3 | Variáveis aleatórias

As variáveis aleatórias são funções que capturam grandezas observáveis nos fenômenos aleatórios como, por exemplo, o tempo de execução de um algoritmo aleatorizado. Uma função $X\colon \Omega \to \mathbb{R}$ é uma variável aleatória (real) se e somente se os conjuntos

$$[X \leqslant t] \coloneqq \{\omega \in \Omega : X(\omega) \leqslant t\} \tag{3.1}$$

são eventos de $(\Omega, \mathcal{A}, \mathbb{P})$ para todo $t \in \mathbb{R}$. Com uma variável aleatória $X\colon (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B})$ associamos a estrutura abstrata $(\Omega, \mathcal{A})$ do modelo probabilístico a uma estrutura conhecida, $(\mathbb{R}, \mathcal{B})$ é a σ-álgebra de Borel[1], e isso nos permite trabalhar nesse espaço conhecido. A σ-álgebra de Borel é suficiente para representar todos eventos de interesse em problemas práticos que envolvam probabilidade.

[1] Dada pela interseção de todas as σ-álgebras que contêm os intervalos $(-\infty, t]$ para todo $t \in \mathbb{R}$. A interseção de uma família de σ-álgebras é σ-álgebra.

Para uma variável aleatória real $X\colon \Omega \to \mathbb{R}$ os eventos do espaço amostral Ω definidos por $[X \leqslant t]$, para qualquer $t \in \mathbb{R}$, são particularmente importantes no estudo de variáveis aleatórias, eles descrevem completamente o comportamento da variável aleatória. A função $F_X(t) := \mathbb{P}[X \leqslant t]$ é chamada **função de distribuição acumulada** de X. De modo análogo podemos definir os subconjuntos $[X = t]$ e $[X \geqslant t]$ de Ω.

3.1 VARIÁVEIS ALEATÓRIAS DISCRETAS

Um exemplo trivial de variável aleatória é dado por uma função constante. Se X é constante, as possibilidades para os conjuntos $[X \leqslant t]$, definidos na equação (3.1), são os eventos $\varnothing$ ou Ω, dependendo do valor de t (verifique). No modelo clássico para o lançamento de um dado equilibrado a função $X\colon \{1,2,3,4,5,6\} \to \{1,2,3,4,5,6\}$ dada por $X(\omega) = \omega$, para todo ω, é variável aleatória. A resposta do Algoritmo 1, página 24, e o tempo de execução do Algoritmo 2, página 43, também são exemplos de variáveis aleatórias.

Exemplo 3.1 (*variável aleatória indicadora*). Este exemplo introduz uma variável aleatória que é muito útil em Probabilidade. 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

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

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

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

ademais $\mathbb{1}_A^{-1}(B)$ vale $\varnothing$, 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 $B \in \mathcal{B}$. ◊

Claramente, em um espaço discreto toda função de Ω em $\mathbb{R}$ satisfaz a equação (3.1), portanto, é uma variável aleatória. A definição de variável aleatória discreta é um pouco mais genérica que funções sobre um espaço discreto.

Exercício 3.2. Mostre que se X é uma variável aleatória que assume valores em $\mathbb{Z}^+$ então

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

Lei de uma variável aleatória Se X é uma variável aleatória do modelo probabilístico $(\Omega, \mathcal{A}, \mathbb{P})$, fica definido um espaço de probabilidade $(\mathbb{R}, \mathcal{B}, \mathbb{P}_X)$ induzido por X da seguinte forma. Definimos para cada evento $B \in \mathcal{B}$ o evento $[X \in B] \in \mathcal{A}$ por

$$[X \in B] = X^{-1}(B) \coloneqq \{\omega \in \Omega : X(\omega) \in B\}$$

e $\mathbb{P}_X(B) = \mathbb{P}[X \in B]$ é uma medida de probabilidade chamada **distribuição de** X, ou **lei de** X.

No caso em que existe um conjunto enumerável $S \subseteq \mathbb{R}$ tal que $\mathbb{P}(X \in S) = 1$, dizemos que X é uma **variável aleatória discreta** e que X tem uma lei, ou distribuição, discreta. A lei de X é definida pela probabilidade dos eventos elementares $\{t\} \subset S$,

$$[X = t] \coloneqq [X \in \{t\}] = X^{-1}(t), \text{ para todo } t \in S.$$

De fato, de $\mathbb{P}_X(\overline{S}) = 0$ e de $\mathbb{P}_X(t) = 0$ para todo $t \notin S$, temos que

$$\mathbb{P}_X(B) = \mathbb{P}_X(B \cap S) = \mathbb{P}\left(\bigcup_{t \in S \cap B} [X = t] \right) = \sum_{t \in S \cap B} \mathbb{P}[X = t] = \sum_{t \in B} \mathbb{P}_X(t) \quad (\forall B \in \mathcal{B}).$$

Claramente, $0 \leqslant \mathbb{P}_X(A) \leqslant 1$, para todo $A \subset S$ e $\mathbb{P}_X(S) = 1$. Ademais, vale para quaisquer $A_1, A_2, \ldots$ subconjuntos disjuntos de S que

$$\mathbb{P}_X\left(\bigcup_{n \geqslant 1} A_n \right) = \mathbb{P}\left(X^{-1}\left(\bigcup_{n \geqslant 1} A_n \right) \right) = \mathbb{P}\left(\bigcup_{n \geqslant 1} X^{-1}(A_n) \right) = \sum_{n \geqslant 1} \mathbb{P}\left(X^{-1}(A_n) \right) = \sum_{n \geqslant 1} \mathbb{P}_X(A_n)$$

de modo que de X obtemos o espaço de probabilidade discreto $(S, \mathbb{P}_X)$.

De um modo geral, se temos uma função de probabilidade $\mathcal{D}$, então temos um espaço de probabilidade sobre $(\mathbb{R}, \mathcal{B})$ pondo $\mathbb{P}_X(B) = \sum_{t \in B} \mathcal{D}(t)$ e no qual está definida a variável aleatória discreta $X(\omega) = \omega$ cuja lei é dada por $\mathcal{D}$. Além disso, se uma variável aleatória discreta Y tem a a mesma lei que X, ou seja $\mathbb{P}_Y(t) = \mathcal{D}(t)$ para todo $t \in S$, vamos dizer que Y tem distribuição $\mathcal{D}$ e nesse caso escrevemos

$$Y \in_{\mathcal{D}} S$$

ou, como é mais comum nos textos, $Y \sim \mathcal{D}$. Duas variáveis aleatórias que seguem a mesma lei são ditas **identicamente distribuídas**. Por exemplo, se X é uma variável aleatória sobre $(\Omega, \mathcal{A}, \mathbb{P})$ e $\tilde{X} \colon \mathbb{R} \to \mathbb{R}$ é a função identidade, então $\tilde{X}$ é uma variável aleatória sobre $(\mathbb{R}, \mathcal{B}, \mathbb{P}_X)$ que segue a mesma lei de X (verifique). Além disso, escrevemos $x \in_{\mathcal{D}} S$ para indicar que $x \in S$ é um elemento de S escolhido de acordo com a distribuição $\mathcal{D}$.

Na maior parte deste texto, trabalhamos com variáveis aleatórias discretas, de modo que omitiremos, a partir de agora, o adjetivo "discreta". Usaremos quase sempre as letras maiúsculas finais do alfabeto, como X, Y, Z, W, para denotar variáveis aleatórias.

Distribuição de Bernoulli É comum ocorrerem situações com experimentos para os quais nos interessa apenas observar duas características dos resultados: *sucesso* ou *fracasso*. Por exemplo, uma peça de uma linha de produção é classificada como *boa* ou *defeituosa*; o resultado de um exame médico é *positivo* ou *negativo*; um entrevistado *concorda* ou *não concorda* com uma afirmação proposta pelo entrevistador; a condição de um laço num algoritmo é *verdadeira* ou *falsa*; um evento A *ocorreu* ou *não ocorreu*. Esses experimentos recebem o nome de **ensaio de Bernoulli** e são modelados como uma variável aleatória $X\colon \Omega \to \mathbb{R}$ com $\mathbb{P}_X(1) = p$, $\mathbb{P}_X(0) = 1 - p$ e $\mathbb{P}_X(t) = 0$ se $t \neq 0, 1$, isto é, definimos a função de probabilidade

$$b_p(t) := p^t(1-p)^{1-t}, \text{ para } t \in \{0,1\}$$

em que $p = \mathbb{P}[X = 1]$ é a *probabilidade de sucesso* do ensaio de Bernoulli, e $b_p(t) = 0$ nos outros casos.

Dizemos que X tem **distribuição de Bernoulli** com parâmetro p se $\mathbb{P}_X(x) = b_p(x)$ e usamos a notação $X \in_{b_p} \{0,1\}$ ou $X \sim \text{Bernoulli}(p)$ para indicar tal fato.

Distribuição uniforme discreta Uma varável aleatória X que assume qualquer valor de um conjunto finito S com a mesma probabilidade

$$\mathcal{U}(t) = \frac{1}{|S|}$$

para todo $t \in S$, tem **distribuição uniforme** sobre S e denotamos esse fato por $X \in_R S$ ou $X \sim \mathcal{U}(S)$.

Distribuição geométrica Uma variável geométrica conta o número de realizações de ensaios de Bernoulli independentes e idênticos até que ocorra um sucesso. Por exemplo, no laço do Algoritmo 2, página 43, que reproduzimos abaixo

repita
  para cada $i \in \{0, \ldots, \lfloor \log_2 M \rfloor\}$ **faça** $d_i \xleftarrow{R} \{0,1\}$;
  $N \leftarrow \sum_i d_i 2^i$;
até que $N < M$;

cada execução das linhas internas é um ensaio de Bernoulli e um *succeso* ocorre quando a condição $N < M$ é verdadeira, estamos interessados no número de repetições até ocorrer um sucesso. Se X é uma variável aleatória que conta o número de ensaios até que ocorra um sucesso, então a lei de X é

$$\mathcal{G}_p(t) = (1-p)^{t-1}p, \text{ para todo inteiro } t \geqslant 1$$

e dizemos que X tem **distribuição geométrica** com parâmetro p, o que denotamos por $X \sim \text{Geom}(p)$ ou $X \in_{\mathcal{G}_p} \mathbb{Z}^+$.

Em alguns textos a distribuição geométrica conta o número de fracassos até o primeiro sucesso, isso faz com que as distribuições sejam diferentes por um fator $1-p$, ou ainda, essa nova distribuição vale $\mathcal{G}_p(t+1)$ em t.

Exercício 3.3. Mostre que se Z tem distribuição geométrica com parâmetro p então $\mathbb{P}[\,Z > n-1\,] = (1-p)^{n-1}$ para todo $n \geqslant 1$.

Distribuição binomial Em n ensaios de Bernoulli idênticos e independentes uma resposta específica $(b_1, b_2, \ldots, b_n)$ ocorre com probabilidade $p^t(1-p)^{n-t}$ sempre que ocorrerem exatamente t sucessos. A probabilidade de ocorrerem exatamente t sucessos é

$$\mathrm{b}_{n,p}(t) := \binom{n}{t} p^t (1-p)^{n-t}, \text{ para todo } t \in \{0,1,\ldots,n\}$$

e uma variável aleatória com tal distribuição é dita ter **distribuição binomial** com parâmetros n e p, o que denotamos por $X \in_{\mathrm{b}_p} \{0,1,\ldots,n\}$ ou $X \sim \text{binomial}(n,p)$. A variável com distribuição binomial de parâmetros n e p conta o número de sucessos em n ensaios de Bernoulli idênticos e independentes.

Proposição 3.4 *Seja $k := \lfloor (n+1)p \rfloor$. A função $\mathrm{b}_{n,p}(t)$ é crescente para $t \in \{0,1,\ldots,k\}$ e é decrescente para $t \in \{k+1, k+2, \ldots, n\}$.*

Demonstração. A razão entre valores sucessivos da função de é, para $t > 0$

$$\frac{\mathrm{b}_{n,p}(t)}{\mathrm{b}_{n,p}(t-1)} = \frac{(n-t+1)p}{t(1-p)}$$

de modo que $\mathrm{b}_{n,p}(t)$ é crescente se, e só se, $(n-t+1)p > t(1-p)$, ou seja, $(n+1)p - t > 0$. □

Se lançamos uma moeda honesta $2n$ vezes, o número de caras é uma variável aleatória binomial $X \sim \text{binomial}(2n, 1/2)$. Com que probabilidade ocorrem exatamente n caras? Usando a aproximação de Stirling (veja (d.3))

$$\mathbb{P}[\,X = n\,] = \binom{2n}{n}\left(\frac{1}{2}\right)^{2n} = \frac{(2n)!}{(n!)^2 4^n} = (1+o(1))\frac{1}{\sqrt{\pi n}}$$

que é uma probabilidade bem pequena, para $n = 50$ temos $\approx 0{,}08$ e para $n = 130$ temos $\approx 0{,}05$.

Porém, de fato, quando lançamos a moeda n vezes, com probabilidade que tende a 1 quando $n \to \infty$ o valor de X/n está no intervalo $(1/2-\varepsilon, 1/2+\varepsilon)$ qualquer que seja $\varepsilon \in (0, 1/2)$ pois, por simetria, temos

$$\mathbb{P}\left[\left|\frac{X}{n} - \frac{1}{2}\right| \geqslant \varepsilon\right] = 2\,\mathbb{P}\left[X \geqslant \left(\frac{1}{2}+\varepsilon\right)n\right]$$

além disso, para $k = \lfloor (n+1)/2 \rfloor = \lceil n/2 \rceil$, como na proposição acima,

$$\binom{n}{k}\frac{1}{2^n} \leqslant \mathbb{P}\left[X \geqslant \left(\frac{1}{2}+\varepsilon\right)n \right] \leqslant (n+1)\binom{n}{k}\frac{1}{2^n}.$$

Usando a aproximação de Stirling

$$\binom{n}{k} = \left(\frac{n}{k}\right)^k \left(\frac{n}{n-k}\right)^{n-k} \sqrt{\frac{n}{k(n-k)}} \frac{1+O(1/n)}{(1+O(1/k))(1+O(1/(n-k)))}$$

nos dois últimos termos desse produto (a raiz e a fração com assintóticos, que denotamos $\varepsilon(n)$), tomando logaritmo e dividindo por n o resultado tende a zero quando $n \to \infty$. Agora

$$\frac{1}{n}\log\frac{1}{2^n}\binom{n}{k} = -\log 2 + \frac{k}{n}\log\left(\frac{n}{k}\right) + \frac{n-k}{n}\log\left(\frac{n}{n-k}\right) + \log\varepsilon(n)$$

que, quando $n \to \infty$, converge para

$$-\log 2 - \frac{1}{2}\log\left(\frac{1}{2}\right) - \frac{1}{2}\log\left(\frac{1}{2}\right) + 0 = 0.$$

Esse resultado pode ser generalizado (veja o Exercício 3.76, página 154) e adiante veremos uma demonstração alternativa e mais geral (veja o Exemplo 6.16, página 174). Ele já era conhecido por volta de 1700 por Jacob Bernoulli e é o primeiro resultado do que veio a ser chamado de Lei Fraca dos Grandes Números. No teorema de Jacob Bernoulli quanto maior n, menor é a incerteza sobre o valor de X/n, a frequência relativa do número de ocorrência de um evento em repetições independentes tende, conforme o número de repetições aumenta, à probabilidade da ocorrência do evento.

Distribuição de Poisson Uma variável aleatória de Poisson expressa a probabilidade de ocorrência de um determinado número de eventos num intervalo de tempo fixo sempre que tais eventos ocorram com uma taxa média conhecida e independentemente do tempo desde a última ocorrência.

Há vários exemplos curiosos de fenômenos aleatórios com essa distribuição na literatura. O seguinte exemplo do célebre *Introdução a Probabilidade* escrito por Feller (1968): na segunda guerra mundial a cidade de Londres foi intensamente bombardeada pelos alemães. Para determinar se as bombas tinham um alvo ou foram lançadas aleatoriamente os ingleses dividiram o sul da cidade em pequenas regiões e determinaram a taxa de 0,9323 bombas por região, ao qual o modelo de Poisson se ajustou impressionantemente bem, o que levou-os a acreditar que o bombardeio foi aleatório. Um outro exemplo, agora clássico, vem de William Sealy Gosset[2], um

[2]Publicou artigos sob o pseudônimo de *Student* porque o seu empregador proibiu as publicações por funcionários depois que segredos comerciais foram divulgados.

químico e matemático formado em Oxford e contratado, em 1899, pela famosa cervejaria *Arthur Guinness and Son* em Dublin. Sua tarefa era aperfeiçoar o processo de produção de cerveja. Gosset trabalhou com o modelo de Poisson para a contagem de células de levedura. Outra aplicação curiosa e conhecida desta distribuição é devida a Ladislau Bortkiewicz que em 1898 publicou dados sobre o número de soldados do exército da Prússia mortos por coices de cavalo, tais números seguiam uma distribuição de Poisson.

Uma variável aleatória de Poisson com parâmetro $\lambda > 0$ conta o número de ocorrências de um determinado evento que ocorre a uma taxa λ e cuja distribuição é dada por

$$\mathrm{Po}_\lambda(t) := \frac{\mathrm{e}^{-\lambda}\lambda^t}{t!}, \text{ para todo inteiro } t \geqslant 0.$$

Gosset, citado acima, observou "como a dispersão nas contagens de colonias de levedura foi semelhante ao limite exponencial da distribuição binomial". De fato, a distribuição de Poisson pode ser derivada como um caso limite da distribuição binomial quando o número de ensaios tende ao infinito e a taxa média de ocorrências permanece fixa (veja o enunciado preciso no Exercício 3.80 no final deste capítulo).

Intuitivamente, podemos dizer que se λ é a taxa de ocorrência de um evento num intervalo e tomamos subintervalos suficientemente pequenos, a ponto da probabilidade de um evento ocorrer duas vezes nesse intervalo é insignificante, então a probabilidade de ocorrência do evento em cada subintervalo é λ/n. Agora, assumimos que as ocorrências do evento em todo o intervalo pode ser visto como n ensaios de Bernoulli com parâmetro λ/n. Em n ensaios independentes de Bernoulli com probabilidade de sucesso $p = p(n) = \lambda/n$, para uma constante $\lambda > 0$, a probabilidade de k sucessos é $\mathrm{b}_{n,p}(k) \approx \lambda^k \mathrm{e}^{-\lambda}/k!$ que é conhecido como a *aproximação de Poisson para a distribuição binomial*.

A seguinte estratégia pode ser transformada numa demonstração desse fato: tome o polinômio

$$\sum_{k=0}^{n} \mathrm{b}_{n,p}(k)x^k = (xp + 1 - p)^n = (1 + (x-1)p)^n$$

pelo binômio de Newton. No limite, quando $n \to \infty$, assumindo que $p = p(n)$ é tal que $p \cdot n = \lambda$, temos (usando (s.8))

$$\sum_{k\geqslant 0} \mathrm{b}_{n,p}(k)x^k = \lim_{n\to\infty}\left(1 + \frac{\lambda(x-1)}{n}\right)^n = \mathrm{e}^{\lambda x}\mathrm{e}^{-\lambda} = \sum_{k\geqslant 0}\frac{\lambda^k \mathrm{e}^{-\lambda}}{k!}x^k$$

e, agora, comparamos os coeficientes de cada lado.

3.1.1 DISTRIBUIÇÃO CONJUNTA E INDEPENDÊNCIA

Vetores aleatórios são sequências finitas de variáveis aleatórias nas quais se considera seu comportamento estatístico conjunto. Uma variável aleatória $\mathbf{X} : \Omega \to \mathbb{R}^n$ ($n > 1$), com $\mathbb{P}(\mathbf{X} \in S) = 1$ para algum S enumerável, é chamada de **vetor aleatório discreto**, e usamos a notação $\mathbf{X} = (X_1, \ldots, X_n)$, onde cada coordenada é uma variável aleatória discreta. Reciprocamente, se $X_1, \ldots, X_n$ são variáveis discretas, então $\mathbf{X} = (X_1, \ldots, X_n)$ é um vetor aleatório discreto. Sua distribuição é a medida de probabilidade $\mathbb{P}_{\mathbf{X}} = \mathbb{P}_{(X_1,\ldots,X_n)}$, definida para todo $\mathbf{a} = (a_1, \ldots, a_n)$ por $\mathbb{P}_{\mathbf{X}}(\mathbf{a}) = \mathbb{P}[(X_1, \ldots, X_n) = (a_1, \ldots, a_n)]$. Essa distribuição é chamada de **distribuição conjunta** das variáveis $X_1, \ldots, X_n$ e, em geral, *não* é determinada pelas distribuições individuais $\mathbb{P}_{X_i}$.

As variáveis aleatórias X e Y definidas em Ω com valores em S são **variáveis aleatórias independentes** se o conhecimento do valor de uma delas não altera a probabilidade da outra assumir qualquer valor, isto é, formalmente, 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 $a, b \in \mathbb{R}$.

As variáveis aleatórias $X_1, X_2, \ldots, X_n$ são **independentes** se para eventos $A_1, \ldots, A_n$

$$\mathbb{P}_{\mathbf{X}}(A_1 \times \cdots \times A_n) = \prod_{i=1}^{n} \mathbb{P}_{X_i}(A_i). \tag{3.2}$$

Notemos que qualquer subconjunto $X_{i_1}, \ldots, X_{i_k}$ dessas variáveis também é independente, basta tomar $A_j = X_j(\Omega)$ para todo $j \neq i_1, \ldots, i_k$ na equação (3.2) acima.

Exercício 3.5. Sejam $X_1, \ldots, X_n$ variáveis aleatórias e $S_1, \ldots, S_n$ conjuntos finitos e não vazios. Mostre que são equivalentes

1. $(X_1, \ldots, X_n) \in_R S_1 \times \cdots \times S_n$
2. $X_1, \ldots, X_n$ são independentes e $X_i \in_R S_i$ para cada i.

Funções de variáveis aleatórias Se $X : \Omega \to \mathbb{R}$ é uma variável aleatória e $f\colon \mathbb{R} \to \mathbb{R}$ é uma função real então a função composta $f(X)$ é uma variável aleatória $Y = f(X)$ cuja distribuição é

$$\mathbb{P}_Y(y) = \mathbb{P}[\, f(X) = y \,] = \sum_{x\colon f(x)=y} \mathbb{P}_X(x).$$

Por exemplo, se X é uma variável aleatória de $(\Omega, \mathbb{P})$, então podemos definir $Z\colon \Omega \to \mathbb{R}$ por $Z(\omega) := X(\omega)^2$ para todo $\omega \in \Omega$. A função $Z = X^2$ também é uma variável aleatória e para todo t não negativo temos $[Z \leqslant t] = [-\sqrt{t} \leqslant X \leqslant \sqrt{t}]$. Se $a, b \in \mathbb{R}$, com $a \neq 0$, então $Y\colon \Omega \to \mathbb{R}$ dada por $Y(\omega) = a \cdot X(\omega) + b$ é uma variável aleatória tal que, para todo real t, $[Y \leqslant t] = [X \leqslant (t-b)/a]$.

Se $(X_1, X_2, \ldots, X_n)$ é um vetor aleatório de um espaço de probabilidade e $f\colon \mathbb{R}^n \to \mathbb{R}$ é função então $f(X_1, X_2, \ldots, X_n)$ é uma variável aleatória. Em particular, se X e Y são variáveis aleatórias definidas em $(\Omega, \mathbb{P})$, a soma é a variável aleatória $X + Y\colon \Omega \to \mathbb{R}$ dada por $\{X+Y\}(\omega) = X(\omega) + Y(\omega)$ e o produto é a variável aleatória $X \cdot Y\colon \Omega \to \mathbb{R}$ dada por $\{X \cdot Y\}(\omega) = X(\omega) \cdot Y(\omega)$. A distribuição de $Z = X + Y$ é

$$\mathbb{P}_Z(z) = \mathbb{P}\Big(Z^{-1}(z)\Big) = \sum_{x \in X(\Omega)} \mathbb{P}\Big(X^{-1}(x) \cap Y^{-1}(z-x)\Big) = \sum_{x \in X(\Omega)} \mathbb{P}_{(X,Y)}((x, z-x)).$$

No caso particular em que X e Y são independentes

$$\mathbb{P}_{X+Y}(z) = \sum_{x \in X(\Omega)} \mathbb{P}_X(x)\mathbb{P}_Y(z-x).$$

Naturalmente, podemos considerar a soma e o produto para $n > 2$ variáveis, denotadas $\sum_{i=1}^n X_i$ e $\prod_{i=1}^n X_i$ respectivamente. Por exemplo, sejam $X_1, \ldots, X_n$ variáveis aleatórias independentes e com distribuição Bernoulli com parâmetro p. Então

$$X = \sum_{i=1}^{n} X_i$$

é uma variável aleatória que conta a quantidade de sucessos nos n ensaios cuja distribuição é

$$\binom{n}{t} p^t (1-p)^{n-t} = b_{n,p}(t)$$

para todo $t \in \{0, 1, \ldots, n\}$.

Também, se $\max\colon \mathbb{R}^n \to \mathbb{R}$ é a função que calcula o maior dentre n valores reais, então temos que $\max(X_1, X_2, \ldots, X_n)$ é uma variável aleatória.

Proposição 3.6 *Sejam X e Y variáveis aleatórias independentes e f e g funções de $\mathbb{R}$ em $\mathbb{R}$. Então as variáveis aleatórias $f(X)$ e $g(Y)$ são independentes.*

Demonstração. Se X e f são como no enunciado e $a \in \mathbb{R}$, então $[f(X) = a] = [X \in A]$ em que $A := \{x \in X(\Omega)\colon f(x) = a\}$. Analogamente, para Y, g e $b \in \mathbb{R}$ dados, temos

$[g(Y) = b] = [Y \in B]$ em que $B := \{x \in X(\Omega)\colon\ g(x) = b\}$. De X e Y independentes temos $\mathbb{P}([X \in A] \cap [Y \in B]) = \mathbb{P}[X \in A] \cdot \mathbb{P}[Y \in B]$, portanto, $\mathbb{P}([f(X) = a] \cap [g(Y) = b]) = \mathbb{P}[f(X) = a] \cdot \mathbb{P}[g(Y) = b]$, ou seja, $f(X)$ e $g(Y)$ são variáveis aleatórias independentes. □

Exercício 3.7. Determine a distribuição da soma de duas variáveis de Poisson e a distribuição da soma de duas variáveis binomiais com mesmo parâmetro p.

3.1.2 VALOR ESPERADO DE UMA VARIÁVEL ALEATÓRIA SIMPLES

Uma função X é chamada de *função simples* se pode ser escrita como combinação linear de funções indicadoras $X = \sum_{i=1}^{k} c_i \mathbb{1}_{A_i}$, onde os conjuntos A_i particionam o domínio da função. Podem haver muitas maneiras de escrever X como uma combinação linear de funções indicadoras, vamos descrever uma forma canônica.

Se X é combinação linear de funções indicadoras, como descrito acima, sua imagem é um conjunto finito (e vice-versa). Tomamos no domínio de X a partição finita $B_1, \ldots, B_m$ dada pela relação de equivalência $\omega_1 \equiv \omega_2$ 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

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

Uma variável aleatória de $(\Omega, \mathbb{P})$ com imagem finita é chamada de **variável aleatória simples**. O **valor médio** ou **valor esperado** ou, ainda, **esperança** da variável aleatória simples X é

$$\mathbb{E}\, X = \mathbb{E}\left[\sum_{i=1}^{m} x_i \mathbb{1}_{B_i}\right] := \sum_{i=1}^{m} x_i \mathbb{P}(B_i) \tag{3.3}$$

que é média dos valores de $X(\Omega) = \{x_1, x_2, \ldots, x_m\}$ ponderada pela probabilidade de cada valor, ou seja, B_i é 0 evento $[X = x_i]$ e

$$\mathbb{E}\, X = \sum_{i=1}^{m} x_i \mathbb{P}_X(x_i). \tag{3.4}$$

Exemplo 3.8. Se A é evento de um espaço de probabilidade então $\mathbb{E}\, \mathbb{1}_A = \mathbb{P}(A)$. ◊

Exemplo 3.9. Consideremos um jogo de azar no qual em cada aposta paga-se R$10,00 e, ou ganhamos R$1.000.010,00 com probabilidade $p \in (0,1)$ ou perdemos R$10,00, que já foram pagos, com probabilidade $1-p$. Se Y é o valor ganho numa aposta, então a esperança de ganho numa aposta é $\mathbb{E}\, Y = 10^6 p - 10(1-p)$. No caso de $p = 1/2$, o prêmios são equiprováveis e o ganho médio é $\mathbb{E}\, Y = 499.995{,}00$. A probabilidade de

ganharmos R$499.995,00 numa aposta é zero. Se $p = 1/100$ então a probabilidade de ganhar o valor alto é muito pequeno quando comparado com a probabilidade de perder 10 reais e o valor esperado de ganho numa única aposta é $\mathbb{E}\, Y = 9.990{,}10$. $\Diamond$

O valor esperado de variáveis aleatórias simples não depende de como escrevemos X como combinação linear de variáveis indicadoras. De fato, seja $\{A_k \colon 1 \leqslant k \leqslant n\}$ uma 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

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

logo

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

portanto

$$\sum_{k=1}^{n} c_k \mathbb{P}(A_k) = \sum_{\ell=1}^{m} x_\ell \mathbb{P}(B_\ell). \tag{3.5}$$

Disso, observamos que se X é simples então $X = \sum_\omega X(\omega)\mathbb{1}_{\{\omega\}}$, portanto,

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

No teorema abaixo daremos algumas propriedades importantes do valor esperado. Na demonstração desses resultados usaremos o seguinte exercício cuja verificação é simples.

Exercício 3.10. 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 (dica: Exercício 3.52, item 4, página 150)

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.

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

Teorema 3.11 (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. *Para todo evento* A, $\mathbb{E}\, \mathbb{1}_A = \mathbb{P}(A)$.

3. Linearidade: *se* Y *é uma variável aleatória simples e* a *e* b *números reais então*

$$\mathbb{E}[\, aX + bY\,] = a\mathbb{E}\, X + b\mathbb{E}\, Y. \tag{3.6}$$

4. 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*

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

5. *Se* f *é uma função real e* $X(\Omega) = \{x_1, x_2, x_3, \ldots, x_n\}$ *então*

$$\mathbb{E}[\, f(X)\,] = \sum_{k=1}^{n} f(x_k)\mathbb{P}_X(x_k).$$

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

Demonstração. As demonstrações dos itens 1 e 2 são imediatas das equações (3.4) e (3.3), respectivamente. Para provarmos os itens 3, 4 e 6 consideramos X e Y tais que $X(\Omega) = \{x_1, \ldots, x_n\}$ e $Y(\Omega) = \{y_1, \ldots, y_m\}$, também as partições $A_k = \{\omega\colon X(\omega) = x_k\}$ e $B_\ell = \{\omega\colon Y(\omega) = y_\ell\}$ de Ω, de modo que $X = \sum_{k=1}^{n} x_k \mathbb{1}_{A_k}$ e $Y = \sum_{\ell=1}^{m} y_\ell \mathbb{1}_{B_\ell}$.

Sejam a e b reais arbitrários. Então, usando a definição (3.4), o item 1 do Exercício 3.10 acima e o item 2 deste teorema, temos

$$\mathbb{E}[\, aX + bY\,] = \sum_{k=1}^{n}\sum_{\ell=1}^{m} (ax_k + by_\ell)\mathbb{P}(A_k \cap B_\ell) = \sum_{k=1}^{n}\sum_{\ell=1}^{m} ax_k\mathbb{P}(A_k \cap B_\ell) + by_\ell\mathbb{P}(A_k \cap B_\ell)$$

e rearranjando as somas deduzimos

$$\begin{aligned}\mathbb{E}[\, aX + bY\,] &= \sum_{k=1}^{n}\sum_{\ell=1}^{m} ax_k\mathbb{P}(A_k \cap B_\ell) + \sum_{\ell=1}^{m}\sum_{k=1}^{n} by_\ell\mathbb{P}(A_k \cap B_\ell) \\ &= \sum_{k=1}^{n} ax_k\mathbb{P}(A_k) + \sum_{\ell=1}^{m} by_\ell\mathbb{P}(B_\ell),\end{aligned}$$

a segunda igualdade segue do fato das famílias de eventos $\{A_1, \ldots, A_n\}$ e $\{B_1, \ldots, B_m\}$ formarem, cada uma, uma partição do espaço amostral, portanto, $\mathbb{E}[\, aX + bY\,] = a\mathbb{E}\, X + b\mathbb{E}\, Y$.

Se $X \leqslant Y$ então $Y - X \geqslant 0$, portanto $\mathbb{E}[\, Y - X\,] \geqslant 0$. Usando a linearidade $\mathbb{E}[\, Y - X\,] = \mathbb{E}\, Y - \mathbb{E}\, X \geqslant 0$, donde deduzimos que $\mathbb{E}\, X \leqslant \mathbb{E}\, Y$.

Se X e Y são independentes então, usando o item 2 do Exercício 3.10 acima,

$$\begin{aligned}\mathbb{E}[XY] &= \sum_{k=1}^{n}\sum_{\ell=1}^{m} x_k y_\ell\mathbb{P}(A_k \cap B_\ell) = \sum_{k=1}^{n}\sum_{\ell=1}^{m} x_k y_\ell\mathbb{P}(A_k)\mathbb{P}(B_\ell) = \sum_{k=1}^{n} x_k\mathbb{P}(A_k)\sum_{\ell=1}^{m} y_\ell\mathbb{P}(B_\ell) \\ &= \mathbb{E}\, X \cdot \mathbb{E}\, Y\end{aligned}$$

o que prova o item 6.

Para o item 5 façamos $Y := f(X)$ de modo que

$$f(X) = \sum_{\ell=1}^{m} y_\ell \, \mathbb{1}_{B_\ell}$$

com $B_\ell = [Y = y_\ell]$. Agora, notemos que para cada $k \in \{1, 2, \ldots, n\}$ existe um único $\ell \in \{1, 2, \ldots, m\}$ tal que $A_k \subseteq B_\ell$, a saber o ℓ tal que $f(x_k) = y_\ell$. Seja $I_\ell := \{k \colon 1 \leqslant k \leqslant n,\ f(x_k) = y_\ell\}$. De fato, temos (verifique)

$$B_\ell = \bigcup_{k \in I_\ell} A_k$$

sendo a união de conjuntos disjuntos, assim

$$f(X) = \sum_{\ell=1}^{m} y_\ell \, \mathbb{1}_{B_\ell} = \sum_{\ell=1}^{m} \sum_{k \in I_\ell} y_\ell \, \mathbb{1}_{A_k} = \sum_{\ell=1}^{m} \sum_{k \in I_\ell} f(x_k) \, \mathbb{1}_{A_k} = \sum_{k=1}^{n} f(x_k) \, \mathbb{1}_{A_k}$$

e pela equação (3.5) temos de

$$f(X) = \sum_{\ell=1}^{m} y_\ell \, \mathbb{1}_{B_\ell} = \sum_{k=1}^{n} f(x_k) \, \mathbb{1}_{A_k}$$

que $\mathbb{E}\, f(X) = \sum_{k=1}^{n} f(x_k) \mathbb{P}(A_k) = \sum_{k=1}^{n} f(x_k) \, \mathbb{P}_X(x_k)$. □

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 3.12 *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 (3.6) usando indução em n. □

Exemplo 3.13 (*esperança das distribuições Bernoulli e binomial*). Se X tem distribuição de Bernoulli então $\mathbb{E}\, X = \mathbb{P}[X = 1] = p$ em que p é a probabilidade de sucesso.

Se $X_1, \ldots, X_n$ são variáveis com distribuição de Bernoulli de parâmetro p, independentes e $X = \sum_i X_i$, então X tem distribuição binomial com parâmetros n e p como vimos acima. Ademais, pela linearidade da esperança, corolário acima, temos

$$\mathbb{E}\, X = \sum_{i=1}^{n} \mathbb{E}\, X_i = np$$

pois $\mathbb{E}\, X_i = p$ para todo i. ◊

Exemplo 3.14. Se $X \in_R \{1,2,3,4,5,6\}$ 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 $Y \in_R S$ 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 no lançamento de dois dados? Pela linearidade da esperança é 7. Alternativamente, a distribuição da soma é mostrada na Tabela 3.1, donde podemos calcular o valor esperado para o produto dos resultados,

$X \cdot Y$	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 3.1: distribuição do produto de dois dados.

$\mathbb{E}[X \cdot Y] = 49/4$. ◊

3.1.3 TABELAS DE ESPALHAMENTO

Em computação um conjunto que é modificado com passar do tempo é, usualmente, chamado de **conjunto dinâmico** e a estrutura de dado que o representa deve contemplar as operações elementares de inserir elementos, remover elementos e testar pertinência.

Uma **tabela de espalhamento**, ou **tabela de 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 N é um vetor $N[i]$ de listas ligadas, indexadas por $M = 0, 1, \ldots, m-1$, e o acesso à tabela ocorre por meio de uma função de *hash* $h : U \to M$. Dado $x \in U$, a inserção, remoção ou busca de x em N é realizada na lista ligada $N[h(x)]$ (veja uma ilustração na Figura 3.1).

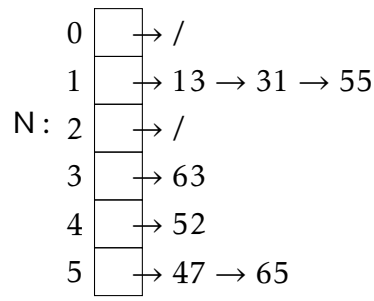

Figura 3.1: tabela de espalhamento com os elementos $\{13, 31, 47, 52, 55, 63, 65\}$ distribuídos de acordo com a função $h(x) = x \bmod 6$.

As operações de busca, inserção e remoção de um elemento numa tabela N que representa um conjunto dinâmico $S \subset U$ com função de *hash* $h\colon U \to M$ 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 a pergunta "$x \in S$?" e ainda, no caso positivo, retorna um apontador para x. Numa tabela de espalhamento isso é resolvido por uma busca sequencial na lista ligada $N[h(x)]$. Uma busca por x em N é dita *com sucesso* caso $x \in N$, 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 colocar o elemento x no fim da lista ligada $N[h(x)]$ e o custo de pior caso dessa operação é constante;

remoção: dado $x \in S$, a remoção de x retira esse elemento da estrutura que representa 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 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 $N[i]$ 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 $n := |S|$, é uma fração pequena de U e o caso interessante é quando $m = O(n)$.

É desejável que os elementos de S fiquem bem espelhados na tabela, pois *colisões* (elementos distintos mapeados para a mesma lista) afetam o desempenho das operações. Como $|U| > m$ não há como evitar colisões. Mais que isso, se o conjunto universo U é suficientemente grande, digamos $|U| > (n-1)m$, então para qualquer função de *hash* $h \in M^U$ sempre haverá um conjunto S de cardinalidade n para o qual uma configuração de pior caso é inevitável.

Das funções de *hashing*, além do bom espalhamento, queremos a função seja fácil de computar e que a representação seja sucinta, isto é, relativamente poucos bits são 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 M^{U} de todas as $|\mathsf{M}|^{|\mathsf{U}|}$ funções de U em M tem, com alta probabilidade, a propriedade de não formar listas com $\mathsf{O}(\log n)$ elementos (veja o Exercício 1.56, página 48), isso garante um pior caso equivalente ao das árvores de buscas, com a vantagem de manutenção mais simples. Além disso, o tempo médio de busca é $\approx n/m = \mathsf{O}(1)$, quando $m = \mathsf{O}(n)$.

Um problema nessa estratégia é que tal função não tem uma representação sucinta. Algumas delas requerem pelo menos $|\mathsf{U}|\log(m)$ 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 $(\mathsf{M}^{\mathsf{U}}, \mathbb{P})$ com $\mathbb{P}(h) = 1/|\mathsf{M}^{\mathsf{U}}|$ para toda função $h \in \mathsf{M}^{\mathsf{U}}$. Notemos que vale

$$\mathbb{P}[\, h(x) = i \,] = \frac{1}{|\mathsf{M}|} \tag{3.7}$$

para todo $x \in \mathsf{U}$ e todo $i \in \mathsf{M}$. De fato, sortear uniformemente uma função é equivalente a sortear uniformemente e independentemente uma imagem para cada elemento do domínio (veja o Exercício 3.5, página 102).

Fixemos os parâmetros $n = |\mathsf{S}|$ e $m = |\mathsf{M}|$. A fração

$$c = c(n, m) := \frac{n}{m} \tag{3.8}$$

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 (3.7). Tomemos

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

a variável aleatória indicadora de colisão dos elementos $x, y \in \mathsf{U}$ sob a escolha $h \in \mathsf{M}^{\mathsf{U}}$. A probabilidade de colisão sempre que $x \neq y$ é

$$\mathbb{P}[\, h(x) = h(y) \,] = \frac{1}{m}$$

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

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

portanto o tamanho esperado da lista $\mathsf{N}[h(x)]$ é, pela linearidade da esperança

$$\sum_{y\in\mathsf{S}} \mathbb{E}\, \mathbb{1}_{[h(x)=h(y)]} = \begin{cases} \frac{n}{m}, & \text{se } x \notin \mathsf{S}, \\ \frac{(n-1)}{m} + 1, & \text{caso contrário.} \end{cases} \tag{3.9}$$

O tempo esperado de uma busca em uma tabela de espalhamento com uma função *hash* h escolhida aleatoriamente em M^{U} é proporcional à carga c da tabela.

Proposição 3.15 *O tempo esperado para uma busca sem sucesso é $c+1$ e o tempo esperado para uma busca com sucesso é $c/2 - 1/(2m) + 1$.*

Demonstração. Consideremos uma busca por $x \in \mathsf{U}$ na tabela de espalhamento N. Se $x \notin \mathsf{S}$ então, pelo primeiro caso da equação (3.9), são necessárias $c+1$ comparações em média numa busca sem sucesso. Agora, suponhamos $x \in \mathsf{S}$ e que os elementos de S foram inseridos sequencialmente. Se x foi o i-ésimo item inserido em N, então o tamanho esperado da lista imediatamente após a inserção é $(i-1)/m+1$ pela equação (3.9), portanto, o tempo médio da busca por x é

$$\frac{1}{n}\sum_{i=1}^{n} \frac{i-1}{m} + 1 = \frac{c}{2} - \frac{1}{2m} + 1$$

que é o número médio de comparações numa busca com sucesso. □

Essa análise não nos dá nenhuma pista sobre o tempo esperado para uma busca no pior caso. No caso $m = \mathsf{O}(n)$ a proposição acima garante que o tempo médio de busca é constante, entretanto, provaremos na seção 6.1.2, página 168, que a maior lista tem $\Theta(\log(n)/\log(\log(n)))$ elementos.

Para S fixo o número esperado de colisões é

$$\binom{n}{2}\frac{1}{m} = \frac{n(n-1)}{2m}$$

assim, se $m = n^2$ então $\mathbb{E}\,\mathsf{C} < 1/2$ e, de fato, não há colisão com probabilidade pelo menos $1/2$ pois

$$\mathbb{P}[\,\mathsf{C} \geqslant 1\,] = \sum_{k=1}^{\binom{n}{2}} \mathbb{P}[\,\mathsf{C} = k\,] \leqslant \sum_{k=0}^{\binom{n}{2}} k\mathbb{P}[\,\mathsf{C} = k\,] = \mathbb{E}\,\mathsf{C} < \frac{1}{2}. \tag{3.10}$$

Exemplo 3.16 (*paradoxo dos aniversários*). Para $n \leqslant m$ há $m(m-1)(m-2)\cdots(m-n+1)$ sequências sem repetições em M^n. Se $(h(x_1), h(x_2), \ldots, h(x_n))$ é uma escolha aleatória

em M^n e C é o número de colisões, então nessa escolha

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

na segunda linha usamos que $\exp(-x) > 1-x$ (veja (d.1)). O conhecido *paradoxo dos aniversários* é o caso $n = 23$ e $m = 365$ 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 $n \leqslant (3/4)m$ temos o seguinte limitante para a probabilidade de não ocorrer colisão

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

de modo que se n é aproximadamente $\sqrt{m}$ então a probabilidade de não haver colisão é maior que $\exp(-1) \approx 0{,}36$, logo com boa probabilidade uma função aleatória de espalhamento consegue espalhar $\sqrt{m}$ itens até que ocorra a primeira colisão (veja o Exemplo 6.17, página 174). ◊

Hashing universal Idealmente, o que procuramos para funções de espalhamento são funções que tenham representação sucinta e imitam uma função aleatória nas estatísticas de interesse, como tempo constante de busca, não formar listas longas, e também que possam ser computadas eficientemente em cada ponto do domínio.

A ideia principal para resolver o problema de armazenamento de funções é reduzir a aleatoriedade em h. Formalmente, em vez de selecionar h de maneira aleatória em M^U, sorteamos em uma coleção menor de funções, $\mathcal{H} \subseteq M^U$. Em vez de exigirmos que a equação (3.7) seja válida para todo x e todo i, pedimos que seja válida a condição

$$\mathbb{P}_{h\in_R\mathcal{H}}[h(x) = h(y)] \leqslant \frac{1}{m}, \tag{3.11}$$

para todo $x \neq y$. Assim, ainda será verdade que o tamanho médio de uma lista é limitado pela equação (3.8), que a sentença enunciada na Proposição 3.15 vale, que o número esperado de colisões é $O(n^2/2m)$ e no caso $m = n^2$ não há colisão com probabilidade pelo menos 1/2 (verifique).

Uma família $\mathcal{H} \subseteq M^U$ de funções que satisfazem a equação (3.11) é chamada de *universal*.

Teorema 3.17 *Se $\mathcal{H}$ é universal, então para qualquer conjunto $S \subseteq U$ de cardinalidade n, para qualquer $x \in U$, se h é escolhido aleatoriamente em $\mathcal{H}$, então número esperado de colisões entre x e outros elementos de S é no máximo n/m.* □

Como não precisamos de funções genuinamente aleatórias para termos as boas propriedades estatísticas dessas funções, vamos procurar por uma família $\mathcal{H}$ universal e não vazia de funções *hashing* com representações sucintas e que são calculadas eficientemente. Por exemplo, a família $\{f, g, h\}$ de funções de $\{1, 2\}$ em $\{0, 1\}$ dadas na Tabela 3.2 é universal (verifique).

	1	2
f	0	0
g	1	0
h	0	1

Tabela 3.2: exemplo de uma família universal.

Para $U := \{0,1\}^u$ e $M := \{0,1\}^m$ podemos definir uma função linear $h\colon U \to M$ sorteando os bits de uma matriz $A = (a_{\ell,c})$ de dimensão $m \times u$ e fazendo $h(x) := A \cdot x$ para todo $x \in U$, com *as operações módulo* 2. Para armazenar a função precisamos apenas da matriz A de tamanho $\log|M| \cdot \log|U|$.

Nesse caso, observamos que a equação (3.7) não vale para todo x e todo i, pois $h(0) = 0$ para qualquer escolha de A. Porém, para elementos distintos $x, y \in U$ teremos $h(x) = h(y)$ se, e somente se, $h(x-y) = A \cdot (x-y) = 0$, pois h é uma função linear, de modo que

$$\sum_{c=1}^{u} a_{\ell,c}(x-y)_c = 0 \text{ para cada } \ell \in \{1,\ldots,m\}.$$

Seja i uma coordenada não nula de $x-y$, digamos que $x_i = 0$ e $y_i = 1$. Se sorteamos todos os bits de A exceto os da coluna i então há uma única escolha para cada bit da coluna i para que a equação acima se verifique, a saber, módulo 2 deve valer

$$a_{\ell,i} = \sum_{\substack{c=1 \\ c \neq i}}^{u} a_{\ell,c}(x-y)_c$$

para cada $\ell \in \{1,\ldots,m\}$, o que ocorre com probabilidade $1/2^m$, ou seja,

$$\mathop{\mathbb{P}}_{A \in_R \{0,1\}^{m \times u}}[A \cdot x = A \cdot y] \leqslant \frac{1}{|M|}.$$

Exemplo 3.18 (Carter e Wegman, 1979). Definimos uma família de funções com domínio $U = \{0, 1, \ldots, N-1\}$ e contradomínio $M = \{0, 1, \ldots, m-1\}$, com $N > m$ e

$N < p \leqslant 2N$ um primo fixo, por

$$\mathcal{H}_{p,m} \coloneqq \left\{ h_{(a,b)} \colon a, b \in \{0, 1, \ldots, p-1\},\ a \neq 0 \right\}$$

com $h_{(a,b)}(x) = (ax + b \bmod p) \bmod m$.

Se $x \neq y$ então $ax + b \bmod p \neq ay + b \bmod p$, pois p é primo. Ademais, para $i, j \in \{0, 1, \ldots, p-1\}$ distintos existe exatamente um par (a, b) tal que

$$\begin{cases} ax + b \bmod p = i \\ ay + b \bmod p = j \end{cases}$$

pois, do sistema acima na forma matricial,

$$\begin{pmatrix} x & 1 \\ y & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} i \\ j \end{pmatrix}$$

temos uma matriz quadrada com determinante diferente de 0 ($x \neq y$). São $p(p-1)$ pares (i, j) distintos e o mesmo tanto de pares (a, b) com $a \neq 0$, logo há uma bijeção entre esses pares. Disso tiramos que, para $x \neq y$ dados, se sorteamos (a, b) em $\{1, 2, \ldots, p-1\} \times \{0, 1, \ldots, p-1\}$ o par (i, j) resultante é qualquer par de valores distintos módulo p com a mesma probabilidade. Resta contarmos a quantidade de pares (i, j) de números distintos côngruos módulo m.

Para cada $i \in \{0, 1, \ldots, p-1\}$, a quantidade de múltiplos de m dentre os p inteiros consecutivos $\{-i, -i+1, \ldots, -i+p-1\}$ é no máximo $(p-1)/m + 1$. Assim, para $x \neq y$, uma escolha aleatória de função nos dá

$$\mathbb{P}\Big[h_{(a,b)}(x) = h_{(a,b)}(y) \Big] \leqslant \frac{p(p-1)/m}{p(p-1)} = \frac{1}{m}.$$

Qualquer membro dessa família pode ser representado unicamente pela tripla (a, b, p) de modo que cada função precisa de no máximo $3 \log(2|U|)$ bits. Além disso, a função contém um número constante de operações que levam tempo constante para avaliar, logo a função pode ser calculada em tempo constante ◊

Exercício 3.19. Para p primo, prove que as p^{r+1} funções de $U = \{0, 1, \ldots, p-1\}^{r+1}$ em $\{0, 1, \ldots, p-1\}$ dadas por

$$h_a(x) = \langle a, x \rangle \pmod p = \sum_{i=0}^{r} a_i x_i \pmod p$$

é universal (sortear uma função equivale a sortear a em U).

Hashing perfeito Dizemos que uma função *hash* é **perfeita**, para S fixo, se as buscas custam $O(1)$ no pior caso.

Se $\mathcal{H}$ é universal e $m = n^2$ então, como já discutimos acima, escolhendo h aleatória em $\mathcal{H}$ há pelo menos 50% de chance de não haver colisões (equação (3.10)). É possível descobrir essa função em tempo $O(n^2)$ supondo que $h(x)$ é calculado em tempo $O(1)$:

1. sorteie h;
2. teste os $O(n^2)$ pares para colisão;
3. repita os passos 1 e 2 até encontrar h perfeita.

Como a probabilidade de encontrar tal h é pelo menos $1/2$, em média 2 sorteios serão suficientes (a condição de parada é uma variável aleatória geométrica). Dessa forma, temos um tabela de *hashing* para um conjunto fixo de tamanho n, com tempo de busca $O(1)$ no pior caso e com espaço $O(n^2)$.

É possível, nesse contexto, tempo de busca constante com espaço linear e o método é o seguinte: primeiro fazemos o espalhamento em uma tabela de tamanho $m = n$ usando *hashing* universal (digamos $(ax+b \bmod p) \bmod m$), que poderá resultar em colisões. Se n_i elementos vão parar na posição i, para cada i, eles serão distribuídos em uma tabela de espalhamento secundária, com n_i^2 posições (digamos, por $h_i(x) = (a_i x + b_i \bmod p) \bmod n_i^2$) como fizemos no parágrafo anterior.

Resumindo, temos uma função de *hash* de primeiro nível h e uma tabela de primeiro nível N e, então, n funções de *hash* de segundo nível $h_1, \ldots, h_n$ e n tabelas de segundo nível $N_1, \ldots, N_n$. Para buscar um elemento x de S em N, primeiro calculamos $i = h(x)$ e depois procuramos o elemento em $N_i[h_i(x)]$.

O espaço total usado no segundo nível é $\sum_i n_i^2$ e $n_i^2 = n_i + n_i(n_i - 1) = n_i + 2\binom{n_i}{2}$. O valor esperado dessa soma é

$$\mathbb{E}\left[\sum_{i=1}^{m} n_i + 2\binom{n_i}{2}\right] = \mathbb{E}\left[\sum_{i=1}^{m} n_i\right] + \mathbb{E}\left[\sum_{i=1}^{m} 2\binom{n_i}{2}\right] = n + \mathbb{E}\left[\sum_{i=1}^{m} 2\binom{n_i}{2}\right]$$

e $\binom{n_i}{2}$ conta o número de colisões em $N_i[h_i(x)]$, portanto, $\mathbb{E}\left[\sum_{i=1}^{m} n_i^2\right] \leqslant n + \frac{n(n-1)}{m} < 2n$. Agora, podemos limitar a probabilidade de ocorrer $[n_i^2 > 4n]$

$$4n\,\mathbb{P}\left[\sum_i n_i^2 > 4n\right] < \sum_{x>4n} 4n\,\mathbb{P}\left[\sum_i n_i^2 = x\right] < \mathbb{E}\left[\sum_i n_i^2\right]$$

logo $\mathbb{P}\left[\sum_i n_i^2 > 4n\right] < 1/2$. Com probabilidade pelo menos $1/2$ a soma do tamanho das tabelas secundárias é linear, no máximo $4n$.

Para finalizar, usamos a estratégia do primeiro parágrafo dessa seção para encontrar h em uma família universal tal que $\sum_i n_i^2 \leqslant 4n$. Sorteios consecutivos até encontrarmos uma tal h toma, em média, tempo $O(n^2)$. Feito isso, fixamos tal h e encontramos as n funções *hash* secundárias h_i em tempo médio $O(n_i^2)$, para cada i, também como no começo dessa seção (veja o exercício 3.67, página 152).

3.2 ESPERANÇA MATEMÁTICA

Se X é uma variável aleatória que assume valores em um subconjunto enumerável de $\mathbb{R}$, então

$$X = \sum_{n=1}^{\infty} x_n \mathbb{1}_{B_n}$$

para alguma sequência $(x_n : n \geqslant 1)$ de números reais e $\{B_n : n \geqslant 1\}$ partição de Ω. Para escrever uma representação, tomamos uma enumeração qualquer $x_1, x_2, x_3, \ldots$ da imagem de X e definimos uma partição do espaço amostral definida pelos eventos formados pela pré-imagem desses valores, isto é, $B_n = X^{-1}(\{x_n\})$ para todo $n \geqslant 1$. Agora, se estendemos de modo natural a definição de valor esperado de uma variável simples dada na equação (3.4), página 104 temos

$$\mathbb{E}\, X = \sum_{n=1}^{\infty} x_n \mathbb{P}(B_n) = \sum_{n=1}^{\infty} x_n \mathbb{P}_X(x_n) \tag{3.12}$$

entretanto esse valor, caso exista já que é um limite, não deve depender da enumeração particular $x_1, x_2, \ldots$ da imagem de X.

A **esperança matemática**, ou **valor médio** ou **valor esperado**, da variável aleatória discreta e real $X : \Omega \to S$ é

$$\mathbb{E}\, X = \sum_{\substack{t>0 \\ t\in S}} t\, \mathbb{P}_X(t) - \sum_{\substack{t<0 \\ t\in S}} |t|\, \mathbb{P}_X(t)$$

se ao menos uma das séries é finita; caso contrário, ela *não está definida*. Quando as duas séries são finitas, dizemos que a variável aleatória X tem **esperança finita** ou dizemos que X é **integrável**.

Se X é integrável, a esperança é um número real e a série converge absolutamente, isto é, $\mathbb{E}\,|X| = \sum_{t\in S} |t|\, \mathbb{P}_X(t)$ converge, e escrevemos $\mathbb{E}\,|X| < +\infty$. Uma variável aleatória X é integrável se, e somente se, $\mathbb{E}\,|X| < +\infty$.

Nos casos em que o valor esperado da variável aleatória X está definido (convergente ou não), podemos reordenar os termos da soma de modo que podemos escrever

$$\mathbb{E}\, X = \sum_{t\in S} t\, \mathbb{P}_X(t) = \sum_{\omega\in\Omega} X(\omega)\mathbb{P}(\omega). \tag{3.13}$$

A esperança de uma variável aleatória depende apenas da sua distribuição, ou seja, se X e Y têm a mesma distribuição e $\mathbb{E}\, X$ está definida, então $\mathbb{E}\, Y = \mathbb{E}\, X$.

Observação 3.20 (*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 série $\sum_n x_n$ pode ser *finita*, isto é convergir para um número real, ou *infinita* (as somas parciais tende ao infinito) o

que denotamos por $\sum_n x_n = +\infty$. Em ambos os casos dizemos que a série é *bem definida*. No caso geral, definimos

$$X^+ := \sum_{n:\, x_n \geqslant 0} x_n \quad \text{e} \quad X^- := \sum_{n:\, x_n < 0} |x_n|$$

e pode acontecer de

- se ambas as séries são finitas então $\sum_n x_n = X^+ - X^-$ e a série *está bem definida*. Além disso, nesse caso a série $\sum_n x_n$ é *absolutamente convergente*, isto é, $\sum_{n\geqslant 1} |x_n|$ converge.
- Se $X^+ = +\infty$ e X^- é finita, então $\sum_n x_n := +\infty$ e, analogamente, se $X^- = +\infty$ e S^+ é finita, então $\sum_n x_n := -\infty$. Em ambos os casos a série $\sum_n x_n$ não é absolutamente convergente, porém *está bem definida* e é infinita.
- Se $X^+ = X^- = +\infty$, então a série $\sum_n x_n$ é *indefinida*.

A propriedade importante para nós é que *sempre que a série $\sum_{n\geqslant 1} x_n$ está bem definida uma permutação π qualquer na ordem dos termos da sequência não altera resultado*, isto é, $\sum_{n\geqslant 1} x_n = \sum_{n\geqslant 1} x_{\pi(n)}$. Isso não vale no caso indefinido, $\sum_n x_{\pi(n)}$ pode dar resultados diferentes para diferentes permutações π (veja equação (3.14) abaixo). Em vista disso, se a série na equação (6.2) está bem definida então *o seu valor não depende da enumeração de* $X(\Omega)$. ◊

Exemplo 3.21. Uma série conhecida pela propriedade de convergir não absolutamente é a série harmônica alternada

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

cuja convergência decorre da série de Taylor para o logaritmo natural. A série formada pelos valores absolutos é a série harmônica, que não converge. O seguinte rearranjo dessa série devido ao matemático Pierre Alphonse Laurent converge para um valor diferente

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

Com isso podemos moldar uma variável aleatória sem valor esperado. Tomemos uma variável aleatória X que assume os valores $x_1, x_2, \ldots$ dados por $x_i := (-1)^{i+1}/i$ para $i = 1, 2, \ldots$ com $\mathbb{P}_X(x_i) = 6/(\pi i)^2$. Aqui usamos que $\sum_{i\geqslant 1} 1/i^2 = \pi^2/6$ (veja (s.7)) para garantir que $\mathbb{P}_X$ seja distribuição. Com essas definições

$$\sum_{i=1}^{\infty} x_{\pi(i)}\, \mathbb{P}_X\big(x_{\pi(i)}\big) = \begin{cases} \frac{6}{\pi^2}\ln(2) & \text{se } \pi \text{ é a permutação identidade} \\ \frac{3}{\pi^2}\ln(2) & \text{se } \pi \text{ é a permutação de Laurent} \end{cases} \tag{3.14}$$

portanto X não tem valor médio bem definido. ◊

Vejamos exemplos de variáveis aleatórias com esperança infinita. Numa urna estão 1 bola branca e 1 bola preta; uma bola é escolhida ao acaso, se for preta ela é devolvida e mais uma bola preta é colocada na urna e o sorteio é repetido, se sair bola branca o experimento termina. Depois de $m \geqslant 0$ sorteios de bolas pretas há na urna $m+2$ bolas com $m+1$ delas da cor preta. No próximo sorteio, a probabilidade de sair uma bola branca, dado que saíram m bolas pretas, é $1/(m+2)$ e a probabilidade de sair uma bola preta é $(m+1)/(m+2)$, portanto, pelo regra da multiplicação (página 29), a probabilidade do experimento terminar no $m+1$-ésimo sorteio, para $m \geqslant 1$, é

$$\left(\prod_{j=1}^{m} \frac{j}{j+1}\right) \frac{1}{m+2} = \frac{1}{2}\frac{2}{3}\frac{3}{4}\cdots\frac{m}{m+1}\frac{1}{m+2} = \frac{1}{(m+1)(m+2)}$$

e a probabilidade do experimento terminar no primeiro sorteio é $1/2$. O número esperado de sorteios no experimento é

$$1\frac{1}{2} + \sum_{m=1}^{\infty} (m+1)\frac{1}{(m+1)(m+2)} = \sum_{m=1}^{\infty} \frac{1}{m} = +\infty$$

essa última é a famosa série harmônica.

O *paradoxo de São Petersburgo* é sobre um jogo de aposta que consiste em pagar uma certa quantia para poder participar de uma rodada. Uma rodada consiste em jogar uma moeda até sair coroa, se o número de lançamentos for igual n então o valor pago ao jogador é 2^n reais. Quanto você estaria disposto a pagar para jogar esse jogo? O ganho esperado é $\sum_{n\geqslant 1} 2^n 2^{-n} = +\infty$. O fato de que a esperança é infinita sugere que um jogador deveria estar disposto a pagar qualquer quantia fixa pelo privilégio de participar do jogo, mas é improvável que alguém esteja disposto a pagar muito e aí está o paradoxo.

Exemplo 3.22 (*esperança da distribuição geométrica e da de Poisson*). Se X é geométrica de parâmetro $p \in (0,1)$, então

$$\mathbb{E}\, X = \sum_{n=1}^{\infty} n\,\mathbb{P}[X=n] = \sum_{n=1}^{\infty} n\,(1-p)^{n-1}p = p\sum_{n=0}^{\infty} n(1-p)^n = \frac{1}{p}$$

usando (s.6a).

Se X tem distribuição de Poisson com parâmetro λ

$$\mathbb{E}\, X = \sum_{n=0}^{\infty} n\,\mathbb{P}[X=n] = \sum_{n=0}^{\infty} n\,\frac{e^{-\lambda}\lambda^k}{k!} = \lambda e^{-\lambda}\sum_{n=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda}e^{\lambda} = \lambda.$$

usando (s.8). ◊

Esperança de funções de variáveis aleatórias Sejam $X\colon \Omega \to S$ uma variável aleatória real de $(\Omega, \mathbb{P})$ e $f\colon S \to \mathbb{R}$ uma função. No espaço de probabilidade $(S, \mathbb{P}_X)$ a função f é uma variável aleatória cuja esperança, caso esteja bem definida, é

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

donde tiramos que f no modelo probabilístico $(S, \mathbb{P}_X)$ é integrável se, e somente se, $f(X)$ em $(\Omega, \mathbb{P})$ é integrável. Se $\mathbb{E}\, f(X)$ está definida então podemos rearranjá-la de modo que

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

e assim provamos o seguinte resultado.

Teorema 3.23 *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 e sua esperança satisfaz*

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

sempre que a soma está bem definida. □

No caso em que Z é geométrica de parâmetro p, por exemplo, o teorema pode ser usado para computar $\mathbb{E}\, Z^2$ em conjunto com (s.6c) do apêndice do seguinte modo: derivando ambos os lados de $\sum_{n\geqslant 1} nx^n = x(1-x)^{-2}$ obtemos (para $|x| < 1$)

$$\sum_{n=1}^{\infty} n^2 x^{n-1} = \frac{1+x}{(1-x)^3}$$

de modo que

$$\mathbb{E}\, Z^2 = \sum_{n=1}^{\infty} n^2\, \mathbb{P}_Z(n) = \sum_{n=1}^{\infty} n^2 (1-p)^{n-1} p = \frac{2-p}{p^3} p = \frac{2-p}{p^2}.$$

3.2.1 PROPRIEDADES DA ESPERANÇA

A seguir veremos propriedades importantes para a média das variáveis aleatórias discretas e reais, em particular veremos que valem as propriedades das variáveis aleatórias simples enunciadas no Teorema 3.11.

Teorema 3.24 *Para todo evento A, $\mathbb{E}\, \mathbb{1}_A = \mathbb{P}(A)$.* □

A prova desse teorema é imediata da equação (3.13).

Teorema 3.25 (linearidade da esperança) *Se* X *e* Y *são variáveis aleatórias integráveis de* $(\Omega, \mathbb{P})$ *e* $a, b \in \mathbb{R}$ *constantes quaisquer 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 (3.13) acima a esperança da variável aleatória $|aX + bY|$ é

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

usando a desigualdade triangular (em (d.4)). De X e Y integráveis temos que o lado direito da equação acima é finito, portanto $aX + bY$ é integrável e

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

logo $\mathbb{E}$ é um funcional linear no conjunto das variáveis aleatórias reais de $(\Omega, \mathbb{P})$. □

Corolário 3.26 *Se* $X_1, \ldots, X_n$ *são integráveis e* $a_1, \ldots, a_n \in \mathbb{R}$ *então*

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

Demonstração. Segue do teorema por indução em $n \geqslant 2$. □

Além da linearidade, outra propriedade importante da esperança é a monotonicidade. Lembremos que $X \leqslant Y$ se $X(\omega) \leqslant Y(\omega)$ para cada $\omega \in \Omega$.

Teorema 3.27 (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 3.28 *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 está definida e 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 3.29 *Se* X *assume valores inteiros e não negativos então* $\mathbb{E}\,X$ *está bem definida e*

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

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

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

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

Proposição 3.30 (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 3.29 acima

$$\sum_{n \geqslant 1} \mathbb{P}[\,\lfloor Y \rfloor \geqslant n\,] \leqslant \mathbb{E}\,Y \leqslant \sum_{n \geqslant 1} \mathbb{P}[\,\lfloor Y \rfloor \geqslant n\,] + 1. \tag{3.15}$$

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 (3.15), $\mathbb{E}\, Y < +\infty$ 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

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

Assim, se X e Y são integráveis, XY também é e a mesma dedução sem os módulos vale, ou seja, concluímos que $\mathbb{E}[XY] = \mathbb{E}\, X \cdot \mathbb{E}\, Y$. Uma proposição mais geral é provada no próximo resultado.

Teorema 3.31 *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 $g : R \to \mathbb{R}$ 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. Da Proposição 3.6 temos que $f(X)$ e $g(Y)$ são variáveis aleatórias independentes. Podemos replicar a dedução acima para provar que $f(X)g(Y)$ é integrável e que, portanto, $\mathbb{E}[\,f(X) \cdot g(Y)\,] = \mathbb{E}[\,f(X)\,] \cdot \mathbb{E}[\,g(Y)\,]$. □

Corolário 3.32 *Se X e Y são variáveis aleatórias independentes então vale a igualdade $\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y]$.* □

Observação 3.33. O Teorema 3.27, 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 estão definidos e pelo menos uma é integrável, então as conclusões dos teoremas valem. Ainda, no Teorema 3.27 a conclusão vale se os valores esperados de X e de Y estão definidos e $\mathbb{E}\, X < +\infty$ ou $\mathbb{E}\, Y > -\infty$. ◊

Exercício 3.34. Prove a **desigualdade de Jensen**: se X é uma variável aleatória integrável e f é uma função real, contínua e convexa[3] então

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

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

[3] $f\colon [a,b] \to \mathbb{R}$ 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, $f(tx + (1-t)y) \leqslant tf(x) + (1-t)f(y)$.

Equação de Wald Se somamos as variáveis aleatórias integráveis $X_1, X_2, \ldots$, porém uma quantidade aleatória $N \geqslant 0$ delas, não podemos usar a linearidade da esperança diretamente (veja Exercício 3.65 no final do capítulo). Em geral

$$\mathbb{E} \sum_{i=1}^{N} X_i \neq \sum_{i=1}^{\mathbb{E}\, N} \mathbb{E}\, X_i .$$

Porém, se N e $X_1, X_2, \ldots$ são independentes, $X_1, X_2, \ldots$ são identicamente distribuídas e $N \leqslant k$, então de $X_1 + \cdots + X_N = X_1 \mathbb{1}_{[N \geqslant 1]} + X_2 \mathbb{1}_{[N \geqslant 2]} + \cdots + X_k \mathbb{1}_{[N \geqslant k]}$ temos

$$\mathbb{E}\left[\sum_{i=1}^{k} X_i \mathbb{1}_{[N \geqslant i]} \right] = \sum_{i=1}^{k} \mathbb{E}[\, X_i \,]\mathbb{E}[\, \mathbb{1}_{[N \geqslant i]} \,] = \mathbb{E}[\, X_1 \,] \sum_{i=1}^{k} \mathbb{E}[\, \mathbb{1}_{[N \geqslant i]} \,] = \mathbb{E}[\, X_1 \,] \sum_{i=1}^{k} \mathbb{P}[\, N \geqslant i \,].$$

portanto,

$$\mathbb{E} \sum_{i=1}^{N} X_i = \mathbb{E}[\, X_1 \,]\mathbb{E}[\, N \,].$$

Essa identidade num caso mais geral, que veremos adiante neste texto, é conhecida como equação de Wald.

Por exemplo, suponha que lançamos uma moeda honesta e sair cara, então lançamos um dado duas vezes, senão lançamos o dado três vezes. Qual é a soma esperada dos dados? Temos uma soma com 2 termos com probabilidade 1/2 cada uma, logo $\mathbb{E}\, N = 5/2$. Cada termo da soma é uma resposta do dado logo a esperança comum é 7/2. Pela equação de Wald, essa soma tem valor médio 35/4.

Exemplo 3.35. Suponhamos que são recebidas $N \leqslant 10$ mensagens eletrônicas por dia, com $\mathbb{E}\, N = 5$ e o tempo gasto lendo-as e respondendo-as são dados pelas variáveis aleatórias $X_1, X_2, \ldots, X_N$ mutuamente independentes e independentes de N, com $\mathbb{E}\, X_i = 8$ minutos, para todo i. O tempo gasto em um dia com mensagens é $X = \sum_{i=1}^{N} X_i$ minutos cujo valor esperado é $8 \cdot \mathbb{E}\, N = 40$ minutos. ◊

Se N não for limitada mas integrável, primeiro verificamos a convergência absoluta

$$\sum_{i=1}^{\infty} \mathbb{E}\Big[\, |X_i| \mathbb{1}_{[N \geqslant i]} \,\Big] = \sum_{i=1}^{\infty} \mathbb{E}[\, |X_i| \,]\mathbb{E}[\, \mathbb{1}_{[N \geqslant i]} \,] = \mathbb{E}[\, |X_1| \,] \sum_{i=1}^{\infty} \mathbb{P}[\, N \geqslant i \,] = \mathbb{E}[\, |X_1| \,]\mathbb{E}[\, N \,] < +\infty$$

de modo que o intercâmbio de somas com esperanças vale, logo

$$\mathbb{E}\left[\sum_{i=1}^{N} X_i \right] = \mathbb{E}\left[\sum_{i=1}^{\infty} X_i \, \mathbb{1}_{[N \geqslant i]} \right] = \sum_{i=1}^{\infty} \mathbb{E}[\, X_i \,]\mathbb{E}[\, \mathbb{1}_{[N \geqslant i]} \,] = \mathbb{E}[\, X_1 \,] \sum_{i=1}^{\infty} \mathbb{P}[\, N \geqslant i \,] = \mathbb{E}[\, X_1 \,]\mathbb{E}[\, N \,].$$

Essa propriedade pode ser bastante útil em situações como a dos algoritmos Las Vegas com laços, em que o i-ésima iteração executa X_i instruções de modo que o número esperado de instruções realizadas durante uma execução é $\mathbb{E} \sum_{i \geqslant 1} X_i$.

Exercício 3.36 (linearidade da esperança, caso geral). Prove que se $X_1, X_2, \ldots$ são variáveis aleatórias e $\sum_{i\geqslant 1} \mathbb{E}\,|X_i|$ converge, então $\mathbb{E}\sum_{i\geqslant 1} X = \sum_{i\geqslant 1} \mathbb{E}\,X_i$.

Para a necessidade da hipótese de convergência no exercício acima, veja o Exercício 3.66 no final do capítulo, página 151.

Lei de potência e um pouco de teoria dos números A distribuição de uma variável aleatória X sobre os inteiros positivos segue uma lei de potência com parâmetro $\alpha > 0$ se tem distribuição

$$\mathbb{P}_X(n) = \frac{1}{n^\alpha} - \frac{1}{(n+1)^\alpha}$$

ou equivalentemente, $\mathbb{P}[X \geqslant n] = 1/n^\alpha$. Distribuições de probabilidade que seguem uma lei de potência são comuns em muitas disciplinas, como a física e a biologia, e mais recentemente ganhou atenção no estudo do que se acostumou de chamar de redes complexas. Usando a Proposição 3.29 temos

$$\mathbb{E}\,X = \sum_{n=1}^{\infty} \mathbb{P}[X \geqslant n] = \sum_{n=1}^{\infty} \frac{1}{n^\alpha}. \tag{3.16}$$

Para $\alpha = 1$ vale que $\mathbb{E}\,X = \sum_{n\geqslant 1} 1/n = +\infty$, já para $\alpha = 2$ obtemos $\mathbb{E}\,X = \pi^2/6$ (veja (s.7)). Se $\alpha \leqslant 1$, a esperança é infinita. Se $\alpha > 1$, a esperança é finita mas nem sempre há uma forma fechada para a soma como no caso $\alpha = 2$.

Observação 3.37. A série no lado direito da equação (3.16) é a Função Zeta de Riemann[4]

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

quando s é qualquer número complexo com parte real maior que 1. A função zeta de Riemann está intimamente relacionada à distribuição dos números primos, fazendo da hipótese de Riemann uma conjectura central na teoria dos números.

Agora, consideramos uma variável aleatória X que assume valor nos inteiros positivos tal que $\mathbb{P}[X = n] = (cn^3)^{-1}$, com $c = \zeta(3) = \sum_{j\geqslant 1} 1/j^3$. Essa variável tem valor esperado $\mathbb{E}\,X = \sum_{n\geqslant 1} 1/cn^2 = \pi^2/6c$. Ainda, X^2 é uma variável aleatória e seu valor esperado é

$$\mathbb{E}\,X^2 = \sum_{n=1}^{\infty} i^2 \mathbb{P}[X = n] = \sum_{n=1}^{\infty} \frac{1}{cn} = +\infty.$$

A distribuição $\mathbb{P}_X(n) = (\zeta(3)n^3)^{-1}$ sobre os inteiros positivos tem a seguinte propriedade interessante (Alexander, Baclawski e Rota, 1993): seja A_p o evento formado

[4] Um dos problemas matemáticos mais importantes sem solução até o momento, conhecido como Hipótese de Riemann, é uma conjectura a respeito dos zeros da função zeta; esses estariam somente nos inteiros negativos pares e nos complexos com parte real 1/2.

pelos múltiplos de p. Se p e q são distintos, os eventos A_p e A_q são independentes $\mathbb{P}_{\mathsf{X}}(\mathsf{A}_p \cap \mathsf{A}_q) = \mathbb{P}_{\mathsf{X}}(\mathsf{A}_p)\,\mathbb{P}_{\mathsf{X}}(\mathsf{A}_q) = 1/(pq)^3$. O mesmo vale para $\mathbb{P}_{\mathsf{X}_s}(n) = (\zeta(s)n^s)^{-1}$ e todo $s > 1$

$$\mathbb{P}_{\mathsf{X}_s}(\mathsf{A}_p) = \frac{\sum_{k\geqslant 1}(pk)^{-s}}{\sum_{n\geqslant 1} n^{-s}} = \frac{1}{p^s} \quad \text{e} \quad \mathbb{P}_{\mathsf{X}_s}\big(\mathsf{A}_p \cap \mathsf{A}_q\big) = \mathbb{P}_{\mathsf{X}_s}\big(\mathsf{A}_{pq}\big) = \frac{1}{p^s}\frac{1}{q^s} = \mathbb{P}_{\mathsf{X}_s}\big(\mathsf{A}_p\big)\mathbb{P}_{\mathsf{X}_s}\big(\mathsf{A}_q\big).$$

Vamos usar essa distribuição para dar um demonstração probabilística de que a série $\sum_{p\text{ primo}} 1/p$ diverge. Defina para todo primo q o conjunto

$$\mathsf{B}_q = \bigcap_{\substack{p\leqslant q\\ p\text{ primo}}} \overline{\mathsf{A}_p}$$

dos inteiros que não têm um divisor primo menor que q. A sequência $(\mathsf{B}_q : q \text{ primo})$ é decrescente e $\bigcap_q \mathsf{B}_q = \{1\}$. Pela continuidade da probabilidade

$$\begin{aligned}\frac{1}{\zeta(s)} = \mathbb{P}_{\mathsf{X}_s}(1) = \mathbb{P}_{\mathsf{X}_s}\left(\bigcap_{q\text{ primo}} \mathsf{B}_q\right) &= \lim_{q\to\infty} \mathbb{P}_{\mathsf{X}_s}\big(\mathsf{B}_q\big) = \lim_{q\to\infty} \prod_{\substack{p\leqslant q\\ p\text{ primo}}} \mathbb{P}_{\mathsf{X}_s}\big(\overline{\mathsf{A}_p}\big)\\ = \prod_{p\text{ primo}} \mathbb{P}_{\mathsf{X}_s}\big(\overline{\mathsf{A}_p}\big) &= \prod_{p\text{ primo}} \left(1 - \frac{1}{p^s}\right)\end{aligned} \tag{3.17}$$

que é a *fórmula do produto de Euler*, descoberta por Euler quando estava na busca por uma prova de que a série $\sum_{p\text{ primo}} 1/p$ diverge. Agora, tomando o logaritmo

$$-\log\zeta(s) = \sum_{p\text{ primo}} \left(1 - \frac{1}{p^s}\right)$$

e de $0 < 1/p^s < 1/2$ temos $\log(1 - 1/p^s) \geqslant -2(1/p^s)$, logo $\log\zeta(s) \leqslant 2\sum_{p\text{ primo}} 1/p^s$ de modo que

$$\frac{1}{2}\log\zeta(s) \leqslant \sum_{p\text{ primo}} \frac{1}{p^s} \leqslant \zeta(s).$$

Agora, a conclusão segue de $\lim_{s\to 1}\zeta(s) = \infty$ pois

$$\lim_{s\to 1} \sum_{k=1}^{n} \frac{1}{k^s} = \sum_{k=1}^{n} \frac{1}{k} > \mathsf{M}$$

para qualquer real $\mathsf{M} \geqslant 1$ se n for suficientemente grande.

De brinde, também temos uma prova da infinitude de números primos. Se assumirmos que a quantidade de números primos é finita, então o produto na direita da equação (3.17) está definido e é positivo para $s > 1$. Fazendo $s \to 1^+$ temos 0 no lado direito e $+\infty$ no lado esquerdo da equação (3.17), uma contradição que demonstra a existência de uma quantidade infinita de números primos.

Ainda, como uma curiosidade a mais, o limite $\lim_{s\to 1}\mathbb{P}_{\mathsf{X}_s}(\mathsf{A})$ é a densidade relativa de A (definida no Exercício 1.60, página 50) $\rho(\mathsf{A}) = \lim_{n\to\infty} |\mathsf{A}\cap\{1,\ldots,n\}|/n$.

3.2.2 *QUICKSORT* PROBABILÍSTICO

O *quicksort* é um algoritmo recursivo de ordenação que, grosso modo, funciona da seguinte maneira: dado uma sequência S de números, dos quais o primeiro é chamado de *pivô*, o *quicksort*(S) compara o pivô com cada outro elemento de S e rearranja a sequência, os elementos menores que o pivô formam a subsequência S_1 (alocados à esquerda do pivô no vetor) e os demais, que não o pivô, formam a subsequência S_2 (alocados à direita do pivô no vetor); o algoritmo devolve a sequência ($quicksort(S_1)$, pivô, $quicksort(S_2)$) ordenada recursivamente.

O *quicksort* foi inventado por C. A. R. Hoare em 1960 e sabemos que é muito rápido em geral, mas é lento em algumas raras instâncias. É um algoritmo que ordena os números em função dos resultados das comparações entre elementos da entrada e, nesse tipo de algoritmo, medimos a eficiência do algoritmo contando o número de comparações realizadas para ordenar a sequência. O algoritmo *quicksort* executa $O(n \log n)$ comparações em média e $O(n^2)$ comparações no pior caso para ordenar sequências de tamanho n. A Figura 3.2 abaixo ilustra um exemplo de pior caso e um exemplo de melhor caso da árvore de recursão para uma sequência de tamanho 6.

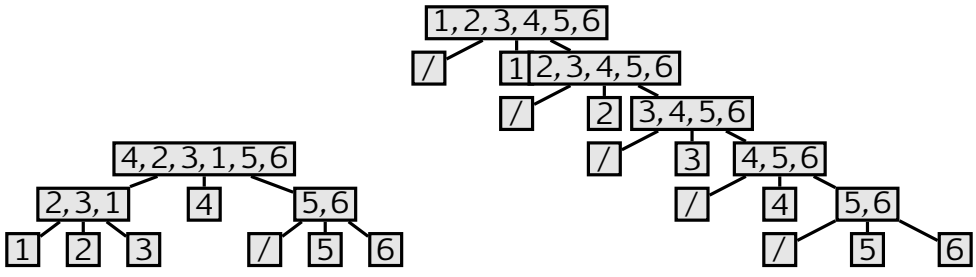

Figura 3.2: na esquerda temos uma árvore de recursão do *quicksort* onde o pivô é a mediana da sequência. A árvore da direita é o caso onde o pivô é sempre o menor elemento da sequência.

É sabido que para garantir $O(n \log n)$ comparações numa execução é suficiente garantir que as subsequências geradas no pivotamento tenham sempre uma fração aproximadamente constante do tamanho da sequência que as origina. Por exemplo, se S_1 (ou S_2) sempre tem 10% do tamanho de S então o número de comparações executadas pelo *quicksort* numa instância de tamanho n, que denotamos $T(n)$, é o número de comparações executadas em S_1 mais o número de comparações executadas em S_2 mais as $O(n)$ comparações executadas no particionamento que cria essas subsequências

$$T(n) = T(0{,}1n) + T(0{,}9n) + O(n)$$

cuja solução é uma função assintoticamente limitada superiormente por $n \log_{10/9} n$. Não há nada de especial na escolha de 10%. De fato, se uma proporção for mantida no particionamento, digamos que a menor parte tem uma fração α da entrada, então a recorrência $t(n) \leqslant t(\lfloor \alpha n \rfloor) + t(\lfloor (1-\alpha)n \rfloor) + cn$ para constantes $\alpha, c > 0$ tem solução $O(n \log n)$. Para conhecer esses resultados com mais detalhes sugerimos ao leitor uma consulta ao texto de Cormen, Leiserson e Rivest, 1990.

Na versão probabilística do *quicksort* a aleatoriedade é usada para descaracterizar o pior caso no sentido de que sorteando o pivô como qualquer elemento da sequência com grande chance evitamos os 10% menores e os 10% maiores elementos do vetor de modo que maior parte do tempo os pivotamentos garantem pelo menos uma fração constante da sequência original no menor lado. O seguinte algoritmo é o *quicksort* aleatorizado.

Instância: uma sequência de números $S = (x_1, x_2, \ldots, x_n)$.
Resposta : os elementos de S em ordem crescente.

```
1  se |S| ⩽ 20 então
2      ordene S na força bruta ;
3      responda a sequência ordenada.
4  senão
5      z ←R S;
6      para cada y ∈ S, y ≠ z faça
7          se y < z então insira y em S1;
8          senão insira y em S2;
9      ordene recursivamente a subsequência S1;
10     ordene recursivamente a subsequência S2;
11     responda (S1, z, S2).
```

Algoritmo 16: *quicksort* aleatorizado.

Para simplificar a análise assumimos que na instância não há repetição, ou seja, todos os elementos de S são distintos. Mostraremos que, com essa estratégia, o número esperado de comparações entre elementos de S é $O(n \log n)$ com alta probabilidade.

A análise desse algoritmo é um detalhamento da seguinte ideia: consideremos um elemento x de uma instância S para o algoritmo. Sejam $S'_0 := S$ e $S'_1, S'_2, \ldots, S'_M$ as subsequências a que x pertence após cada particionamento durante uma execução. O i-ésimo particionamento é *bom* se o pivô não está entre os 10% menores e os 10% maiores elementos de S'_i, o que ocorre com probabilidade 4/5, de modo que $|S'_i|/10 \leqslant |S'_{i+1}| \leqslant 9|S'_i|/10$ e portanto x passa por no máximo $\log_{10/9}(n)$ particionamentos bons. Agora, a variável aleatória M não deve ser muito maior que

$2\log_{10/9}(n)$ pois pelo Exercício 1.51, página 47, a probabilidade de ocorrem menos que $\log_{10/9}(n)$ particionamentos bons em $2\log_{10/9}(n)$ particionamentos é menor que

$$\left(\frac{4}{5}\right)^{2\log_{10/9}(n)} < \left(\frac{9}{10}\right)^{2\log_{10/9}(n)} = \frac{1}{n^2}.$$

A cada particionamento o elemento não-pivô x é comparado uma única vez, com o pivô sorteado. Assim, o numero total de comparações numa execução é a soma para todo x do número de particionamentos pelos quais x passa. Portanto, uma execução realiza mais que $2n\log_{10/9}(n)$ comparações se existe um x que passa por mais que $2\log_{10/9}(n)$ particionamentos o que ocorre com probabilidade $n(1/n^2) = 1/n$. Portanto, com alta probabilidade, o *quicksort* realiza $O(n \log n)$ comparações.

Análise do *quicksort* probabilístico O particionamento (ou pivotamento) de S nas linhas 5, 6, 7 e 8 do algoritmo acima é considerado um *sucesso* se $\min\{|S_1|, |S_2|\} \geqslant (1/10)|S|$, caso contrário, dizemos que o particionamento foi um fracasso. Um sucesso significa que o pivô não está entre os 10% maiores elementos da sequência particionada e nem está entre os 10% menores elementos da sequência particionada. Assim 80% dos elementos de S são boas escolhas para o pivô de modo que a escolha aleatória do pivô é um experimento de Bernoulli com parâmetro p dado por

$$p := \frac{\lfloor 0{,}8|S| \rfloor}{|S|} \tag{3.18}$$

donde

$$0{,}75 < p \leqslant 0{,}8$$

pois $0{,}8|S| - 1 < \lfloor 0{,}8|S| \rfloor \leqslant 0{,}8|S|$ e $|S| > 20$.

Consideremos uma execução do *quicksort* com entrada $S = (x_1, x_2, \ldots, x_n)$. Fixado $i \in \{1, 2, \ldots, n\}$, seja X_i o número de comparações entre o elemento x_i da entrada e algum pivô durante toda uma execução do algoritmo. Em cada particionamento x_i é comparado uma única vez (com o pivô) e se x_i é escolhido pivô então: (1) cada elemento da subsequência da qual x_i pertence é comparado com ele e tais comparações são contabilizadas pelos outros elementos e (2) a partir daí x_i nunca mais participará de um particionamento e nunca mais será comparado com outro elemento de S.

Exercício 3.38. Mostre que x_i participa de no máximo $\lfloor \log_{10/9} n \rfloor$ particionamentos com sucesso.

Denotemos por Y_k o número de particionamentos ocorridos entre o k-ésimo particionamento com sucesso do qual x_i participa (exclusive) e o próximo particionamento com sucesso do qual x_i participa (inclusive), ou seja, é o número de partições realizadas até obter um sucesso, portanto, $Y_k \sim \text{Geom}(p)$, logo $\mathbb{E}\, Y_k = 1/p \leqslant 1/0{,}75 <$

2. Pelo Exercício 3.38 acima, x_i participa de no máximo $\lfloor \log_{10/9} n \rfloor$ particionamentos com sucesso e temos assim que

$$X_i \leqslant \sum_{k=0}^{\lfloor \log_{10/9} n \rfloor} Y_k$$

é o número de particionamentos pelo qual passa x_i, que é, também, o número de vezes que ele é comparado com um pivô. O número total de comparações durante toda execução é dado por

$$T := \sum_{i=1}^{n} X_i.$$

Pela linearidade e monotonicidade da esperança

$$\mathbb{E}\, X_i < 2\lfloor \log_{\frac{10}{9}} n \rfloor \quad \text{e} \quad \mathbb{E}\, T < 2n\lfloor \log_{\frac{10}{9}} n \rfloor. \tag{3.19}$$

Além disso, as no máximo $n/20$ sequências ordenadas na força-bruta na linha 2 fazem, cada uma $O(1)$ comparações, contribuindo no total com $O(n)$ comparações que somadas as $O(n \log n)$ de (3.19) totalizam $O(n \log n)$ comparações.

Lema 3.39 *O número esperado de comparações entre elementos de uma sequência com n elementos numa execução do quicksort aleatorizado é $O(n \log n)$.* □

Da disciplina Análise de Algoritmos sabemos que todo algoritmo de ordenação baseado em comparação realiza $\Omega(n \log n)$ comparações para ordenar n elementos, ou seja, em média o *quicksort* aleatorizado é o melhor possível (Cormen, Leiserson e Rivest, 1990). Isso nos leva a conjeturar que o *quicksort* faz o melhor *quase sempre*.

Porém, saber que um algoritmo é bom em média não nos garante que ele é bom quase sempre. Vamos mostrar que com alta probabilidade o desempenho do algoritmo está próximo da média.

Teorema 3.40 *O número de comparações numa execução do quicksort aleatorizado com uma entrada de tamanho n é $O(n \log n)$ com probabilidade $1 - O(n^{-22})$.*

Demonstração. Definimos

$$L := 12\lfloor \log_{\frac{10}{9}} n \rfloor$$

e consideramos o evento $X_i > L$, ou seja, segundo (3.19) o elemento x_i da entrada foi comparado mais do que seis vezes o valor esperado para o número de comparações. Se $X_i > L$ então em L particionamentos ocorrem menos que $\lfloor \log_{10/9} n \rfloor$ sucessos, ou seja, o número de fracassos em L particionamentos é pelo menos

$$F := 12\lfloor \log_{\frac{10}{9}} n \rfloor - \lfloor \log_{\frac{10}{9}} n \rfloor = 11\lfloor \log_{\frac{10}{9}} n \rfloor.$$

Seja Z_i o número de particionamentos com fracasso pelos quais x_i passa ao longo de L particionamentos. A variável aleatória Z_i tem distribuição binomial com parâmetros L e $1-p$. Vamos estimar a probabilidade do evento $Z_i > F$, com isso, teremos uma limitante superior para a probabilidade de $X_i > L$. A probabilidade de ocorrerem $j > F$ fracassos em L ensaios independentes de Bernoulli é (usando limitante superior usual para o coeficiente binomial, veja (d.2))

$$\binom{L}{j}(1-p)^j p^{L-j} \leqslant \left(\frac{eL}{j}\frac{1-p}{p}\right)^j p^L \leqslant \left(\frac{eL}{F}\frac{1-p}{p}\right)^j p^L.$$

Ademais

$$e \cdot \frac{L}{F} \cdot \frac{1-p}{p} < e \cdot \frac{12}{11} \cdot \frac{0{,}25}{0{,}75} < 0{,}99$$

portanto,

$$\mathbb{P}[Z_i > F] = \sum_{j=F+1}^{L} \binom{L}{j}(1-p)^j p^{L-j} < \sum_{j=F+1}^{L} 0{,}99^j p^L = p^L \sum_{j=F+1}^{L} 0{,}99^j < p^L \sum_{j\geqslant 0} 0{,}99^j = 100p^L$$

no último passo usamos (s.6a), donde segue que

$$\mathbb{P}[X_i > L] < 100(0{,}8)^{12\log_{10/9}(n)-12} = \frac{100}{0{,}8^{12}}(0{,}8)^{\log_{10/9}(n^{12})} < \frac{100}{0{,}8^{12}}(0{,}8)^{c\log_{8/10}(n^{12})}$$

onde $c = 1/\log_{0,8}(10/9) \in (2,3)$ é a constante devido a mudança de base do logaritmo para 0,8, logo

$$100p^L < \frac{100}{0{,}8^{12}} n^{12c} < n^{-23}$$

para todo $n > 20$.

A probabilidade de existir i com $X_i > 12\lfloor \log_{\frac{10}{9}} n \rfloor$ é

$$\mathbb{P}\left(\bigcup_{i=1}^{n}\left[X_i > 12\lfloor \log_{\frac{10}{9}} n \rfloor\right]\right) < n \cdot \frac{1}{n^{23}}$$

logo, com probabilidade $1 - O(n^{-22})$ vale que $X_i \leqslant 12\log_{10/9}(n)$ para todo i e que, com essa probabilidade, o *quicksort* aleatorizado executa $O(n \log n)$ comparações entre elementos da entrada, incluídas aí as $O(n)$ comparações feitas na linha 2 do algoritmo. □

3.3 O MÉTODO PROBABILÍSTICO

O método probabilístico é um método não construtivo para demonstrar a existência de objetos matemáticos e que foi popularizado pelo matemático húngaro Paul Erdős. O princípio básico é baseado no fato de que se algum objeto em uma coleção

de objetos tiver uma determinada propriedade, então com probabilidade positiva uma escolha aleatória com uma distribuição apropriada nessa coleção tem essa propriedade. Em muitos casos o fato demonstrado é uma sentença que não envolve probabilidade, como o exemplo da divergência de $\sum_p 1/p$ que apresentamos na página 124 (com a ressalva de que naquele caso existe demonstração sem usar probabilidade, o que nem sempre é o caso). Há várias técnicas, algumas sofisticadas, que se encaixam nesse método e nessa seção veremos uma muito simples, baseada no seguinte exercício.

Exercício 3.41 (princípio de primeiro momento). Seja X uma variável aleatória real e k um número real. Se $\mathbb{E}\, X \geqslant k$ então $X(\omega) \geqslant k$ para algum $\omega \in \Omega$.

3.3.1 SATISFAZIBILIDADE DE FÓRMULA BOOLEANA

Uma variável booleana assume um de dois possíveis valores: *verdadeiro*, que representaremos por 1, ou *falso*, que representaremos por 0. Uma *fórmula* booleana é uma expressão que envolve variáveis $v_1, v_2, \ldots, v_n$, parênteses e os operadores lógicos $\vee$ (lê-se *ou*), $\wedge$ (lê-se *e*) e $\neg$ (lê-se *não*). Um *literal* é uma variável ou a sua negação. Uma *cláusula* é uma disjunção (*ou*) de literais. Por exemplo,

$$(v_1 \vee \neg v_2 \vee v_3) \wedge (v_2 \vee \neg v_3 \vee v_4) \wedge (v_3 \vee \neg v_4) \tag{3.20}$$

é uma fórmula booleana com três cláusulas, as duas primeiras compostas por três literais e a outra por dois (usamos que o *não* tem precedência sobre os outros operadores de modo que, por exemplo, $v_1 \vee \neg v_2 \vee v_3$ deve ser lido como $v_1 \vee (\neg v_2) \vee v_3$).

Uma fórmula expressa como uma conjunção de cláusulas está na Forma Normal Conjuntiva, abreviada CNF (do inglês *conjunctive normal form*). Para qualquer inteiro $k \geqslant 1$, uma *fórmula k-CNF* é da forma $C_1 \wedge C_2 \wedge \cdots \wedge C_m$ com cada cláusula C_i sendo uma disjunção de no máximo k literais, como na equação (3.20) acima que é uma fórmula 3-CNF.

Uma *valoração* das variáveis é uma atribuição dos valores 0 ou 1 para cada uma delas e tal valoração *satisfaz* a fórmula se o resultado dessas atribuições, de acordo com as regras da lógica booleana[5], é 1. No exemplo dado na equação (3.20) a valoração $v_1 = 1$, $v_2 = 1$, $v_3 = 1$ e $v_4 = 0$ satisfaz a fórmula.

Uma valoração satisfaz uma fórmula CNF se, e somente se, satisfaz cada uma das cláusulas dessa fórmula. Vamos assumir, sem perda de generalidade, que nenhuma cláusula contém um literal e sua negação pois tal cláusula é sempre satisfeita em qualquer valoração.

[5] $u \wedge v = \min\{u, v\}$, $u \vee v = \max\{u, v\}$ e $\neg u = 1 - u$.

O problema Satisfazibilidade é o problema de decidir se uma fórmula booleana admite uma valoração das suas variáveis que a satisfaz.

Problema computacional da satisfazibilidade booleana (SAT):

Instância: uma fórmula booleana CNF.

Resposta : *sim* se existe uma valoração que satisfaz a fórmula, *não* caso contrário.

Esse problema tem um papel central em Complexidade Computacional e não se conhece algoritmo eficiente que o resolva; foi o primeiro problema reconhecido como NP-completo, o que significa que todo problema NP pode ser codificado como uma instância de SAT. Esse notável resultado é conhecido como o Teorema de Cook–Levin.

O 3SAT é o problema de satisfazibilidade obtido quando restringimos as instâncias às fórmulas 3-CNF, também é um problema sem algoritmo eficiente conhecido. Um algoritmo eficiente para SAT também é um algoritmo eficiente para 3SAT e prova-se que um algoritmo eficiente para 3SAT implica em um algoritmo eficiente para SAT. Porém, quando restringimos as instâncias às fórmulas 2-CNF, o 2SAT, conhecemos um teste eficiente de satisfazibilidade (Exercício 3.75 no final do capítulo).

O problema Satisfazibilidade Máxima (MAX-SAT) é um problema de otimização que consiste em determinar o maior número possível de cláusulas de uma fórmula CNF que podem ser satisfeitas por uma valoração.

Problema computacional da satisfazibilidade máxima de uma fórmula k-CNF (MAX-kSAT):

Instância: uma fórmula k-CNF com $k \geqslant 2$.

Resposta : o maior número de cláusulas que podem ser satisfeitas concomitantemente.

Esses problemas são pelo menos tão difíceis de se resolver algoritmicamente quanto SAT e 3SAT, claramente um algoritmo eficiente para o problema MAX-SAT implica em um algoritmo eficiente para o problema SAT.

Seja $\mathcal{C} = \{C_1, C_2, \ldots, C_m\}$ uma instância do SAT e $V = \{v_1, \ldots, v_n\}$ o conjunto das variáveis de $\mathcal{C}$. O conjunto de todas as valorações das variáveis é $\Omega := \{0,1\}^n$ e interpretamos a i-ésima coordenada como o valor atribuído a v_i. Em Ω consideramos a distribuição uniforme. Para cada i definimos a variável aleatória indicadora

$$\mathbb{1}_{[C_i=1]}(x_1, \ldots, x_n) := \begin{cases} 1, & \text{se a cláusula } C_i \text{ foi satisfeita por } (x_1, \ldots, x_n) \in \{0,1\}^n, \\ 0, & \text{caso contrário.} \end{cases}$$

O número de cláusulas satisfeitas por uma valoração é dado pela variável aleatória

$$X := \sum_{i=1}^{m} \mathbb{1}_{[C_i=1]}$$

e o número médio de cláusulas satisfeitas por uma valoração é, pela linearidade da esperança,

$$\mathbb{E}\, X = \sum_{i=1}^{m} \mathbb{P}[\, C_i = 1\,] \geqslant \frac{m}{2} \tag{3.21}$$

pois uma cláusula com k literais não é satisfeita se todos valem 0, o que ocorre com probabilidade 2^{-k}, de modo $[C_i = 1]$ com probabilidade $1 - 2^{-k} \geqslant 1/2$. Na fórmula $v \wedge \neg v$ apenas metade das cláusulas pode ser satisfeita de modo que a garantia dada na equação (3.21) é a melhor possível. Assim, podemos concluir que toda fórmula boolena CNF admite uma valoração que satisfaz pelo menos metade das cláusulas.

Vamos assumir instâncias 3-CNF em que cada cláusula tem exatamente três literais (com a hipótese de não ocorrer um literal e seu complemento); vamos supor que as cláusulas não têm variáveis repetidas. Esse problema é rotulado como MAX-E3SAT.

Uma cláusula não é satisfeita se cada uma de seus literais não o for, o que ocorre com probabilidade 1/8, portanto, uma cláusula é satisfeita com probabilidade $\mathbb{P}[\, C_i = 1\,] = 7/8$, qualquer que seja i. O número de cláusulas satisfeitas por uma valoração é dado pela variável aleatória

$$X := \sum_{i=1}^{m} \mathbb{1}_{[C_i = 1]} \tag{3.22}$$

e o número médio de cláusulas satisfeitas por uma valoração é, pela linearidade da esperança, $\mathbb{E}\, X = (7/8)m$. Se o número médio de cláusulas satisfeitas é $7m/8$, com a média sobre as valorações em Ω, então pelo Exercício 3.41 deve haver uma valoração em Ω que satisfaz pelo menos $7m/8$ cláusulas. Notemos que segue desse resultado que toda instância com no máximo 7 cláusulas é satisfazível (veja o Exemplo 6.2, página 160).

TEOREMA 3.42 *Para toda instância de* MAX-E3SAT, *existe uma valoração que satisfaz pelo menos* 7/8 *de todas as cláusulas.* □

Se um algoritmo sorteia uma valoração e determina quantas cláusulas são satisfeitas por ela, qual o número esperado de sorteios até que seja determinada uma valoração que satisfaça pelo menos 87,5% (ou 7/8) de todas as cláusulas?

Pelo Exercício 3.5, página 102, podemos escolher um valor lógico para cada variável, com as escolhas independentes, ao invés de escolher ao acaso uma valoração

de V. O algoritmo pode ser descrito como abaixo.

Instância: cláusulas $\{C_1, \ldots, C_m\}$ sobre as variáveis $\{v_1, \ldots, v_n\}$ de uma fórmula 3-CNF.
Resposta: uma valoração que satisfaz $\geqslant \frac{7}{8}m$ cláusulas.

1 **repita**
2 **para cada** $i \in \{1, \ldots, n\}$ **faça** $v_i \xleftarrow{R} \{0, 1\}$;
3 avalie $C_1, \ldots, C_m$;
4 **até que** *pelo menos* $\frac{7}{8}m$ *cláusulas estejam satisfeitas*
5 **responda** $v_1, \ldots, v_n$ *e a quantidade de cláusulas satisfeitas.*

Algoritmo 19: 7/8-aproximação para MAX-E3SAT.

Esse algoritmo aleatorizado determina uma valoração das variáveis de uma fórmula booleana de modo a satisfazer pelo menos 7/8 das cláusulas da instância. Como isso garante uma resposta para o problema de otimização que situa-se entre 7/8 do valor ótimo e o próprio valor ótimo de uma instância, dizemos que o algoritmo é uma 7/8-*aproximação* para MAX-E3SAT.

Se p é a probabilidade do evento "pelo menos $7m/8$ cláusulas são satisfeitas" por uma valoração aleatória, então o número de sorteios necessários é uma variável aleatória $Z \sim \text{Geom}(p)$. A probabilidade p é difícil de determinar pois os eventos $[C_i = 1]$ $(i = 1, \ldots, m)$ não são independentes. Vamos estimar um limitante inferior para p, assim teremos um limitante superior para $\mathbb{E}\, Z = 1/p$.

Proposição 3.43 *Se p é a probabilidade com que uma valoração satisfaz pelo menos 7/8 de todas as cláusulas de uma fórmula 3-CNF, então $p \geqslant 1/(m+8)$.*

Demonstração. Para estimar p, seja p_j é a probabilidade de uma valoração aleatória satisfazer exatamente j cláusulas. Se X é o número de cláusulas satisfeitas por uma valoração, definida na equação (3.22) acima, então

$$\frac{7}{8}m = \mathbb{E}\, X = \sum_{j=0}^{m} j p_j. \tag{3.23}$$

Para qualquer inteiro $M < m$

$$\sum_{j=0}^{m} j p_j = \sum_{j=0}^{M} j p_j + \sum_{j=M+1}^{m} j p_j \leqslant \sum_{j=0}^{M} M p_j + \sum_{j=M+1}^{m} m p_j = M \sum_{j=0}^{M} p_j + m \sum_{j=M+1}^{m} p_j \tag{3.24}$$

e se M é o maior natural estritamente menor que $7m/8$, então pela definição de p

$$\sum_{j=0}^{M} p_j = 1 - p \qquad \text{e} \qquad \sum_{j=M+1}^{m} p_j = p. \tag{3.25}$$

Das equações (3.23), (3.24) e (3.25), deduzimos que

$$\frac{7}{8}m \leqslant M(1-p) + mp, \tag{3.26}$$

logo $p \geqslant (7m - 8M)/(8m - 8M)$. Pela escolha de M temos $8M < 7m$ e como ambos são naturais $7m - 8M \geqslant 1$. Também, de $7m/8 \leqslant M + 1$ deduzimos que $8m - 8M \leqslant m + 8$, portanto $p \geqslant 1/(8 + m)$. □

Corolário 3.44 *O número esperado de rodadas do laço do Algoritmo 19 é* $\mathbb{E}\, Z \leqslant m + 8$. □

O número total de instruções executadas até terminar é proporcional ao número de rodadas do **repita**, cujo valor esperado é $\leqslant 8 + m$ pelo corolário acima. Supondo que a avaliação de cada cláusula pode ser feita em tempo constante, cada iteração do **repita** tem custo de sortear n bits e avaliar m cláusulas, portanto, o custo é $O(m+n)$. Esse algoritmo tem custo esperado $O(m(m+n))$, em que m é o número de cláusulas e n o de variáveis. Como consideramos somente variáveis que aparecem explicitamente nas cláusulas temos $n \leqslant 3m$, portanto, o algoritmo acima descobre uma valoração com a propriedade desejada para uma fórmula 3-CNF com m cláusulas usando $O(m^2)$ instruções.

Mais que isso pode ser dito nesse caso. A probabilidade de uma execução exceder muito o tempo esperado é pequena. De fato, usando a estimativa para uma variável geométrica dada no Exercício 3.3, página 99, a probabilidade do número de iterações ultrapassar duas vezes o valor esperado é

$$\mathbb{P}[\, Z \geqslant 2m + 16\,] \leqslant \left(1 - \frac{1}{m+8}\right)^{2m+15} < 0{,}14$$

para todo m. Para $m > 50$, o número de iterações do laço ultrapassa $m\log(m)$ com probabilidade menor que 0,029 e ultrapassa m^2 com probabilidade menor que $1{,}3 \times 10^{-19}$.

Algoritmos aproximativos Muitos problemas de otimização são NP-difíceis, o que quer dizer que eles são, possivelmente, difíceis de resolver de modo eficiente no pior caso. Há casos, no entanto, em que soluções próximas da ótima são razoavelmente boas para a maioria das instâncias.

Para problemas de maximização, *algoritmos aproximativos* são algoritmos que encontram soluções próximas da ótima, para uma determinada garantia $0 < \alpha < 1$, isto é, o valor respondido pelo algoritmo é no máximo o valor ótimo global OPT e pelo menos $\alpha \cdot \text{OPT}$. Nesse caso dizemos que é uma α-aproximação e uma questão é para quais valores de α podemos esperar um algoritmo de aproximação de tempo polinomial. Nesta seção vimos um dos primeiros algoritmos de aproximação considerados, devido a Johnson, 1974.

Os algoritmos aproximativos são muito estudados em Teoria da Computação pois, dentre outros fatos, podem levar a resultados surpreendentes como, por exem-

plo, o de que *uma* $(7/8+\varepsilon)$*-aproximação em tempo polinomial para* MAX-E3SAT, *para qualquer* $\varepsilon > 0$, *implicaria em* $\mathsf{P} = \mathsf{NP}$ (Håstad, 2001). Assim, supondo que $\mathsf{P} \neq \mathsf{NP}$ não temos algoritmo eficiente para o MAX-E3SAT e o melhor que poderemos fazer é encontrar um algoritmo aproximativo eficiente que, no pior caso, garanta uma valoração que satisfaça $7/8$ do número máximo de cláusulas que possam ser satisfeitas, e o algoritmo de Johnson é ótimo dentro de uma constante aditiva positiva, arbitrária.

3.3.2 CORTE GRANDE EM GRAFOS

Dado um grafo $G = (V, E)$, queremos determinar um corte com muitas arestas em G. Recordemos que o corte definido por $A \subset V$ em G é o subconjunto de arestas $\nabla(A) = \{\{i,j\} \in E \colon i \in A \text{ e } j \in \overline{A}\}$. Assumimos, sem perda de generalidade, que $V = \{1, 2, \ldots, n\}$.

Consideramos o espaço amostral $\Omega := \{0,1\}^n$ munido da medida uniforme, assim um subconjunto aleatório $A \subset V$ é identificado pelo ponto amostral $(x_1, \ldots, x_n) \in \Omega$, isto é, para cada vértice $i \in V$, se $x_i = 1$ então $i \in A$, senão $i \in \overline{A}$. O número de arestas no corte é a variável aleatória

$$|\nabla(A)| = \sum_{\{i,j\} \in E} \mathbb{1}_{\{i,j\} \in \nabla(A)}$$

cuja média é $|E| \cdot \mathbb{P}[\{i,j\} \in \nabla(A)]$. Temos $\{i,j\} \in \nabla(A)$ se, e só se, $x_i \neq x_j$ o que ocorre com probabilidade $1/2$, ou seja,

$$\mathbb{E}\,|\nabla(A)| = \frac{|E|}{2}.$$

Se o valor médio do tamanho de um corte em G é $|E|/2$ então G deve conter um corte com pelo menos $|E|/2$ arestas. Com a mesma estratégia da seção anterior, podemos transformar esse resultado num algoritmo aleatorizado para determinar um corte grande em G, isto é, um corte com pelo menos metade das arestas.

Instância: grafo G sobre os vértices $V = \{1, \ldots, n\}$ e com m arestas.
Resposta: $A \subset V(G)$ tal que $\nabla(A)$ tem $\geqslant \frac{m}{2}$ arestas.

```
1 repita
2     para cada i ∈ {1,...,n} faça x_i ←R {0,1};
3     A ← {i : x_i = 1};
4     compute |∇(A)|;
5 até que |∇(A)| ⩾ m/2
6 responda A.
```

Algoritmo 20: 1/2-aproximação para MAX-CUT.

A análise desse algoritmo é análoga à da seção anterior. As linhas 2 a 4 têm custo de execução proporcional ao tamanho do grafo $n+m$. O número total de instruções

executadas é uma variável aleatória que depende do número de rodadas do laço. O número de rodadas do laço até que ocorra um sucesso ($|\nabla(A)| \geqslant \frac{m}{2}$) é uma variável aleatória $Z \sim \text{Geom}(p)$, cuja esperança é $\mathbb{E}\, Z = 1/p$.

Seguindo a mesma dedução da seção anterior, equações (3.23), (3.24), (3.25) e (3.26) com M o maior natural estritamente menor que $m/2$, temos

$$p \geqslant \frac{m - M}{2m - 2M} \geqslant \frac{1}{m+2}$$

portanto, o número esperado de rodadas do laço da linha 1 é $\leqslant m+2$.

O custo esperado para uma rodada do algoritmo é $O(m(m+n))$ para um grafo com n vértices e m arestas. A probabilidade com que o algoritmo realiza pelo menos km rodadas até parar é $\mathbb{P}[Z \geqslant km] = (1 - 1/(m+2))^{km-1}$ para todo k. Para $k = m$ essa probabilidade é menor que 0,00019 para todo grafo com $m \geqslant 10$ arestas. Para $m \geqslant 90$ a probabilidade é menor que 5×10^{-39}.

Problema computacional do corte máximo em grafos (MAX-CUT):

Instância: um grafo G.

Resposta: o tamanho $|\nabla(A)|$, para algum A, do corte máximo.

Para esse problema não é conhecido algoritmo eficiente e o algoritmo acima é uma 1/2-aproximação, pois determina um corte cuja quantidade de arestas está entre $(1/2)\text{OPT}$ e o valor ótimo OPT. Uma $(16/17+\varepsilon)$-aproximação em tempo polinomial para MAX-CUT, para qualquer $\varepsilon > 0$, implica em P = NP (Håstad, 2001).

Existe um algoritmo determinístico eficiente de 1/2-aproximação para esse problema. Começamos com uma partição arbitrária dos vértices do grafo dado $G = (V, E)$ e movemos um vértice de cada vez de um lado da partição para o outro se isso melhora a solução, até que não haja mais movimentos que melhoram a solução. O número de iterações é no máximo $|E|$ porque o algoritmo melhora o corte em pelo menos uma aresta em cada movimento. Quando o algoritmo termina, pelo menos metade das arestas incidentes em cada vértice pertencem ao corte, pois, caso contrário, mover o vértice melhoraria o corte (verifique). Portanto, o corte inclui pelo menos metade das arestas.

3.4 DISTRIBUIÇÃO E ESPERANÇA CONDICIONAIS

Sejam $(\Omega, \mathbb{P})$ um espaço discreto de probabilidade, A um evento desse espaço com probabilidade positiva e X um variável aleatória real e integrável (ou apenas não negativa). A ocorrência de A implica uma nova medida $\mathbb{P}_A$ em Ω e temos uma nova lei, condicionada a A, para a variável X. A **esperança condicional** da variável

X dado A é o valor médio de X com respeito a lei condicional

$$\mathbb{E}[\,X \mid A\,] = \sum_{\omega\in\Omega} X(\omega)\mathbb{P}_A(\omega) = \sum_{\omega\in A} X(\omega)\frac{\mathbb{P}(\omega)}{\mathbb{P}(A)} = \frac{\mathbb{E}[\,X\cdot \mathbb{1}_A\,]}{\mathbb{P}(A)}. \tag{3.27}$$

Pelo Corolário 3.28, item 4, se X é integrável então $X\cdot\mathbb{1}_A$ também é logo $\mathbb{E}[\,X \mid A\,] < +\infty$. Dito isso, se X tem esperança bem definida, então podemos reorganizar os termos da soma e concluir que

$$\mathbb{E}[\,X \mid A\,] = \sum_r r\mathbb{P}[\,X = r \mid A\,] = \sum_r r\,\mathbb{P}_A[X = r] = \sum_r r\,\mathbb{P}_{X|A}(r).$$

Por ser uma média de uma variável aleatória com respeito a uma distribuição, a esperança condicional usufrui das mesmas propriedades da esperança.

Exercício 3.45 (propriedades da esperança condicional). Prove que se X e Y são integráveis, A um evento de probabilidade positiva, B um evento qualquer e $a, b \in \mathbb{R}$, então

1. $\mathbb{E}[\,a \mid A\,] = a$.
2. $\mathbb{E}[\,\mathbb{1}_B \mid A\,] = \mathbb{P}(\,B \mid A\,)$.
3. Se $X \leqslant Y$ então $\mathbb{E}[\,X \mid A\,] \leqslant \mathbb{E}[\,Y \mid A\,]$.
4. $\mathbb{E}[\,aX + bY \mid A\,] = a\mathbb{E}[\,X \mid A\,] + b\mathbb{E}[\,Y \mid A\,]$.
5. Se X e A são independentes então $\mathbb{E}[\,X \mid A\,] = \mathbb{E}\,X$ e $\mathbb{E}[\,XY \mid A\,] = \mathbb{E}[\,X \mid A\,]\cdot\mathbb{E}[\,Y \mid A\,]$.

Nesta seção será conveniente definirmos o **suporte**[6] da variável aleatória $X\colon \Omega \to S$ como os pontos de $(S, \mathbb{P}_X)$ com probabilidade positiva

$$S_X := \left\{ x \in S\colon \mathop{\mathbb{P}}_X(x) > 0 \right\}.$$

Se X e Y são duas variáveis aleatórias de $(\Omega, \mathbb{P})$ e $x \in S_X$ então a **distribuição condicional** da variável aleatória Y dado que ocorre o evento $[X = x]$ é

$$\mathbb{P}_{Y|X=x}(y) := \mathbb{P}[\,Y = y \mid X = x\,]$$

para todo y. Agora, para uma variável aleatória real e integrável Y, a **esperança condicional de** Y **dado** $[X = x]$, para qualquer $x \in S_X$, é dada pela equação (3.27), ou seja, vale que

$$\mathbb{E}[\,Y \mid X = x\,] = \sum_{r\in Y(\Omega)} r\,\mathbb{P}_{Y|X=x}(r).$$

[6] O usual é definir suporte como o S tal que $\mathbb{P}[\,X \in S\,] = 1$.

Claramente, valores diferentes de x podem resultar em distribuições e esperanças condicionais diferentes, temos uma função $\psi\colon \mathbb{R} \to \mathbb{R}$ dada por

$$\psi(x) = \mathbb{E}[\,Y \mid X = x\,]$$

para todo $x \in S_X$ e $\psi(x) = 0$ nos outros casos, de modo que podemos entender $\mathbb{E}[\,Y \mid X\,]$ como uma função de X

$$\mathbb{E}[\,Y \mid X\,] = \psi(X)$$

que é uma variável aleatória

$$\mathbb{E}[\,Y \mid X\,](\omega) = \psi(X(\omega)) = \mathbb{E}[\,Y \mid X = X(\omega)\,].$$

Se $S_X = \{x_1, x_2, \ldots\}$ temos que $\mathbb{E}[\,Y \mid X\,]$ assume o valor $\mathbb{E}[\,Y \mid X = x_i\,]$ com probabilidade $\mathbb{P}[\,X = x_i\,]$. Essa variável aleatória chamamos de **esperança condicional de** X **dado** Y.

Por exemplo, se um dado equilibrado é lançado duas vezes e X_1 e X_2 são os valores obtidos e $Z := X_1 + X_2$ e $P := X_1 \cdot X_2$, então usando a definição temos

$$\mathbb{E}[\,Z \mid X_1 = 2\,] = \sum_{r=2}^{12} r\mathbb{P}[\,Z = r \mid X_1 = 2\,] = 11/2$$

e

$$\mathbb{E}[\,P \mid X_2 = 1\,] = \sum_{r=1}^{36} r\mathbb{P}[\,P = r \mid X_2 = 1\,] = 7/2$$

embora. nesses casos, as esperanças possam ser calculadas de maneira mais fácil usando as propriedades do valor esperado. Por exemplo, para a soma temos

$$\mathbb{E}[\,Z \mid X_1 = 2\,] = \mathbb{E}[\,X_1 + X_2 \mid X_1 = 2\,] = \mathbb{E}[\,X_1 \mid X_1 = 2\,] + \mathbb{E}[\,X_2 \mid X_1 = 2\,] = 2 + \mathbb{E}\,X_2 = 11/2$$

para o produto temos, assumindo independência,

$$\mathbb{E}[\,P \mid X_2 = 1\,] = \mathbb{E}[\,X_1 X_2 \mid X_2 = 1\,] = \mathbb{E}[\,X_1 \mid X_2 = 1\,]\,\mathbb{E}[\,X_2 \mid X_2 = 1\,] = \mathbb{E}[\,X_1 \mid X_2 = 1\,] = 7/2.$$

Ainda, usando a definição, temos $\mathbb{E}[\,X_1 \mid Z = 5\,] = \sum_{r=1}^{6} r\mathbb{P}[\,X_1 = r \mid Z = 5\,] = 5/2$.

O valor de $\mathbb{E}[\,Z \mid P\,](3,5)$ é o valor médio da soma Z sobre todo par (x, y) tal que $P(x, y) = x \cdot y = 15 = P(3,5)$, a saber, os pares $(3,5)$ e $(5,3)$. Como ambos somam 8 o valor médio é $\mathbb{E}[\,Z \mid P\,](3,5) = 8$. O valor de $\mathbb{E}[\,Z \mid P\,](1,4)$ é o valor médio da soma Z sobre todo par (x, y) tal que $P(x, y) = x \cdot y = 4 = P(1,4)$, a saber, os pares $(1,4)$, $(2,2)$ e $(4,1)$. O primeiro e o último somam 5 e o segundo par soma 4, então o valor médio é $\mathbb{E}[\,Z \mid P\,](1,4) = (2/3) \cdot 5 + (1/3) \cdot 4 = 14/3$. Escrevendo de outro modo, se $P = 15$ então os resultados dos lançamentos são $(3,5)$ ou $(5,3)$, para os quais $Z = 8$, de modo que

$$\mathbb{E}[\,Z \mid P\,](3,5) = \psi(P(3,5)) = \psi(15) = \mathbb{E}[\,Z \mid P = 15\,] = \frac{\mathbb{E}\big[\,Z \cdot \mathbb{1}_{[P=15]}\,\big]}{\mathbb{P}[\,P = 15\,]} = \frac{8(2/36)}{2/36} = 8.$$

Se $P = 4$ então os resultados dos lançamentos são $(1,4)$, ou $(2,2)$ ou $(4,1)$, para os quais $Z = 5, 4, 5$, respectivamente, então

$$\mathbb{E}[\,Z \mid P\,](2,2) = \psi(P(2,2)) = \psi(4) = \mathbb{E}[\,Z \mid P = 4\,] = \frac{5(1/36) + 4(1/36) + 5(1/36)}{3/36} = \frac{14}{3}.$$

A lei da variável aleatória $\mathbb{E}[\,Z \mid P\,]$ está dada na Tabela 3.3. Podemos usar a Tabela

$\psi(P)$	2	3	4	$\frac{14}{3}$	6	7	$\frac{15}{2}$	8	9	10	11	12
$\mathbb{P}_{\psi(P)}$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{9}{36}$	$\frac{2}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Tabela 3.3: distribuição da soma de dois dados dado o produto $\mathbb{E}[\,Z \mid P\,]$.

para concluir que $\mathbb{E}[\,\mathbb{E}[\,Z \mid P\,]\,] = \mathbb{E}[\,\psi(P)\,] = 7$. Ainda, pela linearidade da esperança, $\mathbb{E}\,Z = \mathbb{E}\,X_1 + \mathbb{E}\,X_2 = 7$ (recorde o Exemplo 3.14).

Se $Y = aX + b$, para $a, b \in \mathbb{R}$, então $\mathbb{E}[\,Y \mid X = x\,] = ax + b$ e $\mathbb{E}[\,Y \mid X\,] = aX + b$. Ademais, pela linearidade da esperança, $\mathbb{E}[\,\mathbb{E}[\,Y \mid X\,]\,] = a\,\mathbb{E}\,X + b = \mathbb{E}\,Y$.

Se $\mathbb{E}[\,Y \mid X\,]$ é uma variável aleatória, então podemos calcular o seu valor médio quando existir. Considando Y integrável, de modo que $\mathbb{E}[\,Y \mid X = x\,]$ é finita para todo $x \in S_X$, temos

$$\begin{aligned}
\mathbb{E}\,Y &= \sum_y y\mathbb{P}[\,Y = y\,] \\
&= \sum_y y\left(\sum_x \mathbb{P}([\,Y = y\,] \cap [\,X = x\,])\right) \\
&= \sum_y \sum_x y\mathbb{P}[\,Y = y \mid X = x\,]\mathbb{P}[\,X = x\,] \\
&= \sum_x \sum_y y\mathbb{P}[\,Y = y \mid X = x\,]\mathbb{P}[\,X = x\,] \\
&= \sum_x \mathbb{E}[\,Y \mid X = x\,]\mathbb{P}[\,X = x\,] = \sum_x \psi(x)\mathbb{P}[\,X = x\,] = \mathbb{E}\,\psi(X) = \mathbb{E}[\,\mathbb{E}[\,Y \mid X\,]\,]
\end{aligned}$$

pelo teorema de Fubini para séries (s.5). Essa dedução vale sempre que $Y \geqslant 0$ ou Y é integrável, nesse último caso $\mathbb{E}[\,Y \mid X\,]$ também é integrável. Essa dedução demonstra o seguinte resultado.

Teorema 3.46 *Se Y é uma variável aleatória integrável, então*

$$\mathbb{E}\,Y = \sum_{x \in S_X} \mathbb{E}[\,Y \mid X = x\,]\mathbb{P}[\,X = x\,] \; e \tag{3.28}$$

$$\mathbb{E}\,Y = \mathbb{E}[\,\mathbb{E}[\,Y \mid X\,]\,].$$

e a variável aleatória $\mathbb{E}[\,Y \mid X\,]$ é integrável. □

Tomando $X = \mathbb{1}_A$ na equação (3.28) a expressão que obtemos é o teorema da probabilidade total, Teorema 1.26, página 30. A equação (3.28) é chamada **Lei de Esperança Total**.

Como ilustração, vamos usar o teorema acima para calcular o número esperado de lançamentos de uma moeda equilibrada até sair coroa (calculamos essa esperança na página 118) condicionando no resultado do primeiro lançamento. Seja X uma variável aleatória geométrica com parâmetro 1/2, o número de lançamentos até sair coroa usando o Teorema 3.46 é

$$\mathbb{E}\,X = \mathbb{E}[\,X \mid X = 1\,]\mathbb{P}[\,X = 1\,] + \mathbb{E}[\,X \mid X > 1\,]\mathbb{P}[\,X > 1\,] = \frac{1}{2} + \mathbb{E}[\,X \mid X > 1\,]\frac{1}{2}$$

e como uma variável aleatória com distribuição geométrica não tem memória $\mathbb{P}[\,X = n+1 \mid X > 1\,] = \mathbb{P}[X = n]$ (Exercício 3.53, página 150), logo

$$\mathbb{E}[\,X \mid X > 1\,] = \sum_{n\geqslant 1}(n+1)\mathbb{P}[\,X = n+1 \mid X > 1\,] = \sum_{n\geqslant 1}(n+1)\mathbb{P}[X = n] = \mathbb{E}[\,X\,] + 1$$

portanto $\mathbb{E}\,X = (1/2)+(1/2)(\mathbb{E}\,X+1)$ donde concluímos que $\mathbb{E}\,X = 2$. Ainda, podemos usar a mesma estratégia para calcular $\mathbb{E}\,X^2$

$$\begin{aligned}
\mathbb{E}\,X^2 &= \mathbb{E}\Big[X^2 \,\Big|\, X > 1\Big]\frac{1}{2} + \mathbb{E}\Big[X^2 \,\Big|\, X = 1\Big]\frac{1}{2} \\
&= \mathbb{E}\,(X+1)^2\,\frac{1}{2} + \mathbb{E}\Big[X^2 \,\Big|\, X = 1\Big]\frac{1}{2} \\
&= \Big(\mathbb{E}\,X^2 + 2\,\mathbb{E}\,X + 1\Big)\frac{1}{2} + \frac{1}{2} \\
&= \mathbb{E}\,X^2\,\frac{1}{2} + \mathbb{E}\,X + \frac{1}{2} + \frac{1}{2} \\
&= \mathbb{E}\,X^2\,\frac{1}{2} + 3
\end{aligned}$$

portanto $\mathbb{E}\,X^2 = 6$.

Exercício 3.47. Considere o seguinte experimento: lançamos um dado e observamos seu resultado X, em seguida lançamos uma moeda X vezes e Y conta o número de caras. Verifique que os valores esperados $\mathbb{E}[\,Y \mid X\,] = Y/2$ e $\mathbb{E}\,X = \mathbb{E}\,\mathbb{E}[\,X \mid Y\,] = 7/4$.

3.4.1 *DESALEATORIZAÇÃO*: O MÉTODO DAS ESPERANÇAS CONDICIONAIS

Vimos na página 78, no caso do algoritmo para a verificação do produto de matrizes, uma estratégia que reduz a quantidade de bits aleatórios usados naquele caso específico sem comprometer significativamente o desempenho do algoritmo. Nessa seção vamos ver uma estratégia mais genérica que resulta num algoritmo determinístico mas cujo desempenho depende de conseguirmos computar esperanças condicionais de modo eficiente.

A seguir nós usaremos a notação $[X_1 = x_1, X_2 = x_2, \ldots, X_i = x_i]$ com o significado de $[X_1 = x_1] \cap [X_2 = x_2] \cap \cdots \cap [X_i = x_i]$.

Sejam $f\colon S^n \to \mathbb{R}$ uma função em que S é finito e $X_1, \ldots, X_n$ variáveis aleatórias tais que $\mathbb{E}\, f(X_1, \ldots, X_n) \geqslant \mu$. Por exemplo, no MAX-CUT e no MAX-E3SAT temos $(X_1, \ldots, X_n) \in_R \{0,1\}^n$, que correspondem, respectivamente, a uma bipartição nos vértices de um grafo sobre n vértices e a valoração de n variáveis booleanas. No primeiro caso f é o número de arestas no corte e no segundo a quantidade de cláusulas satisfeitas. Em ambos os casos sabemos que $\mathbb{E}\, f$ é limitada inferiormente.

Determinar um ponto $(x_1, \ldots, x_n) \in S^n$ tal que $f(x_1, \ldots, x_n) \geqslant \mu$ fazendo uma busca exaustiva em S^n pode não ser eficiente pois o conjunto pode ser muito grande. Entretanto, pode ser possível determinar tal ponto de modo eficiente quando for possível computar as esperanças condicionais $\mathbb{E}[\, f(X_1, \ldots, X_n) \mid X_1 = x_1, \ldots, X_i = x_i \,]$, para todo i, de modo eficiente.

Tomemos

$$Y = f(X_1, \ldots, X_n)$$

e para cada $i = 1, 2, \ldots, n$, dados $x_1, \ldots, x_i \in \{0,1\}$ tais que $\mathbb{E}[\, Y \mid X_1 = x_1, \ldots, X_i = x_i \,] \geqslant \mu$, temos pelo Teorema 3.46 que a variável aleatória $\mathbb{E}[\, Y \mid X_1 = x_1, \ldots, X_i = x_i, X_{i+1} \,]$ satisfaz

$$\mathbb{E}[\, \mathbb{E}[\, Y \mid X_1 = x_1, \ldots, X_i = x_i, X_{i+1} \,]\,] = \mathbb{E}[\, Y \mid X_1 = x_1, \ldots, X_i = x_i \,] \geqslant \mu$$

logo, pelo princípio de primeiro momento, deve existir algum r em S para o qual temos o limitante para o valor esperado $\mathbb{E}[\, Y \mid X_1 = x_1, \ldots, X_i = x_i, X_{i+1} = r \,] \geqslant \mu$ assim, testamos (de modo eficiente) para cada $r \in S$ se

$$\mathbb{E}[\, Y \mid X_1 = x_1, \ldots, X_i = x_i, X_{i+1} = r \,] \overset{?}{\geqslant} \mu.$$

Repetindo essa estratégia para cada i, ao final, quando $i = n$, teremos determinado valores $(x_1, \ldots, x_n)$ tais que $f(x_1, \ldots, x_n) \geqslant \mu$.

MAX-E3SAT O problema que estudamos na seção 3.3.1 é, dado uma fórmula booleana 3-CNF $\mathcal{C} = \{C_1, \ldots, C_m\}$, determinar uma valoração $(x_1, \ldots, x_n) \in \{0,1\}^n$ para as variáveis $V = \{v_1, \ldots, v_n\}$ da fórmula tal que $(v_1, \ldots, v_n) = (x_1, \ldots, x_n)$ satisfaz pelo menos 7/8 das cláusulas $C_1, \ldots, C_m$. No algoritmo aleatorizado sorteamos a valoração uniformemente em $\{0,1\}^n$ até que descobrirmos uma valoração com a propriedade desejada.

Agora, sejam $X_1, \ldots, X_n \in_R \{0,1\}$ variáveis aleatórias independentes que representam os resultados dos sorteios. Definimos a função $f(x_1, \ldots, x_n)$ como o número de cláusulas satisfeitas pela valoração $(x_1, \ldots, x_n) \in \{0,1\}^n$. Assim, como sabemos da

seção 3.3.1

$$\mathbb{E}\, f(X_1,\ldots,X_n) \geqslant \frac{7m}{8}.$$

Aplicamos a estratégia de *desaleatorização* da esperança condicional para f. Se é dado uma valoração parcial $(x_1,\ldots,x_i)$, então já sabemos os valores lógicos das variáveis $v_1, v_2, \ldots, v_i$. Precisamos computar de modo eficiente

$$\mathbb{E}[\,Y \mid X_1 = x_1,\ldots,X_i = x_i, X_{i+1} = r\,] = \sum_{j=1}^{m} \mathbb{E}\Big[\,\mathbb{1}_{[C_j=1]} \,\Big|\, X_1 = x_1,\ldots,X_i = x_i, X_{i+1} = r\,\Big]$$

para $Y = f(X_1,\ldots,X_n)$, $r = 0$ e para $r = 1$. Para r fixo, cada termo da soma no lado direito é calculada considerando quatro casos:

1. se a valoração parcial deixa 3 variáveis livres na cláusula então o valor esperado condicional de $\mathbb{1}_{[C_j=1]}$ é 7/8;

2. se a valoração parcial deixa 2 variáveis livres na cláusula então se $C_j = 1$ o valor esperado condicional é 1, senão o valor esperado condicional de $\mathbb{1}_{[C_i=1]}$ é 3/4;

3. se a valoração parcial deixa 1 variável livre na cláusula então se $C_j = 1$ o valor esperado condicional é 1, senão o valor esperado condicional de $\mathbb{1}_{[C_j=1]}$ é 1/2;

4. se a valoração parcial não deixa variável livre na cláusula então o valor esperado condicional de $\mathbb{1}_{[C_i=1]}$ é 0 ou 1, valendo 1 se, e só se, a cláusula está satisfeita;

portanto cada termo é avaliado em tempo constante e em $O(m)$ passos a soma fica determinada. Realizada a soma para $r = 0$ e $r = 1$, definimos x_{i+1} como o valor de r para o qual a esperança é pelo menos $7m/8$.

Assim, a esperança é calculada com custo linear em m e determinamos uma valoração para cada uma das n variáveis que, no final, satisfaz pelo menos $7m/8$ cláusulas com custo total $O(nm)$.

Corte grande Vejamos a aplicação desse método no problema da seção 3.3.2 de determinar um corte num grafo com pelo menos metade das arestas. Dado $G = (V, E)$ com vértices $V = \{1,\ldots,n\}$ e m arestas, o algoritmo aleatorizado sorteia $(x_1,\ldots,x_n) \in \{0,1\}^n$ para formar um subconjunto $A = \{\, i \in V \colon x_i = 1\,\}$, se $|\nabla(A)| \geqslant m/2$ então encontrou um corte grande, senão repete o sorteio. Esse algoritmo pode ser *desaleatorizado* com o método da esperança condicional e o resultado é um tão algoritmo eficiente quanto o aleatorizado.

Definimos a função $f(x_1,\ldots,x_n) \coloneqq |\nabla(A)|$ em que $(x_1,\ldots,x_n)$ é o vetor característico do conjunto A definido acima. Sejam $X_1,\ldots,X_n$ variáveis aleatórias independentes com distribuição uniforme em $\{0,1\}$, sabemos da seção 3.3.2 que

$$\mathbb{E}\, f(X_1,\ldots,X_n) \geqslant \frac{m}{2}.$$

Agora, aplicamos a estratégia dada acima para f. Se é dado $(x_1,\ldots,x_i) \in \{0,1\}^i$, para $1 \leqslant i < n$, então já sabemos quais dos vértice dentre $1,2,\ldots,i$ estão em A e vamos decidir se $i+1$ fará parte do conjunto A calculando para $r = 0$ e para $r = 1$ as esperanças $\mathbb{E}[\, f(X_1,\ldots,X_n) \mid X_1 = x_1,\ldots,X_{i+1} = r\,]$; para pelo menos uma dessas duas escolhas devemos ter a esperança condicional pelo menos $m/2$.

Para cada aresta com ambos os extremos em $\{1,\ldots,i\}$ já sabemos se ela está no corte ou não e toda outra aresta tem probabilidade $1/2$ de estar no corte, portanto, para $r \in \{0,1\}$ fixo e $Y = f(X_1,\ldots,X_n)$

$$\mathbb{E}[\, Y \mid X_1 = x_1,\ldots,X_i = x_i, X_{i+1} = r\,] = \sum_{\{j,k\}\in E} \mathbb{E}\Big[\, \mathbb{1}_{[\{j,k\}\in\nabla(A)]} \,\Big|\, X_1 = x_1,\ldots,X_i = x_i, X_{i+1} = r\Big]$$

e cada termos da soma é calculado considerando três casos:

1. se $j,k \in \{1,\ldots,i+1\}$ então o valor esperado condicional para $\mathbb{1}_{[\{j,k\}\in\nabla(A)]}$ é 1 ou 0, valendo 1 se, e só se, $\{j,k\} \in \nabla(A)$;

2. se $j,k \in \{i+2,\ldots,n\}$ então o valor esperado condicional para $\mathbb{1}_{[\{j,k\}\in\nabla(A)]}$ é $1/2$;

3. se $j \in \{1,\ldots,i+1\}$ e $k \in \{i+1,\ldots,n\}$, ou se $k \in \{1,\ldots,i+1\}$ e $j \in \{i+1,\ldots,n\}$, então o valor esperado condicional para $\mathbb{1}_{[\{j,k\}\in\nabla(A)]}$ é $1/2$.

Assim, a esperança $\mathbb{E}[\, f(X_1,\ldots,X_n) \mid X_1 = x_1,\ldots,X_{i+1} = x_{i+1}\,]$ é calculada com custo linear em $|E| = m$ e ao final de n passos determinamos $(x_1,\ldots,x_n)$ (ou seja, um conjunto A) tal que $f(x_1,\ldots,x_n) \geqslant m/2$ (ou seja, $|\nabla(A)| \geqslant m/2$) com custo $O(nm)$.

3.4.2 *SKIP LISTS*

Uma **skip list** é uma estrutura de dados para representar conjuntos dinâmicos (Pugh, 1989) que usa aleatoriedade para se comportar, em desempenho, como uma árvore de busca balanceada[7] e com a vantagem de ser mais eficiente e mais fácil de manter que uma árvore balanceada.

[7] Uma árvore binária balanceada é uma árvore com raiz r cujas subárvores esquerda e direita satisfazem algum critério de balanceamento como, por exemplo, as alturas diferem de no máximo 1. Um critério bom garante que a altura da árvore é logarítmica no número de elementos, o que faz uma busca ser rápida. Em contrapartida, toda alteração exige o esforço de recompor o balanceamento.

Numa *skip list* S mantemos um conjunto, que por abuso de notação também denotamos por S, de elementos de um universo U dotado de ordem total. Uma *skip list* pode ser descrita da seguinte maneira: os elementos de S são mantidos em uma lista ligada ordenada chamada *lista do nível 0* e denotada por S_0. Dada a lista S_i do nível i ($i \geqslant 0$), definimos a *lista do nível* $i+1$, denotada S_{i+1}, tomando um subconjunto aleatório de S_i onde cada elemento é escolhido com probabilidade 1/2, com as escolhas independentes. A lista ligada S_{i+1} é ordenada e cada elemento seu está ligado com a sua cópia imediatamente abaixo no nível S_i. No último nível temos $S_\ell = \varnothing$.

Além dos elementos de S mantemos dois sentinelas em cada nível, o $-\infty$ no começo de todas as listas e o $+\infty$ no fim de todas as listas. A Figura 3.3 descreve um exemplo de uma *skip list*. As operações de dicionário na *skip list* são descritas a

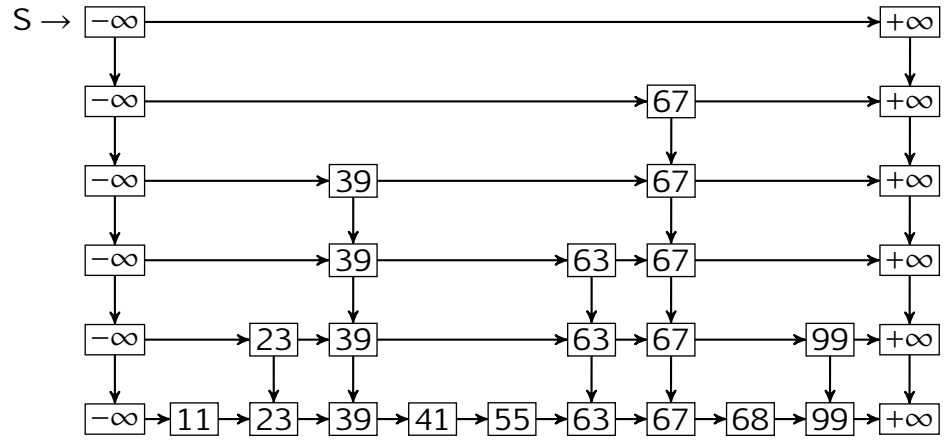

Figura 3.3: exemplo de uma *skip list* com 6 níveis que representa o conjunto $\{11, 23, 39, 41, 55, 63, 67, 68, 99\}$.

seguir.

busca: dado $x \in U$, uma busca devolve um apontador para $x \in S_0$ ou para o menor elemento de S_0 maior que x, caso $x \notin S$. O início da lista é dado por um apontador S para o sentinela $-\infty$ no último nível. A busca começa em S, se o próximo elemento da lista ligada no nível atual é menor ou igual a x então a busca continua nesse nível da lista a partir desse próximo elemento, senão desce um nível se for possível ou termina caso contrário. A Figura 3.4 destaca um exemplo de busca na estrutura da Figura 3.3;

inserção: dado $x \in U$ a inserção de x em S coloca x no lugar apropriado da estrutura caso ele não esteja. Supondo que $x \notin S$, uma busca por x na *skip list* determina a posição do sucessor de x em S_0. A inserção é feita na lista S_0 e

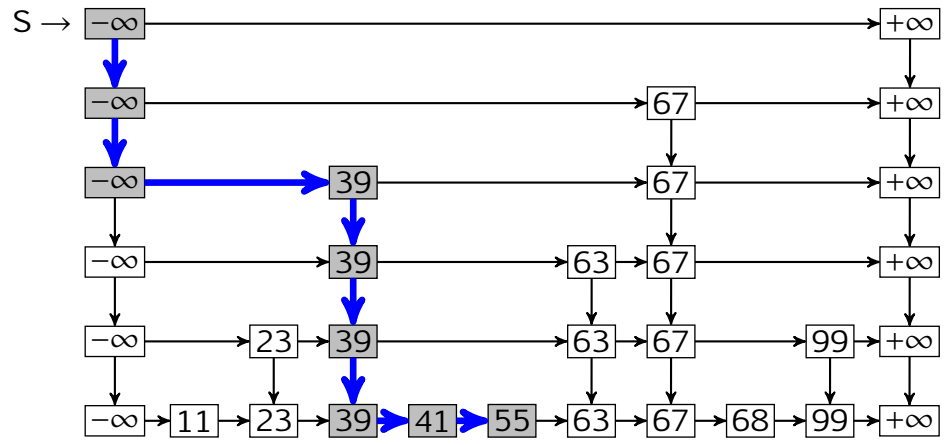

Figura 3.4: o caminho da busca por 55 na *skip list* da Figura 3.3.

em seguida jogamos uma moeda, se der cara replicamos o elemento na lista do nível 1 e jogamos uma moeda para decidir se há inclusão no nível 2, e assim sucessivamente até sair coroa, quando paramos de replicar o elemento novo. Eventualmente, aumentamos o número de níveis da *skip list* numa inserção;

remoção: dado um apontador para $x \in S$, uma remoção retira toda ocorrência desse elemento da estrutura; se o penúltimo nível for uma lista unitária com o elemento removido então o número de níveis diminuirá de um.

O número de comparações realizadas por uma operação de dicionário sobre uma *skip list* depende do número de níveis da estrutura, que é uma variável aleatória. Como S_{i+1} é formado a partir de S_i escolhendo ou não os elementos de S_i com probabilidade 1/2, em média S_{i+1} tem metade do tamanho de S_i, logo esperamos que o número de níveis em S de cardinalidade n seja da ordem de $\log_2(n)$. Como os níveis mais altos são muito esparsos esperamos caminhar pouco na horizontal de modo que, no total, esperamos que a quantidade de passos numa busca seja de ordem logarítmica no tamanho do conjunto. Vamos estudar essa estrutura de dados e para isso começamos definindo alguns parâmetros importantes.

Tomamos $n := |S|$, a quantidade de elementos representados na estrutura de dados. Para todo $x \in S$ definimos a *altura* de x, denotada $h(x)$, como o número de sorteios realizados quando da inserção de x. No exemplo da Figura 3.3, $h(55) = 1$ e $h(63) = 3$. A altura de x também é o número de cópias de x na estrutura. Denotamos por $H = H(S)$ a *altura* da estrutura S dada por

$$H(S) := \max\{h(x) \colon x \in S\}$$

de modo que a *skip list* S é composta pelos níveis $S_0, S_1, \ldots, S_H$. No nosso exemplo $H = h(67) = 5$.

As alturas definidas no parágrafo anterior são variáveis aleatórias. Claramente, $h(x) \sim \text{Geom}(1/2)$ logo $\mathbb{P}[\, h(x) = t \,] = (1/2)^t$ para todo inteiro positivo t. Do Exercício 3.3, página 99, concluímos que $\mathbb{P}[\, h(x) > t \,] = 2^{-t}$. Reunindo essas informações

$$\mathbb{P}[\, h(x) = t \,] = 2^{-t} = \mathbb{P}[\, h(x) > t \,]. \tag{3.29}$$

Podemos computar a esperança da variável aleatória H usando o fato dela assumir valores inteiros positivos. Pela Proposição 3.29, página 121,

$$\mathbb{E}\,\mathsf{H} = \sum_{i \geqslant 1} \mathbb{P}[\,\mathsf{H} \geqslant i\,] = \sum_{i=0}^{\lfloor \log_2 n \rfloor} \mathbb{P}[\,\mathsf{H} > i\,] + \sum_{i > \lfloor \log_2 n \rfloor} \mathbb{P}[\,\mathsf{H} > i\,].$$

Na primeira soma, no lado direto dessa equação, limitamos trivialmente a probabilidade tomando $\mathbb{P}[\,\mathsf{H} > i\,] \leqslant 1$, de modo que

$$\sum_{i=0}^{\lfloor \log_2 n \rfloor} \mathbb{P}[\,\mathsf{H} > i\,] \leqslant \lfloor \log_2 n \rfloor + 1$$

e na segunda soma, para cada i

$$\mathbb{P}[\,\mathsf{H} > i\,] \leqslant \mathbb{P}\left[\bigcup_{x \in \mathsf{S}} [h(x) > i] \right] \leqslant \sum_{x \in \mathsf{S}} \mathbb{P}[\, h(x) > i\,] \leqslant n \left(\frac{1}{2}\right)^i$$

logo

$$\begin{aligned}
\mathbb{E}\,\mathsf{H} &\leqslant \lfloor \log_2 n \rfloor + 1 + \sum_{i > \lfloor \log_2 n \rfloor} n \left(\frac{1}{2}\right)^i \\
&\leqslant \lfloor \log_2 n \rfloor + 1 + n \left(\sum_{i \geqslant 0} \left(\frac{1}{2}\right)^i - \sum_{i=0}^{\lfloor \log_2 n \rfloor} \left(\frac{1}{2}\right)^i \right) \\
&= \lfloor \log_2 n \rfloor + 1 + n \left(2 - \frac{1 - (1/2)^{\lfloor \log_2 n \rfloor + 1}}{1 - (1/2)} \right) \\
&= \lfloor \log_2 n \rfloor + 1 + n \left(\frac{1}{2}\right)^{\lfloor \log_2 n \rfloor} \\
&< \lfloor \log_2 n \rfloor + 3
\end{aligned}$$

ou seja, a estrutura tem altura esperada logarítmica no número de elementos de S, como numa árvore balanceada de busca.

Mais que isso, altura é logarítmica no número de elementos de S com alta probabilidade. Da equação (3.29) deduzimos que a probabilidade de um elemento qualquer de S ter altura maior que $2 \log_2 n$ é $2^{-2 \log_2 n} = n^{-2}$, portanto, a probabilidade de existir algum elemento em S com altura maior que $2 \log_2 n$ é

$$\mathbb{P}\left(\bigcup_{x \in \mathsf{S}} [h(x) > 2 \log_2 n] \right) \leqslant \sum_{x \in \mathsf{S}} \mathbb{P}[\, h(x) > 2 \log_2 n\,] = n \frac{1}{n^2} = \frac{1}{n}$$

ou seja, $\mathbb{P}[\,\mathsf{H} < 2\,\mathbb{E}\,\mathsf{H}\,] = 1 - o(1)$.

Proposição 3.48 *Se* S *é uma* skip list *para um conjunto com* n *elementos então para a altura* $\mathsf{H} = \mathsf{H}(\mathsf{S})$ *valem*

$$\mathbb{E}\,\mathsf{H} < \log_2(n) + 3 \qquad e \qquad \mathbb{P}[\,\mathsf{H} > 2\log_2(n)\,] < 1/n.$$

Observação 3.49 (*sobre o tamanho da estrutura*). Se $\mathsf{N} := \sum_{i=0}^{\mathsf{H}} |\mathsf{S}_i|$ é o tamanho da *skip list*, então de $\mathsf{N} = \sum_{x\in\mathsf{S}} h(x)$. Da linearidade da esperança e de $\mathbb{E}[\,h(x)\,] = 2$

$$\mathbb{E}\,\mathsf{N} = \sum_{x\in\mathsf{S}} \mathbb{E}[\,h(x)\,] = 2n. \tag{3.30}$$

Notemos que de $\mathsf{N} = \sum_{i=1}^{\mathsf{H}} |\mathsf{S}_i|$ não é imediato valer que $\mathbb{E}\,\mathsf{N} = \sum_{i=1}^{\mathbb{E}\,\mathsf{H}} \mathbb{E}\,|\mathsf{S}_i|$ por causa de independência ou não das variáveis envolvidas (veja o Exercício 3.65 no final deste capítulo). Pode-se provar que o número de itens N é $O(n)$ com alta probabilidade e faremos isso na seção 6.2.1. ◊

Para determinar o número de passos esperados numa busca sobre uma *skip list* nós vamos limitar número de comparações feitas durante uma busca em dois casos: as comparações feitas nos níveis mais altos e as comparações feitas nos níveis mais baixos da *skip list*. Nos níveis mais altos o número de comparações é no máximo a quantidade total de elementos nesses níveis.

Proposição 3.50 *O número esperado de comparações feitas nos níveis mais altos de uma* skip list, *do nível* $\lfloor \log_2 n \rfloor$ *até o nível* H, *durante uma busca é* $O(1)$.

Demonstração. Vamos mostrar que esses níveis contribuem pouco com o custo da busca limitando o custo da busca nesses níveis pelo número total de elementos neles. Façamos

$$\ell := \lfloor \log_2 n \rfloor, \quad p' := (1/2)^\ell \quad \text{e} \quad \mathsf{N}' := \sum_{i=\ell}^{\mathsf{H}} |\mathsf{S}_i|.$$

Notemos que em $\mathsf{S}_\ell, \mathsf{S}_{\ell+1}, \mathsf{S}_{\ell+2}, \ldots, \mathsf{S}_{\mathsf{H}}$ temos uma *skip list* para o conjunto S_ℓ logo pela equação (3.30), $\mathbb{E}[\,\mathsf{N}' \,|\, |\mathsf{S}_\ell| = k\,] \leqslant 2k$. Usando o teorema da esperança total, equação (3.28) na página 140, escrevemos

$$\mathbb{E}\,\mathsf{N}' = \sum_{k=0}^{n} \mathbb{E}[\,\mathsf{N}' \,\big|\, |\mathsf{S}_\ell| = k\,]\mathbb{P}[\,|\mathsf{S}_\ell| = k\,]$$

e como qualquer $x \in \mathsf{S}$ ocorre em S_ℓ com probabilidade p', temos $|\mathsf{S}_\ell| \sim \text{binomial}(n, p')$ de modo que

$$\mathbb{E}\,\mathsf{N}' = \sum_{k=0}^{n} \mathbb{E}[\,\mathsf{N}' \,\big|\, |\mathsf{S}_\ell| = k\,]\binom{n}{k} p'^k (1-p')^{n-k} \leqslant 2\sum_{k=0}^{n} k\binom{n}{k} p'^k (1-p')^{n-k}.$$

O somatório corresponde ao valor esperado de uma variável aleatória binomial com parâmetros n e p', então $\mathbb{E}\,\mathsf{N}' \leqslant 2np'$. Ainda,

$$np' = n\left(\frac{1}{2}\right)^\ell = n\left(\frac{1}{2}\right)^{\lfloor \log_2 n \rfloor} \leqslant 2$$

portanto $\mathbb{E}\,N' \leqslant 4$, ou seja, mesmo que uma busca percorra todos os elementos dos níveis $S_\ell, \ldots, S_H$, o número esperado de comparações é limitado superiormente por uma constante. □

O segundo passo da nossa análise é estimar o número de comparações nos níveis mais baixos onde se concentram a maior parte dos elementos. Nesse caso, a probabilidade da busca ficar muito tempo num mesmo nível é pequena, portanto, o tempo gasto é essencialmente determinado pela altura da *skip list*. A ideia chave é explicada no próximo parágrafo.

Consideremos o caminho reverso de uma busca, do posição final para o início. Cada passo é um passo para a esquerda ou para cima. O caminho de uma busca desce um nível na *skip list* sempre que o próximo nó do mesmo nível é maior que o elemento buscado, o que levaria "além" do nó procurado. Assim, quando consideramos o caminho de busca reverso, *ele sempre dará um passo para cima se puder, caso contrário, ele dará um passo para a esquerda.*

De S_0 até S_ℓ, são $\ell = \ell(n) = \lfloor \log_2 n \rfloor$ passos para cima (sucesso). Cada passo para cima é precedido por 0 ou mais passos para a esquerda (fracasso). O número total de passos para cima ou para a esquerda de S_0 até S_ℓ é $\ell(n) + L$, onde $L \sim \mathrm{BN}(\ell(n), 1/2)$.

Uma variável aleatória que conta o número de fracassos até ocorrer o k-ésimo sucesso em uma sequência de experimentos independentes de Bernoulli com parâmetro p tem distribuição $\mathrm{BN}(k, p)$, que chamamos de *distribuição binomial negativa*, com parâmetros $k \in \mathbb{Z}^+$ e $p \in (0, 1)$, cuja lei é:

$$\mathrm{nb}_{k,p}(t) = \binom{t+k-1}{t}(1-p)^t p^k, \quad t = 0, 1, \ldots.$$

Notemos que o caso $k = 1$ é a lei da distribuição geométrica. O valor esperado da variável binomial negativa é $(k/p) - k$.

O custo da busca nos níveis mais baixos da *skip list* é

$$\mathbb{E}[\,\ell(n) + L\,] = \ell(n) + \mathbb{E}[\,L\,] = \ell(n) + \frac{\ell(n)}{1/2} - \ell(n) = 2\ell(n) = O(\log n)$$

e o custo nos outros níveis da estrutura é $O(1)$, logo o custo total é $O(\log n)$.

Teorema 3.51 *O número esperado de comparações feitas por uma busca em uma* skip list *que representa um conjunto com n elementos é* $O(\log n)$. □

No Teorema 6.35, página 187, provamos que o tempo de busca é como dito no teorema acima com alta probabilidade.

3.5 EXERCÍCIOS

Exercício 3.52. Considere os eventos A e B de um espaço de probabilidade. Verifique as identidades

1. $\mathbb{1}_{A\cup B} = \mathbb{1}_A + \mathbb{1}_B - \mathbb{1}_{A\cap B}$.
2. $\mathbb{1}_{\overline{A}} = 1 - \mathbb{1}_A$.
3. $A \subseteq B \Rightarrow \mathbb{1}_A \leqslant \mathbb{1}_B$.
4. $\mathbb{1}_{A\cap B} = \mathbb{1}_A \cdot \mathbb{1}_B$.

Exercício 3.53 (variáveis geométricas não têm memória). Sejam $X \sim \mathrm{Geom}(p)$ uma variável aleatória, $t \geqslant 0$ e $n \geqslant 1$ dois números inteiros. Prove que $\mathbb{P}[\, X = n+t \mid X > t \,] = \mathbb{P}[X = n]$ e que $\mathbb{P}[\, X \geqslant n+t \mid X > t \,] = \mathbb{P}[X \geqslant n]$.

Exercício 3.54. Deduza a lei de uma variável aleatória binomial negativa dado que ela conta o número de fracassos até ocorre o r-ésimo sucesso em qualquer ssequência de experimentos de Bernoulli independentes e mesma lei. Deduza também o valor esperado.

Exercício 3.55. Prove as seguintes propriedades para variáveis aleatórias discretas.

1. Se $\mathbb{P}[X \leqslant Y] = 1$ então $\mathbb{E}\, X \leqslant \mathbb{E}\, Y$.
2. Se $\mathbb{P}[\, a \leqslant X \leqslant b \,] = 1$ então $a \leqslant \mathbb{E}\, X \leqslant b$.
3. Se $X \geqslant 0$ e $\mathbb{E}\, X = 0$ então $\mathbb{P}[X = 0] = 1$.

Exercício 3.56. Considere o espaço produto $(\Omega, \mathbb{P})$ obtido de $(\Omega_i, \mathbb{P}_i)$ para $1 \leqslant i \leqslant n$. Seja X_i a variável aleatória "projeção na i-ésima coordenada" definida em Ω dada por $X_i(\omega_1, \ldots, \omega_n) = \omega_i$. Prove que $X_1, \ldots, X_n$ são mutuamente independentes e que $\mathbb{P}_{X_i} = \mathbb{P}_i$.

Exercício 3.57. Prove que $X_1, \ldots, X_n$ são variáveis aleatórias independentes de Ω em S se, e somente se, a distribuição do vetor aleatório $(X_1, \ldots, X_n)$ sobre S^n é uma medida produto.

Exercício 3.58. Seja π uma permutação de $\{1, 2, \ldots, n\}$. Para todos $1 \leqslant i < j \leqslant n$ o par (i, j) determina um inversão de π se $\pi(i) > \pi(j)$. Determine o número esperado de inversões numa permutação escolhida ao acaso.

Exercício 3.59. Cada um dos $n \geqslant 1$ convidados de uma festa entrega seu chapéu na chapelaria da recepção. Quando a festa acaba o recepcionista devolve os chapéus

aos convidados em uma ordem aleatória. Qual é o número esperado de convidados que recebem seu próprio chapéu de volta?

Exercício 3.60. Nesse exercício vamos deduzir a esperança de uma variável binomial sem recorrer a linearidade da esperança. Primeiro prove que vale a seguinte identidade entre coeficientes binomiais $i\binom{n}{i} = n\binom{n-1}{i-1}$. Em seguida, use essa identidade para provar que $\mathbb{E}[\,Y^k\,] = np\mathbb{E}[\,(X+1)^{k-1}\,]$ vale para $X \sim b(n-1,p)$ e $Y \sim \text{binomial}(n,p)$, com $n > 0$ e $p \in (0,1)$. Conclua que $\mathbb{E}\,Y = np$.

Exercício 3.61. Sejam U e M conjuntos finitos. Uma família de funções $\mathcal{H} \subseteq M^U$ que quando munida da medida uniforme satisfaz

$$\mathop{\mathbb{P}}_{h \in_R \mathcal{H}}[h(u) = i] = \frac{1}{|M|}, \text{ para todo } u \in U \text{ e para todo } i \in M \tag{3.31}$$

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 $|M|$ funções constantes satisfaz (3.31).

Exercício 3.62. Projete um algoritmo aleatorizado que recebe uma lista de n números e devolve o k-ésimo maior elemento da lista. O número esperado de comparações deve ser $\leqslant cn$ para alguma constante positiva c (*dica:* use o particionamento do *quicksort*).

Exercício 3.63. Dado um conjunto de n números e um real positivo ε, uma ε-aproximação da mediana é um elemento cujo posto (posição considerando os mesmo elementos em ordem crescente) está no intervalo $[(1-\varepsilon)n/2, (1+\varepsilon)n/2]$. Projete um algoritmo aleatorizado para computar uma 1/2-aproximação da mediana de um vetor com n elementos com tempo de execução $O(\log n \log\log n)$ e probabilidade de erro $2n^{-2}$ (*dica*: amostre e ordene).

Exercício 3.64. Para dados ε e c positivos, projete um algoritmo aleatorizado para computar uma ε-aproximação da mediana de um vetor com n elementos com tempo de execução $O(\log n \log\log n)$ e probabilidade de erro n^{-c}.

Exercício 3.65. Sejam $X_i \sim \text{Bernoulli}(1/2)$, para $i \geqslant 1$, variáveis aleatórias independentes e $N = \min\{n \geqslant 1 : X_{n+1} = 1\}$. Prove que $N + 1 \sim \text{Geom}(1/2)$ e que $\mathbb{E}\,N = 1$. Assuma que se $N = 0$ então $X_1 + \cdots + X_N = 0$ e prove que

$$\mathbb{E}\sum_{i=1}^{N} X_i \neq \sum_{i=1}^{\mathbb{E}\,N} \mathbb{E}\,X_i\,.$$

Exercício 3.66. Verifique que a hipótese de convergência é necessária no Exercício 3.36, página 124. Para tal, considere o espaço de probabilidade $(\mathbb{N}, \mathbb{P})$ com $\mathbb{P}(n) = 2^{-n}$ e as variáveis aleatórias

$$X_n(j) := \begin{cases} 2^n, & \text{se } j = n, \\ -2^{n+1}, & \text{se } j = n+1, \\ 0, & \text{caso contrário.} \end{cases}$$

1. Mostre que $\mathbb{E}\, X_n = 0$ para todo $n \geqslant 1$.

2. Tome $X = \sum_{n \geqslant 1} X_n$. Determine $X(1), X(2), X(3), \ldots$.

3. Usando o item anterior determine $\mathbb{E}\, X = \sum_{n \geqslant 1} X(n)\mathbb{P}(n)$.

Conclua que $\mathbb{E} \sum_n X_n \neq \sum_n \mathbb{E}\, X_n$.

Exercício 3.67. Na seção 3.1.3, página 112, foi dito que se uma tabela de espalhamento usa uma função aleatória escolhida numa família universal, então o número esperado de colisões para S fixo é $\binom{n}{2}(1/m) = n(n-1)/(2m)$ de modo que se $m = n^2$ então $\mathbb{E}\, C < 1/2$ e não há colisão com probabilidade pelo menos 1/2. Portanto, se S é um conjunto fixo (estático) podemos sortear uma função até que encontremos uma que não ocasiona colisões para elementos de S. Escreva um algoritmo aleatorizado que dado S encontra uma função de *hash h perfeita*. Determine a complexidade desse algoritmo. É possível usar o método das esperanças condicionais nesse caso?

Exercício 3.68. Prove que para cada inteiro $n \geqslant 4$ existe uma coloração das arestas do grafo completo com n vértices com duas cores de modo que o número total de cópias monocromáticas de um grafo completo com 4 vértices é no máximo $\binom{n}{4}2^{-5}$. Escreva um algoritmo aleatorizado de tempo polinomial em n que descobre uma coloração como descrita acima. Mostre como construir tal coloração deterministicamente em tempo polinomial usando o método de esperanças condicionais.

Exercício 3.69. Um **conjunto independente** em um grafo $G = (V, E)$ é um subconjunto de vértices $U \subseteq V$ tal que não há arestas de E formada só por vértices de U, isto é, U não contêm os dois vértices de qualquer aresta do grafo. Por exemplo, no grafo da Figura 1.8, na página 73, $\{2, 6\}$ é um conjunto independente, assim como $\{2, 5\}$, entretanto $\{1, 2, 5\}$ e $\{2, 5, 6\}$ não são conjuntos independentes.

Considere o seguinte algoritmo para computar um conjunto independente: dado $G = (V, E)$ com n vértices e $m := |E| > 0$;

1. inicie com U vazio;

2. para cada $v \in V$, acrescente v a U com probabilidade p;

3. para cada $e \in E$ com ambos extremos em U, remova um dos extremos de U.

Determine um limitante inferior para a cardinalidade esperada para U em função de p, n e m. Conclua que, para uma escolha ótima de p, todo grafo tem um conjunto independente de tamanho pelo menos $n^2/(4m)$ vértices. Escreva e analise um algoritmo Las Vegas para determinar um conjunto independente de cardinalidade maior ou igual ao valor acima.

Exercício 3.70. Considere o seguinte algoritmo para computar um conjunto independente: dado $G = (V, E)$;

1. inicie com U vazio;
2. tome um permutação π de V;
3. percorra os vértices de G na ordem definida pela permutação π e para cada $v \in V$, se v não tem vizinhos em U acrescente v a U.

Prove que U construído dessa maneira é independente em G. Escreva um algoritmo aleatorizado baseado na ideia acima e que determina um conjunto independente cuja cardinalidade esperada é

$$\sum_{u \in V} \frac{1}{d(u)+1}$$

em que $d(u)$ é o grau do vértice u em G.

Exercício 3.71. Prove que numa *skip list*,

$$\mathbb{P}[\, H > 100 \log_2(n) \,] \leqslant \frac{1}{n^{99}}.$$

Exercício 3.72. Prove as seguintes identidades para esperança condicional. Se X, Y e Z são variáveis aleatórias sobre um espaço de probabilidade discreto então

$$\mathbb{E}[\, X \mid Z \,] = \mathbb{E}[\, \mathbb{E}[\, X \mid Y, Z \,] \mid Z \,].$$

Ademais, se f e g são funções de variáveis reais

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

Exercício 3.73. Sejam X e Y variáveis aleatórias discretas e reais. Prove que para qualquer função $h \colon Y(\Omega) \to \mathbb{R}$

$$\mathbb{E}\, (X - \mathbb{E}[\, X \mid Y \,])^2 \leqslant \mathbb{E}\, (X - h(Y))^2$$

ou seja, $\mathbb{E}[\, X \mid Y \,]$ é a função de Y que melhor aproxima X no sentido de ter o menor erro quadrático médio. Agora, prove que vale a igualdade se, e só se, existe uma função $g \colon Y(\Omega) \to \mathbb{R}$ tal que $g(Y) = \mathbb{E}[\, X \mid Y \,]$ com probabilidade 1.

Exercício 3.74. Escreva um algoritmo Las Vegas que recebe uma fórmula booleana CNF $C_1 \wedge \cdots \wedge C_m$ em que cada cláusula tenha exatamente k literais e encontra uma valoração que satisfaz $m(1 - 2^{-k})$ cláusulas. Escreva uma versão desaleatorizada usando o método da esperança condicional. Analise o tempo de execução dos algoritmos.

Exercício 3.75. Um **grafo dirigido** é definido por um par de conjuntos finitos (V, E) em que V é o conjunto de vértices do grafo e E o conjunto de arestas, cada aresta é um par $(u, v) \in V \times V$. Dizemos que x **alcança** y **em** G se existe um caminho dirigido[8] dirigido de x para y no grafo G.

Dada uma fórmula booleana 2-CNF Φ, defina o grafo dirigido $G = G(\Phi)$ com $2n$ vértices dados por x e $\neg x$ para cada variável x de Φ e cada cláusula $C = \ell_1 \vee \ell_2$ contribui com exatamente duas arestas: $(\neg\ell_1, \ell_2)$ e $(\neg\ell_2, \ell_1)$.

Demonstre a seguinte afirmação: *A fórmula 2-CNF Φ não é satisfazível se, e somente se, existe uma variável x de Φ tal que x alcança $\neg x$ e $\neg x$ alcança x.*

Use a afirmação acima para projetar um algoritmo eficiente para 2SAT.

Exercício 3.76 (lei de grandes desvios para variáveis aleatórias binomiais). Sejam $X_i \sim$ Bernoulli(1/2) variáveis aleatórias independentes $1/2 < \alpha \leqslant 1$ um real. Seja $S_n = X_1 + \cdots + X_n$. Verifique que

$$\frac{1}{2^n} \frac{n!}{\lceil \alpha n \rceil! (n - \lceil \alpha n \rceil)!} \leqslant \mathbb{P}[S_n \geqslant \alpha n] \leqslant \frac{n+1}{2^n} \frac{n!}{\lceil \alpha n \rceil! (n - \lceil \alpha n \rceil)!}$$

e conclua, usando a aproximação de Stirling, que

$$\lim_{n \to \infty} \frac{1}{n} \mathbb{P}[S_n \geqslant \alpha n] = -I(\alpha)$$

onde $I(\alpha) = \alpha \log \alpha + (1 - \alpha) \log(1 - \alpha) + \log 2$.

Exercício 3.77. Suponha que temos uma fonte de bits aleatórias que responde 0 com probabilidade $p \in (0, 1)$ e responde 1 com probabilidade $1 - p$, independentemente. Escreva um algoritmo que usa essa fonte responda 0 ou 1 uniforme e independentemente. Calcule o tempo esperado de execução em função de p.

Exercício 3.78. Prove que se distribuímos ao acaso n bolas em m caixas com a condição que $n < (2/e) m \ln(m)$, então com alta probabilidade o maior número de bolas em uma caixa é

$$\frac{4 \ln(n)}{\ln\left(\frac{2n}{me} \ln(n)\right)}.$$

[8] Um caminho dirigido é dado por uma sequência de vértices $x = v_0, v_1, v_2, \ldots, v_k = y$ em que vértices consecutivos formam aresta, $(v_i, v_{i+1}) \in E$.

Exercício 3.79 (Raab e Steger, 1998). Assuma que n bolas são guardadas ao acaso em n caixas. Seja X a quantia de caixas com pelo k bolas e X_i variável aleatória indicadora de ocorrência de mais que k bolas na caixa i, para $i = 1, 2, \ldots, m$.

1. Prove que $\mathbb{E}\, X_i = (1 + o(1))b_{n,\frac{1}{n}}(k)$. Uma sugestão é que $EX_i = \mathbb{P}[X_i = 1] = \sum_{x=k}^{n} b_{n,\frac{1}{n}}(x)$ e a soma acima pode ser estimada da seguinte maneira

$$\sum_{x=k}^{n} b_{n,\frac{1}{n}}(x) = b_{n,\frac{1}{n}}(k)\left(1 + \frac{b_{n,\frac{1}{n}}(k+1)}{b_{n,\frac{1}{n}}(k)} + \frac{b_{n,\frac{1}{n}}(k+2)}{b_{n,\frac{1}{n}}(k+1)}\frac{b_{n,\frac{1}{n}}(k+1)}{b_{n,\frac{1}{n}}(k)} + \cdots + \frac{b_{n,\frac{1}{n}}(n)}{b_{n,\frac{1}{n}}(n-1)}\frac{b_{n,\frac{1}{n}}(n-1)}{b_{n,\frac{1}{n}}(n-2)}\cdots\frac{b_{n,\frac{1}{n}}(k+1)}{b_{n,\frac{1}{n}}(k)}\right)$$

   e se $x \geqslant k$ então $\frac{b_{n,\frac{1}{n}}(x+1)}{b_{n,\frac{1}{n}}(x)} = \varepsilon(n,k) < 1$ (ε não depende de x).

2. Prove, usando as informações em (d.2) e (d.3) do Apêndice, que se tomamos $k = \alpha \ln(n)/\ln(\ln(n))$ então $\mathbb{E}\, X = n^{1-\alpha+o(1)}$ e conclua que

$$\mathbb{E}\, X \to \begin{cases} \infty & \text{se } 0 < \alpha < 1 \\ 0 & \text{se } \alpha > 1 \end{cases}.$$

Exercício 3.80. Sejam $\lambda > 0$ real e (p_n) uma sequência de números reais em $[0,1]$ tal que np_n converge para λ quando $n \to \infty$. Prove que

$$\lim_{n\to\infty} \binom{n}{k} {p_n}^k (1-p_n)^{n-k} = \frac{e^{-\lambda}\lambda^k}{k!}.$$