Introdução à Probabilidade

notas de aula – `2024.11.27.11.37`

Jerônimo Pellegrini

id: `bb4add1ed2dbbbfd190f9497fa7225373aec046a`

Sumário

Nomenclatura

$(X_1, X_2, \dots, X_n)$ variável aleatória multidimensional, página 133

α nível de significância, página 200

$\mathrm{corr}(X, Y)$ correlação entre X e Y, página 165

$\mathrm{cov}(X, Y)$ covariância entre X e Y, página 162

$\mathcal{E}xp(\lambda)$ distribuição exponencial com parâmetro λ, página 113

$\left(\!\binom{n}{k}\!\right)$ combinação de multiconjuntos, página 16

$\mathcal{N}(\mu, \sigma^2)$ distribuição normal com média μ e variância σ^2, página 106

Ω espaço amostral, página 1

$\Phi(x)$ função distribuição de $\mathcal{N}(0, 1)$, página 108

$\mathcal{P}oiss(\lambda)$ distribuição de Poisson com parâmetro λ, página 76

$\Pr(A)$ probabilidade do evento A, página 6

$\rho_{X,Y}$ correlação entre X e Y, página 165

$\mathrm{SD}(X)$ desvio padrão da variável X, página 67

$\mathrm{SD}(X)$ desvio padrão da variável X, página 101

σ_X^2 variância de X, página 66

σ_X desvio padrão da variável X, página 67

σ_X desvio padrão da variável X, página 101

$\overline{A}$ complemento de A, página 7

$\mathcal{U}(a, b)$ distribuição uniforme no intervalo $[a, b]$, página 102

$\mathrm{var}(X)$ variância de X, página 66

$A \times B$ produto cartesiano de A com B, página 11

$f(A)$ frequência do evento A, página 6

$F(x, y)$ $\Pr(X \leq x, Y \leq y$, função acumulada de distribuição conjunta, página 134

$f(x_i, y_j)$ $\Pr[X = x_i, Y = y_j]$, página 134

f_X função de distribuição marginal, página 141

K_n grafo completo com n vértices, página 24

$R(k, k)$ Número de Ramsey diagonal, página 26

r_A definição tentativa de probabilidade usando frequência relativa, página 6

$X \sim D$ X tem distribuição D, página 58

$\binom{n}{k}$ combinação; coeficiente binomial, página 14

Capítulo 1

Introdução

O desenvolvimento da teoria da probabilidade está ligado à observação de jogos de azar – se uma moeda honesta é jogada n vezes, o número de vezes em que o resultado é coroa é próximo de $n/2$; se escolhermos ao acaso uma carta de um baralho n vezes, a quantidade de vezes que a carta será de ouros será próxima de $n/4$; quando um dado honesto é jogado n vezes, a quantidade de vezes em que o resultado é 3 será próxima de $n/6$.

1.1 Experimentos, espaços amostrais e eventos

Para poder modelar fenômenos usando probabilidade, precisamos definir com precisão os objetos matemáticos que usaremos.

Definição 1.1.1 (experimento, espaço amostral)**.** Um **experimento** é um procedimento, passível de repetições por um número indefinido de vezes, com um conjunto não vazio de resultados possíveis.

O conjunto de possíveis resultados de um experimento é chamado de **espaço amostral**. É comum denotar o espaço amostral de um experimento por Ω.

Quando o espaço amostral de um experimento tem somente um elemento, dizemos que o experimento é **determinístico**. Quando tem mais de um elemento, dizemos que o experimento é **não-determinístico**. ◀

Estamos interessados em experimentos não-determinísticos.

Exemplo 1.1.2. O técnico de um time de futebol registrou as quantidades de sucessos de um jogador em cobrança de pênaltis após muitos jogos. Cada chute realizado pelo jogador pode ser modelado como um experimento aleatório, com espaço amostral

$$\Omega = \{\text{sucesso}, \text{falha}\},$$

que são os dois resultados possíveis. ●

Exemplo 1.1.3. Um dado é jogado, e registramos que face do dado ficou voltada para cima.

O espaço amostral é $\{1, 2, 3, 4, 5, 6\}$.

Este é um experimento não-determinístico, porque não modelamos todas as variáveis que influem no resultado. ●

Exemplo 1.1.4. Dois dados diferentes, um branco e um vermelho, são jogados, e registramos os valores das duas faces que ficaram voltadas para cima, sem distinguir o primeiro do segundo dado.

O espaço amostral é

$$\Omega = \left\{ \begin{array}{c} (1,1),(1,2),(1,3),(1,4),(1,5),(1,6), \\ (2,1),(2,2),(2,3),(2,4),(2,5),(2,6), \\ (3,1),(3,2),(3,3),(3,4),(3,5),(3,6), \\ (4,1),(4,2),(4,3),(4,4),(4,5),(4,6), \\ (5,1),(5,2),(5,3),(5,4),(5,5),(5,6), \\ (6,1),(6,2),(6,3),(6,4),(6,5),(6,6) \end{array} \right\}.$$

Da mesma forma que o experimento com um único dado, este é um experimento não-determinístico. ●

Exemplo 1.1.5. Dois dados idênticos são jogados, e registramos os valores das duas faces que ficaram voltadas para cima. Como os dados são idênticos, e não distinguiremos o primeiro do segundo, representaremos cada resultado do experimento como um *conjunto* de números, e não um par ordenado[1]. Teremos menos resultados possíveis que no experimento 1.1.4.

O espaço amostral é

$$\Omega = \left\{ \begin{array}{r} \{1,1\},\{1,2\},\{1,3\},\{1,4\},\{1,5\},\{1,6\}, \\ \{2,2\},\{2,3\},\{2,4\},\{2,5\},\{2,6\}, \\ \{3,3\},\{3,4\},\{3,5\},\{3,6\}, \\ \{4,4\},\{4,5\},\{4,6\}, \\ \{5,5\},\{5,6\}, \\ \{6,6\} \end{array} \right\}.$$

Da mesma forma que o experimento com um único dado, este é um experimento não-determinístico. ●

Exemplo 1.1.6. Selecionamos um ponto no quadrado entre $(0, 0)$ e $(1, 1)$, e registramos o valor das coordenadas.

O espaço amostral é

$$\Omega = \{(x, y) \in \mathbb{R}^2 \ : \ x, y \in [0, 1]\}.$$

Este é um espaço amostral contínuo. ●

[1] Poderíamos ter usado pares ordenados, mas teríamos que incluir, por exemplo, $(2, 3)$ e não $(3, 2)$, e o modelo ficaria ligeiramente mais confuso.

Exemplo 1.1.7. Há quatro candidatos a prefeito em uma cidade, que representaremos como a, b, c e d. Escolhemos um local da cidade e coletamos a intenção de voto de uma pessoa.

O espaço amostral é

$$\Omega = \{\text{branco}, \text{nulo}, a, b, c, d\}.$$ ●

Exemplo 1.1.8. Cartas são selecionadas de um baralho, com reposição, até que uma seja selecionada com naipe vermelho.

O espaço amostral é

$$\Omega = \{V, PV, PPV, PPPV, PPPPV, \ldots\}.$$ ●

Exemplo 1.1.9. Medimos a pressão sistólica de uma pessoa. Rejeitando a possibilidade de pressão negativa, presumimos que a pressão pode ser representada por reais positivos, e o espaço amostral é $\mathbb{R}^+$. Claramente simplificamos muito o experimento ao modelar, já que não cremos poder encontrar, algum dia, qualquer ser humano com pressão igual a dez mil $mmHg$, por exemplo. Ainda assim, o modelo pode ser usado. ●

Exemplo 1.1.10. Um medicamento é administrado a um paciente com cefaléia. Depois de um tempo, verifica-se o resultado (sucesso, caso o paciente tenha melhorado, ou falha, caso a cefaléia permaneça), por isso o espaço amostral é $\{\text{sucesso}, \text{falha}\}$.

Este é um experimento não-determinístico, porque não sabemos quais são os fatores que fariam o paciente reagir bem ao tratamento. ●

Exemplo 1.1.11. Um equipamento eletrônico é ligado, e mantido em funcionamento até falhar. Depois disso o tempo em que permaneceu funcionando é registrado.

Para definir o espaço amostral, definimos antes como será medido o tempo de funcionamento. Podemos escolher horas, dias, ou qualquer outra unidade de tempo. Também devemos decidir se registraremos valores inteiros ou frações. Supondo que tenhamos decidido registrar dias usando números inteiros, o espaço amostral será

$$\{0, 1, 2, \ldots\} = \mathbb{N}.$$

Se quisermos representar o tempo como reais (o que é mais comum), o espaço amostral será $\mathbb{R}$.

Este é um experimento não-determinístico, porque não temos informação detalhada sobre o dispositivo para poder determinar quando ele deixará de funcionar. ●

Exemplo 1.1.12. Um módulo de comunicação transmite um pacote de dados, e aguarda até que a outra ponta responda com um “OK”. O tempo decorrido é registrado.

Podemos definir o espaço amostral como $\mathbb{R}^+$, contendo todos os números positivos.

Neste modelo, fizemos uma simplificação: presumimos que a outra parte sempre responderá, e que podemos esperar indefinidamente por ela. Em uma situação prática, o módulo aguardaria somente um tempo pré-definido e depois presumiria que o pacote de dados se perdeu, enviando-o novamente. ●

Queremos poder discutir não apenas resultados isolados em um espaço amostral, mas conjuntos deles. Por exemplo, será importante poder fazer referência a resultados pares, ímpares, maiores que algum número, etc., por isso é relevante a noção de *evento*.

Definição 1.1.13 (evento)**.** Um subconjunto de um espaço amostral é um **evento**.

Um evento representa um conjunto de possíveis resultados, e o significado de "um evento A ocorreu" é "o resultado obtido, x, pertence ao evento A". ◀

Exemplo 1.1.14. Ao jogar um dado, o espaço amostral é $\Omega = \{1, 2, 3, 4, 5, 6\}$. Alguns eventos em Ω são

- resultado par: $\{2, 4, 6\} \subseteq \Omega$;
- resultado ímpar: $\{1, 3, 5\} \subseteq \Omega$;
- resultado maior que 3: $\{4, 5, 6\} \subseteq \Omega$;
- resultado menor ou igual que 3: $\{1, 2, 3\} \subseteq \Omega$;
- resultado é quadrado perfeito: $\{1, 4\} \subseteq \Omega$. ●

Exemplo 1.1.15. O Exemplo 1.1.2 aborda a situação em que o técnico de um time de futebol registra quantidades de sucessos e falhas de um jogador ao cobrar pênaltis. O espaço amostral usado naquele exemplo é

$$\Omega = \{\text{sucesso}, \text{falha}\},$$

onde podemos definir dois eventos: "SUCESSO"$= \{\text{sucesso}\}$ e "FALHA"$= \{\text{falha}\}$.

Neste exemplo deve ficar claro que os eventos são *conjuntos*, mesmo que unitários. ●

Exemplo 1.1.16. No experimento descrito no Exemplo 1.1.11, em que deixamos um equipamento eletrônico ligado até que ele deixe de funcionar, o espaço amostral é $\Omega = \mathbb{N}$. Para simplificar a exposição, presumimos que todo ano tem 365 dias, e então podemos definir os eventos a seguir.

- duração exatamente igual a 1 ano: $\{365\} \subseteq \Omega$
- duração maior ou igual que 2 anos: $\{730, 731, 732, \ldots\} \subseteq \Omega$
- duração maior ou igual que 10 anos: $\{3650, 3651, 3652, \ldots\} \subseteq \Omega$
- duração exatamente igual a uma quantidade prima de dias: $\{2, 3, 5, 7, 11, \ldots\} \subseteq \Omega$

O último evento pode causar estranheza por não ser útil em situações práticas, mas é, ainda assim, um exemplo de evento definido em um espaço amostral. ●

Exemplo 1.1.17. O Exemplo 1.1.6 descreve um experimento onde selecionamos um ponto em um quadrado. Podemos definir, naquele espaço amostral, alguns eventos:

- os pontos que, além de estarem no quadrado, também ficam no interior do círculo unitário com centro na origem, $\Omega \cap \{(x, y) \ : \ x^2 + y^2 \leq 1\}$;
- os pontos sobre a reta $y = x$, $\Omega \cap \{(x, x) \ : \ x \in \mathbb{R}\}$;
- os pontos abaixo da reta $y = x$, $\Omega \cap \{(x, y) \ : \ x, y \in \mathbb{R}, \ x \geq y\}$. ●

Exemplo 1.1.18. No exemplo da pesquisa de opinião, alguns possíveis eventos são

- votos brancos e nulos, $\{\text{branco}, \text{nulo}\}$;
- votos em candidatos, $\Omega - \{\text{branco}, \text{nulo}\}$;
- votos em um subconjunto específico dos candidatos – por exemplo, $\{a, b\}$.

●

Definição 1.1.19 (eventos mutuamente excludentes)**.** Se A e B são eventos em um espaço amostral, dizemos que são **mutuamente excludentes** se $A \cap B = \emptyset$. ◄

Exemplo 1.1.20. No experimento onde se joga um dado, os eventos "≤ 2" e "≥ 4" são mutuamente excludentes, porque

$$\{1, 2\} \cap \{4, 5, 6\} = \emptyset.$$

Já os eventos "par" e "quadrado perfeito" não são, porque

$$\{2, 4, 6\} \cap \{1, 4\} = \{4\}.$$ ●

Exemplo 1.1.21. No Exemplo 1.1.8, em que o experimento consiste em selecionar cartas de baralho até surgir um naipe vermelho, temos como eventos, por exemplo,

- quantidade ímpar de tentativas, $\{V, PPV, PPPPV, PPPPPPV, \ldots\}$;
- no máximo tres tentativas, $\{V, PV, PPV\}$;
- exatamente tres tentativas, $\{PPV\}$;
- mais que 50 tentativas. ●

1.2 Probabilidade

Tentaremos definir a probabilidade de um evento. Suponha que o experimento seja repetido n vezes. Calculamos em quantas dessas vezes o evento A ocorre, e damos a essa quantidade o nome de "frequência de A", denotada $f(A)$. Poderíamos (mas veremos adiante que isso incorre em problemas) definir probabilidade como *frequência relativa*.

Definição 1.2.1 (frequência relativa)**.** Seja n a quantidade de repetições de um experimento. Seja A um dos resultados possíveis nesse experimento Se a frequêncica de A é $f(A)$, então a **frequência relativa** de A é

$$\frac{f(A)}{n}.$$ ◀

Esperamos que, à medida que aumentamos a quantidade de repetições, a frequência relativa, $f(A)/n$, estabilize em um número. É esse número que nos interessa. A definição para a probabilidade de resultados de experimentos usando frequência relativa, portanto, seria

$$r_A = \lim_{n \to \infty} \frac{f(A)}{n}.$$

Esta definição, no entanto, traz problemas. Em particular, não temos como garantir, em todos os casos, que o limite existe. Precisamos de uma definição que não encontre dificuldades técnicas. Observaremos, inicialmente, propriedades da definição usando frequência relativa, e *usaremos essas propriedades para construir outra definição.*

Proposição 1.2.1. Sejam A em B dois eventos em um espaço amostral.

i) $0 \le r_A \le 1$.

ii) $r_A = 1$ se e somente se A sempre ocorreu.

iii) $r_A = 0$ se e somente se A nunca ocorreu.

iv) se $A \cap B = \emptyset$ então $r_{A \cup B} = r_A + r_B$.

Demonstração. (i) Segue naturalmente da frequência de A é menor ou igual que n, mas maior que zero.

(ii) Como Ω todos os resultados possíveis, $r_\Omega = \frac{n}{n} = 1$.

(iii) Segue trivialmente de (ii)

(iv) Se $A \cap B = \emptyset$, então $r_A + r_B = r_{A \cup B}$, porque $r_A + r_B = \frac{f(A)}{n} + \frac{f(B)}{n} = \frac{f(A)+f(B)}{n} = r_{A \cup B}$ ■

Agora usamos essas propriedades para definir o conceito que realmente usaremos de probabilidade.

Definição 1.2.2 (probabilidade)**.** Sejam Ω um espaço amostral e $A \in \Omega$ um evento. A **probabilidade** do evento A é o número real $\Pr(A)$ tal que

i) $0 \leq \Pr(A) \leq 1$;

ii) $\Pr(\Omega) = 1$;

iii) $A \cap B = \emptyset$ implica que $\Pr(A \cup B) = \Pr(A) + \Pr(B)$;

iv) $\forall i \neq j, \;\; A_i \cap A_j = \emptyset$ implica que $\Pr\left[\bigcup A_i\right] = \sum \Pr(A_i)$. ◀

Exemplo 1.2.3. No experimento em que um dado honesto é jogado, escolhemos o espaço amostral $\Omega = \{1, 2, 3, 4, 5, 6\}$.

Definimos um evento para cada possível face,

$$\begin{array}{ll} A_1 = \{1\} & A_4 = \{4\} \\ A_2 = \{2\} & A_5 = \{5\} \\ A_3 = \{3\} & A_6 = \{6\} \end{array}$$

Como dissemos que o dado é honesto, atribuímos 1/6 de probabilidade a cada evento, porque este é um modelo que cremos ser muito próximo do experimento físico. ●

Exemplo 1.2.4. Os Exemplos 1.1.2 e 1.1.15 tratam do registro, feito por um técnico de um time de futebol, da frequência relativa de sucessos de um jogador em cobrança de pênaltis. Como o jogador obteve sucesso em 150 de 500 chutes, o treinador *atribuiu* as probabilidades $150/500 = 0.3$ para sucesso e 0.7 para falha.

Vale ressaltar que o técnico não “encontrou” a probabilidade de sucesso do jogador, nem “mostrou que ela é” 0.3. Ele realizou observações, e *determinou* que, para fins práticos, este número parece modelar bem o comportamento do jogador. ●

O Teorema 1.2.5, a seguir, estabelece que – como a intuição deve facilmente aceitar – a probabilidade do evento vazio (ou seja, de nenhum resultado ocorrer) é zero.

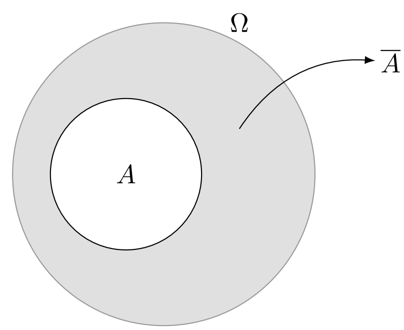
Teorema 1.2.5. Para quaisquer espaços amostrais e experimentos, $\Pr(\emptyset) = 0$.

Demonstração. Seja A um evento.

$$\begin{aligned} A &= A \cup \emptyset \\ \Pr(A) &= \Pr(A \cup \emptyset) = \Pr(A) + \Pr(\emptyset) - \Pr(A \cap \emptyset) \\ &= \Pr(A) + \Pr(\emptyset) \\ \Pr(\emptyset) &= 0. \end{aligned}$$ ■

O Teorema 1.2.6 trata da relação entre a probabilidade de um evento e a de seu complemento. Dizer que um evento A não ocorreu é o mesmo que dizer que “o resultado do experimento *não* é um elemento do conjunto A”, e portanto que o resultado pertence ao conjunto $\overline{A}$, complemento[2] de A.

[2] Com relação ao espaço amostral Ω: falamos de $\overline{A} = \Omega \setminus A$



Teorema 1.2.6. Seja A um evento em um espaço amostral. Então $\Pr(A) = 1 - \Pr(\overline{A})$.

Demonstração. Seja Ω um espaço amostral, e $A \subseteq \Omega$ um evento.

$$\begin{aligned} \Omega &= A \cup \overline{A} \\ \Pr(\Omega) &= \Pr(A \cup \overline{A}) \\ 1 &= \Pr(A \cup \overline{A}) \\ 1 &= \Pr(A) + \Pr(\overline{A}) \\ \Pr(A) &= 1 - \Pr(\overline{A}). \end{aligned}$$ ■

Exemplo 1.2.7. No experimento descrito no Exemplo 1.1.4, dois dados diferentes são jogados. Presumimos que cada resultado tem a mesma probabilidade de ocorrer, e para que isso seja verdade, a probabilidade de qualquer par específico de números é $1/36$, de forma que $\Pr(\Omega)$ seja um.

Seja A o evento "soma par e maior ou igual a oito".

Há dez possibilidades de soma maior que oito:

$$\Omega = \left\{ \begin{array}{r} (6,3),(6,4),(6,5),(6,6), \\ (5,4),(5,5),(5,6), \\ (4,5),(4,6), \\ (3,6) \end{array} \right\}. \tag{1.1}$$

Destes, somente quatro tem soma par: $(6,4),(6,6),(5,5),(4,6)$, logo probabilidade de A ocorrer é

$$\Pr(A) = \frac{4}{36} = \frac{1}{9}.$$

O evento $\overline{A}$ é "soma ímpar *ou* estritamente menor que oito", e

$$\Pr(\overline{A}) = 1 - \frac{1}{9} = \frac{8}{9}.$$

Como o espaço amostral é pequeno, poderíamos ter calculado $\Pr(\overline{A})$ diretamente: somamos

- os 26 resultados com soma menor que oito (36 menos os dez que identificamos na Equação 1.1);
- os seis com soma ímpar maior que oito, $(6,3),(6,5),(5,4),(5,6),(4,5),(3,6)$;

e dividimos pelo tamanho do espaço amostral (36). Com isso, temos novamente

$$\Pr(\overline{A}) = \frac{32}{36} = \frac{8}{9}.$$

●

O Teorema 1.2.8 é essencialmente o princípio da inclusão/exclusão para contagem de cardinalidade de conjuntos, mas traduzido para a linguagem da Probabilidade.

Teorema 1.2.8. Seja Ω um espaço amostral. Para todos os eventos $A, B, C \subseteq \Omega$,

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

Exemplo 1.2.9. No Exemplo 1.1.4, onde são jogados dois dados distintos, os eventos a seguir são definidos:

- A : soma par
- B : soma maior que oito
- C : valor igual nos dois dados

Suas probabilidades são

$$\begin{aligned}
\Pr(A) &= \frac{1}{2} \\
\Pr(B) &= \frac{5}{18} \\
\Pr(C) &= \frac{1}{6} \\
\Pr(A \cap B) &= \frac{4}{36} = \frac{1}{9} \\
\Pr(B \cap C) &= \frac{2}{36} = \frac{1}{18} \\
\Pr(A \cap C) &= \frac{1}{6} \\
\Pr(A \cap B \cap C) &= \frac{2}{36} = \frac{1}{18}.
\end{aligned}$$

A probabilidade do evento $A \cup B \cup C$ é

$$\begin{aligned}
\Pr(A \cup B \cup C) &= \Pr(A) + \Pr(B) + \Pr(C) \\
&- \Pr(A \cap B) - \Pr(A \cap C) - \Pr(B \cap C) \\
&+ \Pr(A \cap B \cap C) \\
&= \frac{1}{2} + \frac{5}{18} + \frac{1}{6} - \frac{1}{9} - \frac{1}{18} - \frac{1}{6} + \frac{1}{18} \\
&= \frac{2}{3}. \qquad \bullet
\end{aligned}$$

Do Teorema 1.2.8 podemos facilmente concluir o Corolário 1.2.1, cuja demonstração é pedida no Exercício 2.

Corolário 1.2.1 (Desigualdade de Boole)**.** Sejam $A_1, A_2, \ldots, A_n$ eventos em um espaço amostral. Então

$$\Pr\left(\bigcup_j A_j\right) \leq \sum_j \Pr(A_j).$$

Teorema 1.2.10. Se $A \subset B$, então $\Pr(A) \leq \Pr(B)$.

Demonstração.

$$\begin{aligned}
B &= A \cup (B \cap \overline{A}) \\
\Pr(B) &= \Pr(A) + \Pr(A \cap \overline{A}) \\
&\geq \Pr(A). \qquad \blacksquare
\end{aligned}$$

O Corolário 1.2.2 pode ser derivado das outras propriedades que estabelecemos – o que é pedido no Exercício 3.

Corolário 1.2.2. Se A e B são eventos em um mesmo espaço amostral, então $\Pr(A \cap B) \geq 1 - \Pr(\overline{A}) - \Pr(\overline{B})$.

1.3 Atribuindo probabilidades

1.3.1 Espaços amostrais finitos

Em muitas situações, presumimos que nenhum evento elementar é mais provável que outro: uma moeda não é viciada; uma carta de um baralho pode ser sorteada tanto quanto qualquer outra; ao retirar uma bola de uma urna, cada uma das bolas tem a mesma probabilidade de ser escolhida. Nestes casos, dizemos que estamos **presumindo a equiprobabilidade dos eventos elementares**.

Definição 1.3.1 (eventos equiprováveis em espaços finitos)**.** Seja Ω finito. Se, para todo $x \in \Omega$, $\Pr(x) = 1/|\Omega|$, então dizemos que os eventos em Ω são **equiprováveis**. ◀

Exemplo 1.3.2. No experimento do jogo de dados, no Exemplo 1.2.3, presumimos que o dado não é viciado – ou seja, que

$$\Pr(1) = \Pr(2) = \Pr(3) = \Pr(4) = \Pr(5) = \Pr(6) = \frac{1}{6}.$$ ●

Exemplo 1.3.3. Sorteamos um número natural qualquer. Qual é a probabilidade deste número ser divisível por sete?

O problema aparentemente exige um espaço amostral infinito (usaríamos $\Omega = \mathbb{N}$), mas é possível construir um modelo com espaço amostral finito.

Representaremos no espaço amostral, não o número sorteado, mas o resto de sua divisão por sete. Assim, $\Omega = \{0, 1, 2, 3, 4, 5, 6\}$. O resultado que nos interessa é o zero, e assim, a probabilidade do número ser divisível por sete é, como a intuição prevê, $1/7$.

Poderíamos também perguntar a probabilidade do número deixar resto maior que quatro quando dividido por sete, e neste caso estamos interessados no evento $\{5, 6\}$, que tem probabilidade $2/7$. ●

Resultados equiprováveis e contagem

Ao presumir resultados equiprováveis, estamos atribuindo a cada resultado x a mesma probabilidade,

$$\Pr(x) = \frac{1}{|\Omega|}.$$

Como para todos A e B disjuntos temos $\Pr(A \cup B) = \Pr(A) + \Pr(B)$, então para um evento $A = \{a_1, a_2, \ldots, a_k\}$,

$$\begin{aligned} \Pr(A) &= \Pr(\{a_1\} \cup \{a_2\} \cup \cdots \cup \{a_k\}) \\ &= \Pr(\{a_1\}) + \Pr(\{a_2\}) + \cdots + \Pr(\{a_k\}) \\ &= \frac{1}{|\Omega|} + \frac{1}{|\Omega|} + \cdots + \frac{1}{|\Omega|} && (k = |A| \text{ vezes}) \\ &= \frac{|A|}{|\Omega|}. \end{aligned}$$

Desta forma, em muitas situações, tendo presumido equiprobabilidade dos resultados, determinar a probabilidade de um evento consiste em determinar os tamanhos do espaço amostral e do evento. A seguir temos, portanto, uma breve discussão de Combinatória básica e suas aplicações em Probabilidade.

Definição 1.3.4 (produto cartesiano)**.** Sejam A e B dois conjuntos. O conjunto

$$A \times B = \{(a, b) \ : \ a \in A,\ b \in B\}$$

é o **produto cartesiano** de A e B.

A definição vale naturalmente para muitos conjuntos:

$$A_1 \times A_2 \times \cdots \times A_k = \{(a_1, a_2, \ldots, a_k) \ : \ a_i \in A_i\}.$$ ◀

Exemplo 1.3.5. Seja $A = \{1, 2, 3, 4, 5, 6\}$ os possíveis resultados ao jogar um dado. Quando jogamos dois dados distintos, o espaço amostral é $A \times A$. ●

Proposição 1.3.1. A quantidade de elementos em $A \times B$ é $|A||B|$. A quantidade de elementos em $A_1 \times A_2 \times \cdots \times A_k$ é $|A_1|\,|A_2| \cdots |A_k|$.

Em particular,

$$\underbrace{|A| \times |A| \times \cdots \times |A|}_{k\text{cópias de } A} = |A|^k.$$

Exemplo 1.3.6. Jogamos um dado e uma moeda. Os possíveis resultados para o dado são $D = \{1, 2, 3, 4, 5, 6\}$, e para a moeda são $M = \{a, o\}$, sendo que "a" representa cara e "o" representa coroa. O espaço amostral para o experimento poderia ser

$$\Omega = M \times D = \left\{ \begin{array}{c} (a,1),(a,2),(a,3),(a,4),(a,5),(a,6), \\ (o,1),(o,2),(o,3),(o,4),(o,5),(o,6) \end{array} \right\},$$

com tamanho $|M||D| = 12$.

Calculamos a probabilidade de obtermos uma cara e uma face par. O evento "cara e face par" é

$$X = \{(a, 2), (a, 4), (a, 6)\}$$

portanto a probabilidade deste evento é

$$\frac{|X|}{|\Omega|} = \frac{3}{12} = \frac{1}{4}.$$ ●

Exemplo 1.3.7. Se selecionarmos um número k entre 1 e 8800, qual é a probabilidade de k ser divisor de 8800?

Primeiro fatoramos $8800 = 2^3 5^2 11$. Há 8800 valores possíveis para k, portanto o espaço amostral é $\Omega = \{1, 2, \ldots, 8800\}$. Um número é divisor de 8800 se é igual a $2^a 5^b 11^c$, com

$$\begin{aligned} 0 \le a \le 3 \\ 0 \le b \le 2 \\ 0 \le c \le 1 \end{aligned}$$

Como podemos escolher a, b e c dentre três conjuntos $A = \{0, 1, 2, 3\}$, $B = \{0, 1, 2\}$ e $C = \{0, 1\}$, a quantidade de divisores de 8800 é

$$|A||B||C| = (4)(3)(2) = 24.$$

Assim, a probabilidade de um número escolhido ao acaso entre 1 e 8800 ser divisor de 8800 é

$$\frac{24}{8800} = \frac{3}{1100} \approx 0.0027.$$ ●

Exemplo 1.3.8. Se posicionarmos duas torres, uma branca e uma preta, em um tabuleiro de xadrez, qual é a probabilidade delas não se atacarem?

Calculamos a probabilidades das duas torres se atacarem: há 64 posições possíveis para a primeira torre; e depois que esta é escolhida, há 15 posições para a segunda, de forma que se ataquem. Assim, há (64)(15) maneiras possíveis de dispor duas torres se atacando mutuamente.

Se não nos importássemos com a possibilidade de ataque, poderíamos escolher 64 posições para a primeira torre, e 63 para a segunda. Assim, a probabilidade de duas torres posicionadas aleatoriamente se atacarem é

$$\frac{(64)(15)}{(64)(63)} = \frac{5}{21} \approx 0.24.$$

A probabilidade de duas torres posicionadas aleatoriamente *não* se atacarem é

$$1 - \frac{5}{21} = \frac{16}{21} \approx 0.76.$$

●

É interessante observar que a Proposição 1.3.1 vale inclusive para conjuntos unitários ou vazios. Se $A = \{x\}$, então $|A| \times |B| = |B|$, para todo B. Da mesma forma, $\emptyset \times B = \emptyset$ (porque não há par ordenado com primeiro elemento pertencente a $\emptyset$, porque o vazio não tem elementos), e $|\emptyset \times B| = 0$.

Proposição 1.3.2. Seja A um conjunto tal que $|A| = n$. A quantidade de tuplas ordenadas $(x_1, x_2, \ldots, x_k)$, com $k \leq n$ e onde $x_i \in A$, *sem elementos repetidos*, é

$$n(n-1)(n-2)\cdots(n-k+1).$$

Exemplo 1.3.9. Cinco prêmios diferentes serão sorteados para cinquenta inscritos em um evento. Qual é a probabilidade de determinadas cinco pessoas obterem os prêmios que desejam, presumindo que não há conflito entre suas vontades (cada um quer um prêmio diferente)?

O espaço amostral é composto pelas tuplas $(a_1, a_2, a_3, a_4, a_5)$, onde cada um dos elementos a_i é uma pessoa: a_1 ganha o primeiro prêmio, a_2 ganha o segundo, e assim por diante.

A quantidade de possíveis tuplas de ganhadores é (50)(49)(48)(47)(46).

O evento que procuramos contém uma única tupla nesse espaço amostral, e portanto tem probabilidade

$$\frac{1}{|\Omega|} = \frac{1}{(50)(49)(48)(47)(46)} = \frac{1}{254251200} \approx 0.0000000039331.$$

●

Corolário 1.3.1. O número de permutações dos elementos de um conjunto A é $|A|!$.

Exemplo 1.3.10. Um atacante conseguiu, ouvindo toques de teclado, determinar que a senha de um usuário tem dez caracteres. Também conseguiu, usando engenharia social, que o usuário comentasse que sua senha não tem caracteres repetidos. Além disso, é sabido que o sistema só permite números nas senhas.

Qual é a probabilidade desse atacante descobrir a senha, tentando uma vez ao acaso?

Cada caracter da senha pode ser um dentre 10 dígitos – mas a senha tem desse usuário dez posições, e não tem repetições. O espaço amostral é, portanto, o conjunto de permutações de $(0, 1, 2, \ldots, 9)$. Como o atacante procura uma única senha, a probabilidade de sucesso é

$$\frac{1}{|\Omega|} = \frac{1}{10!} = \frac{1}{3628800} \approx 0.00000027557. \qquad \bullet$$

Proposição 1.3.3. Seja A um conjunto tal que $|A| = n$. A quantidade de subconjuntos (que *não são ordenados e não admitem elementos repetidos*) $x_1, x_2, \ldots, x_k$, com $k \leq n$ e onde $x_i \in A$, é

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

O número $\binom{n}{k}$ é chamado de **combinação de** n **elementos, tomados** k **de cada vez**, ou de **coeficiente binomial**.

Exemplo 1.3.11. Uma loteria permite escolher quatro números entre 1 e 50; durante o sorteio, quatro números são sorteados, sem repetição (bolas numeradas são retiradas de uma urna, e não são devolvidas). O único prêmio pago é para quem acerta todos os números.

Qual é a probabilidade de uma aposta ser premiada?

O espaço amostral é composto de todos os conjuntos (e *não* tuplas, porque a ordem dos números no sorteio não é relevante) de quatro números entre 1 e 50:

$$\Omega = \Big\{ \; \{a_1, a_2, a_3, a_4\} \; : \; a_i \in \mathbb{Z}, \, 1 \leq a_i \leq 50 \; \Big\}.$$

Ou seja,

$$\Omega = \Big\{ \; \{1, 2, 3, 4\}, \{1, 2, 3, 5\}, \cdots, \{46, 47, 48, 49, 50\} \; \Big\}.$$

O tamanho deste conjunto de conjuntos é

$$\begin{aligned}
|\Omega| &= \binom{50}{4} \\
&= \frac{50!}{4!46!} \\
&= \frac{(50)(49)(48)(47)}{4!} \\
&= \frac{(50)(49)(48)(47)}{24} \\
&= (50)(49)(2)(47) \\
&= 230300.
\end{aligned}$$

A aposta é um desses conjuntos, logo a probabilidade de acerto é

$$\frac{1}{|\Omega|} = \frac{1}{230300} \approx 0.00000434. \qquad \bullet$$

Exemplo 1.3.12. No jogo de pôquer, um jogador tem uma *sequência* quando suas cinco cartas formam uma progressão numérica. Por exemplo, ($\heartsuit$3, $\clubsuit$4, $\heartsuit$5, $\diamondsuit$6, $\diamondsuit$7) é uma sequência. As cartas J, Q, K, A tem valor $11, 12, 13$ e 14, portanto devem aparecer em sequências após o dez, e nessa ordem.

Determinaremos a probabilidade de um jogador obter uma sequência, sendo que nenhum outro jogador recebeu cartas ainda (e portanto o baralho tem todas as cartas).

Um baralho tem 52 cartas. O número de conjuntos de cinco cartas diferentes é

$$|\Omega| = \binom{52}{5} = 2598960.$$

A primeira carta da sequência deve estar entre 2 e 10. Há 10 possibilidades para ela; as outras quatro cartas tem, cada uma, um único valor possivel, e há dez possíveis sequências apenas (uma começando com dois, uma começando com três, etc). No entanto, cada carta na sequência pode ser de um dentre quatro naipes diferentes. Denotamos por V o conjunto de possíveis valores para a primeira carta, e N_i o conjunto de naipes da i-ésima carta. Os tamanhos dos conjuntos são

$$\begin{aligned} |V| &= 10, \\ |N_1| &= 4, \\ |N_2| &= 4, \\ |N_3| &= 4, \\ |N_4| &= 4, \\ |N_5| &= 4, \end{aligned}$$

e o evento que queremos é

$$S = V \times N_1 \times N_2 \times N_3 \times N_4 \times N_5,$$

que tem tamanho $(10)(4)(4)(4)(4)(4) = 10240$. Assim,

$$\Pr(S) = \frac{|S|}{|\Omega|} = \frac{10240}{2598960} \approx 0.00394. \qquad \bullet$$

Adiante será necessário definir multiconjuntos.

Definição 1.3.13 (multiconjunto)**.** Um multiconjunto é um conjunto onde se admite repetição de elementos. ◀

Exemplo 1.3.14. $\{a, b, b, c, d, d, d\}$ é um multiconjunto, contendo um elemento a, dois bs, um c e três ds. ●

Exemplo 1.3.15. Os fatores primos de um número inteiro são naturalmente representados por multiconjuntos: $600 = 2^3(3)5^2$, e os fatores pretencem ao multiconjunto

$$\{2, 2, 2, 3, 5, 5\},$$

já incluída a multiplicidade de cada fator. ●

Proposição 1.3.4. Seja A um conjunto tal que $|A| = n$. A quantidade de submulticonjuntos (que *não são ordenados e admitem elementos repetidos*) $x_1, x_2, \ldots, x_k$, com $k \leq n$ e onde $x_i \in A$, é

$$\left(\!\!\binom{n}{k}\!\!\right) = \binom{n+k-1}{k}.$$

Exemplo 1.3.16. Jogamos três dados, e queremos saber a probabilidade de exatamente dois deles serem iguais a dois.

O evento que definimos é

$$A = \Big\{ \{1, 2, 2\}, \{2, 2, 3\}, \{2, 2, 4\}\{2, 2, 5\}, \{2, 2, 6\} \Big\}.$$

O tamanho do espaço amostral é

$$\begin{aligned} \left(\!\!\binom{6}{3}\!\!\right) &= \binom{6+3-1}{3} \\ &= \frac{8!}{3!5!} \\ &= \frac{(8)(7)(6)}{3!} \\ &= (8)(7) \\ &= 56. \end{aligned}$$

Assim,

$$\Pr(A) = \frac{|A|}{|\Omega|} = \frac{5}{56}.$$

●

1.3.2 Espaços amostrais infinitos e discretos

É possível atribuir probabilidades em espaços amostrais infinitos e discretos, mas é necessário cuidado para não violar as restrições na definição de probabilidade.

Como exemplo, tomamos o experimento onde aguardamos uma quantidade de dias até que um equipamento falhe. Se tentarmos atribuir uma mesma probabilidade positiva para "falha no i-ésimo dia" – valendo para todos os dias, a soma das probabilidades para todos os infinitos dias será $\sum_{i=0}^{\infty} \varepsilon$, que não converge, e portanto nào está entre zero e um. É possível atribuir probabilidades de forma que as restrições sejam respeitadas, *e que faça sentido como modelo de um fenômeno.* Por exemplo, podemos presumir que a probabilidade de falha do equipamento *em um dia qualquer* é $1/k$, ou ainda, que aumenta a cada dia. Nestes casos, chegaremos a uma soma que converge, poderemos calcular a probabilidade de falha depois de algum número finito de dias.

Exemplo 1.3.17. O Exemplo 1.1.8 descreve um experimento onde cartas de baralho são selecionadas, com repetição, até que seja obtida uma carta com naipe vermelho.

Calculamos a probabilidade de obtermos uma carta vermelha após tres ou menos tentativas. Há tres possibilidades de sucesso:

- sucesso na primeira tentativa $(1/2)$;
- falha na primeira $(1/2)$ *e* sucesso na segunda $(1/2)$, logo com probabilidade $1/4$;
- falha nas duas primeiras $(1/4)$ *e* sucesso na terceira $(1/2)$, portanto com probabilidade $1/8$.

$$p = \frac{1}{2} + \left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^3 = \frac{7}{8}.$$

A probabilidade de obtermos uma carta vermelha após mais que tres tentativas é

$$1 - p = \frac{1}{8}.$$ ●

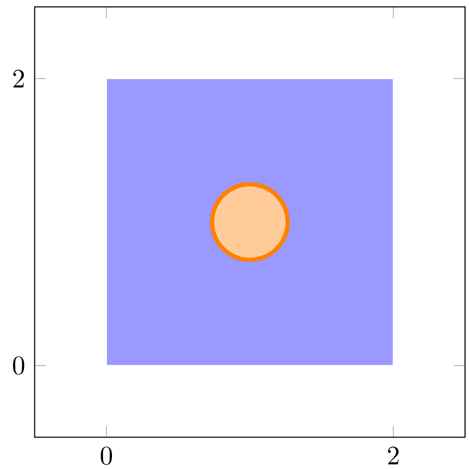
1.3.3 Espaços amostrais contínuos

Não podemos usar para espaços amostrais contínuos a mesma definição que usamos para espaços discretos. Isso porque, se cada elemento tiver uma probabilidade maior que zero, violaremos as restrições da definição de probabilidade (Definição 1.2.2), porque poderemos encontrar alguma partição de Ω em eventos cujas probabilidades somem mais que um. No entanto, a intuição nos propõe que deve ser possível atribuir probabilidades em espaços amostrais contínuos: se um ponto é escolhido ao acaso em um quadrado, é natural pensarmos em atribuir ao evento "na metade superior" a probabilidade $1/2$, e ao evento "na metade inferior" também a mesma probabilidade.

O tratamento rigoroso deste problema fica fora do escopo deste texto, mas podemos usar a seguinte abordagem: podemos atribuir a cada *resultado* em Ω a probabilidade zero, mas a cada *região* uma probabilidade maior que zero – de maneira que possamos calcular probabilidades como integrais. Se Ω é um intervalo, podemos atribuir probabilidades positivas a intervalos com tamanho maior que zero; se Ω está em $\mathbb{R}^2$, atribuímos probabilidades positivas a regiões com áreas maiores que zero (mas *não* a regiões com área zero!)

Exemplo 1.3.18. Um ponto (x, y) é escolhido ao acaso, com $x, y, \in [0, 2]$, e *presumimos equiprobabilidade:* todas as regiões de mesma área tem a mesma probabilidade de serem escolhidos.

Estamos interessados na probabilidade deste ponto cair dentro do círculo com centro em $(1, 1)$ e raio $1/2$.



Como presumimos equiprobabilidade, podemos atribuir probabilidades a regiões, de forma proporcional a área de cada uma. A probabilidade do ponto estar no quadrado, que é o espaço amostral inteiro, é um; a probabilidade de estar em um retângulo de área k é a mesma de estar em outro retângulo de área k. Já a probabilidade de ser um único ponto específico é zero.

Calculamos a área do quadrado Q entre $(0,0)$ e $(2,2)$, e a área da circunferencia C

$$
\begin{aligned}
A_Q &= 2^2 \\
A_C &= \pi \left(\frac{1}{2}\right)^2
\end{aligned}
$$

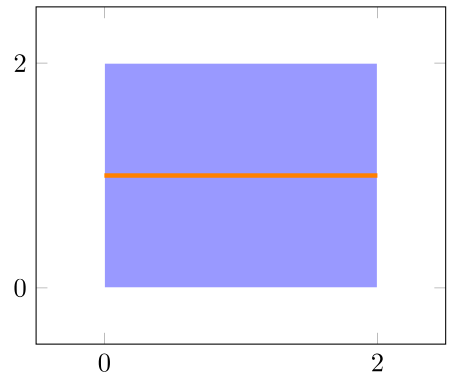
Então

$$
\begin{aligned}
\Pr(C) &= \frac{A_C}{A_Q} \\
&= \frac{\pi}{4}\frac{1}{4} \\
&= \frac{\pi}{16}.
\end{aligned}
$$

●

Exemplo 1.3.19. Um ponto (x,y) é escolhido ao acaso, com $x,y,\in[0,2]$, e *presumimos equiprobabilidade.*

Estamos interessados na probabilidade deste ponto cair exatamente no segmento de reta que vai de $(0,1)$ até $(2,1)$.



A probabilidade que buscamos é igual a zero, porque ela é dada pela *área* do segmento de reta dividida pela *área* do retângulo. Como o segmento de reta não tem área, a probabilidade do evento é $\frac{0}{4} = 0$. ●

Se o espaço amostral é contínuo, a probabilidade de um dentre dois pontos ocorrerem é zero. De maneira mais geral, *se o espaço amostral é contínuo – não enumerável – a probabilidade de qualquer subconjunto enumerável do espaço amostral é zero*

Exemplo 1.3.20. Sorteamos um número real entre zero e um. A probabilidade desse número ser igual a 0.5 é zero. A probabilidade desse número ser um dentre $0.1, 0.2, 0.3, 0.4, 0.5$ também é zero.

A probabilidade do número ser um dentre $10^{-1}, 10^{-2}, \ldots, 10^{-1000}$ é zero, porque o evento é um conjunto enumerável.

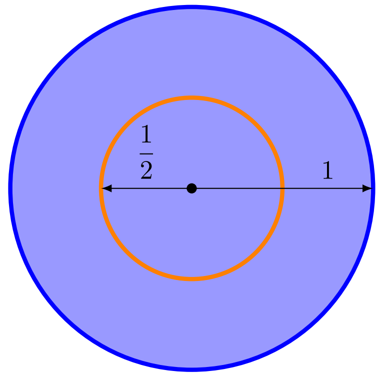
A probabilidade do número sorteado ser racional é, também, zero! Isso porque, mesmo havendo infinitos racionais entre zero e um, os racionais são enumeráveis, e o espaço amostral (os reais entre zero e um) não é. ●

Ao determinar probabilidade em geometria, dividimos o "tamanho" de eventos pelo "tamanho" do espaço amostral. No entanto, sempre comparamos distâncias com distâncias; áreas com áreas; e volumes com volumes. Não se pode comparar, por exemplo, área com distância. Uma maneira rápida e simples de perceber isto é através do seguinte exemplo: se um evento é um segmento de reta com comprimento $2m$, e o espaço amostral é uma área de $4m^2$, a divisão nos daria

$$\frac{2m}{4m^2} = \frac{1}{2m},$$

mas probabilidade não tem unidade de medida (é adimensional), e obtivemos algo com unidade "inverso de metro".

Exemplo 1.3.21. Sorteamos um ponto dentro de uma circunferência de raio um. A probabilidade desse ponto estar exatamente sobre a circunferência de raio 1/2 é zero.



Isso porque o espaço amostral é uma *área*, e o evento tem área zero (a borda da circunferência de raio 1/2 tem comprimento maior que zero, mas não área!) ●

Só faz sentido determinar probabilidades como razão entre áreas quando o espaço amostral está em $\mathbb{R}^2$. Se estivermos lidando com objetos em $\mathbb{R}^3$, com volume, passamos a definir probabilidade como o *volume* de um evento dividido pelo *volume* do espaço amostral. Objetos que tem área, mas não volume (como parte de um plano, ou uma reta) passam a ter probabilidade zero.

Exemplo 1.3.22. Um cilindro de raio r atravessa um cubo de raio l, de forma que o eixo do cilindro é paralelo a um dos lados do cubo. Se eu escolher um ponto qualquer no interior do cubo, qual é a probabilidade dele estar fora do cilindro?

O evento "fora do cilindro" é igual à área do cubo menos a do cilindro, logo a probabilidade é

$$\frac{l^3 - \pi l r^2}{l^3} = 1 - \pi \frac{r^2}{l^2}.$$ ●

Exemplo 1.3.23. Um ponto é sorteado no interior de uma esfera. A probabilidade desse ponto cair em um plano qualquer que tenha interseção com a esfera é zero, porque a probabilidade deve ser calculada dividindo o *volume* do evento pelo *volume* do espaço amostral. E o volume da interseção do plano com a esfera é zero. ●

Há um argumento diferente que ajuda a compreender porque as probabilidades de pontos individuais em $\mathbb{R}$ (ou planos em $\mathbb{R}^3$) devem ser zero.

Com o propósito de construir o argumento, propomos um jogo: um número é selecionado no intervalo $[0, 1]$, e o jogador deve adivinhá-lo. Como o número é real, adivinhá-lo significa adivinhar *todos* os seus dígitos. Note que todo número real tem infinitos dígitos após a vírgula – ainda que em alguns casos, eles sejam sempre zero depois de algum ponto.

A probabilidade do jogador acertar o primeiro dígito após a vírgula deve ser 1/10, porque há finitos dígitos (dez) e nenhum deles é mais provável que outro. Da mesma forma, para dois dígitos a probabilidade deve ser $1/10^2$, e para n dígitos, $1/10^n$.

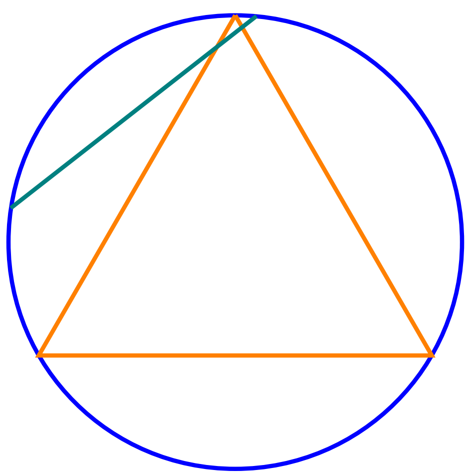
Agora suponha que tenhamos atribuído a probabilidade de acertar o número todo, algum $\varepsilon > 0$. Existe algum k tal que $1/10^k < \varepsilon$, e isso significa que a probabilidade de acertar k dígitos seria *menor* do que a probabilidade de acertar o número inteiro. Além de contrariar a intuição, isso violaria as regras da probabilidade, porque teríamos a probabilidade de um evento A menor que a de um evento $x \subseteq A$ – o que seria uma contradição.

Ainda assim, contrariando a intuição, *o resultado de um experimento como esse é um evento elementar com probabilidade zero*, sempre! (Porque todos tem probabilidade zero.)

1.4 O Paradoxo de Bertrand

Esta Seção trata de um problema conhecido como *Paradoxo de Bertrand*, descrito por Joseph Bertrand em 1889: dependendo do método para resolução, soluções diferentes podem ser encontradas. Por haver mais de uma solução obtida por métodos corretos, o problema foi chamdo por Henri Poincaré de "paradoxo".

Em uma circunfêrencia há, inscrito, um triângulo equilátero. Se selecionarmos equiprovavelmente uma das cordas na circunferência, qual é a probabilidade da corda ser mais longa que o lado do triângulo? Na figura a seguir, por exemplo, a corda é menor que o lado do triângulo.

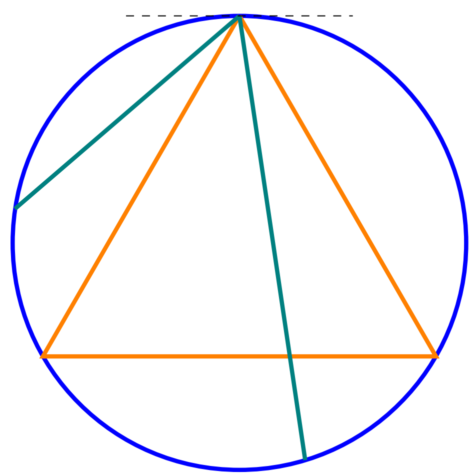


A seguir apresentamos três soluções diferentes para o problema. Em cada uma das soluções, selecionamos a corda usando um método diferente.

1. **Duas extremidades aleatórias:** escolhemos equiprovavelmente dois pontos na borda da circunferência; estes dois pontos serão as extremidades da corda.

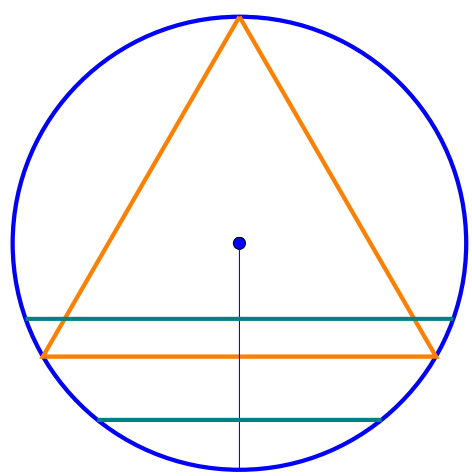
   Podemos rotacionar o triângulo dentro da circunferência sem modificr o resultado (rotacionar o triângulo não muda o comprimento de seus lados!)

– e portanto, podemos presumir que uma das extremidades da corda coincide com um dos vértices do triângulo. Ao escolher a segunda extremidade, estamos escolhendo um ângulo com a tangente no primeiro ponto.



Como os ângulos internos do triângulo são iguais a $\pi/3$, a probabilidade da corda sair dentro do triângulo e interceptar o lado oposto é $1/3$. Ou ainda, uma vez que uma das extermidades é escolhida, há somente $1/3$ da circunferência onde a outra pode ficar para que a corda seja maior que o lado do triângulo.

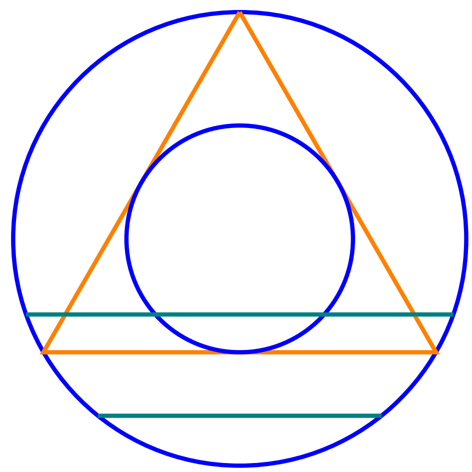
2. **Distância aleatória do centro:** escolha equiprovavelmente uma direção a partir do centro; depois, uma distância entre a corda e o centro.



A direção não influencia no tamanho da corda; observamos portanto a distância escolhida. Como o lado do triângulo divide o raio exatamente na metade, a probabilidade da corda ficar de cada um dos lados é $1/2$.

3. **Ponto médio aleatório:** escolhemos equiprovavelmente um ponto dentro da circunferência como ponto médio da corda; este ponto determina completamente a corda.

   Para a análise, definimos outra circunferência, desta vez inscrita no triângulo.



   A corda será mais longa que o lado do triângulo se o ponto médio estiver dentro da circunferência menor. Como o círculo menor tem raio com comprimento igual à metade do raio maior, a área da circunferencia menor é 1/4 da área da maior, e a probabilidade da corda ser maior que o lado do triângulo é 1.4.

As probabilidades encontradas são diferentes porque dependem de como a corda é "equiprovavelmente" selecionada. Ao tentar tomar uma dentre "todas as possíveis cordas", tivemos que caracterizar a corda de alguma forma: por suas extremidades, pela distância até o centro, ou pelo ponto médio. Acontece que estas formas de seleção não são equivalentes. Interessantemente, as três formas de seleção de corda são equivalentes a:

1. coordenadas polares das extremidades são escolhidas equivalentes;

2. coordenadas cartesianas do ponto médio são escolhidas equiprovavelmente;

3. coordenadas polares do ponto médio são escolhidas equiprovavelmente.

Até mesmo uma simples troca de sistema de coordenadas (que acontece entre os métodos 2 e 3) faz diferença.

Bertrand comentou que a dificuldade vem de haver diferentes maneiras de tomar amostras de um espaço amostral infinito, e como *o problema não especifica como escolher a corda*, ele não está completamente especificado. Cada solução apresentada, ao presumir um método de seleção, completa o enunciado do problema antes de resolvê-lo. São na verdade três problemas diferentes!

Henri Poincaré em 1912 e Edwin Jaynes em 1973 observaram que há uma restrição adicional que poderia ser imposta de maneira natural ao problema: uma solução deve ser válida mesmo que a circunferência e o triângulo passassem por transformações Euclideanas (rotação, translação e dilatação). Quando esta restrição adicional é imposta, somente a segunda solução permanece válida. Jaynes também mencionou que executou um experimento físico, atirando palitos em uma circunferência (cada palito define uma corda), chegando à mesma conclusão. Nem sempre, no entanto, é possível encontrar uma restrição natural como a que Poincaré e Jaynes identificaram para resolver ambiguidades na especificação de problemas.

1.5 Pequeno exemplo do Método Probabilístico

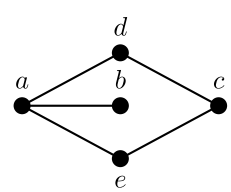
É possível usar Probabilidade para demonstrar fatos que não tem relação com não-determinismo. No Capítulo 4, mostraremos que é possível provar que a soma dos ângulos internos de um triângulo é igual a π, usando um argumento probabilístico – mesmo não havendo nada de aleatório ou não-determinístico a respeito desses ângulos! Esse tipo de raciocínio é chamado de *método probabilístico*, usado em Combinatória moderna.

Para um exemplo do método probabilístico, detalharemos uma demonstração de um fato básico da Teoria de Ramsey. Precisaremos de algumas definições: grafo, clique e conjunto independente, além da definição de número de Ramsey.

Definição 1.5.1 (grafo simples)**.** Um **grafo simples** G é um par (V, E), onde V é um conjunto de **vértices** e E é um conjunto de pares não-ordenados de vértices.

Um grafo é **completo** se cada vértice está conectado com todos os outros por arestas. O grafo completo com n vértices é denotado por K_n ◀

Exemplo 1.5.2. A figura a seguir mostra um grafo com cinco vértices.

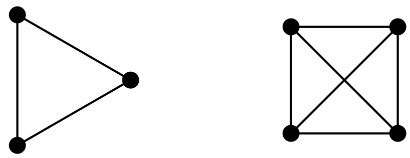


A descrição textual deste grafo é dada a seguir.

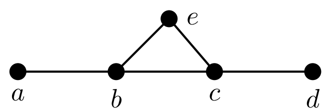
$$V = \{a, b, c, d, e\}$$
$$E = \left\{ \begin{array}{l} \{a, b\}, \{a, d\}, \{a, e\}, \\ \{d, c\}, \{e, c\} \end{array} \right\}.$$
●

Definição 1.5.3 (clique)**.** Uma **clique** em um grafo G é um subgrafo completo de G (ou seja, um subgrafo de G onde cada vértice está conectado com todos os outros). ◀

Exemplo 1.5.4. Os dois grafos a seguir são K_3 e K_4, grafos completos; quando estiverem presentes em outros grafos maiores, diremos que são cliques de tamanho 3 e 4.

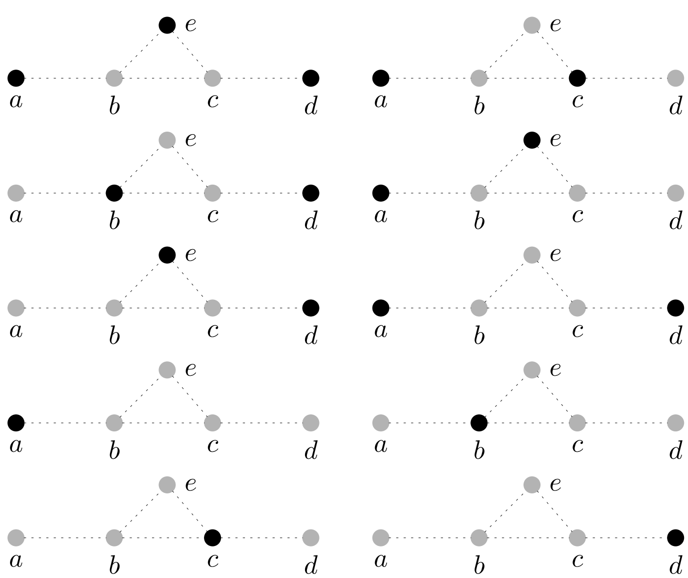


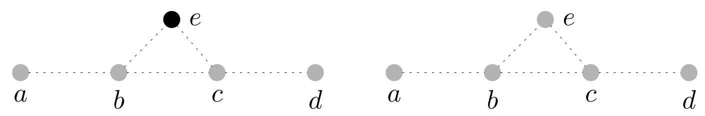
Exemplo 1.5.5. O grafo a seguir contém uma clique de tamanho 3, formada pelos vértices b, c, e; cinco cliques de tamanho dois, formadas pelos conjuntos de vértices $\{a, b\}$, $\{b, c\}$, $\{c, d\}$, $\{b, e\}$, $\{c, e\}$, e cinco cliques de tamanho um, $\{a\}$, $\{b\}$, $\{c\}$, $\{d\}$, $\{e\}$.



Definição 1.5.6 (conjunto independente)**.** Um **conjunto independente** em um grafo G é um subgrafo de G onde não há arestas. ◄

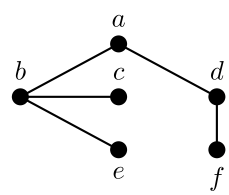
Exemplo 1.5.7. O grafo do Exemplo 1.5.5 tem alguns conjuntos independentes: $\{a, e, d\}$, $\{a, c\}$, $\{b, d\}$, $\{a, e\}$, $\{e, d\}$ e $\{a, d\}$, além, claro, dos vértices isolados $\{a\}$, $\{b\}$, $\{c\}$, $\{d\}$, $\{e\}$ e o conjunto vazio $\{\}$. As figuras a seguir ilustram os conjuntos independentes.





●

Exemplo 1.5.8. O seguinte grafo tem um conjunto independente de tamanho 4 (não há aresta entre os vértices a, c, e, f); alguns de tamanho 3 (a, c, e, por exemplo); e outros de tamanho 2 (b, d, por exemplo).



●

Definição 1.5.9. O **número de Ramsey, denotado** $R(k, t)$**,** é o menor n tal que um grafo com n vértices necessaraimente tem uma clique de tamanho k *ou* um conjunto independente de tamanho t.

Os números $R(k, k)$ são chamados de **números de Ramsey diagonais**. ◀

Exemplo 1.5.10. Claramente, $R(1, k) = R(k, 1) = 1$. Qualquer grafo, tendo um único vértice v, terá também uma clique de tamanho um (composta por v sozinho), e também um conjunto independente de tamanho um (o próprio v, novamente).

Outro fato simples é $R(2, n) = n$. Um grafo com n ou mais vértices tem necessariamente uma aresta (clique de tamanho dois) *ou* ele não tem aresta nenhuma. Mas, se não tem aresta nenhuma, deve então ter um conjunto independente de tamanho n (porque tem n ou mais vértices, e nenhuma aresta entre eles). Com menos de n vértices não podemos ter conjunto independente de tamanho n, e não podemos garantir que um grafo qualquer de tamanho n terá arestas, portanto o número de Ramsey $R(2, n)$ – a quantidade *mínima* de vértices que faz com que o grafo *necessariamente* tenha uma clique K_2 ou um conjunto independente de tamanho n, é n. ●

Exemplo 1.5.11. Outros números de Ramsey são mais difíceis de determinar. É possível provar, por exemplo, que $R(3, 3) = 6$ e $R(4, 4) = 18$. ●

Lema 1.5.1. Seja $k > 2$ um inteiro, e $n \leq 2^{k/2-1}$. Seja também G um grafo com n vértices. Então a probabilidade de G ter uma clique ou conjunto independente de tamanho n é estritamente menor que um.

Demonstração. Crie um grafo G com n vértices, incluindo arestas aleatoriamente (cada aresta tem probabilidade 1/2 de ser escolhida.

A probabilidade de um subgrafo de G com k vértices ser uma clique é

$$\underbrace{\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)\cdots\left(\frac{1}{2}\right)}_{\binom{k}{2}\text{vezes}},$$

porque cada aresta tem probabilidade $1/2$ de ser escolhida. O mesmo vale para a probabilidade de G ser um conjunto independente de tamanho k. Assim, a probabilidade de G ser um conjunto independente *ou* uma clique de tamanho k é

$$2\left(\frac{1}{2^{\binom{k}{2}}}\right) = 2^{1-\binom{k}{2}}.$$

A quantidade de subgrafos de G com k vértices é $\binom{n}{k}$, logo a probabilidade de haver alguma clique ou conjunto independente de tamanho k em G é

$$\begin{aligned}
\binom{n}{k}2^{1-\binom{k}{2}} &< n^k 2^{1-\binom{k}{2}} \\
&= \left(2^{\log_2 n}\right)^k 2^{1-\binom{k}{2}} \\
&= 2^{k\log_2 n}2^{1-\binom{k}{2}} \\
&= 2^{k\log_2 n+1-\binom{k}{2}} \\
&= 2^{k\log_2 n+1-(k(k-1)/2)}.
\end{aligned}$$

Verificamos para quais k este valor é estritamente menor que um:

$$\begin{aligned}
2^{k\log_2 n+1-(k(k-1)/2)} &< 1 \\
k\log_2 n + 1 - (k(k-1)/2) &< 0 && (\log_2(1) = 0) \\
k\log_2\left(2^{k/2-1}\right) + 1 &< (k(k-1)/2) && (\text{maior } n \text{ possível}) \\
\frac{k(k-2)}{2} + 1 &< (k(k-1)/2),
\end{aligned}$$

o que é verdade para todo $k > 2$. ■

Teorema 1.5.12. $R(k,k) > 2^{k/2-1}$.

Demonstração. Como a probabilidade calculada no Lema 1.5.1 é *estritamente* menor que um, deve necessariamente haver algum grafo com $n = 2^{k/2-1}$ vértices que não tem clique nem conjunto independente de tamanho k.

Quanto aos grafos com menos que n vértices, sem cliques ou conjuntos independentes de tamanho k, estes podem ser naturalmente obtidos dos grafos de tamanho n que também são livres dessas cliques e conjuntos independentes, por remoção de vértices.

Como não garantimos que grafos com *mais* que $2^{k/2-1}$ vértices *tem necessariamente* cliques ou conjuntos independentes de tamanho k, podemos dizer somente que $R(k,k)$ é *maior* que $2^{k/2-1}$ (mas não igual). ■

Exercícios

Ex. 1 — Se A e B são eventos mutuamente excludentes, então $\overline{A}$ e $\overline{B}$ também são?

Ex. 2 — Prove o Corolário 1.2.1.

Ex. 3 — Prove o Corolário 1.2.2.

Ex. 4 — Em uma urna há 5 bolas brancas e 7 bolas pretas. Sorteiam-se 3 bolas. Qual é a probabilidade de serem sorteadas mais bolas pretas do que brancas?

Ex. 5 — Sortearei equiprovavelmente três números inteiros, a, b e c, todos entre zero e 999, e calcularei seu produto, $k = abc$. Calcule:

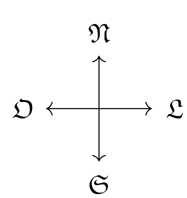
1. $\Pr[k = 0]$.
2. $\Pr[k = 1]$.
3. $\Pr[k = 2n]$ (probabilidade de k ser par).
4. $\Pr[k = 3n]$ (probabilidade de k ser divisível por 3).

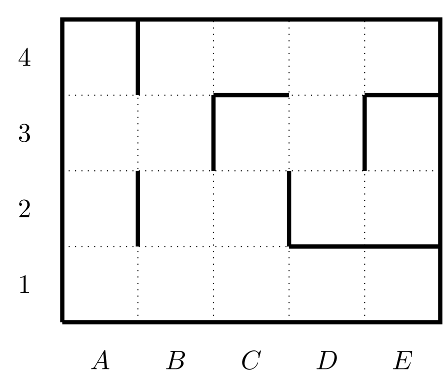
Ex. 6 — No Exercício 5, sorteamos um k, que pode estar entre 0 e 999^3. Usando aquele método, estaremos escolhendo todos os valores possíveis equiprovavelmente?

Ex. 7 — Sorteamos uma matriz 2×2, da seguinte maneira: cada valor na matriz vale 1 ou 0 com probabilidade 1/2.

1. Qual é a probabilidade desta martiz não ter inversa?
2. E se a probabilidade de cada valor ser 1 for 1/3?
3. E se a matriz for 3×3?

Ex. 8 — Um robô está em um andar de um edifício, representado como um labirinto. O robô consegue se mover em quatro direções (N, S, L, O), e tem sensores que ajudam a identificar paredes, também nas quatro direções. Ele tem o mapa do andar, mas não sabe onde está.





Se o robô usa o sensor na direção Norte e percebe que há uma parede, qual é a probabilidade dele estar em alguma local com coordenada numérica maior ou igual a 3 (ou seja, de estar na metade superior do mapa)?

Ex. 9 — Em um lote de vinte produtos há dois com defeito. Podemos tratá-los como distintos, porque todos tem número de série. Selecionamos dois deles ao acaso, sem reposição. Calcule a probabilidade de termos obtido

1. nenhum deles com defeito;
2. pelo menos um com defeito;
3. os dois com defeito.

Ex. 10 — Se dispusermos oito torres aleatoriamente em um tabuleiro de xadrez, qual é a probabilidade de nenhuma delas atacar alguma outra?

Ex. 11 — Se posicionarmos duas rainhas aleatoriamente em um tabuleiro de xadrez, qual é a probabilidade delas não se atacarem?

Ex. 12 — [**Este raciocínio está, de fato, nos fundamentos da Criptografia contemporânea**]
Um sistema informatizado está protegido por uma chave secreta, que é uma sequência de 128 bits. Um atacante não autorizado teria que adivinhar os 128 bits corretos para conseguir acesso ao sistema.

1. Determine a probabilidade de um atacante conseguir acesso ao sistema. Presuma que, não sabendo a chave, ele escolherá uma completamente ao acaso.
2. Se decidirmos que a chave atual não é segura, podemos escolher algum outro número n para o tamanho da chave. Determine a função que dá a probabilidade de sucesso de um atacante, tendo n como parâmetro (ou seja, $f(n) = \Pr(A)$, onde A é o evento "ataque com sucesso".

Ex. 13 — Um banco exige, para acesso em caixas eletrônicos, que o cliente escolha corretamente três códigos de um conjunto de vinte possíveis códigos. Um analista do banco decidiu que a sequência de códigos de um cliente não pode ter códigos repetidos. Determine a probabilidade de um atacante descobrir o código de um cliente tentando uma sequência ao acaso, (i) sem a restrição imposta pelo analista e (ii) com a restrição.

Ex. 14 — Refaça o Exemplo 1.3.10:

1. trocando o sistema que permite somente números por um sistema que também permite letras (sem distinção entre maiúsculas e minúsculas);
2. mantendo somente números nas senhas possíveis, mas permitindo repetições (ou seja, presuma que o atacante não conseguiu obter essa informação do usuário);

Ex. 15 — Selecionaremos aleatoriamente k números inteiros distintos entre 1 e n. Qual é a probabilidade do maior número ser igual a m (com $m \leq n$)?

Ex. 16 — Um usuário esqueceu o último dígito de sua senha para acesso a um sistema. Ele tentará dígitos diferentes, escolhidos ao acaso, mas só pode tentar três vezes. Qual é sua probabilidade de sucesso?

Ex. 17 — (AHSME[3] 1990) Sorteie equiprovavelmente um número a entre 1 e 100; depois sorteie um número b, também equiprovavelmente entre 1 e 100. Qual é a probabilidade do dígito menos significativo de $3^a 7^b$ ser igual a oito?

Ex. 18 — Seleciono aleatoriamente um ponto em uma circunferência.

1. Qual é a probabilidade do ponto estar mais próximo do centro que da borda?
2. Refaça o exercício, com uma esfera ao invés de circunferência.

Ex. 19 — Ao programar um jogo eletrônico de futebol, darei três opções ao usuário: (i) que ele escolha a posição dos seus jogadores; (ii) que as posições sejam dadas por algum esquema pré-determinado; ou (iii) que os jogadores sejam dispostos aleatoriamente no seu lado do campo, sendo que somente o goleiro é posicoinado dentro da pequena área (os outros ficam fora dela).

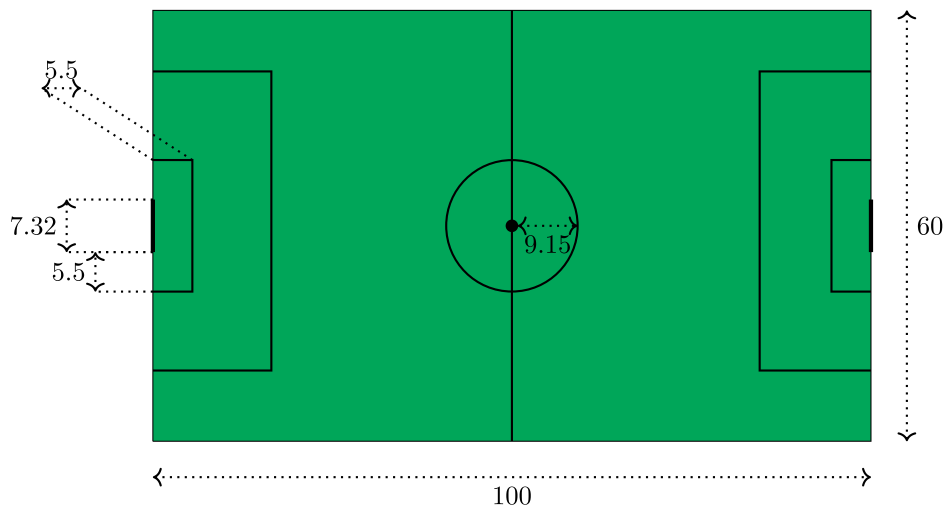
Se o jogador escolher a opção (iii), e eu tratar os jogadores como se fossem pontos,

1. qual é a probabilidade de pelo menos um jogador ser posto dentro da área central?
2. pelo menos dois jogadores dentro da área central?
3. e para n jogadores?

   Para referência[4]:

[3] Americal High School Mathematics Examination

[4] As regras oficiais permitem que a largura do campo fique entre $45m$ e $90m$, e que o comprimento fique entre $90m$ e $120m$.



- A área central é uma circunferência com raio igual a $9.15m$;
- A largura do campo é $60m$;
- O comprimeto do campo é $100m$;
- A distância entre as traves do gol é de $7.32m$;
- A distância da trave, indo na direção da lateral, até o final da pequena área, é $5.5m$;
- A distância da trave, indo para a frente, até o final da pequena área, também é $5.5m$

Ex. 20 — Refaça o Exercício 19 considerando que cada jogador ocupará uma área equivalente a $50cm^2$ na tela, e que não deve ser possível dois jogadores ocuparem o mesmo espaço. Ignore a possibilidade de um jogador ocupar uma região que cruze a fronteira da área central. Suponha também que as áreas dos jogadores podem ser "arranjadas" de forma flexível[5] (eles não tem forma fixa).

Ex. 21 — Um muro de $20m$ se estende de um ponto A até C, passando por B, que fica a $5m$ de A.

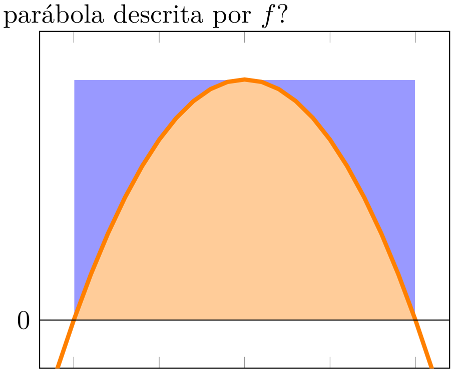
Uma câmera de vigilância fica continuamente se movendo, de forma que seu foco vai de A até C e de volta para C, com velocidade uniforme de $3m/s$.
Um robô invasor será jogado por cima do muro, no ponto B, e tentará chegar até C sem ser notado; este robô anda com velocidade de $2m/s$.
Qual é a probabilidade do robô ser pego pela câmera?

[5] Se, por exemplo, exigíssemos que os jogadores tivessem forma de circunferência,o problema ficaria diferente, e mais difícil.

Ex. 22 — Sorteamos três números reais entre zero e dez. Qual é a probabilidade da soma dos quadrados dos números ser menor que 10?

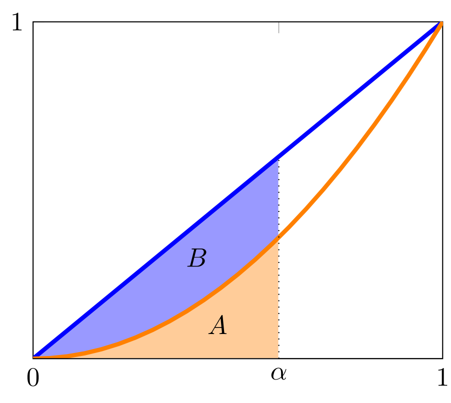
Ex. 23 — Se há 23 pessoas em uma sala, qual é a probabilidade de 2 delas ou mais fazerem aniversário na mesma data? (Ignore anos bissextos, e presuma[6] que nascimentos são distribuídos equiprovavelmente nos dias do ano.)

Ex. 24 — Seja $f(x) = ax^2 + bx + c$ uma função quadrática com duas raízes reais, e $a < 0$. Seja R o menor retângulo que contém a parte positiva do gráfico de f. Se escolhermos um ponto ao acaso neste retângulo, qual a probabilidade dele estar abaixo da parábola descrita por f?



Ex. 25 — Um número real α é escolhido uniformemente no intervalo $[0, 1]$. Considere duas funções reais, x e x^2. Sejam:

- A a área formada pelos pontos com coordenada vertical acima de zero e abaixo de x^2; e coordenada horizontal à direita do zero e à esquerda de α.
- B a área formada pelos pontos com coordenada vertical acima de x^2 e abaixo de x; e coordenada horizontal à direita do zero e à esquerda de α.



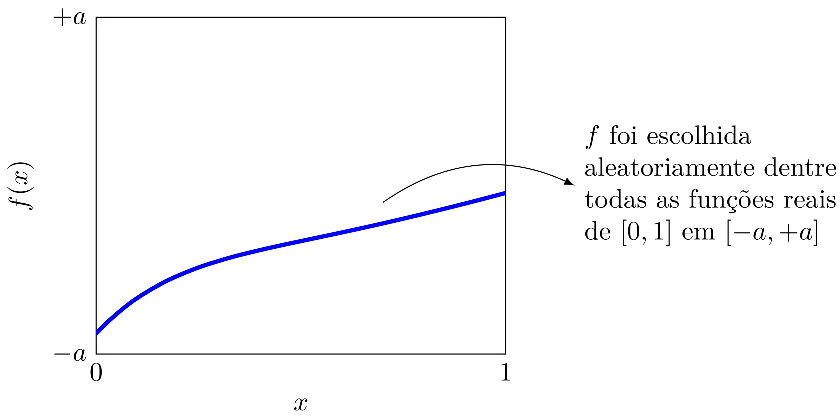
Determine $\Pr[A > B]$.

[6] Incorretamente! Na verdade, há meses com mais e com menos nascimentos.

Ex. 26 — Demonstre a Proposição 2.2.2.

Ex. 27 — Suponha que 75 pontos sejam escolhidos aleatoriamente no plano, de forma que não haja, dentre estes pontos, conjuntos de três pontos coplanares. Prove que, se esclhermos três desses pontos ao acaso, a probabilidade deles formarem um triângulo acutângulo é *no máximo* 0.7.

Ex. 28 — (Isabella[7]) Selecione equiprovavelmente uma função qualquer com domínio $[0, 1]$ e contradomínio $[-a, +a]$, com $a \in \mathbb{R}$. Qual é a probabilidade dessa função ser crescente?



Dica: inspire-se na soma de Riemann.

[7]Adaptado. A pergunta feita pela Isabella em aula é mais difícil (ela perguntou de uma função qualquer; neste exercício, a função é limitada).

Capítulo 2

Probabilidade Condicional e Independência

2.1 Probabilidade Condicional

Antes de definir probabilidade condicional, damos dois exemplos.

Exemplo 2.1.1. No Exemplo 1.1.4, jogamos dois dados diferentes, um vermelho e um branco. Podemos definir naquele espaço amostral os eventos:

- A: "dado vermelho par"
- B: "dado branco ímpar"
- C: "soma dos dois números é maior que 10"

Claramente,

$$\begin{aligned}
\Pr(A) &= \frac{1}{2}, \\
\Pr(B) &= \frac{1}{2}, \\
\Pr(C) &= \frac{3}{36} = \frac{1}{12}.
\end{aligned}$$

Ao calcular $\Pr(C)$, observamos que uma soma maior que dez só é possível com resultados $(5,6), (6,5), (6,6)$ – tres possibilidades no espaço amostral com 36 elementos.

Agora tentamos calcular a probabilidade de C, *dado que A ocorreu.* Como presumimos que A ocorreu, começamos descartando $\Omega \setminus A$, e chamamos A de *espaço amostral reduzido.*

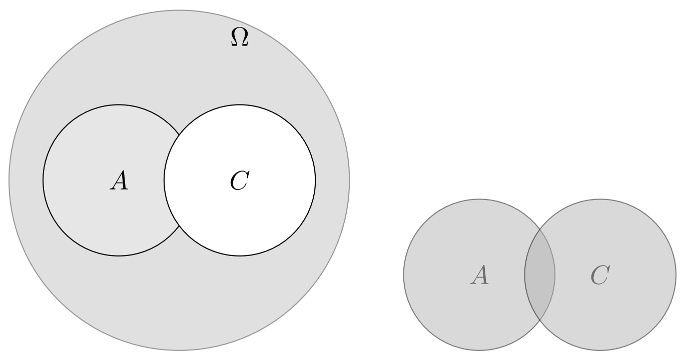
$$A = \left\{ \begin{array}{c} (2,1),(2,2),(2,3),(2,4),(2,5),(2,6), \\ (4,1),(4,2),(4,3),(4,4),(4,5),(4,6), \\ (6,1),(6,2),(6,3),(6,4),(6,5),(6,6) \end{array} \right\}$$

Mas neste espaço amostral reduzido (com 18 elementos) temos duas possibilidades, $(6,5),(6,6)$, com o primeiro dado par. Logo,

$$\Pr(C|A) = \frac{2}{18} = \frac{1}{9},$$

diferente de $\Pr(C)$.

As figuras a seguir ilustram a probabilidade de C (com relação a Ω) e $\Pr(C|A)$ onde tratamos A como espaço amostral reduzido.



A partir da ilustração, a intuição indica que o que queremos é, *para espaços amostrais finitos*, $(|A \cap C|)/|A|$ – mas deveremos estabelecer isto mais adiante.

●

Se repetirmos o experimento do exemplo 2.1.1, poderemos calcular a frequência relativa dos eventos. Nossa definição de probabilidade foi construída a partir das propriedades da frequência relativa, e faremos o mesmo para chegar à definição de probabilidade condicional.

Suponha que o experimento do exemplo 2.1.1 seja repetido n vezes. Sabemos que para determinar a frequência relativa dos eventos, basta calcular o número de vezes que ocorrem e dividir pelo número de repetições do experimento.

Para calcular a probabilidade de C, *dado que A aconteceu*, olhamos para as repetições do experimento, descartamos os casos em que A não acontece, e temos um novo espaço amostral *(onde A sempre ocorre, portanto o evento "A e C" é o mesmo que "C")*. Neste espaço amostral, contabilizamos as ocorrências de "A e C", e dividimos por $|A|$:

$$\begin{aligned}\frac{f(C \cap A)}{f(A)} &= \frac{f(C \cap A)}{f(A)} \left(\frac{n}{n}\right) \\ &= \frac{f(C \cap A)}{n} \frac{n}{f(A)}.\end{aligned}$$

Quando $n \to \infty$, denotando por "$r_{C|A}$" a frequência relativa de C dado que A

ocorreu,

$$\begin{aligned} r_{C|A} &= \lim_{n\to\infty} \frac{f(C\cap A)}{n}\frac{n}{f(A)} \\ &= r_{C\cap A}\,\frac{1}{r_A} \\ &= \frac{r_{C\cap A}}{r_A}. \end{aligned}$$

Definiremos, portanto, probabilidade condicional de B, dado que A cocorreu, como a razão entre $\Pr(A \cap B)$ e $\Pr(A)$

Definição 2.1.2 (probabilidade condicional)**.** Sejam A e B dois eventos em um mesmo espaço amostral. A probabilidade do evento B ter ocorrido, dado que A ocorreu, é chamada de **probabilidade condicional** de A dado B, denotada $\Pr(B|A)$, e definida a seguir.

$$\Pr(B|A) = \frac{\Pr(A\cap B)}{\Pr(A)},$$

desde que $P(A) \neq 0$. ◀

Definimos $\Pr(B|A)$ usando as *probabilidades* de A e de $A \cap B$, e não as cardinalidades dos eventos. A definição usando cardinalidade, $\Pr(B|A) = \frac{|A\cap B|}{|A|}$, funciona para espaços amostrais finitos, mas não para os infinitos (discretos ou contínuos), porque as cardinalidades podem ser infinitas. Ainda assim, observamos que, *em espaços finitos*,

$$\begin{aligned} \Pr(B|A) &= \frac{\Pr(A\cap B)}{\Pr(A)} \\ &= \Pr(A\cap B)\frac{1}{\Pr(A)} \\ &= \left(\frac{|A\cap B|}{|\Omega|}\right)\left(\frac{|\Omega|}{|A|}\right) \\ &= \frac{|A\cap B|}{|A|}. \end{aligned}$$

Exemplo 2.1.3. Em um lote com 80 peças há 20 com defeito. Se selecionarmos duas, uma de cada vez,

- a probabilidade da primeira ser defeituosa é

$$\Pr(A) = \frac{20}{80} = \frac{1}{5};$$

- a probabilidade da segunda peça ser defeituosa, *dado que a primeira já foi selecionada, e tem defeito*, é

$$\Pr(B|A) = \frac{19}{99} \neq \frac{1}{5}.$$

Ao calcular a segunda probabilidade, $\Pr(B|A)$, tínhamos dezenove itens defeituosos restantes (porque o primeiro já era um dos itens com defeito), dentre noventa e nove (porque um já foi selecionado). ●

Exemplo 2.1.4. Suponha que queiramos determinar a probabilidade de, ao selecionar duas cartas de um baralho, as duas serem de naipe vermelho ($\diamondsuit$ ou $\heartsuit$). Denotaremos

- $p =$ probabilidade do naipe da primeira carta ser vermelho.
- $q =$ probabilidade do naipe da segunda carta ser vermelho.

As probabilidades serão diferentes, dependendo de retirarmos as cartas com ou sem reposição.

Com reposição, nos dois casos retiramos alguma carta de um baralho completo, e metade das cartas tem naipe vermelho, logo $p = 1/2$ e $q = 1/2$.

Sem reposição, a probabilidade para a primeira carta é igual à calculada para o caso com reposição,

$$p = \frac{26}{52} = \frac{1}{2}.$$

No entanto, ao retirarmos a segunda carta, temos dois fatos importantes:

- o baralho tem uma carta a menos
- a quantidade de cartas vermelhas que permaneceram no baralho depende da primeira carta ter ou não o naipe vermelho

Se a primeira carta tem o naipe vermelho, sobram 25 cartas vermelhas no baralho, logo

$$q = \frac{25}{51} = 0.49019$$

Se a primeira carta era preta, então sobram todas as 26 vermelhas no baralho, portanto

$$q = \frac{26}{51} = 0.50980$$

O espaço amostral é

$$\Omega = \left\{ \begin{array}{l} (A\clubsuit, A\clubsuit), (A\clubsuit, A\spadesuit), (A\clubsuit, A\heartsuit), (A\clubsuit, A\diamondsuit), \\ \qquad \vdots \\ (K\diamondsuit, K\clubsuit), (K\diamondsuit, K\spadesuit), (K\diamondsuit, K\heartsuit), (K\diamondsuit, K\diamondsuit) \end{array} \right\}.$$

O evento “primeiro naipe vermelho” é A, contendo os pares ordenados onde o naipe do primeiro elemento é vermelho.

$$\Pr(A) = \frac{26}{52} = \frac{1}{2}.$$

O evento "segundo naipe vermelho" é B, contendo os pares ordenados onde o naipe do segundo elemento é vermelho.

$$\Pr(B) = \frac{26}{52} = \frac{1}{2}.$$

Se quisermos determinar a probabilidade do segundo naipe ser vermelho, *dado que o primeiro é*,

- o novo espaço amostral contém somente os eventos com o primeiro naipe vermelho – ou seja, é igual a A. Dizemos que este é o *espaço amostral reduzido*;
- o evento que queremos é aquele em que os *dois* naipes são vermelhos (o segundo vermelho, dado que o primeiro é). Este evento é $A \cap B$. ●

O próximo exemplo ilustra um ponto bastante importante: a probabilidade de um evento depende de detalhes da definição do experimento.

Exemplo 2.1.5. Alguém joga duas moedas diferentes (uma de $0,50 e uma de $1) e nos diz que "o resultado, para *pelo menos uma delas*, é cara". Dado que sabemos isso, qual é a probabilidade das duas moedas terem caído com a face "cara" para cima?

Há quatro resultados possíveis, coroa/coroa, coroa/cara, cara/coroa, cara/cara. Sabemos que um deles (coroa/coroa) é impossível, portanto a probabilidade é 1/3.

E se nos for dito que "para a moeda de $0,50, o resultado é cara"?

Esta situação é diferente da anterior. As moedas são diferentes, e sabemos o resultado para uma delas. *A resposta à pergunta feita só depende da outra moeda*, para a qual só há duas possibilidades, cara ou coroa. A probabilidade é, portanto, 1/2.

As duas probabilidades são diferentes, porque os experimentos são diferentes, apesar de parecerem ser iguais. ●

A definição de probabilidade condicional está de acordo com a definição de probabilidade (ou seja, da maneir aque definimos, "probabilidade condicional" é, também, probabilidade):

- $0 \leq \Pr(A|B) \leq 1$;
- $\Pr(\Omega|A) = 1$;
- Se $A_1 \cap A_2 = \emptyset$, então $\Pr(A_1 \cup A_2|B) = \Pr(A_1|B) + \Pr(A_2|B)$;
- Se $A_i \cap A_j = \emptyset$. então $\Pr(\cup A_i|B) = \sum \Pr(A_i|B)$.

2.2 Independência

Em geral, $\Pr(B) \neq \Pr(B|A)$, a não ser quando os eventos A e B são independentes.

Definição 2.2.1 (eventos independentes)**.** Dois eventos $A, B \subseteq \Omega$ são **independentes** se e somente se $\Pr(A \cap B) = \Pr(A)\Pr(B)$. ◀

Exemplo 2.2.2. Jogamos dois dados distintos e registramos o resultado.

Sejam A B, C os eventos:

- A = o valor do primeiro dado é par;
- B = o valor do segundo dado é par;
- C = o valor do primeiro dado é maior que o do segundo.

Claramente,

$$\begin{aligned}\Pr(A) &= \frac{1}{2},\\ \Pr(B) &= \frac{1}{2},\\ \Pr(C) &= \frac{15}{36} = \frac{5}{12}.\end{aligned}$$

$A \cap B = \{2,4,6\} \times \{2,4,6\}$, e portanto tem 9 elementos. Assim,

$$\Pr(A \cap B) = \frac{9}{36} = \frac{1}{4}.$$

Vemos que

$$\Pr(A)\Pr(B) = \left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = \Pr(A \cap B),$$

e os eventos A e B são independentes.

Há sete casos em que o primeiro valor é par e maior que o segundo, logo

$$\Pr(A \cap C) = \frac{7}{36}.$$

Mas como

$$\Pr(A)\Pr(C) = \left(\frac{1}{2}\right)\left(\frac{5}{12}\right) = \frac{5}{24} \neq \Pr(A \cap C),$$

os eventos A e C não são independentes. ●

Exemplo 2.2.3. O Exemplo 1.3.6 descreve um experimento em que uma moeda e um dado são jogados, e pede a probabilidade de se obter cara na moeda e uma face par no dado.

Se *presumirmos* que os lançamentos da moeda e do dado são independentes (o que parece bastante razoável), podemos calcular as probabilidades individualmente:

$$\begin{aligned}\Pr(CARA) &= \frac{1}{2}\\ \Pr(PAR) &= \frac{1}{2},\end{aligned}$$

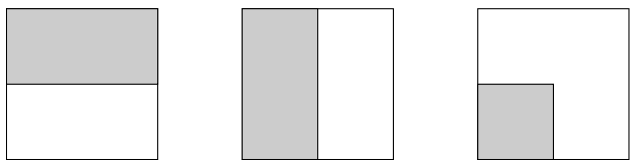
portanto

$$\Pr(CARA \cap PAR) = \left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = \frac{1}{4}.$$ ●

Exemplo 2.2.4. Escolhemos ao acaso um ponto no quadrado entre $(0,0)$ e $(1,1)$. O espaço amostral é $\Omega = \{(x,y) \ : \ x,y \in [0,1]\}$. Definimos tres eventos, A contendo os pontos na metade superior; B contendo os eventos na metade da esquerda; e C contendo os pontos no quadrado entre $(0,0)$ e $(0.5, 0.5)$:

$$
\begin{aligned}
A &= \{(x,y) \ : \ x \in [0,1], y \in [0.5,1]\},\\
B &= \{(x,y) \ : \ x \in [0,0.5], y \in [0,1]\},\\
C &= \{(x,y) \ : \ x \in [0,0.5], y \in [0,0.5]\}.
\end{aligned}
$$

Podemos determinar as probabilidades por simples inspeção das figuras.



A $\qquad$ B $\qquad$ C

$$
\Pr(A) = \frac{1}{2} \qquad \Pr(B) = \frac{1}{2} \qquad \Pr(C) = \frac{1}{4}
$$

Agora verificamos:

$$
\begin{array}{ccc}
\Pr(A \cap B) = \frac{1}{4}, & \Pr(A \cap C) = 0, & \Pr(B \cap C) = \frac{1}{4},\\
\Pr(A)\Pr(B) = \frac{1}{4}, & \Pr(A)\Pr(C) = \frac{1}{8}, & \Pr(B)\Pr(C) = \frac{1}{8}.
\end{array}
$$

Os eventos A e B são independentes; já os outros pares de eventos, A com C e B com C, *não são independentes.* ●

Proposição 2.2.1. *A e B são independentes se e somente se*

$$
\begin{aligned}
P(A|B) &= \Pr(A), \ \textit{e}\\
P(B|A) &= \Pr(B).
\end{aligned}
$$

Exemplo 2.2.5. No exemplo 2.2.4, temos

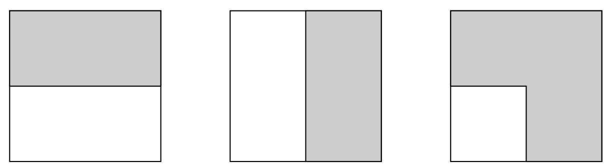
$$
\begin{array}{cccc}
\Pr(A|B) = \frac{1}{2}, & \Pr(B|A) = \frac{1}{2}, & \Pr(A|C) = 0, & \Pr(B|C) = 1,\\
\Pr(A) = \frac{1}{2}, & \Pr(B) = \frac{1}{2}, & \Pr(C|A) = 0. & \Pr(C|B) = \frac{1}{2}.
\end{array}
$$

Novamente concluímos que A e B são independentes. Os outros pares de evnentos, não. ●

Se A e B são independentes, a ocorrencia de B não deve ter efeito sobre a ocorrência de A; logo, A e $\overline{B}$ devem também ser independentes. Isso é aformado na Proposição 2.2.2. A demonstração com rigor é pedida no Exercício 26.

Proposição 2.2.2. A e B são eventos independentes se e somente se A e $\overline{B}$ são independentes.

Exemplo 2.2.6. No Exemplo 2.2.4, A e B são independentes. Também são independentes A e $\overline{B}$. Além disso, A e C são dependentes, asim como A e $\overline{C}$, como verificamos a seguir.



A $\quad$ $\overline{B}$ $\quad$ $\overline{C}$

$$\Pr(A) = \frac{1}{2} \qquad \Pr(\overline{B}) = \frac{1}{2} \qquad \Pr(\overline{C}) = \frac{3}{4}$$

$$\begin{aligned}
\Pr(A \cap \overline{B}) &= \frac{1}{4}, \quad & \Pr(A \cap \overline{C}) &= \frac{1}{2}, \\
\Pr(A)\Pr(\overline{B}) &= \frac{1}{4}, \quad & \Pr(A)\Pr(\overline{C}) &= \frac{3}{8}.
\end{aligned}$$

Da mesma forma, pode-se verificar que B e C são dependentes. ●

Exemplo 2.2.7. Suponha que uma mensagem é enviada por n canais independentes, $c_1, c_2, \ldots, c_n$. Se um dos canais transmitir corretamente a mensagem, diremos que a transmissão aconteceu com sucesso. Cada canal consegue transmitir a mensagem corretamente com probabilidade p_i.

Seja $C = \{c_1, c_2, \ldots c_n\}$ o conjunto de canais. O espaço amostral será o conjunto das partes de C,

$$\Omega = \mathcal{P}(C) = \left\{ \begin{array}{l} \emptyset, \{c_1\}, \{c_2\}, \ldots, \{c_n\}, \\ \{c_1, c_2\}, \{c_1, c_3\}, \ldots, \{c_{n-1}, c_n\}, \\ \quad \vdots \\ C \end{array} \right\},$$

onde cada resultado é o conjunto de canais que conseguiram transmitir a mensagem.

O evento FALHA é o conjunto vazio; o evento sucesso é todo o espaço amostral exceto o vazio (ou seja, $\mathcal{P}(C) \setminus \{\emptyset\}$).

A probabilidade da transmissão ser feita com sucesso é $1 - \Pr(\text{FALHA})$, onde "FALHA" é o evento em que todos os canais falham. Como os eventos são independentes,

$$\begin{aligned}\Pr(\text{SUCESSO}) &= 1 - \Pr(\text{FALHA}) \\ &= 1 - \Pr(\overline{C}_1 \cap \overline{C}_2 \cap \cdots \cap \overline{C}_n) \\ &= 1 - \Pr(\overline{C}_1)\Pr(\overline{C}_2)\cdots\Pr(\overline{C}_n).\end{aligned}$$

Observamos que $\Pr(\text{SUCESSO})$ *não* é igual a $\Pr(C_1) + \Pr(C_2) + \cdots + \Pr(C_n)$, porque os eventos *não* são mutuamente exclusivos! ●

2.3 Teoremas da probabilidade total e de Bayes

O Teorema da Probabilidade Total é uma afirmação simples a respeito de probabailidade condicional.

Teorema 2.3.1 (da probabilidade total)**.** Se $B_1, B_2, \dots, B_n$ são uma partição de Ω, e A é um evento, então

$$\begin{aligned}\Pr(A) &= \sum_i \Pr(A \cap B_i) \\ &= \sum_i \Pr(A|B_i)\Pr(B_i).\end{aligned}$$

Demonstração.

$$\begin{aligned}A &= \bigcup_i A \cap B_i \\ \Pr(A) &= \Pr\left(\bigcup_i A \cap B_i\right) \\ &= \sum_i \Pr(A \cap B_i).\end{aligned}$$

Como $\Pr(B \cap A_j) = \Pr(A_j)Pr(B|A_j)$,

$$\Pr(A) = \sum_i \Pr(A|B_i)\Pr(B_i).$$ ■

Exemplo 2.3.2. O Exemplo 2.1.3 descreve uma situação em que há vinte itens defeituosos em um lote de oitenta, e o experimento consiste em selecionar dois itens, um por vez.

Naquele exemplo, definimos os eventos A (primeiro item defeituoso) e B (segundo item defeituoso), e calculamos

$$\begin{aligned}\Pr(A) &= \frac{1}{5}, \\ \Pr(B|A) &= \frac{19}{99}.\end{aligned}$$

Agora observamos que A e $\overline{A}$ são uma partição de Ω, e portanto podemos calcular a probabilidade de B:

$$\begin{aligned}
\Pr(B) &= \Pr(B|A)\Pr(A) + \Pr(B|\overline{A})\Pr(\overline{A}) \\
&= \left(\frac{19}{99}\right)\left(\frac{1}{5}\right) + \left(\frac{20}{99}\right)\left(\frac{4}{5}\right) \\
&= \frac{19+80}{99(5)} \\
&= \frac{1}{5}.
\end{aligned}$$

●

Exemplo 2.3.3. Estamos estudando um vírus presente em tres regiões.

- na primeira região, o vírus infecta 20% da população;
- na segunda região, 10%;
- na terceira região, 15%.

As populações das regiões são

$$\begin{aligned}
R_1 &: 10000, \\
R_2 &: 5000, \\
R_3 &: 5000.
\end{aligned}$$

Se um paciente chega a um hospital, vindo de uma dessas regiões, sem que saibams qual delas, qual é a probabilidade dele estar infectado?

O espaço amostral é composto de pares ordenados (r, a), onde r é a região de onde o paciente vem, e a é a presença do vírus.

Determinamos alguns eventos: A é o evento que representa o paciente infectado; R_1, R_2, R_3 são os eventos representando a regiào de onde o paciente vem.

Presumindo equiprobabilidade entre as regiões (e isto pode ser muito longe da verdade em situações práticas), as probabilidades do paciente ter vindo de cada região são

$$\begin{aligned}
\Pr(R_1) &: \frac{10000}{20000} = \frac{1}{2}, \\
\Pr(R_2) &: \frac{5000}{20000} = \frac{1}{4}, \\
\Pr(R_3) &: \frac{5000}{20000} = \frac{1}{4}.
\end{aligned}$$

Sabemos as probabilidades condicionais de infecção dada cada região:

$$\begin{aligned}
\Pr(A|R_1) &= 0.20, \\
\Pr(A|R_2) &= 0.10, \\
\Pr(A|R_3) &= 0.15.
\end{aligned}$$

Pelo Teorema de Probabilidade Total,

$$\begin{aligned}\Pr(A) =&\Pr(A|R_1)P(R_1)\\&+\Pr(A|R_2)P(R_2)\\&+\Pr(A|R_3)P(R_3)\\=&(0.2)(0.5)+(0.1)(0.25)+(0.15)(0.25)\\=&0.1625.\end{aligned}$$

●

Outra ferramenta importante relacionada a probabilidade condicional é o Teorema de Bayes, que permite obter $\Pr(B|A)$ a partir de $\Pr(A|B)$.

Teorema 2.3.4 (de Bayes)**.** Sejam $A, B_1, B_2, \ldots,$ eventos em um espaço amostral Ω. Então

$$\Pr(B_i|A) = \frac{\Pr(A|B_i)\Pr(B_i)}{\Pr(A)}.$$

Se B_i é partição de A,

$$\Pr(B_i|A) = \frac{\Pr(A|B_i)\Pr(B_i)}{\sum_j \Pr(A|B_j)\Pr(B_j)}.$$

Demonstração.

$$\begin{aligned}\Pr(B_i|A) &= \frac{\Pr(B_i \cap A)}{\Pr(A)}\\&= \frac{\Pr(B_i)\Pr(A|B_i)}{\sum_i \Pr(B_i)\Pr(A|B_i)}.\end{aligned}$$

■

Exemplo 2.3.5. O exemplo 2.1.1 descreve um experimento onde dois dados diferentes são jogados, e três eventos são definidos:

- A: “dado vermelho par”;
- B: “dado branco ímpar”;
- C: “soma dos dois números é maior que 10”.

Naquele exemplo, calculamos as probabilidades a seguir.

$$\begin{aligned}\Pr(A) &= \frac{1}{2},\\\Pr(B) &= \frac{1}{2},\\\Pr(C) &= \frac{1}{12},\\\Pr(C|A) &= \frac{1}{9}.\end{aligned}$$

O Teorema de Bayes permite obter $\Pr(A|C)$ a partir de $\Pr(C|A)$, $\Pr(A)$ e $\Pr(C)$, que já temos.

$$\begin{aligned}\Pr(A|C) &= \frac{\Pr(C|A)\Pr(A)}{\Pr(C)} \\ &= \frac{(1/9)(1/2)}{1/12} \\ &= \frac{12}{18} \\ &= \frac{2}{3}.\end{aligned}$$

Como o espaço amostral é finito e pequeno, poderíamos também ter observado que $\Pr(A|C)$ é a quantidade de resultados com o primeiro valor par, dado que a soma é maior que dez. Como

$$C = \{(5,6), (6,5), (6,6)\},$$

claramente $\Pr(A|C) = 2/3$. ●

Exemplo 2.3.6. Testes de gravidez normalmente detectam a presença de um hormônio que só se pode perceber quando uma mulher está grávida. O teste pode falhar em detectar gravidez e dar resultado negativo (porque não foi possível detectar o hormônio), mas se o resultado é positivo, há a certeza da gravidez.

Suponha que um novo teste de gravidez, que usa um método diferente da detecção hormonal, foi desenvolvido. Este teste foi aplicado a 145 mulheres selecionadas aleatoriamente na população, com o seguinte resultado:

	positivo	negativo
grávida	85	5
não grávida	10	45

i) Se uma das mulheres for selecionada ao acaso, qual é a probabilidade dela estar grávida?

   Na primeira linha contabilizamos 90 mulheres grávidas, portanto

   $$\Pr(G) = \frac{90}{145} = \frac{18}{29} \approx 0.62.$$

ii) Se uma das mulheres for selecionada ao acaso, qual é a probabilidade de seu teste ter resultado positivo?

   A coluna da esquerda soma 95, e a mulher foi selecionada dentre todas as 156, logo

   $$\Pr(T) = \frac{95}{145} = \frac{19}{29} \approx 0.65.$$

iii) Se uma das mulheres for selecionada ao acaso *dentre as grávidas*, qual é a probabilidade de seu teste ter resultado positivo?

Na primeira linha há 90 mulheres, e dentre estas, 85 tem teste positivo.

$$\Pr(T|G) = \frac{85}{90} = \frac{17}{18} \approx 0.95.$$

iv) Se uma das mulheres for selecionada ao acaso, e o resultado de seu teste for positivo, qual é a probabilidade dela estar grávida?

Aqui o Teorema de Bayes pode ser usado, porque temos $\Pr(T|G)$, $\Pr(T)$ e $\Pr(G)$:

$$\begin{aligned}
\Pr(G|T) &= \frac{\Pr(T|G)\Pr(G)}{\Pr(T)} \\
&= \frac{(17/18)(18/29)}{(19/29)} \\
&= \frac{17}{29}\frac{29}{19} \\
&= \frac{17}{19} \\
&\approx 0.8947.
\end{aligned}$$

A rigor, neste exemplo, não teríamos precisado do Teorema de Bayes: Poderíamos ter calculado $\Pr(G|T)$ observando que há 95 mulheres com teste positivo, e 85 delas estão grávidas, logo

$$\Pr(G|T) = \frac{85}{95} = \frac{17}{19}.$$

Nem sempre, no entanto, teremos um espaço amostral finito e disposto em uma tabela, como neste exemplo. ●

O Exemplo 2.3.7 mostra uma aplicação prática importante de probabilidade condicional e do Teorema de Bayes. Nesse exemplo, não explicitamos o espaço amostral – isso é pedido no Exercício 39.

Exemplo 2.3.7. Suponha que um teste para um tipo específico de tumor tenha as seguintes características

- A sensitividade do teste (probabilidade de resultado positivo em pacientes que de fato desenvolveram o tumor) é 0.95;
- A probabilidade de falso positivo é 0.05.

Sabemos também que a probabilidade de um paciente ter desenvolvido o tumor é 0.02.

Se o teste de um paciente tem resultado positivo, qual é a probabilidade desse paciente ter desenvolvido o tumor?

Definimos dois eventos:

- T: o paciente desenvolveu o tumor;
- P: o resultado do teste é positivo para o paciente.

Listamos agora as probabilidades que temos:

$$\begin{aligned}
\Pr(T) &= 0.02, \\
\Pr(\overline{T}) &= 0.98, && (1 - \Pr(T)) \\
\Pr(P|T) &= 0.95, && \text{(senstividade)} \\
\Pr(P|\overline{T}) &= 0.05. && \text{(falso positivo)}
\end{aligned}$$

Queremos a probabilidade de tumor, dado que o resultado do teste é positivo, $\Pr(T|P)$. Aparentemente, não podemos usar diretamente o Teorema de Bayes: para calcular $\Pr(T|P) = \frac{\Pr(T \cap P)}{\Pr(P)}$, precisaríamos de$\Pr(T \cap P)$ e $\Pr(P)$, que não temos – ainda.

No entanto,

$$\begin{aligned}
\Pr(T|P) &= \frac{\Pr(T \cap P)}{\Pr(P)} \\
&= \frac{\Pr(T \cap P)}{\Pr(T \cap P \cup \overline{T} \cap P)} && \text{(eventos são conjuntos!)} \\
&= \frac{\Pr(T \cap P)}{\Pr(T \cap P) + \Pr(\overline{T} \cap P)} && (T \text{ e } P \text{ são independentes}) \\
&= \frac{\Pr(T)\Pr(P|T)}{\Pr(T)\Pr(P|T) + \Pr(\overline{T})\Pr(P|\overline{T})}. && \text{(definição de } \Pr(.|.))
\end{aligned}$$

Agora temos todos os valores necessários. Substituimos na equação:

$$\Pr(T|P) = \frac{(0.02)(0.95)}{(0.02)(0.95) + (0.98)(0.05)} \approx 0.2794. \qquad \bullet$$

2.4 Monty Hall e soluções "intuitivas"

O *problema de Monty Hall* é um interessante problema de Probabilidade, inspirado no show televisivo *Let's Make a Deal*, que foi ao ar pela primeira vez em 1963. O nome do problema é uma referencia a um dos apresentadores do programa, o Canadense-Americano Monty Hall. Este é um problema importante por ilustrar como a intuição pode estar, algumas vezes, completamente equivocada.

No programa, o apresentador pede que o jogador escolha uma dentre três portas: atrás de uma delas há um carro, e atrás das outras duas, há dois bodes. O jogador ganhará o prêmio atrás da porta que tiver escolhido.

Após o jogador fazer sua escolha sobram duas portas – e certamente uma delas contém um bode. Neste momento, o apresentador abre uma porta não escolhida, e que contém um bode. Em seguida pergunta ao jogador se ele quer mudar de porta, ou se mantém sua escolha.

Como o jogador tem que optar agora por duas portas, uma com o carro e outra com um bode, poderíamos crer que a probabilidade dele ganhar é 1/2 com qualquer das escolhas, e que não faz diferença trocar ou não de porta. E esta resposta, surpreendentemente, está errada!

O problema é que a porta que o apresentador abre não é independente da escolha do jogador, e nem da posição do carro. Apresentamos a seguir uma solução para o problema, usando probabilidade condicional e o Teorema de Bayes.

Cada episódio do show pode ser visto como um experimento aleatório. Defina três variáveis aleatórias

- J, igual ao número da porta escolhida pelo jogador;
- C, igual ao número da porta onde o carro está;
- O, igual ao número da porta aberta pelo apresentador.

Primeiro, notamos que o carro pode estar em qualquer uma das portas, com igual probabilidade:

$$\Pr[C = 1] = \Pr[C = 2] = \Pr[C = 3] = \frac{1}{3}.$$

Agora suponha que o jogador tenha escolhido a porta um. Calculamos a probabilidade do apresentador Monty abrir a porta dois.

$$\begin{aligned}
&\Pr[O = 2|J = 1, C = 1] = \frac{1}{2} && \text{(Monty pode escolher 2 ou 3)}\\
&\Pr[O = 2|J = 1, C = 2] = 0 && \text{(não! o carro está na 2.)}\\
&\Pr[O = 2|J = 1, C = 3] = 1 && \text{(Monty não tem escolha, abre a 2!)}
\end{aligned}$$

Também é evidente que

$$\Pr[C = i, J = i] = \Pr[C = i] \Pr[J = i],$$

porque a escolha do jogador e o local do carro são independentes.

Usaremos o Teorema de Bayes:

$$\Pr(A|B, C) \Pr(B, C) = \Pr(A, B, C) = \Pr(B|A, C) \Pr(A, C),$$

logo

$$\Pr(A|B, C) = \frac{\Pr(B|A, C) \Pr(A, C)}{\Pr(B, C)}.$$

Agora calculamos a probabilidade do jogador ganhar *mudando* de porta, e escolhendo a porta 3 (a dois foi aberta).

$$\begin{aligned}
\Pr[C=3|O=2,J=1] &= \frac{\Pr[O=2|C=3,J=1]\ \Pr[C=3,J=1]}{\Pr[O=2,J=1]} \\
&= \frac{(1)\ \Pr[C=3,J=1]}{\Pr[O=2,J=1]} \\
&= \frac{\Pr[C=3,J=1]}{\Pr[O=2|J=1]\Pr[J=1]} \\
&= \frac{\Pr[C=3]\Pr[J=1]}{\Pr[O=2|J=1]\Pr[J=1]} \\
&= \frac{(1/3)(1)}{1/2} \\
&= \frac{2}{3}.
\end{aligned}$$

Em simulações, pode-se verificar que, quando o jogador sempre troca de portas, e o experimento é repetido muitas vezes, ele ganha em $2/3$ das vezes. Se nunca troca de porta, ele ganha o carro em somente $1/3$ das vezes.

Em um artigo na revista *Decision Line*, Andrew Vazsoini menciona[1] que mesmo Paul Erdös, um dos mais renomados matemáticos do século vinte, demorou a aceitar a solução correta – e só considerou que estava errado após ver resultados de uma simulação em um computador.

Exercícios

Ex. 29 — Dois dados idênticos são lançados.

1. Qual é a probabilidade da soma ser par, dado que um dos dados mostrou a face "3" para cima?
2. Qual é a probabilidade de um dos dados ter resultado par, dado que a distancia entre os resultados é dois?

Ex. 30 — Escolha dois vetores unitários u e v aleatoriamente em $\mathbb{R}^2$. Seja A o ângulo entre eles, medido de u até v no sentido anti-horário. Determine:

1. $\Pr\left[\operatorname{sen}(A)+\cos(A)>0\right]$
2. $\Pr\left[\operatorname{sen}(A)+\cos(A)>0 \mid \operatorname{sen}(A)>0\right]$
3. $\Pr\left[\operatorname{sen}(A)+\cos(A)=0\right]$
4. $\Pr\left[\operatorname{sen}(A)+\cos(A)=0 \mid A=\dfrac{k\pi}{4}\right]$, onde $k\in\mathbb{N}$
5. $\Pr\left[\operatorname{sen}(A)+\cos(A)=0 \mid A=\dfrac{k\pi}{8}\right]$, onde $k\in\mathbb{N}$

[1] Vazsonyi, Andrew (December 1998 – January 1999). *"Which Door Has the Cadillac?"*. Decision Line: 17–19.

Ex. 31 — Sortearei uma matriz A, com duas linhas e duas colunas tendo entradas iguais a zero ou um.

1. Calcule $\Pr[\det A \neq 0]$.
2. Calcule $\Pr[\det A \neq 0 \mid a_{11}a_{12} = 0]$.
3. Calcule $\Pr[a_{11}a_{12} = 0 \mid \det A \neq 0]$.
4. Seja
$$v = \begin{pmatrix} x \\ y \end{pmatrix}$$
um vetor com $x, y \in \mathbb{N}$, escolhidos aletoriamente. Qual a probabilidade de Av ser um vetor somente com entradas pares, *dado que A tem no máximo um zero?*

Ex. 32 — 40% dos estudantes de graduação em uma Universidade recebam bolsa-auxílio para poderem permanecer estudando. Sabe-se que 10% destes estudantes não são do estado onde a Universidade fica (vieram de outro estado para estudar). Se um estudante for selecionado ao acaso, qual é a probabilidade dele ser de outro estado, *e* receber bolsa-auxílio?

Ex. 33 — Jogo uma moeda três vezes. Qual é a probabilidade de *no mínimo* dois dos resultados serem cara, dado que *no mínimo* um é cara?

Ex. 34 — Duas fábricas produzem o mesmo doce, no mesmo formato, mas em sabores ligeiramente diferentes, sempre em sacos de 30 unidades.

- A fábrica A produz sacos de doce com sabores maçã (10 unidades), cereja (5 unidades), abacaxi (5 unidades), e uva (10 unidades).
- A fábrica B produz doces com sabores maçã (8 unidades), cereja (5 unidades), abacaxi (5 unidades), uva (7 unidades), e morango (5 unidades).

Tenho dois pacotes, um de cada fábrica. Abro um deles e ofereço um doce a um amigo, mas não digo de que fábrica ele veio.
Dado que os doces são um de cereja e um de uva, qual é a probabilidade do saco de doces ser da fábrica A?

Ex. 35 — Uma urna tem 5 bolas vermelhas e 8 bolas brancas. Duas bolas são selecionadas, uma por vez, e sem reposição. Qual é a probabilidade da *primeira* bola ser vermelha, dado que a *segunda* é branca?

Ex. 36 — Dois dados diferentes são jogados. Definimos três eventos:

- A = os dois dados resultam em números maiores ou iguais a quatro;
- B = a soma dos dois números é par;
- C = a diferença entre os valores dos dois números é maior ou igual que seis.

Determine, para cada par de eventos, quais são independentes.

Ex. 37 — Qual é a probabilidade de uma mão de pôquer (5 cartas) ser toda de copas, dado que os naipes das cartas são todos vermelhos?

Ex. 38 — Seis bolas são retiradas, sem reposição, de uma urna onda havia três bolas brancas. As outras são azuis. Sabendo que a probabilidade das três bolas brancas serem retiradas é o dobro da probabilidade de nenhuma delas ser removida, determine quantas bolas há na urna.

Ex. 39 — No Exemplo 2.3.7, tratamos de dois eventos diferentes, mas não deixamos explícito qual era o espaço amostral. Proponha um espaço amostral que poderia ter sido usado naquele exemplo.

Ex. 40 — Sejam A e B dois eventos em um espaço amostral. O que se pode dizer da relação entre $\Pr(A|B)$ e $\Pr(A)$, nas situações a seguir?

1. $A \cap B = \emptyset$
2. $A \subseteq B$
3. $B \subseteq A$

Ex. 41 — Calcule as probabilidades de sucesso para o problema de Monty Hall, para os casos em que o jogador sempre troca e nunca troca:

1. quatro portas (o apresentador fecha duas);
2. $n > 4$ portas (o apresentador fecha $n - 2$);
3. quatro portas (o apresentador fecha *uma*, e o jogador escolhe aleatoriamente entre as outras);
4. $n > 4$ (o apresentador fecha *uma*, e o jogador escolhe aleatoriamente entre as outras);

Capítulo 3

Variáveis Aleatórias Discretas

3.1 Variáveis Aleatórias

É comum que tenhamos interesse não no resultado de um experimento, mas em algum número relacionado a ele. Por exemplo, ao jogar duas moedas, o resultado pode ser cara ou coroa, e o espaço amostral mais natural que imaginamos é $\Omega = \{(a, a), (a, b), (b, a), (b, b)\}$, onde a é cara e b é coroa. Mas podemos querer medir, por exemplo, a quantidade de caras, e trabalhar com funções usando esta quantidade. Uma solução para este problema é definir uma função de Ω em um conjunto numérico. Por exemplo, uma função poderia mapear

$$\begin{aligned}
(a, a) &\to 2,\\
(a, b) &\to 1,\\
(b, a) &\to 1,\\
(b, b) &\to 0.
\end{aligned}$$

Definição 3.1.1 (variável aleatória)**.** Seja Ω um espaço amostral. Uma **variável aleatória** é uma função

$$X : \Omega \to \mathbb{R}.$$

Se a quantidade de possíveis valores que X pode assumir é finita ou enumerável, então X é uma variável aleatória **discreta**.

Se $s \in \Omega$ é um possível resultado de um experimento, $X(s)$ é um número real. Usaremos a notação "X" ao invés de "$X(s)$", porque nos interessam mais os valores que $X(s)$ pode ter, e menos quem era o resultado s. ◀

Exemplo 3.1.2. Em um experimento onde dois dados distintos são jogados, determinamos o espaço amostral

$$\Omega = \{(1, 1), (1, 2), \dots, (6, 6)\}.$$

Podemos definir a variável aleatória S, que dá a soma dos dois valores obtidos:

$$S(a,b) = a + b$$

Definimos outra variável aleatória D, que determina o valor absoluto da diferença entre os dois valores obtidos nos dados:

$$D(a,b) = |a - b|$$

Os possíveis valores de D são

$$R_D = \{0, 1, 2, 3, 4, 5\}.$$

Podemos definir um evento contendo os resultados onde a diferença é zero, $D_0 = \{(1,1), (2,2), (3,3), (4,4), (5,5), (6,6)\}$. A probabilidade do evento D_0 é a probabilidade da variável aleatória D ser igual a zero. Usamos a notação $\Pr[D = 0]$.

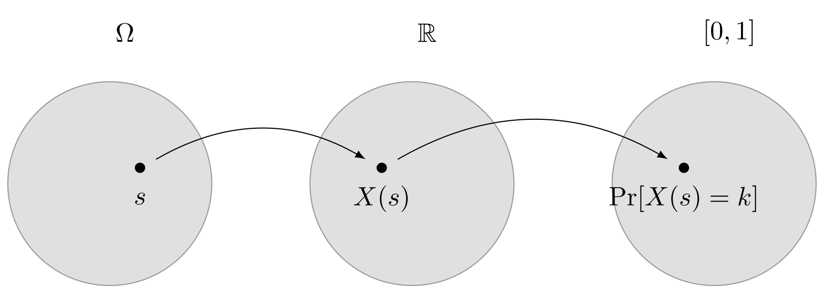
$$\Pr(D_0) = \Pr[D = 0] = \frac{6}{36} = \frac{1}{6}. \qquad \bullet$$

Cada valor que uma variável aleatória pode assumir tem uma probabilidade; se somarmos todos os valores possíveis, estaremos calculando a probabilidade da variável assumir "um de todos os seus possíveis valores" – o que é claramente igual a um.

Podemos fazer uma analogia com uma quantidade de massa igual a um, que distribuímos entre os valores da variável: se, por exemplo, X é $+1$ quando uma moeda resulta em cara, e -1 quando resulta em coroa, e a moeda é honesta, imaginamos que *a massa de probabilidade é dividida igualmente para os casos em que $X = +1$ e $X = -1$*. Por convenção, a massa disponível é um, portanto $\Pr[X = +1] = 1/2$ e $\Pr[X = -1] = 1/2$.

Definição 3.1.3 (função de massa de probabilidade)**.** A **função de massa de probabilidade** de uma variável aleatória discreta é uma função que dá a probabilidade dessa variável ser igual a um dado valor. ◀

A próxima figura ilustra os conceitos de variável aleatória e função de massa de probabilidade. Um resultado do experimento, $s \in \Omega$, é mapeado em um número real pela variável aleatória (função) X. Depois, a função de massa de probabilidade atribui alguma quantidade entre 0 e 1 para a probabilidade de X assumir o valor $X(s)$. Na figura, $s \in \Omega$ pode ser *qualquer* objeto; já $X(s)$ é um número real, e $\Pr[X(s)]$ é um real entre zero e um.



No resto do texto, no entanto, para maior clareza de notação, escreveremos simplesmente “X” e “$X = x$” ao invés de “$X(s)$” e “$X(s) = x$”. Também escrevemos $\Pr[X = k]$ para “a probabilidade de $X(s)$ ser iguala k”.

Exemplo 3.1.4. No experimento onde jogamos um dado, definimos uma variável aleatória X, que siplesmente reflete o número resultante.

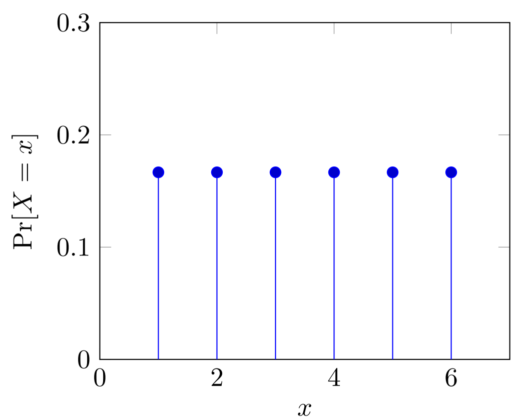
Sendo o dado honesto,

$$\begin{aligned} \Pr[X=1] &= \frac{1}{6} & \Pr[X=4] &= \frac{1}{6} \\ \Pr[X=2] &= \frac{1}{6} & \Pr[X=5] &= \frac{1}{6} \\ \Pr[X=3] &= \frac{1}{6} & \Pr[X=6] &= \frac{1}{6}, \end{aligned}$$

e portanto a função de massa de probabilidade de X é

$$\begin{aligned} f(1) &= \frac{1}{6} & f(4) &= \frac{1}{6} \\ f(2) &= \frac{1}{6} & f(5) &= \frac{1}{6} \\ f(3) &= \frac{1}{6} & f(6) &= \frac{1}{6}. \end{aligned}$$

A figura a seguir ilustra esta função.



●

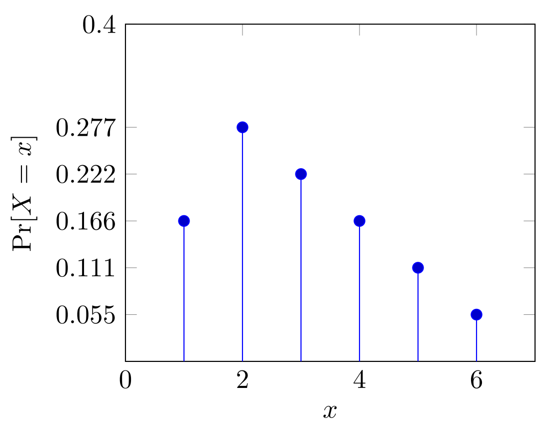
Exemplo 3.1.5. No Exemplo 3.1.2, a variável aleatória X mede a diferença entre os dois valores resultantes quando jogamos dois dados distintos. Para esta variável aleatória,

$$
\begin{aligned}
\Pr[X=0] &= \frac{1}{6} \approx 0.1666 & \Pr[X=3] &= \frac{1}{6} \approx 0.1666 \\
\Pr[X=1] &= \frac{5}{18} \approx 0.2777 & \Pr[X=4] &= \frac{1}{9} \approx 0.1111 \\
\Pr[X=2] &= \frac{2}{9} \approx 0.2222 & \Pr[X=5] &= \frac{1}{18} \approx 0.0555,
\end{aligned}
$$

e portanto a função de massa de probabilidade de X é

$$
\begin{aligned}
f(0) &= \frac{1}{6} \approx 0.1666 & f(3) &= \frac{1}{6} \approx 0.1666 \\
f(1) &= \frac{5}{18} \approx 0.2777 & f(4) &= \frac{1}{9} \approx 0.1111 \\
f(2) &= \frac{2}{9} \approx 0.2222 & f(5) &= \frac{1}{18} \approx 0.0555.
\end{aligned}
$$

A figura a seguir ilustra esta função.



●

3.2 Distribuições de Probabilidade

Já mencionamos que a probabilidade do espaço amostral inteiro é um. Se o espaço amostral Ω for particionado em eventos disjuntos $A_1, A_2, \ldots, A_k$, de forma que $\cup A_i = \Omega$, teremos

$$\Pr(A_i) = \Pr(\Omega) = 1,$$

sendo que *cada evento A_i pode ter probabilidade* $0 \leq p \leq 1$. De maneira simplificada, uma *distribuição de probabilidade* é uma descrição da função que atribui um valor de probabilidade a cada um dos eventos. Para as variáveis discretas de que tratamos neste capítulo, usamos a *função de distribuição de massa*; no Capítulo 4, uma "distribuição" será descrita pela *função de densidade* da variável.

Há algumas distribuições que são encontradas muito frequentemente, e que portanto tem nomes – uniforme, binomial, normal, dentre outras. Neste Capítulo e no próximo descreveremos brevemente algumas delas, mas antes disso damos, sem análise, um exemplo simples de distribuição "com nome": a distribuição de Benford.

Exemplo 3.2.1 (Lei de Benford)**.** Em 1881, o astrônomo americano Simon Newcomb observou que o livro onde ficavam os valores de logaritmos (onde ficavam os números que começam com um) tinhas as páginas do início mais desgastadas por uso do que as do final. Newcomb identificou que os nos números que observava, e cujos logaritmos eram buscados, o dígito inicial era mais comumente igual a um ou a dois do que a outros dígitos. Ele determinou que a probabilidade do primeiro dígito ser igual a d (com d entre 1 e 9) é

$$\log_{10}\left(1 + \frac{1}{d}\right).$$

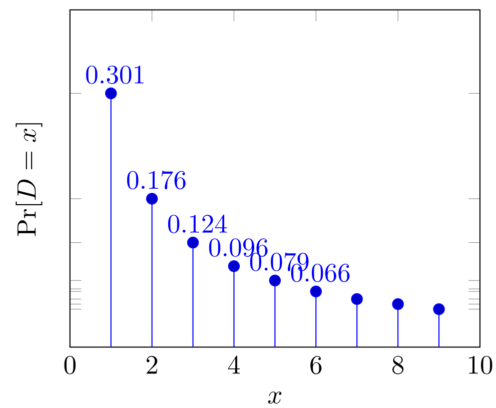
Em 1938, Frank Benford resgatou este fato e o verificou em vários conjuntos de dados: áreas de rios, tamanhos de populações, constantes físicas, e outros. Este fenômeno é conhecido como *Lei de Benford*.

Dificilmente um conjunto de dados obedece perfeitamente a Lei de Benford, mas teoricamente, podemos idealizar a *distribuição de Benford*, cuja função de massa é

$$f(d) = \log_{10}\left(1 + \frac{1}{d}\right).$$

Assim, se D é a variável aleatória que nos idica o primeiro dígito de um número selecionado aleatoriamente de um conjunto de dados que segue a Lei de Benford,

$$\begin{aligned}
\Pr[D = 1] &= \log_{10}(1 + 1) = 0.301 \\
\Pr[D = 2] &= \log_{10}(1 + 1/2) = 0.176 \\
\Pr[D = 3] &= \log_{10}(1 + 1/3) = 0.124 \\
\Pr[D = 4] &= \log_{10}(1 + 1/4) = 0.096 \\
\Pr[D = 5] &= \log_{10}(1 + 1/5) = 0.079 \\
\Pr[D = 6] &= \log_{10}(1 + 1/6) = 0.066 \\
\Pr[D = 7] &= \log_{10}(1 + 1/7) = 0.057 \\
\Pr[D = 8] &= \log_{10}(1 + 1/8) = 0.051 \\
\Pr[D = 9] &= \log_{10}(1 + 1/9) = 0.045
\end{aligned}$$



A distribuição de Benford é, realmente, uma distribuição de massa. Lembramos que

$$10^{\log_{10} x + \log_{10} y} = xy,$$

e verificamos:

$$\begin{aligned}
\sum_{d=1}^{9} \log_{10}\left(1 + \frac{1}{d}\right) &= \log_{10} 10^{\left(\sum_{d=1}^{9} \log_{10}\left(1+\frac{1}{d}\right)\right)} \\
&= \log_{10} \prod_{d=1}^{9} \left(1 + \frac{1}{d}\right) \\
&= \log_{10} 10 \\
&= 1.
\end{aligned}$$

A Lei de Benford é usada, por exemplo, em detecção de fraudes contábeis (porque valores em conjnutos de dados contábeis normalmente seguem, de maneira aproximada, a distribuição de Benford), e em detecção de fraudes eleitorais. ●

Quando uma variável X aleatória tem função de distribuição de massa igual a uma distribuição conhecida D, denotamos "$X \sim D$". Por exemplo, "$X \sim Benf$" significa "X tem função de distribuição de massa igual à distribuição de Benford"; $X \sim Y$ significa "X e Y tem a mesma distribuição de massa". O mesmo valerá também para distribuições contínuas (que ainda não definimos): $X \sim \mathcal{E}xp(0.5)$ significa "X tem distribuição exponencial com parâmetro 0.5".

3.3 Distribuição Acumulada

Definição 3.3.1 (função de distribuição acumulada de probabilidade)**.** A **função de distribuição acumulada de probabilidade** de uma variável aleatória discreta X é a função tal que

$$F_X(x) = \Pr[X \leq x].$$ ◀

Teorema 3.3.2. Se X é uma variável aleatória discreta e F sua função de distribuição acumulada, então

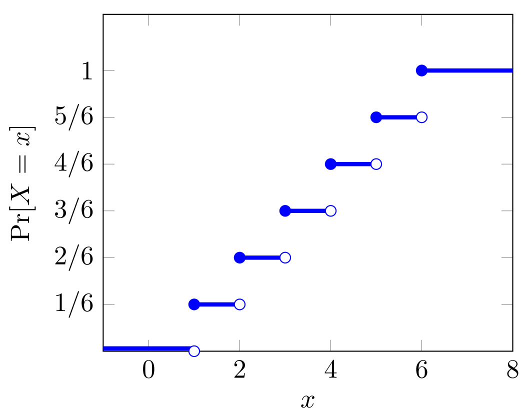
$$F(x) = \sum_{x_j \leq x} \Pr[X = x_j].$$

Exemplo 3.3.3. O Exemplo 3.1.4 ilustra a função de massa de probabilidade, a partir do experimento onde um único dado é jogado. Naquele Exemplo, uma variável aleatória X é definida, sendo seu valor sempre igual ao número resultante quando o dado é jogado.

A função de distribuição acumulada para a variável X é

$$F_X(x) = \begin{cases} 0 & x \in (-\infty, 1) \\ 1/6 & x \in [1, 2) \\ 2/6 & x \in [2, 3) \\ 3/6 & x \in [3, 4) \\ 4/6 & x \in [4, 5) \\ 5/6 & x \in [5, 6) \\ 1 & x \in [6, \infty) \end{cases}$$

O gráfico de F_X é mostrado a seguir.



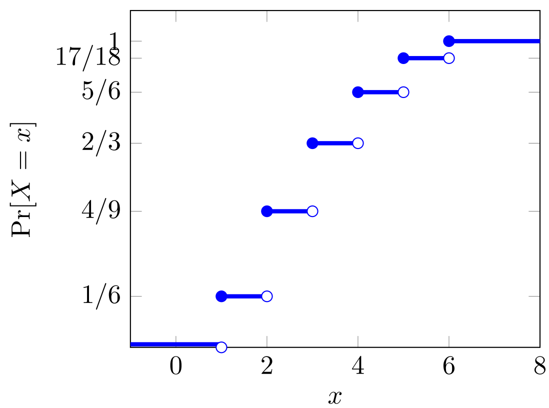
A função F_X é definida para todo número real, porque a pergunta "$\Pr[X \leq x]$" faz sentido para qualquer x real. ●

Exemplo 3.3.4. O Exemplo 3.1.5 trata do experimento em que dois dados distintos são jogados. É definida, ali, uma variável aleatória X, que dá a diferença entre os valores dos dados, e sua função de massa de probabilidade é mostrada.

A função de distribuição acumulada para X é

$$F_X(x) = \begin{cases} 0 & x \in (-\infty, 0) \\ 1/6 \approx 0.1666 & x \in [0, 1) \\ 4/9 \approx 0.4444 & x \in [1, 2) \\ 2/3 \approx 0.6666 & x \in [2, 3) \\ 5/6 \approx 0.8333 & x \in [3, 4) \\ 17/18 \approx 0.9444 & x \in [4, 5) \\ 1 & x \in [5, \infty) \end{cases}$$

O gráfico de F_X é mostrado a seguir.



A função F_X é definida para todo número real, porque a pergunta "Pr[$X \leq x$]" faz sentido para qualquer x real. ●

Definição 3.3.5 (variável indicadora)**.** Uma variável aleatória discreta que pode assumir somente dois valores é chamada de **variável indicadora**. ◀

Em geral, os valores de variáveis indicadoras são zero e um.

Exemplo 3.3.6. Em uma população de 1500 pessoas, 350 fumam. Ao definir um experimento aleatório onde selecionamos uma pessoa e verificamos se ela é fumante, a variável aleatória F valerá 1 se o indivídio selecionado é fumante, e 0 se não é. ●

3.4 Esperança, variância e desvio padrão

Como uma variável aleatória nos dá um valor numérico, podemos querer saber que número estará mais próximo dos valores da variável após muitas repetições. Esse valor é a *esperança* da variável aleatória, a média dos valores obtidos quando a quantidade de repetições do experimento tende a ∞.

Definição 3.4.1 (esperança de variável aleatória discreta)**.** A **esperança** de uma variável aleatória discreta X é

$$\mathbb{E}(X) = \sum x \Pr[X = x],$$

onde o somatório é sobre todos os valores possíveis de x. ◀

Exemplo 3.4.2. No experimento em que uma moeda é jogada, defina a variável aletória X, que assume valor 1 quando o resultado é cara, e 0 quando o resultado é coroa.

A esperança de X é

$$\begin{aligned}
\mathbb{E}[X] &= \sum_x x \Pr[X = x] \\
&= 1\left(\frac{1}{2}\right) + 0\left(\frac{1}{2}\right) \\
&= \frac{1}{2}.
\end{aligned}$$

Se, no entanto, definirmos outra variável, Y, que vale 1 quando o resultado é cara e -1 quando o resultado é coroa, a esperança será

$$\begin{aligned}
\mathbb{E}[Y] &= \sum_y y \Pr[Y = y] \\
&= +1\left(\frac{1}{2}\right) + (-1)\left(\frac{1}{2}\right) \\
&= 0.
\end{aligned}$$

A esperança depende do experimento *e dos valores que a variável aleatória assume para cada evento* – como deve ficar claro a partir dos dois casos acima.

Se definirmos outra variável aleatória, Z, valendo 5 e 10 para cara e coroa, a esperança de Z será 7.5. ●

Exemplo 3.4.3. No Exemplo 3.1.2, definimos a variável aleatória X, que dá a diferença entre os dois números obtidos ao jogar dois dados.

Para determinar as probabilidades de cada possível valor de X, observamos as seguintes tabelas, que indicam, para cada par de possíveis resultados nos dados, a diferença entre os dois valores.

		vermelho					
		1	2	3	4	5	6
branco	1	0	1	2	3	4	5
	2	1	0	1	2	3	4
	3	2	1	0	1	2	3
	4	3	2	1	0	1	2
	5	4	3	2	1	0	1
	6	5	4	3	2	1	0

A tabela a seguir mostra a contagem de possibilidades de ocorrência de cada diferença.

diferença	nº de casos	$P[X = n]$
0	6	$6/36 = 1/6$
1	10	$10/36 = 5/18$
2	8	$8/36 = 2/9$
3	6	$6/36 = 1/6$
4	4	$4/36 = 1/9$
5	2	$2/36 = 1/18$

A esperança dessa variável é

$$\begin{aligned}\mathbb{E}(X) = & 0(1/6) \\ & + 1(5/18) \\ & + 2(2/9) \\ & + 3(1/6) \\ & + 4(1/9) \\ & + 5(1/18) \\ & = \frac{35}{18}.\end{aligned}$$

●

Exemplo 3.4.4. No experimento em que jogamos uma moeda até obter cara, o espaço amostral é

$$\Omega = \{a, oa, ooa, oooa, \ldots\}.$$

A função de massa de probabilidade desta variável é

$$f(n) = \frac{1}{2^n}$$

As probabilidades foram atribuídas corretamente, porque

$$\sum_{j=1}^{\infty} \frac{1}{2^j} = 1.$$

Uma variável aleatória que podemos definir neste espaço é a que contabiliza quantas vezes tivemos que jogar a moeda para obter cara. Denotamos esta variável por X. A esperança de X é

$$\begin{aligned}\mathbb{E}[X] &= \sum_{j=1}^{\infty} x_j \Pr[X = x_j] \\ &= \sum_{j=1}^{\infty} j \left(\frac{1}{2^j}\right) \\ &= 2.\end{aligned}$$

Para este último passo, verificamos o valor da soma parcial

$$\sum_{j=1}^{n} j \left(\frac{1}{2^j}\right) = \frac{1}{2^n}\left(-n + 2^{n+1} - 2\right),$$

e portanto

$$\begin{aligned}\lim_{n\to\infty} \sum_{j=1}^{n} j \left(\frac{1}{2^j}\right) &= \lim_{n\to\infty} \frac{1}{2^n}\left(-n + 2^{n+1} - 2\right) \\ &= \lim_{n\to\infty} \frac{-n}{2^n} + \frac{2^{n+1}}{2^n} - \frac{2}{2^n} \\ &= 0 + 2 + 0.\end{aligned}$$

●

Quando o espaço amostral é infinito, a esperança de uma variável aleatória pode ser infinita, como ilustra o exemplo 3.4.5.

Exemplo 3.4.5. No exemplo 3.4.4, definimos outra variável aleatória, Y, que nos dá uma função exponencial da quantidade de jogadas feita.

$$\begin{aligned}Y(a) &= 2 \\ Y(oa) &= 2^2 \\ Y(ooa) &= 2^3 \\ &\vdots\end{aligned}$$

A esperança de Y é

$$\begin{aligned}
\mathbb{E}[Y] &= \sum_{j=1}^{\infty} y_j \Pr[Y = y_j] \\
&= \sum_{j=1}^{\infty} 2^j \left(\frac{1}{2^j}\right) \\
&= \sum_{j=1}^{\infty} 1. \\
&= \infty.
\end{aligned}$$

●

Se $f(X)$ é uma função f aplicada a uma variável aleatória X, então fica bem definida a esperança de $f(X)$.

Teorema 3.4.6. Se X é uma variável aleatória discreta, e g uma função,

$$\mathbb{E}[g(X)] = \sum_i g(x_i) \Pr(x_i).$$

Demonstração. Definimos, para o mesmo experimento, uma variável Y que assumirá somente valores $g(x)$, para todo possível valor x da variável X. Então,

$$\Pr[Y = y] = \sum_{x:\ g(x)=y} \Pr[X = x].$$

Com isso, temos

$$\begin{aligned}
\mathbb{E}[g(X)] &= \mathbb{E}[Y] \\
&= \sum_y y \Pr[Y = y] \\
&= \sum_y y \sum_{x:\ g(x)=y} \Pr[X = x] \\
&= \sum_y \sum_{x:\ g(x)=y} y \Pr[X = x].
\end{aligned}$$

Mas nesta soma, para "todo valor de y", um termo só será diferente de zero quando o y tomado no primeiro somatório for tal que $g(x) = y$ para algum x. Os dois somatórios estão apenas selecionando x tal que $g(x) = y$, logo podemos reescrevê-los:

$$\begin{aligned}
\mathbb{E}[Y] &= \sum_x y \Pr[X = x] \\
\mathbb{E}[g(X)] &= g(x) \sum_x g(x) \Pr[X = x].
\end{aligned}$$

■

Exemplo 3.4.7. Em um experimento jogamos uma única moeda honesta, e definimos uma variável aleatória:

$$X = \begin{cases} 0 & \text{cara} \\ 10 & \text{coroa} \end{cases}$$

A esperança de X é

$$\mathbb{E}[X] = 0\left(\frac{1}{2}\right) + 10\left(\frac{1}{2}\right) = 5.$$

Já a esperança de X^2 é

$$\mathbb{E}[X] = 0^2\left(\frac{1}{2}\right) + 10^2\left(\frac{1}{2}\right) = 50.$$

Agora calculamos a esperança de $2X - 1$:

$$\begin{aligned} \mathbb{E}[X] &= [2(0) - 1]\left(\frac{1}{2}\right) + [2(10) - 1]\left(\frac{1}{2}\right) \\ &= -\frac{1}{2} + \frac{19}{2} \\ &= 9. \end{aligned}$$

●

Teorema 3.4.8. A esperança é linear: se X é uma variável aleatória discreta,

$$\begin{aligned} \mathbb{E}[aX] &= a\mathbb{E}[X], \\ \mathbb{E}[X + Y] &= \mathbb{E}[X] + \mathbb{E}[Y]. \end{aligned}$$

Demonstração. Para a primeira parte, $\mathbb{E}[aX] = a\mathbb{E}[X]$, usamos o Teorema 3.4.6:

$$\begin{aligned} \mathbb{E}[aX] &= \sum_x ax \Pr[X = x] \\ &= a \sum_x x \Pr[X = x] \\ &= a\mathbb{E}[X]. \end{aligned}$$

Para a segunda parte,

$$\begin{aligned} \mathbb{E}[X + Y] &= \sum_{x,y} (x + y) \Pr[X = x, Y = y] \\ &= \left(\sum_{x,y} x \Pr[X = x, Y = y]\right) + \left(\sum_{x,y} y \Pr[X = x, Y = y]\right) \\ &= \left(\sum_x x \sum_y \Pr[X = x, Y = y]\right) + \left(\sum_y y \sum_x \Pr[X = x, Y = y]\right) \\ &= \left(\sum_x x \Pr[X = x]\right) + \left(\sum_y y \Pr[Y = y]\right) \\ &= \mathbb{E}[X] + \mathbb{E}[Y]. \end{aligned}$$

■

Exemplo 3.4.9. Em um jogo, ganho o valor da diferença entre os dados, se a diferença for par, e perco o valor da diferença, se for ímpar.

Se eu jogar dez vezes, quanto é o valor esperado de meu ganho (que pode ser negativo)?

Seja X_i a variável aleatória que dá o valor ganho na i -ésima jogada. As X_i são todas idênticas, e calculamos a esperança delas. O cálculo é semelhante ao realizado no Exemplo 3.1.2, exceto que subtraímos os valores quando a diferença é ímpar.

$$\begin{aligned}
\mathbb{E}[X_i] &= \Big[2(2/9) + 4(1/9)\Big] - \Big[1(5/18) + 3(1/6) + 5(1/18)\Big] \\
&= \frac{8}{9} - \frac{19}{18} \\
&= -\frac{1}{6}.
\end{aligned}$$

A esperança para o valor ganho em dez rodadas é

$$\begin{aligned}
\mathbb{E}[X] &= \mathbb{E}[X_1 + X_2 + \cdots + X_{10}] \\
&= \mathbb{E}[X_1] + \mathbb{E}[X_2] + \cdots + \mathbb{E}[X_{10}] && (E[\cdot] \text{ é linear}) \\
&= 10\mathbb{E}[X_1] && (\mathbb{E}[X_i] = \mathbb{E}[X_j]) \\
&= -\frac{10}{6} \\
&\approx -16.666,
\end{aligned}$$

e claramente o jogo não é vantajoso. ●

Definição 3.4.10 (variância de variável aleatória discreta)**.** A **variância** de uma variável aleatória discreta X é

$$\operatorname{var}(X) = \mathbb{E}[(X-\mu)^2]$$

onde $\mu = \mathbb{E}(X)$. ◀

Teorema 3.4.11. Se X é uma variável aleatória com esperança igual a μ, então $\operatorname{var}[X] = \mathbb{E}[X^2] - \mu^2$.

Demonstração.

$$\begin{aligned}
\operatorname{var}(X) &= \mathbb{E}[(X-\mu)^2] \\
&= \mathbb{E}[X^2 - 2\mu X + \mu^2] \\
&= \mathbb{E}[X^2] - 2\mu\mathbb{E}[X] + \mu^2 \\
&= \mathbb{E}[X^2] - 2\mu^2 + \mu^2 \\
&= \mathbb{E}[X^2] - \mu^2.
\end{aligned}$$ ■

Para calcular $\mathbb{E}[X^2]$, basta obsrevar que $\mathbb{E}[X^2] = \mathbb{E}[f(X)]$, com $f(k) = k^2$, e aplicar o Teorema 3.4.6:

$$\mathbb{E}[X^2] = \sum_i x_i^2 \Pr[X = x_i].$$

Exemplo 3.4.12. No Exemplo 3.1.2, a variável aleatória D, que dá a diferença entre os valores obtidos em dois dados, tem média $\mu = \frac{35}{18}$. A variância de D é, curiosamente,

$$\begin{aligned}
\operatorname{var}(D) = \mathbb{E}[D^2] - \mu^2 \\
&= 0^2(1/6) \\
&+ 1^2(5/18) \\
&+ 2^2(2/9) \\
&+ 3^2(1/6) \\
&+ 4^2(1/9) \\
&+ 5^2(1/18) \\
&- \left(\frac{35}{18}\right)^2 \\
&= \frac{665}{324} \\
&\approx 2.0524691358.
\end{aligned}$$

●

Definição 3.4.13 (desvio padrão)**.**

$$\sigma_X = \operatorname{SD}(X) = \sqrt{\sigma_X^2}.$$

◀

Exemplo 3.4.14. No Exemplo 3.4.12, calculamos a variância de uma variável aleatória D. Como $\operatorname{var}(D) = \frac{665}{324}$, então o desvio padrão dess variável é

$$\sigma_D = \sqrt{\frac{665}{324}} \approx 1.4326.$$

●

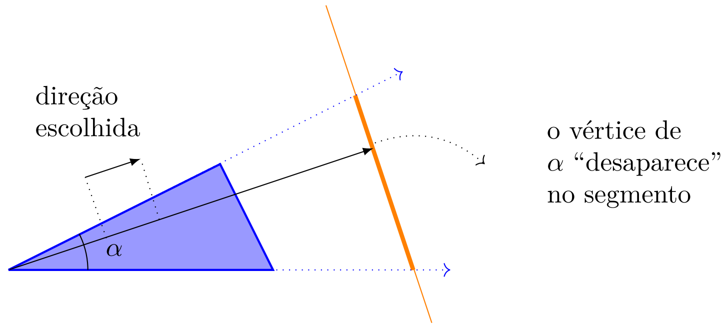
3.5 Método Probabilístico: linearidade da esperança

Retomamos agora o método probabilístico, para desenvolver uma demonstração de que a soma dos ângulos internos de um triângulo é igual a π. Na demonstração, determinamos algumas probabilidades em espaços amostrais contínuos, mas a partir delas definimos variáveis aleatórias discretas.

Exemplo 3.5.1 (Método probabilístico e linearidade da esperança: ângulos internos de um triângulo)**.** Podemos usar a linearidade da esperança para provar que a soma dos ângulos internos de um triângulo é igual a π (ou 180 graus).

Considere o seguinte experimento, realizado com um triângulo qualquer, tendo ângulos internos α, β, γ: escolha aleatoriamente uma direção entre 0 e 2π. Projete a imagem do triângulo em uma reta, ortogonal à direção escolhida.

Seja X a variável aleatória que indica a quantidade de vértices do triângulo que "desapareceram" (ou seja, não caíram na extremidade do segmento de reta projetado).



Claramente, X só poderá ser zero ou um: será zero quando a direção for paralela a um dos lados do triângulo; e será um quando não for. Como só há um número finito de lados, e uma quantidade não enumerável de direções,

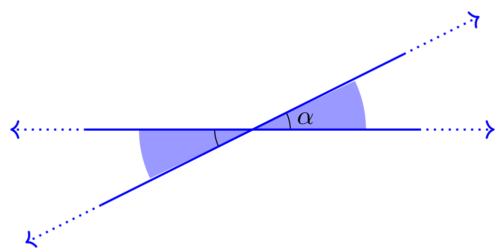
$$\Pr[X = 0] = 0,$$
$$\Pr[X = 1] = 1.$$

Assim,

$$\mathbb{E}[X] = 0 \cdot 0 + 1 \cdot 1 = 1.$$

Agora sejam A, B, C as variáveis que indicam quando cada um dos vértices desaparece ($A = 1$ quando o vértice do ângulo α desaparece, e $A = 0$ se ele não desaparece; semelhantemente para B E C). Tome A, por exemplo.

O vértice do ângulo α desaparece se e somente se a direção escolhida é uma dentre 2α possibilidades:



Como há 2α entre 2π direções que fazem o vértice desaparecer, a probabilidade é

$$\Pr[A = 1] = \frac{2\alpha}{2\pi} = \frac{\alpha}{\pi}.$$

Observamos que *a probabilidade de um vértice desaparecer é definida em espaço contínuo, mas a variável aleatória A é discreta, porque só pode valer zero ou um.*

Calculamos a esperança de A:

$$\mathbb{E}[A] = 1\left(\frac{\alpha}{\pi}\right) + 0\left(1 - \frac{\alpha}{\pi}\right) = \frac{\alpha}{\pi}$$

De maneira semelhante,

$$\mathbb{E}[B] = \frac{\beta}{\pi}$$
$$\mathbb{E}[C] = \frac{\gamma}{\pi}$$

Pela linearidade da esperança,

$$\begin{aligned} A + B + C &= X \\ \mathbb{E}[A + B + C] &= \mathbb{E}[X] \\ \mathbb{E}[A] + \mathbb{E}[B] + \mathbb{E}[C] &= 1 \\ \frac{\alpha}{\pi} + \frac{\beta}{\pi} + \frac{\gamma}{\pi} &= 1 \\ \alpha + \beta + \gamma &= \pi. \end{aligned}$$

A demonstração é válida para qualquer polígono convexo, sendo necessário apenas definir uma variável aleatória por vértice. ●

3.6 Experimentos de Bernoulli e Distribuição Binomial

Alguns exemplos anteriores abordaram a repetição de experimentos que resultam em "sucesso" ou "falha": o Exemplo 1.1.8, onde cartas são retiradas de um baralho até obter um naipe vermelho; e o Exemplo 3.4.4, onde uma moeda é jogada até se obter cara. O tipo de experimento que é repetido nesses exemplos – onde há somente dois resultados possíveis, com probabilidades conhecidas, é chamado de *experimento de Bernoulli*.

Definição 3.6.1 (experimento de Bernoulli)**.** um **experimento de Bernoulli** é um experimento aleatório com exatamente dois resultados possíveis, usualmente chamados de "*sucesso*" e "*falha*", sendo a probabilidade de sucesso a mesma cada vez que o experimento for repetido. ◀

Exemplo 3.6.2. O mais conhecido e mais simples exemplo de experimento de Bernoulli é aquele em que uma moeda – não necessariamente honesta – é jogada. Associamos uma das faces com "*sucesso*" e a outra com "*falha*". ●

Exemplo 3.6.3. No Exemplo 1.1.8, cartas são selecionadas de um baralho, com reposição, *até que uma seja obtida com naipe vermelho*.

Aquele experimento é composto de vários experimentos menores, onde uma carta é selecionada e seu naipe é registrado. Esse experimento menor é um experimento de Bernoulli, onde tratamos "naipe vermelho" como sucesso e "naipe preto" como falha. ●

Se uma variável aleatória X é definida a partir de um experimento de Bernoulli, e seus valores são 1 para sucesso, com probabilidade p, e 0 para falha, com probabilidade $1 - p$, então a função de massa de probabilidade de uma distribuição de Bernoulli é

$$\begin{aligned}\Pr(X = 1) &= p \\ \Pr(X = 0) &= 1 - p.\end{aligned}$$

Além disso, a esperança e a variância são determinadas pelo Teorema 3.6.4.

Teorema 3.6.4.

$$\begin{aligned}\mathbb{E}[X] &= p \\ \mathrm{var}(X) &= p(1 - p).\end{aligned}$$

Demonstração. Trivialmente, a esperança é

$$\mathbb{E}[X] = 1p + 0(1 - p) = p.$$

Já a variância é

$$\begin{aligned}\mathrm{var}(X) &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 \\ &= \mathbb{E}[X^2] - p^2 \\ &= \big(1^2 p + 0^2(1 - p)\big) - p^2 \\ &= p - p^2 \\ &= p(1 - p).\end{aligned}$$

∎

Corolário 3.6.1. Se X é uma variável de Bernoulli, então $0 \leq \sigma_x \leq 1/4$.

Exemplo 3.6.5. Uma moeda não honesta tem, ao ser jogada, probabilidade 0.25 de resultar em cara, e 0.75 de resultar em coroa. Se definirmos uma variável de Bernoulli X, onde sucesso é cara e falha é coroa, então teremos

$$\begin{aligned}\Pr[X = 1] &= 0.25 \\ \Pr[X = 0] &= 0.75\end{aligned}$$

E a esperança de X é

$$\mathbb{E}[X] = 0.25.$$

Já a variância de X é

$$\sigma_X = (0.25)(0.75) = 0.1875,$$

entre 0 e 0.25, como determina o Corolário 3.6.1. ●

Exemplo 3.6.6. Há uma população prestes a escolher o ocupante de um cargo importante, e há dois candidatos, A e B. Selecionaremos um indivíduo dessa população e consultaremos sua intenção de voto. A probabilidade dele escolher

A é 0.7, e a de escolher B é 0.3. Se definirmos uma variável aleatória de Bernoulli V, com sucesso sendo ao voto em A, e falha sendo voto em B, Teremos

$$\begin{aligned}\mathbb{E}[V] &= 0.7\\ \sigma_V &= (0.3)(0.7) = 0.21\end{aligned}$$

O fato da variância ficar entre zero e um quarto será importante na determinação de tamanho mínimo de amostra para uso do Teorema central do Limite, no Capítulo 6. ●

Uma pergunta importante sobre experimentos de Bernoulli é, "em n repetições, quantos sucessos devo esperar conseguir?".

Definição 3.6.7 (distribuição binomial)**.** Seja X a variável aleatória que contabiliza a quantidade de sucessos em uma sequência de n experimentos de Bernoulli, onde a probabilidade de sucesso é p. Dizemos que X tem **distribuição binomial** com parametros (n, p). ◀

Se X uma variável binomial tem parâmetros (n, p), então

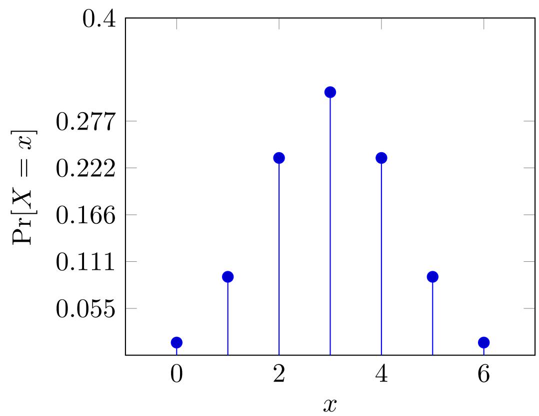
$$\Pr[X = k] = \binom{n}{k} p^k (1-p)^{n-k}.$$

A razão para o nome "binomial" fica clara ao observarmos a fórmula que determina $\Pr[X = k]$, onde há o uso dos coeficientes binomiais $\binom{n}{k}$.

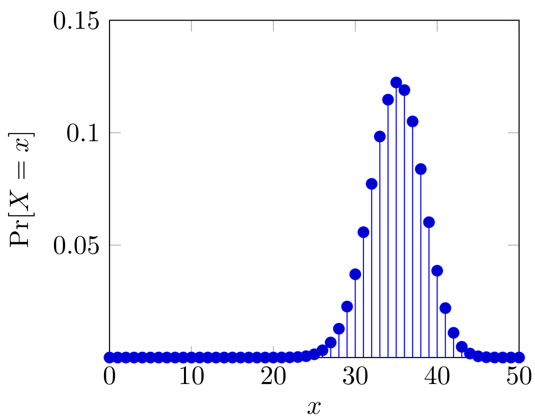
Quando $p = 0.5$, a distribuição binomial é simétrica. Suponha que $n = 6$ e $p = 0.5$. Então

$$\begin{aligned}
\Pr[X = 0] &= \binom{6}{0} 0.5^0 0.5^6 = (1)0.5^6 \approx 0.0156\\
\Pr[X = 1] &= \binom{6}{1} 0.5^1 0.5^5 = (6)0.5^6 \approx 0.0937\\
\Pr[X = 2] &= \binom{6}{2} 0.5^2 0.5^4 = (15)0.5^6 \approx 0.2343\\
\Pr[X = 3] &= \binom{6}{3} 0.5^3 0.5^3 = (20)0.5^6 \approx 0.3125\\
\Pr[X = 4] &= \binom{6}{4} 0.5^4 0.5^2 = (15)0.5^6 \approx 0.2343\\
\Pr[X = 5] &= \binom{6}{5} 0.5^5 0.5^1 = (6)0.5^6 \approx 0.09375\\
\Pr[X = 6] &= \binom{6}{6} 0.5^6 0.5^0 = (1)0.5^6 \approx 0.0156
\end{aligned}$$

O gráfico da função de massa de probabilidade é mostrado a seguir. O gráfico é simétrico ao redor da média, que neste caso é $np = 6(0.5) = 3$.



Quando p é diferente de 0.5, no entanto, a distribuição não é simétrica. A figura a seguir mostra o gráfico da função de massa de probabilidade da distribuição binomial para $p = 0.7$ e $n = 50$.



Exemplo 3.6.8. No Experimento do Exemplo 1.1.8, selecionamos cartas de um baralho até que uma seja obtida com naipe vermelho.

Decrevemos o experimento de outra forma: retiramos cartas de um baralho, com reposição, n vezes, e registramos a sequência de naipes. O espaço amostral é composto por tuplas de tamanho n: se $n = 7$, um elemento do espaço amostral seria (P, P, P, V, V, P, V). Simplificamos a notação, escrevendo a tupla como $PPPVVPV$. para $n = 3$, podemos listar o espaço amostral inteiro.

$$\Omega = \{PPP, PPV, PVP, PVV, VPP, VPV, VVP, VVV\}.$$

Tendo definido n, também definimos a variável X que determina a quantidade de naipes vermelhos na sequência de experimentos. Para $n = 3$, podemos enumerar todas as possíveis sequências de resultados.

$$\begin{aligned} X(PPP) &= 0 & X(VPP) &= 1 \\ X(PPV) &= 1 & X(VPV) &= 2 \\ X(PVP) &= 1 & X(VVP) &= 2 \\ X(PVV) &= 2 & X(VVV) &= 3 \end{aligned}$$

Calculamos $\Pr[V = 2]$. Como $\Pr(V) = \Pr(P) = 1/2$,

$$\begin{aligned} \Pr[X = 2] &= \binom{3}{2} \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^{3-2} \\ &= 3\left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right) \\ &= \frac{3}{8} \\ &= 0.375. \end{aligned}$$

Podemos, como n é pequeno, verificar que realmente, há 3 sequências com dois naipes vermelhos: VVP, VPV, PVV, e a quantidade total de possíveis sequências é oito – o que confirma que $\Pr[X = 2] = 8$. ●

Exemplo 3.6.9. Suponha que no Exemplo 3.6.8 seis cartas com naipe vermelho tenham sido removidas do baralho antes do início do experimento. As probabilidades agora são

$$\begin{aligned} \Pr(P) &= \frac{26}{50} = \frac{13}{25} = 0.52, \\ \Pr(V) &= \frac{22}{50} = \frac{11}{25} = 0.44. \end{aligned}$$

Então,

$$\begin{aligned} \Pr[X = 2] &= \binom{3}{2} \left(\frac{11}{25}\right)^2 \left(\frac{13}{25}\right)^{3-2} \\ &= 3\left(\frac{121}{625}\right)\left(\frac{13}{25}\right) \\ &= 3\frac{1573}{15625} \\ &= 0.302016, \end{aligned}$$

menor que o valor calculado no Experimento 3.6.8, (0.375). ●

Se X é uma variável aleatória com distribuição binomial e parâmetros n e p, então a função de distribuição acumulada de X é dada por

$$F_{n,p}(k) = \Pr(X < k) = \sum_{i=0}^{\lfloor k \rfloor} \binom{n}{i} p^i (1-p)^{n-i}$$

É possível obter uma forma fechada para esta função, mas o desenvolvimento dela se desvia demasiadamente do escopo deste texto.

Teorema 3.6.10. A esperança de uma variável aleatória X que tem distribuição binomial com parâmetros (n, p) é $\mathbb{E}[X] = np$.

Demonstração. Se a distribuição tem parâmetros (n, p), então identificamos n experimentos de Bernoulli, e definimos uma variável aleatória X_I para cada um deles. X_i é a variável igual a um quando o i-ésimo experimento tem sucesso, e zero quando o i-ésimo experimento falha. A variável

$$X = X_1 + X_2 + \cdots + X_n$$

contabiliza, portanto, a quantidade de sucessos, e tem distribuição binomial com parâmetros (n, p).

Para as variáveis X_i, a esperança é simplesmente

$$\mathbb{E}[X_i] = (1)(p) + (0)(1 - p) = p.$$

A esperança de X, portanto, é

$$\begin{aligned}\mathbb{E}[X] &= \mathbb{E}[X_1 + X_2 + \cdots + X_n] \\ &= \mathbb{E}[X_1] + \mathbb{E}[X_2] + \cdots + \mathbb{E}[X_n] \\ &= \underbrace{p + p + \cdots + p}_{n \text{ vezes}} = np.\end{aligned}$$ ■

Exemplo 3.6.11. Se jogarmos quinhentas vezes uma moeda viciada, com probabilidade de resultar em coroa igual a 0.4, a esperança da quantidade de coroas a serem obtidas é

$$np = 500(0.4) = 200.$$ ●

Exemplo 3.6.12. Um remédio tem efeito positivo 70% das vezes em que foi usado em um estudo. Um hospital começa a administrar esse medicamento em 300 pacientes.

Se usarmos o valor sugerido pelo estudo para a probabilidade de sucesso, a esperança da quantidade de pacientes que deve apresentar melhora é

$$np = 300(0.7) = 210.$$ ●

O Teorema 3.6.13 dá uma forma fechada para a variância de variáveis binomiais. Sua demonstração é pedida no Exercício 57.

Teorema 3.6.13. A variância de uma variável aleatória X que tem distribuição binomial com parâmetros (n, p) é $\text{var}(X) = np(1 - p)$.

Exemplo 3.6.14. O Exemplo 3.6.12 considera uma medicação com probabilidade de sucesso 0.7 administrada em 300 pacientes, com esperança de 210 sucessos. A variância será

$$\sigma^2 = np(1 - p) = 300(0.7)(0.3) = 63,$$

e o desvio padrão,

$$\sigma = \sqrt{63} \approx 7.937.$$ ●

3.7 Poisson

Há fenômenos que são semelhantes a experimentos de Bernoulli, mas ao invés de ocorrer em uma quantidade discreta e (relativamente) pequena de experimentos, ocorre ao longo de um contínuo (ou de uma quantidade muito grande de experimentos), com probabilidade baixa de ocorrer. Como exemplos, temos o número de erros por página de um livro; quantidade de chamadas telefônicas recebidas em uma hora; quantidade de pessoas entrando em uma loja em um período de tempo (uma hora, por exemplo). Em todos esses casos, podemos tentar usar a distribuição binomial na modelagem: dividimos o tempo (ou a quantidade de páginas) em partições, e olhamos para cada partição como um expermiento de Bernoulli. Por exemplo, para determminar a probabilidade de alguém entrar em uma loja em um período de duas horas, sabendo a média de pessoas por 30 min, dividimos o período de duas horas em quatro partes de 30 min, e tratamos cada 30 min como um experimento de Bernoulli. Cada período isolado é tratado como independente. Queremos, agora, determinar o que ocorre quando tornamos esses períodos muito pequenos. Teremos a probabilidade de sucesso muito pequena, e a quantidade de experimentos de Bernoulli muito grande. Podemos, portanto, considerar a distribuição binomial com n muito grande e p pequeno.

Por conveniência, denotaremos por λ a esperança de uma variável binomial:

$$\lambda = \mathbb{E}(X) = np.$$

Mas se $\lambda = np$, então

$$p = \frac{\lambda}{n}.$$

Reescrevemos a probabilidade de X ser igual a k:

$$\begin{aligned}
\Pr[X = k] &= \binom{n}{k} p^k (1-p)^{n-k} \\
&= \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\
&= \frac{n!}{k!(n-k)!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}
\end{aligned}$$

Quando n tende a ∞ e p é pequeno,

$$\begin{aligned}
\Pr[X=k] &= \frac{n(n-1)(n-2)\cdots(n-k+1)}{k!}\left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^n \left(1-\frac{\lambda}{n}\right)^{-k} \\
&= \frac{n(n-1)(n-2)\cdots(n-k+1)}{k!}\frac{\lambda^k}{n^k}\left(1-\frac{\lambda}{n}\right)^n \left(1-\frac{\lambda}{n}\right)^{-k} \\
&= \frac{n(n-1)(n-2)\cdots(n-k+1)}{n!}\frac{\lambda^n}{k^n}\left(1-\frac{\lambda}{n}\right)^{n-k} \\
&= \underbrace{\frac{n}{n}\frac{(n-1)}{n}\frac{(n-2)}{n}\cdots\frac{(n-k+1)}{n}}_{\to 1}\underbrace{\frac{\lambda^n}{k^n}}_{\to e^{-\lambda}}\underbrace{\left(1-\frac{\lambda}{n}\right)^{n-k}}_{\to 1}
\end{aligned}$$

Assim, presumindo que p é pequeno,

$$\lim_{n\to\infty}\Pr[X=k] = \frac{\lambda^k}{k!}e^{-\lambda}$$

O que fizemos foi estabelecer que, quando n é miuto grande e p muito pequeno (ou mais precisamente, quando $n\to\infty$, e consequentemente $p\to 0$), sendo $np=\lambda$, as probabilidades determinadas pela distribuição binomial para $\Pr[X=k]$ tendem a

$$\frac{\lambda^k}{k!}e^{-\lambda}$$

Definição 3.7.1 (distribuição de Poisson)**.** Uma variável aleatória X tem **distribuição de Poisson** se

$$\Pr[X=k] = \frac{\lambda^k}{k!}\,e^{-\lambda}$$

Denotamos a distribuição de Poisson com parâmetro λ por $\mathcal{P}oiss(\lambda)$. ◀

Verificamos que a distribuição de Poisson é uma distribuição de probabilidades bem definida: a soma de todas as possíveis probabilidades deve ser um.

$$\begin{aligned}
\sum_{k=0}^{\infty}\Pr[X=k] &= \sum_{k=0}^{\infty}\frac{\lambda^k e^{-\lambda}}{k!} \\
&= e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!} \\
&= e^{-\lambda}e^{\lambda} && (\text{Taylor: } e^x = 1+x+\tfrac{x^2}{2!}+\tfrac{x^3}{3!}+\cdots) \\
&= 1.
\end{aligned}$$

Há algumas condições que devem ser observadas ao presumir que uma variável aletória tem distribuição de Poisson:

- o evento pode ser contabilizado em números inteiros;
- as ocorrências são independentes;
- a frequência média de ocorrências por intervalo de tempo é conhecida;
- dois eventos não ocorrem ao mesmo tempo.

Exemplo 3.7.2 (distribuição de Poisson: ciclistas passando em via)**.** Observamos ciclistas em um determinado ponto de uma ciclovia, e registramos 120 a cada hora. A média de ciclistas por minuto é $120/60 = 2$.

Presumimos distribuição de Poisson.

A probabilidade de nenhum ciclista passar em um período de dois minutos é

$$\begin{aligned}\frac{e^{-\mu}\mu^k}{k!} &= \frac{e^{-4}4^0}{0!}\\ &= e^{-4} \approx 0.0183156.\end{aligned}$$

A probabilidade de exatamente três ciclistas passarem nesse mesmo período é

$$\begin{aligned}\frac{e^{-\mu}\mu^k}{k!} &= \frac{e^{-4}4^3}{3!}\\ &= \frac{64e^{-4}}{3!}\\ &\approx 0.1953668.\end{aligned}$$

●

Exemplo 3.7.3 (distribuição de Poisson: protocolo ALOHA)**.** Entre 1968 e 1971 a Universidade do Havaí desenvolveu um protocolo de rede sem fio chamado ALOHA[1], que foi também usado em redes cabeadas. Do protocolo ALOHA vários outros foram derivados, inclusive alguns dos protocolos modernos para telefonia celular.

A primeira versão do ALOHA é bastante simples: quando uma estação precisa transmitir um pacote de rede, ela o faz imediatamente, sem verificar antes se o canal está em uso. Se, ao transmitir, ela detectar que há outra estação transmitindo, ela desiste e tenta novamente um curto tempo mais tarde. Esse tempo precisa ser aleatório, porque de outra forma as colisões se repetiriam mais tarde, e os pacotes nunca seriam transmitidos.

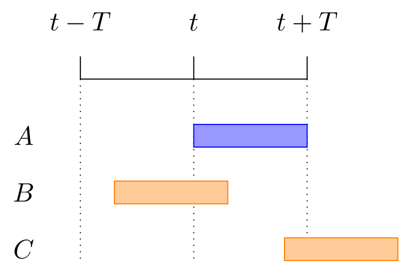
Uma das coisas interessantes que podemos fazer é calcular a vazão máxima (throughput) do protocolo. Para isso é necessário, por simplicidade, presumir que os pacotes sempre são do mesmo tamanho, e o tempo necesário para enviar um deles é fixo, que denotaremos T. Damos o nome de "período" (ou "frame") a esse tempo. Além disso, a quantidade de estações tentando enviar pacotes é modelada por uma distribuição de Poisson.

A probabilidade de k tentativas simultâneas de envio em um período é

$$\frac{\lambda^k e^{-\lambda}}{k!}.$$

[1] "Additive Links On-line Hawaii Area".

Se uma estação começa a transmitir em um tempo t, qualquer transmissão que inicie entre $t-T$ e $t+T$ causará colisão. Este tempo (que tem duração $2T$ – dois períodos) é chamado de *tempo vulnerável*. Isto é ilustrado na figura a seguir: para que a estação A possa transmitir, B não pode iniciar depois de $t-T$, e C não pode iniciar antes de $t+T$.



No período vulnerável (dois períodos consecutivos), a probabilidade de k estações transmitirem é

$$\frac{(2\lambda)^k e^{-2\lambda}}{k!}.$$

Assim, a probabilidade de sucesso de uma transmissão é a de nenhuma outra estação transmitir naquele período:

$$e^{-2\lambda}.$$

O throughput da rede é a quantidade de tentativas (λ) multiplicada pela probabilidade de sucesso, portanto

$$(\lambda)\left(e^{-2\lambda}\right).$$

Esse valor atinge um máximo de 0.1839 envios por período quando $\lambda = 1/2$ (é uma exponencial, portanto tem um único ponto crítico, que é o máximo – basta derivar a função e igualar a zero). Ou seja, no máximo 18.39% do tempo é usado para transmissões com sucesso.

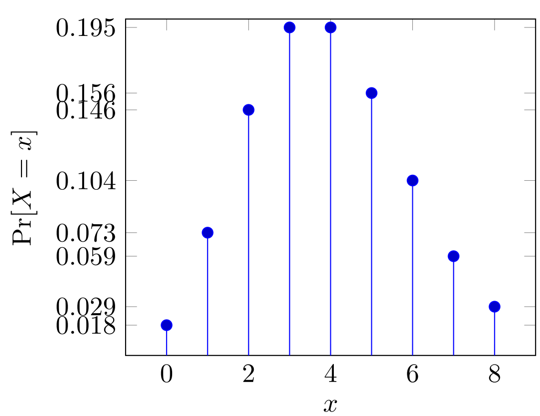
Uma versão posterior do ALOHA mais eficiente foi desenvolvida, dividindo o tempo em *slots* fixos. ●

Calculamos agora os valores de $\Pr[X = k]$, para k entre zero e oito, com

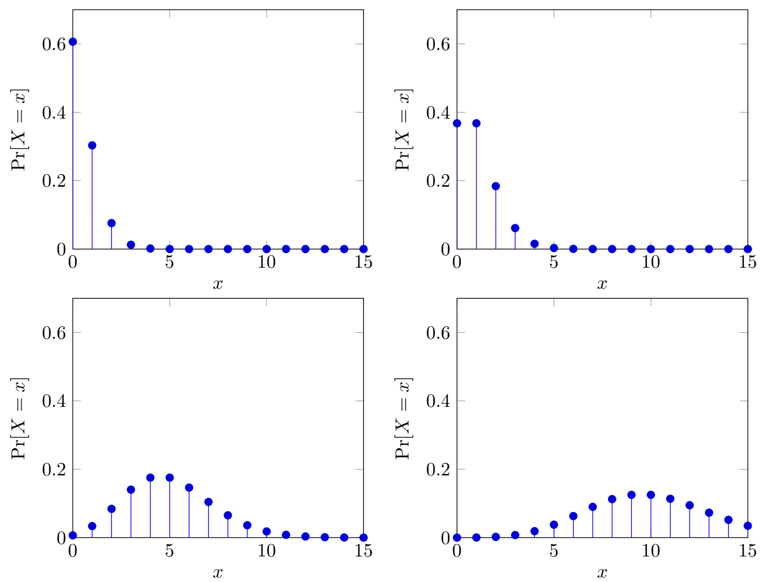
$\lambda = 4$, como no Exemplo 3.7.2.

$$\begin{aligned}
\frac{e^{-4}4^0}{0!} &= e^{-4} \approx 0.0183\\
\frac{e^{-4}4^1}{1!} &= 4e^{-4} \approx 0.0732\\
\frac{e^{-4}4^2}{2!} &= 8e^{-4} \approx 0.1465\\
\frac{e^{-4}4^3}{3!} &= (32/3)e^{-4} \approx 0.1953\\
\frac{e^{-4}4^4}{4!} &= (32/3)e^{-4} \approx 0.1953\\
\frac{e^{-4}4^5}{5!} &= (128/15)e^{-4} \approx 0.1562\\
\frac{e^{-4}4^6}{6!} &= (256/45)e^{-4} \approx 0.1041\\
\frac{e^{-4}4^7}{7!} &= (1024/315)e^{-4} \approx 0.0595\\
\frac{e^{-4}4^8}{8!} &= (512/315)e^{-4} \approx 0.0297
\end{aligned}$$

A seguir está a gráfico da função de massa de probabilidade.



A próxima figura mostra oa gráficos das funções de massa de probabilidade de variáveis com distribuição de Poisson, para λ igual a 0.5, 1, 5 e 10. Quanto maior o valor de λ, mais o gráfico fica baixo e com o valor máximo mais deslocado para a direita.



Teorema 3.7.4. Seja X uma variável aleatória com distribuição de Poisson com parâmetro λ. Então

$$\mathbb{E}(X) = \lambda.$$

Demonstração. Este foi o ponto de partida para definirmos a distribuição de Poisson (*"Por conveniência, denotaremos por λ a esperança de uma variável binomial"* – no início desta seção). ■

Exemplo 3.7.5. Uma cultura em meio líquido contém em média 2000 bactérias por ml. Se tomarmos uma amostra de 1 ml, a probabilidade de não haver bactérias nela é

$$\frac{e^{-2000}2000^0}{0!} = e^{-2000}.$$

Se quisermos saber a probabilidade de haver bactérias em uma amostra de 10 ml, observamos que, se há em média 2000 bactérias por ml, também encontraremos em média 20000 em cada 10 ml. Assim, para determinar a probabilidade de haver bactérias na amostra temos dois caminhos possíveis:

- calcular, para todo $j > 0$, $\Pr[k = j]$. Ou seja, calculamos $\sum_{j=1}^{\infty} \Pr[k = j]$.

$$\begin{aligned}
\sum_{j=1}^{\infty} \frac{e^{-20000} 20000^j}{j!} &= e^{-20000} \sum_{j=1}^{\infty} \frac{20000^j}{j!} \\
&= e^{-20000} \left(\frac{20000^1}{1!} + \frac{20000^2}{2!} + \frac{20000^3}{3!} + \dots \right) \\
&= e^{-20000} \left(-1 + 1 + \frac{20000^1}{1!} + \frac{20000^2}{2!} + \frac{20000^3}{3!} + \dots \right) \\
&= e^{-20000} \left[-1 + \left(1 + \frac{20000^1}{1!} + \frac{20000^2}{2!} + \frac{20000^3}{3!} + \dots \right) \right] \\
&= e^{-20000} (-1 + e^{20000}) \\
&= -e^{-20000} + 1.
\end{aligned}$$

- como só nos interessam os casos diferentes de zero bactérias, calcular a probabilidade do complemento do evento “sem bactérias”:

$$1 - \frac{e^{-20000} 20000^0}{0!} = 1 - e^{-20000},$$

o mesmo resultado que pelo outro método, com muito menos trabalho.

Há uma observação importante a respeito deste exemplo: quando dissemos que o meio líquido “contém em média 2000 bactérias por ml”, e depois procuramos a probabilidade de não haver bactérias, usamos $\lambda = 2000$. Poderíamos ter tentado simplificar o problema, usando 2 ao invés de 2000: o meio “contém em média 2 ’lotes’ de 1000 bactérias por ml”. No entanto, ao usar $\lambda = 2$, estaríamos contando a quantidade de “lotes inteiros” de 1000 bactérias, e nosso modelo não nos permitiria fazer adequadamente afirmações sobre uma quantidade de 500 bactérias por ml, por exemplo. ●

Exemplo 3.7.6. Uma fábrica produz chapas de metal, e é sabido que a média de pequenas irregularidades nelas é de uma a cada $2m^2$. Sabendo que as chapas serão cortadas em pedaços de $50cm^2$, determinaremos a probabilidade de uma chapas ser produzida com exatamente uma ou duas irregularidades.

Presumimos que a distribuição é de Poisson, com média igual a $1/16$ em cada $50cm^2$. A probabilidade que procuramos é

$$\begin{aligned}
\frac{e^{-1/16}(1/16)^1}{1!} + \frac{e^{-1/16}(1/16)^2}{2!} &= e^{-1/16} \left(\frac{1}{16} + \frac{1}{2(16^2)} \right) \\
&= e^{-1/16} \frac{33}{512} \\
&\approx 0.060.
\end{aligned}$$ ●

A função cumulativa de massa da distribuição de Poisson é dada por

$$F_{k,\lambda}(z) \Pr(X \leq z) = \sum_{i=0}^{\lfloor z \rfloor} \frac{\lambda^k}{k!} e^{-\lambda}.$$

Assim como no caso da distribuição binomial, é possível obter uma forma fechada para esta função, cujo desenvolvimento também sai do escopo deste texto.

Teorema 3.7.7. Seja X uma variável aleatória com distribuição de Poisson com parâmetro λ. Então

$$\mathrm{var}(X) = \lambda.$$

Demonstração. Sabemos que

$$\mathrm{var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

Usaremos um fato trivial: como

$$x(x-1) + x = x^2 - x + x = x^2,$$

podemos reescrever

$$\begin{aligned}\mathbb{E}[X^2] &= \mathbb{E}[X(X-1) + X] \\ &= \mathbb{E}[X(X-1)] + \mathbb{E}[X].\end{aligned}$$

Calculamos $\mathbb{E}[X(X-1)]$:

$$\begin{aligned}\mathbb{E}[X(X-1)] &= \sum_{k=0}^{\infty} k(k-1)\frac{\lambda^k e^{-\lambda}}{k!} \\ &= e^{-\lambda}\sum_{k=0}^{\infty} k(k-1)\frac{\lambda^k}{k!} \\ &= e^{-\lambda}\sum_{k=0}^{\infty} \frac{\lambda^k}{(k-2)!} \\ &= \lambda^2 e^{-\lambda}\sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} \\ &= \lambda^2 e^{-\lambda}\sum_{j=0}^{\infty} \frac{\lambda^j}{j!} && (j = k-2) \\ &= \lambda^2 e^{-\lambda} e^{\lambda} \\ &= \lambda^2.\end{aligned}$$

Continuamos, portanto:

$$\begin{aligned}\mathrm{var}[X] &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 \\ &= \mathbb{E}[X(X-1)] + \mathbb{E}[X] - \mathbb{E}[X]^2 \\ &= \lambda^2 + \lambda - \lambda^2 \\ &= \lambda.\end{aligned}$$

■

Exemplo 3.7.8. No Exemplo 3.7.5, a média de bactérias por ml era 20000. Isso significa que a variância é também 20000, e o desvio padrão é

$$\sqrt{20000} = 100\sqrt{2} \approx 141.4213.$$

●

3.8 Geométrica

A distribuição binomial trata da quantidade de sucessos em n experimentos de Bernoulli. Podemos querer saber, ao invés disso, a *quantidade de repetições necessárias* até obter sucesso em um experimento. Se uma variável aleatória tem esse valor, dizemos que ela tem *distribuição geométrica*.

Definição 3.8.1 (distribuição geométrica)**.** Se X é uma variável aleatória que determina a quantidade de experimentos de Bernoulli necessários até que um sucesso ocorra, dizemos que X tem **distribuição geométrica**. ◀

Se X é uma variável aleatória com distribuição geométrica, associada com um experimento de Bernoulli onde a probabilidade de sucesso é p, então

$$\Pr(X = k) = (1-p)^{k-1}p.$$

A equação da função de massa de probabilidade da distribuição geométrica é a de uma função exponencial (e, de fato, a distribuição geométrica é análoga a uma distribuição contínua chamada de "exponencial"). A distribuição tem o nome de "geométrica" porque as probabilidades para $k = 0, 1, 2, \ldots$ formam uma progressão geométrica.

Se quisermos usar uma definição levemente diferente, removendo a possibilidade de $k = 0$, teremos

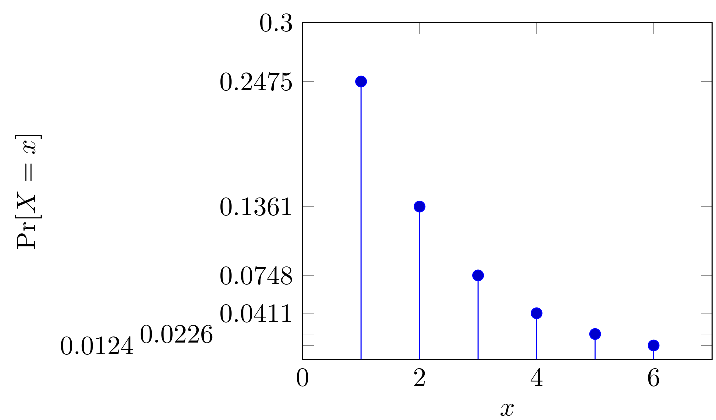
$$\Pr(X = k) = (1-p)^{k}p.$$

A definição a ser usada deve, no entanto, servir ao problema sendo modelado.

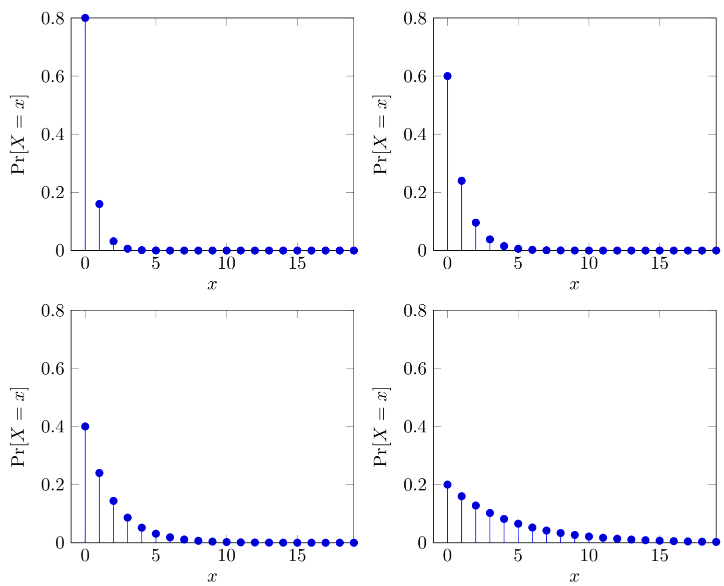
Exemplo 3.8.2. Uma moeda não honesta tem probabilidade 0.45 de, ao ser jogada, resultar em cara (e evidentemente, 0.55 de resultar em coroa). Se quisermos saber a probabilidade de obter cara na quinta vez (e não antes), definimos uma variável aleatória A, que nos dá o número da quantidade de tentantivas até obter cara, *e que tem distribuição geométrica*:

$$\begin{aligned}\Pr[A = 5] &= (1-p)^{k-1}p \\ &= (0.55)^4(0.45) \approx 0.0411778.\end{aligned}$$ ●

A Figura a seguir mostra a função de massa de probabilidade para $p = 0.55$ e $1 - p = 0.45$, como no Exemplo 3.8.2, com k variando de 1 a 6.



A figura aseguir mostra os gráficos das funções de massa de probabilidade de vraiáveis geométricas com p igual a $0.2, 0.4, 0.6$ e 0.8.



Exemplo 3.8.3. Considere o experimento onde jogamos dois dados distintos. Se definirmos que um jogador tem sucesso quando a soma de seus dados é seis, temos um experimento de Bernoulli.

Damos o nome de S ao evento "sucesso":

$$S = \{(1, 5), (2, 4), (3, 3), (4, 2), (5, 1)\}.$$

Como $|S| = 5$ e $|\Omega| = 36$, as probabilidades de sucesso e de falha em uma rodada do experimento são

$$\begin{aligned}\Pr(S) &= \frac{5}{36},\\ \Pr(\overline{S}) &= \frac{31}{36}.\end{aligned}$$

Podemos definir, também, uma variável aleatória X que nos dá a quantidade de rodadas até que o jogador tenha sucesso, e perguntamos qual é a probabilidade do jogador obter sucesso exatamente na quarta tentativa?

A variável X tem distribuição geométrica, portanto

$$\Pr[X = 4] = \left(\frac{31}{36}\right)^3 \left(\frac{1}{36}\right) = \left(\frac{29791}{46656}\right)\left(\frac{1}{36}\right) \approx 0.01773679. \qquad \bullet$$

Se quisermos saber quantas vezes temos que jogar um dado até obter um certo resultado, a intuição nos diz que *não deve ser importante quantas vezes já jogamos o dado, ou quantas vezes já obtivemos sucesso – podemos começar a contar as repetições a qualquer momento.* E de fato, esta é uma propriedade da distribuição geométrica: ela não tem memória.

Teorema 3.8.4. A distribuição geométrica não tem memória. Se X é uma variável com distribuição geométrica, então, para quaisquer a, b inteiros positivos,

$$\Pr[X \geq a + b | X > a] = \Pr[X \geq b].$$

Demonstração.

$$\begin{aligned}\Pr[X \geq a + b | X > a] &= \frac{\Pr[X \geq a + b, X \geq b]}{\Pr[X \geq a]}\\ &= \frac{\Pr[X \geq a + b]}{\Pr[X \geq a]} && (X = a + b \Rightarrow X \geq b)\\ &= \frac{(1-p)^{a+b}}{(1-p)^a}\\ &= (1-p)^b\\ &= \Pr[X \geq b],\end{aligned}$$

Conforme o enunciado. ∎

Exemplo 3.8.5. O Exemplo 3.6.12 descreve uma droga com probabilidade de sucesso igual a 0.7. A probabilidade de que tenhamos que administrar a droga a mais que 4 pacientes até observer efeito positivo é

$$\Pr[X \geq 4] = (0.3)^4(0.7) = 0.00567.$$

Essa probabilidade não depende de quantas vezes já administramos a medicação a outros pacientes. Se já tivermos tratado 50 deles,

$$\Pr[X \geq 54 | X > 50] = \Pr[X \geq 4] = 0.00567. \qquad \bullet$$

Teorema 3.8.6. Se uma variável aleatória X tem distribuição geométrica e parâmetro p, com $\Pr(X = k) = (1-p)^{k-1}p$, então a função de distribuição acumulada de massa para X é

$$F(k) = \Pr(X \leq k) = 1 - (1-p)^k.$$

O Teorema 3.8.6 tem demonstração bastante simples.

Teorema 3.8.7. Se X é uma variável aleatória geométrica relacionada a experimentos de Bernoulli com probabilidade de sucesso $p \in (0,1)$, então

$$\mathbb{E}[X] = \frac{1}{p}.$$

Demonstração. Usaremos a igualdade a seguir: para todo $p \in [0,1]$,

$$\sum_{j=n}^{\infty} (1-p)^j = \frac{(1-p)^n}{p}.$$

Agora calculamos

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{k=1}^{\infty} k \Pr[X = k] \\
&= \sum_{k=1}^{\infty} kp(p-1)^{k-1} \\
&= p \sum_{k=1}^{\infty} k(p-1)^{k-1} \\
&= p \left[\sum_{k=1}^{\infty} (1-p)^{k-1} + \sum_{k=2}^{\infty} (1-p)^{k-1} + \sum_{k=3}^{\infty} (1-p)^{k-1} + \cdots \right] \\
&= p \left[\sum_{j=0}^{\infty} (1-p)^j + \sum_{j=1}^{\infty} (1-p)^j + \sum_{j=2}^{\infty} (1-p)^j + \cdots \right] \\
&= p \left(\frac{1}{p} + \frac{1-p}{p} + \frac{(1-p)^2}{p} + \cdots \right) \\
&= 1 + (1-p) + (1-p)^2 + \cdots \\
&= \sum_{j=0}^{\infty} (1-p)^j \\
&= \frac{1}{p}.
\end{aligned}$$

■

Exemplo 3.8.8. Se usarmos a medicação descrita no Exemplo 3.6.12, que tem probabilidade de sucesso 0.7, repetidamente em diversos pacientes, a esperança da quantidades de tentativas até que um paciente com melhora seja observado é

$$\mathbb{E}[X] = \frac{1}{0.7} \approx 1.428.$$ ●

Teorema 3.8.9. Se X é uma variável aleatória geométrica relacionada a experimentos de Bernoulli com probabilidade de sucesso $p \in (0,1)$, então

$$\mathrm{var}(X) = \frac{1-p}{p^2}.$$

Exemplo 3.8.10. Se usarmos a medicação descrita no Exemplo 3.6.12, que tem probabilidade de sucesso 0.7, repetidamente em diversos pacientes, a esperança da quantidades de tentativas até que um paciente com melhora seja observado é de aproximadamente 1.428, com variância

$$\sigma^2 = \frac{0.3}{0.7^2} \approx 0.612,$$

e desvio padrão

$$\sigma = \sqrt{0.612} \approx 0.782.$$ ●

Exemplo 3.8.11. No experimento do Exemplo 3.8.3, a média de rodadas necessárias para que o jogador obtenha sucesso é

$$\mathbb{E}[X] = \frac{1}{5/36} = 7.2,$$

e a variância é

$$\mathrm{var}(X) = \frac{31/36}{(5/36)^2} = 44.64.$$

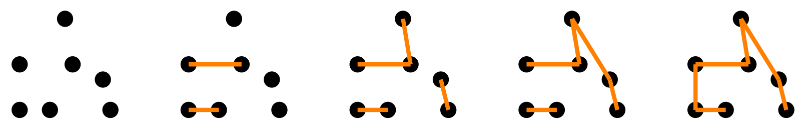
O desvio padrão de X é, portanto $\sqrt{44.64} \approx 6.681317$. ●

3.9 Polímeros e a distribuição de Flory-Schulz

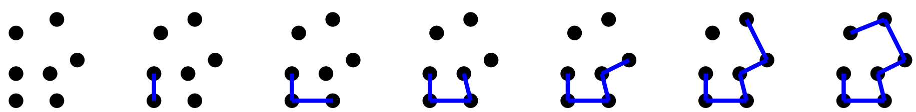
Polimerização é um tipo de processo químico em que moléculas de tamanho relativamente pequeno (monômeros) são agregadas, formando outras moléculas cada vez maiores (oligômeros e, como objetivo final, polímeros são formados). Em 1936 Paul Flory publicou uma análise da distribuição de massa das moléculas em polímeros, determinando a probabilidade da formação de moléculas com cada possível tamanho.

Há vários tipos de reação de polimerização. em uma delas, chamada de *polimerização em etapas*, os monômeros reagem para formar oligômeros (cadeias de diferentes tamanhos), que por sua vez também reagem para formar cadeias maiores, até formar polímeros (cadeias muito longas). Neste tipo de reação de polimerização, oligômeros de qualquer comprimento podem reagir para formar outro maior – ao contrário da *polimerização em cadeia*, onde somente acontece a adição de monômeros a uma cadeia sendo formada.

A primeira figura a seguir ilustra o processo de polimerização em etapas; a segunda ilustra o processo de polimerização em cadeia. Cada ponto representa um monômero, e os pontos ligados por linhas representam oligômeros sendo formados.



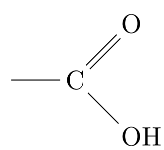
polimerização em etapas



polimerização em cadeia

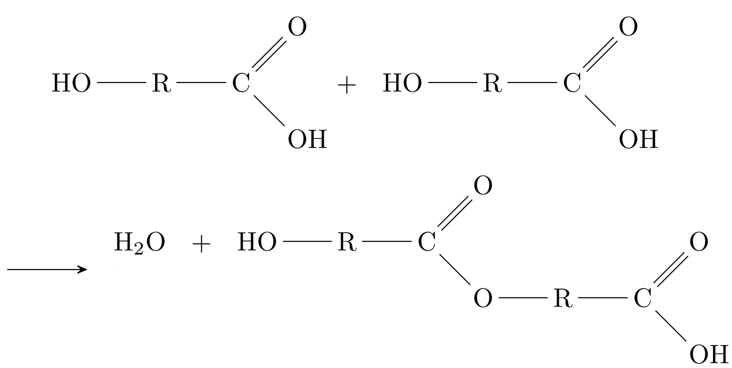
A discussão a seguir vale para os dois tipos de processo de polimerização.

Um grupo funcional é uma estrutura química que confere certas propriedades à molécula em que estão inseridos. Por exemplo, a *carboxila* é a estrutura $-COOH$,



e a *hidroxila* é a estrutura $-OH$.

Quando monômeros tem estes dois grupos funcionais, eles reagem para formar polímeros. Uma reação inicial com dois monômeros é ilustrada a seguir, onde R representa um grupo qualquer que pode ser a cadeia carbônica do monômero ou do próprio polímero, já que só nos interessam os grupos funcionais e a maneira como reagem.



Após várias reações, k monômeros formam um k-oligômero (ou "k-mero"):

$$\mathrm{HO{-}R{-}COOH} \xrightarrow{\text{-}\mathrm{H_2O}} \mathrm{H(O{-}R{-}CO)_k OH}$$

No exemplo acima uma molécula de água é liberada no processo; isto pode ou não acontecer dependendo do processo de polimerização mas no caso de poliésteres sempre ocorrerá. Entretanto, como a molécula liberada tem peso muito pequeno comparado ao das moléculas de polímreos, seu impacto na análise é desprezível. A análise é simplificada presumindo que os grupos funcionais das moléculas (monômeros ou oligômeros, independentemente do tamanho) tem igual probabilidade de reagir – este é chamado de *princípio de igual reatividade* de Flory.

Seguiremos o raciocínio de Flory: os monômeros são chamados de "segmentos" que reagem para formar polímeros. Definimos:

- N_0 é o número total de segmentos
- N é o número de segmentos que não reagiram
- $N_0 - N$ é o número de segmentos que reagiram
- A extensão da reação – a fração dos grupos que reagiram em um dado momento – é
$$p = \frac{N_0 - N}{N_0}$$

Escolhemos um segmento ao acaso. Calcularemos a probabilidade (que Flory denotou por Π_k) deste segmento ser parte de um oligômero com k segmentos.

Como presumimos que os grupos funcionais que não reagiram tem igual probabilidade de reagir, a probabilidade de uma reação ter acontecido ligando dois segmentos em um ponto b é p, e a probabilidade de não ter ocorrido reação ali é $1 - p$.

Seja Π_k^1 a probabilidade de um segmento ser o *primeiro* de um k-oligômero. Π_k^1 é dada por uma distribuição parecida com a geométrica, $\Pi_k^1 = p^{k-1}(p-1)^2$. É necessário computar a probabilidade de não-reação, $p - 1$, duas vezes para levar em conta as *duas* "pontas" ou extremidades da cadeia do oligômero. Mas o segmento pode não ser o primeiro da sequência; há k posições possíveis para ele, portanto
$$\Pi_k = kp^{x-1}(1-p)^2$$
é a probabilidade de um segmento pertencer a um oligômero de tamanho k.

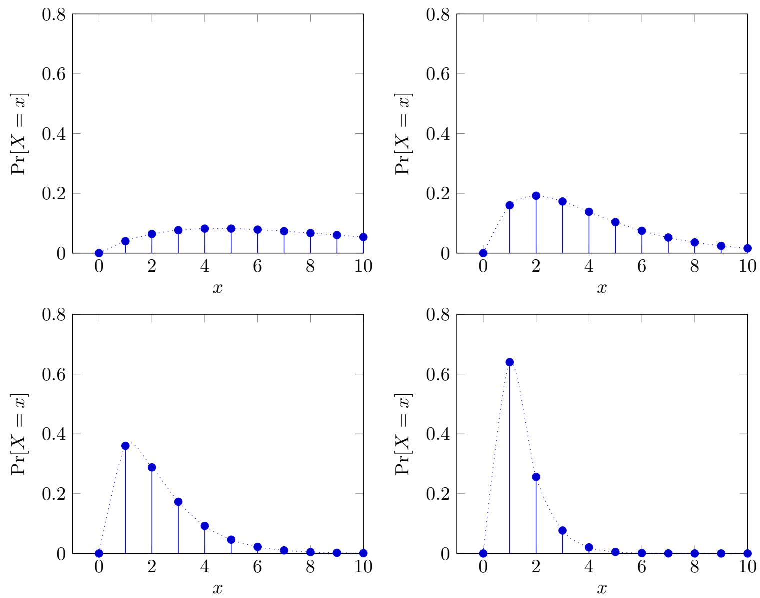
Π_k é de fato uma função de distribuição de massa:

Teorema 3.9.1. $\sum_{j=0}^{\infty} \Pi_k = 1$.

A demonstração do Teorema 3.9.1 é pedida no Exercício 67.

A massa de um oligômero é proporcional à quantidade de monômeros que ele tem; assim, como
$$\Pi_k = \frac{\text{segmentos em } k\text{-oligômeros}}{N_0},$$
podemos tratar Π_k como a fração de massa de k-oligômeros, quando comparada com o total.

A seguir estão gráficos da distribuição para $p = 0.3$, $p = 0.5$ e $p = 0.7$.



Teorema 3.9.2. Se X tem distribuição de Flory-Schulz, então a função acumulada de massa de X é

$$F(x) = 1 - p^x(1 + x - px)$$

A demonstração do Teorema 3.9.2 é pedida no Exercício 68.

Teorema 3.9.3. Se X tem distribuição de Flory-Schulz, então

$$\mathbb{E}[X] = \frac{-p-1}{p-1}.$$

Demonstraremos o Teorema 3.9.3 usando o valor da soma parcial $\sum_{x=1}^{n} x^2 p^{x-1}$ e calculando seu limite quando $x \to \infty$. Esta demonstração, apesar de trabalhosa, é bastante elementar.

Lema 3.9.1. Se $|p| < 1$, então

$$\sum_{x=1}^{n} x^2 p^{x-1} = \frac{-2n^2p^{n+1} + n^2p^{n+2} + n^2p^n - 2np^{n+1} + p^{n+1} + 2np^n + p^n - p - 1}{(p-1)^3}, \tag{3.1}$$

e o limite da soma quando $x \to \infty$ é

$$\lim_{n\to\infty} \frac{-2n^2p^{n+1} + n^2p^{n+2} + n^2p^n - 2np^{n+1} + p^{n+1} + 2np^n + p^n - p - 1}{(p-1)^3} = \frac{-p-1}{(p-1)^3}.$$

A primeira parte do Lema 3.9.1 (equação 3.1) pode ser demonstrada por indução em x. Para a segunda parte, basta separar a soma em dois termos,

$$\lim_{n\to\infty} \frac{-2n^2p^{n+1} + n^2p^{n+2} + n^2p^n - 2np^{n+1} + p^{n+1} + 2np^n + p^n - p - 1}{(p-1)^3}$$
$$= \lim_{n\to\infty} \frac{-2n^2p^{n+1} + n^2p^{n+2} + n^2p^n - 2np^{n+1} + p^{n+1} + 2np^n + p^n}{(p-1)^3} + \frac{-p-1}{(p-1)^3}$$

e verificar que o limite para o numerador do primeiro termo é zero.

A seguir provamos o Teorema 3.9.3.

Demonstração. Usando o Lema 3.9.1:

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{x=1}^{\infty} x f(x) \\
&= \sum_{x=1}^{\infty} x\, x p^{x-1}(1-p)^2 \\
&= (1-p)^2 \sum_{x=1}^{\infty} x^2 p^{x-1} \\
&= (1-p)^2 \lim_{n\to\infty} \sum_{x=1}^{n} x^2 p^{x-1} \\
&= (p-1)^2 \frac{-p-1}{(p-1)^3} \\
&= \frac{-p-1}{p-1}.
\end{aligned}$$

■

Exercícios

Ex. 42 — (Fácil) Prove que para qualquer variável aleatória discreta, $\mathbb{E}[X^2] \geq \mathbb{E}[X]^2$. Identifique quando acontece a igualdade (ou seja, quando $\mathbb{E}[X^2] = \mathbb{E}[X]^2$).

Ex. 43 — Prove que a Lei de Benford não depende da base numérica usada para representar os números: se, por exemplo, um conjunto de dados segue a distribuição de Benford, e o escrevemos na base 8, então

$$\Pr[D = d] = \log_8 \left(1 + \frac{1}{d}\right).$$

Ex. 44 — Seja

$$f(n) = \frac{6}{\pi^2 n^2}.$$

Como

$$\sum_{1}^{\infty} f(n) = 1,$$

esta é uma função de distribuição de massa. Tente calcular a esperança desta função, e explique o que ocorre.

Ex. 45 — No experimento do Exemplo 3.4.7, calcule

$$\mathbb{E}\left[\cos\left(\frac{\pi X}{10}\right)\right].$$

Ex. 46 — No final do Exemplo 3.5.1, mencionamos que a demonstração vale para qualquer polígono convexo. Verifique.

Ex. 47 — A distribuição Soliton, usada em um tipo de código corretor de erros, é definida para todo $K \in \mathbb{N}$ como a seguir.

$$\begin{aligned} \Pr[X = 1] &= \frac{1}{K} \\ \Pr[X = j] &= \frac{1}{j(j-1)} \quad (\text{ para } > 1). \end{aligned}$$

1. Prove que esta é realmente uma função de massa
2. Determine a esperança e a variância da distribuição.

Ex. 48 — (Fácil) Prove o Corolário 3.6.1.

Ex. 49 — Uma máquina produz brinquedos de plástico, e 1% deles é produzido de forma que não passa no controle de qualidade. Sabendo que lotes de 20 são enviados por vez ao processo de controle de qualidade, qual a probabilidade de haver dois brinquedos rejeitados em um lote?

Ex. 50 — Em uma grande cidade, verificou-se que 20% dos trabalhadores do comércio sabe tocar um instrumento musical. Se eu selecionar comerciários aleatoriamente, qual é a esperança da quantidade de pessoas que precisarei selecionar até encontrar uma que saiba tocar? E o desvio padrão?

Ex. 51 — Em um vilarejo com 20 000 pessoas, 2% está infectada com uma doença. A probabilidade de alguém sobreviver após infectado é de 95%, e o chefe do hospital local pediu que calculássemos a probabilidade de haver mais que 3 óbitos decorrentes dessa doença. Dê uma resposta ao chefe.

Ex. 52 — Em dúzia de margaridas, dez abelhas costumo ver. Quão provável seria, ao passar por mil dessas flores, que uma única abelha eu encontrasse?

Ex. 53 — Uma roleta escolhe equiprovavelmente entre os números $1, 2, \ldots, 10$.

1. Qual é a probabilidade de, girando a roleta dez vezes, obter um número divisível por 3 metade das vezes?

2. Qual é a esperança do número de vezes que devo girar a roleta até obter um número primo?

Ex. 54 — Uma pessoa verificou que em seus textos há uma média de 3 erros por página. Qual é a probabilidade dessa pessoa produzir duas páginas sem erros?

Ex. 55 — Sabe-se que uma certa massa de elemento radioativo emite 4.1 partículas por segundo. Qual é a probabilidade de nenhuma partícula ser emitida por três segundos?

Ex. 56 — Um jogador tentará o jogo descrito no Exemplo 3.8.3. Os dados, no entanto, não são honestos. As probabilidades de resultado em cada dado são

$$\begin{aligned}
\Pr(1) &= \frac{1}{6} \\
\Pr(2) &= \frac{1}{6} \\
\Pr(3) &= \frac{1}{6} \\
\Pr(4) &= \frac{3}{20} \\
\Pr(5) &= \frac{3}{20} \\
\Pr(6) &= \frac{1}{5}
\end{aligned}$$

Calcule a probabilidade do jogador obter sucesso em cinco rodadas.

Ex. 57 — Prove o Teorema 3.6.13.

Ex. 58 — Um jogador costuma acertar 60% de suas tentativas de lance livre no jogo de basquete. Ele nos porpôs um desafio: terá direito a vinte lances livres, e registraremos como pontuação a quantidade de cestas que ele fizer acima de dez (por exemplo, se fizer 14, registramos 4; se fizer 7, registramos -3).

1. Se concordarmos em pagar $R\$100$ por ponto neste jogo, qual será o valor esperado que receberemos (ou pagaremos, se for negativo)?
2. Qual é a probabilidade desse jogador errar todas as cestas?
3. Qual é a probabilidade da primeira cesta acontecer após a décima?
4. Qual é a probabilidade de não pagarmos nem recebermos qualquer quantia?
5. Qual é a probabilidade de recebermos algo (diferente de zero)?

Ex. 59 — (Problema de Huygens) Dois jogadores, A e B, jogam dados. Cada um joga um par de dados em sua vez; se A fizer seis pontos antes de B fazer sete, A ganha – senão, B ganha. Se A começar o jogo, qual é sua probabilidade de vencer?

Ex. 60 — Seja $X \sim \mathcal{P}oiss(\lambda)$. Sabendo que $\Pr[X = 0] = 0.3$,

1. determine λ;
2. calcule $\Pr[X > 2]$

Ex. 61 — Uma cultura em meio líquido está sendo estudada. Um biólogo verificou várias amostras de $10ml$ cada, e observou que 15% das amostras estava livre de um tipo de protozoário. Presumindo distribuição de Poisson, estime a média de protozoários por ml.

Ex. 62 — Qual distribuição de Poisson tem o maior valor para $\Pr[X = 3]$?

Ex. 63 — (Rohatgi) Refaça o Exercício 62, trocando $\Pr[X = 3]$ por $\Pr[X = k]$.

Ex. 64 — Seja $\Omega = \{0, 1, 2, 3, \ldots\}$ um espaço amostral, e um experimento onde

$$\Pr(n) = e^{-n} \qquad \text{para } n > 0.$$

1. Defina $\Pr(0)$.
2. Defina duas variáveis aleatórias diferentes neste espaço e calcule suas esperanças.

Ex. 65 — Uma máquina pinta chapas de ferro automaticamente, mas a máquina eventualmente deixa pequenos riscos. A média é de 2 riscos por $10cm^2$. Qual é a probabilidade de selecionarmos duas regiões de $10cm^2$ que

1. tenham, juntas, exatamente 3 riscos?
2. tenham, juntas, no máximo 2 riscos?
3. tenham a mesma quantidade de riscos?

Ex. 66 — Uma fábrica produz barras de chocolate com amendoim, e a média de pedaços de amendoim por barra é dez.

1. Supondo que a quantidade de pedaços de amendoim por barra segue distribuição de Poisson, determine a probabilidade de uma barra não ter qualquer pedaço de amendoim.
2. Dado que a fábrica já produziu mais de um milhão de barras, e que todas tinham mais de sete pedaços de amendoim, comente o item (i) e sua resposta.

Ex. 67 — Demonstre o Teorema 3.9.1.

Ex. 68 — Demonstre o Teorema 3.9.2.

Capítulo 4

Variáveis Aleatórias Contínuas

Este Capítulo aborda variáveis aleatórias contínuas e algumas distribuições contínuas importantes.

4.1 Variáveis contínuas

Definimos variáveis contínuas de forma análoga às variáveis discretas, mas usando integração, e não soma, para calcular quantidades. Como já visto no Capítulo 1, isso implicará que a probabilidade de eventos isolados é igual a zero.

Definição 4.1.1 (variável aleatória contínua, função de densidade de probabilidade)**.** Seja Ω um espaço amostral. Se Ω não é enumerável, então uma função de Ω em $\mathbb{R}$ é uma **variável aleatória contínua**, se, para todo $B \subseteq \Omega$,

$$\Pr[X \in B] = \int_B f(x)dx.$$

Dizemos que f é a **função de densidade de probabilidade** de X. ◀

Exemplo 4.1.2. Seja uma variável aleatória X tal que

$$\Pr[X \leq x] = \begin{cases} \dfrac{1}{e^x} & x \geq 0 \\ 0 & x < 0. \end{cases}$$

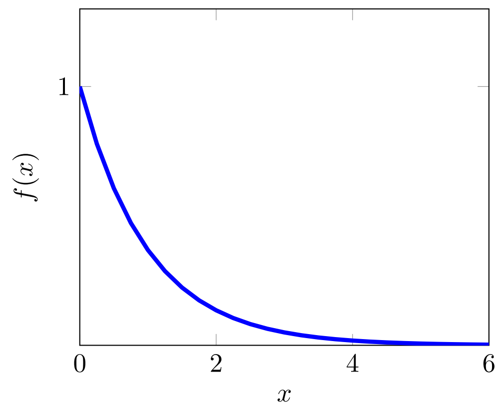
A função de densidade de probabilidade de X será

$$f(x) = \begin{cases} \dfrac{1}{e^x} & x \geq 0 \\ 0 & x < 0, \end{cases}$$

se a integral de $f(x)$ em toda a reta real for igual a um. Mas, de fato,

$$\begin{aligned}\int_{-\infty}^{\infty} f(x)dx &= \int_{-\infty}^{\infty} \frac{1}{e^x}dx \\ &= \int_{0}^{\infty} \frac{1}{e^x}dx \\ &= 1.\end{aligned}$$

A figura a seguir mostra o gráfico de $f(x)$.



●

A função de densidade *não expressa valores de probabilidades em pontos, porque estes são todos iguais a zero.* Usamos funções de densidade apenas para determinar probabilidades de regiões. Por isso, uma função de densidade pode assumir valores fora do intervalo $[0, 1]$.

Exemplo 4.1.3. A função

$$f(x) = \begin{cases} 2 & x \in [0, 1/2] \\ 0 & x \notin [0, 1/2] \end{cases}$$

assume valores maiores que 1, mas é função de densidade:

$$\begin{aligned}\int_{-\infty}^{\infty} f(x) &= \int_{0}^{1/2} f(x) \\ &= 1.\end{aligned}$$

●

Definição 4.1.4 (função de distribuição acumulada, contínua)**.** Se X é uma variável aleatória contínua com função de densidade f, então sua **função de distribuição acumulada** é

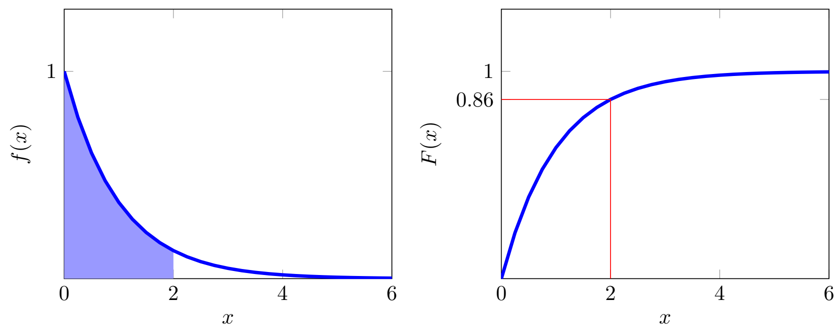
$$F(x) = \Pr[X \leq x] = \int_{\infty}^{x} f(z)dz.$$

◀

Exemplo 4.1.5. No Exemplo 4.1.2 uma função de densidade, $f(x) = e^{-x}$, é apresentada. Se uma variável aleatória X tem densidade dada por f, então terá função de distribuição acumulada

$$\begin{aligned} F(x) &= \int_{\infty}^{x} f(z)dz \\ &= 1 - e^{-x}. \end{aligned}$$

A seguir, o gráfico da esquerda ilustra o gráfico da função de densidade f, onde está destacada a região que representa a probabilidade $\Pr[X < 2]$. A área da região (integral de f no intervalo $[0, 2]$) é o valor de $F(2) = 1 - e^{-2} \approx 0.86$. O gráfico à direita é da função $F(x) = 1 - e^{-x}$.

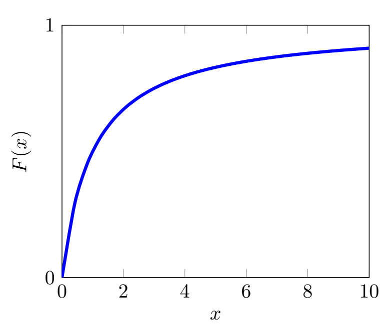


●

Como a função de distribuição acumulada é dada pela integral da função de densidade, é possível obter uma a partir da outra.

Exemplo 4.1.6 (variável contínua – obtendo $f(x)$ a partir de $F(x)$)**.** uponha que uma variável aleatória X tem função de distribuição acumulada

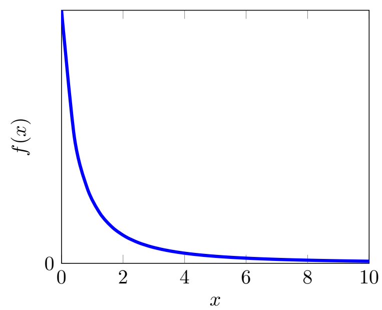
$$F(x) = 1 - \frac{1}{x+1}.$$



Verificamos que $F(x) \to 1$ quando $x \to \infty$ (que deve ser verdade para qualquer função de distribuição acumulada).

A função de densidade de X deve ser

$$f(x) = \frac{d}{dx} F(x) = \frac{1}{(x+1)^2}.$$



Se não soubermos o intervalo onde f é função de densidade de X, podemos determinar um (pode haver vários!) Suponha que queiramos que o intervalo comece em zero.

$$\begin{aligned}\int_0^a f(x)\,dx &= \int_0^a \frac{1}{(x+1)^2}\,dx \\ &- \frac{1}{x+1}\bigg|_0^a \\ &= 1 - \frac{1}{a+1},\end{aligned}$$

que, como esperado, é $F(a)$.

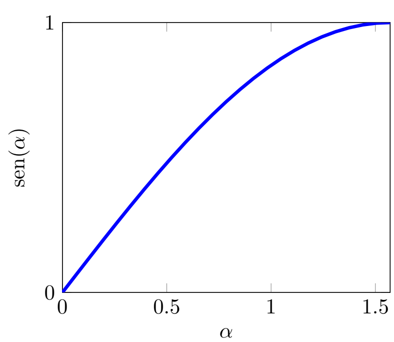
Como a integral deve valer um, queremos $1 - \frac{1}{a+1} = 1$, que só pode acontecer quando $\frac{1}{a+1}$ é zero. O intervalo, portanto é $[0, \infty]$. ●

Exemplo 4.1.7 (variável contínua $\mathrm{sen}(\alpha)$ – obtendo a densidade)**.** Ecolhemos equiprovavelmente um ângulo α em $[0, \pi/2]$, e definimos a variável aleatória

$$S = \mathrm{sen}(\alpha).$$

Determinaremos as funções de densidade e de densidade acumulada de S.

No intervalo que escolhemos, $\mathrm{sen}(\alpha)$ é crescente:



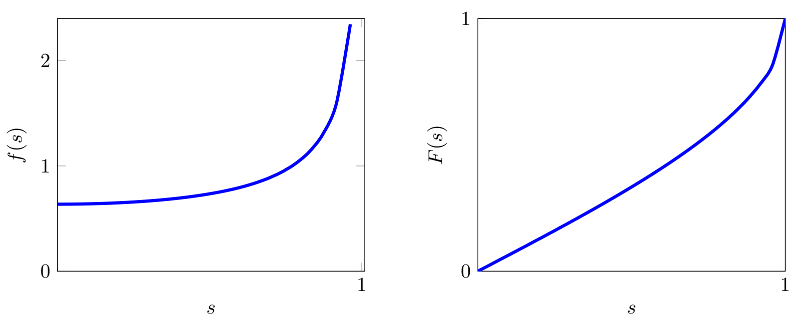
Assim, a probabilidade de $\text{sen}(\alpha) < s$ é

$$\begin{aligned} F(s) &= \Pr[S < s] \\ &= \Pr[\text{sen}(\alpha) < s] \\ &= \Pr[\alpha < \text{arcsen}(s)] \\ &= \frac{\text{arcsen}(s)}{\pi/2} \\ &= \frac{2\,\text{arcsen}(s)}{\pi}. \end{aligned}$$

Então a função de densidade de S é

$$\begin{aligned} f(s) &= \frac{d}{ds}F(s) \\ &= \frac{2}{\pi\sqrt{1-s^2}}. \end{aligned}$$

Os gráficos a seguir ilustram f e F (o de f não está em escala $1:1$).

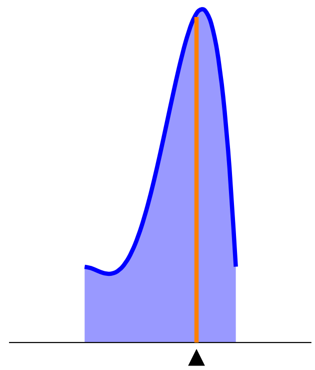


●

4.2 Esperança, variância e desvio padrão

O conceito de esperança é definido para variáveis contínuas de maneira análoga ao que foi feito para variáveis discretas.

Suponha que uma variável aleatória X tenha função de densidade $f(x)$. Ao integrarmos uma região de $f(x)$, estamos calculando uma área, que podemos, para facilitar a discussão, chamar de "volume". A esperança de uma variável contínua representa o centro de massa desse volume – o ponto na reta que divide o volume de $f(x)$ pela metade.



Definição 4.2.1 (esperança de variável aleatória contínua)**.** A **esperança** de uma variável aleatória contínua X com função de densidade de probabilidade $f(x)$ é dada por

$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} x f(x) dx. \qquad \blacktriangleleft$$

Exemplo 4.2.2. No Exemplo 4.1.2, a função de densidade de probabilidade de X era $f(x) = e^{-x}$ quando $x \geq 0$.

A esperança de X é, interessantemente,

$$\begin{aligned}\int_{-\infty}^{\infty} x f(x) dx &= \int_{-\infty}^{\infty} \frac{x}{e^x} dx \\ &= \int_0^{\infty} \frac{x}{e^x} dx \\ &= 1. \qquad \bullet\end{aligned}$$

Assim como para variáveis discretas, a esperança de uma variável contínua pode não ser finita.

Exemplo 4.2.3. Seja X uma variável aleatória contínua, com função de densidade de probabilidade

$$f(x) = \begin{cases} \frac{1}{x^2} & x \geq 1 \\ 0 & x < 1. \end{cases}$$

Esta é, de fato, uma função de densidade de probabilidade, porque

$$\int_1^{\infty} \frac{1}{x^2} dx = 1.$$

A esperança de X, no entanto, é

$$\begin{aligned}
\mathbb{E}[X] &= \int_1^\infty x f(x) dx \\
&= \int_1^\infty \frac{x}{x^2} dx \\
&= \int_1^\infty \frac{1}{x} dx \\
&= +\infty.
\end{aligned}$$

●

Teorema 4.2.4. A esperança é linear: se X é uma variável aleatória contínua,

$$\begin{aligned}
\mathbb{E}[aX] &= a\mathbb{E}[X], \\
\mathbb{E}[X+Y] &= \mathbb{E}[X] + \mathbb{E}[Y].
\end{aligned}$$

Demonstração. A demonstração é idêntica àquela para variáveis discretas, trocando somatórios por integrais (veja o Teorema 3.4.8). ■

Teorema 4.2.5. Seja X uma variável aleatória contínua e $g : \mathbb{R} \to \mathbb{R}$ uma função. Então

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f(x)\, dx.$$

Definição 4.2.6 (variância de variável aleatória contínua)**.** A **variância** de uma variável aleatória contínua X com função de densidade de probabilidade $f(x)$ e média μ é definida da mesma forma que para as variáveis discretas, $\sigma_x^2 = \mathbb{E}[(X-\mu)^2]$. É necessário, no entanto, usar a definição de esperança para variáveis contínuas, obtendo

$$\sigma_X^2 = \mathbb{E}[(X-\mu)^2] = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x) dx.$$

◀

Da mesma forma que para variáveis discretas, vale o Teorema 3.4.11, que reproduzimos aqui.

Teorema 4.2.7 (reprodução do Teorema 3.4.11)**.** Se X é uma variável aleatória com esperança igual a μ, então $\mathrm{var}[X] = \mathbb{E}[X^2] - \mu^2$.

O desvio padrão é definido para variáveis contínuas da mesma forma que para discretas. Repetimos aqui a definição[1]

Definição 4.2.8 (desvio padrão, reprodução da Definição 3.4.13)**.**

$$\sigma_X = \mathrm{SD}(X) = \sqrt{\sigma_X^2}.$$

◀

[1] A Definição 3.4.13 *não* especificava que a variável aleatória deveria ser discreta.

4.3 Uniforme

No Capítulo 1 mencionamos equiprobabilidade, mas somente em espaços discretos finitos e em espaços contínuos (verificamos que nos espaços infinitos e discretos não há como atribuir probabilidades iguais a todos os resultados). Não tratamos de distribuições uniformes no Capítulo 3 porque para o caso finito pouco há que se possa estudar.

No caso contínuo, no entanto, a distribuição uniforme tem parâmetros, e podemos descrever a média e a variância de variáveis uniformes em função desses parâmetros.

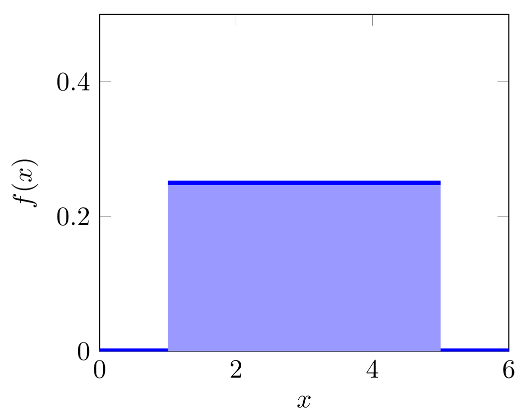
Definição 4.3.1 (distribuição contínua uniforme)**.** Uma variável aleatória contínua X tem **distribuição uniforme** em um intervalo $[a, b]$ se sua função de densidade de probabilidade é

$$f(x) = \frac{1}{b-a}.$$

Denotamos esta distribuição por $\mathcal{U}(a, b)$. ◀

Não existe distribuição uniforme em toda a reta real – somente em intervalos. Se quiséssemos definir $f(x) = k > 0$ constante (e portanto definindo distribuição uniforme em $\mathbb{R}$), ela não seria uma função de densidade de probabilidade, porque

$$\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{\infty} kdx = \infty \neq 1.$$



Exemplo 4.3.2. Seja X uma variável aleatória com distribuição uniforme no

intervalo $[1, 4]$. Calculamos $\Pr[2 < X < 3]$:

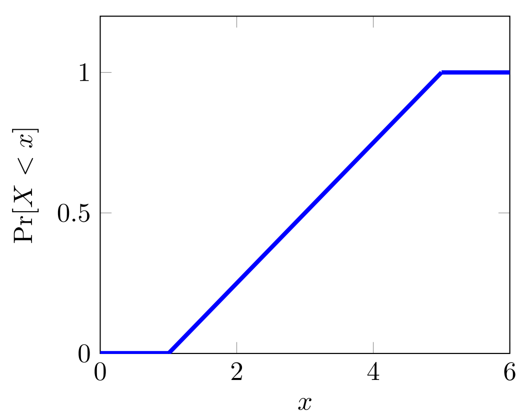
$$\begin{aligned}
\Pr[2 < X < 3] &= \int_2^3 f(x)dx \\
&= \int_2^3 \frac{1}{3}dx \\
&= \frac{1}{3}\int_2^3 dx \\
&= \frac{1}{3} \left. x \right|_2^3 \\
&= \frac{1}{3}(3-2) \\
&= \frac{1}{3}.
\end{aligned}$$

●

A função distribuição da uniforme contínua em $[a, b]$ é

$$F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & x \in [a, b] \\ 1 & x > b. \end{cases}$$

O gráfico a seguir ilustra $F(x)$ para a distribuição uniforme no intervalo $[1, 5]$.



Exemplo 4.3.3. Sorteamos um número qualquer no intervalo $(0, 1)$. Determinaremos a probabilidade dos seus dois primeiros dígitos após a vírgula representarem um número primo.

Seja K o número sorteado. Definimos uma variável aleatória X, que terá o valor igual à parte inteira de $100K$: Por exemplo,

K	X	primo?
$0.\underline{03}215\ldots$	03	sim
$0.\underline{11}238\ldots$	11	sim
$0.\underline{33}212\ldots$	33	não
$\vdots$		$\vdots$

Cada valor de X tem probabilidade $1/100$ de ocorrer:

$$\begin{aligned}
\Pr[X = 00] &= \Pr[0.0 \leq K < 0.1] \\
\Pr[X = 01] &= \Pr[0.1 \leq K < 0.2] \\
\Pr[X = 02] &= \Pr[0.2 \leq K < 0.3] \\
&\vdots
\end{aligned}$$

Como há 25 primos com no máximo dois dígitos[2], logo

$$\Pr[X \text{ é primo}] = \frac{25}{100} = \frac{1}{4}. \qquad \bullet$$

Teorema 4.3.4. A esperança de uma variável contínua no intervalo $[a, b]$ é

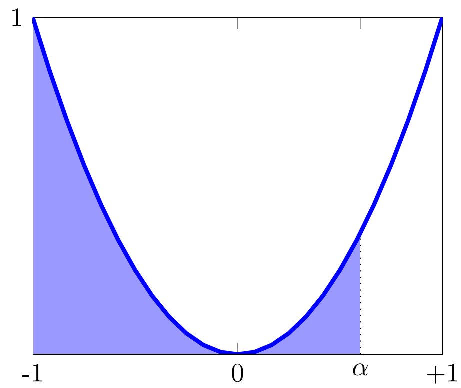
$$\mathbb{E}[X] = \frac{a+b}{2}.$$

Demonstração.

$$\begin{aligned}
\mathbb{E}[X] &= \int_a^b x \frac{1}{b-a} dx \\
&= \frac{1}{b-a} \int_a^b x dx \\
&= \frac{1}{b-a} \left[\frac{x^2}{2} \right]_a^b \\
&= \frac{1}{b-a} \frac{b^2 - a^2}{2} \\
&= \frac{1}{b-a} \frac{(a+b)(a-b)}{2} \\
&= \frac{a+b}{2}.
\end{aligned} \qquad \blacksquare$$

Exemplo 4.3.5. Um ponto α é escolhido equiprovavelmente no intervalo $[-1, +1]$. Definimos uma variável aleatória Y, que será igual à área abaixo da parábola determinada por x^2, *apenas em* $[-1, \alpha]$.

[2] 2,3,5,7,11,13,17,19,23,29,31,37,41,43,47,53,59,61,67,71,73,79,83,89,97.



Calcularemos a esperança de Y de três maneiras diferentes.

Prieiro, sabemos que Y deve variar de zero até a área sob a parábola entre $[-1, +1]$, e que as áreas são equiprováveis. Calculamos o valor máximo de Y:

$$y_{\max} = \int_{-1}^{+1} x^2\ dx = \frac{2}{3}.$$

Como $Y \in [0, 2/3]$, distribuída uniformemente, então a função de densidade para Y é

$$f(y) = \frac{1}{2/3 - 0} = \frac{3}{2}.$$

A esperança de Y é, portanto,

$$\begin{aligned} \mathbb{E}[Y] &= \int_0^{2/3} \frac{3}{2} dy \\ &= \left[\frac{3}{4} x^2\right]_0^2 /3 \\ &= \frac{1}{3}. \end{aligned}$$

Passamos ao segundo método. Definimos uma variável X, que nos dá simplesmente o valor escolhido para α. Os valores de X ficam entre -1 e $+1$, logo sua função de densidade é

$$g(x) = \frac{1}{2}.$$

A variável Y é função de X: a área sob x^2 entre -1 e k é

$$\int_{-1}^{k} x^2 dx = \frac{k^2 + 1}{3},$$

logo

$$Y = \frac{X^2 + 1}{3}.$$

Usamos o Teorema 4.2.5, e calculamos

$$\begin{aligned}\mathbb{E}[Y] &= \mathbb{E}[h(X)] \\ &= \int_{-\infty}^{\infty} h(x)f(x)\,dx \\ &= \int_{-\infty}^{\infty} h(x)\frac{1}{2}\,dx \\ &= \int_{-1}^{+1} x^2 + 13\frac{1}{2}\,dx \\ &= \frac{1}{6}\left[\frac{x^4}{4} + x\right]_{-1}^{+1} \\ &= \frac{1}{3}.\end{aligned}$$

O terceiro método consiste na aplicação direta do Teorema 4.3.4. A variável Y tem distribuição uniforme em $[0, 2/3]$, logo

$$\mathbb{E}[Y] = \frac{0 + 2/3}{2} = \frac{1}{3}.$$ ●

Teorema 4.3.6. Se X tem distribuição uniforme,

$$\operatorname{var}(X) = \frac{(b-a)^2}{12}$$

Exemplo 4.3.7. A variância de Y, no Exemplo 4.3.5, é

$$\sigma_Y^2 = \frac{(2/3)^2}{12} = \frac{1}{27} \approx 0.037,$$

e o desvio padrão é, consequentemente,

$$\sigma_Y = \sqrt{\frac{1}{27}} = \frac{1}{3\sqrt{3}} \approx 0.192.$$ ●

4.4 Normal

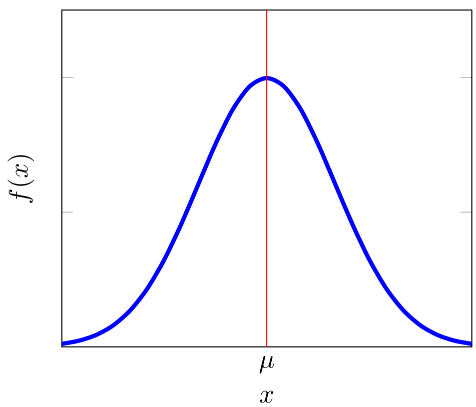
A normal é uma das mais importantes distribuições. Ela surge naturalmente em muitas situações, como é detalhado no Capítulo 6, onde se trata do Teorema Central do Limite.

Definição 4.4.1 (distribuição normal)**.** Uma variável aleatória contínua X tem **distribuição normal** com média μ e variância σ^2 se sua função de densidade de probabilidade é

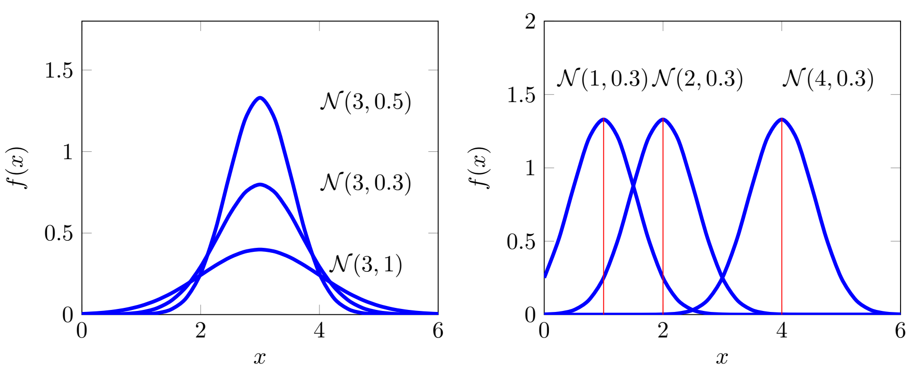
$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}.$$

É usual denotar esta distribuição por $\mathcal{N}(\mu, \sigma^2)$. ◀

O gráfico da função de densidade da normal é simétrico ao redor do valor da média.



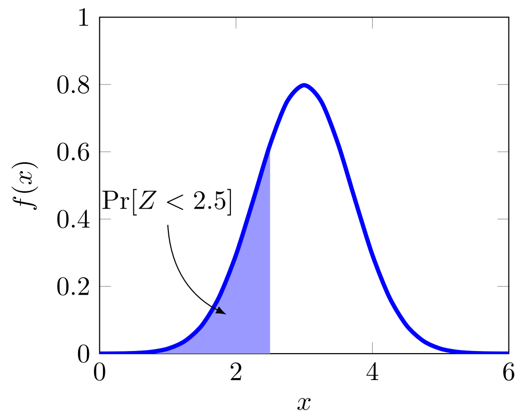
O gráfico da função de densidade da distribuição normal pode variar em sua altura e deslocamento ao longo do eixo das abscissas. Na figura a segiur estão os gráficos das funções de densidade de diferentes distribuições normais: no primeiro gráfico, algumas distribuições com a mesma média; no segundo, algumas com a mesma variância.



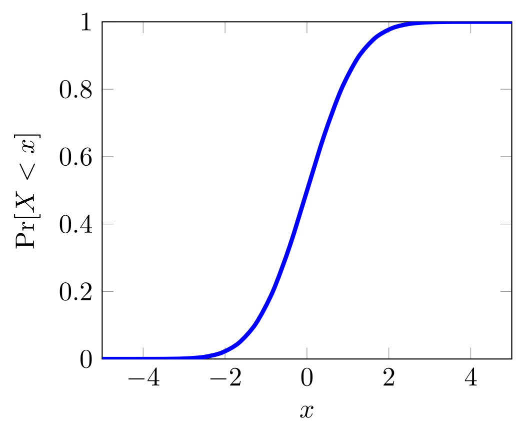
Da Definição 4.4.1 segue que se X tem distribuição $\mathcal{N}(\mu, \sigma)$, então

$$\Pr[X < x] = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx.$$

Por exemplo, a área ilustrada na próxima figura é o valor de $\Pr[X < 2.5]$ para uma das distribuições normais ilustradas antes, $\mathcal{N}(3, 0.5)$.



A seguir está o gráfico de uma aproximação[3] da função distribuição de $\mathcal{N}(0,1)$, usualmente denotada $\Phi(x)$.

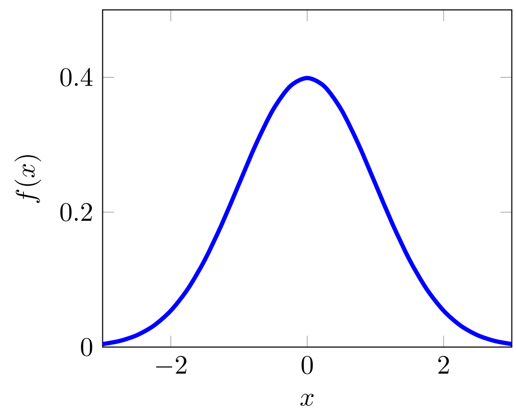


A integral da função de densidade da normal, no entanto, não tem forma fechada, e não podemos resolvê-la analiticamente. Podemos usar uma tabela com os valores da função distribuição de X (ou seja, os valores de $\Pr[X < x]$) – mas precisaríamos de uma tabela para cada par de parâmetros (μ, σ). Usamos uma mudança de variável para obter, a partir de uma única tabela, os valores de $\Pr[X < x]$ para qualquer variável X com distribuição normal.

Definição 4.4.2. A distribuição normal com média zero e variância um, $\mathcal{N}(0,1)$, é chamada de **distribuição normal padrão**. ◀

[3]Conforme relatório técnico de B. Zogheib e M. Hlynka, 2009, "Approximations of the Standard Normal Distribution", WMSR #09-09:

$$\Phi(x) \approx \left(1 + e^{0.0054 - 1.6101x - 0.0674x^3}\right)^{-1}.$$



Exemplo 4.4.3. Seja A uma variável aleatória com distribuição normal padrão. Então $\Pr[A < 1.25]$ é

$$\int_{-\infty}^{1.25} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}}\, dx.$$

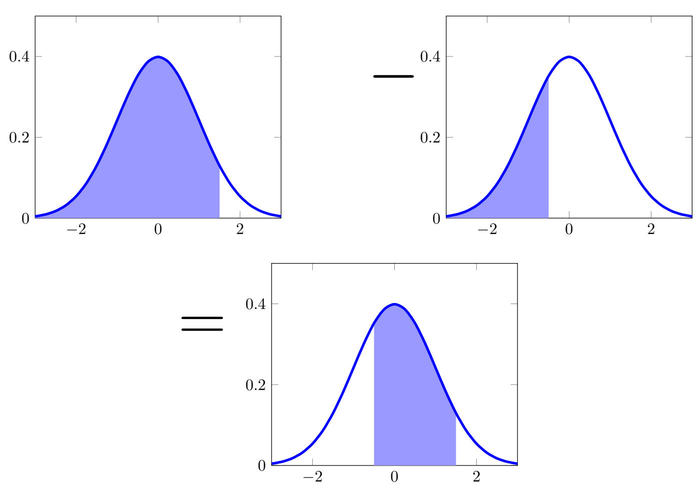
A tabela da função distribuição para a normal padrão pode ser consultada no Apêndice A, e consultando-a verificamos que

$$\Pr[A < 1.25] \approx 0.8944.$$ ●

É também simples calcular $\Pr[X > b]$ e $\Pr[a < X < b]$:

$$\Pr[X > b] = 1 - \Pr[X < b]$$
$$\Pr[a < X < b] = \Pr[X < b] - \Pr[X < a]$$

A segunda parte pode ser facilmente verificada calculando as integrais. A figura a seguir ilustra a obtenção de $\Pr[-0.5 < X < 1.5]$ como $\Pr[X < 1.5] - \Pr[X < -0.5]$.



Exemplo 4.4.4. Seja B uma variável aleatória com distribuição normal padrão. Então

$$\begin{aligned}
\Pr[-0.70 < B < 1.5] &= \int_{-0.7}^{1.5} \frac{1}{\sqrt{2\pi}}\, e^{-\dfrac{x^2}{2}}\, dx \\
&= \left(\int_{-\infty}^{1.5} \frac{1}{\sqrt{2\pi}}\, e^{-\dfrac{x^2}{2}}\, dx \right) - \left(\int_{-\infty}^{-0.7} \frac{1}{\sqrt{2\pi}}\, e^{-\dfrac{x^2}{2}}\, dx \right) \\
&= 0.9332 - \int_{-\infty}^{-0.7} \frac{1}{\sqrt{2\pi}}\, e^{-\dfrac{x^2}{2}}\, dx \\
&= 0.9332 - \left[1 - \int_{-0.7}^{\infty} \frac{1}{\sqrt{2\pi}}\, e^{-\dfrac{x^2}{2}}\, dx \right] \\
&= 0.9332 - \left[1 - \int_{-\infty}^{0.7} \frac{1}{\sqrt{2\pi}}\, e^{-\dfrac{x^2}{2}}\, dx \right] \qquad \text{(simetria)} \\
&= 0.9332 - 0.2420 \\
&= 0.6912.
\end{aligned}$$

●

A utilidade de conseguir calcular probabilidades envolvendo variáveis com distribuição normal padrão é limitada, uma vez que em situações práticas a média e a variância dificilmente serão 0 e 1. Esta dificuldade será resolvida com o Teorema 4.4.5.

Teorema 4.4.5. Seja X uma variável aleatória tem distribuição normal com média μ e variância μ^2. Seja $Y = aX + b$. Então a distribuição de Y é $\mathcal{N}(a\mu + b, a\sigma)$.

O Teorema 4.4.5 nos permite usar a tabela da distribuição normal padrão para cálculo de probabilidades envolvendo *qualquer* variável normal.

Suponha que queiramos trabalhar com uma variável X, que tem distribuição $\mathcal{N}(\mu, \sigma)$. A variável

$$Z = \frac{X}{\sigma} - \frac{\mu}{\sigma},$$

que obtivemos multiplicando X por $1/\sigma$ e somando com $-\mu/\sigma$, tem distribuição

$$\mathcal{N}\left(\frac{\mu}{\sigma} - \frac{\mu}{\sigma},\ \frac{\sigma}{\sigma}\right) = \mathcal{N}(0, 1).$$

Assim, para determinar $\Pr[X < x]$, podemos transformar X em uma variável com distribuição normal padrão, e calcular

$$\Pr[X < x] = \Pr\left[Z < \frac{X - \mu}{\sigma}\right].$$

Exemplo 4.4.6. A média do peso de recém-nascidos em uma determinada população é de $2.9kg$, com desvio padrão $\sigma = 0.4kg$, e a distribuição pode ser modelada como normal.

A probabilidade de um bebê nascer com peso entre $2.8kg$ e $3kg$ é $\Pr[2800 < X < 3000]$.

Seja

$$Z = \frac{X - \mu}{\sigma}.$$

Queremos

$$\begin{aligned}
\Pr[2800 < X < 3000] &= \Pr[2800 - \mu < X - \mu < 3000 - \mu] \\
&= \Pr\left[\frac{2800 - \mu}{\sigma} < \frac{X - \mu}{\sigma} < \frac{3000 - \mu}{\sigma}\right] \\
&= \Pr\left[\frac{2800 - 2900}{400} < \frac{X - 2900}{400} < \frac{3000 - 2900}{400}\right] \\
&= \Pr\left[-\frac{1}{4} < Z < +\frac{1}{4}\right].
\end{aligned}$$

Da tabela da função distribuição da normal padrão (no Apêndice A), vemos que

$$\begin{aligned}
\Pr[Z < 0.25] &= 0.5987 \\
\Pr[Z < -0.25] &= 1 - \Pr[Z < 0.25] = 0.4013.
\end{aligned}$$

Concluímos que

$$\begin{aligned}
\Pr[-0.25 < Z < 0.25] &= \Pr[Z < 0.25] - \Pr[Z < -0.25] \\
&= 0.5987 - 0.4013 \\
&= 0.1974.
\end{aligned}$$

●

4.4.1 Aproximação da Binomial pela Normal

Suponha que X seja uma variável aleatória com distribuição binomial e parâmetros (n, p). Para n bastante grande e p não extremo (não muito perto de zero ou de um), a normal pode ser uma aproximação razoável para a binomial.

Exemplo 4.4.7. Suponha que uma variável X tenha distribuição binomial com parâmetros $(100, 0.4)$. Aproximamos esta distribuição usando uma normal: os parâmetros são

$$\begin{aligned}
\mu &= np = 100(0.4) = 40 \\
\sigma^2 &= np(1 - p) = 100(0.4)(0.6) = 24.
\end{aligned}$$

Agora calculamos $\Pr[X \leq 35]$, primeiro usando a aproximação pela normal, e depois de forma exata, tratando como binomial.

Aproximando como normal,

$$\begin{aligned}
\Pr[X \le 35] &= \Pr\left[Z \le \frac{x-\mu}{\sigma}\right] \\
&= \Pr\left[Z \le \frac{35-40}{\sqrt{24}}\right] \\
&= \Pr[Z \le -1.02] \\
&= \Pr[Z \ge 1.02] \\
&\approx 1 - 0.8461 \\
&\approx 0.1539
\end{aligned}$$

Agora tratamos de forma exata, como binomial:

$$\begin{aligned}
\Pr[X \le 35] &= \sum_{j=0}^{35} \binom{100}{j} (0.4)^j (0.6)^{100-j} \\
&\approx 0.1795.
\end{aligned}$$

A aproximação não é tão boa, mas útil quando n é tão grande que não consigamos calcular de forma exata. ●

Há uma correção que pode ser feita, para ajustar a aproximação pela normal. Consiste em somar $1/2$ a x antes de determinar $\Pr[X \le x]$.

Exemplo 4.4.8. No Exemplo 4.4.7, a aproximação que encontramos não foi boa. Com a "correção de continuidade" da aproximação, temos

$$\begin{aligned}
\Pr[X \le 35] &= \Pr\left[Z \le \frac{x+0.5-\mu}{\sigma}\right] \\
&= \Pr\left[Z \le \frac{35.5-40}{\sqrt{24}}\right] \\
&= \Pr[Z \le -0.91] \\
&= \Pr[Z \ge 0.91] \\
&\approx 1 - 0.91 \\
&\approx 1 - 0.8186 \\
&\approx 0.1814,
\end{aligned}$$

muito melhor que a aproximação anterior, que resultava em 0.1539. ●

4.5 Exponencial

Após ter definido experimentos de Bernoulli, observamos três distribuições:

- binomial, dando a quantidade de sucessos em n repetições de um experimento;

- Poisson, que obtivemos da binomial para n muito grande e p pequeno;
- geométrica, dando a quantidade de tentativas (ou de repetições) do experimento antes de um sucesso

Construiremos uma análoga contínua da distribuição geométrica – distribuição exponencial. Esta distribuição é característica de variáveis aleatórias que dão não o número de repetições, mas o tempo decorrente até que um experimento tenha sucesso.

Suponha que um experimento tenha sido modelado usando a distribuição de Poisson – temos eventos que ocorrem ao longo do tempo, e a média de ocorrências é λ em alguma unidade de tempo. Queremos saber a probabilidade do *primeiro* evento ocorrer após o tempo t, e não antes. Usamos a distribuição de Poisson:

$$\Pr[T > t] = \frac{e^{-\lambda t}\lambda^0}{0!} = e^{-\lambda t}$$

Mas temos somente uma fórmula para $\Pr[T > t]$. Para estudar a distribuição de T gostaríamos de ter sua função de densidade de probabilidade. Podemos calcular sua função distribuição:

$$\begin{aligned} F(t) &= \Pr[T \leq t] \\ &= 1 - \Pr[T > t] \\ &= 1 - e^{-\lambda t}. \end{aligned}$$

Mas $F(t)$ é $\int_t^\infty f(x)dx$, onde $f(x)$ é a função de densidade de probabilidade de T. Assim,

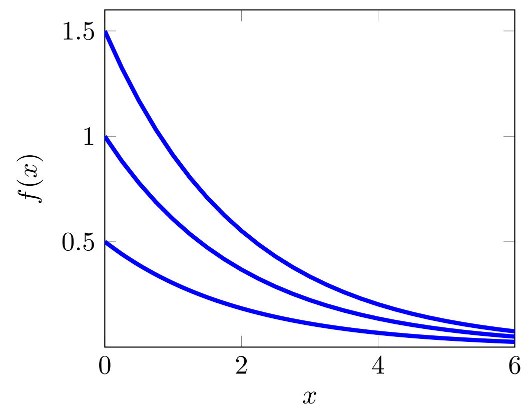
$$\begin{aligned} f(t) &= \frac{d}{dt}F(x) \\ &= \frac{d}{dt}1 - e^{-\lambda t} \\ &= \lambda e^{-\lambda t}. \end{aligned}$$

Definição 4.5.1 (distribuição exponencial)**.** Uma variável aleatória contínua X tem **distribuição exponencial** se sua função de densidade de probabilidade é

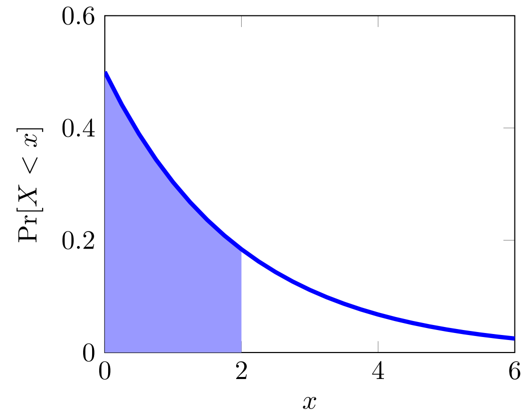
$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0. \end{cases}$$

Quando uma variável X tem distribuição exponencial com parâmetro λ, denotamos $\mathcal{E}xp(\lambda)$. ◀

A figura a seguir mostra os gráficos de três funções de densidade exponenciais – uma com $\lambda = 1.2$, uma com $\lambda = 1$, e uma com $\lambda = 1.5$.



A probabilidade $\Pr[X < x]$ é dada por $\int_0^x \lambda e^{-\lambda t} dt$; a próxima figura ilustra graficamente $\Pr[X < 2]$ com $\lambda = 1/2$.



Teorema 4.5.2. Se uma variável aleatória X tem distribuição exponencial, então

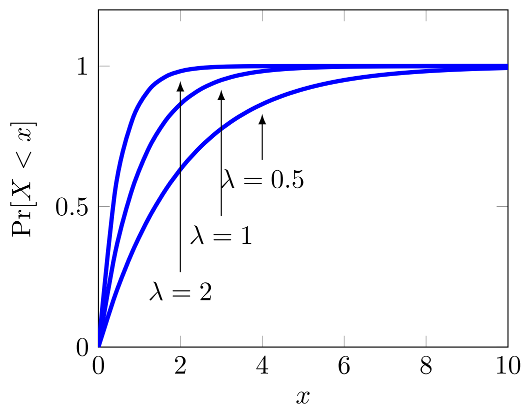
$$\Pr[X < x] = 1 - e^{-\lambda x}.$$

Demonstração.

$$\begin{aligned}
\Pr[X < x] &= \int_0^x \lambda e^{-\lambda t} dt \\
&= -e^{-\lambda x}\Big|_0^x \\
&= -e^{-\lambda x} - (-e^{\lambda 0}) \\
&= 1 - e^{-\lambda x}.
\end{aligned}$$

■

A função distribuição da exponencial é ilustrada para λ igual a 0.5, 1 e 2.



A distribuição exponencial é usada na modelagem da duração entre eventos que seguem distribuição de Poisson. Quando a média de *tempo entre eventos é* μ o parâmetro da exponencial é $\lambda = 1/\mu$ *eventos por unidade de tempo*.

Por exemplo, se um equipamento falha μ vezes por ano, dizemos que sua taxa de falha é de $\lambda = 1/\mu$.

Exemplo 4.5.3. A bateria de um determinado modelo de carro elétrico tem duração média de $1000km$, e temos uma viagem de $750km$ para fazer. Supondo que a variável aleatória que determina a duração da bateria tem distribuição exponencial, qual é a probabilidade da bateria falhar antes de de terminarmos a viagem?

Como $\mu = 1000$, a taxa de falha é $\lambda = 1/1000$.

$$\Pr[X < 750] = 1 - e^{-750/1000} = 1 - e^{-3/4} \approx 0.5276.$$

Isto significa que a probabilidade de conseguirmos realizar a viagem sem que a bateria falhe é

$$\Pr[X > 750] = e^{-3/4} \approx 0.4724.$$

●

Exemplo 4.5.4. A média de duração de conversas em um sistema interativo é de 90 segundos, e o modelamos como exponencial. Determinaremos a probabilidade de uma conversa demorar mais que 3 minutos. O *tempo* de duração médio é de 90 segundos, mas o parâmetro que precisamos para a distribuição exponencial é o inverso disso – a frequência, *por minuto*, com que conversas terminam.

$$\lambda = \frac{1}{90}$$

$$\begin{aligned}
\Pr[D \geq 180] &= 1 - \Pr[D < 180] \\
&= 1 - (1 - e^{-180/90}) \\
&= e^{-180/90} \\
&= e^{-2} \\
&\approx 0.135.
\end{aligned}$$

●

Podemos querer determinar o tempo a partir da probabilidade, ao invés do contrário. O próximo exemplo ilustra como isso pode ser feito.

Exemplo 4.5.5. No Exemplo 4.5.4 expusemos um sistema de chat onde as conversas tem duração média de $90s$. Se há 50 conversas em andamento, em quanto tempo devemos esperar que 10 delas tenham terminado?

Queremos o tempo após o qual 40/50 conversas continua ativa – ou o tempo em que 10/50 das conversas termina.

$$\begin{aligned}
\Pr[T > t] &= 40/50 \\
e^{-t/90} &= 4/5 \\
-\frac{t}{90} &= \log\left(\frac{4}{5}\right) \\
t &= -90\log\left(\frac{4}{5}\right) \\
&\approx 20.08 \text{ segundos.}
\end{aligned}$$

●

Quando expusemos a distribuição de Poisson, mostramos exemplos de eventos raros dispostos no tempo *e no espaço*. A distribuição expoenencial trata dos intervalos de tempo, *ou de espaço*, entre os eventos modelados por uma distribuição de Poisson.

Se andarmos em linha reta e marcarmos pontos aleatoriamente, a quantidade de pontos após um certo tempo *(ou após determinado comprimento – tanto faz, porque andamos em velocidade constante)* seguirá uma distribuição de Poisson, com média μ. O *tempo entre as marcas feitas* é dado por uma distribuição exponencial, com parâmetro $\lambda = 1/\mu$.

Isso pode ser visto claramente ao determinarmos um valor numérico para a média. Suponha que façamos cinco marcas em cada dez metros (ou a cada dez segundos). Então $\mu = 5/10 = 1/2$ marca por metro. Como a média de marcas por metro é 1/2, a distância entre duas marcas deve ser de 2 metros ($\lambda = (1/2)^{-1}$).

Exemplo 4.5.6. Há roteadores sem fio em uma avenida, instalados por pequenos comerciantes e moradores, independentemente; a posição de cada um não tem relação com a posição com as dos outros. Os donos desses roteadores concordaram em usá-los aos domingos para oferecer acesso à rede para as pessoas que passam por ali, em troca de um valor pago pela Prefeitura. Sabemos que há 210 roteadores em 20 quadras – e cada quadra tem $100m$. A quantidade de roteadores por quadra segue distribuição de Poisson. Determinaremos a probabilidade de uma pessoa qualquer, passeando pela avenida em um Domingo, fique a mais de $30m$ de um desses roteadores.

Há $210/20 = 21/2$ roteadores por quadra, logo 21/2 a cada $100m$. O parâmetro que usamos na distribuição exponencial é

$$\lambda = \frac{21}{200}$$

Assim,

$$\begin{aligned}\Pr[D > 30] &= e^{-30(21/200)}\\ &= e^{-63/30}\\ &\approx 0.043.\end{aligned}$$

●

Exemplo 4.5.7. Uma rede de comunicações é composta de dispositivos espalhados em uma determinada área. Presumimos que a distribuição desses dispositivos é aleatória – como se cada um deles tivesse sido posto em um ponto aleatório, sem qualquer relação com a escolha dos outros locais.

Um usuário tentando se comunicar pode ser modelado como um ponto escolhido ao acaso na região. A menor distância desse usuário até um dispositivo da rede segue uma distribuição exponencial. ●

Assim como a geométrica, a distribuição exponenicial não tem memória.

Teorema 4.5.8. A distribuição exponencial não tem memória. Se X é uma variável com distribuição exponencial, então, para quaisquer $0 \leq a, b \in \mathbb{R}$,

$$\Pr[X > a + b | X > a] = \Pr[X > b].$$

Exemplo 4.5.9. No Exemplo 4.5.6 havia $\lambda = 21/200$ roteadores por metro em uma avenida. Se eu percorrer a avenida do começo ao fim, *a posição do próximo roteador não depende da minha posição na avenida, ou da quantidade de roteadores pelos quais já passei*: a probabilidade do próximo roteador estar a mais que $10m$, dado que já andei $50m$, é a área sob a curva $\lambda e^{-\lambda x}$ para x de 60 até ∞, dividida pela área sob a mesma curva, mas para x de 50 até ∞.

$$\begin{aligned}\Pr[D > 60 | D > 50] &= \frac{\int_{60}^{\infty} \lambda e^{-\lambda x} dx}{\int_{50}^{\infty} \lambda e^{-\lambda x} dx}\\ &= \frac{e^{-\lambda 60}}{e^{-\lambda 50}}\\ &= e^{-\lambda(60-50)}\\ &= e^{-21/20}\\ &\approx 0.35,\end{aligned}$$

que é o mesmo que

$$\begin{aligned}\Pr[D > 10] &= e^{-(21/200)10}\\ &= e^{-21/20}\\ &\approx 0.35.\end{aligned}$$

●

Teorema 4.5.10. Se uma variável aleatória X tem distribuição exponencial, então

$$\mathbb{E}[X] = \frac{1}{\lambda}.$$

Demonstração.

$$\begin{aligned}
\mathbb{E}[X] &= \int_0^\infty x\lambda e^{-\lambda x} dx \\
&= \lambda \lim_{a\to\infty} \int_0^a x e^{-\lambda x} dx \\
&= \lambda \lim_{a\to\infty} \left[\left(-\frac{a}{\lambda e^{\lambda a}} - \frac{1}{\lambda^2 e^{\lambda a}} \right) - \left(\frac{0}{\lambda e^0} - \frac{1}{\lambda^2 e^0} \right) \right] \\
&= \lambda \left(\frac{1}{\lambda^2} \right) \\
&= \frac{1}{\lambda}.
\end{aligned}$$

■

Teorema 4.5.11. Se uma variável aleatória X tem distribuição exponencial, então

$$\operatorname{var}(X) = \frac{1}{\lambda^2}.$$

Exemplo 4.5.12. No Exemplo 4.5.4, o tempo médio de duração de uma conversa era de 90s. A variância é, portanto,

$$\sigma^2 = \frac{1}{\lambda^2} = \frac{1}{90^2} =\approx .00012,$$

e o desvio padrão é

$$\sigma = \frac{1}{90} \approx 0.011.$$

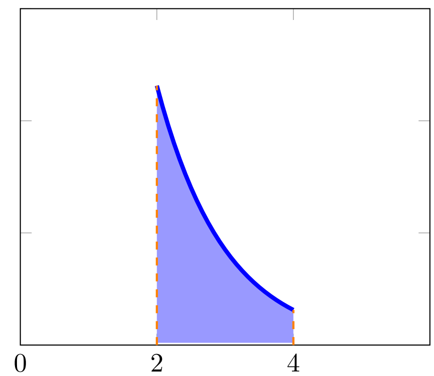
●

4.6 Distribuições truncadas

Há situações em que a distribuição necessária para um modelo é semelhante a uma distribuição conhecida, mas com o domínio restrito. Estas são chamadas de *distribuições truncadas*.

Exemplo 4.6.1 (exponencial truncada)**.** Uma variável aleatória X tem distribuição exponencial truncada, com

$$f(x) = \begin{cases} ae^{-x} & \text{se } x \in [2,4] \\ 0 & \text{se } x \notin [2,4] \end{cases}$$



Note que não podemos usar ae^{-ax}, porque esta não integraria um no intervalo dado, para nenhum valor de a. Para que seja função de densidade,

$$\int_2^4 ae^{-x}dx = a(e^{-2} - e^{-4})$$

deve ser um, portanto

$$a = \frac{e^4}{e^2 - 1}.$$

Podemos calcular, por exemplo, $\Pr[X > 3]$:

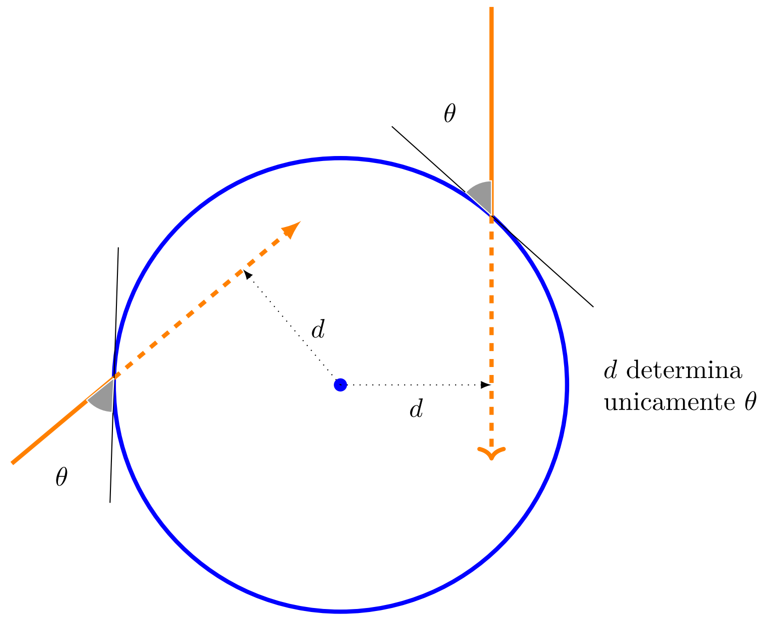
$$\int_3^4 ae^{-x}dx = \frac{e - 1}{e^2 - 1} \approx 0.269.$$

●

Os Exemplos 5.3.2 e 5.5.5 no Capítulo 5 tratarão novamente do tópico de distribuições truncadas.

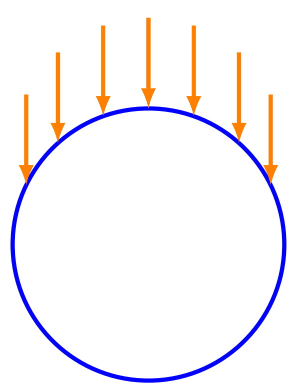
4.7 Crateras lunares e a distribuição do seno

Em 1892 o geólogo Grove Karl Gilbert apresentou uma interessante discussão sobre a formação de crateras lunares, sugerindo que haviam sido formadas por impacto de meteoritos. Gilbert estava interessado na distribuição dos ângulos com que os meteoritos atingiriam a superfície da Lua. Comentava que o ângulo com que cada um atinge a superfície depende da menor distância de sua trajetória até o centro da Lua, e não da sua direção de aproximação.

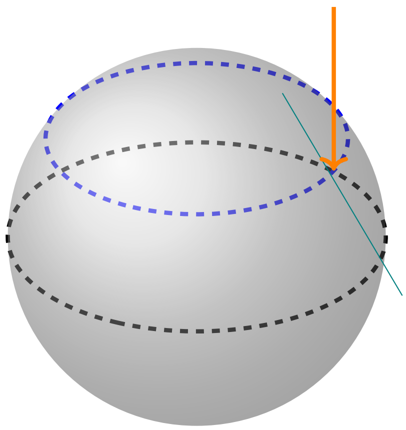


A partir desta observação, Gilbert obteve a probabilidade do ângulo de incidência de um meteorito ser menor ou igual que algum α.

Modelamos a Lua como uma esfera de raio um. Como observado por Gilbert, não faz diferença a direção pela qual o meteorito se aproxima, e sim quão próximo ele passa do centro. Fixamos, portanto, uma direção, simplificando o modelo de forma que os meteoritos cheguem sempre em trajetórias paralelas.



Considere um meteorito que chega formando um ângulo α. Ele define uma calota na esfera, e todos os meteoritos chegando na calota tem angulo de incidência maior ou igual que α.

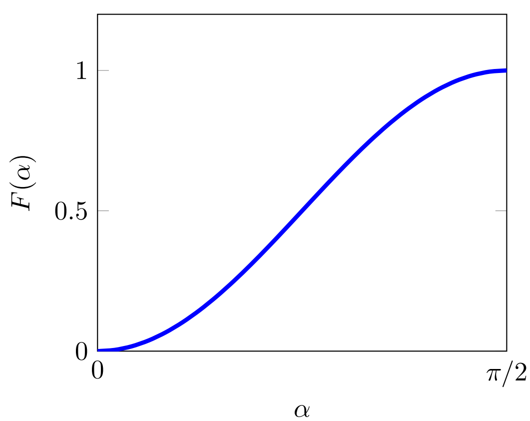


Agora observamos que, se os meteoritos chegam equiprovavelmente em qualquer posição, a probabilidade de um deles formar ângulo maior que α é dada pela área da seção circular menor (e *não* da calota!) dividida pela área da seção maior. A seção maior, que divide a esfera em duas, tem área π. A menor, que tem raio $\operatorname{sen}(\alpha)$, tem área $\pi \operatorname{sen}^2(\alpha)$. A razão entre as duas é

$$\frac{\pi \operatorname{sen}^2(\alpha)}{\pi} = \operatorname{sen}^2(\alpha),$$

que é a probabilidade do ângulo ser menor ou igual que α. Assim, temos uma função de densidade acumulada

$$\Pr[A \leq \alpha] = F(\alpha) = \operatorname{sen}^2(\alpha).$$

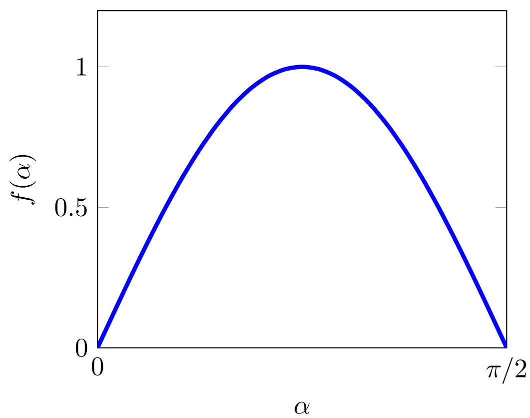


A função de densidade da distribuição que queremos deve ser, portanto,

$$\frac{d}{d\alpha} \operatorname{sen}^2(\alpha) = 2 \operatorname{sen}(\alpha) \cos(\alpha) = \operatorname{sen}(2\alpha)$$

Esta é, de fato, uma função de densidade, porque sua integral é um:

$$\begin{aligned}\int_{-\infty}^{\infty} \operatorname{sen}(2\alpha)d\alpha &= \int_0^{\pi/2} \operatorname{sen}(2\alpha)d\alpha \\ &= 1.\end{aligned}$$



Agora podemos verificar, por exemplo, que a esperança do ângulo de incidência dos meteoritos é

$$\mathbb{E}[A] = \int_0^{\pi/2} \alpha \operatorname{sen}(\alpha)d\alpha = \frac{\pi}{4}.$$

Uma conta simples também nos dá a variância de $\sigma_A^2 = (\pi^2 - 8)/16 \approx 0.1168$.

Exemplo 4.7.1 (probabilidade de ângulos de meteoritos)**.** Com as devidas ressalvas – a Lua não é completamente esférica, e a trajetória dos meteoritos não é completamente retilínea – a probabilidade de um meteorito atingir a Lua fazendo um ângulo menor do que $\pi/4$ com a superfície é

$$F\left(\frac{\pi}{4}\right) = \operatorname{sen}^2\left(\frac{\pi}{4}\right) = \left(\frac{\sqrt{2}}{2}\right)^2 = \frac{1}{2}.$$

Sabemos que $\pi/4$ é a média dos ângulos, e a probabilidade de um ângulo ser menor que a média é, portanto, 1/2. Isso acontece porque a distribuição é simétrica, e a média divide a área sob a curva da função de densidade em duas partes exatamente iguais. Em uma distribuição não simétrica (como a exponencial), isto não acontece.

Já a probabilidade de um ângulo de menos de $\pi/8$ é

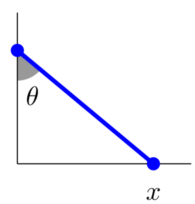
$$F\left(\frac{\pi}{8}\right) = \operatorname{sen}^2\left(\frac{\pi}{8}\right) \approx 0.146.$$

Como esperado, também, $F(0) = 0^2 = 0$, e $F(\pi/4) = 1^2 = 1$. ●

4.8 A distribuição de Cauchy

Esta Seção descreve a distribuição de Cauchy (também chamada de distribuição de Lorenz). O principal objetivo desta Seção é apresentar uma distribuição de densidade que não tem esperança (e nem variância) definida – mas que ainda assim segue rigorosamente a definição de distribuição de probabilidade, e é usada para modelar situações práticas.

Definimos um experimento aleatório: Seja P o ponto com coordenada $(0, 1)$. Escolhemos um ângulo θ de maneira uniforme entre $-\pi$ e $+\pi$. Em seguida, traçamos um segmento de P até o eixo das abscissas, formando ângulo θ com o eixo das ordenadas.



Cada vez que o experimento é executado, a variável aleatória X registrará a distância da origem até o ponto de intersecção do segmento com as abscissas.

Agora determinaremos a distribuição acumulada de densidade para X.

$$\begin{aligned} F(X) &= \Pr[X < x] \\ &= \Pr[\tan(\theta) < x] \\ &= \Pr[\theta < \arctan(x)] \\ &= \frac{1}{2} + \frac{\arctan(x)}{\pi}. \end{aligned}$$

De F obtemos a função de densidade de X,

$$\begin{aligned} f(x) &= \frac{d}{dx} F(X) \\ &= \frac{d}{dx}\left(\frac{1}{2} + \frac{\arctan(x)}{\pi}\right) \\ &= \frac{1}{\pi(1+x^2)}. \end{aligned}$$

Definição 4.8.1. Se uma variável aleatória X tem função de densidade

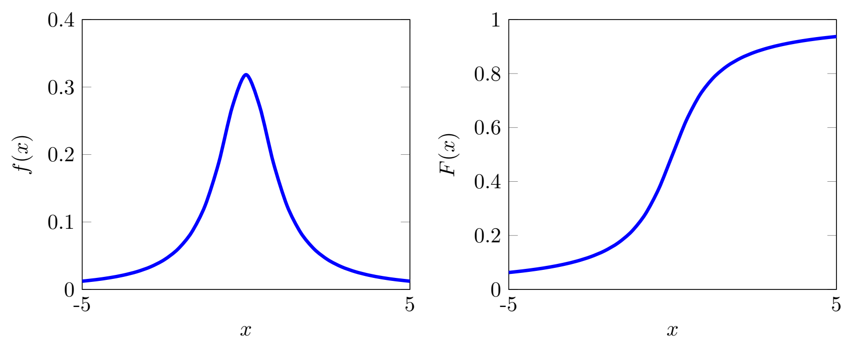
$$f(x) = \frac{1}{\pi(1+x^2)}$$

dizemos que X tem **distribuição de Cauchy** padrão. ◀

A função $f(x)$ apresentada na Definição 4.8.1 é uma função de densidade de probabilidade, porque

$$\begin{aligned}
\int_{-\infty}^{\infty} \frac{1}{\pi(1+x^2)\,dx} &= \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{1}{(1+x^2)}\,dx \\
&= \frac{1}{\pi}\left(2\int_{0}^{\infty} \frac{1}{(1+x^2)}\,dx\right) \qquad (1/(1+x^2) \text{ é função par}) \\
&= \frac{2}{\pi}\lim_{z\to\infty} \arctan(x)\Big|_0^z \\
&= \frac{2}{\pi}\frac{\pi}{2} \\
&= 1.
\end{aligned}$$

A figura a seguir mostra os gráficos de $f(x)$ e $F(x)$.



Teorema 4.8.2. Uma variável X tendo distribuição de Cauchy não tem esperança definida – e consequentemente, não tem variância.

Demonstração. O que definiria a esperança de X é a integral imprópria

$$\int_{-\infty}^{\infty} x\frac{1}{\pi(1+x^2)}\,dx = \lim_{A\to-\infty}\lim_{B\to+\infty}\int_A^B x\frac{1}{\pi(1+x^2)}\,dx,$$

que teria que ser o mesmo que, trocando a ordem de integração,

$$\lim_{B\to+\infty}\lim_{A\to-\infty}\int_A^B x\,\frac{1}{\pi(1+x^2)}\,dx,$$

mas a integral mais interna é divergente. ■

Exercícios

Ex. 69 — Determine k tal que a função f a seguir seja função de densidade de probabilidade.

$$f(x) = \begin{cases} k(-x^3+1) & x \in [-1,0] \\ ke^x & x \in [0,1] \end{cases}$$

Ex. 70 — Verifique que

$$f(x) = \begin{cases} \dfrac{1}{\pi\sqrt{x(1-x)}} & x \in [0,1] \\ 0 & x \notin [0,1] \end{cases}$$

é uma função de densidade de probabilidade.

Ex. 71 — Se uma variável aleatória X tem função de distribuição acumulada

$$F(x) = 1 - \frac{1}{\sqrt{x}}, \quad x \geq k,$$

qual é o valor de k? E qual é a função de densidade de X?

Ex. 72 — Refaça o Exemplo 4.1.7, trocando $\mathrm{sen}(\alpha)$ por $\mathrm{sen}(\alpha)^2$.

Ex. 73 — Em que intervalos a função $\cos(x)$ é função *de distribuição acumulada* de probabilidade?

Ex. 74 — Quero atribuir uma função de densidade de probabilidade f a uma variável aleatória contínua em um intervalo $[a, b]$. A função deve ser decrescente, e deve ser tal que $f(b) = 0$. Descreva uma função que eu possa usar.

Ex. 75 — Suponha que X seja variável aleatória contínua, e que eu tenha escolhido f, a seguir, para ser a função de densidade de probabilidade de X.

$$f(x) = \begin{cases} 3x & x \in [0,1] \\ 0 & x \notin [0,1]. \end{cases}$$

Calculei a esperança e a variância de X, mas encontrei um problema.

$$\begin{aligned} \mathbb{E}[X] &= \int_0^1 x f(x) dx = \int_0^1 3x^2 dx = 3\int_0^1 x^2 dx = 1 \\ \mathbb{E}[X^2] &= \int_0^1 x^2 f(x) dx = 3\int_0^1 x^3 dx = \frac{3}{4} \\ \mathbb{E}[X]^2 &= 1 \\ \mathrm{var}(X) &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{3}{4} - 1 = -\frac{1}{4}. \end{aligned}$$

A variância resultou em valor negativo. Tentei verificar calculando por outro método, e obtive

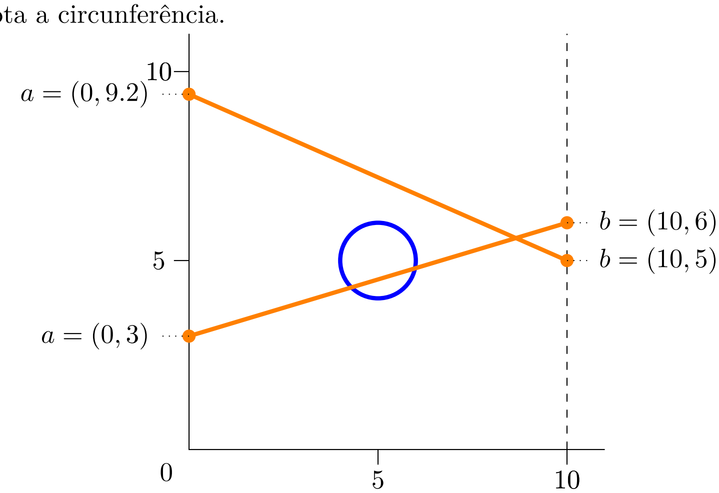
$$\begin{aligned} \mathrm{var}(X) &= \mathbb{E}[(x-\mu)^2] \\ &= \mathbb{E}[(x-1)^2] \\ &= \int_0^1 (x-1)^2 3x dx \\ &= +\frac{1}{4}, \end{aligned}$$

diferente do valor que havia calculado antes. Onde errei?

Ex. 76 — Um número X é sorteado no intervalo $(0, 1)$. Qual é a probabilidade

1. do primeiro dígito após a vírgula ser par?
2. do segundo dígito após a vírgula ser ímpar?
3. da raiz quadrada de X ter o primeiro dígito após a vírgula igual a 4 ou 5?

Ex. 77 — Uma circunferência foi desenhada com centro em $(5, 5)$ e raio 1. Considere o seguinte experimento aleatório: sorteamos equiprovavelmente dois números a e b entre 0 e 10 (são portanto descritos por duas vriáveis uniformes), e traçamos o segmento $(0, a) - (10, b)$. A Figura a seguir ilustra duas rodadas do experimento: na primeira são selecionados $a = 3, b = 6$, e o segmento intercepta a circunferência; na segunda são selecionados $a = 9.2$ e $b = 5$, e o segmento não intercepta a circunferência.



1. Qual é a probabilidade do segmento interceptar a circunferência?
2. Defina uma variável indicadora (definida na Seção 3.1, página 60), $I = 1$ se o segmento intercepta a circunferência, $I = 0$ caso contrário. Determine $\mathbb{E}[I]$ e $\text{var}[I]$.
3. E se a for escolhido entre 0 e 10, mas b for selecionado entre 0 e 5?
4. E se as distribuições de a e b forem exponenciais, com parâmetros θ e λ?

Ex. 78 — Uma fábrica produz bolas de basquete. Nos últimos anos, as bolas produzidas não seguiam a especificação oficial, porque eram destinadas a jogadores ocasionais e pessoas sem forte envolvimento com o esporte. As bolas produzidas tem circunferência com média $75.9cm$ e desvio padrão $2cm$.
De acordo com as regras oficiais, uma bola para competições masculinas deve ter circunferência entre $74.9cm$ e $78.0cm$.

Também de acordo com as regras oficiais, uma bola deve pesar entre $567g$ e $650g$; as bolas da fábrica tem peso médio igual a $590g$, com desvio padrão de $20g$.
Presuma que as distribuições de peso e circunferência das bolas são normais e independentes.

1. Qual é a probabilidade de uma bola produzida por esta fábrica ter circunferência de acordo com a espeficiação oficial?
2. Qual é a probabilidade de uma bola produzida por esta fábrica ter circunferência *e peso* de acordo com a espeficiação oficial?

Ex. 79 — Uma câmara refrigerada mantém a temperatura que for programada, com desvio padrão de 0.3^oC (ou seja, medidas feitas ao longo do tempo seguem uma distribuição normal com esse desvio padrão). Se em 2% das medidas feitas em cada minuto de um dia a temperatura estiver acima de 4^oC, qual deve ter sido a temperatura programada?

Ex. 80 — Seja X uma variável aleatória com distribuição normal, tendo média 3. Sabendo que $\Pr[3 < X < 3.8] = 0.3289$, determine o desvio padrão de X.

Ex. 81 — Uma fábrica produz tubos de cobre. O diâmetro dos tubos, medido em centímetros, segue uma distribuição $\mathcal{N}(4, 0.04)$. Um tubo só passa no controle de qualidade se seu diâmetro está entre $3.9cm$ e $4.1cm$.

1. Qual é a proporção de tubos defeituosos produzida?
2. Qual é a probabilidade de vinte tubos serem produzidos sem defeito?

Ex. 82 — Quero definir uma distribuição que funcione como o quadrado da normal:

$$\hat{f}(x) = \left(\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \right)^2.$$

1. Mostre que $\hat{f}$ não é função de densidade.
2. Mostre que existe $K \in \mathbb{R}$ tal que $f = K\hat{f}$ é função de densidade (ou seja, $\hat{f}$ pode ser transformada em função de densidade multiplicando-a por uma constante).
3. Qual será a média da distribuição que usa f como função de densidade?
4. E a variância?
5. (Trabalhoso!) Quando definimos $\hat{f}$ elevamos a função de densidade da normal ao quadrado. Mostre como ficam a média e a variância se, ao invés de elevarmos ao quadrado, elevarmos a q, com q inteiro positivo.
6. A nova distribuição, tendo f como função de densidade, é normal?

Ex. 83 — Quero definir uma variável com distribuição contínua de probabilidades, de forma que a probabilidade $\Pr[X = x]$ seja dada por uma parábola invertida, com função de densidade de probabilidade

$$f(x) = \max(-4x^2 + c, 0).$$

Que opções tenho?

Ex. 84 — Uma variável aleatória X pode ter função de densidade igual à função abaixo?

$$f(x) = \begin{cases} \dfrac{1}{2\sqrt{e^x}} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Caso seja possível, qual seria sua esperança?

Ex. 85 — Suponha que 10% da superfície de uma esfera tenha sido pintada de uma cor.

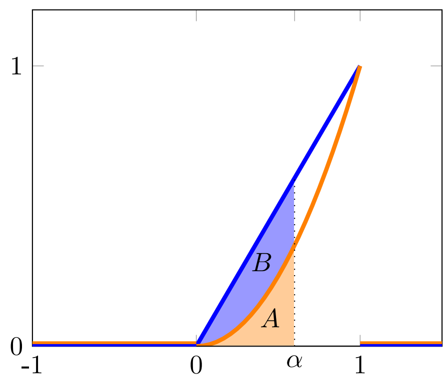
1. Prove que é possível inscrever um cubo nessa esfera de forma que nenhum de seus vértices toque em uma parte pintada.
2. Até que proporção da esfera podemos pintar para que o argumento do item (1) continue funcionando?
3. Refaça o exercício para circunferências em $\mathbb{R}^2$.
   Dicas: (*) ao escolher um ponto na esfera, os outros sete vértices do cubo já estão determinados; (*) tente criar uma variável aleatória indicadora para cada vértice que cai em parte não pintada.

Ex. 86 — No Exemplo 4.5.4 mencionamos um sistema interativo, onde a média de duração de conversas era de 90 segundos. Determine a probabilidade de uma conversa durar entre um e dois minutos.

Ex. 87 — Um modelo de lâmpada de LED tem tempo médio de funcionamento igual a 4 anos. Todas as lâmpadas instaladas em um prédio são deste modelo, e uma norma de segurança determina que 70% deve estar funcionando para que o prédio possa continuar sendo usado. Quanto tempo esperamos que passe antes que lâmpadas tenham que ser trocadas para que o prédio possa continuar funcionando?

Ex. 88 — No Capítulo 1, o Exercício 25 dizia que um número α era escolhido equiprovavelmente entre zero e um. Refaça o exercício, desta vez considerando

1. que α é escolhido de uma distribuição exponencial, com $\lambda = 1$.
2. que α é escolhido de uma distribuição normal, com $\mu = 1/2$ e $\sigma^2 = 0.25$ (como a função de densidade da normal não é integrável, deixe indicadas as integrais, sem resolvê-las). Neste caso, α ficar fora do intervalo $[0, 1]$, e as funções que são iguais a x e x^2 em $[0, 1]$ são iguais a zero fora desse intervalo.



Ex. 89 — Uma bateria de carro tem duração média de 24 meses. Sabendo que o carro já está sendo usado há exatos 24 meses, determine a probabilidade da bateria falhar em seu 36° mês de funcionamento.

Ex. 90 — Pessoas compram passagens aéreas com antecedência média de 20 dias, e com distribuição exponencial.

1. Qual é a probabilidade de uma pessoa comprar uma passagem com dez dias de antecedência, ou menos?
2. Com quanto tempo de antecedência metade das pessoas costuma comprar passagens?

Ex. 91 — Se selecionarmos três pontos em uma circunferência (na borda, não no interior), qual é a probabilidade desses pontos estarem em uma mesma semi-ciecunferência?

Ex. 92 — Prove o Teorema 4.5.11.

Ex. 93 — Sabe-se que o tempo que um aparelho pode funcionar sem manutenção tem distribuição exponencial. Se a probabilidade deste equipamento funcionar mais de um mês sem manutenção é 0.5, qual é a esperança do tempo de funcionamento sem manutenção?

Ex. 94 — Círculos são gerados aleatoriamente, e a área dos círculos (A) segue distribuição exponencial. Determine a função de densidade de probabilidade das variável aleatória R, que representa o raio do círculo.

Ex. 95 — Prove o Teorema 4.3.6.

Ex. 96 — Prove o Teorema 4.5.8,

Ex. 97 — Comprei um aparelho com MTBF ("tempo médio antes de falha") igual a 100 000 horas. Isso significa que a probabilidade dele falhar antes de 100 000 horas é pequena? Quanto, exatamente?

Ex. 98 — Se $X \sim \mathcal{N}(\mu, \sigma)$, determine

1. $\Pr[X > \sigma]$
2. $\Pr[X > 2\sigma]$
3. $\Pr[X > 3\sigma]$

Ex. 99 — Estudos determinaram que o ritmo de batimento cardíaco de recém-nascidos em uma população é de 130.28 por minuto, com desvio padrão de 8.22, seguindo uma distribuição normal. Qual é a probabilidade de um recém-nascido ter batimento com ritmo menor que 120 por minuto?

Ex. 100 — Números complexos são da forma $a+bi$, onde $a, b \in \mathbb{R}$ e $i = \sqrt{-1}$. Usualmente, associamos números complexos com pontos no plano, de forma que o número $a + bi$ seja associado ao ponto (a, b).

Sabe-se que não é possível ordenar completamente os números complexos: há números que não são comapráveis: $2+10i$ e $10+2i$ diferentes, mas não dizemos que um deles é "maior" que o outro. Há casos em que é possível dizer que um número complexo é maior que outro: $2+2i < 10+10i$, porque tanto a parte real como a imaginária do primeiro são menores que as correspondentes no segundo.

Considere o seguinte experimento: um número complexo é sorteado (a e b são escolhidos equiprovavelmente), de forma que a norma do número seja menor ou igual a um ($\sqrt{a^2 + b^2} \leq 1$); em seguida, outro número complexo $\alpha + \beta i$ é sorteado da mesma forma, com a mesma restrição.

Determine a probabilidade do segundo número ser menor que o primeiro.

Ex. 101 — Uma variável X tem distribuição normal. Se $\Pr[X > 30] = 0.7517$, qual a maior média possível para a variável, sabendo que o desvio padrão deve ficar abaixo de 2?

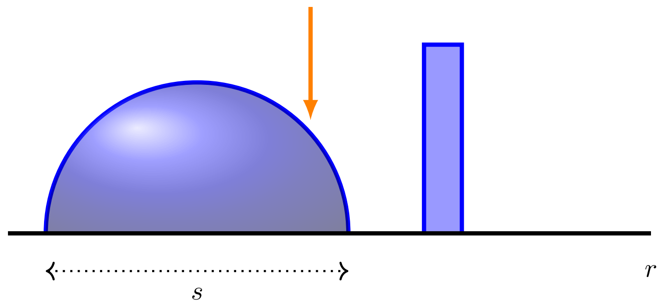
Ex. 102 — Seja $X \sim \mathcal{N}(1, 36)$. Determine $\Pr[0 \leq X^2 \leq 9]$.

Ex. 103 — Seja $X \sim \mathcal{N}(8, 9)$. Determine k tal que $\Pr[X < k] = 4\Pr[X > k]$.

Ex. 104 — Embora não exista forma fechada para a integração da função de densidade da distribuição normal, é possível provar que a integral dela em toda a reta real resulta em um. Mostre como.

Ex. 105 — Uma calota semiesférica está no chão, ao lado de um paralelepípedo. A calota tem raio igual a $20m$, e o paralelepípedo, que tem $25m$ de altura, fica a $10m$ da borda da calota. Seja r a reta que passa pela calota e exatamente pela metade do paralelepípedo.

Se um laser é direcionado, de forma perpendicular ao chão, para um ponto escolhido equiprovavelmente em s, o segmento de r dentro da calota, qual é a probabilidade deste raio atingir o paralelepípedo?



Ex. 106 — Obtenha a função de densidade e a função acumulada de densidade da distribuição generalizada de Cauchy. A distribuição na Seção 4.8 presume que o ponto P está em $(0, 1)$. Para a distribuição generalizada, presuma $P = (x_0, y_0)$.

Ex. 107 — Sejam $X_1, X_2, \ldots, X_n$ variáveis com distribuição generalizada de Cauchy (veja o Exercício 106), cada uma tendo função de densidade $f_1, f_2, \ldots, f_n$. Determine a função de densidade de $X_1 + X_2 + \cdots + X_n$.

Capítulo 5

Variáveis Aleatórias Bidimensionais

No início deste texto – já no Exemplo 1.1.4 – definimos experimentos que naturalmente sugerem a definição de mais de uma variável aleatória. No exemplo mencionado, jogavam-se dois dados, e perguntas eram feitas a respeito dos dois valores, individualmente. Claramente, poderíamos ter definido uma variável aleatória para cada um dos dados. Da mesma forma, no exemplo 2.3.6 podemos visualizar duas variáveis aleatórias: o resultado do teste, que pode ser positivo ou negativo, e a gravidez, que pode ou não existir. Este Capítulo discute o trabalho com mais de uma variável aleatória.

5.1 Vetores aleatórios

Duas variáveis aleatórias, X e Y, podem assumir diferentes valores, dependendo do resultado de um experimento. Conceitualmente, podemos vê-las como um vetor (X, Y), onde cada componente assume, individualmente, os valores de X e de Y. Desta forma, após um experimento, podemos voltar a atenção individualmente para o valor resultante de X ou de Y; ou podemos observar o valor resultante do vetor (X, Y), E, assim como X e Y, o vetor (X, Y) também terá uma função de distribuição de probabilidade. No entanto, enquanto X e Y são funções de Ω em $\mathbb{R}$, a função de distribuição de (X, Y) é $f_{XY} : \Omega \to \mathbb{R}^2$ (e o espaço amostral, Ω, pode ele mesmo ser formado por elementos tipicamente descritos como multidimensionais – por exemplo pontos dentro de uma esfera ou medidas feitas por equipamento hospitalar).

Esta discussão se estende naturalmente a mais de duas variáveis, conforme a Definição 5.1.1.

Definição 5.1.1 (variável aleatória multidimensional)**.** Sejam $X_1, X_2, \ldots, X_n$ variáveis aleatórias definidas em um mesmo espaço amostral. Então o vetor $(X_1, X_2, \ldots, X_n)$ é uma **variável aleatória multidimensional**, ou um **vetor**

aleatório.

Equivalentemente, um vetor aleatório pode ser definido como uma função do espaço amostral, resultando em vários números reais, $f : \Omega \to \mathbb{R}^n$.

Quando o vetor aleatório tem dimensão dois, dizemos que é um **vetor aleatório bidimensional** ◀

Exemplo 5.1.2 (estação meteorológica – variável aleatória bidimensional discreta)**.** Uma estação meteorológica usualmente tem vários sensores para registrar temperatura, pressão atmosférica, umidade do ar, velocidade e orientação do vento, dentre outros. Se tratarmos uma leitura periódica da estação como um experimento aleatório, a leitura de cada um destes sensores pode ser vista como uma variável aleatória – e o conjunto delas, como uma variável multidimensional. ●

Exemplo 5.1.3 (ponto escolhido em círculo – variável aleatória bidimensional contínua)**.** Escolhemos equiprovavelmente um ponto em um círculo C com centro na origem ($C = \{(x, y) \,:\, x^2 + y^2 < 1\}$). As coordenadas X e Y do ponto são variáveis aleatórias, e o *ponto* (X, Y) é um vetor aleatório (ou variável aleatória bidimensional). ●

5.2 Distribuições conjuntas

Um par de varáveis aleatórias tem uma distribuição discreta quando podemos enumerar as probabilidades que elas medem. Isso significa que devemos poder identificar dois conjuntos enumeráveis, A e B, tais que o valor da variável bidimensional sempre será $(a, b) \in A \times B$. A partir disto formulamos a seguinte definição.

Definição 5.2.1 (distribuição conjunta discreta)**.** A variável aleatória (X, Y) é **discreta** se existem A, B, subconjuntos enumeráveis de Ω, tais que para todo ponto $(x_i, y_j) \in A \times B$,

$$f(x_i, y_j) = \Pr[X = x_i, Y = y_j] > 0$$
$$\sum_{i,j} f(x_i, y_j) = 1.$$
◀

Consequentemente, a distribuição acumulada de probabilidade para X, Y é

$$F(x, y) = \sum_{\substack{x_i \leq x \\ y_i \leq y}} p_{ij}.$$

Exemplo 5.2.2 (estação meteorológica – distribuição conjunta discreta)**.** Tomamos dois sensores de uma estação meteorológica: o termômetro e o anemômetro. Deles podemos extrair duas variáveis aleatórias: T, a temperatura local; V, a velocidade aproximada do vento. Suponha que em no local onde a estação foi instalada, num memso horário do dia, a temperatura fica entre 9^oC e 12^oC, e a velocidade do vento sempre fica entre $11m/s$ e $13m/s$. Após 210 dias de medidas verificadas pela estação (dividida em anos diferentes), a tabela a seguir foi compilada (o número na coluna i e linha j é a quantidade de vezes em que a temperatura e velocidade do vento eram aquelas indicadas na linha/coluna correspondente).

		temperatura			
		9	10	11	12
vento	11	20	15	10	5
	12	10	25	25	20
	13	10	20	20	30

Se escolhermos equiprovavelmente um dos dias em que amostras foram feitas (ou um dia no futuro, se presumirmos que probabilidades calculadas a partir da tabela indicam comportamento futuro), a probabilidade da temperatura ser de 9^oC é

$$\Pr[T = 9] = \frac{20 + 10 + 10}{210} = \frac{1}{7}.$$

A probabilidade da temperatura ser 9^oC *e* da velocidade do vento ser de $12m/s$ é

$$\Pr[T = 9, V = 12] = \frac{10}{210} = \frac{1}{21}.$$

A probabilidade da temperatura ter sido menor ou igual a 11 e a velocidade do vento menor ou igual a 12 é

$$\begin{aligned}\Pr[T \le 11, V \le 12] &= \frac{20 + 15 + 10 + 10 + 25 + 25}{210} \\ &= \frac{1}{2}.\end{aligned}$$

É possível conceber experimentos mais interessantes que este. Pode-se, por exemplo, usar uma biruta para determinar a orientação aproximada do vento, e definir mais duas variáveis aleatórias: V_X, para a magnitude da componente da velocidade do vento em uma direção, e V_Y, para a magnitude da componente da velocidade em outra direção. ●

Exemplo 5.2.3 (duas moedas – função conjunta de massa)**.** Jogamos duas moedas duas vezes, sendo que para as duas p é a probabilidade de cara e $q = 1-p$ a probabilidade de coroa. Definimos duas variáveis:

- A, a quantidade de caras;
- O, a quantidade de coroas antes da primeira cara.

Tanto A como O podem assumir valores entre zero e dois. Como A e O foram definidas a partir do mesmo experimento, podemos tratá-las como um vetor aleatório, (A, O).

Os quatro eventos possíveis – e não necessariamente equiprováveis – são

X	$\Pr(X)$	A	O
AA	p^2	2	0
AO	pq	1	0
OA	qp	1	1
OO	q^2	0	2

Listamos agora a função de probabilidade conjunta de A e O, que consiste das probabilidades $\Pr[A = a, O = o]$, para todos os valores possíveis:

$$\begin{aligned} f(A = 1, O = 0) &= pq \\ f(A = 2, O = 0) &= p^2 \\ f(A = 1, O = 1) &= pq \\ f(A = 0, O = 2) &= q^2 \end{aligned}$$

Podemos também representar a função de probabilidade conjunta em uma tabela.

$$\begin{array}{cc|ccc} & & & A & \\ & & 0 & 1 & 2 \\ \hline & 0 & 0 & pq & p^2 \\ O & 1 & 0 & pq & 0 \\ & 2 & q^2 & 0 & 0 \end{array}$$

Verificamos que a soma de todas as probabilidades na tabela é um:

$$\begin{aligned} p^2 + q^2 + 2pq &= p(1-q) + q(1-q) + 2pq \\ &= p - pq + q - pq + 2pq \\ &= p + q \\ &= 1. \end{aligned}$$

●

Observamos as soma das linhas e das colunas no Exemplo 5.2.3. Na primeira linha,

$$\Pr[A = 0, O = 0] + \Pr[A = 1, O = 0] + \Pr[A = 2, O = 0] = pq + p^2 = p.$$

Mas esta soma, pelo Teorema da Probabilidade Total, é $\Pr[O = 0]$.

A função de densidade conjunta de um vetor aleatório contínuo é semelhante à função de densidade de uma variável aleatória contínua: esta é a função $f(x, y)$, tal que $\int\int(x, y)dxdy = \Pr[X \leq x, Y \leq y]$.

Definição 5.2.4 (função de densidade conjunta)**.** Seja (X, Y) um vetor aleatório bidimensional contínuo. Uma função

$$f : \Omega^2 \to \mathbb{R}$$

é uma **função de densidade conjunta** para (X, Y) se

$$f(x, y) \geq 0$$
$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\, dxdy = 1.$$

Como consequência,

$$F(x, y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(u, v)\, dudv.$$

◀

Exemplo 5.2.5 (distribuição conjunta uniforme)**.** Um ponto é selecionado no quadrado entre $(0, 0)$ e $(2, 2)$. Definimos duas variáveis aleatórias, X e Y, que terão o valor das duas coordenadas do ponto.

Como todos os pontos são equiprováveis, a função de densidade conjunta destas variáveis deve ser constante:

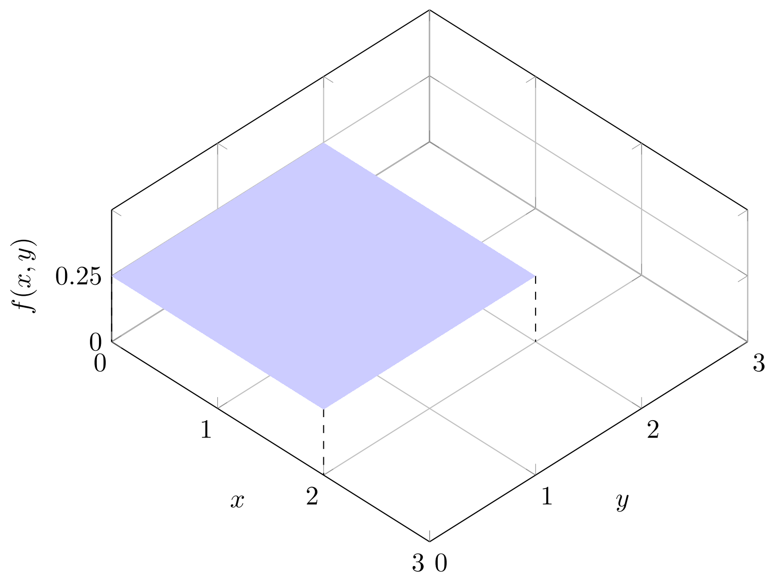
$$f(x, y) = \frac{1}{c}.$$

Para que seja uma função de densidade, sabemos que $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \frac{1}{c}\, dxdy$ deve ser um (ou ainda, $F(+\infty, +\infty) = 1$):

$$\begin{aligned}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \frac{1}{c}\, dxdy &= \int_{0}^{2}\int_{0}^{2} \frac{1}{c}\, dxdy \\ &= \frac{(2)(2)}{c},\end{aligned}$$

o que implica que $c = \frac{1}{4}$. Desta forma,

$$f(x, y) = \frac{1}{4}$$



Esta é uma característica da distribuição uniforme: quando abordamos a distribuição uniforme em intervalos no Capítulo 4, vimos que o valor da função de densidade é constante, e igual ao inverso do comprimento do intervalo,

$$f(x) = \frac{1}{b-a}.$$ ●

Teorema 5.2.6. Sejam X e Y duas variáveis aleatórias contínuas. Se sua a função acumulada conjunta de distribuição é $F(X,Y)$, então a função de densidade conjunta de X e Y será

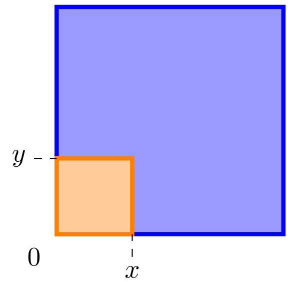
$$f(x,y) = \frac{\partial^2}{\partial x \partial y} F(x,y),$$

se esta derivada existir.

Exemplo 5.2.7 ($f(x,y)$ é derivada de $F(x,y)$)**.** No exemplo 5.2.5, a função de densidade encontrada foi $f(x,y) = 1/4$. A função de densidade acumulada é

$$F(x,y) = \frac{xy}{4},$$

porque esta é a função que dá a probabilidade do ponto sorteado estar no retângulo delimitado por x e y.



O que fizemos agora foi generalizar para duas variáveis, trocando comprimento por área, e usando o método também já visto, no Capítulo 1: se um ponto é escolhido equiprovavelmente em uma região R, a probabilidade do ponto estar dentro de uma outra região $A \subseteq R$ é

$$\frac{|A|}{|R|}.$$

Vemos, agora, que a função de densidade é a derivada da função de distribuição acumulada:

$$\begin{aligned} f(x,y) &= \frac{\partial^2}{\partial x \partial y} F(x,y) \\ &= \frac{\partial^2}{\partial x \partial y} \frac{xy}{4} \\ &= \frac{1}{4}. \end{aligned}$$

Verificamos que de fato f é função de densidade:

$$\begin{aligned} \int_0^2 \int_0^2 f(x,y) dx\, dy &= \int_0^2 \int_0^2 \frac{1}{4} dx\, dy \\ &= \frac{1}{4} \left. xy \right|_0^2 \\ &= 1. \end{aligned}$$

●

Da mesma forma que é ilustrado no Exemplo 4.1.3 para densidades de variáveis aleatórias individuais, a função de densidade deve ter sua integral em todo o domínio igual a um, mas isto não significa que não pode assumir valores maiores que um em partes do domínio.

Exemplo 5.2.8 (densidade maior que um)**.** Sejam X e Y duas variáveis aleatórias contínuas. Então

$$f(x,y) = \begin{cases} x+y & x,y \in [0,1] \\ 0 & x,y \notin [0,1] \end{cases}$$

é uma função de densidade de probabilidade para X e Y, *ainda que* $f(1,1) = 2$, *acima de* 1, *porque* $f(x,y)$ *não representa probabilidades individuais.* A única exigência que existe é que $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y) = 1$. Verificamos:

$$\begin{aligned}\int_0^1\int_0^1 f(x,y)\,dxdy &= \int_0^1\int_0^1 x+y\,dxdy\\ &= \int_0^1 \left[\frac{x^2}{2}+xy\right]_0^1 dy\\ &= \int_0^1 \frac{1}{2}+y\,dy\\ &= 1.\end{aligned}$$

●

Exemplo 5.2.9. Sejam X e Y variáveis aleatórias com a função de densidade conjunta

$$f_{XY}(x,y) = \begin{cases} 2ax^2+ay & x,y \geq 0,\ x+y<2\\ 0 & \text{em outros casos}\end{cases}$$

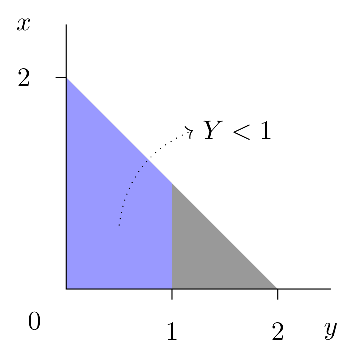
Determinaremos $\Pr[Y > 1]$. Para isso, precismos da constante c. Sabemos que a função de densidade deve intergar um, portanto

$$\begin{aligned}\int_0^2\int_0^{2-y} 2ax^2+ay\,dxdy &= a\int_0^2\int_0^{2-y} 2x^2+y\,dxdy\\ &= a\int_0^2 -\frac{2y^3-9y^2+18y-16}{3}\,dy\\ &= a(4).\end{aligned}$$

Concluimos, portanto, que $a = 1/4$, e temos a função de densidade conjunta

$$f_{XY}(x,y) = \frac{2x^2+y}{4} \quad x,y \geq 0,\ x+y<2.$$

A figura a seguir mostra a região onde devemos integrar f_{XY} e obter $\Pr[Y < 1]$, que depois inverteremos para obter $\Pr[Y > 1]$. Note que neste figura y é abscissa e x, ordenada.



$$\begin{aligned}
\Pr[Y > 1] &= 1 - \Pr[Y < 1] \\
&= 1 - \int_0^1 \int_0^{2-y} f_{XY}(x,y)\,dxdy \\
&= 1 - \frac{19}{24} \\
&\approx 0.208.
\end{aligned}$$

Podemos, equivalentemente, integrar y entre 1 e 2:

$$\begin{aligned}
\Pr[Y > 1] &= \int_1^2 \int_0^{2-y} f_{XY}(x,y)\,dxdy \\
&= \frac{5}{24} \\
&\approx 0.208.
\end{aligned}$$

●

5.3 Distribuições marginais

A distribuição de probabilidade da variável X, *independente do valor que Y possa ter*, é chamada de *distribuição marginal de X*.

As distribuições individuais de duas variáveis aleatórias X e Y não dão informação suficiente para que possamos deduzir sua distribuição conjunta. Na verdade, é possível que mais de uma distribuição conjunta sejam compatíveis com as mesmas distribuições individuais. O contrário já é possível: a partir da distribuição conjunta de X e Y, é possível identificar completamente as distribuições de X e de Y.

Definição 5.3.1 (distribuição marginal)**.** Sejam X e Y duas variáveis aleatórias com distribuição conjunta $f(x,y)$. A probabilidade de $X < x$, independente de qual seja o valor da variável Y, é dada pela **distribuição marginal** de X,

$$F_X(x) = F(x, +\infty),$$

ou seja,

$$F_X(x) = \begin{cases} \sum_{q<x} \sum_r f(q,r) & \text{para o caso discreto} \\ \int_{-\infty}^{x} \int_{-\infty}^{+\infty} f(u,v)\,dvdu & \text{para o caso contínuo} \end{cases}$$

◀

Exemplo 5.3.2 (duas moedas – distribuição marginal discreta)**.** No Exemplo 5.2.3, a distribuição marginal foi mostrada em uma tabela, que é reproduzida a seguir, onde também foram incluídas as somas das linhas e colunas. Verificamos que a soma de todas as probabilidades na tabela é um:

$$\begin{array}{cc|cccl}
 & & & A & & \\
 & & 0 & 1 & 2 & \\
\hline
 & 0 & 0 & pq & p^2 & \dashrightarrow p \\
O & 1 & 0 & pq & 0 & \dashrightarrow pq \\
 & 2 & q^2 & 0 & 0 & \dashrightarrow q^2 \\
 & & \downarrow & \downarrow & \downarrow & \\
 & & q^2 & 2pq & p^2 &
\end{array}$$

A distribuição marginal de A é uma binomial:

$$\begin{aligned}
\binom{2}{0}p^0q^2 &= q^2 \\
\binom{2}{1}p^1q^1 &= 2pq \\
\binom{2}{2}p^2q^0 &= p^2
\end{aligned}$$

Já a distribuição marginal de O é uma geométrica truncada.

$$\begin{aligned}
q^0p &= p \\
q^1p &= pq \\
q^2\cancel{p} &= q^2
\end{aligned}$$

O p não aparece na última linha porque a moeda só é jogada duas vezes, e não faz sentido exigir uma cara depois da segunda vez.

A partir das distribuições marginais, podemos obter a esperança e variância das variáveis individuais:

$$\begin{aligned}
\mathbb{E}[A] &= np = 2p && \text{(binomial)} \\
\mathbb{E}[O] &= 2q^2 + pq \\
&= 2q^2 + (1-q)q \\
&= 2q^2 + q - q^2 \\
&= q^2 + q. \\
\sigma_A^2 &= npq && \text{(binomial)} \\
&= 2pq \\
\sigma_O^2 &= \mathbb{E}[O^2] - \mathbb{E}[O]^2 \\
&= \left[0^2(p) + 1^2(pq) + 2^2(q^2)\right] - \left[(q^2+q)^2\right] \\
&= pq + 4q^2 - (q^2+q)^2 \\
&= pq + 4q^2 - q^4 + 2q^3 + q^2 \\
&= pq - q^4 + 2q^3 + 5q^2 \\
&= -q^4 + 2q^3 + 4q^2 + q.
\end{aligned}$$

●

Exemplo 5.3.3 (distribuição marginal discreta - Poisson–Poisson)**.** Suponha que duas variáveis independentes, A e B, tenham distribuição de Poisson, com médias λ_A e λ_B.

$$\begin{aligned}
f_A(a) &= \frac{e^{-\lambda_a}\lambda_A^a}{a!} \qquad (5.1) \\
f_B(b) &= \frac{e^{-\lambda_b}\lambda_B^b}{b!}
\end{aligned}$$

Naquele Exemplo, ao presumir que as duas variáveis são independentes, chegamos à função conjunta de massa

$$f(a,b) = \frac{e^{-\lambda_A}\lambda_A^a}{a!}\frac{e^{-\lambda_B}\lambda_B^b}{b!}.$$

Se calcularmos a distribuição marginal de A, chegamos à mesma $f_A(a)$ que já conhecemos (Equação 5.1):

$$\begin{aligned}
f_A(a) &= \sum_{b=0}^{\infty}\left(\frac{e^{-\lambda_A}\lambda_A^a}{a!}\right)\left(\frac{e^{-\lambda_B}\lambda_B^b}{b!}\right) \\
&= \left(\frac{e^{-\lambda_A}\lambda_A^a}{a!}\right)\sum_{b=0}^{\infty}\left(\frac{e^{-\lambda_B}\lambda_B^b}{b!}\right) \\
&= \left(\frac{e^{-\lambda_A}\lambda_A^a}{a!}\right) && \text{(porque } \textstyle\sum_{b=0}^{\infty} f_B(b) = 1\text{)}
\end{aligned}$$

O mesmo vale para $f_B(b)$. ●

Exemplo 5.3.4 (distribuição marginal contínua – uniforme)**.** O Exemplo 5.2.5 descreve um experimento onde um ponto é escolhido em um quadrado de lado 2. A função de densidade conjunta para (X, Y) é

$$f(x, y) = \frac{1}{4}.$$

A função de densidade marginal para X é obtida integrando $f(x, y)$ em y de 0 a 2

$$f_X(x) = \int_0^2 f(x, y) dy = \frac{1}{2}.$$

Obtemos também a função de densidade acumulada de X,

$$F_X(x) = \int_0^x \frac{1}{2} du = \frac{x}{2}$$

que claramente está correta: X é escolhida no intervalo $[0, 2]$, portanto a probabilidade de X ser menor que k é exatamente $k/2$. ●

Exemplo 5.3.5 (distribuição marginal contínua – exponencial–exponencial)**.** No Exemplo 5.4.6 as variáveis X e Y são independentes, e o exemplo já traz as distribuições marginais,

$$\begin{aligned} f_X(x) &= \lambda e^{-\lambda x} = \frac{1}{2} e^{-x/2} \\ f_Y(y) &= \theta e^{-\theta y} = \frac{2}{3} e^{-2y/3}. \end{aligned}$$

Verificamos que

$$\begin{aligned} \int_{-\infty}^{\infty} f_{XY}(x, y)\, dy &= \int_0^{\infty} \frac{1}{3} e^{-x/2-2y/3}\, dy \\ &= \lambda e^{-\lambda x} = \frac{1}{2} e^{-x/2} \\ &= f_X(x), \end{aligned}$$

e o mesmo vale para f_Y. ●

5.4 Independência

Já tratamos de independência de eventos em espaços amostrais na Seção 2.2; agora abordamos independência entre variáveis aleatórias. Queremos uma definição que capture o conceito semelhante para eventos.

Dois eventos A e B são independentes se $\Pr(A \cap B) = \Pr(A)\Pr(B)$.

Observamos que podemos definir eventos usando variáveis aleatórias:

$$\begin{aligned} &X \leq 5, \\ &X \geq 10, \\ &Y \leq 1 \end{aligned}$$

são definidos a partir das variáveis aleatórias X e Y, mas cada um deles é válido para um subconjunto do espaço amostral – ou seja, *cada um define um evento*. A definição de independência que procuramos, portanto, deve partir da afirmação de que *duas variáveis X e Y são independentes se quaisquer eventos A_X e A_Y definidos por X e Y são independentes.* Esta afirmação, apesar de definir claramente o que são variáveis independentes, não torna fácil o trabalho que teremos que realizar depois, de *verificar* se variáveis são independentes. A definição a seguir, no entanto, é equivalente a esta, e nos servirá perfeitamente.

Definição 5.4.1 (variáveis aleatórias independentes)**.** Duas variáveis aleatórias X e Y são independentes se e somente se para todos a, b,

$$F(a, b) = F_X(a)F_Y(b). \qquad \blacktriangleleft$$

Fica claro, a partir da motivação para a definição, que, se duas variáveis são definidas a partir de eventos entre os quais não há relação física relevante, podemos tratá-las como independentes (por exemplo, duas repetições diferentes do experimento em que se joga uma moeda).

A definição implica que duas variáveis aleatórias discretas X e Y são independentes se, para todos $x, y \in \Omega$,

$$\Pr[X = x, Y = y] = \Pr[X = x] \Pr[Y = y].$$

Exemplo 5.4.2 (variáveis discretas, dependentes)**.** Uma cidade compilou a incidência de dibetes mellitus em sua população, e obteve os seguintes resultados.

	fumante	não fumante	total
sem diabetes	7500	95700	103200
diabetes	10500	6300	16800
total	18000	102000	120000

A partir desta tabela, podemos identificar:

- a prevalência de tabagismo na cidade, $\Pr(T) = 18000/120000 = 0.15$;
- a prevalência de diabetes mellitus, $\Pr(D) = 16800/120000 = 0.14$;
- $\Pr[D, T] = 10500/120000 = 0.0875$, probabilidade de um cidadão ser tabagista e diabético.

Está implícito que definimos um experimento aleatório, onde selecionamos um indivíduo qualquer dentre os 120000 cidadãos, e que definimos duas variáveis aleatórias, D indicando diabetes, e T indicando tabagismo.

A tabela a seguir traduz as frequências em probabilidades (cada elemento é igual ao elemento da dabela de frequencias dividido por 120000, que é o tamanho do espaço amostral).

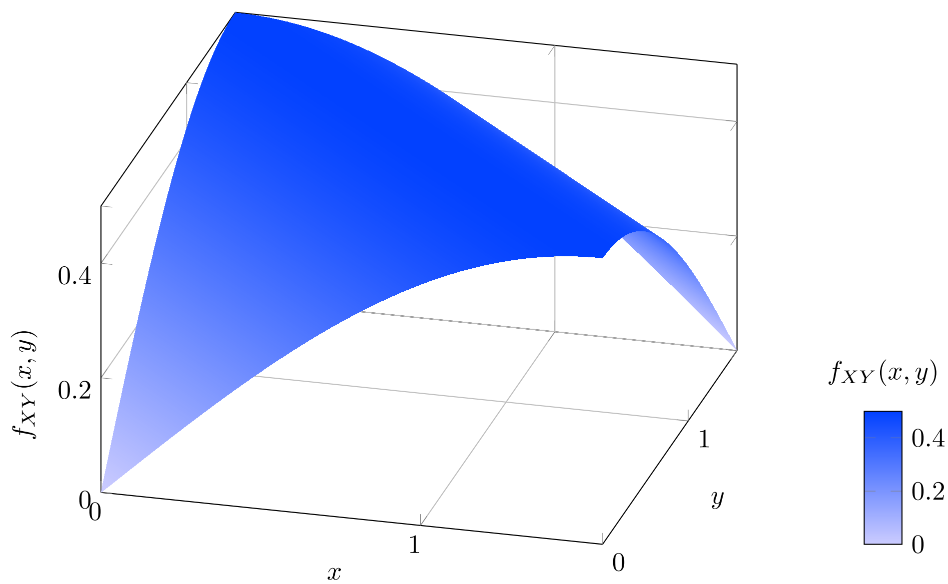
	fumante	não fumante
sem diabetes	0.0625	0.7975
diabetes	0.0875	0.0525

Podemos ver que as variáveis T e D não são independentes, porque

$$\Pr[D,T] = 0.0875$$
$$\Pr(D)\Pr(T) = (0.14)(0.15) = 0.021.$$

Exemplo 5.4.3 (variáveis contínuas, dependentes)**.** Suponha que duas variáveis aleatórias X e Y tem função de densidade conjunta

$$f(x,y) = a\,\text{sen}(x+y).$$



Como $\int\int f(x,y)dxdy$ deve ser um, só há um valor possível para a,

$$a = \frac{1}{2}$$

As distribuições marginais de X e Y são

$$F_X(x) = \frac{\text{sen}(x) - \cos(x) + 1}{2}$$
$$F_Y(y) = \frac{\text{sen}(y) - \cos(y) + 1}{2}$$

Já a função de densidade acumulada conjunta é

$$F_{XY}(x,y) = \frac{\text{sen}(x) + \text{sen}(y) - \text{sen}(x+y)}{2}.$$

No entanto,

$$F_X(x)F_Y(y) = \frac{(\operatorname{sen}(x) - \cos(x) + 1)(\operatorname{sen}(y) - \cos(y) + 1)}{4},$$

diferente de $F_{XY}(x, y)$ – o que pode ser verificado facilmente,

$$\begin{aligned} F_X(1)F_Y(1) &\approx 0.423 \\ F_{XY}(1,1) &\approx 0.386. \end{aligned}$$

Assim, as variáveis X e Y são *dependentes*. ●

Exemplo 5.4.4 (variáveis contínuas, dependentes)**.** Escolhemos um ponto equiprovavelmente na circunferência de raio um centrada na origem.

O espaço amostral para este experimento é

$$\Omega = \{(x, y) \ : \ x^2 + y^2 < 1\}.$$

Definimos duas variáveis aleatórias, X e Y, que nos dão as coordenadas X e Y do ponto. Para determinar se X e Y são dependentes, precisamos de f_X, f_Y e f_{XY}.

Como a distribuição dos pontos é uniforme no espaço amostral, temos

$$f_{XY} = \frac{1}{k}.$$

$$\begin{aligned} \iint_\Omega \frac{1}{k}\, dx\, dy &= 1 \\ \frac{1}{k} \iint_\Omega dx\, dy &= 1 \\ \frac{1}{k}\pi &= 1 && \text{(área de } \Omega \text{ é } \pi(1^2)) \\ k &= \pi. \end{aligned}$$

Já temos $f_{XY}(x, y) = 1/\pi$. Agora,

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f_{XY}(x, y)\, dy \\ &= \int_{-\sqrt{1-x}}^{+\sqrt{1-x}} \frac{1}{\pi}\, dy \\ &= \frac{2\sqrt{1-x^2}}{\pi}. \end{aligned}$$

Por simetria,

$$f_Y(y) = \frac{2\sqrt{1-y^2}}{\pi}.$$

Claramente, X e Y são depentendes, porque

$$\begin{aligned} f_X(x)f_Y(y) &= \frac{1}{\pi} \\ &= \frac{4\sqrt{(1-x^2)(1-y^2)}}{\pi^2} \\ &\neq \frac{1}{\pi} \\ &= f_{XY}(x,y). \end{aligned}$$

●

Exemplo 5.4.5 (variáveis discretas, independentes (Poisson, Binomial))**.** Um computador configurado para funcionar como servidor em uma rede recebe dois tipos de conexão: as do tipo TCP e as do tipo UDP. As conexões chegam de acordo com um processo de Poisson, com média de λ conexões por minuto. Dentre as conexões que chegam em um minuto, a probabilidade de uma conexão qualquer ser do tipo TCP é p, e consequentemente, a probabilidade da conexão ser do tipo UDP é $1-p$. Definimos as variáveis aleatórias T = quantidade de conexões TCP, U = quantidade de conexões UDP, e $N = T + U$, e obteremos a função de distribuição conjunta de T e U.

Queremos $f_{TU} = \Pr[T = t, U = u]$. Podemos somar, para todos os valores de N,

$$\Pr[T = t, U = u] = \sum_{n=0}^{\infty} \Pr[T = t, U = u] \Pr[N = n].$$

Mas como só precisamos considerar o caso em que $n = t + u$, reescrevemos

$$\Pr[T = t, U = u] = \Pr[T = t | N = t + u] \Pr[N = t + u]$$

Se soubermos a quantidade de pacotes ($N = n$, por exemplo), podemos definir T' e U', com

$$\begin{aligned} \Pr[T' = t] &= \Pr[T = t | N = t + u] \\ \Pr[U' = u] &= \Pr[U = u | N = t + u] \end{aligned}$$

e sabemos, pela descrição dada ("dentre as conexões que chegam em um minuto, …"), que tanto T' como U' são binomiais: $T' \sim \mathcal{B}inom(n, p)$, $U' \sim \mathcal{B}inom(n, 1-p)$. Então

$$\begin{aligned} f_{TU}(t,u) &= \Pr[T = t, U = u] \\ &= \overbrace{\Pr[T = t | N = t + u]}^{\mathcal{B}inom(t+u,p)} \overbrace{\Pr[N = t + u]}^{\mathcal{P}oiss(\lambda)} \\ &= \binom{t+u}{t} p^t (1-p)^u \; \frac{e^{-\lambda}\lambda^{t+u}}{(t+u)!} \\ &= \frac{e^{-\lambda p}(\lambda p)^t}{t!} \; \frac{e^{-\lambda(1-p)}\left[\lambda(1-p)\right]^u}{u!} \end{aligned}$$

Agora é interessnate observar que

$$
\begin{aligned}
\Pr[T=t] &= \sum_{u=0}^{\infty} \Pr[T=t, U=u] \\
&= \sum_{u=0}^{\infty} \frac{e^{-\lambda p}(\lambda p)^t}{t!}\ \frac{e^{-\lambda(1-p)}\left[\lambda(1-p)\right]^u}{u!} \\
&= \frac{e^{-\lambda p}(\lambda p)^t}{t!}\ \left(\sum_{u=0}^{\infty} \frac{e^{-\lambda(1-p)}\left[\lambda(1-p)\right]^u}{u!}\right) \qquad (5.2)\\
&= \frac{e^{-\lambda p}(\lambda p)^t}{t!}\ (1),
\end{aligned}
$$

porque a expressão entre parênteses na equação 5.2 é a soma das probabilidades para todos os possíveis valores em uma distribuição de Poisson – que deve ser um. Mas isto mostra que $T \sim \mathcal{P}oiss(\lambda p)$, que faz sentido: a chegada de pacotes TCP segue uma distribuição de Poisson com média λp por minuto (p é a probabilidade do pacote ser TCP, portanto λp é a proporção de pacotes TCP que chegam em um minuto). Da mesma forma se conclui que $U \sim \mathcal{P}oiss(\lambda(1-p))$ – e como a distribuição conjunta é o produto das duas distribuições marginais, T e U são independentes! É claro que, se N é conhecido, então T e U não são independentes, porque um é função linear do outro ($T = N - U$), mas quando consideradas sem conhecimento de N, as duas variáveis são independentes. ●

Exemplo 5.4.6 (variáveis contínuas, independentes – exponencial)**.** Duas baterias em tempo de duração modelado por distribuições exponenciais: a primeira dura 2 anos, em média, e a segunda dura um ano e meio. As garantias das duas são iguais às suas médias de duração.

Os parâmetros das duas distribuições exponenciais são

$$
\begin{aligned}
\lambda &= 1/2 \\
\theta &= 3/2.
\end{aligned}
$$

As variáves são independentes, e as funções de densidade são

$$
\begin{aligned}
f_X(x) &= \lambda e^{-\lambda x} = \frac{1}{2}e^{-x/2} \\
f_Y(y) &= \theta e^{-\theta y} = \frac{2}{3}e^{-2y/3}.
\end{aligned}
$$

A probabilidade des cada bateria ter que ser trocada dentro do seu prazo de garantia é a probabilidade delas durarem menos que a média, portanto

$$
\begin{aligned}
1 - e^{-\lambda u} &= 1 - e^{-2(1/2)} = 1 - e^{-1} \approx 0.632 \\
1 - e^{-\theta t} &= 1 - e^{-(3/2)2/3} 1 - e^{-1} \approx 0.632
\end{aligned}
$$

Como são independentes, a função de densidade conjunta é dada por

$$\begin{aligned} f_{XY}(x,y) &= \left(\frac{1}{2}e^{-x/2}\right)\left(\frac{2}{3}e^{-2y/3}\right) \\ &= \frac{1}{3}e^{-x/2-2y/3}, \end{aligned}$$

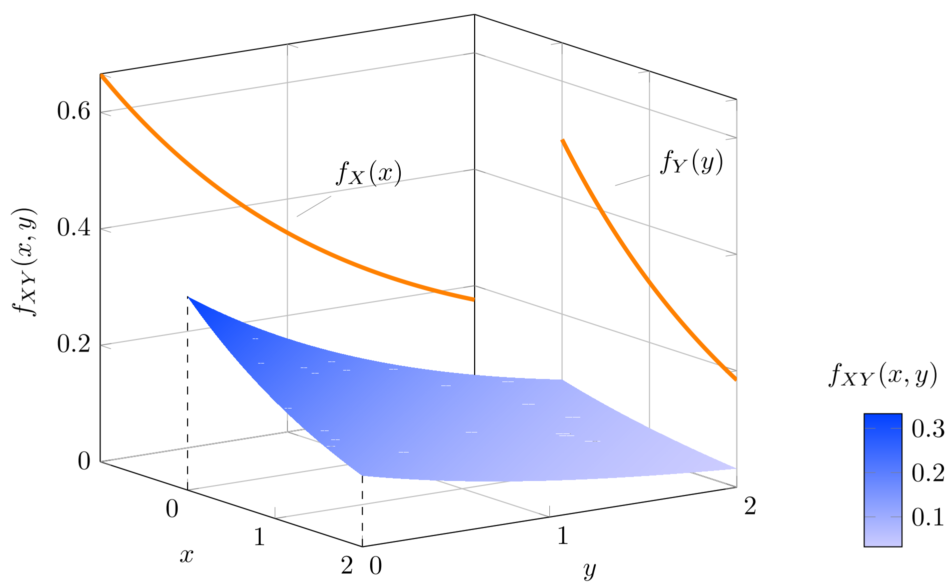
e a função de densidade acumulada é

$$\begin{aligned} F_{XY}(x,y) &= \int_{-\infty}^{y}\int_{-\infty}^{x} \frac{1}{3}e^{-u/2-2v/3}\,dudv \\ &= e^{-2y/3-x/2} - e^{-2y/3} - e^{-x/2} + 1. \end{aligned}$$

A probabilidade de que as duas funcionem por dois anos é

$$F(2,2) = e^{-4/3-1} - e^{-1} - e^{-4/3} + 1 \approx 0.465.$$

A próxima figura mostra a distribuição conjunta $f(x,y)$.



O máximo de f_X é $1/2$, e o máximo de F_Y é $2/3$, porque

$$\begin{aligned} f_X(0) &= \lambda e^{-\lambda 0} = 1/2 \\ f_Y(0) &= \theta e^{-\theta 0} = 2/3. \end{aligned}$$

Mas o máximo de F_{XY} é $1/3$:

$$f_{XY}(0,0) = \lambda e^{-\lambda 0}\theta e^{-\theta 0} = (2)(1.5) = 1/3.$$

Isso significa que os gráficos das distribuições de f_X e f_Y *não* são projeções do gráfico de f. ●

Teorema 5.4.7. Se uma tabela de distribuição conjunta para duas variáveis X e Y contém uma entrada igual a zero sem que sua linha ou sua coluna tenha somente zeros, as variáveis são dependentes.

Demonstração. Suponha que $\Pr[X = x, Y = y] = 0$. Se as variáveis fossem independentes, teríamos $\Pr[X = x] \Pr[X = y] = 0$, o que implicaria em

$$\Pr[X = x] = 0 \quad \text{ou} \quad \Pr[Y = y] = 0.$$

Mas se $\Pr[X = x] = 0$, então linha (ou coluna) de X só poderia conter zeros. ∎

Exemplo 5.4.8. Em uma plantação de mamões há uma pequena incidência de podridão devido a infecção por dois fungos: Fusarium e Geotrichum. mamões selecionados aleatoriamente durante dois períodos (inverno e verão) foram inspecionadas. Se tratarmos todos os mamões como espaço amostral, podemos definir duas variáveis aleatórias:

- E, a estação: -1 para inverno, $+1$ para verão;
- D, a doença: -1 para Geotrichum, 0 para saudável, $+1$ para Fusarium.

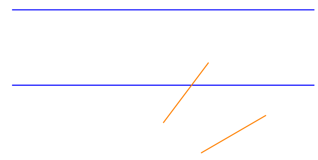
Os dados obtidos estão na tabela a seguir.

	inverno	verão	total
Fusarium	100	400	500
Geotrichum	350	0	350
saudável	20150	20100	40250
total	20600	20500	41100

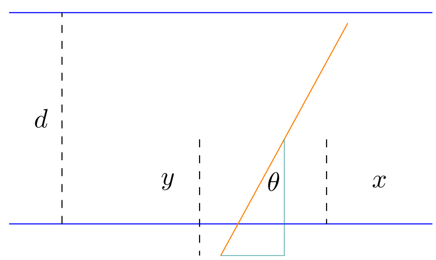
Como não houve incidência de Geotrichum no verão, e o zero (Geotrichum/verão) não está um uma coluna ou linha nula, as duas variáveis são dependentes.

Isto não significa que se pode generalizar e dizer que as duas variáveis sempre são dependentes. Estritamente falando, *o exemplo não trata da probabilidade de mamões serem infectados por fungos, mas de – ao selecionarmos um mamão dentre <u>estes</u> 41100, verificarmos a presença de fungos e a estação em que o mamão foi colhido.* Se um novo experimento fosse conduzido em outro lote de mamões, poderíamos ter dois ou três na posição onde há um zero, e não poderíamos mais usar a regra que usamos para determinar que as variáveis são dependentes. Dito isto, é possível tentar generalizar a afirmação usando métodos da Estatística (e de fato, concluiríamos que é mais comum que Geotrichum ataque mamões no inverno, e Fusarium no verão), mas estes métodos não estão no escopo deste texto. ●

Exemplo 5.4.9 (agulha de Buffon)**.** O problema descrito a seguir foi proposto pelo Conde de Buffon, Georges-Louis Leclerc. Uma agulha de comprimento ℓ é lançada sobre uma superfície listrada (composta por faixas de igual largura), sendo d a largura das faixas. Presumindo que $\ell < d$, determine a probabilidade da agulha cair cruzando uma das faixas. A figura a seguir ilustra duas agulhas – uma cruzando duas faixas ,e outra completamente dentro de uma faixa.



Definimos duas variáveis aleatórias: T o ângulo entre a agulha e as faixas; e X, a menor distância do ponto médio da agulha até alguma das faixas. Denotaremos os valores de T e X por θ e x.



Na figura anterior, x é a distância do ponto médio da agulha até a borda da faixa, e calculamos y:

$$\begin{aligned}\cos\theta &= \frac{y}{\ell/2}\\ \frac{\ell\cos\theta}{2} &= y\end{aligned}$$

Para que a agulha cruze a faixa de que está mais próxima, é necessário que

$$\begin{aligned}x &< y\\ x &< \frac{\ell\cos\theta}{2}.\end{aligned}$$

Sabemos que $X \in [0, d/2]$ e $T \in [0, \pi/2]$. Presumimos que as duas variáveis são independentes (não há motivo para crer que o ângulo está relacionado com a distancia de qualquer forma), portanto a função de densidade conjunta é

$$\begin{aligned}f_{TX}(\theta, x) &= \frac{2}{d}\frac{2}{\pi}\\ &= \frac{4}{d\pi}.\end{aligned}$$

A probabilidade $\Pr[X < \ell\cos(T)/2]$ é

$$\begin{aligned}\Pr\left[X < \ell\cos(T)/2\right] &= \int_0^{\pi/2}\int_0^{(\ell\cos\theta)/2}\frac{4}{d\pi}\,dxd\theta\\ &= \int_0^{\pi/2}\frac{2\ell\cos(\theta)}{d\pi}d\theta\\ &= \frac{2\ell}{d\pi}.\end{aligned}$$

A condição $X < \ell\cos(T)/2$ foi traduzida nos limites de integração: para todo θ, integramos x somente de zero até $\ell\cos(T)/2$. ●

5.5 Distribuições condicionais

Se X e Y são duas variáveis aleatórias discretas, sabemos que a probabilidade condicional de $Y = y$ dado que $X = x$ é

$$\Pr[Y = y|X = x] = \frac{\Pr(x,y)}{\Pr(x)}. \tag{5.3}$$

É possível calcular, portanto, os valores de $\Pr(y|x)$ para cada possível x, desde que $\Pr(x) > 0$ (se a probabilidde de x é zero não faz sentido tentar calcular $\Pr(y|x)$ – e a equação 5.3 nos levaria a uma divisão pode zero neste caso). Se reescrevermos a equação usando funções de massa de probabilidade ($\Pr(x,y) = f_{xy}(x,y)$, e $\Pr(x) = f_X(x)$), obteremos

$$\Pr[Y = y|X = x] = \frac{\Pr(x,y)}{\Pr(x)}.$$

Isto motiva a definição de *distribuição condicional de massa de probabilidade.*

Definição 5.5.1 (Distribuição condicional de massa)**.** Sejam X e Y variáveis aleatórias discreta com funções de massa f_X e f_Y. Então as **distribuições condicionais** de $X|Y$ e a de $Y|X$ são

$$\begin{aligned} g_{X|Y}(x|y) &= \frac{f_{X,Y}(x,y)}{f_Y(y)} \\ g_{Y|X}(y|x) &= \frac{f_{X,Y}(x,y)}{f_X(x)}. \end{aligned}$$ ◀

Demos a $g_{Y|X}$ o nome "função de massa" portanto verificamos que ela é, de fato, uma função de massa de probabilidade.

$$\begin{aligned} \sum_y g_{Y|X}(y|x) &= \sum_y \frac{f_{X,Y}(x,y)}{f_X(x)} \\ &= \frac{\sum_y f_{X,Y}(x,y)}{f_X(x)} \\ &= \frac{f_X(x)}{f_X(x)} \\ &= 1. \end{aligned}$$

Evidentemente, o mesmo vale para $g_{X|Y}$.

Exemplo 5.5.2. Para sortear um número entre 1 e 3, um dado é jogado e o resultado é dividido por 2 e arredondado para cima:

⚀, ⚁ $\rightarrow 1$
⚂, ⚃ $\rightarrow 2$
⚄, ⚅ $\rightarrow 3$

Com três repetições, obtém-se uma sequência de 3 números entre 1 e 3. Ignoraremos a ordem dos números, portanto há dez resultados possíveis:

sequência obtida nos dados	resultado	probabilidade
1, 1, 1	(1, 1, 1)	1/27
1, 1, 2 1, 2, 1 2, 1, 1	(1, 1, 2)	3/27
2, 2, 3 2, 3, 2 3, 2, 2	(2, 2, 3)	3/27
1, 2, 3 1, 3, 2 2, 1, 3 2, 3, 1 3, 1, 2 3, 2, 1	(1, 2, 3)	6/27
2, 2, 2	(2, 2, 2)	1/27
1, 1, 3 1, 3, 1 3, 1, 1	(1, 1, 3)	3/27
1, 3, 3 3, 1, 3 3, 3, 1	(1, 3, 3)	3/27
3, 3, 3	(3, 3, 3)	1/27
1, 2, 2 1, 2, 1 2, 1, 1	(1, 2, 2)	3/27
2, 3, 3 2, 3, 2 3, 2, 2	(2, 3, 3)	3/27

Definimos as variáveis aleatórias

- S, a soma dos valores;
- M, o máximo dos valores.

A função de massa de probabilidade para S é f_S, mostrada na tabela a seguir.

3	4	5	6	7	8	9
$\frac{1}{27}$	$\frac{3}{27}$	$\frac{6}{27}$	$\frac{7}{27}$	$\frac{6}{27}$	$\frac{3}{27}$	$\frac{1}{27}$

A função de massa para M é f_M, na próxima tabela.

1	2	3
$\frac{1}{27}$	$\frac{7}{27}$	$\frac{20}{27}$

A função conjunta de massa f_{SM} é

$$
\begin{array}{cc|ccccccc}
 & & & & & S & & & \\
 & & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\
\hline
 & 1 & \frac{1}{27} & 0 & 0 & 0 & 0 & 0 & 0 \\
M & 2 & 0 & \frac{3}{27} & \frac{3}{27} & \frac{1}{27} & 0 & 0 & 0 \\
 & 3 & 0 & 0 & \frac{3}{27} & \frac{6}{27} & \frac{6}{27} & \frac{3}{27} & \frac{1}{27}
\end{array}
$$

Estamos interessados no evento "maior número igual a dois". Este evento contém apenas

$$(1, 1, 2), (1, 2, 2), (2, 2, 2).$$

A função de massa de $S|M = 2$ é $f_{S|M=2}(s) = F_{SM}(s, 2)/f_M(2)$:

3	4	5	6	7	8	9
0	$\frac{3}{7}$	$\frac{3}{7}$	$\frac{1}{7}$	0	0	0

Os mesmos valores podem ser obtidos consultando a tabela de eventos, sendo necessário observar que o espaço amostral passa a ter tamanho 7 (que são as possíveis sequências de valores com máximo igual a 2). ●

Quando X e Y são contínuas, $\Pr[X = x]$ sempre será zero, e o raciocínio que desenvolvemos para variáveis discretas não pode ser usado. Podemos, no entanto, *definir* a distribuição condicional de densidade de X e Y da mesma forma, tendo a definição do caso discreto como motivação, e em seguida verificar que de fato obtemos uma função de densidade.

Definição 5.5.3 (Distribuição condicional de densidade)**.** Sejam X e Y variáveis aleatórias contínuas com funções de densidade f_X e f_Y. Então as **distribuições condicionais** de $X|Y$ e a de $Y|X$ são

$$
\begin{aligned}
g_{X|Y}(x|y) &= \frac{f_{X,Y}(x, y)}{f_Y(y)} \\
g_{Y|X}(y|x) &= \frac{f_{X,Y}(x, y)}{f_X(x)}.
\end{aligned}
$$
◀

Integramos $g_{Y|X}$ para verificar que de fato é uma função de densidade de probabilidade. Os passos são idênticos aos realizados no caso discreto:

$$
\begin{aligned}
\int_{-\infty}^{\infty} g_{Y|X}(y|x)\, dy &= \int_{-\infty}^{\infty} \frac{f_{X,Y}(x, y)}{f_X(x)}\, dy \\
&= \frac{\int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy}{f_X(x)} \\
&= \frac{f_X(x)}{f_X(x)} \\
&= 1.
\end{aligned}
$$

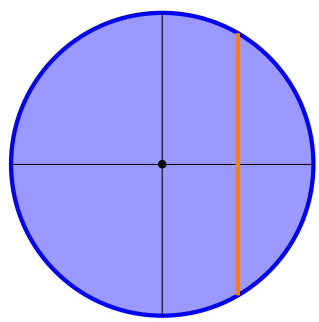
Exemplo 5.5.4 (distribuição condicional contínua)**.** No Exemplo 5.4.4 escolhemos equiprovavelmente um ponto em um círculo. Definimos duas variáveis aleatórias, X e Y, para suas coordenadas, e verificamos que

$$\begin{aligned} f_X(x) &= \frac{2\sqrt{1-x^2}}{\pi} \\ f_{XY}(x,y) &= 1/\pi. \end{aligned}$$

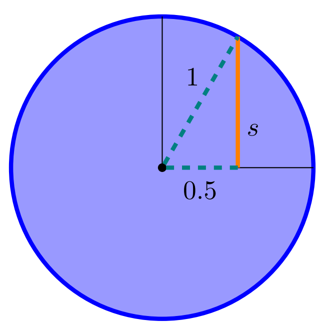
Determinaremos $\Pr[Y < 0.5 | X = 0.5]$. Primiero, calculamos

$$\begin{aligned} f_{Y|X}(x,y) &= \frac{f_{X,Y}}{f_X(x)} \\ &= \frac{1/\pi}{\left(2\sqrt{1-x^2}\right)/\pi} \\ &= \frac{1}{2\sqrt{1-x^2}}. \end{aligned}$$

Queremos a probabilidade de $Y < 0.2$, dado que $X = 0.5$. Se $X = 0.5$, o ponto certamente está no arco perpendicular ao eixo X e que passa por 0.5.



O segmento é menor que $[-1, +1]$. Obtemos o tamanho de cada parte (acima e abaixo de zero) do segmento usando o Teorema de Pitágoras:



Assim, $1^2 = s^2 + 0.5^2$, e $s = \sqrt{0.75}$. O segmento tem tamanho total $2\sqrt{0.75}$.

Assim,

$$
\begin{aligned}
\Pr[Y < 0.2 | X = 0.5] &= \int_{-\infty}^{0.2} f_{Y|X}(x, y)\, dy \\
&= \int_{-\infty}^{0.2} \frac{1}{2\sqrt{1 - x^2}}\, dy \\
&= \frac{1}{2\sqrt{1 - (0.5)^2}} \int_{-\sqrt{0.75}}^{0.2} dy \\
&= \frac{1}{2\sqrt{0.75}}\, y\Big|_{-\sqrt{0.75}}^{0.2} \\
&= \frac{0.2 + \sqrt{0.75}}{2\sqrt{0.75}} \\
&\approx 0.6154.
\end{aligned}
$$

É possível, a fim de convencimento e para este exemplo, realizar o mesmo cálculo geometricamente. Queremos a probabilidade de $X < 0.2$, portanto o espaço amostral é o segmento inteiro, e o evento é a parte do segmento abaixo de 0.2:

$$
\begin{aligned}
\Pr[Y < 0.2 | X = 0.5] &= \frac{\sqrt{0.75} + 0.2}{2\sqrt{0.75}} \\
&\approx 0.6154.
\end{aligned}
$$

É interessante que a integral que calculamos antes dá exatamente o comprimento do segmento. Esta verificação geométrica nem sempre é possível; este exemplo foi iconstruído em um objeto geométrico a fim de permitir realizá-la. ●

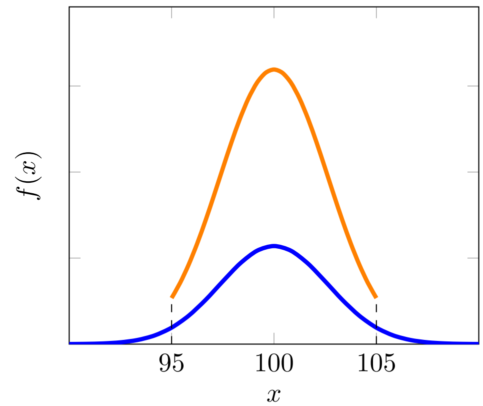
O Exemplo 5.5.5 ilustra o uso de uma distribuição condicional ao tratarmos de uma distribuição normal truncada.

Exemplo 5.5.5 (Distribuição normal truncada – capacitores)**.** [1] Componentes eletrônicos passivos tem um *valor nominal*, que é a esperança do valor. Por exemplo, um capacitor marcado como “$100nF$” (100 nanoFarads) pode não ter a capacitância exatamente igual a $100nF$, mas a média das capacitâncias na produção daquela fábrica (para capacitores com a marcação “$100nF$” é 100.

O valor da capacitância *poderia* ser representado por uma variável com distribuição normal. No entanto, além de valor nominal, estes componentes tem uma marcação de *tolerância*, determinando que o valor não está a mais que uma certa distância da média.

Suponha que uma fábrica produza capacitores de $100nF$, com tolerância de 5%. A capacitância de todos os capacitores produzidos é modelada por uma distribuição normal, com $\mu = 100$ e $\sigma = 7$. No processo de controle de qualidade, os capacitores com valor menor que $95nF$ ou maior que $105nF$ (fora da tolerância) são rejeitados, e a distribuição da capacitância para os lotes após esta fase, é uma distribuição normal truncada.

[1] Adaptado de Soong, T. T., “Fundamentals of Probability and Statistics for Engineers”.



Obteremos as funções de densidade e densidade acumulada. Definimos A como o evento em que $95 < C < 195$.

$$\begin{aligned} F(c|A) &= \Pr[C \leq c | 95 < C < 105] \\ &= \frac{\Pr[C < c \ \ \mathbf{e} \ \ 95 < C < 105]}{\Pr[95 < C < 105]} \end{aligned}$$

Observamos que

$$\Pr[C < c \ \ \mathbf{e} \ \ 95 < C < 105] = \begin{cases} 0 & c < 95 \\ \Pr[95 < C < c] & c \in [95, 105] \\ \Pr[95 < C < 105] & c > 105 \end{cases}$$

$$F_C(c|A) = \begin{cases} 0 & c < 95 \\ \dfrac{\Pr[95 < C < c]}{\Pr[95 < C < 105]} & c \in [95, 105] \\ 1 & c > 115 \end{cases}$$

onde

- $\Pr[95 < C < c] = \int_{95}^{c} f(x)\, dx$,
- $\Pr[95 < C < 105]$ é constante, $\int_{95}^{105} f(x)\, dx$.

A densidade $f_C(c)$ pode ser obtida derivando F_C, e obtemos

$$f_C(c|A) = \frac{dF_C(c|A)}{dc} = \begin{cases} \dfrac{f(c)}{\int_{95}^{105} f(x)\, dx} & c \in [95, 105] \\ 0 & \text{em outros casos.} \end{cases}$$ ●

O Exemplo 5.5.5 apresenta uma distribuição normal truncada de maneira simétrica. É possível também truncá-la de forma assimétrica, ou mesmo truncar somente em um dos lados.

5.6 Funções de duas variáveis aleatórias

Para podermos definir covariância e correlação, será necessário antes definir esperança da função de duas vraiáveis, $\mathbb{E}[g(X,Y)]$.

Primeiro notamos que podemos definir esperança para tais funções, porque, sendo funções de duas variáveis aleatórias, podem ser vistas também como variáveis aleatórias:

$$\begin{aligned} X(e) &: \Omega \to \mathbb{R} \\ Y(e) &: \Omega \to \mathbb{R} \\ g(X,Y) = g(X(e),Y(e)) &: \Omega \to \mathbb{R} \end{aligned}$$

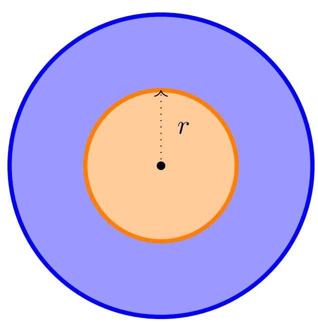
A definição é semelhante à de esperança para função de uma variável aleatória.

Teorema 5.6.1. Sejam X e Y duas variáveis definidas em um mesmo espaço amostral, e $g : \mathbb{R}^2 \to \mathbb{R}$ uma função. Então

$$\mathbb{E}[g(X,Y)] = \begin{cases} \sum_x \sum_y g(x,y)f(x,y) & \text{se } X,Y \text{ são discretas} \\ \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f(x,y)dxdy & \text{se } X,Y \text{ são contínuas} \end{cases}$$

Exemplo 5.6.2 (coordenadas de ponto em círculo – esperança do raio)**.** Escolhemos equiprovavelmente um ponto em um círculo C com centro na origem ($C = \{(x,y) \; : \; x^2 + y^2 < 1\}$). As coordenadas X e Y do ponto são variáveis aleatórias.

O novo ponto define um círculo possivelmente menor que o círculo original.



O raio do círculo menor, R, é função de X e Y, e portanto é também uma variável aleatória:

$$R = \sqrt{X^2 + Y^2}.$$

A esperança de R é

$$\mathbb{E}[R] = \frac{1}{\pi}\int\int_C \sqrt{x^2+y^2}\,dxdy.$$

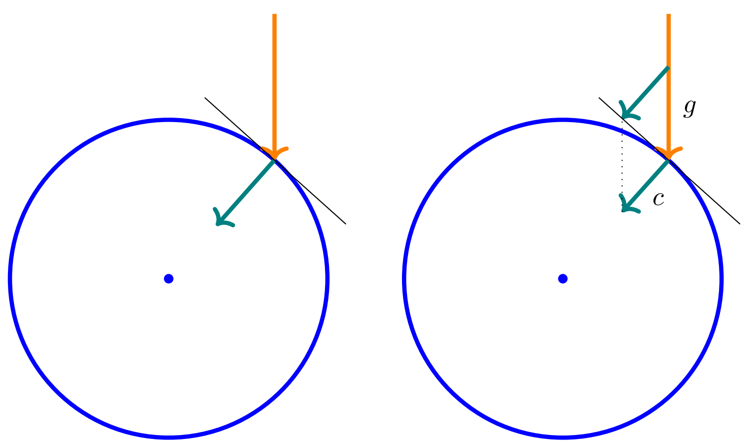
Para integrar, mudamos para coordenadas polares, usando $r = \sqrt{x^2 + y^2}$:

$$\begin{aligned}
\mathbb{E}[R] &= \frac{1}{\pi} \int_0^{2\pi} \int_0^1 r^2 \, dr d\alpha \\
&= \frac{1}{\pi} (2\pi) \frac{r^3}{3} \Big|_0^1 \\
&= \frac{2}{3}.
\end{aligned}$$

Como a esperança do raio do círculo menor é $2/3$, a esperança da área deve ser $\pi(2/3)^2 = 4\pi/9$. ●

Exemplo 5.6.3 (Crateras lunares – esperança de componente de força)**.** A Seção 4.7 discute a distribuição de ângulos formados por meteoros atingindo a superfície da Lua. Abordamos agora um exemplo semelhante: partículas seguem em direção a uma esfera, formando um ângulo ao chegar na superfície. Como já discutido na Seção 4.7, o ângulo não depende da direção de onde a partícula chega, mas da distância entre sua trajetória retilínea e o centro da esfera.

Definimos duas variáveis aleatórias, A para o ângulo de incidência (entre 0 e $\pi/2$, com função de densidade de probabilidade igual a $\text{sen}(2\alpha)$) e G para a magnitude da força aplicada pela partícula (que neste exemplo terá distribuição uniforme entre 0 e 10).



A segunda figura mostra que a magnitude de c é $g\,\text{sen}(\alpha)$.

As funções de densidade das variáveis são

$$\begin{aligned}
f_G(g) &= \frac{1}{10} \\
f_A(\alpha) &= \text{sen}(2\alpha)
\end{aligned}$$

Presumimos que G e A são independentes, portanto

$$f_{GA}(g, \alpha) = \left(\frac{1}{10}\right) (\text{sen}(2\alpha))$$

A magnitude do componente de força na direção do centro da esfera depende de variáveis já definidas no experimento, e pode ser também visto como uma variável aleatória:

$$C = G\sin(A).$$

A esperança de C é

$$\begin{aligned}
\mathbb{E}[C] &= \mathbb{E}[G\sin(A)] \\
&= \int_0^{\pi/2}\int_0^{10} f_{GA}(g,\alpha)\, g\sin(\alpha)\, dg d\alpha \\
&= \int_0^{\pi/2}\int_0^{10} \left(\frac{\text{sen}(2\alpha)}{10}\right) g\sin(\alpha)\, dg d\alpha \\
&= \frac{10}{3}.
\end{aligned}$$

●

Nem sempre o Teorema 5.6.1 é o único caminho para determinar a esperança de uma função de duas variáveis. O Exemplo 5.6.4 ilustra um caso em que o Teorema não é usado.

Exemplo 5.6.4 (componentes em um sistema – mínimo de duas exponenciais)**.** Dois componentes de um sistema tem tempo de falha dado por distribuições exponenciais com parâmetros λ e θ. A esperança do tempo até que o primeiro deles falhe é

$$\mathbb{E}[\min(X,Y)].$$

Seja Z a variável que representa o tempo em que o sistema falhará ($X = \min(X < Y)$). Então,

$$\begin{aligned}
\Pr[Z > t] &= \Pr[\min(X,Y) > t] \\
&= \Pr[X > t, Y > t] \\
&= \Pr[X > t]\Pr[Y > t] \\
&= e^{-\lambda t}e^{-\theta t} \\
&= e^{(-\lambda-\theta)t}.
\end{aligned}$$

A função de densidade acumulada de Z é

$$\begin{aligned}
F_Z(t) &= \Pr[Z \le t] \\
&= 1 - \Pr[Z > t] \\
&= 1 - e^{(-\lambda-\theta)t}.
\end{aligned}$$

Derivamos F_Z, para obter a função de densidade de Z:

$$\begin{aligned}
f_Z(t) &= \frac{d}{dt}F_Z(t) \\
&= \frac{d}{dt}\left(1 - e^{(-\lambda-\theta)t}\right) \\
&= (\lambda+\theta)e^{(-\lambda-\theta)t},
\end{aligned}$$

e f_Z é, portanto, uma função de densidade exponencial com parâmetro $\lambda + \theta$, de onde concluímos que

$$\begin{aligned}\mathbb{E}[Z] &= \mathbb{E}[\min(X,Y)] \\ &= \frac{1}{\lambda + \theta}.\end{aligned}$$

Isto claramente vale para quaisquer duas variáveis aleatórias exponenciais. ●

5.7 Covariância e correlação

A medida de quanto duas variáves são relacionadas linearmente é chamada de *covariância.*

Definição 5.7.1 (covariância)**.** Sejam X e Y duas variáveis aleatórias definidas a partir de um mesmo espaço amostral. A **covariância** entre X e Y é

$$\operatorname{cov}(X,Y) = \mathbb{E}\Big[(X - \mathbb{E}[X])\,(Y - \mathbb{E}[Y])\Big].$$ ◀

Teorema 5.7.2. Se X e Y são duas variáveis aleatórias definidas a partir de um mesmo espaço amostral, então

$$\operatorname{cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

Demonstração. Usando a linearidade da esperança,

$$\begin{aligned}\operatorname{cov}(XY) &= \mathbb{E}[(X - \mu_X)(Y - \mu_Y)] \\ &= \mathbb{E}[XY - \mu_Y X - \mu_X Y + \mu_X \mu_Y] \\ &= \mathbb{E}[XY] - \mathbb{E}[\mu_Y X] - \mathbb{E}[\mu_X Y] + \mathbb{E}[\mu_X \mu_X] \\ &= \mathbb{E}[XY] - \mathbb{E}[\mu_Y X] - \mathbb{E}[\mu_X Y] + \mu_X \mu_X \\ &= \mathbb{E}[XY] - \mu_Y \mathbb{E}[X] - \mu_X \mathbb{E}[Y] + \mu_X \mu_X \\ &= \mathbb{E}[XY] - \mu_Y \mu_X - \mu_X \mu_Y + \mu_X \mu_X \\ &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].\end{aligned}$$ ■

Exemplo 5.7.3 ($x, x^2, 2x$ – covariância)**.** Escolhemos um ponto x no intervalo $[0,1]$, e definimos as seguintes variávis aleatórias:

$$\begin{aligned}X &= x \\ A &= x^2 \\ B &= 2x\end{aligned}$$

Calcularemos a covariância de A com B.

$$\begin{aligned}
\mathbb{E}[X] &= \frac{1}{2} \\
\mathbb{E}[A] = \mathbb{E}[X^2] &= \int_0^1 f_X(x)x^2\,dx = \frac{1}{3} \\
\mathbb{E}[B] = \mathbb{E}[2X] &= 2\mathbb{E}[X] = 1 \\
\mathbb{E}[AB] = \mathbb{E}[2X^2] &= \int_0^1 f_X(x)2x^2\,dx = \frac{1}{6}.
\end{aligned}$$

A covariância é

$$\begin{aligned}
\operatorname{cov}(A,B) &= \frac{1}{2} - \left(\frac{1}{2}\right)\left(\frac{1}{3}\right) \\
&= \frac{1}{6}.
\end{aligned}$$

●

A covariância de uma variável aleatória com ela mesma é sua própria variância:

$$\begin{aligned}
\operatorname{cov}(X,X) &= \mathbb{E}[XX] - \mathbb{E}[X]\mathbb{E}[X] \\
&= \mathbb{E}[X^2] - \mathbb{E}[X]^2 \\
&= \operatorname{var}(X).
\end{aligned}$$

Exemplo 5.7.4 ($\operatorname{cov}(X,X) = \operatorname{var}(X)$)**.** A variável X no Exemplo 5.7.3 tem distribuição uniforme em $[0,1]$. Calculamos a covariância de X com ela mesma.

$$\begin{aligned}
\operatorname{cov}(X,X) &= \mathbb{E}[(X-\mathbb{E}[X])(X-\mathbb{E}[X])] \\
&= \mathbb{E}\left[\left(X-\frac{1}{2}\right)\left(X-\frac{1}{2}\right)\right] \\
&= \mathbb{E}\left[X^2 - 2X\frac{1}{2} + \frac{1}{4}\right] \\
&= \mathbb{E}[X^2 - X] + \frac{1}{4} \\
&= \int_0^1 x^2 - x\,dx + \frac{1}{4} \\
&= -\frac{1}{6} + \frac{1}{4} \\
&= \frac{1}{12},
\end{aligned}$$

a variância de X (a variância de variável uniforme no intervalo $[a,b]$ é $(a-b)^2/12$). ●

O próximo exemplo mostra que, apesar da variância representar a variação conjunta de duas variáveis, ela usualmente não tem valor para a intuição.

Exemplo 5.7.5 (Crateras lunares - covariância de ângulo e força)**.** Na Seção 4.7 tratamos de meteoritos atingindo a Lua. Eles chegam perfazendo um ângulo α com a superfície, e no Exemplo 5.6.3, idenificamos a componente c desta força na direção do centro, e calculamos usa esperança.

Agora, sendo A a variável aleatória que representa o ângulo, e C a que representa a força na direção do centro da Lua, obteremos a covariância das duas, que será

$$\text{cov}(A, V) = \mathbb{E}[AC] - \mathbb{E}[A]\mathbb{E}[C].$$

Já conhecemos $\mathbb{E}[A] = \pi/4$ e $\mathbb{E}[C] = 10/3$; precisamos calcular $\mathbb{E}[AC]$, que, pelo Teorema 5.6.1, é

$$\mathbb{E}[AC] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{AC}(a,c) ac \, dadc.$$

No entanto, não conhecemos f_{AC}. Podemos resolver este problema observando que $C = G\sin(A)$, portanto

$$\begin{aligned}
\mathbb{E}[AC] &= \mathbb{E}[AG\sin(A)] \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{AG}(a,g) ag\sin(a) \, dadg \\
&= \int_{0}^{10}\int_{0}^{\pi/2} f_{AG}(a,g) ag\sin(a) \, dadg \\
&= \int_{0}^{10}\int_{0}^{\pi/2} \frac{\sin(2a)}{10} ag\sin(a) \, dadg \\
&= \frac{15\pi - 20}{9}.
\end{aligned}$$

Assim,

$$\begin{aligned}
\text{cov}(A, C) &= \mathbb{E}[AC] - \mathbb{E}[A]\mathbb{E}[C] \\
&= \mathbb{E}[AG\sin(A)] - \mathbb{E}[A]\mathbb{E}[C] \\
&= \frac{15\pi - 20}{9} - \frac{\pi}{4}\frac{10}{3} \\
&= \frac{15\pi - 40}{18} \\
&\approx 0.3957.
\end{aligned}$$

Temos a covariância de A com C. Mas que significa uma covariância de aproximadamente 0.3957? Não temos como saber se isto é pouco ou muito. ●

A covariância, portanto, pode não ser de grande ajuda, porque sua interpretação depende da comparação com as variâncias das duas variáveis. Suponha que queiramos comparar quatro variáveis tais que

$$\begin{aligned}
\text{cov}(X, Y) &= 50, \\
\text{cov}(A, B) &= 500.
\end{aligned}$$

Mesmo conhecendo as variâncias de todas as variáveis, a comparação será, no mínimo, confusa.

A *correlação* entre duas variáveis é semelhante à covariância, exceto que é normalizada para resultar em um número entre -1 e $+1$.

Definição 5.7.6 (correlação)**.** Sejam X e Y duas variáveis aleatórias definidas a partir de um mesmo espaço amostral. A **correlação** entre X e Y é

$$\operatorname{corr}(X,Y) = \frac{\operatorname{cov}(X,Y)}{\sigma_X \sigma_Y},$$

também denotada $\rho_{X,Y}$. ◀

Teorema 5.7.7. Se X e Y são duas variáveis aleatórias definidas a partir de um mesmo espaço amostral, então

$$\operatorname{corr}(X,Y) \in [-1,+1].$$

Demonstração. Sejam

$$\begin{aligned} A &= X - \mu_X, \\ B &= Y - \mu_Y. \end{aligned}$$

Seja também $k \in \mathbb{R}$. Agora, a esperança do quadrado de uma variável aleatória deve necessariamente ser maior ou igual que zero, logo

$$\begin{aligned} \mathbb{E}\left[(A+kB)^2\right] &\geq 0 \\ \mathbb{E}\left[A^2 + 2kAB + k^2B^2\right] &\geq 0 \\ \mathbb{E}[A^2] + 2k\mathbb{E}[AB] + k^2\mathbb{E}[B^2] &\geq 0 && \text{(linearidade de } \mathbb{E}[.]) \end{aligned}$$

O lado esquerdo é um polinômio quadrático em k, maior ou igual que zero. Mas

$$\begin{aligned} \mathbb{E}[A^2] &= \mathbb{E}[X - \mu_X] = \sigma_X^2 \\ \mathbb{E}[B^2] &= \mathbb{E}[Y - \mu_Y] = \sigma_Y^2 \\ \mathbb{E}[AB] &= \mathbb{E}[(X-\mu_x)(Y-\mu_y)] = \operatorname{cov}(X,Y), \end{aligned}$$

então reescrevemos a desigualdade de segundo grau,

$$\sigma_Y^2 k^2 + 2\operatorname{cov}(X,Y)k + \sigma_x^2 \geq 0,$$

que é claramente $ak^2 + bk + c \geq 0$, com

$$\begin{aligned} a &= \sigma_y^2 \\ b &= 2\operatorname{cov}(X,Y) \\ c &= \sigma_x^2. \end{aligned}$$

A inequação determina que esta parábola fica sempre acima de zero, logo o discriminante deve ser menor ou igual a zero. Assim,

$$\begin{aligned}
b^2 - 4ac &\leq 0 \\
(2\text{cov}(X,Y))^2 - 4(\sigma_y^2)(\sigma_x^2) &\leq 0 \\
\text{cov}(X,Y)^2 - \sigma_y^2\sigma_x^2 &\leq 0 \\
\text{cov}(X,Y)^2 &\leq \sigma_y^2\sigma_x^2 \\
\frac{\text{cov}(X,Y)^2}{\sigma_y^2\sigma_x^2} &\leq 1 \\
\left|\frac{\text{cov}(X,Y)}{\sigma_y\sigma_x}\right| &\leq 1 \\
|\sigma_{X,Y}| &\leq 1. \qquad \blacksquare
\end{aligned}$$

O Exemplo 5.7.8 compara o uso de covariância e correlação.

Exemplo 5.7.8 (uma moeda – covariância e correlação)**.** Seja um experimento onde uma moeda honesta é jogada uma única vez. Definimos duas variáveis aleatórias relacionadas a este experimento.

$$X = \begin{cases} +1 & \text{cara} \\ 0 & \text{coroa} \end{cases}$$
$$Y = \begin{cases} 0 & \text{cara} \\ -1 & \text{coroa} \end{cases}$$

As esperanças das duas variáveis são

$$\mathbb{E}[X] = +\frac{1}{2},$$
$$\mathbb{E}[Y] = -\frac{1}{2}.$$

Calculamos a covariância:

$$
\begin{aligned}
\operatorname{cov}(X,Y) &= \mathbb{E}\Big[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])\Big] \\
&= \mathbb{E}\left[\left(X-\frac{1}{2}\right)\left(Y+\frac{1}{2}\right)\right] \\
&= \sum_{x,y} f(x,y)\Pr[X=x,Y=y] \\
&= \ \left(0-\frac{1}{2}\right)\left(0+\frac{1}{2}\right)[0] \\
&\quad + \left(1-\frac{1}{2}\right)\left(0+\frac{1}{2}\right)\left[\frac{1}{2}\right] \\
&\quad + \left(0-\frac{1}{2}\right)\left(-1+\frac{1}{2}\right)[0] \\
&\quad + \left(1-\frac{1}{2}\right)\left(-1+\frac{1}{2}\right)\left[\frac{1}{2}\right] \\
&= -\frac{1}{4}.
\end{aligned}
$$

A correlação é

$$
\begin{aligned}
\rho_{X,Y} &= \frac{\operatorname{cov}(X,Y)}{\rho_X\rho_Y} \\
&= \frac{-1/4}{(1/2)(-1/2)} \\
&= +1.
\end{aligned}
$$

●

No Exemplo 5.7.9 calculamos a covariância e a correlação para duas variáveis em função das probabilidades p e q (de sucesso e falha) associadas a experimentos de Bernoulli.

Exemplo 5.7.9 (duas moedas – covariância e correlação)**.** Os Exemplos 5.2.3 e 5.3.2 definiam um experimento onde duas moedas são jogdas; definimos para aquele experimento duas variáveis aleatórias, A, quandidade de caras, e O, a quantidade de coroas antes da primeira cara. A tabela com as distribuições marginais, apresentada no Exemplo 5.3.2, é reproduzida a seguir.

$$
\begin{array}{cc|ccc}
 & & & A & \\
 & & 0 & 1 & 2 \\
\hline
 & 0 & 0 & pq & p^2 & \dashrightarrow p \\
O & 1 & 0 & pq & 0 & \dashrightarrow pq \\
 & 2 & q^2 & 0 & 0 & \dashrightarrow q^2 \\
 & & \downarrow & \downarrow & \downarrow & \\
 & & q^2 & 2pq & p^2 &
\end{array}
$$

Determinaremos a covariância e a correlação das duas variáveis. Como $\text{cov}(A, O) = \mathbb{E}[AO] - \mathbb{E}[A]\mathbb{E}[O]$, calculamos $\mathbb{E}[AO]$:

$$
\begin{aligned}
\text{cov}(A, O) &= \mathbb{E}[AO] - \mathbb{E}[A]\mathbb{E}[O] \\
&= pq - 2p(q^2 + q) \\
&= pq - 2pq^2 - 2pq \\
&= -pq - 2pq^2 \\
&= p(q + 2q^2).
\end{aligned}
$$

Sabemos que $\sigma_A^2 = npq$ e $\sigma_O^2 = pq + 4q^2$, logo

$$
\begin{aligned}
\sigma_{AO} &= \frac{\text{cov}(A, O)}{\sigma_A \sigma_O} \\
&= \frac{-p(q + 2q^2)}{\sqrt{2pq(pq + 4q^2)}}.
\end{aligned}
$$

●

Exemplo 5.7.10 (Crateras lunares - correlação entre ângulo e força)**.** No Exemplo 5.7.5 calculamos a covariância de duas variáveis, uma (A) representando o ângulo de incidência de meteoritos e a outra (C) representando a magnitude da força na direção do centro da Lua. Observamos, ali, que o valor calculado não era muito útil, por não ter significado prático. Calcularemos agora a correlação entre A e C, que é

$$
\rho_{AC} = \frac{\text{cov}(A, C)}{\sigma_A \sigma_C}
$$

Já calculamos a variância de A na Seção 4.7, pnde obtivemos $\sigma_A^2 = (\pi^2 - 8)/16$, portanto já temos

$$
\sigma_A = \frac{\sqrt{\pi^2 - 8}}{4}.
$$

Nos falta calcular o desvio padrão de C.

$$
\begin{aligned}
\sigma_C &= \mathbb{E}\left[(C - \mu_C)^2\right] \\
&= \int_0^{10} (c - \mu_c)^2 f_C(c)\, dc,
\end{aligned}
$$

mas como não temos f_C, faremos como no Exemplo 5.7.5, usando $C = G\sin(A)$, e calculamos

$$\begin{aligned}
\sigma_C^2 &= \mathbb{E}\left[(C-\mu_C)^2\right] \\
&= \mathbb{E}\left[(G\sin(A)-\mu_C)^2\right] \\
&= \mathbb{E}\left[\left(G\sin(A)-\frac{10}{3}\right)^2\right] \\
&= \int_0^{10}\int_0^{\pi/2} f_{GA}(g,a)\left(g\sin(a)-\frac{10}{3}\right)^2 dadg \\
&= \int_0^{10}\int_0^{\pi/2} \frac{\sin(2a)}{10}\left(g\sin(a)-\frac{10}{3}\right)^2 dadg \\
&= \int_0^{10} \frac{9g^2-80g+200}{180}\, dg \\
&= \frac{50}{9} \\
&\approx 5.555,
\end{aligned}$$

e $\sigma_c = \sqrt{50/3} = 5\sqrt{2}/3 \approx 2.357$. Finalmente,

$$\begin{aligned}
\rho_{AC} &= \frac{\operatorname{cov}(A,C)}{\sigma_A\sigma_C} \\
&= \frac{\left(\dfrac{15\pi-40}{18}\right)}{\left(\dfrac{\sqrt{\pi^2-8}}{4}\right)\left(\dfrac{5\sqrt{2}}{3}\right)} \\
&= \frac{6\pi-16}{3\sqrt{2\pi^2-16}} \\
&\approx 0.491.
\end{aligned}$$

●

Teorema 5.7.11. Se X e Y são independentes, então $\operatorname{cov}(X,Y) = \rho_{X,Y} = 0$.

Demonstração. Sabemos que

$$\operatorname{cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y],$$

mas por X e Y serem independentes, $\mathbb{E}[XY] = \mathbb{E}[X][Y]$. ■

Exemplo 5.7.12 (variáveis exponenciais, independentes – covariância zero)**.** O Exemplo 5.4.6 trata de duas baterias com tempo de duração modelado por variáveis com distribuição exponencial, com parâmetros λ e θ. Ali, usamos o fato das variáveis serem independentes para determinar a função conjunta de densidade,

$$f_{XY}(x,y) = f_X(x)f_Y(y) = \frac{1}{\lambda\theta}e^{-\lambda x-\theta y}.$$

Sem tratar de valores específicos dos parâmetros daquele Exemplo, calculamos a covariância para duas variáveis independentes distribuidas exponencialmente.

$$\begin{aligned}
\operatorname{cov}(X,Y) &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \\
&= \int_o^\infty \int_0^\infty f_{XY}(x,y)xy\,dxdy - \int_0^\infty f_X(x)x\,dx \int_0^\infty f_Y(y)y\,dy \\
&= \int_o^\infty \int_0^\infty \frac{1}{\lambda\theta} e^{-\lambda x - \theta y}\,xy\,dxdy - \int_0^\infty \lambda e^{-\lambda x}\,x\,dx \int_0^\infty \theta e^{-\theta y}\,y\,dy \\
&= \frac{1}{\lambda\theta} - \left(\frac{1}{\lambda}\right)\left(\frac{1}{\theta}\right) \\
&= 0.
\end{aligned}$$

●

A recíproca do Teorema 5.7.11 *não* é verdadeira, como mostram os Exemplos 5.7.13 e 5.7.14: correlação zero não implica em independência!

Exemplo 5.7.13. Seja X uma variável com distribuição normal padrão: $X \sum \mathcal{N}(0,1)$, e seja $Y = X^2$. A esperança de XY é

$$\mathbb{E}[XY] = \mathbb{E}[X^3] = 0,$$

portanto a covariância entre X e Y é

$$\operatorname{cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = 0.$$

No entanto, X e Y *não* são independentes, porque Y foi definida em função de X. Apenas ocorre que não existe relação *linear* entre as duas variáveis. ●

Exemplo 5.7.14. No seguinte experimento, duas moedas diferentes são jogadas. O espaço amostral para o experimento é

$$\Omega = \{aa, ao, oa, oo\}.$$

Na tentativa de idealizar um jogo a partir do experimento, criamos duas variáveis aleatórias:

- X: quantidade de caras;
- Y: 1 se as duas moedas retornam resultados iguais, e 0 caso contrário.

A próxima tabela mostra os possíveis resultados do experimento, com os valores para os quais as variáveis os mapeiam.

resultado	X	Y
aa	2	1
ao	1	0
oa	1	0
oo	0	1

A seguir estão as funções de massa de probabilidade das variáveis.

x	$\Pr[X = x]$
0	1/4
1	1/2
2	1/4

y	$\Pr[Y = y]$
0	1/2
1	1/2

A função de probabilidade conjunta de X e Y é

$$\begin{aligned} f(0,1) &= 1/4, \\ f(1,0) &= 1/2, \\ f(2,1) &= 1/4, \end{aligned}$$

e zero para quaisquer outros argumentos. A tabela a seguir mostra os todos os valores de $f(x, y)$.

	0	1
0	0	1/4
1	1/2	0
2	0	1/4

Verificamos que as variáveis *não* são independentes:

$$\begin{aligned} \Pr[X \leq 0] &= 1/4 \\ \Pr[Y \leq 0] &= 1/2 \\ \Pr[X \leq 0, Y \leq 0] &= 0 \neq \Pr[X \leq 0] \Pr[Y \leq 0]. \end{aligned}$$

Precisaremos das esperança das duas variáveis, que são triviais de calcular.

$$\begin{aligned} \mathbb{E}[X] &= 1 \\ \mathbb{E}[Y] &= \frac{1}{2} \end{aligned}$$

Finalmente, calculamos a covariância.

$$\begin{aligned}
\operatorname{cov}(X,Y) &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \\
&= \mathbb{E}[XY] - (1)(1/2) \\
&= \left[\sum_{x,y} xy \Pr[X=x, Y=y]\right] - \frac{1}{2} \\
&= (2)(1)\frac{1}{4} && \text{(aa)} \\
&\quad + (1)(0)\frac{1}{2} && \text{(ao,oa)} \\
&\quad + (0)(1)\frac{1}{4} && \text{(oo)} \\
&\quad - \frac{1}{2} \\
&= \frac{2}{4} - \frac{1}{2} \\
&= 0.
\end{aligned}$$

Como a covariância é zero, a correlação também é – mesmo as variáveis sendo dependentes. ●

5.8 Soma de variáveis aleatórias

Se X e Y são duas variáveis aleatórias independentes, então sua soma, $Z = X + Y$, também é (X e Y são funções de Ω em $\mathbb{R}$; $X + Y$ também). A função de distribuição f_Z, de Z, é dada pela *convolução* de f_x e f_y.

Teorema 5.8.1. Se X e Y são variáveis aleatórias independentes, e $Z = X+Y$, então

$$f_Z(z) = \begin{cases} \sum_x f_X(x) f_Y(z-x) & \text{se } X, Y \text{ são discretas} \\ \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\,dx & \text{se } X, Y \text{ são contínuas.} \end{cases}$$

Demonstração. No caso discreto, para cada z, somamos as probabilidades de cada possibilidade de x, y somando z:

$$\begin{aligned}
\sum_{x+y=z} \Pr[X=x]\Pr[Y=y] &= \sum_x \Pr[X=x]\Pr[Y=z-x] \\
&= \sum_x f_X(x) f_Y(z-x).
\end{aligned}$$

Para o caso contínuo, considere a distribuição conjunta (X, Y). Como as duas variáveis são independentes, então a função de densidade da distribuição conjunta é $f_{XY}(x,y) = f_X(x)f_Y(y)$.

$$\begin{aligned}
F_Z(z) &= \Pr[Z < z] \\
&= \Pr[X + Y < z] \\
&= \iint_{x+y<z} f_X(x)f_Y(y)\, dydx \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{z-x} f_X(x)f_Y(y)\, dydx \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{z} f_X(x)f_Y(w-x)\, dwdx \\
&= \int_{-\infty}^{z} \int_{-\infty}^{\infty} f_X(x)f_Y(w-x)\, dxdw.
\end{aligned}$$

Logo,

$$\begin{aligned}
f_Z(z) &= \frac{d}{dz}F_Z(z) \\
&= \int_{-\infty}^{\infty} f_X(x)f_Y(z-x)\, dx. \qquad \blacksquare
\end{aligned}$$

O Teorema 5.8.1 pode ser aplicado a somas de várias distribuições. A seguir abordamos algumas dessas somas, como corolários desse Teorema.

Corolário 5.8.1 (soma de Poissons – distribuição)**.** Se X e Y são variáveis independentes com distribuição de Poisson e parâmetros λ e θ, então $Z = X+Y$ tem distribuição de Poisson com parâmetro $\lambda + \theta$.

Demonstração. As funções de massa de X e Y são

$$\begin{aligned}
f_X(x) &= \frac{\lambda^x}{x!}e^{-\lambda} \\
f_Y(y) &= \frac{\theta^x}{x!}e^{-\theta}
\end{aligned}$$

A função de masa de Z é

$$\begin{aligned}
f_Z(z) &= \sum_x f_X(x) f_Y(z-x) \\
&= \sum_x \frac{\lambda^x}{x!} e^{-\lambda} \frac{\theta^{z-x}}{(z-x)!} e^{-\theta} \\
&= \sum_x \frac{\lambda^x}{x!} \frac{\theta^{z-x}}{(z-x)!} e^{-(\lambda+\theta)} \\
&= e^{-(\lambda+\theta)} \frac{1}{z!} \sum_x \frac{z! \lambda^x \theta^{z-x}}{x!(z-x)!} \\
&= e^{-(\lambda+\theta)} \frac{1}{z!} \sum_x \binom{z}{x} \lambda^x \theta^{z-x} \\
&= \frac{(\lambda+\theta)^z}{z!} e^{-(\lambda+\theta)} && \text{(teorema binomial)}
\end{aligned}$$

Assim, a distribuiçãode $X + Y$ é Poisson, com parâmetro $\lambda + \theta$. ■

Exemplo 5.8.2. A quantidade de infecções por uma bactéria que um hospital atende segue distribuição de Poisson (em média, 2 por ano). Para uma virose com sintomas parecidos, a distribuição também é Poisson (a média é 3 por ano). A quantidade total de atendimentos para as duas patologias segue também distribuição de Poisson, e a média é de 5 casos por ano.

A conclusão pode parecer óbvia se olharmos apenas para a média, mas determinamos a *distribuição* da soma – inclusive sua variância, que é também igual a 5. ●

Corolário 5.8.2 (soma de exponenciais – distribuição)**.** Se X e Y são variáveis com distribuição exponencial e parâmetros λ e θ, então $Z = X + Y$ tem função de densidade

$$f_Z(z) = \frac{\theta\lambda}{\theta - \lambda} \left(e^{-\lambda z} - e^{-\theta z} \right).$$

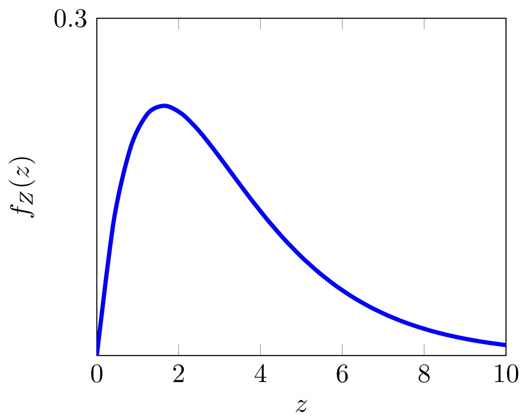
Demonstração. Pelo Teorema 5.8.1,

$$\begin{aligned}
f_Z(z) &= \int_{-\infty}^{z} f_X(x) f_Y(z-x)\, dx \\
&= \int_0^z \left(\lambda e^{-\lambda x}\right) \left(\theta e^{-\theta(z-x)}\right) dx \\
&= \lambda\theta e^{-\theta z} \int_0^z \left(e^{-\lambda x}\right) \left(e^{-\theta - x}\right) dx \\
&= \lambda\theta e^{-\theta z} \int_0^z \left(e^{\theta - \lambda x}\right) dx \\
&= \frac{\theta\lambda}{\theta - \lambda} \left(e^{-\lambda z} - e^{-\theta z} \right).
\end{aligned}$$

É simples confirmar que esta é uma função de densidade – basta calcular

$$\int_0^\infty \frac{\theta\lambda}{\theta-\lambda}\left(e^{-\lambda z}-e^{-\theta z}\right)=1. \qquad \blacksquare$$

O gráfico a seguir ilustra $f_Z(z)$.

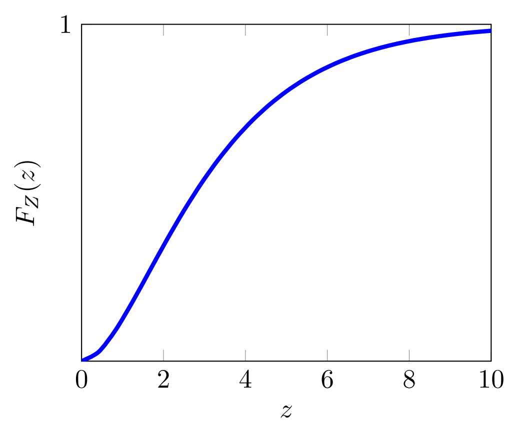


A função acumulada de densidade é

$$\begin{aligned} F_Z(z) &= \int_0^z \frac{\theta\lambda}{\theta-\lambda}e^{-\lambda w}-e^{-\theta w}\,dw \\ &= 1+\frac{\lambda}{\theta-\lambda}e^{-\theta z}-\frac{\theta}{\theta-\lambda}e^{-\lambda z}. \end{aligned}$$

Claramente, quando $z \to \infty$, $F_Z(z) \to 1$.

A próxima figura ilustra $F_Z(z)$.



Exemplo 5.8.3 (soma de exponenciais)**.** Um dispositivo A está sendo usado em uma máquina, e outro dispositivo, B, será usado no lugar de A quando este

falhar. Os dois, embora compatíveis, tem taxas exponenciais de falha diferentes (1.5 ano e 2 anos, respectivamente). Denotamos por X e Y as variáveis aleatórias dando os tempos de falha de A e B:

$$\Pr[X < x] = 1.5e^{-1.5x}$$
$$\Pr[Y < y] = 2e^{-2y}$$

Presumindo que B manterá suas características até o momento de entrar em funcionamento, o tempo total até a máquina parar de funcionar é dado pela variável aleatória $Z = X + Y$, e

$$\begin{aligned}\Pr[Z < z] &= 1 + \frac{\lambda}{\theta - \lambda} e^{-\theta z} - \frac{\theta}{\theta - \lambda} e^{-\lambda z} \\ &= 1 + \frac{1.5}{2 - 1.5} e^{-2z} - \frac{2}{2 - 1.5} e^{-1.5z} \\ &= 1 + 3e^{-2z} - 4e^{-1.5z}.\end{aligned}$$

●

Corolário 5.8.3 (soma de normais – distribuição)**.** Se X e Y são variáveis aleatórias independentes, sendo que $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ e $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2,)$, então a soma $Z = X + Y$ também tem distribuição normal, e

$$Z \sim \mathcal{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2).$$

Observando o Corolário 5.8.3, fica claro que o desvio padrão da soma de normais *não* é a soma dos desvios das variáveis somadas:

$$\sigma_Z = \sqrt{\sigma_X^2 + \sigma_Y^2}.$$

Exemplo 5.8.4 (soma de normais)**.** Cilindros de combustão usados em motores são produzidos por duas máquinas diferentes, com volume obedecendo uma distribuição normal. Na máquina A, a média do volume é $0.5\,l$ e a variância 10^{-2}. Na máquina B, o volume médio é $0.52\,l$, com variância 10^{-4}.

Motores são montados usando um cilindro de cada máquina. A média do volume total de combuistão para os motores é a soma das médias, 1.02, e a variancia é a soma das variancias. O desvio padrão do volume é

$$\begin{aligned}\sigma_C &= \sqrt{\sigma_A + \sigma_B} \\ &= \sqrt{10^{-2} + 10^{-4}} \\ &\approx 0.1005.\end{aligned}$$

Calculamos a seguir a probabilidade do volume total ser menor que $0.99\,l$.

A distribuição total dos dois cilindros é a soma dos volumes, e portanto segue a distribuição da soma das duas variáveis aleatórias – que são normais.

$$\begin{aligned}A &\sim \mathcal{N}(0.5, 10^{-2}) \\ B &\sim \mathcal{N}(0.52, 10^{-4}) \\ A + B \ = \ C &\sim \mathcal{N}(1.02, 10^{-2} + 10^{-4}).\end{aligned}$$

Queremos $Pr[C < 0.99]$.

$$\begin{aligned}
\Pr[C < 0.99] &= \Pr\left[\frac{C - 1.02}{10^{-2} + 10^{-4}} < \frac{0.99 - 1.02}{10^{-2} + 10^{-4}}\right] \\
&= \Pr\left[Z < \frac{-0.03}{0.0101}\right] \\
&= \Phi(-2.97) \\
&= 1 - \Phi(2.97) \\
&\approx 1 - 0.9985 \\
&\approx 0.0015.
\end{aligned}$$

●

Teorema 5.8.5. Para quaisquer duas variáveis aleatórias X, Y com variância definida,

$$\text{var}(X + Y) = \text{var}(X) + \text{var}(Y) + 2\text{cov}(X, Y).$$

Demonstração.

$$\begin{aligned}
\text{var}(X + Y) &= \mathbb{E}[(X + Y)^2] - \mathbb{E}[X + Y]^2 \\
&= \mathbb{E}[X^2 + 2XY + Y^2] - \mathbb{E}[X + Y]^2 \\
&= \mathbb{E}[X^2] + \mathbb{E}[2XY] + \mathbb{E}[Y^2] - \mathbb{E}[X + Y]^2 \\
&= \mathbb{E}[X^2] + \mathbb{E}[2XY] + \mathbb{E}[Y^2] - (\mathbb{E}[X] + \mathbb{E}[Y])^2 \\
&= \mathbb{E}[X^2] + \mathbb{E}[2XY] + \mathbb{E}[Y^2] - (\mathbb{E}[X]^2 + 2\mathbb{E}[X]\mathbb{E}[Y] + \mathbb{E}[Y]^2) \\
&= \mathbb{E}[X^2] - \mathbb{E}[X]^2 + \mathbb{E}[Y^2] - \mathbb{E}[Y]^2 + \mathbb{E}[2XY] - 2\mathbb{E}[X]\mathbb{E}[Y] \\
&= \Big(\mathbb{E}[X^2] - \mathbb{E}[X]^2\Big) + \Big(\mathbb{E}[Y^2] - \mathbb{E}[Y]^2\Big) + \Big(2\mathbb{E}[XY] - 2\mathbb{E}[X]\mathbb{E}[Y]\Big) \\
&= \text{var}(X) + \text{var}(Y) + 2\text{cov}(X, Y).
\end{aligned}$$

■

Exemplo 5.8.6. No Exemplo 5.8.3, duas variáveis exponenciais independentes são somadas. A variância da soma das duas é

$$\begin{aligned}
\text{var}(X + Y) &= \text{var}(X) + \text{var}(Y) + 2\text{cov}(X, Y) \\
&= \frac{1}{\lambda^2} + \frac{1}{\theta^2} + 2\text{cov}(X, Y) \\
&= \frac{1}{\lambda^2} + \frac{1}{\theta^2} && (X,\ Y \text{ independentes}) \\
&= \frac{\lambda^2 + \theta^2}{(\lambda\theta)^2}.
\end{aligned}$$

●

Exemplo 5.8.7. Sorteamos um ângulo α em $[0, \pi/2]$ e definimos as variáveis aleatórias

$$\begin{aligned}
S &= \text{sen}(\alpha)^2 \\
C &= \cos(\alpha)^2
\end{aligned}$$

Para S, usamos a técnica já vista no Exemplo 4.1.7, obtendo

$$\begin{aligned} F_S(x) &= \frac{2\,\mathrm{arcsen}(\sqrt{x})}{\pi} \\ f_S(x) &= \frac{1}{\pi\sqrt{x}\sqrt{1-x}}. \end{aligned}$$

Calculamos a esperança e a variância de S:

$$\begin{aligned} \mathbb{E}[S] &= \int_0^1 x \frac{1}{\pi\sqrt{x}\sqrt{1-x}}\, d\alpha \\ &= \frac{1}{2} \end{aligned}$$

$$\begin{aligned} \mathrm{var}(S) &= \mathbb{E}[S^2] - \mathbb{E}[S]^2 \\ &= \frac{3}{8} - \frac{1}{4} \\ &= \frac{1}{8} \end{aligned}$$

Para C, temos que observar que a função $\cos(x)$ é decrescente no intervalo que escolhemos. Assim,

$$\begin{aligned} F_C(x) &= 1 - \frac{2\arccos(\sqrt{x})}{\pi} \\ f_C(x) &= \frac{1}{\pi\sqrt{x}\sqrt{1-x}} \end{aligned}$$

Como $f_C(x) = f_S(x)$, então as epseranças e variâncias de S e C são idênticas.

$$\begin{aligned} \mathbb{E}[C] &= \frac{1}{2} \\ \mathrm{var}(C) &= \frac{1}{8} \end{aligned}$$

Claramente, $U = \mathrm{sen}(\alpha)^2 + \cos(\alpha)^2 = 1$ é constante, e portanto tem variância zero. Agora podemos usar o Teorema 5.8.5 para obter a covariância de S e C facilmente:

$$\begin{aligned} \mathrm{var}(S+C) &= \mathrm{var}(S) + \mathrm{var}(C) + 2\mathrm{cov}(S,C) \\ 0 &= \frac{1}{8} + \frac{1}{8} + 2\mathrm{cov}(S,C) \\ -\frac{1}{4} &= 2\mathrm{cov}(S,C) \\ \mathrm{cov}(S,C) &= -\frac{1}{8}. \end{aligned}$$

Poderíamos, claro, ter calculado a covariância usando $\text{cov}(S, C) = \mathbb{E}[SC] - \mathbb{E}[S]\mathbb{E}[C]$, como no desenvolvimento a seguir.

$$\begin{aligned}
\mathbb{E}[SC] &= \mathbb{E}[\text{sen}^2(\alpha)\cos^2(\alpha)] \\
&= \mathbb{E}[\text{sen}^2(\alpha)(1 - \text{sen}^2(\alpha))] \\
&= \mathbb{E}[S(1 - S)] \\
&= \int_0^1 (x - x^2) f_S(x)\, dx \\
&= \int_0^1 (x - x^2) \frac{1}{\pi\sqrt{x}\sqrt{1-x}}\, dx \\
&= \frac{1}{8}.
\end{aligned}$$

$$\begin{aligned}
\text{cov}(S, C) &= \mathbb{E}[SC] - \mathbb{E}[S]\mathbb{E}[C] \\
&= \frac{1}{8} - \left(\frac{1}{2}\right)\left(\frac{1}{2}\right) \\
&= -\frac{1}{8}.
\end{aligned}$$

O coeficiente de correlação é

$$\begin{aligned}
\rho_{SC} &= \frac{\text{cov}(S, C)}{\sigma_S \sigma_C} \\
&= \frac{-1/8}{(1/\sqrt{8})(1/\sqrt{8})} && (\sigma_A = \sqrt{\sigma_A^2}) \\
&= -\left(\frac{1}{8}\right)\left(\frac{8}{1}\right) \\
&= -1,
\end{aligned}$$

o que também não deve surpreender, já que $S = 1 - C$. O coeficiente de correlação captura informação, neste exemplo, de forma mais cristalina que a covariância. É interessante também que as duas variáveis tenham a mesma função de densidade, tendo portanto as mesmas média e variância, mas ainda assim sejam correlacionadas negativamente. ●

Teorema 5.8.8. Para quaisquer variáveis aleatórias A, X, Y com variância definida, e qualquer $w \in \mathbb{R}$,

$$\begin{aligned}
\text{cov}(wX, Y) &= \text{cov}(X, wY) = w\text{cov}(X, Y) \\
\text{cov}(A + X, Y) &= \text{cov}(A, Y) + \text{cov}(X, Y) \\
\text{cov}(X, A + Y) &= \text{cov}(X, A) + \text{cov}(X, Y)
\end{aligned}$$

Exemplo 5.8.9 (Crateras Lunares – linearidade da covariância)**.** Quando discutimos crateras lunares, presumimos que a força com que as partículas chegam é distribuída uniformemente no intervalo $[0, 10]$. Definimos as variáveis aleatórias G, uniforme, que representa a força com que a partícula chega na superfície; A, o ângulo da trajetória da partícula com a superfície de Lua; e C, a componente da força na direção do centro da Lua. Tudo isso vale para partículas atingindo qualquer esfera, não somente a Lua. Suponha que o mesmo fenômeno acontece em uma esfera qualquer, também de raio um, para manter a discussão simples, e que a força é três vezes maior que no outro exemplo (uniforme em $[0, 30]$). Para determinar a covariância entre o ângulo A e a força dirigida ao centro, C, não precisamos refazer todas as contas que já fizemos. Como $C = G\operatorname{sen}(A)$, então quando multiplicamos G por 3, temos

$$3G\operatorname{sen}(A) = 3C.$$

Portanto,

$$\begin{aligned}
\operatorname{cov}(3C, A) &= 3\operatorname{cov}(C, A) \\
&= 3\left(\frac{15\pi - 40}{18}\right) \\
&= \frac{15\pi - 40}{6} \\
&\approx 1.187.
\end{aligned}$$

●

Exemplo 5.8.10 (Crateras Lunares – linearidade da covariância (cont.))**.** Retomando o Exemplo 5.8.9, suponha que a força não foi multiplicada por três, mas que a ela adicionamos uma constante 5: $G' = G + 5$. A força, portanto, é uniformemente distribuída em $[5, 15]$. Isto nos dará uma nova variável $C' = (G + 5)\operatorname{sen}(A) = G\operatorname{sen}(A) + 5\operatorname{sen}(A)$. A covariância entre C' e A é

$$\begin{aligned}
\operatorname{cov}(C', A) &= \operatorname{cov}(G\operatorname{sen}(A) + 5\operatorname{sen}(A), A) \\
&= \operatorname{cov}(G\operatorname{sen}(A), A) + \operatorname{cov}(5\operatorname{sen}(A), A). \qquad (5.4)
\end{aligned}$$

Já temos a primeira parcela, $\operatorname{cov}(G\operatorname{sen}(A), A)$, que calculamos no Exercício5.7.5:

$$\operatorname{cov}(C, A) = \frac{15\pi - 40}{18}.$$

Falta calcularmos a covariância de um ângulo em $[0, \pi/2]$ com seu seno multiplicado por 5. Usando $f_A(a) = \operatorname{sen}(2a)$ e o Teorema 4.2.5, que determina que

$\mathbb{E}[g(X)] = \int_{\infty}^{\infty} f_X(x) g(x)\, dx$,

$$\begin{aligned}
\mathrm{cov}(5\,\mathrm{sen}(A), A) &= 5\mathrm{cov}(\mathrm{sen}(A), A) \\
&= 5\left(\mathbb{E}[A\,\mathrm{sen}(A)] - \mathbb{E}[A]\mathbb{E}[\mathrm{sen}(A)]\right) \\
&= 5\left[\int_0^{\pi/2} a\,\mathrm{sen}(a)\ \mathrm{sen}(2a)\, da - \left(\frac{\pi}{4}\right)\left(\int_0^{\pi/2} \mathrm{sen}(a)\ \mathrm{sen}(2a)\, da\right)\right] \\
&\qquad\qquad (\mathbb{E}[A] = \pi/4) \\
&= 5\left[(\pi - 2) - \left(\frac{2\pi}{12}\right)\right] \\
&= \frac{5\pi}{6} - \frac{20}{9} \\
&\approx 0.395.
\end{aligned}$$

Substituímos em 5.4,

$$\begin{aligned}
\mathrm{cov}(C', A) &= \mathrm{cov}(G\,\mathrm{sen}(A), A) + \mathrm{cov}(5\,\mathrm{sen}(A), A) \\
&= \frac{15\pi - 40}{18} + \frac{5\pi}{6} - \frac{20}{9} \\
&\approx 0.791.
\end{aligned}$$

●

Exercícios

Ex. 108 — Uma moeda honesta e um dado viciado são jogados. Para o dado,

$$\begin{aligned}
&\mathrm{Pr}(1) = 1/3 \\
\mathrm{Pr}(2) = \mathrm{Pr}(3) = \mathrm{Pr}(4) = \mathrm{Pr}(r5) = \mathrm{Pr}(6) &= 2/15
\end{aligned}$$

1. Obtenha a distribuição conjunta das variáveis
   $X = 0$ para cara, 1 para coroa
   $Y =$ número da face do dado voltada para cima.
2. Obtenha as distribuições marginais de X e Y.
3. Verifque se X e Y são independentes.
4. Calcule $\mathrm{Pr}[X = 1, Y = 3]$, $\mathrm{Pr}[X = 0, Y < 5]$.

Ex. 109 — Em uma urna há

- 3 bolas vermelhas,
- 4 bolas brancas,
- 5 bolas verdes.

1. Se sortearmos 3 bolas, quais são as probabilidades conjuntas para as possíveis quantidades de boals vermelhas e brancas?
2. Escolha uma quantidade possível para bolas brancas, ou para bolas pretas, e verifique seu valor de duas maneiras diferentes: a partir da distribuição conjunta e da probabilidade marginal; e diretamente.

Ex. 110 — Determine a tal que $f(x,y)$, definida a seguir, seja função de densidade conjunta de probabilidade.

$$f(x,y) = \begin{cases} xy & x,y \in [0,a] \\ 0 & x,y \notin [0,a] \end{cases}$$

Ex. 111 — Em que intervalos a função

$$f_{AB}(a,b) = a\log(b)$$

é densidade conjunta de probabilidade?

Ex. 112 — Seja

$$f(x,y) = \begin{cases} k & x,y \in [-1,1] \\ 0 & x,y \notin [-1,1] \end{cases}$$

a densidade conjunta de duas variáveis, X e Y.

1. Determine k.
2. Determine $F_x(x)$.
3. Encontre $\Pr[X \leq 1/2]$.
4. Encontre $\Pr[X \leq 1/2, Y \geq 1/4]$.
5. Encontre $\Pr[X \geq 2Y]$.

Ex. 113 — A função de probabilidade conjunta de duas variáveis aleatórias X e Y é

$$f_{XY} = c\frac{x}{y+1}.$$

1. Qual é o valor de c?
2. X e Y são independentes?
3. Determine as distribuições marginais.
4. Determine $\mathbb{E}[X]$, $\mathbb{E}[Y]$, ρ_x^2.
5. Determine ρ_y^2.
6. Determine ρ_{XY}.

Ex. 114 — As variáveis X e Y tem distribuição exponencial, com funções de densidade $f(x) = \lambda e^{-\lambda x}$ e $g(y) = \theta e^{-\theta y}$.

1. Determine a função de probabilidade conjunta de X e Y
2. Determine a função de probabilidade marginal que dá $\Pr[X < x]$.

Ex. 115 — sejam

$$f_{AB}(a,b) = \begin{cases} ce^{-3a-b} & a,b>0 \\ 0 & \text{caso contrário} \end{cases}$$
$$g_{XY}(x,y) = \begin{cases} kxy & 0<x<y<e \\ 0 & \text{caso contrário} \end{cases}$$

1. Determine c e k.
2. Determine se A e B, e se X e Y são independentes.

Ex. 116 — Sejam X e Y duas variáveis aleatórias exponenciais com parâmetro $\lambda = 1$:

$$f_X(x) = e^{-x}$$
$$f_Y(y) = e^{-y}$$

Seja $Z = X/Y$. Determine a função de densidade de Z, $f_Z(z)$.

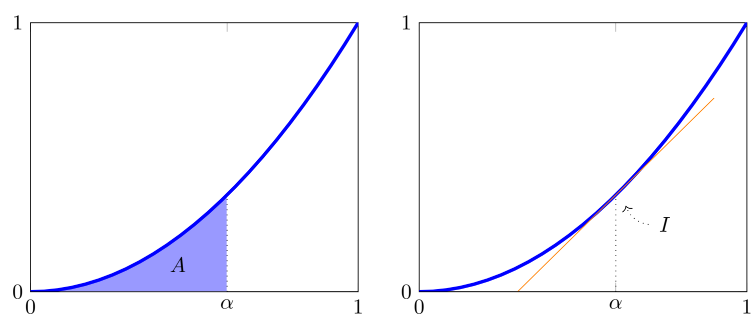
Ex. 117 — No exeperimento do Exemplo 5.4.4 selecionamos um ponto dentro da circunferência $\{(x,y) \ : \ x^2+y^2 < 1\}$. Ali, verificamos que as variáveis aleatórias dando as duas coordenadas cartesianas do ponto são dependentes. Agora defina as variáveis aleatórias Θ e R, dando as coordenadas polares do ponto. Determine $f_{\Theta R}$, f_Θ, f_R, e verifique se as duas são dependentes.

Ex. 118 — Retomamos o exercício em que selecionamos equiprovavelmente um ponto α no intervalo $[0,1]$.

a) Definimos duas variáveis aleatórias:

   $-A$ = área da curva abaixo de x^2, entre 0 e α;

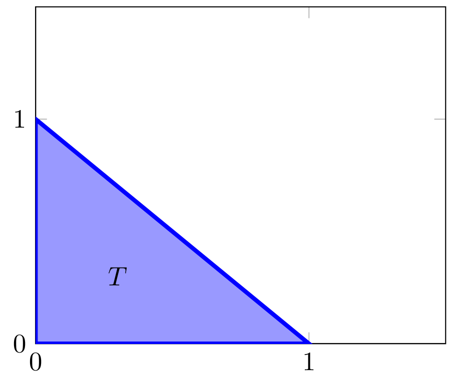
   $-I$ = derivada de x^2 no ponto α.



   Determine $\text{cov}(A,I)$ e $\rho_{A,I}$.

b) Repita o exercício, trocando x^2 por $\text{sen}(x)$, e trocando também o intervalo $[0,1]$ por $[0,\pi]$.

Ex. 119 — Seja T o triângulo delimitado pelos eixos das abscissas e ordenadas, e também pela reta $x + y = 1$.



Selecionamos um ponto em T, usando a função de densidade

$$f(x,y) = \begin{cases} zxy^2 & \text{se } (x,y) \in T \\ 0 & \text{em outros casos.} \end{cases}$$

1. Determine z.
2. Determine $\Pr[X < Y]$.

Ex. 120 — Dois eventos tem duração modelada por distribuições exponenciais: um com parâmetro λ e outro com parâmetro θ. Qual é a probabilidade do primeiro evento ocorrer antes do segundo?

Ex. 121 — As variáveis A, B e C tem função de densidade conjunta dada por

$$f_{ABC}(a,b,c) = \begin{cases} 3ae^{-(a+b+3c)} & a,b,c > 0 \\ 0 & \text{em outros casos} \end{cases}$$

Determine $\Pr[A < B < C]$.

Ex. 122 — No experimento da agulha de Buffon (Exemplo 5.4.9), aparentemente nada muda se trocarmos $\cos\theta$ por $\operatorname{sen}\theta$. Explique.

Ex. 123 — Refaça o Exemplo 5.4.9 para agulhas mais compridas que as faixas ($\ell > d$).

Ex. 124 — Sejam A, B e C variáveis aleatórias, com

$$\begin{aligned} f_A(a) &= 2e^{-2a} \\ F_{AB}(b) &= e^{-2a}\cos(b) \\ C &= AB. \end{aligned}$$

Determine f_B, $\mathbb{E}[C]$, $\operatorname{cov}(B,C)$, e ρ_{BC}.

Ex. 125 — Suponha que A e B são variáveis aleatórias com distribuição binomial, com a mesma probabilidade de sucesso, mas com números de repetições diferentes:

$$\begin{aligned} A &\sim \text{binomial}(n,p) \\ B &\sim \text{binomial}(m,p). \end{aligned}$$

Prove que a distribuição de $A+B$ é binomial, e determine seue parâmetros.

Ex. 126 — Um experimento de Bernoulli é repetido r vezes, com probabilidade de sucesso igual a p. Seja A uma variável aleatória dando a quantidade de sucessos nas primeiras n rodadas (presuma $n \leq r$), e B uma variável dando o total de sucessos. Calcule $\mathbb{E}[A|B=b]$.

Ex. 127 — Se jogarmos uma moeda honesta 10 vezes e um dado honesto 5 vezes, qual é a probabilidade de obtermos k sucessos, onde um sucesso é uma cara como resultado da moeda ou um seis como resultado do dado?

Ex. 128 — A tela de um dispositivo é composta por duas camadas de filme transparente. As duas camadas são produzidas por processos diferentes: o primeiro produz filme com espessura média $0.5mm$, e desvio padrão $0.05mm$; o segundo processo produz uma camada com espessura média de $1mm$ e desvio padrão $0.1mm$. Para que funcione apropriadamente, a primeira camada (mais fina) deve ter espessura entre $0.45mm$ e $0.55mm$; para que caiba no dispositivo, as duas camadas juntas devem ter espessura máxima de $1.57mm$. Os pares de camada (inferior e superior) para telas são tomados aleatoriamente para montagem (se um par não é útil, ele é imediatamente descartado).
Determine a probabilidade de que um par seja útil.

Ex. 129 — Determine $\text{cov}(X,k)$, onde k é constante.

Ex. 130 — Verifique que $\text{cov}(X,Y) = \text{cov}(Y,X)$.

Ex. 131 — Reveja o Exemplo 5.4.5 e calcule a correlação entre N e T.

Ex. 132 — Reveja o Exercício 77, na página 126.

1. Se a for escolhido com distribuição uniforme entre 0 e 10, e b uniforme entre 0 e 5, qual é a correlação entre b e a variável indicadora I?
2. E se a e b tiverem distribuições exponenciais?

Ex. 133 — Reveja os Exemplos 5.7.5 e 5.7.10 e calcule a covariância e a correlação entre a força no momento do impacto do meteorito sobre a Lua na direção da trajetória (G) e a resultante na direção do centro da Lua (C).

Ex. 134 — Nos Exemplos 5.8.9 e 5.8.10 calculamos covariâncias. Calcule os coeficientes de correlação.

Ex. 135 — Sejam X e Y duas variáveis aleatórias com vraiâncias finitas. Determine $\text{var}(X - Y)$. Se X e Y são independentes, qual a diferença entre $\text{var}(X + Y)$ e $\text{var}(X - Y)$?

Capítulo 6

Algumas Desigualdades e o Teorema Central do Limite

Este Capítulo trata das desigualdades de Markov e de Chebychev, que permitem estabelecer valores mínimos e máximos para variáveis aleatórias, sem depender de distribuição. A primeira dessas desigualdades depende apenas da média, e a segunda usa também a variância. Em seguida, são examinados a Lei Fraca dos Grandes Números e o Teorema Central do Limite, afirmações sobre a média e distribuição de variáveis aleatórias que tenham a mesma distribuição.

6.1 Desigualdade de Markov

A desigualdade de Markov permite obter um valor máximo para $\Pr[X = a]$, a partir somente da esperança de X.

Teorema 6.1.1 (desigualdade de Markov)**.** Se $X > 0$, então para todo $a > 0$,

$$\Pr[X \geq a] \leq \frac{\mathbb{E}[X]}{a}.$$

Demonstração. Defina uma variável aleatória I, tal que

$$I = \begin{cases} 1 & X \geq a \\ 0 & X < a. \end{cases}$$

Observamos que

$$\begin{aligned} \mathbb{E}[I] &= (1)\Pr[X \geq a] + (0)\Pr[X < a] \\ &= \Pr[X \geq a]. \end{aligned} \tag{6.1}$$

Como $X > 0$,

$$\begin{aligned} X \geq 1a &\Rightarrow X \geq a \\ X \geq 0a &\Rightarrow X \geq 0, \end{aligned}$$

o que significa que $aI \leq X$. Então,

$$\begin{aligned} aI &\leq X \\ \mathbb{E}[aI] &\leq \mathbb{E}[X] \\ a\mathbb{E}[I] &\leq \mathbb{E}[X] \\ a\Pr[X \geq a] &\leq \mathbb{E}[X] && \text{(pela equação 6.1)} \\ \Pr[X \geq a] &\leq \frac{\mathbb{E}[X]}{a}. && \blacksquare \end{aligned}$$

Exemplo 6.1.2. A média da pressão sistólica em homens adultos em uma determinada região é de $125mmHg$. Um indivíduo é considerado hipertenso se sua pressão sistólica está acima de $140mmHg$.

Se usarmos a desigualdade de Markov, podemos deduzir que

$$\Pr[P > 140] \leq \frac{P}{125}140 = \frac{25}{28} \approx 0.8928,$$

ou seja, a probabilidade de um indivíduo qualquer ser hipertenso é menor que 0.8928. ●

Exemplo 6.1.3 (desigualdade de Boole (demonstração usando desigualdade de Markov))**.** A desigualdade de Boole (Corolário 1.2.1) determina que, se $A_1, A_2, \ldots, A_n$ são eventos em um espaço amostral, então $\Pr\left(\bigcup_j A_j\right) \leq \sum_j \Pr(A_j)$.

Podemos demonstrar a desigualdade de Boole usando a desigualdade de Markov.

Seja K uma variável aleatória igual à quantidade de eventos, dentre os A_i, que ocorrem. A esperança de K é

$$\mathbb{E}[K] = \Pr(A_1) + \Pr(A_2) + \cdots + \Pr(A_n) = \sum_j \Pr(A_j).$$

Pela desigualdade de Markov, com $a = 1$,

$$\begin{aligned} \Pr[K \geq 1] &\leq \frac{\mathbb{E}[K]}{1} \\ \Pr[K \geq 1] &\leq \sum_j \Pr(A_j) \\ \Pr\left(\cup_j A_j\right) &\leq \sum_j \Pr(A_j). \end{aligned}$$ ●

Corolário 6.1.1. Seja X uma variável aleatória discreta não-negativa. Se $\mathbb{E}[X] < k$ então $\Pr[X < k] > 0$.

Demonstração. Usando a desigualdade de Markov,

$$\begin{aligned} \Pr[X \geq k] &\leq \frac{\mathbb{E}[X]}{k} \\ &< 1, \end{aligned}$$

o que significa que o evento $X < t$ tem probabilidade maior que zero. ■

O Coroláio 6.1.1 é facilmente generalizado, resultando no Corolário 6.1.2, a seguir.

Corolário 6.1.2. Seja X uma variável aleatória discreta. Então,
Se $E[X] \leq k$ então $\Pr[X \leq k] > 0$.
Se $E[X] \geq k$ então $\Pr[X \geq k] > 0$.

Exemplo 6.1.4 (Satisfazendo fórmulas lógicas)**.** Uma *variável lógica* é uma variável que pode assumir dois valores: verdadeiro ou falso. Assim, se x é uma variável lógica, podemos ter $x = V$ ou $x = F$.

Se $\{x_1, x_2, \ldots, x_n\}$ é um conjunto de variáveis lógicas, então uma *cláusula lógica* usando as variáveis deste conjunto é uma fórmula lógica (ou uma "afirmação lógica") da forma

$$x_{j_1} \vee x_{j_2} \vee \cdots \vee x_{j_q},$$

onde o símbolo $\vee$ significa "ou" – ou seja,

$$\begin{aligned}
F \vee F &= F \\
F \vee V &= V \\
V \vee F &= V \\
V \vee V &= V
\end{aligned}$$

Por exemplo, sejam $x_1, x_2, x_3, \ldots, x_{10}$ variáveis lógicas. A seguir há quatro cláusulas usando estas variáveis.

$$\begin{aligned}
&x_1 \vee x_2 \vee c_4 \vee c_8 \vee x_{10} \\
&x_2 \\
&x_1 \vee x_7 \vee x_{10}
\end{aligned}$$

Uma *valoração* para um conjunto de variáveis é uma função dess econjunto em $\{V, F\}$, dando um valor a cada uma das variáveis. Por exemplo, para o conjunto $\{x_1, \ldots, x_{10}\}$ que descrevemos, uma possível valoração é

$$\begin{aligned}
x_1 &= F & x_6 &= V \\
x_2 &= F & x_7 &= F \\
x_3 &= V & x_8 &= V \\
x_4 &= V & x_9 &= F \\
x_5 &= F & x_{10} &= F
\end{aligned}$$

Há 2^n possíveis valorações para um conjunto de n variáveis.

Dizemos que uma valoração satisfaz uma cláusula quando ela torna a cláusula verdadeira. Por exemplo, a valoração que apresentamos satisfaz a cláusula $x_1 \vee x_2 \vee c_4 \vee c_8 \vee x_{10}$, porque

$$\begin{aligned}
x_1 \vee x_2 \vee c_4 \vee c_8 \vee x_{10} &= F \vee F \vee V \vee V \vee F \\
&= V,
\end{aligned}$$

já que basta que um dos valores seja V para que o resultado ("ou" lógico) seja V. Mas essa valoração não satisfaz a cláusula $x_1 \vee x_7 \vee x_{10}$, porque

$$\begin{aligned} x_1 \vee x_7 \vee x_{10} &= F \vee F \vee V \\ &= F. \end{aligned}$$

Agora suponha que tenhamos um conjunto de n variáveis, e que tenhamos m cláusulas sobre essas variáveis. Seja q o tamanho (em número de variáveis) da menor das cláusulas. Provaremos, usando a desigualdade de Markov, que existe uma valoração que satisfaz no mínimo

$$m(1 - 2^q)$$

das cláusulas.

Crie uma valoração, determinando equiprovavelmente valores V e F para as variáveis. Seja C_i a variável indicadora que vale 1 se a i-ésima cláusula é safisfeita, ou 0 se a i-ésima cláusula não é satisfeita. Então

$$\Pr[C_i = 1] = 1 - 2^{q_i},$$

porque a única maneira da cláusula valer F é com todas as suas variáveis valendo F.

Então a variável aleatória que determina o número de cláusulas satisfeitas é

$$C = \sum_{i=1}^{m} C_i,$$

com esperança

$$\begin{aligned} \mathbb{E}[C] &= \sum_{i=1}^{m} \mathbb{E}[C_i] \\ &= \sum_{i=1}^{m} \Pr[C_i = 1] \\ &= \sum_{i=1}^{m} (1 - 2^{-q_i}) \end{aligned}$$

Como a esperança é $\sum_{i=1}^{m}(1 - 2^{-q_i})$, então há uma valoração que satisfaz no mínimo esse valor. Se q é o tamanho da menor cláusula, então

$$\sum_{i=1}^{m} (1 - 2^{-q_i}) \geq m(1 - 2^q). \qquad \bullet$$

6.2 Desigualdade de Chebyshev

A desigualdade de Chebychev é, em espírito, semelhante à de Markov, mas por usar a variância, além da média, resulta em valores mais estreitos. Além disso, a desigualdade de Chebychev determina tanto um máximo como um mínimo para o valor da variável.

Teorema 6.2.1 (desigualdade de Chebychev)**.** Se X tem média μ, finita, e variância σ^2, então para todo $k > 0$,

$$\Pr\left[|X - \mu| \geq k\right] \leq \frac{\sigma^2}{k^2}.$$

Demonstração. Defina a variável aleatória $A = (X - \mu)^2$. Claramente, A é não-negativa. Pela desigualdade de Markov, para qualquer $a > 0$,

$$\Pr[A \geq a] \leq \frac{\mathbb{E}[A]}{a}.$$

Escolha $a = k^2$. Então

$$\begin{aligned}
\Pr[(X - \mu)^2 \geq k^2] &\leq \frac{\mathbb{E}[(X - \mu)^2]}{k^2} \\
\Pr[|X - \mu| \geq |k|] &\leq \frac{\mathbb{E}[(X - \mu)^2]}{k^2} \\
\Pr[|X - \mu| \geq k] &\leq \frac{\mathbb{E}[(x - \mu)^2]}{k^2} && \text{(k≥0)} \\
\Pr[|X - \mu| \geq k] &\leq \frac{\mathbb{E}[(X - \mu)^2]}{k^2} \\
\Pr[|X - \mu| \geq k] &\leq \frac{\operatorname{var}(X)}{k^2}.
\end{aligned}$$

■

Exemplo 6.2.2. O Exemplo 6.1.2 descreve uma população onde a média da pressão sistólica em homens adultos é $125mmHg$, e ali a desigualdade de Markov foi usada para deduzir que $\Pr[P > 140] \leq 0.8928$.

Se soubermos não somente a média, mas também a variância, e ignorarmos o fato de que essas medidas tem distribuição normal, poderemos usar a desigualdade de Chebyshev. Suponha que a variância da pressão sistólica é $\sigma^2 = 25$. Então

$$\begin{aligned}
\Pr[\ |P - \mu| \geq k\ ] &\leq \frac{\mu}{k^2} \\
\Pr[\ |P - 125| \geq 15] &\leq \frac{125}{15^2} \\
&= \frac{5}{9} \\
&\approx 0.5555.
\end{aligned}$$

Assim, a probabilidade de um indivíduo ter pressão sistólica acima de $140mmHg$ *ou abaixo* de $110mmHg$ é aproximadamente 0.55.

Para comparação, verificamos o valor de $\Pr[P > 140]$ usando o fato da distribuição ser normal.

A variável

$$Z = \frac{X - 125}{5}$$

tem distribuição $\mathcal{N}(1,0)$, logo

$$\begin{aligned}
\Pr[P > 140] &= \Pr\left[\frac{X-\mu}{\sigma} > \frac{140-125}{5}\right] \\
&= \Pr[Z > 3] \\
&= 0.0013,
\end{aligned}$$

o que não contradiz o resultado obtido pela desigualdade de Chebyshev, apesar de estar bastante distante dele. ●

Exemplo 6.2.3. Quando coeficientes binomiais são muito grandes, seu cálculo pode se tornar difícil. Em particular, o coeficiente $\binom{2n}{n}$ – o coeficiente "do meio", maior dentre os binomiais $\binom{2n}{\cdot}$, é bastante importante – e mostraremos neste exemplo como obter uma cota inferior para ele usando a desigualdade de Chebyshev.

Provaremos que para todo inteiro positivo n,

$$\binom{2n}{n} \geq \frac{2^{2n}}{2+4\sqrt{n}}.$$

Seja X uma variável aleatória, obtida somando $2n$ variáveis aleatórias de Bernoulli,

$$X = X_1 + X_2 + \cdots + X_{2n},$$

onde

$$\begin{aligned}
\Pr[X_i = 0] &= \frac{1}{2} \\
\Pr[X_i = 1] &= \frac{1}{2}.
\end{aligned}$$

Cada X_i é exatamente como uma variável aleatória que representa a jogada de uma moeda honesta.

A esperança e a variância de X são

$$\begin{aligned}
\mathbb{E}[X] &= n \\
\sigma_X^2 &= \frac{n}{2}
\end{aligned}$$

Se aplicarmos a desigualdade de Chebyshev em X, com $k = \sqrt{n}$,

$$\begin{aligned}
\Pr[|X-n| \geq \sqrt{n}] &\leq \frac{1}{2} \\
\Pr[|X-n| < \sqrt{n}] &\geq \frac{1}{2}. && \text{(complemento da anterior)}
\end{aligned}$$

Ou seja, a probabilidade da distância de X até a n ser menor que $\sqrt{n}$ é $\geq 1/2$. Outra maneira de afirmar isto é dizer que "a soma das probabilidades $\Pr[X = n+j]$, para todo$j < \sqrt{n}$, é $\geq 1/2$.

Como X tem distribuição binomial, a probabilidade de $X = n + j$ é

$$\begin{aligned}
\Pr[X = n + j] &= \binom{2n}{n+j} \left(\frac{1}{2}\right)^{n+j} \left(\frac{1}{2}\right)^{2n-(n+j)} \\
&= \binom{2n}{n+j} \left(\frac{1}{2}\right)^{2n} \\
&= \binom{2n}{n+j} 2^{-2n}.
\end{aligned}$$

Obtemos, portanto,

$$\sum_{|j|<\sqrt{n}} \Pr[X = n + j] \leq (2\sqrt{n} + 1)\binom{2n}{n+j} 2^{-2n} \geq \frac{1}{2}, \tag{6.2}$$

porque há $2\sqrt{n} + 1$ valores menores ou iguais que $\sqrt{n}$ (os $\sqrt{n}$ positivos, mais os $\sqrt{n}$, mais o zero).

Mas como $\binom{2n}{n}$ é o maior dentre os coeficientes binomiais $\binom{2n}{\cdot}$, então

$$\binom{2n}{n+j} 2^{-2n} \leq \binom{2n}{n} 2^{-2n}.$$

Reescrevemos a Equação 6.2, portanto:

$$\begin{aligned}
\frac{1}{2} &\leq \sum_{|j|<\sqrt{n}} \Pr[X = n + j] \leq \frac{(1 + 2\sqrt{n})\binom{2n}{n}}{2^{2n}} \\
\frac{1}{2} &\leq \frac{(1 + 2\sqrt{n})\binom{2n}{n}}{2^{2n}} \\
\frac{2^{2n}}{2(1 + 2\sqrt{n})} &\leq \binom{2n}{n},
\end{aligned}$$

que é o que queríamos mostrar. ●

6.3 Lei Fraca dos Grandes Números

O teorema conhecido como “Lei Fraca dos Grandes Números”, provado por Jacob Bernoulli[1], tem grande importância histórica.

Suponha que n variáveis aleatórias independentes $X_1, X_2, \ldots, X_n$ tenham a mesma distribuição e mesma média μ. A Lei Fraca dos Grandes Números[2] determina que, se definirmos uma nova variável aleatória $Y = (X_1 + X_2 + \cdots + X_n)/n$, sua média se aproxima de μ quando $n \to \infty$.

[1] Também Conhecido por “James”, ou “Jacques”. O nome hebraico é “Jacob” (Iacobus em Latim), e é traduzido para o Inglês como James, e em Francês como Jacques. Bernoulli era suíço, mas viajou muito pela Europa.

[2] Há também uma “Lei Forte dos Grandes Números”, cujo enunciado é parecido com o da lei fraca, porém mais forte – e também mais difícil de demonstrar.

Em situações práticas, quase sempre é melhor o uso do Teorema Central do Limite, que faz uma afirmação mais precisa, identificando a distribuição das medições e sua variância.

Teorema 6.3.1 (Lei Fraca dos Grandes Números)**.** Seja $X_1, X_2, \ldots$ uma sequência de variáveis aleatórias independentes com média finita μ, identicamente distribuídas. Então

$$\lim_{n\to\infty} \Pr\left[\ \left|\frac{X_1 + X_2 + \cdots + X_n}{n} - \mu\right| \ \geq \varepsilon\ \right] = 0$$

Demonstração. Seja

$$X = \frac{X_1 + X_2 + \cdots + X_n}{n}.$$

Então a esperança de X é

$$\begin{aligned}
\mathbb{E}[X] &= \mathbb{E}\left[\frac{X_1 + X_2 + \cdots + X_n}{n}\right] \\
&= \frac{1}{n}\mathbb{E}[X_1 + X_2 + \cdots + X_n] \\
&= \frac{1}{n}\Big(\mathbb{E}[X_1] + \mathbb{E}[X_2] + \cdots + \mathbb{E}[X_n]\Big) \\
&= \frac{1}{n}\Big(n\mu\Big) && (\mathbb{E}[X_i] = \mu) \\
&= \mu.
\end{aligned}$$

A variância de X é

$$\begin{aligned}
\operatorname{var}(X) &= \operatorname{var}\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right) \\
&= \operatorname{var}\left(\frac{X_1}{n}\right) + \left(\frac{X_2}{n}\right) + \cdots + \left(\frac{X_n}{n}\right) \\
&= \frac{\sigma^2}{n^2} + \frac{\sigma^2}{n^2} + \cdots + \frac{\sigma^2}{n^2} \\
&= \frac{n\sigma^2}{n^2} \\
&= \frac{\sigma^2}{n}.
\end{aligned}$$

Agora, usando a desigualdade de Chebyshev, temos

$$\begin{aligned}
\Pr(\ |X - \mu| \geq \varepsilon) &\leq \frac{\operatorname{var}(X)}{\varepsilon^2} \\
&= \frac{\sigma^2}{n\varepsilon^2}.
\end{aligned}$$

Observando o limite quando $n \to \infty$,

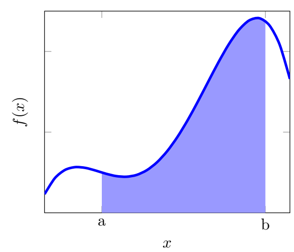
$$\Pr(\ |X - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2}$$
$$\lim_{n\to\infty} \Pr(\ |X - \mu| \geq \varepsilon) \leq \lim_{n\to\infty} \frac{\sigma^2}{n\varepsilon^2}.$$

Como σ^2 é constante e n cresce, o lado direito tende a zero. ■

Exemplo 6.3.2 (Integração de Monte Carlo)**.** Sabemos que a função de densidade da distribuição normal é integrável, mas como ela não tem forma fechada (não conseguimos resolvê-la analiticamente), consultamos seus valores em uma tabela. Há diversas intergais que são difíceis ou impossíveis de resolver, e quando há a necessidade de resolvê-las, usamos métodos numéricos.

Um dos métodos muito usados para integração numérica é o *Método de Monte Carlo*, que consiste em usar amostragem de pontos repetidamente para estimar o valor da integral.

Suponha que queiramos integrar uma função real $f(x)$ no intervalo $[a, b]$.



Seja $p(x)$ uma função de densidade de probabilidade *qualquer*. Então, vemos que

$$\int_a^b f(x)dx = \int_a^b f(x)\frac{p(x)}{p(x)}dx$$
$$= \int_a^b p(x)\ \frac{f(x)}{p(x)}dx.$$

Mas esta integral é exatamente a fórmula para a esperança de $f(x)/p(x)$:

$$\mathbb{E}\left[\frac{f(x)}{p(x)}\right].$$

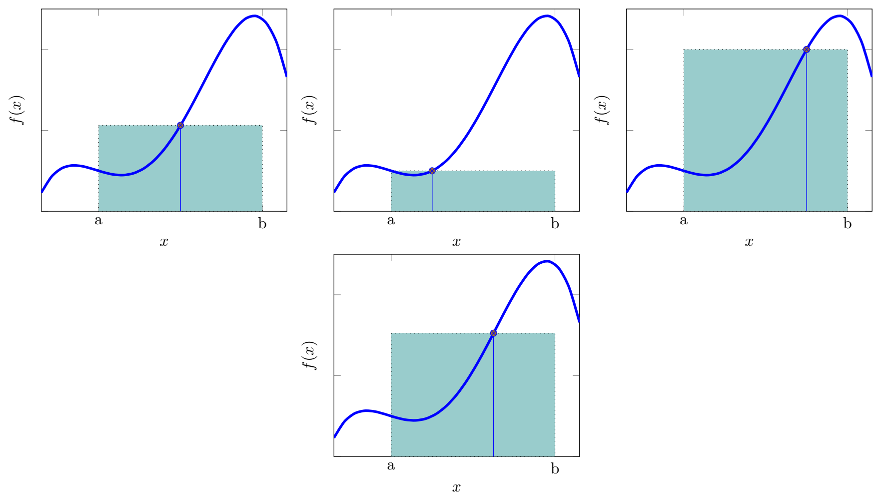
Suponha que $p(x)$ é a densidade uniforme. Então $p(x) = 1/(b - a)$, e podemos calcular

$$\mathbb{E}\left[(b - a)f(x)\right].$$

Sorteamos n pontos $x_1, x_2, \ldots, x_n$, e definimos a variável aleatória

$$A_i = (b - a)f(x_i).$$

Cada vez que sorteamos um ponto e realizamos esse cálculo, estamos selecionando um valor de uma variável aleatória A_i. Cada A_i pode ser visto como o valor da área do retângulo de lado $b - a$ e altura $f(x_i)$, e estamos portanto calculando a média de uma sequência de áreas.



Se obtivermos uma quantidade suficientemente grande de amostras de $f(x)/p(x)$, a média desses valores convergirá para $\mathbb{E}[f(x)/p(x)]$, que é o valor da integral. Isso porque as variáveis A_i são independentes e identicamente distribuídas, e *a Lei Fraca dos Grandes Números garante que a média das variáveis A_i é igual a* $\mathbb{E}\left[\frac{f(x)}{p(x)}\right]$. ●

6.4 Teorema Central do Limite

A distribuição da soma de muitas variáveis independentes aproxima-se da distribuição normal. Por exemplo, seja qual for a distribuição das medidas de altura de pessoas em diferentes amostras de uma população, a soma delas (e consequentemente sua média) tem distribuição normal.

Teorema 6.4.1 (teorema central do limite)**.** Afirmamos o Teorema Central do Limite de duas maneiras:

1. (*Soma de variáveis*) Seja $X_1, X_2, \ldots$ uma sequência de variáveis aleatórias independentes com média finita μ, e variância σ^2, identicamente distribuídas. Seja S a soma de n destas variáveis,

   $$S = X_1 + X_2 + \cdots + X_n.$$

   Então a distribuição de

   $$\frac{S - n\mu}{\sigma\sqrt{n}} \sim \mathcal{N}(0,1)$$

   quando $n \to \infty$.

2. (*Média de variáveis*) Sejam $X_1, X_2, \ldots, X_n$ variáveis aleatórias independentes com média finita μ, e variância σ^2, identicamente distribuídas. Seja $\overline{X}$ a média das n variáveis,

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}.$$

Então a distribuição de

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$

tende a $\mathcal{N}(0,1)$ quando $n \to \infty$.

Ou seja, se $\overline{X} = (X_1 + X_2 + \cdots + X_n)/n$,

$$\overline{X} \sim \mathcal{N}(\mu, \sigma^2/n).$$

quando $n \to \infty$.

O Teorema Central do Limite diz que a distribuição de $(S - n\mu)/\sigma\sqrt{n}$ é a mesma de uma variáevl Z com distribuição $\mathcal{N}(0,1)$, então podemos usar a tabela cumulativa de $\mathcal{N}(0,1)$ para calcular, por exemplo,

$$\Pr\left[\frac{S - n\mu}{\sigma\sqrt{n}} \leq x\right] = \Pr[Z \leq x].$$

Da mesma forma,

$$\Pr\left[\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \leq x\right] = \Pr[Z \leq x].$$

A respeito da diferença entre as duas versões, trivialmente

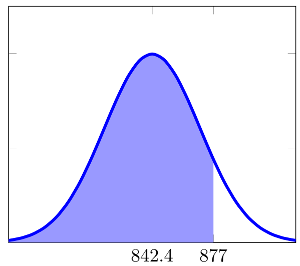
$$\begin{aligned}
\frac{S - n\mu}{\sigma\sqrt{n}} &\sim \mathcal{N}(0,1) \\
\frac{(S - n\mu)/n}{(\sigma\sqrt{n})/n} &\sim \mathcal{N}(0,1) \\
\frac{\overline{X} - \mu}{(\sigma\sqrt{n})/n} &\sim \mathcal{N}(0,1) && (\overline{X} = S/n) \\
\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} &\sim \mathcal{N}(0,1). && (\sqrt{n}/n = 1/\sqrt{n})
\end{aligned}$$

Exemplo 6.4.2 (Teorema Central do Limite (para soma))**.** Há 324 pessoas na minha frente em uma fila[3] para comprar ingressos para um evento. Há 877 ingressos restantes. A quantidade de ingressos comprada por pessoa tem média 2.6 e variância 2.25. Qual é a minha probabilidade de sucesso?

$$\Pr[\text{os 324 compram } \underline{\text{juntos}} \leq 877]$$

[3] Fila online ou fila física – não fará diferença!

A palavra "juntos" indica que estamos somando as variáveis aleatórias X_i, que indicam a quantidade de ingressos que cada pessoa na fila comprará. Posso, portanto, presumir que a distribuição da soma dos X_i é normal (com média $n\mu = 324(2.6)$ e variância $n\sigma^2 = 324(2.25)$).



Pelo Teorema Cental do Limite,

$$\begin{aligned}
\Pr[T \leq 877] &\approx \Pr\left[Z \leq \frac{877 - 324(2.6)}{\sqrt{2.25}\sqrt{324}}\right] \\
&= \Pr\left[Z \leq \frac{877 - 842.4}{1.5(18)}\right] \\
&= \Pr\left[Z \leq \frac{34.60}{27}\right] \\
&= \Pr\left[Z \leq 1.28\right] \\
&= \Phi(1.28) \\
&\approx 0.8997
\end{aligned}$$

●

Exemplo 6.4.3 (Teorema Central do Limite (para soma))**.** A quantidade máxima de unidade de dados que pode ser transmitida por vez (em um "pacote") em diferentes tipos de rede de computadores é chamada de *MTU* – *maximum transmission unit*. Por exemplo, redes do tipo *ethernet* tem *MTU* igual a 1500 bytes.

Em uma rede deste tipo, medimos a quantidade de bytes que são transmitidos com erro, e verificamos que 5% dos bytes chegam com erro (e precisam ser retransmitidos).

Calcularemos a probabilidade de mais do que 100 bytes tenham que ser retransmitidos em um pacote.

Cada byte pode ou não ser transmitido sem erro. Chamamos a transmissão sem erro de "sucesso"; claramente se trata de um experimento de Bernoulli, cmo probabilidad ede sucesso igual a 0.95. Para cada um dos 1500 bytes, criamos uma variável aleatória X_i, de forma que $X_i = 1$ quando o i-ésimo byte foi transmitido com erro, e $X_i = 0$ quando chegou sem erro.

As variáveis são identicamente distribuídas, e presumimos que são independentes. Como cada variável é um experimento de Bernoulli, suas esperanças e variâncias são

$$\begin{aligned}\mathbb{E}[X_i] &= p = 0.05\\ \operatorname{var}(X_i) &= p(1-p) = 0.0475.\end{aligned}$$

Agora podemos usar o Teorema Central do Limite.

$$\begin{aligned}\Pr[X > 100] &= \Pr\left[\frac{X - n\mu}{\sqrt{n}\sigma} > \frac{100 - n\mu}{\sqrt{n}\sigma}\right]\\ &= \Pr\left[Z > \frac{100 - (1500)(0.05)}{\sqrt{1500}\sqrt{0.0475}}\right]\\ &= \Pr\left[Z > \frac{25}{8.44}\right]\\ &= 1 - \Pr\left[Z \leq 2.96\right]\\ &= 1 - \Phi(2.96)\\ &\approx 1 - 0.9985\\ &\approx 0.0015.\end{aligned}$$

●

Exemplo 6.4.4 (Teorema Central do Limite (para média))**.** A embalagem de uma marca de creme dental informa a quantidade de $70g$. O fabricante informa que o desvio padrão é de $2g$. Após realizar medições em 81 unidades dessa marca de creme dental, verifiquei que a média era de $69.5g$.

Qual é a probabilidade da média amostral ficar abaixo ou igual a esse valor, presumindo que o que o fabricante diz é verdade?

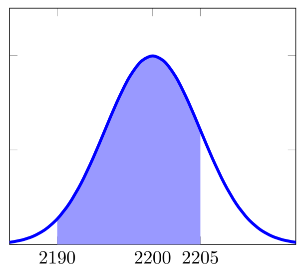
Como esta é uma afirmação sobre uma média de amostra de variáveis, usamos o Teorema Central do Limite para médias.

$$\begin{aligned}\Pr\left[\overline{X} < 69\right] &= \Pr\left[Z < \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}\right]\\ &= \Pr\left[Z < \frac{69.5 - 70}{2/\sqrt{81}}\right]\\ &= \Pr\left[Z < \frac{-0.5}{2/9}\right]\\ &= \Pr\left[Z < -2.25\right]\\ &= \Phi(-2.25)\\ &= 1 - \Phi(2.25)\\ &\approx 1 - 0.9878\\ &\approx 0.012.\end{aligned}$$

A probabilidade de eu ter chegado a essa média amostral, ou algo menor que ela, se o fabricante estivesse correto em sua afirmação, seria de 0.012. A probabilidade da média amostral ser *maior* que 69.5 é aproximadamente 0.9878.

Não podemos dizer que a média amostral de 69.5 não poderia ter acontecido neste experimento, mas podemos dizer que ela era pouco provável: a probabilidade de *qualquer valor* menor que 69.5 era de 0.012. ●

Exemplo 6.4.5 (Teorema Central do Limite (para soma, desigualdade dupla))**.** Um fabricante de lanternas químicas – do tipo que emite luz quando duas substâncias entram em contato – informa que o período de emissão de luz tem média de 22 horas com desvio padrão de 30 minutos. Compramos 100 dessas lanternas, e determinaremos agora a probabilidade do período total de luz (usando as velas sequencialmente) ficar entre 2190 e 2205 horas.



$$
\begin{aligned}
\Pr[2190 < T < 2205] &= \Pr\left[\frac{2190 - n\mu}{\sigma\sqrt{n}} \ < \ Z \ < \frac{2205 - n\mu}{\sigma\sqrt{n}}\ \right] \\
&= \Pr\left[\frac{2190 - (100)22}{0.5\sqrt{100}} \ < \ Z \ < \frac{2205 - (100)22}{0.5\sqrt{100}}\ \right] \\
&= \Pr\left[\frac{-10}{0.5(10)} \ < \ Z \ < \frac{5}{0.5(10)}\ \right] \\
&= \Pr\left[-2 \ < \ Z \ < 1\ \right] \\
&= \Phi(1) - \Phi(-2) \\
&= \Phi(1) - [1 - \Phi(2)] \\
&\approx 0.8413 - 1 + 0.9772 \\
&\approx 0.8185.
\end{aligned}
$$
●

Exemplo 6.4.6. O Teorema Central do Limite não se aplica a variáveis com distribuição de Cauchy, porque elas não tem média definida. Assim, não é possível usar o Teorema Central do Limite para a variável aleatória X do experimento descrito na Seção 4.8 (página 123), que descreve as posições definidas no eixo das abscissas. ●

A seguir definimos nível de confiança e de significância, para em seguida obter uma forma de determinar o tamanho mínimo de amostras quando queremos estimar a média de uma variável aleatória.

Definição 6.4.7 (nível de confiança; nível de significância)**.** Sejam $X_1, X_2, \ldots, X_n$ variáveis aleatórias identicamente distribuídas e independentes; seja $S = X_1 +$

$X_2 + \cdots + X_n$. Então

$$\Pr\left[\left|\frac{S}{n} - \mu\right| \leq \varepsilon\right] \geq 1 - \alpha$$

é a probabilidade da média S/n estar a uma distância menor que ε da verdadeira média das variáveis X_i. Dizemos que *com amostra de tamanho n, podemos estimar a média das variáveis com erro igual a ε e nível de confiança igual a $1 - \alpha$*. ◀

Teorema 6.4.8. Sejam $X_1, X_2, \ldots, X_n$ variáveis aleatórias identicamente distribuídas e independentes; seja $S = X_1 + X_2 + \cdots + X_n$. Então, fixados α e ε,

$$\Pr\left[\left|\frac{S}{n} - \mu\right| \leq \varepsilon\right] \geq 1 - \alpha$$

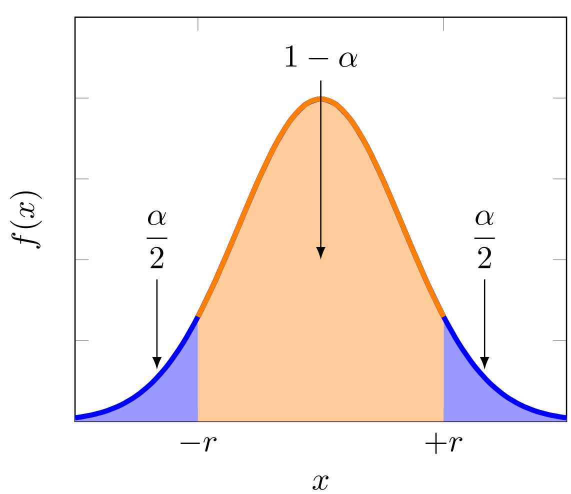
depende apenas de n, que, para atingir o erro e confiança determinados, deve ser

$$n \geq \frac{\sigma^2 r^2}{\varepsilon^2},$$

onde r é tal que

$$\Phi(r) = 1 - \left(\frac{\alpha}{2}\right).$$

Demonstração. A escolha de $\Phi(r) = \alpha/2$ é porque queremos que a probabilidade da média amostral ficar dentro do intervalo especificado é $1 - \alpha$:



Dado um n, usamos o Teorema Central do Limite para média e definimos

$$Z_n = \frac{S/n - \mu}{\sigma/\sqrt{n}}.$$

Então,

$$\begin{aligned}
\Pr\left[-r \leq Z_n \leq +r\right] &= 1-\alpha \\
\Pr\left[-r \leq \frac{S/n-\mu}{\sigma/\sqrt{n}} \leq +r\right] &= 1-\alpha \\
\Pr\left[-\frac{r\sigma}{\sqrt{n}} \leq \frac{S}{n}-\mu \leq +\frac{r\sigma}{\sqrt{n}}\right] &= 1-\alpha \\
\Pr\left[\mu-\frac{r\sigma}{\sqrt{n}} \leq \frac{S}{n} \leq \mu+\frac{r\sigma}{\sqrt{n}}\right] &= 1-\alpha \\
\Pr\left[\frac{S}{n} = \mu \pm \frac{r\sigma}{\sqrt{n}}\right] &= 1-\alpha
\end{aligned}$$

Assim, queremos que

$$\begin{aligned}
\frac{r\sigma}{\sqrt{n}} &\leq \varepsilon \\
r\sigma &\leq \varepsilon\sqrt{n} \\
\frac{r\sigma}{\varepsilon} &\leq \sqrt{n} \\
n &\geq \frac{r^2\sigma^2}{\varepsilon^2}.
\end{aligned}$$

■

Exemplo 6.4.9. Temos a tarefa de realizar uma pesquisa de opinião, a fim de determinar a propoção de pessoas que prefere uma dentre duas opções para uma nova regra para uso de um refeitório.

Queremos que o erro seja de no máximo 0.02, com nível de confiança de 95%. Temos $\alpha = 0.05$, portanto escolhemos $r = 1.96$, já que

$$\Phi(1.96) = 0.975 = 1 - \frac{\alpha}{2}.$$

A quantidade de pessoas que teremos que entrevistar teria que ser

$$n \geq \frac{1.96^2\sigma^2}{0.02^2}.$$

No entanto, não sabemos a variância da distribuição. Nesta situação, podemos fazer uma estimativa para a variância: cada eleitor pode ser modelado como uma variável aleatória de Bernoulli, e o total de votos por A na eleição, por exemplo, teria distribuição binomial. A variância da binomial é $p(1-p)$ – o que parece não ajudar, porque se soubéssemos p, a pesquisa não seria necessária. No entanto, observamos que como $0 \leq p \leq 1$,

$$p(1-p) \leq \frac{1}{4},$$

e podemos usar $1/4$ como uma estimativa pessimista da variância. Dessa forma,

$$\begin{aligned}
n &\geq \frac{1.96^2}{4(0.02^2)} \\
&= 2401.
\end{aligned}$$

●

Exemplo 6.4.10. Conhecemos o desvio padrão de uma variável X, mas não sua média. Tomaremos n amostras, e calcularemos a média das amostras – mas gostaríamos que o erro fosse de no máximo 200. Sendo o desvio padrão igual a 2000, qual é o menor valor de n para que consigamos, com probabilidade 0.99, uma estimativa dentro da margem de erro determinada?

Temos $\alpha = 0.01$, portanto escolhemos $r = 2.57$, já que

$$\Phi(2.57) \approx 0.9949 \approx 1 - \frac{\alpha}{2}.$$

$$\begin{aligned} n &\geq \frac{r^2\sigma^2}{\epsilon^2} \\ &= \frac{2.57^2 2000^2}{200^2} \\ &\approx 660.5 \end{aligned}$$

e precisamos de uma amostra de tamanho $n = 661$. ●

Exercícios

Ex. 136 — Seja X uma variável binomial com parâmetros (n, p). Se $p < k < 1$, encontre um limite superior para

$$\Pr[X \geq kn].$$

Ex. 137 — Suponha que uma variável aleatória X assume valores sempre menores ou iguais a $k \in \mathbb{R}$. Prove que, para todo $x \leq k$,

$$\Pr[X \leq x] \leq \frac{k - \mathbb{E}[X]}{k - x}$$

Ex. 138 — Seja X uma variável uniformemente distribuida no intervalo $[0, 10]$. Calcule $\Pr[|X - 5| > 4]$ usando a desigualdade de Chebychev e compare com o valor exato.

Ex. 139 — Prove que para qualquer variável aleatória X, se $\sigma_X^2 = 0$ então

$$\Pr[X = \mathbb{E}[X]] = 1.$$

Ex. 140 — Mil valores, $x_1, x_2, \ldots, x_{1000}$, são sorteados no intervalo $[-1, 1]$. Qual é a probabilidade de

$$\sum x_i \geq 1000\ ?$$

Ex. 141 — Em uma barreira rodoviária, a carga máxima permitida é de $19800kg$. Um determinado caminhão, sem carga, pesa $10000kg$.
O caminhão está carregado com 49 caixas, sendo o peso médio das caixas $205kg$, com desvio padrão igual a $15kg$. Se cada caixa tiver peso igual à média, o peso total da carga será de $10045kg$, que somado ao do caminhão é maior que o permitido. Poderíamos imaginar que, havendo variância maior que zero, muitas caixas podem estar abaixo do peso, e talvez fosse possível passar.
Qual é a probabilidade deste caminhão conseguir permissão para passar pela barreira?

Ex. 142 — A duração da gestação para mulheres em uma população tem média igual a 268 dias, com desvio padrão de quinze dias. Um posto de saúde em área remota está acompanhando 20 mulheres grávidas. Qual é a probabilidade da média de tempo de gestação *destas 20 mulheres* ser menor que 264 dias?

Ex. 143 — Considere o método de integração numérica a seguir, parecido com o descrito no Exemplo 6.3.2.
Suponha que queiramos calcular

$$\int_a^b f(x)dx$$

para alguma função real $f(x)$. Suponha também que saibamos que $\alpha < f(x) < \beta$.
O método consiste em executar, n vezes, os seguintes passos.

i. escolha $x \in [a, b]$, aleatoriamente;
ii. escolha $y \in [\alpha, \beta]$, aleatoriamete;
iii. calcule $f(x)$;
iv. se $f(x) < y$, aumente o valor do contador c em um.

O resultado é c/n.

1. Prove que o valor obtido neste método tende ao valor da integral quando n tende a infinito.
2. Que desvantagem este método tem, comparado ao outro, apresentado no texto?

Ex. 144 — A Agência Nacional de Vigilância Sanitária determinou, em 2013, que o sal vendido no Brasil seja iodado, e que a quantidade de iodo em cada kg de sal fique entre $15mg$ e $45mg$.
Antes de 2013, a quantidade de iodo em um quilo de sal deveria ficar entre $20mg$ e $60mg$.
Em um estudo feito na época da adaptação, uma fábrica verificou que a quantidade de iodo em cada kg de sal que produzia tinha média igual a $30mg$, com desvio padrão igual a 10. Qual é a probabilidade de um pacote de um quilo de sal produzido por esta fábrica esteja em desacordo com a nova regra? E com a regra antiga?

Ex. 145 — Calcule

$$\lim_{n\to\infty} e^{-n} \sum_{k=0}^{n} \frac{n^k}{k!}.$$

Dica 1: O limite _não_ é um, por mais que isso possa surpreender.
Dica 2: leve a exponencial para dentro do somatório. Estamos, talvez, somando variáveis aleatórias?

Ex. 146 — Proponha uma maneira de estimar o valor de π usando um método de Monte Carlo.

Ex. 147 — (Programação) Use um método de Monte Carlo para terminar de resolver o item 2 do Exercício 88, na página 128 (Capítulo 4).

Ex. 148 — Faremos uma pesquisa de opinião para determinar a intenção de voto em dois candidatos. Quantas pessoas, no mínimo, deveremos consultar, se quisermos que o erro seja de no máximo 1%, com nível de confiança de 99%?

Ex. 149 — Uma variável X tem distribuição exponencial com parâmetro $\lambda = 2$. Seja $Y = X^2$. Agora suponha que haja várias variáveis, $Y_1, Y_2, \ldots, Y_n$, todas com a mesma distribuição de Y. Qual deve ser a distribuição aproximada da média das variáveis Y_i, para n grande?

Ex. 150 — Uma bifurcação em um túnel leva a duas plataformas, A e B, onde pessoas embarcam em composições ferroviárias subterrâneas. Em um certo horário, é sabido que a quantidade de pessoas passando pela bifurcação é de 5000 pessoas a cada cinco minutos (que é o intervalo entre partidas dos trens), e que elas escolhem entre as duas plataformas equiprovavelmente. Quantos lugares cada composição deve ter para que haja lugar para todos se sentarem, com probabilidade 0.95?

Ex. 151 — Suponha que uma sequência de variáveis $X_1, X_2, \ldots$ tenha distribuição $\mathcal{P}oiss(\lambda)$, e que λ não seja conhecido. Se tivesse que determinar a quantidade mínima de experimentos para poder estimar λ, e quisesse tentar usar o Teorema Central do Limite, que dificuldade encontraria?

Capítulo 7

Gerando Distribuições de Probabilidade

Em outros Capítulos tratamos de analisar e descrever distribuições de probabilidade. Neste, mudamos a abordagem e nos aproximamos do problema de obter amostras que tenham uma distribuição desejada. Isto é de grande importância em simulações e outras áreas.

Gostaríamos de poder, por exemplo, simular as durações de conversas em chat, produzindo uma sequência de números, cada um representando uma duração, e não somente descrever que distribuição seguiriam. De maneira semelhante, queremos poder reproduzir, de maneira simulada, o experimento da agulha de Buffon; os três experimentos do paradoxo de Bertrand; ou mesmo os experimentos relacionados aos ângulos de incidência dos meteoros que formaram as das crateras lunares.

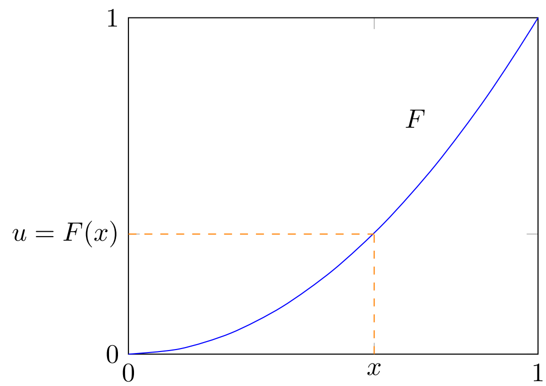
Para conseguirmos realizar essas simulações, precisamos produzir sequências de eventos com distribuições definidas: como produzir uma sequencia de números que poderia ser obtida de uma variável aleatória com distribuição de Poisson, exponencial, ou qualquer outra?

Em geral, podemos presumir que temos à disposição um gerador de números pseudoaleatórios que produz números no intervalo $[0, 1]$, com distribuição indistinguível de uniforme. Podemos portanto dizer, de maneira simplificada, que "temos acesso a amostras de uma variável aleatória uniforme no intervalo $[0, 1]$". A partir deste gerador, conseguiremos simular diferentes variáveis aleatórias, com diferentes distribuições. Há diversos métodos para a geração de amostras de distribuições. Neste Capítulo abordamos dois deles: o método da inversão e o método da rejeição.

7.1 O Método da Inversão

Queremos simular eventos descritos por uma variável aleatória X, que tem alguma distribuição contínua. Sabemos a função de densidade de probabilidade

$f(x)$ e a função de densidade acumulada $F(x)$, e temos acesso a um gerador de números aleatórios com distribuição uniforme em $(0,1)$, que representamos como uma variável aleatória U. A função $F(x)$, em particular, é não decrescente, $0 \leq F(x) \leq 1$.



Escolhemos um valor $u \in [0,1]$ e aplicamos a inversa $F^{-1}(u)$, obtendo x. Os valores obtidos terão a mesma distribuição da variável X. Isso porque a probabilidade de escolhermos $X < k$ é

$$\Pr[X \leq k] = \Pr[U \leq F(k)],$$

mas $\Pr[U \leq F(k)]$ é o valor de $F(k)$, ele mesmo, porque em $(0,1)$, $\Pr[U < z] = z$.

Exemplo 7.1.1. O Exemplo 4.5.4 trata de um sistema interativo onde são medidas as durações de conversas. Essas durações foram modeladas ali como uma variável exponencial com média igual a 90 segundos. Para simular as durações de conversas, podemos usar o método da inversa.

A função de distribuição é $f(x) = 90e^{-90x}$, e a função acumulada de distribuição é

$$F(x) = 1 - e^{-90x}.$$

Calculamos a inversa,

$$F^{-1}(x) = -\frac{\log(1-x)}{90}.$$

Sorteamos números u com distribuição uniforme no intervalo $[0,1]$ e aplicamos

$$F^{-1}(u).$$

A sequência de números gerados será uma sequência com a distribuição das conversas, tendo média igual a 90 segundos. ●

Exemplo 7.1.2. A distribuição de Cauchy tem função de densidade $f(x) = 1/\pi(1+x^2)$, e função de densidade acumulada

$$F(a) = \int_{-\infty}^{a} \frac{1}{\pi(1+x^2)} dx = \frac{1}{2} + \frac{\arctan(a)}{\pi}.$$

A inversa da função de densidade é

$$F^{-1}(u) = \tan\left(\pi u - \frac{\pi}{2}\right),$$

que podemos usar para gerar amostras usando a distribuição de Cauchy. Observamos que $u \in (0, 1)$, e isto claramente é o mesmo que gerar um ângulo $x \in (-\pi/2, +\pi/2)$ e tomar sua tangente (o que está claramente de acordo com a maneira como a distribuição foi definida). ●

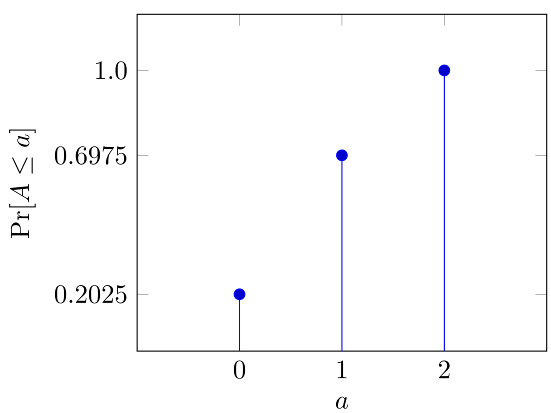
O método da inversão pode ser aplicado também a distribuições discretas, embora com alguns cuidados.

Exemplo 7.1.3. Em um experimento, jogamos uma moeda não honesta duas vezes. A probabilidade de cara é 4.5 e de coroa é 5.5. Definimos uma variável A que contabiliza a quantidade de vezes que o resultado é cara. A função de distribuição de massa de A é

$$\begin{aligned} f(0) &= 0.45^2 = 0.2025 \\ f(1) &= 2(0.45)(0.55) = 0.495 \\ f(2) &= 0.55^2 = 0.3025 \end{aligned}$$

A função de distribuição acumulada é

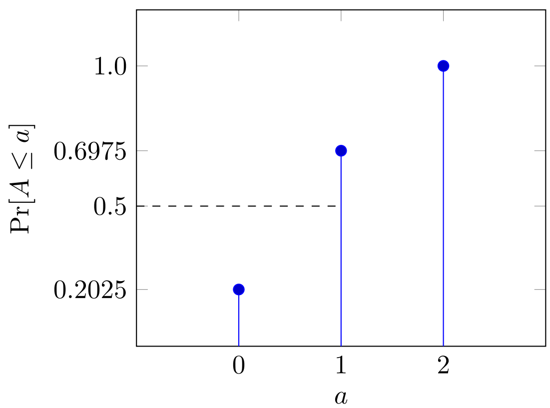
$$\begin{aligned} F(0) &= 0.2025 \\ F(1) &= 0.6975 \\ F(2) &= 1 \end{aligned}$$



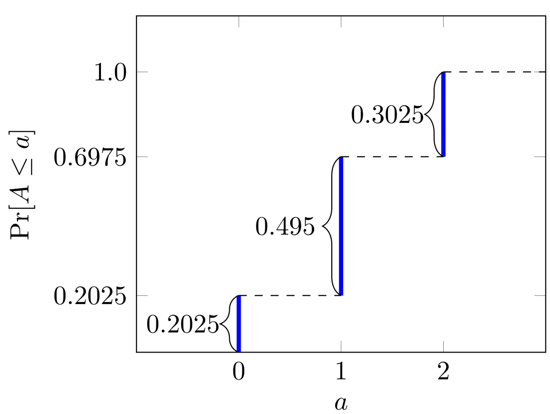
Observamos que a terceira coluna sempre terá altura igual a um, e que portanto *a altura do gráfico sempre será um.*

Se selecionarmos um número aleatório entre zero e um, determinando uma altura no gráfico, e traçarmos uma reta da esquerda para a direita nesta altura, identificamos a primeira coluna que esta reta intercepta.

Por exemplo, se o númdero selecionado for 0.5, a segunda coluna será interceptada.



A probabilidade de selecionarmos a primeira coluna é a área em que ela é atingida antes das outras, $f(0) = 0.2025$; a da segunda coluna, da mesma forma é $f(1) = 0.495$, e a da terceira é $f(2) = 0.3025$, como fica claro na representação a seguir.



A função de massa de probabilidade é $f(a)$, e o que fizemos foi obter a a partir do valor de $f(a)$ – portanto fizemos algo semelhante a inverter a função. Na verdade não a invertemos, porque não é bijetora, mas o nome “inversão” é normalmente usado neste contexto. ●

7.2 O Método da Rejeição

É possível que não seja fácil ou mesmo que não seja possível obter da inversa da função acumulada de probabilidade, e o método da inversão, embora simples e eficaz, pode não ser aplicável. Há diversos métodos que podem ser usados na geração de distribuições, e um deles, também bastante simples, é o da rejeição.

7.2.1 Interprtação geométrica inicial

Inicialmente tratamos de dois exemplos onde, para gerar pontos em uma região R, o método gera valores com uma distribuição A, em uma região S maior e *contendo* a região que nos interessa – $R \subseteq S$, e rejeitar os pontos que foram gerados fora de R (em $S \setminus R$). Isso garante que os pontos gerados em R são distribuídos de acordo com A. Mais adiante aplicamos o mesmo ao gráfico de funções de densidade, chegando ao método da rejeição.

Exemplo 7.2.1. É comum em aplicações computacionais, e em particular em Criptografia, a necessidade de obter números pseudoaleatórios entre zero e um inteiro n.

Computadores trabalham com *palavras de máquina*: os bits não são tratados individualmente, mas armazenados em "blocos" de 8, 16, 32, 64 em sequência. Assim, em uma palavra de oito bits, o número 14 é representado por `0000 1110`, de forma que *o i-ésimo bit da direita para a esquerda representa* 2^i:

$$\begin{aligned}\texttt{0000 1110} &= (0)2^7 + (0)2^6 + (0)2^5 + (0)2^4 + (1)2^3 + (1)2^2 + (1)2^1 + (0)2^0\\ &= 2^3 + 2^2 + 2^1\\ &= 14.\end{aligned}$$

Podemos obter bits aleatórios em palavras. No entanto, suponha que n não seja da forma $2^k - 1$: por exemplo, $n = 100$. O número 100 é representado por

$$\begin{aligned}\texttt{0110 0100} &= (0)2^7 + (1)2^6 + (1)2^5 + (0)2^4 + (0)2^3 + (1)2^2 + (0)2^1 + (0)2^0\\ &= 2^6 + 2^5 + 2^2\\ &= 100.\end{aligned}$$

Se quisermos sortear um número entre zero e 100 (incusive), não podemos somente sortear com distribuição uniforme uma palavra de oito bits, porque o número que ela representa pode ser maior que 100. Mesmo se usássemos somente os sete bits da direita, poderíamos eventualmente obter `0110 0101` $=$ 101. Se tentássemos usar somente seis bits, nunca conseguiríamos chegar a 100.

Usamos o seguinte método:

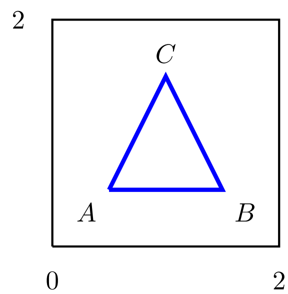
```
repita:
    escolha palavra X de 8 bits com distribuição uniforme
    se X representa número entre 0 e 100
        retorne X
```

Isto nos garante que obteremos um número dentro das restrições desejadas, e com distribuição uniforme. ●

Exemplo 7.2.2 (Geração de pontos em triângulo com método da rejeição)**.** Suponha que queiramos escolher com distribuição uniforme pontos em uma região geométrica – por exemplo, o triângulo ABC, com

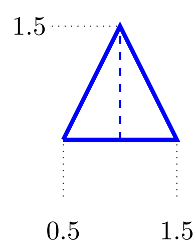
$$
\begin{aligned}
A &= (0.5, 0.5)\\
B &= (1.5, 0.5)\\
C &= (1, 1.5).
\end{aligned}
$$

Podemos escolher facilmente pontos em um retângulo que contenha o triângulo:



Os lados do retângulo tem comprimento 2.

O comprimento da base do triângulo é 1, e ele pode ser dividido em dois triângulos retângulos com base 0.5:

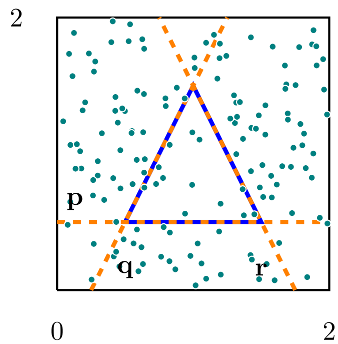


A área do triângulo é, portanto, 0.5.

Faremos isso porque é fácil gerar dois números x, y entre 0 e 2. Repetimos, então

```
repita:
    gere dois números x e y, ambos em [0,2]
    se o ponto (x,y) estiver dentro do triângulo
        retorne (x,y)
```

Determinar se o ponto está dentro do triângulo é simples: basta verificar se ele está acima da reta p e abaixo das duas retas q e r.



A probabildade de conseguirmos um ponto dentro do triângulo é dada pela área do triângulo dividida pela do retângulo. A área do triângulo é 0.5

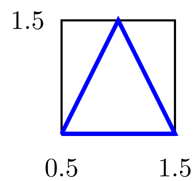
No exemplo dado, a probabilidade será

$$p = \frac{0.5}{4} = 0.125.$$

É interessante que podemos olhar para a tentativa de obter um ponto como um experimento de Bernoulli – e para o processo todo como tendo distribuição geométrica (quantas tentativas precisamos para obter um ponto no triangulo?) com parâmetro $p = 0.125$. Usando a função de distribuição acumulada de massa, verificamos que a de obter um ponto com três tentativas é

$$\Pr(X \leq 3) = 1 - (0.875)^3 \approx 0.330.$$

O retângulo que escolhemos ainda pode ser diminuído, de forma a diminuir a área "inútil" onde sorteamos pontos.



agora a área do retângulo é 1, bem menor que a do anterior (4). A probabilidade de obtermos um ponto bom em cada tentativa passa a ser

$$p = \frac{0.5}{1} = 0.5,$$

bem maior que a anterior (neste caso pode-se também verificar por inspeção, que a área interna do triângulo é metade da do quadrado).

Novamente usamos a função de distribuição acumulada para obter a probabilidade de conseguirmos um ponto em três tentativas:

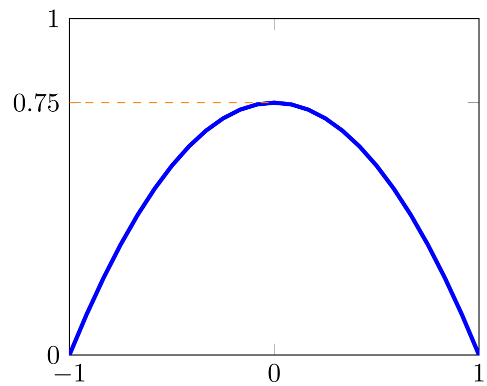
$$\Pr(X \leq 3) = 1 - (0.5)^3 \approx 0.875,$$

bem melhor que a anterior. ●

7.2.2 Gerando distribuições arbitrárias

Suponha que queiramos simular uma variável X com função de densidade $f(x)$. Sabemos que $X \in [a, b]$, e $f(x) \in [0, c]$ (lembramos que para variáveis contínuas c pode ser maior que um). Podemos sortear valores x, y e verificar se o ponto está abaixo do gráfico de $f(x)$. Se estiver, aceitamos os dois valores; senão, rejeitamos e continuamos. O Exemplo 7.2.2. onde obtivemos pontos em um triângulo, poderia ser interpretado desta forma, se o gráfico da função de densidade de fato formar um triângulo.

Exemplo 7.2.3. Simularemos uma variável Z com densidade $f(z) = -3(z-1)(z+1)/4$ sendo Z definida no intervalo $[-1, 1]$.



O máximo de $f(z)$ é $3/4 = 0.75$.

Geramos um par (x, y) onde x é escolhido de uma distribuição uniforme em $[-1, 1]$ e y é escolhido com distribuição uniforme em $[0, 0.75]$. Se o ponto (x, y) estiver abaixo de $f(z)$, aceitamos x. Caso contrário, tentamos novamente. ●

No Exemplo 7.2.3 amostramos pontos em uma distribuição uniforme para depois aceitar ou rejeitar de acordo com uma distribuição específica. Geometricamente, amostramos em um retângulo (a densidade uniforme) e rejeitamos o que estiver acima do gráfico de densidade que queremos.

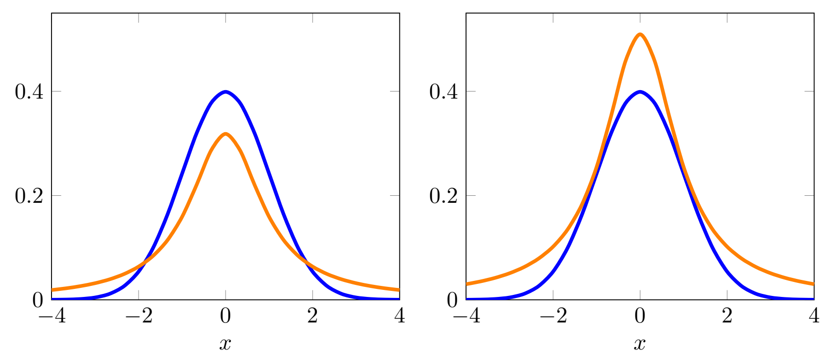
É possível que a área de rejeição seja demasiado grande, e a probabilidade de conseguirmos um ponto muito baixa. Isso tornaria o método muito lento. Ainda mais, pode ser que a distribuição que queiramos gerar não possa ser delimitada por um triângulo (por exemplo, exponencial, normal e Cauchy).

Podemos generalizamos a idéia: queremos gerar amostras de uma variável Y. Se soubermos que há uma variável X com densidade $g(x)$ e uma constante c tais que $cg(x) < f(y)$, podemos amostrar $cg(x)$ e rejeitar os valores maiores que $f(y)$.

Exemplo 7.2.4. Queremos simular uma distribuição normal padrão. Sabemos simular uma distribuição de Cauchy usando inversão, mas não sabemos como inverter a densidade acumulada da distribuição normal.

Mas podemos amostrar de uma variável C distribuição de Cauchy e aceitar somente se o valor estiver abaixo da densidade da normal, e assim simular a variável N com distribuição normal.

No entanto, precisamos multiplicar a variável que amostramos com distribuição de Cauchy por uma constante, porque de outra forma, a densidade de C nem sempre estará acima da de N.



Na segunda figura, a variável C foi multiplicada por 1.6.

repita:
  gere u com distribuição uniforme em $[0,1]$
  $y \leftarrow \tan\left(\pi u - \frac{\pi}{2}\right)$ /* Cauchy */
  se $y < \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}$ /* abaixo da densidade de $\mathcal{N}(0,1)$ */
    retorne y

O valor 1.6 foi usado para ajustar a distribuição de Cauchy. Na verdade, sabe-se que o menor valor possível (e portanto o que maximiza a probabilidade de sucesso) é aproximadamente 1.52, mas a demonstração disso fica fora de nosso escopo.

A distribuição de Cauchy funciona de maneira perfeita para este exemplo porque é unimodal (informalmente, tem um único “pico”) e tem a cauda pesada (e justamente por isso não tem esperança) – com isto é fácil fixá-la acima da normal. ●

Há métodos mais eficientes para simular distribuições normais, mas o exposto aqui deve tornar bastante clara a essência da amostragem pelo método da rejeição.

7.3 Distribuições Bivariadas

É comum que em simulações seja necessário gerar distribuições bivariadas. Esta seção apenas toca a superfície deste tópico, apresentando uma maneira de realizar tais simulações.

7.3.1 Variáveis independentes

Se duas variáveis são independentes, podem ser simuladas separadamente. Fizemos isto no Exemplo 7.2.2, ao escolher as duas coordenadas de cada ponto. Naquele exemplo, as duas coordenadas tinham distribuição uniforme. Damos um exemplo onde duas variáveis tem distribuições diferentes.

Exemplo 7.3.1. Queremos gerar uma imagem com pontos aleatórios, formando algo parecido com uma colina de curvatura leve, mais denso na parte de baixo que na de cima.

Verticalmente, a densidade de pontos deve seguir uma distribuição exponencial, com parâmetro $\lambda = 8$. Horizontalmente, deve ser proporcional a uma senóide.

Os valores para a variável X devem estar no intervalo $[0, 1]$. Como posso gerar pontos (x, y) aleatórios obedecendo essa distribuição?

A distribuição horizontal tem densidade $f(x) = a\,\mathrm{sen}(bx)$. Para que seja função de densidade em $[0, 1]$, $a = \pi/2$ e $b = \pi$, e a função de densidade é

$$f_X(x) = \frac{\pi\,\mathrm{sen}(\pi x)}{2}.$$

A função de distribuição acumulada é

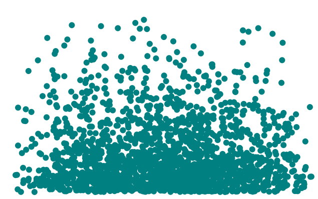
$$F_X(k) = \int_0^k \frac{\pi(\pi x)}{\pi}\,dx = -\frac{\cos(\pi k) - 1}{\pi}.$$

$$F_X^{-1}(x) = \frac{\pi - \arccos(2x - 1)}{\pi}.$$

A distribuição vertical tem densidade $g(x) = 8e^{-8x}$. A distribuição acumulada é $1 - e^{-8y}$, fácil de inverter:

$$F_Y^{-1}(y) = -\frac{\log(1 - y)}{8}$$

As duas distribuições são independentes, portanto geramos os valores de X e Y individualmente!



Nesta figura há dois mil pontos, gerados de acordo com as distribuições expostas. ●

7.3.2 Variáveis dependentes

Suponha que as variáveis aleatórias X e Y tenham distribuição conjunta $f_{X,Y}(x,y)$. Sabemos que

$$f_{X,Y}(x,y) = f_X(x) f_{Y|X}(y|X=x),$$

portanto uma maneira de gerar pares (x,y) de acordo com as distribuições de X e Y é

- gerar um valor x usando a distribuição marginal de X;
- gerar um valor y usando a distribuição condicional $f_{Y|X}$.

Claramente, se as variáveis são independentes podemos simplesmente gerar X e Y separadamente de acordo com suas distribuições, porque $\Pr[y|X=x] = \Pr(y)$. Foi o que fizemos no Exemplo 7.2.2, onde geramos pontos em um quadrado para posteriormente rejeitar os que ficavam fora do triângulo.

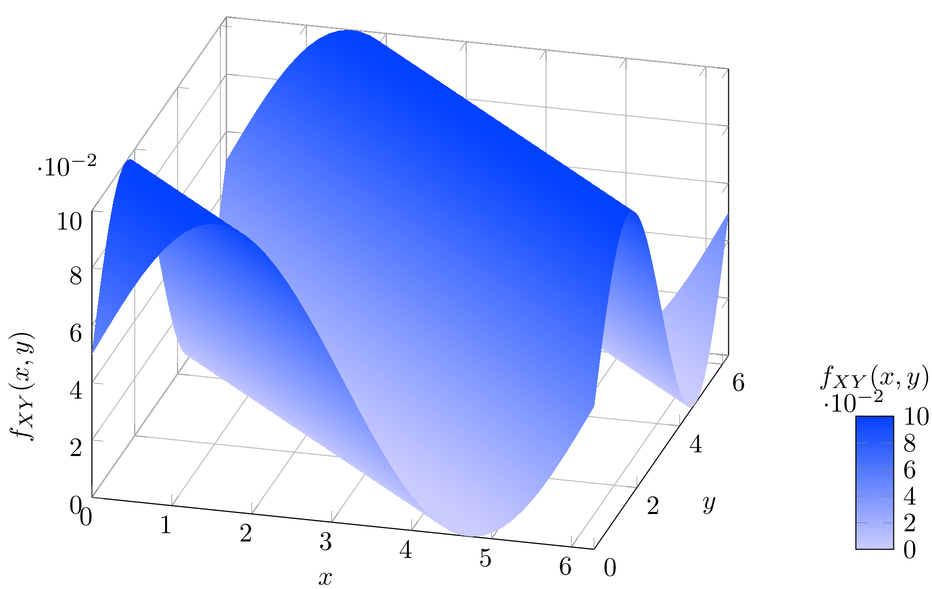
Exemplo 7.3.2. O exemplo 5.4.3 apresenta duas variáveis X e Y, com função de densidade $a\,\text{sen}(x+y)$. Modificaremos esta função para que possamos usá-la no intervalo $[0, 2\pi]$. Como $\text{sen}(x+y)$ nesse intervalo pode chegar a -1, somamos um à função, e calculamos a constante a:

$$f(x,y) = a(1 + \text{sen}(x+y)),$$

Para que $\int_0^{2\pi}\int_0^{2\pi} f(x,y)dydx = 1$, o valor de a deve ser

$$a = 1/4\pi^2.$$

O gráfico da função de densidade é mostrado a seguir.



Para simular a variável, selecionaremos X de acordo com sua distribuição marginal, e depois Y de acordo com a condicional $Y|X$. A marginal de X é

$$\begin{aligned}
F_X(x) &= \int_0^x \int_0^{2\pi} a(1+\operatorname{sen}(x+y))\, dvdu = \frac{x}{2\pi} \quad \text{uniforme!} \\
f_X(x) &= \int_0^{2\pi} a(1+\operatorname{sen}(x+y))\, dv = \frac{1}{2\pi}.
\end{aligned}$$

A distribuição condiconal de Y dado X é

$$\begin{aligned}
g_{Y|X}(x,y) &= \frac{f_{XY}(x,y)}{f_X(x)} \\
&= (a(1+\operatorname{sen}(x+y))2\pi \\
&= \frac{1+\operatorname{sen}(x+y)}{2\pi}
\end{aligned}$$

Podemos gerar X diretamente, porque é uniforme em $[0, 2\pi]$: basta gerar u em $[0, 1]$ e multiplicar por 2π. Para Y, usar o método da inversão tornará o trabalho mais complexo que o necessário para este exemplo, porque teríamos que inverter $(1+\operatorname{sen}(x+y))/2\pi$; ao invés disso usaremos o método da rejeição.

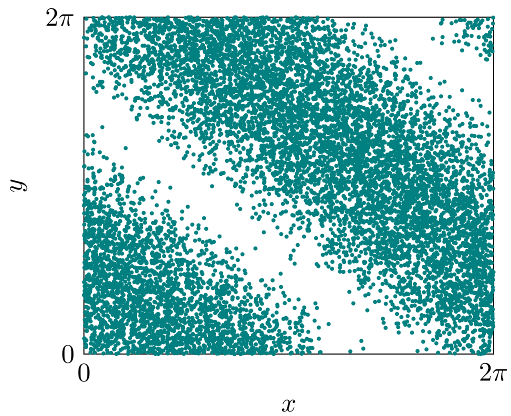
Para podermos gerar os valores, é importante conhecer o mínimo e o máximo da função em que usaremos o método da rejeição:

$$g_{Y|X}(x,y) = \frac{1+\operatorname{sen}(x+y)}{2\pi} \in [0, 0.35]$$

O algoritmo para geração de cada par (x, y) é mostrado a seguir.

gere X em $[0, 2\pi]$
repita:
    gere Y em $[0, 2\pi]$
    gere W em $[0, 0.35]$
    se $W < \dfrac{1+\operatorname{sen}(x+y)}{2\pi}$
        retorne (x, y)

Dez mil pontos obtidos em simulação são mostrados a seguir. Fica bastante clara a relação deste gráfico com o da função de densidade.



7.4 Métodos Específicos

Para algumas distribuições os métodos da inversão e da rejeição podem não ser aplicáveis, ou podem ser ineficientes, e nesses casos pode-se tentar explorar especificidades da distribuição. Esta seção tem o propósito de ilustrar duas situações desse tipo, apresentando algoritmos para simulação de duas distribuições relevantes.

7.4.1 Poisson

Para aplicar o método da inversão em distribuições de Poisson, é necessário calcular a inversa da função acumulada de massa,

$$F(k) = e^{-\lambda} \sum_{i=0}^{\lfloor k \rfloor} \frac{\lambda^i}{i!}.$$

Como a expressão contém um fatorial, haverá dificuldade no cálculo da inversa.

Se observarmos que os tempos entre eventos em um processo de Poisson tem distribuição exponencial, obteremos um método para gerar distribuições de Poisson.

Geramos valores usando uma variável exponencial com parâmetro λ, até que a soma exceda um. Se tomarmos sempre a menor quantidade de variáveis exponenciais somadas com soma menor ou igual a um, a distribuição será de Poisson.

```
soma ← 0
n ← 0
enquanto soma ≤ 1:
    gere ex com distribuição exponencial e parâmetro λ
    soma ← s + ex
```

```
    n  ←  n + 1
retorne n
```

Ou seja, tomamos o menor n tal que

$$\sum_j^n E_j > 1,$$

onde cada E_j é uma variável com distribuição exponencial. Mas isto implica no cálculo de vários logaritmos – o que pode tornar o método lento. Usamos uma propriedade básica de logaritmos para reformular o método: como $\log(a) + \log(b) = \log(ab)$,

$$\begin{aligned}
\sum_j^n E_j &> 1 \\
\sum_j^n \log(U_j) &> -\lambda \\
\log\left(\prod_j^n U_j\right) &> -\lambda \\
\prod_j^n U_j &> e^{-\lambda}.
\end{aligned}$$

Queremos portanto o menor n tal que

$$\prod_j^n U_j > e^{-\lambda}.$$

E o método agora usa somente multiplicações, que são usualmente feitas mais rapidamente que cálculo de logaritmos.

```
prod  ←  1
n ← 0
enquanto prod ≤ e^{-λ}:
    gere u com distribuição uniforme em [0,1]
    prod ← prod * u
    n ← n + 1
retorne n
```

Este método é eficiente quando λ é pequeno; para valores maiores de λ há métodos melhores.

7.4.2 Normal: Box-Müller

A geração de números com distribuição normal usando o método da rejeição pode ser ineficiente, e o método Box-Müller é usualmente aplicado como alternativa mais rápida[1].

Teorema 7.4.1. Sejam U_1 e U_2 variáveis aleatórias com distribuição uniforme. Defina

$$\begin{aligned} X &= \sqrt{-2\log(U_1)}\cos(2\pi U_2), \\ Y &= \sqrt{-2\log(U_1)}\,\mathrm{sen}(2\pi U_2). \end{aligned}$$

Então as duas variáveis, X e Y, tem distribuição normal e são independentes.

Demonstração. Sejam X e Y duas variáveis independentes com distribuição normal padrão. Como são independentes, a função de densidade conjunta das duas é

$$\begin{aligned} f(x,y) &= f_X(x) f_Y(y) \\ &= \frac{1}{\sqrt{2\pi}} e^{-\left(\frac{x^2}{2}\right)} \frac{1}{\sqrt{2\pi}} e^{-\left(\frac{y^2}{2}\right)} \\ &= \frac{1}{2\pi} e^{-\left(\frac{x^2+y^2}{2}\right)}. \end{aligned}$$

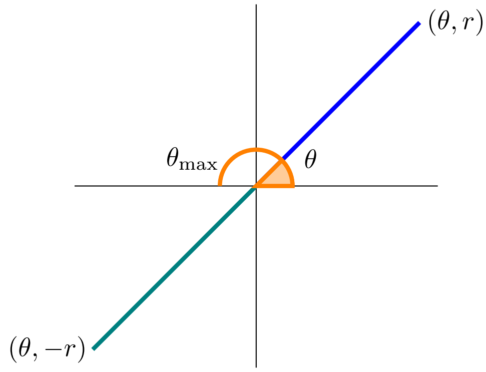
A seguir interpretaremos X e Y como coordenadas no plano: ambas são escolhidas com distribuição normal padrão.

Definimos agora duas outras variáveis independentes, que podem também gerar pontos no plano, mas desta vez usando coordenadas polares:

$$\begin{aligned} \Theta &\sim \mathcal{U}(0,\pi) \\ R^2 &\sim \mathcal{E}xp(1/2) \end{aligned}$$

A variável Θ determina o ângulo, e R^2 determina o quadrado do raio. Definimos Θ entre 0 e π, e *não* entre 0 e 2π. Isso porque R^2 é o *quadrado* do raio, e estamos portanto admitindo raios negativos – isso é ilustrado na figura a seguir.

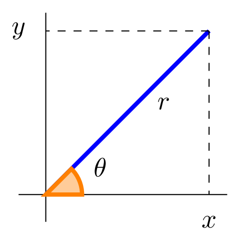
[1] Há variantes mais eficientes do método de Box-Müller, mas o detalhamento delas foge do nosso propósito.



Como Θ e R^2 são independentes, a distribuição conjunta delas é

$$
\begin{aligned}
f_{R^2,\Theta} &= \left(\frac{1}{\pi}\right)\left(\frac{1}{2}e^{-\left[\frac{r^2}{2}\right]}\right) \\
&= \frac{1}{2\pi}e^{-\left(\frac{r^2}{2}\right)}
\end{aligned}
$$

Aogra estabeleceremos uma relação entre os dois pares de variáveis, (X, Y) e (R, Θ).



Como $x^2 + y^2 = r^2$, as distribuições conjuntas de X, Y e Θ, R^2 são iguais. Isso significa que podemos gerar pares Θ, R e transformá-los em X, Y.

Sabemos que podemos gerar uma distribuição exponencial usando a uniforme, portanto se U é uma variável uniforme no intervalo $[0, 1]$, então

$$
\begin{aligned}
R^2 &\sim \frac{-\log(U)}{1/2} \\
R &\sim \sqrt{-2\log(U)}.
\end{aligned}
$$

Assim, podemos descrever X e Y a partir de duas outras variáveis aleatórias, R (o raio) e Θ (o ângulo). Mas como determinamos r como somente a raiz positiva de $-2\log(U)$, passamos a ter raios somente positivos. Assim, mudamos Θ para uniforme em $[0, 2\pi]$. Isso não muda a distribuição dos pontos, porque continuamos com ângulo uniforme (e como a distribuição dos pontos não muda, a de suas coordenadas também não mudará).

Escrevemos as coordenadas cartesianas X e Y em função de R e Θ:

$$X = R\cos(\Theta),$$
$$Y = R\,\mathrm{sen}(\Theta).$$

Finalmente, reescrevemos R e Θ usando U_1 e U_2:

$$X = \sqrt{-2\log(U_1)}\cos(2\pi U_2),$$
$$Y = \sqrt{-2\log(U_1)}\,\mathrm{sen}(2\pi U_2).$$

As variáveis X e Y são independentes com distribuição $\mathcal{N}(0,1)$, conforme determinamos no início da demonstração. ■

Exercícios

Ex. 152 — Se usarmos o método da rejeição para obter pontos uniformemente distribuídos em uma circunferência, circunscrevendo-a em um quadrado (como fizemos com o triângulo no exemplo), qual será a probabilidade de obtermos

i) um ponto dentro da circunferência após 3 tentativas?

ii) um ponto dentro da circunferência após k tentativas?

Ex. 153 — Crie um gerador para distribuição do seno, descrita na Seção 4.7, implemente-o e use seu simulador para verificar os valores calculados naquela seção.

Ex. 154 — Um evento a ser simulado deve ocorrer de acordo com uma variável aleatória X com função de densidade de probabilidade proporcional a x^5, e com $X \in [0,1]$. Assim,

$$f(x) = kx^5.$$

Mostre como usar o método da inversão para simular a variável.

Ex. 155 — Mostre como simular uma variável aleatória com distribuição geométrica.

Ex. 156 — Seja X uma variável aleatória tendo distribuição exponencial com parâmetro λ. Prove que $\lfloor X \rfloor$ tem distribuição geométrica.

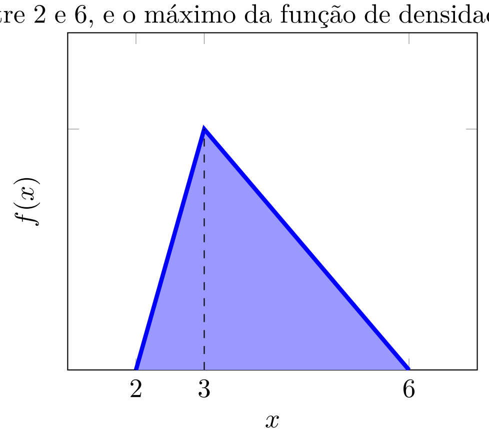
Ex. 157 — Mostre como simular uma variável aleatória com distribuição de Cauchy.

Ex. 158 — Uma variável aleatória contínua X tem função de densidade de probabilidade

$$f(x) = \begin{cases} k(1+x) & x \in (-1,0) \\ k(1-x) & x \in (0,1) \\ 0 & x \notin (-1,1). \end{cases}$$

Determine k, mostre a função de densidade acumulada, e mostre como simular X.

Ex. 159 — Uma variável X tem distribuição triangular, sendo que X pode assumir valores entre 2 e 6, e o máximo da função de densidade se dá em $x = 3$.



Mostre como simular X usando o método da inversão e o da rejeição

Ex. 160 — Mostre como simular o experimento da agulha de Buffon. Implemente sua simulação e verifique as frequências no seu experimento, comparando com os valores obtidos no estudo teórico.

Ex. 161 — Mostre como gerar uma distribuição exponencial com média $\lambda > 0$ truncada em um intervalo $[a, b]$, a partir de uma variável aleatória U uniformemente distribuída em $(0, 1)$.

Ex. 162 — No Exemplo 7.3.1 modelamos a função de densidade de probabilidade de uma distribuição como $f(x) = a \operatorname{sen}(bx)$. Porque é necessário usar duas constantes, a e b? Não seria possível usar somente uma delas?

Ex. 163 — Construa um experimento para aproximar o valor de π a partir de uma simulação.

Ex. 164 — Mostre como simular as distribuições bivariadas com funções de densidade

1. $f(x, y) = ae^{-x-2y}$,
2. $f(x, y) = a \operatorname{sen}(x)e^{-y}$.
   Nos dois casos, $x, y \in [0, 1]$.

Ex. 165 — No Exemplo 7.2.3 tentamos várias vezes obter um ponto abaixo da curva $f(z)$. Qual é a probabilidade de sucesso em cada tentantiva? Qual é a esperança para a quantidade de tentativas até que um ponto seja obtido?

Ex. 166 — No Exemplo 7.3.2 a distribuição marginal de X é uniforme. No entanto, o gráfico da função de densidade e o gráfico com os pontos simulados mostram regiões de densidade muito diferentes. Explique.

Ex. 167 — Seja

$$f(x, y, z) = a\,\mathrm{sen}(x)ye^{-z}$$

uma função conjunta de densidade de probabilidade para variáveis X, Y e Z.

1. Determine a.
2. Mostre como simular as variáveis X, Y e Z.

Ex. 168 — Planeje e execute um experimento que simule o problema de Monty Hall, descrito na Seção 2.4.

Ex. 169 — Simule o protocolo ALOHA, descrito na Seção3.7.3 (página 77), e confirme experimentalmente o resultado obtido naquela Seção.

Ex. 170 — Complemente o Exercício 168 simulando os experimentos descritos no Exercício 41 (na página 52).

Ex. 171 — Sejam A e B duas variáveis independentes com distribuição $\mathcal{N}(0, 1)$. Prove que $A + B$ é independente de $A - B$, e que tanto $A + B$ como $A - B$ tem distribuição $\mathcal{N}(0, 2)$.

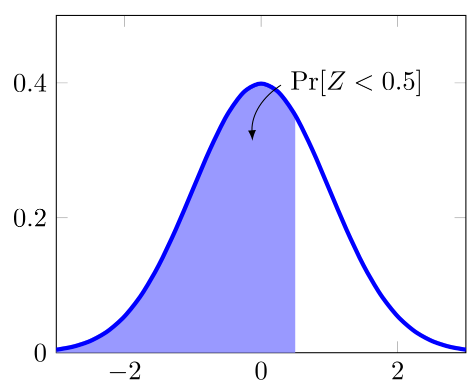
Ex. 172 — Mostre como simular uma variável aleatória com distribuição normal a partir de váris outras com distribuição uniforme, sem usar logaritmos ou funções trigonométricas.

Apêndice A

Função de distribuição da normal padrão

A tabela neste Apêndice traz valores da função de distribuição para $\mathcal{N}(0,1)$,

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}}\, e^{\frac{-y^2}{2}} dy.$$



Para obter $\Pr[Z < k]$, sendo k a soma dos índices de linha e coluna. Por exemplo, o valor de $\Pr[Z < 1.83]$ está na linha 1.8 e coluna 0.03, porque $1.83 = 1.8 + 0.03$, portanto $\Pr[Z < 1.83] = 0.9664$.

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

Apêndice B

Dicas e Respostas

Resp. (Ex. 1) — Não! A seguir há um contraexemplo: sejam

$$\begin{aligned}
\Omega &= \{1,2,3,4,5\},\\
A &= \{1,2\},\\
B &= \{3,4\}.
\end{aligned}$$

Claramente,

$$A \cap B = \emptyset.$$

No entanto,

$$\begin{aligned}
\overline{A} \cap \overline{B} &= \overline{\{1,2\}} \cap \overline{\{3,4\}}\\
&= \{3,4,5\} \cap \{1,2,5\}\\
&= \{5\}.
\end{aligned}$$

Resp. (Ex. 5) —

1. 3/1000.
2. $(1/1000)^3$.
3. 7/8.
4. 19/27.

Resp. (Ex. 6) — Não! As respostas do exercício mostram isso: $\Pr[k = 0] \neq \Pr[k = 1]$, por exemplo.

Resp. (Ex. 7) — (1) 5/8. (2) 65/81.

Resp. (Ex. 8) — 7/9

Resp. (Ex. 9) —

1. Há $\binom{20}{2}$ possibilidades de escolha de dois produtos. Mas como dois deles tem defetios, há $\binom{18}{2}$ escolhas de produtos sem defeitos. Assim, a probabilidade é
$$\frac{\binom{18}{2}}{\binom{20}{2}}.$$
2. Pede-se a probabilidade de "pelo menos um" com defeito – o que é a negação de "nenhum deles com defeito". Assim, o probabilidade buscada é
$$1-\frac{\binom{18}{2}}{\binom{20}{2}}.$$
3. Só há uma possibilidade: a de escolhermos exatamente os dois com defeito. Assim, a probabilidade é
$$\frac{1}{\binom{20}{2}}.$$

Resp. (Ex. 10) — São oito torres que não podem compartilhar linha ecoluna; o tabuleiro tem tamanho 8×8. Suponha que as torres sejam postas uma por vez, começando pela primeira linha e continuando linha por linha. A primeira torre pode ser posta em qualquer coluna, logo há oito possibilidades para sua posição. A segunda pode ser posta em sete posições, porque não pode ocupar a mesma coluna que a primeira. Prosseguindo desta forma, há $8 \cdot 7 \cdot 6 \cdots 1 = 8!$ possíveis configurações onde as torres não se atacam. O espaço amostral contém todas as possíveis configurações de oito torres, atacando-se ou não – ou seja, todas as combinações de oito posições das 64 possíveis,
$$|\Omega| = \binom{64}{8}$$
Assim,
$$\Pr(X) = \frac{8!}{\binom{64}{8}}.$$

Resp. (Ex. 12) —

1. Há 2^{128} chaves possíveis, logo a probabilidade é $\frac{1}{2^128}$, ou 2^{-128}.

2. Quando a chave tem n bits, há 2^n chaves, e a probabilidade de um ataque com sucesso é $1/2^n$, logo

$$f(n) = \frac{1}{2^n}.$$

Resp. (Ex. 15) —

$$\frac{\binom{m-1}{k-1}}{\binom{n}{k}}$$

Resp. (Ex. 16) —

$$\frac{1}{10} + \left(\frac{9}{10}\right)\left(\frac{1}{9}\right) + \left(\frac{8}{10}\right)\left(\frac{1}{8}\right) = \frac{3}{10}$$

Resp. (Ex. 17) — Os dígitos menos significativos de 3^k e 7^k repetem o padrão $3, 9, 7, 1$:

k	3^k	7^k
1	3	7
2	9	49
3	27	343
4	81	2401
5	243	16807
6	729	117649
7	2187	823543
8	6561	5764801
⋮	⋮	⋮

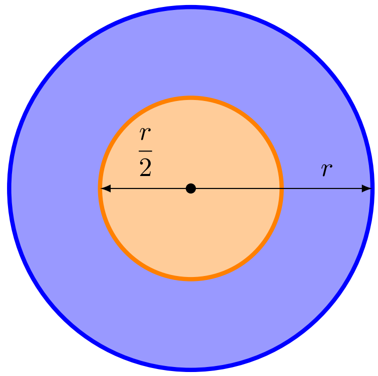
Como 4 divide 100, qualquer um dos quatro é tão provável quanto os outros quando selecionamos a e b.
Há, portanto, 16 pares possíveis – e equiprováveis:

$$\begin{array}{cccc} (3,3) & (9,3) & (7,3) & (1,3) \\ (3,9) & (9,9) & (7,9) & (1,9) \\ (3,7) & (9,7) & (7,7) & (1,7) \\ (3,1) & (9,1) & (7,1) & (1,1) \end{array}$$

Somente 3 delas tem soma terminando em oito: $(7,1), (1,7), (9,9)$. Assim, a probabilidade buscada é

$$\frac{3}{16}.$$

Resp. (Ex. 18) — 1. O ponto deve estar dentro da cicunferência com raio $r/2$.



Assim, se X é o evento em que o ponto fica dentro da circunferencia menor (distante a menos de $r/2$ do centro), então

$$\begin{aligned}
\Pr[X] &= \frac{\pi \left(\frac{r}{2}\right)^2}{\pi r^2} \\
&= \frac{r^2}{2^2}\frac{1}{r^2} \\
&= \frac{1}{4}.
\end{aligned}$$

2. $1/8$.

Resp. (Ex. 21) — Aproximadamente 0.9375

Resp. (Ex. 22) — $\pi/6$.

Resp. (Ex. 23) — ≈ 0.52

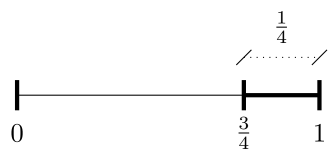
Resp. (Ex. 25) —

$$\begin{aligned}
A &= \int_0^\alpha x^2 \, dx = \frac{\alpha^3}{3} \\
B &= \int_0^\alpha x - x^2 \, dx = \frac{\alpha^2}{2} - \frac{\alpha^3}{3}
\end{aligned}$$

$$\begin{aligned}\Pr[A > B] &= \Pr\left[\frac{\alpha^3}{3} > \frac{\alpha^2}{2} - \frac{\alpha^3}{3}\right] \\ &= \Pr\left[\alpha > \frac{3}{4}\right].\end{aligned}$$

Mas como α é escolhido no intervalo $[0,1]$, a probabilidade de α ser maior que $3/4$ é

$$\Pr[A > B] = \Pr[\alpha > 3/4] = \frac{1}{4}.$$



Resp. (Ex. 26) — (Dica apenas) Comece observando que $\Pr(B|A) = \frac{\Pr(A\cap B)}{\Pr(A)}$. Isole $\Pr(A \cap B)$, e continue.

Resp. (Ex. 28) — Segunda dica: divida o intervalo em n partições e calcule a probabilidade usando um ponto de cada partição. Depois verifique o que acontece quando $n \to \infty$.

Resp. (Ex. 30) — 1. 1/2

2. 3/8
3. 0
4. 1/4
5. 1/8

Resp. (Ex. 31) — 1. 3/8.

2. 1/3.
3. Bayes.

Resp. (Ex. 32) — Damos nomes aos eventos:

- B: recebe bolsa-auxílio;
- E: vem de outro estado.

O enunciado nos deu

$$\begin{aligned} \Pr(B) &= 0.4 \\ \Pr(E|B) &= 0.1. \end{aligned}$$

Basta usarmos a definição de probabilidade condicional:

$$\begin{aligned} \Pr(E|B) &= \frac{\Pr(E \cap B)}{\Pr(B)} \\ \Pr(E|B)\Pr(B) &= \Pr(E \cap B), \end{aligned}$$

portanto

$$\Pr(E \cap B) = (0.1)(0.4) = 0.04.$$

Resp. (Ex. 33) — $\Omega = \{AAA, AAO, AOA, AOO, OAA, OAO, OOA, OOO\}$, logo $|\Omega| = 8$.

- No mínimo uma cara, $X_1 = \Omega - \{OOO\}$, e $\Pr[X_1] = 7/8$.
- No mínimo duas caras, $X_2 = \{AAO, AOA, OAA, AAA\}$, e $\Pr[X_2] = 1/2$.
- $X_1 \cap X_2 = X_2$.

Pela definição de probabilidade condicional,

$$\begin{aligned} \Pr[X_2|X_1] &= \frac{\Pr[X_1 \cap X_2]}{\Pr X_1} \\ &= \frac{4}{7}. \end{aligned}$$

Resp. (Ex. 34) — Obviamente, $\Pr(A) = \Pr(B) = 1/2$. Calcule $\Pr[\text{cereja} \cap \text{uva}|A]$, $\Pr[\text{cereja} \cap \text{uva}|B]$, e use o Teorema de Bayes.

Resp. (Ex. 35) — Sejam B_1 e B_2 as duas bolas selecionadas. É simples obter $\Pr[B_2 = b|B_1 = v]$, mas o exercício pede $\Pr[B_1 = v|B_2 = b]$. Típica aplicação do Teorema de Bayes.

Resp. (Ex. 39) — O espaço amostral poderia ser $\Omega = \{+,-\} \times \{+,-\}$, composto portanto de pares ordenados. Ou seja, $\Omega = \{(-,-), (-,+), (+,-), (+,+)\}$, onde cada resultado tem o significado como descrito a seguir.

- $(-,-)$ paciente sem tumor, resultado negativo;
- $(-,+)$ paciente sem tumor, resultado positivo (falso positivo);
- $(+,-)$ paciente com tumor, resultado negativo (falso negativo);

- $(+,+)$ paciente com tumor, resultado positivo.

Os resultados, no entanto, *não* são equiprováveis (as probabilidades foram dadas no texto do exemplo).

Resp. (Ex. 40) — 1. $\Pr(A|B) = 0 \leq \Pr(A)$.

2. $\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)} = \Pr(A)/\Pr(B) \geq \Pr(A)$, porque $0 \leq \Pr(B) \leq 1$.
3. $\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)} = \Pr(B)/\Pr(B) = 1 \geq \Pr(A)$.

Resp. (Ex. 42) — É simples, $\text{var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$, e $\text{var}(X) \geq 0$. A igualdade ocorre quando $\text{var}(X) = 0$.

Resp. (Ex. 45) —

$$\begin{aligned}
\mathbb{E}\left[\cos\left(\frac{\pi X}{10}\right)\right] &= \cos\left(\frac{0\pi}{10}\right)\left(\frac{1}{2}\right) + \cos\left(\frac{10\pi}{10}\right)\left(\frac{1}{2}\right) \\
&= (1)\left(\frac{1}{2}\right) + (-1)\left(\frac{1}{2}\right) \\
&= 0.
\end{aligned}$$

Resp. (Ex. 49) — ≈ 0.0158.

Resp. (Ex. 50) — 5 e $\sqrt{20} \approx 4.4721$.

Resp. (Ex. 51) —

$$\begin{aligned}
&1 - \Pr[D \leq 3] \\
&= 1 - \sum_{j=0}^{3} \binom{400}{j} (0.05)^j (0.95)^{400-j}
\end{aligned}$$

Resp. (Ex. 52) — $(1000)(10/12)(2/12)^{999}$.

Resp. (Ex. 54) — Presuma Poisson. A média de erros por duas páginas é 6, logo a probabilidade de duas páginas sem erros é dada por

$$\begin{aligned}\frac{e^{-\mu}\mu^k}{k!} &= \frac{e^{-6}6^0}{0!} \\ &= e^{-6} \approx 0.0024787.\end{aligned}$$

Resp. (Ex. 55) — ≈ 0.2238

Resp. (Ex. 58) —

1. O número de acertos segue distribuição binomial. A esperança é

$$\mathbb{E}[X] = np = 20\frac{6}{10} = 12.$$

A quantidade de pontos é $X-10$. A quantidade que pagamos é 100 vezes o número de pontos, logo $100(X-10)$. Como a esperança é linear,

$$\begin{aligned}\mathbb{E}[100(X-10)] &= 100(\mathbb{E}[X]-10) \\ &= 100\,(12-10) \\ &= 200.\end{aligned}$$

2. Novamente, binomial!

$$\Pr[k=0] = \binom{20}{0}(0.4)^{20}(0.6)^0 = (0.4)^{20},$$

algo próximo de 10^{-8}.

3. Soma de geométricas.

$$\sum_{j=10}^{20}\Pr[k=j] = \sum_{j=10}^{20}(0.4)^j(0.6)^{20-j}$$

4. Binomial,

$$\Pr[k=10] = \binom{20}{10}(0.4)^{10}(0.6)^{10}$$

5. Binomial.

$$\begin{aligned}\Pr[k>10] &= \sum_{j=11}^{20}\Pr[k=j] \\ &= \binom{20}{11}\sum_{j=11}^{20}(0.4)^j(0.6)^{20-j}.\end{aligned}$$

Resp. (Ex. 59) — Em cada rodada,

$$\begin{aligned} p_A &= \Pr[A \text{ obtém soma } = 6] = \frac{5}{36} \\ q_A &= \Pr[A \text{ obtém soma } \neq 6] = \frac{31}{36} \\ p_B &= \Pr[A \text{ obtém soma } = 7] = \frac{6}{36} \\ q_B &= \Pr[A \text{ obtém soma } \neq 7] = \frac{30}{36} \end{aligned}$$

E a probabilidade de nem A nem B ganharem é $q_A q_B$. A de A ganhar é p_A. Assim, a probabilidade de A vencer <u>*exatamente*</u> na n-ésima rodada é

$$(q_A q_B)^n p_A$$

(uma geométrica).

A probabilidade de A vencer em qualquer *em qualquer rodada* é a soma, para todo n, da probabilidade de vencer na n-ésima:

$$\begin{aligned} \Pr[A \text{ ganhar }] &= \sum_{n=0}^{\infty} q_A^n q_B^n p_A \\ &= p_A \sum_{n=0}^{\infty} (q_A q_B)^n \\ &= p_A \left(\frac{1}{1 - q_A q_B} \right) && \text{(porque } q_A q_B < 1\text{)} \\ &= \frac{p_A}{1 - q_A q_B}. \end{aligned}$$

Resp. (Ex. 60) — 1. $-\log(3/10) \approx 1.2039$.

2. $\Pr[X > 2] = 1 - \Pr[X \leq 2] \approx 0.4213$.

Resp. (Ex. 61) — A média é λ, e o enunciado dá $\Pr[X = 0] = 0.15$.

$$\Pr[X = 0] = \frac{\lambda^0 e^- \lambda}{0!} = e^{-\lambda},$$

logo

$$\begin{aligned}
e^{-\lambda} &= 0.15 \\
e^{-\lambda} &= \frac{15}{100} \\
\frac{20}{3} &= e^{\lambda} \\
\log\left(\frac{20}{3}\right) &= \lambda \\
\log(6.6666) &= \lambda \\
1.8971 &\approx \lambda.
\end{aligned}$$

Mas esta é a média *por amostra de* $10ml$, logo a média por ml deve ser aproximadamente 0.18971.

Resp. (Ex. 64) — 1.

$$\sum_{n=1}^{\infty} \frac{1}{e^n} = \frac{1}{e-1},$$

que está no intervalo $[0, 1]$, portanto basta definir

$$\Pr(0) = 1 - \frac{1}{e-1}.$$

Resp. (Ex. 65) — 1.

2.

3.

$$\sum_{j=0}^{\infty} \left(\frac{e^{-2}2^j}{j!}\right)^2$$

Resp. (Ex. 66) — 1. Se a variável A, que contabiliza a quantidade de amendoins, seguisse Poisson,

$$\Pr[A = 0] = e^{-10} \approx \frac{1}{22026.46}.$$

2.

$$\Pr[A = 6] = \frac{10^6 e^{-10}}{6!} \approx 0.06,$$

não tão pequena. No entanto, como depois de um milhão de barras não houve sequer uma com menos de sete amendoins, o modelo usado (distribuição de Poisson) não parece adequado. Será mais útil, provavelmente, modelar essa quantidade como seguindo distribuição normal.

Resp. (Ex. 67) — A demonstração é quase a mesma do Teorema 3.8.7.

Resp. (Ex. 68) — Por indução.

Resp. (Ex. 69) —

$$k = \frac{4}{4e+1}$$

Resp. (Ex. 70) — A integral da função deve ser um. Verificamos:

$$\begin{aligned}
\int_0^1 f(x)dx &= \int_0^1 \frac{1}{\pi\sqrt{x(1-x)}}dx \\
&= \frac{\operatorname{arcsen}(2x-1)}{\pi}\Big|_0^1 \\
&= \frac{\operatorname{arcsen}(1)}{\pi} - \frac{\operatorname{arcsen}(-1)}{\pi} \\
&= \frac{(\pi/2)-(-\pi/2)}{\pi} \\
&= \frac{\pi}{\pi} \\
&= 1.
\end{aligned}$$

Resp. (Ex. 71) — $k = 1$, $f(x) = \sqrt{x}/2x^2$.

Resp. (Ex. 73) — $F(x) = \cos(x)$. Derive, $f(x) = \operatorname{sen}(x)$. A integral deve ser igual a um,

$$\int_a^b \operatorname{sen}(x)dx = 1.$$

Isso acontece quando

$$\begin{aligned}
b &= (4k+4)\pi, \\
a &= \left(4k + \frac{3}{2}\right)\pi,
\end{aligned}$$

com $k \in \mathbb{Z}$.

Resp. (Ex. 75) — $f(x)$ *não* é função de densidade!

$$\int_0^1 3xdx = \frac{3}{2} \neq 1.$$

Resp. (Ex. 76) — 1. 1/2
2. 1/3
3. 1/5

Resp. (Ex. 78) — 1. Usando

$$Z = \frac{X-\mu}{\sigma},$$

$$\begin{aligned}
\Pr[749 < C < 780] &= \Pr\left[\frac{749-759}{20} < Z < \frac{780-759}{20}\right] \\
&= \Pr\left[-\frac{1}{2} < Z < \frac{21}{20}\right] \\
&= \Pr[-0.5 < Z < 1.05] \\
&= \Pr[Z < 1.05] - \Pr[Z < -0.5] \\
&= \Pr[Z < 1.05] - \big(1 - \Pr[Z < 0.5]\big) \\
&= 0.8531 - (1 - 0.6915) \\
&= 0.8531 - 0.3085 \\
&= 0.5446.
\end{aligned}$$

Resp. (Ex. 79) — (Dica apenas) O exercício dá $\Pr[T > 4] = 0.02$, $\sigma = 0.3$, e pede μ.

Resp. (Ex. 80) — Se $\Pr[3 < X < 3.8] = 0.3289$, então

$$\Pr[X < 3.8] = 0.8289. \qquad \text{(porque?)}$$

Da tabela de valores de Φ, temos

$$\Phi(0.95) = 0.8289.$$

Queremos, portanto, que a variável X, *quando transformada em normal padrão*, seja tal que

$$\Pr[Z < 0.95] = 0.8289.$$

Ou seja,

$$\begin{aligned}
\Pr\left[\frac{X-3}{\sigma} < \frac{3.8-3}{\sigma}\right] &= 0.8289 \\
\Pr\left[Z < \frac{3.8-3}{\sigma}\right] &= 0.8289.
\end{aligned}$$

Mas como $\Phi(0.95) = 0.8289$, isso implica que

$$\begin{aligned}\frac{3.8-3}{\sigma} &= 0.95\\ \sigma &= \frac{0.8}{0.95}\\ &= 0.8421.\end{aligned}$$

Resp. (Ex. 82) — 1. a integral na reta é $(2\sigma\sqrt{\pi})^{-1}$, diferente de 1.

2. K é o inverso do valor da integral calculada no item (i):

$$K = 2\sigma\sqrt{\pi}$$

3. média

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f(x) dx = \mu$$

Ou, pode-se argumentar que, como a função é não-negativa e simétrica ao redor da média, ao elevá-la ao quadrado tanto $f(\mu+z)$ como $f(\mu-z)$, que eram iguais, continuarão a ter o mesmo valor, $f(\mu+z)^2 = f(\mu-z)^2$. Como os valores da densidade ao redor da média continuam simétricos, a média continua a mesma.

4. variância

$$\begin{aligned}\mathrm{var}[X] &= \mathbb{E}[X^2] - \mathbb{E}[X]^2\\ &= \int_{-\infty}^{\infty} x^2 f(x) dx - \mu^2\\ &= \frac{\sigma^2}{2}.\end{aligned}$$

5. a média permanece μ. a variância será σ^2/q.
6. não! (sabendo quais são média e variância, analise o expoente de e em f – não é normal)

Resp. (Ex. 83) — $\int_{-\infty}^{+\infty} f(x)dx$ deve ser um. Como as raízes são

$$\begin{aligned}r_1 &= \frac{-\sqrt{-4(-4)c}}{2} = -\frac{\sqrt{c}}{2}\\ r_2 &= \frac{+\sqrt{-4(-4)c}}{2} = +\frac{\sqrt{c}}{2},\end{aligned}$$

então

$$\begin{aligned}
\int_{-\infty}^{+\infty} f(x)dx &= \int_{r_1}^{r_2} -4x^2 + cdx \\
&= \left(xc - \frac{4x^3}{3}\right)\Big|_{r_1}^{r_2} \\
&= \frac{2}{3}c\sqrt{c}.
\end{aligned}$$

Igualamos

$$\begin{aligned}
\frac{2}{3}c\sqrt{c} &= 1 \\
c\sqrt{c} &= \frac{3}{2} \\
c^3 &= \frac{3^2}{2^2} \\
c &= \sqrt[3]{\frac{9}{4}}.
\end{aligned}$$

As outras duas raízes de $c^3 = \frac{9}{4}$ não são reais, uma vez que uma equação cúbica $ax^3 + bx^2 + cx + d = 0$, com $b = c = 0$, tem uma única raiz real.

Resp. (Ex. 84) — Sim, porque $\int_{-\infty}^{\infty} f(x) = 1$.
A esperança é $\int_{-\infty}^{\infty} xf(x) = 2$.

Resp. (Ex. 85) — 1. Primeiro observamos que, ao escolher um ponto na esfera, os outros pontos do cubo ficam determinados. O experimento, portanto, consiste em escolher *um* vértice na superfície da esfera, de maneira equiprovável.
Suponha então que $X_1, X_2, \ldots, X_8$ são variáveis aleatórias que determinam quando um vértice está em parte pintada:

$$X_i = \begin{cases} 1 & i\text{-ésimo vértice em parte } \textit{não} \text{ pintada} \\ 0 & i\text{-ésimo vértice em parte pintada} \end{cases}$$

A esperança de cada X_i é

$$\mathbb{E}[X_i] = (1)\frac{9}{10} + (0)\frac{1}{10}.$$

A esperança dos oito X_i é

$$\begin{aligned}
\mathbb{E}\left[\sum_{i=1}^{8} X_i\right] &= \sum_{i=1}^{8} \mathbb{E}[X_i] \\
&= 8\mathbb{E}[X_i] \\
&= 7.2.
\end{aligned}$$

Como a esperança é 7.2, então existe um ponto p tal que $\sum_{i=1}^{8} X_i(p) > 7.2$. Mas a soma dos X_i é inteira, logo existe um ponto p na esfera tal que $X_i(p) \geq 8$.

Resp. (Ex. 86) —

$$\begin{aligned}
\Pr[60 < D < 120] &= \Pr[D < 120] - \Pr[D < 60] \\
&= \left(1 - e^{-120/90}\right) - \left(1 - e^{-60/90}\right) \\
&= \left(1 - e^{-4/3}\right) - \left(1 - e^{-2/3}\right) \\
&= e^{-2/3} - e^{-4/3} \\
&\approx 0.513 - 0.263 \\
&\approx 0.250.
\end{aligned}$$

Resp. (Ex. 87) — Primeiro,

$$\lambda = \frac{1}{4} \text{ falhas/ano.}$$

7/10 das lâmpadas deve continuar funcionando, logo a probabilidade de lâmpadas continuarem funcionando deve ser de 7/10.

$$\begin{aligned}
\Pr[T > t] &= \frac{7}{10} \\
e^{-t/4} &= \frac{7}{10} \\
-\frac{t}{4} &= \log\left(\frac{7}{10}\right) \\
t &= -4\log\left(\frac{7}{10}\right) \\
t &\approx 1.427 \text{ anos.}
\end{aligned}$$

Resp. (Ex. 89) — $\lambda = 1/24$. Use a propriedade "sem memória" da exponencial.

$$\Pr[T > 36 | T > 24] \approx 0.6065.$$

Resp. (Ex. 90) — 1.

$$\Pr[T < 10] = 1 - e^{-10/20} = 1 - e^{1/2} \approx 0.395.$$

2.

$$\begin{aligned}
\Pr[T < t] &= 0.5 \\
1 - e^{-t/20} &= 0.5 \\
e^{-t/20} &= 0.5 \\
-\frac{t}{20} &= \log\left(\frac{1}{2}\right) \\
t &= -20\log\left(\frac{1}{2}\right) \\
&\approx 13.86 \text{ dias.}
\end{aligned}$$

Resp. (Ex. 92) — Um ponto de partida é

$$\mathbb{E}[X^2] = \int_0^\infty x^2 \lambda e^{-\lambda x} dx.$$

Integre duas vezes por partes, obtendo

$$\mathbb{E}[X^2] = \frac{2}{\lambda^2}$$

Depois, sabendo que $\mathbb{E}[X]^2 = \frac{1}{\lambda^2}$, use $\text{var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$.

Resp. (Ex. 94) — Se A tem distribuição exponencial, sua função de densidade é

$$\lambda e^{-\lambda x}, \quad x \geq 0.$$

A função de densidade acumulada de R é

$$\begin{aligned}
G(r) &= \Pr[R \leq r] \\
&= \Pr[A \leq \pi r^2] \\
&= 1 - e^{-\lambda \pi r^2}.
\end{aligned}$$

Assim,

$$\begin{aligned}
g(r) &= \frac{d}{dx} 1 - e^{-\lambda \pi r^2} \\
&= 2\pi\lambda r e^{-\lambda \pi r^2}.
\end{aligned}$$

Resp. (Ex. 96) — Veja o Exemplo 4.5.9.

Resp. (Ex. 97) — Não é pequena!

$$\lambda = \frac{1}{100000}$$

$$\begin{aligned}\Pr[T < 1000000] &= 1 - e^{-100000/100000} \\ &= 1 - e^{-1} \\ &\approx 0.6321.\end{aligned}$$

Resp. (Ex. 98) —
1. 0.1587
2. 0.0228
3. 0.0013

Resp. (Ex. 99) — 0.1056

Resp. (Ex. 101) — 28.86

Resp. (Ex. 102) — É o mesmo que determinar

$$\begin{aligned}\Pr[-3 \leq X \leq 3] &\approx \Phi(0.33) - \Phi(-0.66) \\ &\approx 0.6293 - (1 - \phi(0.66)) \\ &\approx 0.6293 - (1 - 0.7454) \\ &\approx 0.3747.\end{aligned}$$

Resp. (Ex. 103) — Se $X \sim \mathcal{N}(8, 9)$, então $(X - 8)/3 \sim \mathcal{N}(0, 1)$.

$$\begin{aligned}\Pr[X < k] &= \Pr\left[\frac{X-8}{3} < \frac{k-8}{3}\right] \\ &= \Phi\left(\frac{k-8}{3}\right).\end{aligned}$$

Consequentemente,

$$\Pr[X > k] = 1 - \Phi\left(\frac{k-8}{3}\right).$$

O enunciado pede $\Pr[X < k] = 4\Pr[X > k]$. Fazemos $u = \Pr[X < k]$, e

$1 - u = \Pr[X > k]$. Então

$$\begin{aligned} u &= 4(1-u) \\ u &= 4 - 4u \\ 5u &= 4 \\ u &= \frac{4}{5}. \end{aligned}$$

A resposta é tal que

$$\Pr[X < k] = u = \frac{4}{5} = 0.8.$$

Como $\Phi(0.84) \approx 0.7995$ é o mais próximo que temos de 0.8, então determinamos que

$$\begin{aligned} \frac{k-8}{3} &\approx 0.84 \\ k - 8 &\approx 3(0.84) \\ k &\approx 10.52. \end{aligned}$$

Verifique:

$$\begin{aligned} \Pr[X < 10.52] &\approx 0.7995 \\ \Pr[X > 10.52] &\approx 0.2005. \end{aligned}$$

Temos $4(0.2005) = 0.8020$, com erro de 0.0025, devido ao uso da tabela de Φ ao invés de valores exatos – o que é razoável.

Resp. (Ex. 104) — Mude para coordenadas polares e mostre que

$$\left(\int_{-\infty}^{\infty} e^{\left(\frac{-x^2}{2}\right)} dx \right)^2 = 2\pi.$$

Resp. (Ex. 106) —

$$\begin{aligned} f &= \frac{1}{y\pi \left[1 + \left(\frac{x - x_0}{y} \right)^2 \right]} \\ F(x) &= \frac{1}{2} + \frac{1}{\pi} \arctan \left(\frac{x - x_0}{y} \right) \end{aligned}$$

Resp. (Ex. 109) — 1. Definimos duas variáveis aleatórias,

- X = quantidade de bolas vermelhas;
- Y = quantidade de bolas brancas.

A seguir calculamos cada uma das probabilidades $p(x, y) = \Pr[X = x, Y = y]$. Como exemplo, para o caso de uma bola branca e uma vermelha, temos

- uma dentre 4 brancas,
- uma dentre 3 vermelhas,
- uma dentre 5 verdes,

logo há

$$\binom{1}{4}\binom{1}{4}\binom{1}{5}$$

possibilidades para estas três bolas, que dividimos pelo tamanho do espaço amostral,

$$|\Omega| = \binom{12}{3},$$

possibilidades de escolher 3 bolas dentre 12.

$$\begin{aligned}
\Pr(0,0) &= \binom{5}{3}/\binom{12}{3} = \frac{10}{220} = \frac{1}{22} \\
\Pr(0,1) &= \binom{4}{1}\binom{5}{2}/\binom{12}{3} = \frac{40}{220} = \frac{2}{11} \\
\Pr(0,2) &= \binom{4}{2}\binom{5}{1}/\binom{12}{3} = \frac{30}{220} = \frac{3}{22} \\
\Pr(0,3) &= \binom{4}{3}/\binom{12}{3} = \frac{4}{220} = \frac{1}{55} \\
\Pr(1,0) &= \binom{3}{1}\binom{5}{2}/\binom{12}{3} = \frac{30}{220} = \frac{3}{22} \\
\Pr(1,1) &= \binom{3}{1}\binom{4}{1}\binom{5}{1}/\binom{12}{3} = \frac{60}{220} = \frac{3}{11} \\
\Pr(1,2) &= \binom{3}{1}\binom{4}{2}/\binom{12}{3} = \frac{18}{220} = \frac{9}{110} \\
\Pr(2,0) &= \binom{3}{2}\binom{5}{1}/\binom{12}{3} = \frac{15}{220} = \frac{3}{44} \\
\Pr(2,1) &= \binom{3}{2}\binom{4}{1}/\binom{12}{3} = \frac{12}{220} = \frac{3}{55} \\
\Pr(3,0) &= \binom{3}{3}/\binom{12}{3} = \frac{1}{220}
\end{aligned}$$

A tabela a seguir mostra $p(x, y)$. A última linha contém a soma das colunas.

$$
\begin{array}{cc|cccc|}
\multicolumn{6}{c}{Y\,(\text{BRANCAS})} \\[1em]
 & & 0 & 1 & 2 & 3 \\
\cline{3-6}
 & 0 & \frac{10}{220} & \frac{40}{220} & \frac{30}{220} & \frac{4}{220} \\
\text{X (VERMELHAS)} & 1 & \frac{30}{220} & \frac{60}{220} & \frac{18}{220} & 0 \\
 & 2 & \frac{15}{220} & \frac{12}{220} & 0 & 0 \\
 & 3 & \frac{1}{220} & 0 & 0 & 0 \\
\cline{3-6}
\multicolumn{2}{r}{\Sigma} & \frac{56}{220} & \frac{112}{220} & \frac{48}{220} & \frac{4}{220}
\end{array}
$$

2. A soma da coluna onde $Y = 2$ é

$$\frac{30}{220} + \frac{18}{220} = \frac{48}{220},$$

igual a $\Pr[Y = 2]$ – verificamos que $\Pr[Y = 2]$, a possibilidade de obter duas bolas brancas, é, realmente

$$
\begin{aligned}
\Pr[Y = 2] &= \frac{\overbrace{\binom{4}{2}}^{2 \text{ de } 4}\overbrace{\binom{8}{1}}^{1 \text{ de } 8}}{\binom{12}{3}} \qquad \text{(4 brancas, 8 outras)} \\
&= \frac{(6)(8)}{220} \\
&= \frac{48}{220}.
\end{aligned}
$$

Resp. (Ex. 110) —

$$\int_0^a \int_0^a xy \, dxdy = \frac{a^4}{4}.$$

Iguale a um,

$$\begin{aligned}\frac{a^4}{4} &= 1\\ a^4 &= 4\\ a &= \pm\sqrt{2},\end{aligned}$$

Como o exercício pede o intervalo $[0, a]$, entedemos que $a > 0$, logo

$$a = \sqrt{2}.$$

Evidentemente, $f(x, y)$ também seria função de densidade se o intervalo usado fosse $\left[-\sqrt{2}, 0\right]$.

Resp. (Ex. 111) — b deve ser positivo, então queremos

$$\begin{aligned}\int_0^q \int_r^s a\log(b) &= 1\\ \frac{(qs^2 - qr^2)\log(q^2) - 2qs^2 + 2qr^2}{4} &= 1\\ (qs^2 - qr^2)\log(q^2) - 2qs^2 + 2qr^2 - 4 &= 0\end{aligned}$$

f é função de densidade em todos os pares de intervalos, $[0, q]$ para B e $[r, s]$ para A, satisfazendo esta condição.

Resp. (Ex. 113) — (1) $2/\log(2)$. (2) Sim, porque $F_x(x) = x^2$, $F_Y(y) = \log(y + 1).\log(2)$, e $F_{XY} = F_X(x)F_Y(y)$. (3) $f_x(x) = 2x$; $f_y(y) = 1/[\log(2)(y + 1)]$. (4) $\mathbb{E}[X] = 2/3$, $\mathbb{E}[Y] = (1 - \log(2))/\log(2)$. $\rho^2_X = 1/18$. (5)

$$\rho_Y = \frac{2\log(2\log^2(2)) - 4\log(\log(2)) + 3\log^2(2) - 4\log(2)}{2\log^3(2)}$$

(6) Sim, porque são independentes. Ainda assim, pode-se confirmar: $\mathbb{E}[XY] = 2(1 - \log(2))/2\log(2)$, logo

$$\begin{aligned}\operatorname{cov}(X, Y) &= \mathbb{E}[XY] - \mathbb{E}[X][Y]\\ &= 2(1 - \log(2))/2\log(2) - \left(\frac{2}{3}\right)\left(\frac{1 - \log(2)}{\log(2)}\right)\\ &= 0,\end{aligned}$$

portanto a correlação também é zero.

Resp. (Ex. 115) — (1) $c = 3$, $k = e^{-4}$
(2) Obtenha as marginais, e verifique que $f_A(a)f_B(b) = f_{AB}(ab)$, logo A e B são independentes. Já X e Y não são.

Resp. (Ex. 116) — $f_Z(z) = (1+z)^{-2}$.

Resp. (Ex. 118) — (a)

$$\begin{aligned}
\mathbb{E}[A] &= 1/12 \\
\mathbb{E}[O] &= 1 \\
\mathbb{E}[A^2] &= 1/63 \\
\mathbb{E}[I^2] &= 4/3 \\
\mathbb{E}[AO] &= 2/15 \\
\text{cov}(A,O) &= 1/20 = 0.05 \\
\rho_A &= \frac{1}{4\sqrt{7}} \\
\rho_O &= \frac{1}{\sqrt{3}} \\
\rho_{A,O} &= \frac{\sqrt{3}\sqrt{7}}{5} \approx 0.9165
\end{aligned}$$

Resp. (Ex. 119) — (1) 60. (2) $\frac{11}{16}$.

Resp. (Ex. 120) —

$$\begin{aligned}
f(x) &= \lambda e^{-\lambda x} \\
g(y) &= \theta e^{-\theta y} \\
F(x) &= 1 - e^{-\lambda x} \\
G(y) &= 1 - e^{-\theta y}
\end{aligned}$$

Fixe um tempo t.

$$\Pr[A < t, B > t] = F(t)(1 - G(t)).$$

Portanto, para todo t possível,

$$\begin{aligned}
\Pr[A < B] &= \int_0^\infty F(t)(1-G(t))dt \\
&= \frac{\lambda}{\theta^2 + \lambda\theta}.
\end{aligned}$$

Resp. (Ex. 122) — Na segunda figura do exemplo há um triângulo retângulo. A variável aleatória T representa o ângulo θ. Se usarmos o outro ângulo não-reto, trocaremos cos por sen, mas o problema permanece o mesmo – *a distribuição dos dois ângulos é a mesma, e o ângulo é independente da distância.*

Resp. (Ex. 125) — (Dica) A distribuição é $A + B \sim \text{binomial}(m + n, p)$.

Resp. (Ex. 126) —

$$\frac{bn}{r},$$

onde b é o total de sucessos.

Resp. (Ex. 129) — Zero.

Resp. (Ex. 131) — $\sqrt{p}$.

Resp. (Ex. 135) — $\text{var}(X) - 2\text{cov}(X, Y) + \text{var}(Y)$. Por isso, se X e Y são independentes, $\text{var}(X + Y) \;=\; \text{var}(X - Y)$.

Resp. (Ex. 136) — Use a desigualdade de Markov, $\Pr[X \geq kn] \leq \mathbb{E}[X]/kn$.

Resp. (Ex. 137) — Veja que $\Pr[X \leq x] = \Pr[k - R \leq -x]$.

Resp. (Ex. 138) — Como a distribuição é $\mathcal{U}(0, 10)$, então

$$\begin{aligned} \mu &= 5 \\ \sigma^2 &= \frac{25}{3}. \end{aligned}$$

Pela desigualdade de Chebychev,

$$\Pr[|X - 5| > 4] \leq \frac{25}{(3)4^2} \approx 0.52.$$

O valor exato é

$$\begin{aligned} \Pr[|X - 5| > 4] &= \Pr[X < 1] + \Pr[X > 9] \\ &= 0.1 + 0.1 \\ &= 0.2. \end{aligned}$$

De fato, $0.2 \leq 0.52$.

Resp. (Ex. 139) — Chebychev.

Resp. (Ex. 141) — Use o Teorema Central do Limite

$$\begin{aligned}\Pr[C < 9800] &\approx \Pr\left[\frac{X - n\mu}{\sigma\sqrt{n}} < \frac{9800 - n\mu}{\sigma\sqrt{n}}\right] \\ &= \Pr\left[Z < \frac{9800 - n\mu}{15\sqrt{49}}\right] \\ &= \Pr[Z < -2.33] \\ &\approx 0.0099,\end{aligned}$$

e conclua que é muito pouco provável que o caminhão possa passar pela barreira!

Resp. (Ex. 143) — 1.

2. exige saber os valores máximo e mínimo da função (α e β).

Resp. (Ex. 145) — O limite é 1/2. Para começar, imagine uma sequência de variáveis aleatórias $X_1, X_2, \ldots$, todas com distribuição de Poisson, e tendo a mesma média. No entanto, será necessário fixar a média em um para prosseguir.

Resp. (Ex. 146) — A área de um semi-círculo é $(\pi r^2)/2$. Calcule a área do semi-círculo que fica nos quadrantes um e dois, de raio um e centrado na origem. A borda é dada por

$$y(x) = \sqrt{1 - x^2},$$

logo sua área é

$$\int_{-1}^{+1} \sqrt{1 - x^2}\, dx.$$

Basta estimar o valor desta integral, usando o método de Monte Carlo discutido no texto, e multiplicar por dois.

Resp. (Ex. 148) — $\alpha = 0.01$, logo $\alpha/2 = 0.005$.

$$1 - \Phi(2.58) = 0.005.$$

A quantidade pedida é

$$\begin{aligned}n &> \frac{(2.58)^2}{4(0.01)^2} \\ &= 16641.\end{aligned}$$

Resp. (Ex. 151) —

$$\begin{aligned} n &\geq \frac{r^2\sigma^2}{\varepsilon^2} \\ &= \frac{\lambda r^2}{\varepsilon^2}, \end{aligned}$$

e n teria que ser determinado em função de λ.
Há métodos para encontrar algum n para este caso, mas não é pela simples aplicação do Teorema 6.4.8

Resp. (Ex. 154) — Primeiro determine k, $\int_0^1 kx^5 dx = 1$, obtendo $k = 6$. Depois calcule a função de densidade acumulada,

$$F(x) = x^6.$$

Pelo método da inversão, se U tem distribuição uniforme, então X pode ser simulada usando

$$X = \sqrt[6]{U}.$$

Resp. (Ex. 155) — Se u é gerado com distribuição uniforme em $[0, 1]$, então

$$g = \left\lfloor \frac{\log(u)}{\log(1-p)} \right\rfloor$$

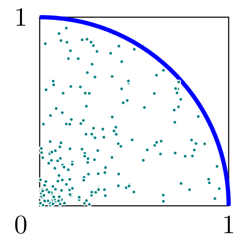
tem distribuição geométrica com parâmetro p. (Método da inversão)

Resp. (Ex. 156) — Comece calculando $\Pr[k - 1 \leq X < k]$.

Resp. (Ex. 157) — A função cumulativa de densidade da distribuição de Cauchy é $F(x) = 1/2 + \arctan(x)/\pi$, facilmente invertível.

Resp. (Ex. 159) — Primeiro verifique que a área deve ser um, e portanto determine $f(3) = 0.5$. Para o método da inversão, a função de densidade acumulada é simples e facilmente invertida. Para o método da rejeição, basta gerar pontos no retângulo $(2, 0) - (6, 0) - (6, 0.5) - (2, 0.5)$ e rejeitar aqueles fora do triângulo.

Resp. (Ex. 163) — Simula X e Y, ambas uniformes em $[0, 1]$. Quando o ponto (x, y) estiver dentro do setor circular da circunferência unitária que fica no primeiro quadrante, aumente a contegem de pontos internos.



Como a área do setor circular é $(\pi r^2)/4$, a probabilidade de um ponto ser escolhido dentro dele é

$$\frac{\text{área do setor circular}}{\text{área do quadrado}},$$

ou seja,

$$\begin{aligned}
\Pr(\text{ponto interno}) &= \frac{(\pi r^2)/4}{1} \\
&= \frac{\pi}{4}. && (r = 1)
\end{aligned}$$

Mas a probabilidade de um ponto ficar dentro do setor circular pode ser aproximada por

$$\frac{\text{pontos internos}}{\text{total de pontos}}.$$

portanto, ao final do experimento basta calcular

$$\pi \approx \frac{4a}{n},$$

onde a é a quantidade de pontos internos e n é a quantidade total de pontos.

Resp. (Ex. 165) — (i) Área abaixo da curva (um) dividida pela área onde os pontos são obtidos (largura 2, altura 0.75, área 1.5). A probabilidade é 2/3. (ii) O exercício pede a esperança de uma distribuição geométrica com parâmetro 2/3; a resposta é portanto $3/2 = 1.5$ tentativas em média.

Resp. (Ex. 171) — Dica: use a transformação Box-Müller (que não é útil apenas em simulações, como podemos ver!)

$$\begin{aligned}
X + Y &= R\cos(2\pi U) + R\,\mathrm{sen}(2\pi U) = \sqrt{2}R\,\mathrm{sen}(2\pi U + \pi/4) \\
X - Y &= R\cos(2\pi U) - R\,\mathrm{sen}(2\pi U) = \sqrt{2}R\cos(2\pi U + \pi/4)
\end{aligned}$$

(A raiz de dois tem influência na variância)

Resp. (Ex. 172) — Use o Teorema Central do Limite.

Ficha Técnica

Este texto foi produzido inteiramente em LaTeX em sistema Debian GNU/Linux. Os diagramas foram criados sem editor gráfico, usando diretamente o pacote Ti*k*Z, incluindo as imagens de variáveis simuladas no Capítulo 7, que foram construídas inteiramente em Ti*k*Z. A exceção é um único gráfico, que usa o método da rejeição e lê uma tabela criada por programa externo, escrito em Scheme (a implementação STklos foi usada). O ambiente Emacs foi usado para edição do texto LaTeX.

Índice Remissivo