

Teoria das Probabilidades

FERNANDO LUCAMBIO PÉREZ
Fevereiro de 2020

Sumário

1	Introdução	1
1.1	Conjuntos e classes de conjuntos	2
1.2	Probabilidade clássica	8
1.2.1	Espaço amostral	8
1.2.2	Combinações e Permutações	10
1.3	Outras definições de probabilidade	16
1.3.1	Probabilidade frequencial	17
1.3.2	Probabilidade geométrica	20
1.4	Exercícios	25
2	Probabilidade axiomática	31
2.1	Álgebra e σ -álgebra de eventos aleatórios	31
2.1.1	σ -álgebra de Borel	38
2.2	Definição axiomática de probabilidade	46
2.2.1	Propriedades da probabilidade axiomática	54
2.3	Probabilidade condicional	60
2.3.1	Independência de eventos	67
2.3.2	Permutabilidade de eventos	73
2.4	Exercícios	76
3	Variáveis aleatórias	83
3.1	Variáveis aleatórias	83
3.1.1	Propriedades das variáveis aleatórias	91
3.2	Função de distribuição	97
3.2.1	Variáveis aleatórias discretas	106
3.2.2	Variáveis aleatórias contínuas	117
3.2.3	Variáveis aleatórias mistas	137
3.3	Funções de variáveis aleatórias	140
3.4	Exercícios	154
4	Momentos	161
4.1	Esperança	162
4.1.1	Propriedades da esperança	171
4.1.2	Esperança de funções de variáveis aleatórias	182
4.1.3	Percentis	191

4.2	Variância	198
4.2.1	Propriedades da variância	201
4.3	Funções geradoras	203
4.3.1	Função geradora de probabilidade	203
4.3.2	Função geradora de momentos	208
4.4	Exercícios	218
5	Vetores aleatórios	223
5.1	Vetores aleatórios	224
5.2	Distribuição conjunta	227
5.3	Funções de vetores aleatórios	228
5.4	Momentos de vetores aleatórios	229
5.4.1	Igualdade de variáveis aleatórias	230
5.4.2	Distribuições de funções de vetores aleatórios	230
5.5	Exercícios	232
6	Distribuição e momentos condicionais	235
6.1	Distribuição condicional	235
6.1.1	Distribuição truncada	238
6.2	Esperança condicional	240
6.2.1	Esperança e distribuição condicionais	242
6.2.2	Caso discreto	245
6.2.3	Caso absolutamente contínuo	247
6.3	Variáveis aleatórias independentes	248
6.3.1	Independência e permutabilidade	248
6.4	Variância condicional	261
6.5	Distribuição normal multivariada	263
6.6	Modelos Copula	265
6.6.1	Propriedades	270
6.6.2	Famílias Copulas	271
6.7	Exercícios	272
7	Convergência estocástica	275
7.1	Função característica	275
7.1.1	Propriedades da função característica	278
7.1.2	Função característica de vetores aleatórios	302
7.1.3	Distribuições infinitamente divisíveis	305
7.2	Desigualdades probabilísticas	312
7.3	Modos de convergência estocástica	320
7.3.1	Convergência em probabilidade	321
7.3.2	Convergência em momentos	326
7.3.3	Convergência quase certamente	332
7.3.4	Convergência em distribuição	335
7.4	Relações entre os modos de convergência	340
7.5	Exercícios	351

8	Teoremas Limites	355
8.1	Lei dos Grandes Números	355
8.1.1	Lei Fraca dos Grandes Números	355
8.1.2	Lei Forte dos Grandes Números	363
8.2	Teorema do Limite Central	370
8.2.1	Ordens de magnitude	371
8.2.2	Teoremas limites	376
8.2.3	Expansões para densidades	390
8.2.4	Velocidade de Convergência	391
8.3	Exercícios	396
9	Referências Bibliográficas	399

Prefácio

Com este livro pretendemos constituir um material de consulta para estudantes de graduação em estatística em particular e, de maneira geral, para qualquer pessoa que queira obter conhecimentos mais do que a nível básico de estatística. Queremos dizer que este material é apropriado para descobrir o assunto assim como também para aprofundar os conhecimentos em na teira das probabilidades. Consideramos então adequado este material para níveis de especialização e mestrado em outras áreas do conhecimento, desde que o interessado conheça álgebra e cálculo.

Os primeiros dois capítulos descrevem os chamados modelos de regressão linear. Apresentamos uma definição moderna destes modelos estatísticos no Capítulo , a estimação e diversos testes de hipóteses no Capítulo . Entendidos estes capítulos podemos aventurarmos em novos conhecimentos. ... o chamado coeficiente de determinação, sua importancia e limitações. entenderemos diferentes resíduos e seas utilidades e

Tópicos mais avançados são tratados dos capítulos e e . No Capítulo abordamos um tema não muito conhecido, é a aplicação do coeficiente de determinação para dizer quanto da resposta é explicado por cada variável dependente. No Capítulo nos dedicamos a um assunto mais ...

O auxilio computacional é baseado na linguagem de programação **R**, versão 2.12. Exemplos mostrados no texto e outros de auxilio na página de modelos de regressão são todos resolvidos utilizando-a. Este texto foi redigido utilizando \LaTeX

Curitiba
Agosto de 2014.

Capítulo 1

Introdução

A teoria das probabilidades está relacionada com a noção matemática do conceito intuitivo de chance ou aleatoriedade o qual, como muitos conceitos, teve sua origem em experiências práticas. Assim podemos dizer que o cálculo de probabilidades teve sua origem em jogos de azar. A curiosidade inicial de apostadores levantaram questões respondidas por matemáticos, as quais produziram um desenvolvimento lento e esporádico da teoria das probabilidades como disciplina matemática. Isso devido a que os matemáticos na época não tinham interesse no desenvolvimento de qualquer teoria somente relacionada com combinações, as quais resolviam a maioria dos problemas em jogos de azar.

Disputas entre jogadores durante a época do Renascimento foram comuns e, em algumas situações, envolveram famosos matemáticos. Em particular, questionamentos de um jogador em 1654 levou a uma troca de correspondências entre dois grandes matemáticos à época: Blaise Pascal¹ e Pierre de Fermat².

Antoine Gombaud³, Chevalier de Méré, um nobre francês com interesse em questões de jogos de azar, chamou a atenção de Pascal e Fermat para uma aparente contradição relativa a um jogo de dados popular. O jogo consistia em atirar um par de dados 24 vezes, o problema era decidir se deve ou não apostar o dinheiro da ocorrência de “pelo menos um duplo seis durante os 24 lances”. A regra do jogo aparentemente bem estabelecida de Méré levou a crer que a aposta em um duplo seis em 24 arremessos seria rentável, mas seus próprios cálculos indicavam justamente o oposto.

Este problema e outros colocados por de Méré levaram a uma troca de correspondência entre Pascal e Fermat em que os princípios fundamentais da teoria das probabilidades foram formuladas pela primeira vez. Apesar de alguns problemas especiais sobre jogos de azar tinham sido resolvidos por alguns matemáticos italianos nos séculos 15 e 16, nenhuma teoria geral foi desenvolvida antes destas famosas correspondências.

O cientista holandês Christian Huygens⁴, sabendo destas correspondências e pouco depois, em 1657, publicou o primeiro livro sobre probabilidade; intitulado “De Ratiociniis no Ludo Aleae”, era um tratado sobre os problemas associados ao jogo. Devido ao apelo inerente de

¹Blaise Pascal (1623 - 1662) foi um físico, matemático, filósofo moralista e teólogo francês.

²Pierre de Fermat (nascido na primeira década do século XVIII - 1665) foi um matemático e cientista francês.

³Antoine Gombaud, denominado Chevalier de Méré (1607 - 1684), foi um nobre e jogador francês

⁴Christiaan Huygens (1629 - 1695) foi um físico, matemático, astrônomo e horologista neerlandês.

jogos de azar, a teoria das probabilidades logo se tornou popular, e o assunto se desenvolveu rapidamente durante o século 18. Os principais contribuintes durante este período foram Jakob Bernoulli⁵ e Abraham de Moivre⁶.

Em 1812, Pierre de Laplace⁷ introduziu uma série de novas ideias e técnicas matemáticas. Antes dele a teoria das probabilidades esteve apenas preocupada com o desenvolvimento de uma análise matemática de jogos de azar e Laplace aplica as ideias probabilísticas para muitos problemas científicos e práticos. A teoria dos erros, a matemática atuarial e a mecânica estatística, são exemplos de algumas das importantes aplicações da teoria das probabilidades desenvolvidas no século 19.

Como a conhecemos atualmente, a teoria matemática das probabilidades é recente, deve-se aos axiomas de Andrei Kolmogorov⁸ que participou das principais descobertas científicas do século XX nas áreas de probabilidades e estatística. Autor da principal teoria científica no campo das probabilidades: a teoria da medida, que revolucionou o cálculo de integrais.

Como tantos outros ramos da matemática, o desenvolvimento da teoria das probabilidades tem sido estimulado pela variedade das suas aplicações. Por outro lado, cada avanço na teoria alargou o âmbito de sua influência. A estatística matemática é um importante ramo da probabilidade aplicada; outras aplicações ocorrem em campos tão diferentes como a genética, psicologia, economia e engenharia. Muitos cientistas têm contribuído para a teoria desde a época de Laplace, entre os mais importantes são Tchebychev⁹, Markov¹⁰, von Mises¹¹ e claro, Kolmogorov.

Primeiro estudaremos os conceitos de conjuntos e classes de conjuntos aplicados à teoria das probabilidades, isso na Seção 1.1. Depois consideramos o conceito clássico de probabilidades, justamente aquela ideia inicial relacionada com jogos de azar, na Seção 1.2 e por último outros dois conceitos relacionados a este são apresentados no Seção 1.3.

1.1 Conjuntos e classes de conjuntos

Neste livro, sempre que a palavra conjunto é usada presume-se designar um subconjunto de um dado conjunto Ω salvo indicação em contrário. Em geral, as letras maiúsculas A , B , C , etc. irão denotar conjuntos e as letras minúsculas u , v , w , x , y , z , etc. representarão pontos ou elementos de conjuntos. Classes de conjuntos são geralmente indicados por letras de roteiro \mathcal{A} , \mathcal{B} , \mathcal{L} , etc.. Lidaremos apenas com classes de conjuntos não vazias.

⁵Jakob Bernoulli (1654 - 1705), foi o primeiro matemático a desenvolver o cálculo infinitesimal.

⁶Abraham de Moivre (1667 - 1754) foi um matemático francês famoso pela Fórmula de Moivre, que relaciona os números complexos com a trigonometria e por seus trabalhos na distribuição normal e na teoria das probabilidades.

⁷Pierre Simon Marquis de Laplace (1749 - 1827) foi um matemático, astrônomo e físico francês que organizou a astronomia matemática.

⁸Andrei Nikolaevich Kolmogorov (1903 - 1987) foi um matemático soviético.

⁹Pafnuti Lvovitch Tchebychev (1821 - 1894) foi um matemático russo.

¹⁰Andrei Andreyevich Markov (1856 - 1922) foi um matemático russo.

¹¹Richard Edler von Mises (1883 - 1953) foi um matemático e engenheiro mecânico austríaco.

Definição 1.1

Sejam A e B dois conjuntos em Ω . Dizemos que A implica na ocorrência de B ou que A é um subconjunto de B , em símbolos escrevemos $A \subset B$, se

$$A \subset B = \{w \in \Omega : \text{se } w \in A \text{ então } w \in B\}.$$

Como é habitual, os símbolos \cup e \cap serão usadas para identificar a união e interseção de conjuntos.

Definição 1.2

Sejam A e B dois conjuntos em Ω . Definimos união dos conjuntos A e B como

$$A \cup B = \{w \in \Omega : w \in A \text{ ou } w \in B\}.$$

Definição 1.3

Sejam A e B dois conjuntos em Ω . A interseção destes conjuntos é definida como

$$A \cap B = \{w \in \Omega : w \in A \text{ e } w \in B\}.$$

Caso não existam elementos de Ω que pertençam a $A \cap B$ dizemos que A e B são disjuntos e escrevemos $A \cap B = \emptyset$, o conjunto \emptyset será chamado de conjunto vazio. Isso significa que A e B não ocorrem simultaneamente. A união dos conjuntos A e B representa o conjunto de que pelo menos um dos dois conjuntos A ou B ocorrem. Por outro lado, a interseção de dois conjuntos A e B representa o conjunto de que ambos A e B ocorrem.

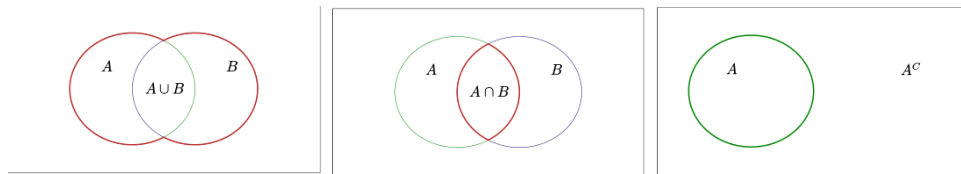


Figura 1.1: A esquerda a situação em que o conjunto A implica na ocorrência de B e a direita conjuntos disjuntos.

A Figura 1.1 a esquerda mostra graficamente o significado do conceito na Definição 1.1 e a direita na mesma figura apresentamos uma situação de conjuntos sem intercepto ou disjuntos. Estes diagramas são conhecidos como Diagramas de Venn¹².

¹²John Venn (1834 - 1923) foi um matemático inglês, estudou e ensinou lógica e teoria das probabilidades.

Exemplo 1.1

Sejam os conjuntos de números reais

$$A = \{w \in \mathbb{R} : 0 \leq w \leq 1\} \quad \text{e} \quad B = \{w \in \mathbb{R} : 1/2 < w \leq 2\}.$$

Então

$$A \cap B = \{w \in \mathbb{R} : 1/2 < w \leq 1\}$$

e

$$A \cup B = \{w \in \mathbb{R} : 0 \leq w \leq 2\}.$$

Teorema 1.1

Para quaisquer conjuntos A, B, C, D temos os seguintes resultados:

- a) $A \cap B \subseteq A \subseteq A \cup B$,
- b) Se $A \subseteq C$ e $B \subseteq D$ então $A \cap B \subseteq C \cap D$ e $A \cup B \subseteq C \cup D$,
- c) $A \subseteq C$ e $B \subseteq C$ se, e somente se, $A \cup B \subseteq C$

Demonstração : Demonstraremos a parte a), os outros resultados serão deixados como exercícios para o leitor. Seja $w \in A \cap B$, então $w \in A$ e $w \in B$. Em particular, $w \in A$, pela definição de interseção. Agora, se $w \in A$ pela definição de união de conjuntos $w \in A \cup B$, isto implica que $A \subseteq A \cup B$. ■

Definição 1.4

Os conjuntos A e B são iguais se $A \subseteq B$ e $B \subseteq A$. Isto é $A = B$ se, e somente se, $A \subseteq B$ e $B \subseteq A$.

Uma vez definido o que entendemos por igualdade de conjuntos podemos estabelecer novas relações e propriedades de conjuntos.

Definição 1.5

Sejam A e B dois conjuntos em Ω . Dizemos que A e B formam uma partição de Ω se $A \cup B = \Omega$ e $A \cap B = \emptyset$.

De acordo com esta definição, se os conjuntos A e B formam uma partição então não podem ter elementos em comum. Por outro lado, uma condição necessária e suficiente para

que $A \cap B \neq \emptyset$ é que A e B tenham, pelo menos, um elemento em comum. Dizemos que se $A \cap B = \emptyset$, os conjuntos A e B não ocorrem simultaneamente. O seguinte teorema resume as principais propriedades da interseção e união.

Teorema 1.2

As operações \cap e \cup são:

- a) *Refletivas*: para todo A , $A \cap A = A = A \cup A$,
- b) *Associativas*: $A \cap (B \cap C) = (A \cap B) \cap C$ e $A \cup (B \cup C) = (A \cup B) \cup C$,
- c) *Comutativas*: $A \cap B = B \cap A$ e $A \cup B = B \cup A$,
- d) *Distributivas*: $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ e $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.

Demonstração : Exercício. ■

Um outro conjunto que tem muita aplicabilidade na teoria das probabilidades é definido a seguir, observe que é uma situação especial de conjuntos que não podem ocorrer simultaneamente.

Definição 1.6

Denotaremos por A^c o complemento do conjunto A , em Ω claro, e será definido como

$$A^c = \{w \in \Omega : w \notin A\}.$$

Observemos que A^c representa o conjunto de que A não ocorre, assim podemos observar que $\emptyset^c = \Omega$ e ,vice-versa, $\Omega^c = \emptyset$. Aplicando este novo conceito podemos definir a diferença entre conjuntos como a seguir.

Definição 1.7

A diferença entre A e B é definida como os elementos de A que não estão em B , denotada por $A \setminus B$, isto é,

$$A \setminus B = \{w \in A : w \notin B\}.$$

Observe que, caso $A = \Omega$, $A \setminus B = \Omega \setminus B = B^c$, também caso $B = \Omega$, então, $A \setminus B = A \setminus \Omega = \emptyset$.

Exemplo 1.2

A diferença entre os conjuntos A e B pode ser escrita como $A \setminus B = A \cap B^c$. Provemos isto,

$$A \cap B^c = \{w \in A : w \notin B\} = A \setminus B.$$

Exemplo 1.3

Sejam $A = \{x \in \mathbb{R} : 0 \leq x \leq 2\}$ e $B = \{x \in \mathbb{R} : 1 < x \leq 2\}$, então $A \setminus B = \{x \in \mathbb{R} : 0 \leq x \leq 1\}$.

A operação de diferença não satisfaz as propriedades da união e interseção listadas no Teorema 1.2. Por exemplo, se $A \neq \emptyset$, $(A \cup A) \setminus A \neq A \cup (A \setminus A)$, isto significa que a colocação de parêntesis em $A \cup A \setminus A$ é importante, o qual não acontece com as operações \cap e \cup . Uma outra diferença é que, enquanto as operações de união e interseção são operações comutativas, a diferença de conjuntos não é.

Teorema 1.3

Para A e B arbitrários, temos que

$$A \cap B = A \setminus (A \setminus B).$$

Demonstração: Provemos que $A \setminus (A \setminus B) \subseteq A \cap B$. Seja $w \in A \setminus (A \setminus B)$, então $w \neq (A \setminus B)$ e isto significa que w não é um elemento de A ou w é um elemento de B mas, dado que $w \in A \setminus (A \setminus B)$, w é um elemento de A . Logo w é elemento de A e também w é um elemento de B , portanto $w \in A \cap B$. A demonstração da outra desigualdade $A \cap B \subseteq A \setminus (A \setminus B)$ é um exercício para o leitor. ■

Estamos considerando relações entre conjuntos e um dos resultados elementais de maior utilidade é o seguinte, conhecido na literatura como Leis de De Morgan¹³. Estas leis serão apresentadas aqui para o caso de dois conjuntos e depois generalizadas para quantidades enumeráveis de conjuntos.

Teorema 1.4 (Leis de De Morgan)

Sejam $A, B \subseteq \Omega$ então:

a) $(A \cup B)^c = A^c \cap B^c$.

b) $(A \cap B)^c = A^c \cup B^c$.

¹³Augustus De Morgan (1806-1871) foi um matemático e lógico britânico. Formulou as Leis de De Morgan e foi o primeiro a introduzir o termo e tornar rigorosa a ideia da indução matemática.

Demonstração: a) Seja $w \in (A \cup B)^c$, significa que $w \in \Omega$ mas $w \notin (A \cup B)$. Agora, $w \notin (A \cup B)$ se, e somente se, $w \notin A$ e $w \notin B$, o qual significa que $w \in A^c$ e $w \in B^c$, logo $w \in A^c \cap B^c$ e vice-versa. Para demonstrar b) observemos que $\Omega \setminus (A^c \cup B^c) = \Omega \setminus A^c \cap \Omega \setminus B^c$, pelo item a) e $\Omega \setminus A^c \cap \Omega \setminus B^c = A \cap B$. Então $(A \cap B)^c = A^c \cup B^c$. ■

Uma situação interessante é que qualquer propriedade apresentada como válida para dois conjuntos, também é válida para uma quantidade finita deles. Assim, por exemplo, se A_1, A_2, \dots, A_n forem subconjuntos de Ω podemos afirmar que

$$(A_1 \cup A_2 \cup \dots \cup A_n)^c = A_1^c \cap A_2^c \cap \dots \cap A_n^c,$$

pelo Teorema 1.4, igualmente válida a afirmação para a afirmação do item b).

Definição 1.8

Sejam A e B dois conjuntos, define-se a diferença simétrica de A e B e denota-se por $A \Delta B$ como

$$A \Delta B = (A \setminus B) \cup (B \setminus A).$$

A diferença simétrica é comutativa e associativa, isto é, $A \Delta B = B \Delta A$ e $(A \Delta B) \Delta C = A \Delta (B \Delta C)$. Ainda temos que

$$(A \Delta B) \Delta (B \Delta C) = (A \Delta C).$$

O conjunto vazio é neutro, e cada conjunto é a sua própria inversa

$$A \Delta \emptyset = A$$

e

$$A \Delta A = \emptyset.$$

Por sua vez a interseção é distributiva sobre a diferença simétrica, já que

$$A \cap (B \Delta C) = (A \cap B) \Delta (A \cap C),$$

e também temos a unicidade da diferença simétrica, demonstrado no teorema a seguir.

Teorema 1.5

A diferença simétrica de dois conjuntos é única, ou seja, se

$$A \Delta B = A \Delta C$$

então $B = C$.

Demonstração: Segundo a Definição 1.4 devemos provar que $B \subseteq C$ e $C \subseteq B$. Primeiro consideremos que $B \subseteq C$. Sabemos que $A \Delta B = (A \setminus B) \cup (B \setminus A) = (A \setminus C) \cup (C \setminus A) = A \Delta C$. Seja $w \in A \Delta B$, então ou $w \in (A \setminus B)$ ou $w \in (B \setminus A)$. Caso $w \in (B \setminus A)$ então $w \in (C \setminus A)$. Se $w \in (A \setminus C)$ então $w \in (A \setminus B)$ e $C \subseteq B$. A outra situação $C \subseteq B$ é similar. ■

1.2 Probabilidade clássica

A teoria da probabilidade teve sua origem em jogos de azar e jogos de azar. Ele deve muito à curiosidade dos jogadores que atormentavam a vida dos seus amigos no mundo matemático com todos os tipos de perguntas. Infelizmente esta associação com o jogo contribuiu para um muito lento crescimento esporádico da teoria da probabilidade como uma disciplina matemática.

1.2.1 Espaço amostral

Definição 1.9

Suponhamos que um experimento seja realizado sob certas condições fixas. Define-se o espaço amostral como o conjunto de todos os resultados possíveis do experimento e denota-se por Ω .

Um espaço amostral é o conjunto de todos os possíveis resultados de um experimento ou de todos os resultados considerados possíveis. Assim, o espaço amostral do lançamento de uma moeda seria uma coleção de resultados que inclui: cara, coroa, a moeda cair em pé, a moeda ser despedaçada por uma bala perdida, um pássaro apanhar a moeda em pleno ar e fugir com ela, a moeda ser acidentalmente engolida pelo experimentador e outros. Excluindo os resultados muito inverossímeis, é perfeitamente razoável considerar apenas cara e coroa, de modo que, no exemplo do lançamento de uma moeda: $\Omega = \{cara, coroa\}$.

Exemplo 1.4

Jogar um dado equilibrado e observar o número da face superior. Neste caso $\Omega = \{1, 2, 3, 4, 5, 6\}$, pois esses serão os únicos resultados possíveis.

Definição 1.10

Seja Ω o espaço amostral de um experimento. Todo subconjunto A de Ω , $A \subset \Omega$ será chamado de evento. Ω será chamado de evento certo e \emptyset o evento impossível. Se $w \in \Omega$, o evento $\{w\}$ será chamado de evento elementar ou simples.

Exemplo 1.5

Medir a pressão arterial sistólica de um indivíduo. A pressão arterial sistólica é o maior valor verificado durante a aferição de pressão arterial. Exemplo: 120x80; onde 120 refere-se à pressão arterial sistólica e 80 refere-se à pressão arterial diastólica, ambas medidas em milímetros de mercúrio (mmHg). O espaço amostral seria uma faixa plausível de valores contínuos, como por exemplo, a faixa de 50 mmHg a 250 mmHg ou $\Omega = \{x \in \mathcal{R} : x \in [50, 250]\}$ medido em mmHg.

Nem sempre é fácil definir ou identificar o espaço amostral de um experimento, o importante é que Ω contenha todo resultado possível, por isso vamos supor:

- (i) a todo resultado possível de um experimento corresponde um, e somente um, ponto $w \in \Omega$;
- (ii) resultados distintos correspondem a pontos distintos em Ω , ou seja, w não pode representar mais do que um resultado.

Quando se realiza um experimento há certos eventos que ocorrem e outros que não ocorrem. No exemplo 1.8, alguns eventos que ocorrem são: (i) $A = \text{"Observa-se um número par"}$; (ii) $B = \text{"Observa-se o número 2"}$ e (iii) $C = \text{"Observa-se um número maior ou igual a 4"}$. Notemos que cada um destes eventos pode ser identificado a um subconjunto de Ω , a saber, $A = \{2, 4, 6\}$, $B = \{2\}$ e $C = \{4, 5, 6\}$. Podemos então perguntar, a quais eventos vamos atribuir um valor de probabilidade?

Definição 1.11

Seja Ω o espaço amostral de um experimento e A um evento ($A \subset \Omega$). Um evento A , ao qual atribuímos um valor de probabilidade, será chamado de evento aleatório.

Perguntamos agora, como vamos atribuir probabilidades a eventos? uma primeira resposta é definindo probabilidade clássica, quando Ω é finito. Esta definição se baseia no conceito de resultado equiprovável.

Definição 1.12 (*Definição Clássica de Probabilidade*)

Seja Ω um espaço amostral finito e $A \subset \Omega$. Definimos probabilidade de ocorrência do evento A como

$$P(A) = \frac{\text{número de resultados favoráveis a } A}{\text{número de resultados possíveis}}.$$

A primeira definição de probabilidade conhecida parece ser devida a De Moivre em 1718, e foi claramente explicitada por Laplace no princípio do século XIX. Laplace adotou o esquema de resultados equiprováveis, isto é, dos resultados igualmente prováveis, comuns às aplicações até então esboçadas para definir probabilidade de um acontecimento como: a relação entre o número de casos favoráveis ao acontecimento e o número total de casos possíveis, supondo todos os casos igualmente possíveis.

Admite-se historicamente que a motivação para a definição do conceito de probabilidades foram baseadas em jogos de azar, dessa forma não causa surpresa o fato de que o conceito de Laplace seja baseado nas propriedades de tais jogos: possibilidade de classificar a priori todos os resultados possíveis num número finito de casos mutuamente exclusivos, simétricos e igualmente possíveis. Apesar das críticas que lhe foram dirigidas a interpretação clássica manteve a sua força até o começo do século XX.

Exemplo 1.6

Uma moeda é lançada duas vezes. O espaço amostral consiste de quatro pontos,

$$\Omega = \{(cara, coroa), (cara, cara), (coroa, cara), (coroa, coroa)\}.$$

Sob a suposição de resultados equiprováveis, cada um dos quatro elementos tem como probabilidade de ocorrência $1/4$.

Em problemas de jogos de azar geralmente obtemos espaços amostrais finitos com eventos equiprováveis e aqui surgiu uma primeira grande crítica: o que são eventos equiprováveis se estamos definindo probabilidade? caso os eventos não sejam equiprováveis, não podemos definir probabilidade? Outra questão é se somente podemos calcular probabilidades quando o espaço amostral seja finito.

Nas situações onde as condições para a aplicação da definição clássica de probabilidade se cumprem o cálculo da probabilidade de qualquer evento se reduz à problemas de contagens. Portanto, estudaremos algumas regras para a contagem do número de permutações e combinações a seguir,

1.2.2 Combinações e Permutações

Esta seção explicam-se as noções básicas de análise combinatória e se desenvolve o fundo probabilístico correspondente. Muitos problemas da teoria da probabilidade exigem contar o número de maneiras que um determinado evento pode ocorrer. Por isso, estudamos as combinações e as permutações. Antes de iniciar a discussão do tema, é útil introduzir uma técnica geral que nos permitirá resolver uma variedade de problemas de contagens, incluindo o problema da contagem do número de permutações de n objetos.

Percebemos, da definição de probabilidade clássica que para calcular a probabilidade de um evento A , temos que dividir o número de pontos de amostragem em A ou casos favoráveis pelo número total de pontos de amostragem ou casos possíveis. Isto é facilitado por um uso sistemático de algumas regras que passamos agora a analisar.

Considere uma experiência que tem lugar em várias fases e é tal que o número de resultados na m -ésima fase é independente dos resultados das etapas anteriores. O número m pode ser diferente para as diferentes fases. Queremos contar o número de maneiras que todo o experimento pode ser realizado.

Teorema 1.6

Com m elementos a_1, \dots, a_m e n elementos b_1, \dots, b_n é possível formar $m \times n$ pares da forma (a_i, b_k) ou arranjos ordenados, contendo cada par um elemento de cada grupo.

Demonstração: Organizar os pares em uma matriz retangular na forma de uma tabela de multiplicação com m linhas e n colunas de modo que (a_i, b_k) fica na intersecção da i -ésima linha e a k -ésima coluna. Observemos que cada par aparece apenas uma única vez e a afirmação torna-se óbvia. ■

Exemplo 1.7

Quantos conjuntos formar se escolhemos um dos quatro naipes e um dos treze valores possíveis? Cada carta é definida por seu naipe e seu valor de face, existem então $4 \times 13 = 52$ dessas combinações ou cartas.

Em situações mais complexas utilizarmos a generalização deste procedimento de contagem, conhecido como princípio multiplicativo da contagem.

Teorema 1.7 (*Princípio Multiplicativo da Contagem*)

Dadas as coleções $a_{11}, a_{12}, \dots, a_{1n_1}$ de n_1 elementos, $a_{21}, a_{22}, \dots, a_{2n_2}$ de n_2 elementos até a coleção $a_{k1}, a_{k2}, \dots, a_{kn_k}$ de n_k elementos é possível formar $n_1 \times n_2 \times \dots \times n_k$ arranjos ordenados da forma $(a_{1j_1}, a_{2j_2}, \dots, a_{kj_k})$ contendo um elemento de cada coleção, $1 \leq j_i \leq n_i, i = 1, 2, \dots, k$.

Demonstração: Exercício. ■

Ao utilizar este princípio é fundamental que o número de maneiras de realizar uma determinada etapa não seja influenciado por nenhuma das etapas predecessoras.

Exemplo 1.8

Um número de telefone é formado por uma sequência de 8 dígitos, porém o primeiro deve ser diferente de 0 e 1. Quantos números de telefone distintos existem? Podemos fazer este cálculo via contagem segundo o Teorema 1.7, onde selecionamos um dígito de cada vez. Temos um total de 8 estágios e escolhemos um de 10 elementos em cada estágio, a menos o primeiro no qual temos somente 8 dígitos.

Então a resposta é

$$8 \times \underbrace{10 \times \dots \times 10}_{7 \text{ vezes}} = 8 \times 10^7.$$

Um detalhe na notação, diremos que o número de elementos no conjunto finito A é $|A|$ ou $\#A$, também chamado de cardinal do conjunto A .

Exemplo 1.9

Um baralho comum consiste de 52 cartas separadas em 4 naipes com 13 cartas de cada um. Para cada naipe, os valores das cartas são 2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K e A. Um baralho comum é embaralhado. Qual é a probabilidade de que as quatro cartas do topo tenham:

(a) valores diferentes?

(b) naipes diferentes?

Se consideramos como relevante a ordem entre as quatro cartas do topo, então o espaço amostral consiste de $52 \times 51 \times 50 \times 49$ resultados. Além disso, existem $52 \times 48 \times 44 \times 40$ resultados em que as cartas têm valores diferentes e $52 \times 39 \times 26 \times 13$ resultados em que

as cartas têm naipes diferentes. Portanto, assumindo que o embaralhamento significa que cada resultado no espaço amostral é igualmente provável, temos que as probabilidades desejadas são

(a)

$$\frac{52 \times 48 \times 44 \times 40}{52 \times 51 \times 50 \times 49} \approx 0,676$$

(b)

$$\frac{52 \times 39 \times 26 \times 13}{52 \times 51 \times 50 \times 49} \approx 0,105.$$

Teorema 1.8 (*Princípio Aditivo da Contagem*)

Se A_1, \dots, A_n são conjuntos dois a dois disjuntos, então

$$\left| \bigcup_{k=1}^n A_k \right| = \sum_{k=1}^n |A_k|.$$

Demonstração: Exercício. ■

Convém recordar uma técnica bastante útil em problemas de contagem: primeiro ignore uma restrição do problema, contando a mais. Depois, desconte o que foi indevidamente contado. Mais geral ainda do que o princípio apresentado no Teorema 1.8 é o chamado princípio da inclusão-exclusão a seguir.

Teorema 1.9 (*Princípio da Inclusão-Exclusão*)

Se A_1, \dots, A_n são conjuntos quaisquer, então

$$\begin{aligned} \left| \bigcup_{k=1}^n A_k \right| &= \sum_{k=1}^n |A_k| - \sum_{i < j} |A_i \cap A_j| + \sum_{i < j < k} |A_i \cap A_j \cap A_k| \\ &\quad + \dots + (-1)^{n+1} |A_1 \cap \dots \cap A_n|. \end{aligned}$$

Demonstração: Exercício. ■

O nome deriva da ideia de que o princípio se baseia em inclusão excessivamente generosa, seguida de compensação por exclusão. Esse conceito é atribuído a Abraham de Moivre num

trabalho de 1718, mas aparece pela primeira vez em um artigo de Daniel da Silva¹⁴ de 1854 e mais tarde em um artigo de J.J. Sylvester¹⁵ publicado em 1883. O princípio é um exemplo do método da peneira amplamente utilizado na teoria dos números e às vezes é chamado de fórmula da peneira.

Exemplo 1.10

Quantas permutações diferentes existem das letras A, B, C, D, E, F

- (a) *que têm as letras A, B juntas em qualquer ordem? Para encontrar este número imaginamos as letras A e B coladas como uma letra só, na ordem AB , o que fornece $5!$ permutações. Como também existem $5!$ permutações nas quais a letra B está imediatamente antes da letra A , obtemos um total de $2 \times 5! = 240$ permutações diferentes;*
- (b) *que têm a letra A em primeiro lugar ou a letra F em último? Sejam \mathcal{A} o conjunto das permutações que começam por A e \mathcal{F} o conjunto das permutações que terminam em F . Pelo Princípio da Inclusão-Exclusão, o número de permutações que começam por A ou terminam em F é*

$$|\mathcal{A} \cup \mathcal{F}| = |\mathcal{A}| + |\mathcal{F}| - |\mathcal{A} \cap \mathcal{F}| = 5! + 5! - 4! = 216;$$

- (c) *em que a letra A vem antes da letra B ? Existe um total de $6! = 720$ permutações possíveis, e existem tantas com A antes de B quantas com B antes de A , logo a resposta é 360;*
- (d) *em que a letra E não é a última? Existem $5!$ permutações em que a letra E é a última, portanto $6! - 5! = 600$ permutações em que E não é a última letra.*

O princípio é mais claramente visto no caso de três conjuntos, A, B e C é dado por

$$|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|. \quad (1.1)$$

Essa fórmula pode ser verificada contando quantas vezes cada região na figura do diagrama de Venn está incluída no lado direito da fórmula. Nesse caso, ao remover as contribuições de elementos com excesso de contagem, o número de elementos na interseção mútua dos três conjuntos foi subtraído com muita frequência; portanto, deve ser adicionado novamente para obter o total correto.

Em um cenário muito abstrato, o princípio de inclusão-exclusão pode ser expresso como o cálculo do inverso de uma determinada matriz. Esse inverso possui uma estrutura especial, tornando o princípio uma técnica extremamente valiosa em combinatória e áreas afins da matemática. Como Gian-Carlo Rota¹⁶ colocou:

“Um dos princípios mais úteis da enumeração em probabilidade discreta e teoria combinatória é o célebre princípio da inclusão-exclusão. Quando aplicado

¹⁴Daniel da Silva (1814 - 1878) foi um metamático português.

¹⁵James Joseph Sylvester (1814 - 1897) era um matemático inglês. Foi fundador do American Journal of Mathematics.

¹⁶Gian-Carlo Rota (1932 - 1999) foi um matemático e filósofo italo-americano.

com habilidade, esse princípio produziu a solução para muitos problemas combinatórios’’

No que segue, vamos nos concentrar principalmente em dois tipos de contagem que envolvem a seleção de k objetos de uma coleção de n objetos. Se a ordem da seleção importa, a seleção é chamada de permutação, caso contrário é chamada de combinação.

Definição 1.13

Consideremos um conjunto de n elementos a_1, a_2, \dots, a_n ao qual chamaremos de população. Qualquer arranjo ordenado $a_{i_1}, a_{i_2}, \dots, a_{i_r}$ de r elementos é chamado de amostra ordenada de tamanho r . Se os elementos da amostra são selecionados um a um, existem duas possibilidades:

- (a) Amostra com reposição: neste caso, as repetições são permitidas e podemos tirar amostras de tamanho arbitrário.
- (b) Amostra sem reposição: neste caso um elemento, uma vez escolhido não é substituído, pelo que não pode haver repetições. É evidente que o tamanho da amostra não pode exceder n .

Teorema 1.10

Seja a_1, \dots, a_n uma população de tamanho n . Se amostras ordenadas de tamanho r são extraídas desta população, então:

- (a) Existem n^r diferentes amostras possíveis com reposição.
- (b) Se as amostras são extraídas sem reposição, existem

$$(n)_r = n \times (n - 1) \cdots \times (n - r + 1)$$

amostras possíveis de tamanho r .

Demonstração: Exercício. ■

Exemplo 1.11

Consideremos uma população de n elementos. Uma amostra de tamanho r é escolhida aleatoriamente com reposição. Então, a probabilidade de que nenhum elemento apareça mais de uma vez é

$$\frac{(n)_r}{n^r}.$$

Definição 1.14

Seja a_1, a_2, \dots, a_n uma população de n elementos. Qualquer arranjo ordenado sem reposição $a_{i_1}, a_{i_2}, \dots, a_{i_n}$ dos n elementos é chamado de permutação.

Al-Khalil (717-786), que foi um matemático e criptografador árabe, escreveu o Livro de Mensagens Criptográficas. Ele contém o primeiro uso de permutações e combinações, para listar todas as possíveis palavras em árabe com e sem vogais (Broemeling, 2011). A regra para determinar o número de permutações de n objetos era conhecida na cultura indiana pelo menos por volta de 1150.

Teorema 1.11

Seja a_1, a_2, \dots, a_n uma população de n objetos quaisquer. O número de permutações possíveis de n objetos é $n!$.

Demonstração: Utilizando o Teorema 1.7, a primeira coleção é a_1, a_2, \dots, a_n , de n elementos.

Uma vez selecionado o elemento a_i a segunda coleção seria $a_1, \dots, a_i, a_{i+1}, \dots, a_n$ de $n - 1$ elementos, até a n -ésima coleção, formada por somente um elemento. Portanto, o número de permutações é $n \cdot (n - 1) \cdot \dots \cdot 2 \cdot 1 = n!$. ■

Exemplo 1.12

Assim, se n bolas devem ser colocadas aleatoriamente em n urnas, a probabilidade de que cada urna seja ocupada é

$$\frac{n!}{n^n}.$$

Definição 1.15

Consideremos a população a_1, a_2, \dots, a_n . Qualquer arranjo não ordenado

$$a_{i_5}, a_{i_1}, \dots, a_{i_r}, \dots, a_{i_2}, \dots, a_{i_{r-1}}$$

de r elementos é chamado de amostra não ordenada de tamanho r . Se os elementos da amostra são selecionados um a um, existem duas possibilidades:

- (a) Amostra não ordenada com reposição.
- (b) Amostra não ordenada sem reposição.

Teorema 1.12

Seja a_1, \dots, a_n uma população de tamanho n . Temos então duas formas de calcular o número de amostras possíveis:

- (a) Se as amostras não ordenadas são extraídas com reposição, existem $\binom{n-r-1}{r}$ amostras possíveis de tamanho r .
- (b) Se amostras não ordenadas de tamanho r são extraídas sem reposição, existem $\binom{n}{r}$ possíveis diferentes destas amostras.

Demonstração : Exercício. ■

Várias críticas foram feitas ao conceito clássico de probabilidade. Por exemplo, o que são casos equiprováveis? na falta de uma definição de casos equiprováveis devemos admitir que é um conceito primitivo? como reconhecer se os casos são equiprováveis? a saída parece ser aceitar que algum princípio apriorístico suporta tal reconhecimento. Nesses casos é comum admitir um dois princípios a seguir:

- (a) Princípio da indiferença, o qual faz apelo às propriedades de simetria ou de homogeneidade da situação experimental. Se o dado é perfeito porque seriam uma das faces preferidas em detrimento de outras?
- (b) Princípio da razão insuficiente: se não há razão para crer que qualquer dos casos é mais provável do que os outros, pode-se admitir que todos os casos são igualmente prováveis.

É bem sabido que não existem moedas perfeitas, dados perfeitos, gases perfeitos, água pura que perfeição além do conceito não existe. Consequentemente o conceito clássico é muitas vezes aplicado em situações idealizadas e não consegue vencer a dificuldades levantadas quando os casos não são igualmente possíveis. Finalmente, como calcular probabilidades quando o número de casos possíveis não é finito nem sequer enumerável?

Apesar de todas as críticas não resta dúvida que a interpretação clássica é aplicável sempre que a simetria dos problemas a justifique e, de fato há numerosos casos em que tal propriedade pode ser aceita. A verdade é que se trata de um modelo probabilístico particular dentro da teoria axiomática a ser desenvolvida, de grande utilidade quando ajustado a uma realidade concreta.

1.3 Outras definições de probabilidade

Até agora definimos probabilidade de um evento segundo a definição clássica, supondo sempre resultados equiprováveis. Outros métodos de definir probabilidade são o da frequência relativa e geométrico.

1.3.1 Probabilidade frequencial

Em muitas situações práticas, os eventos simples do espaço amostral não são equiprováveis ou não podemos supor que sejam equiprováveis. Isso implica que não podemos calcular probabilidades usando a definição clássica. Surgiram então alternativas de cálculo à probabilidade clássica, uma dessas alternativas considera o cálculo de probabilidades como a frequência relativa de ocorrência do evento de interesse.

A abordagem frequencial nos pede para imaginar a repetição do processo físico um número extremamente grande de vezes, conhecidos como “ensaios” do processo físico e, em seguida, olhar para a fração de vezes que o desfecho de interesse ocorre. Essa fração é assintoticamente igual à probabilidade de determinado resultado para esse processo físico.

Definição 1.16 (*Definição Frequencial de Probabilidade*)

Seja Ω um espaço amostral e $A \subset \Omega$. Definimos probabilidade frequencial de ocorrência do evento A como

$$P(A) = \lim_{n \rightarrow \infty} \frac{1}{n} \times \left(\begin{array}{c} \text{número de ocorrências de } A \text{ em } n \\ \text{ensaios independentes do experimento} \end{array} \right).$$

Esta definição nos disse que a probabilidade de ocorrência de A é o limite da frequência relativa da ocorrência do evento A em n repetições independentes do experimento, com n tendendo ao infinito. Baseia-se na experiência, comum a todos nós, da estabilidade da frequência relativa de ocorrência de eventos, quando realizamos muitas repetições do experimento. O problema filosófico com esta abordagem é que habitualmente não se têm a oportunidade de repetir o cenário um grande número de vezes.

Observe-se que, desta forma, não são necessárias as hipóteses de equiprobabilidade dos eventos elementares nem de finitude do espaço dos resultados, superando-se portanto as duas restrições fundamentais da definição clássica. Entretanto, esta nova definição introduz as seguintes dificuldades:

- (i) É necessária certa regularidade da sequência das frequências relativas, no sentido de que a mesma se mantenha estável e convergindo para um valor que seria a probabilidade de A .
- (ii) Mesmo admitindo a existência do limite mencionado em acima, quando parar?

Quanto a (i) é impossível demonstrar concretamente a existência do limite acima mencionado; pode-se testar a estabilidade das frequências relativas como um indício apenas da existência do limite, pois a estabilidade é uma condição necessária para sua existência. Quanto a (ii), se aceitarmos a existência do limite acima ou seja, da probabilidade, em 1713 Bernoulli demonstrou um teorema que garante a convergência da sequência das frequências relativas à probabilidade, de um modo que não precisaremos aqui. Inclusive neste caso é possível estimar a probabilidade de A com precisão e nível de confiança prefixados, se o número de repetições n for suficientemente grande.

Com precisão, se $\epsilon > 0$ então $P(|\text{frequência relativa de } A - P(A)| \geq \epsilon) \leq \frac{1}{4n\epsilon^2}$, ou seja a probabilidade de errar em mais de ϵ ao estimar $P(A)$ através da frequência relativa pode ser tão pequena quanto se quiser, desde que o número de repetições n seja suficientemente grande. Este resultado, cuja demonstração não será apresentada aqui, fundamenta a aproximação da probabilidade através da frequência relativa.

Exemplo 1.13

Considere a experiência de lançar uma moeda honesta. O espaço amostral é $\Omega = \{\text{Cara}, \text{Coroa}\}$. Se o experimento for repetido muitas vezes, a frequência relativa dos resultados costumam estar perto de $1/2$:

- *O naturalista francês Buffon (1707-1788) jogou uma moeda 4040 vezes. Resultando em 2048 caras ou frequência relativa de $2048/4040 = 0,5069$ para caras.*
- *Por volta de 1900 o estatístico inglês Karl Pearson heroicamente lançou uma moeda 24.000 vezes. Resultando em 12.012 caras e uma frequência relativa de 0,5005.*
- *Enquanto estava preso pelos alemães, durante a II Guerra Mundial, o matemático australiano John Kerrich jogou uma moeda de 10.000 vezes. Resultado: 5067 caras, frequência relativa de 0,5067*

Finalmente, deve-se salientar que este método não foi utilizado originalmente como definição, mas como critério empírico destinado a revisar cálculos feitos no contexto dos jogos segundo a definição clássica, na época sujeitos a muitos erros pelos ainda incipientes desenvolvimentos das técnicas de contagem.

Exemplo 1.14

Uma experiência que consiste em observar o sexo de um recém-nascido. Tal experiência já se realizou diversas vezes e existem registros do seu resultado.

A Tabela 1.1 mostra a proporção de meninos entre os nascidos vivos de residentes dos Estados Unidos ao longo de 18 anos. A frequência relativa de meninos entre as crianças recém-nascidas nos Estados Unidos parece ser estável em torno de 0.512. Isto sugere que um modelo razoável para o resultado de um único nascimento é $P(\text{menino}) = 0,512$ e $P(\text{menina}) = 0,488$. Este modelo de nascimentos é equivalente ao sexo de uma criança ser determinado pela escolha ao acaso com reposição de uma caixa de 1.000 bilhetes, contendo 512 bilhetes marcados menino e 488 bilhetes marcados menina. Por outro lado, segundo a definição clássica $P(\text{menino}) = 0,50$ e $P(\text{menina}) = 0,50$.

Vejamos um exemplo para mostrar as diferenças de informações entre a probabilidade clássica e a probabilidade frequencial.

Exemplo 1.15 (*Mega-Sena*)

O jogo da Mega-Sena é um dos jogos de loteria mais conhecidos no Brasil. Existe desde 11 de março de 1996. A Mega-Sena paga milhões para o acertador dos 6 números sorteados dentre os 60 disponíveis no volante de apostas. Pesquisamos o número de ganhadores da Mega-Sena até o sorteio número 1688, o qual aconteceu no dia 21 de março de 2015.

Ano	No. de nascimentos	No. de meninos
1985	3.760.561	0,5126849
1986	3.756.547	0,5124035
1987	3.809.394	0,5121951
1988	3.909.510	0,5121931
1989	4.040.958	0,5121286
1990	4.158.212	0,5121179
1991	4.110.907	0,5112054
1992	4.065.014	0,5121992
1993	4.000.240	0,5121845
1994	3.952.767	0,5116894
1995	3.926.589	0,5084196
1996	3.891.494	0,5114951
1997	3.880.894	0,5116337
1998	3.941.553	0,5115255
1999	3.959.417	0,5119072
2000	4.058.814	0,5117182
2001	4.025.933	0,5111665
2002	4.021.726	0,5117154

Tabela 1.1: Nascimentos vivos de residentes dos Estados Unidos (Fonte: Information Please Almanac).

0	1	2	3	4	5	7	15
1284	298	74	19	9	2	1	1

Tabela 1.2: Número de ganhadores dentre os primeiros 1688 concursos da Mega-Sena.

Utilizando a definição clássica sabemos que a probabilidade de acertar a mega, isto é, de ser o sorteado ao apostar 6 números é de

$$P(\text{Ganhar a Mega-Sena apostando 6 números}) = \frac{1}{\binom{60}{6}} = \frac{1}{50.063.860}.$$

O qual, evidentemente, é um número muito pequeno de acontecer. Agora, utilizando a probabilidade frequencial, observamos o número de ganhadores da mega dentre os 1688 concursos realizados até o dia 21 de março de 2015 e obtivemos os resultados apresentados na Tabela 1.2. Acontece que esta forma de calcular probabilidades não fornece a mesma resposta do que a probabilidade clássica. Com estos resultados não sabemos qual a probabilidade de ser o sortudo apostando somente 6 números, para isto deveríamos saber quantas pessoas diferentes apostaram em cada jogo e esta informação não está disponível. Por outro lado, a probabilidade frequencial fornece outro tipo de informação, por exemplo, nos disse que a probabilidade num jogo da mega não haver acertadores é de $1284/1688 \approx 0,76$ e que a probabilidade de existir somente um acertador num determinado jogo é de $298/1688 \approx 0,18$.

1.3.2 Probabilidade geométrica

Uma outra forma de definir probabilidade é utilizar noções de geometria, conduzindo ao conceito de probabilidade geométrica em situações nas quais os objetos geométricos como pontos, linhas, rotações e áreas são o interesse de estudo.

Probabilidade geométrica é um modelo para calcular as probabilidades da vida real, como a probabilidade de que um ônibus estará esperando fora de um hotel em

Quando da aplicação de uma fórmula para a probabilidade geométrica, é importante que cada ponto sobre o segmento ou na região é a mesma probabilidade de ser escolhidos.

Exemplo 1.16

Suponhamos que nosso experimento agora consiste em escolher, ao acaso, um ponto do círculo de unitário com centro na origem. Acontece que

$$\Omega = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}.$$

Alguns eventos podem ser definidos como interessantes neste experimento, por exemplo:

- $A = \text{“distância entre o ponto escolhido e a origem é } \leq \frac{1}{2}\text{”},$
- $B = \text{“distância entre o ponto escolhido e a origem é } \geq 15\text{”},$
- $C = \text{“Primeira coordenada do ponto escolhido é maior do que a segunda coordenada”}.$

O primeiro problema conhecido relacionado à probabilidade geométrica pode ser encontrado em um manuscrito particular de Isaac Newton¹⁷, anteriormente escrito entre os anos de 1664 a 1666, mas não publicado até o século XX.

Definição 1.17 (*Definição Geométrica de Probabilidade*)

Seja Ω um espaço amostral e $A \subset \Omega$. Definimos probabilidade de ocorrência do evento A como

$$P(A) = \frac{\text{Área de } A}{\text{Área de } \Omega}.$$

A probabilidade geométrica é uma ferramenta para lidar com o problema de resultados infinitos, medindo o número de resultados geometricamente, em termos de comprimento, área ou volume. Na probabilidade clássica, geralmente encontramos problemas finitos, por exemplo, o resultado de um lançamento de dados. No entanto, alguns dos problemas mais interessantes envolvem resultados contínuas, por exemplo, a hora de chegada do seu ônibus. Lidar com resultados contínuas pode ser complicado, mas a probabilidade geométrica fornece uma abordagem útil, permitindo transformar problemas de probabilidade em problemas de geometria.

¹⁷Isaac Newton (1643-1727) foi um astrônomo, alquimista, filósofo natural, teólogo e cientista inglês, mais reconhecido como físico e matemático. Sua obra, *Princípios Matemáticos da Filosofia Natural* é considerada uma das mais influentes na história da ciência.

Acontece que nem todo subconjunto de Ω tem uma área bem definida, o que implica que nem todo evento tem uma probabilidade. Vamos, então, atribuir probabilidades somente aos eventos cuja área estiver bem definida.

Exemplo 1.17 (*Continuação do Exemplo 1.16*)

Ao escolher ao acaso um ponto do círculo unitário com centro na origem temos que $P(A) = \frac{1}{4}$ e $P(B) = 0$. O caso do evento C é mais complexo, mas o resultado é também $P(C) = \frac{1}{2}$.

Neste exemplo, dois eventos têm a mesma probabilidade se, e somente se, eles têm a mesma área, nestas situações definimos probabilidade geométrica.



Figura 1.2: Jogo de dardo.

Exemplo 1.18

O jogo de dardos consiste em lançar um dardo em direção a um alvo, obtendo uma pontuação correspondente ao número atribuído à região na qual o dardo se fixou. Para um jogador novato, parece razoável assumir que a probabilidade de o dardo atingir uma determinada região é proporcional à área da região. Sendo assim, uma região maior apresenta uma maior probabilidade de ser acertada.

Analisando a Figura 1.2, observamos que o alvo tem um raio r e que a distância entre anéis é $r/5$. Supondo que o alvo sempre é atingido temos:

$$P(\text{marcar } i \text{ pontos}) = \frac{\text{Área da região } i}{\text{Área do alvo}}.$$

Por exemplo

$$P(\text{marcar } 1 \text{ ponto}) = \frac{\pi r^2 - \pi(4r/5)^2}{\pi r^2} = 1 - \left(\frac{4}{5}\right)^2.$$

Podemos derivar a fórmula geral e descobrimos que

$$P(\text{marcar } i \text{ pontos}) = \frac{(6-i)^2 - (5-i)^2}{5^2}, \quad i = 1, \dots, 5,$$

independentemente de π e r . A soma das áreas das regiões disjuntas é igual à área do alvo. Portanto, as probabilidades que foram atribuídas aos cinco resultados somam 1.

Nos dois primeiros exemplos utilizam-se figuras similares, isto é, simples coincidência. Qualquer ente geométrico ao qual possamos atribuir alguma medida pode ser utilizado no cálculo de probabilidades. No exemplo 1.19 consideramos uma outra situação.

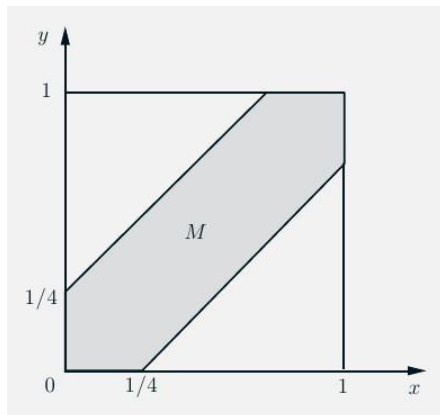


Figura 1.3: Evento M de que Romeu e Julieta se encontram.

Exemplo 1.19

Romeu e Julieta tem um encontro marcado em um dia e hora determinados. Cada um vai chegar no ponto de encontro com um atraso entre 0 e 1 hora, com todos os pares de atrasos sendo igualmente prováveis. O primeiro a chegar vai esperar por 15 minutos e vai sair se o outro ainda não chegou. Qual é a probabilidade de eles se encontrarem?

Definamos o evento de que Romeu e Julieta se encontrem,

$$M = \{(x, y) : |x - y| \leq 1/4, 0 \leq x \leq 1, 0 \leq y \leq 1\}.$$

Como mostrado na Figura 1.3.

A área de M é 1 menos a área dos dois triângulos não sombreados ou $1 - (3/4) \times (3/4) = 7/16$. Assim, a probabilidade de encontro é $7/16$.

Uma vez que a atribuição de uma medida de tais elementos não é um procedimento bastante óbvio, uma série de “paradoxos” tem sido produzidos por incapacidade de distinguir o conjunto de referência. Um destes paradoxos é apresentado no exemplo a seguir, o paradoxo de Bertrand¹⁸.

Exemplo 1.20 (*Paradoxo de Bertrand*)

A corda de um círculo é um segmento de linha geométrica cujos extremos estão sobre o círculo. Imagine que uma corda de um círculo seja desenhado de forma aleatória no círculo unitário. Qual é a probabilidade de que a corda seja maior do que o lado do triângulo equilátero inscrito no círculo?

Apresentamos aqui três das diversas soluções para este problema, dependendo do que interpretamos da frase “de forma aleatória”. Este paradoxo somente será resolvido uma vez definido espaço de probabilidade.

¹⁸Bertrand Arthur William Russell (1872 - 1970) foi um dos mais influentes matemáticos, filósofos e lógicos do Reino Unido que viveram no século XX.

Solução 1 Uma vez que o comprimento da corda é exclusivamente determinada pela posição do seu ponto médio, escolher um ponto C de forma aleatória no círculo e desenhar uma linha através de C e O , o centro do círculo (Figura 1.4). Desenhe a corda por C perpendicular à linha de OC . Se l_1 é o comprimento da corda com C como ponto médio, $l_1 > \sqrt{3}$ se e somente se, C encontra-se no interior do círculo de centro O e raio $1/2$. Assim $P(A) = \pi(\frac{1}{2})^2/\pi = 1/4$. Nesta situação Ω é o círculo com centro em O e raio unitário, o evento de interesse A é o círculo concêntrico com centro em O e raio $1/2$.

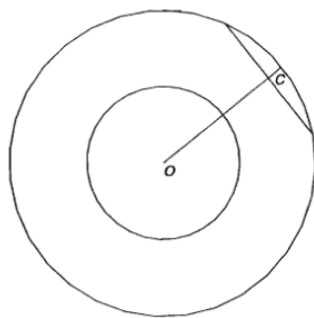


Figura 1.4: Diagrama da solução 1.

Solução 2 Devido a simetria, podemos fixar um ponto final da corda em algum ponto P e, em seguida, escolher o outro ponto final P_1 aleatoriamente. Seja a probabilidade de que o ponto P_1 se encontre num arco de círculo arbitrário ser proporcional ao comprimento desse arco. Agora, o triângulo equilátero inscrito tendo P como um dos seus vértices divide a circunferência em três partes iguais. Uma corda puxada através P será maior do que o lado do triângulo, se e apenas se, o outro ponto final P_1 (Figura 1.5) da corda encontra-se no terço da circunferência que é oposta a P . Segue que a probabilidade requerida é $1/3$. Neste caso $\Omega = [0, 2\pi]$ e $A = [2\pi/2, 4\pi/3]$. Por isso $P(A) = 1/3$.

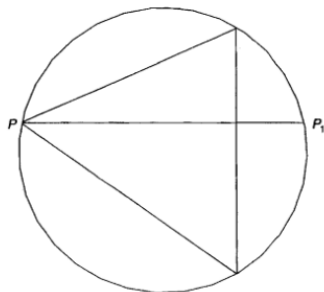


Figura 1.5: Diagrama da solução 2.

Solução 3 Note-se que o comprimento de uma corda é exclusivamente determinada pela distância do seu ponto médio a partir do centro do círculo. Devido à simetria do círculo, assume-se que o ponto médio da corda encontra-se em um raio fixo OM do círculo,

Figura 1.6. A probabilidade de que o centro M situa-se num determinado segmento do raio através de M é então proporcional ao comprimento do segmento. Claramente, o comprimento da corda será mais longo do que o lado do triângulo equilátero se o comprimento de OM é menor do que o raio/2. Segue-se que a probabilidade exigida é $1/2$.

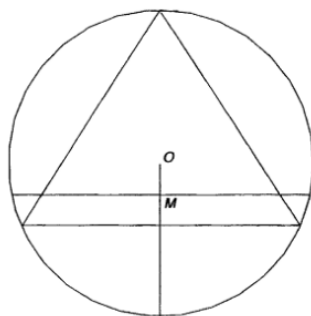


Figura 1.6: Diagrama da solução 3.

1.4 Exercícios

Exercícios da Seção 1.1

- Sejam A, B dois conjuntos quaisquer. Prove que os enunciados a seguir são equivalentes:
 - $A \subseteq B$,
 - $A = A \cap B$,
 - $B = A \cup B$.
- Prove que $A \setminus \emptyset = A$ e que $A \setminus B = A \setminus (A \cap B)$.
- Demonstre que se $A \setminus B = A$, então $A \cap B = \emptyset$.
- Prove que $A \setminus B = \emptyset$ se, e somente se, $A \subseteq B$.
- Sejam A, B conjuntos quaisquer e seja E um conjunto que contenha $A \cup B$. Prove que:
 - $A \setminus B = A \cap (E \setminus B)$.
 - $A \cap (E \setminus A) = \emptyset$.
 - $A \cup (E \setminus A) = E$.
 - $E \setminus (E \setminus A) = A$.
 - $E \setminus \emptyset = E$.
 - $E \setminus E = \emptyset$.
 - $A \subseteq B$ se, e somente se, $E \setminus B \subseteq E \setminus A$.
- Demonstre que se $A \subseteq C$ então $A \cup (B \cap C) = (A \cup B) \cap C$.
- Demonstre que $A \subseteq C$ se, e somente se, $A \cup (B \cap C) = (A \cup B) \cap C$.
- Prove que se $A \neq \emptyset$ então $(A \cup A) \setminus A \neq A \cup (A \setminus A)$.
- Provar que $A \triangle \emptyset = A$ e que $A \triangle A = \emptyset$.
- Prove que para quaisquer dois conjuntos A e C , existe exatamente um conjunto B tal que $A \triangle B = C$. Verifique que este conjunto B é da forma $B = A \triangle C$.
- Verifique que
 - $A \cup B = A \triangle B \triangle (A \cap B)$.
 - $A \setminus B = A \triangle (A \cap B)$.
- Demonstre que $A \triangle B = \emptyset$ se, e somente se, $A = B$.

Exercícios da Seção 1.2

- Um clube deve decidir entre os integrantes A, B, C, D e E qual será o presidente e qual o secretário. Assumindo que um integrante não pode ocupar ambas posições:
 - Descreva o espaço amostral associado à seleção do presidente e secretário;
 - Descreva o evento no qual o integrante A é o secretário.
- Descreva o espaço amostral nos seguintes experimentos:
 - Selecionam-se famílias com três filhos, o sexo dos filhos é registrado em ordem crescente de idade;
 - Um experimento consiste em selecionar 4 itens em uma fábrica e observar se cada item é defeituoso ou não;

- (c) Um livro é aberto em uma página qualquer e conta-se o número de erros de digitação.
3. Sejam A, B, C três eventos arbitrários do espaço amostral Ω .
- Qual é o evento de que somente aconteça A ?
 - Qual é o evento de que ao menos dois dos eventos A, B, C ocorram?
 - Qual é o evento de que ambos A e C ocorram mas B não?
 - Qual é o evento de que ao menos um dos eventos A, B, C ocorram?
4. No jogo de dominós, cada peça é marcada com dois números. As peças são simétricas, de modo que os pares não são ordenados. Quantas diferentes peças podem ser formadas utilizando os números $1, 2, \dots, n$?
5. O conjunto A possui 3 elementos e o conjunto B , 10 elementos. Quantas funções $f : A \rightarrow B$ existem? Quantas delas são injetoras?
6. De quantos modos podemos colocar dois reis diferentes em casas não-adjacentes de um tabuleiro 6×6 ? E se os reis fossem iguais?
7. Cinco moças e cinco rapazes vão posar para uma fotografia, ocupando cinco degraus de uma escadaria, de forma que em cada degrau fique uma moça e um rapaz. De quantas maneiras podemos arrumar este grupo?
8. Em certo estado da federação as placas dos carros mostram quatro números e três letras. Quantos veículos diferentes podem ser licenciados se:
- Os números são colocados depois das letras?
 - Se não houver restrições acerca da ordem em que devem aparecer nas placas os números e as letras?
9. Em um computador digital, um bit é um dos inteiros $\{0, 1\}$ e uma palavra é qualquer sequência de 32 bits. Quantas palavras diferentes são possíveis?
10. Ao organizar pessoas em torno de uma mesa circular, devemos levar em conta os seus lugares em relação uns aos outros e não a posição real de qualquer pessoa. Mostre que n pessoas podem ser organizadas em torno de uma mesa redonda em $(n - 1)!$ maneiras.
11. Um conjunto finito Ω tem n elementos. Mostre que, se contarmos os conjuntos vazio e o próprio Ω como subconjuntos, existem 2^n subconjuntos de Ω .
12. A porta de um centro de informática possui uma trava que tem cinco botões numerados de 1 a 5. A combinação de números que abre a fechadura é uma sequência de cinco números e é repostado a cada semana.
- Quantas combinações são possíveis se cada botão deve ser usado uma vez?
 - Suponha que o bloqueio pode também ter combinações que exigem pressionar duas teclas ao mesmo tempo e depois as outras três, uma de cada vez. Quantas combinações a mais permite que isso?
13. Seja $n \geq 2$ um número natural fixo. Determine la quantidade de pares (x, y) que existem de números naturais x e y tais que
- $$1 \leq x \leq y \leq n.$$
14. Retiram-se 4 cartas, ao acaso, de um baralho de 52 cartas. Registra-se o número de reis na amostra. As retiradas podem ser feitas de duas maneiras:
- sem reposição,
 - com reposição.

Determine em que caso, (a) ou (b), é mais provável obter 4 reis.

15. Suponha que n cartas numeradas de 1 até n sejam embaralhadas e retiradas uma por uma, sem reposição, até todas as cartas serem retiradas. Qual a probabilidade de que para pelo menos uma carta, o número da carta coincida com o número da retirada?
16. Uma empresa de computadores tem quatro candidatos a funcionários, todos igualmente qualificados, dos quais dois são homens e dois são mulheres. A empresa tem de escolher dois candidatos, e ela não discrimina com base no sexo. Se ela escolher dois candidatos aleatoriamente, qual a probabilidade de os dois candidatos escolhidos serem do mesmo sexo? Um aluno responde a essa pergunta da seguinte forma: há três resultados possíveis: duas mulheres, dois homens, uma mulher e um homem. O número de eventos favoráveis é dois. Então a probabilidade é $2/3$. Esse cálculo está correto?
17. Suponha que um grupo de $2n$ adolescentes é dividido em dois subgrupos. Calcule a probabilidade de que dois adolescentes altos estejam em:
 - (a) em grupos diferentes,
 - (b) no mesmo grupo.
18. Qual é a probabilidade de que os aniversários de doze pessoas sejam em meses diferentes? E a probabilidade de que os aniversários de quatro pessoas sejam em dois meses?
19. Três aventureiros devem escolher um deles para uma missão arriscada. Para isso, pegam uma urna com duas bolas brancas e uma bola vermelha, e cada um retira sucessivamente uma bola, sem reposição. Aquele que pegue a bola vermelha será o escolhido para realizar a missão. Mostre que todos têm a mesma probabilidade de ser o escolhido, qualquer que seja a ordem em que realizem as extrações.
20. Se n bolas são colocadas aleatoriamente em n urnas, qual é a probabilidade de que exatamente uma urna permaneça vazia?
21. Um armário contém n pares de sapatos. Se $2r$ sapatos são escolhidos aleatoriamente ($2r < n$), qual é a probabilidade de não haver nenhum par correto na amostra?
22. Duas pessoas lançam, cada uma delas, uma moeda equilibrada n vezes. Qual a probabilidade de que elas obtenham o mesmo número de caras?
23. Noventa alunos, incluindo Joe e Jane, devem ser divididos em três classes de igual tamanho, e isso deve ser feito ao acaso. Qual é a probabilidade de que Joe e Jane acabem na mesma classe?
24. Uma urna contém 3 bolas numeradas de 1 a 3 e outra urna contém 5 bolas numeradas de 1 a 5. Ao retirar-se aleatoriamente uma bola de cada urna, qual é a probabilidade de que a soma dos pontos seja maior do que 4?
25. Um dado equilibrado lança-se 6 vezes consecutivas. Qual é a probabilidade de que:
 - (a) Apareçam as seis caras do dado em ordem crescente ou decrescente?
 - (b) Apareçam as seis caras do dado em qualquer ordem?
 - (c) Somente apareçam números pares?
 - (d) Apareçam números pares e ímpares alternados?
26. Suponha que os motoristas de 8 automóveis estacionam seus carros completamente ao azar num estacionamento de 12 lugares e que o estacionamento é de forma linear. Determine a probabilidade de que os lugares não ocupados sejam adjacentes.
27. Cinco pessoas ficam em um elevador que para a cinco andares. Supondo que cada peso tem igual probabilidade de ir para qualquer andar, qual a probabilidade de que todas elas saiam diferentes andares?
28. **Fofoca** Numa povoado de $n + 1$ habitantes um deles rumora alguma coisa a uma segunda pessoa, esta por sua vez o conta a uma terceira pessoa (que pode ser a primeira pessoa) e assim sucessivamente. Determine a probabilidade de que o rumor se transmita r vezes sem que regresse à primeira pessoa.

29. Um padeiro elabora 100 pães num dia onde 10 deles pesam menos do que deveriam. Um inspetor pesa 5 pães escolhidos ao azar. Calcule a probabilidade de que o inspetor encontre na sua amostra exatamente um pão de peso incorreto.
30. Lançam-se dois dados idênticos e equilibrados ao mesmo tempo. Calcule a probabilidade de que a soma das duas caras seja igual a $2, 3, \dots, 12$. Comprove que a soma de todas essas probabilidades é 1.
31. Uma moeda equilibrada é lançada quatro vezes consecutivas. Calcule a probabilidade de que o número de vezes que se obtém “cara” seja estritamente maior do que o número de vezes que se obtém “coroa”.
32. **Pontos** Suponha que um experimento aleatório têm como espaço amostral o conjunto de pares (x, y) , tais que tanto x quanto y assumem valores em $\{1, 2, \dots, n\}$ e ainda considera-se que quaisquer destes pontos no plano ocorre com mesma probabilidade. Calcule a probabilidade de que ao fazer uma vez o experimento aleatório:
 - a) Se obtenha um ponto na diagonal, isto é, $x = y$;
 - b) Se obtenha um ponto nas bordas, isto é, $x = 1$ ou $x = n$ ou $y = 1$ ou $y = n$;
 - c) Se obtenha um ponto que satisfaça $x \leq y$;
 - d) Se obtenha um ponto tal que $|x - y| \leq 2$.

Exercícios da Seção 1.3

1. Você está visitando São Francisco e está tomando um passeio de bonde a uma loja na Rua do Mercado. Você vai encontrar um amigo na loja às 15:00h. Os bondes partem a cada 10 minutos e a viagem até a loja é de 8 minutos. Você chega na parada de bonde em 14:48h. Qual é a probabilidade de que você vai chegar na loja às 15:00h?
2. Você está esperando um telefonema de um amigo a qualquer momento entre 18:00h e 19:00h. Inesperadamente, você tem que executar uma missão para um parente e se ausentar desde as 17:45 até 18:10. Qual é a probabilidade de você perder o chamado de seu amigo?

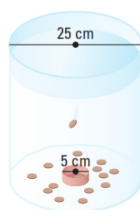
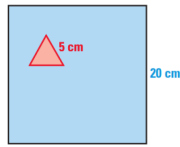
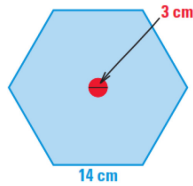


Figura 1.7: Frasco de vidro para arrecadar doações.

3. Imagine que você organiza um evento para levantar fundos em sua escola. Você preenche um grande frasco de vidro que tem um diâmetro de 25 centímetros de água e coloca um prato que tem um diâmetro de 5 centímetros na parte inferior do frasco, como na Figura 1.7. A pessoa doa uma moeda, largando-o na jarra. Se a moeda cai no prato, a pessoa ganha um pequeno prêmio.
 - (a) Calcule a probabilidade de uma moeda cair no prato, tendo a mesma chance de cair em qualquer lugar no fundo do frasco.
 - (b) Suponhamos que, em vez do prato, um círculo com um diâmetro de 5 centímetros é pintado sobre o fundo do frasco e qualquer moeda a tocar o círculo ganha um prêmio. Será que a probabilidade da pessoa ganhar um prêmio muda? Explique.
4. Um alvo quadrado com 20cm de lados inclui uma região triangular com lados iguais, cada um de comprimento 5cm. Um dardo é lançado e atinge o alvo de forma aleatória. Encontre a probabilidade de o dardo atingir o triângulo.



5. Trace as linhas $y = x$ e $y = 3$ em um plano de coordenadas. Um ponto é escolhido aleatoriamente a partir de dentro dos limites $0 < y < 4$ e $0 < x < 4$. Encontre a probabilidade de que as coordenadas do ponto sejam uma solução do sistema de desigualdades $y < 3$ e $y > x$.
6. Um alvo em forma hexagonal regular com lados de 14 centímetros de comprimento tem um olho de boi circular com um diâmetro de três centímetros. Suponha que são lançados dardos para acertar o alvo ao acaso.
- (a) Qual é a probabilidade de que um dardo que atinge o alvo vai bater o olho do boi?
 - (b) Encontre a probabilidade de que um dardo vai bater o olho do boi, se o raio do olho do boi é dobrada.



Capítulo 2

Probabilidade axiomática

Uma das dificuldades em desenvolver uma teoria matemática das probabilidades tem sido a de chegar a uma definição simples e precisa o suficiente para usar em matemática, mas abrangente para ser aplicável a uma vasta gama de fenômenos. A busca de uma definição amplamente aceitável levou quase três séculos e foi marcado por muita controvérsia. A questão foi finalmente resolvida no século 20, tratando a teoria das probabilidades em uma base axiomática.

Em 1933, uma monografia do matemático russo A. Kolmogorov delineou uma abordagem axiomática que forma a base para a moderna teoria¹. Desde então, as ideias foram aperfeiçoadas e um pouco da teoria das probabilidades é agora parte de uma disciplina mais geral conhecida como teoria da medida.

Dedicaremos este capítulo à definição moderna da função de probabilidade ou simplesmente probabilidade. Entendemos por definição moderna a definição axiomática e, portanto, definição matematicamente fundamentada e proposta por A.N.Kolmogorov em 1933.

Para entender a definição moderna de probabilidades devemos compreender primeiro os conceitos de álgebra e σ -álgebra de eventos aleatórios e, ainda, entender uma situação particular destes conjuntos quando aplicados estes conceitos na reta real, a chamada σ -álgebra de Borel.

2.1 Álgebra e σ -álgebra de eventos aleatórios

Vamos supor que a classe dos eventos aleatórios possua certas propriedades básicas intuitivas, as quais serão primordiais para o desenvolvimento posterior da teoria das probabilidades. Se o espaço amostral não é enumerável, muitas vezes não é possível encontrar funções não triviais que meçam todos os conjuntos de Ω , enquanto que a aditividade ainda se mantém. Neste caso, é necessário a medida, pelo menos, ser aplicável a uma grande classe de subconjuntos, esperando que o classe é grande o suficiente e de grande relevância para os conjuntos que surgem na prática. Tal coleção de subconjuntos é chamado de σ -álgebra, e devem satisfazer propriedades fechadas agradáveis.

Lembremos que se dados dois eventos A e B no espaço amostral Ω , dizemos que $A \subset B$

¹A monografia de Kolmogorov, original em russo, está disponível em tradução ao inglês como Fundamentos da Teoria da Probabilidade

se $\omega \in A$ implica que $\omega \in B$. Em palavras, a ocorrência de A implica a ocorrência de B . A união de dois eventos A e B é definida como $A \cup B = \{\omega : \omega \in A \text{ ou } \omega \in B\}$ e representa o evento de que pelo menos um dos dois eventos A ou B ocorrem. A interseção de dois eventos A e B é $A \cap B = \{\omega : \omega \in A \text{ e } \omega \in B\}$ e representa o evento de que ambos A e B ocorrem.

Também, dois eventos A e B são disjuntos ou mutuamente exclusivos se $A \cap B = \emptyset$, sendo \emptyset o evento vazio. Isso significa que A e B não ocorrem simultaneamente. Para qualquer evento A , o complementar de A é $A^c = \{\omega : \omega \notin A\}$ e representa o evento de que A não ocorre. Ainda temos que as operações binárias \cup e \cap satisfazem as leis de distributivas: para quaisquer eventos A , B e C , temos que

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C) \quad \text{e} \quad A \cup (B \cap C) = (A \cup B) \cap (A \cup C).$$

Estas e outras propriedades de eventos foram estudadas na Seção 1.1. Acontece que muitas vezes é conveniente ou mesmo necessário transformar as combinações de eventos em formas alternativas. Nesse sentido propriedades interessantes são as chamadas Leis de Morgan apresentadas no Teorema 1.4. Notamos que (a), no referido teorema, estabelece que o evento de que nenhum A e B ocorre é igual ao complementar do evento de que pelo menos um de A ou B ocorre. Já o item (b), no mesmo teorema, expressa que o complementar do evento de que ambos A e B ocorrem é exatamente o evento de que ao menos um deles não ocorre.

As leis têm o nome de Augustus De Morgan, que introduziu uma versão formal das leis na lógica proposicional clássica. A formulação de De Morgan foi influenciada pela lógica empreendida por George Boole. No entanto, uma observação semelhante foi feita por Aristóteles e era conhecida pelos lógicos gregos e medievais.

Definição 2.1

Seja \mathcal{F} uma classe de eventos aleatórios definidos no espaço amostral Ω , não vazio. Para que \mathcal{F} seja uma álgebra de eventos, deve satisfazer as condições:

- (a) $\Omega \in \mathcal{F}$,
- (b) Se $A \in \mathcal{F}$, então $A^c \in \mathcal{F}$,
- (c) Se $A \in \mathcal{F}$ e $B \in \mathcal{F}$, então $A \cup B \in \mathcal{F}$.

A seguir consideraremos sempre que o espaço amostral Ω é não vazio, a menos que seja explicitado o contrário. Uma questão agora é: num determinado experimento, será que sempre é possível definir uma álgebra de eventos? o seguinte exemplo nos mostra um caminho para chegar à resposta desta questão.

Exemplo 2.1

No caso do lançamento de uma moeda e somente considerando como possíveis resultados $\Omega = \{\text{cara}, \text{coroa}\}$, uma álgebra de eventos seria $\mathcal{F} = \{\emptyset, \Omega, \{\text{cara}\}, \{\text{coroa}\}\}$.

Teorema 2.1

Seja \mathcal{F} uma álgebra de eventos de subconjuntos de Ω . Então valem as seguintes propriedades:

(a) $\emptyset \in \mathcal{F}$,

(b) Para todo n e para todo $A_1, A_2, \dots, A_n \in \mathcal{F}$, temos que

$$\bigcup_{i=1}^n A_i \in \mathcal{F} \quad \text{e} \quad \bigcap_{i=1}^n A_i \in \mathcal{F}.$$

Demonstração: A demonstração de (a) é evidente do item (b) na definição 2.1, se $\Omega \in \mathcal{F}$, então $\Omega^c = \emptyset \in \mathcal{F}$. Pela definição de álgebra de eventos, sabemos que $A_1 \cup A_2 \in \mathcal{F}$, então $(A_1 \cup A_2) \cup A_3 \in \mathcal{F}$. Por indução,

$$\bigcup_{i=1}^n A_i \in \mathcal{F}.$$

Observando que

$$\bigcap_{i=1}^n A_i = \left(\bigcup_{i=1}^n A_i^c \right)^c,$$

segundo o item (b) da Lei de Morgan e, aplicando sucessivamente o item (b) na Definição 2.1, temos que $\bigcap_{i=1}^n A_i \in \mathcal{F}$. ■

Fizemos-no a pergunta de se álgebras de subconjuntos de um espaço amostral não vazio sempre existem. Acontece que, quando Ω é finito sempre é possível definir a álgebra de todas as partes de Ω , isto é, $\mathcal{F} = \mathcal{P}(\Omega)$, este conhecido como conjunto potência. Por exemplo, no Exemplo 2.1 onde $\Omega = \{\text{cara}, \text{coroa}\}$ temos $\mathcal{F} = \mathcal{P}(\Omega) = \{\emptyset, \Omega, \{\text{cara}\}, \{\text{coroa}\}\}$. A classe $\mathcal{P}(\Omega)$ tem $2^2 = 4$ elementos, de modo que há 4 eventos aleatórios associados a este experimento. No caso finito geral, se Ω tem n elementos, $\mathcal{P}(\Omega)$ tem 2^n elementos.

Exemplo 2.2

Se $\Omega = \{1, 2, 3\}$, então $|\mathcal{P}(\Omega)| = 2^{|\Omega|} = 2^3 = 8$, estamos denotando cardinalidade ou número de elementos de um conjunto por $|\cdot|$. Neste caso

$$\mathcal{P}(\Omega) = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \Omega\}.$$

Exemplo 2.3

Seja Ω um espaço amostral. As seguintes classes de subconjuntos de Ω são álgebras:

(a) $\mathcal{F}_1 = \{\emptyset, \Omega\}$,

(b) $\mathcal{F}_2 = \{\emptyset, A, A^c, \Omega\}$, sendo $A \subseteq \Omega$,

- (c) $\mathcal{F}_3 = \{\emptyset, A, A^c, B \setminus A, (B \setminus A)^c, B, B^c, \Omega\}$, sendo $A \subset B$, $A, B \subseteq \Omega$,
 (d) $\mathcal{F}_4 = \mathcal{P}(\Omega)$.

A demonstração que as classes de conjuntos definidos acima são álgebras é um exercício para o leitor.

Quando Ω é finito uma álgebra é uma classe adequada para o cálculo de probabilidades. Isto deve-se a que uma álgebra contém o evento impossível, o evento certo, o evento contrário de qualquer evento que pertença a classe, a união e interseção de eventos que pertençam à classe, isto é, em regra todos os acontecimentos interessantes.

No caso Ω infinito, mesmo que enumerável, uma álgebra deixa de servir para a construção de uma teoria que seja mais forte. Resulta que quando Ω é infinito existem acontecimentos interessantes expressos pela união infinita de outros acontecimentos ou de acontecimentos elementares. Então ao invés de utilizarmos álgebra de eventos, deve-se utilizar σ -álgebra de eventos.

Definição 2.2

Seja \mathcal{F} uma classe de eventos aleatórios definidos no espaço amostral Ω . Para que \mathcal{F} seja uma σ -álgebra de eventos, deve satisfazer as condições:

- (a) $\Omega \in \mathcal{F}$,
 (b) Se $A \in \mathcal{F}$, então $A^c \in \mathcal{F}$,
 (c) Se $A_i \in \mathcal{F}$, para $i = 1, 2, \dots$, então $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

Uma σ -álgebra é sempre uma álgebra. Seja \mathcal{F} uma σ -álgebra, então se $A, B \in \mathcal{F}$

$$A \cup B = A \cup B \cup B \cup \dots \in \mathcal{F},$$

logo \mathcal{F} é álgebra. O contrário não é verdade, nem toda álgebra é σ -álgebra.

Exemplo 2.4

Vamos considerar outro exemplo importante aqui, o espaço amostral associado a um número infinito de jogadas de uma moeda. Seja

$$\Omega = \{\omega = (\omega_1, \omega_2, \dots) : \omega_j = 0 \text{ ou } 1\}.$$

Pensamos em 0 como resultado coroa e 1 como cara. Para cada inteiro positivo n , seja

$$\Omega_n = \{(\omega_1, \omega_2, \dots, \omega_n) : \omega_j = 0 \text{ ou } 1\}.$$

Cada Ω_n é um conjunto finito de 2^n elementos. Denotemos por \mathcal{F}_n as σ -álgebras consistindo de todos os eventos que dependem apenas dos primeiros n lançamentos. Mais

formalmente, definimos \mathcal{F}_n como a coleção de todos os subconjuntos de $A \subset \Omega$, tais que exista um $E \in \mathcal{P}(\Omega)$ com

$$A = \{(\omega_1, \omega_2, \dots) : (\omega_1, \omega_2, \dots, \omega_n) \in E\}.$$

Observemos que \mathcal{F}_n é uma σ -álgebra finita contendo 2^{2^n} subconjuntos e $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \mathcal{F}_3 \subset \dots$. Definamos

$$\mathcal{F}_0 = \bigcup_{n=1}^{\infty} \mathcal{F}_n.$$

Mostrar que \mathcal{F}_0 é álgebra mas não é σ -álgebra fica como exercício ao leitor.

Teorema 2.2

Seja \mathcal{F} uma σ -álgebra definida em Ω . Se $A_1, A_2, \dots \in \mathcal{F}$, então $\bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$.

Demonstração :

$$\bigcap_{i=1}^{\infty} A_i \in \mathcal{F} = \left(\bigcup_{i=1}^{\infty} A_i \right)^c.$$

■

Exemplo 2.5

Seja Ω um espaço amostral não enumerável e seja \mathcal{F} uma classe que consiste de todos os subconjuntos enumeráveis de Ω e de todos os subconjuntos de Ω cujos complementares são enumeráveis. Então \mathcal{F} é uma σ -álgebra.

Primeiro observemos que Ω^c é enumerável (o vazio é enumerável, contém zero elementos), então $\Omega \in \mathcal{F}$. Seja $A \in \mathcal{F}$, se A é enumerável o complementar de A^c é enumerável, logo $A^c \in \mathcal{F}$. Caso A não seja enumerável então A^c é enumerável e, portanto, $A^c \in \mathcal{F}$.

Consideremos $A_1, A_2, \dots \in \mathcal{F}$. Caso todos os A_i sejam enumeráveis, então

$$\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}.$$

Sejam A_{i_1}, A_{i_2}, \dots conjuntos não enumeráveis de \mathcal{F} . Pelo Teorema 1.4 temos que

$$\left(\bigcup_k A_{i_k} \right)^c = \bigcap_k A_{i_k}^c.$$

Observemos que cada $A_{i_1}^c, A_{i_2}^c, \dots$ é enumerável, então

$$\bigcap_k A_{i_k}^c$$

é enumerável e, portanto,

$$\bigcap_k A_{i_k}^c \in \mathcal{F}.$$

Teorema 2.3

Seja \mathcal{F} uma σ -álgebra definida em Ω . Se $A, B \in \mathcal{F}$, então $A \setminus B \in \mathcal{F}$ e $A \Delta B \in \mathcal{F}$.

Demonstração: Exercício. ■

Devemos observar que a união de σ -álgebras não é necessariamente uma σ -álgebra. Vejamos isso no seguinte exemplo.

Exemplo 2.6

Consideremos o espaço amostral definido como $\Omega = \{1, 2, 3\}$ e, definidas em Ω , sejam \mathcal{F}_1 e \mathcal{F}_2 duas σ -álgebras definidas como $\mathcal{F}_1 = \{\emptyset, \{1\}, \{2, 3\}, \Omega\}$ e $\mathcal{F}_2 = \{\emptyset, \{1, 2\}, \{3\}, \Omega\}$.

Observemos que

$$\mathcal{F}_1 \cup \mathcal{F}_2 = \{\emptyset, \{1\}, \{1, 2\}, \{2, 3\}, \{3\}, \Omega\},$$

e que o subconjunto $\{2, 3\} \setminus \{3\} = \{2\} \notin \mathcal{F}_1 \cup \mathcal{F}_2$.

No caso Ω finito $\mathcal{P}(\Omega)$, o conjunto das partes de Ω ou conjunto potência de Ω , é uma σ -álgebra. No caso Ω infinito, em especial contínuo, a construção de uma σ -álgebra de eventos Ω é mais complexa. Este assunto será objeto de estudo na Seção 2.1.1.

Definição 2.3

Uma σ -álgebra \mathcal{F} definida em Ω é mínima em relação a todas as σ -álgebras que contém a classe de conjuntos \mathcal{C} se satisfaz:

(a) $\mathcal{C} \subseteq \mathcal{F}$,

(b) Se \mathcal{G} é uma outra σ -álgebra definida em Ω e $\mathcal{C} \subseteq \mathcal{G}$, então $\mathcal{F} \subseteq \mathcal{G}$.

Exemplo 2.7

Seja Ω um espaço amostral enumerável. Nestas condições sempre podemos construir duas σ -álgebras

$$\mathcal{F}_0 = \{\emptyset, \Omega\}$$

e

$$\mathcal{F}_1 = \mathcal{P}(\Omega).$$

Observemos que se \mathcal{F} é uma outra σ -álgebra definida em Ω , se satisfaz que $\mathcal{F}_0 \subset \mathcal{F} \subset \mathcal{F}_1$. Assim, podemos dizer que \mathcal{F}_0 é a menor σ -álgebra possível e \mathcal{F}_1 é a maior σ -álgebra possível.

Devemos esclarecer que a menor σ -álgebra possível \mathcal{F}_0 sempre existe, mais isso não significa que \mathcal{F}_0 seja sempre a σ -álgebra mínima. Observe que o conceito de σ -álgebra mínima depende da classe de conjuntos \mathcal{C} .

Definição 2.4 (σ -álgebra gerada)

Seja Ω um espaço amostral e \mathcal{C} uma classe de subconjuntos de Ω . Dizemos que $\sigma(\mathcal{C})$ é uma σ -álgebra gerada por \mathcal{C} se $\sigma(\mathcal{C})$ contém \mathcal{C} e é mínima em relação a todas as σ -álgebras que contém \mathcal{C} .

Exemplo 2.8

Sejam $A, B \subseteq \Omega$ tais que $A \cap B = \emptyset$. Definamos a classe $\mathcal{C} = \{A, B\}$. A menor σ -álgebra que contém a classe \mathcal{C} ou σ -álgebra mínima que contém \mathcal{C} é

$$\sigma(\mathcal{C}) = \{\emptyset, A, B, (A \cup B), (A \cup B)^c, A^c, B^c, \Omega\},$$

e, portanto, é a σ -álgebra gerada por \mathcal{C} .

Teorema 2.4

Sejam \mathcal{C}_1 e \mathcal{C}_2 duas classes de subconjuntos de Ω tais que $\mathcal{C}_1 \subseteq \mathcal{C}_2$. Então

$$\sigma(\mathcal{C}_1) \subseteq \sigma(\mathcal{C}_2).$$

Demonstração: Sabemos que $\mathcal{C}_1 \subseteq \mathcal{C}_2 \subseteq \sigma(\mathcal{C}_2)$. Então $\sigma(\mathcal{C}_2)$ é uma σ -álgebra que contém \mathcal{C}_1 , portanto $\sigma(\mathcal{C}_1) \subseteq \sigma(\mathcal{C}_2)$. ■

Exemplo 2.9

Sejam $A, B, C \subseteq \Omega$ disjuntos dois a dois. Definamos as classes $\mathcal{C}_1 = \{A, B\}$ e $\mathcal{C}_2 = \{A, B, C\}$. Do Exemplo 2.8 temos que

$$\sigma(\mathcal{C}_1) = \{\emptyset, A, B, (A \cup B), (A \cup B)^c, A^c, B^c, \Omega\}.$$

É claro que $\mathcal{C}_1 \subseteq \mathcal{C}_2$ e

$$\sigma(\mathcal{C}_2) = \{\emptyset, A, B, C, (A \cup B), (A \cup C), (B \cup C), (A \cup B \cup C), (A \cup B)^c, (A \cup C)^c, (B \cup C)^c, (A \cup B \cup C)^c, A^c, B^c, C^c, \Omega\},$$

portanto, $\sigma(\mathcal{C}_1) \subseteq \sigma(\mathcal{C}_2)$.

Teorema 2.5

Se \mathcal{C} é uma σ -álgebra, então $\sigma(\mathcal{C}) = \mathcal{C}$.

Demonstração: Sabemos que $\mathcal{C} \subseteq \sigma(\mathcal{C})$. Dado que \mathcal{C} é σ -álgebra, então $\sigma(\mathcal{C}) \subseteq \mathcal{C}$, o que demonstra a igualdade. ■

A interseção de σ -álgebras é uma σ -álgebra (ver Exercício 17), isto permite definir de maneira diferente σ -álgebra gerada.

Teorema 2.6

Seja \mathcal{C} uma classe de subconjuntos de Ω . A σ -álgebra gerada por \mathcal{C} é

$$\sigma(\mathcal{C}) = \bigcap \{ \mathcal{F} : \mathcal{F} \text{ é uma } \sigma\text{-álgebra e } \mathcal{C} \subseteq \mathcal{F} \}. \quad (2.1)$$

Demonstração: A σ -álgebra definida em (2.1) contém a coleção \mathcal{C} e se \mathcal{G} for uma σ -álgebra que contenha \mathcal{C} , então $\sigma(\mathcal{C}) \subseteq \mathcal{G}$. Isto mostra que $\sigma(\mathcal{C})$ é uma σ -álgebra mínima em relação a todas as σ -álgebras que contém a classe de conjuntos \mathcal{C} e, portanto, a σ -álgebra em (2.1) é uma σ -álgebra gerada por \mathcal{C} segundo a Definição 2.4. A conclusão da demonstração é exercício para o leitor. ■

2.1.1 σ -álgebra de Borel

Quando o espaço amostral é um conjunto finito de n elementos ou infinito enumerável podemos considerar às σ -álgebras consistindo de todos os seus subconjuntos, com as respectivas potências 2^n ou $2^{\mathbb{N}}$. No entanto, quando o espaço amostral é o conjunto de números reais ou um intervalo nos reais, esta escolha levanta uma série de problemas técnicos. Neste caso, gostaríamos que todos os subconjuntos de um único ponto de Ω e todos os intervalos fechados, abertos ou semi-fechados fossem eventos.

Os problemas técnicos aos quais fizemos referência serão resolvidos utilizando uma σ -álgebra especial nos números reais, conhecida como σ -álgebra de Borel². Excelentes livros deste tema são Halmos (1950), Kolmogorov & Fomin (1961) e Royden (1988), dentre outros.

Recordemos que um conjunto $E \subseteq \mathbb{R}$, onde \mathbb{R} denota o conjunto dos números reais, é dito ser um conjunto aberto se para todo elemento $x \in E$ existe algum $\epsilon > 0$, que depende de x , de maneira que o intervalo $(x - \epsilon, x + \epsilon)$ esteja contido em E . Recordemos também que todo intervalo real da forma (a, b) , $-\infty < a < b < +\infty$ é um conjunto aberto de \mathbb{R} . Um conjunto $F \subseteq \mathbb{R}$ é dito fechado se F^c for um conjunto aberto. Observemos que ambos \mathbb{R} e \emptyset são, simultaneamente, conjuntos abertos e fechados.

Chamemos \mathcal{O} a coleção de todos os conjuntos abertos de \mathbb{R} , não é difícil perceber que \mathcal{O} não é uma σ -álgebra de subconjuntos de \mathbb{R} . Observe que se $A \in \mathcal{O}$, então A é um conjunto aberto e, por definição, A^c é um conjunto fechado, logo $A^c \notin \mathcal{O}$. Entretanto $\sigma(\mathcal{O})$, a σ -álgebra gerada por \mathcal{O} , existe e satisfaz que $\mathcal{O} \subset \sigma(\mathcal{O})$. Isto nos permite a seguinte definição.

²Emile Borel (1871-1953), matemático francês. Um dos principais contribuidores da atual teoria da medida, da qual a teoria das probabilidades moderna é uma situação particular independente.

Definição 2.5 (σ -álgebra de Borel)

Sejam $\Omega = \mathbb{R}$ e \mathcal{O} a classe dos conjuntos abertos em \mathbb{R} . Define-se a σ -álgebra de Borel sobre \mathbb{R} , denotada por $\mathcal{B}(\mathbb{R})$, como

$$\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{O}).$$

Dado que $\mathcal{B}(\mathbb{R})$ é uma σ -álgebra, vemos que necessariamente contém todos os conjuntos abertos, todos os conjuntos fechados, todas as uniões de conjuntos abertos, todas as uniões de conjuntos fechados, todas as interseções de conjuntos fechados e todas as interseções de conjuntos abertos. Os elementos de $\mathcal{B}(\mathbb{R})$ são chamados de Borelianos na reta e, em termos intuitivos, um Boreliano é um conjunto que pode ser obtido de um número enumerável de conjuntos abertos aplicando-se as operações \cap e \cup , assim como o complemento um número enumerável de vezes.

O seguinte teorema caracteriza os conjuntos reais abertos e será de grande utilidade.

Teorema 2.7

Seja $E \subseteq \mathbb{R}$ um conjunto aberto. Então existe, pelo menos, uma quantidade enumerável de intervalos abertos disjuntos I_k , $k = 1, 2, \dots$, tais que

$$E = \bigcup_{k=1}^{\infty} I_k.$$

Demonstração: A ideia é definir a relação de equivalência em E da seguinte maneira. Se $a, b \in E$, dizemos que a é equivalente a b , escrevemos $a \sim b$, se o intervalo aberto (a, b) estiver contido em E . Esta relação de equivalência particiona E na união de conjuntos disjuntos. Não sabemos a priori se existe uma quantidade enumerável destes conjuntos. Assim sendo, denotemos estes conjuntos por I_k , $k \in J$, onde J é um conjunto de índices qualquer. Observemos que I_k , de fato, é um intervalo pelo seguinte motivo: se $a_k, b_k \in I_k$, então $a_k \sim b_k$ de modo que o intervalo aberto (a_k, b_k) está contido em I_k . dado que a_k e b_k são arbitrários vemos que I_k é um intervalo. O próximo passo é provar que I_k é um conjunto aberto. Seja $x \in I_k$, um elemento arbitrário. Dado que $x \in E$ e E é aberto, sabemos existe um $\epsilon > 0$ de maneira que $(x - \epsilon, x + \epsilon) \subseteq E$. Contudo, temos claramente que $a \sim x$ para todo $a \in (x - \epsilon, x + \epsilon)$, o qual implica que a ϵ -vizinhança de x está contida em I_k . Portanto, I_k é um conjunto aberto, como requerido. Por último, vamos mostrar que há, no máximo, um número enumerável de I_k . Isto resulta do fato de que, cada I_k deve conter pelo menos um número racional. Uma vez que existem enumeráveis números racionais, pode haver no máximo enumeráveis intervalos I_k . ■

Em momento algum dizemos que $\mathcal{B}(\mathbb{R})$ seja o conjunto das partes dos reais ou, em outras palavras, nós não provamos que não existam conjuntos reais que não sejam Borelianos.

Exemplos de conjuntos reais não Borelianos podem ser encontrados em Royden (1988). De fato, não existe um procedimento simples para determinar quando um dado conjunto $A \subseteq \mathbb{R}$ é um Boreliano ou não. No entanto, uma maneira para entendermos melhor a σ -álgebra de Borel $\mathcal{B}(\mathbb{R})$ é demonstrando que ela pode ser gerada pelos intervalos da forma $(-\infty, a]$, isso será provado no Teorema 2.9.

Antes disso apresentamos um resultado auxiliar.

Teorema 2.8

Sejam $x, y \in \mathbb{R}$ com $x < y$. Então, podemos escrever:

$$1- (x, y] = (-\infty, y] \setminus (-\infty, x],$$

$$2- \{x\} = \bigcap_{n=1}^{\infty} \left(x - \frac{1}{n}, x \right],$$

$$3- (x, y) = (x, y] \setminus \{y\},$$

$$4- [x, y] = (x, y] \cup \{x\},$$

$$5- [x, y) = \{x\} \cup (x, y] \setminus \{y\},$$

$$6- (x, +\infty) = (-\infty, x]^c.$$

Demonstração: Exercício. ■

Exemplo 2.10

Observemos que, segundo os resultados apresentados no Teorema 2.8, para escrever o intervalo aberto $(x, +\infty)$ podemos utilizar uma sequência de intervalos com extremos direitos divergentes.

Assim, podemos escrever que

$$(x, +\infty) = \bigcup_{n=0}^{\infty} (x, x+n).$$

Significa que qualquer seja $y \in \mathbb{R}$, sendo $y > x$, então

$$y \in \bigcup_{n=0}^{\infty} (x, x+n).$$

Isto deve-se à densidade dos números reais, ou seja, quaisquer sejam os números reais x e y , $x < y$ sempre existirá um número natural n de maneira que $y \in (x, x+n)$.

As relações apresentadas no Teorema 2.8 serão frequentemente usadas nas demonstrações de resultados, isso devido a que o referido teorema nos explica como escrever intervalos reais como união, interseção e demais operações de coleções enumeráveis de intervalos.

A seguir demonstramos um teorema muito importante porque nos mostra uma forma de encontrar elementos de $\mathcal{B}(\mathbb{R})$.

Teorema 2.9

Seja \mathcal{I} a classe dos intervalos $\mathcal{I} = \{(-\infty, a] : a \in \mathbb{Q}\}$, \mathbb{Q} é o conjunto dos números racionais. Então, a σ -álgebra de Borel $\mathcal{B}(\mathbb{R})$ é gerada por \mathcal{I} , isto é,

$$\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{I}).$$

Demonstração: Denotemos por \mathcal{O} a classe todos os intervalos abertos. Dado que todo conjunto aberto em \mathbb{R} é, pelo menos, união enumerável de intervalos abertos (ver Teorema 2.7), nós devemos ter que $\sigma(\mathcal{O}) = \mathcal{B}(\mathbb{R})$. Denotemos por \mathcal{I} a coleção de todos os intervalos da forma $(-\infty, a]$, $a \in \mathbb{Q}$. Seja $(a, b) \in \mathcal{O}$ para algum $b > a$, com $a, b \in \mathbb{Q}$. Seja $a_n = a + \frac{1}{n}$, de modo que $a_n \downarrow a$ quando $n \rightarrow \infty$ e $b_n = b - \frac{1}{n}$, de modo que $b_n \uparrow b$ quando $n \rightarrow \infty$. Portanto

$$(a, b) = \bigcup_{n=1}^{\infty} (a_n, b_n] = \bigcup_{n=1}^{\infty} \{(-\infty, b_n] \cap (-\infty, a_n]^c\},$$

o qual implica que $(a, b) \in \sigma(\mathcal{I})$. Isto é, $\mathcal{O} \subseteq \sigma(\mathcal{I})$ de forma que $\sigma(\mathcal{O}) \subseteq \sigma(\mathcal{I})$. Contudo, todo elemento de \mathcal{I} é um conjunto fechado, logo

$$\sigma(\mathcal{I}) \subseteq \mathcal{B}(\mathbb{R}).$$

Isto dá a cadeia de contenções

$$\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{O}) \subseteq \sigma(\mathcal{I}) \subseteq \mathcal{B}(\mathbb{R}),$$

de forma que $\sigma(\mathcal{I}) = \mathcal{B}(\mathbb{R})$, provando o resultado. ■

Teorema 2.10

Sejam x, y números reais tais que $x < y$. Os intervalos

$$(x, +\infty), (x, y), [x, y], [x, y), (x, y], (x)$$

são todos elementos de $\mathcal{B}(\mathbb{R})$.

Demonstração: Exercício. ■

Segundo este teorema uniões e intercepções de quantidades enumeráveis de quaisquer dos intervalos mencionados assim como complementos dos conjuntos resultantes são todos Borelianos. Isto pode ser utilizado para verificar os seguintes resultados.

Exemplo 2.11

O conjunto \mathbb{Q} , dos números racionais, é Boreliano por ser união enumerável de intervalos degenerados, ou seja, pontos. Devemos lembrar que o conjunto dos números racionais é enumerável. O conjunto dos números irracionais \mathbb{I} também é Boreliano, pois é complementar de união enumerável. Aqui estamos considerando que

$$\mathbb{Q} \cap \mathbb{I} = \emptyset \quad \text{e} \quad \mathbb{Q} \cup \mathbb{I} = \mathbb{R}.$$

De forma equivalente, pode definir-se $\mathcal{B}(\mathbb{R})$ como a σ -álgebra mínima gerada por todos os subconjuntos abertos de \mathbb{R} . Em ambos os casos, a σ -álgebra gerada é $\mathcal{B}(\mathbb{R})$. Além da Definição 2.5 e do comentário anterior, existem outras formas equivalentes de gerar a σ -álgebra de Borel. Esse é o conteúdo do seguinte resultado.

Teorema 2.11

Estas σ -álgebras são todas idênticas a $\mathcal{B}(\mathbb{R})$.

1- $\sigma(\{(x, y) : x \leq y\})$.

4- $\sigma(\{[x, y) : x \leq y\})$.

2- $\sigma(\{[x, y] : x \leq y\})$.

5- $\sigma(\{[x, +\infty) : x \in \mathbb{R}\})$.

3- $\sigma(\{(x, y] : x \leq y\})$.

Demonstração: Apresentamos somente a demonstração da primeira situação, as outras situações demonstram-se de maneira similar. Para demonstrar que $\mathcal{B}(\mathbb{R}) = \sigma(\{(x, y) : x \leq y\})$ verificamos as duas sentenças:

\subseteq Primeiro $(x, y) \in \mathcal{B}(\mathbb{R})$, então $\{(x, y) : x \leq y\} \subseteq \mathcal{B}(\mathbb{R})$, portanto $\sigma(\{(x, y) : x \leq y\}) \subseteq \mathcal{B}(\mathbb{R})$.

\supseteq Sabemos que $(-\infty, y] \in \sigma(\{(x, y) : x \leq y\})$, devido a que

$$(-\infty, y] = (\infty, y) \cup \{y\} \quad \text{e} \quad (-\infty, y) = \bigcup_{n=1}^{\infty} (x - n, y).$$

Então $\{(-\infty, y] : y \in \mathbb{R}\} \subseteq \sigma(\{(x, y) : x \leq y\})$. Portanto $\mathcal{B}(\mathbb{R}) \subseteq \sigma(\{(x, y) : x \leq y\})$. ■

É natural perguntar-se se à coleção $\mathcal{B}(\mathbb{R})$ pertencem todos os subconjuntos de \mathbb{R} , pergunta que nos fizemos atrás. Mesmo com todos os esclarecimentos acerca dos conjuntos que pertencem a σ álgebra de Borel a resposta é negativa, ou seja, pode-se demonstrar que existem subconjuntos dos números reais que não pertencem a $\mathcal{B}(\mathbb{R})$. A construção de tais conjuntos não é simples, alguns exemplos podem ser consultados em Halmos (1950) e Cohn (1980). Nós apresentamos aqui dois exemplos de subconjuntos de números reais que não são borelianos. O primeiro destes exemplos deve-se à Cantor³

³Georg Ferdinand Ludwig Philipp Cantor (1845 - 1918) foi um matemático alemão. Ele criou a teoria

O conjunto do Cantor é um conjunto de pontos situados em um segmento de linha. Ele é criado tomando algum intervalo, por exemplo $[0, 1]$, removendo o terço médio $(\frac{1}{3}, \frac{2}{3})$, removendo o terço médio de cada uma das duas seções restantes $(\frac{1}{9}, \frac{2}{9})$ e $(\frac{7}{9}, \frac{8}{9})$, em seguida removendo o terço médio das quatro seções restantes e assim por diante, ao infinito. Ver a Figura 2.1.

É um conjunto fechado consistindo inteiramente de pontos de fronteira e é um contra exemplo importante na teoria dos conjuntos. Ao aprendermos sobre a cardinalidade, primeiro são mostrados os subintervalos dos números reais, como exemplos de conjuntos não enumeráveis infinitos. Isso pode sugerir que qualquer subconjunto infinito não enumerável deve conter um intervalo; tal afirmação é falsa. Um contra exemplo para essa afirmação é o conjunto Cantor, que não é enumerável, apesar de não conter nenhum intervalo. No Exemplo 2.12 mostramos a construção deste conjunto e propriedades, nossa principal referência é o artigo de Stettin (2019).

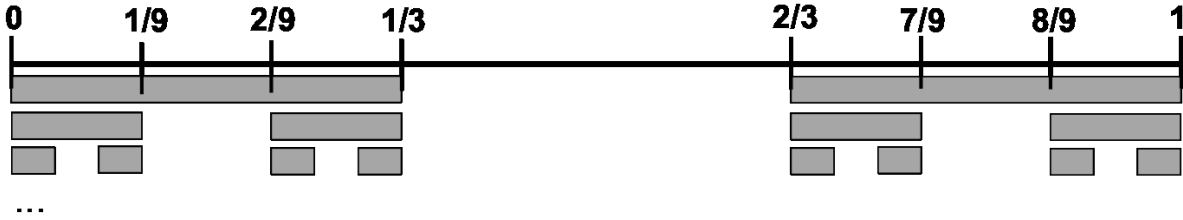


Figura 2.1: Construção do conjunto de Cantor.

Para entendermos a construção do conjunto de Cantor vamos introduzir o conceito de expansões base 3, também chamadas de expansões ternárias, representam números decimais ao usar os dígitos 0, 1, 2. Por exemplo, o número 3 em decimal é 10 em base 3. Frações na base 3 são representadas em termos de potências negativas de 3, onde uma fração N em decimal pode ser representada como

$$N = \frac{c}{b} + \frac{c_2}{b^2} + \frac{c_3}{b^3} + \cdots + \frac{c_n}{b^n} + \cdots, \quad (2.2)$$

onde $0 \leq c_i < 3$ e a fração é escrita em base 3 como $N = 0.c_1c_2c_3 \cdots$. Por exemplo

$$\frac{1}{6} = 0.166666 \cdots = \frac{0}{3} + \frac{1}{3^2} + \frac{1}{3^3} + \frac{1}{3^4} + \frac{1}{3^5} + \frac{1}{3^6} + \frac{1}{3^7} + \cdots = 0.0111111_3 = 0.0\bar{1}_3.$$

Da mesma forma, as frações decimais podem ser convertidas pela multiplicação por 3, obtendo o resultado módulo 3, multiplicando o restante por 3 e, em seguida, tomando o módulo 3

dos conjuntos, que se tornou uma teoria fundamental na matemática. Cantor estabeleceu a importância da correspondência um-a-um entre os membros de dois conjuntos, definiu conjuntos infinitos e bem ordenados, e provou que os números reais são mais numerosos que os números naturais. De fato, o método de prova de Cantor deste teorema implica a existência de uma infinidade de infinitos. Ele definiu os números cardinais e ordinais e sua aritmética.

desse, etc. até que o resto seja 0:

$$\begin{aligned}
 0.166666 \times 3 &= 0 + 0.5, & c_1 &= 0 \\
 0.5 \times 3 &= 1 + 0.5, & c_2 &= 1 \\
 0.5 \times 3 &= 1 + 0.5, & c_2 &= 1 \\
 &\vdots \\
 &= 0.01\bar{1}_3
 \end{aligned} \tag{2.3}$$

Na base 3, alguns pontos possuem mais de uma notação: $\frac{1}{3} \times 3 = 1 + 0$, de maneira que $c_1 = 1$ e $\frac{1}{3} = 0.1_3$. Mas também temos o seguinte:

$$\begin{aligned}
 \frac{1}{3} \times 3 &= 0.3333 \dots \times 3 = 0 + 0.9999 \dots, & c_1 &= 0 \\
 &= 0.9999 \dots \times 3 = 2 + 0.9999 \dots, & c_2 &= 2 \\
 &= 0.9999 \dots \times 3 = 2 + 0.9999 \dots, & c_2 &= 2 \\
 &\vdots \\
 &= 0.0\bar{2}_3
 \end{aligned} \tag{2.4}$$

Desta maneira $\frac{1}{3} = 0.\bar{3} = 0.1_3 = 0.0\bar{2}_3$. Isso às vezes significa que há ambiguidade, um mesmo número pode ser escrito de diversas formas.

Exemplo 2.12 (*Conjunto de Cantor*)

O conjunto de Cantor é construído removendo subintervalos cada vez menores de $[0, 1]$. Na primeira etapa, remova $(\frac{1}{3}, \frac{2}{3})$ de $[0, 1]$. Na segunda etapa remova $(\frac{1}{9}, \frac{2}{9}) \cup (\frac{7}{9}, \frac{8}{9})$ do que resta após o primeiro passo. Como mostrado na Figura 2.1. Em geral, na n -ésima etapa, remova

$$\left(\frac{1}{3^n}, \frac{2}{3^n}\right) \cup \left(\frac{4}{3^n}, \frac{5}{3^n}\right) \cup \dots \cup \left(\frac{3^n - 2}{3^n}, \frac{3^n - 1}{3^n}\right)$$

do que resta após o $(n - 1)$ -ésimo passo. Continuando dessa maneira, obtém-se uma coleção infinita de conjuntos $\{\mathcal{C}_n\}_{n=0}^{\infty}$ tais que $\mathcal{C}_n \subset \mathcal{C}_{n-1}$, para todo $n \geq 1$. Depois de todos os \mathbb{N} passos terem sido dados o que resta é o conjunto do Cantor \mathcal{C} . Então, define-se o conjunto Cantor como sendo

$$\mathcal{C} = \bigcap_{n=0}^{\infty} \mathcal{C}_n. \tag{2.5}$$

Há uma caracterização alternativa que será útil para provar algumas propriedades do conjunto de Cantor. Primeiro apresentamos a caracterização do conjunto de Cantor \mathcal{C} e depois duas propriedades que nos permitirão entender que este conjunto e também seus subconjuntos não pertencem a $\mathcal{B}(\mathbb{R})$.

- O conjunto de Cantor consiste precisamente dos números reais em que as expansões de base 3 contêm apenas os dígitos 0 e 2.

Demonstração: Para provar isto suponhamos que $x \in [0, 1]$ contenha somente os dígitos 0 e 2 na sua expansão em base 3. Seja x_n o truncamento de x em n locais após o ponto

decimal. Por exemplo, se $x = 0.020202 \dots_3$, então $x_1 = 0$, $x_2 = 0.02_3$, $x_4 = 0.0202_3$, etc. Certamente, a sequência $\{x_n\}$ converge a x quando $n \rightarrow \infty$. Em particular, para todo $n \geq 1$, temos que

$$x_n \leq x \leq x_n + \frac{1}{3^n}.$$

Note que os números em $[0, 1]$, cujas expansões em base 3 continuam por exatamente n dígitos após o ponto decimal e que usam apenas os dígitos 0 e 2 são precisamente os pontos finais da esquerda dos intervalos cuja união é \mathcal{C}_n . Assim, o intervalo $[x_n, x_n + 1/3^n]$ está contido em \mathcal{C}_n . Segue-se que $x \in \mathcal{C}_n$ para todos $n \geq 0$ e, portanto $x \in \mathcal{C}$. Por outro lado, suponha $x \in \mathcal{C}$. Então $x \in \mathcal{C}_n$ para todo $n \geq 0$. Observe que os números em \mathcal{C}_n são precisamente aqueles cujo truncamento n -ésimo, ou seja, o número obtido tomando apenas os primeiros n dígitos após o ponto decimal usa apenas 0 e 2 como dígitos na base 3. Segue-se que todo truncamento de x usa apenas 0 e 2 como dígitos. Isso implica x usar apenas 0 e 2 em sua expansão de base 3. ■

- \mathcal{C} não contém nenhum subintervalo de $[0, 1]$.

Demonstração : Seja $[a, b] \subset [0, 1]$ um intervalo arbitrário. Vamos escrever os extremos do intervalo como $a = 0.a_1a_2a_3, \dots_3$ e $b = 0.b_1b_2b_3, \dots_3$. Se algum a_i é igual a 1, então pelo resultado anterior, sabemos que $a \notin \mathcal{C}$. Similarmente, se algum b_i é igual a 1, sabemos que $b \notin \mathcal{C}$. Caso contrário, suponha que k seja o menor índice tal que $a_k \neq b_k$. Estas conclusões forçam que $a_k = 0$ e $b_k = 2$, de maneira que $0.a_1a_2a_3 \dots a_{k-1}1_3 \in [a, b]$ e disto segue que $[a, b] \not\subset \mathcal{C}$. ■

- \mathcal{C} é não enumerável.

Demonstração : Definamos a função $f : \mathcal{C} \rightarrow [0, 1]$ como segue. Se $x = 0.x_1x_2x_3 \dots_3$ é um elemento em \mathcal{C} , seja

$$f(x) = 0.(x_1/2)(x_2/2)(x_3/2) \dots_2.$$

Em outras palavras, na expansão ternária de x , substitua cada dígito 2 com 1 e considere o resultado como um número binário. Agora, suponhamos que \mathcal{C} seja enumerável. Vamos escrever então $\mathcal{C} = \{c_1, c_2, c_3, \dots\}$. Para cada $y \in [0, 1]$, seja $g(y)$ igual ao índice j do elemento c_j com $f(c_j) = y$. Então, podemos construir a sequência $\{d_i\}_{i \in \mathbb{N}}$ tal que $d_{g(y)} = y$ para todo $y \in [0, 1]$. Isso constitui uma bijeção entre $[0, 1]$ e \mathbb{N} , então segue-se que $[0, 1]$ é enumerável. O qual é uma contradição. Portanto, concluímos que \mathcal{C} não é enumerável. ■

Mostramos a seguir mais um exemplo de subconjunto de números reais que não é boreliano.

Exemplo 2.13

Consideremos $x \in \mathbb{R}$ fixo. Podemos definir o conjunto

$$\tilde{\mathbb{Q}}^x = \{q + px \in \mathbb{R} : p, q \in \mathbb{Q}\}.$$

Observemos que o conjunto $\tilde{\mathbb{Q}}^x$ é enumerável, de maneira que existem inumeráveis números

reais que não pertencem a $\tilde{\mathbb{Q}}^x$. Seja então y um de tais números, ou seja, escolhamos $y \in \mathbb{R}$ de maneira que $y \notin \tilde{\mathbb{Q}}^x$ e formamos o conjunto $\tilde{\mathbb{Q}}^y$. Pode-se demonstrar que $x \notin \tilde{\mathbb{Q}}^y$.

Agora, imagine um conjunto de números reais A como se $x, y \in A$, então $x \notin \tilde{\mathbb{Q}}^y$ e $y \notin \tilde{\mathbb{Q}}^x$. Podemos encontrar um de tais conjuntos A como incontáveis elementos e A não sendo um conjunto do Borel.

Também é possível considerar, se interessante for, a σ -álgebra dos conjuntos de Borel restrita a uma porção dos números reais, por exemplo, restrita ao intervalo $[0, 1] \subset \mathbb{R}$. Nesta situação, a σ -álgebra de Borel em $[0, 1]$ é a σ -álgebra gerada pela coleção dos conjuntos abertos em $[0, 1]$, a qual pode ser pensada como a restrição de $\mathcal{B}(\mathbb{R})$ ao intervalo $[0, 1]$, como demonstrado no próximo teorema.

Teorema 2.12

Suponhamos que Ω seja o espaço amostral e que $\Omega' \subseteq \Omega$. Então

- (a) se \mathcal{F} é uma σ -álgebra de subconjuntos de Ω e $\mathcal{F}' = \{A \cap \Omega' : A \in \mathcal{F}\}$, temos que \mathcal{F}' é uma σ -álgebra de subconjuntos de Ω' e
- (b) se a classe de subconjuntos \mathcal{C} gera a σ -álgebra \mathcal{F} em Ω e $\mathcal{C}' = \{A \cap \Omega' : A \in \mathcal{C}\}$, temos que \mathcal{C}' gera a σ -álgebra \mathcal{F}' em Ω' .

Demonstração: Exercício. ■

O conceito de σ -álgebra de Borel pode ser estendido para dimensões mais elevadas.

2.2 Definição axiomática de probabilidade

A noção clássica da teoria da probabilidade, que começa com a noção de casos igualmente prováveis, dominou por 200 anos. Seus elementos foram postos em prática no início do século XVIII e permaneceram assim até o início do século XX. Ainda hoje a probabilidade clássica é utilizada no cálculo de probabilidades.

No início do século XX, muitos matemáticos estavam insatisfeitos com o que viram como uma falta de clareza e rigor no cálculo de probabilidades. A chamada mais célebre de esclarecimento veio de David Hilbert⁴. O sexto dos vinte e três problemas então em aberto que Hilbert apresentou ao Congresso Internacional de Matemáticos, em Paris, em 1900, foi para tratar axiomáticamente a teoria das probabilidades.

A teoria matemática da probabilidade, como a conhecemos hoje, é de origem relativamente recente. Foi A.N. Kolmogorov que axiomatiza a probabilidade em sua obra fundamental

⁴David Hilbert (1862-1943) foi um matemático alemão. David Hilbert é um dos mais notáveis matemáticos, e os tópicos de suas pesquisas são fundamentais em diversos ramos da matemática atual. Hilbert é frequentemente considerado como um dos maiores matemáticos do século XX, no mesmo nível de Henri Poincaré. Devemos a ele principalmente a lista de 23 problemas, alguns dos quais não foram resolvidos até hoje, apresentada em 1900 no Congresso Internacional de Matemáticos em Paris.

“Foundatins of the Theory of Probability” em 1933 (Kolmogorov, 1933). De acordo com este desenvolvimento, eventos aleatórios são representados por conjuntos e probabilidade é apenas uma medida padronizada definida nesses conjuntos.

Este desenvolvimento da teoria não só forneceu uma base logicamente consistente para a teoria das probabilidades, como também e ao mesmo tempo, junto-à corrente principal da matemática moderna. Nesta seção definiremos a função de probabilidade e estudaremos algumas propriedades importantes.

Definição 2.6 (*Definição Axiomática de Probabilidade*)

Seja Ω um espaço amostral e \mathcal{F} uma σ -álgebra definida em Ω . A função $P : \mathcal{F} \rightarrow [0, 1]$ é chamada de medida de probabilidade ou simplesmente probabilidade se satisfaz os seguintes axiomas:

Axioma I: $P(A) \geq 0$, para todo $A \in \mathcal{F}$,

Axioma II: $P(\Omega) = 1$,

Axioma III: Seja $\{A_n\}_{n=1}^{\infty}$ uma sequência de conjuntos disjuntos de \mathcal{F} , isto é, $A_i \cap A_j = \emptyset$ para $i \neq j$. Então

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n).$$

Chamaremos $P(A)$ de probabilidade do evento A . A propriedade contida no Axioma III é chamada de aditividade enumerável. Observemos também que $P(\emptyset) = 0$.

Uma pergunta básica sempre que são definidos novos conceitos é saber se a função de probabilidade sempre existirá, qualquer seja o espaço amostral. Veremos aqui como definir a função de probabilidade quando Ω contiver uma quantidade finita de elementos, quando contiver uma quantidade infinita enumerável de elementos e no caso quando contiver uma quantidade infinita não enumerável de elementos, chamado de caso contínuo.

Teorema 2.13

Seja Ω um espaço amostral não vazio e \mathcal{F} uma σ -álgebra definida em Ω . A função de probabilidade é finitamente aditiva.

Dizemos que uma função de probabilidade é finitamente aditiva se $\{A_k\}_{k=1}^n$, uma sequência de conjuntos disjuntos de \mathcal{F} , então

$$P\left(\bigcup_{k=1}^n A_k\right) = \sum_{k=1}^n P(A_k).$$

Demonstração: Sejam $A_1, A_2, \dots, A_n \in \mathcal{F}$ disjuntos. Notemos inicialmente que $P(\emptyset) = 0$, já que

$$P(\Omega) = P(\Omega \cup \emptyset \cup \emptyset \cup \dots) = P(\Omega) + P(\emptyset) + P(\emptyset) + \dots$$

Definamos $A_k = \emptyset$, para $k = n+1, n+2, \dots$. Então A_1, A_2, \dots são disjuntos, logo

$$P\left(\bigcup_{k=1}^n A_k\right) = P\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} P(A_k) = \sum_{k=1}^n P(A_k) + P(\emptyset) + P(\emptyset) + \dots = \sum_{k=1}^n P(A_k). \quad \blacksquare$$

Exemplo 2.14

Vejam as duas situações nas quais todo evento simples tem probabilidade um. Uma primeira situação é quando $\Omega = \{\omega\}$, $\mathcal{F} = \{\emptyset, \Omega\}$, $P(\emptyset) = 0$ e $P(\{\omega\}) = P(\Omega) = 1$. Uma segunda situação é considerando $\Omega = [0, 1]$, $\mathcal{F} = \{\emptyset, \{0\}, (0, 1], \Omega\}$, $P(\emptyset) = P((0, 1]) = 0$ e $P(\{0\}) = P(\Omega) = 1$.

Percebemos que um modelo probabilístico para um experimento, ou simplesmente um modelo probabilístico, é constituído de um conjunto não vazio Ω de resultados possíveis, o espaço amostral, uma σ -álgebra \mathcal{F} de eventos aleatórios e uma probabilidade P definida em \mathcal{F} . Agora vamos definir o conceito abstrato matemático de espaço de probabilidade.

Definição 2.7

Um espaço de probabilidade é um trio (Ω, \mathcal{F}, P) onde:

- (a) Ω é um conjunto não vazio de resultados possíveis de um experimento,
- (b) \mathcal{F} é uma σ -álgebra de subconjuntos de Ω ,
- (c) P é uma função de probabilidade definida em \mathcal{F} .

Devemos esclarecer que esta definição significa que, uma vez escolhido o espaço amostral, somente aos subconjuntos deste que pertençam à σ -álgebra \mathcal{F} serão atribuídos valores de probabilidade. Um outro esclarecimento que devemos fazer é que a função de probabilidade não é única.

Exemplo 2.15

Seja o experimento o lançamento de uma moeda, o espaço amostral $\Omega = \{\text{cara}, \text{coroa}\}$ e \mathcal{F} a σ -álgebra de todos os subconjuntos de Ω . Definamos a função de probabilidade como

$$P(\{\text{cara}\}) = 1/2, \quad P(\{\text{coroa}\}) = 1/2.$$

Então P é uma probabilidade bem definida. Similarmente, podemos definir

$$P(\{\text{cara}\}) = \frac{2}{3} \quad \text{e} \quad P(\{\text{coroa}\}) = \frac{1}{3}$$

ou

$$P(\{\text{cara}\}) = 1 \quad \text{e} \quad P(\{\text{coroa}\}) = 0.$$

De maneira geral, podemos definir

$$P(\{\text{cara}\}) = p, \quad P(\{\text{coroa}\}) = 1 - p \quad (0 \leq p \leq 1).$$

como sendo a função de probabilidade no espaço amostral Ω .

A importância de todos os elementos do espaço de probabilidade fica claro no exemplo a seguir, no qual todo evento simples têm probabilidade zero.

Exemplo 2.16

Considere o espaço de probabilidade (Ω, \mathcal{F}, P) , onde $\Omega = (0, 1)$, \mathcal{F} denota a σ -álgebra de todos os subintervalos de Ω e P é a função de probabilidade que atribui a cada intervalo da forma (a, b) , $0 < a \leq b < 1$, o comprimento $b - a$. Este espaço de probabilidade está bem definido e será considerado o espaço de probabilidade padrão no intervalo $(0, 1)$. No entanto, para cada número real $x \in (0, 1)$, o evento simples $\{x\}$ têm probabilidade zero, segundo P . Isto significa que $P(\{x\}) = 0$.

Para demonstrar isto, observemos que, para todo $\epsilon > 0$

$$P(\{x\}) \leq P\left(\{(x - \epsilon/2; x + \epsilon/2)\}\right) = \epsilon, \quad \forall \epsilon > 0.$$

Você poderia pensar que, se o espaço amostral Ω fosse enumerável, seguir-se-ia que nem todo evento simples poderia ter probabilidade zero. Caso todo evento simples pertencesse à σ -álgebra, você estaria correto. No entanto, não existe espaço amostral enumerável, nem mesmo finito, no qual todo evento simples têm probabilidade zero. Considere o seguinte caso: espaço amostral $\Omega = \{a, b, c\}$, $\mathcal{F} = \{\emptyset, \{a\}, \{b, c\}, \Omega\}$ e a função de probabilidade

$$P(\emptyset) = 0, \quad P(\{a\}) = 1, \quad P(\{b, c\}) = 0 \quad \text{e} \quad P(\Omega) = 1.$$

É claro que esta axiomática não foi a primeira proposta porém, mostrou-se mais prática e útil do que teorias anteriores. Além disso, proporcionou um espaço de probabilidade preciso para cada experimento aleatório e criou um marco totalmente abstrato o qual possibilitou o desenvolvimento da teoria moderna das probabilidades. Uma situação interessante foi apresentada no Exemplo 1.20. Dependendo do pensamento podemos encontrar distintos valores de probabilidade para o mesmo evento, o qual, evidentemente é uma contradição ao fato de que dado um evento A , $P(A)$ é única.

Definição 2.8

Seja $\{A_n\}_{n=1}^{\infty}$ uma sequência de eventos da σ -álgebra \mathcal{F} , definida no espaço amostral Ω . Dizemos que a sequência decresce para o vazio, denotado por $\{A_n\}_{n=1}^{\infty} \searrow \emptyset$, se $A_n \supseteq A_{n+1} \forall n$ e

$$\bigcap_{n=1}^{\infty} A_n = \emptyset.$$

Duas questões: a probabilidade atribuída a um eventos é única? existe uma única forma de definir probabilidade axiomática? devemos lembrar que temos, pelo menos, três formas diferentes de definir probabilidade clássica.

A resposta para a primeira pergunta é afirmativa e a justificativa é a afirmação (f) do Teorema 2.17. A resposta a segunda pergunta é negativa, ou seja, existem diversas formas equivalentes de definir probabilidade axiomática. Nosso objetivo agora é mostrar que as exigências da Definição 2.6 não são únicas, no sentido de que utilizando outros axiomas também podemos definir a função de probabilidade com as mesmas propriedades. Para chegar a uma destas alternativas à definição axiomática apresentada, utilizaremos a definição acima de sequência que decresce para o vazio.

Definição 2.9

Seja (Ω, \mathcal{F}, P) um espaço de probabilidade e $\{A_n\}_{n=1}^{\infty}$ uma sequência de eventos da σ -álgebra \mathcal{F} que decresce para o vazio. Dizemos que P é contínua no vazio se

$$\lim_{n \rightarrow \infty} P(A_n) = 0. \quad (2.6)$$

Se o espaço amostral Ω for discreto e contiver ao menos $n(< \infty)$ pontos, cada conjunto de um único ponto $\{w_k\}$, $k = 1, \dots, n$ é um evento elementar e é suficiente atribuir probabilidade a cada $\{w_k\}$. Então, se $A \in \mathcal{F}$, onde \mathcal{F} é a classe de todos os subconjuntos de Ω ,

$$P(A) = \sum_{w \in A} P(\{w\}).$$

Uma tal atribuição é a atribuição da mesma probabilidade ou a atribuição de probabilidades uniformes. De acordo com esta atribuição, $P(\{w_k\}) = 1/n$, $k = 1, \dots, n$. Portanto $P(A) = m/n$, se A contiver m eventos elementais, $1 \leq m \leq n$.

Caso Ω seja discreto e contiver um número enumerável de pontos, não se pode fazer uma atribuição da mesma probabilidade a cada evento simples. Basta fazer a atribuição da probabilidade para cada evento elementar. Se $A \in \mathcal{F}$, onde \mathcal{F} é a classe de todos os subconjuntos de Ω , definimos $P(A) = \sum_{w \in A} P(\{w\})$.

Exemplo 2.17

Seja o espaço amostral o conjunto $\Omega = \{1, 2, 3, \dots\}$ ou $\Omega = \mathbb{N}$ e a σ -álgebra potência $\mathcal{F} = \mathcal{P}(\Omega)$, a classe de todos os subconjuntos de Ω . Definamos P em \mathcal{F} como

$$P(\{k\}) = \frac{1}{2^k}, \quad k = 1, 2, \dots$$

Provar que P é função de probabilidade é uma tarefa do leitor. Queremos demonstrar aqui que P , definida acima, é uma função de probabilidade contínua no vazio.

Como a definição sugere, escolhemos uma sequência de eventos de \mathcal{F} que decresce

para o vazio da forma $A_n = \{n, n+1, \dots\}$. Observemos que

$$P(A_n) = \sum_{k=n}^{\infty} \frac{1}{2^k} = \frac{1}{2^{n-1}},$$

logo

$$\lim_{n \rightarrow \infty} P(A_n) = 0.$$

No Exemplo 2.17 utilizamos resultados da soma de séries geométricas da seguinte forma: seja a série $\sum_{k=1}^{\infty} a_1 q^{k-1}$, onde a_1 é o primeiro termo da série e $|q| < 1$. Então,

$$\sum_{k=1}^{\infty} a_1 q^{k-1} = \sum_{k=1}^{n-1} a_1 q^{k-1} + \sum_{k=n}^{\infty} a_1 q^{k-1} = \frac{a_1}{1-q},$$

e a expressão para a soma dos primeiros $n-1$ números é $a_1 \frac{1-q^{n-1}}{1-q}$. Na situação referida, $a_1 = 1/2$ e $q = 1/2$.

Teorema 2.14

Seja (Ω, \mathcal{F}, P) um espaço de probabilidade. Uma função satisfazendo os axiomas I e II na definição de probabilidade e ainda finitamente aditiva é uma probabilidade se, e somente se, é contínua no vazio.

Demonstração: Suponhamos que P é contínua no vazio. Sejam $A_1, A_2, \dots \in \mathcal{F}$ disjuntos, queremos provar que

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n).$$

Seja $A = \bigcup_{n=1}^{\infty} A_n$, então $A = \left(\bigcup_{n=1}^k A_n\right) \bigcup \left(\bigcup_{n=k+1}^{\infty} A_n\right)$ e, pela aditividade finita,

$$P(A) = \sum_{n=1}^k P(A_n) + P\left(\bigcap_{n=k+1}^{\infty} A_n\right).$$

Seja $B_k = \bigcap_{n=k+1}^{\infty} A_n$, então $\{B_k\}_{k=1}^{\infty} \searrow \emptyset$ e, portanto, $\lim_{k \rightarrow \infty} P(B_k) = 0$. Logo $\lim_{k \rightarrow \infty} \sum_{n=1}^k P(A_n) =$

$P(A)$ e então $P(A) = \sum_{n=1}^{\infty} P(A_n)$. Suponhamos agora P uma função de probabilidade, queremos provar que é contínua no vazio. Sejam $A_1, A_2, \dots \in \mathcal{F}$ tais que $A_n \supseteq A_{n+1}$ e $\{A_n\}_{n=1}^{\infty} \searrow \emptyset$, queremos provar que $\lim_{n \rightarrow \infty} P(A_n) = 0$. Temos $A_1 = (A_1 \setminus A_2) \bigcup (A_2 \setminus A_3) \bigcup \dots =$

$\bigcup_{k=1}^{\infty} (A_k \setminus A_{k+1})$. Cada $A_k \setminus A_{k+1}$ são disjuntos, então $P(A_1) = \sum_{k=1}^{\infty} P(A_k \setminus A_{k+1})$, portanto, a série é convergente e

$$\lim_{n \rightarrow \infty} \sum_{k=1}^{n-1} P(A_k \setminus A_{k+1}) = P(A_1).$$

Pela aditividade finita $P(A_k \setminus A_{k+1}) = P(A_k) - P(A_{k+1})$, logo

$$P(A_1) = \lim_{n \rightarrow \infty} \sum_{k=1}^{n-1} P(A_k \setminus A_{k+1}) = \lim_{n \rightarrow \infty} [P(A_1) - P(A_n)],$$

e então $\lim_{n \rightarrow \infty} P(A_n) = 0$. ■

Este teorema afirma que, se P é uma função de probabilidade e se $\{A_n\}_{n \geq 1}$ for uma sequência de eventos que decresce para o vazio (Definição 2.8), então $\lim_{n \rightarrow \infty} P(A_n) = 0$. Estamos agora em condições de apresentar uma definição axiomática alternativa de probabilidade.

Teorema 2.15

Seja (Ω, \mathcal{F}, P) um espaço de probabilidade. Os dois seguintes sistemas de axiomas são equivalentes:

Sistema I: $P(A) \geq 0$, para todo $A \in \mathcal{F}$, $P(\Omega) = 1$ e se $\{A_n\}_{n=1}^{\infty}$ é uma sequência de conjuntos disjuntos de \mathcal{F} , então

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n),$$

Sistema II: $P(A) \geq 0$, para todo $A \in \mathcal{F}$, $P(\Omega) = 1$, se $\{A_k\}_{k=1}^n$ é uma sequência de conjuntos disjuntos de \mathcal{F} , então $P(\bigcup_{k=1}^n A_k) = \sum_{k=1}^n P(A_k)$ e se $\{B_n\}_{n=1}^{\infty}$ é uma outra sequência de eventos em \mathcal{F} que decresce para o vazio, então

$$\lim_{n \rightarrow \infty} P(B_n) = 0.$$

Demonstração: Exercício. ■

Então, para verificar se a função P é uma probabilidade em \mathcal{F} , basta verificar os axiomas do sistema I ou os axiomas do sistema II apresentados no Teorema 2.15.

Se Ω contiver uma quantidade não enumerável de pontos, cada conjunto de um ponto é um evento simples e, novamente, não se pode fazer uma atribuição da mesma probabilidade. De fato, não se pode atribuir probabilidade positiva a cada evento elementar sem violar o axioma $P(\Omega) = 1$. Neste caso, atribuímos probabilidades a eventos compostos constituídos por intervalos. Esta situação será tratada depois na Seção 2.1.1.

Exemplo 2.18

Seja o espaço amostral o conjunto $\Omega = \mathbb{R}$ e a σ -álgebra $\mathcal{F} = \mathcal{B}(\mathbb{R})$, a σ -álgebra de Borel. Para cada intervalo I , definamos

$$P(I) = \int_I \frac{1}{\pi} \frac{1}{1+x^2} dx.$$

No Exemplo 3.33 vamos provar que P assim definida é uma função de probabilidade. Queremos demonstrar aqui que P é uma função de probabilidade contínua no vazio.

Seja $I_n = (n, \infty)$ uma sequência de intervalos que decresce para o vazio. Acontece que

$$P(I_n) = \frac{1}{2} - \frac{1}{\pi} \arctan(n)$$

e

$$\lim_{n \rightarrow \infty} \frac{1}{\pi} \arctan(n) = \frac{\pi}{2}.$$

Portanto, $\lim_{n \rightarrow \infty} P(I_n) = 0$.

Devemos lembrar que, dado um espaço amostral Ω , uma álgebra de eventos \mathcal{F} sempre é possível definir em Ω . No entanto podemos supor, sem perda de generalidade, que \mathcal{F} é uma σ -álgebra em vez de álgebra, pelo Teorema de Extensão de Carathéodory⁵. Este teorema garante que uma probabilidade definida em uma álgebra, e de acordo com os axiomas usuais, pode ser estendida de uma única maneira para a σ -álgebra gerada pela álgebra.

Teorema 2.16 (Teorema de Extensão de Carathéodory)

Seja Ω um espaço amostral e \mathcal{F} uma álgebra de eventos de Ω . Seja P uma probabilidade definida em \mathcal{F} . Então existe uma única função de probabilidade \tilde{P} , definida em $\sigma(\mathcal{F})$, tal que P e \tilde{P} coincidirem em \mathcal{F} .

Demonstração: Seja

$$\mathcal{F}_\sigma = \{A \subseteq \Omega : \exists \{A_n\}_{n=1}^\infty, \text{ cada } A_n \in \mathcal{F}, A_n \nearrow A, \text{ quando } n \rightarrow \infty\}.$$

Na definição da classe de conjuntos \mathcal{F}_σ , $\{A_n\}_{n=1}^\infty$ é uma sequência não decrescente que converge para A , isto é, $\{A_n\}_{n=1}^\infty$ é uma sequência que satisfaz $A_n \subseteq A_{n+1}$ e $\bigcup_{n=1}^\infty A_n = A$.

Definamos $\tilde{P} : \mathcal{F}_\sigma \rightarrow [0, 1]$ de forma que, se $A_n \nearrow A$, quando $n \rightarrow \infty$, então $\tilde{P}(A) = \lim_{n \rightarrow \infty} P(A_n)$. Vejamos que \tilde{P} está bem definida e estende P .

Definamos para $E \subseteq \Omega$ a função

$$\tilde{P}(E) = \inf \left\{ \sum_{i=1}^\infty P(A_i) : E \subseteq \bigcup_{i=1}^\infty A_i, A_i \in \mathcal{F} \right\}. \quad (2.7)$$

Verifiquemos se a função \tilde{P} é uma probabilidade em $\sigma(\mathcal{F})$. ■

⁵Constantin Carathéodory (1873-1950) foi um matemático alemão de origem grega.

Devemos esclarecer que este teorema não garante a existência de alguma σ -álgebra, qualquer seja o espaço amostral. Lembremos que $\sigma(\mathcal{F})$ está definida em termos da existência de σ -álgebras.

2.2.1 Propriedades da probabilidade axiomática

Consideremos que o espaço de probabilidade (Ω, \mathcal{F}, P) está definido num determinado experimento aleatório e chamemos de $A \in \mathcal{F}$ um evento aleatório qualquer na σ -álgebra \mathcal{F} e, portanto, A é um evento ao qual podemos atribuir um valor de probabilidade único, segundo a função P . Nestas condições valem as seguintes propriedades.

Teorema 2.17

Seja (Ω, \mathcal{F}, P) um espaço de probabilidade. A função de probabilidade satisfaz as seguintes propriedades:

- (a) $P(A^c) = 1 - P(A)$,
- (b) $0 \leq P(A) \leq 1$,
- (c) Se $A \subseteq B$, então $P(A) \leq P(B)$ e ainda $P(B \setminus A) = P(B) - P(A)$.
- (d) Sejam A_1, A_2, \dots, A_n elementos da σ -álgebra \mathcal{F} . Então

$$P\left(\bigcup_{k=1}^n A_k\right) \leq \sum_{k=1}^n P(A_k),$$

- (e) Sejam A_1, A_2, \dots elementos da σ -álgebra \mathcal{F} . Então

$$P\left(\bigcup_{k=1}^{\infty} A_k\right) \leq \sum_{k=1}^{\infty} P(A_k),$$

- (f) (Continuidade da probabilidade) Seja $\{A_n\}_{n=1}^{\infty}$ uma sequência de eventos não crescentes em \mathcal{F} tais que convergem para o evento A , isto é, $A_n \supseteq A_{n+1}$ para qualquer valor de n e $\bigcap_{n=1}^{\infty} A_n = A$. Então

$$\lim_{n \rightarrow \infty} P(A_n) = P(A).$$

Demonstração: (a) Como consequência dos axiomas I e III, $A \cup A^c = \Omega$. Então $P(A) + P(A^c) = 1$.

(b) Pelo axioma I, $P(A) \geq 0$. Também $A \subseteq \Omega$, então $P(A) \leq 1$.

(c) Se $A \subseteq B$, então $B = (A \cap B) \cup (B \setminus A) = A \cup (B \setminus A)$ segue então, pela aditividade da função de probabilidade, que $P(B) = P(A) + P(B \setminus A)$.

(d) Observemos que

$$P(A_1 \cup A_2) = P(A_1) + P(A_2 \cap A_1^c) \leq P(A_1) + P(A_2),$$

isto pela propriedade (c), devido a $A_2 \cap A_1^c \subseteq A_2$. A demonstração completa-se por indução.

(e) Exercício.

(f) Observemos que $P(A_n) \geq P(A_{n+1})$, pela propriedade (c) e pela continuidade no vazio (Teorema 2.15), do fato da sequência $\{A_n \setminus A\}_{n \geq 1}$ satisfazer que $(A_{n+1} \setminus A) \subseteq (A_n \setminus A)$ e $\bigcap_{n \geq 1} \{A_n \setminus A\} = \emptyset$, temos que

$$\lim_{n \rightarrow \infty} P(A_n \setminus A) = 0.$$

A aditividade finita implica que $P(A_n \setminus A) = P(A_n) - P(A)$, pois $A \subseteq A_n$. Resumindo, temos que

$$\lim_{n \rightarrow \infty} [P(A_n) - P(A)] = 0$$

e a sequência $\{P(A_n)\}_{n \geq 1}$ é não crescente, logo $\lim_{n \rightarrow \infty} P(A_n) = P(A)$. ■

Gostaríamos de salientar neste ponto que, se $P(A) = 0$ para algum $A \in \mathcal{F}$, chamamos A de um evento com probabilidade zero ou um evento nulo. No entanto, isso não significa que $A = \emptyset$. Por outro lado, se $P(B) = 1$ para algum $B \in \mathcal{F}$, chamamos B de evento certo, mas isso não quer dizer que $B = \Omega$.

Exemplo 2.19

Uma moeda é lançada três vezes. Vamos atribuir igual probabilidade para cada um dos 2^3 eventos elementares em Ω . Seja A o evento de que pelo menos uma cara aparece em três lances. Então

$$\begin{aligned} P(A) &= 1 - P(A^c) = 1 - P(\text{nenhuma cara}) \\ &= 1 - P(\{Coroa\} \cap \{Coroa\} \cap \{Coroa\}) = \frac{7}{8}. \end{aligned}$$

Deste resultado podemos obter algumas outras propriedades, por exemplo, a chamada desigualdade de Boole⁶, também conhecida como união vinculada, diz que, para qualquer conjunto finito ou contável de eventos, a probabilidade de que pelo menos um dos eventos ocorra não é maior que a soma das probabilidades dos eventos individuais. Aa seguir, apresentamos o princípio de inclusão-exclusão, um belo teorema que expressa a relação exata entre a probabilidade de uniões e as probabilidades de eventos individuais, a desigualdade de Bonferroni⁷ e outros.

⁶George Boole (1815 - 1864), matemático e filósofo inglês. Como o inventor da lógica booleana, base do computador moderno, Boole é visto em retrospecto como um dos fundadores da ciência da computação.

⁷Carlo Emilio Bonferroni (1892 - 1960), foi um matemático italiano que trabalhou na teoria das probabilidades.

Teorema 2.18 (*Desigualdade de Boole*)

Seja (Ω, \mathcal{F}, P) um espaço de probabilidade e $\{A_n\}_{n \geq 1}$, $n = 1, 2, \dots$, uma sequência enumerável de eventos em \mathcal{F} . Então

$$P\left(\bigcap_{n=1}^{\infty} A_n\right) \geq 1 - \sum_{n=1}^{\infty} P(A_n^c).$$

Demonstração: Primeiro consideremos dois eventos $A, B \in \mathcal{F}$, então

$$P(A \cap B) = 1 - P((A \cap B)^c) \geq 1 - (P(A^c) + P(B^c)),$$

e definamos

$$B = \bigcap_{n=2}^{\infty} A_n \quad \text{e} \quad A = A_1. \quad \blacksquare$$

Observemos agora que se $A, B \in \mathcal{F}$, isto é, A e B eventos quaisquer da σ -álgebra, temos que

$$A \cup B = (A \setminus B) \cup (B \setminus A) \cup (A \cap B).$$

Os eventos acima a direita são todos disjuntos, então pelo Axioma III

$$P(A \cup B) = P(A \setminus B) + P(B \setminus A) + P(A \cap B)$$

e, dado que $A = (A \cap B) \cup (A \setminus B)$ e $B = (A \cap B) \cup (B \setminus A)$, chegamos a que

$$P(A \cup B) = P(A) + P(B) - P(A \cap B). \quad (2.8)$$

Como generalização desta propriedade temos o seguinte resultado.

Teorema 2.19 (*Princípio de inclusão-exclusão*)

Sejam $A_1, A_2, \dots, A_n \in \mathcal{F}$, eventos definidos no espaço de probabilidade (Ω, \mathcal{F}, P) . Então

$$\begin{aligned} P\left(\bigcup_{k=1}^n A_k\right) &= \sum_{k=1}^n P(A_k) - \sum_{k_1 < k_2}^n P(A_{k_1} \cap A_{k_2}) \\ &\quad + \sum_{k_1 < k_2 < k_3}^n P(A_{k_1} \cap A_{k_2} \cap A_{k_3}) + \dots + (-1)^{n+1} P\left(\bigcap_{k=1}^n A_k\right). \end{aligned}$$

Demonstração: Utilizar repetidas vezes o resultado em (2.8). \blacksquare

Observemos que, da aplicação direta do resultado do Teorema 2.19, temos

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

para o caso de dois eventos aleatórios e que, se tivermos três eventos aleatórios, então

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C).$$

Exemplo 2.20

Suponha que um dado seja lançado duas vezes, observamos o número obtido em cada lance e que a todos os eventos elementares em

$$\Omega = \{(i, j) : i, j = 1, 2, \dots, 6\}$$

atribuímos a mesma probabilidade. Considere A o evento de que o primeiro lance mostre um número ≤ 2 e B , o evento em que o segundo lance mostre pelo menos o número 5. Então

$$A = \{(i, j) : 1 \leq i \leq 2, j = 1, 2, \dots, 6\},$$

$$B = \{(i, j) : 5 \leq j \leq 6, i = 1, 2, \dots, 6\},$$

$$A \cap B = \{(1, 5); (1, 6); (2, 5); (2, 6)\}.$$

Calculando a probabilidade a ocorrência simultânea destes dois eventos, obtemos que

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{1}{3} + \frac{1}{3} - \frac{4}{36} = \frac{5}{9}.$$

Observe que, mesmo sem mencionar, utilizamos o Princípio de inclusão-exclusão para encontrarmos o resultado no exemplo anterior assim como no exemplo a seguir.

Exemplo 2.21

Seja (Ω, \mathcal{F}, P) um espaço de probabilidades e $A, B \in \mathcal{F}$ eventos tais que $P(A) = 0,25$ e $P(B) = 0,8$. Provemos que

$$0,05 \leq P(A \cap B) \leq 0,25.$$

Observe que $A \cap B \subseteq A$ e $A \cap B \subseteq B$, então $P(A \cap B) \leq P(A)$ e $P(A \cap B) \leq P(B)$. Portanto,

$$P(A \cap B) \leq \min\{P(A), P(B)\}$$

do qual temos que $P(A \cap B) \leq 0,25$. Dado que $A \cup B \subseteq \Omega$, temos que $P(A \cup B) \leq 1$, então

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \leq 1 \\ 0,25 + 0,8 - 1 &\leq P(A \cap B), \end{aligned}$$

do qual obtemos que $0,05 \leq P(A \cap B) \leq 0,25$.

Teorema 2.20

Seja (Ω, \mathcal{F}, P) um espaço de probabilidade e $\{A_n\}_{n \geq 1}$ uma sequência enumerável de eventos não decrescente em \mathcal{F} , isto é, $A_n \subseteq A_{n+1}$ para todo n . Então

$$\lim_{n \rightarrow \infty} P(A_n) = P\left(\bigcup_{n=1}^{\infty} A_n\right).$$

Demonstração: Consequência da propriedade de continuidade da probabilidade, item (f) do Teorema 2.17. ■

Exemplo 2.22

Considere o espaço amostral $\Omega = [0, 1]$ e a sequência de eventos em Ω definida como

$$A_n = [0, 1 - 1/n].$$

É claro que $A_n \subseteq A_{n+1}$. Aplicando então o Teorema 2.20 temos que

$$\lim_{n \rightarrow \infty} P(A_n) = 1.$$

Teorema 2.21 (Desigualdade de Bonferroni)

Sejam A_1, A_2, \dots, A_n eventos aleatórios ($n > 1$), então

$$\sum_{k=1}^n P(A_k) - \sum_{k < j} P(A_k \cap A_j) \leq P\left(\bigcup_{k=1}^n A_k\right) \leq \sum_{k=1}^n P(A_k). \quad (2.9)$$

Demonstração: Como consequência do item (d) do Teorema 2.17, será suficiente demonstrar a parte esquerda em (2.9) e isto será realizado por indução. A desigualdade a esquerda é verdadeira para $n = 2$, desde que

$$P(A_1) + P(A_2) - P(A_1 \cap A_2) = P(A_1 \cup A_2).$$

Para $n = 3$ $P\left(\bigcup_{k=1}^3 A_k\right) = \sum_{k=1}^3 P(A_k) - \sum_{i < j} P(A_i \cap A_j) + P(A_1 \cap A_2 \cap A_3)$, e o resultado em (2.9) é válido. Assumindo que (2.9) é válido para $3 < m \leq n - 1$, mostraremos que

também será válido para $m + 1$.

$$\begin{aligned}
 P\left(\bigcup_{i=1}^{m+1} A_i\right) &= P\left[\left(\bigcup_{i=1}^m A_i\right) \cup A_{m+1}\right] \\
 &= P\left(\bigcup_{i=1}^m A_i\right) + P(A_{m+1}) - P\left[A_{m+1} \cap \left(\bigcup_{i=1}^m A_i\right)\right] \\
 &\geq \sum_{i=1}^{m+1} P(A_i) - \sum_{i < j}^m P(A_i \cap A_j) - P\left(\bigcup_{i=1}^m (A_i \cap A_{m+1})\right) \\
 &\geq \sum_{i=1}^{m+1} P(A_i) - \sum_{i < j}^m P(A_i \cap A_j) - \sum_{i=1}^m P(A_i \cap A_{m+1}) \\
 &= \sum_{i=1}^{m+1} P(A_i) - \sum_{i < j}^{m+1} P(A_i \cap A_j).
 \end{aligned}$$

■

Uma melhoria à desigualdade de Bonferroni pode ser encontrada em Worsley (1982).

Exemplo 2.23

Suponha que temos dez eventos A_k , tais que $P(A_k) = 0.99$, para cada $k = 1, 2, \dots, 10$. Queremos estimar a probabilidade $P\left(\bigcap_{k=1}^{10} A_k\right)$.

Observemos que o Teorema 2.21 relaciona somas de probabilidades com probabilidades de uniões de eventos. Então, devemos transformar o evento $\bigcap_{k=1}^{10} A_k$ em uniões de eventos. Para isso utilizamos a Lei de De Morgan, da qual obtemos que

$$P\left(\bigcap_{k=1}^{10} A_k\right) = 1 - P\left(\bigcup_{k=1}^{10} A_k^c\right).$$

Segundo a desigualdade de Bonferroni,

$$P\left(\bigcup_{k=1}^{10} A_k^c\right) \leq \sum_{k=1}^{10} P(A_k^c) = 10 \times 0.01 = 0.10,$$

disto concluímos que $P\left(\bigcap_{k=1}^{10} A_k\right) \geq 0.90$.

Definição 2.10

Seja \mathcal{F} uma σ -álgebra de eventos do espaço amostral Ω e sejam A_1, A_2, \dots eventos aleatórios definidos em \mathcal{F} . Definimos,

$$\begin{aligned}
 \limsup_{n \rightarrow \infty} A_n &= \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k, \\
 \liminf_{n \rightarrow \infty} A_n &= \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k.
 \end{aligned}$$

Nesta definição podemos observar que o evento $\limsup_{n \rightarrow \infty} A_n$ significa a ocorrência de um número infinito dos A_n . Para justificar isto, observemos que

$$w \in \limsup_{n \rightarrow \infty} A_n \iff w \in \bigcup_{k=n}^{\infty} A_k, \quad \forall n.$$

A função de probabilidade é uma função contínua, num sentido particular, a função de probabilidade é uma função de conjuntos contínua. Isso é demonstrado no teorema a seguir.

Teorema 2.22

Seja (Ω, \mathcal{F}, P) um espaço de probabilidade e A_1, A_2, \dots eventos aleatórios definidos em \mathcal{F} . Se $\limsup_{n \rightarrow \infty} A_n = \liminf_{n \rightarrow \infty} A_n$ dizemos que a sequência $\{A_n\}_{n=1}^{\infty}$ tem limite. Se o evento $A \in \mathcal{F}$ é o limite da sequência, isto é, se $\lim_{n \rightarrow \infty} A_n = A$ então

$$\lim_{n \rightarrow \infty} P(A_n) = P(A).$$

Demonstração: Exercício. ■

Na Seção 3.2 vamos apresentar mais propriedades matemáticas da função de probabilidade ao considerarmos derivadas e integrais.

2.3 Probabilidade condicional

Fixemos (Ω, \mathcal{F}, P) como o espaço de probabilidade de interesse e nele consideremos $A, B \in \mathcal{F}$ dois eventos, suponhamos também que $P(B) > 0$. Vimos alguns resultados na seção anterior que trataram de como calcular a probabilidade da união de dois conjuntos, $A \cup B$. Estamos agora interessados em calcular probabilidade da interseção de dois conjuntos, $A \cap B$. Esta probabilidade é referida como a probabilidade conjunta dos conjuntos A e B, $P(A \cap B)$. Nesta seção, vamos perguntar e responder a seguinte pergunta. Se percebemos a ocorrência de um evento B : como devemos mudar as probabilidades dos eventos restantes? Chamaremos a nova probabilidade de um evento A a probabilidade condicional de A dado B .

Queremos estudar então as mudanças na probabilidade de ocorrência do evento quando se conhece que um outro evento ocorreu. Nestes casos, deve-se redefinir o espaço amostral considerando apenas os elementos de B como resultados possíveis.

Exemplo 2.24

Por exemplo, considere o experimento de "jogar um dado" e nos perguntar sobre a probabilidade de obter um seis, sabendo que a cara escolhida mostra um número par. Neste caso, a probabilidade não é $1/6$, uma vez que temos a certeza de que o resultado está no conjunto $\{2, 4, 6\}$. Cada um destes três resultados têm igual probabilidade, então a probabilidade de obtermos um seis sabendo que o resultado é par será $1/3$.

Vamos tentar determinar qual deve ser a probabilidade de um evento A condicionado ao conhecimento de que o evento B ocorreu, usando a interpretação heurística de probabilidade como limite da frequência com que um evento ocorre. Para isso, suponha que fizemos n réplicas independentes da experiência. Denotamos como n_B o número de vezes que o evento B ocorre e por $n_{A \cap B}$ o número de vezes nas quais ocorreu o resultado $A \cap B$.

Heuristicamente a probabilidade condicional de A dado B é o limite da frequência de ocorrência de A nos experimentos onde ocorre B , ou seja, o limite de $n_{A \cap B}/n_B$. Logo, a probabilidade de que ocorra A condicionado a B é dada por

$$\lim_{n \rightarrow \infty} \frac{n_{A \cap B}}{n_B} = \lim_{n \rightarrow \infty} \frac{\frac{n_{A \cap B}}{n}}{\frac{n_B}{n}} = \frac{\lim_{n \rightarrow \infty} \frac{n_{A \cap B}}{n}}{\lim_{n \rightarrow \infty} \frac{n_B}{n}} = \frac{P(A \cap B)}{P(B)}.$$

Isto justifica a seguinte definição.

Definição 2.11

Seja (Ω, \mathcal{F}, P) um espaço de probabilidade e $A, B \in \mathcal{F}$ tal que $P(B) > 0$. Definimos a probabilidade condicional de A dado B , a qual escreveremos como $P(A|B)$, como

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad \forall A \in \mathcal{F}. \quad (2.10)$$

Foi definido um novo conceito de probabilidade mas, surge a pergunta $P(\cdot|B)$, para cada B fixo é uma função de probabilidade? nos referimos à definição axiomática de probabilidade. Isto é demonstrado no teorema a seguir.

Teorema 2.23

Seja (Ω, \mathcal{F}, P) um espaço de probabilidade. Dado um evento $B \in \mathcal{F}$ tal que $P(B) > 0$ fixo definamos a função $\tilde{P} : \Omega \rightarrow [0, 1]$ como

$$\tilde{P}(A) = P(A|B),$$

para todo $A \in \mathcal{F}$. Então $(\Omega, \mathcal{F}, \tilde{P})$ é um espaço de probabilidade.

Demonstração: Observemos que o numerador de $P(A|B)$ é um número real não negativo para qualquer $A \in \mathcal{F}$. Então $P(\cdot|B)$ será sempre positivo ou zero.

(i)

$$\tilde{P}(\Omega) = P(\Omega|B) = \frac{P(\Omega \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1.$$

- (ii) Sejam $A_1, A_2, \dots \in \mathcal{F}$ eventos disjuntos em pares, ou seja, eventos tais que $A_i \cap A_j = \emptyset$, quaisquer sejam $i \neq j$. Então

$$\begin{aligned} \tilde{P}\left(\bigcup_{k=1}^{\infty} A_k\right) &= P\left(\bigcup_{k=1}^{\infty} A_k | B\right) = \frac{P\left(\left(\bigcup_{k=1}^{\infty} A_k\right) \cap B\right)}{P(B)} \\ &= \frac{P\left(\bigcup_{k=1}^{\infty} A_k \cap B\right)}{P(B)} = \frac{\sum_{k=1}^{\infty} P(A_k \cap B)}{P(B)} \\ &= \sum_{k=1}^{\infty} \frac{P(A_k \cap B)}{P(B)} = \sum_{k=1}^{\infty} P(A_k | B) = \sum_{k=1}^{\infty} \tilde{P}(A_k). \quad \blacksquare \end{aligned}$$

O que fizemos foi reduzir o espaço de probabilidade inicial ao considerar realizado o evento B . Com isso o novo espaço amostral consiste do evento B e da σ -álgebra $\widetilde{\mathcal{F}} = \mathcal{F} \cap B$, a qual sabemos consiste dos subconjuntos $A \cap B$, $A \in \mathcal{F}$ pertencentes a B . Neste espaço definimos a função \tilde{P} multiplicando a probabilidade de cada evento por $P(B)^{-1}$. Na verdade $(\Omega, \widetilde{\mathcal{F}}, \tilde{P})$ é um espaço de probabilidade.

Exemplo 2.25

Imagine que três candidatos A , B e C estão concorrendo para o Senado. Decidimos que A e B têm a mesma chance de ganhar e que a chance de C ganhar é de apenas metade da chance de A ganhar. Entendemos então que as probabilidades de ganhar são $P(A) = 2/5$, $P(B) = 2/5$ e $P(C) = 1/5$.

Suponha-se que antes da realização da eleição o candidato A cai fora da corrida. Seria natural atribuir novas probabilidades aos eventos B e C , que sejam proporcionais às probabilidades originais. Assim, teríamos $P(B|A \text{ não é mais candidato}) = 2/3$ e $P(C|A \text{ não é mais candidato}) = 1/3$.

É importante notar que a qualquer momento podemos atribuir probabilidades a eventos da vida real, a probabilidade resultante será útil somente se levarmos em conta toda a informação relevante. Neste exemplo, poderíamos ter o conhecimento de que a maioria dos eleitores irão votar a favor de C dado que A não está mais na corrida. Isto tornará claramente a probabilidade condicional de C ganhar maior do que o valor de $1/3$ atribuído anteriormente.

Teorema 2.24

Seja $C_1, C_2, \dots, C_n \in \mathcal{F}$ uma partição de Ω . Então

$$P(A|B) = \sum_{k=1}^n P(A \cap C_k | B). \quad (2.11)$$

Demonstração: Observemos que $\cup_{k=1}^n (A \cap C_k) = A$ e que $A \cap C_k$ são eventos disjuntos em pares, temos então que

$$P(A|B) = P\left(\bigcup_{k=1}^n (A \cap C_k)|B\right) = \sum_{k=1}^n P(A \cap C_k|B),$$

como queríamos demonstrar. ■

Exemplo 2.26

Uma senhora da alta sociedade dá uma festa em sua mansão. Ao término da festa, ela descobre que sua coleção de joias foi roubada. Após as investigações, a polícia tem certeza de que o ladrão foi precisamente uma das 76 pessoas presentes à festa (entre convidados e garçons). Ademais, os investigadores encontram na cena do crime o perfil de DNA do ladrão, e sabe-se que este perfil de DNA ocorre em 2% de toda população.

Dado que o DNA do Sr. João, o primeiro suspeito cujo DNA é analisado, combina com o perfil achado na cena do crime, qual é a probabilidade de que ele tenha roubado as joias?

Queremos calcular

$$P(\text{Sr. João tenha roubado as joias} \mid \text{perfil de DNA do ladrão}),$$

e para isso sabemos que

$$P(\text{Sr. João tenha roubado as joias e tenha perfil de DNA do ladrão}) = 1/76$$

e que $P(\text{perfil de DNA do ladrão}) = 0,02$.

Utilizando a definição de probabilidade condicional obtemos que a probabilidade de que Sr. João tenha roubado as joias é de 66%, aproximadamente.

Consideremos A e B dois eventos tais que $P(A) > 0$ e $P(B) > 0$. Segue da expressão em (2.10) que

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

a qual pode ser escrita também como

$$P(A \cap B) = P(A)P(B|A).$$

Estas relações podem ser generalizadas para qualquer número finito de eventos. Para vermos isso, consideremos os eventos $A_1, A_2, \dots, A_n \in \mathcal{F}$, $n \geq 2$, tais que $P(\cap_{k=1}^n A_k) > 0$.

Dado que

$$A_1 \supset (A_1 \cap A_2) \supset (A_1 \cap A_2 \cap A_3) \supset \dots \supset (\cap_{k=1}^{n-2} A_k) \supset (\cap_{k=1}^{n-1} A_k),$$

da qual percebemos que

$$P(A_1) > 0, P(A_1 \cap A_2) > 0, \dots, P(\cap_{k=1}^{n-2} A_k) > 0.$$

Segue então que

$$P(A_m \mid \cap_{k=1}^{m-1} A_k),$$

está bem definida para $m = 2, 3, \dots, n$.

Na teoria da probabilidade, a lei ou fórmula da probabilidade total é uma regra fundamental que relaciona probabilidades marginais a probabilidades condicionais. Expressa a probabilidade total de um resultado que pode ser realizado através de vários eventos distintos, daí o nome.

Teorema 2.25 (Teorema da probabilidade Total)

Seja (Ω, \mathcal{F}, P) um espaço de probabilidade e A_1, A_2, \dots uma partição de Ω , tal que $P(A_k) > 0, \forall k$. Então

$$P(B) = \sum_{k=1}^{\infty} P(A_k)P(B|A_k), \quad \forall B \in \mathcal{F}. \quad (2.12)$$

Demonstração: Observemos que $B = \bigcup_{k=1}^{\infty} (B \cap A_k)$ e que $B \cap A_1, B \cap A_2, \dots$ é uma partição de Ω . Então

$$P(B) = \sum_{k=1}^{\infty} P(B \cap A_k) = \sum_{k=1}^{\infty} P(A_k)P(B|A_k). \quad \blacksquare$$

A declaração matemática acima pode ser interpretada da seguinte maneira: dado um evento B , com probabilidades condicionais conhecidas, dado qualquer um dos eventos A_k , cada um com uma probabilidade conhecida, qual é a probabilidade total de que B aconteça? A resposta a esta pergunta é dada por $P(B)$.

Exemplo 2.27

Um experimento consiste em selecionar uma bola de três urnas segundo o seguinte procedimento: a urna 1 contém uma bola branca e duas bolas pretas, a urna 2 contém uma bola preta e duas bolas brancas e a urna 3 contém três bolas pretas e três bolas brancas. Um dado é lançado. Se a face superior mostra os números 1, 2 ou 3 a urna 1 é selecionada; caso mostra 4 a urna 2 é selecionada e se 5 ou 6 são mostrados, a urna 3 é a selecionada. Uma bola é, então, aleatoriamente escolhida a partir da urna selecionada. Seja A o evento que a bola escolhida seja branca e queremos encontrar $P(A)$.

Sejam U, V e W eventos que designam a urna selecionada ter sido 1, 2 ou 3, respectivamente. Então $A = (A \cap U) \cup (A \cap V) \cup (A \cap W)$ e sabemos que

$$P(A \cap U) = P(U)P(A|U) = \frac{3}{6} \times \frac{1}{3}, \quad P(A \cap V) = P(V)P(A|V) = \frac{1}{6} \times \frac{2}{3}$$

e

$$P(A \cap W) = P(W)P(A|W) = \frac{2}{6} \times \frac{1}{2}.$$

Segue, da expressão em (2.12), que

$$P(A) = P(U)P(A|U) + P(V)P(A|V) + P(W)P(A|W) = \frac{4}{9}.$$

Teorema 2.26 (Teorema da Multiplicação)

Seja (Ω, \mathcal{F}, P) um espaço de probabilidade e $A_1, \dots, A_n \in \mathcal{F}$ tais que $P(\cap_{k=1}^n A_k) > 0$. Então

$$P\left(\bigcap_{k=1}^n A_k\right) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P\left(A_n \mid \bigcap_{k=1}^{n-1} A_k\right). \quad (2.13)$$

Demonstração: Exercício. ■

Em algumas situações estamos interessados em consideramos probabilidades condicionais da seguinte forma: Dado o resultado da segunda fase de uma experiência de dois estágios, encontrar a probabilidade de um resultado na primeira etapa. Para os cálculos destas probabilidades utilizamos o teorema a seguir o qual é uma simples consequência da regra da probabilidade total e é conhecido como teorema de Bayes⁸.

Teorema 2.27 (Teorema da Bayes)

Seja (Ω, \mathcal{F}, P) um espaço de probabilidade e $A_1, \dots, A_n \in \mathcal{F}$ uma partição de Ω tal que $P(A_k) > 0, \forall k$. Seja também $B \in \mathcal{F}$ com $P(B) > 0$. Suponhamos conhecidas a priori as probabilidades $P(B|A_k)$ e $P(A_k)$ para todo k . Então

$$P(A_k|B) = \frac{P(A_k)P(B|A_k)}{\sum_{l=1}^n P(A_l)P(B|A_l)}, \quad (2.14)$$

para $k = 1, \dots, n$.

Demonstração: Utilizaremos o Teorema da Probabilidade Total. Observando que $\{A_k\}_{k=1}^n$ é uma partição, aplicando a definição de probabilidade condicional e aplicando o Teorema 2.26 obtemos

$$P(A_k|B) = \frac{P(A_k \cap B)}{P(B)} = \frac{P(A_k)P(B|A_k)}{\sum_{k=1}^n P(A_k)P(B|A_k)}.$$
■

Suponhamos que H_1, H_2, \dots são todas as causas que conduzem para o resultado de uma experiência aleatória. Temos em H_j o conjunto de resultados correspondentes à j -ésima causa. Assume-se que as probabilidades $P(H_j)$, $j = 1, 2, \dots$, chamadas de probabilidades *a priori*, podem ser conhecidas. Agora suponha que o resultado do experimento seja o evento B , de

⁸Thomas Bayes (1701 - 1761) foi um estatístico inglês, filósofo e ministro presbiteriano que é conhecido por formular um caso específico do teorema que leva seu nome: o teorema de Bayes.

probabilidade positiva. Essa informação leva a uma reavaliação das probabilidades anteriores. As probabilidades condicionais $P(H_j|B)$ são chamadas de probabilidades *a posteriori*. A expressão em (2.14) pode ser interpretada como uma regra que dá a probabilidade de que o evento B observado deve-se à ocorrência de H_j .

Exemplo 2.28 (*Continuação do Exemplo 2.27*)

Se no Exemplo 2.27 quisermos calcular a probabilidade de que seja a urna 2 a selecionada se a bola escolhida foi a branca, estamos interessados em calcular a probabilidade condicional $P(V|A)$. Segundo o Teorema de Bayes (Teorema 2.27), temos que

$$P(V|A) = \frac{P(V)P(A|V)}{P(U)P(A|U) + P(V)P(A|V) + P(W)P(A|W)} = \frac{1}{4}.$$

Claro, podemos questionarmos a probabilidade de que seja qualquer urna dado que a bola selecionada foi branca. Assim,

$$P(U|A) = P(W|A) = \frac{3}{8}.$$

O Teorema de Bayes é base para o conceito de Teoria Bayesiana, muito utilizada em modernas aplicações estatísticas. Vejamos agora um outro exemplo, neste percebe-se a utilidade do Teorema de Bayes em situações reais.

Exemplo 2.29

Suponha que a incidência de uma determinada doença em uma população seja de 0.001. Um teste de diagnóstico para essa doença existe, mas tem uma taxa de falso positivo de 0.03 e uma taxa de falso negativo de 0.01; isto é, 3% dos testes em pessoas não doentes indicarão a presença da doença, enquanto 1% dos testes em pessoas com a doença não indicarão a presença da doença. Se uma pessoa sorteada aleatoriamente da população for positiva, qual é a probabilidade de essa pessoa ter a doença?

Definamos o evento D para indicar a presença da doença e A para indicar um teste positivo para a doença. Então

$$P(D) = 0,001, \quad P(A|D^c) = 0,03 \quad \text{e} \quad P(A^c|D) = 0,01.$$

Pelo Teorema de Bayes, temos que

$$\begin{aligned} P(D|A) &= \frac{P(D)P(A|D)}{P(D)P(A|D) + P(D^c)P(A|D^c)} \\ &= \frac{0.001 \times 0.99}{0.001 \times 0.99 + 0.999 \times 0.03} = 0.0319. \end{aligned}$$

Este exemplo ilustra o perigo potencial de testes obrigatórios em uma população onde a taxa de falsos positivos excede a incidência da doença; esse perigo pode ser particularmente agudo em situações em que um teste positivo carrega um estigma social significativo.

2.3.1 Independência de eventos

Se (Ω, \mathcal{F}, P) é um espaço de probabilidades e $A, B \in \mathcal{F}$, com $P(B) > 0$, temos

$$P(A \cap B) = P(B)P(A|B) \quad (2.15)$$

pelo teorema da multiplicação (Teorema 2.26). Em diversas situações a informação fornecida pela evento B não afeta a probabilidade do evento A , isto é, $P(A|B) = P(A)$.

Na prática, geralmente tem-se o sentimento correto de que certos eventos devem ser estocasticamente independente ou então o modelo probabilístico seria um absurdo. Como os exemplos seguintes mostram, no entanto, existem situações em que a independência estocástica pode ser descoberta apenas por cálculo.

A independência de eventos é um conceito especialmente importante e específico da teoria das probabilidades.

Definição 2.12

Sejam A, B dois eventos do espaço de probabilidades (Ω, \mathcal{F}, P) . Os eventos A e B são ditos independentes se, e somente se,

$$P(A \cap B) = P(A)P(B). \quad (2.16)$$

Note que não fizemos qualquer restrição em $P(A)$ ou $P(B)$. Assim, a probabilidade condicional não está definida quando $P(A)$ ou $P(B)$ são zero, mas a independência está. Claramente, se $P(A) = 0$, então A é independente de todos os eventos $E \in \mathcal{F}$ tais que $P(E) > 0$. Além disso, qualquer evento $A \in \mathcal{F}$ com $P(A) > 0$ é independente de \emptyset e Ω .

Observe que esta definição permite-nos encontrar diretamente a probabilidade da interseção de dois eventos como produto das probabilidades respectivas. Significa que, sempre que aconteça a relação em (2.16) os eventos se dirão independentes.

Exemplo 2.30

Uma carta é escolhida aleatoriamente de um baralho de 52 cartas. Seja A o evento que a carta seja um ás e B o evento que a carta seja do naipe Espadas. Então

$$P(A) = \frac{4}{52} = \frac{1}{13}, \quad P(B) = \frac{13}{52} = \frac{1}{4}$$

$$P(A \cap B) = P(\{\text{Ás de Espadas}\}) = \frac{1}{52},$$

de modo que A e B são independentes.

Teorema 2.28

O evento A é independente de si mesmo se, e somente se, $P(A) = 0$ ou $P(A) = 1$.

Demonstração: Primeiro seja A independente de si mesmo, então, observemos que $P(A) = P(A \cap A) = P(A)P(A) = P(A)^2$ somente se $P(A) = 0$ ou $P(A) = 1$. Por outro lado se $P(A) = 0$ ou $P(A) = 1$ temos que $P(A \cap A) = P(A) = P(A)P(A)$. ■

Podemos observar que eventos com probabilidade zero ou um são independentes de qualquer outro. Se $P(A) = 0$, então $P(A \cap B) = 0$ sendo assim A e B independentes, qualquer seja o evento B . Se $P(A) = 1$, então

$$P(A \cap B) = P(B) - P(A^c \cap B),$$

e, dado que $A^c \cap B \subseteq A^c$, implica que

$$P(A^c \cap B) \leq P(A^c) = 0.$$

temos então que $P(A^c \cap B) = 0$ e que

$$P(A \cap B) = P(B) = P(A)P(B),$$

logo A e B são independentes, qualquer seja B .

Teorema 2.29

Sejam A e B dois eventos independentes, então

$$P(A|B) = P(A) \quad \text{se} \quad P(B) > 0,$$

e

$$P(B|A) = P(B) \quad \text{se} \quad P(A) > 0.$$

Demonstração: Segue da observação em (2.15). ■

Exemplo 2.31

Suponha temos uma moeda que, quando arremessada, aparece cara com probabilidades p e aparece coroa com probabilidade q . Suponha agora que esta moeda seja lançada duas vezes. Usando a interpretação frequentista de probabilidade é razoável atribuir ao resultado $(Cara, Cara)$ a probabilidade p^2 , para o resultado $(Cara, Coroa)$ a probabilidade pq e assim por diante. Seja E o evento de obter cara no primeiro lance e F o evento de que aparece coroa no segundo lance. Vamos agora verificar que com as probabilidades acima estes dois eventos são independentes, como esperado. Temos

$$P(E) = p^2 + pq = p \quad \text{e} \quad P(F) = pq + q^2 = q.$$

Finalmente $P(E \cap F) = pq$, logo $P(E \cap F) = P(E)P(F)$.

Exemplo 2.32

Muitas vezes é intuitivamente claro quando dois eventos são independentes, mas nem sempre, que nem nesta situação. No Exemplo 2.31, seja A o evento “o primeiro lance é uma cara” e B o evento “os dois resultados são os mesmos”. Então

$$\begin{aligned} P(B|A) &= \frac{P(A \cap B)}{P(A)} = \frac{P(\{Cara, Cara\})}{P(\{Cara, Cara\}, \{Cara, Coroa\})} \\ &= \frac{1/4}{1/2} = \frac{1}{2} = P(B). \end{aligned}$$

Portanto, A e B são independentes, mas o resultado não foi tão óbvio.

Teorema 2.30

Se os eventos A e B são independentes, então os eventos A e B^c , A^c e B e A^c e B^c também são independentes.

Demonstração :

$$\begin{aligned} P(A^c \cap B) &= P[B \setminus (A \cap B)] \\ &= P(B) - P(A \cap B), \quad \text{devido a } B \supseteq (A \cap B) \\ &= P(B)[1 - P(A)] = P(A^c)P(B). \end{aligned}$$

Similarmente nas outras situações. ■

Gostaríamos de salientar que a independência de eventos não deve ser confundido com eventos disjuntos ou eventos mutuamente exclusivos. Dois eventos, cada um com probabilidade não nula, mutuamente exclusivos, serão dependentes desde que a ocorrência de um interfira na ocorrência do outro. Da mesma forma, se A e B são independentes e $P(A) > 0$, $P(B) > 0$, então A e B não podem ser mutuamente exclusivos.

Exemplo 2.33

Considere famílias com dois filhos e assuma que todas as quatro distribuições possíveis do sexo dos filhos MM , MF , FM , FF , onde M representa menino e F para a menina, sejam igualmente prováveis. Seja E o caso de que uma família escolhida aleatoriamente tenha no máximo uma menina e F o caso em que a família tenha crianças de ambos os sexos. Então

$$P(E) = \frac{3}{4}, \quad P(F) = \frac{1}{2} \quad \text{e} \quad P(E \cap F) = \frac{1}{2},$$

de modo que E e F não são independentes.

Por outro lado, consideremos famílias com três filhos. Assumindo que cada um dos oito possíveis resultados sejam igualmente prováveis, temos que $P(E) = \frac{4}{8}$, $P(F) = \frac{6}{8}$ e $P(E \cap F) = \frac{3}{8}$, de modo que E e F são independentes.

Uma extensão óbvia do conceito de independência entre dois eventos A e B a uma determinada coleção \mathcal{U} de eventos é exigir que todos os eventos distintos em \mathcal{U} sejam independentes.

Definição 2.13

Seja \mathcal{U} uma coleção de eventos de \mathcal{F} . Dizemos que os eventos em \mathcal{U} são independentes a pares se, e somente se, para todo par de eventos distintos $A, B \in \mathcal{U}$,

$$P(A \cap B) = P(A)P(B). \quad (2.17)$$

Exemplo 2.34

Considere as seis permutações das letras a, b, c , bem como as três triplas (a, a, a) , (b, b, b) e (c, c, c) . Tomemos estas nove triplas como os pontos de um espaço amostral e atribuamos probabilidade $1/9$ a cada um. Denote por A_k o evento que o k -ésimo lugar seja ocupado pela letra a . Obviamente cada um destes três eventos tem probabilidade $1/3$, enquanto

$$P(A_1 A_2) = P(A_1 A_3) = P(A_2 A_3) = \frac{1}{9}.$$

Os três eventos são, portanto, independentes a pares, mas os três não são independentes porque também $P(A_1 A_2 A_3) = \frac{1}{9}$. A ocorrência de A_1 e A_2 implica a ocorrência de A_3 e assim A_3 não é independente de $A_1 A_2$.

Um conceito muito mais forte e mais útil é a independência mútua ou completa.

Definição 2.14

A família de eventos \mathcal{U} é dita ser mutualmente ou completamente independente se, e somente se, para toda coleção finita $\{A_{i_1}, A_{i_2}, \dots, A_{i_k}\}$ de elementos de \mathcal{U} , se satisfaz a relação

$$P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = \prod_{j=1}^k P(A_{i_j}). \quad (2.18)$$

No que se segue, vamos omitir o adjetivo mútuo ou completo e falar de eventos independentes. É claro que a partir da Definição 2.14, a fim de verificar a independência de n eventos A_1, A_2, \dots, A_n é preciso verificar as seguintes $2^n - n - 1$ relações.

$$P(A_i \cap A_j) = P(A_i)P(A_j), \quad i \neq j, i, j = 1, 2, \dots, n$$

$$P(A_i \cap A_j \cap A_k) = P(A_i)P(A_j)P(A_k), \quad \begin{array}{l} i \neq j \neq k, \\ i, j, k = 1, 2, \dots, n \end{array}$$

$$\vdots$$

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1)P(A_2) \dots P(A_n).$$

O primeiro desses requisitos é a independência em pares. Portanto independência implica independência de pares, mas não vice-versa.

Exemplo 2.35

Tome quatro bolas idênticas. Na primeira, escreva o símbolo $A_1A_2A_3$ e em cada uma das outras três escreva A_1 , A_2 , A_3 , respectivamente. Coloque as quatro bolas em uma urna e escolha uma aleatoriamente. Seja E_i o evento de que o símbolo A_i apareça na bola escolhida. Então

$$P(E_1) = P(E_2) = P(E_3) = \frac{1}{2},$$

$$P(E_1 \cap E_2) = P(E_2 \cap E_3) = P(E_1 \cap E_3) = \frac{1}{4}$$

e

$$P(E_1 \cap E_2 \cap E_3) = \frac{1}{4}.$$

Segue que, embora os eventos E_1 , E_2 , E_3 não são independentes, eles são independentes a pares.

Vimos que independência a pares não implica independência coletiva. Vejamos agora uma situação diferentes.

Exemplo 2.36

Neste exemplo $P(E_1 \cap E_2 \cap E_3) = P(E_1)P(E_2)P(E_3)$, no entanto os eventos E_1 , E_2 , E_3 não são independentes a pares e, portanto, não são independentes. Seja $\Omega = \{1, 2, 3, 4\}$ e seja p_i a probabilidade atribuída ao evento $\{i\}$, $i = 1, 2, 3, 4$. Se

$$p_1 = \frac{\sqrt{2}}{2} - \frac{1}{4}, \quad p_2 = \frac{1}{4}, \quad p_3 = \frac{3}{4} - \frac{\sqrt{2}}{2}, \quad p_4 = \frac{1}{4}$$

e

$$E_1 = \{1, 3\}, \quad E_2 = \{2, 3\}, \quad E_3 = \{3, 4\},$$

temos que

$$\begin{aligned} P(E_1 \cap E_2 \cap E_3) &= P(\{3\}) = \frac{3}{4} - \frac{\sqrt{2}}{2} = \frac{1}{2} \left(1 - \frac{\sqrt{2}}{2}\right) \left(1 - \frac{\sqrt{2}}{2}\right) \\ &= (p_1 + p_2)(p_2 + p_3)(p_3 + p_4) \\ &= P(E_1)P(E_2)P(E_3). \end{aligned}$$

Mas,

$$P(E_1 \cap E_2) = \frac{3}{4} - \frac{\sqrt{2}}{2} \neq P(E_1)P(E_2)$$

do qual segue que E_1 , E_2 , E_3 não são independentes.

O seguinte resultado tem o nome de Borel e Francesco Paolo Cantelli⁹, que deram declarações ao lema nas primeiras décadas do século XX. Este resultado, conhecido como lema

⁹Francesco Paolo Cantelli (1875 - 1966) foi um matemático italiano.

de Borel-Cantelli afirma que, sob certas condições, um evento terá probabilidade de zero ou um.

Teorema 2.31 (Lema e Borel-Cantelli)

Sejam A_1, A_2, \dots eventos no espaço de probabilidade (Ω, \mathcal{F}, P) e seja $A = \limsup_{n \rightarrow \infty} A_n$.

(a) Se $\sum_{n=1}^{\infty} P(A_n) < \infty$, então $P(A) = 0$.

(b) Se os A_n são independentes e $\sum_{n=1}^{\infty} P(A_n) = \infty$, então $P(A) = 1$.

Demonstração: (a) Para cada n natural, temos que $P(A) \leq P\left(\bigcup_{k=n}^{\infty} A_k\right) \leq \sum_{k=n}^{\infty} P(A_k)$.

Dado que a série $\sum_{n=1}^{\infty} P(A_n)$ é convergente, então a série dos restos $\sum_{k=n}^{\infty} P(A_k)$ tende a zero quando $n \rightarrow \infty$. Assim

$$\lim_{n \rightarrow \infty} P(A) \leq \lim_{n \rightarrow \infty} \sum_{k=n}^{\infty} P(A_k) = 0.$$

Isto implica que $P(A) = 0$.

(b) Nesta situação é suficiente demonstrar que para todo número natural n se satisfaz que $P(\bigcup_{k=n}^{\infty} A_k) = 1$, já que a interseção enumerável de eventos com probabilidade 1 tem probabilidade 1. Para cada $m > n$

$$\begin{aligned} 1 - P\left(\bigcup_{k=n}^{\infty} A_k\right) &\leq 1 - P\left(\bigcup_{k=n}^m A_k\right) = P\left(\bigcap_{k=n}^m A_k^c\right) = \\ &= \prod_{k=n}^m [1 - P(A_k)] \leq \exp\left(-\sum_{k=n}^m P(A_k)\right). \end{aligned}$$

Para obter a última expressão utilizou-se a desigualdade $1 - x \leq \exp(-x)$, válida para qualquer número real x . Dado que $\sum_{n=1}^{\infty} P(A_n) = \infty$ o lado direito tende a zero quando m tende ao infinito. Portanto, $P(\bigcup_{k=n}^{\infty} A_k) = 1$, qualquer seja n . Então $P(A) = 1$. ■

Observemos que o item (b) pode não ser válido caso os eventos A_1, A_2, \dots não forem independentes. Por exemplo, se $A_n = A$, para qualquer n com $0 < P(A) < 1$, temos que $\sum_{n=1}^{\infty} P(A_n) = \infty$. Por outro lado, o evento $\limsup_{n \rightarrow \infty} A_n = A$ mas $P(A) < 1$.

No seguinte exemplo vamos utilizar a chamada função zeta de Riemann¹⁰. A função zeta de Riemann ou função zeta de Euler-Riemann, ζ é uma função de uma variável complexa s

¹⁰Georg Friedrich Bernhard Riemann (1826-1866). Foi um matemático alemão, com contribuições fundamentais para a análise e a geometria diferencial.

continua e definida como a soma da série

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}.$$

Esta série converge quando a parte real de s é maior que 1. Um valor específico desta função que será interessante para nós é quando $\zeta(2) = \pi^2/6$.

Exemplo 2.37

Por exemplo, suponhamos que $\{A_n\}_{n \geq 1}$ seja uma sequência de eventos tais que

$$P(A_n) = \frac{1}{n^2}, \quad \forall n \geq 1.$$

A probabilidade de que A_n ocorra infinitamente para n grande é equivalente à probabilidade da interseção de infinitos eventos da sequência, isto é, equivalente à probabilidade da ocorrência do evento $\bigcap_{n \geq 1} A_n$ para n suficientemente grande. A interseção de tais infinitos eventos é um conjunto de resultados comuns a todos eles.

Entretanto, a soma $P(A_n)$ é uma série convergente, de fato, é uma função zeta de Riemann que tende a $\pi^2/6$ e, então, o Lema de Borel-Cantelli (Teorema 2.31) estabelece que o conjunto de resultados que são comuns a tais infinitamente muitos eventos ocorrem com probabilidade zero. Por isso, a probabilidade de A_n ocorrendo infinitamente é 0. Quase certamente, isto é, com probabilidade 1 $P(A_n)$ é não nula para muitos n finitos.

2.3.2 Permutabilidade de eventos

A noção de permutabilidade foi introduzida por De Finetti¹¹ em 1930 e recuperada por Savage¹² em 1950. Grosso modo, o termo permutabilidade significa que é trocável: diz-se de dois eventos que podem ser alterados sem afetar os resultados. Aqui, dizemos que dois eventos podem ser alterados se a ordem em que eles acontecem é irrelevante para as probabilidades. A decisão de aceitar eventos como permutáveis é uma espécie de confissão do observador que ele não consegue distinguir entre eles.

Diz-se que dois eventos A e B são permutáveis se $P(A^c \cap B) = P(A \cap B^c)$, o que significa que há indiferença em relação à ordem, porque ambas as interseções descrevem a ocorrência de exatamente um dos dois eventos, na primeira ou na segunda instância.

Exemplo 2.38

Imagine que se observa o arremesso de uma moeda 10 vezes, dos quais em 9 deles observou-se cara. A ideia de permutabilidade permitirá uma interpretação objetivista deste resultado. A principal premissa que vamos fazer é que nós não nos importamos quando ocorrem os resultados cara ou coroa, mas apenas quantos deles temos observado.

Intuitivamente, isso significa que em n arremessos da mesma moeda, cada sequência particular de caras e coroas com m caras e $n - m$ coroas tem a mesma probabilidade de ocorrer. A ordem é irrelevante.

¹¹Bruno de Finetti (1906-1985), foi um matemático italiano, notável por seus estudos sobre probabilidade.

¹²Leonard Jimmie Savage (1917-1971) foi um matemático estadunidense.

Definição 2.15

Suponha que $\mathcal{A} = \{A_i : i \in \mathcal{I}\}$ é um conjunto de eventos de um experimento aleatório onde \mathcal{I} é um conjunto de índices enumeráveis. A coleção \mathcal{A} é dita ser permutável se a probabilidade da interseção de um número finito dos eventos depende apenas do número de eventos. Isto é, a coleção \mathcal{A} é permutável se dados \mathcal{J} e \mathcal{K} , subconjuntos finitos de \mathcal{I} tais que $\#\mathcal{J} = \#\mathcal{K}$, temos que

$$P\left(\bigcap_{j \in \mathcal{J}} A_j\right) = P\left(\bigcap_{k \in \mathcal{K}} A_k\right). \quad (2.19)$$

A permutabilidade de n eventos significa que a ocorrência de qualquer interseção de k de tais eventos tem a mesma probabilidade, ou seja, essa probabilidade não depende de sua posição, mas exclusivamente em k . É uma propriedade de simetria. Uma sequência infinita de variáveis aleatórias é infinitamente permutável se toda coleção finita de suas variáveis é permutável.

Teorema 2.32

Seja \mathcal{A} uma sequência permutável. Então se $\mathcal{B} \subseteq \mathcal{A}$, \mathcal{B} é permutável. Inversamente, se cada $\mathcal{B} \subseteq \mathcal{A}$ finito é permutável, então \mathcal{A} é permutável.

Demonstração: Exercício. ■

Para uma coleção de eventos passíveis de troca, o princípio de inclusão-exclusão para a probabilidade de uma união é muito mais simples do que a versão geral.

Teorema 2.33

Suponhamos que $\{A_1, A_2, \dots, A_n\}$ seja uma coleção de eventos permutáveis. Para $\mathcal{J} \subseteq \{1, 2, \dots, n\}$ com $\#\mathcal{J} = k$. Então

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{k=1}^n (-1)^{k-1} \binom{n}{k} P\left(\bigcap_{j \in \mathcal{J}} A_j\right). \quad (2.20)$$

Demonstração: Aplicar o Teorema de Inclusão-Exclusão (Teorema 2.19). ■

Observemos que se os eventos A e B forem independentes, então são permutáveis. Isto deve-se a que

$$P(A \cap B) = P(A)P(B) = P(B \cap A),$$

do qual podemos afirmar que se A_1, A_2, \dots for uma coleção de eventos independentes, então A_1, A_2, \dots é uma coleção de eventos permutáveis. O contrário não é verdade em geral, como veremos no seguinte exemplo que deve à Pólya¹³.

Exemplo 2.39 (Urna de Pólya)

Uma situação ilustrativa de eventos permutáveis porém não independentes é o seguinte, conhecido como modelo urna de Pólya. Imagine que dada uma urna com N_0 bolas pretas e N_1 bolas brancas. O algoritmo seguinte é executado diversas vezes:

- 1- Retire uma bola aleatoriamente da urna e observe sua cor,
- 2- Se a bola é preta definimos o evento $A_i = \text{bola branca}$ e $A_i = \text{bola negra}$ caso contrário,
- 3- Acrescente $i = i + 1$. Coloque n bolas da mesma cor na urna. Ir ao passo 1.

Se $n = 1$, então a sequência A_1, A_2, \dots é de eventos independentes e a amostragem é com reposição, quando $n = 0$ esta é uma amostragem sem reposição. Para ver que os eventos são permutáveis porém não independentes observe o seguinte:

$$\begin{aligned} P(A_1 = \text{bola branca}, A_2 = \text{bola branca}, A_3 = \text{bola negra}, A_4 = \text{bola branca}) &= \\ &= \frac{N_0}{N_0 + N_1} \times \frac{N_0 + n}{N_0 + N_1 + n} \times \frac{N_1}{N_0 + N_1 + 2n} \times \frac{N_0 + 2n}{N_0 + N_1 + 3n}, \end{aligned}$$

também

$$\begin{aligned} P(A_1 = \text{bola branca}, A_2 = \text{bola branca}, A_3 = \text{bola negra}, A_4 = \text{bola branca}) &= \\ &= \frac{N_0}{N_0 + N_1} \times \frac{N_1}{N_0 + N_1 + n} \times \frac{N_0 + n}{N_0 + N_1 + 2n} \times \frac{N_0 + 2n}{N_0 + N_1 + 3n}, \end{aligned}$$

as quais são, evidentemente, iguais. No entanto,

$$P(A_1 = \text{bola branca}, A_2 = \text{bola branca}) \neq P(A_1 = \text{bola branca})P(A_2 = \text{bola branca}),$$

$$\text{já que } P(A_1 = \text{bola branca}) = \frac{N_0}{N_0 + N_1} \text{ e}$$

$$\begin{aligned} P(A_2 = \text{bola branca}) &= \\ &= P(A_1 = \text{bola branca})P(A_2 = \text{bola branca}|A_1 = \text{bola branca}) + \\ &\quad P(A_1 = \text{bola negra})P(A_2 = \text{bola branca}|A_1 = \text{bola negra}) \\ &= \frac{N_0}{N_0 + N_1} \times \frac{N_0 + n}{N_0 + N_1 + n} + \frac{N_1}{N_0 + N_1} \times \frac{N_0}{N_0 + N_1 + n}. \end{aligned}$$

¹³George Pólya (1887-1985) foi um eminente matemático húngaro.

2.4 Exercícios

Exercícios da Seção 2.1

1. Sejam $A = \{x \in \mathbb{R} : 0 \leq x \leq 1\}$ e $B = \{x \in \mathbb{R} : 1/2 \leq x \leq 2\}$. Encontrar $A \setminus B$.
2. Sejam A, B, C e D subconjuntos de Ω . Provar que:
 - (a) $A \cap B \subseteq A \subseteq A \cup B$,
 - (b) Se $A \subseteq C$ e $B \subseteq D$ então $A \cap B \subseteq C \cap D$ e $A \cup B \subseteq C \cup D$,
 - (c) $A \subseteq C$ e $B \subseteq C$ se, e somente se, $A \cup B \subseteq C$.
3. Seja \mathcal{F} uma álgebra de eventos. Prove que \mathcal{F} é fechada para diferenças de eventos, ou seja, se $A, B \in \mathcal{F}$, então $A \setminus B \in \mathcal{F}$, onde $A \setminus B = A \cap B^c$.
4. Prove que no Exemplo 2.3 as classes de subconjuntos definidas são álgebras.
5. Seja Ω um conjunto não vazio.
 - (a) Prove que se \mathcal{A} e \mathcal{B} são σ -álgebras de subconjuntos de Ω , então $\mathcal{A} \cap \mathcal{B}$ é também σ -álgebra de subconjuntos de Ω .
 - (b) Seja \mathcal{C} uma classe de subconjuntos de Ω . Mostre que existe pelo menos uma σ -álgebra que contém \mathcal{C} .
6. Provar que se $A \setminus B = A$ então $A \cap B = \emptyset$.
7. Provar que $A \setminus B = \emptyset$ se, e somente se, $A \subseteq B$.
8. Seja E um conjunto que contém $A \cup B$. Prove que:
 - (a) $A \cap (E \setminus A) = \emptyset$,
 - (b) $A \subseteq B$ se, e somente se, $E \setminus B \subseteq E \setminus A$.
9. Seja $\Omega = \{a, b, c, d, e, f\}$ e $\mathcal{C} = \{\{b, e\}, \{f\}\}$. Qual é a álgebra gerada por \mathcal{C} ? e qual a σ -álgebra gerada por \mathcal{C} ?
10. Suponha que \mathcal{A}_n são álgebras satisfazendo $\mathcal{A}_n \subseteq \mathcal{A}_{n+1}$. Mostre que $\bigcup_{n=1}^{\infty} \mathcal{A}_n$ é álgebra.
11. Prove que no Exemplo 2.4 \mathcal{F}_0 é álgebra mas não é σ -álgebra.
12. Seja \mathcal{A} uma σ -álgebra de subconjuntos de Ω e seja $B \in \mathcal{A}$. Prove que

$$\mathcal{F} = \{A \cap B : A \in \mathcal{A}\},$$

é σ -álgebra de subconjuntos de B .

13. Demonstrar o Teorema 2.3
14. Sejam $A, B \subset \Omega$, tais que $A \subseteq B$. Prove que a classe $\mathcal{F} = \{\emptyset, A, B, A^c, B^c, B \setminus A, (B \setminus A)^c, \Omega\}$ é uma álgebra.
15. Sejam $\Omega = \{a, b, c, d\}$, $A = \{a, b\}$ e $B = \{c, d\}$. Definamos a coleção $\mathcal{C} = \{A, B\}$. Percebemos que \mathcal{C} não é uma álgebra. Encontre $\sigma(\mathcal{C})$.
16. Seja Ω um espaço amostral não vazio e $\mathcal{M} = \{A \subset \Omega : A \text{ ou } A^c\}$ é enumerável. Prove que \mathcal{M} é uma σ -álgebra definida em Ω .
17. Considere $\{\mathcal{F}_i, i \in \mathcal{I}\}$ uma família de σ -álgebras definidas em Ω . Prove que

$$\mathcal{F} = \bigcap_{i \in \mathcal{I}} \mathcal{F}_i$$

é σ -álgebra de subconjuntos de Ω . O conjunto de índices \mathcal{I} pode ser não enumerável.

18. Considere \mathcal{F} uma σ -álgebra de eventos de Ω . Prove que a coleção de eventos $\mathcal{F}^c = \{F^c : F \in \mathcal{F}\}$ é uma σ -álgebra de eventos. Verifique se \mathcal{F} e \mathcal{F}^c coincidem.
19. Seja \mathcal{C} um conjunto não enumerável e \mathcal{F} o conjunto dos subconjuntos finitos de \mathcal{C} . Descreva a σ -álgebra gerada por \mathcal{F} .
20. Demonstre que $\sigma(\sigma(\mathcal{C})) = \sigma(\mathcal{C})$, onde \mathcal{C} é uma coleção de subconjuntos do espaço amostral Ω .
21. Prove que $\sigma(\mathcal{C}_1 \cup \mathcal{C}_2) = \sigma(\sigma(\mathcal{C}_1) \cup \sigma(\mathcal{C}_2))$, onde \mathcal{C}_1 e \mathcal{C}_2 são duas coleções de subconjuntos de Ω .
22. Considere $\{A_i\}_{i=1}^n$ uma coleção de conjuntos e $B_k = \cup_{i=1}^k A_i$, $k = 1, 2, \dots, n$. Mostre que

$$\sigma(\{B_1, B_2, \dots, B_n\}) = \sigma(\{A_1, A_2, \dots, A_n\}).$$

23. Seja Ω um conjunto qualquer e $\mathcal{C} = \{\{x\} : x \in \Omega\}$, ou seja, o conjunto \mathcal{C} é formado por todos os conjuntos simples em Ω . Prove que $\mathcal{F} = \{A \subset \Omega : A \text{ é finito ou } A^c \text{ é finito}\}$ é σ -álgebra de subconjuntos de Ω , contém \mathcal{C} e que seja mínima em relação a todas as outras σ -álgebra que contém \mathcal{C} .
24. Suponhamos que $\{E_1, E_2, \dots\}$ é uma coleção arbitrária de subconjuntos abertos de \mathbf{R} e que $\{F_1, F_2, \dots\}$ é uma coleção arbitrária de subconjuntos fechados de \mathbf{R} . Prove que
 - (a) $\bigcup_{k=1}^{\infty} E_k$ é um conjunto aberto,
 - (b) $\bigcap_{k=1}^{\infty} E_k$ não é necessariamente um conjunto aberto,
 - (c) $\bigcap_{k=1}^{\infty} F_k$ é necessariamente um conjunto fechado,
 - (d) $\bigcup_{k=1}^{\infty} F_k$ não é necessariamente um conjunto fechado.
25. Demonstrar o Teorema 2.8.
26. Demonstrar o Teorema 2.10.
27. Demonstrar o Teorema 2.12.
28. No Exemplo 2.13 provar que $x \notin \tilde{\mathbb{Q}}^y$.

Exercícios da Seção 2.2

1. Um experimento aleatório consiste no lançamento de dois dados equilibrados e como resultado observamos o número mínimo obtido de suas faces. Construa um modelo probabilístico associado.
2. Seja (Ω, \mathcal{F}, P) um espaço de probabilidade. Considere uma sequência de eventos aleatórios $\{A_n\}_{n=1}^{\infty}$ em \mathcal{F} . Defina o evento B_m : “o primeiro evento a ocorrer da sequência $\{A_n\}_{n=1}^{\infty}$ é A_m ”.
 - a) Expresse B_m em função dos A_n . B_m é um evento aleatório? Por quê?
 - b) Os eventos B_m são disjuntos?
 - c) Quem é o evento $\bigcup_{m=1}^{\infty} B_m$?
3. Os eventos A e B no espaço de probabilidades (Ω, \mathcal{F}, P) são tais que $A \subseteq B$. Qual é $P(A \cup B)$? Qual é $P(A \cap B)$? Qual é $P(A \setminus B)$.
4. Uma caixa contém 1.000 lâmpadas. A probabilidade de que exista pelo menos uma lâmpada defeituosa na caixa é de 0,1 e a probabilidade de que existam, pelo menos, duas lâmpadas defeituosas é de 0,05. Determine a probabilidade em cada um dos seguintes casos:
 - (a) A caixa não contém lâmpadas defeituosas.
 - (b) A caixa contém exatamente uma lâmpada defeituosa.
 - (c) A caixa contém no máximo uma lâmpada defeituosa.

5. Em um determinado subúrbio residencial, 60% de todas as casas assinam o jornal metropolitano publicado em uma cidade próxima, 80% assinam o jornal local e 50% assinam ambos. Qual é a probabilidade de que uma família assina:
 - a) pelo menos um dos dois jornais?
 - b) exatamente um dos dois jornais?
 - c) nenhum dos dois jornais?
6. A rota usada por um determinado motorista no deslocamento para o trabalho contém dois cruzamentos com semáforos. As probabilidades de paragem no primeiro sinal e no segundo sinal são 0.4 e 0.5, respectivamente. Além disso, a probabilidade de parar em pelo menos um dos dois sinais é de 0.6. Qual é a probabilidade de que ele deve parar:
 - a) em ambos os sinais?
 - b) no primeiro sinal, mas não no segundo?
 - c) em exatamente um sinal?
7. Seja (Ω, \mathcal{F}, P) um espaço de probabilidades e $A, B, C \in \mathcal{F}$, tais que $A \cap B \subseteq C$. Provar que $P(C^c) \leq P(A^c) + P(B^c)$.
8. Seja (Ω, \mathcal{F}, P) um espaço de probabilidades e $A, B \in \mathcal{F}$ eventos tais $P(A) = P(B) = \frac{1}{2}$ e $P(A^c \cap B^c) = \frac{1}{3}$. Encontre $P(A \cup B^c)$.
9. Prove que $P(A \cap B) - P(A)P(B) = P(A^c)P(B) - P(A^c \cap B)$.
10. Prove que $P(A \cup B \cup C) = P(A) + P(A^c \cap B) + P(A^c \cap B^c \cap C)$.
11. Se $P(E) = 0.9$ e $P(F) = 0.8$, mostre que $P(E \cap F) \geq 0.7$.
12. Demonstrar o Teorema 2.20.
13. Seja $\{A_n\}_{n=1}^\infty$ uma sequência de eventos não decrescentes em \mathcal{F} tais que convergem para o evento A , isto é, $A_n \subseteq A_{n+1}$ para qualquer valor de n e $\bigcup_{n=1}^\infty A_n = A$. Prove que

$$\lim_{n \rightarrow \infty} P(A_n) = P(A).$$

14. Sejam $\{A_n\}_{n=1}^\infty$ e $\{B_n\}_{n=1}^\infty$ duas sequências de eventos aleatórios do mesmo espaço de probabilidade (Ω, \mathcal{F}, P) tais que $\lim_{n \rightarrow \infty} P(A_n) = 1$ e $\lim_{n \rightarrow \infty} P(B_n) = p$. Prove que $\lim_{n \rightarrow \infty} P(A \cap B) = p$.
15. Seja $\{A_n\}_{n=1}^\infty$ uma sequência de eventos aleatórios no espaço de probabilidade (Ω, \mathcal{F}, P) . Prove que:
 - (a) Se $P(A_n) = 0$ para todo $n = 1, 2, \dots$, então $P(\bigcup_{n=1}^\infty A_n) = 0$,
 - (b) Se $P(A_n) = 1$ para todo $n = 1, 2, \dots$, então $P(\bigcap_{n=1}^\infty A_n) = 1$.
16. Considere (Ω, \mathcal{F}, P) um espaço de probabilidade. Seja $\{P_n\}_{n=1}^\infty$ uma sequência de funções de probabilidades definidas em \mathcal{F} e $\{a_n\}_{n=1}^\infty$ uma sequência de números reais não-negativos tal que $\sum_{n=1}^\infty a_n = 1$. Prove que

$$P(E) = \sum_{n=1}^{\infty} a_n P_n(E), \quad \forall E \in \mathcal{F}$$

é também uma função de probabilidade definida em \mathcal{F} .

17. Seja $\{A_n\}_{n \geq 1}$ uma sequência de conjuntos disjuntos 2 a 2 ($A_i \cap A_j = \emptyset$, $i \neq j$) e P uma probabilidade. Demonstre que $\lim_{n \rightarrow \infty} P(A_n) = 0$.
18. Uma moeda com probabilidade p de cara em cada lançamento é lançada infinitas vezes, de maneira independente. Definimos os eventos

A_n : Ocorre pelo menos uma cara nos n primeiros lançamentos.

A: Ocorre pelo menos uma cara.

Prove que

- a) $\{A_n\}_{n=1}^{\infty}$ é uma sequência não decrescente que converge para o evento A ($A_n \nearrow A$), isto é, prove que $A_n \subseteq A_{n+1}$ e $\cup_{n=1}^{\infty} A_n = A$.

b)

$$P(A) = \begin{cases} 0 & \text{se } p = 0 \\ 1 & \text{se } 0 < p \leq 1 \end{cases}$$

19. Prove que

- a) $(\lim_{n \rightarrow \infty} \sup A_n)^c = \lim_{n \rightarrow \infty} \inf A_n^c$,
 b) $(\lim_{n \rightarrow \infty} \inf A_n)^c = \lim_{n \rightarrow \infty} \sup A_n^c$,
 c) $\lim_{n \rightarrow \infty} \sup A \cap B_n = A \cap \lim_{n \rightarrow \infty} \sup B_n$,
 d) $\lim_{n \rightarrow \infty} \sup(A_n \cup B_n) = \lim_{n \rightarrow \infty} \sup A_n \cup \lim_{n \rightarrow \infty} \sup B_n$,
 e) $\lim_{n \rightarrow \infty} \sup(A_n \cap B_n) \subseteq \lim_{n \rightarrow \infty} \sup A_n \cap \lim_{n \rightarrow \infty} \sup B_n$.

20. Seja $\Omega = \mathbb{R}$ e $\mathcal{F} = \mathcal{B}(\mathbb{R})$. Em cada um dos seguintes casos P define uma medida de probabilidade em (Ω, \mathcal{F}) ?

- a) Para cada intervalo I , seja $P(I) = 1$ se I for um intervalo de comprimento finito e $P(I) = 0$ se I for um intervalo infinito.
 b) Para cada intervalo I , seja $P(I) = 0$ se $I \subseteq (-\infty, 1)$ e

$$P(I) = \int_I \frac{1}{2} dx,$$

se $I \subseteq [1, \infty)$. Observe que se $I = I_1 + I_2$, onde $I_1 \subseteq (-\infty, 1)$ e $I_2 \subseteq [1, \infty)$, então $P(I) = P(I_2)$.

21. Seja P uma medida de probabilidade. Verifique se as seguintes funções são também medidas de probabilidade:

- a) $1 - P$,
 b) $(1 + P)/2$,
 c) P^2 ,
 d) $|P|$.

22. O espaço amostral é $\Omega = \{1, 2, \dots, n\}$ e a σ -álgebra $\mathcal{P}(\Omega)$. Nestas condições, verifique em cada caso se a função P é uma probabilidade.

Para cada $A \in \mathcal{P}(\Omega)$ defina:

a)

$$P(A) = \sum_{k \in A} \frac{2k}{n(n+1)},$$

b)

$$P(A) = \prod_{k \in A} \left(1 - \frac{1}{k}\right).$$

23. Desigualdade de Kounias¹⁴. Demonstre que

$$P\left(\bigcup_{i=1}^n A_i\right) \leq \min \left\{ \sum_{\substack{i=1 \\ i \neq j}}^n P(A_i) - P(A_i \cap A_j) \right\}.$$

24. Prove que no Exemplo 2.17 a função P é função de probabilidade.

25. Demonstre o Teorema 2.15.

26. Demonstre o Teorema 2.22

Exercícios da Seção 2.3

1. Se $P(B^c) = 1/4$ e $P(A|B) = 1/2$, quanto é $P(A \cap B)$?
2. Sejam A e B dois eventos tais que $P(A) = p_1 > 0$, $P(B) = p_2 > 0$ e $p_1 + p_2 > 1$. Mostre que $P(B|A) \geq 1 - (1 - p_2)/p_1$.
3. Dois dígitos são escolhidos aleatoriamente sem reposição do conjunto de números inteiros $\{1, 2, 3, 4, 5, 6, 7, 8\}$.
 - (i) Encontrar a probabilidade de que ambos os dígitos sejam maiores do que 5.
 - (ii) Mostre que a probabilidade de que a soma dos dígitos seja igual a 5 é a mesma que a probabilidade de que sua soma exceda 13.
4. Provar o Teorema 2.26.
5. Um dado equilibrado é lançado três vezes consecutivas y come resultado observamos que a soma dos três números é 9. Encontre a probabilidade de que no primeiro lançamentos tivemos obtido o número 5.
6. Seja C_1, C_2, \dots, C_n uma partição de Ω . Mostrar que

$$P(A|B) = \sum_{k=1}^n P(C_k|B)P(A|C_k \cap B).$$

7. Seja (Ω, \mathcal{F}, P) um espaço de probabilidade. Prove que
 - (i) Mostre que se A e B são eventos tais que $P(A) < 1$, $P(B) > 0$ e $P(A|B) = 1$, então $P(B^c|A^c) = 1$.
 - (ii) Prove que se E, F e G são eventos tais que $P(F \cap G) > 0$ e $P(F \cap G^c) > 0$, então

$$P(E|F) = P(E|F \cap G)P(G|F) + P(E|F \cap G^c)P(G^c|F).$$

8. Sejam A e B dois eventos do mesmo espaço de probabilidade e ainda $P(B) > 0$. Prove que
 - (i) $P(A|B) = 0$ se A e B são disjuntos.
 - (ii) $P(A^c|B) = 1 - P(A|B)$.
 - (iii) $P(B|B) = 1$.
 - (iv) $P(A) = P(A|B)P(B) + P(A|B^c)P(B^c)$.
9. Assuma que E e F são dois eventos com probabilidades positivas. Mostre que se $P(E|F) = P(E)$ então $P(F|E) = P(F)$.

¹⁴Kounias, E.G. (1968). Bounds for the probability of a union, with applications. Ann. Math. Statist, 39, pg. 2154-2158.

10. Prove que $P(A|B) \geq 1 - P(A^c)/P(B)$, onde $P(B) > 0$.
11. Encontre $P(A|B)$ se (i) $A \cap B = \emptyset$, (ii) $A \subset B$ e (iii) $A \supset B$.
12. Mostre que se $P(A|B) > P(A)$ então $P(B|A) > P(B)$.
13. Prove que

$$P(A \cap B \cap C) = P(A|B \cap C)P(B|A)P(C).$$

14. Prove que os eventos A e B são independentes se

$$P(A|B) = P(A|B^c).$$

15. Sejam A e B dois eventos tais que $P(B) > 0$. Mostre que

$$P(A^c|B) = 1 - P(A|B).$$

16. Uma urna contém três moedas com cara de cada lado, quatro moedas com uma coroa de cada lado e duas moedas honestas. Se uma dessas nove moedas é selecionada aleatoriamente e joga-se uma vez, qual é a probabilidade de que a cara seja a figura obtida?
17. Seja P uma medida de probabilidade e sejam $P_1(\cdot) = P(\cdot|B)$ e $P_2(\cdot) = P(\cdot|C)$, onde $P(B) > 0$ e $P(C) > 0$. Prove que, qualquer seja o evento A

$$P_2(A) = P(A|B \cap C).$$

18. Mostre que, se $P(A|B) = 1$, então $P(B^c|A^c) = 1$.
19. Provar o Teorema 2.32.
20. Sejam A_1, A_2, \dots, A_n eventos independentes. Prove que

$$P\left(\bigcup_{i=1}^n A_i\right) = 1 - \prod_{i=1}^n [1 - P(A_i)].$$

21. Suponhamos que os eventos A e B sejam independentes. Demonstre que $\sigma(A)$ e $\sigma(B)$ também são eventos independentes.
22. Seja A um evento com probabilidade estritamente positiva. Demonstre que se efetuam-se infinitos ensaios independentes do experimento, a probabilidade de que nunca ocorra o evento A é zero.
23. Sejam A_1, A_2, \dots, A_n eventos independentes com probabilidade de ocorrência

$$p_k = P(A_k), \quad k = 1, 2, \dots, n.$$

Encontre a probabilidade da ocorrência dos seguintes eventos, em termos das probabilidades p_k :

- (a) Que nenhum dos A_k ocorram.
 - (b) Que ocorra pelo menos um dos A_k .
 - (c) Que ocorra exatamente um dos A_k .
 - (d) Que ocorram exatamente dois dos A_k .
 - (e) Que ocorram todos os A_k .
 - (f) Que ocorram, no máximo, $n - 1$ dos eventos A_k .
24. Sejam A, B e C eventos. Dizemos que A e B são condicionalmente independentes, dado C se

$$P(A \cap B|C) = P(A|C)P(B|C).$$

Suponha que:

- (a) A e B são condicionalmente independentes dado C e
- (b) A e B são condicionalmente independentes dado C^c .

Mostre que A e B não são necessariamente independentes, mas que A e B são independentes se C for independente de A ou B .

25. Suponha que A e B sejam eventos independentes. Determinar quais dos seguintes pares de eventos são independentes e quais disjuntos:

- (a) A e B^c .
- (b) $A \cap B^c$ e B .
- (c) A^c e B^c .
- (d) A^c e B .
- (e) $A^c \cap B^c$ e $A \cup B$.

26. Sejam A_1, A_2, \dots eventos independentes com $P(A_k) = p_k$. Defina

$$B = \bigcup_{k=1}^{\infty} A_k.$$

Mostre que $P(B) = 1$ se, e somente se,

$$\sum_{k=1}^{\infty} \ln(1 - p_k) = -\infty.$$

27. Seja X uma variável aleatória se supnhamos que $\{x_n\}$ seja uma sequência de números estritamente crescente, ou seja, $x_n > x_{n+1}$ para todo n . Suponhamos que $\lim_{n \rightarrow \infty} x_n = x_0$. Definamos $A_n = [X \leq x_n]$.

- (a) Mostre que

$$\bigcap_{n=1}^{\infty} A_n = [X \leq x_0].$$

- (b) Prove que $\lim_{n \rightarrow \infty} P(X \leq x_n) = P(X \leq x_0)$.

- (c) Seja agora $\{x_n\}$ uma sequência de números estritamente crescente, ou seja, $x_n < x_{n+1}$ cujo limite é x_0 . Novamente, definindo $A_n = [X \leq x_0]$, mostre que

$$\bigcup_{n=1}^{\infty} A_n = [X < x_0].$$

Prove também que $\lim_{n \rightarrow \infty} P(X \leq x_n) = P(X < x_0)$.

28. Suponha que $\Omega = \{1, 2, \dots, p\}$, onde p é um número primo. Seja $\mathcal{F} = P(\Omega)$ e, para $A \in \mathcal{F}$, defina $P(A) = |A|/p$. Mostre que se A e B são independentes, então ao menos um dos dois eventos é \emptyset ou Ω .
29. Sejam A, B, A_1, A_2, \dots eventos em um espaço de probabilidade. Suponha que $A_n \subseteq A_{n+1}$, $\bigcup_n A_n = A$ e que B é independente de A_n , para todo $n \geq 1$. Prove que A e B são independentes.

Capítulo 3

Variáveis aleatórias

Estudaremos aqui os conceitos de variável aleatória e de funções relacionadas, tais como a função de distribuição, a função de densidade e a função de probabilidade. Acontece que em muitos experimentos é mais fácil trabalhar com uma variável resumo dos resultados de um experimento aleatório do que com a estrutura de probabilidade original e é esse o objetivo do conceito variável aleatória.

Por exemplo, em uma pesquisa de opinião, decidimos perguntar a 20 pessoas se elas concordam ou discordam com um determinado assunto. Quando a pessoa concorda registramos o valor 1 e quando discorda 0; o espaço amostral desse experimento contém $2^{20} = 1048576$ elementos e esse é o número de todos os resultados possíveis que podem ocorrer quando 20 pessoas são assim consultadas. Seria interessante reduzir esse espaço para um de tamanho razoável.

De que forma? Contando nessas 2^{20} sequências de tamanho 20 cada simplesmente o número de pessoas que concordam com o assunto em questão. Dessa forma o espaço amostral seria reduzido à $\{0, 1, 2, \dots, 20\}$, o qual possui 21 elementos e é muito mais fácil de trabalhar. Quando definimos uma tal quantidade X , estamos definindo uma função do espaço amostral original nos números reais.

3.1 Variáveis aleatórias

Uma variável aleatória é uma função do espaço amostral no conjunto dos números reais satisfazendo certas condições. Representa uma tradução de cada um dos resultados do espaço amostral em números reais. Utilizando variáveis aleatórias pode-se considerar que o possível experimento aleatório não produz como resultados elementos de Ω e sim números reais.

Exemplo 3.1

Ao lançar uma moeda n vezes podemos observar como resultado a sequência de caras e coroas obtidas. Os resultados possíveis aqui são sequências de n caras e coroas, assim definirmos

$$\Omega = \{(w_1, \dots, w_n) : w_i = \text{cara ou coroa}, i = 1, \dots, n\}.$$

O número de caras observadas nos n lançamentos é um dos números característicos

da sequência de caras e coroas. Se definimos

$$X = \text{número de caras observadas},$$

o valor de X depende do resultado do experimento e

$$X(w) = \#\{i : w_i = \text{cara}, 1 \leq i \leq n\}.$$

O conceito de variável aleatória é fundamental na teoria das probabilidades e, depois de enunciarmos o conceito, este termo aparecerá com muita frequência.

Definição 3.1 (*Imagem inversa*)

Seja $X : \Omega \rightarrow \Theta$ uma função qualquer. Definimos imagem inversa de X , denotada por X^{-1} , como uma função entre Θ em Ω tal que:

$$X^{-1}(A) = \{\omega \in \Omega : X(\omega) \in A\}, \quad \forall A \subseteq \Theta.$$

Exemplo 3.2 (*Continuação do Exemplo 3.1*)

Encontremos a imagem inversa da função $X : \Omega \rightarrow \{0, 1, 2, \dots, n\}$, onde

$$\Omega = \{(w_1, \dots, w_n) : w_i = \text{cara ou coroa}, i = 1, \dots, n\}$$

e n é o número de vezes que lançamos uma moeda. Seja $A = \{0\} \subset \{0, 1, 2, \dots, n\}$, então

$$X^{-1}(\{0\}) = \{\omega \in \Omega : X(\omega) \in \{0\}\} = \{(\underbrace{\text{coroa}, \dots, \text{coroa}}_{n \text{ vezes}})\} \subset \Omega.$$

Caso $A = \{n\}$, então

$$X^{-1}(\{n\}) = \{(\underbrace{\text{cara}, \dots, \text{cara}}_{n \text{ vezes}})\} \subset \Omega.$$

Mais trabalhoso é encontrar, por exemplo, a imagem inversa de $\{1\}$ já que

$$\begin{aligned} X^{-1}(\{1\}) = & \{(\text{cara}, \underbrace{\text{coroa}, \dots, \text{coroa}}_{n-1 \text{ vezes}}, \text{coroa}), (\text{coroa}, \text{cara}, \underbrace{\text{coroa}, \dots, \text{coroa}}_{n-2 \text{ vezes}}, \text{coroa}), \\ & \dots \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \\ & (\underbrace{\text{coroa}, \dots, \text{coroa}}_{n-2 \text{ vezes}}, \text{cara}, \text{coroa}), (\underbrace{\text{coroa}, \dots, \text{coroa}}_{n-1 \text{ vezes}}, \text{cara})\}. \end{aligned}$$

Logicamente, dado qualquer subconjunto A de Θ , podemos encontrar sua imagem inversa por X .

Teorema 3.1

Para qualquer função $X : \Omega \rightarrow \Theta$, a imagem inversa X^{-1} satisfaz as seguintes propriedades:

- (a) $X^{-1}(A^c) = \left(X^{-1}(A)\right)^c$;
- (b) $X^{-1}\left(\bigcup_n A_n\right) = \bigcup_n X^{-1}(A_n)$;
- (c) $X^{-1}\left(\bigcap_n A_n\right) = \bigcap_n X^{-1}(A_n)$,

onde $A, A_1, A_2, \dots \in \Omega$.

Demonstração : Exercício. ■

Propriedades importantes da imagem inversa foram apresentadas no Teorema 3.1. A utilização do conceito de imagem inversa em probabilidade fica clara na importantíssima definição de variável aleatória, dada a seguir.

Definição 3.2 (Variável aleatória)

Seja (Ω, \mathcal{F}, P) um espaço de probabilidade. Uma variável aleatória real é uma função $X : \Omega \rightarrow \mathbb{R}$ de forma que para qualquer conjunto Boreliano B ($B \in \mathcal{B}(\mathbb{R})$), se satisfaz que o conjunto $X^{-1}(B)$, chamado de imagem inversa, é um elemento de \mathcal{F} , isto é,

$$X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F}, \text{ para todo } B \in \mathcal{B}(\mathbb{R}). \quad (3.1)$$

Graficamente uma variável aleatória pode representar-se como mostrado na Figura 3.1. Significa que uma variável aleatória é uma função de Ω nos reais de maneira que a imagem inversa de qualquer Boreliano é um elemento da σ -álgebra do espaço de probabilidades.

Observemos que, segundo a definição de variável aleatória, se B é um Boreliano, então

$$X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\}.$$

Por exemplo, o conjunto $X^{-1}([0, +\infty)) = \{\omega \in \Omega : X(\omega) \in [0, +\infty)\}$. Note-se que a noção de probabilidade não faz parte da definição de variável aleatória.

Segundo a definição de variável aleatória, para provar que uma função de conjuntos é variável aleatória devemos verificar a condição (3.1) para todo Boreliano e isso é muito trabalhoso. O resultado a seguir simplifica a prova de que uma função seja variável aleatória a somente os intervalos da forma $(-\infty, x]$, $\forall x \in \mathbb{R}$.

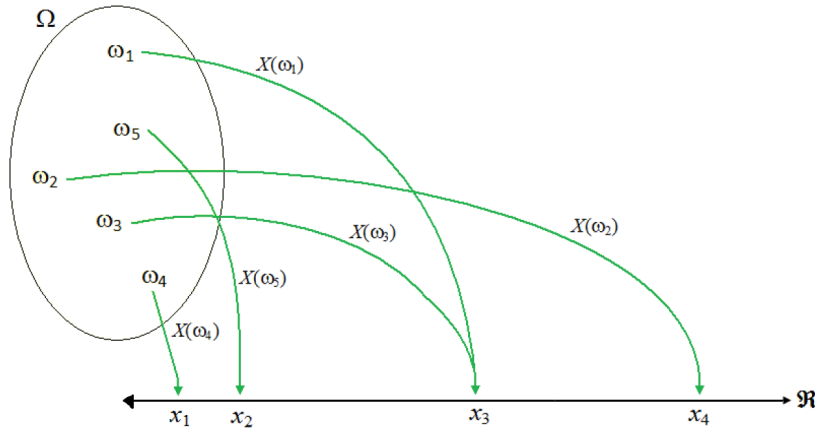


Figura 3.1: Variável aleatória. Imagem inversa de um Boreliano

Teorema 3.2 (*Caracterização de variáveis aleatórias*)

Seja (Ω, \mathcal{F}, P) um espaço de probabilidade e X uma função $X : \Omega \rightarrow \mathbb{R}$. X é uma variável aleatória se, e somente se, para cada $x \in \mathbb{R}$

$$\{w : X(w) \leq x\} = \{X \leq x\} \in \mathcal{F}. \quad (3.2)$$

Demonstração: Se X for variável aleatória a imagem inversa de todo Boreliano pertence a \mathcal{F} .

Em particular, se $B = \{w : -\infty < w \leq x\}$, $x \in \mathbb{R}$, então $X^{-1}(B) = \{w : X(w) \leq x\} \in \mathcal{F}$.

Suponhamos agora que $\{w : X(w) \leq x\} \in \mathcal{F}$, $\forall x \in \mathbb{R}$. Devido a que os intervalos da forma $(-\infty, x]$ geram todos os Borelianos da reta (Teorema 2.9), concluímos que X é variável aleatória. ■

Exemplo 3.3

Seja $\Omega = \{HH, TT, HT, TH\}$ o espaço amostral e $\mathcal{F} = \mathcal{P}(\Omega)$, o conjunto das partes de Ω ou conjunto potência de Ω que é uma σ -álgebra. Definamos X como

$$X(w) = \text{número de H's em } w.$$

Então $X(\{HH\}) = 2$, $X(\{HT\}) = X(\{TH\}) = 1$ e $X(\{TT\}) = 0$. Encontramos que a imagem inversa é

$$X^{-1}(-\infty, x] = \begin{cases} \emptyset & \text{se } x < 0 \\ \{TT\} & \text{se } 0 \leq x < 1 \\ \{HT\}, \{TH\}, \{TT\} & \text{se } 1 \leq x < 2 \\ \Omega & \text{se } x \geq 2 \end{cases}.$$

Por conseguinte, X é uma variável aleatória.

Observemos que se (Ω, \mathcal{F}, P) é um espaço de probabilidade discreto, ou seja, se Ω é um conjunto com uma quantidade enumerável de elementos e se \mathcal{F} é a classe de todos os subconjuntos de Ω , toda função numérica definida em Ω é uma variável aleatória.

Exemplo 3.4

Seja $\Omega = [0, 1]$ e $\mathcal{F} = \mathcal{B}(\mathbb{R}) \cap [0, 1]$, isto é, \mathcal{F} é a σ -álgebra de Borel restrita ao intervalo $[0, 1]$. Definamos X como a função identidade em Ω . Então

$$X(\{w\}) = w, \quad w \in [0, 1].$$

Percebemos que X é variável aleatória. Devido a que todo subconjunto de Borel de Ω é um evento.

Figura 3.2: Variável aleatória. Imagem inversa de um Boreliano.

Explicamos a continuação o motivo técnico pelo qual se definem assim variáveis aleatórias. Lembremos que se (Ω, \mathcal{F}, P) é um espaço de probabilidade e se X é uma variável aleatória, ver Figura 3.2, podemos transladar a função de probabilidade P ao espaço amostral \mathbb{R} com σ -álgebra os Borelianos da seguinte forma: se B é um Boreliano, definimos $P_X(B) = P(X^{-1}(B))$, o que é possível fazer, devido a que $X^{-1}(B)$ é um elemento de \mathcal{F} , domínio de definição de P .

A função $P_X : \mathcal{B}(\mathbb{R}) \rightarrow [0, 1]$ é uma função de probabilidade, o qual será demonstrado no Teorema 3.3, e é conhecida como função de probabilidade induzida pela variável aleatória X .

Com isto queremos dizer que uma vez definida a variável aleatória, esta sempre induz o espaço de probabilidade nos reais $(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_X)$ e, desta forma, não interessa mais o espaço de probabilidades original. Por este motivo, sempre que trabalhemos com variáveis aleatórias não será mais necessário especificar o espaço de probabilidades.

Teorema 3.3 (*Espaço de probabilidade induzido*)

A variável aleatória X , definida no espaço de probabilidade (Ω, \mathcal{F}, P) , induz o espaço de probabilidade $(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_X)$ por meio da relação

$$P_X(B) = P(X^{-1}(B)) = P(\{w \in \Omega : X(w) \in B\}), \quad \forall B \in \mathcal{B}(\mathbb{R}). \quad (3.3)$$

Demonstração : Primeiro observamos que

$$P_X(B) \geq 0, \quad \forall B \in \mathcal{B}(\mathbb{R})$$

e também que

$$P_X(\mathbb{R}) = P(X \in \mathbb{R}) = P(\Omega) = 1.$$

Sejam $\{B_n\}_{n \geq 1}$ eventos disjuntos em Ω . Então

$$\begin{aligned} P_X\left(\bigcup_{n \geq 1} B_n\right) &= P\left(X^{-1}\left(\bigcup_{n \geq 1} B_n\right)\right) = P\left(\left\{w \in \Omega : X(w) \in \bigcup_{n \geq 1} B_n\right\}\right) \\ &= P\left(\bigcup_{n \geq 1} \{w \in \Omega : X(w) \in B_n\}\right) \\ &= \sum_{n \geq 1} P(\{w \in \Omega : X(w) \in B_n\}) = \sum_{n \geq 1} P_X(B_n). \end{aligned}$$

■

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

Tabela 3.1: Diferentes possíveis resultados ao lançar uma moeda três vezes e valores da variável aleatória: número de caras obtidas.

Exemplo 3.5

Consideremos a experiência de jogar uma moeda três vezes de forma independente. A variável aleatória X conta o número de caras obtidos nos três lançamentos. Uma enumeração completa dos valores de X para cada ponto no espaço amostral é dada na Tabela 3.1, nela identificamos por C quando o resultado do lançamento seja cara e por K caso seja coroa.

O espaço amostral de X é $\Omega = \{0, 1, 2, 3\}$ e, partindo do princípio de que todos os oito pontos não elementares em \mathcal{F} sejam equiprováveis, simplesmente contando na Tabela 3.1 vemos que a função de probabilidade induzida em Ω é dada na Tabela 3.2. Resumidamente escrevemos

$$P_X(X = x) = P_X(\{\omega \in \Omega : X(\omega) = x\})$$

ou, de maneira mais específica, escrevemos que

$$P_X(X = 0) = \frac{1}{8}, \quad P_X(X = 1) = \frac{3}{8}, \quad P_X(X = 2) = \frac{3}{8} \quad e \quad P_X(X = 3) = \frac{1}{8}.$$

Vejam agora duas definições auxiliares ambas relacionadas e dedicadas à variáveis aleatórias diretamente, isto é, nestas duas situações definimos diretamente variáveis aleatórias e não mais às consideramos como transformações do espaço amostral original. Esta será a abordagem enfrente, quer dizer, uma vez conhecido o conceito de variável aleatória o modelo

x	0	1	2	3
$P_X(\{\omega \in \Omega : X(\omega) = x\})$	1/8	3/8	3/8	1/8

Tabela 3.2: Função de probabilidade dos diferentes possíveis resultados ao contarmos o número de caras obtidas em três lançamentos.

probabilístico será baseado sempre na suposição da existência de uma determinada variável aleatória.

Definição 3.3 (*Função indicadora*)

Seja (Ω, \mathcal{F}, P) um espaço de probabilidade. Para qualquer conjunto $A \subseteq \Omega$, definamos

$$\mathbf{1}_A(\omega) = \begin{cases} 0 & \text{se } \omega \notin A \\ 1 & \text{se } \omega \in A, \end{cases}.$$

A função $\mathbf{1}_A(\omega)$ é chamada de função indicadora do conjunto A .

Em outros contextos, como na ciência da computação, a função indicado ou função característica é conhecida como função booleana. A função Dirichlet é um exemplo de função indicadora e é o indicador dos números racionais.

Observemos que $\mathbf{1}_A(\omega) = 1$ caso $A = \Omega$ e $\mathbf{1}_A(\omega) = 0$, caso $A = \emptyset$.

Teorema 3.4

Seja (Ω, \mathcal{F}, P) um espaço de probabilidade. A função indicadora $\mathbf{1}_A$ é variável aleatória se, e somente se, $A \in \mathcal{F}$.

Demonstração: Se $\mathbf{1}_A$ é variável aleatória então, para todo $B \in \mathcal{B}(\mathbb{R})$, temos

$$\mathbf{1}_A^{-1}(B) = \{\omega \in \Omega : \mathbf{1}_A(\omega) \in B\} \in \mathcal{F}.$$

Bastaria escolher $B = (\infty, x]$, $x \in \mathbb{R}$. Observe que se $0 < x < 1$ então

$$\mathbf{1}_A^{-1}((-\infty, x]) = \{\omega \in \Omega : \mathbf{1}_A(\omega) \in (-\infty, x]\}$$

e obtemos que $\{\omega \in \Omega : \mathbf{1}_A(\omega) = 0\} = A^c \in \mathcal{F}$, logo $A \in \mathcal{F}$. Caso $x \geq 1$ temos

$$\mathbf{1}_A^{-1}((-\infty, x]) = \{\omega \in \Omega : \mathbf{1}_A(\omega) = 1\} = A \in \mathcal{F}.$$

Vamos supor agora que $A \in \mathcal{F}$ então, pelo Teorema 3.2

$$\mathbf{1}_A((-\infty, x]) = \begin{cases} \emptyset \in \mathcal{F}, & \text{caso } x < 0, \\ \{\omega \in \Omega : \mathbf{1}_A(\omega) = 0\} = A^c \in \mathcal{F} & \text{caso } 0 < x < 1 \\ \{\omega \in \Omega : \mathbf{1}_A(\omega) = 1\} = A \in \mathcal{F} & \text{caso } x \geq 1 \end{cases}.$$

Logo $\mathbf{1}_A$ é variável aleatória. ■

Definição 3.4 (Variável simples)

Uma variável aleatória X em (Ω, \mathcal{F}, P) é chamada *simples* se para qualquer $w \in \Omega$

$$X(w) = \sum_{k=1}^n x_k \mathbf{1}_{A_k}(w),$$

onde $A_1, \dots, A_n \in \mathcal{F}$, x_1, \dots, x_n são números reais e $\mathbf{1}_A$ a função indicadora do conjunto A .

Exemplo 3.6

Todos os exemplos considerados nesta seção são casos particulares de variáveis aleatórias simples. No Exemplo 3.3 a variável aleatória X também é simples, sendo que $x_1 = 0$, $A_1 = \{TT\}$, $x_2 = 1$, $A_2 = \{TH\}$, $x_3 = 1$, $A_3 = \{HT\}$ e $x_4 = 2$, $A_4 = \{HH\}$. Assim

$$X(w) = \sum_{k=1}^4 x_k \mathbf{1}_{A_k}(w), \quad \forall w \in \Omega.$$

Observemos também que, se a classe de eventos \mathcal{F} consistir em todos os subconjuntos de Ω , então o conjunto qualquer A estará sempre em \mathcal{F} e qualquer função de Ω a \mathbb{R} será uma variável aleatória. No entanto, se a classe de eventos \mathcal{F} não consistir em todos os subconjuntos de Ω , algumas funções de Ω a \mathbb{R} podem não ser variáveis aleatórias, conforme ilustrado no exemplo a seguir.

Exemplo 3.7

Este exemplo mostra porque a definição de uma variável aleatória requer que verifiquemos se o conjunto A está em \mathcal{F} . Uma urna contém três bolas. Uma bola é codificada eletronicamente com um rótulo 00. Outra bola é codificada com 01 e a terceira bola tem um rótulo de 10. O espaço amostral para este experimento é $\Omega = \{00, 01, 10\}$. Consideremos a classe de eventos \mathcal{F} consistindo em todas as uniões, interseções e complementos dos eventos $A_1 = \{00, 10\}$ e $A_2 = \{01\}$. Nesta classe de eventos, o resultado 00 não pode ser distinguido a partir do resultado 10. Por exemplo, isso pode resultar de um leitor de etiqueta defeituoso que não consegue distinguir entre 00 e 10. A classe de eventos tem quatro eventos $\mathcal{F} = \{\emptyset, \{00, 10\}, \{01\}, \Omega\}$. A função probabilidade atribuí para os eventos em \mathcal{F} os seguintes valores: $P(A_1) = 2/3$ e $P(A_2) = 1/3$.

Considere a seguinte função X de Ω em \mathbb{R} : $X(00) = 0$, $X(01) = 1$ e $X(10) = 2$. Para encontrar a probabilidade de $\{X = 0\}$ precisarmos da probabilidade de $\{w \in \Omega : X(w) = 0\}$. No entanto, $\{00\}$ não está na classe \mathcal{F} e então X não é uma variável aleatória porque não podemos determinar a probabilidade de $X = 0$.

Como muitos conceitos matemáticos, demonstrar que uma função de conjuntos é variável aleatória não é uma tarefa simples. Por isso destacamos que, uma vez entendido o conceito de variável aleatória, vamos supor sempre sua existência. Agora dedicamos-nos às suas propriedades.

3.1.1 Propriedades das variáveis aleatórias

A continuação demonstramos que algumas operações básicas entre variáveis aleatórias produzem novas variáveis aleatórias. Suponha então que (Ω, \mathcal{F}, P) é um espaço de probabilidade dado. Todas as variáveis aleatórias que consideramos a seguir estão definidas sob o mesmo espaço de probabilidade.

Teorema 3.5

A função $X = c$, onde c é uma constante, é uma variável aleatória em Ω .

Demonstração: Seja $B \in \mathcal{B}(\mathbb{R})$. Para a função $X = c$ temos que,

$$X^{-1}(B) = \begin{cases} \Omega & \text{se } c \in B \\ \emptyset & \text{se } c \notin B \end{cases}.$$

Em ambos os casos $X^{-1}(B) \in \mathcal{F}$ e, portanto, X é variável aleatória. ■

Este teorema nos disse que qualquer número real c é uma variável aleatória a qual chamaremos de determinística.

Teorema 3.6

Seja X uma variável aleatória e c uma constante. Então a função cX é uma variável aleatória.

Demonstração: Provaremos que, para cada número real x , a imagem inversa do conjunto $(-\infty, x]$, obtida utilizando a função cX , é um elemento de \mathcal{F} . Temos três situações:

- Se $c > 0$, então o conjunto $\{cX \leq x\} = \{X \leq x/c\} \in \mathcal{F}$, já X é variável aleatória.
- Se $c < 0$, então novamente o conjunto $\{cX \leq x\} = \{X \geq x/c\} \in \mathcal{F}$, pelo mesmo motivo.
- Se $c = 0$, então $cX = 0$, função constante zero é variável aleatória pelo Teorema 3.5. ■

Observamos existem duas conclusões evidentes. A primeira acontece se escolhemos $c = 0$, resultando sempre em $X = 0$ que, pelo Teorema 3.5 sabemos é variável aleatória. A segunda conclusão evidente acontece escolhendo $c = 1$, situação na qual obtemos como resposta a própria X . Vejamos uma situação na qual $c \neq 0$ e $c \neq 1$.

Exemplo 3.8 (Continuação do Exemplo 3.3)

Consideremos no Exemplo 3.3 o caso $c = 2$, ou seja, pelo Teorema 3.6 sabemos que $2X$ é uma nova variável aleatória de valores

$$2X(\{HH\}) = 4, \quad 2X(\{HT\}) = 2X(\{TH\}) = 2 \quad \text{e} \quad 2X(\{TT\}) = 0.$$

Encontramos que a imagem inversa é

$$2X^{-1}(-\infty, x] = \begin{cases} \emptyset & \text{se } x < 0 \\ \{TT\} & \text{se } 0 \leq x < 2 \\ \{HT\}, \{TH\}, \{TT\} & \text{se } 2 \leq x < 4 \\ \Omega & \text{se } x \geq 4 \end{cases}.$$

Teorema 3.7

Sejam X e Y duas variáveis aleatórias em Ω . Então $X + Y$ é também uma variável aleatória em Ω .

Demonstração: Provaremos que $\forall x \in \mathbb{R}$ o conjunto $\{X + Y > x\} \in \mathcal{F}$. Observemos que para todo número racional podemos escrever

$$\{X + Y > x\} = \bigcup_{r \in \mathbb{Q}} \{X > x\} \cap \{Y > x - r\},$$

isto devido à densidade do conjunto dos racionais. Se esta relação for verdadeira, então $X + Y$ é uma variável aleatória.

\subseteq Seja $\omega \in \Omega$ de maneira que $X(\omega) + Y(\omega) > x$. Então $X(\omega) > x - Y(\omega)$. Como o conjunto dos números racionais são densos então existe um número racional r tal que $X(\omega) > r > x - Y(\omega)$, logo $X(\omega) > r$ e $Y(\omega) > x - r$. Então

$$\omega \in \bigcup_{r \in \mathbb{Q}} \{X > x\} \cap \{Y > x - r\}.$$

\supseteq Seja agora

$$\omega \in \bigcup_{r \in \mathbb{Q}} \{X > x\} \cap \{Y > x - r\}.$$

Então existe um número racional r_0 de forma que $X(\omega) > r_0$ e $Y(\omega) > x - r_0$. Somando, temos que $X(\omega) + Y(\omega) > x$. Logo $\omega \in \{X + Y > x\}$. ■

Exemplo 3.9

Considere um experimento em que uma moeda é lançada até que a primeira cara apareça. O espaço amostral para esta experiência pode ser representado como

$$\Omega = \{C, KC, KKC, KKKC, \dots\},$$

onde C e K representam os eventos de que o resultado do lançamento da moeda seja Cara ou Coroa, respectivamente. Podemos definir uma variável aleatória X que conta o

número de coroas até a primeira cara:

$$X(\{C\}) = 0, \quad X(\{KC\}) = 1, \quad X(\{KKC\}) = 2, \quad \dots$$

e assim por diante. Uma outra variável aleatória que podemos definir seria Y que conta o número de lançamentos da moeda em cada evento, assim

$$Y(\{C\}) = 1, \quad Y(\{KC\}) = 2, \quad Y(\{KKC\}) = 3, \quad \dots$$

de forma que os valores da nova variável aleatória $X+Y$ são os números naturais ímpares.

No exemplo anterior observamos que cada valor da variável aleatória X , assim como da variável Y , correspondem a um único resultado em Ω . Podemos perceber que as variáveis aleatórias podem assumir finitos ou infinitos valores discretos assim como também infinitos valores contínuos.

Teorema 3.8

Seja X uma variável aleatória em Ω e a, b constantes. Então $aX + b$ é também uma variável aleatória.

Demonstração: Exercício. Combinação dos três teoremas anteriores. ■

Uma variável aleatória é uma função, logo dizer que duas variáveis aleatórias X e Y são iguais é afirmar que para todo $\omega \in \Omega$ as imagens coincidem, isto é, $X(\omega) = Y(\omega)$.

Teorema 3.9

Sejam X e Y variáveis aleatórias em Ω . Então $X \times Y$ é também uma variável aleatória.

Demonstração: Suponhamos primeiro que $X = Y$. Então, necessitamos provar que $\{X^2 \leq x\} \in \mathcal{F}$, $\forall x \in \mathbb{R}$. No caso $\{X^2 \leq x\} = \emptyset$ se $x < 0$ e $\{X^2 \leq x\} = \{-\sqrt{x} \leq X \leq \sqrt{x}\}$ se $x > 0$. Em ambos os casos, o conjunto $\{X^2 \leq x\}$ é um elemento de \mathcal{F} . No caso geral, $X \neq Y$, utilizamos a expressão $X \times Y = ((X + Y)^2 - (X - Y)^2)/4$. Portanto, $X \times Y$ é variável aleatória. ■

Exemplo 3.10 (Continuação do 3.3)

Utilizando como base o Exemplo 3.3, definamos a função real Y como

$$Y(w) = \text{número de } T\text{'s em } w.$$

Então $Y(\{HH\}) = 0$, $Y(\{HT\}) = Y(\{TH\}) = 1$ e $Y(\{TT\}) = 2$. Observemos que $Y = 2 - X$, de maneira que Y é variável aleatória pelo Teorema 3.8, nesta situação $a = -1$ e $b = 2$. Encontremos as novas variáveis aleatórias $X + Y$ e $X \times Y$.

Observamos que $X + Y = X + (2 - X) = 2$, ou seja, a soma destas variáveis aleatórias é a variável aleatória constante ou determinística igual a 2. Podemos perceber isto também porque $(X + Y)(\{HH\}) = X(\{HH\}) + Y(\{HH\}) = 2 + 0 = 2$, $(X + Y)(\{HT\}) = X(\{HT\}) + Y(\{HT\}) = 1 + 1 = 2$, $(X + Y)(\{TH\}) = X(\{TH\}) + Y(\{TH\}) = 1 + 1 = 2$ e $(X + Y)(\{TT\}) = X(\{TT\}) + Y(\{TT\}) = 0 + 2 = 2$. No caso do produto $X \times Y$ temos que

$$(X \times Y) = \frac{1}{4}((X + Y)^2 - (X - Y)^2) = 1 - (X - 1)^2,$$

obtendo-se que

$$(X \times Y)(w) = 2X(w) - X^2(w) = \begin{cases} 4 - 4 = 0, & \text{se } w = \{HH\} \\ 2 - 1 = 1, & \text{se } w = \{HT\} \text{ ou } w = \{TH\} \\ 0 - 0 = 0, & \text{se } w = \{TT\} \end{cases}.$$

Resumindo, a nova variável aleatória $X \times Y$ assume somente valores 0 ou 1.

Como consequência se cumpre que, ao multiplicarmos X por ela mesma n vezes temos que X^n é variável aleatória. Portanto, toda função polinomial de uma variável aleatória é também variável aleatória. No seguinte resultado dizemos que $Y \neq 0$, isso significa que $P(Y = 0) = 0$.

Teorema 3.10

Sejam X e Y variáveis aleatórias e $Y \neq 0$. Então X/Y é também uma variável aleatória.

Demonstração: Exercício. ■

Teorema 3.11

Sejam X e Y variáveis aleatórias em Ω . Então as funções $\max\{X, Y\}$ e $\min\{X, Y\}$ são também variáveis aleatórias em Ω .

Demonstração: Para qualquer número real x ,

$$\{\max\{X, Y\} \leq x\} = \{X \leq x, Y \leq x\} = \{X \leq x\} \cap \{Y \leq x\}.$$

Analogamente, $\{\min\{X, Y\} \geq x\} = \{X \geq x, Y \geq x\} = \{X \geq x\} \cap \{Y \geq x\}$. ■

Uma consequência é que tanto $X^+ = \max\{0, X\}$ quanto $X^- = -\min\{0, X\}$ são variáveis aleatórias.

Teorema 3.12

Se X é uma variável aleatória, então $|X|$ é variável aleatória.

Demonstração: Seja $x \geq 0$, então $\{|X| \leq x\} = \{-x \leq X \leq x\} \in \mathcal{F}$ e se $x < 0$, então $\{|X| \leq x\} = \emptyset \in \mathcal{F}$, assim $|X|$ é variável aleatória. Alternativamente, podemos escrever $|X| = X^+ + X^-$, e pelo visto anteriormente obtemos a mesma conclusão. ■

Mostramos a continuação que, em geral, o recíproco do resultado anterior é falso, isto é, se $X : \Omega \rightarrow \mathbb{R}$ é uma função tal que $|X|$ é variável aleatória não necessariamente X é variável aleatória.

Exemplo 3.11

Considere o espaço amostral $\Omega = \{-1, 0, 1\}$ e a σ -álgebra $\mathcal{F} = \{\emptyset, \{0\}, \{-1, 1\}, \Omega\}$. Seja $X : \Omega \rightarrow \mathbb{R}$ a função identidade $X(w) = w$. Então $|X|$ é variável aleatória, devido a que para qualquer Boreliano B ,

$$|X|^{-1}(B) = \begin{cases} \emptyset & \text{se } 0, 1 \notin B \\ \{0\} & \text{se } 0 \in B \text{ e } 1 \notin B \\ \{-1, 1\} & \text{se } 0 \notin B \text{ e } 1 \in B \\ \Omega & \text{se } 0, 1 \in B \end{cases}$$

Isto significa que $|X|^{-1}(B)$ é um elemento de \mathcal{F} . No entanto, X não é variável aleatória devido a que $X^{-1}((-\infty, -1]) = \{-1\}$, o qual não é um elemento de \mathcal{F} .

Os resultados apresentados até aqui permitem-nos perceber que operações básicas entre variáveis aleatórias, como produto por um escalar, soma de escalar, soma, produto e quociente de variáveis aleatórias continua sendo uma variável aleatória. Ainda podemos generalizar estes resultados para um número finito de variáveis aleatórias.

Passando para uma sequência infinita, vamos demonstrar agora dois teoremas com os quais provaremos que a variável aleatória, sob certas condições, é uma função contínua.

Teorema 3.13

Sejam X_1, X_2, \dots variáveis aleatórias em Ω tais que, para cada $\omega \in \Omega$, os números

$$\sup\{X_1(\omega), X_2(\omega), \dots\} \quad \text{e} \quad \inf\{X_1(\omega), X_2(\omega), \dots\}$$

sejam finitos. Então as funções

$$\sup_{n \geq 1} \{X_n\} \quad \text{e} \quad \inf_{n \geq 1} \{X_n\}$$

são também variáveis aleatórias em Ω .

Demonstração: Sabemos que, para cada $\omega \in \Omega$,

$$\left(\sup_{n \geq 1} \{X_n\} \right)(\omega) = \sup_{n \geq 1} \{X_n(\omega) : n \geq 1\},$$

logo é variável aleatória. Similarmente para o caso do ínfimo. ■

Teorema 3.14

Sejam X_1, X_2, \dots variáveis aleatórias em Ω tais que, para cada $\omega \in \Omega$, o número $\lim_{n \rightarrow \infty} X_n(\omega)$ existe e é finito. Então a função

$$\lim_{n \rightarrow \infty} X_n$$

é uma variável aleatória em Ω .

Demonstração: Dado que o limite de X_n existe, os limites superior e inferior da sequência coincidem. Então, pelo demonstrado antes, o limite de X_n é variável aleatória. ■

Exemplo 3.12

Seja (Ω, \mathcal{F}, P) um espaço de probabilidade e a sequência de variáveis aleatórias $\{X_n\}_{n \geq 1}$ tais que, cada $X_n(\omega) = \mathbf{1}_A(\omega)$. Então

$$\lim_{n \rightarrow \infty} X_n(\omega) = \mathbf{1}_A(\omega).$$

Isto significa que o $\lim_{n \rightarrow \infty} X_n$ é uma variável aleatória também, como afirma o Teorema 3.14.

O próximo resultado, apesar de simples, é importante para desenvolvimentos posteriores.

Teorema 3.15

Toda variável aleatória X é limite pontual de uma sequência de variáveis aleatórias simples $\{X_n\}$, tal que $|X_n| \leq |X|$.

Demonstração: Seja $X \geq 0$. Defina para $n = 1, 2, \dots$

$$X_n(w) = \sum_{k=1}^{n2^n} \frac{k-1}{2^n} \mathbf{1}_{A_{n,k}}(w) + n \mathbf{1}_{A_{n,n2^n+1}}(w),$$

onde

$$A_{n,k} = \left\{ w : \frac{k-1}{2^n} \leq X(w) < \frac{k}{2^n} \right\}$$

e

$$A_{n,n2^n+1} = \{w : X(w) \geq n\}.$$

Não é difícil perceber que $\{X_n\}$ é uma sequência não decrescente e

$$\lim_{n \rightarrow \infty} X_n(w) = X(w).$$

Para X arbitrária segue pois $X = X^+ - X^-$, $|X| = X^+ + X^-$ e X^+ e X^- são não negativas. ■

3.2 Função de distribuição

Na seção anterior introduzimos o conceito de variável aleatória e como pode ser notado a medida de probabilidades no espaço amostral não foi envolvida nessa definição. Consideremos agora o espaço de probabilidade (Ω, \mathcal{F}, P) e X uma variável aleatória definida nele. Como faremos os cálculos das probabilidades envolvendo variáveis aleatórias? para responder isso estudaremos uma função associada a toda variável aleatória, chamada de função de distribuição. Nesta seção definimos esta importante função e demonstramos algumas de suas propriedades.

Definição 3.5 (*Função de distribuição*)

A função de distribuição de uma variável aleatória X é a função $F_X : \mathbb{R} \rightarrow [0, 1]$, definida como

$$F_X(x) = P(X \leq x), \quad (3.4)$$

para todo $x \in \mathbb{R}$.

O argumento da função é a letra minúscula x que pode assumir qualquer valor real. Por razões óbvias a esta função se lhe conhece também com o nome de função de acumulação de probabilidades ou simplesmente como função de probabilidade acumulada. Observe que a função de distribuição de uma variável aleatória está definida sob a totalidade do conjunto dos números reais e, sendo uma probabilidade, assume valores no intervalo $[0, 1]$.

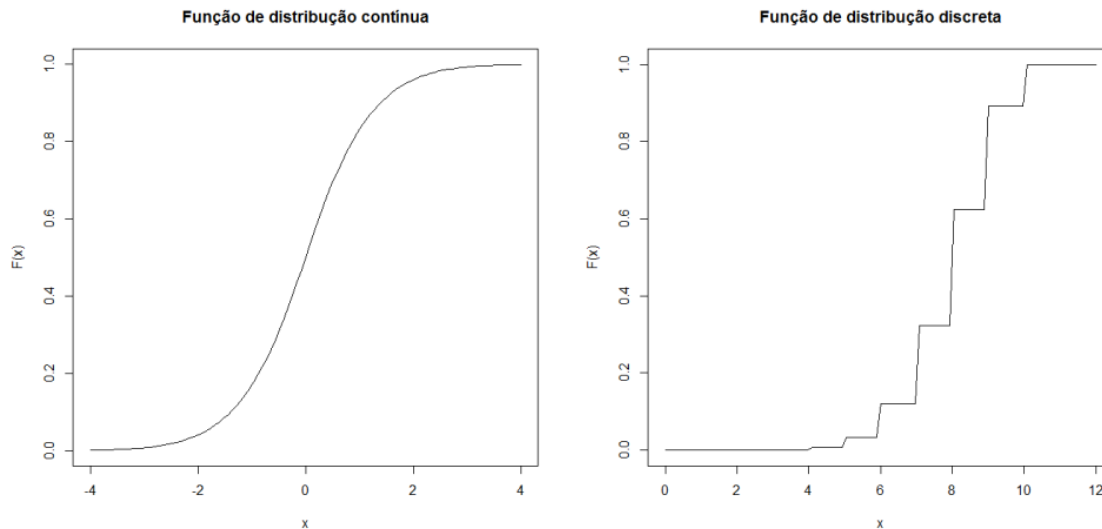


Figura 3.3: Funções de distribuição absolutamente contínua e discreta, respectivamente.

A função de distribuição é importante pois, como será ilustrado abaixo, contém todas as informações da variável aleatória e da medida de probabilidade correspondente. Esta função é apresentada graficamente na Figura 3.3 para duas das diversas possíveis situações: absolutamente contínuo e caso discreto.

Exemplo 3.13

Seja a variável aleatória X definida em (Ω, \mathcal{F}, P) como

$$X(w) = c, \quad \text{para todo } w \in \Omega.$$

$$\text{Então } P(X = c) = 1 \text{ e } F_X(x) = P(X^{-1}(-\infty, x]) = \begin{cases} 0 & \text{se } x < c \\ 1 & \text{se } x \geq c \end{cases}.$$

Veremos a seguir algumas propriedades básicas dessa função e sempre que a expressão $F_X(x^+)$ aparece, significa que tomamos o limite à direita da função F_X no ponto x . Também aparece a expressão de $F_X(x^-)$, o que significa que, de um modo análogo, o limite esquerdo da função F_X no ponto x .

Observemos que da definição de função de distribuição podemos escrever $F_X(x) = P(X \in (-\infty, x])$ ou ainda, se $(\mathbb{R}, \mathcal{B}(\mathbb{R}), P)$ for o espaço de probabilidade original

$$F_X(x) = P(X^{-1}(-\infty, x]),$$

para qualquer número real x .

Como veremos, a importância de F_X é que caracteriza a variável aleatória X , isto é, F_X determina o valor de $P_X(B)$ para todo $B \in \mathcal{B}(\mathbb{R})$.

Teorema 3.16 (Propriedades da função de distribuição)

Seja (Ω, \mathcal{F}, P) um espaço de probabilidade, $X : \Omega \rightarrow \mathbb{R}$ uma variável aleatória e F_X a função de distribuição de X . Então

1-

$$\lim_{x \rightarrow \infty} F_X(x) = 1,$$

2-

$$\lim_{x \rightarrow -\infty} F_X(x) = 0,$$

3- Se $x_1 \leq x_2$, então

$$F_X(x_1) \leq F_X(x_2),$$

4- $F_X(x)$ é contínua pela direita, isto é,

$$F_X(x^+) = F_X(x).$$

Demonstração: Por definição $F_X(x) = P(X \leq x)$.

- 1- Seja x_1, x_2, \dots uma sequência de números reais crescente ao infinito, ou seja, $\lim_{n \rightarrow \infty} x_n = \infty$ e sejam os eventos $A_n = \{w \in \Omega : X(w) \leq x_n\}$. Então $\{A_n\}_{n \in \mathbb{N}}$ é uma sequência de eventos crescente de limite Ω .

Pela continuidade da função de probabilidade

$$\lim_{n \rightarrow \infty} F_X(x_n) = \lim_{x \rightarrow \infty} P(A_n) = P(\Omega) = 1.$$

Isto implica que $F_X(x)$ converge a um quando x tende a infinito.

2- Exercício

3- Sejam $x_1 \leq x_2$, então

$$\begin{aligned} F_X(x_1) &\leq F_X(x_1) + P(x_1 < X \leq x_2) \\ &= P[(X \leq x_1) \cup (x_1 < X \leq x_2)] \\ &= P(X \leq x_2) = F_X(x_2). \end{aligned}$$

4- Seja x_1, x_2, \dots uma sequência de números reais decrescente a x , e sejam os eventos $A_n = \{w \in \Omega : X(w) \leq x_n\}$. Então $\{A_n\}_{n \in \mathbb{N}}$ é uma sequência de eventos decrescente de limite

$$\bigcap_{n \geq 1} \{w \in \Omega : X(w) \leq x_n\} = \{w \in \Omega : X(w) \leq x\}.$$

Pela continuidade da função de probabilidade

$$\lim_{n \rightarrow \infty} F_X(x_n) = \lim_{n \rightarrow \infty} P(X_n \leq x_n) = P(X \leq x) = F_X(x). \quad \blacksquare$$

De maneira geral, a função de distribuição de uma variável aleatória simples, que assumem os valores $x_1 \leq x_2 \leq \dots \leq x_n$ nos conjuntos $A_1, A_2, \dots, A_n \in \mathcal{F}$, é da seguinte forma

$$F_X(x) = \begin{cases} 0 & \text{se } x < x_1 \\ P(A_1) & \text{se } x_1 \leq x < x_2 \\ P(A_1 \cup A_2) & \text{se } x_2 \leq x < x_3 \\ P(A_1 \cup A_2 \cup A_3) & \text{se } x_3 \leq x < x_4 \\ \vdots & \vdots \\ 1 & \text{se } x \geq x_n \end{cases}.$$

Estamos em condições de apresentar uma definição geral de função de distribuição, não fazendo mais referência à variável aleatória nem a espaços de probabilidade particulares.

Definição 3.6

Uma função real $F_X : \mathbb{R} \rightarrow [0, 1]$ é função de distribuição se satisfaz as quatro propriedades no Teorema 3.16.

Um recíproco do último resultado é válido e justifica a importância da função de distribuição. Se enuncia a continuação este interessante resultado do qual omitiremos a demonstração.

Teorema 3.17

Seja $F_X : \mathbb{R} \rightarrow [0, 1]$ uma função de distribuição. Então existe um espaço de probabilidade (Ω, \mathcal{F}, P) e uma variável aleatória X cuja função de distribuição é F_X .

Demonstração: Ver Harris (1966). ■

Desta forma percebemos que é suficiente, em situações práticas, assumir uma função de distribuição específica para saber que existe um determinado espaço de probabilidade e uma variável aleatória definida nele, com função de distribuição a especificada.

Exemplo 3.14

Seja $\Omega = \{H, T\}$ e X uma variável aleatória definida como $X(\{H\}) = 1$ e $X(\{T\}) = 0$. Se P atribui igual probabilidade a cada evento $\{H\}$ e $\{T\}$, temos

$$P(X = 0) = \frac{1}{2} \quad \text{e} \quad P(X = 1) = \frac{1}{2}$$

e a função de distribuição é $F_X(x) = \begin{cases} 0 & \text{se } x < 0 \\ \frac{1}{2} & \text{se } 0 \leq x < 1 \\ 1 & \text{se } x \geq 1 \end{cases}.$

Listamos propriedades que permitem calcular probabilidades utilizando a função de distribuição.

Teorema 3.18

Seja X uma variável aleatória com função de distribuição $F_X(x)$. Para quaisquer números reais $a < b$, se satisfaz que:

- 1- $P(X < a) = F_X(a-),$
- 2- $P(X = a) = F_X(a) - F_X(a-),$
- 3- $P(a < X \leq b) = F_X(b) - F_X(a),$
- 4- $P(a \leq X \leq b) = F_X(b) - F_X(a-),$
- 5- $P(a < X < b) = F_X(b-) - F_X(a),$
- 6- $P(a \leq X < b) = F_X(b-) - F_X(a-).$

Demonstração: Exercício. ■

Observe que como $F_X(x)$ é uma função não decrescente e contínua pela direita, a probabilidade $P(X = x)$ é igual a $F_X(x) - F_X(x-)$, que representa o tamanho do pulo ou descontinuidade da função de distribuição no ponto x . Como consequência, quando $F_X(x)$ é uma função absolutamente contínua e para $a < b$,

$$\begin{aligned} F_X(b) - F_X(a) &= P(a < X \leq b) = P(a \leq X \leq b) \\ &= P(a \leq X < b) = P(a < X < b). \end{aligned}$$

Isto é, quando $F_X(x)$ seja uma função absolutamente contínua, incluir ou excluir os extremos de um intervalo não afeta o valor da probabilidade de dito intervalo. Por isso, para qualquer número real x , o evento $\{X = x\}$ tem probabilidade é zero. Finalizamos esta seção com um resultado interessante de prova surpreendentemente simples.

Teorema 3.19

Toda função de distribuição possui, no máximo, um número enumerável de descontinuidades.

Demonstração: Seja D o conjunto dos pontos de descontinuidade da função de distribuição $F_X(x)$. Para cada número natural n definamos os subconjuntos

$$D_n = \left\{ n \in D : \frac{1}{n+1} < F_X(x) - F_X(x-) \leq \frac{1}{n} \right\}.$$

Cada conjunto D_n possui, no máximo, n elementos. Dado que

$$D = \bigcup_{n=1}^{\infty} D_n,$$

concluimos que D é enumerável. ■

Definição 3.7

Duas variáveis aleatórias X e Y são consideradas identicamente distribuídas se

$$P(X \leq x) = P(Y \leq x), \quad \forall x \in \mathbb{R}.$$

Nada dizemos acerca do espaço de probabilidade ao qual pertencem as variáveis aleatórias X e Y mas, esta definição é aplicável também à situação na qual os espaços de probabilidade sejam diferentes. Suponhamos que $(\Omega_1, \mathcal{F}_1, P_1)$ e $(\Omega_2, \mathcal{F}_2, P_2)$ sejam espaços de probabilidade diferentes, onde $X : \Omega_1 \rightarrow \mathbb{R}$ e $Y : \Omega_2 \rightarrow \mathbb{R}$ então concluimos, desta definição, que se as variáveis X e Y são identicamente distribuídas então $F_X(x) = F_Y(x)$, $\forall x \in \mathbb{R}$. No geral, quando falamos de duas ou mais variáveis aleatórias, estamos assumindo que todas pertencem ao mesmo espaço de probabilidade, ou seja, $X : \Omega \rightarrow \mathbb{R}$ e $Y : \Omega \rightarrow \mathbb{R}$.

Exemplo 3.15

Para todo subintervalo I de $[0, 1]$, seja $P(I)$ o comprimento do intervalo. Então $([0, 1], \mathcal{F}, P)$ é um espaço de probabilidade onde

$$\mathcal{F} = \mathcal{B}(\mathbb{R}) \cap [0, 1]$$

é a σ -álgebra de Borel restrita ao intervalo $[0, 1]$.

A função de distribuição da variável aleatória $X(w) = w$, $w \in [0, 1]$ é dada por

$$F_X(x) = P(\{w : X(w) \leq x\}) = \begin{cases} 0, & \text{se } x < 0, \\ P([0, x]) = x, & \text{se } x \in [0, 1], \\ 1, & \text{se } x \geq 1 \end{cases}$$

Teorema 3.20

As duas seguintes afirmações são equivalentes:

- (a) As variáveis aleatórias X e Y são identicamente distribuídas.
- (b) $F_X(x) = F_Y(x)$ para todo x .

Demonstração : Para demonstrar a equivalência devemos mostrar que cada afirmação implica a outra. Em primeiro lugar, vamos mostrar que se a afirmação em (a) se cumpre, então temos o afirmado em (b). Devido a X e Y serem igualmente distribuídas, para qualquer conjuntos $A \in \mathcal{B}(\mathbb{R})$,

$$P(X \in A) = P(Y \in A).$$

Em particular, para cada x , o conjunto $(-\infty, x] \in \mathcal{B}(\mathbb{R})$ e

$$F_X(x) = P(X \in (-\infty, x]) = P(Y \in (-\infty, x]) = F_Y(x).$$

O argumento acima mostrou que, se as funções de probabilidade de X e Y concordam em todos os conjuntos, então também coincidem nos intervalos que geram a σ -álgebra de Borel. Para demonstrar que a afirmação em (b) implica (a), devemos provar que se X e Y coincidem em todos os intervalos então coincidem em todos os conjuntos de $\mathcal{B}(\mathbb{R})$. Os detalhes desta parte da demonstração podem ser consultados em (Chung, 2001, Seção 2.2). ■

Vamos agora falar acerca dos diferentes tipos de funções de distribuição que existem. Três tipos de funções serão considerados os mais importantes: discretas, contínuas e mistas, uma mistura das duas anteriores. Seja $\{a_i\}$ um conjunto enumerável de pontos de salto da função de distribuição F e seja b_i do tamanho em salto a_i , então

$$F(a_i) - F(a_i^-) = b_i$$

uma vez que $F(a_i^+) = F(a_i)$. Considere a função

$$F_d(x) = \sum_i b_i \delta(x - a_i), \quad (3.5)$$

onde δ é a chamada função delta, definida como

$$\delta(z) = \begin{cases} 0, & z < 0, \\ 1, & z \geq 0 \end{cases} \quad (3.6)$$

Com esta representação da função F_d , observamos que esta função representa a soma de todos os saltos de F na meia reta $(-\infty, x]$. Esta função é claramente crescente, contínua a direita com

$$F_d(-\infty) = 0 \quad \text{e} \quad F_d(+\infty) = \sum_i b_i = 1.$$

Por isso, F_d é uma função crescente limitada. Deve constituir a “parte de salto” de F e, se for subtraída de F , a função restante deve ser positiva, não conter mais saltos e assim ser contínua. Essas declarações plausíveis serão comprovadas: são fáceis de compreender mas não são realmente triviais.

Teorema 3.21

Seja F uma função de distribuição. Então se F pode ser escrita como

$$F_c(x) = F(x) - F_d(x),$$

sendo F_d definida em (3.5). Temos por consequência que F_c é uma função positiva, crescente e contínua.

Demonstração: Seja $x < x'$, então nós temos

$$\begin{aligned} F_d(x') - F_d(x) &= \sum_{x < a_i \leq b} b_i = \sum_{x < a_i \leq b} [F(a_i) - F(a_i^-)] \\ &\leq F(x') - F(x). \end{aligned}$$

Segue-se que ambos F_d e F_c são crescentes e, se colocarmos $x = -\infty$ acima, vemos que $F_d < F$ e então F_c é realmente positiva. Em seguida, F_d é contínuo à direita desde que cada δ_{a_i} pertence à série que define F_d como uma função que converge uniformemente em x ; o mesmo argumento cede

$$F_d(x) - F_d(x^-) = \begin{cases} b_i, & \text{se } x = a_i \\ 0, & \text{caso contrário} \end{cases}.$$

Agora, esta avaliação também é válida se F_d for substituída por F de acordo com a definição de a_i e b_i , daí obtemos para cada x :

$$F_c(x) - F_c(x^-) = F(x) - F(x^-) - [F_d(x) - F_d(x^-)] = 0.$$

Isso mostra que F_c é uma função contínua; uma vez que também é contínua à direita, sendo a diferença de duas dessas funções, é contínua. ■

Até agora sabemos que qualquer função de distribuição pode ser escrita como uma combinação linear de duas funções, uma discreta e a outra contínua. O seguinte resultado complementa este o estudo.

Teorema 3.22

Seja F uma função de distribuição. Suponha que exista uma função contínua G_c e uma função G_d da forma

$$G_d(x) = \sum_i b_i \delta(x - a_i),$$

onde $\{a_i\}$ é um conjunto enumerável de números reais e $\sum_i b_i = 1$, $b_i \geq 0 \forall i$, de maneira que

$$F = G_c + G_d,$$

então

$$G_c = F_c, \quad G_d = F_d,$$

sendo F_c e F_d definidos como antes (Teorema 3.21).

Demonstração: Se $F_d \neq G_d$ então, os conjuntos $\{a_i\}$ e $\{c_i\}$ não são idênticos ou podemos renomear um c_i para que $c_i = a_i$ para todo i mas $d_i \neq b_i$ para algum i . Em ambos os casos temos, para pelo menos um i , um $\tilde{a} = a_i$ ou c_i de forma que

$$[F_d(\tilde{a}) - F_d(\tilde{a}^-)] - [G_d(\tilde{a}) - G_d(\tilde{a}^-)] \neq 0.$$

Desde que $F_c - G_c = G_d - F_d$, isto implica que

$$F_c(\tilde{a}) - G_c(\tilde{a}) - [F_c(\tilde{a}^-) - G_c(\tilde{a}^-)] \neq 0,$$

contradizendo o fato de que $F_c - G_c$ é uma função contínua. Consequentemente $F_d = G_d$ e, por conseguinte, $F_c = G_c$. ■

Exemplo 3.16

Suponhamos que $X \sim F_X$, sendo que

$$F_X(x) = \begin{cases} 0, & x < 0 \\ \frac{x}{4}, & 0 \leq x < 1 \\ \frac{1}{2} + \frac{x-1}{4}, & 1 \leq x < 2 \\ \frac{11}{12}, & 2 \leq x < 3 \\ 1, & x \geq 3 \end{cases}$$

Queremos a expressão F_X como a soma $G_c + G_d$, identificando-as.

Observemos a forma desta função de distribuição na Figura 3.4, podemos afirmar que a função F_X é de distribuição porque satisfaz as quatro propriedades listadas no

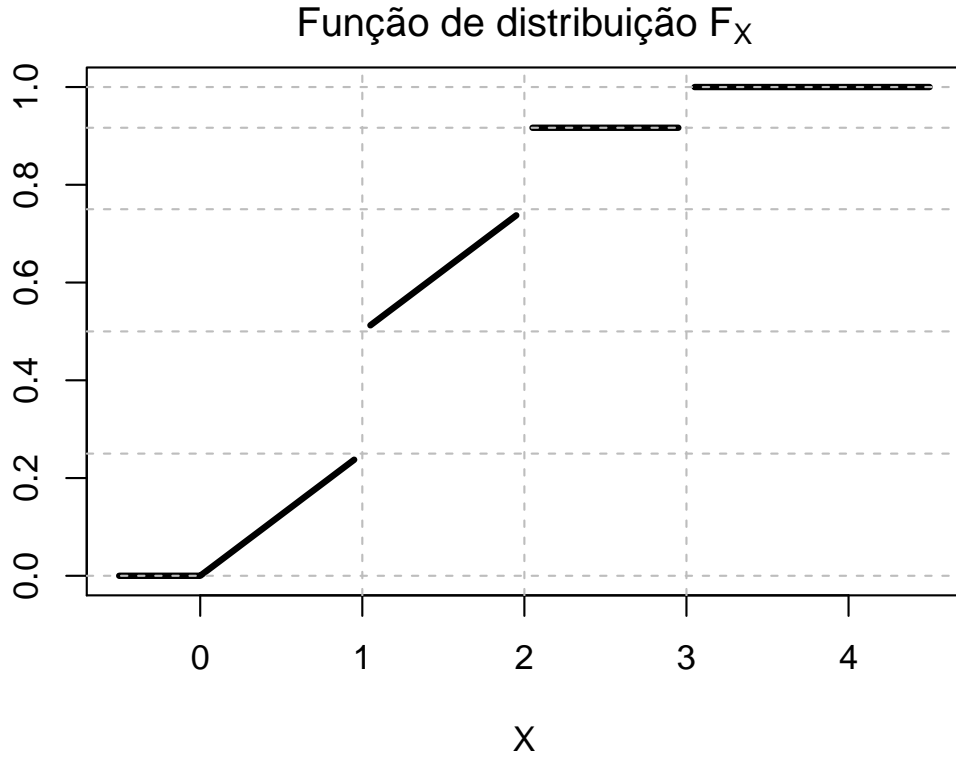


Figura 3.4: Função de distribuição do Exemplo 3.16.

Teorema 3.16. Percebemos também que, nesta situação, temos três pontos claros de descontinuidade: 1, 2 e 3. Por isso, escolhemos

$$F_d(x) = \frac{1}{4}\delta(x-1) + \frac{3}{12}\delta(x-2) + \frac{1}{12}\delta(x-3),$$

como a parte discreta no Teorema 3.21.

Corresponde assim, o restante da função F_X à parte contínua, podendo ser escrita como

$$F_c = \begin{cases} 0, & x < 0 \\ \frac{x}{4}, & 0 \leq x < 1 \\ \frac{1}{4} + \frac{x-1}{4}, & 1 \leq x < 2 \\ \frac{1}{2}, & 2 \leq x < 3 \\ \frac{1}{2}, & x \geq 3 \end{cases}.$$

A decomposição única de F_X em suas partes discreta F_d e contínua F_c , disponíveis neste exemplo, estão apresentadas na Figura 3.5.

Com estes dois teoremas, Teorema 3.21 e Teorema 3.22, estamos concluindo que toda função de distribuição pode ser escrita como uma parte discreta e uma outra contínua. Agora

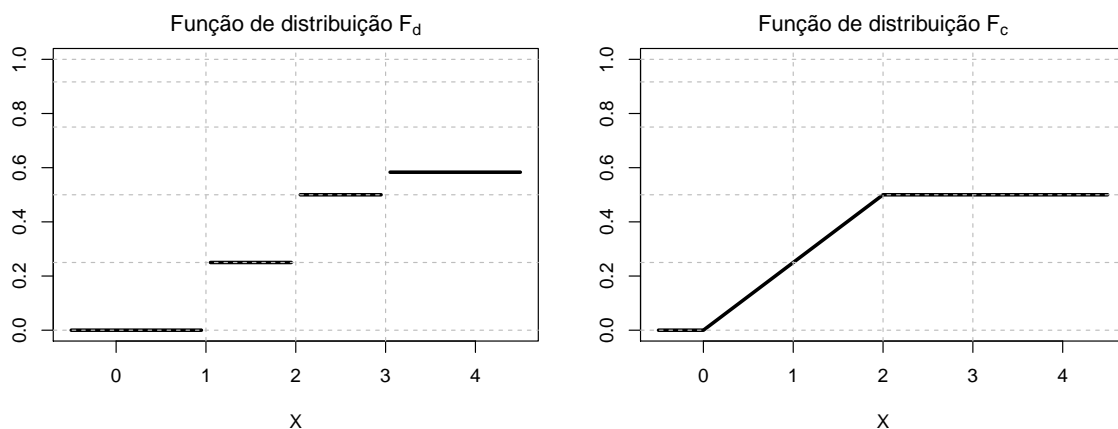


Figura 3.5: Função de distribuição do Exemplo 3.16.

estamos em condições de classificar as variáveis aleatórias dependendo da característica da correspondente função de distribuição. Embora existam dois tipos de funções de distribuição existem, pelo menos, três tipos de variáveis aleatórias: discretas, contínuas e mistas. As variáveis aleatórias discretas estão relacionadas com as funções de distribuição discretas, já as variáveis aleatórias contínuas estão relacionadas às funções de distribuição contínuas e as mistas são discretas em alguns pontos e contínuas em intervalos. Dentre as variáveis aleatórias contínuas dedicaremos nossa atenção a um subgrupo importantíssimo, as variáveis aleatórias absolutamente contínuas. Vejamos a definição de cada uma delas nas seguintes seções.

3.2.1 Variáveis aleatórias discretas

Definição 3.8

Consideremos que a função de distribuição F que pode ser representada na forma

$$F(x) = \sum_i b_i \delta(x - a_i),$$

onde $\{a_i\}$ é um conjunto enumerável de números reais, $b_i > 0$ para cada i , $\sum_i b_i = 1$ e δ a função delta, definida em (3.6). Nestas condições F é chamada de função de distribuição discreta.

Uma definição verbal plausível de uma função de distribuição discreta pode ser dado assim: “É uma função de distribuição que tem saltos e é constante entre os saltos”. Essa função às vezes é chamada de “função por etapas”, embora o significado deste termo não pareça estar bem estabelecido. O que há de errado com isso? Mas suponha que o conjunto de pontos de salto seja “discreto” na topologia euclidiana, então a definição é válida, além da nossa convenção de continuidade à direita.

Definição 3.9 (Variável aleatória discreta)

A variável aleatória X é dita ser discreta se sua correspondente função de distribuição $F_X(x)$ é discreta.

Exemplo 3.17

As variáveis aleatórias simples são discretas. A forma da função de distribuição nesta situação foi apresentada no Exemplo 3.13.

Observemos que nada foi afirmado acerca da quantidade de valores diferentes que uma variável aleatória discreta pode assumir, isto é, uma variável aleatória discreta pode assumir tanto uma quantidade finita de valores como uma quantidade infinita enumerável.

Caso x_1, x_2, \dots sejam os pontos de descontinuidade de F_X , em cada um destes pontos o tamanho da descontinuidade de F_X é

$$P(X = x_i) = F_X(x_i) - F_X(x_i-) > 0.$$

Definição 3.10

Seja X uma variável aleatória discreta e F_X sua função de distribuição. A função $P(X = x)$, que indica o incremento de F_X em cada ponto de descontinuidade, se conhece como função de probabilidade de X e se define como:

X	x_1	x_2	x_3	\dots
$P(X = x)$	$P(X = x_1)$	$P(X = x_2)$	$P(X = x_3)$	\dots

A função de distribuição se reconstrói da seguinte forma $F_X(x) = \sum_{u \leq x} P(X = u)$.

A função de distribuição mais simples é a correspondente à variável aleatória degenerada num ponto. Esta é outra forma de chamarmos às variáveis aleatórias simples.

Exemplo 3.18 (Variável aleatória degenerada)

Uma variável aleatória X , assumindo somente um único valor, é chamada degenerada e sua função de probabilidade é definida por

$$P(X = c) = 1 \quad \text{e} \quad P(X \neq c) = 0. \quad (3.7)$$

Com o auxílio da função delta, podemos escrever a função de distribuição F_X , a função de distribuição da variável degenerada em c , da forma

$$F_X(x) = \delta(x - c),$$

a qual é crescente com um salto quando $x = c$, logicamente então uma função de distribuição discreta.

Observe que, no exemplo acima, temos somente um único ponto de salto ou descontinuidade.

Teorema 3.23 (*Propriedades da função de probabilidade*)

Seja X uma variável aleatória discreta assumindo valores em $\{x_1, x_2, \dots\}$ e P sua função de probabilidade. Então

(a) $P(X = x) \geq 0$, para todo $x \in \{x_1, x_2, \dots\}$,

(b) $\sum_{x \in \{x_1, x_2, \dots\}} P(X = x) = 1$.

Demonstração : Exercício. ■

Vamos agora considerar diferentes novas funções de distribuição discretas mais complexas, estas novas funções definirão modelos probabilísticos discretos. Observe que, até o momento consideremos modelos probabilísticos nos quais a variável aleatória assume somente um ou dois valores. Uma situação mais geral acontece quando considerarmos a possibilidade de que a variável aleatória discreta assuma uma quantidade finita de pontos, todos com a mesma probabilidade de acontecer. Este é o caso da variável aleatória uniforme discreta.

Definição 3.11 (*Distribuição uniforme discreta*)

Seja X uma variável aleatória podendo assumir valores em $\{x_1, x_2, \dots, x_n\}$. Dizemos que X tem distribuição uniforme discreta se

$$P(X = x_i) = \frac{1}{n}, \quad i = 1, 2, \dots, n. \quad (3.8)$$

Fazendo uso da função indicadora definida em (3.3), podemos escrever

$$X = \sum_{i=1}^n x_i \mathbf{1}_{\{X=x_i\}}$$

também, a função de distribuição é da forma

$$F_X(x) = \frac{1}{n} \sum_{i=1}^n \delta(x - x_i).$$

Um pouco mais complexa do que a distribuição degenerada, uma variável aleatória que segue esta distribuição assume somente dois valores.

Definição 3.12 (*Distribuição Bernoulli*)

Dizemos que uma variável aleatória discreta X tem distribuição Bernoulli de parâmetro p , se assumir somente dois valores x_1 e x_2 , com probabilidades

$$P(X = x_1) = p \quad \text{e} \quad P(X = x_2) = 1 - p, \quad 0 < p < 1. \quad (3.9)$$

Uma variável aleatória com distribuição Bernoulli¹ pode ser escrita como

$$X(\omega) = x_1 \mathbf{1}_{\{X=x_1\}}(\omega) + x_2 \mathbf{1}_{\{X=x_2\}}(\omega),$$

onde a função $\mathbf{1}_{\{X=x\}}$ é conhecida como função indicadora da ocorrência do evento $\{X = x\}$, isto é,

$$\mathbf{1}_{\{X=x\}}(\omega) = \begin{cases} 0, & \text{se } X(\omega) \neq x, \\ 1, & \text{se } X(\omega) = x \end{cases} \quad (3.10)$$

A partir desta forma de escrever podemos perceber que a função de distribuição é dada por

$$F_X(x) = p\delta(x - x_1) + (1 - p)\delta(x - x_2). \quad (3.11)$$

Uma outra forma de definir variável aleatória discreta é verificando se existe um conjunto enumerável $E \subseteq \mathbb{R}$, de forma que

$$P(X \in E) = 1. \quad (3.12)$$

Os valores em E são chamados de descontinuidade de F_X , pontos de pulo ou ponto de acréscimo da função de distribuição de X . Observemos que $E \in \mathcal{B}(\mathbb{R})$ uma vez que cada conjunto de um ponto pertence a $\mathcal{B}(\mathbb{R})$. Na verdade, se $x \in \mathbb{R}$, então

$$\{x\} = \bigcap_{n=1}^{\infty} \left\{ \left(x - \frac{1}{n} < x \leq x + \frac{1}{n} \right) \right\}.$$

Então $\{X \in E\}$ é um evento. Seja X uma variável aleatória assumindo valores x_i com probabilidade p_i , $i = 1, 2, \dots$. Temos então que

$$P(\{w : X(w) = x_i\}) = p_i,$$

para $i = 1, 2, \dots$ e $p_i \geq 0$ para todo i . Logo

$$\sum_{i=1}^{\infty} p_i = 1.$$

Observemos que isto significa que toda serie de probabilidades é convergente e, mais ainda, de soma 1.

¹Jacob Bernoulli(1655-1705) foi um matemático suíço e um dos muitos matemáticos proeminentes da família Bernoulli.

Como indicado na Definição 3.9, a função de distribuição da variável aleatória discreta X é

$$F_X(x) = P(X \leq x) = \sum_{x_i \leq x} p_i.$$

Se $\mathbf{1}_A$ denota a função indicadora do conjunto A , podemos escrever qualquer variável aleatória discreta como

$$X(w) = \sum_{i=1}^{\infty} x_i \mathbf{1}_{\{X=x_i\}}(w), \quad \forall w \in \Omega.$$

com função de distribuição corresponde dada por

$$F_X(x) = \sum_{i=1}^{\infty} p_i \delta(x - x_i). \quad (3.13)$$

Uma situação um pouco mais complexa a apresentamos no seguinte exemplo. Lembremos que na definição de variáveis aleatórias discretas dizemos que o conjunto de valores pode ser até infinito enumerável como o caso a seguir.

Exemplo 3.19

Seja X uma variável aleatória com função de probabilidade

$$P(X = x) = \frac{6}{\pi^2} \frac{1}{x^2}, \quad x = 1, 2, \dots$$

Então, a função de distribuição é

$$F_X(x) = \frac{6}{\pi^2} \sum_{k=1}^{\infty} \frac{1}{k^2} \delta(x - k).$$

Em teoria das probabilidades e na estatística a distribuição binomial é a distribuição de probabilidade do número de sucessos em uma sequência de n ensaios independentes, cada um dos quais produz como resposta sucesso com probabilidade p ou fracasso com probabilidade $1 - p$. Um experimento de resultado sucesso ou fracasso também é chamado de experimento Bernoulli; quando $n = 1$, a distribuição binomial é uma distribuição de Bernoulli.

Definição 3.13 (*Distribuição Binomial*)

Dizemos que uma variável aleatória discreta X tem distribuição Binomial de parâmetro p se sua função de probabilidade é da forma

$$P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}, \quad x = 0, 1, 2, \dots, n; \quad 0 < p < 1. \quad (3.14)$$

Escrevemos resumidamente $X \sim \text{Binomial}(n, p)$. Observemos que n é um inteiro positivo conhecido. Esta é a distribuição que atribuímos, por exemplo, ao número de caras obtido em

n lançamentos de uma moeda tendo probabilidade p de obtermos cara em cada lançamento. Mais geral, consideremos um experimento básico no qual estamos interessados na ocorrência ou não de determinado evento, o qual pode acontecer com probabilidade p , como a obtenção do número 6 na face superior no lançamento de um dado equilibrado. Se repetimos o experimento básico n vezes independentes e contamos o número de ocorrências do evento de interesse, esse número terá por distribuição $Binomial(n, p)$.

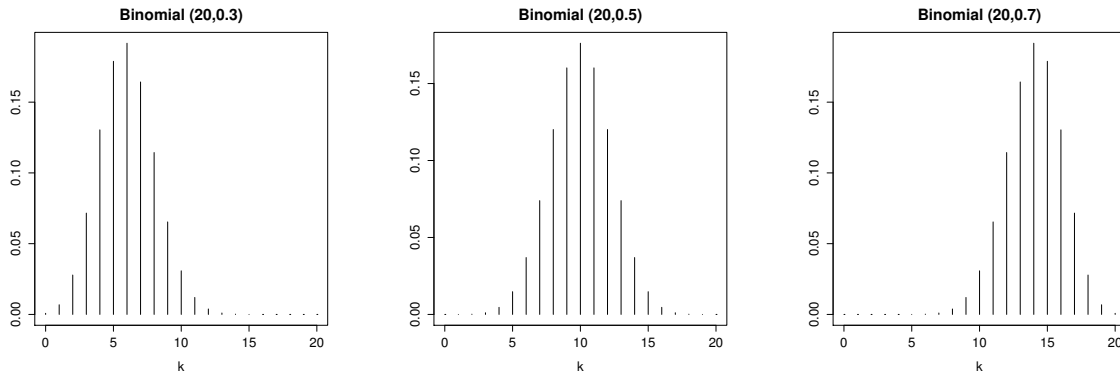


Figura 3.6: Representação da função de probabilidade Binomial para $n = 20$ e três valores do parâmetro $p = 0.3$, $p = 0.5$ e $p = 0.7$.

A Figura 3.6 representa três situações da função de probabilidade Binomial para o caso $n = 20$ e valores da probabilidade de sucesso 0.3; 0.5 e 0.7. Podemos perceber que esta função é simétrica no caso $p = 0.5$, nas outras situações não.

Na Definição 3.13 observamos que

$$\sum_{x=0}^n P(X = x) = [p + (1 - p)]^n = 1,$$

logo a expressão em (3.14) define uma função de probabilidade. Observemos que a função de distribuição pode ser escrita como

$$F_X(x) = \sum_{k=0}^n \binom{n}{k} p^k (1 - p)^{n-k} \delta(x - k),$$

onde $\delta(\cdot)$ é a função delta definida em (3.6).

Exemplo 3.20

Considere a situação na qual r bolas sejam distribuídas em n células, de modo que cada uma das n^r disposições possíveis tenha probabilidade $\frac{1}{n^r}$. Nós estamos interessados na probabilidade p_k que uma célula especificada tenha exatamente k bolas.

Então, a distribuição de cada bola pode ser considerada como uma tentativa. Um sucesso resulta se a bola vai para a célula especificada, com probabilidade $\frac{1}{n}$; caso contrário, a tentativa resultará em uma falha, com probabilidade de $1 - \frac{1}{n}$.

Seja X o número de sucessos em r tentativas. Então

$$p_k = P(X = k) = \binom{r}{k} \left(\frac{1}{n}\right)^k \left(1 - \frac{1}{n}\right)^{r-k}, \quad k = 0, 1, 2, \dots, r.$$

Uma outra distribuição discreta foi descoberta por Siméon-Denis Poisson² e publicada em 1838 num trabalho de focava em certas variáveis aleatórias que contavam, entre outras coisas, o número de ocorrências contínuas que tinham lugar durante um intervalo de tempo de determinado comprimento.

Definição 3.14 (*Distribuição Poisson*)

Seja X uma variável aleatória assumindo valores inteiros não negativos. Dizemos que X tem por função de probabilidade Poisson de parâmetro $\theta > 0$ se

$$P(X = x) = \frac{e^{-\theta} \theta^x}{x!}, \quad x = 0, 1, 2, \dots \quad (3.15)$$

Resumidamente escrevemos $X \sim \text{Poisson}(\theta)$, caso a função de probabilidade da variável aleatória seja como em (3.15). Observemos que X neste caso é o resultado de uma contagem sem limite teórico conhecido. Verifiquemos que a função descrita em (3.15) é de probabilidade, para isso

$$\sum_{x=0}^{\infty} P(X = x) = \sum_{x=0}^{\infty} \frac{e^{-\theta} \theta^x}{x!} = e^{-\theta} \sum_{x=0}^{\infty} \frac{\theta^x}{x!} = e^{-\theta} e^{\theta} = 1.$$

Nesta demonstração utilizamos o Teorema de Taylor³ para escrever

$$e^{\theta} = \sum_{x=0}^{\infty} \frac{\theta^x}{x!}.$$

O Teorema de Taylor afirma que qualquer função que satisfaça certas condições pode ser representada como:

$$f(x) = f(0) + x f'(0) + \frac{x^2}{2!} f''(0) + \dots + \frac{x^{n-1}}{(n-1)!} f^{(n-1)}(0) + \int_0^x \frac{(x-u)^{n-1}}{(n-1)!} f^{(n)}(u) du,$$

conhecida como série de Taylor.

Uma forma diferente de escrever o caso $X \sim \text{Poisson}(\theta)$ é fazendo

$$X = \sum_{x=0}^{\infty} x \mathbf{1}_{\{X=x\}},$$

²Siméon Denis Poisson (1781-1840) foi um matemático e físico francês.

³O Teorema de Taylor foi criado por Brook Taylor em 1712 e publicado em 1715. Brook Taylor (1685 - 1731) foi um matemático inglês.

onde $\mathbf{1}_A(w) = \begin{cases} 0 & \text{se } \omega \notin A \\ 1 & \text{se } \omega \in A, \end{cases}$ é a função indicadora do conjunto A . Podemos escrever também

$$F_X(x) = \sum_{k=0}^{\infty} e^{-\theta} \frac{\theta^k}{k!} \delta(x - k),$$

como a função de distribuição de probabilidades de X .

Observemos que, nem sempre, as variáveis aleatórias discretas assumem uma quantidade finita de valores e como primeira situação temos a variável aleatória com distribuição ou função de probabilidade Poisson.

Definição 3.15 (*Distribuição Geométrica*)

Seja X uma variável aleatória assumindo valores inteiros não negativos. Dizemos que X tem por função de probabilidade Geométrica de parâmetro p , $0 < p < 1$, se

$$P(X = x) = p(1 - p)^x, \quad x = 0, 1, 2, \dots \quad (3.16)$$

Utiliza-se esta função de probabilidade quando contamos o número de tentativas independentes de um experimento até o primeiro sucesso, cada tentativa tem por probabilidade de sucesso p . Observamos que, para $X = n$ devemos ter que as primeiras $n - 1$ tentativas foram fracassos e, quando $X = n$, aconteceu um sucesso. É claro que $P(X = x) > 0$ para qualquer $x \geq 0$, também

$$\sum_{x=1}^{\infty} P(X = x) = \sum_{x=0}^{\infty} p(1 - p)^x = p \sum_{x=0}^{\infty} (1 - p)^x = \frac{p}{1 - (1 - p)} = 1, \quad (3.17)$$

do qual concluímos que, com probabilidade 1, um sucesso acabará acontecendo. Escrevemos resumidamente $X \sim \text{Geométrica}(p)$ sempre que a expressão da função de probabilidade de X for dada por (3.16). Para concluirmos que a soma das probabilidades em (3.17) é um, fizemos uso do resultado acerca da soma infinita de uma série geométrica, resultado apresentado em (3.19). Em matemática, uma série geométrica é uma série com uma relação constante entre termos sucessivos. Assim, a soma dos primeiros n termos de uma série geométrica é

$$a + ar + ar^2 + ar^3 + \dots + ar^{n-1} = \sum_{k=0}^{n-1} ar^k = a \left(\frac{1 - r^n}{1 - r} \right), \quad (3.18)$$

para $r \neq 1$ conhecido como proporção comum e a sendo o primeiro termo da série. Fazendo $n \rightarrow \infty$, o valor absoluto de r deve ser menor que um para a série convergir. Nessas condições, a soma torna-se então

$$a + ar + ar^2 + ar^3 + ar^4 + \dots = \sum_{k=0}^{\infty} ar^k = \frac{a}{1 - r}, \quad |r| < 1. \quad (3.19)$$

Exemplo 3.21

Suponha que $X \sim \text{Geométrica}(p)$. Vamos mostrar que, para qualquer inteiro não negativo k ,

$$P(X \geq k) = (1 - p)^k.$$

Utilizaremos o resultado em (3.18), o qual trata acerca de soma finita de uma série geométrica. Por definição

$$\begin{aligned} P(X \geq k) &= 1 - P(X \leq k - 1) = \sum_{x=0}^{k-1} p(1-p)^x \\ &= p + p(1-p) + p(1-p)^2 + \cdots + p(1-p)^{k-1} = p \left(\frac{1 - (1-p)^k}{1 - (1-p)} \right) \\ &= 1 - [1 - (1-p)^k] = (1-p)^k. \end{aligned}$$

Uma outra variável aleatória discreta bastante utilizada é a chamada de Binomial Negativa a qual definimos a continuação. Seja (Ω, \mathcal{F}, P) um espaço de probabilidade de um dado experimento estatístico e consideremos $A \in \mathcal{F}$ com $P(A) = p$. Em qualquer desempenho do experimento, se A acontece, chamamos isso de sucesso, caso contrário, um fracasso. Considere uma sucessão de tentativas deste experimento e calculemos a probabilidade de observar exatamente r sucessos, onde $r \geq 1$ é um número inteiro fixo. Se X denota o número de falhas que precedem r sucessos, $X + r$ é o número total de replicações necessárias para produzir r sucessos. Isso acontecerá se, e somente se, a última tentativa resultar em um sucesso e entre os $(r + X - 1)$ testes anteriores existirem exatamente X falhas. Segue que

$$P(X = x) = \binom{x+r-1}{x} p^r (1-p)^x, \quad x = 0, 1, 2, \dots \quad (3.20)$$

Reescrevendo a expressão em (3.20) na forma

$$P(X = x) = \binom{-r}{x} p^r (-q)^x, \quad x = 0, 1, 2, \dots; \quad q = 1 - p,$$

temos que $\sum_{x=0}^{\infty} \binom{-r}{x} (-q)^x = (1-q)^{-r} = p^{-r}$, do qual segue que $\sum_{x=0}^{\infty} P(X = x) = 1$.

Definição 3.16 (Distribuição Binomial Negativa)

Para o inteiro positivo fixo $r \geq 1$ e $0 < p < 1$, uma variável aleatória X com função de probabilidade dada por (3.20) é dita ter uma distribuição Binomial Negativa, ou seja, $X \sim BN(r; p)$ se a função de probabilidade é

$$P(X = x) = \binom{x+r-1}{x} p^r (1-p)^x, \quad x = 0, 1, 2, \dots$$

Podemos escrever a variável aleatória com distribuição Binomial Negativa como $X = \sum_{x=0}^{\infty} x \mathbf{1}_{\{X=x\}}$ e a função de distribuição como $F_X(x) = \sum_{k=0}^{\infty} \binom{k+r-1}{k} p^r (1-p)^k \delta(x-k)$.

Exemplo 3.22

Um dado é rolado repetidamente. Vamos calcular a probabilidade do evento A definido como “o número 2 aparecerá antes do número 5”. Com esse objetivo vamos considerar que A_j o evento que um 2 apareça na j -ésima tentativa, $j = 1, 2, \dots$, pela primeira vez e um 5 não aparece nos $j-1$ ensaios anteriores.

Então

$$P(A) = \sum_{j=1}^{\infty} P(A_j),$$

onde

$$P(A_j) = \frac{1}{6} \left(\frac{4}{6}\right)^{j-1}.$$

Segue que

$$P(A) = \sum_{j=1}^{\infty} \frac{1}{6} \left(\frac{4}{6}\right)^{j-1} = \frac{1}{2}.$$

Da mesma forma, a probabilidade de um 2 aparecer antes de um 5 ou um 6 é $1/3$ e assim por diante.

Se, no entanto, somos interessantes na distribuição do número de tentativas necessárias para obtermos r sucessos, temos, escrevendo $Y = X + r$ que

$$P(Y = y) = \binom{y-1}{r-1} p^r (1-p)^{y-r}, \quad y = r, r+1, \dots \quad (3.21)$$

Seja X uma variável aleatória com distribuição Binomial de parâmetros n e p e seja Y o variável aleatório definida em (3.21). Se houver r ou mais sucessos nas primeiras n tentativas, no máximo n tentativas foram necessárias para obter os primeiros r desses sucessos. Nós temos

$$P(X \geq r) = P(Y \leq n)$$

e também que

$$P(X < r) = P(Y > n)$$

No caso especial quando $r = 1$, a distribuição de X é chamada de Geométrica (Definição 3.15).

Exemplo 3.23 (*Problema da caixa de fósforos de Banach*)

Um matemático carrega uma caixa de fósforos em seus bolsos direito e esquerdo. Quando ele quer um fósforo, ele seleciona o bolso esquerdo com probabilidade p e o bolso direito com probabilidade $1-p$. Suponha que inicialmente cada caixa contenha N fósforos. Considere o momento em que o matemático descobre que uma caixa está vazia. Nesse momento, a outra caixa pode conter $0, 1, 2, \dots, N$ fósforos. Vamos identificar o sucesso

com a escolha do bolso esquerdo. A caixa no bolso esquerdo estará vazia no momento em que a caixa no bolso direito contiver exatamente r fósforos se, e somente se, exatamente $N - r$ falhas precederem o $(N + 1)$ -ésimo sucesso. Um argumento similar se aplica ao bolso direito e nós temos vamos chamar P_r a probabilidade de o matemático descobrir que uma caixa num bolso estive vazia enquanto a outra contiver r fósforos. Desta maneira vemos que

$$P_r = \binom{2N-r}{N-r} p^{N+1} q^{N-r} + \binom{2N-r}{N-r} q^{N+1} p^{N-r}.$$

Por fim, vejamos como definirmos mais uma distribuição discreta, chamada de hipergeométrica. Suponhas temos uma caixa que contém N bolinhas de gude. Desas, M são retirados aleatoriamente, marcadas e devolvidos à caixa. O conteúdo da caixa é então completamente misturado. Em seguida, n bolinhas de gude são sorteadas aleatoriamente da caixa e marcadas são contadas. Se X denota o número de bolinhas de gude marcadas, então

$$P(X = x) = \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}}. \quad (3.22)$$

Como x não pode exceder M ou n , devemos ter

$$x \leq \min(M, n). \quad (3.23)$$

Também, $x \geq 0$ e $N - M \geq n - x$, de maneira que

$$x \geq \max(0, M + n - N). \quad (3.24)$$

Definição 3.17 (*Distribuição Hipergeométrica*)

Uma variável aleatória X com função de probabilidade dada em (3.22) satisfazendo as restrições (3.23) e (3.24) é chamada de variável aleatória hipergeométrica.

Observemos que

$$\sum_{k=0}^n \binom{a}{k} \binom{b}{n-k} = \binom{a+b}{n},$$

para a e b números arbitrários e n um inteiro positivo. Segue então que

$$\sum_x P(X = x) = \frac{\sum_x \binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}} = 1.$$

Exemplo 3.24

Um lote composto por 50 lâmpadas é inspecionado tomando aleatoriamente 10 lâmpadas e testando-as. Se o número de lâmpadas defeituosas for no máximo 1, o lote é aceito; caso contrário, será rejeitado. Se houver, de fato, 10 lâmpadas defeituosas no lote, a probabilidade de aceitar o lote é

$$\frac{\binom{10}{1}\binom{40}{9}}{\binom{50}{10}} + \frac{\binom{40}{10}}{\binom{50}{10}} = 0.3487103.$$

De maneira geral podemos concluir a afirmação do teorema seguinte como resumo das funções de probabilidades.

Teorema 3.24

Seja $\{p_n\}_{n \geq 0}$ uma série de números reais não negativos tal que $\sum_{n=0}^{\infty} p_n = 1$. Então, os $\{p_n\}_{n \geq 0}$ correspondem aos valores de alguma função de probabilidade, para alguma variável aleatória X .

Demonstração: Exercício. ■

Neste caso dizemos também que a função de distribuição é discreta. A função de probabilidade P sempre existe, e se lhe conhece também como função de massa de probabilidade. Costuma-se utilizar o termo função de densidade como uma analogia com o caso de variáveis aleatórias contínuas e absolutamente contínuas, definidas mais adiante. Observe que a função de probabilidade P é uma função não negativa de soma um $\sum_i P(X = x_i) = 1$. Reciprocamente, como demonstrado no Teorema 3.24 toda função da forma como na Definição 3.10 que satisfaça estas duas propriedades chama-se função de probabilidade sem que, necessariamente, tenha sido definida uma variável aleatória.

Vejamos agora o caso contínuo.

3.2.2 Variáveis aleatórias contínuas**Definição 3.18**

Uma função de distribuição que é contínua em todos os pontos é chamada de função de distribuição contínua.

Sabemos que F_d é uma função crescente limitada que constitui a parte de salto de F , uma função de distribuição qualquer, e se for subtraída de F o restante deve ser positivo, não

conter mais saltos e assim ser contínua. Isso foi provado nos Teoremas 3.21 e Teorema 3.22. Vamos agora resumir esses dois teoremas, como segue.

Teorema 3.25

Toda função de distribuição F pode ser escrita como uma combinação lineal convexa de uma função de distribuição discreta F_d e outra contínua F_c , isto é, admite-se a seguinte representação

$$F(x) = \alpha F_d(x) + (1 - \alpha) F_c(x), \quad \forall x \in \mathbb{R},$$

onde $0 \leq \alpha \leq 1$ e esta tal decomposição é única.

Demonstração: Suponhamos que $F_c \neq 0$ e $F_d \neq 0$ no Teorema 3.21. Podemos então definir $\alpha = F_d(\infty)$ de modo que $0 < \alpha < 1$,

$$F_1 = \frac{1}{\alpha} F_d, \quad F_2 = \frac{1}{1 - \alpha} F_c$$

e escrever

$$F = \alpha F_1 + (1 - \alpha) F_2. \quad (3.25)$$

Agora, F_1 é uma função de distribuição discreta e F_2 é uma função de distribuição contínua e F é uma combinação convexa deles. Se $F_c = 0$, então F é discreta e estabelecemos $\alpha = 1$, $F_1 = F$, $F_2 = 0$; se $F_d = 0$ então F é contínua e nós temos $\alpha = 0$, $F_1 = 0$, $F_2 = F$, em qualquer caso extremo (3.25) permanece válido. ■

Definição 3.19 (Variável aleatória contínua)

A variável aleatória X disse-se contínua se sua correspondente função de distribuição é uma função contínua.

Exemplo 3.25

A classe das funções de distribuição contínuas é extensa, escolhemos agora duas delas, as quais apresentam expressões matemáticas simples e mostradas na Figura 3.7.

A continuação as expressões matemáticas das primeiras funções de distribuição contínuas escolhidas.

$$(a) \quad F(x) = \begin{cases} 0, & \text{caso } x \leq 0 \\ x, & \text{caso } 0 < x < 1 \\ 1, & \text{caso } x \geq 1 \end{cases} \quad e \quad (b) \quad F(x) = \begin{cases} 0, & \text{caso } x \leq 0 \\ 1 - e^{-x}, & \text{caso } 0 < x \end{cases}.$$

As distribuições contínuas se classificam, por sua vez, em distribuições absolutamente contínuas e distribuições singulares.

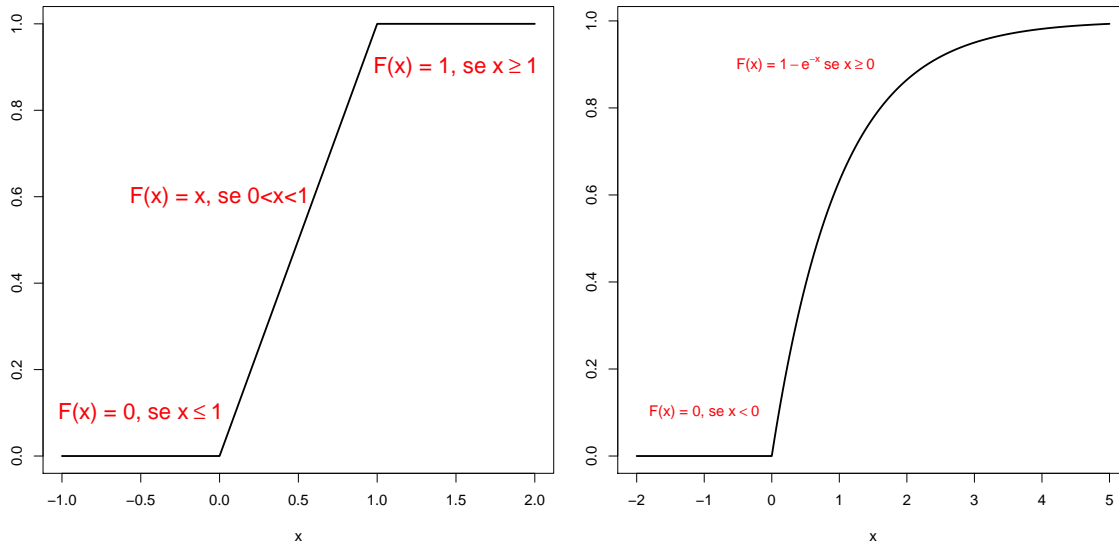


Figura 3.7: Funções de distribuição apresentadas no Exemplo 3.25. À esquerda a distribuição em (a) e a direita a distribuição definida em (b).

Definição 3.20 (*Variável aleatória absolutamente contínua*)

Uma variável aleatória contínua X , com função de distribuição $F_X(x)$ se disse absolutamente contínua, se existe uma função não negativa e integrável f tal que, para qualquer valor de $x \in \mathbb{R}$ se cumpre

$$F_X(x) = \int_{-\infty}^x f(u) du. \quad (3.26)$$

Em tais situações diz-se que a função $f(x)$ é função de densidade de X .

Mesmo quando exista uma função não negativa e integrável f que satisfaça (3.26), esta pode não ser única, pois basta modifica-la num ponto para que seja ligeiramente diferente e mesmo assim seguir cumprindo (3.26). Embora isso, nos referiremos à função de densidade como si fosse única e justifica-se pelo fato de que as probabilidades são as mesmas, seja utilizando uma função de densidade ou modificações dela que cumpram (3.26).

É claro que a função de densidade de uma variável aleatória absolutamente contínua é não negativa e sua integral sobre toda a recta real é um. Reciprocamente, toda função f não negativa que integre um em \mathbb{R} chama-se função de densidade. Se X é absolutamente contínua com função de distribuição F_X e função de densidade contínua f , então teorema fundamental do cálculo estabelece que, da relação em (3.26), $F'(x) = f(x)$, $\forall x \in \mathbb{R}$. Além disso, a probabilidade de que X tome um valor no intervalo (a, b) é a área baixo a função de densidade sobre o intervalo.

Isto se ilustra na Figura 2.6, a probabilidade é a mesma incluindo ou excluindo os extremos do intervalo. Podem construir-se exemplos de variáveis aleatórias contínuas que não tem função de densidade, isto é, que não exista uma função f não negativa e integrável sa-

tisfazendo (3.26) para qualquer número real x . Em tais situações diz-se que a distribuição é singular.

Definição 3.21 (*Variável aleatória singular*)

A variável aleatória contínua X ou sua correspondente função de distribuição contínua F_X , chama-se singular se $F'_X(x) = 0$ quase certamente, isto é, o conjunto nos quais $F'_X(x) = 0$ tem probabilidade 1 de ocorrer.

Podemos escrever diferente a afirmação neste definição: dizemos que X é uma variável aleatória contínua singular se

$$P(\{x \in \mathbb{R} : F'_X(x) = 0\}) = 1. \quad (3.27)$$

Uma situação interessante das variáveis aleatórias singulares, isto é, de variáveis aleatórias cujas funções de distribuição são contínuas porém não têm função de densidade é o caso da distribuição de Cantor. Uma excelente referência neste assunto é Dvlogoshey, Martio, Ryazanov & Vuorinen (2006).

Exemplo 3.26 (*Distribuição de Cantor*)

A distribuição de Cantor é a distribuição de probabilidade cuja função de distribuição é a função Cantor. Esta distribuição não tem nem uma função de densidade nem uma função de probabilidade, pois embora sua função de distribuição seja uma função contínua, a distribuição não é absolutamente contínua nem possui probabilidades positivas pontuais. Portanto, não é uma distribuição de probabilidade discreta nem absolutamente contínua, nem é uma mistura dessas. Pelo contrário, é um exemplo de uma distribuição singular. Esta função de distribuição é contínua em todos os lugares, mas horizontal em quase todos os lugares.

A distribuição de Cantor é a distribuição de probabilidade única para a qual em qualquer \mathcal{C}_n , $n \in \{0, 1, 2, 3, \dots\}$, que definem o conjunto de Cantor (ver Exemplo 2.5), a probabilidade de um intervalo particular em \mathcal{C}_n contendo a variável aleatória distribuída por Cantor é identicamente 2^{-n} em cada um dos 2^n intervalos.

Seja $x \in [0, 1]$, a expansão ternária de x é

$$x = \sum_{n=1}^{\infty} \frac{a_n(x)}{3^n}, \quad a_n(x) \in \{0, 1, 2\}. \quad (3.28)$$

Denotemos por N_x o menor n com $a_n(x) = 1$ se existir e escolhemos $N_x = \infty$ se não houver tal $a_n(x)$. Então, a função de Cantor $F : [0, 1] \rightarrow \mathbb{R}$ pode ser definida como

$$F(x) = \frac{1}{2^{N_x}} + \frac{1}{2} \sum_{n=1}^{N_x-1} \frac{a_n(x)}{2^n}. \quad (3.29)$$

Vamos demonstrar que F , definida em (3.29) é contínua e crescente, mas não absolutamente contínua.

Demonstração: Segue-se diretamente de (3.29) que F é uma função crescente, e além disso (3.29) implica que F é constante em todo intervalo $J \subseteq [0, 1] \setminus \mathcal{C}$, sendo \mathcal{C} o conjunto de Cantor. Observe também que se $x, y \in [0, 1]$, $x \neq y$, e x tende para y , então podemos tomar representações ternárias (3.28) para que

$$\min\{n : |a_n(x) - a_n(y)| \neq 0\} \rightarrow \infty.$$

Assim, a continuidade de F também decorre de (3.29). O fato de F ser monótona e a constância de F em cada intervalo J implica que F é singular. Devida a que F é singular e não constante então não é absolutamente contínua. Outro detalhe é que $F(x) \leq 1$, para todo $x \in [0, 1]$. ■

Mostramos a forma da distribuição de Cantor na Figura 3.8.

Um detalhe interessante é que o conjunto de Cantor \mathcal{C} pode ser construído como o conjunto de todos os pontos de $[0, 1]$ que tenham expansão como em (3.28), utilizando somente os dígitos 0 e 2. No caso $x \in \mathcal{C}$, temos $a_n(x) \in \{0, 2\}$ e a função de Cantor assume a forma

$$F(x) = \frac{1}{2} \sum_{n=1}^{\infty} \frac{a_n(x)}{2^n}.$$

Observe que existe uma sutil diferença entre uma função de distribuição discreta e a função de distribuição de Cantor. O detalhe é que toda função de distribuição discreta tem descontinuidade em pontos fixos, em todos os exemplos mostrados anteriormente de variáveis aleatórias X discretas, $P(X = x) > 0$ para valores fixos de x , por exemplo, nas distribuições Poisson e Geométrica em todo $x \in \mathbb{N}$, no caso Bernoulli de $X = x_1$ e $X = x_2$ fixos, no caso Binomial quando $X \in \{1, 2, \dots, n\}$. Isso não acontece na distribuição de Cantor, embora pareça discreta não é, é contínua e constante em todo intervalo $J \subseteq [0, 1] \setminus \mathcal{C}$. Pode-se demonstrar que $P(\mathcal{C}) = 0$, logo a distribuição de Cantor tem derivada $F'(x) = 0$ num conjunto de probabilidade 1.

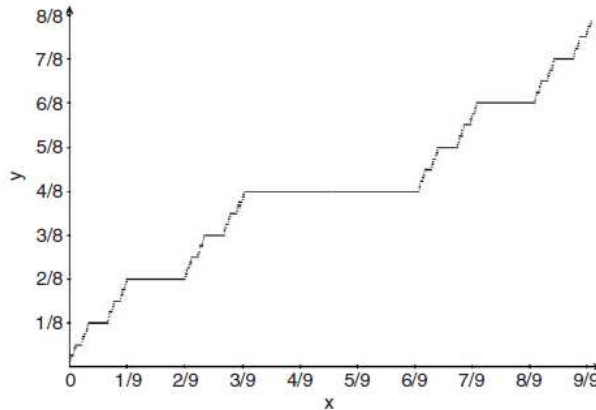


Figura 3.8: Função de distribuição de Cantor. Reprodução da figura mostrada no artigo de Dovgoshey *et al.* (2006).

Lembremos que todo conjunto da σ -álgebra de Borel pode ser obtido por uma quantidade enumerável de operações de uniões, interseções, complementos e diferenças de intervalos. Por isso, pode ser demonstrado o seguinte resultado.

Teorema 3.26

Seja X uma variável aleatória contínua com função de distribuição F_X . Então existe uma função real contínua de forma que

$$P(B) = \int_B f(t) dt,$$

para todo Boreliano B . Ainda temos que $F'_X(x) = \frac{dF_X(x)}{dx} = f(x)$.

Demonstração: Exercício. ■

Exemplo 3.27

Seja X uma variável aleatória com função de densidade triangular, isto é, cuja função de densidade é da forma

$$f(x; \theta) = \frac{\theta - |x|}{\theta^2}, \quad -\theta \leq x \leq \theta, \quad \theta > 0. \quad (3.30)$$

Não é difícil provar que esta é uma densidade. Observemos que $f \geq 0$ e que

$$F_X(x) = \begin{cases} 0, & \text{se } x < -\theta \\ \frac{1}{\theta^2} \int_{-\theta}^x (\theta + u) du = \frac{1}{2} + \frac{1}{\theta^2} \left(\theta x + \frac{x^2}{2} \right), & \text{se } -\theta \leq x < 0 \\ \frac{1}{2} + \frac{1}{\theta^2} \int_0^x (\theta - u) du = \frac{1}{2} + \frac{1}{\theta^2} \left(\theta x - \frac{x^2}{2} \right), & \text{se } 0 \leq x < \theta \\ 1, & \text{se } x \geq \theta \end{cases}$$

No caso discreto $P(X = a)$ é a probabilidade de que a variável aleatória X assuma o valor a . No caso contínuo $f(a)$ não é a probabilidade de X assumir o valor a . De fato, se X é do tipo contínuo, então assume cada um dos diferentes valores com probabilidade 0.

Teorema 3.27

Seja X uma variável aleatória de qualquer tipo. Então

$$P(X = a) = \lim_{\substack{t \rightarrow a \\ t < a}} P(t < X \leq a).$$

Demonstração: Seja $t_1 < t_2 < \dots < a$, $t_n \rightarrow a$ e escrevamos

$$A_n = \{t_n < X \leq a\}.$$

Então, $\{A_n\}_{n \geq 1}$ é uma sequência não-crescente de eventos que converge para

$$\bigcap_{n=1}^{\infty} A_n = \{X = a\}.$$

Segue então que

$$\lim_{n \rightarrow \infty} P(A_n) = P(X = a). \quad \blacksquare$$

Observemos que $P(t < X \leq a) = F_X(a) - F_X(t)$, do qual segue que

$$\begin{aligned} \lim_{\substack{t \rightarrow a \\ t < a}} P(t < X \leq a) &= P(X = a) = F_X(a) - \lim_{\substack{t \rightarrow a \\ t < a}} F_X(t) \\ &= F_X(a) - F_X(a^-). \end{aligned}$$

Assim, F_X tem uma descontinuidade de salto em a se, e apenas se, $P(X = a) > 0$, ou seja, F_X é contínua em a se, e somente se, $P(X = a) = 0$. Caso X uma variável aleatória do tipo contínuo, $P(X = a) = 0$ para todo $a \in \mathbb{R}$. Além, disso,

$$P(\mathbb{R} \setminus \{a\}) = 1.$$

Desejamos lembrar mais uma vez que, se $P(A) = 0$ para algum $A \in \mathcal{B}(\mathbb{R})$, chamamos de A um evento com probabilidade zero ou um evento nulo. No entanto, não significa que $A = \emptyset$. Da mesma forma, se $P(B) = 1$ para algum $B \in \mathcal{B}(\mathbb{R})$, chamamos B de evento certo, mas não segue por isso que $B = \Omega$.

Exemplo 3.28

Seja X uma variável aleatória com função de distribuição da forma

$$F_X(x) = \begin{cases} 0, & \text{se } x \leq 0, \\ x, & \text{se } 0 < x \leq 1, \\ 1, & \text{se } 1 < x \end{cases}.$$

Diferenciando F_X em relação a x , nos pontos de continuidade, obtemos a forma da função de densidade como

$$f(x) = \frac{dF_X(x)}{dx} = \begin{cases} 0, & \text{se } x \notin (0, 1) \\ 1, & \text{se } 0 < x < 1 \end{cases}.$$

A função f não é contínua em $x = 0$ ou $x = 1$. Podemos definir $f(0)$ e $f(1)$ de qualquer maneira. Escolhendo $f(0) = f(1) = 0$, temos que

$$f(x) = \begin{cases} 1, & \text{se } 0 < x < 1 \\ 0, & \text{caso contrário} \end{cases}.$$

Então, por exemplo, $P(0.3 < X \leq 0.55) = F_X(0.55) - F_X(0.3) = 0.25$.

Vejamos agora duas propriedades importantes da função de densidade. Normalmente, é mais conveniente especificar o comportamento distribucional de uma variável aleatória contínua por meio de sua função de densidade do que por meio de sua função de distribuição. A função de densidade descreve essencialmente a probabilidade de X assumir valores em uma vizinhança de um ponto x e, portanto, é análoga à função de probabilidades de uma variável aleatória discreta. Da Definição 3.20 acima, temos

$$P(x < X < x + \epsilon) = \int_x^{x+\epsilon} f(u) \, du \approx \epsilon f(x) \geq 0, \quad \forall x,$$

se ϵ for pequeno e f contínua em x . Assim temos demonstrado a primeira propriedade da função de densidade, ela é sempre não negativa. Uma segunda propriedade da função de densidade é que

$$\int_{-\infty}^{\infty} f(x) \, dx = \lim_{x \rightarrow \infty} F_X(x) - \lim_{x \rightarrow -\infty} F_X(x) = 1.$$

É importante notar que as funções de densidade não são exclusivamente determinadas; isto é, dada uma função de densidade f é sempre possível encontrar uma outra função de densidade f^* tal que

$$\int_a^b f(x) \, dx = \int_a^b f^*(x) \, dx,$$

para todos a e b , mas para os quais f^* e f diferem em um finito número de pontos. Esta não-singularidade não coloca problemas, em geral, para cálculos de probabilidade, mas pode representar problemas sutis em determinadas aplicações estatísticas.

Se a função de distribuição F for diferenciável em x , podemos encontrar a função de densidade f como sendo a derivada de F em x : $f(x) = F'(x)$, segundo nos disse o Teorema 3.26. Para fins de manipulação, podemos assumir que $f(x) = F'(x)$ quando a derivada é bem definida e definimos $f(x)$ arbitrariamente quando não é. Vejamos isto no seguinte exemplo.

Exemplo 3.29

Suponhamos que X tenha por função de densidade

$$f(x) = \begin{cases} cx^3, & \text{para } 0 \leq x \leq 1 \\ 0, & \text{caso contrário} \end{cases},$$

onde c é uma constante positiva. Para determinar o valor de c , notamos que

$$1 = P(-\infty < X < \infty) = \int_{-\infty}^0 0 \, dx + \int_0^1 cx^3 \, dx + \int_1^{\infty} 0 \, dx = \frac{c}{4},$$

do qual obtemos que $c = 4$. Podemos também determinar a função de distribuição integrando a densidade de $-\infty$ a qualquer ponto x ; nós obtemos

$$F_X(x) = \begin{cases} 0, & \text{caso } x < 0 \\ x^4, & \text{caso } 0 \leq x \leq 1 \\ 1, & \text{caso } x > 1 \end{cases}.$$

Como observamos acima, podemos modificar a função de densidade em um número enumerável de pontos e ainda obter a mesma função de distribuição. Por exemplo, se

definirmos

$$f^*(x) = \begin{cases} cx^3, & \text{para } 0 < x < 1 \\ 0, & \text{caso contrário} \end{cases},$$

então obtemos a mesma função de distribuição, embora f^* difere de f em $x = 0$ e $x = 1$.

Observemos que se $f \geq 0$ e satisfaz

$$\lim_{x \rightarrow \infty} F_X(x) = F_X(+\infty) = \int_{-\infty}^{+\infty} f(t) dt = 1.$$

Sejam a e b dois números reais tais que $a < b$. Então

$$P(a < X \leq b) = F_X(b) - F_X(a) = \int_a^b f(t) dt.$$

Teorema 3.28 (*Caracterização de função de densidade*)

Toda função real integrável f sobre \mathbb{R} , satisfazendo:

(a) $f \geq 0$ e

(b) $\int_{-\infty}^{+\infty} f(x) dx = 1$

é a função de densidade de alguma aleatória X de tipo absolutamente contínuo.

Demonstração: Segundo o Teorema 3.26 basta mostrar que a f corresponde uma função de distribuição F . Definimos

$$F(x) = \int_{-\infty}^x f(t) dt, \quad x \in \mathbb{R}.$$

■

Exemplo 3.30

Segundo o Teorema 3.28, qualquer seja a função real integrável que satisfaça os itens (a) e (b) é uma função de densidade. Vejamos a seguinte situação:

$$f(x) = \frac{1}{\pi \sqrt{x(1-x)}}, \quad 0 < x < 1.$$

Observemos que $f \geq 0$, para todo $x \in (0, 1)$ e verifica-se que o item (b) também vale. Esta é chamada de densidade arco-seno. Verificamos que f , acima definida, é uma função de densidade sem mesmo dizermos qual é a variável aleatória.

Estudaremos agora distribuições absolutamente contínuas usadas com mais frequência, às quais chamaremos de modelos probabilísticos contínuos, e descreveremos algumas de suas propriedades importantes. Devemos deixar claro que a classe das variáveis aleatórias absolutamente contínuas contém muito mais modelos do que a classe das variáveis aleatórias

singulares, por este motivo, muitas vezes confundem-se estas duas classes e consideram-se as variáveis absolutamente contínuas como sendo toas as variáveis aleatórias contínuas. Em todas as situações definidas a continuação a função que descreve o comportamento da variável aleatória particular é uma função de densidade. Também de vemos mencionar que os modelos probabilísticos contínuos definem-se a partir da função de densidade. No caso discreto modelos foram descritos a partir da função de probabilidade.

Definição 3.22 (*Distribuição Uniforme*)

Uma variável aleatória X diz-se satisfaz a distribuição $Uniforme(\alpha, \beta)$ contínua se tem por função de densidade

$$f(x) = \begin{cases} \frac{1}{\beta - \alpha}, & \text{quando } \alpha \leq x \leq \beta \\ 0, & \text{caso contrário} \end{cases}. \quad (3.31)$$

Escrevemos $X \sim Uniforme(\alpha, \beta)$.

Observa-se que nesta definição nada dizemos acerca das constantes α e β , entendemos que $\alpha < \beta$ e nada mais. Significa que ambas podem ser números negativos, α pode ser negativa enquanto β um número positivo ou ambas quantidades podem ser positivas. Qualquer seja a situação se $X \sim Uniforme(\alpha, \beta)$ a densidade é como em (3.31). Nesta situação a função de distribuição é

$$F_X(x) = \begin{cases} 0, & \text{se } x < \alpha \\ \frac{x - \alpha}{\beta - \alpha}, & \text{se } \alpha \leq x \leq \beta \\ 1, & \text{se } x > \beta \end{cases}.$$

Assumindo que $(\alpha, \beta) = (0, 1)$ temos a distribuição $Uniforme(0, 1)$, também chamada Uniforme padrão, apresentamos esta função de densidade assim como a função de distribuição correspondente na Figura 3.9.

Teorema 3.29 (*Transformação Integral de Probabilidade*)

Seja X uma variável aleatória absolutamente contínua com função de distribuição F_X . Então, $F_X(X)$ é uma variável aleatória com distribuição $Uniforme(0, 1)$, ou seja, $F_X(X) \sim Uniforme(0, 1)$.

Demonstração: Definamos $Y \sim F_X(X)$. Então

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(F_X(X) \leq y) \\ &= P(X \leq F_X^{-1}(y)) = F_X(F_X^{-1}(y)) = y, \quad 0 \leq y \leq 1. \end{aligned}$$

F_Y é a função de distribuição de uma variável aleatória $Uniforme(0, 1)$. ■

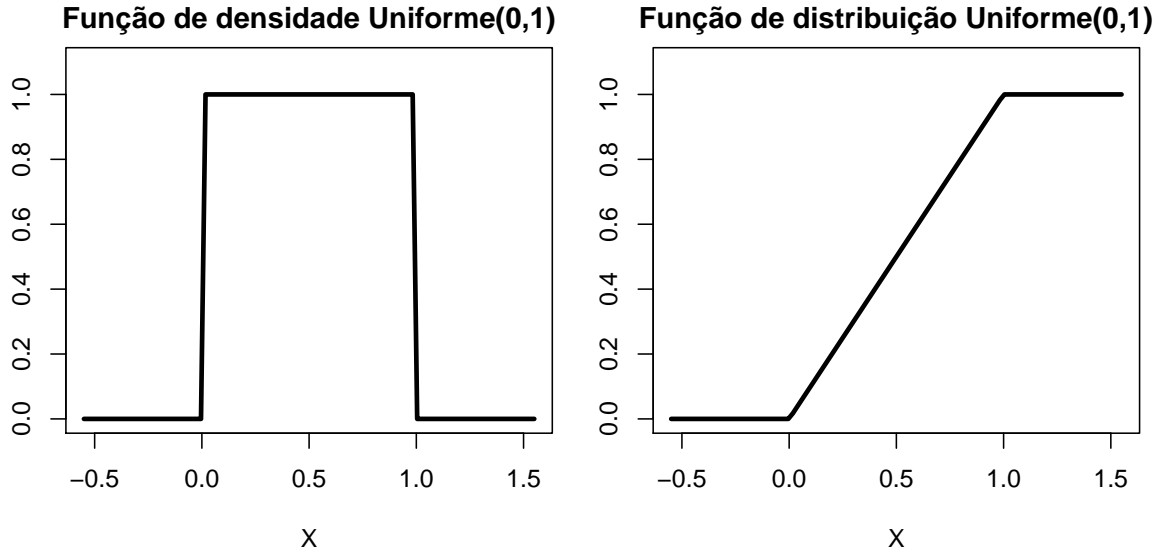


Figura 3.9: Função de densidade (esquerda) e de distribuição (direita) *Uniforme*(0,1).

Exemplo 3.31

Suponhas que X tenha a função de densidade dada por

$$f(x) = \begin{cases} 0, & \text{se } x < 0 \\ e^{-x}, & \text{se } x \geq 0 \end{cases}.$$

Então,

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

Seja agora $Y = F_X(X) = 1 - e^{-X}$. Encontremos a função de densidade de Y .

Com esse objetivo, vemos que a função de distribuição da variável aleatória Y é

$$F_Y(y) = \begin{cases} 0, & \text{se } y < 0 \\ P(1 - e^{-X} \leq y), & \text{se } 0 < y \leq 1 \\ 1, & \text{se } y > 1 \end{cases} = \begin{cases} 0, & \text{se } y < 0 \\ 1 - e^{\ln(1-y)}, & \text{se } 0 < y \leq 1 \\ 1, & \text{se } y > 1 \end{cases},$$

do qual concluímos que $F_Y(y) = y$, quando $y \in (0, 1)$ e zero caso contrário, portanto, Y é uma variável aleatória com distribuição *Uniforme*(0,1).

Pede-se ao leitor que considere o que acontece quando F é a função de distribuição de uma variável aleatória discreta. Uma aplicação da propriedade apresentada no Teorema 3.29, da distribuição Uniforme, é o chamada resíduo quantil nos modelos de regressão generalizados e outros modelos de regressão avançados (Dunn & Smyth, 1996). Podemos perceber que, mesmo sendo a distribuição *Uniforme*(0,1) simples, como função de densidade, a aplicabilidade é grande. Na direção inversa, o seguinte resultado é válido.

Teorema 3.30

Seja F uma função de distribuição e X uma variável aleatória com distribuição $Uniforme(0, 1)$. Então existe uma função h de maneira que $h(X)$ têm por distribuição F , ou seja,

$$P(h(X) \leq x) = F(x), \quad \forall x \in (-\infty, +\infty).$$

Demonstração: Temos três situações:

- (i) Seja F uma função de distribuição discreta. Consideremos que $Y \sim F$ e que $P(Y = y_k) = p_k$, $k = 1, 2, 3, \dots$. Definamos h como segue:

$$h(x) = \begin{cases} y_1, & \text{se } 0 \leq x < p_1 \\ y_2, & \text{se } p_1 \leq x < p_1 + p_2 \\ y_3, & \text{se } p_1 + p_2 \leq x < p_1 + p_2 + p_3 \\ \vdots & \vdots \end{cases}.$$

Então

$$\begin{aligned} P(h(X) = y_1) &= P(0 \leq X < p_1) = p_1 \\ P(h(X) = y_2) &= P(p_1 \leq X < p_1 + p_2) = p_2 \\ P(h(X) = y_3) &= P(p_1 + p_2 \leq X < p_1 + p_2 + p_3) = p_3 \\ \vdots & \quad \quad \quad \vdots \end{aligned}$$

e, em geral, $P(h(X) = y_k) = p_k$, $k = 1, 2, 3, \dots$. Portanto, $h(X)$ é uma variável aleatória discreta com distribuição F .

- (ii) Seja F uma função de distribuição contínua e estritamente crescente. Neste caso F^{-1} está bem definida e podemos escolher $h(X) = F^{-1}(X)$. Temos então que

$$\begin{aligned} P(h(X) \leq x) &= P(F^{-1}(X) \leq x) \\ &= P(X \leq F(x)) = F(x). \end{aligned}$$

- (iii) No caso geral, definimos

$$F^{-1}(y) = \min\{x : F(x) \geq y\} \quad (3.32)$$

e seja $h(X) = F^{-1}(X)$. Acontece que, se y satisfaz que $F^{-1}(y) \leq x$, isto implica que para todo $\epsilon > 0$, $y \leq F(x + \epsilon)$. Desde que $\epsilon > 0$ é arbitrário e F contínua a direita, fazendo $\epsilon \rightarrow 0$ concluímos que $y \leq F(x)$. Dado que $y \leq F(x)$ implica que $F^{-1}(y) \leq x$, pela definição em (3.32) segue que

$$\{y \in \mathbb{R} : F^{-1}(y) \leq x\} = \{y \in \mathbb{R} : y \leq F(x)\}.$$

Então

$$P(F^{-1}(X) \leq x) = P(X \leq F(x)) = F(x).$$



Este teorema é bastante útil na geração de amostras com a ajuda da distribuição uniforme: escolhemos um $y \in (0, 1)$ arbitrário e, em seguida, usamos o inverso da função de distribuição desejada $x = F_X^{-1}(y)$ para encontrar um elemento x da distribuição correspondente F_X .

Exemplo 3.32

Vejam os o procedimento teórico de geração de dados na distribuição logística. Consideremos a função de densidade logística de parâmetros $\mu \in \mathbb{R}$ e $\sigma > 0$, definida como

$$f(x) = \frac{1}{\sigma} \frac{e^{-(x-\mu)/\sigma}}{(1 + e^{-(x-\mu)/\sigma})^2}, \quad x \in \mathbb{R}.$$

A função de distribuição logística é da forma

$$F_X(x) = \frac{1}{1 + e^{-(x-\mu)/\sigma}}.$$

Queremos encontrar um ponto da distribuição F_X .

Para isso escolhemos $y \in (0, 1)$ qualquer e fazemos

$$y = \frac{1}{1 + e^{-(x-\mu)/\sigma}},$$

de uma simples álgebra, obtemos que

$$\ln\left(\frac{y}{1-y}\right) = \frac{x-\mu}{\sigma},$$

de maneira que o valor procurado de F_X é $x = \mu + \sigma \ln\left(\frac{y}{1-y}\right)$.

Uma caracterização interessante da distribuição uniforme é dada pelo seguinte teorema, este resultado será utilizado para caracterizar algumas outras distribuições.

Teorema 3.31

Seja X uma variável aleatória absolutamente contínua com valores em $(0, 1)$ ou suporte o intervalo $(0, 1)$. Se $P(x < X \leq y)$ depende somente do comprimento $y - x$ para todo $0 \leq x \leq y < 1$, então

$$X \sim \text{Uniforme}(0, 1).$$

Demonstração: Seja $P(x < X \leq y) = f(y - x)$, sendo f alguma função. Então,

$$f(x+y) = P(0 < X \leq x+y) = P(0 < X \leq x) + P(x < X \leq x+y) = f(x) + f(y). \quad (3.33)$$

Observemos que f é contínua à direita. Temos então que, pela relação em (3.33), $f(x) = f(x) + f(0)$, de maneira que $f(0) = 0$. Também

$$0 = f(0) = f(x - x) = f(x) + f(-x),$$

do qual vemos que $f(-x) = -f(x)$. Isso mostra que $f(x) = cx$ para alguma constante c . É suficiente mostrar o resultado para x positivo. Seja m um inteiro, então

$$f(\underbrace{x + x + \cdots + x}_{m \text{ vezes}}) = \underbrace{f(x) + f(x) + \cdots + f(x)}_{m \text{ vezes}} = mf(x).$$

Escolhendo $x = n/m$, sendo $n < m$, temos que

$$f\left(m \times \frac{n}{m}\right) = mf\left(\frac{n}{m}\right),$$

de maneira que

$$f\left(\frac{n}{m}\right) = \frac{1}{m}f(n) = \frac{n}{m}f(1),$$

para n e m inteiros positivos. Escolhendo $f(1) = c$, provamos que $f(x) = cx$, para número racionais x . Para completarmos a demonstração vamos considerar o caso onde x é um número irracional positivo. Podemos encontrar uma sequência decrescente de números racionais positivos x_1, x_2, \dots tais que $\lim_{n \rightarrow \infty} x_n = x^-$. Como f é contínua à direita,

$$f(x) = \lim_{x_n \rightarrow x^-} f(x_n) = \lim_{x_n \rightarrow x^-} cx_n = cx.$$

Agora, para $0 \leq x \leq 1$,

$$\begin{aligned} F(x) &= P(X \leq 0) + P(0 < X \leq x) \\ &= F(0) + P(0 < X \leq x) = f(x) = cx. \end{aligned}$$

Dado que $F(1) = 1$, nós devemos ter $c = 1$, de modo que $F(x) = x$, $0 \leq x \leq 1$. ■

Definição 3.23 (*Distribuição Exponencial*)

Seja X uma variável aleatória assumindo valores reais não negativos. Dizemos que X tem por função de densidade Exponencial de parâmetro $\theta > 0$ se

$$f(x) = \theta e^{-\theta x}, \quad x > 0. \quad (3.34)$$

Não é difícil demonstrarmos que se X tem densidade como em (3.34) a integral definida, no intervalo de variação da variável $[0, \infty)$, é um e ainda temos que a função de distribuição é $F_X(x) = 1 - e^{-\theta x}$. Dado que a função de distribuição F_X é estritamente crescente sobre o conjunto onde $F_X(x) > 0$, podemos determinar a inversa F_X^{-1} resolvendo a equação

$$1 - \exp\left(-\theta F^{-1}(x)\right) = x,$$

que produz $F^{-1}(x) = -\frac{1}{\theta} \ln(1 - x)$. Então, se $U \sim \text{Uniforme}(0, 1)$ temos que

$$X = -\theta^{-1} \ln(1 - U) \sim \text{Exponencial}(\theta).$$

Observe que $1 - U$ tem a mesma distribuição que U , também temos que $-\theta^{-1} \ln(U) \sim \text{Exponencial}(\theta)$.

Uma observação interessante é que uma função de distribuição pode corresponder a várias variáveis aleatórias no mesmo espaço de probabilidade (Ω, \mathcal{F}, P) . Por exemplo, seja X uma variável aleatória com função de distribuição $F_X(x) = e^{-x^2/2}/\sqrt{2\pi}$, definida para todo $x \in \mathbb{R}$. Então, a variável aleatória $-X$ tem a mesma função de distribuição e, conseqüentemente, $F_X = F_{-X}$. No entanto, $P(X = -X) = P(2X = 0) = P(X = 0) = 0$.

Uma função interessante é a chamada generalização do fatorial ou fatorial de um número real. Esta função é a gama, definida como

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx, \quad (3.35)$$

quando $\alpha > 0$. Em particular, se $\alpha = 1$, $\Gamma(1) = 1$. Caso $\alpha > 1$, por integração por partes prova-se que

$$\Gamma(\alpha) = (\alpha - 1) \int_0^\infty x^{\alpha-2} e^{-x} dx = (\alpha - 1)\Gamma(\alpha - 1).$$

Caso $\alpha = n$ seja um inteiro, temos que $\Gamma(n) = (n - 1)!$. Utilizaremos a função gama para definirmos a seguinte função de densidade.

Existe uma relação especial entre a distribuição Poisson e a função gama incompleta⁴. Seja $X \sim \text{Poisson}(\lambda)$, então

$$P(X \leq x) = \frac{1}{x!} \int_\lambda^\infty e^{-u} u^x du, \quad (3.36)$$

de maneira que expressamos a função de distribuição Poisson em termos da função gama incompleta. Para percebermos que a relação acima é verdadeira, observemos que

$$\frac{d}{dx} P(X \leq x) = \sum_{k=0}^x \frac{1}{k!} \left(k e^{-\lambda} \lambda^{k-1} - \lambda^k e^{-\lambda} \right) = \frac{-\lambda^x e^{-\lambda}}{x!},$$

do qual segue que a expressão em (3.36).

Definição 3.24 (*Distribuição Gama*)

Seja X uma variável aleatória assumindo valores positivos. Dizemos que X tem por função de densidade gama de parâmetros $\alpha > 0$ e $\beta > 0$ se

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}, \quad x \geq 0, \quad (3.37)$$

e zero caso contrário. Escrevemos $X \sim \text{Gama}(\alpha, \beta)$.

⁴A função gama incompleta é uma variação da função gama completa definida em (3.35). No caso da gama incompleta define-se por uma integral com um dos limites, superior ou inferior, finito e não infinito.

Um caso especial acontece quando $\alpha = 1$, nesta situação chegamos à distribuição exponencial de parâmetro $1/\beta$. A função de distribuição de $X \sim Gama(\alpha, \beta)$ é da forma

$$F_X(x) = \begin{cases} 0, & \text{se } x \leq 0 \\ \frac{1}{\Gamma(\alpha)\beta^\alpha} \int_0^x u^{\alpha-1} e^{-u/\beta} du, & \text{se } x > 0 \end{cases}. \quad (3.38)$$

Na Figura 3.10 apresentamos diferentes formas das densidade exponencial, gama e χ^2 , esta última definida logo a depois da figura. Nas definições destas distribuições assim como no caso uniforme, estas e todas as funções de densidade apresentadas e que depois iremos definir dependem de valores reais que, se não especificados são chamados de parâmetros. Desta forma a densidade $Uniforme(\alpha, \beta)$ depende dos parâmetros $\alpha < \beta$, $\alpha, \beta \in \mathbb{R}$; a densidade $Exponencial(\theta)$, depende do parâmetro $\theta > 0$ e a densidade $Gama(\alpha, \beta)$ depende dos parâmetros $\alpha > 0$ e $\beta > 0$. Podemos dizer que, mais do uma função de densidade, definimos famílias de funções de densidade em todas as situações anteriores. O mesmo acontece com as funções de probabilidade.

Dependendo de diferentes valores dos parâmetros que definem as densidades exponencial, gama e χ^2 mostramos nas Figuras 3.10 e 3.11 alguns dos elementos que formam as famílias mencionadas.

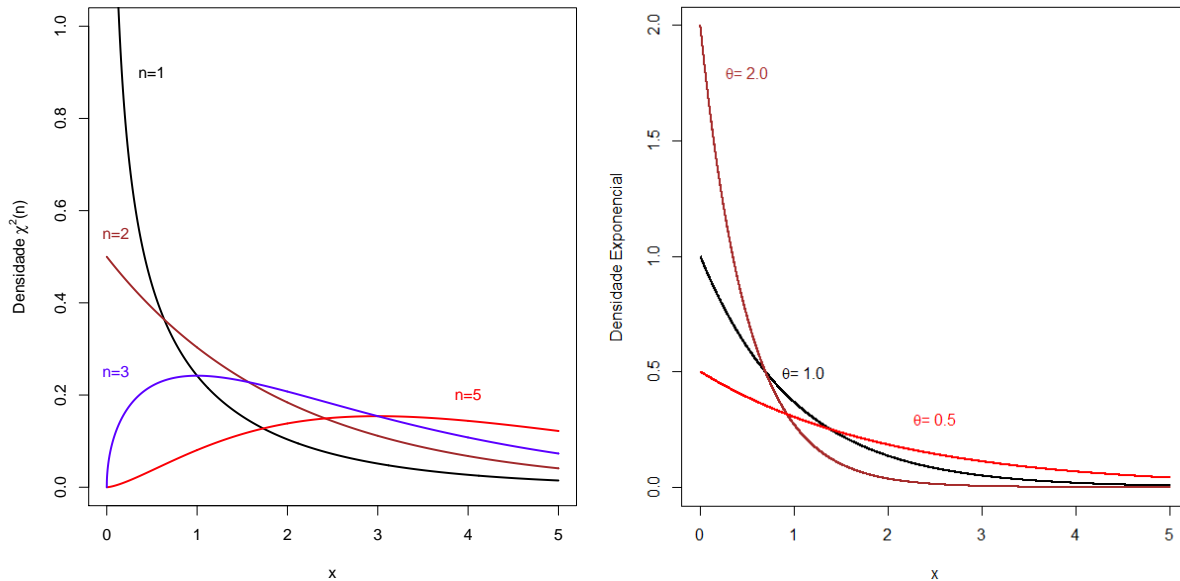


Figura 3.10: À esquerda mostramos diferentes formas que pode assumir a função de densidade χ^2 , com valores do parâmetro $n = 1, 2, 3, 5$. À direita mostramos a forma correspondente assumida pela distribuição $Exponencial(\theta)$ com valores do parâmetro $\theta = 0.5, 1.0$ e $\theta = 2.0$. Perceba como são parecidas estas densidades, conforme os valores dos parâmetros.

Um outro caso especial da densidade gama acontece quando $\alpha = n/2$, $n > 0$, $n \in \mathbb{N}$ e $\beta = 2$, esta situação chama-se distribuição qui-quadrado.

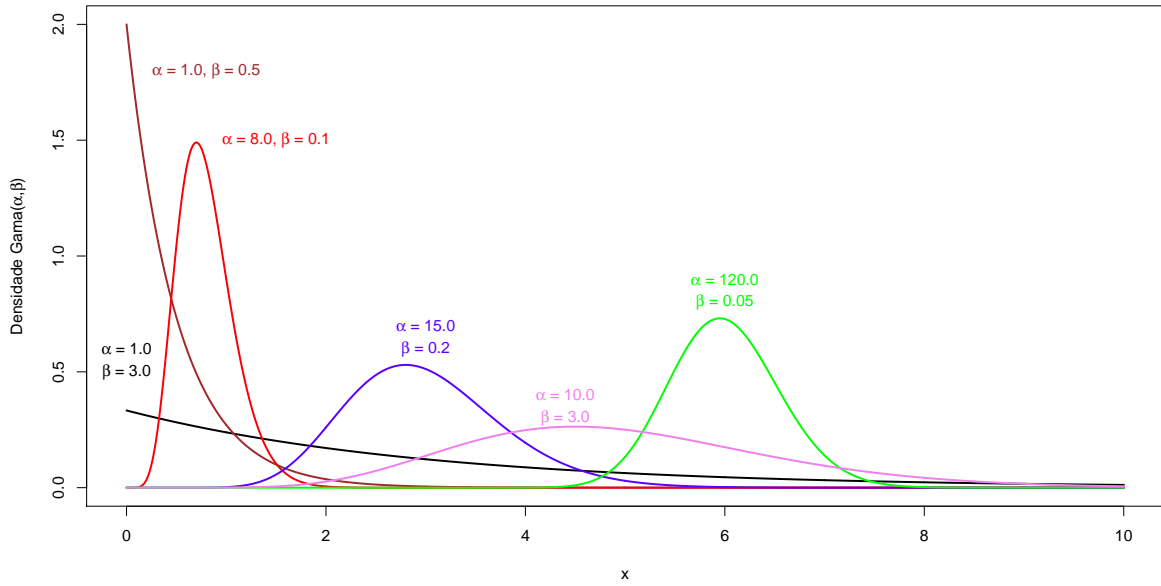


Figura 3.11: No gráfico mostramos formas assumidas pela densidade $Gama(\alpha, \beta)$, para os valores dos parâmetros mostrados.

Definição 3.25 (*Distribuição χ^2*)

Dizemos que X tem por função de densidade χ^2 com n graus de liberdade se

$$f(x) = \begin{cases} 0, & \text{se } x \leq 0 \\ \frac{1}{\Gamma(n/2)2^{n/2}} x^{n/2-1} e^{-x/2}, & \text{se } x > 0 \end{cases}. \quad (3.39)$$

Resumidamente escrevemos $X \sim \chi^2(n)$.

Acontece que se $X \sim Poisson(\lambda)$, sua função de distribuição pode ser escrita em termos da função gama incompleta ou em termos da distribuição χ^2 da seguinte maneira:

$$\begin{aligned} \frac{d}{d\lambda} P(X \leq x) &= \sum_{k=0}^x \frac{1}{k!} (k e^{-k} \lambda^{k-1} - \lambda^k e^{-\lambda}) \\ &= \frac{-\lambda^x e^{-\lambda}}{x!}, \end{aligned}$$

disto segue que

$$P(X \leq x) = \frac{1}{x!} \int_{\lambda}^{\infty} u^x e^{-u} du. \quad (3.40)$$

A distribuição qui-quadrado é uma das distribuições de probabilidade mais amplamente usadas nas estatísticas inferenciais, principalmente no teste de hipóteses ou na construção de intervalos de confiança. A distribuição definida em (3.39) é às vezes chamada de distribuição

qui-quadrado central. Essa distribuição foi descrita pela primeira vez pelo estatístico alemão Friedrich Robert Helmert em artigos de 1875 a 1876, onde ele calculou a distribuição amostral da variação da amostra de uma população normal.

Uma outra forma de escrever a expressão em (3.40) é a seguinte:

$$P(X \leq x) = P(Y \geq 2\lambda),$$

onde $X \sim \text{Poisson}(\lambda)$ e $Y \sim \chi^2(2x + 2)$.

No seguinte exemplo definimos uma distribuição chamada de Cauchy⁵.

Exemplo 3.33 (*Distribuição Cauchy*)

A variável aleatória X diz-se ter distribuição Cauchy de parâmetros $\sigma > 0$ e $\mu \in \mathbb{R}$ se tiver por função de densidade

$$f(x) = \frac{\sigma}{\pi} \frac{1}{\sigma^2 + (x - \mu)^2}, \quad x \in \mathbb{R}. \quad (3.41)$$

Escrevemos $X \sim \text{Cauchy}(\mu, \sigma)$. Vamos demonstrar que a expressão em (3.41) é uma função de densidade. Para isso notemos que $f \geq 0$, portanto devemos provar somente que a integral é 1. Fazendo a substituição $y = (x - \mu)/\sigma$, temos que

$$\int_{-\infty}^{\infty} f(x) dx = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{dy}{1 + y^2} = \frac{2}{\pi} \tan^{-1}(y) \Big|_{-\infty}^{\infty} = 1.$$

Observemos que para provarmos que a função em (3.41) é função de densidade, transformamos $X \sim \text{Cauchy}(\mu, \sigma)$ em uma nova variável aleatória $Y \sim \text{Cauchy}(0, 1)$. Nesta situação, a função de distribuição assume a forma

$$F_Y(y) = \frac{1}{2} + \frac{1}{\pi} \tan^{-1}(y), \quad -\infty < y < \infty.$$

Diversas propriedades da distribuição Cauchy podem ser encontradas em Pitman & Williams (1967).

Definição 3.26 (*Distribuição Beta*)

Uma variável aleatória X é dita ter por distribuição Beta de parâmetros $\alpha > 0$ e $\beta > 0$ se sua função de densidade tiver a forma

$$f(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1 - x)^{\beta-1}, \quad 0 < x < 1. \quad (3.42)$$

Escrevemos $X \sim \text{Beta}(\alpha, \beta)$.

⁵Augustin-Louis Cauchy (1789 - 1857) foi um matemático, engenheiro e físico francês que fez contribuições pioneiras para vários ramos da matemática, incluindo análise matemática e mecânica contínua.

A função de distribuição $Beta(\alpha, \beta)$ é da forma

$$F_X(x) = \begin{cases} 0, & \text{se } x \leq 0 \\ \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^x u^{\alpha-1}(1-u)^{\beta-1} du, & \text{se } 0 < x < 1 \\ 1, & \text{se } 1 \leq x \end{cases}.$$

Observemos que no caso especial $\alpha = \beta = 1$, temos a distribuição $Uniforme(0, 1)$. Uma característica importante é se $X \sim Beta(\alpha, \beta)$, então $1 - X \sim Beta(\beta, \alpha)$. Em particular, $X \sim Beta(\alpha, \alpha)$ se, e somente se, $1 - X \sim Beta(\alpha, \alpha)$. Estas afirmações implicam que a função de densidade em (3.42) satisfaz que

$$f(x) = f(1 - x), \quad 0 < x < 1.$$

Devemos destacar que se X e $1 - X$ tem a mesma distribuição, isto não significa que $X \sim Beta(\alpha, \alpha)$. Para vermos isso basta assumirmos que a densidade de X seja

$$f(x) = 2 \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \left((1-x)^{\alpha-1}x^{\beta-1} + (1-x)^{\alpha-1}x^{\beta} - 1 \right),$$

a qual satisfaz que $f(x) = f(1 - x)$, mas não é a densidade $Beta(\alpha, \beta)$.

Definição 3.27 (*Distribuição Normal*)

Seja X uma variável aleatória assumindo valores reais. Dizemos que X tem por função de densidade Normal se

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad x \in \mathbb{R}, \quad (3.43)$$

resumidamente escrevemos $X \sim N(0, 1)$.

Esta função de densidade é conhecida como densidade Normal padrão, isso porque não depende de parâmetros. Outra forma de definir a densidade Normal será apresentada na Seção 3.3. Vamos demonstrar agora que a função em (3.43) é uma densidade. Como f é não-negativa uma ideia então é verificar que o quadrado da integral é igual a 1. Temos que

$$\begin{aligned} \left(\int_{-\infty}^{+\infty} f(z) dz \right)^2 &= \left(\int_{-\infty}^{+\infty} f(x) dx \right) \times \left(\int_{-\infty}^{+\infty} f(y) dy \right) \\ &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-(x^2+y^2)/2} dx dy. \end{aligned}$$

Mudemos para coordenadas polares (θ, ρ) , onde $x = \rho \cos(\theta)$ e $y = \rho \sin(\theta)$, com $\theta \in (0, 2\pi)$ e $\rho \in (0, +\infty)$. Desta maneira $\rho^2 = x^2 + y^2$ e $dx dy = \rho d\rho d\theta$. Então

$$\begin{aligned} \frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-(x^2+y^2)/2} dx dy &= \frac{1}{2\pi} \int_0^{+\infty} \int_0^{2\pi} \rho e^{-\rho^2/2} d\theta d\rho \\ &= \int_0^{+\infty} \rho e^{-\rho^2/2} d\rho, \end{aligned}$$

fazendo a transformação $u = \rho^2/2$, obtemos $\int_0^{+\infty} e^{-u} du = 1$, como queríamos demonstrar.

Mais geral ainda seria considerar a variável aleatória com distribuição $X \sim N(\mu, \sigma^2)$. Estas quantidades μ e σ^2 são chamados de parâmetros da distribuição, definem a família de distribuições Normal e determinam a forma específica da função de densidade quando especificamos valores particulares. Assim, a família de densidades $Normal(\mu, \sigma^2)$ é da forma

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad (3.44)$$

onde $\mu \in \mathbb{R}$ e $\sigma^2 > 0$.

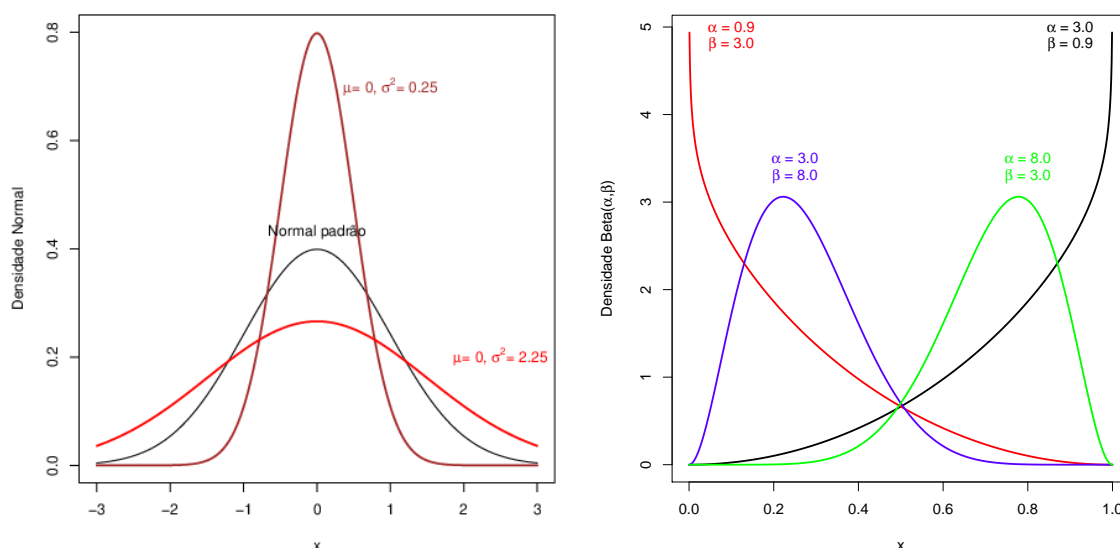


Figura 3.12: À esquerda mostramos diferentes formas que pode assumir a função de densidade Normal, sempre com valor do parâmetro $\mu = 0$ e modificando o valor do parâmetro σ^2 , podendo este ser $\sigma^2 = 1$ no caso Normal padrão, nas outras curvas $\sigma^2 = 0,25$ e $\sigma^2 = 2,25$. Percebe que quanto menor o valor do parâmetro σ^2 mais concentrada no zero apresenta-se a densidade Normal. À direita mostramos a forma correspondente assumida pela distribuição Beta com valores dos parâmetros mostrados.

Nos exemplos a seguir mostramos duas distribuições derivadas da densidade Normal.

Exemplo 3.34 (*Distribuição Normal inversa*)

A distribuição Normal inversa, é uma família de distribuições de dois parâmetros $\mu > 0$ e $\sigma > 0$, com função de densidade

$$f(x) = \sqrt{\frac{\sigma}{2\pi x^3}} \exp\left(-\sigma \frac{(x - \mu)^2}{2\mu^2 x}\right), \quad x > 0.$$

Esta distribuição também conhecida como distribuição de Wald, quando σ tende ao infinito, a distribuição Normal inversa torna-se mais parecida com uma distribuição Normal. A distribuição Normal inversa tem várias propriedades análogas à distribuição Normal.

Exemplo 3.35 (*Distribuição log-Normal*)

Uma distribuição log-Normal é uma distribuição de probabilidade contínua, cujo logaritmo é normalmente distribuído. Assim, se a variável aleatória X tiver distribuição log-Normal, então $Y = \ln(X)$ tem uma distribuição Normal. Da mesma forma, se Y tem uma distribuição Normal, então a função exponencial de Y , $X = \exp(Y)$, tem uma distribuição log-Normal. Então, a função de densidade log - Normal(μ, σ^2) é da forma

$$f(x) = \frac{1}{x} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln(x) - \mu)^2}{2\sigma^2}\right), \quad x > 0.$$

3.2.3 Variáveis aleatórias mistas

Esta aqui definida é uma outra das diversas variáveis aleatórias contínuas. Quando misturamos uma variável discreta com uma outra contínua o resultado, tomando os devidos cuidados, é uma variável mista: em alguns pontos discreta e contínua em intervalos. Uma situação diferente foi considerada no Exemplo 3.36. Também podemos misturar duas ou mais variáveis contínuas.

Definição 3.28

Seja X uma variável aleatória com função de densidade na forma

$$f(x) = \sum_{k=1}^m \alpha_k f_k(x), \quad x \in \mathbb{R}, \quad (3.45)$$

onde cada f_k é uma função de densidade ou de probabilidade, cada α_k é positivo e $\sum_{k=1}^m \alpha_k = 1$. A variável aleatória X com densidade como em (3.45) é chamada de mista.

Estas variáveis aleatórias podem admitir alguns pontos de salto, como as discretas, porém os restantes valores são números reais.

Exemplo 3.36

Uma situação de distribuições contínuas acontece quando a variável aleatória tem por distribuição normal inversa com ajuste no zero, definida como

$$f(x; \mu, \sigma, \nu) = \nu \quad \text{se } x = 0$$

e

$$f(x; \mu, \sigma, \nu) = (1 - \nu) \sqrt{\frac{\sigma}{2\pi x^3}} \exp\left(-\sigma \frac{(x - \mu)^2}{2\mu^2 x}\right)$$

caso contrário, isto para $x \in (0, \infty)$, $\mu > 0$, $\sigma > 0$ e $0 < \nu < 1$. A forma funcional esta apresentada na Figura 3.13.

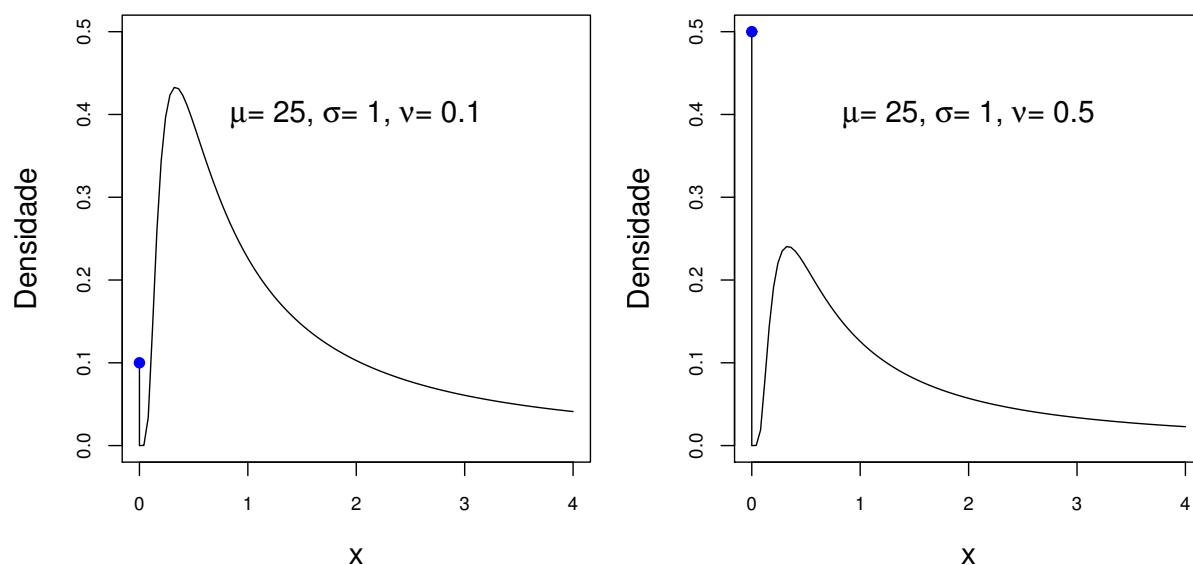


Figura 3.13: Função de densidade contínua normal inversa com ajuste no zero.

Segundo a definição acima podemos ter misturas de três tipos: misturas de densidade propriamente, misturas de funções de probabilidade e misturas de densidades e probabilidades. Observemos que a variável considerada no Exemplo 3.36 é também chamada de mista, porém de natureza diferente às definidas. As variáveis mistas mais utilizadas em aplicações são aquelas aqui definidas. Na Figura 3.14 apresentamos diferentes possíveis formas de funções de densidade mista, sempre no caso particular de duas densidades ou probabilidades. No quadro superior esquerdo dessa figura mostramos, na curva em preto, a misturas de duas distribuições normais uma de parâmetros $\mu = -3$ e $\sigma^2 = 1$ e outra com $\mu = 1$ e $\sigma^2 = 0.64$, por isso no denominador aparecer a quantidade $2\sigma^2 = 2 \times 0.64 = 1.28$; no caso da linha em vermelho temos a mistura de também duas densidade normais de parâmetros $\mu = -3$ e $\sigma^2 = 0.36$ no primeiro caso e $\mu = 1$ e $\sigma^2 = 3.24$, em ambas situações $\alpha_1 = 0.3$ e $\alpha_2 = 1 - \alpha_1 = 0.7$.

Ainda acerca da Figura 3.14, no gráfico acima à direita temos a mistura de uma densidade normal, de parâmetros $\mu = 2$ e $\sigma^2 = 0.16$, e uma densidade exponencial de parâmetro $\theta = 0.6$ no caso da linha em vermelho e a mistura de uma densidade exponencial de parâmetro $\theta = 0.4$ e uma densidade normal de parâmetros $\mu = 4$ e $\sigma^2 = 0.49$, no caso da linha preta; em ambas situações os parâmetros α_1 e α_2 da combinação linear foram escolhidos como 0.4 e 0.6. Nos gráficos abaixo temos dois gráficos com misturas de funções de probabilidade; a esquerda a mistura é de duas binomiais, uma com $n = 6$ e $p = 0.6$ a outra com $n = 8$ e $p = 0.8$. Por último, a mistura foi de duas probabilidades Poisson de parâmetros $\theta = 2$ e $\theta = 10$, na linha preta e $\theta = 6$ e $\theta = 15$, no caso da linha em vermelho.

Evidentemente, os gráficos mostrados na Figura 3.14, são somente alguns poucos exemplos das diversas possibilidades de misturas. Devemos considerar que podemos misturar também funções de densidade com funções de probabilidade. Mais ainda, os exemplos aqui mostrados referem-se sempre a mistura de duas funções de densidades ou de duas funções de probabilidades quando na verdade, segundo a definição, a mistura pode ser de uma quantidade finita de densidades ou probabilidades e continuarmos tendo misturas.

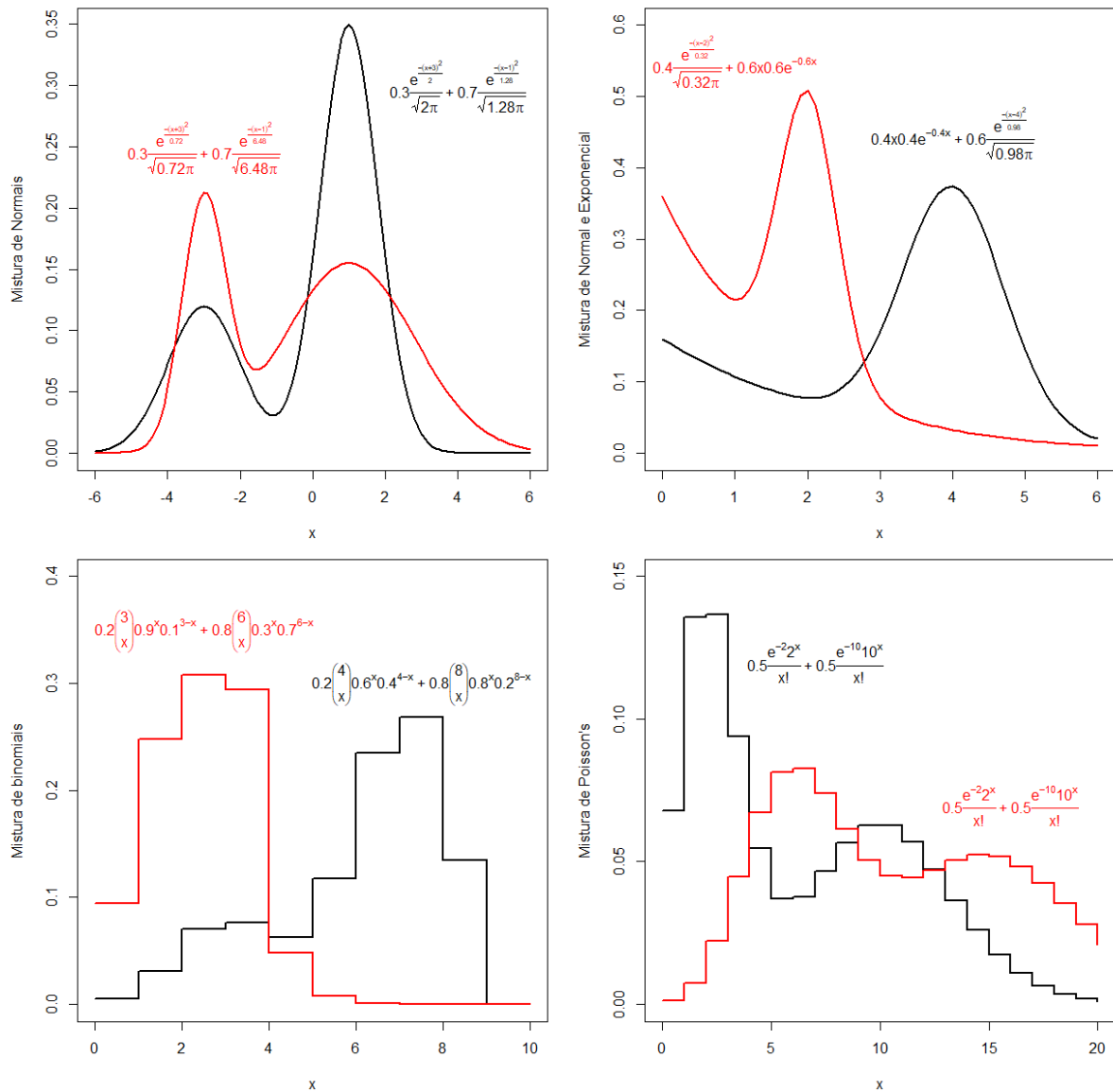


Figura 3.14: Diferentes formas de funções de densidade correspondentes à variáveis aleatórias mistas nos dois quadros superiores, à esquerda misturamos duas densidades Normais enquanto a direita misturamos uma densidade Normal e uma densidade Exponencial. Abaixo temos misturas de funções de probabilidade pertencentes às mesmas famílias: à esquerda temos uma mistura de binomiais e a direita misturas de Poisson. Em cada situação consideramos misturas de duas densidades ou probabilidades, por isso, sempre aparece um coeficiente à frente de cada expressão e perceba que sempre somam um, sendo estes coeficientes escolhidos diferentes a propósito em cada diferente quadro mas não nas combinações diferentes. Um outro detalhe é que decidimos mostrar a mistura de probabilidades utilizando função em degraus e não simplesmente pontos, foi decidido assim para aparentar continuidade nas curvas e melhor diferenciá-las. Percebe-se a grande quantidade de misturas possíveis com somente as poucas densidades e probabilidades conhecidas até o momento. Devemos mencionar também que as funções de densidade ou de probabilidade mistas satisfazem todas as propriedades correspondentes, ou seja, somam ou integram 1 e são sempre não negativas.

O detalhe continua sendo a aplicabilidade destes modelos. Também associamos a estas variáveis funções de distribuição, porém geralmente são de difícil obtenção analítica sendo utilizadas, quando necessário, de maneira numérica.

3.3 Funções de variáveis aleatórias

Sejam X uma variável aleatória de qualquer tipo com uma distribuição conhecida F_X e g uma função real. Queremos identificar a distribuição de $g(X)$, uma função de X , desde que esta função, ou seja, desde que $g(X)$ seja também uma variável aleatória. Perguntamos então: qualquer função de uma variável aleatória é também uma variável aleatória? caso $g(X)$ seja uma variável aleatória, sempre será possível encontrar a distribuição transformada?

Definição 3.29

Dizemos que Y é uma função da variável aleatória X se

$$Y = g(X), \quad (3.46)$$

para alguma função real g .

Logicamente, qualquer seja a função real g , Y é uma transformação da variável aleatória original. Vejamos agora um exemplo no qual fazemos modificações numa variável discreta.

Exemplo 3.37

Consideremos $X \sim \text{Bernoulli}(p)$, ou seja, X tem função de probabilidade Bernoulli de parâmetro p , o qual significa que

$$P(X = x_1) = p, \quad P(X = x_2) = 1 - p,$$

sendo $x_1, x_2 \in \mathbb{R}$ diferentes e $0 < p < 1$. Seja agora $g(x) = x + a$, para $\forall x \in \mathbb{R}$ e " a " um número real fixo. Nestas condições, $Y = g(X)$ é tal que $X = g^{-1}(Y)$, sendo $y_1 = x_1 + a$ e $y_2 = x_2 + a$, logo $x_1 = y_1 - a$ e $x_2 = y_2 - a$.

Portanto,

$$P(Y = y_1) = P(X + a = y_1) = P(X = y_1 - a) = P(X = x_1) = p$$

e

$$P(Y = y_2) = P(X + a = y_2) = P(X = y_2 - a) = P(X = x_2) = 1 - p.$$

Concluimos então que Y é também uma variável aleatória Bernoulli(p).

Definimos Y como função de uma variável aleatória mas não dizemos se esta é uma nova variável aleatória, para isso a função g não pode ser qualquer uma, deve ser uma função Borel mensurável. Às funções Borel mensuráveis são dedicadas os próximos conceitos.

Definição 3.30

Seja $g : \mathbb{R} \rightarrow \mathbb{R}$ e seja $A \subseteq \mathbb{R}$. Entendemos por imagem de A através da função g os elementos y de \mathbb{R} tais que $y = g(x)$, para algum $x \in A$. Denotamos esta imagem como $g(A)$. Se $B \subseteq \mathbb{R}$, definimos a imagem inversa de B por g ao conjunto dos $y \in \mathbb{R}$ para os quais $g^{-1}(y) \in B$ e escrevemos

$$g^{-1}(B) = \{x \in \mathbb{R} : g(x) \in B\}.$$

Aqui definimos imagem inversa de um conjunto A por uma função real g , de maneira similar à Definição 3.1. No caso do Exemplo 3.37 a função real é a translação $g(x) = x + a$, sendo "a" um número real fixo. Propriedades similares ao Teorema 3.1 é apresentado a seguir para o caso de imagem inversa de uma função real.

Teorema 3.32

Seja $g : \mathbb{R} \rightarrow \mathbb{R}$ uma função real e B, B_1, B_2, \dots subconjuntos em \mathbb{R} . Então

$$a) \ g^{-1}(B^c) = \left(g^{-1}(B)\right)^c,$$

$$b) \ g^{-1}\left(\bigcup_{k=1}^{\infty} B_k\right) = \bigcup_{k=1}^{\infty} g^{-1}(B_k),$$

$$c) \ g^{-1}\left(\bigcap_{k=1}^{\infty} B_k\right) = \bigcap_{k=1}^{\infty} g^{-1}(B_k).$$

Demonstração: Exercício. ■

Agora estamos em condições para definir quando diremos que uma função é Borel mensurável.

Definição 3.31 (Função Borel mensurável)

A função $g : \mathbb{R} \rightarrow \mathbb{R}$ é chamada Borel mensurável se

$$g^{-1}(B) \in \mathcal{B}(\mathbb{R}), \quad \forall B \in \mathcal{B}(\mathbb{R}).$$

Nesta definição dizemos que g é uma função Borel mensurável se transforma elementos de da sigma álgebra $\mathcal{B}(\mathbb{R})$ em elementos de $\mathcal{B}(\mathbb{R})$. É claro que uma função real g da variável

real x é Borel mensurável se o conjunto

$$\{x : -\infty < g(x) \leq y\} \in \mathcal{B}(\mathbb{R}),$$

para qualquer número real y . Vejamos se isso acontece no Exemplo 3.37. Nessa situação

$$\{x : -\infty < g(x) \leq y\} = \{x : -\infty < x + a \leq y\} = \{x : -\infty < x \leq y - a\} \in \mathcal{B}(\mathbb{R}).$$

Teorema 3.33

Toda função $g : \mathbb{R} \rightarrow \mathbb{R}$ contínua é Borel mensurável.

Demonstração : Exercício. ■

Lembremos alguns detalhes acerca de funções diferenciáveis e contínuas. Primeiro, o fato da função g ser contínua não implica seja diferenciável e, por último, devemos lembrar ainda existem funções que não são diferenciáveis em nenhum ponto. Tudo isto porque se g é diferenciável então é contínua e, pelo teorema acima, g é Borel mensurável (Spivak, 1994).

Teorema 3.34

Seja X uma variável aleatória e g uma função contínua. Então $g(X)$ é uma variável aleatória.

Demonstração :

$$\{g(X) \leq y\} = \{X \in g^{-1}(-\infty, y]\},$$

e, dado que g é contínua, $g^{-1}(-\infty, y] \in \mathcal{B}(\mathbb{R})$. Logo, $\{g(X) \leq y\} \in \mathcal{B}(\mathbb{R})$. ■

Exemplo 3.38

Seja X o número de falantes ativos em um grupo de n falantes. Seja p a probabilidade de um falante estar ativo. Sabemos que $X \sim \text{Binomial}(n, p)$. Suponha que um sistema de transmissão de voz possa transmitir até m sinais de voz de cada vez e que, quando X excede m , os sinais selecionados aleatoriamente de $X - m$ são descartados.

Seja Y o número de sinais descartados, então

$$Y = (X - m)^+.$$

A nova variável aleatória Y assume valores no conjunto $\{0, 1, 2, \dots, n - m\}$. Y será igual a zero sempre que X for menor ou igual a m e Y será igual quando X for igual a $k > 0$, quando X for igual a $m + k$, ou seja,

$$Y = \begin{cases} 0, & \text{se } X \leq m \\ X, & \text{se } X \in \{m + 1, \dots, n\} \end{cases}.$$

Assim sendo

$$P(Y = 0) = P(X \in \{0, 1, 2, \dots, m\}) = \sum_{x=0}^m P(X = x)$$

e

$$P(Y = k) = P(X = m + k), \quad 0 < k \leq n - m.$$

Teorema 3.35

Dada a variável aleatória X com função de distribuição conhecida, a distribuição da variável aleatória $Y = g(X)$, sendo g uma função Borel mensurável, é conhecida.

Demonstração: Temos que

$$P(Y \leq y) = P(X \in g^{-1}(-\infty, y]). \quad (3.47) \quad \blacksquare$$

No que segue, sempre assumiremos que as funções em consideração são Borel mensuráveis. Na verdade, a restrição de um valor único para o inverso de g não é necessária. Caso g tenha um número finito ou inclusive um número enumerável de inversas para cada valor y , da propriedade de σ -aditividade da função de probabilidade P temos que

$$\begin{aligned} P(Y = y) &= P(g(X) = y) = P\left(\bigcup_a [X = a, g(a) = y]\right) \\ &= \sum_a P(X = a, g(a) = y). \end{aligned}$$

Exemplo 3.39

Seja X uma variável aleatória com função de probabilidade

$$P(X = -2) = \frac{1}{5}, \quad P(X = -1) = \frac{1}{6}, \quad P(X = 0) = \frac{1}{5}, \quad P(X = 1) = \frac{1}{15}$$

e

$$P(X = 2) = \frac{11}{30}.$$

Seja $Y = X^2$. Então $A = \{-2, -1, 0, 1, 2\}$ e $B = \{0, 1, 4\}$. Temos

$$P(Y = y) = \begin{cases} \frac{1}{5} = \frac{6}{30}, & \text{se } y = 0, \\ \frac{1}{6} + \frac{1}{15} = \frac{7}{30}, & \text{se } y = 1, \\ \frac{1}{5} + \frac{11}{30} = \frac{17}{30}, & \text{se } y = 4. \end{cases}$$

Exemplo 3.40

Seja X uma variável aleatória com função de distribuição F_X . Então

- (a) $|X|$,
- (b) $aX + b$, onde $a \neq 0$ e b são constantes,
- (c) X^n , onde $n \geq 0$ é um inteiro e
- (d) $|X|^\alpha$, sendo $\alpha > 0$

são todas variáveis aleatórias. Definamos

$$X^+ = \begin{cases} X, & \text{caso } X \geq 0, \\ 0, & \text{caso } X < 0, \end{cases}$$

e

$$X^- = \begin{cases} X, & \text{caso } X \leq 0, \\ 0, & \text{caso } X > 0. \end{cases}$$

Então, X^+ e X^- são também variáveis aleatórias. Vejamos cada situação:

- (a) Temos que, para $y > 0$,

$$\begin{aligned} P(|X| \leq y) &= P(-y \leq X \leq y) = P(X \leq y) - P(X < -y) \\ &= F_X(y) - F_X(-y) + P(X = -y). \end{aligned}$$

$$(b) P(aX + b \leq y) = P(aX \leq y - b) = \begin{cases} P\left(X \leq \frac{y-b}{a}\right), & \text{se } a > 0 \\ P\left(X \geq \frac{y-b}{a}\right), & \text{se } a < 0 \end{cases}$$

- (c) Dado que n é um inteiro,

$$P(X^n \leq y) = P(X \leq \sqrt[n]{y}).$$

- (d) $|X| = X^+ - X^-$ e

$$P(X^+ \leq y) = \begin{cases} 0, & \text{se } y < 0, \\ P(X \leq 0), & \text{se } y = 0, \\ P(X < 0) + P(0 \leq X \leq y), & \text{se } y > 0. \end{cases}$$

Similarmente

$$P(X^- \leq y) = \begin{cases} 1, & \text{se } y \geq 0, \\ P(X \leq y), & \text{se } y < 0, \end{cases}$$

e

$$P(|X|^\alpha \leq y) = P(|X| \leq y^{1/\alpha}).$$

Exemplo 3.41

Seja $X \sim \text{Poisson}(\theta)$ o qual implica que a função de probabilidade da variável aleatória é

$$P(X = x) = \begin{cases} \frac{\theta^x}{x!} e^{-\theta}, & x = 0, 1, 2, \dots, \quad \theta > 0 \\ 0, & \text{caso contrário} \end{cases}$$

Agora seja $Y = X^2 + 3$, logo $g(x) = x^2 + 3$ é uma função contínua e $Y = g(X)$ é uma variável aleatória. Então, $y = x^2 + 3$ é uma transformação entre os conjuntos

$$A = \{0, 1, 2, 3, 4, 5, \dots\} \quad \text{e} \quad B = \{3, 4, 7, 12, 19, 28, \dots\}.$$

A inversa é $x = \sqrt{y - 3}$ e como não existem números negativos em A , escolhemos a raiz quadrada positiva de $y - 3$. Temos então,

$$P(Y = y) = P(X = \sqrt{y - 3}) = \frac{e^{-\theta} \theta^{\sqrt{y-3}}}{\sqrt{y-3}!}, \quad y \in B,$$

e $P(Y = y) = 0$ caso contrário.

O caso em que X é uma variável aleatória do tipo contínuo não é tão simples. Observamos que, se X é uma variável aleatória de tipo contínuo e g é alguma função Borel mensurável tal que $Y = g(X)$, pode acontecer de Y não ser uma variável aleatória do tipo contínuo.

Exemplo 3.42

Seja $X \sim \text{Uniforme}(-1, 1)$, isto é, a função de densidade de X é da forma

$$f(x) = \begin{cases} \frac{1}{2}, & \text{se } -1 < x < 1 \\ 0, & \text{caso contrário.} \end{cases}$$

Seja $Y = X^+$. Então, pelo Exemplo 3.40, temos que

$$P(Y \leq y) = \begin{cases} 0, & \text{se } y < 0 \\ \frac{1}{2}, & \text{se } y = 0 \\ \frac{1}{2} + \frac{1}{2}y, & \text{se } 0 < y < 1 \\ 1, & \text{se, } y > 1 \end{cases}.$$

Vemos que a função de distribuição de Y tem um salto em $y = 0$ e que Y não é nem discreta nem contínuo. Note que, nesta situação nossa escolha foi

$$g(x) = \begin{cases} 0, & \text{se } x \leq 0 \\ x, & \text{se } x \geq 0 \end{cases},$$

a qual é uma função contínua.

No Exemplo 3.42 mostramos que precisamos de algumas condições em g para garantir que $g(X)$ seja também uma variável aleatória do tipo contínuo, sempre que X for contínuo. Este será o caso quando g é uma função monotônica contínua. Uma condição suficiente é dada no seguinte teorema conhecido como transformação do Jacobiano⁶.

Teorema 3.36 (*Jacobiano*)

Seja X uma variável aleatória contínua com função de densidade f . Seja $y = g(x)$ uma função derivável para todo x e seja $g'(x) > 0$ para todo x ou $g'(x) < 0$ para todo x . Então $Y = g(X)$ é também uma variável aleatória contínua com função de densidade

$$h(y) = \begin{cases} f(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right| & \text{se } \alpha < y < \beta, \\ 0 & \text{caso contrário,} \end{cases} \quad (3.48)$$

onde $\alpha = \min\{g(-\infty), g(+\infty)\}$ e $\beta = \max\{g(-\infty), g(+\infty)\}$.

Demonstração: Se g é uma função diferenciável para todo x e $g'(x) > 0$ para todo x , então g é contínua e estritamente crescente. Os limites α e β existem, podendo ser infinitos, e a função inversa $x = g^{-1}(y)$ existe, sendo estritamente crescente e diferenciável. A função de distribuição de Y , para $\alpha < y < \beta$ é dada por

$$P(Y \leq y) = P(X \leq g^{-1}(y)).$$

A função de densidade de g é obtida por diferenciação. Temos que

$$h(y) = \frac{d}{dy} P(Y \leq y) = f(g^{-1}(y)) \frac{d}{dy} g^{-1}(y).$$

Similarmente, se $g' < 0$ então g é estritamente decrescente e temos

$$P(Y \leq y) = P(X \geq g^{-1}(y)) = 1 - P(X \leq g^{-1}(y)),$$

lembrando que X é uma variável aleatória do tipo contínuo, de modo que

$$h(y) = -f(g^{-1}(y)) \frac{d}{dy} g^{-1}(y).$$

Uma vez que g e g^{-1} são ambas estritamente decrescentes,

$$\frac{d}{dy} g^{-1}(y) < 0$$

e a relação em (3.48) é válida. ■

⁶Carl Jacob Jacob Jacobi (1804 - 1851) foi um matemático alemão, que fez contribuições fundamentais para funções elípticas, dinâmica, equações diferenciais e teoria dos números. Jacobi foi o primeiro matemático judeu a ser nomeado professor em uma universidade alemã.

Observemos que

$$\frac{d}{dy}g^{-1}(y) = \frac{1}{\frac{d}{dx}g(x)} \bigg|_{x=g^{-1}(y)},$$

de maneira que (3.48) pode ser escrita como

$$h(y) = \frac{f(x)}{\frac{d}{dx}g(x)} \bigg|_{x=g^{-1}(y)}, \quad \alpha < y < \beta. \quad (3.49)$$

A chave para encontrar a distribuição induzida de $Y = g(X)$ da distribuição de X é (3.47). Se as condições do Teorema 3.36 são satisfeitas, somos capazes de identificar o conjunto

$$\{X \in g^{-1}(-\infty, y]\}$$

como

$$\{X \leq g^{-1}(y)\} \quad \text{ou} \quad \{X \geq g^{-1}(y)\},$$

de acordo com g , se está crescendo ou decrescendo. Em situações práticas, o Teorema 3.36 é bastante útil, mas sempre que as condições sejam violadas deve-se retornar a (3.47) para calcular a distribuição induzida. Caso a função de densidade f de X for zero fora do intervalo $[a, b]$ de comprimento finito basta supor que g é diferenciável em (a, b) e $g'(x) > 0$ ou $g'(x) < 0$ ao longo do intervalo. Então, tomamos

$$\alpha = \min(g(a), g(b)) \quad \text{e} \quad \beta = \max(g(a), g(b))$$

no Teorema 3.36.

Exemplo 3.43

Seja $X \sim \text{Exponencial}(2)$, ou seja, X é uma variável aleatória com função de densidade

$$f(x) = \begin{cases} 2e^{-2x} & \text{se } x > 0 \\ 0 & \text{caso contrário.} \end{cases}$$

Consideremos uma nova variável aleatória $Y = X^3$. Significa que $y = g(x) = x^3$ e então $X = \sqrt[3]{Y}$. A função inversa é $g^{-1}(y) = \sqrt[3]{y}$ do qual obtemos que

$$h(y) = \begin{cases} f(\sqrt[3]{y}) \left| \frac{d}{dy} \sqrt[3]{y} \right|, & \text{se } y > 0 \\ 0 & \text{caso contrário} \end{cases}$$

Observe que $g'(x)3x^2 > 0$, quando $x > 0$. Portanto, a função de densidade de Y é

$$h(y) = \begin{cases} \frac{2}{3}y^{-2/3}e^{-2\sqrt[3]{y}}, & \text{se } y > 0 \\ 0 & \text{caso contrário} \end{cases}$$

Vamos encontrar agora a função de densidade quando a transformação é $Y = X^\alpha$ e o suporte da função de densidade de X seja a reta positiva. Nestas situações consideramos uma grande classe de densidades. Então, seja X uma variável aleatória com densidade $f(x)$ quando $x \geq 0$, zero noutras situações e $\alpha > 0$. Do qual temos que

$$P(Y \leq y) = P(X^\alpha \leq y) = \begin{cases} P(X \leq y^{1/\alpha}), & \text{se } y \geq 0 \\ 0, & \text{se } y < 0 \end{cases}.$$

A função de densidade de Y é dada por

$$h(y) = f(y^{1/\alpha}) \left| \frac{d}{dy} y^{1/\alpha} \right| = \begin{cases} \frac{1}{\alpha} y^{1/\alpha-1} f(y^{1/\alpha}) & \text{se } y > 0 \\ 0 & \text{caso contrário,} \end{cases} \quad (3.50)$$

Uma caso particular da situação descrita em (3.50) acontece quando $X \sim \text{Exponencial}(\theta)$, ou seja, quando a função de densidade de X é

$$f(x) = \theta e^{-\theta x}, \quad \text{para } x \geq 0,$$

e zero caso contrário. Dado que $Y = X^\alpha$, a função de densidade de Y é

$$h(y) = \frac{1}{\alpha} y^{1/\alpha-1} \theta e^{-\theta y^{1/\alpha}}$$

sempre que $y \geq 0$ e $\alpha > 0$.

Exemplo 3.44

Consideremos agora uma variável aleatória X com função de densidade

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad -\infty < x < \infty,$$

conhecida como densidade Normal padrão. Seja $Y = X^2$. Nesta situação $g'(x) = 2x$, a qual é positiva se $x > 0$ e negativa caso $x < 0$. Portanto, as condições do Teorema 3.36 não se satisfazem. Mas, para $y > 0$

$$P(Y \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y}) = F(\sqrt{y}) - F(-\sqrt{y}),$$

onde F é a função de distribuição de X . Então, a função de densidade de Y é dada por

$$h(y) = \begin{cases} \frac{1}{2\sqrt{y}} [f(\sqrt{y}) + f(-\sqrt{y})], & \text{se } y > 0 \\ 0, & \text{se } y \leq 0 \end{cases}.$$

Desta maneira,

$$h(y) = \begin{cases} \frac{1}{\sqrt{2\pi y}} e^{-y/2}, & \text{se } y > 0 \\ 0, & \text{se } y \leq 0 \end{cases}.$$

Exemplo 3.45 (*Distribuição arcoseno*)

Seja $\Theta \sim \text{Uniforme}(-\pi/2, \pi/2)$ e a transformação é $Y = \sin(\Theta)$. Note que, em geral, $y = \sin(\theta)$ não é uma transformação monotônica, mas sob a restrição $-\pi/2 < \theta < \pi/2$, essa transformação está de fato aumentando monotonicamente. Observe também que com esta transformação a variável aleatória resultante Y , deve assumir valores no intervalo $(-1, 1)$. Portanto, qualquer que seja a função de densidade obtida para Y , deve-se entender que esta função é zero fora do intervalo $(-1, 1)$.

Aplicando os resultados em (3.48) proporciona

$$h(y) = \left. \frac{f(\theta)}{\cos(\theta)} \right|_{\theta=\sin^{-1}(y)} = \frac{1}{\pi \cos(\sin^{-1}(y))} = \frac{1}{\pi \sqrt{1-y^2}}, \quad -1 < y < 1.$$

Esta é conhecida como a distribuição arcoseno.

Exemplo 3.46

A função de densidade da variável aleatória X é

$$f(x) = \begin{cases} \frac{2x}{\pi^2}, & \text{se } 0 < x < \pi \\ 0, & \text{caso contrário} \end{cases}.$$

Seja $Y = \sin(X)$. Nesta situação $g'(x) = \cos(x) > 0$ para $x \in (0, \pi/2)$ e negativa quando $x \in (\pi/2, \pi)$ de maneira que as condições do Teorema 3.36 não são satisfeitas. Para calcularmos a função de densidade de Y voltamos nosso olhar à (3.47) e vemos que a função de distribuição de Y é dada por

$$\begin{aligned} P(Y \leq y) &= P(\sin(X) \leq y), \quad 0 < y < 1, \\ &= P(\{0 \leq X \leq x_1\} \cup \{x_2 \leq X \leq \pi\}), \end{aligned}$$

onde $x_1 = \sin^{-1}(y)$ e $x_2 = \pi - \sin^{-1}(y)$. Obtemos que

$$P(Y \leq y) = \int_0^{x_1} f(x) dx + \int_{x_2}^{\pi} f(x) dx = \left(\frac{x_1}{\pi}\right)^2 + 1 - \left(\frac{x_2}{\pi}\right)^2,$$

e a função de densidade de Y é dada por

$$\begin{aligned} h(y) &= \frac{d}{dy} \left\{ \left(\frac{\sin^{-1}(y)}{\pi}\right)^2 + \left[1 - \left(\frac{\pi - \sin^{-1}(y)}{\pi}\right)^2\right] \right\} \\ &= \begin{cases} \frac{2}{\pi \sqrt{1-y^2}}, & 0 < y < 1, \\ 0, & \text{caso contrário} \end{cases}. \end{aligned}$$

No Exemplo 3.44 e no Exemplo 3.46 a função $y = g(x)$ pode ser escrita como soma de duas funções monótonas. Aplicamos então o Teorema 3.36 a cada uma dessas duas funções monótonas. Esses dois exemplos são situações especiais do seguinte resultado.

Teorema 3.37 (Generalização do Jacobiano)

Seja X uma variável aleatória contínua com função de densidade f e $y = g(x)$ uma função diferenciável para todo x . Consideremos que $g'(x)$ é contínua e diferente de zero em quase tudo, exceto um número finito de valores de x . Então, para todo número real y

(a) existe um número inteiro $n = n(y)$ e números reais $x_1(y), x_2(y), \dots, x_n(y)$ tais que

$$g(x_k(y)) = y, \quad g'(x_k(y)) \neq 0, \quad k = 1, 2, \dots, n(y),$$

ou

(b) não existe x tal que $g(x) = y$, $g'(x) \neq 0$, caso em que escrevemos $n(y) = 0$.

Então, Y é uma variável aleatória contínua com função de densidade dada por

$$h(y) = \begin{cases} \sum_{k=1}^n f(x_k(y)) |g'(x_k(y))|^{-1} & \text{se } n > 0 \\ 0 & \text{se } n = 0 \end{cases}.$$

Demonstração: Exercício. ■

Exemplo 3.47

Seja X uma variável aleatória absolutamente contínua com função de densidade f e função de distribuição F_X . Queremos fazer

$$Y = X^2$$

e encontrar a função de densidade de Y e sua função de distribuição

Primeiro devemos observar que $Y = g(X)$, onde $g(x) = x^2$ é uma função derivável com derivada $g'(x) = 2x$, a qual satisfaz que $g'(x) < 0$, se $x < 0$ e $g'(x) > 0$ se $x > 0$. Significa que $n(y) = 2$, sendo que $x_1(y) = \sqrt{y}$ e $x_2(y) = -\sqrt{y}$, para $y > 0$, com $g'(x_1(y)) < 0$ e $g'(x_2(y)) > 0$. Logo, estão satisfeitas as exigências do Teorema 3.37. Assim,

$$h(y) = \frac{1}{2\sqrt{y}} f(\sqrt{y}) + \frac{1}{2\sqrt{y}} f(-\sqrt{y})$$

é a função de densidade de Y . A função de distribuição F_Y é então

$$F_Y(y) = \begin{cases} 0, & \text{se } y \leq 0 \\ F_X(\sqrt{y}) - F_X(-\sqrt{y}), & \text{se } y > 0 \end{cases}.$$

Uma outra forma sempre possível é observando, na Figura 3.15, que o evento $\{Y \leq y\}$

ocorre quando $\{X^2 \leq y\}$ ou, equivalentemente, quando $\{-\sqrt{y} \leq X \leq \sqrt{y}\}$ para y não negativo. O evento é nulo quando y é negativo. Assim obtemos as expressões de $h(y)$ e $F_Y(y)$ no exemplo anterior.

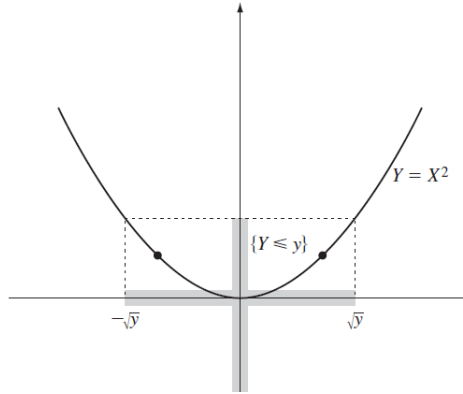


Figura 3.15: O evento equivalente para o evento $\{Y \leq y\}$ é o evento $\{-\sqrt{y} \leq X \leq \sqrt{y}\}$, se $y \geq 0$.

Exemplo 3.48

Seja X uma variável aleatória positiva com função de densidade f , ou seja, afirmar que X é positiva significa que

$$P(X < 0) = 0 \quad \text{e} \quad P(X \geq 0) = 1.$$

Seja $Y = |X|$. Nesta situação $n(y) = 2$, $x_1(y) = y$, $x_2(y) = -y$ para $y > 0$ e

$$h(y) = \begin{cases} f(y) + f(-y), & y > 0 \\ 0, & y \leq 0 \end{cases}.$$

(i) Se $f(x) = \begin{cases} \frac{1}{2}, & \text{se } -1 \leq x \leq 1 \\ 0, & \text{caso contrário} \end{cases}$, então

$$h(y) = \begin{cases} 1, & 0 \leq y \leq 1 \\ 0, & \text{caso contrário} \end{cases}.$$

(ii) Se $f(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$, quando $-\infty < x < \infty$, temos que

$$h(y) = \begin{cases} \frac{2}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}, & y > 0 \\ 0, & \text{caso contrário} \end{cases}.$$

Exemplo 3.49

Consideremos novamente uma variável aleatória contínua positiva com função de densidade f . Seja $Y = X^{2m}$ onde m é um inteiro positivo. Nesta situação

$$g(x) = x^{2m}$$

e

$$g'(x) = \begin{cases} 2mx^{2m-1} > 0, & \text{para } x > 0 \\ g'(x) < 0, & \text{para } x < 0 \end{cases}.$$

Escrevendo $n = 2m$, vemos que, para qualquer $y > 0$, $n(y) = 2$,

$$x_1(y) = -y^{1/n} \quad \text{e} \quad x_2(y) = y^{1/n}.$$

Segue que

$$\begin{aligned} h(y) &= f[x_1(y)] \frac{1}{ny^{1-1/n}} + f[x_2(y)] \frac{1}{ny^{1-1/n}} \\ &= \begin{cases} \frac{1}{ny^{1-1/n}} [f(y^{1/n}) + f(-y^{1/n})], & y > 0, \\ 0, & y \leq 0 \end{cases}. \end{aligned}$$

Neste mesmo exemplo, caso a função de densidade seja como àquela no Exemplo 3.44, temos que

$$h(y) = \begin{cases} \frac{2}{\sqrt{2\pi}} \frac{1}{ny^{1-1/n}} \exp \left\{ -\frac{y^{2/n}}{2} \right\}, & y > 0 \\ 0, & y \leq 0 \end{cases}.$$

A fórmula básica em (3.47) e a σ -aditividade das probabilidades nos permite calcular a distribuição de $Y = g(X)$ em alguns casos, mesmo que g tenha um número enumerável de inversas. Seja $A \subseteq \mathbb{R}$ e g uma função que transforma A em $B \subseteq \mathbb{R}$. Suponhamos que A possa ser representado como a união enumerável de conjuntos disjuntos A_k , $k = 1, 2, \dots$, tais que g define uma transformação um-a-um de cada A_k em B . Então, a função de distribuição de Y é definida por

$$\begin{aligned} P(Y \leq y) &= P(X \in g^{-1}(-\infty, y]) \\ &= P\left(X \in \bigcup_{k=1}^{\infty} [g^{-1}(-\infty, y] \cap A_k\right) \\ &= \sum_{k=1}^{\infty} P(A_k \cap \{g^{-1}(-\infty, y]\}). \end{aligned}$$

Caso as condições do Teorema 3.36 sejam cumpridas pelas restrições de g a cada A_k , obtemos a função de densidade de Y diferenciando a função de distribuição de Y . Lembramos ao leitor que a diferenciação termo-a-termo é permitida se a série diferenciada for uniformemente convergente (Spivak, 1994).

Vejamos um exemplo no qual a variável aleatória é positiva e no qual temos uma transformação com um número enumerável de inversas.

Exemplo 3.50

Seja $X \sim \text{Exponencial}(\theta)$, ou seja, $f(x) = \begin{cases} \theta e^{-\theta x}, & \text{se } x > 0 \\ 0, & \text{se } x \leq 0 \end{cases} \quad \theta > 0$. Escolhemos $Y = \sin(X)$ e seja $\sin^{-1}(y)$ o valor principal. Então, para $0 < y < 1$

$$\begin{aligned}
 P(\sin(X) \leq y) &= \\
 &= P\left(0 < X \leq \sin^{-1}(y) \quad \text{ou} \quad (2n-1)\pi - \sin^{-1}(y) \leq X \leq 2n\pi + \sin^{-1}(y), \right. \\
 &\qquad\qquad\qquad \left. \text{para algum inteiro } n \geq 1\right) \\
 &= P\left(0 < X \leq \sin^{-1}(y)\right) + \sum_{n=1}^{\infty} P\left((2n-1)\pi - \sin^{-1}(y) \leq X \leq 2n\pi + \sin^{-1}(y)\right) \\
 &= 1 - e^{-\theta \sin^{-1}(y)} + \sum_{n=1}^{\infty} \left\{ e^{-\theta((2n-1)\pi - \sin^{-1}(y))} - e^{-\theta(2n\pi + \sin^{-1}(y))} \right\} \\
 &= 1 - e^{-\theta \sin^{-1}(y)} + \left(e^{\theta\pi + \theta \sin^{-1}(y)} - e^{-\theta \sin^{-1}(y)} \right) \sum_{n=1}^{\infty} e^{-2\theta\pi n} \\
 &= 1 - e^{-\theta \sin^{-1}(y)} + \left(e^{\theta\pi + \theta \sin^{-1}(y)} - e^{-\theta \sin^{-1}(y)} \right) \left(\frac{e^{-2\theta\pi}}{1 - e^{-2\theta\pi}} \right) \\
 &= 1 + \frac{e^{-\theta\pi + \theta \sin^{-1}(y)} - e^{-\theta \sin^{-1}(y)}}{1 - e^{-2\theta\pi}}.
 \end{aligned}$$

Um cálculo similar pode ser feito para $y < 0$. Segue que a função de densidade de Y é dada por

$$h(y) = \begin{cases} \frac{\theta e^{-\theta\pi}}{1 - e^{-2\theta\pi}} \frac{e^{\theta \sin^{-1}(y)} + e^{-\theta\pi - \theta \sin^{-1}(y)}}{\sqrt{1 - y^2}}, & \text{se } -1 < y < 0 \\ \frac{\theta}{1 - e^{-2\theta\pi}} \frac{e^{-\theta \sin^{-1}(y)} + e^{-\theta\pi + \theta \sin^{-1}(y)}}{\sqrt{1 - y^2}}, & \text{se } 0 < y < 1 \\ 0, & \text{caso contrário} \end{cases}.$$

3.4 Exercícios

Exercícios da Seção 3.1

1. Demonstrar o Teorema 3.1.
2. Demonstre o resultado contrário no Teorema 3.4, ou seja, prove que $A \in \mathcal{F}$ então I_A é variável aleatória.
3. Demonstrar o Teorema 3.8.
4. Demonstrar o Teorema 3.10.
5. Suponha X uma variável aleatória não negativa com média $\mu > 0$ e variância $\sigma^2 < \infty$. O coeficiente de variação de X é definido como $CV(X) = \sigma/\mu$. Considere X e Y duas variáveis aleatórias não negativas com coeficientes de variação finito $CV(X)$ e $CV(Y)$, respectivamente. Prove que

$$CV(X + Y) \leq CV(X) + CV(Y).$$

6. Demonstre que se a variável aleatória X é independente de si mesma, então X é uma variável degenerada.
7. Suponhamos que X e Y são duas variáveis aleatórias discretas assumindo um número finito de valores x_1, \dots, x_m e y_1, \dots, y_n , respectivamente. Prove que X e Y são independentes se

$$P(X = x_i, Y = y_j) = P(X = x_i)P(Y = y_j)$$

para $1 \leq i \leq m-1$ e $1 \leq j \leq n-1$. Em outras palavras, para provar que X e Y são independentes, basta verificar que $(m-1)(n-1)$ equações.

8. Mostre que se as variáveis aleatórias (X, Y) são permutáveis, com densidade conjunta $f(x, y)$, então

$$P(X < Y) = P(X > Y) = \frac{1}{2},$$

com $P(X = Y) = 0$.

9. Considere as variáveis aleatórias contínuas permutáveis X_1, \dots, X_n , com função de densidade $f(x_1, \dots, x_n)$. Prove que

$$P(X_1 < X_2 < \dots < X_n) = P(X_{\pi_1} < X_{\pi_2} < \dots < X_{\pi_n}) = \frac{1}{n!},$$

com $P(X_i = X_j) = 0$ para algum par (i, j) , tal que $i \neq j$.

10. Seja $\Omega = \{-1, 0, 1\}$ um espaço amostral e $\mathcal{F} = \{\emptyset, \{0\}, \{-1, 1\}, \Omega\}$ uma σ -álgebra definida em Ω . Considere $X(\omega) = \omega$. Prove que X^2 é variável aleatória porém X não é variável aleatória.
11. Um dado é lançado cinco vezes. Seja X a soma dos valores de face. Escreva os eventos $\{X = 4\}$, $\{X = 6\}$, $\{X = 30\}$, $\{X \geq 29\}$.
12. Sejam X_1, X_2, \dots, X_n variáveis aleatórias. Prove que:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad \text{e} \quad S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2,$$

são variáveis aleatórias.

13. Seja X uma variável aleatória. Prove que as funções X^2 , e^X , $\sin X$, $\cos(X)$ e $1/X$ são variáveis aleatórias, onde $\{X = 0\} = \emptyset$.
14. Seja X o número de caras em três lançamentos de uma moeda. Qual é o espaço amostral? Quais valores X faz corresponder a cada elemento de Ω ? Descreva os eventos $\{X \leq 2.75\}$ e $\{0.5 \leq X \leq 1.72\}$

15. Seja $\Omega = [0, 1]$ e $\mathcal{B}(\mathbb{R})$ a σ -álgebra dos Borelianos em Ω . Defina X em Ω como

$$X(\omega) = \begin{cases} \omega & \text{se } 0 \leq \omega \leq 1/2 \\ \omega - 1/2 & \text{se } 1/2 < \omega \leq 1 \end{cases}.$$

Prove que X é variável aleatória. Qual é o evento $\{\omega : X(\omega) \in (1/4, 1/2)\}$?

16. Seja Ω um espaço amostral e $\mathcal{F} = \{\emptyset, \Omega\}$ uma σ -álgebra definida em Ω . Prove que a função $X : \Omega \rightarrow \mathbb{R}$ é variável aleatória se, e somente se, X é constante.
17. Prove que X é variável aleatória se, e somente se, $\{X < x\} \in \mathcal{F}$, $\forall x \in \mathbb{R}$.
18. Prove que X é variável aleatória se, e somente se, $\{X \geq x\} \in \mathcal{F}$, $\forall x \in \mathbb{R}$.
19. Prove que X é variável aleatória se, e somente se, $\{X > x\} \in \mathcal{F}$, $\forall x \in \mathbb{R}$.
20. Prove que X é variável aleatória se, e somente se, $\{a < X < b\}$ pertence à σ -álgebra \mathcal{F} para todo intervalo (a, b) em \mathbb{R} .
21. Seja c uma constante e X uma variável aleatória. Prove que $\max\{X, c\}$ e $\min\{X, c\}$ também são variáveis aleatórias.
22. Seja X uma variável aleatória qualquer. Prove que a parte inteira de X , denotada por $[X]$, é uma variável aleatória discreta.
23. Prove que a função $X : \Omega \rightarrow \mathbb{R}$ é variável aleatória se, e somente se, as funções $X^+ = \max\{0, X\}$ e $X^- = -\min\{0, X\}$ forem também variáveis aleatórias.
24. Seja Ω um espaço amostral e $\mathcal{F} = \{\emptyset, A, A^c, \Omega\}$ uma σ -álgebra em Ω , com $A \subseteq \Omega$. Prove que $X : \Omega \rightarrow \mathbb{R}$ constante em A e em A^c é variável aleatória. Desta forma, toda variável aleatória nesta σ -álgebra assume somente dois valores distintos.
25. Seja A_1, \dots, A_n uma partição finita de Ω , isto é, $A_i \cap A_j = \emptyset$ para $i \neq j$ e $\cup_{i=1}^n A_i = \Omega$. Considere ainda $\mathcal{F} = \sigma\{A_1, \dots, A_n\}$, ou seja, \mathcal{F} a σ -álgebra gerada pela partição. Prove que $X : \Omega \rightarrow \mathbb{R}$ é variável aleatória se, e somente se, X é constante para cada elemento da partição. Como consequência, X assume no máximo n valores distintos.
26. Seja X uma variável aleatória e $a < b$ duas constantes. Prove que as seguintes funções são variáveis aleatórias:

a)

$$Y = \begin{cases} X & \text{se } X < a \\ a & \text{se } X \geq a \end{cases}.$$

b)

$$Y = \begin{cases} a & \text{se } X < a \\ X & \text{se } a \leq X \leq b \\ b & \text{se } X \geq b \end{cases}.$$

c)

$$Y = \begin{cases} X & \text{se } |X| \leq a \\ 0 & \text{se } |X| > a, \end{cases}$$

supondo que $a > 0$.

Exercícios da Seção 3.2

1. Cinco pontos são escolhidos, independentemente e ao acaso, do intervalo $[0, 1]$. Seja X o número de pontos que pertencem ao intervalo $[0, c]$, onde $0 < c < 1$. Qual é a distribuição de X ?
2. Um ponto é selecionado, ao acaso, do quadrado unitário $[0, 1] \times [0, 1]$. Seja X a primeira coordenada do ponto selecionado. Qual a função de distribuição dessa variável aleatória. Represente-a graficamente.
3. Demonstre o item 2 do Teorema 3.16.
4. Seja X uma variável aleatória com função de distribuição

$$F_X(x) = \begin{cases} 0 & \text{se } x < 0 \\ x & \text{se } 0 \leq x \leq 1/2 \\ 1 & \text{se } x > 1/2 \end{cases}.$$

Encontre $P(X > 1/4)$ e $P(1/3 < X \leq 3/8)$.

5. Seja X uma variável aleatória com função de distribuição

$$F(x) = \begin{cases} 0 & \text{se } x < 0 \\ 1 - e^{-x} & \text{se } x \geq 0 \end{cases}.$$

Encontre $P(-\infty < X < 2)$.

6. Seja X uma variável aleatória com função de distribuição

$$F(x) = \begin{cases} 0 & \text{se } x < 2 \\ 1 - \frac{4}{x^2} & \text{se } x \geq 2 \end{cases}.$$

Calcule $P(X \leq 1)$, $P(X = 1)$, $P(0 < X < 3)$ e $P(X \geq 3)$.

7. Prove que as seguintes funções são de distribuição:

a)

$$F(x) = \begin{cases} 0 & \text{se } x \leq 0 \\ 1 - e^{-x} & \text{se } x > 0 \end{cases}.$$

b)

$$F(x) = \begin{cases} 0 & \text{se } x \leq 0 \\ 1 - (1+x)e^{-x} & \text{se } x > 0 \end{cases}.$$

c)

$$F(x) = \begin{cases} 0 & \text{se } x < -1 \\ \frac{x+1}{2} & \text{se } -1 \leq x \leq 1 \\ 1 & \text{se } x > 1 \end{cases}.$$

8. Sejam $F(x)$ e $G(x)$ duas funções de distribuição contínuas e estritamente crescentes. Prove que:

a) Se $F(x) \geq G(x)$, então $F^{-1}(x) \leq G^{-1}(x)$,

b) Se X tem distribuição $F(x)$, então $Y = G^{-1}(F(X))$ tem função de distribuição $G(x)$,

c) Se $F(x) = G(x)$, então existem variáveis aleatórias X e Y de funções de distribuição são $F(x)$ e $G(x)$, respectivamente, e tais que $X = Y$. Sugestão: Use o resultado anterior.

9. Seja X uma variável aleatória com função de distribuição $F(x)$. Prove que $F(x)$ é contínua em $x = x_0$ se, e somente se, $P(X = x_0) = 0$.
10. Seja $F(x)$ uma função de distribuição contínua. Prove que para qualquer inteiro $n \geq 1$, as seguintes funções também são de distribuição:

- a) $[F(x)]^n$,
- b) $1 - [1 - F(x)]^n$.

11. Prove que

$$f(x) = \begin{cases} 0 & \text{se } x \leq 0 \\ \theta^2 x e^{-\theta x} & \text{se } x > 0 \end{cases},$$

onde $\theta > 0$, define uma função de densidade. Encontre a função de distribuição de probabilidades e se X tiver esta como função de densidade, encontre $P\{X \geq 1\}$.

12. Para quais valores da constante c as seguintes funções são de probabilidade para alguma variável aleatória?

- a) $f(x) = c\lambda^x/x!$, para $x = 0, 1, 2, \dots$ e $\lambda > 0$.
- b) $f(x) = c/N$, para $x = 1, 2, \dots, N$.

13. Seja $f(x)$ uma função de densidade e c uma constante qualquer. Prove que $f(x+c)$ é também uma função de densidade.

14. Prove que as seguintes funções são de distribuição, encontre as correspondentes funções de densidade e prove que, de fato, são funções de densidade. Apresente-as graficamente.

a)

$$F(x) = \begin{cases} 0 & \text{se } x < 0 \\ 1 & \text{se } x \geq 0, \end{cases}$$

b)

$$F(x) = \begin{cases} 0 & \text{se } x \leq 0 \\ x & \text{se } 0 < x < 1 \\ 1 & \text{se } x \geq 1, \end{cases}$$

c)

$$F(x) = e^x/(1 + e^x),$$

d)

$$F(x) = \frac{1}{2} \int_{-\infty}^x e^{|u|} du.$$

15. Seja X uma variável aleatória com densidade

$$f(x) = \begin{cases} cx^2 & \text{se } -1 \leq x \leq 1 \\ 0 & \text{caso contrário.} \end{cases}$$

Determine o valor da constante c para que $f(x)$ seja efetivamente função de densidade e ache o valor de α tal que $F_X(\alpha) = 1/4$, o valor de α que satisfaz a equação anterior é conhecido como primeiro quartil da distribuição.

16. Uma variável aleatória X tem função de distribuição

$$F_X(x) = \begin{cases} 0 & \text{se } x < 0, \\ x^3 & \text{se } 0 \leq x \leq 1 \\ 1 & \text{se } x > 1. \end{cases}$$

Qual é a densidade de X ?

Exercícios da Seção 3.3

1. Seja $X \sim \text{Binomial}(n, p)$, ou seja, uma variável aleatória com função de probabilidade

$$P(X = r) = \binom{n}{r} p^r (1-p)^{n-r}, \quad r = 0, 1, 2, \dots, n, \quad 0 \leq p \leq 1.$$

Encontrar as funções de probabilidade das variáveis aleatórias:

- (a) $Y = aX + b$, sendo a e b constantes conhecidas.
 - (b) $Y = X^2$.
 - (c) $Y = \sqrt{X}$.
2. Seja X uma variável aleatória não negativa com densidade f_X e $Y = X^n$. Encontre a expressão da função de densidade de Y em termos de f_X .
3. Seja X uma variável aleatória com densidade

$$f_X(x) = \begin{cases} 0 & \text{se } x \leq 0 \\ \frac{1}{(x+1)^2} & \text{se } x > 0 \end{cases}.$$

Seja $Y = \max\{X, c\}$, onde c é uma constante estritamente positiva.

- a) Ache a função de distribuição de Y .
 - b) Decomponha F_Y em suas partes discreta, absolutamente contínua e singular.
4. Seja X uma variável aleatória com função de densidade f_X . Seja $Y = a + bX$, onde a e b são duas constantes. Prove que, se $a \neq 0$ então

$$f_Y(y) = \frac{1}{|a|} f_X((y-b)/a).$$

5. Seja X uma variável aleatória com função de densidade

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

Seja $Y = (X - 1/\theta)^2$. Encontre a função de densidade de Y .

6. Seja $X \sim \text{Cauchy}(0, \sigma)$. Prove que $c/X \sim \text{Cauchy}(0, |c|/\sigma)$, sendo c uma constante.
7. Prove que $X \sim \text{Cauchy}(0, 1)$ se, e somente se, $1/X \sim \text{Cauchy}(0, 1)$.
8. Seja $X \sim \text{Uniforme}(-\pi/2, \pi/2)$. Prove que $Y = \tan(X)$ tem distribuição Cauchy.
9. Seja X uma variável aleatória com função de densidade

$$f_X(x) = \begin{cases} \frac{1}{2\pi} & \text{se } 0 < x < 2\pi \\ 0 & \text{caso contrário.} \end{cases}$$

Seja $Y = \sin(X)$. Encontre a função de densidade de Y .

10. Seja X uma variável aleatória com função de densidade

$$f_X(x) = \begin{cases} \frac{1}{3} & \text{se } -1 < x < 2\pi \\ 0 & \text{caso contrário.} \end{cases}$$

Seja $Y = |X|$. Encontre a função de densidade de Y .

11. Seja X uma variável aleatória com função de densidade

$$f_X(x) = \begin{cases} \frac{1}{2\theta} & \text{se } -\theta < x < \theta \\ 0 & \text{caso contrário.} \end{cases}$$

Seja $Y = 1/X^2$. Encontre a função de densidade de Y .

12. Determine a função de densidade de $Y = (b-a)X + a$, sendo $X \sim Uniforme(0, 1)$.
13. Se X tem densidade $f_X(x) = e^{-|x|}/2$, para $-\infty < x < \infty$. Qual a distribuição de $Y = |X|$?
14. Seja X uma variável aleatória com função de densidade f_X . Encontre a função de densidade de $U = \frac{X}{1+X}$. Se, em particular, $X \sim Uniforme(0, 1)$, qual é a função de densidade de U ?
15. Seja X uma variável aleatória com função de densidade f_X e seja $Y = g(X)$, g definida como segue:

$$\begin{aligned} \text{(a)} \quad g(x) &= \begin{cases} 1 & \text{se } x > 0 \\ -1 & \text{se } x \leq 0 \end{cases} \\ \text{(b)} \quad g(x) &= \begin{cases} b & \text{se } x \geq b \\ x & \text{se } |x| < b \\ -b & \text{se } x \leq -b \end{cases} \\ \text{(c)} \quad g(x) &= \begin{cases} x & \text{se } |x| \geq b \\ 0 & \text{se } |x| < b \end{cases} \end{aligned}$$

Encontre a função de distribuição de Y em cada caso.

16. Seja $X \sim Normal(2, 4)$. Encontre a função de densidade de $Y = X^+$.
17. Seja $X \sim Normal(0, \sigma^2)$, sendo $\sigma > 0$ uma quantidade real qualquer. Seja $Y = X^3$ e encontre f_Y a função de densidade de Y .
18. A tensão X é uma variável aleatória $Normal(1, 2)$. Encontre a função de densidade da potência dissipada P por um resistor de R -ohm, sendo a potência calculada como $P = RX^2$.
19. A amplitude de um sinal de rádio X é uma variável aleatória com função de densidade $Rayleigh(\alpha)$ de expressão:

$$f_X(x) = \frac{x}{\alpha^2} e^{-x^2/2\alpha^2}, \quad x > 0, \alpha > 0.$$

- (a) Encontre a função de densidade de $Y = (X - r)^+$, sendo $r > 0$ um número real.
- (b) Encontre a função de densidade de $Y = X^2$.
20. Consideremos as constantes $\beta > 0$ e $\lambda > 0$. Dizemos que a variável aleatória Y tem distribuição $Weibull(\beta, \lambda)$ se:

$$F_X(x) = \begin{cases} 1 - e^{-(x/\lambda)^\beta} & \text{se } x \geq 0 \\ 0 & \text{se } x < 0 \end{cases}$$

Seja $Y = (X/\lambda)^\beta$. Encontre a função de densidade de Y .

Capítulo 4

Momentos

O estudo das funções de distribuição de variáveis aleatórias é essencialmente o estudo de algumas características numéricas associadas. Esses parâmetros que caracterizam as funções de distribuição e, portanto, as funções de densidade ou de probabilidade associadas têm um papel muito importante no cálculo das probabilidades.

Aqui queremos estudar medidas pontuais que permitirão reduzir os valores das variáveis aleatórias a números especiais com os quais as identificaremos.

Por exemplo, se X é uma variável aleatória com função de distribuição F_X ; dado $F_X(x)$ para todos os valores de x , sabemos tudo o que precisamos sobre X . No entanto, às vezes é útil resumir a distribuição de X por certas características, como esperança ou valor esperado da distribuição. Estes valores são uma forma de caracterizar a distribuição, essencialmente o valor esperado de X é o centro de massa de sua distribuição. Este conceito será melhor compreendido no seguinte exemplo.

Exemplo 4.1

Definamos por X uma variável aleatória discreta de valores x_1, \dots, x_n com probabilidades p_1, \dots, p_n . Imagine, por exemplo, certas vigas com massas p_1, \dots, p_n suspensas pelos pontos x_1, \dots, x_n . Suponha que tentamos equilibrar as vigas sobre um ponto de apoio μ . A força para baixo exercida pela massa colocada em x_i é proporcional à $p_i|x_i - \mu|$ e, portanto, para o feixe estar em equilíbrio no fulcro μ , devemos ter

$$\sum_{x_i < \mu} p_i |x_i - \mu| = \sum_{x_i > \mu} p_i |x_i - \mu|$$

ou

$$\sum_{x_i < \mu} p_i (\mu - x_i) = \sum_{x_i > \mu} p_i (x_i - \mu). \quad (4.1)$$

Resolvendo esta equação para μ , obtemos

$$\mu = \sum_{i=1}^n x_i p_i = \sum_{i=1}^n x_i P(X_i = x_i),$$

como expressão do centro de massa μ , chamado de esperança ou valor esperado de X .

4.1 Esperança

Certamente um dos conceitos mais conhecidos na teoria das probabilidade é a esperança de uma variável aleatória, mas não com esse nome e sim com os nomes de média ou valor esperado. Durante todo este estudo consideraremos X uma variável aleatória qualquer com função de distribuição F_X .

Estender o conceito de centro de massa para distribuições mais gerais é possível, embora conte com algumas sutilezas matemáticas: não é claro que o centro de massa existirá se X assumir infinitos valores, sejam estes enumeráveis ou não.

Se, por exemplo,

$$P(X = x) = \frac{6}{\pi^2 x^2},$$

para $x = 1, 2, \dots$ temos que

$$\mu = \sum_{x=1}^{\infty} xP(X = x) = \frac{6}{\pi^2} \sum_{x=1}^{\infty} \frac{1}{x} = \infty,$$

isto seguindo o desenvolvimento no Exemplo 4.1. Significa que, para esta função de probabilidade o centro de massa não existe.

Definição 4.1 (*Esperança*)

Suponha X uma variável aleatória não negativa com função de distribuição F_X . A esperança ou valor esperado de X , denotado por $E(X)$, é definido como

$$E(X) = \int_0^{\infty} (1 - F_X(x)) dx, \quad (4.2)$$

quando a integral esteja bem definida.

Observemos que uma variável aleatória X é não negativa se $X(w) \geq 0$, $\forall w \in \Omega$ ou, de maneira equivalente, se $P(X \geq 0) = 1$. Nesta situação, a função de distribuição satisfaz que $F_X(x) = 0$, para todo $x < 0$. Ainda observamos que

$$E(X) = \int_0^{\infty} (1 - F_X(x)) dx = \int_0^{\infty} P(X > x) dx. \quad (4.3)$$

Na Definição 4.1 com a frase “quando a integral esteja bem definida” queremos dizer que a integral em (4.2) ou em (4.3) possa ser calculada, sendo o resultado finito.

Exemplo 4.2 (*Distribuição Exponencial*)

Seja $X \sim \text{Exponencial}(\theta)$. Sabemos que a função de distribuição assume a forma $F_X(x) = 1 - e^{-\theta x}$. Então

$$E(X) = \int_0^{\infty} (1 - F_X(x)) dx = \int_0^{\infty} (1 - (1 - e^{-\theta x})) dx = \frac{1}{\theta}.$$

Observemos também que

$$\int_0^{\frac{1}{\theta}} \left(\frac{1}{\theta} - x \right) \theta e^{-\theta x} dx = e^{-\theta}, \quad (4.4)$$

a que, por outro lado

$$\int_{\frac{1}{\theta}}^{\infty} \left(x - \frac{1}{\theta} \right) \theta e^{-\theta x} dx = e^{-\theta}. \quad (4.5)$$

Isto significa que, de fato, $1/\theta$ é o centro de massa da distribuição *Exponencial*(θ). As expressões em (4.4) e (4.5) acima representam as equações similares às equações em (4.1), que o centro de massa ou esperança deve satisfazer, ou seja, estas são as equivalentes contínuas das expressões em (4.1).

Observando a definição de esperança, os exemplos mais simples são àqueles nos quais a função de distribuição assume uma expressão funcional conhecida, como no exemplo anterior. A seguinte é uma situação onde a variável é discreta.

Exemplo 4.3 (*Distribuição Bernoulli*)

No caso em que X é Bernoulli de parâmetro p , temos que

$$P(X = x) = p\mathbf{1}_{\{x_1\}}(x) + (1 - p)\mathbf{1}_{\{x_2\}}(x),$$

onde x_1 e x_2 são os valores que assume a variável aleatória e $x \in \mathbb{R}$. Então

$$E(X) = \int_0^{\infty} (1 - F_X(x)) dx = x_1 + (1 - p) \times (x_2 - x_1) = px_1 + (1 - p)x_2.$$

Não sabemos muito acerca da esperança, mas vamos considerar que $x_1 < \mu < x_2$, sendo $\mu = E(X)$. Aceita esta consideração vamos verificar se as relações em (4.1) são válidas nesta situação. Então

$$\begin{aligned} \sum_{x_i < \mu} p_i(\mu - x_i) &= P(X = x_1)(\mu - x_1) = p(px_1 + (1 - p)x_2 - x_1) \\ &= p(1 - p)(x_2 - x_1) \end{aligned}$$

e

$$\begin{aligned} \sum_{x_i > \mu} p_i(x_i - \mu) &= P(X = x_2)(x_2 - \mu) = (1 - p)(x_2 - px_1 - (1 - p)x_2) \\ &= p(1 - p)(x_2 - x_1). \end{aligned}$$

Verificamos assim que $\mu = px_1 + (1 - p)x_2$ é o centro de massa e, como definido, é o valor da esperança nesta situação.

A questão agora é: realmente $x_1 < \mu < x_2$? verificar isso é o mesmo que verificar se

$$x_1 \leq px_1 + (1 - p)x_2 \leq x_2. \quad (4.6)$$

A solução desta questão é deixada ao leitor.

Uma situação particular importante acontece quando $x_1 = 0$ e $x_2 = 1$, neste caso $\mu = 1 - p$.

Teorema 4.1

Se a variável aleatória X assume somente valores inteiros não negativos, então

$$E(X) = \sum_{n=0}^{\infty} P(X > n) = \sum_{n=1}^{\infty} P(X \geq n). \quad (4.7)$$

Demonstração: Consideramos valores inteiros não negativos $1, 2, 3, \dots$. Então

$$P(X > n) = P(X \geq n+1) \quad n = 1, 2, 3, \dots,$$

portanto

$$\sum_{n=0}^{\infty} P(X > n) = \sum_{n=0}^{\infty} P(X \geq n+1) = \sum_{n=1}^{\infty} P(X \geq n).$$

Pela definição de esperança e segundo a Definição 3.8, de função de distribuição discreta, temos que

$$\begin{aligned} E(X) &= \int_0^{\infty} (1 - F_X(x)) dx = \int_0^{\infty} \left(1 - \sum_{n=1}^{\infty} P(X = n) \mathbf{1}_{\{n\}}(x)\right) dx \\ &= \sum_{n=0}^{\infty} (1 - F_X(n)) = \sum_{n=0}^{\infty} P(X > n). \end{aligned} \quad \blacksquare$$

Para entendermos a demonstração do Teorema 4.1 prestemos atenção à Figura 4.1. Mostramos dois gráficos, em (a) a função de distribuição $F_X(n)$ numa situação específica e em (b) $1 - F_X(n)$. Vejamos os detalhes no Exemplo 4.4.

Exemplo 4.4 (Distribuição Geométrica)

Seja $X \sim \text{Geométrica}(\theta)$, ou seja, X é uma variável aleatória discreta, com valores nos números naturais incluindo o zero e que tem por função de probabilidade Geométrica de parâmetro θ . Significa que

$$P(X = x) = \theta(1 - \theta)^x, \quad x = 0, 1, 2, \dots, \quad 0 < \theta < 1.$$

A função de distribuição geométrica de parâmetro $\theta = 0.6$ é mostrada para os primeiros 6 valores da variável no gráfico (a) da Figura 4.1. Lembremos que a função de distribuição é a soma das probabilidades até o n , assim

$$F_X(n) = \sum_{x=0}^n P(X = x) = \sum_{x=0}^n \theta(1 - \theta)^x, \quad (4.8)$$

é a expressão matemática da função de distribuição geométrica.

Na Figura 4.1, gráfico (b), mostramos a área de $1 - F_X(n)$, cujo valor nos permite

encontrar o valor da esperança. Neste caso

$$\begin{aligned} E(X) &= \sum_{n=0}^{\infty} P(X \geq n) = \sum_{n=0}^{\infty} (1 - F_X(n)) \\ &= \sum_{n=0}^{\infty} \left(1 - \sum_{x=0}^n 0.6 \times 0.4^x \right) = 1.66666. \end{aligned}$$

Aqui utilizamos os resultados de uma progressão geométrica, isto é, de uma coleção de números que satisfazem a relação ar^x , onde $|r| < 1$. Nesta situação a soma dos n -ésimos primeiros termos é

$$\frac{a(1 - r^n)}{1 - r}, \quad \text{e a soma da série é} \quad \frac{a}{1 - r}.$$

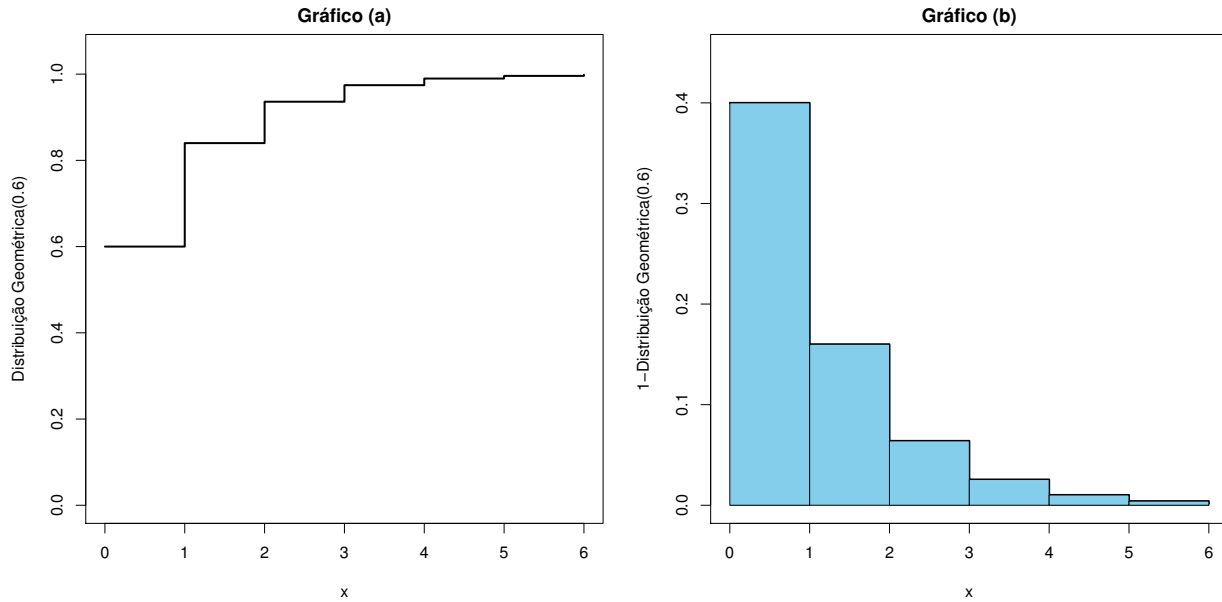


Figura 4.1: Gráfico (a) Função de distribuição Geométrica(0.6), a área por cima desta curva é $E(X)$. Gráfico (b) A direita mostramos a função $1 - F_X(n)$, sendo $F_X(n)$ a função de distribuição Geométrica definida em (4.8) e aqui apresentada somente para os primeiros 6 valores. A área sombreada é o valor de $E(X)$.

Exemplo 4.5

Dizemos que uma variável aleatória tem distribuição uniforme discreta nos primeiros N números naturais se

$$P(X = n) = \frac{1}{N}, \quad n = 1, 2, \dots, N.$$

Esta é uma situação de variável aleatória discreta positiva. Então, segundo o Teorema

4.1 devemos calcular a soma

$$\begin{aligned}\sum_{n=1}^N P(X \geq n) &= 1 + P(X \geq 2) + P(X \geq 3) + \cdots + P(X \geq N) \\ &= 1 + \frac{N-1}{N} + \frac{N-2}{N} + \cdots + \frac{1}{N} \\ &= \frac{N}{N} + \frac{N-1}{N} + \frac{N-2}{N} + \cdots + \frac{1}{N}.\end{aligned}$$

Observemos que o numerador na expressão acima corresponde à soma dos N primeiros números naturais, que é

$$1 + 2 + \cdots + N - 2 + N - 1 + N = \frac{N(N+1)}{2},$$

do qual obtemos que

$$E(X) = \frac{N+1}{2}.$$

Fomos convencidos de que o centro de massa, quando existe, ou seja, quando é finito corresponde à esperança de uma variável aleatória positiva. Vejamos agora como calcular a integral em (4.2) para cada caso: na situação de variáveis discretas e caso estas sejam contínuas.

Teorema 4.2 (Cálculo da esperança no caso discreto)

Seja X uma variável aleatória discreta assumindo valores x_1, x_2, \dots com probabilidades p_1, p_2, \dots . Então a esperança de X calcula-se como

$$E(X) = \sum_{i=1}^{\infty} x_i P(X = x_i) = \sum_{i=1}^{\infty} x_i p_i, \quad (4.9)$$

quando a série é convergente.

Demonstração: Seja $I = \int_0^{\infty} (1 - F_X(x)) dx$. Então

$$I = \sum_{k=1}^{\infty} \int_{\frac{k-1}{n}}^{\frac{k}{n}} P(X > x) dx$$

e, desde que $P(X > x)$ é uma função não crescente, temos que

$$\frac{1}{n} \sum_{k=1}^{\infty} P\left(X > \frac{k}{n}\right) \leq I \leq \frac{1}{n} \sum_{k=1}^{\infty} P\left(X > \frac{k-1}{n}\right),$$

para todo inteiro positivo n . Sejam

$$U_n = \frac{1}{n} \sum_{k=1}^{\infty} P\left(X > \frac{k-1}{n}\right) \quad \text{e} \quad L_n = \frac{1}{n} \sum_{k=1}^{\infty} P\left(X > \frac{k}{n}\right).$$

Obtemos

$$L_n = \frac{1}{n} \sum_{k=1}^{\infty} \sum_{i=k}^{\infty} P\left(\frac{i}{n} < X \leq \frac{i+1}{n}\right) = \frac{1}{n} \sum_{k=2}^{\infty} (k-1) P\left(\frac{k-1}{n} < X \leq \frac{k}{n}\right),$$

reagrupando as séries. Então

$$\begin{aligned} L_n &= \sum_{k=2}^{\infty} \frac{k}{n} P\left(\frac{k-1}{n} < X \leq \frac{k}{n}\right) - \sum_{k=2}^{\infty} \frac{1}{n} P\left(\frac{k-1}{n} < X \leq \frac{k}{n}\right) \\ &= \sum_{k=2}^{\infty} \frac{k}{n} \sum_{\frac{k-1}{n} < x_i \leq \frac{k}{n}} p_i - \frac{1}{n} P\left(X > \frac{1}{n}\right) \\ &\geq \sum_{k=1}^{\infty} \sum_{\frac{k-1}{n} < x_i \leq \frac{k}{n}} x_i p_i - \frac{1}{n} = E(X) - \frac{1}{n}. \end{aligned}$$

De maneira similar, pode-se demonstrar que, para todo inteiro positivo n ,

$$U_n \leq E(X) + \frac{1}{n}.$$

Então, temos que

$$E(X) - \frac{1}{n} \leq \int_0^{\infty} [1 - F_X(x)] dx \leq E(X) + \frac{1}{n}$$

para todo inteiro positivo n . Tomando limite quando $n \rightarrow \infty$, vemos que

$$E(X) = \int_0^{\infty} (1 - F_X(x)) dx. \quad \blacksquare$$

Este valor está bem definido quando a soma não depende da ordem dos termos, em particular quando a série converge absolutamente, isto é, quando

$$\sum_{i=1}^{\infty} |x_i| P(X = x_i) < \infty. \quad (4.10)$$

Devemos observar que a série

$$\sum_{i=1}^{\infty} x_i P(X = x_i)$$

pode ser finita mas a série em (4.10) não. Nestas situações dizemos que $E(X)$ não existe. Vejamos no seguinte exemplo os detalhes desta particularidade.

Exemplo 4.6

Consideremos a variável aleatória X com função de probabilidade dada por

$$p_k = P\left(X = (-1)^{k+1} \frac{3^k}{k}\right) = \frac{2}{3^k}, \quad k = 1, 2, \dots$$

Então

$$\sum_{k=1}^{\infty} |x_k| p_k = \sum_{k=1}^{\infty} \frac{2}{k} = \infty,$$

e $E(X)$ não existe. Por outro lado, a série

$$\sum_{k=1}^{\infty} x_k p_k = \sum_{k=1}^{\infty} (-1)^{k+1} \frac{2}{k} = \log(4) \approx 1.3862 \dots$$

é convergente. Mas ainda, se a série acima é convergente, então a podemos decompor como a soma dos número pares mais a soma dos números ímpares, assim vemos que

$$\sum_{k=1}^{\infty} (-1)^{2k+1} \frac{2}{2k} + \sum_{k=1}^{\infty} (-1)^{2k+2} \frac{2}{2k+2} \neq \sum_{k=1}^{\infty} (-1)^{k+1} \frac{2}{k},$$

isto porque as somas das números pares e ímpares são divergentes.

Teorema 4.3 (Cálculo da esperança no caso contínua)

Seja X uma variável aleatória contínua assumindo valores positivos com função de densidade de f_X . Então a esperança de X calcula-se como

$$E(X) = \int_0^{\infty} x f_X(x) dx, \quad (4.11)$$

quando a integral é convergente.

Demonstração: Consideremos que $E(X) < \infty$, então

$$E(X) = \int_0^{\infty} x f(x) dx = \lim_{n \rightarrow \infty} \int_0^n x f(x) dx.$$

Integrando por partes obtemos

$$\int_0^n x f(x) dx = n F_X(n) - \int_0^n F_X(x) dx = -n(1 - F_X(n)) \Big|_0^n + \int_0^n (1 - F_X(x)) dx.$$

Mas,

$$n(1 - F_X(n)) \Big|_0^n = n \int_n^{\infty} f(x) dx < \int_n^{\infty} x f(x) dx$$

e, como $E(X) < \infty$, segue que $\lim_{n \rightarrow \infty} n(1 - F_X(n)) \rightarrow 0$. Temos então que

$$E(X) = \lim_{n \rightarrow \infty} \int_0^n x f(x) dx = \lim_{n \rightarrow \infty} [1 - F_X(x)] dx = \int_0^{\infty} (1 - F_X(x)) dx. \quad \blacksquare$$

Caso X seja absolutamente contínua com função de densidade f_X dizemos que $E(X)$ existe e é igual a $\int_{-\infty}^{\infty} x f_X(x) dx$, desde que

$$\int_{-\infty}^{\infty} |x| f(x) dx < \infty. \quad (4.12)$$

Enfatizamos que a condição em (4.12) deve ser verificada antes se pode concluir que $E(X)$ existe e é igual a $\int_{-\infty}^{\infty} xf_X(x) dx$. Além disso, vale a pena recordar neste ponto que a integral $\int_{-\infty}^{\infty} g(x) dx$ existe, desde que exista o limite $\lim_{a \rightarrow \infty} \int_{-a}^a g(x) dx$. É bem possível que o limite $\lim_{a \rightarrow \infty} \int_{-a}^a g(x) dx$ exista sem a existência de $\int_{-\infty}^{\infty} g(x) dx$, como no Exemplo 4.8.

Exemplo 4.7

Seja X uma variável aleatória com função de densidade $f_X(x) = \begin{cases} \frac{2}{x^3}, & x \geq 1 \\ 0, & x < 1 \end{cases}$.

Então

$$E(X) = \int_1^{\infty} \frac{2}{x^2} dx = 2.$$

Consideremos agora X assumindo valores reais quaisquer. Como vamos calcular a esperança uma variável deste tipo? a resposta aparece no resultado a seguir o qual demonstra-se em parte.

Teorema 4.4

Seja X uma variável aleatória com função de distribuição F_X . Então

$$E(X) = \int_0^{\infty} (1 - F_X(x)) dx - \int_{-\infty}^0 F_X(x) dx, \quad (4.13)$$

quando a integral é convergente.

Demonstração: Caso a variável aleatória assuma somente valores positivos, a expressão em (4.13) coincide com a expressão em (4.2) da definição de esperança de X . Interessa-nos então a situação mais geral, quando a variável assume valores reais. Considerando a variável aleatória contínua a demonstração segue utilizando integração por partes. O caso discreto é deixado como exercício para o leitor. ■

Devemos esclarecer duas coisas: a primeira é que no Teorema 4.2 não foi dito, mas somente foram considerados valores não negativos da variável discreta. O segundo esclarecimento tem a ver com a forma de cálculo da esperança, mesmo somente tendo sido demonstrado para situações positivas os teoremas 4.2 e 4.3 valem para as situações mais gerais nas quais as variáveis aleatórias assumem valores reais. Podemos então concluir que

$$E(X) = \int_{-\infty}^{+\infty} xf(x) dx, \quad (4.14)$$

caso a variável aleatória X seja absolutamente contínua, com função de densidade f e tenha

esperança finita. Caso a variável aleatória X seja discreta, então

$$E(X) = \sum_x xP(X = x), \quad (4.15)$$

caso a esperança seja finita.

Exemplo 4.8 (*Distribuição Cauchy*)

Seja X uma variável aleatória com distribuição *Cauchy*(0,1), ou seja, com função de densidade

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

Acontece que

$$\lim_{a \rightarrow \infty} \int_{-a}^a \frac{x}{\pi} \frac{1}{1+x^2} dx = 0. \quad (4.16)$$

Contudo, $E(X)$ não existe, desde que a integral

$$\int_{-\infty}^{\infty} \frac{|x|}{\pi} \frac{1}{1+x^2} dx$$

diverge. Observe que esta é uma situação muito especial porque, além das integrais em (4.16) serem finitas mas $E(X)$ não existir, a densidade *Cauchy*(0,1) é simétrica com centro de simetria em 0, mas 0 não é a esperança de X , que não existe.

Exemplo 4.9

Seja X uma variável aleatória com função de probabilidade dada por

$$P(X = -2) = P(X = 0) = \frac{1}{4}, \quad P(X = 1) = \frac{1}{3} \quad \text{e} \quad P(X = 2) = \frac{1}{6}.$$

A função de distribuição é da forma

$$F_X(x) = \begin{cases} 0 & \text{se } x \leq -2 \\ \frac{1}{4} & \text{se } -2 \leq x < 0 \\ \frac{1}{4} + \frac{1}{4} = \frac{1}{2} & \text{se } 0 \leq x < 1 \\ \frac{1}{2} + \frac{1}{3} = \frac{5}{6} & \text{se } 1 \leq x < 2 \\ 1 & \text{se } x \geq 2 \end{cases}.$$

Sabemos que para valores positivos se satisfaz que

$$\int_0^{\infty} [1 - F_X(x)] dx = \sum_{i=2}^4 x_i p_i,$$

onde $x_1 = -2$, $x_2 = 0$, $x_3 = 1$ e $x_4 = 2$ são os valores da variável, dos quais somente desconsideramos o $x_1 = -2$, por ser negativo e então $\sum_{i=2}^4 x_i p_i = \frac{2}{3}$, sendo cada p_i a

probabilidade da variável assumir cada valor. Vejamos a integral correspondente à parte negativa.

$$\int_{-\infty}^0 F_X(x) dx = \int_{-\infty}^{-2} 0 dx + \int_{-2}^0 \frac{1}{4} dx = \frac{1}{2},$$

com o qual temos que

$$E(X) = \frac{2}{3} - \frac{1}{2} = \frac{1}{6} = \sum_{i=1}^4 x_i p_i.$$

4.1.1 Propriedades da esperança

Vamos considerar nesta seção que a esperança das variáveis aleatórias sempre existem.

Teorema 4.5

Seja X uma variável aleatória que satisfaz $P(X \geq 0) = 1$. Então $E(X) \geq 0$.

Demonstração: Exercício. ■

Exemplo 4.10 (*Distribuição Binomial Negativa*)

Seja $X \sim BN(r; p)$ uma variável aleatória com função de probabilidade Binomial Negativa de parâmetros $r \geq 1$ e $0 < p < 1$, segundo a Definição (3.16). Encontremos $E(X)$.

A distribuição Binomial Negativa recebe esse nome da relação

$$\binom{x+r-1}{x} = (-1)^x \binom{-r}{x} = (-1)^x \frac{(-r)(-r-1)\cdots(-r-x+1)}{x(x-1)\cdots 2 \cdot 1}, \quad (4.17)$$

que é a equação que define o coeficiente binomial com números inteiros negativos. Juntamente com (4.17), temos $\sum_{x=0}^{\infty} P(X=x) = 1$. Da expansão binomial negativa que afirma que

$$\frac{1}{(1+t)^r} = \sum_{k=0}^{\infty} \binom{-r}{k} t^k = \sum_{k=0}^{\infty} (-1)^k \binom{k+r-1}{k} t^k.$$

Temos então

$$\begin{aligned} E(X) &= \sum_{x=0}^{\infty} x \binom{x+r-1}{x} p^r (1-p)^x = \sum_{x=1}^{\infty} \frac{(x+r-1)!}{(x-1)!(r-1)!} p^r (1-p)^x \\ &= \frac{r(1-p)}{p} \sum_{x=1}^{\infty} \binom{x+r-1}{x-1} p^{r+1} (1-p)^{x-1} \\ &= \frac{r(1-p)}{p} \sum_{z=0}^{\infty} \binom{r+1+z-1}{z} p^{r+1} (1-p)^z = r \frac{1-p}{p} > 0. \end{aligned}$$

Esta primeira propriedade nos conta que a esperança de variáveis aleatórias positivas são estritamente positivas. Como exemplos podemos mencionar quase todos os exemplos de variáveis aleatórias discretas considerados até o momento, assim como diversas variáveis contínuas. No caso de variáveis aleatórias contínuas positivas podemos mencionar a distribuição Exponencial.

Teorema 4.6

Seja X uma variável aleatória. Se $X = c$, ou seja, $X(\omega) = c$, para todo $\omega \in \Omega$. Então $E(X) = c$.

Demonstração: Exercício. ■

Esta propriedade pode ser traduzida de várias maneiras. Uma é que a esperança de uma constante é ela própria ou que a esperança de uma variável determinística (degenerada) é ela própria. Na próxima propriedade utilizamos a relação $X \leq Y$, isto significa que $X(\omega) \leq Y(\omega)$, $\forall \omega \in \Omega$.

Teorema 4.7

Sejam X e Y duas variáveis aleatórias. Se $X \leq Y$, ou seja, se $P(X \leq Y) = 1$, então $E(X) \leq E(Y)$.

Demonstração: Para esta propriedade ser válida basta que as esperanças estejam bem definidas e, com isto, queremos dizer que alguma das esperanças seja finita, ou $E(X) = -\infty$ ou $E(Y) = +\infty$. Se $Y \leq z$ então $X \leq z$, logo

$$\{\omega \in \Omega : Y(\omega) \leq z\} \subseteq \{\omega \in \Omega : X(\omega) \leq z\}.$$

Portanto, $F_Y(z) \leq F_X(z)$, sendo F_X e F_Y as funções de distribuição das variáveis aleatórias X e Y , respectivamente. Também $1 - F_Y(z) \geq 1 - F_X(z)$. Pela propriedade em (4.13),

$$E(Y) = \int_0^\infty (1 - F_Y(z)) dz - \int_{-\infty}^0 F_Y(z) dz \geq \int_0^\infty (1 - F_X(z)) dz - \int_{-\infty}^0 F_X(z) dz = E(X). \quad \blacksquare$$

Uma questão em aberto no Teorema anterior é se existem variáveis X e Y , tais que $X \leq Y$. A resposta está também no seguinte exemplo.

Exemplo 4.11

Sejam $Y \sim U(0, 1)$ e $X = \min(Y, 1/2)$, logo $X \leq Y$. Então

$$\begin{aligned} E(X) &= \int_0^{1/2} x dx + \int_{1/2}^1 \frac{1}{2} P\left(X = \frac{1}{2}\right) dx \\ &= \frac{1}{8} + \frac{1}{2} P\left(Y \geq \frac{1}{2}\right) = \frac{1}{8} + \frac{1}{4} = \frac{3}{8} < \frac{1}{2} = E(Y). \end{aligned}$$

Como consequência dos Teoremas 4.6 e 4.7 concluímos que, se $\alpha \leq X(\omega) \leq \beta$, para todo $\omega \in \Omega$, então

$$\alpha \leq E(X) \leq \beta,$$

desde que a esperança exista, ou seja, se a variável aleatória for limitada e a esperança exista, afirmamos que a esperança é também limitada. Isto ficará claro no Teorema 4.9, chamado de Teorema de Existência.

Teorema 4.8 (*Linearidade*)

Seja X uma variável aleatória para a qual $E(X) < \infty$ e sejam a e b dois números reais quaisquer. Então

$$E(aX + b) = a E(X) + b.$$

Demonstração: (a) Caso $a = 0$, temos que $E(aX + b) = E(b) = b = a E(X) + b$.

(b) Agora, caso $a > 0$ e $b \neq 0$ definamos $Y = aX + b$, então

$$F_Y(x) = P(aX + b \leq x) = P\left(X \leq \frac{x-b}{a}\right) = F_X\left(\frac{x-a}{b}\right).$$

$$\begin{aligned} E(Y) &= \int_0^\infty (1 - F_Y(x)) dx - \int_{-\infty}^0 F_Y(x) dx \\ \text{Logo} \quad &= \int_0^\infty \left(1 - F_X\left(\frac{x-a}{b}\right)\right) dx - \int_{-\infty}^0 F_X\left(\frac{x-a}{b}\right) dx, \quad \text{escrevendo } y = \\ &\quad (x-a)/b \text{ e pela mudança de variáveis vemos que} \end{aligned}$$

$$\begin{aligned} E(Y) &= a \int_{-b/a}^\infty (1 - F_X(y)) dy - a \int_{-\infty}^{-b/a} F_X(y) dy \\ &= a \int_0^\infty (1 - F_X(y)) dy - a \int_{-\infty}^0 F_X(y) dy + \\ &\quad + a \int_{-b/a}^0 (1 - F_X(y)) dy + a \int_{-b/a}^0 F_X(y) dy \\ &= a E(X) + a \int_{-b/a}^0 dy = a E(X) + b. \end{aligned}$$

(c) Por último, caso $a < 0$ é análogo ao item (b). ■

Exemplo 4.12

Consideremos $X \sim U(0, 1)$. Nesta situação, sabemos que $E(X) = 1/2$. Caso $Y \sim U(4, 5)$, acontece que $Y = X + 4$, do qual obtemos que

$$E(Y) = E(X + 4) = E(X) + 4 = 4 + 1/2.$$

Observe que este exemplo é uma aplicação do Teorema 4.8 quando $a = 1$ e $b = 4$.

Teorema 4.9 (*Existência*)

Seja X uma variável aleatória limitada, isto é,

$$P(|X| < M) = 1,$$

para $0 < M < \infty$. Então, $E(X)$ existe.

Demonstração: Isto é consequência direta dos Teoremas 4.6 e 4.7. Como a variável aleatória é limitada, então $-M < X < M$, logo $E(X) \leq M$. ■

Exemplo 4.13 (*Distribuição Uniforme*)

Seja $X \sim \text{Uniforme}(\alpha, \beta)$, sendo $\alpha < \beta$ duas quantidades reais. A função de densidade é

$$f(x) = \begin{cases} \frac{1}{\beta - \alpha}, & \text{se } x \in (\alpha, \beta) \\ 0, & \text{se } x \notin (\alpha, \beta) \end{cases}$$

Queremos verificar a existência de $E(X)$ e, caso afirmativo, o valor da esperança.

Logicamente,

$$P(|X| < \max\{|\alpha|, |\beta|\}) = 1.$$

Assim, denotando como $M = \max\{|\alpha|, |\beta|\}$ temos que $|X| < M$ e, portanto, $E(X)$ existe pela afirmação do Teorema 4.9. Mais ainda, um cálculo simples fornece-nos que

$$E(X) = \frac{\alpha + \beta}{2}.$$

Uma outra distribuição naturalmente limitada é quando $X \sim \text{Beta}(\alpha, \beta)$, a expressão de sua função de densidade foi apresentada na Definição 3.26, nesta situação $|X| = X < 1$. O cálculo não é simples, mas podemos afirmar que

$$E(X) = \frac{\alpha}{\alpha + \beta}.$$

Definição 4.2 (*Variável aleatória simétrica*)

Dizemos que uma variável aleatória X é simétrica a respeito do ponto α se

$$P(X \geq \alpha + x) = P(X \leq \alpha - x), \quad \forall x \in \mathbb{R}.$$

Podemos escrever esta definição em termos da função de distribuição F_X de X , isso significa que se

$$F_X(\alpha - x) = 1 - F_X(\alpha + x) + P(X = \alpha + x),$$

vale para todo $x \in \mathbb{R}$, dizemos que a função de distribuição F_X da variável aleatória X é simétrica com α o centro de simetria. Caso $\alpha = 0$, então para todo x

$$F_X(-x) = 1 - F_X(x) + P(X = x).$$

Em particular, se X é contínua dizemos que é simétrica com centro de simetria α se, e somente se, a função de densidade de X satisfizer que

$$f(\alpha - x) = f(\alpha + x), \quad \forall x \in \mathbb{R}.$$

A situação $\alpha = 0$ implica que X seja simétrica com relação à origem. Acontece que o conceito de esperança matemática está relacionado com o valor central da distribuição como veremos no Teorema 4.12. Em algumas situações as distribuições têm um valor central natural, essas são as chamadas distribuições simétricas.

Exemplo 4.14 (*Distribuição t-Student*)

Uma variável aleatória diz-se ter distribuição $t - Student(n)$ se sua função de densidade é da forma

$$f(x) = \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})\sqrt{n\pi}} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}, \quad -\infty < x < \infty. \quad (4.18)$$

Caso $n = 1$ a distribuição $t - Student$ se reduz à distribuição *Cauchy*(0, 1), isto implica que, como vimos no Exemplo 4.8 não para todo valor de n a esperança existe.

A função de densidade em (4.18) é simétrica, para vermos isso observemos que $f(x) = f(-x)$, o qual implica que a distribuição $t - Student$ é simétrica com relação à origem. Para n grande, a distribuição $t - Student$ está próxima da distribuição normal. De fato,

$$\lim_{n \rightarrow \infty} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} = e^{-x^2/2} \quad \text{e} \quad \lim_{n \rightarrow \infty} \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})\sqrt{n\pi}} = \frac{1}{\sqrt{2\pi}}.$$

Para o caso de n pequeno, entretanto, desvia consideravelmente da normal. De fato,

$$P(|X| \geq x_0) \geq P(|Z| \geq x_0), \quad Z \sim N(0, 1),$$

ou seja, para n pequeno ($n < 30$) há mais probabilidade na cauda da distribuição $t - Student$ do que na cauda da normal padrão.

Mostramos no exemplo anterior que a distribuição $t - Student$ é simétrica em torno da origem, outras distribuições conhecidas são também simétricas, como a distribuição Normal e Cauchy.

Teorema 4.10

A variável aleatória X tem distribuição simétrica a respeito do 0 se, e somente se, $F_X = F_{-X}$.

Demonstração: Consideremos X uma variável aleatória simétrica no 0. Então, pela Definição 4.2 temos que $P(X \geq x) = P(X \leq -x)$, $\forall x \in \mathbb{R}$ então

$$P(0 < X < x) = P(-x < X < 0), \quad \forall x > 0. \quad (4.19)$$

Isto implica que

$$P(0 < X < x) + P(X = x) = F_X(x) - F_X(0) \quad (4.20)$$

e

$$P(-x < X < 0) + P(X = x) = P(0 < -X \leq x) = F_{-X}(x) - F_{-X}(0). \quad (4.21)$$

Portanto, de (4.19), (4.20) e (4.21) resulta que, se X tem distribuição simétrica no zero então

$$F_X(x) - F_X(0) = F_{-X}(x) - F_{-X}(0), \quad \forall x > 0. \quad (4.22)$$

Resulta que

$$\begin{aligned} \lim_{x \rightarrow \infty} F_X(x) - F_X(0) &= \lim_{x \rightarrow \infty} F_{-X}(x) - F_{-X}(0) \\ 1 - F_X(0) &= 1 - F_{-X}(0) \\ F_X(0) &= F_{-X}(0). \end{aligned} \quad (4.23)$$

Procedendo de maneira similar caso $x < 0$ temos, dos resultados em (4.21) e (4.23), que se X é simétrica no zero, então

$$F_X(x) = F_{-X}(x), \quad \forall x \in \mathbb{R}. \quad (4.24)$$

Vamos supor agora que a relação em (4.24) é válida. Então, para todo $x \in \mathbb{R}$ vale que

$$P(X \leq x) = F_X(x) = F_{-X}(x) = P(-X \leq x) = P(X \geq x).$$

Em particular

$$P(X \leq 0) = P(X \geq 0),$$

do qual concluímos que X é simétrica a respeito do ponto 0. ■

Exemplo 4.15 (*Distribuição Exponencial dupla*)

Dizemos que a variável aleatória X tem por distribuição exponencial dupla ou Laplace se sua função de densidade é da forma

$$f(x) = \frac{1}{2}e^{-|x|}, \quad x \in \mathbb{R}.$$

A distribuição de Laplace é fácil de integrar. Sua função de distribuição é a seguinte:

$$F_X(x) = \begin{cases} \frac{1}{2}e^x, & \text{se } x \leq 0 \\ 1 - \frac{1}{2}e^{-x}, & \text{se } x \geq 0 \end{cases}.$$

Vamos utilizar o Teorema 4.10 para demonstrarmos que esta é uma distribuição

simétrica. Observemos que

$$\begin{aligned} F_{-X}(x) &= P(-X \leq x) = P(X \geq -x) \\ &= 1 - P(X \leq -x) = 1 - F_X(-x), \end{aligned}$$

a qual pode ser escrita como

$$F_X(x) = \begin{cases} 1 - \frac{1}{2}e^{-x}, & \text{se } x \geq 0 \\ \frac{1}{2}e^x, & \text{se } x \leq 0 \end{cases},$$

a qual é exatamente a mesma função. Concluimos então que a distribuição Laplace é simétrica com relação à origem.

Devemos observar que a maioria das variáveis aleatórias simétricas são absolutamente contínuas, quase nenhuma variável aleatória discreta é simétrica. Podemos mencionar como exemplos de variáveis aleatórias simétricas discretas o caso $X \sim \text{Binomial}(p)$ se assumir como valores $-x$ e x , sendo $x > 0$, ou seja,

$$P(X = -x) = 1 - p \quad \text{e} \quad P(X = x) = p.$$

Uma outra situação discreta simétrica seria a uniforme discreta, caso esta assuma um número par de valores e simétricos.

Teorema 4.11

A variável aleatória simétrica X tem centro de simetria α se, e somente se, $Y = X - \alpha$ tem centro de simetria 0.

Demonstração: Segundo a Definição 4.2, temos que $P(X \geq \alpha + x) = P(X \leq \alpha - x)$, $\forall x \in \mathbb{R}$.

Então

$$\begin{aligned} P(X \geq x + \alpha) &= P(X - \alpha \geq x) = P(Y \geq x). \\ P(X \leq \alpha - x) &= P(X - \alpha \leq -x) = P(Y \leq -x), \end{aligned}$$

então Y é simétrica com centro de simetria α . ■

Uma consequência da variável aleatória X ser simétrica com centro de simetria α é que o centro de simetria coincide com a esperança, se esta existe, resultado fornecida pelo teorema a seguir.

Teorema 4.12 (*Esperança de variáveis simétricas*)

Seja X uma variável aleatória com distribuição simétrica com esperança finita, de centro de simetria α . Então $E(X) = \alpha$.

Demonstração: Consideremos primeiro que $\alpha = 0$. Nesta situação, pelo Teorema 4.8, escolhendo $a = -1$ e $b = 0$ temos

$$E(-X) = -E(X). \quad (4.25)$$

Segundo vimos no Teorema 4.10 $F_X = F_{-X}$ e

$$\begin{aligned} E(X) &= \int_0^\infty [1 - F_X(x)] dx - \int_{-\infty}^0 F_X(x) dx \\ &= \int_0^\infty [1 - F_{-X}(x)] dx - \int_{-\infty}^0 F_{-X}(x) dx = E(-X), \end{aligned}$$

do qual concluímos que $E(X) = E(-X)$, então $E(X) = 0$. Na situação geral $\alpha \neq 0$ sabemos, pelo resultado no Teorema 4.11, que $X - \alpha$ tem distribuição simétrica em 0. Utilizando novamente o Teorema 4.8 vemos que

$$0 = E(X - \alpha) = E(X) - \alpha. \quad \blacksquare$$

Vejamos uma última propriedade, chamada de critério de integralidade. Lembremos que consideramos uma variável integrável quando tem esperança finita. Agora vamos entender quando podemos afirmar que a esperança de uma variável aleatória é finita.

Teorema 4.13 (*Critério de integrabilidade*)

Seja X uma variável aleatória qualquer. Então

$$\sum_{n=1}^{\infty} P(|X| \geq n) \leq E(|X|) \leq 1 + \sum_{n=1}^{\infty} P(|X| \geq n),$$

portanto, X é integrável se, e somente se, $\sum_{n=1}^{\infty} P(|X| \geq n) < \infty$.

Demonstração: Caso $x \geq 0$, seja $[x]$ a parte inteira de x .¹ Então, a variável aleatória $[|X|]$ satisfaz que

$$0 \leq [|X|] \leq |X| \leq 1 + [|X|].$$

Pelos Teoremas 4.7 e 4.8 obtemos que

$$0 \leq E([|X|]) \leq E(|X|) \leq 1 + E([|X|]).$$

A nova variável aleatória $[|X|]$ assume somente valores inteiros não negativos, então vale o Teorema 4.1 do qual concluímos que

$$\sum_{n=1}^{\infty} P(|X| \geq n) \leq E(|X|) \leq 1 + \sum_{n=1}^{\infty} P(|X| \geq n), \quad \blacksquare$$

¹ A função matemática parte inteira é definida como

$$[x] = \max\{n \in \mathbb{Z} : n \leq x\}.$$

Exemplo 4.16 (Continuação do Exemplo 4.2)

Na situação referida encontramos que se $X \sim \text{Exponencial}(\theta)$, então $E(X) = 1/\theta$. Vamos utilizar o critério no Teorema 4.13. Sabemos que $P(|X| \geq n) = 1 - F_X(n) = e^{-n\theta}$. Com isso, chegamos à conclusão que

$$\sum_{n=1}^{\infty} P(|X| \geq n) = \sum_{n=1}^{\infty} e^{-n\theta} = \frac{1}{e^{\theta} - 1},$$

logo

$$\frac{1}{e^{\theta} - 1} \leq \frac{1}{\theta} \leq \frac{e^{\theta}}{e^{\theta} - 1}.$$

Vamos utilizar o Teorema 4.12, que nos diz qual é a esperança de variáveis aleatórias simétricas caso existam, e o critério de integrabilidade (Teorema 4.13) para encontrarmos a esperança de distribuições simétricas.

Por exemplo, caso $X \sim \text{Laplace}$, como no Exemplo 4.15, temos que

$$\begin{aligned} P(|X| > n) &= 1 - \int_{-n}^n f_X(x) dx \\ &= 1 - \int_{-n}^0 \frac{1}{2} e^x dx - \int_0^n \frac{1}{2} e^{-x} dx \\ &= 1 - (1 - e^{-n}) = e^{-n}. \end{aligned}$$

Desta forma

$$P(|X| > n) = e^{-n}$$

e, portanto,

$$\sum_{n=1}^{\infty} P(|X| > n) = \sum_{n=1}^{\infty} e^{-n} = \frac{1}{e - 1} \approx 0.581976.$$

Provamos assim que X é integrável e, pelo Teorema 4.12, $E(X) = 0$. Vejamos agora como proceder caso a distribuição seja normal padrão.

Teorema 4.14

Seja $Z \sim \text{Normal}(0, 1)$. Então

$$P(Z > x) \approx \frac{1}{x\sqrt{2\pi}} e^{-x^2/2}, \quad \text{quando } x \rightarrow \infty. \quad (4.26)$$

Mais precisamente, para todo $x > 1/\sqrt{2}$ vale que

$$\max \left\{ 0, \frac{1}{\sqrt{2\pi}} \left(\frac{1}{x} - \frac{1}{x^3} \right) e^{-x^2/2} \right\} < P(Z > x) < \frac{1}{x\sqrt{2\pi}} e^{-x^2/2}. \quad (4.27)$$

Demonstração: Temos que

$$\frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-u^2/2} \left(1 - \frac{3}{u^4}\right) du = \frac{1}{\sqrt{2\pi}} \left(\frac{1}{x} - \frac{1}{x^3}\right) e^{-x^2/2}$$

e

$$\frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-u^2/2} \left(1 + \frac{1}{u^2}\right) du = \frac{1}{\sqrt{2\pi}} \frac{1}{x} e^{-x^2/2}.$$

Juntando estes dois resultados temos as expressões em (4.26) e (4.27). ■

Na Figura 4.2 apresentamos o comportamento dos limites da função $P(Z > x)$ encontrados no Teorema 4.14. No gráfico à direita vemos que a aproximação em (4.26) apresenta resultado satisfatório quando $x > 2$. Quanto aos limites em (4.27) vemos que, embora sempre se cumpram, somente para valores $x > 1$ o limite inferior é maior do zero.

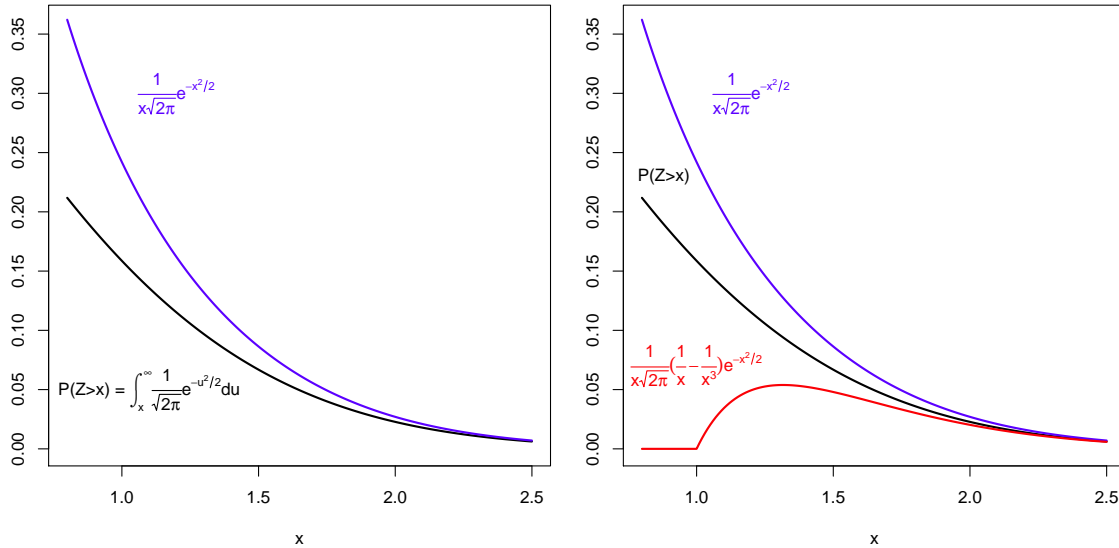


Figura 4.2: Aproximação à $P(|Z| > x)$ no gráfico à esquerda, expressão (4.26). Limites inferior e superior para $P(|Z| > x)$ a direita, expressões em (4.27).

Qual é o objetivo de dedicarmos tanto esforço a encontrarmos expressões aproximadas da distribuição normal padrão? ou formulando a pergunta de uma outra forma, qual o objetivo em encontrarmos resultados como o apresentado no Teorema 4.14?

Queremos encontrar expressões que possamos utilizar para provarmos propriedades desta distribuição. Devemos esclarecer que para a função de distribuição normal padrão não temos uma expressão matemática analítica, então como vamos demonstrar, por exemplo, que a distribuição normal padrão têm esperança finita? vamos utilizar os resultados no Teorema 4.14 e o critério de integrabilidade no Teorema 4.13 para demonstrarmos isso.

Segundo o critério de integrabilidade devemos demonstrar que

$$\sum_{n=1}^{\infty} P(|Z| \geq n) < \infty,$$

quando $Z \sim \text{Normal}(0, 1)$. Mas,

$$\sum_{n=1}^{\infty} P(|Z| > n) < \sum_{n=1}^{\infty} \frac{2}{n\sqrt{2\pi}} e^{-n^2/2}.$$

Acontece que $2/\sqrt{2\pi} = \sqrt{2/\pi} \approx 0.79788$ e que a série

$$\sum_{n=1}^{\infty} \frac{1}{n} e^{-n^2/2} \approx 0.677986$$

é convergente, de maneira que

$$\sum_{n=1}^{\infty} P(|Z| > n) < 0.540951.$$

Concluimos então que para a distribuição normal padrão vale o critério de integrabilidade, logo esta distribuição têm média finita e é zero, ou seja, $E(Z) = 0$ devido a ser simétrica. Devemos lembrar que o fato da distribuição de X ser simétrica em torno de α , não significa tenha esperança finita e que esta seja α , deve ser demonstrado a existência de esperança para podermos afirmar que $E(X) = \alpha$. O seguinte exemplo é uma outra situação de função de densidade simétrica em torno do zero e que têm esperança finita.

Exemplo 4.17

Veja agora como demonstrar que se $X \sim t\text{-Student}(n)$ então $E(X) = 0$ quando $n > 1$. Lembremos que no Exemplo 4.18 demonstramos que esta é uma distribuição simétrica, logo se $E(X) < \infty$, então esta é zero.

Provemos que a integral em (4.12) é finita nesta situação, ou seja, provemos que

$$\int_{-\infty}^{\infty} |x| \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})\sqrt{n\pi}} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} dx < \infty.$$

Seguiremos o desenvolvimento apresentado em Psarakis & Panaretos (1990).

Observemos que

$$E(|X|) = \int_0^{\infty} x C(n) \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} dx - \int_{-\infty}^0 x C(n) \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} dx,$$

onde $C(n) = \Gamma(\frac{n+1}{2}) / (\Gamma(\frac{n}{2})\sqrt{n\pi})$. Desenvolvendo cada integral, obtemos que

$$\begin{aligned} \int_0^{\infty} x C(n) \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} dx &= -\frac{n}{n-1} \int_0^{\infty} x \frac{d}{dx} \left(1 + \frac{x^2}{n}\right)^{-\frac{n-1}{2}} C(n) dx \\ &= \frac{n}{n-1} C(n) \end{aligned}$$

e

$$\begin{aligned} \int_{-\infty}^0 x C(n) \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} dx &= -\frac{n}{n-1} \int_{-\infty}^0 x \frac{d}{dx} \left(1 + \frac{x^2}{n}\right)^{-\frac{n-1}{2}} C(n) dx \\ &= -\frac{n}{n-1} C(n), \end{aligned}$$

do qual obtemos que $E(|X|) = \frac{2n}{n-1}C(n)$. Esta expressão significa que a distribuição t -Student têm esperança finita e, devido a ser uma função de densidade simétrica, podemos afirmar que a esperança é zero. Um detalhe, a distribuição da variável aleatória $|X|$, quando $X \sim t - Student(n)$ é conhecida como distribuição t -Student dobrada.

4.1.2 Esperança de funções de variáveis aleatórias

As funções $g(x) = x^n$, onde n é um inteiro positivo e $g(x) = |x|^\alpha$, onde α é um número real positivo, são de especial importância. Queremos aqui encontrarmos a esperança de funções de variáveis aleatórias e logicamente surgem perguntas: como calcular a esperança dessas transformações? a esperança de X é uma média ponderada dos valores possíveis de X , onde os pesos são determinados pela sua distribuição agora, será que a esperança de $g(X)$ continuará sendo uma média ponderada com pesos determinados pela distribuição de X ?

Teorema 4.15

Seja X uma variável aleatória e g uma função Borel mensurável nos reais e $Y = g(X)$. Então

$$E(Y) = \sum_{i=1}^{\infty} g(x_i)P(X = x_i) \quad (4.28)$$

no sentido de que, se um dos lados de (4.28) existe, o mesmo acontece com o outro, e então os dois são iguais.

Demonstração: Seja X uma variável aleatória discreta e suponhamos que $P(X \in A) = 1$. Se $y = g(x)$, onde $g : A \rightarrow B$ é uma função um-para-um, então

$$P(Y = y) = P(X = g^{-1}(y)), \quad y \in B.$$

Temos que

$$\sum_{x \in A} g(x)P(X = x) = \sum_{y \in B} yP(Y = y).$$

Caso X seja contínua e g satisfizer as condições do Teorema 3.36, temos que

$$\int g(x)f(x) dx = \int_{\alpha}^{\beta} f(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right| dy,$$

mudando de variáveis para $y = g(x)$. Então

$$\int g(x)f(x) dx = \int_{\alpha}^{\beta} yh(y) dy. \quad \blacksquare$$

Claro que esta demonstração é restrita à situação de funções um-para-um mesmo sendo o enunciado do teorema escrito para qualquer funções Borel mensurável. Para uma demonstração mais geral, incluindo o caso onde g não é uma função um-para-um, pode-se consultar o livro Loève (1977).

Exemplo 4.18

Seja $X \sim \text{Exponencial}(\theta)$. Isto implica que $P(X > x) = e^{-\theta x}$, quando $x \geq 0$ e

$$E(X^n) = n \int_0^\infty x^{n-1} e^{-\frac{x}{\theta}} dx.$$

Sabemos que $E(X) = \theta$. Então,

$$\begin{aligned} E(X^2) &= 2 \int_0^\infty x e^{-\frac{x}{\theta}} dx = 2\theta \int_0^\infty \frac{x}{\theta} e^{-\frac{x}{\theta}} dx \\ &= 2\theta \int_0^\infty x f(x) dx = 2\theta E(X) = 2\theta^2, \end{aligned}$$

e

$$\begin{aligned} E(X^3) &= 3 \int_0^\infty x^2 e^{-\frac{x}{\theta}} dx = 3\theta \int_0^\infty \frac{x^2}{\theta} e^{-\frac{x}{\theta}} dx \\ &= 3\theta \int_0^\infty x^2 f(x) dx = 3\theta E(X^2) = 6\theta^3. \end{aligned}$$

Por indução, temos que $E(X^k) = k!\theta^k$.

Acontece que nem sempre as variáveis X e Y são do mesmo tipo, ou seja, pode ser que X seja discreta e Y não ou reciprocamente. Queremos dizer que, nem sempre a variável X e sua transformada $Y = g(X)$, nem sempre são ambas discretas ou ambas contínuas; pode acontecer que, por exemplo, a transformada Y seja de natureza diferente de X . No seguinte exemplo, utilizando novamente o conceito de parte inteira de x , mostramos uma situação na qual X é contínua mas sua transformada, Y é discreta. O conceito de parte inteira de um número foi definido na nota de rodapé 1.

Exemplo 4.19

Seja X uma variável aleatória com distribuição $\text{Exponencial}(\theta)$. Definamos

$$Y = [X] + 1, \tag{4.29}$$

onde $[X]$ denota a parte inteira de X . Queremos obter a distribuição de Y .

A variável Y assume valores naturais, ou seja, assume valores $\{1, 2, \dots\}$. Observemos que, para $n \in \mathbb{N}^+$

$$P(Y = n) = P(n-1 \leq X < n)$$

e, dado que $F_X(x) = 1 - e^{-\theta x}$, temos por resultado que

$$P(Y = n) = F_X(n^-) - F_X(n-1) = e^{-\theta n}(e^\theta - 1), \tag{4.30}$$

a qual é uma progressão geométrica de soma 1. Significa que Y , definida em (4.29) é uma variável aleatória discreta com valores em \mathbb{N}^+ ou suporte \mathbb{N}^+ e com função de probabilidade dada em (4.30), $\theta > 0$.

Durante a resolução deste exemplo não dizemos, mas não é difícil perceber que a função $g(x) = [x] + 1$ é Borel mensurável.

Definição 4.3 (Momentos finitos de variáveis aleatórias discretas)

Seja X é uma variável aleatória discreta e n um número natural qualquer. Definimos a esperança da função de variável aleatória $g(X) = X^n$ como

$$E(X^n) = \sum_{i=1}^{\infty} x_i^n P(X = x_i),$$

onde x_1, x_2, \dots são os valores assumidos pela variável aleatória X e p_1, p_2, \dots as probabilidades correspondentes.

Caso $E(X^n)$ exista para algum inteiro positivo n , dizemos que $E(X^n)$ é o momento de n -ésima ordem de X ou o momento de n -ésima ordem da distribuição de X .

Exemplo 4.20 (Continuação do Exemplo 4.5)

Caso X tenha por distribuição uniforme discreta nos primeiros N números naturais, isto é, se

$$P(X = n) = \frac{1}{N}, \quad n = 1, 2, \dots, N,$$

obtivemos que

$$E(X) = \frac{N+1}{2}.$$

Encontremos $E(X^2)$.

$$E(X^2) = \sum_{n=1}^N \frac{n^2}{N} = \frac{(N+1)(2N+1)}{6}.$$

Definição 4.4 (Momentos finitos de variáveis aleatórias contínuas)

Seja X é uma variável aleatória contínua com função de densidade f e n um número natural qualquer. Definimos a esperança da função de variável aleatória $g(X) = X^n$ como

$$E(X^n) = \int_{-\infty}^{\infty} x^n f(x) dx.$$

Uma situação interessante é apresentada no exemplo a seguir, é o caso de uma variável aleatória com esperança finita mas com momento de ordem superior infinito.

Exemplo 4.21

Seja X uma variável aleatória com função de densidade $f(x) = \begin{cases} \frac{2}{x^3}, & \text{se } x \geq 1 \\ 0, & \text{se } x < 1 \end{cases}$.

Então

$$E(X) = \int_1^{\infty} \frac{2}{x^2} dx = 2.$$

Por outro lado,

$$E(X^2) = \int_1^{\infty} \frac{2}{x} dx,$$

a qual não existe. Mais ainda, todos os momentos de ordem superior, ou seja, quando $n \geq 2$ não existem.

Exemplo 4.22 (*Distribuição Uniforme*)

Consideremos que $X \sim \text{Uniforme}(0,1)$, esta é uma situação simples para obtermos todos os momentos finitos. Nesta situação

$$E(X^n) = \int_0^1 x^n dx = \frac{x^{n+1}}{n+1} \Big|_0^1 = \frac{1}{n+1}.$$

Observemos que, como consequência do resultado acima, quando $n = 1$, $E(X) = 1/2$ e que, quando $n \rightarrow \infty$ acontece que

$$\lim_{n \rightarrow \infty} E(X^n) = \lim_{n \rightarrow \infty} \frac{1}{n+1} = 0.$$

Devemos observar que, embora quando $n \rightarrow \infty$ os momentos de ordem superior desta distribuição tendem a zero, para cada n finito todos os momentos existem.

Na verdade, é possível construir exemplos de variáveis aleatórias para as quais existem todos os momentos até uma ordem especificada, mas não há momentos de ordem superior à especificada. Este é o caso da distribuição t -Student(n), para ela existem todos os momentos até ordem menor do que n mas, os momentos de ordem superior à n não existem.

Exemplo 4.23 (*Distribuição t-Student*)

Segundo o Exemplo 4.14, se $X \sim t\text{-Student}(n)$ a esperança existe e é zero. Um resultado relevante é que $E(X^r)$ existe para $r < n$. Em particular, temos que

$$E(X^r) = \begin{cases} 0, & \text{se } r < n, r \text{ par} \\ n^{r/2} \frac{\Gamma\left(\frac{r+1}{2}\right) \Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{1}{2}\right) \Gamma\left(\frac{n}{2}\right)}, & \text{se } r < n, r \text{ ímpar} \end{cases}$$

Como consequência, vemos que $E(X) = 0$, se $n > 2$.

Para a demonstração do resultado no Exemplo 4.23 utiliza-se um conceito que será estudo na Seção 6.2 e, por isso, mostraremos o cálculo no momento adequado. Mais ainda, este um exemplo para no qual momentos de ordem superior existem, mas não todos. Observe que a existência dos momentos estão limitados à $r < n$, significa que o momento de ordem n também não existe. Nada garante a existência de todos os momentos de ordem superior.

Teorema 4.16

Seja X uma variável aleatória. Suponhamos que o momento de X de ordem t existe. Então existem os momentos de X de ordem $0 < s < t$.

Demonstração: Seja X uma variável contínua com função de densidade f . Temos então

$$\begin{aligned} E(|X|^s) &= \int_{\{w \in \Omega : |x|^s \leq 1\}} |x|^s f(x) dx + \int_{\{w \in \Omega : |x|^s > 1\}} |x|^s f(x) dx \\ &\leq P(|X|^s \leq 1) + E(|X|^t) < \infty. \end{aligned}$$

Demonstração similar pode ser realizada no caso de X discreta. ■

No Exemplo 4.21 foi visto que a existência dos momentos de ordem r nada garantem acerca da existência de momentos de ordem superiores. Por outro lado, o Teorema 4.16 garante a existência dos momentos de ordem inferior. Vejamos no exemplo a seguir duas distribuições; a distribuição qui-quadrado e sua raiz quadrada, a distribuição qui. Vamos encontrar as esperanças correspondentes como exemplos da afirmação no Teorema 4.16.

Exemplo 4.24 (Distribuição qui)

Na Definição 3.25 apresentamos a função de densidade $\chi^2(n)$. Então, se $X \sim \chi^2(n)$, para encontrarmos os momentos de ordem superior desta distribuição fazemos

$$E(X^r) = \int_0^\infty \frac{x^r}{\Gamma(n/2)2^{n/2}} x^{n/2-1} e^{-x/2} dx.$$

Escrevendo $m = 2r + n$, obtemos

$$E(X^r) = \frac{\Gamma(m/2)}{\Gamma(n/2)} 2^r \int_0^\infty \frac{1}{\Gamma(m/2)2^{m/2}} x^{m/2-1} e^{-x/2} dx = \frac{\Gamma(m/2)}{\Gamma(n/2)} 2^r. \quad (4.31)$$

Temos, portanto, uma expressão para os momentos de ordem superior caso a distribuição seja $\chi^2(n)$. Por outro lado, o Teorema 4.16 trata acerca de momentos de ordem inferior.

Por exemplo, $E(\sqrt{X})$, a qual sabemos existe dado que $E(X) = \frac{\Gamma((n+2)/2)}{\Gamma(n/2)} 2$.

Acontece que, a raiz quadrada de uma variável aleatória com distribuição $\chi^2(n)$ é uma variável aleatória com distribuição $\chi(n)$, chamada de distribuição qui. Para encontrarmos a função de densidade qui utilizaremos o Teorema 3.36. A transformação $Y = g(X) = \sqrt{X}$ é uma transformação um-a-um com inversa $X = g^{-1}(Y) = Y^2$ e Jacobiano $dg^{-1}(y)/dy = 2y$. Portanto, pela técnica de transformação, a função densidade de Y é

$$\begin{aligned} f(y) &= f_X(g^{-1}(y)) \left| \frac{dx}{dy} \right| = \frac{1}{2^{n/2} \Gamma(n/2)} (y^2)^{n/2-1} e^{-y^2/2} |2y| \\ &= \frac{1}{2^{n/2-1} \Gamma(n/2)} y^{n-1} e^{-y^2/2}, \quad y > 0. \end{aligned}$$

Similarmente ao desenvolvimento em (4.31), obtemos que $E(Y) = \sqrt{2} \frac{\Gamma((n+1)/2)}{\Gamma(n/2)}$.

Teorema 4.17

Seja X uma variável aleatória e $E(|X|^k) < \infty$, para algum $k > 0$. Então

$$\lim_{n \rightarrow \infty} n^k P(|X| > n) = 0.$$

Demonstração: Faremos a demonstração somente para o caso contínuo, com função de densidade f . Temos que

$$\lim_{n \rightarrow \infty} \int_{|x| \leq n} |x|^k f(x) dx = \int |x|^k f(x) dx < \infty.$$

Segue que

$$\lim_{n \rightarrow \infty} \int_{|x| > n} |x|^k f(x) dx = 0.$$

Mas,

$$n^k P(|X| > n) \leq \int_{|x| > n} |x|^k f(x) dx. \quad \blacksquare$$

Probabilidades do tipo $P(|X| > n)$ ou qualquer um de seus componentes, $P(X > n)$ ou $P(X < -n)$, são chamadas probabilidades de cauda. O resultado do Teorema 4.17, portanto, fornece a taxa na qual $P(|X| > n)$ converge para 0 quando $n \rightarrow \infty$. O inverso do Teorema 4.17 não se sustenta em geral, isto é, caso

$$\lim_{n \rightarrow \infty} n^k P(|X| > n) = 0,$$

para algum k não implica necessariamente que $E(|X|^k) < \infty$.

Exemplo 4.25

Vamos considerar uma variável aleatória discreta com função de probabilidade dada por

$$P(X = n) = \frac{c}{n^2 \log(n)}, \quad n = 2, 3, \dots,$$

onde c é uma constante determinada pela restrição

$$\sum_{n=2}^{\infty} \frac{c}{n^2 \log(n)} = 1.$$

Queremos provar que o inverso do Teorema 4.17 pode não ser verdade. Para isso, observamos que

$$P(X > n) \approx c \int_n^{\infty} \frac{c}{x^2 \log(x)} dx \approx \frac{c}{n \log(n)} \quad (4.32)$$

e, então, $\lim_{n \rightarrow \infty} nP(X > n) = 0$ e ainda $E(X) = \sum_{n=1}^{\infty} \frac{c}{n \log(n)} = \infty$.

Na expressão em (4.32), com o símbolo \approx , queremos dizer que o quociente dos dois lados converge a 1 quando $n \rightarrow \infty$. Observamos que precisamos

$$\lim_{n \rightarrow \infty} n^{k+\delta} P(|X| > n) = 0,$$

para algum $\delta > 0$ a fim de garantir que $E(|X|^k) < \infty$. Essa é chamada de condição de momentos. Podemos então afirmar que para uma variável aleatória X , $E(|X|) < \infty$ se, e somente se, as integrais

$$\int_{-\infty}^0 P(X \leq x) dx \quad \text{e} \quad \int_0^{\infty} P(X > x) dx$$

forem ambas convergentes e, nesse caso

$$E(X) = \int_0^{\infty} P(X > x) dx - \int_{-\infty}^0 P(X \leq x) dx.$$

Teorema 4.18

Seja X uma variável aleatória com momento absoluto de ordem $\alpha > 0$ finito. Então

$$E(|X|^\alpha) = \int_0^{\infty} P(|X|^\alpha > x) dx = \alpha \int_0^{\infty} x^{\alpha-1} P(|X| > x) dx.$$

Demonstração : Mudança de variáveis. ■

Com este resultado podemos afirmar que uma variável aleatória possui momento absoluto de ordem $\alpha > 0$ se, e somente se,

$$|x|^{\alpha-1} P(|X| > x)$$

é integrável no intervalo $(0, \infty)$. Mais ainda, podemos afirmar que

$$E(|X|^\alpha) < \infty \quad \text{se, e somente se,} \quad \sum_{n=1}^{\infty} P(|X| > n^{1/\alpha}) < \infty.$$

Teorema 4.19

Seja X uma variável aleatória com função de distribuição satisfazendo que

$$\lim_{n \rightarrow \infty} n^\alpha P(|X| > n) = 0, \tag{4.33}$$

para algum $\alpha > 0$. Então, $E(|X|^\beta) < \infty$, para algum $0 < \beta < \alpha$.

Demonstração: Dado que o limite em (4.33) é zero então, para qualquer $\epsilon > 0$, podemos escolher $N = N(\epsilon)$, de maneira que

$$P(|X| > n) < \frac{\epsilon}{n^\alpha}, \quad \forall n \geq N.$$

Segue que, para $0 < \beta < \alpha$

$$\begin{aligned} E(|X|^\beta) &= \beta \int_0^N x^{\beta-1} P(|X| > x) dx + \beta \int_N^\infty x^{\beta-1} P(|X| > x) dx \\ &\leq N^\beta + \beta \epsilon \int_N^\infty x^{\beta-\alpha-1} dx < \infty. \end{aligned}$$

■

Utilizando os resultados dos Teoremas 4.17 e 4.19 vamos mostrar no seguinte exemplo que existem variáveis aleatórias para quais momentos de qualquer ordem não existem, ou seja, vamos mostrar que existe uma variável aleatória satisfazendo que

$$\lim_{n \rightarrow \infty} n^\alpha P(|X| > n) = 0$$

para todo $\alpha > 0$ mas, para a qual

$$E(|X|^\alpha) = \infty, \quad \forall \alpha > 0.$$

Exemplo 4.26

Seja X uma variável aleatória com função de densidade

$$f(x) = \begin{cases} \frac{1}{2|x| \log^2(|X|)}, & \text{caso } |x| > e \\ 0, & \text{caso contrário} \end{cases}.$$

A função de distribuição de X é dada por

$$F_X(x) = \begin{cases} \frac{1}{2 \log(|x|)}, & \text{caso } x < -e \\ \frac{1}{2}, & \text{caso } -e < x < e \\ 1 - \frac{1}{2 \log(x)}, & \text{caso } x \geq e \end{cases}.$$

Então, para $x > e$, temos que

$$P(|X| > x) = 1 - F_X(x) + F_X(x^-) = \frac{1}{2 \log(x)},$$

e

$$\lim_{x \rightarrow \infty} x^\alpha P(|X| > x) = \infty, \quad \forall \alpha > 0.$$

Segue então que $E(|X|) = \infty$, qualquer seja $\alpha > 0$.

Na próxima propriedade vamos utilizar funções convexas. Uma função g é dita convexa se podemos afirmar que, para quaisquer x e y pertencentes ao domínio e $\forall \alpha \in [0, 1]$, tem-se:

$$g(\alpha x + (1 - \alpha)y) \leq \alpha g(x) + (1 - \alpha)g(y).$$

Caso g seja duas vezes diferenciável e $g''(x) \geq 0$ para todo x , então g é convexa. Pode-se demonstrar que se g é convexa, então g encontra-se acima de qualquer linha que atinge g em algum ponto, chamada de linha tangente.

Caso a função g seja convexa, temos que $-g$ é côncava. Muitas funções bem conhecidas são convexas, como exemplos podemos mencionar $g(x) = x^2$, $g(x) = |x|$ e $g(x) = e^x$. Caso a função seja côncava, além dos exemplos evidentes, a função $g(x) = \log(x)$ é um exemplo típico de função côncava.

Teorema 4.20 (*Desigualdade de Jensen*)

Seja g uma função real convexa. Se a variável aleatória X têm média finita, então

$$g(E(X)) \leq E(g(X)). \quad (4.34)$$

Demonstração: Seja $l(x) = a + bx$ a linha tangente a g no ponto $E(X)$. Dado que g é convexa, fica acima da linha l . Assim,

$$E(g(X)) \geq E(l(X)) = E(a + bX) = a + bE(X) = l(E(X)) = g(E(X)). \quad \blacksquare$$

Pode-se utilizar a desigualdade de Jensen em diversas situações. Por exemplo, se $g(x) = |x|$, como consequência da afirmação em (4.34) temos que

$$|E(X)| \leq E(|X|).$$

Mais ainda, caso $g(x) = x^2$ obtemos

$$(E(X))^2 \leq E(X^2),$$

da qual também deduzimos que, escolhendo $g(x) = |x|^p$ onde $p \geq 1$ temos

$$|E(X)|^p \leq E(|X|^p).$$

Como última situação particular de função convexa consideramos $g(x) = 1/x$, da qual concluímos que

$$\frac{1}{E(X)} \leq E\left(\frac{1}{X}\right).$$

Foi mencionado que caso g fosse uma função convexa então $-g$ é côncava e, nesta situação, vale a desigualdade de Jensen contrária, isto é

$$g(E(X)) \geq E(g(X)),$$

para função côncavas. Um exemplo típico desta situação acontece quando $g(x) = \log(x)$ e, portanto

$$E(\log(X)) \leq \log(E(X)).$$

4.1.3 Percentis

Vimos que, para algumas distribuições, a média não existe. Consideramos a seguir outras características, chamadas de percentis, os quais sempre existem.

Exemplo 4.27 (*Apostas Justas*)

Suponha que X seja a quantidade de chuva que cairá amanhã e que X tenha função de distribuição F_X . Suponha ainda que queremos colocar uma aposta de dinheiro igual em X da seguinte forma: Se $X \leq x_0$, ganhamos um Real e se $X > x_0$ perdemos um Real. Para tornar esta aposta justa, precisamos que

$$P(X \leq x_0) = P(X > x_0) = 1/2.$$

Podemos pesquisar através de todos os números reais x tentando encontrar um tal que $F_X(x) = 1/2$, e então vamos considerar x_0 igual ao valor que encontramos. Se F_X é uma função um-para-um, então F_X tem inversa e

$$x_0 = F_X^{-1}(1/2).$$

O valor x_0 que procuramos no Exemplo 4.27 é chamado de quantil 0.5 de X ou do 50% percentil de X , porque 50% da distribuição de X é igual ou inferior a x_0 .

Definição 4.5 (*Percentil*)

O número x satisfazendo que

$$P(X \leq x) \geq p \quad \text{e} \quad P(X \geq x) \geq 1 - p, \quad 0 < p < 1,$$

é chamado de percentil de ordem p ou $100p$ -ésimo percentil para a variável aleatória X ou percentil da distribuição F_X de X . Escrevemos que δ_p é o percentil de ordem p da variável aleatória X .

Exemplo 4.28

Seja $X \sim \text{Exponencial}(\theta)$. Então, a função de distribuição de X é

$$F_X(x) = 1 - e^{-\frac{x}{\theta}}, \quad x \geq 0.$$

Para encontrarmos o percentil desta distribuição fazemos $1 - e^{-\frac{x}{\theta}} \geq p$ e $e^{-\frac{x}{\theta}} \geq 1 - p$, do qual temos que $e^{-\frac{x}{\theta}} = 1 - p$ e obtemos que a expressão do percentil da distribuição Exponencial é

$$\delta_p = -\theta \ln(1 - p).$$

No próximo exemplo utilizaremos um conceito empregado nas apólices ou contratos de seguros, conhecido como valor em risco².

²Segundo a SUSEP, Valor em Risco (VR) é o valor total de reposição dos bens segurados imediatamente

Exemplo 4.29 (Valor em risco)

Valor em risco. O gerente de uma carteira de investimentos está interessado em quanto dinheiro a carteira pode perder em um horizonte de tempo fixo. Seja X a mudança no valor da carteira dada durante um período de um mês. Suponha que X seja uma variável aleatória. O gerente calcula uma quantidade conhecida no mundo da gestão de risco como valor em risco, denotado por VR . Para ser específico, consideremos $Y = -X$ representar a perda incorrida pela carteira ao longo de um mês. O gerente quer ter um nível de confiança sobre o tamanho do Y . Neste exemplo, o gerente especifica um nível de probabilidade como 0.99 e, em seguida, localiza y_0 , o quantil 0.99 de Y .

O gerente agora tem 99% de certeza de que $Y \leq y_0$ e y_0 é chamado de valor em risco ou VR . Se X tem uma distribuição contínua, então é fácil ver que y_0 está intimamente relacionado ao quantil 0.01 da distribuição de X . O quantil x_0 de 0.01 tem a propriedade que

$$P(X < x_0) = 0.01.$$

Mas

$$P(X < x_0) = P(Y > -x_0) = 1 - P(Y \leq -x_0).$$

Portanto, $-x_0$ é um quantil de 0.99 de Y .

No exemplo utilizamos o conceito de quantil. Quantis são pontos estabelecidos em intervalos regulares a partir da função distribuição, de uma variável aleatória. Os quantis dividem os dados ordenados em q subconjuntos de dados de dimensão essencialmente igual dando origem a q -quantis.

De outra forma, o k -ésimo q -quantil é o valor x tal que a probabilidade da variável aleatória ser inferior a x é, no máximo, k/q e a probabilidade da variável aleatória ser superior ou igual a x é, pelo menos, $(q - k)/q$. Temos então $q - 1$ quantis, sendo k um inteiro satisfazendo $0 < k < q$.

Mais especificamente ainda, quantis são alguns dos percentis assim temos que $\delta_{0.01}$ é o quantil de 0.01. Se, em vez de usar inteiros k e q , o p -quantil é baseado em um número real p com $0 < p < 1$, então, p substitui k/q nas fórmulas acima e concluímos que percentil e quantil são dois nomes para o mesmo conceito.

Acontece que, se δ_p é o percentil de ordem p da variável aleatória X , com função de distribuição F_X , temos que

$$p \leq F_X(\delta_p) \leq p + P(X = \delta_p).$$

Observemos ainda que, se $P(X = \delta_p) = 0$, como no caso de variáveis aleatórias contínuas, o percentil de ordem p é solução da equação

$$F_X(\delta_p) = p. \quad (4.35)$$

Se a função de distribuição F_X é estritamente crescente a equação em (4.35) tem solução única, como a situação apresentada no Exemplo 4.28. Caso contrário, pode haver várias soluções, inclusive um número não enumerável, uma de tais soluções será chamada de percentil de ordem p .

antes da ocorrência do sinistro. A SUSEP é o órgão responsável pelo controle e fiscalização dos mercados de seguro, previdência privada aberta, capitalização e resseguro. Autarquia vinculada ao Ministério da Fazenda.

Exemplo 4.30

Seja $X \sim \text{Bernoulli}(p)$. Utilizando a expressão da função de probabilidade mostrada em (3.11), a função de distribuição correspondente é

$$F_X(x) = \begin{cases} 0, & \text{se } x < x_1 \\ p, & \text{se } x_1 \leq x < x_2 \\ 1, & \text{se } x_2 \leq x \end{cases}.$$

Vejamos agora como obter os percentis. Acontece que somente podemos definir percentis de ordem p e qualquer seja o ponto entre x_1 e x_2 satisfaz que $F_X(\delta_p) = p$.

No seguinte Exemplo 4.32 utilizaremos que se c é uma constante qualquer a distribuição da variável aleatória cX continua sendo Uniforme contínua no intervalo $[ac, bc]$. Resulta que, se $X \sim U(a, b)$, ou seja, se $f(x) = \frac{1}{b-a}$ para $a \leq x \leq b$ e zero em caso contrário temos que

$$\begin{aligned} P(cX \leq x) &= P\left(X \leq \frac{x}{c}\right) = F_X\left(\frac{x}{c}\right) \\ &= \frac{\frac{x}{c} - a}{b - a} = \frac{x - ca}{cb - ca}, \quad \text{para } x \in [ca, cb]. \end{aligned}$$

Exemplo 4.31

No Exemplo 3.22 foi encontrado que, caso a distribuição de X seja Uniforme contínua, então

$$F_X(x) = \frac{x - a}{b - a}, \quad x \in [a, b].$$

Isto implica que o percentil de ordem p desta distribuição é da forma

$$\delta_p = pb + (1 - p)a,$$

sendo esta expressão obtida como solução da equação em (4.35).

Vejamos um exemplo da utilidade dos percentis, também associado a risco de investimentos, como no Exemplo 4.29.

Exemplo 4.32 (Investidor)

Um investidor está tentando escolher entre duas ações possíveis para comprar um investimento de três meses. Uma ação custa R\$50,00 por ação e tem uma taxa de retorno de R_1 reais por ação para o período de três meses, em que R_1 é uma variável aleatória. O segundo estoque custa R\$30,00 por ação e tem uma taxa de retorno de R_2 por ação para o mesmo período de três meses. O investidor tem um total de R\$6.000,00 para investir. Para este exemplo, suponha que o investidor compre ações de apenas uma ação. Suponha que R_1 tem a distribuição uniforme contínua no intervalo $[-10, 20]$ e que R_2 tem a distribuição uniforme contínua no intervalo $[-4, 5, 10]$. Devemos primeiro calcular o valor esperado em reais de investir em cada uma das duas ações. Para o primeiro estoque, os R\$6.000,00 serão aplicados para comprar 120 ações, então o retorno será $120 \times R_1$, cuja

esperança é $120 E(R_1) = R\$600,00$. Para o segundo estoque, os $R\$6.000,00$ vai comprar 200 ações, então o retorno será $200 E(R_2) = R\$550,00$.

O primeiro estoque tem um retorno esperado maior. Além de calcular o retorno esperado, devemos também perguntar qual dos dois investimentos é mais arriscado. Vamos agora calcular o valor em risco (VR) a uma probabilidade de 0.97 para cada investimento. VR será o negativo do quantil $1 - 0.97 = 0.03$ para o retorno de cada investimento. Para o primeiro estoque, o retorno $120 \times R_1$ tem a distribuição uniforme contínua no intervalo $[-1.200, 2.400]$, cujo quantil 0.03 é $0.03 \times 2.400 + 0,97 \times (-1.200) = -1.092$. Então $VR = 1.092$. Para o segundo estoque, o retorno $200 \times R_2$ tem a distribuição uniforme contínua no intervalo $[-900, 2.000]$ cujo quantil 0.03 é $0.03 \times 2.000 + 0.97 \times (-900) = -813$, segundo o resultado obtido no Exemplo 4.31. Então $VR = 813$. Até embora o primeiro estoque tenha maior retorno esperado, o segundo estoque parece ser ligeiramente menos arriscado em termos de VR. Como devemos equilibrar risco e retorno esperado e escolher entre as duas compras? Uma maneira de responder a essa pergunta é ilustrada no Exemplo 4.37, depois de aprendermos sobre utilidade.

Uma desvantagem da esperança é ser instável, isto é, é muito sensível à pequenas perturbações, pequenas modificações na distribuição da variável se refletem em importantes mudanças nos valores da esperança. Outra desvantagem da esperança é que nem sempre existe. Inclusive isto pode ocorrer em distribuições simétricas. Portanto, a simetria não garante a existência da esperança. Neste sentido, não é uma boa medida de centralidade, dado que qualquer medida de centralidade deveria coincidir com o centro de simetria da função de densidade f , no caso deste existir.

Para entendermos o porquê pequenas perturbações na distribuição de uma variável aleatória são refletidas na esperança consideremos, no seguinte exemplo, o caso de uma variável aleatória Y com distribuição Uniforme modificada.

Exemplo 4.33

Seja $X \sim \text{Uniforme}(0, 1)$, com função de densidade segundo a Definição 3.22. Faremos agora uma pequena modificação na função de densidade $\text{Uniforme}(0, 1)$ da seguinte forma

$$f(y) = \begin{cases} 1, & \text{se } 0 \leq y < 0.9 \\ 0.8, & \text{se } 0.9 \leq y < 0.95 \\ 1.2, & \text{se } 0.95 \leq y \leq 1 \end{cases} \quad (4.36)$$

Diremos que Y é uma variável aleatória Uniforme modificada se tiver como função de densidade a explicitada em (4.36). Pode-se verificar que a expressão acima realmente define uma função de densidade.

Não é muito difícil perceber que $E(X) = 1/2$ enquanto $E(Y) = 0.412$.

Para eliminarmos, se possível, esta deficiência surge a ideia de definirmos outras medidas de centralidade, por exemplo, a mediana. Se existe um valor que tenha a mesma probabilidade à sua direita do que a esquerda, esse valor é a mediana. Esse número poder ser obtido sempre no caso de uma variável aleatória contínua. Se X é simétrica então a mediana coincide com o centro de simetria. Uma definição geral da mediana é a seguinte.

Definição 4.6 (*Mediana*)

Dizemos que m é uma mediana da variável aleatória X , se satisfaz que

$$P(X \geq m) \geq \frac{1}{2} \quad \text{e que} \quad P(X \leq m) \geq \frac{1}{2}.$$

Observemos que, caso X seja discreta, então a mediana m deve satisfazer que

$$\frac{1}{2} \leq P(X \leq m) \leq \frac{1}{2} + P(X = m).$$

Exemplo 4.34

A densidade *Cauchy*(0, 1) é dada por

$$f(x) = \frac{1}{\pi} \frac{1}{1 + x^2}, \text{ para } x \in \mathbb{R}.$$

Esta função é par e por isso simétrica. Esta distribuição não tem esperança, provaremos isso fazendo $y = 1 + x^2$, obtendo $dy = 2x dx$, logo

$$\int_0^\infty \frac{1}{\pi} \frac{x}{1 + x^2} dx = \frac{1}{2\pi} \int_1^\infty \frac{1}{y} dy = \frac{1}{2\pi} \log(y) \Big|_1^\infty = +\infty.$$

Agora, a mediana desta variável aleatória é $m = 0$.

Acontece que a mediana sempre existe e, se não é única, o conjunto das medianas é conexo, quer dizer, o conjunto de todas as medianas de uma variável aleatória é um intervalo em \mathbb{R} . Para demonstrarmos isto, primeiro provamos o seguinte resultado.

Teorema 4.21

Seja X uma variável aleatória com função de distribuição F_X . Então,

$$P\left(X \geq F_X^{-1}(y)\right) \geq 1 - y.$$

Demonstração : Definimos o conjunto

$$F^{-1}(y) = \inf\{x : F_X(x) \geq y\}. \quad (4.37)$$

Seja $x < F_X^{-1}(y)$, então como $F_X^{-1}(y)$ é mínimo do conjunto em (4.37), temos que $F_X(x) < y$. Vamos assumir, sem perda de generalidade, que

$$x = F_X^{-1}(y) - \frac{1}{n}$$

a qual satisfaz que $x < F_X^{-1}(y)$. Dessa maneira temos que

$$F_X^{-1}(x) = F_X^{-1}\left(F_X^{-1}(y) - \frac{1}{n}\right) < y,$$

o qual significa que $P(X \leq F_X^{-1}(y) - \frac{1}{n}) < y$. Definamos os eventos

$$A_n = \left\{ \omega \in \Omega : X(\omega) \leq F_X^{-1}(y) - \frac{1}{n} \right\}.$$

A sequência de eventos $\{A_n\}$ é não decrescente e satisfaz que

$$\bigcup_{n=1}^{\infty} A_n = \{\omega \in \Omega : X(\omega) < F_X^{-1}(y)\}.$$

Portanto

$$\lim_{n \rightarrow \infty} P\left(X \leq F_X^{-1}(y) - \frac{1}{n}\right) = P\left(X < F_X^{-1}(y)\right) < y,$$

do qual concluímos que $P(X \geq F_X^{-1}(y)) \geq 1 - y$. ■

A mediana é especial. A mediana de uma distribuição é uma das várias características que as pessoas gostam de usar ao sumarizar a distribuição de uma variável aleatória. Como a mediana é um resumo tão popular, precisamos observar que há várias definições diferentes, porém semelhantes, de mediana. Lembre-se de que o quantil $1/2$ é o menor número x tal que $F_X(x) \leq 1/2$. Para algumas distribuições, geralmente distribuições discretas, haverá um intervalo de números $[x_1, x_2)$ tal que para todo $x \in (x_1, x_2)$, $F_X(x) = 1/2$. Nesses casos, é comum referir para todos esses x (incluindo x_2) como medianas da distribuição. Outra convenção popular é chamar $(x_1 + x_2)/2$ a mediana. Esta última é provavelmente a convenção mais comum. Os leitores devem estar cientes de que, sempre que encontrarem uma mediana, pode ser qualquer uma das coisas que acabamos de discutir. Felizmente, todos eles significam quase a mesma coisa, ou seja, que o número divide a distribuição pela metade, tanto quanto possível.

Agora vamos demonstrar o grande resultado desta subseção, com o seguinte teorema garantimos que a mediana sempre existe e dizemos como encontrar-la.

Teorema 4.22 (*Propriedades da mediana*)

Seja X uma variável aleatória com função de distribuição F_X . Então

- (a) $F_X^{-1}(\frac{1}{2})$, é uma mediana.
- (b) Se m é uma mediana de X , então $F_X^{-1}(\frac{1}{2}) \leq m$.
- (c) Caso m_a e m_b sejam medianas de X , então todo $m \in (m_a, m_b)$, é também mediana de X .

Demonstração: (a) Pelo Teorema 4.21 sabemos que $P(X \geq F_X^{-1}(\frac{1}{2})) \geq \frac{1}{2}$. Baste agora escolher $y = 1/2$ e substituir em (4.37).

(b) Caso m^* seja uma outra mediana temos que $P(X \leq m^*) \geq 1/2$, do qual resulta que

$$m^* \in \{x : F_X(x) \geq 1/2\}.$$

Devido a que $F_X^{-1}(1/2) = \inf\{x : F_X(x) \geq 1/2\}$, resulta que $F_X^{-1}(1/2) \leq m^*$.

(c) Seja $m \in (m_a, m_b)$. Então, $P(X \geq m_a) \geq P(X \geq m) \geq P(X \geq m_b) \geq 1/2$. Por outro lado, $1/2 \leq P(X \leq m_a) \leq P(X \leq m) \leq P(X \leq m_b)$. ■

Exemplo 4.35

Seja X uma variável aleatória discreta com função de probabilidade

$$P(X = -2) = P(X = 0) = \frac{1}{4}, \quad P(X = 1) = \frac{1}{3}, \quad P(X = 2) = \frac{1}{6}.$$

Então

$$P(X \leq 0) = \frac{1}{2} \quad e \quad P(X \geq 0) = \frac{3}{4} > \frac{1}{2}.$$

De fato, se x é qualquer número tal que $0 < x < 1$, temos que

$$P(X \leq x) = P(X = -2) + P(X = 0) = \frac{1}{2}$$

e

$$P(X \geq x) = P(X = 1) + P(X = 2) = \frac{1}{2}.$$

Concluimos então que qualquer x , $0 \leq x < 1$, é mediana da variável aleatória X . Utilizando a relação em (4.37) escolhemos como mediana o ponto 0.

Teorema 4.23

Seja X uma variável aleatória com esperança finita e seja m uma mediana. Então m minimiza $E(|X - c|)$, $c \in \mathbb{R}$, isto é,

$$E(|X - m|) = \min_{c \in \mathbb{R}} E(|X - c|).$$

Demonstração: Vamos supor que $m < c$, a demonstração para o caso contrário $c < m$ é análoga. Desejamos provar que $E(|X - c|) \geq E(|X - m|)$. Escolhemos então $\lambda = c - m > 0$ e

- se $x \leq m$, então $|x - c| = |x - m| + \lambda$;
- se $x > m$, então $|x - c| \geq |x - m| - \lambda$.

Disto obtemos que

- caso $X \leq m$, então $|X - c| - |X - m| = \lambda$;
- caso $X > m$, então $|X - c| - |X - m| \geq -\lambda$.

Como m é mediana, $P(X \leq m) \geq 1/2$ e $P(X > m) = 1 - P(X \leq m) \leq 1/2$. Então, a variável aleatória $Y = |X - c| - |X - m|$ tem esperança não negativa, pois assume $\lambda > 0$ com probabilidade maior do que $1/2$ e assume valores maiores do que $-\lambda$ com probabilidade menor do que $1/2$. Podemos dizer que

$$Y \geq \lambda \mathbf{1}_{[X \leq m]} - \lambda \mathbf{1}_{[X > m]}$$

do qual obtemos que

$$\begin{aligned} E(Y) &\geq \lambda E(\mathbf{1}_{[X \leq m]}) - \lambda E(\mathbf{1}_{[X > m]}) \\ &\geq \lambda P(X \leq m) - \lambda P(X > m) = \lambda(P(X \leq m) - P(X > m)) \geq \lambda\left(\frac{1}{2} - \frac{1}{2}\right) = 0. \end{aligned}$$

Concluimos que $E(Y) \geq 0$, logo $E(|X - c|) \geq E(|X - m|)$. ■

4.2 Variância

Um valor central é importante porém pouco informa se não consideramos alguma medida de quão dispersos estiverem os valores da variável ao redor dele. A utilidade do valor esperado, como uma previsão para o resultado de um experimento, é maior quando o resultado não é provável que se afaste demasiado do valor esperado. Nesta seção, vamos introduzir uma medida deste desvio, chamado variância. Antes, vejamos algumas relações.

Se $E(X^n)$ existe para algum inteiro positivo n , chamaremos $E(X^n)$ o n -ésimo momento de X na origem. Utilizaremos a seguinte notação

$$\mu_n = E(X^n).$$

Definição 4.7

Sejam k um número inteiro positivo e c uma constante. Se a esperança $E(X - c)^k$ existe o chamaremos de momento de ordem k no ponto c . Escolhendo $c = E(X) = \mu$, chamaremos $E(X - \mu)^k$ o momento central do ordem k ou momento de ordem k na média. Escrevemos

$$m_k = E(X - \mu)^k. \quad (4.38)$$

Se conhecermos os coeficientes m_1, \dots, m_k podemos calcular μ_1, \dots, μ_k e reciprocamente. Temos que

$$m_k = E(X - \mu)^k = \mu_k - \binom{k}{1} m \mu_{k-1} + \binom{k}{2} m^2 \mu_{k-2} - \dots + (-1)^k m^k$$

e

$$\mu_k = E(X - \mu + \mu)^k = m_k + \binom{k}{1} m \mu_{k-1} + \binom{k}{2} m^2 \mu_{k-2} + \dots + m^k.$$

O caso $k = 2$ na Definição 4.7 é de especial importância, é a chamada variância. Então, a variância de uma variável aleatória é uma medida do grau de dispersão dos diferentes valores assumidos pela variável. Sua definição é a seguinte.

Definição 4.8 (*Variância*)

A variância de uma variável aleatória X , denotada por $\text{Var}(X)$, define-se como a esperança a seguir, caso exista,

$$\text{Var}(X) = E\left((X - E(X))^2\right). \quad (4.39)$$

Observemos que se $E(X^2)$ existe, a variância é momento de ordem 2 na média. Quando X é discreta com função de probabilidade $P(X)$ e esperança finita μ a variância de X , quando existe, se calcula como

$$\text{Var}(X) = \sum_x (x - \mu)^2 P(X = x).$$

Se X é absolutamente contínua com função de densidade $f(x)$ e esperança finita μ , então a variância de X , quando existe, é calculada como

$$\text{Var}(X) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx.$$

Observemos que, para qualquer variável aleatória X , temos que $\text{Var}(X) \geq 0$. A variância denota-se regularmente pelo símbolo σ^2 . À raiz quadrada positiva de $\text{Var}(X)$ se le conhece como desvio padrão e se denota por σ . Novamente há situações nas quais a variância não é finita, e nestas situações se disse que a variável aleatória não tem variância. Observe que para calcular a variância é necessário conhecer primeiro a esperança. Uma observação também é que, em geral,

$$\text{Var}(X + Y) \neq \text{Var}(X) + \text{Var}(Y).$$

Para ver isto, podemos escolher $Y = X$, com $\text{Var}(X) \neq 0$ e verificamos a não igualdade.

Teorema 4.24

Seja X uma variável aleatória tal que $E(X^2)$ seja finita. Então

$$\text{Var}(X) = E(X^2) - E^2(X).$$

Demonstração: Observemos que

$$\text{Var}(X) = E([X - E(X)]^2) = E[X^2 - 2XE(X) + E^2(X)] = E(X^2) - E^2(X).$$

■

Exemplo 4.36 (*Distribuição Bernoulli*)

Seja X uma variável aleatória com função de probabilidade Bernoulli(θ). Encontremos $\text{Var}(X)$. Sabemos que $E(X) = \theta$, então devemos calcular $E(X^2)$.

$$E(X^2) = 0^2P(X = 0) + 1^2P(X = 1) = \theta,$$

logo $E(X^2) = \theta$ e, portanto, $\text{Var}(X) = \theta - \theta^2 = \theta(1 - \theta)$.

Exemplo 4.37 (*Investimento*)

No Exemplo 4.32 comparamos duas possíveis compras de ações com base em seus retornos esperados e valor em risco, VR. Suponha que o investidor tenha uma função de utilidade não linear para reais. Para ser específico, suponha que a utilidade de um retorno de x seja uma função da forma

$$U(x) = \begin{cases} x^\alpha, & \text{para } x \geq 0 \\ x, & \text{para } x \leq 0 \end{cases}.$$

Podemos calcular a utilidade esperada do retorno de cada uma das duas possíveis compras de ações no Exemplo 4.32 para decidir qual é mais favorável. Se R é o retorno por ação e nós compramos s ações, então o retorno é $X = sR$ e a utilidade esperada do retorno é

$$E(U(sR)) = \int_{-\infty}^0 sr f(r) dr + \int_0^{\infty} (sr)^\alpha f(r) dr, \quad (4.40)$$

onde f é a função de densidade de R e selecionamos $\alpha = 0.8$. Para o primeiro estoque, o retorno por ação é R_1 com distribuição uniforme no intervalo $[-10, 20]$ e o número de ações seria $s_1 = 120$. Isto faz (4.40) igual a

$$E(U(120R_1)) = \int_{-10}^0 \frac{120r}{30} dr + \int_0^{20} \frac{(120r)^{0.8}}{30} dr = 12.6.$$

Para o segundo estoque, o retorno por ação é R_2 com distribuição uniforme no intervalo $[-4, 5, 10]$ e o número de ações seria $s_2 = 200$. Isto faz (4.40) igual a

$$E(U(200R_2)) = \int_{-4.5}^0 \frac{200r}{14.5} dr + \int_0^{10} \frac{(200r)^{0.8}}{14.5} dr = 27.9.$$

Com a função de utilidade, a utilidade esperada da compra de ações é negativa porque os grandes ganhos, aqui considerados assim àqueles até $120 \times 20 = 2400$, adicionam menos à utilidade $2400^{0.8} = 506$ do que as grandes perdas, àqueles até $120 \times -10 = -1200$ tire do utilitário. A segunda compra de ações tem utilidade esperada positiva, por isso seria a escolha preferida neste exemplo.

4.2.1 Propriedades da variância

Teorema 4.25

Seja X uma variável aleatória constante, isto é, $X = c$ sendo c uma constante real qualquer. Então $\text{Var}(X) = 0$.

Demonstração: Sabemos que $E(X) = c$, então

$$\text{Var}(X) = E[(X - c)^2] = E(0) = 0. \quad \blacksquare$$

Significa que a variância de qualquer número real é zero e assim dizemos que ausência de variância significa que a variável aleatória é degenerada ou não aleatória, é determinística nesta situação.

Teorema 4.26

Seja X uma variável aleatória. Então, para todo $a, b \in \mathbb{R}$, temos que $\text{Var}(X+b) = \text{Var}(X)$ e $\text{Var}(aX + b) = a^2 \text{Var}(X)$.

Demonstração: Sabemos que $E(aX + b) = aE(X) + b$, então

$$\begin{aligned} \text{Var}(aX + b) &= E[aX + b - aE(X) - b]^2 = E\{a^2[X - E(X)]^2\} \\ &= a^2 \text{Var}(X). \quad \blacksquare \end{aligned}$$

Exemplo 4.38

Assumiremos que $E|X|^2 < \infty$ e definamos

$$Z = \frac{X - E(X)}{\sqrt{\text{Var}(X)}} = \frac{X - \mu}{\sigma}.$$

Percebemos então que $E(Z) = 0$ e $\text{Var}(Z) = 1$. Chamaremos Z de variável aleatória padronizada. Na expressão anterior assumimos $\text{Var}(x) = \sigma^2$, sendo esta uma notação comum para a variância.

Teorema 4.27

Seja X uma variável aleatória. Então $\text{Var}(X) \leq E(X - c)^2$, para qualquer $c \neq E(X)$.

Demonstração: Temos que

$$\text{Var}(X) = E(X - \mu)^2 = E(X - c)^2 + (c - \mu)^2.$$

■

Vejamos uma situação simples, onde possamos fazer os cálculos para mostrar esta desigualdade.

Exemplo 4.39

Consideremos a situação em que $X \sim \text{Uniforme}(\alpha, \beta)$. Neste caso $E(X) = \frac{\alpha + \beta}{2}$ e $E(X^2) = \frac{\beta^2 + \alpha\beta + \alpha^2}{3}$. Obtemos então que

$$\text{Var}(X) = \frac{\alpha^2 + \alpha\beta + \beta^2}{3} - \frac{(\alpha + \beta)^2}{4}.$$

Continuando nossos cálculos, encontramos que

$$E(X - c)^2 = \frac{3c^2 - 3c + 1}{3}.$$

Isto significa que $E(X - c)^2 \geq \frac{\alpha^2 + \alpha\beta + \beta^2}{3} - \frac{(\alpha + \beta)^2}{4}$.

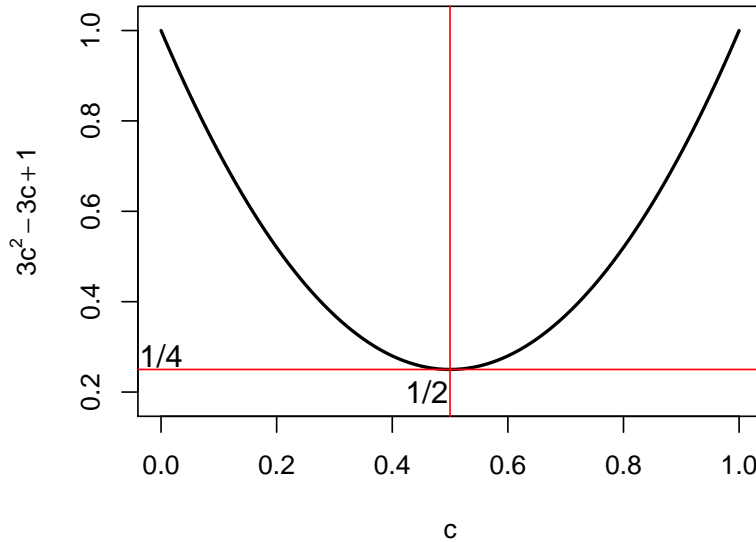


Figura 4.3: Gráfico da função $E(X - c)^2$ em termos de c , especificamente no caso em que $X \sim \text{Uniforme}(0, 1)$. Este é um caso particular daquele considerado no Exemplo 4.39.

No Teorema 4.27 foi dito que $c \neq E(X)$, sendo c uma constante qualquer. Acontece que c deve pertencer ao suporte ou conjunto de valores da variável aleatória tais que a função de densidade seja estritamente positiva. Caso contrário não faz sentido porque simplesmente

não temos como encontrar os momentos de X . Na Figura 4.3 apresentamos o comportamento de $E(X - c)^2$, como função de c para o caso em que $X \sim \text{Uniforme}(0, 1)$.

4.3 Funções geradoras

Vamos agora considerar uma maneira diferente de caracterizar a distribuição de uma variável aleatória mais intimamente relacionada com seus momentos do que com sua função de distribuição.

4.3.1 Função geradora de probabilidade

Nesta seção estudaremos funções que geram probabilidades ou momentos de variáveis aleatórias. A situação mais simples destas funções na teoria das probabilidades está associada a variáveis aleatórias discretas com valores inteiros. Seja X uma variável aleatória discreta com função de probabilidade como

$$p_X(x) = P(X = x), \quad (4.41)$$

para $x = 0, 1, 2, \dots$ e $\sum_{x=0}^{\infty} p_X(x) = 1$.

Definição 4.9

Seja X uma variável aleatória discreta com função de probabilidade como em (4.41). A função definida como

$$G_X(s) = E(s^X) = \sum_{x=0}^{\infty} p_X(x) s^x, \quad (4.42)$$

é chamada de função geradora de probabilidade da variável aleatória X .

Observemos que se a variável aleatória X assume somente um número finito de valores, atribuímos probabilidade zero a todos os valores que não ocorrem. Desta forma, a série em (4.42) sempre está bem definida. Percebemos também que

$$G_X(0) = P(X = 0) \quad \text{e} \quad G_X(1) = \sum_{x=0}^{\infty} P(X = x) = 1.$$

Exemplo 4.40

Seja $X \sim \text{Bernoulli}(\theta)$, $0 < \theta < 1$, ou seja, $P(X = 0) = 1 - \theta$ e $P(X = 1) = \theta$. Então

$$G_X(s) = (1 - \theta)s^0 + \theta s^1 = 1 - \theta + \theta s,$$

é a expressão matemática da função geradora de probabilidade neste caso. Observe que, $G_X(0) = 1 - \theta = P(X = 0)$ e que $G_X(1) = 1$.

A função $G_X(\cdot)$ é uma série de potências que, para qualquer distribuição específica, é conhecida como a função geradora de probabilidade associada. As funções de geradoras de probabilidades têm propriedades interessantes e geralmente podem reduzir bastante a quantidade de trabalho árduo envolvido na análise de uma distribuição.

Teorema 4.28

Seja X uma variável aleatória discreta com função de probabilidade como em (4.41). A função geradora de probabilidade $G_X(s)$ é uma série convergente para $|s| \leq 1$.

Demonstração: A série em (4.42) é conhecida como série de potências. Observemos que se $|s| = 1$, $G_X(s) = 1$ e se $|s| < 1$, então

$$\sum_{x=0}^{\infty} p_X(x) |s|^x \leq \sum_{x=0}^{\infty} p_X(x) = 1.$$

Logo $G_X(s)$ é convergente quando $|s| \leq 1$. ■

O Teorema 4.28 se refere à situação em que a variável aleatória discreta assume infinitos valores.

Exemplo 4.41

Seja X uma variável aleatória com função de probabilidade Poisson(λ), ou seja,

$$P(X = x) = e^{-\lambda} \frac{\lambda^x}{x!}, \quad \text{para } x = 0, 1, 2, \dots$$

Então

$$G_X(s) = \sum_{x=0}^{\infty} p_X(x) s^x = \sum_{x=0}^{\infty} \frac{e^{-\lambda}}{x!} (s\lambda)^x = e^{-\lambda} e^{s\lambda} = e^{-\lambda(1-s)}, \quad |s| \leq 1.$$

Teorema 4.29

Seja X uma variável aleatória discreta com valores inteiros e seja $G_X(s)$ a função geradora de probabilidade. A fórmula

$$P(X = n) = \frac{1}{n!} \left. \frac{d^n G_X(s)}{ds^n} \right|_{s=0}, \quad n = 0, 1, 2, \dots \quad (4.43)$$

permite construir a função de probabilidades de X a partir de derivadas da função geradora de probabilidade.

Demonstração: Observemos primeiro que

$$G_X(s) = p_X(0)s^0 + p_X(1)s^1 + p_X(2)s^2 + p_X(3)s^3 + \dots,$$

logo $G_X(0) = p_X(0)$ e, portanto, $P(X = 0) = G_X(0)$. Por outro lado

$$\frac{dG_X(s)}{ds} = p_X(1) + 2p_X(2)s + 3p_X(3)s^2 + \dots,$$

logo $dG_X(0)/ds = p_X(1)$. Também

$$\frac{d^2G_X(s)}{ds^2} = 2p_X(2) + 3!p_X(3)s + 4 \times 3p_X(4)s^2 + \dots,$$

logo $d^2G_X(0)/ds^2 = 2!p_X(2)$, portanto, $p_X(2) = (d^2G_X(0)/ds^2)/2!$ e assim sucessivamente. ■

Este teorema mostra uma forma de obtermos, via derivação, da função de probabilidade. Uma outra forma de obtermos esta função seria expandir em série de potências de s a função geradora de probabilidade e fazer $p_X(x)$ igual ao coeficiente respectivo da potência s^x na série resultante.

Exemplo 4.42

Seja X uma variável aleatória com função geradora de probabilidade $G_X(s) = \frac{s}{5}(2 + 3s^2)$.

Encontremos a distribuição de X .

Utilizando o Teorema 4.29:

- $G_X(0) = \frac{0}{5}(2 + 3 \times 0^2) = P(X = 0) = 0$,
- $G'_X(s) = \frac{2}{5} + \frac{9}{5}s^2$, implica que $G'_X(0) = P(X = 1) = \frac{2}{5}$,
- $G''_X(s) = \frac{18}{5}s$, implica que $\frac{1}{2}G''_X(0) = P(X = 2) = 0$,
- $G'''_X(s) = \frac{18}{5}$, implica que $\frac{1}{3!}G'''_X(0) = P(X = 3) = \frac{3}{5}$
- $G_X^{(r)}(s) = 0$, para todo $r \geq 4$, do qual temos $\frac{1}{r!}G_X^{(r)}(0) = P(X = r) = 0$.

Então

$$X = \begin{cases} 1, & \text{com probabilidade } 2/5 \\ 3, & \text{com probabilidade } 3/5 \end{cases}.$$

Propriedades da função geradora de probabilidade

Estudamos como a través da função geradora de probabilidade obtemos a função de probabilidade de uma variável aleatória discreta. Vejamos agora outras propriedades esta função.

Teorema 4.30

Sejam X e Y duas variáveis aleatórias assumindo valores em $\{0, 1, \dots\}$ tais que $G_X(s)$ e $G_Y(s)$ existem e coincidem em algum intervalo ao redor do ponto $s = 0$. Então X e Y tem a mesma distribuição de probabilidade.

Demonstração: Para cada $x \geq 0$, sejam $a_x = P(X = x)$ e $b_x = P(Y = x)$. A igualdade

$$G_X(s) = G_Y(s) \text{ se escreve da forma: } \sum_{x=0}^{\infty} a_x s^x = \sum_{x=0}^{\infty} b_x s^x.$$

Para que estas duas séries de potências em s coincidirem, em algum intervalo não trivial ao redor do zero, seus coeficientes devem coincidir, isto é, $a_x = b_x$ para cada $x \geq 0$. Isto significa que as distribuições de probabilidade coincidem. ■

O Teorema 4.30 significa que: se podemos mostrar que duas variáveis aleatórias têm a mesma função geradora de probabilidade em algum intervalo contendo 0, então mostramos que as duas variáveis aleatórias tem a mesma distribuição. Outra maneira de expressar isso é dizer que a função geradora de probabilidade de X nos diz tudo o que há para saber sobre a distribuição de X .

Além de calcular probabilidades, também podemos usar a função geradora de probabilidade para calcular os momentos da distribuição de X . Os momentos de uma distribuição são a média, variância etc.

Teorema 4.31

Se o n -ésimo momento fatorial da variável aleatória X existe, então

$$\lim_{s \rightarrow 1} \frac{d^n}{ds^n} G_X(s) = E(X(X-1) \cdots (X-n+1)). \quad (4.44)$$

Demonstração: Como as séries de potências absolutamente convergentes podem ser derivadas termo a termo conservando-se o mesmo radio de convergência, temos que

$$\begin{aligned} G_X^{(n)}(s) &= \frac{d^n}{ds^n} \sum_{x=0}^{\infty} s^x p_x = \sum_{x=0}^{\infty} p_x \frac{d^n}{ds^n} s^x \\ &= \sum_{x=0}^{\infty} p_x x(x-1) \cdots (x-n+1) s^{x-n}. \end{aligned}$$

Como por hipóteses a esperança existe, pelo Teorema de Abel³ obtemos o resultado em (4.44). ■

³Teorema de Abel: Seja a série de constantes $\sum_{n=0}^{\infty} c_n$ convergente e $f(x) = \sum_{n=0}^{\infty} c_n x^n$, $-1 < x < 1$. Então $\lim_{x \rightarrow 1} f(x) = \sum_{n=0}^{\infty} c_n$. Para demonstração ver Teorema 8.2, p.174, Rudin (1976)

Como consequência deste teorema obtemos que

$$E(X) = G'_X(1) \quad \text{e} \quad \text{Var}(X) = G''_X(1) + G'_X(1) - (G'_X(1))^2. \quad (4.45)$$

Isto deriva do fato que

$$G'_X(s) = \sum_{x=0}^{\infty} xP(X=x)s^{x-1} = \sum_{x=1}^{\infty} xP(X=x)s^{x-1},$$

logo $G'_X(1) = E(X)$. Lembrando que $\text{Var}(X) = E(X^2) - E^2(X)$, encontremos primeiro $E(X^2)$. Acontece que

$$E(X^2) = G''_X(1) + G'_X(1), \quad (4.46)$$

já que

$$G''_X(s) = \sum_{x=2}^{\infty} x(x-1)P(X=x)s^{x-2} = \sum_{x=2}^{\infty} (x^2 - x)P(X=x)s^{x-2},$$

desde que a série $\sum_{x=2}^{\infty} x^2 P(X=x)s^{x-2}$ seja convergente obtém-se (4.46). Como consequência provamos os resultados em (4.45).

Exemplo 4.43 (*Distribuição Geométrica*)

Seja $X \sim \text{Geometrica}(\theta)$. Significa que $P(X=x) = \theta(1-\theta)^x$, para $x = 0, 1, 2, \dots$ e $0 < \theta < 1$. Segundo a Definição 4.9

$$\begin{aligned} G_X(s) &= \sum_{x=0}^{\infty} s^x \theta (1-\theta)^x = \theta \sum_{x=0}^{\infty} ((1-\theta)s)^x \\ &= \frac{\theta}{1 - (1-\theta)s}, \quad \text{para todo } s \text{ tal que } |(1-\theta)s| < 1. \end{aligned}$$

Portanto $G_X(s) = \frac{\theta}{1 - (1-\theta)s}$, sempre que $|s| < \frac{1}{1-\theta}$.

Por simples derivação obtemos que

$$G'_X(s) = \frac{d}{ds} \frac{\theta}{1 - (1-\theta)s} = \frac{\theta(1-\theta)}{(1 - (1-\theta)s)^2},$$

avaliando em $s = 1$, temos $E(X) = G'_X(1) = \frac{1-\theta}{\theta}$. Também

$$G''_X(s) = \frac{2\theta(1-\theta)^2}{(1 - (1-\theta)s)^3},$$

avaliando em $s = 1$, temos $\text{Var}(X) = \frac{1-\theta}{\theta^2}$.

Devemos esclarecer que existe uma outra parametrização da Distribuição Geométrica, desta vez a mudança é que a contagem começa em 1 e não em zero, como no caso do Exemplo 4.43. Caso seja esta a situação

$$E(X) = \frac{1}{\theta} \quad \text{e} \quad \text{Var}(X) = \frac{1 - \theta}{\theta^2}.$$

4.3.2 Função geradora de momentos

Esta é uma outra função que pode associar-se a algumas distribuições de probabilidade. Sua existência não está garantida em todos os casos, quando existe, pode identificar a distribuição associada e possui propriedades semelhantes à função geradora de probabilidade, estudada na Seção 4.3.1.

Definição 4.10 (*Função geradora de momentos*)

Seja X uma variável aleatória definida no espaço de probabilidades (Ω, \mathcal{F}, P) . A função definida como

$$M_X(s) = E(e^{sX}), \quad (4.47)$$

é chamada de função geradora de momentos da variável aleatória X , caso esta esperança existe na vizinhança da origem.

É importante observar que esta função deve existir numa vizinhança não trivial ao redor do zero. Observemos também que a função geradora de momentos e a função geradora de probabilidades estão relacionadas, quando existem, pela igualdade

$$M_X(s) = G_X(e^s). \quad (4.48)$$

Exemplo 4.44 (*Função geradora de momentos da distribuição Gama*)

Seja X uma variável aleatória com distribuição Gama(n, λ). Então a função geradora de momentos de X calcula-se da seguinte forma

$$\begin{aligned} M_X(s) &= \int_0^\infty e^{sx} \frac{(\lambda x)^{n-1}}{\Gamma(n)} \lambda e^{-\lambda x} dx \\ &= \lambda^n (\lambda - s)^{-n} \int_0^\infty \frac{[(\lambda - s)x]^{n-1}}{\Gamma(n)} (\lambda - s) e^{-(\lambda - s)x} dx = \left(\frac{\lambda}{\lambda - s} \right)^n. \end{aligned}$$

Neste cálculo, a última integral vale um devido à que o integrando é a função de densidade Gama. Observe que $M_X(s)$ está definida somente quando $s < \lambda$.

Exemplo 4.45

Suponhamos que a função geradora de momentos da variável aleatória X seja

$$M_X(s) = e^{3(e^s - 1)}.$$

Qual o valor de $P(X = 0)$?

Para respondermos esta questão lembremos a relação entre a função geradora de momentos e a função geradora de probabilidade em (4.48). Pelo enunciado do problema sugere-se que X seja uma variável aleatória discreta, vamos então escrever

$$G_X(s) = e^{3s-1}, \quad (4.49)$$

do qual obtemos que

$$P(X = 0) = \frac{1}{0!} \left. \frac{d^0 G_X(s)}{ds^0} \right|_{s=0} = e^{-3}.$$

Uma outra forma é identificarmos a função geradora de momentos em (4.49) como uma situação particular da função geradora de probabilidade encontrada no Exemplo 4.41. Observe que, se no exemplo referido, $\lambda = 3$ temos por resposta a função geradora de probabilidade, em (4.49).

Na próxima seção demonstramos algumas propriedades básicas da função geradora de momentos e mostramos sua utilidade a través de exemplos.

Propriedades da função geradora de momentos

É importante notar que os momentos não identificam de maneira unívoca a distribuição de probabilidade ou densidade. Além de não termos um teorema que garanta isso veremos, no Exemplo 4.50 duas distribuições diferentes com mesmo momentos.

Teorema 4.32

Seja X uma variável aleatória com função geradora de momentos $M_X(s)$, finita para cada $s \in (-t, t)$ e algum $t > 0$. Então todos os momentos de X são finitos.

Demonstração : Esta prova baseia-se nas identidades:

$$E(|X|^n) = n \int_0^\infty (1 - F_X(x))x^{n-1} dx + n \int_{-\infty}^0 F_X(x)|x|^{n-1} dx,$$

e

$$M_X(s) = 1 + s \int_0^\infty (1 - F_X(x))e^{sx} dx - s \int_{-\infty}^0 F_X(x)e^{sx} dx,$$

onde, por hipóteses, as duas integrais em $M_X(s)$ são finitas para qualquer $s \in (-t, t)$. Demonstraremos que cada integral da expressão de $E(|X|^n)$ é menor ou igual à correspondente integral em $M_X(s)$. Para o caso $x > 0$ escolhemos $s \in (0, t)$ e então

$$\frac{(sx)^n}{n!} \leq e^{sx}.$$

Isto é, $x^n \leq (n!/s^n)e^{sx}$. Desta forma vemos que a primeira integral de $E(|X|^n)$ é menor ou igual do que a primeira integral em $M_X(s)$, sendo esta última finita, a primeira também. Para o caso $x < 0$ é conveniente escolher $s \in (-t, 0)$, pois neste caso $sx > 0$ e então

$$\frac{|sx|^n}{n!} \leq e^{|sx|} = e^{sx}.$$

Novamente, isto significa que $|x|^n \leq (n!/|s|^n)e^{sx}$. Agora a segunda integral de $E(|X|^n)$ é menor ou igual à segunda integral em $M_X(s)$, sendo esta última finita, a primeira também. Desta forma todos os momentos de X existem quando $M_X(s)$ é finita em algum intervalo não trivial ao redor do zero. ■

O fato do n -ésimo momento de uma variável aleatória exista, não implica que este pode ser calculado a través da n -ésima derivada da função geradora de momentos avaliada em zero. Significa que é necessário conhecer a existência da função geradora de momentos para que esta possa ser utilizada na obtenção dos momentos. Vejamos melhor isto no seguinte exemplo.

Exemplo 4.46 (*Função geradora de momentos da distribuição t – Student*)

Seja X uma variável aleatória com densidade t – Student(n). Então, a função geradora de momentos é

$$M_X(s) = \int_{-\infty}^{+\infty} e^{sx} \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} dx.$$

Esta integral não é finita, já que $\lim_{x \rightarrow \infty} e^{sx}(1 + x^2/n)^{-(n+1)/2} = \infty$. Portanto a função geradora de momentos não existe para $s \neq 0$. Por outro lado, sabemos que $E(X) = 0$ e $\text{Var}(X) = n/(n-2)$ para $n > 2$.

Exemplo 4.47

Segundo o Exemplo 4.14, se $X \sim t$ – Student(n) a esperança existe e é zero. Um resultado relevante é que $E(X^r)$ existe para $r < n$. Em particular, temos que

$$E(X^r) = \begin{cases} 0, & \text{se } r < n, r \text{ par} \\ n^{r/2} \frac{\Gamma\left(\frac{r+1}{2}\right) \Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{1}{2}\right) \Gamma\left(\frac{n}{2}\right)}, & \text{se } r < n, r \text{ ímpar} \end{cases}$$

Como consequência, vemos que $E(X) = 0$, se $n > 2$.

Teorema 4.33

Seja X uma variável aleatória com função geradora de momentos $M_X(s)$, finita para cada $s \in (-t, t)$ e algum $t > 0$. Então

$$M_X(s) = \sum_{n=0}^{\infty} \frac{s^n}{n!} E(X^n).$$

Demonstração : Lembremos novamente que

$$E(X^n) = n \int_0^\infty (1 - F_X(x))x^{n-1} dx - n \int_{-\infty}^0 F_X(x)x^{n-1} dx.$$

Então para qualquer $s \in (-t, t)$ e $m \geq 1$,

$$\begin{aligned} \sum_{n=0}^m \frac{s^n}{n!} E(X^n) &= 1 + \sum_{n=1}^m \frac{s^n}{n!} n \int_0^\infty (1 - F_X(x))x^{n-1} dx \\ &\quad - \sum_{n=1}^m \frac{s^n}{n!} n \int_{-\infty}^0 F_X(x)x^{n-1} dx \\ &= 1 + s \int_0^\infty (1 - F_X(x)) \sum_{n=0}^{m-1} \frac{s^n}{n!} x^n dx \\ &\quad - s \int_{-\infty}^0 F_X(x) \sum_{n=0}^{m-1} \frac{s^n}{n!} x^n dx. \end{aligned}$$

Utilizando o teorema da convergência monótona ou de convergência dominada, dependendo dos valores de s e x , cada uma das integrais acima é convergente para qualquer $s \in (-t, t)$, quando m tende ao infinito. De modo que

$$\sum_{n=0}^\infty \frac{s^n}{n!} E(X^n) = 1 + s \int_0^\infty (1 - F_X(x))e^{sx} dx - s \int_{-\infty}^0 F_X(x)e^{sx} dx = M_X(s). \quad \blacksquare$$

É claro que, segundo o Teorema 4.33, somente podemos ter função geradora de momentos se estes, os momentos, existirem. A situação especial da função de densidade Cauchy serve de indicativo de que em nem toda situação existe esta função.

Exemplo 4.48

Segundo o Exemplo 4.22,

$$E(X^n) = \frac{1}{n+1},$$

para $n \geq 1$, no caso em que $X \sim \text{Uniforme}(0, 1)$. Encontremos a função geradora de momentos utilizando o Teorema 4.33.

Simplesmente escrevendo

$$\sum_{n=0}^\infty \frac{s^n}{n!} E(X^n),$$

e substituindo pelos respectivos valores de $E(X^n)$, temos que

$$M_X(s) = \sum_{n=0}^\infty \frac{s^n}{(n+1)!}, \quad (4.50)$$

é a expressão para a função geradora de momentos de uma variável aleatória esta distribuição. Segundo o Teorema Weierstrass M-test (Spivak, 1994, Teorema 4, Parte IV, Seção 24, pg.499) a série em (4.50) é uniformemente convergente qualquer seja o número real t e $|s| < t$.

Observemos que, tanto a função geradora de probabilidade $G_X(s)$ quanto a função geradora de momentos $M_X(s)$ são séries de funções, por isso a necessidade de Cálculo avançado a entendermos estes conceitos assim como àqueles que serão considerados no Capítulo 7.

Teorema 4.34

Seja X uma variável aleatória com função geradora de momentos $M_X(s)$, finita para cada $s \in (-t, t)$ e algum $t > 0$. Então $M_X(s)$ tem derivadas de qualquer ordem em $s \in (-t, t)$ e

$$\left. \frac{d^n}{ds^n} M_X(s) \right|_{s=0} = E(X^n).$$

Demonstração: Dado que $M_X(s)$ pode ser expressa como série de potências de s , diferenciando e avaliando em zero são obtidos os coeficientes de $E(X^n)$. ■

Exemplo 4.49 (Função geradora de momentos da densidade Beta)

Seja $X \sim \text{Beta}(\alpha, \beta)$, para $\alpha, \beta > 0$, segundo a Definição 3.26. Significa que a função de densidade de X é

$$f(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} = \frac{1}{\text{Beta}(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1}, \quad 0 < x < 1.$$

Encontremos a função geradora de momentos $M_X(s)$.

$$\begin{aligned} M_X(s) &= E(e^{sX}) = \frac{1}{\text{Beta}(\alpha, \beta)} \int_0^1 e^{sx} x^{\alpha-1} (1-x)^{\beta-1} dx \\ &= \frac{1}{\text{Beta}(\alpha, \beta)} \int_0^1 \left(\sum_{n=0}^{\infty} \frac{(sx)^n}{n!} \right) x^{\alpha-1} (1-x)^{\beta-1} dx \end{aligned}$$

Utilizamos a expansão em série de potências para a função exponencial

$$= \frac{1}{\text{Beta}(\alpha, \beta)} \sum_{n=0}^{\infty} \frac{(s)^n}{n!} \int_0^1 x^{n+\alpha-1} (1-x)^{\beta-1} dx$$

A série de potências é integrável termo a termo no raio de convergência

$$= \sum_{n=0}^{\infty} \frac{(s)^n}{n!} \left(\frac{B(n + \alpha, \beta)}{B(\alpha, \beta)} \right)$$

Por definição da função Beta

$$= \frac{B(\alpha, \beta)}{B(\alpha, \beta)} \frac{(s)^0}{0!} \sum_{n=1}^{\infty} \frac{(s)^n}{n!} \left(\frac{B(n + \alpha, \beta)}{B(\alpha, \beta)} \right).$$

Agora faremos um trabalho de uma álgebra complicada e aplicaremos uma identidade

de recorrência da função gama com um salto de comprimento n :

$$\Gamma(\alpha + n) = \Gamma(\alpha) \prod_{r=0}^n (\alpha + r). \quad (4.51)$$

$$M_X(s) = 1 + \sum_{n=1}^{\infty} \left(\frac{\Gamma(n + \alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta + n)} \times \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \right) \frac{s^n}{n!}$$

Pela definição de função Beta

$$= 1 + \sum_{n=1}^{\infty} \left(\frac{\Gamma(n + \alpha)}{\Gamma(\alpha)} \times \frac{\Gamma(\alpha + \beta)}{\Gamma(n + \alpha + \beta)} \right) \frac{s^n}{n!}$$

$$1 + \sum_{n=1}^{\infty} \left(\frac{\Gamma(\alpha) \prod_{r=0}^n (\alpha + r)}{\Gamma(\alpha)} \times \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha + \beta) \prod_{r=0}^n (\alpha + \beta + r)} \right) \frac{s^n}{n!}$$

Aplicando identidade de recorrência (4.51)) da função gama

$$= 1 + \sum_{n=1}^{\infty} \left(\prod_{r=0}^{n-1} \frac{\alpha + r}{\alpha + \beta + r} \right) \frac{s^n}{n!}. \quad \text{Produto de produtos}$$

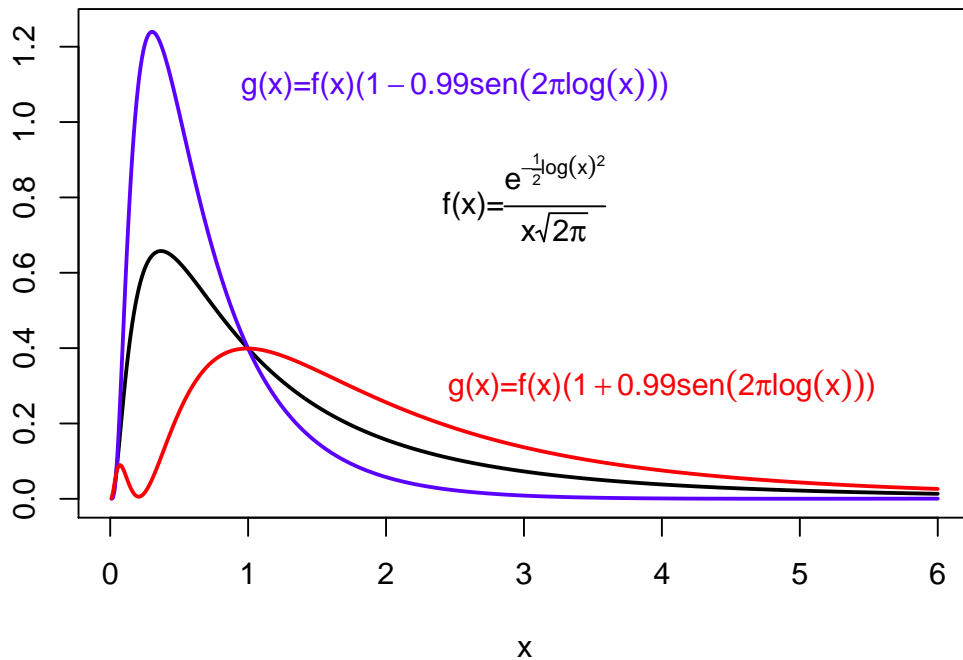


Figura 4.4: Funções de densidade mostradas no Exemplo 4.50, situação exemplificando que diferentes funções de densidade podem ter a mesma função geradora de momentos. Mostramos a função de densidade f da variável aleatória X e duas da família de densidade de Y , quando $a = -0.99$ e $a = 0.99$, estas avaliadas em x .

Uma outra situação interessante acontece no seguinte exemplo apresentado por Heyde (1963), neste caso prova-se que os momentos não definem univocamente a função de densidade.

Exemplo 4.50

Seja X uma variável aleatória positiva com função de densidade

$$f_0(x) = \frac{e^{-\frac{1}{2} \log^2(x)}}{\sqrt{2\pi x}}, \quad x > 0, \quad (4.52)$$

e seja uma outra variável aleatória Y com função de densidade

$$f_a(y) = f_0(y)(1 + a \sin(2\pi \log(y))) \quad \text{para} \quad -1 \leq a \leq 1, y > 0. \quad (4.53)$$

Observemos que a densidade de X é a log-Normal (Exemplo 3.35). Os momentos destas distribuições coincidem, isto foi demonstrado em Heyde (1963). Mas, como mostrado na Figura 4.4 as funções de densidade de X e Y são diferentes.

Neste exemplo afirmamos que as densidades em (4.52) e em (4.53) possuem os mesmos momentos. Para demonstrar isso é suficiente mostrar que

$$\int_0^\infty y^r f_0(y) \sin(2\pi \log(y)) dy = 0, \quad \text{para} \quad r = 0, 1, 2, \dots$$

Fazendo a mudança de variáveis $y = \exp(s+r)$, $s = \log(y) - r$, $ds = dy/y$ a integral torna-se

$$\begin{aligned} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty \exp(rs + r^2) \exp(-(s+r)^2/2) \sin(2\pi(s+r)) ds &= \\ &= \frac{1}{\sqrt{2\pi}} \exp(r^2/2) \int_{-\infty}^\infty \exp(-s^2/2) \sin(2\pi s) ds = 0. \end{aligned}$$

As duas equações mantêm-se porque r é um número inteiro e a integral é ímpar. A partir da prova, deve ficar claro que podíamos escolher

$$g(x) = f_0(x) \left\{ 1 + \sum_{k=1}^\infty a_k \sin(k\pi \log(x)) \right\}, \quad \text{se} \quad \sum_{k=1}^\infty |a_k| \leq 1$$

para obter uma grande família de densidades tendo os mesmos momentos que a log-Normal.

Exemplo 4.51 (Momentos da distribuição log-Normal)

Perguntamo-nos agora, quais são os momentos da distribuição log-Normal? recordemos que se a variável aleatória Z tiver distribuição Normal, então $X = \exp(Z)$ tem por distribuição log-Normal. Assim,

$$E(X^n) = E(\exp(nZ)) = \int_{-\infty}^\infty \frac{e^{nx}}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = e^{\frac{n^2}{2}} \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-n)^2}{2}} dx = e^{\frac{n^2}{2}},$$

uma vez que a última integral é a densidade do normal com esperança n e variância 1.

Vamos notar que existe uma família de variáveis aleatórias discretas com estes mesmos momentos. Seja $a > 0$ e

$$P(Y_a = ae^k) = \frac{1}{c_a a^k} \exp\left(-\frac{k^2}{2}\right), \quad k \in \mathbb{Z},$$

onde c_a é uma constante normalizadora, ou seja, $c_a = \sum_{k \in \mathbb{Z}} \exp(-k^2/2)/a^k$. Então,

$$\begin{aligned} \exp\left(-\frac{n^2}{2}\right) E(Y_a^n) &= \exp\left(-\frac{n^2}{2}\right) \sum_{k \in \mathbb{Z}} (ae^k)^n \frac{1}{c_a a^k} \exp\left(-\frac{k^2}{2}\right) \\ &= \sum_{k \in \mathbb{Z}} \frac{1}{c_a a^{k-n}} \exp\left(-\frac{(k-n)^2}{2}\right) = 1, \end{aligned}$$

pela definição de c_a . É claro que isto significa que os momentos de Y são dados por

$$E(Y_a^n) = \exp\left(\frac{n^2}{2}\right).$$

A densidade log-Normal decai segundo $\exp(-\log^2(x)/2)$ quando $|x| \rightarrow \infty$. O próximo exemplo decaia mais rápido. Uma vez que a distribuição exponencial e^{-x} , para $x \geq 0$, é determinada pelos seus momentos não podemos esperar fazer muito melhor do que isto.

Exemplo 4.52

Seja $\lambda \in (0, 1)$ e para $a \in (-1, 1)$ seja

$$f_{a,\lambda}(x) = c_\lambda \exp(-|x|^\lambda) \left(1 + a \sin(\beta|x|^\lambda \operatorname{sgn}(x))\right), \quad x \in \mathbb{R},$$

onde $\beta = \tan(\lambda\pi/2)$, $1/c_\lambda = \int_{-\infty}^{\infty} \exp(-|x|^\lambda) dx$ e $\operatorname{sgn}(x)$ é a função signo, ou seja,

$$\operatorname{sgn}(x) = \begin{cases} -1, & \text{se } x \leq 0 \\ 1, & \text{se } x > 0 \end{cases}.$$

Para provar que se trata de funções de densidade e que, para um valor fixo de λ têm os mesmos momentos, basta mostrar

$$\int_{-\infty}^{\infty} x^n \exp(-|x|^\lambda) \sin(\beta|x|^\lambda \operatorname{sgn}(x)) dx = 0, \quad n = 0, 1, 2, \dots$$

Isto é claro para o n , uma vez que o integrando é estranho. Para provar o resultado para o ímpar n é suficiente integrar mais de $[0, 1)$. Utilizando a identidade

$$\int_0^\infty t^{p-1} e^{-qt} dt = \frac{\Gamma(p)}{q^p}, \quad \text{quando } \operatorname{Re}(q) > 0,$$

com $p = (n+1)/\lambda$, $q = 1 + i\beta$, $\operatorname{Re}(q)$ indicando a parte inteira do número complexo q e

fazendo a mudança de variáveis $t = x^\lambda$, temos

$$\begin{aligned} \frac{\Gamma\left(\frac{n+1}{\lambda}\right)}{(1+i\beta)^{\frac{n+1}{\lambda}}} &= \int_0^\infty x^{\lambda\left(\frac{n+1}{\lambda}-1\right)} \exp\left(-(1+i\beta)x^\lambda\right) \lambda x^{\lambda-1} dx \\ &= \lambda \int_0^\infty x^n \exp(-x^\lambda) \cos(\beta x^\lambda) dx - i\lambda \int_0^\infty x^n \exp(-x^\lambda) \sin(\beta x^\lambda) dx. \end{aligned}$$

Desde que $\beta = \tan(\lambda\pi/2)$,

$$(1+i\beta)^{\frac{n+1}{\lambda}} = \left(\cos(\lambda\pi/2)\right)^{-\frac{n+1}{\lambda}} \left(\exp(i\lambda\pi/2)\right)^{\frac{n+1}{\lambda}}.$$

O lado direito é real, uma vez que $\lambda < 1$ e $(n+1)$ é par, então

$$\int_0^\infty x^n \exp(-x^\lambda) \sin(\beta x^\lambda) dx = 0.$$

Como mencionamos antes, nem todas as distribuições de probabilidade permitem calcular a função geradora de momentos num intervalo não trivial ao redor do zero e nem todos os cálculos são tão simples. Por exemplo, a função geradora de momentos da distribuição Cauchy não existe para valores de s distintos de zero. Por outro lado, quando duas variáveis X e Y tem a mesma distribuição, suas funções geradoras de momentos coincidem. Pelo contrário, se $M_X(s) = M_Y(s)$ numa vizinhança não trivial ao redor do zero, pode demonstrar-se que suas distribuições não necessariamente coincidem, como visto no Exemplo 4.50.

Mais ainda, seja $X \sim F$ sendo F uma função de distribuição contínua não simétrica, ou seja $F(x) \neq 1 - F(-x)$ para ao menos um $x \in (-\infty, \infty)$, para a qual todos os momentos de ordem ímpar são zero. Podemos verificar que

$$G(x) = \frac{1}{2}F(x) + \frac{1}{2}(1 - F(-x)),$$

é uma função de distribuição de uma variável aleatória possuindo os mesmos momentos do que X . Também verifica-se que G é uma função simétrica, ou seja, $G(x) = 1 - G(-x)$, para todo $x \in (-\infty, \infty)$.

Devemos prestar atenção a um detalhe importante no Exemplo 4.50. Nele consideramos a situação de duas funções de densidade f e g diferentes mas com os mesmos momentos. O detalhe é que dizemos que os momentos das variáveis aleatórias correspondentes coincidem e não as funções geradoras de momentos. Acontece que, no caso da densidade log-Normal, a função geradora de momentos não existe (Heyde, 1963), mas os momentos sim. Todos os momentos da distribuição $\log - Normal(\mu, \sigma^2)$ existem e são

$$E(X^n) = e^{n\mu + n^2\sigma^2/2}.$$

Para começar, não é difícil dar exemplos de funções de probabilidade diferentes que têm a mesma média e a mesma variância. Por exemplo, suponha que X e Y são duas variáveis aleatórias com funções de probabilidade

x	1	2	3	4	5	6
$P(X=x)$	0	1/4	1/2	0	0	1/4

e

y	1	2	3	4	5	6
$P(Y=y)$	1/4	0	0	1/2	1/4	0

Calculando as médias e variâncias obtemos que $E(X) = E(Y) = 7/2$ e $\text{Var}(X) = \text{Var}(Y) = 9/4$ mas, certamente as funções probabilidade de X e Y são bastante diferentes.

4.4 Exercícios

Exercícios das Seções 4.1

1. Verifique se a desigualdade em (4.6), no Exemplo 4.3 é verdadeira.
2. Um jogador vai lançando uma moeda honesta. Ele para depois de lançar ou duas caras ou duas coroas sucessivas. Qual a esperança do número de lançamentos?
3. Suponha que $X \sim U[0, 1]$. Determine os valores de t ($t \in \mathbb{R}$) tais que $E(X^t)$ é finita. Qual o valor da esperança nesse caso?
4. Seja $X(\omega) = \mathbf{1}_A(\omega)$, para algum $A \in \mathcal{F}$. Prove que $E(X) = P(A)$.
5. Suponha que a variável aleatória X têm densidade triangular, da forma

$$f(x) = \begin{cases} 1+x & \text{se } -1 \leq x \leq 0 \\ 1-x & \text{se } 0 < x \leq 1 \\ 0 & \text{se } x < -1 \text{ ou } x > 1 \end{cases}.$$

Calcule $E(X)$ e $\text{Var}(X)$.

6. Suponha que a variável aleatória X têm densidade da forma

$$f(x) = \begin{cases} \frac{1}{2}x & \text{se } 0 \leq x < 1 \\ 1/2 & \text{se } 1 < x \leq 2 \\ \frac{1}{2}(3-x) & \text{se } 2 < x \leq 3 \\ 0 & \text{se } x < 0 \text{ ou } x > 3 \end{cases}.$$

Calcule $E(X)$ e $\text{Var}(X)$.

7. Demonstre o Teorema 4.5.
8. Demonstre o Teorema 4.6.
9. Prove que se $E(X) = 0$ e $P(X \geq 0) = 1$, então $P(X = 0) = 1$.
10. Seja $X \sim U(0, 1)$, isto é, X é uniforme contínua no intervalo $[0, 1]$. Calcule $E(X^n)$, sendo n um número natural qualquer.
11. Para variáveis aleatórias positivas X definimos os momentos negativos de ordem n como $E(X^{-n})$, onde $n > 0$ é um inteiro. Encontre $E(1/(X+1))$ se a distribuição da variável aleatória é:
 - a) *Binomial*(θ), isto é, $P(X = x) = \binom{m}{x}\theta^x(1-\theta)^{m-x}$, $x = 0, 1, 2, \dots, m$ e $0 < \theta < 1$.
 - b) *Poisson*(λ), isto é, $P(X = x) = e^{-\lambda} \frac{\lambda^x}{x!}$, $x = 0, 1, 2, \dots$ e $\lambda > 0$.
12. Prove que $\text{Var}(X) = E[X(X-1)] - E(X)[E(X)-1]$.
13. Suponha que a variável aleatória X assume valores somente no intervalo $[a, b]$. Prove que $a \leq E(X) \leq b$ e

$$\text{Var}(X) \leq \frac{(b-a)^2}{4}.$$

14. Prove que $\text{Var}(X) = 0$ se, e somente se, $P[X = E(X)] = 1$.
15. Suponhamos que X seja uma variável aleatória com função de distribuição F_X e inversa, ou função quantil, F_X^{-1} . Prove que

$$E(X) = \int_0^1 F_X^{-1}(x) dx,$$

se $E(X)$ estiver bem definida.

16. Seja $X \geq 0$ com esperança finita e suponha que, para algum $p \in (0, 1)$, se cumpre a desigualdade $P(X \geq k) \leq p^k$, para cada $k = 0, 1, \dots$. Prove que $E(X) \leq 1/(1 - p)$.
17. Uma caixa contém N cartões idênticos numerados entre 1 e N . Dessa caixa são selecionados n cartões com reposição. Seja X o maior dos números selecionados. Encontre $E(X)$.
18. Prove que se a variável aleatória X é limitada, isto é, uma variável aleatória é dita limitada se X é tal que $P(a \leq X \leq b) = 1$, $a < b$ reais qualquer, então X tem momentos finitos de toda ordem.
19. Seja A um evento aleatório. Calcule todos os momentos absolutos da variável aleatória $X = I_A$.
20. Sejam X e Y duas variáveis aleatórias ambas com esperança finita. Prove que
 - a) $E(\min\{X, Y\}) \leq \min\{E(X), E(Y)\} \leq E(X)$,
 - b) $E(\max\{X, Y\}) \geq \max\{E(X), E(Y)\} \geq E(X)$.
21. Seja X uma variável aleatória não negativa com função de distribuição contínua F_X e com esperança finita μ . Demonstre que a seguinte função é de distribuição.

$$G(y) = \begin{cases} 1 - \frac{1}{\mu} \int_y^\infty [1 - F_X(x)] dx & \text{se } y > 0 \\ 0 & \text{se } y \leq 0 \end{cases}.$$

Demonstre que a esperança desta distribuição é $2E(X^2)/\mu$, supondo que o segundo momento de X é finito.

22. Seja X uma variável aleatória tal que $E(e^{aX})$ existe para algum $a > 0$. Prove que $P(X \geq \epsilon) \leq E(e^{aX})/e^{a\epsilon}$.
23. Seja X uma variável aleatória com variância finita e seja c uma constante. Prove que

$$E[(X - c)^2] = \text{Var}(X) + [E(X) - c]^2.$$

24. Duas bolas são escolhidas aleatoriamente de uma urna contendo 4 bolas azuis, 3 vermelhas e 2 laranjas. Suponha que ganhamos 10 reais para cada bola azul selecionada, ganhamos 1 real para cada bola laranja, porém perdemos 8 reais para cada bola vermelha. Seja X o nosso lucro.
 - a) Determine a função de probabilidade de X ,
 - b) Obtenha o valor esperado e a variância de X .

25. Seja X uma variável aleatória com distribuição logística da forma

$$f(x) = \frac{e^{-x}}{(1 + e^{-x})^2}, \quad x \in \mathbb{R}.$$

- i) Prove que a distribuição de X é simétrica em torno de zero, isto é, prove que $f(x) = f(-x)$.
 - ii) Determine se X tem esperança finita. Se tiver, ache o valor. Encontre o valor de $\text{Var}\{X\}$.
 - iii) Obtenha a densidade de $Y = e^X$ e ache $E\{Y\}$.
26. De uma caixa contendo N bilhetes idênticos, numerados de 1 a N , n bilhetes são sorteados com reposição. Seja X o maior número sorteado. Encontre $E(X)$.
27. Seja $X \sim \text{Geométrica}(p)$. Mostre que

$$E\left(\frac{1}{X}\right) = \frac{-p \log(p)}{1 - p}.$$

Dica: Você vai precisar avaliar expressões do tipo $\sum_{n=1}^\infty a^n/n$. Para isso, escreva $a^n/n = \int_0^a u^{n-1} du$ e troque a soma e a integral.

Exercícios da Seção 4.2

- Suponha que a variável aleatória X tenha esperança μ , variância σ^2 e que $Y = aX + b$. Determinar os valores de a e b para os quais $E(Y) = 0$ e $\text{Var}(Y) = 1$.
- Considere que X seja uma variável aleatória com esperança $E(X) = 1$ e variância $\text{Var}(X) = 4$. Encontre:
 - $E(2X - 4)$,
 - $E^2(X)$,
 - $E((2X - 4)^2)$.

- A variável aleatória X tem por função de densidade

$$f(x) = \begin{cases} \alpha + \beta x^2, & \text{para } 0 \leq x \leq 1 \\ 0, & \text{caso contrário} \end{cases}.$$

Sabemos que $E(X) = 3/5$. Encontre:

- $P(X < \frac{1}{2})$.
 - $\text{Var}(X)$.
- Mostre que todos os momentos, qualquer seja a ordem, existem caso a função de densidade da variável aleatória X seja

$$f(x) = \begin{cases} \frac{1}{2}x, & \text{para } 0 \leq x < 1 \\ \frac{1}{2}, & \text{para } 1 < x \leq 2 \\ \frac{1}{2}(3-x), & \text{para } 2 < x \leq 3 \end{cases}.$$

Encontre a esperança e variância de X .

- Considere que a variável aleatória X tem por função de densidade triangular, dada por

$$f(x) = \begin{cases} 1+x, & \text{para } -1 \leq x \leq 0 \\ 1-x, & \text{para } 0 < x \leq 1 \\ 0, & \text{para } x < -1 \text{ ou } 1 < x \end{cases}.$$

Encontrar $E(X)$ e $\text{Var}(X)$.

- Considere que a variável aleatória assume valores somente no intervalo $[\alpha, \beta]$, $\alpha < \beta$. Prove que

$$\alpha < E(X) < \beta \quad \text{e} \quad \text{Var}(X) \leq \frac{(\beta - \alpha)^2}{2}.$$

- Seja X uma variável aleatória com distribuição $\text{Pareto}(\alpha, \beta)$, isto é, a função de densidade de X é da forma

$$f(x) = \begin{cases} 0, & \text{caso } x < \alpha \\ \frac{\beta \alpha^\beta}{x^{\beta+1}}, & \text{caso } x \geq \alpha \end{cases}.$$

- Mostre que todos os momentos de ordem n existem se, e somente se, $n < \beta$.
 - Para $\beta > 2$ encontre a esperança e variância desta distribuição.
- Suponha que a variável aleatória X tenha esperança μ e variância σ^2 . Mostre que o terceiro momento central de X pode ser expresso como $E(X^3) - 3\mu\sigma^2 - \mu^3$.
 - Seja $X \sim N(\mu, \sigma^2)$. Prove que o desvio médio é dado por

$$E(|X - \mu|) = \sigma \sqrt{\frac{2}{\pi}} = 0.7978846\sigma.$$

10. Diz-se que X tem por distribuição $\log - Normal(\mu, \sigma^2)$ se a função de densidade tiver a expressão

$$f(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\ln(x) - \mu)^2}{2\sigma^2}\right) \quad (4.54)$$

para $x > 0$ e zero em caso contrário. Prove que $E(X) = \exp(\mu + \sigma^2/2)$ e $\text{Var}(X) = \exp(2\mu + 2\sigma^2) - \exp(2\mu + \sigma^2)$.

11. Seja X uma variável aleatória e c uma constante real qualquer. Prove que

$$E(X - c)^2 = \text{Var}(X) + (E(X) - c)^2.$$

12. Seja $X \sim U(\alpha, \beta)$. Demonstre que $E(X^n) = \frac{\beta^{n+1} - \alpha^{n+1}}{(n+1)(\beta - \alpha)}$.

13. Considere (Ω, \mathcal{F}, P) um espaço de probabilidade e $X = \mathbf{1}_A$, ou seja, X é a função indicado do evento $A \in \mathcal{F}$. Prove que $\text{Var}(X) = P(A)[1 - P(A)] \leq 1/4$.

14. Demonstre que se X é uma variável aleatória limitada quase certamente, isto é, que se existe $k > 0$ de maneira que $P(|X| \leq k) = 1$, então todos os momentos centrais de X existem.

15. Seja X uma variável aleatória com segundo momento finito e seja m uma mediana. Prove que

$$|m - E(X)| \leq \sqrt{2 \text{Var}(X)}.$$

16. Seja $X \sim N(\mu, \sigma^2)$. Prove que para cada $n = 0, 1, 2, \dots$ se cumpre a relação

$$E(|X - \mu|^n) = \begin{cases} 1 \times 3 \times 5 \times \dots \times (n-1)\sigma^n, & \text{caso } n \text{ par} \\ 0, & \text{caso } n \text{ ímpar} \end{cases}.$$

17. Seja $X \sim N(0, 1)$. Prove que para cada $n = 0, 1, 2, \dots$ se cumpre a relação

$$E(X^n) = \begin{cases} \frac{n!}{2^{n/2}(n/2)!}, & \text{caso } n \text{ par} \\ 0, & \text{caso } n \text{ ímpar} \end{cases}.$$

Exercícios da Seção 4.3

1. Encontre a função geradora de probabilidades G_X para cada uma das seguintes funções de probabilidade:

(a) $P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$, para $x = 0, 1, 2, \dots, n$ e $0 < p < 1$.

(b) $P(X = x) = \frac{\lambda^x}{x!} \frac{e^{-\lambda}}{1 - e^{-\lambda}}$, para $x = 1, 2, \dots$ e $\lambda > 0$.

(c) $P(X = x) = \frac{pq^x}{1 - q^{N+1}}$, para $x = 0, 1, 2, \dots, N$, $0 < p < 1$ e $q = 1 - p$.

2. Prove que $M_X(s) = G_X(e^s)$ caso existirem estas duas funções para a variável aleatória X .
3. Seja X uma variável aleatória discreta com valores em $\{0, 1, 2, 3, \dots\}$. Mostre que

$$\sum_{n=0}^{\infty} s^n P(X \leq n) = \frac{G_X(s)}{1-s}.$$

4. Seja X uma variável aleatória com função geradora de momentos M_X e a e b constantes quaisquer. Prove que

$$M_{aX+b}(s) = e^{bt} M_X(as)$$

5. Seja X uma variável aleatória discreta com função geradora de probabilidades G_X . Sejam α e β dois números inteiros não negativos e $Y = \alpha X + \beta$. Encontre a função geradora de probabilidades G_Y de Y .
6. Seja X um avariável aleatória com função de densidade f_X e função geradora de momentos M_X definida numa vizinhança $(-\delta, \delta)$ do zero, para algum $\delta > 0$. Prove que

$$P(X > a) \leq e^{at} M_X(s), \quad 0 < s < \delta.$$

7. Considere X uma variável aleatória com distribuição Exponencial Dupla ou Laplace com função de densidade

$$f(x) = \frac{1}{2\lambda} e^{-\frac{1}{\lambda}|x-\mu|}, \quad -\infty < x < \infty, \quad \lambda > 0 \quad -\infty < \mu < \infty.$$

Prove que a função geradora de momentos M_X existe e é igual a

$$M_X(s) = \frac{e^{\mu s}}{1 - \lambda^2 s^2}, \quad s < \frac{1}{\lambda}.$$

8. Para uma variável aleatória X , com função geradora de momentos M_X , define-se a função geradora de cumulantes K_X como

$$K_X(s) = \ln(M_X(s)).$$

Prove que:

$$(a) \quad K'_X(0) = E(X).$$

$$(b) \quad K''_X(0) = \text{Var}(X)$$

9. Suponha que a variável aleatória Y tenha função geradora de momentos M_Y e que a variável aleatória X tenha função geradora de momentos M_X dada por

$$M_X(s) = \frac{1}{3}(2e^{3s} + 1)M_Y(s).$$

Dado que $E(Y) = 10$ e que $\text{Var}(Y) = 12$ determine $E(X)$ e $\text{Var}(X)$.

10. Seja X uma variável aleatória discreta com função de probabilidade $P(X = x) = p_x$, $x = 0, 1, 2, \dots$ e seja $P(X > x) = q_x$, $x = 0, 1, 2, \dots$. Logicamente $q_x = p_{x+1} + p_{x+2} + \dots$, para $x \geq 0$. Consideremos a série $Q(s) = \sum_{x=0}^{\infty} q_x s^x$. Esta série é convergente quando $|s| < 1$. Mostre que

$$Q(s) = \frac{1 - G_X(s)}{1 - s},$$

onde G_X é a função geradora de probabilidades de X . Encontre $E(X)$ e $\text{Var}(X)$, se existirem, em termos de Q e suas derivadas.

11. Seja X uma variável com função geradora de momentos M_X , a qual existe numa vizinhança $(-\delta, \delta)$ do zero, para $\delta > 0$. Prove que

$$E(|X|) < \frac{n!}{s^n} (M_X(s) + M_X(-s)),$$

para qualquer $s \in (-\delta, \delta)$, fixo e qualquer inteiro $n \geq 1$.

Capítulo 5

Vetores aleatórios

Até este ponto, consideramos apenas variáveis aleatórias unidimensionais definidas em um espaço amostral. Naturalmente, é possível definir mais de uma variável aleatória para um determinado experimento. Considere, por exemplo, a seleção criteriosa de N indivíduos, cada indivíduo selecionado tem um número de atributos que podem ser medidos: altura, peso, idade e assim por diante. Claramente, podemos definir variáveis aleatórias unidimensionais que medem cada um desses atributos para um indivíduo escolhido aleatoriamente. Assim, podemos considerar as distribuições unidimensionais de cada uma dessas variáveis ou podemos começar a considerar as que chamaremos de distribuições conjuntas ou funções de distribuição de várias variáveis.

Um outro exemplo simples seria o caso de escolher, ao acaso, um ponto do círculo unitário. Nesse caso devemos considerar um par de variáveis aleatórias univariadas ou unidimensionais X e Y , correspondendo às coordenadas cartesianas do ponto escolhido. Formalmente, temos

$$\mathbf{w} = (x, y) = (X(w), Y(w)), \quad \mathbf{w} \in \Omega = \{(x, y) : \sqrt{x^2 + y^2} \leq 1\}.$$

Em situações como essa o interesse está em um vetor de variáveis aleatórias, todas definidas no mesmo espaço de probabilidade.

Significa assim que a noção de uma variável aleatória pode ser generalizada para o caso em que várias quantidades são de interesse. Podemos então supor que uma variável aleatória vetorial X ou um vetor aleatório seja uma função que atribui um vetor de números reais para cada resultado em Ω ou $\mathbf{\Omega}$, o espaço amostral do experimento aleatório.

Chamemos atenção acerca da notação. Observemos que $\omega \in \Omega$ são notações correspondentes aos pontos ω do espaço amostral Ω de dimensão um, enquanto $\mathbf{\omega}$ será a notação dos pontos do espaço amostral $\mathbf{\Omega}$ de dimensão superior; desta forma vamos escrever

$$\mathbf{\omega} = (\omega_1, \omega_2, \dots, \omega_n) \in \mathbf{\Omega} = (\Omega_1, \Omega_2, \dots, \Omega_n),$$

sendo $\mathbf{\Omega}$ um espaço amostral de dimensão n . Desta mesma maneira, $\mathbf{X}, \mathbf{Y}, \dots$ serão as notações para variáveis aleatórias de dimensão mais do que um e \mathcal{F} e \mathbf{P} denotarão σ -álgebras e funções de probabilidade multidimensionais, respectivamente.

Primeiro, na Seção 5.1, vamos definir vetores aleatórios e apresentamos as noções básicas envolvidas. Na Seção 5.2 estudamos as distribuições conjuntas e algumas de suas propriedades. Nas seções 5.3 e 5.4 examinamos funções de vetores aleatórias e definimos momentos vetoriais.

5.1 Vetores aleatórios

Podemos generalizar a noção de variável aleatória para o caso em que várias quantidades são de interesse. Intuitivamente, dizemos que uma variável aleatória vetorial \mathbf{X} é uma função que atribui um vetor de números reais de dimensão m para cada resultado em Ω , o espaço amostral da experiência aleatória multidimensional.

Vamos esclarecer melhor a notação. Isto é necessário porque ao trabalharmos com diversas dimensões, em muitas situações, devemos esclarecer quantas dimensões têm os espaços nos quais trabalhamos. Primeiro dizemos que

$$\Omega^n = (\underbrace{\Omega, \dots, \Omega}_{n \text{ vezes}}),$$

representa o espaço amostral de dimensão n ou o espaço amostral n -dimensional. Quando aparecer somente a notação Ω sem o índice n significa que a dimensão é entendida. Sejam agora $\mathcal{F}_1, \mathcal{F}_2, \dots, \mathcal{F}_n$ σ -álgebras de eventos definidas nos espaços amostrais $\Omega_1, \dots, \Omega_n$, respectivamente. Significa que cada classe de eventos aleatórios $\mathcal{F}_1, \mathcal{F}_2, \dots$ satisfaz as condições na Definição 2.2. Dizemos então que

$$\mathcal{F}^n = (\mathcal{F}_1, \dots, \mathcal{F}_n), \quad (5.1)$$

é uma σ -álgebra de eventos de Ω^n de dimensão n ou n -dimensional. A σ -álgebra definida acima permite que em cada dimensão a σ -álgebra possa ser diferente, mas o mais comum será termos a mesma σ -álgebra em cada dimensão. Assim, vamos dizer que

$$\mathcal{F}^n = (\underbrace{\mathcal{F}, \dots, \mathcal{F}}_{n \text{ vezes}}),$$

representa a σ -álgebra de dimensão n , onde \mathcal{F} é uma σ -álgebra em Ω .

Definição 5.1 (*Vetor aleatório*)

Seja $(\Omega^n, \mathcal{F}^n, P^n)$ um espaço de probabilidade. Uma função $\mathbf{X} : \Omega^n \rightarrow \mathbb{R}^n$, definida como

$$\mathbf{X}(\omega) = (X_1(\omega_1), X_2(\omega_2), \dots, X_n(\omega_n)), \quad \omega = (\omega_1, \dots, \omega_n) \in \Omega^n,$$

é chamada de variável aleatória n -dimensional se a imagem inversa de cada intervalo de dimensão n

$$\mathbf{I} = \{(x_1, x_2, \dots, x_n) : -\infty < x_i \leq a_i, a_i \in \mathbb{R}, i = 1, 2, \dots, n\}$$

pertence a \mathcal{F}^n , isto é, se

$$\mathbf{X}^{-1}(\mathbf{I}) = \{\omega \in \Omega^n : X_1(\omega_1) \leq a_1, \dots, X_n(\omega_n) \leq a_n\} \in \mathcal{F}^n.$$

Observemos que o evento $\{X_1(\omega_1) \leq a_1, \dots, X_n(\omega_n) \leq a_n\}$ representa a interseção dos eventos $\{X_1(\omega_1) \leq a_1\}, \{X_2(\omega_2) \leq a_2\}, \dots, \{X_n(\omega_n) \leq a_n\}$, ou seja,

$$\{X_1(\omega_1) \leq a_1, \dots, X_n(\omega_n) \leq a_n\} = \bigcap_{i=1}^n \{X_i(\omega_i) \leq a_i\}. \quad (5.2)$$

Nada dizemos até agora acerca de \mathbf{P}^n , o qual será definido como o vetor de funções de probabilidades

$$\mathbf{P}^n = (\underbrace{P, \dots, P}_{n \text{ vezes}}),$$

sendo cada P uma função de probabilidade em (Ω, \mathcal{F}) .

Exemplo 5.1

Os pacotes chegam a cada uma das três portas de entrada de um comutador¹ de pacotes de acordo com as tentativas independentes de Bernoulli. Cada pacote que chega é igualmente provável que seja destinado a qualquer uma das três portas de saída. Deixe onde está o número total de pacotes que chegam para a porta de saída i . X é uma variável aleatória de vetor cujos valores são determinados pelo padrão de chegadas nas portas de entrada.

Teorema 5.1

Sejam X_1, X_2, \dots, X_n n variáveis aleatórias. Então, $\mathbf{X} = (X_1, X_2, \dots, X_n)$ é um vetor aleatório de dimensão n .

Demonstração: Seja $\mathbf{I} = \{(x_1, x_2, \dots, x_n) : -\infty < x_i \leq a_i, a_i \in \mathbb{R}, i = 1, 2, \dots, n\}$. Então

$$\begin{aligned} \{(X_1, \dots, X_n) \in \mathbf{I}\} &= \{\omega \in \Omega : X_1(\omega_1) \leq a_1, X_2(\omega_2) \leq a_2, \dots, X_n(\omega_n) \leq a_n\} \\ &= \bigcap_{k=1}^n \{\omega \in \Omega : X_k(\omega) \leq a_k\} \in \mathcal{F}, \end{aligned}$$

como afirmado. ■

Em muitas aplicações, é necessário lidar com um grande número de variáveis aleatórias. Muitas vezes, o número de variáveis pode ser arbitrário. Neste capítulo, os conceitos desenvolvidos anteriormente para variáveis aleatórias únicas e pares de variáveis aleatórias são estendidos para permitir um número arbitrário de variáveis aleatórias. Grande parte do foco deste capítulo é sobre variáveis aleatórias Gaussianas multidimensionais, uma vez que a maioria das variáveis aleatórias não-Gaussianas são difíceis de lidar em muitas dimensões. Um dos nossos principais objetivos aqui é desenvolver uma notação vetorial / matriz que nos permita representar sequências potencialmente grandes de variáveis aleatórias com uma notação compacta. Muitos dos conceitos desenvolvidos no Capítulo 5 podem ser estendidos para múltiplas dimensões de uma maneira muito direta; Assim, dedicaremos tempo mínimo a esses conceitos em nossa discussão atual. Em vez disso, a atenção é focada nas idéias que exigem mais do que uma extensão trivial do trabalho feito nos capítulos anteriores.

Não surpreendentemente, muitas vezes é conveniente pensar os vetores aleatórios como elementos (aleatórios) de um espaço vetorial; isso nos permite, por exemplo, manipular vetores aleatórios através das operações de álgebra linear. Quando este for o caso, assumiremos (por defeito) que o valor aleatório é, de fato, um vetor coluna, a menos que explicitamente declarado de outra forma. Esperamos que isso fique claro no contexto.

Sabemos que toda variável aleatória induz um espaço de probabilidade e, para isto, precisamos definir os Borelianos em \mathbb{R}^n . A σ -álgebra de Borel no \mathbb{R}^n será a menor σ -álgebra contendo todo retângulo n -dimensional, ou seja, a σ -álgebra n -dimensional será a σ -álgebra gerada pelos retângulos. Como exemplo, caso tenhamos duas variáveis, observamos que qualquer região aberta A no plano pode ser escrita como união enumerável de retângulos e, portanto, faz sentido falar em calcular a probabilidade do vetor aleatório (X, Y) pertencer a A , com isto justificamos que faz sentido perguntamo-nos qual seria o valor de $P((X, Y) \in A)$.

Teorema 5.2 (*Caracterização da função de distribuição conjunta*)

Seja $F : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função que satisfaz as seguintes propriedades:

- (i) F é não decrescente em cada um das variáveis. Significa que, para a i -ésima variável, temos que, se $x \leq x_i \leq y$, então

$$F(x_1, \dots, x_{i-1}, x, x_{i+1}, \dots, x_n) \leq F(x_1, \dots, x_{i-1}, y, x_{i+1}, \dots, x_n).$$

- (ii) F é contínua à direita em cada variável. Seja $\{y_m\}$ uma sequência numérica satisfazendo que $\lim_{m \rightarrow \infty} y_m = x_i$, $i = 1, 2, \dots, n$. Então

$$\lim_{m \rightarrow \infty} F(x_1, \dots, x_{i-1}, y_m, x_{i+1}, \dots, x_n) = F(x_1, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_n).$$

- (iii) Para todo $i = 1, \dots, n$ se satisfaz que

$$\lim_{x_i \rightarrow -\infty} F(x_1, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_n) = 0$$

e

$$\lim_{x_i \rightarrow \infty} F(x_1, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_n) = 1.$$

- (iv) $\delta F \geq 0$.

Então, F é uma função de distribuição conjunta.

Demonstração: content...

■

Observemos que, no caso uni-dimensional, as três primeiras propriedades no Teorema 5.2 seriam suficientes para definir a função de distribuição. Acontece que no caso n -variado não é tão simples. O seguinte exemplo nos mostra a necessidade da propriedade (iv) na situação multivariada.

Acontece que os tipos discreto e absolutamente contínuos de variáveis aleatórias unidimensionais têm análogos no caso multivariado, definidos a seguir.

Definição 5.2 (*Vetores discretos e absolutamente contínuos*)

Seja $\mathbf{X} = (X_1, \dots, X_n)$ um vetor aleatório definido no espaço de probabilidades $(\Omega_n, \mathcal{F}_n, P_n)$

Exemplo 5.2

conteúdo...

Definição 5.3

Uma função $F : \mathbb{R} \leftarrow \mathbb{R}$ que satisfaz as propriedades anteriores é conhecida como função de distribuição n -dimensional ou n -variada.

Exemplo 5.3

conteúdo...

5.2 Distribuição conjunta

Definição 5.4 (*Distribuição conjunta*)

A função $F : \mathbb{R}^n \rightarrow [0, 1]$, definida como

$$F(x_1, x_2, \dots, x_n) = P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n), \quad \forall (x_1, x_2, \dots, x_n) \in \mathbb{R}^n,$$

é conhecida como a função de distribuição conjunta do vetor aleatório (X_1, X_2, \dots, X_n) .

Não é difícil demonstrar que as seguintes propriedades verificam-se:

- (i) $F(x_1, x_2, \dots, x_n)$ é não decrescente e contínua à direita com respeito a cada coordenada,

$$\begin{aligned} \lim_{x_1 \rightarrow -\infty} F(x_1, x_2, \dots, x_n) &= 0, \quad \forall (x_2, \dots, x_n) \\ \vdots & \\ \vdots & \end{aligned}$$

$$\lim_{x_n \rightarrow -\infty} F(x_1, x_2, \dots, x_n) = 0, \quad \forall (x_1, x_2, \dots, x_{n-1})$$

Mas, as condições (i), (ii) e (iii) não são suficientes para fazer qualquer função $F(\dots)$ uma função de distribuição. Vejamos isto no seguinte exemplo.

Exemplo 5.4

content...

5.3 Funções de vetores aleatórios

Vamos considerar primeiramente algumas funções simples de vetores aleatórios. Sejam X e Y variáveis aleatórias definidas em algum espaço de probabilidade (Ω, \mathbf{F}, P) . Definamos

Teorema 5.3

Sejam X e Y variáveis aleatórias em (Ω, \mathbf{F}, P) . Então $X + Y$ e $X \times Y$ também são variáveis aleatórias.

Demonstração: content...

■

Teorema 5.4

Sejam X e Y duas variáveis aleatórias independentes e identicamente distribuídas. Então $X - Y$ é uma variável aleatória simétrica.

Demonstração: Vamos demonstrar este teorema para o caso no qual ambas variáveis X e Y são discretas. Temos que

$$\begin{aligned}
 P(X - Y \leq x) &= \sum_{i=1}^{\infty} P(-Y \leq x - x_i) P(X = x_i) \\
 &= \sum_{i=1}^{\infty} P(-X \leq x - x_i) P(X = x_i) \\
 &= 1 - \sum_{i=1}^{\infty} P(-X > x - x_i) P(X = x_i).
 \end{aligned}$$

Assim sendo

$$\begin{aligned}
 P(X - Y \leq -x) &= 1 - \sum_{i=1}^{\infty} P(-X > -x - x_i) P(X = x_i) \\
 &= 1 - \sum_{i=1}^{\infty} P(X \leq x + x_i) P(Y = x_i) \\
 &= 1 - \sum_{i=1}^{\infty} P(X - Y < x) = 1 - P(X - Y \leq x) + P(X - Y = x),
 \end{aligned}$$

e segue-se que $X - Y$ é uma variável aleatória simétrica.

■

Exemplo 5.5

content...

Teorema 5.5

Seja (X, Y) um vetor aleatório contínuo com função de densidade f . Sejam

$$Z = X + Y, \quad U = X - Y, \quad V = X \times Y \quad \text{e} \quad W = \frac{X}{Y}$$

caso $P(Y = 0) = 0$. Então, as funções de densidade de Z , U , V e W são, respectivamente, dadas por

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{\infty} f(x, z - x) \, dx, \\ f_U(u) &= \int_{-\infty}^{\infty} f(u + y, y) \, dy, \\ f_V(v) &= \int_{-\infty}^{\infty} f(x, y/x) \frac{1}{|x|} \, dx, \\ f_W(w) &= \int_{-\infty}^{\infty} f(xw, x) |x| \, dx. \end{aligned} \tag{5.3}$$

Demonstração: Exercício. ■

5.4 Momentos de vetores aleatórios

Seja (X, Y) um vetor de variáveis aleatórias e g uma função real, Borel mensurável, definida no plano \mathbb{R}^2 .

Corolário 5.6

$$E(aX + bY) = aE(X) + bE(Y).$$

Demonstração: As variáveis aleatórias X e Y satisfazem as condições do Teorema 4.8. Sabemos que,

$$E(aX + c) = aE(X) + c \quad \text{e} \quad E(bY - c) = bE(Y) - c,$$

logo

$$E(aX + bY) = E([aX + c] + [bY - c]) = ()$$
■

5.4.1 Igualdade de variáveis aleatórias

Duas variáveis aleatórias X e Y são estritamente iguais se para cada ω se satisfaz que $X(\omega) = Y(\omega)$. Existem, no entanto, outras formas de definir igualdade.

Definição 5.5 (*Igualdade de variáveis aleatórias*)

Diz-se que duas variáveis aleatórias X e Y são:

- a) *Iguais quase certamente, escreve-se como $X \stackrel{q.c.}{=} Y$, se $P\{X = Y\} = 1$. Mais geral, um evento ocorre quase certamente se sua probabilidade é um.*
- b) *Duas variáveis são iguais em distribuição, escreve-se como $X \stackrel{d}{=} Y$, se suas correspondentes funções de distribuição coincidirem, isto é, se $F_X(x) = F_Y(x)$ para todo número real x .*

É interessante observar que a igualdade quase certa é mais forte que a igualdade em distribuição, isto é, se X e Y são iguais quase certamente, então são iguais em distribuição. No entanto, se X e Y possuem a mesma distribuição, não necessariamente são iguais quase certamente. De maneira geral, quando apareça uma expressão de igualdade entre variáveis aleatórias, será considerado que a igualdade é no sentido forte, isto é, quase certamente, caso não seja indicado o contrário.

Exemplo 5.6

Sejam X e Y duas variáveis aleatórias. Demonstrar que o conjunto $\{X = Y\}$ é um evento. Consequentemente, faz sentido calcular a probabilidade deste conjunto.

Exemplo 5.7

Demonstrar que se $X \stackrel{q.c.}{=} Y$, então $X \stackrel{d}{=} Y$. Pelo contrário, para demonstrar que se $X \stackrel{d}{=} Y$, não necessariamente $X \stackrel{q.c.}{=} Y$, bastaria apresentar um exemplo. Neste sentido, considere a variável X tal que $P(X = -1) = P(X = 1) = 1/2$, e definamos $Y = -X$.

5.4.2 Distribuições de funções de vetores aleatórios

Vamos considerar $X = (X_1, \dots, X_n)$ um vetor aleatório no espaço de probabilidade (Ω, \mathbb{F}, P) e aqui consideraremos o problema de determinar a distribuição de $Y = g(X_1, \dots, X_n)$, sendo g uma função Borel mensurável. Formalmente o problema é de solução simples, dado que a distribuição de Y é

$$F_Y(y) = P(Y \leq y) = P(g(X_1, \dots, X_n) \leq y),$$

e esta probabilidade pode ser calculada utilizando a distribuição conjunta de X_1, \dots, X_n : definindo

$$B = \{(x_1, \dots, x_n) : g(x_1, \dots, x_n) \leq y\}$$

então

$$F_Y(y) = P((X_1, \dots, X_n) \in B) = P_{X_1, \dots, X_n}(B).$$

Em outras palavras, conhecendo a distribuição conjunta de X_1, \dots, X_n , podemos obter a distribuição de qualquer função Borel mensurável do vetor X_1, \dots, X_n .

Caso X_1, \dots, X_n seja um vetor de variáveis discretas

Exemplo 5.8

content...

Exemplo 5.9

content...

5.5 Exercícios

Exercícios da Seção 5.1

1. Seja (Ω, \mathcal{F}, P) um espaço de probabilidade e sejam $A, B \in \mathcal{F}$ eventos tais que $P(A) > 0$ e $P(B) > 0$. Definamos as variáveis aleatórias X e Y como

$$X(\omega) = 1_A(\omega) \quad \text{e} \quad Y(\omega) = 1_B(\omega), \quad \forall \omega \in \Omega.$$

Prove que X e Y são independentes se, e somente se, os eventos A e B forem independentes.

2. Sejam X_1, X_2 variáveis aleatórias independentes com função de probabilidade comum

$$P(X = \pm 1) = \frac{1}{2}.$$

Definimos $X_3 = X_1 X_2$. Prove que X_1, X_2, X_3 são independentes em pares mas não independentes.

3. Sejam X e Y variáveis aleatórias independentes tais que XY é uma variável aleatória degenerada. Mostre que X e Y são também degeneradas.
4. Demonstrar o Lema 6.9.
5. Demonstrar o Teorema 6.12.
6. Demonstrar o Teorema 6.14.
7. Sejam X e Y variáveis aleatórias independentes. Mostre que se X e $X - Y$ são independentes, então X deve ser uma variável degenerada.
8. Consideremos o vetor aleatório (X_1, X_2, X_3) com função de densidade conjunta

$$f(x_1, x_2, x_3) = \begin{cases} \frac{1}{4} & \text{se } (x_1, x_2, x_3) \in A \\ 0 & \text{caso contrário,} \end{cases}$$

onde $A = \{(1, 0, 0), (0, 1, 0), (0, 0, 1), (1, 1, 1)\}$. São X_1, X_2 e X_3 independentes? São X_1, X_2 e X_3 independentes aos pares? E $X_1 + X_2$ independente de X_3 ?

9. Sejam X e Y duas variáveis aleatórias independentes. Prove que a covariância entre elas é zero, isto é, prove que $\text{Cov}(X, Y) = 0$. O inverso não é verdadeiro: duas variáveis aleatórias que têm covariância zero não são necessariamente independentes.
10. Sejam X e Y duas variáveis aleatórias independentes. Prove que se $Y = g(X)$ então $\text{Corr}[g(X), Y] = 1$ mesmo que $\text{Corr}(X, Y) = 0$.
11. Suponha que X e Y sejam variáveis aleatórias com função de densidade conjunta

$$f(x, y) = \frac{1}{\pi} \quad \text{se } x^2 + y^2 \leq 1.$$

Então o vetor (X, Y) tem distribuição Uniforme no conjunto $\{(x, y) : x^2 + y^2 \leq 1\}$.

- a) As variáveis X e Y são independentes?
 - b) Prove que $\text{Corr}(X, Y) = 0$.
 - c) Prove que $\text{Corr}(X^2, -Y^2) = 1/3$.
12. Suponha que U e R sejam duas variáveis aleatórias contínuas independentes onde U tem distribuição Uniforme no intervalo $[0, 1]$ e R tem como distribuição

$$f_R(r) = e \exp(-r^2/2), \quad \text{para } r \geq 0.$$

- a) Prove que R^2 tem a Exponencial como distribuição de probabilidades.

- b) Definamos $X = R \cos(2\pi U)$ e $Y = R \sin(2\pi U)$. Prove que X e Y são variáveis aleatórias independentes com distribuição Normal padrão.

13. Provar a afirmação no Exemplo 6.16.

14. Demonstrar o Teorema 6.15.

15. Suponha X uma variável aleatória não negativa com média $\mu > 0$ e variância $\sigma^2 < \infty$. O coeficiente de variação de X é definido como $CV(X) = \sigma/\mu$. Considere X e Y duas variáveis aleatórias não negativas com coeficientes de variação finito $CV(X)$ e $CV(Y)$, respectivamente. Prove que

$$CV(X + Y) \leq CV(X) + CV(Y).$$

16. Demonstre que se a variável aleatória X é independente de si mesma, então X é uma variável degenerada.

17. Suponhamos que X e Y são duas variáveis aleatórias discretas assumindo um número finito de valores x_1, \dots, x_m e y_1, \dots, y_n , respectivamente. Prove que X e Y são independentes se

$$P(X = x_i, Y = y_j) = P(X = x_i)P(Y = y_j)$$

para $1 \leq i \leq m - 1$ e $1 \leq j \leq n - 1$. Em outras palavras, para provar que X e Y são independentes, basta verificar que $(m - 1)(n - 1)$ equações.

18. Mostre que se as variáveis aleatórias (X, Y) são permutáveis, com densidade conjunta $f(x, y)$, então

$$P(X < Y) = P(X > Y) = \frac{1}{2},$$

com $P(X = Y) = 0$.

19. Considere as variáveis aleatórias contínuas permutáveis X_1, \dots, X_n , com função de densidade conjunta $f(x_1, \dots, x_n)$. Prove que

$$P(X_1 < X_2 < \dots < X_n) = P(X_{\pi_1} < X_{\pi_2} < \dots < X_{\pi_n}) = \frac{1}{n!},$$

com $P(X_i = X_j) = 0$ para algum par (i, j) , tal que $i \neq j$.

Exercícios da Seção 5.4

1. Seja X

Capítulo 6

Distribuição e momentos condicionais

Como de costume, o nosso ponto de partida é um experimento aleatório no qual definimos o espaço de probabilidades (Ω, \mathcal{F}, P) e sejam X e Y duas variáveis aleatórias definidas no referido espaço. Neste capítulo vamos estudar a distribuição condicional assim como a esperança condicional, conceitos de fundamental importância na probabilidade. Como vamos ver, a esperança Y , dado X é a função de X que melhor se aproxima Y , no sentido da média quadrática. Note-se que X é uma variável aleatória geral, não necessariamente assumindo valores reais.

Ferramentas mais gerais precisam ser desenvolvidas para lidar com as relações entre variáveis aleatórias dependentes. O conceito de probabilidade condicional, a distribuição de um conjunto de variáveis aleatórias dada a informação relativa aos valores observados de outro conjunto, se revelará uma ferramenta muito útil.

6.1 Distribuição condicional

Primeiro considere o problema: Qual é a probabilidade de um evento B dado que A ocorreu? Se sabemos que $w \in A$, então nosso novo espaço amostral é A e a probabilidade de B é proporcional à probabilidade de que parte dele se encontre em A . Assim, dado o espaço de probabilidade, para eventos $A, B \in \mathcal{F}$, tais que $P(A) > 0$, lembremos que a probabilidade condicional do evento B ocorrer dado que A ocorreu é dada por $P(B|A) = P(A \cap B)/P(A)$. Isso se estende imediatamente a probabilidade condicional por variáveis aleatórias, quando esta assume apenas um número enumerável de valores.

Definição 6.1

Seja X uma variável aleatória discreta assumindo valores x_1, x_2, \dots . A probabilidade condicional de A dado $X = x_k$ é definida por

$$P(A | X = x_k) = \frac{P(A, X = x_k)}{P(X = x_k)},$$

se $P(X = x_k) > 0$ e, arbitrariamente definida como zero, caso $P(X = x_k) = 0$.

Devemos lembrar que $P(X = x_k) = P(\{\omega \in \Omega : X(\omega) = x_k\})$, ou seja, tecnicamente a probabilidade condicional definida na Seção 2.3 e a probabilidade condicional de $A|X = x_k$, na Definição 6.1, é o mesmo conceito. Lembremos também que

$$P(A, X = x_k) = P(A \cap \{\omega \in \Omega : X(\omega) = x_k\}).$$

O que precisa ser feito é generalizar a definição, de modo a ser capaz de lidar com variáveis aleatórias que assumem valores não numeráveis. Veja o caso mais simples: é definida a seguir.

Definição 6.2

Suponha que a variável aleatória X seja definida em (Ω, \mathcal{F}, P) e $A \in \mathcal{F}$. Se $B \in \mathcal{B}(\mathbb{R})$ é tal que $P(X \in B) > 0$, definimos a probabilidade condicional de A dado $X \in B$ como

$$P(A|X \in B) = \frac{P(A, X \in B)}{P(X \in B)}.$$

Exemplo 6.1

Consideremos $X \sim \text{Uniforme}(-1, 1)$ e $A = \{\omega \in \Omega : X(\omega) \geq 0\}$. Queremos encontrar $P(A|X \in A^c)$.

É claro que o evento B , na Definição 6.2, é A^c nesta situação. Também sabemos que $P(A \cap A^c) = 0$, logo

$$P(A|X \in A^c) = P(A|A^c) = \frac{P(A \cap A^c)}{P(A^c)} = 0.$$

Observe que $P(A^c) = 1/2$.

Mas suponha que queremos dar sentido à probabilidade condicional de A dado $X(\omega) = x$. Claro que, se $P(X = x) > 0$, não temos problemas e procedemos como na Definição 6.1. Mas muitas das variáveis aleatórias interessantes têm a propriedade que $P(X = x) = 0$ para todo x , como as variáveis contínuas. Isto causa um alvoroço. Uma coisa óbvia a tentar é tomar limites, ou seja, tentar definir

$$P(A|X = x) = \lim_{h \downarrow 0} \frac{P(A, X \in (x - h, x + h))}{P(X \in (x - h, x + h))}. \quad (6.1)$$

Em geral, isto não é bom. Se $P(X = x_0) = 0$, então não há garantia, a menos que coloquemos condições mais restritivas em P e X , que o limite acima existirá para $x = x_0$. Portanto, ou adicionamos estas restrições, muito desagradáveis, ou olhamos para o problema de uma forma diferente. Olhe para o limite em (6.1) globalmente em função de x . Intuitivamente, parece que estamos a tentar tomar a derivada de uma medida em relação a outra. Isto tem um anel familiar; olhamos para trás para ver o que pode ser feito.

Interessa-nos agora a distribuição de probabilidade de uma variável aleatória dado o conhecimento de algum evento A . Se o evento A tiver probabilidade positiva, então podemos

definir distribuições condicionais, funções de densidade condicional, funções de probabilidade condicional e as respectivas funções marginais. Para definirmos este novo conceito seguiremos a definição de probabilidade condicional na Seção 2.3.

Definição 6.3

Sejam X e Y duas variáveis aleatórias definidas no mesmo espaço de probabilidade $(\mathbb{R}, \mathcal{B}(\mathbb{R}), P)$ e $A \in \mathcal{B}(\mathbb{R})$ de maneira que $P(X \in A) > 0$. Definimos a distribuição condicional de Y dado X , resumidamente $Y|X$, como

$$F_{Y|X}(y | X \in A) = \frac{P(\{\omega \in \mathbb{R} : X(\omega) \in A, Y(\omega) \leq y\})}{P(\{\omega \in \mathbb{R} : X(\omega) \in A\})}. \quad (6.2)$$

Nesta definição, quando escrevemos $\{\omega \in \mathbb{R} : X(\omega) \in A, Y(\omega) \leq y\}$, estamos indicando a ocorrência simultânea dos eventos envolvidos, ou seja,

$$\{\omega \in \mathbb{R} : X(\omega) \in A, Y(\omega) \leq y\} = \{\omega \in \mathbb{R} : X(\omega) \in A\} \cap \{\omega \in \mathbb{R} : Y(\omega) \leq y\}.$$

Observemos que somente temos definida distribuição condicional quando $P(X \in A) > 0$, ou seja, quando a probabilidade do evento $\{A \in \mathcal{B}(\mathbb{R})\}$ é estritamente positiva. Caso $P(X \in A) = 0$, dizemos que $F_{Y|X}(y | X \in A) = F_Y(y)$ resultado similar acontece caso $P(X \in A) = 1$, neste caso é evidente que $F_{Y|X}(y | X \in A) = F_Y(y)$.

Deduzir probabilidade condicional e depois provar que ..

Agora devemos demonstrar que a função definida em (6.2) é de fato uma medida de probabilidade. Vejamos isso no seguinte teorema.

Teorema 6.1

Sejam X e Y duas variáveis aleatórias definidas no mesmo espaço de probabilidade $(\mathbb{R}, \mathcal{B}(\mathbb{R}), P)$ e $A \in \mathcal{B}(\mathbb{R})$ de maneira que $P(X \in A) > 0$. Então,

- (i) $F_{Y|X}(y | x) \geq 0$.
- (ii) $F_{Y|X}(y | x) = 1$,

Demonstração: conteúdo...



Exemplo 6.2

content...

Exemplo 6.3

O número total de defeitos X em um chip é uma variável aleatória de Poisson com

média θ . Cada defeito tem uma probabilidade p de cair em uma região específica S e a localização de cada defeito é independente das localizações de outros defeitos. Encontre o pmf do número de defeitos Y que caem na região R . Podemos imaginar a realização de um teste de Bernoulli cada vez que um defeito ocorre com um "sucesso" ocorrendo quando o defeito cai na região R . Se o número total de defeitos for então Y é uma variável aleatória binomial com os parâmetros k e p :

Então $Y \sim \text{Poisson}(\theta p)$.

6.1.1 Distribuição truncada

Uma distribuição truncada é uma distribuição condicional que resulta da restrição do domínio da variável aleatória. As distribuições truncadas surgem em situações práticas nos casos em que a capacidade de registrar ou mesmo de ter conhecimento de, ocorrências se limita a valores superiores ou inferiores a um determinado limiar ou dentro de um intervalo especificado.

Definição 6.4

Seja X uma variável aleatória definida no espaço de probabilidade $(\mathbb{R}, \mathcal{B}(\mathbb{R}), P)$ e seja $T \in \mathcal{B}(\mathbb{R})$ tal que $P(X \in T) > 0$. Então, a distribuição condicional $P(X \leq x | X \in T)$, definida para todo x , é chamada de distribuição truncada de X .

Observemos que, na definição somente exigimos que $P(X \in T) > 0$; caso $P(X \in T) = 1$ não teremos motivos para o truncamento da variável X , isto significa que

$$P(X \leq x | X \in T) = P(X \leq x),$$

caso $P(x \in T) = 1$.

Suponhamos que temos uma variável aleatória X com função de distribuição F absolutamente contínua e seja f a função de densidade correspondente, ambas com suporte infinito. Suponhamos que desejamos saber a densidade da variável aleatória depois de restringir o suporte para estar entre duas constantes de modo que o suporte seja $(a, b]$. Ou seja, suponhamos que queremos saber como o X é distribuída, dado que $a < X \leq b$,

$$f(x | a < X \leq b) = \frac{g(x)}{F(b) - F(a)}, \quad (6.3)$$

onde $g(x) = f(x)$, para todo $x \in (a, b]$ e zero em todos os outros pontos. Similarmente no caso discreto.

Exemplo 6.4

Sejam $X \sim \text{Cauchy}(0, 1)$. Dizemos que X tem por densidade $\text{Cauchy}(0, 1)$, restrita ao intervalo $(a, b]$ ou *Cauchy padrão truncada*, se sua função de densidade for

$$f(x | a < X \leq b) = \frac{1}{\pi} \frac{(1 + x^2)^{-1}}{F(b) - F(a)}, \quad a < x \leq b$$

e F a função de distribuição Cauchy padrão que, como dada no Exemplo 3.33, assume a forma $F(x) = \frac{1}{2} + \frac{1}{\pi} \tan^{-1}(x)$, para $-\infty < x < \infty$. Podemos escrever a função de densidade Cauchy truncada de maneira geral como

$$f(x) = \frac{1}{\theta} \left(\arctan\left(\frac{b-\mu}{\theta}\right) - \arctan\left(\frac{a-\mu}{\theta}\right) \right)^{-1} \left(1 + \left(\frac{x-\mu}{\theta}\right)^2 \right)^{-1}, \quad (6.4)$$

para $-\infty < a < x \leq b < \infty$, $-\infty < \mu < \infty$ e $\theta > 0$. Encontramos então que, a função de distribuição de uma variável aleatória Cauchy truncada pelo intervalo $(a, b]$, assume a forma

$$F(x) = \frac{\arctan\left(\frac{x-\mu}{\theta}\right) - \arctan\left(\frac{a-\mu}{\theta}\right)}{\arctan\left(\frac{b-\mu}{\theta}\right) - \arctan\left(\frac{a-\mu}{\theta}\right)},$$

também para $-\infty < a < x \leq b < \infty$, $-\infty < \mu < \infty$ e $\theta > 0$.

Uma distribuição truncada onde apenas a parte inferior da distribuição foi removida é a seguinte:

$$f(x | X > a) = \frac{g(x)}{1 - F(a)},$$

onde $g(x) = f(x)$, para todo $x > a$ e zero em todos os outros pontos e $F(a)$ é a função de distribuição da variável não truncada avaliada em a . Por outro lado, uma distribuição truncada onde o topo da distribuição foi removido é a seguinte

$$f(x | X \leq b) = \frac{g(x)}{F(b)},$$

onde $g(x) = f(x)$, para todo $x \leq b$ e zero em todos os outros pontos e $F(b)$ é a função de distribuição da variável não truncada avaliada em b . Os extremos do intervalo $(a, b]$ podem eventualmente serem $\pm\infty$.

Exemplo 6.5

Seja $X \sim N(0, 1)$. Sabemos que o suporte da distribuição Normal é toda a reta real, podemos então querer restringir esta distribuição ao caso positivo, ou seja, vamos encontrar a forma da densidade Normal padrão restrita ao eixo positivo ou truncada pelo intervalo $[0, \infty)$. Pela Definição 6.4 e, mais especificamente, pelo restado em (6.3) temos que

$$f(x | 0 \leq X < \infty) = \frac{1}{\sqrt{2\pi}} \frac{e^{-x^2/2}}{1 - \Phi(0)}, \quad x > 0$$

e Φ denotando a função de distribuição Normal padrão. Como sabido $\Phi(0) = 1/2$, então

$$f(x | 0 \leq X < \infty) = \frac{2}{\sqrt{2\pi}} e^{-x^2/2}, \quad x > 0.$$

A esperança de uma variável aleatória discreta truncada pode ser escrita como

$$\text{content.} \quad (6.5)$$

É claro que podemos obter de forma similar variáveis discretas truncadas. Nesta situação o truncamento pode ser em intervalos ou mesmo num único ponto

$$\text{content.} \quad (6.6)$$

6.2 Esperança condicional

Definir distribuições condicionais quando o evento condicionado tem probabilidade 0 é muito mais difícil, mas ainda assim um problema importante. Por exemplo, se temos duas variáveis aleatórias contínuas X e Y , podemos estar interessados na distribuição condicional de Y dado $X = x$, onde $P(X = x) = 0$, já que X é contínua.

Definição 6.5

Suponha que o vetor aleatório (X_1, \dots, X_n) tenha função de densidade conjunta $g(x_1, \dots, x_n)$. A função de densidade condicional de

$$X_1, \dots, X_k \mid X_{k+1} = x_{k+1}, \dots, X_n = x_n$$

é definida como

$$f(x_1, \dots, x_k \mid x_{k+1}, \dots, x_n) = \frac{g(x_1, \dots, x_k, x_{k+1}, \dots, x_n)}{h(x_{k+1}, \dots, x_n)}, \quad (6.7)$$

desde que $h(x_{k+1}, \dots, x_n)$, a densidade conjunta de X_{k+1}, \dots, X_n , seja estritamente positiva.

Esta função de densidade condicional, vista como uma função de x_1, \dots, x_k para x_{k+1}, \dots, x_n fixos tem, as mesmas propriedades que qualquer outra função de densidade, podemos usar esta função de densidade condicional avaliar probabilidades condicionais (dado $X_{j+1} = x_{j+1}, \dots, X_k = x_k$) por integração, bem como avaliar os valores esperados condicionais.

É imediato então que a variável aleatória $E[h(X)|Y]$ satisfaz propriedades usuais da esperança. Por exemplo, o teorema a seguir mostra diversas destas propriedades.

Sejam X e Y duas variáveis aleatórias definidas no mesmo espaço de probabilidade. Então se verificam as seguintes propriedades da esperança condicional:

Teorema 6.2

- (i) $E(c|Y) = c$, onde c é uma constante qualquer;
- (ii) $E(aX + b|Y) = a E(X|Y) + b$, onde a e b são duas constantes;
- (iii)
- (iv) Se $X \leq 0$, então $E(X|Y) \leq 0$;
- (v) Se $X_1 \leq X_2$, então $E(X_1|Y) \leq E(X_2|Y)$.

Demonstração: Vejamos que ...

■

Teorema 6.3

Se $E(X^2) < \infty$, então

$$\text{Var}(X) = \text{Var}[E(X|Y)] + E[\text{Var}(X|Y)]. \quad (6.8)$$

Demonstração: A parte direita em (6.8) é igual a, por definição de variância

$$\begin{aligned} \{ E[E^2(X|Y)] - E^2[E(X|Y)] \} + E[E(X^2|Y) - E^2(X|Y)] &= \\ \{ E[E^2(X|Y)] - E^2(X) \} + E(X^2) - E[E^2(X|Y)] &= \text{Var}(X). \end{aligned}$$

■

Corolário 6.4

Se $E(X^2) < \infty$, então

$$\text{Var}(X) \geq \text{Var}[E(X|Y)],$$

e a igualdade acontece se, e somente se, X for uma função de Y .

Demonstração: Este resultado segue imediatamente de (6.8). A igualdade acontece se, e somente se,

$$E[\text{Var}(X|Y)] = E[X - E(X|Y)]^2 = 0,$$

o qual acontece se, e somente se,

$$X = E(X|Y).$$

■

Devemos lembrar que, se X e Y forem independentes,

$$F_{X|Y}(x|y) = F_X(x), \quad \forall x$$

e que

$$F_{Y|X}(y|x) = F_Y(y), \quad \forall y.$$

Segue, portanto, que se $E[h(X)]$ existir e se X e Y forem independentes, então

$$E[h(X)|Y] = E[h(X)].$$

Similarmente

$$E[h(Y)|X] = E[h(Y)],$$

desde que $E[h(Y)]$ exista e X e Y sejam independentes. Observe que a relação em (6.8) se torna trivial no caso X e Y independentes.

6.2.1 Esperança e distribuição condicionais

Estendemos aqui o conceito de probabilidade condicional em (2.11) para obter a distribuição condicional e, posteriormente, definimos esperança condicional. Estes conceitos são extremamente complicados em situações gerais, porém nos casos discretos e absolutamente contínuos são mais intuitivos, justamente estas duas situações serão nosso objeto de estudo.

Definição 6.6

Seja Y uma variável aleatória no espaço de probabilidade $(\Omega, \mathfrak{F}, P)$ e seja $A \in \mathfrak{F}$ um evento aleatório tal que $P(A) > 0$. Definimos a probabilidade condicional de Y dado o evento A por

$$P(Y \in B|A) = \frac{P([Y \in B] \cap A)}{P(A)}, \quad (6.9)$$

para $B \in \mathcal{B}$, a σ -álgebra dos Borelianos na reta.

A função definida em (6.9) define, de fato, uma função de probabilidade na reta como demonstrado no teorema a seguir.

Teorema 6.5

Seja $(\Omega, \mathfrak{S}, P)$ um espaço de probabilidade e $A \in \mathfrak{S}$ tal que $P(A) > 0$. Então $P(\cdot|A)$ é uma probabilidade nos Borelianos $\mathcal{B}(\mathbb{R})$ na reta, satisfazendo:

- (i) $P(Y \in B|A) \geq 0$ para qualquer evento $B \in \mathcal{B}$.
- (ii) $P(Y \in \mathbb{R}|A) = 1$.
- (iii) Sejam B_1, B_2, \dots Borelianos disjuntos dois a dois, então

$$P\left(Y \in \bigcup_{k=1}^{\infty} B_k | A\right) = \sum_{k=1}^{\infty} P(Y \in B_k | A).$$

Demonstração: Exercício. ■

Associado ao conceito na definição 6.6 temos as probabilidades acumuladas condicionais ou função de distribuição condicional. Podemos interpretar a distribuição condicional de Y dado o evento A como a nova distribuição que se atribui a Y quando se sabe da ocorrência do evento A .

Definição 6.7

Seja $(\Omega, \mathfrak{S}, P)$ um espaço de probabilidade e $A \in \mathfrak{S}$ tal que $P(A) > 0$. A função de distribuição associada à probabilidade condicional é chamada de função de distribuição condicional de Y dado A e definida por

$$F_{Y|A}(y|A) = P(Y \leq y|A) = \frac{P([Y \leq y] \cap A)}{P(A)}, \quad (6.10)$$

para todo $y \in \mathbb{R}$.

Definição 6.8

Seja $(\Omega, \mathfrak{S}, P)$ um espaço de probabilidade e $A \in \mathfrak{S}$ tal que $P(A) > 0$. A esperança condicional de Y dado A é a esperança da distribuição condicional, definida por

$$E(Y|A) = \int_{\mathbb{R}} y \, dF_{Y|A}(y|A), \quad (6.11)$$

se esta integral existe.

Mais geral ainda, podemos definir a esperança condicional de uma variável aleatória dada uma outra.

Definição 6.9

Sejam X e Y duas variáveis aleatórias definidas no espaço de probabilidade $(\Omega, \mathfrak{F}, P)$ e seja h uma função Borel mensurável. Consideremos que $E\{h(Y)\}$ existe. Então, a esperança condicional de $h(Y)$ dado X , escrita como $E\{h(Y)|X\}$ é uma variável aleatória assumindo os valores $E\{h(Y)|x\}$ e definida por

$$E(h(Y)|x) = \int_{\mathbb{R}} h(y) dF_{Y|X}(y|x), \quad (6.12)$$

se esta integral existe.

Uma definição semelhante pode ser dada para a esperança condicional $E\{h(X)|Y\}$, desde que $E\{h(X)\}$ exista. Os momentos da distribuição condicional são definidas da maneira usual. Assim, se $E\{Y^r\}$ existir para algum inteiro r , então $E\{Y^r|X\}$ define o momento r -ésimo da distribuição condicional. Podemos definir os momentos centrais da distribuição condicional e, em particular, a variância. Não há maiores dificuldades em generalizar estes conceitos para as distribuições n -dimensionais quando $n \geq 2$. Deixamos ao leitor fornecer os detalhes.

O problema nestas definições é como calcular as integrais em (6.11) e (6.12). Em duas situações podemos calcular esta integral sem recorrer a novos conceitos matemáticos.

Teorema 6.6

Suponhamos que g seja uma função Borel mensurável e (X, Y) um vetor aleatório. Consideremos que $E(g(X))$ exista. Então,

$$E(g(X)) = E(E(g(X)|Y)). \quad (6.13)$$

Demonstração: Vamos demonstrar para cada uma das duas principais situações:

(i) Caso (X, Y) um vetor de variáveis aleatórias discretas. Temos que

$$\begin{aligned} E(g(X)) &= \sum_y \left(\sum_x g(x) P(X = x|Y = y) \right) P(Y = y) \\ &= \sum_y \left(\sum_x g(x) P(X = x, Y = y) \right) \\ &= \sum_x g(x) \sum_y P(X = x, Y = y) = E(g(X)). \end{aligned}$$

(ii) Caso (X, Y) um vetor de variáveis aleatórias absolutamente contínuas. Temos

$$\begin{aligned} E(g(X)) &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} g(x) f(x|y) dy \right) f(y) dy \\ &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} g(x) f(x, y) dx \right) dy \\ &= \int_{-\infty}^{\infty} g(x) \left(\int_{-\infty}^{\infty} f(x, y) dy \right) dx = E(g(X)). \end{aligned}$$

■

Exemplo 6.6 (*Continuação do Exemplo 6.3*)

Vamos encontrar a $E(Y)$ no Exemplo 6.3 usando a esperança condicional.

Lembremos que para que o Teorema 6.6 seja válido consideramos que $E(g(X))$ exista, em outras palavras, estamos considerando que $E|g(X)| < \infty$. Vejamos no seguinte exemplo uma situação interessante na qual $E(g(X)) = \infty$ e, portanto, não podemos utilizar o Teorema 6.6. Mas, $E(E(g(X))|Y)$ existe e é finita.

Exemplo 6.7

6.2.2 Caso discreto

Sejam X e Y duas variáveis aleatórias discretas definidas no mesmo espaço de probabilidade $(\Omega, \mathfrak{F}, P)$. Sejam $R_X = \{x \in \mathbb{R} : p_X(x) > 0\}$ e $R_Y = \{y \in \mathbb{R} : p_Y(y) > 0\}$, onde $p_X(\cdot)$ e $p_Y(\cdot)$ denotam as funções de probabilidade marginais de X e Y , respectivamente. Logo, para cada $x \in R_X$ definimos a função de probabilidade condicional de Y dado $X = x$ como

$$p_{Y|X}(y|x) = \frac{p_{XY}(x, y)}{p_X(x)}, \quad (6.14)$$

isto segundo a definição de distribuição acumulada condicional em (6.7).

Para cada $x \in R_X$ fixo, a função em (6.14), é uma função de probabilidade devido a que

$$\sum_{y \in R_Y} p_{Y|X}(y|x) = \sum_{y \in R_Y} \frac{p_{XY}(x, y)}{p_X(x)} = \frac{1}{p_X(x)} \sum_{y \in R_Y} p_{XY}(x, y) = \frac{p_X(x)}{p_X(x)} = 1,$$

e representa a distribuição condicional de Y uma vez conhecido o valor de $X = x$.

No caso vetorial a definição é semelhante. Sejam $\mathbf{X} = \{X_1, \dots, X_k\}$ e $\mathbf{Y} = \{Y_1, \dots, Y_h\}$ vetores aleatórios discretos e $R_{\mathbf{X}} = \{\mathbf{x} \in \mathbb{R}^k : p_{\mathbf{X}}(\mathbf{x}) > 0\}$, podemos então definir

$$p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) = \frac{p_{\mathbf{XY}}(\mathbf{x}, \mathbf{y})}{p_{\mathbf{X}}(\mathbf{x})}, \quad (6.15)$$

para todo $\mathbf{x} \in \mathbb{R}^k$. Isto permite calcular as probabilidades envolvendo \mathbf{Y} quando sabemos que o evento $\{\mathbf{X} = \mathbf{x}\}$ aconteceu. De fato, se $B \in \mathbb{B}^h$ (os Borelianos em \mathbb{R}^h) definimos

$$P(\mathbf{Y} \in B | \mathbf{X} = \mathbf{x}) = \sum_{\mathbf{y} \in R_{\mathbf{Y}} \cap B} p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}). \quad (6.16)$$

Seja agora Y uma variável aleatória e \mathbf{X} um vetor aleatório de dimensão k , ambos discretos. A esperança condicional de Y dado que $\mathbf{X} = \mathbf{x}$ define-se utilizando a distribuição determinada em (6.16) e dada por

$$E(Y|\mathbf{X} = \mathbf{x}) = \sum_{y \in R_Y} y p_{Y|\mathbf{X}}(y|\mathbf{x}), \quad (6.17)$$

e este valor representa a esperança condicional da variável Y quando se conhece que o vetor \mathbf{X} assume o valor \mathbf{x} .

Observemos que se $g(\mathbf{x}) = E(Y|\mathbf{X} = \mathbf{x})$ temos que $g(\mathbf{x}) : \mathbb{R}^k \rightarrow \mathbb{R}$. Vamos definir agora uma variável aleatória que chamaremos de esperança condicional de Y dado vetor aleatório \mathbf{X} , a qual denotaremos por $E(Y|\mathbf{X})$. Esta nova variável aleatória define-se por

$$E(Y|\mathbf{X}) = g(\mathbf{X}). \quad (6.18)$$

O seguinte teorema relaciona as esperanças definidas em (6.17) e (6.18).

Teorema 6.7

Sejam X e Y duas variáveis aleatórias discretas definidas no mesmo espaço de probabilidade. Se Y tiver esperança finita, temos então que

$$E\{E(Y|\mathbf{X})\} = E(Y). \quad (6.19)$$

Demonstração: Segundo a equação em (6.18) temos que

$$E\{E(Y|\mathbf{X})\} = E(g(\mathbf{X})) = \sum_{\mathbf{x} \in \mathbb{R}^k} g(\mathbf{x}) p_{\mathbf{X}}(\mathbf{x}).$$

Utilizando o fato que $g(\mathbf{x})$ é definida por (6.17), temos que

$$\begin{aligned} E\{E(Y|\mathbf{X})\} &= \sum_{\mathbf{x} \in R_{\mathbf{X}}} \left(\sum_{y \in R_Y} y p_{Y|\mathbf{X}}(y|\mathbf{x}) \right) p_{\mathbf{X}}(\mathbf{x}) \\ &= \sum_{\mathbf{x} \in R_{\mathbf{X}}} \left(\sum_{y \in R_Y} y \frac{p_{\mathbf{X}Y}(\mathbf{x}, y)}{p_{\mathbf{X}}(\mathbf{x})} \right) p_{\mathbf{X}}(\mathbf{x}) \\ &= \sum_{\mathbf{x} \in R_{\mathbf{X}}} \left(\sum_{y \in R_Y} y p_{\mathbf{X}Y}(\mathbf{x}, y) \right) \\ &= \sum_{y \in R_Y} y \left(\sum_{\mathbf{x} \in R_{\mathbf{X}}} p_{\mathbf{X}Y}(\mathbf{x}, y) \right) \\ &= \sum_{y \in R_Y} y p_Y(y) = E(Y). \end{aligned}$$

Trocar a ordem na soma é justificado pelo fato da soma ser convergente por hipóteses. ■

Exemplo 6.8

Suponhamos que fazemos uma primeira série de n lançamentos de uma moeda e seja X o número de caras obtidas. Com base no resultado da primeira série de lançamentos, inciamos uma segunda série de X lançamentos. Seja Y o número de caras obtidas nesta segunda série de lançamentos. Calcular $E(Y)$.

Se $X = x$, a distribuição de condicional de Y dado que $X = x$ é Binomial($0.5, x$). Logo $E(Y|X = x) = 0.5x$. Construimos então a variável aleatória $E(Y|X) = g(X) = 0.5X$, e pela expressão (6.19) no Teorema 6.7 temos que $E(Y) = E\{E(Y|X)\} = 0.5E(X)$. Dado que $X \sim B(0.5, n)$ então $E(X) = 0.5n$. Concluimos então que $E(Y) = 0.25n$.

Teorema 6.8

Seja Y uma variável aleatória tal que $P(Y = c) = 1$, onde c é uma constante qualquer. Então, qualquer seja o vetor \mathbf{X} temos que

- (i) $p_{Y|\mathbf{X}}(c|\mathbf{x}) = 1$,
- (ii) $E(Y|\mathbf{X} = \mathbf{x}) = c$.

Demonstração: Exercício. ■

Exemplo 6.9

Uma urna contém três bolas vermelhas e duas bolas verdes. Uma amostra aleatória de duas bolas é atraído (a) com reposição e (b) sem reposição. Seja $X = 0$ se a primeira bola tirada é verde e $X = 1$ se a primeira bola tirada é vermelho. Definamos também $Y = 0$ se a segunda bola tirada é verde e $Y = 1$ se a segunda bola tirada é vermelha. Vamos obter as funções de probabilidade conjunta e as funções de probabilidade e esperanças condicionais.

A função de probabilidade conjunta em cada caso é apresentada nas tabelas a seguir:

As funções de probabilidade condicional correspondentes assim como as esperanças condicionais são apresentadas a seguir.

6.2.3 Caso absolutamente contínuo

Aqui veremos como ...

Também

6.3 Variáveis aleatórias independentes

Isso levanta uma questão: Se X é uma variável aleatória com função de probabilidade p_X ou de densidade f_X e se sabemos que sua média $E(X) = \mu$ e sua variância $\text{Var}(X) = \sigma^2$ então, o que mais precisamos saber para determinar p_X ou f_X completamente?

Considere uma sequência $\{a_i\}_{i=0}^{\infty}$ de números reais, uma função geradora genérica pode ser definida como

$$G(s) = \sum_{i=0}^{\infty} a_i s^i, \quad (6.20)$$

para os valores do parâmetro s nos quais a soma em (6.20) converge. Para uma dada sequência, existe um raio de convergência $R \geq 0$ tal que, a soma converge absolutamente se $|s| \leq R$ e diverge se $|s| > R$. $G(s)$ pode ainda ser diferenciada ou integrada termo a termo qualquer número de vezes quando $|s| \leq R$. Em muitas situações $G(s)$ pode ser escrita de forma fechada, em outras situações isso não será possível.

6.3.1 Independência e permutabilidade

Recordamos que a distribuição conjunta de um vector aleatório determina univocamente as distribuições marginais das variáveis aleatórias componentes mas, em geral, o conhecimento das distribuições marginais não é suficiente para determinar a distribuição conjunta. Na verdade, é perfeitamente possível ter uma coleção infinita de densidades conjuntas com determinadas densidades marginais.

Exemplo 6.10

Sejam f_1, f_2 e f_3 três funções de densidade com as correspondentes funções de distribuição F_1, F_2 e F_3 e seja α uma constante tal que $|\alpha| \leq 1$. Definimos

$$f_{\alpha}(x_1, x_2, x_3) = f_1(x_1)f_2(x_2)f_3(x_3) \times \\ \{1 + \alpha[2F_1(x_1) - 1][2F_2(x_2) - 1][2F_3(x_3) - 1]\}$$

Mostraremos que f_{α} é uma função de densidade para cada $\alpha \in [-1, 1]$ e que a coleção de densidades $\{f_{\alpha} : -1 \leq \alpha \leq 1\}$ possui as mesmas funções de densidade marginais f_1, f_2 e f_3 .

Primeiro observemos que

$$|[2F_1(x_1) - 1][2F_2(x_2) - 1][2F_3(x_3) - 1]| \leq 1,$$

de modo que

$$1 + \alpha[2F_1(x_1) - 1][2F_2(x_2) - 1][2F_3(x_3) - 1] \geq 0.$$

Também

$$\begin{aligned}
 & \int \int \int f_\alpha(x_1, x_2, x_3) dx_1 dx_2 dx_3 \\
 &= 1 + \alpha \left(\int [2F_1(x_1) - 1] f_1(x_1) dx_1 \right) \left(\int [2F_2(x_2) - 1] f_2(x_2) dx_2 \right) \\
 & \quad \left(\int [2F_3(x_3) - 1] f_3(x_3) dx_3 \right) \\
 &= 1 + \alpha \left\{ [F_1^2(x_1)]_{-\infty}^{+\infty} - 1 \right\} [F_2^2(x_2)]_{-\infty}^{+\infty} - 1 \left\{ [F_3^2(x_3)]_{-\infty}^{+\infty} - 1 \right\} = 1.
 \end{aligned}$$

Se segue que f_α é uma função da densidade. Demonstrar que f_1 , f_2 e f_3 são as funções de densidades marginais de f_α segue de forma semelhante.

Nesta seção lidaremos com uma classe muito especial de distribuições, nas quais as distribuições marginais determinam com exclusividade a distribuição conjunta de um vetor aleatória. Primeiro vamos considerar o caso bivariado.

Seja $F(x, y)$ a função de distribuição conjunta de (X, Y) e $F_1(x)$ e $F_2(y)$, as funções de distribuição marginais de X e Y , respectivamente.

Definição 6.10 (Variáveis aleatórias independentes)

Dizemos que as variáveis aleatórias X e Y são independentes se, e somente se,

$$F(x, y) = F_1(x)F_2(y) \quad \forall (x, y) \in \mathbb{R}_2. \quad (6.21)$$

Esta constitui a maneira mais geral de definir quando duas variáveis aleatórias serão consideradas independentes.

Lema 6.9

Se as variáveis aleatórias X e Y são independentes e $a < c$, $b < d$ são números reais, então

$$P(a < X \leq c, b < Y \leq d) = P(a < X \leq c)P(b < Y \leq d). \quad (6.22)$$

Demonstração: Exercício. ■

Podemos observar que a independência não é propriamente uma propriedade dos eventos ou das variáveis aleatórias envolvidas é uma propriedade da função de probabilidade.

Exemplo 6.11

Consideremos duas variáveis aleatórias discretas X e Y com função de probabilidade

conjunta dada na tabela a seguir. Observemos que X e Y não são independentes, pois

$$\frac{1}{5} = P(X = 2, Y = 2) \neq P(X = 2)P(Y = 2) = \frac{9}{25}.$$

Basta identificar um par de valores das variáveis nos quais não se satisfaz a Definição 6.3 para poder concluir que as variáveis X e Y não são independentes.

Teorema 6.10

Sejam X e Y duas variáveis aleatórias:

- (a) A condição necessária e suficiente para que as variáveis aleatórias discretas X e Y sejam independentes é que

$$P(X = x_i, Y = y_i) = P(X = x_i)P(Y = y_i)$$

para todo par (x_i, y_i) .

- (b) Duas variáveis aleatórias X e Y absolutamente contínuas são independentes se, e somente se,

$$f(x, y) = f_1(x)f_2(y)$$

para todo par $(x, y) \in \mathbb{R}_2$ onde f é a função de densidade conjunta e f_1, f_2 as funções de densidade marginais de X e Y , respectivamente.

Demonstração: (a) Sejam X e Y independentes. No Lema 6.9, fazendo $a \rightarrow c$ e $b \rightarrow d$ obtemos

$$P(X = c, Y = d) = P(X = c)P(Y = d).$$

Reciprocamente,

$$F(x, y) = \sum_B P(X = x_i, Y = y_i),$$

onde $B = \{(i, j) : x_i \leq x, y_j \leq y\}$. Então

$$\begin{aligned} F(x, y) &= \sum_B P(X = x_i)P(Y = y_i) \\ &= \sum_{x_i \leq x} \left\{ \sum_{y_j \leq y} P(Y = y_j) \right\} P(X = x_i) = F(x)F(y). \end{aligned}$$

(b) Exercício. ■

Corolário 6.11

Sejam X e Y variáveis aleatórias independentes. Então $F_{Y|X}(y|x) = F_Y(y)$, para todo y e $F_{X|Y}(x|y) = F_X(x)$ para todo x .

Voltando ao Lema 6.9. Recordemos novamente que todo conjunto de Borel na reta real pode ser obtido por um número enumerável de operações de uniões, interseções e diferenças de intervalos semi-fechados.

Teorema 6.12

As variáveis aleatórias X e Y são independentes se, e somente se

$$P(X \in A_1, Y \in A_2) = P(X \in A_1)P(Y \in A_2), \quad (6.23)$$

para todo conjunto de Borel A_1 no eixo x e todo conjunto de Borel A_2 no eixo y .

Demonstração: Exercício. ■

Teorema 6.13

Sejam X e Y variáveis aleatórias independentes e f e g duas funções Borel medíveis. Então $f(X)$ e $g(Y)$ são também variáveis aleatórias independentes.

Observe que o Teorema 6.12 é numa situação particular do Teorema 6.13.

Demonstração: Temos que

$$\begin{aligned} P(\{f(X) \leq x, g(Y) \leq y\}) &= P(\{X \in f^{-1}(-\infty, x], Y \in g^{-1}(-\infty, y]\}) \\ &= P(\{X \in f^{-1}(-\infty, x]\}P\{Y \in g^{-1}(-\infty, y]\}) \\ &= P(f(X) \leq x)P(g(Y) \leq y). \end{aligned} \quad \blacksquare$$

Observe que a variável aleatória degenerada é independente de qualquer outra variável aleatória.

Exemplo 6.12

Sejam X e Y variáveis aleatórias com função de densidade conjunta

$$f(x, y) = \begin{cases} \frac{1 + xy}{4}, & |x| < 1, |y| < 1, \\ 0, & \text{caso contrário} \end{cases}$$

Então X e Y não são independentes uma vez que $f_1(x) = 1/2$, quando $|x| < 1$ e $f_2(y) =$

$1/2$, quando $|y| < 1$ são as densidades marginais de X e Y , respectivamente. Contudo, as variáveis X^2 e Y^2 são independentes. De fato,

$$\begin{aligned}
 P(X^2 \leq u, Y^2 \leq \nu) &= \int_{-\nu^{1/2}}^{\nu^{1/2}} \int_{-u^{1/2}}^{u^{1/2}} f(x, y) \, dx \, dy \\
 &= \frac{1}{4} \int_{-\nu^{1/2}}^{\nu^{1/2}} \left[\int_{-u^{1/2}}^{u^{1/2}} (1 + xy) \, dx \right] dy \\
 &= u^{1/2} \nu^{1/2} \\
 &= P(X^2 \leq u) P(Y^2 \leq \nu).
 \end{aligned}$$

Consideremos agora situações mais gerais que envolvam mais do que duas variáveis aleatórias e, inclusive, sequências infinitas de variáveis aleatórias. Algumas das demonstrações de resultados que serão apresentados a continuação constituem simples extensões daqueles já mostrados, por isso, o leitor será instruído a demonstra-los como exercício.

Começemos com coleções finitas de variáveis aleatórias e depois generalizamos no caso infinito enumerável.

Definição 6.11

A coleção de variáveis aleatórias X_1, X_2, \dots, X_n é dita ser mutuamente ou completamente independente se, e somente se

$$F(x_1, \dots, x_n) = \prod_{i=1}^n F_i(x_i), \quad \forall (x_1, \dots, x_n) \in \mathbb{R}_n, \quad (6.24)$$

onde F é a função de distribuição conjunta de (X_1, \dots, X_n) e F_i é a função de distribuição marginal de X_i , $i = 1, \dots, n$. As variáveis X_1, \dots, X_n são ditas independentes em pares se, e somente se, qualquer par delas forem independentes.

Um resultado análogo ao Teorema 6.10 é possível, mas por constituir basicamente uma repetição do mesmo instruímos ao leitor a considera-lo como verídico e dedicaremos este espaço para outros resultados.

Exemplo 6.13

No exemplo 6.10 não podemos escrever

$$f_\alpha(x_1, x_2, x_3) = f_1(x_1) f_2(x_2) f_3(x_3),$$

a menos que $\alpha = 0$. Segue então que X_1, X_2 e X_3 não são independentes excepto quando $\alpha = 0$.

Os seguintes resultados são fáceis de serem demonstrados.

Teorema 6.14

Se X_1, \dots, X_n forem variáveis aleatórias independentes, qualquer subcoleção X_{i_1}, \dots, X_{i_k} de X_1, \dots, X_n são também independentes.

Demonstração: Exercício. ■

Exemplo 6.14

É bastante possível para variáveis aleatórias X_1, \dots, X_n serem independentes aos pares sem serem mutuamente independentes. É o caso vetor aleatório (X, Y, Z) com a seguinte função de probabilidade conjunta

$$P(X = x, Y = y, Z = z) = \begin{cases} \frac{3}{16} & \text{se } (x, y, z) \in \{(0, 0, 0), (0, 1, 1), \\ & (1, 0, 1), (1, 1, 0)\} \\ \frac{1}{16} & \text{se } (x, y, z) \in \{(0, 0, 1), (0, 1, 0), \\ & (1, 0, 0), (1, 1, 1)\} \end{cases}$$

Percebe-se que X, Y e Z não são mutuamente independentes, devido a

$$P(X = x, Y = y) = \frac{1}{4}, \quad (x, y) \in \{(0, 0), (0, 1), (1, 0), (1, 1)\}$$

$$P(Y = y, Z = z) = \frac{1}{4}, \quad (y, z) \in \{(0, 0), (0, 1), (1, 0), (1, 1)\}$$

$$P(X = x, Z = z) = \frac{1}{4}, \quad (x, z) \in \{(0, 0), (0, 1), (1, 0), (1, 1)\}$$

$$P(X = x) = \frac{1}{2}, \quad x = 0, x = 1$$

$$P(Y = y) = \frac{1}{2}, \quad y = 0, y = 1$$

$$P(Z = z) = \frac{1}{2}, \quad z = 0, z = 1$$

Segue então que X e Y , Y e Z e X e Z são independentes aos pares.

Exemplo 6.15

Um outro exemplo para mostrar que verificar independência em pares é insuficiente para garantir a independência. Sejam X_1, X_2 e X_3 independentes tais que $P(X_i = +1) = P(X_i = -1) = 1/2$. Sejam $Y_1 = X_2X_3$, $Y_2 = X_1X_3$ e $Y_3 = X_1X_2$. Então Y_1, Y_2 e Y_3 são independentes em pares, mas não mutuamente independentes.

Definição 6.12

As variáveis aleatórias na sequência $\{X_n\}$ são ditas independentes se para qualquer $n = 2, 3, 4, \dots$ as variáveis aleatórias X_1, \dots, X_n são independentes.

Exemplo 6.16

Pode-se provar que n variáveis aleatórias X_1, \dots, X_n são mutuamente independentes se, e somente se

$$\mathbb{E} \left[\prod_{i=1}^n g_i(X_i) \right] = \prod_{i=1}^n \mathbb{E}[g_i(X_i)], \quad (6.25)$$

para quaisquer n funções g_1, \dots, g_n Borel mensuráveis, sempre que os valores esperados acima existam e estejam bem definidos.

Definição 6.13

Dizemos que as variáveis aleatórias X e Y são identicamente distribuídas se X e Y possuem a mesma função de distribuição, isto é,

$$F_X(x) = F_Y(y) \quad \forall x, y \in \mathbb{R},$$

onde F_X e F_Y são as funções de distribuição de X e Y , respectivamente.

Definição 6.14

Dizemos que $\{X_n\}$ é uma sequência de variáveis aleatórias independentes e identicamente distribuídas (resumidamente iid) com função de distribuição comum F_X , se $\{X_n\}$ é uma sequência de variáveis aleatórias independentes e a distribuição de X_n ($n = 1, 2, \dots$) é a mesma do que a distribuição de X .

De acordo com a Definição 6.13, X e Y são identicamente distribuídas se, e somente se, tiverem a mesma distribuição. Isso não significa que $X = Y$ com probabilidade 1. Se $P(X = Y) = 1$, dizemos que X e Y são variáveis aleatórias equivalentes. Tudo que a Definição 6.13 nos disse é que X e Y são identicamente distribuídas se, e somente se

$$P(X \in A) = P(Y \in A), \quad \forall A \in \mathbb{B},$$

os Borelianos na reta real. Nada é dito acerca da igualdade dos eventos $\{X \in A\}$ e $\{Y \in A\}$.

Definição 6.15

Dois vetores aleatórios (X_1, X_2, \dots, X_m) e (Y_1, Y_2, \dots, Y_n) são ditos independentes se

$$F(x_1, x_2, \dots, x_m, y_1, y_2, \dots, y_n) = F_1(x_1, x_2, \dots, x_m) F_2(y_1, y_2, \dots, y_n) \quad (6.26)$$

para todo $(x_1, x_2, \dots, x_m, y_1, y_2, \dots, y_n) \in \mathbb{R}_{n+m}$ onde F , F_1 e F_2 são as funções de distribuição conjuntas dos vetores $(X_1, \dots, X_m, Y_1, \dots, Y_n)$, (X_1, \dots, X_m) e (Y_1, \dots, Y_n) , respectivamente.

Logicamente, a independência dos vetores (X_1, \dots, X_m) e (Y_1, \dots, Y_n) não implica na independência de suas respectivas componentes.

Exemplo 6.17

Seja (X, Y) um vetor aleatório com função de densidade conjunta

$$f(x, y) = \frac{1}{4}[1 + xy(x^2 - y^2)], \quad \text{se } |x| \leq 1, |y| \leq 1,$$

e seja (U, V) um outro vetor aleatório com função de densidade conjunta

$$g(u, v) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)}(u^2 - 2\rho uv + v^2) \right\},$$

quando $-\infty < u < \infty$ e $-\infty < v < \infty$ e $|\rho| < 1$ é uma constante conhecida. A função de densidade conjunta de (X, Y, U, V) é dada por

$$h(x, y, u, v) = \frac{1}{8\pi\sqrt{1-\rho^2}}[1 + xy(x^2 - y^2)] \times \\ \exp \left\{ -\frac{1}{2(1-\rho^2)}(u^2 - 2\rho uv + v^2) \right\},$$

para $|x| \leq 1, |y| \leq 1, -\infty < u, v < \infty$. Então (X, Y) e (U, V) são independentes mas X e Y não o são, assim como também U e V não são independentes.

Teorema 6.15

Sejam $\mathbf{X} = (X_1, \dots, X_m)$ e $\mathbf{Y} = (Y_1, \dots, Y_n)$ dois vetores aleatórios independentes. Então a componente X_j de \mathbf{X} , $j = 1, 2, \dots, m$ e a componente Y_k de \mathbf{Y} , $k = 1, 2, \dots, n$ são variáveis aleatórias independentes. Se h e g são duas funções Borel medíveis então $h(X_1, \dots, X_m)$ e $g(Y_1, \dots, Y_n)$ são duas variáveis aleatórias independentes.

Demonstração: Exercício. ■

É possível que uma variável aleatória X seja independente de Y e também de Z mas X não seja independente do vetor aleatório (Y, Z) . Veja o Exemplo 6.14.

Teorema 6.16

Seja K um conjunto arbitrário e $I_k, k \in K$, arbitrários conjuntos indicadores mutuamente disjuntos. Definamos $I = \bigcap_{k \in K} I_k$. Se as variáveis aleatórias na família $\{X_k\}_{k \in I}$ forem independentes, então as σ -álgebras na família $\{\sigma(X_j)\}_{j \in I_k, k \in K}$ são independentes.

Demonstração: Definamos o conjunto

■

Concluimos esta parte mencionando que a noção de independência de vetores aleatórios pode ser estendida a qualquer número de vetores aleatórios. Vejamos agora um conceito relacionado.

Definição 6.16

Dizemos que as variáveis aleatórias X_1, \dots, X_n são permutáveis ou simétricas se a distribuição conjunta $F(x_1, \dots, x_n)$ é simétrica, isto é, se F é invariante para permutações de seus argumentos, de modo que

$$F(x_{\pi_1}, \dots, x_{\pi_n}) = F(x_1, \dots, x_n), \quad (6.27)$$

para toda permutação (π_1, \dots, π_n) do vetor $(1, \dots, n)$.

Observemos que se X_1, \dots, X_n são independentes e igualmente distribuídas então estas variáveis serão também permutáveis. O contrário não é verdade, por isso, o conceito de variáveis permutáveis é mais fraco do que o de independência.

Por outro lado, se X e Y forem permutáveis então são identicamente distribuídas, mas não o contrário. Afirmamos então que permutabilidade é um conceito mais forte do que distribuição idêntica.

Exemplo 6.18

Sejam as X e Y variáveis aleatórias assumindo valores no mesmo espaço amostral $\{0, 1, 2\}$ com função de probabilidade conjunta. Verificando as distribuições marginais mostra-se que X e Y estão distribuídas de forma idêntica, mas

$$P(X = 0, Y = 2) = 1/7 \neq P(Y = 0, X = 2) = 0/7.$$

Isto é, as variáveis aleatórias X e Y têm a mesma distribuição, mas os vectores aleatórios (X, Y) e (Y, X) não.

Exemplo 6.19

Suponha que uma urna contém 10 bolas vermelhas e 15 bolas verdes. Seja X_K o número de bolas vermelhas na k -ésima retirada sem substituição, para $k = 1, \dots, 12$. Se você tiver 10

bolas vermelhas nas primeiras 10 retiradas, então você sabe que a 11 será uma bola verde, de modo que não é independente. Mas não é difícil mostrar que eles são permutáveis. Ao contrário do exemplo anterior, este não pode ser alargada a uma infinitamente longo sequência de variáveis aleatórias permutáveis.

Teorema 6.17

Sejam X_1, \dots, X_n variáveis aleatórias independentes com funções geradoras de probabilidade $G_{X_i}(s)$, $i = 1, \dots, n$, respectivamente. Então

$$G_{X_1+\dots+X_n}(s) = \prod_{i=1}^n G_{X_i}(s).$$

Demonstração: Sendo X_1, \dots, X_n independentes,

$$G_{X_1+\dots+X_n}(s) = E(s^{X_1+\dots+X_n}) = E(s^{X_1} \times \dots \times s^{X_n}) = \prod_{i=1}^n E(s^{X_i}) = \prod_{i=1}^n G_{X_i}(s). \quad \blacksquare$$

Exemplo 6.20

Encontramos que a função geradora de probabilidade de uma variável aleatória X com distribuição $Poisson(\lambda)$ é $G_X(s) = e^{-\lambda(1-s)}$. Utilizando esta função encontremos a esperança e variância de X . Derivando uma vez obtemos $G'_X(s) = \lambda e^{-\lambda(1-s)}$ e avaliando em $s = 1$ temos, $E\{X\} = G'_X(1) = \lambda$. Derivando pela segunda vez, $G''_X(s) = \lambda^2 e^{-\lambda(1-s)}$ e em $s = 1$ obtemos $E\{X(X-1)\} = G''_X(1) = \lambda^2$. Portanto, $\text{Var}\{X\} = E\{X^2\} - E^2\{X\} = \lambda^2 + \lambda - \lambda^2 = \lambda$.

Devido esta segunda propriedade, a função geradora de probabilidade também é conhecida como função geradora de momentos fatoriais. Mostramos agora a utilização desta função para determinar a distribuição de uma variável aleatória, o procedimento é elegante e simples.

Exemplo 6.21

Suponhamos que X e Y são variáveis aleatórias independentes com distribuição $Poisson(\lambda_1)$ e $Poisson(\lambda_2)$, respectivamente. Então

$$\begin{aligned} G_{X+Y}(s) &= G_X(s)G_Y(s) \\ &= \exp\{-\lambda_1(1-s)\} \exp\{-\lambda_2(1-s)\} \\ &= \exp\{-(\lambda_1 + \lambda_2)(1-s)\}. \end{aligned}$$

Esta expressão corresponde à função geradora de probabilidade da distribuição $Poisson$ com parâmetro $\lambda_1 + \lambda_2$. Devido à unicidade, podemos afirmar que $X + Y$ tem distribuição $Poisson(\lambda_1 + \lambda_2)$.

Exemplo 6.22

Sejam X_1, \dots, X_n variáveis aleatórias independentes *Bernoulli*(θ), isto é,

$$P\{X_i = 0\} = 1 - \theta \quad \text{e} \quad P\{X_i = 1\} = \theta, \quad i = 1, \dots, n.$$

Seja

$$X = \sum_{i=1}^n X_i.$$

Encontremos a distribuição de probabilidade de X . Primeiramente observemos que

$$G_X(s) = G_{X_1}(s) \times \dots \times G_{X_n}(s),$$

e cada função geradora de probabilidade é da forma

$$G_{X_i}(s) = 1 - \theta + \theta s,$$

para $i = 1, \dots, n$. Então

$$\begin{aligned} G_X(s) &= (1 - \theta + \theta s)^n = \sum_{x=0}^n \binom{n}{x} (\theta s)^x (1 - \theta)^{n-x} \\ &= \sum_{x=0}^n \binom{n}{x} \theta^x (1 - \theta)^{n-x} s^x. \end{aligned}$$

Então

$$P\{X = x\} = \binom{n}{x} \theta^x (1 - \theta)^{n-x},$$

para $x = 0, 1, \dots$ exatamente igual ao x -ésimo coeficiente da soma $G_X(s)$, ou seja, X tem como distribuição *Binomial*(n, θ).

Teorema 6.18

Sejam N, X_1, X_2, \dots variáveis aleatórias independentes e X_1, X_2, \dots variáveis aleatórias identicamente distribuídas, cada uma com função geradora de probabilidade $G_X(s)$. Então

$$S_N = X_1 + \dots + X_N,$$

tem função geradora de probabilidade da forma

$$G_{S_N}(s) = G_N(G_X(s)).$$

Demonstração :

$$\begin{aligned}
 G_{S_N}(s) &= E\{s^{S_N}\} \\
 &= \sum_{n=0}^{\infty} E\{s^{S_N} | N = n\} P\{N = n\} \\
 &= \sum_{n=0}^{\infty} G_{X_1+\dots+X_n}(s) P\{N = n\} \\
 &= \sum_{n=0}^{\infty} G_X^n(s) P\{N = n\} \\
 &= G_N(G_X(s)). \quad \blacksquare
 \end{aligned}$$

A definição de função geradora de probabilidade pode ser estendida ao caso de vetores aleatórios da seguinte forma. A função geradora de probabilidade do vetor (X, Y) é a função $G_{X,Y}(s, t) = E\{s^X t^Y\}$, para valores reais de s e t onde esta esperança seja absolutamente convergente. Pode demonstrar-se que as variáveis X e Y são independentes se, e somente se, $G_{X,Y}(s, t) = G_X(s)G_Y(t)$. O caso de vetores de dimensão maior do que dois é análogo.

Novamente é simples demonstrar que a função geradora de momentos da soma de variáveis aleatórias independentes é o produto das funções geradoras de momentos individuais.

Teorema 6.19

Sejam X e Y duas variáveis aleatórias independentes de funções geradoras de momentos que existem numa vizinhança não trivial ao redor do zero. Então para qualquer $s \in (-t, t)$ para algum $t > 0$,

$$M_{X+Y}(s) = M_X(s) \times M_Y(s).$$

Demonstração :

$$M_{X+Y}(s) = E\{e^{s(X+Y)}\} = E\{e^{sX} e^{sY}\} = E\{e^{sX}\} E\{e^{sY}\} = M_X(s) \times M_Y(s). \quad \blacksquare$$

É interessante observar que a condição $M_{X+Y}(s) = M_X(s) \times M_Y(s)$ não é suficiente para concluir que as variáveis aleatórias X e Y são independentes.

Exemplo 6.23

Seja (X, Y) um vetor aleatório com função de densidade conjunta

$$f(x, y) = \frac{1 + xy(x^2 - y^2)}{4}, \quad \text{para } -1 < x, y < 1.$$

Demonstremos X e Y não são independentes e, no entanto, se satisfaz a identidade $M_{X+Y}(s) = M_X(s) \times M_Y(s)$.

A desigualdade de Hoeffding¹. Na teoria da probabilidade, a desigualdade de Hoeffding

¹Wassily Hoeffding (1914-1991) foi um estatístico e probabilista finlandês. Hoeffding foi um dos fundadores das estatísticas não paramétricas, nas quais Hoeffding contribuiu com a ideia e os resultados básicos das estatísticas U .

fornece um limite superior sobre a probabilidade de a soma das variáveis aleatórias se desviar do seu valor esperado é semelhante em espírito à desigualdade de Markov, mas é uma desigualdade mais acentuada. Começamos com o seguinte resultado importante.

Teorema 6.20 (*Desigualdade Hoeffding*)

Suponha que X seja uma variável aleatória satisfazendo $a \leq X \leq b$. Então,

$$\mathbb{E}(e^{tX}) \leq e^{t\mu} e^{\frac{t^2(b-a)^2}{8}},$$

onde $\mu = \mathbb{E}(X)$.

Demonstração :

$$\mathbb{E}(X) = c \cdot \text{Var}(X) = \mathbb{E}((X - c)^2) = \mathbb{E}(0) = 0.$$



6.4 Variância condicional

Assim como definimos a expectativa condicional de X , dado o valor de Y , também podemos definir a variância condicional de X , dado que $Y = y$, que é definido como segue.

Definição 6.17 (*Variância condicional*)

Sejam X e Y duas variáveis aleatórias definidas no mesmo espaço de probabilidade. Definimos por $\text{Var}(X|Y)$ a variância condicional de X dado $Y = y$, como sendo

$$\text{Var}(X|Y) = E\left((X - E(X|Y))^2|Y\right),$$

sempre que esta esperança exista.

Isto é, $\text{Var}(X|Y)$ é igual ao quadrado de expectativa condicional da diferença entre X e sua esperança condicional quando o valor de Y é dado. Ou, em outras palavras, $\text{Var}(X|Y)$ é exatamente análogo à definição usual de variância, mas agora todas as esperanças são condicionais sobre o fato de que Y é conhecido.

Exemplo 6.24

Seja (X, Y) um vetor aleatório com função de densidade conjunta

$$f_{X,Y}(x, y) = \begin{cases} 2, & \text{se } 0 < x < y < 1 \\ 0, & \text{caso contrário} \end{cases}$$

Então

$$f_X(x) = \int_0^1 2 \, dy = \begin{cases} 2 - 2x, & \text{se } 0 < x < 1 \\ 0, & \text{caso contrário} \end{cases}$$

e

$$f_Y(y) = \int_0^1 2 \, dx = \begin{cases} 2y, & \text{se } 0 < y < 1 \\ 0, & \text{caso contrário} \end{cases},$$

são as funções de densidade marginais.

Existe uma relação muito grande entre $\text{Var}(X)$, a variância incondicional de X e $\text{Var}(X|Y)$, a variância condicional de X dada por Y , que pode ser aplicada para computar $\text{Var}(X)$. A relação mencionada está apresentada a continuação.

Teorema 6.21

Suponhamos que $E(X^2) < \infty$. Então, a variância condicional, satisfaz que

$$\text{Var}(X) = E(\text{Var}(X|Y)) + \text{Var}(E(X|Y)). \quad (6.28)$$

Demonstração: Para obter a relação em (6.28), observe primeiro que, pelo mesmo raciocínio que produz $\text{Var}(X) = E(X^2) - E^2(X)$, temos que

$$\text{Var}(X|Y) = E(X^2|Y) - E^2(X|Y), \quad (6.29)$$

de maneira que

$$\begin{aligned} E(\text{Var}(X|Y)) &= E(E(X^2|Y)) - E(E^2(X|Y)) \\ &= E(X^2) - E(E^2(X|Y)). \end{aligned} \quad (6.30)$$

Também, dado que $E(E(X|Y)) = E(X)$, temos que

$$\text{Var}(E(X|Y)) = E(E(X|Y)^2) - E^2(X). \quad (6.31)$$

Consequentemente, utilizando os resultados em (6.30) e (6.31) obtemos a expressão em (6.28). ■

Exemplo 6.25

Suponha que a qualquer momento t o número de pessoas que chegaram a uma estação ferroviária é uma variáveis aleatórias com distribuição $\text{Poisson}(\lambda t)$. Se o trem inicial chega à estação em um momento, independente de quando os passageiros chegam, que é uniformemente distribuído sobre $(0, T)$, qual é a média e variância do número de passageiros que entram no trem?

Seja $N(t)$, para cada $t \geq 0$, o número de chegadas no tempo t e seja Y o tempo ao qual o trem chega. A variável aleatória de interesse é, então, $N(Y)$. Condicionando em Y temos

$$\begin{aligned} E(N(Y)|Y = t) &= E(N(t)|Y = t) \\ &= E(N(t)) \quad \text{pela independência de } Y \text{ e } N(t) \\ &= \lambda t \quad \text{desde que } N(t) \text{ é Poisson com média } \lambda t \end{aligned}$$

Por consequência

$$E(N(Y)|Y) = \lambda Y,$$

então, calculando as esperanças temos

$$E(N(Y)) = \lambda E(Y) = \frac{\lambda T}{2}.$$

Para obtermos $\text{Var}(N(Y))$ utilizamos a expressão da variância condicional, da forma,

$$\begin{aligned} \text{Var}(N(Y) | Y = t) &= \text{Var}(N(t) | Y = t) \\ &= \text{Var}(N(t)) \quad \text{pela independência de } Y \text{ e } N(t) \\ &= \lambda t, \end{aligned}$$

de maneira que $\text{Var}(N(Y)|Y) = \lambda Y$ e $E(N(Y)|Y) = \lambda Y$. Então, da expressão da variância condicional

$$\begin{aligned} \text{Var}(N(Y)) &= E(\lambda Y) + \text{Var}(\lambda Y) \\ &= \lambda \frac{T}{2} + \lambda^2 \frac{T^2}{12}, \end{aligned}$$

do qual obtemos que $\text{Var}(Y) = T^2/12$.

Exemplo 6.26

content...

6.5 Distribuição normal multivariada

De maneira geral, dizemos que um vetor aleatório tem distribuição normal multivariada se possuir a mesma distribuição de uma transformação linear de variáveis aleatórias normal padrão independentes. Significa o seguinte: se X_1, X_2, \dots, X_n forem independentes e tiverem distribuição normal padrão então, o vetor aleatório $Y = (Y_1, \dots, Y_n)$, definida cada componente como

$$Y_i = a_{i,1}X_1 + \dots + a_{i,n}X_n + \mu_i,$$

para $i = 1, \dots, n$, possui distribuição normal multivariada, sendo as constantes $a_{i,1}, \dots, a_{i,n}$ e μ_1, \dots, μ_n números reais. Matricialmente escrevemos

$$Y = AX^\top + \mu, \quad (6.32)$$

onde

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ a_{2,1} & a_{2,2} & \dots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \dots & a_{n,n} \end{pmatrix}, \quad X^\top = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} \quad \text{e} \quad \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix}.$$

Podemos observar que cada Y_i , sendo uma combinação linear de normais independentes, também tem distribuição normal. De fato,

$$Y_i \sim N\left(\mu_i, \sum_{k=1}^n a_{i,k}^2\right). \quad (6.33)$$

Além disso, a matriz de variâncias e covariâncias do vetor Y é $\Sigma = AA^\top$, isto é,

$$\Sigma = (\text{Cov}(Y_i, Y_j) : 1 \leq i, j \leq n),$$

devido a

$$\begin{aligned} \text{Cov}(Y_i, Y_j) &= E((Y_i - \mu_i)(Y_j - \mu_j)) \\ &= E((a_{i,1}X_1 + \dots + a_{i,n}X_n)(a_{j,1}X_1 + \dots + a_{j,n}X_n)) \\ &= \sum_{k=1}^n \sum_{l=1}^n a_{i,k}a_{j,l} E(X_kX_l) \\ &= \sum_{k=1}^n a_{i,k}a_{j,k} = (AA^\top)_{i,j}. \end{aligned}$$

Definição 6.18 (*Distribuição Normal multivariada*)

Sejam X_1, X_2, \dots, X_n variáveis aleatórias independentes identicamente distribuídas $N(0, 1)$ e seja $Y = AX^\top + \mu$, como em (6.32). Dizemos que $Y \sim N(\mu, \Sigma)$, tem distribuição normal n -variada com esperança μ e matriz de variâncias e covariâncias Σ .

6.6 Modelos Copula

Durante muito tempo os estatísticos têm se interessado na relação entre uma função de distribuição multivariada e suas distribuições marginais. Féchet (1951); Fréchet (1958), Dall’Aglio (1959) e outros pesquisadores fizeram alguns trabalhos interessantes sobre o assunto na década de 1950, estudando as funções bivariadas e trivariadas com determinadas funções marginais univariadas. Recordemos que a distribuição conjunta de um vetor aleatório determina unicamente as distribuições marginais das variáveis aleatórias componentes mas, em geral, o conhecimento das distribuições marginais não é suficiente para determinar a distribuição conjunta. Na verdade, é perfeitamente possível ter uma coleção infinita de densidades conjuntas com determinadas densidades marginais.

Vejamos o exemplo a seguir, apresentado por Gumbel (1958).

Exemplo 6.27

Sejam f_1 , f_2 e f_3 funções de densidade com funções de distribuição correspondentes F_1 , F_2 e F_3 . Seja ainda α uma constante $|\alpha| \leq 1$. Definamos

$$f_\alpha(x_1, x_2, x_3) = f_1(x_1)f_2(x_2)f_3(x_3) \left(1 + \alpha(2F_1(x_1) - 1)(2F_2(x_2) - 1)(2F_3(x_3) - 1)\right).$$

Mostraremos que f_α é uma função de densidade conjunta para cada $\alpha \in [-1, +1]$ e que a coleção de densidades conjuntas $\{f_\alpha : -1 \leq \alpha \leq +1\}$ têm as mesmas funções de densidade marginais f_1 , f_2 e f_3 .

Primeiro observe que

$$|(2F_1(x_1) - 1)(2F_2(x_2) - 1)(2F_3(x_3) - 1)| \leq 1,$$

de maneira que

$$1 + \alpha(2F_1(x_1) - 1)(2F_2(x_2) - 1)(2F_3(x_3) - 1) \geq 0.$$

Também

$$\begin{aligned} \int \int \int f_\alpha(x_1, x_2, x_3) dx_1 dx_2 dx_3 &= 1 + \\ &\alpha \left(\int (2F_1(x_1) - 1) f_1(x_1) dx_1 \right) \left(\int (2F_2(x_2) - 1) f_2(x_2) dx_2 \right) \left(\int (2F_3(x_3) - 1) f_3(x_3) dx_3 \right) \\ &= 1 + \alpha \left(\left(F_1^2(x_1) - 1 \right) \Big|_{-\infty}^{+\infty} \left(F_2^2(x_2) - 1 \right) \Big|_{-\infty}^{+\infty} \left(F_3^2(x_3) - 1 \right) \Big|_{-\infty}^{+\infty} \right) = 1. \end{aligned}$$

Disto segue que f_α é função de densidade. Provar que f_1 , f_2 e f_3 são as funções de densidade marginais de f_α segue similarmente.

A principal vantagem desta construção do Exemplo 6.27 é que, logicamente, podemos obter uma função de densidade conjunta a partir de densidades marginais conhecidas e, além disso, as densidades marginais podem ser de famílias de densidades diferentes. Por outro lado, temos por desvantagens que a forma funcional de f_α somente está definida para o caso de densidades marginais, ou seja, nada podemos fazer caso as variáveis aleatórias marginais

sejam discretas, também não temos controle acerca da estrutura de correlação e podemos ter expressões funcionais muito complexas, difíceis de serem implementadas e, ainda, não temos uma única função de densidade conjunta.

A resposta a este problema para o caso de marginais univariadas foi dada por Abe Sklar em 1959: a criação de uma nova classe de funções que ele chamou de Copulas. A Copula é uma função que relaciona uma função de distribuição multivariada com as funções de distribuição marginais, geralmente funções unidimensionais. Este conceito foi introduzido no artigo de Sklar (1959) mas, recentemente, tem se tornado uma ferramenta estatística muito importante por suas diversas aplicações.

Para iniciarmos ao conceito Copula primeiro vamos notar que, para as variáveis aleatórias U e V , com funções de distribuição contínuas F_U e F_V , respectivamente; as funções de variáveis aleatórias $F_U(U)$ e $F_V(V)$ são ambas uniformemente distribuídas no intervalo $[0, 1]$.

Daí para alguma função de distribuição conjunta $F_{U,V}$ com função de distribuição marginais F_U e F_V , temos que

$$\begin{aligned} F_{U,V}(u, v) &= P(U \leq u, V \leq v) \\ &= P(F_U(U) \leq F_U(u), F_V(V) \leq F_V(v)) \\ &= C_{U,V}(F_U(u), F_V(v)). \end{aligned}$$

Isto é a Copula C acima é exatamente a função de distribuição do vetor (U, V) com distribuições marginais uniforme. Definimos assim a função Copula como a função de distribuição multivariada de vetores aleatórios com marginais uniformes. Começaremos com a definição de Copula para o caso bivariado.

Definição 6.19 (*Modelo Copula*)

A função Copula bivariada ou modelo Copula bivariado é uma função $C_{U,V} : [0, 1] \times [0, 1] \rightarrow [0, 1]$ que satisfaz as seguintes condições:

- (i) Para todo $(u, v) \in [0, 1]$, $C_{U,V}(u, 0) = C(0, v) = 0$, $C_{U,V}(u, 1) = u$ e $C_{U,V}(1, v) = v$,
- (ii) Para todo $u_1, u_2, v_1, v_2 \in [0, 1]$ tais que $u_1 \leq u_2$ e $v_1 \leq v_2$, temos que

$$C_{U,V}(u_2, v_2) - C_{U,V}(u_2, v_1) - C_{U,V}(u_1, v_2) + C_{U,V}(u_1, v_1) \geq 0.$$

A primeira das condições nesta definição é chamada de restrições de contorno e a segunda de propriedade não decrescente. Observemos que toda função de distribuição conjunta é uma Copula. O que interessa a mais nesta situação é que as variáveis aleatórias U e V sejam dependentes, mesmo não conhecendo a expressão funcional da distribuição conjunta. Somente será importante o conhecimento das distribuições marginais, com isto queremos dizer que a diferença fundamental entre a função Copula e a distribuição conjunta é o fato de na Copula não importar a forma funcional conjunta sendo o mais importante o conhecimento das funções marginais.

Vejamos agora alguns exemplos de funções Copula bivariadas.

Exemplo 6.28

Caso as variáveis aleatórias U e V sejam independentes, a função

$$C_{U,V}(u, v) = uv,$$

para cada $(u, v) \in [0, 1] \times [0, 1]$ é conhecida como Copula independente.

Mais interessante ainda são funções que dependam de parâmetros, apresentamos a seguir exemplos de funções Copula que dependem de parâmetros.

Exemplo 6.29

Família de Copula descrita em Farlie (1960). Seja $\theta \in [-1, +1]$, então a função

$$C_{U,V}(u, v) = uv + \theta uv(1 - u)(1 - v),$$

é uma família uniparamétrica Copula. Em trabalhos posteriores podem ser encontradas generalizações desta função (Quesada-Molina & Rodriguez-Lallena, 1995).

Mostramos nos seguintes exemplos outras funções Copula paramétricas.

Exemplo 6.30

Família Copula Clayton. Se $\theta > 0$ então a família de funções Copula Clayton é definida como

$$C_{U,V}(u, v) = (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}.$$

Exemplo 6.31

Família Copula Ali-Mikhail-Haq. Se $-1 < \theta < 1$ então esta família de funções Copula é definida como

$$C_{U,V}(u, v) = uv(1 - \theta(1 - u)(1 - v))^{-1}.$$

Mencionaremos aqui algumas particularidades básicas destas funções.

- (i) A função Copula C é dita simétrica se $C_{U,V}(u, v) = C_{U,V}(v, u)$ para todo $(u, v) \in [0, 1] \times [0, 1]$. Observemos que todos os exemplos mencionados até agora são de funções Copula simétricas. Claro que nem todas as funções Copula são simétricas, por exemplo,

$$C_{U,V}(u, v) = \min \{u, \max\{0, v - \theta(1 - u)\}\},$$

para $\theta \in (0, 1)$ é Copula porém não é simétrica.

- (ii) A Copula transposta de uma Copula bivariada, denotada por C^T , é dada por

$$C_{U,V}^T(u, v) = C_{U,V}(v, u),$$

para todo $(u, v) \in [0, 1] \times [0, 1]$.

Teorema de Sklar

Será de grande utilidade o seguinte conceito. Para uma variável aleatória contínua U , com função de distribuição $F_U(u) = P(U \leq u)$, denotamos a imagem inversa da função de distribuição como

$$F_U^{-1}(u) = \inf\{z : F_U(z) \geq u\}.$$

A história da Copula pode-se dizer começou com Fréchet no seu artigo de 1951, mas foi em 1959 que Sklar obteve o resultado mais importante a este respeito, introduzindo ainda esta noção e o nome. Demonstrou também o teorema que hoje leva seu nome, apresentado a seguir.

Teorema 6.22 (Teorema de Sklar)

Sejam U e V duas variáveis aleatórias com função de distribuição conjunta $F_{U,V}$ e distribuições marginais F_U e F_V , respectivamente. Então existe uma única Copula C de forma que

$$C_{U,V}(u, v) = F_{U,V}(F_U^{-1}(u), F_V^{-1}(v)),$$

para todo $(u, v) \in [0, 1] \times [0, 1]$ e que conecta $F_{U,V}$ com F_U e F_V via

$$F_{U,V}(u, v) = C_{U,V}(F_U^{-1}(u), F_V^{-1}(v)),$$

para todo $(u, v) \in [-\infty, +\infty] \times [-\infty, +\infty]$.

Demonstração: content...

■

Este importante teorema garante a existência e unicidade da Copula caso as variáveis sejam contínuas. Caso contrário C é univocamente determinado em $\text{Suporte}(F_U) \times \text{Suporte}(F_V)$. Por outro lado, se C é uma Copula e F_U e F_V são funções de distribuição, então $F_{U,V}(u, v) = C_{U,V}(F_U^{-1}(u), F_V^{-1}(v))$ é uma função de distribuição conjunta com marginais F_U e F_V .

Exemplo 6.32

Sejam U e V duas variáveis aleatórias com função de distribuição conjunta

$$F_{U,V}(u, v) = \frac{1}{1 + \exp(-u) + \exp(-v)},$$

para todo $(u, v) \in (-\infty, +\infty)^2$, isto é, $F_{U,V}$ é a função de distribuição logística Gumbel bivariada. As distribuições marginais são, respectivamente,

$$F_U(u) = \frac{1}{1 + \exp(-u)} \quad e \quad F_V(v) = \frac{1}{1 + \exp(-v)}.$$

As imagens inversas destas funções são, respectivamente

$$F_U^{-1}(u) = -\ln\left(\frac{1-u}{u}\right) \quad e \quad F_V^{-1}(v) = -\ln\left(\frac{1-v}{v}\right).$$

Então, de acordo com o Teorema de Sklar

$$C_{U,V}(u, v) = F_{U,V}(F_U^{-1}(u), F_V^{-1}(v)) = \frac{uv}{u + v - uv}, \quad (u, v) \in (0, 1)^2.$$

O Teorema de Sklar deve ser usado com algum cuidado quando as distribuições marginais têm saltos. De fato, mesmo se não existir uma representação para a cópula não contínuos de conjuntos de, já não é único. Em tais casos, a modelagem e interpretação de dependência através de cópulas precisa de algum cuidado. Os leitores interessados devem se referir ao papel (Marshall, 1996) e para a discussão em profundidade por Genest e Nelehová (2007).

A prova do teorema de Sklar não foi dada em Sklar (1959), somente foi apresentado um esboço do que foi fornecido em Sklar (1973), de modo que por alguns anos profissionais da área tiveram que reconstruí-lo contando com as notas escritas à mão pelo próprio Sklar. Por cerca de 15 anos, todos os resultados em matéria de Copulas foram obtidos no âmbito da teoria dos espaços probabilísticos (Schweizer e Sklar, 1974). Somente em meados dos anos setenta a comunidade estatística interessou-se em Copulas.

Exemplo 6.33

Sejam U e V duas variáveis aleatórias com função de distribuição conjunta

$$F_{U,V}(u, v) = \begin{cases} \frac{(u+1)(\exp(v)-1)}{u+2\exp(v)-1}, & \text{se } (u, v) \in (-1, 1) \times (0, +\infty) \\ 1 - \exp(-v), & \text{se } (u, v) \in (1, +\infty) \times (0, +\infty) \\ 0, & \text{caso contrário} \end{cases},$$

então as distribuições marginais são

$$F_U(u) = \frac{u+1}{2}, \quad u \in (-1, 1) \quad \text{e} \quad F_V(v) = 1 - \exp(-v), \quad v \geq 0.$$

uniformemente distribuídas em $(-1, 1)$ e a distribuição exponencial, respectivamente. As imagens inversas destas funções são,

$$F_U^{-1}(u) = 2u - 1 \quad \text{e} \quad F_V^{-1}(v) = -\ln(1 - v).$$

Então, de acordo com o Teorema de Sklar

$$C_{U,V}(u, v) = F_{U,V}(F_U^{-1}(u), F_V^{-1}(v)) = \frac{uv}{u + v - uv}, \quad (u, v) \in (0, 1)^2,$$

isto é, a mesma do Exemplo 6.32.

Os resultados acima implicam que os modelos Copula podem ser utilizados para expressar uma distribuição multivariada em termos das suas distribuições marginais. Muitas vezes, sabe-se muito sobre as distribuições marginais das variáveis individuais, mas pouco sobre o seu comportamento comum. Copulas permitem aos investigadores reconstituir distribuições conjuntas, quando apenas as distribuições marginais são conhecidas com certeza.

6.6.1 Propriedades

O resultado básico para começar a análise remonta para a década de 1940 ao depararmos com os limites de Fréchet-Hoeffding para Copulas.

Teorema 6.23 (*Limites de Fréchet-Hoeffding*)

Para qualquer Copula $C_{U,V}$ e quaisquer $(u, v) \in [0, 1]^2$, se satisfaz que

$$\max(u + v - 1, 0) \leq C_{U,V}(u, v) \leq \min(u, v).$$

Demonstração: content... ■

O lado direito é sempre uma função de distribuição o assim chamado modelo co-monotônico, enquanto que o lado esquerdo é apenas uma função de distribuição no caso considerado, ou seja, no caso bivariado em que este modelo é chamado de contra-monotônico. Como uma conclusão importante que podemos aprender é que as funções de distribuição marginais restringem as possíveis funções de distribuição conjunta construídas em cima delas, segundo o Teorema de Sklar, tal como consta no Teorema 1.2.

Grande parte da utilidade da Copula em modelos estatísticos deriva dos fatos expressos no seguinte resultado.

Teorema 6.24

Sejam U e V duas variáveis aleatórias com Copula $C_{U,V}(u, v)$ e sejam g e h funções monótonas definidas no suporte destas variáveis, respectivamente. Então

(i) Se g e h forem estritamente crescentes, então

$$C_{g(U),h(V)}(u, v) = C_{U,V}(u, v).$$

(ii) Se g é estritamente crescente e h estritamente decrescente, então

$$C_{g(U),h(V)}(u, v) = u - C_{U,V}(u, 1 - v).$$

(iii) Se g é estritamente decrescente e h estritamente crescente, então

$$C_{g(U),h(V)}(u, v) = v - C_{U,V}(1 - u, v).$$

(iv) Se g e h forem estritamente decrescentes, então

$$C_{g(U),h(V)}(u, v) = u + v - 1 + C_{U,V}(1 - u, 1 - v).$$

Demonstração: content...



Densidade Copula

De acordo com sua definição, uma Copula é uma função de distribuição acumulada. É bastante típico que estas funções sejam monótonas crescentes e que, apesar de serem teoricamente muito poderosas, seus gráficos são difíceis de interpretar. Devido a isso, as funções de densidade Copula são utilizadas para ilustrar a distribuição, em vez da própria Copula. Claro, não em todos os casos Copulas têm densidades e vamos ver exemplos muito em breve. No entanto, se a Copula é suficientemente diferenciável a densidade Copula pode ser calculada.

Definição 6.20

Seja $Y = (Y_1, \dots, Y_K)^\top$ um vetor de variáveis aleatórias dependentes e $y = (y_1, \dots, y_K)^\top$ uma realização deste e C a função Copula correspondente. Então a densidade Copula, denotada por c caso exista, calcula-se como

$$c(y; \theta) = \frac{\partial^n C(y_1, \dots, y_K; \theta)}{\partial y_1 \cdots \partial y_K}. \quad (6.34)$$

Como mencionado, em muitas situações, pode ser encontrada a densidade Copula por simples diferenciação. Vejamos o seguinte exemplo.

Exemplo 6.34

ofhçgh ... hh . ohoho

Distribuição condicional

Como apontado no exemplo introdutório, a dependência é um conceito importante para inferir resultados de uma variável aleatória com base no conhecimento de um fator relacionado. Para uma ilustração, considere duas variáveis aleatórias uniformes U e V com Copula conhecida C e U é observada. O objetivo é deduzir a distribuição condicional de $V|U$ que pode então ser utilizada para estimar ou prever V .

Medidas de associação

6.6.2 Famílias Copulas

Funções implícitas

Funções explícitas

6.7 Exercícios

Exercícios da Seção 6.1

- 1.

Exercícios da Seção 6.2

1. Seja X uma variável aleatória com densidade

$$f(x; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \frac{(x - \mu)^2}{\sigma^2} \right\},$$

para $-\infty < x, \mu < +\infty$ e $\sigma > 0$.

Encontre $E(X|a < X < b)$, onde a e b são constantes.

2. Encontre $E[Y - E(Y|X)]^2$.
3. Seja (X, Y) um vetor aleatório com função de densidade

$$f(x, y) = \begin{cases} \frac{y}{(1+x)^4} e^{-\frac{y}{1+x}}, & x, y \geq 0 \\ 0, & \text{caso contrário} \end{cases}$$

Encontre $E(Y|X)$.

4. Seja (X, Y) um vetor aleatório com função de densidade

$$f(x, y) = \begin{cases} \frac{4}{5}(x + 3y)e^{-(x+2y)}, & x, y \geq 0 \\ 0, & \text{caso contrário} \end{cases}$$

Encontre $E(Y|X)$.

5. Suponha que r seja uma função real. Prove que

$$E[r(X) E(Y|X)] = E[r(X)Y].$$

6. Prove que $E[E(Y|X)] = E(Y)$.
7. Seja s uma função real. Prove que $E[s(X)Y|X] = s(X) E(Y|X)$.
8. Suponhamos que X e Y sejam duas variáveis aleatórias independentes. Prove que

$$E(Y|X) = E(Y).$$

9. Provar que $E[E(Z|X, Y)|X] = E[E(Z|X)|X, Y] = E(Z|X)$.
10. Seja X uma variável aleatória com densidade normal de parâmetros μ e σ^2 . Encontre $E(X|a < X < b)$, onde a e b são constantes.
11. Encontrar a expressão de $E\{Y - E(Y|X)\}^2$.
12. Sejam X e Y variáveis aleatórias e $\varphi(X)$ uma outra variável aleatória. Assuma que $E(Y)$ e $E\{\varphi(Y)\}$ existem. Mostre que

$$(i) \quad E\{\varphi(X)|X\} = \varphi(X),$$

$$(ii) \quad E\{\varphi(X)Y|X\} = \varphi(X) E(Y|X).$$

13. Sejam X e Y variáveis aleatórias tais que $E(X^2) < \infty$ e $E(Y^2) < \infty$. Demonstre que

$$\text{Cov}(X, Y) = \text{Cov}\{X, E(Y|X)\}.$$

14. Para todas as variáveis aleatórias X , Y e Z , seja $\text{Cov}(X, Y|Z = z)$ a covariância de X e Y na sua distribuição conjunta condicional dada $Z = z$. Prove que

$$\text{Cov}(X, Y) = E(\text{Cov}(X, Y|Z)) + \text{Cov}(E(X|Z), E(Y|Z)).$$

Exercícios da Seção 6.3

1. Seja $(X_{(1)}, X_{(2)}, \dots, X_{(n)})$ o conjunto das estatísticas de ordem de n variáveis aleatórias independentes X_1, X_2, \dots, X_n com função de densidade comum

$$f(x) = \begin{cases} \beta e^{-x\beta}, & \text{se } x \geq 0 \\ 0, & \text{caso contrário} \end{cases}.$$

- a) Mostre que $X_{(s)}$ e $X_{(r)} - X_{(s)}$ são independentes para quaisquer $r > s$.
 b) Encontre a função de densidade de $X_{(r+1)} - X_{(r)}$.
 c) Seja $Z_1 = nX_{(1)}$, $Z_2 = (n-1)(X_{(2)} - X_{(1)})$, $Z_3 = (n-2)(X_{(3)} - X_{(2)})$, ..., $Z_n = ((X_{(n)} - X_{(n-1)}))$. Prove que (Z_1, Z_2, \dots, Z_n) e (X_1, X_2, \dots, X_n) são identicamente distribuídas.
2. Sejam X_1, X_2, \dots, X_n variáveis aleatórias independentes e igualmente distribuídas com função de densidade comum

$$f(x) = \begin{cases} \frac{1}{\sigma} e^{-[(x-\theta)/\sigma]}, & \text{se } x > \theta \\ 0, & \text{se } x \leq \theta \end{cases}$$

Mostre que $X_{(1)}, X_{(2)} - X_{(1)}, X_{(3)} - X_{(2)}, \dots, X_{(n)} - X_{(n-1)}$ são independentes.

3. Sejam X_1, X_2, \dots, X_n variáveis aleatórias independentes e igualmente distribuídas com função de distribuição acumulada comum

$$F_X(t) = \begin{cases} 0, & \text{se } t \leq 0 \\ t^\alpha, & \text{se } 0 < t < 1 \\ 1, & \text{se } t \geq 1 \end{cases}$$

para $\alpha > 0$. Mostre que $X_{(i)}/X_{(n)}$, $i = 1, 2, \dots, n-1$ e $X_{(n)}$ são independentes.

4. Sejam X_1 e X_2 duas variáveis aleatórias discretas independentes com função de probabilidade comum

$$P(X = x) = \theta(1 - \theta)^{x-1}, \quad x = 1, 2, \dots; \quad 0 < \theta < 1.$$

Mostre que $X_{(1)}$ e $X_{(2)} - X_{(1)}$ são independentes.

5. Sejam X_1, \dots, X_n duas variáveis aleatórias independentes com função de densidade comum f . Encontre a função de densidade de $X_{(1)}$ e de $X_{(n)}$.
 6. Sejam $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ as estatísticas de ordem de n variáveis aleatórias independentes e igualmente distribuídas X_1, X_2, \dots, X_n com função de densidade comum

$$f(x) = \begin{cases} 1 & \text{se } 0 < x < 1 \\ 0, & \text{caso contrário} \end{cases}.$$

Prove que $Y_1 = X_{(1)}/X_{(2)}$, $Y_2 = X_{(2)}/X_{(3)}$, ..., $Y_{n-1} = X_{(n-1)}/X_{(n)}$ e $Y_n = X_{(n)}$ são independentes. Encontre a função de densidade conjunta de Y_1, Y_2, \dots, Y_n .

7. Sejam X_1, X_2, \dots, X_n variáveis aleatórias independentes identicamente distribuídas não negativas contínuas. Prove que se $E|X| < \infty$, então $E|X_{(r)}| < \infty$. Definamos $M_n = X_{(n)} = \max(X_1, X_2, \dots, X_n)$. Mostre que

$$E(M_n) = E(M_{n-1}) + \int_0^\infty F^{n-1}(x)[1 - F(x)] dx, \quad n = 2, 3, \dots$$

Encontre $E(M_n)$ em cada uma das seguintes situações:

- a) X_k tem como função de distribuição comum $F(x) = 1 - e^{-x\beta}$, se $x \geq 0$.
 b) X_k tem como função de distribuição comum $F(x) = x$, se $0 < x < 1$.

8. Seja (Ω, \mathcal{F}, P) um espaço de probabilidade e sejam $A, B \in \mathcal{F}$ eventos tais que $P(A) > 0$ e $P(B) > 0$. Definamos as variáveis aleatórias X e Y como

$$X(\omega) = 1_A(\omega) \quad \text{e} \quad Y(\omega) = 1_B(\omega), \quad \forall \omega \in \Omega.$$

Prove que X e Y são independentes se, e somente se, os eventos A e B forem independentes.

9. Sejam X_1, X_2 variáveis aleatórias independentes com função de probabilidade comum

$$P(X = \pm 1) = \frac{1}{2}.$$

Definimos $X_3 = X_1 X_2$. Prove que X_1, X_2, X_3 são independentes em pares mas não independentes.

10. Sejam X e Y variáveis aleatórias independentes tais que XY é uma variável aleatória degenerada. Mostre que X e Y são também degeneradas.
11. Sejam X e Y variáveis aleatórias independentes. Mostre que se X e $X - Y$ são independentes, então X deve ser uma variável degenerada.
12. Consideremos o vetor aleatório (X_1, X_2, X_3) com função de densidade conjunta

$$f(x_1, x_2, x_3) = \begin{cases} \frac{1}{4} & \text{se } (x_1, x_2, x_3) \in A \\ 0 & \text{caso contrário,} \end{cases}$$

onde $A = \{(1, 0, 0), (0, 1, 0), (0, 0, 1), (1, 1, 1)\}$. São X_1, X_2 e X_3 independentes? São X_1, X_2 e X_3 independentes aos pares? É $X_1 + X_2$ independente de X_3 ?

13. Sejam X e Y duas variáveis aleatórias independentes. Prove que a covariância entre elas é zero, isto é, prove que $\text{Cov}(X, Y) = 0$. O inverso não é verdadeiro: duas variáveis aleatórias que têm covariância zero não são necessariamente independentes.
14. Sejam X e Y duas variáveis aleatórias independentes. Prove que se $Y = g(X)$ então $\text{Corr}[g(X), Y] = 1$ mesmo que $\text{Corr}(X, Y) = 0$.
15. Suponha que X e Y sejam variáveis aleatórias com função de densidade conjunta

$$f(x, y) = \frac{1}{\pi} \quad \text{se } x^2 + y^2 \leq 1.$$

Então o vetor (X, Y) tem distribuição Uniforme no conjunto $\{(x, y) : x^2 + y^2 \leq 1\}$.

- a) As variáveis X e Y são independentes?
- b) Prove que $\text{Corr}(X, Y) = 0$.
- c) Prove que $\text{Corr}(X^2, -Y^2) = 1/3$.
16. Sejam $\{X_n\}$ variáveis aleatórias definidas no espaço de probabilidade $(\Omega, \mathcal{B}(\mathbb{R}), P)$, sendo $\Omega = (0, 1)$. Prove que se $X_n = \sin(2\pi n\omega)$, $n = 1, 2, \dots$; então estas variáveis são não correlacionadas e não independentes.
17. Suponha que U e R sejam duas variáveis aleatórias contínuas independentes onde U tem distribuição Uniforme no intervalo $[0, 1]$ e R tem como distribuição

$$f_R(r) = e \exp(-r^2/2), \quad \text{para } r \geq 0.$$

- a) Prove que R^2 tem a Exponencial como distribuição de probabilidades.
- b) Definamos $X = R \cos(2\pi U)$ e $Y = R \sin(2\pi U)$. Prove que X e Y são variáveis aleatórias independentes com distribuição Normal padrão.

Exercícios da Seção 6.4

1. Demonstrar a relação em (6.29), ou seja, provar que $\text{Var}(X|Y) = E(X^2|Y) - (E(X|Y))^2$.
- 2.

Capítulo 7

Convergência estocástica

Neste capítulo vamos investigar modos de convergência de vetores de variáveis aleatórias. Assim como na análise matemática é possível distinguir vários tipos de convergência. Diversos modos de convergência são introduzidos na Seção 7.3.

Neste capítulo, apresentamos uma série de resultados de aproximação que simplificam a análise de grandes amostras aleatórias. Na primeira seção, damos dois exemplos para ilustrar os tipos de análises que podemos desejar realizar e como ferramentas adicionais podem ser necessárias para poder realizá-las.

Antes de tudo isso, estudaremos a chamada função característica na Seção 7.1. Esta função será extremamente importante na demonstração de diversos resultados. Introduzimos na sub-seção 7.1.3 uma classe de variáveis aleatórias, chamadas infinitamente divisíveis, a qual constitui a classe das distribuições limite.

7.1 Função característica

Para qualquer tipo de variável aleatória o conceito de função geradora de momentos não define univocamente a função de probabilidade ou de densidade. Embora mais complexo, o conceito de função característica permite-nos esta identificação. Ainda este conceito será útil para provar resultados de convergência em distribuição de variáveis e vetores aleatórios.

Mesmo que as funções características assumem valores nos números complexos, não é preciso ter muita familiaridade com estes números para poder trabalhar com elas. Isso ficará claro no decorrer desta discussão.

Nesta seção, o símbolo i representará o número imaginário $\sqrt{-1}$. Se X e Y forem duas variáveis aleatórias em (Ω, \mathcal{F}, P) , então $Z = X + iY$ é uma variável aleatória complexa. Esta nova variável é definida em Ω porém assume valores complexos, com $Z(\omega) = X(\omega) + iY(\omega)$ para $\omega \in \Omega$, ou seja, $Z : \Omega \rightarrow \mathbb{C}$ mas X e Y são variáveis reais. Logicamente, se a nova variável assume valores complexos, sua esperança também. A esperança de Z é definida por linearidade,

$$E(Z) = E(X) + iE(Y), \quad (7.1)$$

no caso de $E(X)$ e $E(Y)$ finitas. Justifica-se essa forma e cálculo lembrando que $i = \sqrt{-1}$.

Definição 7.1

Seja X uma variável aleatória. A função característica de X é a função complexa $\varphi_X : \mathbb{R} \rightarrow \mathbb{C}$, definida como

$$\varphi_X(t) = E(e^{itX}), \quad (7.2)$$

para todo $t \in \mathbb{R}$.

Devemos observar que

$$E(e^{itX}) = E[\cos(tX)] + iE[\sin(tX)], \quad \forall t \in \mathbb{R},$$

pela fórmula de Euler¹. Observamos também que a variável aleatória complexa

$$e^{itX} = \cos(tX) + i\sin(tX)$$

sempre possui esperança finita, para toda variável aleatória X , pois as variáveis aleatórias $\cos(X)$ e $\sin(X)$ são limitadas. Assim, a esperança na Definição 7.1 é finita e devido a isso garantimos que a função característica sempre está bem definida.

Ainda devemos perceber que segundo a Definição 4.1, de esperança de uma variável aleatória, temos que

$$\varphi_X(t) = \int_{-\infty}^{+\infty} \cos(tx)f(x) dx + i \int_{-\infty}^{+\infty} \sin(tx)f(x) dx,$$

e, pela linearidade da integral, obtemos que

$$\varphi_X(t) = \int_{-\infty}^{+\infty} e^{itx} f(x) dx, \quad t \in \mathbb{R}.$$

Claro que estas expressões forem escritas pensando em variáveis aleatórias contínuas, caso seja discreta a variável X então substituímos a integral pela somatória e a função de densidade pela função de probabilidade. Esta última integral é chamada de transformada de Fourier de F e fornece uma definição alternativa de função característica.

Em situações nas quais a esperança matemática não existe, temos função característica. Um exemplo é a situação da distribuição Cauchy. Este caso é tratado no exemplo a seguir, o leitor poderá perceber o complicado do cálculo.

Exemplo 7.1

Suponhamos $X \sim \text{Cauchy}(\mu, \theta)$. Encontremos a expressão da função característica. Com esse objetivo provaremos que

$$\int_{-\infty}^{+\infty} \frac{e^{itx}}{\mu^2 + (x - \theta)^2} dx = \frac{\pi}{\mu} e^{i\theta t} e^{-\mu|t|}, \quad \forall t \in \mathbb{R}. \quad (7.3)$$

Deste resultado vamos obter que

$$\varphi_X(t) = e^{i\theta t} e^{-\mu|t|}, \quad \forall t \in \mathbb{R}.$$

¹A fórmula de Euler, que mostra uma relação entre as funções trigonométricas e a função exponencial, é dada por: $e^{iz} = \cos(z) + i\sin(z)$, para $z \in \mathbb{R}$

Considere a função $f(t) = e^{-|t|}$, então a transformada de Fourier de $f(t)$ é dada por

$$\begin{aligned} F(x) = \mathbf{F}[f(t)] &= \int_{-\infty}^{\infty} f(t)e^{-itx} dt = \int_{-\infty}^{\infty} e^{-|t|}e^{-itx} dt \\ &= \int_{-\infty}^0 e^t e^{-itx} dt + \int_0^{\infty} e^{-t} e^{-itx} dt \\ &= \lim_{u \rightarrow -\infty} \left. \frac{e^{(1-ix)t}}{1-ix} \right|_{t=u}^0 - \lim_{v \rightarrow \infty} \left. \frac{e^{-(1+ix)t}}{1+ix} \right|_{t=0}^v \\ &= \frac{1}{1-ix} + \frac{1}{1+ix} = \frac{2}{x^2+1}. \end{aligned}$$

Agora, a transformação inversa de $F(t)$ é

$$f(t) = \mathbf{F}^{-1}[F(t)] = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(t)e^{itx} dx = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{2}{x^2+1} e^{itx} dx,$$

o qual significa que

$$\pi e^{-|t|} = \int_{-\infty}^{\infty} \frac{e^{itx}}{x^2+1} dx. \quad (7.4)$$

Para encontrarmos a integral em (7.3) fazemos a transformação de variável $x = \nu + \theta$, do qual obtemos que

$$\int_{-\infty}^{+\infty} \frac{e^{itx}}{\mu^2 + (x - \theta)^2} dx = \int_{-\infty}^{+\infty} \frac{e^{it(\nu+\theta)}}{\mu^2 + \nu^2} d\nu = e^{it\theta} \int_{-\infty}^{+\infty} \frac{e^{it\nu}}{\mu^2 + \nu^2} d\nu.$$

Mais ainda, fazendo agora a substituição $\nu = \mu u$ na última integral, temos que

$$\int_{-\infty}^{+\infty} \frac{e^{itx}}{\mu^2 + (x - \theta)^2} dx = \frac{e^{it\theta}}{\mu} \int_{-\infty}^{+\infty} \frac{e^{it\mu u}}{u^2 + 1} du = \frac{\pi}{\mu} e^{it\theta} e^{-\mu|t|},$$

com o qual demonstramos a função característica de uma variável aleatória *Cauchy*(μ, θ) é da forma

$$\varphi_X(t) = e^{it\theta} e^{-\mu|t|}, \quad \forall t \in \mathbb{R}.$$

Notemos que a função característica é determinada pela função de distribuição e, mais adiante, veremos que a função de distribuição é determinada pela função característica. Vejamos agora um exemplo mais simples do cálculo desta função.

Exemplo 7.2

Suponhamos que $X \sim \text{Poisson}(\lambda)$. Encontremos a expressão da função característica. Pela definição, temos que,

$$\varphi_X(t) = \sum_{x=0}^{\infty} e^{itx} e^{-\lambda} \frac{\lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda e^{it})^x}{x!} = e^{-\lambda} e^{\lambda e^{it}} = e^{\lambda(e^{it}-1)}.$$

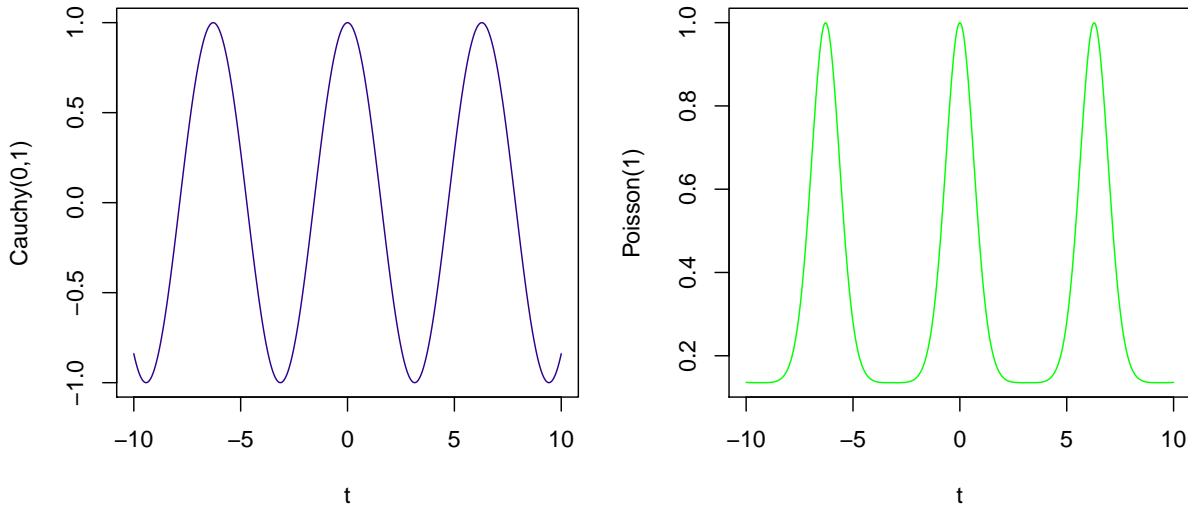


Figura 7.1: Gráfico das funções características: (a) densidade Cauchy(0,1) obtida no Exemplo 7.1 e (b) função de probabilidade Poisson(1) a direita, obtida no Exemplo 7.2

7.1.1 Propriedades da função característica

Os seguintes resultados, que serão provados utilizando resultados padrão de análise, nos permitirão entender a importância da função característica. Uma excelente referência neste assunto é o livro Gnedenko & Kolmogorov (1954).

Uma observação importante para demonstrar a primeira das propriedades da função característica é que o valor absoluto de um número complexo $z = x + iy$ é dado por $|z| = \sqrt{x^2 + y^2}$.

Teorema 7.1

A função característica é limitada por 1, isto é, $|\varphi_X(t)| \leq 1, \forall t \in \mathbb{R}$.

Demonstração: Pela definição de valor absoluto de um número complexo obtemos

$$|\varphi_X(t)| = |E(e^{itX})| = \sqrt{E^2[\cos(tX)] + E^2[\sin(tX)]}.$$

Agora, pela desigualdade de Jensen em (4.34), temos

$$\sqrt{E^2[\cos(tX)] + E^2[\sin(tX)]} \leq \sqrt{E[\cos^2(tX)] + E[\sin^2(tX)]} = 1.$$

Este último resultado é devido ao fato de que

$$E[\cos^2(tX)] + E[\sin^2(tX)] = E[\cos^2(tX) + \sin^2(tX)]$$

e que $\cos^2(tX) + \sin^2(tX) = 1$. ■

Lembremos que se $c = x + iy$ for um número complexo qualquer, o seu complexo conjugado é definido como $\bar{c} = x - iy$.

Teorema 7.2

A função característica assume o valor 1 no ponto zero, ou seja, $\varphi_X(0) = 1$ e $\overline{\varphi_X(t)} = \varphi_X(-t)$, $\forall t \in \mathbb{R}$, onde $\overline{\varphi_X(t)}$ é o complexo conjugado de $\varphi_X(t)$.

Demonstração :

$$\varphi_X(0) = E(e^{i0X}) = E(1) = 1,$$

e

$$\overline{\varphi_X(t)} = E[\cos(tX)] - i E[\sin(tX)] = E[\cos(-tX)] + i E[\sin(-tX)] = \varphi_X(-t),$$

lembrando que o cosseno é uma função par e o seno é uma função ímpar. ■

Teorema 7.3

A função característica sempre existe e é um núcleo definido ou kernel não negativo, ou seja, para qualquer n , quaisquer números reais t_1, t_2, \dots, t_n e quaisquer números complexos z_1, z_2, \dots, z_n , temos

$$\sum_{j=1}^n \sum_{k=1}^n z_j \varphi(t_j - t_k) \bar{z}_k \geq 0. \quad (7.5)$$

Demonstração : A função $e^a \geq 0$, então $\varphi_X(t) \geq 0$. Como foi provado no Teorema 7.1 $E(|e^{itX}|^2) = 1$, do qual temos que

$$E(|e^{itX}|) \leq \sqrt{E(|e^{itX}|^2)} = 1,$$

portanto a função e^{itX} sempre é de esperança finita. A demonstração da desigualdade em (7.5) será realizada para o caso absolutamente contínuo.

$$\begin{aligned} \sum_{j=1}^n \sum_{k=1}^n z_j \varphi(t_j - t_k) \bar{z}_k &= \sum_{j=1}^n \sum_{k=1}^n \int_{-\infty}^{\infty} z_j e^{i(t_j - t_k)x} \bar{z}_k f(x) dx \\ &= \sum_{j=1}^n \sum_{k=1}^n \int_{-\infty}^{\infty} z_j e^{it_j x} \overline{(z_k e^{it_k x})} f(x) dx \\ &= E\left(\sum_{j=1}^n \sum_{k=1}^n z_j e^{it_j X} \overline{z_k e^{it_k X}}\right) \geq E\left(\sum_{j=1}^n |z_j e^{it_j X}|^2\right) \geq 0. \end{aligned} \quad \blacksquare$$

O significado do Teorema 7.3 pode não ser aparente à primeira vista. No entanto, essas três propriedades nos Teoremas 7.1, 7.2 e 7.3 são consideradas as propriedades definidoras de uma função característica, porque essas propriedades também são suficientes para que uma

função arbitrária seja a função característica de alguma variável aleatória. Esse importante resultado é conhecido como Teorema de Bochner², que está além do nosso escopo de estudo.

Teorema 7.4

Se $Y = aX + b$, sendo $a, b \in \mathbb{R}$ quaisquer, então $\varphi_Y(t) = e^{itb}\varphi_X(at)$.

Demonstração :

$$\varphi_Y(t) = E[e^{it(aX+b)}] = e^{itb} E(e^{iatX}) = e^{itb}\varphi_X(at). \quad \blacksquare$$

Exemplo 7.3

Seja $Z \sim N(0, 1)$. Encontremos a expressão da função característica. Pela definição, temos que

$$\varphi_Z(t) = \int_{-\infty}^{+\infty} e^{itz} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz.$$

Completando quadrado na integral, temos que

$$\varphi_Z(t) = e^{-\frac{t^2}{2}} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(z-it)^2}{2}} dz.$$

Sabemos que $\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2}} dx = 1$, pela definição de distribuição Normal, para qualquer constante μ .

Então,

$$\varphi_Z(t) = e^{-\frac{t^2}{2}} \quad (7.6)$$

é a expressão da função característica de uma variável aleatória Normal padrão. Agora, utilizando o Teorema 7.4 obtemos que, se $Y = \sigma Z + \mu$ então $Y \sim N(\mu, \sigma^2)$ e

$$\varphi_Y(t) = e^{it\mu}\varphi_Z(\sigma t) = e^{it\mu - \frac{\sigma^2 t^2}{2}} \quad (7.7)$$

é a expressão da função característica de uma variável aleatória Normal de média μ e variância σ^2 .

Devemos fazer distinção, neste exemplo, entre a variável aleatória normal padrão Z e a variável aleatória complexa Z . Embora a função característica é uma função complexa, não estamos interessados em estudar variáveis aleatórias complexas. As variáveis aleatórias complexas foram introduzidas somente para entendermos o conceito de função característica.

Teorema 7.5

A função característica $\varphi_X(t)$ é uniformemente contínua na reta.

²Salomon Bochner (1899-1982). Foi um matemático austro-húngaro naturalizado estadunidense.

Para demonstrar esta propriedade devemos lembrar que uma função uniformemente contínua é contínua em todo ponto. A recíproca deste resultado não vale, por exemplo, a função $f(x) = x^2$ é contínua na reta mas não é uniformemente contínua. Por definição, uma função φ é uniformemente contínua se para $\epsilon > 0$, existe $\delta = \delta(\epsilon) > 0$ tal que $|\varphi(t) - \varphi(s)| < \epsilon$, quando $|t - s| < \delta$. Notemos que δ precisa depender apenas de ϵ .

Demonstração: Observemos que

$$\begin{aligned} |\varphi_X(t) - \varphi_X(s)| &= \left| \int_{-\infty}^{+\infty} (e^{itx} - e^{isx}) f(x) dx \right| \leq \int_{-\infty}^{+\infty} |e^{itx} - e^{isx}| f(x) dx \\ &= \int_{-\infty}^{+\infty} |e^{isx}| |e^{i(t-s)x} - 1| f(x) dx \leq \int_{-\infty}^{+\infty} |e^{i(t-s)x} - 1| f(x) dx. \end{aligned}$$

Definamos esta última integral como $h(t-s)$, isto é, definamos

$$h(t-s) = \int_{-\infty}^{+\infty} |e^{i(t-s)x} - 1| f(x) dx.$$

Observemos que $h(u) = E|e^{iuX} - 1|$, também $0 \leq |e^{iuX} - 1| \leq 2$, 2 é integrável e

$$\lim_{u \rightarrow 0} |e^{iuX(\omega)} - 1| = 0, \quad \forall \omega \in \Omega.$$

Então, pelo Teorema da Convergência Dominada³, $\lim_{u \rightarrow 0} h(u) = 0$ e, neste caso, para dado $\epsilon > 0$, existe $\delta > 0$ tal que $|t - s| < \delta$, o qual implica que

$$|\varphi_X(t) - \varphi_X(s)| \leq h(t-s) < \epsilon. \quad \blacksquare$$

Teorema 7.6

Sejam X e Y forem duas variáveis aleatórias independentes com funções características respectivas $\varphi_X(t)$ e $\varphi_Y(t)$, então $\varphi_{X+Y}(t) = \varphi_X(t) \times \varphi_Y(t)$, $\forall t \in \mathbb{R}$.

Demonstração:

$$\varphi_{X+Y}(t) = E[e^{it(X+Y)}] = E(e^{itX} e^{itY}) = E(e^{itX}) E(e^{itY}) = \varphi_X(t) \times \varphi_Y(t). \quad \blacksquare$$

Esta propriedade implica que o produto de duas funções características também é uma função característica. A demonstração deste teorema também pode ser realizada usando senos e cossenos. Por indução percebemos que

$$\varphi_{X_1 + \dots + X_n}(t) = \prod_{k=1}^n \varphi_{X_k}(t), \quad \forall t \in \mathbb{R}, \quad (7.8)$$

caso as variáveis aleatórias X_1, \dots, X_n sejam independentes, ou seja, a função característica da soma de n variáveis aleatórias independentes é o produto das funções características associadas à cada variável aleatória.

³Teorema da Convergência Dominada: Seja $\{f_n\}$ uma sequência de funções contínuas que convergem pontualmente à certa função f e suponhamos existe uma função integrável g , tal que $|f_n(x)| \leq g(x)$. Então temos $\int f = \lim_{n \rightarrow \infty} \int f_n$.

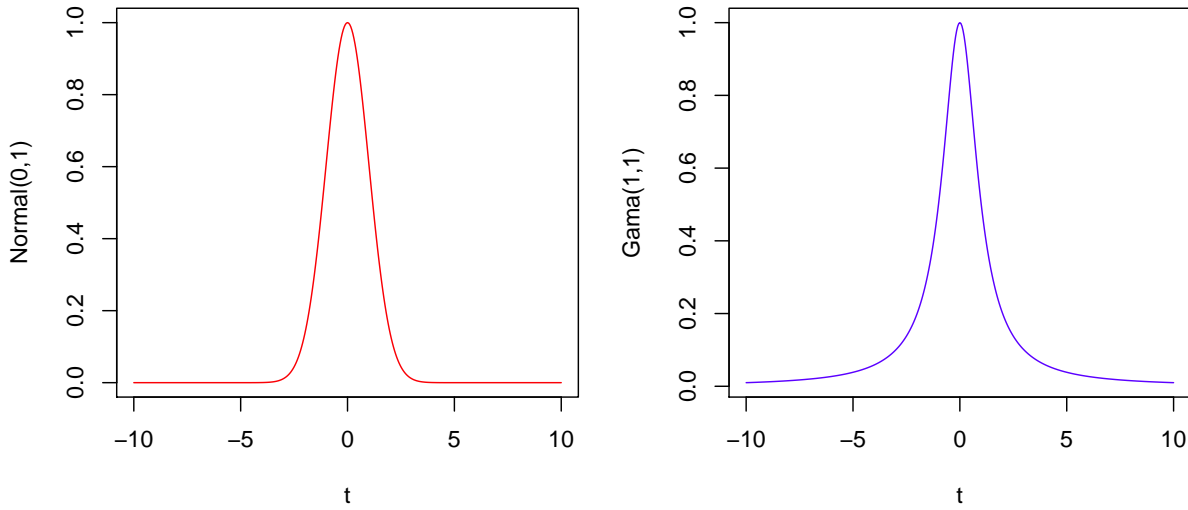


Figura 7.2: Gráfico das funções características: (a) densidade Normal a esquerda obtida no Exemplo 7.3, (b) função de densidade Gama a direita, obtida no Exemplo 7.4. Em ambos os casos utilizamos a situação canônica, ou seja, no caso $Normal(0, 1)$ a função característica é $\varphi_X(t) = e^{-t^2/2}$ e no caso $Gama(1, 1)$, a função característica é da forma $\varphi_X(t) = 1/(1 - it)$.

Exemplo 7.4

Caso $X \sim Gama(\alpha, \nu)$, com função de densidade

$$f(x; \alpha, \nu) = \frac{1}{\Gamma(\nu)} \alpha^\nu x^{\nu-1} e^{-\alpha x},$$

quando $x \geq 0$ e zero caso contrário a função característica é

$$\begin{aligned} \varphi_X(t) &= E(e^{-itX}) = \int_0^\infty \frac{e^{itx}}{\Gamma(\nu)} \alpha^\nu x^{\nu-1} e^{-\alpha x} dx \\ &= \frac{\alpha^\nu}{\Gamma(\nu)} \int_0^\infty x^{\nu-1} e^{-(\alpha - it)x} dx = \left(1 - \frac{it}{\alpha}\right)^{-\nu}. \end{aligned}$$

Corolário 7.7

Se $\varphi_X(t)$ é uma função característica, então $|\varphi_X(t)|^2$ também é função característica.

Demonstração: Sabemos que $\varphi_X(t)$ é a função característica de X . Então, existe uma variável aleatória Y , independente de X que têm mesma função de distribuição e, portanto, tendo a mesma função característica. Então, a função característica de $X - Y$ é

$$E(e^{it(X-Y)}) = E(e^{itX}) E(e^{-itY}) = \varphi_X(t) \varphi_X(-t) = |\varphi_X(t)|^2.$$

■

Corolário 7.8

Sejam F_1, \dots, F_n funções de distribuição com funções características correspondentes $\varphi_1, \dots, \varphi_n$ e $\lambda \geq 0$ tais que $\lambda_1 + \dots + \lambda_n = 1$. Então,

$$\sum_{k=1}^n \lambda_k F_k, \quad \text{tem por função característica} \quad \sum_{k=1}^n \lambda_k \varphi_k.$$

Demonstração: Exercício. ■

$\sum_{k=1}^n \lambda_k F_k$ é uma função de distribuição? a resposta é sim, esta combinação linear de funções de distribuição é uma função de distribuição satisfazendo as propriedades do Teorema 3.16.

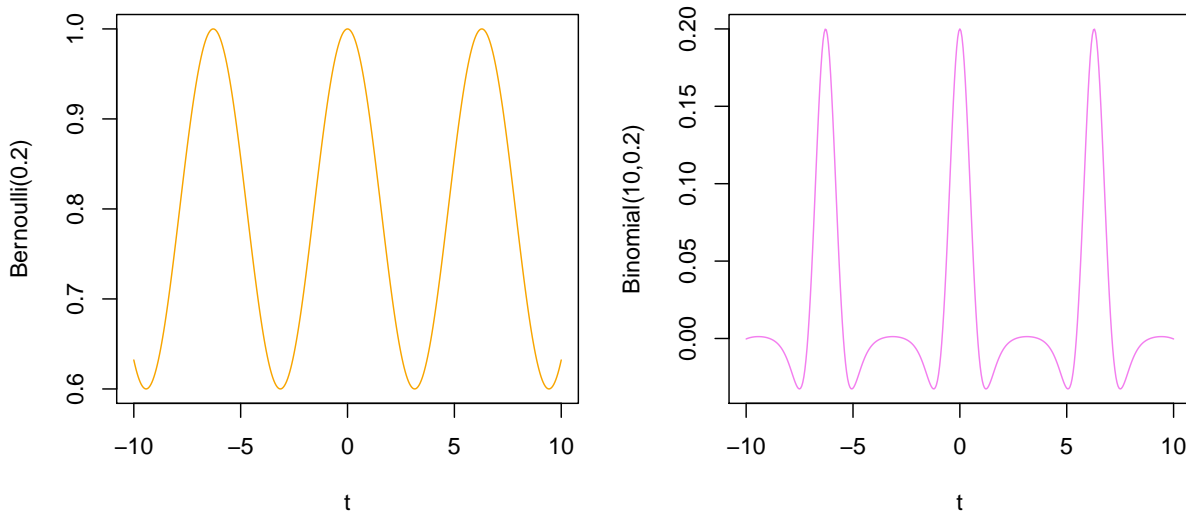


Figura 7.3: Gráfico das funções características: (a) função de probabilidade Bernoulli(0.2) a esquerda obtida no Exemplo 7.5 e (b) função de probabilidade Binomial(10,0.2) a direita e obtida no mesmo exemplo.

Exemplo 7.5

Sabemos que se X_1, X_2, \dots, X_n forem n variáveis aleatórias independentes cada uma Bernoulli(p), então $Y = X_1 + \dots + X_n$ tem por distribuição $Y \sim \text{Binomial}(n, p)$.

Queremos encontrar a função característica de uma variável com distribuição Binomial. Sabemos agora que podemos encontrar a função característica φ_Y como produto das funções características φ_{X_k} . Utilizando a definição, temos que $\varphi_{X_k}(t) = (1 - p) + pe^{it}$. Agora, pelo resultado em (7.8), obtemos que

$$\varphi_Y(t) = \prod_{k=1}^n \varphi_{X_k}(t) = \prod_{k=1}^n ((1 - p) + pe^{it}) = \sum_{k=0}^n e^{itk} \binom{n}{k} p^k (1 - p)^{n-k},$$

lembrando que $(pe^{it})^k = p^k e^{itk}$. Como deveríamos esperar, o resultado corresponde com a função característica de uma variável aleatória *Binomial*(n, p).

Demonstremos uma das propriedades mais importantes da função característica. Comentamos, logo depois de definir a função característica, que devido a $\varphi_X(t) = \int e^{itx} f(x) dx$ e $\varphi_X(t) = \sum e^{itx} P(X = x)$, a função característica é determinada pela função de distribuição. A propriedade no Teorema 7.10 é uma recíproca disto, ou seja, essa próxima propriedade nos disse que a função característica determina sua função de distribuição univocamente.

O resultado a seguir implica que duas funções de distribuição são iguais se, e somente se, suas funções características o forem e como consequência a função característica é uma representação da função de distribuição. Esta é a principal razão do nosso interesse em funções características, elas descrevem exclusivamente a função de distribuição.

Antes de disso, precisamos de um resultado auxiliar, a chamada fórmula de inversão a qual permite calcularmos probabilidades de intervalos a partir da função característica. Significa que a fórmula de inversão permite recuperar a densidade ou a função de probabilidade de uma variável aleatória X de sua função característica e usa essa fórmula para estabelecer o fato de que, em geral, a função característica de X caracteriza exclusivamente a distribuição de X .

Teorema 7.9 (*Fórmula de Inversão*)

Se a variável aleatória X tem como função característica φ_X , então para qualquer intervalo (a, b) temos

$$P(a < X < b) + \frac{P(X = a) + P(X = b)}{2} = \lim_{T \rightarrow \infty} \frac{1}{2\pi} \int_{-T}^T \frac{e^{-ita} - e^{-itb}}{it} \varphi_X(t) dt.$$

Demonstração: Considere a integral

$$\begin{aligned} \frac{1}{2\pi} \int_{-T}^T \frac{e^{-ita} - e^{-itb}}{it} \varphi_X(t) dt &= \\ &= \frac{1}{2\pi} \int_{-T}^T \frac{e^{-ita} - e^{-itb}}{it} \left[\int_{-\infty}^{+\infty} e^{itx} f(x) dx \right] dt \\ &= \int_{-T}^T \int_{-\infty}^{+\infty} \frac{e^{it(x-a)} - e^{it(x-b)}}{2\pi it} f(x) dx dt \\ &= \int_{-\infty}^{+\infty} \int_{-T}^T \frac{e^{it(x-a)} - e^{it(x-b)}}{2\pi it} dt f(x) dx. \end{aligned}$$

Para qualquer número real c temos que

$$\int_{-T}^T \frac{e^{itc}}{2it} dt = \int_{-T}^T \left[\frac{\cos(tc)}{2it} + \frac{i \sin(tc)}{2it} \right] dt.$$

Acontece que $\frac{\cos(tc)}{t}$ é uma função ímpar e $\frac{\sen(tc)}{t}$ é uma função par, logo

$$\int_{-T}^T \frac{\cos(tc)}{t} dt = 0 \quad \text{e} \quad \int_{-T}^T \frac{\sen(tc)}{t} dt = 2 \int_0^T \frac{\sen(tc)}{t} dt.$$

Então

$$\int_{-T}^T \frac{e^{itc}}{2it} dt = \int_0^T \frac{\sen(tc)}{t} dt,$$

e deste resultado temos

$$\begin{aligned} \frac{1}{2\pi} \int_{-T}^T \frac{e^{-ita} - e^{-itb}}{it} \varphi_X(t) dt &= \\ &= \int_{-\infty}^{+\infty} \frac{1}{\pi} \left\{ \int_0^T \frac{\sen[t(x-a)]}{t} dt - \int_0^T \frac{\sen[t(x-b)]}{t} dt \right\} f(x) dx. \end{aligned}$$

A integral

$$\lim_{T \rightarrow \infty} \int_0^T \frac{\sen(t)}{t} dt = \frac{\pi}{2},$$

é conhecida como integral clássica de Dirichlet e, a partir dela, podemos calcular

$$\lim_{T \rightarrow \infty} \frac{1}{\pi} \int_0^T \frac{\sen(tc)}{t} dt = \begin{cases} \lim_{T \rightarrow \infty} \frac{1}{\pi} \int_0^T \frac{\sen(t)}{t} dt = \frac{1}{2}, & \text{se } c > 0 \\ \lim_{T \rightarrow \infty} -\frac{1}{\pi} \int_0^T \frac{\sen(-tc)}{t} dt = -\frac{1}{2}, & \text{se } c < 0 \end{cases}$$

e zero caso $x = c$, satisfazendo-se $\forall c \in \mathbb{R}$.

Utilizando o Teorema da Convergência Dominada temos que

$$\begin{aligned} \lim_{T \rightarrow \infty} \int_{-\infty}^{+\infty} \frac{1}{\pi} \left\{ \int_0^T \frac{\sen[t(x-a)]}{t} dt - \int_0^T \frac{\sen[t(x-b)]}{t} dt \right\} f(x) dx &= \\ &= \int_{-\infty}^{+\infty} \lim_{T \rightarrow \infty} \frac{1}{\pi} \left\{ \int_0^T \frac{\sen[t(x-a)]}{t} dt - \int_0^T \frac{\sen[t(x-b)]}{t} dt \right\} f(x) dx. \end{aligned}$$

Aplicando o resultado da integral de Dirichlet, esta última integral é igual a

$$\int_{-\infty}^{+\infty} g(x) f(x) dx,$$

onde

$$g(x) = \begin{cases} \frac{1}{2}, & \text{se } x = a \\ 1, & \text{se } a < x < b \\ \frac{1}{2}, & \text{se } x = b \\ 0, & \text{caso contrário} \end{cases}$$

e esta integral é igual a

$$P(a < X < b) + \frac{P(X = a) + P(X = b)}{2}.$$

■

Devemos observar que, caso os pontos a e b forem de continuidade da função de distribuição F , a fórmula de inversão se reduz à expressão

$$F(b) - F(a) = \lim_{T \rightarrow \infty} \frac{1}{2\pi} \int_{-T}^T \frac{e^{-ita} - e^{-itb}}{it} \varphi_X(t) dt. \quad (7.9)$$

Isto significa que a função característica φ_X determina a função de distribuição F e, portanto, a função de distribuição pode ser calculada a partir da função característica.

O seguinte exemplo trata acerca da situação na qual a função característica φ_X corresponde a uma variável aleatória X discreta.

Exemplo 7.6 (Fórmula de Inversão para probabilidades pontuais)

Seja X uma variável aleatória discreta com função característica φ_X . Mostremos que para cada $x \in \mathbb{R}$,

$$P(X = x) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T e^{-itx} \varphi_X(t) dt.$$

Nesta situação $\varphi_X(t) = \sum_x P(X = x) e^{itx}$, então

$$\begin{aligned} \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T e^{-itx} \varphi_X(t) dt &= \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T e^{-itx} \left(\sum_y P(X = y) e^{ity} \right) dt \\ &= \lim_{T \rightarrow \infty} \frac{1}{2T} \sum_y P(X = y) \int_{-T}^T e^{-it(x-y)} dt \\ &= \lim_{T \rightarrow \infty} \frac{1}{2T} \begin{cases} P(X = x) \int_{-T}^T dt & \text{caso } x = y \\ \sum_y P(X = y) \int_{-T}^T e^{-i(x-y)t} dt & \text{caso } x \neq y \end{cases} \\ &= P(X = x). \end{aligned}$$

Este resultado é obtido já que $\lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T e^{-i(x-y)t} dt = \lim_{T \rightarrow \infty} \frac{2 \operatorname{sen}((x-y)T)}{2T} = 0$.

Acontece que não costuma ser prático obter a função de distribuição através da fórmula de inversão. A utilidade desta fórmula é na demonstração do seguinte teorema, conhecido como Teorema de Unicidade.

Teorema 7.10 (Teorema de Unicidade)

A função característica de uma variável aleatória X determina univocamente sua função de distribuição.

Demonstração: É uma consequência direta do Teorema 7.9, a Fórmula de Inversão, pois esta

implica que, $\forall x \in \mathbb{R}$

$$F(x) = \lim_{b \rightarrow x} \lim_{a \rightarrow -\infty} \lim_{T \rightarrow \infty} \frac{1}{2\pi} \int_{-T}^T \frac{e^{-ita} - e^{-itb}}{it} \varphi_x(t) dt, \quad (7.10)$$

sendo x um ponto de continuidade de F e o limite é único. ■

Este resultado permanece válido mesmo não sendo x um ponto de continuidade de F . Mais ainda, os Teoremas 7.9 e 7.10 são válidos caso F seja uma função contínua a direita, arbitrária, de variação limitada, restrita à condição $F(-\infty) = 0$. De fato, tal função pode ser representada na forma

$$F(x) = \alpha_1 F_1(x) + \alpha_2 F_2(x),$$

onde F_1 e F_2 são funções de distribuição. Para a função característica obtemos, de forma óbvia, a equação correspondente

$$\varphi_x(x) = \alpha_1 \varphi_{x_1}(x) + \alpha_2 \varphi_{x_2}(x).$$

Dos resultados em (7.9) e (7.10), aplicados a F_1 e F_2 separadamente, obtemos imediatamente (7.9) e (7.10) para F .

Teorema 7.11

Sejam X e Y dois vetores aleatórios tais que $\varphi_x(t) = \varphi_y(t)$, $\forall t \in \mathbb{R}^n$. Então, X e Y tem a mesma função de distribuição.

Demonstração: Consequência do Teorema 7.10. ■

Consideramos alguns exemplos da aplicação destes últimos resultados.

Exemplo 7.7

Se as variáveis independentes X_1 e X_2 são normalmente distribuídas, então sua soma $Y = X_1 + X_2$ é também normalmente distribuída. De fato, se $E(X_1) = \mu_1$, $\text{Var}(X_1) = \sigma_1^2$, $E(X_2) = \mu_2$ e $\text{Var}(X_2) = \sigma_2^2$, então as funções características das variáveis X_1 e X_2 são, respectivamente,

$$\varphi_{x_1}(t) = \exp \left\{ i\mu_1 t - \frac{1}{2}\sigma_1^2 t^2 \right\} \quad \text{e} \quad \varphi_{x_2}(t) = \exp \left\{ i\mu_2 t - \frac{1}{2}\sigma_2^2 t^2 \right\}.$$

Pelo Teorema 7.6, a função característica φ_Y da soma $Y = X_1 + X_2$ é

$$\varphi_Y(t) = \varphi_{x_1}(t) \times \varphi_{x_2}(t) = \exp \left\{ i(\mu_1 + \mu_2)t - \frac{1}{2}(\sigma_1^2 + \sigma_2^2)t^2 \right\}. \quad (7.11)$$

A expressão em (7.11) é a função característica da distribuição normal com esperança $\mu = \mu_1 + \mu_2$ e variância $\sigma = \sigma_1^2 + \sigma_2^2$. Com base no Teorema 7.11, o teorema de unicidade,

concluimos que a função de distribuição da variável aleatória Y é normal. Demonstramos no Exemplo 7.7 que soma de duas variáveis aleatórias independentes normais é também uma variável aleatória normal e encontramos os momentos correspondentes, este resultado pode-se estender para a soma finita de n variáveis aleatórias independentes normais. Assim, podemos afirmar que se $x_i \sim N(\mu_i, \sigma_i^2)$ forem normais independentes, então

$$Y = X_1 + \cdots + X_n \sim N(\mu_1 + \cdots, \mu_n, \sigma_1^2 + \cdots + \sigma_n^2). \quad (7.12)$$

É interessante mencionar que a proposição inversa também é verdadeira: se a soma de duas variáveis aleatórias independentes é normalmente distribuída, então cada somando é normalmente distribuído (Cramér, 1936).

Exemplo 7.8

Caso as variáveis $X_1 \sim \text{Poisson}(\lambda_1)$ e $X_2 \sim \text{Poisson}(\lambda_2)$ sejam independentes, no Exemplo 7.2 foi obtido que a função característica de cada uma é

$$\varphi_{X_1}(t) = \exp \{ \lambda_1 (e^{it} - 1) \} \quad \text{e} \quad \varphi_{X_2}(t) = \exp \{ \lambda_2 (e^{it} - 1) \}.$$

A função característica da soma $Y = X_1 + X_2$ é

$$\varphi_Y(t) = \varphi_{X_1}(t) \times \varphi_{X_2}(t) = \exp \{ (\lambda_1 + \lambda_2) (e^{it} - 1) \},$$

isto é, a função característica de certa variável aleatória com distribuição Poisson. De acordo com o teorema de unicidade, a variável Y tem distribuição Poisson de parâmetro $\lambda = \lambda_1 + \lambda_2$, ou seja,

$$P(Y = k) = \frac{(\lambda_1 + \lambda_2)^k e^{-(\lambda_1 + \lambda_2)}}{k!}, \quad (k > 0).$$

Vale também aqui afirmar que soma de n variáveis aleatórias independentes com distribuição cada uma $\text{Poisson}(\lambda_i)$ tem distribuição $\text{Poisson}(\lambda_1 + \cdots, \lambda_n)$. Raikov (1938) provou a proposição inversa: se uma soma de variáveis aleatórias independentes tem por distribuição a Poisson, então cada uma das variáveis da soma tem distribuição Poisson.

Teorema 7.12

Seja F uma função de distribuição e $\varphi(t) = \int_{-\infty}^{\infty} e^{itx} f(x) dx$ a função característica da correspondente função de distribuição. Consideremos $\varphi(t)$ absolutamente integrável na reta real, ou seja, $\int_{-\infty}^{\infty} |\varphi(t)| dt < \infty$. Então a função de distribuição F tem função de densidade contínua e limitada f e a seguinte fórmula de inversão ocorre

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{itx} \varphi(t) dt, \quad x \in \mathbb{R}. \quad (7.13)$$

Demonstração: Dado que $\varphi(t)$ é integrável e $|e^{itx}| = 1$, f em (7.13) está bem definida. Para $a < b$, temos

$$\begin{aligned} \int_a^b f(x) dx &= \frac{1}{2\pi} \int_a^b \left(\int_{-\infty}^{\infty} e^{itx} \varphi(t) dt \right) dx \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \varphi(t) \left(\int_a^b e^{itx} dx \right) dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-ita} - e^{-itb}}{it} \varphi(t) dt \\ &= \lim_{T \rightarrow \infty} \frac{1}{2\pi} \int_{-T}^T \frac{e^{-ita} - e^{-itb}}{it} \varphi(t) dt = F(b) - F(a). \end{aligned}$$

■

Um detalhe importantíssimo neste teorema é a exigência de $\int_{-\infty}^{\infty} |\varphi(t)| dt < \infty$. Vejamos no exemplo a seguir um caso interessante no qual esta restrição não se cumpre.

Exemplo 7.9

Seja $\varphi_x(t) = \frac{\lambda}{\lambda - it}$, $\lambda > 0$ a função característica de uma variável aleatória X . Queremos identificar a distribuição de X .

Acontece que nesta situação não podemos utilizar o resultado em (7.13) para encontrarmos a função de densidade de X . Resulta que a função φ_x não é integrável, isso porque $|\varphi_x(t)| \sim 1/|t|$, quando $t \rightarrow \infty$. Então, o Teorema 7.12 não é aplicável mas, se $f(x) = \lambda e^{-\lambda x}$, para $0 < x < \infty$ temos que

$$\varphi_x(t) = \int_{-\infty}^{\infty} e^{itx} \lambda e^{-\lambda x} dx = \lambda \int_0^{\infty} e^{-(\lambda - it)x} dx = \frac{\lambda}{\lambda - it},$$

ou seja, a função de distribuição φ_x corresponde à função de densidade Exponencial(λ).

Mostramos agora uma situação na qual sim podemos utilizar o Teorema 7.12.

Exemplo 7.10

Seja $\varphi_x(t) = \frac{\lambda^2}{\lambda^2 + t^2}$, $\lambda > 0$ a função característica de uma variável aleatória X . Queremos identificar a distribuição de X utilizando para isso a expressão em (7.13). Primeiro percebemos que não é difícil provar que

$$\int_{-\infty}^{\infty} \frac{\lambda^2}{\lambda^2 + t^2} dt = \lambda\pi, \quad (7.14)$$

logo, o Teorema 7.12 é aplicável. Então,

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} e^{itx} \frac{\lambda^2}{\lambda^2 + t^2} dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{itx}}{1 + (t/\lambda)^2} dt = \frac{\lambda}{2\pi} \int_{-\infty}^{\infty} \frac{e^{i(\lambda x)v}}{1 + v^2} dv,$$

isto pela mudança de variáveis $v = t/\lambda$. Utilizando agora a expressão em (7.4) concluímos que

$$f(x) = \frac{\lambda}{2} e^{-\lambda|x|}, \quad \lambda > 0,$$

ou seja, a função de distribuição acima corresponde à função de densidade exponencial dupla de parâmetro $\lambda > 0$.

Teorema 7.13

A variável aleatória X tem distribuição simétrica em torno de zero se, e somente se, φ_X é uma função real $\forall t \in \mathbb{R}$.

Demonstração: Uma variável aleatória tem distribuição simétrica em torno de zero se

$$P(X \leq x) = P(X \geq -x), \quad \forall x \in \mathbb{R}.$$

Esta igualdade pode ser escrita como

$$P(X \leq x) = P(-X \leq x), \quad \forall x \in \mathbb{R},$$

isto significa que X e $-X$ são igualmente distribuídas, logo

$$\varphi_X(t) = \varphi_{-X}(t) = \overline{\varphi_{-X}(-t)} = \overline{E[e^{i(-t)(-X)}]} = \overline{E[e^{itX}]} = \varphi_X(t),$$

Mas, $c = \bar{c}$ se, e somente se, c é real. Então X é simétrica em torno de zero se, e somente se, $\varphi_X(t)$ é real $\forall t \in \mathbb{R}$. ■

Exemplo 7.11

Seja $\varphi(t) = \cos(at)$, sendo $a > 0$. Mostremos que φ é função característica para alguma variável aleatória X . Para demonstrar isto encontremos a função de distribuição correspondente a φ . Primeiro observamos que φ é real, então X é uma variável aleatória simétrica em torno de zero. Mais ainda, a função coseno é par, então $\cos(at) = \cos(-at)$. Significa que a variável aleatória X da qual φ é função característica assume somente os valores a e $-a$. Do fato de ser simétrica ambos valores são assumidos com a mesma probabilidade, assim temos que

$$P(X = a) = P(X = -a) = \frac{1}{2}.$$

Teorema 7.14

Se $E|X|^n < \infty$, então φ_X possui n derivadas contínuas e

$$\varphi_X^{(k)}(t) = \int_{-\infty}^{+\infty} (ix)^k e^{itx} f(x) dx,$$

para $k = 1, 2, \dots, n$.

Esta propriedade nos disse que $\varphi_x^{(k)}(0) = i^k \mathbb{E}(X^k)$, de modo que a função característica é uma espécie de função geradora de momentos. Aqui

$$\varphi_x^{(k)}(0) = \left. \frac{d^k \varphi_x(t)}{dt^k} \right|_{t=0}$$

é a k -ésima derivada de φ_x avaliada no ponto $t = 0$.

Demonstração: Dado que X é integrável queremos provar que $\varphi'_x(t) = \int_{-\infty}^{+\infty} (ix)e^{itx} f(x) dx$, restando somente justificar a diferenciação dentro da integral. Para quaisquer $t, h \in \mathbb{R}$, $h \neq 0$ temos que

$$\begin{aligned} \frac{\varphi_x(t+h) - \varphi_x(t)}{h} &= \int_{-\infty}^{+\infty} \frac{e^{i(t+h)x} - e^{itx}}{h} f(x) dx = \\ &= \int_{-\infty}^{+\infty} e^{itx} \frac{(e^{ihx} - 1)}{h} f(x) dx = \mathbb{E} \left[e^{itx} \frac{(e^{ihx} - 1)}{h} \right]. \end{aligned}$$

Dado que

$$\lim_{h \rightarrow 0} \left[\frac{(e^{ihx} - 1)}{h} \right] = ix, \quad \forall x \in \mathbb{R}$$

temos que

$$\lim_{h \rightarrow 0} \left[e^{itx} \frac{(e^{ihx} - 1)}{h} \right] = ix e^{itx}.$$

Por outro lado, $\forall x \in \mathbb{R}$,

$$\begin{aligned} \left| \frac{e^{ihx} - 1}{h} \right| &= \left| \frac{\int_0^h ix e^{iux} du}{h} \right| = |x| \left| \frac{\int_0^h e^{iux} du}{h} \right| \leq \\ &\leq |x| \frac{\int_0^h |e^{iux}| du}{h} = |x| \frac{\int_0^h 1 du}{h} = |x|, \end{aligned}$$

então

$$\left| e^{itx} \frac{(e^{ihx} - 1)}{h} \right| \leq |x|.$$

Logo, pelo Teorema da Convergência Dominada temos

$$\begin{aligned} \varphi'_x(t) &= \lim_{h \rightarrow 0} \frac{\varphi_x(t+h) - \varphi_x(t)}{h} = \lim_{h \rightarrow 0} \mathbb{E} \left[e^{itx} \frac{(e^{ihx} - 1)}{h} \right] \\ &= \mathbb{E}(iX e^{itX}) = \int_{-\infty}^{\infty} ix e^{itx} f(x) dx. \end{aligned}$$

A continuidade de $\varphi_x(t)$ no ponto t decorre de

$$\lim_{s \rightarrow t} ix e^{isx} = ix e^{itx} \quad \text{e que} \quad |ix e^{isx}| = |x|.$$

O restante da demonstração é obtida por indução em n . ■

Exemplo 7.12 (Continuação Exemplo 7.2)

Vamos obter a média e variância de X utilizando o Teorema 7.14. Sabemos que existem e com isto queremos dizer que são finitas, as esperanças de X para qualquer n . Então

$$i E(X) = \varphi'_X(0) = i\lambda e^{it} e^{\lambda(e^{it}-1)} \Big|_{t=0} = i\lambda.$$

Também

$$\begin{aligned} i^2 E(X^2) = \varphi''_X(0) &= \frac{d}{dt} \left[i\lambda e^{it} e^{\lambda(e^{it}-1)} \right] \Big|_{t=0} \\ &= i\lambda e^{-\lambda} e^{it+\lambda e^{it}} (i + i\lambda e^{it}) \Big|_{t=0} = i\lambda e^{-\lambda} e^{\lambda} (i + i\lambda) = i^2 \lambda(1 + \lambda), \end{aligned}$$

portanto, $E(X) = \lambda$, $E(X^2) = \lambda + \lambda^2$ e $\text{Var}(X) = \lambda$.

Aqui estabelecemos novamente o fato importante de que uma medida de probabilidade em \mathbb{R} é determinada exclusivamente por sua função característica.

Teorema 7.15 (Teorema da unicidade de funções geradoras de momentos)

Sejam X e Y variáveis aleatórias reais, com as respectivas funções geradoras de momentos reais M_X e M_Y e funções de distribuição F_X e F_Y . Se $M_X(u) = M_Y(u) < \infty$ para todos os u , em algum intervalo aberto não vazio, então $F_X(x) = F_Y(x)$, para todos $x \in \mathbb{R}$.

Demonstração: Chamemos $(a; b)$ o intervalo aberto não vazio no qual $M_X(u) = M_Y(u) < \infty$ para todos os u e seja

$$\mathcal{D} = \{z \in \mathbb{C} : a < \text{Re}(z) < b\},$$

sendo aqui $\text{Re}(z)$ a parte real do número complexo z .

Caso 1: $0 \in (a; b)$. Nesse caso, o eixo imaginário $I = \{it : t \in \mathbb{R}\}$ está contido em \mathcal{D} ($I \subseteq \mathcal{D}$). As funções característica complexas φ_X e φ_Y de X e Y existem, são diferenciáveis em \mathcal{D} e coincidem em $\{u + 0i : u \in (a; b)\}$ por suposição. Pelo Teorema 7.11, φ_X e φ_Y coincidem em todos os pontos de \mathcal{D} e, em particular, com I . Em outras palavras,

$$E(e^{itX}) = \varphi_X(t) = \varphi_Y(t) = E(e^{itY}), \quad \text{para todo } t \in \mathbb{R}$$

Assim, X e Y têm a mesma função característica e, portanto, a mesma distribuição.

Caso 2: $0 \notin (a; b)$. Tratamos isso por inclinação exponencial, da seguinte maneira. Para simplificar a exposição, suponha que X e Y tenham densidades f e g respectivamente. Coloque $\theta = (a + b)/2$ e definamos

$$f_\theta(x) = \frac{e^{\theta x} f(x)}{M_X(\theta)} \quad \text{e} \quad g_\theta(x) = \frac{e^{\theta x} g(x)}{M_Y(\theta)}. \quad (7.15)$$

As funções f_θ e g_θ são funções de densidade e as correspondentes funções geradoras de momentos são

$$M_\theta(u) = \frac{M_X(u + \theta)}{M_X(\theta)} \quad \text{e} \quad N_\theta(u) = \frac{M_Y(u + \theta)}{M_Y(\theta)}.$$

M_θ e N_θ coincidem e são finitos no intervalo $(a - \theta; b - \theta)$. Desde que esse intervalo contém 0, o argumento anterior mostra que as distribuições com densidades f_θ e g_θ coincidem. Resulta de (7.15) que as distribuições com densidades f e g coincidem. ■

Devemos lembrar o Exemplo 4.50. Nele apresentamos duas funções de densidade diferentes mas, com os mesmos momentos finitos. As funções de densidades utilizadas foram a densidade log-Normal, definida no Exemplo 4.54 como

$$f(x) = \frac{1}{\sqrt{2\pi}} \frac{1}{x^{1+\log(x)/2}}, \quad (7.16)$$

no caso padrão e uma família de densidade $g(x) = f(x)(1 + \sin(\log(x)))$. Para ambas variáveis o k -ésimo momento assume a forma

$$E(X^k) = e^{k^2/2}. \quad (7.17)$$

Um detalhe importantíssimo nesta situação é que, mesmo que os momentos existem e são finitos, no caso da densidade g a função geradora de momentos não existe.

Vejamos agora como identificar as funções de distribuição a partir das funções geradoras de momentos univocamente.

Teorema 7.16 (Teorema da unicidade por momentos)

Sejam X e Y variáveis aleatórias reais com funções de distribuição F_X e F_Y , respectivamente. Se

- (i) X e Y têm momentos finitos de todas as ordens,
- (ii) $E(X^k) = E(Y^k) = \alpha_k$ para todos os $k \in \mathbb{R}$ e
- (iii) o raio R de convergência da série de potências

$$\sum_{k=1}^{\infty} \frac{\alpha_k u^k}{k!}$$

é diferente de zero.

Então

$$F_X(x) = F_Y(x), \quad \text{para todo } x \in \mathbb{R}.$$

Demonstração: Sejam M_X e M_Y as funções geradoras de momentos de X e Y , respectivamente. Pelo Teorema 7.11

$$M_X(u) = \sum_{k=1}^{\infty} \frac{\alpha_k u^k}{k!} = M_Y(u),$$

para todos os u no intervalo aberto $(-R, R)$. Pelo Teorema 7.15 concluímos que $F_X = F_Y$. ■

No seguinte exemplo utilizaremos dois conceitos: um deles chamado de fatorial duplo⁴ e o segundo conhecido como teste da razão de convergência se séries (Spivak, 1994).

Exemplo 7.13

Suponha que X seja uma variável aleatória de forma que

$$\alpha_k = E(X^k) = \begin{cases} (k-1) \times (k-3) \times \cdots \times 5 \times 3 \times 1 & \text{caso } k \text{ seja par,} \\ 0 & \text{caso } k \text{ seja ímpar.} \end{cases}$$

Então, $X \sim N(0, 1)$, porque uma variável aleatória normal padrão tem esses momentos e a série

$$\sum_{k=1}^{\infty} \frac{\alpha_k u^k}{k!} \quad (7.18)$$

tem um raio infinito de convergência.

Provemos que a série em (7.18) converge. Primeiro observemos que podemos escrever $\alpha_k = (k-1)!!$, chamado de fatorial duplo, do qual resulta que

$$\sum_{k=1}^{\infty} \frac{\alpha_k u^k}{k!} = \sum_{n=1}^{\infty} \frac{(2n)! u^{2n}}{(2n)! 2^n n!} = \sum_{n=1}^{\infty} \frac{u^{2n}}{2^n n!}.$$

Qualquer seja o valor $u_0 \in \mathbb{R}$, a série

$$\sum_{n=1}^{\infty} \frac{(u_0^2)^n}{2^n n!}$$

é convergente segundo o teste da razão de convergência, resultando que a série de potências

$$\sum_{n=1}^{\infty} \frac{u^{2n}}{2^n n!} \text{ é uniformemente convergente em } [-u_0^2, u_0^2].$$

O detalhe no Exemplo 4.50 é que a série de potências

$$\sum_{k=1}^{\infty} \frac{\alpha_k u^k}{k!} = \sum_{k=1}^{\infty} \frac{e^{k^2/2} u^k}{k!}$$

converge apenas para $u = 0$. No referido exemplo, a esperança $E(X^k) = e^{k^2/2}$ corresponde aos momentos de uma variável aleatória com função de densidade $g(x) = f(x)(1 + \sin(\log(x)))$ e f a densidade log-Normal.

O próximo resultado, chamado de Critério de Polya⁵, é útil para construir exemplos de funções características. Uma excelente referência nesta parte é o livro de Durrett (1996).

⁴O produto de todos os números inteiros ímpares até um número inteiro ímpar positivo $k-1$ é chamado fatorial duplo de $k-1$ e definido por $(k-1)!! = \frac{k!}{2^{k/2}(k/2)!}$, sendo k um número inteiro par.

⁵George Pólya (1887-1985). Foi um matemático húngaro. Trabalhou com uma variedade de tópicos matemáticos, incluindo séries, teoria dos números, análise matemática, geometria, álgebra, combinatória e probabilidade.

Devemos esclarecer que o Critério de Polya nos permitirá, de maneira indireta, encontrar funções de densidade das quais não temos ideia da forma funcional.

Antes de nos perdermos nos detalhes da prova, devemos observar vários detalhes: funções convexas⁶, limites à esquerda e à direita⁷ e as partes positivas e negativas de uma função⁸

Teorema 7.17 (Critério de Polya)

Seja φ uma função real não negativa satisfazendo $\varphi(0) = 1$, $\varphi(t) = \varphi(-t)$, decrescente e convexa em $(0, \infty)$ com

$$\lim_{t \downarrow 0} \varphi(t) = 1, \quad \lim_{t \uparrow \infty} \varphi(t) = 0.$$

Então existe uma medida de probabilidade f em $(0, 1)$, de forma que

$$\varphi(t) = \int_0^\infty \left(1 - \left|\frac{t}{s}\right|\right)^+ f(s) ds, \quad (7.19)$$

e, portanto, φ é uma função característica.

A suposição de que $\lim_{t \downarrow 0} \varphi(t) = 1$ é necessária porque a função $\varphi(t) = 1_0(t)$ que é 1 em 0 e 0 caso contrário, satisfaz todas as outras hipóteses. Poderíamos permitir $\lim_{t \downarrow 0} \varphi(t) = c \geq 0$ tendo uma massa pontual de tamanho c em 0, mas deixamos essa extensão para o leitor.

Demonstração: Seja φ' a derivada à direita de φ , isto é,

$$\varphi'(t) = \lim_{h \downarrow 0} \frac{\varphi(t+h) - \varphi(t)}{h}.$$

Como φ é convexa, a derivada acima existe e é contínua e crescente. Podemos então escolher uma função F em $(0, \infty)$ com $F(b) - F(a) = \varphi'(b) - \varphi'(a)$ para todo $0 \leq a < b \leq \infty$ e seja $F' = f$.

Dado que $\lim_{t \rightarrow \infty} \varphi'(t) = 0$, então

$$-\varphi'(s) = \int_s^\infty r^{-1} f(r) dr.$$

6

Dizemos que uma função $f : I \rightarrow \mathbb{R}$ é convexa se para cada $x, y \in I$ e para cada $t \in [0, 1]$ tivermos que $f(tx + (1-t)y) \leq tf(x) + (1-t)f(y)$.

7

Suponhamos que $f : A \rightarrow \mathbb{R}$ seja uma função e c um ponto de acumulação de A . O limite à esquerda de f em c é L_1 , denotado $\lim_{x \uparrow c} f(x) = L_1$, se $\forall \epsilon > 0$, $\exists \delta > 0$ tal que se $x \in A$ e $x < c$, então $|f(x) - L_1| < \epsilon$. Da mesma forma, o limite à direita de f em c é L_2 , denotado $\lim_{x \downarrow c} f(x) = L_2$, se $\forall \epsilon > 0$, $\exists \delta > 0$ tal que se $x \in A$ e $x > c$ então $|f(x) - L_2| < \epsilon$.

8

Seja f uma função f no intervalo I . A Parte Positiva de f denotada f^+ é a função definida para todos $x \in I$ por $f^+(x) = \max\{f(x), 0\}$. Da mesma forma, a Parte Negativa de f denotada f^- é a função definida para todos $x \in I$ por $f^-(x) = \max\{-f(x), 0\}$. É importante notar que $f^+(x) \geq 0$ e $f^-(x) \geq 0$ para todos $x \in I$. Analisaremos mais detalhadamente algumas das propriedades dessas funções no seguinte teorema. Mais ainda $f = f^+ - f^-$ e $|f| = f^+ + f^-$.

Integrando, temos que, para $t \geq 0$

$$\begin{aligned}\varphi(t) &= \int_t^\infty \int_s^\infty r^{-1} f(r) \, dr \, ds = \int_t^\infty r^{-1} \int_t^r ds f(r) \, dr \\ &= \int_t^\infty \left(1 - \frac{t}{r}\right) f(r) \, dr = \int_0^\infty \left(1 - \frac{t}{r}\right)^+ f(r) \, dr.\end{aligned}$$

Usando que $\varphi(?t) = \varphi(t)$ estendemos a expressão acima para $t \leq 0$, temos (7.19). Escolhendo $t = 0$ provamos que $\int_{-\infty}^\infty f(s) \, ds = 1$. ■

Algumas considerações acerca da demonstração do Teorema 7.17. Primeiro, o teorema foi escrito pensando em funções de densidade e assim foi realizado o desenvolvimento da demonstração. Ainda temos que considerar que a função de distribuição F pode ser linear por partes e, portanto, F deve possuir um número finito de pontos de descontinuidade, sendo válido mesmo assim o desenvolvimento apresentado. Por último, para provar o resultado geral, seja F_n uma sequência de medidas de probabilidade em $(0, 1)$ com um número finito de pontos de descontinuidade que converge para cada $x \in \mathbb{R}$ e seja

$$\varphi_n(t) = \int_0^\infty \left(1 - \left|\frac{t}{s}\right|\right)^+ f_n(s) \, ds.$$

Devido a que $(1 - |t/s|)^+$ é limitada e contínua, $\lim_{n \rightarrow \infty} \varphi_n(t) = \varphi(t)$ do qual pode-se demonstrar o teorema anterior no caso geral.

Exemplo 7.14

Mostremos que $\varphi(t) = \exp(-|t|^\alpha)$ é função característica para $0 < \alpha \leq 2$.

Duas situações importantes:

(i) Caso $\alpha = 1$, corresponde à função característica da densidade *Cauchy*(0, 1).

(ii) Caso $\alpha = 2$, corresponde à função característica da densidade *Normal*(0, 1).

É claro que para qualquer $0 < \alpha < 2$ e $\alpha \neq 1$, a cada função característica da forma $\exp(-|t|^\alpha)$, corresponde uma função de densidade (Teorema 7.9, Fórmula de Inversão) da qual não temos ideia da forma funcional.

Vejamos agora o caso geral $0 < \alpha < 2$. Um pequeno cálculo mostra que, para qualquer β e $|x| < 1$, temos

$$(1 - x)^\beta = \sum_{n=0}^{\infty} \binom{\beta}{n} (-x)^n,$$

onde $\binom{\beta}{n} = \frac{\beta(\beta-1)\cdots(\beta-n+1)}{1 \cdot 2 \cdots n}$. Seja agora

$$\psi(t) = 1 - (1 - \cos(t))^{\frac{\alpha}{2}} = \sum_{n=1}^{\infty} c_n \cos^n(t),$$

sendo que $c_n = \binom{\alpha/2}{n} (-1)^{n+1}$, $c_n \geq 0$ e $\sum_{n=1}^{\infty} c_n = 1$, isto percebe-se assumindo $t = 0$.

Pelo Exemplo 7.11 percebemos que $\varphi(t) = \cos(t)$ é função característica e ainda vemos que

$$\lim_{t \rightarrow 0} (1 - \cos(t)) = \frac{t^2}{2},$$

de maneira que

$$1 - \cos(t2^{1/2}n^{-1/\alpha}) \cong \frac{t}{n^{2/\alpha}}.$$

Disto segue que

$$\lim_{n \rightarrow \infty} \left(\varphi(t2^{1/2}n^{-1/\alpha}) \right)^n = \exp(-|t|^\alpha)$$

é função característica.

Nosso objetivo agora é demonstrar que se a sequência de funções de distribuição $\{F_n\}_{n=1}^\infty$ converge pontualmente a F , uma função de distribuição se, e somente se, a sequência de funções características $\{\varphi_n\}_{n=1}^\infty$ converge a uma função característica φ . A importância dos resultados referidos será melhor entendida na Seção 7.3 e no Capítulo 8, chamados de resultados de convergência.

Provaremos isto em duas partes, a necessidade no Teorema 7.18, conhecido como Teorema de Helly⁹-Bray¹⁰ e a suficiência com o Teorema 7.19, o Teorema da Continuidade de Paul Lévy¹¹.

Teorema 7.18 (Helly-Bray)

Sejam F e $\{F_n\}_{n=1}^\infty$ funções de distribuição. Se $\lim_{n \rightarrow \infty} F_n(x) = F(x)$, para todo ponto de continuidade de F , então

$$\lim_{n \rightarrow \infty} \int_{-\infty}^{\infty} g(x) f_n(x) d(x) = \int_{-\infty}^{\infty} g(x) f(x) dx,$$

para toda função real $g : \mathbb{R} \rightarrow \mathbb{R}$, contínua e limitada, sendo $\{f_n\}_{n=1}^\infty$ a sequência de funções de densidade das variáveis aleatórias X , $\{X_n\}_{n=1}^\infty$ corresponde às funções de distribuição.

Demonstração: Para $-\infty < a < b < \infty$,

$$\begin{aligned} \left| \int_{-\infty}^{\infty} g(x) f_n(x) dx - \int_{-\infty}^{\infty} g(x) f(x) dx \right| &\leq \left| \int_{-\infty}^{\infty} g(x) f_n(x) dx - \int_a^b g(x) f_n(x) dx \right| \\ &+ \left| \int_a^b g(x) f_n(x) dx - \int_a^b g(x) f(x) dx \right| + \left| \int_a^b g(x) f(x) dx - \int_{-\infty}^{\infty} g(x) f(x) dx \right|. \end{aligned}$$

⁹Eduard Helly (1884-1943). Foi um matemático suíço.

¹⁰Hubert Evelyn Bray (1889-1978). Foi um matemático norte-americano.

¹¹Paul Pierre Lévy (1886-1971). Foi um matemático francês. Trabalhou entre outros assuntos com a teoria das probabilidades e os processos estocásticos.

Seja $c = \sup_{x \in \mathbb{R}} |g(x)| < \infty$, o qual acontece por hipótese, e seja $\epsilon > 0$. Temos então que

$$\begin{aligned} \left| \int_a^b g(x)f(x) dx - \int_{-\infty}^{\infty} g(x)f(x) dx \right| &= \left| \int_{-\infty}^a g(x)f(x) dx + \int_b^{\infty} g(x)f(x) dx \right| \\ &\leq \int_{-\infty}^a cf(x) dx + \int_b^{\infty} cf(x) dx \\ &= c[F(a) + (1 - F(b))]. \end{aligned}$$

Podemos escolher a e b , pontos de continuidade de F , tais que

$$\left| \int_a^b g(x)f(x) dx - \int_{-\infty}^{\infty} g(x)f(x) dx \right| \leq c[F(a) + (1 - F(b))] < \epsilon, \quad (7.20)$$

devido a que

$$\lim_{\substack{a \rightarrow -\infty \\ b \rightarrow \infty}} [F(a) + (1 - F(b))] = 0,$$

e para que a segunda desigualdade em (7.20) seja satisfeita bastaria escolher a suficientemente pequeno e b suficientemente grande, lembrando que os pontos de continuidade de F são densos. Para esses valores de a e b , temos

$$\begin{aligned} \left| \int_{-\infty}^{\infty} g(x)f_n(x) dx - \int_a^b g(x)f_n(x) dx \right| &\leq \int_{-\infty}^a cf_n(x) dx + \int_b^{\infty} cf_n(x) dx \\ &= c[F_n(a) + (1 - F_n(b))], \end{aligned}$$

o qual satisfaz que

$$\lim_{n \rightarrow \infty} c[F_n(a) + (1 - F_n(b))] = c[F(a) + (1 - F(b))]$$

e, logo

$$\left| \int_{-\infty}^{\infty} g(x)f_n(x) dx - \int_a^b g(x)f_n(x) dx \right| + \left| \int_a^b g(x)f_n(x) dx - \int_a^b g(x)f(x) dx \right| < 2\epsilon,$$

para n suficientemente grande. Provemos agora que

$$\left| \int_a^b g(x)f_n(x) dx - \int_a^b g(x)f(x) dx \right| \leq 2\epsilon,$$

para todo n , suficientemente grande. Para isso, sejam a e b os pontos escolhidos anteriormente e observemos que g é contínua no intervalo $[a, b]$, devido a isso podemos escolher x_0, x_1, \dots, x_N pontos de continuidade de F , tais que $a = x_0 < x_1 < \dots < x_N = b$ e nos quais $|g(x) - g(x_i)| < \epsilon$, para todo $x \in [x_i, x_{i+1}]$, $i = 0, 1, 2, \dots, N-1$. Então

$$\begin{aligned} m_{n_i} &= (g(x_i) - \epsilon)(F_n(x_{i+1}) - F_n(x_i)) \leq \int_{x_i}^{x_{i+1}} g(x)f_n(x) dx \\ &\leq (g(x_i) + \epsilon)(F_n(x_{i+1}) - F_n(x_i)) = M_{n_i}, \\ m_i &= (g(x_i) - \epsilon)(F(x_{i+1}) - F(x_i)) \leq \int_{x_i}^{x_{i+1}} g(x)f(x) dx \\ &\leq (g(x_i) + \epsilon)(F(x_{i+1}) - F(x_i)) = M_i. \end{aligned}$$

Logo,

$$m_{n_i} - M_i \leq \int_{x_i}^{x_{i+1}} g(x)f_n(x) dx - \int_{x_i}^{x_{i+1}} g(x)f(x) dx \leq M_{n_i} - m_i,$$

para todo $i = 0, 1, 2, \dots, N-1$. Somando temos

$$\sum_{i=0}^{N-1} (m_{n_i} - M_i) \leq \int_a^b g(x)f_n(x) dx - \int_a^b g(x)f(x) dx \leq \sum_{i=0}^{N-1} (M_{n_i} - m_i).$$

Dado que os pontos x_i são de continuidade de F e F_n converge a F , temos que

$$\lim_{n \rightarrow \infty} m_{n_i} = m_i \quad \text{e} \quad \lim_{n \rightarrow \infty} M_{n_i} = M_i,$$

para todo $i = 0, 1, 2, \dots, N-1$. Logo

$$\lim_{n \rightarrow \infty} \sum_{i=0}^{N-1} (m_{n_i} - M_i) = \sum_{i=0}^{N-1} (m_i - M_i) = -2\epsilon (F(b) - F(a)) \geq -2\epsilon$$

e

$$\lim_{n \rightarrow \infty} \sum_{i=0}^{N-1} (M_{n_i} - m_i) = \sum_{i=0}^{N-1} (M_i - m_i) = 2\epsilon (F(b) - F(a)) \leq 2\epsilon.$$

Temos então que, para todo n suficientemente grande

$$\left| \int_a^b g(x)f_n(x) dx - \int_a^b g(x)f(x) dx \right| \leq 2\epsilon. \quad \blacksquare$$

O enunciado apresentado do Teorema de Helly-Bray assim como a demonstração foram específicos para variáveis aleatórias contínuas mas, este resultado vale qualquer seja o tipo de variável aleatória considerado (Billingsley, 1968). Em particular, como as funções $\cos(tx)$ e $\sin(tx)$ são contínuas e limitadas para t conhecido, concluímos que

$$\lim_{n \rightarrow \infty} E(\cos(tX)) = E(\cos(tX)) \quad \text{e} \quad \lim_{n \rightarrow \infty} E(\sin(tX)) = E(\sin(tX)),$$

logo

$$\lim_{n \rightarrow \infty} \varphi_{X_n}(t) = \varphi_X(t), \quad \forall t \in \mathbb{R},$$

ou seja, sequência de funções características converge, para cada t , a uma função característica.

Para fins da teoria da probabilidade, uma das propriedades mais importantes da função característica é dada a seguir, chamado de Teorema da Continuidade, devido a Paul Lévy¹² e Harald Cramér¹³.

¹²Paul Pierre Lévy (1886-1971). Foi um matemático francês. Trabalhou entre outros assuntos com a teoria das probabilidades, Martingales, etc.

¹³Carl Harald Cramér (1893-1985). Foi um matemático e estatístico sueco. Especialista em estatística matemática e teoria das probabilidades.

Teorema 7.19 (Teorema de Continuidade)

Seja $\{F_n\}_{n=1}^\infty$ uma sequência de funções de distribuição e $\{\varphi_n\}_{n=1}^\infty$ a sequência das correspondentes funções características. Se $\lim_{n \rightarrow \infty} \varphi_n(t) = \varphi(t)$, qualquer seja t e se φ é contínua no ponto zero, então

(i) existe uma função de distribuição F , tal que, $\lim_{n \rightarrow \infty} F_n(x) = F(x)$, sendo x um ponto de continuidade de F e

(ii) φ é a função característica de F .

Demonstração: Pelas hipóteses do teorema, (i) implica (ii), segundo o Teorema 7.18, Teorema de Helly-Bray. O detalhe agora é demonstrar que a sequência $\{F_n\}_{n=1}^\infty$ de funções de distribuição converge a alguma função de distribuição. Neste sentido, vamos demonstrar que existe uma subsequência F_{n_k} de funções de distribuição e uma função de distribuição F , de forma que $\lim_{k \rightarrow \infty} F_{n_k}(x) = F(x)$, sendo x um ponto de continuidade de F .

Demonstremos que existe uma subsequência F_{n_1}, F_{n_2}, \dots da sequência $\{F_n\}_{n=1}^\infty$, e uma função $F : \mathbb{R} \rightarrow [0, 1]$, de forma que F seja não decrescente, contínua à direita e tal que $\lim_{k \rightarrow \infty} F_{n_k}(x) = F(x)$, para todo ponto de continuidade de F . Sejam q_1, q_2, \dots os números racionais. Escolhemos uma sequência $1 \leq n_1 < n_2 < \dots$ de números inteiros positivos tais que $F_{n_k}(q_r)$ converge, quando $k \rightarrow \infty$, para cada r fixo. Seja este limite $F(q_r)$, de modo que $\lim_{k \rightarrow \infty} F_{n_k}(q_r) = F(q_r)$, para todo r . É claro que $0 \leq F(q_r) \leq 1$ e que F é não decrescente nos racionais. Seja agora x um número irracional qualquer e definamos

$$F(x) = \lim_{\substack{q \rightarrow x \\ q \text{ racionais}}} F(q).$$

A função assim definida F é não decrescente, mas não necessariamente contínua à direita. Para provarmos isso, consideremos x um ponto de continuidade de F e sejam q_1 e q_2 números racionais tais que $q_1 < x < q_2$ e $F(q_2) - \epsilon < F(x) < F(q_1) + \epsilon$, $\forall \epsilon > 0$. Então,

$$\begin{aligned} F(x) - \epsilon < F(q_1) &= \lim_{k \rightarrow \infty} F_{n_k}(q_1) \leq \liminf_{k \rightarrow \infty} F_{n_k}(x) \leq \limsup_{k \rightarrow \infty} F_{n_k}(x) \\ &\leq \lim_{k \rightarrow \infty} F_{n_k}(q_2) = F(q_2) < F(x) + \epsilon. \end{aligned}$$

Como ϵ é arbitrário, então $\lim_{k \rightarrow \infty} F_{n_k}(x) = F(x)$. Podemos redefinir F , nos seus pontos de descontinuidade para torná-la contínua à direita.

Restaria demonstrar que a função F satisfaz que $F(+\infty) = 1$ e que $F(-\infty) = 0$. Seja g uma função característica qualquer e G a função de distribuição associada. Então, a função característica integral é definida como

$$\begin{aligned} \int_0^t g(s) ds &= \int_0^t \left(\int_{-\infty}^{\infty} e^{isx} g(x) dx \right) ds = \int_{-\infty}^{\infty} \left(\int_0^t e^{isx} g(x) dx \right) ds \\ &= \int_{-\infty}^{\infty} \frac{e^{itx} - 1}{ix} g(x) dx. \end{aligned}$$

A ordem de integração pode ser trocada porque o integrando é limitado, isso é provado definindo $(e^{itx} - 1)/ix = t$, quando $x = 0$; para $x \neq 0$ esta função é limitada e contínua. Além disso, $\lim_{|x| \rightarrow \infty} (e^{itx} - 1)/ix = 0$.

Seja agora f_{n_1}, f_{n_2}, \dots a sequência de funções de densidade correspondentes à sub-sequência F_{n_1}, F_{n_2}, \dots . Então

$$\lim_{k \rightarrow \infty} \int_{-\infty}^{\infty} \frac{e^{itx} - 1}{ix} f_{n_k}(x) dx = \int_{-\infty}^{\infty} \frac{e^{itx} - 1}{ix} f(x) dx,$$

portanto

$$\lim_{k \rightarrow \infty} \int_0^t \varphi_{n_k}(s) ds = \int_0^t \left(\int_{-\infty}^{\infty} e^{isx} f(x) dx \right) ds.$$

Mas $\lim_{k \rightarrow \infty} \varphi_{n_k}(s) = \varphi(s)$, $\forall s$ concluindo-se que φ é contínua no zero, o qual implica que φ é limitada e integrável, logo pelo Teorema da Convergência Dominada¹⁴

$$\lim_{k \rightarrow \infty} \int_0^t \varphi_{n_k}(s) ds = \int_0^t \varphi(s) ds.$$

Resulta que

$$\frac{1}{t} \int_0^t \varphi(s) ds = \frac{1}{t} \int_0^t \left(\int_{-\infty}^{\infty} e^{isx} f(x) dx \right) ds, \quad t \neq 0,$$

disto concluímos que

$$\varphi(0) = \lim_{t \rightarrow 0} \frac{1}{t} \int_0^t \varphi(s) ds = \int_{-\infty}^{\infty} \left(\lim_{t \rightarrow 0} \frac{1}{t} \int_0^t e^{isx} \right) ds f(x) dx = F(+\infty) - F(-\infty).$$

Dado que $\varphi(0) = \lim_{k \rightarrow \infty} \varphi_{n_k}(0) = 1$, concluímos que $F(+\infty) = 1$ e $F(-\infty) = 0$. ■

As aplicações destes dois últimos teoremas ficaram claras nas próximas seções. Devemos esclarecer que na demonstração anterior, some provamos que existe um sub-sequência F_{n_1}, F_{n_2}, \dots da sequência $\{F_n\}_{n=1}^{\infty}$ que converge pontualmente, ou seja, $\lim_{k \rightarrow \infty} F_{n_k}(x) = F(x)$, $\forall x \in \mathbb{R}$. Agora, isso é suficiente para a demonstração?

A resposta é afirmativa. Para vermos isso, suponhamos que a sequência F_n não convirja a F , mas a subsequência F_{n_k} sim converge a F , $\forall x \in \mathbb{R}$. Significa que existe x , ponto de continuidade de F e uma outra subsequência F_{m_k} , tais que $\lim_{k \rightarrow \infty} F_{m_k} = a \neq F(x)$. Como essa nova subsequência também satisfaz as condições do Teorema 7.19, existirá uma subsequência F_{p_k} da subsequência F_{m_k} e uma função de distribuição G , tais que $\lim_{k \rightarrow \infty} F_{p_k}(x) = G(x)$. Resulta que F e G terão a mesma função característica, logo $F(x) = G(x)$, $\forall x$, logo $\lim_{k \rightarrow \infty} F_{p_k}(x) = G(x) = F(x)$, o qual é um absurdo, portanto $F_n(x)$ não converge a $F(x)$, $\forall x$.

¹⁴Teorema da Convergência Dominada: seja $\{f_n\}$ uma sequência de funções integráveis tais que $\lim_{n \rightarrow \infty} f_n(x) = f(x)$ e dominada, ou seja, $|f_n(x)| \leq g(x)$. Então

$$\lim_{n \rightarrow \infty} \int_{-\infty}^{\infty} f_n(x) dx = \int_{-\infty}^{\infty} f(x) dx.$$

7.1.2 Função característica de vetores aleatórios

Discutiremos muito brevemente a função característica no espaço de mais de uma dimensão. O conceito de função característica pode ser generalizado à vetores aleatórios de duas ou mais dimensões na seguinte definição.

Definição 7.2

Seja $\mathbf{X} = (X_1, \dots, X_n)$ um vetor aleatório de dimensão n . A função característica do vetor \mathbf{X} é a função $\varphi_{\mathbf{X}} : \mathbb{R}^n \rightarrow \mathbb{C}$, definida por

$$\varphi_{\mathbf{X}}(\mathbf{t}) = \mathbb{E} \left(\exp \left(i \sum_{k=1}^n t_k X_k \right) \right), \quad \forall \mathbf{t} = (t_1, \dots, t_n) \in \mathbb{R}^n. \quad (7.21)$$

Exemplo 7.15

Consideremos o vetor aleatório $\mathbf{X} = (X_1, X_2)$ com função de densidade conjunta exponencial, queremos encontrar a função característica vetorial. Segundo a Definição 7.2

$$\varphi_{\mathbf{X}}(\mathbf{t}) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp(i\mathbf{t}\mathbf{x}^T) f(\mathbf{x}) d\mathbf{x},$$

sendo que $\mathbf{t} = (t_1, t_2)$ e $\mathbf{x} = (x_1, x_2)$ as estimativas do vetor aleatório \mathbf{X} . Esta expressão é específica no caso absolutamente contínuo, considerando f a função de densidade vetorial correspondente.

Sejam W_0 , W_1 e W_2 variáveis aleatórias exponenciais independentes com parâmetros λ_0 , λ_1 e λ_2 , respectivamente. A função de densidade exponencial bivariada, proposta em Ghosh & Alzaatreh (2015) assume a forma

$$f(\mathbf{x}) = k \exp(-\lambda_1 x_1 - \lambda_2 x_2 - \lambda_0 \max\{x_1, x_2\}), \quad (7.22)$$

onde $k = \frac{\lambda(\lambda_1 + \lambda_0)(\lambda_2 + \lambda_0)}{2\lambda_0 + \lambda_1 + \lambda_2}$, $\lambda = \lambda_0 + \lambda_1 + \lambda_2$ e cada $\lambda_i > 0$, $i = 0, 1, 2$.

Por simples integração, a função característica da distribuição exponencial bivariada em (7.22) assume a forma

$$\varphi_{\mathbf{X}}(\mathbf{t}) = \begin{cases} \frac{\lambda(\lambda_0 + \lambda_1)(\lambda_0 + \lambda_2)}{(\lambda + \lambda_0)(\lambda_0 + \lambda_1 - it_1)(\lambda - it_1 - it_2)} & \text{caso } \max\{x_1, x_2\} = x_1 \\ \frac{\lambda(\lambda_0 + \lambda_1)(\lambda_0 + \lambda_2)}{(\lambda + \lambda_0)(\lambda_0 + \lambda_2 - it_2)(\lambda - it_1 - it_2)} & \text{caso } \max\{x_1, x_2\} = x_2 \end{cases}.$$

Percebemos que a função característica do vetor \mathbf{X} pode ser escrita matricialmente como $\varphi_{\mathbf{X}}(\mathbf{t}) = \mathbb{E}(e^{i\mathbf{t}\mathbf{X}^T})$. A função característica de um vetor aleatório tem propriedades análogas às propriedades de uma função característica de uma variável aleatória unidimensional.

Por exemplo, a Fórmula de Inversão bidimensional no caso absolutamente contínuo pode ser formulada da seguinte forma

$$P(x_1 < X < x_2, y_1 < Y < y_2) = \lim_{T \rightarrow \infty} \frac{1}{(2\pi)^2} \int_{-T}^T \int_{-T}^T \frac{e^{-isx_1} - e^{-isx_2}}{is} \frac{e^{-itx_1} - e^{-itx_2}}{it} f(s, t) ds dt,$$

em cada retângulo de continuidade $\{(x, y) : x_1 < x < x_2, y_1 < y < y_2\}$ da função de densidade bidimensional. A prova é inteiramente semelhante à de uma dimensão. Segue-se, como lá, que f determina unicamente a função característica.

A seguir apresentamos um teorema que nos fornece um critério para verificar independência de vetores aleatórios utilizando suas funções características.

Teorema 7.20

Sejam $\mathbf{X} = (X_1, \dots, X_n)$ e $\mathbf{Y} = (Y_1, \dots, Y_m)$ vetores aleatórios. Os vetores \mathbf{X} e \mathbf{Y} são independentes se, e somente se,

$$\varphi_{(\mathbf{X}, \mathbf{Y})}(\mathbf{s}, \mathbf{t}) = \varphi_{\mathbf{X}}(\mathbf{s}) \times \varphi_{\mathbf{Y}}(\mathbf{t}), \quad \forall (\mathbf{s}, \mathbf{t}) \in \mathbf{R}^n \times \mathbf{R}^m. \quad (7.23)$$

Demonstração: Suponhamos que os vetores \mathbf{X} e \mathbf{Y} sejam independentes. Então, as variáveis aleatórias $e^{is\mathbf{X}^\top}$ e $e^{it\mathbf{Y}^\top}$ são independentes, logo

$$\begin{aligned} \varphi_{(\mathbf{X}, \mathbf{Y})}(\mathbf{s}, \mathbf{t}) &= E\left(e^{i(\mathbf{s}\mathbf{X}^\top + \mathbf{t}\mathbf{Y}^\top)}\right) = E\left(e^{is\mathbf{X}^\top} e^{it\mathbf{Y}^\top}\right) \\ &= E\left(e^{is\mathbf{X}^\top}\right) E\left(e^{it\mathbf{Y}^\top}\right) = \varphi_{\mathbf{X}}(\mathbf{s}) \times \varphi_{\mathbf{Y}}(\mathbf{t}). \end{aligned} \quad \blacksquare$$

No Teorema 7.20 demonstramos um resultado válido para duas variáveis aleatórias, é claro que podemos generalizar este resultado para um número finito de vetores aleatórios. Assim, se X_1, X_2, \dots, X_n forem variáveis aleatórias temos que X_1, X_2, \dots, X_n serão independentes se, e somente se,

$$\varphi_{(X_1, X_2, \dots, X_n)}(\mathbf{t}) = \prod_{k=1}^n \varphi_{X_k}(t_k), \quad \forall \mathbf{t} = (t_1, t_2, \dots, t_n) \in \mathbf{R}^n. \quad (7.24)$$

Observando o Exemplo 7.15 vamos demonstrar que o vetor (X_1, X_2) não é composto de variáveis aleatórias independentes. Para isso primeiro vemos que no artigo de Ghosh & Alzaatreh (2015), os autores encontraram que as funções de densidade marginais de X_1 e X_2 assumem a forma

$$f_{X_1}(x_1) = a_1(\lambda_0 + \lambda_1)e^{-(\lambda_0 + \lambda_1)x_1} - a_2\lambda e^{-\lambda x_1}, \quad x_1 > 0$$

e

$$f_{X_2}(x_2) = a_3(\lambda_0 + \lambda_2)e^{-(\lambda_0 + \lambda_2)x_2} - a_4\lambda e^{-\lambda x_2}, \quad x_2 > 0,$$

onde

$$a_1 = \frac{k}{\lambda_2(\lambda_0 + \lambda_1)}, \quad a_2 = \frac{k\lambda_0}{\lambda\lambda_2(\lambda_0 + \lambda_2)}, \quad a_3 = \frac{k}{\lambda_1(\lambda_0 + \lambda_2)}, \quad a_4 = \frac{k\lambda_0}{\lambda\lambda_1(\lambda_0 + \lambda_1)},$$

k e λ , como definidos no Exemplo 7.15.

Encontremos as funções características marginais

$$\varphi_{X_1}(s) = \frac{a_1(\lambda_0 + \lambda_1)}{\lambda_0 + \lambda_1 - is} - \frac{a_2\lambda}{\lambda - is}$$

e

$$\varphi_{X_2}(t) = \frac{a_3(\lambda_0 + \lambda_2)}{\lambda_0 + \lambda_2 - it} - \frac{a_4\lambda}{\lambda - it}.$$

Fica claro que

$$\varphi_{(X_1, X_2)}(s, t) \neq \varphi_{X_1}(s) \times \varphi_{X_2}(t).$$

Então, segundo o Teorema 7.20, as variáveis X_1 e X_2 não são independentes.

Exemplo 7.16

Nosso objetivo aqui será encontrar a função característica da distribuição normal multivariada.

Seja \mathbf{X} um vetor aleatório de dimensão n , como definido na Seção 6.32, dizemos que o vetor \mathbf{X} tem a distribuição normal multivariada se existe um vetor $\mu \in \mathbb{R}^n$ e uma matriz semi-definida positiva simétrica $\Sigma \in \mathbb{R}^{n \times n}$, tal que a função de densidade conjunta é da forma

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} \exp \left(-\frac{1}{2} (\mathbf{x} - \mu) \Sigma^{-1} (\mathbf{x} - \mu)^\top \right),$$

onde $\mathbf{x} = (x_1, \dots, x_n)$ é o vetor de valores do vetor aleatório X e $\mu = (\mu_1, \dots, \mu_n)$ o vetor de esperanças.

Pela Definição 7.2, obtemos que a função característica do vetor \mathbf{X} é da forma

$$\varphi_{\mathbf{X}}(\mathbf{t}) = \exp \left(it\mu^\top - \frac{1}{2} \mathbf{t} \Sigma \mathbf{t}^\top \right),$$

onde $\mathbf{t} = (t_1, \dots, t_n) \in \mathbb{R}^n$ e $\mu = (\mu_1, \dots, \mu_n)$ representa o vetor de esperanças de cada variável que compõe o vetor \mathbf{X} .

O seguinte resultado reduz o problema de determinação de uma distribuição multivariada à situação univariada.

Teorema 7.21

Seja $\mathbf{X} = (X_1, \dots, X_n)$ um vetor aleatório de dimensão n . A distribuição do vetor \mathbf{X} é completamente determinada pelas distribuições das combinações lineares

$$\mathbf{t} \mathbf{X}^\top = \sum_{k=1}^n t_k X_k, \quad \mathbf{t} = (t_1, \dots, t_n) \in \mathbb{R}^n.$$

Demonstração: Pelo Teorema 7.11 sabemos que a distribuição de \mathbf{X} é determinada pela função característica $\varphi_{\mathbf{X}}$, então

$$\varphi_{\mathbf{X}}(\mathbf{t}) = \mathbb{E}\left(e^{it\mathbf{X}^\top}\right) = \varphi_{\mathbf{tX}^\top}(1), \quad \forall \mathbf{t} \in \mathbb{R}^n.$$

■

7.1.3 Distribuições infinitamente divisíveis

Teoremas de limites clássicos assim como generalizações, que constituem o conteúdo principal desta capítulo, têm a ver com somas do tipo

$$X = \xi_1 + \xi_2 + \cdots + \xi_n,$$

de um número crescente de termos de uma sequência $\xi_1, \xi_2, \dots, \xi_n, \dots$ de variáveis aleatórias independentes. Neste capítulo vai aparecer que as distribuições infinitamente divisíveis desempenham um papel fundamental, mesmo nos problemas clássicos de teoremas limite para somas de variáveis aleatórias discretas independentes.

Definição 7.3

Dizemos que a variável aleatória X é infinitamente divisível se, para todo número natural n , X pode ser escrita como a soma

$$X = \xi_1 + \xi_2 + \cdots + \xi_n,$$

de n variáveis aleatórias $\xi_1, \xi_2, \dots, \xi_n$ independentes e igualmente distribuídas.

As variáveis aleatórias infinitamente divisíveis podem ser discretas ou contínuas e ainda devemos observar que a distribuição da variável X não é necessariamente a mesma das variáveis $\xi_1, \xi_2, \dots, \xi_n$. Um outro detalhe interessante é que se X for infinitamente divisível não significa que todas as variáveis $\xi_1, \xi_2, \dots, \xi_n$ também o sejam.

As funções de distribuição de variáveis aleatórias infinitamente divisíveis serão chamadas de funções de distribuição infinitamente divisíveis. Obviamente, a função de distribuição F é infinitamente divisível se, e somente se, sua função característica φ é.

Teorema 7.22

Seja X uma variável aleatória infinitamente divisível com função característica φ_X . Então, para todo número natural n , φ_X é a n -ésima potência de alguma função característica φ_{ξ_n} , isto é,

$$\varphi_X(t) = [\varphi_{\xi_n}(t)]^n. \quad (7.25)$$

Demonstração: Sabemos, pela Definição 7.3, que para todo número natural n , X pode ser escrita como a soma $X = \xi_1 + \xi_2 + \cdots + \xi_n$, onde $\xi_1, \xi_2, \dots, \xi_n$ são independentes e igualmente distribuídas cada uma com função característica φ_{ξ_n} . Pelo resultado em (7.8), sabemos que a função característica da soma de variáveis aleatórias independentes e igualmente distribuídas é

$$\varphi_X(t) = \prod_{k=1}^n \varphi_{\xi_k}(t),$$

mas, as variáveis $\xi_1, \xi_2, \dots, \xi_n$ são independentes e igualmente distribuídas. Logo, todas têm a mesma função característica φ_{ξ_n} , então

$$\varphi_X(t) = [\varphi_{\xi_n}(t)]^n.$$

■

Exemplo 7.17

A variável aleatória X , normalmente distribuída, é infinitamente divisível. Suponhamos que $E(X) = \mu$ e $\text{Var}(X) = \sigma^2$, sabemos que a função característica de X é

$$\varphi_X(t) = e^{i\mu t - \frac{\sigma^2}{2}t^2}.$$

Dado que, para todo $n > 0$, $\varphi_{\xi_n}(t) = e^{i\frac{\mu}{n}t - \frac{\sigma^2}{2n}t^2}$ é a função característica de uma variável aleatória ξ_n com distribuição normal de média μ/n e variância σ^2/n .

A expressão inversa de (7.25)

$$\varphi_{\xi_n}(t) = \sqrt[n]{\varphi_X(t)},$$

não determina exclusivamente os valores de $\varphi_{\xi_n}(t)$ em termos dos valores de $\varphi_X(t)$, a n -ésima raiz tem n valores. Requisitos adicionais tornarão possível determinar exclusivamente $\varphi_{\xi_n}(t)$ em todo intervalo de t .

Exemplo 7.18

Foi obtido no Exemplo 7.4 que se uma variável aleatória X tem por distribuição Gama de parâmetros α e ν , sua função característica é da forma

$$\varphi_X(t) = \left(1 - \frac{it}{\alpha}\right)^{-\nu}.$$

Observemos que com o produto de n funções características $(1 - it/\alpha)^{-\nu/n}$ reproduzimos φ_X , da forma

$$\varphi_X(t) = \prod_{k=1}^n \left(1 - \frac{it}{\alpha}\right)^{-\frac{\nu}{n}} = \left[\left(1 - \frac{it}{\alpha}\right)^{-\frac{\nu}{n}}\right]^n = \left(1 - \frac{it}{\alpha}\right)^{-\nu}.$$

Logo, escolhendo cada $\xi_n \sim \text{Gama}(\alpha, \nu/n)$, para qualquer n natural, temos que X é infinitamente divisível.

Exemplo 7.19

A variável aleatória X , com distribuição Poisson, é infinitamente divisível. Suponhamos que os possíveis valores de X sejam da forma $\lambda + kh$, $k = 0, 1, 2, \dots$ e que

$$P(X = \mu + kh) = \frac{\lambda^k e^{-\lambda}}{k!}.$$

Então, a função característica é

$$\varphi_X(t) = e^{i\mu t + \lambda(e^{ith} - 1)}.$$

Daí concluímos que, para todo $n > 0$,

$$\sqrt[n]{\varphi_X(t)} = e^{i\frac{\mu}{n}t + \frac{\lambda}{n}(e^{ith} - 1)},$$

é a função característica de uma variável aleatória, também com distribuição Poisson.

Teorema 7.23

A função característica de uma variável aleatória infinitamente divisível nunca é zero.

Demonstração: Seja X uma variável aleatória infinitamente divisível e $\varphi_X(t)$ sua função característica. Logo satisfaz a relação (7.25) e vamos escrever

$$\varphi_X(t) = [\varphi_{\xi_n}(t)]^n,$$

e dado que $\varphi_{\xi_n}(t) \geq 0$, segue

$$\varphi_{\xi_n}(t) = [\varphi(t)]^{1/n}, \quad \forall t. \quad (7.26)$$

Mas $0 \leq \varphi(t) \leq 1$, conseqüentemente

$$\lim_{n \rightarrow \infty} [\varphi(t)]^{1/n} = \begin{cases} 0, & \text{quando } \varphi(t) = 0 \\ 1, & \text{quando } \varphi(t) \neq 0 \end{cases}.$$

Então $\lim_{n \rightarrow \infty} [\varphi_n(t)]^{1/n}$ existe para todo t e a função limite, digamos $h(t)$, pode assumir pelo menos dois possíveis valores 0 ou 1. Além disso, dado que φ é contínua em $t = 0$, com $\varphi(t) = 1$, existe $t_0 > 0$ de tal modo que $\varphi(t) \neq 0$ para $|t| \leq t_0$. Segue que $h(t) = 1$, para $|t| \leq t_0$. Portanto, a sequência de funções características φ_n converge para a função h , que acaba de ser mostrado ser contínua na origem. Pelo teorema da convergência, h deve ser uma função característica e portanto contínua em todos os lugares. Portanto, h é identicamente igual a 1 e, logo, pela observação depois de termos

$$|f(t)|^2 = \varphi(t) \neq 0, \quad \forall t.$$



Exemplo 7.20

A variável aleatória X , com distribuição Cauchy é infinitamente divisível. A função de distribuição, para $\sigma > 0$ assume a forma

$$F(x) = \frac{1}{\pi} \left(\frac{\pi}{2} + \arctan \left(\frac{x - \mu}{\sigma} \right) \right).$$

Obtemos, então, que a função característica nesta situação é da forma

$$\varphi_X(t) = e^{i\mu t - \sigma|t|},$$

a qual é, evidentemente, sempre maior do que zero.

Um detalhe interessante é que o Teorema 7.23 não diz que toda variável aleatória cuja função característica nunca é zero seja infinitamente divisível. Caso a função característica seja

$$\frac{2 + \cos(t)}{3}, \quad (7.27)$$

percebemos que ela nunca é zero mas, a relação em (7.25) falha. Vejamos a resposta a isso no seguinte exemplo.

Exemplo 7.21 (Generalização do Exemplo 7.11)

Consideremos X uma variável aleatória discreta com valores $-a, 0, a$, sendo $a > 0$ uma constante. A função de probabilidades de X é $P(X = -a) = P(X = a) = p$ e $P(X = 0) = 1 - 2p$. Encontremos a função característica φ_X .

$$\begin{aligned} E(e^{itX}) &= e^{-ita}P(X = -a) + P(X = 0) + e^{ita}P(X = a) \\ &= e^{-ita}p + (1 - 2p) + e^{ita}p = (e^{-ita} + e^{ita})p + (1 - 2p). \end{aligned}$$

Sabemos que $e^{-ita} = \cos(-ta) + i \sin(-ta)$ e $e^{ita} = \cos(ta) + i \sin(ta)$ assim como que a função cosseno é par, logo $\cos(-ta) = \cos(ta)$. Então

$$\varphi_X(t) = 2p \cos(ta) + (1 - 2p).$$

Segundo o resultado mostrado no Exemplo 7.21 vemos que a função característica em (7.27) corresponde a uma variável aleatória discreta com valores $-1, 0, 1$ e função de probabilidade $P(X = -1) = P(X = 1) = 1/6$ e $P(X = 0) = 2/3$, a qual não é infinitamente divisível.

Uma outra situação interessante acontece quando $X \sim U(-1, 1)$. Pode-se demonstrar que a função característica é

$$\varphi_X(t) = \frac{\sin(t)}{t},$$

a qual é zero múltiplas vezes, como por observado na Figura 7.4. Por outro lado $X \sim U(-1, 1)$ não é infinitamente divisível, pois como acabamos de perceber sua função característica é zero muitas vezes embora ela tenha infinitamente muitos divisores, isto porque que pode ser escrita

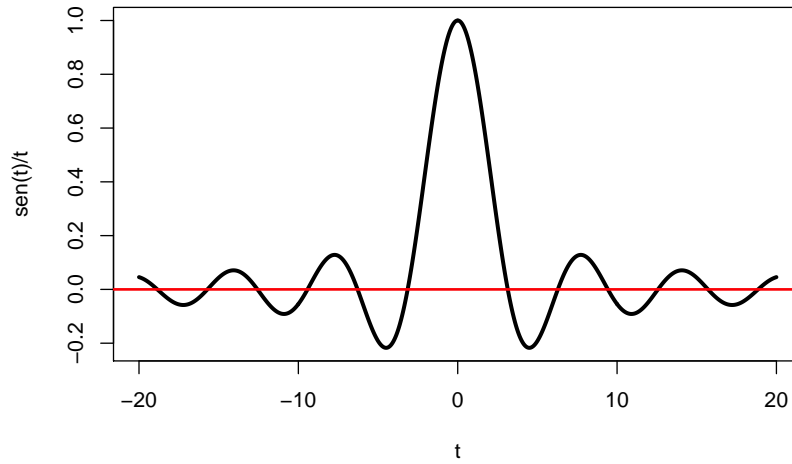


Figura 7.4: Gráfico da função característica $\text{sen}(t)/t$.

como $\frac{\text{sen}(t)}{t} = \prod_{n=1}^{\infty} \cos\left(\frac{t}{2^n}\right)$ e, inclusive, pode ser escrita como

$$\frac{\sin(t)}{t} = \left(\prod_{k=1}^{\infty} \cos\left(\frac{t}{2^{2k-1}}\right) \right) \left(\prod_{k=1}^{\infty} \cos\left(\frac{t}{2^{2k}}\right) \right).$$

Agora descrevemos um procedimento para gerar distribuições infinitamente divisíveis de distribuições arbitrárias.

Teorema 7.24

Seja φ uma função característica e $\lambda > 0$ uma constante real. Então, $e^{-\lambda(1-\varphi(t))}$ é uma função característica infinitamente divisível.

Demonstração: Observemos que, para cada t , $\lim_{n \rightarrow \infty} \left(1 - \frac{\lambda(1-\varphi(t))}{n}\right)^n = e^{-\lambda(1-\varphi(t))}$. Observemos que $1 - \frac{\lambda(1-\varphi(t))}{n} = \left(1 - \frac{\lambda}{n}\right) + \frac{\lambda}{n}\varphi(t)$. Consequentemente, como $\varphi(t) = \int e^{itx} f(x) dx$, no caso absolutamente contínuo, então

$$1 - \frac{\lambda(1-\varphi(t))}{n} = \int e^{itx} g(x) dx,$$

onde $g(x) = \left(1 - \frac{\lambda}{n}\right)\delta_0 + \frac{\lambda}{n}f(x)$. Dado que g é uma medida de probabilidade segue que $1 - \frac{\lambda(1-\varphi(t))}{n}$ é uma função característica e, portanto, $\left(1 - \frac{\lambda(1-\varphi(t))}{n}\right)^n$ também é uma função característica; pelo Teorema de Continuidade ■

Os seguintes resultados são as propriedades mais profundas das funções características infinitamente divisíveis. Para uma prova completa do teorema, vários casos especiais e desenvolvimentos adicionais, consulte o livro de Gnedenko & Kolmogorov (1962).

Teorema 7.25

Seja $\{kf\}$, $k \geq 1$ uma sequência de funções características infinitamente divisíveis convergindo em todos os pontos para a função característica f . Então f é infinitamente divisível.

Demonstração: A dificuldade é provar primeiro que f nunca desvanece. Considere, $g = |f|^2$, logo $kg = |kf|^2$. Para cada $n > 1$, considere $x^{1/n}$ a enésima raiz real positiva de um x , positivo real. Então temos, pela hipótese da convergência e da continuidade de $x^{1/n}$ em função de x ,

$$\lim_{k \rightarrow \infty} (kg(t))^{1/n} = (g(t))^{1/n}.$$

Pelo Teorema 7.22 conclui-se que $(g(t))^{1/n}$ é uma função característica. Na expressão acima, o lado à direita é contínua em qualquer lugar. Segue-se do Teorema da Convergência para funções características que $(g(\cdot))^{1/n}$ é uma função características. Como g é sua enésimo potência, e isso é verdade para cada $n > 1$, provamos que g é infinitamente divisível e, portanto, nunca desvanece. Portanto, f nunca desvanece e possui um logaritmo definido em todos os pontos. Como a convergência de kf a f é necessariamente uniforme em cada intervalo finito, segue-se que $\log(kf)$ convirja a $\log(f)$ em todos os pontos e, consequentemente,

$$\lim_{k \rightarrow \infty} \left(\exp(\log(kf(t))/n) \right) = \exp(\log(f(t))/n), \quad (7.28)$$

para todo t . A função à esquerda em (7.28) é uma função característica, como consequência do Teorema 7.22, a função à direita acima é contínua pela definição de $\log(f)$. Portanto, segue-se como antes que $e^{\log(f(t))/n}$ é uma função característica e f é infinitamente divisível. ■

Usando o teorema anterior, podemos construir uma ampla classe de funções características que sejam infinitamente divisíveis. Para cada a e u reais, a função

$$e^{a(e^{iut}-1)}$$

é uma função característica infinitamente divisível, obtida da função característica da distribuição Poisson. Observe que provamos acima que toda função infinitamente divisível está na classe das funções características limitadas, embora não tenhamos estabelecido a representação canônica.

Teorema 7.26

Se $|\varphi| \geq 0$ então φ é integrável se, e somente se, a distribuição F correspondente tiver uma densidade limitada.

**Demonstração :**

Estamos agora em posição de declarar o teorema fundamental de funções características infinitamente divisíveis, devido a Paul Levy e Aleksandr Khinchin¹⁵.

A representação canônica de Lévy - Khinchin foi proposta por A.Ya. Khinchin (1937) e é equivalente a uma fórmula proposta um pouco antes por P. Lévy (1934) e denominada representação canônica de Lévy. Para cada distribuição infinitamente divisível corresponde um conjunto único de características φ_X e f na representação canônica de Lévy ? Khinchin e, inversamente, para qualquer φ_X e f como em (7.29), a representação canônica de Lévy ? Khinchin determina a função característica de uma distribuição infinitamente divisível.

Teorema 7.27

A função característica φ_X da variável aleatória infinitamente divisíveis X tem a seguinte representação canônica

$$\varphi_X(t) = \exp \left\{ ait + \int_{-\infty}^{\infty} \left(e^{itu} - 1 - \frac{itu}{1+u^2} \right) \frac{1+u^2}{u^2} f(u) du \right\}, \quad (7.29)$$

onde a é uma constante real, f é a função de densidade ou de probabilidade de X e a integral em $u = 0$ é definida pela equação

$$\left[\left(e^{itu} - 1 - \frac{itu}{1+u^2} \right) \frac{1+u^2}{u^2} \right]_{u=0} = -\frac{t^2}{2}.$$

A representação da função característica em (7.29) é única.

Demonstração : content...

¹⁵Aleksandr Yakovlevich Khinchin (1894-1959). Foi um matemático russo e um dos cientistas soviéticos mais importantes na escola da teoria das probabilidades.

7.2 Desigualdades probabilísticas

Desigualdades são úteis para delimitar quantidades que poderiam ser difíceis de calcular. Diversas desigualdades que serão apresentadas nesta seção permitirão demonstrar resultados importantes, por isso o cuidado de dedicar um espaço para desigualdades que envolvam o cálculo de probabilidades. Por exemplo, desigualdades serão utilizadas para provarmos resultados de convergência a serem vistos na Seção 7.3 e no Capítulo 8.

Teorema 7.28 (*Desigualdade de Chebyshev*)

Seja X uma variável aleatória não negativa com esperança $E(X)$ finita. Então, para todo $\epsilon > 0$,

$$P(X > \epsilon E(X)) \leq \frac{1}{\epsilon}. \quad (7.30)$$

Demonstração: Observemos que

$$\begin{aligned} E(X) &= \int_0^{\infty} (1 - F(x)) \, dx \\ &= \int_0^{\epsilon E(X)} (1 - F(x)) \, dx + \int_{\epsilon E(X)}^{\infty} (1 - F(x)) \, dx \\ &\geq \int_{\epsilon E(X)}^{\infty} (1 - F(x)) \, dx = \int_{\epsilon E(X)}^{\infty} P(X > x) \, dx \\ &\geq \epsilon E(X) \int_1^{\infty} P(X > \epsilon E(X)) \, dx \geq \epsilon E(X) P(X > \epsilon E(X)), \end{aligned}$$

então, dividindo ambos os lados por $\epsilon E(X)$, obtemos a expressão em (7.30). ■

A desigualdade recebeu o nome do matemático russo Pafnuty Chebyshev¹⁶, que primeiro declarou a desigualdade sem provas em 1874. Dez anos depois, a desigualdade foi comprovada por Markov em sua dissertação de doutorado. Devido a variações de como representar o alfabeto russo em inglês e também em português, Chebyshev também é escrito como Tchebysheff.

Teorema 7.29 (*Desigualdade de Markov*)

Seja X uma variável aleatória não negativa com r -ésimo momento $E(X^r)$ finito, para algum $r > 0$. Então, para todo $\epsilon > 0$, temos que

$$P(X > \epsilon) \leq \frac{E(X^r)}{\epsilon^r}. \quad (7.31)$$

¹⁶Pafnuty Lvovich Chebyshev, (1821-1894). Matemático Russo, fundador da escola de matemática de São Petersburgo às vezes chamada de escola de Chebyshev. Lembrado principalmente por seu trabalho na teoria dos números primos e na aproximação de funções.

Demonstração: Fazemos $U = X^r$, de maneira que $E(U) < \infty$. Então, observe que o evento $\{\omega \in \Omega : X(\omega) > \epsilon\}$ é equivalente a $\{\omega \in \Omega : U(\omega) > \epsilon^r\}$. Segue da desigualdade de Chebyshev que

$$P(X > \epsilon) = P(U > \epsilon^r) \leq \frac{E(X^r)}{\epsilon^r},$$

onde $\epsilon = \epsilon^r / E(X^r)$ em (7.30). ■

A desigualdade de Markov¹⁷ e a desigualdade de Chebyshev são extremamente importantes em uma ampla variedade de provas teóricas, especialmente nos teoremas de limites. Na prática, essas desigualdades têm sido algumas vezes usadas para encontrar limites nas probabilidades que surgem na prática, esperando que os limites se aproximem das probabilidades de interesse. Isso era especialmente verdade quando as probabilidades envolviam distribuições para as quais somente tínhamos disponíveis resultados imprecisos ou quando as distribuições eram fora do padrão ou complicadas. Geralmente, este não é um método muito bom de aproximação para fins práticos.

Atualmente, os pacotes de software permitem cálculos exatos para uma ampla variedade de distribuições usadas com frequência e permitem a aproximação por simulação em casos complicados e não padronizados.

Teorema 7.30 (*Desigualdade generalizada de Chebyshev*)

Seja X uma variável aleatória integrável, isto é, $E(X) < \infty$. Então,

$$P(|X - E(X)| \geq \epsilon) \leq \frac{\text{Var}(X)}{\epsilon^2}, \quad \forall \epsilon > 0.$$

Demonstração: Utilizando a desigualdade de Markov, definindo $E(X) = \mu$, temos que

$$P(|X - \mu| \geq \epsilon) = P(|X - \mu|^2 \geq \epsilon^2) \geq \frac{E(X - \mu)^2}{\epsilon^2} = \frac{\text{Var}(X)}{\epsilon^2}. \quad \blacksquare$$

Exemplo 7.22

Sejam X_1, \dots, X_n variáveis aleatórias independentes identicamente distribuídas, tais que $X_1 \sim \text{Bernoulli}(p)$. Então $\bar{X}_n = \sum_{k=1}^n X_k/n$, $\text{Var}(\bar{X}_n) = \text{Var}(X_1)/n = p(1-p)/n$ e

$$P(|\bar{X}_n - p| \geq \epsilon) \leq \frac{\text{Var}(\bar{X}_n)}{\epsilon^2} = \frac{p(1-p)}{n\epsilon^2} \leq \frac{1}{4n\epsilon^2},$$

desde que $p(1-p) \leq \frac{1}{4}$, para todo p .

¹⁷Andrey Andrey Andreyevich Markov, (1856-1922), foi um matemático russo que ajudou a desenvolver a teoria dos processos estocásticos, especialmente as chamadas Cadeias Markov. Seu trabalho tem sido desenvolvido e amplamente aplicado nas ciências biológicas e sociais.

Se soubermos mais sobre a distribuição com a qual estamos trabalhando, podemos garantir que mais dados é um certo número de desvios padrão longe da média. Por exemplo, se sabemos que temos uma distribuição normal, então 95% dos dados são dois desvios padrão da média. A desigualdade de Chebyshev diz que nesta situação sabemos que pelo menos 75% dos dados são dois desvios padrão da média. Como podemos ver neste caso, pode ser muito mais do que estes 75%.

O valor da desigualdade é que nos dá um cenário do pior caso em que as únicas coisas que sabemos sobre a distribuição de probabilidade são a média e o desvio padrão. Quando não sabemos mais nada sobre os nossos dados, a desigualdade de Chebyshev fornece alguma informação adicional sobre a distribuição da variável aleatória.

Exemplo 7.23

Suponha que T , o tempo necessário para concluir um processo seja a soma de três distribuições independentes conhecidas:

- (1) $W \sim \text{Uniforme}(0, 12)$, com $E(W) = 6$ e $\text{Var}(W) = 12$, ver Exemplo 4.39,
- (2) $X \sim \text{Exponencial}(1/5)$, $E(X) = 5$ e $\text{Var}(X) = 25$ e
- (3) $Y \sim \text{Normal}(10, 4)$, com $E(Y) = 10$ e $\text{Var}(Y) = 4$.

Então $T = W + X + Y$ tem uma distribuição para a qual a função de densidade e a função de distribuição são muito difíceis de encontrar, mas sabemos que $E(T) = 21$ e $\text{Var}(T) = 41$. Buscamos $P(T > 35)$, a probabilidade de que o processo geral leve mais de 35 unidades de tempo para conclusão. Mencionamos vários métodos possíveis para limitar ou aproximar essa probabilidade.

Desigualdade de Markov *Por causa do componente normal, há uma pequena probabilidade de que T seja negativo e, portanto, estritamente falando, a desigualdade de Markov não se aplica a T . Mas a parte positiva T^+ de T é essencialmente a mesma que T , e ignoramos essa dificuldade. Então a desigualdade de Markov diz*

$$P(T > 35) \leq 21/35 = 0.6.$$

Desigualdade de Chebyshev *Pela desigualdade de Chebyshev, sabemos que $P(T > \epsilon E(T)) \leq 1/\epsilon$. Escolhendo $\epsilon = 1.666667$, sabemos que $21 \times 1.666667 = 35.00001$, temos que*

$$P(T > 35) \leq 1/\epsilon = 0.5999999.$$

Percebemos que os resultados são consistentes.

A desigualdade de Hoeffding¹⁸ é semelhante em espírito à desigualdade de Markov, mas é uma desigualdade mais acentuada. Começamos com os seguintes resultados importantes.

¹⁸Wassily Hoeffding (1914-1991). Foi estatístico e probabilista finlandês. Hoeffding foi um dos fundadores da estatística não paramétrica, na qual Hoeffding contribuiu com a idéia e com os resultados básicos das U -estatísticas. Na teoria da probabilidade, a desigualdade de Hoeffding fornece um limite superior à probabilidade da soma das variáveis aleatórias se desviar do seu valor esperado.

Lema 7.31

Suponhamos que $a \leq X \leq b$. Então

$$\mathbb{E}(e^{tX}) \leq e^{t\mu} e^{t^2(b-a)^2/8},$$

sendo que $\mathbb{E}(X) = \mu$.

Antes de iniciarmos a prova, lembre-se de que uma função g é convexa se para cada x, y e cada $\alpha \in [0, 1]$,

$$g(\alpha x + (1 - \alpha)y) \leq \alpha g(x) + (1 - \alpha)g(y).$$

Demonstração: Vamos assumir que $\mu = 0$. Dado que $a \leq X \leq b$, podemos escrever X como uma combinação convexa de a e b , seja esta, $X = \alpha b + (1 - \alpha)a$ onde $\alpha = (X - a)/(b - a)$. Pela convexidade da função $g(y) = e^{ty}$, temos

$$e^{tX} \leq \alpha e^{tb} + (1 - \alpha)e^{ta} = \frac{X - a}{b - a}e^{tb} + \frac{b - X}{b - a}e^{ta}.$$

Tomando esperança de ambos os lados e usando o fato de que $\mathbb{E}(X) = 0$, obtemos

$$\mathbb{E}(e^{tX}) \leq -\frac{a}{b - a}e^{tb} + \frac{b}{b - a}e^{ta} = e^{g(u)},$$

onde $u = t(b - a)$, $g(u) = -\gamma u + \log(1 - \gamma + \gamma e^u)$ e $\gamma = -a/(b - a)$. Observe que $g(0) = g'(0) = 0$, também $g''(u) \leq 1/4$ para todo $u > 0$. Pelo Teorema de Taylor¹⁹, existe um $\xi \in (0, u)$ tal que

$$g(u) = g(0) + ug'(0) + \frac{u^2}{2}g''(\xi) = \frac{u^2}{2}g''(\xi) \leq \frac{u^2}{8} = \frac{t^2(b - a)^2}{8}.$$

Consequentemente $\mathbb{E}(e^{tX}) \leq e^{g(u)} \leq e^{t^2(b-a)^2/8}$. ■

Lema 7.32

Seja X uma variável aleatória. Então,

$$P(X > \epsilon) \leq \inf_{t \geq 0} \{e^{-t\epsilon} \mathbb{E}(e^{tX})\}.$$

¹⁹Teorema de Taylor: Seja $k \geq 1$ um inteiro e $f : \mathbb{R} \rightarrow \mathbb{R}$ uma função real com derivada até ordem k no ponto $a \in \mathbb{R}$. Então, existe uma função $h_k : \mathbb{R} \rightarrow \mathbb{R}$ de forma que

$$f(x) = f(a) + f'(a)(x - a) + \frac{f''(a)}{2!}(x - a)^2 + \cdots + \frac{f^{(k)}(a)}{k!}(x - a)^k + h_k(x)(x - a)^k,$$

e $\lim_{x \rightarrow \infty} h_k(x) = 0$ (Spivak, 1994).

Demonstração: Para qualquer $t > 0$,

$$P(X > \epsilon) = P(e^X > e^\epsilon) = P(e^{tX} > e^{t\epsilon}) \leq e^{-t\epsilon} E(e^{tX}),$$

isto pela desigualdade de Markov (Teorema 7.29). Como isso é verdadeiro para cada $t \geq 0$, o resultado segue. ■

Agora estamos em condições de apresentar e logo em seguida demonstrar a chamada desigualdade de Hoeffding.

Teorema 7.33 (*Desigualdade de Hoeffding*)

Seja X uma variável aleatória satisfazendo $a \leq X \leq b$ e $E(X) = \mu$, finita. Então, para qualquer $\epsilon > 0$, temos

$$P(|X| \geq \epsilon) \leq 2e^{-2\epsilon^2/(b-a)^2}.$$

Demonstração: Sem perda de generalidade vamos assumir que $\mu = 0$. Primeiro vemos que

$$\begin{aligned} P(|X| \geq \epsilon) &= P(X \geq \epsilon) + P(X \leq -\epsilon) \\ &= P(X \geq \epsilon) + P(-X \geq \epsilon). \end{aligned}$$

Sabemos que $P(X \geq \epsilon) = P(e^X \geq e^\epsilon) = P(e^{tX} \geq e^{t\epsilon}) \leq e^{-t\epsilon} E(e^{tX})$, isto último pela desigualdade de Markov e $E(e^{tX}) \leq e^{t^2(b-a)^2/8}$, pelo Lema 7.31. De maneira que

$$P(X \geq \epsilon) \leq e^{-t\epsilon} e^{t^2(b-a)^2/8}.$$

O mínimo na expressão acima é obtido escolhendo $t = 4\epsilon/(b-a)^2$, do qual

$$P(X \geq \epsilon) \leq e^{-2\epsilon^2/(b-a)^2}.$$

Aplicando argumento similar a $P(-X \geq \epsilon)$ chegamos ao resultado. ■

A desigualdade de Hoeffding cumpre-se caso as variáveis aleatórias X_1, \dots, X_n sejam independentes identicamente distribuídas, cada uma satisfazendo $a \leq X_k \leq b$ e $E(X_k) = \mu$, $k = 1, \dots, n$. Nesta situação geral, a desigualdade de Hoeffding assume a forma

$$P(|\bar{X}_n - \mu| \geq \epsilon) \leq 2e^{-2n\epsilon^2/(b-a)^2}. \quad (7.32)$$

Exemplo 7.24

Sejam X_1, \dots, X_n variáveis aleatórias independentes e identicamente distribuídas *Bernoulli*(p). Da desigualdade de Hoeffding temos que

$$P(|\bar{X}_n - p| > \epsilon) \leq 2e^{-2n\epsilon^2}.$$

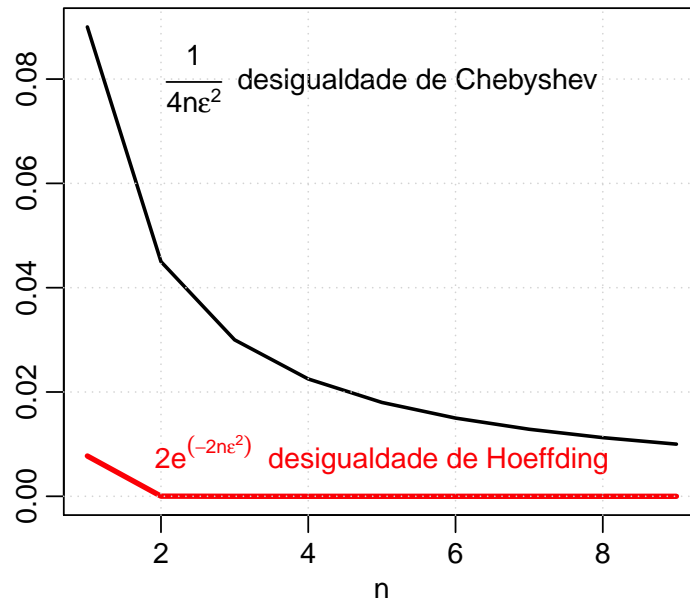


Figura 7.5: Limites superior das desigualdades de Chebyshev e Hoeffding para a probabilidade considerada no Exemplo 7.22 e no Exemplo 7.24.

Mostramos na Figura 7.5 as linhas, segundo o valor de n , dos limites superior obtidos nas desigualdades de Chebyshev e de Hoeffding. Fica claro que a desigualdade de Hoeffding é muito mais acentuada do que a de Chebyshev. Devemos esclarecer que estes limites foram obtidos para a mesma probabilidade sob mesmas condições.

Apresentamos a seguir dois resultados que serão de grande importância para demonstrar resultados de convergência no Capítulo 8.

Teorema 7.34 (*Desigualdade de Kolmogorov*)

Seja $\{X_n\}_{n \geq 1}$ uma sequência de variáveis aleatórias independentes tais que $E(X_n) = 0$ e $\text{Var}(X_n) < \infty$. Então, para todo $\lambda > 0$,

$$P\left(\max_{1 \leq k \leq n} |S_k| \geq \lambda\right) \leq \frac{1}{\lambda^2} \text{Var}(S_n),$$

sendo que $S_k = \sum_{s=1}^k X_s$.

Demonstração: Seja

$$A = \left\{ \omega \in \Omega : \max_{1 \leq k \leq n} S_k^2(\omega) \geq \lambda^2 \right\},$$

vamos decompor este conjunto da seguinte forma

$$\begin{aligned} A_1 &= \{\omega \in \Omega : S_1^2(\omega) \geq \lambda^2\}, \\ A_2 &= \{\omega \in \Omega : S_1^2(\omega) < \lambda^2, S_2^2(\omega) \geq \lambda^2\}, \\ &\vdots \\ A_k &= \{\omega \in \Omega : S_1^2(\omega) < \lambda^2, \dots, S_{k-1}^2(\omega) < \lambda^2, S_k^2(\omega) \geq \lambda^2\}, \quad \text{para } 2 \leq k \leq n. \end{aligned}$$

Então, os conjuntos A_k são disjuntos dois a dois e $A = \bigcup_{k=1}^n A_k$, logo $I_A = \sum_{k=1}^n I_{A_k}$, então

$$S_n^2 = \left(\sum_{k=1}^n X_k \right)^2 \geq S_n^2 I_A = \sum_{k=1}^n S_n^2 I_{A_k}, \text{ o qual implica em}$$

$$E(S_n^2) \geq \sum_{k=1}^n E(S_n^2) I_{A_k}. \quad (7.33)$$

Observemos que

$$\begin{aligned} S_n^2 &= (S_n - S_k + S_k)^2 = (S_n - S_k)^2 + S_k^2 + 2(S_n - S_k)S_k \\ &\geq S_k^2 + 2(S_n - S_k)S_k. \end{aligned}$$

Portanto, $E(S_n^2 I_{A_k}) \geq E(S_k^2 I_{A_k}) + 2E((S_n - S_k)S_k I_{A_k})$. Agora $S_n - S_k = X_{k+1} + \dots + X_n$ então

$$E((S_n - S_k)S_k I_{A_k}) = E((X_{k+1} + \dots + X_n)S_k I_{A_k}) = \sum_{s=k+1}^n E(X_s S_k I_{A_k}) = 0,$$

isto porque as variáveis X_n , $n = 1, \dots$, são independentes. Logo,

$$E(S_n^2 I_{A_k}) \geq E(S_k^2 I_{A_k}) \geq E(\lambda^2 I_{A_k}) = \lambda^2 P(A_k).$$

Pelo resultado em (7.33), vemos que $E(S_n^2) \geq \sum_{k=1}^n \lambda^2 P(A_k) = \lambda^2 P(A)$, logo

$$P(A) \leq \frac{1}{\lambda^2} E(S_n^2) = \frac{1}{\lambda^2} \text{Var}(S_n). \quad \blacksquare$$

Na demonstração do seguinte resultado vamos utilizar a seguinte desigualdade numérica²⁰:

$$\sum_{n=k}^{\infty} \frac{1}{n^2} \leq \frac{2}{k}, \quad \text{para } k = 1, 2, \dots \quad (7.34)$$

²⁰Demonstração da desigualdade em (7.34): para $n = 1, 2, \dots$, temos que

$$\frac{1}{n^2} \leq \frac{1}{n(n-1)} = \frac{1}{n-1} - \frac{1}{n},$$

então

$$\sum_{n=k}^{\infty} \frac{1}{n^2} = \frac{1}{k^2} + \sum_{n=k+1}^{\infty} \frac{1}{n^2} \leq \frac{1}{k^2} + \sum_{n=k+1}^{\infty} \left(\frac{1}{n-1} - \frac{1}{n} \right) = \frac{1}{k^2} + \frac{1}{k} \leq \frac{2}{k}.$$

Teorema 7.35

Seja X uma variável aleatória integrável com função de distribuição F . Então,

$$\sum_{n=1}^{\infty} \left(\frac{1}{n^2} \int_{-n}^n x^2 f(x) dx \right) < \infty,$$

caso F fosse absolutamente contínua e f a função de densidade correspondente.

Demonstração: Vamos escrever $\int_{-n}^n x^2 f(x) dx = \sum_{k=-n+1}^n \int_{k-1}^k x^2 f(x) dx$, resulta que

$$\begin{aligned} \sum_{n=1}^{\infty} \left(\frac{1}{n^2} \int_{-n}^n x^2 f(x) dx \right) &= \sum_{n=1}^{\infty} \sum_{k=-n+1}^n \left(\frac{1}{n^2} \int_{k-1}^k x^2 f(x) dx \right) \\ &= \sum_{k=1}^{\infty} \sum_{n=k}^{\infty} \left(\frac{1}{n^2} \int_{k-1}^k x^2 f(x) dx \right) \\ &\quad + \sum_{k=-\infty}^0 \sum_{n=|k|+1}^{\infty} \left(\frac{1}{n^2} \int_{k-1}^k x^2 f(x) dx \right), \end{aligned}$$

agora, pelo resultado em (7.34),

$$\sum_{n=1}^{\infty} \left(\frac{1}{n^2} \int_{-n}^n x^2 f(x) dx \right) \leq 2 \sum_{k=1}^{\infty} \int_{k-1}^k \frac{x^2}{k} f(x) dx + 2 \sum_{k=-\infty}^0 \int_{k-1}^k \frac{x^2}{|k|+1} f(x) dx.$$

Observamos que

$$\begin{aligned} \frac{x^2}{k} &\leq x, \quad \text{caso } x \in (k-1, k], k \geq 1 \\ \frac{x^2}{|k|+1} &\leq |x|, \quad \text{caso } x \in (k-1, k], k \leq 0, \end{aligned}$$

temos,

$$\begin{aligned} \sum_{n=1}^{\infty} \left(\frac{1}{n^2} \int_{-n}^n x^2 f(x) dx \right) &\leq 2 \sum_{k=1}^{\infty} \int_{k-1}^k x^2 f(x) dx + 2 \sum_{k=-\infty}^0 \int_{k-1}^k |x| f(x) dx \\ &= 2 \sum_{k=-\infty}^0 \int_{k-1}^k |x| f(x) dx = 2 \int_{-\infty}^{\infty} |x| f(x) dx = 2E(|X|) < \infty. \quad \blacksquare \end{aligned}$$

Caso no Teorema 7.35 a função de distribuição seja discreta então

$$\int_{-n}^n x^2 f(x) dx = \sum_{x=-n}^n x^2 P(X=x)$$

e a demonstração segue de maneira similar.

7.3 Modos de convergência estocástica

Em probabilidade e estatística é muitas vezes necessário examinar a distribuição de uma variável aleatória, que é ela própria uma função de várias variáveis aleatórias, escrita como $Y = g(X_1, \dots, X_n)$. Como exemplos simples podemos pensar na soma amostral $S_n = \sum_{k=1}^n X_k$ e na soma dos quadrados $S_n^2 = \sum_{k=1}^n X_k^2$.

Infelizmente, encontrar a distribuição exata é frequentemente muito difícil ou muito demorada mesmo se a distribuição conjunta das variáveis aleatórias seja conhecida exatamente. Em outros casos, podemos ter apenas informações parciais sobre a distribuição conjunta de X_1, \dots, X_n caso em que é impossível determinar a distribuição de Y . No entanto, quando n é grande, pode ser possível obter aproximações à distribuição de Y , mesmo quando apenas uma informação parcial sobre X_1, \dots, X_n esteja disponível. Em muitos casos, estas aproximações podem ser extremamente precisas.

A abordagem padrão para aproximar uma função de distribuição é considerar a função de distribuição como parte de uma sequência infinita de funções de distribuição; então tentamos encontrar uma distribuição limite para a sequência e usamos essa distribuição limite para aproximar a distribuição da variável aleatória em questão.

Esta abordagem é comum em matemática. Por exemplo, se n é grande em comparação com x , pode-se aproximar $(1 + x/n)^n$ por $\exp(x)$ desde que

$$\lim_{n \rightarrow \infty} \left(1 + \frac{x}{n}\right)^n = e^x.$$

No entanto, esta aproximação pode ser muito pobre se x/n não estiver perto de 0. Um exemplo mais interessante é a aproximação de Stirling²¹, que é usada para aproximar $n!$ para grandes valores de n :

$$n! \approx \sqrt{2\pi} e^{-n} n^{n+1/2}.$$

Em certo sentido, a aproximação de Stirling mostra que as aproximações assintóticas podem ser úteis num contexto mais geral. Na prática estatística, aproximações assintóticas, tipicamente justificadas para amostras de grandes dimensões, são muito comumente usadas mesmo em situações onde o tamanho da sequência de variáveis aleatórias é pequeno. É claro, nem sempre, a utilização de tais aproximações é justificada mas, no entanto, existe um conjunto suficientemente rico de exemplos onde se justifica fazer o estudo da convergência de variáveis aleatórias que valham a pena.

Uma questão importante é: qual o espaço de probabilidade da variável limite? o espaço de probabilidade de cada uma das variáveis X_1, X_2, \dots é (Ω, \mathcal{F}, P) mas, se a sequência $\{X_n\}$ convergir a uma variável aleatória X , esta não necessariamente tem que estar definida no mesmo espaço de probabilidade.

Observemos que se $\Omega^n = \underbrace{(\Omega, \dots, \Omega)}_{n \text{ vezes}}$, for o espaço amostral do vetor (X_1, \dots, X_n) então o espaço amostral limite é Ω^∞ , a σ -álgebra limite satisfaz que

$$\mathcal{F}^n = \underbrace{(\mathcal{F}, \dots, \mathcal{F})}_{n \text{ vezes}}$$

e a função de probabilidade limite é $P^{n+1} = \underbrace{(P, \dots, P)}_{n \text{ vezes}}, \infty$.

²¹James Stirling (1692-1770). Foi um matemático escocês.

7.3.1 Convergência em probabilidade

Suponha-se que X_1, X_2, \dots é uma sequência de variáveis aleatórias. A grosso modo, diz-se que esta sequência converge para um determinado número c , se a distribuição de probabilidade de X_n torna-se cada vez mais concentrada em torno de c , quando $n \rightarrow \infty$. Para ser mais precisos, damos a seguinte definição.

Definição 7.4

Seja $\{X_n\}$ uma sequência de variáveis aleatórias definidas no mesmo espaço de probabilidade. Dizemos que a sequência $\{X_n\}$ converge em probabilidade à variável aleatória X se, para todo $\epsilon > 0$,

$$\lim_{n \rightarrow \infty} P(|X_n - X| > \epsilon) = 0. \quad (7.35)$$

Nesta situação escrevemos $X_n \xrightarrow{P} X$.

Ressaltamos que esta definição não diz nada sobre a convergência das variáveis aleatórias X_n à variável aleatória X , no sentido em que é entendido em cálculo. Então se $X_n \xrightarrow{P} X$ isto não implica que, X_n e X serão aproximadamente iguais quando n seja suficientemente grande. Esta definição somente nos disse acerca da convergência da sequência de probabilidades $P(|X_n - X| > \epsilon)$ a zero. Uma outra observação que devemos fazer é que esta definição também nada informa acerca do espaço de probabilidade no qual a variável limite está definida.

Teorema 7.36

Seja $\{X_n\}$ uma sequência de variáveis aleatórias com $\mu_n = E(X_n)$ e $\sigma_n^2 = \text{Var}(X_n)$ tais que $\lim_{n \rightarrow \infty} \mu_n = c$, onde c é uma constante e $\lim_{n \rightarrow \infty} \sigma_n^2 = 0$. Então,

$$X_n \xrightarrow{P} c.$$

Demonstração: Pela desigualdade de Markov em (7.31), temos que

$$\lim_{n \rightarrow \infty} P(|X_n - c| \geq \epsilon) \leq \lim_{n \rightarrow \infty} \frac{E[(X_n - c)^2]}{\epsilon^2} = 0,$$

devido a que $E[(X_n - c)^2] = \sigma_n^2 + (\mu_n - c)^2$. ■

Exemplo 7.25

Suponhamos que X_1, \dots, X_n sejam variáveis aleatórias independentes identicamente distribuídas $U(0, 1)$. Definamos

$$X_{(n)} = \max\{X_1, X_2, \dots, X_n\}.$$

Intuitivamente percebemos que $X_{(n)}$ deve ser aproximadamente 1 para n grande. Queremos demonstrar que $X_{(n)} \xrightarrow{P} 1$, quando $n \leftarrow \infty$.

A função de distribuição de $X_{(n)}$ é

$$\begin{aligned} F_{X_{(n)}}(x) &= P(X_{(n)} \leq x) = P(X_1 \leq x, X_2 \leq x, \dots, X_n \leq x) \\ &= \prod_{k=1}^n P(X_k \leq x) = x^n, \quad \text{para } 0 \leq x \leq 1. \end{aligned}$$

Então, para $0 < \epsilon < 1$

$$P(|X_{(n)} - 1| > \epsilon) = P(X_{(n)} < 1 - \epsilon) = (1 - \epsilon)^n,$$

da qual temos que $\lim_{n \rightarrow \infty} (1 - \epsilon)^n = 0$ desde que $|1 - \epsilon| < 1$. Obtendo-se que $X_{(n)} \xrightarrow{P} 1$.

Deve-se ressaltar que a convergência em probabilidade é uma propriedade sobre as distribuições das variáveis aleatórias X_1, X_2, \dots, X em vez das próprias variáveis aleatórias. Em particular, a convergência em probabilidade, não implica que a sequência X_1, X_2, \dots convirja pontualmente. Isto fica claro no exemplo a seguir.

Exemplo 7.26

Definamos a sequência de variáveis aleatórias X_1, X_2, \dots da seguinte forma.

$$\begin{aligned} X_1(\omega) &= \begin{cases} 0, & \omega \in [\frac{1}{2}, 1] \\ 1, & \omega \in [0, \frac{1}{2}) \end{cases}, \\ X_2(\omega) &= \begin{cases} 0, & \omega \in [0, \frac{1}{2}) \\ 1, & \omega \in [\frac{1}{2}, 1] \end{cases}, \\ X_3(\omega) &= \begin{cases} 0, & \omega \in [0, 1] \setminus [0, \frac{1}{4}) \\ 1, & \omega \in [0, \frac{1}{4}) \end{cases}, \\ X_4(\omega) &= \begin{cases} 0, & \omega \in [0, 1] \setminus [\frac{1}{4}, \frac{1}{2}) \\ 1, & \omega \in [\frac{1}{4}, \frac{1}{2}) \end{cases}, \\ X_5(\omega) &= \begin{cases} 0, & \omega \in [0, 1] \setminus [\frac{1}{2}, \frac{3}{4}) \\ 1, & \omega \in [\frac{1}{2}, \frac{3}{4}) \end{cases}, \\ X_6(\omega) &= \begin{cases} 0, & \omega \in [0, 1] \setminus [\frac{3}{4}, 1] \\ 1, & \omega \in [\frac{3}{4}, 1] \end{cases}, \\ &\vdots \end{aligned}$$

Qual o limite pontual desta sequência? Qual o limite em probabilidade? Podemos perceber que esta sequência numérica $X_1(\omega), X_2(\omega), \dots$ não converge a nenhum número,

qualquer seja $\omega \in [0, 1]$ quando n cresce. Veja que $X_1(0), X_2(0), \dots$ as vezes é 0 e em outras situações é 1, o mesmo acontece com a sequência numérica $X_1(1), X_2(1), \dots$. Por outro lado, a sequência numérica $X_1(\omega), X_2(\omega), \dots$ converge a zero quando $\omega \in (0, 1)$.

Por outro lado, se $P(X_n = 0) = 1 - \frac{1}{n}$ e $P(X_n = 1) = \frac{1}{n}$ temos que

$$P(|X_n| > \epsilon) = \begin{cases} P(X_n = 1) = \frac{1}{n}, & \text{se } 0 < \epsilon < 1 \\ 0, & \text{caso contrário} \end{cases}.$$

Isto significa, pelo demonstrado no Teorema 7.36, que $\lim_{n \rightarrow \infty} P(|X_n| > \epsilon) = 0$, ou seja, $X_n \xrightarrow{P} 0$. Se, numa outra situação, $P(X_n = 0) = \frac{1}{n}$ e $P(X_n = 1) = 1 - \frac{1}{n}$, concluímos que $X_n \xrightarrow{P} 1$.

Então, o que a convergência em probabilidade realmente significa, afinal? significa que a probabilidade de que X_n não esteja perto de X vai para zero quando $n \rightarrow \infty$ e nada mais do que isso.

Estritamente falando, a definição de convergência em probabilidade se aplica quando todos os X_n e X são finitamente avaliados. Assim, por exemplo, se $X_n(\omega) = +\infty$ e $X(\omega) = +\infty$ para algum ω , então $X_n(\omega) - X(\omega)$ não está definido e, portanto, um ω desse tipo não pode pertencer ao conjunto $\{|X_n - X| > \epsilon\}$ figurando em (7.35).

Teorema 7.37

Seja $\{X_n\}$ uma sequência de variáveis aleatórias que converge em probabilidade à variável aleatória X e g uma função contínua definida em \mathbb{R} . Então

$$g(X_n) \xrightarrow{P} g(X), \quad (7.36)$$

quando $n \rightarrow \infty$.

Demonstração: Dado que X é variável aleatória, temos que, dado $\epsilon > 0$ podemos encontrar uma constante $k = k(\epsilon)$ de maneira que

$$P(|X| > k) \leq \frac{\epsilon}{2}.$$

Também, g é contínua em \mathbb{R} , logo é uniformemente contínua em $[-k, k]$. Segue então que existe $\delta = \delta(\epsilon, k)$ tal que

$$|g(x_n) - g(x)| < \epsilon,$$

sempre que $|x| \leq k$ e $|x_n - x| < \delta$. Sejam os eventos

$$A = \{\omega \in \Omega : |X(\omega)| \leq k\}, \quad B = \{\omega \in \Omega : |X_n(\omega) - X(\omega)| < \delta\}$$

e

$$C = \{\omega \in \Omega : |g(X_n(\omega)) - g(X(\omega))| \leq \epsilon\}.$$

Então, $A \cap B \subseteq C$. Segue que

$$P(C^c) \leq P(A^c) + P(B^c),$$

isto é,

$$P(|g(X_n) - g(X)| \geq \epsilon) \leq P(|X| > k) + P(|X_n - X| \geq \delta),$$

para $n \geq N(\epsilon, \delta, k)$, onde $N(\epsilon, \delta, k)$ é escolhida satisfazendo $P(|X_n - X| \geq \delta) < \frac{\epsilon}{2}$, para $n \geq N(\epsilon, \delta, k)$. ■

Teorema 7.38

Sejam $\{X_n\}$ e $\{Y_n\}$ duas sequências de variáveis aleatórias tais que $X_n \xrightarrow{P} X$ e $Y_n \xrightarrow{P} Y$. Então:

- (a) $X_n \pm Y_n \xrightarrow{P} X \pm Y$.
- (b) Caso $X_n \xrightarrow{P} 1$, então $X_n^{-1} \xrightarrow{P} 1$.
- (c) Se c é uma constante qualquer, então $cX_n \xrightarrow{P} cX$.
- (d) Caso $X_n \xrightarrow{P} c$, então $X_n^2 \xrightarrow{P} c^2$.
- (e) Caso $X_n \xrightarrow{P} a$ e $Y_n \xrightarrow{P} b$ sendo a e b constantes, então $X_n \times Y_n \xrightarrow{P} ab$.
- (f) $X_n - X_m \xrightarrow{P} 0$, quando $n, m \rightarrow \infty$.
- (g) Seja Z uma variável aleatória, então $X_n \times Z \xrightarrow{P} X \times Z$.
- (h) $X_n \times Y_n \xrightarrow{P} X \times Y$.
- (i) Se $\lim_{n \rightarrow \infty} c_n = c$ então $c_n X_n \xrightarrow{P} cX$.

Demonstração: (a) Exercício.

(b) Temos que, para

$$\begin{aligned} P(|X_n^{-1} - 1| \geq \epsilon) &= P(X_n^{-1} \geq 1 + \epsilon) + P(X_n^{-1} \leq 1 - \epsilon) \\ &= P(X_n^{-1} \geq 1 + \epsilon) + P(X_n^{-1} \leq 0) + P(0 < X_n^{-1} \leq 1 - \epsilon) \end{aligned}$$

e cada um dos três termos à direita converge a 0 quando $n \rightarrow \infty$.

(c) Exercício.

(d) Exercício.

(e) Podemos escrever o produto como

$$X_n \times Y_n = \frac{(X_n + Y_n)^2 - (X_n - Y_n)^2}{4},$$

o qual converge para $\frac{1}{4}((a+b)^2 - (a-b)^2) = a \times b$.

(f)

$$P(|X_n - X_m| > \epsilon) \leq P(|X_n - X| > \frac{\epsilon}{2}) + P(|X_m - X| > \frac{\epsilon}{2}),$$

do qual segue que $\lim_{n \rightarrow \infty} P(|X_n - X_m| > \epsilon) = 0$.

(g) Z é variável aleatória e dado $\delta > 0$, existe um $k > 0$ tal que $P(|Z| > k) < \frac{\delta}{2}$. Então,

$$\begin{aligned} P(|X_n Z - X Z| > \epsilon) &= P(|X_n - X||Z| > \epsilon, |Z| > k) + P(|X_n - X||Z| > \epsilon, |Z| \leq k) \\ &< \frac{\delta}{2} + P(|X_n - X| > \frac{\epsilon}{k}). \end{aligned}$$

(h) Observemos que $(X_n - X) \times (Y_n - Y) \xrightarrow{P} 0$ e utilizando o resultado em (e).

(i) Considerar no item (e) que $X_n - X \xrightarrow{P} 0$ e a sequência de variáveis aleatórias $Y_n = c_n$, depois utilizar o resultado em (g). ■

Exemplo 7.27

Sejam X_1, X_2, \dots variáveis aleatórias independentes e identicamente distribuídas, tais que $X_k \sim \text{Exponencial}(1)$, ou seja,

$$F_{X_n}(x) = \begin{cases} 1 - e^{-x}, & \text{para } x \geq 0 \\ 0, & \text{para } x < 0 \end{cases}.$$

Provemos que a sequência de variáveis aleatórias independentes $Y_n = \frac{X_n}{\log(n)}$, converge em probabilidade a 0 para $n > 1$.

Observemos que

$$P(|Y_n| > \epsilon) = P(X_n > \epsilon \log(n)) = e^{-\epsilon \log(n)} = \frac{1}{n^\epsilon}$$

e que $\lim_{n \rightarrow \infty} n^{-\epsilon} = 0$. Por outro lado, podemos considerar o resultado do Teorema 7.38

(i), para o qual escolhemos a sequência de constantes $c_n = \frac{1}{\log(n)}$, para $n > 1$. Então

$Y_n = c_n X_n \xrightarrow{P} 0$, devido a que $\lim_{n \rightarrow \infty} c_n = 0$, como mencionado, isto utilizando o resultado (i) do Teorema 7.38.

Definição 7.5

conteúdo...

O limite em distribuição é único, o que significa que, se \mathbf{X}_n converge em distribuição para \mathbf{X} e \mathbf{Y} , \mathbf{X} e \mathbf{Y} têm a mesma distribuição, isso é $F_X = F_Y$ no contexto de \mathbb{R}^k .

7.3.2 Convergência em momentos

A situação aqui é considerar X, X_1, X_2, \dots variáveis aleatórias definidas no mesmo espaço de probabilidade (Ω, \mathcal{F}, P) e suponha que X_n convirja pontualmente para X , quando $n \rightarrow \infty$. Como a convergência é pontual significa que $X_n(\omega) \rightarrow X(\omega)$, para todo $\omega \in \Omega$.

A questão é a seguinte: sob quais condições a esperança do limite é o limite das esperanças? Isto é, queremos saber quando

$$E(X) = \lim_{n \rightarrow \infty} E(X_n).$$

Infelizmente, nem sempre $E(X_n)$ converge a $E(X)$, quando $n \rightarrow \infty$.

Definição 7.6

Consideremos uma sequência de variáveis aleatórias $\{X_n\}$ tais que $E(|X_n|^r) < \infty$, para algum $r > 0$. Dizemos que a sequência X_n converge em momento de ordem r à variável aleatória X se $E(|X|^r) < \infty$ e

$$\lim_{n \rightarrow \infty} E(|X_n - X|^r) = 0.$$

Escrevemos $X_n \xrightarrow{r} X$.

Exemplo 7.28

Seja $X \sim \text{Cauchy}(0, 1)$ e cada X_n definida como

$$X_n = X 1_{[-n \leq X \leq n]} = \begin{cases} X, & \text{se } -n \leq X \leq n \\ 0, & \text{se } |X| > n \end{cases}.$$

Isto significa que cada X_n é uma variável Cauchy truncada. Podemos perceber que $X_n(\omega) \rightarrow X(\omega)$, pois $X_n(\omega) = X(\omega)$ para $n \geq |X(\omega)|$. Também percebemos que X_n é integrável, porque é limitada e ainda obtemos que $E(X_n) = 0$, isto por ser simétrica. Mas $E(X_n)$ não converge para $E(X)$, porque $E(X)$ não existe.

Exemplo 7.29

Consideremos a sequência $\{X_n\}$ de variáveis aleatórias discretas com função de probabilidade da forma

$$P(X_n = 0) = 1 - \frac{1}{n}, \quad P(X_n = 1) = \frac{1}{n}, \quad n = 1, 2, 3, \dots$$

Obtemos que

$$E(|X_n|^2) = \frac{1}{n},$$

logo $\lim_{n \rightarrow \infty} E(|X_n|^2) = 0$ e, então $X_n \xrightarrow{2} 0$.

Teorema 7.39

Seja $\{X_n\}$ uma sequência de variáveis aleatórias tais que $X_n \xrightarrow{2} X$. Então

$$\lim_{n \rightarrow \infty} E(X_n) = E(X) \quad e \quad \lim_{n \rightarrow \infty} E(X_n^2) = E(X^2).$$

Demonstração :

$$|E(X_n - X)| \leq E(|X_n - X|) \leq E^{1/2}(|X_n - X|^2),$$

do qual sabemos que $\lim_{n \rightarrow \infty} E^{1/2}(|X_n - X|^2) = 0$. Escrevendo

$$E(X_n^2) = E((X_n - X)^2) + E(X^2) + 2E(X(X_n - X))$$

e note que

$$|E(X(X_n - X))| \leq \sqrt{E(X^2)E((X_n - X)^2)},$$

pela desigualdade de Cauchy-Schwarz. Tomando limite obtemos que $E(X_n^2) \xrightarrow{2} E(X^2)$. ■

Obtemos, como consequência deste teorema, que se $X_n \xrightarrow{2} X$ então

$$\lim_{n \rightarrow \infty} \text{Var}(X_n) = \text{Var}(X).$$

Corolário 7.40

Sejam $\{X_n\}$ e $\{Y_m\}$ duas sequências de variáveis aleatórias tais que $X_n \xrightarrow{2} X$ e $Y_m \xrightarrow{2} Y$. Então,

$$\lim_{n,m \rightarrow \infty} E(X_n Y_m) = E(XY).$$

Demonstração : Exercício. ■

Como consequências do Teorema 7.39 e do corolário acima temos que, se $X_n \xrightarrow{2} X$ e $Y_m \xrightarrow{2} Y$ conjuntamente, então

$$\lim_{n,m \rightarrow \infty} \text{Cov}(X_n, Y_m) = \text{Cov}(XY).$$

Corolário 7.41

Seja X uma variável aleatória e r, s números reais tais que $0 < s < r < \infty$. Então,

$$E(|X|^s)^{\frac{1}{s}} \leq E(|X|^r)^{\frac{1}{r}}.$$

Demonstração: Para duas variáveis aleatórias Z e Y , utilizamos a desigualdade de Holder²² e obtemos

$$E(|ZY|) \leq E(|Z|^p)^{\frac{1}{p}} E(|Y|^q)^{\frac{1}{q}}.$$

Escolhendo $Y = 1$ e $Z = |X|^s$, temos $E(|X|^s) \leq E(|X|^{sp})^{\frac{1}{s}}$. Dado que $p \geq 1$, temos $sp \leq s$. Seja $sp = r$, então $E(|X|^s) \leq E(|X|^s)^{\frac{r}{s}}$ ou $E(|X|^s)^{\frac{1}{s}} \leq E(|X|^r)^{\frac{1}{r}}$. ■

Teorema 7.42

Seja $\{X_n\}$ uma sequência de variáveis aleatórias tais que $X_n \xrightarrow{r} X$. Então

$$\lim_{n \rightarrow \infty} E(|X_n|^r) = E(|X|^r).$$

Demonstração: Seja $0 < r \leq 1$. Então

$$E(|X_n|^r) = E(|X_n - X + X|^r),$$

de maneira que

$$E(|X_n|^r) - E(|X|^r) \leq E(|X_n - X|^r).$$

Similarmente temos que

$$E(|X|^r) - E(|X_n|^r) \leq E(|X_n - X|^r),$$

do qual segue que

$$\lim_{n \rightarrow \infty} |E(|X|^r) - E(|X_n|^r)| \leq \lim_{n \rightarrow \infty} E(|X_n - X|^r) = 0.$$

Seja $r > 1$. Então, utilizando a desigualdade de Minkowski para obtermos que

$$\left(E(|X_n|^r)\right)^{\frac{1}{r}} \leq \left(E(|X_n - X|^r)\right)^{\frac{1}{r}} + \left(E(|X|^r)\right)^{\frac{1}{r}}$$

e

$$\left(E(|X|^r)\right)^{\frac{1}{r}} \leq \left(E(|X_n - X|^r)\right)^{\frac{1}{r}} + \left(E(|X_n|^r)\right)^{\frac{1}{r}}.$$

Segue então que

$$\lim_{n \rightarrow \infty} \left| E^{\frac{1}{r}}(|X_n|^r) - E^{\frac{1}{r}}(|X|^r) \right| \leq \lim_{n \rightarrow \infty} E^{\frac{1}{r}}(|X_n - X|^r) = 0. \quad \blacksquare$$

²²Desigualdade de Holder: Sejam p, q números reais tais que $1 \leq p, q < \infty$ relacionados pela igualdade $\frac{1}{p} + \frac{1}{q} = 1$. Então

$$\int |u(x)v(x)| \, dx \leq \sqrt[p]{\int |u(x)|^p \, dx} \sqrt[q]{\int |v(x)|^q \, dx}.$$

Teorema 7.43

Seja $r > s$. Então, se $X_n \xrightarrow{r} X$ temos que $X_n \xrightarrow{s} X$.

Demonstração: Do Corolário 7.41 segue que, para $s < r$

$$\lim_{n \rightarrow \infty} E(|X_n - X|^s) \leq \lim_{n \rightarrow \infty} E^{\frac{s}{r}}(|X_n - X|^r) = 0,$$

dado que $X_n \xrightarrow{r} X$. ■

Claramente, o inverso ao Teorema 7.43 pode não ser verdade, uma vez que $E(|X|^s) < \infty$ para $s < r$, não implica que $E(|X|^r) < \infty$. Devemos observar também que, pelo Teorema 7.42, segue que

$$\text{se } X_n \xrightarrow{r} X \text{ então } \lim_{n \rightarrow \infty} E(|X_n|^s) = E(|X|^s), \text{ para } s \leq r.$$

Investigaremos respostas à questão anteriormente levantada, ou seja, vamos investigar sob quais situações o inverso ao Teorema 7.43 acontece. A primeira resposta é apresentada no teorema a seguir, o qual nos disse que a esperança do limite é o limite das esperanças quando as variáveis são não negativas e a sequência cresce monotonamente.

Teorema 7.44 (Teorema da convergência monótona)

Sejam X, X_1, X_2, \dots variáveis aleatórias definidas no espaço de probabilidade (Ω, \mathcal{F}, P) . Suponhamos que, para todo $\omega \in \Omega$, a sequência $\{X_n\}$ satisfaz:

- (a) $0 \leq X_n(\omega)$,
- (b) $X_n(\omega) \leq X_{n+1}(\omega)$ e
- (c) $\lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)$.

Então

$$\lim_{n \rightarrow \infty} E(X_n) = E(X),$$

sendo que $E(X_n) \leq E(X_{n+1})$.

Demonstração: Sabemos, pelo Teorema 4.7, que $0 \leq E(X_n) \leq E(X)$ e $E(X_n) \leq E(X_{n+1})$.

Logo $\lim_{n \rightarrow \infty} E(X_n) \leq E(X)$. Restaria agora prover que $\lim_{n \rightarrow \infty} E(X_n) \geq E(X) - \epsilon$, $\forall \epsilon > 0$.

Com este objetivo, definamos o evento $B_n = \{\omega \in \Omega : n\epsilon < X(\omega) \leq (n+1)\epsilon\}$, para $n = 0, 1, 2, \dots$ e a variável aleatória

$$Y(\omega) = \sum_{n=0}^{\infty} n\epsilon \mathbf{1}_{B_n}(\omega), \quad \forall \omega \in \Omega.$$

Observemos que $X - \epsilon \leq Y \leq X$ e, então $E(X) - \epsilon \leq E(Y) \leq E(X)$. Vamos provar agora que $E(Y) \leq \lim_{n \rightarrow \infty} E(X_n)$, ou seja, provemos que $\lim_{n \rightarrow \infty} E(X_n) \geq E(X) - \epsilon$. Seja agora $A_m = \{\omega \in \Omega : X_m(\omega) \geq Y\}$ e observemos que $A_m \subseteq A_{m+1}$ e $\bigcup_{m=1}^{\infty} A_m = \Omega$. Isto deve-se a que, se $X_m(\omega) \geq Y(\omega)$, para algum $\omega \in \Omega$, então $X_{m+1}(\omega) \geq Y(\omega)$. Sabemos por hipóteses que a sequência $\{X_n\}$ é monótona não decrescente assim, por tanto, como a sequência de conjuntos $\{A_m\}$. Dado que a sequência $\{X_n\}$ converge a X , então para algum m suficientemente grande devemos ter que $X_m(\omega) \geq Y(\omega)$. Para m fixo, temos que

$$Y(\omega)\mathbf{1}_{A_m}(\omega) = \begin{cases} Y(\omega) \leq X_m(\omega), & \text{se } \omega \in A_m \\ 0 \leq X_m(\omega), & \text{se } \omega \notin A_m \end{cases},$$

logo $0 \leq Y\mathbf{1}_{A_m} \leq X_m$ e $0 \leq E(Y\mathbf{1}_{A_m}) \leq E(X_m)$. Para calcular $E(Y\mathbf{1}_{A_m})$ devemos notar que

$$Y(\omega)\mathbf{1}_{A_m}(\omega) = \begin{cases} n\epsilon, & \text{se } \omega \in B_n \cap A_m, \quad n = 0, 1, 2, \dots \\ 0, & \text{se } \omega \notin \bigcup_{n=0}^{\infty} (B_n \cap A_m) \end{cases}.$$

Então,

$$E(X_m) \geq E(Y\mathbf{1}_{A_m}) = \sum_{n=0}^{\infty} n\epsilon P(B_n \cap A_m) \geq \sum_{n=0}^N n\epsilon P(B_n \cap A_m), \quad \forall N.$$

Mas, a sequência $\{P(B_n \cap A_m)\}_{m=0}^{\infty}$ é não decrescente e tal que $\lim_{m \rightarrow \infty} P(B_n \cap A_m) = P(B_n)$, logo

$$\lim_{m \rightarrow \infty} E(X_m) \geq \lim_{m \rightarrow \infty} \sum_{n=0}^{\infty} n\epsilon P(B_n \cap A_m) = \sum_{n=0}^N n\epsilon P(B_n), \quad \forall N.$$

Portanto,

$$\lim_{m \rightarrow \infty} E(X_m) \geq \sum_{n=0}^N n\epsilon P(B_n) = E(Y).$$

■

Exemplo 7.30

Consideremos uma sequência de variáveis aleatórias cada uma com distribuição Poisson truncada da seguinte forma: seja $X \sim \text{Poisson}(\theta)$ e

$$X_n(\omega) = X(\omega)\mathbf{1}_{A_n}(\omega),$$

onde $A_n = \{\omega \in \Omega : X(\omega) \leq n\}$ para $n = 0, 1, 2, 3, \dots$ e para todo $\omega \in \Omega$. Esta sequência satisfaz os itens (a), (b) e (c) do Teorema 7.44. Então vale que $\lim_{n \rightarrow \infty} E(X_n) = E(X)$. Vamos verificar isso diretamente.

Pelo resultado em (6.5) temos que $E(X_n) = \sum_x x \frac{P(X = x, X \in A_n)}{P(X \in A_n)}$, para $x = 0, 1, 2, 3, \dots$. Desta forma temos $E(X_0) = 0$, $E(X_1) = (1 + \theta)e^{-\theta}$, etc. Mostramos na Figura 7.6 o comportamento da sequências de esperanças a qual, evidentemente, converge para $E(X)$.

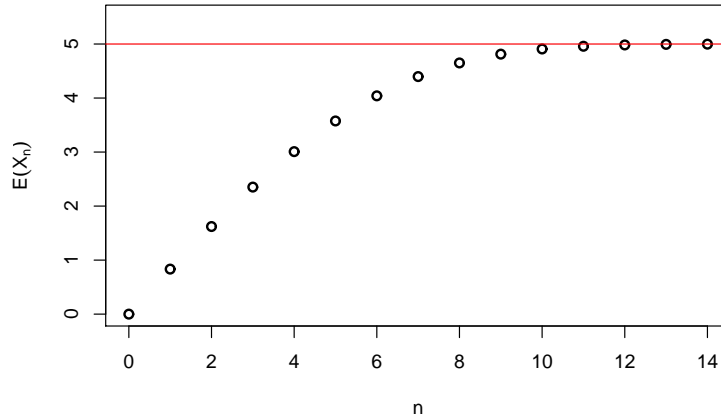


Figura 7.6: Sequência das esperanças das variáveis $\{X_n\}$ no Exemplo 7.30, cada uma desdas variáveis com distribuição Poisson truncada. Mostramos que esta sequência converge ao valor de $E(X)$, escolhido como $\theta = 5$.

Uma outra resposta à questão levantada no início desta seção é também apresentada na forma de teorema, conhecido como de convergência dominada. Resumidamente, a esperança do limite é o limite das esperanças quando a sequência seja uniformemente limitada ou dominada por uma variável cuja esperança seja finita.

Teorema 7.45 (Teorema da convergência dominada)

Sejam Y, X, X_1, X_2, \dots variáveis aleatórias definidas no espaço de probabilidade (Ω, \mathcal{F}, P) tais que $E(Y)$ existe, $|X_n| \leq Y$ e $\lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)$ para todo $\omega \in \Omega$. Então, $E(X)$ e $E(X_n) \forall n$ existem e satisfazem que

$$\lim_{n \rightarrow \infty} E(X_n) = E(X).$$

Demonstração: Pela propriedade no Teorema 4.7 sabemos que, tanto X quanto todas as X_n , são dominadas por Y , observemos que $|X| = \lim_{n \rightarrow \infty} |X_n| \leq Y$, e por causa disto temos a existência de $E(X)$ e $E(X_n)$. Definamos

$$Y_n = \inf_{k \geq n} X_k,$$

observemos que Y_n é variável aleatória devido a que $[Y_n \leq y] = \bigcap_{k \geq n} [X_k \leq y]$. Então,

$Y_n \leq Y_{n+1}$ e ainda temos

$$X(\omega) = \lim_{n \rightarrow \infty} X_n(\omega) = \lim_{n \rightarrow \infty} \inf_{k \geq n} X_k(\omega) = \lim_{n \rightarrow \infty} Y_n(\omega), \quad \forall \omega \in \Omega.$$

Logo $\{(Y_n + Y)\}$ formam um sequência que satisfaz $\lim_{n \rightarrow \infty} (Y_n + Y) = (X + Y)$ e não decrescente, ou seja, $(Y_n + Y) \leq (Y_{n+1} + Y)$. Temos que $\forall n, X_n \geq -Y$ o qual implica que $Y_n \geq -Y$ que,

por sua vez, implica em $(Y_n + Y) \geq 0$. Pelo Teorema 7.44 obtemos que $\lim_{n \rightarrow \infty} E(Y_n + Y) = E(X + Y)$ de maneira monótona, ou seja, sendo ainda que $E(Y_n + Y) \leq E(Y_{n+1} + Y)$. O qual implica que

$$\lim_{n \rightarrow \infty} E(Y_n) = E(X) \quad (7.37)$$

de maneira monótona, ou seja, $E(Y_n) \leq E(Y_{n+1})$. Similarmente, definamos

$$Z_n = \sup_{k \geq n} X_k,$$

então a sequência $\{Z_n\}$ decresce de maneira monótona a X , ou seja, $Z_n(\omega) \geq Z_{n+1}(\omega)$ e $\lim_{n \rightarrow \infty} Z_n(\omega) = X(\omega)$, $\forall \omega$. Isto significa que $\{(Y - Z_n)\}$ formam um sequência que satisfaz $\lim_{n \rightarrow \infty} (Y - Z_n) = (Y - X)$ e não decrescente, ou seja, $(Y - Z_n) \leq (Y - Z_{n+1})$. Temos que $\forall n$, $X_n \leq Y$ o qual implica que $Z_n \leq Y$ que, por sua vez, implica em $(Y - Z_n) \geq 0$. Pelo Teorema 7.44 obtemos que $\lim_{n \rightarrow \infty} E(Y - Z_n) = E(Y - X)$ de maneira monótona, ou seja, sendo ainda que $E(Y - Z_n) \leq E(Y - Z_{n+1})$. O qual implica que

$$\lim_{n \rightarrow \infty} E(Z_n) = E(X) \quad (7.38)$$

de maneira monótona, ou seja, $E(Z_n) \geq E(Z_{n+1})$. Devido a que $Y_n = \inf_{k \geq n} X_k \leq X_n \leq \sup_{k \geq n} X_k = Z_n$, temos $E(Y_n) \leq E(X_n) \leq E(Z_n)$. Utilizando os resultados em (7.37) e em (7.38) concluímos que $\lim_{n \rightarrow \infty} E(X_n) = E(X)$. ■

7.3.3 Convergência quase certamente

Mencionamos aqui um outro tipo de convergência para sequências de variáveis aleatórias, a saber, a convergência com probabilidade 1 ou quase certamente. No interesse da completude, discutiremos este tipo de convergência, embora não nos referiremos a ele subsequentemente no texto.

Definição 7.7

Seja $\{X_n\}_{n \geq 1}$ uma sequência de variáveis aleatórias e X uma outra variável aleatória todas definidas no mesmo espaço de probabilidade (Ω, \mathcal{F}, P) . Dizemos que a sequência $\{X_n\}_{n \geq 1}$ converge quase certamente à variável aleatória X se

$$P\left(\left\{\omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\right\}\right) = 1. \quad (7.39)$$

Nesta situação escrevemos $X_n \xrightarrow{q.c.} X$.

O que exatamente a convergência com probabilidade 1 ou quase certamente significa? Pela definição acima, se $X_n \xrightarrow{q.c.} X$ então $X_n(\omega) \longleftrightarrow X(\omega)$ para todos os elementos $\omega \in A$ com $P(A) = 1$. Para um dado $\omega \in A$ e $\epsilon > 0$, existe um número $n_\epsilon(\omega)$ tal que $|X_n(\omega) - X(\omega)| \leq \epsilon$, para todo $n \geq n_\epsilon(\omega)$.

Observemos que convergência quase certamente é convergência pontual, com probabilidade 1. Dizemos então que, se $X_n \xrightarrow{q.c.} X$, significa que $X_n(\omega)$ converge para $X(\omega)$, para quase tudo ω .

Exemplo 7.31

Suponhamos que X_1, X_2, \dots sejam variáveis aleatórias independentes identicamente distribuídas com distribuição $U(0, 1)$ e

$$X_{(n)} = \max(X_1, X_2, \dots, X_n).$$

Observemos que $1 \geq X_{(n+1)}(\omega) \geq X_{(n)}(\omega)$, para todo ω e então

$$\begin{aligned} \{\omega \in \Omega : |X_{(n+1)} - 1| > \epsilon\} &= \{\omega \in \Omega : X_{(n+1)} < 1 - \epsilon\} \\ &\subseteq \{\omega \in \Omega : X_{(n)} < 1 - \epsilon\} \\ &= \{\omega \in \Omega : |X_{(n)} - 1| > \epsilon\}. \end{aligned}$$

Portanto

$$\lim_{n \rightarrow \infty} P \left(\bigcup_{k=n}^{\infty} \{\omega \in \Omega : |X_{(k)} - 1| > \epsilon\} \right) = \lim_{n \rightarrow \infty} P \left(\{\omega \in \Omega : |X_{(n)} - 1| > \epsilon\} \right) = 0.$$

Então $X_{(n)} \xrightarrow{q.c.} 1$.

Teorema 7.46

Dizemos que $X_n \xrightarrow{q.c.} X$ se, e somente se,

$$\lim_{n \rightarrow \infty} P \left(\sup_{m \geq n} |X_m - X| > \epsilon \right) = 0, \quad \text{para todo } \epsilon > 0.$$

Demonstração: Observemos que se $X_n \xrightarrow{q.c.} X$, então $X_n - X \xrightarrow{q.c.} 0$. É suficiente mostrar a equivalência entre as afirmações:

$$(a) X_n \xrightarrow{q.c.} 0 \quad \text{e} \quad (b) \lim_{n \rightarrow \infty} P \left(\sup_{m \geq n} |X_m| > \epsilon \right) = 0, \quad \forall \epsilon > 0.$$

Primeiro provemos que (a) implica (b). Seja $\epsilon > 0$ e definamos

$$A_n(\epsilon) = \left\{ \omega \in \Omega : \sup_{m \geq n} |X_m(\omega)| > \epsilon \right\} \quad \text{e} \quad C = \left\{ \omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = 0 \right\}.$$

Vamos definir também a sequência $B_n(\epsilon) = C \cap A_n(\epsilon)$, observemos que $B_{n+1}(\epsilon) \subset B_n(\epsilon)$ e ainda que $\bigcap_{n \geq 1} B_n(\epsilon) = \emptyset$. Segue que

$$\lim_{n \rightarrow \infty} P(B_n(\epsilon)) = P \left(\bigcap_{n \geq 1} B_n(\epsilon) \right) = 0.$$

Dado que $P(C) = 1$, $P(C^c) = 0$ e

$$\begin{aligned} P(B_n(\epsilon)) &= P(A_n(\epsilon) \cap C) = P(C^c \cup A_n(\epsilon)^c) \\ &= 1 - P(C^c) - P(A_n(\epsilon)^c) + P(C^c \cap A_n(\epsilon)^c) \\ &= P(A_n(\epsilon)) + P(C^c \cap A_n(\epsilon)^c) = P(A_n(\epsilon)). \end{aligned}$$

Disto segue que (b) é válido. Reciprocamente, seja $\lim_{n \rightarrow \infty} P(A_n(\epsilon)) = 0$ e definamos

$$D(\epsilon) = \left\{ \omega \in \Omega : \lim_{n \rightarrow \infty} |X_n(\omega)| > \epsilon \right\}, \quad \forall \epsilon > 0.$$

Observemos que $D(\epsilon) \subset A_n(\epsilon)$ para $n = 1, 2, \dots$, do qual concluímos que $P(D(\epsilon)) = 0$. Também

$$C^c = \left\{ \omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) \neq 0 \right\} \subset \bigcup_{k=1}^{\infty} \left\{ \omega \in \Omega : \lim_{n \rightarrow \infty} |X_n(\omega)| > \frac{1}{k} \right\},$$

de maneira que

$$1 - P(C) \leq \sum_{k=1}^{\infty} P(D(1/k)) = 0,$$

e, portanto (a) é válido. ■

Exemplo 7.32

Suponha que $X_n \xrightarrow{q.c.} X$ e que g seja uma função contínua. Então $g(X_n) \xrightarrow{q.c.} g(X)$.

Para ver isso, seja A o conjunto dos $\omega \in \Omega$ para os quais $\lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)$. Desde que g é uma função contínua, a convergência pontual de $\lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)$ implica a convergência pontual de $\lim_{n \rightarrow \infty} g(X_n(\omega)) = g(X(\omega))$. Isso ocorre para todos os ω no conjunto A e como para qualquer $\omega \notin A$, $\lim_{n \rightarrow \infty} X_n(\omega) \neq X(\omega)$, então $P(A) = 1$ e assim $g(X_n) \xrightarrow{q.c.} g(X)$.

Devemos esclarecer que $X_n \xrightarrow{q.c.} 0$ significa que, para $\epsilon > 0$ e $\eta > 0$ arbitrário, podemos encontrar um n_0 de maneira que

$$P\left(\sup_{n \geq n_0} |X_n| > \epsilon\right) < \eta. \quad (7.40)$$

De fato, podemos escrever, de forma equivalente, que

$$\lim_{n_0 \rightarrow \infty} P\left(\bigcup_{n \geq n_0} \{|X_n| > \epsilon\}\right) = 0. \quad (7.41)$$

Podemos alargar a Definição 7.7 ao caso multivariado de uma forma totalmente simples, como mostrado a continuação.

Definição 7.8

Seja $\{\mathbb{X}_n\}$ uma sequência de variáveis aleatórias vetoriais e \mathbb{X} uma outra variável aleatória vetorial, da mesma dimensão do que a sequência, todas definidas no mesmo espaço de probabilidade (Ω, \mathcal{F}, P) . Dizemos que a sequência $\{\mathbb{X}_n\}_{n \geq 1}$ converge quase certamente à variável aleatória vetorial \mathbb{X} se

$$P\left(\left\{\omega \in \Omega : \lim_{n \rightarrow \infty} \mathbb{X}_n(\omega) = \mathbb{X}(\omega)\right\}\right) = 1. \quad (7.42)$$

Nesta situação escrevemos $\mathbb{X}_n \xrightarrow{q.c.} \mathbb{X}$.

Em alternativa, uma vez que a demonstração do Teorema 7.46 também se aplica a vectores aleatórios, dizemos que $\mathbb{X}_n \xrightarrow{q.c.} \mathbb{X}$ se para qualquer $\epsilon > 0$,

$$\lim_{n \rightarrow \infty} P\left(\sup_{m \geq n} \|\mathbb{X}_m - \mathbb{X}\| > \epsilon\right) = 0.$$

Vimos nos Teoremas 2.24 e 2.30 que as funções contínuas preservam tanto a convergência na probabilidade como a convergência na distribuição. No entanto, estes factos eram bastante difíceis de provar. Felizmente, o resultado análogo para a convergência com a probabilidade é bastante fácil de provar. De facto, como a convergência quase certa é definida em termos de convergência de sequências de vectores reais (e não aleatórios), o teorema seguinte pode ser provado utilizando o mesmo método de prova utilizado para o teorema 1.16.

7.3.4 Convergência em distribuição

Consideremos agora mais uma forma de convergência. Esta será a última forma de convergência e para entendermos do que se trata, vejamos el seguinte exemplo.

Exemplo 7.33

Considere a sequência de funções de distribuição

$$F_n(x) = \begin{cases} 0, & \text{se } x < n \\ 1, & \text{se } x \geq n \end{cases}.$$

Aqui F_n é a função de distribuição da variável aleatória X , degenerada em $x = n$. Observamos que F_n converge a uma função F que é identicamente igual a zero a qual, portanto, não é uma função de distribuição.

Como deve ter ficado claro, tratamos aqui de convergência de função de distribuição. Digamos assim, se $\{F_n\}_{n \geq 1}$ for uma sequência de funções de distribuição, se existe uma função de distribuição F , de modo que

$$\lim_{n \rightarrow \infty} F_n(x) = F(x), \quad (7.43)$$

em todo ponto x no qual F é contínua, dizemos que F_n converge em distribuição para F .

Devemos observar que é possível uma determinada sequência de funções de distribuição converja para uma função que não seja função de distribuição, como visto no Exemplo 7.33. É importante lembrar que se a sequência $\{F_n\}$ converge para F , isso somente implica a convergência de funções de distribuição e não das próprias variáveis aleatórias. Desta maneira fundamental, a convergência na distribuição é bem diferente da convergência em probabilidade ou convergência quase certamente.

Definição 7.9

Seja $\{X_n\}_{n \geq 1}$ uma sequência de variáveis aleatórias e $\{F_{X_n}\}_{n \geq 1}$ a correspondente sequência de funções de distribuição. Dizemos que X_n converge em distribuição se existe uma variável aleatória X , com função de distribuição F_X , de maneira que

$$\lim_{n \rightarrow \infty} F_n(x) = F(x), \quad (7.44)$$

em todo ponto x no qual F é contínua. Escrevemos $X_n \xrightarrow{D} X$.

Notemos que convergência em distribuição define-se em termos das funções de distribuição e formalmente, não necessariamente as variáveis aleatórias correspondentes devem estar definidas no mesmo espaço de probabilidade. No seguinte exemplo mostramos uma situação na qual a sequência de variáveis aleatórias converge para uma variável aleatória, mas a distribuição limite é completamente diferente da distribuição das componentes da sequência.

Exemplo 7.34

Sejam X_1, X_2, \dots variáveis aleatórias igualmente distribuídas $Uniforme(0, \theta)$, $\theta > 0$. Definamos a variável aleatória

$$X_{(n)} = \max(X_1, X_2, \dots, X_n).$$

Então, a função de densidade é

$$f_{X_{(n)}}(x) = \begin{cases} \frac{xx^{n-1}}{\theta^n}, & \text{caso } 0 < x < \theta \\ 0, & \text{caso contrário} \end{cases},$$

com função de distribuição de cada $X_{(n)}$ dada por

$$F_n(x) = \begin{cases} 0, & \text{caso } x < 0 \\ \left(\frac{x}{\theta}\right)^n, & \text{caso } 0 \leq x < \theta \\ 1, & \text{caso } x \geq \theta \end{cases}.$$

Podemos perceber que, quando $n \rightarrow \infty$ obtemos que

$$\lim_{n \rightarrow \infty} F_n(x) = \lim_{n \rightarrow \infty} \begin{cases} 0, & \text{caso } x < 0 \\ \left(\frac{x}{\theta}\right)^n, & \text{caso } 0 \leq x < \theta \\ 1, & \text{caso } x \geq \theta \end{cases} = F(x) = \begin{cases} 0, & \text{caso } x < \theta \\ 1, & \text{caso } x \geq \theta \end{cases},$$

a qual é uma função de distribuição. Por isso afirmamos que $X_{(n)} \xrightarrow{D} X$, onde $X = \theta$.

Teorema 7.47

Seja $\{X_n\}$ uma sequência de variáveis aleatórias tal que $X_n \xrightarrow{D} X$. Então

(a) $X_n + c \xrightarrow{D} X + c$, sendo c uma constante real qualquer.

(b) $cX_n \xrightarrow{D} cX$, sendo $c \neq 0$.

Demonstração : Exercício. ■

Notemos que convergência em distribuição é definida em termos das funções de distribuição por isso, formalmente, as variáveis aleatórias não precisam estar definidas no mesmo espaço de probabilidade. O seguinte exemplo ilustra o motivo pelo qual requeremos convergência apenas nos pontos de continuidade de F .

Exemplo 7.35

Seja $\{X_n\}$ uma sequência de variáveis aleatórias definidas em (Ω, \mathcal{F}, P) tais que

$$X_n(\omega) = \frac{1}{n}, \quad \omega \in \Omega,$$

para todo $n = 1, 2, \dots$ e X a variável aleatória constante zero, ou seja, $P(X = 0) = 1$. A função de distribuição de X_n é

$$F_{X_n}(x) = \begin{cases} 1, & \text{se } x \geq \frac{1}{n} \\ 0, & \text{se } x < \frac{1}{n} \end{cases},$$

e a função de distribuição de X é

$$F_X(x) = \begin{cases} 1, & \text{se } x \geq 0 \\ 0, & \text{se } x < 0 \end{cases}. \quad (7.45)$$

Acontece que

- Se $x > 0$ temos que $\lim_{n \rightarrow \infty} F_{X_n}(x) = 1$.
- Se $x < 0$ temos que $\lim_{n \rightarrow \infty} F_{X_n}(x) = 0$, de modo que $\{F_{X_n}(x)\}$ converge para $F(x)$, nos pontos de continuidade de F .
- No ponto de descontinuidade de F , $x = 0$, temos que

$$\lim_{n \rightarrow \infty} F_{X_n}(0) = 0,$$

mas $F(0) \neq 0$, $F(0) = 1$ como podemos observar em (7.45), diferente do valor ao qual converge a sequência.

Vejamos no seguinte exemplo, que deve-se a Curtiss (1942), uma outra sequência de variáveis aleatórias que convergem a mesma função de distribuição.

Exemplo 7.36

Considere a sequência de funções de distribuição

$$F_n(x) = \begin{cases} 0, & \text{caso } x < -n, \\ \frac{1}{2} + c_n \arctan(nx), & \text{caso } -n \leq x < n, \\ 1, & \text{caso } x \geq n, \end{cases}$$

onde $c_n = \frac{1}{2 \arctan(n^2)}$. Pode-se observar que

$$\lim_{n \rightarrow \infty} F_n(x) = \begin{cases} 0, & x < 0, \\ 1, & x \geq 0, \end{cases}$$

em todos os pontos de continuidade da função de distribuição F .

Geralmente, é difícil verificar a convergência em distribuição diretamente por vários motivos. Por exemplo, muitas vezes é difícil trabalhar com as funções de distribuição F_{X_n} . Além disso, em muitos casos, a função de distribuição F_{X_n} pode não ser especificada exatamente, mas pode pertencer a uma classe mais ampla. Podemos saber, por exemplo, que a média e a variância correspondem a F_{X_n} , mas pouco mais sobre F_{X_n} . Do ponto de vista prático, os casos onde F_{X_n} não é conhecida exatamente são os mais interessantes; se F_{X_n} é conhecida exatamente, não há realmente nenhuma razão para se preocupar com uma distribuição limitante F a menos que F_{X_n} seja difícil de trabalhar computacionalmente.

Por estas razões, gostaríamos de ter métodos alternativos para estabelecer convergência na distribuição. Felizmente, existem várias outras condições suficientes para a convergência na distribuição que são úteis na prática para verificar se uma sequência de variáveis aleatórias converge na distribuição e determina a distribuição da variável aleatória limitante.

Nosso objetivo agora é provar que $X_n \xrightarrow{D} X$ se, e somente se, $\varphi_{X_n}(t) \xrightarrow{D} \varphi_X(t)$, ou seja, queremos provar que

$$\lim_{n \rightarrow \infty} F_{X_n}(t) = F_X(t) \quad \text{se, e somente se,} \quad \lim_{n \rightarrow \infty} \varphi_{X_n}(t) = \varphi_X(t), \quad \forall t \in \mathbb{R}.$$

Isto é justamente a afirmação do Teorema 7.19, ou seja, o Teorema de Continuidade de Paul Lévy justifica a afirmação acima.

Exemplo 7.37

Sejam X_1, X_2, \dots variáveis aleatórias tais que $X_n \sim \text{Geométrica}(\lambda/n)$, onde $\lambda > 0$ é uma

constante. Definamos uma nova sequência Y_n , sendo que

$$Y_n = \frac{1}{n}X_n, \quad n = 1, 2, \dots$$

Vamos demonstrar que Y_n converge em distribuição à *Exponencial*(λ).

Seja $W \sim \text{Geométrica}(p)$, então para qualquer inteiro positivo m , temos

$$\begin{aligned} P(W \leq m) &= \sum_{k=1}^m (1-p)^{k-1} p = p \sum_{k=1}^m (1-p)^{k-1} \\ &= p \frac{1 - (1-p)^m}{1 - (1-p)} = 1 - (1-p)^m. \end{aligned}$$

Agora, dado que $Y_n = \frac{1}{n}X_n$, para qualquer número real, podemos escrever

$$P(Y_n \leq y) = P(X_n \leq ny) = 1 - \left(1 - \frac{\lambda}{n}\right)^{[ny]},$$

sendo que $[ny]$ representa a maior inteiro menor ou igual a ny . Podemos escrever agora

$$\begin{aligned} \lim_{n \rightarrow \infty} F_{Y_n}(y) &= \lim_{n \rightarrow \infty} 1 - \left(1 - \frac{\lambda}{n}\right)^{[ny]} \\ &= 1 - \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^{[ny]} = 1 - e^{-\lambda y}. \end{aligned}$$

A última igualdade vale porque $ny - 1 \leq [ny] \leq ny$ e $\lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^{[ny]} = e^{-\lambda y}$.

7.4 Relações entre os modos de convergência

Queremos entender nesta seção como estão relacionados as diferentes formas de convergência estocástica definidas. Isto significa que investigaremos em quais situações coincidem, em quais não e se existe uma forma mais fraca e alguma mais forte. Entenderemos como convergência mais fraca àquela induzida por qualquer uma das outras e como convergência mais forte àquela que induz as outras. Confuso? vamos entender isso aqui.

Teorema 7.48

Seja $\{X_n\}_{n \geq 1}$ uma sequência de variáveis aleatórias e X uma outra variável aleatória tais que $X_n \xrightarrow{P} X$. Então, $X_n \xrightarrow{D} X$.

Demonstração: Sejam F_n e F as funções de distribuição de X_n e X , respectivamente. Temos

então que

$$\begin{aligned}\{\omega : X(\omega) \leq x'\} &= \{\omega : X_n(\omega) \leq x, X(\omega) \leq x'\} \cup \{\omega : X_n(\omega) > x, X(\omega) \leq x'\} \\ &\subseteq \{\omega : X(\omega) \leq x\} \cup \{\omega : X_n(\omega) > x, X(\omega) \leq x'\}.\end{aligned}$$

Segue então que

$$F(x') \leq F_n(x) + P(X_n > x, X \leq x').$$

Dado que $X_n - X \xrightarrow{P} 0$ temos que, para $x' < x$

$$P(X_n > x, X \leq x') \leq P(|X_n - X| > x - x') \xrightarrow{n \rightarrow \infty} 0.$$

Portanto

$$F(x') \leq \liminf_{n \rightarrow \infty} F_n(x), \quad x' < x.$$

Similarmente, trocando X e X_n assim como x e x' , temos que

$$\limsup_{n \rightarrow \infty} F_n(x) \leq F(x''), \quad x < x''.$$

Então, para $x' < x < x''$, temos que

$$F(x') \leq \liminf_{n \rightarrow \infty} F_n(x) \leq \limsup_{n \rightarrow \infty} F_n(x) \leq F(x'').$$

Devido a F ter somente um número enumerável de pontos de descontinuidades, escolhemos x como um ponto de continuidade de F e fazemos $x'' \rightarrow x^+$ e $x' \rightarrow x^-$, do qual temos que

$$F(x) = \lim_{n \rightarrow \infty} F_n(x)$$

em todo ponto de continuidade de F . ■

Devemos observar que o resultado contrário nem sempre é verdade, ou seja, em geral convergência em distribuição não implica convergência em probabilidade. Vejamos isto no seguinte exemplo.

Exemplo 7.38

Sejam X, X_1, X_2, X_3, \dots variáveis aleatórias igualmente distribuídas com distribuição conjunta do (X_n, X) como:

$$\begin{aligned}P(X_n = 0, X = 0) &= 0, & P(X_n = 0, X = 1) &= \frac{1}{2}, \\ P(X_n = 1, X = 0) &= \frac{1}{2}, & P(X_n = 1, X = 1) &= 0.\end{aligned}$$

Observemos que

$$F(x) = F_n(x) = \begin{cases} 0, & \text{caso } x < 0 \\ \frac{1}{2}, & \text{caso } 0 \leq x < 1 \\ 1, & \text{caso } x \geq 1 \end{cases}$$

é a forma das funções de distribuição das variáveis aleatórias, respectivamente. Fica claro

que

$$\lim_{n \rightarrow \infty} F_n(x) = F(x), \text{ ou seja, } X_n \xrightarrow{D} X.$$

Por outro lado, vamos demonstrar que $X_n \not\xrightarrow{P} X$.

$$\begin{aligned} P(|X_n - X| > \tfrac{1}{2}) &\geq P(|X_n - X| = 1) \\ &= P(X_n = 0, X = 1) + P(X_n = 1, X = 0) = 1 \xrightarrow{P} 0. \end{aligned}$$

Portanto, $X_n \not\xrightarrow{P} X$, mas $X_n \xrightarrow{D} X$.

Teorema 7.49

Seja $\{X_n\}_{n \geq 1}$ uma sequência de variáveis aleatórias tal que $X_n \xrightarrow{D} c$, onde c é uma constante. Então, $X_n \xrightarrow{P} c$ se, e somente se, $X_n \xrightarrow{D} c$.

Demonstração: Exercício. ■

Retornemos ao Exemplo 7.29. Esse exemplo vai nos auxiliar para percebermos que convergência em probabilidade não implica em convergência em momentos. Observemos que

$$E(X_n^k) = n^k \frac{1}{n} = n^{k-1},$$

sendo k um número inteiro positivo. Também $E(X^k) = 0$ de maneira que, qualquer seja k , $E(X_n^k) \not\xrightarrow{P} E(X^k)$. Mas, a sequência no referido exemplo é tal que $X_n \xrightarrow{P} 0$.

O seguinte resultado nos mostra que a convergência em momentos é mais forte que a convergência em probabilidade. Podemos provar isso usando a desigualdade de Markov.

Teorema 7.50

Seja $\{X_n\}_{n=1}^{\infty}$ uma sequência de variáveis aleatórias tais que $X_n \xrightarrow{r} X$ para algum $r > 0$. Então, $X_n \xrightarrow{P} X$.

Demonstração: Para qualquer $\epsilon > 0$, temos

$$P(|X_n - X| \geq \epsilon) = P(|X_n - X|^r \geq \epsilon^r) \leq \frac{E(|X_n - X|^r)}{\epsilon^r},$$

pela Desigualdade de Markov, Teorema 7.29. Dado que assumimos $\lim_{n \rightarrow \infty} E(|X_n - X|^r) = 0$, concluímos que $P(|X_n - X| \geq \epsilon) = 0$. ■

O inverso do Teorema 7.50 não é verdade em geral, ou seja, existem sequências que convergem em probabilidade, mas não em momentos. Para isso voltamos a considerar o Exemplo 7.29. Naquela situação consideramos a sequência $\{X_n\}$ de variáveis aleatórias definidas como

$$P(X_n = 0) = 1 - \frac{1}{n^r}, \quad P(X_n = n) = \frac{1}{n^r}, \quad r > 0, \quad n = 1, 2, \dots$$

Então, $E(|X_n|^r) = 1$, de maneira que $X_n \not\xrightarrow{r} 0$. Verificamos agora que $X_n \xrightarrow{P} 0$. Para isto vemos

$$P(|X_n| > \epsilon) = \begin{cases} P(X_n = n), & \text{se } \epsilon < n \\ 0, & \text{se } \epsilon > n \end{cases} \rightarrow 0 \quad n \rightarrow \infty.$$

Como visto, podemos ter uma sequência que convirja em probabilidade mas isso não significa que a mesma será convergente em momentos.

Exemplo 7.39

Seja $\{X_n\}_{n=1}^{\infty}$ uma sequência de variáveis aleatórias definidas como

$$P(X_n = 0) = 1 - \frac{1}{n}, \quad P(X_n = 1) = \frac{1}{n}, \quad n = 1, 2, \dots$$

Então

$$E(|X_n|^2) = \frac{1}{n}, \quad \text{logo} \quad \lim_{n \rightarrow \infty} E(|X_n|^2) = \lim_{n \rightarrow \infty} \frac{1}{n} = 0,$$

e, vemos que, $X_n \xrightarrow{2} X$, sendo X uma variável aleatória degenerada em 0. Pelo Teorema 7.50, como consequência, $X_n \xrightarrow{P} 0$.

Teorema 7.51

Seja $\{X_n\}_{n=1}^{\infty}$ uma sequência de variáveis aleatórias tal que $X_n \xrightarrow{q.c.} X$. Então $X_n \xrightarrow{P} X$.

Demonstração: Pela relação em (7.40), $X_n \xrightarrow{q.c.} X$ implica que, para arbitrários $\epsilon > 0$ e $\eta > 0$, podemos escolher $n_0 = n_0(\epsilon, \eta)$ de maneira que

$$P\left(\bigcap_{n=n_0}^{\infty} \{|X_n - X| \leq \epsilon\}\right) \geq 1 - \eta.$$

Claro que

$$\bigcap_{n=n_0}^{\infty} \{|X_n - X| \leq \epsilon\} \subset \{|X_n - X| \leq \epsilon\}, \quad \text{para } n \geq n_0.$$

Segue então que, para $n \geq n_0$,

$$P(\{|X_n - X| \leq \epsilon\}) \geq P\left(\bigcap_{n=n_0}^{\infty} \{|X_n - X| \leq \epsilon\}\right) \geq 1 - \eta,$$

isto é

$$P(\{|X_n - X| > \epsilon\}) < \eta, \quad \text{para } n \geq n_0,$$

o que é o mesmo de afirmar $X_n \xrightarrow{P} X$. ■

O resultado contrário ao Teorema 7.51 não é válido e podemos ver isso no seguinte exemplo, ou seja, vamos construir um exemplo que convirja em probabilidade mas não quase certamente.

Exemplo 7.40

Para cada inteiro positivo n existem inteiros m and k , univocamente determinados, de maneira que

$$n = 2^k + m, \quad 0 \leq m < 2^k, \quad k = 0, 1, 2, \dots$$

Por exemplo, para $n = 1$, $k = 0$ e $m = 0$; para $n = 5$, $k = 2$ e $m = 1$ e assim por diante. Definamos as variáveis aleatórias X_n , para $n = 1, 2, \dots$, em $\Omega = [0, 1]$ como

$$X_n(\omega) = \begin{cases} 2^k, & \text{caso } \frac{m}{2^k} \leq \omega < \frac{m+1}{2}, \\ 0, & \text{caso contrário} \end{cases}$$

Definamos a função de probabilidade de X_n sendo

$$P(\{I\}) = \text{comprimento do intervalo } I \subset \Omega. \quad (7.46)$$

Portanto $P(X_n = 2^k) = \frac{1}{2^k}$, $P(X_n = 0) = 1 - \frac{1}{2^k}$. O limite $\lim_{n \rightarrow \infty} X_n(\omega)$ não existe para nenhum $\omega \in \Omega$, de maneira que X_n não converge quase certamente. Mas,

$$P(|X_n| > \epsilon) = P(X_n > \epsilon) = \begin{cases} 0, & \text{caso } \epsilon \geq 2^k \\ \frac{1}{2^k}, & \text{caso } 0 < \epsilon < 2^k \end{cases}$$

e vemos que $\lim_{n \rightarrow \infty} P(|X_n| > \epsilon) = 0$.

No exemplo anterior somente deixamos ao leitor demonstrar que a função em (7.46) é de probabilidade. Vejamos uma outra situação na qual existe convergência em probabilidade mas não quase certamente.

Exemplo 7.41

Consideremos o espaço amostral como o intervalo semiaberto $\Omega = (0, 1]$ e, para todo intervalo $A \subset \Omega$, da forma $A = (a, b]$, seja $P(A) = b - a$ o comprimento desse intervalo.

Definamos a sequência de intervalos A_1, A_2, \dots como segue:

$$\begin{aligned} A_1 &= (0, 1] \\ A_2, A_3, A_4 &= \left(0, \frac{1}{2}\right] \cup \left(\frac{1}{3}, \frac{2}{3}\right] \cup \left(\frac{2}{3}, 1\right] \\ A_5, A_6, A_7, A_8, A_9 &= \left(0, \frac{1}{5}\right] \cup \left(\frac{1}{5}, \frac{2}{5}\right] \cup \left(\frac{2}{5}, \frac{3}{5}\right] \cup \left(\frac{3}{5}, \frac{4}{5}\right] \cup \left(\frac{4}{5}, 1\right] \\ &\vdots \quad \vdots \quad \vdots \\ A_{m^2+1} \text{ até } A_{(m+1)^2} &= \left(0, \frac{1}{2m+1}\right] \cup \dots \cup \left(\frac{2m}{2m+1}, 1\right] \\ &\vdots \quad \vdots \quad \vdots \end{aligned}$$

Note-se em particular que

$$P(A_n) = \frac{1}{2m+1},$$

em que $m = [\sqrt{n-1}]$ é o maior inteiro não superior a $\sqrt{n-1}$, ou seja, m é a parte inteira de $\sqrt{n-1}$. Agora, definamos $X_n = \mathbf{1}_{\{A_n\}}$ e vamos escolher $0 < \epsilon < 1$. Então $P(|X_n - 0| < \epsilon)$ é o mesmo que $1 - P(A_n)$. Desde que

$$\lim_{n \rightarrow \infty} P(A_n) = 0 = \lim_{n \rightarrow \infty} \frac{1}{2[\sqrt{n-1}] + 1} = 0,$$

concluimos que $X_n \xrightarrow{P} 0$.

No entanto, não é verdade que $X_n \xrightarrow{q.c.} 0$. Uma vez que cada $\omega \in \Omega$ está contida em infinitamente muitos A_n , o conjunto $\{\omega \in \Omega : |X_k(\omega) - X(\omega)| < \epsilon \text{ para todo } k \geq n\}$ é vazio $\forall n$. Em alternativa, considere o conjunto $S = \{\omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = 0\}$. Para qualquer ω , $X_n(\omega)$ não tem limite porque $X_n(\omega) = 1$ e $X_n(\omega) = 0$ ambos ocorrem para infinitamente muitos n . Assim o conjunto S é vazio. Isto não é convergência com probabilidade um; é convergência com probabilidade zero.

Teorema 7.52

Seja $\{X_n\}_{n=1}^{\infty}$ uma sequência de variáveis aleatórias positivas estritamente decrescente e suponhamos que $X_n \xrightarrow{P} 0$. Então $X_n \xrightarrow{q.c.} 0$.

Demonstração: Exercício. ■

Exemplo 7.42

Consideremos a sequência de variáveis aleatórias independentes $\{X_n\}_{n=1}^{\infty}$ definidas como

$$P(X_n = 0) = 1 - \frac{1}{n}, \quad P(X_n = 1) = \frac{1}{n}, \quad n = 1, 2, \dots$$

Então $E(|X_n - 0|^2) = E(|X_n|^2) = \frac{1}{n} \xrightarrow{n \rightarrow \infty} 0$, de maneira que $X_n \xrightarrow{2} 0$. Também

$$P(\{X_n = 0 \text{ para todo } m \leq n \leq n_0\}) = \prod_{n=m}^{n_0} \left(1 - \frac{1}{n}\right) = \frac{m-1}{n_0},$$

que diverge à zero quando $n_0 \rightarrow \infty$ para todos os valores de m . Portanto, X_n não converge a 0 com probabilidade 1.

Exemplo 7.43

Seja $\{X_n\}_{n=1}^{\infty}$ uma sequência de variáveis aleatórias tais que

$$P\left(X_n = \pm \frac{1}{n}\right) = \frac{1}{2}.$$

Implica que $\lim_{n \rightarrow \infty} E(|X_n|^r) = \lim_{n \rightarrow \infty} \frac{1}{n^r} = 0$, logo $X_n \xrightarrow{r} 0$. Para $j < k$, $|X_k| < |X_j|$, de maneira que

$$\{\omega \in \Omega : |X_k(\omega)| > \epsilon\} \subset \{\omega \in \Omega : |X_j(\omega)| > \epsilon\}$$

e disto segue que

$$\bigcup_{j=n}^{\infty} \{\omega \in \Omega : |X_j(\omega)| > \epsilon\} = \{\omega \in \Omega : |X_n(\omega)| > \epsilon\}.$$

Escolhendo $n > 1/\epsilon$, vemos que

$$P\left(\bigcup_{j=n}^{\infty} \{\omega \in \Omega : |X_j(\omega)| > \epsilon\}\right) = P(\{\omega \in \Omega : |X_k(\omega)| > \epsilon\}) \leq P\left(|X_n| > \frac{1}{n}\right),$$

dado que $\lim_{n \rightarrow \infty} P(|X_n| > 1/n) = 0$ e, pela relação em (7.41), implica que $X_n \xrightarrow{q.c.} 0$.

Exemplo 7.44

Definamos a sequência de variáveis aleatórias $\{X_n\}_{n=1}^{\infty}$ tais que

$$P(X_n = 0) = 1 - \frac{1}{n^r}, \quad P(X_n = n) = \frac{1}{n^r}, \quad r \geq 2, \quad n = 1, 2, \dots$$

Então $P(\{X_n = 0 \text{ para todo } m \leq n \leq n_0\}) = \prod_{n=m}^{n_0} \left(1 - \frac{1}{n^r}\right)$. Acontece que, quando $n_0 \rightarrow \infty$ o produto infinito converge para uma quantidade diferente de zero, que por si só converge para 1 quando $m \rightarrow \infty$. Deste modo $X_n \xrightarrow{q.c.} 0$. Contudo, $E(|X_n|^r) = 1$ e $X_n \not\xrightarrow{r} 0$ quando $n \rightarrow \infty$.

Teorema 7.53

Seja $\{X_n, Y_n\}_{n=1}^{\infty}$ uma sequência de pares de variáveis aleatórias. Então, se $|X_n - Y_n| \xrightarrow{P} 0$ e $Y_n \xrightarrow{D} Y$ temos por conclusão que $X_n \xrightarrow{D} Y$.

Demonstração: Seja x um ponto de continuidade de F_Y , a função de distribuição de Y , e $\epsilon > 0$. Então

$$\begin{aligned} P(X_n \leq x) &= P(Y_n \leq x + Y_n - X_n) \\ &= P(Y_n \leq x + Y_n - X_n; Y_n - X_n \leq \epsilon) + P(Y_n \leq x + Y_n - X_n; Y_n - X_n > \epsilon) \\ &\leq P(Y_n \leq x + \epsilon) + P(Y_n - X_n > \epsilon). \end{aligned}$$

Segue que $\lim_{n \rightarrow \infty} P(X_n \leq x) \leq \lim_{n \rightarrow \infty} P(Y_n \leq x + \epsilon)$, similarmente $\lim_{n \rightarrow \infty} P(X_n \leq x) \geq \lim_{n \rightarrow \infty} P(Y_n \leq x - \epsilon)$. Dado que $\epsilon > 0$ é arbitrário e x um ponto de continuidade de $P(Y \leq x)$, temos que o resultado é válido quando $\epsilon \rightarrow 0$. ■

Teorema 7.54

Seja $\{X_n\}_{n=1}^{\infty}$ uma sequência de variáveis aleatórias tal que $X_n \xrightarrow{P} X$. Então $X_n \xrightarrow{D} X$.

Demonstração: Suponhamos que $X_n \xrightarrow{P} X$ e seja x um ponto de continuidade de F_X . Queremos provar que $\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x)$.

Para todo $\epsilon > 0$, sabemos que $X_n \leq x$ implica que $X \leq x + \epsilon$ ou que $X - X_n > \epsilon$, pela convergência em probabilidade, logo

$$\{\omega \in \Omega : X_n(\omega) \leq x\} \subset \{\omega \in \Omega : X(\omega) \leq x + \epsilon\} \cup \{\omega \in \Omega : |X_n(\omega) - X(\omega)| > \epsilon\}.$$

Logo,

$$F_{X_n}(x) = P(X_n \leq x) \leq F_X(x + \epsilon) + P(|X_n - X| > \epsilon).$$

Por outro lado, $X \leq x - \epsilon$ implica que $X_n \leq x$ ou $X_n - X > \epsilon$, de modo que

$$F_X(x - \epsilon) \leq F_{X_n}(x) + P(|X_n - X| > \epsilon).$$

Destas duas desigualdades, temos

$$F_X(x - \epsilon) - P(|X_n - X| > \epsilon) \leq F_{X_n}(x) \leq F_X(x + \epsilon) + P(|X_n - X| > \epsilon)$$

para todo $\epsilon > 0$ e todo n .

$$\lim_{\substack{n \rightarrow \infty \\ \epsilon \rightarrow 0}} F_X(x - \epsilon) \leq \liminf_{n \rightarrow \infty} F_{X_n}(x) \leq \limsup_{n \rightarrow \infty} F_{X_n}(x) \leq \lim_{\substack{n \rightarrow \infty \\ \epsilon \rightarrow 0}} F_X(x + \epsilon)$$

e, portanto,

$$F_X(x) = \lim_{n \rightarrow \infty} F_{X_n}(x). \quad \blacksquare$$

Exemplo 7.45

Seja X uma variável aleatória e $X_n = X + Y_n$, onde $E(Y_n) = 1/n$ e $\text{Var}(Y_n) = \sigma^2/n$, onde $\sigma > 0$ é uma constante. Vamos mostrar que $X_n \xrightarrow{P} X$.

Primeiro notamos que, para todo $a, b \in \mathbb{R}$, temos $|a + b| \leq |a| + |b|$. Escolhendo $a = Y_n - E(Y_n)$ e $b = E(Y_n)$, obtemos

$$|Y_n| \leq |Y_n - E(Y_n)| + \frac{1}{n}.$$

Agora, para qualquer $\epsilon > 0$, temos

$$\begin{aligned} P(|X_n - X| \geq \epsilon) &= P(|Y_n| \geq \epsilon) \leq P\left(|Y_n - E(Y_n)| + \frac{1}{n} \geq \epsilon\right) \\ &= P\left(|Y_n - E(Y_n)| \geq \epsilon - \frac{1}{n}\right) \leq \frac{\sigma^2}{(\epsilon - 1/n)^2} \end{aligned}$$

isto último pela desigualdade de Chebyshev. Obtemos assim que

$$\lim_{n \rightarrow \infty} P(|X_n - X| \geq \epsilon) \leq \lim_{n \rightarrow \infty} \frac{\sigma^2}{n(\epsilon - \frac{1}{n})^2} = 0.$$

Portanto, concluímos que $X_n \xrightarrow{P} X$ e, pelo resultado no Teorema 7.54, $X_n \xrightarrow{D} X$.

Embora o recíproco do Teorema 7.54 não seja válido de maneira geral é válido na seguinte situação.

Corolário 7.55

Seja $\{X_n\}_{n=1}^{\infty}$ uma sequência de variáveis aleatórias tais que $X_n \xrightarrow{D} c$, então $X_n \xrightarrow{P} c$.

Demonstração: Dado que a variável aleatória limite é uma constante, a distribuição limite é

$$F_c(x) = \begin{cases} 1, & \text{se } x \geq c \\ 0, & \text{se } x < c \end{cases}$$

Observamos que c não é ponto de continuidade, então se $x \neq c$ segue, pela convergência em distribuição, que

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = \begin{cases} 1, & \text{se } x > c \\ 0, & \text{se } x < c \end{cases}$$

Para $\epsilon > 0$, temos que

$$\begin{aligned} P(|X_n - c| \leq \epsilon) &= P(c - \epsilon \leq X_n \leq c + \epsilon) \geq P(c - \epsilon < X \leq c + \epsilon) \\ &\geq F_{X_n}(c + \epsilon) - F_{X_n}(c - \epsilon), \end{aligned}$$

do qual obtemos que $\lim_{n \rightarrow \infty} F_{X_n}(c + \epsilon) - F_{X_n}(c - \epsilon) = 1$, o qual significa que

$$\lim_{n \rightarrow \infty} P(|X_n - c| > \epsilon) = 0, \quad \forall \epsilon > 0 \quad \text{e} \quad X_n \xrightarrow{P} c. \quad \blacksquare$$

Como dizemos, em geral convergência em distribuição não implica convergência em probabilidade. Vejamos isto no seguinte exemplo.

Exemplo 7.46

Sejam X, X_1, X_2, \dots variáveis aleatórias independentes com distribuição comum $N(0, 1/2)$.

Significa que $X_n \xrightarrow{D} X$, devido a todas as distribuições serem iguais.

Observamos que

$$|X_n - X| \sim N(0, 1) \quad \text{e} \quad P(|X_n - X| \geq \epsilon) = 2 - 2\Phi(\epsilon),$$

sendo Φ a notação da função de distribuição normal padrão. Logo, para $\epsilon > 0$,

$$\lim_{n \rightarrow \infty} P(|X_n - X| \geq \epsilon) \neq 0, \quad \text{portanto} \quad X_n \not\xrightarrow{P} X.$$

Teorema 7.56

Seja $\{X_n, Y_n\}$ uma sequência de pares de variáveis aleatórias e c uma constante. Então

(a) se $X_n \xrightarrow{D} X$ e $Y_n \xrightarrow{P} c$ temos que $X_n \pm Y_n \xrightarrow{D} X \pm c$,

(b) se $X_n \xrightarrow{D} X$ e $Y_n \xrightarrow{P} c$ temos que $\begin{cases} X_n Y_n \xrightarrow{D} cX, & \text{se } c \neq 0, \\ X_n Y_n \xrightarrow{P} 0, & \text{se } c = 0 \end{cases}$,

(c) se $X_n \xrightarrow{D} X$ e $Y_n \xrightarrow{P} c$ temos que $\frac{X_n}{Y_n} \xrightarrow{D} \frac{X}{c}$, se $c \neq 0$.

Demonstração: (a) $X_n \xrightarrow{D} X$, implica que $X_n \pm c \xrightarrow{D} X \pm c$. Também, $Y_n - c = (Y_n \pm X_n) \mp (X_n \pm c) \xrightarrow{P} 0$. Utilizando o Teorema 7.53 mostramos que

$$X_n \pm Y_n \xrightarrow{D} X \pm c.$$

(b) Primeiro vamos considerar o caso $c = 0$. Para qualquer número fixo $k > 0$,

$$\begin{aligned} P(|X_n Y_n| > \epsilon) &= P(|X_n Y_n| > \epsilon, |Y_n| \leq \epsilon/k) + P(|X_n Y_n| > \epsilon, |Y_n| > \epsilon/k) \\ &\leq P(|X_n| > k) + P(|Y_n| > \epsilon/k). \end{aligned}$$

Dado que $Y_n \xrightarrow{P} 0$ e $X_n \xrightarrow{D} X$, segue que

$$\lim_{n \rightarrow \infty} P(|X_n Y_n| > \epsilon) \leq P(|X| > k).$$

Desde k é arbitrário, podemos fazer $P(|X| > k)$ tão pequeno quanto desejamos, escolhendo k grande. Segue então que $X_n Y_n \xrightarrow{P} 0$. Agora, seja $c \neq 0$. Então

$$X_n Y_n - c X_n = X_n (Y_n - c)$$

e, desde que, $X_n \xrightarrow{D} X$, $Y_n \xrightarrow{P} c$ e $X_n (Y_n - c) \xrightarrow{P} 0$. Utilizando novamente o Teorema 7.53, temos o resultado

$$X_n Y_n \xrightarrow{D} cX.$$

(c) $Y_n \xrightarrow{P} c$ e $c \neq 0$, então $Y_n^{-1} \xrightarrow{P} c^{-1}$. Dado que $X_n \xrightarrow{D} X$ e $Y_n \xrightarrow{P} c$, implica $X_n Y_n^{-1} \xrightarrow{D} c^{-1} X$. ■

Exemplo 7.47

Seja $\{X_n\}_{n=1}^{\infty}$ uma sequência de variáveis aleatórias independentes com distribuição comum $N(0, 1)$. Vamos determinar a distribuição limite da variável aleatória

$$W_n = \sqrt{n} \frac{X_1 + X_2 + \cdots + X_n}{X_1^2 + X_2^2 + \cdots + X_n^2}.$$

Escolhemos

$$U_n = \frac{1}{\sqrt{n}} (X_1 + X_2 + \cdots + X_n) \quad \text{e} \quad V_n = \frac{1}{n} (X_1^2 + X_2^2 + \cdots + X_n^2),$$

assim $W_n = \frac{U_n}{V_n}$.

Pelo Exemplo 7.7, sabemos que $U_n \sim N(0, 1)$, logo $U_n \xrightarrow{D} Z$, sendo $Z \sim N(0, 1)$. Encontremos o limite em probabilidade da sequência de variáveis aleatórias V_n . Segundo o Exemplo 3.44, se $X_1 \sim N(0, 1)$ então $X_1^2 \sim \chi^2(1)$, conhecida como distribuição qui-quadrado ou Gama(1/2, 2).

7.5 Exercícios

Exercícios da Seção 7.1

1. Seja ϕ_X uma função característica de alguma variável aleatória X . Prove que $\varphi_Y(t) = e^{\phi_X(t)-1}$ é também uma função característica para alguma variável aleatória Y .
2. Sejam F e G duas funções de distribuição absolutamente contínuas, f e g as funções de densidade correspondentes, $\phi(t) = \int_{-\infty}^{\infty} e^{itx} f(x) dx$ e $\psi(t) = \int_{-\infty}^{\infty} e^{ity} g(y) dy$ as funções características respectivas. Prove que se cumpre a seguinte identidade

$$\int_{-\infty}^{\infty} e^{ity} \phi(t) g(t) dt = \int_{-\infty}^{\infty} \psi(x-y) f(x) dx, \quad (7.47)$$

chamada de identidade de Parseval²³.

3. Use a definição de função característica para encontrar a distribuição de $Y = aX^2$, se X tiver distribuição gaussiana com média zero e variância $\sigma^2 > 0$, a um número real positivo.
4. Use a definição de função característica para encontrar a distribuição de $Y = \sin(X)$, se X estiver uniformemente distribuído em $(-\pi/2, \pi/2)$.
5. Sejam U_1 e U_2 variáveis aleatórias independentes cada uma com distribuição uniforme no intervalo $(-1/2; 1/2)$. Seja $S = U_1 + U_2$. Encontre a função característica φ_S da soma aleatória. Qual é a densidade de S ?
6. Demonstrar o resultado apresentado no Exemplo 7.16.
7. Sejam X_1, X_2, \dots variáveis aleatórias independentes identicamente distribuídas com função característica $\exp(-|t|^\alpha)$, $0 < \alpha < 2$. Prove que $(X_1 + X_2 + \dots + X_n)/n^{1/\alpha}$ têm a mesma distribuição que X_1 .
8. Seja φ_X a função característica da variável aleatória X . Prove que se φ_X for uma função real, então as variáveis aleatórias X e $-X$ têm a mesma distribuição.
9. Prove que se $\lim_{t \downarrow 0} (\varphi_X(t) - 1)/t^2 = c$, c constante, então $E(X) = 0$ e $E(|X|^2) = -2c < \infty$.
10. Prove que se X_1, \dots, X_n são variáveis aleatórias independentes uniformemente distribuídas no intervalo $(-1, 1)$, então para $n \geq 2$, $X_1 + \dots + X_n$ têm por densidade

$$f(x) = \int_0^\infty \left(\frac{\sin(t)}{t} \right)^n \cos(tx) dt.$$

Embora não pareça claro na expressão acima, f é um polinômio em cada intervalo $(k, k+1)$, $k \in \mathbb{Z}$ e é zero em $[-n, n]^c$.

11. Mostre que a distribuição de uma variável aleatória limitada Z é infinitamente divisível se, e somente se, Z for constante.
Dica: Mostrar que $\text{Var}(Z) = 0$ e lembrar que $P(|Z| < M) = 1$, $M > 0$ constante.
12. Prove que $\varphi(t) = (1-b)/(1-be^{it})$, $0 < b < 1$, é uma função característica infinitamente divisível.
Dica: Use a forma canônica.
13. Mostre que o função de distribuição com densidade, $\beta^\alpha \Gamma^{-1}(\alpha) x^{\alpha-1} e^{-\beta x}$, $\alpha, \beta > 0$ definida em $(0, \infty)$ e 0, caso contrário, é infinitamente divisível.
14. Seja φ_X uma função característica infinitamente divisível. Prove que, para todo $c > 0$ a função φ_X^c é uma função característica.

²³Marc-Antoine Parseval (1755-1836), foi um matemático francês.

15. Sejam X e Y duas variáveis aleatórias. Encontre $\text{Var}(Z)$, onde $Z = X + iY$.
 16. Seja F_X uma função de distribuição para alguma variável aleatória X e seja

$$G(x) = 1 - F_X(-x^-),$$

onde x^- denota o limite pela esquerda.

- a) Prove que $F_X * G$ é uma função simétrica.
 b) Prove que $F_X * G = G * F_x$.
 17. Prove, utilizando função característica, que se $\lim_{n \rightarrow \infty} F_n = F$ e $\lim_{n \rightarrow \infty} G_n = G$, então $\lim_{n \rightarrow \infty} F_n * G_n = F * G$.
 18. Prove que se $\varphi(t)$ é uma função característica, então $|\varphi(t)|^2$ também é uma função característica.
 19. Seja $\{X_n\}_{n \geq 1}$ uma sequência de variáveis aleatórias com função de probabilidade dada por

$$P(X_n = 1) = 1/n \quad \text{e} \quad P(X_n = 0) = 1 - 1/n, \quad \forall n \geq 1.$$

Prove que $X_n = o_P(1)$.

20. Encontre a expressão da função característica se a distribuição é $\text{Cauchy}(\mu, \sigma)$.

Exercícios da Seção 7.2

1. Demonstrar a desigualdade de Hoeffding geral em (7.32).

Exercícios da Seção 7.3

1. Seja $\{X_n\}_{n \geq 1}$ uma sequência de variáveis aleatórias cada uma com função de probabilidade dada por $P(X_n = 1) = 1/n$ e $P(X_n = 0) = 1 - 1/n$, para todo $n \geq 1$. Prove que $X_n \xrightarrow{P} 0$.
 2. Seja $\{X_n\}_{n \geq 1}$ uma sequência de variáveis aleatórias com função de distribuição dada por

$$F_n(x) = \begin{cases} 0, & \text{caso } x < -n \\ \frac{x+n}{2n}, & \text{caso } -n \leq x < n \\ 0, & \text{caso } x \geq n \end{cases}$$

A sequência $\{F_n\}_{n \geq 1}$ converge a uma função de distribuição?

3. Prove o Corolário 7.8.
 4. Considere que as variáveis aleatórias X_1, X_2, \dots tenham por distribuição Normal padrão e $\{\bar{X}_n\}_{n \geq 1}$ uma sequência onde $\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k$. Seja F_n a função de distribuição de \bar{X}_n para $n = 1, 2, \dots$. Encontre $\lim_{n \rightarrow \infty} F_n(x)$. O limite da sequência $F_n(x)$ é uma função de distribuição?
 5. Seja $\{X_n\}_{n \geq 1}$ uma sequência de variáveis aleatórias tal que $X_n \xrightarrow{P} c$, onde c é uma constante. Prove que, se g for uma função contínua, então $g(X_n) \xrightarrow{P} g(c)$, quando $n \rightarrow \infty$.
 6. Prove que se $X_n \xrightarrow{P} X$, então $X_n - X \xrightarrow{P} 0$, quando $n \rightarrow \infty$.
 7. Prove que se $X_n \xrightarrow{P} X$, então $X_n^{-1} \xrightarrow{P} X^{-1}$.
 8. Seja Z_1, Z_2, \dots uma sequência de variáveis aleatórias e suponhamos que a distribuição de Z_n , para cada $n = 1, 2, \dots$, é dada por

$$P(Z_n = n^2) = \frac{1}{n} \quad \text{e} \quad P(Z_n = 0) = 1 - \frac{1}{n}.$$

Mostre que

$$\lim_{n \rightarrow \infty} E(Z_n) = \infty \quad \text{mas} \quad Z_n \xrightarrow{P} 0.$$

9. Para cada n , seja X_n uma variável aleatória não negativa com esperança finita μ_n . Prove que se $\lim_{n \rightarrow \infty} \mu_n = 0$, então $X_n \xrightarrow{P} 0$.
10. Suponha que $x_n \leq Y_n \leq Z_n$, sendo $\{X_n\}$, $\{Y_n\}$ e $\{Z_n\}$ seqüências de variáveis aleatórias tais que $X_n \xrightarrow{P} X$, $Y_n \xrightarrow{P} Y$ e $Z_n \xrightarrow{P} Z$. Se $\lim_{n \rightarrow \infty} E(X_n) = E(X)$ e $\lim_{n \rightarrow \infty} E(Z_n) = E(Z)$, mostre que $\lim_{n \rightarrow \infty} E(Y_n) = E(Y)$.
11. Seja X_n uma variável aleatória com função de densidade

$$f_n(x) = \frac{n}{\pi(1 + n^2 x^2)}, \quad n \geq 1,$$

e $x \in \mathbb{R}$. Com relação a quais modos de convergência X_n converge quando $n \rightarrow \infty$?

12. Sejam X_1, X_2, \dots variáveis aleatórias satisfazendo que $\text{Var}(X_n) < c$ para todo n e c alguma constante real positiva. Mostre que $\frac{1}{n} \sum_{k=1}^n (X_k - E(X_k)) \xrightarrow{P} 0$ se o coeficiente de correlação satisfaz uma das seguintes afirmações:
- (a) $\rho(X_i, X_j) \leq 0$ para todo $i \neq j$ ou
- (b) $\lim_{|i-j| \rightarrow \infty} \rho(X_i, X_j) \leq 0$.
13. Demonstre o Corolário 7.54
14. Considere que as variáveis aleatórias X_1, X_2, \dots sejam independentes com função de densidade comum

$$f(x) = \begin{cases} 0, & \text{caso } |x| \leq 2, \\ \frac{c}{x^2 \log(|x|)}, & \text{caso } |x| > 2 \end{cases},$$

onde c é uma constante. Prove que X_n não têm esperança mas $\frac{1}{n} \sum_{k=1}^n X_k \xrightarrow{P} 0$.

15. Seja $\{X_n\}_{n \geq 1}$ uma seqüência de variáveis aleatórias independentes igualmente distribuídas tais que $E(X_1) = \mu$ e $\text{Var}(X_1) = \sigma^2 < \infty$. Seja ainda $\{\bar{X}_n\}_{n \geq 1}$ a seqüência correspondente das médias aritméticas, ou seja

$$\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k.$$

Prove que $\{\bar{X}_n\}_{n \geq 1}$ converge em momentos e encontre o limite.

16. Seja $\{X_n\}_{n=1}^\infty$ uma seqüência de variáveis aleatórias. Prove que
- (a) $X_n \xrightarrow{q.c.} X$ se, e somente se, $P(\{X_n = X, \text{ para todos, mas finitamente muitos } n \in \mathbb{N}\}) = 1$.
- (b) $X_n \xrightarrow{P} X$ se, e somente se, $\lim_{n \rightarrow \infty} P(X_n \neq X) = 0$.
17. Sejam X_1, X_2, \dots variáveis aleatórias independentes identicamente distribuídas $Uniforme(0, 1)$. Definamos a seqüência

$$Y_n = \min\{X_1, X_2, \dots, X_n\}.$$

Prove os seguintes resultados de convergência:

- (i) $Y_n \xrightarrow{D} 0$,
- (ii) $Y_n \xrightarrow{P} 0$,
- (iii) $Y_n \xrightarrow{r} 0$, para todo $r \geq 1$,
- (iv) $Y_n \xrightarrow{q.c.} 0$.

18. Seja $\{X_n\}_{n \geq 1}$ uma sequência de variáveis aleatórias com função de probabilidade $P(X_n = \pm 1) = \frac{1}{2}$, $n = 1, 2, \dots$. Seja $Z_n = \frac{1}{2^n} \sum_{k=1}^n X_k$. Mostre que $Z_n \xrightarrow{D} Z$, onde $Z \sim \text{Uniforme}(-1, 1)$.
19. Prove que se X e Y forem independentes e se X e $X + Y$ tiverem mesma distribuição, então $Y = 0$ quase certamente, ou seja, $P(Y = 0) = 1$.

Capítulo 8

Teoremas Limites

Neste capítulo vamos investigar em quais situações sequências de variáveis aleatórias convergem a números e em quais situações a convergência é a variáveis aleatórias. Os grandes resultados que serão apresentados aqui podem ser reunidos em dois grandes grupos: diferentes leis de grandes números e teoremas do limite central. Estes são sem dúvida, a classe dos mais importantes teoremas e formam a espinha dorsal da teoria das probabilidades.

8.1 Lei dos Grandes Números

Nesta seção, vamos investigar respostas para a seguinte questão: existem sequências de constantes $\{a_n\}_{n=1}^{\infty}$ e $\{b_n\}_{n=1}^{\infty}$ tais que a sequência de variáveis aleatórias $\{b_n^{-1}(Y_n - a_n)\}_{n=1}^{\infty}$ convirja, quando $n \rightarrow \infty$, sendo $Y_n = g(X_1, \dots, X_n)$?

Uma primeira resposta é considerada na Seção 8.1.1 Lei Fraca dos Grandes Números e uma outra na Seção 8.1.2 Lei Forte dos Grandes Números. A diferença fundamental entre estas repostas é no tipo de convergência. A lei fraca considera convergência em probabilidade enquanto a lei forte refere-se à convergência quase certa.

Agora estamos em posição de provar nosso primeiro teorema fundamental da probabilidade. Vimos que uma maneira intuitiva de ver a probabilidade de um determinado resultado é a frequência com que esse resultado ocorre a longo prazo, quando o experimento é repetido várias vezes. Também definimos probabilidade matematicamente como um valor de uma função de distribuição para a variável aleatória que representa o experimento. A Lei dos Grandes Números, que é um teorema comprovado sobre o modelo matemático de probabilidade, mostra que esse modelo é consistente com a interpretação da frequência da probabilidade.

8.1.1 Lei Fraca dos Grandes Números

Se considerarmos $\{X_n\}_{n=1}^{\infty}$, uma sequência de variáveis aleatórias e $S_n = \sum_{k=1}^n X_k$ a lei fraca, considerada aqui, trata das condições necessárias nas sequências de números $\{a_n\}_{n=1}^{\infty}$ e $\{b_n\}_{n=1}^{\infty}$ para que a sequência de variáveis aleatórias $\{b_n^{-1}(S_n - a_n)\}_{n=1}^{\infty}$ convirja em probabilidade para zero, quando $n \rightarrow \infty$.

Definição 8.1 (*Lei Fraca dos Grandes Números*)

Seja $\{X_n\}_{n=1}^{\infty}$ uma sequência de variáveis aleatórias e $S_n = \sum_{k=1}^n X_k$, $n \geq 1$. Dizemos que a sequência $\{X_n\}_{n=1}^{\infty}$ obedece a Lei Fraca dos Grandes Números com relação à sequência de constantes $\{b_n\}_{n=1}^{\infty}$, $b_n > 0$ e $\lim_{n \rightarrow \infty} b_n = \infty$, se existir uma sequência de números reais $\{a_n\}_{n=1}^{\infty}$ tal que

$$\frac{S_n - a_n}{b_n} \xrightarrow{P} 0. \quad (8.1)$$

Os números a_n são chamados de constantes de centralização e os b_n são chamados de constantes de normalização.

Claro que os diversos exemplos na Seção 7.3.1 servem como situações particulares de sequências que satisfazem a Lei Fraca dos Grandes Números, interessa-nos então apresentar resultados mais gerais.

Teorema 8.1 (*Lei dos Grandes Números de Chebyshev*)

Seja $\{X_n\}_{n=1}^{\infty}$ uma sequência de variáveis aleatórias definidas no mesmo espaço de probabilidade tais que $E(X_k) = \mu_k$, $\text{Var}(X_k) = \sigma_k^2$ e $\text{Cov}(X_r, X_s) = 0$, $r \neq s$. Se

$$\lim_{n \rightarrow \infty} \sum_{k=1}^n \sigma_k^2 = +\infty$$

e escolhendo $a_n = \sum_{k=1}^n \mu_k$ e $b_n = \sum_{k=1}^n \sigma_k^2$, temos que

$$\frac{S_n - E(S_n)}{\text{Var}(S_n)} \xrightarrow{P} 0. \quad (8.2)$$

Demonstração: Utilizando a Desigualdade de Chebyshev temos que, para todo $\epsilon > 0$,

$$P\left(\left|S_n - \sum_{k=1}^n \mu_k\right| \geq \epsilon \sum_{k=1}^n \sigma_k^2\right) \leq \frac{E\left[\sum_{k=1}^n (X_k - \mu_k)\right]^2}{\epsilon^2 \left(\sum_{k=1}^n \sigma_k^2\right)^2} = \frac{1}{\epsilon^2 \sum_{k=1}^n \sigma_k^2} \xrightarrow{n \rightarrow \infty} 0.$$

Logo, concluímos que a convergência em (8.2) é válida. ■

Note que, neste caso, as variáveis aleatórias não precisam ter esperanças e variâncias iguais para ser válido este resultado, nem mesmo todas tiverem a mesma distribuição. Claro que,

caso a sequência $\{X_n\}_{n \geq 1}$ fosse composta de variáveis aleatórias independentes e igualmente distribuídas com média $E(X_1) = \mu$ e $\text{Var}(X_1) = \sigma^2 < \infty$, então $\text{Var}(S_n)/n^2 = n\sigma^2/n^2 = \sigma^2/n \rightarrow 0$, quando $n \rightarrow \infty$ e, portanto, $\{X_n\}_{n \geq 1}$ satisfaz a Lei dos Grandes Números de Chebyshev.

Corolário 8.2

Sejam $\{X_n\}_{n=1}^{\infty}$ variáveis aleatórias identicamente distribuídas tais que $E(X_1) = \mu$ e $\text{Var}(X_1) = \sigma^2$ finita, não correlacionadas a pares. Então, se satisfaz a Lei dos Grandes Números de Chebyshev.

Demonstração: Escolhendo $a_n = n\mu$ e $b_n = n\sigma^2$ se satisfaz o Teorema 8.1. ■

Uma situação particular deste corolário acontece caso $\sigma^2 = 1$. Escolhendo $a_n = n\mu$ e $b_n = n$ se satisfaz o Teorema 8.1 desde que $n\sigma^2/n^2 \rightarrow 0$, quando $n \rightarrow \infty$. Percebemos que se a sequência $\{X_n\}_{n=1}^{\infty}$ fosse de variáveis aleatórias não correlacionadas a pares, com mesma esperança $E(X_1) = \mu$ e tiverem variância finita, então $S_n/n \xrightarrow{P} \mu$.

Exemplo 8.1

Seja $\{X_n\}_{n=1}^{\infty}$ uma sequência de variáveis aleatórias independentes identicamente distribuídas Bernoulli(p). Então, $E(X_1) = p$, $\text{Var}(X_1) = p(1-p)$ e temos

$$\frac{S_n}{n} \xrightarrow{P} p, \quad n \rightarrow \infty.$$

Observe que S_n/n é a proporção de sucessos em n tentativas

Teorema 8.3

Seja $\{X_n\}_{n=1}^{\infty}$ uma sequência de variáveis aleatórias tais que $E(X_n) = \mu_n$ e seja

$$Y_n = \frac{1}{n} \sum_{k=1}^n (X_k - \mu_k).$$

Uma condição necessária e suficiente para que a sequência $\{X_n\}$ satisfaça a lei fraca de grandes números é que

$$\lim_{n \rightarrow \infty} E \left(\frac{Y_n^2}{1 + Y_n^2} \right) = 0. \quad (8.3)$$

Demonstração: Para quaisquer números reais positivos a, b , $a \geq b > 0$, temos

$$\left(\frac{a}{1+a} \right) \left(\frac{1+b}{b} \right) \geq 1. \quad (8.4)$$

Seja $A = \{\omega \in \Omega : |Y_n(\omega)| \geq \epsilon\}$. Então, se $\omega \in A$, implica que $|Y_n(\omega)|^2 \geq \epsilon^2 > 0$. Utilizando o resultado em (8.4) vemos que, se $\omega \in A$, então

$$\left(\frac{Y_n(\omega)^2}{1 + Y_n(\omega)^2} \right) \left(\frac{1 + \epsilon^2}{\epsilon^2} \right) \geq 1.$$

Observemos que se $\omega \in A$, então $\omega \in \left(\frac{Y_n(\omega)^2}{1 + Y_n(\omega)^2} \right) \left(\frac{1 + \epsilon^2}{\epsilon^2} \right) \geq 1$, do qual concluímos que

$$P(A) = P(\{\omega \in \Omega : |Y_n(\omega)| \geq \epsilon\}) = P\left(\omega \in \Omega : \left(\frac{Y_n(\omega)^2}{1 + Y_n(\omega)^2} \right) \left(\frac{1 + \epsilon^2}{\epsilon^2} \right) \geq 1\right).$$

Disto segue que

$$\begin{aligned} P(A) &= P\left(\left(\frac{Y_n^2}{1 + Y_n^2} \right) \geq \left(\frac{\epsilon^2}{1 + \epsilon^2} \right)\right) \\ &\leq \frac{E\left(\frac{Y_n^2}{1 + Y_n^2}\right)}{\frac{\epsilon^2}{1 + \epsilon^2}}, \end{aligned}$$

isto último pela desigualdade de Markov. Pelas condições do teorema $\lim_{n \rightarrow \infty} P(|Y_n| \geq \epsilon) = 0$, logo $Y_n \xrightarrow{P} 0$, quando $n \rightarrow \infty$.

Em contrapartida, mostraremos que, para todo $\epsilon > 0$

$$P(|Y_n| \geq \epsilon) \geq E\left(\frac{Y_n^2}{1 + Y_n^2}\right). \quad (8.5)$$

Vamos provar o resultado em (8.5) para o caso Y_n absolutamente contínua. Caso seja f a função de densidade de Y_n , então

$$\begin{aligned} E\left(\frac{Y_n^2}{1 + Y_n^2}\right) &= \int_{-\infty}^{\infty} \frac{y^2}{1 + y^2} f(y) dy = \int_{|y| > \epsilon} \frac{y^2}{1 + y^2} f(y) dy + \int_{|y| < \epsilon} \frac{y^2}{1 + y^2} f(y) dy \\ &\leq \int_{|y| > \epsilon} f(y) dy + \int_{|y| < \epsilon} \frac{y^2}{1 + y^2} f(y) dy \\ &= P(|Y_n| > \epsilon) + \int_{-\epsilon}^{\epsilon} \left(1 - \frac{1}{1 + y^2}\right) f(y) dy, \end{aligned}$$

aqui utilizamos o fato de $y^2/(1 + y^2) < 1$.

Observemos que $\left(1 - \frac{1}{1 + y^2}\right) < 1$ e que $\int_{-\epsilon}^{\epsilon} f(y) dy > 0$, então

$$E\left(\frac{Y_n^2}{1 + Y_n^2}\right) \leq P(|Y_n| > \epsilon),$$

e, a sequência $\{Y_n\}_{n \geq 1}$ converge em probabilidade a zero, ou seja,

$$\lim_{n \rightarrow \infty} P(|Y_n| > \epsilon) = 0; \quad \text{logo} \quad \lim_{n \rightarrow \infty} E\left(\frac{Y_n^2}{1 + Y_n^2}\right) = 0. \quad \blacksquare$$

A condição (8.3) não se aplica às variáveis individuais mas à sua soma, por isso o Teorema 8.3 é de utilidade limitada. No entanto, notamos que todas as leis fracas de grandes números obtidos até agora decorrem deste teorema.

Exemplo 8.2

Sejam X_1, X_2, \dots, X_n variáveis aleatórias conjuntamente normais tais que $E(X_k) = 0$, $E(X_k^2) = 1$ para todo k e $\rho(X_k, X_l) = \rho$, caso $|k - l| = 1$ e $\rho(X_k, X_l) = 0$, caso contrário. Então

$$S_n = \sum_{k=1}^n X_k \sim N(0, \sigma^2),$$

sendo $\sigma^2 = \text{Var}(S_n) = n + 2(n-1)\rho$.

$$E\left(\frac{Y_n^2}{1 + Y_n^2}\right) = E\left(\frac{S_n^2}{n^2 + S_n^2}\right) = \frac{2}{\sigma\sqrt{2\pi}} \int_0^\infty \frac{x^2}{n^2 + x^2} e^{-\frac{x^2}{2\sigma^2}} dx.$$

Fazemos agora a mudança de variáveis $y = x/\sigma$ e obtemos

$$\begin{aligned} E\left(\frac{Y_n^2}{1 + Y_n^2}\right) &= \frac{2}{\sqrt{2\pi}} \int_0^\infty \frac{y^2(n + 2(n-1)\rho)}{n^2 + y^2(n + 2(n-1)\rho)} e^{-\frac{y^2}{2}} dy \\ &\leq \frac{n + 2(n-1)\rho}{n^2} \frac{2}{\sqrt{2\pi}} \int_0^\infty y^2 e^{-\frac{y^2}{2}} dy \xrightarrow{n \rightarrow \infty} 0. \end{aligned}$$

Utilizando a afirmação do Teorema 8.3 segue que $\frac{1}{n}S_n \xrightarrow{P} 0$.

Seja $\{X_n\}_{n \geq 1}$ uma sequência de variáveis aleatórias arbitrárias e

$$S_n = \sum_{k=1}^n X_k. \quad (8.6)$$

Seja $\{X_n^c\}_{n \geq 1}$ a sequência das variáveis correspondentes truncadas, como definidas na Seção 6.1.1. Sejam também,

$$S_n^c = \sum_{k=1}^n X_k^c \quad \text{e} \quad \mu_n^c = \sum_{k=1}^n E(X_k^c), \quad (8.7)$$

as somas das primeiras n variáveis aleatórias truncadas e das n primeiras esperanças das variáveis truncadas, respectivamente.

Teorema 8.4

Para qualquer $\epsilon > 0$,

$$P(|S_n - \mu_n^c| > \epsilon) \leq P(|S_n^c - \mu_n^c| > \epsilon) + \sum_{k=1}^n P(|X_k| > c),$$

onde $\{X_n\}_{n \geq 1}$ é uma sequência de variáveis aleatórias arbitrárias, S_n como definida em (8.6), S_n^c e μ_n^c como definidas em (8.7).

Demonstração :

$$\begin{aligned}
 P\left(\{\omega \in \Omega : |S_n - \mu_n^c| > \epsilon\}\right) &= \\
 &= P\left(\{\omega \in \Omega : |S_n(\omega) - \mu_n^c| > \epsilon\} \cap \{\omega \in \Omega : |X_k(\omega)| \leq c, \text{ para } k = 1, 2, \dots, n\}\right) + \\
 &\quad P\left(\{\omega \in \Omega : |S_n(\omega) - \mu_n^c| > \epsilon\} \cap \{\omega \in \Omega : |X_k(\omega)| > c, \text{ para algum } k, 1 \leq k \leq n\}\right) \\
 &\leq P\left(\{\omega \in \Omega : |S_n^c(\omega) - \mu_n^c| > \epsilon\}\right) + P\left(\bigcup_{k=1}^n \{\omega \in \Omega : |X_k(\omega)| > c\}\right) \\
 &\leq P\left(\{\omega \in \Omega : |S_n^c(\omega) - \mu_n^c| > \epsilon\}\right) + \sum_{k=1}^n P\left(\{\omega \in \Omega : |X_k(\omega)| > c\}\right).
 \end{aligned}$$

■

Corolário 8.5

Seja $\{X_n\}_{n \geq 1}$ uma sequência de variáveis aleatórias identicamente distribuídas. Então,

$$P(|S_n - \mu_n^c| > \epsilon) \leq P(|S_n^c - \mu_n^c| > \epsilon) + nP(|X_1| > c).$$

para qualquer $\epsilon > 0$.

Demonstração : Exercício.

■

Se, em particular, as variáveis aleatórias na sequência $\{X_n\}_{n \geq 1}$ foram independentes, temos que

$$P(|S_n - \mu_n^c| > \epsilon) \leq \frac{n}{\epsilon^2} E(X_1^c)^2 + nP(|X_1| > c). \quad (8.8)$$

Sejam X_1, X_2, \dots variáveis aleatórias independentes *Cauchy*(0, 1) identicamente distribuídas. Como demonstrar se $\{n^{-1}S_n\}_{n \geq 1}$ satisfaz a Lei dos Grandes Números? esclarecemos que não sabemos demonstrar que não existem sequências de números $\{b_n\}_{n=1}^\infty$ e $\{a_n\}_{n=1}^\infty$ satisfazendo as condições da Definição 8.1. O que vamos fazer é demonstrar que, especificamente, a sequência $\{n^{-1}S_n\}_{n \geq 1}$ não satisfaz a Lei dos Grandes Números, utilizando a consequência do Corolário 8.5 mostrada na relação (8.8).

Observemos que nenhum dos resultados apresentados até o momento, nem mesmo os que serão apresentados depois, podem ser utilizados nesta situação. Lembremos que $E(X_1)$ não existe. Então, vamos verificar se

$$\lim_{n \rightarrow \infty} P(|n^{-1}S_n| > \epsilon) = 0, \quad (8.9)$$

para todo $\epsilon > 0$. Para isso, vamos considerar a distribuição truncada definida restringindo a distribuição Cauchy padrão ao intervalo $[-10^m, 10^m]$, sendo m algum inteiro positivo. Essa distribuição truncada tem todos os momentos, como vimos na Seção 6.1.1; contudo, para quase todos os fins práticos, ela se comporta como uma distribuição de Cauchy.

Exemplo 8.3

Seja $\{X_n\}_{n \geq 1}$ uma sequência de variáveis aleatórias independentes identicamente distribuídas $Cauchy(0, 1)$. Queremos ter ideia se a Lei dos Grandes Números vale nesta situação.

Lembremos que esta é uma situação emblemática, isto porque $E(X_1)$ não existe e, mesmo assim, queremos ver se o limite em (8.9) é válido. Nossa ideia é visualizar isso utilizando o resultado no Corolário 8.5, especificamente a consequência em (8.8). Para isso, observemos primeiro que $E(X_1^n) = 0$ e, portanto $\mu_n^n = 0$, escolhendo $c = n$. Significa que devemos demonstrar somente se $\lim_{n \rightarrow \infty} nP(|X_1| > n) = 0$.

Observemos na Figura 8.1 que o limite acima não se verifica, ou seja, vemos que

$$\lim_{n \rightarrow \infty} nP(|X_1| > n) \approx 0.6366198.$$

É claro então que não podemos utilizar os resultados anteriormente mencionados para concluir acerca do cumprimento da Lei dos Grandes Números pela sequência de variáveis Cauchy padrão. Teoricamente é o máximo que podemos fazer, realizando procedimento de simulação de amostras Cauchy padrão pode-se observar que o limite em (8.9) não é válido nesta situação.

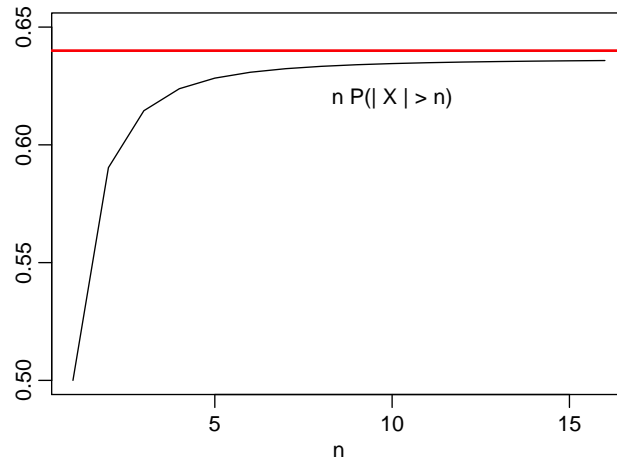


Figura 8.1: Comportamento de $nP(|X| > n)$, sendo $X \sim Cauchy(0, 1)$.

Teorema 8.6 (Lei dos Grandes Números de Khintchine)

Sejam X_1, X_2, \dots variáveis aleatórias definidas no mesmo espaço de probabilidade independentes e igualmente distribuídas tais que $E(X_k) = \mu < \infty$ existe. Então

$$\bar{X}_n - \mu \xrightarrow{P} 0, \quad (8.10)$$

quando $n \rightarrow \infty$.

Demonstração: Escolhendo $c = n$ em (8.8) e substituindo ϵ por ϵn , temos que

$$P(|S_n - \mu_n^c| > n\epsilon) \leq \frac{1}{n\epsilon^2} E(X_1^n)^2 + nP(|X_1| > n).$$

Observe que $E(|X_1|) < \infty$ implica que $\lim_{n \rightarrow \infty} nP(|X_1| > n) = 0$.

$$\begin{aligned} E(X_1^n)^2 &= 2 \int_0^n xP(|X_1| > x) dx \\ &= 2 \int_0^a xP(|X_1| > x) dx + 2 \int_a^n xP(|X_1| > x) dx, \end{aligned}$$

onde a é escolhido suficientemente grande para que $xP(|X_1| > x) < \frac{\delta}{2}$, para todo $x \geq a$, $\delta > 0$ arbitrário. Então,

$$E(X_1^n)^2 \leq c + \delta \int_a^n dx \leq c + n\delta,$$

onde c é uma constante. Disto segue que

$$\frac{1}{n\epsilon^2} E(X_1^n)^2 \leq \frac{c}{n\epsilon^2} + \frac{\delta}{\epsilon^2}$$

e, desde que, δ é arbitrário $\frac{1}{n\epsilon^2} E(X_1^n)^2$ pode ser feita arbitrariamente pequena para n suficientemente grande. ■

É interessante observar que a existência do segundo momento não é assumida, como no Teorema 8.1 de Chebyshev, mas as observações são consideradas independentes e igualmente distribuídas. A prova deste teorema, apresentada a seguir, segue apenas resultados elementares e foi desenvolvida por Markov.

Vimos que, se uma sequência de funções de distribuição cumulativas F_n converge pontual a um limite, a função de limitar a F não é necessariamente uma função de distribuição cumulativa. Para garantir que ele é, é necessário que as distribuições estão apertados. Da mesma forma, se uma sequência de funções características convergem para cada t , o limite não é, necessariamente, uma função característica de uma distribuição de probabilidades. No entanto, neste caso, o aperto de uma sequência traduz numa condição muito simples na função característica limitativa.

Exemplo 8.4

Sejam X_1, X_2, \dots variáveis aleatórias independentes ideticamente distribuídas tais que

$$E(|X_1^r|) < \infty,$$

para algum inteiro positivo r . Então,

$$\sum_{k=1}^n \frac{X_k^r}{n} \xrightarrow{P} E(X_1^r), \quad n \rightarrow \infty.$$

Portanto, se

Embora o caso de variáveis aleatórias identicamente distribuídas seja o mais frequente, observe que essa condição não é necessária, bastando que todas as variáveis tenham a mesma média

Exemplo 8.5

Sejam X_1, X_2, \dots variáveis aleatórias independentes ideticamente distribuídas com densidade comum

$$f(x) = \begin{cases} \frac{1+\delta}{x^{2+\delta}}, & \text{caso } x \geq 1 \\ 0, & \text{caso } x < 1 \end{cases},$$

sendo que $\delta > 0$. Então,

$$E(|X|) = (1+\delta) \int_1^\infty \frac{1}{x^{1+\delta}} dx = \frac{1+\delta}{\delta} < \infty,$$

e a lei dos grandes números se satisfaz, isto é, $\frac{1}{n}S_n \xrightarrow{P} \frac{1+\delta}{\delta}$, quando $n \rightarrow \infty$.

8.1.2 Lei Forte dos Grandes Números

Se considerarmos $\{X_n\}$ uma sequência de variáveis aleatórias e $S_n = \sum_{k=1}^n X_k$ a lei forte, considerada aqui, trata das condições necessárias nas sequências de números $\{a_n\}_{n \geq 1}$ e $\{b_n\}_{n \geq 1}$ para que a sequência de variáveis aleatórias $b_n^{-1}(S_n - a_n)$ convirja quase certamente para zero, quando $n \rightarrow \infty$.

Definição 8.2 (*Lei Forte dos Grandes Números*)

Seja $\{X_n\}_{n \geq 1}$ uma sequência de variáveis aleatórias e seja $S_n = \sum_{k=1}^n X_k$, $n = 1, 2, \dots$. Dizemos que a sequência obedece a Lei Fraca dos Grandes Números com relação à sequência de constantes $\{b_n\}_{n \geq 1}$, $b_n > 0$ e $\lim_{n \rightarrow \infty} b_n = \infty$ se existir uma sequência de números reais $\{a_n\}_{n \geq 1}$ tal que

$$\frac{(S_n - a_n)}{b_n} \xrightarrow{q.c.} 0, \quad (8.11)$$

quando $n \rightarrow \infty$. Os números a_n são chamadas de constantes de centralização e os b_n de constantes de normalização.

Lembremos o Teorema 2.31, chamado de Lema de Borel-Cantelli. Sejam $\{A_n\}_{n \geq 1}$ eventos no espaço de probabilidade (Ω, \mathcal{F}, P) . Vamos recordar que

$$\limsup_{n \rightarrow \infty} A_n = \lim_{n \rightarrow \infty} \bigcup_{k=n}^{\infty} A_k = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k.$$

Dizemos que a sequência $\{A_n\}_{n=1}^{\infty}$ tem limite se o evento $A \in \mathcal{F}$ é o limite da sequência, isto

é, se $\lim_{n \rightarrow \infty} A_n = A$. Então, pelo Teorema 2.22

$$\lim_{n \rightarrow \infty} P(A_n) = P(A).$$

Note que A , evento limite da sequência $\{A_n\}_{n \geq 1}$, é o evento em que infinitamente muitos dos A_n ocorrem. Tendo em conta o Teorema 7.46 e a observação em (7.41), nós temos que

$$X_n \xrightarrow{q.c.} 0, \quad \text{se, e somente se,} \quad \lim_{n \rightarrow \infty} P\left(\bigcup_n \{|X_n| > \epsilon\}\right) = 0, \quad \forall \epsilon > 0.$$

Como consequência, se a sequência $\{A_n\}_{n \geq 1}$ for de eventos independentes então, $P(A) = 0$ ou $P(A) = 1$.

Como uma simples aplicação do Lema de Borel-Cantelli, obtemos a seguir uma versão da Lei Forte dos Grandes Números.

Teorema 8.7 (*Lei Forte dos Grandes Números*)

Seja $\{X_n\}_{n \geq 1}$ uma sequência de variáveis aleatórias independentes identicamente distribuídas com média comum $E(X_1) = \mu$ e até quarto momento finito. Então,

$$P\left(\lim_{n \rightarrow \infty} \frac{S_n}{n} = \mu\right) = 1,$$

onde $S_n = \sum_{k=1}^n X_k$.

Demonstração: Observemos primeiro que

$$E\left(\sum_{k=1}^n (X_k - \mu)\right)^4 = n E(X_1 - \mu)^4 + 6 \binom{n}{2} \sigma^4 \leq cn^2.$$

Pela desigualdade de Markov obtemos

$$P\left(\left|\sum_{k=1}^n (X_k - \mu)\right| > n\epsilon\right) \leq \frac{E\left(\sum_{k=1}^n (X_k - \mu)\right)^4}{(n\epsilon)^4} \leq \frac{cn^2}{(n\epsilon)^4} = \frac{c}{\epsilon^4 n^2}.$$

Portanto,

$$\sum_{n=1}^{\infty} P(|S_n - \mu| > n\epsilon) \leq \infty,$$

e segue-se pelo Lema Borel-Cantelli que, com probabilidade 1, apenas se verificam finitamente muitos dos eventos $\left\{\omega \in \Omega : \left|\frac{S_n}{n} - \mu\right| > \epsilon\right\}$, ou seja, em que $P(A_\epsilon) = 0$ onde

$$A_\epsilon = \limsup_{k \rightarrow \infty} \left\{\left|\frac{S_n}{n} - \mu\right| > \epsilon\right\}.$$

Os conjuntos A_ϵ crescem, quando $\epsilon \rightarrow 0$, para o conjunto em que $\frac{S_n}{n} \rightarrow \mu$. Deixando $\epsilon \rightarrow 0$ através de um conjunto enumerável de valores, temos

$$P\left(\left|\frac{S_n}{n} - \mu\right| \rightarrow 0\right) = P\left(\bigcup_k A_{\frac{1}{k}}\right) = 0. \quad \blacksquare$$

Exemplo 8.6 (*Integração Monte Carlo*)

Seja $f : [0, 1] \rightarrow \mathbb{R}$ uma função e assumamos que queremos determinar o valor de sua integral

$$I = \int_0^1 f(x) dx$$

numericamente. Suponha que o computador gere os números X_1, X_2, \dots que podem ser considerados como números aleatórios independentes, distribuídos uniformemente em $[0, 1]$. Para $n \in \mathbb{N}$, defina o valor estimado

$$\hat{I}_n = \frac{1}{n} \sum_{k=1}^n f(X_k).$$

Assumindo que f seja integrável, a Lei dos Grandes Números nos disse que

$$P\left(\lim_{n \rightarrow \infty} \hat{I}_n = I\right) = 1,$$

ou $\hat{I}_n \xrightarrow{q.c.} I$, quando $n \rightarrow \infty$.

Devemos esclarecer que no exemplo anterior cumprem-se as exigências do Teorema 8.7. A sequência X_1, X_2, \dots é composta de variáveis independentes identicamente distribuídas $Uniforme(0, 1)$. Então $E(X_1) = 1/2$, $E(X_1^2) = 1/3$, $E(X_1^3) = 1/4$ e $E(X_1^4) = 1/5$, ou seja, os momentos até ordem quatro são finitos.

Note-se que o teorema mencionado (Teorema 8.7) não faz qualquer declaração sobre a velocidade da convergência. Isso significa que não temos controle sobre a quantidade

$$P(|\hat{I}_n - I| > \epsilon).$$

Para obter estimativas mais precisas para a integral, precisamos de informações adicionais; por exemplo, o valor $V = \int_0^1 f^2(x) dx - I^2$ se f^2 for integrável; pela desigualdade de Chebyshev. De fato, neste caso,

$$P\left(|\hat{I}_n - I| > \frac{\epsilon}{\sqrt{n}}\right) \leq \frac{V}{\epsilon^2}.$$

Portanto, o erro é, no máximo, da ordem $n^{-1/2}$. O Teorema do Limite Central mostrará que o erro é de fato exatamente desta ordem. Se f for uma função suave em algum sentido, então os procedimentos numéricos usuais produzem melhores ordens de convergência. Portanto, a simulação Monte Carlo só deve ser aplicada se todos os outros métodos falharem. Este é o caso em particular se $[0, 1]$ for substituído por $G \subset \mathbb{R}^d$ para d muito grande.

Corolário 8.8

Seja $\{X_n\}_{n \geq 1}$ uma sequência de variáveis aleatórias independentes identicamente distribuídas tais que $P(|X_n| < M) = 1$, para todo n , onde M é uma constante positiva. Então,

$$\frac{S_n}{n} \xrightarrow{q.c.} \mu,$$

quando $n \rightarrow \infty$.

Demonstração: Exercício. ■

Exemplo 8.7

Sejam X_1, X_2, \dots variáveis aleatórias independentes cada uma com mesma distribuição *Cauchy*(0, 1). Também, seja a sequência

$$Y_n = \begin{cases} X_n, & \text{se } |X_n| \leq M \\ 0, & \text{se } |X_n| > M \end{cases},$$

sendo $M > 0$ uma constante real. Então, $\frac{1}{n} \sum_{k=1}^n Y_k \xrightarrow{q.c.} 0$.

A função de distribuição empírica, definida a seguir, é uma aproximação amostral da função de distribuição que gerou as variáveis aleatórias. Vamos provar, no Teorema 8.9, que a distribuição empírica converge com probabilidade 1 para aquela distribuição subjacente.

Definição 8.3 (Função de distribuição empírica)

Sejam X_1, X_2, \dots variáveis aleatórias. A função $F_n : \mathbb{R} \rightarrow [0, 1]$, definida como

$$F_n(x) = \frac{1}{n} \sum_{k=1}^n \mathbf{1}_{(-\infty, x]}(X_k), \quad \forall x \in \mathbb{R},$$

é chamada de Função de distribuição empírica de X_1, \dots, X_n .

A Figura 8.2 dedica-se a mostrar o resultado do Teorema 8.9. O referido teorema nos disse que a função F_n converge quase certamente ou com probabilidade 1 à função de distribuição das variáveis aleatórias. A figura mostra como, conforme aumenta n , a distribuição empírica de X_1, \dots, X_{50} aproxima-se mais da função de distribuição *Poisson*(5) do que quando consideramos as variáveis X_1, \dots, X_{20} .

O objetivo do seguinte teorema, o qual deve-se a Glivenko¹ e Cantelli, é mostrar que

¹Valery Ivanovich Glivenko (1896-1940). Foi um era um matemático soviético russo. Ele trabalhou em fundamentos de matemática, análise real, teoria da probabilidade e estatística matemática.

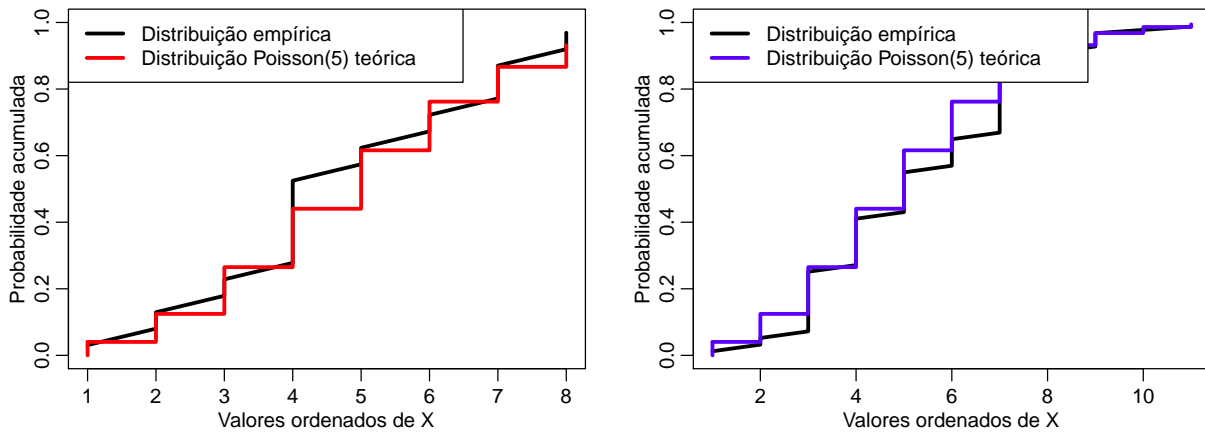


Figura 8.2: Gráfico à esquerda: mostra a função de distribuição empírica numa amostra das 20 observações $Poisson(5)$: 1 5 3 8 4 4 2 8 3 4 7 4 6 2 4 7 6 7 5 4. Gráfico à direita: mostrando a função de distribuição empírica das 50 observações $Poisson(5)$: 3 9 6 4 8 1 4 3 3 3 5 4 7 6 5 5 3 7 5 6 7 6 8 9 4 7 3 4 9 4 2 3 3 11 7 4 5 7 8 5 7 7 3 3 6 4 5 2 7 8.

a distribuição empírica converge à distribuição teórica, isto como consequência da Lei dos Grandes Números.

Teorema 8.9 (*Teorema de Glivenko-Cantelli*)

Sejam X_1, X_2, \dots variáveis aleatórias independentes identicamente distribuídas com função de distribuição F e seja F_n , $n \in \mathbb{N}$ a função de distribuição empírica. Então,

$$\limsup_{n \rightarrow \infty} \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| = 0, \quad \text{quase certamente.}$$

Demonstração: Consideremos $x \in \mathbb{R}$ fixo e seja $Y_n(x) = \mathbf{1}_{(-\infty, x]}(X_n)$ e $Z_n(x) = \mathbf{1}_{(-\infty, x)}(X_n)$, para $n \in \mathbb{N}$. Adicionalmente, definamos a limite à esquerda $F(x^-) = \lim_{y \uparrow x} F(y)$ e similarmente para F_n . Então, cada uma das famílias $\{Y_n(x)\}_{n \geq 1}$ e $\{Z_n(x)\}_{n \geq 1}$ são independentes. Além disso, $E(Y_n(x)) = P(X_n \leq x) = F(x)$ e $E(Z_n(x)) = P(X_n < x) = F(x^-)$. Pela Lei Forte dos Grandes Números, temos, portanto

$$F_n(x) = \frac{1}{n} \sum_{k=1}^n Y_k(x) \xrightarrow{q.c.} F(x), \quad \text{quando } n \rightarrow \infty$$

e

$$F_n(x^-) = \frac{1}{n} \sum_{k=1}^n Z_k(x) \xrightarrow{q.c.} F(x^-), \quad \text{quando } n \rightarrow \infty.$$

Formalmente, definamos $F(-\infty) = 0$ e $F(\infty) = 1$. Fixando algum $N \in \mathbb{N}$ e definamos

$$x_k = \inf\{x \in \mathbb{R} : F(x) \geq k/N\}, \quad k = 0, 1, \dots, N,$$

e

$$R_n = \max_{k=1, \dots, N-1} \{|F_n(x_k) - F(x_k)| + |F_n(x_k^-) - F(x_k^-)|\}.$$

Como mostrado acima, $R_n \xrightarrow{q.c.} 0$. Para $x \in (x_{k-1}, x_k)$ nós temos, por definição de x_k

$$F_n(x) \leq F_n(x_k^-) \leq F(x_k^-) + R_n \leq F(x) + R_n + \frac{1}{N}$$

e $F_n(x) \geq F_n(x_{k-1}) \geq F(x_{k-1}) - R_n \geq F(x) - R_n - \frac{1}{N}$. Por conseguinte

$$\limsup_{n \rightarrow \infty} \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \leq \frac{1}{N} + \limsup_{n \rightarrow \infty} R_n = \frac{1}{N}.$$

Tomando $N \rightarrow \infty$, a reivindicação segue. ■

Teorema 8.10

Seja $\{X_n\}_{n \geq 1}$ uma sequência de variáveis aleatórias independentes. Então, $X_n \xrightarrow{q.c.} 0$ se, e somente se,

$$\sum_{n=1}^{\infty} P(|X_n| > \epsilon) < \infty, \quad \forall \epsilon > 0.$$

Demonstração: Escrevendo $A_\epsilon = \{|X_n| > \epsilon\}$, vemos que a sequência $\{A_n\}$ é composta de eventos independentes. Desde que $X_n \xrightarrow{q.c.} 0$, $X_n \rightarrow 0$ no conjunto E^c , sendo que $P(E) = 0$. O ponto $\omega \in E^c$ pertence somente a um número finito de A_n . Segue então que

$$\limsup_{n \rightarrow \infty} A_n \subset E,$$

por isso, $P\left(\limsup_{n \rightarrow \infty} A_n\right) = 0$. Pelo Lema de Borel-Cantelli, Teorema 2.31, devemos ter

$$\sum_{n=1}^{\infty} P(A_n) < \infty, \text{ caso contrário, teríamos que } \sum_{n=1}^{\infty} P(A_n) = \infty \text{ e, então, } P\left(\limsup_{n \rightarrow \infty} A_n\right) = 1.$$

No sentido contrário, utilizamos o argumento do Teorema 8.7 e, para isso, definimos

$$A_{\frac{1}{k}} = \limsup_{n \rightarrow \infty} \{|X_k| > 1/k\}.$$
■

Definição 8.4 (*Sequências de variáveis aleatórias equivalentes na cauda*)

Duas sequências de variáveis aleatórias $\{X_n\}_{n \geq 1}$ e $\{Y_n\}_{n \geq 1}$ são ditas equivalentes na cauda se elas diferem quase certamente apenas por um número finito de termos; isto é, se para quase todos $\omega \in \Omega$ existe um $n(\omega)$ tal que, para $n > n(\omega)$, as sequências de números $\{X_n(\omega)\}_{n \geq 1}$ e $\{Y_n(\omega)\}_{n \geq 1}$ são as mesmas. Em símbolos, escrevemos $X_n \neq Y_n \xrightarrow{q.c.} 0$, quando $n \rightarrow \infty$.

Lembramos que, duas variáveis aleatórias X e Y são ditas equivalentes se $P(X \neq Y) = 0$; isto é, se para qualquer $\epsilon > 0$ existe um $\delta > 0$ tal que $P(|X - Y| > \epsilon) < \delta$.

Teorema 8.11

Cada uma das seguinte três condições implicam as outras dois:

(a) $\varphi'(0) = i\mu$.

(b) Quando $t \rightarrow \infty$,

$$t[1 - F(t) + F(t)] \rightarrow 0, \quad \int_{-t}^t x f(x) dx \rightarrow \mu. \quad (8.12)$$

(c) A média

$$\frac{1}{n}(X_1 + X_2 + \cdots + X_n)$$

tende em probabilidade a μ quando $n \rightarrow \infty$.

Demonstração: Definamos um par de novas variáveis aleatórias

$$\begin{aligned} Y_i &= X_i, & Z_i &= 0, & \text{se } |X_i| < \delta n, \\ Y_i &= 0, & Z_i &= X_i, & \text{se } |X_i| \geq \delta n, \end{aligned}$$

para $i = 1, \dots, n$ e $\delta > 0$ fixo. Então $X_i = Y_i + Z_i$. Seja $E(Y_i) = \mu_n$ para $i = 1, \dots, n$.

Dado que $E(X_i) = \mu$

$$|\mu_n - \mu| < \epsilon$$

para qualquer ϵ dado, se n é escolhido suficientemente grande. ■

8.2 Teorema do Limite Central

Diferentemente da Lei dos Grandes Números, a qual trata da convergência de sequências de variáveis aleatórias a números, o Teorema do Limite Central trata da convergência de sequências de variáveis aleatórias a uma variável aleatória.

Podemos realizar o seguinte agrupamento de resultados fundamentais da teoria das probabilidades. No primeiro grupo consideramos aqueles que tratam da lei dos Grandes Números. O segundo grupo fundamental das probabilidades são aqueles que tratam do Teorema do Limite Central. Estes teoremas dizem que se S_n é a soma de n variáveis aleatórias independentes a função de distribuição de S_n é bem aproximada por uma determinada função de densidade ou de probabilidade. Existem diversas versões destes teoremas, dentre elas, versões aplicáveis a sequências de variáveis aleatórias dependentes. Esses resultados embora importantes também, não serão abordados aqui. Dentre muitas referências, o leitor interessado pode consultar Billingsley (1961) ou Teicher & Chow (1978), para maiores detalhes.

Este teorema tem uma história interessante. A primeira versão deste teorema foi postulada pelo matemático francês Abraham de Moivre, que em um notável artigo publicado em 1733, usou a distribuição normal para aproximar a distribuição do número de caras resultantes de muitos lançamentos de uma moeda não viciada. Esse pensamento foi muito à frente de seu tempo, mas fora esquecido até que o famoso matemático francês Pierre Simon de Laplace resgatou-o da obscuridade em sua monumental obra *Théorie des Analytique probabilites*, que foi publicada em 1812. Laplace expandiu a descoberta de De Moivre e encontrou a aproximação da distribuição binomial a partir distribuição normal. Mas, como a descoberta de Moivre, a descoberta de Laplace recebeu pouca atenção naquela época.

Basicamente, Moivre e Laplace demonstraram que se X_1, X_2, \dots forem variáveis aleatórias independentes identicamente distribuídas tais que

$$P(X_1 = 1) = P(X_1 = -1) = \frac{1}{2} \quad \text{e} \quad S_n = \sum_{k=1}^n X_k,$$

então

$$\lim_{n \rightarrow \infty} P\left(a \leq \frac{S_n}{\sqrt{n}} \leq b\right) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx.$$

Somente no final do século XIX, que a importância do Teorema do Limite Central foi discernida, quando em 1901, o matemático russo Aleksandr Lyapunov definiu em termos gerais e provou exatamente como o teorema funcionava matematicamente. Hoje em dia, o Teorema do Limite Central é considerado o soberano não oficial da teoria da probabilidade.

A versão mais geral é conhecida como Teorema Central do Limite de Lindeberg², devido a ter sido ele quem o demonstrou com estas suposições em Lindeberg (1922). Uma excelente referência acerca da história deste teorema assim como também a história de diferentes versões pode ser consultada em Fischer (2010), algumas das quais serão consideradas aqui.

Primeiro vamos demonstrar a forma de geral do Teorema do Limite Central ou Teorema Central do Limite de Lindeberg, depois consideramos a versão de Lyapunov que envolve o terceiro momento.

Começamos apresentando o conceito de ordens de magnitude O e o .

²Jarl Waldemar Lindeberg (1876-1932). Foi um matemático finlandês conhecido pelo seu trabalho na demonstração da versão mais geral conhecida do Teorema do Limite Central.

8.2.1 Ordens de magnitude

A notação O e o é uma notação matemática que descreve o comportamento limitador de uma função quando o argumento tende para um determinado valor ou infinito, é uma família de notações assintótica. Esta notação caracteriza as funções de acordo com suas taxas de crescimento: diferentes funções com a mesma taxa de crescimento podem ser representadas usando a mesma notação O . As letras O e o são usadas porque a taxa de crescimento de uma função também é referida como a ordem da função.

Definição 8.5 (*Ordem de magnitude de O*)

Consideremos f e g duas funções e assumamos que $g(x) > 0$ para um x suficientemente grande. Dizemos que $f(x)$ é no máximo da ordem de $g(x)$ quando $x \rightarrow \infty$ e escrevemos $f(x) = O(g(x))$ quando $x \rightarrow \infty$ se existe um x_0 e uma constante $c > 0$ tal que

$$|f(x)| < cg(x), \quad \forall x \geq x_0.$$

Assim, $f(x) = O(g(x))$ significa que $|f(x)|/g(x)$ está limitado a x grande. Escrevemos $f(x) = O(1)$ para expressar o fato de f ser limitada para x grande.

Teorema 8.12

Sejam f e g duas funções e assumamos que $g(x) > 0$ para um x suficientemente grande. Então:

- (a) se $f(x) = O(g_1(x))$ e $f(x) = O(g_2(x))$, obtemos que $f(x) = O(g_1(x) + g_2(x))$,
- (b) se $\alpha > 0$ é uma constante, se $f(x) = O(g(x))$, obtemos que $f(x) = O(\alpha g(x))$,
- (c) se $f_1(x) = O(g_1(x))$ e $f_2(x) = O(g_2(x))$, obtemos que

$$f_1(x)f_2(x) = O(g_1(x)g_2(x)).$$

Demonstração: Exercício. ■

No uso típico a notação O é assintótica, ou seja, refere-se a um x muito grande. Neste contexto, a contribuição dos termos que crescem mais rapidamente acabará por tornar os outros irrelevantes. Como resultado, as seguintes regras de simplificação podem ser aplicadas:

- (a) se f é uma soma de vários termos, se houver um com maior taxa de crescimento, pode ser mantido, e todos os outros omitidos,
- (b) se f é um produto de vários fatores, os termos no produto que não dependem de x pode ser omitida, ou seja, qualquer constante pode ser omitida.

Exemplo 8.8

Por exemplo, seja $f(x) = 6x^4 - 2x^3 + 5$ e suponhamos que desejamos simplificar esta função, usando a notação O , para descrever sua taxa de crescimento à medida que x se aproxima do infinito. Esta função é a soma de três termos: $6x^4$, $-2x^3$ e 5 . Destes três termos, o de maior taxa de crescimento é o de maior expoente em função de x , ou seja, $6x^4$. Agora pode-se aplicar a segunda regra: $6x^4$ é um produto de 6 e x^4 em que o primeiro fator não depende de x , ou seja, é constante. Omitir este fator resulta na forma simplificada x^4 . Matematicamente, podemos escrever $f(x) = O(x^4)$. Pode-se confirmar este cálculo usando a definição formal: sejam $f(x) = 6x^4 - 2x^3 + 5$ e $g(x) = x^4$. Aplicando a definição formal, a afirmação de que $f(x) = O(x^4)$ é equivalente a $|f(x)| \leq cx^4$, para alguma escolha adequada de x_0 e $c > 0$ e para todos os $x \geq x_0$. Para provar isto, vamos escolher $x_0 = 1$ e $c = 13$. Então, para todos os $x \geq x_0$,

$$|6x^4 - 2x^3 + 5| \leq 6x^4 + 2x^3 + 5 \leq 6x^4 + 2x^4 + 5x^4 = 13x^4,$$

portanto $|6x^4 - 2x^3 + 5| \leq 13x^4$.

Lema 8.13

Seja λ um número real positivo. Então,

$$\sum_{k=1}^n k^\lambda = O(n^{\lambda+1}).$$

Demonstração: Suponhamos $\lambda > 0$. Vamos demonstrar que

$$\lim_{n \rightarrow \infty} \frac{1}{n^{\lambda+1}} \sum_{k=1}^n k^\lambda = \frac{1}{\lambda+1}.$$

Observemos que $\begin{cases} x^\lambda \leq k^\lambda & \text{se } k-1 \leq x \leq k \\ k^\lambda \leq x^\lambda & \text{se } k \leq x \leq k+1 \end{cases}$, então

$$\begin{aligned} \int_{k-1}^k x^\lambda dx &\leq \int_{k-1}^k k^\lambda dx = k^\lambda = \int_k^{k+1} k^\lambda dx \leq \int_k^{k+1} x^\lambda dx \\ \sum_{k=1}^n \int_{k-1}^k x^\lambda dx &\leq \sum_{k=1}^n k^\lambda \leq \sum_{k=1}^n \int_k^{k+1} x^\lambda dx \\ \int_0^n x^\lambda dx &= \frac{n^{\lambda+1}}{\lambda+1} \leq \sum_{k=1}^n k^\lambda \leq \frac{(n+1)^{\lambda+1}}{\lambda+1} - \frac{1}{\lambda+1}. \end{aligned}$$

Logo

$$\frac{1}{\lambda+1} \leq \frac{1}{n^{\lambda+1}} \sum_{k=1}^n k^\lambda \leq \frac{1}{\lambda+1} \left(\frac{n+1}{n} \right)^{\lambda+1}.$$

■

Definição 8.6 (*Ordem de magnitude de o*)

Sejam f e g funções ambas definidas e positivas para x grande. Dizemos que f é de uma ordem menor do que g quando $x \rightarrow \infty$, e escrevemos $f(x) = o(g(x))$ quando $x \rightarrow \infty$ se

$$\lim_{x \rightarrow \infty} \frac{f(x)}{g(x)} = 0.$$

Escrevemos $f(x) = o(1)$ para expressar o fato de que $\lim_{x \rightarrow \infty} f(x) = 0$. Os símbolos O e o também são usados se x tende a um valor finito.

Teorema 8.14

Sejam f e g duas funções e assumamos ambas positivas para um x suficientemente grande. Então:

(a) se $f(x) = o(g(x))$ implica que $f(x) = O(g(x))$,

(b) se $f_1(x) = O(g_1(x))$ e $f_2(x) = o(g_2(x))$, obtemos que

$$f_1(x)f_2(x) = o(g_1(x)g_2(x)).$$

Demonstração : Exercício. ■

Lema 8.15

Dizemos que $f(x) = o(x)$ se $\lim_{x \rightarrow 0} \frac{f(x)}{x} = 0$. Temos então que

$$\lim_{n \rightarrow \infty} \left(1 + \frac{a}{n} + o\left(\frac{1}{n}\right) \right)^n = e^a,$$

para cada real a .

Demonstração : Pela expansão de Taylor³, temos

$$f(x) = f(0) + xf'(\delta x) = f(0) + xf'(0) + (f'(\delta x) - f'(0))x, \quad 0 < \delta < 1.$$

³Expansão de Taylor (Spivak, 1994). Seja n um inteiro não negativo e seja $f^{(n+1)}$ a $(n+1)$ derivada da

Caso $f'(x)$ seja contínua em $x = 0$, então

$$f(x) = f(0) + xf'(0) + o(x).$$

Escolhendo $f(x) = \log(1+x)$, temos $f'(x) = (1+x)^{-1}$ a qual é contínua em $x = 0$, de maneira que

$$\log(1+x) = x + o(x).$$

Então,

$$\begin{aligned} n \log \left(1 + \frac{a}{n} + o\left(\frac{1}{n}\right) \right) &= n \left(\frac{a}{n} + o\left(\frac{1}{n}\right) + o\left(\frac{a}{n} + o\left(\frac{1}{n}\right)\right) \right) \\ &= a + no\left(\frac{1}{n}\right) = a + o(1). \end{aligned}$$

Disto segue que

$$\left(1 + \frac{a}{n} + o\left(\frac{1}{n}\right) \right)^n = e^{a+o(1)}$$

e, portanto, $\lim_{n \rightarrow \infty} \left(1 + \frac{a}{n} + o\left(\frac{1}{n}\right) \right)^n = e^a$. ■

Exemplo 8.9

Vamos mostrar alguns exemplos da notação O e o :

(a) Podemos escrever, pela expansão de Taylor, que $e^x = 1 + x + \frac{x^2}{2} + O(x^3)$, quando $x \approx 0$.

(b) Se $f(x) = 10 + 7x + 2x^2 + 6x^3$, então $f(x) = O(x^3)$.

(c) $\sum_{k=1}^n k = \frac{n(n+1)}{2} = \frac{n^2}{2} + \frac{n}{2}$, então $\sum_{k=1}^n k = O(n^2)$.

(d) $\sum_{k=1}^n k^2 = n(n+1)(2n+1)$, então $\sum_{k=1}^n k^2 = O(n^3)$.

função f . Se $f^{(n+1)}(x)$ é contínua em $|x-a| \leq h$, então

$$f(x) = f(a) + \sum_{k=1}^n \frac{(x-a)^k}{k!} f^{(k)}(a) + r_n,$$

onde o termo do resto r_n , o qual depende de n , a e h , é dado por

$$r_n = \frac{1}{n!} \int_a^x (x-t)^n f^{(n+1)}(t) dt.$$

Outras duas formas para o termo do resto:

(a) Resto de Lagrange: $r_n = f^{(n+1)}(x_0) \frac{(x-a)^{n+1}}{(n+1)!}$, para algum $x_0 \in (a, x)$,

(b) Resto de Cauchy: $r_n = f^{(n+1)}(x_0) \frac{(x-x_0)^n}{n!} (x-a)$, para algum $x_0 \in (a, x)$.

Mostraremos uma forma de demonstrarmos que uma transformação de variáveis aleatórias independentes converge em distribuição a uma variável aleatória que sempre pode utilizada mas, não é um procedimento prático. Resultados práticos os mostraremos na Seção 8.2.2.

Exemplo 8.10

Sejam X_1, X_2, \dots variáveis aleatórias independentes *Bernoulli*(θ). Sejam também

$$S_n = \sum_{k=1}^n X_k, \quad a_n = E(S_n) = n\theta \quad e \quad b_n = \sqrt{\text{Var}(S_n)} = \sqrt{n\theta(1-\theta)}.$$

Provemos que

$$Z_n = \frac{S_n - n\theta}{\sqrt{n\theta(1-\theta)}} \xrightarrow{D} Z,$$

sendo que $Z \sim N(0, 1)$.

Para isso encontremos a função característica de Z_n .

$$\begin{aligned} \varphi_{Z_n}(t) &= E(e^{itZ_n}) = E\left(\exp\left(it \frac{S_n - n\theta}{\sqrt{n\theta(1-\theta)}}\right)\right) \\ &= \prod_{k=1}^n E\left(it \frac{X_k - \theta}{\sqrt{n\theta(1-\theta)}}\right) \\ &= \exp\left(-it \frac{n\theta}{\sqrt{n\theta(1-\theta)}}\right) \left(1 - \theta + \theta \exp\left(\frac{it}{\sqrt{n\theta(1-\theta)}}\right)\right)^n \\ &= \left((1-\theta) \exp\left(-it \frac{\theta}{\sqrt{n\theta(1-\theta)}}\right) + \theta \exp\left(\frac{it}{\sqrt{n\theta(1-\theta)}}\right)\right)^n. \end{aligned}$$

Segue, pelo Lema 8.15, que $\lim_{n \rightarrow \infty} \varphi_{Z_n}(t) = \lim_{n \rightarrow \infty} \left(1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right)\right)^n = e^{-t^2/2}$. Desde que $e^{-t^2/2}$ é a função característica de uma variável $N(0, 1)$ temos, pelo teorema de continuidade, que

$$\lim_{n \rightarrow \infty} P\left(\frac{S_n - n\theta}{\sqrt{n\theta(1-\theta)}} \leq x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt, \quad \forall x \in \mathbb{R}.$$

A ideia neste exemplo e nos próximos exemplos é padronizar a soma de variáveis aleatórias. Com padronizar a soma de variáveis aleatórias queremos dizer que será diminuída a esperança da soma e dividimos pelo desvio padrão da soma. O interessante é que, para amostras grandes, a soma assim padronizada converge em distribuição à uma variável aleatória normal padrão.

Um detalhe interessante e importantíssimo é que o denominador seja o desvio padrão da soma, caso o denominador seja a variância da soma a convergência seria em probabilidade à uma constante, como vimos anteriormente.

Exemplo 8.11

Sejam X_1, X_2, \dots variáveis aleatórias independentes identicamente distribuídas $\chi^2(1)$ e $S_n = \sum_{k=1}^n X_k \sim \chi^2(n)$, $E(S_n) = n$ e $\text{Var}(S_n) = 2n$.

Definamos $Z_n = \frac{S_n - n}{\sqrt{2n}}$, queremos demonstrar que $Z_n \xrightarrow{D} Z \sim N(0, 1)$ e vamos usar a função característica de Z_n .

$$\begin{aligned}\varphi_{Z_n}(t) &= E(e^{itZ_n}) = \exp\left(-it\sqrt{\frac{n}{2}}\right) \left(1 - it\sqrt{\frac{1}{2n}}\right)^{-\frac{n}{2}}, \quad 2t < \sqrt{2n}, \\ &= \left(\exp\left(it\sqrt{\frac{2}{n}}\right) - it\sqrt{\frac{2}{n}}\exp\left(it\sqrt{\frac{2}{n}}\right)\right)^{-\frac{n}{2}}, \quad t < \sqrt{\frac{n}{2}}.\end{aligned}$$

Utilizando a expansão de Taylor, temos

$$\exp\left(it\sqrt{\frac{2}{n}}\right) = 1 + it\sqrt{\frac{2}{n}} - i\frac{t^2}{2}\left(\sqrt{\frac{2}{n}}\right)^2 + \frac{e^{\theta_n}}{6}\left(it\sqrt{\frac{2}{n}}\right)^3,$$

onde $0 < \theta_n < t/\sqrt{2/n}$. Segue então que

$$\varphi_{Z_n}(t) = \left(1 - \frac{t^2}{n} + \frac{\xi(n)}{n}\right)^{-\frac{n}{2}},$$

onde $\xi(n) = t^3\sqrt{\frac{2}{n}} - \left(\frac{it^3}{3}\sqrt{\frac{2}{n}} + \frac{2t^4}{3n}\right)e^{\theta_n}$ e observemos que $\lim_{n \rightarrow \infty} \xi(n) = 0$, para todo t fixo. Do Lema 8.15 temos que $\lim_{n \rightarrow \infty} \varphi_{Z_n}(t) = e^{-\frac{t^2}{2}}$, para todo $t \in \mathbb{R}$, do qual segue que $Z_n \xrightarrow{D} Z$, sendo que $Z \sim N(0, 1)$.

8.2.2 Teoremas limites

Os exemplos 8.10 e 8.11 sugerem que, se tomarmos uma sequência de variáveis aleatórias independentes identicamente distribuídas com variância finita, e tomarmos $a_n = E(S_n)$ e $b_n = \text{Var}(S_n)$, então $b_n^{-1/2}(S_n - a_n) \xrightarrow{D} Z$, onde $Z \sim N(0, 1)$. Este é o resultado do limite central, que nós agora provaremos. O leitor deve notar que em ambos os exemplos mencionados usamos mais do que apenas a existência de $E|X|^2$.

O principal resultado desta seção pode ser resumido aproximadamente ao efeito de que quanto menor a cauda de uma distribuição F , mais lisa é a sua função característica φ ; inversamente, quanto mais lisa F , melhor será o comportamento φ no infinito. A maioria das estimativas ligadas às funções características dependem de uma avaliação do erro cometido na aproximação de e^{it} por finitos muitos termos de sua expansão de Taylor. O próximo lema diz que este erro é dominado pelo primeiro termo omitido.

Lema 8.16

Para $n = 1, 2, \dots$ e $t > 0$ se satisfaz que

$$\left| e^{it} - 1 - \frac{it}{1!} - \dots - \frac{(it)^{n-1}}{(n-1)!} \right| \leq \frac{t^n}{n!}.$$

Demonstração: Vamos denotar a expressão dentro dos sinais de valor absoluto por $\rho_n(t)$. Então

$$\rho_1(t) = i \int_0^t e^{is} ds, \quad (8.13)$$

de onde, $|\rho_1(t)| \leq t$. Além disso, para $n > 1$, $\rho_n(t) = i \int_0^t \rho_{n-1}(s) ds$, e a expressão em (8.13) segue por indução. ■

A mesma prova mostra que quando o desenvolvimento de Taylor para $\sin(t)$ ou $\cos(t)$ é interrompido após finitos termos, o erro é do mesmo sinal, e menor em valor absoluto, do que o primeiro termo omitido. Por exemplo, o erro é do mesmo sinal, $1 - \cos(t) \leq t^2/2$.

Lema 8.17

Sejam $c_{n,k}$ números complexos tais que $\lim_{n \rightarrow \infty} \sum_{k=1}^n c_{n,k} = c$. Se

$$\lim_{n \rightarrow \infty} \max_{1 \leq k \leq n} |c_{n,k}| = 0 \quad (8.14)$$

e

$$\sum_{k=1}^n |c_{n,k}| \leq M < \infty, \quad (8.15)$$

onde M é uma constante que não depende de n . Então $\lim_{n \rightarrow \infty} \prod_{k=1}^n (1 + c_{n,k}) = e^c$.

Demonstração: Pela condição em (8.14) existe um número n_0 de maneira que $n \geq n_0$, então $|c_{n,k}| \leq \frac{1}{2}$ para todo k , de modo que $1 + c_{n,k} \neq 0$. Vamos considerar somente grandes valores de n a continuação e denotaremos por $\log(1 + c_{n,k})$ a determinação do logaritmo com um ângulo em $(-\pi, \pi]$. Assim $\log(1 + c_{n,k}) = c_{n,k} + \Lambda |c_{n,k}|^2$, onde Λ é um número complexo que depende de várias variáveis, mas limitado por uma constante absoluta que não depende de nada, e não é necessariamente o mesmo em cada aspecto. No caso em apreço, temos de

fato

$$\begin{aligned} |\log(1 + c_{n,k}) - c_{n,k}| &= \left| \sum_{m=2}^{\infty} \frac{(-1)^{m-1}}{m} c_{n,k}^m \right| \leq \sum_{m=2}^{\infty} \frac{|c_{n,k}|^m}{m} \\ &\leq \frac{|c_{n,k}|^2}{2} \sum_{m=1}^{\infty} \left(\frac{1}{2}\right)^{m-2} = |c_{n,k}|^2 \leq 1, \end{aligned}$$

para que a constante absoluta acima referida possa ser considerada como 1. Desta forma,

$$\sum_{k=1}^{k_n} \log(1 + c_{n,k}) = \sum_{k=1}^{k_n} c_{n,k} + \Lambda \sum_{k=1}^{k_n} |c_{n,k}|^2.$$

Este Λ não é o mesmo de antes, mas limitado pelo mesmo 1. Decorre das alíneas (8.14)

e (8.15) que $\lim_{n \rightarrow \infty} \sum_{k=1}^{k_n} |c_{n,k}|^2 \leq \lim_{n \rightarrow \infty} \max_{1 \leq k \leq n} |c_{n,k}| \sum_{k=1}^{k_n} |c_{n,k}| \leq M \lim_{n \rightarrow \infty} \max_{1 \leq k \leq n} |c_{n,k}| = 0$, e,

consequentemente, $\lim_{n \rightarrow \infty} \sum_{k=1}^{k_n} \log(1 + c_{n,k}) = c$. ■

Teorema 8.18 (Teorema do Limite Central de Lindeberg)

Sejam X_1, X_2, \dots variáveis aleatórias independentes não degeneradas com funções de distribuição respectivas F_1, F_2, \dots . Assumiremos que $E(X_k) = \mu_k$, $\text{Var}(X_k) = \sigma_k^2$ e definiremos $\sigma_{S_n}^2 = \sigma_1^2 + \sigma_2^2 + \dots + \sigma_n^2$. Se F_k é absolutamente contínua com função de densidade f_k , assumiremos que a relação

$$\lim_{n \rightarrow \infty} \frac{1}{\sigma_{S_n}^2} \sum_{k=1}^n \int_{|x - \mu_k| > \epsilon S_n} (x - \mu_k)^2 f_k(x) dx = 0, \quad (8.16)$$

é válida $\forall \epsilon > 0$. Se X_k é do tipo discreta com probabilidade positiva p_{kl} em x_{kl} , $l = 1, 2, \dots$, assumiremos que a relação

$$\lim_{n \rightarrow \infty} \frac{1}{\sigma_{S_n}^2} \sum_{k=1}^n \sum_{|x_{kl} - \mu_k| > \epsilon S_n} (x_{kl} - \mu_k)^2 p_{kl} = 0, \quad (8.17)$$

é válida $\forall \epsilon > 0$. Então, a distribuição da soma padronizada

$$S_n^* = \frac{S_n - \sum_{k=1}^n \mu_k}{\sigma_{S_n}} = \frac{\sum_{k=1}^n (X_k - \mu_k)}{\sigma_{S_n}} \quad (8.18)$$

converge em distribuição à uma variável aleatória normal padrão, isto é,

$$\frac{\sum_{k=1}^n (X_k - \mu_k)}{\sigma_{S_n}} \xrightarrow{D} Z, \quad (8.19)$$

onde $Z \sim N(0, 1)$.

Demonstração: Sem perda de generalidade vamos supor que $\mu_k = 0$ e $\sigma_k^2 = 1$. Mostraremos que a sequência de funções características das somas parciais padronizadas convergem para a função característica correspondente à distribuição Normal padrão. Seja

$$Z_n = \frac{S_n}{\sqrt{n}},$$

queremos demonstrar então que, para todo $t \in \mathbb{R}$,

$$\lim_{n \rightarrow \infty} \varphi_{S_n}(t) = \lim_{n \rightarrow \infty} \prod_{k=1}^n \mathbb{E}(e^{itX_k/\sqrt{n}}) = e^{-t^2/2}. \quad (8.20)$$

Utilizaremos duas versões da Expansão de Taylos aplicadas à função e^{itx} e o Lema 8.16. Fixemos $t \in \mathbb{R}$, então

$$\begin{aligned} (i) \quad e^{itx} &= 1 + itx + \theta_1(x) \frac{t^2 x^2}{2}, \quad \text{sendo } |\theta_1(x)| \leq 1 \\ (ii) \quad e^{itx} &= 1 + itx - \frac{t^2 x^2}{2} + \theta_2(x) \frac{t^3 x^3}{6}, \quad \text{sendo } |\theta_1(x)| \leq 1. \end{aligned}$$

Seja $\epsilon > 0$. Usando a expansão em (i) para $|x| > \epsilon$ e a segunda, em (ii), para $|x| \leq \epsilon$, podemos escrever e^{itx} da seguinte forma

$$\begin{aligned} e^{itx} &= 1 + itx + r_\epsilon(x), \\ \text{onde } r_\epsilon(x) &= \begin{cases} (1 + \theta_1(x)) \frac{t^2 x^2}{2}, & \text{caso } |x| > \epsilon \\ \theta_2(x) \frac{t^3 x^3}{6}, & \text{caso } |x| \leq \epsilon \end{cases}. \text{ Portanto,} \\ \mathbb{E}(e^{itX_k/\sqrt{n}}) &= \int_{-\infty}^{\infty} e^{itx/\sqrt{n}} f_k(x) dx = \\ &= \int_{-\infty}^{\infty} \left(1 + it \frac{x}{\sqrt{n}} - \frac{t^2}{2} \left(\frac{x}{\sqrt{n}} \right)^2 + r_\epsilon \left(\frac{x}{\sqrt{n}} \right) \right) f_k(x) dx \\ &= 1 + it \mathbb{E} \left(\frac{X_k}{\sqrt{n}} \right) - \frac{t^2}{2} \mathbb{E} \left(\frac{X_k}{\sqrt{n}} \right)^2 + \\ &\quad + \frac{t^2}{2} \int_{|x| > \epsilon\sqrt{n}} \left(1 + \theta_1 \left(\frac{x}{\sqrt{n}} \right) \right) \left(\frac{x}{\sqrt{n}} \right)^2 f_k(x) dx + \\ &\quad + \frac{t^3}{6} \int_{|x| \leq \epsilon\sqrt{n}} \theta_2 \left(\frac{x}{\sqrt{n}} \right) \left(\frac{x}{\sqrt{n}} \right)^3 f_k(x) dx, \end{aligned}$$

isto último pela linearidade da esperança e f_k a função de densidade de X_k . Então,

$$\mathbb{E} \left(it \frac{X}{\sqrt{n}} \right) = 1 - \frac{t^2}{2n} + r_{n,k},$$

onde $r_{n,k}$ satisfaz que

$$\begin{aligned} |r_{n,k}| &\leq t^2 \int_{|x| > \epsilon\sqrt{n}} \left(\frac{x}{\sqrt{n}} \right)^2 f_k(x) dx + \frac{|t|^3}{6} \int_{|x| \leq \epsilon\sqrt{n}} \epsilon \left(\frac{x}{\sqrt{n}} \right)^2 f_k(x) dx \\ &\leq \frac{t^2}{n} \int_{|x| > \epsilon\sqrt{n}} x^2 f_k(x) dx + \frac{\epsilon |t|^3}{6n} \int_{-\infty}^{\infty} x^2 f_k(x) dx. \end{aligned}$$

Temos que $\sum_{k=1}^n |r_{n,k}| \leq \frac{t^2}{n} \sum_{k=1}^n \int_{|x| > \epsilon \sqrt{n}} x^2 f_k(x) dx + \frac{\epsilon |t|^3}{6}$. Pela condição de lindeberg, a primeira parcela do termo à direita tende a zero quando $n \rightarrow \infty$. Logo, para n suficientemente grande, $\sum_{k=1}^n |r_{n,k}| \leq \frac{\epsilon |t|^3}{6}$.

Escolhemos então uma sequência de ϵ 's que converge a zero, por exemplo $\epsilon = 1/m$. Neste caso, existe n_m tal que, para $n > n_m$, $\sum_{k=1}^n |r_{n,k}| \leq \frac{|t|^3}{3m}$ e, portanto $\lim_{n \rightarrow \infty} \sum_{k=1}^n |r_{n,k}| = 0$. Substituindo em (8.20), vemos que

$$\varphi_{S_n}(t) = \prod_{k=1}^n \left(1 - \frac{t^2}{2n} + r_{n,k}\right).$$

Vamos utilizar agora o Lema 8.17, nosso caso $c_{n,k} = -\frac{t^2}{2n} + r_{n,k}$ e $c = -\frac{t^2}{2}$. Então,

$$\lim_{n \rightarrow \infty} \sum_{k=1}^n |c_{n,k}| \leq \lim_{n \rightarrow \infty} \left(\frac{t^2}{2} + \sum_{k=1}^n |r_{n,k}| \right) = \frac{t^2}{2},$$

logo $\sum_{k=1}^n |c_{n,k}|$ é uniformemente limitado, ou seja, existe um $M < \infty$ tal que, para todo n , $\sum_{k=1}^n |c_{n,k}| < M$. Para aplicarmos o Lema 8.17 basta verificarmos a condição sobre o máximo,

$$\lim_{n \rightarrow \infty} \max_{1 \leq k \leq n} |c_{n,k}| \leq \lim_{n \rightarrow \infty} \max_{1 \leq k \leq n} \frac{t^2}{2n} + \lim_{n \rightarrow \infty} \max_{1 \leq k \leq n} |r_{n,k}| = 0. \quad \blacksquare$$

Exemplo 8.12

Sejam X_1, X_2, \dots variáveis aleatórias independentes tais que $X_k \sim \text{Uniforme}(-k, k)$, $k = 1, 2, \dots$ e é claro que não são identicamente distribuídas. Pelo desenvolvimento do Exemplo 4.39, obtemos que

$$E(X_k) = 0 \quad \text{e} \quad \text{Var}(X_k) = \frac{k^2}{3}.$$

Então $\sigma_{S_n}^2 = \sum_{k=1}^n \frac{k^2}{3} = \frac{n(n+1)(2n+1)}{18}$, sendo $S_n = \sum_{k=1}^n X_k$. Verifiquemos se a Condição de Lindeberg, limite em (8.16), é satisfeita.

Para a k -ésima variável na sequência temos

$$\int_{|x| > \epsilon \sigma_{S_n}} x^2 f_k(x) dx = \frac{1}{2k} \int_{\{x \in \mathbb{R} : |x| > \epsilon \sigma_{S_n}\} \cap \{x \in \mathbb{R} : |x| < k\}} x^2 dx.$$

Pelo Lema 8.13, sabemos que $\sigma_{S_n} = O(n^{3/2})$. Observemos que $k \leq n$ sempre e, se

$\epsilon\sigma_{S_n} > n$, a integral acima à direita é zero. Então,

$$\begin{aligned} \frac{1}{\sigma_{S_n}^2} \sum_{k=1}^n \int_{|x| > \epsilon\sigma_{S_n}} x^2 f_k(x) dx &\leq \frac{1}{\sigma_{S_n}^2} \sum_{k=1}^n \frac{1}{2k} \int_{\{x \in \mathbb{R} : |x| > \epsilon\sigma_{S_n}\} \cap \{x \in \mathbb{R} : |x| < n\}} n^2 dx \\ &\leq \frac{n^2}{\sigma_{S_n}^2} \sum_{k=1}^n P(|X_k| > \epsilon\sigma_{S_n}) \leq \frac{n^2}{\sigma_{S_n}^2} \sum_{k=1}^n \frac{\text{Var}(X_k)}{\epsilon^2 \sigma_{S_n}^2}, \end{aligned}$$

disto, obtemos que

$$\lim_{n \rightarrow \infty} \frac{1}{\sigma_{S_n}^2} \sum_{k=1}^n \int_{|x| > \epsilon\sigma_{S_n}} x^2 f_k(x) dx \leq \lim_{n \rightarrow \infty} \frac{n^2}{\sigma_{S_n}^2} = 0,$$

e, portanto, a Condição de Lindeberg é satisfeita.

Um comentário importante acerca do Exemplo 8.12. Suponhamos que X_1, X_2, \dots sejam variáveis aleatórias independentes tais que $X_k \sim \text{Uniforme}(-a_k, a_k)$, $k = 1, 2, \dots$. No referido exemplo $a_k = k$ e, $\lim_{n \rightarrow \infty} \sum_{k=1}^n a_k^2 = \infty$. Disto concluímos que a Condição de Lindeberg é satisfeita nestas circunstâncias sempre que $\lim_{n \rightarrow \infty} \sum_{k=1}^n a_k^2 = \infty$. Consideremos agora a situação contrária, ou seja, caso $\lim_{n \rightarrow \infty} \sum_{k=1}^n a_k^2 < \infty$, a Condição de Lindeberg continuará sendo satisfeita?

Primeiro vemos que se $\lim_{n \rightarrow \infty} \sigma_{S_n}^2 = a^2 < \infty$ então, para k fixo, podemos encontrar ϵ_k de maneira que $\epsilon_k a < a_k$ e $P(|X_k| > \epsilon_k \sigma_{S_n}) \geq P(|X_k| > \epsilon_k a) > 0$. Para $n \geq k$, temos

$$\begin{aligned} \frac{1}{\sigma_{S_n}^2} \sum_{l=1}^n \int_{|x| > \epsilon_k \sigma_{S_n}} x^2 f_l(x) dx &\geq \frac{\epsilon_k^2 \sigma_{S_n}^2}{\sigma_{S_n}^2} \sum_{l=1}^n P(|X_l| > \epsilon_k \sigma_{S_n}) \\ &\geq \epsilon_k P(|X_k| > \epsilon_k \sigma_{S_n}) > 0, \end{aligned}$$

implicando que a Condição de Lindeberg não é satisfeita.

De fato, sejam X_1, X_2, \dots variáveis aleatórias independentes tais que existe uma constante a satisfazendo $P(|X_n| \leq a) = 1$, para todo n , as Condição de Lindeberg (8.16) e (8.39) são satisfeitas desde que $\lim_{n \rightarrow \infty} \sigma_{S_n}^2 = \infty$.

Para demonstrarmos esta afirmação suponhamos que $\lim_{n \rightarrow \infty} \sigma_{S_n}^2 = \infty$. Desde que as variáveis X_k 's são uniformemente limitadas, então as variáveis $X_k - E(X_k)$ também são limitadas. Segue que, para todo $\epsilon > 0$ podemos encontrar um N_ϵ tal que, para $n \geq N_\epsilon$, $P(|X_k - E(X_k)| < \epsilon\sigma_{S_n}, k = 1, 2, \dots, n) = 1$. As Condições de Lindeberg seguem imediatamente.

O inverso também é válido, se $\lim_{n \rightarrow \infty} \sigma_{S_n}^2 < \infty$ e a Condição de Lindeberg se satisfaz, existe uma constante $a < \infty$ de maneira que $\lim_{n \rightarrow \infty} \sigma_{S_n}^2 = a^2$. Para qualquer k fixo, podemos encontrar um $\epsilon > 0$, de maneira que $P(|X_k - E(X_k)| > \epsilon a) > 0$. Então, para $n \geq k$,

$$\begin{aligned} \frac{1}{\sigma_{S_n}^2} \sum_{l=1}^n \int_{|x - E(X_l)| > \epsilon\sigma_{S_n}} (x - E(X_l))^2 f_l(x) dx &\geq \epsilon^2 \sum_{l=1}^n P(|X_l - E(X_l)| > \epsilon\sigma_{S_n}) \\ &\geq \epsilon^2 P(|X_k - E(X_k)| > \epsilon a) > 0, \end{aligned}$$

implicando que a Condição de Lindeberg não é satisfeita. Esta contradição mostra que $\lim_{n \rightarrow \infty} \sigma_{S_n}^2 = \infty$ é também uma condição necessária; ou seja, para uma sequência de variáveis aleatórias independentes uniformemente limitadas, $\lim_{n \rightarrow \infty} \sigma_{S_n}^2 = \infty$ é uma condição necessária e suficiente para que o Teorema do Limite Central se satisfaça.

Exemplo 8.13

Sejam X_1, X_2, \dots variáveis aleatórias independentes tais que $E(|X_k|)^{2+\delta} < \infty$, para algum $\delta \geq 0$ e

$$E(|X_1|)^{2+\delta} + E(|X_2|)^{2+\delta} + \dots + E(|X_n|)^{2+\delta} = o(\sigma_{S_n}^{2+\delta}).$$

Então, a Condição de Lindeberg é satisfeita e o Teorema do Limite Central é válido. Para verificar isso, vemos que

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{\sigma_{S_n}^2} \sum_{k=1}^n \int_{|x - E(X_k)| > \epsilon \sigma_{S_n}} (x - E(X_k))^2 f_k(x) dx \\ \leq \lim_{n \rightarrow \infty} \frac{1}{\epsilon^\delta \sigma_{S_n}^{2+\delta}} \sum_{k=1}^n \int_{-\infty}^{\infty} |x - E(X_k)|^{2+\delta} f_k(x) dx \\ = \lim_{n \rightarrow \infty} \frac{\sum_{k=1}^n E(|X_k|)^{2+\delta}}{\epsilon^\delta \sigma_{S_n}^{2+\delta}} = 0. \end{aligned}$$

Um argumento semelhante é aplicável no caso discreto.

Se as variáveis aleatórias na sequência $\{X_n\}_{n \geq 1}$ forem independentes é bem possível que o Teorema do Limite Central se aplique, mas não a Lei dos Grandes Números. Vejamos isso no seguinte exemplo.

Exemplo 8.14

Seja $\{X_n\}_{n \geq 1}$ uma sequência de variáveis aleatórias independentes com função de probabilidade

$$P(X_k = k^\delta) = P(X_k = -k^\delta) = \frac{1}{2}, \quad k = 1, 2, \dots,$$

para algum $\delta \geq 0$. Então, $E(X_k) = 0$ e $\text{Var}(X_k) = k^{2\delta}$.

Observemos que

$$\sigma_{S_n}^2 = 1^{2\delta} + 2^{2\delta} + 3^{2\delta} + \dots + n^{2\delta} = O\left(\frac{n^{2\delta+1}}{2\delta+1}\right),$$

isto porque

$$\sigma_{S_n}^2 \leq \int_0^n x^{2\delta} dx = \frac{n^{2\delta+1}}{2\delta+1}$$

e também pela Lema 8.13.

Em termos gerais, uma condição suficiente para que a Lei dos Grandes Números possa ser aplicada é que

$$\lim_{n \rightarrow \infty} \frac{\sigma_{S_n}}{n} = 0. \quad (8.21)$$

Para demonstrarmos isto utilizamos a desigualdade de Chebyshev da seguinte forma

$$\lim_{n \rightarrow \infty} P(|S_n| > \epsilon n) \leq \lim_{n \rightarrow \infty} \frac{\text{Var}(S_n)}{\epsilon^2 n^2} = \lim_{n \rightarrow \infty} \frac{\sigma_{S_n}^2}{\epsilon^2 n^2} = 0,$$

desde que $0 < \delta < \frac{1}{2}$, isto porque, neste caso o limite em (8.21) se cumpre e, portanto, a Lei dos Grandes Números. Vamos verificar que, para $\delta > 0$, a Condição de Lindeberg se satisfaz.

Observemos que $k^\delta < n^\delta$, de maneira que a soma $\sum_{k=1}^n \sum_{|x_{kl}| > \epsilon \sigma_{S_n}} x_{kl}^2 p_{kl}$ pode ser diferente de zero se

$$n^\delta > \epsilon \sigma_{S_n} \approx \epsilon \frac{n^{\delta+1/2}}{\sqrt{2\delta+1}}.$$

Acontece que a relação acima é válida quando $(2\delta+1)\epsilon^{-2} > n$, daí decorre que, desde que, $n > (2\delta+1)\epsilon^{-2}$,

$$\frac{1}{\sigma_{S_n}^2} \sum_{k=1}^n \sum_{|x_{kl}| > \epsilon \sigma_{S_n}} x_{kl}^2 p_{kl} = 0$$

e a Condição de Lindeberg é satisfeita. Por conseguinte, o Teorema do Limite Central é válido para $\lambda > 0$. Significa que

$$\lim_{n \rightarrow \infty} P\left(\alpha < \sqrt{\frac{2\delta+1}{n^{2\delta+1}}} S_n \leq \beta\right) = \frac{1}{\sqrt{2\pi}} \int_{\alpha}^{\beta} e^{-\frac{x^2}{2}} dx.$$

Implica que

$$\lim_{n \rightarrow \infty} P\left(\alpha \frac{n^{\delta-1/2}}{\sqrt{2\delta+1}} < \frac{S_n}{n} \leq \beta \frac{n^{\delta-1/2}}{\sqrt{2\delta+1}}\right) = \frac{1}{\sqrt{2\pi}} \int_{\alpha}^{\beta} e^{-\frac{x^2}{2}} dx,$$

quando $\delta > 0$ e a Lei dos Grandes Números não se cumpre para $\delta \geq \frac{1}{2}$.

Acontece que, nem sempre, a convergência é para a distribuição Normal padrão. Vejamos o seguinte exemplo.

Exemplo 8.15 (Erdős e Kac (1946))

Sejam X_1, X_2, \dots variáveis aleatórias independentes identicamente distribuídas tais que $E(X_1) = 0$ e $E(X_1^2) = 1$. Sejam $S_0 = 0$, $S_n = X_1 + \dots + X_n$ e $\xi_n = \max\{S_0, S_1, \dots, S_n\}$, para $n = 1, 2, \dots$. Então

$$\lim_{n \rightarrow \infty} P\left(\frac{\xi_n}{\sqrt{n}} \leq x\right) = \begin{cases} 2\Phi(x) - 1, & \text{para } x \geq 0 \\ 0, & \text{para } x < 0 \end{cases}, \quad (8.22)$$

sendo que Φ denota a função de distribuição Normal padrão.

Vamos supor que $x \geq 0$ e primeiro vamos demonstrar (8.22) no caso particular

$$P(X_k = 1) = P(X_k = -1) = \frac{1}{2}, \quad (8.23)$$

para $k = 1, 2, \dots$. Pode-se demonstrar que

$$P(\xi_n \leq a) = P(S_n \leq a) - P(S_n \leq -a), \quad (8.24)$$

para $a = 1, 2, \dots$. Vamos escolher $a = a_n = x\sqrt{n}$ na expressão em (8.24) e, segundo o resultado do Exemplo 8.14, obtemos que

$$\lim_{n \rightarrow \infty} P\left(\frac{\xi_n}{\sqrt{n}} \leq x\right) = \lim_{n \rightarrow \infty} P\left(-x \leq \frac{S_n}{\sqrt{n}} \leq x\right) = \Phi(x) - \Phi(-x),$$

para $x \geq 0$ e demonstramos o resultado em (8.22) para este caso particular. Vamos supor que Y_1, Y_2, \dots sejam variáveis aleatórias independentes para as quais $P(Y_n \leq x) = \Phi(x)$. Sejam também $S'_0 = 0$, $S'_n = Y_1 + \dots + Y_n$ e $\eta_n = \max\{S'_0, S'_1, \dots, S'_n\}$, para $n = 1, 2, \dots$. Demonstramos que, para $k = 1, 2, \dots$ e $\epsilon > 0$, temos a seguinte desigualdade

$$\begin{aligned} P(\eta_k \leq (x - \epsilon)\sqrt{k}) - \frac{1}{\epsilon^2 k} &\leq \liminf_{n \rightarrow \infty} P(\xi_n \leq x\sqrt{n}) \\ &\leq \limsup_{n \rightarrow \infty} P(\xi_n \leq x\sqrt{n}) \leq P(\eta_k \leq x\sqrt{n}). \end{aligned} \quad (8.25)$$

Para cada $n = 1, 2, \dots$ e cada $k = 1, 2, \dots$ definamos $n_j = [nj/k]$, isto é, a parte inteira do quociente nj/k para $j = 0, 1, 2, \dots, k$ e vamos escrever

$$G_{nk}(x) = P(\max\{S_{n_0}, S_{n_1}, \dots, S_{n_k}\} \leq x\sqrt{n}).$$

Acontece que

$$\lim_{n \rightarrow \infty} G_{nk}(x) = P(\max\{S'_0, S'_1, \dots, S'_k\} \leq x\sqrt{k}). \quad (8.26)$$

Pelo Exemplo 8.14 temos

$$\begin{aligned} \lim_{n \rightarrow \infty} P(S_{n_1} \leq x_1\sqrt{n}, S_{n_2} - S_{n_1} \leq x_2\sqrt{n}, \dots, S_{n_k} - S_{n_{k-1}} \leq x_k\sqrt{n}) &= \\ &= \Phi(x_1\sqrt{k}) \Phi(x_2\sqrt{k}) \dots \Phi(x_k\sqrt{k}) \\ &= P(Y_1 \leq x_1\sqrt{k}, Y_2 \leq x_2\sqrt{k}, \dots, Y_k \leq x_k\sqrt{k}) \\ &= P(S'_1 \leq x_1\sqrt{k}, S'_2 - S'_1 \leq x_2\sqrt{k}, \dots, S'_k - S'_{k-1} \leq x_k\sqrt{k}), \end{aligned}$$

para quaisquer x_1, x_2, \dots, x_k . Isto implica em (8.26). Para $r = 1, 2, \dots, n$, seja agora

$$Q_r(x) = P(S_0 \leq x\sqrt{n}, S_1 \leq x\sqrt{n}, \dots, S_{r-1} \leq x\sqrt{n}, S_r \leq x\sqrt{n}).$$

Então $\sum_{r=1}^n Q_r(x) = 1 - G_n(x) \leq 1$. Para $n_i < r \leq n_{i+1}$, $i = 0, 1, \dots, k-1$ temos que

$$\begin{aligned} Q_r(x) &= P(S_0 \leq x\sqrt{n}, \dots, S_{r-1} \leq x\sqrt{n}, S_r > x\sqrt{n}, |S_{n_{i+1}} - S_r| \geq \epsilon\sqrt{n}) + \\ &\quad P(S_0 \leq x\sqrt{n}, \dots, S_{r-1} \leq x\sqrt{n}, S_r > x\sqrt{n}, |S_{n_{i+1}} - S_r| < \epsilon\sqrt{n}). \end{aligned} \quad (8.27)$$

Para qualquer $\epsilon > 0$, o primeiro termo à direita em (8.27) é

$$Q_r(x) P(|S_{n_{i+1}} - S_r| \geq \epsilon\sqrt{n}) \leq Q_r(x) \frac{1}{k\epsilon^2} \quad (8.28)$$

que decorre de $E\left((S_{n_{i+1}} - S_r)^2\right) = n_{i+1} - r \leq \frac{n}{k}$. Assim, a partir de (8.27) e (8.28) conclui-se que

$$\begin{aligned} 1 - G_n(x) &= \sum_{r=1}^n Q_r(x) \leq \frac{1}{k\epsilon^2} + \\ &+ \sum_{i=0}^{n-1} \sum_{n_i < r \leq n_{i+1}} P\left(S_0 \leq x\sqrt{n}, \dots, S_{r-1} \leq x\sqrt{n}, S_r > x\sqrt{n}, |S_{n_{i+1}} - S_r| \geq \epsilon\sqrt{n}\right) \leq \\ &\leq \frac{1}{k\epsilon^2} + P\left(\max\{S_{n_0}, S_{n_1}, \dots, S_{n_k}\} > (x - \epsilon)\sqrt{n}\right), \end{aligned}$$

isto é,

$$1 - G_n(x) \leq \frac{1}{k\epsilon^2} + 1 - G_{nk}(x - \epsilon),$$

para qualquer x e $\epsilon > 0$. Uma vez que, evidentemente $G_n(x) < G_{nk}(x)$, segue que

$$G_{nk}(x - \epsilon) - \frac{1}{k\epsilon^2} < G_n(x) < G_{nk}(x), \quad (8.29)$$

por todo x e $\epsilon > 0$. Tomando $n \rightarrow \infty$ na desigualdade em (8.29) obtemos a desigualdade em (8.25). Se aplicarmos (8.25) às variáveis aleatórias em (8.23), então obtemos que

$$P\left(S'_k \leq (x - \epsilon)\sqrt{k}\right) - \frac{1}{k\epsilon^2} \leq G(x) \leq P\left(S'_k \leq x\sqrt{k}\right), \quad (8.30)$$

onde

$$G(x) = \begin{cases} 2\Phi(x) - 1, & \text{para } x \geq 0 \\ 0, & \text{para } x < 0 \end{cases}.$$

Se substituirmos x por $x + \epsilon$ em (8.30), obtemos

$$G(x) \leq P\left(S'_{n_k} \leq x\sqrt{k}\right) \leq G(x + \epsilon) + \frac{1}{k\epsilon^2} \quad (8.31)$$

e, por conseguinte, por (8.25), (8.30) e (8.31)

$$G(x - \epsilon) - \frac{1}{k\epsilon^2} \leq \liminf_{n \rightarrow \infty} G_n(x) \leq \limsup_{n \rightarrow \infty} G_n(x) \leq G(x + \epsilon) + \frac{1}{k\epsilon^2} \quad (8.32)$$

para qualquer x , $\epsilon > 0$ e $k = 1, 2, \dots$. Sejam $k \rightarrow \infty$ e $\epsilon \rightarrow 0$ em (8.32). Dado que G é contínua, obtemos que

$$\lim_{n \rightarrow \infty} G_n(x) = G(x),$$

para qualquer x .

No exemplo acima foi demonstrado que se X_1, X_2, \dots forem variáveis aleatórias independentes distribuídas de forma idêntica com $E(X_1) = 0$ e $\text{Var}(X_1) = 1$, a distribuição limite em (8.22) não depende da função de distribuição $P(X_1 \leq x)$, a distribuição de cada variável aleatória na sequência e, inclusive, nem sempre a distribuição limite é Normal.

Corolário 8.19 (Teorema do Limite Central)

Sejam X_1, X_2, \dots variáveis aleatórias independentes, igualmente distribuídas tais que $E(X_1) = \mu$ e $\text{Var}(X_1) = \sigma^2 > 0$. Então

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{D} Z, \quad (8.33)$$

quando $n \rightarrow \infty$, sendo $S_n = \sum_{k=1}^n X_k$ e $Z \sim N(0, 1)$.

Demonstração: Verifiquemos que as condições do Teorema 8.18 ou condições de Lindeberg são satisfeitas, isto significa que vamos verificar se a condição em (8.16) é cumprida, ou seja, vamos verificar se

$$\lim_{n \rightarrow \infty} \frac{1}{n\sigma^2} \sum_{k=1}^n \int_{|x-\mu| > \epsilon\sigma\sqrt{n}} (x-\mu)^2 f(x) dx = 0.$$

Demonstremos o contrário, observemos que

$$1 - \frac{1}{n\sigma^2} \sum_{k=1}^n \int_{|x-\mu| \leq \epsilon\sigma\sqrt{n}} (x-\mu)^2 f(x) dx = \frac{1}{n\sigma^2} \sum_{k=1}^n \int_{|x-\mu| > \epsilon\sigma\sqrt{n}} (x-\mu)^2 f(x) dx.$$

Então,

$$\begin{aligned} \lim_{n \rightarrow \infty} \sum_{k=1}^n \frac{1}{n\sigma^2} \int_{|x-\mu| \leq \epsilon\sigma\sqrt{n}} (x-\mu)^2 f(x) dx &= \lim_{n \rightarrow \infty} \frac{1}{\sigma^2} \int_{|x-\mu| \leq \epsilon\sigma\sqrt{n}} (x-\mu)^2 f(x) dx \\ &= \frac{1}{\sigma^2} \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx = \frac{\sigma^2}{\sigma^2} = 1, \end{aligned}$$

satisfazendo-se assim a condição de Lindeberg em (8.16). ■

Exemplo 8.16

Sejam X_1, X_2, \dots variáveis aleatórias independentes identicamente distribuídas, cada uma com distribuição Geométrica(θ). Disto temos,

$$E(X_1) = \frac{1-\theta}{\theta} \quad \text{e} \quad \text{Var}(X_1) = \frac{1-\theta}{\theta^2}.$$

Pelo Teorema do Limite Central, Corolário 8.19, vemos que

$$\frac{\sum_{k=1}^n X_k - n\frac{1-\theta}{\theta}}{\sqrt{n\frac{1-\theta}{\theta^2}}} \xrightarrow{D} Z \sim N(0, 1), \quad \text{quando} \quad n \rightarrow \infty.$$

Exemplo 8.17

Sejam X_1, X_2, \dots variáveis aleatórias independentes identicamente distribuídas, cada uma com distribuição $Beta(\alpha, \beta)$. Então

$$E(X_1) = \frac{\alpha}{\alpha + \beta} \quad e \quad \text{Var}(X_1) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}.$$

Pelo Teorema do Limite Central, Corolário 8.19, vemos que

$$\frac{\sum_{k=1}^n X_k - n \frac{\alpha}{\alpha + \beta}}{\sqrt{n \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}}} \xrightarrow{D} Z,$$

quando $n \rightarrow \infty$, onde $Z \sim N(0, 1)$.

Vamos considerar que a sequência de variáveis aleatórias independentes identicamente distribuídas $\{X_n\}_{n \geq 1}$ é tal que $E(X_1) = 0$ e $E(X_1^2) = 1$. Nesta situação, vale o resultado anterior, Corolário 8.19 e, portanto, $\frac{S_n}{\sqrt{n}} \xrightarrow{D} Z \sim N(0, 1)$, sendo $S_n = X_1 + X_2 + \dots + X_n$. Isto equivale a afirmar que

$$\lim_{n \rightarrow \infty} \varphi^n(x/\sqrt{n}) = e^{-\frac{x^2}{2}}, \quad (8.34)$$

para todo $x \in \mathbb{R}$. Lembremos que φ representa a função característica de X_1 . Demonstramos isto utilizando a expansão de Taylor da seguinte forma

$$\varphi(z) = \varphi(0) + z\varphi'(0) + \frac{1}{2}z^2\varphi''(0) + o(z^2), \quad \text{quando} \quad z \approx 0. \quad (8.35)$$

Escolhendo z arbitrário e fazendo $x = z/\sqrt{n}$ concluímos que

$$\varphi\left(\frac{z}{\sqrt{n}}\right) = 1 - \frac{1}{2n}z^2 + o\left(\frac{1}{n}\right), \quad \text{quando} \quad n \rightarrow \infty.$$

Tomando potência n -ésima temos por resultado (8.34).

É natural esperar que, quando F possua uma densidade f , a densidade de S_n/\sqrt{n} tenha tendência para a densidade Normal padrão ϕ . Isto nem sempre é verdade, mas as exceções são, felizmente, bastante patológicas. O teorema seguinte abrange as situações que ocorrem na prática comum.

Teorema 8.20

Seja $\{X_n\}_{n \geq 1}$ uma sequência de variáveis aleatórias independentes identicamente distribuídas tais que $E(X_1) = 0$ e $E(X_1^2) = 1$. Então se, $|\varphi|$, a função característica de X_1 é integrável, temos que S_n/\sqrt{n} tem por densidade f_n que tende uniformemente à densidade Normal padrão ϕ .

Demonstração: Utilizando a fórmula de inversão é válida para ambos, f_n e ϕ e, portanto,

$$|f_n(x) - \phi(x)| \leq \frac{1}{2\pi} \int_{-\infty}^{\infty} \left| \varphi^n \left(\frac{z}{\sqrt{n}} \right) - e^{-\frac{z^2}{2}} \right| dz. \quad (8.36)$$

O lado direito é independente de x e devemos mostrar que tende a 0 quando $n \rightarrow \infty$. Tendo em conta que a expansão de Taylor em (??) é possível escolher $\delta > 0$ de maneira que

$$|\varphi(z)| \leq e^{-\frac{z^2}{4}}, \quad \text{quando} \quad |z| < \delta. \quad (8.37)$$

Dividimos agora o integral em três partes e provaremos que cada uma é $< \epsilon$ para n suficientemente grande. (1) Como vimos na última prova, dentro de um intervalo fixo $-a \leq z \leq a$ o integral tende uniformemente a zero e, por conseguinte, a contribuição de $\overline{-a, a}$ tende a zero. (2) ■

Exemplo 8.18

conteúdo...

Corolário 8.21 (*Teorema do Limite Central de Lyapunov*)

Sejam X_1, X_2, \dots variáveis aleatórias independentes. Sejam $E(X_k) = \mu_k$, $E(X_k - \mu_k)^2 = \sigma_k^2 \neq 0$ e $E(|X_k - \mu_k|^3) = \beta_k$, os quais existem para cada $k = 1, 2, \dots$.

$$\sigma_{S_n}^2 = \sigma_1^2 + \sigma_2^2 + \dots + \sigma_n^2.$$

Se F_k é absolutamente contínua com função de densidade f_k , assumiremos que a relação

$$\lim_{n \rightarrow \infty} \frac{1}{\sigma_{S_n}^2} \sum_{k=1}^n \int_{|x - \mu_k| > \epsilon S_n} (x - \mu_k) f_k(x) dx = 0, \quad (8.38)$$

é válida $\forall \epsilon > 0$. Se X_k é do tipo discreta com probabilidade positiva p_{kl} em x_{kl} , $l = 1, 2, \dots$, assumiremos que a relação

$$\lim_{n \rightarrow \infty} \frac{1}{\sigma_{S_n}^2} \sum_{k=1}^n \sum_{|x_{kl} - \mu_k| > \epsilon S_n} (x_{kl} - \mu_k) p_{kl} = 0, \quad (8.39)$$

é válida $\forall \epsilon > 0$. Então, a distribuição da soma padronizada

$$S_n^* = \frac{S_n - \sum_{k=1}^n \mu_k}{\sigma_{S_n}} = \frac{\sum_{k=1}^n (X_k - \mu_k)}{\sigma_{S_n}} \quad (8.40)$$

converge em distribuição à uma variável aleatória normal padrão, isto é,

$$\frac{\sum_{k=1}^n (X_k - \mu_k)}{\sigma_{S_n}} \xrightarrow{D} Z, \quad (8.41)$$

onde $Z \sim N(0, 1)$.

Demonstração : conteúdo... ■

Demonstração : Verifiquemos se a condição de Lindeberg se satisfaz. ■

Corolário 8.22 (*Lindeberg-Feller-Rao pg.128*)

Sejam X_1, X_2, \dots variáveis aleatórias independentes não degeneradas com funções de distribuição respectivas F_1, F_2, \dots . Assumiremos que $E(X_k) = \mu_k$, $\text{Var}(X_k) = \sigma^2$ e definiremos

$$\sigma_{S_n}^2 = \sigma_1^2 + \sigma_2^2 + \dots + \sigma_n^2.$$

Se F_k é absolutamente contínua com função de densidade f_k , assumiremos que a relação

$$\lim_{n \rightarrow \infty} \frac{1}{\sigma_{S_n}^2} \sum_{k=1}^n \int_{|x - \mu_k| > \epsilon S_n} (x - \mu_k) f_k(x) dx = 0, \quad (8.42)$$

é válida $\forall \epsilon > 0$. Se X_k é do tipo discreta com probabilidade positiva p_{kl} em x_{kl} , $l = 1, 2, \dots$, assumiremos que a relação

$$\lim_{n \rightarrow \infty} \frac{1}{\sigma_{S_n}^2} \sum_{k=1}^n \sum_{|x_{kl} - \mu_k| > \epsilon S_n} (x_{kl} - \mu_k) p_{kl} = 0, \quad (8.43)$$

é válida $\forall \epsilon > 0$. Então, a distribuição da soma padronizada

$$S_n^* = \frac{S_n - \sum_{k=1}^n \mu_k}{\sigma_{S_n}} = \frac{\sum_{k=1}^n (X_k - \mu_k)}{\sigma_{S_n}} \quad (8.44)$$

converge em distribuição à uma variável aleatória normal padrão, isto é,

$$\frac{\sum_{k=1}^n (X_k - \mu_k)}{\sigma_{S_n}} \xrightarrow{D} Z, \quad (8.45)$$

onde $Z \sim N(0, 1)$.

8.2.3 Expansões para densidades

O trabalho relacionado com o Teorema do Limite Central influenciou grandemente o desenvolvimento e a afiação das ferramentas agora geralmente utilizadas na teoria da probabilidade, pelo que uma comparação de diferentes provas é esclarecedora. Até recentemente, o método das funções características, utilizado pela primeira vez por P. Lévy, era incomparavelmente mais simples do que a abordagem direta concebida por Lindeberg, para não falar de outras abordagens. A versão moderna racionalizada desta última não é mais complicada e tem, além disso, outros méritos. Por outro lado, o método das funções características conduz a aperfeiçoamentos que, atualmente, não são exequíveis por métodos diretos. Entre estes encontram-se o Teorema do Limite Local nesta secção, bem como as estimativas de erro e expansões assintóticas desenvolvidas no próximo capítulo. Separamos o caso das variáveis com uma distribuição comum, em parte devido à sua importância, e em parte para explicar a essência do método na situação mais simples.

O Teorema do Limite Central para as densidades pode ser consideravelmente reforçado quando existem alguns momentos mais elevados. O pressuposto importante é que

$$\int_{-\infty}^{\infty} |\varphi(z)|^{\nu} dz < \infty, \quad (8.46)$$

para algum $\nu \geq 1$.

Para arredondar o quadro, passamos para a distribuição em malha, ou seja, supomos que as variáveis X_j estão restritas a valores da forma $x, x \pm h, x \pm 2h, \dots$. Assumimos que h é a extensão ou vão da distribuição F , ou seja, h é o maior número positivo com a propriedade declarada.

Teorema 8.23

Se F é uma distribuição em malha com vão h , então como n

Demonstração : conteúdo...



8.2.4 Velocidade de Convergência

Uso da Lei dos Grandes Números como aproximação

Na lei fraca dos grandes números, tivemos uma declaração sobre a velocidade da convergência (Teorema 5.14). Na lei forte dos grandes números, porém, não o fizemos. Como exigimos apenas os primeiros momentos, em geral, não podemos esperar obter nenhum estado útil. No entanto, se assumirmos a existência de momentos mais altos, ficamos razoáveis estimativas sobre a taxa de convergência.

O núcleo da lei fraca de grandes números é a desigualdade de Chebyshev. Aqui nós apresentamos uma desigualdade mais forte que reivindica o mesmo limite, mas agora para o máximo sobre todas as somas parciais até um tempo fixo.

Teorema 8.24 (*Generalização da desigualdade de Kolmogorov*)

Seja $n \in \mathbb{N}$ e sejam X_1, \dots, X_n variáveis aleatórias independentes tais que $E(X_k) = 0$ e $\text{Var}(X_k) < \infty$, para $k = 1, \dots, n$. Mais ainda, sejam $S_m = X_1 + \dots + X_m$ para $m = 1, \dots, n$. Então, qualquer seja $\epsilon > 0$,

$$P\left(\max_{m=1, \dots, n} \{S_m\} \geq \epsilon\right) \leq \frac{\text{Var}(S_n)}{\epsilon^2 + \text{Var}(S_n)}.$$

Demonstração: Vamos decompor o espaço de probabilidade de acordo com a primeira vez τ em que as somas parciais excedem o valor t . Por conseguinte, seja

$$\tau = \min_{m \in \{1, 2, \dots, n\}} \{S_m \geq t\}$$

e $A_m = \{\tau = m\}$ para $m = 1, 2, \dots, n$. Além disso, sejam as variáveis aleatórias $(S_m + c)\mathbf{1}_{A_m}$ e $S_n - S_m$. Pelo Teorema 6.16, as duas variáveis aleatórias são independentes e

$$E((S_m + c)\mathbf{1}_{A_m}(S_n - S_m)) = E((S_m + c)\mathbf{1}_{A_m})E(S_n - S_m) = 0.$$

Os eventos A_1, \dots, A_n são disjuntos a pares; por conseguinte $\sum_{k=1}^n \mathbf{1}_{A_k} = \mathbf{1}_A \leq 1$. Obtemos então

$$\begin{aligned} \text{Var}(S_n) + c^2 &= E((S_n + c)^2) \geq E\left(\sum_{k=1}^n (S_n + c)^2 \mathbf{1}_{A_k}\right) = \sum_{k=1}^n E((S_n + c)^2 \mathbf{1}_{A_k}) \\ &= \sum_{k=1}^n E\left((S_k + c)^2 + 2(S_k + c)(S_n - S_m) + (S_n - S_m)^2 \mathbf{1}_{A_k}\right) \\ &= \end{aligned}$$

■

Da desigualdade de Kolmogorov, derivamos a seguinte apuração da forte lei dos grandes números.

Uso do Teorema do Limite Central como aproximação

Em matemática, é frequentemente feita uma distinção entre teoremas de limites e teoremas de aproximação; o primeiro especifica simplesmente o limite de uma sequência, enquanto o segundo fornece uma estimativa ou um limite da diferença entre um elemento da sequência e o seu limite. Por exemplo, é sabido que $\ln(1+x)$ pode ser aproximado por x quando x é pequeno; um limite bruto sobre a diferença absoluta $|\ln(1+x) - x|$ é $x^2/2$ quando $x \geq 0$ e $x^2/(2(1+x)^2)$ quando $x < 0$. Para efeitos práticos, os teoremas de aproximação são mais úteis, uma vez que permitem uma estimativa do erro cometido na aproximação por o limite.

O Teorema do Limite Central, tal como aqui referido, não é um teorema de aproximação. Esse teorema não nos diz qual deve ser o tamanho de n para a distribuição de S_n se aproximar da distribuição Normal padrão. No entanto, com pressupostos adicionais, este mesmo teorema pode ser expresso como um teorema da aproximação. O seguinte resultado que foi demonstrado de maneira independentemente por Berry (1941) e Esseen (1942), trata acerca da velocidade da convergência no Teorema do Limite Central.

Teorema 8.25 (*Berry-Esseen*)

Sejam X_1, X_2, \dots variáveis aleatórias independentes identicamente distribuídas tais que $E(X_1) = 0$, $E(X_1^2) = \sigma^2$ e $E(|X_1^3|) < \infty$. Sejam também $S_n = \frac{1}{\sqrt{n}\sigma^2}(X_1 + X_2 + \dots + X_n)$ e Φ a função de distribuição Normal padrão. Então, para todo $n \in \mathbb{N}$,

$$\sup_{x \in \mathbb{R}} |P(S_n \leq x) - \Phi(x)| \leq \frac{0.8 E(|X_1^3|)}{\sigma^3 \sqrt{n}}.$$

Demonstração: Uma vez que nenhum dos lados da desigualdade é afectado pela escalada, podemos supor sem perda de generalidade que $\sigma = 1$. A primeira fase do argumento consiste em obter uma desigualdade, Lemma 3.4.11, que relaciona a diferença entre as duas distribuições com a distância entre os seus ch.f. A densidade do Polya (ver Exemplo 3.3.8 e utilização (e) de Teorema 3.3.1) $h_L(x) = 1 - \cos Lx$ tem ch.f. $L(\cdot) = (1 - \cdot/L) +$ para $0 \leq \cdot \leq L$. Utilizaremos HL para a sua função de distribuição. Convolveremos as distribuições em consideração com HL para obter ch.f. que tem suporte compacto. O primeiro passo é mostrar que a convolução com a HL não reduz a diferença entre as distribuições é demasiado grande. ■

O traço marcante da desigualdade é que ela depende apenas dos três primeiros momentos. A expansão proporciona uma melhor estimativa assintótica, mas a velocidade da convergência depende de propriedades mais delicadas da distribuição subjacente. O fator 3 da direita poderia ser substituído por um C de limite superior melhor, mas não é feita qualquer tentativa na nossa configuração para alcançar resultados ótimos.

Exemplo 8.19

Considere as variáveis aleatórias independentes identicamente distribuídas com função de

densidade

$$f(x) = \frac{1}{2\alpha} \frac{1}{x^{1+1/\alpha}}, \quad \text{quando } |x| \geq 1.$$

Neste caso $E(X_1) = 0$ e $\text{Var}(X_1) = 1/(1 - 2\alpha) < \infty$ se $\alpha < 1/2$. Denotemos como F_n a função de distribuição de S_n e esta como definida no Teorema 8.25. Observemos que $0 < E(|X_1^3|) = 1/(1 - 3\alpha) < \infty$ quando $\alpha < 1/3$.

Seja F_n a função de distribuição de S_n e Φ a função de distribuição normal padrão. Para ganhar algum conhecimento dos factores que afetam a rapidez da convergência da F_n para Φ , utilizaremos as expansões da Edgeworth⁴. Para se ter uma ideia dos factores que afetam a velocidade de convergência de F_n a Φ , utilizaremos as expansões da Edgeworth. Suponha que F_n seja uma função de distribuição contínua e que $E(X_k^4) < \infty$, definamos

$$\gamma = \frac{E(X_k - \mu)^3}{\sigma^3} \quad \text{e} \quad \kappa = \frac{E(X_k - \mu)^4}{\sigma^4} - 3,$$

sendo γ e κ os coeficientes de skewness (enviesamento) e kurtosis (curtose), respectivamente. Estes coeficientes assumem valor 0 caso a distribuição seja a Normal padrão.

Teorema 8.26 (*Expansão de Edgeworth*)

Teorema do Limite Central!Aproximação de Edgeworth Seja X_1, X_2, \dots uma sequência de variáveis aleatórias independentes identicamente distribuídas. Então,

$$F_n(x) = \Phi(x) - \phi(x) \left(\frac{\gamma}{6\sqrt{n}} p_1(x) + \frac{\kappa}{24n} p_2(x) + \frac{\gamma^2}{72n} p_3(x) \right) + r_n(x),$$

onde ϕ é a função de densidade Normal padrão e p_1, p_2 e p_3 são polinômios, definidos como $p_1(x) = x^2 - 1$, $p_2(x) = x^3 - 3x$ e $p_3(x) = x^5 - 10x^3 + 15x$ e o termo resto $r_n(x)$ satisfaz que $nr_n(x) \rightarrow 0$.

Demonstração: conteúdo...



Os Hermite⁵

A partir desta expansão, parece claro que o erro de aproximação $|F_n(x) - \Phi(x)|$ depende somente do skewness e kurtosis, ou seja, de γ e κ dos X_k 's. Os coeficientes de skewness e kurtosis são medidas simples de que quanto uma distribuição difere da normalidade; o skewness é uma medida da assimetria da distribuição, $\gamma = 0$ se a distribuição for simétrica em torno da sua esperança, enquanto a kurtosis é uma medida da espessura da cauda da distribuição, $\kappa > 0$ indica caudas mais pesadas do que a da distribuição normal enquanto $\kappa < 0$ indica caudas mais leves.

⁴Francis Ysidro Edgeworth (1845-1926). Foi um filósofo e economista político anglo-irlandês que contribuiu significativamente para os métodos estatísticos durante a década de 1880. A partir de 1891, foi nomeado editor fundador do The Economic Journal.

⁵Charles Hermite (1822-1901). Foi um matemático francês.

Por exemplo, uma distribuição *Uniforme* tem $\gamma = 0$ e $\kappa = -1.2$ enquanto uma distribuição *Exponencial* tem $\gamma = 2$ e $\kappa = 6$. Assim, devemos esperar que a convergência ocorra mais rapidamente para somas de variáveis aleatórias *Uniformes* do que para somas de variáveis aleatórias *Exponenciais*. Realmente, isto é verdade; de fato, a distribuição de uma soma de poucas variáveis aleatórias *Uniformes* está suficientemente próxima da distribuição de uma variável *Normal* padrão. Para ilustrar a diferença na precisão da aproximação *Normal*, consideramos a distribuição de $S_n = X_1 + \cdots + X_{10}$ quando as X_k 's são uniformes e exponenciais; a Figura 8.3 (a) e (b) dão as densidades exatas e as suas aproximações Normais nestes dois casos.

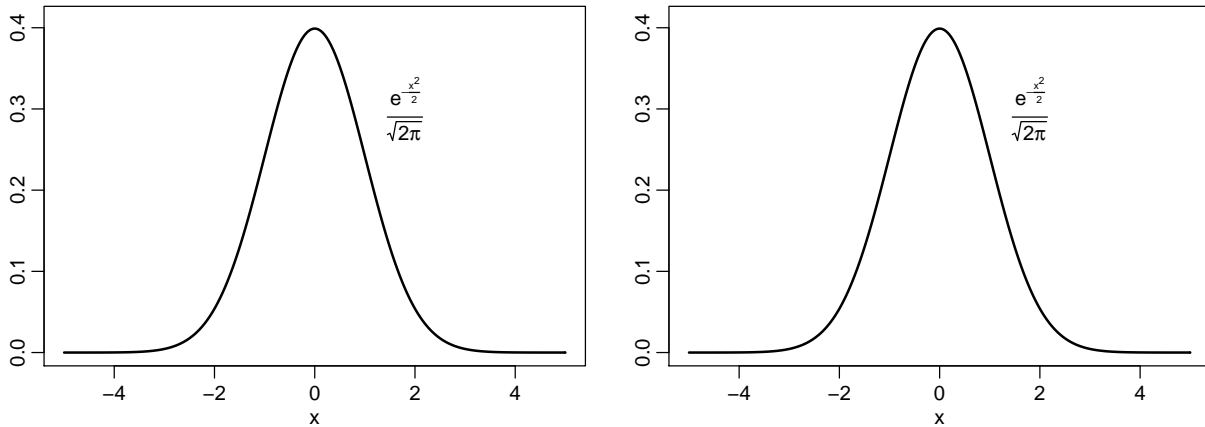


Figura 8.3: (a) Comportamento de $nP(|X| > n)$, sendo $X \sim \text{Cauchy}(0, 1)$ e (b).

A velocidade de convergência do Teorema do Limite Central e, por conseguinte, a bondade da aproximação pode frequentemente ser melhorada através da aplicação de transformações para reduzir a skewness e a kurtosis de \bar{X}_n . Recorde-se que se $\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{D} Z$, então

$$\sqrt{n}(g(\bar{X}_n) - g(\mu)) \xrightarrow{D} g'(\mu)Z.$$

Se g for escolhido de forma a que a distribuição de $g(\bar{X}_n)$ seja mais simétrica e tenha caudas mais leves do que a de \bar{X}_n , o Teorema do Limite Central deverá fornecer uma aproximação mais precisa para a distribuição de $\sqrt{n}(g(\bar{X}_n) - g(\mu))$ do que para a distribuição de $\sqrt{n}(\bar{X}_n - \mu)$.

Embora a expansão da Edgeworth nem sempre seja válida quando os X_k 's são discretos, os comentários anteriores sobre velocidade de convergência e precisão da aproximação normal continuam a ser, em geral, verdadeiras. No entanto, quando os X_k 's são discretos, há uma técnica simples que pode melhorar a precisão da aproximação normal. Vamos ilustrar esta técnica para a distribuição *Binomial*. Suponha que X tenha distribuição *Binomial* com parâmetros n e θ ; X pode ser considerada como uma soma de n variáveis aleatórias independentes identicamente distribuídas *Bernoulli* para que a distribuição de X possa ser aproximado por uma distribuição *Normal*, se n for suficientemente grande. Mais especificamente, a distribuição de

$$\frac{X - n\theta}{\sqrt{n\theta(1 - \theta)}}$$

é aproximadamente normal para n grandes. Suponhamos que queremos avaliar $P(a \leq X \leq b)$ para alguns inteiros a e b . Ignorando o fato de X ser uma variável aleatória discreta e a distribuição Normal é uma distribuição contínua, uma aplicação ingênua do Teorema do Limite Central dá

$$\begin{aligned} P(a \leq X \leq b) &= P\left(\frac{a - n\theta}{\sqrt{n\theta(1-\theta)}} \leq \frac{X - n\theta}{\sqrt{n\theta(1-\theta)}} \leq \frac{b - n\theta}{\sqrt{n\theta(1-\theta)}}\right) \\ &\approx \Phi\left(\frac{b - n\theta}{\sqrt{n\theta(1-\theta)}}\right) - \Phi\left(\frac{a - n\theta}{\sqrt{n\theta(1-\theta)}}\right). \end{aligned}$$

Como pode esta aproximação ser melhorada? A resposta é clara se compararmos a distribuição exacta de X com a sua aproximação normal. A distribuição de X pode ser convenientemente representada como um histograma de probabilidade, como na Figura 3.6, com a área de cada barra que representa a probabilidade de X tomar um determinado valor. A aproximação normal ingênua acima apresentada apenas integra a densidade normal aproximada de $a = 8$ a $b = 17$; esta A probabilidade é representada pela área sombreada na figura 3.7. É . parece que a aproximação normal ingênua irá subestimar a verdadeira probabilidade e as Figuras 3.7 e 3.8 sugerem que uma melhor a aproximação pode ser obtida através da integração de $-0,5$ a $7,5$

8.3 Exercícios

Exercícios da Seção 8.1

1. Seja $\{X_n\}$ uma sequência de variáveis aleatórias independentes e igualmente distribuídas com segundo momento finito. Seja

$$Y_n = \frac{2}{n(n+1)} \sum_{i=1}^n iX_i.$$

Prove que $Y_n \xrightarrow{P} E(X_1)$.

2. Seja $\{X_n\}_{n \geq 1}$ uma sequência de variáveis aleatórias com função de probabilidade dada por

$$P(X_n = 1) = \frac{1}{n} \quad \text{e} \quad P(X_n = 0) = 1 - \frac{1}{n}, \quad \forall n \geq 1.$$

Prove que $X_n = o_P(1)$.

3. Para as seguintes sequências de variáveis aleatórias independentes $\{X_n\}_{n \geq 1}$ a Lei dos Grandes Números se verifica?

$$(a) P(X_n = \pm 2^n) = \frac{1}{2},$$

$$(b) P(X_n = \pm 2^n) = \frac{1}{2^{2n+1}}, P(X = 0) = 1 - \frac{1}{2^{2n}},$$

$$(c) P(X_n = \pm n) = \frac{1}{2\sqrt{n}}, P(X_n = 0) = 1 - \frac{1}{\sqrt{n}}.$$

4. Provar a relação em (8.5) no caso discreto.
5. Demonstrar o Corolário 8.5.
6. **Medida de Lévy.** Sejam F e G duas funções de distribuição e definimos $\rho(F, G)$ da seguinte forma

$$\rho(F, G) = \inf_{h>0} \{F(x-h) - h \leq G(x) \leq F(x+h) + h\}, \quad \forall x \in \mathbf{R}.$$

Prove que:

(a) ρ , definida acima, é uma função de distância⁶.

(b) $\lim_{n \rightarrow \infty} F_n(x) = F(x)$, $\forall x \in \mathbf{R}$ se, e somente se, $\lim_{n \rightarrow \infty} \rho(F_n(x), F(x)) = 0$.

7. Demonstrar o Corolário 8.8.

8.

⁶Função de distância: uma função $d : \mathbf{X} \times \mathbf{X} \rightarrow \mathbf{R}$ é uma distância se satisfaz os seguintes axiomas:

- (i) ser positivamente definida, ou seja, $d(x, y) \geq 0$, para todos $x, y \in \mathbf{R}$,
- (ii) deve ser simétrica, ou seja, $d(x, y) = d(y, x)$, para todos $x, y \in \mathbf{R}$,
- (iii) satisfazer a desigualdade triangular, ou seja, $d(x, z) \leq d(x, y) + d(y, z)$, para todos $x, y, z \in \mathbf{R}$,
- (iv) é nula quanto os pontos coincidirem, ou seja, $d(x, y) = 0$ se, e somente se, $x = y$.

Exercícios da Seção 8.2

1. Demonstrar os Teoremas 8.12 e 8.14.
2. Seja $\{X_n\}_{n \geq 1}$ uma sequência de variáveis aleatórias independentes com as seguintes funções de probabilidade. Em cada caso mostre se a condição de Lindeberg é válida?
 - i) $P(X_n = \pm 1/2^n) = 1/2$.
 - ii) $P(X_n = \pm 2^{n+1}) = 1/2^{n+3}$ e $P(X_n = 0) = 1 - 1/2^{n+2}$.
 - iii) $P(X_n = \pm 1) = (1 - 2^{-n})/2$ e $P(X_n = \pm 2^{-n}) = 1/2^{n+1}$.
 - iv) $\{X_n\}_{n \geq 1}$ é uma sequência de variáveis aleatórias independentes $Poisson(\lambda_n)$, $n = 1, 2, \dots$, tais que $\sum_{n=1}^{\infty} \lambda_n \rightarrow \infty$.
 - v) $P(X_n = \pm 2^n) = 1/2$.
 - vi) $P(X_n = \pm 2^n) = 1/2^{n+1}$ e $P(X_n = \pm 1) = 1/2(1 - 1/2^n)$.
3. Seja $\{X_n\}_{n \geq 1}$ uma sequência de variáveis aleatórias independentes cada uma com função de densidade $f_k(x) = \frac{x^{a_k-1}}{\Gamma(a_k)e^x}$, onde $a_k \rightarrow \infty$, para $k = 1, 2, \dots$. A variância de S_n é $\sigma_{S_n}^2 = a_1 + \dots + a_n$. Mostre que a Condição de Lindeberg é satisfeita se

$$\lim_{n \rightarrow \infty} \frac{1}{\sigma_{S_n}^2} \sum_{k=1}^n a_k^2 = 0.$$

4. Seja $\{X_n\}_{n \geq 1}$ uma sequência de variáveis aleatórias independentes com função de probabilidade $Binomial(n, \theta)$, $0 < \theta < 1$. Se $S_n = X_n/n$, para $n \geq 1$,
 - i) prove que $\sqrt{n}(S_n - \theta) \xrightarrow{D} Z$, onde $Z \sim N(0, \theta(1 - \theta))$.
 - ii) determine o limite em distribuição das sequências de variáveis aleatórias $\sqrt{n}(1/S_n - 1/\theta)$ e $\sqrt{n}[S_n(1 - T_n) - \theta(1 - \theta)]$.
5. Seja $\{X_k\}$ uma sequência de variáveis aleatórias independentes com distribuição simétrica. Para cada n e $k \leq n$ seja $X_{k,n}$ a variável obtida pelo truncamento de X_k em $\pm a_n$. Suponha que os limites $\lim_{n \rightarrow \infty} \sum_{k=1}^n P(|X_k| > a_n) = 0$ e

$$\lim_{n \rightarrow \infty} \sum_{k=1}^n \int_{|x| > t} x^2 f_{k,n}(x) dx = 0,$$

para cada $t > 0$, sejam satisfeitos. Mostre que a distribuição de S_n/a_n tende à distribuição Normal padrão.

6. Seja X_1, X_2, \dots uma sequência de variáveis aleatórias de $X \sim Uniforme(0, \theta)$, $\theta > 0$. Prove que $\sqrt{n}(\log(2\bar{X}_n) - \log(\theta))$ converge em distribuição. Encontre o limite.
7. Seja X_1, X_2, \dots uma sequência de variáveis aleatórias com mesma média μ e variância σ^2 . Também seja, $\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k$ e $S^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X})^2$. Mostre que $\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{D} Z$, onde $Z \sim N(0, 1)$.
8. Seja $\{X_k\}_{k \geq 1}$ uma sequência de variáveis aleatórias independentes igualmente distribuídas cada uma com função de distribuição

$$F(x) = \begin{cases} 1 - \frac{1}{x \ln^2(x)}, & \text{para } x \geq e, \\ 0, & \text{para } x < e. \end{cases}$$

Encontrar a distribuição limite de $\{(S_n - a_n)/b_n\}_{n \geq 1}$ quando $n \rightarrow \infty$ e $\{a_n\}_{n \geq 1}$ e $\{b_n\}_{n \geq 1}$ são sequências de constantes normalizadoras escolhidas de forma adequada.

9. Seja $\{X_n\}_{n \geq 1}$ uma sequência de variáveis aleatórias com $X_n \sim \text{Gama}(n, \beta)$, onde $\beta > 0$ é uma constante independente de n . Encontre a distribuição limite de X_n/n .
10. Seja $\{X_n\}_{n \geq 1}$ uma sequência de variáveis aleatórias com $X_n \sim \chi^2(n)$, onde $n = 1, 2, \dots$. Encontre a distribuição limite de X_n/n^2 .
11. Seja $\{X_k\}_{k \geq 1}$ uma sequência de variáveis aleatórias independentes igualmente distribuídas cada uma com função de distribuição

$$F(x) = \begin{cases} 1 - \frac{1}{2x \ln^2(x)}, & \text{para } x \geq e, \\ \frac{1}{2e}, & \text{para } -e \leq x < e, \\ 1 - \frac{1}{3|x| \ln^2(|x|)}, & \text{para } x \leq -e. \end{cases}$$

Encontrar a distribuição limite de $\{(S_n - a_n)/b_n\}_{n \geq 1}$ quando $n \rightarrow \infty$ e $\{a_n\}_{n \geq 1}$ e $\{b_n\}_{n \geq 1}$ são sequências de constantes normalizadoras escolhidas de forma adequada.

12. (Kallianpur & Robbins, 1954) Seja $\{X_k\}_{k \geq 1}$ uma sequência de variáveis aleatórias independentes igualmente distribuídas cada uma com função de densidade

$$f(x) = \begin{cases} \frac{1}{2} \frac{\ln(x)}{x^2}, & \text{para } |x| \geq 1, \\ 0, & \text{para } |x| \leq 1. \end{cases}$$

Encontrar a distribuição limite de $\{(S_n - a_n)/b_n\}_{n \geq 1}$ quando $n \rightarrow \infty$ e $\{a_n\}_{n \geq 1}$ e $\{b_n\}_{n \geq 1}$ são sequências de constantes normalizadoras escolhidas de forma adequada.

13. Sejam X_1, X_2, \dots, X_n variáveis aleatórias com distribuição conjunta Normal com $E(X_k) = V$, $E(X_k^2) = 1$, para todo k e $\text{Cov}(X_k, X_l) = \rho$, $k, l = 1, 2, \dots$ e $k \neq l$. Qual é a distribuição limite de $\frac{1}{n} \sum_{k=1}^n X_k$.

Referências Bibliográficas

- Berry, A.C. (1941). The accuracy of the Gaussian approximation to the sum of independent variates. *Transactions of the American Mathematical Society*, **49**, 122–136.
- Billingsley, P. (1961). *Statistical Inference for Markov Processes*. Chicago: University Press.
- Billingsley, P. (1968). *Convergence of Probability Measures*. Wiley, New York.
- Broemeling, L.D. (2011). An Account of Early Statistical Inference in Arab Cryptology. *The American Statistician*, **65**(4), 255–257.
- Chung, K.L. (2001). *A course in probability theory*. Academic Press, Inc., third edition edition.
- Cohn, D.L. (1980). *Measure Theory*. Birkhäuser.
- Cramér, H. (1936). Ueber eine Eigenschaft der normalen Verteilungsfunktion. *Mathematische Zeitschrift*, **41**, 405–414.
- Curtiss, J.H. (1942). A note on the theory of moment generating functions. *Annals of Mathematical Statistics*, **13**, 430–433.
- Dall’Aglio, G. (1959). Sulla compatibilità delle funzioni di ripartizione doppia. *Rendiconti di Matematica e delle sue Applicazioni*, **18**, 385–413.
- Dovgoshey, O., Martio, O., Ryazanov, V. & Vuorinen, M. (2006). The Cantor function. *Expositiones Mathematicae*, **24**, 1–37.
- Dunn, K.P. & Smyth, G.K. (1996). Randomized quantile residuals. *Journal of Computational and Graphical Statistics*, **5**, 1–10.
- Durrett, R. (1996). *Probability; theory and examples*. Duxbury Press, second edition edition.
- Erdős, P. & Kac, M. (1946). On certain limit theorems of the theory of probability. *Bulletin of the American Mathematical Society*, **52**(4), 292–302.
- Esseen, C.G. (1942). *On the Liapounoff Limit of Error in the Theory of Probability*. Arkiv för matematik, astronomi och fysik. Almqvist & Wiksell.
- Farlie, D.J.G. (1960). The performance of some correlation coefficients for a general bivariate distribution. *Biometrika*, **47**, 307–323.
- Féchet, M.R. (1951). Sur les tableaux de corrélation dont les marges sont données. *Annales de l’Université de Lyon*, **4**, 53–84.
- Fischer, H. (2010). *A History of the Central Limit Theorem*. Springer.
- Fréchet, M.R. (1958). Remarques au sujet de la note précédente. *Comptes Rendus de l’Académie des Sciences, Paris*, **246**, 2719–2720.

- Ghosh, I. & Alzaatreh, A. (2015). A new class of bivariate and multivariate exponential distributions. *Far East Journal of Theoretical Statistics*, **20**(2), 77–98.
- Gnedenko, B.V. & Kolmogorov, A.N. (1954). *Limit distributions for sums of independent random variables*. Addison-Wesley Publishing Co., Inc.
- Gnedenko, B.V. & Kolmogorov, A.N. (1962). *Limits Distributions for Sums of Independent Random Variables*. Addison-Wesley, Reading, Massachusetts, second edition.
- Gumbel, E.J. (1958). Distributions à plusieurs variables dont les marges sont données. *Comptes Rendus de l'Académie des Sciences, Paris*, **246**, 2717–2720.
- Halmos, P.R. (1950). *Measure Theory*. Van Nostrand, New York.
- Harris, B. (1966). *Theory of Probability*. Addison-Wesley.
- Heyde, C. (1963). On a property of the lognormal distribution. *jrss*, **25**(2), 392–393.
- Kallianpur, G. & Robbins, H. (1954). The sequence of sums of independent random variables. *Duke Mathematical Journal, Duke University Press*, **21**(2), 285–307.
- Kolmogorov, A.N. (1933). *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Berlin: Springer.
- Kolmogorov, A.N. & Fomin, S.V. (1961). *Elements of the Theory of Functions and Functional Analysis*, volume 2. Garylock Press, New York.
- Lindeberg, J.W. (1922). Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung. *Mathematische Zeitschrift*, **15**, 211–225.
- Loève, M. (1977). *Probability Theory*. Springer-Verlag, fourth edition.
- Pitman, E.J.G. & Williams, E.J. (1967). Cauchy-distributed functions of Cauchy variates. *Annals of Mathematical Statistics*, **38**, 916–918.
- Psarakis, S. & Panaretos, J. (1990). The folded t distribution. *Communications in Statistics - Theory and Methods*, **19**(7), 2717–2734.
- Quesada-Molina, J.J. & Rodríguez-Lallena, J.A. (1995). Bivariate copulas with quadratic sections. *Journal of Nonparametric Statistics*, **5**, 323–337.
- Raikov, D.A. (1938). On the decomposition of Gauss and Poisson laws. *Izv. Akad. Nauk SSSR*, **2**, 91–124.
- Royden, H.L. (1988). *Real Analysis*. Prentice Hall, New Jersey, third edition.
- Rudin, W. (1976). *Principles of Mathematical Analysis*. McGraw-Hill, Inc., third edition.
- Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de statistique de l'Université de Paris*, **8**, 229–231.

Spivak, M. (1994). *Calculus*. Publish or Perish, Inc, third edition.

Stettin, R. (2019). *Generalizations and Properties of the Ternary Cantor Set and Explorations in Similar Sets*. Ph.D. thesis, Electronic Thesis or Dissertation. Ashland University. Retrieved from <https://etd.ohiolink.edu/>.

Teicher, Y.S. & Chow, H. (1978). *Probability Theory: Independence, Interchangeability, Martingales*. New York: Springer-Verlag.

Worsley, K. J. (1982). An Improved Bonferroni Inequality and Applications. *Biometrika*, **69**(2), 297–302.

Índice Remissivo

- σ -álgebra de Borel
 - Definição, 39
 - Elementos, 41
- σ -álgebra de eventos, 31
 - σ -álgebra de Borel, 38
 - σ -álgebra gerada, 37
 - σ -álgebra mínima, 36
 - Definição de, 34
- Cantor
 - Conjunto de , 44
- Conjuntos, 2
- Convergência em distribuição, 336
- Convergência em momentos, 326
- Convergência em probabilidade, 321
- Convergência quase certamente, 332
- Densidade Exponencial
 - Conjunta, 302
- Desigualdades probabilísticas, 312
 - Chebyshev, 312
 - Hoeffding, 316
 - Markov, 312
- Distribuição
 - χ^2 , 133
 - Bernoulli, 109
 - Beta, 134
 - Binomial, 110
 - Binomial Negativa, 114
 - Cantor, 120
 - Cauchy, 134
 - Exponencial, 130, 183
 - Exponencial dupla, 176
 - Gama, 131
 - Geométrica, 113
 - Hipergeométrica, 116
 - Laplace, 176
 - log-Normal, 137
 - Logística, 219
 - Normal, 135
 - Normal inversa, 136
 - Pareto, 220
 - Poisson, 112
 - qui, 186
 - Rayleigh, 159
 - t-Student, 175
 - Uniforme, 126
 - Uniforme discreta, 108
 - Weibull, 159
- Distribuição condicional, 237
- Distribuição conjunta, 226, 227
- Distribuição Normal
 - Multivariada, 263, 264
- Distribuição truncada, 238
- Distribuições de vetores aleatórios, 230
- Espaço amostral, 8
- Esperança, 162
 - Bernoulli, 163
 - Binomial Negativa, 171
 - Cauchy, 170
 - Densidade Beta, 212
 - Desigualdade de Jensen, 190
 - Distribuição Geométrica, 207
 - Exponencial, 162
 - Geométrica, 164
 - Linearidade, 173
 - log-Normal, 214
 - Propriedades, 171
 - t-Student, 181, 185
 - Uniforme, 174, 185, 202
- Esperança condicional, 240
 - Definição, 240
- Esperança condicional
 - Propriedades, 241

- Esperança de funções de variáveis aleatórias, 182
- Função Borel mensurável, 141
- Função característica, 276
 - Infinitamente divisíveis, 305
 - Propriedades, 278
 - Fórmula de Inversão, 284
 - Teorema de Unicidade, 286
- Função característica vetorial, 302
- Função de distribuição, 97
 - Propriedades, 98
- Função geradora de cumulantes, 222
- Função geradora de momentos, 208
 - Propriedades, 209
- Função geradora de probabilidade, 203
 - Propriedades, 205
- Função indicadora, 109
- Funções de variáveis aleatórias, 140
- Funções de vetores aleatórios, 228
- Infinitamente divisíveis, 305
- Lei dos Grandes Números, 355
 - Chebyshev, 356
 - Khintchine, 361
- Lei Forte dos Grandes Números, 363
 - Função de distribuição empírica, 366
 - Integração Monte Carlo, 365
 - Teorema de Glivenko-Cantelli, 367
- Lei Fraca dos Grandes Números, 355
- Modelo Copula, 266
- Modelos Copula, 265
- Momentos, 184
- Momentos de vetores, 229
- Operações entre conjuntos, 3
- Percentil
 - Mediana, 195
- Percentis, 191
- Princípio Aditivo da Contagem, 12
- Princípio da Inclusão-Exclusão, 12
- Princípio Multiplicativo da Contagem, 11
- Probabilidade axiomática, 46
 - Definição axiomática de probabilidade, 47
- Espaço de probabilidade, 48
- Propriedades, 54
 - Desigualdade de Bonferroni, 58
 - Desigualdade de Boole, 56
 - Princípio de inclusão-exclusão, 56
- Teorema de Extensão, 53
- Probabilidade clássica, 8, 9
- Probabilidade condicional, 60
 - Definição, 61
- Independência, 67
 - Eventos independentes, 67
 - Lema de Borel-Cantelli, 72
- Permutabilidade, 73
- Teorema da multiplicação, 65
- Teorema da probabilidade total, 64
- Teorema de Bayes, 65
- Probabilidade frequencial, 17
- Probabilidade geométrica, 20
- Teorema
 - Carathéodory, 53
 - Continuidade, 300
 - Convergência dominada, 331
 - Convergência monótona, 329
 - Critério de Polya, 295
 - Helly-Bray, 297
 - Inversão, 284
 - Unicidade, 286
- Teorema do Limite Central, 386
 - Aproximação de Barry-Essen, 392
 - de Lindeberg, 378
 - Lindeberg-Feller, 389
 - Lyapunov, 388
- Transformação do Jacobiano, 146
 - Generalização, 150
- Variáveis aleatórias, 83
 - Absolutamente contínuas, 119
 - Contínuas, 118
 - Discretas, 107
 - Propriedades, 91
 - Simples, 90
 - Singular, 120
- Variáveis aleatórias discretas, 106

- Degenerada, 107
- Variáveis aleatórias independentes, 249
- Variáveis infinitamente divisíveis, 305
- Variáveis permutáveis, 256
- Variável aleatória, 85
 - Simétrica, 174
- Variância, 198
 - Densidade Beta, 212
 - Distribuição Bernoulli, 200
 - Distribuição geométrica, 207
 - Distribuição Uniforme, 202
 - Propriedades, 201
- Variância condicional, 261
- Vetor aleatório, 224
- Vetores aleatórios, 224
 - Variáveis aleatórias independentes, 248