Método Probabilístico em Combinatória

Franco Severo

O que você encontrará aqui é moralmente uma tradução resumida do material "Expected Uses of Probability", do Evan Chen.

1 Definições e Propriedades

Antes de qualquer coisa, um aviso: a formalização da *teoria geral da probabilidade* é feita via *teoria da medida*, mas dentro do contexto que trabalharemos aqui, tudo será suficientemente intuitivo para que não precisemos entrar nesses detalhes mais técnicos. Portanto, se você tiver um censo crítico matemático muito forte e achar que algo aqui está mal definido ou mal explicado, não se desespere: tudo pode ser construído com o mais alto rigor matemático.

Uma **variável aleatória** é um objeto que toma valores variando aleatoriamente. Dado um **evento** A, denotamos por $\mathbb{P}(A)$ a **probabilidade** de A. Por exemplo, o resultado de jogar um dado justo de seis lados é uma variável aleatória. Chame este variável de D, então neste caso temos:

$$\mathbb{P}(D=1)=\mathbb{P}(D=2)=\cdots=\mathbb{P}(D=6)=1/6$$

ou ainda $\mathbb{P}(D=0)=0$ e $\mathbb{P}(D\geq 4)=1/2$.

Podemos também definir a **esperança** (ou **valor esperado**) de uma variável aleatória X, que é simplesmente a média de X. Formalmente:

$$\mathbb{E}(X)=\sum_x x\mathbb{P}(X=x) \tag{1}$$

No exemplo do dado:

$$\mathbb{E}(D)=1\cdot\frac{1}{6}+2\cdot\frac{1}{6}+\cdots+6\cdot\frac{1}{6}=3,5$$

Problema 1 Escolha uma permutação σ de $\{1,2,3,4\}$ aleatória e uniformemente (isto é, cada permutação tem a mesma probabilidade $1/4!=1/24$ de ser escolhida). Seja F o número de pontos fixos de σ. Calcule $\mathbb{E}(F)$ usando a definição (1). E se, em vez de 4, tivermos n elementos?

Teorema (Linearidade da Esperança) Dadas variáveis aleatórias $X_1, X_2, \cdots, X_n$, temos sempre:

$$\mathbb{E}[X_1+X_2+\cdots+X_n]=\mathbb{E}[X_1]+\mathbb{E}[X_2]+\cdots+\mathbb{E}[X_n] \tag{2}$$

Prova: Primeiramente, note que basta mostras o resultado quando temos apenas duas variáveis. Neste caso, por definição:

$$\begin{aligned}
\mathbb{E}[X+Y] &= \sum_z z\mathbb{P}[X+Y=z] \\
&= \sum_z z\left(\sum_{x+y=z}\mathbb{P}[X=x,Y=y]\right) \\
&= \sum_{x,y}(x+y)\mathbb{P}[X=x,Y=y] \\
&= \sum_{x,y}x\mathbb{P}[X=x,Y=y]+\sum_{x,y}y\mathbb{P}[X=x,Y=y] \\
&= \sum_x x\left(\sum_y\mathbb{P}[X=x,Y=y]\right)+\sum_y y\left(\sum_x\mathbb{P}[X=x,Y=y]\right) \\
&= \sum_x x\mathbb{P}[X=x]+\sum_y y\mathbb{P}[Y=y] \\
&= \mathbb{E}[X]+\mathbb{E}[Y]
\end{aligned}$$

□

Note que o teorema acima nos permite calcular a esperança da variável $X = X_1 + \cdots + X_n$ sem precisar saber a *distribuição* dela (isto é, saber os valores exatos de $\mathbb{P}(X = x)$ para cada x possível): precisamos apenas saber o valor esperado de cada X_i. Somente com a definição (1) isso não seria possível.
De fato, descobrir uma forma explícita para a distribuição de F do Problema 1 para qualquer n é difícil (e um bom exercício, aliás), mas podemos calcular $\mathbb{E}(F)$ de maneira muito mais simples usando a linearidade da esperança: note que $F = F_1 + F_2 + \cdots + F_n$, onde

$$F_i = \begin{cases} 1\ , & \text{se i é um ponto fixo} \\ 0\ , & \text{caso contrário} \end{cases}$$

Mas é fácil ver que $\mathbb{E}(F_i) = \mathbb{P}(\text{i é um ponto fixo}) = \mathbb{P}(\sigma_i = i) = 1/n$, logo segue de (2) que $\mathbb{E}(F) = 1$.

Problema 2 (HMMT 2006) Em uma creche, 2006 crianças estão sentadas em círculo. De repente, cada criança cutuca aleatoriamente ou a criança à sua esquerda ou à sua direita. Qual o número esperado de crianças que não foram cutucadas?

Problema 3 (AHSME 1989) Suponha que 7 garotos e 13 garotas formam aleatoriamente uma fila. Seja S o número de lugares na fila onde um garoto e uma garota estão em pé um depois do outro. Por exemplo, para a fila $GBBGGGBGBGGGBGBGGBGG$ temos $S = 12$. Calcule a esperança de S.

Problema 4 Dado n um inteiro positivo, seja a_k o número de permutações de $\{1, 2, ..., n\}$ com exatamente k pontos fixos. Calcule $a_1 + 4a_2 + \cdots + n^2 a_n$.

2 O Método

2.1 Forma básica

Em sua forma mais simples, podemos usar o conceito de esperança para provar existência via um fato muito simples: se $\mathbb{E}(X) > n$ então existe uma probabilidade positiva de que $X > n$. Pode parecer estranho calcular a esperança de uma variável aleatória X para depois obter uma informação sobre seus possíveis valores, já que na definição (1) precisamos conhecer a distribuição de X a priori para calcular a esperança. Porém, como foi observado na seção anterior, com a linearidade da esperança em mãos talvez possamos calcular $\mathbb{E}(X)$ sem necessariamente saber sua distribuição. E de fato é isso que faremos!

Exemplo 1 (SJSU Midterm) Prove que qualquer subgrafo de $K_{n,n}$ (o grafo bipartido completo de $2n$ vértices) com pelo menos $n^2 - n + 1$ arestas tem um pareamento perfeito.

Solução: Fixe G um subgrafo qualquer de $K_{n,n}$ com pelo menos n^2-n+1 arestas. Tome $(i_1, j_1), (i_2, j_2), \cdots, (i_n, j_n)$ um pareamento escolhido aleatória e uniformemente entre todos os pareamentos possíveis e defina a variável aleatória $X = |\{k\,;\, \{i_k, j_k\} \in E(G)\}|$. Queremos então mostrar que $X = n$ com probabilidade positiva e para isso basta mostrar que $\mathbb{E}(X) > n - 1$. Note que $X = X_1 + X_2 + \cdots + X_n$ onde:

$$X_k = \begin{cases} 1\ , & \text{se } \{i_k, j_k\} \in E(G) \\ 0\ , & \text{caso contrário} \end{cases}$$

Como cada par (i_k, j_k) é escolhido uniformemente entre todos os n^2 possíveis, temos que $\mathbb{E}(X_k) = |E(G)|/n^2 \geq \frac{n^2-n+1}{n^2}$ para todo k. Pela linearidade da esperança segue então que $\mathbb{E}(X) \geq \frac{n^2-n+1}{n} = n - 1 + 1/n > n - 1$, como queríamos.

$\square$

Problema 5 (IMC2002) Uma olimpíada tem 6 problemas e 200 competidores. Cada competidor é bem esperto, então cada problema é resolvido por pelo menos 120 deles. Prove que existem dois competidores tais que cada problema foi resolvido por pelo menos um deles.

Problema 6 Numa tabela $n \times n$, cada um dos números $1, 2, ..., n$ aparece exatamente n vezes. Prove que existe uma linha ou uma coluna com pelo menos $\sqrt{n}$ números distintos.

Problema 7 (Número de Ramsey) Seja $k \geq 3$ e n inteiros tais que $\binom{n}{k} 2^{1-\binom{k}{2}} < 1$. Mostre que é possível colorir K_n (o grafo completo de n vértices) com duas cores de modo que *não* é possível encontrar k vértices tais que todas as $\binom{k}{2}$ arestas entre eles têm a mesma cor.

OBS: Na linguagem dos números de Ramsey, isto diz que $R(k,k) > n$. Usando a desigualdade $\binom{n}{k} < \frac{1}{e}(\frac{en}{k})^k$ (onde $e = 2,718...$ é a constante de Euler) e escolhendo n melhor possível, obtemos de fato:

$$R(k,k) > \frac{1}{e\sqrt{2}}(1+o(1))k2^{k/2}$$

2.2 Técnicas mais sofisticadas

A primeira técnica que veremos aqui é chamada de **alteração**. Ela consiste basicamente de selecionar (sempre de forma aleatória) o que desejamos e depois retirar os possíveis problemas que encontrarmos. O exemplo abaixo deixará isso mais claro:

Exemplo 2 (Turán Fraco) Seja G um grafo com grau médio d. Prove que existe um *conjunto independente* (isto é, um conjunto de vértices tal que G não possui nenhuma aresta entre eles) com cardinalidade pelo menos $n/2d$.

Solução: Em vez de selecionar aleatoriamente um subconjunto de $n/2d$ vértices, e torcer para o número esperado de arestas é menor que 1 (verifique que isso não funciona!), faremos algo mais esperto: vamos selecionar cada vértice independentemente e com probabilidade p (para uma boa escolha de p). Isto nos dará um conjunto de vértices cuja cardinalidade será aleatória. Depois retiraremos os vértices que "dão problema"e provaremos que o número esperado de vértices que sobram é pelo menos $n/2d$, terminando a demonstração.
Se selecionamos cada vértice com probabilidade p, a número esperado de vértices selecionados será np. Agora, como G tem $nd/2$ arestas, o número esperado de "problemas"(isto é, pares de vértices que selecionamos e que são ligados por uma aresta em G) será $ndp^2/2$. Portanto, se para cada "aresta problemática"retiramos um dos seus extremos do conjunto de vértices selecionados, ao fim teremos um *conjunto independente* aleatório A que tem cardinalidade esperada igual a

$$np - \frac{1}{2}ndp^2 = np\left[1 - \frac{1}{2}dp\right]$$

Tomando $p = 1/d$ segue o resultado.

□

Teorema (Desigualdade de Markov) Seja X é uma variável aleatória tomando valores não negativos. Suponha que $\mathbb{E}(X) = c$. Então

$$\mathbb{P}(X \geq r) \leq \frac{c}{r}$$

Prove o teorema acima. O próximo resultado é bem mais técnico, por isso omitirei a prova.

Teorema (Lema Local de Lovász) Considere $A_1, A_2, \cdots, A_n$ eventos com $\mathbb{P}(A_i) \leq p$ para todo i. Suponha que cada evento é independente de todos os demais exceto no máximo d deles, com $d \geq 1$. Então se $epd \leq 1$, temos necessariamente que $\mathbb{P}(A_1 \cup A_2 \cup \cdots \cup A_n) < 1$ (isto é, com probabilidade positiva nenhum dos eventos A_i ocorre).

Se todos os eventos A_i são independentes, então é claro que basta termos $p < 1$: neste caso sabemos que $\mathbb{P}(A_1 \cup A_2 \cup \cdots \cup A_n) = 1 - \mathbb{P}(A_1^c \cap A_2^c \cap \cdots \cap A_n^c) \geq 1 - (1-p)^n < 1$. Por outro lado, se permitirmos que todos os eventos A_i dependam uns dos outros (em particular, eles poderiam ser todos disjuntos), então só garantimos o resultado se tivermos $p < 1/n$: neste caso $\mathbb{P}(A_1 \cup A_2 \cup \cdots \cup A_n) \leq \sum_{k=1}^{n} \mathbb{P}(A_k) \leq np < 1$. Portanto, podemos entender o **Lema Local de Lovász (LLL)** como um caso intermediário, entres os dois extremos descritos acima.

Exemplo 3 (Rússia 2006) Num acampamento de turismo, cada pessoa tem pelo menos 50 e no máximo 100 amigos entre as demais pessoas no acampamento. Mostre que é possível dar uma camisa para cada pessoa de forma que as camisas tem (no máximo) 1331 cores diferentes e qualquer pessoa tem pelo menos 20 amigos cujas camisas tem cores todas diferentes.

Solucão: O número de cores $C = 1331$ é extremamente fraco. Provaremos de fato que C=48 é suficiente.
Dê a cada pessoa uma cor de camisa entre C possíveis, de forma aleatória e uniforme. Para cada pessoa P seja E_P o evento "os amigos de P têm no máximo 19 cores diferentes de camisa". Queremos então mostrar que, com probabilidade positiva, nenhum dos eventos E_P ocorre. Para isso usaremos o LLL. Primeiro encontramos uma cota para d: note que se duas pessoas A e B não são amigas nem têm amigos em comum, então E_A e E_B são independentes; como cada pessoa tem no máximo 100 amigos, cada evento E_P é independente de todos exceto

no máximo $100 + 100 \cdot 99 = 100^2$ outros eventos. Logo podemos tomar $d = 100^2$.
Agora, se uma pessoa P tem exatamente k amigos, $50 \leq k \leq 100$, então podemos calcular:

$$\mathbb{P}(E_P) = \binom{C}{19} \cdot \left(\frac{19}{C}\right)^k$$

Usando agora as desigualdades $\binom{n}{m} < \frac{1}{e}\left(\frac{en}{m}\right)^m$ e $k \geq 50$, vemos que isso é menor que

$$\frac{1}{e}\left(\frac{eC}{19}\right)^{19} \cdot \left(\frac{19}{C}\right)^{50} = e^{18}\left(\frac{19}{C}\right)^{31}$$

Logo podemos tomar $p = e^{18}\left(\frac{19}{C}\right)^{31}$. Pelo LLL, precisamos apenas que:

$$epd = e^{19}\left(\frac{19}{C}\right)^{31} \cdot 100^2 \leq 1$$

O menor C para o qual a desigualdade acima é satisfeita é $C = 48$, mas verificar que vale para $C = 1331$ é trivial.

$\square$

Problema 8 Seja n um inteiro positivo. Suponha que tenhamos $11n$ pontos em um círculo, coloridos com n cores, tal que cada cor aparece exatamente 11 vezes. Prove que podemos escolher um ponto de cada cor de forma que não existem dois pontos adjacentes entre eles.

Problema 9 (Ramsey com alteração) Use alteração para mostrar que $R(k,k) > n - \binom{n}{K}2^{1-\binom{k}{2}}$ para todos k, n.
OBS: Otimizando em n, esta cota nos dá:

$$R(k,k) > \frac{1}{e}(1 + o(1))k2^{k/2}$$

Problema 10 (Ramsey com LLL) Use LLL para mostrar que se $e\left(\binom{k}{2}\binom{n}{k-2} + 1\right)2^{1-\binom{n}{k}} < 1$, então $R(k,k) > n$.
OBS: Otimizando em n, esta cota nos dá:

$$R(k,k) > \frac{\sqrt{2}}{e}(1 + o(1))k2^{k/2}$$

Esta é a melhor cota inferior para $R(k,k)$ conhecida até hoje!

3 Mais Problemas

Problema 11 Mostre que, para n suficientemente grande, é possível construir um torneio com mais de n pessoas tal que em qualquer conjunto de 1000 delas, algum competidor ganhou de todos os outros.

Problema 12 (BAMO 2004) Mostre que, dados n números reais que somam 0, podemos ordená-los como $a_1, a_2, \cdots, a_n$ de modo que

$$a_1a_2 + a_2a_3 + \cdots + a_{n-1}a_n < 0$$

Problema 13 (Rússia 1996) Na Duma existem 1600 delegados, que devem formar 16000 comitês de 80 pessoas cada. Prove que é possível encontrar dois comitês com pelo menos 4 membros em comum.

Problema 14 (Romênia 2004) Prove que para quaisquer números complexos $z_1, z_2, \cdots, z_n$ satisfazendo $|z_1|^2 + \cdots |z_n|^2 = 1$, podemos encontrar $\epsilon_1, \epsilon_2, \cdots, \epsilon_n \in \{-1, 1\}$ tais que

$$\left|\sum_{k=1}^{n} \epsilon_k z_k\right| \leq 1$$

Problema 15 (Shortlist 1994 C4) Seja A um conjunto de n restos (mod n^2). Prove que existe um conjunto B de n restos (mod n^2) tal que $A+B = \{a+b \mid a \in A, b \in B\}$ contém pelo menos metade de todos os restos (mod n^2).

Problema 16 (USAMO 2012/6) Para um inteiro $n \geq 2$, seja $x_1, x_2, \cdots, x_n$ números reais tais que $x_1 + x_2 + \cdots + x_n = 0$ e $x_1^2 + x_2^2 + \cdots + x_n^2 = 1$. Para cada subconjunto $A \subseteq \{1, 2, \cdots, n\}$ defina

$$S_A = \sum_{i \in A} x_i$$

com a convenção de que $S_\emptyset = 0$. Prove que para cada número positivo λ, o número de conjuntos A tais que $S_A \geq \lambda$ é no máximo $2^{n-3}/\lambda^2$. Para que escolhas de $x_1, x_2, \cdots, x_n, \lambda$ a igualdade ocorre?

Problema 17 (Irã TST 2008/6) Suponha que 799 times participam de um torneio. Prove que podemos encontrar dois conjuntos disjuntos de A e B de 7 elementos cada tais que todos os times de A ganharam de todos os times de B.

Problema 18* (Teorema de Caro-Wei) Dado G um grafo, prove que podemos encontrar um *conjunto independente* com cardinalidade pelo menos

$$\sum_{v \in V(G)} \frac{1}{deg\, v + 1}$$

OBS: Aplicando Jensen, nosso conjunto independente tem tamanho pelo menos $\frac{n}{d+1}$, onde d é o grau médio, o que melhora a cota obtida no Exemplo 2. Este resultado é chamado Teorema de Turán.

Problema 19* (Rússia 1999) Numa certa escola, todo garoto gosta de pelo menos uma garota. Prove que podemos achar um conjunto S com pelo menos metade dos estudantes da escola de forma que cada garoto em S gosta de um número ímpar de garotas em S.

Problema 20* (Suécia 2010, adaptado) Numa cidade existem n pessoas, quaisquer duas se conhecem ou conhecem alguém em comum. Prove que podemos encontrar um grupo de pelo menos $\sqrt{n \log n} + 1$ pessoas tal que todas as demais pessoas conhecem pelo menos uma pessoa deste grupo.

Problema 21* (Sperner) Considere N subconjuntos distintos $S_1, S_2, \cdots, S_N$ de $\{1, 2, ..., n\}$ tais que $S_i \subsetneq S_j$ para todos $i \neq j$. Prove que

$$N \leq \binom{n}{\lfloor n/2 \rfloor}$$

Problema 22* (Erdös-Ko-Rado) Considere N subconjuntos de distintos $S_1, S_2, \cdots, S_N$ de $\{1, 2, ..., n\}$ tais que $|S_i \cap S_j| \neq \emptyset$ e $|S_i| = k$ para todos i, j. Prove que

$$N \leq \binom{n-1}{k-1}$$

Problema 23* (Erdös) Prove que em qualquer conjunto S de n inteiros positivos distintos podemos sempre encontrar um subconjunto T com $n/3$ ou mais elementos com a propriedade que $a + b \neq c$ para todos $a, b, c \in T$. (Tal conjunto é dito *livre de soma*).