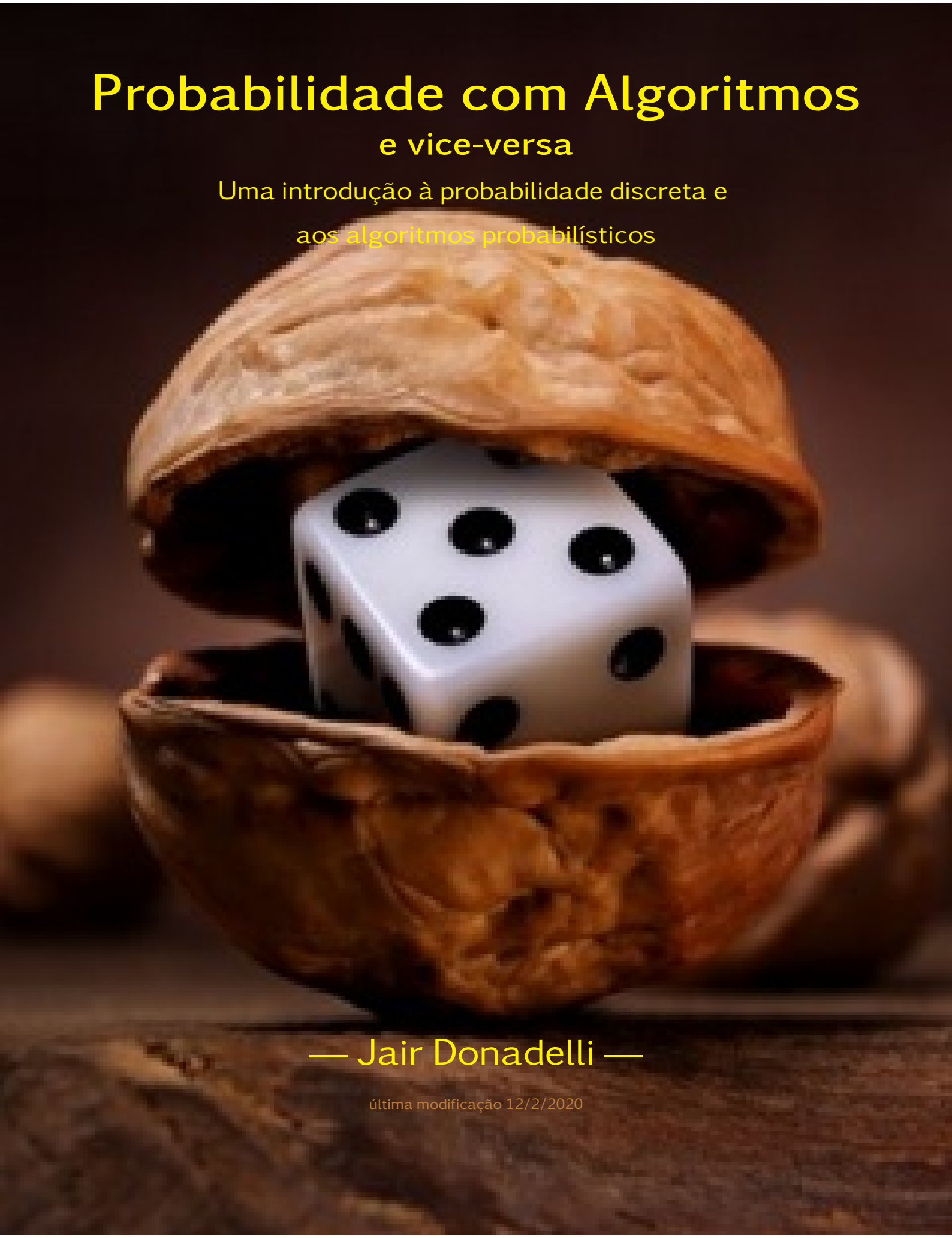
Probabilidade com Algoritmos

e vice-versa

Uma introdução à probabilidade discreta e aos algoritmos probabilísticos



— Jair Donadelli —

última modificação 12/2/2020

[*Rosencrantz and Guildenstern are riding horses down a path - they pause*]
R: Umm, uh...
[*Guildenstern rides away, and Rosencrantz follows. Rosencrantz spots a gold coin on the ground*]
R: Whoa - whoa, whoa.
[*Gets off horse and starts flipping the coin*] **R**: Hmmm. Heads. Heads. Heads. Heads. Heads. Heads. Heads. Heads. Heads. Heads. Heads. Heads. Heads. Heads. Heads. Heads. Heads. Heads.
[*Guildenstern grabs the coin, checks both sides, then tosses it back to Rosencrantz*]
R: Heads.
[*Guildenstern pulls a coin out of his own pocket and flips it*]
R: Bet? Heads I win?
[*Guildenstern looks at coin and tosses it to Rosencrantz*]
R: Again? Heads.

[...]

R: Heads
G: A weaker man might be moved to re-examine his faith, if in nothing else at least in the law of probability.
R: Heads
G: Consider. One, probability is a factor which operates within natural forces. Two, probability is not operating as a factor. Three, we are now held within um...sub or supernatural forces. Discuss!
R: What?
[...]
R: Heads, getting a bit of a bore, isn't it?

[...]

R: 78 in a row. A new record, I imagine.
G: Is that what you imagine? A new record?
R: Well...
G: No questions? Not a flicker of doubt?
R: I could be wrong.
G: No fear?
R: Fear?
G: Fear!
R: Seventy nine.

[...]

G: I don't suppose either of us was more than a couple of gold pieces up or down. I hope that doesn't sound surprising because its very unsurprisingness is something I am trying to keep hold of. The equanimity of your average tosser of coins depends upon a law, or rather a tendency, or let us say a probability, or at any rate a mathematically calculable chance, which ensures that he will not upset himself by losing too much nor upset his opponent by winning too often. This made for a kind of harmony and a kind of confidence. It related the fortuitous and the ordained into a reassuring union which we recognized as nature. The sun came up about as often as it went down, in the long run, and a coin showed heads about as often as it showed tails.

Tom Stoppard, *Rosenkrantz and Guildenstern are dead* (1996).

1 | ESPAÇOS DE PROBABILIDADE

Intuitivamente, uma medida de probabilidade é uma forma quantitativa de expressar a chance com que um conjunto de resultados de um experimento *aleatório*, um processo que nos fornece resultados os quais não podem ser previamente determinados, ocorra. A Probabilidade é a disciplina dedicada a modelagem desses fenômenos com condições de incerteza. Um modelo probabilístico é um modelo matemático de um experimento aleatório. A modelagem probabilística tem sido importante em praticamente todas as áreas do conhecimento e o desenvolvimento da Teoria da Probabilidade tem sido estimulada pela ampla variedade de suas aplicações. Neste capítulo introduzimos o tratamento axiomático moderno da probabilidade introduzido pelo matemático russo Andrei Nikolaevich Kolmogorov (1903-1987) por volta de 1930 e que, ao contrário da interpretações que estabelecem uma forma explícita de calcular probabilidades, o modelo axiomático estuda as propriedades que uma probabilidade deve satisfazer. Veremos a importante noção de probabilidade condicional e a independência de eventos.

1.1 ESPAÇOS DE PROBABILIDADE DISCRETOS

Monty Hall é o nome do apresentador de um concurso televisivo exibido na década de 1970 nos Estados Unidos chamado *Let's Make a Deal*, e é o nome de um problema agora clássico em probabilidade. O jogo consistia em o apresentador Monty Hall apresentar três portas a um espectador que concorre a um prêmio escondido pela porta escolhida através de um processo de escolhas que será descrito a seguir. O protocolo da brincadeira é: Monty Hall escolhe, ao acaso com igual probabilidade, uma das portas para esconder um carro; nas outras duas esconde um bode cada. Na primeira etapa o concorrente escolhe uma porta ao acaso (que ainda não é aberta); em seguida Monty Hall abre uma das outras duas portas que o concorrente não escolheu, sabendo que ela esconde um bode. Se são duas possibilidades, ele escolhe uma ao acaso. Com duas portas fechadas apenas, e sabendo que o carro está atrás de uma delas, o apresentador oferece ao concorrente a oportunidade de trocar de porta. O concorrente tem que decidir se permanece com a porta que escolheu no início do jogo ou se muda para a outra porta que ainda está fechada; feita a escolha, o apresentador abre a porta escolhida e o concorrente leva o prêmio escondido pela porta.

Assumindo que o objetivo do jogador é ganhar o carro, o problema é determinar uma estratégia de decisão que

maximiza a chance de ganhar o carro. A resposta para esse problema será dada mais a frente no texto, no momento convidamos o leitor a refletir um pouco sobre o problema antes de passar adiante na leitura, para, ao menos, identificar os experimentos aleatórios escondidos na descrição feita no parágrafo acima.

Um modelo probabilístico para um experimento aleatório é caracterizado por um *espaço amostral* — conjunto dos resultados possíveis — um *espaço de eventos* — família[1] dos subconjuntos de resultados que admitem uma probabilidade — e uma *(medida de) probabilidade* — uma função que associa um valor numérico a cada evento.

Espaço amostral O espaço amostral de um experimento aleatório, quase sempre denotado por Ω, é um conjunto não vazio em que cada elemento representa um resultado possível do experimento e cada resultado tem um representante que pertence ao conjunto. Um elemento de Ω é chamado de **ponto amostral** e a escolha de algum ponto amostral representa uma realização do experimento.Ω

Exemplo 1.1. São experimentos com respectivos espaços amostrais

1. um dado é lançado e observamos a face para cima, $\Omega = \{1,2,3,4,5,6\}$;
2. uma moeda é lançada e observamos sua face para cima, $\Omega = \{\text{Ca}, \text{Co}\}$;
3. uma moeda é lançada e observamos o resultado até sair coroa. Cada ponto amostral é representado por uma sequência de Ca que termina com Co e por ∞ que representa a eventualidade de nunca ocorrer coroa, assim $\Omega = \{(\text{Co}), (\text{Ca}, \text{Co}), (\text{Ca}, \text{Ca}, \text{Co}), \ldots, (\text{Ca}, \text{Ca}, \ldots, \text{Ca}, \text{Co}), \ldots, \infty\}$;
4. dois amigos competem, cada um com uma sequência de três caras ou coroas, num jogo em que uma moeda honesta é lançada sucessivamente até que saia uma das duas sequências. Como pode ser necessário lançar a moeda uma quantidade arbitrariamente grande de vezes um espaço amostral é considerar todas as sequências $(a_i \in \{\text{Ca}, \text{Co}\} \colon i \geqslant 1)$ de resultados possíveis, isto é, $\Omega = \{\text{Ca}, \text{Co}\}^{\mathbb{N}}$;
5. observamos tempo de vida de uma lâmpada, $\Omega = \{t \in \mathbb{R} \colon t \geqslant 0\}$;
6. um brasileiro é escolhido e medimos sua altura, $\Omega = \{h \in \mathbb{R} \colon h > 0\}$;
7. um dardo é lançado num alvo circular de raio 1 e observamos o ponto atingido, um espaço amostral é obtido usando um sistema de coordenadas cartesianas com a origem no centro do alvo de modo que $\Omega = \{(x,y) \in \mathbb{R}^2 \colon x^2 + y^2 \leqslant 1\}$. $\Diamond$

O espaço amostral de um experimento aleatório reflete a observação do resultado de um experimento e não é único; no item 3 do exemplo acima, podemos escrever o espaço amostral $\{1,2,3,\ldots,\infty\}$ para representar os resultados do experimento. No item 7 do exemplo acima, podemos escrever o espaço amostral com pontos dados em coordenadas polares $\{(r,\theta) \in \mathbb{R}^2 \colon 0 \leqslant r \leqslant 1 \text{ e } -\pi < \theta \leqslant \pi\}$.

EXERCÍCIO 1.2. Identifique os experimentos aleatórios e descreva um espaço amostral para o problema de Monty Hall.

Espaço de eventos Intuitivamente um evento aleatório de um experimento aleatório é um acontecimento observável ao final da realização do experimento. Quando da realização do experimento deve ser sempre possível dizer se tal fenômeno ocorreu ou não ocorreu. Por exemplo, se um dado é lançado então o resultado "é um número par" e o resultado "é um número maior que 3" são eventos do experimento de lançar um dado. Um evento E é representado no modelo probabilístico por um subconjunto E_Ω do espaço amostral Ω o qual fica definido pela coleção de resultados *possíveis* do experimento que satisfazem a descrição do evento. Nos exemplo do dado, se $\Omega = \{1,2,3,4,5,6\}$ então "é um número

[1] Família é usado como sinônimo de conjunto.

par" e "é um número maior que 3" são modelados por $\{2,4,6\}$ e $\{4,5,6\}$, respectivamente. Usualmente, por abuso de notação, usamos E para denotar E_Ω embora o modelo para um evento depende do espaço amostral construído.

Assim, um modelo de um evento aleatório é subconjunto do espaço amostral Ω também chamado de **evento aleatório**. Na realização de um experimento o evento $A \subset \Omega$ *ocorre* se o resultado observado é representado por um elemento de A, caso contrário o evento A *não ocorre*. Em especial, $\varnothing$ é o evento *impossível*; Ω é o evento *certo*; $\{\omega\}$ é um evento *elementar* para cada elemento $\omega \in \Omega$; o *complemento* do evento A é o evento *não*-A dado por

$$\overline{A} = \Omega \setminus A := \{\omega \in \Omega : \omega \notin A\}.$$

Em um lançamento de dados $\Omega = \{1,2,3,4,5,6\}$ e são exemplos de eventos

- $A = \{2,4,6\}$, ou seja, A representa o evento "número par";
- $\overline{A} = \{1,3,5\}$, ou seja, $\overline{A}$ representa o evento "não é número par";
- $B = \{4,5,6\}$, ou seja, B representa o evento "número maior que 3";
- $C = \{4\}$, ou seja, C representa o evento "número 4";
- $A \cap \overline{A} = \varnothing$, ou seja, $A \cap \overline{A}$ representa o evento "número par *e* número ímpar", que é o evento impossível;
- $A \cup \overline{A} = \Omega$, ou seja, $A \cup \overline{A}$ representa o evento "número par *ou* número ímpar", que é o evento certo;
- $B \cap C = \{4\}$, ou seja, $B \cap C$ representa o evento "número maior que 3 *e* número 4" que é o mesmo evento que "número 4";
- $B \cap A = \{4,6\}$, ou seja, $B \cap A$ representa o evento "número maior que 3 *e* número par";
- o evento "múltiplo de 2 *ou* múltiplo de 3 mas não múltiplo de ambos" é representado pela diferença simétrica $\{2,4,6\} \triangle \{3,6\} = (\{2,4,6\} \cup \{3,6\}) \setminus (\{2,4,6\} \cap \{3,6\}) = \{2,3,4\}$.

Dizemos que A e B são eventos **disjuntos** ou **eventos mutuamente exclusivos** quando não têm elementos em comum, isto é, $A \cap B = \varnothing$. Os eventos $A_1, A_2, \ldots, A_n$ são ditos **mutuamente exclusivos** se são disjuntos tomados dois-a-dois, isto é, $A_i \cap A_j = \varnothing$ sempre que $i \neq j$. Embora eventos sejam conjuntos e a Teoria dos Conjuntos tem uma linguagem tradicional e bem aceita a Probabilidade tem um linguagem peculiar para os eventos e a descrevemos na tabela 1.1 abaixo.

Denotemos por $\mathcal{A}$ um conjunto de eventos aleatórios que podem ocorrer num experimento aleatório. Para ser consistente com a intuição $\mathcal{A}$ deve ter $\varnothing \in \mathcal{A}$ e $\Omega \in \mathcal{A}$ entre seus elementos, ser fechado para as operações usuais de conjunto (as linhas da tabela acima descrevem eventos) e, também, pedimos que satisfaça o seguinte: se $A_i \in \mathcal{A}$ para todo $i \geqslant 1$, então $\bigcup_{i \geqslant 1} A_i \in \mathcal{A}$ e uma justificativa para isso é dada adiante.

Um **espaço de eventos** é um conjunto $\mathcal{A}$ de eventos aleatórios de um experimento aleatório. Quais são as famílias de subconjuntos de Ω que podem ser tomadas como espaço de eventos é um assunto que não trataremos. Uma escolha óbvia é o conjunto 2^Ω das partes de Ω, mas acontece que em muitos casos é preciso restringir essa família a um subconjunto próprio de 2^Ω para que questões probabilísticas façam sentido. Por ora, chamamos atenção ao fato de ser possível haver subconjuntos de um espaço amostral Ω que não são eventos aleatórios, como é o caso dado no exemplo 1.10 na página 11. Esse fenômeno só é importante quando Ω é muito grande, infinito e não enumerável.

EXERCÍCIO 1.3. Descreva, segundo a solução dada no exercício 1.2, o evento de interesse no problema de Monty Hall, isto é, o subconjunto que modela o evento "o espectador concorrente ganha o carro".

Notação	*Eventos*	*Conjunto*
Ω	espaço amostral, evento certo	universo
$\varnothing$	evento impossível	vazio
$\{\omega\}$	evento elementar	conjunto unitário
A	evento	subconjunto
A	ocorre A	$\omega \in A$
$\overline{A}$	não ocorre A	$\omega \notin A$ (complemento)
$A \cap B$	ocorre A e B	$\omega \in A \cap B$ (intersecção)
$A \cup B$	ocorre A ou B	$\omega \in A \cup B$ (união)
$A \setminus B$	ocorre A e não ocorre B	$\omega \in A$ e $\omega \notin B$ (diferença)
$A \triangle B$	ocorre A ou B, não ambos	$\omega \in A \cup B$ e $\omega \notin A \cap B$ (diferença simétrica)
$A \subset B$	se ocorre A, então ocorre B	$\omega \in A \Rightarrow \omega \in B$ (inclusão)

Tabela 1.1: termos da Probabilidade.

Medida de probabilidade Uma medida de probabilidade sobre um espaço de eventos $\mathcal{A}$ de um espaço amostral Ω é uma função, genericamente denotada por $\mathbb{P}$, que atribui a cada evento aleatório $A \in \mathcal{A}$ um número real $\mathbb{P}(A)$ satisfazendo os seguintes axiomas

A1 – *não negatividade*: $\mathbb{P}(A) \geqslant 0$;

A2 – *normalização*: $\mathbb{P}(\Omega) = 1$;

A3 – *aditividade enumerável*: $\mathbb{P}\left(\bigcup_{i\geqslant 1} A_i\right) = \sum_{i\geqslant 1} \mathbb{P}(A_i)$ sempre que $\{A_i : i \geqslant 1\}$ é um conjunto de eventos mutuamente exclusivos.[2]

As primeiras consequências importantes desses axiomas são enunciadas na proposição a seguir.

Proposição 1.4 *Algumas consequências desses axiomas são dadas abaixo.*

1. *A probabilidade do evento impossível é* $\mathbb{P}(\varnothing) = 0$.
2. *Aditividade finita: se* $A_1, A_2, \ldots, A_n$ *são eventos mutuamente exclusivos então*
$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathbb{P}(A_i).$$
3. *A probabilidade do complemento satisfaz* $\mathbb{P}(A) + \mathbb{P}(\overline{A}) = 1$, *para todo evento* A.
4. *Monotonicidade: se* $A \subset B$ *então* $\mathbb{P}(A) \leqslant \mathbb{P}(B)$.
5. *Regra da adição:* $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$ *para quaisquer eventos* A *e* B.

Demonstração. Fazendo $A_1 = \Omega$ e $A_i = \varnothing$ para todo $i \geqslant 2$ temos, pela aditividade enumerável, que
$$\mathbb{P}(\Omega) = \mathbb{P}(\Omega \cup \varnothing \cup \varnothing \cup \cdots \cup \varnothing \cup \cdots) = \mathbb{P}(\Omega) + \sum_{i\geqslant 2} \mathbb{P}(\varnothing)$$

[2] O lado esquerdo da igualdade não depende de uma enumeração particular dos conjuntos A_i e, nesse caso, o mesmo vale para o lado direito, veja (s.1).

portanto, pela não-negatividade, resta que $\mathbb{P}(\varnothing) = 0$. Agora, definindo $A_i = \varnothing$ para todo $i > n$

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \mathbb{P}\left(\bigcup_{i\geqslant 1} A_i\right) = \sum_{i\geqslant 1} \mathbb{P}(A_i) = \sum_{i=1}^{n} \mathbb{P}(A_i) + \sum_{i>n} \mathbb{P}(\varnothing) = \sum_{i=1}^{n} \mathbb{P}(A_i)$$

que é o resultado afirmado.

A probabilidade da complemento segue do item anterior e da normalização. Os detalhes ficam a cargo do leitor.

Para monotonicidade, consideremos A e B eventos tais que $A \subset B$. Usamos que B pode ser escrito como a união disjunta $A \cup (\overline{A} \cap B)$, donde $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(\overline{A} \cap B)$ e como $\mathbb{P}(\overline{A} \cap B) \geqslant 0$ temos $\mathbb{P}(B) \geqslant \mathbb{P}(A)$. Notemos que, como consequência imediata, temos para todo evento A $\mathbb{P}(A) \leqslant 1$.

Finalmente, a união $A \cup B$ pode ser escrita como duas uniões disjuntas $(A \setminus B) \cup (B \setminus A) \cup (A \cap B)$ donde concluímos que

$$\mathbb{P}(A \cup B) = \mathbb{P}(A \setminus B) + \mathbb{P}(B \setminus A) + \mathbb{P}(A \cap B). \tag{1.1}$$

Agora, A pode ser escrito como a união disjunta $(A \setminus B) \cup (A \cap B)$ e, analogamente, $B = (B \setminus A) \cup (A \cap B)$, portanto $\mathbb{P}(A) = \mathbb{P}(A \setminus B) + \mathbb{P}(A \cap B)$ assim como $\mathbb{P}(B) = \mathbb{P}(B \setminus A) + \mathbb{P}(A \cap B)$. Isolando $\mathbb{P}(A \setminus B)$ e $\mathbb{P}(B \setminus A)$ nessas duas igualdades e substituindo na equação (1.1) prova a regra da adição. □

O seguinte limitante é bastante útil e podes ser facilmente provado usando indução e a regra da adição.

Corolário 1.5 *Se $A_1, A_2, \ldots, A_n$ são eventos então*

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

Exemplo 1.6 (*lançamento de uma moeda equilibrada*). O modelo probabilístico para o lançamento de uma moeda equilibrada é $\Omega = \{\mathrm{Ca}, \mathrm{Co}\}$ e $\mathbb{P}(\varnothing) = 0$, $\mathbb{P}(\{\mathrm{Ca}\}) = \mathbb{P}(\{\mathrm{Co}\}) = 1/2$, $\mathbb{P}(\Omega) = 1$. ◊

Exemplo 1.7 (*lançamento de um dado equilibrado*). No caso do lançamento de um dado equilibrado atribuímos a probabilidade $1/6$ a cada uma das faces, o que é interpretado como todas as faces serem equiprováveis. A partir disso qualquer subconjunto $A \subset \Omega$ de faces do dado é um evento que tem probabilidade de ocorrência dada por

$$\mathbb{P}(A) = \frac{|A|}{6} \tag{1.2}$$

de modo que os axiomas de probabilidade ficam satisfeitos. Esse exemplo, bem como o exemplo anterior, é o modo clássico de interpretar probabilidade no caso finito. ◊

Exemplo 1.8. Quando escolhemos um inteiro positivo ao acaso com a probabilidade de escolher i dada por $(1/2)^i$ e estendemos a probabilidade a qualquer subconjunto A de inteiros positivos pondo

$$\mathbb{P}(A) := \sum_{a\in A} \mathbb{P}(\{a\}) \tag{1.3}$$

temos um modelo probabilístico. De fato, temos (veja (s.6a))

$$\mathbb{P}(\Omega) = \sum_{i\geqslant 1} \left(\frac{1}{2}\right)^i = 1$$

e a convergência absoluta dessa série implica que toda subsérie dela é convergente (veja (s.4)), assim temos que a probabilidade dada na equação (1.3) está bem definida, isto é, $\mathbb{P}(A)$ como definido acima é um número real não negativo menor ou igual a 1.

Também segue da convergência absoluta que um rearranjo da série resulta noutra série que converge para o mesmo resultado donde obtemos a aditividade enumerável da medida de probabilidade,

$$\mathbb{P}\left(\bigcup_{i\geq 1} A_i\right) = \sum_{a\in\bigcup_i A_i} \mathbb{P}(\{a\}) = \sum_{i\geq 1}\sum_{a\in A_i} \mathbb{P}(\{a\}) = \sum_{i\geq 1} \mathbb{P}(A_i) \tag{1.4}$$

para qualquer conjunto $\{A_i : i \geqslant 1\}$ de eventos mutuamente exclusivos.

A probabilidade de escolher um número par é

$$\sum_{a \text{ par}} \mathbb{P}(\{a\}) = \sum_{k\geqslant 1}\left(\frac{1}{2}\right)^{2k} = \sum_{k\geqslant 1}\left(\frac{1}{4}\right)^{k} = \frac{1}{3}$$

portanto, calculando a probabilidade do complemento, a probabilidade de escolher um número ímpar é 2/3. A probabilidade de escolha de um múltiplo de 3 é 1/7 e a probabilidade da escolha de um múltiplo de 6 é 1/31 (verifique). Usando a regra da adição e o fato de que ser múltiplo de 6 equivale a ser múltiplo de 2 *e* múltiplo de 3, temos que a probabilidade de escolha de um múltiplo de 2 *ou* um múltiplo de 3 é a probabilidade de escolha de um múltiplo de 2 mais a probabilidade de escolha de um múltiplo de 3 menos a probabilidade de escolha de um múltiplo de 6, ou seja, a probabilidade de escolha de um múltiplo de 2 *ou* um múltiplo de 3 é $1/3+1/7-1/31 = 289/651 \approx 0{,}444$. ◊

Exemplo 1.9. No intervalo $\Omega = [0,1]$ da reta real podemos definir uma medida de probabilidade $\mathbb{P}$ de modo que os intervalos (a,b), $(a,b]$, $[a,b)$, $[a,b]$ tenham probabilidade $|b-a|$, entretanto não há tal medida de modo que $\mathbb{P}(A)$ esteja definida para todo $A \subset \Omega$, ou seja, nem todo subconjunto do espaço amostral é evento (veja, e.g., Rosenthal, 2006, proposição 1.2.6). O conjunto dos eventos aleatórios é subconjunto próprio do conjunto das partes do intervalo. ◊

Exemplo 1.10 (*probabilidade geométrica*). Consideremos o experimento 7 do exemplo 1.1. É possível definir uma medida de probabilidade para $A \subset \Omega$ como a área de A proporcionalmente a de Ω, i.e.,

$$\mathbb{P}(A) = \frac{\text{Área}(A)}{\pi}$$

Assim, a probabilidade de um lançamento aleatório acertar o círculo de mesmo centro do alvo e raio 1/2 é 1/4.

Como no exemplo anterior, há subconjuntos de Ω que não têm uma probabilidade associada pois não é possível definir área para todo subconjunto do plano (veja, e.g., Gelbaum e Olmsted, 1964, capítulo 11). ◊

Nesses dois últimos casos, além do fato de não poder se atribuir uma medida de probabilidade para qualquer subconjunto, pode haver mais de um modo natural de definir probabilidade para um evento, diferente de (1.2) por exemplo. No caso do alvo, como o raio do círculo menor é 1/2 e o raio do círculo maior é 1, a probabilidade procurada pode ser definida como 1/2, ou seja, como o centro do alvo é o objetivo, podemos definir a probabilidade como proporcional à distância. Ao contrário do caso finito, no caso contínuo uma escolha aleatória não define, ao menos intuitivamente, unicamente o modelo probabilístico.

Exemplo 1.11 (*paradoxo de Bertrand*). O seguinte problema, conhecido como o paradoxo de Bertrand mas que a rigor não é um paradoxo, é passível de mais de uma interpretação para a palavra *aleatório*. Numa circunferência de raio 1, um triângulo equilátero inscrito tem lado $\sqrt{3}$ (figura 1.1). Qual é a probabilidade de que uma corda AB escolhida ao acaso tem comprimento maior que $\sqrt{3}$?

Na primeira interpretação a escolha da corda é por tomarmos A e B escolhidos ao acaso dentre os pontos da circunferência. Imaginemos, o triângulo rotacionado de modo que um de seus vértices coincida com o ponto A. A corda tem comprimento maior que o lado do triângulo se B está no arco da circunferência entre os dois outros vértices do triângulo, o que ocorre com probabilidade 1/3 (os vértices dividem a circunferência em três arcos de mesmo comprimento, figura 1.2).

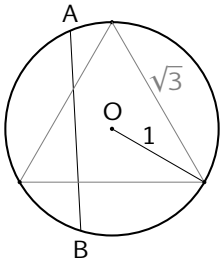


Figura 1.1: paradoxo de Bertrand.

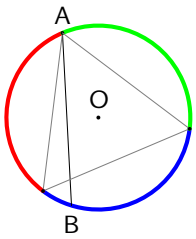


Figura 1.2: a corda é dada por uma escolha aleatória de A e de B na circunferência. A probabilidade procurada é 1/3.

Na segunda interpretação, a corda é obtida por uma escolha de P no interior da circunferência e AB é a corda cujo ponto médio é P (figura 1.3). A corda é maior que o lado do triângulo se P está no interior da circunferência de centro O e raio 1/2, o que ocorre com probabilidade 1/4 (como discutimos acima, poderíamos atribuir probabilidade 1/2, mas isso não será interessante por causa no próximo caso).

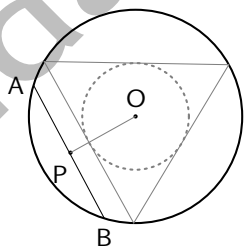


Figura 1.3: a corda AB é definida por P, seu ponto médio. A probabilidade procurada é 1/4.

Na terceira é última interpretação para uma corda aleatória nós fixamos um raio. A corda é obtida escolhendo um ponto P no raio e tomando a corda que passa por P e perpendicular ao raio (figura 1.11). A corda é maior do que um lado do triângulo, se o ponto escolhido está mais próximo do centro do círculo, que o ponto onde o lado do triângulo intersecta o raio, logo se $|OP| \in (0, 1/2)$ o que ocorre com probabilidade 1/2. ◊

Há uma diferença fundamental entre os modelos probabilísticos dos exemplos 1.7 e 1.8 e os modelos dos exemplos 1.9 e 1.10. Nos dois primeiros é possível atribuir probabilidade a todo subconjunto do espaço amostral, o que não é possível nos outros dois. A explicação desse fenômeno é muito técnica para ser dada aqui, mas está relacionada a cardinalidade do espaço amostral. Um espaço amostral enumerável (finito ou infinito) é chamado de **espaço amostral discreto**. Um espaço que têm a mesma cardinalidade dos reais, que também é o caso do item 4 do exemplo 1.1, é chamado de **espaço amostral contínuo**. São espaços discretos os espaços amostrais dos experimentos 1, 2 e 3 dados no exemplo 1.1; os experimentos restantes do exemplo 1.1 têm espaços contínuos.

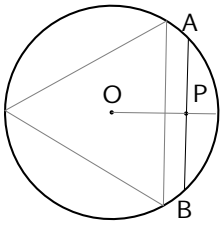


Figura 1.4: a corda é dada pela escolha de um raio e pela escolha de um ponto P nesse raio. A probabilidade procurada é 1/2.

A respeito do espaço de eventos de um espaço amostral contínuo vale uma afirmação mais forte do que o fato descrito nos exemplos 1.9 e 1.10 acima, é verdade que *não há medida de probabilidade que possa ser definida para todo subconjunto* desses espaços. Esse é um resultado difícil para demonstrarmos aqui e que só ocorre em espaços amostrais não enumeráveis. Em resumo, o espaço de eventos $\mathcal{A}$ é uma necessidade técnica e sua compreensão vai muito além do que precisamos neste texto que é dedicado ao caso discreto.

EXERCÍCIO 1.12. Determine uma medida de probabilidade para os eventos do problema de Monty Hall.

Espaço de probabilidade Probabilidade pode ser estudada do ponto de vista abstrato sem se referir a experimentos aleatórios e sem que os números associados aos eventos tenham qualquer interpretação. Formalmente, exigimos que qualquer medida de probabilidade $\mathbb{P}$ esteja definida sobre uma família $\mathcal{A}$ de subconjuntos de Ω que deve satisfazer: (i) $\Omega \in \mathcal{A}$; (ii) se $A \in \mathcal{A}$ então $\overline{A} \in \mathcal{A}$; (iii) se $A_i \in \mathcal{A}$ para todo $i \geqslant 1$, então $\bigcup_{i\geqslant 1} A_i \in \mathcal{A}$. Uma família de subconjuntos como acima é dita σ-álgebra de subconjuntos de Ω. Um **espaço de probabilidade**, assim como um **modelo probabilístico**, é uma terna $(\Omega, \mathcal{A}, \mathbb{P})$ tal que Ω é um conjunto não vazio, chamado **espaço amostral**; $\mathcal{A}$ é uma σ-álgebra de subconjuntos de Ω ditos **eventos**; e $\mathbb{P}\colon \mathcal{A} \to [0,1]$ é uma **medida de probabilidade**.

Deixamos para a reflexão do leitor o fato de que todo modelo probabilístico de um experimento aleatório corresponde a um espaço de probabilidades e todo espaço de probabilidades corresponde ao modelo probabilístico de um experimento e usaremos essas terminologias sem distinção.

1.1.1 MODELO PROBABILÍSTICO DISCRETO

Um **modelo probabilístico discreto**, ou **espaço de probabilidade discreto**, é um espaço $(\Omega, 2^\Omega, \mathbb{P})$ em que Ω é enumerável (finito ou infinito). No caso de espaço amostral discreto, todo experimento tem seu modelo probabilístico especificado quando estabelecemos

(D1) um espaço amostral enumerável Ω que pode ser finito ou infinito;

(D2) uma função de probabilidade $p\colon \Omega \to [0,1]$ tal que $\sum_{\omega\in\Omega} p(\omega) = 1$.

De fato, dado (Ω, p) como acima podemos definir uma função sobre 2^Ω tomando

$$\mathbb{P}(A) := \sum_{\omega\in A} p(\omega)$$

que é um número real positivo para qualquer $A \subset \Omega$, como já observamos no exemplo 1.8 (veja (s.4)).

Convencionamos a notação $\mathbb{P}(\omega) := \mathbb{P}(\{\omega\})$ para os eventos elementares.

Claramente, $\mathbb{P}(A) \geqslant 0$ e $\mathbb{P}(\Omega) = \sum_i \mathbb{P}(\omega_i) = 1$. Ainda, se A_i para $i \geqslant 1$ são eventos mutuamente exclusivos então $\mathbb{P}\big(\bigcup_{i\geqslant 1} A_i\big) = \sum_{i\geqslant 1} \mathbb{P}(A_i)$, como na equação (1.4), segue da convergência absoluta e da exclusão mútua.

Para registro, enunciamos o seguinte resultado sem prova.

TEOREMA *Se $\Omega \neq \varnothing$ é enumerável e $p\colon \Omega \to [0,1]$ é tal que $\sum_{\omega\in\Omega} p(\omega) = 1$ então $(\Omega, 2^\Omega, \mathbb{P})$ com $\mathbb{P}(A) = \sum_{\omega\in A} p(\omega)$ para todo $A \in 2^\Omega$ é um espaço de probabilidade. Reciprocamente, se $(\Omega, 2^\Omega, \mathbb{P})$ é um espaço de probabilidade sobre Ω enumerável então (Ω, p) com $p(\omega) := \mathbb{P}(\{\omega\})$ satisfaz os itens (D1) e (D2) acima.*

Exemplo 1.13. No item 3 do exemplo 1.1 uma moeda equilibrada é lançada e observamos o resultado até sair coroa. Esse experimento é modelado pelo espaço amostral $\Omega = \{(\text{Co}), (\text{Ca},\text{Co}), (\text{Ca},\text{Ca},\text{Co}), \ldots, \infty\}$ munido da função de probabilidade $p((c_1, c_2, \ldots, c_i)) = 2^{-i}$ onde $c_j = \text{Co}$ se $j = i$ e $c_j = \text{Ca}$ caso contrário, e $p(\infty) = 0$. Da argumentação feita no exemplo 1.8, página 10, deduzimos igualmente que Ω e p definem um modelo probabilístico discreto para o experimento. ◊

Exemplo 1.14 (*um modelo probabilístico para Monty Hall*). No caso do problema de Monty Hall, consideremos o experimento que consiste das seguintes três etapas

1. o apresentador esconde o carro atrás de uma das portas escolhida com probabilidade 1/3;
2. com probabilidade 1/3, uma porta é escolhida pelo jogador;
3. o apresentador revela, dentre as duas que o jogador não escolheu, aquela que não esconde o carro. Se houver duas possibilidades então o apresentador escolhe uma delas com probabilidade 1/2.

O espaço amostral é definido pelas ternas (e_1, e_2, e_3) em que e_i é a porta escolhida no passo i descrito acima e se a portas estão numeradas por 1,2 e 3 então definimos um modelo probabilístico discreto com

$$\Omega := \{(1,2,3),(1,3,2),(2,1,3),(2,3,1),(3,1,2),(3,2,1),(1,1,2),(1,1,3),(2,2,1),(2,2,3),(3,3,1),(3,3,2)\}.$$

e probabilidades de acordo com o diagrama de árvore mostrado na figura 1.5 abaixo; um caminho seguido pelo jogador numa rodada do jogo corresponde a um caminho na árvore, a partir da raiz (o ponto mais alto) até uma folha (um dos pontos mais baixos). A primeira ramificação corresponde a escolha de porta para esconder o carro, as segundas ramificações correspondem a escolha do jogador e as terceiras ramificações correspondem a escolha de porta para abrir feita pelo apresentador. Os eventos que interessam, a saber "o jogador vence trocando de porta" e "o jogador vence não trocando de porta", são complementares e denotados por A e $\overline{A}$ respectivamente, de modo que $A = \{(1,2,3),(1,3,2),(2,1,3),(2,3,1),(3,1,2),(3,2,1)\}$ e $\overline{A}$ é dada pelas ternas restantes de Ω. O jogador ganha o carro trocando de porta com probabilidade

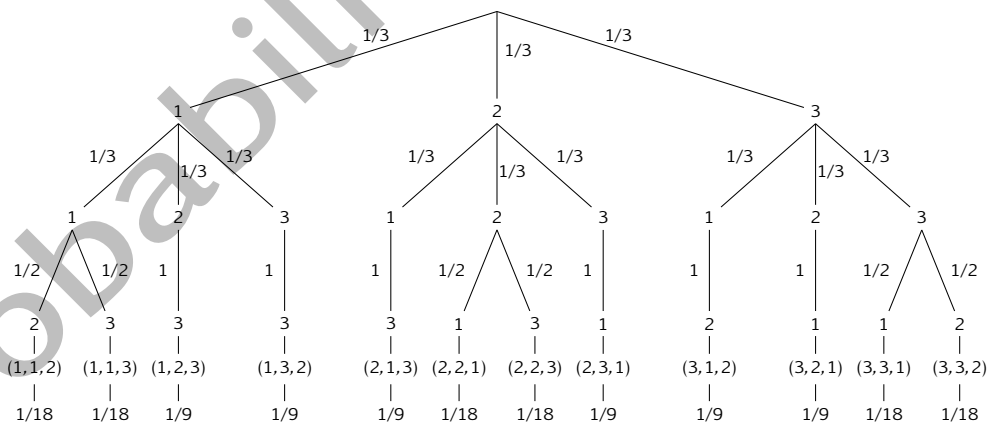


Figura 1.5: diagrama de árvore de um modelo para Monty Hall.

$$\mathbb{P}(A) = \mathbb{P}(\{(1,2,3),(1,3,2),(2,1,3),(2,3,1),(3,1,2),(3,2,1)\}) = \frac{2}{3}$$

portanto, ganha sem trocar de porta com probabilidade $1-2/3=1/3$, que corresponde à probabilidade de ter escolhido a porta certa já na primeira oportunidade de escolha. Portanto, a melhor estratégia é trocar de porta quando é oferecida essa oportunidade. ◊

1.1.2 CONTINUIDADE DE UMA MEDIDA DE PROBABILIDADE

Consideremos novamente o lançamento de uma moeda equilibrada até sair coroa, citado no exemplo 1.13, modelado por $\Omega=\{(\mathrm{Co}),(\mathrm{Ca},\mathrm{Co}),(\mathrm{Ca},\mathrm{Ca},\mathrm{Co}),\ldots,\infty\}$ munido da função de probabilidade $p((c_1,c_2,\ldots,c_i))=2^{-i}$ onde $c_j=\mathrm{Co}$ se $j=i$ e $c_j=\mathrm{Ca}$ caso contrário, e $p(\infty)=0$. Como cada resultado de um lançamento é igualmente provável e não depende dos resultados dos outros lançamentos deve ser intuitivamente claro[3] que devamos assumir que $(\mathrm{Ca},\mathrm{Ca},\ldots,\mathrm{Ca},\mathrm{Co})$ tenha probabilidade de ocorrer igual a

$$\left(\frac{1}{2}\right)^{\text{número de lançamentos}} \tag{1.5}$$

e como cada ponto amostral em $\Omega\setminus\{\infty\}$ está associado a um único inteiro positivo $\mathbb{P}(\Omega\setminus\{\infty\})=\sum_{n\geqslant 1}2^{-n}=1$ o que nos obriga a tomar como 0 a probabilidade para o evento "nunca sair coroa". Nessa seção veremos que essa obrigação condiz com a proposta intuitiva para finitos lançamentos tomada em (1.5) acima.

Consideremos o evento A_n definido por "não sai coroa até o n-ésimo lançamento" que ocorre com probabilidade 2^{-n}. Falando inda de modo intuitivo, queremos que o evento "nunca sair coroa", representado por $\lim_{n\to\infty}A_n$, tenha probabilidade $\lim_{n\to\infty}\mathbb{P}(A_n)=\lim_{n\to\infty}2^{-n}=0$. Essa "passagem ao limite", $\mathbb{P}(\lim_{n\to\infty}A_n)=\lim_{n\to\infty}\mathbb{P}(A_n)$, é garantida pela aditividade enumerável.

Uma sequência qualquer $(A_n\colon n\geqslant 1)$, de eventos em um espaço de probabilidade $(\Omega,\mathcal{A},\mathbb{P})$ é dita **monótona** se vale um dos casos

crescente: $A_1\subset A_2\subset\cdots\subset A_n\subset A_{n+1}\subset\cdots$ e definimos

$$\lim_{n\to\infty}A_n:=\bigcup_{n\geqslant 1}A_n.$$

decrescente: $A_1\supset A_2\supset\cdots\supset A_n\supset A_{n+1}\supset\cdots$ e definimos

$$\lim_{n\to\infty}A_n:=\bigcap_{n\geqslant 1}A_n.$$

Se $(A_n\colon n\geqslant 1)$ é uma sequência crescente, então o limite pode ser escrito como uma união de eventos disjuntos $A_1\cup(A_2\setminus A_1)\cup(A_3\setminus A_2)\cup\cdots$ de modo que se tomamos $A_0:=\varnothing$ então

$$\mathbb{P}\Big(\lim_{n\to\infty}A_n\Big)=\lim_{n\to\infty}\sum_{i=1}^{n}\mathbb{P}(A_i\setminus A_{i-1})=\lim_{n\to\infty}\sum_{i=1}^{n}(\mathbb{P}(A_i)-\mathbb{P}(A_{i-1}))=\lim_{n\to\infty}\mathbb{P}(A_n).$$

No caso em que $(A_n\colon n\geqslant 1)$ é decrescente tomamos os complementos e temos que $(\overline{A_n}\colon n\geqslant 1)$ é crescente, portanto,

$$\mathbb{P}\Big(\lim_{n\to\infty}\overline{A_n}\Big)=\mathbb{P}\left(\bigcup_{n\geqslant 1}\overline{A_n}\right)=\mathbb{P}\left(\overline{\bigcap_{n\geqslant 1}A_n}\right)=\mathbb{P}\Big(\overline{\lim_{n\to\infty}A_n}\Big)=1-\mathbb{P}\Big(\lim_{n\to\infty}A_n\Big)$$

por outro lado

$$\mathbb{P}\Big(\lim_{n\to\infty}\overline{A_n}\Big)=\lim_{n\to\infty}\mathbb{P}\Big(\overline{A_n}\Big)=\lim_{n\to\infty}\Big(1-\mathbb{P}\Big(\overline{A_n}\Big)\Big)=1-\lim_{n\to\infty}\mathbb{P}(A_n)$$

[3] Essa noção intuitiva é formalizada na seção 1.4. Por ora, notemos que em n lançamentos a probabilidade de ocorrer um resultado específico é $(1/2)^n$ quando todos os resultados são igualmente prováveis.

portanto

$$\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}\Big(\lim_{n\to\infty} A_n\Big).$$

Em resumo, se $(A_n\colon n \geqslant 1)$, é uma sequência monótona de eventos então

$$\mathbb{P}\Big(\lim_{n\to\infty} A_n\Big) = \lim_{n\to\infty} \mathbb{P}(A_n). \tag{1.6}$$

De volta ao exemplo do início da seção, consideremos o evento A_n definido por "não sai coroa até o n-ésimo lançamento". Então a sequência $(A_n\colon n \geqslant 1)$ é monótona pois $A_n \supset A_{n-1}$ para todo $n > 1$, portanto,

$$\mathbb{P}(\infty) = \mathbb{P}\Big(\lim_{n\to\infty} A_n\Big) = \lim_{n\to\infty} \mathbb{P}(A_n) = 0.$$

Exemplo 1.15. Consideremos o lançamento de uma moeda equilibrada infinitas vezes, o que pode ser modelado pelo espaço amostral contínuo $\Omega = \{\mathrm{Ca}, \mathrm{Co}\}^{\mathbb{N}}$. Intuitivamente, parece ser claro que devemos esperar que a probabilidade de nunca sair cara deve ser zero: se lançarmos n vezes, a probabilidade de nunca sair cara é 2^{-n}, então no limite a probabilidade é 0. A propriedade dada na equação (1.6) permite a passagem ao limite: A_n é o evento "nos primeiros n lançamentos ocorre pelo menos uma cara"; para $n \geqslant 1$ temos uma sequência monótona de eventos. O limite é o evento "em algum momento, ocorre cara" cuja probabilidade é

$$\lim_{n\to\infty} \mathbb{P}(A_n) = \lim_{n\to\infty} 1 - 2^{-n} = 1.$$

Portanto, de fato, a probabilidade de nunca sair cara é zero. ◊

Vimos que a propriedade dada na equação (1.6) segue dos axiomas A1 (não negatividade), A2 (normalização) e A3 (aditividade enumerável) para uma medida de probabilidade. Se tomarmos por axiomas de probabilidade os axiomas A1, A2, a aditividade finita (isto é, o item 2 da proposição 1.4) e a propriedade dada na equação (1.6) para sequências monótonas de eventos, então vale a aditividade enumerável.

De fato, assumindo os axiomas A1 e A2, a equação (1.6) e o item 2 da proposição 1.4, se $(A_n\colon n \geqslant 1)$ é qualquer sequência de eventos mutuamente exclusivos então

$$B_n := \bigcup_{i\geqslant n} A_i$$

é uma sequência monótona decrescente e $\lim_{n\to\infty} B_n = \varnothing$. Usando a aditividade finita de $\mathbb{P}$

$$\mathbb{P}\left(\bigcup_{i\geqslant 1} A_i\right) = \mathbb{P}\left(\bigcup_{i=1}^{n-1} A_i\right) + \mathbb{P}\left(\bigcup_{i\geqslant n} A_i\right) = \sum_{i=1}^{n-1} \mathbb{P}(A_i) + \mathbb{P}(B_n)$$

e se tomamos o limite quando n tende ao infinito

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

Diferente do que fizemos no exemplo 1.8, página 10, a demonstração para espaços contínuos de que uma função candidata a medida de probabilidade é enumeravelmente aditiva é difícil. Usualmente, o que é feito é verificar que uma função é finitamente aditiva e contínua para toda sequência $(B_n\colon n \geqslant 1)$ tal que $\lim_{n\to\infty} B_n = \varnothing$, isso é suficiente para garantir que é enumeravelmente aditiva e, em geral, uma tarefa mais fácil de realizar.

1.2 CONVENÇÕES DE NOTAÇÃO

Consideremos E um evento aleatório e E_Ω o subconjunto que o modela em $(\Omega, \mathcal{E}, \mathbb{P})$. Denotamos por $\mathbb{P}[E]$ a probabilidade do evento E, isto é, $\mathbb{P}[E] := \mathbb{P}(E_\Omega)$. Por exemplo, suponha que uma moeda equilibrada é lançada até sair coroa, então

a probabilidade do evento "o número de lançamentos é par" com essa convenção fica $\mathbb{P}[\text{o número de lançamentos é par}]$ que é o mesmo que $\mathbb{P}(\{(c_1,\ldots,c_i)\colon i \text{ é par}\})$. Caso haja a necessidade de evidenciar o espaço amostral escreveremos

$$\mathop{\mathbb{P}}_{\Omega}[\mathsf{E}] \quad \text{ou} \quad \mathop{\mathbb{P}}_{\omega\in\Omega}[\omega \text{ satisfaz } \mathsf{E}] \quad \text{ou} \quad \mathop{\mathbb{P}}_{\omega\in\Omega}[\omega \in \mathsf{E}] \tag{1.7}$$

com o mesmo sentido, o de $\mathbb{P}(\mathsf{E}_\Omega)$.

Caso Ω seja finito e a menos que seja dada explicitamente outra medida, então a notação na equação (1.7) significa que estamos assumindo a medida de **probabilidade uniforme**: $\mathbb{P}(\omega) = 1/|\Omega|$ para todo $\omega \in \Omega$. Por exemplo, seja $p(x)$ um polinômio não nulo com coeficientes inteiros e Ω um conjunto finito de números inteiros. A probabilidade de que o sorteio de um elemento de Ω resulte numa raiz do polinômio é descrita por

$$\mathop{\mathbb{P}}_{x\in\Omega}[p(x) = 0]$$

que é a probabilidade do evento $\mathsf{R} = \{\omega \in \Omega\colon p(\omega) = 0\}$ e que, caso não seja dito nada a respeito da medida, é dada por $\mathbb{P}(\mathsf{R}) = |\mathsf{R}|/|\Omega|$.

Nos algoritmos assumiremos a possibilidade de se fazer escolhas aleatórias, ou seja, assumiremos que os algoritmos dispõem de uma fonte de bits aleatórios e escrevemos a instrução

$$a \leftarrow_{\mathrm{R}} \{0,1\}$$

para denotar o fato de que a é uma variável do algoritmo e que após a execução da atribuição $\leftarrow_{\mathrm{R}}$ o valor da variável a é um elemento qualquer de $\{0,1\}$ com probabilidade $1/2$. De um modo geral, se Ω é um conjunto finito, então escrevemos a instrução

$$a \leftarrow_{\mathrm{R}} \Omega$$

chamada de atribuição por uma **escolha aleatória uniforme** em Ω, o que significa que a assume qualquer um dos elementos de Ω com igual probabilidade, a saber $1/|\Omega|$.

1.2.1 SIGILO PERFEITO

Vejamos como aplicação dos conceitos elementares de probabilidade uma das contribuições do grande matemático americano Claude Shannon (1916 – 2001) que é considerado fundador da Teoria da Informação.

Um *sistema de codificação* é definido por uma quina $(\mathcal{P},\mathcal{C},\mathcal{K},\mathcal{E},\mathcal{D})$ de conjuntos, onde $\mathcal{P}$ é o conjunto dos textos comuns (ou legíveis); $\mathcal{C}$ é o conjunto dos textos codificados (ou ilegíveis); $\mathcal{K}$ é o espaço das chaves que são usadas para codificar e decodificar um texto; $\mathcal{E}$ é o conjunto das funções de codificação $\mathsf{E}_k\colon \mathcal{P} \to \mathcal{C}$ para $k \in \mathcal{K}$; $\mathcal{D}$ é o conjunto das funções de decodificação $\mathsf{D}_k\colon \mathcal{C} \to \mathcal{P}$ para $k \in \mathcal{K}$. Essas funções são tais que para cada $e \in \mathcal{K}$ existe $d \in \mathcal{K}$ para as quais vale $\mathsf{D}_d(\mathsf{E}_e(p)) = p$.

Exemplo 1.16 (*cifra de César*). Essa técnica identifica o alfabeto $\{a,b,\ldots,z\}$ com o conjunto $\{0,\ldots,25\}$ dos restos da divisão inteira por 26 e $\mathcal{K} = \mathcal{P} = \mathcal{C} := \{0,\ldots,25\}^\ell$ em que ℓ é o comprimento da mensagem. Para uma chave $e \in \mathcal{K}$ a mensagem $\mathsf{x} = x_1x_2\ldots x_\ell$ é codificada como $\mathsf{E}_e(\mathsf{x}) = y_1y_2\ldots y_\ell$ com $y_i = (x_i + e) \bmod 26$, para todo i, e é decodificada como $\mathsf{D}_e(\mathsf{x}) = y_1y_2\ldots y_\ell$ com $y_i = (x_i - e) \bmod 26$ para todo i. Por exemplo, para a chave $e = 3$ o texto "essaauladasono" é codificado como "hvvddzodgdvrqr".

A cifra de César deve seu nome ao imperador romano Júlio César que a usou com a chave fixa $e = 3$. Tal codificação é facilmente decifrada não oferecendo segurança na comunicação e sua efetividade na época de César deveu-se principalmente ao fato de que a maioria das pessoas eram analfabetas.

No caso $\ell = 1$ conseguimos um cifra segura se escolhemos uma chave aleatoriamente. Tomemos $(\mathcal{K},\mathbb{P})$ com $\mathbb{P}$ a medida uniforme. Dadas duas mensagens legíveis $m_1, m_2 \in \mathcal{P}$ quaisquer e uma mensagem codificada $y \in \mathcal{C}$ qualquer, temos

$$\mathbb{P}(\{k \in \mathcal{K}\colon \mathsf{E}_k(m_1) = y\}) = \frac{1}{26} = \mathbb{P}(\{k \in \mathcal{K}\colon \mathsf{E}_k(m_2) = y\})$$

ou seja, o conhecimento do texto codificado não dá nenhuma informação a respeito do texto legível. No caso $\ell = 2$ a situação é outra. Se $ab, az \in \mathcal{P}$ e $bc \in \mathcal{C}$ então $\mathbb{P}(\{k \in \mathcal{K}\colon \mathsf{E}_k(ab) = bc\}) = 1/26$ pois podemos tomar $k = 1$ e essa é a única chave que codifica ab em bc, por outro lado não existe chave que codifica az em bc de modo que $\mathbb{P}(\{k \in \mathcal{K}\colon \mathsf{E}_k(az) = bc\}) = 0$. Agora, o conhecimento do texto codificado dá alguma informação a respeito do texto legível. ◊

Um sistema de codificação tem **sigilo perfeito** se para quaisquer $m_1,\ m_2 \in \mathcal{P}$ de mesmo comprimento e para todo $\mathsf{C} \in \mathcal{C}$ vale

$$\mathop{\mathbb{P}}_{k\in\mathcal{K}}[\mathsf{E}_k(m_1) = \mathsf{C}] = \mathop{\mathbb{P}}_{k\in\mathcal{K}}[\mathsf{E}_k(m_2) = \mathsf{C}].$$

Pelas convenções de notação, o espaço amostral é o conjunto das chaves, a descrição $[\mathsf{E}_k(m_1) = \mathsf{C}]$ corresponde ao evento $\{k \in \mathcal{K}\colon \mathsf{E}_k(m_1) = \mathsf{C}\}$ formado por todas as chaves que codificam m_1 como C e, também, está implícito que probabilidade de uma chave qualquer é $1/|\mathcal{K}|$.

Em outras palavras, o sigilo perfeito requer que, dado um texto cifrado, qualquer texto legível tem a mesma probabilidade de ser o texto legível subjacente ao texto cifrado.

Exemplo 1.17 (*one-time pad*). O seguinte sistema de codificação, conhecido por *one-time pad*, foi descrito pela primeira vez em 1882 por Frank Miller e reinventado, também patenteado em 1919, por Gilbert Sandford Vernam e aperfeiçoado por Joseph Mauborgne, que reconheceu que se a chave fosse aleatória e usada uma única vez o sistema seria muito seguro.

Tomamos $\mathcal{P} = \mathcal{K} = \mathcal{C} := \{0,1\}^n$ e para uma chave k escolhida previamente definimos

$$\mathsf{E}_k(x) := x \oplus k \qquad \text{e} \qquad \mathsf{D}_k(y) := y \oplus k. \tag{1.8}$$

em que $x \oplus y$ é o *ou exclusivo* (ou soma módulo 2) coordenada-a-coordenada das sequências binárias x e y. O leitor pode verificar que em (1.8) vale $\mathsf{D}_k(\mathsf{E}_k(x)) = x$. ◊

Claude Shannon provou que o *one-time pad* é uma codificação "inviolável" no sentido de que o sistema tem sigilo perfeito. O *one-time pad* não é o único sistema que possui sigilo perfeito, mas foi o primeiro a ser descoberto.

A codificação do texto legível $m \in \mathcal{P}$ usando a chave $k \in \mathcal{K}$ é o texto cifrado $\mathsf{C} = m \oplus k$, logo $m \oplus \mathsf{C} = m \oplus (m \oplus k) = (m \oplus m) \oplus k = k$, portanto, dados m e C existe uma única chave $k \in \mathcal{K}$ tal que $\mathsf{E}_k(m) = \mathsf{C}$, de modo que

$$\mathop{\mathbb{P}}_{k\in\mathcal{K}}[m_1 \oplus k = \mathsf{C}] = \frac{|\{k \in \mathcal{K}\colon k \oplus m_1 = \mathsf{C}\}|}{|\mathcal{K}|} = \frac{1}{|\mathcal{K}|} = \mathop{\mathbb{P}}_{k\in\mathcal{K}}[m_2 \oplus k = \mathsf{C}]$$

para todos os textos legíveis $m_1,\ m_2 \in \mathcal{P}$ e todo texto cifrado $\mathsf{C} \in \mathcal{C}$. Isso demonstra o seguinte resultado.

Teorema 1.18 *O* one-time pad *tem sigilo perfeito.* □

1.2.2 TESTE DE IDENTIDADE POLINOMIAL

Nosso primeiro exemplo de um algoritmo aleatorizado é um teste de identidade entre polinômios: dados dois polinômios p e q, decidir de modo eficiente se eles são idênticos.

Há muitas questões a serem esclarecidas nessa formulação do problema: o que significam "eficiente", "dado um polinômio" e "idênticos"? Isso será tratado mais tarde, na seção 2.2.3, por ora basta saber que "dado um polinômio p" significa que p é um polinômio com coeficientes inteiros dado por uma caixa preta da qual a função polinomial $p(x)$ pode ser avaliada em qualquer número x. Além disso, vamos considerar o problema equivalente de decidir se um polinômio dado f é identicamente nulo. Uma algoritmo que resolve esse problema também resolve o problema original, basta tomarmos $f(x) = p(x) - q(x)$.

Um algoritmo para essa versão do problema funciona do seguinte modo: dado $f(x)$ de grau no máximo d, escolhemos aleatoriamente $a \in \{1, 2, \ldots, 4d\}$ e avaliamos $f(a)$; se $f(a) = 0$, então respondemos *sim*, caso contrário respondemos *não*. A resposta *sim* significa que $f(x)$ é o polinômio nulo e a resposta *não* significa que $f(x)$ não é o polinômio nulo. Esse

algoritmo pode responder errado dependendo de f e da escolha aleatória a e devemos tentar garantir que a probabilidade de ocorrer o erro seja pequena.

Se o polinômio f é nulo então a resposta está sempre certa. Suponhamos que f não é nulo. Nesse caso, se a escolha aleatória a for uma raiz do polinômio então $f(a) = 0$ e a resposta a resposta *sim* dada pelo algoritmo está errada e se a escolha aleatória a não for uma raiz do polinômio então $f(a) \neq 0$ e a resposta a resposta *sim* dada pelo algoritmo está correta. Em resumo, uma resposta *não* dada pelo algoritmo está correta e uma resposta *sim* pode estar errada. O algoritmo descrito abaixo resume essa estratégia.

Instância : d inteiro positivo e f polinômio de grau no máximo d.
Resposta : *não* se f não é nulo, senão *sim* com probabilidade de erro no máximo 1/4.

1 $a \leftarrow_R \{1, 2, \ldots, 4d\}$;
2 **se** $f(a) = 0$ **então responda** *sim*.
3 **senão responda** *não*.

Algoritmo 1: teste de identidade entre polinômios.

Para determinar um limitante para a probabilidade do algoritmo responder errado, seja f um polinômio não nulo e de grau no máximo d e consideremos o evento E formado pelas raízes de f que pertencem ao espaço amostral $\Omega = \{1, 2, \ldots, 4d\}$. Então $|E| \leqslant \text{grau}(f) \leqslant d$ pois f tem no máximo grau(f) raízes distintas pelo teorema fundamental da álgebra. O algoritmo erra se a escolha aleatória resulta num elemento de E, portanto,

$$\mathbb{P}[\text{erro}] \leqslant \frac{1}{4}.$$

Proposição 1.19 *Sejam d um inteiro positivo, f um polinômio não nulo de grau no máximo d e $\Omega \subset \mathbb{Z}$ finito. Então a probabilidade com que uma escolha aleatória uniforme em Ω seja raiz de f é no máximo $d/|\Omega|$, portanto, o algoritmo 1 erra com probabilidade no máximo 1/4.* □

Veremos mais adiante que é possível fazer essa probabilidade arbitrariamente pequena ao custo de um pouco mais computação. Intuitivamente, imagine tal algoritmo sendo executado em dois computadores diferentes concomitantemente. Basta um deles responder *não* para que a resposta definitiva seja *não*. Agora se f é não nula, com que probabilidade todos respondem *sim*? Uma resposta em $E \times E \subset \Omega \times \Omega$ (uma coordenada para cada computador) ocorre com probabilidade no máximo $(1/4)^2$. Se são dez computadores a probabilidade de erro é no máximo $(1/4)^{10} < 10^{-6}$ (em metros é menos que o diâmetro do fio da teia de aranha).

A versão geral do problema da identidade polinomial tem importância central em Complexidade Computacional. Em resumo, podemos dizer que o algoritmo aleatorizado dado acima resolve esse problema e é muito eficiente (executa poucas instruções), enquanto que um algoritmo eficiente que resolve esse problema sem usar aleatoriedade não é conhecido e não se sabe se pode existir, a existência de tal algoritmo determinístico e eficiente teria implicações profundas em Teoria da Computação. Esse fato e outros pontos importante serão detalhados no capítulo 7.

1.3 PROBABILIDADE CONDICIONAL

Lançamos dois dados equilibrados, um deles é vermelho e tem doze faces numeradas de 1 a 12 e o outro preto com vinte faces numeradas de 1 a 20.

Suponhamos que temos a informação de que a soma dos resultados é 15, e isso é tudo que sabemos. Qual é a probabilidade do dado vermelho ter resultado 6 dado que a soma dos resultados dos dois dados é 15?

Definimos um modelo discreto para o experimento tomando o espaço amostral Ω composto pelos $12 \cdot 20 = 240$ pontos amostrais, dados pelos pares ordenados de resultados de cada dado, com a medida uniforme de probabilidade. Sejam Q_Ω o subconjunto que representa o evento "a soma é 15" e S_Ω o evento "o valor do dado vermelho é 6". Se é certo que ocorre Q_Ω então vamos renormalizar a probabilidade de cada ponto amostral $\omega \in Q_\Omega$ para $\mathbb{P}_Q(\omega) = (1/|\Omega|)/(|Q_\Omega|/|\Omega|) = 1/12$ de modo que Q_Ω tenha probabilidade 1. Ademais $S_Q = S_\Omega \cap Q_\Omega = \{(6,9)\}$, portanto $\mathbb{P}_Q(S_Q) = 1/12$. Essa é a probabilidade de ocorrer um 6 vermelho sob a condição de que a soma dos dados é 15.

Em um espaço de probabilidade discreto definido por Ω e $\mathbb{P}$, a **probabilidade condicional** do evento A dado que ocorre o evento E, em que $\mathbb{P}(E) > 0$, é definida por

$$\mathbb{P}(A \mid E) := \frac{\mathbb{P}(A \cap E)}{\mathbb{P}(E)} \tag{1.9}$$

e $\mathbb{P}(A \mid E)$ é lido como a **probabilidade de** A **dado** E.

Por exemplo, se Ω é finito com medida de probabilidade uniforme e $E \neq \varnothing$ então

$$\mathbb{P}(A \mid E) = \frac{\mathbb{P}(A \cap E)}{\mathbb{P}(E)} = \frac{|A \cap E|}{|E|}$$

que é, essencialmente, a medida uniforme em E. No exemplo acima, do par de dados, usando a definição de probabilidade condicional com S e Q eventos do modelo probabilístico discreto $(\Omega, \mathbb{P})$ para o lançamento dos dados

$$\mathbb{P}(S \mid Q) = \frac{\mathbb{P}(S \cap Q)}{\mathbb{P}(Q)} = \frac{|\{(6,9)\}|}{|\{(i,j)\colon i+j=15\}|} = \frac{1}{12}.$$

EXERCÍCIO 1.20. Considere um espaço de probabilidade $(\Omega, \mathcal{A}, \mathbb{P})$ e $E \in \mathcal{A}$ um evento com probabilidade positiva. Verifique que $\mathbb{P}_E(A) := \mathbb{P}(A \mid E)$ é uma medida de probabilidade para os eventos em $\mathcal{A}$ (i.e, satisfaz os axiomas de probabilidade da página 9). Verifique, também, que $(E, \{A \cap E\colon A \in \mathcal{A}\}, \mathbb{P}_E)$ é um espaço de probabilidade.

Exemplo 1.21. Uma urna tem 20 bolas azuis e 10 bolas brancas. Das bolas azuis, 5 têm a letra X e 15 têm a letra Y; das bolas brancas, 1 têm a letra X e 9 tem a letra Y. Uma bola é escolhida ao acaso. Qual é a probabilidade dessa bola ser azul e com a letra X?

Se A representa o evento "bola azul" e X o evento "letra X" então $\mathbb{P}(X \mid A) = 5/20 = 1/4$, que é a proporção de bolas azuis com a letra X. Usando a equação (1.9) podemos deduzir que a probabilidade de sortear uma bola azul e com a letra X é $\mathbb{P}(A \cap X) = \mathbb{P}(X \mid A) \cdot \mathbb{P}(A) = (1/4) \cdot (20/30) = 1/6$. ◊

O TEOREMA DA MULTIPLICAÇÃO E EXPERIMENTOS COMPOSTOS A igualdade $\mathbb{P}(A \cap E) = \mathbb{P}(A \mid E) \cdot \mathbb{P}(E)$ usada no exemplo 1.21 é consequência direta da definição de probabilidade condicional e é conhecida como **teorema da multiplicação** ou regra da multiplicação. Um caso geral dela é dado no exercício 1.24 abaixo. Nos próximos exemplos ilustramos como o teorema da multiplicação pode ser usado para definir um modelo probabilístico para um experimento composto por dois ou mais experimentos aleatórios, como foi o caso do modelo probabilístico para o problema de Monty Hall, exemplo 1.14.

Consideremos três urnas, digamos A, B e C, cada uma com a mesma probabilidade de ser escolhida, 1/3. Em cada uma das urnas há seis bolas, cada uma com a mesma probabilidade de ser escolhida, 1/6. Na urna A temos três bolas pretas e três bolas vermelhas; na urna B temos duas bolas pretas e quatro vermelhas; na urna C todas as bolas são pretas. Uma urna é escolhida aleatoriamente e, em seguida, uma bola é escolhida aleatoriamente e observamos a cor dessa bola. Vamos definir um modelo probabilístico discreto $(\Omega, \mathbb{P})$ para esse *experimento composto* de modo que $\mathbb{P}$ respeita as probabilidades em cada experimento num sentido que ficará claro abaixo.

Temos dois experimentos aleatórios, o primeiro consiste de sortear uma urna e o segundo e sortear uma bola da urna que foi escolhida. Para o primeiro experimento temos o modelo discreto $(\Omega_1, \mathbb{P}_1)$ dado pelo espaço amostral $\Omega_1 = \{A, B, C\}$ e a medida de probabilidade uniforme $\mathbb{P}_1$. Para o segundo experimento tomamos $(\Omega_2, \mathbb{P}_2)$ com o espaço amostral $\Omega_2 = \{V, P\}$ em que usamos os eventos atômicos (pontos amostrais) $V \in \Omega_2$ para modelar "bola vermelha" e $P \in \Omega_2$ para modelar "bola preta" e cujas probabilidades dependem da urna e são dadas na tabela 1.2 abaixo, mas que por abuso de notação escrevemos $\mathbb{P}_2$ em todos os casos.

urna	$\mathbb{P}_2(V)$	$\mathbb{P}_2(P)$
A	1/2	1/2
B	2/3	1/3
C	0	1

Tabela 1.2: probabilidade dos eventos "bola vermelha" e "bola preta" em cada urna.

Um espaço amostral para o experimento composto pelos dois sorteios é $\Omega := \Omega_1 \times \Omega_2 = \{A, B, C\} \times \{V, P\}$. Agora, por exemplo, o evento "urna A" é modelado no primeiro experimento e no experimento composto por, respectivamente

$$U_{\Omega_1} = \{A\} \quad \text{e} \quad U_\Omega = \{A\} \times \Omega_2 = \{(A, V), (A, P)\}$$

de modo que para a medida $\mathbb{P}_\Omega$ em Ω temos

$$\mathbb{P}(U_\Omega) = \mathbb{P}(\{A\} \times \Omega_2) = \mathbb{P}_1(U_{\Omega_1}) = \frac{1}{3}.$$

O evento "bola preta" é modelado em Ω por $E_\Omega = \Omega_1 \times \{P\} = \{(A, P), (B, P), (C, P)\}$ de modo que

$$\mathbb{P}(E_\Omega) = \mathbb{P}(\Omega_1 \times \{P\}) = \mathbb{P}((A, P)) + \mathbb{P}((B, P)) + \mathbb{P}((C, P))$$

entretanto, diferente do caso anterior, o resultado do segundo experimento depende do resultado do primeiro; o que conhecemos são as probabilidades condicionais de "cor" dado "urna". O ponto amostral (A, P) de Ω é dado por

$$(\Omega_1 \times \{P\}) \cap (\{A\} \times \Omega_2) = E_\Omega \cap U_\Omega \tag{1.10}$$

e tem probabilidade dada pelo teorema da multiplicação da seguinte forma

$$\mathbb{P}(E_\Omega \cap U_\Omega) = \mathbb{P}(E_\Omega \mid U_\Omega)\mathbb{P}(U_\Omega) = \mathbb{P}_2(E_{\Omega_2})\mathbb{P}_1(U_{\Omega_1}) = \frac{1}{2} \cdot \frac{1}{3} \tag{1.11}$$

pois dado que ocorre "urna A", a probabilidade de "bola preta" é $\mathbb{P}(E_\Omega \mid U_\Omega) = \mathbb{P}_2(P_{\Omega_2}) = 1/2$.

Podemos determinar de maneira análoga a probabilidade de todo ponto amostral de Ω, cada um é dado por um interseção e pode ser escrito como na equação (1.10) e a probabilidade é calculada pelo teorema da multiplicação como na equação (1.11).

Quando o espaço amostral é pequeno, como nesse exemplo, pode ser conveniente descrevermos o modelo probabilístico através de um diagrama de árvore como o da figura 1.6 abaixo e como já fizemos para Monty Hall (figura 1.5, pág. 14). O diagrama de árvore da figura 1.6 representa cada etapa do experimento em um nível da árvore, com as respectivas probabilidades nas ramificações correspondentes aos resultados de cada etapa. A partir do segundo nível essas probabilidades são condicionadas ao que ocorreu na etapas anteriores. Uma maneira prática de atribuir probabilidade a um ponto amostral é tomar o produto das probabilidades no caminho até ele nessa árvore, por exemplo, $\mathbb{P}((A, P)) = 1/3 \cdot 1/2$.

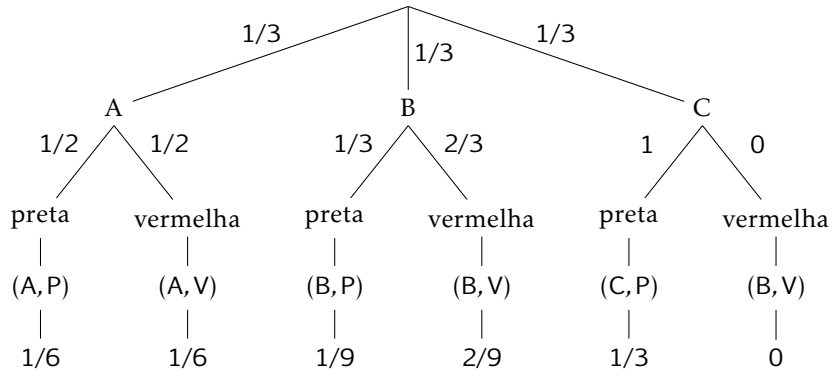


Figura 1.6: diagrama de árvore.

Como o modelo é discreto, estendemos a probabilidade para qualquer subconjunto de $\Omega_1 \times \Omega_2$ somando a probabilidade de seus elementos. Por exemplo, o evento dado por "a bola sorteada é preta" ocorre com probabilidade

$$\mathbb{P}((A,P)) + \mathbb{P}((B,P)) + \mathbb{P}((C,P)) = \frac{11}{18}.$$

Notemos que essa probabilidade não depende do número de bolas pretas na urna C, portanto, embora a medida de probabilidade em cada experimento seja uniforme a probabilidade de "a bola sorteada é preta" não é a quantidade de bolas pretas dividido pelo número total de bolas, que é um erro cometido frequentemente nesse caso. ◊

Exemplo 1.22. Numa cômoda há três gavetas e em cada gaveta um par de meias. Na primeira gaveta há um par de meias brancas, na segunda um par de meias pretas e na terceira gaveta um par com um pé de cada cor. Uma gaveta é escolhida uniformemente e, sem olhar para o interior da gaveta, um pé de meia é escolhido uniformemente e em seguida a gaveta é fechada. O pé de meia retirado é branco. Qual a probabilidade de o outro pé que ficou sozinho na gaveta ser preto?

Vamos denotar por B o evento "retirou uma meia branca" e por T o evento "ficou uma meia preta", ambos eventos do experimento composto por dois experimentos realizados consecutivamente. Queremos determinar $\mathbb{P}(T \mid B) = \mathbb{P}(T \cap B)\mathbb{P}(B)$.

Em metade dos resultados possíveis a meia escolhida é branca, portanto $\mathbb{P}(B) = 1/2$. A probabilidade de ocorrer ambos os eventos é a probabilidade de um ponto amostral específico que corresponde a abrir a terceira gaveta e retirar uma meia branca, o que ocorre com probabilidade $\mathbb{P}(T \cap B) = \mathbb{P}(B \mid T)\mathbb{P}(T)$ pela regra da multiplicação. A probabilidade condicional $\mathbb{P}(B \mid T)$ corresponde a sortear no segundo experimento uma meia branca na terceira gaveta, o que ocorre com probabilidade 1/2 e $\mathbb{P}(T)$ corresponde a probabilidade de sortear a terceira gaveta no primeiro experimento, o que ocorre com probabilidade 1/3. Logo, $\mathbb{P}(T \cap B) = 1/2 \cdot 1/3 = 1/6$. Portanto, segue da definição que $\mathbb{P}(T \mid B) = 1/3$. ◊

EXERCÍCIO 1.23. Verifique a seguinte igualdade para o teorema da multiplicação com três eventos

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A) \cdot \mathbb{P}(B \mid A) \cdot \mathbb{P}(C \mid A \cap B) \tag{1.12}$$

e identifique seu uso no exemplo 1.14, o modelo probabilístico para o problema de Monty Hall.

*EXERCÍCIO 1.24 (*teorema da multiplicação*).* Sejam $A_1, A_2, \ldots, A_n$ eventos de um modelo probabilístico. Prove que

$$\mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right) = \mathbb{P}(A_1) \prod_{i=2}^{n} \mathbb{P}\left(A_i \,\middle|\, \bigcap_{j=1}^{i-1} A_j\right)$$

sempre que as probabilidades condicionais estão definidas (é suficiente pedir que $\mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_{n-1}) > 0$).

EXERCÍCIO 1.25. Sejam A, B e C eventos de um mesmo espaço amostral. Verifique que vale a seguinte igualdade

$$\mathbb{P}(\,C\cap A\mid B\,)=\mathbb{P}(\,C\mid A\cap B\,)\mathbb{P}(\,A\mid B\,) \tag{1.13}$$

sempre que as condicionais estão definidas.

1.3.1 OS TEOREMAS DA PROBABILIDADE TOTAL E DE BAYES

Se E e A são eventos, com $0<\mathbb{P}(E)<1$, então o evento A ocorre se, e somente se, ocorre (A e E) ou (A e $\overline{E}$) e esses eventos entre parênteses são disjuntos; mais que isso $\{A\cap E, A\cap\overline{E}\}$ é uma partição de A, portanto,

$$\begin{aligned}\mathbb{P}(A)&=\mathbb{P}\Big((A\cap E)\cup(A\cap\overline{E})\Big)\\&=\mathbb{P}(A\cap E)+\mathbb{P}(A\cap\overline{E})\\&=\mathbb{P}(\,A\mid E\,)\mathbb{P}(E)+\mathbb{P}(\,A\mid\overline{E}\,)\mathbb{P}(\overline{E}).\end{aligned} \tag{1.14}$$

No problema de Monty Hall se o convidado fica com a porta que escolheu inicialmente, então a probabilidade de ganhar um carro é 1/3, que é a probabilidade dele ter escolhido a porta certa logo de início. Agora, vamos supor que o convidado troca de porta. Nesse caso, denotamos por A o evento "ganha o carro" e por E o evento "a porta escolhida na primeira etapa esconde o carro". Claramente, $\mathbb{P}(\,A\mid E\,)=0$ e $\mathbb{P}(E)=1/3$. Se a primeira escolha não era a correta então o convidado ganha o carro, ou seja, $\mathbb{P}(\,A\mid\overline{E}\,)=1$. Com isso temos por (1.14)

$$\mathbb{P}(A)=\mathbb{P}(\,A\mid E\,)\mathbb{P}(E)+\mathbb{P}(\,A\mid\overline{E}\,)\mathbb{P}(\overline{E})=0\frac{1}{3}+1\frac{2}{3}=\frac{2}{3}$$

portanto, é melhor trocar de porta.

O caso geral dessa igualdade é conhecido como o Teorema da Probabilidade Total. Segue da dedução acima e usando indução em n que se $\{E_1,E_2,\ldots,E_n\}$ é um conjunto de eventos que particionam o espaço amostral Ω com $\mathbb{P}(E_i)>0$[1] para todo $i\geqslant 1$, então vale

$$\mathbb{P}(A)=\sum_{i=1}^{n}\mathbb{P}(A\cap E_i)=\sum_{i=1}^{n}\mathbb{P}(\,A\mid E_i\,)\mathbb{P}(E_i) \tag{1.15}$$

para qualquer evento A. Deixamos a prova por conta do leitor, abaixo provamos um pouco mais, considerando partições enumeráveis.

Teorema 1.26 (Teorema da Probabilidade Total) *Seja* $\{E_i : i\in\mathbb{N}\}$ *um conjunto de eventos que particionam o espaço amostral* Ω *com* $\mathbb{P}(E_i)>0$ *para todo* i. *Então*

$$\mathbb{P}(A)=\sum_{i\geqslant 1}\mathbb{P}(A\cap E_i)=\sum_{i\geqslant 1}\mathbb{P}(\,A\mid E_i\,)\mathbb{P}(E_i) \tag{1.16}$$

para qualquer evento A.

Demonstração. Os conjuntos $A\cap E_i$, para $i\in\mathbb{N}$, são disjuntos dois a dois e a da união deles resulta A. A primeira igualdade em (1.16) segue da aditividade enumerável. A segunda igualdade segue de $\mathbb{P}(A\cap E_i)=\mathbb{P}(\,A\mid E_i\,)\mathbb{P}(E_i)$ para todo i. □

Exemplo 1.27 (Ross, 2010). As seguradoras de automóveis classificam motoristas em *propensos a acidentes* e *não propensos a acidentes*; estimam que os propensos são 30% da população. As estatísticas mostram que os propensos a acidentes se envolvem em acidente no período de um ano com probabilidade 0,4 e os não propensos a acidentes se envolvem em acidente no período de um ano com probabilidade 0,2. Denotemos por A o evento definido pelos motoristas propensos a acidentes. Então a probabilidade de um novo segurado se envolver em acidente em um ano é

$$\mathbb{P}(\,A_1\,)=\mathbb{P}(\,A_1\mid A\,)\mathbb{P}(A)+\mathbb{P}(\,A_1\mid\overline{A}\,)\mathbb{P}(\overline{A})=0{,}4\cdot 0{,}3+0{,}2\cdot 0{,}7=0{,}26$$

e se um novo segurado se envolve em acidente nesse prazo, a probabilidade dele ser propenso a acidentes é

$$\mathbb{P}(\,A \mid A_1\,) = \frac{\mathbb{P}(A_1 \cap A)}{\mathbb{P}(A_1)} = \frac{\mathbb{P}(\,A_1 \mid A\,)\mathbb{P}(A)}{\mathbb{P}(A_1)} = \frac{0{,}4 \cdot 0{,}3}{0{,}26} = \frac{6}{13}.$$

Portanto, a probabilidade de ser propenso a acidente dado que se envolve em acidente em um ano é 0,46, aproximadamente. Dado que o motorista não se envolve em acidente em um ano, a probabilidade de ser propenso a acidente é $\mathbb{P}(\,A \mid \overline{A}_1\,) \approx 0{,}24$. ◊

Exemplo 1.28 (*urna de Pólya*). Uma urna contém duas bolas, uma branca e uma preta. Em cada instante $t \in \{1, 2, \ldots\}$ sorteamos uma bola da urna. A bola sorteada é devolvida para a urna junto com uma outra bola da mesma cor dessa sorteada. Assim, o t-ésimo sorteio ($t \geqslant 1$) ocorre com $t+1$ bolas na urna. Neste exemplo nós vamos calcular a a probabilidade com que uma bola preta é sorteada em cada instante.

Seja P_t ($t \geqslant 1$) o evento "a t-ésima bola sorteada é preta"; se não é sorteada uma bola preta então é sorteada uma bola branca, cujo evento é denotado por B_t. Certamente,

$$\mathbb{P}(P_1) = \frac{1}{2}.$$

Pelo teorema de probabilidade total, como na equação (1.14), $\mathbb{P}(P_2) = \mathbb{P}(\,P_2 \mid P_1\,)\mathbb{P}(P_1) + \mathbb{P}(\,P_2 \mid B_1\,)\mathbb{P}(B_1)$ e se ocorre P_1, então para o segundo sorteio há 2 bolas pretas dentre 3 bolas, portanto, $\mathbb{P}(\,P_2 \mid P_1\,) = 2/3$ e, analogamente, $\mathbb{P}(\,P_2 \mid B_1\,) = 1/3$, de modo que

$$\mathbb{P}(P_2) = \frac{2}{3} \cdot \frac{1}{2} + \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{2}.$$

Para computar $\mathbb{P}(P_3)$ precisamos de um pouco mais de esforço. Pelo teorema da probabilidade total, equação (1.15), temos

$$\mathbb{P}(P_3) = \mathbb{P}((P_1 \cap P_2) \cap P_3) + \mathbb{P}((P_1 \cap B_2) \cap P_3) + \mathbb{P}((B_1 \cap P_2) \cap P_3) + \mathbb{P}((B_1 \cap B_2) \cap P_3).$$

e cada termo dessa soma pode ser computado pelo teorema da multiplicação (especificamente, equação (1.12) na página 22), por exemplo

$$\mathbb{P}(P_1 \cap P_2 \cap P_3) = \mathbb{P}(P_1)\mathbb{P}(\,P_2 \mid P_1\,)\mathbb{P}(\,P_3 \mid P_1 \cap P_2\,) = \frac{1}{2} \cdot \frac{2}{3} \cdot \frac{3}{4}$$

de modo que temos

$$\begin{aligned} \mathbb{P}(P_3) &= \mathbb{P}((P_1 \cap P_2) \cap P_3) + \mathbb{P}((P_1 \cap B_2) \cap P_3) + \mathbb{P}((B_1 \cap P_2) \cap P_3) + \mathbb{P}((B_1 \cap B_2) \cap P_3) \\ &= \frac{1}{2} \cdot \frac{2}{3} \cdot \frac{3}{4} + \frac{1}{2} \cdot \frac{1}{3} \cdot \frac{2}{4} + \frac{1}{2} \cdot \frac{1}{3} \cdot \frac{2}{4} + \frac{1}{2} \cdot \frac{2}{3} \cdot \frac{1}{4} = \frac{1}{2}. \end{aligned} \tag{1.17}$$

Até aqui, $\mathbb{P}(P_1) = \mathbb{P}(P_2) = \mathbb{P}(P_3) = 1/2$.

Notemos que $\mathbb{P}(B_1 \cap B_2 \cap P_3) = \mathbb{P}(P_1 \cap P_2 \cap B_3)$ e que $\mathbb{P}(B_1 \cap P_2 \cap P_3) = \mathbb{P}(P_1 \cap B_2 \cap B_3)$ de modo que, substituindo em na equação (1.17) acima

$$\mathbb{P}(P_3) = \mathbb{P}(P_1 \cap (P_2 \cap P_3)) + \mathbb{P}(P_1 \cap (B_2 \cap P_3)) + \mathbb{P}(P_1 \cap (P_2 \cap B_3)) + \mathbb{P}(P_1 \cap (B_2 \cap B_3))$$

que por (1.15) é $\mathbb{P}(P_1)$ e, então, $\mathbb{P}(P_1) = \mathbb{P}(P_3)$. Tal simetria vale para qualquer t de modo que $\mathbb{P}(P_t) = \mathbb{P}(P_1)$ para todo $t \geqslant 1$. Vamos demonstrar esse fato.

Consideremos uma sequência de eventos $X_1, X_2, \ldots, X_{t-1}, P_t$ em que para cada i, $1 \leqslant i < t$, temos $X_t \in \{P_t, B_t\}$. As 2^{t-1} possíveis sequências de eventos $X_1, \ldots, X_{t-1}$ particionam o espaço amostral de modo que, pelo teorema da probabilidade total e o caso geral do teorema da multiplicação (exercício 1.24 na página 22), a probabilidade de P_t é

$$\sum_{(X_1,\ldots,X_{t-1})} \mathbb{P}(X_1 \cap X_2 \cap \cdots \cap X_{t-1} \cap P_t) = \sum_{(X_1,\ldots,X_{t-1})} \mathbb{P}(X_1) \cdot \left(\prod_{i=2}^{t-1} \mathbb{P}\left(X_i \;\middle|\; \bigcap_{j=1}^{i-1} X_j \right) \right) \cdot \mathbb{P}\left(P_t \;\middle|\; \bigcap_{j=1}^{t-1} X_j \right) \tag{1.18}$$

em que a soma é sobre todas as 2^{t-1} sequências $X_1,\ldots,X_{t-1}$ de eventos. Os somandos no lado direito da equação (1.18) são

$$\mathbb{P}(X_1)\mathbb{P}(X_2 \mid X_1)\mathbb{P}(X_3 \mid X_1 \cap X_2)\cdots\mathbb{P}(P_t \mid X_1 \cap \cdots \cap X_{t-1}) = \prod_{i=1}^{t} \frac{n_i}{i+1} \tag{1.19}$$

onde n_i é a quantidade de bolas da cor X_i sorteada no i-ésimo sorteio e os denominadores são $i+1$ porque em cada sorteio o número total de bolas aumenta de 1.

Fixemos um instante t e suponhamos que até esse instante tenham sido sorteadas m bolas brancas e, portanto, $t-m$ bolas pretas. Sejam $1 \leqslant t_1 < t_2 < \cdots < t_m \leqslant t$ os instantes em que ocorrem sorteio de bola branca. Quando foi realizado o primeiro sorteio de uma bola branca, havia uma bola branca de modo que $n_{t_1} = 1$, no segundo sorteio $n_{t_2} = 2$, e assim por diante, até o último sorteio de bola branca no instante t_m quando $n_{t_m} = m$. No momentos em que não foram sorteados bolas brancas foram sorteados bolas pretas, sejam $1 \leqslant s_1 < s_2 < \cdots < s_{t-m} \leqslant t$ tais instantes em que ocorrem sorteio de bolas pretas. De modo análogo temos que $n_{s_1} = 1$, $n_{s_2} = 2$, $\ldots$, $n_{s_{t-m}} = t-m$.

Agora, notemos que nos numeradores no lado direito da equação (1.19) ocorrem os números $1,2,\ldots,m$ e $1,2,\ldots,t-m$ de modo que o fator determinante no cálculo é a probabilidade de ocorrer m sorteios de bolas brancas e $t-m$ sorteios de bolas pretas, a ordem não importa. Dessa observação concluímos que os somandos no lado direito da equação (1.18) são

$$\frac{1}{2}\cdot\frac{2}{3}\cdots\frac{m}{m+1}\cdot\frac{1}{m+2}\cdot\frac{2}{m+3}\cdots\frac{t-m}{t+1} = \frac{m!(t-m)!}{(t+1)!} = \frac{1}{(t+1)\binom{t}{m}}. \tag{1.20}$$

Há $\binom{t-1}{m}$ sequências $X_1,X_2,\ldots,X_{t-1}$ de eventos com m posições correspondentes ao sorteio de bola branca e cada uma tem probabilidade dada pela equação (1.20), portanto

$$\mathbb{P}(P_t) = \sum_{m=0}^{t-1}\binom{t-1}{m}\frac{1}{(t+1)\binom{t}{m}} = \sum_{m=0}^{t-1}\frac{1}{(t+1)}\frac{t-m}{t} = \frac{1}{t(t+1)}\sum_{m=1}^{t} m = \frac{1}{2}$$

ou seja, $\mathbb{P}(P_t) = 1/2$ para todo $t \geqslant 1$. ◊

Um fato interessante que deduzimos do exemplo acima é que usando equação (1.20) podemos concluir que a probabilidade de haver m bolas brancas após t-ésimo sorteio é

$$\mathbb{P}[\text{há } m \text{ bolas brancas após } t\text{-ésimo sorteio}] = \binom{t}{m}\frac{1}{(t+1)\binom{t}{m}} = \frac{1}{t+1}$$

que não depende de m.

Teorema de Bayes Seja $E_1,\ldots,E_n$ uma partição do espaço amostral. Se soubermos que o evento A ocorre, então qual é a probabilidade (condicionada) com que E_j tenha ocorrido? Para todo j

$$\mathbb{P}(E_j \mid A) = \frac{\mathbb{P}(A \mid E_j)\mathbb{P}(E_j)}{\mathbb{P}(A)}$$

usando o teorema da probabilidade total obtemos

$$\mathbb{P}(E_j \mid A) = \frac{\mathbb{P}(A \mid E_j)\mathbb{P}(E_j)}{\sum_{i=1}^{n}\mathbb{P}(A \mid E_i)\mathbb{P}(E_i)}$$

para todo evento com probabilidade positiva A. Esse resultado é conhecido como teorema de Bayes.

Teorema 1.29 (Teorema de Bayes) *Se $\{E_i : i \in \mathbb{N}\}$ é uma partição do espaço amostral com $\mathbb{P}(E_i) > 0$ para todo i e $\mathbb{P}(A) > 0$, então*

$$\mathbb{P}(E_j \mid A) = \frac{\mathbb{P}(A \mid E_j)\mathbb{P}(E_j)}{\sum_{i\geqslant 1}\mathbb{P}(A \mid E_i)\mathbb{P}(E_i)}.$$

Suponha que *probabilite* é um vírus que afeta 10% da população de estudantes universitários. Um professor de Probabilidade aplica um teste que detecta *probabilite* mas eventualmente se engana: 3% de falsos positivos e 1% de falsos negativos. Se for detectado *probabilite* em um indivíduo escolhido ao acaso, qual é a probabilidade que ele tenha o vírus? Queremos determinar $\mathbb{P}(\,\mathsf{B}\mid\mathsf{A}\,)$ onde A é o evento "foi detectado *probabilite*" e B o evento "tem *probabilite*". Usando o teorema de Bayes com $\mathbb{P}(\,\mathsf{A}\mid\mathsf{B}\,)=0{,}99$, pois $\mathbb{P}(\,\overline{\mathsf{A}}\mid\mathsf{B}\,)=0{,}01$ é a chance de ocorrer um falso negativo, e $\mathbb{P}(\,\mathsf{A}\mid\overline{\mathsf{B}}\,)=0{,}03$ é a chance de ocorrer um falso positivo

$$\mathbb{P}(\,\mathsf{B}\mid\mathsf{A}\,)=\frac{\mathbb{P}(\,\mathsf{A}\mid\mathsf{B}\,)\mathbb{P}(\mathsf{B})}{\mathbb{P}(\,\mathsf{A}\mid\mathsf{B}\,)\mathbb{P}(\mathsf{B})+\mathbb{P}(\,\mathsf{A}\mid\overline{\mathsf{B}}\,)\mathbb{P}(\overline{\mathsf{B}})}=\frac{0{,}99\cdot 0{,}1}{0{,}99\cdot 0{,}1+0{,}03\cdot 0{,}9}\approx 0{,}78.$$

Agora, supondo que *probabilite* seja uma contaminação muito rara, que afeta só 1,05% da população universitária, e que o teste seja um pouco mais acurado, só há 1% de chance de falsos positivos e falsos negativos, então

$$\mathbb{P}(\,\mathsf{B}\mid\mathsf{A}\,)=\frac{0{,}99\cdot 0{,}0105}{0{,}99\cdot 0{,}0105+0{,}01\cdot 0{,}9895}\approx 0{,}51$$

logo o teste é essencialmente tão efetivo quanto decidir lançando uma moeda. Se o vírus for ainda mais raro, digamos que apena 0,5% da população universitária tenha o vírus. Assim,

$$\mathbb{P}(\,\mathsf{B}\mid\mathsf{A}\,)=\frac{0{,}99\cdot 0{,}005}{0{,}99\cdot 0{,}005+0{,}01\cdot 0{,}995}\approx 0{,}34$$

ou seja, se o teste do professor detectou *probabilite* em um estudante é duas vezes mais provável que o indivíduo não tenha *probabilite*. Esse resultado aparentemente paradoxal ocorre porque o número de indivíduos não contaminados pelo *probabilite* é muito grande em relação ao número de contaminados, de modo que a quantidade de falsos positivos supera a quantidade positivos verdadeiros. Se 100.000 indivíduos forem testados, esperamos que aproximadamente 99.500 não tenham *probabilite* mas que ocorram $0{,}01\cdot 99.500\approx 1.000$ falsos positivos; também, esperamos que 500 indivíduos tenham *probabilite* e deles $0{,}99\cdot 500\approx 500$ verdadeiros positivos, portanto, só 1/3 dos resultados positivos são genuínos.

Exemplo 1.30 (*filtro bayesiano para mensagens eletrônicas indesejáveis* (Spam)). O uso de técnicas baseadas no teorema de Bayes para classificar mensagens eletrônicas (*emails*) surgiu em 1996 num trabalho de Jason Rennie chamado *Ifile* e ganhou impulso com o ensaio de Graham (2002), que descrevemos abaixo sem entrar em detalhes de implementação.

Previamente identificamos algumas características das mensagens que estão classificadas em dois conjuntos, as que são *spam* e as que não são *spam*, da seguinte forma. A frequência com que ocorre uma palavra no conjunto das mensagens que são *spam* define uma probabilidade da palavra condicionada à mensagem ser um *spam* e as palavras características de *spam* são as mais relevantes de acordo com essas probabilidades. Por exemplo, nas minhas mensagens muitos dos *spam*s têm a palavra *watch* enquanto que muitos dos não *spam*s têm a palavra "reunião"; a maioria das mensagens têm a palavra "a", tantos *spam*s quanto não *spam*s, logo "a" não deve ser uma característica classificatória; separamos as palavras tais que $\mathbb{P}[\,\text{palavra}\mid \mathit{spam}\,]$ seja bem maior que 1/2 e $\mathbb{P}[\,\text{palavra}\mid \text{não }\mathit{spam}\,]$ seja bem menor que 1/2. Ao final temos algumas características classificatórias, digamos n características, que podem estar ou não estar presentes nas mensagens futuras e que vão ajudar a classificá-las.

Dadas as n características, cada mensagem fica associada uma sequência binária de $\Omega := \{0,1\}^n$ em que cada coordenada da sequência correspondente indica se a mensagem tem ou não tem uma determinada característica. A primeira coordenada, especificamente, é 1 se a mensagem é *spam* e 0 caso, contrário. Assim, $(1,0,0,1,1)$ corresponde a uma mensagem que é *spam* não tem as características 2 e 3, mas tem as características 4 e 5. Denotemos por S o evento "*spam*" e por C_i o evento "tem a característica i". Na classificação prévia contamos a quantidade k_i de mensagens *spam* que têm a característica i dentre as K mensagens classificadas. Também, determinamos a quantidade ℓ_i de mensagens não *spam* que têm a característica i dentre as L mensagens classificadas. Com essa informação determinamos

$$\mathbb{P}(\,\mathsf{C}_i\mid\mathsf{S}\,)=\frac{k_i}{\mathsf{K}}\quad\text{e}\quad\mathbb{P}(\,\mathsf{C}_i\mid\overline{\mathsf{S}}\,)=\frac{\ell_i}{\mathsf{L}}$$

para cada característica $i > 1$. Pelo teorema de Bayes, a probabilidade de uma mensagem que apresenta a característica i ser *spam* é

$$p_i := \mathbb{P}(\,\mathsf{S} \mid \mathsf{C}_i\,) = \frac{\mathbb{P}(\,\mathsf{C}_i \mid \mathsf{S}\,)\mathbb{P}(\mathsf{S})}{\mathbb{P}(\,\mathsf{C}_i \mid \mathsf{S}\,)\mathsf{P}(\mathsf{S}) + \mathbb{P}(\,\mathsf{C}_i \mid \overline{\mathsf{S}}\,)\mathbb{P}(\overline{\mathsf{S}})}.$$

Se assumimos que, a priori, temos a mesma chance de receber um *spam* quanto um não *spam* então $\mathbb{P}(\mathsf{S}) = \mathbb{P}(\overline{\mathsf{S}}) = 1/2$ e assumindo $\mathsf{K} = \mathsf{L}$, para simplificar, a equação acima se resume a

$$p_i = \frac{(k_i/\mathsf{K})\mathbb{P}(\mathsf{S})}{(k_i/\mathsf{K})\mathbb{P}(\mathsf{S}) + (\ell_i/\mathsf{L})\mathbb{P}(\overline{\mathsf{S}})} = \frac{k_i}{k_i + \ell_i}.$$

Recebida uma mensagem como classificá-la? Determinamos quais das n características estão presentes na mensagem, digamos que para algum subconjunto de índices I a mensagem tem C_i para todo $i \in \mathsf{I}$ e com essa informação calculamos $\mathbb{P}(\,\mathsf{S} \,\big|\, \bigcap_{i\in\mathsf{I}} \mathsf{C}_i\,)$. Se essa probabilidade for maior que um limiar $\varepsilon \in (0,1)$ estabelecido, então a mensagem recebida é classificada como *spam*, senão é classificada como não *spam*. No que segue vamos provar que a probabilidade dessa mensagem ser *spam* é dada por

$$\mathbb{P}\left(\,\mathsf{S} \,\middle|\, \bigcap_{i\in\mathsf{I}} \mathsf{C}_i\right) = \frac{\prod_{i\in\mathsf{I}} p_i}{\prod_{i\in\mathsf{I}} p_i + \prod_{i\in\mathsf{I}}(1-p_i)}. \tag{1.21}$$

Para isso vamos assumir que valem

$$\mathbb{P}\left(\,\bigcap_{i\in\mathsf{I}} \mathsf{C}_i \,\middle|\, \mathsf{S}\right) = \prod_{i\in\mathsf{I}} \mathbb{P}(\,\mathsf{C}_i \mid \mathsf{S}\,) \tag{1.22}$$

$$\mathbb{P}\left(\,\bigcap_{i\in\mathsf{I}} \mathsf{C}_i \,\middle|\, \overline{\mathsf{S}}\right) = \prod_{i\in\mathsf{I}} \mathbb{P}(\,\mathsf{C}_i \mid \overline{\mathsf{S}}\,), \tag{1.23}$$

para todo $\mathsf{I} \subset \{2,3,\ldots,n\}$, o que pode não ser uma hipótese muito realista. Usando a definição de probabilidade condicional

$$\mathbb{P}\left(\,\mathsf{S} \,\middle|\, \bigcap_{i\in\mathsf{I}} \mathsf{C}_i\right) = \frac{\mathbb{P}(\,\bigcap_{i\in\mathsf{I}} \mathsf{C}_i \mid \mathsf{S}\,)\mathbb{P}(\mathsf{S})}{\mathbb{P}(\,\bigcap_{i\in\mathsf{I}} \mathsf{C}_i\,)}$$

e da lei da probabilidade total

$$\frac{\mathbb{P}(\,\bigcap_{i\in\mathsf{I}} \mathsf{C}_i \,\big|\, \mathsf{S}\,)\mathbb{P}(\mathsf{S})}{\mathbb{P}(\,\bigcap_{i\in\mathsf{I}} \mathsf{C}_i\,)} = \frac{\mathbb{P}(\,\bigcap_{i\in\mathsf{I}} \mathsf{C}_i \,\big|\, \mathsf{S}\,)\mathbb{P}(\mathsf{S})}{\mathbb{P}(\,\bigcap_{i\in\mathsf{I}} \mathsf{C}_i \,\big|\, \mathsf{S}\,)\mathbb{P}(\mathsf{S}) + \mathbb{P}(\,\bigcap_{i\in\mathsf{I}} \mathsf{C}_i \mid \overline{\mathsf{S}}\,)\mathbb{P}(\overline{\mathsf{S}})}$$

e pelas equações (1.22) e (1.23)

$$\frac{\mathbb{P}(\,\bigcap_{i\in\mathsf{I}} \mathsf{C}_i \,\big|\, \mathsf{S}\,)\mathbb{P}(\mathsf{S})}{\mathbb{P}(\,\bigcap_{i\in\mathsf{I}} \mathsf{C}_i \,\big|\, \mathsf{S}\,)\mathbb{P}(\mathsf{S}) + \mathbb{P}(\,\bigcap_{i\in\mathsf{I}} \mathsf{C}_i \mid \overline{\mathsf{S}}\,)\mathbb{P}(\overline{\mathsf{S}})} = \frac{\prod_{i\in\mathsf{I}} \mathbb{P}(\,\mathsf{C}_i \mid \mathsf{S}\,)\mathbb{P}(\mathsf{S})}{\prod_{i\in\mathsf{I}} \mathbb{P}(\,\mathsf{C}_i \mid \mathsf{S}\,)\mathbb{P}(\mathsf{S}) + \prod_{i\in\mathsf{I}} \mathbb{P}(\,\mathsf{C}_i \mid \overline{\mathsf{S}}\,)\mathbb{P}(\overline{\mathsf{S}})}$$

e, usando que $\mathbb{P}(\,\mathsf{C}_i \mid \mathsf{S}\,) = \mathbb{P}(\,\mathsf{S} \mid \mathsf{C}_i\,)\mathbb{P}(\mathsf{C}_i)/\mathbb{P}(\mathsf{S})$ e a igualdade análoga para $\overline{\mathsf{S}}$

$$\frac{\prod_{i\in\mathsf{I}} \mathbb{P}(\,\mathsf{C}_i \mid \mathsf{S}\,)\mathbb{P}(\mathsf{S})}{\prod_{i\in\mathsf{I}} \mathbb{P}(\,\mathsf{C}_i \mid \mathsf{S}\,)\mathbb{P}(\mathsf{S}) + \prod_{i\in\mathsf{I}} \mathbb{P}(\,\mathsf{C}_i \mid \overline{\mathsf{S}}\,)\mathbb{P}(\overline{\mathsf{S}})} = \frac{\prod_{i\in\mathsf{I}} \mathbb{P}(\,\mathsf{S} \mid \mathsf{C}_i\,)}{\prod_{i\in\mathsf{I}} \mathbb{P}(\,\mathsf{S} \mid \mathsf{C}_i\,) + \prod_{i\in\mathsf{I}} \mathbb{P}(\,\overline{\mathsf{S}} \mid \mathsf{C}_i\,)}$$

donde seque a equação (1.21). ◊

As hipóteses assumidas nas equações (1.22) e (1.23) significam, grosso modo, que o conhecimento de algumas das características não dá nenhuma pista sobre a presença ou não das outras características; estamos assumindo independência dos eventos e o significado preciso disso é o assunto da próxima seção.

1.4 INDEPENDÊNCIA DE EVENTOS

Se um dado é lançado duas vezes, então temos

$$\mathbb{P}[\text{ a soma é } 7 \mid \text{o primeiro resultado é } 4\ ] = \frac{1}{6} = \mathbb{P}[\text{a soma é } 7\ ]$$

entretanto

$$\mathbb{P}[\text{ a soma é } 12 \mid \text{o primeiro resultado é } 4\ ] = 0 \neq \mathbb{P}[\text{a soma é } 12\ ].$$

O condicionamento de ocorrência de um evento A à ocorrência de B pode afetar ou não a probabilidade de ocorrência de A e quando $\mathbb{P}(\,A \mid B\,) = \mathbb{P}(A)$. Definimos que o evento A é independente do evento B se

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$$

Notemos que se A é independente de B então B é independente de A de modo que dizemos A e B são **eventos independentes**. .

Se os eventos têm probabilidade não nula então decorre da definição acima que a independência de A e B equivale a $\mathbb{P}(\,A \mid B\,) = \mathbb{P}(A)$ e $\mathbb{P}(\,B \mid A\,) = \mathbb{P}(B)$. Como vimos, se um dado é lançado duas vezes então a probabilidade do evento "soma do dois lançamentos resultar 7" é 1/6, por sua vez, o evento "a soma ser 7" e "o primeiro lançamento resultar 4" tem probabilidade $1/36 = 1/6 \cdot 1/6$, logo são eventos independentes. Por outro lado, "a soma ser 5" e "o primeiro lançamento resultar 4" tem probabilidade 1/36, mas "soma ser 5" tem probabilidade $4/36 \neq 1/6 \cdot 1/6$, logo não são eventos independentes.

É imediato da definição o seguinte fato.

Proposição 1.31 *Todo evento* A *de um modelo probabilístico é independente do evento certo* Ω *e do evento impossível* $\varnothing$. □

A independência dos eventos A e B resulta na independência entre seus complementos, como enunciado a seguir.

Proposição 1.32 *Se* A *e* B *são eventos independentes de um modelo probabilístico então* A *e* $\overline{B}$ *são eventos independentes,* $\overline{A}$ *e* B *são eventos independentes e* $\overline{A}$ *e* $\overline{B}$ *são eventos independentes.*

Demonstração. Deduzimos de $\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap \overline{B})$, usando a independência de A e B, que

$$\mathbb{P}(A \cap \overline{B}) = \mathbb{P}(A) - \mathbb{P}(A \cap B) = \mathbb{P}(A) - \mathbb{P}(A)\mathbb{P}(B) = \mathbb{P}(A)(1 - \mathbb{P}(B)) = \mathbb{P}(A)\mathbb{P}(\overline{B})$$

portanto são eventos independentes. Os outros casos são demonstrados de modo análogo. □

Para investigar o caso de três eventos, voltemos ao experimento de um dado lançado duas vezes. Sejam A o evento "a soma do dois lançamentos é 7", B o evento "o primeiro lançamento resulta 4" e C o evento "o segundo lançamento resulta 2". Como vimos, A e B são eventos independentes. Por razão análoga A e C são independentes. Porém $\mathbb{P}(\,A \mid B \cup C\,) = 2/11 \neq \mathbb{P}(A)$ e $\mathbb{P}(\,A \mid B \cap C\,) = 0 \neq \mathbb{P}(A)$, ou seja, A não é independente de [B *ou* C] e não é independente de [B *e* C].

Para três eventos, digamos A, B e C, queremos que A seja independente do par de eventos $\{B, C\}$ quando o conhecimento de qualquer informação a respeito da ocorrência de B, de C, ou de uma combinação deles pelas operações elementares de conjuntos não altere a probabilidade de ocorrer A.

EXERCÍCIO 1.33. Assuma, como definição de "A *é independente de* $\{B, C\}$" se vale a equação equação (1.24) a seguir

$$\mathbb{P}(\,A \mid B \cap C\,) = \mathbb{P}(\,A \mid B\,) = \mathbb{P}(\,A \mid C\,) = \mathbb{P}(A). \tag{1.24}$$

Prove que se A é independente de $\{B, C\}$ então A é independente de cada um dos eventos da família

$$\left\{\varnothing, B, C, \overline{B}, \overline{C}, B \cup C, B \cap C, B \cup \overline{C}, B \cap \overline{C}, \overline{B} \cup C, \overline{B} \cap C, \overline{B} \cup \overline{C}, \overline{B} \cap \overline{C}, \Omega\right\}$$

que chamamos de **espaço de eventos gerado** por $\{B, C\}$.

Em vista disso, definimos que A **é independente de** $\{B, C\}$ se for independente de todo evento do espaço de eventos gerado por $\{B, C\}$. Tal definição é equivalente a (veja a equação (1.24)): A *é independente de* $\{B, C\}$ *se, e somente se, é independente de* B, *é independente de* C *e é independente de* $B \cap C$, isto é,

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B),\ \mathbb{P}(A \cap C) = \mathbb{P}(A)\mathbb{P}(C) \text{ e } \mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B \cap C).$$

Ademais, notemos que essa definição é compatível com a definição de "A independente de B" dada anteriormente pois, pelas proposições 1.31 e 1.32, o evento A é independente de todo evento do espaço de eventos gerado por $\{B\}$, o qual é $\{\varnothing, B, \overline{B}, \Omega\}$.

Em geral, estamos interessados no caso em que cada evento é independente dos outros dois eventos restantes e, nesse caso, dizemos que os eventos A, B e C são **mutuamente independentes** o que é equivalente a

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B),\ \mathbb{P}(A \cap C) = \mathbb{P}(A)\mathbb{P}(C),\ \mathbb{P}(B \cap C) = \mathbb{P}(B)\mathbb{P}(C) \text{ e } \mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C).$$

Consideremos três lançamentos de uma moeda equilibrada e os eventos E_{12} dado por "o resultado do primeiro e do segundo lançamentos coincidem", E_{13} dado por "o resultado do primeiro e do terceiro coincidem" e E_{23} dado por "o resultado do segundo e do terceiro coincidem". Cada um desses eventos tem probabilidade 1/2. Os eventos são independentes quando tomados dois-a-dois: $\mathbb{P}(E_{12} \cap E_{13}) = \mathbb{P}(\{(Ca, Ca, Ca), (Co, Co, Co)\}) = 1/4$ e, analogamente, os eventos $E_{12} \cap E_{23}$ e $E_{13} \cap E_{23}$ têm probabilidade 1/4. Entretanto esses eventos não são mutuamente independentes pois $\mathbb{P}(E_{12} \cap E_{13} \cap E_{23}) = 1/4$ enquanto que $\mathbb{P}(E_{12})\mathbb{P}(E_{13})\mathbb{P}(E_{23}) = 1/8$.

Agora, num lançamento de dados tomamos os eventos $A = \{1, 2, 3, 4\}$ e $B = C = \{4, 5, 6\}$. Os eventos B e C não são independentes. Também A e B não são independentes pois $\mathbb{P}(A \cap B) = 1/6$ enquanto que $\mathbb{P}(A)\mathbb{P}(B) = 1/3$, logo A e C não são eventos independentes. Porém $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$.

Uma coleção enumerável de eventos $\mathcal{E} = \{E_n\}$ é dita **mutuamente independente** se para todo subconjunto finito $J \subset \mathbb{N}$ vale que

$$\mathbb{P}\left(\bigcap_{\ell \in J} E_\ell\right) = \prod_{\ell \in J} \mathbb{P}(E_\ell).$$

Para $k \in \{2, \ldots, n\}$ fixo, dizemos que a coleção $\mathcal{E}$ é k-**a**-k **independente** se todo subconjunto de índices $J \subset \mathbb{N}$ com $|J| \leqslant k$ define uma subcoleção de eventos mutuamente independentes.

Independência condicional Dizemos que A_1 e A_2 são **condicionalmente independentes** dado B se

$$\mathbb{P}(A_2 \cap A_1 \mid B) = \mathbb{P}(A_1 \mid B)\mathbb{P}(A_2 \mid B).$$

Essa definição estende-se naturalmente, como acima, para um coleção com mais que dois eventos.

As equações (1.22) e (1.23) no exemplo para filtros anti-*spam* pede que as características $C_1, \ldots, C_n$, que são usadas para classificar as mensagens, sejam independentes quando condicionamos à ocorrência de *spam* e quando condicionamos à ocorrência de não *spam*, respectivamente.

No contexto do exemplo 1.27, página 23, qual a probabilidade de um motorista se envolver num acidente no segundo ano dado que tenha se envolvido em acidente no primeiro ano de contrato? Denotemos por A_2 o evento "acidente no 2º ano de contrato". Assumiremos que A_1 e A_2 são condicionalmente independentes dado A, ou seja, $\mathbb{P}(A_2 \cap A_1 \mid A) = \mathbb{P}(A_1 \mid A)\mathbb{P}(A_2 \mid A)$. Se definirmos a medida de probabilidade $\mathbb{Q}(X) := \mathbb{P}(X \mid A_1)$, queremos determinar $\mathbb{Q}(A_2)$. Pelo teorema de probabilidade total $\mathbb{Q}(A_2) = \mathbb{Q}(A_2 \mid A)\mathbb{Q}(A) + \mathbb{Q}(A_2 \mid \overline{A})\mathbb{Q}(\overline{A})$. Mas

$$\mathbb{Q}(A_2 \mid A) = \frac{\mathbb{Q}(A_2 \cap A)}{\mathbb{Q}(A)} = \frac{\mathbb{P}(A_2 \cap A \mid A_1)}{\mathbb{P}(A \mid A_1)} = \mathbb{P}(A_2 \mid A \cap A_1) = \mathbb{P}(A_2 \mid A)$$

em que a última igualdade segue da independência condicional assumida (verifique). Lembremos que os motoristas propensos a acidentes se envolvem em acidente no período de um ano com probabilidade 0,4, logo $\mathbb{P}(A_2 \mid A) = 0{,}4$. Ainda, calculamos no exemplo 1.27 que $\mathbb{Q}(A) = 6/13$. Desse modo temos que

$$\mathbb{Q}(A_2) = \mathbb{Q}(A_2 \mid A)\mathbb{Q}(A) + \mathbb{Q}(A_2 \mid \overline{A})\mathbb{Q}(\overline{A}) = 0{,}4 \cdot \frac{6}{13} + 0{,}2 \cdot \frac{7}{13} \approx 0{,}29.$$

Exemplo 1.34. Suponhamos que numa caixa há duas moedas, uma delas com duas caras e a outra é uma moeda comum. Uma moeda é sorteada e lançada duas vezes. Sejam A e B os eventos "o primeiro lançamento é cara" e "o segundo lançamento é cara", respectivamente. Condicionados ao evento C definido por "a moeda normal foi a escolhida" os eventos A e B são independentes:

$$\mathbb{P}(A \cap B \mid C) = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = \mathbb{P}(A \mid C) \cdot \mathbb{P}(B \mid C).$$

Entretanto os eventos A e B não são independentes pois

$$\mathbb{P}(A) = \mathbb{P}(A \mid C)\mathbb{P}(C) + \mathbb{P}(A \mid \overline{C})\mathbb{P}(\overline{C}) = \frac{1}{2} \cdot \frac{1}{2} + 1\frac{1}{2} = \frac{3}{4}$$
$$\mathbb{P}(B) = \mathbb{P}(B \mid C)\mathbb{P}(C) + \mathbb{P}(B \mid \overline{C})\mathbb{P}(\overline{C}) = \frac{1}{2} \cdot \frac{1}{2} + 1\frac{1}{2} = \frac{3}{4}$$

porém

$$\begin{aligned}\mathbb{P}(A \cap B) &= \mathbb{P}(A \cap B \mid C)\mathbb{P}(C) + \mathbb{P}(A \cap B \mid \overline{C})\mathbb{P}(\overline{C}) \\ &= \mathbb{P}(A \mid C)\mathbb{P}(B \mid C)\mathbb{P}(C) + \mathbb{P}(A \mid \overline{C})\mathbb{P}(B \mid \overline{C})\mathbb{P}(\overline{C}) = \frac{5}{8}\end{aligned}$$

por causa da independência condicional, usada para deduzir a segunda linha da equação acima. ◊

Agora, retomemos o exemplo do vírus da *probabilite*, dado na página 26, que é um vírus que contamina 0,5% da população universitária e que o teste para detectar o vírus tem 1% de chance de acusar falsos positivos e falsos negativos. Vimos que $\mathbb{P}(B \mid A) \approx 0{,}34$ onde A é o evento "foi detectado *probabilite*" e B o evento "tem *probabilite*". Agora, suponha que o estudante estava com dor de cabeça (o teste foi feito em véspera de prova). É sabido que 95% dos indivíduos que tem *probabilite* apresentam dor de cabeça, enquanto que 10% da população não contaminada apresenta dor de cabeça; também é sabido que o evento "ter dor de cabeça", que denominamos C, não afeta a precisão do teste no sentido de que A e C são condicionalmente independentes $\mathbb{P}(A \cap C \mid B) = \mathbb{P}(A \mid B)\mathbb{P}(C \mid B)$. Com esse fato, a probabilidade de ter *probabilite* dado que o teste deu positivo e o estudante tem dor de cabeça é

$$\begin{aligned}\mathbb{P}(B \mid A \cap C) &= \frac{\mathbb{P}(A \cap C \mid B)\mathbb{P}(B)}{\mathbb{P}(A \cap C \mid B)\mathbb{P}(B) + \mathbb{P}(A \cap C \mid \overline{B})\mathbb{P}(\overline{B})} \\ &= \frac{\mathbb{P}(A \mid B)\mathbb{P}(C \mid B)\mathbb{P}(B)}{\mathbb{P}(A \mid B)\mathbb{P}(C \mid B)\mathbb{P}(B) + \mathbb{P}(A \mid \overline{B})\mathbb{P}(C \mid \overline{B})\mathbb{P}(\overline{B})} \\ &= \frac{0{,}99 \cdot 0{,}95 \cdot 0{,}005}{0{,}99 \cdot 0{,}95 \cdot 0{,}005 + 0{,}01 \cdot 0{,}1 \cdot 0{,}995} \approx 0{,}82.\end{aligned}$$

1.4.1 ESPAÇO PRODUTO

Retomando o exemplo no final da seção 1.2.2, executamos o algoritmo 1 em dois computadores e os resultados foram considerados independentes, o probabilidade de uma execução errar não altera a probabilidade do outro computador errar de modo que a probabilidade de ambos errarem é o produto das probabilidade de cada um errar. Nessa seção vamos formalizar essa ideia.

Dados os espaços de probabilidade discretos $(\Omega_i, \mathbb{P}_i)$, para $1 \leqslant i \leqslant n$, podemos definir um espaço de probabilidade discreto cujo espaço amostral é dado pelas sequências $(\omega_1, \omega_2, \ldots, \omega_n)$ do produto cartesiano $\Omega_1 \times \Omega_2 \times \cdots \times \Omega_n$ e a medida de probabilidade em cada ponto amostral é

$$\mathbb{P}(\omega) := \mathbb{P}_1(\omega_1)\mathbb{P}_2(\omega_2)\cdots\mathbb{P}_n(\omega_n)$$

para todo $\omega = (\omega_1, \omega_2, \ldots, \omega_n) \in \Omega_1 \times \Omega_2 \times \cdots \times \Omega_n$, a qual se estende do modo usula para todo $A \subset \Omega$. Esse espaço de probabilidade é chamado **espaço produto** e é o modelo probabilístico de um experimento sendo repetido n vezes sob condições idênticas. Não é difícil verificar que

$$\sum_{\omega\in\Omega} \mathbb{P}(\omega) = \sum_{\omega_1\in\Omega_1} \sum_{\omega_2\in\Omega_2} \cdots \sum_{\omega_n\in\Omega_n} \mathbb{P}_1(\omega_1)\mathbb{P}_2(\omega_2)\cdots\mathbb{P}_n(\omega_n) = 1$$

o que garante que o espaço produto é um espaço de probabilidade. Além disso, para $A_i \subset \Omega_i$, $1 \leqslant i \leqslant n$, temos (verifique) $\mathbb{P}(A_1 \times A_2 \times \cdots \times A_n) = \mathbb{P}_1(A_1)\mathbb{P}_2(A_2)\cdots\mathbb{P}_n(A_n)$.

Um modelo probabilístico para n lançamentos de uma moeda equilibrada em que os resultados dos lançamentos são mutuamente independentes é dado pelo espaço produto $(\Omega^n, \mathbb{P}^n)$, em que $(\Omega, \mathbb{P})$ é o modelo para um lançamento dado no exemplo 1.6.

Exemplo 1.35. Sortear um número inteiro entre 0 e 999 é um experimento aleatório cujo espaço amostral é o conjunto dos números naturais até 999 e cuja medida de probabilidade é a uniforme, isto, é cada número ocorre com probabilidade 1/1.000. Se temos disponível os algarismos $0, 1, \ldots, 9$ podemos gerar uniformemente um número entre 0 e 999 se sortearmos $d_1, d_2, d_3 \in \{0, 1, \ldots, 9\}$, com os resultados mutuamente independentes, e tomarmos $n := d_1 \times 10^2 + d_2 \times 10^1 + d_3 \times 10^0$, então o espaço amostral é dado pelas ternas de algarismos (d_1, d_2, d_3) e a probabilidade de n é $(1/10)^3 = 1/1.000$. Esse modelo probabilístico é dado pelo espaço produto. Há uma correspondência bijetiva (dada por $n = n(d_1, d_2, d_3)$) entre esses dois espaços amostrais que preserva a probabilidade dos pontos amostrais, com isso os eventos aleatórios têm a mesma probabilidade no dois modelos e, nesse sentido, sortear uniformemente um número de três algarismos e sortear uniformemente cada um de três algarismos são experimentos aleatórios equivalentes. ◊

EXERCÍCIO 1.36. Considere n repetições de um experimento modelado por $(\Omega, \mathbb{P})$. Sejam $A_1, A_2, \ldots, A_n$ eventos de Ω^n tais que a j-ésima rodada sozinha determina se A_j ocorre, ou seja, existe um $E_j \subset \Omega$ tal que $A_j = \Omega^{j-1} \times B_j \times \Omega^{n-j}$. Se em Ω^n tomarmos a medida produto, então os eventos $A_1, A_2, \ldots, A_n$ são mutuamente independentes. (Dica: comece com a prova de que os eventos são dois a dois independentes.)

Repetições independentes de um algoritmo Uma ideia central na utilidade dos algoritmos probabilísticos é a possibilidade de reduzir o erro das respostas executando o algoritmo com a mesma entrada várias vezes: se um algoritmo erra com probabilidade ε, então em duas execuções errará com probabilidade ε^2, em $r \in \mathbb{N}$ execuções a probabilidade de erro é ε^r.

Nos algoritmos probabilísticos assumimos independência dos resultados nos sorteios. Primeiro, assumimos que todos os sorteios feitos durante uma rodada do algoritmo são independentes, isto é, o resultado de um ou mais sorteios não altera a probabilidade do resultado de um outro sorteio. Também, assumimos que os sorteios feitos durante uma execução não altera a probabilidade dos sorteios nas outras execuções de modo que se justifica e decaimento exponencial no erro descrito no parágrafo anterior.

Lembremos que o algoritmo 1 erra quando declara um polinômio não nulo como nulo, o que pode ter ocorrido pela escolha de uma raiz do polinômio pelo algoritmo. Fixada uma instância do problema, suponhamos r rodadas independentes desse algoritmo (com a mesma instância). Se em alguma dessas rodadas o o algoritmo 1 responde *não*, então essa é a resposta definitiva para o problema com essa instância. Se todas as r respostas forem *sim* então a resposta

definitiva é *sim* e essa resposta estará errada se todas as r respostas de cada rodada estiverem erradas, o que ocorre com probabilidade 4^{-r}, pela independência dos eventos "resposta errada na i-ésima execução". Assim, se precisamos de uma garantia na resposta para o problema, por exemplo com probabilidade de erro menor que ε, para algum $\varepsilon > 0$ fixo, então basta escolher r de modo que $4^{-r} < \varepsilon$, ou seja, $r > \log_2 \sqrt{1/\varepsilon}$ rodadas.

Proposição 1.37 *Dado um real positivo* ε*, o problema* teste de identidade de polinômios em uma variável *pode ser resolvido por um algoritmo aleatorizado com probabilidade de erro menor que* ε. □

1.4.2 GERADOR DE NÚMEROS ALEATÓRIOS

Suponhamos que temos disponível uma fonte que gera bits aleatórios de modo uniforme e independente e queremos projetar um algoritmo que recebe um inteiro positivo M e nos devolve uma escolha aleatória em $\{0, 1, \ldots, M-1\}$. Se M é uma potência de 2, digamos que $M = 2^k$, então a resposta é simples: basta sortearmos k bits aleatórios $d_0, d_1, \ldots d_{k-1} \in \{0, 1\}$ que o resultado é o número $\sum_{i=0}^{k-1} d_i 2^i$ no domínio desejado com probabilidade $1/M$. No caso em que M não é potência de 2, digamos que $2^{k-1} < M < 2^k$ (o que significa que precisamos de k bits aleatório) usamos o mesmo processo descrito no parágrafo anterior com a exceção de que se o resultado for maior ou igual a M, o processo é reiniciado e é repetido até que um número entre 0 e $M-1$ seja obtido. O algoritmo descrito concisamente fica como abaixo, no algoritmo 2. O resultado das escolhas aleatórias na linha 2 do algoritmo 2, a sequência $d_{k-1} d_{k-2} \ldots d_0$, é um evento elementar do espaço produto $(\{0,1\}^k, \mathbb{P}^k)$, em que $\mathbb{P}(0) = \mathbb{P}(1) = 1/2$, que tem probabilidade $\mathbb{P}^k(d_{k-1} d_{k-2} \ldots d_0) = (1/2)^k$. Essa sequência é a representação binária do número $\sum_{i=0}^{k-1} d_i 2^i$ que pertence a $\{0, 1, \ldots, 2^k - 1\}$.

Instância : inteiro positivo $M \geqslant 2$.
Resposta : uma escolha aleatória uniforme em $\{0, 1, \ldots, M-1\}$.

1 **repita**
2 **para cada** $i \in \{0, \ldots, \lfloor \log_2 M \rfloor\}$ **faça** $d_i \leftarrow_R \{0, 1\}$;
3 $N \leftarrow \sum_i d_i 2^i$;
4 **até que** $N < M$;
5 **responda** N.

Algoritmo 2: gerador de números aleatórios.

Por exemplo, se $M = 7$ então $k = 3$. Com três bits $d_2 d_1 d_0$ temos as representações binárias dos naturais de 0 a 7. O laço da linha 1 gera qualquer um desses números com a mesma probabilidade, a saber $1/2^3 = 1/8$. Porém o algoritmo só termina se o sorteio for diferente de 7, isto é, não ocorre o evento $d_2 d_1 d_0 = 111$. Dado que esse evento não ocorre, qual a probabilidade do algoritmo responder 4? Usando probabilidade condicional $\mathbb{P}[\, N = 4 \mid N \neq 7 \,] = (1/8)/(7/8) = 1/7$ e, de fato, o algoritmo escolhe $N \in \{0, \ldots, 6\}$ com probabilidade 1/7.

No caso geral, definimos o evento $A = \{0, 1, \ldots, M-1\}$ e para qualquer $t \in A$ a probabilidade do algoritmo responder t é dada por

$$\mathop{\mathbb{P}}_{N \in \{0, \ldots, 2^k - 1\}} [\, N = t \mid N \in A \,] = \frac{\mathbb{P}(\{t\} \cap A)}{\mathbb{P}(A)} = \frac{(1/2)^k}{M/2^k} = \frac{1}{M}.$$

Portanto, se o algoritmo termina, ou seja, dado que o sorteio N satisfaz $N < M$, então ele responde com um número entre 0 e $M-1$ de modo uniforme. Resta provarmos que o algoritmo termina, isto é, eventualmente a condição $N < M$ na linha 4 é satisfeita. Nesse caso assumimos que os sorteios e os eventos que eles definem em rodadas diferentes do laço da linha 1 são independentes.

Fixamos uma instância M com $2^{k-1} < M < 2^k$ e $k = \lfloor \log_2 M \rfloor + 1$ é o número de bits sorteados. A probabilidade com que uma rodada do laço da linha 1 resulte em um inteiro N que pertença ao conjunto $\overline{A} = \{M, \ldots, 2^k - 1\}$ é

$$\mathbb{P}(\overline{A}) = \frac{2^k - M}{2^k} = 1 - \frac{M}{2^k}.$$

Definimos para todo $n \geqslant 1$ o evento A_n por "o algoritmo leva mais que n rodadas para terminar". Tal evento ocorre se nas n primeiras tentativas do laço na linha 1 ocorre um sorteio em $\overline{A}$ e daí pra diante pode ocorrer qualquer um dos dois casos, A ou $\overline{A}$, logo $\mathbb{P}(A_n) = (1 - M/2^k)^n$ pela independência da ocorrência dos eventos $\overline{A}$ em cada rodada do laço.

Os eventos A_n formam uma sequência decrescente $A_n \supset A_{n+1}$ e $\lim_{n\to\infty} A_n = \bigcap_{n\geqslant 1} A_n$ é o evento "o algoritmo não termina", cuja probabilidade é, por continuidade (equação (1.6) na página 16),

$$\mathbb{P}[\text{o algoritmo não termina}] = \mathbb{P}\Big(\lim_{n\to\infty} A_n\Big) = \lim_{n\to\infty} \mathbb{P}(A_n) = \lim_{n\to\infty} \left(1 - \frac{M}{2^k}\right)^n = 0$$

pois $M < 2^k$, portanto, o algoritmo termina com probabilidade 1.

Notemos que, diferente do exemplo do algoritmo 1, nesse caso a resposta está sempre correta e a aleatoriedade influencia na duração das rodadas, isto é, no tempo que leva para o algoritmo terminar. Esse algoritmo não só termina como, de fato, termina rápido, em poucas rodadas do laço da linha 1 com alta probabilidade. De $M > 2^{k-1}$ temos $\mathbb{P}(\overline{A}) < 1/2$, portanto, em n rodadas do laço todos os inteiros sorteados pertencem a $\overline{A}$ com probabilidade menor que 2^{-n}. A probabilidade de não terminar em, por exemplo, 4 rodadas é menor que 0,07, em 10 rodadas é menor que 0,00098.

1.5 EXERCÍCIOS

EXERCÍCIO 1.38. Sejam A, B e C eventos aleatórios. Determine expressões que envolvem somente conjuntos e operações sobre conjuntos para

1. somente A ocorre;
2. A e B mas não C ocorrem;
3. os três eventos ocorrem;
4. pelo menos um evento ocorre;
5. pelo menos dois eventos ocorrem;
6. exatamente um evento ocorre;
7. exatamente dois eventos ocorrem;
8. nenhum evento ocorre;
9. não mais que dois eventos ocorrem.

EXERCÍCIO 1.39. Considere o lançamento repetido de uma moeda equilibrada até sair coroa, como descrito no exemplo 1.13. Com que probabilidade o número de lançamentos é par?

EXERCÍCIO 1.40 (Princípio da inclusão–exclusão). Prove que para eventos $A_1, A_2, \ldots, A_n$ vale

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathbb{P}(A_i) - \sum_{i=1}^{n}\sum_{j=i+1}^{n} \mathbb{P}\big(A_i \cap A_j\big) + \sum_{i=1}^{n}\sum_{j=i+1}^{n}\sum_{k=j+1}^{n} \mathbb{P}\big(A_i \cap A_j \cap A_k\big) + \cdots + (-1)^n \mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right).$$

EXERCÍCIO 1.41. Prove que corolário 1.5, página 10, admite a seguinte extensão: para qualquer conjunto enumerável $\{E_i : i \geqslant 1\}$ de eventos num espaço discreto vale

$$\mathbb{P}\left(\bigcup_{i\geqslant 1} E_i\right) \leqslant \sum_{i\geqslant 1} \mathbb{P}(E_i).$$

EXERCÍCIO 1.42. Seja p primo e tome $\Omega := \{1, 2, \ldots, p\}$ com medida $\mathbb{P}(A) = |A|/p$, para todo $A \subset \Omega$. Prove que se A e B são independentes então pelo menos um desses eventos deve ser $\varnothing$ ou Ω.

	$E_0(m)$	$E_1(m)$
α	a	b
β	b	a

Tabela 1.3: função de codificação.

EXERCÍCIO 1.43. Defina um sistema de codificação com $\mathcal{P} = \{\alpha, \beta\}$, $\mathcal{C} = \{a, b\}$ e $\mathcal{K} = \{0, 1\}$ tais que $\mathbb{P}_{\mathcal{K}}(0) = 1/10$ e $\mathbb{P}_{\mathcal{K}}(1) = 9/10$. A codificação $E_k(m)$, para cada $m \in \{\alpha, \beta\}$ é dada na tabela 1.3 abaixo. Prove que o sistema não tem sigilo perfeito.

EXERCÍCIO 1.44 (Teorema de Shannon). Prove o seguinte resultado: dados um sistema de codificação com $\mathcal{P}, \mathcal{K}, \mathcal{C}$ finitos e $|K| = |C|$ e dada uma medida de probabilidade $\mathbb{P}_{\mathcal{P}}$ sobre $\mathcal{P}$ tal que $\mathbb{P}_{\mathcal{P}}(P) > 0$, para todo $P \in \mathcal{P}$, esse sistema tem sigilo perfeito se e somente se as chaves são equiprováveis e se para todo $P \in \mathcal{P}$ e todo $C \in \mathcal{C}$ existe um único $K \in \mathcal{K}$ tal que $E_K(P) = C$.

EXERCÍCIO 1.45. Considere $\Omega = \{0, 1\}^n$ e suponha que x é o resultado de um sorteio uniforme em Ω e y é o resultado de um sorteio em Ω em que os resultados ocorrem de acordo com alguma medida de probabilidade, possivelmente diferente da uniforme. Mostre que $y \oplus x$ (ou-exclusivo coordenada-a-coordenada) é qualquer elemento de Ω com probabilidade $(1/2)^n$.

EXERCÍCIO 1.46. Vamos provar uma generalização do exercício 1.45. Seja $(G, \circ)$ um grupo abeliano finito, T um conjunto finito e $f\colon T \to G$ uma função qualquer. Suponha que x é o resultado de um sorteio uniforme em G e y é o resultado de um sorteio em T em que os resultados ocorrem de acordo com alguma medida de probabilidade, possivelmente diferente da uniforme. Prove que $x \circ f(y)$ é qualquer elemento do grupo G com probabilidade uniforme. Prove também que para quaisquer $i \in T, j \in G$ os eventos definidos por "$y = i$" e "$x \circ f(y)$=j" são independentes.

EXERCÍCIO 1.47. No exercício anterior, suponha que f seja invertível e T um conjunto de textos legíveis. A codificação de um texto legível $m \in T$ é feita transformando-o num elemento do grupo com $f(t)$, sorteando uma chave $k \in G$ uniformemente e calculando $c = k \circ f(t)$. A decodificação é feita conhecendo-se a chave k e calculando $f^{-1}(c \circ (-k))$, em que $-k$ é o elemento inverso de k em G. Verifique que tal sistema de codificação tem sigilo perfeito.

EXERCÍCIO 1.48. Suponha que você tem três moedas e uma delas é viciada de modo que $\mathbb{P}(\text{cara}) = 2/3$. Escolhendo uma delas ao acaso a probabilidade de acertar qual é a viciada é um terço. Agora, suponha que o resultado do lançamento de cada uma delas, sem conhecer qual é a viciada, resulta em (cara, cara, coroa). Mostre, usando o Teorema de Bayes, que a probabilidade da primeira moeda ser a viciada é 2/5.

EXERCÍCIO 1.49 (Saldanha, 1997). Dois amigos querem decidir quem pagará a conta da pizzaria com uma aposta. Cada um deles escolhe uma sequência de três resultados do lançamento de uma moeda honesta, em seguida eles jogam uma moeda até que saia uma das duas sequências: aquele que tiver escolhido a primeira sequência a sair ganhou a aposta. Por exemplo, André é o primeiro e fica com a sequência (coroa, coroa, cara) enquanto Renato responde com (cara, coroa, coroa). Eles jogam a moeda obtendo coroa, cara, coroa, cara, coroa, cara, coroa, coroa e neste momento Renato é o vencedor. Mostre que nesse caso a probabilidade do Renato ganhar o jogo é 3/4. Prove que o segundo jogador sempre tem uma escolha mais vantajosa pra ele.

EXERCÍCIO 1.50. Três convidados chegaram numa festa vestindo chapéu e os entregaram na recepção. O funcionário, pouco cuidadoso, não identificou os chapéus e no final da festa os entregou aleatoriamente para as mesmas três pessoas.

Use o princípio da inclusão–exclusão para mostrar que ninguém recebe o próprio chapéu com probabilidade 1/3. Generalize o resultado para n convidados e use a série de potências para a função exponencial (veja (s.8)) para mostrar que a probabilidade de ninguém pegar o próprio chapéu converge, quando $n \to \infty$, para 1/e.

EXERCÍCIO 1.51. Num treino um grupo de n paraquedistas estão enfileirados e um paraquedista é escolhido ao acaso no seu grupo. O paraquedista escolhido cumprimenta todos os paraquedistas do seu grupo e pula; o grupo é dividido em dois: um grupo formado pelos paraquedistas à esquerda do que pulou e o outro grupo formado pelos paraquedistas à direita do que pulou. O procedimento é repetido nos grupos restantes até sobrarem grupos de um único paraquedista, que pulam um a um. Note que paraquedistas que em algum momento ficam em grupos diferentes não se cumprimentaram e não se cumprimentarão desse momento em diante. A ordem da fila dentro de cada grupo é sempre mantida. Prove que os paraquedistas das posições i e j, sem perda de generalidade $j > i$, se cumprimentam com probabilidade $2/(j-i+1)$.

EXERCÍCIO 1.52. Considere uma moeda que resulta em cara com probabilidade $p \in (0,1)$. Prove que a probabilidade de que em n lançamentos (independentes) temos mais que k caras é no máximo $\binom{n}{k}p^k$.

EXERCÍCIO 1.53. Considere uma moeda que resulta em cara com probabilidade 1/5 e coroa com probabilidade 4/5. Justifique que a probabilidade de sair menos que k coroas em $2k$ lançamentos (independentes) é

$$\sum_{0}^{k-1}\binom{2k}{i}(4/5)^i(1/5)^{2k-i} < (1/5)^{2k}4^k\sum_{0}^{k-1}\binom{2k}{i}.$$

Use o teorema do binômio de Newton e prove que a probabilidade de sair menos que k coroas em $2k$ lançamentos dessa moeda é menor que $(4/5)^{2k}$.

EXERCÍCIO 1.54. Considere o seguinte procedimento para gerar uma permutação do vetor $a = (1,2,\ldots,n)$, para qualquer $n \in \mathbb{N}$:

- para cada coordenada $i = 1,2,\ldots,n$ do vetor a
  - sorteie uniformemente uma coordenada $j \in \{1,2,\ldots,n\}$
  - troque os componentes das coordenadas: coloque o número da coordenada j na coordenada i e o da coordenada i na coordenada j.

O vetor resultante é uma permutação do vetor inicial com probabilidade uniforme?

EXERCÍCIO 1.55. Considere um algoritmo que recebe um inteiro positivo $M > 0$ e devolve um inteiro escolhido aleatoriamente em $\{0,1,\ldots,M-1\}$ da seguinte forma: (1) gere um número N de k bits como no algoritmo acima; (2) devolva o resto da divisão inteira de N por M. Isso evitaria a execução de mais que uma iteração do laço do algoritmo acima. Mostre que isso não resulta em uma resposta em $\{0,\ldots,M-1\}$ com probabilidade uniforme.

EXERCÍCIO 1.56. O seguinte gerador de números aleatórios é adaptado do exercício anterior.

Instância : inteiros positivos M e t.
Resposta : uma escolha aleatória uniforme em $\{0,1,\ldots,M-1\}$.

1 seja k o número de bits de M;
2 $(d_0, d_1, \ldots, d_{k+t-1}) \leftarrow_R \{0,1\}^{k+t}$;
3 $N \leftarrow \sum_i d_i 2^i$;
4 **responda** N mod M.

No caso $t = 0$ a probabilidade da resposta não é uniforme (exercício 1.55 acima). Prove que quanto maior é t mais próximo a probabilidade de N está da uniforme, no seguinte sentido

$$\sum_{n=0}^{M-1} \left| \mathbb{P}[N = n] - \frac{1}{M} \right| \leqslant \frac{1}{2^{t-1}}.$$

EXERCÍCIO 1.57. Distribuímos uniformemente e independentemente m bolas idênticas em n caixas distintas. Qual é a probabilidade com que a i-ésima caixa fica vazia? Qual é a probabilidade com que a j-ésima e a i-ésima caixas ficam vazias? Qual é a probabilidade com que nenhuma fica vazia? E de exatamente uma ficar vazia?

Prove que no caso $n = m$ o maior número de bolas em qualquer caixa é no máximo $2e \log_2 n$ com probabilidade $1 - n^{-4}$ (dica: estime a probabilidade de uma caixa ter muitas bolas, a fórmula de Stirling (d.3) pode ser útil nos cálculos, e use o corolário 1.5).

EXERCÍCIO 1.58. Considere o conjunto dos números naturais e defina para todo subconjunto E e cada $n \geqslant 1$ a densidade relativa de E

$$P_n(E) := \frac{|\{1, 2, \ldots, n\} \cap E|}{n}.$$

Defina $p(E) := \lim_{n\to\infty} \mathbb{P}_n(E)$ quando o limite existe e seja $\mathcal{A}$ a família de subconjunto E para os quais o limite existe. Prove que se A e B são elementos disjuntos de $\mathcal{A}$ então $A \cup B \in \mathcal{A}$ e $p(A \cup B) = p(A) + p(B)$ e que esse não é o caso se os eventos não são disjuntos. Prove que p não é enumeravelmente aditiva. Finalmente, prove que p é invariante por translação, ou seja, se $p(A)$ existe então $p(\{a + 1 : a \in A\}) = p(A)$.

EXERCÍCIO 1.59. Prove que o seguinte algoritmo não termina com probabilidade 1.

```
1 j ← 0;
2 repita
3 |   j ← j + 1;
4 |   para cada i ∈ {1, 2, ..., j} faça d_i ←_R {0, 1};
5 até que d_i = 1 para todo i.
```

EXERCÍCIO 1.60. Alice e Bob têm, cada um, um enorme banco de dados que eles querem saber se são iguais. Podemos assumir, sem perda de generalidade, que cada banco de dados é um vetor de n bits $a_1 a_2 \ldots a_n$ para Alice e $b_1 b_2 \ldots b_n$ para Bob. Alice e Bob podem enviar mensagens um ao outro. Uma saída trivial é: Alice envia os n bits para Bob, então Bob verifica se os dois vetores são os mesmos e envia o resultado (sim ou não) para Alice. Toda a comunicação usa $n + 1$ mensagens de um bit.

O seguinte protocolo pressupõe-se mais econômico:

1. Alice escolhe um primo $p \in [n^2, 2n^2]$ e manda para Bob;
2. Alice constrói o polinômio $A(x) = a_1 + a_2 x + a_3 x^2 + \cdots + a_n x^{n-1}$;
3. Bob constrói o polinômio $B(x) = b_1 + b_2 x + b_3 x^2 + \cdots + b_n x^{n-1}$;
4. Alice sorteia $\alpha \in \mathbb{Z}_p = \{0, 1, 2, \ldots, p - 1\}$, calcula $A(\alpha) \bmod p$ e manda α e $A(\alpha) \bmod p$ para Bob;
5. Bob calcula $B(\alpha)$ e manda para Alice se $A(\alpha) = B(\alpha)$ ou não.

Tal protocolo erra se $A \neq B$, porém $A(\alpha) = B(\alpha)$. Verifique que

$$\mathbb{P}[\text{erro}] \leqslant \mathop{\mathbb{P}}_{\alpha\in\mathbb{Z}_p} [(A - B)(\alpha) = 0] \leqslant \frac{n-1}{p}$$

e conclua que $\mathbb{P}[\text{erro}] < 1/n$. Mostre que são usados $O(\log n)$ bits.

EXERCÍCIO 1.61 (Lema do isolamento, (Mulmuley, Vazirani e Vazirani, 1987)). O seguinte resultado diz que, independentemente da natureza de uma família $\mathcal{F}$ de conjuntos, uma atribuição aleatória de pesos aos elementos de $\bigcup_{F\in\mathcal{F}} F$ isola o elemento da família menos pesado com grande probabilidade. Este lema tem muitas aplicações na teoria da computação, em particular, Mulmuley e seus coautores o usaram para projetar um algoritmo aleatorizado paralelizável para encontrar emparelhamento de peso máximo em um grafo (exercícios 2.66 e 2.72).

Sejam E um conjunto finito e $\mathcal{F}$ uma família de subconjuntos de E. Uma *m-ponderação* de E é uma função $p\colon E \to \{1,\dots,m\}$ que atribui pesos inteiros para os elementos de E. O peso de um subconjunto não vazio $S \subset E$ é $p(S) = \sum_{e\in S} p(e)$. A ponderação p é *isolante para* $\mathcal{F}$ se o peso mínimo, $\min_{S\in\mathcal{F}} p(S)$, for alcançado em um único elemento de $\mathcal{F}$. O lema do isolamento é o seguinte resultado

Dado $m \in \mathbb{N}$, para todo conjunto finito E e toda família $\mathcal{F} \subset 2^E$ temos

$$\mathop{\mathbb{P}}_{p\colon E\to\{1,\dots,m\}}[p \text{ é isolante para } \mathcal{F}] \geqslant \left(1-\frac{1}{m}\right)^{|E|}.$$

Para provar esse resultado, suponha que nenhum elemento de $\mathcal{F}$ é um superconjunto de outro elemento de $\mathcal{F}$. Considere P o conjunto de todas as m-ponderações de E e $P^{>1}$ o conjunto de todas as m-ponderações de E que não atribuem o peso 1 a qualquer elemento de E. Defina $\phi\colon P^{>1} \to P$ da seguinte forma: dada a ponderação $p \in P$ tome algum $S_p \in \mathcal{F}$ de peso mínimo de acordo com p e, fazendo $p' = \phi(p)$, defina

$$p'(i) = \begin{cases} p(i)-1 & \text{se } i \in S_p \\ p(i) & \text{se } i \notin S_p. \end{cases}$$

1. Prove que se $p \in P^{>1}$ então p' é isolante em $\mathcal{F}$.
2. Prove que ϕ é injetiva.
3. Prove que $\mathbb{P}_p[p \text{ é isolante em } \mathcal{F}] \geqslant |\phi(P^{>1})|/|P|$.
4. Conclua a demonstração do lema do isolamento.

Bibliografia

Agrawal, Manindra e Somenath Biswas (2003). "Primality and identity testing via Chinese remaindering". Em: *J. ACM* 50.4, 429–443 (electronic) (ver pp. 75, 76).

Agrawal, Manindra, Neeraj Kayal e Nitin Saxena (2004). "PRIMES is in P". Em: *Ann. of Math. (2)* 160.2, pp. 781–793 (ver pp. 4, 67).

Aleliunas, Romas et al. (1979). "Random walks, universal traversal sequences, and the complexity of maze problems". Em: *Proceedings of the 20th Annual Symposium on Foundations of Computer Science*. SFCS '79. Washington, DC, USA: IEEE Computer Society, pp. 218–223. DOI: http://dx.doi.org/10.1109/SFCS.1979.34 (ver p. 156).

Alexander, K. S., K. Baclawski e G. C. Rota (1993). "A stochastic interpretation of the Riemann zeta function". Em: *Proceedings of the National Academy of Sciences of the United States of America* 2.90, 697–699 (ver p. 105).

Arvind, V. e Partha Mukhopadhyay (2008). "Derandomizing the Isolation Lemma and Lower Bounds for Circuit Size". Em: *Approximation, Randomization and Combinatorial Optimization. Algorithms and Techniques*. Ed. por Ashish Goel et al. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 276–289 (ver p. 86).

Aspvall, Bengt, Michael F. Plass e Robert Endre Tarjan (1979). "A linear-time algorithm for testing the truth of certain quantified Boolean formulas". Em: *Inform. Process. Lett.* 8.3, pp. 121–123 (ver p. 144).

Bach, Eric e Jeffrey Shallit (1996). *Algorithmic Number Theory*. Cambridge, MA, USA: MIT Press (ver pp. 61, 62).

Bartle, Robert G. (1976). *The elements of real analysis*. Second. John Wiley & Sons, New York-London-Sydney, pp. xv+480 (ver p. 187).

Bernstein, Daniel J. (1998). "Detecting perfect powers in essentially linear time". Em: *Math. Comput.* 67.223, pp. 1253–1283. DOI: http://dx.doi.org/10.1090/S0025-5718-98-00952-1 (ver p. 76).

Billingsley, P. (1979). *Probability and measure*. Wiley series in probability and mathematical statistics. Probability and mathematical statistics. Wiley (ver p. 81).

Cormen, Thomas H., Charles E. Leiserson e Ronald L. Rivest (1990). *Introduction to algorithms*. The MIT Electrical Engineering and Computer Science Series. Cambridge, MA: MIT Press, pp. xx+1028 (ver pp. 108, 110).

DeMillo, Richard A. e Richard J. Lipton (1978). "A Probabilistic Remark on Algebraic Program Testing". Em: *Inf. Process. Lett.* 7.4, pp. 193–195 (ver p. 57).

Diffie, Whitfield e Martin E. Hellman (1976). "New directions in cryptography". Em: *IEEE Trans. Information Theory* IT-22.6, pp. 644–654 (ver pp. 59, 137).

Erdős, P. (1956). "On pseudoprimes and Carmichael numbers". Em: *Publ. Math. Debrecen* 4, pp. 201–206 (ver p. 69).

Feller, William (1968). *An introduction to probability theory and its applications. Vol. I*. Third edition. New York: John Wiley & Sons Inc., pp. xviii+509 (ver pp. 92, 167, 182).

Freivalds, Rusins (1977). "Probabilistic Machines Can Use Less Running Time". Em: *IFIP Congress*, pp. 839–842 (ver p. 54).

Gao, Shuhong e Daniel Panario (1997). "Tests and Constructions of Irreducible Polynomials over Finite Fields". Em: *Foundations of Computational Mathematics*. Ed. por Felipe Cucker e Michael Shub. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 346–361 (ver p. 66).

Garfinkel, Simson (1994). *PGP: Pretty Good Privacy*. O'Reilly (ver p. 70).

Gathen, Joachim von zur e Jürgen Gerhard (2013). *Modern Computer Algebra*. 3ª ed. Cambridge University Press. DOI: 10.1017/CBO9781139856065 (ver p. 63).

Gelbaum, B.R. e J.M.H. Olmsted (1964). *Counterexamples in Analysis*. Dover books on mathematics. Holden-Day (ver p. 11).

Golub, Gene H. e Charles F. Van Loan (1989). *Matrix computations*. Second. Vol. 3. Johns Hopkins Series in the Mathematical Sciences. Baltimore, MD: Johns Hopkins University Press, pp. xxii+642 (ver p. 171).

Graham, Paul (2002). *A Plan for Spam*. http://www.paulgraham.com/spam.html. Acesso em 06/04/2009 (ver p. 26).

Graham, Ronald L., Donald E. Knuth e Oren Patashnik (1994). *Concrete mathematics*. Second. A foundation for computer science. Reading, MA: Addison-Wesley Publishing Company, pp. xiv+657 (ver p. 78).

Harman, Glyn (2005). "On the Number of Carmichael Numbers up to x". Em: *Bulletin of the London Mathematical Society* 37.5, pp. 641–650. DOI: 10.1112/S0024609305004686. eprint: https://londmathsoc.onlinelibrary.wiley.com/doi/pdf/10.1112/S0024609305004686 (ver p. 69).

Harvey, David e Joris Van Der Hoeven (mar. de 2019). "Integer multiplication in time $O(n \log n)$". working paper or preprint (ver p. 47).

Håstad, Johan (jul. de 2001). "Some Optimal Inapproximability Results". Em: *J. ACM* 48.4, pp. 798–859. DOI: 10.1145/502090.502098 (ver pp. 114, 115).

Haveliwala, Taher e Sepandar Kamvar (2003). *The Second Eigenvalue of the Google Matrix*. Rel. técn. 20. Stanford University (ver p. 171).

Ireland, Kenneth e Michael Rosen (1990). *A classical introduction to modern number theory*. Second. Vol. 84. Graduate Texts in Mathematics. New York: Springer-Verlag, pp. xiv+389 (ver p. 137).

Johnson, David S. (1974). "Approximation algorithms for combinatorial problems". Em: *J. Comput. System Sci.* 9. Fifth Annual ACM Symposium on the Theory of Computing (Austin, Tex., 1973), pp. 256–278 (ver p. 112).

Karger, David R. (1993). "Global min-cuts in RNC, and other ramifications of a simple min-cut algorithm". Em: *Proceedings of the Fourth Annual ACM-SIAM Symposium on Discrete Algorithms (Austin, TX, 1993)*. New York: ACM, pp. 21–30 (ver p. 52).

Kimbrel, Tracy e Rakesh Kumar Sinha (1993). "A probabilistic algorithm for verifying matrix products using $O(n^2)$ time and $\log_2 n + O(1)$ random bits". Em: *Inform. Process. Lett.* 45.2, pp. 107–110. DOI: 10.1016/0020-0190(93)90224-W (ver p. 55).

Klivans, Adam R. e Daniel A. Spielman (2001). "Randomness efficient identity testing of multivariate polynomials". Em: *Proceedings on 33rd Annual ACM Symposium on Theory of Computing, July 6-8, 2001, Heraklion, Crete, Greece*, pp. 216–223. DOI: 10.1145/380752.380801 (ver p. 86).

Knuth, Donald E. (1981). *The art of computer programming. Vol. 2*. Second. Seminumerical algorithms, Addison-Wesley Series in Computer Science and Information Processing. Addison-Wesley Publishing Co., Reading, Mass., pp. xiii+688 (ver p. 54).

Lehmann, Daniel e Michael O. Rabin (1981). "On the advantages of free choice: a symmetric and fully distributed solution to the dining philosophers problem". Em: *POPL '81: Proceedings of the 8th ACM SIGPLAN-SIGACT symposium on Principles of programming languages*. Williamsburg, Virginia: ACM, pp. 133–138. DOI: http://doi.acm.org/10.1145/567532.567547 (ver p. 80).

Lidl, Rudolf e Harald Niederreiter (1997). *Finite fields*. Second. Vol. 20. Encyclopedia of Mathematics and its Applications. With a foreword by P. M. Cohn. Cambridge: Cambridge University Press, pp. xiv+755 (ver pp. 63, 65, 66).

Maltese, George (1986). "A Simple Proof of the Fundamental Theorem of Finite Markov Chains". Em: *The American Mathematical Monthly* 93.8, pp. 629–630 (ver p. 148).

Menezes, Alfred J., Paul C. van Oorschot e Scott A. Vanstone (1997). *Handbook of applied cryptography*. CRC Press Series on Discrete Mathematics and its Applications. With a foreword by Ronald L. Rivest. Boca Raton, FL: CRC Press, pp. xxviii+780 (ver p. 80).

Miller, Gary L. (1975). "Riemann's hypothesis and tests for primality". Em: *Seventh Annual ACM Symposium on Theory of Computing (Albuquerque, N.M., 1975)*. Assoc. Comput. Mach., New York, pp. 234–239 (ver p. 71).

Mitzenmacher, Michael e Eli Upfal (2005). *Probability and computing*. Randomized algorithms and probabilistic analysis. Cambridge: Cambridge University Press, pp. xvi+352 (ver p. 55).

Monier, Louis (1980). "Evaluation and comparison of two efficient probabilistic primality testing algorithms". Em: *Theoretical Computer Science* 12.1, pp. 97 –108. DOI: https://doi.org/10.1016/0304-3975(80)90007-9 (ver p. 67).

Mulmuley, Ketan, Umesh V. Vazirani e Vijay V. Vazirani (1987). "Matching is as easy as matrix inversion". Em: *Combinatorica* 7.1, pp. 105–113 (ver p. 37).

Nightingale, Edmund B., John R. Douceur e Vince Orgovan (2011). "Cycles, Cells and Platters: An Empirical Analysisof Hardware Failures on a Million Consumer PCs". Em: *Proceedings of the Sixth Conference on Computer Systems*. EuroSys '11. Salzburg, Austria: ACM, pp. 343–356. DOI: 10.1145/1966445.1966477 (ver p. 5).

O'Gorman, T. J. et al. (jan. de 1996). "Field Testing for Cosmic Ray Soft Errors in Semiconductor Memories". Em: *IBM J. Res. Dev.* 40.1, pp. 41–50. DOI: 10.1147/rd.401.0041 (ver p. 5).

Page, Lawrence et al. (1998). *The PageRank Citation Ranking: Bringing Order to the Web*. Rel. técn. Stanford Digital Library Technologies Project (ver pp. 169, 171).

Prasolov, V. V. (2006). *Elements of Combinatorial and Differential Topology*. Graduate Studies in Mathematics 74. American Mathematical Society (ver p. 148).

Pugh, William (1989). "Skip lists: a probabilistic alternative to balanced trees". Em: *Algorithms and data structures (Ottawa, ON, 1989)*. Vol. 382. Lecture Notes in Comput. Sci. Berlin: Springer, pp. 437–449 (ver p. 124).

Raab, Martin e Angelika Steger (1998). ""Balls into Bins- A Simple and Tight Analysis". Em: *Proceedings of the Second International Workshop on Randomization and Approximation Techniques in Computer Science*. RANDOM '98. Berlin, Heidelberg: Springer-Verlag, pp. 159–170 (ver p. 132).

Rabin, Michael O. (1980a). "Probabilistic algorithm for testing primality". Em: *J. Number Theory* 12.1, pp. 128–138 (ver p. 71).

— (1980b). "Probabilistic algorithms in finite fields". Em: *SIAM J. Comput.* 9.2, pp. 273–280 (ver p. 66).

Reingold, Omer (2008). "Undirected connectivity in log-space". Em: *J. ACM* 55.4, Art. 17, 24 (ver p. 156).

Ribenboim, Paulo (1996). *The new book of prime number records*. New York: Springer-Verlag, pp. xxiv+541 (ver p. 79).

Rosenthal, Jeffrey S. (2006). *A first look at rigorous probability theory*. Second. World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, pp. xvi+219 (ver p. 11).

Ross, Sheldon M. (2010). *A first course in Probability*. 8th. New Jersey: Prentice Hall (ver pp. 23, 161).

Saldanha, Nicolau (1997). "Precisa-se de alguém para ganhar muito dinheiro". Disponível em http://www.mat.puc-rio.br/~nicolau/publ/papers/otario.pdf. Acesso em 07/2018 (ver p. 34).

Schroeder, Bianca, Eduardo Pinheiro e Wolf-Dietrich Weber (2009). "DRAM Errors in the Wild: A Large-scale Field Study". Em: *Proceedings of the Eleventh International Joint Conference on Measurement and Modeling of Computer Systems*. SIGMETRICS '09. Seattle, WA, USA: ACM, pp. 193–204. DOI: 10.1145/1555349.1555372 (ver p. 5).

Schwartz, Jacob T. (1979). "Probabilistic algorithms for verification of polynomial identities". Em: *Symbolic and algebraic computation (EUROSAM '79, Internat. Sympos., Marseille, 1979)*. Vol. 72. Lecture Notes in Comput. Sci. Berlin: Springer, pp. 200–215 (ver p. 57).

Shpilka, Amir e Amir Yehudayoff (2010). "Arithmetic Circuits: A Survey of Recent Results and Open Questions". Em: *Foundations and Trends® in Theoretical Computer Science* 5.3–4, pp. 207–388. DOI: 10.1561/0400000039 (ver p. 57).

We knew the web was big... (2008). http://googleblog.blogspot.com/2008/07/we-knew-web-was-big.html (ver p. 172).

Wills, Rebecca S. (2006). "Google's PageRank: the math behind the search engine". Em: *Math. Intelligencer* 28.4, pp. 6–11 (ver p. 172).

Zippel, Richard (1979). "Probabilistic algorithms for sparse polynomials". Em: *Symbolic and algebraic computation (EUROSAM '79, Internat. Sympos., Marseille, 1979)*. Vol. 72. Lecture Notes in Comput. Sci. Berlin: Springer, pp. 216–226 (ver p. 57).