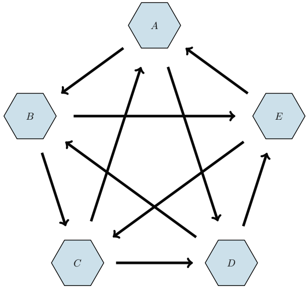
Jornada de pesquisa - Dados intransitivos

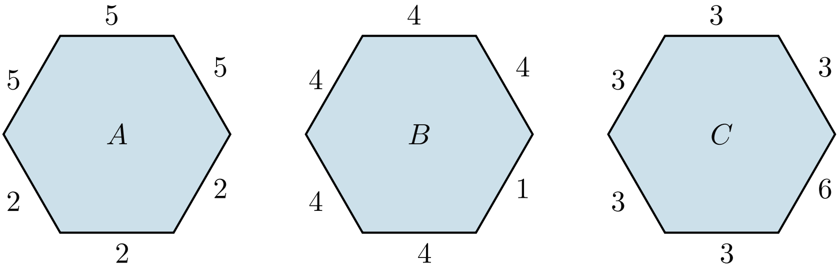


João Victor A. P.[1] João Pedro C. P.[2] Lael V. Lima[3] Luis G. Coelho[4]

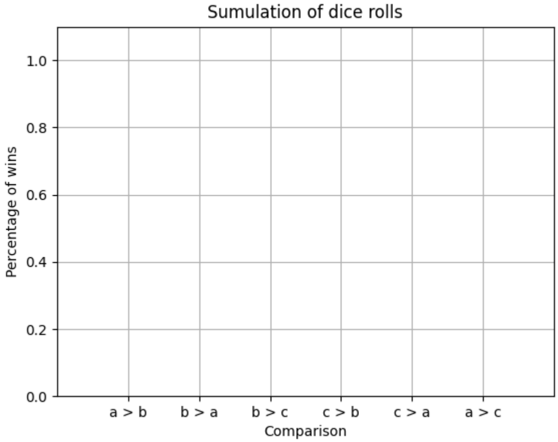
[1]IFSC,[4]FFCLRP - USP

[2, 3]IMECC - Unicamp

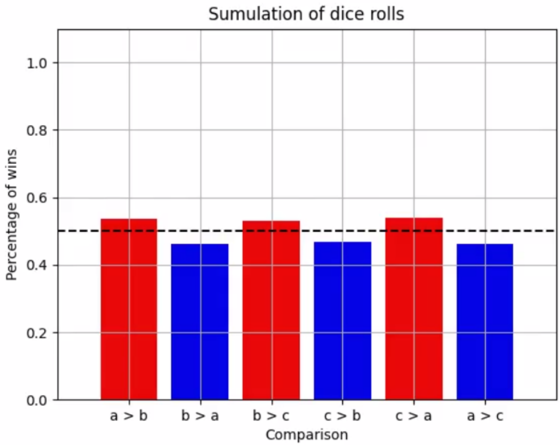
Dados Intransitivos

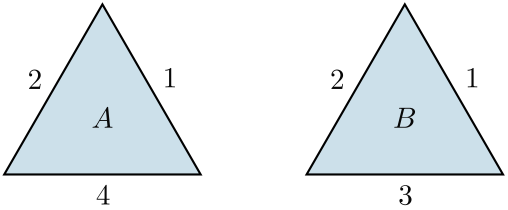

5
5
5
A
2
2
2
4
4
4
B
4
1
4
3
3
3
C
3
6
3

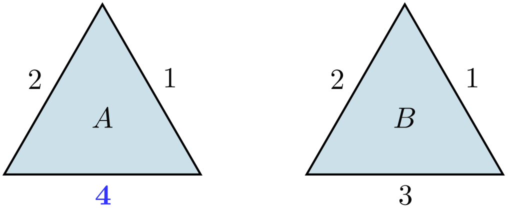
Número de Vitórias



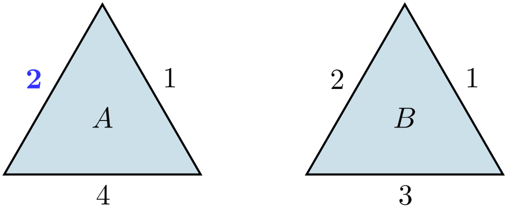
Número de Vitórias



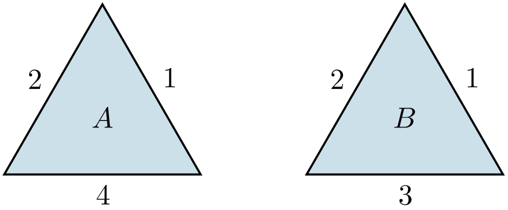

2
1
A
4
2
1
B
3



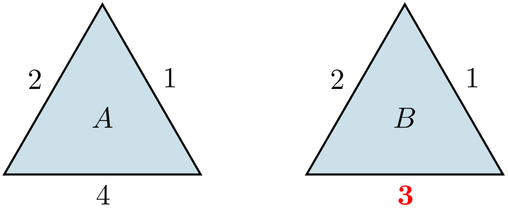
3



$3 + \mathbf{1}$



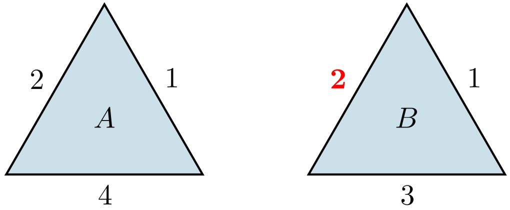
$3 + 1 = 4$



$3 + 1 = 4$

2

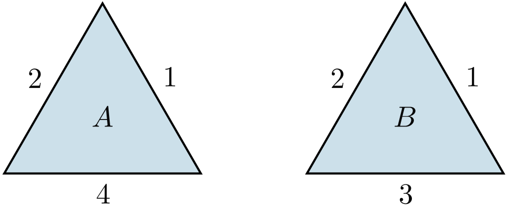
O Jogo



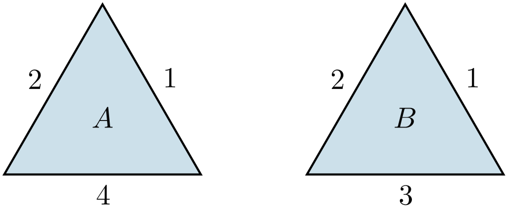
$3 + 1 = 4$

$2 + \mathbf{1}$

O Jogo



$$3 + 1 = 4$$
$$2 + 1 = 3$$



$3 + 1 = 4$
$2 + 1 = 3$

$$A \triangleright B$$

"Dado A melhor que o dado B"

Possíveis perguntas

- Condições de existência?
- Probabilidade?
- Possíveis modelos?

Quanta magazine — Physics | Mathematics | Biology | Computer Science | Topics | Archive

COMBINATORICS

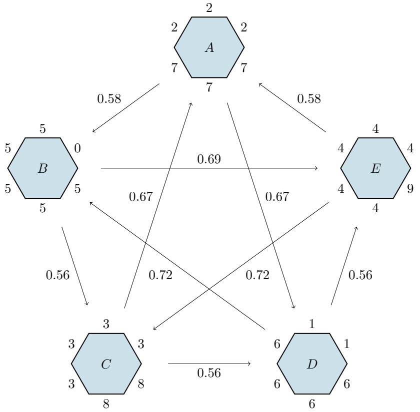
Mathematicians Roll Dice and Get Rock-Paper-Scissors

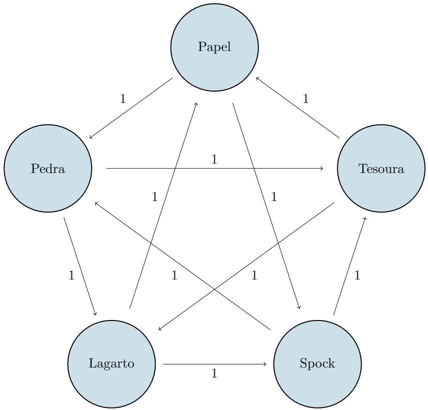
3

Mathematicians have uncovered a surprising wealth of rock-paper-scissors-like patterns in randomly chosen dice.

Referência do Quanta ao Pedra Papel Tesoura

Pedra, Papel e ... Tesoura?





É possível criar uma tripla de dados A, B e C similar ao jogo de pedra, papel e tesoura?

Isto é, A ganha quase sempre de B, B ganha quase sempre de C e C ganha quase sempre de A?

É possível criar uma tripla de dados A, B e C similar ao jogo de pedra, papel e tesoura?

Isto é, A ganha quase sempre de B, B ganha quase sempre de C e C ganha quase sempre de A?

NÃO!

Se para os dados A, B e C tivermos

$$\mathbb{P}(\rho(A) > \rho(B)) = 1 - \varepsilon$$

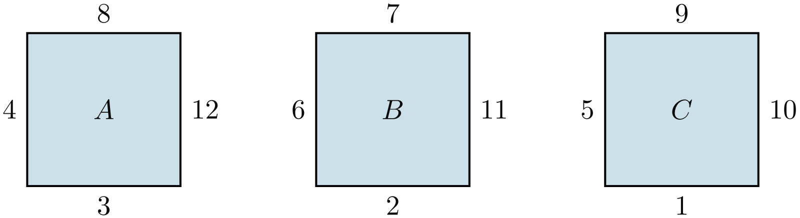
e

$$\mathbb{P}(\rho(B) > \rho(C)) = 1 - \varepsilon$$

teremos necessariamente

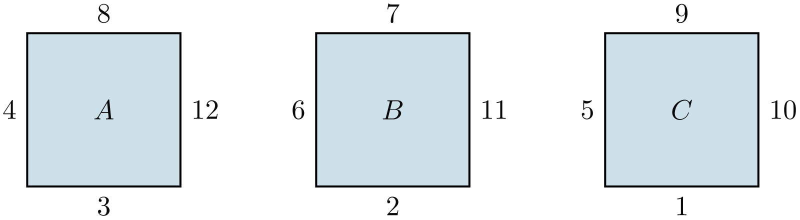
$$\mathbb{P}(\rho(C) > \rho(A)) < 2\varepsilon.$$

Dados & Palavras

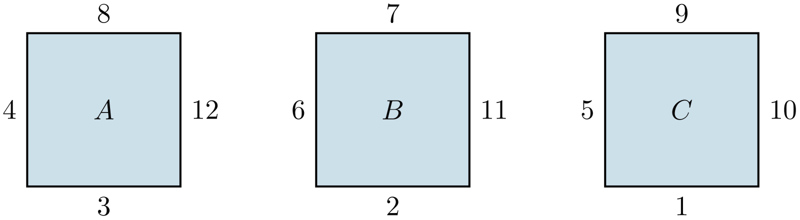


$\boldsymbol{A}$												
$\boldsymbol{B}$												
$\boldsymbol{C}$												

Equivalência

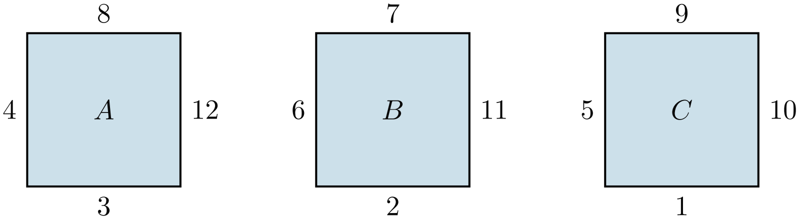


$\boldsymbol{A}$	12											
$\boldsymbol{B}$												
$\boldsymbol{C}$												
	A											

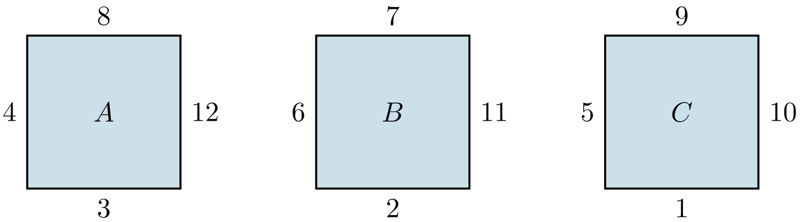


A	12											
B		11										
C												
	A	B										

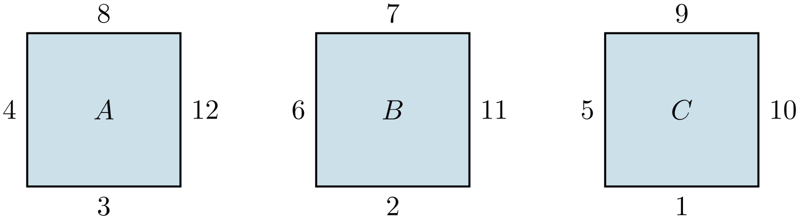
Equivalência



A	12											
B		11										
C			10									
	A	B	C									

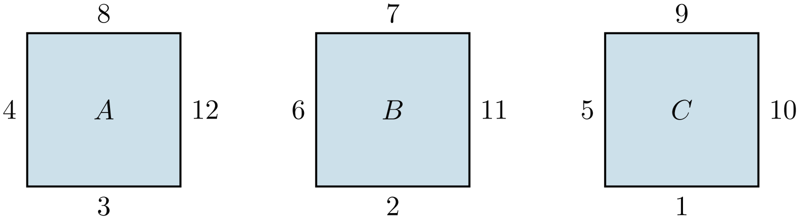


A	12											
B		11										
C			10	9								
	A	B	C	C								



A	12				8							
B		11										
C			10	9								
	A	B	C	C	A							

Equivalência



A	12				8				4	3		
B		11				7	6				2	
C			10	9				5				1
	A	B	C	C	A	B	B	C	A	A	B	C

Teorema

Problema dos Dados Intransitivos,
$PDI(m,n)$: Existem m dados com n faces que são intransitivos?

Problema dos Dados Intransitivos,
$PDI(m,n)$: Existem m dados com n faces que são intransitivos?

Proposição

Se $PDI(m,n)$ *tem solução, então* $PDI(m+1,n)$ *e* $PDI(m,n+2)$ *têm solução.*

Problema dos Dados Intransitivos,
$PDI(m, n)$: Existem m dados com n faces que são intransitivos?

Proposição

Se $PDI(m, n)$ tem solução, então $PDI(m + 1, n)$ e $PDI(m, n + 2)$ têm solução.

Teorema

$PDI(m, n)$ tem solução para todo $m, n \geq 3$.

Tabela

$n \backslash m$	1	2	3	4	5	$\dots$
1	$\times$	$\times$	$\times$	$\times$	$\times$	$\dots$
2	$\times$	$\times$	$\times$	$\times$	$\times$	$\dots$
3	$\times$	$\times$	$\checkmark$	$\checkmark$	$\checkmark$	$\dots$
4	$\times$	$\times$	$\checkmark$	$\checkmark$	$\checkmark$	$\dots$
5	$\times$	$\times$	$\checkmark$	$\checkmark$	$\checkmark$	$\dots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

Exemplo

Partimos de $ABCCABBCAABC$.

Partimos de $ABCCABBCAABC$.

Se quisermos adicionar um novo dado ao ciclo tal que ele seja melhor que B e pior que A, podemos inserir um D ao lado de cada A

$$A\mathbf{D}BCCA\mathbf{D}BBCA\mathbf{D}A\mathbf{D}BC$$

Partimos de $ABCCABBCAABC$.

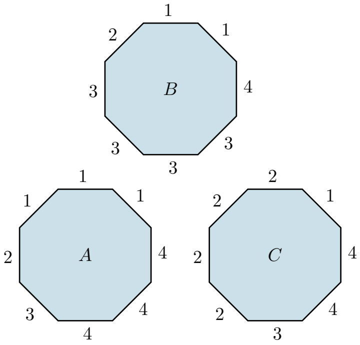
Se quisermos adicionar um novo dado ao ciclo tal que ele seja melhor que B e pior que A, podemos inserir um D ao lado de cada A

$$A\mathbf{D}BCCA\mathbf{D}BBCA\mathbf{D}A\mathbf{D}BC$$

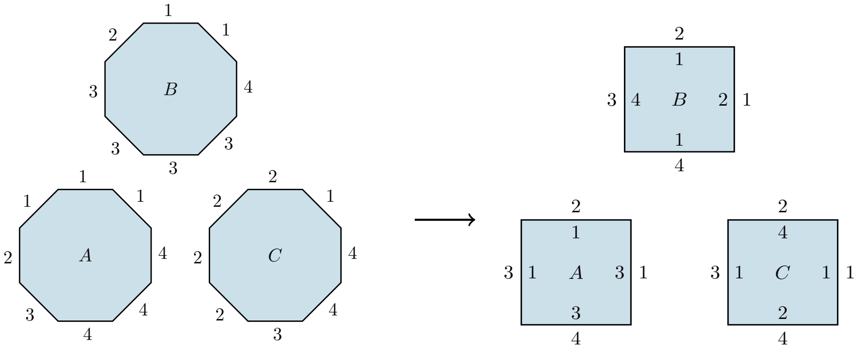
Se quisermos adicionar duas faces extras à cada um dos dados podemos adicionar uma *palavra neutra*.

$$\mathbf{ABCCBA}ABCCABBCAABC$$

Repetição?


1
2
1
3
B
4
3
3
3
1
1
1
2
A
4
3
4
4
2
2
1
2
C
4
2
4
3

Pesos e Repetição



Sejam A e B dados desonestos com n faces enumeradas de 1 a n.

Os dados A e B têm faces k com peso p_k e q_k.

Sejam A e B dados desonestos com n faces enumeradas de 1 a n.

Os dados A e B têm faces k com peso p_k e q_k.

Se $S_a = \sum_{k=1}^{n} p_k$ e $S_b = \sum_{k=1}^{n} q_k$, temos que

$$\mathbb{P}(\rho(A) > \rho(B)) = \frac{1}{S_a S_b} \sum_{i=2}^{n} \sum_{j=1}^{i-1} p_i q_j$$

Proposição

Não existe solução para $PDI(m, 3, \mathcal{M}_p)$*, onde* $\mathcal{M}_p$ *representa o modelo de pesos.*

Proposição

Não existe solução para $PDI(m, 3, \mathcal{M}_p)$*, onde* $\mathcal{M}_p$ *representa o modelo de pesos.*

Lema

Se $PDI(m, n, \mathcal{M}_p)$ *tem solução, então* $PDI(m+1, n, \mathcal{M}_p)$ *e* $PDI(m, n+1, \mathcal{M}_p)$ *tem soluções.*

Proposição

Não existe solução para $PDI(m, 3, \mathcal{M}_p)$*, onde* $\mathcal{M}_p$ *representa o modelo de pesos.*

Lema

Se $PDI(m, n, \mathcal{M}_p)$ *tem solução, então* $PDI(m+1, n, \mathcal{M}_p)$ *e* $PDI(m, n+1, \mathcal{M}_p)$ *tem soluções.*

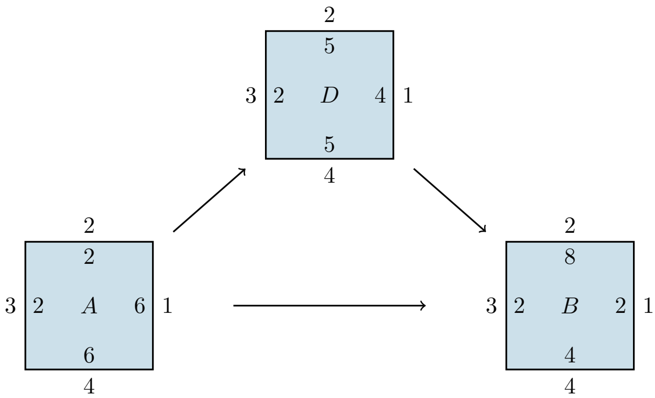
Teorema

$PDI(m, n, \mathcal{M}_p)$ *tem solução para qualquer* $m \geq 3$ *e* $n \geq 4$*.*

Tabela de existência

$n \backslash m$	1	2	3	4	5	$\dots$
1	$\times$	$\times$	$\times$	$\times$	$\times$	$\dots$
2	$\times$	$\times$	$\times$	$\times$	$\times$	$\dots$
3	$\times$	$\times$	$\times$	$\times$	$\times$	$\dots$
4	$\times$	$\times$	$\checkmark$	$\checkmark$	$\checkmark$	$\dots$
5	$\times$	$\times$	$\checkmark$	$\checkmark$	$\checkmark$	$\dots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

Construções Exemplo



Perceba que a função $f : \mathbb{R}^{2n+2} \to \mathbb{R}$ definida por

$$f(x, y) = \sum_{i=2}^{n+1} \sum_{j=1}^{i-1} x_i y_j - x_j y_i.$$

é contínua e, se $p = (p_1, p_2, \ldots, p_n, 0)$ e $q = (q_1, q_2, \ldots, q_n, 0)$, então $f(p, q) > 0$.

Perceba que a função $f : \mathbb{R}^{2n+2} \to \mathbb{R}$ definida por

$$f(x,y) = \sum_{i=2}^{n+1} \sum_{j=1}^{i-1} x_i y_j - x_j y_i.$$

é contínua e, se $p = (p_1, p_2, \ldots, p_n, 0)$ e $q = (q_1, q_2, \ldots, q_n, 0)$, então $f(p,q) > 0$. Por continuidade existe $\varepsilon > 0$ tal que A', de pesos $p = (p_1, p_2, \ldots, p_n, \varepsilon)$ e B', de pesos $q = (q_1, q_2, \ldots, q_n, \varepsilon)$ tal que $f(p,q) > 0$.

Distribuições

São realizações de eventos com probabilidades associadas. Exemplos:

- A face resultante do lançamento de uma moeda.
- O número resultante do lançamento de um dado.

Variáveis aleatórias representadas por números tem, além de probabilidades associadas à seus valores, outras propriedades

Variáveis aleatórias representadas por números tem, além de probabilidades associadas à seus valores, outras propriedades

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

Variáveis aleatórias representadas por números tem, além de probabilidades associadas à seus valores, outras propriedades

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

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

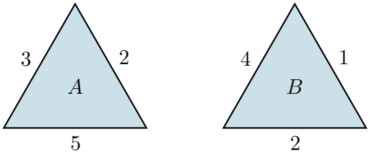
Variáveis aleatórias representadas por números tem, além de probabilidades associadas à seus valores, outras propriedades

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

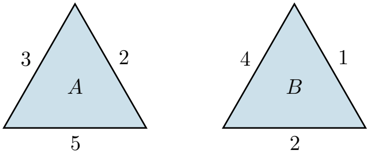
$$Var(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$$

$$Cov(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$$

Fixados dois dados de n faces, A e B



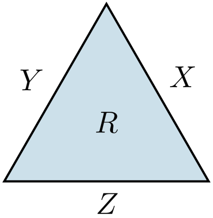
Fixados dois dados de n faces, A e B



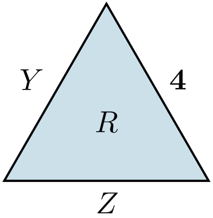
Podemos contar quantas vezes primeiro ganha do segundo com a seguinte expressão

$$N_{A>B} = \sum_{i=1}^{n} \sum_{j=1}^{n} \chi_{A_i > B_j}$$

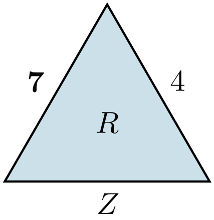
São dados cujas faces são variáveis aleatórias



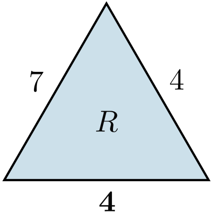
São dados cujas faces são variáveis aleatórias



São dados cujas faces são variáveis aleatórias



São dados cujas faces são variáveis aleatórias



Quando A e B são aleatórios, a quantidade de vitórias de um sobre outro também é.

https://www.overleaf.com/project/63d7dab32194b5ad2a6ba277

Considere $N_{A>B}$ como a variável aleatória que conta o número de vitórias do dado A sobre o dado B, cada dado com n faces com entradas aleatórias no intervalo $(0, 1)$.
Conseguimos determinar para nossa variável

Considere $N_{A>B}$ como a variável aleatória que conta o número de vitórias do dado A sobre o dado B, cada dado com n faces com entradas aleatórias no intervalo $(0, 1)$.
Conseguimos determinar para nossa variável

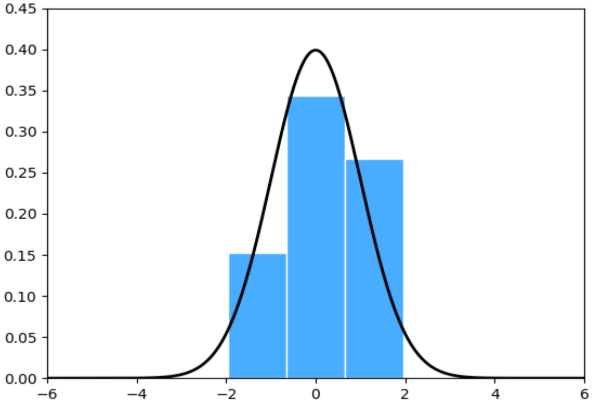
$$\mathbb{E}[N_{A>B}] = \frac{n^2}{2}.$$

Considere $N_{A>B}$ como a variável aleatória que conta o número de vitórias do dado A sobre o dado B, cada dado com n faces com entradas aleatórias no intervalo $(0, 1)$.
Conseguimos determinar para nossa variável

$$\mathbb{E}[N_{A>B}] = \frac{n^2}{2}.$$

$$\text{Var}\left(N_{A>B}\right) = \frac{n^2(2n+1)}{12}.$$

Histogramas de $\bar{N}_{A>B}$ quando n cresce



Teorema

A variável aleatória $\bar{N}_{A>B} = \frac{N_{A>B} - \mathbb{E}[N_{A>B}]}{\sqrt{\operatorname{Var} N_{A>B}}}$ *converge em distribuição para* $X \sim \mathcal{N}(0,1)$ *quando* $n \to \infty$.

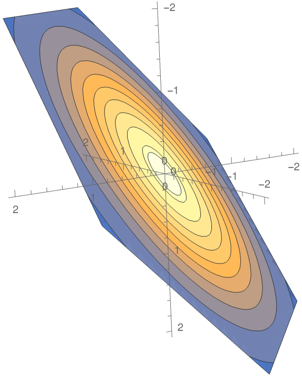
Vetor Normal

Considere, agora, o vetor $\bar{N} = (\bar{N}_{A>B}, \bar{N}_{B>C}, \bar{N}_{C>A})$, Considere, agora, o vetor $\bar{N} = (\bar{N}_{A>B}, \bar{N}_{B>C}, \bar{N}_{C>A})$,

Teorema

O vetor aleatório $\bar{N}$ converge em distribuição para (X, Y, Z) quando $n \to \infty$, onde $X, Y, Z \sim \mathcal{N}(0, 1)$ e

$$\mathrm{Cov}(X, Y) = \mathrm{Cov}(X, Z) = \mathrm{Cov}(Y, Z) = -1/2.$$



Função de distribuição

$$\frac{1}{3\pi}\mathrm{e}^{-\frac{2}{9}(x^2+y^2+z^2-xy-xz-yz)}$$

Plano $G: x+y+z=0$

Para mostrar

$$\bar{N}_{A>B} \to X,$$

Para mostrar

$$\bar{N}_{A>B} \to X,$$

é suficiente mostrar que

$$\mathbb{E}[(\bar{N}_{A>B})^t] \to \mathbb{E}[X^t],$$

para todo inteiro positivo t.

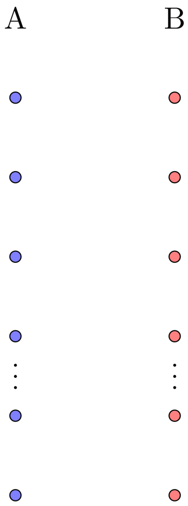
Ideia da Demonstração

$$\mathbb{E}[(\bar{N}_{A>B})^{2k}] = \sum_{i_1,j_1} \cdots \sum_{i_{2k},j_{2k}} \mathbb{E}[e_{i_1,j_1} \ldots e_{i_{2k},j_{2k}}]$$

$$\mathbb{E}[(\bar{N}_{A>B})^{2k}] = \sum_{i_1,j_1} \cdots \sum_{i_{2k},j_{2k}} \mathbb{E}[e_{i_1,j_1} \dots e_{i_{2k},j_{2k}}]$$

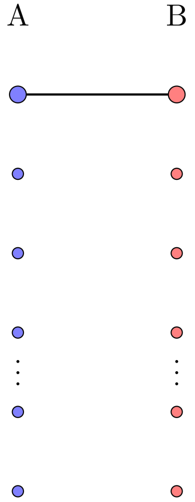
$$e_{1,1}\ e_{2,3}\ e_{2,1}\ \dots\ e_{i,j}$$

Ideia da Demonstração



$$\mathbb{E}[(\bar{N}_{A>B})^{2k}] = \sum_{i_1,j_1} \cdots \sum_{i_{2k},j_{2k}} \mathbb{E}[e_{i_1,j_1} \ldots e_{i_{2k},j_{2k}}]$$

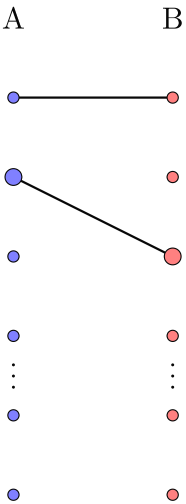
$$e_{1,1}\ e_{2,3}\ e_{2,1}\ \ldots\ e_{i,j}$$



$$\mathbb{E}[(\bar{N}_{A>B})^{2k}] = \sum_{i_1,j_1} \cdots \sum_{i_{2k},j_{2k}} \mathbb{E}[e_{i_1,j_1} \ldots e_{i_{2k},j_{2k}}]$$

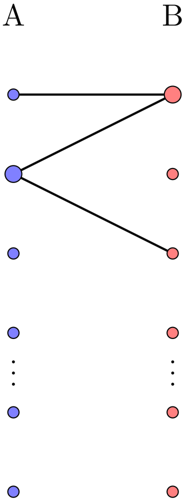
$$\boldsymbol{e_{1,1}} \; e_{2,3} \; e_{2,1} \; \ldots \; e_{i,j}$$

Ideia da Demonstração



$$\mathbb{E}[(\bar{N}_{A>B})^{2k}] = \sum_{i_1,j_1} \cdots \sum_{i_{2k},j_{2k}} \mathbb{E}[e_{i_1,j_1} \ldots e_{i_{2k},j_{2k}}]$$

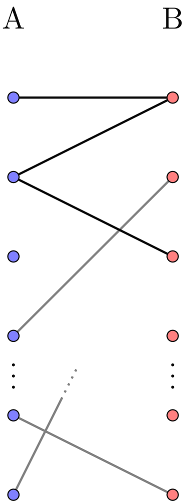
$$e_{1,1} \; \boldsymbol{e_{2,3}} \; e_{2,1} \; \ldots \; e_{i,j}$$



$$\mathbb{E}[(\bar{N}_{A>B})^{2k}] = \sum_{i_1,j_1} \cdots \sum_{i_{2k},j_{2k}} \mathbb{E}[e_{i_1,j_1} \ldots e_{i_{2k},j_{2k}}]$$

$$e_{1,1}\ e_{2,3}\ \boldsymbol{e_{2,1}}\ \ \ldots\ \ e_{i,j}$$

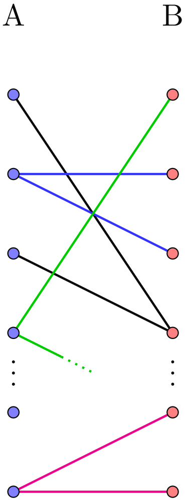
Ideia da Demonstração



$$\mathbb{E}[(\bar{N}_{A>B})^{2k}] = \sum_{i_1,j_1} \cdots \sum_{i_{2k},j_{2k}} \mathbb{E}[e_{i_1,j_1} \ldots e_{i_{2k},j_{2k}}]$$

$$e_{1,1}\ e_{2,3}\ e_{2,1}\ \ldots\ e_{i,j}$$

Ideia da Demonstração



$$\mathbb{E}[(\bar{N}_{A>B})^{2k}] = \sum_{i_1,j_1} \cdots \sum_{i_{2k},j_{2k}} \mathbb{E}[e_{i_1,j_1} \ldots e_{i_{2k},j_{2k}}]$$

$$e_{1,4}\ e_{4,3}\ e_{2,2}\ \ldots\ e_{2k,2k}$$

$$\mathbb{E}[(\bar{N}_{A>B})^{2k}] = \frac{(2k)!}{2^k k!}(1 + o(1)) = \mathbb{E}[X^{2k}](1 + o(1))$$

Distribuição de Palavras Intransitivas

$\mathcal{D}(n)$ é o número de palavras com n letras A, B e C.

$\mathcal{D}(n)$ é o número de palavras com n letras A, B e C.

$\mathcal{I}(n)$ é o número dessas palavras que são intransitivas.

$\mathcal{D}(n)$ é o número de palavras com n letras A, B e C.

$\mathcal{I}(n)$ é o número dessas palavras que são intransitivas.

Qual a proporção de palavras intransitivas

$$\frac{\mathcal{I}(n)}{\mathcal{D}(n)}?$$

É possível determinar que

$$\frac{\mathcal{I}(n)}{\mathcal{D}(n)} = (k + o(1))\mathrm{e}^{\,n(L - 3\log 3 + o(1))}.$$

É possível determinar que

$$\frac{\mathcal{I}(n)}{\mathcal{D}(n)} = (k + o(1))\mathrm{e}^{\,n(L-3\log 3+o(1))}.$$

O valor de L é um problema em aberto, mas sabemos que

$$2{,}320 \leq L \leq 3\log 3 \approx 3{,}296.$$

É possível determinar que

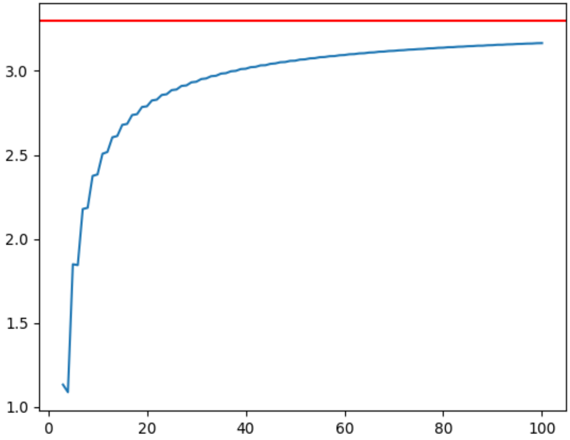
$$\frac{\mathcal{I}(n)}{\mathcal{D}(n)} = (k + o(1))\mathrm{e}^{\,n(L - 3\log 3 + o(1))}.$$

O valor de L é um problema em aberto, mas sabemos que

$$2{,}320 \leq L \leq 3\log 3 \approx 3{,}296.$$

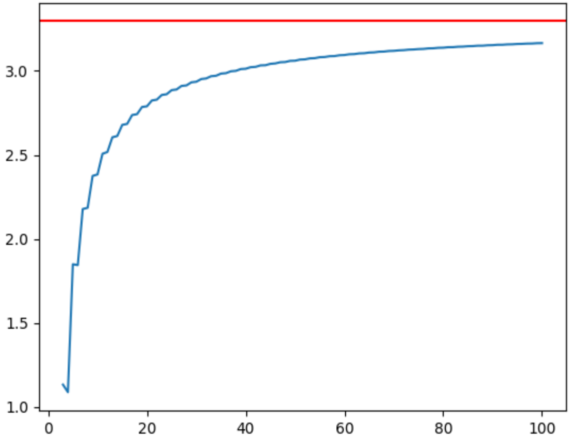
Simulações computacionais sugerem que $L = 3\log 3$.

Gráfico de convergência



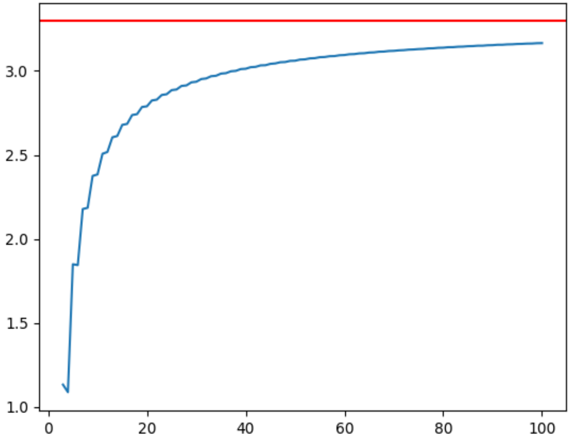
$$\frac{\log \mathcal{I}(n)}{n}$$

Gráfico de convergência



$$\frac{\log \mathcal{I}(n)}{n} \to L$$

Gráfico de convergência



$$\frac{\log \mathcal{I}(n)}{n} \to L$$

Simulações mostram que

$$\frac{\log \mathcal{I}(10000)}{10000} \approx 3{,}294$$

Mas

$$\frac{\mathcal{I}(n)}{\mathcal{D}(n)} = \mathbb{P}_{\text{oct1}} + \mathbb{P}_{\text{oct8}},$$

Mas

$$\frac{\mathcal{I}(n)}{\mathcal{D}(n)} = \mathbb{P}_{\text{oct1}} + \mathbb{P}_{\text{oct8}},$$

onde

$$\mathbb{P}_{\text{oct1}} = \mathbb{P}(\bar{N}_{A>B} > 0, \bar{N}_{B>C} > 0, \bar{N}_{C>A} > 0)$$

e

Mas

$$\frac{\mathcal{I}(n)}{\mathcal{D}(n)} = \mathbb{P}_{\text{oct1}} + \mathbb{P}_{\text{oct8}},$$

onde

$$\mathbb{P}_{\text{oct1}} = \mathbb{P}(\bar{N}_{A>B} > 0, \bar{N}_{B>C} > 0, \bar{N}_{C>A} > 0)$$

e

$$\mathbb{P}_{\text{oct8}} = \mathbb{P}(\bar{N}_{A>B} < 0, \bar{N}_{B>C} < 0, \bar{N}_{C>A} < 0)$$

Mas

$$\frac{\mathcal{I}(n)}{\mathcal{D}(n)} = \mathbb{P}_{\text{oct1}} + \mathbb{P}_{\text{oct8}},$$

onde

$$\mathbb{P}_{\text{oct1}} = \mathbb{P}(\bar{N}_{A>B} > 0, \bar{N}_{B>C} > 0, \bar{N}_{C>A} > 0)$$

e

$$\mathbb{P}_{\text{oct8}} = \mathbb{P}(\bar{N}_{A>B} < 0, \bar{N}_{B>C} < 0, \bar{N}_{C>A} < 0)$$

A convergência do vetor normal implica que a proporção de palavras intransitivas tende a 0 para quando n cresce

Mas

$$\frac{\mathcal{I}(n)}{\mathcal{D}(n)} = \mathbb{P}_{\text{oct1}} + \mathbb{P}_{\text{oct8}},$$

onde

$$\mathbb{P}_{\text{oct1}} = \mathbb{P}(\bar{N}_{A>B} > 0, \bar{N}_{B>C} > 0, \bar{N}_{C>A} > 0)$$

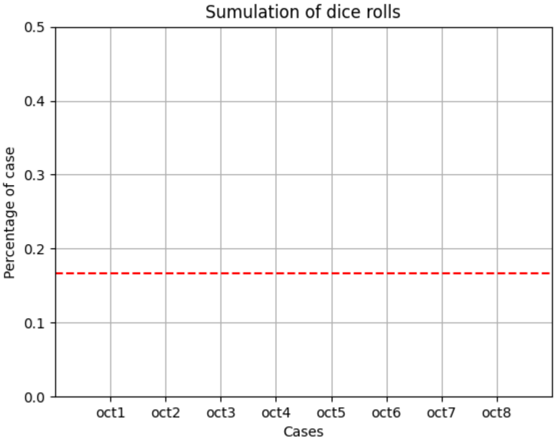
e

$$\mathbb{P}_{\text{oct8}} = \mathbb{P}(\bar{N}_{A>B} < 0, \bar{N}_{B>C} < 0, \bar{N}_{C>A} < 0)$$

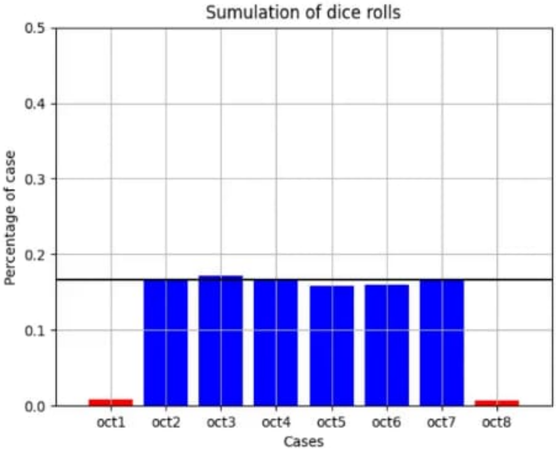
A convergência do vetor normal implica que a proporção de palavras intransitivas tende a 0 para quando n cresce

$$\frac{\mathcal{I}(n)}{\mathcal{D}(n)} \to 0.$$

Distribuição nos octantes



Distribuição nos octantes



Obrigado!