O método probabilístico e algumas aplicações

Lucas Mendes Marques Gonçalves
Orientador: Yoshiharu Kohayakawa

Fevereiro de 2010

Parte I
Parte Objetiva

Introdução

O método probabilístico é um método para prova de teoremas. Tomando-se um espaço de probabilidade adequado, pode-se provar que uma determinada estrutura existe provando-se que ela ocorre, com probabilidade não nula, nesse espaço.

Esse método é amplamente utilizado em diversas áreas do conhecimento, incluindo especialmente tópicos relacionados à ciência da computação e à combinatória.

Nosso projeto foi a leitura de um livro a esse respeito, chamado "O Método Probabilístico". Com isso, o aluno pôde adquirir aptidão com o uso do método, e também ter contato com áreas diversas da Ciência da Computação, o que será útil na escolha da área de um futuro mestrado.

Após a leitura, selecionamos três tópicos que o livro apresenta, e os desenvolvemos em maior detalhe. É esse desenvolvimento que apresentamos aqui.

O primeiro resultado que apresentamos é, historicamente, o primeiro resultado da área de grafos aleatórios. Serve como introdução a essa área de estudo, e ilustra o uso de ferramentas clássicas para o uso do método probabilístico.

O segundo resultado que apresentamos é o teorema de Shannon. Esse teorema é um dos teoremas fundamentais de Teoria da Informação. Ele estabelece qual a mínima perda de velocidade necessária para se reduzir a probabilidade de erro num determinado meio físico de comunicação.

Nosso terceiro assunto é relacionado a geometria computacional. Abordamos um problema clássico da área, o problema de *range searching*, mas com um forte viés combinatório. O problema em si é bastante fundamental, e nosso estudo o aborda utilizando a dimensão de Vapnik–Chervonenkis, o que nos permite revelar algumas propriedades interessantes dos espaços que estudamos.

1 Grafos aleatórios e subgrafos pequenos

O estudo de grafos aleatórios é relevante, não só nas suas aplicações matemáticas, mas em problemas de teoria da computação, de redes, e mesmo em ciências sociais.

O problema que vamos explorar nas páginas que seguem é (uma variação) do primeiro problema da área: a determinação de quando um grafo aleatório contem um determinado subgrafo pequeno.

Primeiro, teremos que definir algumas coisas: Nosso modelo de grafo aleatório é $G(n,p)$. Isso quer dizer que estamos tomando grafos de n vértices, e incluindo cada aresta com probabilidade p. (ou, de forma mais precisa, que para todo grafo G de n vértices, temos $P(G) = p^{e_g}(1-p)^{\binom{n}{2}-e_g}$, onde e_g é o número de arestas de G)

Nosso problema é determinar quando $G(n,p)$ contém um determinado subgrafo H (fixo). Mais precisamente, queremos determinar um comportamento assintótico, ou seja, saber se H surge em G quando n vai a infinito. Note que esse problema é trivial se p é fixo (Como o tamanho de H é fixo e o de G vai a infinito, G sempre conterá H se $p > 0$ é fixo). Assim, nos preocuparemos com probabilidades que sejam funções decrescentes de n. De fato, pelo mesmo motivo, lidaremos apenas com p que vão a zero.

As provas dessa primeira sessão, a não ser quando dizemos explicitamente o contrário, são derivadas das provas de [NA08]

1.1 Alguns fatos necessários

Antes de enfrentarmos o problema em si, será necessário enunciar alguns fatos. Presumimos que o leitor se recorda da definição de esperança de uma variável aleatória, e de mais alguns conceitos básicos relacionados.

1.1.1 Desigualdades de Markov e Chebyshev

Seja $X > 0$ uma variável aleatória discreta. Apenas sabendo a esperança de X, já podemos fazer uma afirmação sobre a sua distribuição:

Teorema 1. $P[X \geq a] \leq \frac{E(X)}{a}$

Demonstração.

$$\begin{aligned} E(X) &= \sum_i p(X=i)i \geq \sum_{i \geq a} p(X=i)i \\ &\geq \sum_{i \geq a} p(X=i)a = a\sum_{i \geq a} p(X=i) = a.p(X \geq a) \end{aligned}$$

Assim,

$$E(X) \geq a.p(X \geq a) \text{ , ou seja, } P(X \geq a) \leq \frac{E(X)}{a}$$

□

Essa é a chamada Desigualdade de Markov. Utilizaremos essa desigualdade com $a = 1$ e X inteiro (X será o número de vezes que H aparece em G). Teremos, então:

$$P(X \neq 0) = P(X \geq 1) \leq E(X) \tag{1}$$

Assim, poderemos, provando que $E(X)$ é pequena, obter que $P(X \neq 0)$ também o é.

Infelizmente, como veremos em breve, não conseguiremos fazer algo análogo para $P(X \geq 1)$ e $E(X)$ grande. Nem mesmo quando $E(X)$ vai a infinito poderemos implicar que $P(X \geq 1)$ vai a um. Precisaremos estudar a variância para fazer alguma afirmação nesse sentido.

Lembremos a definição de variância,

$$\text{Var}(X) = \sigma^2 = E[(X-\mu)^2] \text{ , onde } \mu = E(X)$$

Aplicando a desigualdade de Markov, temos $P[(X-\mu)^2 \geq a^2] \leq \sigma^2/a^2$ e portanto

Teorema 2. $P[|X-\mu| \geq a] \leq \sigma^2/a^2$

Essa é a chamada desigualdade de Chebyshev.

Podemos aplicá-la com $a = \mu$, obtendo

$$P(X=0) \leq P[|X-\mu| \geq \mu] \leq \sigma^2/\mu^2 \tag{2}$$

Temos, assim, um limitante superior para a $P(X=0)$. Junto com o limitante para $P(X \neq 0)$, poderemos descrever bastante bem o comportamento de $G(n,p)$ em relação a conter ou não H. Mas, primeiro, precisaremos de algum trabalho técnico para utilizar a desigualdade acima.

1.1.2 Lidando com a variância

Nosso objetivo aqui é estabelecer uma maneira conveniente de lidar com a desigualdade (2). Vamos assumir algumas hipóteses sobre a variável aleatória X, e depois descreveremos melhor como elas se aplicam a nosso problema.

Suponhamos que $X = \sum X_i$, sendo X_i a variável indicadora[1] do evento A_i. Então, temos:

$$\sigma^2 = \sum_i \text{Var}(X_i) + \sum_{i \neq j} \text{Cov}[X_i, X_j]$$

Mas $\text{Var}(X_i) = E(X_i^2) - [E(X_i)^2] = p_i - p_i^2 = p_i(1 - p_i) \leq p_i$ (onde p_i é a probabilidade de A_i, i.e. a probabilidade de $X_i = 1$). Assim

$$\begin{aligned} \sigma^2 &\leq \sum_i p_i + \sum_{i \neq j} \text{Cov}[X_i, X_j] \\ &= E[X] + \sum_{i \neq j} \text{Cov}[X_i, X_j] \end{aligned}$$

Mas

$$\begin{aligned} &\text{Cov}[X_i, X_j] = 0 \text{ se } X_i \text{ e } X_j \text{ forem independentes} \\ &\text{Cov}[X_i, X_j] = E[X_i X_j] - E[X_i]E[X_j] \text{ , caso contrário} \end{aligned}$$

Usando que

$$E[X_i X_j] - E[X_i]E[X_j] \leq E[X_i X_j] = P[X_i = 1 \wedge X_j = 1] = P[A_i \wedge A_j]$$

E dizendo que $i \sim j$ se $i \neq j$ e A_i não é independente de A_j. Temos,

$$\begin{aligned} \sigma^2 &\leq E[X] + \sum_{i \neq j} \text{Cov}[X_i, X_j] \\ &= E[X] + \sum_{i \sim j} \text{Cov}[X_i, X_j] \\ &\leq E[X] + \sum_{i \sim j} P[A_i \wedge A_j] = E[X] + \Delta \end{aligned} \tag{3}$$

(Onde Δ é $\sum_{i \sim j} P[A_i \wedge A_j]$)

Mas,

$$\Delta = \sum_{i \sim j} P[A_i \wedge A_j] = \sum_i P[A_i] \sum_{j/j \sim i} P[A_j | A_i]$$

Se tivermos que $\sum_{j/j \sim i} P[A_j|A_i]$ independe de i, então chamaremos essa soma de Δ^* e teremos

$$\begin{aligned} \sigma^2 &\leq E[X] + \Delta \\ &= E[X] + E[X]\Delta^* \end{aligned} \tag{4}$$

Aplicando (2), obtemos:

$$P(X = 0) \leq \frac{\sigma^2}{E[X]^2} \leq \frac{E[X] + E[X]\Delta^*}{E[X]^2} \tag{5}$$

Fazendo essa ultima expressão ir a zero, obteremos o seguinte

[1] X_i é 1 se A_i ocorre e 0 caso contrário

Teorema 3. *Se $X = \sum X_i$, X_i variáveis indicadoras de eventos A_i, e temos que $\sum_{j\sim i} P[A_j|A_i]$ independe de i, e ainda supondo que $\lim_{n\to\infty} E[X] = \infty$ e $\Delta^* = o(E(X))$, temos*

$$\lim_{n\to\infty} P(X = 0) = 0$$

1.2 Achando a função limiar

Agora, já podemos provar nosso primeiro resultado. Estamos interessados em saber para quais $p(n)$ temos que $G(n,p)$ contém um determinado grafo H. Mais precisamente, para quais $p(n)$ temos que $\lim_{n\to\infty} P[H \subset G(n,p)] = 1$, e para quais $\lim_{n\to\infty} P[H \subset G(n,p)] = 0$, onde $H \subset G$ significa que há em G um grafo K, não necessariamente induzido, isomorfo a H

Dizemos que $f(n) \gg g(n)$ se $\lim_{n\to\infty} f(n)/g(n) = \infty$, e que $f(n) \ll g(n)$ se $\lim_{n\to\infty} f(n)/g(n) = 0$

Desejamos achar uma função $f(n)$ tal que, se $p(n) \gg f(n)$, $\lim_{n\to\infty} P[H \subset G(n,p)] = 1$ e, caso $p(n) \ll f(n)$, $\lim_{n\to\infty} P[H \subset G(n,p)] = 0$. A uma tal função chamaremos de função limiar.

Vamos, nesse momento, nos restringir a um tipo especial de grafo H, chamado balanceado. Posteriormente, generalizaremos nosso teorema para todo H fixo.

Definição 1. *A densidade de um grafo: $\rho(H) = e(H)/v(H)$*

Definição 2. *Um grafo H é dito balanceado se, $\rho(H) \geq \rho(L)$ para todo grafo $L \subseteq H$, e é dito estritamente balanceado se $\rho(H) > \rho(L)$ para todo $L \subset H$, com $L \neq H$*

Teorema 4. *Se H é um grafo balanceado, a função $f(n) = (1/n)^{1/\rho(H)}$ é uma função limiar para a propriedade de G conter H. Ou seja, se $p(n) \gg f(n)$ temos $\lim_{n\to\infty} P[H \subset G(n,p)] = 1$ e se $p(n) \ll f(n)$ temos $\lim_{n\to\infty} P[H \subset G(n,p)] = 0$*

Vamos provar cada limite do teorema acima separadamente. No que se segue, utilizaremos $\rho = \rho(H)$, $v = v(H)$, $e = e(H)$ e $p = p(n)$.

Fato. *Se $p(n) \ll f(n)$, então $\lim_{n\to\infty} P[H \subset G(n,p)] = 0$*

Demonstração. Vamos utilizar aqui a desigualdade (1), onde X será o número de cópias de H em $G(n,p)$

Como dissemos acima, uma cópia é um subgrafo K de $G(n,p)$ isomorfo a H.

Notemos que há, num grafo completo com n vértices, $\binom{n}{v}.v!/\text{aut}(H)$ tais subgrafos (onde $\text{aut}(H)$ é o número de automorfismos de H). Enumeremos os (possíveis) subgrafos correspondentes em $G(n,p)$ como $K_1, K_2, \ldots, K_j$. Vamos definir X_i, que será 1 se K_i for de fato subgrafo de $G(n,p)$ (i.e. se todas as arestas de K_i estiverem em $G(n,p)$) e 0 caso contrário.

Claramente, $E(X_i) = P(X_i = 1) = p^e$. Assim, temos,

$$E(X) = E(\sum_i X_i) = \sum_i E(X_i) = p^e \binom{n}{v}.v!/\text{aut}(H) = (n)_v p^e/\text{aut}(H) = \Theta(n^v p^e)$$

Mas $p \ll f(n)$. Assim

$$0 = \lim_{n\to\infty} p/f(n) = \lim_{n\to\infty} p/(1/n)^{v/e}$$

Portanto

$$0 = \lim_{n\to\infty}[p/(1/n)^{v/e}]^e = \lim_{n\to\infty} p^e/(1/n)^v = \lim_{n\to\infty} n^v p^e$$

De (1), temos

$\lim_{n\to\infty} P(X > 0) \leq \lim_{n\to\infty} E(X) \leq lim_{n\to\infty} Cn^v p^e = 0$ (usando a constante superior C da definição de Θ).

□

De maneira completamente análoga (e, em particular, sem utilizar a hipótese de balanceamento) obtemos que, se $p \gg f(n)$, $\lim_{n\to\infty} E[X] = \infty$

Ainda nem sequer provamos o teorema no caso restrito, mas já podemos antever o problema do caso geral. Tomando H um grafo não balanceado, sabemos que ele contém um H_m tal que $\rho(H) < \rho(H_m)$. A prova acima mostra que, tomando $p(n)$ tal que $(1/n)^{1/\rho(H)} \ll p(n) \ll (1/n)^{1/\rho(H_m)}$ (ver [2]) , a probabilidade de haver cópia de H_m em $G(n,p)$ vai a zero. Assim, a probabilidade de obtermos uma cópia de H vai a zero também.

Então, temos os seguintes fatos interessantes;

1. A esperança do número ocorrências de H vai a infinito, mas a probabilidade da ocorrência de H vai a 0
2. Nossa função limiar para o caso balanceado não pode ser a função limiar do caso geral
3. A esperança do número de ocorrências de H é maior do que a esperança do número de ocorrências de H_m, um subgrafo de H

Após essa digressão, retomamos a demonstração (em particular, voltamos a assumir H balanceado)

Fato. *Se* $p(n) \gg f(n)$*, então* $\lim_{n\to\infty} P[H \subset G(n,p)] = 1$

Demonstração. Vamos usar o teorema 3, usando o mesmo X do lema anterior. Verifiquemos que todas as condições são respeitadas:

Pela definição, X é soma de variáveis indicadoras.

Como já dissemos, $\lim_{n\to\infty} E[X] = \infty$.

Tomemos A_i o evento associado à variável X_i, ou seja, o evento de K_i ser subgrafo de $G(n,p)$. De fato, $\sum_{j\sim i} P[A_j|A_i]$ independe de A_i. (Tomando G o grafo completo de n vértices, K_i uma cópia de H em G e l um número natural, o número de K_a, cópias de H em G tais que $E(K_i) \cap E(K_a) = l$ independe de K_i)

Assim, basta verificar que $\Delta^* = o(E[X])$ e finalizaremos a prova.

Vamos analisar melhor a expressão $\sum_{j/j\sim i} P[A_j|A_i]$. No nosso problema, $i \sim j$ significa que o K_i e K_j são distintos e tem aresta em comum, e $P[A_j|A_i]$ é a probabilidade de que as arestas de K_j estarem em $G(n,p)$ caso as de K_i já estejam, ou seja, $p^{|E(K_j)-E(K_i\cap K_j)|}$.

Como $K_i \cap K_j$ é um subgrafo de K_j (e K_j é balanceado), temos que $e(K_i \cap K_j) \leq v(K_i \cap K_j)e/v$. Assim, $P[A_j|A_i] \leq p^{e-v(K_i\cap K_j)e/v}$

Vamos reescrever a soma da seguinte forma:

$\sum_{j/j\sim i} P[A_j|A_i] = \sum_{l=2}^{v(H)} \sum_{K_j/v(K_i\cap K_j)=l} P[A_j|A_i]$

[2]Note que é possível tomar tal $p(n)$, pois $(1/n)^{1/\rho(H)} \ll (1/n)^{1/\rho(H_m)}$

Mas o número de K_j com $v(K_i \cap K_j) = l$ é

$$\binom{v}{l}\binom{n}{v-l}v!/\text{aut}(H) = O\left(\binom{n}{v-l}\right) = O(n^{v-l})$$

Assim,

$$\begin{aligned}\Delta^* &\leq \sum_{l=2}^{v(H)} O(n^{v-l})p^{e-le/v} = \sum_{l=2}^{v(H)} O(n^{v-l}p^{e-le/v}) \\ &= \sum_{l=2}^{v(H)} O[(n^v p^e)^{1-l/v}] = \sum_{l=2}^{v(H)} o(n^v p^e) \leq vo(n^v p^e) = o(n^v p^e)\end{aligned}$$

Mas $E[X] = \Theta(n^v p^e)$,
Assim,

$$\lim_{n\to\infty} \frac{\Delta^*}{E[X]} = 0$$

como queríamos.

□

1.2.1 Notação

Alguns comentários breves sobre nossa notação:

- Dado um evento E, se temos que $\lim_{n\to\infty}[P(E \text{ ocorre})] = 1$, dizemos que esse evento ocorre assintoticamente quase certamente (a.q.c.)

  Evitamos usar esse termo na nossa primeira prova, porque cremos que assim o texto fica mais claro (em particular, as contas com limites surgem mais naturalmente). No restante do texto, porém, usaremos essa notação.

  Em particular, uma função limiar para uma propriedade P será uma $f(n)$ tal que $p(n) \gg f(n)$ implica que P ocorre a.q.c, e $p(n) \ll f(n)$, que a não ocorrência de P é assintoticamente quase certa.

- Quando dizemos que $H \subset G$, H e G grafos, podemos querer dizer duas coisas distintas (mas muito semelhantes). Ou de fato temos que $V(H) \subset V(G)$ e $E(H) \subset E(G)$, ou temos J isomorfo a H tal que $V(J) \subset V(G)$ e $E(J) \subset E(G)$

- De forma estrita, $G(n,p)$ não é um grafo, mas sim um espaço amostral. Quando falamos de $G(n,p)$ no texto, mais adequado seria dizer "um elemento de $G(n,p)$". Fazemos um abuso de notação, mas um abuso de notação muito usual na literatura.

1.3 Generalizando o teorema

Será que há uma função limiar que se aplique a um grafo não balanceado H ? Como discutimos na sessão anterior, certamente não poderá ser $f(n) = (1/n)^{1/\rho(H)}$. Na verdade, sabemos que não pode ser menos que $f(n) = (1/n)^{1/m(H)}$, onde $m(H) = \max_{L\subset H}\rho(L)$.

Veremos que essa é, de fato, uma função limiar:

Teorema 5. *Se H é um grafo (qualquer) $f(n) = (1/n)^{1/m(H)}$ é uma função limiar para a propriedade de $G(n,p)$ conter H*

Dedicaremos essa sessão à prova desse fato. Cumpre enfatizar, porém, que essa generalização não foi trivial. De fato, houve 20 anos entre a prova do teorema para grafos balanceados e a prova do caso geral.

1.3.1 Uma prova construtiva

Dado um grafo não balanceado H, construiremos um grafo balanceado F, que contem H, e tal que $\rho(F) = m(F) = m(H)$.

Pelo teorema (4), e tomando $p(n) \gg (1/n)^{1/\rho(F)} = (1/n)^{1/m(H)}$, obtemos que $G(n,p)$ contem F a.q.c.. Assim, $G(n,p)$ contem H a.q.c. . Já estabelecemos na sessão anterior, que $p(n) \ll (1/n)^{1/m(H)}$ implica que $G(n,p)$ a.q.c. não contém H. Assim, a existência de tal F será suficiente para provar que $(1/n)^{1/m(H)}$ é função limiar para $G(n,p)$ conter H.

Essa abordagem é devida a [EG85], e apresentamos aqui um detalhamento da prova do artigo.

Provemos, então, o teorema:

Teorema 6. *Para todo grafo H, existe um grafo F tal que $H \subset F$, F é balanceado e $\rho(F) = m(F) = m(H)$*

Vamos dividir a prova em 3 casos;
Se $m(H) < 1$

Demonstração. Todo ciclo tem densidade 1. Assim, H não contém ciclos, i.e. H é uma floresta.

Primeiro, notemos que, se temos duas árvores A_1 e A_2, e A_1 tem maior número de vértices, ela é a mais densa. (pois $v_1 > v_2$ implica $\frac{v_1-1}{v_1} > \frac{v_2-1}{v_2}$). Diremos que uma árvore é maior que outra quando tiver mais vértices.

Vamos provar agora que a densidade de toda floresta é menor ou igual à densidade de sua maior árvore:

Demonstração. A prova será por indução no número de árvores da floresta.

Se a floresta tem uma só árvore, não há o que provar.

Se a floresta C tem mais de uma árvore, separemos a floresta em A, uma árvore com $|A|$ máximo e B, o restante da floresta. Tomando A_b a arvore de tamanho máximo em B, temos, pela H.I.,

$\rho(B) \leq \rho(A_b) \leq \rho(A)$

Assim, $e_A/v_A \geq e_B/v_B$ (ou, de uma forma mais conveniente: $e_A \frac{v_B}{v_A} \geq e_B$)

Portanto, $\rho(A) = \frac{e_A}{v_A} = \frac{e_A + e_A \frac{v_B}{v_A}}{v_A + v_A \frac{v_B}{v_A}} \geq \frac{e_A+e_B}{v_A+v_B} = \rho(C)$

□

Assim, é imediato que $m(H) = \rho(A)$, onde A é a maior árvore de H.

É fácil completar as outras árvores de H para que tenham o mesmo número de vértices que A, obtendo F. Todo subgrafo F' de F é uma floresta, com árvore máxima menor (ou igual) a A, e, portanto, com $\rho(F') \leq \rho(A) = \rho(F)$

□

O caso $m(H) = 1$ também é simples. Não faremos uma prova completa.

$m(H) = 1$ implica que H contém um ciclo, e que H não tem nenhuma componente com dois ciclos. Para gerar F, adicionaremos a cada componente sem ciclo um ciclo (adicionando vértices e/ou arestas). $\rho(F) = 1$, e $\rho(F') \leq 1$ para todo F' contido em F.

Agora, chegamos ao caso difícil, $m > 1$.

Demonstração. Lembremos que temos H, desbalanceado, e nosso objetivo é obter um F que contem H e tal que $\rho(F) = m(F) = m(H)$. Chamaremos $m(H)$ de m

Inicialmente, notemos que, se $\rho(H_1) = \rho(H_2) = m$, então $\rho(H_1 \cup H_2) = m$

Demonstração. $|E(H_1 \cup H_2)| = |(E(H_1)| + |(E(H_2)| - |E(H_1 \cap H_2)|$

Mas $|(E(H_1)| = m|V(H_1)|$, $|(E(H_2)| = m|V(H_2)|$ e $|E(H_1 \cap H_2)| \leq m|V(H_1 \cap H_2)|$

Assim, $|(E(H_1 \cup H_2)| \geq m|V(H_1)| + m|V(H_2)| - m|V(H_1 \cap H_2)| = m|V(H_1 \cup H_2)|$.

Temos $\frac{|(E(H_1 \cup H_2)|}{|(V(H_1 \cup H_2)|} \geq m$, e já tínhamos (da definição de m) $\frac{|(E(H_1 \cup H_2)|}{|(V(H_1 \cup H_2)|} \leq m$. □

Assim, tomando todo $H_i \subset H$ tal que $\rho(H_i) = m$, definimos $\bar{H} = \cup H_i$ e notamos que $\rho(\bar{H}) = m$

Agora, desejamos construir uma sequencia de grafos F_i tal que:

- $H = F_0$
- $F_{i-1} \subset F_i$
- $m(F_{i-1}) = m(F_i)$
- $V(F_i) - V(\bar{F}_i) \subsetneq V(F_{i-1}) - V(\bar{F}_{i-1})$

(onde $\bar{F}_i$ é definido analogamente a $\bar{H}$)

Basicamente, estamos obtendo grafos que contêm H, com $m(F_i) = m(H)$ e estamos diminuindo o tamanho do conjunto de vértices de F_i fora de $\bar{F}_i$, até que $\bar{F}_i = F_i$, i.e., $\rho(F_i) = m$. Tal F_i será nosso F. No que se segue, faremos a construção de F_1, pois as seguintes são completamente análogas, e isso nos permite usar notação menos carregada.

Definamos defeito de $L \subset H$ como:

$\epsilon(L) = m|V(L)| - |E(L)|$

Temos que:

a. $\epsilon(\bar{H}) = 0$

b. $\epsilon(L_1 \cup L_2) = \epsilon(L_1) + \epsilon(L_2) - \epsilon(L_1 \cap L_2)$ (basta expandir para verificar)

c. $\epsilon(L)$ é positivo para todo $L \subset H$

d. se $\epsilon(L)$ é positivo, $\rho(L) \leq m$

Tomemos um L subgrafo de H tal que $L \not\subset \bar{H}$ tal que $\epsilon(L)$ é minimo. Podemos escolher L de forma que $\bar{H} \subset L$ (pois $\epsilon(L \cup \bar{H}) = \epsilon(L) + 0 - \epsilon(L \cap \bar{H})$) Chamaremos tal L de L^* e $\epsilon(L^*)$ de ϵ^*

Se $\epsilon^* \geq 1$, podemos adicionar a H uma aresta e que tenha uma ponta fora de $\bar{H}$, sem que $m(H + e) > m(H)$: Claramente não criaremos L com $\epsilon(L) < 0$ (i.e. , não criamos subgrafo com densidade muito grande) e, pela definição de $\bar{H}$, tem de haver um par de vértices disponível para adicionar essa aresta[3]. Assim, assumiremos que $\epsilon^* < 1$

[3]Se um vértice v fora de $\bar{H}$ tivesse aresta com todo membro de $\bar{H}$, $\rho(v + \bar{H}) \geq \rho(\bar{H})$

Tomemos k = $\lfloor \frac{1-\epsilon^*}{m-1} \rfloor + 1 = \frac{1-\epsilon^*}{m-1} + \delta$ (com $0 < \delta \leq 1$). Tomemos $m = p/q$, p e q primos entre si (a definição de m claramente implica que ele é racional). Temos que $\epsilon^* = r/q$ (pela definição de ϵ esse denominador é de fato possível)

Tomemos, ainda, inteiros M e N, tais que:

$mN - M = k(1-m) - \epsilon^* + m$

Podemos ver que eles existem usando a equação equivalente: $pN - qM = k(q-p) - r + p$, pois p e q são primos entre si e o lado direito é inteiro. Assim, podemos aplicar o teorema de Bézout. De fato, há infinitos M e N positivos, de forma que podemos escolher M e N com algumas propriedades convenientes:

i $N > 2m + 4$

ii $|\lfloor M/N \rfloor - m| \leq 1$
isso é possível pois $mN - M = J$ (onde $J = k(1-m) - \epsilon^* + m$, uma constante positiva) e, assim, $m - \frac{J}{N} = \frac{M}{N}$, com N tão grande quanto se queira.

iii $m > M/N$ (de fato, $m = M/N + J/N$)

iv $M/N > 1$ (pelo mesmo argumento de ii, usando $m > 1$)

Vamos agora construir um grafo auxiliar B, com M arestas e N vértices. Os vértices serão os números de 1 a N. Vamos definir as arestas: Tomando $s = \lfloor M/N \rfloor$, todo vértice x será adjacente a $x-1, x-2 \ldots x-s$ e a $x+1, x+2 \ldots x+s$. Faltam f arestas, onde f é o resto da divisão de M por N. Essas serão escolhidas da seguinte forma: Quando

$$\lfloor x\frac{f}{N} \rfloor > \lfloor (x-1)\frac{f}{N} \rfloor, \tag{6}$$

adicionaremos uma aresta entre x e $x - s - 1$. (note que todas as somas e subtrações aqui são mod N)

Não escolhemos a mesma aresta duas vezes, pois $m > M/N > s$ e N é grande (ver i, iii)

Vamos agora juntar B a H. Adicionemos o caminho $x_0 x_1 \ldots x_k$ com $x_0 \in B$ e $x_k \in L^* - \bar{H}$, e $\{x_1 \ldots x_{k-1}\} \cap (V(B) \cup V(H)) = \emptyset$

Terminamos nossa construção de F_1. Agora resta provar que ele é como queremos.

Tomemos $F^\star$ o subgrafo de F_1 induzido por $V(L^*) \cup V(B) \cup \{x_1 \ldots x_{k-1}\}$. Ele tem defeito 0, ou seja, densidade m

$$\begin{aligned} e(F') &= m(|V(L^*)| + |V(B)| + k - 1) - (|E(L^*)| + k + |E(B)|) \\ &= (m|V(L^*)| - |E(L^*)|) + m|V(B)| - E(B) + mk - m - k \\ &= \epsilon^* + (mN - M) + mk - m - k \\ &= \epsilon^* + [k(1-m) - \epsilon^* + m] + mk - m - k \\ &= 0 \end{aligned}$$

Assim, $F^\star \subset \bar{F_1}$, e portanto, $V(F_1) - V(\bar{F_1}) \subsetneq V(F_0) - V(\bar{F_0})$ (onde, $F_0 = H$).

Só precisamos provar que $m(F_1) = m(F_0 = H)$ e teremos terminado. Faremos isso provando que todo $F' \subset F_1$ tem $\epsilon(F') \geq 0$ (vide d.). Na verdade, basta provar esse fato para todo F' conexo, pois, se F' tem duas partes desconexas ${F'}_1$ e ${F'}_2$, temos $\epsilon(F') = \epsilon({F'}_1) + \epsilon({F'}_2) - \epsilon({F'}_1 \cap {F'}_2) = \epsilon({F'}_1) + \epsilon({F'}_2)$

Se $F' \subset H$, $\epsilon(F') \geq 0$ obviamente.
Se $F' = B$,

$$\begin{aligned}\epsilon(F') &= \epsilon(B) = mN - M \\ &= k(1-m) - \epsilon^* + m \\ &= (1-m)\frac{1-\epsilon^*}{m-1} + (1-m)\delta - \epsilon^* + m \\ &= -1 + \epsilon^* + \delta - m\delta - \epsilon^* + m \\ &= -1 + \delta + m(-\delta + 1) \\ &\geq -1 + \delta - \delta + 1 = 0 \text{ (usando } m > 1, \delta \leq 1)\end{aligned}$$

Se $F' \subsetneq B$, teremos que estudar dois casos. Para isso, definimos arco, um subgrafo de B induzido por vértices consecutivos (mod N). Seja c o comprimento do maior arco em F'. Note que de fato só estamos interessados em F' induzidos.

Um vértice de B pode ter grau $2s$ (se não está contido em nenhuma das últimas f arestas que colocamos), $2s+1$ (se está contido em uma delas) ou $2s+2$ (se está contido em duas). Vamos chamar as f ultimas arestas de "arestas do tipo 2", e as anteriores, "arestas do tipo 1".

Se $c < s$:

Todo vértice x em F' está num arco de comprimento c. Nos extremos desse arco, há dois vértices x_a e x_b, adjacentes a x em B, que não pertencem a F'. Assim, o grau máximo de um vértice em F' é $2s$

Um vértice i no extremo de um arco (i.e., i com $i, i+1 \ldots i+c-1 \in F'$ e $i-1 \notin F'$) tem grau máximo $2s-1$. Provemos esse fato: Existe $j \in W = \{i-2 \ldots i-s-1\}$ que não está em F' (se todo j em W pertencesse a F', teríamos que W é um arco de tamanho maior do que o permitido). Assim, se $\delta_B(i) = 2s+2$, $\delta_{F'}(x_i) = 2s-1$. Mas se $\delta_B(x_i) \leq 2s+1$, já tínhamos o fato.

Assim, temos $E(F') \leq s|V(F')| - 1$, e portanto
$\epsilon(F') \geq m|V(F')| - s|V(F')| + 1 \geq 1$ (pois $m \geq M/N \geq s$)

Se $c \geq s$:

Vamos tomar F' com $\epsilon(F')$ mínimo e, dentre esses, F' com c máximo. Suponhamos que $x, x+1, \ldots x+c-1$ induzem um arco máximo A, e que $x-i \in V(F')$ $(1 < i \leq s+1)$. Vamos criar F'' adicionando o vértice $x-1$ a F', e as arestas de $x-1$ a $x-i, x, x+1 \ldots x+s-1$. Adicionamos $s+1$ arestas e um vértice, assim, $\epsilon(F'') = \epsilon(F') + m - s - 1 \leq \epsilon(F')$ (ver ii e a definição de s). Assim, a presença de $x-i$ gerou uma contradição com a escolha de F'. Portanto, para $1 < i \leq s+1$, $x-i$ e (analogamente) $x+c-1+i$ não estão em F'. Pela definição de B, A é uma componente conexa de F'. Lembrando que pudemos assumir F' conexo, $F' = A$. Assim, temos:

$$\epsilon(F') = \epsilon(A) \geq mc - [(c-1) + (c-2) + \cdots + (c-s) + \lfloor (x+c-1)\frac{f}{N} \rfloor - \lfloor x\frac{f}{N} \rfloor]$$

$(c-1) + (c-2) + \cdots + (c-s)$ é a quantidade de arestas do tipo 1 em F'. Para verificar esse fato, vamos somar, para cada vértice v_i, as arestas a sua esquerda. Há $(c-1)$ v_i com uma aresta a esquerda, $(c-2)$ v_i com duas, $(c-s)$ com s e nenhum com mais de s.

$\lfloor (x+c-1)\frac{f}{N} \rfloor - \lfloor (x)\frac{f}{N} \rfloor$ é o motivo pelo qual dissemos $\geq$ e não $=$. É um limite superior para a quantidade de arestas do tipo 2 em F'. Esse é o número de mudanças de valor de $\lfloor a\frac{f}{N} \rfloor$ no intervalo $[x+c-1, x]$, e, portanto, o número de vértices que respeitam (6) em A. Cada

vértice desses tem uma aresta do tipo 2 associada, mas é possível que a outra ponta dessa aresta não esteja em A, por isso um limite superior.

Assim, temos:

$$\begin{aligned}\epsilon(F') &\geq mc - [(c-1)+(c-2)+\cdots+(c-s)+\lfloor(x+c-1)\frac{f}{N}\rfloor - \lfloor x\frac{f}{N}\rfloor] \\ &\geq mc - cs + \frac{s(s+1)}{2} - [(x+c-1)\frac{f}{N}] + x\frac{f}{N} - 1 \\ &= c(m-s-\frac{f}{N}) + \frac{s(s+1)}{2} + \frac{f}{N} - 1 \\ &\geq m - s - \frac{f}{N} + s + \frac{f}{N} - 1 = m - 1 \geq 0\end{aligned}$$

(observando que $m - s - \frac{f}{N} = m - \frac{M}{N} \geq 0$, e que $m > 1$)

Terminamos, então, os F' em B. Falta ainda provar que $\epsilon(F') \geq 0$, se $F' \not\subset B$ e $F' \not\subset H$. Como F' é conexo, essas restrições resultam que F' contém algum x_i (lembremos, x_i são os vértices que colocamos para formar o caminho de B a H)

Podemos assumir que F' tem mais de um vértice (caso contrário, $\epsilon(F') = m$). Assim, Se F' contém x_i com $\delta(x_i) = 1$. podemos retirar x_i de F', obtendo F'', que tem menor defeito (dado que $m > 1$ e aplicando a definição de defeito).

Assim, F' contém $x_0, x_1 \ldots x_k$

Tomemos, então, $H_0 = H \cap F'$ ($\neq \emptyset$, pois $x_k \in H$) e $B_0 = B \cap F'$ ($\neq \emptyset$, pois $x_0 \in B$). Sabemos que $\epsilon(H_0) \geq \epsilon^*$ (pois $x_k \notin \bar{H}$, e usando a definição de ϵ^*). Temos, então

$$\begin{aligned}\epsilon(F') &= m(|V(H_0)| + |V(B_0)| + k - 1) - |E(H_0)| - |E(B_0)| - k \\ &= m|V(H_0)| - |E(H_0)| + m|V(B_0)| - |E(B_0)| + m(k-1) - k \\ &\geq \epsilon^* + \epsilon(B_0) + m(k-1) - k\end{aligned}$$

Notando que B_0 é conexo (pois assumimos que F' é), podemos usar os fatos anteriores sobre $\epsilon(B_0)$. Definindo c, o comprimento do maior arco de B_0, tínhamos:

Se $c < s$, $\epsilon(B_0) \geq 1$. Assim,

$$\epsilon(F') \geq \epsilon^* + 1 + m(k-1) - k = \epsilon^* + (m-1)(k-1) \geq \epsilon^* \geq 0$$

(lembrando que $m > 1$, e examinando a definição de k, que implica $k \geq 1$)

Se $c \geq s$, $\epsilon(B_0) \geq m - 1$. Assim,

$$\epsilon(F') \geq \epsilon^* + m - 1 + m(k-1) - k = \epsilon^* + (m-1)k - 1$$

Mas $k \geq \frac{1-\epsilon^*}{m-1}$. Assim,

$$\epsilon(F') \geq \epsilon^* + 1 - \epsilon^* - 1 = 0$$

Provamos, então, que não há subgrafo F' de F_1 com defeito negativo, ou seja, com $\rho(F') > m$. F_1 é como prometemos, e finalizamos a prova.

□

1.3.2 Uma prova probabilística

Cremos que a prova construtiva é bastante interessante e instrutiva. Ela é, porém, bastante trabalhosa e cheia de detalhes. Faremos agora uma prova do teorema 5 usando ferramentas probabilísticas.

Precisaremos do seguinte teorema:

Teorema 7. *Seja H um grafo qualquer. H ocorre a.q.c. em $G(n,p)$ se tivermos que, para todo F_l subgrafo de H, $E[X_l] \to \infty$, onde X_l é o número de cópias de F_l em $G(n,p)$.*

Demonstração. Usaremos a variável X, o número de ocorrências de H, e o teorema 3.

Basta verificar que $\Delta^* = o(E(X))$. (as outras verificações são como no segundo fato do teorema 4)

Lembremos o que isso quer dizer. Tomando K_i a i-ésima cópia possível de H em $G(n,p)$, o evento A_i é o evento de $K_i \subset G(n,p)$

Temos, então

$$\Delta^* = \sum_{j/j\sim i} P[A_j|A_i]$$

Lembremos também que $j \sim i$ se A_i e A_j não são independentes. No nosso caso, isso quer dizer que K_i e K_j têm aresta em comum.

Tomemos

$$\Delta^*_{F_l} = \sum_{j/K_i\cap K_j = F_l} P[A_j|A_i]$$

Note que $K_i \cap K_j$ é um grafo. Temos

$$\Delta^* = \sum_{F_l/e(H)>|E(F_l)|\geq 1} \Delta^*_{F_l}$$

Vamos, então, tentar limitar superiormente $\Delta^*_{F_l}$

Supondo que $|V(F_l)| = a$, $|E(F_l)| = b$, $|V(H)| = v$ e $|E(H)| = e$, temos

$$\Delta^*_{F_l} \leq \binom{v}{a}\binom{n-v}{v-a} v! p^{e-b}$$

Assim,

$$\Delta^*_{F_l} = O(n^{v-a}p^{e-b}) = O\left(\frac{n^v p^e}{n^a p^b}\right)$$

Mas $E(X) = \Theta(n^v p^e)$ e $E(X_l) = \Theta(n^a p^b)$. Assim,

$$\Delta^*_{F_l} = O\left(\frac{E(X)}{E(X_l)}\right)$$

$$\frac{\Delta^*}{E(X)} = O\left(\frac{\sum_{F_l/e(H)>|E(F_l)|\geq 1} \frac{E(X)}{E(X_l)}}{E(X)}\right) = O\left(\sum_{F_l/e(H)>|E(F_l)|\geq 1} \frac{1}{E(X_l)}\right)$$

Como há um número finito de F_l, e $E(F_l) \to \infty$ para todo F_l, temos que $\lim_{n\to\infty} \sum_{F_l/e(H)>|E(X_l)|\geq 1} \frac{1}{E(X_l)} = 0$, e assim, $\Delta^* = o(E(X))$, o que conclui a prova. □

Provemos agora, novamente, o teorema 5.

Demonstração. Tomemos $f(n) = (1/n)^{1/m(H)}$.

Já sabemos que, se $p(n) \ll f(n)$, H não ocorre em $G(n,p)$ a.q.c.

Quando $p(n) \gg f(n)$, temos que, para todo F_l subgrafo de H, $E(X_l) \to \infty$

Assim, pelo teorema 7, H ocorre a.q.c. em $G(n,p)$ □

1.4 No limiar

A definição de função limiar não é muito intuitiva. Parece natural que nos perguntemos se não é possível fazer melhor. Em particular, se não podemos determinar o comportamento de $G(n,p)$ em relação a conter H quando $p(n) = \Theta(f(n))$.

No que se segue, diremos que a função limiar descreve bem o comportamento de $G(n,p)$, no sentido que, se $p(n) = \Theta(f(n))$,

$$\lim_{n\to\infty} P(H \subset G(n,p)) \notin \{0,1\}$$

1.4.1 Limitando a probabilidade

De início, precisamos definir o seguinte:

Definição 3. $\phi_H = min\{E(X_K) : K \subseteq H, e_k > 0\}$

Onde X_K é como definido anteriormente, ou seja, a esperança do número de ocorrências de K em $G(n,p)$ (por enquanto, não estamos tomando nenhum limite, ou mandando n a infinito)

Notemos que, na nossa prova do teorema 7, ϕ_H teve um papel essencial. De fato, o que obtivemos foi:

$$P(X=0) \leq \frac{1}{E(X)} + \sum_{F_l/e(H)>|E(F_l)|\geq 1} \frac{1}{E(X_l)}$$

(para verificar isso, veja a prova dos teoremas 7 e 3)

Obtivemos que o lado direito ia a 0, usando que todas as esperanças envolvidas iam a ∞. Mas, de fato, temos:

$$P(X=0) \leq C \cdot \max\{\frac{1}{E(X_l)} : K_l \subseteq H\} = C\frac{1}{\phi_h}$$

(onde C é o número de subgrafos K_l de H com $e(K_l) \geq 1$, incluindo H)

Observando a prova do primeiro fato do teorema 4 (e o comentário posterior) notamos que nosso raciocínio foi

$$P(X>0) \leq P(X_l > 0) \leq E(X_l)$$

de onde temos

$$P(X \neq 0) = P(X>0) \leq \phi_H \text{ e portanto, } P(X=0) \geq 1-\phi_H$$

Até agora, obtivemos, então, que:

$$1-\phi_H \leq P(X=0) \leq C\frac{1}{\phi_h}$$

Ocorre, porém, que é possível fazer uma restrição bem melhor de $P(X=0)$:

Teorema 8.

$$e^{\frac{-\phi_H}{1-p(n)}} \leq P(X=0) \leq e^{-\Theta(\phi_H)}$$

Isso nos interessa porque, se $\phi_H = \Theta(1)$ (e $p(n)$ for a zero), $\lim_{n\to\infty} P(X = 0)$ não será 0 nem 1. A verificação desse fato é trivial.

Verifiquemos, então, que:

Fato. *Se* $p(n) = \Theta((\frac{1}{n})^{\frac{1}{m(H)}})$, $\phi_H = \Theta(1)$.

Demonstração. Tomemos K subgrafo de H com $\rho(K) = m(H)$

Pela definição de ϕ, temos

$$\phi_H \leq E(K) = \Theta(n^{v(K)}p^{e(K)}) = \Theta[(np^{m(H)})^{v(K)}]$$

Também temos que, para algum $J \subset H$

$$\phi_H = E(J) = \Theta(n^{v(J)}p^{e(J)}) = \Theta[(np^{\rho(J)})^{v(J)}] \geq \Theta[(np^{m(H)})^{v(J)}]$$

Não afirmamos que J é o mesmo para todo n. Mesmo que J varie com n, temos que $\Theta[(np^{m(H)})^{v(J)}] = \Theta(1)$, se $p(n) = \Theta((\frac{1}{n})^{\frac{1}{m(H)}})$.

Também temos que $\Theta[(np^{m(H)})^{v(K)}] = \Theta(1)$

Assim, $\phi_H = \Theta(1)$.

□

Dessa forma, temos que se $p(n) = \Theta((\frac{1}{n})^{\frac{1}{m(H)}})$, H não ocorre a.q.c., nem deixa de ocorrer a.q.c.

1.4.2 Provando as desigualdades exponenciais

Apresentamos uma prova do teorema (8), baseada na de [JS].

Começamos provando que $e^{\frac{-\phi_H}{1-p(n)}} \leq P(X = 0)$. Para isso, será necessário o seguinte teorema:

Teorema 9. *(desigualdade FKG) Seja f uma função de $[0,1]^n$ para $\mathbb{R}$. Dizemos que uma tal função f é crescente se $A \subset B$ implica $f(A) \leq f(B)$, e que uma tal função é decrescente se $A \subset B$ implica $f(A) \geq f(B)$.*

Tomemos $Y = (y_1, y_2 \dots y_n)$, uma string de variáveis aleatórias y_i, que têm valor 1 com probabilidade p_i e zero com probabilidade $(1 - p_i)$, e f e g duas funções crescentes, ou f e g duas funções decrescentes.

Temos, então: $E(f(Y)g(Y)) \geq E(f(Y))E(g(Y))$.

Uma prova pode ser encontrada em [eDRS].

Podemos ver os Y descritos acima como subconjuntos aleatórios de $[1, n]$. É frequente que, dado um tal subconjunto aleatório, desejemos determinar se ele contém algum conjunto de uma determinada família. Por exemplo, usaremos esse Y como um conjunto aleatório de arestas, e nos preocuparemos com quando ele contém uma determinada cópia de H, nosso grafo de interesse.

Para formalizar essa noção, consideremos S um subconjunto de $2^{[1,n]}$, e para cada $A \in S$, vamos definir uma variável indicadora I_A, que é 1 se Y é 1 em todos os elementos de A, e 0 caso contrário, e definamos $I_S = \sum I_A$.

Temos, então:

Fato.

$$P(I_S = 0) \geq e^{-\frac{E(I_S)}{1-\max p_i}}$$

Demonstração. A propriedade P_a, de Y não ser 1 em todos os elementos de A, é claramente decrescente.[4] Assim, temos, da desigualdade FKG:

$$E[\prod_{A\in S} P_a] \geq \prod_{A\in S} E[P_a]$$

Mas $E[\prod_{A\in S} P_a] = P(I_S = 0)$ e $E[P_a] = 1 - E[I_A]$. Assim, temos

$$P(I_S = 0) \geq \prod_{A\in S} 1 - E[I_A]$$

Todos os $E[I_A]$ são ≤ 1. Mas, com $x \leq 1$, temos que $1 - x \geq e^{-x/(1-x)}$. Assim

$$P(I_S = 0) \geq \prod_{A\in S} e^{-\frac{E[I_A]}{1-E[I_A]}} \geq \prod_{A\in S} e^{-\frac{E[I_A]}{1-\max_{A\in S} E[I_A]}} = e^{-\frac{E[I_S]}{1-\max_{A\in S} E[I_A]}} \geq e^{-\frac{E[I_S]}{1-\max_i p_i}}$$

□

Usando esse resultado, provaremos $e^{\frac{-\phi_H}{1-p(n)}} \leq P(X = 0)$, a desigualdade que queremos:

Seja H_m um subgrafo de H tal que $E(H_m)$ é minimo (ou seja, igual a $E(H_m) = \phi_H$).

Tomemos n como o número de arestas máximo de $G(n, p)$ (ou seja, $\binom{n}{2}$) e os elementos $A \in S$ como conjuntos de arestas de cópias de H_m

Temos que $P(X = 0) = P(H \text{ não ocorre}) \geq P(H_m \text{ não ocorre}) = P(I_S = 0) \geq e^{-\frac{E[I_S]}{1-\max_i p_i}} = e^{-\frac{\phi_h}{1-p}}$, como queríamos.

Provemos, agora, $P(X = 0) \leq e^{-\Theta(\phi_H)}$. Para isso, utilizaremos o seguinte teorema, cuja prova pode ser encontrada em [JS]:

Teorema 10. *Com $I_S = \sum I_A$, como definido acima, temos*

$$P(I_S = 0) \leq e^{\frac{E(I_S)^2}{\sum\sum_{A,B\in S, A\cap B\neq\emptyset} E(I_A I_B)}}$$

Novamente, tomamos n como o número de arestas máximo de $G(n, p)$, mas, desta vez, e os elementos $A \in S$ serão conjuntos de arestas de cópias de H.

Dois conjuntos A e B de S tem intersecção não nula somente se representam cópias de H com arestas em comum. Tomemos L o menor subgrafo de H isomorfo a esse conjunto de arestas. Então, temos:

$$\sum_{A,B\in S, A\cap B\neq\emptyset} E(I_A I_B) = \sum_{L\subset G, e_l\geq 1} \sum_{G} \sum_{G'\cap G} p^{2e_h - e_l}$$

Ou seja, para toda intersecção L possível, tomamos todos os pares de grafos isomorfos a H com intersecção L (aos quais chamamos G e G') e somamos a probabilidade de esses pares ocorrerem em $G(n, p)$.

Mas,

$$\sum_{L\subset G, e_h\geq 1} \sum_{G} \sum_{G'\cap G} p^{2e_h - e_l} = \sum_{L\subset G, e_h\geq 1} \Theta(n^{2v_h - v_l} p^{2e_h - e_l}) = \sum_{L\subset G, e_h\geq 1} \Theta(E(H)^2/E(L))$$

Tomando k o número de subgrafos de H, temos $(E(H)^2/\phi_H) \leq \sum_{L\subset G, e_h\geq 1}(E(H)^2/E(L)) \leq k(E(H)^2/\phi_H)$. Como k independe de n, segue que $\sum_{L\subset G, e_h\geq 1} \Theta(E(H)^2/E(L)) = \Theta(E(H)^2/\phi_H)$.

[4] Descrevemos uma propriedade como uma função, que é 1 quando a propriedade é respeitada e 0 caso contrário

Assim, usando $E(I_S) = E(H)$, temos:

$$e^{\frac{E(I_S)^2}{\sum\sum_{A,B\in S, A\cap B\neq\emptyset} E(I_A I_B)}} = e^{\Theta(\Phi_H)}$$

que é o que necessitávamos provar.

2 Teorema de Shannon

Nosso segundo problema é o seguinte:

Como transmitir dados com confiabilidade alta, usando um canal físico que tem confiabilidade baixa ?

Podemos ter, por exemplo, um sistema de envio de dados digitais via radio, que, por suas limitações físicas, apresenta uma chance de 0.1 de inverter cada um dos bits enviados. Nós podemos melhorar esse desempenho, sem alterar as características físicas do sistema, usando uma codificação antes do envio e uma decodificação após o recebimento.

2.1 Uma solução "ingênua"

Uma abordagem possível é simplesmente repetir os bits.

Utilizemos como codificação repetir cada bit 3 vezes, e como decodificação, tomar o bit mais frequente em cada tripla.

Se enviamos o bit 1 por nosso canal, qual a chance de ele ser recebido como 1 ? O decodificador aceita como 1 as strings em

$$\{(1,1,1),(1,1,0),(1,0,1),(0,1,1)\}$$

A primeira dessas strings ocorre com probabilidade $0.9^3 = 0.729$, e as demais, com $p = 0.9^2 \cdot 0.1 = 0.081$. Assim, nossa probabilidade de acerto subiu para 0.972

De fato, podemos usar essa ideia para reduzir a probabilidade de erro o quanto queiramos. Tornemos essa noção mais precisa, e provemos esse fato:

Definição 4. *Um canal binário simétrico é um canal para transmissão de dados digitais que inverte um bit com uma dada probabilidade p.*

Aplicaremos esse conceito com a seguinte notação: Seja x uma string binária, $y = C(x)$ é uma string aleatória com

$$P(y_i = 1|x_i = 1) = P(y_i = 0|x_i = 0) = 1 - p$$
$$P(y_i = 1|x_i = 0) = P(y_i = 0|x_i = 1) = p$$

Definição 5. *O código de repetição R_n é um par de funções: $F : \{0,1\} \to \{0,1\}^n$ e $G : \{0,1\}^n \to \{0,1\}$.*

$F(1)$ é a string de n 1s, $F(0)$ é a string de n zeros.

$G(A) = 1$ se A é uma string com mais uns do que zeros, e 0 caso contrário.

Fato. *Tomando um canal binário simétrico, com $p < \frac{1}{2}$ temos que*

$$\lim_{n\to\infty} P[G(C(F(x_i))) \neq x_i] = 0,$$

com F e G da definição de R_n, e i uma posição qualquer do vetor de bits transmitido

Demonstração. Tomemos o código R_n, X_i o número de bits invertidos em $C(F(x_i))$. Temos:

$$P[G(C(F(x_i))) \neq x_i] = P[X_i \geq n/2] \leq P[|X_i - np| \geq \frac{n}{2} - np]$$

Como $E(X_i) = np$, pela desigualdade de Chebychev,

$$P[|X_i - np| \geq \frac{n}{2} - np] \leq \frac{\sigma^2_{X_i}}{(np - \frac{n}{2})^2} = \frac{n[p(1-p)]}{n^2(p - \frac{1}{2})^2}$$

Como $\lim_{n\to\infty} \frac{n[p(1-p)]}{n^2(p-\frac{1}{2})^2} = 0$, provamos o fato. □

Notemos que a hipótese de $p < 1/2$ é realmente necessária. Na prova, haveria uma divisão por zero se isso não fosse verdade. Isso acontece porque,de fato, um canal de comunicação com $p = 1/2$ não transmite informação nenhuma.

Para termos uma noção intuitiva desse fato, basta notar que um canal que ignora a mensagem e transmite uma string aleatória uniforme é um canal binário simétrico com $p = 1/2$.

2.1.1 Por que ingênua ?

Na solução apresentada, conseguimos aumentar $P[G(C(F(x_i))) = x_i]$ o quanto quisemos, mas isso teve um custo: Conforme $P[G(C(F(x_i))) = x_i]$ ia a um, a velocidade de transmissão dos dados ia a zero.

O resultado que desejamos provar é que podemos fazer melhor do que isso: É possível, usando um código adequado, transformar um canal bastante impreciso em um quão preciso quanto se queira, e mesmo assim perder uma fração fixa (independente da precisão desejada!) da velocidade.

Vamos precisar de algumas definições para poder deixar essa afirmação mais precisa.

2.2 Definições

Sejam X e Y variáveis aleatórias discretas, que assumem, respectivamente, valores $x_1, x_2 \ldots x_n$ e $y_1, y_2 \ldots y_m$

Denotaremos $p(X = x_i)$ como $p(x_i)$, $p(Y = y_j)$ como $p(y_j)$ e analogamente para as probabilidades condicionais.

Definição 6. *Entropia de X:*

$$H(X) = \sum_{i=1}^{n} p(x_i) \log \left(\frac{1}{p(x_i)} \right)$$

Podemos interpretar a entropia como uma medida da quantidade de informação fornecida por uma variável aleatória.

Definição 7. *Entropia condicional de X dado $Y = y_j$:*

$$H(X|Y = y_j) = \sum_{i=1}^{n} p(x_i|y_j) \log \left(\frac{1}{p(x_i|y_j)} \right)$$

Definição 8. *Entropia condicional de X dado Y:*

$$\begin{aligned} H(X|Y) &= \sum_{j=1}^{m} p(y_j) H(X|y_j) \\ &= \sum_{j=1}^{m} p(y_j) \sum_{i=1}^{n} p(x_i|y_j) \log\left(\frac{1}{p(x_i|y_j)}\right) \\ &= \sum_{j=1}^{m} \sum_{i=1}^{n} p(x_i, y_j) \log\left(\frac{1}{p(x_i|y_j)}\right) \end{aligned}$$

Definição 9. *Entropia conjunta de X, Y:*

$$H(X,Y) = \sum_{j=1}^{m} \sum_{i=1}^{n} p(x_i, y_j) \log\left(\frac{1}{p(x_i, y_j)}\right)$$

Fato. $H(X,Y) = H(Y) + H(X|Y)$

Demonstração.

$$\begin{aligned} H(X|Y) + H(Y) &= \sum_{j=1}^{m} p(y_j) \sum_{i=1}^{n} p(x_i|y_j) \log\left(\frac{1}{p(x_i|y_j)}\right) + \sum_{j=1}^{m} p(y_j) \log\left(\frac{1}{p(y_j)}\right) \\ &= \sum_{j=1}^{m} p(y_j) \sum_{i=1}^{n} p(x_i|y_j) \log\left(\frac{1}{p(x_i|y_j)}\right) + \sum_{j=1}^{m} p(y_j) \sum_{i=1}^{n} p(x_i|y_j) \log\left(\frac{1}{p(y_j)}\right) \\ &= \sum_{j=1}^{m} p(y_j) \sum_{i=1}^{n} p(x_i|y_j) \log\left(\frac{1}{p(x_i|y_j)p(y_j)}\right) \\ &= \sum_{j=1}^{m} \sum_{i=1}^{n} p(x_i, y_j) \log\left(\frac{1}{p(x_i, y_j)}\right) \end{aligned}$$

□

Definição 10. *A informação mútua de X e Y:*

$$I(X,Y) = H(X) - H(X|Y)$$

Temos que $I(X,Y) = I(Y,X)$ (do fato acima) e que $I(X,Y) > 0$

Essa definição é fundamental para o que se segue. Podemos interpretar a informação mútua como uma medida de quanto uma variável aleatória revela sobre outra.

Há um porém com essas definições: precisamos excluir as probabilidade nulas dos somatórios. Para evitar carregar a notação, mencionamos esse fato aqui, e não nas definições acima.

2.2.1 Códigos de bloco

Definição 11. *Um código de bloco (N,K) é um conjunto ordenado S de $2^K = |S|$ strings em um alfabeto A_x, cada string com comprimento N.*

A taxa de um código de bloco é definida como $\frac{K}{N}$. O tamanho do código é N.

R_n, o nosso exemplo de código "ingênuo", tem $N = n$ e $|S| = 2$, assim sua taxa é $\frac{1}{n}$

Definição 12. *Um decodificador (associado a um código de bloco) é uma função das strings de comprimento N em A_y para o conjunto $S \cup f$, onde $f \notin S$*

Adicionamos esse elemento extra para indicar uma decodificação que falhou.

No nosso exemplo "ingênuo", o decodificador era uma função que associava cada string de comprimento n a seu bit mais frequente. No caso de empate, o decodificador no nosso exemplo devolve zero, mas talvez fosse mais adequado devolver f, pois é igualmente provável que o bit de entrada tenha sido zero ou um.

Definição 13. *Um canal discreto sem memória Q é caracterizado por dois alfabetos: um de entrada (A_x) e um de saída (A_y), e, para cada elemento de A_x, uma distribuição de probabilidades $p(y|x)$ sobre os elementos de A_y.*

Definiremos, como no nosso exemplo, uma variável aleatória $C(x)$, que associa a cada string no alfabeto A_x uma string no alfabeto A_y com probabilidade $p(y_1y_2 \dots y_n|x_1x_2 \dots x_n) = p(y_1|x_1)p(y_2|x_2) \dots p(y_n|x_n)$

Definição 14. *A probabilidade de falha de uma string $S_i \in S$ é*

$$P_f(S_i) = P[G(C(S_i)) \neq S_i]$$

Nosso objetivo é escolher um código (e um decodificador) que minimize P_{max}, a máxima $P_f(S_i)$.

2.3 O Teorema de Shannon[5]

Teorema 11. *Para todo canal discreto sem memória, existe um número D tal que, para todo $\epsilon > 0$ e $R < D$, tomando N grande o bastante, existe um código de bloco de tamanho N e taxa maior que R tal que $P_{max} < \epsilon$*

Apresentaremos uma prova devida a [Mac05], com algumas alterações.

Tomemos A_x e A_y os alfabetos de entrada e saída do nosso canal. Seja N o tamanho do nosso código, e $p(x)$ uma distribuição de probabilidade sobre os elementos de A_x, definindo a variável aleatória X. Sorteando $x_i \in A_x$, definimos a variável aleatória $Y = C(x_i)$ (i.e. Y é uma variável aleatória cuja distribuição depende de X).

Vale a pena enfatizar que essa distribuição $p(x)$ que estamos tomando tem como finalidade estabelecer uma distribuição nos códigos possíveis. Não estamos colocando nenhum tipo de restrição na distribuição das mensagens que posteriormente se enviará pelo canal.

Chamaremos as strings de comprimento N em A_x de $j_1, j_2 \dots j_{|A_x|^N}$, e às de comprimento N em A_y, $k_1, k_2 \dots k_{|A_y|^N}$. J será uma variável aleatória nas strings de comprimento N em A_x, com $p(j = x_1, x_2 \dots x_N) = p(x_1)p(x_2) \dots p(x_N)$. Definiremos também uma variável aleatória $K = C(J)$.

[5] O teorema que apresentamos é basicamente o mesmo de [NA08], mas generalizado para canais quaisquer. Uma diferença relevante é o fato de que não assumimos nada sobre a distribuição dos caracteres de entrada. O teorema de [NA08] pode ser adaptado para ter essa característica, bastando para isso não usar todas as strings possíveis do alfabeto de entrada, como na prova que apresentamos

Definição 15. *Diremos que j_i e k_l são **típicas em conjunto**[6] para uma tolerância β se:*

$$\left|\frac{1}{N}\log\left(\frac{1}{p(j_i)}\right) - H(X)\right| < \beta \tag{7}$$

$$\left|\frac{1}{N}\log\left(\frac{1}{p(k_l)}\right) - H(Y)\right| < \beta \tag{8}$$

$$\left|\frac{1}{N}\log\left(\frac{1}{p(j_i,k_l)}\right) - H(X,Y)\right| < \beta \tag{9}$$

Onde as probabilidades são as determinadas pelas distribuições J e K.

Fato. *A probabilidade de que j_i e k_l, sorteadas como em J e K, sejam **típicas em conjunto** vai a um conforme N vai a infinito.*

Demonstração. Vamos avaliar a esperança e a variância de $\frac{1}{N}\log\left(\frac{1}{p(j_i)}\right)$, com o intuito de aplicar a desigualdade de Chebychev.

$$E\left(\frac{1}{N}\log\left(\frac{1}{p(j_i)}\right)\right) = \frac{1}{N}E\left(\log\left(\frac{1}{p(j_i)}\right)\right) = \frac{1}{N}H(J) = \frac{1}{N}(NH(X)) = H(X)$$

Onde $H(J) = NH(X)$ vem de aplicações sucessivas de $H(A,B) = H(A) + H(B|A)$

$$\text{var}\left(\frac{1}{N}\log\left(\frac{1}{p(j_i)}\right)\right) = \frac{1}{N^2}\text{var}\left(\log\left(\frac{1}{p(j_i)}\right)\right) = \frac{1}{N^2}N\text{var}\left(\log\left(\frac{1}{p(x_i)}\right)\right) = \frac{1}{N}\sigma^2$$

(onde σ^2 independe de N)
Assim,

$$p\left(\left|\frac{1}{N}\log\left(\frac{1}{p(j_i)}\right) - H(X)\right| \geq \beta\right) \leq \frac{\frac{\sigma^2}{N}}{\beta^2}$$

que, com β fixo, vai a zero conforme N vai a infinito.

Assim, provamos que a probabilidade de (7) se verificar quando N vai a infinito vai a 1. Podemos provar esse fato para (8) e (9) de forma inteiramente análoga. Daí, temos que (7), (8) e (9) se verificam ao mesmo tempo com probabilidade que vai a 1 □

Com um N fixo, há um conjunto fixo de pares **típicos em conjunto**, ao qual chamaremos $T_{N\beta}$. Quando sorteamos um par de acordo com J e K, há uma probabilidade de que esse par caia nesse conjunto. O que acabamos de afirmar é que essa probabilidade vai a um conforme $N \to \infty$

Vamos agora limitar o tamanho de $T_{N\beta}$.

Fato. $|T_{N\beta}| \leq 2^{N(H(X,Y)+\beta)}$

[6] em inglês, jointly typical

Demonstração. Se $(j_i, k_l) \in T_{N\beta}$, temos

$$\begin{aligned}\left|\frac{1}{N}\log\left(\frac{1}{p(j_i,k_l)}\right) - H(X,Y)\right| &< \beta \\ -\frac{1}{N}\log(p(j_i,k_l)) - H(X,Y) &< \beta \\ \log(p(j_i,k_l)) &> -N(\beta + H(X,Y)) \\ p(j_i,k_l) &> 2^{-N(\beta+H(X,Y))}\end{aligned}$$

Assim, fazendo um sorteio de acordo com as distribuições J e K, temos

$$p[(j_i,k_l) \in T_{N\beta}] = \sum_{(j_a,k_b)\in T_{N\beta}} p(j_a,k_b) \geq 2^{-N(\beta+H(X,Y))}|T_{N\beta}|$$

Mas $1 \geq p[(j_i,k_l) \in T_{N\beta}]$. Assim,

$$2^{N(\beta+H(X,Y))} \geq |T_{N\beta}|$$

□

Agora, sortearemos j_i de acordo com a distribuição J, mas também outro j_g, e então k_l de acordo com $C(j_g)$. (i.e. mantivemos as probabilidades de um dado j_i ou k_l, mas os tornamos independentes)

Fato. $P((j_i,k_l) \in T_{N\beta}) \leq 2^{-N(I(X,Y)-3\beta)}$

Demonstração. O fato de j_i e k_l pertencerem a algum par em $T_{N\beta}$ já nos permite deduzir algo sobre as probabilidades de j_i e k_l ocorrerem: $|\frac{1}{N}\log\left(\frac{1}{p(j_i)}\right) - H(X)| < \beta$ implica que $p(j_i) < 2^{-N(-\beta+H(X))}$ e, analogamente, temos $p(k_l) < 2^{-N(-\beta+H(Y))}$

$$\begin{aligned}P((j_i,k_l) \in T_{N\beta})) &= \sum_{(j_i,k_l)\in T_{N\beta}} P(j_i)P(k_j) \\ &\leq |T_{N\beta}|2^{-N(-\beta+H(X))}2^{-N(-\beta+H(Y))} \\ &\leq 2^{N(H(X,Y)+\beta)}2^{-N(-\beta+H(X))}2^{-N(-\beta+H(Y))} \\ &= 2^{N(3\beta+H(X,Y)-H(X)-H(Y))}\end{aligned}$$

De $H(X,Y) = H(Y) + H(X|Y)$ e $I(X,Y) = H(X) - H(X|Y)$ temos

$$I(X,Y) = H(Y) + H(X) - H(X,Y)$$

Assim,

$$P((j_i,k_l) \in T_{N\beta})) \leq 2^{N(3\beta-I(X,Y))}$$

□

O decodificador G associará uma string k_l a uma S_i tal que S_i e k_l são **típicas em conjunto**, se houver apenas uma tal $S_i \in S$. Caso k_l não seja **típica em conjunto** com nenhuma S_i, ou seja **típica em conjunto** com S_i e S_j, $i \neq j$, k_l será associada a f.

Dizemos que houve falha na transmissão de uma string S_i se $G(C(S_i)) \neq S_i$, e chamamos a probabilidade desse evento de $p_f(S_i)$. Ocorre falha de duas maneiras distintas: Talvez $C(S_i)$ não seja **típica em conjunto** com S_i, ou talvez haja alguma S_j que seja **típica em conjunto** com $C(S_i)$.

Faremos o seguinte experimento: Utilizando $|S| = 2^{NR'}$, tomamos um código aleatório V_t, ou seja, um conjunto de S strings, com cada string escolhida independentemente de acordo com J. Escolhemos aleatoriamente, então, uma string S_i dentre as $|S|$ strings do código, e calculamos a probabilidade de ocorrer falha na transmissão dessa string.

A probabilidade de $C(S_i)$ não ser tipica em conjunto com S_i é a probabilidade de tomarmos j_i de acordo com J, e termos que $C(j_i)$ não é **típica em conjunto** com j_i. Como já estabelecemos, podemos levar essa chance a zero aumentando N.

Vamos agora estimar a chance de algum outro $S_j \in V$ ser típico em conjunto com $C(S_i)$. Fixando um tal S_j, já sabemos que essa probabilidade não é maior que $2^{-N(I(X,Y)-3\beta)}$, do terceiro fato. Como temos apenas $2^{NR'} - 1$ escolhas para j, então, a chance de haver um tal j é menor ou igual a $2^{NR'}2^{-N(I(X,Y)-3\beta)} = 2^{-N(I(X,Y)-R'-3\beta)}$

Se tivermos $I(X,Y) > R' + 3\beta$, levar N a infinito também leva essa probabilidade de erro a zero. Assim, se $I(X,Y)$ respeitar essa desigualdade, podemos tomar uma probabilidade de erro do experimento menor que δ, para qualquer δ que queiramos.

Mas, sendo V o conjunto dos códigos possíveis, P_{V_t} a probabilidade de termos selecionado o código V_t, P_{S_i} a probabilidade de termos selecionado o símbolo S_i e $P(V_t, S_i)$ a probabilidade de erro na decodificação de S_i, usando o código V_t, essa probabilidade que acabamos de estimar é

$$\sum_{V_t \in V, S_i \in V_t} P_{V_i} P_{S_i} P(V_t, S_i) = \sum_{V_t \in V} P_{V_i} \sum_{S_i \in V_t} P_{S_i} P(V_t, S_i)$$

Temos, então, $\sum_{V_t \in V} P_{V_i} \sum_{S_i \in V_t} P_{S_i} P(V_t, S_i) \leq \delta$. Assim, há um código W com

$$\sum_{S_i \in W} P_{S_i} P(W, S_i) \leq \delta.$$

Nesse código, obviamente não há mais do que metade das strings com $P(W, S_i) > 2\delta$. Retiremos essas strings. Obtemos um código W' onde $P(W', S_i) \leq 2\delta$ (toda string $S_a \in W'$ já tinha $P(W', S_a) \leq 2\delta$ em W, e a retirada de elementos de W somente fez $P(W', S_a)$ diminuir)

W' é um código com $\frac{2^{NR'}}{2}$ strings. Assim, sua taxa é $\frac{NR'-1}{N} = R' - \frac{1}{N}$

O teorema está provado! Tomemos $D = I(X,Y)$, $R < D$ $\epsilon > 0$. Podemos construir W' com P_{max} menor que ϵ, e taxa $T = R' - \frac{1}{N}$, onde a única restrição sobre R' é $R' < I(X,Y) - 3\beta$. Assim, podemos tomar T quão próxima de D quanto queiramos (aumentando N e diminuindo β). Em particular, para todo $R < D$, podemos tomar $T > R$.

$I(X,Y)$ varia conforme a distribuição $p(x)$ que adotamos, então, é conveniente tomar $p(x)$ que maximize $I(X,Y)$ e, portanto D. Chamaremos esse $I(X,Y)$ máximo de capacidade do canal.

3 Um problema geométrico

Nessa sessão, trataremos do seguinte problema geométrico: Temos um conjunto A de n pontos em um plano, e, para uma grande quantidade de regiões distintas, queremos saber (ou estimar) quantos desses pontos estão em uma determinada região. Seria desejável reduzir o número de operações necessárias para estimar o número de pontos em uma dada região.

Digamos que queremos saber quantos elementos de A estão dentro de um triângulo. Ao invés de testarmos todos os pontos de A, é possível escolher para testar um conjunto $B \subset A$, tal que, para todo triângulo T, a proporção de elementos de B em T é próxima da proporção de elementos de A em T. O resultado surpreendente que discutiremos a seguir é que o tamanho de B pode ser fixo, *independente do tamanho de* A, e mesmo assim obteremos uma boa estimativa dessa proporção.

Na verdade, podemos obter um tal B se nossas regiões forem quadriláteros. Ou hexágonos. Ou se nossos pontos estiverem em $\mathbb{R}^4$, e nossas regiões forem semi-espaços. Utilizaremos um importante conceito de combinatória, a dimensão de Vapnik e Chervonenkis (que chamaremos **dimensão VC**) para caracterizar exatamente quando será possível obter um tal B.

3.1 A dimensão VC

Vamos começar definindo exatamente o que se espera do conjunto B.

Definição 16. *Um espaço de regiões* [7] (X, R) *é um par, onde* X *é um conjunto, e* R *é um conjunto de subconjuntos de* X

No nosso exemplo, X é $\mathbb{R}^2$ e R é o conjunto de todos os triângulos em $\mathbb{R}^2$.

Definição 17. *Dado um espaço de regiões* (X, R) *e um conjunto* $A \subseteq X$*, finito, diremos que* $B \subseteq A$ *é uma* ϵ*-aproximação de* A *se, para todo* $r \in R$

$$\left| \frac{|A \cap r|}{|A|} - \frac{|B \cap r|}{|B|} \right| \leq \epsilon$$

O que desejamos determinar é quais espaços de regiões (X, R) são tais que, para todo A subconjunto finito de X, temos uma ϵ-aproximação B, com $|B| < t$, onde t é um inteiro que depende de ϵ e (X, R), mas independe de A.

Começamos exibindo um espaço que não nos permite tais B: $(\mathbb{R}^2, P)$, onde P é o conjunto de todos os polígonos convexos em $\mathbb{R}^2$.

Tomemos $(\mathbb{R}^2, P)$, e A um conjunto de pontos distintos sobre um circulo. Para todo $C \subseteq A$, podemos tomar o polígono que tem os elementos de C como vértices. Chamando esse polígono de r, é claro que $r \cap A = C$.

Isso gera um problema: Se B é uma ϵ-aproximação de A, $r \in R$ e $r \cap A = C$, segue direto da definição que

$$\frac{|A \cap r|}{|A|} = \frac{|C|}{|A|} > \epsilon \Rightarrow |B \cap r| > 1,$$

como temos que $|B \cap r| = |B \cap A \cap r| = |B \cap C|$, segue que B tem que intersectar todos os subconjuntos C de A com $|C| > \epsilon|A|$. Isso só é possível com $|B| \geq (1 - \epsilon)|A|$. Como A pode ter um tamanho arbitrário, claramente não há o t que desejávamos.

[7] em inglês, range space

O fundamental do argumento acima foi construir um conjunto A, de tamanho arbitrário, tal que, por meio de intersecções com conjuntos de R, conseguimos gerar todos os subconjuntos de A. Isso é suficiente para impedir um espaço de ter as ϵ-aproximações que buscamos. Esse fato motiva as seguintes definições:

Definição 18. *Dado um espaço de regiões (X, R), e um subconjunto $A \subseteq X$, definimos a projeção de A como*

$$\Pi_R(A) = \{A \cap r | r \in R\}.$$

Dizemos que um conjunto A é particionado em um espaço de regiões se $|\Pi_R(A)| = 2^{|A|}$ (i.e. se a projeção de A contém todos os subconjuntos de A)

A dimensão VC de um espaço de regiões (X, R) é o maior natural d tal que há A, com $|A| = d$, e A particionado. Se não houver tal d (i.e. para todo natural n houver A, com $|A| \geq n$, particionado) diremos que a dimensão de (X, R) é ∞

Como vimos em $(\mathbb{R}^2, P)$, espaços de dimensão infinita não permitem ϵ-aproximações de tamanho limitado. Provaremos em breve que espaços de dimensão finita permitem tais aproximações.

3.2 Alguns espaços de Dimensão VC finita

Vamos agora mostrar alguns espaços de regiões naturais com dimensão finita.

Definição 19. *Dado um ponto x em $\mathbb{R}^n$, dizemos que o conjunto de pontos y tais que $(x, y) = 1$ é o hiperplano definido por x em $\mathbb{R}^n$ (onde (x, y) é o produto escalar usual).*

Dizemos também que esse hiperplano define dois semi-espaços: $\{y|(x, y) > 1\}$ e $\{y|(x, y) < 1\}$. Ao primeiro, chamaremos de semi-espaço positivo de h, e ao segundo, de negativo. Chamaremos o conjunto de todos os semi-espaços em uma dada dimensão de H^n

Teorema 12. *(R^n, H^n) é um espaço de regiões de dimensão $n + 1$*

Para provar esse fato, precisaremos mostrar que não há conjunto de $n + 2$ elementos particionado nesse espaço, mas há tal conjunto com $n + 1$ elementos.

Fato. *Não há conjunto de $n + 2$ pontos que possa ser particionado por hiperplanos em $\mathbb{R}^n$*

Demonstração. Tomemos um conjunto S de $n + 2$ pontos em $\mathbb{R}^n$. Vamos mostrar um subconjunto C de S tal que não há $h \in H^n$ com $S \cap h = C$.

Tomando $S = \{s_1, s_2, \ldots, s_{n+2}\}$, temos $n + 2$ reais $\lambda_1, \lambda_2 \ldots \lambda_{n+2}$, não todos nulos, tais que:

$$\sum_{i=1}^{n+2} \lambda_i s_i = 0, \sum_{i=1}^{n+2} \lambda_i = 0$$

(para verificar a existência de tais λ, basta tomar o conjunto $L = \{l_1, l_2, \ldots, l_{n+2}\}$ de $n + 2$ pontos no espaço de dimensão $n + 1$, onde l_i tem as primeiras n coordenadas de s_i, e 1 em sua última coordenada, e o fato segue diretamente de L ser um conjunto linearmente dependente de pontos)

Tomamos, então, C, o conjunto dos s_i tais que λ_i é positivo. Notemos que $C \neq \{\}$ e $C \neq S$ pela definição dos λ. Suponhamos, por absurdo, que há um hiperplano que contenha todos os pontos de C, e nenhum de $C - S$. Seja x o vetor que define esse hiperplano. Temos, então, que, para todo s_i em C, $(x, s_i) > 1$, e para os demais, $(x, s_i) \leq 1$.

Porém, é fácil verificar que

$$\frac{\sum_{s_i \in C} \lambda_i s_i}{\sum_{s_i \in C} \lambda_i} = \frac{\sum_{s_i \notin C} -\lambda_i s_i}{\sum_{s_i \notin C} -\lambda_i} \tag{10}$$

e, fazendo o produto escalar desses dois vetores com x, obtemos que

$$\frac{\sum_{s_i \in C} \lambda_i (x, s_i)}{\sum_{s_i \in C} \lambda_i} = \frac{\sum_{s_i \notin C} -\lambda_i (x, s_i)}{\sum_{s_i \notin C} -\lambda_i}$$

Mas o lado da esquerda é uma média ponderada de números maiores que 1, e o da direita, uma média ponderada de números menores que 1, o que é um absurdo.

(Vale a pena observar que o lado esquerdo de (10) é um ponto no fecho convexo de C, e o lado direito, de $S - C$)

□

Fato. *Há um conjunto particionado por hiperplanos, com $n+1$ pontos, em $\mathbb{R}^n$*

Demonstração. Tomemos S o conjunto que contém o vetor nulo, e os vetores com 1 em uma coordenada e 0 nas demais.

Para todo subconjunto C de S que não contém a origem, tomemos o vetor $(x_1, x_2, \ldots, x_n)$ com $x_i = 2$ se a o vetor da coordenada i está em C, e zero caso contrário. Tal vetor define um hiperplano h. Chamando o semi-espaço positivo de h de h^+, temos $h^+ \cap S = C$. Chamando o semi-espaço negativo de h de h^-, temos $h^- \cap S = S - C$. Assim, podemos, por intersecções com semi-espaços, obter todos os subconjuntos de S que não contém a origem, e também todos os que contém a origem, o que finaliza a demonstração. □

Temos ainda que $(\mathbb{R}^n, H_k^n)$ tem dimensão finita, onde H_k^n é o conjunto das intersecções de no máximo k elementos de H^n. Assim, os triângulos em $\mathbb{R}^2$ (ou qualquer conjunto de polígonos com numero fixo de lados) geram um espaço de dimensão finita. Esse fato segue do seguinte teorema:

Teorema 13. *Dado um espaço de regiões (X, R), de dimensão d, finita, com $d > 2$, e tomando $R_h = \{r_1 \cap r_2 \cap \ldots r_h : r_1, r_2 \ldots r_h \in R\}$, temos que (X, R_h) também tem dimensão finita. De fato, a dimensão de (X, R_h) é menor ou igual a $2dh \log(dh)$.*

Provaremos esse teorema na sessão seguinte.

Observemos que, para o caso dos triângulos, o limite superior que esse teorema fornece é bastante frouxo. A dimensão de $(\mathbb{R}^2, T)$, onde T é o conjunto dos triângulos em $\mathbb{R}^2$, é 7. Verifiquemos que, de fato, a dimensão de $(\mathbb{R}^2, T)$ é menor que 8:

Tomemos 8 pontos em $\mathbb{R}^2$. Se um ponto está no fecho convexo dos demais, é impossível selecionar os demais sem selecionar esse ponto. Caso contrário, os pontos são vértices de um polígono convexo é impossível selecionar um conjunto de 4 vértices sem vértices vizinhos.

3.3 Uma consequência de uma dimensão VC finita

Nessa sessão, vamos estabelecer uma propriedade essencial dos espaços de dimensão finita: Como a dimensão afeta $\Pi_R(A)$, para subconjuntos A maiores do que a dimensão do espaço. Na verdade, determinaremos que $|\Pi_R(A)| \leq |A|^d + 1$, para tais A. Esse fato será fundamental para a prova de que há ϵ-aproximações nesses espaços.

Seja $g(n, d) = \sum_{i=0}^{d} \binom{n}{i}$, e observemos que, diretamente de $\binom{n}{i} = \binom{n-1}{i} + \binom{n-1}{i-1}$ segue que $g(n, d) = g(n-1, d) + g(n-1, d-1)$ .Temos, então:

Teorema 14. *Se (X, R) é um espaço de regiões de dimensão d, e $|X| = n$, então $|R| < g(n, d)$*

Demonstração. Faremos essa prova por indução. De início, se $n = 0$, o único conjunto que R pode conter é o vazio. Assim, $|R| \leq 1 = g(d, n)$. Se $d = 0$, o teorema também se verifica: R não pode conter mais do que um conjunto, pois, caso contrário, haveria um subconjunto unitário de X particionado.

Agora, vamos assumir que o teorema se verifica para espaços com $|X| = n - 1$ e dimensão d ou $d - 1$. Vamos agora fixar um elemento $x \in X$ e, com base nele, definir dois espaços de regiões auxiliares:

- S_1, retirando x de X, e de todo $r \in R$ que o contenha.Ou seja,

$$S_1 = (X_1, R_1), \text{com } X_1 = X - \{x\} \text{ e } R_1 = \{r - x | r \in R\})$$

- $S_2 = (X_2, R_2)$, retirando x de X, e selecionando em R apenas conjuntos r "atropelados"em S_1, ou seja, conjuntos $r \in R$ tais que $x \notin r$ e $r \cup \{x\} \in R$.

Claramente, $|R_1| + |R_2| = |R|$, e $|R_1| \leq g(n-1, d)$. Também temos que a dimensão de S_2 é no máximo $d-1$, pois, para todo subconjunto particionado C de S_2, $C \cup \{x\}$ é um subconjunto particionado de S. Assim, $|R_2| \leq g(n-1, d-1)$, e temos

$$|R| \leq g(n-1, d-1) + g(n-1, d) = g(n, d)$$

□

Nosso interesse no teorema acima reside no seguinte fato: Seja (X, R) um espaço de regiões de dimensão d qualquer (i.e., X pode ser infinito) e A um subconjunto finito de X. $(A, \Pi_R(A))$ é um espaço de regiões, de dimensão no máximo d. Assim, $|\Pi_R(A)| \leq g(|A|, d)$.

Resta provar que $g(|A|, d) \leq |A|^d + 1$.

Na verdade, se $d \geq 2$, $g(|A|, d) \leq |A|^d$. Notemos que $g(|A|, d)$ é o número de subconjuntos de A de tamanho máximo d. Vamos mapear as strings de comprimento d no alfabeto A para os subconjuntos não vazios. Tomemos uma tal string, J, e seja $a_J \in A$ o primeiro símbolo repetido dessa string. Vamos associar J ao conjunto dos símbolos que antecedem a segunda aparição de a_J (e se não houver tal a_J, associaremos a string ao conjunto de todos os seus d símbolos). Lidemos com com o vazio da seguinte maneira: escolhemos um conjunto de d elementos,e associamos alguma de suas ordenações ao vazio. Construímos, assim, uma função sobrejetora, de um conjunto de cardinalidade $|A|^d$ para um de cardinalidade $g(|A|, d)$, o que prova $|A|^d \geq g(|A|, d)$. No caso $d = 1$ verifica-se trivialmente que $g(|A|, 1) = |A| + 1$

Registremos esse resultado

Teorema 15. $|\Pi_R(A)| \leq |A|^d + 1$

Estamos prontos para provar nosso resultado principal. Antes, porém, devemos provar o teorema (13):

Demonstração. Temos que $d \geq 2$, assim, $|\Pi_R(A)| < |A|^d$. Mas os elementos de $\Pi_{R_h}(A)$ são intersecções de no máximo h elementos de $\Pi_R(A)$. Assim, $\Pi_{R_h}(A) \leq \sum_{i=1}^{h} \binom{|A|^d}{h} \leq |A|^{dh}$. Assim, se $|A|^{dh} < 2^{|A|}$, A não pode ser particionado. Sabemos que há um inteiro $n = |A|$ a partir do qual essa desigualdade vale, e pode-se verificar que a desigualdade é verdade para $n > 2dh \log(dh)$ (lembrando que $d \geq 2$ e $h \geq 2$). □

3.4 Existência de ϵ-aproximações para espaços de regiões

Teorema 16. *Dado (X, R), um espaço de regiões de dimensão VC d, para todo A subconjunto de X, existe uma ϵ-aproximação de A de cardinalidade $\min(|A|, m)$, onde m é função de ϵ e d, mas não de A.*

Tomemos uma sequencia aleatória (com possíveis repetições) de m elementos de A. Chamando o multiconjunto[8] dos elementos que aparecem na sequencia de L, seja P_1 a probabilidade de L não ser uma ϵ-aproximação de A, ou seja, a probabilidade de existir $r \in R$ tal que $\left|\frac{|A\cap r|}{|A|} - \frac{|L\cap r|}{|L|}\right| > \epsilon$.

Tomemos também uma sequencia aleatória de $2m$ elementos de A. Chamemos o multiconjunto dos m primeiros termos de B, e o dos m seguintes, de C. Diremos que uma tal sequencia é desequilibrada de houver $r \in R$ tal que $\left|\frac{|B\cap r|}{|B|} - \frac{|C\cap r|}{|C|}\right| \geq \frac{\epsilon}{2}$. Chamaremos o evento de uma tal sequencia ser desequilibrada de E_2, e a probabilidade de esse evento ocorrer, de P_2.

Fato. *Se $m > \frac{2}{\epsilon^2}$, $P_2 \geq \frac{1}{2}P_1$*

Demonstração. Tomando uma sequencia aleatória de $2m$ elementos de A, com B e C definidos como acima, chamaremos E_3 ao seguinte evento:

(*i*) B não é uma ϵ-aproximação, pois, para alguns r em R, $\left|\frac{|A\cap r|}{|A|} - \frac{|B\cap r|}{|B|}\right| > \epsilon$.

(*ii*) Para um dos r acima, temos que $\left|\frac{|A\cap r|}{|A|} - \frac{|C\cap r|}{|C|}\right| \leq \frac{\epsilon}{2}$

Claramente, E_3 implica E_2, e, portanto, $P_2 \geq P_3$.

Mas, a chance de (i) ocorrer é P_1, e vamos provar que a chance de (ii) ocorrer, para um r qualquer, é maior que $1/2$. Assim, teremos $P_3 \geq P_1/2$.

Provemos que a chance de (ii) ocorrer para um r qualquer é maior que $1/2$: Seja p a probabilidade de, sorteando um elemento de A uniformemente, obtermos um elemento de r. Pela desigualdade de Chebychev, temos $P\left(\left|\frac{|A\cap r|}{|A|} - \frac{|C\cap r|}{|C|}\right| > \frac{\epsilon}{2}\right) < \frac{\sigma^2}{\frac{\epsilon^2}{4}} = \frac{\frac{p(1-p)}{m}}{\frac{\epsilon^2}{4}} < 1/2$. (usando que $p(p-1) \leq 1/4$ e $m > 2/\epsilon^2$).

Assim, (ii) ocorre com probabilidade $\geq 1/2$, $P_3 \geq P_1/2$ e, portanto, $P_2 \geq P_1/2$. □

Fato. *Dada uma sequencia aleatória de $2m$ elementos de A, a probabilidade de a sequencia ser desbalanceada para é menor que $g(2m, d)e^{-m\epsilon^2/8}$*

Antes de partir para a demonstração, precisamos enunciar um teorema, devido a [Hoe63].

Teorema 17. *Sejam $X_1 \ldots X_m$ variáveis aleatórias independentes, tais que $a_i \leq X_i \leq b_i$, e $t > 0$, temos*

$$P[|\bar{X} - \mu| \geq t] \leq 2e^{\frac{-2t^2m^2}{\sum_{i=1}^{n}(b_i-a_i)^2}}$$

□

Demonstração. (do fato) Faremos um sorteio em duas etapas, que produzirá as sequencias de tamanho $2m$ de forma equiprovável.

[8] i.e., um conjunto generalizado que permite a repetição de elementos. Nesse trabalho, quando lidamos com $|F \cap r|$, e F é um multiconjunto, estamos nos referindo à soma do número de ocorrências de cada elemento de r em F

De início, sorteamos uma sequencia de $2m$ elementos de A, elemento a elemento, com reposição, obtendo uma sequencia S_1. Chamemos ao i-ésimo elemento dessa sequencia de a_i. Usando S_1, produzimos uma sequencia aleatória S_2, da seguinte forma: Para cada $1 \leq i \leq m$, trocamos de posição a_i e a_{m+i}, com probabilidade $1/2$.

S_2 claramente tem distribuição equiprovável dentre todas as sequencias de $2m$ elementos de A

Depois de feito o sorteio de S_1, já sabemos quais subconjuntos de A estão em $\Pi_R(S_2) = \Pi_R(S_1)$, e que há no máximo $g(2m, d)$ deles. S_2 será desbalanceada para um r somente se for desbalanceada para seu representante em $\Pi_R(S_2)$, pela definição.

Nosso intuito, então, é achar um limite superior para a probabilidade de S_2 ser desbalanceada para um dos conjuntos de $\Pi_R(S_2)$. Fixemos um tal conjunto, chamando-o w.

Definamos as variáveis aleatórias $X_1 \dots X_m$ da seguinte maneira: Se o i-ésimo elemento de S_2 pertence a w, mas o $m+1$-ésimo não pertence, X_i é 1. Se o inverso ocorrer, X_i é -1. Se ambos (ou nenhum) pertencerem a w, X_i é 0. Assim, X_i são variáveis aleatórias, no intervalo $[-1, 1]$ (de fato, algumas dessas "variáveis aleatórias"assumem apenas o valor zero, mas isso não nos impede de aplicar o teorema (17)). O valor esperado da soma é zero. A probabilidade de S_2 ser desbalanceada é $P[|\bar{X} - 0| > \epsilon/2]$. Mas, pelo teorema (17):

$$P[|\bar{X} - 0| > \epsilon/2] \leq 2e^{\frac{-2(\frac{\epsilon}{2})^2 m^2}{\sum_{i=1}^{n}(b_i - a_i)^2}} = e^{\frac{-\frac{\epsilon^2}{2}m^2}{\sum_{i=1}^{n}(2)^2}} = e^{\frac{-\frac{\epsilon^2}{2}m^2}{4m}} = e^{-\frac{\epsilon^2}{8}m}$$

Assim, $P_2 \leq e^{-\frac{\epsilon^2}{8}m}|\Pi_R(S_2)| = e^{-\frac{\epsilon^2}{8}m}g(2m, d)$

□

A prova do teorema (16) segue: $e^{-\frac{\epsilon^2}{8}m}g(2m, d) \leq e^{-\frac{\epsilon^2}{8}m}[(2m)^d + 1]$, e essa última expressão vai a zero conforme m cresce. Assim, P_2 vai a zero, e, portanto, P_1 vai a zero. Existe, portanto, b tal que, se $m > b$, $P_1 < 1$. Isso implica que há $L \subset A$ tal que o evento de P_1 não ocorre. Tal L é uma ϵ-aproximação.

A prova do teorema nos dá mais informações: Podemos escolher m de tal forma que, ao pegarmos $B \subset A$ de cardinalidade m, B aleatório, a probabilidade de B não ser uma ϵ-aproximação é quão pequena quanto queiramos. De fato, segue da desigualdade que obtivemos que, com

$$m = \frac{c}{\epsilon^2}\left(d \log \frac{d}{\epsilon} + \log \frac{1}{\delta}\right)$$

um subconjunto aleatório de m elementos de A é uma ϵ-aproximação com probabilidade $1 - \delta$. (c é uma constante positiva, independente de ϵ, δ e d)

Os resultados dessa sessão, provados em um contexto diferente, vêm de [eAYC71]

3.5 Rumo a uma solução exata (e barata)

Acabamos de obter uma maneira razoável de dar soluções aproximadas para o problema geométrico que nos propomos a estudar. Nessa seção, estabelecemos um resultado útil para construir uma solução exata.

Antes de mais nada, precisamos pensar a respeito da solução exata trivial: Simplesmente pegar cada ponto de A e testar se ele está em r. Essa é uma solução linear em $|A|$. Desejamos soluções melhores: Soluções cujo tempo de execução seja $O(|A|^{\alpha})$, com $\alpha < 1$.

Para isso, utilizaremos a seguinte estrutura:

Definição 20. *Uma árvore de partição para um conjunto A é uma árvore binária, com $|A|$ folhas. Cada nó n da árvore é caracterizado por um conjunto $E(n)$ de elementos de A. Para*

cada subconjunto unitário x de A, associamos uma folha (i.e. n, com $E(n) = x$). Quando n não é uma folha, ele tem dois filhos e temos $E(n) = E(n_e) \cup E(n_d)$, onde n_e é o filho esquerdo de n, e n_d, o direito. Alem disso, $E(n_e)$ e $E(n_d)$ são disjuntos e não vazios. A raiz ra tem $E(ra) = A$.

Dada uma tal árvore, o problema de determinar $|r \cup A|$ pode ser resolvido recursivamente: Definimos uma função $f(r, n)$ que retorna 0 se r e $E(n)$ são disjuntos, $|E(n)|$ se $E(n) \subset r$, e $f(r, n_d) + f(r, n_e)$ quando $E(n)$ tem elementos dentro e fora de r. Claramente, $f(r, ra)$ nos retorna $|r \cap A|$. O tempo necessário para executar f depende do número de nós visitados, e também do tempo necessário para determinar qual tipo de intersecção ocorre entre r e $E(n)$.

Nós, porém, só nos preocuparemos com o número de nós visitados. Em quais espaços de regiões é possível, para todo A, construir uma árvore de partição, de forma que, para todo $r \in R$, o algoritmo descrito acima visite $O(|A|^\alpha)$ nós ? Claramente, essa é uma condição necessária para conseguirmos usar árvores de partição para descobrir $|A \cap r|$ rapidamente .Para alguns espaços de regiões naturais, essa condição será suficiente.

Os resultados que apresentamos a seguir, e sua aplicação para gerar algoritmos eficientes, podem ser encontrados em [eEW89].

3.5.1 A dimensão VC e as árvores de partição

Dizemos que um conjunto r atravessa um conjunto $E(n)$ quando existem $x, y \in E(n)$ tais que $x \in r$ e $y \notin r$.

Dado um espaço de regiões (X, R), $A \subset X$ e T uma árvore de partições de A, dizemos que $r \in R$ visita um nó n de T se n é a raiz, ou se o pai de n é atravessado por r. O número de visitações de uma dada árvore é o máximo número de nós visitado por um $r \in R$.

Nosso objetivo é, então, descrever quais espaços (X, R) permitem árvores de partições com número de visitação $O(|A|^\alpha)$. Novamente, provaremos que esses espaços são os espaços de dimensão VC finita.

De início, provemos que num espaço de dimensão VC infinita, toda árvore tem número de visitação em $\Omega(|A|)$.

Se A é particionado, então, para toda árvore de A, temos que o número de visitação é $\geq |A|$: Tomemos uma árvore T. Enumerando suas folhas da direita para a esquerda, e escolhendo as folhas com numeração impar, obtemos um conjunto $I \subset A$. Tomemos r tal que $r \cap A = I$. r atravessa todo nó tal que seus dois filhos são folhas, e também todo nó que tem um descendente tal que seus dois filhos são folhas. Assim, r atravessa todo nó interno de T, e visita todos os nós de T, incluindo as $|A|$ folhas. Assim, num espaço de dimensão VC infinita não podemos construir as árvores que desejamos.

No que se segue, provaremos que todo espaço de dimensão VC finita permitirá árvores de partições com número de visitação $O(|A|^\alpha)$, $\alpha < 1$.

3.5.2 O espaço de regiões dual

Dado um espaço de regiões (X, R), vamos definir seu espaço de regiões dual como o espaço (R, X^*), onde $X^* = \{R_x | x \in X\}$ e $R_x = \{R | x \in R\}$. Ou seja, passamos a ver os conjuntos $r \in R$ como pontos, e os pontos $x \in X$ como conjuntos R_x que contém todos os r que continham x.

Se (X, R) é um espaço de regiões, definamos a função de particionamento [9] como $\pi(m) = \max_{|A|=m} |\Pi_R(A)|$. Essa noção não é nova, e o fato de que ela é polinomial em m quando um

[9] em inglês, shatter function

espaço tem dimensão finita foi bastante importante na prova da existência das ϵ-aproximações.

Definimos, então a função dual de particionamento de um espaço (X, R), $\pi^*(m)$, como a função de particionamento de (R, X^*).

Há também uma definição alternativa, que nos ajuda a entender melhor essa função: Dado um conjunto $Q \subset R$, chamamos de *célula* um conjunto maximal de elementos de X que não é atravessado por nenhum $r \in Q$ (i.e. uma classe de equivalência em relação à pertinência em elementos de Q). Definindo $\Pi^*(Q)$ como o número de células de Q, temos $\pi^*(m) = \max_{|Q|=m} \Pi^*(Q)$

O espaço de regiões dual tem a seguinte propriedade importante:

Teorema 18. *(X, R) tem dimensão finita sse (R, X^*) também tem dimensão finita.*

Demonstração. Provaremos que (X, R) tem dimensão *infinita* sse (R, X^*) também o tem. De fato, provaremos apenas que se (X, R) tem dimensão infinita, (R, X^*) também o tem, pois a volta é análoga.

Se (X, R) tem dimensão infinita, então, para todo k, (X, R) contém um conjunto particionado com 2^k elementos[10]. Tomemos um tal conjunto A, e numeremos seus elementos. Tomemos, então a representação binária desses números.

Para todo $0 \leq i \leq k-1$, seja S_i um conjunto em R tal que $S_i \cap A$ contém todos os elementos com 1 na i-ésima posição de sua representação binária, e nenhum com 0 nessa posição. Como A é particionado, R de fato contém um S_i para cada i. Se houver mais de um candidato a S_i para um dado i, simplesmente escolhemos arbitrariamente.

Seja S o conjunto dos S_i. Claramente, ele é particionado em (R, X^*).

Assim, obtivemos um conjunto particionado de R, com k elementos, para qualquer k. Temos, então que, (R, X^*) tem dimensão infinita. □

3.5.3 Construindo uma boa árvore de partição

Nós afirmamos que, para todo espaço de regiões (X, R) de dimensão finita, e todo subconjunto $A \in X$, é possível produzir uma árvore com número de visitação em $O(|A|^\alpha)$, com $\alpha < 1$. Nessa sessão, provaremos esse fato, construindo uma tal árvore. Precisamos de mais dois fatos antes de iniciar a construção:

Teorema 19. *Se (X, R) tem dimensão finita, $(X, \hat{R})$ também tem (onde $\hat{R} = \{(r \cup r') \backslash (r \cap r') | r, r' \in R\}$)*

Para constatar esse fato, basta notar que $\Pi_{\hat{R}} \leq (\Pi_R(A))^2$ e, assim, se $\pi_R(m) \leq m^d + 1$ (isso é o teorema (15) traduzido para a linguagem de funções de particionamento) então $\pi_{\hat{R}}(m) \leq (m^d + 1)^2$. Mas, a partir de algum j, $(j^d + 1)^2 \leq 2^j$, e assim, conjuntos A tais que $|A| \geq j$ não são particionados em $(X, \hat{R})$

Definição 21. *Dado um espaço de regiões (X, R), e um multiconjunto A de elementos de X, com cardinalidade finita, uma ϵ-rede [11] é um multiconjunto $B \in A$ tal que*

$$|A \cap r| > \epsilon |A| \rightarrow r \cap B \neq \emptyset$$

Essa definição, e o teorema a seguir, são essenciais para a construção que vamos fazer.

[10] de fato, sabemos que há um conjunto particionado com mais de 2^k elementos, e podemos tomar um subconjunto de exatamente 2^k elementos que, obviamente, também é particionado

[11] em inglês, ϵ-net

Teorema 20. *Se (X, R) é um espaço de regiões de dimensão d finita, para todo A multiconjunto de elementos de X, existe B uma ϵ-rede de A, com $|B| \leq \lceil \frac{8d}{\epsilon} \log \frac{8d}{\epsilon} \rceil$*

A prova desse teorema é bastante similar à prova que já apresentamos, da existência de ϵ-aproximações. Assim, a omitimos. Ela pode ser encontrada em [NA08] [12] Registramos apenas que toda ϵ-aproximação é uma ϵ-rede.

Faremos a construção da árvore de partição em etapas: Definindo uma aresta de A como um par de elementos de A, provaremos que, em determinadas condições, existe uma aresta que atravessa poucos conjuntos. Utilizaremos tais arestas para construir um caminho cujas arestas atravessam poucos conjuntos (caminho no sentido usual para grafos). Esse caminho, então, nos permitirá construir uma árvore de partição.

Preste atenção nas hipóteses dos teoremas a seguir: Basicamente, eles assumem que (X, R) tem dimensão finita, e que (R, X^*) também o tem. Pelo teorema (18), podemos deduzir essas hipóteses apenas do fato de que (X, R) tem dimensão finita.

Teorema 21. *Seja (X, R) um espaço de regiões, $\pi^*(m) \leq gm^d$. Seja A um subconjunto de X com n pontos, e Q um multiconjunto de elementos de R com m conjuntos. Existe um par de pontos de A que é atravessado por no máximo $cm(\log n)/n^{1/d}$ conjuntos $r \in Q$, onde c é uma constante.*

Esse teorema é o coração da prova. Tentaremos escrevê-lo com cuidado, e sugerimos cuidado ao lê-lo.

Demonstração. Como $\pi^*(m) \leq gm^d$, o espaço dual (R, X^*) tem dimensão finita. Lembremos que os elementos de X^* são os R_x, conjuntos que contém como elementos todos os $r \in R$ que contém x.

A diferença simétrica de R_x e R_y é o conjunto de todos os conjuntos de R que contém x ou y, mas não ambos. Ou seja, o conjunto de todos os conjuntos de R que atravessam $\{x, y\}$. Assim, $\hat{X^*}$ é um conjunto de subconjuntos de R, e para cada desses subconjuntos, há um par $\{x, y\}$ que todos os seus elementos atravessam, e mais ninguém atravessa.

Mas, como já vimos em no teorema (19), $(R, \hat{X^*})$ tem dimensão finita. Resulta que podemos construir uma ϵ-rede de $(R, \hat{X^*})$, para qualquer valor de ϵ que queiramos. Construamos N, uma ϵ-rede, com $\epsilon = c(\log n)/n^{1/d}$. Pelo teorema (20), e usando $\pi^*(m) \leq gm^d$, podemos obter $\pi^*(|N|) < n$, com uma escolha apropriada de c.

Então, construímos N, um conjunto de elementos de R, com menos de n células. Assim, há dois pontos de P numa mesma célula. Ou seja, esse par não é atravessado nenhum dos conjuntos de N. Mas todo conjunto grande de $\hat{X^*}$ está representado em N. Ou seja: todo par que é atravessado por mais de $\epsilon m = cm(\log n)/n^{1/d}$ conjuntos de Q é atravessado por um conjunto de N. Assim, esse par que está numa mesma célula é atravessado por menos de $cm(\log n)/n^{1/d}$ conjuntos de Q.

□

Teorema 22. *Seja (X, R) um espaço de regiões, de dimensão u, com $\pi^*(m) \leq gm^d$ e A um subconjunto finito de x. Há um $C = C(g, d)$ e um caminho hamiltoniano em A tal que todo $r \in R$ atravessa não mais do que $Cn^{1-1/n} \log n$ arestas.*

[12] A prova que consta nesse livro é de um teorema mais forte que esse: Não só se afirma a existência de uma ϵ-rede, como também se elabora um procedimento para obtê-la com probabilidade alta. Por isso, o tamanho da ϵ-rede no livro também é função de um parâmetro δ. Tomando um δ apropriado, nossa versão do teorema segue imediatamente

Demonstração. Nosso intuito é aplicar o teorema (21) várias vezes, para construir uma árvore com todos os elementos de A. Apliquemos o teorema com A e $Q = \Pi_R(A)$, obtendo um par x_0, y_0 que não é atravessado por mais do que $cm(\log n)/n^{1/d}$ conjuntos de Q. Apliquemos novamente o teorema, dessa vez, a $A - \{x_0\}$ e $Q_1 = Q \cup \{r \in Q | \text{r atravessa } x_0y_0\}$[13]. Obtemos um par x_1y_1, que não é atravessado por mais do que $c|Q_1|(\log n - 1)/(n-1)^{1/d}$ conjuntos de Q.

Aplicando o teorema $n-1$ vezes, obtemos uma árvore (composta pelas arestas devolvidas pelo teorema), e um conjunto Q_{n-1} tal que

$$|Q_{n-1}| \leq \Pi_R(A) \prod_{i=0}^{n-1} \left(1 + \frac{c \log n}{(n-i)^{1/t}}\right) \leq (n^u + 1) e^{c \log n \sum_{i=0}^{n-1} \frac{1}{(n-i)^{1/t}}}$$

$$\leq 2.2^{u \log n} e^{c \log n \sum_{i=0}^{n} \frac{1}{(n-i)^{1/t}}} \leq 2^{c' n^{1-1/t} \log n}$$

onde c' é uma função de u, t e c.

Temos que um conjunto $r \in R$ que atravesse s arestas da nossa árvore aparece em Q_{n-1} 2^s vezes. Assim, temos $2^s \leq 2^{c' n^{1-1/t} \log n}$, e portanto, $s \leq c' n^{1-1/t} \log n$

Transformemos a árvore que obtivemos num caminho: Dobramos cada aresta da árvore, e tomamos um circuito euleriano no grafo resultante. Isso nos fornece uma string de vértices. Mantendo apenas a primeira ocorrência de cada vértice, obtemos um caminho hamiltoniano. Esse é o caminho que desejamos. Afirmamos que, se um dado conjunto $r \in R$ atravessa l arestas na árvore que obtivemos, ele atravessa no máximo $2l$ arestas nesse novo caminho hamiltoniano. Associemos as arestas do caminho hamiltoniano aos caminhos correspondentes no circuito euleriano (i.e. associamos uma aresta xy ao caminho percorrido entre a primeira ocorrência de x e a primeira de y no caminho euleriano). Uma aresta do caminho hamiltoniano só pode ser atravessada por r se alguma aresta do caminho correspondente o for. Como há $2l$ arestas atravessadas por r no circuito euleriano, há no máximo $2c' n^{1-1/t} \log n$ arestas atravessadas no caminho hamiltoniano. □

Agora, estabelecemos como transformar nosso caminho hamiltoniano numa árvore de partições adequada.

Teorema 23. *Dado um espaço de regiões (X, R), e um conjunto finito $A \subset X$, de cardinalidade n, se há um caminho hamiltoniano em A tal que todo $r \in R$ atravessa no máximo p arestas do caminho, então, há uma árvore de partição para A, com número de visitações no máximo* $2p\lceil \log n \rceil + 1$

Demonstração. Nossa árvore será uma árvore binária balanceada, com n folhas. Tomando as folhas, da direita para a esquerda, vamos associá-las aos conjuntos unitários de A na ordem em que eles aparecem no caminho hamiltoniano.

Tomemos um $q \in R$ o conjunto com número de visitações máximo. Se um nó interno da árvore for atravessado por q, a subárvore desse nó contém duas folhas consecutivas x e y tais que xy é atravessada por q. Dessa forma se houver p tais folhas, haverá no máximo $p\lceil \log n \rceil$ nós internos atravessados por q. Assim, q visitará no máximo $2p\lceil \log n \rceil + 1$ nós (os filhos de quem ele atravessa, e mais a raiz). □

Assim, obtivemos uma árvore de partição com número de visitações no máximo $4c' n^{1-1/t} \log n \lceil \log n \rceil + 1$, e, portanto, que pertence a $O(|A|^\alpha)$.

[13] os Q_i são multiconjuntos, e estamos aqui aumentando a multiplicidade de alguns de seus elementos, "penalizando"os conjuntos que atravessam x_0y_0

Referências

[eAYC71] V. N. Vapnik e A. YA. Chervonenkis, *On the uniform convergence of relative frequencies of events to their probabilities*, Theory Probab. Appl. **16** (1971), 264–280.

[eDRS] Geoffrey R Grimmett e David R. Stirzaker, *One thousand exercises in probability*, Oxford University Press.

[eEW89] Bernard Chazelle e Emo Welzl, *Quasi-optimal range searching in spaces of finite vc-dimension*, Discrete and Computational Geometry **4** (1989), 467–489.

[EG85] A. Rucinski E. Györi, B. Rothschild, *Every graph is contained in a sparsest possible balanced graph*, Math. Proc. Camb. Phil. Soc. **98** (1985).

[Hoe63] Wassily Hoeffding, *Probability inequalities for sums of bounded random variables*, Journal of the American Statistical Association **58** (1963).

[JS] Rucinski A Janson S., Luczak T., *Random graphs*, Wiley.

[Mac05] David J. C. MacKay, *Information theory, inference and learning algorithms*, Wiley-Interscience, 2005.

[NA08] J. Spencer N. Alon, *The probabilistic method*, Wiley-Interscience, 2008.

Parte II
Parte Subjetiva

Acho que o maior problema que tive durante o TCC (e talvez, durante todo o BCC) foi estabelecer uma rotina de estudo. Ocorre que eu me cobro muito quando estudo, o que não só torna o trabalho menos produtivo, mas também menos prazeroso do que poderia e acaba sendo um forte incentivo para deixar para depois. Acho que isso está melhorando (e, em particular, melhorou bastante por causa do TCC) mas ainda é um problema.

O livro "O Método Probabilístico"não é fácil de se ler. Ele enfatiza bastante o método de prova, e não se detém muito sobre cada resultado. O próprio autor diz que o objetivo é enfatizar o método de prova, e não os resultados em si. Isso foi um desafio muito enriquecedor: Eu tive que buscar me aprofundar nos resultados, para poder entender porque eles eram interessantes. Saí um pouco da postura passiva de ser convencido da relevância dos fatos, para uma postura mais ativa, que acredito que será muito positiva nas minhas futuras atividades acadêmicas. A variedade dos resultados com os quais tive contato também foi um ponto positivo da escolha do livro.

O trabalho indicou vários possíveis caminhos para meus estudos futuros. Em particular, a leitura do "information theory"é muito agradável, e pretendo ler o resto o livro e estudar mais sobre teoria da informação. Pretendo fazer um mestrado em 2010, e tenho certeza que o método probabilístico me será útil em qualquer área da ciência da computação que eu vá estudar. Há algumas técnicas probabilísticas de prova com as quais gostaria de me familiarizar mais, e que suspeito que serão muito úteis no mestrado. Em particular, gostaria de estudar mais sobre desigualdades de correlação.

Vale a pena enfatizar que a leitura de "O Método Probabilístico"me levou a estudar (em menos detalhes, claro) vários resultados que não pude incluir nessa monografia.

É necessário destacar que o apoio dos meus pais e dos meus colegas de BCC foi essencial durante a feitura dessa monografia. Também cumpre agradecer o professor Yoshi, por seu apoio e paciência.

A recuperação

Essa é a segunda versão dessa monografia. A grande mudança em relação à primeira é a terceira parte, que aborda um problema de geometria computacional. Essa terceira parte foi bastante bem pesquisada e (creio eu) está bastante bem escrita. Acho que marca bem o meu amadurecimento no correr desse trabalho.

Os problemas de organização para o trabalho persistem, e tiveram seu impacto nessa versão da monografia. Me empolguei bastante com essa terceira parte, e acabei negligenciando um pouco as anteriores.

Disciplinas do BCC mais relevantes para o trabalho

- Algoritmos em Grafos, Matemática Concreta (grafinhos) : Foram muito úteis, por servir como introdução a problemas de grafos e combinatória, e, em particular, por ilustrar bastante bem técnicas construtivas de prova.

- Análise de Algoritmos: Sem os conceitos de análise assintótica, seria absolutamente impraticável usar o método probabilístico. O conhecimento da disciplina também ajudou a motivar diversos dos resultados estudados.

- Álgebra I: Essa é uma disciplina que eu acho que aproveitei particularmente bem, e tem sido muito útil desde então. O ferramental para lidar com os inteiros acaba ficando tão natural que a gente nem nota que está usando. Em particular, na monografia, utiliza-se o teorema de Bezout.

- Álgebra Linear para Computação: De fato, cursei essa disciplina na engenharia. Suspeito que teria sido bom cursá-la no IME mesmo, porque tenho oportunidade de a utilizar muito frequentemente. Em particular, álgebra linear é bastante útil quando lidando com problemas que envolvem geometria e combinatória.

- Cálculo Diferencial e Integral I: É absolutamente essencial para o raciocínio assintótico. Como álgebra I, é uma daquelas disciplinas que se tornaram naturais para mim, e nem noto muito quando uso. Também cursei essa disciplina na poli, mas supri as deficiências cursando Análise Real.

- Álgebra II: É uma disciplina muito interessante, e é essencial para entender os enunciados de alguns dos teoremas cuja prova li durante o trabalho (mas para nenhum dos que foram inclusos nessa monografia)

- Tópicos de Matemática Discreta: Fiz essa matéria com a Cris, e o assunto estudado foi algoritmos probabilísticos. A matéria introduziu vários conceitos que eu utilizaria no TCC, e também me deu algum traquejo com o tipo de prova e raciocínio necessários. De fato, tive a oportunidade de assistir a várias matérias de "Tópicos", e considero que elas foram muito interessantes para minha formação. Porém, a matricula nelas não foi possível, por causa do reuso das siglas "coringa". De fato, infelizmente, isso fez com que, em alguns momentos, tais matérias, que eu considero muito importantes, fossem preteridas em favor de outras, menos interessantes mas "pra nota".

  Já comentei esse problema com alguns professores, e gostaria muito que ele fosse corrigido. Talvez com a inclusão de mais siglas coringa e um uso mais organizado das mesmas.