5 | Leis de desvio e de Concentração

5.1 Desigualdade de Markov 159
5.1.1 Satisfazibilidade de fórmula booleana aleatória 160
5.1.2 Métodos probabilísticos de 2º momento 162
5.1.3 Triângulos em grafo aleatório 165
5.1.4 Distribuição de bolas em caixas 166
5.2 Momentos, Variância e a desigualdade de Chebyshev 170
5.2.1 Desigualdade de Chebyshev 173
5.2.2 *Hashing* 2-universal 177
5.3 Desigualdades de Bernstein–Chernoff–Hoeffding. 182
5.3.1 *Skip list* revisitada 187
5.3.2 *Treaps* 189
5.4 Martingais 192
5.5 Exercícios 192

Neste capítulo temos ferramentas úteis para estudar a distribuição dos valores que uma variável aleatória assume, os desvios das variáveis em relação a valores típicos. Por exemplo, se X é uma variável aleatória real e integrável e $f\colon \mathbb{R} \to \mathbb{R}$ uma função que assume valores não negativos com $f(x) \geqslant t > 0$ sempre que $x \in S$, então

$$\mathbb{E}\, f(X) \geqslant t\,\mathbb{P}[\,X \in S\,] + 0\,\mathbb{P}[\,X \notin S\,]$$

logo $\mathbb{P}[\,X \in S\,] \leqslant \mathbb{E}\, f(X)/t$. As leis de concentração tratam os desvios dos valores de uma variável aleatória ao redor do valor esperado e as leis de grandes desvios trata de desvios exponencialmente raros, fornecendo estimativas precisas sobre a probabilidade de concentração ao redor do valor esperado. Os primeiros resultados desse tipo, dos idos de 1950, tiveram interesse renovado e novas técnicas foram desenvolvidas, impulsionadas por aplicações em algoritmos e otimização.

5.1 DESIGUALDADE DE MARKOV

A desigualdade de Markov[1] tem um enunciado simples e a vantagem de valer em situações bem gerais, sem muitas hipóteses sobre a variável aleatória, o que a torna amplamente aplicável. Por outro lado, ela pode fornecer limitantes fracos que, as vezes, são pouco úteis.

Teorema 5.1 (desigualdade de Markov) *Se X é uma variável aleatória integrável*

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

para todo $t > 0$. Em particular, se X assume valores não negativos,

$$\mathbb{P}[\,X \geqslant t\,] \leqslant \frac{\mathbb{E}\,X}{t} \tag{5.1}$$

para todo $t > 0$.

Demonstração. De X integrável $|X|$ também é integrável, logo a desigualdade segue de $\mathbb{E}[\,|X|\,] \geqslant \mathbb{E}[\,|X|\mathbb{1}_{|X|\geqslant t}\,] \geqslant \mathbb{E}[\,t\mathbb{1}_{|X|\geqslant t}\,] \geqslant t\,\mathbb{E}[\,\mathbb{1}_{[|X|\geqslant t]}\,] = t\mathbb{P}[\,|X| \geqslant t\,]$. □

Há casos onde essa desigualdade é justa, como mostra o seguinte exemplo. Seja X uma variável aleatória que assume o valor 0 com probabilidade 99/100 e assume o valor 10 com probabilidade 1/100. Então $\mathbb{E}\,X = 1/10$ e

$$\mathbb{P}[\,X \geqslant 10\,] \leqslant \frac{1/10}{10} = \frac{1}{100}$$

que é exatamente a probabilidade do evento $[X \geqslant 10]$.

Como exemplo de aplicação dessa desigualdade, vamos ver uma prova alternativa de que em uma execução do *quicksort* (seção 2.2.2) o número esperado de comparações entre elementos de uma instância com n elementos é $O(n \log n)$ com alta probabilidade. Consideremos um elemento x de uma instância S para o algoritmo 16 e $S_0 = S, S_1, S_2, \ldots, S_M$ as subsequências 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). O tamanho esperado das subsequências pode ser estimado por

$$\mathbb{E}[\,|S_{i+1}|\,|\,|S_i| = k\,] \leqslant p\frac{9k}{10} + (1-p)k < 0,8\frac{9k}{10} + 0,25k = 0,97k$$

já que com probabilidade $p \leqslant 0,8$ temos $|S_{i+1}| \leqslant 9|S_i|/10$ e com probabilidade $1-p < 0,25$ temos, trivialmente, $|S_{i+1}| \leqslant |S_i|$. Agora, podemos usar a Lei de Esperança

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

Total, Teorema 2.46, equação (2.28), para escrever

$$\begin{aligned}\mathbb{E}\,|S_{i+1}| &= \sum_k \mathbb{E}[\,|S_{i+1}|\,|\,|S_i| = k\,]\mathbb{P}[\,|S_i| = k\,] \\ &< \sum_k 0{,}97k\mathbb{P}[\,|S_i| = k\,] \\ &= 0{,}97\sum_k k\mathbb{P}[\,|S_i| = k\,] \\ &= 0{,}97\,\mathbb{E}\,|S_i|\end{aligned}$$

iterando essa desigualdade concluímos $\mathbb{E}\,|S_M| < (0{,}97)^M n$. Para $M = a\log_{100/97} n$ com $a \geqslant 3$ constante, temos $\mathbb{E}\,|S_M| < n^{1-a}$. Pela desigualdade de Markov

$$\mathbb{P}[\,|S_M| > 20\,] \leqslant \frac{1}{20n^{a-1}}$$

de modo que a probabilidade de que exista um $x \in S$ que passe por mais que $a\log_{100/97} n$ particionamentos é

$$< n\frac{1}{20n^{a-1}} = \frac{1}{20n^{a-2}}$$

ou seja, o número de comparações efetuadas pelo *quicksort* é $O(n\log n)$ com probabilidade pelo menos $1 - 1/(20n^{a-2}) \geqslant 1 - 1/(20n)$.

Exemplo 5.2 (*satisfatibilidade de fórmulas CNF*). Seja $C_1 \wedge C_2 \wedge \cdots \wedge C_m$ uma fórmula booleana CNF e consideremos uma valoração aleatória das variáveis da fórmula. Tomemos por X o número de cláusulas falsas com essa atribuição. Se c_i é o número de literais na cláusula C_i então a probabilidade de C_i ser falsa com a valoração é 2^{-c_i}, portanto $\mathbb{E}\,X = \sum_{i=1}^m 2^{-c_i}$. Pela desigualdade de Markov

$$\mathbb{P}[\,X \geqslant 1\,] \leqslant \sum_{i=1}^m 2^{-c_i}$$

de modo que se $\sum_{i=1}^m 2^{-c_i} < 1$, então $\mathbb{P}[\,X \geqslant 1\,] < 1$ ou seja, o evento $[X = 0]$ (a fórmula é satisfazível) tem probabilidade positiva. Em particular, para fórmulas em que cada cláusula tem k-literais $\sum_{i=1}^m 2^{-c_i} = m/2^k$ logo se o número de cláusulas m é menor que 2^k então a fórmula admite uma valoração que a torna verdadeira. ◊

5.1.1 SATISFAZIBILIDADE DE FÓRMULA BOOLEANA ALEATÓRIA

Para todo $k \in \mathbb{N}$, podemos gerar uma fórmula booleana aleatória $\Phi_{n,m,k}$ na forma k-CNF com $n \in \mathbb{N}$ variáveis e $m = m(n) \in \mathbb{N}$ cláusulas, sorteando cláusulas de modo aleatório e independente dentre toda as $\binom{n}{k}2^k$ possíveis cláusulas não triviais (i.e., sem literais complementares) com k literais. A *densidade* de uma fórmula é m/n, o número de cláusulas por variável.

Por volta dos anos de 1990, uma parte significativa do esforço para compreender o problema da satisfazibilidade foi dedicada ao estudo de instâncias geradas aleatoriamente, pois alguns resultados sugeriram que decidir SAT é, tipicamente, fácil. Isso se deve à experiência empírica de que uma fórmula aleatória com uma certa densidade de cláusulas é quase certamente satisfazível ou quase certamente insatisfazível, dependendo se a densidade está abaixo ou acima de um certo valor crítico. Durante vários anos, buscou-se comprovar que o problema de satisfazibilidade de fórmula booleana apresenta o fenômeno de transição de fase: existe uma densidade crítica α_c tal que, se a densidade $\alpha = m/n$ de uma fórmula $\Phi_{n,m,k}$ for menor que α_c, então a fórmula é quase certamente satisfazível, enquanto, se $\alpha > \alpha_c$ a fórmula é quase certamente insatisfazível. Tipicamente, esse estudo é feito com k fixo enquanto n é arbitrariamente grande.

Um dos primeiros fatos conhecidos a esse respeito é que uma fórmula $C_1 \wedge \cdots \wedge C_m$, com $m = \alpha n$ para $\alpha > 0$ constante, é quase certamente não-satisfazível para $\alpha > 2^k \ln(2)$ e satisfazível com probabilidade uniformemente positiva para $\alpha < 2^k/k$ (o leitor interessado pode encontrar o histórico desse problema dado por Achlioptas, 2021). Notemos que os limiares $2^k \ln(2)$ e $2^k/k$ estão bem distantes.

O limiar inferior deduzimos da desigualdade de Markov. Para uma fórmula $\Phi = \Phi_{n,m,k}$ com $m = \alpha n$ e k fixo, temos 2^k valorações para as k variáveis de uma cláusula C_i qualquer de Φ. Como apenas uma dessas atribuições torna a cláusula falsa, com probabilidade 2^{-k} uma cláusula não é satisfeita. Como a fórmula possui m cláusulas, a probabilidade de todas serem verdadeiras é $(1 - 2^{-k})^m$.

Se $X = X(\Phi)$ é o número de valorações que satisfazem a fórmula Φ, o número esperado de valorações que tornam a fórmula Φ verdadeira é

$$\mathbb{E}\, X = 2^n \left(1 - \frac{1}{2^k}\right)^m = \left(2\left(1 - \frac{1}{2^k}\right)^\alpha\right)^n \tag{5.2}$$

pois $m = m(n) = \alpha n$, onde $\alpha > 0$ é uma constante. Ainda, de $\ln(1 - x) < -x$, para todo $x \in (0, 1)$,

$$\ln\left(2\left(1 - \frac{1}{2^k}\right)^\alpha\right) < \ln 2 + \alpha\left(-\frac{1}{2^k}\right)$$

donde deduzimos que $2\left(1 - (1/2)^k\right)^\alpha < 1$ se, e só se, $\alpha > 2^k \ln(2)$, ou seja, *se* $\alpha > 2^k \ln(2)$ *então o número esperado de atribuições que torna a fórmula* $\Phi_{n,m,k}$ *satisfazível é* ε^n, *para* $\varepsilon \in (0, 1)$, *que tende a* 0 *quando* $n \to \infty$.

Pela desigualdade de Markov $\mathbb{P}[X = 0] > 1 - \mathbb{E}\, X$ e se $\mathbb{E}\, X \to 0$ o evento $[X = 0]$ ocorre quase certamente, ou seja, a quantidade de cláusulas satisfeitas e 0 quase certamente

$$\lim_{n \to \infty} \mathbb{P}[\,\Phi_{n,m,k} \text{ é satisfazível}\,] = 0$$

sempre que $\alpha > 2^k \ln(2)$.

Achlioptas e Peres (2004) reduziram a distância entre $2^k \ln(2)$ e $2^k/k$, eles demonstraram que o limiar justo é $2^k \log(2) - O(k)$. O problema de determinar o limiar justo foi completamente resolvido por Ding, Sly e Sun (2022). $\Diamond$

5.1.2 MÉTODOS PROBABILÍSTICOS DE 2º MOMENTO

Vimos que pela desigualdade de Markov

$$\mathbb{P}[X = 0] > 1 - \mathbb{E}\,X = 1 - o(1), \text{ se } \mathbb{E}\,X \to 0$$

onde $\mathbb{E}\,X$ está parametrizada por alguma variável que tende ao infinito. Por outro lado, caso $\mathbb{P}[X \neq 0] \geqslant \mathbb{E}\,X$ e $\mathbb{E}\,X \to \infty$, não há garantia que $[X \neq 0]$ ocorra quase certamente. Por exemplo, se $\mathbb{P}[X = 0] = 1 - 1/n$ e $\mathbb{P}[X = n^2] = 1/n$, então $\mathbb{E}\,X = n$ e X tem maior probabilidade de valer 0 para n grande. Nesses casos, uma estratégia de segundo momento pode ajudar. Os métodos de segundo momento são técnicas usadas para analisar uma distribuição de probabilidade baseados em cálculos que dependem do quadrado de uma distribuição.

Sejam X e Y duas variáveis aleatórias reais com quadrado integráveis, portanto, elas mesmas integráveis pelo seguinte corolário do Teorema 2.27: *Se X^2 é integrável então X é integrável.* De fato, notemos que $\mathbb{E}\,X^2$ e $\mathbb{E}\,|X|$ sempre estão definidas por serem séries com termos não negativos e o enunciado segue da desigualdade $0 \leqslant |x| \leqslant x^2 + 1$, que vale para todo real x. Da monotonicidade e da linearidade da esperança $0 \leqslant \mathbb{E}\,|X| \leqslant \mathbb{E}\,X^2 + 1$ e como o lado direito é finito segue que $\mathbb{E}\,|X| < +\infty$.

Desigualdades de Cauchy–Bunyakovsky–Schwarz e de Paley–Zygmund A famosa desigualdade de Cauchy–Bunyakovsky–Schwarz, na versão para variáveis aleatórias, afirma que a seguinte desigualdade é válida

$$(\mathbb{E}[XY])^2 \leqslant \mathbb{E}\,[X^2] \cdot \mathbb{E}\,[Y^2]. \tag{5.3}$$

Para demonstrá-la tomamos a variável aleatória $Z = X - \lambda Y$, onde λ é o valor de que minimiza $\mathbb{E}\,[Z^2] = \mathbb{E}\,[(X-\lambda Y)^2] = \mathbb{E}\,[X^2 - 2\lambda XY + \lambda^2 Y^2] = \mathbb{E}\,[X^2] - 2\lambda\,\mathbb{E}\,[XY] + \lambda^2\,\mathbb{E}\,[Y^2]$, que é o vértice da parábola parametrizada por λ (assumimos que $\mathbb{E}\,[Y^2] \neq 0$ pois caso contrário a desigualdade vale trivialmente), ou seja,

$$\lambda = \frac{\mathbb{E}\,[XY]}{\mathbb{E}\,[Y^2]}.$$

Agora, substituímos esse valor de λ em $\mathbb{E}\,[Z^2]$ e, como $\mathbb{E}\,[Z^2] \geqslant 0$, obtemos

$$\mathbb{E}\,[Z^2] = \mathbb{E}\,[X^2] - 2\frac{\mathbb{E}\,[XY]^2}{\mathbb{E}\,[Y^2]} + \frac{\mathbb{E}\,[XY]^2}{\mathbb{E}\,[Y^2]} = \mathbb{E}\,[X^2] - \frac{\mathbb{E}\,[XY]^2}{\mathbb{E}\,[Y^2]} \geq 0$$

que pode ser rearranjado para obter a desigualdade da equação (5.3).

Teorema 5.3 (desigualdade Cauchy–Bunyakovsky–Schwarz) *Seja X e Y são variáveis aleatórias reais com os seus quadrados integráveis, então XY é integrável e*

$$|\mathbb{E}[XY]| \leqslant \sqrt{\mathbb{E}[X^2]} \cdot \sqrt{\mathbb{E}[Y^2]}$$

com igualdade se, e somente se, $X = \lambda Y$ com probabilidade 1, para algum $\lambda \in \mathbb{R}$. □

Demonstração. De X^2 e Y^2 integráveis e $2|XY| \leqslant |X|^2 + |Y|^2$ deduzimos $\mathbb{E}|XY| < \infty$. Ademais, na dedução acima, vale a igualdade se existe λ_0 tal que $\mathbb{E}[(X - \lambda_0 Y)^2] = 0$, portanto, $\mathbb{P}[X - \lambda_0 Y = 0] = 1$ pelo Exercício 2.55, item 3. □

Observação 5.4. Para um espaço de probabilidade $(\Omega, \mathcal{A}, \mathbb{P})$, o conjunto de todas as variáveis aleatórias $X\colon \Omega \to \mathbb{R}$ tais que $\mathbb{E}|X|^2 < \infty$ é um espaço vetorial. Nesse espaço vetorial, denotado por $L^2(\Omega, \mathcal{A}, \mathbb{P})$, tem-se o produto interno $\mathbb{E}[XY]$ e a norma induzida por esse produto interno $\|X\| = \sqrt{\mathbb{E}[X^2]}$, de modo que em $L^2(\Omega, \mathcal{A}, \mathbb{P})$ temos as noções de ângulo e distância. Além disso, toda sequência de Cauchy[2] converge. Portanto, esse espaço vetorial é um *espaço de Hilbert*.

Como aplicação da desigualdade e Cauchy–Schwarz derivamos a seguir uma lei de desvio.

Corolário 5.5 (desigualdade de Paley–Zygmund) *Seja X uma variável aleatória que assume valores não negativos e com quadrado integrável. Então*

$$\mathbb{P}[X > \alpha \mathbb{E}X] \geqslant (1-\alpha)^2 \frac{(\mathbb{E}X)^2}{\mathbb{E}[X^2]} \tag{5.4}$$

para todo real $\alpha \in [0,1]$.

Demonstração. Da igualdade $X = X\mathbb{1}_{[X \leqslant \alpha \mathbb{E}X]} + X\mathbb{1}_{[X > \alpha \mathbb{E}X]}$, da linearidade da esperança e da desigualdade de Cauchy–Schwarz deduzimos

$$\begin{aligned}
\mathbb{E}X &= \mathbb{E}[X\mathbb{1}_{[X \leqslant \alpha \mathbb{E}X]}] + \mathbb{E}[X\mathbb{1}_{[X > \alpha \mathbb{E}X]}] \\
&\leqslant \alpha \mathbb{E}X + \mathbb{E}[X\mathbb{1}_{[X > \alpha \mathbb{E}X]}] \\
&\leqslant \alpha \mathbb{E}X + \sqrt{\mathbb{E}[X^2] \cdot \mathbb{E}[\mathbb{1}_{[X > \alpha \mathbb{E}X]}]} \\
&\leqslant \alpha \mathbb{E}X + \sqrt{\mathbb{E}[X^2] \cdot \mathbb{P}[X > \alpha \mathbb{E}X]}
\end{aligned}$$

donde segue (5.4). □

Corolário 5.6 *Seja X uma variável aleatória que assume valores não negativos, com quadrado integrável e $\mathbb{E}[X^2] \neq 0$. Então*

$$\mathbb{P}[X > 0] \geqslant \frac{(\mathbb{E}X)^2}{\mathbb{E}[X^2]}. \tag{5.5}$$

[2] Em matemática, uma sequência de Cauchy é uma sucessão tal que a distância (na norma induzida) entre os termos vai se aproximando de zero.

Observe que pela desigualdade de Jensen, Exercício 2.34, $\mathbb{E}[\,X^2\,] \geqslant (\mathbb{E}\,X)^2$.

Exemplo 5.7 (*hashing 2-a-2 independente*). Tomemos uma família de funções de hash $\mathcal{H} \subset N^U$ 2-a-2 independentes, isto é, tais que

$$\mathop{\mathbb{P}}_{h\in\mathcal{H}}([h(x) = i] \cap [h(y) = j]) = \frac{1}{|N|^2},$$

para quaisquer $x, y \in U$ distintos e para quaisquer $i, j \in N = \{0, \ldots, n-1\}$. Decorre da equação acima que $\mathbb{P}_{h\in\mathcal{H}}[h(x) = i] = 1/n$.

Sejam $S \subset U$ fixo de cardinalidade m, X_i a variável aleatória indicadora do evento $[h(i) = 0]$ e X a quantidade de elementos de S que são mapeados em 0. Pela desigualdade de Markov (eq. (5.1))

$$\mathbb{P}[\,X > 0\,] \leqslant \mathbb{E}\,X = \frac{|S|}{n}.$$

Agora,

$$\mathbb{E}[\,X^2\,] = \sum_{i\in S}\sum_{j\in S}\mathbb{E}\Big[\,X_iX_j\,\Big] = \sum_{i\in S}\mathbb{P}[\,h(i) = 0\,]\sum_{j\in S}\mathbb{P}[\,h(j) = 0 \mid h(i) = 0\,]$$

usando o fato das imagens serem independentes 2-a-2

$$\mathbb{E}[\,X^2\,] = \sum_{i\in S}\frac{1}{m}\left(1 + \sum_{\substack{j\in S\\ j\neq i}}\frac{1}{n}\right) = \frac{m}{n}\left(1 + \frac{m-1}{n}\right) \leqslant \frac{m}{n}\left(1 + \frac{m}{n}\right)$$

e da equação (5.5)

$$\mathbb{P}[\,X > 0\,] \geqslant \frac{(\mathbb{E}\,X)^2}{\mathbb{E}[\,X^2\,]} = \frac{m/n}{1 + m/n} = \frac{m}{m+n}.$$

Reunindo as desigualdades

$$\frac{m}{m+n} \leqslant \mathbb{P}[\,X \neq 0\,] \leqslant \frac{m}{n}.$$

Na seção 5.2.2 estudaremos com mais detalhes as famílias de funções de *hash* 2-a-2 independentes. ♢

Primeiro e segundo momentos Se X é variável aleatória com valores inteiros não negativos, então

$$\frac{(\mathbb{E}\,X)^2}{\mathbb{E}[\,X^2\,]} \leqslant \mathbb{P}[\,X \neq 0\,] \leqslant \mathbb{E}\,X \tag{5.6}$$

e se $\mathbb{E}\,X \to 0$, então a segunda desigualdade (de primeiro momento) garante que ocorre $[X = 0]$ com alta probabilidade, ou *quase certamente*. Por outro lado, $\mathbb{E}\,X \to \infty$ não significa que $X > 0$ com alta probabilidade, entretanto, se além disso $\mathbb{E}[\,X^2\,] = (1 + o(1))(\mathbb{E}\,X)^2$, então pela primeira desigualdade em (5.6) (de segundo momento)

temos $\mathbb{P}(X \neq 0) = 1 - o(1)$, ou seja, $[X \neq 0]$ ocorre com alta probabilidade, ou *quase certamente*, ou seja

$$\mathbb{P}[X \neq 0] = \begin{cases} o(1), & \text{se } \mathbb{E}\,X \to 0 \\ 1 - o(1), & \text{se } \mathbb{E}\,X \to \infty \text{ e } \mathbb{E}[X^2] = (1 + o(1))(\mathbb{E}\,X)^2 \end{cases}$$

onde $o(1)$ representa uma função (não negativa) que tende a 0.

5.1.3 TRIÂNGULOS EM GRAFO ALEATÓRIO

Denotamos por $\mathbb{G}_{n,p}$ o modelo binomial de grafo aleatório obtido considerando o conjunto de vértices $V = \{1, 2, \ldots, n\}$ e cada par mão ordenado $\{u, v\}$ de vértices está presente no conjunto de arestas E com probabilidade p.

Um triângulo em $\mathbb{G}_{n,p}$ é uma tripla de vértices $T = \{u, v, w\} \subset V$ com as tês arestas presentes em E. Uma tripla fixa de vértices T forma um triângulo no $\mathbb{G}_{n,p}$ com probabilidade p^3.

Consideremos uma enumeração $T_1, \ldots, T_{\binom{n}{3}}$ das triplas de vértices e X_i a variável aleatória indicadora de presença do triângulo T_i no grafo aleatório. Se X é a quantidade de triângulos no $\mathbb{G}_{n,p}$, o valor esperado para número de triângulos é

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

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 (Exercício 155), em particular, $\lim_{n\to\infty} \mathbb{P}[X = 0] = e^{-c^3/6}$.

Se $p \ll n^{-1}$, então $EX = \Theta(n^3 p^3) \to 0$ e pela desigualdade de Markov $\mathbb{P}[X > 0] \leqslant \mathbb{E}\,X = o(1)$. Por exemplo, se $p = p(n) = (n \log n)^{-1}$, então

$$\mathbb{E}\,X = \left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)\frac{1}{6\log^3 n}$$

e pela desigualdade de Markov $\mathbb{P}[X \neq 0] \leqslant \mathbb{E}\,X < \log^{-3} n$. Portanto, $\mathbb{P}[X = 0] \geqslant 1 - o(1)$, isto é, com alta probabilidade $\mathbb{G}_{n,p}$ não tem triângulo.

Por outro lado, se $p \gg n^{-1}$, então $\mathbb{E}\,X \to \infty$ quando $n \to \infty$. A condição $\mathbb{E}\,X \to \infty$ não garante que que $X > 0$ com alta probabilidade (veja o Exercício 5.43 no final do capítulo). Aqui entra o princípio de segundo momento. Vamos estimar a somatória

$$\mathbb{E}[X^2] = \mathbb{E}\left[\left(\sum_{i=1}^{\binom{n}{3}} X_i\right)^2\right] = \mathbb{E}\left[\left(\sum_{i=1}^{\binom{n}{3}} X_i^2 + \sum_{i\neq j} X_i \cdot X_j\right)\right] = \sum_{i=1}^{\binom{n}{3}} \mathbb{E}[X_i^2] + \sum_{i\neq j} \mathbb{E}[X_i \cdot X_j] \quad (5.7)$$

em que o último somatório é sobre todo par i, j 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 $\mathbb{P}([X_i = 1] \cap [X_j = 1]) = \mathbb{P}[X_i = 1]\mathbb{P}[X_j = 1] = p^6$;

3. as duas triplas são disjuntas, nesse caso $\mathbb{P}([X_i = 1] \cap [X_j = 1]) = \mathbb{P}[X_i = 1]\mathbb{P}[X_j = 1] = p^6$.

A contribuição do primeiro caso para a soma $\sum \mathbb{E}[X_i \cdot X_j]$ 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, $\binom{n}{4}\binom{4}{2}p^5 = O(n^4p^5) = \frac{1}{n^2p}O(n^6p^6) = o(1)(\mathbb{E}\,X)^2$. Nos outros dois casos, a contribuição para $\sum \mathbb{E}[X_i \cdot X_j]$ 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 direito da equação (5.7) usamos que $X_i^2 = X_i$, logo

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

portanto

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

assim $\mathbb{G}_{n,p}$ contém triângulo quase certamente.

5.1.4 DISTRIBUIÇÃO DE BOLAS EM CAIXAS

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

Lema 5.8 *Suponha que em n caixas sejam distribuídas $m = m(n)$ bolas de forma aleatória, uniforme e independente. Para qualquer constante $\varepsilon > 0$, se $m > (1+\varepsilon)n\ln(n)$, então, com alta probabilidade, não há caixas vazias; por outro lado, se $m < (1-\varepsilon)n\ln(n)$, então, com alta probabilidade, há pelo menos uma caixa vazia.*

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

$$\mathbb{E}\, X = \sum_{i=1}^{n} \mathbb{P}[\, X_i = 1\,] = n\left(1 - \frac{1}{n}\right)^m$$

usando que $(1 - 1/n)^n$ converge para 1/e, obtemos $\mathbb{E}\, X = \Theta\left(ne^{-m/n}\right)$.

Se $m > (1+\varepsilon) n \ln(n)$ então $\mathbb{E}\, X = O(n^{-\varepsilon})$, portanto, pela desigualdade de Markov (5.1) vale $\mathbb{P}[\, X > 0\,] \leqslant C\, n^{-\varepsilon}$ para alguma constante $C > 0$, ou seja, a probabilidade com que nenhuma caixa está vazia é alta

$$\mathbb{P}[\, X = 0\,] > 1 - C n^{-\varepsilon}. \tag{5.8}$$

Por outro lado, se $m < (1-\varepsilon) n \ln(n)$, então $\mathbb{E}\, X = \Omega(n^{\varepsilon})$ e, nesse caso, usamos segundo momento

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

em que o último somatório é sobre todo par de inteiros distintos $1 \leqslant i, j \leqslant n$, como em (5.7). Agora, $X_i^2 = X_i$, por ser variável indicadora, logo $\sum_i \mathbb{E}[\, X_i^2\,] = \mathbb{E}[\, X\,]$. Também, $X_i \cdot X_j$ é uma variável aleatória indicadora do evento "ambas caixas estão vazias" e $\mathbb{E}[\, X_i \cdot X_j\,] = \mathbb{P}([\, X_i = 1\,] \cap [\, X_j = 1\,]) = (1 - 2/n)^m$, portanto

$$\sum_{i \neq j} \mathbb{E}[\, X_i \cdot X_j\,] = n(n-1)(1-2/n)^m \leqslant n(n-1)(1-1/n)^{2m} = \left(1 - \frac{1}{n}\right)(\mathbb{E}\, X)^2$$

de modo que

$$\frac{(\mathbb{E}\, X)^2}{\mathbb{E}[\, X^2\,]} \geqslant \frac{(\mathbb{E}\, X)^2}{\mathbb{E}\, X + (1 - 1/n)(\mathbb{E}\, X)^2} = \frac{1}{1/\mathbb{E}\, X + (1 - 1/n)} = 1 - o(1)$$

quando $n \to \infty$. Usando a desigualdade (5.5), de Paley–Zygmund, concluímos que a probabilidade de ter não ter pelo menos uma caixa vazia é

$$\mathbb{P}[\, X = 0\,] < 1 - \frac{(\mathbb{E}\, X)^2}{\mathbb{E}[\, X^2\,]} = o(1). \tag{5.9}$$

Em resumo, temos das equações (5.8) e (5.9) que a probabilidade de não haver caixas vazias é

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

para todo $\varepsilon > 0$. □

Dos exemplos anteriores, o dos triângulos em grafos aleatórios e o das caixas vazias na distribuição de bolas em caixas, tiramos o seguinte extrato. Se $X_1, \ldots, X_n$ são variáveis aleatórias de Bernoulli identicamente distribuídas e $X = \sum_i X_i$, então

$$\mathbb{E}\, X^2 = \mathbb{E}\, X + \sum_{i \neq j} \mathbb{E}[\, X_i \cdot X_j \,] \tag{5.10}$$

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 $\mathbb{P}[\, X = 0 \,] = o(1)$.

Carga máxima no caso $n = m$ Agora, suponha que n bolas sejam distribuídas de forma aleatória, uniforme e independente em n caixas. Seja X a quantidade 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 $i = 1, 2, \ldots, n$. 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 1.6, página 12, 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

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

Para a esperança de X temos que $\mathbb{E}\, X \approx n(k)^{-k}$ que é constante quando $k^k \approx n$. Qual a função $k = k(n)$ para a qual $k^k = n$? Não há uma fórmula fechada mas conseguimos uma boa aproximação da seguinte forma. Tomando logaritmo e, novamente, tomando logaritmo, temos

(i) $k \ln(k) = \ln(n)$

(ii) $\ln(k) + \ln\ln(k) = \ln\ln(n)$.

Por um lado, $2\ln(k) \geqslant \ln(k) + \ln\ln(k) = \ln\ln(n)$, então $\ln(n)/\ln\ln(n) \geqslant k\ln(k)/(2\ln(k)) = k/2$. Por outro lado, $\ln(k) + \ln\ln(k) = \ln\ln(n) \geqslant \ln(k)$, então $\ln(n)/\ln\ln(n) \leqslant k\ln(k)/\ln(k) = k$. Com isso, concluímos que

$$k = k(n) = \Theta\left(\frac{\log(n)}{\log\log(n)}\right).$$

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

$$\left(\frac{e}{k}\right)^k = \left(\frac{e\ln\ln(n)}{4\ln(n)}\right)^{\frac{4\ln(n)}{\ln\ln(n)}} \leqslant \left(\frac{\ln\ln(n)}{\ln(n)}\right)^{\frac{4\ln(n)}{\ln\ln(n)}} = \left(e^{\ln\ln\ln(n) - \ln\ln(n)}\right)^{\frac{4\ln(n)}{\ln\ln(n)}} =$$

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

pois $4\ln\ln\ln(n)/\ln\ln(n)$ tem valor máximo 4/e em $n = e^{e^e}$. Portanto, $\mathbb{P}(X > 0) \leqslant n \cdot n^{-2}$, isto é, *a probabilidade de existir uma caixa com pelo menos k bolas é no máximo* $1/n$ ou ainda

$$\mathbb{P}[\,X = 0\,] = 1 - o(1)$$

de modo que com alta probabilidade a carga máxima numa caixa é $O(\ln n/\ln(\ln n))$.

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

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

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

Para calcular $\mathbb{E}\,X^2$ usamos a linearidade da esperança como na equação (5.10) e temos

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

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

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

Começamos com algumas simplificações

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

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

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

e o valor esperado de X^2 é limitado por

$$\mathbb{E}[X^2] = \mathbb{E}\,X + \left(1-\frac{1}{n}\right)^{-2k}\sum_{i\neq j}(\mathbb{E}\,X_i)(\mathbb{E}\,X_j) = \mathbb{E}\,X + \left(1-\frac{1}{n}\right)^{-2k}(\mathbb{E}\,X)^2.$$

Finalmente, usando a desigualdade (5.5),

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

mas $1/\mathbb{E}\,X \to 0$ e $(1-1/n)^{-2k} \approx e^{2k/n} \to 1$, 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.

Resumindo, a probabilidade de não haver caixas com pelo menos k bolas é

$$\mathbb{P}[X=0] = \begin{cases} 1-o(1), & \text{se } k \geqslant 4\ln(n)/\ln\ln(n), \\ o(1), & \text{se } k \leqslant \ln(n)/3\ln\ln(n). \end{cases}$$

5.2 MOMENTOS, VARIÂNCIA E A DESIGUALDADE DE CHEBYSHEV

Dado um inteiro positivo k, definimos o momento de ordem k, ou o k-**ésimo momento** da variável aleatória $X\colon \Omega \to \mathbb{R}$ como a esperança de X^k

$$\mathbb{E}\,X^k = \sum_t t^k\,\mathbb{P}_X(t)$$

quando está definida. De $|x|^k \leqslant |x|^{k+1}+1$ para todo $x \in \mathbb{R}$, temos que se o momento de ordem $k+1$ é finito então o de ordem k também é finito. Ainda, usando a desigualdade de Markov podemos provar facilmente que, dado um inteiro positivo k, vale

$$\mathbb{P}[|X| \geqslant t] \leqslant \mathbb{P}\Big[|X|^k \geqslant t^k\Big] \leqslant \frac{\mathbb{E}\,|X|^k}{t^k} \tag{5.11}$$

para toda constante $t>0$, sempre que $|X|^k$ é integrável.

A **função geradora de momentos** da variável X é dada por

$$M_X(s) := \mathbb{E}[e^{sX}]$$

para todo $s \in \mathbb{R}$ onde a esperança é finita. Se a função está definida em algum intervalo $(-x_0, x_0)$ da reta real, então a variável aleatória X tem os seus momentos finitos e, usando a série de Taylor, vale que

$$M_X(s) = \sum_{k\geqslant 0}\frac{s^k}{k!}\,\mathbb{E}[X^k]$$

para todo $s \in (-x_0, x_0)$.

Se X tem esperança finita, definimos o k-**ésimo momento central** por $\mathbb{E}\,(X - \mathbb{E}\,X)^k$. A **variância** de uma variável aleatória integrável é o segundo momento central dessa variável

$$\mathrm{Var}[X] := \mathbb{E}\,(X - \mathbb{E}\,X)^2 = \sum_t (t - \mathbb{E}\,X)^2\, \mathbb{P}_X(t) \tag{5.12}$$

a qual é finita sempre que X^2 é integrável. De fato, dado $\mu \in \mathbb{R}$ temos $(x^2 - \mu)^2 \leqslant 2x^2 + 2\mu^2$ para qualquer real x. Notemos que desenvolvendo o quadrado da diferença em $(X - \mathbb{E}\,X)^2$ e usando a linearidade da esperança chegamos na identidade

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

para a variância de X. A variância de uma variável aleatória é sempre não negativa e a sua raiz quadrada positiva é chamada de **desvio padrão**, tradicionalmente denotado por σ_X. A variância e o desvio padrão são medidas fundamentais de dispersão em estatística, elas nos ajudam a entender o quanto os valores de uma variável aleatória se afastam da média. O desvio padrão tema vantagem de estar na mesma unidade de medida da variável original.

Da equação (5.12) temos que a variância depende apenas da distribuição da variável aleatória.

Exemplo 5.9. Se $Z \sim \mathrm{Be}(p)$ então $\mathbb{E}\,Z = p$ e $\mathbb{E}\,Z^2 = 0^2(1-p) + 1^2 p = p$, logo $\mathrm{Var}[Z] = p - p^2 = p(1-p) = pq$. Notemos que a variância é maximizada quando $p = 1/2$. Se $Y \sim \mathrm{Geom}(p)$ então $\mathbb{E}\,Y = 1/p$ e $\mathbb{E}\,Y^2 = (2-p)/p^2$, como vimos na equação (2.2), logo $\mathrm{Var}[Y] = (1-p)/p^2$.

Em muitas aplicações a variável X é uma soma de variáveis aleatórias independentes e com mesma distribuição. Comecemos supondo que $X = \sum_{i=1}^n X_i$, com X_i^2 integrável para todo i e, portanto, de variância finita. Com essas hipóteses vale que

$$\mathbb{E}\left(\sum_{i=1}^n X_i\right)^2 - \left(\sum_{i=1}^n \mathbb{E}\,X_i\right)^2 = \sum_{i=1}^n \sum_{k=1}^n \mathbb{E}\,X_i X_k - \sum_{i=1}^n \sum_{k=1}^n \mathbb{E}\,X_i\,\mathbb{E}\,X_k$$

portanto a variância de X é

$$\mathrm{Var}[X] = \sum_{i=1}^n \sum_{k=1}^n (\mathbb{E}\,X_i X_k - \mathbb{E}\,X_i\,\mathbb{E}\,X_k) = \sum_{i=1}^n \sum_{k=1}^n \mathrm{Cov}(X_i, X_k), \tag{5.13}$$

o termo

$$\mathrm{Cov}(X_i, X_k) := \mathbb{E}\,X_i X_k - \mathbb{E}\,X_i\,\mathbb{E}\,X_k = \mathbb{E}\,(X_i - \mathbb{E}\,X_i)(X_k - \mathbb{E}\,X_k)$$

é conhecido como a **covariância** das variáveis aleatórias X_i e X_k. Quando $i = k$ a covariância coincide com variância.

Considerando os casos $i = k$ e $i \neq k$, separadamente, o lado direito da equação acima fica reescrito como

$$\sum_{i=1}^{n} \Big(\mathbb{E}[\,\mathsf{X}_i^2\,] - (\mathbb{E}\,\mathsf{X}_i)^2 \Big) + \sum_{i=1}^{n} \sum_{i \neq k=1}^{n} (\mathbb{E}\,\mathsf{X}_i\mathsf{X}_k - \mathbb{E}\,\mathsf{X}_i\,\mathbb{E}\,\mathsf{X}_k)$$

e reescrevendo a equação (5.13),

$$\operatorname{Var}[\,\mathsf{X}\,] = \sum_{i=1}^{n} \operatorname{Var}[\,\mathsf{X}_i\,] + 2 \sum_{i=1}^{n} \sum_{k=i+1}^{n} \operatorname{Cov}(\mathsf{X}_i, \mathsf{X}_k). \tag{5.14}$$

A covariância $\operatorname{Cov}(\mathsf{X}_i, \mathsf{X}_k)$ é finita sempre que as variáveis aleatórias tenham quadrado integrável e, mais que isso, $\operatorname{Cov}(\mathsf{X}_i, \mathsf{X}_k) = 0$ sempre que X_i e X_k são independentes por consequência do Teorema 2.31, página 122.

Proposição 5.10 *Se $\mathsf{X}_1, \ldots, \mathsf{X}_n$ são variáveis aleatórias integráveis e 2-a-2 independentes então*

$$\operatorname{Var}\left[\sum_{i=1}^{n} \mathsf{X}_i \right] = \sum_{i=1}^{n} \operatorname{Var}[\,\mathsf{X}_i\,]. \tag{5.15}$$

Demonstração. A afirmação segue imediatamente da equação (5.14) e da observação acima de que independência implica $\operatorname{Cov}(\mathsf{X}_i, \mathsf{X}_k) = 0$ para todos i e k distintos. □

Uma aplicação imediata do teorema acima é o cômputo da variância da distribuição binomial. Se $\mathsf{X} \sim \mathsf{b}(n, p)$ então a sua variância é $\operatorname{Var}[\,\mathsf{X}\,] = \sum_{i=1}^{n} \operatorname{Var}[\,\mathsf{X}_i\,] = \sum_{i=1}^{n} \mathbb{E}[\,\mathsf{X}_i\,] - (\mathbb{E}\,\mathsf{X}_i)^2 = np(1-p)$.

Exemplo 5.11 (*hashing 2-a-2 independente*). Suponha que numa tabela *hashing* os m elementos de S, subconjunto de um universo U, foram distribuídos por h dentre as n listas ligadas de uma tabela de espalhamento H de modo uniforme e 2-a-2 independente (Exemplo 5.7). Sabemos do capítulo anterior, página 110, que o número esperado de elementos numa lista é $c = m/n$. Nesse exemplo, vamos calcular a variância para o número de colisões e o tamanho de uma lista.

Sabemos, de capítulo anterior, que o número esperado de colisões é $\mathbb{E}\,\mathsf{C} = \binom{m}{2}/n$. Para calcular a variância, escrevemos C como a soma de variáveis aleatórias 2-a-2 independentes indicadoras de colisão, $\mathsf{C} = \sum_{i<j} \mathbb{1}_{[h(i)=h(j)]}$, assim

$$\operatorname{Var}[\,\mathsf{C}\,] = \sum_{i<j} \operatorname{Var}\Big[\,\mathbb{1}_{[h(i)=h(j)]}\,\Big] = \sum_{i<j} \frac{1}{n} - \left(\frac{1}{n}\right)^2 = \binom{m}{2}\left(\frac{1}{n} - \frac{1}{n^2}\right).$$

Agora, denotemos por $\ell_i = \ell_i(h)$ o tamanho da i-ésima lista $\ell_i = \sum_{s \in \mathsf{S}} \mathbb{1}_{[h(s)=i]}$ e se fizermos

$$\mathsf{X}_s := \mathbb{1}_{[h(s)=i]} - \frac{1}{n}$$

então $\mathbb{E}\, X_s = 0$ e $\sum_{s\in S} X_s = \ell_i - c$, para todo i. Essa soma é uma variável aleatória cuja esperança é 0. Ademais

$$\operatorname{Var}[\ell_i] = \mathbb{E}\,(\ell_i - c)^2 = \mathbb{E}\left(\sum_{s\in S} X_s\right)^2 = \sum_s \mathbb{E}[X_s^2] + \sum_{s\neq t} \mathbb{E}\, X_s X_t = \sum_s \mathbb{E}[X_s^2]$$

pois, caso $s \neq t$, vale $\mathbb{E}\, X_s X_t = \mathbb{E}\, X_s\, \mathbb{E}\, X_t$ pela independência das variáveis e a esperança é zero. Agora

$$\mathbb{E}[X_s^2] = \mathbb{E}\left[\left(\mathbb{1}_{[h(s)=i]} - \frac{1}{n}\right)^2\right] = \mathbb{E}\,\mathbb{1}^2_{[h(s)=i]} - 2\frac{1}{n}\mathbb{E}\,\mathbb{1}_{[h(s)=i]} + \left(\frac{1}{n}\right)^2 = \frac{1}{n} - \frac{1}{n^2}$$

portanto, a variância de ℓ_i é $\operatorname{Var}[\ell_i] = m(1/n - 1/n^2) = c(1 - 1/n)$. ◊

Exercício 5.12. Mostre que se X^2 é integrável e $a, b \in \mathbb{R}$ então

$$\operatorname{Var}[aX + b] = a^2 \operatorname{Var}[X].$$

Exercício 5.13 (variância condicional). Podemos definir de maneira natural o conceito de **variância condicional** por

$$\operatorname{Var}[X \mid Y] := \mathbb{E}[X^2 \mid Y] - \mathbb{E}[X \mid Y]^2.$$

Prove a **lei de variância total**:

$$\operatorname{Var}[X] = \mathbb{E}[\operatorname{Var}[X \mid Y]] + \operatorname{Var}[\mathbb{E}[X \mid Y]]$$

onde, no lado direito da igualdade, a esperança no primeiro termo da soma e a variância no segundo termo da soma são computadas sobre a distribuição de Y.

5.2.1 DESIGUALDADE DE CHEBYSHEV

A próxima desigualdade é conhecida por desigualdade de Chebyshev, também conhecida por desigualdade de Bienaymé–Chebyshev[3]. Da equação (5.11) obtemos a **desigualdade de Chebyshev**

$$\mathbb{P}[|X| \geqslant t] \leqslant \frac{\mathbb{E}[X^2]}{t^2} \tag{5.16}$$

que é mais conhecida na forma enunciada a seguir.

[3] Pafnuty Lvovich Chebyshev (1821–1894) foi um matemático russo conhecido por suas contribuição à Probabilidade, dentre seus alunos vários são matemáticos reconhecidos e um deles foi Andrey Markov. Irénée-Jules Bienaymé (1796–1878) foi um estatístico francês, contribuiu para a Probabilidades e o desenvolvimento da Estatística e suas aplicações.

Teorema 5.14 (desigualdade de Chebyshev) *Para toda constante $t > 0$ e toda variável aleatória X com quadrado integrável temos*

$$\mathbb{P}[\,|X - \mathbb{E}\,X| \geqslant t\,] \leqslant \frac{\operatorname{Var}[X]}{t^2}. \tag{5.17}$$

Demonstração. Segue da desigualdade (5.16) com a variável aleatória $X - \mathbb{E}\,X$. □

A desigualdade (5.17) pode ser justa como no seguinte exemplo. Seja X uma variável aleatória que assume o valor 0 com probabilidade $q \in (0,1)$ e os valores 1 e -1 com probabilidade $p/2$ cada, com $p = 1 - q$. A esperança de X é 0 e a variância é p. A probabilidade do evento $[|X| \geqslant 1]$ é p de modo que

$$p = \mathbb{P}[\,|X - 0| \geqslant 1\,] \leqslant \frac{\operatorname{Var}[X]}{1^2} = p.$$

Tomando a constante $t\sigma_X > 0$ na desigualdade, obtemos

$$\mathbb{P}[\,|X - \mathbb{E}\,X| \geqslant t\sigma_X\,] \leqslant \frac{1}{t^2}$$

e com $t\,\mathbb{E}\,X > 0$, obtemos

$$\mathbb{P}[\,|X - \mathbb{E}\,X| \geqslant t\,\mathbb{E}\,X\,] \leqslant \frac{\operatorname{Var}[X]}{t^2(\mathbb{E}\,X)^2} \leqslant \frac{\mathbb{E}[X^2]}{t^2(\mathbb{E}\,X)^2} \tag{5.18}$$

onde a segunda desigualdade segue do fato de

$$\frac{\operatorname{Var}[X]}{(\mathbb{E}\,X)^2} = \frac{\mathbb{E}[X^2]}{(\mathbb{E}\,X)^2} - 1.$$

2º momento A desigualdade de Chebyshev também é usada para mostrar a existência de subestruturas, como no caso de triângulos no $\mathbb{G}_{n,p}$ ou caixas com muitas bolas, com alta probabilidade, por meio do método de segundo momento. Se $X \geqslant 0$ é uma variável aleatória com quadrado integrável, provamos que $\mathbb{P}[\,X > 0\,] \to 1$ sempre que conseguimos mostrar

$$\frac{\operatorname{Var}[X]}{(\mathbb{E}\,X)^2} \to 0 \quad \text{ou} \quad \frac{\mathbb{E}[X^2]}{(\mathbb{E}\,X)^2} \to 1 \tag{5.19}$$

pois, como corolário da desigualdade de Chebyshev, obtemos da equação (5.18) com $t = 1$ a desigualdade

$$\mathbb{P}[\,X > 0\,] > 1 - \frac{\operatorname{Var}[X]}{(\mathbb{E}\,X)^2} \tag{5.20}$$

para uma variável aleatória $X \geqslant 0$ que assume valores inteiros. Usando que $\operatorname{Var}[\,aX\,] = a^2 \operatorname{Var}[X]$, para $a \in \mathbb{R}$, reescrevemos (5.19) como

$$\frac{\operatorname{Var}[X]}{(\mathbb{E}\,X)^2} = \operatorname{Var}\left[\frac{X}{\mathbb{E}\,X}\right] \to 0 \quad \text{ou} \quad \frac{\mathbb{E}[X^2]}{(\mathbb{E}\,X)^2} = \mathbb{E}\left[\left(\frac{X}{(\mathbb{E}\,X)}\right)^2\right] \to 1$$

o que, por Chebyshev, implica em

$$\mathbb{P}\left[\left|\frac{X}{\mathbb{E}\,X}-1\right|\leqslant\varepsilon\right]\to 1$$

ou seja, $\mathbb{P}[\,(1-\varepsilon)\mathbb{E}\,X<X<(1+\varepsilon)EX\,]\to 1$.

Em contraste com (5.20) acima, o Corolário 5.6 da desigualdade de Paley–Zygmund

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

fornece uma estimativa melhor pois

$$\frac{(\mathbb{E}\,X)^2}{\mathbb{E}[\,X^2\,]}=1-\frac{\mathrm{Var}[\,X\,]}{(\mathbb{E}\,X)^2+\mathrm{Var}[\,X\,]}\geqslant 1-\frac{\mathrm{Var}[\,X\,]}{(\mathbb{E}\,X)^2}.$$

Exemplo 5.15 (*colisões em tabelas de espalhamento*). Retomando o Exemplo 5.11, em uma tabela de espalhamento H com função de *hashing* uniforme e 2-a-2 independente (Exemplo 5.7), sabemos que o número esperado de colisões C tem variância

$$\mathrm{Var}[\,C\,]=\binom{m}{2}\left(\frac{1}{n}-\frac{1}{n^2}\right).$$

Pela desigualdade de Markov (já fizemos essa conta na página 111)

$$\mathbb{P}[\,C=0\,]\geqslant\frac{1}{2},\text{ se }m\leqslant\sqrt{n}$$

e por Chebyshev

$$\mathbb{P}[\,C=0\,]\leqslant\frac{\mathrm{Var}[\,C\,]}{\mathbb{E}[\,C\,]^2}=\frac{\binom{m}{2}\left(\frac{1}{n}-\frac{1}{n^2}\right)}{\left(\binom{m}{2}\frac{1}{n}\right)^2}=\frac{2(n-1)}{m(m-1)}\leqslant\frac{1}{2},\text{ se }m>2\sqrt{n-1}.$$

Exemplo 5.16 (*carga em tabelas de espalhamento*). Ainda no mesmo contexto do exemplo anterior, se ℓ_i é o tamanho da i-ésima lista e vamos estimar a probabilidade de ℓ_i desviar da média $c=m/n$. Computamos no Exemplo 5.11 a variância

$$\mathrm{Var}[\,\ell_i\,]=m\left(\frac{1}{n}-\frac{1}{n^2}\right).$$

e usando a desigualdade de Chebyshev com $t=(k-1)c$, para algum $k>1$,

$$\mathbb{P}[\,\ell_i\geqslant k\,\mathbb{E}\,\ell_i\,]\leqslant\mathbb{P}[\,|\ell_i-c|\geqslant(k-1)c\,]\leqslant\frac{c(1-1/n)}{(k-1)^2c^2}=\frac{1}{(k-1)^2c}\left(1-\frac{1}{n}\right).$$

Lei Fraca de Grandes Números Seja $(X_n\colon n\geqslant 1)$ uma sequência de variáveis aleatórias 2-a-2 independentes, identicamente distribuídas e quadrado integráveis. Então, estão bem definidos o valor esperado μ e variância σ^2 para toda X_i.

Definimos a média amostral

$$\overline{X}_n:=\frac{X_1+X_2+\cdots+X_n}{n}=\frac{S_n}{n}.$$

Do Exercício 5.12, junto com a aditividade da variância (eq. (5.15)) e a linearidade da esperança, deduzimos

$$\mathbb{E}\, S_n = n\mu \;\text{ e }\; \mathrm{Var}[\,S_n\,] = n\sigma^2$$
$$\mathbb{E}\, \overline{X}_n = \mu \;\text{ e }\; \mathrm{Var}[\,\overline{X}_n\,] = \frac{1}{n^2}\,\mathrm{Var}[\,S_n\,] = \frac{\sigma^2}{n}$$

e, notemos, em particular, que o desvio padrão de $\overline{X}_n$ é $\sigma/\sqrt{n}$ e decresce com $n \to \infty$.

Usando a desigualdade de Chebyshev para $\overline{X}_n$ e $\varepsilon > 0$ fixo

$$\mathbb{P}\Big[\,\big|\overline{X}_n - \mu\big| \geqslant \varepsilon\,\Big] \leqslant \frac{\sigma^2}{\varepsilon^2 n}$$

e notemos que o lado direito tende a zero quando n tende ao infinito, ou seja, para n suficientemente grande, é pouco provável que a média amostral esteja longe da média μ da população. Portanto

$$\lim_{n\to\infty} \mathbb{P}\Big(\Big\{\omega \in \Omega\colon \big|\overline{X}_n(\omega) - \mu\big| < \varepsilon\Big\}\Big) = 1$$

ou, ainda,

$$\lim_{n\to\infty} \mathbb{P}\Big[\,\big|\overline{X}_n - \mu\big| < \varepsilon\,\Big] = 1$$

para todo $\varepsilon > 0$, e dizemos que a média amostral **converge *em probabilidade*** ao valor esperado. Esse resultado é uma versão da Lei Fraca dos Grandes Números, chamada **Lei dos Grandes Números de Khintchine**[4].

Em particular, se $(X_n\colon n \geqslant 1)$ é uma sequência de variáveis aleatórias de Bernoulli com as hipóteses acima, temos $\mu = p$ e $\sigma^2 = p(1-p)$, logo

$$\mathbb{P}\Big[\,\big|\overline{X}_n - p\big| \geqslant \varepsilon\,\Big] \leqslant \frac{p(1-p)}{\varepsilon^2 n}.$$

Exemplo 5.17. Uma pesquisa será realizada no Rio de Janeiro para determinar a porcentagem da população que intenciona votar no Macaco Tião[5] nas próximas eleições para prefeito. Assumimos que os votos são sim ou não e são independentes.

Denotamos por n a quantidade de entrevistados e por p $(0 \leqslant p \leqslant 1)$ a fração, a ser estimada, dos votantes em Macaco Tião. Sejam X_i, para $i = 1, 2, \ldots, n$, variáveis aleatórias indicadoras dos votos, $X_i \in_{\mathrm{b}_p} \{0, 1\}$. Uma aproximação para p é dada pela variável aleatória

$$\frac{S_n}{n} := \frac{X_1 + \cdots + X_n}{n}$$

[4] Aleksandr Khinchin (1894–1959) ou, em francês, Alexandre Khintchine, foi outro notável matemático russo com grande contribuições para a Probabilidade.

[5] Macaco Tião foi um chimpanzé do Jardim Zoológico do Rio de Janeiro que virou uma celebridade por ter sido lançado candidato a prefeito em 1988, como voto de protesto. A votação era em cédulas de papel e foram registrados em torno 400 mil votos [fonte: verbete "Macaco Tião" na Wikipedia].

cuja esperança é p e variância é $p(1-p)/n \leqslant 1/(4n)$. Queremos que este valor aproximado seja diferente do valor real de p por no máximo $\pm 0{,}05$ (margem de erro) com probabilidade maior ou igual a 95%, isto é

$$\mathbb{P}\left[\left|\frac{S_n}{n} - p\right| < 0{,}05\right] \geqslant 0{,}95.$$

Pela desigualdade de Chebyshev é suficiente que tenhamos

$$1 - \frac{\mathrm{Var}[S_n/n]}{(0{,}05)^2} = 1 - \frac{p(1-p)}{(0{,}05)^2 n} \geqslant 1 - \frac{1}{4(0{,}05)^2 n} \geqslant 0{,}95$$

ou seja, $n \geqslant 2000$ entrevistados. ◊

Desigualdade de Chebyshev para soma de variáveis aleatórias indicadoras e 2-a-2 independentes Da equação (5.15), na página 172, para a soma de variáveis aleatórias 2-a-2 independentes $X_1, \ldots, X_n$ que assumem valores em $\{0, 1\}$, temos

$$\mathrm{Var}[X_i] = \mathbb{E}[X_i^2] - (\mathbb{E}\, X_i)^2 = \mathbb{E}\, X_i (1 - \mathbb{E}\, X_i) \leqslant \mathbb{E}\, X_i$$

para todo $1 \leqslant i \leqslant n$, portanto, $\mathrm{Var}[\sum_i X_i] = \sum_i \mathrm{Var}[X]_i \leqslant \sum_i \mathbb{E}\, X_i = \mathbb{E} \sum_i X_i$ e segue da desigualdade de Chebyshev, equação (5.17), que

$$\mathbb{P}[|X - \mathbb{E}\, X| \geqslant t] \leqslant \frac{\mathbb{E}\, X}{t^2}. \tag{5.21}$$

Em particular, se $X \sim \text{binomial}(n, p)$ então

$$\mathbb{P}[(1-\varepsilon)np < X < (1+\varepsilon)np] > 1 - \frac{1}{\varepsilon^2 np}.$$

5.2.2 *HASHING* 2-UNIVERSAL

No capítulo 2 fizemos várias estimativas sobre probabilidades de alguns eventos relacionados a uma função de *hash* aleatória como, por exemplo, a probabilidade com que dois elementos distintos colidem. Em seguida, usamos o fato de qualquer função de uma família que cumpra estatísticas relevantes, como a da probabilidade de colisão, dará o mesmo número médio de colisões que uma função de *hash* aleatória. As funções aleatórias não têm uma descrição pequena e a ideia chave para contornar esse problema foi dada por Carter e Wegman em 1979, e que exploramos na página 113, Exemplo 2.18: a função de *hash* é escolhida de uma família relativamente pequena de funções que são fáceis de computar, precisam de pouco espaço para serem descritas e têm algumas propriedades estatísticas importantes das funções aleatórias (Carter e Wegman, 1979). Por exemplo, para $U = N = \{0, 1, \ldots, p-1\}$ com p primo, tomemos a família $\{h_a \in N^U : a \in U\}$ das funções dadas por $h_a(x) := ax \bmod p$. Uma escolha aleatória $a \in_R U$ define unicamente

uma função e dados dois elementos x e y distintos no domínio, eles colidem com probabilidade $1/p$.

O problema é, então, encontrar função que se comporte como uma função aleatória, que não seja realmente uma função aleatória e que pode ser computada sem muito custo de tempo e memória. Esse problema é comum em muitos contextos na Ciência da Computação Teórica, é um dos principais problemas da teoria da desaleatorização. Nessa seção, usaremos *independência de pares* para caracterizar função que se comporta como uma função aleatória. De fato, essa é uma característica de uma família de funções, não é uma propriedade de uma única função.

Para uma família $\mathcal{H} = \{ h_\lambda \in \mathsf{N}^{\mathsf{U}} \colon \lambda \in \Lambda \}$ de funções de *hash*, pedimos que cada h_λ seja computável eficientemente, tenha descrição curta, que para todo $x \in \mathsf{U}$ e $\lambda \in_{\mathrm{R}} \Lambda$ a variável aleatória $h(x)$ dada por

$$h(x)(\lambda) \coloneqq h_\lambda(x)$$

tenha distribuição uniforme em N, que os eventos

$$[h(x) = i] \coloneqq \{ \lambda \in \Lambda \colon h_\lambda(x) = i \} \quad \text{e} \quad [h(y) = j] \coloneqq \{ \lambda \in \Lambda \colon h_\lambda(y) = j \}$$

sejam independentes para $x \neq y$ quaisquer e, finalmente, que $|\Lambda|$ não seja muito grande para que a escolha aleatória de λ também seja eficiente. Uma família de funções $\mathcal{H} \subset \mathsf{N}^{\mathsf{U}}$, com $|\mathsf{U}| \geqslant |\mathsf{N}|$, é uma família **2-universal de funções de *hash*** se munida da distribuição uniforme em Λ temos

$$\mathop{\mathbb{P}}_{\lambda \in \Lambda} ([h(x) = i] \cap [h(y) = j]) = \frac{1}{|\mathsf{N}|^2},$$

para quaisquer $x, y \in \mathsf{U}$ distintos e para quaisquer $i, j \in \mathsf{N}$.

Notemos que $[h(x) = i]$ e $[h(y) = j]$ são eventos independentes de Λ. A distribuição da variável aleatória $h(x)$ é a uniforme, pois dados x e $y \in \mathsf{U}$ distintos e $i \in \mathsf{N}$

$$\mathbb{P}_{h(x)}(i) = \mathop{\mathbb{P}}_{\lambda \in \Lambda} [h(x) = i] = \sum_{j \in \mathsf{N}} \mathop{\mathbb{P}}_{\lambda \in \Lambda} ([h(x) = i] \cap [h(y) = j]) = \sum_{j \in \mathsf{N}} \frac{1}{|\mathsf{N}|^2} = \frac{1}{|\mathsf{N}|}.$$

Ainda, para $\mathsf{S} \subset \mathsf{U}$ fixo e qualquer $i \in \mathsf{N}$ vale que o número esperado para o tamanho do subconjunto dos elementos de S que são mapeados para um único elemento do conjunto N é

$$c \coloneqq \mathbb{E} \sum_{x \in \mathsf{S}} \mathbb{1}_{[h(x)=i]} = \sum_{u \in \mathsf{S}} \frac{1}{|\mathsf{N}|} = \frac{|\mathsf{S}|}{|\mathsf{N}|} = \frac{m}{n}$$

onde $m \coloneqq |\mathsf{S}|$ e $n \coloneqq |\mathsf{N}|$.

A família de funções sobre $\mathsf{U} = \mathsf{N} = \{0, 1, \ldots, p-1\}$

$$\left\{ h_{(a,b)} \in \mathsf{N}^{\mathsf{U}} \colon a, b \in \mathsf{U},\ a \neq 0 \right\}$$

com p primo, dadas por $h_{(a,b)}(x) := ax + b \bmod p$ é 2-universal. Se $x \neq y$ então $h_{(a,b)}(x) = i$ e $h_{(a,b)}(y) = j$ para um único par $(a, b) \in \mathsf{U}^2$ com $a \neq 0$, portanto, para uma escolha aleatória uniforme de (a, b) ocorre $[h_{(a,b)}(x) = i] \cap [h_{(a,b)}(y) = j]$ com probabilidade $1/p^2$ para quaisquer i e j em M.

Exemplo 5.18. Seja $\mathbb{F}$ um corpo finito com q elementos e tomemos a família de funções de $\mathbb{F}^n$ em $\mathbb{F}$

$$\mathcal{H} := \left\{ h_{(a,b)}(x) := \langle a, x \rangle + b \colon\ a, x \in \mathbb{F}^n,\ b \in \mathbb{F},\ a \neq 0 \right\}$$

onde $\langle \mathsf{u}, \mathsf{v} \rangle$ denota o produto escalar usual $\langle \mathsf{u}, \mathsf{v} \rangle = \sum_k u_k v_k$. Uma escolha aleatória de uma função é dada por $(a, b) \in_{\mathrm{R}} \mathbb{F}^n \times \mathbb{F}$ com os sorteios $a_1, \ldots, a_n, b$ em $\mathbb{F}$ são independentes.

Se $x \neq 0$ então, para uma escolha aleatória de $a \in \mathbb{F}^n$, a probabilidade do evento $[\langle a, x \rangle = j]$ é $1/q$, pelo princípio da decisão adiada, pois devemos ter $a_k = (1/x_k)(j - \sum_{l \neq k} a_l x_l)$ para algum k tal que $x_k \neq 0$.

Dados $x, y \in \mathbb{F}^k$ distintos e $i, j \in \mathbb{F}$ o evento

$$[\langle a, x \rangle + b = i] \cap [\langle a, y \rangle + b = j]$$

definido pelos pares $(a, b) \in \mathbb{F}^n \times \mathbb{F}$ é dado também por

$$[\langle a, x \rangle - \langle a, y \rangle = i - j] \cap [b = \langle a, x \rangle - i].$$

Dos parágrafos anteriores temos $\mathbb{P}[\langle a, x \rangle - \langle a, y \rangle = i - j] = \mathbb{P}[\langle a, x - y \rangle = i - j] = 1/q$ e $\mathbb{P}[b = \langle a, x \rangle - i] = 1/q$. Logo

$$\mathbb{P}([\langle a, x \rangle + b = i] \cap [\langle a, y \rangle + b = j]) = \frac{1}{q^2}$$

e concluímos que $\mathcal{H}$ é 2-universal. ◊

No exemplo acima com o caso particular de $\mathcal{H}$ ser o conjunto das 2^3 funções de $\{0,1\}^2$ em $\{0,1\}$, cada função é dada por 3 bits, dos quais 2 são do parâmetro a e 1 do parâmetro b. Assim, sorteando 3 bits temos uma sequência de 4 bits com distribuição uniforme e 2-a-2 independentes: $h_{(a,b)}(00)$, $h_{(a,b)}(01)$, $h_{(a,b)}(10)$ e $h_{(a,b)}(11)$.

No caso geral do corpo binário, $\mathbb{F} = \{0,1\}$, sorteando $n + 1$ bits obtemos uma sequência de 2^n bits com distribuição uniforme e 2-a-2 independentes. Essa sequência pode ser usada para a desaleatorização de algoritmos que usam bits aleatórios com independência 2-a-2, como é o caso do algoritmo para achar cortes grandes em grafos (Algoritmo 20, página 136). Mais detalhes desse caso são dados na página **??**.

Lema de misturabilidade do *hashing* universal Uma propriedade útil de famílias de funções independentes em pares é chamada de misturabilidade[6], do inglês *mixing*. Fixados $S \subseteq U$ e $T \subseteq N$, suponha que escolhemos uma função aleatória $h \in N^U$. Para quaisquer $x \in S$ e $y \in T$ temos $h(x) = y$ com probabilidade $1/|N|$, assim esperamos que $|S||T|/|N|$ elementos de S têm imagem em T. Para uma família 2-universal de *hashing* essa propriedade, com um pequeno erro, também vale quando uma função é escolhida aleatoriamente.

No caso de uma função aleatória $h \in_R N^U$, se X é a quantidade de elementos de $S \subset U$ que são mapeados em $i \in N$, então $c = |S|/|N|$ é o valor esperado de X. A desigualdade de Chebyshev nos diz que

$$\mathbb{P}[\,|X - c| < \varepsilon c\,] > 1 - \frac{1}{\varepsilon^2 c}$$

ou seja, $1 - 1/\varepsilon^2 c$ das $|N|^{|U|}$ funções mapeiam os elementos de S de modo que $(1-\varepsilon)c < X < (1+\varepsilon)c$, isto é, a quantidade de elementos de S que são mapeados em i por um escolha aleatória de h é próximo a média. Esse resultado pode ser transferido para famílias 2-universal de funções de *hash*, para todo S suficientemente grande, quase toda função de uma família 2-universal distribui os elementos de S de modo aproximadamente uniforme no contradomínio.

Lema 5.19 *Seja $\mathcal{H} \subset N^U$ uma família 2-universal de funções* hash. *Para todo $\varepsilon > 0$, para todo $i \in N$ e todo $S \subset U$, se $|S| > \varepsilon^{-3}|N|$ então para pelo menos uma fração $1-\varepsilon$ das funções h em $\mathcal{H}$ vale*

$$(1-\varepsilon)\frac{|S|}{|N|} < \Big|\{\, u \in S \colon h(u) = i \,\}\Big| < (1+\varepsilon)\frac{|S|}{|N|}. \tag{5.22}$$

Demonstração. Para demonstrar esse resultado, vamos provar que para uma escolha aleatória uniforme de função em $\mathcal{H} = \{h_\lambda : \lambda \in \Lambda\}$ o evento correspondente a equação (5.22) tem probabilidade pelo menos $1 - \varepsilon$.

Sejam ε e S como no enunciado. Fixados $u \in U$ e $i \in N$, consideremos as variáveis aleatórias indicadoras $\mathbb{1}_{[h(u)=i]}$ do evento "a função sorteada mapeia u em i" e a variável aleatória X dada por

$$X(\lambda) := \sum_{u \in S} \mathbb{1}_{[h(u)=i]}(\lambda) = \Big|\{\, u \in S \colon h_\lambda(u) = i \,\}\Big|$$

cujo valor esperado, como já vimos, é $c = |S|/|N|$.

[6] Esse termo refere-se, usualmente em matemática, ao comportamento de processos nos quais, com o passar do tempo, as dependências entre os estados diminuem ou desaparecem. A propriedade de *mixing* indica que, para eventos distantes no tempo, a ocorrência de um deles torna-se quase independente da ocorrência do outro. Essa propriedade é importante em contextos nos quais se deseja garantir que o processo "esqueça" suas condições iniciais, movendo-se em direção a um equilíbrio estatístico.

Pela desigualdade de Chebyshev, equação (5.21) na página 177

$$\mathbb{P}[\,|X - c| \geqslant \varepsilon c\,] \leqslant \frac{c}{\varepsilon^2 c^2} \leqslant \varepsilon$$

a segunda desigualdade decorre de $c > \varepsilon^{-3}$. Daí, temos que para uma escolha aleatória uniforme em $\mathcal{H}$ vale com probabilidade maior que $1-\varepsilon$ que $|X - c| < \varepsilon c$ ou seja, para uma fração maior que $1-\varepsilon$ das funções

$$(1-\varepsilon)c < X < (1+\varepsilon)c$$

donde segue a afirmação do lema. □

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

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

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

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

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

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

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