Probabilidade com Algoritmos

e vice-versa

Uma introdução à probabilidade discreta e aos algoritmos probabilísticos

— Jair Donadelli —

última modificação 12/2/2020

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

[...]

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

[...]

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

[...]

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

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

3 | VARIÁVEIS ALEATÓRIAS

Uma função que modela alguma grandeza observável em um fenômeno aleatório como, por exemplo, o número de lançamentos de uma moeda até sair cara, e o tempo de vida de uma lâmpada escolhida na linha de produção, é uma *variável aleatória*. No caso mais geral, uma *variável aleatória* em um espaço de probabilidade $(\Omega, \mathcal{A}, \mathbb{P})$ é uma função $X\colon (\Omega, \mathcal{A}) \to (S, \mathcal{B})$ definida em Ω e que assume valores num conjunto S munido de uma σ-álgebra $\mathcal{B}$ e tal que $X^{-1}(B) \in \mathcal{A}$ para todo $B \in \mathcal{B}$. Uma vantagem dessa função é que ela induz uma medida de probabilidade no contradomínio e assim podemos usar X para associar a estrutura abstrata $(\Omega, \mathcal{A})$ do modelo probabilístico a uma estrutura conhecida como, por exemplo, $(\mathbb{R}, \mathcal{B})$ com $\mathcal{B}$ gerado pelos intervalos abertos da reta e trabalhar com a medida de probabilidade induzida por X nesse espaço conhecido. Essa abordagem funcional de probabilidade proporciona ferramentas poderosas para o cálculo de probabilidades.
Neste capítulo estudaremos as variáveis aleatórias *discretas*, isto é, da forma $X\colon \Omega \to S$ para a qual existe $B = \{x_1, x_2, \ldots\} \subset S$ enumerável com $\mathbb{P}(\{\omega\colon X(\omega) \in B\}) = 1$. Veremos, formalmente, o conceito de media ponderada desses valores, a esperança matemática, e uma introdução aos primeiro e segundo momentos e uma técnica de desaleatorização de algoritmos, genérica mas que depende de computação eficiente de médias condicionadas.

3.1 VARIÁVEIS ALEATÓRIAS DISCRETAS

Uma função definida sobre um espaço amostral de um espaço de probabilidade discreto é uma **variável aleatória discreta**. Na maior parte deste texto estaremos interessados em variáveis aleatórias com contradomínio no conjunto dos números reais e uma variável desse tipo é chamada de **variável aleatória real** de modo que na maioria das ocorrências teremos uma variável aleatória discreta e real e omitiremos os adjetivos "discreta" e "real" ao nos referirmos às variáveis aleatórias. Usaremos, em geral, as letras maiúsculas finais do alfabeto, como por exemplo X, Y, Z, W, para denotar as variáveis aleatórias.

Um exemplo trivial de variável aleatória é dado por uma função constante, por exemplo, quando para todo $\omega \in \Omega$ há $c \in \mathbb{R}$ tal que $X(\omega) = c$. No modelo clássico para o lançamento de um dado equilibrado a variável aleatória $X\colon \{1,2,3,4,5,6\} \to \{1,2,3,4,5,6\}$ dada por $X(\omega) = \omega$, para todo ω, é o resultado do lançamento. A resposta do algoritmo

1 e o tempo de execução do algoritmo 2 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 $(\Omega, \mathbb{P})$, 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** do evento A. Toda variável aleatória discreta e real pode ser representada em termos de variáveis aleatórias indicadoras. Para escrever essa representação nós tomamos uma enumeração qualquer $x_1, x_2, x_3, \ldots$ dos valores que a variável aleatória X assume e definimos uma partição do espaço amostral definida pelos eventos $A_n := \{\omega \in \Omega \colon X(\omega) = x_n\}$ de modo que

$$X(\omega) = \sum_{n \geqslant 1} x_n \mathbb{1}_{A_n}(\omega)$$

para todo $\omega \in \Omega$. ◊

Distribuição Seja $X \colon \Omega \to S$ uma variável aleatória do modelo probabilístico discreto $(\Omega, \mathbb{P})$. Denotamos por $X(\Omega)$ a imagem da variável aleatória $X \colon \Omega \to S$ e definimos para cada $B \subset X(\Omega)$ o evento

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

e $[X = t] := [X \in \{t\}] = X^{-1}(t)$. A função X não é necessariamente invertível, embora possamos fazê-la sobrejetiva restringindo S à imagem ela pode não ser injetiva. No entanto, X^{-1} é bem definida para subconjuntos.

A **distribuição de** X, ou **lei** de X, é a medida de probabilidade $\mathbb{P}_X$ definida em $2^{X(\Omega)}$ por

$$\underset{X}{\mathbb{P}}(t) := \mathbb{P}[X = t]$$

para todo $t \in X(\Omega)$ e estendida para todo subconjunto de $X(\Omega)$ da maneira usual de modo que vale

$$\underset{X}{\mathbb{P}}(B) = \mathbb{P}[X \in B], \text{ para todo } B \subset X(\Omega).$$

Claramente $0 \leqslant \mathbb{P}_X(B) \leqslant 1$. Ainda $\mathbb{P}_X(S) = \mathbb{P}(\Omega) = 1$ e vale a aditividade

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

Além disso, consideramos $(S, \mathbb{P}_X)$ com a medida de probabilidade sendo uma extensão[1] da definição acima para todo $B \subset S$ dada por $\mathbb{P}_X(B) = \mathbb{P}[X \in B \cap X(\Omega)]$.

Por exemplo, se X é o resultado do lançamento de um dado equilibrado, então a lei de X é a medida uniforme $\mathbb{P}_X(\omega) = 1/6$, para todo $\omega \in \{1,2,3,4,5,6\}$. Se A é um evento de um modelo probabilístico $(\Omega, \mathbb{P})$ então a lei de $\mathbb{1}_A$ é $\mathbb{P}_{\mathbb{1}_A}$ dada por $\mathbb{P}_{\mathbb{1}_A}(1) = \mathbb{P}(A)$ e $\mathbb{P}_{\mathbb{1}_A}(0) = 1 - \mathbb{P}(A)$.

Exemplo 3.2 (*distribuição de Bernoulli*). É comum ocorrerem situações com experimentos que interessam apenas 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 a afirmação feita; a condição de um laço num algoritmo é *verdadeira* ou *falsa*; um evento A *ocorreu* ou *não ocorreu*.

[1] Eventualmente temos $S = \mathbb{R}$, porém não estamos considerando aqui espaços contínuos o que torna essa consideração um abuso de notação.

Esses experimentos recebem o nome de **ensaio de Bernoulli**. Nessas situações podemos modelar a observação com uma variável aleatória $X\colon \Omega \to \{0,1\}$ e definimos

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

em que $p = \mathbb{P}[X = 1]$ é a probabilidade de *sucesso*. Dizemos que X tem **distribuição de Bernoulli** com parâmetro p se $\mathbb{P}_X(x) = b_p(x)$ e usamos a notação $X \in_{b(p)} \{0,1\}$ para indicar tal fato. ◊

De um modo geral, se $\mathcal{D}$ associa a cada elemento $t \in S$ um real não-negativo $\mathcal{D}(t)$, então chamamos $\mathcal{D}$ uma **distribuição** de probabilidade sobre o conjunto enumerável S se $\sum_t \mathcal{D}(t) = 1$. A variável aleatória $X\colon \Omega \to S$ tem distribuição $\mathcal{D}$ se $\mathbb{P}_X(t) = \mathcal{D}(t)$ para todo $t \in S$ e nesse caso escrevemos

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

e escrevemos $x \in_{\mathcal{D}} S$ para dizer que $x \in S$ é um elemento fixo de S escolhido de acordo com a distribuição $\mathcal{D}$.

Exemplo 3.3 (*distribuição uniforme*). 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 $X \in_{\mathcal{U}} S$. ◊

Exemplo 3.4 (*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 que reproduzimos abaixo

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

$$\mathbb{P}_X(t) = (1-p)^{t-1}p, \text{ para todo } t \geqslant 1$$

e dizemos que X tem **distribuição geométrica** com parâmetro p, o que denotamos por $X \in_{\mathcal{G}(p)} \mathbb{N}$. ◊

EXERCÍCIO 3.5. Mostre que se Z tem distribuição geométrica com parâmetro p então $\mathbb{P}[Z > n-1] = (1-p)^{n-1}$ para todo $n \geqslant 1$.

Exemplo 3.6 (*distribuição binomial*). Em n ensaios de Bernoulli idênticos e independentes uma resposta $(b_1, b_2, \ldots, b_n)$ ocorre com probabilidade $p^t(1-p)^{n-t}$ sempre que ocorrerem t sucessos. A probabilidade de ocorrerem exatamente t sucessos é

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

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

Proposição 3.7 *Seja* $k := \lfloor (n+1)p \rfloor$. *A função* $b_{n,p}(t)$ *é crescente em* $\{0,1,\ldots,k\}$ *e é decrescente em* $\{k+1, k+2, \ldots, n\}$.

Demonstração. A razão de valores sucessivos 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, com que probabilidade ocorrem exatamente n caras? O número de caras é uma variável aleatória binomial $\mathrm{X} \in_{\mathrm{b}(1/2)} \{0, 1, \ldots, 2n\}$ de modo que

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

usando a aproximação de Stirling (veja (d.3)). É um probabilidade bem pequena, para $n = 50$ temos $\approx 0{,}08$ e para $n = 130$ temos $\approx 0{,}05$. Porém, de fato, 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

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

além disso, para k como na proposição acima

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

Usando a aproximação de Stirling ingenuamente

$$\binom{n}{k} \approx \frac{n^{n+\frac{1}{2}}}{(n-k)^{n-k+\frac{1}{2}}k^{k+\frac{1}{2}}}$$

tomando logaritmo e dividindo por n

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

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

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

Esse resultado pode ser feito rigoroso (veja o exercício 3.70) e veremos uma demonstração alternativa e mais geral (veja o exemplo 4.4). 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.

Exemplo 3.8 (*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. Começemos com o seguinte exemplo do célebre *Introdução a Probabilidade* de 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. Se n_k é o número de regiões que receberam k bombas, a contagem foi

k	0	1	2	3	4	5 ou mais
n_k	229	211	93	35	7	1

ao qual o modelo de Poisson se ajusta 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 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 num livro 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

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

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 3.71 no final deste capítulo). 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. Tomamos o limite quando $n \to \infty$ assumindo que $p = p(n)$ é tal que $p \cdot n = \lambda$ e 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. Por isso, $\mathrm{Poi}_\lambda(k)$ pode ser usada como uma aproximação da distribuição binomial $\mathrm{b}_{n,p}(k)$ se n for suficientemente grande e p suficientemente pequena.

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*. ◊

Para uma variável aleatória real $\mathsf{X}\colon \Omega \to \mathbb{R}$ os eventos do espaço amostral Ω definidos por

$$[\mathsf{X} \leqslant r] := \{\omega \in \Omega \colon \mathsf{X}(\omega) \leqslant r\}$$

para qualquer $r \in \mathbb{R}$, são particularmente importantes no estudo de variáveis aleatórias mais geralmente. Eles descrevem completamente o comportamento da variável aleatória, mesmo no caso geral, além do discreto. A função $\mathsf{F}_\mathsf{X}(r) := \mathbb{P}[\mathsf{X} \leqslant r]$ é chamada **função de distribuição acumulada** de X. De modo análogo nós definimos os eventos como $[\mathsf{X} < r]$ e $[\mathsf{X} \geqslant r]$.

Distribuição conjunta e Independência Uma variável aleatória discreta $\mathsf{Z}\colon \Omega \to \mathbb{R}^n$ $(n > 1)$ é chamada de **vetor aleatório** e usamos a notação $\mathsf{Z} = (\mathsf{X}_1, \ldots, \mathsf{X}_n)$, nesse caso cada coordenada é uma variável aleatória. Sua distribuição sobre $\mathsf{Z}(\Omega) \subset \mathbb{R}^n$ é a medida de probabilidade $\mathbb{P}_\mathsf{Z} = \mathbb{P}_{(\mathsf{X}_1,\ldots,\mathsf{X}_n)}$ definida por $\mathbb{P}_\mathsf{Z}((a_1,\ldots,a_n)) = \mathbb{P}[\mathsf{Z} = (a_1,\ldots,a_n)] = \mathbb{P}[(\mathsf{X}_1,\ldots,\mathsf{X}_n) =$

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

$(a_1,\dots,a_n)]$. Essa distribuição, chamada **distribuição conjunta** das variáveis $X_1,\dots,X_n$, *não* fica determinada pelas distribuições $\mathbb{P}_{X_i}$ a não ser com a hipótese de independência das variáveis 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 de Ω

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

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

Naturalmente, podemos estender essa definição para qualquer quantidade finita de variáveis aleatórias. As variáveis aleatórias $X_1,X_2,\dots,X_n$ são **independentes** se para quaisquer eventos $A_1,\dots,A_n$

$$\mathop{\mathbb{P}}_{(X_1,X_2,\dots,X_n)}(A_1\times\cdots\times A_n) = \prod_{i=1}^{n}\mathop{\mathbb{P}}_{X_i}(A_i). \tag{3.1}$$

Notemos que qualquer subconjunto $X_{i_1},\dots,X_{i_k}$ dessas variáveis também é independente, basta tomar $A_j = X_j(\Omega)$ para todo $j\neq i_1,\dots,i_k$ na equação (3.1) acima.

EXERCÍCIO 3.9. Sejam $X_1,\dots,X_n$ variáveis aleatórias e $S_1,\dots,S_n$ conjuntos finitos e não vazios. Mostre que são equivalentes

1. $(X_1,\dots,X_n)\in_{\mathcal{U}} S_1\times\cdots\times S_n$
2. $X_1,\dots,X_n$ são independentes e $X_i\in_{\mathcal{U}} S_i$ para cada i.

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

$$\mathop{\mathbb{P}}_{Y}(y) = \mathbb{P}[f(X)=y] = \sum_{x:\,g(x)=y}\mathop{\mathbb{P}}_{X}(x).$$

Por exemplo, se X é uma variável aleatória de $(\Omega,\mathbb{P})$, então podemos definir $Z:\Omega\to\mathbb{R}$ por $Z(\omega) := X(\omega)^2$ para todo $\omega\in\Omega$. A função Z 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 $Y:\Omega\to\mathbb{R}$ 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 $(X_1,X_2,\dots,X_n)$ é um vetor aleatório de um espaço de probabilidade e $f:\mathbb{R}^n\to\mathbb{R}$ é função então $f(X_1,X_2,\dots,X_n)$ é uma variável aleatória real. Logo operações aritméticas elementares com variáveis aleatórias reais resultam em variáveis aleatórias. Em particular, se X e Y são variáveis aleatórias definidas em $(\Omega,\mathbb{P})$, a soma é a variável aleatória $X+Y:\Omega\to\mathbb{R}$ dada por $\{X+Y\}(\omega) = X(\omega)+Y(\omega)$ e o produto é a variável aleatória $X\cdot Y:\Omega\to\mathbb{R}$ dada por $\{X\cdot Y\}(\omega) = X(\omega)\cdot Y(\omega)$. A distribuição de $Z = X+Y$ é

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

No caso particular em que X e Y são independentes

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

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 real.

Proposição 3.10 *Sejam* X *e* Y *variáveis aleatórias reais 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 3.11. Determine a distribuição da soma de duas variáveis de Poisson e. da soma de duas variáveis binomiais com mesmo p.

3.1.1 VALOR ESPERADO DE UMA VARIÁVEL ALEATÓRIA SIMPLES

Uma variável aleatória de $(\Omega, \mathbb{P})$ com imagem finita é chamada de **variável aleatória simples**. O **valor médio** de X é a média dos valores de $X(\Omega) = \{x_1, x_2, \ldots, x_n\}$ ponderada pela probabilidade de cada valor

$$\mathbb{E}X := x_1 \mathop{\mathbb{P}}_{X}(x_1) + x_2 \mathop{\mathbb{P}}_{X}(x_2) + \cdots + x_n \mathop{\mathbb{P}}_{X}(x_n) \tag{3.2}$$

e também chamado de **valor esperado** ou, ainda, **esperança** da variável aleatória simples.

Exemplo 3.12. Consideremos um jogo de azar no qual em cada aposta ou ganhamos R$1.000.000,00 com probabilidade $p \in (0,1)$ ou perdemos R$10,00 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 $\mathbb{E}Y = 9.990{,}10$. ◊

Se tomamos a partição de Ω dada por $A_k = [X = x_k]$ para todo k, então podemos escrever

$$X = \sum_{k=1}^{n} x_k \mathbb{1}_{A_k}$$

e o valor médio de X é dada em função de $\mathbb{P}$ ao invés de $\mathbb{P}_X$ por

$$\mathbb{E}\,X = x_1\mathbb{P}(A_1) + x_2\mathbb{P}(A_2) + \cdots + x_n\mathbb{P}(A_n). \tag{3.3}$$

O valor esperado não depende de como escrevemos X como combinação linear de variáveis indicadoras. Seja $\{B_\ell \colon 1 \leqslant \ell \leqslant m\}$ uma partição *qualquer* de Ω tal que

$$\sum_{k=1}^{n} x_k \mathbb{1}_{A_k} = \sum_{\ell=1}^{m} y_\ell \mathbb{1}_{B_\ell}$$

com possíveis valores repetidos para os y_ℓ's. Então para todo k

$$A_k = \bigcup_{\ell\colon\, y_\ell = x_k} B_\ell$$

logo

$$\sum_{k=1}^{n} x_k \mathbb{P}(A_k) = \sum_{k=1}^{n} x_k \sum_{\ell\colon\, y_\ell = x_k} \mathbb{P}(B_\ell) = \sum_{k=1}^{n} \sum_{\ell\colon\, y_\ell = x_k} y_\ell \mathbb{P}(B_\ell)$$

portanto

$$\sum_{k=1}^{n} x_k \mathbb{P}(A_k) = \sum_{\ell=1}^{m} y_\ell \mathbb{P}(B_\ell). \tag{3.4}$$

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

EXERCÍCIO 3.13. 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 4, página 129)

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 3.14 (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*

$$\mathbb{E}[\,aX + bY\,] = a\mathbb{E}\,X + b\mathbb{E}\,Y. \tag{3.5}$$

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 (3.2) 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 (3.2), o item 1 do exercício 3.13 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

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

onde 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 3.13 acima,

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

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 $A_k \subset B_\ell$, 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 (3.4) 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 3.15 *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 (3.5) usando indução em n. □

Exemplo 3.16 (*esperança das distribuições Bernoulli e binomial*). Se X tem distribuição de Bernoulli $\mathbb{E}X = \mathbb{P}[X = 1] = p$ em que p.

Se $X_1, \ldots, X_n$ são variáveis Bernoulli independentes e $X = \sum_i X_i$ então X é 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 3.17. Se $X \in_{\mathcal{U}} \{1,2,3,4,5,6\}$ 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 $Y \in_{\mathcal{U}} S$ 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 abaixo, donde podemos

$X+Y$	2	3	4	5	6	7	8	9	10	11	12
$\mathbb{P}_{X+Y}$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Tabela 3.1: distribuição da soma de dois dados.

calcular o valor esperado da soma $\mathbb{E}[X+Y] = (2\cdot 1 + 3\cdot 2 + \cdots + 11\cdot 2 + 12\cdot 1)/36 = 7$. O produto XY tem valor esperado $7/2 \cdot 7/2$, pelo item 6 do teorema acima já que as variáveis são independentes. Alternativamente, a distribuição do produto é mostrada na tabela 3.2, donde podemos calcular o valor esperado para o produto dos resultados, $\mathbb{E}[X \cdot Y] = 49/4$. ◊

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

Tabela 3.2: distribuição do produto de dois dados.

3.1.2 TABELAS DE ESPALHAMENTO

Em computação um conjunto que é modificado com passar do tempo é usualmente chamado de **conjunto dinâmico**. Uma solução bastante conhecida e estudada para a representação computacional de um conjunto dinâmico S de modo que possamos inserir elementos, remover elementos e testar pertinência é conhecida como **tabela de espalhamento** ou **tabela *hashing***. Uma maneira de implementar uma tabela de espalhamento N é construir um vetor $N[i]$ de listas ligadas indexadas por $M = \{0, 1, \ldots, m-1\}$ e o acesso à tabela dá-se por uma função de *hash* $h\colon U \to M$: dado $x \in U$, a inserção, remoção ou busca por x em N é feita na lista ligada $N[h(x)]$ (veja uma ilustração na figura 3.1).

As operações 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;

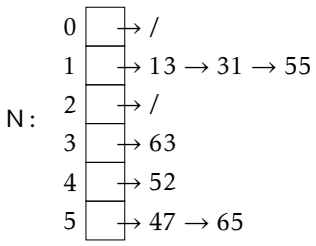


Figura 3.1: tabela de espalhamento com os elementos $\{13, 31, 47, 52, 55, 63, 65\}$ distribuídos de acordo com a função $h(x) = x \bmod 6$.

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, S é uma fração pequena de U e o caso interessante é quando $|S| \geqslant |M|$. Como $|U| > |M|$ não há como evitar *colisões* (elementos distintos mapeados para a mesma lista). 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 *hashing* queremos que a probabilidade de colisão seja pequena, que a função seja fácil de computar e que a representação seja sucinta, isto é, relativamente poucos bits são necessários para armazenar a função. Essas funções, além de uma grande ferramenta prática como descrevemos abaixo, é bastante útil em Complexidade Computacional e em Criptografia e voltaremos a estudar funções de *hash* em outras seções adiante neste texto.

Uma função h escolhida uniformemente no conjunto M^U de todas as $|M|^{|U|}$ funções de U em M tem, com alta probabilidade, a propriedade de não formar listas longas (veja o exercício 1.57, página 36), entretanto, tal função pode não ter uma representação sucinta, alguma delas requer pelo menos $|U| \log(|M|)$ bits para ser representada. Idealmente, o que procuramos são funções que tenham representação sucinta e imitam uma função aleatória na propriedade de não formar listas longas e, também, que possam ser computadas eficientemente em cada ponto do domínio.

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 $(M^U, \mathbb{P})$ com $\mathbb{P}(h) = 1/|M^U|$ para toda função $h \in M^U$. Notemos que vale

$$\mathbb{P}[\, h(x) = i \,] = \frac{1}{|M|} \tag{3.6}$$

para todo $x \in U$ e todo $i \in M$. De fato, sortear uniformemente uma função é equivalente a sortear uniformemente e independentemente uma imagem para cada elemento do domínio (veja o exercício 3.9, página 94), isto é, $h \in_{\mathcal{U}} M^U$ é equivalente a $h(x) \in_{\mathcal{U}} M$ com as escolhas $h(x)$ independentes, para todo $x \in U$.

Fixemos os parâmetros $n = |S|$ e $m = |M|$. A fração

$$\mu = \mu(n, m) := \frac{n}{m} \tag{3.7}$$

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 (3.6). Tomemos

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

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

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

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

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

portanto o tamanho esperado da lista $N[h(x)]$ é, pela linearidade da esperança (corolário 3.15)

$$\sum_{y \in S} \mathbb{E}\,\mathbb{1}_{[h(x)=h(y)]} = \begin{cases} \frac{n}{m}, & \text{se } x \notin S, \\ \frac{(n-1)}{m} + 1, & \text{caso contrário.} \end{cases} \tag{3.8}$$

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 μ da tabela.

Proposição 3.18 *O tempo esperado para uma busca sem sucesso é $\mu + 1$ e o tempo esperado para uma busca com sucesso é $\mu/2 - 1/(2m) + 1$.*

Demonstração. Consideremos uma busca por $x \in U$ na tabela de espalhamento N. Se $x \notin S$ então, pelo primeiro caso da equação (3.8), são necessárias $\mu + 1$ comparações em média numa busca sem sucesso. Agora, suponhamos $x \in S$ 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 (3.8), portanto, o tempo médio da busca por x é

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

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

Essa análise não nos dá nenhuma pista sobre o tempo médio de pior caso de uma busca. O tempo médio de pior caso é $O(1 + L)$ com L o maior tamanho de lista na tabela. No caso $n = m$ a proposição acima garante que o tempo médio de busca é constante, entretanto, vimos no exercício 1.57, página 36, que a maior lista tem $O(\log n)$ com alta probabilidade. De fato, o tempo médio pior caso para uma função aleatória é dado por $\mathbb{E}\,L = \Theta(\log(n)/\log(\log(n)))$ (veja seção 3.3.1, página 121).

Exemplo 3.19 (*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. Notemos que 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 $\mathbb{E}\,C < 1/2$ e, de fato, não há colisão com probabilidade pelo menos 1/2 pois

$$\frac{1}{2} > \mathbb{E}\,C = \sum_{k=0}^{\binom{n}{2}} k\mathbb{P}[\,C = k\,] \geqslant \sum_{k=1}^{\binom{n}{2}} \mathbb{P}[\,C = k\,] = \mathbb{P}[\,C \geqslant 1\,].$$

Hashing universal Se em vez de exigirmos que seja válida a equação (3.6) para todo x e todo i, pedirmos que numa família de funções $\mathcal{H} \subset M^U$ seja válida a condição

$$\mathbb{P}_{h\in\mathcal{H}}[h(x) = h(y)] \leqslant \frac{1}{m} \tag{3.9}$$

para todo $x \neq y$ ainda será verdade que o tamanho médio de uma lista é limitado pela equação (3.7), que a sentença enunciada na proposição 3.18 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. Uma família $\mathcal{H} \subset M^U$ de funções que satisfazem a equação (3.9) é chamada de *universal*.

Como não precisamos de funções genuinamente aleatórias para termos as boas propriedades estatísticas dessas funções nos resta procurar por uma família $\mathcal{H}$ universal e não vazia de funções *hashing* com representações sucintas e que são calculadas eficientemente. Um exemplo é dado a seguir. Tomamos $U := \{0,1\}^u$ e $M := \{0,1\}^m$ e definimos uma função $h\colon U \to M$ 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 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. Nesse caso

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

Seja i uma coordenada não nula de $x-y$. 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

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

para cada $\ell \in \{1,\ldots,m\}$, o que ocorre com probabilidade $1/2^m$, ou seja, $\mathbb{P}_{h\in\mathcal{H}}[h(x) = h(y)] \leqslant 1/|M|$. Observemos que $h(0) = 0$ para qualquer escolha de A o que faz com que a equação (3.6) não seja válida para todo x e todo i.

3.1.3 ESPERANÇA MATEMÁTICA

Seja X uma variável aleatória real sobre o espaço discreto de probabilidade $(\Omega, \mathbb{P})$ com imagem enumerável infinita. Se estendemos de modo natural a definição de valor esperado de uma variável simples dada na equação (3.2), página 95, tomamos $X(\Omega) = \{x_1, x_2, \dots\}$ e temos

$$\mathbb{E}X = \sum_{n \geqslant 1} x_n \mathop{\mathbb{P}}_{X}(x_n) \tag{3.10}$$

entretanto esse valor, caso exista já que é um limite (a definição e algumas propriedades importantes podem ser vistas na página 187), não deve depender da enumeração particular $x_1, x_2, \dots$ da imagem de X. Uma leitura atenta nas demonstrações das propriedades de valor esperado de variável simples revela que a reordenação dos termos da soma (sem alterar o resultado) é uma operação importante e que queremos preservar.

Seja $(x_n \colon n \geqslant 1)$ uma sequência de números reais. Se $x_n \geqslant 0$ para todo n então a série $\sum_{n \geqslant 1} x_n$ pode ser *finita*, isto é convergir para um número real, ou *infinita*, isto é tender ao infinito, o que denotamos por $\sum_n x_n = +\infty$. Em ambos os casos dizemos que a série é *bem definida*. Agora, no caso geral, definimos

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

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 (3.13) abaixo). Em vista disso, se a série na equação (3.10) está bem definida então a série não depende da enumeração de $X(\Omega)$ e expressa a esperança $\mathbb{E}X$ de X.

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

$$\mathbb{E}X := \sum_{r \in S} r \mathop{\mathbb{P}}_{X}(r). \tag{3.11}$$

sempre que a série (3.10) está bem definida. Ademais, a esperança é um número real quando a série converge absolutamente, isto é, quando quando $\mathbb{E}|X| := \sum_{r \in S} |r| \, \mathbb{P}_X(r)$ converge. Nesse último caso escrevemos $\mathbb{E}|X| < +\infty$ e dizemos que a variável aleatória X tem **esperança finita** ou que X é **somável**.

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

$$\mathbb{E}X = \sum_{\omega \in \Omega} X(\omega)\mathbb{P}(\omega) \tag{3.12}$$

que é útil na prática. Assim, X é somável sempre que $\mathbb{E}|X| = \sum_\omega |X(\omega)|\mathbb{P}(\omega) < +\infty$.

No que segue chamamos genericamente de *soma* tanto um somatório com finitos termos quanto uma série.

Exemplo 3.20 (*esperança da distribuição geométrica*). Se X é geométrica de parâmetro p, então

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

usando (s.6a). ◊

Os próximos exemplos são 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 teorema da multiplicação (página 22), 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

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

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.

Uma série conhecida pela propriedade de convergir não-absolutamente é a série harmônica alternada

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

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

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

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

$$\sum_{i\geqslant 1} x_{\pi(i)} \mathbb{P}_X(x_{\pi(i)}) = \begin{cases} \frac{6}{\pi^2}\log(2) & \text{se } \pi \text{ é a permutação identidade} \\ \frac{3}{\pi^2}\log(2) & \text{se } \pi \text{ é a permutação de Laurent} \end{cases} \tag{3.13}$$

portanto X não tem valor médio bem definido.

Proposição 3.21 *Se X assume valores inteiros e positivos então $\mathbb{E}X$ está bem definida e*

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

Demonstração. Se X assume valores positivos então podemos rearranjar a soma

$$\sum_{r\geqslant 1} r\mathbb{P}_X(r) = \sum_{r\geqslant 1} \mathbb{P}_X(r) + \sum_{r\geqslant 2} \mathbb{P}_X(r) + \sum_{r\geqslant 3} \mathbb{P}_X(r) + \cdots$$

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

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

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

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

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

e assim provamos o seguinte resultado.

Teorema 3.22 *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*

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

sempre que a soma está bem definida. □

Vejamos um exemplo. Se Z é geométrica de parâmetro p, então para computar $\mathbb{E} Z^2$ usamos (s.6c) do seguinte modo: derivando ambos os lados de $\sum_{n \geqslant 1} n x^n = x(1-x)^{-2}$ obtemos (para $|x| < 1$)

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

de modo que, pelo teorema 3.22 e a série acima

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

Exemplo 3.23 (*lei de potência*). A distribuição de uma variável aleatória X sobre os inteiros positivos segue uma lei de potência com parâmetro $\alpha > 0$ se tem distribuição

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

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 3.21 temos

$$\mathbb{E} X = \sum_{n \geqslant 1} \mathbb{P}[X \geqslant n] = \sum_{n \geqslant 1} \frac{1}{n^\alpha}. \tag{3.14}$$

Para $\alpha = 1$ temos $\mathbb{E} X = \sum_{n \geqslant 1} 1/n = +\infty$. Para $\alpha = 2$ temos $\mathbb{E} X = \pi^2/6$ (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 $\alpha = 2$. A soma no lado direito da equação (3.14) quando α é qualquer número complexo com parte real maior que 1 é conhecida como Função Zeta de Riemann[3] e denotada por $\zeta(\alpha)$.

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, pelo teorema acima,

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

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

◊

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 pelos múltiplos de p. Se p e q são distintos, os eventos A_p e A_q são independentes $\mathbb{P}_X(A_p \cap A_q) = \mathbb{P}_X(A_p)\,\mathbb{P}_X(A_q) = 1/(pq)^3$. O mesmo vale para

$$\mathop{\mathbb{P}}_{X_s}(n) = \frac{n^{-s}}{\zeta(s)}$$

e todo $s > 1$

$$\mathop{\mathbb{P}}_{X_s}(A_p) = \frac{\sum_{k\geqslant 1}(pk)^{-s}}{\sum_{n\geqslant 1} n^{-s}} = \frac{1}{p^s} \quad \text{e} \quad \mathop{\mathbb{P}}_{X_s}(A_p \cap A_q) = \mathop{\mathbb{P}}_{X_s}(A_{pq}) = \frac{1}{p^s}\frac{1}{q^s} = \mathop{\mathbb{P}}_{X_s}(A_p)\mathop{\mathbb{P}}_{X_s}(A_q).$$

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

$$B_q = \bigcap_{\substack{p\leqslant q \\ p\text{ primo}}} \overline{A_p}$$

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

$$\frac{1}{\zeta(s)} = \mathop{\mathbb{P}}_{X_s}(1) = \mathop{\mathbb{P}}_{X_s}\left(\bigcap_{q\text{ primo}} B_q\right) = \lim_{q\to\infty}\mathop{\mathbb{P}}_{X_s}(B_q) = \lim_{q\to\infty}\prod_{\substack{p\leqslant q \\ p\text{ primo}}}\mathop{\mathbb{P}}_{X_s}(\overline{A_p}) = \prod_{p\text{ primo}}\mathop{\mathbb{P}}_{X_s}(\overline{A_p}) = \prod_{p\text{ primo}}\left(1-\frac{1}{p^s}\right)$$

que é a *fórmula do produto de Euler*, descoberta por Euler quando na busca de uma prova de que $\sum_{p\text{ primo}} 1/p$ 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

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

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

Ainda, como uma curiosidade a mais, $\lim_{s\to 1}\mathbb{P}_{X_s}(A) = \rho(A)$ em que $\rho(A) = \lim_{n\to\infty}|A\cap\{1,\ldots,n\}|/n$ é a densidade relativa de A (definida no exercício 1.58, página 36).

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 3.14.

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

A prova desse teorema é imediata da definição de valor esperado, equação (3.11).

Teorema 3.25 (linearidade da esperança) *Se* X *e* Y *são variáveis aleatórias somáveis de* $(\Omega,\mathbb{P})$ *e* $a, b \in \mathbb{R}$ *constantes quaisquer então* $aX + bY$ *é somável e sua esperança satisfaz* $\mathbb{E}[aX+bY] = a\,\mathbb{E}\,X + b\,\mathbb{E}\,Y$.

Demonstração. Sejam X e Y variáveis aleatórias somáveis e $a, b \in \mathbb{R}$. Pela equação (3.12) acima a esperança da variável aleatória $|aX + bY|$ é

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

usando a desigualdade triangular (veja (d.4)). De X e Y somáveis temos que o lado direito da equação acima é finito, portanto $aX + bY$ é somável 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 somáveis 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 somáveis 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 $a \leqslant X \leqslant b$, então $a \leqslant \mathbb{E}X \leqslant b$.*
4. *Se X é somável então $X \cdot \mathbb{1}_A$ é somável 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 somável, então, por monotonicidade, $X\mathbb{1}_A$ é somável.

Se X está definida e não é somável então $|\mathbb{E}X| = +\infty = \mathbb{E}|X|$ e o item 5 vale com igualdade. Senão, X é somável 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, $|\mathbb{E}X| \leqslant \mathbb{E}|X|$. □

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

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

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

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

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

Teorema 3.29 *Sejam $X\colon \Omega \to S$ e $Y\colon \Omega \to R$ variáveis aleatórias independentes em $(\Omega, \mathbb{P})$ e sejam $f\colon S \to \mathbb{R}$ e $g\colon R \to \mathbb{R}$ funções tais que $f(X)$ e $g(Y)$ são variáveis aleatórias somáveis. Então $f(X)\cdot g(Y)$ é somável e vale $\mathbb{E}[\,f(X)\cdot g(Y)\,] = \mathbb{E}[\,f(X)\,]\mathbb{E}[\,g(Y)\,]$.*

Demonstração. Da proposição 3.10 temos que $f(X)$ e $g(Y)$ são variáveis aleatórias independentes. Podemos replicar a dedução acima para provar que $f(X)g(Y)$ é somável e que, portanto, $\mathbb{E}[\,f(X)\cdot g(Y)\,] = \mathbb{E}[\,f(X)\,]\cdot\mathbb{E}[\,g(Y)\,]$. □

Observação 3.30. O teorema 3.25, da linearidade, ainda vale com as hipóteses dos valores esperados $\mathbb{E}X$ e $\mathbb{E}Y$ definidos e $a\mathbb{E}X + b\mathbb{E}Y$ definido, isto é, não resulta nos indeterminados $+\infty - \infty$ ou $-\infty + \infty$. Analogamente, 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 somável, 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 $\mathbb{E}Y > -\infty$. ◊

3.1.4 *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; os elementos menores que o pivô formam a subsequência S_1 e os demais que não o pivô formam a subsequência S_2; o algoritmo devolve a sequência ordenada recursivamente (*quicksort*(S_1), pivô, *quicksort*(S_2)).

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.

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 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

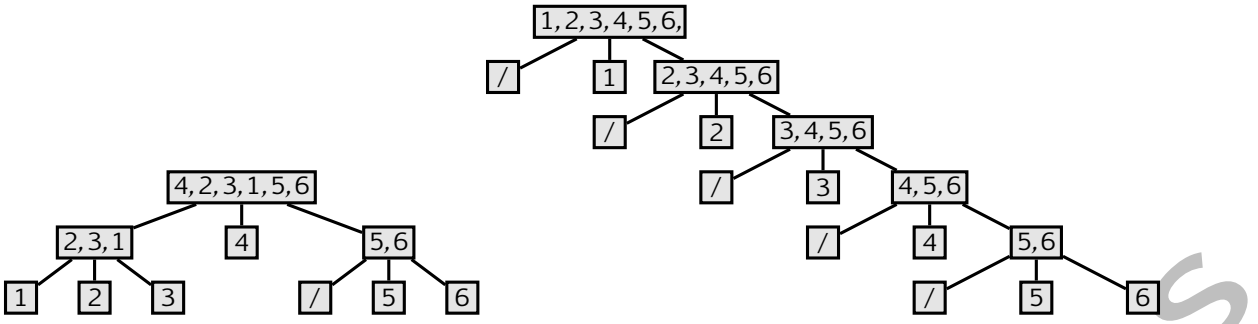


Figura 3.2: na esquerda temos uma árvore de recursão do *quicksort* onde o pivô é a mediana da sequência. A árvore da direita é o caso onde o pivô é sempre o menor elemento da sequência.

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 fazendo todas as comparações entre os elementos;
3      responda a sequência ordenada.
4  senão
5      x ←_R S;
6      para cada y ∈ S, y ≠ x faça
7          se y < x então insira y em S_1;
8          senão insira y em S_2;
9      ordene recursivamente S_1;
10     ordene recursivamente S_2;
11     responda (S_1, x, S_2).
```

Algoritmo 30: *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, M não deve ser muito maior que $2\log_{10/9}(n)$ pois pelo exercício 1.53, página 35, 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 $2n\log_{10/9}(n)$ se existe um x que passa por mais que $2\log_{10/9}(n)$ particionamentos o que ocorre com probabilidade $n(1/n^2) = 1/n$.

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

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

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 os demais elementos da subsequência a qual ele pertence é comparado com ele e desde então x_i nunca mais participará de um particionamento e nunca mais será comparado com outro elemento de S.

EXERCÍCIO 3.31. 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, $Y_k \in_{\mathcal{G}(p)} \mathbb{N}$, logo $\mathbb{E}\, Y_k = 1/p \leqslant 1/0{,}75 < 2$. Pelo exercício 3.31 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

$$\mathbb{E}\, X_i < 2\lfloor \log_{\frac{10}{9}} n \rfloor \quad \text{e} \quad \mathbb{E}\, T < 2n\lfloor \log_{\frac{10}{9}} n \rfloor. \tag{3.15}$$

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 (3.15) totalizam $O(n\log n)$ comparações.

Lema 3.32 *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 3.33 *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 (3.15) 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 (d.2) para estimar o coeficiente binomial,

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

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

$$\mathbb{P}[X_i > L] < 100(0{,}8)^{12\log_{10/9}(n)-1}.$$

Mudando a base do logaritmo obtemos

$$\log_{10/9} n^{12} = c \log_{8/10} n^{12} = \log_{8/10} n^{12c}$$

com $1/c = \log_{0,8}(10/9) = -0{,}472164734$ de modo que

$$100(0{,}8)^{12\log_{10/9}(n)-1} = 125n^{12c} < n^{-23}$$

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. □

3.2 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 em uma coleção de objetos algum tiver uma determinada propriedade, então a probabilidade de uma escolha aleatória com uma distribuição apropriada nessa coleção ter essa propriedade é diferente de zero, não importa quão pequena seja a probabilidade, basta que seja estritamente positiva. Em muitos casos o fato demonstrado é uma sentença que não envolve probabilidade, embora a demonstração use probabilidade, como o exemplo da divergência de $\sum_p 1/p$ que apresentamos na página 105 (com a ressalva de que naquele caso existe demonstração sem usar probabilidade, que nem sempre é o caso). Há várias técnicas, algumas sofisticadas, que se encaixam nesse método e neste momento veremos uma simples, baseada no seguinte exercício.

EXERCÍCIO 3.34 (princípio de primeiro momento). Seja X uma variável aleatória real e k real. Se $\mathbb{E}X \geqslant k$ então $X(\omega) \geqslant k$ para algum $\omega \in \Omega$.

3.2.1 MAX-3-SAT

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,

$$(v_1 \vee \neg v_2 \vee v_3) \wedge (v_2 \vee \neg v_3 \vee v_4) \wedge (v_3 \vee \neg v_4 \vee v_1) \tag{3.16}$$

é uma fórmula booleana com três cláusulas e cada uma delas é composta por três literais. Consideramos 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 (*e*) de cláusulas está na Forma Normal Conjuntiva, abreviada CNF (do inglês *conjunctive normal form*). Uma *fórmula 3-CNF* é da forma $C_1 \wedge C_2 \wedge \cdots \wedge C_m$ com cada cláusula C_i sendo uma disjunção de 3 literais, como na equação (3.16) acima.

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, é 1. No exemplo dado na equação (3.16) a valoração $v_1 = 1$, $v_2 = 1$, $v_3 = 1$ e $v_4 = 0$ satisfaz a fórmula.

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

Problema computacional da satisfazibilidade booleana (SAT):
Instância : uma fórmula booleana Φ e suas variáveis.
Resposta : *sim* se existe uma valoração das variáveis que satisfaz Φ, *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 NP-completo conhecido e esse notável resultado é o conhecido Teorema de Cook–Levin.

O 3-SAT é o problema de satisfazibilidade obtido quando restringimos as instâncias às fórmulas 3-CNF, também é um problema sem algoritmo eficiente conhecido.

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

Problema computacional da satisfazibilidade máxima de uma fórmula 3-CNF (MAX-3-SAT):
Instância : uma fórmula 3-CNF Φ e suas variáveis.
Resposta : valoração do maior número possível de cláusulas dentre todas valorações.

Esses problemas são pelo menos tão difíceis de se resolver algoritmicamente quanto SAT e 3-SAT.

Por exemplo, $\mathcal{C} = \{\{v_1, \neg v_2, v_3\}, \{v_2, \neg v_3, v_4\}, \{v_3, \neg v_4, v_1\}\}$ é a descrição de uma instância de 3-SAT e de MAX-3-SAT que corresponde a fórmula booleana dada na equação (3.16) acima. Como vimos, todas as cláusulas são satisfeitas para uma valoração. No que segue, vamos supor que as cláusulas não têm variáveis repetidas.

Sejam $\mathcal{C} = \{C_1, C_2, \ldots, C_m\}$ uma instância do MAX-3-SAT e $V = \{v_1, \ldots, v_n\}$ o conjunto das variáveis de $\mathcal{C}$. O conjunto de todas as valorações das variáveis é $\Omega := \{0,1\}^n$ em que interpretamos a i-ésima coordenada como o valor atribuído a v_i e em Ω consideramos a distribuição uniforme. Para cada i definimos a variável aleatória indicadora

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

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

$$X := \sum_{i=1}^{m} \mathbb{1}_{[C_i=1]} \tag{3.17}$$

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 3.34 deve haver uma valoração em Ω que satisfaz pelo menos $7m/8$ cláusulas.

THEOREMA 3.35 (JOHNSON, 1974) *Para toda instância de* MAX 3-SAT, *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 3.9, página 94, podemos escolher um valor lógico para cada variável, com as escolhas independentes, ao invés de escolher ao acaso uma valoração de V.

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

```
1 repita
2     para cada i ∈ {1,...,n} faça  v_i ←_R {0,1};
3     avalie C_1,...,C_m;
4 até que pelo menos (7/8)m cláusulas estejam satisfeitas
5 responda v_1,...,v_n.
```

Algoritmo 33: 7/8-aproximação para MAX-3-SAT.

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-3-SAT.

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 $Z \in_{\mathcal{G}(p)} \mathbb{N}$. 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 3.36 *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 (3.17) acima, então

$$\frac{7}{8}m = \mathbb{E}X = \sum_{j=0}^{m} j p_j. \tag{3.18}$$

Para qualquer inteiro $M < m$

$$\sum_{j=0}^{m} j p_j = \sum_{j=0}^{M} j p_j + \sum_{j=M+1}^{m} j p_j \leqslant \sum_{j=0}^{M} M p_j + \sum_{j=M+1}^{m} m p_j = M \sum_{j=0}^{M} p_j + m \sum_{j=M+1}^{m} p_j \tag{3.19}$$

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

$$\sum_{j=0}^{M} p_j = 1 - p \qquad \text{e} \qquad \sum_{j=M+1}^{m} p_j = p. \tag{3.20}$$

Das equações (3.18), (3.19) e (3.20), deduzimos que

$$\frac{7}{8}m \leqslant M(1-p) + mp, \tag{3.21}$$

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 $7m/8 < M + 1$ deduzimos que $8m - 8M \leqslant m + 8$, portanto $p \geqslant 1/(8+m)$. □

COROLÁRIO 3.37 *O número esperado de rodadas do laço do algoritmo 33 é $\mathbb{E}Z \leqslant m + 8$.* □

O algoritmo 33 não responde errado, entretanto o número total de instruções executadas depende dos sorteios, execuções independentes desse algoritmo com a mesma instância podem resultar em tempos diferentes até terminar.

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 3.5, página 91, 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 0,00000000000000000013.

Algoritmos aproximativos São algoritmos que encontram soluções aproximadas para problemas de otimização. Para $0 < \alpha < 1$, uma α-aproximação é uma garantia de que o valor respondido pelo algoritmo é no máximo o valor ótimo OPT e pelo menos $\alpha \cdot \text{OPT}$. Os algoritmos aproximativos são muito estudados em teoria da computação pois podem levar a resultados surpreendentes como, por exemplo, uma $(7/8+\varepsilon)$-aproximação em tempo polinomial para MAX-3-SAT, para qualquer $\varepsilon > 0$, implicaria em P = NP (Håstad, 2001) respondendo o mais importante problema em aberto na Teoria da Computação e um dos maiores problemas matemáticos em aberto segundo o Instituto Clay de Matemática.

3.2.2 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 34: 1/2-aproximação para MAX-CUT.

A análise desse algoritmo também 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 $Z \in_{\mathcal{G}(p)} \mathbb{N}$, cuja esperança é $\mathbb{E}Z = 1/p$.

Seguindo a mesma dedução da seção anterior, equações (3.18), (3.19), (3.20) e (3.21) 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 : um subconjunto de vértices A tal que $|\nabla(A)|$ é 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 $\mathsf{P} = \mathsf{NP}$ (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.

3.3 DISTRIBUIÇÃO E ESPERANÇA CONDICIONAIS

Sejam $(\Omega, \mathbb{P})$ um espaço de probabilidade, A um evento desse espaço com probabilidade positiva e X um variável aleatória real e somável (ou apenas não negativa). A **esperança condicional** da variável X dado A é o valor médio de X com respeito a lei condicional $\mathbb{P}_A$, usando a equação (3.12)

$$\mathbb{E}[X \mid A] = \sum_{\omega \in \Omega} X(\omega)\mathbb{P}_A(\omega) = \sum_{\omega \in \Omega} X(\omega)\mathbb{P}(\{\omega\} \mid A) = \sum_{\omega \in A} X(\omega)\frac{\mathbb{P}(\omega)}{\mathbb{P}(A)} = \frac{\mathbb{E}[X \cdot \mathbb{1}_A]}{\mathbb{P}(A)}. \tag{3.22}$$

Pelo corolário 3.28, item 4, se X é somável 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

$$\mathbb{E}[X \mid A] = \sum_r r\mathbb{P}[X = r \mid A].$$

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

EXERCÍCIO 3.38 (propriedades da esperança condicional). Prove que se X e Y são somáveis, 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 $\mathbb{E}[X \mid A] = \mathbb{E}X$.
6. Se X e Y são independentes então $\mathbb{E}[XY \mid A] = \mathbb{E}[X \mid A] \cdot \mathbb{E}[Y \mid A]$.

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

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

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

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

para todo x. Claramente, valores diferentes de y podem resultar em distribuições condicionais diferentes. Agora, para uma variável aleatória real e somável X, a **esperança condicional de** X **dado** $[Y = y]$, para qualquer $y \in R_Y$, é dada pela equação (3.22), ou seja, vale que

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

Se um dado equilibrado é lançado duas vezes e X e Y são os valores obtidos e $S := X + Y$ e $P := X \cdot Y$, então usando a definição temos

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

e

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

embora as esperanças possam ser calculadas de maneira mais fácil usando as propriedades do valor médio. Por exemplo, para a soma temos $\mathbb{E}[\, S \mid X = 2 \,] = \mathbb{E}[\, X \mid X = 2 \,] + \mathbb{E}[\, Y \mid X = 2 \,] = 2 + \mathbb{E}\, Y = 2 + 7/2$; para o produto temos $\mathbb{E}[\, P \mid Y = 1 \,] = \mathbb{E}[\, XY \mid Y = 1 \,] = \mathbb{E}[\, X \mid Y = 1 \,]\mathbb{E}[\, Y \mid Y = 1 \,] = \mathbb{E}[\, X \mid Y = 1 \,] = 7/2$. Ainda usando a definição temos $\mathbb{E}[\, X \mid S = 5 \,] = \sum_{r=1}^{6} r\mathbb{P}[\, X = r \mid S = 5 \,] = 5/2$.

Se X é somável, $y \in R_Y$ e se olharmos para $\mathbb{E}[\, X \mid Y = y \,]$ como um função de y, digamos $\psi\colon \mathbb{R} \to \mathbb{R}$ dada por

$$\psi(y) := \begin{cases} \mathbb{E}[\, X \mid Y = y \,], & \text{se } y \in R_Y \\ 0, & \text{caso contrário,} \end{cases}$$

então a função

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

é uma variável aleatória que passamos a chamar de **esperança condicional de** X **dado** Y e para todo $\omega \in \Omega$

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

O valor $\mathbb{E}[\, X \mid Y \,](\omega)$ é a média ponderada dos valores $X(\upsilon)$ sobre todo ponto amostral $\upsilon \in \Omega$ tal que $Y(\upsilon) = Y(\omega)$. Por exemplo, voltemos ao caso de dois lançamentos de um dado (X, Y) e consideremos $\mathbb{E}[\, S \mid P \,] = \mathbb{E}[\, X + Y \mid X \cdot Y \,]$. O valor de $\mathbb{E}[\, S \mid P \,](3, 5)$ é o valor médio da soma S 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 $\mathbb{E}[\, S \mid P \,](3, 5) = 8$. O valor de $\mathbb{E}[\, S \mid P \,](1, 4)$ é o valor médio da soma S 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 $\mathbb{E}[\, S \mid P \,](1, 4) = (2/3) \cdot 5 + (1/3) \cdot 4 = 14/3$. Escrevendo de outro modo, se $P = 15$ então os resultados dos lançamentos são $(3, 5)$ ou $(5, 3)$, para os quais $S = 8$, de modo que

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

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

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

$\psi(P) = \mathbb{E}[S \mid P]$	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.

A lei da variável aleatória $\mathbb{E}[S \mid P]$ está dada na tabela 3.3 abaixo.

Se $\mathbb{E}[X \mid Y]$ é uma variável aleatória então podemos calcular o seu valor médio. No exemplo acima, podemos usar a tabela 3.3 para concluir que $\mathbb{E}[\mathbb{E}[S \mid P]] = \mathbb{E}[\psi(P)] = 7$. Lembremos o exemplo 3.17, onde determinamos que $\mathbb{E}S = 7$. Esse fenômeno não é uma coincidência e o investigaremos agora.

Suponhamos que X é somável de modo que $\mathbb{E}[X \mid Y = y]$ é finita para todo $y \in R_Y$.

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

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

Teorema 3.39 *Se X é uma variável aleatória somável, então a variável aleatória $\mathbb{E}[X \mid Y]$ é somável e $\mathbb{E}[\mathbb{E}[X \mid Y]] = \mathbb{E}X$.* □

Na dedução do teorema acima passamos pela seguinte igualdade que também é bastante útil. De fato, o resultado anterior é uma maneira mais abstrata de dizer o seguinte.

Teorema 3.40 (teorema da esperança total) *Se X e Y são variáveis aleatórias*

$$\mathbb{E}X = \sum_{y \in R_Y} \mathbb{E}[X \mid Y = y]\mathbb{P}[Y = y] \tag{3.23}$$

sempre que X é uma variável a aleatória somável. □

Notemos que fazendo $X = \mathbb{1}_A$ na equação (3.23) a expressão que obtemos é o teorema da probabilidade total, teorema 1.26, página 23.

Por exemplo, podemos usar o teorema acima para calcular o número esperado de lançamentos de uma moeda equilibrada até sair coroa (já calculamos essa esperança no exemplo 3.20, página 102) condicionando no resultado do primeiro lançamento. Seja X o número de lançamentos até sair coroa, então usando o teorema 3.40

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

e como uma variável aleatória com distribuição geométrica não tem memória (exercício 3.51, página 129)

$$\begin{aligned}\mathbb{E}[\,X \mid X>1\,] &= \sum_{n\geqslant 1} n\mathbb{P}[\,X=n \mid X>1\,] = \sum_{n\geqslant 2} n\mathbb{P}[\,X=n \mid X>1\,] \\ &= \sum_{n\geqslant 1}(n+1)\mathbb{P}[\,X=n+1 \mid X>1\,] = \sum_{n\geqslant 1}(n+1)\mathbb{P}[X=n] = \mathbb{E}[X]+1\end{aligned}$$

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

Exemplo 3.41. Suponhamos que são recebidas N 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 com mensagens num determinado dia é $X = \sum_{i=1}^{N} X_i$ minutos cuja esperança é $\mathbb{E}X = \mathbb{E}\mathbb{E}[\,X \mid N\,] = \mathbb{E}\psi(N)$. Porém, por causa da independência, da linearidade e do fato de todas as variáveis X_i, $1 \leqslant i \leqslant n$, terem a mesma esperança

$$\psi(n) = \mathbb{E}\left[\sum_{i=1}^{N} X_i \,\middle|\, N=n\right] = \mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n}\mathbb{E}X_i = 8n$$

portanto

$$\mathbb{E}\,\psi(N) = \sum_{n\geqslant 1}\psi(n)\mathbb{P}_N(n) = \sum_{n\geqslant 1} 8n\,\mathbb{P}_N(n) = 8\mathbb{E}N.$$

logo $\mathbb{E}X = 40$ minutos é o tempo médio gasto por dia com mensagens eletrônicas. ◊

3.3.1 O MÉTODO PROBABILÍSTICO REVISITADO

Seja X uma variável aleatória que assume valores não negativos e com quadrado somável (X^2 é somável, logo X também é). Pelo teorema da esperança total,

$$\mathbb{E}[\,X^2\,] = \mathbb{E}[\,X^2 \mid X>0\,]\mathbb{P}[\,X>0\,] + \mathbb{E}[\,X^2 \mid X=0\,]\mathbb{P}[\,X=0\,].$$

No entanto $\mathbb{E}[\,X^2 \mid X=0\,] = 0$ o que nos leva a concluir que $\mathbb{E}[\,X^2\,] = \mathbb{E}[\,X^2 \mid X>0\,]\mathbb{P}[\,X>0\,]$. Ademais, usando a linearidade da esperança condicionada nós obtemos de $\mathbb{E}[\,(X-\mathbb{E}[\,X \mid X>0\,])^2 \mid X>0\,] \geqslant 0$ que $\mathbb{E}[\,X^2 \mid X>0\,] \geqslant \mathbb{E}[\,X \mid X>0\,]^2$. Logo

$$\mathbb{E}[\,X^2\,]\mathbb{P}[\,X>0\,] \geqslant (\mathbb{E}[\,X \mid X>0\,]\mathbb{P}[\,X>0\,])^2 = (\mathbb{E}X)^2$$

donde concluímos a seguinte desigualdade.

Teorema 3.42 (princípio de segundo momento) *Seja* X *uma variável aleatória que assume valores não negativos e com quadrado somável. Então*

$$\mathbb{P}[\,X>0\,] \geqslant \frac{(\mathbb{E}X)^2}{\mathbb{E}[\,X^2\,]} \tag{3.24}$$

se $\mathbb{E}[\,X^2\,] > 0$. □

Ademais, se $X \geqslant 0$ assume valores inteiros, então $\mathbb{E}X = \mathbb{E}[X \mid X > 0]\mathbb{P}[X > 0]$ donde

$$\mathbb{P}[X > 0] \leqslant \mathbb{E}X \tag{3.25}$$

que é uma versão quantitativa do princípio de primeiro momento (exercício 3.34).

De um modo geral, se $\mathbb{E}X \to 0$, então o primeiro momento garante que $\mathbb{P}[X > 0] = o(1)$ e dizemos que $X = 0$ com alta probabilidade, ou *quase certamente*. Por outro lado, $\mathbb{E}X \to \infty$ não significa que $X > 0$ com alta probabilidade, entretanto se $\mathbb{E}X^2 = (1 + o(1))(\mathbb{E}X)^2$, então pelo segundo momento temos $\mathbb{P}(X > 0) = 1 - o(1)$ e dizemos que $X > 0$ com alta probabilidade, ou *quase certamente*. O exercício 3.75 no final do capítulo dá um contra-exemplo para a afirmação "se $\mathbb{E}X \to \infty$ então $X > 0$ com alta probabilidade".

Exemplo 3.43 (*triângulos em grafos aleatórios*). Três vértices de um grafo definem um triângulo, denotado genericamente por K^3, se todas as três arestas definidas pelos vértice estão presentes. Denotamos por $\mathcal{G}_{n,p}$ o modelo binomial de grafo aleatório obtido considerando o conjunto de vértices $V = \{1, 2, \ldots, n\}$ e cada par $\{u, v\}$ está presente no conjunto de arestas com probabilidade p. Dada uma tripla de vértices $\{u, v, w\} \subset V$, eles formam um triângulo no $\mathcal{G}_{n,p}$ com probabilidade p^3.

Se X é a quantidade de triângulos no $\mathcal{G}_{n,p}$, o valor esperado para número de triângulos é

$$\mathbb{E}X = \sum_{i=1}^{\binom{n}{3}} \mathbb{E}X_i = \sum_{i=1}^{\binom{n}{3}} p^3 = \binom{n}{3} p^3 = \left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)\frac{n^3 p^3}{6},$$

onde X_i é variável aleatória indicadora de presença de triângulo na i-ésima tripla de vértices, considerando uma enumeração qualquer das $\binom{n}{3}$ tais triplas. Tomemos $p = n^{-\alpha}$ para uma constante $\alpha > 0$.

Se $\alpha > 1$, então $\mathbb{E}X = \Theta(n^{3-3\alpha}) = o(1)$ quando $n \to \infty$, logo $\mathbb{P}[X > 0] = o(1)$ pelo primeiro momento, ou seja, quase certamente $\mathcal{G}_{n,p}$ não tem triângulo nesse caso.

Se $\alpha < 1$ então $\mathbb{E}X \to \infty$ quando $n \to \infty$. Nesse caso vamos estimar a somatória

$$\mathbb{E}X^2 = \mathbb{E}\left[\left(\sum_{i=1}^{\binom{n}{3}} X_i\right)^2\right] = \mathbb{E}\left[\left(\sum_{i=1}^{\binom{n}{3}} X_i^2 + \sum_{i \neq j} X_i \cdot X_j\right)\right] = \sum_{i=1}^{\binom{n}{3}} \mathbb{E}[X_i^2] + \sum_{i \neq j} \mathbb{E}[X_i \cdot X_j] \tag{3.26}$$

em que o último somatório é sobre todo par de inteiros distintos $1 \leqslant i, j \leqslant m$. Para cada $i \neq j$, $\mathbb{E}[X_i \cdot X_j] = \mathbb{P}([X_i = 1] \cap [X_j = 1])$ é a probabilidade das i-ésima e j-ésima triplas de vértice formarem triângulos. Isso pode ocorrer de três modos de acordo com o número de vértices em comum entre as triplas

1. as duas triplas de vértices têm juntas 4 vértices e os triângulos uma aresta em comum, nesse caso $\mathbb{P}([X_i = 1] \cap [X_j = 1]) = p^5$;
2. as duas triplas têm juntas 5 vértices, um vértice em comum, nesse caso temos $\mathbb{P}([X_i = 1] \cap [X_j = 1]) = p^6$;
3. as duas triplas são disjuntas, nesse caso $\mathbb{P}([X_i = 1] \cap [X_j = 1]) = p^6$.

A contribuição do primeiro caso para $\sum_{i \neq j} \mathbb{E}[X_i \cdot X_j]$ é de no máximo $\binom{n}{4}$ subconjuntos de quatro vértices, em cada um há $\binom{4}{2}$ modos de escolher a aresta em comum, e as arestas ocorrem com probabilidade p^5; assintoticamente, $\binom{n}{4}\binom{4}{2}p^5 = O(n^4 p^5) = o(1)(\mathbb{E}X)^2$. Nos outros dois casos $\mathbb{E}[X_i \cdot X_j] = \mathbb{E}X_i \cdot \mathbb{E}X_j = p^6$ pois, como não há aresta em comum, as variáveis são independentes (teorema 3.14, item 6); a contribuição para $\sum_{i \neq j} \mathbb{E}[X_i \cdot X_j]$ é no máximo $\sum_i \mathbb{E}X_i \sum_j \mathbb{E}X_j \leqslant (\mathbb{E}X)^2$. Para o primeiro somatório no lado dirito da equação (3.26) usamos que $X_i^2 = X_i$, logo

$$\sum_{i=1}^{\binom{n}{3}} \mathbb{E}[X_i^2] + \sum_{i \neq j} \mathbb{E}[X_i \cdot X_j] \leqslant \mathbb{E}X + (1 + o(1))(\mathbb{E}X)^2$$

então do segundo momento obtemos

$$\mathbb{P}[\,X=0\,]<1-\frac{(\mathbb{E}X)^2}{\mathbb{E}X^2}<1-\frac{(\mathbb{E}X)^2}{\mathbb{E}X+(1+o(1))(\mathbb{E}X)^2}=o(1)$$

pois $\mathbb{E}X\to\infty$, assim $\mathcal{G}_{n,p}$ contém triângulo quase certamente.

Se $p=cn^{-1}$ ($c>0$ constante) então $\mathbb{E}X\to c^3/6$, quando $n\to\infty$. Nesse caso é possível demonstrar que X converge para a distribuição de Poisson, em particular, $\lim_{n\to\infty}\mathbb{P}[\,X=0\,]=e^{-c^3/6}$. ◊

Suponha que n bolas são distribuídas uniforme e independentemente em m caixas e seja X_i variável aleatória indicadora do evento "i-ésima caixa vazia", para $i=1,2,\ldots,m$. A quantidade de caixas vazias X tem esperança $\mathbb{E}X=\sum_{i=1}^{m}\mathbb{P}[\,X_i=1\,]=m(1-1/m)^n$ logo (usando (d.1)) $\mathbb{E}X\approx me^{-n/m}=1$ se $n=n^*(m)=m\log(m)$. Vamos usar as desigualdades de momento dadas acima para mostrar que se n é suficientemente menor que $n^*(m)$ então quase certamente ocorre caixa vazia e se n é suficientemente maior que $n^*(m)$ então quase certamente não ocorre caixa vazia.

Lema 3.44 *Suponha que n bolas são distribuídas aleatoriamente, uniforme e independentemente, em m caixas. Dado qualquer constante $\varepsilon>0$, se $n>(1+\varepsilon)m\log(m)$ então não há caixa vazia com alta probabilidade e se $n<(1-\varepsilon)m\log(m)$ então há caixa vazia com alta probabilidade.*

Demonstração. Dado $\varepsilon>0$, seja X a quantidade de caixas vazias e X_i variável aleatória indicadora de "i-ésima caixa vazia", para todo $i=1,2,\ldots,m$. O número de caixas vazias é $X=\sum X_i$ e

$$\mathbb{E}X=\sum_{i=1}^{m}\mathbb{P}[\,X_i=1\,]=m\Big(1-\frac{1}{m}\Big)^n=\Theta\big(me^{-n/m}\big)$$

e se $n>(1+\varepsilon)m\log(m)$ então $\mathbb{E}X=O(m^{-\varepsilon})$, portanto, pela equação (3.25) vale $\mathbb{P}[\,X>0\,]\leqslant Cm^{-\varepsilon}$ para alguma constante $C>0$, ou seja, a probabilidade com que nenhuma caixa está vazia é alta

$$\mathbb{P}[\,X=0\,]>1-Cm^{-\varepsilon}. \tag{3.27}$$

Por outro lado, se $n<(1-\varepsilon)m\log(m)$ então $\mathbb{E}X=\Omega(m^{\varepsilon})$ mas isso não assegura que há caixas vazias com alta probabilidade. Nesse caso usamos segundo memento

$$\mathbb{E}X^2=\sum_{i=1}^{m}\mathbb{E}[\,X_i^2\,]+\sum_{i\neq j}\mathbb{E}[\,X_i\cdot X_j\,]$$

em que o último somatório é sobre todo par de inteiros distintos $1\leqslant i,j\leqslant m$, como em (3.26). Agora, $X_i^2=X_i$, por ser variável indicadora, logo $\mathbb{E}[\,X_i^2\,]=\mathbb{P}[\,X_i=1\,]=(1-1/m)^n$. Também, $X_i\cdot X_j$ é uma variável aleatória indicadora, do evento "ambas caixas estão vazias" e $\mathbb{E}[\,X_i\cdot X_j\,]=\mathbb{P}([\,X_i=1\,]\cap[\,X_j=1\,])=(1-2/m)^n$, portanto

$$\begin{aligned}\frac{(\mathbb{E}X)^2}{\mathbb{E}X^2}&=\frac{m^2\big(1-\frac{1}{m}\big)^{2n}}{m\big(1-\frac{1}{m}\big)^{n}+m(m-1)\big(1-\frac{2}{m}\big)^{n}}\geqslant\frac{m\big(1-\frac{1}{m}\big)^{2n}}{\big(1-\frac{1}{m}\big)^{n}+m\big(1-\frac{1}{m}\big)^{2n}}\\&=\frac{1}{\frac{1}{m}\big(1-\frac{1}{m}\big)^{-n}+1}>\frac{1}{\frac{1}{m}\big(1-\frac{1}{m}\big)^{-m\log(m^{1-\varepsilon})}+1}>\frac{1}{m^{-\varepsilon}+1}\end{aligned}$$

que tende a 1 quando $m\to\infty$. Usando a desigualdade (3.24) concluímos que a probabilidade de ter não ter pelo menos uma caixa vazia é

$$\mathbb{P}[\,X=0\,]<1-\frac{1}{(m^{-\varepsilon}+1)}. \tag{3.28}$$

Em resumo, temos das equações (3.27) e (3.28) que a probabilidade de não haver caixas vazias é

$$\mathbb{P}[\,X=0\,]=\begin{cases}1-o(1), & \text{se } n>(1+\varepsilon)m\log(m),\\ o(1), & \text{se } n<(1-\varepsilon)m\log(m),\end{cases}$$

onde $o(1)$ expressa uma função (não negativa) de m que tende a 0 quando m tende ao infinito. □

Do desenvolvimento acima temos que (1) se $\mathbb{E}X = o(1)$ então $\mathbb{P}(X = 0) = 1 - o(1)$ e (2) se $X_1, \ldots, X_n$ são variáveis aleatórias de Bernoulli de mesmo parâmetro e $X = \sum_i X_i$ então

$$\mathbb{E}X^2 = \mathbb{E}X + \sum_{i \neq j} \mathbb{E}[X_i \cdot X_j] \tag{3.29}$$

e se $\sum_{i \neq j} \mathbb{E}[X_i \cdot X_j] \leqslant (1 + o(1))(\mathbb{E}X)^2$ enquanto $\mathbb{E}X \to \infty$, então $\mathbb{P}[X = 0] = o(1)$. Ademais

$$\mathbb{P}[X = 0] = \begin{cases} 1 - o(1), & \text{se } \mathbb{E}X = o(1), \\ o(1), & \text{se } \mathbb{E}X \to \infty. \end{cases}$$

Vamos aplicar esse princípio no próximo exemplo.

Carga máxima no caso $n = m$ Agora, suponhamos que n bolas são distribuídas aleatoriamente, uniforme e independentemente, em $m = n$ caixas. Seja X a quantia de caixas com pelo menos k bolas e X_i variável aleatória indicadora de ocorrência de k ou mais bolas na caixa i, para todo $i = 1, 2, \ldots, m$. Assim, $X = \sum X_i$.

Há $\binom{n}{k}$ modos de escolher um conjunto de k bolas, as quais estão numa caixa específica com probabilidade $(1/n)^k$. Para as bolas restantes ignoramos o destino, portanto pelo corolário 1.5, página 10, a probabilidade de uma caixa ter pelo menos k bolas é

$$\mathbb{E}X_i = \mathbb{P}[X_i = 1] \leqslant \binom{n}{k}\left(\frac{1}{n}\right)^k \leqslant \left(\frac{en}{k}\right)^k \frac{1}{n^k} = \left(\frac{e}{k}\right)^k.$$

Ainda,

$$\mathbb{P}[X_i = 1] \geqslant \binom{n}{k}\left(\frac{1}{n}\right)^k \left(1 - \frac{1}{n}\right)^{n-k} \geqslant \frac{n^k}{k^k}\frac{1}{n^k}\frac{1}{e} \geqslant \frac{1}{ek^k}$$

de modo que

$$\frac{1}{ek^k} \leqslant \mathbb{E}X_i \leqslant \left(\frac{e}{k}\right)^k.$$

Ingenuamente, a transição de $X = 0$ quase certamente para $X > 0$ quase certamente se dá quando $\mathbb{E}X \approx n(e/k)^k \approx 1$, ou seja, quando $k\log(k) \approx \log(n)$, portanto quando $k \approx \log(n)/\log(\log(n))$.

Fazendo $k = 4\log(n)/(\log\log(n))$ temos $\mathbb{E}X = o(1)$, para $n \to \infty$. De fato,

$$\left(\frac{e}{k}\right)^k = \left(\frac{e\log\log(n)}{4\log(n)}\right)^k \leqslant \left(\frac{\log\log(n)}{\log(n)}\right)^{\frac{4\log(n)}{\log\log(n)}} = \left(e^{\log\log\log(n) - \log\log(n)}\right)^{\frac{4\log(n)}{\log\log(n)}}$$
$$= e^{-4\log(n)\left(1 - \frac{\log\log\log(n)}{\log\log(n)}\right)} = n^{-4 + \frac{4\log\log\log(n)}{\log\log(n)}} \leqslant n^{-4(1+1/e)} < n^{-2}$$

pois $4\log\log\log(n)/\log\log(n)$ tem valor máximo 4/e em $n = e^{e^e}$. Portanto, por (3.25), $\mathbb{P}(X > 0) \leqslant \mathbb{E}X \leqslant n \cdot n^{-2}$, isto é, $\mathbb{P}(X = 0) = 1 - o(1)$ de modo que com alta probabilidade a carga máxima numa caixa é $O(\log n/\log(\log n))$.

Agora, se $k = \log(n)/(3\log\log(n))$ então $\mathbb{E}X \to \infty$, para $n \to \infty$. De fato,

$$\left(\frac{1}{k}\right)^k = \left(\frac{3\log\log(n)}{\log(n)}\right)^{\frac{\log(n)}{3\log\log(n)}} \geqslant \left(\frac{\log\log(n)}{\log(n)}\right)^{\frac{\log(n)}{3\log\log(n)}} = \left(e^{\log\log\log(n) - \log\log(n)}\right)^{\frac{\log(n)}{3\log\log(n)}}$$
$$= e^{-\frac{\log(n)}{3}\left(1 - \frac{\log\log\log(n)}{\log\log(n)}\right)} = n^{-\frac{1}{3}\left(1 - \frac{\log\log\log(n)}{\log\log(n)}\right)}$$

de modo que $\mathbb{E}X \geqslant n \cdot n^{-1/3 + \log\log\log(n)/\log\log(n)} = n^{2/3 + o(1)}$.

Para calcular $\mathbb{E}X^2$ usamos a linearidade da esperança como na equação (3.29) e temos

$$\mathbb{E}X^2 = \mathbb{E}X + n(n-1)\mathbb{E}[X_i \cdot X_j]$$

para quaisquer $i \neq j$ e precisamos estimar a probabilidade

$$\mathbb{E}[\,X_i \cdot X_j\,] = \mathbb{P}\Big([X_i = 1] \cap [X_j = 1]\Big) = \sum_{k_1=k}^{n-k} \sum_{k_2=k}^{n-k_1} \binom{n}{k_1}\binom{n-k_1}{k_2}\Big(\frac{1}{n}\Big)^{k_1+k_2}\Big(1-\frac{2}{n}\Big)^{n-k_1-k_2}.$$

Começamos com algumas simplificações

$$\begin{aligned}
\sum_{k_1=k}^{n-k} \sum_{k_2=k}^{n-k_1} \binom{n}{k_1}\binom{n-k_1}{k_2}\Big(\frac{1}{n}\Big)^{k_1+k_2}\Big(1-\frac{2}{n}\Big)^{n-k_1-k_2} &\leqslant \sum_{k_1=k}^{n} \sum_{k_2=k}^{n} \binom{n}{k_1}\binom{n}{k_2}\Big(\frac{1}{n}\Big)^{k_1+k_2}\Big(1-\frac{1}{n}\Big)^{2(n-k_1-k_2)} \\
&= \sum_{k_1=k}^{n}\left(\binom{n}{k_1}\Big(\frac{1}{n}\Big)^{k_1}\Big(1-\frac{1}{n}\Big)^{n-2k_1} \sum_{k_2=k}^{n}\binom{n}{k_2}\Big(\frac{1}{n}\Big)^{k_2}\Big(1-\frac{1}{n}\Big)^{n-2k_2}\right) \\
&= \sum_{k_1=k}^{n}\left(\mathrm{b}_{n,\frac{1}{n}}(k_1)\Big(1-\frac{1}{n}\Big)^{-k_1} \sum_{k_2=k}^{n}\mathrm{b}_{n,\frac{1}{n}}(k_2)\Big(1-\frac{1}{n}\Big)^{-k_2}\right) \\
&\leqslant \Big(1-\frac{1}{n}\Big)^{-2k} \sum_{k_1=k}^{n}\left(\mathrm{b}_{n,\frac{1}{n}}(k_1) \sum_{k_2=k}^{n}\mathrm{b}_{n,\frac{1}{n}}(k_2)\right).
\end{aligned}$$

Porém, $\mathbb{E}\,X_i = \sum_{x=k}^{n} \mathrm{b}_{n,\frac{1}{n}}(x)$, logo

$$\mathbb{E}[\,X_i \cdot X_j\,] \leqslant \Big(1-\frac{1}{n}\Big)^{-2k} \sum_{k_1=k}^{n}\left(\mathrm{b}_{n,\frac{1}{n}}(k_1) \sum_{k_2=k}^{n}\mathrm{b}_{n,\frac{1}{n}}(k_2)\right) = \Big(1-\frac{1}{n}\Big)^{-2k} (\mathbb{E}\,X_i)(\mathbb{E}\,X_j)$$

e o valor esperado de X^2 é limitado por

$$\mathbb{E}\,X^2 \leqslant \mathbb{E}\,X + n^2\Big(1-\frac{1}{n}\Big)^{-2k}(\mathbb{E}\,X_i)(\mathbb{E}\,X_j) = \mathbb{E}\,X + \Big(1-\frac{1}{n}\Big)^{-2k}(\mathbb{E}\,X)^2.$$

Finalmente, usando a desigualdade de segundo momento, equação (3.24),

$$\mathbb{P}[\,X = 0\,] < 1 - \frac{\mathbb{E}[\,X\,]^2}{\mathbb{E}\,X^2} \leqslant 1 - \frac{1}{\frac{1}{\mathbb{E}\,X} + \Big(1-\frac{1}{n}\Big)^{-2k}}$$

mas $1/\mathbb{E}\,X \to 0$ e $(1-1/n)^{-2k} \approx \mathrm{e}^{2k/n} \to 1$, pela escolha de k, donde concluímos que $\mathbb{P}[\,X = 0\,] = o(1)$, ou seja, a carga máxima é $\Omega(\log(n)/\log(\log(n)))$ com alta probabilidade.

3.3.2 *DESALEATORIZAÇÃO*: O MÉTODO DAS ESPERANÇAS CONDICIONAIS

Vimos na página 55, 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 $[X_1 = x_1] \cap [X_2 = x_2] \cap \cdots \cap [X_i = x_i]$.

Seja $f\colon S^n \to \mathbb{R}$ uma função em que S é finito e que $\mathbb{E}\,f(X_1,\ldots,X_n) \geqslant \mu$. Por exemplo, no MAX-CUT $(X_1,\ldots,X_n) \in_{\mathcal{U}} \{0,1\}^n$ corresponde a uma bipartição nos vértices de um grafo de n vértices e f é o número de arestas no corte, nesse caso sabemos que $\mathbb{E}\,f$ é pelo menos metade do número de arestas do grafo.

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 a esperança condicional $\mathbb{E}[\,f(X_1,\ldots,X_n) \mid X_1 = x_1,\ldots,X_i = x_i\,]$ de modo eficiente. Para cada $i = 1, 2, \ldots, n$, dados $x_1,\ldots,x_i$ tais que $\mathbb{E}[\,f(X_1,\ldots,X_n) \mid X_1 = x_1,\ldots,X_i = x_i\,] \geqslant \mu$, tomamos

$$\psi(r) \coloneqq \mathbb{E}[\,f(X_1,\ldots,X_n) \mid X_1 = x_1,\ldots,X_i = x_i, X_{i+1} = r\,]$$

pelo teorema 3.39 o valor esperado $\mathbb{E}\,\psi(X_{i+1})$ é

$$\mathbb{E}[\,\mathbb{E}[\,f(X_1,\ldots,X_n)\,|\,X_1=x_1,\ldots,X_i=x_i,X_{i+1}\,]\,]=\mathbb{E}[\,f(X_1,\ldots,X_n)\,|\,X_1=x_1,\ldots,X_i=x_i\,]\geqslant\mu$$

logo, pelo princípio de primeiro momento, para algum r temos $\psi(r)\geqslant\mu$; testamos para cada $r\in S$ se

$$\mathbb{E}[\,f(X_1,\ldots,X_n)\,|\,X_1=x_1,\ldots,X_i=x_i,X_{i+1}=r\,]\stackrel{?}{\geqslant}\mu$$

Repetindo essa estratégia para cada i ao final teremos $(x_1,\ldots,x_n)$ tal que $f(x_1,\ldots,x_n)\geqslant\mu$.

MAX-3-SAT O problema que estudamos na seção 3.2.1 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 $X_1,\ldots,X_n\in_{\mathcal{U}}\{0,1\}$ 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 3.2.1

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

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

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

para $r=0$ 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 $\mathbb{1}_{[C_i=1]}$ é 7/8;
2. se a valoração parcial deixa 2 variáveis livres na cláusula então 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 o valor esperado condicional de $\mathbb{1}_{[C_i=1]}$ 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 3.2.2 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 $f(x_1,\ldots,x_n):=|\nabla(A)|$ em que $(x_1,\ldots,x_n)$ é o vetor característico de A, como definido acima. Sejam $X_1,\ldots,X_n$ variáveis aleatórias independentes com distribuição uniforme em $\{0,1\}$, sabemos da seção 3.2.2 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 as esperanças $\mathbb{E}[\, f(X_1,\ldots,X_n) \,|\, X_1 = x_1,\ldots,X_{i+1} = r\,]$ para $r = 0$ e para $r = 1$; 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

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

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 $\mathbb{E}[\, f(X_1,\ldots,X_n) \,|\, X_1 = x_1,\ldots,X_{i+1} = x_{i+1}\,]$ 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 $O(nm)$.

3.3.3 *SKIP LISTS*

Uma **skip list** também é uma estrutura de dados para representar conjuntos dinâmicos (Pugh, 1989). Essa estrutura de dados é uma coleção de listas ligadas que usa aleatoriedade para se comportar, em desempenho, como uma árvore de busca balanceada[4] e com a vantagem de ser mais eficiente e mais fácil de manter que uma árvore balanceada.

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 aponta para sua cópia imediatamente abaixo no nível S_i. No último nível temos $S_\ell = \varnothing$. Além dos elementos de S mantemos dois sentinelas, o $-\infty$ no começo de todas as listas e o $+\infty$ no fim de todas as listas. A figura 3.3 abaixo descreve um exemplo de uma *skip list*.

As operações de dicionário na *skip list* são descritas a seguir.

busca: dado $x \in U$ uma busca devolve um apontador para $x \in S$ ou para o menor elemento de S 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. 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

[4] 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.

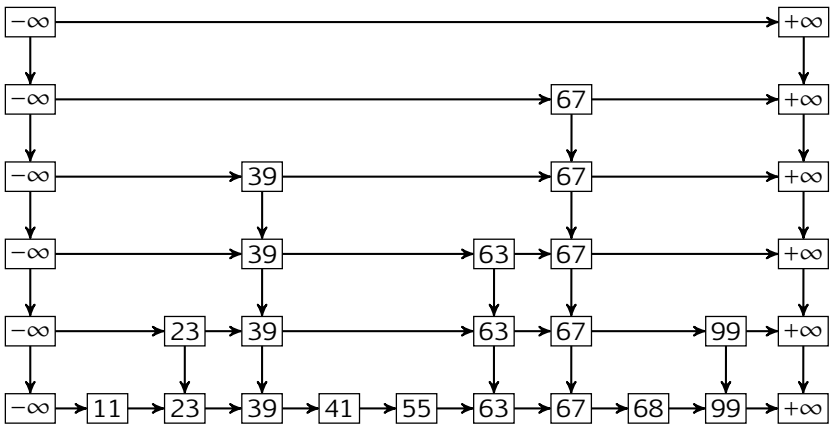


Figura 3.3: exemplo de uma *skip list* com 6 níveis que representa o conjunto $\{11, 23, 39, 41, 55, 63, 67, 68, 99\}$.

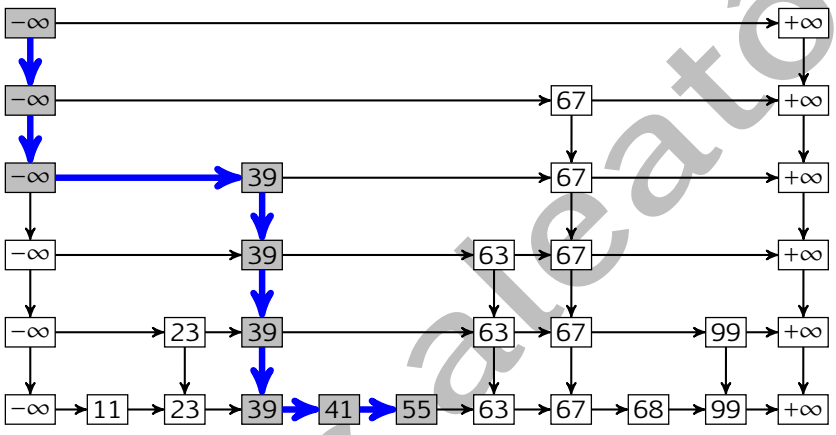


Figura 3.4: o caminho da busca por 55 na *skip list* da figura 3.3.

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, teremos que aumentar 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 $\log_2(n)$. Vamos estudar essa estrutura de dados e para isso começamos definindo alguns parâmetros importantes. Façamos $n := |S|$, 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. Podemos computar a esperança da variável aleatória H usando o fato dela assumir valores inteiros positivos, pela proposição 3.21, página 103,

$$\mathbb{E}\,\mathsf{H} = \sum_{i\geqslant 1} \mathbb{P}[\,\mathsf{H} \geqslant i\,] = \sum_{i=0}^{\lceil \log_2 n\rceil} \mathbb{P}[\,\mathsf{H} > i\,] + \sum_{i>\lceil \log_2 n\rceil} \mathbb{P}[\,\mathsf{H} > i\,]. \tag{3.30}$$

Claramente, $h(x) \in_{\mathcal{G}(\frac{1}{2})} \mathbb{N}$ logo $\mathbb{P}[\,h(x) = t\,] = (1/2)^{t-1}(1/2)$ para todo inteiro positivo t. Do exercício 3.5, página 91, concluímos que $\mathbb{P}[\,h(x) > t\,] = 2^{-t}$. Reunindo essas informações

$$\mathbb{P}[\,h(x) = t\,] = 2^{-t} = \mathbb{P}[\,h(x) > t\,]. \tag{3.31}$$

De volta na equação (3.30), na primeira soma no lado direto dessa equação limitaremos trivialmente a probabilidade tomando $\mathbb{P}[\,\mathsf{H} > i\,] \leqslant 1$, de modo que

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

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

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

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.

Proposição 3.45 *Se* S *é uma* skip list *sobre* n *elementos então para a altura* $\mathsf{H} = \mathsf{H}(\mathsf{S})$ *valem*

$$\mathbb{E}\,\mathsf{H} \leqslant \log_2(4n) \qquad e \qquad \mathbb{P}[\,\mathsf{H} > 2\log_2(n)\,] < 1/n.$$

Demonstração. Para a esperança de H basta observar que $\log_2(n) + 2 = \log_2(4n)$. Da equação (3.31) deduzimos que a probabilidade de um elemento qualquer de S ter altura maior que $c\log_2 n$ é $2^{-c\log_2 n} = n^{-c}$ para qualquer constante positiva c, portanto, a probabilidade de existir algum elemento em S com altura maior que $c\log_2 n$ é

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

portanto, fazendo $c = 2$ segue que $\mathbb{P}[\,\mathsf{H} > 2\log_2(n)\,] \leqslant 1/n$. □

Observação 3.46 (*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 $\mathbb{E}[\,h(x)\,] = 2$

$$\mathbb{E}\,\mathsf{N} = \sum_{x\in \mathsf{S}} \mathbb{E}[\,h(x)\,] = 2n. \tag{3.32}$$

Notemos que de $N = \sum_{i=1}^{H} |S_i|$ não é imediato valer que $\mathbb{E}N = \sum_{i=1}^{\mathbb{E}H} \mathbb{E}|S_i|$ por causa de independência ou não das variáveis envolvidas (veja o exercício 3.61 no final deste capítulo). Pode-se provar que o número de itens N é $O(n)$ com alta probabilidade e faremos isso na seção 4.3.1. ◊

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 3.47 *O número esperado de comparações feitas nos níveis mais altos de uma* skip list, *do nível* $\lceil \log_2 n \rceil + 1$ *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

$$\ell := \lceil \log_2 n \rceil + 1, \quad p := (1/2)^{\ell} \quad \text{e} \quad N' := \sum_{i=\ell}^{H} |S_i|.$$

Notemos que em $S_\ell, S_{\ell+1}, S_{\ell+2}, \ldots, S_H$ temos uma *skip list* para o conjunto S_ℓ logo pela equação (3.32), $\mathbb{E}[N' \,|\, |S_\ell| = k] \leqslant 2k$. Usando o teorema da esperança total, equação (3.23) na página 117, escrevemos

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

e como qualquer $x \in S$ ocorre em S_ℓ com probabilidade p, temos $|S_\ell| \in_{b(p)} \{0, 1, \ldots, n\}$ de modo que

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

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

$$np = n\left(\frac{1}{2}\right)^{\ell} = n\left(\frac{1}{2}\right)^{\lceil \log_2 n \rceil + 1} \leqslant n\left(\frac{1}{2}\right)^{\log_2 n} = 1$$

portanto $\mathbb{E}N' \leqslant 2$, 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, veremos que a probabilidade da busca ficar muito tempo num nível é pequena, portanto, o tempo gasto é essencialmente determinado pela altura da *skip list*. A ideia chave é explicada no próximo parágrafo.

Seja $x_1 < x_2 < \cdots < x_n$ a lista S_0 e denotemos por y o elemento que está sendo buscado. Suponhamos que a busca tenha chegado em S_i a partir de S_{i+1} pelo elemento x_j. Notemos que $x_j \in S_{i+1}$ e seja $x_m \in S_{i+1}$ o próximo elemento de S_{i+1} depois de x_j. Com essas hipóteses $x_j \leqslant y < x_m$ e nenhum dentre os elementos consecutivos $x_{j+1}, x_{j+2}, \ldots, x_{m-1} \in S_0$ pertence a S_{i+1}. Se a busca realiza t passos em S_i então temos uma sequência $x_{a_1}, x_{a_2}, \ldots, x_{a_t}$ com os t maiores elementos que são menores ou iguais a y em S_i, precedidos por x_j e que não ocorrem em S_{i+1} (veja a figura 3.5); a sequência $x_j, x_{a_1}, x_{a_2}, \ldots, x_{a_t}$ com a configuração descrita acima na *skip list* corresponde a lançamentos de moeda que resultam em uma cara seguida de t coroas, já que só o primeiro elemento da sequência é replicado em S_{i+1}.

Lema 3.48 *Numa busca sem sucesso em* S, *o número esperado de comparações feitas no nível* i *é* $O(1)$ *para cada* i, $0 \leqslant i \leqslant \lceil \log_2 n \rceil$.

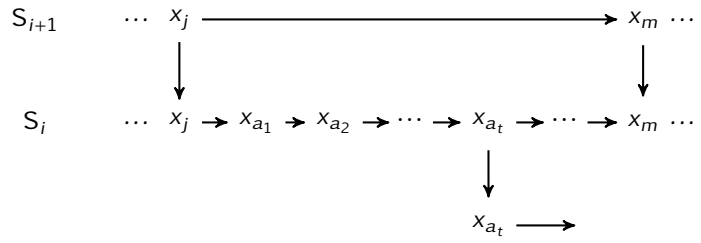


Figura 3.5: visão de uma busca no nível S_i.

Demonstração. Seja y o elemento buscado, S uma *skip list* e suponhamos que $y \notin S$. Seja X_i o número de passos dados pela busca em S_i, para $0 \leqslant i \leqslant \lceil \log_2 n \rceil$. No exemplo de busca ilustrado na figura 3.4, página 125, $X_5 = X_4 = X_2 = X_1 = 0$, $X_3 = 1$ e $X_0 = 2$.

A quantidade elementos em S_i menores ou iguais a y é uma variável aleatória que denotamos por $M_i(y)$. Fixado i e condicionando a $M = M_i(y)$ o número esperado de passos no nível i é

$$\mathbb{E}\, X_i = \sum_{k=1}^{n} \mathbb{E}[\, X_i \mid M = k\,]\mathbb{P}[\, M = k\,]$$

Para todo k e todo t tal que $1 \leqslant k \leqslant n$ e $0 \leqslant t \leqslant k$ temos $X_i = t$, dado que $M = k$, se e só se os t maiores elementos menores ou iguais a y em S_i não ocorrem em S_{i+1}, mas o elemento que precede esses t ocorre em S_{i+1} (tal predecessor pode ser o sentinela), logo

$$\mathbb{P}[\, X_i = t \mid M = k\,] \leqslant \left(\frac{1}{2}\right)^t$$

de modo que usando (s.6c) obtemos

$$\mathbb{E}[\, X_i \mid M = k\,] = \sum_{j=0}^{k} j\mathbb{P}[\, X_i = j \mid M = k\,] \leqslant \sum_{j=0}^{k} \frac{j}{2^j} \leqslant 2$$

e, então,

$$\mathbb{E}\, X_i = \sum_{k=1}^{n} \mathbb{E}[\, X_i \mid M = k\,]\mathbb{P}[\, M = k\,] \leqslant \sum_{k=1}^{n} 2\mathbb{P}[\, M = k\,] = 2.$$

ou seja, o número esperado de passos no nível i é no máximo 2. □

Como o número de comparações em cada nível é constante, o número total de comparações é proporcional ao número de níveis, a saber $\lceil \log_2 n \rceil + 1$.

Teorema 3.49 *O número esperado de comparações feitas por uma busca em uma* skip list *que representa um conjunto com n elementos é* $O(\log n)$.

Demonstração. Dados uma *skip list* S e y, consideremos uma busca por y em S. Podemos assumir $y \notin S$, pois isso resulta num limitante superior para o custo de uma busca.

Com a notação do resultado anterior o número de comparações feitas pela busca no nível i, contando com a comparação que faz o apontador descer um nível (ou descobrir que $x \notin S$ achando alguém maior) é $X_i + 1$. O custo da busca nos $O(\log n)$ níveis mais baixos da *skip list* é

$$\mathbb{E} \sum_{i=0}^{\lceil \log_2 n \rceil} (X_i + 1) = \sum_{i=0}^{\lceil \log_2 n \rceil} \mathbb{E}\, X_i + 1 \leqslant 3 \log_2 n$$

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

3.4 EXERCÍCIOS

EXERCÍCIO 3.50. 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. $A \subset B \Rightarrow \mathbb{1}_A \leqslant \mathbb{1}_B$.
4. $\mathbb{1}_{A\cap B} = \mathbb{1}_A \cdot \mathbb{1}_B$.

EXERCÍCIO 3.51 (variáveis geométricas não têm memória). Sejam $X \in_{\mathcal{G}(p)} \mathbb{N}$ 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 3.52. 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 3.53. 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 3.54. 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 3.55. 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 3.56. 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 3.57. 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 $i\binom{n}{i} = n\binom{n-1}{i-1}$. Agora, use essa identidade para provar que $\mathbb{E}[\, Y^k \,] = np\mathbb{E}[\,(X+1)^{k-1}\,]$ vale para $X \in_{b(p)} \{0, 1, \ldots, n-1\}$ e $Y \in_{b(p)} \{0, 1, \ldots, n\}$, com $n > 0$ e $p \in (0, 1)$. Conclua que $\mathbb{E}Y = np$.

EXERCÍCIO 3.58. Sejam U e M conjuntos finitos. Uma família de funções $\mathcal{H} \subset M^U$ que quando munida da medida uniforme satisfaz

$$\mathop{\mathbb{P}}_{h\in\mathcal{H}}[h(u) = i] = \frac{1}{|M|}, \text{ para todo } u \in U \text{ e para todo } i \in M \tag{3.33}$$

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 (3.33). Nesse caso, toda $h \in \mathcal{H}$ resulta no pior caso?

EXERCÍCIO 3.59. 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 $O(n)$. Determine a probabilidade com que o algoritmo faz mais que $O(n \log n)$ comparações (*dica:* use o particionamento do *quicksort*).

EXERCÍCIO 3.60. Dado um conjunto de n números e um real positivo ε, tome uma amostra aleatória de $\Theta(\varepsilon^{-1}\log n)$ elementos e ordene-os usando $O(\log n \log\log n)$ comparações. Prove que o resultado é uma ε-aproximação da mediana dos números dados com probabilidade de erro no máximo n^{-2}.

EXERCÍCIO 3.61. Considere o espaço amostral

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

e as variáveis aleatórias $X_i(\omega)$ que dá a i-ésima coordenada de ω, para $i = 1,2,3$ e todo $\omega \in \Omega$. Prove que

- para todos $i, j \in \{1,2,3\}$, vale $\mathbb{P}[X_i = j] = 1/3$;
- as variáveis X_i não são independentes.

Seja N uma variável aleatória sobre $\{1,2,3\}$ com a mesma distribuição de X_2. Prove que

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

EXERCÍCIO 3.62 (linearidade da esperança para soma enumerável). Prove que se $X = \sum_{i\geqslant 1} X_i$ e $\sum_{i\geqslant 1}\mathbb{E}|X_i|$ converge, então $\mathbb{E}X = \sum_{i\geqslant 1}\mathbb{E}X_i$.

EXERCÍCIO 3.63. Verifique que a hipótese de convergência é necessária no exercício anterior. 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 3.64. Na seção 3.1.2, página 101, foi dito que se um tabela de espalhamento usa um função aleatória 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, de fato, não há colisão com probabilidade pelo menos $1/2$. Portanto, se S é um conjunto 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 3.65. 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 3.66. Um **conjunto independente** em um grafo $G = (V, E)$ é um subconjunto de vértices $U \subset V$ 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 2.4, na página 51, $\{2, 6\}$ 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 $G = (V, E)$;

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

$$\sum_{u \in V} \frac{1}{d(u) + 1} \tag{3.34}$$

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

Prove o teorema de Turán: *todo grafo* G *admite um conjunto independente de cardinalidade pelo menos* (3.34).

EXERCÍCIO 3.67. 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 3.68. 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 3.69. Prove, usando o método probabilístico, que toda fórmula booleana CNF $C_1 \wedge \cdots \wedge C_k$ sem variáveis repetidas em cada cláusula, admite uma valoração que satisfaz metade das cláusulas. Derive desse fato um algoritmo que recebe uma fórmula CNF e devolva uma valoração que satisfaça metade das cláusulas.

EXERCÍCIO 3.70 (lei de grandes desvios para variáveis aleatórias binomiais). Sejam X_i variáveis aleatórias independentes Bernoulli com parâmetro $1/2$ e $1/2 < \alpha \leqslant 1$. 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 3.71. Sejam $\lambda > 0$ real e (p_n) uma sequência de reais em $[0,1]$. Prove que

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

EXERCÍCIO 3.72. 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 3.73 (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), que se $k = \alpha\log(n)/\log(\log(n))$ então $\mathbb{E}X = n^{1-\alpha+o(1)}$ e conclua que

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

3. Prove se $k = \alpha\log(n)/\log(\log(n))$ então

$$\mathbb{P}[X > 0] = \begin{cases} 1-o(1) & \text{se } 0 < \alpha < 1 \\ o(1) & \text{se } \alpha > 1 \end{cases}.$$

EXERCÍCIO 3.74. Prove que se distribuímos ao acaso n bolas em m caixas e $n < (2/e)m\log(m)$ então com alta probabilidade o maior número de bolas em uma caixa é

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

EXERCÍCIO 3.75. Para uma estrutura aleatória $\mathcal{S}_{n,p}$, a função de probabilidade $p^*(n)$ é uma **função limiar** para a propriedade $\mathcal{P}$ se $\mathcal{S}_{n,p}$ se

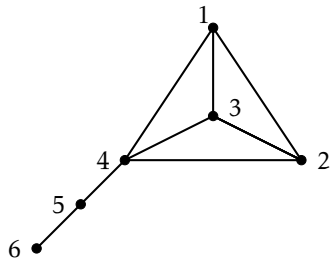
$$\mathbb{P}\left[\mathcal{S}_{n,p} \text{ tem } \mathcal{P}\right] = \begin{cases} o(1) & \text{se } p(n) = o(p^*(n)) \\ 1-o(1) & \text{se } p^*(n) = o(p(n)) \end{cases}$$

ou seja, $\mathcal{S}_{n,p}$ quase certamente não tem $\mathcal{P}$ quando $p \ll p^*$ e quase certamente tem $\mathcal{P}$ quando $p \gg p^*$. Ou ainda, a função é limiar se ocorre o inverso, ou seja, $\mathcal{S}_{n,p}$ quase certamente tem $\mathcal{P}$ quando $p \ll p^*$ e quase certamente não tem $\mathcal{P}$ quando $p \gg p^*$

$$\mathbb{P}\left[\mathcal{S}_{n,p} \text{ tem } \mathcal{P}\right] = \begin{cases} 1-o(1) & \text{se } p(n) = o(p^*(n)) \\ o(1) & \text{se } p^*(n) = o(p(n)). \end{cases}$$

A função p^* marca uma transição de fase entre não ter a propriedade e ter a propriedade.

No exemplo 3.43 a função $p^*(n) = n^{-1}$ é chamada de função limiar para propriedade "conter triângulo". Seguindo o exemplo 3.43, mostre que $n^{-2/3}$ é uma função limiar para "$\mathcal{G}_{n,p}$ contém subgrafo completo com 4 vértices". Em seguida, considere o grafo H dado pela figura abaixo. Prove que se $n^{-2/3} \gg p = p(n) \gg n^{-3/4}$ então o número esperado de cópias de H em $\mathcal{G}_{n,p}$ tende ao infinito, entretanto, o número esperado de subgrafos completos com 4 vértices em $\mathcal{G}_{n,p}$ tende a 0. Conclua que quase certamente $\mathcal{G}_{n,p}$ não contém H.



Bibliografia

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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