vez que realizamos um experimento, observamos uma função amostra $x(t, s)$ e esta função amostra especifica o valor de $x(t_1, s)$. Cada vez que realizamos o experimento, temos um novo s e observamos um novo $x(t_1, s)$. Portanto,

cada $x(t_1, s)$ é uma amostra de uma variável aleatória. Aqui é usada a notação $X(t_1)$ para esta variável aleatória. Como qualquer outra variável aleatória, tem ou uma fdp $f_{X(t_1)}(x)$ ou uma fmp $p_{X(t_1)}(x)$. Note que a notação $X(t)$ pode se referir tanto a um processo estocástico como a uma variável aleatória, correspondente ao valor do processo estocástico no instante t . Nas seções subsequentes, irá ficar claro a partir do contexto se estamos nos referindo ao processo inteiro ou à uma variável aleatória.

Exemplo 8.1. *Seja $X(t) = R|\cos(2\pi ft)|$ um sinal cossenoidal retificado com amplitude aleatória R com fdp exponencial*

$$f_R(r) = \begin{cases} \frac{1}{10} e^{-r/10}, & r \geq 0 \\ 0, & \text{caso contrário} \end{cases}$$

Qual é a fdp de $f_{X(t_1)}(x)$?

Solução. Desde que $X(t) \geq 0$ para todo t , $P[X(t) \leq x] = 0$ para $x < 0$. Quando $x \geq 0$ e $\cos(2\pi ft) \neq 0$,

$$P[X(t) \leq x] = P[R \leq x/|\cos(2\pi ft)|] = \int_0^{x/|\cos(2\pi ft)|} f_R(r) dr = 1 - e^{-x/10|\cos(2\pi ft)|}$$

Quando $\cos(2\pi ft) \neq 0$, a cdf completa de $X(t)$ é

$$F_{X(t)}(x) = \begin{cases} 0, & x < 0 \\ 1 - e^{-x/10|\cos(2\pi ft)|}, & x \geq 0 \end{cases}$$

Quando $\cos(2\pi ft) \neq 0$, a fdp completa de $X(t)$ é

$$f_{X(t)}(x) = \frac{dF_{X(t)}(x)}{dx} = \begin{cases} 0, & x < 0 \\ \frac{1}{10|\cos(2\pi ft)|} e^{-x/10|\cos(2\pi ft)|}, & x \geq 0 \end{cases}$$

Quando $|\cos(2\pi ft)| = 0$, o que corresponde a $t = \pi/2 + k\pi$, $X(t) = 0$ independente de quão grande R possa ser. Neste caso, $f_{X(t)}(x) = \delta(x)$. Neste exemplo, existe uma variável aleatória diferente para cada valor de t .

Quando $X(t)$ é um processo de tempo discreto, toda informação sobre o mesmo está contida no valor da constante T na Definição 8.3 e a sequência de variáveis aleatórias, $X(nT)$, $n = \dots, -2, -1, 0, 1, 2, \dots$

Exemplo 8.2. Suponha que nos instantes de tempo $T = 0, 1, 2, \dots$, jogamos um dado e anotamos o resultado N_T , onde $1 \leq N_T \leq 6$. Definimos um processo estocástico $X(t)$ tal que para $T \leq t < T + 1$, $X(t) = N_T$. Neste caso, o experimento consiste em uma sequência infinita de jogadas, e uma função amostra é apenas uma forma de onda correspondente à sequência particular dos resultados observados.

Seja $X_n = X(nT)$. Qual é a fmp de X_3 ?

Solução. a variável aleatória X_3 é o valor da jogada do dado no instante 3. Neste caso,

$$p_{X_3}(x) = \begin{cases} 1/6, & x = 1, 2, \dots, 6 \\ 0, & \text{caso contrário} \end{cases}$$

Vimos no Capítulo 2 que a fdp $f_X(x)$ é um modelo probabilístico completo para a variável aleatória X . Similarmente, para um par de variáveis aleatórias X_1, X_2 , precisamos da fdp conjunta $f_{X_1, X_2}(x_1, x_2)$. Vimos também que as fdps marginais $f_{X_1}(x_1)$ e $f_{X_2}(x_2)$ não são suficientes para descrever este par de variáveis aleatórias.

Para processos estocásticos, a situação é similar. Se amostramos um processo em k instantes de tempo t_1, \dots, t_k , obtemos k variáveis aleatórias $X(t_1), \dots, X(t_k)$. É possível ver esta coleção de variáveis aleatórias como um vetor k -dimensional $[X(t_1), X(t_2), \dots, X(t_k)]$, chamado de vetor aleatório.

8.4 Sequências aleatórias independentes e identicamente distribuídas

Definição 8.6. Sequências iid. Uma sequência aleatória independente e identicamente distribuída (iid) é uma sequência aleatória X_n para a qual $\dots, X_{-2}, X_{-1}, X_0, X_1, X_2, \dots$ são variáveis aleatórias iid.

Uma sequência aleatória ocorre quando realizamos tentativas independentes de um experimento a uma taxa constante. Uma sequência aleatória pode assumir tanto valores discretos quanto contínuos. No caso discreto, cada variável aleatória X_i tem fmp $p_{X_i}(x) = p_X(x)$, enquanto que no caso contínuo, cada X_i tem fdp $f_{X_i}(x) = f_X(x)$.

Exemplo 8.3. Em uma linha de produção de resistores de 1000Ω , a resistência real de cada resistor é uma variável aleatória R com distribuição uniforme entre 950Ω e 1050Ω . Assuma que os valores das resistências dos diferentes resistores são independentes. A companhia tem uma encomenda de resistores de 1% de tolerância (resistências entre 990Ω e 1010Ω). Um testador automático toma um resistor por segundo e mede sua resistência exata. Seja R_n igual ao número de resistores com tolerância de 1% encontrados durante o minuto n . Assim, a variável aleatória R_n tem fmp binomial

$$p_{R_n}(r) = \begin{cases} \binom{60}{r} p^r (1-p)^{60-r}, & r = 0, 1, \dots, 60 \\ 0, & \text{caso contrário} \end{cases}$$

Desde que cada resistor é um resistor de tolerância 1% independentemente de todos os outros resistores, o número de resistores com tolerância de 1% encontrados a cada minuto é independente do número encontrado em outros minutos. Então, R_1, R_2, \dots é uma sequência aleatória iid.

Para uma sequência aleatória, a distribuição conjunta de um vetor amostra X_1, X_2, \dots, X_n é fácil de escrever desde que é o produto das fdp ou fmp individuais.

Teorema 8.1. *Seja X_n uma sequência aleatória iid. Para um processo de valor discreto, o vetor amostra X_{n_1}, \dots, X_{n_k} tem fmp conjunta*

$$p_{X_{n_1}, \dots, X_{n_k}}(x_1, \dots, x_k) = p_X(x_1)p_X(x_2) \cdots p_X(x_k) = \prod_{i=1}^k p_X(x_i)$$

Se o processo assume valores contínuos, então a fdp conjunta de X_{n_1}, \dots, X_{n_k} é dada por

$$f_{X_{n_1}, \dots, X_{n_k}}(x_1, \dots, x_k) = f_X(x_1)f_X(x_2) \cdots f_X(x_k) = \prod_{i=1}^k f_X(x_i)$$

De todas as sequências iid, talvez a mais simples seja a sequência aleatória de Bernoulli.

Definição 8.7. *Um processo de Bernoulli X_n com probabilidade de sucesso p é uma sequência aleatória na qual cada X_n é uma variável aleatória com distribuição de Bernoulli tal que $P[X_n = 1] = p = 1 - P[X_n = 0]$.*

Exemplo 8.4. *Para o processo do resistor do Exemplo 8.3, seja $Y_n = 1$ se no i -ésimo segundo encontramos um resistor de 1%, caso contrário $Y_n = 0$. A sequência aleatória Y_n é um processo de Bernoulli.*

Exemplo 8.5. *Para um processo de Bernoulli X_n com probabilidade de sucesso p , encontre a fmp conjunta de X_1, \dots, X_n .*

Solução. Para uma única amostra X_i , podemos escrever a fmp de Bernoulli da seguinte maneira

$$p_{X_i}(x_i) = \begin{cases} p^{x_i}(1-p)^{1-x_i}, & x_i \in \{0, 1\} \\ 0, & \text{caso contrário} \end{cases}$$

Quando $x_i \in \{0, 1\}$ para $i = 0, \dots, n$, a fmp conjunta pode ser escrita como

$$p_{X_1, \dots, X_n}(x_1, \dots, x_n) = \prod_{i=1}^n p^{x_i}(1-p)^{1-x_i} = p^k(1-p)^{n-k}$$

onde $k = x_1 + \dots + x_n$. A expressão completa para a fmp conjunta é

$$p_{X_1, \dots, X_n}(x_1, \dots, x_n) = \begin{cases} p^{x_1 + \dots + x_n}(1-p)^{n-(x_1 + \dots + x_n)}, & x_i \in \{0, 1\}, i = 1, 2, \dots, n \\ 0, & \text{caso contrário} \end{cases}$$

8.5 Processo de Contagem

Um processo de contagem $N(t)$ começa no instante $t = 0$ e conta a ocorrência de eventos. Estes eventos são em geral chamados de *chegadas* desde que os processos de contagem são mais usados para modelar a chegada de clientes a um determinado servidor.

Desde que iniciamos em $t = 0$, $n(t, s) = 0$ para todo $t \leq 0$. Ainda, o número de chegadas até um instante $t > 0$ qualquer é um número inteiro que não decresce com o tempo.

Definição 8.8. Processo de contagem. Um processo estocástico $N(t)$ é um processo de contagem se para cada função amostra, $n(t, s) = 0$ para $t \leq 0$ e $n(t, s)$ assume valores inteiros e não decrescentes com o tempo.

Podemos imaginar $N(t)$ como o número de clientes que chega a um sistema no intervalo $(0, t]$. Uma função amostra típica de um processo de contagem é mostrada na Figura 8.4. Os saltos na função amostra de um processo de contagem marcam as chegadas e o número de chegadas no intervalo $(t_0, t_1]$ é simplesmente $N(t_1) - N(t_0)$.

Podemos usar um processo de Bernoulli X_1, X_2, \dots para derivar um processo de contagem simples. Considere um intervalo de tempo de tamanho Δ de modo que exista uma chegada na intervalo $(n\Delta, (n+1)\Delta]$ se e somente se $X_n = 1$. Para uma constante arbitrária $\lambda > 0$, podemos escolher Δ pequeno o suficiente para assegurar que $\lambda\Delta < 1$. Neste caso, escolhemos a probabilidade de sucesso de X_n como sendo $\lambda\Delta$. Isto implica que o número de chegadas N_m no instante $T = m\Delta$ tem fmp binomial

$$P_{N_m}(n) = \begin{cases} \binom{m}{n} (\lambda T/m)^n (1 - \lambda T/m)^{m-n}, & n = 0, 1, \dots, m \\ 0, & \text{caso contrário} \end{cases} \quad (8.1)$$

Pode-se mostrar que à medida que $m \rightarrow \infty$, ou equivalentemente à medida que $\Delta \rightarrow 0$ a fmp de N_m aproxima-se da fmp de uma v.a. com distribuição de Poisson $N(T)$ com fmp

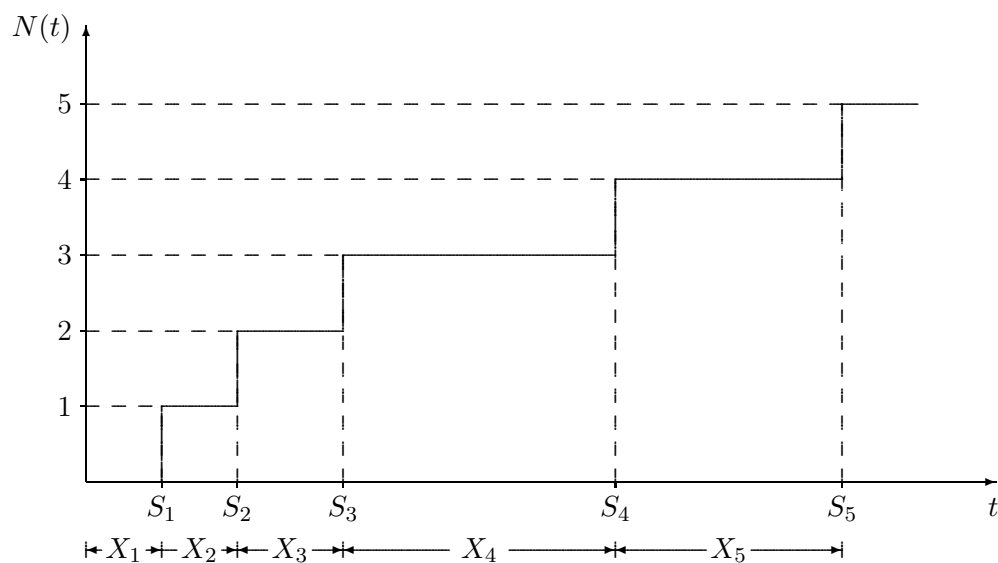


Figura 8.4: Função amostra de um processo de contagem

$$P_{N(t)}(n) = \begin{cases} (\lambda T)^n e^{-\lambda T} / n!, & n = 0, 1, 2, \dots \\ 0, & \text{caso contrário} \end{cases}$$

Podemos generalizar este argumento e dizer que para qualquer intervalo $(t_0, t_1]$, o número de chegadas poderia ter uma fmp de Poisson com parâmetro λT onde $T = t_1 - t_0$. Além disso, o número de chegadas em $(t_0, t_1]$ depende das tentativas independentes de Bernoulli correspondentes àquele intervalo. Então o número de chegadas em intervalos não sobrepostos irá ser independente. No limite à medida que $\Delta \rightarrow 0$, obtemos um processo de contagem no qual o número de chegadas em qualquer intervalo é uma variável aleatória com distribuição de Poisson independente das chegadas em qualquer outro intervalo não sobreposto. Chamamos este processo limite de um processo de Poisson. Na próxima seção iremos examinar o processo de Poisson com mais detalhes.

8.6 Processo de Poisson

Considere uma situação na qual os eventos ocorrem em instantes de tempo aleatórios a uma taxa média de λ eventos por segundo. Por exemplo, um evento poderia representar a chegada de um cliente a uma estação de serviço ou a falha de um componente em algum sistema. Seja $N(t)$ o número de ocorrências destes eventos no intervalo de tempo $[0, t]$. $N(t)$ é então um processo estocástico contínuo no tempo, não decrescente e que assume apenas valores inteiros, como mostrado na Figura 8.4.

Suponha agora que o intervalo $[0, t]$ seja dividido em n subintervalos de duração infinitesimal $\delta = t/n$. Assuma também que as seguintes condições sejam também observadas:

1. A probabilidade da ocorrência de mais de um evento em um destes subintervalos é desprezível comparada à probabilidade de observar zero ou um eventos.

2. A ocorrência de um evento em um dado subintervalo independe dos resultados observados nos outros subintervalos.

A primeira suposição implica que o resultado em cada subintervalo pode ser visto como o resultado de um teste de Bernoulli. A segunda suposição implica que estes testes de Bernoulli são independentes. Então, estas duas suposições juntas implicam que o processo de contagem $N(t)$ pode ser aproximado pelo processo de contagem binomial, que conta o número de sucessos em n testes de Bernoulli.

Se a probabilidade de ocorrência de um evento em cada subintervalo é p , então o número esperado de eventos no intervalo $[0, t]$ é np . Desde que os eventos ocorrem a uma taxa de λ eventos por segundo, o número médio de eventos no intervalo $[0, t]$ é também λt . Então devemos ter

$$\lambda t = np$$

Se fizermos agora $n \rightarrow \infty$ (isto é, $\delta \rightarrow 0$), e $p \rightarrow 0$ enquanto mantemos $\lambda t = np$ fixo, então a distribuição binomial tende a uma distribuição de Poisson com parâmetro λt . Podemos concluir então que o número de ocorrências $N(t)$ de eventos no intervalo $[0, t]$ tem uma distribuição de Poisson de média λt :

$$P[N(t) = k] = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, k = 0, 1, 2, \dots \quad (8.2)$$

Por esta razão, $N(t)$ é conhecido como processo de Poisson. Formalmente, podemos definir um processo de Poisson como:

Definição 8.9. Processo de Poisson. Um processo de contagem $N(t)$ é um processo de Poisson de taxa λ se

- O número de chegadas em qualquer intervalo $(t_0, t_1]$, $N(t_1) - N(t_0)$, é uma variável aleatória com distribuição de Poisson com valor esperado $\lambda(t_1 - t_0)$.
- Para qualquer par de intervalos não sobrepostos, $(t_0, t_1]$ e $(t'_0, t'_1]$, o número de chegadas em cada intervalo, $N(t_1) - N(t_0)$ e $N(t'_1) - N(t'_0)$ respectivamente, são variáveis aleatórias independentes.

Chamamos λ de taxa do processo pois o número esperado de chegadas por unidade de tempo é $E[N(t)]/t = \lambda$. Pela definição da variável aleatória de Poisson, $M = N(t_1) - N(t_0)$ tem fmp

$$P_M(m) = \begin{cases} \frac{[\lambda(t_1 - t_0)]^m}{m!} e^{-\lambda(t_1 - t_0)}, & m = 0, 1, \dots \\ 0, & \text{caso contrário} \end{cases} \quad (8.3)$$

Para um conjunto de instantes de tempo $t_1 < t_2 < \dots < t_k$, podemos usar a propriedade de que o número de chegadas em intervalos não sobrepostos são independentes para escrever a fmp conjunta de $N(t_1), \dots, N(t_k)$ como um produto de probabilidades.

Teorema 8.2. Para um processo de Poisson $N(t)$ de taxa λ , a fmp conjunta de $N(t_1), \dots, N(t_k)$, $t_1 < t_2 < \dots < t_k$, é dada por

$$p_{N(t_1), \dots, N(t_k)}(n_1, \dots, n_k) = \begin{cases} \frac{\alpha_1^{n_1} e^{-\alpha_1}}{n_1!} \frac{\alpha_2^{n_2 - n_1} e^{-\alpha_2}}{(n_2 - n_1)!} \dots \frac{\alpha_k^{n_k - n_{k-1}} e^{-\alpha_k}}{(n_k - n_{k-1})!}, & 0 \leq n_1 \leq \dots \leq n_k \\ 0, & \text{caso contrário} \end{cases}$$

onde $\alpha_i = \lambda(t_i - t_{i-1})$.

Demonstração. Seja $M_1 = N(t_1)$ e para $i = 2, \dots, k$, seja $M_i = N(t_i) - N(t_{i-1})$. Pela definição do processo de Poisson, M_1, \dots, M_k é uma coleção de variáveis aleatórias independentes com distribuição de Poisson tal que $E[M_i] = \alpha_i$.

$$\begin{aligned} p_{N(t_1), \dots, N(t_k)}(n_1, \dots, n_k) &= p_{M_1, M_2, \dots, M_k}(n_1, n_2 - n_1, \dots, n_k - n_{k-1}) \\ &= p_{M_1}(n_1) p_{M_2}(n_2 - n_1) \dots p_{M_k}(n_k - n_{k-1}) \end{aligned}$$

Substituindo a Equação (8.3) por $p_{M_i}(n_i - n_{i-1})$, completamos a prova. \square

Exemplo 8.6. Um sistema de mensagens gravadas recebe acessos de acordo com um processo de Poisson de taxa 15 acessos por minuto. Encontre a probabilidade de, em um intervalo de tempo de 1 minuto, 3 acessos sejam feitos nos primeiros 10 segundos e 2 acessos sejam feitos nos últimos 15 segundos.

Solução. A taxa de chegada em segundos é $\lambda = 15/60 = 1/4$ acessos por segundo. Escrevendo o tempo em segundos, a probabilidade de interesse é

$$P[N(10) = 3 \text{ e } N(60) - N(45) = 2]$$

Aplicando as propriedades de incrementos independentes e incrementos estacionários,

$$\begin{aligned} P[N(10) = 3 \text{ e } N(60) - N(45) = 2] &= P[N(10) = 3] P[N(60) - N(45) = 2] \\ &= P[N(10) = 3] P[N(60 - 45) = 2] \\ &= \frac{(10/4)^3 e^{-10/4}}{3!} \frac{(15/4)^2 e^{-15/4}}{2!} = 0,035 \end{aligned}$$

É importante lembrar que a propriedade dos intervalos independentes do processo de Poisson precisa se manter mesmo para intervalos bastante pequenos. Por exemplo, o número de chegadas em $(t, t + \delta]$ precisa ser independente do processo de chegada sobre $[0, t]$ independentemente de quão pequeno escolhamos $\delta > 0$. Essencialmente, a probabilidade de uma chegada em qualquer instante é independente da história passada do processo. Neste sentido, o processo de Poisson é *sem memória*.

Esta propriedade de ser sem memória pode também ser vista quando examinamos os instantes entre as chegadas. Como mostrado na Figura 8.4, o tempo aleatório X_n entre a chegada $n - 1$ e a chegada n é chamado de n -ésimo tempo entre chegadas. Adicionalmente, chamamos o instante X_1 , da primeira chegada, como o primeiro tempo entre chegadas, mesmo não havendo chegadas anteriores.

Teorema 8.3. *Para um processo de Poisson de taxa λ , os tempos entre chegadas X_1, X_2, \dots são uma sequência aleatória iid com fdp exponencial*

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & \text{caso contrário} \end{cases}$$

Demonstração. Dado $X_1 = x_1, X_2 = x_2, \dots, X_{n-1} = x_{n-1}$, a chegada $n - 1$ ocorre no instante

$$t_{n-1} = x_1 + \dots + x_{n-1}$$

Para $x > 0$, $X_n > x$ se e somente se $\frac{1}{2}$ se não ocorrerem chegadas no intervalo $(t_{n-1}, t_{n-1} + x]$. O número de chegadas em $(t_{n-1}, t_{n-1} + x]$ é independente da história passada descrita por X_1, \dots, X_{n-1} . Isto implica

$$P[X_n > x | X_1 = x_1, \dots, X_{n-1} = x_{n-1}] = P[N(t_{n-1} + x) - N(t_{n-1}) = 0] = e^{-\lambda x}$$

Então X_n é independente de X_1, \dots, X_{n-1} e tem fdc exponencial

$$F_{X_n}(x) = 1 - P[X_n > x] = \begin{cases} 1 - e^{-\lambda x}, & x > 0 \\ 0, & \text{caso contrário} \end{cases}$$

Tomando a derivada da fdc, podemos ver que X_n tem fdp exponencial $f_{X_n}(x) = f_X(x)$, o que demonstra o teorema. \square

Exemplo 8.7. *Encontre a média e a variância do tempo até o décimo acesso no Exemplo 8.6.*

Solução. A taxa de chegada é de $\lambda = 1/4$ acessos por segundo, de modo que os tempos entre chegadas são variáveis aleatórias com distribuição exponencial de parâmetro λ .

Para a distribuição exponencial, a média e a variância são, respectivamente, $1/\lambda$ e $1/\lambda^2$ (veja Apêndice E). O instante da décima chegada é a soma destas variáveis aleatórias iid, então

$$\begin{aligned} E[S_{10}] &= 10E[T] = \frac{10}{\lambda} = 40 \text{ segundos} \\ \text{Var}[S_{10}] &= 10 \text{Var}[T] = \frac{10}{\lambda^2} = 160 \text{ segundos}^2. \end{aligned}$$

A propriedade de ser sem memória do processo de Poisson pode também ser vista nos tempos entre chegadas exponenciais. Desde que $P[X_n > x] = e^{-\lambda x}$, a probabilidade condicional de que $X_n - x' > x$ dado que $X_n > x'$, é

$$P[X_n - x' > x | X_n > x'] = \frac{P[X_n > x + x', X_n > x']}{P[X_n > x']} = e^{-\lambda x} \quad (8.4)$$

A interpretação da Equação (8.4) é que dado que a chegada não ocorreu no instante x' , o tempo adicional até a chegada, $X_n - x'$, tem a mesma distribuição exponencial de X_n . Isto é, não importa o quanto esperamos para a chegada, o tempo restante até a chegada tem sempre uma distribuição exponencial com média $1/\lambda$.

A partir de uma função amostra de $N(t)$, podemos identificar os tempos entre chegadas X_1, X_2 e assim por diante. Similarmente, a partir dos tempos entre chegadas X_1, X_2, \dots , podemos construir a função amostra do processo de Poisson $N(t)$. Isto implica que uma representação equivalente do processo de Poisson é uma sequência aleatória iid X_1, X_2, \dots de tempos entre chegadas exponencialmente distribuídos.

Teorema 8.4. *Um processo de contagem com tempos entre chegadas exponenciais independentes X_1, X_2, \dots com média $E[X_i] = 1/\lambda$ é um processo de Poisson de taxa λ .*

8.7 Processo sinal telegráfico aleatório

Considere um processo aleatório $X(t)$ que assume os valores ± 1 . Suponha que $X(0) = \pm 1$ com probabilidade $1/2$, e suponha que $X(t)$ mude de polaridade com cada evento de um processo de Poisson de taxa α . A Figura 8.5 mostra uma função amostra de $X(t)$.

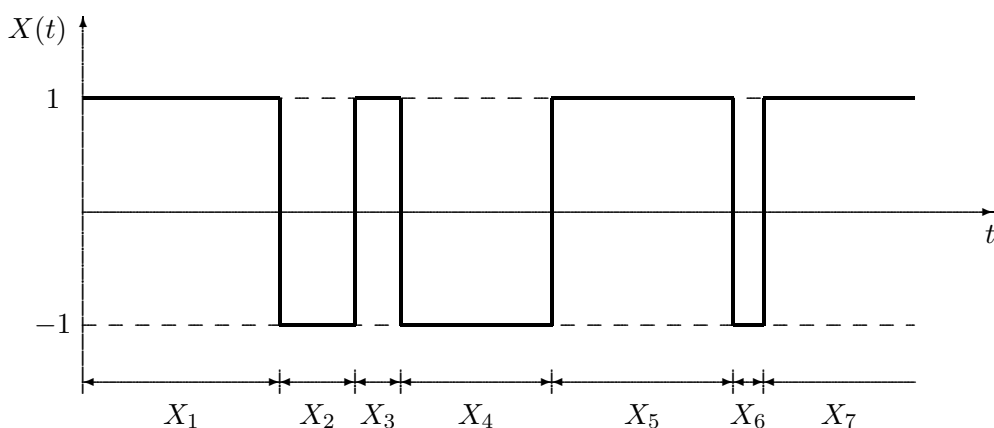


Figura 8.5: Função amostra de um processo telegráfico aleatório

A fmp de $X(t)$ é dada por

$$P[X(t) = \pm 1] = P[X(t) = \pm 1 | X(0) = 1]P[X(0) = 1] + P[X(t) = \pm 1 | X(0) = -1]P[X(0) = -1] \quad (8.5)$$

Podemos encontrar as fmp's condicionais notando que $X(t)$ irá ter a mesma polaridade de $X(0)$ somente quando ocorrer um número par de eventos no intervalo $(0, t]$. Então

$$\begin{aligned}
 P[X(t) = \pm 1 | X(0) = \pm 1] &= P[N(t) = \text{inteiro par}] \\
 &= \sum_{j=0}^{\infty} \frac{(\alpha t)^{2j}}{(2j)!} e^{-\alpha t} \\
 &= e^{-\alpha t} \frac{1}{2} \{e^{\alpha t} + e^{-\alpha t}\} \\
 &= \frac{1}{2} \{1 + e^{-2\alpha t}\}
 \end{aligned} \tag{8.6}$$

$X(t)$ e $X(0)$ irão ter sinais opostos quando o número de eventos no intervalo $(0, t]$ for ímpar:

$$\begin{aligned}
 P[X(t) = \pm 1 | X(0) = \mp 1] &= P[N(t) = \text{inteiro ímpar}] \\
 &= \sum_{j=0}^{\infty} \frac{(\alpha t)^{2j+1}}{(2j+1)!} e^{-\alpha t} \\
 &= e^{-\alpha t} \frac{1}{2} \{e^{\alpha t} - e^{-\alpha t}\} \\
 &= \frac{1}{2} \{1 - e^{-2\alpha t}\}
 \end{aligned} \tag{8.7}$$

Substituindo estes resultados na Equação (8.5), temos:

$$\begin{aligned}
 P[X(t) = 1] &= \frac{1}{2} \frac{1}{2} \{1 + e^{-2\alpha t}\} + \frac{1}{2} \frac{1}{2} \{1 - e^{-2\alpha t}\} = \frac{1}{2} \\
 P[X(t) = -1] &= 1 - P[X(t) = 1] = \frac{1}{2}
 \end{aligned} \tag{8.8}$$

Então, o sinal telegráfico assume os valores -1 e +1 com a mesma probabilidade. A média e a variância de $X(t)$ são dadas por:

$$E[X(t)] = 1P[X(t) = 1] + (-1)P[X(t) = -1] = 0$$

$$\text{Var}[X(t)] = E[X^2(t)] - E^2[X(t)] = (1)^2 P[X(t) = 1] + (-1)^2 P[X(t) = -1] - 0 = 1$$

E a função de autocorrelação de $X(t)$ é dada por:

$$\begin{aligned}
 C_X(t_1, t_2) &= E[X(t_1), X(t_2)] \\
 &= 1P[X(t_1) = X(t_2)] + (-1)P[X(t_1) \neq X(t_2)] \\
 &= \frac{1}{2} \left\{ 1 + e^{-2\alpha|t_2-t_1|} \right\} - \frac{1}{2} \left\{ 1 - e^{-2\alpha|t_2-t_1|} \right\} \\
 &= e^{-2\alpha|t_2-t_1|}
 \end{aligned}$$

Pode-se ver então que as amostras de $X(t)$ tornam-se cada vez menos correlacionadas à medida que o tempo entre elas aumenta.

8.8 Processo movimento Browniano

O processo de Poisson e o processo telegráfico são exemplos de processos de tempo contínuo e valor discreto. Agora iremos examinar o processo movimento Browniano, que é um processo de tempo e valores contínuos.

Definição 8.10. Processo movimento Browniano. Um processo movimento Browniano $X(t)$ tem a propriedade de que $X(0) = 0$ e para $\tau > 0$, $X(t + \tau) - X(t)$ é uma variável aleatória gaussiana com média 0 e variância $\alpha\tau$ que é independente de $X(t')$ para todo $t \leq t'$.

Para o movimento Browniano, podemos ver $X(t)$ como a posição de uma partícula em uma linha. Para um pequeno incremento de tempo δ ,

$$X(t + \delta) = X(t) + [X(t + \delta) - X(t)] \quad (8.9)$$

Embora esta expansão possa parecer trivial, pela definição de movimento Browniano, o incremento $Y_\delta = X(t + \delta) - X(t)$, é independente de $X(t)$ e é Gaussiano de média zero e variância $\alpha\delta$. Esta propriedade do movimento Browniano é chamada de *incrementos independentes*. Então, depois de um intervalo de tempo δ , a posição da partícula moveu-se de uma quantidde Y_δ que é independente da posição anterior $X(t)$. A mudança de posição Y_δ pode ser positiva ou negativa. Por esta razão, o movimento Browniano é também chamado de uma caminhada aleatória unidimensional.

Teorema 8.5. Para o processo movimento Browniano $X(t)$, a fdp conjunta de $X(t_1), \dots, X(t_k)$ é

$$f_{X(t_1), \dots, X(t_k)}(x_1, \dots, x_k) = \prod_{n=1}^k \frac{1}{\sqrt{2\pi\alpha(t_n - t_{n-1})}} e^{-(x_n - x_{n-1})^2 / [2\alpha(t_n - t_{n-1})]} \quad (8.10)$$

Demonstração. Desde que $X(0) = 0$, $X(t_1) = X(t_1) - X(0)$ é uma variável aleatória Gaussiana. Dados os instantes de tempo t_1, \dots, t_k , definimos $t_0 = 0$ e, para $n = 1, \dots, k$, $Y_n = X(t_n) - X(t_{n-1})$. Note que Y_1, \dots, Y_k são variáveis aleatórias Gaussianas independentes de média zero tais que Y_n é $N(0, \alpha(t_n - t_{n-1}))$.

$$f_{Y_n}(y) = \frac{1}{\sqrt{2\pi\alpha(t_n - t_{n-1})}} e^{-y^2 / [2\alpha(t_n - t_{n-1})]} \quad (8.11)$$

Note que $X(t_1) = x_1, \dots, X(t_n) = x_n$ se e somente se $Y_1 = x_1, Y_2 = x_2 - x_1, \dots, Y_k = x_k - x_{k-1}$.

Depois de alguma manipulação, chegamos a

$$f_{X(t_1), \dots, X(t_k)}(x_1, \dots, x_k) = \prod_{n=1}^k f_{Y_n}(x_n - x_{n-1}) \quad (8.12)$$

Substituindo (8.11) em (8.12), completamos a prova. \square

8.9 Médias estatísticas de processos aleatórios

Assim como definimos médias estatísticas para v.a.'s, podemos de forma similar, definir médias estatísticas para um processo estocástico. Tais médias são também chamadas de **médias de conjunto**. Estas serão então utilizadas para especificar um processo aleatório.

Se adotarmos a idéia de que um processo aleatório $X(t)$ é uma v.a. X que é uma função do tempo, chegaremos à conclusão de que $X(t)$ é completamente especificada se a fdp de X é especificada para cada valor de t . Iremos ver rapidamente que as coisas não são assim tão simples. Entretanto, comecemos com a idéia de especificar uma v.a. X para cada valor de t .

Para o processo estocástico $X(t)$ representando a temperatura de uma cidade, isto implicará em considerar as amplitudes das funções amostra em algum instante $t = t_1$. O valor $X(t_1; \zeta_1)$ representa a temperatura no instante t_1 no ζ_i -ésimo dia e é o resultado da ζ_i -ésima tentativa. Então, todas as amplitudes das funções amostra em $t = t_1$ representam valores tomados pela v.a. X em $t = t_1$, isto é $X(t_1)$.

Podemos fazer isto para cada valor de t . A fdp de X pode ser diferente para diferentes valores de t em geral. Para indicar este fato, a fdp de X no instante t é expressa por $f_X(x; t)$.

Exemplo 8.8. Limiar de detecção. Sobre um canal binário, as mensagens $m = 0$ e $m = 1$ são transmitidas com probabilidades iguais usando um pulso positivo e negativo, respectivamente. O pulso transmitido correspondente à mensagem 1 é $p(t)$, mostrado na Figura 8.6, e o pulso transmitido correspondente à mensagem 0 será $-p(t)$. Determine as estatísticas de primeira ordem.

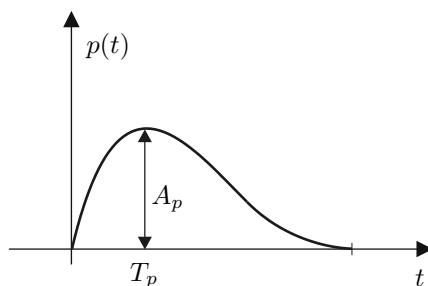


Figura 8.6: Forma de onda do pulso $p(t)$.

Solução. Seja A_p a amplitude de pico de $p(t)$ em $t = T_p$. Por causa do ruído de canal $n(t)$, os pulsos recebidos serão $\pm p(t) + n(t)$ (Figura 8.7).

Para detectar os pulsos no receptor, cada pulso é amostrado em sua amplitude de pico. Na ausência de ruído, a saída do amostrador é A_p (para $m = 1$) ou $-A_p$ (para $m = 0$). Por causa do ruído do canal, a saída do amostrador é $\pm A_p + n$, onde n , a amplitude do sinal de ruído no instante de amostragem, é uma v.a. No receptor, utiliza-se um limiar de detecção igual a 0, isto é, se o pulso amostrado em seu pico tem um valor positivo, a decisão é 1, e se é menor que 0, a decisão é 0.

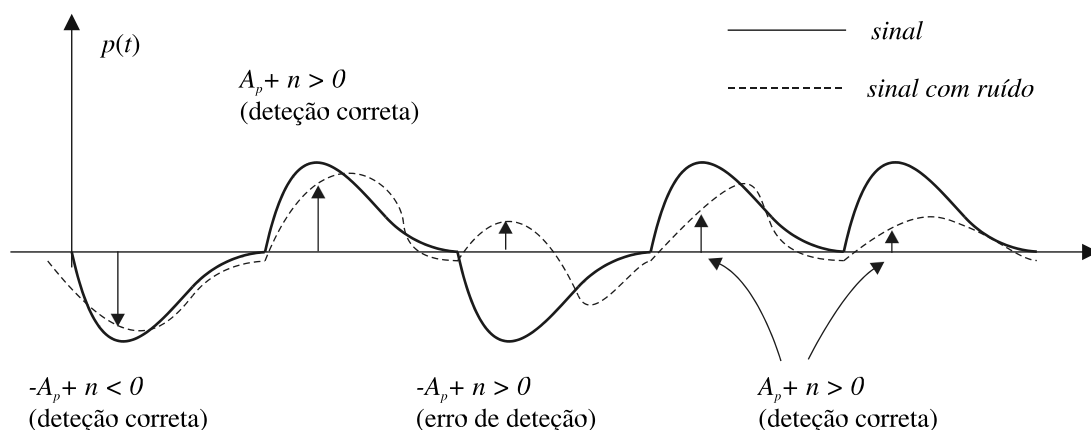


Figura 8.7: Erro de detecção devido ao ruído.

Vamos tentar interpretar $P[\varepsilon|1]$, a probabilidade de erro dado que 1 foi transmitido. Se 1 é transmitido, a saída do amostrador no receptor é $A_p + n$. Se $A_p + n > 0$, fazemos uma decisão correta, e se $A_p + n < 0$, ou equivalentemente $n < -A_p$, tomamos uma decisão errada.

Interpretando a probabilidade em termos de frequência relativa, se repetimos o experimento (de transmitir e receber o símbolo 1) N vezes ($N \rightarrow \infty$), e se N_ε vezes a amostra do ruído foi negativa o suficiente para que $A_p + n < 0$, então

$$P[\varepsilon|1] = \frac{N_\varepsilon}{N}$$

Vamos examinar o sinal de ruído no instante t_s . Em cada tentativa, temos um novo sinal de ruído (função amostra) do conjunto de ruído e um valor diferente de n no instante de amostragem t_s , e se $n < -A_p$ digamos 100 vezes em 100 milhões de tentativas, então a probabilidade de erro é dada por $P[\varepsilon|1] = 100/100 \cdot 10^6 = 10^{-6}$. Mas este número é também a probabilidade de $n < -A_p$, onde n é uma v.a. formada pelas amplitudes em $t = t_s$ das funções amostra do conjunto do processo estocástico $n(t)$. Esta é a v.a. $n(t_s)$ cuja fdp é $f_n(n; t_s)$.

A fdp $f_X(x; t)$ é conhecida como fdp de primeira ordem. Infelizmente o conhecimento da fdp de primeira ordem é insuficiente para especificar um processo aleatório. Para entender o porque, um exemplo é bastante instrutivo.

Exemplo 8.9. Seja um processo estocástico $X(t)$ cujo conjunto é mostrado na Figura 8.8a). Suponha que a distribuição de amplitudes em qualquer instante t é a mesma, isto é $f_X(x; t)$ é independente de t , e $f_X(x; t) = f_X(x)$, como mostrado na Figura 8.9.

Se comprimirmos no tempo o processo $X(t)$ por um fator k ($k > 1$), formamos um novo processo $Y(t)$, como mostrado na Figura 8.8b). Verifique porque as estatísticas de primeira ordem não são suficientes para diferenciar $X(t)$ e $Y(t)$.

Solução. Pode-se ver facilmente que a distribuição de amplitudes de $X(t)$ é idêntica à de $Y(t)$ e, desta forma, a fdp de primeira ordem de $X(t)$ é idêntica à de $Y(t)$.

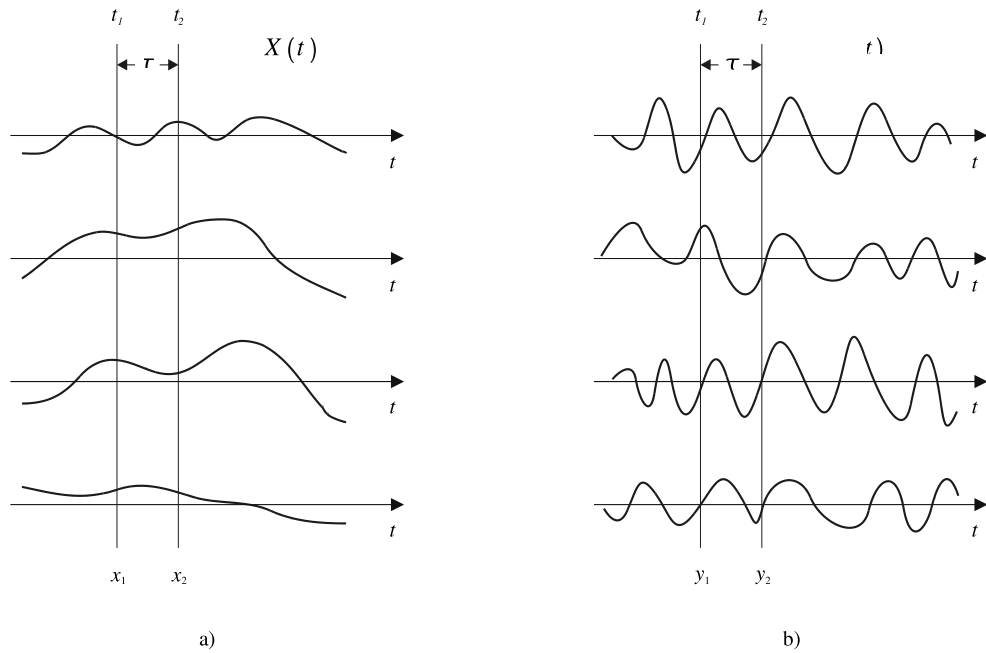


Figura 8.8: Processo estocástico comprimido no tempo.

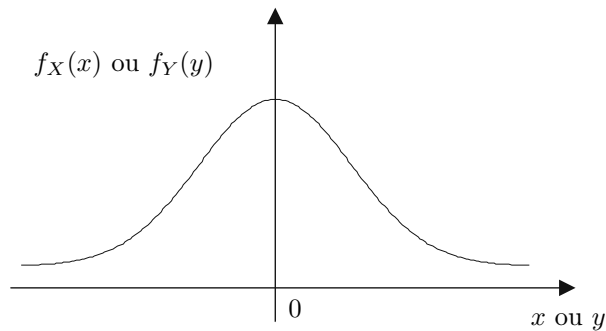


Figura 8.9: Fdp dos processos x e y .

Entretanto estes processos são bastante diferentes entre si pois o processo $Y(t)$ contém componentes em frequências mais altas do que as de $X(t)$. De fato, o espectro de $Y(t)$ será o espectro de $X(t)$ expandido por um fator k .

Este exemplo mostra claramente que a fdp de primeira ordem não é suficiente para especificar completamente um processo estocástico. O conteúdo de frequências de um processo depende da velocidade com que as amplitudes variam com o tempo. Isto pode ser medido correlacionando amplitudes em t_1 e $t_1 + \tau$. Se o processo varia lentamente, as amplitudes em t_1 e $t_1 + \tau$ devem ser similares (Figura 8.8a). Por outro lado, se o processo varia rapidamente, as amplitudes em t_1 e $t_1 + \tau$ não terão nenhuma semelhança (Figura 8.8b). Podemos usar a correlação para medir a similaridade das amplitudes em t_1 e $t_2 = t_1 + \tau$.

Definição 8.11. Função de autocorrelação. Se as variáveis aleatórias $X(t_1)$ e $X(t_2)$ são denotadas por X_1 e X_2 respectivamente, então para um processo estocástico real, a função de autocorrelação $R_X(t_1, t_2)$ é definida como

$$R_X(t_1, t_2) = E[X(t_1)X(t_2)]$$

Esta é a correlação das v.a.'s $X(t_1)$ e $X(t_2)$ e é calculada multiplicando-se as amplitudes em t_1 e t_2 de uma função amostra e então fazendo a média deste produto sobre o conjunto.

No exemplo acima, pode-se ver que para um valor pequeno de τ , o produto X_1X_2 será positivo para a maioria das funções amostra de $X(t)$, mas o produto Y_1Y_2 poderá ser tanto positivo como negativo. Desta forma, $E[X_1X_2]$ será maior que $E[Y_1Y_2]$. Além disso, X_1 e X_2 irão mostrar correlação para valores de τ consideravelmente grandes, enquanto que Y_1 e Y_2 irão perder a correlação rapidamente mesmo para valores pequenos de τ , como mostrado na Figura 8.10.

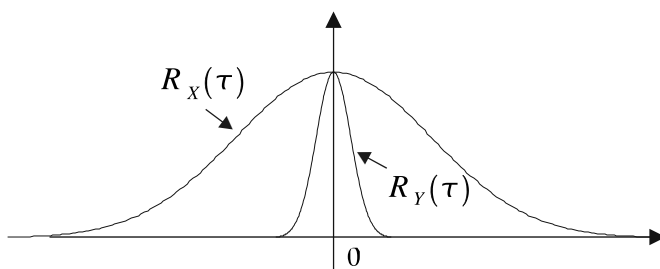


Figura 8.10: Funções de autocorrelação para os processos $X(t)$ e $Y(t)$.

Então $R_X(t_1, t_2)$, a função de autocorrelação de $X(t)$, fornece informações importantes sobre o conteúdo de freqüências do processo, e pode ser derivada da fdp conjunta de X_1 e X_2 . Esta é a fdp de segunda ordem.

Em resumo, para especificar um processo aleatório, precisamos não só da fdp de primeira ordem mas também da fdp de segunda ordem.

Em geral, precisamos da medida de interdependência de n variáveis x_1, x_2, \dots, x_n nos instantes t_1, t_2, \dots, t_n . Esta informação é fornecida pela fdp de ordem n , $f_{X_1X_2\dots X_n}(x_1, x_2, \dots, x_n)$. A determinação desta fdp é uma tarefa formidável, mas felizmente, na maioria dos casos, iremos precisar apenas das estatísticas de primeira e segunda ordens.

Podemos sempre derivar a fdp de ordem inferior a partir de fdp's de ordem superior por simples integração. Por exemplo,

$$f_{X_1}(x_1) = \int_{-\infty}^{+\infty} f_{X_1X_2}(x_1, x_2) dx_2$$

Então, quando temos a fdp de ordem n , não é necessário especificar as fdp's de ordem menor que n .

8.9.1 Momentos

Definição 8.12. *Seja um processo estocástico $X(t)$ e seja também $X_{t_i} \equiv X(t_i)$. O n -ésimo momento da v.a. X_{t_i} é definido como*

$$E[X_{t_i}^n] = \int_{-\infty}^{\infty} x_{t_i}^n f_X(x_{t_i}) dx_{t_i}$$

O primeiro momento é chamado de média de um processo estocástico, e é definido como

Definição 8.13. *A **média** $E[X(t)]$ de um processo estocástico pode ser determinada da fdp de primeira ordem usando a seguinte expressão*

$$E[X(t)] = \int_{-\infty}^{+\infty} x f_X(x; t) dx$$

Em geral, o valor do n -ésimo momento irá depender do instante de tempo t_i se a fdp de X_{t_i} depender de t_i . Quando o processo é estacionário, entretanto, $f_X(x_{t_i}) = f_X(x_{t_i} + t)$ para todo t . Portanto, a fdp é independente do tempo e, como consequência, o n -ésimo momento também o é.

8.9.2 Função de autocovariância

Definição 8.14. *A função de autocovariância é definida como*

$$K_X(t_1, t_2) = E[(X_{t_1} - m(t_1))(X_{t_2} - m(t_2))] = R_X(t_1, t_2) - m(t_1)m(t_2)$$

onde $m(t_1)$ e $m(t_2)$ são as médias de X_{t_1} e X_{t_2} , respectivamente.

Momentos conjuntos de ordem superior de duas ou mais v.a.'s derivadas de um processo estocástico são definidos da mesma maneira. Entretanto, com a possível exceção do processo gaussiano, para o qual os momentos de ordem superior podem ser expressos em termos do primeiro e segundo momentos, estes momentos de ordem superior são encontrados com pouca frequência na prática.

8.10 Classificação dos processos estocásticos

8.10.1 Processos estocásticos estacionários e não estacionários

Definição 8.15. *Um processo estocástico cujas características estatísticas não variam com o tempo é classificado como um **processo estocástico estacionário**. Para um processo estacionário, podemos dizer que uma mudança da origem de tempo será impossível de detectar; o processo irá parecer o mesmo.*

Suponha que determinemos $f_{X_1}(x_1; t_1)$, desloquemos a origem para t_0 e calculemos novamente $f_{X_1}(x_1; t_1)$. O instante t_0 na nova referência é dado por $t_2 = t_1 + t_0$ na referência antiga. Desta forma, as fdp's de X em t_1 e $t_2 = t_1 + t_0$ precisam ser as mesmas. Portanto, para um processo estacionário, $f_{X_1}(x_1; t_1)$ e $f_{X_2}(x_2; t_2)$ precisam ser idênticas. Isto é possível somente se $f_{X_1}(x_1; t_1)$ é independente de t . Então, a densidade de primeira ordem de um processo estocástico estacionário pode ser expressa como

$$f_X(x; t) = f_X(x)$$

De forma similar, podemos ver que a função de autocorrelação $R_X(t_1, t_2)$ precisa ser função apenas de $t_2 - t_1$. Desta forma, para um processo estacionário real

$$R_X(t_1, t_2) = R_X(t_1 - t_2) = R_X(\tau), \quad \tau = t_1 - t_2$$

$$R_X(\tau) = E[X(t)X(t + \tau)] \quad (8.13)$$

Também a função de autocovariância pode ser simplificada para

$$K_X(t_1, t_2) = K_X(t_1 - t_2) = K_X(\tau) = R_X(\tau) - E^2[X]$$

onde $\tau = t_1 - t_2$.

Para um processo estacionário, a função densidade de probabilidade conjunta para x_1 e x_2 precisa também depender somente de $t_2 - t_1$. Similarmente, funções densidade de probabilidade de ordem mais alta tais como $f_{X_1 X_2 \dots X_n}(x_1, x_2, \dots, x_n)$ onde $x_1 = X(t_i)$, são todas independentes da escolha da origem.

Exemplo 8.10. *O processo aleatório $X(t)$ que representa a temperatura de uma cidade é um exemplo de processo estocástico não estacionário, pois as estatísticas da temperatura (valor médio por exemplo) dependem da hora do dia. Por outro lado, um processo estocástico representado por um ruído branco é um processo estacionário, porque suas estatísticas não se alteram com o tempo.*

Em geral não é fácil determinar se um processo é estacionário, pois isto envolve a investigação das estatísticas de ordem n ($n \rightarrow \infty$). Na prática, podemos determinar a estacionariedade se não houver mudanças no mecanismo de geração do sinal.

8.10.2 Processos estacionários no sentido amplo

Um processo pode não ser estacionário no sentido estrito, mas pode ainda apresentar estacionariedade para as estatísticas de primeira e segunda ordem. Quando isto acontece, temos um processo estocástico estacionário no sentido amplo. Abaixo tem-se uma definição formal.

Definição 8.16. *Processos estacionários no sentido amplo (ou fracamente estacionários) são aqueles que têm um valor médio e uma função de autocorrelação que são independentes de deslocamento na origem de tempo, ou seja*

$$E[X(t)] = \text{constante}$$

$$R_X(t_1, t_2) = R_X(\tau), \quad \tau = t_1 - t_2$$

Note que a estacionariedade é um condição muito mais forte do que a estacionariedade no sentido amplo: todos os processos estacionários são estacionários no sentido amplo, mas o inverso não é necessariamente verdade.

Assim como não existem sinais senoidais na prática, não existem também processos estacionários. Todos os processos reais são não estacionários desde que têm duração finita, isto é, têm um início e um final. Um processo estacionário precisaria iniciar em $t = -\infty$ e durar para sempre. Entretanto, muitos processos apresentam-se como estacionários para o intervalo de tempo de interesse, e a suposição de estacionariedade permite que usemos um modelo matemático tratável.

Exemplo 8.11. *Mostre que o processo aleatório $X(t) = A \cos(\omega_c t + \theta)$, onde θ é uma v.a. uniformemente distribuída na faixa $(0, 2\pi)$, é um processo estacionário no sentido amplo.*

Solução. O conjunto da Figura 8.11 consiste de senóides de amplitude A e frequência ω_c constantes, mas a fase θ é aleatória. Para qualquer função amostra a fase pode ter qualquer valor no intervalo $(0, 2\pi)$, com distribuição uniforme.

Pelo fato de θ ser uma v.a. uniformemente distribuída sobre a faixa $(0, 2\pi)$, podemos determinar $f_X(x; t)$ e, portanto, $E[X(t)]$.

Para este caso particular, entretanto, $E[X(t)]$ pode ser determinada diretamente:

$$E[X(t)] = AE[\cos(\omega_c t + \theta)]$$

E como $\cos(\omega_c t + \theta)$ é uma função de uma v.a. θ , temos

$$E[\cos(\omega_c t + \theta)] = \int_0^{2\pi} \cos(\omega_c t + \theta) f_\theta(\theta) d\theta$$

mas $f_\theta(\theta) = 1/2\pi$ no intervalo $(0, 2\pi)$ e 0 fora dele, de modo que podemos reescrever a equação acima como

$$E[\cos(\omega_c t + \theta)] = \frac{1}{2\pi} \int_0^{2\pi} \cos(\omega_c t + \theta) d\theta$$

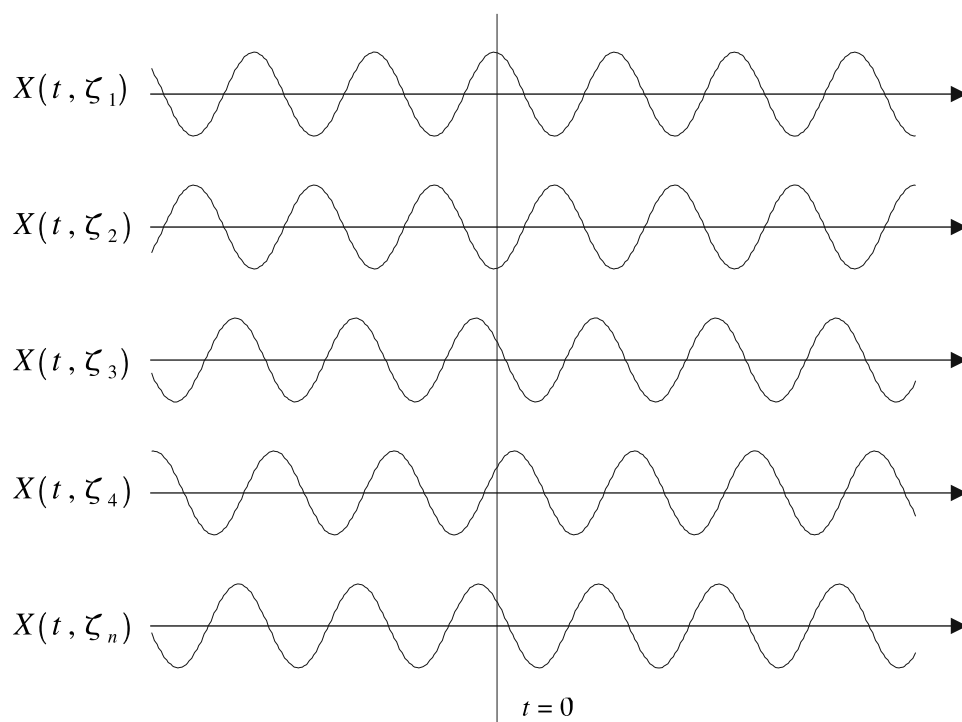


Figura 8.11: Processo aleatório $X(t) = A \cos(\omega_c t + \theta)$.

E portanto

$$E[X(t)] = 0$$

Desta forma, a média do conjunto das amplitudes das funções amostra em qualquer instante t é zero.

A função de autocorrelação para este processo pode também ser determinada diretamente a partir da Equação 8.13

$$\begin{aligned} R_X(t_1, t_2) &= E[A^2 \cos(\omega_c t_1 + \theta) \cos(\omega_c t_2 + \theta)] \\ &= A^2 E[\cos(\omega_c t_1 + \theta) \cos(\omega_c t_2 + \theta)] \\ &= \frac{A^2}{2} E[\cos[\omega_c(t_1 - t_2)] + \cos[\omega_c(t_1 + t_2) + 2\theta]] \end{aligned}$$

onde usamos a seguinte propriedade: $\cos A \cos B = \frac{1}{2} [\cos(A - B) + \cos(A + B)]$.

O primeiro termo do lado direito não contém v.a.'s, e desta forma

$$E[\cos[\omega_c(t_1 - t_2)]] = \cos[\omega_c(t_1 - t_2)]$$

O segundo termo é função da v.a. θ , e sua média é

$$E[\cos[\omega_c(t_1 + t_2) + 2\theta]] = \frac{1}{2\pi} \int_0^{2\pi} \cos[\omega_c(t_1 + t_2) + 2\theta] d\theta = 0$$

Portanto,

$$R_X(t_1, t_2) = \frac{A^2}{2} \cos[\omega_c(t_1 - t_2)]$$

ou

$$R_X(\tau) = \frac{A^2}{2} \cos(\omega_c \tau), \quad \tau = t_1 - t_2$$

E portanto $X(t)$ é um processo estacionário no sentido amplo.

Propriedades da função de autocorrelação para processos estacionários no sentido amplo:

1. A função de autocorrelação é par: $R_X(\tau) = R_X(-\tau)$

$$\textit{Demonstração. } R_X(\tau) = E[X(t)X(t+\tau)] \Rightarrow R_X(-\tau) = E[X(t)X(t-\tau)]$$

$$\text{Fazendo } \sigma = t - \tau, \text{ temos } R_X(-\tau) = E[X(\sigma + \tau)X(\sigma)] = R_X(\tau)$$

□

2. $R_X(0) = E[X^2]$

$$\textit{Demonstração. } R_X(0) = E[X(t)X(t+0)] = E[X(t)X(t)] = E[X^2(t)] = E[X^2]$$

□

3. $R_X(0) \geq 0$

$$\textit{Demonstração. } R_X(0) = E[X(t)X(t)] = E[X^2(t)].$$

$$\text{Desde que } X^2(t) \geq 0, E[X^2(t)] \geq 0, \text{ devemos ter } E[X^2(t)] \geq 0.$$

□

4. Se $Z(t) = X(t) + Y(t)$ então $R_Z(\tau) = R_X(\tau) + R_Y(\tau) + R_{XY}(\tau) + R_{YX}(\tau)$

Demonstração.

$$\begin{aligned} R_Z(\tau) &= E[Z(t)Z(t+\tau)] \\ &= E[(X(t) + Y(t))(X(t+\tau) + Y(t+\tau))] \\ &= E[X(t)X(t+\tau) + X(t)Y(t+\tau) + Y(t)X(t+\tau) + Y(t)Y(t+\tau)] \\ &= R_X(\tau) + R_{XY}(\tau) + R_{YX}(\tau) + R_Y(\tau) \end{aligned}$$

□

5. Se um processo estocástico tem um componente periódico, então a função de autocorrelação também é periódica:

$$X(t) = X(t + nT) \Rightarrow R_X(\tau) = R_X(\tau + nT)$$

6. Se $X(t)$ não tem componentes periódicos, então

$$\lim_{\tau \rightarrow \infty} R_X(\tau) = E^2[X]$$

Demonstração. A demonstração é bastante complexa, mas podemos dar uma justificativa plausível: se $X(t)$ não tem componentes periódicos, então podemos considerar que as variáveis aleatórias em $X(t)$ e $X(t + \tau)$, $\tau \rightarrow \infty$ são independentes. Desta forma:

$$\lim_{\tau \rightarrow \infty} R_X(\tau) = E[X(t)X(t + \tau)] = E[X(t)]E[X(t + \tau)] = E^2[X]$$

□

7. $R_X(0) \geq |R_X(\tau)|$.

$$\textit{Demonstração. } E[(X(t) \pm X(t + \tau))^2] \geq 0$$

Expandindo o quadrado, temos

$$E[X^2(t) \pm 2X(t)X(t + \tau) + X^2(t + \tau)] \geq 0$$

que pode ser reescrita em termos da função de autocorrelação como

$$2R_X(0) \pm 2R_X(\tau) \geq 0 \Rightarrow R_X(0) \geq |R_X(\tau)|$$

□

8.10.3 Processos ergódicos

Até agora estudamos a média e a função de autocorrelação de um processo aleatório. Estas são médias de conjunto de algum tipo. Por exemplo, $\overline{X(t)}$ é a média de conjunto das amplitudes das funções amostra em t , e é a média de conjunto do produto das amplitudes das funções amostra $X(t_1)$ e $X(t_2)$. Podemos também definir médias temporais para cada função amostra.

Definição 8.17. A média temporal, $\overline{\overline{X(t, \zeta_i)}}$, de uma função amostra $X(t, \zeta_i)$ é dada por

$$\overline{\overline{X(t, \zeta_i)}} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} X(t, \zeta_i) dt$$

Similarmente, temos

Definição 8.18. A função de autocorrelação temporal $\mathcal{R}_X(\tau, \zeta_i)$ é dada por

$$\mathcal{R}_X(\tau, \zeta_i) = \overline{\overline{X(t, \zeta_i)X(t + \tau, \zeta_i)}} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} X(t, \zeta_i) X(t + \tau, \zeta_i) dt$$

Agora temos condições para definir um processo ergódico.

Definição 8.19. Processos ergódicos são aqueles para os quais as médias de conjunto são iguais às médias temporais de qualquer função amostra. Então para um processo ergódico $X(t)$

$$E[X(t)] = \overline{\overline{X(t, \zeta_i)}}$$

$$R_X(\tau) = \mathcal{R}_X(\tau, \zeta_i)$$

Estas são apenas duas das médias possíveis. Para um processo ergódico, todas as possíveis médias de conjunto são iguais às médias temporais correspondentes de uma de suas funções amostra. Pelo fato de uma média temporal não poder ser uma função do tempo, é evidente que um processo ergódico é necessariamente um processo estacionário, mas o inverso não é verdadeiro. Na Figura 8.12 tem-se um diagrama com a classificação dos processos estocásticos quanto à estacionariedade e ergodicidade.



Figura 8.12: Classificação dos processos estocásticos.

A exemplo da estacionariedade, é difícil testar se um processo é ergódico ou não, pois precisamos testar as médias de conjunto e temporais para todas as ordens possíveis. Contudo, na prática muitos dos processos estacionários são usualmente ergódicos com relação pelo menos às estatísticas de segunda ordem, tais como a média e a função de autocorrelação.

Exemplo 8.12. Mostre que o processo do exemplo anterior é ergódico para estatísticas de até segunda ordem.

Solução.

$$\begin{aligned}\overline{\overline{X(t)}} &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} X(t) dt \\ &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} A \cos(\omega_c t + \theta) dt \\ &= 0\end{aligned}$$

$$\begin{aligned}\mathcal{R}_X(\tau) &= \overline{\overline{X(t)X(t+\tau)}} \\ &= \overline{A^2 \cos(\omega_c t + \theta) \cos(\omega_c(t+\tau) + \theta)} \\ &= \frac{A^2}{2} \left[\overline{\cos(2\omega_c t + \omega_c \tau + 2\theta) + \cos(-\omega_c \tau)} \right] \\ &= \frac{A^2}{2} \left[\overline{\cos(2\omega_c t + \omega_c \tau + 2\theta)} + \cos(\omega_c \tau) \right] \\ &= \frac{A^2}{2} \cos(\omega_c \tau)\end{aligned}$$

Exemplo 8.13. O conceito de ergodicidade pode ser explicado por um exemplo simples de semáforos de trânsito em uma cidade.

Suponha que uma cidade é bem planejada, com todas as suas ruas nas direções norte-sul e leste-oeste, e com semáforos em cada interseção. Assuma que cada semáforo permaneça verde 0,75 minutos na direção leste-oeste e 0,25 minutos na direção norte-sul, e que a mudança em um semáforo é independente de outro.

Se consideramos uma certa pessoa dirigindo um carro e que chega a um semáforo aleatoriamente na direção leste-oeste a probabilidade de encontrar um farol verde será de 0,75, ou seja, na média, 75% do tempo ele irá observar uma luz verde.

Por outro lado, se considerarmos um grande número de motoristas que chegam aleatoriamente em um semáforo na direção leste-oeste simultaneamente em algum instante t , então 75% dos motoristas irá encontrar um farol verde, e os 25% restantes irão encontrar um farol vermelho.

Então, a experiência de um único motorista chegando aleatoriamente várias vezes em um farol irá conter a mesma informação estatística (estatísticas de funções amostra) da experiência de um grande número de motoristas chegando simultaneamente em vários semáforos (estatísticas de conjunto para um dado instante).

A noção de ergodicidade é extremamente importante, porque na prática não temos um grande número de funções amostra disponível para calcular estatísticas de conjunto. Se sabemos que um processo é ergódico, então precisamos apenas de uma função amostra para calcular as estatísticas de conjunto.

8.11 Exercícios

1. Seja o processo estocástico definido por

$$x(t) = ax + b$$

onde b é uma constante e a é uma variável aleatória uniformemente distribuída na faixa $(0,100)$.

- (a) Esboce o conjunto deste processo.
- (b) Apenas observando o conjunto, determine se este é um processo estocástico estacionário ou não estacionário. Justifique sua resposta.

2. Desenhe algumas funções amostra do processo estocástico definido por

$$x(t) = A \cos(\omega t + \phi)$$

- (a) Se A é a variável aleatória uniformemente distribuída na faixa $(-1,1)$.
- (b) Se ω é a variável aleatória uniformemente distribuída na faixa $(0,10)$.
- (c) Se ϕ é a variável aleatória uniformemente distribuída na faixa $(-\pi, \pi)$.

3. Mostre que para o processo

$$x(t) = k \cos(\omega_0 t + \theta)$$

onde θ é uma variável aleatória uniformemente distribuída sobre o intervalo $(0, 2\pi)$, a função de autocorrelação temporal é dada por

$$R_X(\tau) = \overline{x(t)x(t+\tau)} = \frac{k^2}{2} \cos(\omega_0 \tau)$$

4. Para o processo estocástico

$$x(t) = \zeta \sin(\omega t + \varphi)$$

onde ω e φ são constantes e ζ é uma variável aleatória qualquer:

- (a) Calcule a média, a função de autocorrelação e a função de autocovariância.
- (b) Este processo é estacionário no sentido amplo?

Resp:

- (a) $E[X(t)] = E[\zeta] \sin(\omega t + \varphi)$
 $R_X(t_1, t_2) = E[\zeta^2] \sin(\omega t_1 + \varphi) \sin(\omega t_2 + \varphi)$
 $K_X(\tau) = \sigma_\zeta^2 \sin(\omega t_1 + \varphi) \sin(\omega t_2 + \varphi)$
- (b) não

5. Encontre uma expressão para $E[(X_{t_2} - X_{t_1})^2]$ em termos da função de autocorrelação.

$$\text{Resp: } E[(X_{t_2} - X_{t_1})^2] = R_X(t_2, t_2) - 2R_X(t_2, t_1) + R_X(t_1, t_1)$$

6. No receptor de um rádio AM, o sinal recebido contém uma portadora cossenoidal com frequência f_c e fase aleatória Θ que é uniformemente distribuída no intervalo $[0, 2\pi]$. O sinal de portadora recebido é

$$X(t) = A \cos(2\pi f_c t + \Theta)$$

- (a) Determine o valor esperado e a função de autocorrelação do processo $X(t)$.
- (b) Este processo é estacionário no sentido amplo?

Resp:

- (a) $E[X(t)] = 0$
 $R_X(t_1, t_2) = \frac{A^2}{2} \cos(2\pi f_c(t_1 - t_2))$
- (b) sim

7. Seja o processo estocástico

$$x(t) = A \cos(\omega t + \theta)$$

onde ω e θ são constantes e A é uma variável aleatória uniformemente distribuída no intervalo $(-1, 1)$.

- (a) Esboce o conjunto deste processo.
- (b) Apenas pela observação do conjunto, determine se este processo é estacionário ou não estacionário. Justifique sua resposta.

Resp:

- (a)
- (b) Não. Em $\omega t + \theta = \pi/2 + n\pi$, o processo vale 0. Nos demais pontos, vale $A \cos(\omega t + \theta)$

8. Seja o processo estocástico definido por

$$X(t) = A \cos \omega t + B \sin \omega t$$

onde A e B são v.a.'s iid de média zero.

- (a) Mostre que $X(t)$ é estacionário no sentido amplo.
- (b) Mostre que $X(t)$ não é estacionário no sentido estrito. *Dica: Considere $E[X^3(t)]$.*

Resp:

- (a) $E[X(t)] = 0$
 $R_X(t_1, t_2) = E[A^2] \cos(\omega(t_1 - t_2))$
- (b) $E[X^3(t)] = E[A^3](\cos^3(\omega t) + \sin^3(\omega t))$

9. Seja um processo estocástico $X(t)$ dado por

$$X(t) = Y \cos \omega t, \quad t \geq 0$$

onde ω é uma constante e Y é uma v.a. distribuída uniformemente no intervalo $(0, 1)$. Para este processo, calcule:

- (a) A média $E[X(t)]$.
- (b) A função de autocorrelação $R_X(t_1, t_2)$.
- (c) A função de autocovariância $K_X(t_1, t_2)$.
- (d) Este processo é estacionário?

Resp:

- (a) $E[X(t)] = \frac{1}{2} \cos(\omega t)$
- (b) $R_X(t_1, t_2) = \frac{1}{3} \cos(\omega t_1) \cos(\omega t_2)$
- (c) $K_X(t_1, t_2) = \frac{1}{12} \cos(\omega t_1) \cos(\omega t_2)$
- (d) não

10. Em uma linha de produção de resistores de 1000Ω , a resistência real de cada resistor é uma variável aleatória R com distribuição uniforme entre 950Ω e 1050Ω . Assuma que os valores das resistências dos diferentes resistores são independentes. A companhia tem uma encomenda de resistores de 1% de tolerância (resistências entre 990Ω e 1010Ω). Um testador automático toma um resistor por segundo e mede sua resistência exata (este teste demora 1 segundo). O processo estocástico $N(t)$ denota o número de resistores com tolerância de 1% encontrados em t segundos. A variável aleatória T_r segundos é o tempo decorrido até encontrarmos r resistores com tolerância de 1%.

- (a) Calcule p , a probabilidade de um resistor ter tolerância de 1%.
- (b) Qual é a fmp de $N(t)$?
- (c) Calcule $E[T_1]$, o tempo esperado para encontrar o primeiro resistor com tolerância de 1%.
- (d) Qual é a probabilidade de o primeiro resistor com tolerância de 1% ser encontrado em exatamente 5 segundos?
- (e) $E[T_2|T_1 = 10]$, a esperança condicional do tempo necessário para encontrar o segundo resistor com tolerância de 1%, dado que o primeiro foi encontrado em 10 segundos.

Resp:

- (a) 0.2
- (b) $p_{N(t)}(n) = \begin{cases} \binom{t}{n} p^n (1-p)^{t-n}, & n = 0, 1, \dots, t \\ 0, & \text{caso contrário} \end{cases}$
- (c) 5
- (d) $(0,8)^4(0,2) \approx 0,08192$
- (e) 15

11. Para uma sequência de variáveis aleatórias Gaussianas iid X_n de média zero e variância unitária, encontre a fdp conjunta de X_1, \dots, X_m .

$$\text{Resp: } f_{X(1), \dots, X(m)}(x_1, \dots, x_m) = \frac{1}{(2\pi)^{m/2}} e^{-(x_1^2 + \dots + x_m^2)/2}$$

12. Pacotes de dados transmitidos por um modem sobre uma linha telefônica formam um processo de Poisson de taxa 10 pacotes/segundo. Usando M_k para denotar o número de pacotes transmitidos na k -ésima hora, encontre a fmp conjunta de M_1 e M_2 .

$$\text{Resp: } p_{M_1, M_2}(m_1, m_2) = \begin{cases} \frac{\alpha^{m_1+m_2} e^{-2\alpha}}{m_1! m_2!}, & m_1 = 0, 1, \dots; \quad m_2 = 0, 1, \dots \\ 0, & \text{caso contrário} \end{cases}$$

13. Seja $X(t)$ um processo movimento Browniano com variância $\text{Var}[X(t)] = \alpha t$. Mostre que $Y(t) = X(t)/\sqrt{\alpha}$ é um processo movimento Browniano com variância $\text{Var}[Y(t)] = t$.
14. Sejam dois processos estocásticos $X(t)$ e $Y(t)$ dados por:

$$X(t) = A \cos(\omega t + \Theta) \quad Y(t) = A \sin(\omega t + \Theta)$$

onde A e ω são constantes e Θ é uma v.a. com distribuição uniforme no intervalo $0, 2\pi$. Calcule $R_{XY}(\tau)$, $R_{YX}(\tau)$, e mostre que $R_{XY}(\tau) = R_{YX}(-\tau)$.

Resp:

$$R_{XY}(t_1, t_2) = \frac{A^2}{2} \sin(\omega(t_2 - t_1))$$

$$R_{YX}(t_1, t_2) = -\frac{A^2}{2} \sin(\omega(t_2 - t_1))$$

15. Seja $X(t)$ um processo estacionário no sentido estrito.

- (a) $Y(t) = X(t + a)$ é também um processo estocástico estacionário?
 (b) $Z(t) = X(at)$, $a \neq 0$, é também um processo estocástico estacionário?

Justifique suas respostas.

Resp: (a) sim (b) sim.

16. Considere um processo estocástico $X(t)$ definido por

$$X(t) = U \cos t + V \sin t, \quad -\infty < t < \infty$$

onde U e V são variáveis aleatórias independentes, e cada uma assume os valores -2 e 1 com probabilidades 1/3 e 2/3, respectivamente.

- (a) Calcule $E[X(t)]$.
 (b) Calcule $R_X(t_1, t_2)$.
 (c) Este processo é estacionário no sentido amplo?

Resp: (a) 0 (b) $2 \cos(t_2 - t_1)$ (c) sim.

17. Pacientes chegam a um consultório de acordo com um Processo de Poisson de taxa $\lambda = 1/10$ pacientes por minuto. O doutor não irá atender um paciente até que pelo menos três pacientes estejam na sala de espera.

- (a) Encontre o tempo médio de espera até que o primeiro paciente seja admitido pelo doutor.
 (b) Qual é a probabilidade de que ninguém seja atendido na primeira hora?

Resp: (a) 30 minutos (b) $25 \cdot e^{-6}$

18. Considere o processo estocástico $X(t) = Y \cos(\omega t)$, $t \geq 0$, onde ω é uma constante, e Y uma variável aleatória uniformemente distribuída no intervalo $(0, 1)$.

- (a) Calcule a média $E[X(t)]$.
 (b) Calcule a função de autocorrelação $R_X(t_1, t_2)$.
 (c) Calcule a função de autocovariância $K_X(t_1, t_2)$.
 (d) Este processo é estacionário no sentido amplo? Justifique sua resposta.

Resp: (a) $\frac{1}{2} \cos(\omega t)$ (b) $\frac{1}{3} \cos(\omega t_1) \cos(\omega t_2)$ (c) $\frac{1}{12} \cos(\omega t_1) \cos(\omega t_2)$
 (d) não

19. Um processo estocástico $v(t)$ é formado pela soma de um processo estocástico estacionário no sentido amplo $\xi(t)$ com um processo determinístico $s(t) = S_0 e^{\alpha t}$. $v(t)$ é estacionário no sentido amplo? Justifique sua resposta.

Resp: não.

20. Seja um processo estocástico $v(t) = \xi(t) + \eta$, onde $\xi(t)$ é um processo estocástico ergódico, e η é uma variável aleatória. Verifique se $v(t)$ é ou não estacionário no sentido amplo.

Resp: sim.

21. Suponha que uma secretária receba chamadas que chegam de acordo com um processo de Poisson a uma taxa de 10 chamadas por hora. Qual é a probabilidade de a secretária atender a todas as chamadas, dado que ela está fora de seu escritório nos 15 minutos iniciais e finais de cada hora?

Resp: e^{-5} .

22. Considere os seguintes processos autorregressivos:

$$W_n = 2W_{n-1} + X_n, W_0 = 0$$

$$Z_n = \frac{1}{2}Z_{n-1} + X_n, Z_0 = 0$$

Encontre W_n e Z_n em termos de X_n, X_{n-1}, \dots, X_1 , e então encontre $E[W_n]$ e $E[Z_n]$.

Resp:

$$W_n = \sum_{i=1}^n 2^{n-i} X_i \quad E[W_n] = (2^n - 1)E[X]$$

$$Z_n = \sum_{i=1}^n \frac{1}{2} 2^{n-i} X_i \quad E[Z_n] = 2(1 - (1/2)^{n-1})E[X]$$

23. Seja Z_1, Z_2, \dots, Z_n um conjunto de variáveis aleatórias iid, com $P[Z_n = 1] = p$ e $P[Z_n = -1] = q = 1 - p$ para todo n . Seja

$$X_n = \sum_{i=1}^n Z_i, n = 1, 2, \dots$$

e $X_0 = 0$. A coleção de variáveis aleatórias $\{X_n, n \geq 0\}$ é um processo aleatório, conhecido como caminhada aleatória simples em uma dimensão $X(n)$.

- (a) Construa uma sequência amostral típica de $X(n)$.
 (b) Sabendo que para este processo, a fdp de primeira ordem é dada por:

$$p_n(k) = \binom{n}{(n+k)/2} p^{(n+k)/2} q^{(n-k)/2}$$

calcule a probabilidade de $X(n) = -2$ depois de 4 passos.

- (c) Comprove o resultado do item b) enumerando todas as sequências possíveis que levam ao valor $X(n) = -2$ depois de 4 passos.

Resp: $P[X(4) = -2] = 4pq^3$

24. Seja $X_n, n \geq 0$ uma sequência de variáveis aleatórias iid com média 0 e variância 1. Mostre que $\{X_n, n \geq 0\}$ é um processo estacionário no sentido amplo.

Capítulo 9

Processamento de Sinais Aleatórios

Neste capítulo vamos utilizar os modelos do Capítulo 8 para representar sinais elétricos como funções amostra de processos estocásticos estacionários no sentido amplo. Usamos esta representação para descrever os efeitos de filtros lineares. Em particular vamos derivar a função de autocorrelação do processo estocástico na saída de um filtro em termos da função de autocorrelação do processo de entrada e da resposta a impulso do filtro. Vamos definir também a função espectro densidade de potência de um processo estocástico.

9.1 Sistemas lineares e invariantes no tempo

Antes de entrarmos no escopo da matéria, vamos definir alguns conceitos essenciais à compreensão do assunto:

Definição 9.1. Linearidade: Um sistema é linear se atende ao Teorema da Superposição, isto é

$$T[ax_1(t) + bx_2(t)] = aT[x_1(t)] + bT[x_2(t)] \quad (9.1)$$

onde $x_1(t)$ e $x_2(t)$ são sinais de entrada arbitrários, e a e b são constantes arbitrárias.

Definição 9.2. Invariância no Tempo: Se $y(t)$ é a resposta à entrada $x(t)$, então o sistema é dito invariante no tempo se para $x(t - \tau)$ temos $y(t - \tau)$.

Definição 9.3. Resposta Impulsiva: A resposta impulsiva $h(t)$ de um sistema linear e invariante no tempo é definida por

$$h(t) = T[\delta(t)] \quad (9.2)$$

onde $\delta(t)$ é uma função impulso unitário aplicada no instante $t = 0$.

A resposta do sistema para uma entrada arbitrária $x(t)$ é então a convolução de $x(t)$ com $h(t)$:

$$y(t) = h(t) \star x(t) = \int_{-\infty}^{+\infty} h(s)x(t-s) ds = \int_{-\infty}^{+\infty} h(t-s)x(s) ds \quad (9.3)$$

para sinais contínuos no tempo, e

$$y[n] = h[n] \star x[n] = \sum_{j=-\infty}^{\infty} h[j]x[n-j] = \sum_{j=-\infty}^{\infty} h[n-j]x[j] \quad (9.4)$$

para sinais discretos no tempo.

Definição 9.4. Causalidade: Um sistema é causal se a resposta no instante t depende apenas de valores de entrada passados, isto é, se

$$h(t) = 0, \quad \forall t < 0 \quad (9.5)$$

9.2 Filtragem linear de um processo estocástico

Em muitas aplicações de processamento de sinais é interessante representar os sinais como funções amostra de um processo estocástico. Nestas aplicações, é impossível saber antecipadamente qual sinal irá aparecer. Entretanto, podemos obter informações sobre os modelos probabilísticos destes sinais. Nas aplicações mais frequentes, os processos estocásticos são estacionários no sentido amplo, e desta forma as informações disponíveis consistem das estatísticas de primeira e segunda ordem, ou seja, a fdp ou fmp $f_X(x)$ e a função de autocorrelação $R_X(\tau)$.

Consideremos um filtro linear invariante no tempo com resposta a impulso $h(t)$. Se a entrada é um sinal determinístico $v(t)$, a saída $w(t)$ é dada pela integral de convolução

$$w(t) = \int_{-\infty}^{\infty} h(u)v(t-u) du \quad (9.6)$$

Esta relação pode também ser expressa no domínio da frequência em termos da transformada de Fourier.

Definição 9.5. Transformada de Fourier. As funções $g(t)$ e $G(f)$ são chamadas de um par de transformadas de Fourier se

$$G(f) = \int_{-\infty}^{\infty} g(t) e^{-j2\pi ft} dt \quad g(t) = \int_{-\infty}^{\infty} G(f) e^{j2\pi ft} df \quad (9.7)$$

Se um filtro linear tem resposta a impulso $h(t)$, a transformada de Fourier $H(f)$ é chamada de **resposta em frequência** do filtro. A convolução entre a entrada $v(t)$ do filtro e a sua resposta a impulso $h(t)$ no domínio do tempo torna-se uma multiplicação o domínio da frequência, isto é, se $v(t)$ é a entrada de um filtro linear invariante no

tempo com resposta a impulso $h(t)$, a transformada de Fourier da saída do filtro $W(f)$, está relacionada à transformada da entrada, $V(f)$ e à resposta em frequência do filtro $H(f)$ por

$$W(f) = H(f)V(f) \quad (9.8)$$

Se as possíveis entradas do filtro são funções amostras de um processo estocástico $X(t)$, então para uma entrada particular $x(t; s)$, a saída será dada pela convolução

$$y(t, s) = \int_{-\infty}^{\infty} h(u)x(t - u; s) du \quad (9.9)$$

Pelo fato de $y(t; s)$ estar associada a um resultado s de um experimento, $y(t; s)$ é uma função amostra de um processo estocástico $Y(t)$. Portanto, o modelo de filtragem linear completo consiste dos seguintes passos

- Realização do experimento e observação de um resultado s .
- Para o processo estocástico $X(t)$ estacionário no sentido amplo, usa-se a função amostra $x(t; s)$ como entrada para um filtro linear invariante no tempo com resposta a impulso $h(t)$.
- Observação da saída $y(t; s)$ do filtro.

Definição 9.6. *Processo de saída de um filtro linear invariante no tempo.* $X(t)$ é a **entrada** de um filtro linear invariante no tempo com resposta a impulso $h(t)$, e $Y(t)$ é a **saída** se todas as entradas do filtro são funções amostra de $X(t)$ e as saídas são funções amostra de $Y(t)$. $Y(t)$ está relacionado com $X(t)$ pela integral de convolução

$$Y(t) = \int_{-\infty}^{\infty} h(u)X(t - u) du = \int_{-\infty}^{\infty} h(t - u)X(u) du \quad (9.10)$$

A notação matemática da Definição 9.6 indica que a v.a. $Y(t_0) = \int_{-\infty}^{\infty} h(t_0 - u)X(u) du$ é uma função de todas as v.a.'s $X(u)$, para $-\infty < u < \infty$. Desde que $Y(t_0)$ é uma v.a., tem valor esperado

$$E[Y(t_0)] = E \left[\int_{-\infty}^{\infty} h(u)X(t_0 - u) du \right]$$

Para avaliar o valor esperado desta integral, lembremos que esta corresponde ao limite

$$Y(t_0) = \lim_{\Delta \rightarrow \infty} \sum_n h(n\Delta)X(t_0 - n\Delta)\Delta$$

Desde que a esperança da soma é igual à soma das esperanças, temos para valores pequenos de Δ ,

$$E[Y(t_0)] \approx E \left[\sum_n h(n\Delta) X(t_0 - n\Delta) \Delta \right] = \sum_n h(n\Delta) E[X(t_0 - n\Delta)] \Delta$$

Isto sugere que à medida que $\Delta \rightarrow 0$, temos

$$E[Y(t_0)] = E \left[\int_{-\infty}^{\infty} h(u) X(t_0 - u) du \right] = \int_{-\infty}^{\infty} h(u) E[X(t_0 - u)] du \quad (9.11)$$

Embora o argumento acima não seja uma prova, contém a idéia básica que uma integral é o limite de uma soma a qual podemos trocar de posição com a esperança. O seguinte Teorema usa a Equação (9.11) para relacionar o valor médio μ_Y e a função de autocorrelação $R_Y(\tau)$ com $h(t)$ e os parâmetros correspondentes de $X(t)$.

Teorema 9.1. *Se a entrada de um filtro linear invariante no tempo com resposta a impulso $h(t)$ é um processo estacionário no sentido amplo $X(t)$, a saída é um processo estacionário no sentido amplo $Y(t)$ com valor médio e função de autocorrelação dados por*

$$\mu_Y = \mu_X \int_{-\infty}^{\infty} h(t) dt = \mu_X H(0) \quad (9.12)$$

$$R_Y(\tau) = \int_{-\infty}^{\infty} h(u) \int_{-\infty}^{\infty} h(v) R_X(\tau + u - v) dv du \quad (9.13)$$

Demonstração. Primeiramente, observemos que a média de $Y(t)$ é

$$\mu_Y = E \left[\int_{-\infty}^{\infty} h(\tau) X(t - \tau) d\tau \right] = \int_{-\infty}^{\infty} h(u) E[X(t - u)] du$$

Desde que $E[X(t)] = \mu_X$ para todo t (pois $X(t)$ é estacionário no sentido amplo), $\mu_Y = \int_{-\infty}^{\infty} h(u) \mu_X du = \mu_X H(0)$. Para encontrar $R_Y(t, \tau) = E[Y(t)Y(t + \tau)]$, usamos a Definição 9.6 para escrever

$$\begin{aligned} R_Y(t, \tau) &= E \left[\int_{-\infty}^{\infty} h(u) X(t - u) du \int_{-\infty}^{\infty} h(v) X(t + \tau - v) dv \right] \\ &= \int_{-\infty}^{\infty} h(u) \int_{-\infty}^{\infty} h(v) E[X(t - u) X(t + \tau - v)] dv du \end{aligned}$$

Como $X(t)$ é estacionário no sentido amplo, $E[X(t - u) X(t + \tau - v)] = R_X(\tau - v + u)$ de modo que

$$R_Y(t, \tau) = R_Y(\tau) = \int_{-\infty}^{\infty} h(u) \int_{-\infty}^{\infty} h(v) R_X(\tau - v + u) dv du$$

□

Quando a entrada e a saída de um filtro são determinísticas, a relação no domínio da frequência $W(f) = H(f)V(f)$ avaliada em $f = 0$ leva a $W(0) = H(0)V(0)$. Para sinais determinísticos, $V(0)$ e $W(0)$ são conhecidas como as componentes DC (frequência zero) de $v(t)$ e $w(t)$.

Por analogia, podemos interpretar a Equação (9.12) no Teorema 9.1 chamando μ_X e μ_Y de componentes DC dos processos $X(t)$ e $Y(t)$.

A interpretação da segunda parte do Teorema 9.1 é menos direta. Além disso, usando o Teorema 9.1 para calcular $R_Y(\tau)$ a partir de $R_X(\tau)$ e $h(u)$ é extremamente difícil. Neste caso, é mais fácil trabalhar no domínio da frequência.

Exemplo 9.1. $X(t)$, um processo estocástico estacionário no sentido amplo com valor esperado $\mu_X = 10$ volts, é a entrada de um filtro linear invariante no tempo. A resposta a impulso do filtro é

$$h(t) = \begin{cases} e^{t/0,2} & 0 \leq t \leq 0,1 \\ 0 & \text{caso contrário} \end{cases}$$

Qual é o valor esperado do processo $Y(t)$ de saída do filtro?

Solução. Aplicando o Teorema 9.1 temos

$$\mu_Y = \mu_X \int_{-\infty}^{\infty} h(t) dt = 10 \int_0^{0,1} e^{t/0,2} dt = 2 \left. e^{t/0,2} \right|_0^{0,1} = 2(e^{0,5} - 1) = 1,30 \text{ volts}$$

9.3 Espectro densidade de potência

Assim como para sinais determinísticos, é instrutivo considerar a filtragem linear de processos estocásticos no domínio da frequência.

Definição 9.7. Espectro densidade de potência. Para um processo estocástico $X(t)$ estacionário no sentido amplo, a função de autocorrelação e o espectro densidade de potência $S_X(f)$ são o par de transformadas de Fourier

$$S_X(f) = \int_{-\infty}^{\infty} R_X(\tau) e^{-j2\pi f\tau} d\tau \quad R_X(\tau) = \int_{-\infty}^{\infty} S_X(f) e^{j2\pi f\tau} df$$

Pelo fato de $S_X(f)$ e $R_X(\tau)$ serem um par de transformadas de Fourier, se tivermos a expressão de uma, podemos sempre derivar a expressão da outra. O espectro densidade de potência tem algumas propriedades importantes.

Teorema 9.2. Para um processo estocástico $X(t)$ estacionário no sentido amplo, o espectro densidade de potência $S_X(f)$ tem as seguintes propriedades:

- a) $E[X^2(t)] = R_X(0) = \int_{-\infty}^{\infty} S_X(f) df$
 b) $S_X(-f) = S_X(f)$

Demonstração. A primeira propriedade é demonstrada considerando $\tau = 0$ para $R_X(\tau)$ na Definição 9.7.

Para provar a segunda propriedade, observemos que $R_X(\tau) = R_X(-\tau)$ implica

$$S_X(f) = \int_{-\infty}^{\infty} R_X(-\tau) e^{-j2\pi f\tau} d\tau$$

Fazendo $\tau' = -\tau$ temos

$$S_X(f) = \int_{\infty}^{-\infty} R_X(\tau') e^{-j2\pi f(-\tau')} (-d\tau') = \int_{-\infty}^{\infty} R_X(\tau') e^{-j2\pi(-f)\tau'} d\tau' = S_X(-f)$$

□

Quando interpretamos $E[X^2(t)]$ como a potência média de $X(t)$, a primeira parte do Teorema 9.2 sugere que $S_X(f)$ é uma medida da potência por unidade de frequência de $X(t)$. Quando passamos $X(t)$ através de um filtro linear $h(t)$, encontramos o espectro densidade de potência de $Y(t)$.

Teorema 9.3. Quando um processo $X(t)$ estacionário no sentido amplo é a entrada de um filtro linear invariante no tempo com resposta em frequência $H(f)$, a densidade espectral de potência da saída $Y(t)$ é

$$S_Y(f) = |H(f)|^2 S_X(f) \quad (9.14)$$

Demonstração. Do Teorema 9.1, podemos escrever

$$S_Y(f) = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(u)h(v)R_X(\tau + v - u) dudv \right) e^{-j2\pi f\tau} d\tau$$

Fazendo $\tau' = \tau + v - u$ temos

$$\begin{aligned} S_Y(f) &= \underbrace{\int_{-\infty}^{\infty} h(u) e^{-j2\pi fu} du}_{H(f)} \underbrace{\int_{-\infty}^{\infty} h(v) e^{j2\pi fv} dv}_{H^*(f)} \underbrace{\int_{-\infty}^{\infty} R_X(\tau') e^{-j2\pi f\tau'} d\tau'}_{S_X(f)} \\ &= |H(f)|^2 S_X(f) \end{aligned}$$

□

Estamos prontos agora para fazer novas interpretações sobre o espectro densidade de potência. Como mostrado na Figura 9.1, suponha que $H(f)$ é um filtro passa faixa ideal com largura de banda B centrada em f_0 , isto é

$$H(f) = \begin{cases} 1 & |f \pm f_0| \leq B/2 \\ 0 & \text{caso contrário} \end{cases}$$

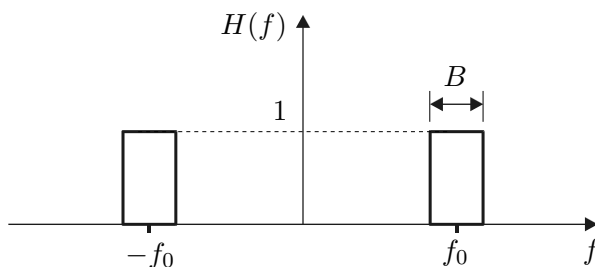


Figura 9.1: Filtro passa faixa ideal $H(f)$ com frequência central f_0 e largura de banda B Hz.

Neste caso, se passamos um processo estocástico $X(t)$ através do filtro $H(f)$ teremos na saída uma forma de onda $Y(t)$ que está na banda de passagem do filtro $H(f)$. Como mostrado acima, o espectro densidade de potência da saída do filtro é

$$S_Y(f) = |H(f)|^2 S_X(f)$$

Além disso, a potência média de $Y(t)$ satisfaz

$$E[Y^2(t)] = \int_{-\infty}^{\infty} S_Y(f) df = \int_{-f_0-B/2}^{-f_0+B/2} S_X(f) df + \int_{f_0-B/2}^{f_0+B/2} S_X(f) df$$

Desde que $S_X(f) = S_X(-f)$, quando B é pequeno, temos¹

$$E[Y^2(t)] \approx 2BS_X(f_0) \quad (9.15)$$

Podemos ver que a potência média da saída do filtro é aproximadamente o espectro densidade de potência da entrada na frequência central do filtro vezes a largura de faixa do filtro. Desta forma podemos concluir que $S_X(f_0)$ caracteriza a potência por unidade de frequência de $X(t)$ nas frequências próximas de f_0 .

Além disso, $E[Y^2(t)] \geq 0$ para qualquer frequência f_0 e largura de banda B não nula. No limite para B arbitrariamente pequeno, a aproximação da Equação (9.15) torna-se uma igualdade. Isto implica que $BS_X(f_0) \geq 0$ para todo B não nulo. Segue então que $S_X(f) \geq 0$ para todo f . Embora este argumento não seja uma prova, fornece uma intuição para o seguinte teorema:

Teorema 9.4. *Para um processo estocástico $X(t)$ estacionário no sentido amplo, o espectro densidade de potência $S_X(f) \geq 0$ para todo f .*

¹ $S_X(f)$ é aproximadamente constante quando B é pequeno.

Exemplo 9.2. Um processo estacionário $X(t)$ no sentido amplo com função de autocorrelação $R_X(\tau) = e^{-b|\tau|}$ é aplicado a um filtro RC com resposta a impulso

$$h(t) = \begin{cases} e^{-t/(RC)} & t \geq 0 \\ 0 & \text{caso contrário} \end{cases}$$

Assumindo que $b > 0$ e $b \neq 1/(RC)$, encontre $S_Y(f)$ e $R_Y(\tau)$ da saída $Y(t)$ do filtro. Qual é a potência média do processo estocástico na saída do filtro?

Solução. Por conveniência, façamos $a = 1/(RC)$. Desta forma, a função de transferência do filtro é

$$H(f) = \int_0^{\infty} e^{-at} e^{-j2\pi ft} dt = \frac{1}{a + j2\pi f}$$

Portanto

$$|H(f)|^2 = H(f)H^*(f) = \frac{1}{a + j2\pi f} \frac{1}{a - j2\pi f} = \frac{1}{a^2 + (2\pi f)^2}$$

O espectro densidade de potência do sinal de entrada é

$$\begin{aligned} S_X(f) &= \int_{-\infty}^{\infty} e^{-b|\tau|} e^{-j2\pi f\tau} d\tau \\ &= \int_{-\infty}^0 e^{b\tau} e^{-j2\pi f\tau} d\tau + \int_0^{\infty} e^{-b\tau} e^{-j2\pi f\tau} d\tau \\ &= \frac{1}{b - j2\pi f} + \frac{1}{b + j2\pi f} \\ &= \frac{2b}{(2\pi f)^2 + b^2} \end{aligned}$$

Usando o Teorema 9.3, escrevemos

$$S_Y(f) = \frac{2b}{[(2\pi f)^2 + a^2][(2\pi f)^2 + b^2]} = \frac{2b/(b^2 - a^2)}{(2\pi f)^2 + a^2} - \frac{2b/(b^2 - a^2)}{(2\pi f)^2 + b^2}$$

onde a última igualdade foi obtida através de frações parciais.

Reconhecendo que para qualquer constante $c > 0$, $e^{-c|\tau|}$ e $2c/((2\pi f)^2 + c^2)$ são pares de transformadas de Fourier, obtemos a expressão para a função de autocorrelação de $Y(t)$

$$R_Y(\tau) = \frac{b/a}{b^2 - a^2} e^{-a|\tau|} - \frac{1}{b^2 - a^2} e^{-b|\tau|}$$

A potência média é obtida pelo Teorema 9.2

$$E[Y^2(t)] = R_Y(0) = \frac{b/a - 1}{b^2 - a^2} = \frac{1}{a(b + a)}$$

9.4 Correlações cruzadas

Vimos que quando passamos um processo estocástico $X(t)$ através de um filtro linear $H(f)$, a saída $Y(t)$ é um novo processo estocástico. Para duas v.a.'s X e Y , a fdp ou fmp conjunta é um modelo de probabilidade completo. Para dois processos estocásticos $X(t)$ e $Y(t)$, um modelo de probabilidade completo consiste de uma fdp ou fmp conjunta das v.a.'s

$$X(t_1), X(t_2), \dots, X(t_n), Y(t'_1), Y(t'_2), \dots, Y(t'_k)$$

para todo $n, k, t_1, t_2, \dots, t_n$ e t'_1, t'_2, \dots, t'_k . Tal função de probabilidade conjunta contém informação suficiente para responder qualquer questão de engenharia sobre os processos estocásticos combinados $X(t)$ e $Y(t)$. Entretanto, encontrar e trabalhar com tal função é em geral extremamente custoso e difícil. A exceção principal é o caso de **processos independentes**.

Definição 9.8. Processos independentes. Os processos estocásticos $X(t)$ e $Y(t)$ são independentes se para qualquer coleção de amostras de tempo, t_1, t_2, \dots, t_n e t'_1, t'_2, \dots, t'_m

$$\begin{aligned} f_{X(t_1), \dots, X(t_n), Y(t'_1), \dots, Y(t'_m)}(x_1, \dots, x_n, y_1, \dots, y_m) \\ = f_{X(t_1), \dots, X(t_n)}(x_1, \dots, x_n) f_{Y(t'_1), \dots, Y(t'_m)}(y_1, \dots, y_m) \end{aligned}$$

9.4.1 Função de correlação cruzada

Para obter ferramentas úteis para analisar um par de processos dependentes, lembremos que a covariância e a correlação de um par de v.a.'s fornecem informações valiosas sobre a relação entre as v.a.'s. Portanto, para os processos $X(t)$ e $Y(t)$, trabalhamos com a correlação e a covariância das v.a.'s $X(t)$ e $Y(t + \tau)$. Desde que as v.a.'s dependem das suas variáveis temporais t e τ , a correlação das duas variáveis é uma função do tempo.

Definição 9.9. Função de correlação cruzada. A correlação cruzada dos processos $X(t)$ e $Y(t)$ é dada por

$$R_{XY}(t, \tau) = E[X(t)Y(t + \tau)]$$

Definida a correlação cruzada, vamos agora apresentar dois conceitos importantes no estudo dos processos estocásticos:

Definição 9.10. Processos descorrelacionados. Dois processos $X(t)$ e $Y(t)$, estacionários no sentido amplo, são ditos descorrelacionados se sua função de correlação cruzada é igual ao produto de suas médias, isto é

$$R_{XY}(\tau) = \overline{X(t)Y(t+\tau)} = \overline{X}\overline{Y}$$

Isto implica que as v.a.'s $x(t)$ e $y(t+\tau)$ são descorrelacionadas para todo t e τ .

Definição 9.11. Processos incoerentes ou ortogonais. Dois processos $X(t)$ e $Y(t)$, estacionários no sentido amplo, são ditos incoerentes ou ortogonais se

$$R_{XY}(\tau) = 0$$

Observe que os processo ortogonais são processos descorrelacionados com $\overline{X} = 0$ e/ou $\overline{Y} = 0$.

Assim como para a autocorrelação, existem muitas aplicações práticas nas quais a correlação cruzada depende somente da diferença entre dois instantes de tempo τ .

Definição 9.12. Processos conjuntamente estacionários no sentido amplo. Os processos estocásticos $X(t)$ e $Y(t)$ são conjuntamente estacionários no sentido amplo se cada um deles é estacionário no sentido amplo, e a correlação cruzada satisfaz

$$R_{XY}(t, \tau) = R_{XY}(\tau)$$

Propriedades da função de correlação cruzada

Vimos anteriormente que a função de autocorrelação é par, ou seja, $R_X(\tau) = R_X(-\tau)$. A correlação cruzada de processos estocásticos conjuntamente estacionários tem uma simetria ligeiramente diferente:

Teorema 9.5. Se $X(t)$ e $Y(t)$ são conjuntamente estacionários no sentido amplo então

$$R_{XY}(\tau) = R_{YX}(-\tau)$$

Demonstração. Da Definição 9.9, $R_{XY}(\tau) = E[X(t)Y(t+\tau)]$. Fazendo $u = t + \tau$, temos

$$R_{XY}(\tau) = E[X(u - \tau)Y(u)] = E[Y(u)X(u - \tau)] = R_{YX}(u, -\tau)$$

Desde que $X(t)$ e $Y(t)$ são conjuntamente estacionários no sentido amplo, podemos concluir que $R_{YX}(u, -\tau) = R_{YX}(-\tau)$ \square

Teorema 9.6. *Se $X(t)$ e $Y(t)$ são conjuntamente estacionários no sentido amplo então*

$$|R_{XY}(\tau)| \leq \{R_X(0)R_Y(0)\}^{1/2}$$

Demonstração. Usando a Desigualdade de Cauchy-Schwarz (Equação (3.34)), segue que

$$\{E[X(t)Y(t + \tau)]\}^2 \leq E[X^2(t)]E[Y^2(t + \tau)]$$

Reescrevendo esta equação em termos da função de autocorrelação, temos:

$$[R_{XY}(\tau)]^2 \leq R_X(0)R_Y(0) \Rightarrow |R_{XY}(\tau)| \leq \sqrt{R_X(0)R_Y(0)}$$

\square

Teorema 9.7. *Se $X(t)$ e $Y(t)$ são conjuntamente estacionários no sentido amplo então*

$$|R_{XY}(\tau)| \leq \frac{1}{2} [R_X(0) + R_Y(0)]$$

Demonstração.

$$E \{ [X(t) - Y(t + \tau)]^2 \} \geq 0$$

Expandindo o quadrado, temos

$$E [X^2(t) - 2X(t)Y(t + \tau) + Y^2(t + \tau)] \geq 0$$

$$E [X^2(t)] - 2E [X(t)Y(t + \tau)] + E [Y^2(t + \tau)] \geq 0$$

Reescrevendo esta equação em termos das funções de autocorrelação e correlação cruzada, temos

$$R_X(0) - 2R_{XY}(\tau) + R_Y(0) \geq 0 \Rightarrow R_{XY}(\tau) \leq \frac{1}{2} [R_X(0) + R_Y(0)]$$

\square

Teorema 9.8. *Se X e Y são v.a.'s independentes, então*

$$R_{XY}(\tau) = R_{YX}(\tau) = \overline{X} \overline{Y}$$

Demonstração.

$$R_{XY}(\tau) = E[X(t)Y(t + \tau)]$$

Como X e Y são independentes, podemos escrever

$$E[X(t)Y(t + \tau)] = E[X(t)]E[Y(t + \tau)] = \overline{X} \overline{Y}$$

□

9.4.2 Densidade espectral cruzada

Quando $X(t)$ e $Y(t)$ são conjuntamente estacionários no sentido amplo, podemos estudar a correlação cruzada no domínio da frequência.

Definição 9.13. Densidade espectral cruzada. *Para processos $X(t)$ e $Y(t)$ conjuntamente estacionários no sentido amplo, a transformada de Fourier da correlação cruzada leva à densidade espectral cruzada*

$$S_{XY}(f) = \int_{-\infty}^{\infty} R_{XY}(\tau) e^{-j2\pi f\tau} d\tau$$

Como a densidade espectral cruzada é a transformada de Fourier da função de correlação cruzada, podemos mostrar o seguinte teorema:

Teorema 9.9. *Para os processos $X(t)$ e $Y(t)$ conjuntamente estacionários no sentido amplo, a densidade espectral cruzada apresenta a seguinte simetria*

$$S_{XY}(f) = S_{YX}(-f)$$

Encontramos correlações cruzadas em experimentos que envolvem observações ruidosas de um processo estocástico $X(t)$ estacionário no sentido amplo.

Exemplo 9.3. Suponha que estejamos interessados em $X(t)$ mas só podemos observar

$$Y(t) = X(t) + N(t)$$

onde $N(t)$ é um processo estacionário no sentido amplo com média zero, que interfere com nossa observação de $X(t)$. Assumimos que $X(t)$ e $N(t)$ são conjuntamente estacionários no sentido amplo. Para caracterizar $Y(t)$, encontre a média $E[Y(t)]$, a função de autocorrelação $R_Y(\tau)$, e o espectro densidade de potência $S_Y(f)$.

Solução. Desde que o valor esperado de uma soma é igual à soma dos valores esperados,

$$E[Y(t)] = E[X(t)] + E[N(t)] = E[X(t)]$$

desde que $E[N(t)] = 0$ (dado do problema).

Para a função de autocorrelação, temos

$$\begin{aligned} R_Y(t, \tau) &= E[Y(t)Y(t + \tau)] \\ &= E[(X(t) + N(t))(X(t + \tau) + N(t + \tau))] \\ &= R_X(\tau) + R_{XN}(t, \tau) + R_{NX}(t, \tau) + R_N(\tau) \end{aligned}$$

Quando $X(t)$ e $N(t)$ são conjuntamente estacionários no sentido amplo $R_{XN}(t, \tau) = R_{XN}(\tau)$ e $R_{NX}(t, \tau) = R_{NX}(\tau)$. Então podemos reescrever a equação acima como

$$R_Y(t, \tau) = R_X(\tau) + R_{XN}(\tau) + R_{NX}(\tau) + R_N(\tau)$$

O lado direito desta equação indica que $R_Y(t, \tau)$ depende somente de τ . Isto implica que $Y(t)$ é estacionário no sentido amplo com função de autocorrelação $R_Y(t, \tau) = R_Y(\tau)$. Tomando a transformada de Fourier de ambos os lados, obtemos a densidade espectral de potência de $Y(t)$

$$S_Y(f) = S_X(f) + S_{XN}(f) + S_{NX}(f) + S_N(f)$$

Exemplo 9.4. Continuando o Exemplo 9.3, suponha que $N(t)$ seja um processo de média zero, independente de $X(t)$. Encontre a função de autocorrelação e a densidade espectral de potência da observação $Y(t)$.

Solução. Neste caso,

$$R_{XN}(t, \tau) = E[X(t)N(t + \tau)] = E[X(t)]E[N(t + \tau)] = 0$$

Similarmente, $R_{NX}(t, \tau) = 0$. Isto implica

$$R_Y(\tau) = R_X(\tau) + R_N(\tau)$$

$$S_Y(f) = S_X(f) + S_N(f)$$

9.4.3 Filtragem de processos estocásticos

A função de autocorrelação e a densidade espectral de potência são particularmente úteis na caracterização da entrada e saída de um filtro linear invariante no tempo. Quando $X(t)$ e $Y(t)$ são os processos de entrada e saída de um filtro linear invariante no tempo $h(t)$, podemos usar a Definição 9.6 para calcular a correlação cruzada $R_{XY}(t, \tau)$.

Teorema 9.10. *Quando um processo $X(t)$ estacionário no sentido amplo é a entrada de um filtro linear invariante no tempo $h(t)$, a correlação cruzada entre entrada e saída é dada por*

$$R_{XY}(t, \tau) = R_{XY}(\tau) = \int_{-\infty}^{\infty} h(u) R_X(\tau - u) du$$

Demonstração. Da Definição 9.6, $Y(t + \tau) = \int_{-\infty}^{\infty} h(u) X(t + \tau - u) du$. Isto implica que a correlação cruzada entre a entrada e a saída do filtro é

$$\begin{aligned} R_{XY}(t, \tau) &= E \left[X(t) \int_{-\infty}^{\infty} h(u) X(t + \tau - u) du \right] \\ &= \int_{-\infty}^{\infty} h(u) E[X(t) X(t + \tau - u)] du \\ &= \int_{-\infty}^{\infty} h(u) R_X(\tau - u) du \end{aligned}$$

□

Quando a entrada $X(t)$ de um filtro linear invariante no tempo é um processo estacionário no sentido amplo, o Teorema 9.1 diz que a saída $Y(t)$ é também um processo estacionário no sentido amplo, e o Teorema 9.10 diz que a correlação cruzada $R_{XY}(t, \tau)$ depende somente de τ . Estes dois resultados implicam no seguinte teorema.

Teorema 9.11. *Quando um processo $X(t)$ estacionário no sentido amplo é a entrada de um filtro linear invariante no tempo, a entrada $X(t)$ e a saída $Y(t)$ são conjuntamente estacionárias no sentido amplo.*

No Teorema 9.10 vimos que a correlação cruzada entre a entrada e a saída é dada pela convolução entre a função de autocorrelação $R_X(\tau)$ da entrada e a resposta a impulso $h(t)$ do filtro. Então podemos pensar em $R_{XY}(\tau)$ como a saída do filtro $h(t)$ quando $R_X(\tau)$ é a entrada. No exemplo a seguir veremos que calcular a correlação cruzada através de convoluções tende a ser um processo complicado.

Exemplo 9.5. *Um processo $X(t)$ estacionário no sentido amplo com função de autocorrelação $R_X(\tau) = e^{-b|\tau|}$ é a entrada de um filtro RC com resposta impulsiva*

$$h(t) = \begin{cases} e^{-t/(RC)} & t \geq 0 \\ 0 & \text{caso contrário} \end{cases}$$

Assumindo que $b > 0$, encontre a correlação cruzada $R_{XY}(\tau)$ entre a entrada e a saída.

Solução. Seja $a = 1/(RC)$. Do Teorema 9.10, a correlação cruzada é

$$R_{XY}(\tau) = \int_{-\infty}^{\infty} h(u) R_X(\tau - u) du = \int_0^{\infty} e^{-au} e^{-b|\tau-u|} du$$

Para $\tau \geq 0$, esta integral pode ser escrita como

$$R_{XY}(\tau) = \int_0^{\tau} e^{-(a-b)u-b\tau} du + \int_{\tau}^{\infty} e^{-(a+b)u+b\tau} du = \frac{e^{-b\tau}}{a-b} - \frac{2be^{-a\tau}}{a^2-b^2}$$

Quando $\tau < 0$ e $u \geq 0$, então $|\tau - u| = u - \tau$ e

$$R_{XY}(\tau) = \int_0^{\infty} e^{-au} e^{-b(u-\tau)} du = \frac{e^{b\tau}}{a+b}$$

Uma expressão completa para a correlação cruzada entre a entrada e a saída é

$$R_{XY}(\tau) = \begin{cases} \frac{e^{b\tau}}{a+b} & \tau < 0 \\ \frac{e^{-b\tau}}{a-b} - \frac{2be^{-a\tau}}{a^2-b^2} & \tau \geq 0 \end{cases}$$

O Teorema 9.10 nos encoraja a reexaminar o Teorema 9.1 desde que a integral dupla para $R_Y(\tau)$ pode ser expressa em termos da correlação cruzada $R_{XY}(\tau)$

Teorema 9.12. Quando um processo $X(t)$ estacionário no sentido amplo é a entrada de um filtro linear $h(t)$ invariante no tempo, a função de autocorrelação da saída $Y(t)$ é dada por

$$R_Y(\tau) = \int_{-\infty}^{\infty} h(-w) R_{XY}(\tau - w) dw$$

Demonstração.

$$R_Y(\tau) = \int_{-\infty}^{\infty} h(u) \underbrace{\int_{-\infty}^{\infty} h(v) R_X(\tau + u - v) dv}_{R_{XY}(\tau+u)} du = \int_{-\infty}^{\infty} h(u) R_{XY}(\tau + u) du$$

A substituição $w = -u$ na integral acima completa a prova. □

O Teorema 9.12 diz que ao passarmos o sinal determinístico $R_{XY}(\tau)$ através de um filtro linear invariante no tempo $h(-t)$ obtemos a função de autocorrelação $R_Y(\tau)$.

Observemos que um filtro com resposta a impulso $h(-t)$ pode também ser representado como um filtro de resposta em frequência $H^*(f)$. No domínio da frequência, os Teoremas 9.10 e 9.12 têm as seguintes consequências

Teorema 9.13. *Seja $X(t)$ uma entrada estacionária no sentido amplo para um filtro linear invariante no tempo $H(f)$. A entrada $X(t)$ e a saída $Y(t)$ satisfazem*

$$S_{XY}(f) = H(f)S_X(f) \quad S_Y(f) = H^*(f)S_{XY}(f)$$

As relações entre $R_X(\tau)$, $R_{XY}(\tau)$ e $R_Y(\tau)$, bem como entre $S_X(f)$, $S_{XY}(f)$ e $S_Y(f)$ são mostradas na Figura 9.2.

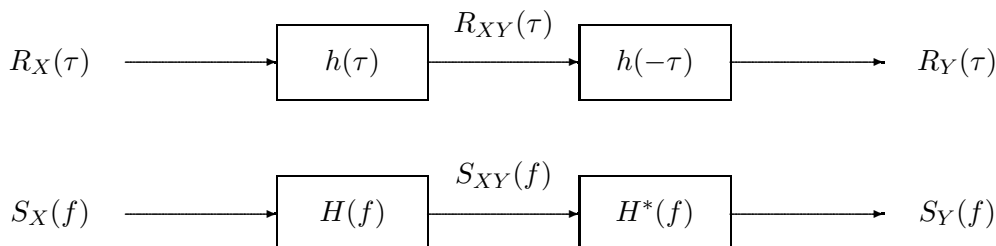


Figura 9.2: A correlação cruzada entre a entrada e a saída de um filtro linear invariante no tempo é a convolução da resposta a impulso do filtro com a função de autocorrelação da entrada. A densidade espectral cruzada entre a entrada e a saída é o produto do espectro de densidade de potência da entrada com a função de transferência do filtro. A densidade espectral de potência da saída é o produto da densidade espectral cruzada da entrada e da saída e o complexo conjugado da função de transferência do filtro.

9.5 Processos gaussianos

Um processo gaussiano tem a propriedade de que toda coleção de valores de amostras é descrita pela fdp Gaussiana multidimensional. Isto é, uma coleção de amostras $X(t_1), X(t_2), \dots, X(t_k)$, tem uma fdp conjunta descrita por um vetor $\mu_X = [\mu_X(t_1), \mu_X(t_2), \dots, \mu_X(t_k)]^t$ e uma matriz \mathbf{C} cujo i, j -ésimo elemento

$$C_{i,j} = C_X(t_i, t_j - t_i) = R_X(t_i, t_j - t_i) - \mu_X(t_i)\mu_X(t_j)$$

é a covariância entre $X(t_i)$ e $X(t_j)$. Usando o vetor $\mathbf{x} = [x_1, \dots, x_k]^t$, o vetor de valores médios μ_X , a matriz de covariância \mathbf{C} e seu determinante $|\mathbf{C}|$, podemos definir a fdp Gaussiana multidimensional.

Definição 9.14. Processo Gaussiano: $X(t)$ é um processo estocástico Gaussiano se a fdp conjunta de $X(t_1), \dots, X(t_k)$ tem densidade Gaussiana multidimensional

$$f_{X(t_1) \dots X(t_k)}(x_1, \dots, x_k) = \frac{1}{(2\pi)^{k/2} |\mathbf{C}|^{1/2}} e^{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_X)^t \mathbf{C}^{-1} (\mathbf{x} - \boldsymbol{\mu}_X)}$$

Embora esta expressão possa parecer bastante complicada, pode ser reduzida para expressões familiares em vários casos. Por exemplo, quando $k = 1$, a matriz \mathbf{C} é simplesmente o escalar $C_X(t_1, 0) = \text{Var}(X(t_1)) = \sigma_1^2$, o vetor $\boldsymbol{\mu}_X$ é o escalar $E[X(t_1)] = \mu_1$ e a fdp conjunta pode ser simplificada para a densidade Gaussiana ordinária

$$f_{X(t_1)}(x_1) = \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{(x_1 - \mu_1)^2}{2\sigma_1^2}}$$

Similarmente, para $k = 2$, $X(t_1)$ e $X(t_2)$ apresentam distribuição Gaussiana bidimensional

$$f_{X(t_1)X(t_2)}(x_1, x_2) = \frac{\exp \left[-\frac{\left(\frac{x_1 - \mu_1}{\sigma_1}\right)^2 - \frac{2\rho(x_1 - \mu_1)(x_2 - \mu_2)}{\sigma_1\sigma_2} + \left(\frac{x_2 - \mu_2}{\sigma_2}\right)^2}{2(1 - \rho^2)} \right]}{2\pi\sigma_1\sigma_2\sqrt{1 - \rho^2}}$$

onde $X(t_1)$ e $X(t_2)$ têm coeficiente de correlação $\rho = C_X(t_1, t_2 - t_1)/(\sigma_1\sigma_2)$ e

$$E[X(t_1)] = \mu_1 \quad E[X(t_2)] = \mu_2 \quad \text{Var}[X(t_1)] = \sigma_1^2 \quad \text{Var}[X(t_2)] = \sigma_2^2$$

Um último caso importante para a fdp Gaussiana conjunta ocorre quando $X(t_1), \dots, X(t_k)$ são mutuamente independentes. Neste caso, o elemento (i, j) da matriz de covariância \mathbf{C} é dado por

$$C_{ij} = C_X(t_i, t_j - t_i) = \begin{cases} \text{Var}[X(t_i)] & i = j \\ 0 & \text{caso contrário} \end{cases}$$

Isto é, a matriz \mathbf{C} é uma matriz diagonal. Neste caso, \mathbf{C}^{-1} é também diagonal, com o i -ésimo elemento da diagonal dado por $C_{ii}^{-1} = 1/\text{Var}[X(t_i)]$. Usando μ_i e σ_i^2 para denotar a média e a variância de $X(t_i)$, observamos que o vetor de valores médios é $\boldsymbol{\mu}_X = [\mu_1, \dots, \mu_k]^t$ e que o expoente da distribuição Gaussiana conjunta é

$$-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_X)^t \mathbf{C}^{-1} (\mathbf{x} - \boldsymbol{\mu}_X) = -\frac{1}{2} \left(\frac{(x_1 - \mu_1)^2}{\sigma_1^2} + \dots + \frac{(x_k - \mu_k)^2}{\sigma_k^2} \right)$$

Neste caso, a fdp conjunta torna-se

$$\begin{aligned} f_{X(t_1), \dots, X(t_k)}(x_1, \dots, x_k) &= \frac{e^{-(x_1 - \mu_1)^2/(2\sigma_1^2)}}{\sqrt{2\pi\sigma_1^2}} \dots \frac{e^{-(x_k - \mu_k)^2/(2\sigma_k^2)}}{\sqrt{2\pi\sigma_k^2}} \\ &= f_{X(t_1)}(x_1) \dots f_{X(t_k)}(x_k) \end{aligned}$$

Um fato importante a ser observado da distribuição Gaussiana multidimensional geral é que a fdp é completamente especificada pelas médias $\mu_{X(t_1)}, \dots, \mu_{X(t_k)}$ e as covariâncias $C_X(t_i, t_j - t_i)$. Ou seja, um processo estocástico Gaussiano é completamente especificado pelas estatísticas de primeira e segunda ordens ($\mu_{X(t)}$ e $C_X(t, \tau)$).

Nosso interesse principal está nos processos Gaussianos estacionários no sentido amplo. Neste caso, $E[X(t_i)] = \mu_X$ para cada t_i e $C_X(t_i, t_j - t_i) = R_X(t_j - t_i) - \mu_X^2$. Isto é, quando o processo Gaussiano é estacionário no sentido amplo, sua distribuição é completamente especificada pela média μ_X e a função de autocorrelação $R_X(\tau)$.

Teorema 9.14. *Se $X(t)$ é um processo Gaussiano estacionário no sentido amplo, então $X(t)$ é um processo Gaussiano estacionário no sentido estrito.*

Demonstração. Sejam $\bar{\mu}$ e \mathbf{C} o vetor média e a matriz de covariância do vetor aleatório $[X(t_1), \dots, X(t_k)]^t$. Sejam $\bar{\mu}'$ e \mathbf{C}' as mesmas quantidades para o vetor aleatório deslocado no tempo $[X(t_1 + T), \dots, X(t_k + T)]^t$. Desde que $X(t)$ é estacionário no sentido amplo,

$$E[X(t_i)] = E[X(t_i + T)] = \mu_X$$

O elemento (i, j) de \mathbf{C} é

$$C_{ij} = C_X(t_i, t_j) = C_X(t_j - t_i) = C_X(t_j + T - (t_i + T)) = C_X(t_i + T, t_j + T) = C'_{ij}$$

Então, $\bar{\mu} = \bar{\mu}'$ e $\mathbf{C} = \mathbf{C}'$, o que implica em

$$f_{X(t_1), \dots, X(t_k)}(x_1, \dots, x_k) = f_{X(t_1+T), \dots, X(t_k+T)}(x_1, \dots, x_k)$$

Portanto $X(t)$ é um processo estacionário no sentido estrito. \square

A Definição 9.14 é bastante difícil de usar na prática. Uma definição equivalente de um processo Gaussiano refere-se a uma v.a. que é um **funcional linear** de um processo estocástico $X(t)$. Especificamente, se integramos $X(t)$ ponderada por uma função $g(t)$ sobre um intervalo $(0, T)$, obtemos a v.a.

$$Y = \int_0^T g(t)X(t) dt$$

Teorema 9.15. *$X(t)$ é um processo estocástico Gaussiano se $Y = \int_0^T g(t)X(t) dt$ é uma v.a. Gaussiana para todo $g(t)$ tal que $E[Y^2] < \infty$.*

Este teorema nos permite mostrar facilmente que a filtragem linear de um processo Gaussiano gera um outro processo Gaussiano.

Teorema 9.16. *Passando um processo $X(t)$ estacionário Gaussiano através de um filtro linear $h(t)$, gera-se na saída um processo estocástico Gaussiano $Y(t)$ com média e função de autocorrelação dados pelo Teorema 9.1.*

Demonstração. A saída $Y(t)$ é dada pela integral de convolução

$$Y(t) = \int_{-\infty}^{\infty} h(t - \tau) X(\tau) d\tau$$

Para mostrar que $Y(t)$ é um processo Gaussiano, mostramos que um funcional linear de $Y(t)$ é sempre Gaussiano pois é um funcional linear de $X(t)$, isto é,

$$\int_0^T Y(t)g(t) dt = \int_0^T \int_{-\infty}^{\infty} h(t - \tau) X(\tau) d\tau g(t) dt = \int_{-\infty}^{\infty} X(\tau) \left[\int_0^T h(t - \tau)g(t) dt \right] d\tau$$

No lado direito temos um funcional linear de $X(t)$ o qual é uma v.a. Gaussiana. Desta forma mostramos que um funcional linear de $Y(t)$ é uma v.a. Gaussiana, o que implica que $Y(t)$ é um processo estocástico Gaussiano. \square

9.6 Processo ruído branco gaussiano

Em engenharia elétrica é comum o estudo de ruído: ruído térmico em resistores, ruído em sistemas de comunicações, etc. O ruído é uma forma de onda imprevisível que é normalmente modelado por um processo estocástico Gaussiano estacionário $W(t)$. O ruído não tem componente DC, de modo que

$$E[W(t_1)] = \mu_W = 0$$

Além disso, para enfatizar a natureza imprevisível do processo de ruído, assumimos que para qualquer coleção de instantes de tempo distintos t_1, \dots, t_k , $W(t_1), \dots, W(t_k)$ é um conjunto de v.a.'s independentes. Neste caso, o valor do ruído no instante t_1 não diz nada sobre o valor do mesmo no instante $t_j, j \neq i$. Uma consequência desta independência é que para $\tau \neq 0$,

$$R_W(\tau) = E[W(t)W(t + \tau)] = E[W(t)]E[W(t + \tau)] = 0$$

Para completar nosso modelo de $W(t)$, temos que encontrar $R_W(0)$. Para isto, vamos considerar a função densidade espectral de potência $S_W(f)$ da Definição 9.7

$$S_W(f) = \int_{-\infty}^{\infty} R_W(\tau) e^{-j2\pi f\tau} d\tau$$

Com $R_W(\tau) = 0$ para $\tau \neq 0$, $S_W(f)$ é uma constante para todo f . Ainda, a constante é igual a zero a menos que $R_W(\tau) = \delta(\tau)$. Portanto, N_0 é a potência por unidade de largura de banda do processo estocástico Gaussiano branco. Embora o processo ruído branco Gaussiano seja um modelo matemático bastante útil, ele não se conforma com nenhum sinal real. Note que a potência média do ruído é

$$E[W^2(t)] = R_W(0) = \int_{-\infty}^{\infty} S_W(f) df = \int_{-\infty}^{\infty} \frac{N_0}{2} df = \infty$$

Isto é, o ruído branco tem potência infinita, o que é fisicamente impossível. O modelo é útil quando se imagina que é um modelo de ruído na entrada de um sistema físico. Todo sinal de ruído Gaussiano observado na prática pode ser visto como um sinal de ruído branco Gaussiano filtrado. Passando um processo ruído branco através de um filtro $h(t)$ geramos um processo de ruído

$$Y(t) = \int_0^t h(t - \tau)W(\tau) d\tau$$

Ao contrário do processo branco $W(t)$, o processo de ruído $Y(t)$ tem potência média finita.

Exemplo 9.6. Um processo Gaussiano branco com $N_0 = 10^{-15}$ W/Hz é inserido em um filtro linear invariante no tempo com resposta a impulso

$$h(t) = \begin{cases} 2\pi 10^6 e^{-2\pi 10^6 t} & t \geq 0 \\ 0 & \text{caso contrário} \end{cases}$$

Encontre as seguintes propriedades do processo de saída $Y(t)$.

- (a) A função densidade espectral de potência $S_Y(f)$.
- (b) A função de autocorrelação $R_Y(\tau)$.
- (c) A potência média $E[Y^2(t)]$.

Solução. Resolvemos este problema usando o Teorema 9.3. A função densidade espectral de potência da entrada é $S_X(f) = 10^{-15}/2$ W/Hz para todo f .

A magnitude ao quadrado da resposta em frequência do filtro é dada por

$$|H(f)|^2 = \frac{(2\pi 10^6)^2}{(2\pi f)^2 + (2\pi 10^6)^2}$$

Portanto, a função densidade espectral de potência da saída é dada por

$$S_Y(f) = |H(f)|^2 S_X(f) = \frac{10^{-15}}{2} \frac{(2\pi 10^6)^2}{(2\pi f)^2 + (2\pi 10^6)^2} = \frac{\pi 10^{-9}}{2} \frac{2(2\pi 10^6)}{(2\pi f)^2 + (2\pi 10^6)^2}$$

A transformada inversa de Fourier de $\frac{2(2\pi 10^6)}{(2\pi f)^2 + (2\pi 10^6)^2}$ é dada por $e^{-2\pi 10^6 |\tau|}$. Isto implica que

$$R_Y(\tau) = \frac{\pi 10^{-9}}{2} e^{-2\pi 10^6 |\tau|}$$

A potência média no processo de saída é, portanto, $R_Y(0) = \pi/2 \times 10^{-9}$ W.

9.7 Exercícios

1. Mostre que se o espectro densidade de potência de um processo estocástico é limitado em banda a B Hz, e se as amostras do sinal são descorrelacionadas em $\tau = n/(2B)$, para todos os valores integrais de n , então o processo precisa ter um espectro densidade de potência com distribuição uniforme sobre a banda $(0, B)$. Em outras palavras, o processo precisa ser um ruído branco limitado em banda.
2. Suponha que em um sistema de comunicação existem dois sinais sendo transmitidos: $x(t)$ e $y(t)$. Na transmissão, devido ao ruído de canal, $n(t)$, chegam ao receptor os sinais $x(t) + n(t)$ e $y(t) + n(t)$. Explique como podemos decidir qual sinal foi recebido, se o receptor conhece as formas de onda de $x(t)$ e $y(t)$.
3. Um processo estocástico $Y(t)$ é relacionado ao processo estocástico $X(t)$ por

$$Y(t) = X(t) \cos(\omega_0 t + \theta)$$

onde θ é uma variável aleatória independente uniformemente distribuída sobre o intervalo $(0, 2\pi)$. Mostre que

$$R_Y(\tau) = \frac{1}{2} R_X(\tau) \cos(\omega_0 \tau)$$

$$S_Y(\omega) = \frac{1}{4} [S_X(\omega + \omega_c) + S_X(\omega - \omega_c)]$$

Esta é a extensão do teorema da modulação para processos estocásticos.

Dica: se dois processos estocásticos $x(t)$ e $y(t)$ são independentes, então

$$\overline{x(t)g(t)x(t+\tau)g(t+\tau)} = \overline{x(t)x(t+\tau)} \overline{g(t)g(t+\tau)} = R_X(\tau)R_g(\tau)$$

4. Dois processos estocásticos são dados por

$$x(t) = A \cos(\omega_1 t + \phi) \quad \text{e} \quad y(t) = B \cos(\omega_2 t + \theta)$$

onde A , B , ω_1 e ω_2 são constantes. As fases iniciais ϕ e θ estão relacionadas pela equação $\theta = 2\phi$ e a variável aleatória ϕ é uniformemente distribuída sobre o intervalo $(0, 2\pi)$. Mostre que a função de correlação cruzada e o espectro densidade de potência cruzada dos dois processos são zero.

5. Sejam os processos estocásticos

$$x(t) = A \cos(\omega_0 t + \varphi) \quad \text{e} \quad y(t) = B \cos(n\omega_0 t + n\varphi)$$

onde A , B e ω_0 são constantes e φ é uma variável aleatória uniformemente distribuída no intervalo $(0, 2\pi)$. Mostre que os dois processos são incoerentes.

6. Seja $h(t)$ um filtro passa baixas com resposta a impulso

$$h(t) = \begin{cases} e^{-t} & t \geq 0 \\ 0 & \text{caso contrário} \end{cases}$$

A entrada do filtro é $X(t)$, um processo estacionário no sentido amplo com valor esperado $\mu_X = 2$ e função de autocorrelação $R_X(\tau) = \delta(\tau)$. Calcule a média e a função de autocorrelação do processo $Y(t)$ na saída deste filtro.

Resp: $E[Y(t)] = 2 \quad R_Y(\tau) = \frac{1}{2} e^{-|\tau|}$

7. Seja um processo $X(t)$ estacionário no sentido amplo e de média zero com função de autocorrelação dada por $R_X(\tau) = \delta(\tau)$. Se passarmos este sinal por um filtro linear invariante no tempo com resposta a impulso

$$h(t) = \begin{cases} e^{-2t} & t \geq 0 \\ 0 & \text{caso contrário} \end{cases}$$

qual será a densidade espectral de potência da saída $Y(t)$?

Resp: $S_Y(f) = \frac{1}{4 + 4\pi^2 f^2}$

8. O processo $X(t)$ estacionário no sentido amplo é a entrada de um filtro “tapped delay line”

$$H(f) = a_1 e^{-j2\pi f t_1} + a_2 e^{-j2\pi f t_2}$$

Encontre a densidade espectral cruzada $S_{XY}(f)$ e a correlação cruzada $R_{XY}(\tau)$.

Resp:

$$S_{XY}(f) = a_1 e^{-j2\pi f t_1} S_X(f) + a_2 e^{-j2\pi f t_2} S_X(f)$$

$$R_{XY}(\tau) = a_1 R_X(\tau - t_1) + a_2 R_X(\tau - t_2)$$

9. $X(t)$ é um processo estocástico Gaussiano de média zero com função de autocorrelação $R_X(\tau) = 2^{-|\tau|}$. Qual é a fdp conjunta de $X(t)$ e $X(t+1)$?

Resp: $f_{X(t), X(t+1)}(x_0, x_1) = \frac{1}{\sqrt{3\pi^2}} e^{-\frac{2}{3}(x_0^2 - x_0 x_1 + x_1^2)}$

10. Um processo ruído branco Gaussiano $N(t)$ com densidade espectral de potência de α W/Hz é passado através de um integrador gerando a saída $Y(t) = \int_0^t N(u) du$. Calcule a função de autocorrelação $R_Y(t, \tau)$.

Resp: $R_Y(t, \tau) = \alpha \min\{t, t + \tau\}$

11. Verifique quais das funções abaixo podem ser consideradas espectro densidade de potência de um processo estocástico real. Em caso positivo, calcule a potência do processo.

$$\begin{array}{lll}
 \text{(a)} \frac{1}{\omega^2 + 16} & \text{(b)} j[\delta(\omega - \omega_0) + \delta(\omega + \omega_0)] & \text{(c)} \frac{1}{\omega^4 + 9\omega^2 + 18} \\
 \text{(d)} \frac{\omega}{\omega^2 + 16} & \text{(e)} \frac{j\omega^2}{\omega^2 + 16} & \text{(f)} \frac{\omega^3}{\omega^4 + 9\omega^2 + 18}
 \end{array}$$

Resp:

- (a) Sim. $P = 1/8$.
- (b) Não.
- (c) Sim. $P = \frac{\sqrt{6} - \sqrt{3}}{18\sqrt{2}} \approx 0,0282$.
- (d) Não.
- (e) Não.
- (f) Não.

12. A função de autocorrelação de um sinal telegráfico é dada por

$$R_X(\tau) = e^{-2\alpha|\tau|}$$

Calcule o espectro densidade de potência deste processo.

$$\text{Resp: } S_X(f) = \frac{4\alpha}{4\alpha^2 + (2\pi f)^2}$$

13. Um processo estocástico $X(t)$, estacionário no sentido amplo, com função de autocorrelação

$$R_X(\tau) = e^{-a|\tau|}$$

onde a é uma constante positiva real, é aplicado à entrada de um sistema linear invariante no tempo com resposta a impulso

$$h(t) = e^{-bt}u(t)$$

onde b é uma constante real positiva. Encontre a função de autocorrelação da saída $Y(t)$ do sistema.

$$\text{Resp: } R_Y(\tau) = \frac{1}{(a^2 - b^2)b} \left[ae^{-b|\tau|} - be^{-a|\tau|} \right]$$

14. Seja um processo ruído branco cujas componentes de frequência são limitadas à faixa $-W \leq f \leq W$. Determine:

- (a) O espectro densidade de potência.
- (b) A função de autocorrelação.
- (c) A potência média do processo.

Resp:

$$(a) S_X(f) = \begin{cases} \frac{N_0}{2}, & |f| \leq W \\ 0, & \text{caso contrário} \end{cases}$$

$$(b) R_X(\tau) = N_0 W \operatorname{sinc}(2W\tau)$$

$$(c) P = N_0 W$$

15. Dois processos estocásticos $X(t)$ e $Y(t)$ são dados por

$$X(t) = A \cos(\omega t + \Theta) \quad Y(t) = A \sin(\omega t + \Theta)$$

onde A e ω são constantes e Θ é uma v.a. com distribuição uniforme sobre o intervalo $(0, 2\pi)$.

(a) Encontre a correlação cruzada entre $X(t)$ e $Y(t)$.

(b) Mostre que $R_{XY}(\tau) = R_{YX}(-\tau)$

Resp:

$$(a) R_{XY}(t, t + \tau) = \frac{A^2}{2} \sin(\omega\tau)$$

$$R_{YX}(t, t + \tau) = -\frac{A^2}{2} \sin(\omega\tau)$$

(b)

16. Mostre que o espectro densidade de potência de um sinal real é real e par.

17. Seja $Y(n) = X(n) + W(n)$, onde $X(n) = A$ (para todo n) e A é uma v.a. com média zero e variância σ_A^2 , e $W(n)$ é um ruído branco discreto de potência média σ^2 . Assuma também que $X(n)$ e $Y(n)$ são independentes.

(a) Mostre que $Y(n)$ é estacionário no sentido amplo.

(b) Encontre o espectro densidade de potência de $Y(n)$.

Resp:

$$(a) E[Y(n)] = 0$$

$$R_Y(n, n + k) = \sigma_A^2 + \sigma^2 \delta(k)$$

$$(b) S_Y(\omega) = 2\pi\sigma_A^2 \delta(\omega) + \sigma^2, \quad -\pi \leq \omega \leq \pi$$

18. Um processo estocástico $Y(t)$ é definido por

$$Y(t) = AX(t) \cos(\omega_c t + \Theta)$$

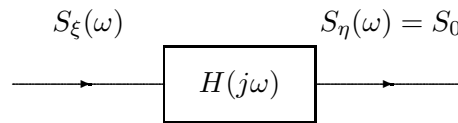
onde A e ω_c são constantes, Θ é uma v.a. com distribuição uniforme no intervalo $(-\pi, \pi)$, e $X(t)$ é um processo estocástico de média zero, função de autocorrelação $R_X(\tau)$, e espectro densidade de potência $S_X(\omega)$. Ainda, $X(t)$ e Θ são independentes. Mostre que $Y(t)$ é estacionário no sentido amplo e encontre o espectro densidade de potência de $Y(t)$.

Resp:

$$S_Y(\omega) = \frac{A^2}{4} [S_X(\omega - \omega_c) + S_X(\omega + \omega_c)]$$

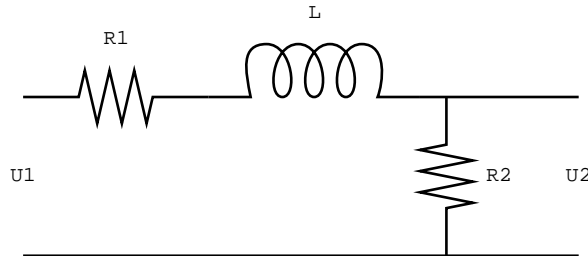
19. Na entrada de um filtro, tem-se um processo estocástico com espectro densidade de potência $S_\xi(\omega)$.

- Determine a resposta em frequência (amplitude) de um filtro para que a saída seja um ruído branco, ou seja, $S_\eta(\omega) = S_0$.
- Idem para um processo de entrada com $S_\xi(\omega) = S_0 \exp[-2\beta(\omega - \omega_0)^2]$.
- Idem para um processo de entrada com $S_\xi(\omega) = \frac{S_0\alpha^2}{\omega^2 + \alpha^2}$.



Resp: (a) $\sqrt{\frac{S_0}{S_\xi(\omega)}}$ (b) $e^{\beta(\omega - \omega_0)^2}$ (c) $\frac{1}{\alpha} \sqrt{\omega^2 + \alpha^2}$

20. Na entrada do circuito mostrado na Figura abaixo tem-se um ruído branco com $S_0 = 120\mu\text{V}^2/\text{Hz}$. Dados $R_1 = R_2 = 10^4\Omega$ e $L = 10^{-2}\text{H}$, calcule o espectro densidade de potência, a função de autocorrelação e a potência do processo de saída.



Dica: a função de transferência deste filtro é dada por

$$H(\omega) = \frac{U_2(\omega)}{U_1(\omega)} = \frac{R_2}{R_1 + R_2 + j\omega L} = \frac{\alpha}{1 + j\omega T}, \quad \alpha = \frac{R_2}{R_1 + R_2}, \quad T = \frac{L}{R_1 + R_2}.$$

Resp: $S_Y(\omega) = \frac{S_0\alpha^2}{1 + (\omega T)^2}$ $R_Y(\tau) = \frac{\alpha^2 S_0}{2T} e^{-\frac{|\tau|}{T}}$ $E[Y^2(t)] = \frac{\alpha^2 S_0}{T}$

21. Seja $Y(t) = X(t - d)$, onde d é um atraso constante e $X(t)$ é um processo estacionário no sentido amplo. Calcule $R_{YX}(\tau)$, $S_{YX}(f)$, $R_Y(\tau)$ e $S_Y(f)$, em função de $R_X(\tau)$ e $S_X(f)$.

Resp:

$$\begin{aligned} R_{YX}(\tau) &= R_X(\tau + d) & S_{YX}(f) &= S_X(f) \cos(2\pi f d) - j S_X(f) \sin(2\pi f d) \\ R_Y(\tau) &= R_X(\tau) & S_Y(f) &= S_X(f) \end{aligned}$$

22. Seja $X(t)$ um processo estocástico diferenciável, estacionário no sentido amplo. Seja também

$$Y(t) = \frac{d}{dt}X(t)$$

Encontre uma expressão para $S_Y(f)$ e $R_Y(\tau)$ em função de $S_X(f)$ e $R_X(\tau)$.

Dica: Para este sistema: $H(f) = j2\pi f$.

$$\text{Resp: } S_Y(f) = 4\pi^2 f^2 S_X(f) \quad R_Y(\tau) = -\frac{d^2}{d\tau^2} R_X(\tau)$$

23. Dois processos estocásticos $X(t)$ e $Y(t)$ são dados por

$$X(t) = A \cos(\omega t + \theta) \quad Y(t) = A \sin(\omega t + \theta)$$

onde A e ω são constantes, e θ é uma variável aleatória com distribuição uniforme no intervalo $(0, 2\pi)$.

- (a) Encontre a função de correlação cruzada entre $X(t)$ e $Y(t)$.
- (b) Mostre que $R_{XY}(-\tau) = R_{YX}(\tau)$.

$$\text{Resp: } R_{XY}(\tau) = \frac{A^2}{2} \sin(\omega\tau)$$

24. Em relação ao espectro densidade de potência $S_X(\omega)$:

- (a) Mostre que $S_X(\omega)$ é real.
- (b) Mostre que $S_X(\omega)$ é par.

Dica: use a identidade de Euler: $e^{j\theta} = \cos(\theta) + j \sin(\theta)$ e os conceitos de funções pares e ímpares.

Capítulo 10

Cadeias de Markov

Em geral, uma variável aleatória dentro de um conjunto, definindo um processo estocástico, não é independente e de fato pode ser estatisticamente dependente de várias formas complexas. Neste capítulo será introduzida a classe dos processos aleatórios de Markov que tem uma forma simples de dependência e é bastante utilizada em modelamento de problemas encontrados na prática.

10.1 Processos de Markov

Definição 10.1. Um processo aleatório $X(t)$ é um **processo de Markov** se o futuro, dado o presente, é independente do passado, isto é, para instantes arbitrários $t_1 < t_2 < \dots < t_n < t_{n+1}$,

$$P[X(t_{n+1}) = x_{n+1} | X(t_n) = x_n, X(t_{n-1}) = x_{n-1}, \dots, X(t_1) = x_1] = P[X(t_{n+1}) = x_{n+1} | X(t_n) = x_n] \quad (10.1)$$

se $X(t)$ assume valores discretos, e

$$P[a < X(t_{n+1}) \leq b | X(t_n) = x_n, X(t_{n-1}) = x_{n-1}, \dots, X(t_1) = x_1] = P[a < X(t_{n+1}) \leq b | X(t_n) = x_n] \quad (10.2)$$

se $X(t)$ assume valores contínuos.

Se as amostras de $X(t)$ são conjuntamente contínuas, então a equação (10.2) é equivalente a

$$f_{X(t_{n+1})}(x_{n+1} | X(t_n) = x_n, \dots, X(t_1) = x_1) = f_{X(t_{n+1})}(x_{n+1} | X(t_n) = x_n) \quad (10.3)$$

Chamaremos as equações (10.1), (10.2) e (10.3) como a **propriedade de Markov**. Nas expressões acima t_n é o presente, t_{n+1} o futuro, e t_1, \dots, t_{n-1} , o passado.

Desta maneira, para os processos de Markov, as fmp's e fdp's que são condicionadas a vários instantes de tempo, sempre se reduzirão a fmp's e fdp's condicionadas apenas ao mais recente instante de tempo. Por esta razão nos referimos ao valor de $X(t)$ no instante t como o **estado** do processo no instante t .

Exemplo 10.1. *Verifique se o processo de soma*

$$S_n = X_1 + X_2 + \cdots + X_n = S_{n-1} + X_n$$

onde os X_i 's são uma sequência de variáveis aleatórias independentes e identicamente distribuídas e onde $S_0 = 0$, é um processo de Markov.

Solução. S_n é um processo de Markov, pois

$$\begin{aligned} P[S_{n+1} = s_{n+1} | S_n = s_n, S_{n-1} = s_{n-1}, \dots, S_1 = s_1] &= P[X_{n+1} = S_{n+1} - S_n] \\ &= P[S_{n+1} = s_{n+1} | S_n = s_n] \end{aligned}$$

Exemplo 10.2. *Considere média móvel de uma sequência de Bernoulli*

$$Y_n = \frac{1}{2}(X_n + X_{n-1})$$

onde os X_i são sequências independentes de Bernoulli, com $p = 1/2$. Verifique se Y_n é ou não um processo de Markov.

Solução. A fmp de Y_n é

$$\begin{aligned} P[Y_n = 0] &= P[X_n = 0, X_{n-1} = 0] = \frac{1}{4} \\ P[Y_n = 1/2] &= P[X_n = 0, X_{n-1} = 1] + P[X_n = 1, X_{n-1} = 0] = \frac{1}{2} \\ P[Y_n = 1] &= P[X_n = 1, X_{n-1} = 1] = \frac{1}{4} \end{aligned}$$

Consideremos agora as seguintes probabilidades condicionais para dois valores consecutivos de Y_n :

$$\begin{aligned} P[Y_n = 1 | Y_{n-1} = 1/2] &= \frac{P[Y_n = 1, Y_{n-1} = 1/2]}{P[Y_{n-1} = 1/2]} \\ &= \frac{P[X_n = 1, X_{n-1} = 1, X_{n-2} = 0]}{1/2} \\ &= \frac{(1/2)^3}{1/2} = \frac{1}{4} \end{aligned}$$

Suponhamos agora que temos um conhecimento adicional sobre o passado:

$$\begin{aligned}
P\left[Y_n = 1 | Y_{n-1} = \frac{1}{2}, Y_{n-2} = 0\right] &= \frac{P[Y_n = 1, Y_{n-1} = 1/2, Y_{n-2} = 0]}{P[Y_{n-1} = 1/2, Y_{n-2} = 0]} \\
&= \frac{P[X_n = 1, X_{n-1} = 1, X_{n-2} = 0, X_{n-3} = 0]}{P[X_{n-1} = 1, X_{n-2} = 0, X_{n-3} = 0]} \\
&= \frac{1/16}{1/8} = \frac{1}{2}
\end{aligned}$$

Desta forma,

$$P\left[Y_n = 1 | Y_{n-1} = \frac{1}{2}, Y_{n-2} = 0\right] \neq P\left[Y_n = 1 | Y_{n-1} = \frac{1}{2}\right]$$

e este não é um processo de Markov.

Definição 10.2. *Um processo de Markov que assume somente valores inteiros é chamado de **Cadeia de Markov**.*

No restante deste capítulo iremos nos ater às Cadeias de Markov.

Se $X(t)$ é uma cadeia de Markov, então a fmp conjunta para três instantes de tempo arbitrários é

$$\begin{aligned}
&P[X(t_3) = x_3, X(t_2) = x_2, X(t_1) = x_1] = \\
&= P[X(t_3) = x_3 | X(t_2) = x_2, X(t_1) = x_1] P[X(t_2) = x_2, X(t_1) = x_1] \\
&= P[X(t_3) = x_3 | X(t_2) = x_2] P[X(t_2) = x_2, X(t_1) = x_1] \\
&= P[X(t_3) = x_3 | X(t_2) = x_2] P[X(t_2) = x_2 | X(t_1) = x_1] P[X(t_1) = x_1]
\end{aligned} \tag{10.4}$$

onde usamos a definição de probabilidade condicional e a propriedade de Markov. Em geral, a fmp conjunta para $n + 1$ instantes de tempo arbitrários é

$$\begin{aligned}
&P[X(t_{n+1}) = x_{n+1}, X(t_n) = x_n, \dots, X(t_1) = x_1] \\
&= P[X(t_{n+1}) = x_{n+1} | X(t_n) = x_n] P[X(t_n) = x_n | X(t_{n-1}) = x_{n-1}] \cdots P[X(t_1) = x_1]
\end{aligned} \tag{10.5}$$

Desta forma a fmp conjunta de $X(t)$ em instantes de tempo arbitrários é dada pelo produto da fmp do instante de tempo inicial e as probabilidades para as transições de estado subsequentes. Evidentemente, as probabilidades de transição de estado determinam o comportamento estatístico de uma cadeia de Markov.

10.2 Cadeias de Markov de Tempo discreto

Seja X_n uma cadeia de Markov de tempo discreto, que começa em $n = 0$ com a seguinte fmp

$$p_j(0) \triangleq P[X_0 = j], \quad j = 0, 1, 2, \dots \quad (10.6)$$

Da equação (10.3) a fmp conjunta para os primeiros $n+1$ valores do processo é dada por

$$\begin{aligned} P[X_n = i_n, X_{n-1} = i_{n-1}, \dots, X_0 = i_0] = \\ P[X_n = i_n | X_{n-1} = i_{n-1}] P[X_{n-1} = i_{n-1} | X_{n-2} = i_{n-2}] \cdots \\ P[X_1 = i_1 | X_0 = i_0] P[X_0 = i_0] \end{aligned} \quad (10.7)$$

Desta forma a fmp conjunta para uma sequência particular é simplesmente o produto da probabilidade para o estado inicial com as probabilidades para as transições de um passo subsequentes.

Definição 10.3. Probabilidades de transição homogêneas: Uma cadeia de Markov X_n tem probabilidades de transição homogêneas se as probabilidades de transição para um passo são fixas e não variam com o tempo, isto é

$$P[X_{n+1} = j | X_n = i] = p_{ij}, \forall n \quad (10.8)$$

A fmp conjunta para X_n, X_{n-1}, \dots, X_0 é então dada por

$$P[X_n = i_n, X_{n-1} = i_{n-1}, \dots, X_0 = i_0] = p_{i_{n-1}, i_n} \cdots p_{i_0, i_1} p_{i_0}(0) \quad (10.9)$$

Desta forma X_n é completamente especificado pela fmp inicial $p_i(0)$ e pela matriz de probabilidades de transição de um passo P

$$P = \begin{bmatrix} p_{00} & p_{01} & p_{02} & \cdots \\ p_{10} & p_{11} & p_{12} & \cdots \\ \vdots & \vdots & \vdots & \cdots \\ p_{i-1,0} & p_{i-1,1} & p_{i-1,2} & \cdots \\ p_{i,0} & p_{i,1} & p_{i,2} & \cdots \\ \vdots & \vdots & \vdots & \cdots \end{bmatrix} \quad (10.10)$$

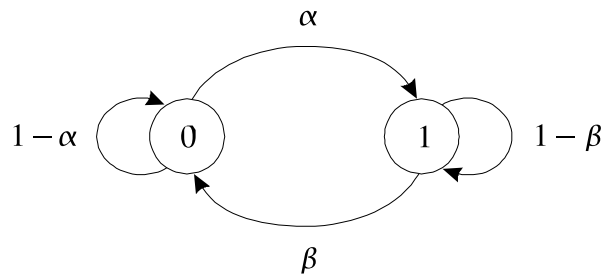
A matriz P é chamada de **matriz de probabilidade de transição**. Note que a soma de cada linha de P deve ser igual a 1

$$1 = \sum_j P[X_{n+1} = j | X_n = i] = \sum_j p_{ij} \quad (10.11)$$

Exemplo 10.3. Um modelo de Markov para transmissão de voz por pacotes assume que se o n -ésimo pacote contém silêncio, a probabilidade de silêncio no próximo pacote é $(1 - \alpha)$ e a probabilidade do pacote conter voz é α .

Similarmente, se o n -ésimo pacote contiver atividades de voz, a probabilidade do próximo pacote conter voz é $(1 - \beta)$, e a probabilidade de silêncio é β . Esboce uma cadeia de Markov para este problema.

Solução. Supondo X_n a função que indica a atividade voz em um determinado pacote no instante n , então X_n é uma cadeia de Markov de 2 estados e matriz de probabilidade de transição como mostrado abaixo.



$$P = \begin{bmatrix} 1 - \alpha & \alpha \\ \beta & 1 - \beta \end{bmatrix}$$

10.2.1 Probabilidade de transição para n passos

Para avaliar a fmp conjunta em instantes de tempo arbitrários (veja equação 10.5), precisamos conhecer as probabilidades de transição para um número arbitrário de passos. Seja $P(n) = \{p_{ij}(n)\}$ a matriz de probabilidades de transição para n passos, onde

$$p_{ij}(n) = P[X_{n+k} = j | X_k = i], \quad n, i, j \geq 0 \quad (10.12)$$

Note que $P[X_{n+k} = j | X_k = i] = P[X_n = j | X_0 = i] \quad \forall n \geq 0, \forall k \geq 0$, desde que as probabilidades de transição não dependem do tempo.

Consideremos primeiramente as probabilidades de transição para dois passos. A probabilidade de ir do estado i em $t = 0$, passando pelo estado k em $t = 1$, e terminando no estado j em $t = 2$ é

$$\begin{aligned}
 P[X_2 = j, X_1 = k | X_0 = i] &= \frac{P[X_2 = j, X_1 = k, X_0 = i]}{P[X_0 = i]} \\
 &= \frac{P[X_2 = j | X_1 = k] P[X_1 = k | X_0 = i] P[X_0 = i]}{P[X_0 = i]} \\
 &= P[X_2 = j | X_1 = k] P[X_1 = k | X_0 = i] \\
 &= p_{ik}(1) p_{kj}(1)
 \end{aligned}$$

Note que $p_{ik}(1)$ e $p_{kj}(1)$ são componentes de P , a matriz de transição de um passo. Obtemos $p_{ij}(2)$, a probabilidade de ir do estado i em $t = 0$ para o estado j em $t = 2$, somando sobre todos os possíveis estados intermediários k

$$p_{ij}(2) = \sum_k p_{ik}(1)p_{kj}(1) \quad \forall i, j \quad (10.13)$$

O conjunto de equações fornecido pela equação (10.13) afirma que a matriz $P(2)$ é obtida pela multiplicação das matrizes de transição de um passo

$$P(2) = P(1)P(1) = P^2 \quad (10.14a)$$

Através dos mesmos argumentos utilizados acima, verifica-se que $P(n)$ é encontrada multiplicando-se $P(n-1)$ por P

$$P(n) = P(n-1)P \quad (10.14b)$$

As equações (10.14a) e (10.14b) juntas implicam que

$$P(n) = P^n \quad (10.15)$$

isto é, a n -ésima matriz de probabilidades de transição é a n -ésima potência da matriz de probabilidades de transição de um passo.

10.2.2 Probabilidades dos estados

Consideremos agora as probabilidades dos estados no instante n . Seja $\mathbf{p}(n) = p_j(n)$ o vetor (linha) de **probabilidades de estados** no instante n . A probabilidade $p_j(n)$ relaciona-se a $\mathbf{p}(n-1)$ através da expressão

$$\begin{aligned} p_j(n) &= \sum_i P[X_n = j | X_{n-1} = i] P[X_{n-1} = i] \\ &= \sum_i p_{ij} p_i(n-1) \end{aligned} \quad (10.16)$$

A equação (10.16) afirma que $\mathbf{p}(n)$ é obtida pela multiplicação do vetor linha $\mathbf{p}(n-1)$ pela matriz P

$$\mathbf{p}(n) = \mathbf{p}(n-1)P \quad (10.17)$$

Similarmente, $p_j(n)$ está relacionada a $\mathbf{p}(0)$ por

$$\begin{aligned} p_j(n) &= \sum_i P[X_n = j | X_0 = i] P[X_0 = i] \\ &= \sum_i p_{ij}(n) p_i(0) \end{aligned} \quad (10.18)$$

e em notação matricial

$$\mathbf{p}(n) = \mathbf{p}(0)P(n) = \mathbf{p}(0)P^n \quad n = 1, 2, \dots \quad (10.19)$$

Então a fmp de um estado no instante n é obtida multiplicando-se a fmp do estado inicial por P^n .

Exemplo 10.4. *Seja $\alpha = 1/10$ e $\beta = 1/5$ no Exemplo 10.3. Encontre $P(n)$ para $n = 2, 4, 8, 16$*

Solução.

$$P^2 = \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix}^2 = \begin{bmatrix} 0.83 & 0.17 \\ 0.34 & 0.66 \end{bmatrix}$$

$$P^4 = \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix}^4 = \begin{bmatrix} 0.7467 & 0.2533 \\ 0.5066 & 0.4934 \end{bmatrix}$$

$$P^8 = \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix}^8 = \begin{bmatrix} 0.6859 & 0.3141 \\ 0.6282 & 0.3718 \end{bmatrix}$$

$$P^{16} = \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix}^{16} = \begin{bmatrix} 0.6678 & 0.3322 \\ 0.6644 & 0.3356 \end{bmatrix}$$

Existe uma clara tendência aqui: à medida que $n \rightarrow \infty$,

$$P^n \rightarrow \begin{bmatrix} 2/3 & 1/3 \\ 2/3 & 1/3 \end{bmatrix}$$

De fato, podemos mostrar com um pouco de álgebra linear que

$$P^n = \frac{1}{\alpha + \beta} \begin{bmatrix} \beta & \alpha \\ \beta & \alpha \end{bmatrix} + \frac{(1 - \alpha - \beta)^n}{\alpha + \beta} \begin{bmatrix} \alpha & -\alpha \\ -\beta & \beta \end{bmatrix}$$

que claramente aproxima

$$\frac{1}{\alpha + \beta} \begin{bmatrix} \beta & \alpha \\ \beta & \alpha \end{bmatrix} = \begin{bmatrix} 2/3 & 1/3 \\ 2/3 & 1/3 \end{bmatrix}$$

Exemplo 10.5. *No exemplo 10.4 sejam as probabilidades iniciais para os estados dadas por*

$$P[X_0 = 0] = p_0(0) \quad e \quad P[X_0 = 1] = 1 - p_0(0)$$

Encontre as probabilidades dos estados à medida que $n \rightarrow \infty$.

Solução. O vetor de probabilidades de estados no instante n é

$$\mathbf{p}(n) = [p_0(0), 1 - p_0(0)]P^n$$

À medida que $n \rightarrow \infty$, temos que

$$\mathbf{p}(n) = [p_0(0), 1 - p_0(0)] \begin{bmatrix} 2/3 & 1/3 \\ 2/3 & 1/3 \end{bmatrix} = \left[\frac{2}{3}, \frac{1}{3} \right]$$

Podemos ver que as probabilidades dos estados não dependem das probabilidades do estado inicial, à medida que $n \rightarrow \infty$.

10.2.3 Probabilidades em regime

O exemplo 10.5 é típico de cadeias de Markov que entram em regime estacionário depois que o processo está em vigor durante um longo tempo. À medida que $n \rightarrow \infty$, a matriz de transição de n passos aproxima-se de uma matriz para a qual todas as linhas são iguais à mesma fmp, isto é

$$p_{ij}(n) \rightarrow \pi_j, \forall i \quad (10.20)$$

À medida que $n \rightarrow \infty$, a equação (10.18) torna-se

$$p_j(n) \rightarrow \sum_i \pi_j p_i(0) = \pi_j \quad (10.21)$$

Definição 10.4. Sistema em equilíbrio ou regime permanente. *Uma cadeia de Markov está em equilíbrio ou regime permanente quando, à medida que $n \rightarrow \infty$, a probabilidade do estado j aproxima-se de uma constante independente do tempo e das probabilidades do estado inicial:*

$$p_j(n) \rightarrow \pi_j, \forall j \quad (10.22)$$

Podemos encontrar a fmp $\pi \triangleq \{\pi_j\}$ (onde π é um vetor linha) na equação (10.22) (quando existir) notando que à medida que $n \rightarrow \infty$, $p_j(n) \rightarrow \pi_j$ e $p_i(n-1) \rightarrow \pi_i$, de modo que a equação (10.16) aproxima

$$\pi_j = \sum_i p_{ij} \pi_i \quad (10.23a)$$

que em notação matricial fica

$$\pi = \pi P \quad (10.23b)$$

Em geral, a equação (10.23b) tem $(n-1)$ equações linearmente independentes. A equação adicional necessária é dada por

$$\sum_i \pi_i = 1 \quad (10.23c)$$

Nos referimos a π como a **fmp de regime permanente** da cadeia de Markov. Se iniciamos a cadeia de Markov com fmp de estado inicial $\mathbf{p}(0) = \pi$, então pelas equações (10.19) e (10.23b) temos que o vetor de probabilidades de estados é dado por

$$\mathbf{p}(n) = \pi P^n = \pi, \quad \forall n \quad (10.24)$$

O processo resultante é estacionário, desde que a probabilidade da sequência de estados i_0, i_1, \dots, i_n iniciando no instante k é, pela equação (10.7)

$$\begin{aligned}
P[X_{n+k} = i_n, \dots, X_k = i_0] &= \\
&= P[X_{n+k} = i_n | X_{n+k-1} = i_{n-1}] \cdots P[X_{1+k} = i_1 | X_k = i_0] P[X_k = i_0] \quad (10.25) \\
&= p_{i_{n-1}, i_n} \cdots p_{i_0, i_1} \pi_{i_0}
\end{aligned}$$

a qual é independente do instante inicial k . Então as probabilidades são independentes da escolha da origem dos tempos, e o processo é estacionário.

Observação:

Note que, como o processo está em regime, as Equações (10.23) e (10.24) são equivalentes. Em outras palavras, em regime permanente, as probabilidades dos estados são sempre as mesmas, independentemente do número de transições efetuadas.

Exemplo 10.6. *Encontre a fmp estacionária de estados para o processo do exemplo 10.3*

Solução. A equação (10.23a) fornece

$$\begin{aligned}
\pi_0 &= (1 - \alpha)\pi_0 + \beta\pi_1 \\
\pi_1 &= \alpha\pi_0 + (1 - \beta)\pi_1
\end{aligned}$$

o que implica que $\alpha\pi_0 = \beta\pi_1 = \beta(1 - \pi_0)$ desde que $\pi_0 + \pi_1 = 1$. Então, para $\alpha = 1/10$ e $\beta = 1/5$, temos

$$\pi_0 = \frac{\beta}{\alpha + \beta} = \frac{2}{3} \quad \pi_1 = \frac{\alpha}{\alpha + \beta} = \frac{1}{3}$$

10.3 Cadeias de Markov em tempo contínuo

Na seção 10.2 vimos que a matriz de probabilidades de transição determina o comportamento de uma cadeia de Markov de tempo discreto. Nesta seção iremos ver que o mesmo acontece com cadeias de Markov de tempo contínuo.

A fmp conjunta para $(k+1)$ instantes de tempo arbitrários de uma cadeia de Markov é dada pela equação (10.5)

$$\begin{aligned}
P[X(t_{n+1}) = x_{n+1}, X(t_n) = x_n, \dots, X(t_1) = x_1] \\
= P[X(t_{n+1}) = x_{n+1} | X(t_n) = x_n] \cdots P[X(t_2) = x_2 | X(t_1) = x_1] P[X(t_1) = x_1] \quad (10.26)
\end{aligned}$$

Este resultado vale independente do processo ser de tempo discreto ou de tempo contínuo. No caso contínuo, a equação (10.26) requer que saibamos as probabilidades de transição no intervalo entre um instante de tempo arbitrário s e outro instante de tempo arbitrário $s + t$:

$$P[X(s + t) = j | X(s) = i], \quad t \geq 0$$

Assumimos aqui que as probabilidades de transição dependem somente da diferença entre os dois instantes de tempo:

$$P[X(s+t) = j | X(s) = i] = P[X(t) = j | X(0) = i] = p_{ij}(t), \quad t \geq 0, \forall s \quad (10.27)$$

Dizemos que $X(t)$ tem **probabilidades de transição homogêneas**.

Teorema 10.1. *Seja $P(t) = \{p_{ij}(t)\}$ a matriz de probabilidades de transição em um intervalo de comprimento t . Desde que $p_{ii}(0) = 1$ e $p_{ij}(0) = 0$ para $i \neq j$, temos*

$$P(0) = \mathbf{I} \quad (10.28)$$

onde \mathbf{I} é a matriz identidade.

Exemplo 10.7. *Para o processo de Poisson, as probabilidades de transição satisfazem*

$$\begin{aligned} p_{ij}(t) &= P[j - i \text{ eventos em } t \text{ segundos}] \\ &= p_{0, j-i}(t) \\ &= \frac{(\alpha t)^{j-i}}{(j-i)!} e^{-\alpha t}, \quad j \geq i \end{aligned}$$

Portanto

$$P = \begin{bmatrix} e^{-\alpha t} & \alpha t e^{-\alpha t} & (\alpha t)^2 e^{-\alpha t}/2! & (\alpha t)^3 e^{-\alpha t}/3! & \dots \\ 0 & e^{-\alpha t} & \alpha t e^{-\alpha t} & (\alpha t)^2 e^{-\alpha t}/2! & \dots \\ 0 & 0 & e^{-\alpha t} & \alpha t e^{-\alpha t} & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}$$

À medida que $t \rightarrow 0$, $e^{-\alpha t} \approx 1 - \alpha t$. Então para um intervalo de tempo pequeno δ ,

$$P \approx \begin{bmatrix} 1 - \alpha \delta & \alpha \delta & 0 & \dots \\ 0 & 1 - \alpha \delta & \alpha \delta & \dots \\ 0 & 0 & 1 - \alpha \delta & \dots \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}$$

onde todos os termos de ordem δ^2 ou superior foram negligenciados. Então a probabilidade de mais de uma transição em um intervalo de tempo bastante curto é desprezível.

Exemplo 10.8. *Para um processo telegráfico aleatório, $X(t)$ muda com cada ocorrência de um evento em um processo de Poisson. Vimos na seção 8.7 que as probabilidades de transição para este processo são*

$$P[X(t) = a | X(0) = a] = \frac{1}{2} \{1 + e^{-2\alpha t}\}$$

$$P[X(t) = a | X(0) = b] = \frac{1}{2} \{1 - e^{-2\alpha t}\}, \quad \text{se } a \neq b$$

Então a matriz de probabilidade de transição é

$$P(t) = \begin{bmatrix} 1/2\{1 + e^{-2\alpha t}\} & 1/2\{1 - e^{-2\alpha t}\} \\ 1/2\{1 - e^{-2\alpha t}\} & 1/2\{1 + e^{-2\alpha t}\} \end{bmatrix}$$

10.3.1 Tempos de ocupação de estados

Desde que o sinal telegráfico aleatório muda de polaridade com cada ocorrência de um evento em um processo de Poisson, segue que o tempo em que o sistema permanece em cada estado é uma variável aleatória exponencial. Desta forma esta é uma propriedade do **tempo de ocupação de estados** para todas as cadeias de Markov de tempo contínuo, isto é: $X(t)$ permanece em um dado valor (estado) para um intervalo de tempo aleatório exponencialmente distribuído.

Para ver como isto acontece, seja T_i o tempo gasto no estado i . A probabilidade de gastar mais de t segundos neste estado é então

$$P[T_i > t]$$

Suponha agora que o processo já tenha estado no estado i por s segundos; então a probabilidade de gastar mais t segundos neste estado é

$$P[T_i > t + s | T_i > s] = P[T_i > t + s | X(s') = i, 0 \leq s' \leq s],$$

desde que $\{T_i > s\}$ implica que o sistema tem estado no estado i durante o intervalo de tempo $(0, s)$. A propriedade de Markov implica que se $X(s) = i$, então o passado é irrelevante e podemos ver o sistema como sendo reiniciado no estado i no instante s :

$$P[T_i > t + s | T_i > s] = P[T_i > t] \quad (10.29)$$

Somente a variável aleatória exponencial satisfaz esta propriedade de ser sem memória. Então o tempo gasto no estado i é uma variável aleatória exponencial com alguma média $1/v_i$:

$$P[T_i > t] = e^{-v_i t} \quad (10.30)$$

O **tempo médio de ocupação de estado** $1/v_i$ irá geralmente ser diferente para cada estado.

O resultado acima nos dá uma outra maneira de olhar para cadeias de Markov de tempo contínuo. A cada vez que um estado i é alcançado, seleciona-se um tempo de ocupação de estado T_i exponencialmente distribuído. Quando o tempo se esgota, o próximo estado j é selecionado de acordo com uma cadeia de Markov de tempo discreto, com probabilidades de transição \tilde{q}_{ij} . Então o novo tempo de ocupação de estado é selecionado de acordo com T_j , e assim por diante. Chamamos \tilde{q}_{ij} de uma **cadeia de Markov embutida**.

Exemplo 10.9. *O sinal telegráfico aleatório do exemplo 10.8 gasta um tempo exponencialmente distribuído com média $1/\alpha$ em cada estado. Quando uma transição ocorre, a transição é sempre do estado presente para um único outro estado, então a cadeia de Markov embutida é*

$$\begin{array}{ll} \tilde{q}_{00} = 0 & \tilde{q}_{01} = 1 \\ \tilde{q}_{10} = 1 & \tilde{q}_{11} = 0 \end{array}$$

10.3.2 Taxas de transição e probabilidades de estados dependentes de tempo

Considere as probabilidades de transição em um intervalo de tempo bastante curto de duração δ segundos. A probabilidade de o processo permanecer no estado i durante o intervalo é

$$\begin{aligned} P[T_i > \delta] &= e^{-v_i \delta} \\ &= 1 - \frac{v_i \delta}{1!} + \frac{v_i^2 \delta^2}{2!} - \dots \\ &= 1 - v_i \delta + o(\delta) \end{aligned}$$

onde $o(\delta)$ denota os termos que se tornam desprezíveis em relação a δ à medida que $\delta \rightarrow 0^1$. As distribuições exponenciais para os tempos de ocupação de estados implicam que a probabilidade de duas ou mais transições em um intervalo de duração δ é $o(\delta)$. Então para δ pequeno, $p_{ii}(\delta)$ é aproximadamente igual à probabilidade de o processo permanecer no estado i por δ segundos:

$$p_{ii}(\delta) \approx P[T_i > \delta] = 1 - v_i \delta + o(\delta)$$

ou equivalentemente,

$$1 - p_{ii}(\delta) = v_i \delta - o(\delta) \quad (10.31)$$

Chamamos v_i a taxa na qual o processo $X(t)$ deixa o estado i .

Uma vez que o processo deixa o estado i , ele entra no estado j com probabilidade \tilde{q}_{ij} . Então

$$\begin{aligned} p_{ij}(\delta) &= (1 - p_{ii}(\delta))\tilde{q}_{ij} \\ &= v_i \tilde{q}_{ij} \delta - o(\delta) \\ &= \gamma_{ij} \delta - o(\delta) \end{aligned} \quad (10.32)$$

Chamamos $\gamma_{ij} = v_i \tilde{q}_{ij}$ a taxa na qual o processo $X(t)$ entra no estado j partindo do estado i . Definimos $\gamma_{ii} = -v_i$, e pela equação (10.31),

¹Uma função $g(h)$ é $o(h)$ se $\lim_{h \rightarrow 0} \frac{g(h)}{h} = 0$, isto é, se $g(h)$ tende a zero mais rápido do que h .

$$p_{ii}(\delta) - 1 = \gamma_{ii}\delta - o(\delta) \quad (10.33)$$

Se dividirmos ambos os lados das equações (10.32) e (10.33) por δ e tomarmos o limite $\delta \rightarrow 0$, obtemos

$$\lim_{\delta \rightarrow 0} \frac{p_{ij}(\delta)}{\delta} = \gamma_{ij}, \quad i \neq j \quad (10.34a)$$

e

$$\lim_{\delta \rightarrow 0} \frac{p_{ii}(\delta) - 1}{\delta} = \gamma_{ii}, \quad (10.34b)$$

desde que

$$\lim_{\delta \rightarrow 0} \frac{o(\delta)}{\delta} = 0$$

pois $o(\delta)$ é de ordem superior a δ .

Podemos então desenvolver um conjunto de equações para encontrar as probabilidades dos estados no instante t , que serão denotados por

$$p_j(t) \triangleq P[X(t) = j].$$

Para $\delta > 0$, temos (veja Figura 10.1)

$$\begin{aligned} p_j(t + \delta) &= P[X(t + \delta) = j] \\ &= \sum_i P[X(t + \delta) = j | X(t) = i] P[X(t) = i] \\ &= \sum_i p_{ij}(\delta) p_i(t) \end{aligned} \quad (10.35)$$

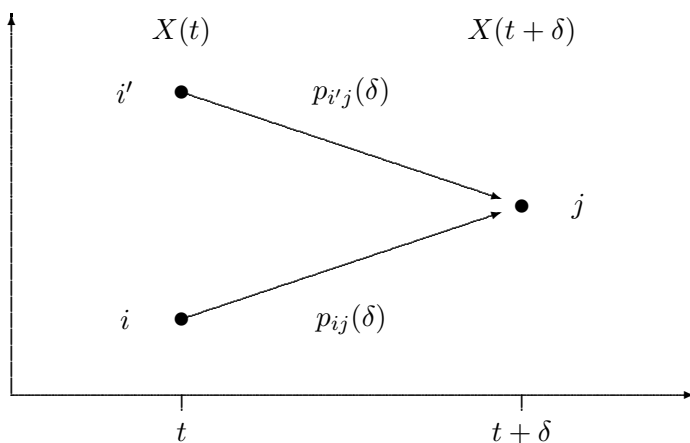


Figura 10.1: Transições para o estado j .

Se subtrairmos $p_j(t)$ de ambos os lados, obtemos

$$\begin{aligned}
p_j(t + \delta) - p_j(t) &= \sum_i p_{ij}(\delta) p_i(t) - p_j(t) \\
&= \sum_{i \neq j} p_{ij}(\delta) p_i(t) + p_{jj}(\delta) p_j(t) - p_j(t) \\
&= \sum_{i \neq j} p_{ij}(\delta) p_i(t) + (p_{jj}(\delta) - 1) p_j(t) \quad (10.36)
\end{aligned}$$

Se dividirmos ambos os membros por δ , aplicarmos (10.34a) e (10.34b), e fizermos $\delta \rightarrow 0$, obtemos

$$p'_j(t) = \sum_i \gamma_{ij} p_i(t) \quad (10.37)$$

A Equação (10.37) é uma das formas das **Equações de Chapman-Kolmogorov** para cadeias de Markov de tempo contínuo. Para encontrar $p_j(t)$ precisamos resolver este sistema de equações diferenciais com condições iniciais especificadas pela fmp de estado inicial $\{p_j(0), j = 0, 1, \dots\}$.

Importante:

Note que se resolvemos a Equação (10.37) com a suposição de que o sistema estava no estado i no instante inicial, isto é, com condição inicial $p_i(0) = 1$ e $p_j(0) = 0$ para todo $j \neq i$, então a solução é de fato $p_{ij}(t)$, a componente ij de $P(t)$. Então a Equação (10.37) pode também ser utilizada para encontrar a matriz de probabilidades de transição. Veja o exemplo abaixo:

Exemplo 10.10. *Um sistema de filas alterna entre dois estados. No estado 0, o sistema está livre e esperando a chegada de um cliente. Este tempo desocupado é uma v.a. exponencial com média $1/\alpha$. No estado 1, o sistema está ocupado servindo um usuário. O tempo no estado ocupado é uma v.a. exponencial com média $1/\beta$. Encontre as probabilidades dos estados $p_0(t)$ e $p_1(t)$ em termos das probabilidades dos estados iniciais $p_0(0)$ e $p_1(0)$.*

Solução. O sistema passa do estado 0 para o estado 1 a uma taxa α , e do estado 1 para o estado 0 a uma taxa β :

$$\begin{aligned}
\gamma_{00} &= -\alpha & \gamma_{01} &= \alpha \\
\gamma_{10} &= \beta & \gamma_{11} &= -\beta
\end{aligned}$$

A Equação (10.37) fornece então

$$\begin{aligned}
p'_0(t) &= -\alpha p_0(t) + \beta p_1(t) \\
p'_1(t) &= \alpha p_0(t) - \beta p_1(t)
\end{aligned}$$

Desde que $p_0(t) + p_1(t) = 1$, a primeira equação torna-se

$$p_0'(t) = -\alpha p_0(t) + \beta(1 - p_0(t))$$

que é uma equação diferencial de primeira ordem:

$$p_0'(t) + (\alpha + \beta)p_0(t) = \beta \quad p_0(0) = p_0$$

A solução geral desta equação é

$$p_0(t) = \frac{\beta}{\alpha + \beta} + Ce^{-(\alpha + \beta)t}$$

Obtemos C fazendo $t = 0$ e resolvendo em termos de $p_0(0)$. Assim, encontramos

$$p_0(t) = \frac{\beta}{\alpha + \beta} + \left(p_0(0) - \frac{\beta}{\alpha + \beta} \right) e^{-(\alpha + \beta)t}$$

Similarmente, temos que

$$p_1(t) = \frac{\alpha}{\alpha + \beta} + \left(p_1(0) - \frac{\alpha}{\alpha + \beta} \right) e^{-(\alpha + \beta)t}$$

Note que à medida que $t \rightarrow \infty$

$$p_0(t) \rightarrow \frac{\beta}{\alpha + \beta} \quad \text{e} \quad p_1(t) \rightarrow \frac{\alpha}{\alpha + \beta}$$

Então, à medida que $t \rightarrow \infty$, as probabilidades dos estados se aproximam de valores constantes que são independentes das probabilidades iniciais dos estados.

Exemplo 10.11. *Encontre as probabilidades dos estados para o processo de Poisson.*

Solução. O processo de Poisson move-se somente do estado i para o estado $i + 1$ a uma taxa α . Então

$$\gamma_{ii} = -\alpha \quad \gamma_{i,i+1} = \alpha$$

A Equação (10.37) fornece então

$$\begin{aligned} p_0'(t) &= -\alpha p_0(t), & j &= 0 \\ p_j'(t) &= -\alpha p_j(t) + \alpha p_{j-1}(t), & j &\geq 1 \end{aligned}$$

A condição inicial para o processo de Poisson é $p_0(0) = 1$, de modo que a solução para a primeira equação é

$$p_0(t) = e^{-\alpha t}$$

Para a segunda equação, temos

$$p_1'(t) = -\alpha p_1(t) + \alpha e^{-\alpha t}, \quad p_1(0) = 0$$

que também é uma equação diferencial de primeira ordem, cuja solução é

$$p_1(t) = \frac{\alpha t}{1!} e^{-\alpha t}$$

Adicionalmente pode-se mostrar através de indução que a solução para o estado j é dada por

$$p_j(t) = \frac{(\alpha t)^j}{j!} e^{-\alpha t}$$

Note que para qualquer j , $p_j(t) \rightarrow 0$ à medida que $t \rightarrow \infty$. Então para o processo de Poisson, a probabilidade de qualquer estado finito tende a zero à medida que $t \rightarrow \infty$. Isto é consistente com o fato de que o processo cresce de forma constante com o tempo.

10.4 Probabilidades de Estados em Regime e Equações de Balanço Globais

À medida que $t \rightarrow \infty$, as probabilidades dos estados do sistema de filas do Exemplo 10.10 convergem para uma fmp que não depende das condições iniciais. Este comportamento é típico de sistemas que alcançam uma condição de “equilíbrio” ou “regime permanente”. Para tais sistemas, $p_j(t) \rightarrow p_j$ e $p'_j(t) \rightarrow 0$, de modo que a Equação (10.37) torna-se

$$0 = \sum_i \gamma_{ij} p_i, \quad \forall j, \quad (10.38a)$$

ou equivalentemente, lembrando que $\gamma_{jj} = -v_j$,

$$v_j p_j = \sum_{i \neq j} \gamma_{ij} p_i, \quad \forall j \quad (10.38b)$$

A Equação (10.38b) pode ser reescrita como

$$p_j \left(\sum_{i \neq j} \gamma_{ji} \right) = \sum_{i \neq j} \gamma_{ij} p_i \quad (10.38c)$$

desde que

$$v_j = \sum_{i \neq j} \gamma_{ji}$$

O sistema de equações lineares dado pelas Equações (10.38b) ou (10.38c) é chamado de **Equações de Balanço Global**. Estas equações afirmam que, em equilíbrio, a probabilidade do fluxo para fora do estado j , dada por $v_j p_j$, é igual à probabilidade do fluxo para dentro do estado j , como mostrado na Figura 10.2. Resolvendo este conjunto de equações lineares podemos obter a fmp dos estados do sistema em regime permanente (quando existir).

Referimo-nos a $\mathbf{p} = \{p_i\}$ como a **fmp estacionária dos estados** da cadeia de Markov. Desde que \mathbf{p} satisfaz a Equação (10.37), se iniciamos a cadeia de Markov com uma fmp inicial dada por \mathbf{p} , então as probabilidades dos estados serão

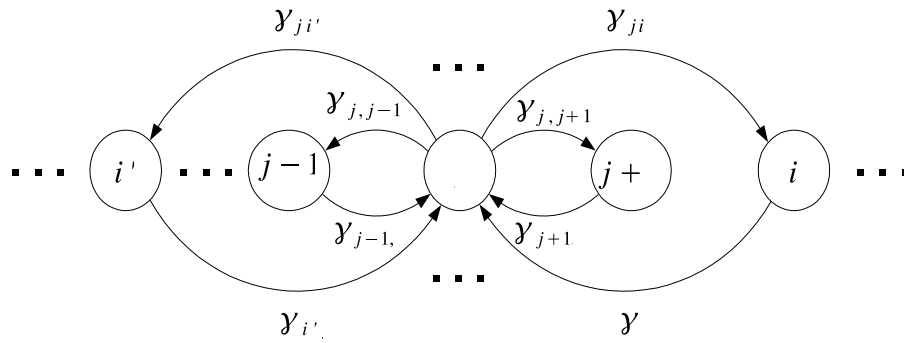


Figura 10.2: Balanço global de fluxo de probabilidade.

$$p_i(t) = p_i, \quad \forall t$$

O processo resultante é estacionário, desde que a probabilidade da sequência de estados i_0, i_1, \dots, i_n nos instantes $t < t_1 + t < \dots < t_n + t$ é, pela Equação (10.26),

$$P[X(t) = i_0, X(t_1 + t) = i_1, \dots, X(t_n + t) = i_n] = \\ P[X(t_n + t) = i_n | X(t_{n-1} + t) = i_{n-1}] \cdots \times P[X(t_1 + t) = i_1 | X(t) = i_0] P[X(t) = i_0]$$

As probabilidades de transição dependem somente da diferença entre os tempos associados. Então a probabilidade conjunta acima depende da escolha da origem apenas através de $P[X(t) = i_0]$. Mas $P[X(t) = i_0] = p_{i_0}$ para todo t . Portanto concluímos que a probabilidade conjunta acima é independente da escolha da origem dos tempos e que o processo é estacionário.

Exemplo 10.12. Encontre a fmp de estado estacionário para o sistema de filas de dois estados do Exemplo 10.10.

Solução. A Equação (10.38b) para este sistema fornece

$$\alpha p_0 = \beta p_1 \text{ e } \beta p_1 = \alpha p_0$$

Notando que $p_0 + p_1 = 1$, obtemos

$$p_0 = \frac{\beta}{\alpha + \beta} \text{ e } p_1 = \frac{\alpha}{\alpha + \beta}$$

Exemplo 10.13. Sistema de filas de servidor único $M/M/1$. Considere um sistema de filas no qual os clientes são servidos um de cada vez pela ordem de chegada. O tempo entre chegadas de clientes é exponencialmente distribuído com taxa λ , e o tempo requerido para atender um cliente é exponencialmente distribuído com taxa μ . Encontre a fmp para o número de clientes no sistema quando este está em regime permanente.

Solução. As taxas de transição de estados são as seguintes. Os clientes chegam a uma taxa λ , então

$$\gamma_{i,i+1} = \lambda \quad i = 0, 1, 2, \dots$$

Quando o sistema não está vazio, os clientes saem a uma taxa μ . Então

$$\gamma_{i,i-1} = \mu \quad i = 1, 2, 3, \dots$$

O diagrama de taxa de transição é mostrado na Figura 10.3.

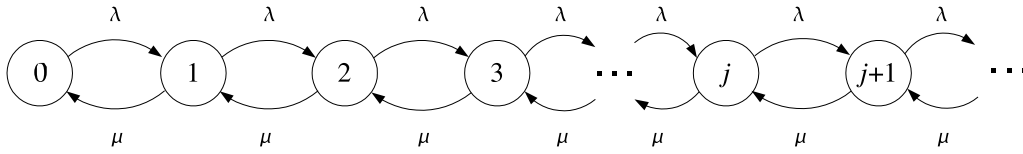


Figura 10.3: Diagrama de transição de estados para o sistema M/M/1.

As Equações de balanço global são

$$\lambda p_0 = \mu p_1, \quad j = 0 \quad (10.39a)$$

$$(\lambda + \mu)p_j = \lambda p_{j-1} + \mu p_{j+1}, \quad j = 1, 2, \dots \quad (10.39b)$$

Podemos reescrever a Equação (10.39b) como segue:

$$\lambda p_j - \mu p_{j+1} = \lambda p_{j-1} - \mu p_j, \quad j = 1, 2, \dots$$

o que implica que

$$\lambda p_{j-1} - \mu p_j = \text{constante}, \quad j = 1, 2, \dots \quad (10.40)$$

A Equação (10.40) com $j = 1$, juntamente com a Equação (10.39a), implica que

$$\text{constante} = \lambda p_0 - \mu p_1 = 0$$

Então a Equação (10.40) torna-se

$$\lambda p_{j-1} = \mu p_j$$

ou equivalentemente,

$$p_j = \rho p_{j-1}, \quad j = 1, 2, \dots$$

e por indução

$$p_j = \rho^j p_0$$

onde $\rho = \lambda/\mu$. Obtemos p_0 notando que a soma das probabilidades precisa ser igual a um:

$$1 = \sum_{j=0}^{\infty} p_j = (1 + \rho + \rho^2 + \dots) p_0 = \frac{1}{1 - \rho} p_0$$

onde a série converge se e somente se $\rho < 1$. Então

$$p_j = (1 - \rho)\rho^j, \quad j = 1, 2, \dots \quad (10.41)$$

A condição para a existência de uma solução de regime permanente tem uma explicação simples. A condição $\rho < 1$ é equivalente a

$$\lambda < \mu$$

isto é, a taxa na qual os clientes chegam precisa ser menor que a taxa na qual o sistema possa atendê-los. Caso contrário, a fila cresce sem limite à medida que o tempo passa.

Exemplo 10.14. Um *processo de nascimento e morte* é uma cadeia de Markov para a qual ocorrem transições apenas entre estados adjacentes, como mostrado na Figura 10.4. O sistema de filas discutido no exemplo anterior é um exemplo de um processo de nascimento e morte. Repita o exercício anterior para um processo de nascimento e morte geral.

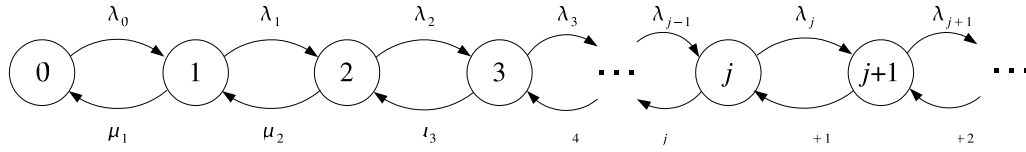


Figura 10.4: Diagrama de taxa de transição para um processo de nascimento e morte geral.

Solução. As Equações de balanço global para um processo de nascimento e morte geral são

$$\lambda_0 p_0 = \mu_1 p_1, \quad j = 0 \quad (10.42a)$$

$$\lambda_j p_j - \mu_{j+1} p_{j+1} = \lambda_{j-1} p_{j-1} - \mu_j p_j, \quad j = 1, 2, \dots \quad (10.42b)$$

Como no exemplo anterior, segue que

$$p_j = r_j p_{j-1}, \quad j = 1, 2, \dots$$

e

$$p_j = r_j r_{j-1} \cdots r_1 p_0, \quad j = 1, 2, \dots$$

onde $r_j = (\lambda_{j-1})/\mu_j$. Se definirmos

$$R_j = r_j r_{j-1} \cdots r_1 \text{ e } R_0 = 1,$$

então encontramos p_0 através de

$$1 = \left(\sum_{j=0}^{\infty} R_j \right) p_0.$$

Se a série da Equação acima converge, então a fmp estacionária é dada por

$$p_j = \frac{R_j}{\sum_{i=0}^{\infty} R_i} \quad (10.43)$$

Se a série não converge, então uma fmp estacionária não existe, e $p_j = 0$ para todo j .

10.5 Classes de estados, propriedades de recorrência e probabilidades limite

Nesta seção iremos olhar mais de perto a relação entre o comportamento de uma cadeia de Markov e sua matriz de probabilidade de transições de estados. Primeiramente iremos ver que os estados de uma cadeia de Markov de tempo discreto podem ser divididos em uma ou mais classes separadas e que estas podem ser de diferentes tipos. Iremos então mostrar que o comportamento de longo prazo de uma cadeia de Markov está relacionada aos tipos de suas classes de estados. Finalmente, usaremos estes resultados para relacionar o comportamento de longo prazo de cadeias de Markov de tempo contínuo com o de sua cadeia de Markov embutida.

10.5.1 Classes de estados

Definição 10.5. Acessibilidade. Dizemos que o estado j é acessível a partir do estado i se para algum $n \geq 0$, $p_{ij}(n) > 0$, isto é, se existe uma sequência de transições de i para j com probabilidade não nula.

Definição 10.6. Comunicabilidade. Os estados i e j se comunicam se i é acessível a partir de j e j é acessível a partir de i . Representamos este fato com a seguinte notação: $i \leftrightarrow j$.

Note que um estado se comunica consigo mesmo desde que $p_{ii}(0) = 1$.

Se o estado i se comunica com o estado j , e o estado j se comunica com o estado k , isto é, se $i \leftrightarrow j$ e $j \leftrightarrow k$, então o estado i se comunica com o estado k . Para verificar isto, note que $i \leftrightarrow j$ implica que existe um caminho de probabilidade não nula de i para j , e $j \leftrightarrow k$ implica que existe um caminho subsequente de probabilidade não nula de j

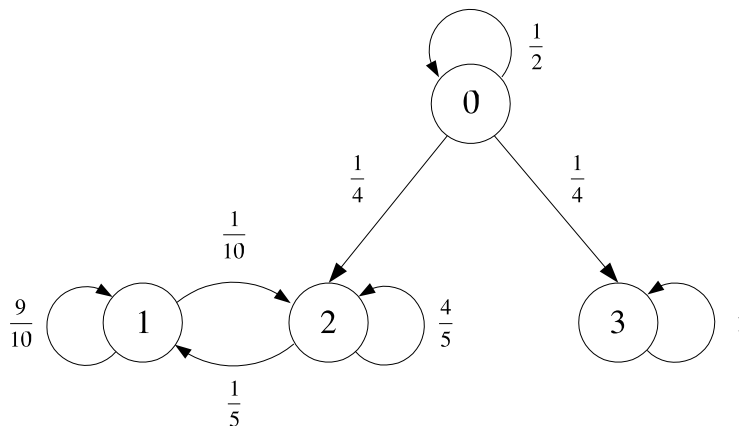
para k . Os caminhos combinados formam um caminho de probabilidade não nula de i para k . Existe um caminho de probabilidade não nula na direção reversa pelas mesmas razões.

Definição 10.7. Classes de estados: dizemos que dois estados pertencem a uma mesma classe se estes se comunicam entre si.

Note que duas classes de estados diferentes precisam ser disjuntas desde que se tiverem um estado em comum, isto implicaria que os estados de ambas as classes se comunicariam entre si. Então os estados de uma cadeia de Markov consistem de uma ou mais classes de comunicação disjuntas.

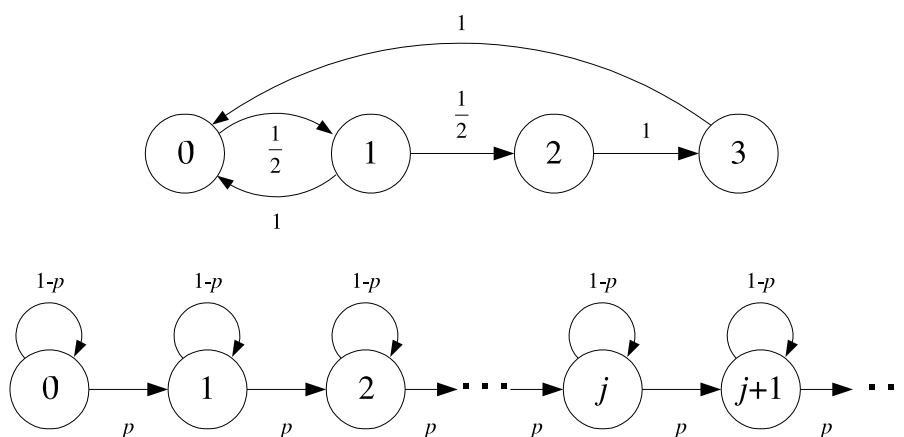
Definição 10.8. Cadeia Irredutível: Uma cadeia de Markov que consiste de uma única classe é dita irredutível.

Exemplo 10.15. A figura abaixo mostra o diagrama de transição de estados para uma cadeia de Markov com três classes: $\{0\}$, $\{1, 2\}$ e $\{3\}$

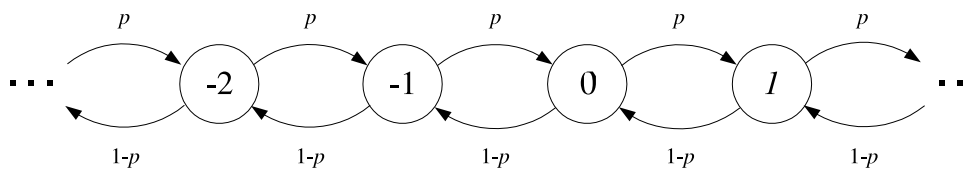


Exemplo 10.16. Abaixo tem-se o diagrama de transição de estados para uma cadeia de Markov periódica com apenas uma classe $\{0, 1, 2, 3\}$. Então esta cadeia é irredutível.

Exemplo 10.17. Neste exemplo, temos o diagrama de transição de estados para um processo de contagem binomial. Pode-se ver que as classes são: $\{0\}$, $\{1\}$, $\{2\}$, \dots



Exemplo 10.18. A figura abaixo mostra o diagrama de transição de estados para o processo de caminhada aleatória. Se $p > 0$, então o processo tem apenas uma classe $\{0, \pm 1, \pm 2, \dots\}$, de modo que é irredutível.



10.5.2 Propriedades de recorrência

Definição 10.9. Estado recorrente: O estado i é chamado **recorrente** se o processo retorna a ele com probabilidade um, isto é,

$$f_i = P[\text{alguma vez retornar ao estado } i] = 1$$

Definição 10.10. Estado transiente: O estado i é chamado **transiente** se

$$f_i < 1$$

Se iniciamos uma cadeia de Markov em um estado recorrente i , então o estado ocorre novamente um número infinito de vezes. Se iniciamos uma cadeia de Markov em um estado transiente, o estado não ocorre novamente depois de algum número finito de retornos. Cada nova ocorrência do estado pode ser vista como uma falha em uma tentativa de Bernoulli. A probabilidade de falha é f_i . Então o número de retornos ao estado i terminando com um sucesso (não retorno) é uma variável aleatória geométrica com média $(1 - f_i)^{-1}$. Se $f_i < 1$, então a probabilidade de um número infinito de

sucessos é zero. Portanto um estado transiente ocorre novamente um número finito de vezes.

Seja X_n uma cadeia de Markov com estado inicial i , $X_0 = i$. Seja $I_i(x)$ uma função indicadora para o estado i , isto é, $I_i(x) = 1$ se $X = i$, e $I_i(x) = 0$ caso contrário. O número esperado de retornos para o estado i é então

$$E \left[\sum_{n=1}^{\infty} I_i(X_n) | X_0 = i \right] = \sum_{n=1}^{\infty} E[I_i(X_n) | X_0 = i] = \sum_{n=1}^{\infty} p_{ii}(n) \quad (10.44)$$

desde que

$$E[I_i(X_n) | X_0 = i] = P[X_n = i | X_0 = i] = p_{ii}(n)$$

Um estado é recorrente se e somente se ele ocorre novamente um número infinito de vezes, então da Equação (10.44), o estado i é recorrente se e somente se

$$\sum_{n=1}^{\infty} p_{ii}(n) = \infty \quad (10.45)$$

Similarmente, o estado i é transiente se e somente se

$$\sum_{n=1}^{\infty} p_{ii}(n) < \infty \quad (10.46)$$

Exemplo 10.19. Dado o diagrama de transição de estados do Exemplo 10.15 verifique que o estado 0 é transiente, e o estado 1 é recorrente.

Solução. O estado 0 é transiente desde que $p_{00}(n) = (1/2)^n$, de modo que

$$\sum_{n=1}^{\infty} p_{00}(n) = \frac{1}{2} + \left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^3 + \cdots = 1 < \infty$$

Por outro lado, se o processo se iniciar no estado 1, teríamos o processo de dois estado discutidos no Exemplo 10.4. Para este processo mostramos que

$$p_{11}(n) = \frac{\beta + \alpha(1 - \alpha - \beta)^n}{\alpha + \beta} = \frac{1/2 + 1/4(7/10)^n}{3/4}$$

de modo que

$$\sum_{n=1}^{\infty} p_{11}(n) = \sum_{n=1}^{\infty} \left(\frac{2}{3} + \frac{(7/10)^n}{3} \right) = \infty$$

Portanto o estado 1 é recorrente.

Exemplo 10.20. Mostre que para um processo binomial de contagem todos os estados são transientes.

Solução. Para este processo, $p_{ii}(n) = (1 - p)^n$, de modo que para $p > 0$,

$$\sum_{n=1}^{\infty} p_{ii}(n) = \sum_{n=1}^{\infty} (1 - p)^n = \frac{1 - p}{p} < \infty$$

Exemplo 10.21. Para o processo de caminhada aleatória, verifique se o estado 0 é transiente ou recorrente.

Solução. O estado 0 ocorre novamente a cada $2n$ passos se e somente se os estados $n + 1$ e $n - 1$ ocorrem durante os $2n$ passos. Isto ocorre com probabilidade

$$p_{00}(2n) = \binom{2n}{n} p^n (1 - p)^n$$

A fórmula de Stirling para $n!$ pode ser utilizada para mostrar que

$$\binom{2n}{n} p^n (1 - p)^n \sim \frac{(4p(1 - p))^n}{\sqrt{\pi n}}$$

onde $a_n \sim b_n$ quando $\lim_{n \rightarrow \infty} \frac{a_n}{b_n} = 1$.

Então a Equação (10.44) para o estado 0 é

$$\sum_{n=1}^{\infty} p_{00}(2n) \sim \sum_{n=1}^{\infty} \frac{(4p(1 - p))^n}{\sqrt{\pi n}}$$

Se $p = 1/2$, então $4p(1 - p) = 1$ e a série diverge. Segue então que o estado 0 é recorrente. Se $p \neq 1/2$, então $4p(1 - p) < 1$, e a série acima converge. Isto implica que o estado 0 é transiente. Então quando $p = 1/2$, o processo de caminhada aleatória mantém um balanço precário em torno do estado 0. Logo que $p \neq 1/2$, uma perturbação positiva ou negativa é introduzida e o processo cresce ao redor de $\pm\infty$.

Se o estado i é recorrente então todos os estados de sua classe irão eventualmente ser visitados à medida que o processo retorna repetidamente a i . De fato, todos os outros estados em sua classe são visitados um número infinito de vezes. Então recorrência é uma propriedade de classe, isto é, se o estado i é recorrente e $i \leftrightarrow j$, então o estado j também é recorrente. Similarmente, a transitoriedade também é uma propriedade de classe.

Se uma cadeia de Markov é irredutível, isto é, se consiste de uma única classe de comunicação, então todos os seus estados são ou transientes ou recorrentes. Se o número de estados na cadeia é finito, é impossível para todos os estados serem transientes. Então, os estados de uma cadeia de Markov irredutível com número de estados finito são todos recorrentes.

A informação sobre quando o estado i pode ocorrer novamente está contido em $p_{ii}(n)$, a probabilidade de transição de n passos do estado i para ele mesmo.

Definição 10.11. Período de um estado. Dizemos que o estado i tem **período** d se ele puder ocorrer nos instantes que são múltiplos de d , isto é, $p_{ii}(n) = 0$ quando n não é múltiplo de d , onde d é o maior inteiro com esta propriedade. Pode-se mostrar que todos os estados de uma classe têm o mesmo período.

Definição 10.12. Cadeia de Markov aperiódica. Uma cadeia de Markov irreduzível é dita **aperiódica** se os estados em sua classe única têm período unitário.

Exemplo 10.22. Verifique qual é o período da cadeia de Markov do Exemplo 10.15.

Solução. Para esta cadeia, $p_{ii}(n) > 0$ para todos os estados, $n = 1, 2, \dots$. Portanto todas as três classes na cadeia têm período unitário.

Exemplo 10.23. Para a cadeia de Markov do Exemplo 10.16, verifique o valor de seu período.

Solução. Para esta cadeia, os estados 0 e 1 podem ocorrer novamente nos instantes 2, 4, 6, \dots e os estados 2 e 3 nos instantes 4, 6, 8, \dots . Portanto esta cadeia tem período 2.

Exemplo 10.24. Verifique o período do processo de caminhada aleatória do Exemplo 10.18.

Solução. Para este processo, um estado ocorre novamente quando o número de sucessos (+1s) é igual ao número de falhas (-1s). Isto acontece somente depois de um número par de eventos, e portanto este processo tem período 2.

10.5.3 Probabilidades limite

Se todos os estados em uma cadeia de Markov são transientes, então as probabilidades de todos os estados tendem a zero à medida que $n \rightarrow \infty$. Se uma cadeia de Markov tem algumas classes transientes e outras classes recorrentes, como a cadeia do Exemplo 10.15, então eventualmente o processo irá entrar e permanecer em uma das classes recorrentes. Assim, podemos nos concentrar nas classes recorrentes para o estudo das probabilidades limite de uma cadeia. Por esta razão iremos assumir nesta seção que estamos lidando com uma cadeia de Markov irreduzível.

Suponha que iniciemos uma cadeia de Markov em um estado recorrente i no instante $n = 0$. Sejam $T_i(1), T_i(1) + T_i(2), \dots$ os instantes aonde o processo retorna ao estado i , onde $T_i(k)$ é o tempo decorrido entre o $(k - 1)$ -ésimo e o k -ésimo retornos (veja

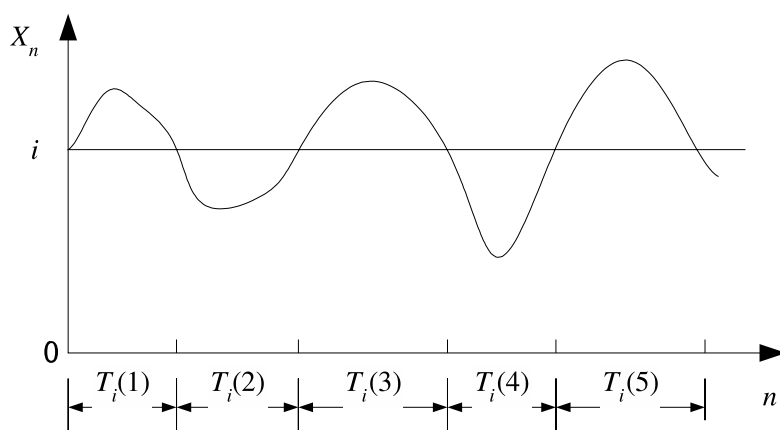


Figura 10.5: Instantes de recorrência para o estado i .

Figura 10.5). Os T_i formam uma sequência iid desde que cada instante de retorno é independente dos instantes de retorno anteriores.

A proporção de tempo gasto no estado i depois de k retornos a i é

$$\text{proporção de tempo no estado } i = \frac{k}{T_i(1) + T_i(2) + \cdots + T_i(k)} \quad (10.47)$$

Desde que o estado é recorrente, o processo retorna ao estado i um número infinito de vezes. Então a Lei dos Grandes Números implica que, com probabilidade um, o recíproco da expressão acima aproxima-se do **tempo médio de recorrência** $E[T_i]$ de modo que a proporção de longo prazo do tempo gasto no estado i aproxima

$$\text{proporção de tempo no estado } i \rightarrow \frac{1}{E[T_i]} = \pi_i \quad (10.48)$$

onde π_i é a proporção de longo prazo de tempo gasto no estado i ,

Se $E[T_i] < \infty$, então dizemos que o estado i é **recorrente positivo**. A Equação (10.48) implica então que

$$\pi_i > 0, \text{ se o estado } i \text{ é recorrente positivo}$$

Se $E[T_i] = \infty$, então dizemos que o estado i é **recorrente nulo**. A Equação (10.48) implica então que

$$\pi_i = 0, \text{ se o estado } i \text{ é recorrente nulo}$$

Pode-se mostrar que recorrência positiva e nula são propriedades de classe.

Estados recorrentes, aperiódicos e recorrentes nulos são chamados de **ergódicos**. Uma cadeia de Markov ergódica é definida como uma cadeia irredutível, aperiódica e recorrente positiva.

Exemplo 10.25. Para o processo do Exemplo 10.16, calcule $E[T_0]$ e π_0 .

Solução. Este processo retorna ao estado 0 em dois passos com probabilidade 1/2 e em quatro passos com probabilidade 1/2. Portanto o tempo de recorrência média para o estado 0 é

$$E[T_0] = \frac{1}{2}(2) + \frac{1}{2}(4) = 3$$

Portanto o estado 0 é recorrente positivo e a proporção de longo prazo de tempo em que o sistema permanece no estado 0 é

$$\pi_0 = \frac{1}{3}$$

Exemplo 10.26. No Exemplo 10.21 foi mostrado que o processo de caminhada aleatória é recorrente se $p = 1/2$. Entretanto, pode-se mostrar que o tempo médio de recorrência é infinito quando $p = 1/2$ ([Fel68], p.314). Então todos os estados da cadeia são recorrentes nulos.

Os π_j 's na Equação (10.48) satisfazem a equação que define a fmp de estado estacionário

$$\pi_j = \sum_i \pi_i P_{ij}, \quad \forall j \quad (10.49a)$$

$$\sum_i \pi_i = 1 \quad (10.49b)$$

Para ver isto, note que desde que π_i é a proporção de tempo gasto no estado i , então $\pi_i P_{ij}$ é a proporção de tempo na qual o estado j segue o estado i . Se somarmos sobre todos os estados i , obteremos a proporção de longo prazo do tempo no estado j , π_j .

Exemplo 10.27. Encontre a fmp de estado estacionário para a cadeia de Markov do Exemplo 10.16.

Solução. Temos das Equações (10.49a) e (10.49b) que

$$\pi_0 = \frac{1}{2}\pi_1 + \pi_3, \quad \pi_1 = \pi_0, \quad \pi_2 = \frac{1}{2}\pi_1, \quad \pi_3 = \pi_2$$

Estas equações implicam que $\pi_1 = \pi_0$ e $\pi_2 = \pi_3 = \pi_0/2$. Usando o fato de que a soma das probabilidades deve ser um, obtemos

$$\pi_1 = \pi_0 = \frac{1}{3} \quad \text{e} \quad \pi_2 = \pi_3 = \frac{1}{3}$$

Note que $\pi_0 = 1/3$ foi obtida do tempo de recorrência médio, calculado no Exemplo 10.26.

Na Seção 10.2 vimos que para cadeias de Markov que exigem um comportamento estacionário, a matriz de transição de n passos aproxima-se de uma matriz fixa de linhas iguais à medida que $n \rightarrow \infty$ (veja Equação 10.20). Vimos também que as linhas desta matriz limite consistiam de uma fmp que satisfaz (10.49a) e (10.49b). Iremos agora definir sob quais condições isto ocorre.

Teorema 10.2. *Para uma cadeia de Markov irredutível, aperiódica e recorrente positiva,*

$$\lim_{n \rightarrow \infty} p_{ij}(n) = \pi_j, \quad \forall j$$

onde π_j é a única solução não negativa das Equações (10.49a) e (10.49b).

Uma prova deste teorema pode ser encontrada em [Ros83]. O Teorema 10.5.3 afirma que para cadeias de Markov irredutíveis, aperiódicas e recorrente positivas, as probabilidades dos estados aproximam-se de valores de estado de regime permanente que são independentes da condição inicial. Estas probabilidades de regime permanente correspondem às probabilidades estacionárias obtidas nas Equações (10.49a) e (10.49b) e portanto correspondem à proporção de longo prazo do tempo gasto no estado dado. Esta é a razão pela qual cadeias de Markov irredutíveis, aperiódicas e recorrente positivas são chamadas de ergódicas.

Para processos periódicos, temos o seguinte resultado:

Teorema 10.3. *Para uma cadeia de Markov irredutível, periódica e recorrente positiva com período d ,*

$$\lim_{n \rightarrow \infty} p_{jj}(nd) = d\pi_j \quad \forall j$$

onde π_j é a única solução não negativa das Equações (10.49a) e (10.49b).

Como antes, π_j a proporção de tempo gasto no estado j . Entretanto, o fato de o estado j ocorrer apenas em múltiplos de d passos implica que a probabilidade de ocorrência do estado j é d vezes maior nos instantes permitidos e zero para os demais.

Exemplo 10.28. *Calcule as probabilidades de longo prazo para os estados 0 e 2 para a cadeia de Markov do Exemplo 10.16*

Solução. Nos Exemplos 10.25 e 10.27 vimos que proporção de longo prazo de tempo gasto no estado 0 é $\pi_0 = 1/3$. Se começamos no estado 0, então só podem ocorrer estados pares nos instantes de tempo pares. Então nestes instantes de tempo pares a probabilidade do estado 0 é $2/3$ e a probabilidade do estado 2 é $1/3$. Em instantes de tempo ímpares, as probabilidades dos estados 0 e 2 são zero.

10.5.4 Probabilidades limite para as cadeias de Markov de tempo contínuo

Vimos na Seção 10.3 que uma cadeia de Markov de tempo contínuo $X(t)$ pode ser vista como sendo constituída de uma sequência de estados determinada por alguma cadeia de Markov discreta X_n com probabilidades de transição \tilde{q}_{ij} e uma sequência de tempos de

ocupação de estados correspondente exponencialmente distribuída. Nesta seção, iremos mostrar que se a cadeia discreta associada é irredutível e recorrente positiva com fmp estacionária π_j , então a proporção de tempo de longo prazo gasta por $X(t)$ no estado i é

$$p_i = \frac{\pi_i/v_i}{\sum_j \pi_j/v_j}$$

onde $1/v_i$ é o tempo médio de ocupação no estado i . Além disso, mostramos que os p_i são as soluções únicas das equações de balanço global (10.38b) e (10.38c).

Suponha que a cadeia de Markov embutida X_n é irredutível e recorrente positiva, de modo que a Equação (10.48) seja válida. Seja $N_i(n)$ o número de vezes que o estado i ocorre nas primeiras n transições, e seja $T_i(j)$ o tempo de ocupação da j -ésima vez que o estado i ocorre. A proporção de tempo gasto no estado i depois das primeiras n transições é

$$\begin{aligned} \frac{\text{tempo gasto no estado } i}{\text{tempo gasto em todos os estados}} &= \frac{\sum_{j=1}^{N_i(n)} T_i(j)}{\sum_i \sum_{j=1}^{N_i(n)} T_i(j)} \\ &= \frac{\frac{N_i(n)}{n} \frac{1}{N_i(n)} \sum_{j=1}^{N_i(n)} T_i(j)}{\sum_i \frac{N_i(n)}{n} \frac{1}{N_i(n)} \sum_{j=1}^{N_i(n)} T_i(j)} \end{aligned} \quad (10.50)$$

À medida que $n \rightarrow \infty$, pelas Equações (10.48), (10.49a) e (10.49b), com probabilidade um,

$$\frac{N_i(n)}{n} \rightarrow \pi_i \quad (10.51)$$

a fmp estacionária da cadeia de Markov embutida. Adicionalmente, temos que $N_i(n) \rightarrow \infty$ à medida que $n \rightarrow \infty$, de modo que pela lei forte dos números grandes, com probabilidade um,

$$\frac{1}{N_i(n)} \sum_{j=1}^{N_i(n)} T_i(j) \rightarrow E[T_i] = \frac{1}{v_i} \quad (10.52)$$

onde usamos o fato de que o tempo de ocupação de estado no estado i tem média $1/v_i$. As Equações (10.51) e (10.52) quando aplicadas a (10.50) implicam que, com probabilidade um, a proporção de longo prazo do tempo gasto no estado i aproxima

$$p_i = \frac{\pi_i/v_i}{\sum_j \pi_j/v_j} = c\pi_i/v_i \quad (10.53)$$

onde π_j é a fmp solução única para

$$\pi_j = \sum_i \pi_i \tilde{q}_{ij}, \quad \forall j \quad (10.54)$$

e c é uma constante de normalização.

Obtemos a equação de balanço global (10.38b), substituindo $\pi_i = v_i p_i / c$ da Equação (10.53) e $\tilde{q}_{ij} = \gamma_{ij} / v_i$ na Equação (10.54):

$$v_j p_j = \sum_{i \neq j} p_i \gamma_{ij}, \quad \forall j$$

Então os p_i 's são a solução únicas das equações de balanço global.

Exemplo 10.29. Encontre as probabilidades de longo prazo para os estados da cadeia de Markov do Exemplo 10.10.

Solução. Para este sistema

$$[\tilde{q}_{ij}] = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

A equação $\boldsymbol{\pi} = \boldsymbol{\pi}[\tilde{q}_{ij}]$ implica que

$$\pi_0 = \pi_1 = \frac{1}{2}$$

Adicionalmente, $v_0 = \alpha$ e $v_1 = \beta$. Então

$$p_0 = \frac{1/2(1/\alpha)}{1/2(1/\alpha + 1/\beta)} = \frac{\beta}{\alpha + \beta} \quad \text{e} \quad p_1 = \frac{\alpha}{\alpha + \beta}$$

10.6 Exercícios

1. Seja T_n o tempo de chegada do n -ésimo cliente a uma estação de serviço. Seja Z_n o intervalo de tempo entre as chegadas do cliente n e do cliente $n - 1$, isto é

$$Z_n = T_n - T_{n-1}, \quad n \geq 1$$

e $T_0 = 0$. Seja $\{X(t), t \geq 0\}$ o processo de contagem associado com $\{T_n, n \geq 0\}$. Mostre que se $X(t)$ tem incrementos estacionários, então $Z_n, n = 1, 2, \dots$ são v.a.'s identicamente distribuídas.

2. Desenhe os diagramas de transição de estados e classifique os estados das Cadeias de Markov para as seguintes matrizes de transição.

$$P = \begin{bmatrix} 0 & 0.5 & 0.5 \\ 0.5 & 0 & 0.5 \\ 0.5 & 0.5 & 0 \end{bmatrix} \quad P = \begin{bmatrix} 0 & 0 & 0.5 & 0.5 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \quad P = \begin{bmatrix} 0.3 & 0.4 & 0 & 0 & 0.3 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.6 & 0.4 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

3. Considere uma cadeia de Markov com espaço de estados $\{0, 1\}$ e matriz de probabilidades de transição

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

Mostre que o estado 0 é recorrente e que o estado 1 é transiente.

4. Considere uma cadeia de Markov de dois estados com matriz de probabilidade de transição

$$P = \begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix}, \quad 0 < a < 1, \quad 0 < b < 1$$

- (a) Encontre P^n .
 (b) Encontre P^n para $n \rightarrow \infty$.

Resp:

$$(a) \quad P^n = \frac{1}{a+b} \left\{ \begin{bmatrix} b & a \\ b & a \end{bmatrix} + (1-a-b)^n \begin{bmatrix} a & -a \\ -b & b \end{bmatrix} \right\}$$

$$(b) \quad \lim_{n \rightarrow \infty} P^n = \frac{1}{a+b} \begin{bmatrix} b & a \\ b & a \end{bmatrix}$$

5. Um modelo de Markov para transmissão de voz por pacotes assume que se o n -ésimo pacote contém silêncio, a probabilidade de silêncio no próximo pacote é $(1 - \alpha)$ e a probabilidade do pacote conter voz é α .

Similarmente, se o n -ésimo pacote contiver atividades de voz, a probabilidade do próximo pacote conter voz é $(1 - \beta)$, e a probabilidade de silêncio é β .

- (a) Esboce uma cadeia de Markov para este problema.
 (b) Para $\alpha = 1/10$ e $\beta = 1/5$, determine a matriz de transição de estados de um passo.
 (c) Dadas as probabilidades iniciais dos estados $p_0 = p_1 = 0,5$, determine as probabilidades dos estados depois de 2 passos.

Resp:

$$(a)$$

$$(b) \quad P = \begin{bmatrix} 0,9 & 0,1 \\ 0,2 & 0,8 \end{bmatrix}$$

$$(c) \quad \mathbf{p}(2) = [0,585 \quad 0,415]$$

6. Considere uma cadeia de Markov com dois estados e matriz de probabilidade de transição de estados dada por

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

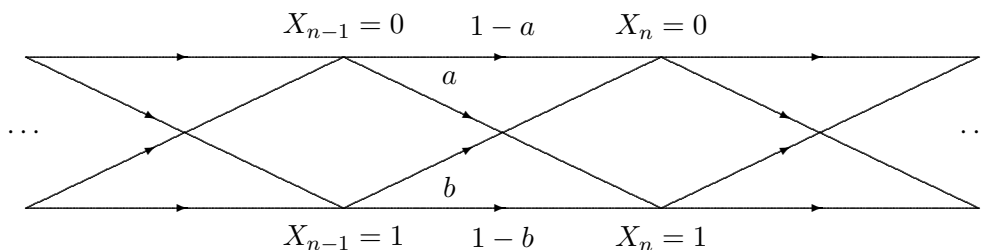
- (a) Encontre a distribuição estacionária de estados $\hat{\mathbf{p}}$ para esta cadeia.
- (b) Encontre $\lim_{n \rightarrow \infty} P^n$.

Resp:

(a) $\hat{\mathbf{p}} = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} \end{bmatrix}$

(b) $P^n = \begin{bmatrix} 2/3 & 1/3 \\ 2/3 & 1/3 \end{bmatrix}$

7. Um exemplo de uma cadeia de Markov de dois estados é um sistema consistindo de sequências de estágios em cascata de canais de comunicação binários, como mostrado na figura abaixo.



Aqui, X_n denota o dígito que deixa o n -ésimo estágio do canal, e X_0 denota o dígito que entra no primeiro estágio. A matriz de probabilidades de transição deste sistema de comunicação é geralmente chamado de *matriz de canal*, e é dada por

$$P = \begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix}, \quad 0 < a, b < 1$$

Assuma que $a = 0,1$ e $b = 0,2$, e que a distribuição inicial é $P[X_0 = 0] = P[X_0 = 1] = 0,5$.

- (a) Encontre a distribuição de X_n .
- (b) Encontre a distribuição de X_n quando $n \rightarrow \infty$.

Dica:

$$P^n = \frac{1}{a+b} \left\{ \begin{bmatrix} b & a \\ b & a \end{bmatrix} + (1-a-b)^n \begin{bmatrix} a & -a \\ -b & b \end{bmatrix} \right\}$$

Resp:

$$(a) \begin{bmatrix} \frac{2}{3} - \frac{(0,7)^n}{6} & \frac{1}{3} - \frac{(0,7)^n}{6} \end{bmatrix}$$

$$(b) \begin{bmatrix} \frac{2}{3} & \frac{1}{3} \end{bmatrix}$$

8. Considere uma cadeia de Markov com dois estados e matriz de transição dada por

$$P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

(a) Encontre a distribuição de estado estacionário π .

(b) Mostre que $\lim_{n \rightarrow \infty} P^n$ não existe.

Dica: Calcule P^2, P^3, P^4, \dots e faça a prova por indução.

Resp: (a) $[1/2 \quad 1/2]$

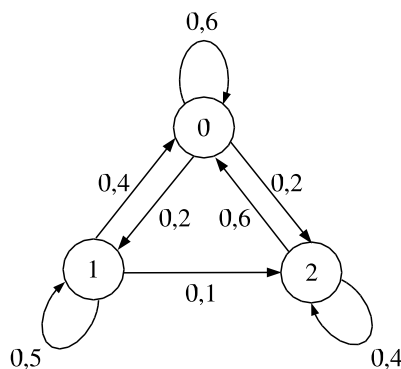
9. Um elétron pode estar em uma de três possíveis órbitas. A transição da órbita i para a órbita j ($i, j = 1, 2, 3$) ocorre em uma unidade de tempo com probabilidade

$$C_i e^{-\lambda|i-j|}, \lambda > 0$$

Esboce uma cadeia de Markov para este problema e calcule as constantes C_i .

$$\text{Resp: } C_1 = \frac{1}{1 + e^{-\lambda} + e^{-2\lambda}} \quad C_2 = \frac{1}{1 + 2e^{-\lambda}} \quad C_3 = \frac{1}{1 + e^{-\lambda} + e^{-2\lambda}}$$

10. Dada a cadeia de Markov abaixo, calcule as probabilidades dos estados em regime permanente (se existirem).



$$\text{Resp: } \pi = \begin{bmatrix} \frac{5}{9} & \frac{2}{9} & \frac{2}{9} \end{bmatrix}$$

11. Uma cadeia de Markov com probabilidades de transição p_{ij} possui um estado particular k para o qual $p_{ik} = q$ para todos os estados i . Mostre que $p_k(n) = q, \forall n$.

12. Uma urna contém inicialmente 5 bolas brancas e 5 bolas pretas. O seguinte experimento é repetido indefinidamente: uma bola é retirada da urna; se a mesma é branca ela é recolocada na urna, caso contrário é deixada de fora. Seja X_n o número de bolas pretas que permanecem na urna depois de n testes.
- (a) X_n é um processo de Markov? Se sim, esboce uma cadeia para este processo.
 - (b) As probabilidades de transição dependem de n ?
 - (c) Calcule $P(n)$, $n \rightarrow \infty$, e encontre uma explicação para o resultado obtido.

Resp:

- (a) Sim. Para $k = 1, 2, \dots, 5$: $P[k-1|k] = k/(5+k) = 1 - P[k|k]$, $P[0|0] = 1$.
 - (b) Não.
13. Seja $X(n)$ um processo de caminhada aleatória unidimensional.
- (a) Mostre que $X(n)$ é um processo de Markov.
 - (b) Encontre a matriz de transição de um passo para este processo.

$$\text{Resp: } P = \begin{cases} p, & j = i + 1 \\ q, & j = i - 1 \\ 0, & \text{caso contrário} \end{cases}$$

Apêndice A

Tabelas Matemáticas

A.1 Identidades trigonométricas

$$\operatorname{sen}^2(\alpha) + \cos^2(\alpha) = 1$$

$$\operatorname{sen}(\alpha + \beta) = \operatorname{sen} \alpha \cos \beta + \cos \alpha \operatorname{sen} \beta$$

$$\operatorname{sen}(\alpha - \beta) = \operatorname{sen} \alpha \cos \beta - \cos \alpha \operatorname{sen} \beta$$

$$\cos(\alpha + \beta) = \cos \alpha \cos \beta - \operatorname{sen} \alpha \operatorname{sen} \beta$$

$$\cos(\alpha - \beta) = \cos \alpha \cos \beta + \operatorname{sen} \alpha \operatorname{sen} \beta$$

$$\operatorname{sen} 2\alpha = 2 \operatorname{sen} \alpha \cos \alpha$$

$$\cos 2\alpha = \cos^2 \alpha - \operatorname{sen}^2 \alpha = 2 \cos^2 \alpha - 1 = 1 - 2 \operatorname{sen}^2 \alpha$$

$$\operatorname{sen} \alpha \operatorname{sen} \beta = \frac{1}{2} [\cos(\alpha - \beta) - \cos(\alpha + \beta)]$$

$$\cos \alpha \cos \beta = \frac{1}{2} [\cos(\alpha - \beta) + \cos(\alpha + \beta)]$$

$$\operatorname{sen} \alpha \cos \beta = \frac{1}{2} [\operatorname{sen}(\alpha + \beta) + \operatorname{sen}(\alpha - \beta)]$$

$$\cos \alpha \operatorname{sen} \beta = \frac{1}{2} [\operatorname{sen}(\alpha + \beta) - \operatorname{sen}(\alpha - \beta)]$$

$$\operatorname{sen}^2 \alpha = \frac{1}{2} (1 - \cos 2\alpha)$$

$$\cos^2 \alpha = \frac{1}{2} (1 + \cos 2\alpha)$$

$$e^{j\alpha} = \cos \alpha + j \operatorname{sen} \alpha$$

$$\cos \alpha = \frac{e^{j\alpha} + e^{-j\alpha}}{2}$$

$$\operatorname{sen} \alpha = \frac{e^{j\alpha} - e^{-j\alpha}}{2j}$$

$$\operatorname{sen} \alpha = \cos(\alpha - \pi/2)$$

A.2 Coeficientes Binomiais

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \binom{n}{n-k}$$

$$\binom{n}{k} = 0 \text{ para } n < k$$

$$\binom{n}{n} = \binom{n}{1} = \binom{n}{0} = 1$$

$$\binom{n}{k} = \binom{n}{p} \Leftrightarrow k = p \text{ ou } k + p = n \text{ (binomiais complementares)}$$

$$\binom{n}{k} + \binom{n}{k+1} = \binom{n+1}{k+1} \text{ (relação de Stiffel)}$$

A.3 Derivadas

Nas expressões a seguir, u , v e w são funções de x ; a , b e c são constantes.

$$\frac{d}{dx}(c) = 0$$

$$\frac{d}{dx}(cx) = c$$

$$\frac{d}{dx}(cx^n) = ncx^{n-1}$$

$$\frac{d}{dx}(u \pm v \pm w \pm \dots) = \frac{du}{dx} \pm \frac{dv}{dx} \pm \frac{dw}{dx} \pm \dots$$

$$\frac{d}{dx}(cu) = c \frac{du}{dx}$$

$$\frac{d}{dx}(uv) = u \frac{dv}{dx} + v \frac{du}{dx}$$

$$\frac{d}{dx}(uvw) = \frac{du}{dx}vw + u \frac{dv}{dx}w + uv \frac{dw}{dx}$$

$$\frac{d}{dx} \left(\frac{u}{v} \right) = \frac{v(du/dx) - u(dv/dx)}{v^2}$$

$$\frac{d}{dx}(u^n) = nu^{n-1} \frac{du}{dx}$$

$$\frac{dy}{dx} = \frac{dy}{du} \frac{du}{dx}$$

$$\frac{d}{dx} \text{sen}(u) = \cos(u) \frac{du}{dx}$$

$$\frac{d}{dx} \cos(u) = -\text{sen}(u) \frac{du}{dx}$$

$$\frac{d}{dx} \log_a(u) = \frac{\log_a(e)}{u} \frac{du}{dx}, \quad a > 0 \text{ e } a \neq 1$$

$$\frac{d}{dx} \ln(u) = \frac{d}{dx} \log_e(u) = \frac{1}{u} \frac{du}{dx}$$

$$\frac{d}{dx} a^u = a^u \ln(a) \frac{du}{dx}, \quad a > 0$$

$$\frac{d}{dx} e^u = e^u \frac{du}{dx}$$

$$\frac{d}{dx} u^v = \frac{d}{dx} e^{v \ln(u)} = e^{v \ln(u)} \frac{d}{dx} (v \ln(u)) = v u^{v-1} \frac{du}{dx} + u^v \ln(u) \frac{dv}{dx}$$

$$\frac{d}{dx} \operatorname{arctg}(u) = \frac{1}{1+u^2} \frac{du}{dx}$$

A.4 Integrais indefinidas

$$\int u dv = uv - \int v du, \text{ onde } u \text{ e } v \text{ são funções de } x.$$

$$\int x^n dx = \frac{x^{n+1}}{n+1}, \text{ exceto para } n = -1$$

$$\int x^{-1} dx = \ln x$$

$$\int e^{ax} dx = \frac{e^{ax}}{a}$$

$$\int \ln x dx = x \ln x - x$$

$$\int \frac{1}{a^2 + x^2} dx = \frac{1}{a} \tan^{-1} \frac{x}{a}$$

$$\int \frac{(\ln x)^n}{x} dx = \frac{1}{n+1} (\ln x)^{n+1}$$

$$\int x^n \ln(ax) dx = \frac{x^{n+1}}{n+1} \ln(ax) - \frac{x^{n+1}}{(n+1)^2}$$

$$\int x e^{ax} dx = \frac{e^{ax}(ax - 1)}{a^2}$$

$$\int x^2 e^{ax} dx = \frac{e^{ax}(a^2 x^2 - 2ax + 2)}{a^3}$$

$$\int \operatorname{sen}(ax) dx = -\frac{1}{a} \cos(ax)$$

$$\int \cos(ax) dx = \frac{1}{a} \operatorname{sen}(ax)$$

$$\int \operatorname{sen}^2(ax) dx = \frac{x}{2} - \frac{\operatorname{sen}(2ax)}{4a}$$

$$\int x \operatorname{sen}(ax) dx = \frac{1}{a^2} (\operatorname{sen}(ax) - ax \cos(ax))$$

$$\int x^2 \operatorname{sen}(ax) dx = \frac{2ax \operatorname{sen}(ax) + 2\cos(ax) - a^2 x^2 \operatorname{sen}(ax)}{a^3}$$

$$\int \cos^2(ax) dx = \frac{x}{2} + \frac{\operatorname{sen}(2ax)}{4a}$$

$$\int x \cos(ax) dx = \frac{1}{a^2} (\cos(ax) + ax \operatorname{sen}(ax))$$

$$\int x^2 \cos(ax) dx = \frac{1}{a^3} (2ax \cos(ax) - 2 \operatorname{sen}(ax) + a^2 x^2 \operatorname{sen}(ax))$$

A.5 Integrais definidas

$$\int_0^\infty t^{n-1} e^{-(a+1)t} dt = \frac{\Gamma(n)}{(a+1)^n} \quad n > 0, a > -1$$

$\Gamma(n) = (n-1)!$ se n é um inteiro positivo

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$

$$\Gamma\left(n + \frac{1}{2}\right) = \frac{1 \cdot 3 \cdot 5 \cdots (2n-1)}{2^n} \sqrt{\pi} \quad n = 1, 2, 3, \dots$$

$$\int_0^\infty e^{-\alpha^2 x^2} dx = \frac{\sqrt{\pi}}{2\alpha}$$

$$\int_0^\infty x e^{-\alpha^2 x^2} dx = \frac{1}{2\alpha^2}$$

$$\int_0^\infty x^2 e^{-\alpha^2 x^2} dx = \frac{\sqrt{\pi}}{4\alpha^3}$$

$$\int_0^\infty x^n e^{-\alpha^2 x^2} dx = \Gamma\left(\frac{n+1}{2}\right) / (2\alpha^{n+1})$$

$$\int_0^\infty \frac{a}{a^2 + x^2} dx = \frac{\pi}{2}, \quad a > 0$$

$$\int_0^\infty \frac{\operatorname{sen}^2(ax)}{x^2} dx = |a| \frac{\pi}{2}, \quad a > 0$$

$$\int_0^\infty \frac{\cos(mx)}{x^2 + a^2} dx = \frac{\pi}{2a} e^{-ma}$$

$$\int_{-\infty}^\infty e^{-(ax^2+bx+c)} dx = \sqrt{\frac{\pi}{a}} e^{\frac{b^2-4ac}{4a}}$$

Apêndice B

Tabelas de transformadas de Fourier

B.1 Definição

$$G(f) = \mathcal{F}\{g(t)\} = \int_{-\infty}^{\infty} g(t)e^{-j2\pi ft} dt$$

$$g(t) = \mathcal{F}^{-1}\{G(f)\} = \int_{-\infty}^{\infty} G(f)e^{j2\pi ft} df$$

B.2 Propriedades

Linearidade: $\mathcal{F}\{ag_1(t) + bg_2(t)\} = aG_1(f) + bG_2(f)$

Escalonamento no tempo: $\mathcal{F}\{g(at)\} = G(f/a)/|a|$

Dualidade: se $\mathcal{F}\{g(t)\} = G(f)$ então $\mathcal{F}\{G(t)\} = g(-f)$

Deslocamento no tempo: $\mathcal{F}\{g(t - t_0)\} = G(f)e^{-j2\pi ft_0}$

Deslocamento em frequência: $\mathcal{F}\{g(t)e^{j2\pi f_0 t}\} = G(f - f_0)$

Diferenciação: $\mathcal{F}\{g'(t)\} = j2\pi f G(f)$

Integração: $\mathcal{F}\left\{\int_{-\infty}^t g(s)ds\right\} = G(f)/(j2\pi f) + (G(0)/2)\delta(f)$

Multiplicação no tempo: $\mathcal{F}\{g_1(t)g_2(t)\} = G_1(f) * G_2(f)$

Convolução no tempo: $\mathcal{F}\{g_1(t) * g_2(t)\} = G_1(f)G_2(f)$

B.3 Pares de transformadas

$g(t)$	$G(f)$
$\begin{cases} 1, & -T/2 \leq t \leq T/2 \\ 0, & \text{caso contrário} \end{cases}$	$T \frac{\text{sen}(\pi f T)}{\pi f T}$
$2W \frac{\text{sen}(2\pi W t)}{2\pi W t}$	$\begin{cases} 1, & -W \leq f \leq W \\ 0, & \text{caso contrário} \end{cases}$
$\begin{cases} 1 - \frac{ t }{T}, & t < T \\ 0, & \text{caso contrário} \end{cases}$	$T \frac{\text{sen}^2(\pi f T)}{(\pi f T)^2}$
$e^{-at}u(t), \quad a > 0$	$\frac{1}{a + j2\pi f}$
$e^{-a t }, \quad a > 0$	$\frac{2a}{a^2 + (2\pi f)^2}$
$e^{-\pi t^2}$	$e^{-\pi f^2}$
$\delta(t)$	1
1	$\delta(f)$
$\delta(t - t_0)$	$e^{-j2\pi f t_0}$
$e^{j2\pi f_0 t}$	$\delta(f - f_0)$
$\cos(2\pi f_0 t)$	$\frac{1}{2} [\delta(f - f_0) + \delta(f + f_0)]$
$\text{sen}(2\pi f_0 t)$	$\frac{1}{2j} [\delta(f - f_0) - \delta(f + f_0)]$
$u(t)$	$\frac{1}{2} \delta(f) + \frac{1}{j2\pi f}$

Apêndice C

Séries de Taylor

C.1 Série de Taylor para funções de uma variável

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)(x-a)^2}{2!} + \cdots + \frac{f^{(n-1)}(a)(x-a)^{(n-1)}}{(n-1)!} + R_n$$

onde R_n , o resto após n termos, é dado por qualquer das formas seguintes:

Forma de Lagrange $R_n = \frac{f^{(n)}(\xi)(x-a)^n}{n!}$

Forma de Cauchy $R_n = \frac{f^{(n)}(\xi)(x-\xi)^{n-1}(x-a)}{(n-1)!}$

O valor ξ , que pode ser diferente nas duas formas, fica entre a e x . O resultado determina se $f(x)$ tem derivadas contínuas de ordem n pelo menos.

Se $\lim_{n \rightarrow \infty} R_n = 0$, a série é infinita, chamada de *Série de Taylor* para $f(x)$ em $x = a$. Se $x = 0$, a série é frequentemente chamada de *Série de Maclaurin*. Estas séries geralmente convergem para todos os valores de x em algum intervalo de convergência e divergem para todos os x fora deste intervalo.

C.2 Expansões mais utilizadas

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots, -\infty < x < \infty$$

$$a^x = e^{x \ln(a)} = 1 + x \ln(a) + \frac{(x \ln(a))^2}{2!} + \frac{(x \ln(a))^3}{3!} + \cdots, -\infty < x < \infty$$

$$\ln(x) = \left(\frac{x-1}{x}\right) + \frac{1}{2} \left(\frac{x-1}{x}\right)^2 + \frac{1}{3} \left(\frac{x-1}{x}\right)^3 + \cdots, x \geq \frac{1}{2}$$

$$\sin(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots, -\infty < x < \infty$$

$$\cos(x) = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots, -\infty < x < \infty$$

$$\sinh(x) = x + \frac{x^3}{3!} + \frac{x^5}{5!} + \frac{x^7}{7!} + \cdots, -\infty < x < \infty$$

$$\cosh(x) = 1 + \frac{x^2}{2!} + \frac{x^4}{4!} + \frac{x^6}{6!} + \cdots, -\infty < x < \infty$$

$$e^{\operatorname{sen}(x)} = 1 + x + \frac{x^2}{2} - \frac{x^4}{8!} - \frac{x^5}{15!} + \cdots, -\infty < x < \infty$$

$$e^{\cos(x)} = e \left(1 - \frac{x^2}{2} + \frac{x^4}{6} - \frac{31x^6}{720} + \cdots \right), -\infty < x < \infty$$

Apêndice D

Variáveis aleatórias discretas

D.1 Bernoulli

$$S_X = \{0, 1\}$$

$$p_0 = q = 1 - p \quad p_1 = p \quad 0 \leq p \leq 1$$

$$E[X] = p \quad \text{Var}[X] = p(1 - p)$$

$$G_X(z) = (q + pz)$$

Observações: a variável aleatória de Bernoulli é o valor da função indicadora I_A para algum evento A ; $X = 1$ se A ocorre, e 0 caso contrário.

D.2 Binomial

$$S_X = \{0, 1, \dots, n\}$$

$$p_k = \binom{n}{k} p^k (1 - p)^{n-k} \quad k = 0, 1, \dots, n$$

$$E[X] = np \quad \text{Var}[X] = np(1 - p)$$

$$G_X(z) = (q + pz)^n$$

Observações: X é o número de sucessos em n testes de Bernoulli, e portanto a soma de n variáveis aleatórias iid com distribuição de Bernoulli.

D.3 Geométrica

Primeira versão

$$S_X = \{0, 1, 2, \dots\}$$

$$p_k = p(1 - p)^k \quad k = 0, 1, \dots$$

$$E[X] = \frac{1 - p}{p} \quad \text{Var}[X] = \frac{1 - p}{p^2}$$

$$G_X(z) = \frac{p}{1 - qz}$$

Observações: X é o número de falhas antes do primeiro sucesso em uma sequência de

testes de Bernoulli independentes. A variável aleatória geométrica é a única variável aleatória discreta sem memória.

Segunda versão

$$S'_X = \{1, 2, \dots\}$$

$$p_k = p(1-p)^{k-1} \quad k = 1, 2, \dots$$

$$E[X'] = \frac{1}{p} \quad \text{Var}[X'] = \frac{1-p}{p^2}$$

$$G'_X(z) = \frac{pz}{1-qz}$$

Observações: $X' = X + 1$ é o número de tentativas antes do primeiro sucesso em uma sequência de testes de Bernoulli independentes.

D.4 Binomial negativa

$S_X = \{r, r+1, \dots\}$ onde r é um inteiro positivo

$$p_k = \binom{k-1}{r-1} p^r (1-p)^{k-r} \quad k = r, r+1, \dots, n$$

$$E[X] = \frac{r}{p} \quad \text{Var}[X] = \frac{r(1-p)}{p^2}$$

$$G_X(z) = \left(\frac{pz}{1-qz} \right)^r$$

Observações: X é o número de tentativas até o r -ésimo sucesso em uma sequência de testes de Bernoulli independentes.

D.5 Poisson

$$S_X = \{0, 1, 2, \dots\}$$

$$p_k = \frac{\alpha^k}{k!} e^{-\alpha}, \quad k = 0, 1, \dots \quad \alpha > 0$$

$$E[X] = \alpha \quad \text{Var}[X] = \alpha$$

$$G_X(z) = e^{\alpha(z-1)}$$

Observações: X é o número de eventos que ocorrem em uma unidade de tempo quando o tempo entre os eventos segue uma distribuição exponencial de média $1/\alpha$.

Apêndice E

Variáveis aleatórias contínuas

E.1 Uniforme

$$S_X = [a, b]$$

$$f_X(x) = \frac{1}{b-a} \quad a \leq x \leq b$$

$$E[X] = \frac{a+b}{2} \quad \text{Var}[X] = \frac{(b-a)^2}{12}$$

$$\psi_X(j\omega) = \frac{e^{j\omega b} - e^{j\omega a}}{j\omega(b-a)}$$

E.2 Exponencial

$$S_X = [0, \infty)$$

$$f_X(x) = \lambda e^{-\lambda x} \quad \lambda > 0$$

$$E[X] = \frac{1}{\lambda} \quad \text{Var}[X] = \frac{1}{\lambda^2}$$

$$\psi_X(j\omega) = \frac{\lambda}{\lambda - j\omega}$$

Observações: A variável aleatória exponencial é a única variável aleatória contínua sem memória. Em geral é usada para modelar o tempo entre eventos consecutivos em um processo de Poisson.

E.3 Gaussiana (Normal)

$$S_X = (-\infty, \infty)$$

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \sigma > 0$$

$$E[X] = \mu \quad \text{Var}[X] = \sigma^2$$

$$\psi_X(j\omega) = e^{j\mu\omega - \sigma^2\omega^2/2}$$

Observações: Sob uma grande gama de condições, X pode ser utilizada para aproximar a soma de um grande número de variáveis aleatórias independentes.

E.4 Gama

$$S_X = (0, \infty)$$

$$f_X(x) = \frac{\lambda(\lambda x)^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)} \quad \alpha > 0, \lambda > 0$$

$$E[X] = \frac{\alpha}{\lambda} \quad \text{Var}[X] = \frac{\alpha}{\lambda^2}$$

$$\psi_X(j\omega) = \frac{1}{(1 - j\omega/\lambda)^\alpha}$$

E.5 m-Erlang

$$S_X = (0, \infty)$$

$$f_X(x) = \frac{\lambda e^{-\lambda x} (\lambda x)^{m-1}}{(m-1)!} \quad \lambda > 0, \quad m \text{ inteiro positivo.}$$

$$E[X] = \frac{m}{\lambda} \quad \text{Var}[X] = \frac{m}{\lambda^2}$$

$$\psi_X(j\omega) = \left(\frac{\lambda}{\lambda - j\omega} \right)^m$$

Observações: Uma variável aleatória m-Erlang é obtida pela adição de m variáveis aleatórias iid com distribuição exponencial de parâmetro λ . Pode ser obtida a partir da distribuição gama, fazendo $\alpha = m$, onde m é um inteiro positivo.

E.6 Chi-Quadrado (χ^2)

$$S_X = (0, \infty)$$

$$f_X(x) = \frac{x^{(k-2)/2} e^{-x/2}}{2^{k/2} \Gamma(k/2)} \quad \text{onde } k \text{ é um inteiro positivo.}$$

$$E[X] = k \quad \text{Var}[X] = 2k$$

$$\psi_X(j\omega) = \left(\frac{1}{1 - j2\omega} \right)^{k/2}$$

Observações: A soma do quadrado de k variáveis aleatórias gaussianas de média zero e variância unitária corresponde a uma variável aleatória com distribuição χ^2 com k graus de liberdade.

E.7 Rayleigh

$$S_X = [0, \infty)$$

$$f_X(x) = \frac{x}{\alpha^2} e^{-x^2/(2\alpha^2)} \quad \alpha > 0.$$

$$E[X] = \alpha \sqrt{\frac{\pi}{2}} \quad \text{Var}[X] = \left(2 - \frac{\pi}{2}\right) \alpha^2$$

E.8 Cauchy

$$S_X = (-\infty, \infty)$$

$$f_X(x) = \frac{\alpha}{\pi(x^2 + \alpha^2)} \quad \alpha > 0$$

A média e a variância não existem.

$$\psi_X(j\omega) = e^{-\alpha|\omega|}$$

E.9 Laplace

$$S_X = (-\infty, \infty)$$

$$f_X(x) = \frac{\alpha}{2} e^{-\alpha|x|} \quad \alpha > 0.$$

$$E[X] = 0 \quad \text{Var}[X] = \frac{2}{\alpha^2}$$

$$\psi_X(j\omega) = \frac{\alpha^2}{\omega^2 + \alpha^2}$$

Apêndice F

Valores da distribuição normal

Nas tabelas a seguir são listados os valores da função distribuição cumulativa $\Phi(x)$ de uma variável aleatória com distribuição normal $N(0, 1)$.

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
-4.00	0.000031671	-3.50	0.000232629	-3.00	0.001349898	-2.50	0.006209665
-3.99	0.000033036	-3.49	0.000241510	-2.99	0.001394887	-2.49	0.006387154
-3.98	0.000034457	-3.48	0.000250706	-2.98	0.001441241	-2.48	0.006569119
-3.97	0.000035936	-3.47	0.000260229	-2.97	0.001488998	-2.47	0.006755652
-3.96	0.000037474	-3.46	0.000270087	-2.96	0.001538195	-2.46	0.006946850
-3.95	0.000039075	-3.45	0.000280293	-2.95	0.001588869	-2.45	0.007142810
-3.94	0.000040740	-3.44	0.000290857	-2.94	0.001641061	-2.44	0.007343630
-3.93	0.000042472	-3.43	0.000301790	-2.93	0.001694810	-2.43	0.007549411
-3.92	0.000044274	-3.42	0.000313105	-2.92	0.001750156	-2.42	0.007760253
-3.91	0.000046148	-3.41	0.000324814	-2.91	0.001807143	-2.41	0.007976260
-3.90	0.000048096	-3.40	0.000336929	-2.90	0.001865813	-2.40	0.008197535
-3.89	0.000050122	-3.39	0.000349463	-2.89	0.001926209	-2.39	0.008424186
-3.88	0.000052228	-3.38	0.000362429	-2.88	0.001988375	-2.38	0.008656319
-3.87	0.000054417	-3.37	0.000375840	-2.87	0.002052358	-2.37	0.008894042
-3.86	0.000056693	-3.36	0.000389712	-2.86	0.002118205	-2.36	0.009137467
-3.85	0.000059058	-3.35	0.000404057	-2.85	0.002185961	-2.35	0.009386705
-3.84	0.000061517	-3.34	0.000418891	-2.84	0.002255676	-2.34	0.009641869
-3.83	0.000064071	-3.33	0.000434229	-2.83	0.002327400	-2.33	0.009903075
-3.82	0.000066725	-3.32	0.000450087	-2.82	0.002401182	-2.32	0.010170438
-3.81	0.000069483	-3.31	0.000466479	-2.81	0.002477074	-2.31	0.010444077
-3.80	0.000072348	-3.30	0.000483424	-2.80	0.002555130	-2.30	0.010724110
-3.79	0.000075323	-3.29	0.000500936	-2.79	0.002635402	-2.29	0.011010658
-3.78	0.000078414	-3.28	0.000519035	-2.78	0.002717944	-2.28	0.011303844
-3.77	0.000081623	-3.27	0.000537737	-2.77	0.002802814	-2.27	0.011603791
-3.76	0.000084956	-3.26	0.000557061	-2.76	0.002890068	-2.26	0.011910625
-3.75	0.000088417	-3.25	0.000577025	-2.75	0.002979763	-2.25	0.012224472
-3.74	0.000092010	-3.24	0.000597648	-2.74	0.003071959	-2.24	0.012545461
-3.73	0.000095739	-3.23	0.000618951	-2.73	0.003166716	-2.23	0.012873721
-3.72	0.000099611	-3.22	0.000640952	-2.72	0.003264095	-2.22	0.013209383
-3.71	0.000103629	-3.21	0.000663674	-2.71	0.003364160	-2.21	0.013552581
-3.70	0.000107799	-3.20	0.000687137	-2.70	0.003466973	-2.20	0.013903447
-3.69	0.000112127	-3.19	0.000711363	-2.69	0.003572600	-2.19	0.014262118
-3.68	0.000116616	-3.18	0.000736375	-2.68	0.003681108	-2.18	0.014628730
-3.67	0.000121275	-3.17	0.000762194	-2.67	0.003792562	-2.17	0.015003422
-3.66	0.000126107	-3.16	0.000788845	-2.66	0.003907032	-2.16	0.015386334
-3.65	0.000131120	-3.15	0.000816352	-2.65	0.004024588	-2.15	0.015777607
-3.64	0.000136319	-3.14	0.000844739	-2.64	0.004145301	-2.14	0.016177383
-3.63	0.000141710	-3.13	0.000874031	-2.63	0.004269243	-2.13	0.016585806
-3.62	0.000147301	-3.12	0.000904255	-2.62	0.004396488	-2.12	0.017003022
-3.61	0.000153098	-3.11	0.000935436	-2.61	0.004527111	-2.11	0.017429177
-3.60	0.000159108	-3.10	0.000967603	-2.60	0.004661188	-2.10	0.017864420
-3.59	0.000165338	-3.09	0.001000782	-2.59	0.004798796	-2.09	0.018308899
-3.58	0.000171797	-3.08	0.001035002	-2.58	0.004940015	-2.08	0.018762766
-3.57	0.000178490	-3.07	0.001070293	-2.57	0.005084925	-2.07	0.019226172
-3.56	0.000185427	-3.06	0.001106684	-2.56	0.005233608	-2.06	0.019699270
-3.55	0.000192615	-3.05	0.001144206	-2.55	0.005386145	-2.05	0.020182215
-3.54	0.000200063	-3.04	0.001182890	-2.54	0.005542623	-2.04	0.020675162
-3.53	0.000207779	-3.03	0.001222768	-2.53	0.005703126	-2.03	0.021178269
-3.52	0.000215773	-3.02	0.001263873	-2.52	0.005867741	-2.02	0.021691693
-3.51	0.000224053	-3.01	0.001306238	-2.51	0.006036558	-2.01	0.022215594

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
-2.00	0.022750131	-1.50	0.066807201	-1.00	0.158655253	-0.50	0.308537538
-1.99	0.023295467	-1.49	0.068112117	-0.99	0.161087059	-0.49	0.312066949
-1.98	0.023851764	-1.48	0.069436623	-0.98	0.163543059	-0.48	0.315613696
-1.97	0.024419185	-1.47	0.070780876	-0.97	0.166023246	-0.47	0.319177508
-1.96	0.024997895	-1.46	0.072145036	-0.96	0.168527607	-0.46	0.322758110
-1.95	0.025588059	-1.45	0.073529259	-0.95	0.171056126	-0.45	0.326355220
-1.94	0.026189844	-1.44	0.074933699	-0.94	0.173608780	-0.44	0.329968553
-1.93	0.026803418	-1.43	0.076358509	-0.93	0.176185542	-0.43	0.333597820
-1.92	0.027428949	-1.42	0.077803840	-0.92	0.178786379	-0.42	0.337242726
-1.91	0.028066606	-1.41	0.079269841	-0.91	0.181411254	-0.41	0.340902973
-1.90	0.028716559	-1.40	0.080756659	-0.90	0.184060125	-0.40	0.344578258
-1.89	0.029378980	-1.39	0.082264438	-0.89	0.186732943	-0.39	0.348268273
-1.88	0.030054038	-1.38	0.083793322	-0.88	0.189429654	-0.38	0.351972707
-1.87	0.030741908	-1.37	0.085343450	-0.87	0.192150202	-0.37	0.355691245
-1.86	0.031442762	-1.36	0.086914961	-0.86	0.194894521	-0.36	0.359423566
-1.85	0.032156774	-1.35	0.088507991	-0.85	0.197662543	-0.35	0.363169348
-1.84	0.032884118	-1.34	0.090122672	-0.84	0.200454193	-0.34	0.366928263
-1.83	0.033624969	-1.33	0.091759135	-0.83	0.203269391	-0.33	0.370699981
-1.82	0.034379502	-1.32	0.093417508	-0.82	0.206108053	-0.32	0.374484165
-1.81	0.035147893	-1.31	0.095097917	-0.81	0.208970087	-0.31	0.378280478
-1.80	0.035930319	-1.30	0.096800484	-0.80	0.211855398	-0.30	0.382088577
-1.79	0.036726955	-1.29	0.098525329	-0.79	0.214763884	-0.29	0.385908118
-1.78	0.037537980	-1.28	0.100272567	-0.78	0.217695437	-0.28	0.389738752
-1.77	0.038363570	-1.27	0.102042315	-0.77	0.220649946	-0.27	0.393580126
-1.76	0.039203903	-1.26	0.103834681	-0.76	0.223627292	-0.26	0.397431886
-1.75	0.040059156	-1.25	0.105649773	-0.75	0.226627352	-0.25	0.401293674
-1.74	0.040929508	-1.24	0.107487697	-0.74	0.229649997	-0.24	0.405165128
-1.73	0.041815137	-1.23	0.109348552	-0.73	0.232695092	-0.23	0.409045884
-1.72	0.042716220	-1.22	0.111232437	-0.72	0.235762497	-0.22	0.412935577
-1.71	0.043632936	-1.21	0.113139446	-0.71	0.238852068	-0.21	0.416833836
-1.70	0.044565462	-1.20	0.115069670	-0.70	0.241963652	-0.20	0.420740290
-1.69	0.045513977	-1.19	0.117023196	-0.69	0.245097093	-0.19	0.424654565
-1.68	0.046478657	-1.18	0.119000107	-0.68	0.248252230	-0.18	0.428576284
-1.67	0.047459681	-1.17	0.121000484	-0.67	0.251428895	-0.17	0.432505068
-1.66	0.048457226	-1.16	0.123024403	-0.66	0.254626914	-0.16	0.436440537
-1.65	0.049471468	-1.15	0.125071935	-0.65	0.257846110	-0.15	0.440382307
-1.64	0.050502583	-1.14	0.127143150	-0.64	0.261086299	-0.14	0.444329995
-1.63	0.051550748	-1.13	0.129238112	-0.63	0.264347292	-0.13	0.448283213
-1.62	0.052616138	-1.12	0.131356881	-0.62	0.267628893	-0.12	0.452241573
-1.61	0.053698928	-1.11	0.133499513	-0.61	0.270930903	-0.11	0.456204687
-1.60	0.054799291	-1.10	0.135666060	-0.60	0.274253117	-0.10	0.460172162
-1.59	0.055917402	-1.09	0.137856572	-0.59	0.277595324	-0.09	0.464143607
-1.58	0.057053433	-1.08	0.140071090	-0.58	0.280957308	-0.08	0.468118627
-1.57	0.058207555	-1.07	0.142309654	-0.57	0.284338849	-0.07	0.472096829
-1.56	0.059379940	-1.06	0.144572299	-0.56	0.287739718	-0.06	0.476077817
-1.55	0.060570758	-1.05	0.146859056	-0.55	0.291159686	-0.05	0.480061194
-1.54	0.061780176	-1.04	0.149169950	-0.54	0.294598516	-0.04	0.484046563
-1.53	0.063008364	-1.03	0.151505002	-0.53	0.298055965	-0.03	0.488033526
-1.52	0.064255487	-1.02	0.153864230	-0.52	0.301531787	-0.02	0.492021686
-1.51	0.065521712	-1.01	0.156247645	-0.51	0.305025730	-0.01	0.496010643

Referências Bibliográficas

- [AZW89] Daniel Tabak Alexander Zayezdny and Dov Wulich, *Engineering applications of stochastic processes - theory, problems and solutions*, Research Studies Press, Taunton, Somerset, England, 1989.
- [Fel68] W. Feller, *An introduction to probability theory and its implications*, vol. I, John Wiley and Sons, New York, 1968.
- [Hay01] Simon Haykin, *Communication systems*, John Wiley and Sons, 2001.
- [Hsu96] Hwei P. Hsu, *Probability, random variables and stochastic processes*, McGraw-Hill, 1996.
- [Jr.87] Wilber B. Davenport Jr., *Probability and random processes. an introduction for applied scientists and engineers.*, McGraw-Hill, 1987.
- [Lat89] Bhagwandas Pannalal Lathi, *Modern digital and analog communication systems*, Saunders College Publishing, 1989.
- [LG94] Alberto Leon-Garcia, *Probability and random processes for electrical engineering - second edition*, Addison-Wesley, 1994.
- [Lip93] Seymour Lipschutz, *Probabilidade*, Makron Books, São Paulo, 1993.
- [Pap84] Athanasios Papoulis, *Probability, random variables and stochastic processes*, McGraw-Hill, 1984.
- [Pro95] John G. Proakis, *Digital communications*, McGraw-Hill, 1995.
- [Ros83] Sheldon. M. Ross, *Stochastic processes*, John Wiley and Sons, New York, 1983.
- [Spi78] Murray R. Spiegel, *Probabilidade e estatística*, McGraw-Hill, 1978.
- [SW94] Henry Stark and John W. Woods, *Probability, random processes and estimation theory for engineers - second edition*, Prentice Hall, New Jersey, 1994.
- [Swo94] Earl W. Swokowsky, *Cálculo com geometria analítica*, Makron Books, São Paulo, 1994.
- [YG98] Roy D. Yates and David J. Goodman, *Probability and stochastic processes - a friendly introduction for electrical and computer engineers*, John Wiley and Sons, New York, 1998.