

Análise Estatística de dados

Inteligência Artificial



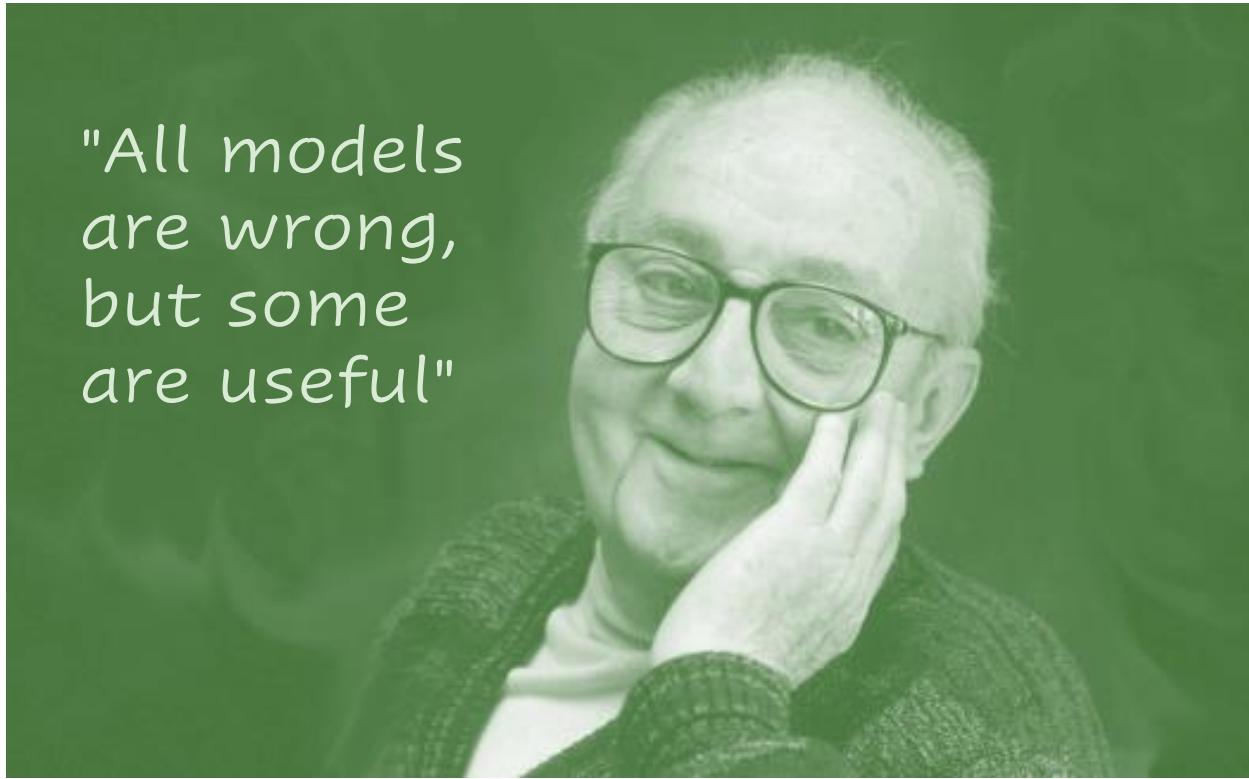
AULA 08 – MODELOS DISCRETOS DE PROBABILIDADE

Larissa Driemeier

PROGRAMA DO CURSO

Aula	Data	Conteúdo da Aula
01	18/02	Aula Inaugural
02	25/02	Introdução ao Curso. Noções de Álgebra Linear , Geometria Analítica Parte I
03	11/03	Noções de Álgebra Linear , Geometria Analítica Parte II
04	18/03	Decomposição de valor singular (SVD)
05	25/03	Otimização: derivadas, derivadas parciais (operadores gradiente, Jacobiano, Hessiano e Laplaciano), algoritmos de gradiente, noções de algoritmos genéticos
06	01/04	Variáveis independentes e não independentes. Estatística Descritiva e Indutiva. Definições de medidas de dispersão e tendência central.
07	08/04	Probabilidade e Teorema de Bayes
08	15/04	Modelos de probabilidade discretos
09	22/04	Modelos de probabilidade contínuos
10	29/04	Modelo de Markov. Modelo de Markov oculto.

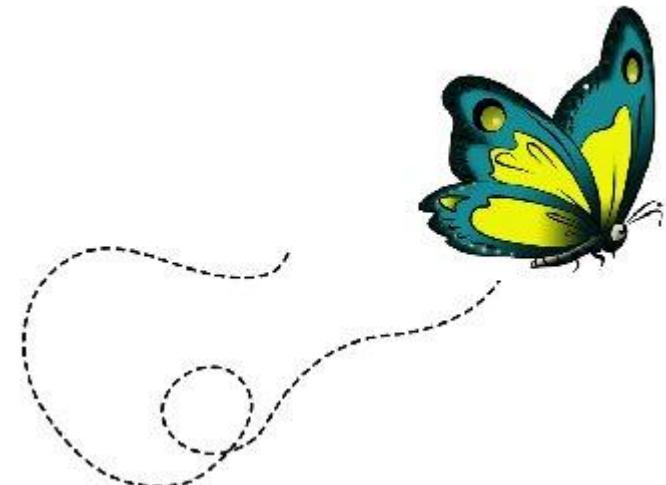
PRECISAMOS DE UM MODELO...



George Edward Pelham Box (1919 – 2013) foi um estatístico britânico, que trabalhou nas áreas de controle de qualidade, desenho de experimentos, inferência bayesiana. Ele foi chamado de *uma das grandes mentes estatísticas do século XX*.

É importante ressaltar que nenhum modelo estatístico está correto, são apenas aproximações da descrição de diferentes propriedades do mundo em certos contextos.

Utilidade e parâmetros fisicamente consistentes são as principais métricas ao projetar/selecionar um modelo.



Como vimos na aula anterior...

HAPPY

MEMORIES

VOCÊ SE LEMBRA???

Considere que numa grande loja de departamentos, em **10%** dos dias algum cliente rouba alguma peça.

Construir a distribuição de probabilidades para a variável aleatória $X =$ **número de dias com roubos na loja**, considerando o período de observação de **dois** dias (suponha independência).



E SE O PERÍODO DE OBSERVAÇÃO FOSSE 20 DIAS, 1000 DIAS????





MODELOS PROBABILÍSTICOS DISCRETOS

Distribuições de Bernoulli,
Binomial, Multinomial,
Geométrica e Poisson.

PROPRIEDADES DAS DISTRIBUIÇÕES DISCRETAS

Variáveis aleatórias discretas são descritas com uma função massa de probabilidade (PMF, do inglês *probability mass function*). Uma PMF mapeia cada valor no espaço amostral da variável para uma probabilidade.

Seja $P(X)$ a função massa de probabilidade de um certo evento X , esta deve apresentar as seguintes propriedades:

- $P(X = x) > 0$, para todos **possíveis valores** de x ;
- $\sum P(x) = 1$, considerando todo o espaço amostral

DISTRIBUIÇÕES DISCRETAS

Entre as variáveis aleatórias discretas, provavelmente as distribuições de probabilidade mais importantes são as distribuições de **Bernoulli, Binomial, Poisson**.

Ainda temos a distribuição **Multinomial, Geométrica, ...**

BERNOULLI E BINOMIAL

Na medicina, uma nova droga sendo testada em pacientes, podendo ser efetiva ou não;

Verificar se um e-mail é ou não é spam;

Uma moeda não honesta – por exemplo, uma com 80% de probabilidade de aparecer cara;

Escolher aleatoriamente as pessoas na rua para responder uma pergunta de sim ou não;

Tentativa de convencer os visitantes de um site a comprar um produto - o resultado é se cada visitante comprou ou não.

Todas as situações acima exibem uma característica de dualidade.

Muitas vezes, tem-se situações onde um experimento tem dois resultados possíveis





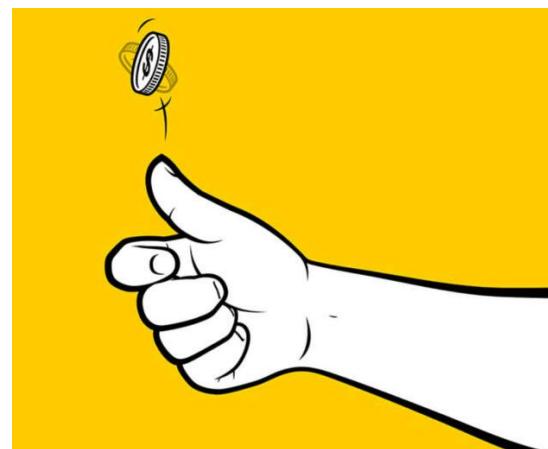
BERNOULLI



Sucesso (1) ou fracasso (0)

MODELO DE BERNOULLI

Em um experimento Bernoulli repetimos o experimento uma única vez.



Vou jogar a moeda 1 vez,
qual a probabilidade de
sucesso (dar cara)?

MODELO DE BERNOULLI – DEFINIÇÃO

A distribuição de Bernoulli é a distribuição de probabilidade discreta de uma variável aleatória que obtém uma saída binária: 1 com probabilidade p e 0 com probabilidade $(1 - p)$.

$$P(X = x) = \begin{cases} p^x(1 - p)^{1-x} & \text{se } x = 0 \text{ ou } 1 \\ 0 & \text{caso contrário} \end{cases} = \begin{cases} p & \text{se } x = 1 \\ 1 - p & \text{se } x = 0 \end{cases}$$

A probabilidade de sucesso p é o parâmetro da distribuição de Bernoulli e, se uma variável aleatória discreta X seguir essa distribuição,

$$X \sim Be(p)$$

VALOR ESPERADO E VARIÂNCIA DO MODELO

Valor esperado e a variância da distribuição são definidas pelas relações:

$$E[x] = p$$
$$\sigma_x^2 = p(1 - p)$$

POR EXEMPLO, IMAGINE...

... que seu experimento consiste em jogar uma moeda e você ganhará se a saída for de *Cara*: Além disso, como a moeda é honesta, você sabe que a probabilidade de ter uma *Cara* é $p = 1/2$. Portanto, depois de definir *Cara* = 1 e *Coroa* = 0, você pode calcular a **probabilidade de sucesso** da seguinte maneira:

$$P(X = 1) = p = \frac{1}{2}$$

... que você está prestes a jogar um dado e apostar seu dinheiro no número 5: portanto, o número 5 será o seu sucesso (definido como 1), enquanto qualquer outro número será um fracasso (definido como 0). A probabilidade de sucesso é $1/6$. Se você quiser calcular a **probabilidade de falha**,

$$P(X = 0) = 1 - p = \frac{5}{6}$$

SIMULANDO EM PYTHON

```
from scipy.stats import bernoulli
```

```
# Variáveis de entrada:  
# Probabilidade de sucesso de cada experimento, p  
p = 0.3  
X = bernoulli(p)  
prob1 = X.pmf(1)  
prob0 = X.pmf(0)  
print('A probabilidade de sucesso em 1 evento é {:.0%} e probabilidade  
de fracasso é {:.0%}'.format(prob1, prob0))
```

```
A probabilidade de sucesso em 1 evento é 30% e probabilidade de fracasso é 70%
```

SIMULAÇÃO DO EVENTO 10000 VEZES

```
# Número de repetições do evento
k = 10000
# Probabilidade de sucesso de cada experimento, p
p = 0.3
X = bernoulli(p)
X_samples = X.rvs(size=k) # Random Variable Sampling
print('Veja que a resposta é sucesso ou fracasso, para cada uma das %d simulações do evento:'%(k))
print(X_samples)
heads = [np.sum(X_samples) for i in range(len(X_samples)) if i == 1]
tails = k - heads[0]
print('Número de caras: ', '{:6.1f}'.format(heads[0]), ' Número de coroas: ',
'{:6.1f}'.format(tails))
# Valor esperado e variância
print("Valores esperados para Distribuição de Bernoulli:")
print('Valor esperado={:5.3f}, Variância={:5.3f}'.format(p, p*(1-p)))
print("Modelo randômico de dados com distribuição de Bernoulli:")
print('Média={:5.3f}, Variância={:5.3f}'.format(X_samples.mean(), (X_samples.std())**2))
# Plotagem dos resultados
sns.set_style("darkgrid")
fig, ax = plt.subplots(figsize=(8,5))
ax = sns.histplot(X_samples, bins=np.linspace(0,2,3), color=colors[0])
ax.set_xlabel("Número de caras", fontsize=16)
ax.set_ylabel("Frequência", fontsize=16)
```

RESULTADOS

Veja que a resposta é sucesso ou fracasso, para cada uma das 10000 simulações do evento:

[0 0 0 ... 1 0 0]

Número de caras: 3039.0

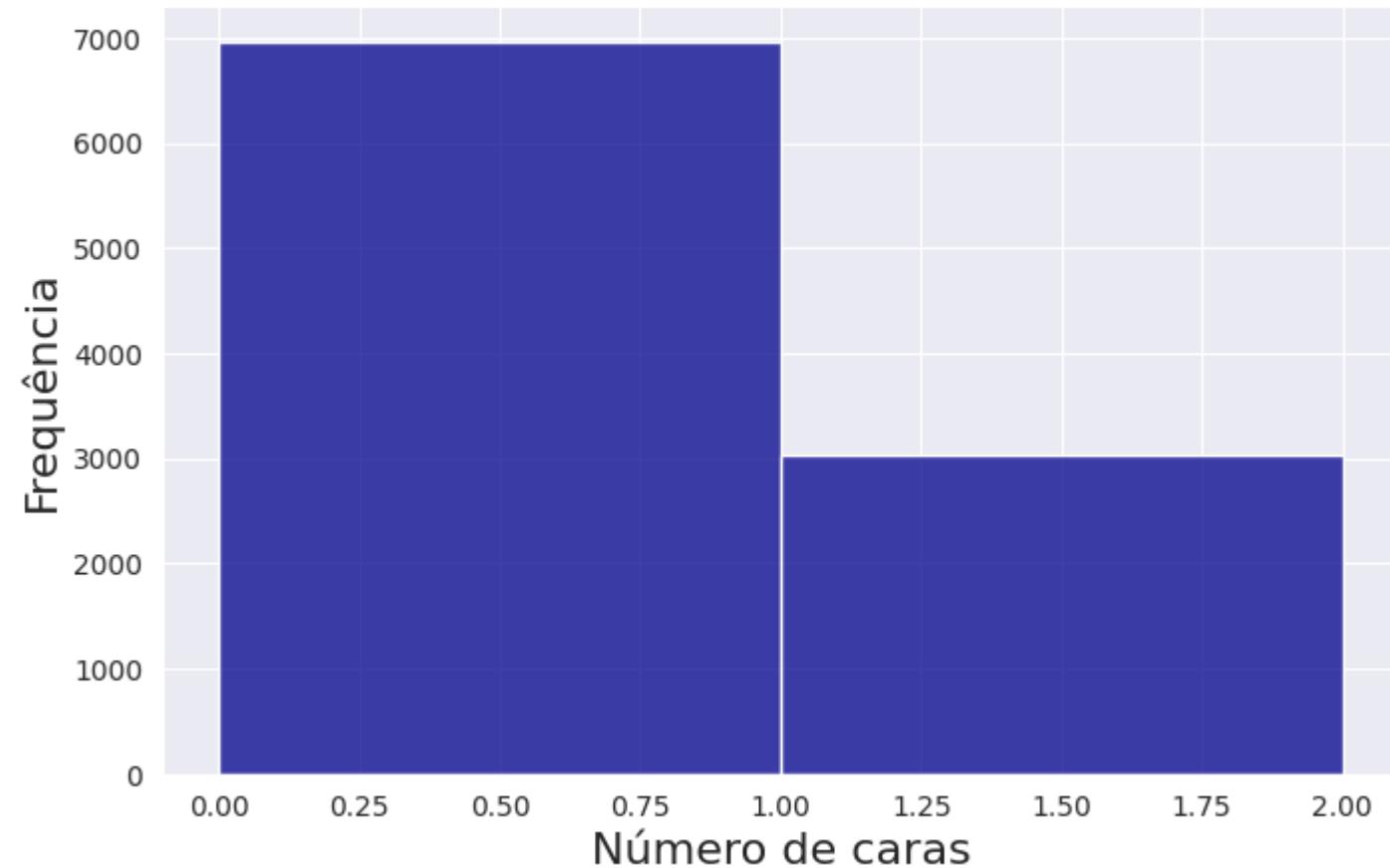
Número de coroas: 6961.0

Valores esperados para Distribuição de Bernoulli:

Valor esperado=0.300, Variância=0.210

Modelo randômico de dados com distribuição de Bernoulli:

Média=0.304, Variância=0.212





MODELO BINOMIAL



Sucesso (1) ou Fracasso (0)
repetindo o experimento n
vezes.

O QUE É DISTRIBUIÇÃO BINOMIAL?

A distribuição binomial é a distribuição de probabilidade de uma sequência de experimentos onde cada experimento produz um resultado binário e onde cada um dos resultados é independente de todos os outros.



Vou jogar a moeda 4 vezes, qual a probabilidade de 3 sucessos (dar cara 3x)?

MODELO BINOMIAL

O modelo binomial pressupõe:

Em cada evento são efetuados n experimentos iguais e independentes (n eventos de Bernoulli).

Cada um dos experimentos tem apenas 2 resultados possíveis e excludentes (**sucesso** e **fracasso**).

Conseqüentemente, a probabilidade de **sucesso** (p) e **fracasso** ($1 - p$) **para cada experimento** são constantes.

A variável aleatória de interesse é o número de sucessos obtidos nos n experimentos.



Na Estatística é prática comum tratar eventos como independentes quando amostras pequenas são retiradas de grandes populações. Orientação: assumir independência quando o tamanho da amostra for, no máximo, 5% do tamanho da população.

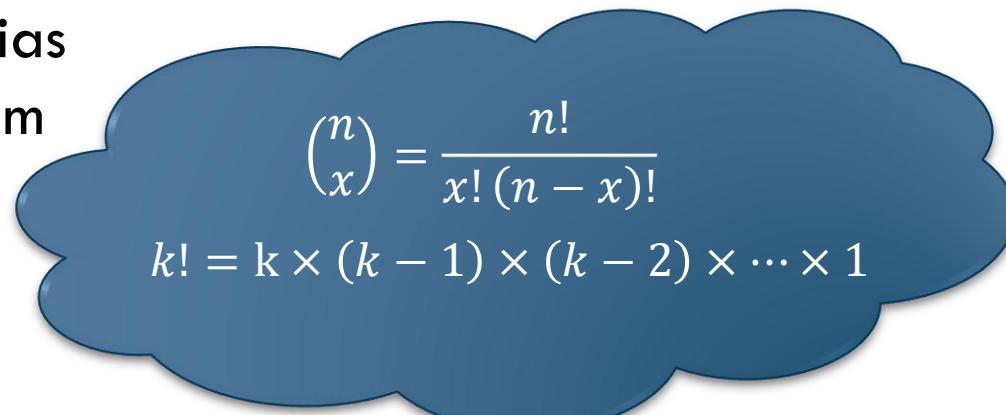
DISTRIBUIÇÃO BINOMIAL

Modelo utilizado para descrever variáveis aleatórias que representam o número x de sucessos obtidos em n provas (tentativas) independentes, quando cada tentativa tem uma probabilidade de sucesso constante p e uma probabilidade de insucesso, também constante $(1 - p)$.

$$P(x|n,p) = \binom{n}{x} p^x (1-p)^{n-x}$$

para $x \geq 0$ e inteiro.

A distribuição é usualmente referenciada pelo símbolo $X \sim B(n, p)$.


$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$
$$k! = k \times (k-1) \times (k-2) \times \cdots \times 1$$



VALOR ESPERADO E VARIÂNCIA DO MODELO

Valor esperado e a variância da distribuição são definidas pelas relações:

$$E[x] = np$$
$$\sigma_x^2 = np(1 - p)$$

COMO DETERMINAR UMA DISTRIBUIÇÃO BINOMIAL

1. Identifique um sucesso;
2. Determine p , a probabilidade de sucesso;
3. Determine n , o número de repetições do experimento;
4. Determine x , a variável aleatória número de sucessos;
5. Substitua valores na fórmula,

$$P(x) = \binom{n}{x} p^x (1 - p)^{n-x}$$

VAMOS AO EXEMPLO DE NOSSA MOEDA

Eu joguei uma moeda 10 vezes, qual a probabilidade de tirar 6 caras?

1. Identifique um sucesso: *Cara*
2. Determine p , a probabilidade de sucesso: $p = 1/2$
3. Determine n , o número de repetições do experimento: $n = 10$
4. Determine x , a variável aleatória número de sucessos: $x = 6$
5. Substitua valores na fórmula,

$$P(x) = \binom{n}{x} p^x (1 - p)^{n-x}$$

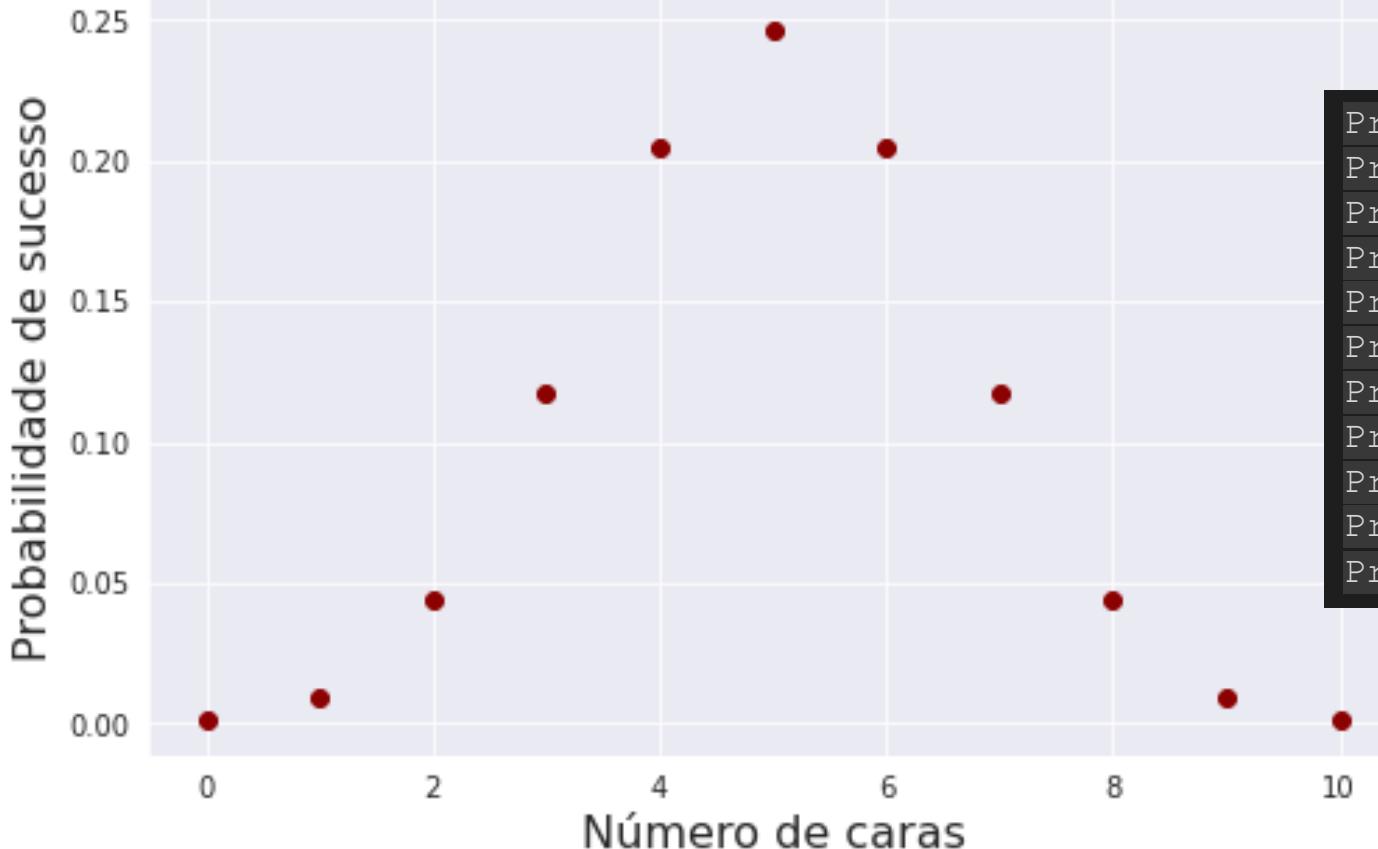
$$P(6) = \binom{10}{6} p^6 (1 - p)^{10-6} = \frac{10 \times 9 \times 8 \times 7 \times 6!}{6!4!} 0,5^6 (1 - 0,5)^4 = 210 \times 0.015625 \times 0.0625 \cong 20,5\%$$

EM PYTHON

```
from scipy.stats import binom
```

```
# Variáveis de entrada:  
# Número de eventos independentes, n  
n = 10  
# Probabilidade de sucesso de cada experimento, p  
p = 0.5  
Y = binom(n,p)  
  
probabilities = []  
for i in range(n+1):  
    # Número de sucessos, x = i  
    print('Probabilidade de {:3d} sucesso(s) no evento: {:.2%}'.format(i,Y.pmf(i)))  
    probabilities.append([i,Y.pmf(i)])  
prob = np.array(probabilities)  
fig, ax = plt.subplots(figsize=(8,5))  
plt.scatter(prob[:,0],prob[:,1], color = colors[1])  
plt.xlabel("Número de caras", fontsize=16)  
plt.ylabel("Probabilidade de sucesso", fontsize=16)  
plt.show()
```

RESPOSTA



```
Probabilidade de 0 sucesso(s) no evento: 0.10%
Probabilidade de 1 sucesso(s) no evento: 0.98%
Probabilidade de 2 sucesso(s) no evento: 4.39%
Probabilidade de 3 sucesso(s) no evento: 11.72%
Probabilidade de 4 sucesso(s) no evento: 20.51%
Probabilidade de 5 sucesso(s) no evento: 24.61%
Probabilidade de 6 sucesso(s) no evento: 20.51%
Probabilidade de 7 sucesso(s) no evento: 11.72%
Probabilidade de 8 sucesso(s) no evento: 4.39%
Probabilidade de 9 sucesso(s) no evento: 0.98%
Probabilidade de 10 sucesso(s) no evento: 0.10%
```

EXEMPLO

Suponha que em uma população, com milhões de pessoas, 3% delas sejam canhotas. Qual a probabilidade p de encontrarmos 4 ou mais canhotos, dentre 120 pessoas escolhidas ao acaso desta população?

$$P(4) = \binom{120}{4} 0,03^4 (0,97)^{120-4} = \frac{120 \times 119 \times 118 \times 117 \times 116!}{4! 116!} 0,03^4 (0,97)^{116} \approx 19,43\%$$

$$P(x \geq 4) = P(4) + P(5) + P(6) + P(7) + \dots$$

$$P(x \geq 4) = 1 - [P(0) + P(1) + P(2) + P(3)]$$

```

# Variáveis de entrada:
#Identifique o sucesso: Canhoto
# Probabilidade de sucesso de cada experimento, p
p = 0.03
# Número de eventos independentes (repetições do experimento), n
n = 120
Y = binom(n,p)

```

```

x_min = 4
prob = []
for i in range(x_min):
    # Número de sucessos, x = i
    print('Probabilidade de {:3d} sucesso(s) no evento: {:.3%}'.format(i,Y.pmf(i)))
    prob.append(Y.pmf(i))
prob_tot = sum(prob)
print('A probabilidade de encontrarmos até {:3d} canhotos : {:.2%}'.format(x_min-1,prob_tot))

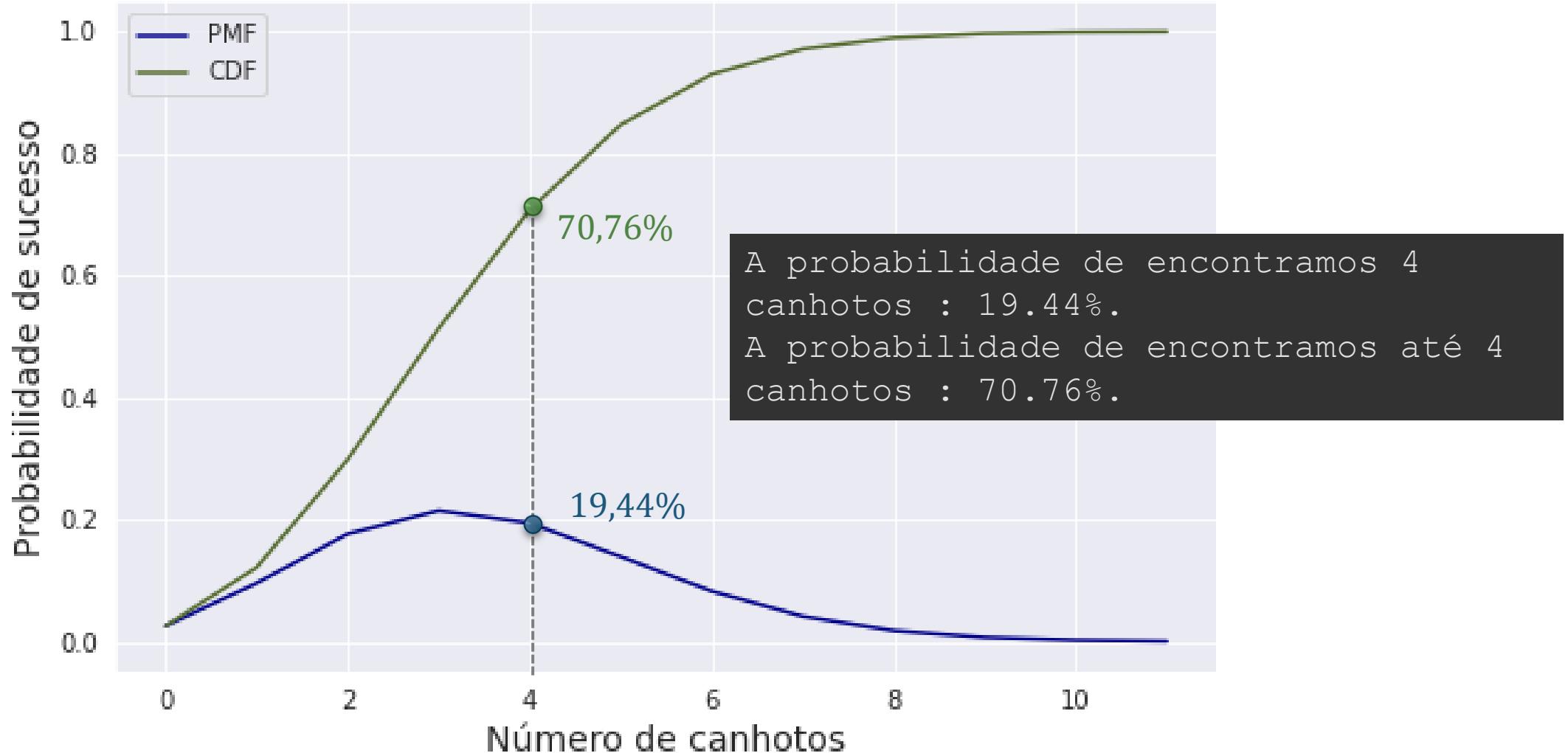
```

```

Probabilidade de 0 sucesso(s) no evento: 2.586%
Probabilidade de 1 sucesso(s) no evento: 9.597%
Probabilidade de 2 sucesso(s) no evento: 17.661%
Probabilidade de 3 sucesso(s) no evento: 21.484%
A probabilidade de encontrarmos até 3 canhotos : 51.33%.

```

```
prob_tot = Y.cdf(x_min-1)
print('A probabilidade de encontramos {:3d} canhotos : {:.2%}'.format(x_min,Y.pmf(x_min)))
print('A probabilidade de encontramos até {:3d} canhotos : {:.2%}'.format(x_min,prob_tot+Y.pmf(x_min)))
```



EXEMPLO

Retira-se uma amostra de 12 e-mails. Sabendo-se que 2% dos e-mails são spams, qual é a probabilidade de haver um único spam dentre os 12? Adicionalmente, qual a probabilidade de não haver mais de um spam em toda amostra?

Para o problema em estudo tem-se $n = 12$ e $p = 0,02$, considerando como sucesso o resultado **spam**.

A distribuição de número de spams é dada pela relação:

$$P(x) = \binom{12}{x} 0,02^x (0,98)^{12-x}$$

A chance de se obter um spam é:

$$P(x = 1) = P(1) = \binom{12}{1} 0,02^1 (0,98)^{11} = 0,192$$

A chance de se obter não mais de um spam é:

$$P(x \leq 1) = P(0) + P(1)$$

$$P(0) = \binom{12}{0} 0,02^0 (0,98)^{12} = 0,785$$

$$P(0) + P(1) = 0,192 + 0,785 = 0,977$$

PARA VOCÊ TREINAR

Uma empresa envia código de desconto para 9 pessoas, cada uma com uma probabilidade estimada de sucesso de compra de 0,1. Nenhuma das 9 pessoas compra.

1. Qual é a probabilidade disso acontecer?
2. Qual a probabilidade de, pelo menos, 2 pessoas comprarem?

DISTRIBUIÇÃO BINOMIAL REPETINDO k VEZES O EVENTO

Rode mais de uma vez.

```
n=3  
p=0.5  
Z = binom(n,p)
```

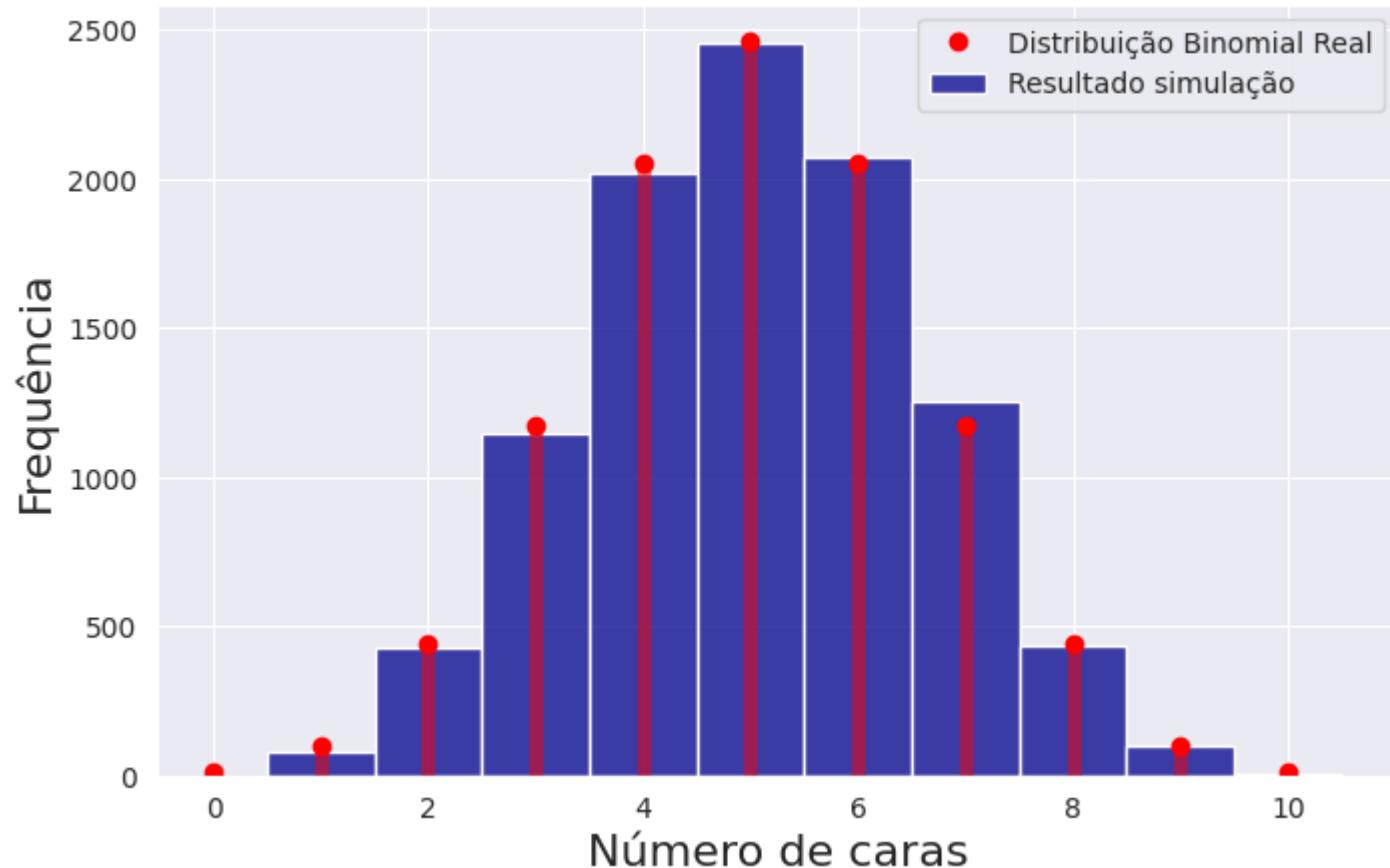
```
#size = k = número de experimentos (eventos)  
k=5;  
# veja que você pode usar a opção: success = binomial(n, p, size=k), que  
importamos no inicio  
success = Z.rvs(size=k)  
print('Total de sucessos em cada evento:', success)  
# Momentos: média e variância  
mean, var = Z.stats(moments='mv')  
print('Média={:6.3f}, Variância={:6.3f}'.format(mean, var))
```

```
Total de sucessos em cada evento: [1 2 0 1 2]  
Média= 1.500, Variância= 0.750
```

TESTE VOCÊ...

Rode 1000 tentativas,
cada uma composta de
10 eventos
independentes, e
probabilidade de sucesso
de cada evento de 0,5.

Rode mais de uma vez.



SIGA O EXEMPLO DO VENDEDOR EM SEU NOTEBOOK

Um vendedor que atende via *call center* recebe em média 50 ligações por dia.

A probabilidade de uma conversão (venda) para cada chamada é de 4%.

A receita média para a empresa em cada conversão é de R\$ 50.

O *call center* que você está analisando possui 150 vendedores.

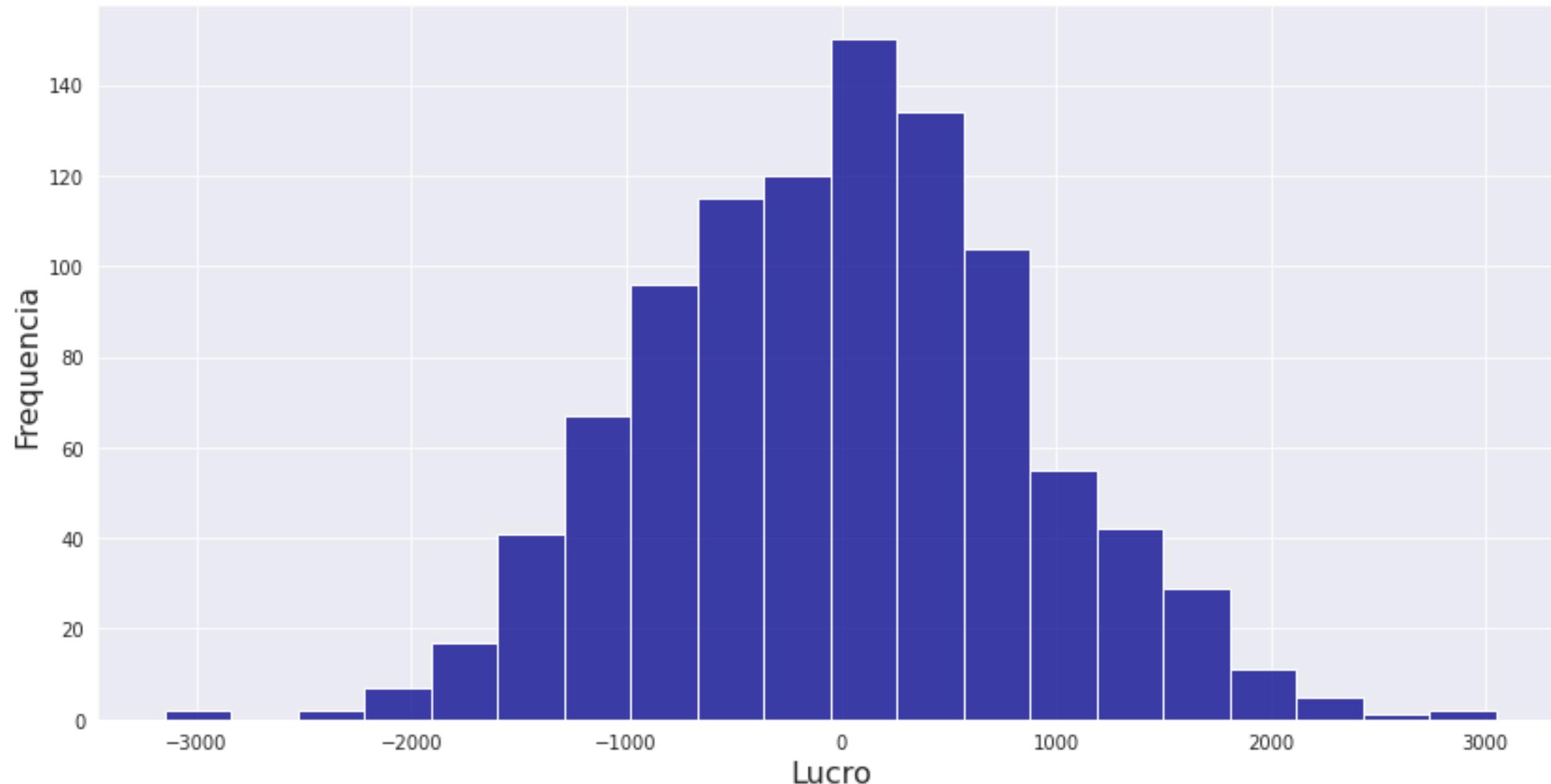
Cada vendedor custa R\$ 100 por dia de trabalho.

Analise o lucro da empresa. Analise a resposta ao:

- aumentar a taxa de conversão de 4% para 5%;
- aumentar o número de funcionários de 150 para 200.

Qual seria a atitude mais eficiente para aumentar os lucros da empresa?







DISTRIBUIÇÃO MULTINOMIAL



Uma distribuição multinomial é a generalização de uma distribuição binomial.

DISTRIBUIÇÃO MULTINOMIAL COM k VARIÁVEIS DE ESTADO

A distribuição multinomial é uma generalização da distribuição binomial para k categorias em vez de apenas binário (sucesso/fracasso). Considere n tentativas independentes, cada uma das quais leva ao sucesso de uma das k categorias. Cada categoria tendo uma determinada probabilidade $p_i, i = 1, 2 \dots k$ fixa de sucesso. A distribuição multinomial fornece a probabilidade de qualquer combinação específica de números de sucessos para as várias categorias, através da fórmula



$$P(x_1 x_2 \cdots x_k | n, p_1 p_2 \cdots p_k) = \binom{n}{x_1 x_2 \cdots x_k} \prod_{i=1}^k p_i^{x_i}$$

$$\binom{n}{m_1 m_2 \cdots m_i} = \frac{n!}{m_1! m_2! \cdots m_i!} ; \sum_i p_i = 1$$

Lejeune Dirichlet (1805-1859)

VALOR ESPERADO E VARIÂNCIA DO MODELO

Valor esperado e a variância da distribuição são definidas pelas relações:

$$E_i[x] = np_i$$
$$\sigma_i^2 = np_i(1 - p_i)$$

GENERALIZAÇÃO DO BINOMIAL

```
from scipy.stats import binom
from scipy.stats import multinomial as multinom
```

```
print('Solução por multinomial:', multinom.pmf([3, 4], n=7, p=[0.4, 0.6]))
print('Solução por binomial:', binom.pmf(3, 7, 0.4))
```

```
Solução por multinomial: 0.29030399999999973
Solução por binomial: 0.290304
```

EXEMPLO

Suponha que, em uma eleição, de um grande país, com três candidatos, o candidato A receba 20% dos votos, o candidato B receba 30% dos votos e o candidato C receba 50% dos votos. Se seis eleitores forem selecionados aleatoriamente, qual é a probabilidade de haver exatamente um partidário do candidato A , dois partidários do candidato B e três partidários do candidato C na amostra?

$$P(x_1 x_2 \cdots x_k | n, p_1 p_2 \cdots p_k) = \binom{n}{x_1 x_2 \cdots x_k} \prod_{i=1}^k p_i^{x_i}$$

$$P(A = 1, B = 2, C = 3 | 6, p_A = 0,20; p_B = 0,30; p_C = 0,50) = \frac{6!}{1!2!3!} 0,2^1 0,3^2 0,5^3$$

$$P(A = 1, B = 2, C = 3 | 6, p_A = 0,20; p_B = 0,30; p_C = 0,50) = 13,5\%$$

```
print('Solução por multinomial:', '{:.2%}'.format(multinom.pmf([1,2,3], n=6, p=[0.2,0.3, 0.5])))
```

Solução por multinomial: 13.50%

E SE...

Supõe-se, agora, que 1000 pessoas são entrevistadas. Quantas, provavelmente, votarão em cada candidato?

```
p = [0.2, 0.3, 0.5]
k = 1000
# Rodar a simulaçao
cases = multinom.rvs(k, p)
# summarize cases
for i in range(len(cases)):
    print('Candidato %d: %d' % (i+1, cases[i]))
```

```
Candidato 1: 179
Candidato 2: 308
Candidato 3: 513
```

TORNEIO DE XADREZ

Em um torneio de xadrez, queremos determinar qual é a probabilidade de, após 12 jogos, o jogador 1 ter 7 vitórias, o jogador 2 ter 2 vitórias e os jogos restantes terminarem empatados. Para isso, suponha que a probabilidade do Jogador 1 vencer seja 40%, o Jogador 2 seja 35% e o empate tenha probabilidade 25%. Portanto temos,

```
multinom.pmf( [7, 2, 3] , n=12 , p=[ 0.4 , 0.35 , 0.25 ] )
```

```
Probabilidade final: 0.02483711999999996
```

VAMOS TESTAR VÁRIAS PARTIDAS...

```
n = 12                      # número de partidas
pvals = [0.4, 0.35, 0.25]    # probabilidades em uma única partida

# número de partidas jogadas
sizes = []
# uma lista para conter proporções (devem ser convergentes para pvals)
# em que o jogador 1 vence 7 vezes, o jogador 2 vence 2 vezes e ocorrem 3 empates:
p = []
```

Vamos testar várias simulações de 12 partidas

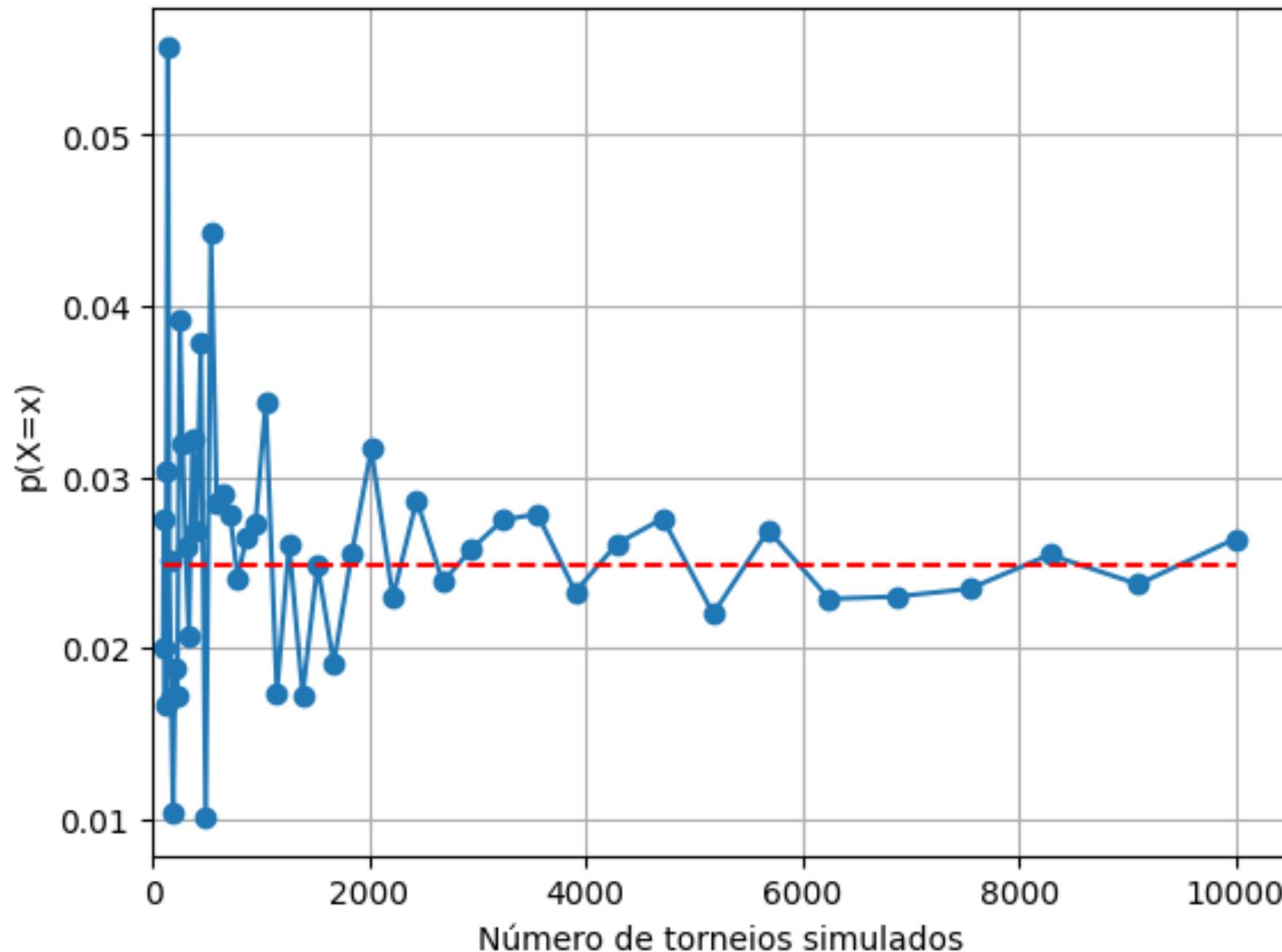
Vamos verificar quantas vezes conseguimos que o jogador 1 vença 7 partidas, o jogador 2 vença 2 e ocorram 3 empates

```
sizes = [100, 109, 120, 132, 145, 159, 175, 193, 212, 232, 255,
281, 308, 339, 372, 409, 449, 494, 542, 596, 655, 719, 790, 868,
954, 1048, 1151, 1264, 1389, 1526, 1676, 1842, 2023, 2222, 2442,
2682, 2947, 3237, 3556, 3906, 4291, 4714, 5179, 5689, 6250,
6866, 7543, 8286, 9102, 10000]
```

```
for size in np.logspace(2,4):
    # variáveis aleatórias discretas são geradas de acordo com a distribuição multinomial:
    outcomes = multinom.rvs(n, pvals, size=int(size))
    # vamos contar a proporção do resultado esperado sobre todos os resultados
    # deverá convergir para a probabilidade
    prob = sum((outcomes[:,0]==7) & (outcomes[:,1]==2) & (outcomes[:,2]==3))/len(outcomes)
    p.append(prob)
    sizes.append(int(size))

# Gráfico
fig1 = plt.figure()
plt.plot(sizes,p,'o-')
plt.plot(sizes,[0.0248]*len(sizes),'--r')
plt.grid()
plt.xlim(xmin=0)
plt.xlabel('Número de partidas')
plt.ylabel('p(X=K)')
plt.title('Teórico p(X=K) = 0.0248');
```

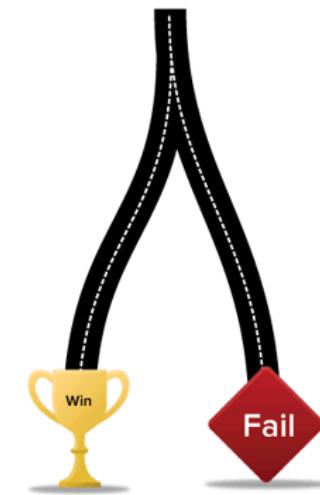
Teórico $p(X=x) = 0.0248$



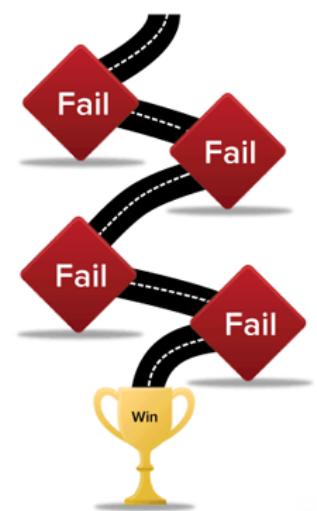


DISTRIBUIÇÃO GEOMÉTRICA

What Most
People Think



What Successful
People Know



Quantas tentativas eu preciso
para chegar ao sucesso?

DISTRIBUIÇÃO GEOMÉTRICA

A distribuição geométrica fornece a probabilidade de que a primeira ocorrência de sucesso em uma série de testes de Bernoulli exija k tentativas independentes, cada uma com probabilidade de sucesso p . Então a probabilidade de que a k -ésima tentativa seja o primeiro sucesso é dada por,

$$P(X = k) = p(1 - p)^{k-1}$$

para $k = 1, 2, 3 \dots$

As três suposições necessárias para aplicação do modelo são:

- Existem dois resultados possíveis para cada tentativa (sucesso ou fracasso).
- As tentativas são independentes.
- A probabilidade de sucesso é a mesma em todas as tentativas.

VALOR ESPERADO E VARIÂNCIA DO MODELO

Valor esperado e a variância da distribuição são definidas pelas relações:

$$E[x] = \frac{1}{p}$$

$$\sigma_x^2 = \frac{1-p}{p^2}$$

EXEMPLO

Um médico está procurando um antidepressivo para um paciente recém-diagnosticado. Suponha que, dos antidepressivos disponíveis, a probabilidade de que um medicamento em particular seja eficaz para um paciente em particular seja $p = 0,4$.

- Qual é o número esperado de medicamentos que serão testados para encontrar um que seja eficaz?
- Qual é a probabilidade de que o primeiro medicamento considerado eficaz para esse paciente seja o primeiro medicamento testado, o segundo medicamento testado e assim por diante?

$$E[x] = \frac{1}{p} = \frac{1}{0,4} = 2,5$$

$$P(X = k) = p(1 - p)^{k-1} = 0,4(0,6)^{k-1}$$

k	$P(X = k) = 0,4(0,6)^{k-1}$
1	$P(1) = 0,4 \times 1 = 0,4$
2	$P(2) = 0,4 \times 0,6 = 0,24$
3	$P(3) = 0,4 \times 0,6^2 = 0,144$

```
from scipy.stats import geom
```

```
# Probabilidade de sucesso, p; e número de tentativas até o sucesso, k
p=0.4
k=3
print('A chance de acertar na terceira tentativa é:', '{:.2%}'.format(geom.pmf(k,p)))
mean, var = geom.stats(p, moments='mv')
print('Média=%.3f, Variância=%.3f' % (mean, var))
```

```
A chance de acertar na terceira tentativa é: 14.40%
Média=2.500, Variância=3.750
```

PARA VOCÊ FAZER...

Um paciente está aguardando um doador de rim compatível para transplante. Se a probabilidade de um doador selecionado aleatoriamente ser compatível é $p = 0,1$, qual é o **número esperado de doadores** que serão testados antes que um compatível seja encontrado?

BASQUETE

Nos esportes, é comum os jogadores fazerem várias tentativas de marcar pontos. Cada tentativa única pode ter dois resultados possíveis: pontuação ou não pontuação. Essas situações podem ser modeladas com distribuições geométricas. Responda,

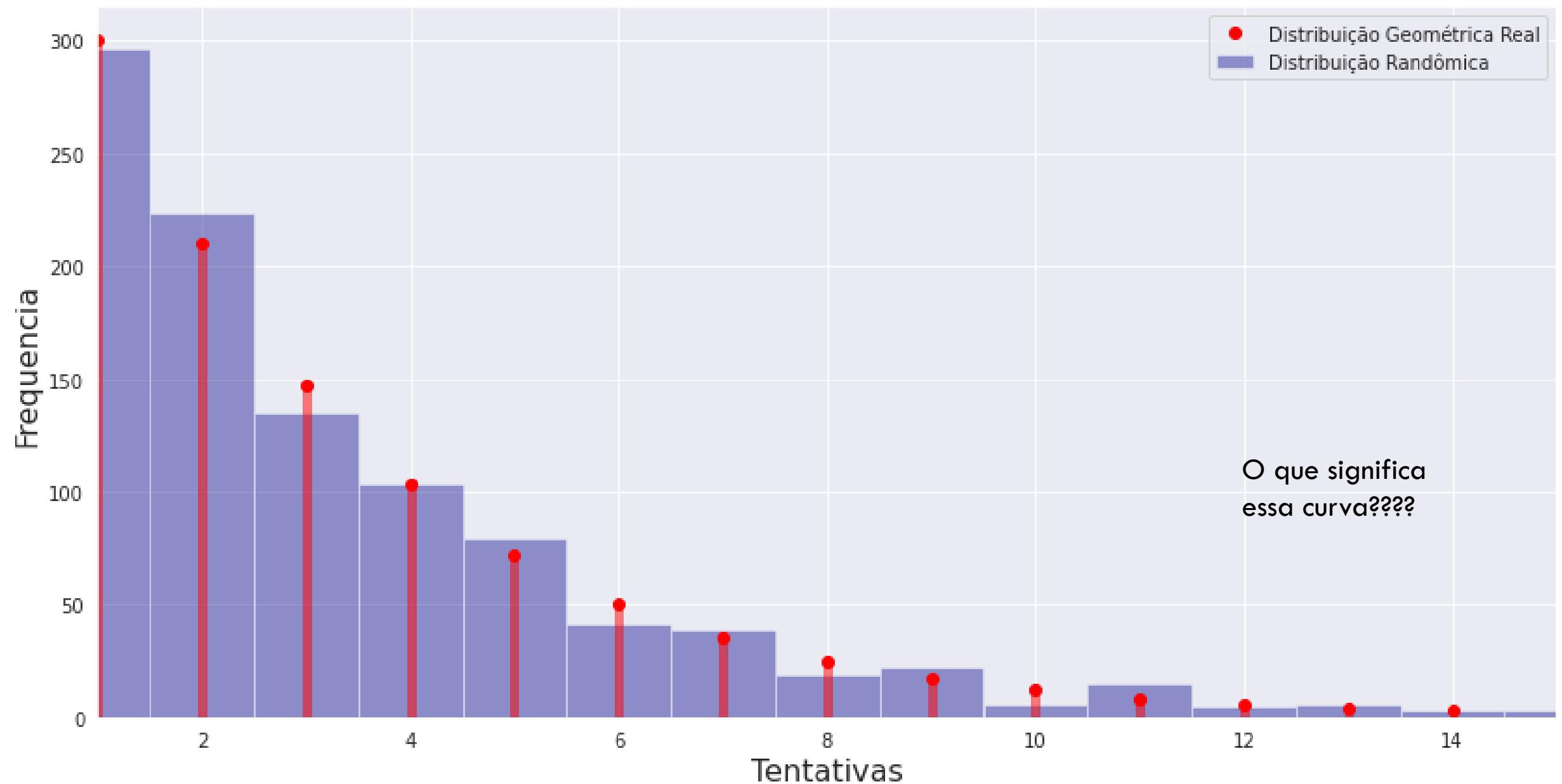
- Qual a probabilidade de um jogador de basquete acertar uma cesta de 3 pontos na terceira tentativa, dado que a probabilidade de sucesso é de 30%?
- Mostre a curva de tentativas k até o sucesso para 1000 repetições do evento.

```
# Probabilidade de sucesso, p; e número de tentativas até o sucesso, k
p=0.3
k=3
print('A chance de acertar na terceira tentativa é: ', '{:.2%}'.format(geom.pmf(k,p)))
mean, var = geom.stats(p, moments='mv')
print('Média=%.3f, Variância=%.3f' % (mean, var))
```

```
A chance de acertar na terceira tentativa é: 14.70%
Média=3.333, Variância=7.778
```

```
# Criação da amostra
p=0.3
trials = 1000
sample = geom.rvs(p, size=trials)
# Plotagem dos resultados
sns.set_style("darkgrid")
fig, ax = plt.subplots(figsize=(14,7))
ax = sns.distplot(sample, bins = np.linspace(0,15,16)+0.5, kde=False, label='Distribuição Randômica', color=colors[0])
k = range(1,15)
ax.plot(k, geom.pmf(k, p)*trials, 'ro', label='Distribuição Geométrica Real')
print(geom.pmf(k, p))
ax.vlines(k, 0, geom.pmf(k, p)*trials, colors='r', lw=5, alpha=0.5)
plt.legend()
plt.xlim(1,15)
ax.set_xlabel("Tentativas até o sucesso", fontsize=16)
ax.set_ylabel("Frequencia", fontsize=16)
plt.savefig(fname='Basquete_Hist', dpi=150)
```

```
[0.3 0.21 0.147 0.1029 0.07203 0.050421 0.0352947 0.02470629 0.0172944 0.01210608 0.00847426 0.00593198
0.00415239 0.00290667]
```





DISTRIBUIÇÃO DE POISSON

NEM SEMPRE FUNCIONA...

Em muitos casos conhecemos o número de sucessos, porém o número de fracassos seria difícil de ser determinado, ou muitas vezes não faria nenhum sentido prático.

Por exemplo, número de descargas elétricas em um dia normal de chuvas.

Podemos, num determinado espaço de tempo determinar o número de descargas elétricas que ocorreram, porém, o número de descargas elétricas que deixaram de ocorrer não poderá ser determinado.

DISTRIBUIÇÃO DE POISSON

Utilizada para determinar o número de sucessos em um intervalo **contínuo** de avaliação, o qual pode ser tempo, distância, área superficial, entre outros.

A variável aleatória (número de ocorrências) é **discreta!**

A distribuição tem como parâmetro a frequência média de sucessos do fenômeno estudado (λ),

$$P(x) = \frac{e^{-\lambda t} (\lambda t)^x}{x!}$$

onde t representa a magnitude do intervalo contínuo e $x \geq 0$ e inteiro.

A distribuição é usualmente referenciada pelo símbolo $X \sim Po(\lambda)$.

VALOR ESPERADO E VARIÂNCIA DO MODELO

Valor esperado e a variância da distribuição são definidas pelas relações:

$$E[x] = \lambda$$
$$\sigma_x^2 = \lambda$$

Única distribuição em que a média é igual à variância...

EXEMPLO

Um digitador faz em média 3 erros por página. Qual a probabilidade que em um teste de digitação de uma página, este não cometa mais de um erro.

Fonte: Costa Neto, P. L. ; Cymbalista, M.
“Probabilidades”, 2^a edição, Editora Edgard
Blucher, 2006.

Aplicando-se a distribuição de Poisson, tem-se:

$$P(x) = \frac{e^{-3 \times 1} (3 \times 1)^x}{x!}$$

onde x é o número de erros em uma página digitada.

O termo não mais de um erro significa que podem ser cometidos 0 erro ou 1 erro.
Esta probabilidade é expressa pela relação:

$$P(x \leq 1) = P(0) + P(1)$$

$$P(x \leq 1) = \frac{e^{-3 \times 1} (3 \times 1)^0}{0!} + \frac{e^{-3 \times 1} (3 \times 1)^1}{1!}$$

$$P(x \leq 1) = 0,0498 + 0,1494 = 0,199$$

OUTRO EXEMPLO

Um telefone recebe em média 0,25 chamadas por hora. Qual a probabilidade de, em 4 horas, receber:

- a) No máximo 2 chamadas?
- b) Exatamente três chamadas?
- c) No mínimo 3 chamadas?

λ : 0,25 chamadas/hora (taxa de ocorrência do evento)

$t = 4$ horas

$$P(x) = \frac{e^{-\lambda t}(\lambda t)^x}{x!} = \frac{e^{-0,25 \times 4}(0,25 \times 4)^x}{x!} = \frac{e^{-1}}{x!}$$

a. $P(x \leq 2) = P(x = 0) + P(x = 1) + P(x = 2)$

$$P(x = 0) = \frac{e^{-1}}{0!} = 0,3679$$

$$P(x = 1) = \frac{e^{-1}}{1!} = 0,3679$$

$$P(x = 2) = \frac{e^{-1}}{2!} = 0,1839$$

$$P(x \leq 2) = 2 \times 0,3679 + 0,1839$$

$$P(x \leq 2) = 0,9197 = 91,97\%$$

λ : 0,25 chamadas/hora (taxa de ocorrência do evento)

$t = 4$ horas

$$P(x) = \frac{e^{-\lambda t}(\lambda t)^x}{x!} = \frac{e^{-0,25 \times 4}(0,25 \times 4)^x}{x!} = \frac{e^{-1}}{x!}$$

b. $P(x = 3) = \frac{e^{-1}}{3!} = 0,0613$

$$P(x = 3) = 6,13\%$$

λ : 0,25 chamadas/hora (taxa de ocorrência do evento)

$t = 4$ horas

$$P(x) = \frac{e^{-\lambda t}(\lambda t)^x}{x!} = \frac{e^{-0,25 \times 4}(0,25 \times 4)^x}{x!} = \frac{e^{-1}}{x!}$$

c. $P(x \geq 3) = 1 - P(x \leq 2)$

$$P(x \geq 3) = 1 - 0,9197 = 0,0803$$

$$P(x \geq 3) = 8,03\%$$

VOLTEMOS AO NOTEBOOK

```
from scipy.stats import poisson
```

```
lamb_value=0.25
tot=4
mu=lamb_value*tot
iterable = (poisson.pmf(i, mu) for i in range(0,3))
values = np.fromiter(iterable, float)
print(values)
aux = np.sum(values)
#Pode-se usar a acumulada cdf
aux2 = poisson.cdf(2, mu)
print('A probabilidade de receber, no máximo, 2 chamadas, é:
', '{:.2%}'.format(aux), 'ou', '{:.2%}'.format(aux2))
```

```
[0.36787944 0.36787944 0.18393972]
```

```
A probabilidade de receber, no máximo, 2 chamadas, é: 91.97% ou 91.97%
```

```
x = 3
```

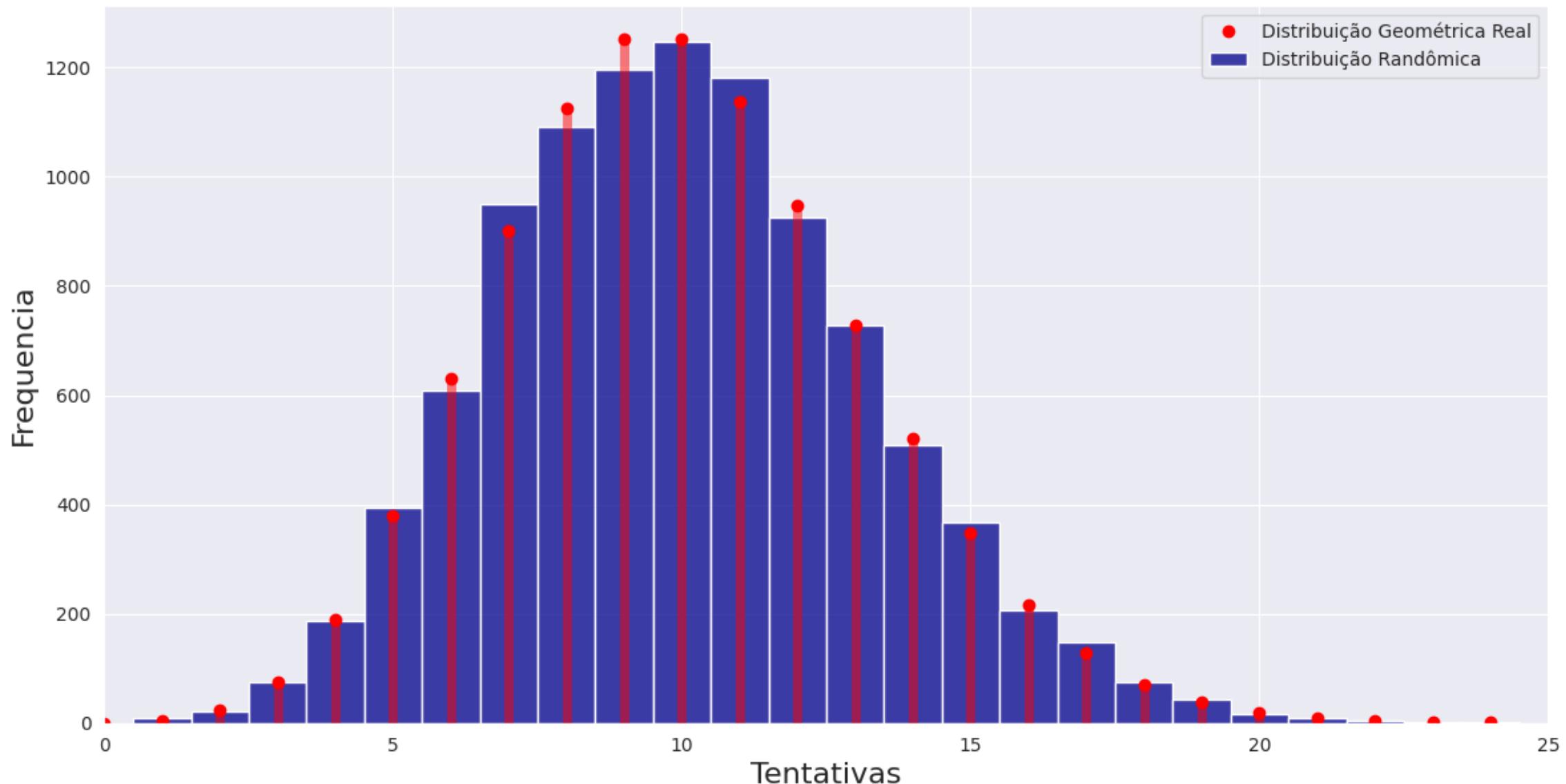
```
print('A probabilidade de receber 3 chamadas, é: ', '{:.2%}'.format(poisson.pmf(x,mu)))
print('A probabilidade de receber no mínimo 3 chamadas, é: ', '{:.2%}'.format(1-aux))
```

```
A probabilidade de receber 3 chamadas, é: 6.13%
```

```
A probabilidade de receber no mínimo 3 chamadas, é: 8.03%
```

ANÁLISE RANDÔMICA

```
# Distribuição randômica
lamb_value=2.5 #troque os valores de lambda e verifique a forma da curva
tot=4
mu=lamb_value*tot
trials = 10000
sample_poisson = poisson.rvs(mu, size=trials)
# Plotagem dos resultados
sns.set_style("darkgrid")
fig, ax = plt.subplots(figsize=(14, 7))
ax = sns.distplot(sample_poisson, bins = np.linspace(0,26,27)+0.5, kde=False,
label='Distribuição Randômica', color=colors[0])
k = range(0,25)
ax.plot(k, poisson.pmf(k, mu)*trials, 'ro', label='Distribuição Geométrica Real')
print(poisson.pmf(k, mu))
ax.vlines(k, 0, poisson.pmf(k, mu)*trials, colors='r', lw=5, alpha=0.5)
plt.legend()
plt.xlim(0,25)
ax.set_xlabel("Tentativas", fontsize=16)
ax.set_ylabel("Frequencia", fontsize=16)
plt.savefig(fname='Basquete_Hist', dpi=150)
```





SPOILER

AS LOCADORAS DE CARROS
DO JACK

JACK'S CAR RENTAL PROBLEM

Definição do problema: Jack possui duas locadoras de veículos como parte de uma agência nacional de locação. As duas locadoras têm níveis variados de demanda (aluguel de carros por dia) e de devolução (quantos carros são devolvidos por dia). Como uma delas tem mais demanda do que devolução, Jack move os carros durante a noite de um local para outro. Ele quer ter certeza de que tem carros suficientes em cada local para maximizar seu retorno.

As questões que ajudaremos Jack a responder são:

Mas, isto é um spoiler do curso de Aprendizado por Reforço!



POISSON

O número de carros em cada local de aluguel de carros deve nos fornecer um **retorno esperado**, quanto dinheiro esperamos ganhar. Esse retorno é **influenciado por uma distribuição aleatória de Poisson que usaremos para ponderar todos os possíveis retornos/alugueis que podem ocorrer.**

Taxa de ocorrência λ

	Locadora 1	Locadora 2
Solicitação (saída)	3	4
Devolução (entrada)	3	2

A Locadora 2 tem mais alugueis do que devoluções, enquanto a Locadora 1 tem número igual de alugueis e devoluções.

EXEMPLO – LOCADORA 2

Suponhamos que o aluguel e a devolução de carros na locadora seguem um processo de Poisson, com médias iguais a 4 e 2, respectivamente.

- a) Qual é a probabilidade de exatamente 5 carros serem alugados em 1 dia?
- b) Qual é a probabilidade de pelo menos 2 carros serem devolvidos em 1 dia?

Seja X o número de carros alugados em 1 dia. Como a taxa média de alugueis por dia é de $\lambda = 4$, $X \sim Po(4)$.

$$P(x) = \frac{e^{-4t}(4t)^x}{x!}$$

Seja Y o número de carros devolvidos em 1 dia. Como a taxa média de devoluções por dia é de $\lambda = 2$, $Y \sim Po(2)$.

$$P(y) = \frac{e^{-2t}(2t)^y}{y!}$$

$$a) P(x = 5) = \frac{e^{-4 \times 1} (2 \times 1)^5}{5!}$$

$$P(x = 5) = 15,63\%$$

```
lamb_value=4
tot=1
mu=lamb_value*tot
poisson.pmf(5, mu)
```

Poisson não
tem limite
superior.

$$b) P(y \geq 2) = 1 - P(y = 0) - P(y = 1)$$

$$= 1 - \frac{e^{-2 \times 1} (2 \times 1)^0}{0!} - \frac{e^{-2 \times 1} (2 \times 1)^1}{1!}$$

$$P(y \geq 2) = 1 - 0,1353 - 0,2707 = 0,5940 = 59,40\%$$



THINK BAYESIAN



VEROSSIMILHANÇA BINOMIAL

A distribuição Binomial é a distribuição de probabilidade que descreve a probabilidade de obter x sucessos em n tentativas, se a probabilidade de sucesso em cada tentativa for p . Esta distribuição é apropriada para dados de prevalência onde você sabe que teve x resultados positivos em n amostras. Seu modelo estima a probabilidade de sucesso, P , que depende dos parâmetros do modelo.

$$P(x|n, p) = \binom{n}{x} p^x (1 - p)^{n-x}$$

VOCÊ GOSTA DE COCA ZERO OU COCA NORMAL?

Que atire a
primeira latinha
quem nunca bebeu
refrigerante na
vida!

$$P(x|n, p) = \binom{n}{x} p^x (1 - p)^{n-x}$$



SERÁ QUE AS PESSOAS PREFEREM ZERO A NORMAL?

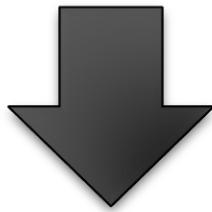
$$P(x|n, p) = \binom{n}{x} p^x (1 - p)^{n-x}$$

$$P(x = 4|n = 7, p = 0.5) = \left(\frac{7!}{4! (7-4)!} \right) 0.5^4 (1 - 0.5)^{7-4} = 0.273$$



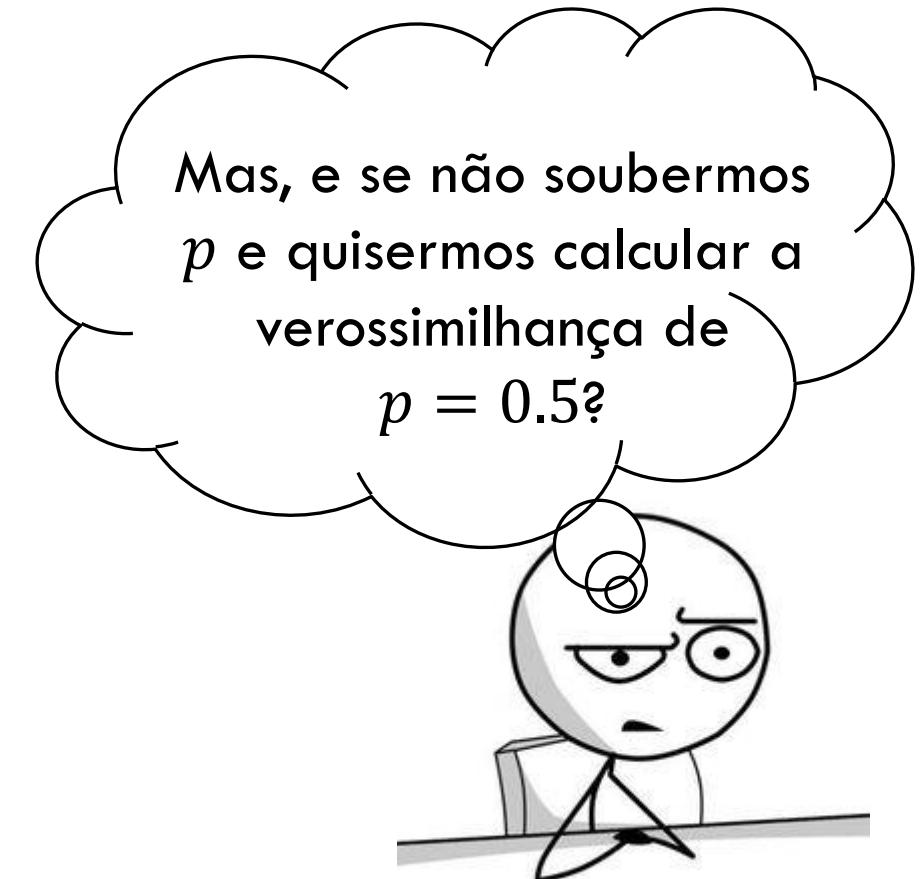
A probabilidade de ocorrer x , dado n e p

$$P(x = 4 | n = 7, p = 0.5)$$



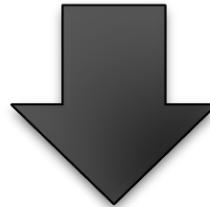
$$P(p = 0.5 | n = 7, x = 4)$$

A verossimilhança de p , dado n e x



A probabilidade de ocorrer x , dado n e p

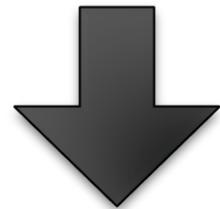
$$P(x = 4|n = 7, p = 0.5)$$



$$P(p = 0.5|n = 7, x = 4) = \left(\frac{7!}{4!(7-4)!} \right) 0.5^4 (1 - 0.5)^{7-4} = 0.273$$

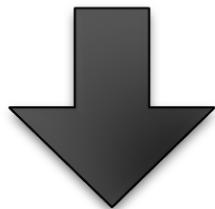
A verossimilhança de p , dado n e x

Podemos variar p



$$P(p = 0.5 | n = 7, x = 4) = \left(\frac{7!}{4! (7 - 4)!} \right) 0.5^4 (1 - 0.5)^{7-4} = 0.273$$

Enquanto n, x
permanecem
constantes



$$P(p = 0.5 | n = 7, x = 4) = \left(\frac{7!}{4! (7 - 4)!} \right) 0.5^4 (1 - 0.5)^{7-4} = 0.273$$

Em outras palavras, podemos calcular a verossimilhança para diferentes valores de p , dado que 4 em 7 pessoas preferem coca zero.

POR EXEMPLO

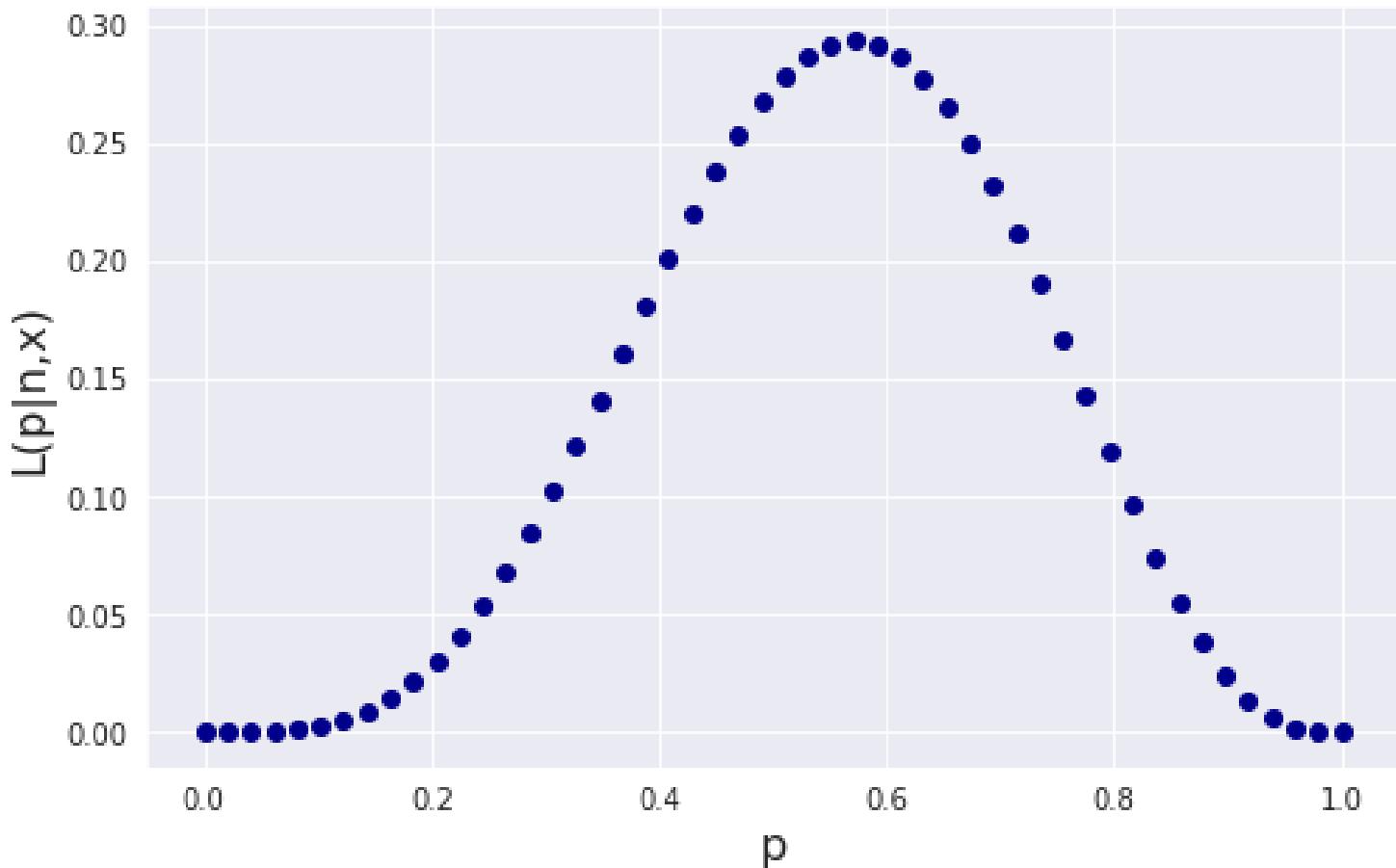
$$P(p = 0.25 | n = 7, x = 4) = \left(\frac{7!}{4! (7 - 4)!} \right) 0.25^4 (1 - 0.25)^{7-4} = 0.058$$

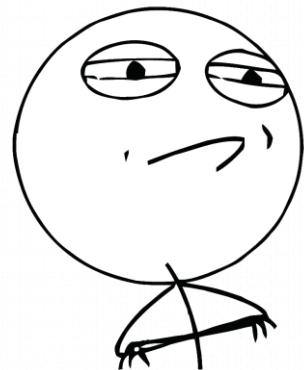
Veja que a verossimilhança de $p = 0.25$, dado que 4 em 7 preferem coca zero é menor que a verossimilhança de $p = 0.5$.

$$P(p = 0.57 | n = 7, x = 4) = \left(\frac{7!}{4! (7 - 4)!} \right) 0.57^4 (1 - 0.57)^{7-4} = 0.294$$

Veja que a verossimilhança de $p = 0.57$, dado que 4 em 7 preferem coca zero é maior que a verossimilhança de $p = 0.5$.

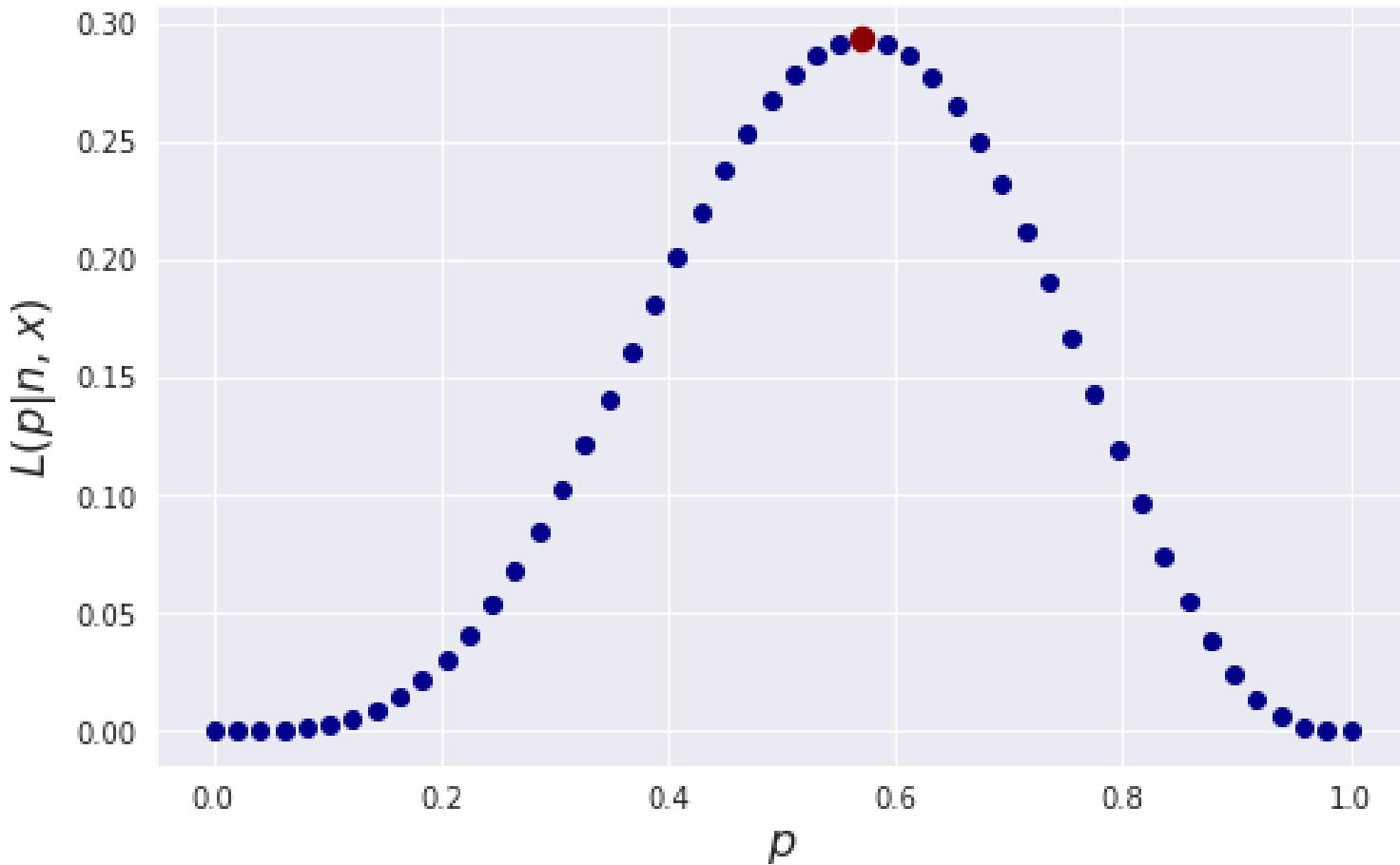
CURVA



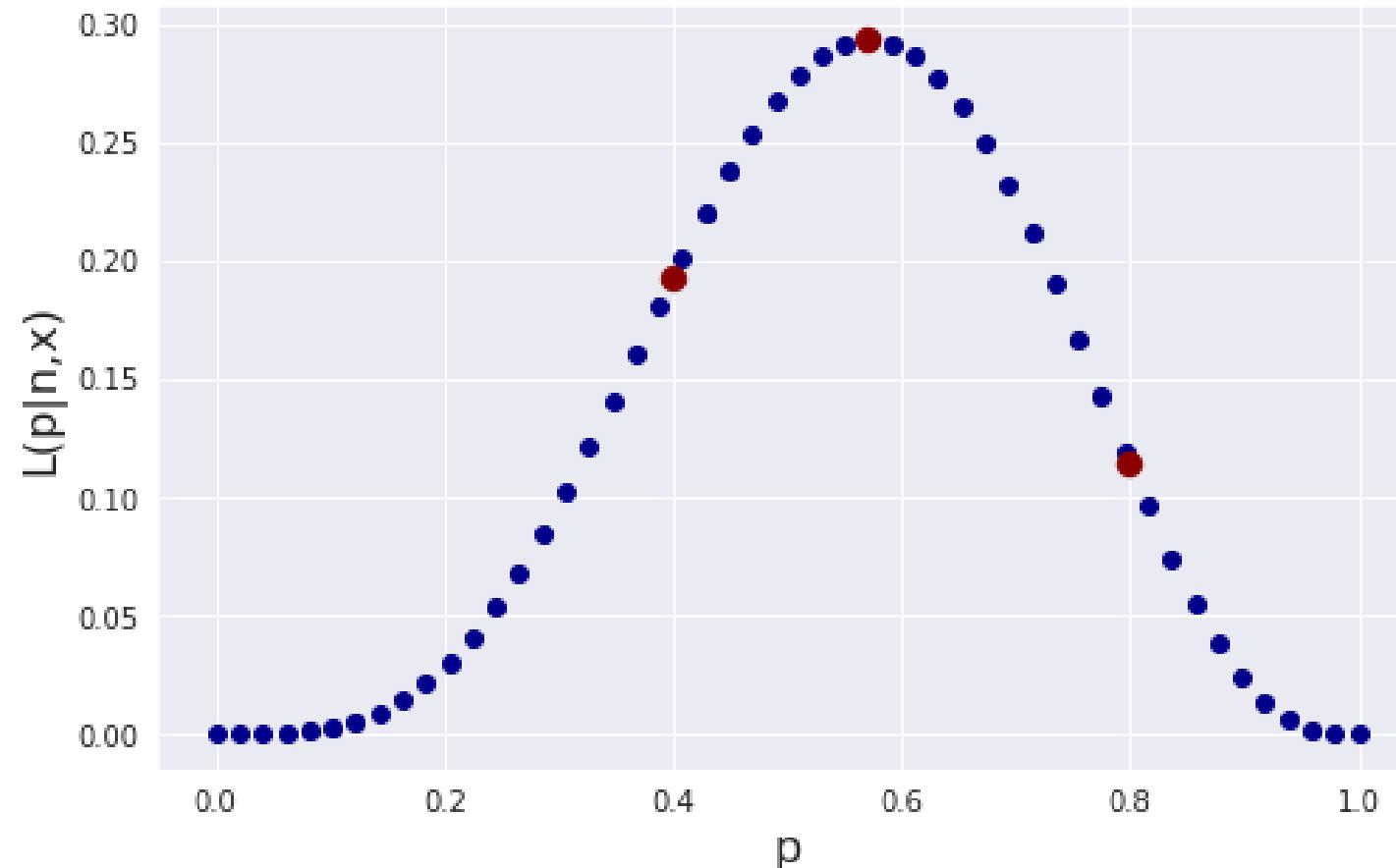


Como calcular MLE?

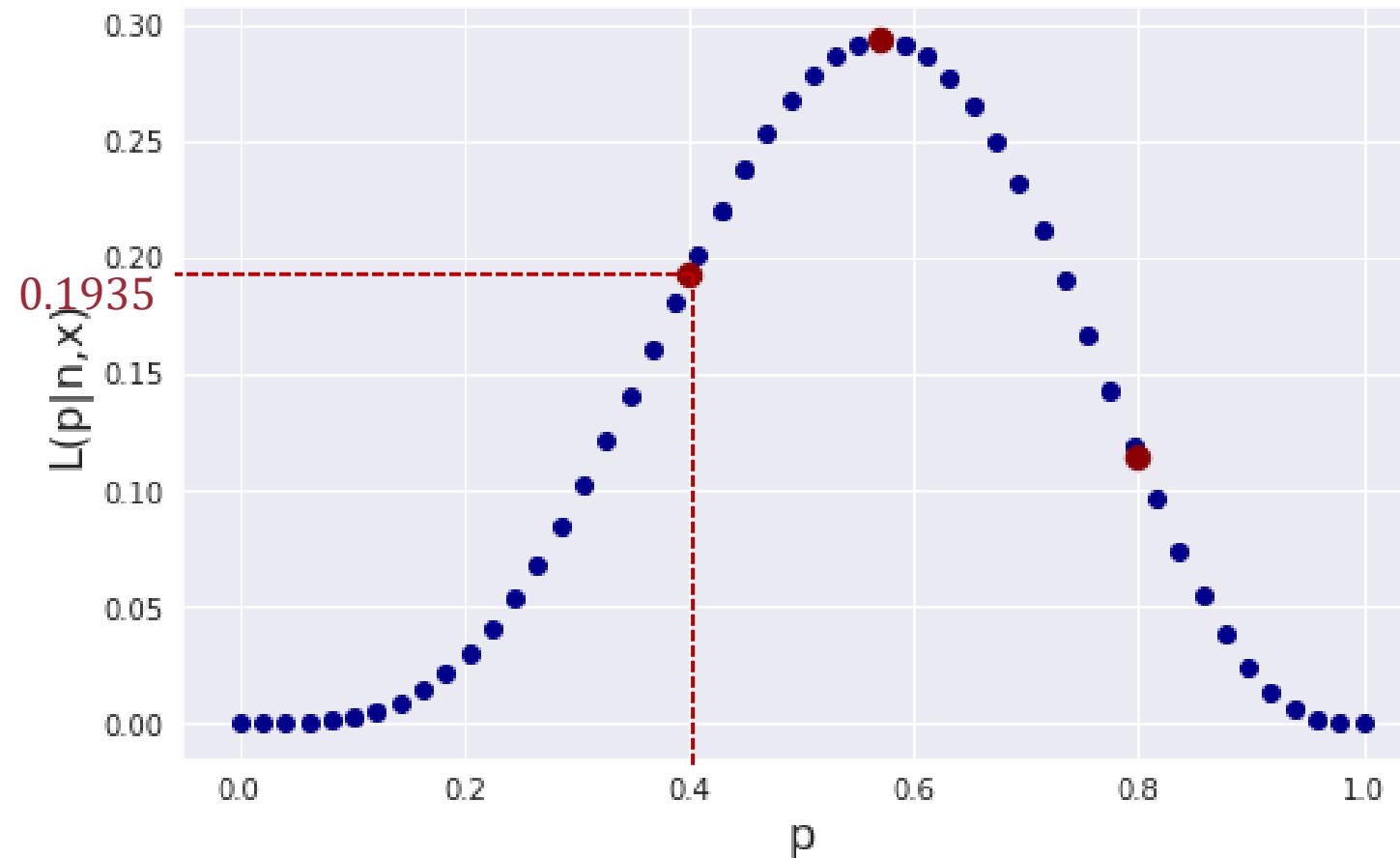
(do inglês, Maximum Likelihood Estimation)

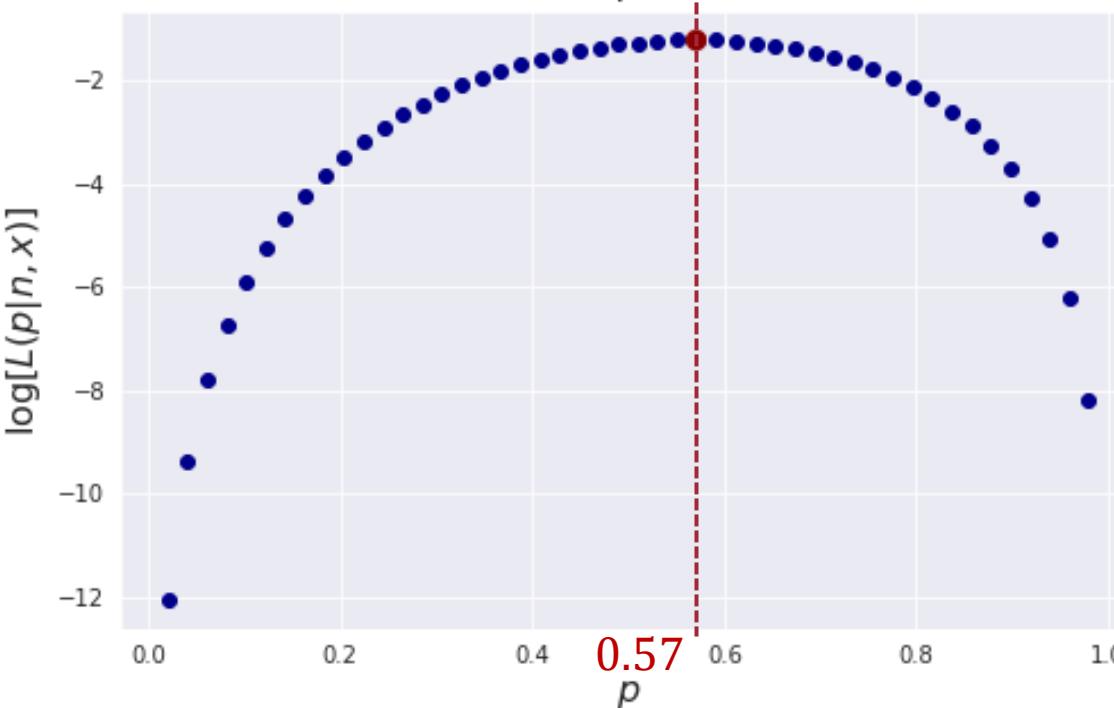
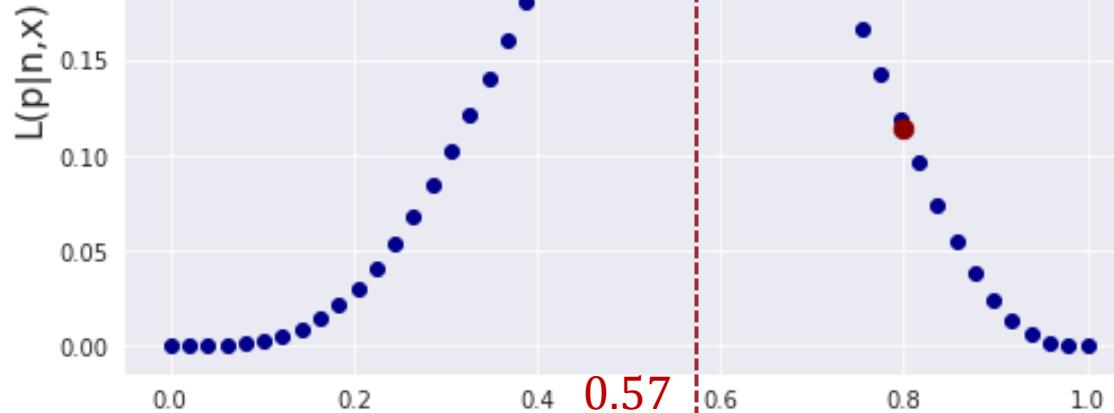


Mas... O mais importante é
entender o que significa cada
ponto nessa curva???



$$\mathcal{L}(p = 0.4|n, x) = 0.1935 \cong 20\%$$





Pode-se provar que:

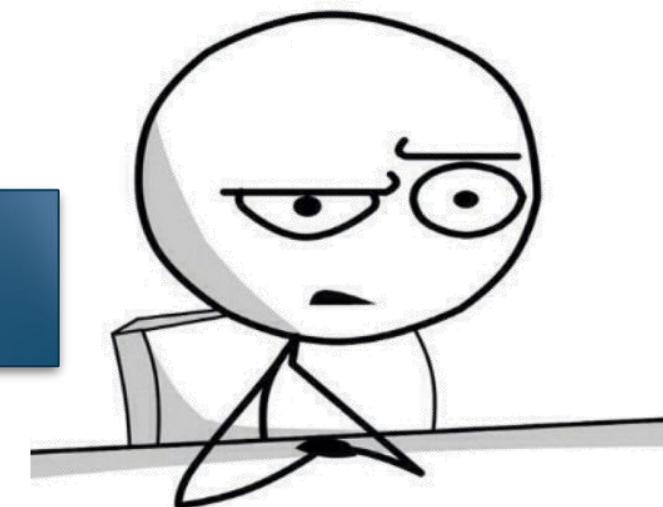
$$\mathcal{L}(p|n, x) = \binom{n}{x} p^x (1-p)^{n-x}$$

ou:

$$\ln \mathcal{L}(p|n, x) = \ln \binom{n}{x} \ln p^x \ln(1-p)^{n-x}$$

têm seu máximo em,

$$p = \frac{x}{n}$$



MLE

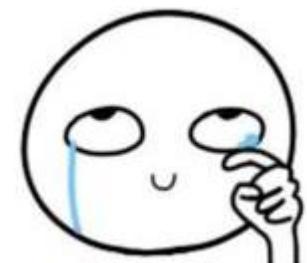
$$\mathcal{L}(p|n=7, x=4) = \left(\frac{7!}{4!(7-4)!} \right) p^4 (1-p)^{7-4}$$

Vamos usar logaritmo

$$\ln \mathcal{L}(p|n=7, x=4) = \ln \left(\frac{7!}{4!(7-4)!} p^4 (1-p)^{7-4} \right)$$

$$\ln \mathcal{L}(p|n=7, x=4) = \ln \left(\frac{7!}{4!(7-4)!} \right) + 4 \ln p + (7-4) \ln(1-p)$$

Pronto para derivar e igualar a zero???



$$\ln \mathcal{L}(p|n=7, x=4) = \boxed{\ln \left(\frac{7!}{4! (7-4)!} \right)} + \boxed{4 \ln p} + \boxed{(7-4) \ln(1-p)}$$

$$\frac{d \ln \mathcal{L}(p|n=7, x=4)}{dp} = 0 + 4 \frac{1}{p} - 3 \frac{1}{1-p} = 0$$

Multiplicando os dois lados por $p(1-p)$:

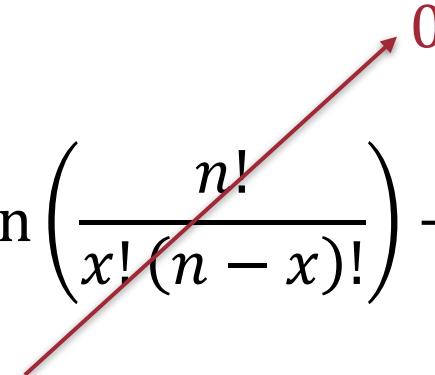
$$4(1-p) - 3p = 0$$

Chega-se a,

$$p = \frac{4}{7} = 0.57$$

GENERALIZANDO

$$\frac{d \ln \mathcal{L}(p|n,x)}{dp} = \ln\left(\frac{n!}{x!(n-x)!}\right) + x \ln p + (n-x) \ln(1-p)$$



$$\frac{d \ln \mathcal{L}(p|n,x)}{dp} = x \frac{1}{p} - (n-x) \frac{1}{1-p} = 0$$

$$x(1-p) + xp - np = 0$$

$$x - xp + xp - np = 0$$

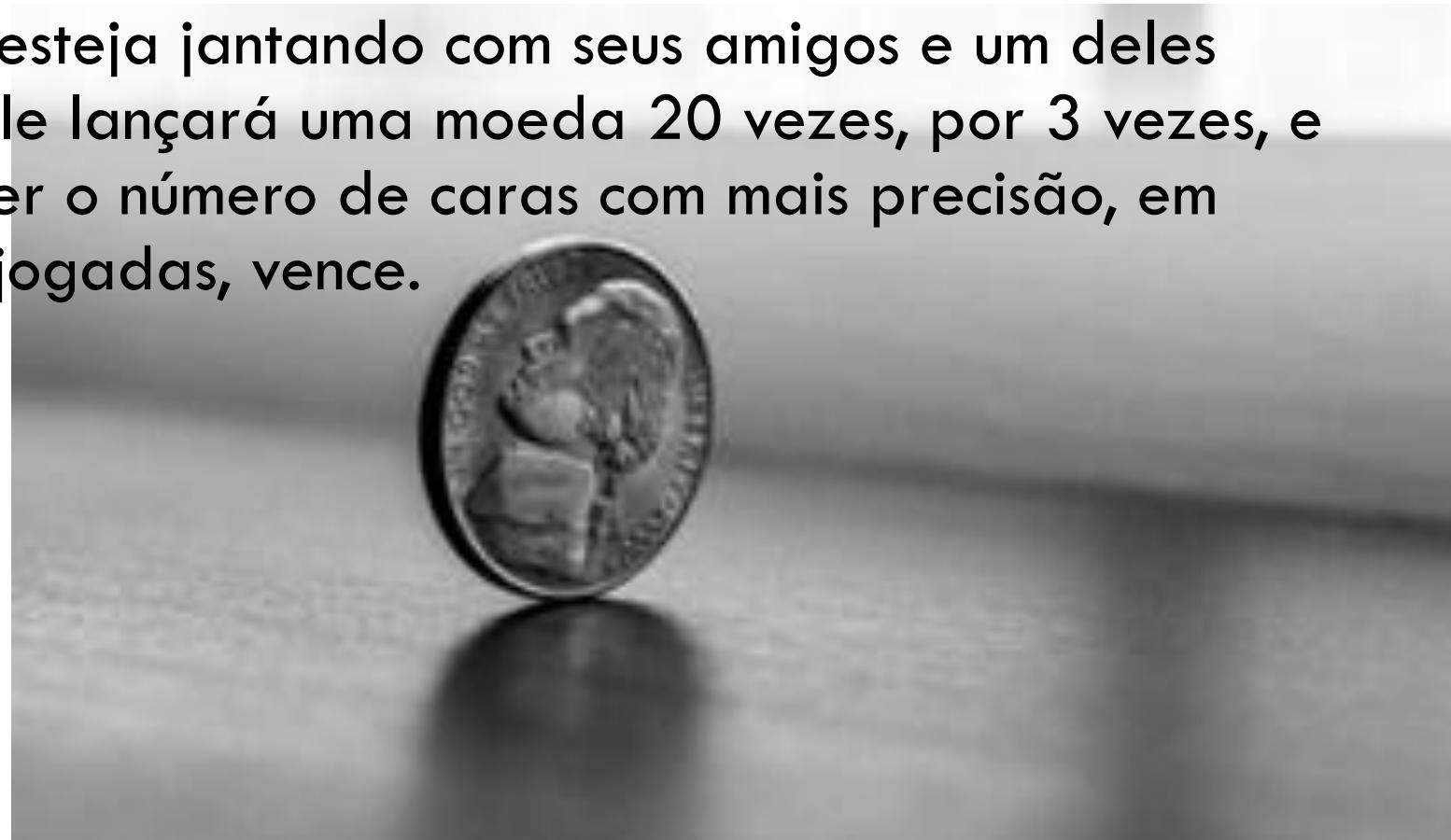
$$p = \frac{x}{n}$$

\rightarrow Número de sucessos
 \rightarrow Número de tentativas

$$p = \frac{1}{n} \sum_i x_i$$

SEU PROBLEMA

Suponha que você esteja jantando com seus amigos e um deles sugere um jogo – ele lançará uma moeda 20 vezes, por 3 vezes, e a pessoa que prever o número de caras com mais precisão, em qualquer uma das jogadas, vence.



VAMOS ANALISAR ENTÃO...

A Posteriori: nossa crença refinada da hipótese, quando levamos em conta a evidência.

Quão honesta é a moeda do seu amigo, dado que você viu um determinado número de caras para as 20 tentativas.

$$P(\text{Hipótese} | \text{Evidência}) = P(\text{Hipótese})$$

A Priori: aquilo que acreditamos antes de qualquer experimento. É uma espécie de potência da inferência bayesiana. Sua crença em quão honesta a moeda de seu amigo é. Será baseando-se nesta hipótese que você irá fazer seu primeiro chute.

Verossimilhança: é a probabilidade condicionada da evidência, dada a hipótese. Dado o nível de honestidade da moeda de seu amigo, quantas caras se espera para as 20 tentativas.

$$\frac{P(\text{Evidência} | \text{Hipótese})}{P(\text{Evidência})}$$

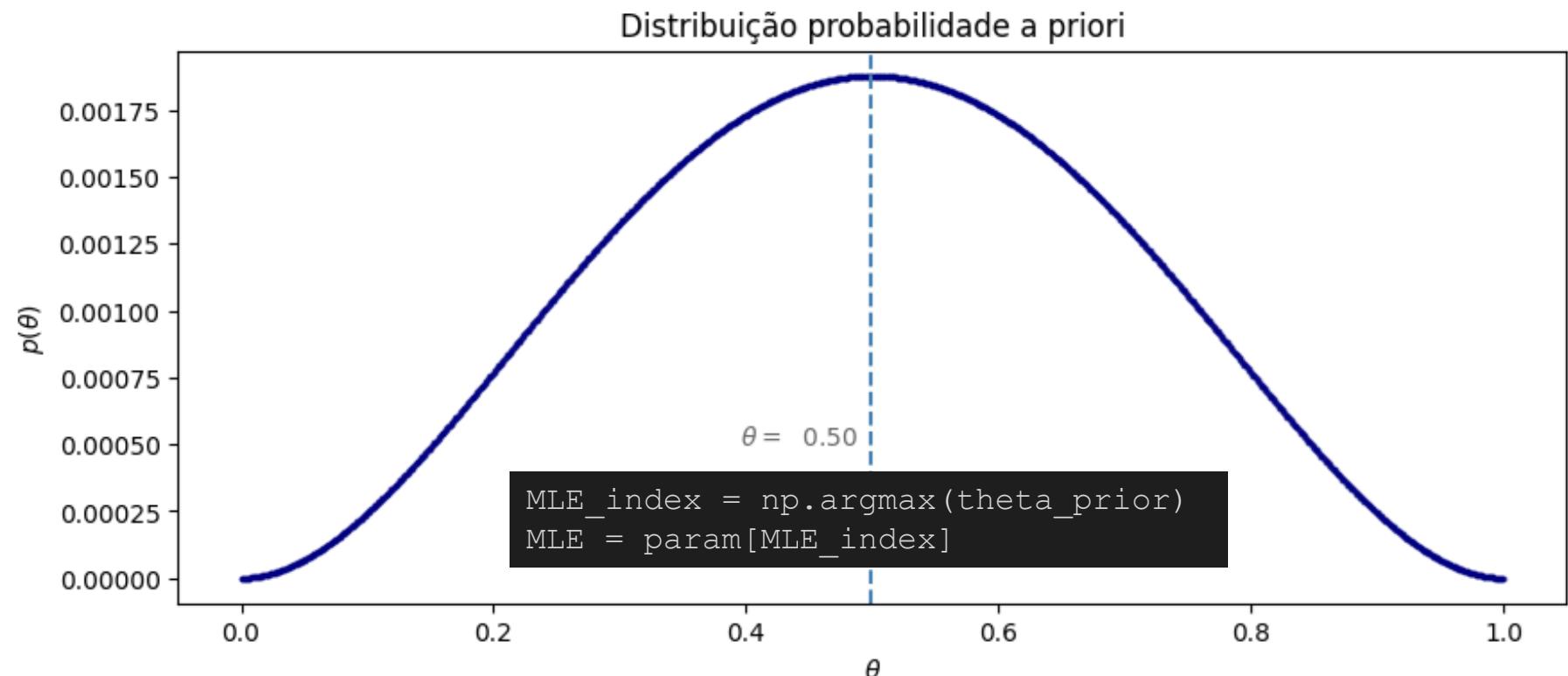
Precisamos modelar todas essas probabilidades. Essa é a principal característica da Estatística Bayesiana: define a probabilidade como uma crença (que pode ser forte ou fraca).

Evidência: soma das probabilidades ponderadas de todos os valores possíveis das hipóteses. Se tivéssemos várias observações de níveis de honestidade da moeda (mas não sabíamos ao certo), este valor nos diz a probabilidade de ver uma certa sequência de movimentos (evidência) para todas as possibilidades de nossa crença na honestidade da moeda.

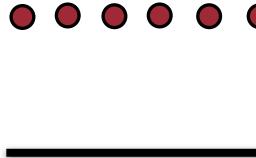
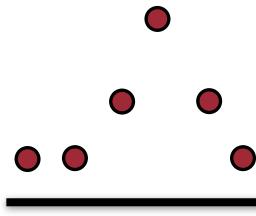
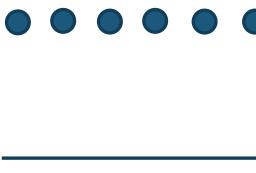
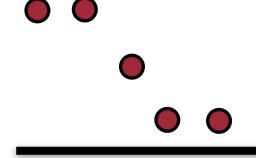
A PRIORI

Sem nem mesmo jogar a moeda uma vez, você usa sua experiência passada em observações de moedas e, portanto, espera que a probabilidade de observar caras seja 0,5.

```
moeda_prior = [0]*2+[1]*2 # moeda_prior = [0, 0, 1, 1]
param = linspace(0,1,1000)
theta_prior = np.array([np.prod(bernoulli.pmf(moeda_prior, p)) for p in param])
# Normalizamos o valor, dado que não nos preocuparemos com p(E)
theta_prior = theta_prior/np.sum(theta_prior)
theta_initial = theta_prior
```



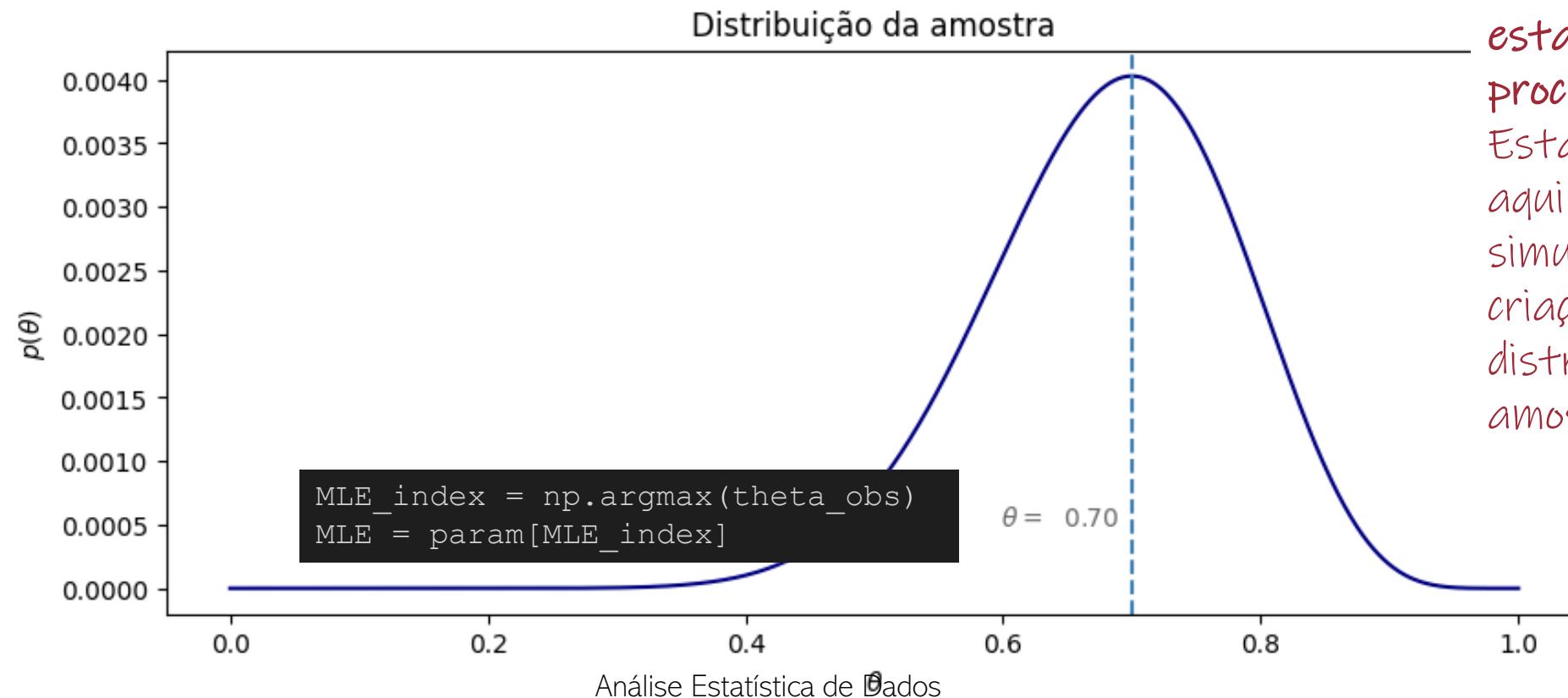
A CRENÇA A PRIORI É SUA...

Modelo	Intervalo	Forma	Forte	Baixo
Constante	$[a, b]$			
Binomial	$[0, +\infty]$			
Poisson	$[0, +\infty]$			

VEROSSIMILHANÇA

```
theta_obs = np.array([np.prod(bernoulli.pmf(moeda_obs, p)) for p in param])
theta_obs = theta_obs/np.sum(theta_obs)
```

Vamos supor que a nossa moeda, na realidade, não seja equilibrada ($\theta = 0.77$).



Veja que o valor real $\theta=0.77$ é o valor que não temos e estamos procurando.
 Estamos usando aqui para simularmos a criação da distribuição de amostragem.

P(EVIDÊNCIA)

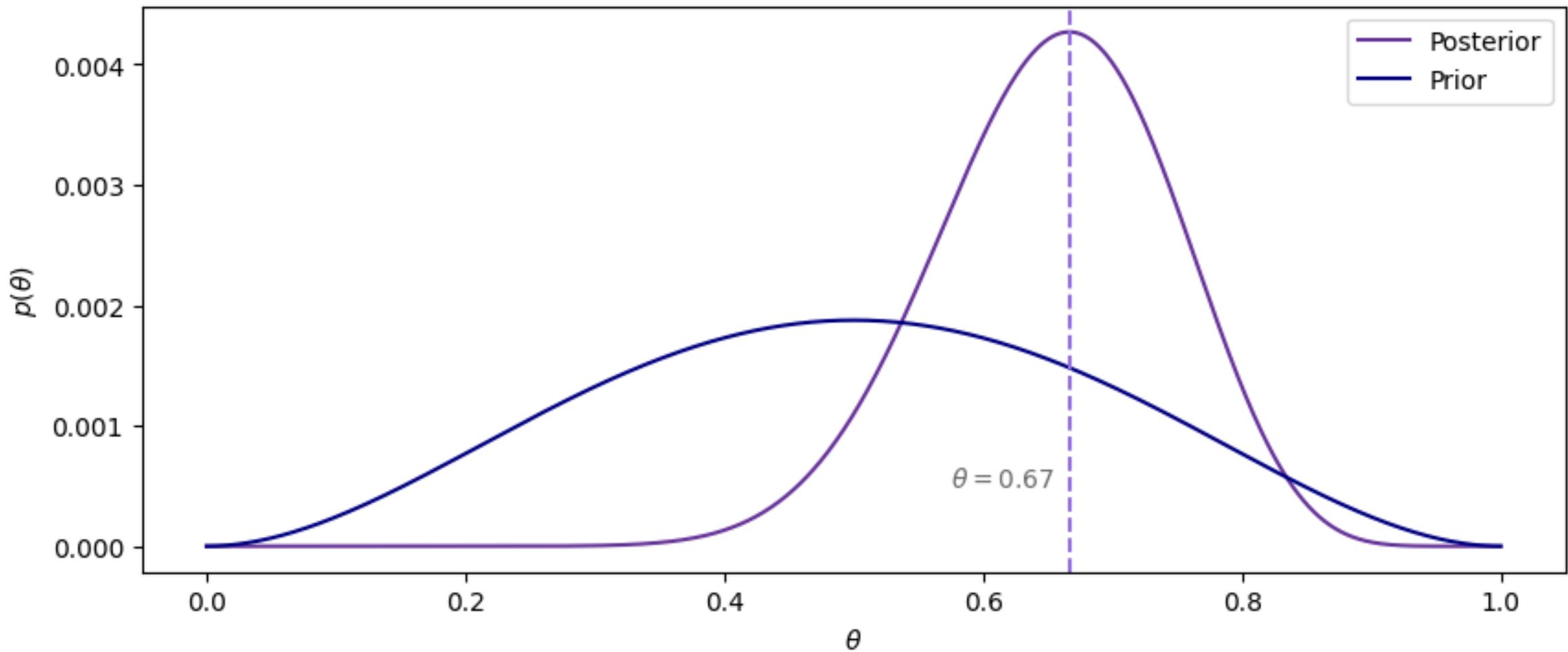
Como mencionado, *P(Evidência)* atua como uma constante normalizadora para tornar a densidade posterior adequada. Como esse denominador simplesmente escala a densidade posterior para torná-la uma densidade adequada, o Teorema de Bayes para distribuições de probabilidade é frequentemente indicado como:

$$P(\text{Hipótese}|\text{Evidência}) \propto P(\text{Hipótese})P(\text{Evidência}|\text{Hipótese})$$

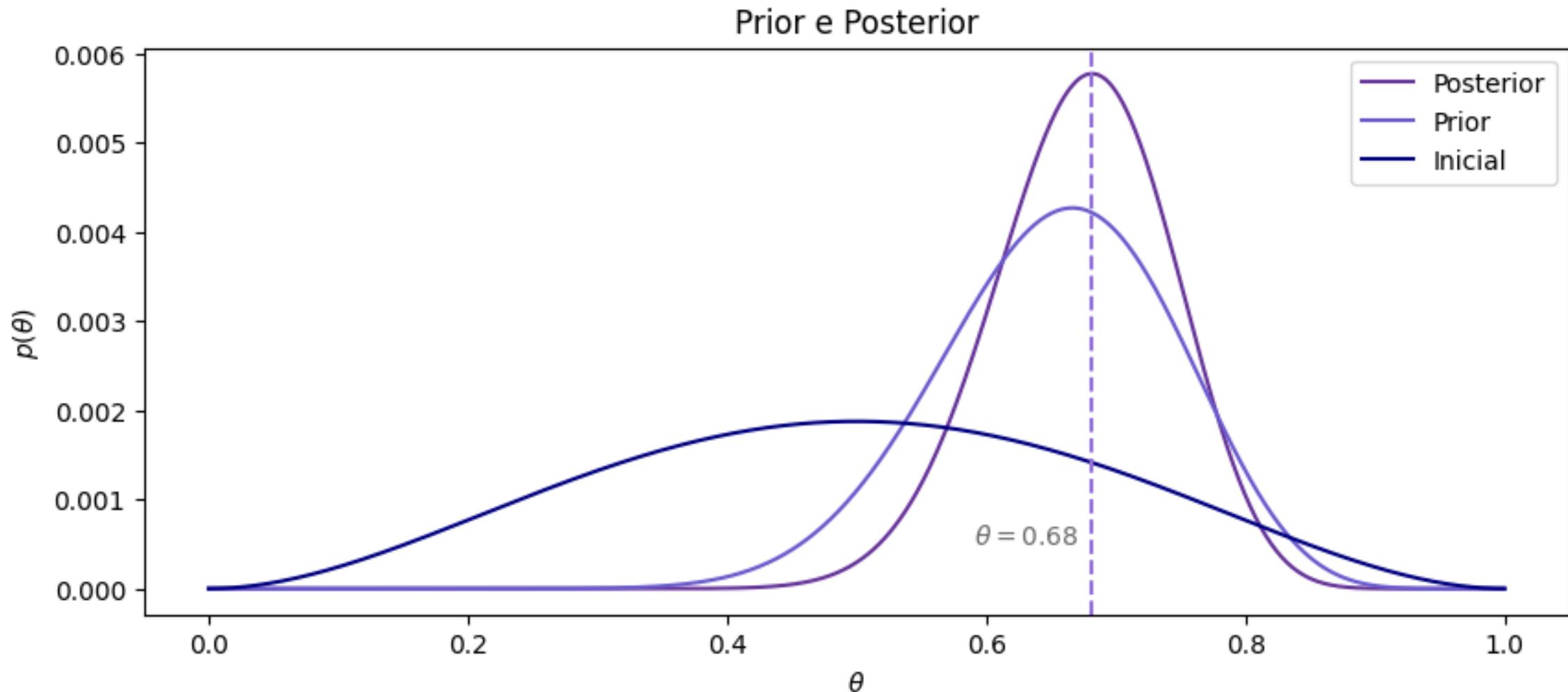
O que realmente importa aqui é o que chamamos de função de verossimilhança e a priori.

$p(H|E) \propto p(E|H) * p(H)$ OU Posterior \propto Verossimilhança * Prior

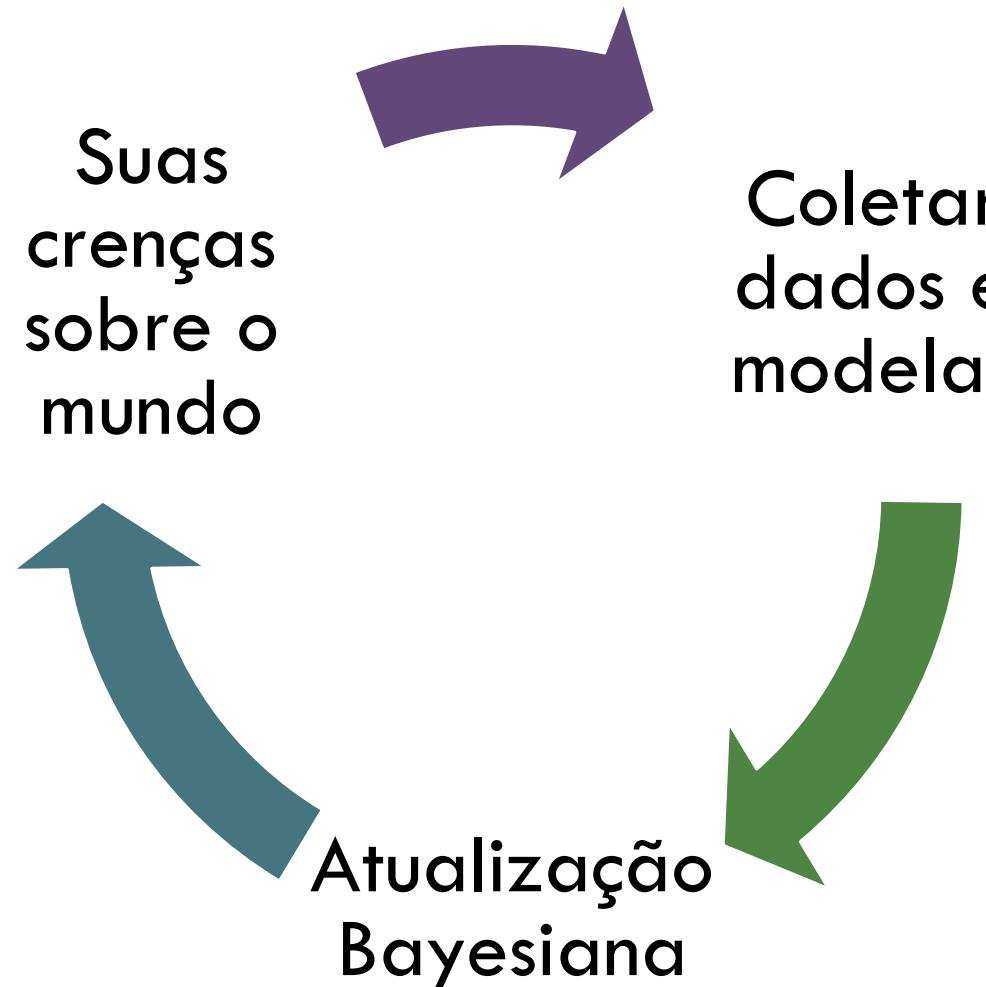
Prior e Posterior



E ASSIM, ITERATIVAMENTE...



MÉTODO BAYESIANO



Passos para uma atualização Bayesiana:

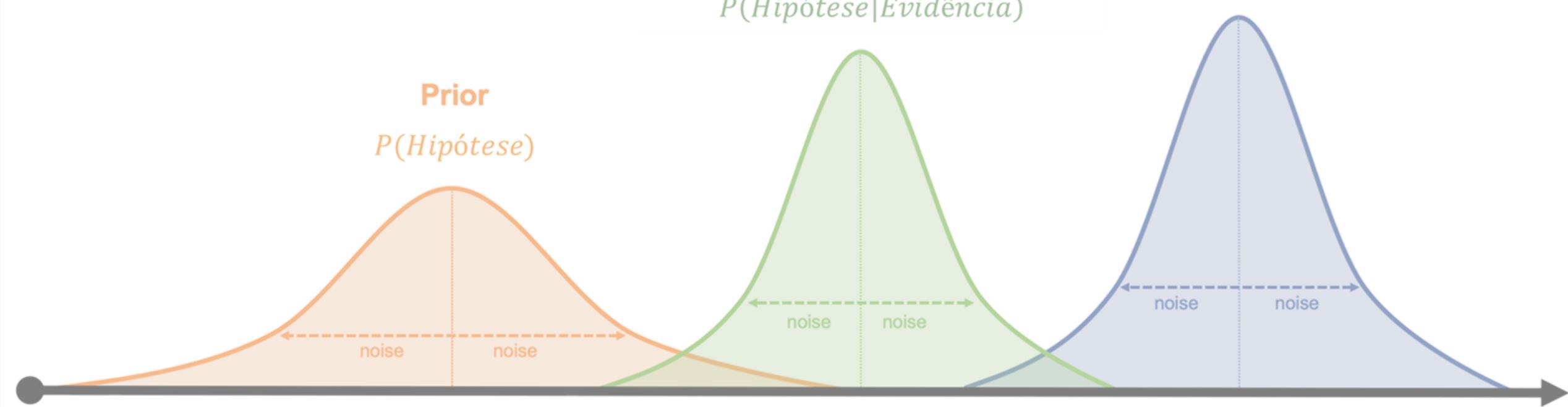
1. Estabeleça uma crença – a primeira distribuição de probabilidade a priori;
2. Colete dados e modele;
3. Atualize suas crenças usando os novos dados para definir sua distribuição de probabilidade a posteriori;
4. Repita os passos 2 e 3 usando o posterior do passo 3 como seu novo priori.

Likelihood

Posterior
 $P(\text{Hipótese}|\text{Evidência})$

Likelihood
 $P(\text{Evidência}|\text{Hipótese})$

Prior
 $P(\text{Hipótese})$



$$P(\text{Hipótese}|\text{Evidência}) = P(\text{Hipótese}) \frac{P(\text{Evidência}|\text{Hipótese})}{P(\text{Evidência})}$$

SUA LIÇÃO DE CASA

Você terá uma questão sobre um dos modelos estudados (1) e outra que permitirá fazer upload (2).

No Moodle, será definida uma crença inicial e o valor valor p de probabilidade de número de caras da moeda que seu amigo irá jogar.

Quanto à crença inicial, serão definidos dois dados:

1. Se a moeda é honesta, é viesada para maior número de caras, ou é viesada para maior número de coroas;
2. Se a sua crença será forte ou fraca.

Com esses dados, escreva um texto que deve ser ilustrado por, no mínimo, 3 gráficos. O texto deve conter a definição, com justificativa, de quais seriam seus "chutes" inicial, e depois de cada vez que seu amigo joga 20x a moeda; escreva também as suas conclusões a respeito do desempenho de seus resultados. Os 3 gráficos sugeridos são:

1. crença inicial;
2. sua crença atualizada após a primeira vez que seu amigo joga a moeda 20x,
3. sua crença atualizada após a segunda vez que seu amigo joga a moeda 20x.

NEVER
GIVE UP



ACABOU...

Reveja a aula antes
de resolver os
exercícios.