PROBABILIDADE

Aplicações à Estatística

PAUL L. MEYER

2ª edição

LIVROS TÉCNICOS E CIENTÍFICOS EDITORA

PROBABILIDADE

Aplicações à Estatística

PROBABILIDADE

Aplicações à Estatística

PUCRS
BIBLIOTECA CENTRAL

PAUL L. MEYER

519.2
M613pd
BC

Tradução
Ruy de C. B. Lourenço Filho
Professor da Universidade Federal do Rio de Janeiro e da Escola Nacional de Ciências Estatísticas (ENCE/IBGE)

2ª EDIÇÃO

PUCRS / BCE
0.528.348-4

LTC
EDITORA

BIBLIOTECA CENTRAL DA PUCRS
PRESERVE SUA FONTE DE CONHECIMENTO

1.ª edição: 1969 — Reimpressões: 1970, 1972, 1974, 1975, 1976 (duas), 1977,
1978 (duas), 1980, 1981 e 1982
2.ª edição: 1983 — Reimpressões: 1991, 1994 (duas), 1995, 1997, 1999 e 2000

Copyright © 1969 por Ao Livro Técnico
Título do original em inglês: Introductory Probability and Statistical Applications
Copyright © 1965 e 1969 por Addison-Wesley Publishing Company, Inc.
All rights reserved. Authorized translation from english language edition published by
Addison-Wesley Publishing Company, Inc.

Direitos exclusivos para a língua portuguesa
Copyright © 1983 by
LTC — Livros Técnicos e Científicos Editora S.A.
Travessa do Ouvidor, 11
Rio de Janeiro, RJ — CEP 20040-040
Tel.: 21–221-9621
Fax: 21–221-3202

Reservados todos os direitos. É proibida a duplicação ou
reprodução deste volume, no todo ou em parte,
sob quaisquer formas ou por quaisquer meios
(eletrônico, mecânico, gravação, fotocópia ou outros),
sem permissão expressa da Editora.

Para Alan e David

NOTA DA EDITORA

Temos por norma, nas traduções que editamos, converter as unidades para o sistema legal no Brasil.

No presente caso, abrimos uma exceção. O livro possui problemas nos sistemas inglês, CGS e MKS, que foram mantidos, a conselho de especialistas no assunto, visando a dar ao estudante maior flexibilidade, pela oportunidade de praticar nos diferentes sistemas.

A EDITORA

PREFÁCIO DA SEGUNDA EDIÇÃO

Devido ao considerável número de observações favoráveis que recebi durante os anos passados, tanto de alunos como de professores que empregaram a primeira edição deste livro, relativamente poucas alterações foram feitas. Durante a minha própria utilização repetida do livro, verifiquei que a organização básica do conteúdo e o nível geral de apresentação (por exemplo, a mistura de demonstrações matemáticas rigorosas com explanações mais informais e exemplos) estão bastante adequados ao tipo de estudante que se inscreve no curso.

No entanto, diversas modificações e acréscimos foram feitos. Antes de mais nada, foi realizado um esforço para eliminar os diversos enganos e erros de impressão que existiam na primeira edição. O autor é extremamente grato aos muitos leitores que não somente descobriram alguns deles, como foram bastante interessados em me apontá-los.

Em segundo lugar, foi feito um esforço para lançar maior esclarecimento na relação entre várias distribuições de probabilidade, de modo que o estudante pudesse alcançar maior compreensão de como diversos modelos probabilísticos podem ser empregados para com um obter aproximação de outro.

Finalmente, alguns problemas novos foram acrescentados à já extensa lista incluída na primeira edição.

O autor deseja agradecer, mais uma vez, à Addison-Wesley Publishing Company, pela sua cooperação em tudo quanto contribuiu para esta nova edição.

P. L. M.

Pullman, Washington

PREFÁCIO DA PRIMEIRA EDIÇÃO

Se temer que suspeitem ser sua narrativa inverídica, lembre-se da probabilidade.

JOHN GAY

Este livro é destinado a cursos de um semestre ou dois quadrimestres, de Introdução à Teoria da Probabilidade e algumas de suas aplicações. O pré-requisito é um ano de Cálculo Diferencial e Integral. Não se supõe qualquer conhecimento prévio de Probabilidade ou de Estatística. Na Washington State University, o curso, para o qual este livro foi preparado, vem sendo lecionado há alguns anos, principalmente a alunos orientados para a Engenharia ou às Ciências Naturais. A maioria desses alunos pode dedicar somente um semestre ao estudo desta matéria, porém, já que esses alunos estão familiarizados com o Cálculo, estão em condições de começar o estudo desta matéria por um nível além daquele estritamente elementar.

Muitos tópicos da Matemática podem ser apresentados em diferentes estágios de dificuldade, e isto é certamente verdade para a Probabilidade. Neste livro, faz-se um esforço para tirar proveito da base matemática do leitor, sem ultrapassá-la. Linguagem matemática rigorosa é empregada, mas toma-se o cuidado de não se ficar excessivamente mergulhado em minúcias matemáticas desnecessárias. Este não é, seguramente, um "livro de receitas". Muito embora alguns conceitos sejam introduzidos e explicados de maneira não formal, as definições e os teoremas são enunciados cuidadosamente. Quando uma demonstração pormenorizada de um teorema não é factível ou desejável, ao menos um esboço das idéias importantes é oferecido. Um traço peculiar deste livro são os "Comentários", que se seguem à maioria dos teoremas e definições. Nesses Comentários, o particular conceito ou resultado que esteja sendo apresentado é explicado de maneira intuitiva.

Em virtude da restrição que nos impusemos, de escrever um livro relativamente conciso sobre um domínio muito extenso, algumas escolhas tiveram de ser feitas, quanto à inclusão ou exclusão de determinados tópicos. Não parece existir maneira óbvia de resolver esta questão. Certamente, não sustento que para alguns dos tópicos excluídos, não se pudesse encontrar um lugar; nem pretendo que alguma parte da matéria não se pudesse omitir. Não obstante, para a maior parte dela, deu-se destaque às noções fundamentais, apresentadas bastante pormenorizadamente. Somente o Cap. 11, sobre confiabilidade, poderia ser considerado "artigo supérfluo", mas ainda aqui, sinto que as noções associadas às questões de confiabilidade são de interesse fundamental para muitas pessoas. Além disso, conceitos de confiabilidade constituem veículo excelente para se ilustrarem muitas das idéias anteriormente introduzidas ao livro.

Muito embora a cobertura seja limitada pelo tempo disponível, uma seleção razoavelmente ampla dos assuntos foi conseguida. De um exame rápido do Sumário, fica evidenciado que cerca de três quartos do livro são dedicados a assuntos probabilísticos, enquanto o último quarto é dedicado a uma explanação da Inferência Estatística. Apesar de nada haver de extraordinário nesta particular divisão de importância entre Probabilidade e Estatística, creio que um sólido conhecimento dos fundamentos da Probabilidade é obrigatório para uma compreensão adequada dos métodos estatísticos. Idealmente, um curso de Probabilidade deveria ser seguido de outro, de Teoria e Metodologia Estatísticas. No entanto, como já mencionei anteriormente, a maioria dos alunos que toma este curso não tem tempo para uma exposição de dois semestres nesse domínio e, por isso, senti-me compelido a explanar ao menos alguns dos mais importantes aspectos do tema geral da Inferência Estatística.

O sucesso potencial de determinada apresentação de um assunto não deve ser avaliado apenas em termos das idéias específicas aprendidas e das técnicas específicas adquiridas. A apreciação final deve também levar em conta quão bem o estudante ficará preparado para continuar seu estudo do assunto, seja por si mesmo, seja através do trabalho em um curso complementar. Se este critério for considerado importante, então se tornará evidente que os conceitos básicos e as técnicas fundamentais devam ser salientados, enquanto métodos e tópicos altamente especializados devam ser relegados a um papel secundário. Isto se torna também um importante fator na seleção de quais tópicos incluir.

A importância da teoria da Probabilidade é difícil de se exagerar. O modelo matemático apropriado para o estudo de um grande número de fenômenos observáveis é mais um modelo probabilístico do que um determinístico. Além disso, todo o assunto da Inferência Estatística é baseado em considerações probabilísticas. Técnicas estatísticas estão entre as mais importantes ferramentas dos cientistas e engenheiros. A fim de empregar inteligentemente essas técnicas, um profundo conhecimento dos conceitos probabilísticos é exigido.

Espera-se que, além dos vários métodos e conceitos específicos com os quais o leitor venha a se familiarizar, ele também desenvolva uma certa atitude: a de pensar probabilisticamente, substituindo questões como "Quanto tempo este componente funcionará?" por "Quão provável é que este componente funcione mais do que 100 horas?" Em muitas situações, a segunda questão poderá ser não somente a mais apropriada, mas de fato a única que tenha sentido fazer-se.

Tradicionalmente, muitos dos importantes conceitos de probabilidade são ilustrados com o auxílio de diferentes "jogos de azar"; jogadas de moedas ou dados, extração de cartas de um baralho, giração de uma roleta etc. Muito embora eu não tenha evitado inteiramente a referência a tais jogos, já que eles servem para ilustrar as noções fundamentais, um esforço foi feito para colocar o estudante em contato com ilustrações mais adequadas das aplicações da probabilidade: a emissão de partículas α por uma fonte radioativa, amostragem de lotes, duração da vida de dispositivos eletrônicos, e os problemas relacionados de confiabilidade de componentes e de sistemas etc.

Estou relutante em mencionar o mais óbvio traço de qualquer livro de Matemática: os problemas. E, no entanto, parece-me proveitoso salientar que a resolução de problemas deve ser considerada parte integrante do curso. Somente ao se tornar pessoalmente interessado em propor e resolver os exercícios, poderá realmente o estudante desenvolver uma compreensão e apreciação das idéias e uma familiaridade com as técnicas adequadas. Por isso, mais de 330 problemas foram incluídos no livro e, para mais de metade deles, respostas são dadas ao fim do livro. Além dos problemas propostos ao leitor, há numerosos exemplos resolvidos, espalhados através do livro.

Este livro foi escrito de maneira bem encadeada: o entendimento da maioria dos capítulos exige conhecimento profundo dos capítulos anteriores. Contudo, é possível examinar os Caps. 10 e 11 um tanto despreocupadamente, particularmente se alguém estiver interessado em dedicar mais tempo às aplicações estatísticas que são explanadas nos Caps. 13 a 15.

Tal como deve ser certo para qualquer um que escreva um livro, os débitos que tenho são muito numerosos: para com meus colegas, por muitas conversas estimulantes e úteis; para com meus próprios professores, pelo conhecimento e interesse neste assunto; para com os críticos das versões anteriores do manuscrito, por muitas sugestões e críticas úteis; para com a Addison-Wesley Publishing Company, por sua grande ajuda e cooperação desde as fases iniciais até o fim mesmo deste projeto; para com a Srª Carol Sloan, por ser uma datilógrafa eficiente e atenta; para com D. Van Nostrand, Inc., The Free Press, Inc. e Macmillan Publishing Company, por sua permissão para reproduzir as Tábuas 3, 6 e 1, respectivamente; para com McGraw-Hill Book Co., Inc., Oxford University Press Inc., Pergamon Press, Ltda. e Prentice-Hall, Inc., por suas permissões para citar determinados exemplos no texto, e, finalmente para com minha esposa, não somente por me amparar no esforço, como também por "deixar-me" e levar nossos dois filhos com ela a visitarem os avós, por dois cruciais meses de verão, durante os quais fui capaz de transformar nossa casa em uma desordenada, porém tranqüila oficina, da qual surgiu, miraculosamente, a última versão final deste livro.

P. L. M.

Pullman, Washington
Abril, 1965

SUMÁRIO

Introdução à Probabilidade

Capítulo 1

1.1. Modelos Matemáticos

Neste capítulo examinaremos o tipo de fenômeno que estudaremos por todo este livro. Além disso, formularemos um modelo matemático que nos ajudará a investigar, de maneira bastante precisa, esse fenômeno.

De início, é muito importante distinguir o próprio fenômeno e o modelo matemático para esse fenômeno. Naturalmente, não exercemos influência sobre aquilo que observamos. No entanto, ao escolher um modelo, podemos lançar mão de nosso julgamento crítico. Isto foi especialmente bem expresso pelo Prof. J. Neyman, que escreveu:*

> "Todas as vezes que empregarmos Matemática a fim de estudar alguns fenômenos de observação, deveremos essencialmente começar por construir um modelo matemático (determinístico ou probabilístico) para esses fenômenos. Inevitavelmente, o modelo deve simplificar as coisas e certos pormenores devem ser desprezados. O bom resultado do modelo depende de que os pormenores desprezados sejam ou não realmente sem importância na elucidação do fenômeno estudado. A resolução do problema matemático pode estar correta e, não obstante, estar em grande discordância com os dados observados, simplesmente porque as hipóteses básicas feitas não sejam confirmadas. Geralmente é bastante difícil afirmar com certeza se um modelo matemático especificado é ou não adequado, *antes* que alguns dados de observação sejam obtidos. A fim de verificar a validade de um modelo, deveremos *deduzir* um certo número de conseqüências de nosso modelo e, a seguir, comparar esses resultados *previstos* com observações."

Deveremos nos lembrar das idéias acima enquanto estivermos estudando alguns fenômenos de observação e modelos apropriados

* *University of California Publications in Statistics*, Vol. I, University of California Press, 1954.

para sua explicação. Vamos examinar, inicialmente, o que se pode adequadamente denominar *modelo determinístico*. Por essa expressão pretendemos nos referir a um modelo que estipule que as condições sob as quais um experimento seja executado *determinem* o resultado do experimento. Por exemplo, se introduzirmos uma bateria em um circuito simples, o modelo matemático que, presumivelmente, descreveria o fluxo de corrente elétrica observável seria $I = E/R$, isto é, a Lei de Ohm. O modelo prognostica o valor de I tão logo os valores de E e R sejam fornecidos. Dizendo de outra maneira, se o experimento mencionado for repetido um certo número de vezes, toda vez utilizando-se o mesmo circuito (isto é, conservando-se fixados E e R), poderemos presumìvelmente esperar observar o mesmo valor para I. Quaisquer desvios que pudessem ocorrer seriam tão pequenos que, para a maioria das finalidades, a descrição acima (isto é, o modelo) seria suficiente. O importante é que a bateria, fio, e amperômetro particulares utilizados para gerar e observar a corrente elétrica, e a nossa capacidade de empregar o instrumento de medição, determinam o resultado em cada repetição. (Existem determinados fatores que bem poderão ser diferentes de repetição para repetição, que, no entanto, não influenciarão de modo digno de nota o resultado. Por exemplo, a temperatura e a umidade no laboratório, ou a estatura da pessoa que lê o amperômetro, pode-se razoavelmente admitir, não terão influência no resultado.)

Na natureza, existem muitos exemplos de "experimentos", para os quais modelos determinísticos são apropriados. Por exemplo, as leis da gravitação explicam bastante precisamente o que acontece a um corpo que cai sob determinadas condições. As leis de Kepler nos dão o comportamento dos planetas. Em cada situação, o modelo especifica que as condições, sob as quais determinado fenômeno acontece, determinam o valor de algumas variáveis observáveis: a *grandeza* da velocidade, a *área* varrida durante determinado período de tempo etc. Esses números aparecem em muitas das fórmulas com as quais estamos familiarizados. Por exemplo, sabemos que, sob determinadas condições, a distância percorrida (verticalmente, acima do solo) por um objeto é dada por $s = -16t^2 + v_0t$, onde v_0 é a velocidade inicial e t o tempo gasto na queda. O ponto, no qual desejamos fixar nossa atenção, não é a forma particular da equação acima (que é quadrática), mas antes o fato de que existe uma relação definida entre t e s, a qual determina univocamente a quantidade no primeiro membro da equação, se aquelas no segundo membro forem fornecidas.

Para um grande número de situações, o modelo matemático determinístico apresentado acima é suficiente. Contudo, existem também muitos fenômenos que requerem um modelo matemático diferente para sua investigação. São os que denominaremos modelos *não-determinísticos* ou *probabilísticos*. (Outra expressão muito comumente empregada é modelo *estocástico*.) Mais adiante neste capítulo, estudaremos muito minuciosamente, como tais modelos probabilísticos podem ser apresentados. Por ora, examinaremos alguns exemplos.

Suponhamos que se tenha um fragmento de material radioativo que emita partículas *alfa*. Com o auxílio de um dispositivo de contagem, poderemos registrar o número dessas partículas emitidas durante um intervalo de tempo especificado. É evidente que não poderemos antecipar precisamente o número de partículas emitidas, ainda que se conheçam de modo exato a forma, a dimensão, a composição química e a massa do objeto em estudo. Por isso, parece não existir modelo determinístico razoável que forneça o número de partículas emitidas, por exemplo n, como uma função de várias características pertinentes ao material fonte. Deveremos considerar, em seu lugar, um modelo probabilístico.

Como outro exemplo, considere-se a seguinte situação meteorológica. Deseja-se determinar qual a precipitação de chuva que cairá como resultado de uma tempestade particular, que ocorra em determinada localidade. Dispõe-se de instrumentos para registrar a precipitação. Observações meteorológicas podem nos fornecer considerável informação relativa à tempestade que se avizinhe: pressão barométrica em vários pontos, variações de pressão, velocidade do vento, origem e direção da tormenta, e várias leituras referentes a altitudes elevadas. Contudo, quão valiosas essas informações possam ser para o prognóstico da natureza geral da precipitação (digamos, fraca, média ou forte), simplesmente não tornam possível dizer-se *quanta* chuva irá cair. Novamente estaremos nos ocupando de um fenômeno que não se presta a um tratamento determinístico. Um modelo probabilístico explica a situação mais rigorosamente.

Em princípio, poderemos ser capazes de dizer quanta chuva caiu se uma teoria tiver sido desenvolvida (o que não foi). Por isso, empregaremos um modelo probabilístico. No exemplo que trata de desintegração radioativa, deveremos empregar um modelo probabilístico *invariavelmente em princípio.*

Arriscando-nos a adiantarmos demais na apresentação de um conceito que será definido posteriormente, vamos apenas afirmar que, em um modelo determinístico, admite-se que o resultado efetivo (numérico ou de outra espécie) seja determinado pelas condições sob as quais o experimento ou o procedimento seja executado. Em um modelo não-determinístico, no entanto, as condições da experimentação determinam somente o comportamento probabilístico (mais especificamente, a lei probabilística) do resultado observável.

Em outras palavras, em um modelo determinístico empregamos "considerações físicas" para prever o resultado, enquanto em um modelo probabilístico empregamos a mesma espécie de considerações para especificar uma distribuição de probabilidade.

1.2. Introdução aos Conjuntos

A fim de expor os conceitos básicos do modelo probabilístico que desejamos desenvolver, será conveniente conhecer algumas idéias e conceitos da teoria matemática dos conjuntos. Este é um assunto dos mais extensos e muito se tem escrito sobre ele. Contudo, necessitaremos apenas de algumas noções fundamentais.

Um *conjunto* é uma coleção de objetos. Usualmente, conjuntos são representados por letras maiúsculas A, B etc. Existem três maneiras de descrever que objetos estão contidos no conjunto A:

(*a*) Poderemos fazer uma lista dos elementos de A. Por exemplo, $A = \{1, 2, 3, 4\}$ descreve o conjunto formado pelos inteiros positivos 1, 2, 3, 4.

(*b*) Poderemos descrever o conjunto A por meio de palavras. Por exemplo, poderemos dizer que A é formado de todos os números reais entre 0 e 1, inclusive.

(*c*) Para descrever o conjunto acima poderemos simplesmente escrever $A = \{x | 0 \leq x \leq 1\}$; isto é, A é o conjunto de todos os x, onde x é um número real entre 0 e 1, inclusive.

Os objetos que individualmente formam a coleção ou conjunto A são denominados *membros* ou *elementos* de A. Quando "a" for um elemento de A, escreveremos $a \in A$, e quando "a" não for um elemento de A, escreveremos $a \notin A$.

Existem dois conjuntos especiais que, freqüentemente, nos interessarão. Em muitos problemas nos dedicaremos a estudar um conjunto definido de objetos, e não outros. Por exemplo, poderemos nos interessar por todos os números reais, por todas as peças que saem de uma linha de produção durante um período de 24 horas etc. Definiremos o *conjunto fundamental* como o conjunto de todos os objetos que estejam sendo estudados. Este conjunto é, geralmente representado pela letra U.

O outro conjunto que deve ser destacado pode surgir da seguinte maneira: Suponha-se que o conjunto A seja descrito como o conjunto de todos os números *reais* x, que satisfaçam à equação $x^2 + 1 = 0$. Naturalmente, sabemos que não existem tais números; isto é, o conjunto A não contém qualquer elemento. Esta situação ocorre tão freqüentemente que se justifica a introdução de um nome especial para esse conjunto. Por isso, definiremos o conjunto *vazio* ou *nulo* como o conjunto que não contenha qualquer elemento. Geralmente se representa esse conjunto por $\emptyset$.

Pode acontecer que, quando dois conjuntos A e B sejam considerados, ser elemento de A implique ser elemento de B. Nesse caso, diremos que A é um *subconjunto* de B, e escreveremos $A \subset B$. Interpretação semelhante será dada para $B \subset A$. Diremos que dois conjuntos constituem o mesmo conjunto, $A = B$, se, e somente se, $A \subset B$ e $B \subset A$. Desse modo, dois conjuntos serão *iguais* se, e somente se, eles contiverem os mesmos elementos.

As duas seguintes propriedades do conjunto vazio e do conjunto fundamental são imediatas:

(*a*) Para todo conjunto A, temos que $\emptyset \subset A$.

(*b*) Desde que se tenha definido o conjunto fundamental, então, para todo conjunto A, considerado na composição de U, teremos $A \subset U$.

Exemplo 1.1. Suponha-se que U = todos os números reais, $A = \{x | x^2 + 2x - 3 = 0\}$, $B = \{x | (x - 2)(x^2 + 2x - 3) = 0\}$ e $C = \{x | x = -3, 1, 2\}$. Então, $A \subset B$ e $B = C$.

A seguir, estudaremos a importante idéia de *combinar* conjuntos dados, a fim de formarmos um novo conjunto. Há duas operações fundamentais, e essas operações se assemelham, em certos aspectos, às operações de adição e multiplicação de números. Sejam dois conjuntos A e B.

Definiremos C como a *união* de A e B (algumas vezes denominada a soma de A e B), da seguinte maneira:

$$C = \{x | x \in A \text{ ou } x \in B \text{ (ou ambos)}\}.$$

Escreveremos a união de A e B, assim: $C = A \cup B$. Desse modo, C será formado de todos os elementos que estejam em A, ou em B, ou em ambos.

Definiremos D como a *interseção* de A e B (algumas vezes denominada o produto de A e B), da seguinte maneira:

$$D = \{x \mid x \in A \text{ e } x \in B\}.$$

Escreveremos a interseção de A e B, assim: $D = A \cap B$. Portanto, D será formado de todos os elementos que estão em A e em B.

Finalmente, introduziremos a noção de *complemento* de um conjunto A, na forma seguinte: O conjunto denotado por $\bar{A}$, constituído por todos os elementos que *não* estejam em A (mas que estejam no conjunto fundamental U) é denominado complemento de A. Isto é, $\bar{A} = \{x \mid x \notin A\}$.

Um recurso gráfico, conhecido como *Diagrama de Venn*, poderá ser vantajosamente empregado quando estivermos combinando conjuntos, na maneira indicada acima. Em cada diagrama na Fig. 1.1, a região *sombreada* representa o conjunto sob exame.

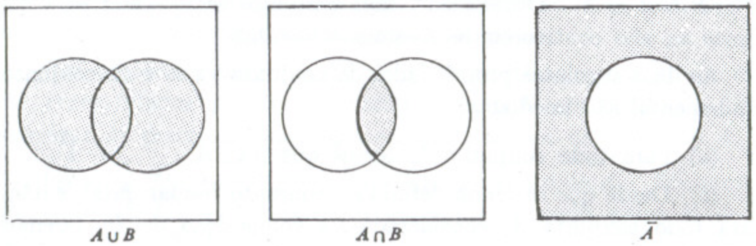


Fig. 1.1

Exemplo 1.2. Suponha-se que $U = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$, $A = \{1, 2, 3, 4\}$, $B = \{3, 4, 5, 6\}$. Então, encontraremos que $\bar{A} = \{5, 6, 7, 8, 9, 10\}$, $A \cup B = \{1, 2, 3, 4, 5, 6\}$ e $A \cap B = \{3, 4\}$. Observe-se que, ao descrever um conjunto (tal como $A \cup B$), cada elemento é relacionado apenas uma vez.

As operações de união e interseção, definidas acima para dois conjuntos, podem ser estendidas, intuitivamente, para qualquer número finito de conjuntos. Assim, definiremos $A \cup B \cup C$ como $A \cup (B \cup C)$ ou $(A \cup B) \cup C$, o que é equivalente, como se poderá verificar facilmente. De modo análogo, definiremos $A \cap B \cap C$ como sendo $A \cap (B \cap C)$ ou $(A \cap B) \cap C$, o que também se pode verificar serem equivalentes. É evidente que poderemos continuar essas composições de conjuntos para *qualquer número finito* de conjuntos dados.

Afirmamos que alguns conjuntos são equivalentes, por exemplo, $A \cap (B \cap C)$ e $(A \cap B) \cap C$. Conclui-se que existe um certo número de tais conjuntos *equivalentes*, alguns dos quais estão relacionados abaixo. Se nos lembrarmos de que dois conjuntos são o mesmo conjunto sempre que eles contenham os mesmos elementos, será fácil mostrar que as afirmações feitas são verdadeiras. O leitor poderá se convencer disso, com a ajuda dos Diagramas de Venn.

$$(a)\ A \cup B = B \cup A, \qquad (b)\ A \cap B = B \cap A, \tag{1.1}$$
$$(c)\ A \cup (B \cup C) = (A \cup B) \cup C, \quad (d)\ A \cap (B \cap C) = (A \cap B) \cap C.$$

Denominaremos (*a*) e (*b*) *leis comutativas*, e (*c*) e (*d*) *leis associativas*.

Há outras *identidades de conjuntos* encerrando união, interseção e complementação. As mais importantes delas estão enumeradas a seguir. A validade de cada uma delas poderá ser verificada com a ajuda de um Diagrama de Venn.

$$(e)\ A \cup (B \cap C) = (A \cup B) \cap (A \cup C),$$
$$(f)\ A \cap (B \cup C) = (A \cap B) \cup (A \cap C),$$
$$(g)\ A \cap \emptyset = \emptyset, \tag{1.2}$$
$$(h)\ A \cup \emptyset = A, \qquad (i)\ \overline{(A \cup B)} = \bar{A} \cap \bar{B},$$
$$(j)\ \overline{(A \cap B)} = \bar{A} \cup \bar{B}, \qquad (k)\ \bar{\bar{A}} = A.$$

Observe-se que (*g*) e (*h*) mostram que $\emptyset$ se comporta entre os conjuntos (relativamente às operações $\cup$ e $\cap$) da maneira que o número zero (com relação às operações de adição e multiplicação) o faz entre os números.

Uma outra maneira de formar conjuntos, quando forem dados dois (ou mais) conjuntos, será necessária a seguir.

Definição. Sejam dois conjuntos A e B. Denominaremos *produto cartesiano* de A e B, denotando-o por $A \times B$, o conjunto $\{(a, b), a \in A, b \in B\}$, isto é, o conjunto de todos os pares ordenados nos quais o primeiro elemento é tirado de A e o segundo, de B.

Exemplo 1.3. Suponha-se que $A = \{1, 2, 3\}$; $B = \{1, 2, 3, 4\}$. Então, $A \times B = \{(1, 1), (1, 2), \ldots, (1, 4), (2, 1) \ldots, (2, 4), (3, 1), \ldots, (3, 4)\}$.

Observação. Em geral, $A \times B \neq B \times A$.

A noção acima pode ser estendida da seguinte maneira: Se $A_1, \ldots, A_n$ forem conjuntos, então, $A_1 \times A_2 \times \ldots \times A_n = \{(a_1, a_2, \ldots a_n), a_i \in A_i\}$, ou seja, o conjunto de todas as ênuplas ordenadas.

Um caso especial importante surge quando consideramos o produto cartesiano de um conjunto por ele próprio, isto é, $A \times A$ ou $A \times A \times A$. Exemplos disso surgem quando tratamos do plano euclideano, $R \times R$, onde R é o conjunto de todos os números reais, e do espaço euclideano tridimensional, representado por $R \times R \times R$.

O *número de elementos* de um conjunto terá grande importância para nós. Se existir um número finito de elementos no conjunto A, digamos $a_1, a_2, \ldots, a_n$, diremos que A é *finito*. Se existir um número infinito de elementos em A, os quais possam ser postos em *correspondência biunívoca* com os inteiros positivos, diremos que A é *numerável* ou *infinito numerável*. (Pode-se mostrar, por exemplo, que o conjunto de todos os números racionais é numerável.) Finalmente, deveremos considerar o caso de um conjunto infinito não-numerável; este tipo de conjunto possui um número infinito de elementos que não podem ser enumerados. Pode-se mostrar, por exemplo, que para quaisquer dois números reais $b > a$, o conjunto $A = \{x \mid a \leq x \leq b\}$ contém um número não-numerável de elementos. Já que poderemos associar um ponto da reta dos números reais a cada número real, o que dissemos acima afirma que qualquer intervalo (não degenerado) contém mais do que um número contável de pontos.

Os conceitos apresentados acima, muito embora representem apenas um rápido exame da teoria dos conjuntos, são suficientes para nossos objetivos: expor, com razoável rigor e precisão, as idéias fundamentais da teoria da probabilidade.

1.3. Exemplos de Experimentos Não-Determinísticos

Estamos agora em condições de examinar o que entendemos por um experimento "aleatório" ou "não-determinístico". (Mais precisamente, daremos exemplos de fenômenos, para os quais modelos não-determinísticos são apropriados. Esta é uma distinção que o leitor deverá guardar. Portanto, nos referiremos freqüentemente a experimentos não-determinísticos ou aleatórios, quando de fato estaremos falando de um *modelo* não-determinístico para um experimento.) Não nos esforçaremos em dar uma definição precisa deste conceito. Em vez disso, citaremos um grande número de exemplos que ilustrarão o que temos em mente.

E_1: Jogue um dado e observe o número mostrado na face de cima.

E_2: Jogue uma moeda quatro vezes e observe o número de caras obtido.

E_3: Jogue uma moeda quatro vezes e observe a seqüência obtida de caras e coroas.

E_4: Em uma linha de produção, fabrique peças em série e conte o número de peças defeituosas produzidas em um período de 24 horas.

E_5: Uma asa de avião é fixada por um grande número de rebites. Conte o número de rebites defeituosos.

E_6: Uma lâmpada é fabricada. Em seguida é ensaiada quanto à duração da vida, pela colocação em um soquete e anotação do tempo decorrido (em horas) até queimar.

E_7: Um lote de 10 peças contém 3 defeituosas. As peças são retiradas uma a uma (sem reposição da peça retirada) até que a última peça defeituosa seja encontrada. O número total de peças retiradas do lote é contado.

E_8: Peças são fabricadas até que 10 peças perfeitas sejam produzidas. O número total de peças fabricadas é contado.

E_9: Um míssil é lançado. Em um momento especificado t, suas três velocidades componentes, v_x, v_y e v_z são observadas.

E_{10}: Um míssil récem-lançado é observado nos instantes $t_1, t_2, \ldots, t_n$. Em cada um desses instantes, a altura do míssil acima do solo é registrada.

E_{11}: A resistência à tração de uma barra metálica é medida.

E_{12}: De uma urna, que só contém bolas pretas, tira-se uma bola e verifica-se sua cor.

E_{13}: Um termógrafo registra a temperatura continuamente, num período de 24 horas. Em determinada localidade e em uma data especificada, esse termógrafo é lido.

E_{14}: Na situação descrita em E_{13}, x e y, as temperaturas mínima e máxima, no período de 24 horas considerado, são registradas.

O que os experimentos acima têm em comum? Os seguintes traços são pertinentes à nossa caracterização de um *experimento aleatório*:

(*a*) Cada experimento poderá ser repetido indefinidamente sob condições essencialmente inalteradas.

(*b*) Muito embora não sejamos capazes de afirmar que resultado *particular* ocorrerá, seremos capazes de descrever o conjunto de *todos os possíveis* resultados do experimento.

(*c*) Quando o experimento for executado repetidamente, os resultados individuais parecerão ocorrer de uma forma acidental. Contudo, quando o experimento for repetido um *grande* número de vezes, uma configuração definida ou regularidade surgirá. É esta regularidade que torna possível construir um modelo matemático preciso, com o qual se analisará o experimento. Mais tarde, teremos muito que dizer sobre a natureza e a importância desta regularidade. Por ora, o leitor necessita apenas pensar nas repetidas jogadas de uma moeda equilibrada. Muito embora caras e coroas apareçam sucessivamente, em uma maneira quase arbitrária, é fato empírico bem conhecido que, depois de um grande número de jogadas, a proporção de caras e a de coroas serão aproximadamente iguais.

Deve-se salientar que todos os experimentos descritos acima satisfazem a essas características gerais. (Evidentemente, a última característica mencionada somente pode ser verificada pela experimentação; deixaremos para a intuição do leitor acreditar que se o experimento fosse repetido um grande número de vezes, a regularidade referida seria evidente. Por exemplo, se um grande número de lâmpadas, de um mesmo fabricante, fosse ensaiado, presumivelmente o número de lâmpadas que se queimaria após 100 horas poderia ser previsto com precisão considerável.) Note-se que o experimento E_{12} apresenta o traço peculiar de que somente um resultado é possível. Em geral, tais experimentos não nos interessarão, porque, realmente, o fato de não sabermos qual particular resultado virá a ocorrer, quando um experimento for realizado, é que torna um experimento interessante para nós.

Comentário: Ao descrever os diversos experimentos, nós especificamos não somente o procedimento que tem que ser realizado, mas também aquilo que estaremos interessados em observar (veja, por exemplo, a diferença entre E_2 e E_3, citados anteriormente). Este é um ponto muito importante, ao qual novamente nos referiremos mais tarde, quando explicarmos variáveis aleatórias. Por ora, vamos apenas comentar que, em conseqüência de um procedimento experimental isolado ou a ocorrência de um fenômeno único, *muitos* valores numéricos diferentes poderiam ser calculados. Por exemplo, se uma pessoa for escolhida de um grupo grande de pessoas (e a escolha real seria o procedimento experimental previamente mencionado), poderíamos estar interessados na altura daquela pessoa, no seu peso, na sua renda anual, no número de filhos dela etc. Naturalmente, na maioria dos casos, nós saberemos, antes de iniciar nossa experimentação, quais serão as características numéricas em que iremos estar interessados.

1.4. O Espaço Amostral

Definição. Para cada experimento $\mathcal{E}$ do tipo que estamos considerando, definiremos o *espaço amostral* como o conjunto de todos os resultados possíveis de $\mathcal{E}$. Geralmente representaremos esse conjunto por S. (Neste contexto, S representa o conjunto fundamental, explicado anteriormente.)

Vamos considerar cada um dos experimentos acima e descrever um espaço amostral para cada um deles. O espaço amostral S_i se referirá ao experimento E_i.

S_1: $\{1, 2, 3, 4, 5, 6\}$.

S_2: $\{0, 1, 2, 3, 4\}$.

S_3: {todas as seqüências possíveis da forma a_1, a_2, a_3, a_4}, onde cada $a_i = H$ ou T, conforme apareça cara ou coroa na i-ésima jogada.

S_4: $\{0, 1, 2, \ldots, N\}$, onde N é o número máximo que pode ser produzido em 24 horas.

S_5: $\{0, 1, 2, \ldots, M\}$, onde M é o número de rebites empregado.

S_6: $\{t \mid t \geq 0\}$.

S_7: $\{3, 4, 5, 6, 7, 8, 9, 10\}$.

S_8: $\{10, 11, 12, \ldots\}$.

S_9: $\{v_x, v_y, v_z \mid v_x, v_y, v_z$ números reais$\}$.

S_{10}: $\{h_1, \ldots, h_n \mid h_i \geq 0,\ i = 1, 2, \ldots, n\}$.

S_{11}: $\{T \mid T \geq 0\}$.

S_{12}: {bola preta}.

S_{13}: Este espaço amostral é o mais complexo de todos os considerados aqui. Podemos admitir, com realismo, que a temperatura em determinada localidade nunca possa ocorrer acima ou abaixo de certos valores M e m. Afora esta restrição, poderemos aceitar a possibilidade de que qualquer gráfico apareça com determinadas restrições. Presumivelmente, o gráfico não terá saltos (isto é, ele representará uma função contínua). Além disso, o gráfico terá certas características de regularização, que podem ser resumidas matematicamente dizendo-se que o gráfico representa uma função derivável. Deste modo, poderemos finalmente afirmar que o espaço amostral será:

$$\{f \mid f \text{ uma função derivável, que satisfaça a } m \leq f(t) \leq M, \text{ para todo } t\}.$$

S_{14}: $\{(x, y) | m \leq x \leq y \leq M\}$. Isto é, S_{14} é formado por todos os pontos dentro e sobre um triângulo, no plano x, y bidimensional.

(Neste livro não cuidaremos de espaços amostrais da complexidade encontrada em S_{13}. No entanto, tais espaços amostrais podem surgir, mas exigem para seu estudo mais Matemática avançada do que estamos admitindo aqui.)

A fim de descrever um espaço amostral associado a um experimento, devemos ter uma idéia bastante clara daquilo que estamos mensurando ou observando. Por isso, devemos falar de "um" espaço amostral associado a um experimento, e não de "o" espaço amostral. A esse respeito, note-se a diferença entre S_2 e S_3.

Saliente-se, também, que o resultado de um experimento não é necessariamente, um número. Por exemplo, em E_3, cada resultado é uma seqüência de caras (H) e coroas (T). Em E_9 e E_{10} cada resultado é formado por um vetor, enquanto em E_{13}, cada resultado constitui uma função.

Será também importante estudar o *número* de resultados em um espaço amostral. Surgem três possibilidades: o espaço amostral pode ser finito, infinito numerável, ou infinito não-numerável. Relativamente aos exemplos acima, observamos que S_1, S_2, S_3, S_4, S_5, S_7 e S_{12} são finitos, S_8 é infinito numerável, e S_6, S_9, S_{10}, S_{11}, S_{13} e S_{14} são infinitos não-numeráveis.

Neste ponto poderá ser valioso comentar a diferença entre um espaço amostral "idealizado" matematicamente e um espaço realizável experimentalmente. Com este objetivo, consideremos o experimento E_6 e seu espaço amostral associado S_6. É evidente que, quando estivermos realmente registrando o tempo total t, durante o qual uma lâmpada funcione, seremos "vítimas" da precisão de nosso instrumento de medir. Suponha-se que temos um instrumento que seja capaz de registrar o tempo com duas casas decimais, por exemplo, 16,43 horas. Com esta restrição imposta, nosso espaço amostral se tornará *infinito numerável*: $\{0{,}00,\ 0{,}01,\ 0{,}02, \ldots\}$. Além disso, é bastante próximo da realidade admitir que nenhuma lâmpada possa durar mais do que H horas, onde H pode ser um número muito grande. Conseqüentemente, parece que se formos completamente realistas na descrição deste espaço amostral, estaremos realmente tratando com um espaço amostral *finito*: $\{0{,}00,\ 0{,}01,\ 0{,}02, \ldots, H\}$. O número total de resultados seria $(H/0{,}01) + 1$, que poderá ser um número muito grande, mesmo que H seja moderadamente grande, por exemplo se $H = 100$. Torna-se bem mais simples e, matematicamente, conveniente, admitir que *todos* os valores de $t \geq 0$ sejam resultados possíveis e, portanto, tratar o espaço amostral S_6 tal como foi originalmente definido.

Diante desses comentários, alguns dos espaços amostrais descritos são idealizados. Em todas as situações subseqüentes, o espaço amostral considerado será aquele que for matematicamente mais conveniente. Na maioria dos problemas, pouca dúvida surge quanto à escolha adequada do espaço amostral.

1.5. Eventos

Outra noção fundamental é o conceito de *evento*. Um evento A (relativo a um particular espaço amostral S, associado a um experimento $\mathcal{E}$) é simplesmente um conjunto de resultados possíveis. Na terminologia dos conjuntos, um evento é um *subconjunto* de um espaço amostral S. Considerando nossa exposição anterior, isto significa que o próprio S constitui um evento, bem como o é o conjunto vazio $\emptyset$. Qualquer resultado individual pode também ser tomado como um evento.

Alguns exemplos de eventos são dados a seguir. Novamente, nos referimos aos experimentos relacionados acima: A_i se referirá ao evento associado ao experimento E_i:

A_1: Um número par ocorre, isto é, $A_1 = \{2, 4, 6\}$.

A_2: $\{2\}$; isto é, duas caras ocorrem.

A_3: $\{HHHH, HHHT, HHTH, HTHH, THHH\}$; isto é, mais caras do que coroas ocorreram.

A_4: $\{0\}$; isto é, todas as peças são perfeitas.

A_5: $\{3, 4, \ldots, M\}$; isto é, mais do que dois rebites eram defeituosos.

A_6: $\{t \mid t < 3\}$; isto é, a lâmpada queima em menos de 3 horas.

A_{14}: $\{(x, y) \mid y = x + 20\}$; isto é, a temperatura máxima é 20° maior do que a mínima.

Quando o espaço amostral S for finito ou infinito numerável, *todo* subconjunto poderá ser considerado um evento. [Constitui um exercício fácil de provar, e o faremos resumidamente, que se S contiver n elementos, existirão exatamente 2^n subconjuntos (eventos).] Contudo, se S for infinito não-numerável, surgirá uma dificuldade teórica. Verifica-se que *nem* todo subconjunto imaginável poderá ser considerado um evento. Determinados subconjuntos "não admissíveis" deverão ser excluídos por motivos que ultrapassam o nível desta explanação. Felizmente, tais subconjuntos não-admissíveis não surgem nas aplicações e, por isso, não cuidaremos deles aqui. Na exposição subseqüente, será admitido tacitamente que sempre que nos referirmos a um evento, ele será da espécie que já admitimos considerar.

Agora, poderemos empregar as várias técnicas de combinar conjuntos (isto é, eventos) e obter novos conjuntos (isto é, eventos), os quais já apresentamos anteriormente.

(*a*) Se A e B forem eventos, $A \cup B$ será o evento que ocorrerá se, e somente se, A ou B (ou ambos) ocorrerem.

(*b*) Se A e B forem eventos, $A \cap B$ será o evento que ocorrerá se, e somente se, A e B ocorrerem.

(*c*) Se A for um evento, $\overline{A}$ será o evento que ocorrerá se, e somente se, *não* ocorrer A.

(*d*) Se $A_1, \ldots, A_n$ for qualquer coleção finita de eventos, então, $\bigcup_{i=1}^{n} A_i$ será o evento que ocorrerá se, e somente se, *ao menos* um dos eventos A_i ocorrer.

(*e*) Se $A_1, \ldots, A_n$ for qualquer coleção finita de eventos, então $\bigcap_{i=1}^{n} A_i$ será o evento que ocorrerá se, e somente se, *todos* os eventos A_i ocorrerem.

(*f*) Se $A_1, \ldots, A_n, \ldots$ for qualquer coleção infinita (numerável) de eventos, então, $\bigcup_{i=1}^{\infty} A_i$ será o evento que ocorrerá se, e somente se, *ao menos* um dos eventos A_i ocorrer.

(*g*) Se $A_1, \ldots, A_n, \ldots$ for qualquer coleção infinita (numerável) de eventos, então, $\bigcap_{i=1}^{\infty} A_i$ será o evento que ocorrerá se, e somente se, *todos* os eventos A_i ocorrerem.

(*h*) Suponha-se que S represente o espaço amostral associado a algum experimento $\mathcal{E}$, e que nós executemos $\mathcal{E}$ duas vezes. Então, $S \times S$ poderá ser empregado para representar todos os resultados dessas duas repetições. Portanto, $(s_1, s_2) \in S \times S$ significa que s_1 resultou quando $\mathcal{E}$ foi executado a primeira vez e s_2, quando $\mathcal{E}$ foi executado a segunda vez.

(*i*) O exemplo contido em (*h*) pode, obviamente, ser generalizado. Considerem-se n repetições de um experimento $\mathcal{E}$ cujo espaço amostral seja S:

$$S \times S \times \ldots \times S = \{(s_1, s_2, \ldots, s_n),\ s_i \in S,\ i = 1, \ldots, n\}$$

representa o conjunto de todos os possíveis resultados, quando $\mathcal{E}$ for executado n vezes. De certo modo, $S \times S \times \ldots \times S$ é ele próprio um espaço amostral, a saber, o espaço amostral associado a n repetições de $\mathcal{E}$.

Definição. Dois eventos, A e B, são denominados *mutuamente excludentes*, se eles não puderem ocorrer juntos. Exprimiremos isso escrevendo $A \cap B = \emptyset$, isto é, a interseção de A e B é o conjunto vazio.

Exemplo 1.4. Um dispositivo eletrônico é ensaiado e o tempo total de serviço t é registrado. Admitiremos que o espaço amostral seja $\{t \mid t \geq 0\}$. Sejam A, B e C três eventos definidos da seguinte maneira:

$$A = \{t \mid t < 100\}; \quad B = \{t \mid 50 \leq t \leq 200\}; \quad C = \{t \mid t > 150\}.$$

Conseqüentemente,

$$A \cup B = \{t \mid t \leq 200\}; \quad A \cap B = \{t \mid 50 \leq t < 100\};$$
$$B \cup C = \{t \mid t \geq 50\}; \quad B \cap C = \{t \mid 150 < t \leq 200\}; \quad A \cap C = \emptyset;$$
$$A \cup C = \{t \mid t < 100 \text{ ou } t > 150\}; \quad \overline{A} = \{t \mid t \geq 100\}; \quad \overline{C} = \{t \mid t \leq 150\}.$$

Uma das características fundamentais do conceito de "experimento", como foi apresentado na seção anterior, é que nós não sabemos qual resultado particular ocorrerá quando o experimento for realizado. Por outras palavras, se A for um evento associado a um experimento, então, não poderemos afirmar com certeza que A irá ocorrer ou não. Por isso, torna-se muito importante tentar associar um número ao evento A, o qual medirá de alguma maneira quão verossímil é que o evento A venha a ocorrer. Essa tarefa nos leva à teoria da probabilidade.

1.6. Freqüência Relativa

A fim de motivar a maneira de tratar o assunto, considere-se o seguinte procedimento: Suponha-se que repetimos n vezes o experimento $\mathcal{E}$, e sejam A e B dois eventos associados a $\mathcal{E}$. Admitamos que sejam, respectivamente, n_A e n_B o número de vezes que o evento A e o evento B ocorram nas n repetições.

Definição. $f_A = n_A/n$ é denominada *freqüência relativa* do evento A nas n repetições de $\mathcal{E}$. A freqüência relativa f_A apresenta as seguintes propriedades, de fácil verificação:

(1) $0 \leq f_A \leq 1$.

(2) $f_A = 1$ se, e somente se, A ocorrer em *todas* as n repetições.

(3) $f_A = 0$ se, e somente se, A *nunca* ocorrer nas n repetições.

(4) Se A e B forem eventos mutuamente excludentes, e se $f_{A \cup B}$ for a freqüência relativa associada ao evento $A \cup B$, então, $f_{A \cup B} = f_A + f_B$.

(5) f_A, com base em n repetições do experimento e considerada como uma função de n, "converge" em certo sentido probabilístico para $P(A)$, quando $n \to \infty$.

Comentário: A Propriedade (5) está evidentemente expressada um tanto vagamente, nesta seção. Somente mais tarde (Seç. 12.2), estaremos aptos a tornar esta idéia mais precisa. Por enquanto, podemos apenas afirmar que a Propriedade (5) envolve a noção nitidamente intuitiva de que a freqüência relativa, baseada em um número crescente de observações, tende a se "estabilizar" próximo de algum valor definido. Este *não* é o mesmo conceito usual de convergência encontrado em alguma parte da Matemática. De fato, tal como afirmamos aqui, esta não é, de modo algum, uma conclusão matemática, mas apenas um fato empírico.

A maioria de nós está intuitivamente a par deste fenômeno de estabilização, muito embora nunca o tenhamos verificado. Fazê-lo exige considerável porção de tempo e de paciência, porque inclui um grande número de repetições de um experimento. Contudo, algumas vezes, poderemos ser ingênuos observadores deste fenômeno, como ilustra o seguinte exemplo:

Exemplo 1.5. Admitamos que estejamos postados na calçada e fixemos nossa atenção em dois blocos de meio-fio adjacentes. Suponha-se que comece a chover de tal maneira que sejamos realmente capazes de distinguir pingos isolados de chuva e registrar se esses pingos caem num meio-fio ou noutro. Ficamos a observar os pingos e a anotar seu ponto de impacto. Denotando o i-ésimo pingo por X_i, onde $X_i = 1$ se o pingo cair no primeiro meio-fio, e igual a 0 se cair no outro, poderemos observar uma seqüência como, por exemplo, 1, 1, 0, 1, 0, 0, 0, 1, 0, 0, 1. É evidente que não seremos capazes de prever onde um particular pingo irá cair. (Nosso experimento consta de alguma espécie de situação meteorológica que causa a queda dos pingos de chuva.) Se calcularmos a freqüência relativa do evento $A = \{$o pingo cai no meio-fio 1$\}$, então, a seqüência de resultados acima dará origem às seguintes freqüências relativas (baseadas na observação de 1, 2, 3, . . . pingos): 1, 1, 2/3, 3/4, 3/5, 3/6, 3/7, 4/8, 4/9, 4/10, 5/11, . . . Esses números evidenciam um elevado grau de variação, especialmente no início. É intuitivamente evidente que, se o experimento acima continuasse indefinidamente, essas freqüências relativas iriam se estabilizar próximas do valor 1/2. Conseqüentemente, teríamos toda razão em acreditar que, depois de algum tempo decorrido, os dois meios-fios estariam igualmente molhados.

Esta propriedade de estabilidade da freqüência relativa é, por enquanto, uma noção inteiramente intuitiva, porém mais tarde estaremos aptos a torná-la matematicamente precisa. A essência desta propriedade é que, se um experimento for executado um grande número de vezes, a freqüência relativa da ocorrência de algum evento A tenderá a variar cada vez menos à medida que o número de repetições for aumentada. Esta característica é também conhecida como *regularidade estatística.*

Nós fomos um tanto vagos em nossa definição de experimento. Quando um procedimento ou mecanismo constituirá, em nossa acepção, um experimento capaz de ser estudado matematicamente por meio de um modelo não-determinístico? Já afirmamos, anteriormente, que um experimento deve ser capaz de ser realizado repetidamente, sob condições essencialmente inalteradas. Agora, podemos acrescentar outra condição. Quando o experimento for repetidamente realizado, ele deverá apresentar a regularidade estatística mencionada acima. Mais adiante, estudaremos um teorema (denominado Lei dos Grandes Números) que mostrará que a regularidade estatística é, de fato, uma *conseqüência* da primeira condição: reprodutibilidade.

1.7. Noções Fundamentais de Probabilidade

Voltemos agora ao problema proposto acima: atribuir um número a cada evento A, o qual avaliará quão verossímil será a ocorrência de A quando o experimento for realizado. Uma *possível* maneira de tratar a questão seria a seguinte: repetir o experimento um grande número de vezes, calcular a freqüência relativa f_A e utilizar esse número. Quando recordamos as propriedades de f_A, torna-se evidente que este número *fornece* uma informação muito precisa de quão verossímil é a ocorrência de A. Além disso, sabemos que à medida que o experimento se repetir mais e mais vezes, a freqüência relativa f_A se estabilizará próxima de algum número, suponhamos p. Há, contudo, duas sérias objeções a esta maneira de tratar a questão: (*a*) Não está esclarecido quão grande deva ser n, antes que se conheça o número: 1.000 ? 2.000 ? 10.000 ? (*b*) Uma vez que o experimento tenha sido completamente descrito e o evento A especificado, o número que estamos procurando não deverá depender do experimentador ou da particular veia de sorte que ele possua. (Por exemplo, é possível que uma moeda perfeitamente equilibrada, quando jogada 10 vezes, venha a apresentar 9 caras e 1 coroa. A freqüência relativa do evento $A = \{$ocorrer cara$\}$ seria, nesse caso, igual a 9/10.

No entanto, é evidente que nas próximas 10 jogadas o padrão de caras e coroas possa se inverter.) O que desejamos é um meio de obter tal número, sem recorrer à experimentação. Naturalmente, para que o número que convencionarmos tenha significado, qualquer experimentação subseqüente deverá produzir uma freqüência relativa que seja "próxima" do valor convencionado, particularmente se o número de repetições, no qual a freqüência relativa calculada se tenha baseado, for muito grande. Nós procederemos, formalmente, da seguinte maneira:

Definição. Seja $\mathcal{E}$ um experimento. Seja S um espaço amostral associado a $\mathcal{E}$. A cada evento A associaremos um número real representado por $P(A)$ e denominado *probabilidade de A*, que satisfaça às seguintes propriedades:

(1) $0 \leq P(A) \leq 1$.

(2) $P(S) = 1$. (1.3)

(3) Se A e B forem eventos mutuamente excludentes, $P(A \cup B) = P(A) + P(B)$.

(4) Se $A_1, A_2, \ldots, A_n, \ldots$ forem, dois a dois, eventos mutuamente excludentes, então,

$$P(\cup_{i=1}^{\infty} A_i) = P(A_1) + P(A_2) + \ldots + P(A_n) + \ldots$$

Observe-se que da Propriedade 3, *decorre* imediatamente que, para qualquer *n finito*,

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

A Propriedade 4 *não* se seguirá; no entanto, quando considerarmos o espaço amostral idealizado, esta propriedade será imposta e, por isso, foi incluída aqui.

A escolha das propriedades da probabilidade acima relacionadas é, obviamente, sugerida pelas correspondentes características da freqüência relativa. A propriedade, antes mencionada como regularidade estatística, será mais adiante vinculada a esta definição de probabilidade. Por enquanto, nós apenas afirmamos que se pode mostrar que os números $P(A)$ e f_A são "próximos" um do outro (em determinado sentido), se f_A for baseado em um grande número de repetições. É este fato que nos dá a justificativa da utilização de $P(A)$ para avaliarmos quão verossímil é a ocorrência de A.

Por enquanto não sabemos *como* calcular $P(A)$. Nós apenas arrolamos algumas propriedades gerais que $P(A)$ possui. O leitor

terá que ser um pouco mais paciente (até o próximo capítulo), antes que aprenda como avaliar $P(A)$. Antes de voltarmos a esta questão, vamos enunciar e demonstrar várias conseqüências relacionadas a $P(A)$, que decorrem das condições acima e que não dependem da maneira pela qual nós realmente calculamos $P(A)$.

Teorema 1.1. Se $\emptyset$ for o conjunto vazio, então $P(\emptyset) = 0$.

Demonstração: Para qualquer evento A, podemos escrever $A = A \cup \emptyset$. Uma vez que A e $\emptyset$ são mutuamente excludentes, decorre da Propriedade 3, que $P(A) = P(A \cup \emptyset) = P(A) + P(\emptyset)$. Daqui, a conclusão do teorema é imediata.

Comentário: Mais tarde, teremos ocasião de ver que a recíproca do teorema acima não é verdadeira. Isto é, se $P(A) = 0$, não poderemos, em geral, concluir que $A = \emptyset$, porque existem situações nas quais atribuímos probabilidade zero a um evento que pode ocorrer.

Teorema 1.2. Se $\overline{A}$ for o evento complementar de A, então

$$P(A) = 1 - P(\overline{A}). \tag{1.4}$$

Demonstração: Podemos escrever $S = A \cup \overline{A}$ e, empregando as Propriedades 2 e 3, obteremos $1 = P(A) + P(\overline{A})$.

Comentário: Este é um resultado particularmente útil, porque ele significa que sempre que desejarmos avaliar $P(A)$, poderemos calcular $P(\overline{A})$ e, depois, obtermos o resultado desejado por subtração. Veremos mais tarde que, em muitos problemas, é muito mais fácil calcular $P(\overline{A})$ do que $P(A)$.

Teorema 1.3. Se A e B forem dois eventos *quaisquer*, então

$$P(A \cup B) = P(A) + P(B) - P(A \cap B). \tag{1.5}$$

Demonstração: A idéia desta demonstração é decompor $A \cup B$ e B em dois eventos mutuamente excludentes e, em seguida, aplicar a Propriedade 3. (Veja o Diagrama de Venn na Fig. 1.2.)

Desse modo escreveremos

$$A \cup B = A \cup (B \cap \overline{A}),$$
$$B = (A \cap B) \cup (B \cap \overline{A}).$$

Conseqüentemente,

$$P(A \cup B) = P(A) + P(B \cap \overline{A}),$$
$$P(B) = P(A \cap B) + P(B \cap \overline{A}).$$

Subtraindo a segunda igualdade da primeira, obtém-se

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

e daí chega-se ao resultado.

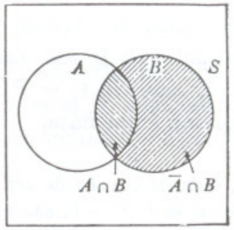


Fig. 1.2

Comentário: Este teorema representa uma extensão imediata da Propriedade 3, porque se $A \cap B = \emptyset$, obteremos do enunciado acima a Propriedade 3.

Teorema 1.4. Se A, B e C forem três eventos quaisquer, então

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - $$
$$- P(B \cap C) + P(A \cap B \cap C). \qquad (1.6)$$

Demonstração: A demonstração consiste em escrever $A \cup B \cup C$ na forma $(A \cup B) \cup C$ e aplicar o resultado do teorema anterior. Deixa-se ao leitor completar a demonstração.

Comentário: Uma extensão óbvia do teorema é sugerida. Sejam $A_1, \ldots, A_k$, quaisquer k eventos. Então,

$$P(A_1 \cup A_2 \cup \ldots \cup A_k) = \sum_{i=1}^{k} P(A_i) - \sum_{i<j=2}^{k} P(A_i \cap A_j) +$$
$$+ \sum_{i<j<r=3}^{k} P(A_i \cap A_j \cap A_r) + \ldots + (-1)^{k-1} P(A_1 \cap A_2 \cap \ldots \cap A_k). \qquad (1.7)$$

Este resultado pode ser facilmente estabelecido por indução matemática.

Teorema 1.5. Se $A \subset B$, então $P(A) \leq P(B)$.

Demonstração: Podemos decompor B em dois eventos mutuamente excludentes, na seguinte forma: $B = A \cup (B \cap \bar{A})$. Conseqüentemente, $P(B) = P(A) + P(B \cap \bar{A}) \geq P(A)$, porque $P(B \cap \bar{A}) \geq 0$, pela Propriedade 1.

Comentário: Este resultado é, certamente, de conhecimento intuitivo, pois ele afirma que se B *deve* ocorrer sempre que A ocorra, conseqüentemente, B é mais provável do que A.

1.8. Algumas Observações

(*a*) Cabe aqui uma palavra de advertência. Da exposição anterior poderia ser (incorretamente) inferido que quando escolhermos um modelo probabilístico para a descrição de algum fenômeno de observação, estaremos abandonando todas as relações determinísticas. Nada poderia estar mais distante da verdade. Nós ainda utilizamos o fato de que, por exemplo, a Lei de Ohm $I = E/R$ vale em determinadas circunstâncias. A diferença seria uma diferença de interpretação. Em vez de afirmar que a relação acima determina I para E e R dados, admitiremos que E ou R (ou ambos) possam variar de alguma maneira aleatória imprevisível e que, em conseqüência, I variará também de alguma forma aleatória. Para E e R *dados*, I será ainda determinado pela relação acima. O importante é que, quando se adotar um modelo probabilístico para a descrição de um circuito, considera-se a possibilidade de que E e R possam variar de alguma maneira imprevisível, a qual somente pode ser descrita probabilisticamente. Portanto, desde que tenha sentido considerar somente a *probabilidade* de que E e R tomem certos valores, torna-se significativo falar somente da probabilidade de que I venha a tomar certos valores.

(*b*) Algumas vezes, pode ser difícil realizar a escolha entre a adoção de um modelo determinístico ou um modelo probabilístico. Poderá depender da complicação de nossa técnica de mensuração e da exatidão associada a ela. Por exemplo, se medidas exatas forem tão difíceis de obter que leituras repetidas da mesma quantidade conduzam a resultados variados, um modelo probabilístico será sem dúvida mais adequado para descrever a situação.

(*c*) Indicaremos resumidamente que, sob certas circunstâncias, teremos condições de fazer hipóteses adicionais sobre o comportamento probabilístico de nossos resultados experimentais, as quais nos conduzirão a um método de avaliação das probabilidades básicas. A escolha dessas hipóteses adicionais pode ser baseada em considerações físicas do experimento (por exemplo, propriedades de simetria), evidência empírica ou, em alguns casos, apenas julgamento pessoal, baseado em experiência anterior de uma situação similar. A freqüência relativa f_A pode desempenhar um importante papel em nossa deliberação sobre a atribuição numérica de $P(A)$. Contudo, é importante compreender que qualquer suposição que façamos sobre $P(A)$ deve ser tal, que sejam satisfeitos os axiomas básicos desde (1) até (4) da Definição (1.3).

(*d*) No curso do desenvolvimento das idéias básicas da teoria da probabilidade, faremos algumas referências a determinadas analogias mecânicas. A primeira delas pode ser apropriada aqui. Em Mecânica, atribuímos a cada corpo B sua massa, digamos $m(B)$. Em seguida, faremos diversos cálculos e obteremos várias conclusões sobre o comportamento de B e suas relações com outros corpos, muitas das quais envolvem sua massa $m(B)$. O fato de que nós poderemos ter que recorrer a alguma aproximação para obter realmente $m(B)$ para um corpo especificado não diminui a utilidade do conceito de massa. Semelhantemente, estipularemos para cada evento A associado ao espaço amostral de um experimento um número $P(A)$, denominado *probabilidade de A*, e satisfazendo nossos axiomas básicos. Ao calcular realmente $P(A)$ para um evento específico, nós teremos que fazer hipóteses adicionais ou que obter uma aproximação baseada em evidência empírica.

(*e*) É muito importante compreender que nós tenhamos *postulado* a existência do número $P(A)$, e que tenhamos postulado determinadas propriedades que esse número possui. A validade das várias conseqüências (teoremas), decorrentes desses postulados, de modo algum depende da maneira pela qual iremos obter um valor numérico para $P(A)$. É essencial que este ponto fique claro. Por exemplo, admitimos que $P(A \cup B) = P(A) + P(B)$. A fim de empregar esta relação para a *avaliação* concreta de $P(A \cup B)$, deveremos conhecer os valores de $P(A)$ e de $P(B)$. Explicaremos, resumidamente, que sob certas circunstâncias, nós poderemos fazer suposições adicionais que conduzam a um método de avaliação dessas probabilidades. Se essas (ou outras) suposições não forem fundamentadas, poderemos ter de recorrer à experimentação a fim de alcançar o valor de $P(A)$ a partir de dados reais. A freqüência relativa f_A desempenhará nisso um importante papel e será, de fato, utilizada para aproximar $P(A)$.

Contudo, é importante saber que f_A e $P(A)$ não são a mesma coisa; que nós apenas utilizaremos f_A para aproximar $P(A)$ e que, sempre que nos referirmos a $P(A)$, estaremos nos referindo ao valor postulado. Se nós "identificarmos" f_A com $P(A)$, deveremos compreender que estaremos tão-somente substituindo um valor postulado por uma aproximação obtida experimentalmente. Quão boa ou má essa aproximação possa ser, de modo algum influencia a estrutura lógica de nosso modelo. Muito embora o fenômeno que o modelo tente representar tenha sido levado em conta na construção do modelo, nós nos teremos distanciado do próprio fenômeno (ao menos temporariamente), quando entrarmos no reino do modelo.

Problemas

1.1. Suponha que o conjunto fundamental seja formado pelos inteiros positivos de 1 a 10. Sejam $A = \{2, 3, 4\}$, $B = \{3, 4, 5\}$, e $C = \{5, 6, 7\}$. Enumere os elementos dos seguintes conjuntos:

(*a*) $\overline{A} \cap B$. (*b*) $\overline{A} \cup B$. (*c*) $\overline{\overline{A} \cap \overline{B}}$. (*d*) $\overline{A \cap \overline{(B \cap C)}}$. (*e*) $\overline{A \cap (B \cup C)}$.

1.2. Suponha que o conjunto fundamental U seja dado por $U = \{x \mid 0 \leq x \leq 2\}$. Sejam os conjuntos A e B definidos da forma seguinte: $A = \{x \mid 1/2 < x \leq 1\}$ e $B = \{x \mid 1/4 \leq x < 3/2\}$. Descreva os seguintes conjuntos:

(*a*) $\overline{A \cup B}$. (*b*) $A \cup \overline{B}$. (*c*) $\overline{A \cap B}$. (*d*) $\overline{A} \cap B$.

1.3. Quais das seguintes relações são verdadeiras?

(*a*) $(A \cup B) \cap (A \cup C) = A \cup (B \cap C)$. (*b*) $(A \cup B) = (A \cap \overline{B}) \cup B$.
(*c*) $\overline{A} \cap B = A \cup B$. (*d*) $\overline{(A \cup B)} \cap C = \overline{A} \cap \overline{B} \cap C$.
(*e*) $(A \cap B) \cap (\overline{B} \cap C) = \emptyset$.

1.4. Suponha que o conjunto fundamental seja formado por todos os pontos (x, y) de coordenadas ambas inteiras, e que estejam dentro ou sobre a fronteira do quadrado limitado pelas retas $x = 0$, $y = 0$, $x = 6$ e $y = 6$. Enumere os elementos dos seguintes conjuntos:

(*a*) $A = \{(x, y) \mid x^2 + y^2 \leq 6\}$. (*b*) $B = \{(x, y) \mid y \leq x^2\}$.
(*c*) $C = \{(x, y) \mid x \leq y^2\}$. (*d*) $B \cap C$. (*e*) $(B \cup A) \cap \overline{C}$.

1.5. Empregue diagramas de Venn para estabelecer as seguintes relações: (*a*) $A \subset B$ e $B \subset C$ implica que $A \subset C$. (*b*) $A \subset B$ implica que $A = A \cap B$. (*c*) $A \subset B$ implica que $\overline{B} \subset \overline{A}$. (*d*) $A \subset B$ implica que $A \cup C \subset B \cup C$. (*e*) $A \cap B = \emptyset$ e $C \subset A$ implicam que $B \cap C = \emptyset$.

1.6. Peças que saem de uma linha de produção são marcadas defeituosa (D) ou não defeituosa (N). As peças são inspecionadas e sua condição registrada. Isto é feito até que duas peças defeituosas consecutivas sejam fabricadas ou que quatro peças tenham sido inspecionadas, aquilo que ocorra em primeiro lugar. Descreva um espaço amostral para este experimento.

1.7. (*a*) Uma caixa com N lâmpadas contém r lâmpadas ($r < N$) com filamento partido. Essas lâmpadas são verificadas uma a uma, até que uma lâmpada defeituosa seja encontrada. Descreva um espaço amostral para este experimento.

(*b*) Suponha que as lâmpadas acima sejam verificadas uma a uma, até que todas as defeituosas tenham sido encontradas. Descreva o espaço amostral para este experimento.

1.8. Considere quatro objetos, a, b, c e d. Suponha que a *ordem* em que tais objetos sejam listados represente o resultado de um experimento. Sejam os eventos A e B definidos assim: $A = \{a$ está na primeira posição$\}$; $B = \{b$ está na segunda posição$\}$.

(*a*) Enumere todos os elementos do espaço amostral.

(*b*) Enumere todos os elementos dos eventos $A \cap B$ e $A \cup B$.

1.9. Um lote contém peças pesando 5, 10, 15, ..., 50 gramas. Admitamos que ao menos duas peças de cada peso sejam encontradas no lote. Duas peças são retiradas do lote. Sejam X o peso da primeira peça escolhida e Y o peso da segunda. Portanto, o par de números (X, Y) representa um resultado simples do experimento. Empregando o plano XY, marque o espaço amostral e os seguintes eventos:

(*a*) $\{X = Y\}$. (*b*) $\{Y > X\}$.

(*c*) A segunda peça é duas vezes mais pesada que a primeira.

(*d*) A primeira peça pesa menos 10 gramas que a segunda peça.

(*e*) O peso médio de duas peças é menor do que 30 gramas.

1.10. Durante um período de 24 horas, em algum momento X, uma chave é posta na posição "ligada". Depois, em algum momento futuro Y (ainda durante o mesmo período de 24 horas), a chave é virada para a posição "desligada". Suponha que X e Y sejam medidas em horas, no eixo dos tempos, com o início do período na origem da escala. O resultado do experimento é constituído pelo par de números (X, Y).

(*a*) Descreva o espaço amostral.

(*b*) Descreva e marque no plano XY os seguintes eventos:

(i) O circuito está ligado por uma hora ou menos.

(ii) O circuito está ligado no tempo z, onde z é algum instante no período dado de 24 horas.

(iii) O circuito é ligado antes do tempo t_1 e desligado depois do tempo t_2 (onde também $t_1 < t_2$ são dois instantes durante o período de 24 horas especificado).

(iv) O circuito permanece ligado duas vezes mais tempo do que desligado.

1.11. Sejam A, B e C três eventos associados a um experimento. Exprima em notações de conjuntos, as seguintes afirmações verbais:

(*a*) Ao menos um dos eventos ocorre.

(*b*) Exatamente um dos eventos ocorre.

(*c*) Exatamente dois dos eventos ocorrem.

(*d*) Não mais de dois dos eventos ocorrem simultaneamente.

1.12. Demonstre o Teor 1.4.

1.13. (*a*) Verifique que para dois eventos quaisquer, A_1 e A_2, temos que $P(A_1 \cup A_2) \leq P(A_1) + P(A_2)$.

(*b*) Verifique que para quaisquer n eventos $A_1, \ldots, A_n$, temos que

$$P(A_1 \cup \ldots \cup A_n) \leq P(A_1) + \ldots + P(A_n).$$

[*Sugestão*: Empregue a indução matemática. O resultado enunciado em (*b*) é denominado desigualdade de Boole.]

1.14. O Teor. 1.3 trata da probabilidade de que *ao menos um* de dois eventos A ou B ocorra. O seguinte enunciado se refere à probabilidade de que *exatamente um* dos eventos A ou B ocorra. Verifique que

$$P[(A \cap \overline{B}) \cup (B \cap \overline{A})] = P(A) + P(B) - 2P(A \cap B).$$

1.15. Um certo tipo de motor elétrico falha se ocorrer uma das seguintes situações: emperramento dos mancais, queima dos enrolamentos, desgaste das escovas. Suponha que o emperramento seja duas vezes mais provável do que a queima, esta sendo quatro vezes mais provável do que o desgaste das escovas. Qual será a probabilidade de que a falha seja devida a cada uma dessas circunstâncias?

1.16. Suponha que A e B sejam eventos tais que $P(A) = x$, $P(B) = y$, e $P(A \cap B) = z$. Exprima cada uma das seguintes probabilidades em termos de x, y e z.

(*a*) $P(\overline{A} \cup \overline{B})$. (*b*) $P(\overline{A} \cap B)$. (*c*) $P(\overline{A} \cup B)$. (*d*) $P(\overline{A} \cap \overline{B})$.

1.17. Suponha que A, B e C sejam eventos tais que $P(A) = P(B) = P(C) = 1/4$, $P(A \cap B) = P(C \cap B) = 0$ e $P(A \cap C) = 1/8$. Calcule a probabilidade de que ao menos um dos eventos A, B ou C ocorra.

1.18. Uma instalação é constituída de duas caldeiras e uma máquina. Admita que o evento A seja que a máquina esteja em boas condições de funcionamento, enquanto os eventos B_k ($k = 1, 2$) são os eventos de que a k-ésima caldeira esteja em boas condições. O evento C é que a instalação possa funcionar. Se a instalação puder funcionar sempre que a máquina e pelo menos uma das caldeiras funcionar, expresse os eventos C e $\overline{C}$, em termos de A e dos B_k.

1.19. Um mecanismo tem dois tipos de unidades: I e II. Suponha que se disponha de duas unidades do tipo I e três unidades do tipo II. Defina os eventos A_k, $k = 1, 2$ e B_j, $j = 1, 2, 3$ da seguinte maneira: A_k: a k-ésima unidade do tipo I está funcionando adequadamente; B_j: a j-ésima unidade do tipo II está funcionando adequadamente. Finalmente, admita que C represente o evento: o mecanismo funciona. Admita que o mecanismo funcione se ao menos uma unidade do tipo I e ao menos duas unidades do tipo II funcionarem; expresse o evento C em termos dos A_k e dos B_j.

Espaços Amostrais Finitos

Capítulo 2

2.1. Espaço Amostral Finito

Neste capítulo nos ocuparemos unicamente de experimentos para os quais o espaço amostral S seja formado de um número *finito* de elementos. Isto é, admitiremos que S possa ser escrito sob a forma $S = \{a_1, a_2, \ldots, a_k\}$. Se nos reportarmos aos exemplos de espaços amostrais da Seç. 1.4, observaremos que S_1, S_2, S_3, S_4, S_5, S_7 e S_{12} são todos finitos.

A fim de caracterizar $P(A)$ para este modelo, deveremos inicialmente considerar o evento formado por um *resultado simples*, algumas vezes denominado evento *simples* ou *elementar*, $A = \{a_i\}$. Procederemos da seguinte maneira:

A cada evento simples $\{a_i\}$ associaremos um número p_i, denominado probabilidade de $\{a_i\}$, que satisfaça às seguintes condições:

(*a*) $p_i \geq 0, \quad i = 1, 2, \ldots, k,$

(*b*) $p_1 + p_2 + \ldots + p_k = 1.$

[Porque $\{a_i\}$ é um evento, essas condições devem ser coerentes com aquelas postuladas para as probabilidades dos eventos em geral, como foi feito nas Eq. (1.3). É fácil verificar que isso se dá.]

Em seguida, suponha-se que um evento A seja constituído por r resultados, $1 \leq r \leq k$, a saber

$$A = \{a_{j_1}, a_{j_2}, \ldots, a_{j_r}\},$$

onde $j_1, j_2, \ldots, j_r$ representam um qualquer dos r índices, de 1 até k. Conseqüentemente, conclui-se da Eq. (1.3), Propriedade 4, que

$$P(A) = p_{j_1} + p_{j_2} + \ldots + p_{j_r}. \tag{2.1}$$

Para resumir: a atribuição de probabilidades P_i a cada evento elementar $\{a_i\}$, sujeito às condições (*a*) e (*b*) citadas anteriormente, determina unicamente $P(A)$ para cada evento $A \subset S$, onde $P(A)$ é dado pela Eq. (2.1).

Para avaliarmos os p_j individuais, alguma hipótese referente aos resultados individuais deve ser feita.

Exemplo 2.1. Suponha-se que somente três resultados sejam possíveis em um experimento, a saber, a_1, a_2 e a_3. Além disso, suponha-se que a_1 seja duas vezes mais provável de ocorrer que a_2, o qual por sua vez é duas vezes mais provável de ocorrer que a_3.

Portanto, $p_1 = 2p_2$ e $p_2 = 2p_3$. Já que $p_1 + p_2 + p_3 = 1$, teremos $4p_3 + 2p_3 + p_3 = 1$, o que finalmente dá

$$p_3 = \frac{1}{7}, \quad p_2 = \frac{2}{7} \quad \text{e} \quad p_1 = \frac{4}{7}.$$

Comentário: Na exposição que se segue, empregaremos a expressão "igualmente verossímeis" para significar "igualmente prováveis".

2.2. Resultados Igualmente Verossímeis

A hipótese mais comumente feita para espaços amostrais finitos é a de que todos os resultados sejam igualmente verossímeis. Esta hipótese não pode ser, contudo, tomada como segura; ela deve ser cuidadosamente justificada. Existem muitos experimentos para os quais tal hipótese é assegurada, mas existem também muitas situações experimentais nas quais seria absolutamente errôneo aceitar-se essa suposição. Por exemplo, seria bastante irreal supor que seja igualmente verossímil não ocorrerem chamadas telefônicas em um centro entre 1 e 2 horas da madrugada e entre 17 e 18 horas da tarde.

Se todos os k resultados forem igualmente verossímeis, segue-se que cada probabilidade será $p_i = 1/k$. Conseqüentemente, a condição $p_1 + \ldots + p_k = 1$ torna-se $kp_i = 1$ para todo i. Disto decorre que, para qualquer evento A formado de r resultados, teremos

$$P(A) = r/k.$$

Este método de avaliar $P(A)$ é freqüentemente enunciado da seguinte maneira:

$$P(A) = \frac{\text{número de casos favoráveis a } A \text{ pelos quais } \mathcal{E} \text{ pode ocorrer}}{\text{número total de casos pelos quais } \mathcal{E} \text{ pode ocorrer}}$$

É muito importante compreender que a expressão de $P(A)$ acima é apenas uma conseqüência da suposição de que todos os resultados sejam igualmente verossímeis, e ela é aplicável somente quando essa suposição for atendida. Ela certamente não serve como uma definição geral de probabilidade.

Exemplo 2.2. Um dado é lançado e todos os resultados se supõem igualmente verossímeis. O evento A ocorrerá se, e somente se, um número maior do que 4 aparecer, isto é, $A = \{5, 6\}$. Conseqüentemente, $P(A) = 1/6 + 1/6 = 2/6$.

Exemplo 2.3. Uma moeda equilibrada é atirada duas vezes. Seja A o evento: {aparece uma cara}. Na avaliação de $P(A)$, a análise do problema poderia ser a seguinte: O espaço amostral é $S = \{0, 1, 2\}$ onde cada resultado representa o número de caras que ocorre. Portanto, seria encontrada $P(A) = 1/3$! Esta análise é obviamente incorreta, porque no espaço amostral considerado acima, todos os resultados *não* são igualmente verossímeis. A fim de aplicar os métodos expostos, deveremos considerar em seu lugar o espaço amostral $S' = \{HH, HT, TH, TT\}$, onde H representa cara, e T coroa. Neste espaço amostral todos os resultados são igualmente verossímeis e, por isso, obteremos como solução correta de nosso problema: $P(A) = 2/4 = 1/2$. Poderíamos empregar corretamente o espaço S da seguinte maneira: Os resultados 0 e 2 são igualmente verossímeis, enquanto o resultado 1 é duas vezes mais provável que qualquer um dos outros. Portanto, $P(A) = 1/2$, o que concorda com a resposta anterior.

Este exemplo ilustra dois aspectos. Primeiro, deveremos estar bastante seguros de que todos os resultados possam supor-se igualmente verossímeis, antes de empregar o procedimento acima. Segundo, poderemos freqüentemente, por uma escolha apropriada do espaço amostral, reduzir o problema a outro, em que todos os resultados *sejam* igualmente verossímeis. Sempre que possível, isto deve ser feito porque geralmente torna o cálculo mais simples. Este aspecto será de novo mencionado em exemplos subseqüentes.

Muito freqüentemente, a maneira pela qual o experimento é executado determina se os resultados possíveis são igualmente verossímeis ou não. Por exemplo, suponha-se que retiremos um parafuso de uma caixa que contenha três parafusos de tamanhos diferentes. Se simplesmente escolhermos o parafuso estendendo a mão dentro da caixa e apanhando aquele que tocarmos primeiro, é óbvio que o parafuso maior terá maior probabilidade de ser escolhido que os outros dois. No entanto, etiquetando cuidadosamente cada parafuso com um número, escrevendo o número em um cartão, e escolhendo um cartão, tentaremos garantir que cada parafuso tenha de fato a mesma probabilidade de ser escolhido. Assim, poderemos nos meter em enorme trabalho a fim de assegurarmos que a suposição matemática de resultados igualmente verossímeis seja de fato apropriada.

Nos exemplos já vistos e em muitos que se seguirão, trataremos da escolha ao acaso de um ou mais objetos de uma dada coleção de objetos. Definamos esta noção mais precisamente. Suponhamos que se tenha N objetos, a saber $a_1, a_2, \ldots, a_N$.

(*a*) *Escolher ao acaso um objeto,* dentre N objetos, significa que cada objeto tem a mesma probabilidade de ser escolhido, isto é,

$$\text{Prob (escolher } a_i) = 1/N, \quad i = 1, 2, \ldots, N.$$

(*b*) *Escolher ao acaso dois objetos,* dentre N objetos, significa que *cada par* de objetos (deixada a ordem à parte) tem a mesma probabilidade de ser escolhido que qualquer outro par. Por exemplo, se devemos escolher ao acaso dois objetos dentre (a_1, a_2, a_3, a_4), obter a_1 e a_2 é então tão provável quanto obter a_2 e a_3 etc. Esta formulação levanta imediatamente a questão de *quantos* pares diferentes existem. Admita-se que existam K desses pares. Então, a probabilidade de cada par seria $1/K$. Logo, veremos como calcular K.

(*c*) *Escolher ao acaso n objetos* ($n \leq N$) dentre N objetos significa que cada ênupla, a saber $a_{i_1}, a_{i_2}, \ldots, a_{i_n}$ é tão provável de ser escolhida quanto qualquer outra ênupla.

Comentário: Já sugerimos acima que se deve tomar extremo cuidado durante o procedimento experimental, para assegurarmos que a suposição matemática de "escolher ao acaso" seja atendida.

2.3. Métodos de Enumeração

Deveremos fazer uma digressão, a esta altura, para aprendermos como enumerar. Considere-se novamente a forma já vista de $P(A)$, a saber $P(A) = r/k$, onde k é igual ao número total de maneiras pelas quais $\mathcal{E}$ pode ocorrer, enquanto r é igual ao número de maneiras pelas quais A pode ocorrer. Nos exemplos apresentados até aqui, pequena dificuldade foi encontrada para calcular r e k. Mas nós precisamos estudar situações apenas um pouco mais complicadas, para percebermos a necessidade de alguns procedimentos sistemáticos de contagem ou enumeração.

BIBLIOTECA CENTRAL PUCRS

Exemplo 2.4. Uma partida de cem peças é composta de 20 peças defeituosas e 80 peças perfeitas. Dez dessas peças são escolhidas ao acaso, sem reposição de qualquer peça escolhida antes que a seguinte seja escolhida. Qual é a probabilidade de que exatamente metade das peças escolhidas seja defeituosa?

Para analisarmos este problema, consideremos o seguinte espaço amostral S. Cada elemento de S é constituído de dez possíveis peças da partida, $(i_1, i_2, \ldots, i_{10})$. Quantos resultados desses existem? E dentre esses resultados, quantos têm a característica de que exatamente a metade das peças seja defeituosa? Nós, evidentemente, precisamos ter condições de responder a tais questões a fim de resolvermos o problema em estudo. Muitos problemas semelhantes dão origem a questões análogas. Nas poucas seções seguintes, apresentaremos algumas técnicas sistemáticas de enumeração.

A. *Regra da Multiplicação*. Suponha-se que um procedimento designado por 1 possa ser executado de n_1 maneiras. Admita-se que um segundo procedimento, designado por 2, possa ser executado de n_2 maneiras. Suponha-se, também, que cada maneira de executar 1 possa ser seguida por qualquer daquelas para executar 2. Então, o procedimento formado por 1 seguido de 2 poderá ser executado de $n_1 \cdot n_2$ maneiras. Para indicar a validade deste princípio, é mais fácil considerar o seguinte tratamento sistemático.

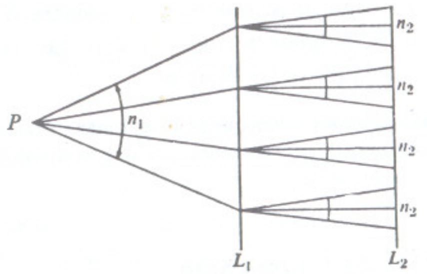


Fig. 2.1

Considerem-se um ponto P e duas retas L_1 e L_2. Admita-se que o procedimento 1 consista em ir de P até L_1, enquanto o procedimento 2 consista em ir de L_1 até L_2. A Fig. 2.1 indica como o resultado final é obtido.

Comentário: Obviamente, esta regra pode ser estendida a qualquer número de procedimentos. Se existirem k procedimentos e o i-ésimo procedimento puder ser executado de n_i maneiras, $i = 1, 2, \ldots, k$, então o procedimento formado por 1, seguido por 2, ..., seguido pelo procedimento k, poderá ser executado de $n_1 n_2 \ldots n_k$ maneiras.

Exemplo 2.5. Uma peça manufaturada deve passar por três estações de controle. Em cada estação, a peça é inspecionada para determinada característica e marcada adequadamente. Na primeira estação, três classificações são possíveis, enquanto nas duas últimas, quatro classificações são possíveis. Conseqüentemente, existem $3 \cdot 4 \cdot 4 = 48$ maneiras pelas quais uma peça pode ser marcada.

B. *Regra da Adição*. Suponha-se que um procedimento, designado por 1, possa ser realizado de n_1 maneiras. Admita-se que um segundo procedimento, designado por 2, possa ser realizado de n_2 maneiras. Além disso, suponha-se que *não* seja possível que *ambos* os procedimentos 1 e 2 sejam realizados em conjunto. Então, o número de maneiras pelas quais poderemos realizar ou 1 ou 2 será $n_1 + n_2$.

Novamente, empregaremos um tratamento esquemático para nos convencermos da validade da regra da adição, como a Fig. 2.2 indica.

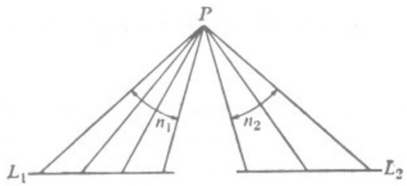


Fig. 2.2

Comentário: Esta regra também pode ser generalizada da seguinte maneira: Se existirem k procedimentos e o i-ésimo procedimento puder ser realizado de n_i maneiras, $i = 1, 2, \ldots, k$, então, o número de maneiras pelas quais poderemos realizar *ou* o procedimento 1, *ou* o procedimento 2, *ou*..., *ou* o procedimento k, é dado por $n_1 + n_2 + \ldots + n_k$, supondo-se que dois quaisquer deles não se possam realizar conjuntamente.

Exemplo 2.6. Suponha-se que estejamos planejando uma viagem e devamos escolher entre o transporte por ônibus ou por trem. Se existirem três rodovias e duas ferrovias, então existirão $3 + 2 = 5$ caminhos disponíveis para a viagem.

C. *Permutações e Arranjos*. (*a*) Suponha-se que nós temos n objetos diferentes. De quantas maneiras ${}_nP_n$ poderemos dispor (permutar) esses objetos? Por exemplo, se tivermos os objetos a, b e c, poderemos considerar as seguintes permutações: *abc*, *acb*, *bac*, *bca*, *cab* e *cba*. Portanto, a resposta é 6. Considere-se, em geral, o seguinte esquema: Permutar os n objetos equivale a colocá-los dentro de uma caixa com n compartimentos, em alguma ordenação:

1	2	·	·	·	n

O primeiro compartimento pode ser ocupado por qualquer uma das n maneiras, o segundo compartimento por qualquer uma das $(n-1)$ maneiras,..., e o último compartimento apenas por uma maneira. Portanto, aplicando-se a regra da multiplicação, vista acima, verifica-se que a caixa poderá ser carregada de $n(n-1)(n-2)\ldots 1$ maneiras. Este número aparece tão freqüentemente em Matemática que se adotam um nome e um símbolo especiais para ele.

Definição. Sendo n um inteiro positivo, definimos $n! = (n)(n-1)(n-2)\ldots 1$ e o denominamos *fatorial de n.* Também definimos $0! = 1$.

Dessa maneira, o número de permutações de n objetos diferentes é dado por

$$_nP_n = n!$$

(*b*) Considerem-se novamente n objetos *diferentes.* Agora desejamos escolher r desses objetos, $0 \leq r \leq n$ *e* permutar os r escolhidos. Denotaremos o número de maneiras de fazer isso (arranjos) por $_nA_r$. Recorremos novamente ao esquema acima, de encher uma caixa de n compartimentos; desta vez simplesmente paramos depois que o compartimento de ordem r tenha sido ocupado. Assim, o primeiro compartimento pode ser preenchido de n maneiras, o segundo de $(n-1)$ maneiras,... e o de ordem r de $n-(r-1)$ maneiras. Portanto, o procedimento completo poderá ser executado, novamente aplicando-se a regra da multiplicação, de

$$n(n-1)(n-2)\ldots(n-r+1)$$

maneiras. Empregando a notação de fatorial, introduzida acima, poderemos escrever

$$_nA_r = \frac{n!}{(n-r)!}.$$

D. Combinações. Considerem-se, novamente, n objetos diferentes. Agora, trataremos da contagem do número de maneiras de

escolher r dentre esses n objetos *sem* considerarmos a ordem. Por exemplo, temos os objetos a, b, c e d, e $r = 2$; desejamos contar ab, ac, ad, bc, bd e cd; por outras palavras, *não* contaremos ab *e* ba, porque os mesmos objetos estão incluídos e somente a ordem é diversa.

Para obtermos o resultado geral, recordaremos a fórmula deduzida acima: o número de maneiras de escolher r objetos dentre n, *e* permutar os r escolhidos é $n!/(n-r)!$ Seja C o número de maneiras de escolher r dentre os n, não considerada a ordem. (Isto é, C é o número procurado.) Observe-se que, uma vez que r objetos tenham sido escolhidos, existirão $r!$ maneiras de permutá-los. Conseqüentemente, aplicando-se novamente a regra da multiplicação, juntamente com esse resultado, obteremos

$$Cr! = \frac{n!}{(n-r)!}.$$

Portanto, o número de maneiras de escolher r dentre n objetos diferentes, não se considerando a ordem, é dado por

$$C = \frac{n!}{r!(n-r)!}.$$

Este número surge em muitas passagens da Matemática e, por isso, um símbolo especial é empregado para ele. Escreveremos

$$\frac{n!}{r!(n-r)!} = \binom{n}{r}$$

Para nossos objetivos atuais, $\binom{n}{r}$ somente fica definido para n inteiro positivo e r um inteiro tal que $0 \leq r \leq n$. Contudo, poderemos definir $\binom{n}{r}$ de modo mais geral, para qualquer número real n e para qualquer inteiro não negativo r, na forma seguinte:

$$\binom{n}{r} = \frac{n(n-1)(n-2)\cdots(n-r+1)}{r!}.$$

Os números $\binom{n}{r}$ são freqüentemente denominados *coeficientes binomiais*, porque eles aparecem como coeficientes no desenvolvimento da expressão binomial $(a+b)^n$. Se n for um inteiro positivo, $(a+b)^n = (a+b)(a+b)\ldots(a+b)$. Quando a multiplicação tiver sido executada, cada termo será formado de k elementos a, e de $(n-k)$ elementos b, $k = 0, 1, 2, \ldots, n$. Quantos termos da forma $a^k b^{n-k}$

existirão? Simplesmente contaremos o número de maneiras possíveis de escolher k dentre os n elementos a, deixando de lado a ordem. Mas isto é justamente dado por $\binom{n}{k}$. Daí obtermos o que é conhecido como o *teorema binomial:*

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}. \tag{2.2}$$

Os números $\binom{n}{r}$ apresentam muitas propriedades interessantes, apenas duas das quais mencionaremos aqui: (A menos que se diga expressamente de modo diverso, admitiremos que n seja inteiro positivo e r um inteiro, $0 \leq r \leq n$.)

$$(a)\ \binom{n}{r} = \binom{n}{n-r}, \qquad (b)\ \binom{n}{r} = \binom{n-1}{r-1} + \binom{n-1}{r}$$

É fácil verificar algebricamente as duas identidades acima. Basta desenvolverem-se, em cada uma, o primeiro e o segundo membros, e verificar que são iguais.

Existe, contudo, outra maneira de verificar essas identidades, que emprega a interpretação que demos para $\binom{n}{r}$, a saber, o número de maneiras de escolher r dentre n coisas.

(*a*) Quando escolhemos r dentre n coisas, estamos ao mesmo tempo deixando $(n-r)$ coisas não escolhidas, e, por isso, escolher r dentre n é equivalente a escolher $(n-r)$ dentre n. Ora, isso é exatamente a primeira identidade a verificar.

(*b*) Vamos fixar um qualquer dos n objetos, por exemplo o primeiro, a_1. Ao escolher r objetos, a_1 estará incluído ou estará excluído, mas não ambas as coisas. Portanto, ao contar o número de maneiras de escolher r objetos, poderemos aplicar a Regra da Adição, explicada anteriormente.

Se a_1 for excluído, então deveremos escolher os r objetos desejados dentre os restantes $(n-1)$ objetos, e existem $\binom{n-1}{r}$ maneiras de se fazer isso.

Se a_1 for incluído, então somente $(r-1)$ mais objetos devem ser escolhidos dentre os restantes $(n-1)$ objetos e isto pode ser feito de $\binom{n-1}{r-1}$ maneiras. Conseqüentemente, o número procurado é a *soma* desses dois, o que verifica a segunda identidade.

Comentário: Neste contexto, os coeficientes binomiais $\binom{n}{k}$ têm sentido somente se n e k forem inteiros não-negativos, com $0 \leqslant k \leqslant n$. Todavia, se escrevermos

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!} = \frac{n\,(n-1)\ldots(n-k+1)}{k!},$$

observaremos que a última expressão tem sentido se n for qualquer número real e k for qualquer inteiro não-negativo. Portanto,

$$\binom{-3}{5} = \frac{(-3)(-4)\ldots(-7)}{5!}$$

e assim por diante.

Empregando esta versão estendida dos coeficientes binomiais, poderemos estabelecer a *forma generalizada do teorema binomial*:

$$(1+x)^n = \sum_{k=0}^{\infty} \binom{n}{k} x^k$$

Esta série tem significado para qualquer n real e para todo x tal que $|x| < 1$. Observe-se que, se n for um inteiro positivo, a série infinita se reduz a um número finito de termos, porque, neste caso, $\binom{n}{k} = 0$, se $k > n$.

Exemplo 2.7. (*a*) Dentre oito pessoas, quantas comissões de três membros podem ser escolhidas? Desde que duas comissões sejam a mesma comissão se forem constituídas pelas mesmas pessoas (não se levando em conta a ordem em que sejam escolhidas), teremos $\binom{8}{3} = 56$ comissões possíveis.

(*b*) Com oito bandeiras diferentes, quantos sinais feitos com três bandeiras se podem obter? Este problema parece-se muito com o anterior. No entanto, aqui a ordem acarreta diferença e, por isso, obteremos $8!/5! = 336$ sinais.

(*c*) Um grupo de oito pessoas é formado de cinco homens e três mulheres. Quantas comissões de três pessoas podem ser constituídas, incluindo exatamente dois homens? Aqui deveremos fazer duas coisas: escolher dois homens (dentre cinco) e escolher uma mulher (dentre três). Daí obtermos como número procurado $\binom{5}{2} \cdot \binom{3}{1} = 30$ comissões.

(*d*) Agora poderemos verificar uma afirmação feita anteriormente, a saber, a de que o número de subconjuntos (ou partes) de um conjunto constituído de n elementos é igual a 2^n (contados o conjunto vazio e o próprio conjunto). Simplesmente associemos a cada elemento o valor um ou zero, conforme esse elemento deva ser incluído ou excluído do subconjunto. Existem duas maneiras de rotular cada elemento e existem ao todo n desses elementos. Daí a regra da multiplicação nos dizer que existem $2 \cdot 2 \cdot 2 \cdots 2 = 2^n$ rotulações possíveis. Mas cada rotulação particular representa uma escolha de um subconjunto. Por exemplo, $(1, 1, 0, 0, 0, \ldots, 0)$ constituiria o subconjunto formado exatamente por a_1 e a_2. Ainda, $(1, 1, \ldots, 1)$ representaria o próprio S, e $(0, 0, \ldots, 0)$ representaria o conjunto vazio.

(*e*) Poderíamos obter o resultado acima, pelo emprego da Regra da Adição, na forma seguinte: Para obter subconjuntos, deveremos escolher o conjunto vazio, aqueles subconjuntos constituídos exatamente por um elemento, aqueles constituídos exatamente por dois elementos, ..., e o próprio conjunto constituído por todos os n elementos. Isto seria feito de

$$\binom{n}{0}+\binom{n}{1}+\binom{n}{2}+\cdots+\binom{n}{n}$$

maneiras. Ora, a soma desses coeficientes binomiais é exatamente o desenvolvimento de $(1+1)^n = 2^n$.

Voltemos agora ao Ex. 2.4. De uma partida formada por 20 peças defeituosas e 80 peças perfeitas, escolhemos ao acaso 10 (sem reposição). O número de maneiras de fazer isso é $\binom{100}{10}$. Daí, a probabilidade de achar exatamente 5 peças defeituosas e 5 perfeitas entre as 10 escolhidas ser dada por

$$\frac{\binom{20}{5}\binom{80}{5}}{\binom{100}{10}}.$$

Por meio de logaritmos de fatoriais (os quais se acham tabulados), a expressão acima pode ser avaliada como igual a 0,021.

Exemplo 2.8. Vamos generalizar o problema acima. Admitamos que temos N peças. Se escolhermos ao acaso n delas, sem reposição, teremos $\binom{N}{n}$ diferentes amostras possíveis, todas elas com a mesma probabilidade de serem escolhidas. Se as N peças forem formadas por r_1 da classe A e r_2 da classe B (com $r_1 + r_2 = N$), então, a probabilidade de que as n peças escolhidas sejam exatamente s_1 da classe A e $(n - s_1)$ da classe B será dada por

$$\frac{\binom{r_1}{s_1}\binom{r_2}{n-s_1}}{\binom{N}{n}}.$$

(A expressão acima se denomina *probabilidade hipergeométrica*, e será ainda reencontrada.)

Comentário: É muito importante especificar, quando falarmos de peças extraídas ao acaso, se a escolha é *com* ou *sem reposição*. Na maioria dos casos concretos, pretenderemos a última. Por exemplo, quando inspecionamos certo número de peças manufaturadas a fim de descobrirmos quantas defeituosas poderão existir, geralmente não tencionaremos examinar a mesma peça duas vezes. Já dissemos que o número de maneiras de escolher r coisas dentre n, não considerada a ordem, é dado por $\binom{n}{r}$. O número de maneiras de escolher r coisas dentre n, com reposição, é dado por n^r. Neste caso, *estaremos* interessados na ordem em que as peças sejam escolhidas.

Exemplo 2.9. Admitamos que se escolham ao acaso dois objetos, dentre os quatro denominados *a, b, c* e *d*.

(*a*) Se escolhermos sem reposição, o espaço amostral S poderá ser representado da forma abaixo:

$$S = \{(a, b);\ (a, c);\ (b, c);\ (b, d);\ (c, d);\ (a, d)\}.$$

Existem $\binom{4}{2} = 6$ resultados possíveis. Cada um desses resultados indica somente *quais* os dois objetos que foram escolhidos e *não* a ordem em que eles foram escolhidos.

(*b*) Se escolhermos com reposição, o espaço amostral S' poderá ser representado por:

$$S' = \left\{\begin{matrix}(a, a); (a, b); (a, c); (a, d); (b, a); (b, b); (b, c); (b, d);\\ (c, a); (c, b); (c, c,); (c, d); (d, a); (d, b); (d, c); (d, d)\end{matrix}\right\}.$$

Existem $4^2 = 16$ resultados possíveis. Aqui, cada um desses resultados indica quais objetos foram escolhidos *e* a ordem em que eles o foram. Escolher ao acaso implica que, se escolhermos sem reposição, todos os resultados em S serão igualmente verossímeis, enquanto se escolhermos com reposição, então todos os resultados em S' serão igualmente verossímeis. Portanto, se A for o evento {o objeto c é escolhido}, então teremos: de S, $P(A) = 3/6 = 1/2$ se escolhermos sem reposição; e de S', $P(A) = 7/16$ se escolhermos com reposição.

E. Permutações com Alguns Elementos Repetidos. Em todas as técnicas de enumeração já apresentadas, admitimos que todos os objetos considerados fossem diferentes (isto é, distinguíveis). No entanto, não é sempre essa a situação que ocorre.

Suponha-se, a seguir, que temos n objetos, tais que n_1 sejam de uma primeira espécie, n_2 de uma segunda espécie,..., n_k de uma k-ésima espécie, com $n_1 + n_2 + \ldots + n_k = n$. Nesse caso, o número de permutações possíveis desses n objetos é dado por

$$\frac{n!}{n_1! n_2! \ldots n_k!}.$$

Deixa-se ao leitor a dedução dessa fórmula. Note-se que, se todos os objetos *fossem* diferentes, teríamos $n_i = 1$, $i = 1, 2, \ldots, k$, e, conseqüentemente, a fórmula acima se reduziria a $n!$, que é o resultado obtido anteriormente.

Comentário: Devemos salientar mais uma vez que a atribuição realística de probabilidades a resultados individuais de um espaço amostral (ou a uma coleção de resultados, isto é, um evento) constitui alguma coisa que não pode ser deduzida matematicamente, mas que deve ser originada de outras considerações. Por exemplo, poderemos recorrer a determinados traços simétricos do experimento para averiguar se todos os resultados são igualmente prováveis. Além disso, poderemos construir um procedimento de amostragem (por exemplo, escolhendo um ou vários indivíduos de uma população especificada) de tal maneira que seja razoável admitir que todas as escolhas sejam igualmente prováveis. Em muitos outros casos, quando nenhuma suposição básica natural seja apropriada, deveremos recorrer à aproximação da freqüência relativa. Nós repetiremos o experimento n vezes e, em seguida, calcularemos a proporção de vezes em que o resultado (ou evento) em estudo tenha ocorrido. Ao empregar isto como uma aproximação, sabemos que é bastante improvável que esta freqüência relativa difira da "verdadeira" probabilidade (cuja existência tenha sido especificada por nosso modelo teórico), de um valor apreciável, se n for suficientemente grande. Quando for impossível estabelecer suposições razoáveis sobre a probabilidade de um resultado e também impossível repetir o experimento um grande número de vezes (em virtude de considerações de custo ou de tempo, por exemplo), será realmente bastante sem sentido prosseguir com um estudo probabilístico do experimento, exceto em uma base puramente teórica. (Para um comentário adicional sobre este mesmo ponto, veja a Seção 13.5).

Problemas

2.1. O seguinte grupo de pessoas está numa sala: 5 homens maiores de 21 anos; 4 homens com menos de 21 anos de idade; 6 mulheres maiores de 21 anos, e

3 mulheres menores. Uma pessoa é escolhida ao acaso. Definem-se os seguintes eventos: A = {a pessoa é maior de 21 anos}; B = {a pessoa é menor de 21 anos}; C = {a pessoa é homem}; D = {a pessoa é mulher}. Calcule:

(*a*) $P(B \cup D)$, (*b*) $P(\overline{A} \cap \overline{C})$.

2.2. Em uma sala, 10 pessoas estão usando emblemas numerados de 1 até 10. Três pessoas são escolhidas ao acaso e convidadas a saírem da sala simultaneamente. O número de seu emblema é anotado.

(*a*) Qual é a probabilidade de que o menor número de emblema seja 5?

(*b*) Qual é a probabilidade de que o maior número de emblema seja 5?

2.3. (*a*) Suponha que os três dígitos 1, 2 e 3 sejam escritos em ordem aleatória. Qual a probabilidade de que ao menos um dígito ocupe seu lugar próprio?

(*b*) O mesmo que em (*a*), com os dígitos 1, 2, 3 e 4.

(*c*) O mesmo que em (*a*), com os dígitos 1, 2, 3, ..., n.

Sugestão: Empregue (1.7).

(*d*) Examine a resposta a (*c*), quando n for grande.

2.4. Uma remessa de 1.500 arruelas contém 400 peças defeituosas e 1.100 perfeitas. Duzentas arruelas são escolhidas ao acaso (sem reposição) e classificadas.

(*a*) Qual a probabilidade de que sejam encontradas exatamente 90 peças defeituosas?

(*b*) Qual a probabilidade de que se encontrem ao menos 2 peças defeituosas?

2.5. Dez fichas numeradas de 1 até 10 são misturadas em uma urna. Duas fichas, numeradas (X, Y), são extraídas da urna, sucessivamente e sem reposição. Qual é a probabilidade de que seja $X + Y = 10$?

2.6. Um lote é formado de 10 artigos bons, 4 com defeitos menores e 2 com defeitos graves. Um artigo é escolhido ao acaso. Ache a probabilidade de que:

(*a*) Ele não tenha defeitos.

(*b*) Ele não tenha defeitos graves.

(*c*) Ele ou seja perfeito ou tenha defeitos graves.

2.7. Se do lote de artigos descrito no Probl. 2.6, dois artigos forem escolhidos (sem reposição), ache a probabilidade de que:

(*a*) Ambos sejam perfeitos. (*b*) Ambos tenham defeitos graves. (*c*) Ao menos um seja perfeito. (*d*) No máximo um seja perfeito. (*e*) Exatamente um seja perfeito. (*f*) Nenhum deles tenha defeitos graves. (*g*) Nenhum deles seja perfeito.

2.8. Um produto é montado em três estágios. No primeiro estágio, existem 5 linhas de montagem; no segundo estágio, existem 4 linhas de montagem e no terceiro estágio, existem 6 linhas de montagem. De quantas maneiras diferentes poderá o produto se deslocar durante o processo de montagem?

2.9. Um inspetor visita 6 máquinas diferentes durante um dia. A fim de evitar que os operários saibam quando ele os irá inspecionar, o inspetor varia a ordenação de suas visitas. De quantas maneiras isto poderá ser feito?

2.10. Um mecanismo complexo pode falhar em 15 estágios. De quantas maneiras poderá ocorrer que ele falhe em 3 estágios?

2.11. Existem 12 categorias de defeitos menores de uma peça manufaturada, e 10 tipos de defeitos graves. De quantas maneiras poderão ocorrer 1 defeito menor e 1 grave? E 2 defeitos menores e 2 graves?

2.12. Um mecanismo pode ser posto em uma dentre quatro posições: a, b, c e d. Existem 8 desses mecanismos incluídos em um sistema.

(*a*) De quantas maneiras esse sistema pode ser disposto?

(*b*) Admita que esses mecanismos sejam instalados em determinada ordem (linear) preestabelecida. De quantas maneiras o sistema poderá ser disposto, se dois mecanismos adjacentes não estiverem em igual posição?

(*c*) Quantas maneiras de dispor serão possíveis, se somente as posições a e b forem usadas, e o forem com igual freqüência?

(*d*) Quantas maneiras serão possíveis, se somente duas posições forem usadas, e dessas posições uma ocorrer três vezes mais freqüentemente que a outra?

2.13. Suponha que de N objetos, n sejam escolhidos ao acaso, *com* reposição. Qual será a probabilidade de que nenhum objeto seja escolhido mais do que uma vez? (Admita $n < N$.)

2.14. Com as seis letras a, b, c, d, e, f quantas palavras-código de 4 letras poderão ser formadas se:

(*a*) Nenhuma letra puder ser repetida?

(*b*) Qualquer letra puder ser repetida qualquer número de vezes?

2.15. Supondo que $\binom{99}{5} = a$ e $\binom{99}{4} = b$, expresse $\binom{100}{95}$ em termos de a e b. (*Sugestão*: *Não* calcule as expressões acima, para resolver o problema.)

2.16. Uma caixa contém etiquetas numeradas $1, 2, \ldots, n$. Duas etiquetas são escolhidas ao acaso. Determine a probabilidade de que os números das etiquetas sejam inteiros consecutivos se:

(*a*) As etiquetas forem escolhidas sem reposição.

(*b*) As etiquetas forem escolhidas com reposição.

2.17. Quantos subconjuntos se podem formar, contendo ao menos um elemento, de um conjunto de 100 elementos?

2.18. Um inteiro é escolhido ao acaso, dentre os números $1, 2, \ldots, 50$. Qual será a probabilidade de que o número escolhido seja divisível por 6 ou por 8?

2.19. Dentre 6 números positivos e 8 negativos, escolhem-se ao acaso 4 números (sem reposição) e multiplicam-se esses números. Qual será a probabilidade de que o produto seja um número positivo?

2.20. Determinado composto químico é obtido pela mistura de 5 líquidos diferentes. Propõe-se despejar um líquido em um tanque e, em seguida, juntar os outros líquidos sucessivamente. Todas as seqüências possíveis devem ser ensaiadas, para verificar-se qual delas dará o melhor resultado. Quantos ensaios deverão ser efetuados?

2.21. Um lote contém n peças, das quais se sabe serem r defeituosas. Se a ordem da inspeção das peças se fizer ao acaso, qual a probabilidade de que a peça inspecionada em k-ésimo lugar ($k \geq r$) seja a última peça defeituosa contida no lote?

2.22. Dentre os números $0, 1, 2, \ldots, 9$ são escolhidos ao acaso (sem reposição) r números ($0 < r < 10$). Qual é a probabilidade de que não ocorram dois números iguais?

Probabilidade Condicionada e Independência

Capítulo 3

3.1. Probabilidade Condicionada

Vamos reexaminar a diferença entre extrair uma peça de um lote, ao acaso, com ou sem reposição. No Ex. 2.4, o lote estudado tinha a seguinte composição: 80 não-defeituosas e 20 defeituosas. Suponha-se que escolhemos duas peças desse lote: (*a*) com reposição; (*b*) sem reposição.

Definamos os dois eventos seguintes:

$A = \{$a primeira peça é defeituosa$\}$; $B = \{$a segunda peça é defeituosa$\}$.

Se estivermos extraindo *com* reposição, $P(A) = P(B) = 20/100 = 1/5$, porque cada vez que extrairmos do lote, existirão 20 peças defeituosas no total de 100. No entanto, se estivermos extraindo *sem* reposição, os resultados não serão tão imediatos. É ainda verdade, naturalmente, que $P(A) = 1/5$. Mas e sobre $P(B)$? É evidente que, a fim de calcularmos $P(B)$, deveremos conhecer a composição do lote *no momento de se extrair a segunda peça*. Isto é, deveremos saber se A ocorreu ou não. Este exemplo mostra a necessidade de se introduzir o seguinte importante conceito.

Sejam A e B dois eventos associados ao experimento $\mathcal{E}$. Denotaremos por $P(B \mid A)$ a *probabilidade condicionada* do evento B, *quando* A tiver ocorrido.

No exemplo acima, $P(B \mid A) = 19/99$, porque se A tiver ocorrido, então para a segunda extração restarão somente 99 peças, das quais 19 delas serão defeituosas.

Sempre que calcularmos $P(B \mid A)$, estaremos essencialmente calculando $P(B)$ em relação ao *espaço amostral reduzido* A, em lugar de

fazê-lo em relação ao espaço amostral original S. Consideremos o Diagrama de Venn da Fig. 3.1.

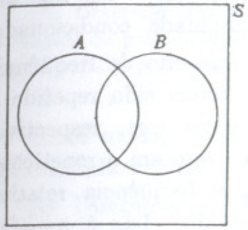


Fig. 3.1

Quando calcularmos $P(B)$ estaremos nos perguntando quão provável será estarmos em B, sabendo que devemos estar em S. E quando calcularmos $P(B \mid A)$ estaremos perguntando quão provável será estarmos em B, sabendo que devemos estar em A. (Isto é, o espaço amostral ficou *reduzido* de S para A.) Logo, daremos uma definição rigorosa de $P(B \mid A)$. Por enquanto, contudo, empregaremos nossa noção intuitiva de probabilidade condicionada e daremos um exemplo.

Exemplo 3.1. Dois dados equilibrados são lançados, registrando-se o resultado como (x_1, x_2), onde x_i é o resultado do i-ésimo dado, $i = 1, 2$. Por isso, o espaço amostral S pode ser representado pela seguinte lista de 36 resultados igualmente prováveis.

$$S = \left\{ \begin{array}{cccc} (1,1) & (1,2) & \dots & (1,6) \\ (2,1) & (2,2) & \dots & (2,6) \\ \vdots & & & \vdots \\ (6,1) & (6,2) & \dots & (6,6) \end{array} \right\}.$$

Consideremos os dois eventos seguintes:

$$A = \{(x_1, x_2) \mid x_1 + x_2 = 10\}, \qquad B = \{(x_1, x_2) \mid x_1 > x_2\}.$$

Assim, $A = \{(5,5), (4,6), (6,4)\}$ e $B = \{(2,1), (3,1), (3,2), \dots, (6,5)\}$. Portanto, $P(A) = \dfrac{3}{36}$ e $P(B) = \dfrac{15}{36}\cdot$ E $P(B \mid A) = \dfrac{1}{3}$, uma vez que o espaço amostral é, agora, formado por A (isto é, três resultados), e somente um desses três resultados é coerente com o evento B. De modo semelhante, poderemos calcular $P(A \mid B) = 1/15$.

Finalmente, vamos calcular $P(A \cap B)$. O evento $A \cap B$ ocorre se, e somente se, a soma dos dois dados for 10 *e* se o primeiro dado tiver apresentado um valor maior que o segundo dado. Existe apenas *um* desses resultados e, por isso, $P(A \cap B) = 1/36$. Se fizermos um exame cuidadoso dos vários números já calculados, concluiremos que

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \quad \text{e} \quad P(B \mid A) = \frac{P(A \cap B)}{P(A)}.$$

Essas relações não surgiram apenas do particular exemplo que consideramos. Ao contrário, elas são bastante gerais, e nos dão um caminho para *definir rigorosamente* a probabilidade condicionada.

Para sugerir essa definição, voltemos ao conceito de freqüência relativa. Admitamos que um experimento $\mathcal{E}$ tenha sido repetido n vezes. Sejam n_A, n_B e $n_{A\cap B}$ o número de vezes que, respectivamente, os eventos A, B e $A \cap B$ tenham ocorrido em n repetições. Qual o significado de $n_{A\cap B}/n_A$? Representa a freqüência relativa de B naqueles resultados em que A tenha ocorrido. Isto é, $n_{A\cap B}/n_A$ é a freqüência relativa de B, condicionada a que A tenha ocorrido. Poderemos escrever $n_{A\cap B}/n_A$, da seguinte forma:

$$\frac{n_{A\cap B}}{n_A} = \frac{n_{A\cap B}/n}{n_A/n} = \frac{f_{A\cap B}}{f_A},$$

onde $f_{A\cap B}$ e f_A são as freqüências relativas dos eventos $A \cap B$ e A, respectivamente. Como já dissemos (e explicaremos mais tarde) se n, o número de repetições for grande, $f_{A\cap B}$ será próxima de $P(A \cap B)$ e f_A será próxima de $P(A)$. Conseqüentemente, a relação acima sugere que $n_{A\cap B}/n_A$ será próxima de $P(B|A)$. Por isso, estabeleceremos a seguinte definição:

Definição:

$$P(B|A) = \frac{P(A \cap B)}{P(A)}, \text{ desde que } P(A) > 0. \tag{3.1}$$

Comentários: (*a*) É importante compreender que isso não é um teorema (nós não demonstramos coisa alguma), nem é um axioma. Apenas introduzimos a noção intuitiva de probabilidade condicionada e, depois, estabelecemos uma definição formal daquilo que essa noção significa. O fato de que nossa definição formal corresponde à nossa noção intuitiva é fundamentado pelo parágrafo que precede à definição.

(*b*) É assunto simples verificar que $P(B|A)$ para A fixado, satisfaz aos vários postulados de probabilidade das Eq. (1.3). (Ver Probl. 3.22.) Isto é, temos

(1') $0 \leq P(B|A) \leq 1$,

(2') $P(S|A) = 1$,

(3') $P(B_1 \cup B_2|A) = P(B_1|A) + P(B_2|A)$ se $B_1 \cap B_2 = \emptyset$, (3.2)

(4') $P(B_1 \cup B_2 \cup \cdots |A) = P(B_1|A) + P(B_2|A) + \cdots$ se $B_i \cap B_j = \emptyset$ para $i \neq j$.

(*c*) Se $A = S$, $P(B \mid S) = P(B \cap S) \mid P(S) = P(B)$.

(*d*) A cada evento $B \subset S$ poderemos associar dois números, $P(B)$, a probabilidade (não-condicionada) de B, e $P(B \mid A)$, a probabilidade condicionada de B, desde que algum evento A (para o qual $P(A) > 0$) tenha ocorrido. Em geral, essas duas medidas de probabilidade atribuirão probabilidades diferentes ao evento B, como indicaram os exemplos precedentes. Dentro em breve, estudaremos um caso especial importante, para o qual $P(B)$ e $P(B \mid A)$ serão iguais.

(*e*) Observe-se que a probabilidade condicionada está definida em termos da medida de probabilidade não-condicionada P, isto é, se conhecermos $P(B)$ para todo $B \subset S$, poderemos calcular $P(B \mid A)$ para todo $B \subset S$.

Deste modo, temos duas maneiras de calcular a probabilidade condicionada $P(B|A)$:

(*a*) Diretamente, pela consideração da probabilidade de B em relação ao espaço amostral reduzido A.

(*b*) Empregando a definição acima, onde $P(A \cap B)$ e $P(A)$ são calculados em relação ao espaço amostral original S.

Comentário: Se $A = S$, obteremos $P(B|S) = P(B \cap S)/P(S) = P(B)$, porque $P(S) = 1$ e $B \cap S = B$. Isto é como seria de se esperar, porque dizer que S ocorreu é apenas dizer que o experimento *foi* realizado.

Tab. 3.1

	E	M	
N	40	30	70
U	20	10	30
	60	40	100

Exemplo 3.2. Suponha-se que um escritório possua 100 máquinas de calcular. Algumas dessas máquinas são elétricas (E), enquanto outras são manuais (M); e algumas são novas (N), enquanto outras são muito usadas (U). A Tab. 3.1 dá o número de máquinas de cada categoria. Uma pessoa entra no escritório, pega uma máquina ao acaso, e descobre que é nova. Qual será a probabilidade de que seja elétrica? Em termos da notação introduzida, desejamos calcular $P(E|N)$.

Considerando-se somente o espaço amostral reduzido N (isto é, as 70 máquinas novas), temos $P(E|N) = 40/70 = 4/7$. Empregando a definição de probabilidade condicionada, temos que

$$P(E|N) = \frac{P(E \cap N)}{P(N)} = \frac{40/100}{70/100} = \frac{4}{7}.$$

A mais importante conseqüência da definição de probabilidade condicionada acima, é obtida ao se escrever:

$P(A \cap B) = P(B|A)P(A)$ ou, equivalentemente,

$$P(A \cap B) = P(A \mid B)P(B) \qquad (3.3.a)$$

Isto é, algumas vezes, mencionado como o *teorema da multiplicação* de probabilidades.

Podemos aplicar esse teorema para calcular a probabilidade da ocorrência conjunta dos eventos A e B.

Exemplo 3.3. Consideremos novamente o lote formado de 20 peças defeituosas e 80 não-defeituosas, estudado no início da Seç. 3.1. Se escolhermos ao acaso duas peças, sem reposição, qual será a probabilidade de que ambas as peças sejam defeituosas?

Como anteriormente, definamos os eventos A e B, na seguinte forma.

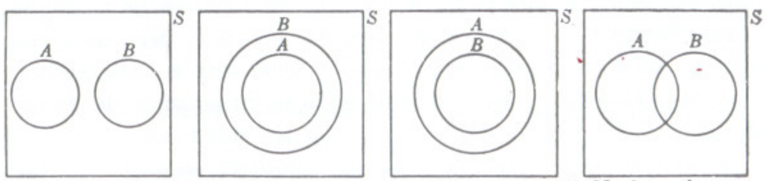
$A = \{$a primeira peça é defeituosa$\}$; $B = \{$a segunda peça é defeituosa$\}$.

Conseqüentemente, pediremos $P(A \cap B)$, que poderemos calcular, de acordo com a fórmula acima, como $P(B \mid A)\,P(A)$. Mas, $P(B \mid A) = 19/99$, enquanto $P(A) = 1/5$. Portanto, $P(A \cap B) =$ $= 19/495$.

Comentário: O teorema da multiplicação de probabilidades (3.3.*a*) pode ser generalizado para mais de dois eventos, da seguinte maneira:

$$P[A_1 \cap A_2 \cap \ldots \cap A_n] =$$

$$= P(A_1)P(A_2 \mid A_1)P(A_3 \mid A_1, A_2)\ldots P(A_n \mid A_1, \ldots A_{n-1}). \qquad (3.3.b)$$



(a) $A \cap B = \emptyset$ (b) $A \subset B$ (c) $B \subset A$ (d) Nenhum desses casos

Fig. 3.2

Examinemos agora, rapidamente, se poderemos fazer uma afirmação geral sobre a grandeza relativa de $P(A \mid B)$ e $P(A)$. Consideraremos quatro casos, que estão ilustrados pelos Diagramas de Venn, na Fig. 3.2. Teremos:

(*a*) $P(A \mid B) = 0 \leq P(A)$, porque A não poderá ocorrer se B tiver ocorrido.

(*b*) $P(A \mid B) = P(A \cap B)/P(B) = [P(A)/P(B)] \geq P(A)$, já que $0 \leq P(B) \leq 1$.

(*c*) $P(A \mid B) = P(A \cap B)/P(B) = P(B)/P(B) = 1 \geq P(A)$.

(*d*) Neste caso nada poderemos afirmar sobre a grandeza relativa de $P(A \mid B)$ e $P(A)$.

Observe-se que em dois dos casos acima, $P(A) \leq P(A \mid B)$; em um caso, $P(A) \geq P(A \mid B)$; e no quarto caso, não podemos fazer qualquer comparação.

Até aqui, empregamos o conceito de probabilidade condicionada a fim de avaliar a probabilidade de ocorrência conjunta de dois eventos. Poderemos aplicar esse conceito em outra maneira de calcular a probabilidade de um evento simples A. Necessitaremos da seguinte definição:

Definição. Dizemos que os eventos $B_1, B_2, \ldots, B_k$ representam uma *partição* do espaço amostral S, quando

(*a*) $B_i \cap B_j = \emptyset$, para todo $i \neq j$.

(*b*) $\bigcup_{i=1}^{k} B_i = S$.

(*c*) $P(B_i) > 0$ para todo i.

Explicando: Quando o experimento $\mathcal{E}$ é realizado *um, e somente um*, dos eventos B_i ocorre.

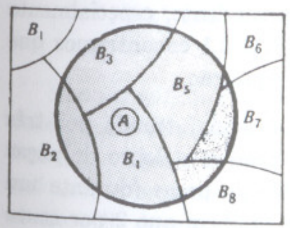


Fig. 3.3

(Por exemplo: na jogada de um dado, $B_1 = \{1, 2\}$, $B_2 = \{3, 4, 5\}$ e $B_3 = \{6\}$ representariam uma partição do espaço amostral, enquanto $C_1 = \{1, 2, 3, 4\}$ e $C_2 = \{4, 5, 6\}$ não o representariam.)

Consideremos A um evento qualquer referente a S, e $B_1, B_2, \ldots, B_k$ uma partição de S. O Diagrama de Venn na Fig. 3.3 ilustra isso para $k = 8$. Portanto, poderemos escrever

$$A = A \cap B_1 \cup A \cap B_2 \cup \ldots \cup A \cap B_k.$$

Naturalmente, alguns dos conjuntos $A \cap B_j$ poderão ser vazios, mas isso não invalida essa decomposição de A. O ponto importante é que todos os eventos $A \cap B_1, \ldots, A \cap B_k$ são dois a dois mutuamente excludentes. Por isso, poderemos aplicar a propriedade da adição de eventos mutuamente excludentes [Eq. (1.3)], e escrever

$$P(A) = P(A \cap B_1) + P(A \cap B_2) + \ldots + P(A \cap B_k).$$

Contudo, cada termo $P(A \cap B_j)$ pode ser expresso na forma $P(A \mid B_j) \cdot P(B_j)$ e, daí, obteremos o que se denomina o teorema da *probabilidade total*:

$$P(A) = P(A \mid B_1)P(B_1) + P(A \mid B_2)P(B_2) + \ldots + P(A \mid B_k)P(B_k). \quad (3.4)$$

Este resultado representa uma relação extremamente útil, porque freqüentemente, quando $P(A)$ é pedida, pode ser difícil calculá-la diretamente. No entanto, com a informação adicional de que B_j tenha ocorrido, seremos capazes de calcular $P(A \mid B_j)$ e, em seguida, empregar a fórmula acima.

Exemplo 3.4. Consideremos (pela última vez) o lote de 20 peças defeituosas e 80 não-defeituosas, do qual extrairemos duas peças, *sem reposição*. Novamente definindo-se A e B como iguais a

$$A = \{\text{a primeira peça extraída é defeituosa}\},$$
$$B = \{\text{a segunda peça extraída é defeituosa}\},$$

poderemos, agora, calcular $P(B)$, assim:

$$P(B) = P(B \mid A)P(A) + P(B \mid \bar{A})P(\bar{A}).$$

Empregando alguns dos cálculos realizados no Ex. 3.3, encontramos que

$$P(B) = \frac{19}{99} \cdot \frac{1}{5} + \frac{20}{99} \cdot \frac{4}{5} = \frac{1}{5}$$

Este resultado pode ser um tanto surpreendente, especialmente se o leitor se recordar de que no início da Seç. 3.1 encontramos que $P(B) = 1/5$, quando extraímos as peças *com* reposição.

Exemplo 3.5. Uma determinada peça é manufaturada por três fábricas, digamos 1, 2 e 3. Sabe-se que 1 produz o dobro de peças que 2, e 2 e 3 produziram o mesmo número de peças (durante um período de produção especificado). Sabe-se também que 2 por cento das peças produzidas por 1 e por 2 são defeituosas, enquanto 4 por cento daquelas produzidas por 3 são defeituosas. Todas as peças produzidas são colocadas em um depósito, e depois uma peça é extraída ao acaso. Qual é a probabilidade de que essa peça seja defeituosa?

Vamos introduzir os seguintes eventos: $A = \{$a peça é defeituosa$\}$, $B_1 = \{$a peça provém de 1$\}$, $B_2 = \{$a peça provém de 2$\}$, $B_3 = \{$a peça provém de 3$\}$.

Pede-se $P(A)$, e empregando-se o resultado acima, poderemos escrever:

$$P(A) = P(A \mid B_1)P(B_1) + P(A \mid B_2)P(B_2) + P(A \mid B_3)P(B_3).$$

Ora, $P(B_1) = 1/2$, enquanto $P(B_2) = P(B_3) = 1/4$. Também, $P(A \mid B_1) = P(A \mid B_2) = 0{,}02$, enquanto $P(A \mid B_3) = 0{,}04$. Levando-se esses valores à expressão acima, encontraremos $P(A) = 0{,}025$.

Comentário: A seguinte analogia com o teorema da probabilidade total é observada em Química: Suponha-se que temos k frascos contendo diferentes soluções de um mesmo sal totalizando, digamos, um litro. Seja $P(B_i)$ o volume do i-ésimo frasco e seja $P(A \mid B_i)$ a concentração da solução no i-ésimo frasco. Se reunirmos todas as soluções em um só frasco e se $P(A)$ denotar a concentração da solução resultante, teremos:

$$P(A) = P(A \mid B_1)P(B_1) + \cdots + P(A \mid B_k)P(B_k).$$

3.2. Teorema de Bayes

Poderemos empregar o Ex. 3.5 para sugerir outro importante resultado. Suponha-se que uma peça seja retirada do depósito e se verifique ser ela defeituosa. Qual é a probabilidade de que tenha sido produzida na fábrica 1?

Empregando a notação já introduzida, pede-se $P(B_1 \mid A)$. Poderemos calcular esta probabilidade como uma conseqüência da seguinte exposição: Seja $B_1, B_2, \ldots, B_k$ uma partição do espaço amostral S e seja A um evento associado a S. Aplicando-se a definição de probabilidade condicionada, poderemos escrever

$$P(B_i \mid A) = \frac{P(A \mid B_i)P(B_i)}{\sum_{j=1}^{k} P(A \mid B_j)P(B_j)} \qquad i = 1, 2, \ldots, k. \quad (3.5)$$

Este resultado é conhecido como *Teorema de Bayes*. É também denominado fórmula da probabilidade das "causas" (ou dos "antecedentes"). Desde que os B_i constituam uma partição do espaço amostral um, e somente um, dos eventos B_i ocorrerá. (Isto é, *um* dos eventos B_i deverá ocorrer e *somente um poderá ocorrer.*) Portanto, a expressão acima nos dá a probabilidade de um particular B_i (isto é, uma "causa"), dado que o evento A *tenha* ocorrido. A fim de aplicar esse teorema, deveremos conhecer os valores das $P(B_i)$. Muito freqüentemente, esses valores são desconhecidos, e isso limita a aplicabilidade do teorema. Tem havido considerável controvérsia sobre o Teorema de Bayes; ele é perfeitamente correto matematicamente; somente a escolha imprópria dos $P(B_i)$ pode tornar o resultado discutível.

Voltando ao problema proposto acima, e agora aplicando a Eq. (3.5), obtemos:

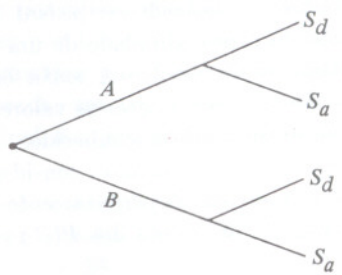
$$P(B_1 | A) = \frac{(0{,}02)(1/2)}{(0{,}02)(1/2) + (0{,}02)(1/4) + (0{,}04)(1/4)} = 0{,}40.$$

Comentário: De novo, podemos encontrar para o Teorema de Bayes, uma analogia da Química. Em k frascos, temos soluções do mesmo sal, porém de concentrações diferentes. Admita-se que o volume total das soluções seja um litro. Denotando por $P(B_i)$ o volume da solução do i-ésimo frasco, e a concentração do sal nesse i-ésimo frasco por $P(A | B_i)$, verificaremos que a Eq. (3.5) fornece a proporção da quantidade total do sal que é encontrada no i-ésimo frasco.

O seguinte exemplo do Teorema de Bayes nos dará uma oportunidade para introduzir a idéia do *diagrama de árvore*, um esquema bastante útil para analisar determinados problemas.

Suponha-se que um grande número de caixas de bombons sejam compostas de dois tipos, A e B. O tipo A contém 70 por cento de bombons doces e 30 por cento de bombons amargos, enquanto no tipo B essas percentagens de sabor são inversas. Além disso, suponha-se que 60 por cento de todas as caixas de bombons sejam do tipo A, enquanto as restantes sejam do tipo B.

Você agora se defronta com o seguinte problema de decisão: uma caixa do tipo desconhecido lhe é oferecida. Você terá permissão para tirar uma amostra de bombom (uma situação reconhecidamente irrealística, mas que nos permitirá introduzir idéias importantes, sem ficar muito complicado), e com esta informação você deve decidir se adivinha que a caixa que lhe foi oferecida é do tipo A ou se do tipo B. O seguinte "diagrama de árvore" (assim denominado por causa dos vários passos ou ramos que aparecem) nos ajudará a analisar o problema. (S_d e S_a correspondem, respectivamente, a escolher um bombom de sabor doce ou um bombom de sabor amargo.)



Façamos alguns cálculos:

$P(A) = 0{,}6; P(B) = 0{,}4; P(S_d | A) = 0{,}7;$

$P(S_a | A) = 0{,}3; P(S_d | B) = 0{,}3; P(S_a | B) = 0{,}7.$

Desejamos realmente saber:

$$P(A | S_d),\quad P(A | S_a),\quad P(B | S_d) \text{ e } P(B | S_a).$$

Suponha-se que realmente retiremos um bombom de sabor doce. Qual decisão seríamos mais tentados a tomar? Vamos comparar

$$P(A | S_d) \text{ e } P(B | S_d).$$

Empregando a fórmula de Bayes, teremos

$$P(A | S_d) = \frac{P(S_d | A)P(A)}{P(S_d | A)P(A) + P(S_d | B)P(B)} =$$

$$= \frac{(0{,}7)(0{,}6)}{(0{,}7)(0{,}6) + (0{,}3)(0{,}4)} = \frac{7}{9}.$$

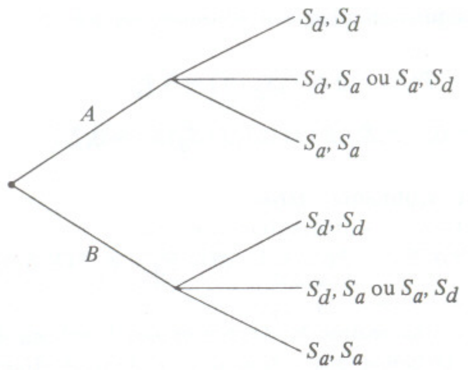
Cálculo semelhante dará

$$P(B | S_d) = 2/9.$$

Dessa maneira, baseados na evidência que tivemos (isto é, a tirada de um bombom de sabor doce) é $2\frac{1}{2}$ vezes mais provável que nós estejamos diante de uma caixa do tipo A, em vez de uma do tipo B. Conseqüentemente, poderíamos presumivelmente decidir que uma caixa do tipo A foi apresentada. (Naturalmente, nós poderíamos estar errados. A sugestão desta análise é que estaremos escolhendo aquela alternativa que pareça a mais provável, com base na evidência limitada que tivermos.)

Em termos do diagrama da árvore, o que era realmente necessário (e foi feito) era uma análise para o passado. Assim, dado o que foi observado S_d, neste caso qual a probabilidade de que o tipo A seja o envolvido?

Uma situação mais interessante surge, se nos for permitido tirar *dois* bombons antes de decidir se se trata do tipo A ou do tipo B. Neste caso, o diagrama de árvore aparece assim:



No problema 3.26, você será chamado a decidir de qual dos dois tipos, A ou B, você tirou a amostra, na dependência de qual seja observado dentre três resultados experimentais possíveis.

3.3. Eventos Independentes

Ja consideramos eventos A e B que não podem ocorrer conjuntamente, isto é, $A \cap B = \emptyset$. Tais eventos são denominados mutuamente excludentes, ou eventos incompatíveis. Observamos anteriormente que se A e B forem mutuamente excludentes, então $P(A \mid B) = 0$, porque a ocorrência dada de B impede a ocorrência de A. No outro extremo, temos a situação já estudada, na qual $B \supset A$ e, conseqüentemente, $P(B \mid A) = 1$.

Em cada uma das situações mencionadas, saber que B já ocorreu nos dá alguma informação bastante definida referente à probabilidade de ocorrência de A. Existem, porém, muitas situações nas quais saber que algum evento B ocorreu não tem qualquer interesse quanto à ocorrência ou não ocorrência de A.

Exemplo 3.6. Suponhamos que um dado equilibrado seja jogado duas vezes. Definamos os eventos A e B, da seguinte forma:

$$A = \{\text{o primeiro dado mostra um número par}\},$$
$$B = \{\text{o segundo dado mostra um 5 ou um 6}\}.$$

É intuitivamente compreensível que os eventos A e B são inteiramente não relacionados. Saber que B ocorreu não fornece qualquer informação sobre a ocorrência de A. De fato, o seguinte cálculo mostra isso. Tomando como nosso espaço amostral os 36 resul-

tados igualmente prováveis, considerados no Ex. 3.1, encontraremos que $P(A) = 18/36 = 1/2$, $P(B) = 12/36 = 1/3$, enquanto $P(A \cap B) = 6/36 = 1/6$. Conseqüentemente, $P(A \mid B) = P(A \cap B) / P(B) = (1/6)/(1/3) = 1/2$.

Deste modo encontramos, como seria de se esperar, que a probabilidade absoluta (ou não condicionada) é igual à probabilidade condicionada $P(A \mid B)$. Semelhantemente,

$$P(B \mid A) = \frac{P(B \cap A)}{P(A)} = \frac{\left(\frac{1}{6}\right)}{\left(\frac{1}{2}\right)} = \frac{1}{3} = P(B).$$

Daí, poderíamos ser tentados a dizer que A e B serão independentes se, e somente se, $P(A \mid B) = P(A)$ e $P(B \mid A) = P(B)$. Muito embora isso pudesse ser essencialmente apropriado, existe outra forma de colocar a questão que contorna a dificuldade encontrada aqui, a saber, que tanto $P(A)$ como $P(B)$ devem ser não-nulos para que as igualdades acima tenham significado.

Consideremos $P(A \cap B)$, supondo que as probabilidades condicionadas sejam iguais às correspondentes probabilidades absolutas. Teremos:

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

Desse modo, desde que nem $P(A)$ nem $P(B)$ sejam iguais a zero, verificamos que as probabilidades absolutas serão iguais às probabilidades condicionadas se, e somente se, $P(A \cap B) = P(A)\,P(B)$. Em conseqüência, formulamos a seguinte definição, a qual será também válida quer $P(A)$ ou $P(B)$ seja nulo:

Definição: A e B serão *eventos independentes* se, e somente se,

$$P(A \cap B) = P(A)P(B). \tag{3.6}$$

Comentário: Esta definição é, essencialmente, equivalente àquela sugerida acima, a saber, que A e B são independentes quando $P(B \mid A) = P(B)$ e $P(A \mid B) = P(A)$. Esta última forma é ligeiramente mais intuitiva, porque diz precisamente o que se tinha tentado dizer antes: que A e B serão independentes se o conhecimento da ocorrência de A de nenhum modo influenciar a probabilidade da ocorrência de B.

Pelo exame do seguinte exemplo, vê-se que a definição formal acima adotada apresenta também uma certa atração intuitiva.

Exemplo 3.7. Consideremos novamente o Ex. 3.2. Inicialmente examinaremos apenas a tabela abaixo, em que são fornecidos somente os valores marginais. Isto é, existem 60 máquinas elétricas e 40 manuais, e delas 70 são novas enquanto 30 são usadas.

	E	M	
N			70
U			30
	60	40	100

Existem muitas maneiras de preencher as casas da tabela, concordantes com os totais marginais dados. A seguir apresentaremos algumas dessas possibilidades.

	E	M	
N	60	10	70
U	0	30	30
	60	40	100

(a)

	E	M	
N	30	40	70
U	30	0	30
	60	40	100

(b)

	E	M	
N	42	28	70
U	18	12	30
	60	40	100

(c)

Consideremos a Tab. (*a*). Aqui *todas* as máquinas elétricas são novas e *todas* as máquinas usadas são manuais. Desse modo, existe uma conexão óbvia (não necessariamente causal) entre a característica de ser elétrica e a de ser nova. Semelhantemente, na Tab. (*b*), *todas* as máquinas manuais são novas e *todas* as máquinas usadas são elétricas. Também, uma conexão definida existe entre essas características. No entanto, quando chegamos à Tab. (*c*), a situação fica bem diferente: aqui, nenhuma relação evidente existe. Por exemplo, 60 por cento de todas as máquinas são elétricas, e exatamente 60 por cento das máquinas usadas são elétricas. Semelhantemente, 70 por cento de todas as máquinas são novas, enquanto exatamente 70 por cento das máquinas manuais são novas etc. Portanto, nenhuma indicação está evidente de que a característica de "ser nova" e de "ser elétrica" tenham qualquer conc-terística de "ser nova" e de "ser elétrica" tenham qualquer cone-xão uma com a outra. Naturalmente, esta tabela foi construída justamente de modo a apresentar essa propriedade. Como foram obtidos os valores das casas da tabela? Apenas com o emprego da Eq. (3.6); isto é, porque $P(E) = 60/100$ e $P(N) = 70/100$, deveremos ter, para independência, $P(E \cap N) = P(E)\,P(N) = 42/100$. Daí, a casa na tabela que indique o número de máquinas elétricas novas deverá conter o número 42. As outras casas seriam obtidas de maneira análoga.

Na maioria das aplicações, teremos que *adotar a hipótese* de independência de dois eventos A e B, e depois empregar essa suposição para calcular $P(A \cap B)$ como igual a $P(A)\,P(B)$. Geralmente,

condições físicas sob as quais o experimento seja realizado tornarão possível decidir se tal suposição será justificada ou ao menos aproximadamente justificada.

Exemplo 3.8. Consideremos um lote grande de peças, digamos 10.000. Admitamos que 10 por cento dessas peças sejam defeituosas e 90 por cento perfeitas. Duas peças são extraídas. Qual é a probabilidade de que ambas sejam perfeitas?

Definamos os eventos A e B, assim:

$$A = \{\text{a primeira peça é perfeita}\},$$
$$B = \{\text{a segunda peça é perfeita}\}.$$

Se admitirmos que a primeira peça seja reposta, antes que a segunda seja escolhida, então os eventos A e B podem ser considerados independentes e, portanto, $P(A \cap B) = (0{,}9)\,(0{,}9) = 0{,}81$. Na prática, contudo, a segunda peça é escolhida sem a reposição da primeira peça; neste caso,

$$P(A \cap B) = P(B|A)P(A) = \frac{8999}{9999}\,(0{,}9)$$

que é aproximadamente igual a 0,81. Assim, muito embora A e B não sejam independentes no segundo caso, a hipótese de independência (que simplifica consideravelmente os cálculos) acarreta apenas um erro desprezível. (Recorde-se o objetivo de um modelo matemático, tal como foi apresentado na Seç. 1.1.) Se existissem somente poucas peças no lote, digamos 30, a hipótese de independência teria acarretado um erro grande. Por isso, torna-se importante verificar cuidadosamente as condições sob as quais o experimento é realizado, a fim de estabelecer a validade de uma suposição de independência entre os vários eventos.

Exemplo 3.9. Admitamos que um mecanismo seja constituído por dois componentes montados em série, como indicado na Fig. 3.4. Cada componente tem uma probabilidade p de não funcionar. Qual será a probabilidade de que o mecanismo funcione?

Fig. 3.4

É evidente que o mecanismo funcionará se, e somente se, *ambos* os componentes estiverem funcionando. Por isso,

Prob (o mecanismo funcione) = Prob (C_1 funcione e C_2 funcione).

A informação fornecida não nos permite continuar sem que se saiba (ou se suponha) que os dois mecanismos trabalhem independentemente um do outro. Isto pode, ou não, ser uma suposição realista, dependendo de como as duas partes sejam engatadas. Se admitirmos que as duas partes trabalhem independentemente, obteremos para a probabilidade pedida o valor $(1 - p)^2$.

Será importante para nós, estendermos a noção de independência para mais de dois eventos. Consideremos, inicialmente, três eventos associados a um experimento, digamos A, B e C. Se A e B, A e C, B e C forem independentes *dois a dois* (no sentido acima), então não se concluirá, em geral, que não exista dependência entre os três eventos. O exemplo seguinte (um tanto artificial) ilustra esse ponto.

Exemplo 3.10. Suponha-se que joguemos dois dados. Definam-se os eventos A, B e C da seguinte forma:

$A = \{$o primeiro dado mostra um número par$\}$,

$B = \{$o segundo dado mostra um número ímpar$\}$,

$C = \{$ambos os dados mostram números ímpares ou ambos mostram números pares$\}$.

Temos $P(A) = P(B) = P(C) = 1/2$. Além disso, $P(A \cap B) = P(A \cap C) = P(B \cap C) = 1/4$. Portanto, os três eventos são todos independentes dois a dois. Contudo, $P(A \cap B \cap C) = 0 \neq P(A)\,P(B)\,P(C)$.

Este exemplo sugere a seguinte definição.

Definição. Diremos que os três eventos A, B e C são *mutuamente independentes* se, e somente se, *todas* as condições seguintes forem válidas:

$$P(A \cap B) = P(A)P(B), \qquad P(A \cap C) = P(A)P(C), \tag{3.7}$$
$$P(B \cap C) = P(B)P(C), \qquad P(A \cap B \cap C) = P(A)P(B)P(C).$$

Finalmente, generalizaremos esta noção para n eventos, na seguinte definição:

Definição. Os n eventos $A_1, A_2, \ldots, A_n$ serão mutuamente independentes se, e somente se, tivermos para $k = 2, 3, \ldots, n$:

$$P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = P(A_{i_1})P(A_{i_2}) \cdots P(A_{i_k}). \tag{3.8}$$

(Existem ao todo $2^n - n - 1$ condições aí arroladas; veja o Probl. 3.18.)

Comentário: Na maioria das aplicações, não precisaremos verificar todas essas condições, porque nós geralmente *admitimos* a independência (baseada naquilo que conhecermos do experimento). Depois, empregaremos essa suposição para calcular, digamos $P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k})$ como $P(A_{i_1})P(A_{i_2}) \cdots P(A_{i_k})$.

Exemplo 3.11. A probabilidade de fechamento de cada relé do circuito apresentado na Fig. 3.5 é dada por p. Se todos os relés funcionarem independentemente, qual será a probabilidade de que haja corrente entre os terminais L e R?

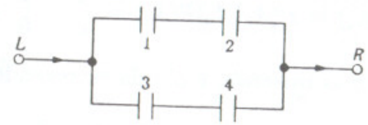


Fig. 3.5

Represente-se por A_i o evento $\{$o relé i está fechado$\}$, $i = 1, 2, 3, 4$. Represente-se por E o evento $\{$a corrente passa de L para $R\}$. Em conseqüência, $E = (A_1 \cap A_2) \cup (A_3 \cap A_4)$. (Observe-se que $A_1 \cap A_2$ e $A_3 \cap A_4$ *não* são mutuamente excludentes.) Portanto,

$$P(E) = P(A_1 \cap A_2) + P(A_3 \cap A_4) - P(A_1 \cap A_2 \cap A_3 \cap A_4)$$
$$= p^2 + p^2 - p^4 = 2p^2 - p^4.$$

Exemplo 3.12. Suponhamos novamente que, para o circuito da Fig. 3.6, a probabilidade de que cada relé esteja fechado é p, e que todos os relés funcionem independentemente. Qual será a probabilidade de que exista corrente entre os terminais L e R?

Empregando a mesma notação do Ex. 3.11, teremos que

$$\begin{aligned} P(E) = {} & P(A_1 \cap A_2) + P(A_5) + P(A_3 \cap A_4) - P(A_1 \cap A_2 \cap A_5) \\ & - P(A_1 \cap A_2 \cap A_3 \cap A_4) - P(A_5 \cap A_3 \cap A_4) \\ & + P(A_1 \cap A_2 \cap A_3 \cap A_4 \cap A_5) \\ = {} & p^2 + p + p^2 - p^3 - p^4 - p^3 + p^5 = p + 2p^2 - 2p^3 - p^4 + p^5. \end{aligned}$$

Vamos encerrar este capítulo com a indicação de uma bastante comum, mas errônea, resolução de um problema.

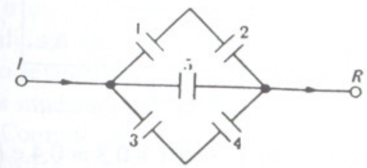


Fig. 3.6

Exemplo 3.13. Admita-se que dentre seis parafusos, dois sejam menores do que um comprimento especificado. Se dois dos parafusos forem escolhidos ao acaso, qual será a probabilidade de que os dois parafusos mais curtos sejam extraídos? Seja A_i o evento {o *i*-ésimo parafuso escolhido é curto}, $i = 1, 2$.

Portanto, desejamos calcular $P(A_1 \cap A_2)$. A solução correta é obtida, naturalmente, escrevendo

$$P(A_1 \cap A_2) = P(A_2 | A_1)P(A_1) = \frac{1}{5} \cdot \frac{2}{6} = \frac{1}{15}.$$

A solução comum, mas *incorreta*, é obtida escrevendo-se

$$P(A_1 \cap A_2) = P(A_2)P(A_1) = \frac{1}{5} \cdot \frac{2}{6} = \frac{1}{15}.$$

Naturalmente, o importante é que, muito embora a resposta esteja numericamente correta, a identificação de 1/5 com $P(A_2)$ é incorreta; 1/5 representa $P(A_2 | A_1)$. Para calcular $P(A_2)$ corretamente, escreveremos

$$P(A_2) = P(A_2 | A_1)P(A_1) + P(A_2 | \overline{A}_1)P(\overline{A}_1) = \frac{1}{5} \cdot \frac{2}{6} + \frac{2}{5} \cdot \frac{4}{6} = \frac{1}{3}.$$

3.4. Considerações Esquemáticas; Probabilidade Condicionada e Independência

A abordagem esquemática seguinte poderá ser útil para compreender a probabilidade condicionada. Suponhamos que A e B sejam dois eventos associados a um espaço amostral para o qual as várias probabilidades estão indicadas no Diagrama de Venn, dado na Fig. 3.7.

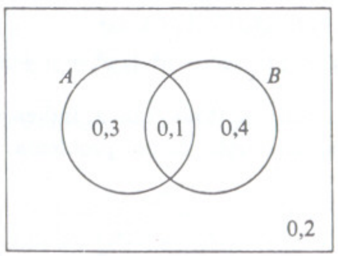


Fig. 3.7

Tem-se $P(A \cap B) = 0{,}1$; $P(A) = 0{,}1 + 0{,}3 = 0{,}4$ e $P(B) = 0{,}1 + 0{,}4 = 0{,}5$.

Em seguida, representaremos as várias probabilidades pelas *áreas* dos retângulos, como na Fig. 3.8. Em cada caso, as regiões sombreadas indicam o evento B: no retângulo da esquerda, estamos representando $A \cap B$ e, no da direita, $A' \cap B$.

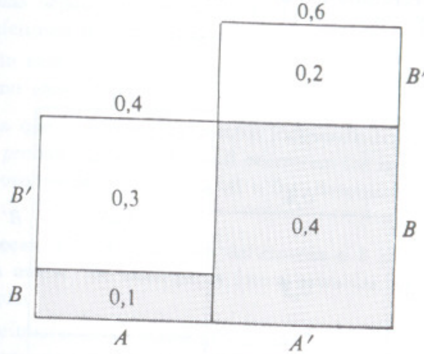


Fig. 3.8

Agora, admitamos que se deseje calcular $P(B \mid A)$. Por isso, necessitamos somente considerar A, isto é, A' pode ser ignorado no cálculo. Observamos que a proporção de B em A é 1/4. (Poderemos também verificar isso pela aplicação da Eq. (3.1): $P(B \mid A) = P(A \cap B) \mid P(A) = 0{,}1/0{,}4 = 1/4$.) Portanto, $P(B' \mid A) = 3/4$, e nosso diagrama representando essa probabilidade condicionada seria dado pela Fig. 3.9.

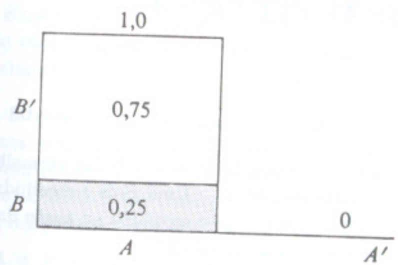


Fig. 3.9

Observe-se, também, que se A for dado como tendo ocorrido, toda a probabilidade (isto é, 1) deverá ser associada ao evento A, enquanto nenhuma probabilidade (isto é, 0) estará associada a A'. Além disso, observe-se que, no retângulo da esquerda, representando A, somente os valores individuais mudaram na Fig. 3.8 para a Fig. 3.9 (cuja soma é 1, em lugar de 0,4). Contudo, as proporções dentro do retângulo permaneceram as mesmas (isto é, 3:1).

Vamos também ilustrar a noção de independência, empregando a abordagem esquemática introduzida anteriormente. Suponhamos que A e B sejam como indicado na Fig. 3.10. Nesse caso, as proporções nos dois retângulos, representando A e A', são *as mesmas*: 3:1 nos dois casos. Por isso, teremos $P(B) = 0{,}1 + 0{,}15 = 0{,}25$ e $P(B \cap A) = 0{,}1/0{,}4 = 0{,}25$.

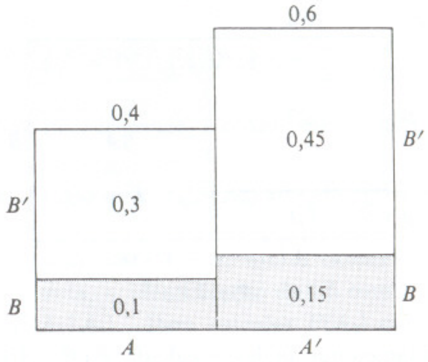


Fig. 3.10

Finalmente, observe-se que, simplesmente olhando a Fig. 3.8, poderemos também calcular as outras probabilidades condicionadas: $P(A \mid B) = 1/5$ (desde que 1/5 da área total retangular representando B esteja ocupada por A); $P(A' \mid B) = 4/5$.

Problemas

3.1. A urna 1 contém x bolas brancas e y bolas vermelhas. A urna 2 contém z bolas brancas e v bolas vermelhas. Uma bola é escolhida ao acaso da urna 1 e posta na urna 2. *A seguir*, uma bola é escolhida ao acaso da urna 2. Qual será a probabilidade de que esta bola seja branca?

3.2. Duas válvulas defeituosas se misturam com duas válvulas perfeitas. As válvulas são ensaiadas, uma a uma, até que ambas as defeituosas sejam encontradas.

(*a*) Qual será a probabilidade de que a última válvula defeituosa seja encontrada no segundo ensaio?

(*b*) Qual será a probabilidade de que a última válvula defeituosa seja encontrada no terceiro ensaio?

(*c*) Qual será a probabilidade de que a última válvula defeituosa seja encontrada no quarto ensaio?

(*d*) Some os números obtidos em (*a*), (*b*) e (*c*) acima. O resultado é surpreendente?

3.3. Uma caixa contém 4 válvulas defeituosas e 6 perfeitas. Duas válvulas são extraídas juntas. Uma delas é ensaiada e se verifica ser perfeita. Qual a probabilidade de que a outra válvula também seja perfeita?

3.4. No problema anterior, as válvulas são verificadas extraindo-se uma válvula ao acaso, ensaiando-a e repetindo-se o procedimento até que todas as 4 válvulas defeituosas sejam encontradas. Qual será a probabilidade de que a quarta válvula defeituosa seja encontrada:

(*a*) No quinto ensaio?

(*b*) No décimo ensaio?

3.5. Suponha que A e B sejam eventos independentes associados a um experimento. Se a probabilidade de A ou B ocorrerem for igual a 0,6, enquanto a probabilidade da ocorrência de A for igual a 0,4, determine a probabilidade da ocorrência de B.

3.6. Vinte peças, 12 das quais são defeituosas e 8 perfeitas, são inspecionadas uma após a outra. Se essas peças forem extraídas ao acaso, qual será a probabilidade de que:

(*a*) As duas primeiras peças sejam defeituosas?

(*b*) As duas primeiras peças sejam perfeitas?

(*c*) Das duas primeiras peças inspecionadas, uma seja perfeita e a outra defeituosa?

3.7. Suponha que temos duas urnas 1 e 2, cada uma com duas gavetas. A urna 1 contém uma moeda de ouro em uma gaveta e uma moeda de prata na outra gaveta; enquanto a urna 2 contém uma moeda de ouro em cada gaveta. Uma urna é escolhida ao acaso; a seguir uma de suas gavetas é aberta ao acaso. Verifica-se que a moeda encontrada nessa gaveta é de ouro. Qual a probabilidade de que a moeda provenha da urna 2?

3.8. Um saco contém três moedas, uma das quais foi cunhada com duas caras, enquanto as duas outras moedas são normais e não viciadas. Uma moeda é tirada ao acaso do saco e jogada quatro vezes, em seqüência. Se sair cara *toda* vez, qual será a probabilidade de que essa seja a moeda de duas caras?

3.9. Em uma fábrica de parafusos, as máquinas A, B e C produzem 25, 35 e 40 por cento do total produzido, respectivamente. Da produção de cada máquina, 5, 4 e 2 por cento, respectivamente, são parafusos defeituosos. Escolhe-se ao acaso um parafuso e se verifica ser defeituoso. Qual será a probabilidade de que o parafuso venha da máquina A? Da B? Da C?

3.10. Sejam A e B dois eventos associados a um experimento. Suponha que $P(A) = 0{,}4$, enquanto $P(A \cup B) = 0{,}7$. Seja $P(B) = p$.

(*a*) Para que valor de p, A e B serão mutuamente excludentes?

(*b*) Para que valor de p, A e B serão independentes?

3.11. Três componentes C_1, C_2 e C_3, de um mecanismo são postos em série (em linha reta). Suponha que esses componentes sejam dispostos em ordem aleatória. Seja R o evento $\{C_2$ está à direita de $C_1\}$, e seja S o evento $\{C_3$ está à direita de $C_1\}$. Os eventos R e S são independentes? Por quê?

3.12. Um dado é lançado e, independentemente, uma carta é extraída de um baralho completo (52 cartas). Qual será a probabilidade de que:

(a) O dado mostre um número par e a carta seja de um naipe vermelho?

(b) O dado mostre um número par ou a carta seja de um naipe vermelho?

3.13. Um número binário é constituído apenas dos dígitos zero e um. (Por exemplo, 1 011, 1 100 etc.) Esses números têm importante papel na utilização de computadores eletrônicos. Suponha que um número binário seja formado de n dígitos. Suponha que a probabilidade de um dígito incorreto aparecer seja p e que os erros em diferentes dígitos sejam independentes uns dos outros. Qual será a probabilidade de formar-se um *número* incorreto?

3.14. Um dado é atirado n vezes. Qual é a probabilidade de que "6" apareça ao menos uma vez em n jogadas?

3.15. Cada uma de duas pessoas joga três moedas equilibradas. Qual é a probabilidade de que elas obtenham o mesmo número de caras?

3.16. Jogam-se dois dados. Desde que as faces mostrem números diferentes, qual é a probabilidade de que uma face seja 4?

3.17. Sabe-se que na fabricação de um certo artigo, defeitos de um tipo ocorrem com probabilidade 0,1 e defeitos de outro tipo com probabilidade 0,05. Qual será a probabilidade de que:

(a) Um artigo não tenha ambos os tipos de defeitos?

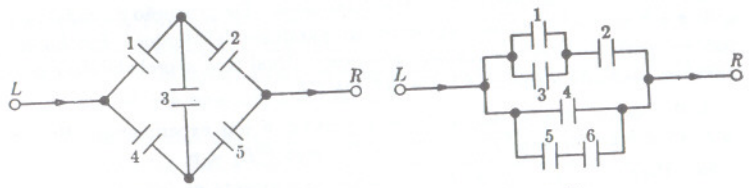
(b) Um artigo seja defeituoso?

(c) Um artigo tenha apenas um tipo de defeito, sabido que é defeituoso?

3.18. Verifique que o número de condições impostas pela Eq. (3.8) é dado por $2^n - n - 1$.

3.19. Demonstre que, se A e B forem eventos independentes, também o serão A e $\overline{B}$, $\overline{A}$ e B, $\overline{A}$ e $\overline{B}$.

3.20. Na Fig. 3.11 (a) e (b), suponha que a probabilidade de que cada relé esteja fechado seja p, e que cada relé seja aberto ou fechado independentemente um do outro. Em cada caso, determine a probabilidade de que a corrente passe de L para R.



(a) (b)

Fig. 3.11

3.21. Duas máquinas A e B, sendo operadas independentemente, podem ter alguns desarranjos cada dia. A Tab. 3.2 dá a distribuição de probabilidades dos desarranjos para cada máquina. Calcule as seguintes probabilidades:

(a) A e B tenham o mesmo número de desarranjos.

(b) O número total de desarranjos seja menor que 4; menor que 5.

(c) A tenha mais desarranjos que B.

(d) B tenha duas vezes mais desarranjos que A.

(e) B tenha 4 desarranjos, quando se saiba que B já tenha tido 2 desarranjos.

(f) O número mínimo de desarranjos das duas máquinas seja 3; seja menor do que 3.

(g) O número máximo de desarranjos das máquinas seja 3; seja maior que 3.

Tab. 3.2

Número de desarranjos	0	1	2	3	4	5	6
A	0,1	0,2	0,3	0,2	0,09	0,07	0,04
B	0,3	0,1	0,1	0,1	0,1	0,15	0,15

3.22. Verifique pelas Eqs. (3.2) que, sendo A fixo, $P(B|A)$ satisfaz aos vários postulados da probabilidade.

3.23. Se cada elemento de um determinante de segunda ordem for zero ou um, qual será a probabilidade de que o valor do determinante seja positivo? (Admita que os elementos do determinante sejam escolhidos independentemente, a cada valor se atribuindo a probabilidade 1/2.)

3.24. Verifique que o teorema da multiplicação $P(A \cap B) = P(A|B)P(B)$, estabelecido para dois eventos, pode ser estendido para três eventos, da seguinte maneira:

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

3.25. Uma montagem eletrônica é formada de dois subsistemas A e B. De procedimentos de ensaio anteriores, as seguintes probabilidades se admitem conhecidas:

$P(A$ falhe$) = 0{,}20$, $P(A$ e B falhem$) = 0{,}15$, $P(B$ falhe *sozinho*$) = 0{,}15$.

Calcule as seguintes probabilidades:

(a) $P(A$ falhe $|$ B tenha falhado$)$. (b) $P(A$ falhe sozinho$)$.

3.26. Conclua a análise do exemplo dado na Seção 3.2, pela decisão de qual dos dois tipos de caixa de bombons, A ou B, foi apresentada, baseando-se na evidência dos dois bombons que foram tirados na amostra.

3.27. Sempre que um experimento é realizado, a ocorrência de um particular evento A é igual a 0,2. O experimento é repetido independentemente, até que A ocorra. Calcule a probabilidade de que seja necessário levar a cabo o experimento até a quarta vez.

3.28. Suponha que um equipamento possua N válvulas, todas necessárias para seu funcionamento. A fim de localizar uma válvula com mau funcionamento, faz-se a substituição de cada válvula, sucessivamente, por uma válvula nova. Calcule a probabilidade de que seja necessário trocar N válvulas, se a probabilidade (constante) de uma válvula estar desarranjada por p.

3.29. Demonstre: Se $P(A \mid B) > P(A)$, então, $P(B \mid A) > P(B)$.

3.30. Uma válvula a vácuo pode provir de três fabricantes, com probabilidades $p_1 = 0{,}25$, $p_2 = 0{,}50$ e $p_3 = 0{,}25$. As probabilidades de que, durante determinado período de tempo, a válvula funcione bem são, respectivamente, 0,1; 0,2 e 0,4 para cada um dos fabricantes. Calcule a probabilidade de que uma válvula escolhida ao acaso funcione bem durante o período de tempo especificado.

3.31. Um sistema elétrico é composto de dois comutadores do tipo A, um do tipo B, e quatro do tipo C, ligados como indica a Fig. 3.12. Calcule a probabilidade de que uma pane no circuito não possa ser eliminada com a chave K, se os comutadores A, B e C estiverem abertos (isto é, desligados) com probabilidades 0,3; 0,4 e 0,2, respectivamente, e se eles operarem independentemente.

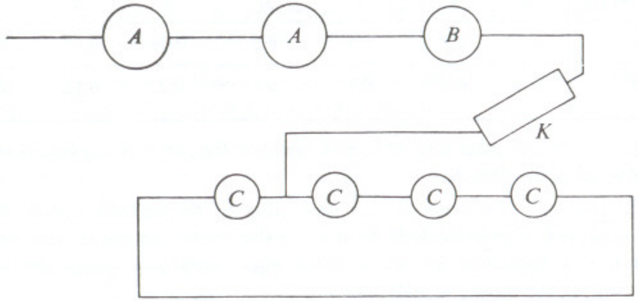


Fig. 3.12

3.32. A probabilidade de que um sistema fique sobrecarregado é 0,4 durante cada etapa de um experimento. Calcule a probabilidade de que o sistema deixe de funcionar em três tentativas independentes do experimento, se as probabilidades de falhas em 1, 2 ou 3 tentativas forem iguais, respectivamente, a 0,2; 0,5 e 0,8.

3.33. Quatro sinais de rádio são emitidos sucessivamente. Se a recepção de cada um for independente da recepção de outro, e se essas probabilidades forem 0,1; 0,2; 0,3 e 0,4, respectivamente, calcule a probabilidade de que k sinais venham a ser recebidos para $k = 0, 1, 2, 3, 4$.

3.34. A seguinte (de algum modo simplória) previsão de tempo é empregada por um amador. O tempo, diariamente, é classificado como "seco" ou "úmido", e supõe-se que a probabilidade de que qualquer dia dado seja igual ao dia anterior seja uma constante p $(0 < p < 1)$. Com base em registros passados, admite-se que 1º de janeiro tenha probabilidade β de ser dia "seco". Fazendo β_n = probabilidade (de que o n-ésimo dia do ano seja "seco"), pede-se obter uma expressão para β_n em termos de β e de p. Calcule também $\lim_{n \to \infty} \beta_n$ e interprete o seu resultado [*Sugestão:* Exprima β_n em termos de β_{n-1}.]

3.35. Três jornais A, B e C são publicados em uma cidade e uma recente pesquisa entre os leitores indica o seguinte: 20 por cento lêem A; 26 por cento lêem B; 14 por cento lêem C; 8 por cento lêem A e B; 5 por cento lêem A e C; 2 por cento lêem A, B e C; e 40 por cento lêem B e C. Para um adulto escolhido ao acaso, calcule a probabilidade de que: (a) ele não leia qualquer dos jornais; (b) ele leia exatamente um dos jornais; (c) ele leia ao menos A e B, se se souber que ele lê ao menos um dos jornais publicados.

3.36. Uma moeda equilibrada é jogada $2n$ vezes. (a) Obtenha a probabilidade de que ocorrerá um igual número de caras e coroas; (b) Mostre que a probabilidade calculada em (a) é uma função decrescente de n.

3.37. Cada uma das n urnas: Urna 1, Urna 2, . . . , Urna n, contém α bolas brancas e β bolas pretas. Uma bola é retirada da Urna 1 e posta na Urna 2; em seguida, uma bola é retirada da Urna 2 e posta na Urna 3, e assim por diante. Finalmente, uma bola é retirada da Urna n. Se a primeira bola transferida for branca, qual será a probabilidade de que a última bola escolhida seja branca? Que acontece, se $n \to \infty$? [*Sugestão:* Faça p_n = Prob (a n-ésima bola transferida seja branca) e exprima p_n em termos de p_{n-1}.]

3.38. A Urna 1 contém α bolas brancas e β bolas pretas, enquanto a Urna 2 contém β bolas brancas e α pretas. Uma bola é extraída (de uma das urnas) e é em seguida reposta naquela urna. Se a bola extraída for branca, escolha a próxima bola da Urna 1; se a bola extraída for preta, escolha a próxima bola da Urna 2. Continue a operar dessa maneira. Dado que a primeira bola escolhida venha da Urna 1, calcule Prob (n-ésima bola escolhida seja branca) e também o limite dessa probabilidade, quando $n \to \infty$.

3.39. Uma máquina impressora pode imprimir n letras, digamos $\alpha_1, \alpha_2, \ldots, \alpha_n$. Ela é acionada por impulsos elétricos, cada letra sendo produzida por um impulso diferente. Suponha que exista uma probabilidade constante p de imprimir a letra correta e também suponha independência. Um dos n impulsos, escolhido ao acaso, foi alimentado na máquina duas vezes e, em ambas, a letra α_1 foi impressa. Calcule a probabilidade de que o impulso escolhido tenha sido para imprimir α_1.

Variáveis Aleatórias Unidimensionais

Capítulo 4

4.1. Noção Geral de Variável Aleatória

Ao descrever o espaço amostral de um experimento, não especificamos que um resultado individual necessariamente seja um número. De fato, apresentamos alguns exemplos nos quais os resultados do experimento não eram uma quantidade numérica. Por exemplo, ao descrever uma peça manufaturada, podemos empregar apenas as categorias "defeituosa" e "não defeituosa". Também, ao observar a temperatura durante o período de 24 horas, podemos simplesmente registrar a curva traçada pelo termógrafo. Contudo, em muitas situações experimentais, estaremos interessados na mensuração de alguma coisa e no seu registro como um *número*. Mesmo nos casos mencionados acima, poderemos atribuir um número a cada resultado (não numérico) do experimento. Por exemplo, poderemos atribuir o valor um às peças perfeitas e o valor zero às defeituosas. Poderemos registrar a temperatura máxima do dia, ou a temperatura mínima, ou a média das temperaturas máxima e mínima.

Os exemplos acima são bastante típicos de uma classe muito geral de problemas: em muitas situações experimentais, desejamos atribuir um número real x a todo elemento s do espaço amostral S. Isto é, $x = X(s)$ é o valor de uma função X do espaço amostral no espaço dos números reais. Com isto em mente, formulamos a seguinte definição.

Definição. Sejam $\mathcal{E}$ um experimento e S um espaço amostral associado ao experimento. Uma *função* X, que associe a cada elemento $s \in S$ um número real, $X(s)$, é denominada *variável aleatória.*

Comentários: (a) A terminologia acima é um tanto infeliz, mas é tão universalmente aceita, que não nos afastaremos dela. Tornamos tão claro quanto possível que X é uma *função*, e contudo, a denominamos uma variável (aleatória)!

(*b*) É evidente que *nem* toda função imaginável pode ser considerada uma variável aleatória. Um requisito (embora não seja o mais geral) é que, para todo número real x, o evento $[X(s) = x]$ e, para todo intervalo I, o evento $[X(s) \in I]$ têm probabilidades bem definidas, consistentes com os axiomas básicos. Na maioria das aplicações, essa dificuldade não surge e nós não voltaremos a nos referir a ela.

(*c*) Em algumas situações, o resultado s do espaço amostral já constitui a característica numérica que desejamos registrar. Simplesmente tomaremos $X(s) = s$, a função identidade.

(*d*) Na maior parte de nossa subseqüente exposição sobre variáveis aleatórias, não necessitaremos indicar a natureza funcional de X. Geralmente, estaremos interessados nos *valores* possíveis de X, mais do que de onde eles se originam. Por exemplo, suponha-se que atiremos duas moedas e consideremos o espaço associado a este experimento. Isto é,

$$S = \{HH, HT, TH, TT\}.$$

Definamos a variável aleatória da seguinte maneira: X é o número de caras (H) obtidas nas duas moedas. Daí, $X(HH) = 2$, $X(HT) = X(TH) = 1$ e $X(TT) = 0$.

S = espaço amostral de $\mathcal{E}$ R_X = valores possíveis de X

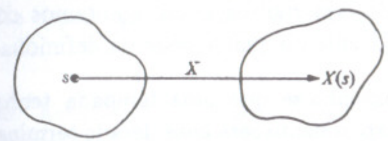


Fig. 4.1

(*e*) É muito importante compreender uma exigência fundamental de uma função (unívoca): A *cada* $s \in S$ corresponderá *exatamente um* valor $X(s)$. Isto está apresentado esquematicamente na Fig. 4.1. Diferentes valores de s *podem* levar ao mesmo valor de X. Por exemplo, na ilustração acima, verificamos que $X(HT) = X(TH) = 1$.

O espaço R_X, conjunto de todos os valores possíveis de X, é algumas vezes denominado *contradomínio*. De certo modo, poderemos considerar R_X como um outro espaço amostral. O espaço amostral (original) S corresponde ao resultado (possivelmente não-numérico) do experimento, enquanto R_X é o espaço amostral associado à variável aleatória X, representando a característica numérica que nos poderá interessar. Se for $X(s) = s$, teremos $S = R_X$.

Muito embora estejamos prevenidos do perigo didático inerente a dar muitas explicações para uma mesma coisa, vamos salientar que poderemos pensar em uma variável aleatória X, de duas maneiras:

(*a*) Realizamos o experimento $\mathcal{E}$ que dá um resultado $s \in S$; *a seguir* calculamos o número $X(s)$.

(*b*) Realizamos $\mathcal{E}$, obtemos o resultado s, e (imediatamente) calculamos $X(s)$. Neste caso, o número $X(s)$ é pensado como o próprio resultado do experimento e R_X se torna *o* espaço amostral do experimento.

A diferença entre as interpretações (*a*) e (*b*) é percebida com dificuldade; é relativamente secundária, mas merecedora de atenção. Em (*a*), o experimento essencialmente termina com a observação de s. A avaliação de $X(s)$ é considerada alguma coisa que é feita posteriormente, e que não é influenciada pela aleatoriedade de $\mathcal{E}$. Em (*b*), o experimento não é considerado concluído até que o número $X(s)$ tenha sido realmente calculado, desse modo se originando o espaço amostral R_X. Muito embora a primeira interpretação, (*a*), seja aquela geralmente pretendida, a segunda interpretação, (*b*), poderá ser muito útil e o leitor deverá lembrar-se dela.

Aquilo que estamos dizendo, e isso ficará cada vez mais claro nas seções posteriores, é que no estudo das variáveis aleatórias estaremos mais interessados nos valores que X toma do que em sua forma funcional. Conseqüentemente, em muitos casos, ignoramos completamente o espaço amostral subjacente no qual X pode ser definido.

Exemplo 4.1. Suponha-se que uma lâmpada tenha sido posta em um soquete. O experimento será considerado terminado quando a lâmpada se queimar. Qual será um possível resultado, s? Uma das maneiras de descrever s seria apenas registrar o dia e a hora em que a lâmpada se queimou, por exemplo: 19 de maio, 16 h e 32 min. Em conseqüência, o espaço amostral poderia ser representado por $S = \{(d, t) \mid d = \text{dia},\ t = \text{momento do dia}\}$. Presumivelmente, a variável aleatória que interessa é X, a duração até queimar. Observe-se que, uma vez que $s = (d, t)$ tenha sido observado, o *cálculo* de $X(s)$ não inclui qualquer aleatoriedade. Quando s é especificado, $X(s)$ fica completamente determinado.

As duas interpretações explicadas acima podem ser aplicadas a este exemplo, como se segue. Em (*a*), consideramos o experimento terminado com a observação $s = (d, t)$, o dia e a hora. O cálculo de $X(s)$ é realizado depois, abrangendo uma operação aritmética simples. Em (*b*), consideramos que o experimento somente estará terminado *depois* que $X(s)$ tenha sido calculado e um número, por exemplo, $X(s) = 107$ horas seja então considerado o resultado do experimento.

Pode-se salientar que análise semelhante se aplicaria a qualquer outra variável que interessasse, por exemplo, $Y(s)$, a temperatura da sala no momento em que a lâmpada se tenha queimado.

Exemplo 4.2. Três moedas são atiradas sobre a mesa. Tão logo as moedas repousem, a fase "aleatória" do experimento terminou. Um resultado simples s poderia consistir na descrição detalhada de como e onde as moedas pousaram. Presumivelmente, estaremos somente interessados em certas características numéricas associadas a este experimento. Por exemplo, poderíamos avaliar:

$X(s)$ = número de caras que apareceram,

$Y(s)$ = distância máxima entre duas moedas quaisquer,

$Z(s)$ = distância mínima das moedas a um bordo qualquer da mesa.

Se for a variável X que interesse, poderemos, como se explicou no exemplo anterior, incluir a avaliação de $X(s)$ na descrição de nosso experimento e, depois, simplesmente afirmar que o espaço amostral associado ao experimento é $\{0, 1, 2, 3\}$, correspondendo aos valores de X. Conquanto muito freqüentemente venhamos a adotar esta interpretação, é importante compreender que a contagem do número de caras é feita *depois* que os aspectos aleatórios do experimento tenham terminado.

Comentário: Referindo-nos a variáveis aleatórias, empregamos quase sem exceção letras maiúsculas, como X, Y, Z etc. Contudo, quando falamos do *valor* que essas variáveis aleatórias tomam, usaremos, em geral, letras minúsculas, como x, y, z etc. Esta é uma *distinção muito importante* a ser feita e o estudante pode bem parar para considerá-la. Por exemplo, quando nós falamos em escolher uma pessoa ao acaso, de alguma população designada, e medimos sua altura (em centímetros, por exemplo), poderemos nos referir aos resultados *possíveis* como uma variável aleatória X. Poderemos então formular várias questões sobre X, como indagar se $P(X \geqslant 60)$. No entanto, uma vez que tenhamos escolhido uma pessoa e medido sua altura, obteremos um valor específico de X, digamos x. Por isso, não teria sentido indagar se $P(x \geqslant 60)$, uma vez que x é ou não é $\geqslant 60$. Esta distinção entre uma variável aleatória e seu valor é importante e nós voltaremos a fazer referência a ela.

Quando estivermos interessados nos eventos associados a um espaço amostral S, verificaremos a necessidade de examinar os eventos relativamente à variável aleatória X, isto é, subespaços do contradomínio R_X. Bastante freqüentemente, certos eventos associados a S são "relacionados" (em um sentido a ser explicado) a eventos associados com R_X, na seguinte forma:

Definição. Sejam um experimento $\mathcal{E}$ e seu espaço amostral S. Seja X uma variável aleatória definida em S e seja R_X seu contradomínio. Seja B um evento definido em relação a R_X, isto é, $B \subset R_X$.

Então, A será definido assim:

$$A = \{s \in S \mid X(s) \in B\}. \tag{4.1}$$

Explicando: A será constituído por todos os resultados em S, para os quais $X(s) \in B$ (veja Fig. 4.2). Neste caso, diremos que A e B são *eventos equivalentes*.

Fig. 4.2

Comentários: (*a*) Dizendo a mesma coisa, com menos rigor: A e B serão equivalentes sempre que ocorram juntos. Isto é, quando A ocorre, B ocorre, e inversamente. Porque se A tiver ocorrido, então um resultado s terá ocorrido, para o qual $X(s) \in B$ e, portanto, B ocorreu. Reciprocamente, se B ocorreu, um valor $X(s)$ terá sido observado, para o qual $s \in A$ e, portanto, A ocorreu.

(*b*) É importante compreender que, em nossa definição de eventos equivalentes, A e B são associados a espaços amostrais *diferentes*.

Exemplo 4.3. Considere-se a jogada de duas moedas. Daí, $S = \{HH, HT, TH, TT\}$. Seja X o número de caras obtido. Portanto, $R_X = \{0, 1, 2\}$. Seja $B = \{1\}$. Já que $X(HT) = X(TH) = 1$ se, e somente se, $X(s) = 1$, temos que $A = \{HT, TH\}$ é equivalente a B.

Agora, daremos a seguinte importante definição.

Definição. Seja B um evento no contradomínio R_X. Nesse caso, *definimos* $P(B)$ da seguinte maneira

$$P(B) = P(A), \quad \text{onde} \quad A = \{s \in S \mid X(s) \in B\}. \tag{4.2}$$

Explicando: Definimos $P(B)$ igual à probabilidade do evento $A \subset S$, o qual é equivalente a B, no sentido da Eq. (4.1).

Comentários: (*a*) Estamos admitindo que probabilidades possam ser associadas a eventos em S. Portanto, a definição acima torna possível atribuir probabilidades a eventos associados a R_X em termos de probabilidades definidas sobre S.

(*b*) É realmente possível *demonstrar* que $P(B)$ deve ser definida tal como o fizemos. Contudo, isto envolveria algumas dificuldades teóricas que desejamos evitar e, por isso, procedemos como acima.

(*c*) Desde que na formulação da Eq. (4.2) os eventos A e B se referem a espaços amostrais diferentes, deveríamos realmente empregar notação diferente quando nos referíssemos a probabilidades definidas sobre S e àquelas definidas sobre R_X, digamos alguma coisa tal como $P(A)$ e $P_X(B)$. No entanto, não faremos isso, mas continuaremos simplesmente a escrever $P(A)$ e $P(B)$. O contexto em que tais expressões apareçam tornará clara a interpretação.

(*d*) As probabilidades associadas a eventos no espaço amostral (original) S são, de certo modo, determinadas por "forças fora de nosso controle", ou como às vezes se diz "pela Natureza". A composição de uma fonte radioativa que emita partículas, a disposição de um grande número de pessoas que façam chamadas telefônicas durante certa hora, e a agitação térmica que dê origem a um fluxo ou as condições atmosféricas que dêem origem a uma tempestade, ilustram esse aspecto. Quando introduzimos uma variável aleatória X e seu contradomínio R_X estamos *induzindo* probabilidades nos eventos associados a R_X, as quais serão estritamente determinadas se as probabilidades associadas a eventos em S forem especificadas.

Exemplo 4.4. Se as moedas consideradas no Ex. 4.3 forem "equilibradas", teremos $P(HT) = P(TH) = 1/4$. Portanto, $P(HT, TH) = 1/4 + 1/4 = 1/2$. (Os cálculos acima são uma conseqüência direta de nossa suposição fundamental referente à propriedade de equilíbrio ou simetria das moedas.) Visto que o evento $\{X = 1\}$ é equivalente ao evento $\{HT, TH\}$, empregando a Eq. (4.1), teremos que $P(X = 1) = P(HT, TH) = 1/2$. [Na realidade não existe escolha para o valor de $P(X = 1)$ coerente com a Eq. (4.2), uma vez que $P(HT, TH)$ tenha sido determinada. É neste sentido que probabilidades associadas a eventos de R_X são *induzidas*.]

Comentário: Agora que já estabelecemos a existência de uma função de probabilidade induzida sobre o contradomínio de X — Eqs. (4.1 e 4.2) — achamos conveniente *suprimir* a natureza funcional de X. Por isso, escreveremos (como fizemos no exemplo acima) $P(X = 1) = 1/2$. O que se quer dizer é que, um certo evento no espaço amostral S, a saber $\{HT, TH\} = \{s \mid X(s) = 1\}$ ocorre com probabilidade 1/2. Daí atribuirmos essa mesma probabilidade ao evento $\{X = 1\}$ no contradomínio. Continuaremos a escrever expressões semelhantes a $P(X = 1)$, $P(X \leq 5)$ etc. É *muito importante* para o leitor compreender o que essas expressões realmente representam.

Uma vez que as probabilidades associadas aos vários resultados (ou eventos) no contradomínio R_X tenham sido determinadas (mais precisamente, induzidas), ignoraremos freqüentemente o espaço amostral original S, que deu origem a essas probabilidades. Assim, no exemplo anterior, simplesmente estaremos interessados em $R_X = \{0, 1, 2\}$ e as probabilidades associadas (1/4, 1/2, 1/4). O fato, de que essas probabilidades sejam *determinadas* por uma função de probabilidade definida sobre o espaço amostral original S, não nos interessa, quando estamos apenas interessados em estudar os *valores* da variável aleatória X.

Ao apresentar, em minúcias, muitos dos importantes conceitos referentes a variáveis aleatórias, julgamos conveniente distinguir dois casos importantes: as variáveis aleatórias discretas e as variáveis aleatórias contínuas.

4.2. Variáveis Aleatórias Discretas

Definição. Seja X uma variável aleatória. Se o número de valores possíveis de X (isto é, R_X, o contradomínio) for finito ou infinito numerável, denominaremos X de ***variável aleatória discreta***. Isto é, os *valores* possíveis de X, podem ser postos em lista como $x_1, x_2, \ldots, x_n$. No caso finito, a lista acaba, e no caso infinito numerável, a lista continua indefinidamente.

Exemplo 4.5. Uma fonte radioativa está emitindo partículas α. A emissão dessas partículas é observada em um dispositivo contador, durante um período de tempo especificado. A variável aleatória seguinte é a que interessa:

$$X = \text{número de partículas observadas.}$$

Quais são os valores possíveis de X? Admitiremos que esses valores são todos os inteiros não negativos, isto é, $R_X = \{0, 1, 2 \ldots, n, \ldots\}$. Uma objeção com que já nos defrontamos uma vez pode, novamente, ser levantada neste ponto. Pode-se argumentar que durante um especificado intervalo (finito) de tempo, é impossível observar mais do que, digamos N partículas, onde N pode ser um inteiro positivo muito grande. Conseqüentemente, os valores possíveis para X realmente seriam: $0, 1, 2, \ldots, N$. Contudo, torna-se matematicamente mais simples considerar a descrição idealizada feita acima. De fato, sempre que admitirmos que os valores possíveis de uma variável aleatória X sejam infinito numerável, estaremos realmente considerando uma representação idealizada de X.

À vista de nossas explicações anteriores da descrição probabilística de eventos com um número finito ou infinito numerável de elementos, a descrição probabilística de uma variável aleatória discreta não apresentará qualquer dificuldade. Procederemos da seguinte maneira:

Definição. Seja X uma variável aleatória discreta. Portando, R_X, o contradomínio de X, será formado no máximo por um número infinito numerável de valores $x_1, x_2, \ldots$ A cada possível resultado x_i associaremos um número $p(x_i) = P(X = x_i)$, denominado probabilidade de x_i. Os números $p(x_i)$, $i = 1, 2, \ldots$ devem satisfazer às

seguintes condições:

$$(a)\quad p(x_i) \geq 0 \text{ para todo } i,$$

$$(b)\quad \sum_{i=1}^{\infty} p(x_i) = 1. \tag{4.3}$$

A função p, definida acima, é denominada *função de probabilidade* (ou função de probabilidade no ponto) da variável aleatória X. A coleção de pares $[x_i, p(x_i)]$, $i = 1, 2, \ldots$, é algumas vezes denominada *distribuição de probabilidade* de X.

Comentários: (*a*) A escolha particular dos números $p(x_i)$ é presumivelmente determinada a partir da função de probabilidade associada aos eventos no espaço

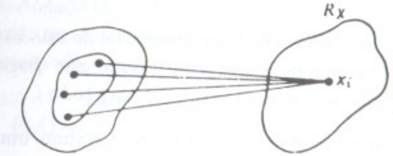


Fig. 4.3

amostral S, no qual X seja definida. Isto é, $p(x_i) = P[s \mid X(s) = x_i]$. [Veja as Eqs. (4.1 e 4.2).] Contudo, já que estamos interessados apenas nos valores de X, isto é, R_X, e as probabilidades associadas a estes valores, estaremos novamente suprimindo a natureza funcional de X. (Veja a Fig. 4.3.) Muito embora, na maioria dos casos, os números sejam de fato determinados a partir da distribuição de probabilidades em algum espaço amostral subjacente S, *qualquer* conjunto de números $p(x_i)$, que satisfaçam às Eqs. (4.3), pode servir como descrição probabilística apropriada de uma variável aleatória discreta.

(*b*) Se X tomar apenas um número finito de valores, digamos $x_1, \ldots, x_N$, então $p(x_i) = 0$ para $i > N$, e, portanto, a série infinita na Eq. (4.3) se transforma em uma soma finita.

(*c*) Podemos salientar, novamente, uma analogia com a Mecânica, ao considerarmos a massa total de uma unidade distribuída sobre a reta real, com a massa total concentrada nos pontos $x_1, x_2, \ldots$ Os números $p(x_i)$ representam a quantidade de massa localizada no ponto x_i.

(*d*) A interpretação geométrica (Fig. 4.4) de uma distribuição de probabilidade é freqüentemente útil.

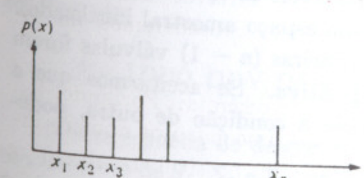


Fig. 4.4

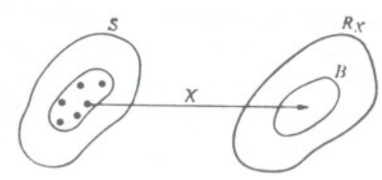


Fig. 4.5

Seja B um evento associado à variável aleatória X; isto é, $B \subset R_X$ (Fig. 4.5). Suponha-se, especificamente, que $B = \{x_{i_1}, x_{i_2}, \ldots\}$. Daí,

$$P(B) = P[s \mid X(s) \in B] \text{ (porque esses eventos são equivalentes)}$$

$$= P[s \mid X(s) = x_{i_j}, j = 1, 2, \ldots] = \sum_{j=1}^{\infty} p(x_{i_j}). \qquad (4.4)$$

Explicando: A probabilidade de um evento B é igual à soma das probabilidades dos resultados individuais associados com B.

Comentários: (*a*) Suponhamos que a variável aleatória discreta X possa tomar somente um número finito de valores, $x_1, \ldots, x_N$. Se os resultados forem igualmente prováveis, então teremos obviamente $p(x_1) = \ldots = p(x_N) = 1/N$.

(*b*) Se X tomar um número infinito numerável de valores, então é *impossível* ter todos os valores igualmente prováveis; porque não poderemos satisfazer à condição $\sum_{i=1}^{\infty} p(x_i) = 1$, se tivermos $p(x_i) = c$ para todo i.

(*c*) Em todo intervalo finito, existirá no máximo um número finito de valores possíveis de X. Se algum desses intervalos não contiver qualquer desses valores possíveis, nós atribuiremos a ele probabilidade zero. Assim, se $R_X = [x_1, x_2, \ldots, x_n]$ e se nenhum $x_i \in [a, b]$, então $P[a \leqslant X \leqslant b] = 0$.

Exemplo 4.6. Suponhamos que uma válvula eletrônica seja posta em um soquete e ensaiada. Admitamos que a probabilidade de que o teste seja positivo seja 3/4; daí, a probabilidade de que seja negativo é igual a 1/4. Admitamos também que estejamos ensaiando uma partida grande dessas válvulas. Os ensaios continuam até que a primeira válvula positiva apareça. Definamos a variável aleatória, assim: X é o número de testes necessários para concluir o experimento. O espaço amostral associado a este experimento é:

$$S = \{+, -\,+, -\,-\,+, -\,-\,-\,+, \ldots\}.$$

Para determinarmos a distribuição de probabilidade de X, raciocinaremos da seguinte forma: os valores possíveis de X são $1, 2, \ldots n, \ldots$ (estamos, obviamente, tratando com um espaço amostral idealizado). E será $X = n$ se, e somente se, as primeiras $(n-1)$ válvulas forem negativas e a n-ésima válvula for positiva. Se aceitarmos que a condição de uma válvula não influencie a condição de outra, poderemos escrever

$$p(n) = P(X = n) = \left(\frac{1}{4}\right)^{n-1}\left(\frac{3}{4}\right), \qquad n = 1, 2, \ldots.$$

Para verificarmos que esses valores de $p(n)$ satisfazem à Eq. (4.3) observaremos que

$$\sum_{n=1}^{\infty} p(n) = \frac{3}{4}\left(1 + \frac{1}{4} + \frac{1}{16} + \cdots\right)$$

$$= \frac{3}{4}\,\frac{1}{1 - \frac{1}{4}} = 1.$$

Comentário: Estamos empregando aqui o resultado de que a *série geométrica* $1 + r + r^2 + \ldots$ converge para $1/(1-r)$ sempre que $|r| < 1$. Este é um resultado que será mencionado muitas vezes. Suponha-se que desejemos calcular $P(A)$, onde A é definido como: {O experimento termina depois de um número par de repetições.} Empregando a Eq. (4.4), teremos:

$$P(A) = \sum_{n=1}^{\infty} p(2n) = \frac{3}{16} + \frac{3}{256} + \cdots$$

$$= \frac{3}{16}\left(1 + \frac{1}{16} + \cdots\right) = \frac{3}{16}\,\frac{1}{1 - \frac{1}{16}} = \frac{1}{5}.$$

4.3. A Distribuição Binomial

Nos próximos capítulos, estudaremos pormenorizadamente algumas variáveis discretas importantes. Agora estudaremos apenas uma delas e, em seguida, a empregaremos para ilustrar alguns conceitos importantes.

Exemplo 4.7. Suponha que peças saiam de uma linha de produção e sejam classificadas como defeituosas (D) ou como não-defeituosas (N), isto é, perfeitas. Admita que três dessas peças, da produção de um dia, sejam escolhidas ao acaso e classificadas de acordo com esse esquema. O espaço amostral para esse experimento, S, pode ser assim, apresentado:

$$S = \{DDD, DDN, DND, NDD, NND, NDN, DNN, NNN\}.$$

(Outra maneira de descrever S é como $S = S_1 \times S_2 \times S_3$, o produto cartesiano de S_1, S_2 e S_3, onde cada $S_i = \{D, N\}$.) Suponhamos que seja 0,2 a probabilidade de uma peça ser defeituosa e 0,8 a de ser não-defeituosa. Admitamos que essas probabilidades sejam as *mesmas* para cada peça, ao menos enquanto durar o nosso estudo. Finalmente, admita-se que a classificação de qualquer peça em particular, seja independente da classificação de qualquer outra peça. Empregando essas suposições, segue-se que as probabilidades associadas aos vários resultados do espaço amostral S, como se explicou acima, são:

$$(0{,}2)^3, (0{,}8)(0{,}2)^2, (0{,}8)(0{,}2)^2, (0{,}8)(0{,}2)^2, (0{,}2)(0{,}8)^2, (0{,}2)(0{,}8)^2, (0{,}2)(0{,}8)^2, (0{,}8)^3.$$

Geralmente, nosso interesse não está dirigido para os resultados individuais de S. Ao contrário, desejamos tão-somente conhecer *quantas* peças defeituosas seriam encontradas (não interessando a ordem em que tenham ocorrido). Isto é, desejamos estudar a variável aleatória X, a qual atribui a cada resultado $s \in S$ o número de peças defeituosas encontradas em s. Conseqüentemente, o conjunto dos valores possíveis de X é $\{0, 1, 2, 3\}$.

Poderemos obter a distribuição de probabilidade de X, $p(x_i) = P(X = x_i)$, da seguinte maneira:

$X = 0$ se, e somente se, ocorrer NNN;

$X = 1$ se, e somente se, ocorrer DNN, NDN, ou NND;

$X = 2$ se, e somente se, ocorrer DDN, DND, ou NDD;

$X = 3$ se, e somente se, ocorrer DDD.

(Note-se que $\{NNN\}$ é equivalente a $\{X = 0\}$ etc.) Então,

$$p(0) = P(X = 0) = (0{,}8)^3 \qquad p(1) = P(X = 1) = 3(0{,}2)(0{,}8)^2,$$
$$p(2) = P(X = 2) = 3(0{,}2)^2(0{,}8), \quad p(3) = P(X = 3) = (0{,}2)^3.$$

Observe que a soma dessas probabilidades é igual a 1, porque a soma pode ser escrita como igual a $(0{,}8 + 0{,}2)^3$.

Comentário: A explicação dada ilustra como as probabilidades em um contradomínio R_X (neste caso $\{0, 1, 2, 3\}$) são *induzidas* pelas probabilidades definidas sobre o espaço amostral S. Porque a hipótese de que os oito resultados de

$$S = \{DDD, DDN, DND, NDD, NND, NDN, DNN, NNN\}$$

tenham as probabilidades dadas no Ex. 4.7, *determinou* o valor de $p(x)$ para todo $x \in R_X$.

Vamos agora generalizar as noções introduzidas no ex. anterior.

Definição: Consideremos um experimento $\mathcal{E}$ e seja A algum evento associado a $\mathcal{E}$. Admita-se que $P(A) = p$ e conseqüentemente $P(\bar{A}) = 1 - p$. Considerem-se n repetições de $\mathcal{E}$. Daí, o espaço amostral será formado por todas as seqüências possíveis $\{a_1, a_2, \ldots, a_n\}$, onde cada a_i é ou A ou $\bar{A}$, dependendo de que tenha ocorrido A ou $\bar{A}$ na i-ésima repetição de $\mathcal{E}$. (Existem 2^n dessas seqüências.) Além disso, suponha-se que $P(A) = p$ permaneça a mesma para todas as repetições. A variável aleatória X será assim definida: X = número de vezes que o evento A tenha ocorrido. Denominaremos X de variável aleatória *binomial*, com parâmetros n e p. Seus valores possíveis são evidentemente $0, 1, 2, \ldots, n$. (De maneira equivalente, diremos que X tem uma *distribuição binomial*.)

As repetições individuais de $\mathcal{E}$ serão denominadas *Provas de Bernouilli*.

***Teorema* 4.1. Seja X uma variável binomial, baseada em n repetições. Então,**

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \qquad k = 0, 1, \ldots, n. \tag{4.5}$$

Demonstração: Considere-se um particular elemento do espaço amostral de $\mathcal{E}$ satisfazendo à condição $X = k$. Um resultado como esse poderia surgir, por exemplo, se nas primeiras k repetições de $\mathcal{E}$ ocorresse A, enquanto nas últimas $n - k$ repetições ocorresse $\bar{A}$, isto é,

$$\underbrace{AAA\cdots A}_{k}\ \underbrace{\bar{A}\bar{A}\bar{A}\cdots\bar{A}}_{n-k}.$$

Como todas as repetições são independentes, a probabilidade desta seqüência particular seria $p^k(1 - p)^{n-k}$, mas exatamente essa mesma probabilidade seria associada a qualquer outro resultado para o qual $X = k$. O número total de tais resultados é igual a $\binom{n}{k}$, porque deveremos escolher exatamente k posições (dentre n) para o evento A. Ora, isso dá o resultado acima, porque esses $\binom{n}{k}$ resultados são todos mutuamente excludentes.

Comentários: (a) Para verificar nosso resultado, observemos que empregando o teorema binomial temos

$$\sum_{k=0}^{n} P(X = k) = \sum_{k=0}^{n} \binom{n}{k} p^k(1-p)^{n-k} = [p + (1-p)]^n = 1^n = 1,$$

como era de se esperar. Como as probabilidades $\binom{n}{k} p^k(1-p)^{n-k}$ são obtidas pelo desenvolvimento da expressão binomial $[p + (1-p)]^n$, ela recebe a denominação de distribuição binomial.

(b) Sempre que realizarmos repetições independentes de um experimento e estivermos interessados somente em uma dicotomia — defeituoso *ou* não-defeituoso (perfeito); dureza acima *ou* abaixo de certo padrão; nível de ruído em um sistema de comunicações acima *ou* abaixo de um limiar preestabelecido — estaremos virtualmente tratando com um espaço amostral no qual podemos definir uma variável aleatória binomial. Enquanto as condições da experimentação permaneçam suficientemente estáveis, de modo que a probabilidade de algum atributo, digamos A, permaneça constante, poderemos empregar o modelo acima.

(c) Se n for pequeno, os termos individuais da distribuição binomial serão relativamente fáceis de calcular. Contudo, se n for relativamente grande, os cálculos se tornam bastante incômodos. Felizmente, foram preparadas tábuas de probabilidades binomiais; existem várias dessas tábuas. (Veja o Apêndice.)

Exemplo 4.8. Suponha-se que uma válvula eletrônica, instalada em determinado circuito, tenha probabilidade 0,2 de funcionar mais do que 500 horas. Se ensaiarmos 20 válvulas, qual será a probabilidade de que delas, exatamente k, funcionem mais que 500 horas, $k = 0, 1, 2, \ldots, 20$?

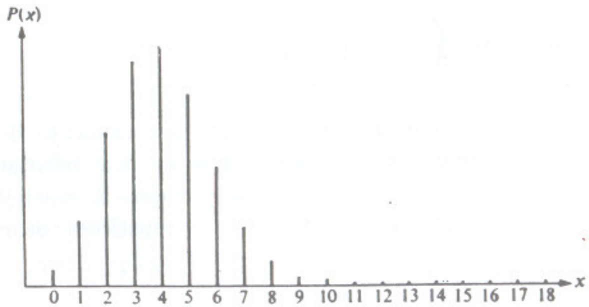


Fig. 4.6

Se X for o número de válvulas que funcionem mais de 500 horas, admitiremos que X tenha uma distribuição binomial. Então, $P(X = k) = \binom{20}{k}(0,2)^k(0,8)^{20-k}$. Os valores podem ser lidos na Tab. 4.1.

Se marcarmos os valores dessa distribuição, obteremos o gráfico apresentado na Fig. 4.6. A configuração que observamos aqui é bastante geral. As probabilidades binomiais crescem monotonicamente, até que atingem um valor máximo e, depois, decrescem monotonicamente. (Veja o Probl. 4.8.)

Tab. 4.1

$P(X = 0) = 0{,}012$	$P(X = 4) = 0{,}218$	$P(X = 8) = 0{,}022$
$P(X = 1) = 0{,}058$	$P(X = 5) = 0{,}175$	$P(X = 9) = 0{,}007$
$P(X = 2) = 0{,}137$	$P(X = 6) = 0{,}109$	$P(X = 10) = 0{,}002$
$P(X = 3) = 0{,}205$	$P(X = 7) = 0{,}055$	$P(X = k) = 0^+$ para $k \geq 11$

(As probabilidades restantes são menores do que 0,001.)

Exemplo 4.9. Ao operar determinada máquina, existe alguma probabilidade de que o operador da máquina cometa um erro. Pode-se admitir, razoavelmente, que o operador aprenda, no sentido de que decresça a probabilidade de cometer um erro, se ele usar repetidamente a máquina. Suponha que o operador faça n tentativas e que as n repetições sejam estatisticamente independentes. Suponhamos, especificamente, que P (um erro ser cometido na i-ésima repetição) $= 1/(i+1)$, $i = 1, 2, \ldots, n$. Admitamos que se pretendam 4 tentativas (isto é, $n = 4$) e definamos a variável aleatória X como o número de operações da máquina, executadas sem erro. Note-se que X não tem distribuição binomial, porque a probabilidade de "sucesso" não é constante.

Para calcular a probabilidade de que $X = 3$, por exemplo, procede-se do seguinte modo: $X = 3$ se, e somente se, houver exatamente uma tentativa mal sucedida. Isto pode ocorrer na primeira, segunda, terceira ou quarta tentativas. Portanto,

$$P(X = 3) = \frac{1}{2}\,\frac{2}{3}\,\frac{3}{4}\,\frac{4}{5} + \frac{1}{2}\,\frac{1}{3}\,\frac{3}{4}\,\frac{4}{5} + \frac{1}{2}\,\frac{2}{3}\,\frac{1}{4}\,\frac{4}{5} + \frac{1}{2}\,\frac{2}{3}\,\frac{3}{4}\,\frac{1}{5} = \frac{5}{12}.$$

Exemplo 4.10. Considere-se uma situação semelhante àquela apresentada no Ex. 4.9. Agora, admitiremos que exista uma probabilidade constante p_1 de não cometer um erro na máquina, durante cada uma das n_1 tentativas, e uma probabilidade constante $p_2 \leq p_1$ de não cometer um erro em cada uma das n_2 repetições subseqüentes. Seja X o número de operações bem sucedidas da máquina durante as $n = n_1 + n_2$ tentativas independentes. Vamos procurar a expressão geral de $P(X = k)$. Pelo mesmo motivo dado no exemplo precedente, X não tem distribuição binomial. Para obter $P(X = k)$, procede-se da seguinte maneira:

Sejam Y_1 o número de operações corretas durante as primeiras n_1 tentativas, e Y_2 o número de operações corretas durante as n_2 tentativas subseqüentes. Portanto, Y_1 e Y_2 são variáveis aleatórias independentes e $X = Y_1 + Y_2$. Assim, $X = k$ se, e somente se,

$Y_1 = r$ e $Y_2 = k - r$, para qualquer inteiro r que satisfaça às condições $0 \leq r \leq n_1$ e $0 \leq k - r \leq n_2$.

As restrições acima, sobre r, são equivalentes a $0 \leq r \leq n_1$ e $k - n_2 \leq r \leq k$. Combinando-as, poderemos escrever máx. $(0, k - n_2) \leq r \leq$ mín. (k, n_1). Portanto, teremos

$$P(X = k) = $$

$$= \sum_{r = \text{máx.}(0, k - n_2)}^{r = \text{mín.}(k, n_1)} \binom{n_1}{r} p_1^r (1 - p_1)^{n_1 - r} \binom{n_2}{k - r} p_2^{k-r} (1 - p_2)^{n_2 - (k - r)}.$$

Com nossa convenção usual de que $\binom{a}{b} = 0$ sempre que $b > a$ ou $b < 0$, poderemos escrever a probabilidade acima como

$$P(X = k) = \sum_{r=0}^{n_1} \binom{n_1}{r} p_1^r (1 - p_1)^{n_1 - r} \binom{n_2}{k - r} p_2^{k-r} (1 - p_2)^{n_2 - k + r} \tag{4.6}$$

Por exemplo, se $p_1 = 0{,}2$, $p_2 = 0{,}1$, $n_1 = n_2 = 10$ e $k = 2$, a probabilidade acima fica, depois de um cálculo direto:

$$P(X = 2) = \sum_{r=0}^{2} \binom{10}{r} (0{,}2)^r (0{,}8)^{10-r} \binom{10}{2 - r} (0{,}1)^{2-r} (0{,}9)^{8+r} = 0{,}27.$$

Comentário: Suponha-se que $p_1 = p_2$. Neste caso, a Eq. (4.6) se reduz a $\binom{n}{k} p_1^k (1 - p_1)^{n-k}$, porque agora a variável aleatória X *tem* uma distribuição binomial. Para verificar que é assim, note-se que poderemos escrever, (desde que $n_1 + n_2 = n$):

$$P(X = k) = p_1^k (1 - p_1)^{n-k} \sum_{r=0}^{n_1} \binom{n_1}{r} \binom{n_2}{k - r}$$

Para verificar que a soma acima é igual a $\binom{n}{k}$, basta comparar os coeficientes das potências de x^k em ambos os membros da identidade $(1 + x)^{n_1} (1 + x)^{n_2} = (1 + x)^{n_1 + n_2}$.

4.4. Variáveis Aleatórias Contínuas

Suponha-se que o contradomínio de X seja formado por um número finito muito grande de valores, digamos todos os valores x no intervalo $0 \leq x \leq 1$, da forma: 0; 0,01; 0,02; ...; 0,98; 0,99; 1,00. A cada um desses valores está associado um número não-negativo $p(x_i) = P(X = x_i)$, $i = 1, 2, \ldots$, cuja soma é igual a 1. Esta operação está representada geometricamente na Fig. 4.7.

Já salientamos anteriormente que poderia ser matematicamente mais fácil idealizar a apresentação probabilística de X, pela suposição de que X pudesse tomar *todos* os valores possíveis, $0 \leq x \leq 1$. Se fizermos isso, que acontecerá às probabilidades no ponto $p(x_i)$? Como os valores possíveis de X não são numeráveis, não podemos realmente falar do i-ésimo valor de X, e, por isso, $p(x_i)$ se torna sem sentido. O que faremos é substituir a função p definida somente para $x_1, x_2, \ldots$ por uma função f definida (neste contexto) para *todos* os valores de x, $0 \leq x \leq 1$. As propriedades da Eq. (4.3) serão substituídas por $f(x) \geq 0$ e $\int_0^1 f(x) dx = 1$. Vamos proceder formalmente como se segue.

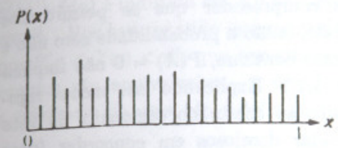


Fig. 4.7

Definição: Diz-se que X é uma *variável aleatória contínua*, se existir uma função f, denominada *função densidade de probabilidade* (*fdp*) de X que satisfaça às seguintes condições:

(*a*) $f(x) \geqslant 0$ para todo x,

(*b*) $\int_{-\infty}^{+\infty} f(x)\, dx = 1$, (4.7)

(*c*) para quaisquer a, b, com $-\infty < a < b < +\infty$, teremos $P(a \leqslant X \leqslant b) = \int_a^b f(x)\, dx$. (4.8)

Comentários: (*a*) Estaremos essencialmente dizendo que X é uma variável aleatória contínua, se X puder tomar todos os valores em algum intervalo (c, d), onde c e d podem ser $-\infty$ e $+\infty$, respectivamente. A existência estipulada de uma fdp constitui um artifício matemático, que possui considerável apelo intuitivo e torna nossos cálculos mais simples. Em relação a isso, também devemos salientar que, quando supomos que X seja uma variável aleatória contínua, estamos tratando com uma descrição *idealizada* de X.

(*b*) $P(c < X < d)$ representa a área sob a curva no gráfico da Fig. 4.8, da fdp f, entre $x = c$ e $x = d$.

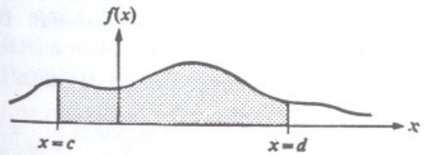


Fig. 4.8

(*c*) Constitui uma conseqüência da descrição probabilística de X, acima, que, para *qualquer* valor especificado de X, digamos x_0, teremos $P(X = x_0) = 0$, porque $P(X = x_0) = \int_{x_0}^{x_0} f(x)\,dx = 0$. Este resultado pode parecer muito contrário à nossa intuição. Contudo, devemos compreender que se permitirmos que X tome *todos* os valores em algum intervalo, então a probabilidade zero não é equivalente à impossibilidade. Por isso, no caso contínuo, $P(A) = 0$ não implica ser $A = \emptyset$, o conjunto vazio. (Veja o Teor. 1.1.) Explicando isso menos rigorosamente, considere-se a escolha de um ponto ao acaso, no segmento de reta $\{x \mid 0 \leq x \leq 2\}$. Muito embora possamos estar desejosos em concordar (para objetivos matemáticos) que cada ponto imaginável no segmento possa ser resultado de nosso experimento, ficaríamos completamente surpreendidos quanto a isso, se de fato escolhessemos precisamente o ponto médio do segmento ou qualquer outro ponto *especificado*. Quando expressamos isto em linguagem matemática rigorosa, dizemos que o evento tem "probabilidade zero". Tendo em vista essas observações, as seguintes probabilidades serão todas *iguais*, se X for uma variável aleatória contínua:

$$P(c \leq X \leq d), \quad P(c \leq X < d), \quad P(c < X \leq d), \quad \text{e} \quad P(c < X < d).$$

(*d*) Apesar de não verificarmos aqui os detalhes, pode-se mostrar que essa atribuição de probabilidades a eventos em R_X satisfaz aos axiomas básicos da probabilidade [Eq. (1.3)], onde poderemos tomar $\{x \mid -\infty < x < +\infty\}$ como nosso espaço amostral.

(*e*) Se uma função f^* satisfizer às condições $f^*(x) \geq 0$ para todo x, e $\int_{-\infty}^{+\infty} f^*(x)\,dx = K$, onde K é um número real positivo (não necessariamente igual a 1), então f^* não satisfaz a todas as condições para ser uma fdp. No entanto, poderemos facilmente definir uma nova função, digamos f, em termos de f^*, assim:

$$f(x) = \frac{f^*(x)}{K} \quad \text{para todo } x.$$

Em conseqüência, f satisfará a todas as condições de uma fdp.

(*f*) Se X tomar valores somente em algum intervalo finito $[a, b]$, poderemos simplesmente pôr $f(x) = 0$ para todo $x \notin [a, b]$. Em conseqüência, a fdp ficará definida para *todos* os valores reais de x, e poderemos exigir que $\int_{-\infty}^{+\infty} f(x)\,dx = 1$. Sempre que a fdp for especificada somente para determinados valores de x, deveremos supor que seja zero em todos os demais.

(*g*) $f(x)$ não representa a probabilidade de coisa alguma! Anteriormente já salientamos que, por exemplo, $P(X = 2) = 0$ e, conseqüentemente, $f(2)$ certamente não representa essa probabilidade. Somente quando a função for integrada entre dois limites, ela produzirá uma probabilidade. Poderemos, contudo, dar uma interpretação de $f(x)\Delta x$, da seguinte maneira: Do teorema do valor médio, em Cálculo, tem-se que

$$P(x \leq X \leq x + \Delta x) = \int_x^{x+\Delta x} f(s)\,ds = \Delta x f(\xi), \qquad x \leq \xi \leq x + \Delta x.$$

Se Δx for pequeno, $f(x)\Delta x$ será *aproximadamente* igual a $P(x \leq X \leq x + \Delta x)$. (Se f for contínua à direita, esta aproximação se tornará mais exata quando $\Delta x \to 0$.)

(*h*) Devemos novamente salientar que a distribuição de probabilidade (neste caso a fdp) é induzida em R_X pela probabilidade subjacente associada com eventos em S. Por isso, quando escrevemos $P(c < X < d)$, queremos significar como sempre $P[c < X(s) < d]$, que por sua vez é igual a $P[s \mid c < X(s) < d]$, já que esses eventos são equivalentes. A definição anterior, Eq. (4.8), estipula essencialmente a existência de uma fdp f definida sobre R_X tal que

$$P[s \mid c < X(s) < d] = \int_c^d f(x)\,dx.$$

Novamente suprimiremos a natureza funcional de X e, por isso, trataremos somente com R_X e a fdp f.

(*i*) No caso contínuo, também poderemos considerar a seguinte *analogia com a Mecânica*: Suponha-se que temos uma massa total de uma unidade continuamente distribuída sobre o intervalo $a \leq x \leq b$. Nesse caso, $f(x)$ representa a densidade de massa no ponto x e $\int_c^d f(x)\,dx$ representa a massa total contida no intervalo $c \leq x \leq d$.

Exemplo 4.11. A existência de uma fdp foi admitida na exposição de uma variável aleatória contínua. Vamos considerar um exemplo simples, no qual poderemos facilmente determinar a fdp, fazendo uma suposição apropriada sobre o comportamento probabilístico da variável aleatória. Suponhamos que um ponto seja escolhido no intervalo (0,1). Representemos por X a variável aleatória cujo valor seja a abscissa x do ponto escolhido.

Supor: Se I for qualquer intervalo em (0,1), então Prob $[X \in I]$ será diretamente proporcional ao cumprimento de I, digamos $L(I)$. Isto é, Prob $[X \in I] = kL(I)$, onde k é a constante de proporcionalidade. (É fácil verificar, tomando-se $I = (0,1)$ e observando-se que $L[(0,1)] = 1$ e Prob $[X \in (0,1)] = 1$, que $k = 1$.)

Obviamente, X torna todos os valores em (0,1). Qual é sua fdp? Assim, podemos encontrar uma função f tal que

$$P(a < X < b) = \int_a^b f(x)\,dx?$$

Note que, se $a < b < 0$ ou $1 < a < b$, $P(a < X < b) = 0$ e, por isso, $f(x) = 0$. Se $0 < a < b < 1$, $P(a < X < b) = b - a$ e, conseqüentemente, $f(x) = 1$. Portanto, encontramos

$$f(x) = \begin{cases} 1, & 0 < x < 1 \\ 0, & \text{para quaisquer outros valores.} \end{cases}$$

Exemplo 4.12. Suponhamos que a variável aleatória X seja contínua. (Veja a Fig. 4.9.) Seja a fdp f dada por

$$f(x) = 2x, \quad 0 < x < 1,$$
$$= 0, \quad \text{para quaisquer outros valores.}$$

Evidentemente, $f(x) \geq 0$ e $\int_{-\infty}^{+\infty} f(x)dx = \int_0^1 2x\,dx = 1$. Para calcular $P(X \leqslant 1/2)$, deve-se apenas calcular a integral $\int_0^{1/2} (2x)\,dx = 1/4$.

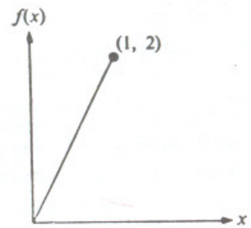


Fig. 4.9

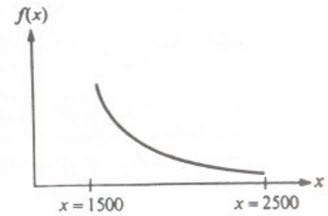


Fig. 4.10

O conceito de probabilidade condicionada, explicado no Cap. 3, pode ser significativamente aplicado a variáveis aleatórias. Assim, no exemplo acima, podemos calcular $P\left(X \leq \frac{1}{2} \middle| \frac{1}{3} \leq X \leq \frac{2}{3}\right)$. Aplicando-se diretamente a definição de probabilidade condicionada, teremos

$$P\left(X \leq \frac{1}{2} \middle| \frac{1}{3} \leq X \leq \frac{2}{3}\right) = \frac{P\left(\frac{1}{3} \leq X \leq \frac{1}{2}\right)}{P\left(\frac{1}{3} \leq X \leq \frac{2}{3}\right)}$$
$$= \frac{\int_{1/3}^{1/2} 2x\,dx}{\int_{1/3}^{2/3} 2x\,dx} = \frac{5/36}{1/3} = \frac{5}{12}.$$

Exemplo 4.13. Seja X a duração da vida (em horas) de um certo tipo de lâmpada. Admitindo que X seja uma variável aleatória contínua, suponha-se que a fdp f de X seja dada por

$$f(x) = a/x^3, \quad 1.500 \leq x \leq 2.500,$$
$$= 0, \text{ para quaisquer outros valores.}$$

(Isto é, está se atribuindo probabilidade zero aos eventos $\{X < 1.500\}$ e $\{X > 2.500\}$.) Para calcular a constante a, recorre-se à condição $\int_{-\infty}^{\infty} f(x)\,dx = 1$, que neste caso se torna $\int_{1.500}^{2.500} (a/x^3)\,dx = 1$. Daí se obtém $a = 7.031.250$. O gráfico de f está apresentado na Fig. 4.10.

Em capítulo posterior estudaremos, pormenorizadamente, muitas variáveis aleatórias importantes, tanto discretas como contínuas.

Sabemos, de nosso emprego de modelos determinísticos, que certas funções gozam de papel mais importante que outras. Por exemplo, as funções linear, quadrática, exponencial e trigonométrica têm papel vital na explicação de modelos determinísticos. Ao desenvolver modelos não-determinísticos (isto é, probabilísticos) verificaremos que certas variáveis aleatórias são de notável importância.

4.5. Função de Distribuição Acumulada

Vamos introduzir outro importante conceito geral, neste capítulo.

Definição. Seja X uma variável aleatória, discreta ou contínua. Define-se a função F como a *função de distribuição acumulada* da variável aleatória X (abreviadamente indicada fd) como $F(x) = P(X \leq x)$.

Teorema 4.2. (*a*) Se X for uma variável aleatória discreta

$$F(x) = \sum_j p(x_j), \tag{4.9}$$

onde o somatório é estendido a todos os índices j que satisfaçam à condição $x_j \leq x$.

(*b*) Se X for uma variável aleatória contínua com fdp f,

$$F(x) = \int_{-\infty}^{x} f(s)ds. \tag{4.10}$$

Demonstração. Ambos os resultados decorrem imediatamente da definição.

Exemplo 4.14. Suponhamos que a variável aleatória X tome os três valores 0, 1 e 2, com probabilidades 1/3, 1/6 e 1/2, respectivamente. Então,

$$F(x) = 0 \quad \text{se } x < 0,$$
$$= \frac{1}{3} \quad \text{se } 0 \leq x < 1,$$
$$= \frac{1}{2} \quad \text{se } 1 \leq x < 2,$$
$$= 1 \quad \text{se } x \geq 2.$$

(Observe-se que é muito importante indicar a inclusão ou a exclusão dos limites, na descrição dos diversos intervalos.) O gráfico de F está apresentado na Fig. 4.11.

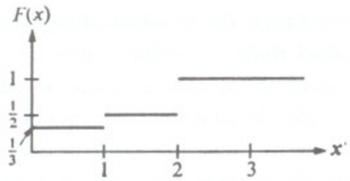


Fig. 4.11

Exemplo 4.15. Suponhamos que X seja uma variável contínua com fdp

$$f(x) = 2x, \quad 0 < x < 1,$$
$$= 0, \text{ para quaisquer outros valores.}$$

Portanto, a fdp de F é dada por

$$F(x) = 0 \text{ se } x \leq 0,$$
$$= \int_0^x 2s\, ds = x^2$$
$$\text{se } 0 < x \leq 1,$$
$$= 1 \text{ se } x > 1.$$

O gráfico está apresentado na Fig. 4.12.

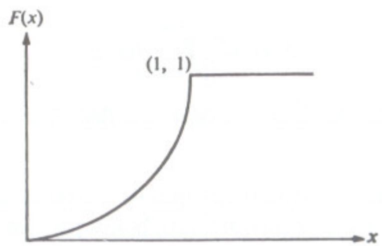


Fig. 4.12

Os gráficos apresentados nas Figs. 4.11 e 4.12 para as fd são, em cada caso, bastante típicos, no seguinte sentido:

(*a*) Se X for uma variável aleatória discreta, com um número finito de valores possíveis, o gráfico da fd será constituído por segmentos de reta horizontais (nesse caso, a fd se denomina *função em degraus*). A função F é contínua, exceto nos valores possíveis de X: $x_1, \ldots, x_n, \ldots$ No valor x_j o gráfico apresenta um "salto" de magnitude $p(x_j) = P(X = x_j)$.

(*b*) Se X for uma variável aleatória contínua, F será uma função contínua para todo x.

(*c*) A fd F é definida para *todos* os valores de x, o que é um motivo importante para considerá-la.

Existem duas outras importantes propriedades da fd, que resumiremos no teorema seguinte:

Teorema 4.3. (*a*) A função F é não-decrescente. Isto é, se $x_1 \leq x_2$, teremos $F(x_1) \leq F(x_2)$.

(*b*) $\lim_{x \to -\infty} F(x) = 0$ e $\lim_{x \to \infty} F(x) = 1$. [Freqüentemente, escrevemos isto como $F(-\infty) = 0$ e $F(\infty) = 1$.]

Demonstração: (*a*) Definamos os eventos A e B, assim: $A = \{X \leq x_1\}$, $B = \{X \leq x_2\}$. Portanto, como $x_1 \leq x_2$, teremos $A \subset B$ e, pelo Teor. 1.5, $P(A) \leq P(B)$, que é o resultado desejado.

(*b*) No caso contínuo, teremos:

$$F(-\infty) = \lim_{x \to -\infty} \int_{-\infty}^{x} f(s)\, ds = 0,$$

$$F(\infty) = \lim_{x \to \infty} \int_{-\infty}^{x} f(s)\, ds = 1.$$

No caso discreto, o raciocínio é análogo.

A função de distribuição (acumulada) é importante por muitas razões. Isto é particularmente verdadeiro quando tratarmos com uma variável aleatória contínua, porque nesse caso não poderemos estudar o comportamento probabilístico de X através do cálculo de $P(X = x)$. Aquela probabilidade é sempre igual a zero no caso contínuo. Contudo, poderemos indagar de $P(X \leqslant x)$ e, como demonstra o teorema seguinte, obter a fdp de X.

Teorema 4.4. (*a*) Seja F a fd de uma variável aleatória contínua, com fdp f. Então,

$$f(x) = \frac{d}{dx} F(x),$$

para todo x no qual F seja derivável.

(*b*) Seja X uma variável aleatória discreta, com valores possíveis $x_1, x_2, \ldots$, e suponha-se que esses valores tenham sido indexados de modo que $x_1 < x_2 < \ldots$ Seja F a fd de X. Então,

$$p(x_j) = P(X = x_j) = F(x_j) - F(x_{j-1}). \tag{4.12}$$

Demonstração: (*a*) $F(x) = P(X \leq x) = \int_{-\infty}^{x} f(s)\, ds$. Por isso, aplicando-se o teorema fundamental do Cálculo, obteremos $F'(x) = f(x)$.

(b) Como admitimos $x_1 < x_2 < \ldots$, teremos

$$F(x_j) = P(X = x_j \cup X = x_{j-1} \cup \cdots \cup X = x_1)$$
$$= p(j) + p(j-1) + \cdots + p(1).$$

E

$$F(x_{j-1}) = P(X = x_{j-1} \cup X = x_{j-2} \cup \cdots \cup X = x_1)$$
$$= p(j-1) + p(j-2) + \cdots + p(1).$$

Portanto,

$$F(x_j) - F(x_{j-1}) = P(X = x_j) = p(x_j).$$

Comentário: Vamos resumidamente reconsiderar a parte (*a*) do Teorema 4.4. Recordemos a definição de derivada de uma função F:

$$F'(x) = \lim_{h \to 0} \frac{F(x+h) - F(x)}{h}$$

$$= \lim_{h \to 0^+} \frac{P(X \leqslant x+h) - P(X \leqslant x)}{h}$$

$$= \lim_{h \to 0^+} \frac{1}{h}\,[P(x < X \leqslant x+h)].$$

Portanto, se h for pequeno e positivo,

$$F'(x) = f(x) \cong \frac{P(x < X \leqslant x+h)}{h}.$$

Assim, $f(x)$ é aproximadamente igual à "quantidade de probabilidade no intervalo $(x, x+h)$ pelo comprimento h". Daí o nome ***função densidade de probabilidade***.

Exemplo 4.16. Suponha-se que uma variável aleatória contínua tenha a fd F dada por

$$F(x) = 0, \quad x \leq 0,$$
$$= 1 - e^{-x}, \quad x > 0.$$

Nesse caso, $F'(x) = e^{-x}$ para $x > 0$, e, por isso, a fdp será dada por

$$f(x) = e^{-x}, \quad x \geq 0,$$
$$= 0, \text{ para quaisquer outros valores.}$$

Comentário: É oportuno dizer uma palavra final sobre a terminologia. Esta terminologia, muito embora ainda não uniforme, tornou-se bastante padronizada. Quando falamos da *distribuição de probabilidade* de uma variável aleatória X, nos referimos à sua fdp se X for contínua, ou à sua função de probabilidade no ponto, p, definida para $x_1, x_2, \ldots$ se X for discreta. Quando falamos da função de distribuição acumulada, ou algumas vezes apenas *função de distribuição* (ou *função de repartição*), queremos sempre nos referir a F, onde $F(x) = P(X \leqslant x)$.

4.6. Distribuições Mistas

Restringimos nossa explanação tão-somente a variáveis aleatórias que sejam discretas ou contínuas. Tais variáveis são certamente as mais importantes nas aplicações. Contudo, há situações em que poderemos encontrar um tipo *misto*: a variável aleatória X pode tomar alguns valores diferentes, digamos $x_1, \ldots x_n$, com probabilidade não-nula, e também tomar todos os valores em algum intervalo, digamos $a \leq x \leq b$. A distribuição de probabilidade de tal variável aleatória seria obtida pela combinação das idéias já examinadas na descrição de variáveis aleatórias discretas e de contínuas, como se verá a seguir. A cada valor x_i associa-se um número $p(x_i)$ tal que $p(x_i) \geq 0$ para todo i, e tal que $\sum_{i=1}^{n} p(x_i) = p < 1$. Em seguida, define-se uma função f, satisfazendo a $f(x) \geq 0$, $\int_a^b f(x)\,dx = 1 - p$. Para todo a, b, com $-\infty < a < b < +\infty$, $P(a \leqslant X \leqslant b) = \int_a^b f(x)\,dx + \sum_{\{i:\, a \leqslant x_i \leqslant b\}} p(x_i)$. Desta maneira, atenderemos à condição $P(S) = P(-\infty < X < \infty) = 1$.

Uma variável aleatória de tipo misto poderia surgir da maneira explicada a seguir. Suponha-se que estejamos ensaiando algum equipamento e façamos igual a X o tempo de funcionamento. Em muitos problemas, descreveremos X como uma variável aleatória contínua, com valores possíveis $x \geq 0$. No entanto, podem surgir situações nas quais exista uma probabilidade não-nula de que a peça não funcione de modo algum, isto é, falhe no momento $X = 0$. Nesse caso, desejaríamos modificar nosso modelo e atribuir uma probabilidade $p > 0$ ao resultado $X = 0$. Conseqüentemente, teríamos $P(X = 0) = p$ e $P(X > 0) = 1 - p$. Deste modo, p descreveria a distribuição de X no ponto 0, enquanto a fdp f descreveria a distribuição para valores de $X > 0$. (Veja a Fig. 4.13.)

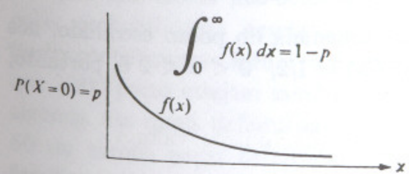


Fig. 4.13

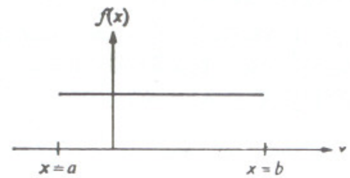


Fig. 4.14

4.7. Variáveis Aleatórias Uniformemente Distribuídas

Nos Caps. 8 e 9, estudaremos minuciosamente muitas variáveis aleatórias discretas e contínuas importantes. Já introduzimos a importante variável aleatória binomial. Vamos agora examinar, resumidamente, uma importante variável aleatória contínua.

Definição. Suponha-se que X seja uma variável aleatória contínua, que tome todos os valores no intervalo $[a, b]$, no qual a e b sejam ambos finitos. Se a fdp de X for dada por

$$f(x) = \frac{1}{b-a}, \quad a \leq x \leq b, \tag{4.13}$$
$$= 0, \text{ para quaisquer outros valores, diremos}$$

que X é *uniformemente distribuída* sobre o intervalo $[a, b]$. (Veja a Fig. 4.14.)

Comentários: (*a*) Uma variável aleatória uniformemente distribuída tem uma fdp que é *constante* sobre o intervalo de definição. A fim de satisfazer à condição $\int_{-\infty}^{+\infty} f(x)\,dx = 1$, essa constante deve ser igual ao inverso do comprimento do intervalo.

(*b*) Uma variável aleatória uniformemente distribuída representa o análogo contínuo dos resultados igualmente prováveis, no seguinte sentido. Para qualquer subintervalo $[c, d]$, onde $a \leq c < d \leq b$, $P(c \leq X \leq d)$ é a *mesma* para todos os subintervalos que tenham o mesmo comprimento. Isto é,

$$P(c \leq X \leq d) = \int_c^d f(x)\,dx = \frac{d-c}{b-a}$$

e, por isso, depende unicamente do comprimento do intervalo e não da posição desse intervalo.

(*c*) Agora podemos tornar mais precisa a noção intuitiva de *escolher ao acaso um ponto P*, em um intervalo $[a, b]$. Por isto simplesmente queremos dizer que a coordenada x do ponto escolhido, digamos X, é uniformemente distribuída sobre $[a, b]$.

Exemplo 4.17. Um ponto é escolhido ao acaso no segmento de reta $[0, 2]$. Qual será a probabilidade de que o ponto escolhido esteja entre 1 e 3/2?

Fazendo-se X representar a coordenada do ponto escolhido, nós temos que a fdp de X é dada por $f(x) = 1/2$, $0 < x < 2$ e, portanto, $P(1 \leq X \leq 3/2) = 1/4$.

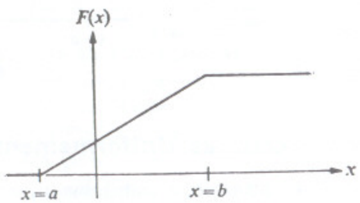


Fig. 4.15

Exemplo 4.18. A dureza H de uma peça de aço (avaliada na escala Rockwell) pode-se supor ser uma variável aleatória contínua uniformemente distribuída sobre o intervalo [50, 70], da escala B. Conseqüentemente,

$$f(h) = \frac{1}{20}, \quad 50 < h < 70,$$
$$= 0, \text{ para quaisquer outros valores.}$$

Exemplo 4.19. Vamos obter a expressão da fd de uma variável aleatória uniformemente distribuída.

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(s)ds$$
$$= 0 \quad \text{se } x < a,$$
$$= \frac{x-a}{b-a} \quad \text{se } a \leq x < b,$$
$$= 1 \quad \text{se } x \geq b.$$

O gráfico está apresentado na Fig. 4.15.

4.8 Uma Observação

Repetidamente temos salientado que em algum estágio de nosso desenvolvimento de um modelo probabilístico, algumas probabilidades devem ser atribuídas a resultados, com base ou em evidência experimental (como as freqüências relativas, por exemplo) ou em alguma outra consideração, como a experiência passada com o fenômeno que esteja em estudo. A seguinte questão pode ocorrer ao estudante: Por que nós não podemos obter todas as probabilidades em que estejamos interessados por tais meios não-dedutivos? A resposta é que muitos eventos cujas probabilidades desejamos conhecer são tão complicados que nosso conhecimento intuitivo é insuficiente. Por exemplo, suponhamos que 1000 peças estejam saindo diariamente de uma linha de produção, algumas das quais defeituosas. Desejamos saber a probabilidade de ter 50 ou menos peças defeituosas em certo dia. Mesmo que estejamos familiarizados com o comportamento geral do processo de produção, poderá ser difícil para nós associarmos uma medida quantitativa com o evento: 50 ou menos peças são defeituosas. No entanto, poderemos ser capazes de fazer a afirmação de que qualquer peça individual tenha probabilidade de 0,10 de ser defeituosa. (Assim, a experiência passada nos dá a informação de que cerca de 10 por cento das peças são defeituosas.) Além disso, poderemos estar inclinados a admitir que, individualmente, as peças sejam defeituosas ou perfeitas independentemente uma da outra. *Agora,* poderemos proceder dedutivamente e obter a probabilidade do evento em estudo. Assim, se X = número de peças defeituosas,

$$P(X \leqslant 50) = \sum_{k=0}^{50} \binom{1000}{k} (0{,}10)^k \, (0{,}90)^{1000-k}$$

O que se quer destacar aqui é que os vários métodos que nós deduzimos para calcular probabilidades (e outros que serão estudados subseqüentemente) são de enorme importância, porque com eles poderemos avaliar probabilidades associadas a eventos bastante complicados, as quais seriam difíceis de obter por meios intuitivos ou empíricos.

Problemas

4.1. Sabe-se que uma determinada moeda apresenta cara três vezes mais freqüentemente que coroa. Essa moeda é jogada três vezes. Seja X o número de caras que aparece. Estabeleça a distribuição de probabilidade de X e também a fd. Faça um esboço do gráfico de ambas.

4.2. De um lote que contém 25 peças, das quais 5 são defeituosas, são escolhidas 4 ao acaso. Seja X o número de defeituosas encontradas. Estabeleça a distribuição de probabilidade de X, quando:

(*a*) As peças forem escolhidas com reposição.

(*b*) As peças forem escolhidas sem reposição.

4.3. Suponha que a variável aleatória X tenha os valores possíveis 1, 2, 3, ..., e $P(X = j) = 1/2^j$, $j = 1, 2, \ldots$

(*a*) Calcule $P(X$ ser par).

(*b*) Calcule $P(X \geq 5)$.

(*c*) Calcule $P(X$ ser divisível por 3).

4.4. Considere uma variável aleatória X com resultados possíveis: 0, 1, 2, ... Suponha que $P(X = j) = (1 - a)a^j$, $j = 0, 1, 2, \ldots$

(*a*) Para que valores de a o modêlo acima tem sentido?

(*b*) Verifique que essa expressão representa uma legítima distribuição de probabilidade.

(*c*) Mostre que, para quaisquer dois inteiros positivos s e t,

$$P(X > s + t \mid X > s) = P(X \geq t).$$

4.5. Suponha que a máquina 1 produza (por dia) o dobro das peças que são produzidas pela máquina 2. No entanto, 4% das peças fabricadas pela máquina 1 tendem a ser defeituosas, enquanto somente cerca de 2% de defeituosas produz a máquina 2. Admita que a produção diária das duas máquinas seja misturada. Uma amostra aleatória de 10 peças é extraída da produção total. Qual será a probabilidade de que essa amostra contenha 2 peças defeituosas?

4.6. Foguetes são lançados até que o primeiro lançamento bem sucedido tenha ocorrido. Se isso não ocorrer até 5 tentativas, o experimento é suspenso e o equipamento inspecionado. Admita que exista uma probabilidade constante de 0,8 de haver um lançamento bem sucedido e que os sucessivos lançamentos sejam independentes. Suponha que o custo do primeiro lançamento seja K dólares, enquanto os lançamentos subseqüentes custam $K/3$ dólares. Sempre que ocorre um lançamento bem sucedido, uma certa quantidade de informação é obtida, a qual pode ser expressa como um ganho financeiro de C dólares. Sendo T o custo líquido desse experimento, estabeleça a distribuição de probabilidade de T.

4.7. Calcule $P(X = 5)$, onde X é a variável aleatória definida no Ex. 4.10. suponha que $n_1 = 10$, $n_2 = 15$, $p_1 = 0{,}3$ e $p_2 = 0{,}2$.

4.8. (*Propriedades das Probabilidades Binomiais.*) Na explanação do Ex. 4.8, um padrão geral para as probabilidades binomiais $\binom{n}{k}p^k(1-p)^{n-k}$ foi sugerido. Vamos denotar essas probabilidades por $p_n(k)$.

(*a*) Mostre que, para $0 \leq k < n$, temos

$$p_n(k+1)/p_n(k) = [(n-k)/(k+1)]\,[p/(1-p)].$$

(*b*) Empregando (*a*), mostre que

(i) $p_n(k+1) > p_n(k)$ se $k < np - (1-p)$,
(ii) $p_n(k+1) = p_n(k)$ se $k = np - (1-p)$,
(iii) $p_n(k+1) < p_n(k)$ se $k > np - (1-p)$.

(*c*) Mostre que se $np - (1-p)$ for um inteiro, $p_n(k)$ toma seu valor máximo para dois valores de k, a saber, $k_0 = np - (1-p)$ e $k_0' = np - (1-p) + 1$.

(*d*) Mostre que se $np - (1-p)$ não for um inteiro, então $p_n(k)$ toma seu valor máximo quando k for igual ao menor inteiro maior que k_0.

(*e*) Mostre que se $np - (1-p) < 0$, $p_n(0) > p_n(1) > \ldots > p_n(n)$, enquanto se $np - (1-p) = 0$, $p_n(0) = p_n(1) > p_n(2) > \ldots > p_n(n)$.

4.9. A variável aleatória contínua X tem para fdp: $f(x) = x/2$, $0 \leq x \leq 2$. São feitas duas determinações independentes de X. Qual será a probabilidade de que ambas essas determinações sejam maiores do que 1? Se três determinações independentes forem feitas, qual a probabilidade de que exatamente duas delas sejam maiores do que 1?

4.10. Seja X a duração da vida de uma válvula eletrônica e admita-se que X possa ser representada por uma variável aleatória contínua, com fdp $f(x) = be^{-bx}$, $x \geq 0$. Seja $p_j = P(j \leq X < j + 1)$. Verifique que p_j é da forma $(1-a)a^j$ e determine a.

4.11. A variável aleatória contínua X tem fdp $f(x) = 3x^2$, $-1 \leq x \leq 0$ se b for um número que satisfaça a $-1 < b < 0$, calcule $P(X > b \mid X < b/2)$.

4.12. Suponha que f e g sejam fdp no mesmo intervalo $a \leq x \leq b$.

(*a*) Verifique que $f + g$ não é uma fdp nesse intervalo.

(*b*) Verifique que, para todo número β, $0 < \beta < 1$, $\beta f(x) + (1 - \beta)g(x)$ é uma fdp nesse intervalo.

4.13. Suponha que o gráfico na Fig. 4.16 represente a fdp de uma variável aleatória X.

(*a*) Qual será a relação entre a e b?

(*b*) Se $a > 0$ e $b > 0$, que se pode dizer do maior valor que b pode tomar? (Veja a Fig. 4.16.)

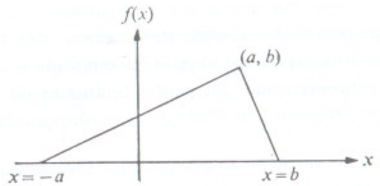


Fig. 4.16

4.14. A percentagem de álcool (100 X) em certo composto pode ser considerada uma variável aleatória, onde X, $0 < X < 1$, tem a seguinte fdp:

$$f(x) = 20x^3(1 - x), \qquad 0 < x < 1.$$

(*a*) Estabeleça a expressão da fd F e esboce seu gráfico.

(*b*) Calcule $P(X \leq 2/3)$.

(*c*) Suponha que o preço de venda desse composto dependa do conteúdo de álcool. Especificamente, se $1/3 < X < 2/3$, o composto se vende por C_1 dólares/galão; caso contrário, ele se vende por C_2 dólares/galão. Se o custo for C_3 dólares/galão, calcule a distribuição de probabilidade do lucro líquido por galão.

4.15. Seja X uma variável aleatória contínua, com fdp dada por

$$\begin{aligned} f(x) &= ax, & 0 \leq x \leq 1,\\ &= a, & 1 \leq x \leq 2,\\ &= -ax + 3a, & 2 \leq x \leq 3,\\ &= 0, & \text{para quaisquer outros valores.} \end{aligned}$$

(*a*) Determine a constante a. (*b*) Determine a fd F e esboce o seu gráfico. (*c*) Se X_1, X_2 e X_3 forem três observações independentes de X, qual será a probabilidade de, exatamente, um desses três números ser maior do que 1,5?

4.16. O diâmetro X de um cabo elétrico supõe-se ser uma variável aleatória contínua X, com fdp $f(x) = 6x(1 - x)$, $0 \leq x \leq 1$.

(*a*) Verifique que essa expressão é uma fdp e esboce o seu gráfico.

(*b*) Obtenha uma expressão para a fd de X e esboce o seu gráfico.

(*c*) Determine um número b tal que $P(X < b) = 2P(X > b)$.

(*d*) Calcule $P(X \leq 1/2 \mid 1/3 < X < 2/3)$.

4.17. Cada uma das seguintes funções representa a fd de uma variável aleatória contínua. Em cada caso, $F(x) = 0$ para $x < a$ e $F(x) = 1$ para $x > b$, onde $[a, b]$ é o intervalo indicado. Em cada caso, esboce o gráfico da função F, determine a fdp f e faça o seu gráfico. Também verifique que f é uma fdp.

(*a*) $F(x) = x/5$, $0 \leq x \leq 5$. (*b*) $F(x) = (2/\pi)\,\text{sen}^{-1}(\sqrt{x})$, $0 \leq x \leq 1$.

(*c*) $F(x) = e^{3x}$, $-\infty < x \leq 0$. (*d*) $F(x) = x^3/2 + \dfrac{1}{2}$, $-1 \leq x \leq 1$.

4.18. Seja X a duração da vida (medida em horas) de um dispositivo eletrônico. Suponha que X seja X variável aleatória contínua com fdp $f(x) = k/x^n$, $2.000 \leq x \leq 10.000$.

(*a*) Para $n = 2$, determine k. (*b*) Para $n = 3$, determine k. (*c*) Para n em geral, determine k. (*d*) Qual a probabilidade de que o dispositivo falhe antes que 5.000 horas se tenham passado? (*e*) Esboce a fd $F(t)$ para a letra (*c*) e determine sua forma algébrica.

4.19. Seja X uma variável aleatória com distribuição binomial, baseada em 10 repetições de um experimento. Se $p = 0{,}3$, calcule as seguintes probabilidades, empregando a tábua da distribuição binomial do Apêndice:

(*a*) $P(X \leq 8)$; (*b*) $P(X = 7)$; (*c*) $P(X > 6)$.

4.20. Suponha que X seja uniformemente distribuída sobre $[-\alpha, +\alpha]$, onde $\alpha > 0$. Quando possível, determine α de modo que as seguintes relações sejam satisfeitas:

(*a*) $P(X > 1) = \dfrac{1}{3}$. (*b*) $P(X > 1) = \dfrac{1}{2}$. (*c*) $P(X < \dfrac{1}{2}) = 0{,}7$.

(*d*) $P(X < \dfrac{1}{2}) = 0{,}3$. (*e*) $P(|X| < 1) = P(|X| > 1)$.

4.21. Suponha que X tenha distribuição uniforme sobre $[0, \alpha]$, $\alpha > 0$. Responda às perguntas do Probl. 4.20.

4.22. Um ponto é escolhido ao acaso, sobre uma reta de comprimento L. Qual é a probabilidade de que o quociente do segmento mais curto para o mais longo seja menor do que 1/4?

4.23. Uma fábrica produz 10 recipientes de vidro por dia. Deve-se supor que exista uma probabilidade constante $p = 0{,}1$ de produzir um recipiente defeituoso. Antes que esses recipientes sejam estocados, eles são inspecionados e os defeituosos são separados. Admita que exista uma probabilidade constante $r = 0{,}1$ de que um recipiente defeituoso seja mal classificado. Faça X igual ao número de recipientes classificados como defeituosos ao fim de um dia de produção. (Admita que todos os recipientes fabricados em um dia sejam inspecionados naquele dia.)

(*a*) Calcule $P(X = 3)$ e $P(X > 3)$. (*b*) Obtenha a expressão de $P(X = k)$.

4.24. Suponha que 5 por cento de todas as peças que saiam de uma linha de fabricação sejam defeituosas. Se 10 dessas peças forem escolhidas e inspecionadas, qual será a probabilidade de que no máximo 2 defeituosas sejam encontradas?

4.25. Suponha que a duração da vida (em horas) de uma certa válvula seja uma variável aleatória contínua X com fdp $f(x) = 100/x^2$, para $x > 100$, e zero para quaisquer outros valores de x.

(*a*) Qual será a probabilidade de que uma válvula dure menos de 200 horas, se soubermos que ela ainda está funcionando após 150 horas de serviço?

(*b*) Se três dessas válvulas forem instaladas em um conjunto, qual será a probabilidade de que exatamente uma delas tenha de ser substituída após 150 horas de serviço?

(*c*) Qual será o número máximo de válvulas que poderá ser colocado em um conjunto, de modo que exista uma probabilidade de 0,5 de que após 150 horas de serviço todas elas ainda estejam funcionando?

4.26. Um experimento consiste em n tentativas independentes. Deve-se admitir que por causa da "aprendizagem", a probabilidade de obter um resultado favorável cresça com o número de tentativas realizadas. Especificamente, suponha que P (sucesso na i-ésima repetição) $= (i + 1)/(i + 2)$, $i = 1, 2, \ldots, n$.

(*a*) Qual será a probabilidade de ter ao menos 3 resultados favoráveis, em 8 repetições?

(*b*) Qual será a probabilidade de que o primeiro resultado favorável ocorra na oitava repetição?

4.27. Com referência ao Ex. 4.10:

(*a*) Calcule $P(X = 2)$, se $n = 4$.

(*b*) Para n arbitrário, verifique que $P(X = n - 1) = P$ (exatamente uma tentativa mal sucedida) é igual a $[1/(n + 1)] \sum_{i=1}^{n} (1/i)$.

4.28. Se a variável aleatória K fôr uniformemente distribuída sobre (0, 5), qual será a probabilidade de que as raízes da equação $4x^2 + 4xK + K + 2 = 0$ sejam reais?

4.29. Suponha que a variável aleatória X tenha valores possíveis, $1, 2, 3, \ldots$ e que

$$P(X = r) = k(1 - \beta)^{r-1}, 0 < \beta < 1.$$

(*a*) Determine a constante k.

(*b*) Ache a *moda* desta distribuição (isto é, o valor de r que torne $P(X = r)$ a maior de todas).

4.30. Uma variável aleatória X pode tomar quatro valores, com probabilidades $(1 + 3x)/4$, $(1 - x)/4$, $(1 + 2x)/4$ e $(1 - 4x)/4$. Para que valores de x é esta uma distribuição de probabilidade?

Funções de Variáveis Aleatórias

Capítulo 5

5.1. Um Exemplo

Suponhamos que o raio do orifício de um tubo calibrado com precisão X seja considerado uma variável aleatória contínua com fdp f. Seja $A = \pi X^2$ a área da secção transversal do orifício. É intuitivamente evidente que, uma vez que o valor de X é o resultado de um experimento aleatório, o valor de A também o é. Quer dizer, A é uma variável aleatória (contínua) e desejamos obter sua fdp, que denotaremos g. Esperamos, uma vez que A é função de X, que a fdp g seja de algum modo deduzível do conhecimento da fdp f. Neste capítulo, trataremos de problemas desse tipo geral. Antes porém de nos familiarizarmos com algumas das técnicas específicas necessárias, vamos exprimir os conceitos acima mais rigorosamente.

5.2. Eventos Equivalentes

Seja $\mathcal{E}$ um experimento e seja S um espaço amostral associado a $\mathcal{E}$. Seja X uma variável aleatória definida em S. Suponha-se que $y = H(x)$ seja uma função real de x. Então, $Y = H(X)$ é uma

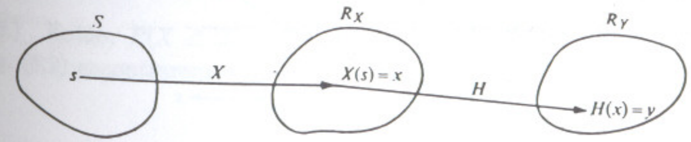


Fig. 5.1

variável aleatória, porque para todo $s \in S$, um valor de Y fica determinado, a saber $y = H[X(s)]$. Esquematicamente, teremos a Fig. 5.1.

Como anteriormente, denominaremos R_X o contradomínio de X, o conjunto de todos os valores possíveis da função X. Semelhantemente, definiremos R_Y como o *contradomínio da variável aleatória* Y, o conjunto de todos os valores possíveis de Y. Anteriormente, já definimos [Eq. (4.1)] a noção de eventos equivalentes em S e em R_X. Agora, estenderemos esse conceito na seguinte forma natural.

Definição. Seja C um evento (subconjunto) associado ao contradomínio R_Y, de Y, como se explicou acima. Seja $B \subset R_X$ definido assim:

$$B = \{x \in R_X : H(x) \in C\}. \tag{5.1}$$

Em palavras: B é o conjunto de todos os valores de X, tais que $H(x) \in C$. Se B e C forem relacionados desse modo, os denominaremos *eventos equivalentes*.

Comentários: (*a*) Como anteriormente, a interpretação não formal disso é que B e C serão eventos equivalentes se, e somente se, B e C ocorrerem conjuntamente. Isto é, quando B ocorrer, C ocorrerá, e inversamente.

(*b*) Suponha-se que A seja um evento associado a S, o qual é equivalente a um evento B associado a R_X. Então, se C for um evento associado a R_Y o qual é equivalente a B, teremos que A será equivalente a C.

(*c*) É também importante compreender que quando falamos de eventos equivalentes (no sentido acima), esses eventos são associados a diferentes espaços amostrais.

Exemplo 5.1. Suponha-se que $H(x) = \pi x^2$, tal como na Seç. 5.1. Então, os eventos B: $\{X > 2\}$ e C: $\{Y > 4\pi\}$ são equivalentes. Porque, se $Y = \pi X^2$, então $\{X > 2\}$ ocorrerá se, e somente se, $\{Y > 4\pi\}$ ocorrer, desde que X não pode tomar valores negativos no caso presente. (Veja a Fig. 5.2.)

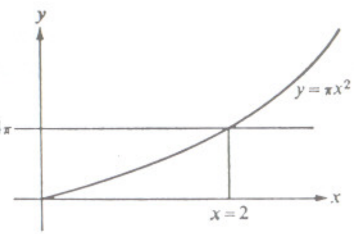


Fig. 5.2

Comentário: É também importante salientar que uma notação abreviada está sendo empregada quando escrevemos expressões tais como $\{X > 2\}$ e $\{Y > 4\pi\}$. Aquilo a que nos estaremos referindo, naturalmente, são os valores de X e os valores de Y, isto é, $\{s \mid X(s) > 2\}$ e $\{x \mid Y(x) > 4\pi\}$.

Tal como fizemos no Cap. 4, [Eq. (4.2)], daremos a seguinte definição.

Definição: Seja uma variável aleatória X definida no espaço amostral S. Seja R_X o contradomínio de X. Seja H uma função real e considere-se a variável aleatória $Y = H(X)$ com contradomínio R_Y. Para qualquer evento $C \subset R_Y$, *definiremos* $P(C)$ assim:

$$P(C) = P[\{x \in R_X : H(x) \in C\}]. \tag{5.2}$$

Em linguagem corrente: A probabilidade de um evento associado ao contradomínio de Y é definida como a probabilidade do evento equivalente (em termos de X), como indicado pela Eq. (5.2).

Comentários: (*a*) A definição acima torna possível calcular probabilidades que envolvam eventos associados a Y, se conhecermos a distribuição de probabilidade de X e se pudermos determinar o evento equivalente em apreço.

(*b*) Uma vez que explicamos anteriormente [Eq. (4.1 e 4.2)] como relacionar probabilidades associadas a R_X com probabilidades associadas a S, podemos reescrever a Eq. (5.2) assim:

$$P(C) = P[\{x \in R_X : H(x) \in C\}] = P[\{s \in S : H[X(s)] \in C\}].$$

Exemplo 5.2. Seja X uma variável contínua com fdp

$$f(x) = e^{-x}, \qquad x > 0.$$

(Uma integração simples confirma que $\int_0^\infty e^{-x}\,dx = 1$.)

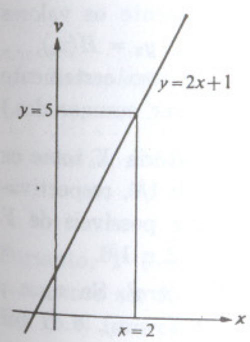


Fig. 5.3

Suponha-se que $H(x) = 2x + 1$. Em conseqüência, $R_X = \{x \mid x > 0\}$, enquanto $R_Y = \{y \mid y > 1\}$. Suponha-se que o evento C seja definido deste modo: $C = \{Y \geq 5\}$. Então, $y \geq 5$ se, e somente se, $2x + 1 \geq 5$, o que por sua vez acarreta $x \geq 2$. Daí, C é equivalente a $B = \{X \geq 2\}$. (Veja Fig. 5.3.) Então, $P(X \geq 2) = \int_2^\infty e^{-x}\,dx = 1/e^2$. Aplicando-se então a Eq. (5.2) encontraremos que

$$P(Y \geq 5) = 1/e^2.$$

Comentários: (*a*) É novamente proveitoso salientar que poderemos considerar a incorporação de ambas as avaliações de $x = X(s)$ e de $y = H(x)$ em nosso experimento e, conseqüentemente, considerar apenas R_Y, o contradomínio de Y, como o espaço amostral de nosso experimento.

Rigorosamente falando, o espaço amostral de nosso experimento é S e o resultado do experimento é s. Tudo o que se faz subseqüentemente não é influenciado pela natureza aleatória do experimento. A determinação de $x = X(s)$ e a avaliação de $y = H(x)$ são processos rigorosamente determinísticos depois que s tenha sido observado. Contudo, como já explicamos, podemos incorporar esses cálculos na descrição de nosso experimento e, deste modo, tratar diretamente com o contradomínio R_Y.

(*b*) Exatamente do modo como a distribuição de probabilidade foi induzida em R_X pela distribuição de probabilidade sobre o espaço amostral original S, a distribuição de probabilidade de Y será determinada quando a distribuição de probabilidade de X for conhecida. Assim, no Ex. 5.2 acima, a distribuição especificada de X determinou completamente o valor de $P(Y \geq 5)$.

(*c*) Ao considerar uma função de uma variável aleatória X, digamos $Y = H(X)$, devemos observar que nem toda função H concebível poderá ser aceita. Contudo, as funções que surgem nas aplicações estão infalivelmente entre aquelas que podemos considerar e, por isso, não nos referiremos mais a esta pequena dificuldade.

5.3. Variáveis Aleatórias Discretas

Caso 1. X é uma variável aleatória discreta. Se X for uma variável aleatória discreta e $Y = H(X)$, nesse caso segue-se imediatamente que Y será também uma variável aleatória discreta.

Porque supor que os valores possíveis de X possam ser enumerados como $x_1, x_2, \ldots, x_n, \ldots$ acarreta que certamente os valores possíveis de Y sejam enumerados como $y_1 = H(x_1)$, $y_2 = H(x_2), \ldots$ (Alguns desses valores de Y poderão ser iguais, mas isso certamente não perturba o fato de que esses valores possam ser enumerados.)

Exemplo 5.3. Suponhamos que a variável aleatória X tome os três valores -1, 0 e 1, com probabilidades 1/3, 1/2 e 1/6, respectivamente. Seja $Y = 3X + 1$. Nesse caso os valores possíveis de Y são -2, 1 e 4, tomados com probabilidades 1/3, 1/2 e 1/6.

Este exemplo sugere o seguinte *procedimento geral*: Se $x_1, \ldots, x_n, \ldots$ forem os valores possíveis de X, $p(x_i) = P(X = x_i)$, e H for uma função tal que, a cada valor y corresponda exatamente um valor x, então a distribuição de probabilidade de Y será obtida do seguinte modo:

Valores possíveis de Y: $\quad y_i = H(x_i),\ i = 1, 2, \ldots, n, \ldots;$

Probabilidades de Y: $\quad q(y_i) = P(Y = y_i) = p(x_i).$

Muito freqüentemente a função H não possui a característica acima, e poderá acontecer que vários valores de X levem ao mesmo valor de Y, como ilustra o exemplo seguinte.

Exemplo 5.4. Suponha-se que consideramos a mesma variável aleatória X, como no Ex. 5.3 acima. Contudo, introduzimos $Y = X^2$.

Portanto, os valores possíveis de Y são zero e um, tomados com probabilidades 1/2 e 1/2, porque $Y = 1$ se, e somente se, $X = -1$ ou $X = 1$ e a probabilidade deste último evento é $1/3 + 1/6 = 1/2$. Em termos de nossa terminologia preliminar, os eventos B: $\{X = \pm 1\}$ e C: $\{Y = 1\}$ são eventos equivalentes e, em conseqüência, pela Eq. (5.2) têm iguais probabilidades.

O *procedimento geral* para situações como a apresentada no exemplo acima é o seguinte: Sejam $x_{i_1}, x_{i_2}, \ldots, x_{i_k}, \ldots$ os valores de X que tenham a propriedade $H(x_{i_j}) = y_i$ para todo j. Então,

$$q(y_i) = P(Y = y_i) = p(x_{i_1}) + p(x_{i_2}) + \ldots$$

isto é, para calcular a probabilidade do evento $\{Y = y_i\}$, acha-se o evento equivalente em termos de X (no contradomínio R_X) e em seguida adicionam-se todas as probabilidades correspondentes. (Veja a Fig. 5.4.)

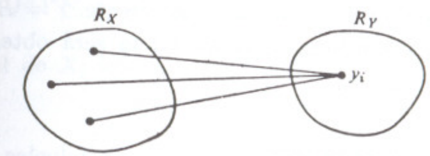


Fig. 5.4

Exemplo 5.5. Admita-se que X tenha os valores possíveis $1, 2, \ldots, n, \ldots$ e suponha-se que $P(X = n) = (1/2)^n$. Seja

$$Y = 1 \quad \text{se } X \text{ for par,}$$
$$Y = -1 \quad \text{se } X \text{ for ímpar.}$$

Portanto, Y toma os dois valores -1 e $+1$. Desde que $Y = 1$ se, e somente se, $X = 2$, ou $X = 4$, ou $X = 6$, ou $\ldots$, a aplicação da Eq. (5.2) fornece

$$P(Y = 1) = \frac{1}{4} + \frac{1}{16} + \frac{1}{64} + \ldots = \frac{1}{3}.$$

Conseqüentemente:

$$P(Y = -1) = 1 - P(Y = 1) = \frac{2}{3}$$

Caso 2. X é uma variável aleatória contínua. Pode acontecer que X seja uma variável aleatória contínua enquanto Y seja discreta. Por exemplo, suponha-se que X possa tomar todos os valores reais, enquanto Y seja definido igual a $+1$ se $X \geq 0$, e

$Y = -1$ se $X < 0$. A fim de obter a distribuição de probabilidade de Y, determina-se apenas o evento equivalente (no contradomínio R_X) correspondente aos diferentes valores de Y. Neste caso, $Y = 1$ se, e somente se, $X \geq 0$, enquanto $Y = -1$ se, e somente se, $X < 0$. Por isso, $P(Y = 1) = P(X \geq 0)$, enquanto $P(Y = -1) = P(X < 0)$. Se a fdp de X for conhecida, essas probabilidades poderão ser calculadas. No caso geral, se $\{Y = y_i\}$ for equivalente a um evento, por exemplo A, no contradomínio de X, então

$$q(y_i) = P(Y = y_i) = \int_A f(x)\,dx.$$

5.4. Variáveis Aleatórias Contínuas

O caso mais importante (e mais freqüentemente encontrado) aparece quando X for uma variável aleatória contínua com fdp f e H for uma função contínua. Conseqüentemente $Y = H(X)$ será uma variável aleatória contínua, e nossa tarefa será obter sua fdp, que denotaremos por g.

O *procedimento geral* será:

(*a*) Obter G, a fd de Y, na qual $G(y) = P(Y \leq y)$, achando-se o evento A (no contradomínio de X) o qual é equivalente ao evento $\{Y \leq y\}$.

(*b*) Derivar $G(y)$ em relação a y, a fim de obter $g(y)$.

(*c*) Determinar aqueles valores de y no contradomínio de Y, para os quais $g(y) > 0$.

Exemplo 5.6. Suponhamos que X tenha fdp

$$f(x) = 2x, \quad 0 < x < 1,$$
$$= 0, \text{ para outros quaisquer valores.}$$

Seja $H(x) = 3x + 1$. Daí, para obter a fdp de $Y = H(X)$, teremos (veja a Fig. 5.5).

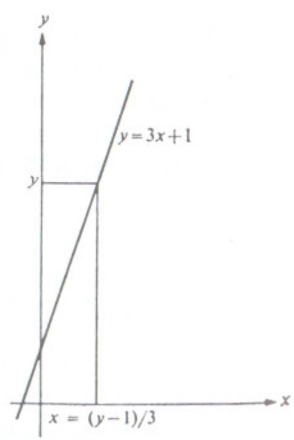


Fig. 5.5

$$G(y) = P(Y \leq y) = P(3X + 1 \leq y)$$
$$= P[X \leq (y-1)/3]$$
$$= \int_0^{(y-1)/3} 2x\,dx = [(y-1)/3]^2.$$

Daí

$$g(y) = G'(y) = \frac{2}{9}(y-1).$$

Desde que $f(x) > 0$ para $0 < x < 1$, encontramos que $g(y) > 0$ para $1 < y < 4$.

Comentário: O evento A, referido acima, equivalente ao evento $\{Y \leq y\}$ é apenas $\{X \leq (y-1)/3\}$.

Existe uma outra maneira, ligeiramente diferente, de obter o mesmo resultado, a qual será de utilidade mais tarde. Consideremos novamente

$$G(y) = P(Y \leq y) = P\left(X \leq \frac{y-1}{3}\right) = F\left(\frac{y-1}{3}\right),$$

onde F é a fd de X; isto é,

$$F(x) = P(X \leq x).$$

A fim de calcular a derivada de G, $G'(y)$, empregaremos a regra de derivação de função, como segue:

$$\frac{dG(y)}{dy} = \frac{dG(y)}{du} \cdot \frac{du}{dy}, \quad \text{onde} \quad u = \frac{y-1}{3}.$$

Portanto,

$$G'(y) = F'(u) \cdot \frac{1}{3} = f(u) \cdot \frac{1}{3} = 2\left(\frac{y-1}{3}\right) \cdot \frac{1}{3},$$

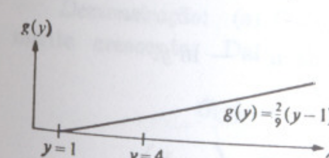


Fig. 5.6

como anteriormente. A fdp de Y tem o gráfico apresentado na Fig. 5.6. (Para verificar o cálculo, observe que $\int_1^4 g(y)dy = 1$.)

Exemplo 5.7. Suponhamos que uma variável aleatória contínua tem a fdp como foi dada no Ex. 5.6. Seja $H(x) = e^{-x}$. Para achar a fdp de $Y = H(X)$, procederemos como se indica a seguir (veja a Fig. 5.7):

$$G(y) = P(Y \leq y) = P(e^{-X} \leq y)$$
$$= P(X \geq -\ln y) = \int_{-\ln y}^{1} 2x\,dx$$
$$= 1 - (-\ln y)^2.$$

Daí, $g(y) = G'(y) = -2 \ln y / y$. Visto que $f(x) > 0$ para $0 < x < 1$, encontramos que $g(y) > 0$ para $1/e < y < 1$. [Observe que o sinal algébrico para $g(y)$ está correto, pois que $\ln y < 0$ para $1/e < y < 1$.] O gráfico de $g(y)$ está esboçado na Fig. 5.8.

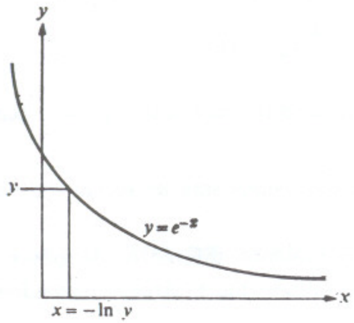


Fig. 5.7

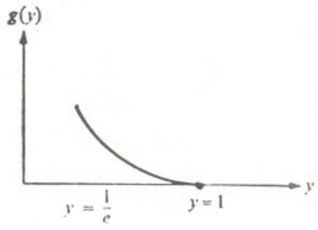


Fig. 5.8

Poderemos também obter o resultado acima por um tratamento um pouco diferente, que esboçaremos resumidamente. Tal como anteriormente

$$G(y) = P(Y \leq y) = P(X \geq -\ln y)$$
$$= 1 - P(X \leq -\ln y) = 1 - F(-\ln y),$$

onde F é a fd de X, como antes. A fim de obter a derivada de G, aplicaremos também a regra de derivação de função de função, como se segue:

$$\frac{dG(y)}{dy} = \frac{dG}{du}\,\frac{du}{dy}, \qquad \text{onde } u = -\ln y.$$

Deste modo

$$G'(y) = -F'(u)\left(-\frac{1}{y}\right) = +2 \ln y \cdot \left(-\frac{1}{y}\right),$$

tal como anteriormente.

Vamos agora generalizar o tratamento sugerido pelos exemplos acima. O passo mais importante em cada um dos exemplos foi dado quando substituímos o evento $\{Y \leq y\}$ pelo evento equivalente em termos da variável aleatória X. Nos problemas acima, isso foi relativamente fácil porque em cada caso a função de X era estritamente crescente ou estritamente decrescente.

Na Fig. 5.9, y é uma função estritamente crescente de x. Por isso, poderemos resolver $y = H(x)$ em termos de y, isto é, $x = H^{-1}(y)$, onde H^{-1} é denominada função inversa de H. Portanto, se H for estritamente crescente $\{H(X) \leq y\}$ será equivalente a $\{X \leq H^{-1}(y)\}$, enquanto se H for estritamente decrescente, $\{H(X) \leq y\}$ será equivalente a $\{X \geq H^{-1}(y)\}$.

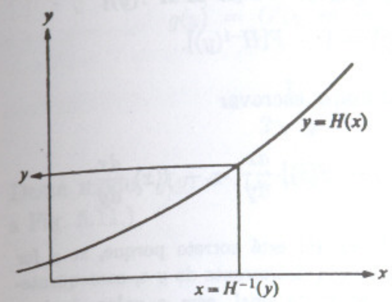


Fig. 5.9

O processo empregado nos exemplos acima pode agora ser generalizado, na seguinte forma:

Teorema 5.1. Seja X uma variável aleatória contínua com fdp f, onde $f(x) > 0$, para $a < x < b$. Suponha-se que $y = H(x)$ seja uma função de x estritamente monótona (ou crescente ou decrescente). Admita-se que essa função seja derivável (e, portanto, contínua) para todo x. Então, a variável aleatória Y, definida como $Y = H(X)$ possui a fdp g dada por

$$g(y) = f(x)\left|\frac{dx}{dy}\right|, \tag{5.3}$$

onde x é expresso em termos de y. Se H for crescente, então g será não-nula para aqueles valores de y que satisfaçam $H(a) < y < H(b)$. Se H for decrescente, então g será não-nula para aqueles valores de y que satisfaçam $H(b) < y < H(a)$.

Demonstração: (a) Suponha-se que H seja uma função estritamente crescente. Daí

$$G(y) = P(Y \leq y) = P[H(X) \leq y]$$
$$= P[X \leq H^{-1}(y)] = F[H^{-1}(y)].$$

Derivando $G(y)$ em relação a y, obteremos com o emprego da regra da derivada de função de função:

$$\frac{dG(y)}{dy} = \frac{dG(y)}{dx}\,\frac{dx}{dy}, \qquad \text{onde } x = H^{-1}(y).$$

Portanto,

$$G'(y) = \frac{dF(x)}{dx}\,\frac{dx}{dy} = f(x)\,\frac{dx}{dy}.$$

(*b*) Suponha-se que H seja uma função decrescente. Daí

$$G(y) = P(Y \leq y) = P[H(X) \leq y] = P[X \geq H^{-1}(y)]$$
$$= 1 - P[X \leq H^{-1}(y)] = 1 - F[H^{-1}(y)].$$

Procedendo tal como acima, poderemos escrever

$$\frac{dG(y)}{dy} = \frac{dG(y)}{dx}\frac{dx}{dy} = \frac{d}{dx}[1 - F(x)]\frac{dx}{dy} = -f(x)\frac{dx}{dy}.$$

Comentário: O sinal algébrico obtido em (*b*) está correto porque, se y for uma função decrescente de x, x será uma função decrescente de y e, conseqüentemente, $dx/dy < 0$. Deste modo, pelo emprego do sinal, com o valor absoluto em torno de dx/dy, poderemos combinar o resultado de (*a*) e de (*b*) e obter a forma final do teorema.

Exemplo 5.8. Vamos reexaminar os Exs. 5.6 e 5.7 pela aplicação do Teor. 5.1.

(*a*) No Ex. 5.6 tivemos $f(x) = 2x$, $0 < x < 1$, e $y = 3x + 1$. Conseqüentemente, $x = (y - 1)/3$ e $dx/dy = 1/3$. Por isso, $g(y) = 2[(y - 1)/3](1/3) = (2/9)(y - 1)$, $1 < y < 4$, o que confirma o resultado obtido anteriormente.

(*b*) No Ex. 5.7, tivemos $f(x) = 2x$, $0 < x < 1$ e $y = e^{-x}$. Em conseqüência, $x = -\ln y$ e $dx/dy = -1/y$. Deste modo, $g(y) = -2(\ln y)/y$, $1/e < y < 1$, o que também confirma o resultado já obtido.

Se $y = H(x)$ não for uma função monótona de x, não poderemos aplicar diretamente o processo acima. Em vez disso, voltaremos ao processo geral esquematizado acima. O exemplo seguinte ilustra esse procedimento.

Exemplo 5.9. Suponhamos que

$$f(x) = 1/2, \quad -1 < x < 1,$$
$$= 0, \quad \text{fora desse intervalo.}$$

Seja $H(x) = x^2$. Obviamente, esta *não* é uma função monótona sobre o intervalo $[-1, 1]$ (Fig. 5.10). Por isso, obteremos a fdp de $Y = X^2$ do seguinte modo:

$$G(y) = P(Y \leq y) = P(X^2 \leq y)$$
$$= P(-\sqrt{y} \leq X \leq \sqrt{y})$$
$$= F(\sqrt{y}) - F(-\sqrt{y}),$$

onde F é a fd da variável aleatória X. Logo,

$$g(y) = G'(y) = \frac{f(\sqrt{y})}{2\sqrt{y}} - \frac{f(-\sqrt{y})}{-2\sqrt{y}}$$
$$= \frac{1}{2\sqrt{y}}[f(\sqrt{y}) + f(-\sqrt{y})].$$

Deste modo, $g(y) = (1/2\sqrt{y})(1/2 + 1/2) = 1/2\sqrt{y}, 0 < y < 1$. (Veja a Fig. 5.11.)

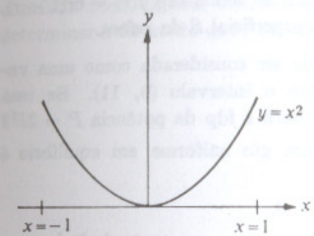


Fig. 5.10

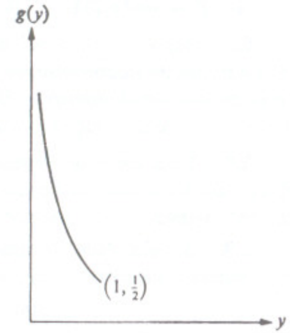


Fig. 5.11

O processo empregado no exemplo acima fornece o seguinte resultado geral.

Teorema 5.2. Seja X uma variável aleatória contínua com fdp f. Façamos $Y = X^2$. Então, a variável aleatória Y tem a fdp dada por

$$g(y) = \frac{1}{2\sqrt{y}}[f(\sqrt{y}) + f(-\sqrt{y})].$$

Demonstração: Veja o Ex. 5.9.

Problemas

5.1. Suponha que X seja uniformemente distribuída sobre $(-1, 1)$. Seja $Y = 4 - X^2$. Achar a fdp de Y, $g(y)$, e fazer seu gráfico. Verifique também que $g(y)$ é a fdp adequada.

5.2. Suponha que X seja uniformemente distribuída sobre $(1, 3)$. Ache a fdp das seguintes variáveis aleatórias:

(*a*) $Y = 3X + 4$, (*b*) $Z = e^X$.

Verifique em cada caso que a função obtida é a fdp. Esboce os gráficos.

5.3. Suponha que a variável aleatória contínua X tenha fdp $f(x) = e^{-x}$, $x > 0$. Ache a fdp das seguintes variáveis aleatórias:

(*a*) $Y = X^3$, (*b*) $Z = 3/(X+1)^2$.

5.4. Suponha que a variável aleatória discreta X tome os valores 1, 2 e 3 com igual probabilidade. Ache a distribuição de probabilidade de $Y = 2X + 3$.

5.5. Suponha que X seja uniformemente distribuída sobre o intervalo (0, 1). Ache a fdp das seguintes variáveis aleatórias:

(*a*) $Y = X^2 + 1$, (*b*) $Z = 1/(X+1)$.

5.6. Suponha que X seja uniformemente distribuída sobre $(-1, 1)$. Ache a fdp das seguintes variáveis aleatórias:

(*a*) $Y = \operatorname{sen}(\pi/2)X$, (*b*) $Z = \cos(\pi/2)X$, (*c*) $W = |X|$.

5.7. Suponha que o raio de uma esfera seja uma variável aleatória contínua. (Em virtude de imprecisões do processo de fabricação, os raios das diferentes esferas podem ser diferentes.) Suponha que o raio R tenha fdp $f(r) = 6r(1-r)$, $0 < r < 1$. Ache a fdp do volume V e da área superficial S da esfera.

5.8. Uma corrente elétrica oscilante I pode ser considerada como uma variável aleatória uniformemente distribuída sobre o intervalo (9, 11). Se essa corrente passar em um resistor de 2 ohms, qual será a fdp da potência $P = 2I^2$?

5.9. A velocidade de uma molécula em um gás uniforme em equilíbrio é uma variável aleatória V cuja fdp é dada por

$$f(v) = av^2 e^{-bv^2}, \quad v > 0,$$

onde $b = m/2kT$ e k, T e m denotam respectivamente a constante de Boltzman, a temperatura absoluta e a massa da molécula.

(*a*) Calcular a constante a (em termos de b). [*Sugestão*: Considere o fato de que $\int_0^\infty e^{-x^2}dx = \sqrt{\pi}/2$ e integre por partes.]

(*b*) Estabeleça a distribuição da variável aleatória $W = mV^2/2$, a qual representa a energia cinética da molécula.

5.10. A tensão elétrica aleatória X é uniformemente distribuída sobre o intervalo $(-k, k)$. Se Y for a entrada de um dispositivo não-linear, com as características indicadas na Fig. 5.12, ache a distribuição de probabilidade de Y, nos três casos seguintes: (*a*) $k < a$, (*b*) $a < k < x_0$, (*c*) $k > x_0$.

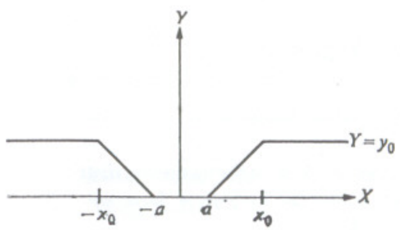


Fig. 5.12

Comentário: A distribuição de probabilidade de Y constitui um exemplo de uma distribuição *mista*. Y toma o valor zero com probabilidade não-nula e também toma todos os valores em certos intervalos. (Veja a Seç. 4.6.)

5.11. A energia radiante (em Btu/hora/pé²) é dada pela seguinte função da temperatura T (em escala Fahrenheit): $E = 0{,}173\,(T/100)^4$. Suponha que a temperatura T seja considerada uma variável aleatória contínua como fdp

$$f(t) = 200\,t^{-2}, \quad 40 \le t \le 50,$$
$$= 0, \quad \text{para outros quaisquer valores.}$$

Estabeleça a fdp da energia radiante E.

5.12. Para medir velocidades do ar, utiliza-se um tubo (conhecido como tubo estático de Pitot), o qual permite que se meça a pressão diferencial. Esta pressão diferencial é dada por $P = (1/2)\,dV^2$, onde d é a densidade do ar e V é a velocidade do vento (mph). Achar a fdp de P, quando V for uma variável aleatória uniformemente distribuída sobre (10, 20).

5.13. Suponha que $P(X \leqslant 0{,}29) = 0{,}75$, onde X é uma variável aleatória contínua com alguma distribuição definida sobre (0, 1). Quando $Y = 1 - X$, determinar k de modo que $P(Y \leqslant k) = 0{,}25$.

CENTRAL
PUCRS