Esse resultado pode ser generalizado de modo que para todo $T \subset N$

$$(1-\varepsilon)\frac{|S|}{|N|}|T| < \left|\{\, u \in S \colon h(u) \in T\}\right| < (1+\varepsilon)\frac{|S|}{|N|}|T| \tag{5.23}$$

para quase toda h numa família 2-universal de funções, dado que $|S||T|$ é grande o suficiente. De fato, agora tomamos Y dada por

$$Y(\lambda) := \sum_{u \in S} \mathbb{1}_{[h(u)\in T]}(\lambda) = \left|\{\, u \in S \colon h_\lambda(u) \in T\}\right| \tag{5.24}$$

cuja média é $\mathbb{E}\, Y = |S||T|/|N| = c|T|$, portanto,

$$\mathbb{P}[\,|Y - c|T||\,] \geqslant \varepsilon c|T|\,] \leqslant \varepsilon$$

dado que $|S||T| \geqslant \varepsilon^{-3}|N|$. Daí segue (5.23) como na demonstração do lema anterior.

Agora, consideremos o subconjunto das funções $h_\lambda \in \mathcal{H}$ tais que

$$\left|\frac{Y(\lambda)}{|U|} - \frac{|S|}{|U|}\frac{|T|}{|N|}\right| \leqslant \varepsilon \tag{5.25}$$

em que Y é a variável aleatória definida em (5.24). A probabilidade com que uma escolha uniforme em $\mathcal{H}$ não satisfaça a condição da equação (5.25) é

$$\mathbb{P}\left[\left|Y - \frac{|S||T|}{|N|}\right| \geqslant \varepsilon|U|\right] \leqslant \frac{|S||T|/|N|}{\varepsilon^2|U|^2} = \frac{|S||T|/|U||N|}{\varepsilon^2|U|}$$

de acordo com a desigualdade de Chebyshev.

Podemos interpretar a equação (5.25) como a probabilidade de um escolha uniforme u estar em S e ter imagem $h(u)$ em T é, aproximadamente, a probabilidade de escolhas uniformes e independentes de uma elemento de S e um imagem dele em T

$$\left|\mathop{\mathbb{P}}_{u\in U}([u \in S] \cap [h(u) \in T]) - \mathop{\mathbb{P}}_{U}(S)\mathop{\mathbb{P}}_{N}(T)\right| \leqslant \varepsilon.$$

Nesse caso, dizemos que a função de *hash* $h \colon U \to N$ é ε-*independente* para (S, T).

Lema 5.22 (Hash mixing lemma, Nisan, 1992) *Seja $\mathcal{H} \subset N^U$ uma família 2-universal. Então,*

$$\mathop{\mathbb{P}}_{h\in\mathcal{H}}[h \text{ não é } \varepsilon\text{-independente para } (S,T)] \leqslant \frac{1}{\varepsilon^2|U|}\frac{|S|}{|U|}\frac{|T|}{|N|}$$

para quaisquer $\varepsilon > 0$, $S \subset U$ e $T \subset M$. □

5.3 DESIGUALDADES DE BERNSTEIN–CHERNOFF–HOEFFDING.

As desigualdades de Chernoff formam uma classe de desigualdades derivadas da desigualdade de Markov e de limitantes para a função geradora de momentos.

Essa técnica foi introduzida por Bernstein, 1924[7], que desenvolveu limitantes para a soma de variáveis aleatórias com variâncias finitas. Chernoff, 1952, derivou limitantes para soma de variáveis aleatórias independentes

> "Devo mencionar que Cramér (1938) havia obtido resultados muito mais elegantes e gerais sobre grandes desvios. Descobri isso depois de derivar meus próprios resultados. No entanto, Cramér exigia uma condição que não era satisfeita pelas variáveis aleatórias de valores inteiros no meu problema. Shannon havia publicado um artigo usando o Teorema Central do Limite como uma aproximação para grandes desvios e foi criticado por isso. Meu artigo permitiu que ele modificasse seus resultados e gerou muita publicidade na literatura de ciência da computação para o que ficou conhecido como o limitante de Chernoff, que, na verdade, era o resultado de Rubin." Herman Chernoff, (Chernoff, 2014).

Por fim, Hoeffding, 1994, introduziu uma desigualdade para soma a de variáveis aleatórias independentes e limitadas.

Se X é uma variável aleatória, $f\colon \mathbb{R} \to \mathbb{R}^+$ uma função crescente e $t \in \mathbb{R}$, então vale $[X \geqslant t] = [f(X) \geqslant f(t)]$ e pela desigualdade de Markov deduzimos

$$\mathbb{P}[X \geqslant t] = \mathbb{P}[f(X) \geqslant f(t)] \leqslant \frac{\mathbb{E}\, f(X)}{f(t)}$$

donde derivamos, anteriormente, a desigualdade de Chebyshev usando $f(x) = x^2$. Agora, neste contexto, tomamos a função $f(x) = \exp(sx)$ para algum parâmetro $s \in \mathbb{R}$ e temos que

- se $s > 0$, então $[X \geqslant t] = [sX \geqslant st] = [\exp(sX) \geqslant \exp(st)]$, logo

$$\mathbb{P}[X \geqslant t] = \mathbb{P}[e^{sX} \geqslant e^{st}] \leqslant \frac{\mathbb{E}\, e^{sX}}{e^{st}}, \tag{5.26}$$

- se $s < 0$, então $[X \leqslant t] = [sX \geqslant st] = [\exp(sX) \geqslant \exp(st)]$, logo

$$\mathbb{P}[X \leqslant t] = \mathbb{P}[e^{sX} \geqslant e^{st}] \leqslant \frac{\mathbb{E}\, e^{sX}}{e^{st}}. \tag{5.27}$$

Recordemos que $M_X(s) = \mathbb{E}[e^{sX}]$ é a função geadora de momentos de X, definida para todo $s \in (-x_0, x_0)$ onde X tem os seus momentos finitos. Como (5.26) e (5.27) valem para todo s em que os momentos de X são finitos, obtemos dessas desigualdades o seguinte resultado.

[7] Sergei Natanovich Bernstein, 1880–1968 foi um matemático soviético. Em 1917, sugeriu a primeira base axiomática da teoria de probabilidade. Conhecido também por resolver o décimo nono problema de Hilbert sobre a solução analítica de equações diferenciais elípticas.

Teorema 5.23 (método de Chernoff–Cramér) *Se* X *é uma variável aleatória e* $t \in \mathbb{R}$

$$\mathbb{P}[X \geqslant t] \leqslant \inf_{s>0} \frac{\mathbb{E}\, e^{sX}}{e^{st}}$$
$$\mathbb{P}[X \leqslant t] \leqslant \inf_{s<0} \frac{\mathbb{E}\, e^{sX}}{e^{st}}.$$

O método de Chernoff–Cramér consiste em estimar o lado direito dessas desigualdades e é especialmente útil no caso de variáveis aleatórias que podem ser escritas como soma de variáveis aleatórias independentes, pois nesse caso vale $M_{X+Y}(s) = M_X(s)M_Y(s)$.

Soma de variáveis aleatórias de Rademacher Uma variável aleatória tem **distribuição de Rademacher** se assume os valores $\{-1, 1\}$ com probabilidade uniforme.

Teorema 5.24 *Se* $X = \sum_{i=1}^{n} X_i$ *com* $X_1, \ldots, X_n \in_R \{-1, 1\}$ *independentes e* $t > 0$ *então*

$$\mathbb{P}[|X| \geqslant t] \leqslant 2\exp\left(\frac{-t^2}{2n}\right).$$

Demonstração. Seja X é uma variável aleatória como enunciado. Pela independência das variáveis aleatórias X_i, para todo $s \in \mathbb{R}$

$$\mathbb{E}\, e^{sX} = \prod_{i=1}^{n} \mathbb{E}\, e^{sX_i} = \prod_{i=1}^{n} \frac{e^s + e^{-s}}{2} = (\cosh(s))^n$$

$\cosh(s)$ é a função cosseno hiperbólico. Usando a série de Taylor para o cosseno hiperbólico e o fato de que, para todo inteiro $k \geqslant 0$, $(2k)! \geqslant 2^k k!$ (segue, facilmente, por indução em k)

$$\frac{1}{2}(e^s + e^{-s}) = \sum_{i=0}^{\infty} \frac{s^{2i}}{(2i)!} \leqslant \sum_{i=0}^{\infty} \left(\frac{s^2}{2}\right)^i \frac{1}{i!} = e^{s^2/2}$$

portanto, $\mathbb{P}[X \geqslant t] \leqslant \exp(s^2 n/2 - st)$ e o mínimo no lado direito é atingido para $s = t/n$, logo

$$\mathbb{P}[X \geqslant t] \leqslant e^{-t^2/2n}. \tag{5.28}$$

De $\mathbb{E}\, e^{-sX} = (\cosh(s))^n$ obtemos $\mathbb{P}[-X \geqslant t] \leqslant \exp(s^2 n/2 - st)$, portanto, pela subaditividade (equação (1.6))

$$\mathbb{P}[|X| \geqslant t] \leqslant 2e^{-t^2/2n}$$

para todo $t > 0$. □

Exemplo 5.25 (*discrepância em hipergrafos*). Dados o conjunto $V = \{1, 2, \ldots, n\}$ e uma coleção $\mathcal{S} = \{S_1, S_2, \ldots, S_m\}$ de subconjuntos de V, a discrepância é uma medida do desequilíbrio na distribuição dos elementos de V em dois subconjuntos disjuntos.

Seja $(V, \mathcal{S})$ um sistema de conjuntos (ou, hipergrafo). Dada uma coloração $\mathcal{X} : V \to \{-1, +1\}$ dos elementos em V, a discrepância em S_j é $\mathcal{X}(S_j) \coloneqq \sum_{x \in S_j} \mathcal{X}(x)$ e a discrepância da coleção $\mathcal{S}$ é

$$\operatorname{disc}(\mathcal{S}) = \min_{\mathcal{X} \in \{-1,+1\}^V} \max_{j \in \{1,\ldots,m\}} \left| \mathcal{X}(S_j) \right|$$

onde o mínimo é sobre toda coloração e o máximo sobre todo elemento de $\mathcal{S}$.

Tomando uma coloração aleatória $\mathcal{X} \in_R \{-1, +1\}$, então em cada conjunto S_j a probabilidade com que a soma $\left| \sum_{x \in S_j} \mathcal{X}(x) \right|$ exceda $2\sqrt{n \ln m}$ é, pela equação (5.28) no máximo $2\exp(-2\ln m) = 2/m^2$. Por subaditividade, a probabilidade de que todo subconjunto do sistema $\mathcal{S}$ tenha discrepância menor que $2\sqrt{n \ln m}$ é pelo menos $1 - 2/m$. Esse resultado foi melhorado por Spencer, 1985, que mostrou que cada soma pode ser limitada a $6\sqrt{n}$ e Bansal, 2010, deu uma versão construtiva desse resultado. ◊

Compreender a discrepância de vários sistemas de conjuntos tem sido um problema de pesquisa importante tanto em matemática quanto em ciência da computação (Chazelle, 2000). A discrepância tem uma gama de aplicações em vários tópicos da ciência da computação, como particionamento balanceado de dados, alocação de recursos e distribuição de carga, aprendizado de máquina e *streaming*.

Exemplo 5.26 (*balanceamento de um conjunto (*set balancing*)*). O problema de balanceamento de um conjunto pode ser formulado de forma matricial, o que facilita sua análise e implementação computacional. Vamos expressar o problema usando uma matriz binária $A \in \{0,1\}^{m \times n}$ que representa os subconjuntos $\mathcal{S} = \{S_1, S_2, \ldots, S_m\}$, um vetor de sinais $\sigma \in \{-1, +1\}^n$.

A matriz A é dada por $a_{i,j} = 1$ se, e só se, $j \in S_i$. O objetivo é minimizar o maior desbalanceamento

$$\min_{\sigma \in \{-1,+1\}^V} \|A\sigma\|_\infty.$$

Por exemplo, para os subconjuntos $S_1 = \{1, 2\}$, $S_2 = \{3, 4\}$ e $S_3 = \{1, 3, 4\}$ de $V \coloneqq \{1, 2, 3, 4\}$ temos

$$A = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{pmatrix} \quad \text{e} \quad A\sigma = \begin{pmatrix} \sigma_1 + \sigma_2 \\ \sigma_3 + \sigma_4 \\ \sigma_1 + \sigma_3 + \sigma_4 \end{pmatrix}$$

e para $\sigma = (+1, -1, +1, -1)$ temos $\|A\sigma\|_\infty = \max(|0|, |0|, |1|) = 1$, que é uma solução ótima.

Um algoritmo probabilístico para esse problema simplesmente sorteia uniformemente e independentemente as coordenadas de σ. O resultado do produto escalar

da i-ésima linha de A com σ, denotado por $(A\sigma)_i$, para $i = 1, 2, \ldots, n$, é desbalanceado se $|(A\sigma)_i| \geqslant 2\sqrt{n\ln(m)}$ o que ocorre com probabilidade

$$\mathop{\mathbb{P}}_{\sigma \in_R \{-1,+1\}^n}\left[|(A\sigma)_i| \geqslant 2\sqrt{n\ln(m)}\right] \leqslant 2\exp\left(-\frac{4n\ln(m)}{2n}\right) = \frac{2}{m^2}.$$

logo, existe uma linha desbalanceada com probabilidade menor que $2/m$, ou seja, com alta probabilidade nenhuma linha é desbalanceada. ♢

O problema de balanceamento de conjuntos é NP-difícil na maioria de suas variantes, especialmente quando há restrições adicionais nos subconjuntos ou pesos associados aos sinais. Esse problema surge em diversas aplicações, como design de experimentos e aprendizado de máquina. No aprendizado de máquina, por exemplo, ele aparece em contextos práticos de aprendizado supervisionado com classes desbalanceadas, como na detecção de fraudes em que a classe "fraude" é geralmente muito menor que a classe "não fraude". Nesse caso, o objetivo é reamostrar os dados para criar subconjuntos balanceados. Cada subconjunto pode representar as instâncias de uma classe, e os sinais determinam se as instâncias devem ser incluídas ou excluídas no treinamento. No design de redes neurais, durante o treinamento, as atualizações dos pesos podem ser interpretadas como subconjuntos balanceados. O objetivo, nesse caso, é minimizar o desbalanceamento no gradiente acumulado ao longo das camadas para evitar problemas como o "desvanecimento do gradiente". O desvanecimento do gradiente ocorre quando os gradientes usados para atualizar os pesos diminuem exponencialmente à medida que são propagados para trás no tempo, tornando difícil para a rede capturar dependências de longo prazo.

Exercício 5.27. Prove que se $X_1, \ldots, X_n$ são variáveis aleatórias independentes com distribuição de Rademacher e $a_1, \ldots, a_n \in \mathbb{R}$ então

$$\mathbb{P}\left[\sum_{i=1}^{n} a_i X_i \geqslant t\right] \leqslant \exp\left(-\frac{t^2}{2\|a\|^2}\right).$$

Exercício 5.28. Prove que se $X_1, \ldots, X_n$ são variáveis aleatórias independentes, assumem valores no intervalo $[-1, 1]$ e têm esperança 0, então $\mathbb{E}\, e^{sX_i} \leqslant \cosh(s)$ e vale (5.28) (*dica*: por convexidade, $\exp(tx) \leqslant ((1-x)\exp(-t) + (1+x)\exp(t))/2$ para todo $x \in [-1, 1]$).

Soma de variáveis aleatórias de Bernoulli uniformes Dados $X_1, \ldots, X_n \in_R \{-1, 1\}$ independentes e $t > 0$, então $Y_i := (X_i + 1)/2 \in_R \{0, 1\}$ são variáveis independentes e $Y := \sum_i Y_i \sim \text{binomial}(n, 1/2)$. Aplicando o resultado do Teorema 5.24 com $X = 2Y + n$ e $t = 2a$, para todo $a > 0$,

$$\mathbb{P}[|X| \geqslant 2a] = \mathbb{P}[|2Y - n| \geqslant 2a] = \mathbb{P}[|Y - n/2| \geqslant t] \leqslant e^{-2t^2/n}$$

para todo $t > 0$. Disso concluímos o seguinte resultado.

Teorema 5.29 (Chernoff para v.a. binomial padrão) *Se* $X \sim \text{binomial}(n, 1/2)$ *então*

$$\mathbb{P}[\,|X - \mathbb{E}\,X| \geqslant t\,] \leqslant 2\exp\left(\frac{-2t^2}{n}\right).$$

para todo $t > 0$. □

Exemplo 5.30 (*Markov, Chebyshev e Chernoff no lançamento de moedas*). Consideremos S_n a quantidade de caras no lançamento de n moedas equilibradas. O número esperado de caras é $n/2$, portanto, pela desigualdade de Markov temos

$$\mathbb{P}\left[\,S_n \geqslant \frac{3}{4}n\,\right] \leqslant \frac{n/2}{3n/4} = \frac{2}{3}.$$

Usando a desigualdade de Chebyshev

$$\mathbb{P}\left[\,\left|S_n - \frac{n}{2}\right| \geqslant \frac{n}{4}\,\right] \leqslant \frac{n/4}{(n/4)^2} = \frac{4}{n}$$

em particular, $\mathbb{P}[\,S_n \geqslant 3n/4\,] \leqslant 4/n$. Pela desigualdade de Chernoff

$$\mathbb{P}\left[\,S_n \geqslant \frac{3n}{4}\,\right] = \mathbb{P}\left[\,S_n \geqslant \frac{n}{2} + \frac{n/2}{2}\,\right] \leqslant e^{-\frac{(n/2)^2}{2n}} = e^{-\frac{n}{8}}.$$

Em resumo, os limitantes superiores para a probabilidade de $S_n \geqslant 1{,}5\,\mathbb{E}\,S_n$

	Markov	Chebyshev	Chernoff
$\mathbb{P}[\,S_n \geqslant 3n/4\,]$	$\leqslant 2/3$	$\leqslant 4/n$	$\leqslant \exp(-n/8)$

embora a desigualdade de Chernoff seja muito mais poderosa devemos lembrar que essas desigualdades assumem hipóteses diferentes, principalmente, com relação a independência quando temos soma de variáveis aleatórias. ◊

Vamos derivar uma desigualdade do tipo Chernoff para soma de variáveis aleatórias com distribuição geométrica. Sejam $X_i \sim \text{Geom}(1/2)$, para $1 \leqslant i \leqslant n$, variáveis aleatórias independentes, e $X := \sum_{i=1}^{n} X_i$, claramente, $\mathbb{E}\,X = 2n$. Para todo $s > 0$

$$\mathbb{E}\,e^{sX} = \prod_{i=1}^{n} \mathbb{E}\,e^{sX_i} = \prod_{i=1}^{n}\left(\sum_{k\geqslant 1} e^{sk}\mathbb{P}[\,X_i = k\,]\right) = \prod_{i=1}^{n}\left(\sum_{k\geqslant 1}\left(\frac{e^s}{2}\right)^k\right).$$

Assumindo $|s| < \ln 2$ a série converge

$$\sum_{k\geqslant 1}\left(\frac{e^s}{2}\right)^k = \frac{e^s}{2 - e^s}$$

portanto

$$\mathbb{E}\,e^{sX} = \left(\frac{e^s}{2 - e^s}\right)^n$$

usando a desigualdade (5.26) com $t = (2+a)n$, para qualquer $a > 0$ temos

$$\mathbb{P}[\,X \geqslant (2+a)n\,] \leqslant e^{-s(2+a)n}\left(\frac{e^s}{2-e^s}\right)^n = \left(\frac{e^{-(1+a)s}}{2-e^s}\right)^n.$$

Tomando $s = \ln(1 + a/(2+a))$ o valor no lado direito acima é mínimo

$$\mathbb{P}[\,X \geqslant (2+a)n\,] \leqslant \left(1+\frac{a}{2}\right)^n \left(1-\frac{a}{2+2a}\right)^{(1+a)n},$$

usando que $1 - x \leqslant \exp(-x)$, para todo x,

$$\left(1-\frac{a}{2+2a}\right)^{(1+a)n} \leqslant e^{-an/2}$$

e que $1 + a/2 \leqslant \exp(a/4)$ para $a \geqslant 3$

$$\left(1+\frac{a}{2}\right)^n \leqslant e^{an/4},$$

finalmente, para todo $a \geqslant 3$ vale que $\mathbb{P}[\,X \geqslant (2+a)n\,] \leqslant \exp(-an/4)$.

Teorema 5.31 (Chernoff para v.a. geométrica padrão) *Sejam* $X_i \sim \text{Geom}(1/2)$, *para* $1 \leqslant i \leqslant n$, *variáveis aleatórias independentes e* $X = \sum_{i=1}^n X_i$

$$\mathbb{P}[\,X \geqslant (1+t)\mathbb{E}\,X\,] \leqslant \exp\left(-\frac{tn}{2}\right) \tag{5.29}$$

para todo $t \geqslant 1{,}5$. □

Exemplo 5.32 (*o tamanho de uma* skip list). Na seção 2.4.2 definimos uma estrutura de dados aleatorizada, chamada *Skip List*, onde cada elemento x de um conjunto S com n elementos é replicado $h(x)$ vezes, com $h(x) \sim \text{Geom}(1/2)$, e $N = \sum_{x\in S} h(x)$ é o tamanho da *Skip List*, cuja esperança é $\mathbb{E}\,N = \sum_{x\in S}\mathbb{E}[\,h(x)\,] = 2n$. Usando a desigualdade (5.29) com $t = 1{,}5$ temos $\mathbb{P}[\,N \geqslant 5n\,] \leqslant \exp(-3n/4)$, ou seja, com probabilidade maior que $1 - \exp(-3n/4) = 1 - o(1)$ uma *skip list* que representa um conjunto com cardinalidade n usando espaço $O(n)$. ◊

Soma de variáveis aleatórias Bernoulli não identicamente distribuídas A estratégia do método dos momentos pode ser usada na soma de variáveis aleatórias que são independentes mas não têm, necessariamente, a mesma distribuição. Se X é a soma de variáveis aleatórias $X_i \sim \text{Bernoulli}(p_i)$ independentes, temos

$$\mathbb{E}\,e^{sX} = \prod_{i=1}^n \mathbb{E}\,e^{sX_i} = \prod_i (p_i e^s + (1-p_i)e^0) \leqslant \left(\frac{1}{n}\sum_i (p_i e^s + (1-p_i)e^0)\right)^n$$

pela desigualdade entre as médias aritméticas e geométricas. Pondo

$$\hat{p} := \frac{1}{n}\sum_{i=1}^n p_i \quad \text{e} \quad \hat{q} := 1 - \hat{p}$$

a desigualdade acima fica

$$\mathbb{P}[\,X \geqslant (\hat{p}+t)n\,] \leqslant \frac{\mathbb{E}\, e^{sX}}{e^{s(\hat{p}+t)n}} = \frac{\prod_i \mathbb{E}[\,e^{sX_i}\,]}{e^{s(\hat{p}+t)n}} = \left(\frac{\hat{p}e^{s}+\hat{q}}{e^{s(\hat{p}+t)}}\right)^{n}. \tag{5.30}$$

Minimizando o lado direito para $s > 0$: queremos minimizar a função de x

$$\frac{pe^{x}+(1-p)}{e^{x(p+t)}}.$$

A derivada, com relação a x é $p(1-(p+t))e^{x(1-(p+t))}-(1-p)(p+t)e^{-x(p+t)}$, cujos pontos críticos são os x tais que

$$p(1-(p+t))e^{x(1-(p+t))} = (1-p)(p+t)e^{-x(p+t)}.$$

Multiplicando ambos os lados por $e^{x(p+t)}$, obtemos

$$p(1-(p+t))e^{x} = (1-p)(p+t).$$

Resolvendo para x e tomando o logaritmo natural

$$x = \ln\left(\frac{(1-p)(p+t)}{p(1-(p+t))}\right) = \ln\left(\frac{q(p+t)}{p(q-t)}\right).$$

Lema 5.33 *Se X é a soma das variáveis aleatórias independentes $X_i \sim$ Bernoulli(p_i), $1 \leqslant i \leqslant n$, $\hat{p} = \frac{1}{n}\sum_{i=1}^{n} p_i$ e $\hat{q} := 1-\hat{p}$, então*

$$\mathbb{P}[\,X \geqslant (\hat{p}+t)n\,] \leqslant \left(\left(\frac{\hat{p}}{\hat{p}+t}\right)^{\hat{p}+t}\left(\frac{\hat{q}}{\hat{q}-t}\right)^{\hat{q}-t}\right)^{n} \qquad (0 \leqslant t < \hat{q}).$$

Demonstração. Da discussão acima, o lado direito da desigualdade (5.30) é mínimo para $s = \ln((\hat{p}+t)\hat{q}/(\hat{p}(\hat{q}-t)))$ e $0 \leqslant t < \hat{q}$. Substituindo tal valor para s em (5.30) prova o lema. □

Fazendo $s = \ln(1+t)$ e $t = \varepsilon\hat{p}$, para todo $\varepsilon > 0$, em (5.30), ao invés de tomarmos o ótimo como acima, obtemos uma desigualdade mais fraca, mas que pode ser mais fácil de aplicar

$$\mathbb{P}[\,X \geqslant (\hat{p}+t)n\,] = \mathbb{P}[\,X \geqslant (1+\varepsilon)\mathbb{E}\,X\,] \leqslant \left(\frac{1+\varepsilon\hat{p}}{(1+\varepsilon)^{(1+\varepsilon)\hat{p}}}\right)^{n} \leqslant \left(\frac{e^{\varepsilon}}{(1+\varepsilon)^{(1+\varepsilon)}}\right)^{\mathbb{E}\,X},$$

pois $n\hat{p} = \sum_i p_i = \mathbb{E}\,X$.

Exercício 5.34. Deduza que, para todo $\varepsilon \in (0,1)$ vale que

$$\mathbb{P}[\,X \leqslant (1-\varepsilon)\mathbb{E}\,X\,] \leqslant \left(\frac{e^{-\varepsilon}}{(1-\varepsilon)^{1-\varepsilon}}\right)^{\mathbb{E}\,X}.$$

Lema 5.35 *Para todo $1 \leqslant i \leqslant n$, $p_i \in [0,1]$ e $X_i \sim \text{Bernoulli}(p_i)$ são variáveis aleatórias independentes. Se $X = \sum_{i=1}^{n} X_i$, então*

$$\begin{aligned} \mathbb{P}[X \geqslant (1+\varepsilon)\mathbb{E}\,X] &\leqslant \left(\frac{e^{\varepsilon}}{(1+\varepsilon)^{1+\varepsilon}}\right)^{\mathbb{E}\,X} \qquad \text{para todo } \varepsilon > 0, \text{ e} \\ \mathbb{P}[X \leqslant (1-\varepsilon)\mathbb{E}\,X] &\leqslant \left(\frac{e^{-\varepsilon}}{(1-\varepsilon)^{1-\varepsilon}}\right)^{\mathbb{E}\,X} \qquad \text{para todo } \varepsilon \in (0,1). \end{aligned} \tag{5.31}$$

□

A seguir vamos trabalhar os limitantes dos lemas acima para obtermos desigualdade mais fáceis de serem usadas.

Reescrevendo o lado direito da desigualdade demonstrada no Lema 5.33 acima

$$\mathbb{P}[X \geqslant (\hat{p}+t)n] \leqslant \exp(-f(t)\cdot n) \tag{5.32}$$

onde

$$f(t) = (\hat{p}+t)\ln\frac{\hat{p}+t}{\hat{p}} + (\hat{q}-t)\ln\frac{\hat{q}-t}{\hat{q}}.$$

Agora, vamos estudar essa função f. A primeira e segunda derivadas de f são, respectivamente,

$$f'(t) = \ln\left(\frac{(\hat{p}+t)\hat{q}}{\hat{p}(\hat{q}-t)}\right) \quad \text{e} \quad f''(t) = \frac{1}{(\hat{p}+t)(\hat{q}-t)}$$

de modo que $f'(0) = f''(0) = 0$ e $f''(t) \geqslant 4$ para todo $0 \leqslant t \leqslant \hat{q}$, pois o denominador é o produtos de dois reais não negativos que somam 1. Pela fórmula de Taylor com resto

$$f(t) = f(a) + f'(a)(t-a) + \frac{f''(\xi)}{2!}(t-a)^2$$

para algum $\xi \in (a,t)$. Com $a = 0$ temos $f(t) = f''(\xi)t^2/2 \geqslant 2t^2$.

Substituindo na equação (5.32) obtemos $\mathbb{P}[X \geqslant (\hat{p}+t)n] \leqslant \exp(-2t^2 n)$.

Replicando essa dedução para $Y_i = 1 - X_i$ obtemos

$$\mathbb{P}[X \leqslant (\hat{p}-t)n] = \mathbb{P}\left[\sum_i Y_i \geqslant ((1-\hat{p})+t)n\right] \leqslant \exp\left(-2t^2 n\right)$$

para todo $t \geqslant 0$

Corolário 5.36 *Com as hipótese do Lema 5.33, $\mathbb{E}\,X = n\hat{p}$ e*

$$\mathbb{P}[|X - \mathbb{E}\,X| \geqslant tn] \leqslant 2\exp\left(-2t^2 n\right), \quad \text{para todo } t \geqslant 0. \qquad \square$$

A seguir, usamos a seguinte desigualdade cuja demonstração postergamos para o final dessa seção.

Proposição 5.37 *Para todo $x > 0$, $(1+x)\ln(1+x) - x \geqslant 3x^2/(6+2x)$.*

A desigualdade para o limitante superior na equação (5.31) fica, usando a proposição acima,

$$\mathbb{P}[\,X \geqslant (1+\varepsilon)\mathbb{E}\,X\,] \leqslant \exp(-((1+\varepsilon)\ln(1+\varepsilon)-\varepsilon)\mathbb{E}\,X) \leqslant \left(-\frac{\varepsilon^2\,\mathbb{E}\,X}{2(1+\varepsilon/3)}\right) \leqslant \left(-\frac{\varepsilon^2\,\mathbb{E}\,X}{3}\right).$$

Por outro lado, tomando $h(x) = f(-x\hat{p})$, então $h'(x) = -\hat{p}f'(-x\hat{p})$, $h'(0) = h(0) = 0$ e

$$h''(x) = p^2 f''(-x\hat{p}) = \frac{\hat{p}^2}{(\hat{p}-x\hat{p})(\hat{q}+x\hat{p})} = \frac{\hat{p}}{(1-x)(\hat{q}+\hat{p}x)} \geqslant \frac{\hat{p}}{(1-x)} \geqslant \hat{p}$$

portanto, $h(x) \geqslant \hat{p}x^2/2$ pela fórmula de Taylor com resto. Agora,

$$\mathbb{P}[\,X \leqslant (1-\varepsilon)\mathbb{E}\,X\,] = \mathbb{P}[\,X \leqslant (\hat{p}-\varepsilon\hat{p})n\,] \leqslant \exp(-f(-\varepsilon\hat{p})\cdot n) = \exp\left(-\frac{t^2\,\mathbb{E}\,X}{2}\right)$$

para todo $\varepsilon \in (0,1)$.

Reunindo os resultados acima, a partir do Corolário 5.36 escrevemos o teorema a seguir.

Teorema 5.38 *Se* $X_i \sim \text{Bernoulli}(p_i)$ *são independentes com* $p_i \in [0,1]$ *e* $X = \sum_{i=1}^{n} X_i$, *então*

1. *para todo* $t > 0$
$$\mathbb{P}[\,|X - \mathbb{E}\,X| > t\,] \leqslant 2\exp\left(\frac{-2t^2}{n}\right);$$
2. *para todo* $\varepsilon > 0$
$$\mathbb{P}[\,X \geqslant (1+\varepsilon)\mathbb{E}\,X\,] \leqslant \exp\left(-\frac{t^2\,\mathbb{E}\,X}{3}\right)$$
3. *para todo* $0 < \varepsilon < 1$
$$\mathbb{P}[\,X \leqslant (1-\varepsilon)\mathbb{E}\,X\,] \leqslant \exp\left(\frac{-t^2\,\mathbb{E}\,X}{2}\right)$$

□

Observação 5.39. Na prática podemos ter uma estimativa $\alpha \leqslant \mathbb{E}\,X \leqslant \beta$ para a esperança e usar as desigualdades para todo $t \in (0,1)$

$$\mathbb{P}[\,X \geqslant (1+t)\beta\,] \leqslant \exp\left(-\frac{t^2\beta}{3}\right) \quad \text{e} \quad \mathbb{P}[\,X \leqslant (1-t)\alpha\,] \leqslant \exp\left(\frac{-t^2\alpha}{3}\right).$$

Exemplo 5.40 (*tabelas de espalhamento*). Assumindo $t > 2e-1$ podemos simplificar a expressão a direita da desigualdade em (5.31) e temos

$$\mathbb{P}[\,X \geqslant (1+t)\mathbb{E}\,X\,] \leqslant \left(\frac{e^t}{(2e)^{1+t}}\right)^{\mathbb{E}\,X} < 2^{-t\,\mathbb{E}\,X}.$$

De volta ao problema do tamanho das listas ligadas numa tabela de espalhamento, se os m elementos são distribuídos uniforme e independentemente pelas n listas, então a desigualdade acima, para todo $i \in \mathsf{N}$, nos dá $\mathbb{P}[\,\ell_i \geqslant (1+t)c\,] \leqslant 2^{-tc}$, em que ℓ_i é o tamanho da lista i e $c = m/n$ o tamanho esperado. Escolhendo $t = 2(\log_2 m)/c$

$$\mathbb{P}[\,\ell_i \geqslant 2\log_2 m + c\,] \leqslant 2^{-2\log_2 m} = \frac{1}{m^2}.$$

Dessa forma, a probabilidade de haver uma lista na tabela com mais que $2\log_2 m + c$ elementos é

$$\mathbb{P}\left[\bigcup_{i\in\mathsf{N}}\left[\ell_i \geqslant 2\log_2(m) + \frac{m}{n}\right]\right] \leqslant \frac{1}{m}.$$

$\Diamond$

Demonstração da proposição 5.37. Basta mostrar que $h(x) = (6 + 8x + 2x^2)\ln(1+x) - 8x - 5x^2$ é não negativa para todo $x \geqslant 0$. A derivada é $h'(x) = 4(2+x)\ln(1+x) - 8x$, a segunda derivada é $h''(x) = 4(\ln(1+x) - x/(1+x))$ e a terceira derivada $h'''(x) = 4x/(1+x)^2$.

Para todo $x \geqslant 0$, a terceira derivada $h'''(x) \geqslant 0$, logo $h(x)$ é não decrescente e como $h''(0) = 0$, temos $h''(x) \geqslant 0$. Com isso, $h(x)$ é não decrescente e como $h'(0) = 0$, temos $h'(x) \geqslant 0$. Portanto $h(x)$ é não decrescente e como $h(0) = 0$, concluímos que $h(x) \geqslant 0$ para todo $x \geqslant 0$. □

Soma de variáveis aleatórias limitadas Esse método do momento admite ainda outra generalização, agora para soma de variáveis limitadas. Esse caso é conhecido como extensão de Hoeffding, as desigualdades são chamadas de Chernoff–Hoeffding. e nele as variáveis nem precisam ser discretas.

Vejamos como fica a desigualdade (5.30) para variáveis aleatórias independentes e limitadas, $0 \leqslant \mathsf{X}_i \leqslant 1$, com média $\mathbb{E}\,\mathsf{X}_i = p_i$. Pela convexidade da função exponencial $\exp(sx) \leqslant (e^s - 1)x + 1$, para todo $x \in (0,1)$, logo

$$\mathbb{E}[\,e^{s\mathsf{X}}\,] = \prod_i \mathbb{E}[\,e^{s\mathsf{X}_i}\,] \leqslant \prod_i \mathbb{E}[\,(e^s - 1)\mathsf{X}_i + 1\,] \leqslant \prod_i (p_i e^s + (1 - p_i)) \leqslant \left(\frac{pe^s + q}{e^{s(p+t)n}}\right)^n$$

e estamos de volta em (5.30). Para finalizar essa seção, vamos mostrar o seguinte resultado.

Teorema 5.41 (desigualdade de Chernoff–Hoeffding, 1963) *Para $1 \leqslant i \leqslant n$, sejam X_i variáveis aleatórias independentes com $0 \leqslant \mathsf{X}_i \leqslant 1$. então*

1. *para todo $t > 0$*

$$\mathbb{P}[\,|\mathsf{X} - \mathbb{E}\,\mathsf{X}| > t\,] \leqslant 2\exp\left(\frac{-2t^2}{n}\right);$$

2. *para todo* $\varepsilon > 0$

$$\mathbb{P}[X \geqslant (1+\varepsilon)\mathbb{E}\,X] \leqslant \exp\left(-\frac{t^2\,\mathbb{E}\,X}{3}\right)$$

3. *para todo* $0 < \varepsilon < 1$

$$\mathbb{P}[X \leqslant (1-\varepsilon)\mathbb{E}\,X] \leqslant \exp\left(\frac{-t^2\,\mathbb{E}\,X}{2}\right)$$

□

A desigualdade de Hoeffding, 1994 permite que apliquemos a desigualdade de Chernoff para qualquer soma variáveis aleatórias limitadas.

Teorema 5.42 (desigualdade de Hoeffding) *Sejam* $X_1, \ldots, X_n$ *variáveis aleatórias independentes com* $a_i \leqslant X_i \leqslant b_i$ *para* a_i, b_i *constantes, para todo* i. *Se* $X = \sum_i X_i$ *então para todo* $t > 0$

$$\mathbb{P}[|X - \mathbb{E}\,X| \geqslant t] \leqslant 2\exp\left(\frac{-2t^2}{\sum_i (b_i - a_i)^2}\right).$$

A demonstração segue do seguinte resultado: *Para toda variável aleatória* X *com* $\mathbb{E}\,X = 0$ *e* $a \leqslant X \leqslant b$, *para* a *e* b *constantes,*

$$\mathbb{E}\ e^{sX} \leqslant \exp\left(\frac{s^2(b-a)^2}{8}\right)$$

para todo $s > 0$.

5.3.1 ALGUNS EXEMPLOS REVISITADOS

Vamos retomar o caso de distribuir m bolas em n caixas ao acaso. O número de bolas na caixa i é $X_i = Y_1 + Y_2 + \cdots + Y_m$ onde Y_j indica se a bola j foi alocada na caixa i. Evidentemente, $\mathbb{E}\,X_i = m/n$.

Quando $m < 2n\log_2 n$, o Exercício 2.79 pede para mostrar que a carga máxima é $4\ln(n)/\ln\left(\frac{2n}{me}\ln(n)\right)$ com alta probabilidade.

No caso $m \geqslant 2n\log_2 n$, usando que para $\varepsilon \geqslant e - 1$ vale

$$\frac{e^\varepsilon}{(1+\varepsilon)^{1+\varepsilon}} = \frac{e^\varepsilon}{(1+\varepsilon)^\varepsilon}\frac{1}{1+\varepsilon} \leqslant \frac{1}{1+\varepsilon} \leqslant \frac{1}{2}$$

portanto, usando (5.31)

$$\mathbb{P}\left[X_i > e\frac{m}{n}\right] < 2^{-\frac{m}{n}} \leqslant \frac{1}{n^2}$$

portanto, por subaditividade,

$$\mathbb{P}\left[\text{maior carga} > e\frac{m}{n}\right] \leqslant \sum_{k=1}^{n} \mathbb{P}\left[X_i > e\frac{m}{n}\right] \leqslant \frac{1}{n}$$

ou, com probabilidade pelo menos $1 - 1/n$ toda caixa tem no máximo em/n bolas.

Ordenação com *Quicksort* em $O(\log n)$ com alta probabilidade Vejamos outra dedução de que o *quicksort* probabilístico (seção 2.2.2) faz, em média, $O(n \log n)$ comparações entre elementos de uma instância S com n elementos com alta probabilidade. Como sempre, fixamos um elemento $x \in S$ e olhamos para as subsequências $S_0 = S, S_1, S_2, \ldots, S_M$ da instância S a que x pertence após cada particionamento durante uma execução.

Sorteando um pivô, o i-ésimo particionamento é *bom* se $|S_i|/10 \leqslant |S_{i+1}| \leqslant 9|S_i|/10$ o que ocorre com probabilidade $0,75 < p \leqslant 0,8$ dada na equação (2.18). De $|S_{i+1}| \leqslant 9|S_i|/10$, obtemos que $|S_M| \leqslant (9/10)^M n$, portanto, x passa por no máximo $\log_{10/9} n < 10 \ln(n)$ boas partições.

Seja $X = X_1 + \cdots + X_M$ a quantidade de partições *ruins* pelas quais x passa, com $X_i \in \{0,1\}$ variável aleatória indicadora de partição ruim, cuja esperança é $\mathbb{E}\, X \geqslant M/5 > 4 \ln(n)$, portanto

$$\mathbb{P}[\, X > 10 \ln(n)\,] = \mathbb{P}[\, X > \mathbb{E}\, X + 6 \ln(n)\,] \leqslant \exp\left(-\frac{72 \ln^2(n)}{20 \ln(n)}\right) = \frac{1}{n^{3,6}}$$

de modo que a probabilidade de existir $x \in S$ que passa por mais que $20 \ln(n)$ partições é $n^{-2,6}$.

Busca em *Skip List* em $O(\log n)$ com alta probabilidade Seja S um conjunto com n elementos representado por uma *skip list* de níveis $S_0 \supseteq S_1 \supseteq \cdots \supseteq S_H$ e tamanho $N = \sum_{i=1}^{H} |S_i|$. Provamos, na seção 2.4.2, que $H = O(\log n)$ com alta probabilidade e que $N = O(n)$ com alta probabilidade, no Exemplo 5.32. Também vimos na seção 2.4.2 que uma busca tem tempo médio $O(\log n)$. Agora vamos provar que uma busca realiza $O(\log n)$ passos com alta probabilidade.

Como antes, consideremos o percurso reverso de uma busca: a partir da posição final de uma busca em S_0, se possível suba de nível (sucesso), senão vá para o nó antecessor no mesmo nível (fracasso).

Seja X número de passos numa busca por $x \in S$. Vamos limitar a probabilidade do evento $[X > 14 \log_2 n]$ para uma constante > 0. Fixamos $h := 3\lfloor \log_2(n) \rfloor$ e pela Lei de Probabilidade Total

$$\begin{aligned}
\mathbb{P}[\, X > 14 \log_2 n\,] &= \mathbb{P}[\, X > 14 \log_2 n \mid H \leqslant h\,]\mathbb{P}[H \leqslant h] + \mathbb{P}[\, X > 14 \log_2 n \mid H > h\,]\mathbb{P}[H > h] \\
&\leqslant \mathbb{P}[\, X > 14 \log_2 n \mid H \leqslant h\,] + \mathbb{P}[H > h] \\
&\leqslant \mathbb{P}[\vec{X} > 11 \log_2 n\,] + n\frac{2}{n^3}
\end{aligned}$$

em que $\vec{X}$ é a quantidade de passos "$\rightarrow$" (horizontais) da busca na estrutura. Em $14 \log_2 n$ passos $\mathbb{E}\vec{X} = 7 \log_2 n$ e

$$\mathbb{P}[\vec{X} > 11 \log_2 n\,] = \mathbb{P}[\vec{X} > 7 \log_2 n + 4 \log_2(n)\,] \leqslant \exp\left(-2\frac{(4 \log_2 n)^2}{14 \log_2 n}\right) = \exp(-2,2 \log_2 n)$$

logo

$$\mathbb{P}[\,X > 14\log_2 n\,] \leqslant \mathbb{P}[\,\vec{X} > 11\log_2 n\,] + n\frac{2}{n^3} \leqslant \frac{1}{n^{2,2}} + \frac{2}{n^2}$$

e a probabilidade de existir um $x \in S$ para o qual uma busca leva mais que $14\log_2(n)$ passos é $< 3/n$.

Teorema 5.43 *Uma busca numa* Skip List *que representa um conjunto com n elementos termina com $O(\log n)$ comparações com probabilidade $1 - o(1)$.* □

5.3.2 LEMA DE JOHNSON–LINDSTRAUSS

5.3.3 *SAMPLING* E PROBLEMAS DE CONTAGEM

Estimação para #DNF

Estimação para o grau médio de um grafo

5.3.4 *TREAPS*

Como no caso de *skip lists*, denotamos por U o conjunto universo e $S \subset U$ deve ser representado de modo a permitir as operações de dicionário: inserção, busca e remoção. Uma **treap** (Vuillemin, 1980) é árvore binária onde a cada nó v está associado $\mathit{chave}(v)$ e $\mathit{prioridade}(v)$ de modo a árvore se comporta como árvore binária de busca com relação a chave e uma *heap* com relação a $\mathit{prioridade}$. Um nó da árvore tem a propriedade *heap* se a sua prioridade é maior que de seus descendentes e tem a propriedade de árvore de busca se sua chave é maior que a chave do filho esquerdo e menor que a chave do filho direito.

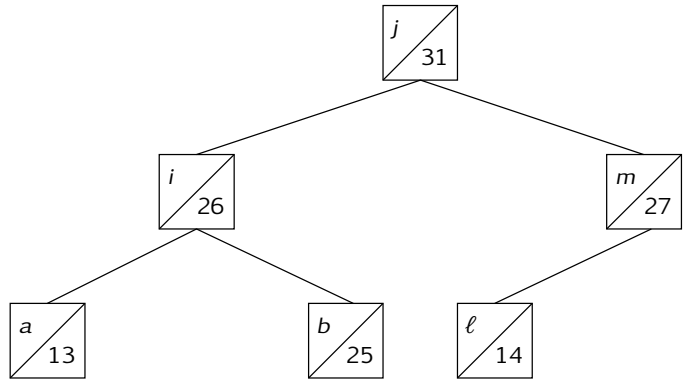


Figura 5.1: Exemplo de *treap* com 6 elementos. As letras são as chaves e os números as prioridades.

Seja S um conjunto de elementos de um universo U representados por uma *treap*, as operações de dicionário são descritas a seguir.

busca — dado $x \in U$ queremos saber se $x \in S$. A busca é feita como em árvore binária de busca, a partir da raiz comparamos x com a chave do nó, se x for maior então continuamos na subárvore esquerda, senão continuamos na subárvore direita; o custo dessa operação é proporcional a altura da árvore;

inserção — dado $x \in U$ queremos inserir x em S. Fazemos uma busca por x em S e caso falhe inserimos x na folha resultante da busca na árvore com alguma escolha para a prioridade de x. Essa operação mantém a propriedade de árvore de busca binária mas destrói a propriedade de *heap*, que pode ser refeita por uma série de rotações na árvore. Por exemplo, na Figura 5.2 abaixo, se x é filho de y mas tem maior prioridade então uma rotação a direita em y refaz a propriedade de *heap* entre esses nós;

remoção — dado $x \in S$ queremos remover x de S. Após uma busca em S rotacionamos a árvore no filho menos prioritário de x sucessivamente até que x seja uma folha e possa ser removido.

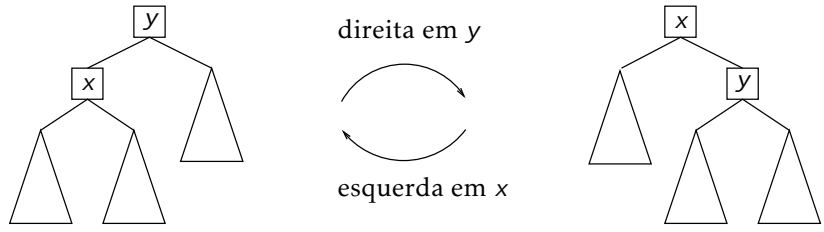


Figura 5.2: Rotações a esquerda e a direita em um nó numa árvore binária.

Essas operações têm custo proporcional a altura da árvore e usamos aleatoriedade para garantir altura logarítmica: vamos assumir que a prioridade atribuída a v quando ele é inserido em S é uma escolha aleatória no intervalo $[0, 1]$. A prioridade de um nó não precisa ser um número real, basta garantir que as prioridades sejam distintas.

Seja x_k o k-ésimo menor elemento do conjunto S que está representado por uma *treap*. Vamos estimar a que distância esse nó está da raiz. Seja $\mathbb{1}_{[x_j \prec x_k]}$ variável aleatória indicadora de que x_j é ancestral próprio de x_k na *treap*, fato que denotamos por $x_j \prec x_k$. Então a profundidade de x_k é o seu número de ancestrais próprios e

$$\mathbb{E}\,\text{profundidade}(x_k) = \sum_{j=1}^{n} \mathbb{P}[\,\mathbb{1}_{[x_j \prec x_k]} = 1\,]$$

e precisamos estimar a probabilidade de $x_j \prec x_k$. Para $j < k$ defina $\mathsf{X}(j,k) = \{x_j, x_{j+1}, \ldots, x_k\}$ e $\mathsf{X}(k,j) = \mathsf{X}(j,k)$.

Proposição 5.44 *Seja $j \neq k$. Então, x_j é um ancestral de x_k se e somente se x_j tem prioridade* $\min \mathsf{X}(j,k)$.

Demonstração. Suponha $j \leqslant k$ e que $\mathit{prioridade}(x_j) > \mathit{prioridade}(x_i)$ para todo $j < i \leqslant k$. Pela propriedade *heap* x_j não pode estar numa subárvore de x_k, mais ainda, não pode haver um ancestral x_i de x_k com x_j e x_k em subárvores de x_i distintas (justifique) logo x_j deve ser um ancestral de x_k se $\mathit{prioridade}(x_j) > \mathit{prioridade}(x_i)$ para todo $j < i \leqslant k$. Por outro lado, tomemos um ancestral x_j de x_k e suponhamos que existe existe algum x_i com $j < i \leqslant k$ com prioridade maior do que x_j. Pela propriedade *heap*, x_i não pode estar numa subárvore enraizada em x_j. Mas x_i não pode ser ancestral de x_j: teríamos $x_j < x_i \leqslant x_k$, assim x_j seria armazenado na subárvore esquerda de x_i e x_k não poderia estar armazenado naquela subárvore, contradizendo que x_j é um ancestral de x_k. Resta o caso que existe um ancestral x_ℓ de x_j que tem x_j em uma de suas subárvores e x_i na outra. Mas junto com $x_j < x_i \leqslant x_k$ novamente contradiz que x_j um ancestral de x_k, pois x_k estaria na mesma subárvore de x_ℓ que x_i. Portanto, se x_j é um ancestral de x_k então $\mathit{prioridade}(x_j) > \mathit{prioridade}(x_i)$ para todo $j < i \leqslant k$. □

Agora, qualquer elemento de $\mathsf{X}(j,k)$, $j \neq k$, tem a mesma probabilidade de ser o elemento de maior prioridade, logo

$$\mathbb{P}[\,\mathbb{1}_{[x_j \prec x_k]} = 1\,] = \frac{1}{|k-j|+1}$$

e

$$\begin{aligned}
\mathbb{E}\ \text{profundidade}(x_k) &= \sum_{j=1}^{k} \frac{1}{k-j+1} + \sum_{j=k}^{n} \frac{1}{j-k+1} - 1 \\
&= \sum_{i=1}^{k} \frac{1}{i} + \sum_{1}^{n-k+1} \frac{1}{i} - 1 \\
&= \mathsf{H}_k + \mathsf{H}_{n-k+1} - 1
\end{aligned}$$

onde H_n denota o n-ésimo número harmônico, $\mathsf{H}_n = \sum_{i=1}^{n} 1/i = \ln n + \gamma + \Theta(n^{-1})$, onde $\gamma \approx 0{,}577$ é a constante de Euler–Mascheroni (veja Graham, Knuth e Patashnik, 1994). Logo, a profundidade de qualquer nó tem valor esperado

$$\mathbb{E}\ \text{profundidade}(x_k) = \mathsf{O}(\log n).$$

Como aplicação da desigualdade de Chernoff obtemos o seguinte limitante para a altura de uma *treap* e, consequentemente, um limitante para as operações de dicionário na *treap*.

Teorema 5.45 *A altura de uma* treap *com n elementos é $O(\log n)$ com probabilidade $1 - O(n^{-5})$.*

Demonstração. Podemos provar que a profundidade de todo nó é proporcional a $\log n$ observando o fato de que $\mathbb{1}_{[x_j<x_k]}$ é independente de $\mathbb{1}_{[x_i<x_k]}$ se $i \neq j$ e podemos aplicar (5.31) com $t = e^2 - 1 > 6$ e obtermos de $H_k + H_{n-k+1} - 1 \geqslant H_n$ que

$$\mathbb{P}[\,\text{profundidade}(x_k) > (1+t)(H_k + H_{n-k+1} - 1)\,] \leqslant \left(\frac{e^t}{(1+t)^{1+t}}\right)^{H_n} \leqslant \left(\frac{1}{e}\right)^{tH_n} \leqslant \left(\frac{1}{n}\right)^6.$$

Agora, a probabilidade de existir um nó com profundidade grande, maior que $1+t$ vezes o valor esperado, é limitada por

$$\begin{aligned}\mathbb{P}[\bigcup_k [\text{profundidade}(x_k) > (1+t)(H_k + H_{n-k+1} - 1)]\,] \leqslant \\ \sum_{k=1}^{n} \mathbb{P}[\,\text{profundidade}(x_k) > (1+t)(H_k + H_{n-k+1} - 1)\,] \quad \leqslant \left(\tfrac{1}{n}\right)^5\end{aligned}$$

portanto, não existe tal nó com probabilidade pelo menos $1 - n^{-5}$. □

5.4 MARTINGAIS

5.5 EXERCÍCIOS

Exercício 5.46. Seja X uma variável aleatória quadrado integrável. Calcule a esperança e a variância da variável aleatória $Y := \frac{X - \mathbb{E}\,X}{\sqrt{\text{Var}[\,X\,]}}$.

Exercício 5.47. Prove que $\text{Var}[\,X\,] = 0$ então $\mathbb{P}[\,X = \mathbb{E}\,X\,] = 1$.

Exercício 5.48 (propriedades da covariância). Sejam X e Y variáveis aleatórias quadrado integráveis, $a, b \in \mathbb{R}$. Verifique que valem

1. $\text{Cov}(X, X) = \text{Var}[\,X\,]$;
2. $\text{Cov}(X, Y + a) = \text{Cov}(X, Y)$;
3. $\text{Cov}(X, Y) = \text{Cov}(Y, X)$;
4. $\text{Cov}(X, a \cdot Y + b \cdot Z) = a \cdot \text{Cov}(X, Y) + b \cdot \text{Cov}(X, Z)$;
5. $|\text{Cov}(X, Y)| \leqslant \sqrt{\text{Var}[\,X\,]}\sqrt{\text{Var}[\,Y\,]}$.

Exercício 5.49. Prove que se $X_1, \ldots, X_n$ são variáveis aleatórias independentes com distribuição de Rademacher, $t \in (0,1)$ e $s > 0$, então

$$\mathbb{P}\left[\frac{1}{n}\sum_{i=1}^{n} X_i \geqslant t\right] \leqslant \exp\left(-\frac{n}{2}((1+t)\ln(1+t) + (1-t)\ln(1-t))\right).$$

(*dica*: minimize $f(s) = \cosh(s)\exp -st$ usando as técnicas de Cálculo Diferencial.)

Exercício 5.50. Considere a família $\mathcal{H}$ de funções de *hashing* do Exemplo 5.20 e fixe $S \subset \mathbb{F}^n$. Prove que se $|S| \leqslant \sqrt{|\mathbb{F}|}$ então para uma escolha aleatória de $h \in \mathcal{H}$ a probabilidade de não ocorrer colisão para os elementos de S é pelo menos $1/2$.

Exercício 5.51. Suponha que $\mathcal{H}$ seja uma família 2-universal de funções $\{0,1\}^n \to \{0,1\}^n$. Para $m < n$ podemos definir uma família $\mathcal{H}'$ de funções $\{0,1\}^n \to \{0,1\}^m$ fazendo para cada $h \in \mathcal{H}$

$$h'(x) := h(x) \restriction_m$$

em que $h(x) \restriction_m$ denota a restrição de $h(x)$ aos primeiros m bits. $\mathcal{H}'$ é uma família 2-universal?

Exemplo 5.52. A família de funções $h\colon \{0,1\}^n \to \{0,1\}^m$ indexada pelas matriz binárias $M \in \{0,1\}^{(n+1)\times m}$ definida para todo $x \in \{0,1\}^n$ por

$$h_M(x) := (x,1) \cdot M$$

Mostre que essa família é 2-universal.

Exercício 5.53. Uma família de funções $\mathcal{H} = \{h_\lambda \in N^U : \lambda \in \Lambda\}$ tais que para quaisquer $x \neq y$

$$\mathop{\mathbb{P}}_{\lambda\in\Lambda}[h_\lambda(x) = h_\lambda(y)] \leqslant \frac{1}{|N|}.$$

é chamada de *hash fracamente 2-universal*. Verifique que a família de funções de Carter e Wegman, 1979, dada no Exemplo 2.18, página 113 não é 2-universal, mas é fracamente 2-universal.

*Exercício 5.54 (*desaleatorização de corte grande*).* Seja $G = (V,E)$ uma grafo sobre $V = \{0,2,\ldots,n-1\}$. Uma função $h\colon V \to \{0,1\}$ determina o subconjunto $A \subset V$ dado por $A = \{i \in V : h(i) = 1\}$ o qual, por sua vez, determina o corte $\nabla(A)$. Prove que se se sorteamos h em uma família de funções de *hash* fracamente 2-universal então $\mathbb{E}\,|\nabla(A)| \geqslant |E|/2$. Escreva e analise um algoritmo aleatorizado baseado nessa ideia.

Escreva e analise um algoritmo (determinístico) que *desaleatoriza* o algoritmo que você escreveu usando a família fracamente 2-universal do Exemplo 2.18, página 113.

Exercício 5.55. Prove que $L \in \mathsf{BPP}$ se, e só se, L é decidida por algoritmo probabilístico de tempo polinomial com probabilidade de erro $(1/2) - (1/p(n))$ nas entradas de tamanho n, para algum polinômio p (*dica:* simule um algoritmo que garante pertinência em BPP uma quantidade suficientemente grande de vezes e use a desigualdade de Chernoff para estimar a probabilidade de erro).

Exercício 5.56 (árvore binária de busca). Uma árvore binária de busca é uma estrutura de dados para representa um conjunto S munido de uma ordem total. Vamos assumir aqui que $S \subset \mathbb{Z}$. Uma árvore binária de busca é uma árvore binária com raiz e tal que a cada nó está associado um inteiro de S de modo que se $v \in S$ está em um nó então todos os nós da subárvore esquerda estão associados a elementos de S menores que v e todos os nós da subárvore direita estão associados a elementos de S maiores que v. A Figura 5.3 mostra uma árvore binária de busca para $S = \{3, 8, 9, 11, 13, 17, 19\}$. Uma busca por x numa árvore binária de busca começa

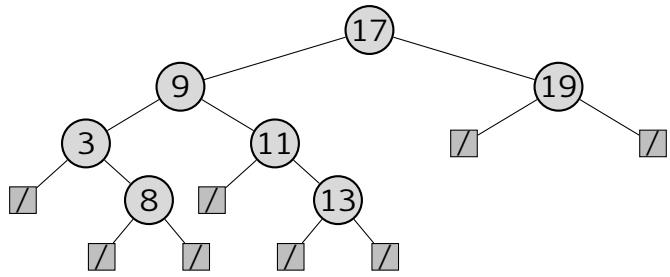


Figura 5.3: exemplo de uma árvore binária de busca.

examinando o nó raiz. Se o valor x é igual ao da raiz, a busca termina. Se o valor x é menor do que o da raiz então a busca segue pela subárvore esquerda e se o valor x é maior do que o da raiz, a busca segue pela subárvore direita. Esse processo é repetido até o valor ser encontrado ou a subárvore ser vazia. O tempo de busca é, no pior caso, proporcional à altura da árvore, isto é, o maior número de arestas percorridas num caminho da raiz até alguma folha. No exemplo da Figura 5.3 a altura é 3.

A inserção de x numa árvore binária de busca começa com uma busca até chegar numa subárvore vazia, é nesse local que o elemento é inserido. Por exemplo, a inserção de 2, 15 e 20 na árvore mostrada na Figura 5.3 resulta na árvore da Figura 5.4 abaixo.

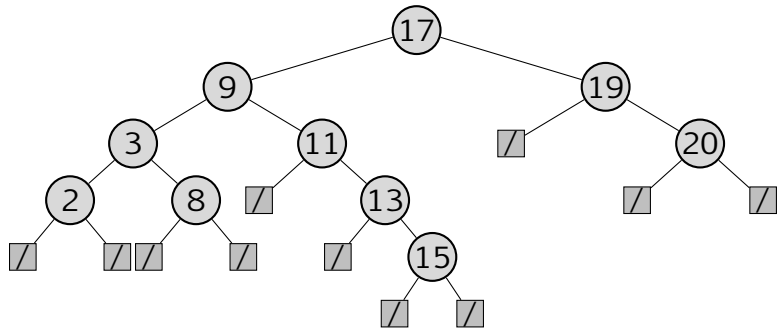


Figura 5.4: árvore de busca binária obtida a partir do exemplo acima após a inserção de 2, 15 e 20.

A árvore da Figura 5.4 pode ser obtida a partir da árvore vazia e inserindo-se os elementos de S, um a um, na seguinte ordem: $17, 9, 3, 2, 8, 11, 13, 15, 19, 20$.

1. Qual seria a árvore resultante se os elementos fossem inseridos na ordem: $2, 3, 8, 9, 11, 13, 15, 17, 19, 20$?

2. Seja $S \subset \mathbb{Z}$ finito, não vazio e de cardinalidade n. Prove que se uma árvore de busca binária é construída tomando-se uma permutação aleatória dos elementos de S e inserindo-se os elementos na ordem definida pela permutação, então a altura da árvore obtida é $O(\log n)$ com probabilidade $1 - n^{-c}$, para alguma constante $c > 0$. Conclua que com alta probabilidade a árvore é construída em tempo $O(n \log n)$.

Exercício 5.57. Dizemos que X é uma (ε, δ)-aproximação para a quantidade V se $\mathbb{P}[\,|X - V| \leqslant \varepsilon V\,] \geqslant 1 - \delta$. Sejam $X_1, \ldots, X_m$ variáveis aleatórias independentes e com mesma distribuição sobre $\{0, 1\}$ e $\mathbb{E}\, X_i = \mu$ para todo i. Denotemos por $\overline{X}_m$ a média amostral, $\overline{X}_m = (1/m)\sum_{i=1}^{m} X_i$. Prove que se $m \geqslant 3\ln(2/\delta)/(\varepsilon^2 \mu)$ então

$$\mathbb{P}[\,\left|\overline{X}_m - \mu\right| \geqslant \varepsilon\mu\,] \leqslant \delta$$

e disso segue que a média amostral é uma (ε, δ)-aproximação para μ.

Exercício 5.58. (Karp e Luby, 1983) Suponha que sejam dados conjuntos $S_1, S_2, \ldots, S_n$ com cardinalidades conhecidas. O algoritmo a seguir estima a cardinalidade da união desses conjuntos. Defina $\mathrm{cov}(x) := |\{i : x \in S_i\}|$ e $s := \sum_{i=1}^{n} |S_i|$.

Instância: conjuntos $S_1, S_2, \ldots, S_n$.
Resposta: uma estimativa para $S_1 \cup S_2 \cup \cdots \cup S_n$.

escolha $S \in_{\mathcal{D}} \{S_1, S_2, \ldots, S_n\}$ de acordo com lei $\mathcal{D}(S_i) = |S_i|/s$
$x \leftarrow_{\mathrm{R}} S$
compute $\mathrm{cov}(x)$
$X \leftarrow \frac{s}{\mathrm{cov}(x)}$
responda X.

Algoritmo 22: Algoritmo de Karp–Luby

Prove que $\mathbb{E}\, X = \left|\bigcup_{i=1}^{n} S_i\right|$ e que $\mathrm{Var}[X] \leqslant \mathbb{E}[X]^2$.

Exercício 5.59. Sejam $X_1, X_2, \ldots, X_m$ as respostas de m rodadas independentes do Algoritmo 22 com a mesma entrada e $\overline{X}_m$ a média amostral dessas rodadas. Prove que se $m \geqslant (9/2)(n-1)(1/\varepsilon^2)\ln(2/\delta)$ então temos uma (ε, δ)-aproximação para $\left|\bigcup_{i=1}^{n} S_i\right|$.

Exercício 5.60. Uma fórmula DNF é uma fórmula booleana com n variáveis e m cláusulas disjuntivas $C_1, \ldots, C_m$; uma cláusula é uma disjunção se for um "e" de

literais, a fórmula é um "ou" dessas cláusulas. O problema #DNF, ou o problema de contagem DNF, toma como entrada uma fórmula DNF e retorna o número de valorações que satisfazem a fórmula.

Escreva um algoritmo que, para dados $\varepsilon > 0$, $\delta > 0$ e φ fórmula DNF, obtenha uma (ε, δ)-aproximação para estimar o número de valorações que satisfazem uma fórmula DNF em $O(m\varepsilon^{-2}\ln(1/\delta))$ rodadas e tempo de execução total $O(nm^2\varepsilon^{-2})\ln(1/\delta))$.

Exercício 5.61. Seja $G = (V, E)$ um grafo sobre $V = \{1, 2, \ldots, n\}$. Assuma que G tenha as seguintes propriedades: (1) todo vértice tem grau d; (2) para todo $W \subseteq V$ com $|W| \leqslant n/2$ vale que $|N(W)| \geqslant c|W|$, onde $N(W) = \{x \in V : \{x, w\} \in E \text{ para algum } w \in W\}$. Um *passeio aleatório* em G é uma sequência $(X_i : i \in \mathbb{N})$ de variáveis aleatórias com $X_0 \in_R V$ e $X_{i+1} \in_R N(X_i)$.

Prove que se $B \subset V$ é tal que $|B| = \beta n$ para alguma constante $\beta > 0$, então para todo $\delta > 0$

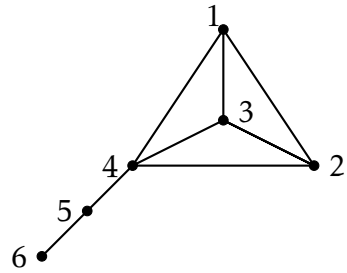
$$\mathbb{P}\left[\left|\frac{1}{k}\sum_{i=1}^{k} \mathbb{1}_{[X_i \in B]} - \beta\right| > \delta\right] < 2\exp\left(-\frac{(1-c)\delta^2 k}{4}\right).$$

Exercício 5.62. Para uma estrutura aleatória $\mathcal{S}_{n,p}$, a função de probabilidade $p^*(n)$ é uma **função limiar** para a propriedade $\mathcal{P}$ se

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

ou seja, $\mathcal{S}_{n,p}$ quase certamente não tem $\mathcal{P}$ quando $p \ll p^*$ e quase certamente tem $\mathcal{P}$ quando $p \gg p^*$. A função p^* marca uma transição de fase entre não ter a propriedade e ter a propriedade. Na seção 5.1.3 mostramos que a função $p^*(n) = n^{-1}$ é função limiar para propriedade "conter triângulo".

Seguindo o exemplo de triângulos, 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 $p = p(n)$ é tal que $n^{-2/3} \gg p \gg n^{-3/4}$ 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.



Exercício 5.63 (Raab e Steger, 1998). Esse exercício continua o desenvolvimento feito no Exercício 2.80, página 155. Prove que, nas condições daquele enunciado, se $k = \alpha \log(n)/\log(\log(n))$ então

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