Aplicações inesperadas do valor esperado

Thiago Landim

Janeiro 2020

1 Definições e Exemplos

Uma **variável aleatória** pode ser vista apenas como um valor que varia aleatoriamente. Mais especificamente, qualquer função pode ser vista como uma variável aleatória. Por exemplo, um dado D pode ter como valores 1, 2, 3, 4, 5 ou 6, uma moeda M pode dar *cara* ou *coroa*, e uma mão do Poker pode ser qualquer conjunto de 5 cartas do baralho.

A um conjunto de valores (que chamamos de **evento**), nós podemos associar uma **probabilidade**, denotada por $\mathbb{P}(\cdot)$. Por exemplo, se nossa moeda não está viciada, então

$$\mathbb{P}(M = \text{cara}) = \mathbb{P}(M = \text{coroa}) = \frac{1}{2},$$

se nosso dado não está viciado, então

$$\mathbb{P}(D \text{ é par}) = \frac{1}{2},$$

e a probabilidade de nossa mão no Poker ser o Royal Flush é $\frac{1}{649.740}$.

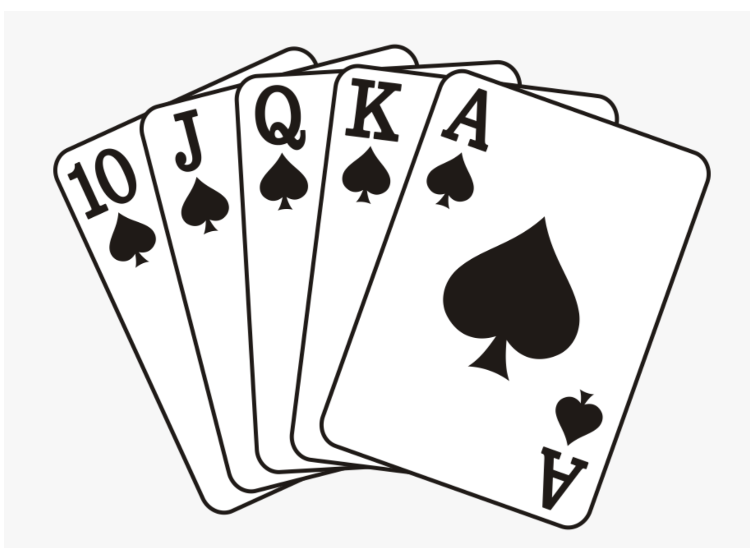

Figura 1: A probabilidade de você ter um Royal Flush no Poker é $\frac{1}{649.740}$.

Nosso objeto de estudo será o que chamamos de **valor esperado**. Digamos que João proponha a Maria a seguinte aposta. Ele lançará uma moeda e, se der cara, João pagará 10 reais; se der coroa, Maria pagará 5 reais. Como ela deve julgar se aceita ou não a proposta? Bem, imagine que, no lugar de uma única moeda, João lançasse 200. Em aproximadamente 100 delas, o resultado seria coroa, e em aproximadamente 100 delas, o resultado seria cara. Então Maria teria de pagar 500 reais, mas receberia 1000. Ou seja, teria um lucro de 500 reais! Assim, ela deve apostar com João. (Bem, mas será que João realmente faria essa aposta se a moeda fosse justa?...)

Para decidirmos se Maria deveria ou não aceitar a proposta, nós imaginamos MUITAS moedas sendo lançadas. Isso ocorre pois imaginamos (e esse fato recebe o nome pomposo de *Lei dos Grandes Números*) que, ao repetirmos o mesmo processo diversas vezes, a distribuição dos eventos será bem aproximada pela nossa probabilidade. Assim, é natural acreditar que podemos caracterizar o processo acima por meio apenas da variável aleatória e da probabilidade.

Se temos uma variável aleatória X (e sempre denotamos por letras maiúsculas), então a **valor esperado** de X é a média de todos os seus valores (sendo o peso de cada evento a sua probabilidade):

$$\mathbb{E}[X] = \sum_x x\mathbb{P}(X = x).$$

Essa fórmula pode parecer feia, mas vejamos como ela se comporta com os exemplos anteriores. Para o nosso dado D, o valor esperado é

$$\mathbb{E}[D] = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 3,5,$$

enquanto que o valor esperado para Maria em sua aposta A é

$$\mathbb{E}[A] = 10 \cdot \frac{1}{2} - 5 \cdot \frac{1}{2} = 2,5.$$

(É claro, assumindo a honestidade de João.)

Por outro lado, imagine que temos N casais em Recife, e que há n_k casais com k filhos. Quantas crianças nós temos em Recife?

$$C = 1 \cdot n_1 + 2 \cdot n_2 + 3 \cdot n_3 + \cdots .$$

E temos, em média, $\frac{C}{N}$ filhos por casal. Se F é a variável aleatória que a cada casal associa a quantidade de filhos que ele possui, então

$$\mathbb{E}[F] = \frac{C}{N} = 1 \cdot \frac{n_1}{N} + 2 \cdot \frac{n_2}{N} + 3 \cdot \frac{n_3}{N} + \cdots ,$$

mais um indício que nossa definição faz sentido.

Para um exemplo mais matemático, vejamos uma **variável aleatória de Bernoulli**. Seja E um evento qualquer (por exemplo, se no dia x choveu) e considere a variável aleatória X definida por

$$X = \begin{cases} 1, & \text{se } E \text{ ocorre,} \\ 0, & \text{caso contrário.} \end{cases}$$

Então $\mathbb{E}[X] = 1 \cdot \mathbb{P}[X = 1] + \cancel{0 \cdot \mathbb{P}[X = 0]} = \mathbb{P}[E \text{ ocorre}]$.

Esse exemplo é muito importante, pois é um caso em que sabemos calcular o valor esperado, e vamos em geral tentar reduzir nossas variáveis aleatórias a essa.

Problema 1.1 (Problema dos dois filhos 1.0)**.** Suponha que Clara tem dois filhos, e que o mais velho é um menino. Qual a probabilidade de que ambos os filhos sejam meninos?

Problema 1.2 (Problema dos dois filhos 2.0)**.** Suponha que Clara tem dois filhos, um dos quais é um menino. Qual a probabilidade de que ambos os filhos sejam meninos?

Problema 1.3 (Problema dos dois filhos 3.0)**.** Suponha que Clara tem dois filhos, um dos quais é um menino que nasceu em uma segunda-feira. Qual a probabilidade de que ambos os filhos sejam meninos?

Problema 1.4 (Dois envelopes)**.** Suponha que você receba dois envelopes, um com duas vezes mais dinheiro que o outro. Você pode abrir um envelope, olhar o que tem dentro e depois decidir se troca ou não de envelope. Qual a melhor estratégia?

Problema 1.5 (Ao trabalho!)**.** Suponha que José vá ao trabalho todo dia de ônibus, que passa em intervalos de 15min ou 45min, com igual probabilidade. Qual o tempo médio que João passa esperando o ônibus?

2 Propriedades Básicas

Vejamos agora como se comporta o valor esperado e como podemos utilizá-lo para resolver alguns problemas.

Exemplo 2.1 (Amigo Secreto)

No Natal, é comum as famílias fazerem um amigo secreto. Em geral, os nomes são escritos em pequenos pedaços de papel e cada um pega um papel de maneira aleatória. Infelizmente, é possível que você pegue o seu próprio nome (e terá de pegar um outro papel). Em uma família composta por 20 pessoas, qual a quantidade média de pessoas que irá pegar o próprio nome?

Solução. A chave é olhar o sorteio como uma permutação dos familiares. Vamos fazer para três pessoas, e esperamos que o leitor consiga adaptar o argumento para 20 familiares. Listamos todas as $3! = 6$ permutações, e vemos que queremos somar todas as vezes em que B aparece na coluna do B, e assim por diante.

B	G	T	Σ
B	G	T	3
B	T	G	1
G	B	T	1
T	B	G	0
G	T	B	0
T	G	B	1
2	2	2	6

No lugar de somar cada linha (que parecem irregulares), nós podemos somar cada coluna. Neste caso, a soma fica fácil, pois temos $2! = 2$ B's em cada coluna, e o mesmo para G e T. Vemos assim que ficamos com um total de $2 + 2 + 2 = 6$ pessoas sorteando si mesmas. Porém temos 6 sorteios. Logo, em média, temos uma única pessoa sorteando si mesma. Isso ocorrerá também para 20 familiares. □

2.1 Linearidade

A propriedade mais importante do valor esperado é a seguinte.

Teorema 2.2 (Linearidade do Valor Esperado)

Dadas $X_1,\ X_2, \ldots, X_n$ variáveis aleatórias, então

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

Em geral, buscamos separar nossa variável aleatória X em outras X_k cujo valor esperado é fácil de calcular. Para dar nosso primeiro exemplo, vamos abstrair o anterior para o seguinte.

Exemplo 2.3 (Amigo Secreto 2.0)

Calcule a quantidade média de pontos fixos de uma permutação aleatória de n elementos.

Solução. Suponhas que o conjunto dado seja $\{1, 2, \dots, n\}$, e seja

$$X_k = \begin{cases} 1, & \text{se } k \text{ é fixado na permutação,} \\ 0, & \text{caso contrário.} \end{cases}$$

Então a quantidade de pontos fixos é dada por

$$P = X_1 + X_2 \cdots + X_n.$$

Note, então, que $\mathbb{E}[X_k] = \frac{1}{n}$, pois das $n!$ permutações, $(n-1)!$ fixam k. Assim,

$$\mathbb{E}[P] = \mathbb{E}[X_1] + \mathbb{E}[X_2] + \cdots + \mathbb{E}[X_n] = n \cdot \frac{1}{n} = 1.$$

□

Antes de apresentar o próximo problema, vamos relembrar um resultado conhecido. Uma progressão geométrica (PG) é uma sequência cuja razão entre dois termos consecutivos é constante. Ela pode ser escrita, de maneira mais geral, na forma a, ar, ar^2, ar^3, $\dots$. Se $-1 < r < 1$, podemos calcular a soma de uma PG infinita. Note que, se S representa essa soma, então

$$\begin{aligned} S &= a + ar + ar^2 + ar^3 + \cdots \\ -\,(rS &= \phantom{a+{}} ar + ar^2 + ar^3 + \cdots) \\ \hline (1-r)S &= a, \end{aligned}$$

de onde segue que $S = \frac{a}{1-r}$.

Existe um outro tipo de sequência menos conhecida, chamada progressão aritmético-geométrica (PAG). Ela é definida como o produto de uma PA com uma PG. Eu irei apresentar o caso mais simples, onde a PA é dada por $1, 2, 3, \dots$. Assim, nossa sequência terá a cara a, $2ar$, $3ar^2$, $\dots$. Novamente queremos calcular o valor da soma S quando $-1 < r < 1$. Usaremos o mesmo truque:

$$\begin{aligned} S &= a + 2ar + 3ar^2 + 4ar^3 + \cdots \\ -\,(rS &= \phantom{a+{}} ar + 2ar^2 + 3ar^3 + \cdots) \\ \hline (1-r)S &= a + \ ar + \ ar^2 + \ ar^3 + \cdots = \frac{a}{1-r} \end{aligned}$$

de onde segue que $S = \frac{a}{(1-r)^2}$.

Exemplo 2.4 (Problema do Colecionador de Figurinhas)

Suponha que uma criança possua um álbum de figurinhas da Copa com 682 espaços e que, a cada 5 dias, ela compre um pacote com 5 figurinhas. Em quantos dias é esperado que ela complete o álbum?

Esse problema também pode ser enunciado da seguinte forma (alterando o valor 682 para 365): qual a quantidade média de amizades que eu devo fazer para ter festas de aniversário em todos os dias do ano?

Solução. Vamos supor a criança descobre uma única figurinha todo dia. Seja T o tempo necessário para completar o álbum e T_k o tempo necessário para, já possuindo $k-1$ figurinhas, encontrar uma diferente. Então

$$T = T_1 + T_2 + \cdots + T_{682}.$$

Vamos estudar cada T_k separadamente. Já tendo colado $k-1$ figurinhas, a probabilidade da próxima não ser repetida é $p_k = 1 - \frac{k-1}{682} = \frac{682-k+1}{682}$. Assim, a figurinha nova aparece no n-ésimo dia se ele encontra $n-1$ figurinhas repetidas consecutivas, e a outra não é mais repetida. Isso ocorre com probabilidade $(1-p_k)^{n-1}p_k$. Assim, pela definição de valor esperado, vemos que

$$\mathbb{E}[T_k] = 1 \cdot p_k + 2 \cdot (1-p_k)p_k + 3 \cdot (1-p_k)^2 p_k + \cdots$$

é exatamente a soma de uma PAG onde $a = p_k$ e $r = 1 - p_k$, e sabemos que o valor é dado por

$$\mathbb{E}[T_k] = \frac{a}{(1-r)^2} = \frac{p_k}{p_k^2} = \frac{1}{p_k}.$$

Usando a linearidade, concluímos que o tempo esperado é

$$\mathbb{E}[T] = \frac{1}{p_1} + \frac{1}{p_2} + \cdots + \frac{1}{p_{682}} = 682 \cdot \left(\frac{1}{682} + \frac{1}{681} + \cdots + \frac{1}{1} \right) \approx 4450 \text{ dias.}$$

□

Problema 2.5 (HMMT 2006)**.** Em um berçário, 2006 bebês sentam em um círculo. De repente, cada bebê aleatoriamente cutuca o bebê à sua esquerda ou à sua direita. Qual a quantidade esperada de bebês não cutucados?

Problema 2.6 (Romênia 2004)**.** Mostre que, para quaisquer números complexos z_1, z_2, …, z_n satisfazendo $|z_1|^2 + |z_2|^2 + \cdots + |z_n|^2 = 1$, é possível selecionar $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n \in \{-1, 1\}$ tais que

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

2.2 Valor Esperado Como Média

Por ser uma média, o valor esperado nos permite garantir a existência de objetos que satisfazem algumas desigualdades.

Lema 2.7

Se X é uma variável aleatória, então existe um ponto tal que $X \geq \mathbb{E}[X]$ e um ponto tal que $X \leq \mathbb{E}[X]$.

Exemplo 2.8 (MOP 2007)

Em uma tabela 100×100, são escritos os números 1, 2, 3, …, 100, cada um aparecendo exatamente 100 vezes. Mostre que existe uma linha ou uma coluna da tabela com pelo menos 10 números diferentes.

Solução. Como desejamos, contar a quantidade de números diferentes, considere as variáveis aleatórias

$$X_n = \begin{cases} 1, & \text{se } n \text{ está na coluna/linha;} \\ 0, & \text{caso contrário.} \end{cases}$$

Então $X = X_1 + \cdots + X_{100}$ conta a quantidade de números distintos na linha ou na coluna, e

$$\mathbb{E}[X] = \mathbb{E}[X_1] + \cdots + \mathbb{E}[X_{100}].$$

Agora note que $\mathbb{E}[X_n] \geq \frac{20}{200} = \frac{1}{10}$. De fato, se n aparece em a_n linhas e em b_n colunas, então as 100 vezes que n aparece estão nas $a_n \cdot b_n$ casas determinadas por essas linhas e colunas, e daí segue que

$$\frac{a_n + b_n}{2} \geq \sqrt{a_n \cdot b_n} \geq 10.$$

Por fim, temos que

$$\mathbb{E}[X] = \mathbb{E}[X_1] + \cdots + \mathbb{E}[X_{100}] \geq \frac{1}{10} + \cdots + \frac{1}{10} = 10,$$

e, pelo lema, existe uma coluna ou uma linha tal que $X \geq \mathbb{E}[X] \geq 10$. □

Uma ferramenta muito útil quando desejamos usar o valor esperado é o que chamamos de **convexidade**. O exemplo canônico que o leitor deve ter em mente é a função $f(x) = x^2$.

Um função f é dita **convexa** se, para qualquer par de pontos no gráfico de f, o segmento de reta que liga esses dois pontos está acima do gráfico.

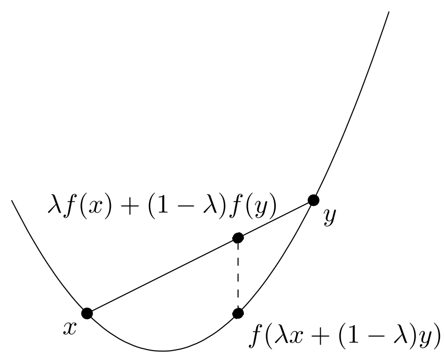

Figura 2: A função $f(x) = x^2$ é um exemplo de uma função convexa.

Podemos escrever matematicamente a definição usando desigualdades. Uma função é convexa se e somente se, para quaisquer x, y no domínio e $\lambda \in [0, 1]$,

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y).$$

Em outras palavras, se a função aplicada na média é menor ou igual do que a média da função. Isso se generaliza para o seguinte resultado.

Teorema 2.9 (Desigualdade de Jensen)

Seja f uma função convexa e $x_1, x_2, \ldots, x_n$ números reais. Se $\lambda_1, \lambda_2, \ldots, \lambda_n$ são reais positivos tais que $\lambda_1 + \cdots + \lambda_n = 1$, então

$$f(\lambda_1 x_1 + \cdots + \lambda_n x_n) \leq \lambda_1 f(x_1) + \cdots + \lambda_n f(x_n).$$

Escrito de maneira mais probabilística: para qualquer variável aleatória X,

$$f(\mathbb{E}[X]) \leq \mathbb{E}[f(X)].$$

Esse enunciado pode parecer obscuro, mas precisaremos apenas do nosso exemplo canônico para resolver o exemplo (embora há um exercício que utiliza uma função diferente). Se $x_1, \ldots, x_n$ são números reais, então

$$\frac{x_1^2 + \cdots + x_n^2}{n} \geq \left(\frac{x_1 + \cdots + x_n}{n}\right)^2.$$

Mais geralmente, isso é verdadeiro para toda função quadrática (cujo coeficiente líder é positivo). Assim, se $\binom{n}{2} = \frac{n(n-1)}{2}$ e $n_1, \ldots, n_m$ são números naturais que têm como média $\bar{n}$, então

$$\binom{n_1}{2} + \cdots + \binom{n_m}{2} \geq m\binom{\bar{n}}{2}.$$

Exemplo 2.10 (Rússia, 1998, adaptado)

Na Câmara dos Deputados do Brasil, há 513 deputados, que formam 2.052 alianças, cada uma com 50 pessoas. Mostre que é possível encontrar um par de alianças com cinco membros em comum.

Solução. Vamos seguir a mesma ideia das questões anteriores. Vamos escolher aleatoriamente um par de alianças (A_1, A_2) das $\binom{2052}{2}$ possíveis. Nossa variável aleatória será dada por $X =$ quantidade de membros em comum de A_1 e A_2. Assim, se

$$X_k = \begin{cases} 1, & \text{se o } k\text{-ésimo deputado estiver em } A_1 \text{ e em } A_2 \\ 0, & \text{caso contrário,} \end{cases}$$

então $X = X_1 + \cdots + X_{513}$. Agora nos resta calcular $\mathbb{E}[X_k]$. Para tanto, seja n_k a quantidade de alianças da qual participa o k-ésimo deputado. Então $\mathbb{E}[X_k] = \binom{n_k}{2}/\binom{2052}{2}$ e

$$\mathbb{E}[X] = \frac{1}{\binom{2052}{2}} \sum_{k=1}^{513} \binom{n_k}{2}.$$

Por outro lado, sabemos que $n_1 + n_2 + \cdots + n_{513} = 2.052 \cdot 50$, pois é a quantidade de pares (A, d) tais que o deputado d está na aliança A. Assim, $\bar{n} = 200$ e

$$\mathbb{E}[X] \geq \frac{1}{\binom{2052}{2}} \cdot 513\binom{200}{2} = 513\frac{200 \cdot 199}{2052 \cdot 2051} = 4.851\ldots$$

Pelo lema, isso implica que existe um par (A_1, A_2) tal que $X > 4.85$. Como esse valor precisa ser inteiro, nós temos um par de alianças com pelo menos 5 deputados em comum. □

Problema 2.11 (Rússia, 1999)**.** Em uma escola, cada menino é amigo de pelo menos uma menina. Mostre que existe um grupo com pelo menos metade dos estudantes onde cada menino do grupo é amigo de uma quantidade ímpar de meninas do grupo.

Problema 2.12. Em uma festa de formatura com n meninos e n meninas, há no máximo $n - 1$ relações de inimizade. Mostre que é possível, na hora do baile, cada menino convidar para dançar uma menina com a qual não possui inimizade.

Problema 2.13 (IMC 2002)**.** Uma olimpíada matemática tem 6 problemas e 200 participantes. Como os participantes eram muito habilidosos, cada questão foi resolvida por pelo menos 120 participantes. Mostre que existem dois participantes tais que cada problema foi resolvido por algum deles.

Problema 2.14 (Irã TST, 2008)**.** Suponha que 799 times participam de um torneio no qual cada time joga exatamente uma vez com todos os outros times. Mostre que existe dois grupos disjuntos A e B com 7 times cada tal que todos os times de A ganharam de todos os times de B.

3 Aplicações Inesperadas

Apresentaremos, aqui, dois problemas (o segundo talvez um pouco teórico) que podem ser apresentados a alunos do ensino médio.

Definimos o piso de um número real x como o maior interior $\leq x$, e denotamos por $\lfloor x \rfloor$. Por exemplo, $\lfloor \pi \rfloor = 3$ e $\left\lfloor \frac{17}{4} \right\rfloor = 4$. Semelhantemente, podemos definir a parte fracionária de x por $\{x\} := x - \lfloor x \rfloor$. Assim, $\{\pi\} = 0,14159...$ e $\left\{ \frac{17}{4} \right\} = 0,25$. Note que $0 \leq \{x\} < 1$.

Exemplo 3.1 (MOP 2008)
Sejam a, b e c números reais positivos tais que para todo inteiro n

$$\lfloor na \rfloor + \lfloor nb \rfloor = \lfloor nc \rfloor .$$

Prove que pelo menos um entre a, b e c é inteiro.

Solução. Primeiramente, note que podemos transferir esse problema para a parte fracionária. De fato, nós temos,

$$(na - \{na\}) + (nb - \{nb\}) = nc - \{nc\}$$

e podemos organizar os termos e dividir por n, para ficarmos com

$$a + b - c = \frac{\{nc\} - \{na\} - \{nb\}}{n}.$$

O lado esquerdo não depende de n, mas podemos deixar n arbitrariamente grande, de modo a deixar o lado direito arbitrariamente pequeno. Assim, $a + b = c$. Além disso, observe que nós não podemos considerar um número inteiro escolhido aleatoriamente (de maneira uniforme). Então vamos adaptar e, para cada N, iremos considerar um número aleatoriamente escolhido em $[-N, N]$ e construiremos as variáveis aleatórias A, B e C dadas por: se o número sorteado é n, então $A = \{na\}$, $B = \{nb\}$ e $C = \{nc\}$. Suponha que a, b e c **não** são inteiros; tentaremos chegar em um absurdo. Se a é irracional, então $\{na\}$ nunca é inteiro e $\{-na\} = 1 - \{na\}$. Assim,

$$\sum_{k=-N}^{N} \{na\} = \sum_{k=1}^{N} (\cancel{\{na\}} + (1 - \cancel{\{na\}})) = N,$$

e $\mathbb{E}[A] = \frac{N}{2N+1}$ que se aproxima de $\frac{1}{2}$ quando N cresce arbitrariamente. Por outro lado, se $a = \frac{p}{q}$ é um irracional, então $\{na\}$ pode dar os valores 0, $\frac{1}{q}$, $\frac{2}{q}$, $\ldots$, $\frac{q-1}{q}$. Assim, quando N cresce arbitrariamente, todos esses valores serão igualmente prováveis, e teremos que $\mathbb{E}[A]$ se aproxima de

$$\frac{\frac{1}{q} + \frac{2}{q} + \cdots + \frac{q-1}{q}}{q} = \frac{q(q-1)}{2q^2} = \frac{1}{2} - \frac{1}{2q}.$$

Portanto, $\mathbb{E}[A]$ se aproxima de $\frac{1}{2}$, se a é irracional; e $\mathbb{E}[A]$ se aproxima de um valor em $[\frac{1}{4}, \frac{1}{2})$, se a é racional, e

$$\mathbb{E}[C] = \mathbb{E}[A + B] = \mathbb{E}[A] + \mathbb{E}[B] \to \ell_a + \ell_b \in \left[\frac{1}{2}, 1\right].$$

Mas isso só é possível se $\mathbb{E}[C] \to \frac{1}{2}$ e $\mathbb{E}[A]$, $\mathbb{E}[B] \to \frac{1}{4}$. Mas então a e b são racionais, c é irracional, e $a + b = c$, o que é um absurdo! Portanto algum dos valores a, b ou c é inteiro. $\square$

Dizemos que um conjunto S é **livre de somas** se a soma de quaisquer dois elementos do conjunto não está no conjunto. Matematicamente, escrevemos que $a + b = c$ não tem solução com a, b, $c \in S$. Por exemplo, podemos tomar o conjunto dos números ímpares. Como a soma de dois ímpares é sempre um par, esse conjunto é livre de soma. Por outro lado, qualquer conjunto que contenha o 0 **não** é livre de somas, pois $0+0=0$. O conjunto dos quadrados $\{1,4,9,16,\ldots,\}$ não é livre de quadrados, porém o conjunto das potências cúbicas $\{1,8,27,64,\ldots\}$ é - esse é um caso particular do Último Teorema de Fermat.

Como comentamos acima, podemos formar um subconjunto de $\{1,2,\ldots,n\}$ livre de somas de tamanho $\lfloor \frac{n}{2} \rfloor$, basta tomar todos os ímpares ou o conjunto $\{\lfloor \frac{n}{2} \rfloor, \ldots, n-1, n\}$. Porém, esse conjunto é muito particular. Digamos que temos um conjunto A qualquer de números naturais e saber o quão grande pode ser um subconjunto livre de somas $B \subseteq A$. Esse é o tema do nosso próximo resultado.

Teorema 3.2 (Erdös, 1965)
Seja $A \subseteq \mathbb{N}$ um conjunto com n elementos. Então existe $B \subseteq A$ livre de somas com $> \frac{n}{3}$ elementos.

Solução. A ideia chave é usar aritmética modular para introduzir permutações. Seja $\bar{a}$ o maior elemento de A e tome $p > 2\bar{a}$ um número primo. Dessa forma, para $a,b,c \in A$,

$$a+b=c \iff a+b \equiv c \pmod{p},$$

o que nos permite nos preocupar apenas com aritmética modular, o que é bem mais simples! Suponha, para simplificar as contas, que $p = 3k+2$. Da mesma forma como fizemos para $\{1,2,\ldots,n\}$, nós podemos considerar o conjunto livre de somas $S = \{k+1, k+2, \ldots, 2k+1\}$ com $k+1$ elementos. E nós vamos permutar esse conjunto usando o produto por algum $x \in \mathbb{Z}/p\mathbb{Z}^*$ aleatório, pois

$$xa + xb \equiv xc \pmod{p} \iff a+b \equiv c \pmod{p}.$$

Considere, então, a variável aleatória dada por

$$X(x) = |xS \cap A|,$$

onde $xS = \{x(k+1), x(k+2), \ldots, x(2k+1)\}$. Novamente, podemos separar a nossa variável aleatória em $X = \sum_{a \in A} X_a$ onde

$$X_a = \begin{cases} 1, & \text{se } a \in xS, \\ 0, & \text{caso contrário.} \end{cases}$$

Agora, note que é fácil calcular o valor esperado de cada X_a, pois

$$a \in xS \iff x^{-1}a \in S,$$

de onde segue que $\mathbb{E}[X_a] = \mathbb{P}[a \in xS] = \frac{k+1}{3k+1} > \frac{1}{3}$. Portanto $\mathbb{E}[X] > \frac{n}{3}$ e existe um x tal que $|xS \cap A| > \frac{n}{3}$. Se tomarmos $B = xS \cap A$, temos o conjunto livre de somas desejado. □

4 Problemas Difíceis

Problema 4.1. Prove que é possível colorir os inteiros de 1 a 2014 com duas cores sem que haja 18 inteiros formando uma PA monocromática.

Problema 4.2 (Desigualdade LYM)**.** Sejam $A_1, \ldots, A_s$ subconjuntos de $\{1,2,\ldots,M\}$ tais que nenhum A_i e subconjunto de um outro A_j. Para cada i, chame $a_i = |A_i|$. Mostre que

$$\sum_{i=1}^{s} \frac{1}{\binom{M}{a_i}} \leq 1$$

Problema 4.3 (Lema de Sperner)**.** Mostre que, se $\mathcal{F}$ é uma coleção de subconjuntos de $\{1,2,\ldots,n\}$ tal que nenhum elemento de $\mathcal{F}$ é subconjunto de um outro elemento de $\mathcal{F}$, então $|\mathcal{F}| \leq \binom{n}{\lfloor n/2 \rfloor}$.

Problema 4.4 (Lema de Littlewood-Offord)**.** Sejam x_1, x_2, $\ldots$, x_n números reais não nulos não necessariamente distintos. Suponha que c_1, c_2, $\ldots$, c_n são variáveis aleatórias independentes, cada qual é ± 1 com igual probabilidade. Mostre que existe $C > 0$ tal que

$$\mathbb{P}[c_1 x_1 + \cdots + c_n x_n = 0] \leq \frac{C}{\sqrt{n}}.$$

Referências

[1] Alon, N. Spencer, J. *The Probabilistic Method*, Wiley, 2015. Capítulos iniciais disponíveis em: https://cs.nyu.edu/cs/faculty/spencer/nogabook/.

[2] Bellos, A. *Can you solve it? The two child problem*, The Guardian, 18 Nov 2019. Disponível em: https://amp.theguardian.com/science/2019/nov/18/can-you-solve-it-the-two-child-problem?.

[3] Chen, E. Critch, A. *Unexpected Expectations*, 2015. Disponível em: http://acritch.com/media/math/Andrew_Critch_and_Evan_Chen_-_Unexpected_Expectations.pdf.

[4] Chen, E. *Expected Uses of Probability*, 2014. Disponível em: https://web.evanchen.cc/handouts/ProbabilisticMethod/ProbabilisticMethod.pdf.

[5] Feller, W. *An Introduction to Probability Theory and Its Applications*, John Wiley & Sons, 1968. Disponível em: https://www.ime.usp.br/~fmachado/MAE5709/FellerV1.pdf

[6] Loh, P. S. *Probabilistic Methods in Combinatorics*, 2009. Disponível em: https://www.math.cmu.edu/~ploh/docs/math/mop2009/prob-comb.pdf.