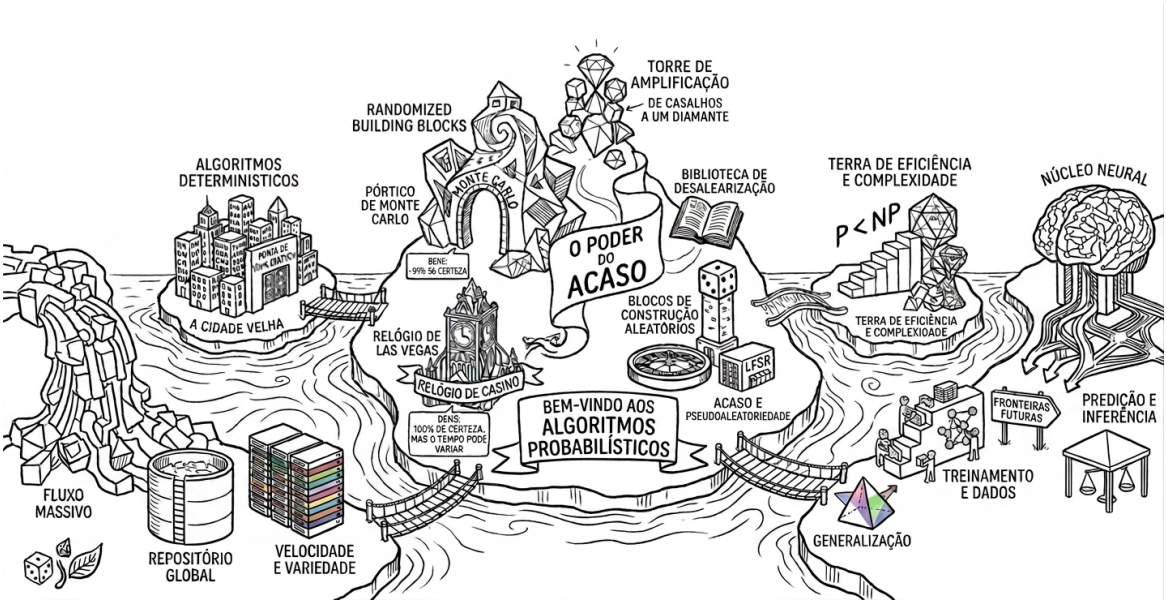
NOTAS DE AULA

ALGORITMOS PROBABILÍSTICOS



MCZA035-17 ALGORITMO PROBABILÍSTICOS (Q3 2026)

PROF.: JAIR DONADELLI

http://professor.ufabc.edu.br/~jair.donadelli/algprob

Sumário

Notação

1. $\mathbb{N} = \{0, 1, 2, \dots\}$ denota o conjunto dos números naturais; $\mathbb{Z}$ denota o conjunto dos números inteiros e $\mathbb{Z}_+ = \{1, 2, \dots\}$ os inteiros positivos; $\mathbb{R}$ denota o conjunto dos números reais.

2. ":=" tem o significado de "é definido igual a".

3. O conjunto de todos os subconjuntos de B com cardinalidade k é denotado $\binom{B}{k}$.

   O conjunto de todos os subconjuntos de B, ou seja, o ***conjunto das partes*** de B, é denotado por 2^B.

   Se A e B são conjuntos então B^A é o c***onjunto de todas as funções de*** A ***em*** B.

   $\overline{A}$ para o ***complemento*** $\Omega \setminus A$.

4. ln é a função logaritmo natural, ou seja, logaritmo na base e $=$ 2,71 . . . . Em base a, usamos $\log_a$ e omitimos a no caso $a = 2$.

5. Usamos $\sum_{i \geqslant 1}$ com o mesmo significado de $\sum_{i=1}^{\infty}$ para séries.

   Também, usamos $\bigcup_{i \geqslant 1}$ tanto quanto $\bigcup_{i=1}^{\infty}$ para operações sobre conjuntos.

1
Noções de probabilidade

Intuitivamente, uma medida de probabilidade é uma forma quantitativa de expressar a chance de ocorrência de eventos associados a um experimento aleatório, cujos resultados não podem ser previamente determinados. A teoria da probabilidade é a área da matemática dedicada à modelagem e ao estudo rigoroso desses fenômenos sob condições de incerteza.

Uma medida de probabilidade $\mathbb{P}$ é 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 que satisfaz essas propriedades é chamada de ***σ-álgebra de subconjuntos de Ω***.

Um ***espaço de probabilidade***, ou 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; e $\mathbb{P}\colon \mathcal{A} \to [0,1]$ é uma ***medida de probabilidade***, que atribui a cada evento aleatório um número real satisfazendo

Neste texto, usaremos os termos *espaço de probabilidade* e *modelo probabilístico* de maneira intercambiável, sem distinção.

Não-negatividade $\mathbb{P}(A) \geqslant 0$ para todo $A \in \mathcal{A}$.
Normalização $\mathbb{P}(\Omega) = 1$.
Aditividade enumerável se $\{A_i \in \mathcal{A} \colon i \geqslant 1\}$ é um conjunto enumerável de eventos mutuamente exclusivos, então

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

O lado esquerdo dessa igualdade não depende de uma enumeração particular dos conjuntos A_i e o mesmo vale para o lado direito.

Proposição 1.1: Propriedades de $\mathbb{P}$

Para uma medida de probabilidade $\mathbb{P}$ sobre um espaço de eventos $\mathcal{A}$ do espaço amostral Ω valem as seguintes propriedades.

1. A probabilidade do ***evento impossível*** é
$$\mathbb{P}(\varnothing) = 0.$$
2. ***Aditividade finita:*** se $A_1, A_2, \ldots, A_n \in \mathcal{E}$ são eventos mutuamente exclusivos então
$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathbb{P}(A_i). \tag{1.1}$$
3. ***A probabilidade do complemento*** é
$$\mathbb{P}(\overline{A}) = 1 - \mathbb{P}(A)$$
para todo evento $A \in \mathcal{E}$.
4. ***Monotonicidade***: se $A \subseteq B$, então $\mathbb{P}(A) \leqslant \mathbb{P}(B)$. Também, $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A)$.
5. ***Regra da adição***: Para quaisquer A, B
$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).$$
6. ***Subaditividade***: Se $A_1, A_2, \ldots, A_n$ são eventos, então
$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) \leqslant \sum_{i=1}^{n} \mathbb{P}(A_i). \tag{1.2}$$

Demonstração. Fazendo $A_1 = \Omega$ e $A_i = \varnothing$ para todo $i \geqslant 2$ temos, pela aditividade enumerável, que

$$\mathbb{P}(\Omega) = \mathbb{P}(\Omega \cup \varnothing \cup \varnothing \cup \cdots \cup \varnothing \cup \cdots) = \mathbb{P}(\Omega) + \sum_{i=2}^{\infty} \mathbb{P}(\varnothing)$$

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

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

que é o resultado afirmado no item 2.

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

Para monotonicidade, consideremos A e B eventos tais que $A \subseteq B$. Usamos que B pode ser escrito como a união disjunta $A \cup (B \setminus A)$,

donde $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A)$ e como $\mathbb{P}(\overline{A} \cap B) \geqslant 0$ temos $\mathbb{P}(B) \geqslant \mathbb{P}(A)$. Notemos que, como consequência imediata, temos $\mathbb{P}(A) \leqslant 1$ para todo evento A.

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

$$\mathbb{P}(A \cup B) = \mathbb{P}(A \setminus B) + \mathbb{P}(B \setminus A) + \mathbb{P}(A \cap B). \tag{1.3}$$

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 (1.3) prova a regra da adição. □

1.1 *Convenções notacionais*

Sejam A uma descrição de um evento aleatório e A_Ω o subconjunto que o representa em $(\Omega, \mathcal{A}, \mathbb{P})$. Denotamos por $\mathbb{P}[A]$ a probabilidade do evento descrito por A, isto é, $\mathbb{P}[A] \coloneqq \mathbb{P}(A_\Omega)$. Caso haja a necessidade de evidenciar o espaço amostral escreveremos

$$\mathop{\mathbb{P}}_{\Omega}[A] \quad \text{ou} \quad \mathop{\mathbb{P}}_{\omega \in \Omega}[\text{ocorre } A] \quad \text{ou} \quad \mathop{\mathbb{P}}_{\omega \in \Omega}[\omega \in A] \tag{1.4}$$

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

Por exemplo, no caso em que uma moeda equilibrada é lançada até sair coroa, o evento caracterizado por "o número de lançamentos é par" é definido em $(\Omega, \mathcal{A}, \mathbb{P})$ por $A_\Omega \coloneqq \{(c_1, \ldots, c_i) : i \text{ é par}\}$ e $\mathbb{P}[\text{número par de lançamentos}]$ denota o mesmo que $\mathbb{P}(A_\Omega)$.

Quando Ω é finito e a menos que seja dada explicitamente outra medida, então a notação na equação (1.4) 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 e Ω um conjunto finito de números. A probabilidade de que o sorteio de um elemento de Ω, todos com a mesma probabilidade de ocorrência, resulte numa raiz do polinômio é descrita por

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

que é a probabilidade do evento $R = \{\omega \in \Omega : p(\omega) = 0\}$.

Nos algoritmos assumiremos a possibilidade de se fazer escolhas aleatórias, ou seja, assumiremos que os algoritmos dispõem da instrução

$$a \xleftarrow{R} \Omega$$

chamada de atribuição por uma ***escolha aleatória uniforme*** em Ω (finito), o que significa que a assume qualquer um dos elementos de

Ω com igual probabilidade, a saber $1 / |\Omega|$. Por exemplo, a partir de uma fonte de bits aleatórios, a instrução

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

denota o fato de que a é uma variável do algoritmo e que após a execução da atribuição $\xleftarrow{\mathrm{R}}$ o valor da variável a é um elemento qualquer de $\{ 0, 1 \}$ com probabilidade $1/2$.

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)
$A \subseteq B$	se ocorre A, então ocorre B	$\omega \in A \Rightarrow \omega \in B$ (inclusão)

Tabela 1.1: pequeno dicionário de termos da probabilidade.

1.2 *Probabilidade discreta*

Um ***modelo probabilístico discreto***, ou ***espaço de probabilidade discreto***, é um espaço $(\Omega, \mathcal{E}, \mathbb{P})$ em que Ω é enumerável (finito ou infinito) e $\mathcal{E} = 2^{\Omega}$.

Se A é conjunto, 2^A denota o conjunto das partes de A.

Nesse caso, todo experimento tem seu modelo probabilístico especificado quando estabelecemos

1. um espaço amostral enumerável Ω;
2. uma função $p \colon \Omega \to [0, 1]$ tal que $\sum_{\omega \in \Omega} p(\omega) = 1$, chamada ***função massa de probabilidade***.

De fato, dado (Ω, p) como acima, podemos definir uma função $\mathbb{P}$ em 2^{Ω} tomando

$$\mathbb{P}(A) := \sum_{\omega \in A} p(\omega), \qquad A \subseteq \Omega. \tag{1.5}$$

Essa soma está bem definida e $0 \leqslant \mathbb{P}(A) \leqslant 1$ para todo $A \subseteq \Omega$. Claramente, $\mathbb{P}(\Omega) = \sum_{\omega} \mathbb{P}(\{ \omega \}) = 1$. Além disso, se A_i para $i \geqslant 1$ são eventos mutuamente exclusivos, então

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

segue da convergência absoluta e da exclusão mútua. Portanto, $\mathbb{P}$ é uma medida de probabilidade sobre 2^{Ω}.

Assim, em um espaço de probabilidade discreto, a função massa de probabilidade determina completamente a medida de probabilidade.

Convencionamos a notação

$$\mathbb{P}(\,\omega\,) := \mathbb{P}(\,\{\,\omega\,\}\,)$$

para os eventos elementares.

Exemplo 1.1: Moeda equilibrada

Uma moeda equilibrada é lançada repetidamente até sair coroa. No espaço amostral definimos a função massa de probabilidade

$$p((c_1, c_2, \ldots, c_i)) = \begin{cases} 2^{-i} & \text{para } c_j = \text{Ca se } j < i \text{ e } c_i = \text{Co}, \\ 0 & \text{nos outros casos.} \end{cases}$$

Ω e p definem um modelo probabilístico discreto para o experimento.

Monty Hall

Vejamos um modelo probabilístico para o problema de Monty Hall: o apresentador Monty Hall mostra três portas a um espectador que concorre ao prêmio escondido atrás das portas.

Monty Hall é o nome do apresentador de um concurso televisivo que era exibido na década de 1970, nos Estados Unidos, chamado *Let's Make a Deal*.

O protocolo é o seguinte: Monty Hall escolhe, ao acaso, uma das portas para esconder um carro; nas outras duas esconde um bode em 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 e escolhendo ao acaso se houver mais de uma possibilidade. 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. Qual é a estratégia que maximiza a chance de ganhar o carro?

O experimento 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.

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

$$\{(1,2,3),(1,3,2),(2,1,3),(2,3,1),(3,1,2),(3,2,1),$$
$$(1,1,2),(1,1,3),(2,2,1),(2,2,3),(3,3,1),(3,3,2)\}.$$

e probabilidades de acordo com o diagrama de árvore mostrado na Figura 1.1 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

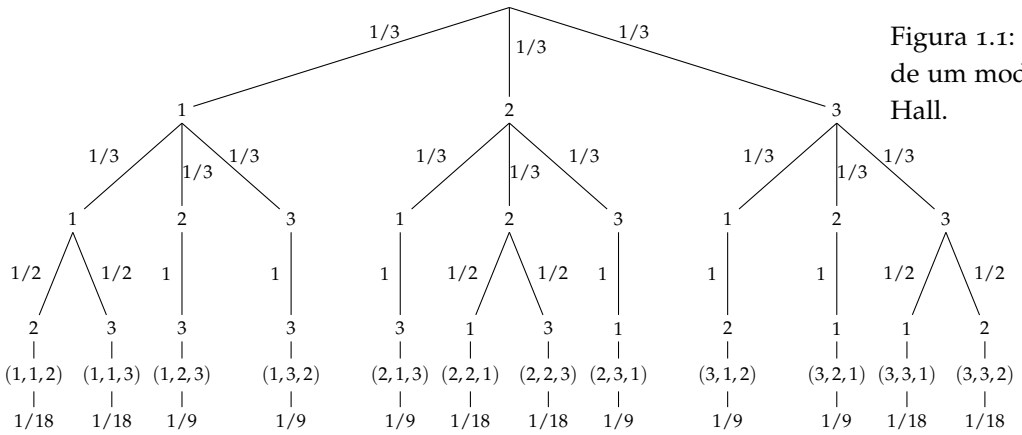


Figura 1.1: diagrama de árvore de um modelo para Monty Hall.

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

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 pois isso dobra a probabilidade de ganhar o carro.

1.3 *Probabilidade condicional*

Em um espaço de probabilidade discreto definido por Ω e $\mathbb{P}$, a probabilidade de ocorrência do evento A condicionada à ocorrência o evento E, ou como dizemos a ***probabilidade condicional*** de A dado E, em que $\mathbb{P}(E) \neq 0$, é definida por

$$\mathbb{P}(A \mid E) := \frac{\mathbb{P}(A \cap E)}{\mathbb{P}(E)} \tag{1.6}$$

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

$$\mathbb{P}(A \mid E) = \frac{\mathbb{P}(A \cap E)}{\mathbb{P}(E)} = \frac{|A \cap E|}{|E|} \tag{1.7}$$

Ideia importante

Considere um espaço de probabilidade $(\Omega, \mathcal{A}, \mathbb{P})$ e um evento com probabilidade positiva $E \in \mathcal{A}$.
$\mathbb{P}_E(A) := \mathbb{P}(A \mid E)$ é uma medida de probabilidade para os eventos em $\mathcal{A}$ (i.e, satisfaz os axiomas de probabilidade).
Também, $(E, \{A \cap E : A \in \mathcal{A}\}, \mathbb{P}_E)$ é um espaço de probabilidade.

Regra do produto: A igualdade, que decorre da definição,

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

é consequência direta da definição de probabilidade condicional e é conhecida como ***teorema da multiplicação*** ou ***regra do produto***.

Essa regra é facilmente estendida para calcular $\mathbb{P}(A \cap B \cap C)$:

$$\begin{aligned}\mathbb{P}(A \cap B \cap C) &= \mathbb{P}(A) \cdot \mathbb{P}(B \cap C \mid A) \\ &= \mathbb{P}(A) \cdot \mathbb{Q}(B \cap C) \\ &= \mathbb{P}(A) \cdot \mathbb{Q}(B) \cdot \mathbb{Q}(C \mid B)\end{aligned}$$

onde $\mathbb{Q}$ é a medida $\mathbb{P}(\cdot \mid A)$. Mas

$$\mathbb{Q}(C \mid B) = \frac{\mathbb{P}(B \cap C \mid A)}{\mathbb{P}(B \mid A)} = \frac{\frac{\mathbb{P}(A \cap B \cap C)}{\mathbb{P}(A)}}{\frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}} = \mathbb{P}(C \mid A \cap B)$$

assim

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A) \cdot \mathbb{P}(B \mid A) \cdot \mathbb{P}(C \mid A \cap B) \tag{1.8}$$

(identifique o uso dessa regra no modelo probabilístico para o problema de Monty Hall, página 9).

Um caso geral desse teorema é dado abaixo, cuja prova é exercício.

Teorema 1.1: Teorema da multiplicação

Sejam $A_1, A_2, \ldots, A_n$ eventos de um modelo probabilístico. Prove que

$$\begin{aligned}\mathbb{P}(\, A_1 \cap \cdots \cap A_n \,) = &\mathbb{P}(\, A_1 \,) \cdot \mathbb{P}(\, A_2 \mid A_1 \,) \cdot \\ &\mathbb{P}(\, A_3 \mid A_2 \cap A_1 \,) \cdots \\ &\mathbb{P}(\, A_n \mid A_{n-1} \cap A_{n-2} \cap \cdots \cap A_1 \,)\end{aligned} \tag{1.9}$$

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

A Lei da Probabilidade Total: 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,

$$\begin{aligned}\mathbb{P}(\, A \,) &= \mathbb{P}\big(\, (A \cap E) \cup (A \cap \overline{E}) \,\big) \\ &= \mathbb{P}(\, A \cap E \,) + \mathbb{P}(\, A \cap \overline{E} \,) &(1.10)\\ &= \mathbb{P}(\, A \mid E \,)\, \mathbb{P}(\, E \,) + \mathbb{P}(\, A \mid \overline{E} \,)\, \mathbb{P}(\, \overline{E} \,). &(1.11)\end{aligned}$$

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 (1.11)

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

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 $\mathbb{P}(\, E_i \,) > 0$ para todo $i \geqslant 1$, então vale

$$\mathbb{P}(\, A \,) = \sum_{i=1}^{n} \mathbb{P}(\, A \cap E_i \,) = \sum_{i=1}^{n} \mathbb{P}(\, A \mid E_i \,) \mathbb{P}(\, E_i \,) \tag{1.12}$$

para qualquer evento A. Deixamos a prova por conta do leitor do resultado abaixo considerando partições enumeráveis.

Teorema 1.2: Lei da probabilidade total

Seja $\{ E_i : i \geqslant 1 \}$ uma partição (enumerável) do espaço amostral Ω com $\mathbb{P}(E_i) > 0$ para todo i. Então

$$\mathbb{P}(A) = \sum_{i=1}^{\infty} \mathbb{P}(A \cap E_i) = \sum_{i=1}^{\infty} \mathbb{P}(A \mid E_i)\,\mathbb{P}(E_i) \tag{1.13}$$

para qualquer evento A.

Demonstração. Os conjuntos $A \cap E_i$ são disjuntos dois a dois e da união deles resulta A, logo a primeira igualdade em (1.13) segue de $\mathbb{P}$ ser medida de probabilidade. 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 1.2

As seguradoras de automóveis classificam motoristas em *mais propensos a acidentes* e *menos propensos a acidentes*. Com isso estimam que os mais propensos são 30% da população e que esses se envolvem em acidente no período de um ano com probabilidade 0,4, enquanto que os menos propensos a acidentes se envolvem em acidente no período de um ano com probabilidade 0,2. Denotemos por A^+ o evento definido pelos motoristas mais propensos a acidentes. Então a probabilidade de um novo segurado se envolver em acidente em um ano é

$$\begin{aligned}\mathbb{P}(A_1) &= \mathbb{P}(A_1 \mid A^+)\,\mathbb{P}\big(A^+\big) + \mathbb{P}(A_1 \mid \overline{A^+})\,\mathbb{P}(\overline{A^+}) \\ &= 0{,}4 \cdot 0{,}3 + 0{,}2 \cdot 0{,}7 = 0{,}26\end{aligned}$$

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

$$\begin{aligned}\mathbb{P}(A^+ \mid A_1) &= \frac{\mathbb{P}(A_1 \cap A^+)}{\mathbb{P}(A_1)} = \frac{\mathbb{P}(A_1 \mid A^+)\,\mathbb{P}(A^+)}{\mathbb{P}(A_1)} \\ &= \frac{0{,}4 \cdot 0{,}3}{0{,}26} = \frac{6}{13}.\end{aligned}$$

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 $\mathbb{P}(A^+ \mid \overline{A}_1) \approx 0{,}24$.(Ross, 2010)

1.3.1 *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.

Definição 1.1: Independência

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***.

É imediato da Definição 1.1 o seguinte fato.

Proposição 1.2

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

Se os eventos têm probabilidade não nula então decorre da definição acima que a independência de A e B equivale a

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

Por exemplo, de

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

temos que "a soma é 7" é um evento independente de "o primeiro resultado é 4" se o experimento é lançar um dado duas vezes.

Exemplo 1.3: 2 lançamentos de uma moeda

Considere

$$\Omega = \{\,\mathrm{CaCa}, \mathrm{CaCo}, \mathrm{CoCa}, \mathrm{CoCo}\,\},$$

onde todos os resultados têm probabilidade $1/4$.
Considere os eventos

$$A = [\text{o primeiro lançamento é cara}] = \{\,\mathrm{CaCa}, \mathrm{CaCo}\,\},$$

$$B = [\text{os dois lançamentos são iguais}] = [\text{CaCa}, \text{CoCo}].$$

Intuitivamente, pode parecer que os eventos são dependentes. De fato, saber que os dois lançamentos são iguais restringe os resultados possíveis a $\{\,\text{CaCa}, \text{CoCo}\,\}$. Assim, parece que essa informação deveria afetar a probabilidade de o primeiro lançamento ser cara. Entretanto,

$$\mathbb{P}(\,A\,) = \frac{1}{2} \qquad \text{e} \qquad \mathbb{P}(\,B\,) = \frac{1}{2}.$$

Além disso, $A \cap B = \{\,\text{CaCa}\,\}$, de modo que

$$\mathbb{P}(\,A \cap B\,) = \frac{1}{4}.$$

Portanto, $\mathbb{P}(\,A \cap B\,) = \mathbb{P}(\,A\,)\mathbb{P}(\,B\,)$. Logo, A e B são independentes.
Equivalentemente,

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

Assim, embora a ocorrência de B elimine metade dos resultados possíveis, ela não altera a probabilidade de A. Esse exemplo ilustra que a independência não significa que dois eventos não tenham qualquer relação ou que a ocorrência de um não forneça nenhuma informação sobre o experimento. O que importa é se a ocorrência de um evento altera ou não a probabilidade do outro. Embora saber que os dois lançamentos são iguais forneça informação sobre os resultados possíveis, essa informação não altera a probabilidade de o primeiro lançamento ser cara.

Uma família enumerável de eventos $\mathcal{E} = \{\,E_n : n \in I\,\}$ é dita ***mutuamente independente*** se para todo subconjunto finito $J \subseteq I$ vale que

$$\mathbb{P}\left(\bigcap_{\ell \in J} E_\ell\right) = \prod_{\ell \in J} \mathbb{P}(\,E_\ell\,). \tag{1.14}$$

Para $k \geqslant 2$ fixo, dizemos que a coleção $\mathcal{E}$ é ***k-a-k independente*** se todo subconjunto de índices $J \subseteq I$ com $|J| \leqslant k$ define uma subcoleção de eventos mutuamente independentes.

1.3.2 Teste de identidade polinomial

Nosso primeiro exemplo de um algoritmo aleatorizado é um teste de identidade entre polinômios: dados dois polinômios p e q sobre x,

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"? Por ora, deixemos de lado tais questões e vamos ver como decidir, se dado um polinômio f por uma "caixa preta", da qual a função polinomial $p(x)$ pode ser avaliada em qualquer número x, o polinômio dado f é identicamente nulo. Um algoritmo que resolve esse problema também resolve o problema original, basta tomarmos $f(x) = p(x) - q(x)$.

Fato 1. Esse problema aparentemente simples pode ser generalizado para polinômios em várias variáveis. Um algoritmo determinístico eficiente para o caso multivariado tem consequências profundas em Teoria da Computação, especificamente, em Complexidade Computacional.

Um algoritmo para esse 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 *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.

Algoritmo 1.1: teste de identidade entre polinômios

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

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,

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

Podemos fazer essa probabilidade arbitrariamente pequena ao custo de mais computação. Intuitivamente, como no caso dos lançamentos de moedas, imagine tal algoritmo sendo executado k vezes com a mesma entrada.

Basta uma rodada responder *não* para que a resposta definitiva seja *não*. Se todas as rodadas respondem *sim*, qual a probabilidade de todas estarem erradas?

$$\mathbb{P}[\,\text{erro em } k \text{ rodadas independentes}\,] \leqslant \left(\frac{1}{4}\right)^k.$$

Se $k = 10$ a probabilidade de erro é a probabilidade de 10 respostas *sim*, ou seja, no máximo $(1/4)^{10} < 10^{-6}$ (em metros é menos que o diâmetro do fio da teia de uma aranha), valor desprezível na prática.

Ideia importante

Assumimos que os sorteios realizados durante a execução de um algoritmo são independentes. Além disso, sorteios realizados em diferentes execuções do mesmo algoritmo são independentes entre si.

Algoritmos Monte Carlo Algoritmos aleatorizados com tempo de execução que depende somente da entrada e não dos sorteios e que podem responder errado, são conhecidos na literatura como algoritmos ***Monte Carlo***. O teste de identidade polinomial é um algoritmo que pode errar, no entanto, somos capazes de delimitar a probabilidade de tal solução incorreta ocorrer.

Observamos uma propriedade útil de um algoritmo Monte Carlo: se o algoritmo for executado repetidamente com escolhas aleatórias independentes a cada vez, a probabilidade de falha pode se tornar arbitrariamente pequena, às custas do tempo de execução.

Observação 1 (*erros de hardware*). Para um base de comparação, sem a intenção de ser rigoroso, *hardware* apresentam falhas no processamento, transmissão e armazenamento de informação.Estudos da IBM realizados na década de 1990 indicaram que os efeitos de radiação cósmica podem produzir soft errors em memórias semicondutoras.

T. J. O'Gorman et. al. (1996). *Field testing for cosmic ray soft errors in semiconductor memories*. IBM J. Res. Dev., 40(1):41–50.

Uma estimativa frequentemente associada aos estudos da IBM é de aproximadamente um erro induzido por raios cósmicos para cada 256 MB de memória por mês. Essa taxa corresponde, aproximadamente, a $\approx 1{,}4 \times 10^{-15}$ por byte-segundo.

Schroeder, Pinheiro e Weber (2009), em um estudo de larga escala realizado em uma grande frota de servidores do Google, analisaram erros de DRAM durante aproximadamente 2,5 anos. Os autores encontraram taxas de erro muito superiores às previstas por estudos laboratoriais anteriores, da ordem de 25.000 a 70.000 erros por bilhão de horas de dispositivo por Mbit, além de observarem que os erros registrados eram predominantemente associados a falhas permanentes (*hard errors*), e não a *soft errors* causados por radiação[1].

Bianca Schroeder et. al. (2009). *DRAM errors in the wild: A large-scale field study*. Proceedings of the Eleventh International Joint Conference on Measurement and Modeling of Computer Systems, SIGMETRICS '09, pages 193–204.

[1] Assim, esses resultados não devem ser interpretados diretamente como uma taxa de erros induzidos por raios cósmicos.

Como ilustração da grande variabilidade das taxas de erro reportadas na literatura, taxas de 120 a 81.000 FIT por Mbit foram posteriormente compiladas em trabalhos sobre confiabilidade de hardware. Como 1 FIT corresponde a uma falha por 10^9 horas, essas taxas correspondem aproximadamente a um erro a cada 950 anos e a cada 1,4 anos, respectivamente, para cada Mbit de memória. Para uma máquina com 4 GiB de DRAM, essas taxas corresponderiam, respectivamente, a aproximadamente 3 e 2.000 erros por mês, supondo uma taxa constante e proporcional à quantidade de memória.

Nightingale et al. (2011), em um estudo da Microsoft envolvendo aproximadamente 950.000 computadores, encontraram uma probabilidade de aproximadamente 1/2700 de uma primeira falha classificada como DRAM one-bit flip entre as máquinas com pelo menos cinco dias de tempo acumulado de CPU (TACT). Para máquinas com pelo menos 30 dias de TACT, essa probabilidade era de aproximadamente 1/1700. Esses valores representam probabilidades de falha observada no sistema, e não uma taxa de erro por bit ou por unidade de tempo.

Edmund B. Nightingale et. al. (2011). *Cycles, cells and platters: An empirical analysisof hardware failures on a million consumer PCs*. Proceedings of the Sixth Conference on Computer Systems, EuroSys '11, pages 343–356.

Amplificação

Para ***problemas de decisão*** — problemas computacionais para os quais a resposta para uma instância é "sim" ou "não"), — existem dois tipos de algoritmos Monte Carlo:

- com ***erro unilateral*** (*one-sided error*), se a probabilidade de errar for zero para pelo menos uma das respostas possíveis;
- com ***erro bilateral*** (*two-sided error*), se houver uma probabilidade não nula de ele errar tanto quando responde "sim" quanto quando responde "não".

Para algoritmos Monte Carlo de erro unilateral, com probabilidade de erro $\leqslant \varepsilon$, para $\varepsilon \in (0,1)$, podemos usar a estratégia de repetições

independentes para atingir um limiar $\delta \in (0,1)$ para a probabilidade de erro.

Se A é um algoritmo Monte Carlo que erra apenas nas respostas "sim", então em t rodadas do algoritmo com a mesma entrada o erro só se dá *se todas as t respostas forem* "sim". A probabilidade de erro será no máximo δ para todo t tal que $\varepsilon^t \leqslant \delta$, portanto, para todo

$$t \geqslant \frac{\ln(1/\delta)}{\ln(1/\varepsilon)}.$$

Para A que erra apenas nas respostas "não" o resultado é análogo.

Proposição 1.3: Amplificação unilateral

Dados um limiar $\delta \in (0,1)$ e um algoritmo Monte Carlo A de erro unilateral limitado por ε, com $\lceil \ln(1/\delta)/\ln(1/\varepsilon) \rceil$ ou mais repetições de A temos um algoritmo Monte Carlo de erro (unilateral) no máximo δ. □

Assim, dado conhecido um algoritmo probabilístico A de erro unilateral ε, com $O(\ln(1/\delta))$ repetições baixamos o erro para $< \delta$.

No caso de erro bilateral nenhuma das duas respostas é definitiva, a estratégia é repetir o algoritmo t vezes e devolver a resposta de maior ocorrência.

Proposição 1.4: Amplificação bilateral

Seja $\varepsilon < 1/2$ fixo. Dados um limiar $\delta \in (0,1)$ e um algoritmo Monte Carlo A de erro bilateral no máximo ε, um número $O(\ln(1/\delta))$ de repetições independentes de A fornece um algoritmo Monte Carlo de erro bilateral no máximo δ. A constante implícita depende de ε.

Demonstração. Se $\varepsilon = 0$, o próprio A já tem erro $0 \leqslant \delta$ e nada há a fazer. Suponha, então, $0 < \varepsilon < \frac{1}{2}$.

O algoritmo amplificado executa A independentemente $t = 2k+1$ vezes e retorna a resposta majoritária, que é bem definida por ser t ímpar. Ao final da prova escolhemos k em função de δ e de ε.

Seja S o número dessas t execuções que produzem a resposta correta. Cada execução está correta com probabilidade $p \geqslant 1-\varepsilon > \frac{1}{2}$, independentemente das demais, de modo que

$$\mathbb{P}[X = i] = \binom{t}{i} p^i (1-p)^{t-i}$$

que é não crescente em p, e portanto basta tratar o pior caso $p = 1-\varepsilon$; suporemos isso daqui em diante. Escrevendo

$$p = 1-\varepsilon = \frac{1}{2} + q, \qquad \text{onde } q := \frac{1}{2} - \varepsilon \in \left(0, \tfrac{1}{2}\right),$$

temos $1 - p = \frac{1}{2} - q = \varepsilon$ e

$$\mathbb{P}[\, S = i \,] = \binom{2k+1}{i} \left(\frac{1}{2} + q\right)^{i} \left(\frac{1}{2} - q\right)^{2k+1-i}.$$

Como t é ímpar, não há empates: o algoritmo amplificado erra exatamente quando a resposta correta é minoritária, isto é, quando $S \leqslant k$. Sua probabilidade de erro é, pois,

$$\mathbb{P}[\, S \leqslant k \,] = \sum_{i=0}^{k} \binom{2k+1}{i} \left(\frac{1}{2} + q\right)^{i} \left(\frac{1}{2} - q\right)^{2k+1-i}.$$

Estimamos os dois fatores separadamente. Para $i \leqslant k$ vale $2k + 1 - i \geqslant k + 1$ e, como $\frac{1}{2} + q > \frac{1}{2} - q$, trocar um fator $\left(\frac{1}{2} - q\right)$ por $\left(\frac{1}{2} + q\right)$ só pode aumentar o produto; aplicando essa troca $k - i$ vezes,

$$\left(\frac{1}{2} + q\right)^{i} \left(\frac{1}{2} - q\right)^{2k+1-i} \leqslant \left(\frac{1}{2} + q\right)^{k} \left(\frac{1}{2} - q\right)^{k+1}.$$

Quanto aos coeficientes binomiais, a identidade

$$\binom{2k+1}{i} = \binom{2k+1}{2k+1-i}$$

estabelece uma bijeção entre os índices $0, \ldots, k$ e os índices $k + 1, \ldots, 2k + 1$ que preserva o valor do coeficiente. Como esses dois conjuntos particionam $\{\, 0, 1, \ldots, 2k+1 \,\}$, cada um deles responde por metade da soma total, e o Teorema Binomial dá

$$\sum_{i=0}^{k} \binom{2k+1}{i} = \frac{1}{2} \sum_{i=0}^{2k+1} \binom{2k+1}{i} = \frac{2^{2k+1}}{2} = 2^{2k}.$$

Combinando as duas estimativas e usando $\frac{1}{2} - q = \varepsilon$,

$$\mathbb{P}[\, S \leqslant k \,] \leqslant 2^{2k} \left(\frac{1}{2} + q\right)^{k} \left(\frac{1}{2} - q\right)^{k+1} = \varepsilon \left[4 \left(\frac{1}{2} + q\right) \left(\frac{1}{2} - q\right)\right]^{k} = \varepsilon \left(1 - 4q^2\right)^{k} \leqslant \left(1 - 4q^2\right)^{k}.$$

Escreva

$$\alpha := 1 - 4q^2 = 1 - 4 \left(\frac{1}{2} - \varepsilon\right)^{2}.$$

De $0 < q < \frac{1}{2}$ segue $0 < \alpha < 1$, e o que provamos é que a probabilidade de o algoritmo amplificado produzir uma resposta incorreta é no máximo $\alpha^k = \alpha^{(t-1)/2}$. Essa cota é no máximo δ sempre que

$$k \geqslant \frac{\ln(1/\delta)}{\ln(1/\alpha)},$$

de modo que basta tomar

$$k = \left\lceil \frac{\ln(1/\delta)}{\ln(1/\alpha)} \right\rceil, \qquad t = 2k + 1.$$

Sendo ε fixo, $\ln(1/\alpha)$ é uma constante positiva e, consequentemente, $t = O(\ln(1/\delta))$. □

1.3.3 *Continuidade de uma medida de probabilidade*

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

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

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

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

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

Se $(A_n : n \geqslant 1)$ 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 $(A_n : n \geqslant 1)$ é decrescente tomamos os complementos e temos que $(\overline{A_n} : n \geqslant 1)$ é crescente, portanto,

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

por outro lado

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

portanto

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

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

$$\mathbb{P}\Big(\lim_{n\to\infty} A_n\Big) = \lim_{n\to\infty} \mathbb{P}(A_n). \tag{1.15}$$

A propriedade (1.15) é chamada de ***continuidade da medida de probabilidade***.

Consideremos novamente o exemplo do lançamento de uma moeda equilibrada até sair coroa

$$\Omega = \{\, (\mathrm{Co}), (\mathrm{Ca}, \mathrm{Co}), \ldots, (\mathrm{Ca}, \mathrm{Ca}, \ldots) \,\}$$

munido da função massa de probabilidade $p((c_1, c_2, \ldots, c_n)) = 2^{-n}$ para a sequência que termina em coroa após n lançamentos.

Definimos o evento A_n definido por "não sai coroa até o n-ésimo lançamento". Esse evento ocorre com probabilidade

$$1 - \mathbb{P}(\,\overline{A_n}\,) = 1 - \sum_{i=1}^{n} \left(\frac{1}{2}\right)^i = 1 - \frac{1/2 - (1/2)^{n+1}}{1/2} = \left(\frac{1}{2}\right)^n.$$

Intuitivamente, "nunca sair coroa" tem probabilidade $\lim_{n\to\infty} 2^{-n} = 0$. Essa passagem ao limite, $\mathbb{P}(\lim_{n\to\infty} A_n) = \lim_{n\to\infty} \mathbb{P}(A_n)$, é garantida pela continuidade.

A sequência $(A_n : n \geqslant 1)$ é monótona pois $A_n \supseteq A_{n-1}$ para todo $n > 1$, portanto,

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

Ideia importante

A interpretação correta desse fato é que "sair cara para sempre" é um resultado possível, mas tem probabilidade zero. Isso porque, nesse caso, o espaço amostral não é discreto: o resultado $(\mathrm{Ca}, \mathrm{Ca}, \ldots)$ é um elemento do espaço amostral, mas o conjunto que contém apenas esse resultado tem probabilidade zero. Em contraste, *se o espaço amostral é discreto e todo resultado elementar tem probabilidade positiva*, então

$$\mathbb{P}(A) = 0 \;\Rightarrow\; A = \varnothing.$$

De fato, se $A \neq \varnothing$, existe algum resultado elementar $\omega \in A$ e, pela monotonicidade da probabilidade, $\mathbb{P}(A) \geqslant \mathbb{P}(\omega) > 0$. *Nesse contexto, portanto, um evento de probabilidade zero é necessariamente impossível.*

1.4 Um gerador de números aleatórios

Dada uma fonte que gera bits aleatórios de modo uniforme e independente, como projetar um algoritmo que recebe um inteiro positivo M e 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 $k = \log M$ 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 $k = \lceil \log M \rceil$ bits aleatórios), 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.

Algoritmo 1.2: Gerador de números aleatórios

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

1 **repita**
2 | **para cada** $i \in \{0, \ldots, \lceil \log M \rceil - 1\}$ **faça** $d_i \xleftarrow{R} \{0, 1\}$;
3 | $N \leftarrow \sum_i d_i 2^i$;
4 **até que** $N < M$;
5 **responda** N.

O resultado das escolhas aleatórias na linha 2 do Algoritmo 1.2, digamos que seja a sequência $d_{k-1}d_{k-2}\ldots d_0$, é um evento 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 ao conjunto $\{0, 1, \ldots, 2^k - 1\}$.

O ALGORITMO TERMINA?

Isto é, eventualmente a condição $N < M$ na linha 4 é satisfeita?

Fixamos uma instância M com

$$2^{k-1} < M \leqslant 2^k \quad \text{e} \quad k = \lceil \log M \rceil.$$

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

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

pois $M > 2^{k-1}$.

Para todo $n \geqslant 1$,

A_n é o evento "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, pela independência da ocorrência dos eventos $\overline{A}$ em cada rodada do laço

$$\mathbb{P}(A_n) = \left(1 - \frac{M}{2^k}\right)^n$$

portanto,

$$\mathbb{P}(A_n) < \left(\frac{1}{2}\right)^n.$$

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 (1.15) na página 21),

$$\begin{aligned}\mathbb{P}[\,\text{o algoritmo não termina}\,] &= \mathbb{P}\left(\bigcap_{n\geqslant 1} A_n\right)\\ &= \lim_{n\to\infty} \mathbb{P}(\,A_n\,)\\ &= \lim_{n\to\infty} \left(1-\frac{M}{2^k}\right)^n\\ &= 0.\end{aligned} \tag{1.16}$$

Consequentemente, $\mathbb{P}[\,\text{o algoritmo termina}\,] = 1$. O algoritmo termina ***quase certamente***.

Se termina, quão rápido termina?

A aleatoriedade influencia na duração das rodadas, isto é, no número de rodadas do laço da linha 1 necessárias para que ele termine.

O algoritmo não apenas termina quase certamente, como, de fato, termina em poucas rodadas 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.

Se o algoritmo termina, responde certo?

Por exemplo, se $M = 7$ então $k = 3$. Com três bits $d_2d_1d_0$ 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 $d_2d_1d_0 = 111$. Dado que esse evento não ocorre, qual a probabilidade do algoritmo responder 4?

Seja N a resposta do algoritmo. Usando probabilidade condicional $\mathbb{P}[\,N = 4 \mid N \neq 7\,] = (1/8)\,/\,(7/8) = 1/7$ e, de fato, o algoritmo escolhe qualquer número em $\{\,0,\ldots,6\,\}$ com probabilidade $1/7$.

No caso geral, definimos o evento $A = \{\,0,1,\ldots,M-1\,\}$ e para qualquer $t \in A$ a probabilidade da resposta N do algoritmo ser t dado que $N \in A$ é

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

Portanto, se o algoritmo termina, ou seja, dado que o sorteio N satisfaça $N < M$, então ele responde com um número entre 0 e $M-1$ de modo uniforme.

Diferente do algoritmo para identidade polinomial, o Algoritmo 1.2 não responde errado. Em contrapartida, o tempo de execução depende dos sorteios.

Algoritmos Las Vegas O algoritmo gerador de números aleatórios sempre fornece a solução correta. A única variação de uma execução para outra é o seu tempo de execução. Chamamos esse tipo de algoritmo de ***algoritmo Las Vegas***. Um algoritmo Las Vegas é, por definição, um algoritmo Monte Carlo com probabilidade de erro igual a o cujo tempo de execução é uma variável aleatória (nosso próximo tema). Em geral, deseja-se que esses algoritmos tenham tempo de execução esperado pequeno e que a probabilidade de uma execução demorar muito seja pequena.

Dizemos que um algoritmo Las Vegas é um algoritmo ***Las Vegas eficiente*** se, para qualquer entrada, seu *tempo de execução esperado*[2] for delimitado por uma função polinomial do tamanho da entrada.

[2] A esperança da variável aleatória para o tempo

Dizemos que um algoritmo Monte Carlo é um algoritmo ***Monte Carlo eficiente*** se, para qualquer entrada, seu tempo de execução no pior caso for delimitado por uma função polinomial do tamanho da entrada.

Exercícios

Exercício 1. Há três urnas e em cada urna um par de bolas. Na primeira urna há um par de bolas brancas, na segunda um par de bolas pretas e na terceira um par com uma bola de cada cor. Uma urna é escolhida uniformemente e, sem olhar para o interior da urna, uma bola é escolhida uniformemente é retirada. A bola retirada é branca. Qual a probabilidade da bola que ficou sozinha na urna ser preta?

Exercício 2. Considere a seguinte proposta de algoritmo que recebe um inteiro positivo $M > 0$ e devolve um inteiro escolhido aleatoriamente em $\{0, 1, \ldots, M-1\}$:

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

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

Prove que as respostas não são equiprováveis quando M não é potência de 2.

Exercício 3. Agora, considere o seguinte gerador de números aleatórios adaptado do anterior.

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

1 seja k o número de bits de M;
2 $(d_0, d_1, \ldots, d_{k+t-1}) \xleftarrow{R} \{0,1\}^{k+t}$;
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 (como vimos acima). Prove que

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

Isso mostra que, à medida que t cresce, a distribuição de N mod M converge exponencialmente rápido para a distribuição uniforme em $\{0, \ldots, M-1\}$, o que justifica usar valores modestos de t (por exemplo $t = O(\log(1/\delta))$) para obter uma amostragem "quase uniforme" com erro estatístico δ arbitrariamente pequeno.

Exercício 4. Prove que no seguinte algoritmo a probabilidade p_n do laço executar pelo menos n vezes, para todo $n \geqslant 2$, é maior que 0 e, mais que isso, essa probabilidade é maior que 0 no limite, ou seja, $\lim p_n > 0$ quando $n \to \infty$ (pode ser útil a desigualdade $1 - x \geqslant \exp(-2x)$ para $x \in [0, 1/2]$).

1 $j \leftarrow 0$;
2 **repita**
3 | $j \leftarrow j + 1$;
4 | **para cada** $i \in \{1, 2, \ldots, j\}$ **faça** $d_i \xleftarrow{R} \{0, 1\}$;
5 **até que** $d_i = 1$ para todo i.

2
Variáveis aleatórias

As variáveis aleatórias são funções que descrevem grandezas observáveis em fenômenos aleatórios, como o tempo de execução de um algoritmo aleatorizado, o número de comparações realizadas por um procedimento de busca, ou o tamanho de uma estrutura de dados gerada aleatoriamente. No Algoritmo 1.2, o número de rodadas do laço é uma variável aleatória. Note que execuções diferentes do algoritmo, mesmo que com a entrada M fixa, a quantidade de rodadas do laço pode ser diferente.

Definição 2.1: Variável aleatória

Se $(\Omega, \mathcal{A}, \mathbb{P})$ é um espaço de probabilidade, então uma função

$$X \colon \Omega \to \mathbb{R}$$

é uma ***variável aleatória discreta*** se sua imagem, que denotamos por $X(\Omega)$, é enumerável.
Em particular, se Ω é enumerável, então qualquer função $X \colon \Omega \to \mathbb{R}$ é uma ***variável aleatória discreta***.

A definição mais geral para v.a. discreta é: existe um enumerável S tal que $\mathbb{P}[\, X \in S \,] = 1$ Note que isso permite $\omega \in \Omega$ tal que $X(\omega) \notin S$, desde que $\mathbb{P}(\, \omega \,) = 0$. Entretanto, neste texto, raramente será preciso dessa generalidade. Quase sempre trabalharemos com Ω enumerável.

Exemplo 2.1: variável aleatória indicadora

Para qualquer evento A de um espaço de probabilidade $(\Omega, \mathcal{A}, \mathbb{P})$, definimos a variável aleatória $\mathbb{1}_A \colon \Omega \to \{\, 0, 1 \,\}$ por

$$\mathbb{1}_A(\omega) \coloneqq \begin{cases} 1, & \text{se } \omega \in A \\ 0, & \text{caso contrário,} \end{cases}$$

e que chamamos de ***variável aleatória indicadora*** da ocorrência do evento A.

Proposição 2.1

Toda variável aleatória discreta pode ser escrita como uma combinação linear de variáveis indicadoras.

Demonstração. Seja X uma variável aleatória discreta e seja $\{x_1, x_2, \dots\} \subseteq \mathbb{R}$ o conjunto dos valores que X pode assumir. Para cada i, defina a variável aleatória indicadora

$$I_i = \mathbb{1}_{[X=x_i]}.$$

Como X assume, com probabilidade 1, um dos valores x_i, para todo ω pertencente a um evento de probabilidade 1 existe um único i tal que $X(\omega) = x_i$. Consequentemente,

$$\sum_i x_i I_i(\omega) = \sum_i x_i \mathbb{1}_{[X=x_i]}(\omega) = X(\omega).$$

Logo,

$$X = \sum_i x_i \mathbb{1}_{[X=x_i]}$$

No caso em que X assume apenas um número finito de valores, a soma é finita; no caso enumerável, trata-se de uma soma enumerável (série). □

Proposição 2.2

Se X é uma variável aleatória que assume valores nos inteiros positivos $\{1, 2, \dots\}$ então

$$X = \sum_{k=1}^{\infty} \mathbb{1}_{[X \geqslant k]}. \tag{2.1}$$

Demonstração. Para cada $\omega \in \Omega$, como X assume valores nos inteiros positivos, existe um único $m = X(\omega) \in \{1, 2, \dots\}$. Então

$$\mathbb{1}_{[X \geqslant k]}(\omega) = \begin{cases} 1, & k \leqslant m, \\ 0, & k > m. \end{cases}$$

Logo,

$$\sum_{k=1}^{\infty} \mathbb{1}_{[X \geqslant k]}(\omega) = \underbrace{1 + \cdots + 1}_{m \text{ termos}} = m = X(\omega).$$

Como vale para todo $\omega \in \Omega$, concluímos (2.1) □

Usaremos quase sempre as letras maiúsculas finais do alfabeto, como X, Y, Z, W, para denotar variáveis aleatórias.

Se X é uma v.a., $[X = a]$ denota o evento

$$[X = a] = \{\omega \in \Omega \colon X(\omega) = a\}$$

que ocorre com probabilidade

$$\mathbb{P}_X(a) := \mathbb{P}[X = a]$$

chamada ***lei***, ou ***distribuição*** de X.

Analogamente, $[X \leqslant a]$ denota o evento $\{\omega \in \Omega\colon X(\omega) \leqslant a\}$.

Para a v.a. indicadora, como no Exemplo 2.1 acima, temos

$$[\mathbb{1}_A \leqslant t] = \begin{cases} \varnothing, & \text{se } t < 0 \\ \overline{A}, & \text{se } 0 \leqslant t < 1 \\ \Omega, & \text{se } t \geqslant 1 \end{cases}$$

Ainda, a notação $[X \in A]$ é usada para o evento dado pela pré-imagem de A, i.e., $X^{-1}(A) = \{\omega \in \Omega\colon X(\omega) \in A\}$, para todo $A \subseteq X(\Omega)$.

Por exemplo, as pré-imagens $\mathbb{1}_A^{-1}(B)$ são $\varnothing$, A, $\overline{A}$ e Ω nos casos, respectivamente, $\{0,1\} \cap B = \varnothing$, $\{0,1\} \cap B = \{1\}$, $\{0,1\} \cap B = \{0\}$ e $\{0,1\} \cap B = \{0,1\}$ para todo $B \subseteq \mathbb{R}$.

De volta ao laço do Algoritmo 1.2, se denotarmos por X o número de rodadas até a condição $N < M$ ser satisfeita, então X é uma variável aleatória e, sem mais nenhuma informação, o melhor que podemos fazer para estimar (no sentido de distância quadrática, ou erro quadrático médio) o seu valor é tomar a sua média ponderada

$$\sum_{t \geqslant 1} t \cdot \mathbb{P}[X = t] = \sum_{t=1}^{\infty} t q^{t-1} p = p \sum_{t=1}^{\infty} t q^{t-1} = p \cdot \frac{1}{(1-q)^2} = p \cdot \frac{1}{p^2} = \frac{1}{p},$$

onde $p = M/2^k$ é a probabilidade de $N < M$ e $q = 1 - p$. Note que essa média é < 2.

2.0.1 *Independência*

As variáveis aleatórias X e Y definidas em Ω ***variáveis aleatórias independentes*** se para quaisquer eventos A e B

$$\mathbb{P}([X \in A] \cap [Y \in B]) = \mathbb{P}[X \in A] \cdot \mathbb{P}[Y \in B].$$

Considerando as leis $\mathbb{P}_{(X,Y)}$ do vetor aleatório (X, Y), $\mathbb{P}_X$ de X e $\mathbb{P}_Y$ de Y temos que independência como definido acima é equivalente a

1. $\mathbb{P}_{(X,Y)}(A \times B) = \mathbb{P}_X(A) \cdot \mathbb{P}_Y(B)$;
2. $[X = a]$ e $[Y = b]$ são independentes, para quaisquer $a \in X(\Omega)$ e $b \in Y(\Omega)$;
3. $\mathbb{P}_{(X,Y)}((a,b)) = \mathbb{P}_X(a) \cdot \mathbb{P}_Y(b)$, para quaisquer $a \in X(\Omega)$ e $b \in Y(\Omega)$;

4. $\mathbb{P}(\,[X \leqslant a] \cap [Y \leqslant b]\,) = \mathbb{P}[\,X \leqslant a\,] \cdot \mathbb{P}[\,Y \leqslant b\,]$ para quaisquer $a, b \in \mathbb{R}$.

Mais geralmente, as variáveis aleatórias $X_1, X_2, \ldots, X_n$ são ***independentes*** se para subconjuntos $A_1, \ldots, A_n \subseteq \mathbb{R}$

$$\mathbb{P}\left(\bigcap_{i=1}^{n} [X_i \in A_i]\right) = \prod_{i=1}^{n} \mathbb{P}_{\mathbf{X}}(A_1 \times \cdots \times A_n) = \prod_{i=1}^{n} \mathbb{P}_{X_i}(A_i) = \prod_{i=1}^{n} \mathbb{P}[\,X_i \in A_i\,].$$

Equivalentemente,

$$\mathbb{P}_{\mathbf{X}}(A_1 \times \cdots \times A_n) = \prod_{i=1}^{n} \mathbb{P}_{X_i}(A_i).$$

2.1 *Valor esperado de uma variável aleatória simples*

Uma variável aleatória X é chamada de ***simples*** se pode ser escrita como combinação linear finita de funções indicadoras, ou seja, existem valores $x_1, x_2, \ldots, x_k$ e eventos $A_1, A_2, \ldots, A_k$ tais que

$$X = \sum_{i=1}^{k} x_i \mathbb{1}_{A_i}.$$

Podem existir muitas maneiras de escrever uma variável aleatória simples como uma combinação linear de funções indicadoras, descreveremos a seguir uma forma canônica. Tomamos no domínio de X a partição finita $B_1, \ldots, B_m$ dada pela relação de equivalência $\omega_1 \equiv \omega_2$ se, e somente se, $X(\omega_1) = X(\omega_2)$. Assim, se $X(\omega) = x_i$ para todo $\omega \in B_i$, então $x_i \neq x_j$ sempre que $i \neq j$. Com isso, a representação canônica é

$$X = \sum_{i=1}^{m} x_i \mathbb{1}_{B_i}$$

O ***valor médio*** ou ***valor esperado*** ou, ainda, ***esperança*** da variável aleatória simples X é definida por

$$\mathbb{E}X \coloneqq \sum_{i=1}^{m} x_i \, \mathbb{P}(\,B_i\,) \tag{2.2}$$

que é média dos valores de $X(\Omega) = \{\, x_1, x_2, \ldots, x_m \,\}$ ponderada pela probabilidade de cada valor. Como B_i é o evento $[X = x_i]$ segue que

$$\mathbb{E}X = \sum_{i=1}^{m} x_i \, \mathbb{P}[\,X = x_i\,] = \sum_{i=1}^{m} x_i \, \mathbb{P}_X(\,x_i\,). \tag{2.3}$$

Exemplo 2.2: Esperança da indicadora

Se A é evento de um espaço de probabilidade então

$$\mathbb{E}[\,\mathbb{1}_A\,] = \mathbb{P}(\,A\,).$$

O valor esperado de variáveis aleatórias simples não depende de como escrevemos X como combinação linear de variáveis indicadoras. De fato, seja $\{\, A_k : 1 \leqslant k \leqslant n \,\}$ uma outra partição *qualquer* de Ω tal que

$$\sum_{k=1}^{n} c_k \mathbb{1}_{A_k} = \sum_{\ell=1}^{m} x_\ell \mathbb{1}_{B_\ell}.$$

Então, para todo ℓ

$$B_\ell = \bigcup_{k:\, x_\ell = c_k} A_k$$

logo

$$\sum_{\ell=1}^{m} x_\ell \mathbb{P}(\,B_\ell\,) = \sum_{\ell=1}^{m} x_\ell \sum_{k:\, x_\ell = c_k} \mathbb{P}(\,A_k\,) = \sum_{k=1}^{n} \sum_{\ell:\, c_k = x_\ell} c_k \mathbb{P}(\,A_k\,)$$

portanto

$$\mathbb{E}X = \sum_{k=1}^{n} c_k \mathbb{P}(\,A_k\,) = \sum_{\ell=1}^{m} x_\ell \mathbb{P}(\,B_\ell\,). \tag{2.4}$$

Exemplo 2.3: Média pode ser pouco representativa

Consideremos um jogo de azar no qual em cada aposta pagamos R$10,00 e, ou ganhamos R$1.000.010,00 com probabilidade $p \in (0,1)$ ou não ganhamos nada, ou seja, perdemos os R$10,00 que já foram pagos, com probabilidade $1 - p$. Se Y é o ganho líquido em uma aposta, então a esperança do ganho por aposta é $\mathbb{E}Y = 10^6 p - 10(1-p) = 1.000.010p - 10$.
Se $p = 1/100$ então a probabilidade de ganhar o valor alto é muito pequeno quando comparado com a probabilidade de perder 10 reais, mas o prêmio é tão grande que o ganho esperado por aposta ainda é positivo: $\mathbb{E}Y = 9.990{,}10$. A probabilidade de receber esse valor numa aposta é 0.

Teorema 2.1: propriedades do valor esperado

Seja X uma variável aleatória simples definida no espaço amostral Ω munido da medida de probabilidade $\mathbb{P}$.

1. Se $X(\omega) = c$ para todo $\omega \in \Omega$ então $\mathbb{E}X = c$.
2. ***Linearidade:*** se Y é uma variável aleatória simples e a e b

números reais então

$$\mathbb{E}[\,aX + bY\,] = a\mathbb{E}\,X + b\mathbb{E}\,Y\,. \tag{2.5}$$

3. ***Monotonicidade:*** se Y é uma variável aleatória simples e $X \leqslant Y$, isto é, $X(\omega) \leqslant Y(\omega)$ para todo $\omega \in \Omega$, então

$$\mathbb{E}\,X \leqslant \mathbb{E}\,Y\,. \tag{2.6}$$

4. Se $f\colon \mathbb{R} \to \mathbb{R}$ é uma função real, então

$$\mathbb{E}[\,f(X)\,] = \sum_{x \in X(\Omega)} f(x)\mathbb{P}_X(\,x\,).$$

5. Se X e Y são variáveis aleatórias simples e independentes então $\mathbb{E}[\,XY\,] = \mathbb{E}[\,X\,]\,\mathbb{E}[\,Y\,]$.

Demonstração. Exercício. □

Em geral, $\mathbb{E}[\,X \cdot Y\,] \neq \mathbb{E}X \cdot \mathbb{E}Y$. Por exemplo, se X é o resultado do lançamento de um dado equilibrado então, pelo item 5 do teorema acima, $\mathbb{E}[\,X \cdot X\,] = \sum_{n=1}^{6} n^2\,\mathbb{P}_X(\,n\,) = 91/6 \neq 49/4 = \mathbb{E}X \cdot \mathbb{E}X$.

Corolário 2.1

Se $X_1, \ldots, X_n$ são variáveis aleatórias simples e $a_1, \ldots, a_n$ números reais então

$$\mathbb{E}\left[\sum_{i=1}^{n} a_i X_i\right] = \sum_{i=1}^{n} a_i \mathbb{E}X_i.$$

Demonstração. Segue da equação (2.5) usando indução em n. □

Exemplo 2.4

Se $X \in_{\mathrm{R}} \{\,1,2,3,4,5,6\,\}$ é o resultado de um lançamento de um dado, então

$$\mathbb{E}X = 1\frac{1}{6} + 2\frac{1}{6} + 3\frac{1}{6} + 4\frac{1}{6} + 5\frac{1}{6} + 6\frac{1}{6} = \frac{7}{2}$$

que também é a esperança de qualquer variável aleatória $Y \in_{\mathrm{R}} S$ para qualquer conjunto S com $|S| = 6$. Agora, se Y é o resultado de outro lançamento de dado, qual é a esperança da soma $X + Y$ dos pontos nos dois lançamentos do dado?
Pela linearidade da esperança é 7. A distribuição do produto $X \cdot Y$ é mostrada na Tabela 2.1, donde podemos calcular o valor esperado para o produto usando o item 5 do teorema

acima, ou seja, $\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y] = 49/4$.

$X \cdot Y$	1	2	3	4	5	6	8	9	10	12	15	16	18	20	24	25	30	36
$\mathbb{P}_{X \cdot Y}$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{4}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{4}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Tabela 2.1: distribuição do produto de dois dados.

Exemplo 2.5: Esperança da distribuição de Bernoulli

Em muitos experimentos aleatórios estamos interessados apenas em duas possibilidades para o resultado: *sucesso* ou *fracasso*. Esses experimentos recebem o nome de ***ensaio de Bernoulli***. Modelamos um ensaio de Bernoulli por uma variável aleatória indicadora $X\colon \Omega \to \mathbb{R}$ que assume apenas os valores 0 e 1, onde $X = 1$ representa *sucesso* e $X = 0$ representa *fracasso*.
Se, para $p \in [0,1]$, temos

$$\mathbb{P}_X(1) = p$$

e, portanto, $\mathbb{P}_X(0) = 1 - p$, então dizemos que X tem ***distribuição de Bernoulli*** com parâmetro p.
Usamos $X \sim \text{Bernoulli}(p)$ para indicar que X tem distribuição de Bernoulli com parâmetro p. Claramente,

$$\mathbb{E}X = \mathbb{P}[X = 1] = p.$$

Exemplo 2.6: Esperança da distribuição binomial

Em n ensaios de Bernoulli idênticos e independentes, uma resposta específica $(b_1, b_2, \ldots, b_n)$ com exatamente $t \in \{0, \ldots, n\}$ sucessos e $n - t$ fracassos ocorre com probabilidade $p^t(1-p)^{n-t}$.
A quantidade de sequências de respostas distintas com exatamente t sucessos é $\binom{n}{t}$.
Se X é a quantidade de sucessos em n ensaios de Bernoulli, então a probabilidade de ocorrerem exatamente t sucessos, independente da ordem, é

$$\mathbb{P}_X(t) := \binom{n}{t} p^t (1-p)^{n-t}, \quad t \in \{0, 1, \ldots, n\}$$

chamada ***distribuição binomial*** com parâmetros n e p. Nesse caso, escrevemos $X \sim \text{Binomial}(n, p)$.
Se $X_1, \ldots, X_n \sim \text{Bernoulli}(p)$ são as variáveis aleatórias indica-

doras de sucesso, então

$$X = \sum_{i=1}^{n} X_i \sim \text{Binomial}(n, p)$$

e, pela linearidade da esperança,

$$\mathbb{E}X = \mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbb{E}[\,X_i\,] = np.$$

2.2 *Tabelas de espalhamento*

Uma ***tabela de espalhamento***, ou ***tabela hashing***, é uma solução bastante conhecida e estudada para a representação computacional de um conjunto, seja ele estático ou dinâmico. Tabelas de espalhamento podem ser implementadas da seguinte maneira: uma tabela de espalhamento é um vetor H cujas entradas são listas ligadas, indexadas por $N = \{0, 1, \ldots, n-1\}$, e o acesso à tabela é feito por meio de uma função de *hash* $h : U \to N$. Dado $x \in U$, a inserção, remoção ou busca de x em H é realizada na lista ligada $H[h(x)]$ (veja uma ilustração na Figura 2.1).

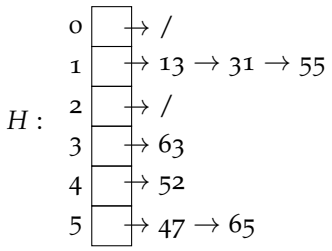


Figura 2.1: tabela de espalhamento com os elementos $\{13, 31, 47, 52, 55, 63, 65\}$ distribuídos de acordo com a função $h(x) = x \bmod 6$.

As operações de busca, inserção e remoção de um elemento numa tabela H que representa um conjunto dinâmico $S \subset U$ com função de ***hash*** $h\colon U \to N$ são descritas a seguir e são chamadas de ***operações de dicionário***:

busca: dado $x \in U$, uma operação de busca por x em S responde à pergunta "$x \in S$?" e ainda, no caso positivo, retorna um apontador para x. Numa tabela de espalhamento isso é realizado por uma busca sequencial na lista ligada $H[h(x)]$. Uma busca por x em H é dita *com sucesso* caso $x \in H$, senão é dita *sem sucesso*. O custo (ou tempo) de pior caso de uma busca é proporcional ao número de elementos na maior lista;

inserção: dado $x \in U$, uma operação de inserção acrescenta na estrutura que representa S o elemento x, caso esse ainda não faça

parte de S. A inserção propriamente dita numa tabela de espalhamento é simplesmente inserir o elemento x no início da lista ligada $H[h(x)]$ e o custo dessa operação é constante;

remoção: dado $x \in U$, a remoção de x retira esse elemento da estrutura que representa S, caso $x \in S$. A remoção propriamente dita numa tabela de espalhamento, dada a posição de x na lista ligada através de um apontador, é a remoção do elemento de uma lista ligada e o custo de pior caso dessa operação também é constante.

Com essas descrições, é natural constatar que, para uma análise do desempenho das operações de dicionário nessa estrutura de dados, é relevante o tempo de busca de um item que, no pior caso, é o tempo de busca na lista $H[i]$ com o maior número de elementos. Ademais, a pior configuração que podemos ter para o desempenho desses algoritmos ocorre quando todos os elementos de S são mapeados para a mesma lista ligada.

No emprego de tabelas de espalhamento, usualmente, temos a seguinte situação: o ***conjunto universo*** U é muito grande, o que significa que o custo de uma representação de S usando, por exemplo, um vetor de bits de tamanho $|U|$ é proibitivo. Se fosse viável, teríamos uma estrutura com tempo de busca $O(1)$. O conjunto S, de cardinalidade $m := |S|$, é uma fração pequena de U e o caso interessante é quando $n = O(m)$.

É desejável que os elementos de S fiquem bem espalhados na tabela, pois ***colisões*** (elementos distintos mapeados para a mesma lista) afetam o desempenho das operações. Como U tem muito mais que n elementos não há como evitar colisões.

Das funções de *hashing*, além do bom espalhamento, queremos que a função seja fácil de computar e que a representação seja sucinta, isto é, relativamente poucos bits sejam necessários para armazenar a função. Esse tipo de função, além de uma grande ferramenta prática como descrevemos abaixo, é bastante útil em Complexidade Computacional e em Criptografia e voltaremos a estudá-las adiante neste texto.

Hashing com funções aleatórias Uma função h escolhida uniformemente no conjunto N^U de todas as $|N|^{|U|}$ funções de U em N produz, quando $n = \Theta(m)$, listas de tamanho no máximo $O(\log m)$ com alta probabilidade. Isso garante, com alta probabilidade, um tempo de busca equivalente ao das árvores de buscas, com a vantagem de manutenção mais simples. Além disso, o tempo médio de busca é $\approx m/n = O(1)$, quando $n = O(m)$.

Nesse modelo probabilístico, a escolha de uma função aleatória equivale a escolher, independentemente para cada elemento $x \in U$,

uma posição $h(x)$ uniformemente em N.

Um problema nessa estratégia é que tal função não tem uma representação sucinta. Algumas delas requerem pelo menos $|U| \log(n)$ bits para serem representadas. Por ora analisaremos o caso de uma função aleatória para um conjunto S fixo, porém arbitrário. Nesse caso o tamanho de uma lista é uma variável aleatória.

Vamos considerar o modelo probabilístico $(N^U, \mathbb{P})$ com $\mathbb{P}(h) = 1/|N^U|$ para toda função $h \in N^U$. Notemos que vale

$$\mathbb{P}[\, h(x) = i \,] = \frac{1}{|N|} \tag{2.7}$$

para todo $x \in U$ e todo $i \in N$. De fato, sortear uniformemente uma função é equivalente a sortear uniformemente e independentemente uma imagem para cada elemento do domínio.

Fixamos os parâmetros $m = |S|$ e $n = |N|$ e então a fração

$$c = c(n, m) \coloneqq \frac{m}{n} \tag{2.8}$$

é denominada ***carga*** da tabela. A carga da tabela é o valor esperado para a quantidade de elementos de S numa mesma posição da tabela, segundo a medida da equação (2.7). Tomemos a variável aleatória

$$\mathbb{1}_{[h(x)=h(y)]} \colon N^U \to \{\, 0, 1 \,\}$$

indicadora de colisão dos elementos $x, y \in U$ sob a escolha $h \in N^U$. A probabilidade de colisão é

$$\begin{aligned} \mathbb{P}[\, h(x) = h(y) \,] &= \sum_{i \in N} \mathbb{P}[\, h(x) = i \,] \mathbb{P}[\, h(y) = i \,] \qquad (x \neq y) \\ &= n \cdot \frac{1}{n} \cdot \frac{1}{n} = \frac{1}{n}. \end{aligned} \tag{2.9}$$

assim, o número de itens na lista $H[h(x)]$ é dado por $\sum_{y \in S} \mathbb{1}_{[h(x)=h(y)]}$ e

$$\mathbb{E}\mathbb{1}_{[h(x)=h(y)]} = \begin{cases} \frac{1}{n}, & \text{se } y \neq x \\ 1, & \text{caso contrário,} \end{cases}$$

portanto o tamanho esperado da lista $H[h(x)]$ é, pela linearidade da esperança

$$\sum_{y \in S} \mathbb{E}\mathbb{1}_{[h(x)=h(y)]} = \begin{cases} \frac{m}{n}, & \text{se } x \notin S, \\ \frac{(m-1)}{n} + 1, & \text{caso contrário.} \end{cases} \tag{2.10}$$

O tempo esperado de uma busca em uma tabela de espalhamento com uma função *hash* h escolhida aleatoriamente em N^U é proporcional à carga c da tabela e, portanto, é constante quando $n = \Theta(m)$.

Proposição 2.3

Sejam $S \subseteq U$, $|S| = m$, e H uma tabela de espalhamento com n posições, usando uma função de *hash* $h\colon U \to N$ escolhida aleatoriamente. Suponha que os elementos de S são inseridos no início da lista ligada correspondente a cada célula de H. Então:

1. o tempo esperado de uma busca *sem sucesso* é $c + 1$,
2. o tempo esperado de uma busca *com sucesso* é $1 + \frac{c}{2} - \frac{1}{2n}$,

onde $c = m/n$ é a carga média da tabela.

Demonstração. Se $x \notin S$, a busca percorre todos os elementos na lista $H[h(x)]$. Pelo modelo de função aleatória (equação (2.10)), o tamanho esperado dessa lista é $m/n = c$, assim chegamos ao final da lista com $c + 1$ comparações em média.

Suponha que $x \in S$ e que foi o i-ésimo elemento inserido. Como inserimos sempre no início da lista, x estará na frente da lista no momento da inserção. Dos $m - i$ elementos restantes que ainda serão inseridos, em média $(m-i)/n$ colidirão com x, de modo que o tempo esperado para achar x nessa lista será $(m-i)/n + 1$ comparações. A média sobre todos os possíveis valores de i é

$$\frac{1}{m}\sum_{i=1}^{m}\left(1 + \frac{m-i}{n}\right) = 1 + \frac{1}{mn}\sum_{i=1}^{m}(m-i) = 1 + \frac{m-1}{2n}.$$

Assim, o tempo médio de busca com sucesso é $1 + c/2 - 1/(2n)$. □

Até aqui determinamos o tamanho esperado de uma lista, mas essa informação não permite controlar diretamente a probabilidade de uma lista ser excepcionalmente grande. Para isso, precisamos estudar a variância da variável aleatória que representa o tamanho de uma lista e utilizar desigualdades de concentração.

Andrey Markov (1856–1922) foi um matemático russo conhecido principalmente por suas contribuições no que foi posteriormente conhecido como Cadeias de Markov. Seu filho Andrey Markov Jr (1903–1979) ficou conhecido pelas contribuições em Teoria da Computação.

Teorema 2.2: Desigualdade de Markov

Se X é uma variável aleatória integrável que assume valores não negativos,

$$\mathbb{P}[\,X \geqslant t\,] \leqslant \frac{\mathbb{E}X}{t} \tag{2.11}$$

para todo $t > 0$.

Demonstração. De X integrável, a desigualdade segue de $X \geqslant 0$ e $X \geqslant X\mathbb{1}_{[X \geqslant t]}$ por monotonicidade da esperança

$$\mathbb{E}[\,X\,] \geqslant \mathbb{E}[\,X\mathbb{1}_{[X\geqslant t]}\,] \geqslant \mathbb{E}[\,t\mathbb{1}_{[X\geqslant t]}\,] \geqslant t \cdot \mathbb{E}[\,\mathbb{1}_{[X\geqslant t]}\,].$$

Basta notar que $\mathbb{E}[\,\mathbb{1}_{[X \geqslant t]}\,] = \mathbb{P}[\,X \geqslant t\,]$. □

Para S fixo o número esperado de colisões C é

$$\mathbb{E}C = \binom{m}{2}\frac{1}{n} = \frac{m(m-1)}{2n} \tag{2.12}$$

assim, usando (2.11)

$$\mathbb{P}[\,C = 0\,] \geqslant 1 - \mathbb{E}[\,=\mid 1\,] - \frac{m(m-1)}{2n}$$

e se $n = m^2$ então $\mathbb{E}C < 1/2$ e $\mathbb{P}[\,C = 0\,] \geqslant 1/2$.

A próxima desigualdade é conhecida por desigualdade de Chebyshev, também conhecida por desigualdade de Bienaymé–Chebyshev. Da equação (2.11) obtemos a ***desigualdade de Chebyshev***

$$\mathbb{P}\Big[\,X^2 \geqslant t^2\,\Big] \leqslant \frac{\mathbb{E}[\,X^2\,]}{t^2} \tag{2.13}$$

que é mais conhecida na forma enunciada a seguir.

Variância

Definimos a ***variância*** da variável aleatória X por

$$\mathrm{Var}(\,X\,) := \mathbb{E}\Big[\,(X - \mathbb{E}X)^2\,\Big] = \mathbb{E}[\,X^2\,] - \mathbb{E}[\,X\,]^2 \tag{2.14}$$

sempre que X é quadrado-integrável, isto é, X^2 é integrável.

Proposição 2.4: var da soma de v.a. independentes

Se $X_1, \ldots, X_n$ são variáveis aleatórias integráveis e 2-a-2 independentes então

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(\,X_i\,). \tag{2.15}$$

Demonstração. A afirmação segue da expansão dos termos em (2.14)

$$\begin{aligned}
\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) &= \mathbb{E}\left[\left(\sum_{i=1}^{n} X_i\right)^2\right] - \mathbb{E}\left[\sum_{i=1}^{n} X_i\right]^2 \\
&= \mathbb{E}\left[\left(\sum_{i=1}^{n}\sum_{j=1}^{n} X_i X_j\right)\right] - \left(\sum_{i=1}^{n} \mathbb{E}X_i\right)^2 \\
&= \left(\sum_{i=1}^{n}\sum_{j=1}^{n} \mathbb{E}[\,X_i X_j\,]\right) - \left(\sum_{i=1}^{n}\sum_{j=1}^{n} \mathbb{E}[\,X_i\,]\mathbb{E}[\,X_j\,]\right) \\
&= \sum_{i=1}^{n} \mathbb{E}\Big[\,X_i^2\,\Big] - \sum_{i=1} \mathbb{E}[\,X_i\,]^2
\end{aligned}$$

pois, da independência das variáveis aleatórias, $\mathbb{E}[\,X_i X_j\,] = \mathbb{E}[\,X_i\,]\mathbb{E}[\,X_j\,]$ e termos se cancelam. □

Pafnuty Lvovich Chebyshev (1821–1894) foi um matemático russo conhecido por suas contribuição à Probabilidade, dentre seus alunos vários são matemáticos reconhecidos e um deles foi Andrey Markov.

Teorema 2.3: desigualdade de Chebyshev

Para toda constante $t > 0$ e toda variável aleatória X com quadrado integrável temos

$$\mathbb{P}[\,|X - \mathbb{E}X| \geqslant t\,] \leqslant \frac{\mathrm{Var}(X)}{t^2}. \tag{2.16}$$

Demonstração. Segue da desigualdade (2.13), do fato de $t > 0$ vale $[X^2 \geqslant t^2] = [|X| \geqslant t]$ e usamos a variável aleatória $X - \mathbb{E}X$ no lugar de X. □

Seja X_i o número de elementos de S que caem na posição i

$$X_i = \sum_{x \in S} \mathbb{1}_{[h(x)=i]}.$$

Sabemos que $\mathbb{E}[X_i] = \frac{m}{n} = c$. Como as variáveis indicadoras são independentes,

$$\mathrm{Var}(X_i) = \mathrm{Var}\left(\sum_{x \in S} \mathbb{1}_{[h(x)=i]}\right) = \sum_{x \in S} \mathrm{Var}\left(\mathbb{1}_{[h(x)=i]}\right)$$

mas $\mathrm{Var}(\mathbb{1}_{[h(x)=i]}) = \frac{1}{n}\left(1 - \frac{1}{n}\right)$, assim temos

$$\mathrm{Var}(X_i) = m\frac{1}{n}\left(1 - \frac{1}{n}\right) = c\left(1 - \frac{1}{n}\right) \leq c.$$

Assim, a desigualdade de Chebyshev dá

$$\mathbb{P}[\,|X_i - c| \geqslant t\,] \leqslant \frac{c(1 - 1/n)}{t^2},$$

e, por exemplo, tomando $t = c$,

$$\mathbb{P}[\,X_i \geqslant 2c\,] \leqslant \mathbb{P}[\,|X_i - c| \geqslant c\,] \leqslant \frac{1}{c} - \frac{1}{cn}.$$

Exemplo 2.7: O paradoxo dos aniversários

Para $m \leqslant n$ há $n(n-1)(n-2)\cdots(n-m+1)$ sequências sem repetições em N^m. Se $(h(x_1), h(x_2), \ldots, h(x_m))$ é uma escolha aleatória em N^m e C é o número de colisões, então nessa

escolha

$$\begin{aligned}\mathbb{P}[\,C=0\,] &= \frac{n(n-1)(n-2)\cdots(n-m+1)}{n^m}\\ &= \left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{m-2}{n}\right)\left(1-\frac{m-1}{n}\right)\\ &< \exp\left(-\frac{1}{n}\right)\exp\left(-\frac{2}{n}\right)\cdots\exp\left(-\frac{m-1}{n}\right)\\ &= \exp\left(-\frac{m(m-1)}{2n}\right)\end{aligned}$$

na segunda linha usamos que $\exp(-x) > 1-x$ (veja (d.1)).
O ***paradoxo dos aniversários*** é o caso $m = 23$ e $n = 365$ na equação acima: $\mathbb{P}[\,C > 0\,] > 1-\mathbb{P}[\,C = 0\,] > 0{,}5$, ou seja, apenas 23 pessoas são suficientes para que duas delas façam aniversário no mesmo dia com probabilidade maior que $1/2$, supondo que os nascimentos ocorram uniformemente ao longo do ano.
Para todo real $x \in [0, 3/4]$ vale que $1-x > \exp(-2x)$, portanto para $m \leqslant (3/4)n$ temos o seguinte limitante para a probabilidade de não ocorrer colisão

$$\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{m-1}{n}\right) > \exp\left(-\frac{m(m-1)}{n}\right)$$

de modo que se $m \approx \sqrt{n}$ então a probabilidade de não haver colisão é maior que $\approx \exp(-1) \approx 0{,}36$, de modo que, para $m \approx \sqrt{n}$, a probabilidade de inserir m elementos sem colisão permanece limitada inferiormente por uma constante positiva, próxima de e^{-1}.

2.3 *Esperança matemática de variáveis discretas*

Estendendo de modo natural a definição de valor esperado de uma variável simples, dada na equação (2.3), para uma variável aleatória discreta X que assume valores em $\{\,x_1, x_2, \ldots\,\} \subseteq \mathbb{R}$ com probabilidade 1, podemos escrever

$$X = \sum_{n=1}^{\infty} x_n \mathbb{1}_{[X=x_n]}, \tag{2.17}$$

e obter, formalmente,

$$\mathbb{E}X = \sum_{n=1}^{\infty} x_n\, \mathbb{P}[\,X = x_n\,]. \tag{2.18}$$

Para que essa definição faça sentido, é necessário que o valor da soma não dependa da ordem em que os termos são somados. Como

veremos a seguir, isso é garantido quando a série é absolutamente convergente.

Exemplo 2.8: Rearranjos podem convergir para somas distintas

Por volta de 1833, Cauchy observou que rearranjos da série harmônica alternada $\sum_{i\geqslant 1}(-1)^{i+1}/i$ resultam em limites diferentes. Por exemplo

$$1-\frac{1}{2}+\frac{1}{3}-\cdots+\frac{(-1)^{n+1}}{n}+\cdots=\ln(2)$$

enquanto que, no rearranjo atribuído a Laurent

$$1-\frac{1}{2}-\frac{1}{4}+\frac{1}{3}-\frac{1}{6}-\frac{1}{8}+\cdots+\frac{1}{2k-1}-\frac{1}{2(2k-1)}-\frac{1}{4k}+\cdots=\frac{\ln(2)}{2}.$$

Se a variável aleatória X assume os valores $x_i := (-1)^{i+1}/i$, para $i=1,2,\ldots$, com probabilidade $\mathbb{P}_X(x_i)=6/(\pi i)^2$, então

$$\sum_{i=1}^{\infty} x_{\sigma(i)}\mathbb{P}\left[X=x_{\sigma(i)}\right]=\begin{cases}\frac{6}{\pi^2}\ln(2) & \text{se } \sigma \text{ é a permutação identidade}\\ \frac{3}{\pi^2}\ln(2) & \text{se } \sigma \text{ é a permutação de Laurent}\end{cases}, \tag{2.19}$$

portanto, X não tem um valor médio bem definido.
Um teorema atribuído a Riemann (veja em Rudin, 1976, teorema 3.54) estabelece que, se $\sum_n a_n$ converge e $\sum_{i\geqslant 1}|a_i| = \infty$, então, para qualquer $S \in \mathbb{R}\cup\{-\infty,+\infty\}$, existe uma permutação σ dos inteiros positivos tal que $\sum_{i\geqslant 1} a_{\sigma(i)} = S$.

Definição 2.2: Valor esperado

Seja X uma variável aleatória definida sobre um espaço discreto. Seu ***valor esperado***, ou ***valor médio***, ou ***esperança***, é o número

$$\mathbb{E}X=\sum_{\omega\in\Omega}X(\omega)\,\mathbb{P}(\omega) \tag{2.20}$$

desde de que a série convirja absolutamente.
Dizemos que X é ***integrável*** se $\sum_{\omega\in\Omega}|\,X(\omega)\,|\,\mathbb{P}(\omega)<\infty$.
Dizemos que $\mathbb{E}X$ está ***definida*** se X for integrável ou $\mathbb{E}X\in\{\infty,-\infty\}$.

Seja $X : \Omega \to \mathbb{R}$ uma variável aleatória discreta e integrável. Então,

$$\begin{aligned}
\mathbb{E}X &= \sum_{\omega\in\Omega} X(\omega)\,\mathbb{P}(\omega) \\
&= \sum_{\omega\in\Omega}\left(\sum_{t\in\mathrm{Im}(X)} t\,\mathbb{1}_{[X(\omega)=t]}\right)\mathbb{P}(\omega) \\
&= \sum_{t\in\mathrm{Im}(X)} t\left(\sum_{\omega\in\Omega}\mathbb{1}_{[X(\omega)=t]}\,\mathbb{P}(\omega)\right) \\
&= \sum_{t\in\mathrm{Im}(X)} t \sum_{\omega:\,X(\omega)=t}\mathbb{P}(\omega) \\
&= \sum_{t\in\mathrm{Im}(X)} t\,\mathbb{P}[\,X=t\,].
\end{aligned}$$

A passagem da segunda para a terceira linha usa troca da ordem de somação, justificada pela convergência absoluta. Portanto,

Fato 2. Se a variável aleatória discreta X é integrável

$$\mathbb{E}X = \sum_{t\in\mathrm{Im}(X)} t\,\mathbb{P}_X(t), \tag{2.21}$$

ou, equivalentemente, a série em (2.18) é a esperança de X.

Ideia importante

A esperança de uma variável aleatória depende apenas da sua distribuição, ou seja, se X e Y têm a mesma distribuição, então $\mathbb{E}Y = \mathbb{E}X$.
Em particular, a esperança pode ser vista como uma propriedade da distribuição de X, não do espaço amostral subjacente.

Esperança de funções de variáveis aleatórias Sejam $X\colon \Omega \to S$ uma variável aleatória em $(\Omega, \mathbb{P})$ com valores em $S \subset \mathbb{R}$ enumerável e $f\colon S \to \mathbb{R}$ uma função. No espaço de probabilidade discreto $(S, \mathbb{P}_X)$, a função f é uma variável aleatória cuja lei é a medida imagem de $\mathbb{P}_X$ por f, isto é, $\mathbb{P}_f(\,B\,) = \mathbb{P}_X(f^{-1}(B))$ para $B \subset \mathbb{R}$. Assim, quando bem definida,

$$\mathbb{E}_S\, f \;= \sum_y y\,\mathbb{P}_f(\,y\,) = \sum_y y\,\mathbb{P}[\,f(X)=y\,] = \mathbb{E}_\Omega\, f(X)$$

donde tiramos que f é integrável no espaço de probabilidade $(S, \mathbb{P}_X)$ se, e somente se, $f(X)$ é integrável em $(\Omega, \mathbb{P})$.

Assim, tomando $f(x) = |x|$, concluímos que

$$X \text{ é integrável se, e somente se, } \mathbb{E}|X| < +\infty.$$

Se $f(X)$ é integrável, então podemos rearranjar a série que define $\mathbb{E}f(X)$ de modo que

$$\mathbb{E}f(X) = \sum_y y\,\mathbb{P}_{f(X)}(y) = \sum_y y\,\mathbb{P}_X(\{s\colon f(s) = y\}) = \sum_y \sum_{\substack{s\in S \\ f(s)=y}} y\,\mathbb{P}_X(s) = \sum_{s\in S} f(s)\mathbb{P}_X(s) \quad (2.22)$$

e assim provamos o seguinte resultado.

O termo "Lei do Estatístico Inconsciente" foi cunhado por Sheldon Ross em seu famoso livro "A First Course in Probability" (1976). A piada interna da comunidade acadêmica é que mudar o parâmetro de x para $f(x)$ dentro do somatório ou da integral parece tão óbvio e natural que as pessoas fazem isso mecanicamente, como se fosse uma mera definição, quando na verdade trata-se de um teorema que exige demonstração.

Teorema 2.4: Lei do Estatístico Inconsciente

Se $X\colon \Omega \to S$ é uma variável aleatória e $f\colon S \to \mathbb{R}$ uma função, então $f(X)\colon \Omega \to \mathbb{R}$ é variável aleatória. Além disso, se $f(X)$ é integrável, então

$$\mathbb{E}f(X) = \sum_{s\in S} f(s)\,\mathbb{P}_X(s).$$

2.4 *Propriedades da esperança*

A seguir veremos propriedades importantes para a média das variáveis aleatórias discretas a valores reais.

Teorema 2.5: linearidade da esperança

Se X e Y são variáveis aleatórias integráveis em $(\Omega, \mathbb{P})$ e $a, b \in \mathbb{R}$, então $aX + bY$ é integrável e sua esperança satisfaz

$$\mathbb{E}[aX + bY] = a\,\mathbb{E}X + b\,\mathbb{E}Y.$$

Demonstração. Sejam X e Y variáveis aleatórias integráveis e $a, b \in \mathbb{R}$. Pela equação (2.20) acima a esperança da variável aleatória $|aX + bY|$ é

$$\begin{aligned}\mathbb{E}|aX + bY| &= \sum_\omega |aX(\omega) + bY(\omega)|\,\mathbb{P}(\omega) \\ &\leqslant \sum_\omega \big(|aX(\omega)| + |bY(\omega)|\big)\,\mathbb{P}(\omega) \\ &= \sum_\omega |aX(\omega)|\,\mathbb{P}(\omega) + \sum_\omega |bY(\omega)|\,\mathbb{P}(\omega).\end{aligned}$$

Pela desigualdade triangular e como X e Y são integráveis, o lado

direito é finito. Logo, $aX + bY$ é integrável. Além disso,

$$\begin{aligned}\mathbb{E}aX + bY &= \sum_{\omega}(aX + bY)(\omega)\,\mathbb{P}(\,\omega\,)\\ &= \sum_{\omega}\big(aX(\omega) + bY(\omega)\big)\,\mathbb{P}(\,\omega\,)\\ &= \sum_{\omega}aX(\omega)\,\mathbb{P}(\,\omega\,) + \sum_{\omega}bY(\omega)\,\mathbb{P}(\,\omega\,)\\ &= a\,\mathbb{E}X + b\,\mathbb{E}Y.\end{aligned}$$

□

Corolário 2.2: caso geral da linearidade

Se $X_1, \ldots, X_n$ são integráveis e $a_1, \ldots, a_n \in \mathbb{R}$ então

$$\mathbb{E}\left[\sum_{i=1}^{n} a_i X_i\right] = \sum_{i=1}^{n} a_i\,\mathbb{E}[\,X_i\,].$$

Teorema 2.6: monotonicidade da esperança

Se X e Y são integráveis tais que $X \leqslant Y$ então $\mathbb{E}X \leqslant \mathbb{E}Y$.

Demonstração. Se $X \leqslant Y$ então $Y - X \geqslant 0$, logo $\mathbb{E}[\,Y - X\,] \geqslant 0$. Da linearidade da esperança $\mathbb{E}Y - \mathbb{E}X \geqslant 0$ donde segue o teorema. □

Corolário 2.3: consequências da linearidade e monotonicidade

As seguintes propriedades decorrem dos teoremas acima.

1. Se $X = c$ então $\mathbb{E}X = c$, para qualquer $c \in \mathbb{R}$.
2. Se $\mathbb{E}X$ está definida, então $\mathbb{E}[\,aX + b\,] = a\mathbb{E}X + b$.
3. Se $\mathbb{P}[\,a \leqslant X \leqslant b\,] = 1$, então $a \leqslant \mathbb{E}X \leqslant b$.
4. Se X é integrável então $X \cdot \mathbb{1}_A$ é integrável para todo evento A.
5. Se $\mathbb{E}X$ está definida então $|\mathbb{E}X| \leqslant \mathbb{E}|X|$.

Demonstração. Vamos demonstrar apenas os dois últimos itens, os outros ficam para verificação do leitor. Notemos que $X\,\mathbb{1}_A \leqslant X$ e que $|X\mathbb{1}_A| = |X|\mathbb{1}_A$, logo, se X é integrável, então, por monotonicidade, $X\mathbb{1}_A$ é integrável.

Se X não é integrável então $|\mathbb{E}X| = +\infty = \mathbb{E}|X|$ e o item 5 vale com igualdade. Senão, X é integrável e de $-|x| \leqslant x \leqslant |x|$, para todo real x, temos $-|X| \leqslant X \leqslant |X|$ donde $-\mathbb{E}|X| \leqslant \mathbb{E}X \leqslant \mathbb{E}|X|$, ou seja, $|\mathbb{E}X| \leqslant \mathbb{E}|X|$. □

Proposição 2.5

Se X assume valores inteiros e não negativos então $\mathbb{E}X$ está bem definida e

$$\mathbb{E}X = \sum_{n=0}^{\infty} \mathbb{P}[\, X > n \,].$$

Demonstração. Se X assume valores não negativos, segue da definição que $\mathbb{E}X$ está definida. Rearranjando os termos, a esperança é

$$\begin{aligned}
\sum_{t=0}^{\infty} t\, \mathbb{P}_X(\,t\,) = &1\,\mathbb{P}_X(\,1\,) + 2\,\mathbb{P}_X(\,2\,) + 3\,\mathbb{P}_X(\,3\,) + 4\,\mathbb{P}_X(\,4\,) + \cdots \\
&+ 2\,\mathbb{P}_X(\,2\,) + 3\,\mathbb{P}_X(\,3\,) + 4\,\mathbb{P}_X(\,4\,) + \cdots \\
&+ 3\,\mathbb{P}_X(\,3\,) + 4\,\mathbb{P}_X(\,4\,) + \cdots \\
&+ 4\,\mathbb{P}_X(\,4\,) + \cdots \\
&\vdots \\
= &\sum_{n=0}^{\infty} \sum_{t=n+1}^{\infty} \mathbb{P}_X(\,t\,)
\end{aligned}$$

como $\sum_{t>n} \mathbb{P}_X(t) = \mathbb{P}[\, X > n \,]$, segue a proposição. □

Proposição 2.6: condição de integrabilidade

Uma variável aleatória X é integrável se, e somente se,

$$\sum_{n=1}^{\infty} \mathbb{P}[\, |X| \geqslant n \,] < +\infty.$$

Demonstração. Tomemos $Y = |X|$ e temos que X é integrável se, e somente se, $\mathbb{E}Y < +\infty$. De $0 \leqslant \lfloor Y \rfloor \leqslant Y \leqslant \lfloor Y \rfloor + 1$ segue que $\mathbb{E}[\, \lfloor Y \rfloor \,] \leqslant \mathbb{E}Y \leqslant \mathbb{E}[\, \lfloor Y \rfloor \,] + 1$ por monotonicidade. Como $\lfloor Y \rfloor$ assume valores inteiros e não negativos, da Proposição 2.5 acima

$$\sum_{n \geqslant 1} \mathbb{P}[\, \lfloor Y \rfloor \geqslant n \,] \leqslant \mathbb{E}Y \leqslant \sum_{n \geqslant 1} \mathbb{P}[\, \lfloor Y \rfloor \geqslant n \,] + 1. \tag{2.23}$$

Se $Y \geqslant n$ então $\lfloor Y \rfloor > Y - 1 \geqslant n - 1$, logo $\lfloor Y \rfloor \geqslant n$. Por outro lado, se $\lfloor Y \rfloor \geqslant n$ então $Y \geqslant \lfloor Y \rfloor \geqslant n$, logo $Y \geqslant n$. Portanto, $\mathbb{P}[\, \lfloor Y \rfloor \geqslant n \,] = \mathbb{P}[\, Y \geqslant n \,]$ e da equação (2.23), $\mathbb{E}Y < +\infty$ se e só se $\sum_{n \geqslant 1} \mathbb{P}[\, |X| \geqslant n \,] < +\infty$. □

Se $X \colon \Omega \to S$ e $Y \colon \Omega \to R$ são variáveis aleatórias independentes, então

$$\mathbb{E}[\, |XY| \,] = \sum_{(x,y) \in S \times R} |xy| \mathop{\mathbb{P}}_{(X,Y)}(\,(x,y)\,) = \sum_{(x,y) \in S \times R} |x|\,|y| \mathop{\mathbb{P}}_{X}(\,x\,) \mathop{\mathbb{P}}_{Y}(\,y\,)$$

e como os termos são não-negativos, podemos rearranjar a soma de modo que

$$\mathbb{E}[\,|XY|\,] = \left(\sum_{x\in S} |x|\, \mathbb{P}_X(x)\right) \cdot \left(\sum_{y\in R} |y|\, \mathbb{P}_Y(y)\right) = \mathbb{E}[\,|X|\,] \cdot \mathbb{E}[\,|Y|\,]$$

Assim, se X e Y são integráveis, XY também é e a mesma dedução sem os módulos vale, ou seja, concluímos que $\mathbb{E}[\,XY\,] = \mathbb{E}X \cdot \mathbb{E}Y$.

Teorema 2.7: produto de variáveis aleatórias independentes

Sejam $X\colon \Omega \to S$ e $Y\colon \Omega \to R$ variáveis aleatórias independentes em $(\Omega, \mathbb{P})$ e sejam $f\colon S \to \mathbb{R}$ e $g\colon R \to \mathbb{R}$ funções tais que $f(X)$ e $g(Y)$ são variáveis aleatórias integráveis.
Então $f(X)\cdot g(Y)$ é integrável e vale

$$\mathbb{E}[\,f(X)\cdot g(Y)\,] = \mathbb{E}[\,f(X)\,]\,\mathbb{E}[\,g(Y)\,].$$

Demonstração. De X e Y independentes, temos que $f(X)$ e $g(Y)$ são variáveis aleatórias independentes (justifique). Podemos replicar a dedução acima para provar que $f(X)g(Y)$ é integrável e que, portanto, $\mathbb{E}[\,f(X)\cdot g(Y)\,] = \mathbb{E}[\,f(X)\,]\cdot\mathbb{E}[\,g(Y)\,]$. □

Se X e Y são variáveis aleatórias independentes, então tomamos f, g como a função identidade e vale a igualdade

$$\mathbb{E}[\,XY\,] = \mathbb{E}[\,X\,]\,\mathbb{E}[\,Y\,]. \tag{2.24}$$

Observação 2. O Teorema 2.6, da monotonicidade, ainda vale com as hipóteses de $\mathbb{E}X$ e $\mathbb{E}Y$ definidos e $\mathbb{E}X - \mathbb{E}Y$ definido. Em particular, se as duas variáveis aleatórias, X e Y, têm valor esperado definidos e pelo menos uma é integrável, então as conclusões dos teoremas valem. Ainda, no Teorema 2.6 a conclusão vale se os valores esperados de X e de Y estão definidos e $\mathbb{E}X < +\infty$ ou $\mathbb{E}Y > -\infty$.

Exemplo 2.9: Esperança da distribuição geométrica

Uma variável geométrica conta o número de realizações de ensaios de Bernoulli independentes e identicamente distribuídos realizados até, e incluindo, o primeiro sucesso.
Se X é uma variável aleatória geométrica, então para $X = t$ é necessário que nas primeiras $t - 1$ iterações ocorra fracasso e na t-ésima iteração ocorra sucesso. Como os ensaios são independentes, a lei de X é

$$\mathbb{P}_X(t) := (1-p)^{t-1}p, \quad t \in \{1, 2, \dots\}$$

e dizemos que X tem ***distribuição geométrica*** com parâmetro p, o que denotamos por $X \sim \mathrm{Geom}(p)$.

Em alguns textos a distribuição geométrica conta o número de fracassos antes do primeiro sucesso. Nesse caso a variável aleatória $Y = X - 1$ assume valores em $\{0, 1, 2, \dots\}$ e tem distribuição $\mathbb{P}_Y(t) = \mathbb{P}_X(t+1)$.
Seja X o número de tentativas de Bernoulli até o primeiro sucesso, com probabilidade de sucesso p. Condicionando no primeiro resultado, seja A o evento "sucesso",

- A ocorre com probabilidade p na primeira tentativa, então $X = 1$;
- $\overline{A}$ ocorre com probabilidade $1 - p$ na primeira tentativa, nesse caso, já gastamos 1 tentativa, então $X = 1 + Y$, com $Y \sim \text{Geom}(p)$ (o processo recomeça do zero). Portanto,

Assim,

$$\begin{aligned}
\mathbb{E}[X] &= \sum_x x\mathbb{P}[X = x] \\
&= \sum_x x\big(\mathbb{P}[X = x \mid A]\mathbb{P}(A) + \mathbb{P}\big[X = x \mid \overline{A}\big]\mathbb{P}(\overline{A})\big) \\
&= \left(\sum_x x \underset{A}{\mathbb{P}}[X = x]\right)\mathbb{P}(A) + \left(\sum_x x \underset{\overline{A}}{\mathbb{P}}[X = x]\right)\mathbb{P}(\overline{A}) \\
&= \mathbb{E}[X \mid A]\mathbb{P}(A) + \mathbb{E}\big[X \mid \overline{A}\big]\mathbb{P}(\overline{A})
\end{aligned}$$

Agora, $\mathbb{E}[X \mid A] = \mathbb{E}[1] = 1$ e $\mathbb{E}\big[X \mid \overline{A}\big] = \mathbb{E}[Y] = \mathbb{E}[X] + 1$, portanto $\mathbb{E}[X] = 1 \cdot p + (\mathbb{E}[X] + 1)(1 - p)$ e resolvendo para $\mathbb{E}X$

$$\mathbb{E}[X] = \frac{1}{p}.$$

ANÁLISE DO ALGORITMO 1.2

Dado $M \geqslant 2$, tome $k = \lceil \log_2 M \rceil$. Seja X o número de rodadas do laço `repita`. O tempo de execução das instruções dentro do laço é $O(k) = O(\ln M)$. Assim, se $T(M)$ é o tempo de execução, então

$$\mathbb{E}T(M) = \mathbb{E}[X] \cdot O(k) = \frac{1}{p} \cdot O(\ln M)$$

onde $p = \mathbb{P}[N < M] = \frac{M}{2^k}$ é o parâmetro da variável geométrica X. De $p > \frac{1}{2}$ obtemos $\frac{1}{p} < 2$ logo

$$\mathbb{E}T(M) = O(\ln M).$$

Vimos no Exemplo 2.3 que o valor esperado pode ser pouco representativo, mas aqui sabemos mais. Usando o Exercício 11 no final deste capítulo temos que

$$\mathbb{P}[\text{no. de rodadas} > 10 \times \text{no. esperado de rodadas}] = (1-p)^{10 \times \text{no. esperado de rodadas}}$$

$$< \left(\frac{1}{2}\right)^{10/p} < \left(\frac{1}{2}\right)^{10} = 0{,}0009765625.$$

Exercícios

Exercício 5. Sejam $X = \sum_{k=1}^{n} x_k \mathbb{1}_{A_k}$ e $Y = \sum_{\ell=1}^{m} y_\ell \mathbb{1}_{B_\ell}$ duas variáveis aleatórias simples de $(\Omega, \mathbb{P})$. Verifique que valem as seguintes identidades

1. para quaisquer $a, b \in \mathbb{R}$

$$aX + bY = \sum_{k=1}^{n} \sum_{\ell=1}^{m} (ax_k + by_\ell) \mathbb{1}_{A_k \cap B_\ell};$$

2.

$$X \cdot Y = \sum_{k=1}^{n} \sum_{\ell=1}^{m} (x_k \cdot y_\ell) \mathbb{1}_{A_k \cap B_\ell}.$$

Exercício 6. Prove que se o conjunto universo U é suficientemente grande, digamos $|U| > (m-1)n$, então para qualquer função de *hash* $h \in N^U$ sempre haverá um conjunto $S \subset U$ de cardinalidade m para o qual uma configuração de pior caso é inevitável, ou seja, todos os elementos de S colidem.

Exercício 7 (***desigualdade de Jensen***). Sejam X uma variável aleatória integrável, $I \subseteq \mathbb{R}$ um intervalo tal que $\mathbb{P}[\, X \in I \,] = 1$ e $f \colon I \to \mathbb{R}$ uma função convexa em I, então $\mathbb{E}[\, f(X) \,]$ está definida e

$$\mathbb{E}[\, f(X) \,] \geqslant f(\mathbb{E}X). \tag{2.25}$$

$f \colon [a, b] \to \mathbb{R}$ é dita convexa se para quaisquer $x, y \in [a, b]$ e para todo $t \in [0, 1]$, tem-se que o segmento de reta que une x e y fica acima do gráfico de f entre x e y, isto é, $f(tx + (1-t)y) \leqslant tf(x) + (1-t)f(y)$

Exercício 8. Conclua do teorema acima que $\mathbb{E}|\, X \,| \geqslant |\, \mathbb{E}X \,|$ e que $\mathbb{E}[\, X^2 \,] \geqslant (\mathbb{E}X)^2$.

Exercício 9. Sejam U e N conjuntos finitos. Uma família de funções $\mathcal{H} \subseteq N^U$ que quando munida da medida uniforme satisfaz

$$\mathop{\mathbb{P}}_{h \in_{\mathrm{R}} \mathcal{H}} [\, h(u) = i \,] = \frac{1}{|N|}, \text{ para todo } u \in U \text{ e para todo } i \in N \tag{2.26}$$

não é suficiente para garantir um bom comportamento da família de funções de *hash* numa tabela de espalhamento. Verifique que a família $\mathcal{H}$ formada pelas $|N|$ funções constantes satisfaz (2.26).

Exercício 10. Sejam U e N conjuntos finitos. Uma família de funções $\mathcal{H} \subseteq N^U$ que quando munida da medida uniforme satisfaz

$$\mathop{\mathbb{P}}_{h \in_{\mathrm{R}} \mathcal{H}} [\, [h(u) = i] \cap [h(v) = j] \,] = \frac{1}{|N|^2}, \tag{2.27}$$

para todo $u, v \in U$, com $u \neq v$, e para todo $i, j \in N$, é dita uma família de funções de ***hash 2-independente***.

1. Verifique que essa condição implica a condição (2.26) para todo $u \in U$ e todo $i \in N$.

2. Explique por que a família formada pelas $|N|$ funções constantes não satisfaz a condição de 2-independência.

3. Verifique a Proposição 2.3 para qualquer escolha de função numa família 2-independente.

Exercício 11. Mostre que se Z tem distribuição geométrica com parâmetro p então $\mathbb{P}[\, Z > n-1 \,] = (1-p)^{n-1}$ para todo $n \geqslant 1$.

Exercício 12. Este exercício fornece uma maneira de derivar um algoritmo Las Vegas a partir de um algoritmo Monte Carlo.

Considere um algoritmo Monte Carlo A para um problema computacional Π cujo tempo de execução esperado é de no máximo $T(n)$ em qualquer instância de tamanho n e que produz uma solução correta com probabilidade $\gamma(n)$. Suponha ainda que, dada uma solução para Π, possamos verificar sua corretude no tempo $t(n)$. Mostre como obter um algoritmo Las Vegas que sempre fornece uma resposta correta para Π e roda em um tempo esperado de no máximo $(T(n) + t(n))/\gamma(n)$.

A
Sobre convergência de séries

Seja $(x_n : n \geqslant 1)$ uma sequência de números reais. Se $x_n \geqslant 0$ para todo n, então a soma parcial $\sum_{n=1}^{m} x_n$ é crescente e, portanto, sempre tende a um limite S, possivelmente infinito. Assim, a série $\sum_{n=1}^{\infty} x_n$ de termos não negativos

- ***converge*** para um número real (finito), caso em que escrevemos $\sum_{n\geqslant 1} x_n < \infty$; ou
- ***diverge*** para $+\infty$ (isto é, as somas parciais crescem sem limite), e escrevemos $\sum_{n\geqslant 1} x_n = \infty$.

Em ambos os casos, o valor da série não se altera se a ordem dos termos for modificada. No caso geral, onde $x_n \in \mathbb{R}$, definimos

$$X^+ := \sum_{n\colon x_n \geqslant 0} x_n \quad \text{e} \quad X^- := \sum_{n\colon x_n < 0} |x_n| \tag{A.1}$$

e podem ocorrer as seguintes situações.

1. Ambas as séries em (A.1) são finitas; nesse caso, $\sum_n x_n = X^+ - X^-$, a série é finita e tem um valor ***bem definido***. Além disso, a série $\sum_n x_n$ é ***absolutamente convergente***, isto é, $\sum_{n\geqslant 1} |x_n| < \infty$, e qualquer rearranjo dos termos converge para o mesmo valor.
2. Se $X^+ = +\infty$ e $X^- < \infty$, definimos

   $$\sum_{n=1}^{\infty} x_n := +\infty.$$

   De forma análoga, se $X^- = +\infty$ e $X^+ < \infty$, definimos

   $$\sum_{n=1}^{\infty} x_n := -\infty.$$

   Nesses casos, a série é ***divergente unilateralmente***, e qualquer rearranjo dos termos converge para o mesmo valor (infinito).
3. Nos outros casos, a série não é absolutamente convergente. Nesse caso, a soma pode depender da ordem dos termos, e diferentes rearranjos podem levar a valores distintos ou mesmo à divergência.

Quando os termos da série são números reais quaisquer e a série é absolutamente convergente, podemos alterar arbitrariamente a ordem dos termos sem alterar a propriedade de convergência absoluta, nem a soma da série.

A.1 Sequências e séries

(s.1) Uma série $\sum_{i\geqslant 1} x_i$ ***converge*** se existe o limite das somas parciais $S_n := \sum_{i=1}^{n} x_i$

$$\lim_{n\to\infty} S_n = \lim_{n\to\infty} \sum_{i=1}^{n} x_i \tag{A.2}$$

e ***converge absolutamente*** se $\sum_{i\geqslant 1} |x_i|$ converge. Uma prova dos seguintes resultados pode ser encontrada em [1].

(a) Uma série que converge absolutamente também converge.

(b) Se $\sum_n x_n$ converge e $\sum_n y_n$ é a série obtida de $\sum_n x_n$ eliminando-se os termos $x_n = 0$, então $\sum_n y_n$ converge para o mesmo valor.

(c) Se uma série converge absolutamente então qualquer série obtida rearranjando os termos dessa série converge absolutamente para o mesmo valor. Ademais, vale que se $\{ A_i : i \in I \}$ é uma partição finita ou enumerável de $\mathbb{N}$ e $y_i = \sum_{j\in A_i} x_j$, então

$$\sum_n x_n = \sum_{i\in I} y_i.$$

(s.2) Se $\sum_{i\geqslant 1} x_i$ é tal que $x_i \geqslant 0$ para todo i, então a série converge ou tende ao infinito. A série converge se, e só se, converge absolutamente.

(s.3) Teste de convergência por comparação: se $0 \leqslant x_n \leqslant y_n$ para todo n suficientemente grande e $\sum_n y_n$ converge então $\sum_n x_n$ converge.

(s.4) a convergência absoluta de uma série implica que toda subsérie da série é convergente. Isso segue de (s.1b) acima e do teste de convergência por comparação.

(s.5) Seja $(x_{n,m} : n, m \in \mathbb{N})$ uma sequência duplamente indexada.

(Tonelli) Se $x_{n,m} \geqslant 0$ para todo (n, m), então as somas iteradas e a soma dupla coincidem:

$$\sum_{n\in\mathbb{N}} \sum_{m\in\mathbb{N}} x_{n,m} = \sum_{(n,m)\in\mathbb{N}\times\mathbb{N}} x_{n,m} = \sum_{m\in\mathbb{N}} \sum_{n\in\mathbb{N}} x_{n,m} \in [0, +\infty].$$

(Fubini) Se $\sum_{(n,m)\in\mathbb{N}\times\mathbb{N}} |x_{n,m}| < \infty$, então a série $\sum_{(n,m)} x_{n,m}$ é absolutamente convergente e

$$\sum_{n\in\mathbb{N}} \sum_{m\in\mathbb{N}} x_{n,m} = \sum_{(n,m)\in\mathbb{N}\times\mathbb{N}} x_{n,m} = \sum_{m\in\mathbb{N}} \sum_{n\in\mathbb{N}} x_{n,m}, \tag{A.3}$$

com todas as somas convergindo para o mesmo valor real.

(s.6) se $|x| < 1$ então

(a) $\sum_{n \geqslant 0} x^n = (1-x)^{-1}$;
(b) $\sum_{n \geqslant k} x^n = x^k (1-x)^{-1}$;
(c) $\sum_{n \geqslant 1} n x^n = x(1-x)^{-2}$.

(s.7) $\sum_{n \geqslant 1} n^{-2} = \pi^2/6$.

A convergência dessa série foi estabelecida pela primeira vez por Euler e é um dos resultados responsáveis por lançar à notoriedade, na epóca, o matemático.

(s.8) A sequência $(1 + 1/n)^n$ é estritamente crescente e a sequência $(1 - 1/n)^{-n}$ é estritamente decrescente, ambas convergem para a constante $e = 2{,}71\ldots$ quando $n \to \infty$. Além disso, para todo x real

$$e^x = \sum_{n \geqslant 0} \frac{x^n}{n!}.$$

B
Desigualdades

(d.1) De (s.8), para todo $n \geqslant 0$ vale

$$\left(1-\frac{1}{n}\right)^{-n} \geqslant \mathrm{e} \geqslant \left(1+\frac{1}{n}\right)^{n}$$

e para qualquer $x > 0$

$$1 - x + \frac{x^2}{2} > \mathrm{e}^{-x} > (1-x).$$

(d.2) O coeficiente binomial tem os seguintes limitantes triviais

$$0 \leqslant \binom{n}{i} \leqslant 2^n.$$

Para um limitante inferior melhor podemos escrever

$$\binom{n}{i} = \prod_{j=0}^{i-1} \frac{n-j}{j-i} \geqslant \left(\frac{n}{i}\right)^{i}$$

e para um limitante superior usamos o teorema binômial de Newton e (s.8)

$$\mathrm{e}^{nx} \geqslant (1+x)^n = \sum_{i=0}^{n} \binom{n}{i} x^i \geqslant \binom{n}{i} x^i$$

para todo $x > 0$ e todo $i \in \{\, 0, 1, \ldots, n \,\}$. Logo

$$\binom{n}{i} \leqslant \mathrm{e}^{nx} x^{-i}.$$

Em particular, fazendo $x = i/n$, temos a segunda desigualdade de

$$\left(\frac{n}{i}\right)^{i} \leqslant \binom{n}{i} \leqslant \left(\mathrm{e}\frac{n}{i}\right)^{i}.$$

(d.3) A seguinte igualdade assintótica conhecida como fórmula de Stirling

$$n! = \left(1 + O\left(\frac{1}{n}\right)\right) \sqrt{2\pi n} \left(\frac{n}{\mathrm{e}}\right)^{n} = (1 + o(1)) \sqrt{2\pi n} \left(\frac{n}{\mathrm{e}}\right)^{n} \qquad \text{(B.1)}$$

A seguinte tabela que exibe exemplos para números pequenos nos dá ideia da qualidade dessa aproximação

n	$n!$	Stirling
1	1	0.922137
3	6	5.83621
7	5040	4980.396
10	3628800	3598696
20	$2.432902e+18$	$2.422787e+18$
50	$3.041409e+64$	$3.036345e+64$
75	$2.480914e+109$	$2.478159e+109$
100	$9.332622e+157$	$9.324848e+157$
142	$2.695364e+245$	$2.693783e+245$

também valem os limitantes

$$\sqrt{2\pi n}\Big(\frac{n}{\mathrm{e}}\Big)^n \leqslant n! \leqslant \mathrm{e}\sqrt{n}\Big(\frac{n}{\mathrm{e}}\Big)^n$$

Nos limitantes em (d.2) para $\binom{n}{i}$ usamos que

$$\prod_{j=0}^{i-1}(n-j) > (n-i)^i = \left(1-\frac{i}{n}\right)^i n^i \geqslant \left(1-\frac{i^2}{n}\right)n^i = (1+o(1))n^i$$

e se $i = o(\sqrt{n})$

$$\binom{n}{i} = (1+o(1))\frac{n^i}{i!} = (1+o(1))\frac{1}{\sqrt{2\pi i}}\Big(\frac{\mathrm{e}n}{i}\Big)^i.$$

(d.4) Para quaisquer números reais $x_1, x_2, \ldots$, para todo natural n

$$|x_1 + \cdots + x_n| \leqslant |x_1| + \cdots + |x_n|.$$

Ademais, $|x_1| + \cdots + |x_n| \leqslant |x_1| + |x_2| + \cdots$, portanto, $|x_1 + \cdots + x_n| \leqslant |x_1| + |x_2| + \cdots$ e pela continuidade da função valor absoluto $\lim_{n\to\infty} |x_1 + \cdots + x_n| = |x_1 + x_2 + \cdots|$ portanto

$$|x_1 + x_2 + \cdots| \leqslant |x_1| + |x_2| + \cdots$$

sempre que $\sum_n x_n$ converge.

C
Resolução dos exercícios

1. Sejam os eventos:

   - U_1: urna escolhida é a das duas bolas brancas;
   - U_2: urna escolhida é a das duas bolas pretas;
   - U_3: urna escolhida é a mista (uma branca, uma preta);
   - Br: a bola retirada é branca.

   Queremos a probabilidade de a bola que ficou na urna ser preta, o que só pode acontecer se a urna escolhida foi a urna U_3 (mista). Ou seja, queremos $\mathbb{P}(\, U_3 \mid Br\,)$.

   Cada urna é escolhida com probabilidade $1/3$: $\mathbb{P}(\, U_1\,) = \mathbb{P}(\, U_2\,) = \mathbb{P}(\, U_3\,) = \frac{1}{3}$.

   A probabilidade de sair bola branca dado cada urna:

   $$\mathbb{P}(\, Br \mid U_1\,) = 1, \qquad \mathbb{P}(\, Br \mid U_2\,) = 0, \qquad \mathbb{P}(\, Br \mid U_3\,) = \frac{1}{2}.$$

   Probabilidade total de sair branca

   $$\begin{aligned}\mathbb{P}(\, Br\,) &= \mathbb{P}(\, Br \mid U_1\,)\mathbb{P}(\, U_1\,) + \mathbb{P}(\, Br \mid U_2\,)\mathbb{P}(\, U_2\,) + \mathbb{P}(\, Br \mid U_3\,)\mathbb{P}(\, U_3\,) \\ &= 1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{3} + \frac{1}{6} = \frac{1}{2}.\end{aligned}$$

   Agora,

   $$\mathbb{P}(\, U_3 \mid Br\,) = \frac{\mathbb{P}(\, Br \mid U_3\,)\,\mathbb{P}(\, U_3\,)}{\mathbb{P}(\, Br\,)} = \frac{\frac{1}{2} \cdot \frac{1}{3}}{\frac{1}{2}} = \frac{1/6}{1/2} = \frac{1}{3}.$$

   Portanto $\mathbb{P}[\,\text{bola restante é preta} \mid \text{bola retirada é branca}\,] = \frac{1}{3}$.

2.

3. Seja k o número de bits de M, ou seja, $2^{k-1} \leqslant M < 2^k$. No algoritmo, $N = \sum_i d_i 2^i$ é obtido a partir de $k+t$ bits uniformes e independentes, logo N é uniforme sobre $\{0, 1, \ldots, K-1\}$, onde $K = 2^{k+t}$. A saída do algoritmo é $N \bmod M$.

   Na divisão euclidiana de K por M

$$K = qM + r, \quad 0 \leqslant r < M, \quad q = \left\lfloor \frac{K}{M} \right\rfloor$$

   e para cada resto $n \in \{0, \ldots, M-1\}$, os inteiros em $\{0, \ldots, K-1\}$ congruentes a n módulo M são

$$n,\ n+M,\ n+2M,\ \ldots$$

   e há exatamente $q+1$ deles se $n < r$, e exatamente q deles se $n \geq r$ (isso decorre diretamente de $K = qM + r$).

   Como N é uniforme em $\{0, \ldots, K-1\}$,

$$\mathbb{P}[\, N = n \,] = \begin{cases} \dfrac{q+1}{K}, & 0 \leq n < r, \\[2ex] \dfrac{q}{K}, & r \leq n \leq M-1. \end{cases}$$

   Agora, calculando os desvios em relação a $1/M$.

   De $K = qM + r$ obtemos $q = \dfrac{K-r}{M}$, logo:

$$\frac{1}{M} - \frac{q}{K} = \frac{1}{M} - \frac{K-r}{MK} = \frac{r}{MK},$$

$$\frac{q+1}{K} - \frac{1}{M} = \frac{K-r+M}{MK} - \frac{1}{M} = \frac{M-r}{MK}.$$

   Ambas as quantidades são não negativas (pois $0 \leq r < M$), então

$$\left| \mathbb{P}[\, N = n \,] - \frac{1}{M} \right| = \begin{cases} \dfrac{M-r}{MK}, & 0 \leq n < r, \\[2ex] \dfrac{r}{MK}, & r \leq n \leq M-1. \end{cases}$$

   Há r valores de n no primeiro caso e $M-r$ valores no segundo, logo

$$\sum_{n=0}^{M-1} \left| \mathbb{P}[\, N = n \,] - \frac{1}{M} \right| = r \cdot \frac{M-r}{MK} + (M-r) \cdot \frac{r}{MK} = \frac{2r(M-r)}{MK}.$$

   A função $r \mapsto r(M-r)$ é uma parábola côncava em r, com máximo em $r = M/2$, onde vale $M^2/4$. Assim, para todo $r \in [0, M]$ (em particular para o inteiro r em questão),

$$r(M-r) \leqslant \frac{M^2}{4}.$$

Substituindo, usando que $M < 2^k$ e $K = 2^{k+t}$

$$\sum_{n=0}^{M-1} \left| \mathbb{P}[\, N = n \,] - \frac{1}{M} \right| \leqslant \frac{2}{MK} \cdot \frac{M^2}{4} = \frac{M}{2K} < \frac{1}{2^{t+1}}.$$

4. seja p_n a probabilidade de o laço `repeat–until` executar pelo menos n vezes, para $n \geqslant 1$.

```
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.
```

Na j-ésima execução do corpo do laço, sorteiam-se j bits $d_1, \ldots, d_j$ uniformes e independentes em $\{0,1\}$, e o laço termina se, e somente se, todos eles forem iguais a 1. Como esses bits são *redefinidos do zero* a cada execução, o evento

$$E_j := [\text{“todos os } j \text{ bits sorteados na iteração } j \text{ são iguais a 1”}]$$

tem probabilidade

$$\mathbb{P}[\, E_j \,] = 2^{-j},$$

e os eventos $E_1, E_2, E_3, \ldots$ são mutuamente independentes, pois dizem respeito a sorteios distintos e independentes entre si.

O laço executa pelo menos n vezes se, e somente se, ele não é interrompido em nenhuma das iterações $j = 1, \ldots, n-1$, ou seja, se nenhum dos eventos $E_1, \ldots, E_{n-1}$ ocorre. Pela independência,

$$p_n = \mathbb{P}[\, \overline{E_1} \cap \cdots \cap \overline{E_{n-1}} \,] = \prod_{j=1}^{n-1} \left(1 - 2^{-j}\right),$$

com a convenção usual de que o produto vazio (caso $n = 1$) vale 1.

Para cada $j \geqslant 1$ temos $2^{-j} \in (0,1)$, logo cada fator $1 - 2^{-j}$ pertence a $(0,1)$. Como p_n é um produto *finito* de números estritamente positivos, segue que $p_n > 0$.

Como

$$p_{n+1} = p_n \cdot \left(1 - 2^{-n}\right) \quad \text{e} \quad 0 < 1 - 2^{-n} < 1,$$

temos $p_{n+1} \leqslant p_n$ para todo n. Sendo decrescente e limitada inferiormente por 0, a sequência (p_n) converge para algum limite $L \geqslant 0$. Resta mostrar que $L > 0$.

Como $2^{-j} \leqslant \frac{1}{2}$ para todo $j \geqslant 1$, podemos aplicar a desigualdade $1 - x \geqslant e^{-2x}$, válida para $x \in [0, 1/2]$, com $x = 2^{-j}$:

$$1 - 2^{-j} \;\geqslant\; \exp\left(-2 \cdot 2^{-j}\right) = \exp\left(-2^{1-j}\right).$$

Multiplicando essas desigualdades para $j = 1, \ldots, n-1$ (todos os fatores são positivos, logo a desigualdade é preservada pelo produto):

$$p_n = \prod_{j=1}^{n-1} \left(1 - 2^{-j}\right) \geqslant \prod_{j=1}^{n-1} \exp\left(-2^{1-j}\right) = \exp\left(-\sum_{j=1}^{n-1} 2^{1-j}\right) = \exp\left(-2\sum_{j=1}^{n-1} 2^{-j}\right).$$

A soma geométrica satisfaz

$$\sum_{j=1}^{n-1} 2^{-j} = 1 - 2^{-(n-1)} < 1 \qquad \text{para todo } n \geqslant 1,$$

donde

$$p_n \geqslant \exp\left(-2\left(1 - 2^{-(n-1)}\right)\right) > \exp(-2),$$

para todo $n \geqslant 1$. Ou seja, obtemos uma cota inferior *uniforme*, independente de n:

$$p_n > e^{-2} \qquad \text{para todo } n.$$

Como $(p_n)_{n\geqslant 1}$ é decrescente, converge para $L = \lim_{n\to\infty} p_n$, e como $p_n > e^{-2}$ para todo n, segue que

$$L = \lim_{n\to\infty} p_n \geqslant e^{-2} > 0.$$

Portanto $\lim_{n\to\infty} p_n > 0$.

5.

6. A prova é uma aplicação direta do princípio da casa dos pombos: se cada lista $h^{-1}(i)$ tivesse no máximo $m-1$ elementos, teríamos

$$|U| = \sum_{i=0}^{n-1} |h^{-1}(i)| \leqslant n(m-1),$$

contradizendo $|U| > (m-1)n$. Logo, alguma lista contém pelo menos m elementos, e podemos escolher S com m deles.

7. De f convexa em I, deduzimos que para todo $a \in I$ existe um real m_a tal que

$$\frac{f(a) - f(x)}{a - x} \leqslant m_a \leqslant \frac{f(y) - f(a)}{y - a}$$

para todo $x < a$ e todo $y > a$. Portanto, $f(x) \geqslant f(a) + m_a(x - a)$ para todo $x \in I$ e, tomando $a = \mathbb{E}X$, $x = X$ e a esperança nos dois lados dessa desigualdades

$$\mathbb{E}[\, f(X)\,] \geqslant f(\mathbb{E}X) + m_a \mathbb{E}(X - \mathbb{E}[\, X\,])$$

donde segue a desigualdade que queremos demonstrar.

8.

9. Seja $\mathcal{H} = \{h_j : j \in N\}$ a família das funções constantes, onde para cada $j \in N$ a função $h_j \colon U \to N$ é dada por $h_j(u) = j$ para todo $u \in U$.

   Note que $|\mathcal{H}| = |N|$, uma função constante para cada elemento de N, e há uma bijeção natural entre $\mathcal{H}$ e N (a função h_j corresponde ao elemento j).

   Fixe $u \in U$ e $i \in N$ arbitrários. Como $h \in_{\mathrm{R}} \mathcal{H}$ significa escolher $h = h_j$ com $j \in_{\mathrm{R}} N$ (pela bijeção entre $\mathcal{H}$ e N), temos

   $$\mathop{\mathbb{P}}_{h \in_{\mathrm{R}} \mathcal{H}}\big[\, h(u) = i \,\big] = \mathop{\mathbb{P}}_{j \in_{\mathrm{R}} N}\big[\, h_j(u) = i \,\big].$$

   Mas $h_j(u) = j$ para todo u (por definição, independente de u), logo

   $$h_j(u) = i \iff j = i.$$

   Portanto,

   $$\mathop{\mathbb{P}}_{j \in_{\mathrm{R}} N}\big[\, h_j(u) = i \,\big] = \mathop{\mathbb{P}}_{j \in_{\mathrm{R}} N}[\, j = i \,] = \frac{1}{|N|},$$

   pois j é escolhido uniformemente em N e há exatamente um valor de j (a saber, $j = i$) que satisfaz a condição.

   Como $u \in U$ e $i \in N$ foram arbitrários, concluímos que

   $$\mathop{\mathbb{P}}_{j \in_{\mathrm{R}} N}\big[\, h_j(u) = i \,\big] = \frac{1}{|N|} \qquad \text{para todo } u \in U \text{ e todo } i \in N,$$

   ou seja, $\mathcal{H}$ satisfaz (2.26).

   ### Por que isso não é suficiente na prática

   Embora a condição (2.26) seja satisfeita, essa família é péssima para uma tabela de espalhamento: uma vez escolhida (uniformemente) uma função $h = h_j \in \mathcal{H}$, todos os elementos de U são mapeados no mesmo balde j:

   $$h(u_1) = h(u_2) = \cdots = h(u_{|U|}) = j.$$

   Isso mostra que a condição (2.26) — que garante apenas que cada elemento individual, isoladamente, tem posição marginal uniforme — é fraca demais: ela nada diz sobre a relação conjunta entre $h(u_1)$ e $h(u_2)$ para $u_1 \neq u_2$. O que realmente se deseja de uma boa família de hash é algo como independência par a par (ou universalidade), isto é, algo do tipo

   $$\mathop{\mathbb{P}}_{h \in_{\mathrm{R}} \mathcal{H}}\big[\, h(u_1) = i_1 \wedge h(u_2) = i_2 \,\big] \approx \frac{1}{|N|^2} \qquad \text{para } u_1 \neq u_2,$$

condição que a família de funções constantes viola drasticamente (nesse caso essa probabilidade conjunta é $\frac{1}{|N|}$, não $\frac{1}{|N|^2}$), pois conhecer $h(u_1)$ determina completamente $h(u_2)$.

10. Seja $n = |N|$. A condição de 2-independência diz que, para quaisquer $u, v \in U$, com $u \neq v$, e quaisquer $i, j \in M$,

$$\mathop{\mathbb{P}}_{h \in_R \mathcal{H}}[\,(u) = i \text{ e } h(v) = j\,] = \frac{1}{n^2}.$$

(a) Fixemos $u \in U$ e $i \in N$. Assuma $|U| \geqslant 2$ e escolha um elemento $v \in U$, $v \neq u$. Como $\{\, h \in \mathcal{H} \colon [h(v) = j]\ j \in N \,\}$ particiona $\mathcal{H}$, temos

$$\begin{aligned}\mathop{\mathbb{P}}_{h \in_R \mathcal{H}}[\,h(u) = i\,] &= \sum_{j \in N} \mathop{\mathbb{P}}_{h \in_R \mathcal{H}}[\,h(u) = i \text{ e } h(v) = j\,] \\ &= \sum_{j \in N} \frac{1}{n^2} = \frac{n}{n^2} = \frac{1}{n}.\end{aligned}$$

para todo $u \in U$ e todo $i \in n$.

(b) Considere a família $\mathcal{H}$ formada pelas $n = |N|$ funções constantes. Para cada $i \in M$, seja h_i a função constante dada por

$$h_i(u) = i, \qquad u \in U.$$

Fixemos $u, v \in U$, com $u \neq v$, e escolhase $i, j \in N$ com $i \neq j$. Como uma função constante não pode satisfazer simultaneamente $h(u) = i$ e $h(v) = j$, temos

$$\mathop{\mathbb{P}}_{h \in_R \mathcal{H}}[\,h(u) = i \text{ e } h(v) = j\,] = 0.$$

(c)

11.

12. Construiremos um algoritmo Las Vegas B a partir do algoritmo Monte Carlo A da seguinte maneira.

Dada uma instância I de tamanho n, o algoritmo B repete independentemente os seguintes passos:

(a) Execute A sobre I e obtenha uma solução s.

(b) Verifique a corretude de s em tempo $t(n)$.

(c) Se s for correta, retorne s; caso contrário, repita.

Como a solução só é retornada depois de ser verificada, o algoritmo B sempre fornece uma resposta correta. Resta analisar seu tempo esperado de execução.

Cada tentativa consiste em executar A e verificar a solução obtida. Como o tempo esperado de execução de A é no máximo $T(n)$, o tempo esperado de uma tentativa é no máximo $T(n) + t(n)$.

Além disso, pela hipótese, cada execução de A produz uma solução correta com probabilidade $\gamma(n)$. Como as execuções são independentes, o número N de tentativas até obter a primeira solução correta tem distribuição geométrica com parâmetro $\gamma(n)$. Assim, $\mathbb{E}N = \frac{1}{\gamma(n)}$.

Seja X_k o tempo gasto na k-ésima tentativa. O tempo total de B é $X_1 + \cdots + X_N$.

Como cada tentativa tem tempo esperado no máximo $T(n) + t(n)$, obtemos

$$\mathbb{E}\left[\sum_{k=1}^{N} X_k\right] \leqslant \mathbb{E}[\,N\,]\big(T(n) + t(n)\big).$$

Portanto,

$$\mathbb{E}[\text{tempo}(B)] \leqslant \frac{T(n) + t(n)}{\gamma(n)}.$$

Logo, B é um algoritmo Las Vegas: ele sempre retorna uma solução correta e seu tempo esperado de execução é, no máximo,

$$\frac{T(n) + t(n)}{\gamma(n)}.$$

Referências Bibliográficas

Robert G. Bartle. *The elements of real analysis*. John Wiley & Sons, New York-London-Sydney, second edition, 1976.

Sheldon Ross. *Probabilidade: Um Curso Moderno com Aplicações*. Porto Alegre: Bookman, 8.º ed., 2010.

Walter Rudin. *Principles of Mathematical Analysis*. International Series in Pure and Applied Mathematics. McGraw-Hill, New York, 3rd edition, 1976.

Devdatt Dubhashi e Alessandro Panconesi. *Concentration of Measure for the Analysis of Randomized Algorithms*. Cambridge: Cambridge University Press, 2009.

Juraj Hromkovic. *Design and Analysis of Randomized Algorithms: Introduction to Design Paradigms*. Berlin: Springer, 2005. (online)

Sanjeev Arora e Boaz Barak. *Computational Complexity: A Modern Approach*. Cambridge: Cambridge University Press, 2009.

Avrim Blum, John Hopcroft e Ravindran Kannan. *Foundations of Data Science*. Cambridge: Cambridge University Press, 2020. (online)

Roman Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge: Cambridge University Press, 2018. (online, soluções)

Olle Häggström. *Finite Markov Chains and Algorithmic Applications*. Cambridge: Cambridge University Press, 2002.

Michael Mitzenmacher e Eli Upfal. *Probability and Computing: Randomized Algorithms and Probabilistic Analysis*. Cambridge: Cambridge University Press, 2005.

R. Motwani e P. Raghavan. *Randomized Algorithms*. Cambridge: Cambridge University Press, 1995.

Índice Remissivo