Análise Estatística de dados

Inteligência Artificial

PECE Programa de Educação Continuada
Escola Politécnica da USP

USP

AULA 10A – MODELOS DE PROBABILIDADE CONTÍNUOS CONT...

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.

Distribuições: Poisson, Exponencial, Gama e Beta

PECE Programa de Educação Continuada
Escola Politécnica da USP

DISTRIBUIÇÃO DE POISSON

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),

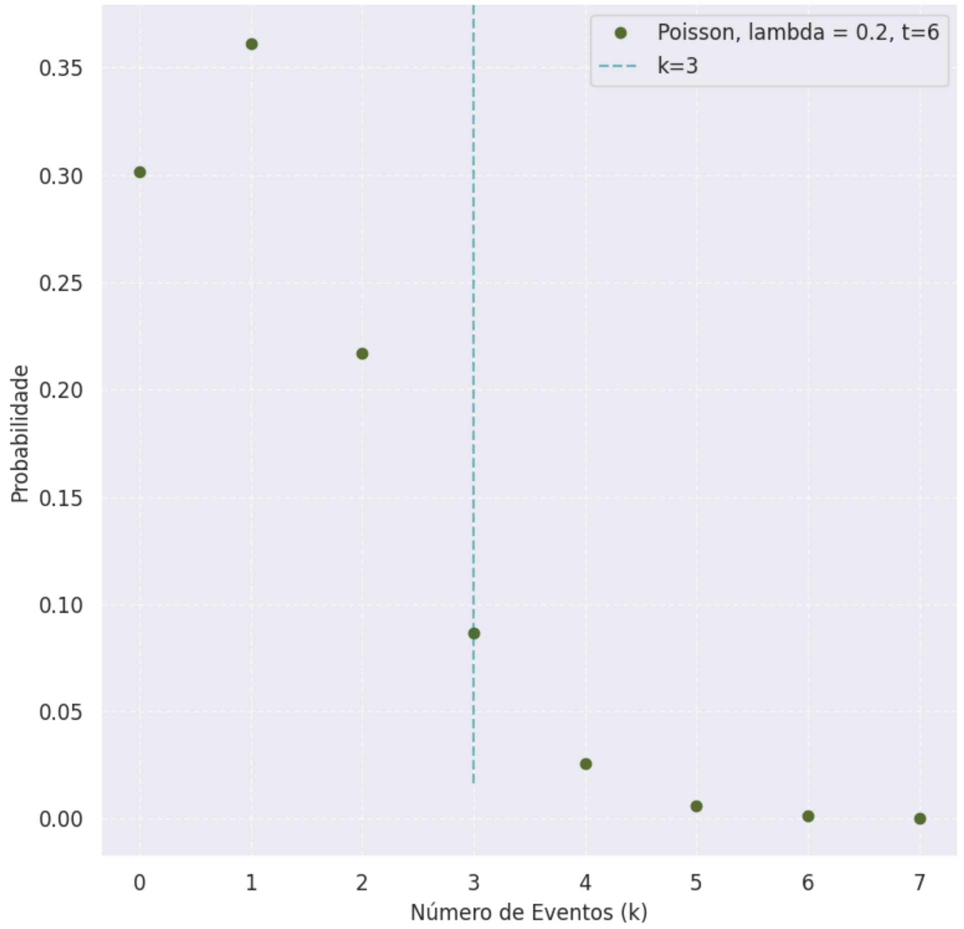
$$P(x) = \frac{e^{-\lambda t}(\lambda t)^x}{x!}$$

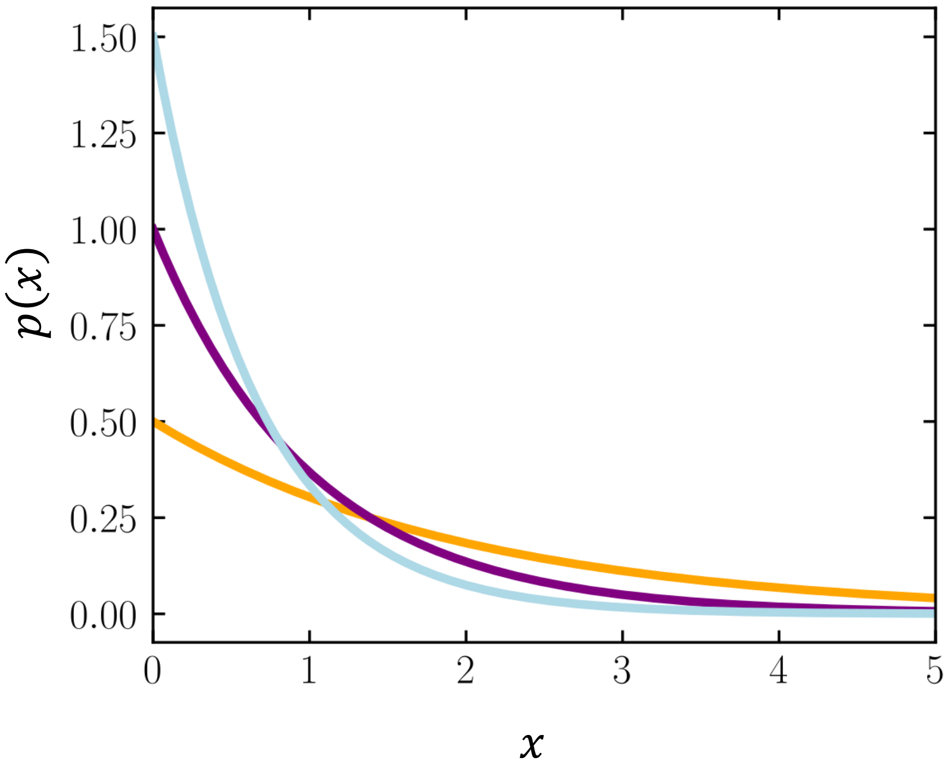
onde t representa a magnitude do intervalo contínuo e $x \geq 0$ e inteiro.

A distribuição é usualmente referenciada pelo símbolo $X \sim Po\ (\lambda)$. Valor esperado e a variância da distribuição são definidos, respectivamente, pelas relações: $E[x] = \lambda t$ e $\sigma_x^2 = \lambda t$.

EXEMPLO

Algumas lâmpadas duram, em média, cinco anos. Qual a probabilidade de que, em 6 anos, as 3 lâmpadas da sua garagem falhem.





DISTRIBUIÇÃO EXPONENCIAL

$p(x;\lambda)$

DISTRIBUIÇÃO EXPONENCIAL

Uma variável aleatória x contínua, que tome todos os valores possíveis não negativos, tem uma distribuição exponencial com uma taxa média constante $\lambda > 0$ se a sua função densidade de probabilidade for dada por:

$$p(x) = \lambda e^{-\lambda x}$$

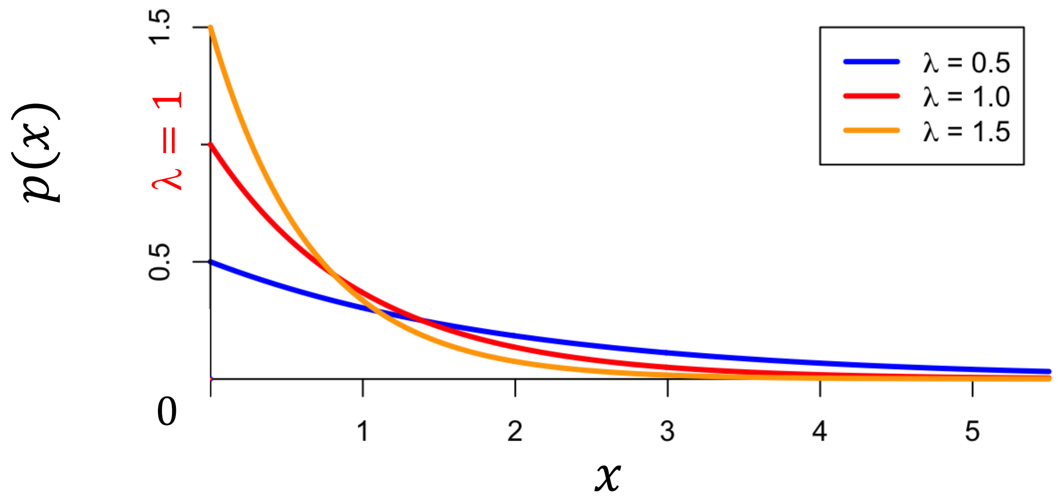
A média e variância são dadas por, respectivamente, $\mu = \frac{1}{\lambda}$ e $\sigma^2 = \frac{1}{\lambda^2}$

Frequentemente usada para calcular a quantidade de tempo entre eventos.

Durante uma chuva, o intervalo de tempo entre raios, por exemplo, tem uma distribuição exponencial. Outro exemplo é o tempo de duração de uma bateria de carro em meses.

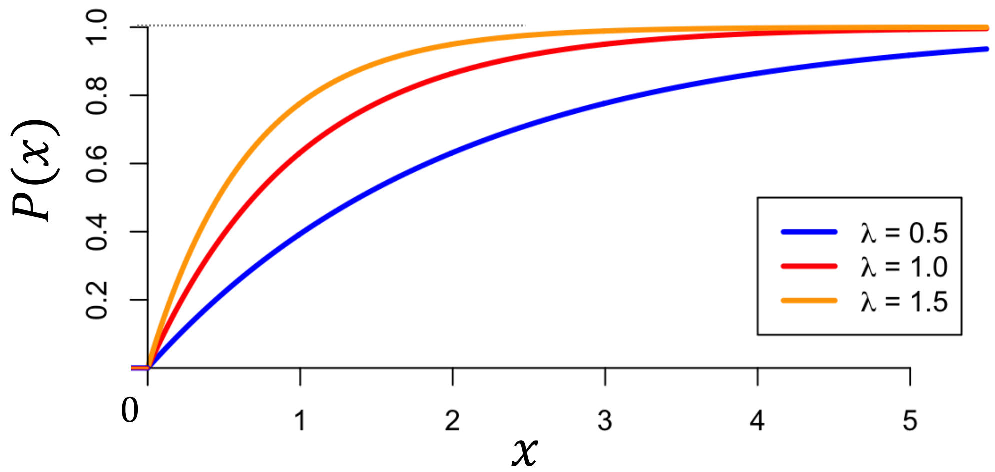
Função densidade de probabilidade da distribuição exponencial:

$$p(x) = \lambda e^{-\lambda x}$$



A distribuição acumulada da distribuição exponencial é facilmente obtida a partir da integração da função densidade de probabilidade:

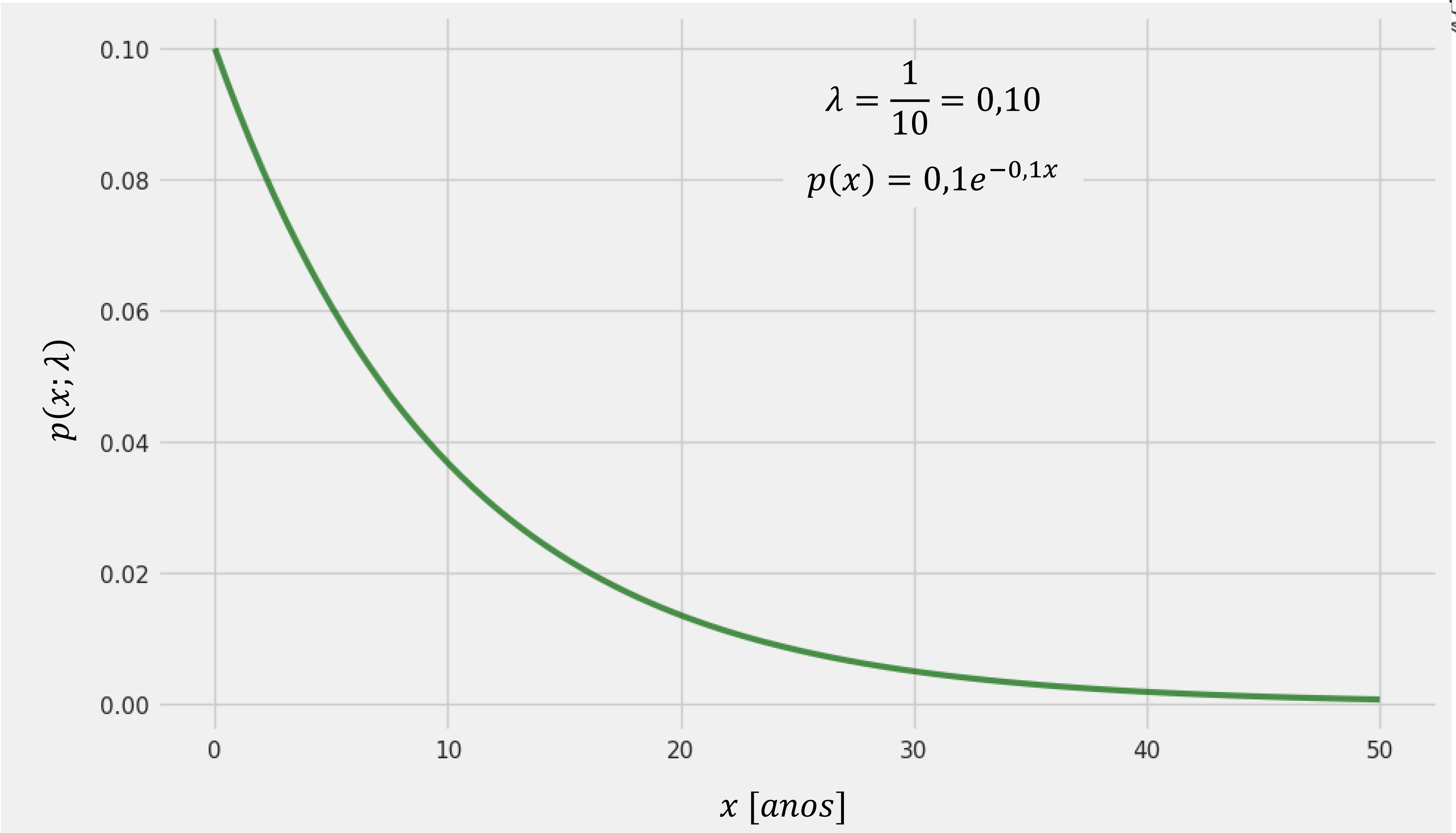
$$P(x) = 1 - e^{-\lambda x}$$



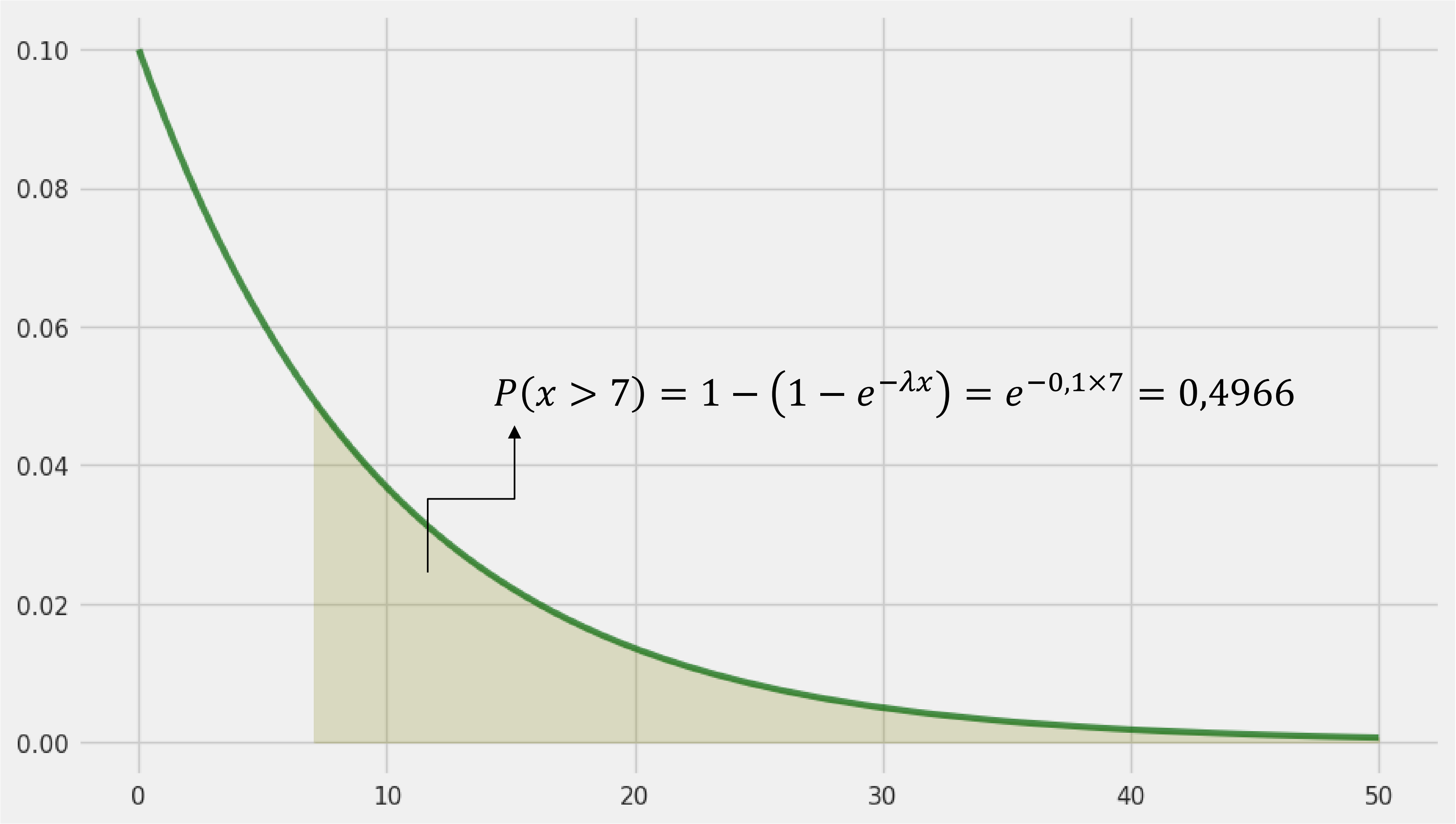
EXEMPLO

Em média, uma determinada peça do computador dura, em média, dez anos. O período de duração da peça do computador é distribuído exponencialmente.

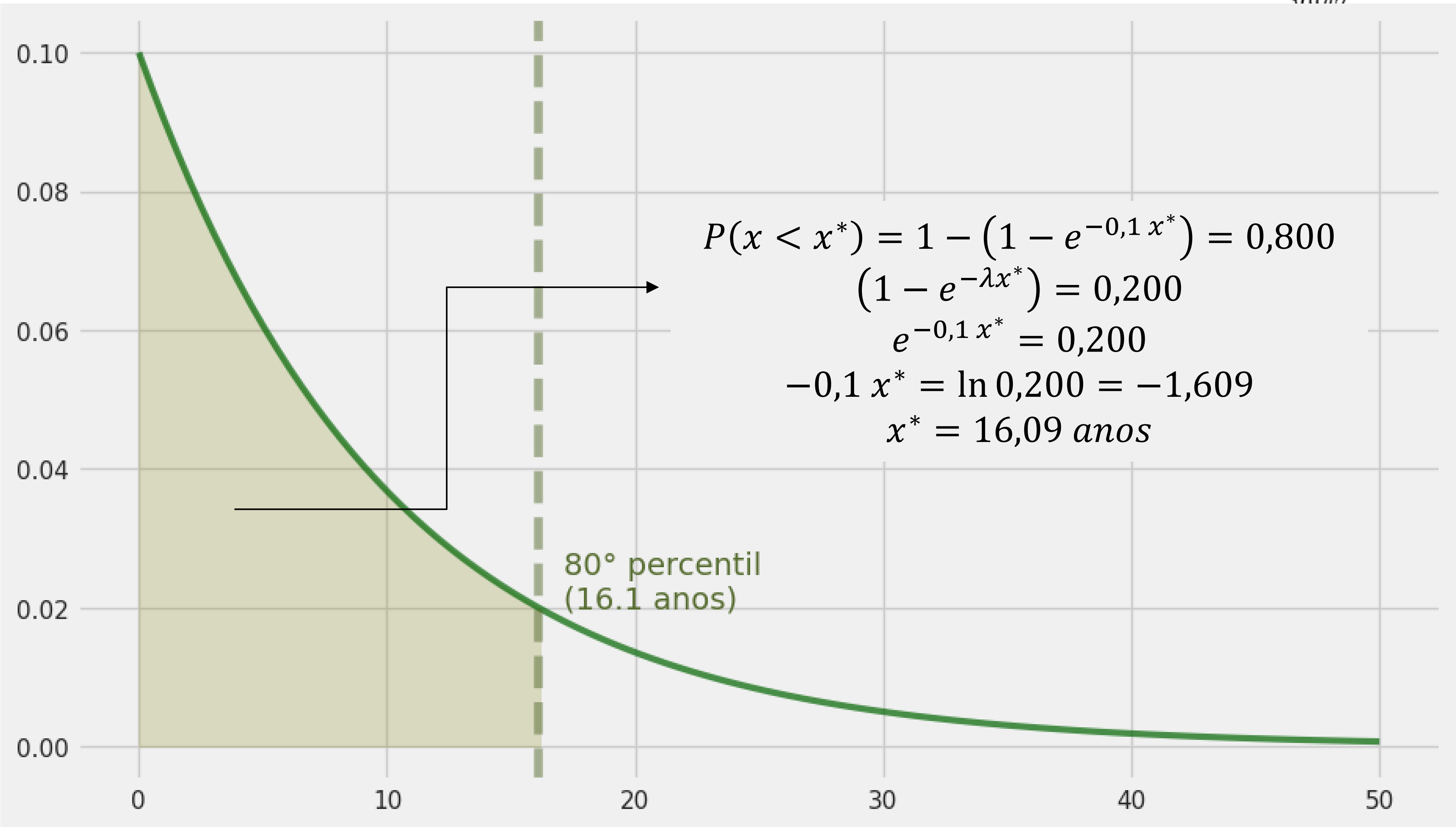
1. Desenhe a curva exponencial que modela o problema.
2. Qual é a probabilidade de uma peça de computador durar mais de 7 anos?
3. Quanto durará 80% das peças?
4. Qual é a probabilidade de uma peça de computador durar entre nove e 11 anos?


$\lambda = \frac{1}{10} = 0{,}10$
$p(x) = 0{,}1e^{-0{,}1x}$
$p(x;\lambda)$
0.10
0.08
0.06
0.04
0.02
0.00
0
10
20
30
40
50
$x\ [anos]$

Qual é a probabilidade de uma peça de computador durar mais de 7 anos?

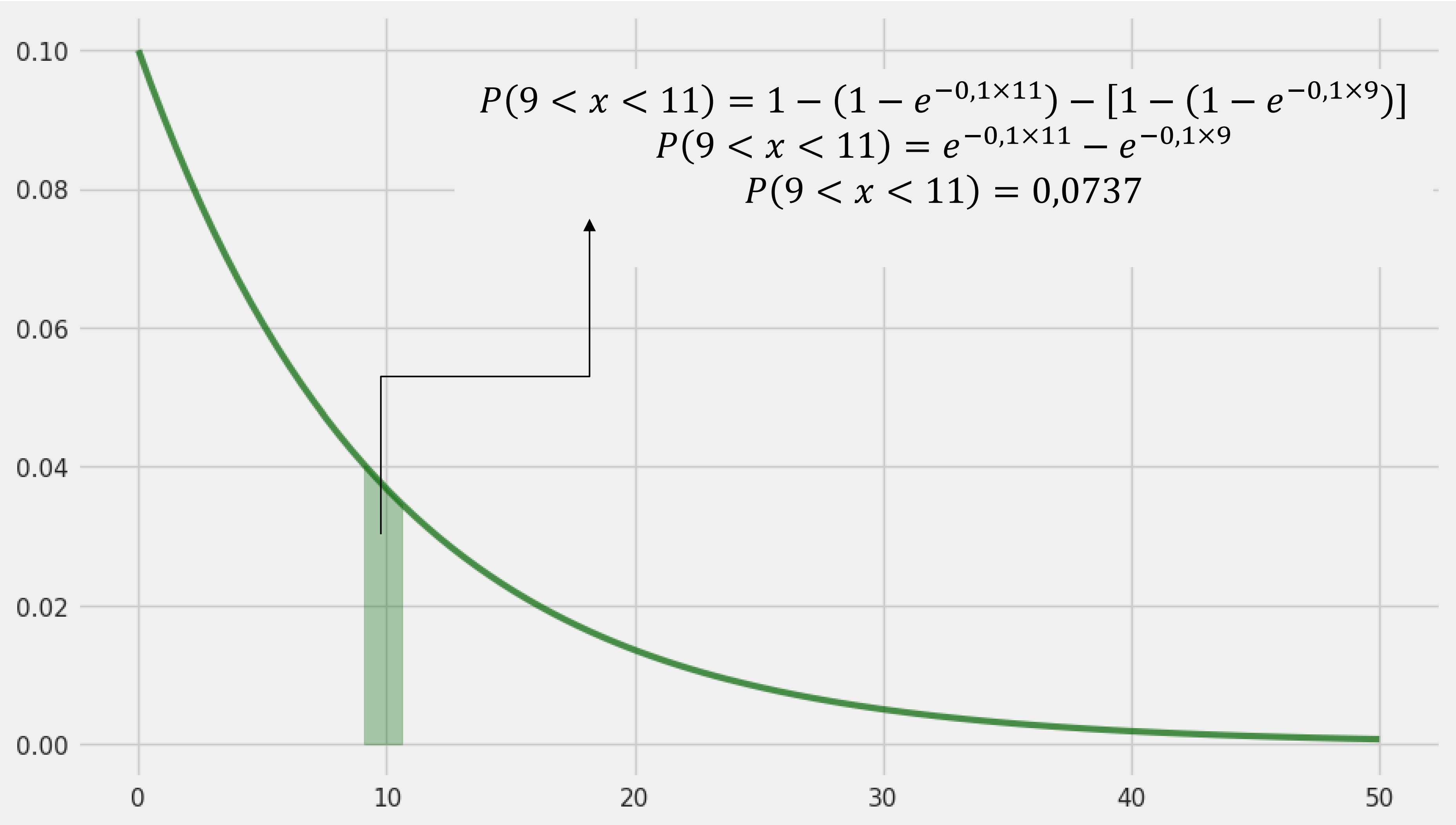


Quanto durará 80% das peças?



$$P(x < x^*) = 1 - \left(1 - e^{-0{,}1\,x^*}\right) = 0{,}800$$
$$\left(1 - e^{-\lambda x^*}\right) = 0{,}200$$
$$e^{-0{,}1\,x^*} = 0{,}200$$
$$-0{,}1\,x^* = \ln 0{,}200 = -1{,}609$$
$$x^* = 16{,}09\ anos$$

Qual é a probabilidade de uma peça de computador durar entre nove e 11 anos?



$$P(9 < x < 11) = 1 - (1 - e^{-0,1\times 11}) - [1 - (1 - e^{-0,1\times 9})]$$
$$P(9 < x < 11) = e^{-0,1\times 11} - e^{-0,1\times 9}$$
$$P(9 < x < 11) = 0,0737$$

ESSE EXEMPLO É SIMPLES...

Algumas lâmpadas duram, em média, cinco anos. Portanto, a vida útil X da lâmpada é uma variável aleatória **exponencial** com parâmetro $\lambda = \frac{1}{5} = 0{,}2$ falhas por ano.

Qual a média do tempo até queimar?

$$E[x] = \frac{1}{\lambda} = 5\ anos$$

Qual a variância do tempo de vida?

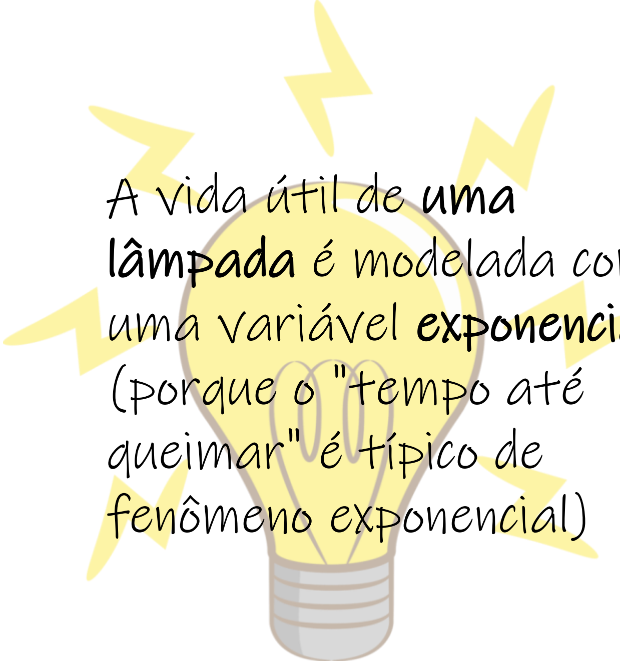
$$Var[x] = \frac{1}{\lambda^2} = 25\ anos^2$$

Qual a função densidade de probabilidade de X?

$$p(x) = \lambda e^{-\lambda x} = 0{,}2\ e^{-0{,}2x}$$

Qual a probabilidade da lâmpada durar mais que 6 anos?

$p(X > x) = e^{-\lambda x} \rightarrow p(X > 6) = e^{-0{,}2 \times 6} = 0{,}3012$ →30% de chance de durar mais de 6 anos.



A vida útil de **uma lâmpada** é modelada como uma variável **exponencial** (porque o "tempo até queimar" é típico de fenômeno exponencial)

VAMOS UM POUCO ADIANTE...

Algumas lâmpadas duram, em média, cinco anos. Seja X o tempo de duração de uma garagem com seis lâmpadas. Qual o tempo esperado para troca das lâmpadas, dado que **trocamos todas** quando a **primeira lâmpada queimar**?

Cada lâmpada individualmente tem uma distribuição exponencial com taxa de falha $\lambda = 1/5$.
E você está esperando a primeira queimar.

O tempo até o primeiro evento **também é Exponencial**, mas com taxa:

$$\lambda = 6 \times 1/5$$

Portanto,

$$E[x] = \frac{1}{\lambda} = \frac{5}{6} \approx 0{,}83\ anos$$

Como **todas** as lâmpadas estão funcionando ao mesmo tempo, a chance de *qualquer uma* queimar **é maior** do que a de uma única lâmpada. De fato, **as chances somam**, porque são processos independentes.

MAS...

Algumas lâmpadas duram, em média, cinco anos. Seja X o tempo de duração de uma garagem com seis lâmpadas. Qual o tempo esperado para troca das lâmpadas, dado que **trocamos todas** quando a **última lâmpada queimar**?

$\lambda = 1/5$ para **cada** lâmpada que queima a cada 5 anos, em média.

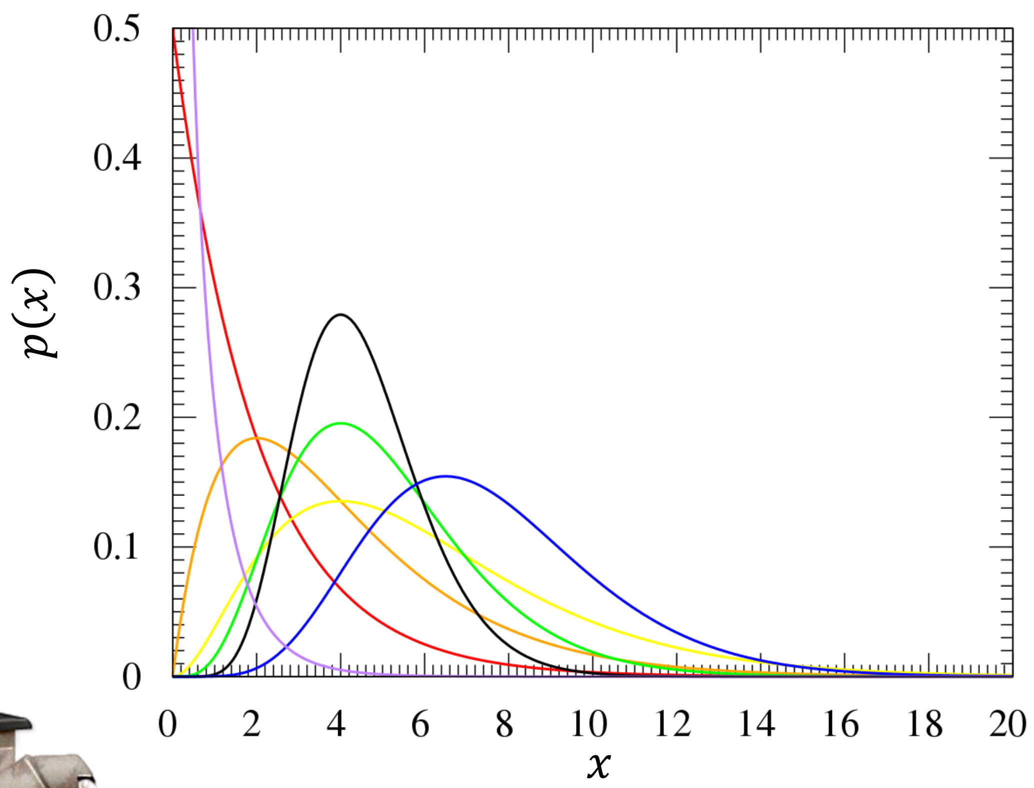
$$X_i \sim Exp(\lambda), i = 1,2,\cdots,6$$

são **variáveis aleatórias exponenciais independentes.**

Agora, para o **tempo total de falha de 6 lâmpadas,** ou seja, o somatório dos tempos de vida de 6 lâmpadas,

$$Z = X_1 + X_2 + \cdots + X_6$$

não resulta em uma distribuição exponencial, mas sim em uma distribuição **gama**. Isso ocorre porque a **distribuição gama** descreve o tempo até o k-ésimo evento em um processo de Poisson, o que se encaixa perfeitamente no contexto em que você está somando os tempos de vida de várias lâmpadas que queimam ao mesmo tempo.



DISTRIBUIÇÃO GAMA

$$p(x; \alpha, \beta)$$

DISTRIBUIÇÃO GAMA

Suponha que os eventos ocorram de forma independente e aleatória. O tempo de espera até que os primeiros α eventos ocorram é uma variável aleatória gama(α, β). Uma variável aleatória x tem distribuição Gama com parâmetros $\alpha > 0,\ \beta = \lambda > 0$ se sua função densidade de probabilidade é dada por X$\sim Gama(\alpha, \beta)$,

$$p(x; \alpha, \beta) = \begin{cases} \dfrac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)} & \text{se } x > 0 \\ 0 & cc \end{cases}$$

onde, $\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1} e^{-x} dx$ é a função gama completa. Se α é um número natural, $\Gamma(\alpha) = (\alpha - 1)!$ Além disso

Quando $\alpha = 1$, Gama se reduz à distribuição exponencial. O valor esperado e variância são dados, respectivamente, por, $\mu = \frac{\alpha}{\beta}$ e $\sigma^2 = \frac{\alpha}{\beta^2}$.

FUNÇÃO DE DISTRIBUIÇÃO ACUMULADA DE GAMA

$$P(t < x; \alpha, \beta) = \frac{\gamma(\alpha, x\beta)}{\Gamma(\alpha)}$$

Tem uma dedução dessa fórmula disponível aos interessados...

$\boxed{\gamma(\alpha, x\beta) = \int_0^{x\beta} t^{\alpha-1} e^{-t} dt \text{ é}}$ a função gama incompleta inferior e x é o valor do limite inferior que queremos acumular.

EXEMPLO

Carros chegam a um cruzamento a uma taxa, em média, de **1 a cada 2 minutos**. Seja X o tempo até que **5** carros cheguem.

Aqui, a chegada dos carros é **um processo de Poisson** — ou seja, os eventos (chegadas dos carros) são aleatórios, independentes e têm uma taxa constante $\lambda = 0{,}5\ carro/minuto$.

Para tempo até k eventos em Poisson, a variável que modela isso não é mais uma simples Exponencial.

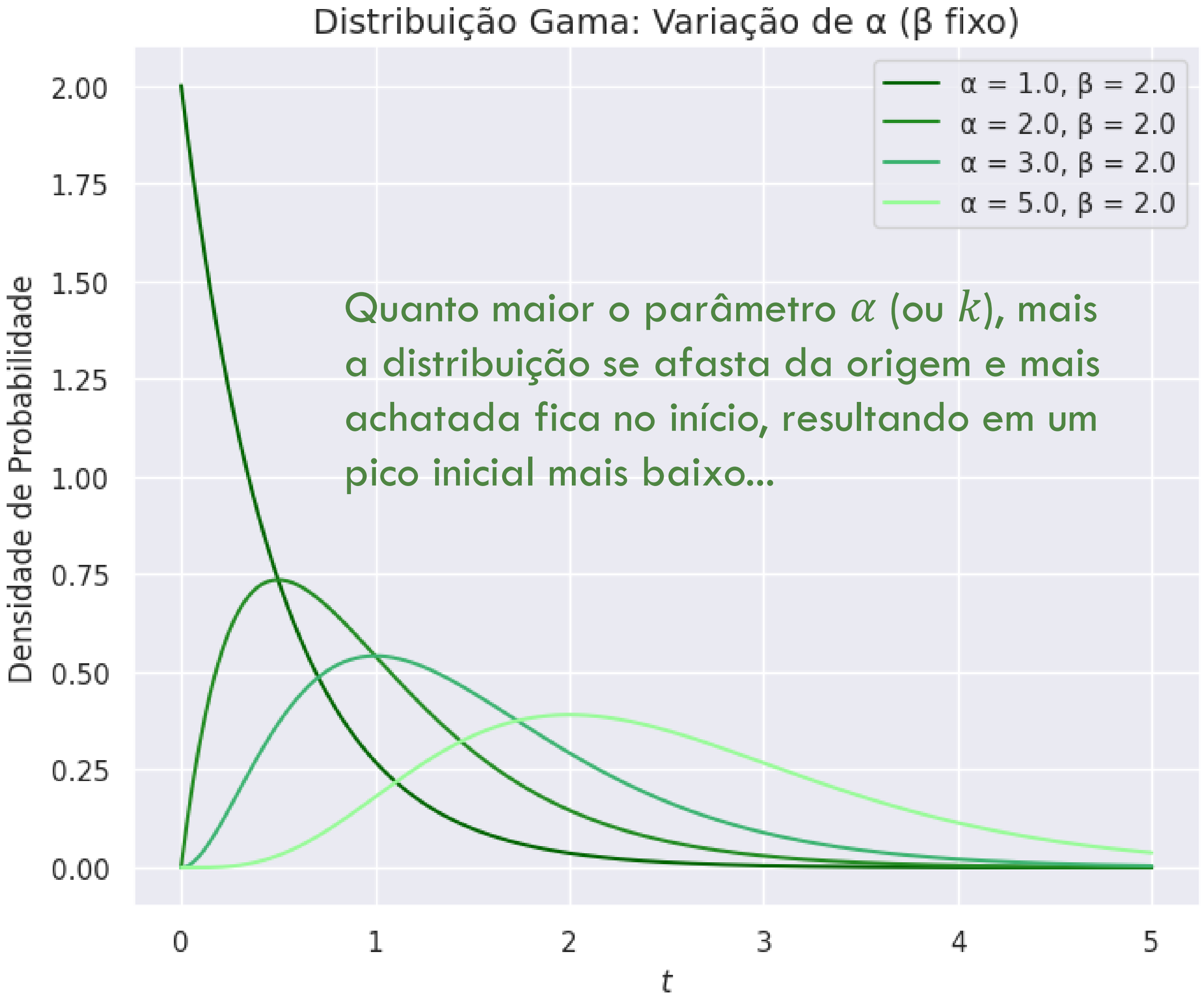
Ela vira uma **distribuição Gama (Gamma)**! Mais precisamente:

- α **(alfa):** é o número de eventos esperados (no seu caso, $\alpha = 5$ carros).
- β **(beta):** é a taxa. Como é 1 carro a cada 2 minutos, a taxa $\beta = \lambda$ é 0,5 carro por minuto.

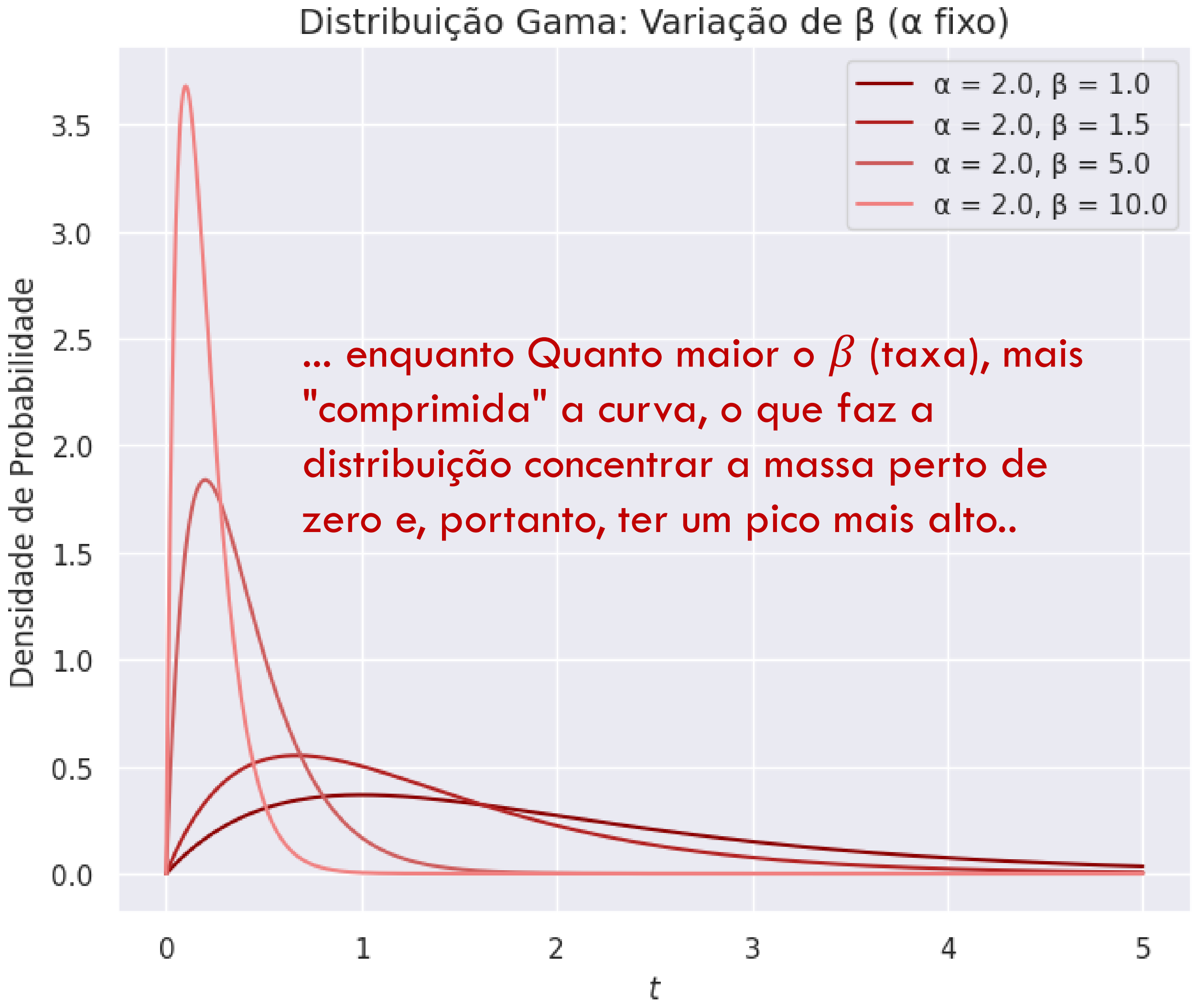
O parâmetro α é conhecido como parâmetro de forma (“formato” geral), e o parâmetro β é chamado de parâmetro de escala (controla o "ritmo de decaimento“)

$$p(x;\alpha,\beta) = \frac{\beta^{\alpha} x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)}$$

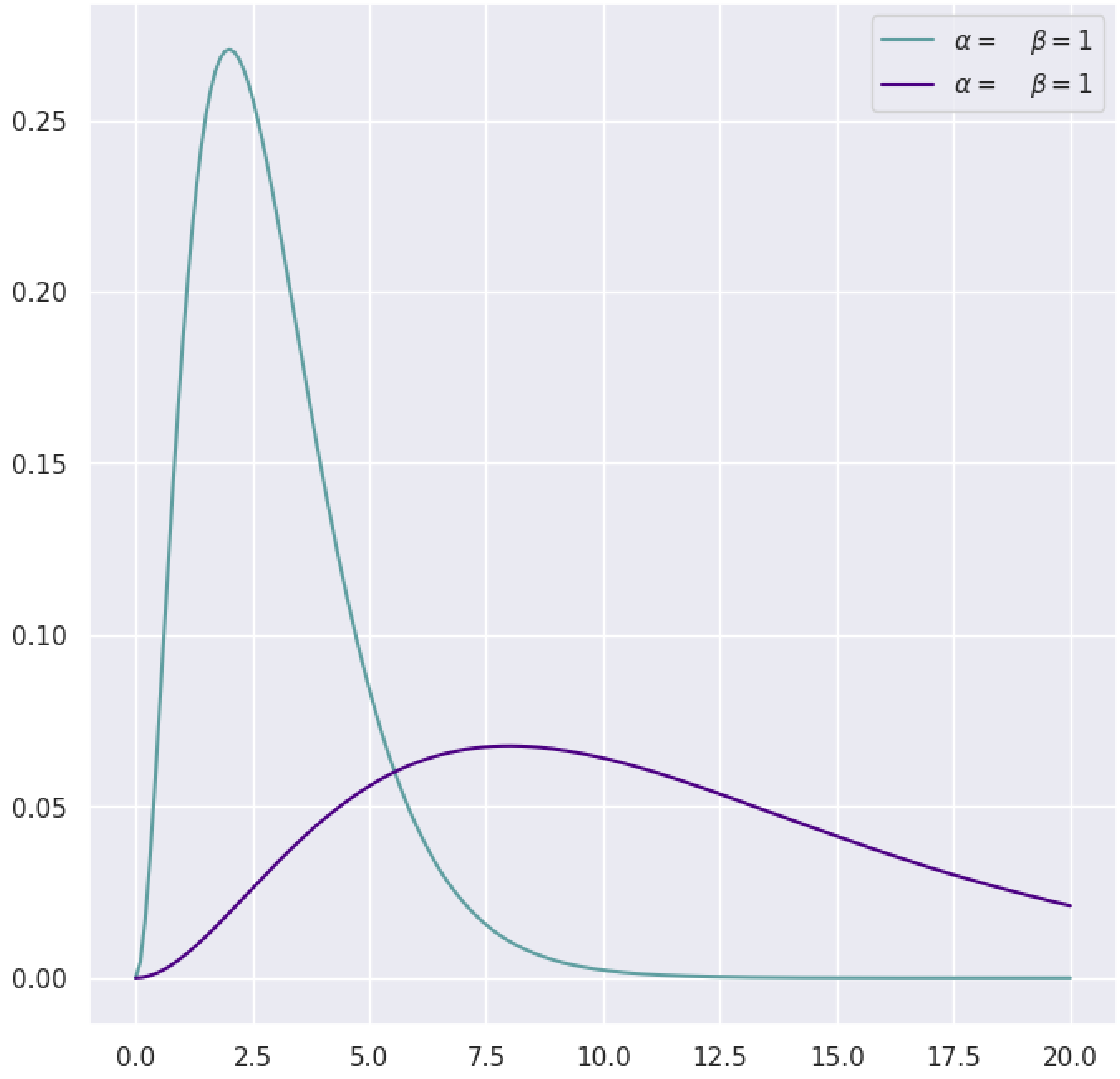
PECE Programa de Educação Continuada
Escola Politécnica da USP



Quanto maior o parâmetro α (ou k), mais a distribuição se afasta da origem e mais achatada fica no início, resultando em um pico inicial mais baixo...



... enquanto Quanto maior o β (taxa), mais "comprimida" a curva, o que faz a distribuição concentrar a massa perto de zero e, portanto, ter um pico mais alto..



Qual α vale 3 e qual vale 5?

EXEMPLO

Carros chegam a um cruzamento a uma taxa, em média, de um a cada dois minutos. Seja X o tempo de espera até que cinco carros cheguem. Modele o problema com a função Gama.

$\boldsymbol{\alpha = 5}$ (Número de carros desejados) e

$\boldsymbol{\beta = 0,5}$ ($\lambda = 0{,}5$ carros/minuto)

Pede-se,

1. Qual é a probabilidade de espera de até 8 minutos para o evento?
2. Plote a distribuição Gama do problema.

$\alpha = 5; \beta = 0{,}5$
$x = 8 \text{ min}$

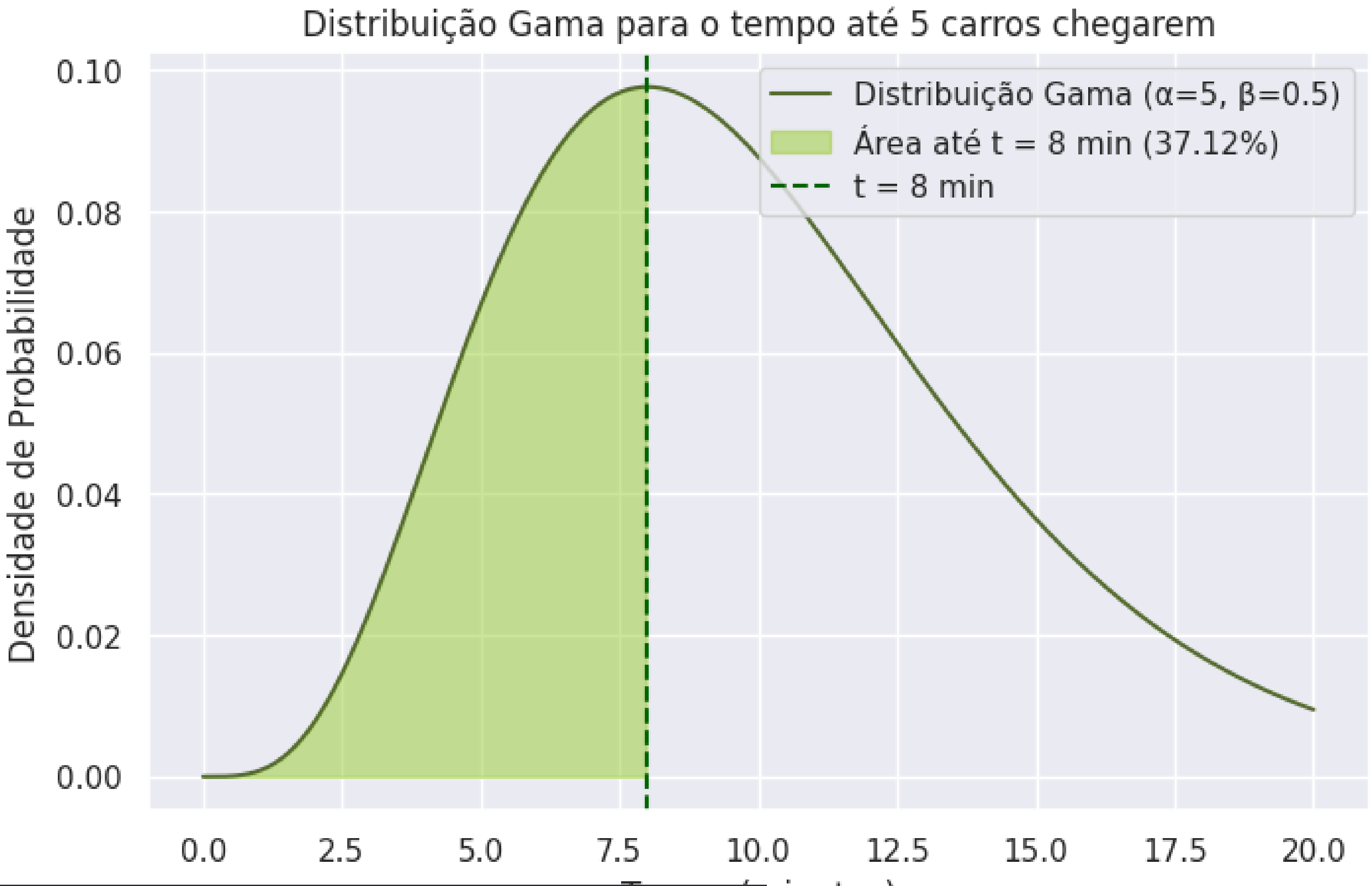
$$P(t < x; \alpha = 5, \beta = 0{,}5) = \frac{\gamma(\alpha, \beta x)}{\Gamma(\alpha)} = \frac{\int_0^{\beta x} t^{\alpha-1} e^{-t} dt}{(\alpha - 1)!}$$

PECE Programa de Educação Continuada
Escola Politécnica da USP

$$P(t < 8) = \frac{\int_0^{4} t^{5-1} e^{-8} dt}{24}$$

Esse valor não dá pra integrar na mão facilmente

A probabilidade de esperar até 8 minutos para que 5 carros cheguem é aproximadamente 37,12%.




```
prob = gamma.cdf(x, a=alpha, scale=1/beta)
```

DISTRIBUIÇÃO DE GAMA VS EXPONENCIAL VS POISSON

As falhas em CPU's de computadores usualmente são modeladas por processos de Poisson. Isso porque, tipicamente, as falhas não são causadas por desgaste, mas por eventos externos ao sistema. Assuma que as unidades que falham sejam imediatamente reparadas e que a taxa média seja 0,0001 falhas por hora .

Determine as probabilidades de que:

1. o tempo entre falhas sucessivas exceda 10.000 horas;
2. o tempo até a quarta falha exceda 40.000 horas;
3. ocorram mais que 3 falhas em 20.000 horas.

POR QUE A DISTRIBUIÇÃO GAMA É IMPORTANTE NO APRENDIZADO DE MÁQUINA?

Capacidade de modelar dados com valores positivos: A Distribuição Gama é adequada para representar variáveis que não podem ser negativas, tornando-a útil para modelar grandezas como tempos de resposta ou contagens em tarefas de aprendizado de máquina.

Lidando com variabilidade e assimetria: Com seus parâmetros de forma e escala, a Distribuição Gama pode capturar efetivamente níveis variados de assimetria e forma em distribuições de dados, permitindo a modelagem precisa de diversos conjuntos de dados.

Importância na inferência bayesiana: Servindo como um prior conjugado para certas funções de verossimilhança, a Distribuição Gama simplifica a estimativa de parâmetros bayesianos e a modelagem probabilística, aumentando a eficiência e a precisão dos algoritmos bayesianos de aprendizado de máquina.

Modeling and Characterizing Social Media Topics Using the Gamma Distribution

Connie Yee, Nathan Keane, Liang Zhou
Text Analytics and Machine Learning
Thomson Reuters
New York, NY 10036, USA
{connie.yee,nathan.keane,l.zhou}@thomsonreuters.com

Abstract

We present a novel technique to identify emerging or important topics mentioned on social media. A sudden increase in related posts can indicate an occurrence of an external event. Assuming that the sequence of posts is a homogeneous Poisson process, this sudden change can be modeled using the Gamma distribution. Our Gamma curve fitter is used to return a set of emerging topics. We demonstrate our algorithm on Twitter data and evaluate empirically using the Reuters News Archive and manual inspection. Our experimental results show that our algorithm provides a good picture of the emerging topics discussed on Twitter.

We are interested in discovering events related to both content from news outlets and content that originates on social media. An event occurrence can be detected by the volume and sudden change in volume of posts. After examining the distributions of the volumes of topics in Twitter, we observe two main categories of topics:

- Long-lasting topics that Twitter users frequently discuss in their daily lives, such as the foods they ate and the activities they are currently doing
- Emerging topics[1], or topics of importance to the general public, such as sporting events and natural disasters

Date	Topic	Top Words
2014-06-14	Stanley Cup	game kings cup win hockey
2014-06-15	Wonder Goal	goal messi argentina france #worldcup
2014-06-19	Biting	england rooney suarez goal uruguay
2014-06-27	Player Contract	money pay million shaw united

Table 1: Selected topics.

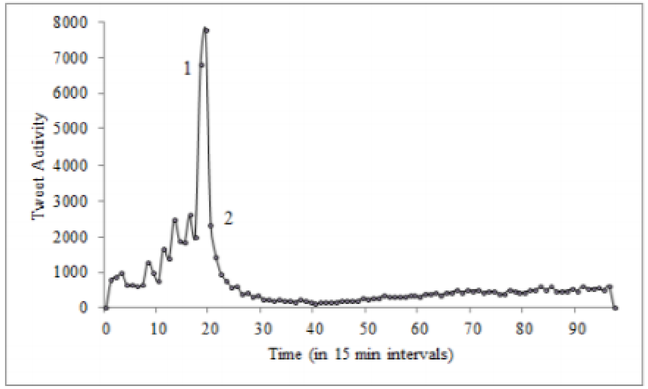
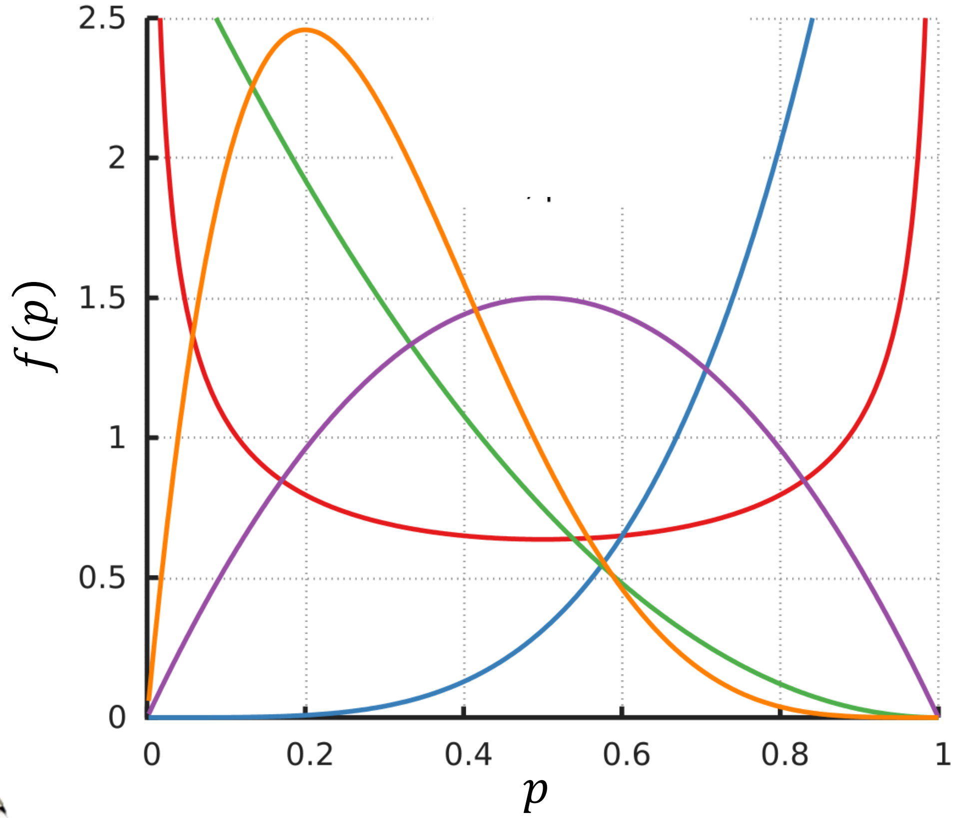


Figure 1: Frequency distribution of the “Stanley Cup” topic.



FUNÇÃO BETA

$$f(p;\alpha,\beta)$$

DISTRIBUIÇÃO BETA

Beta é uma distribuição de probabilidade contínua definida no intervalo $[0,1]$ parametrizada por dois parâmetros positivos de forma, α e β . A PDF da distribuição beta é:

$$f(p;\alpha,\beta) = \frac{1}{B(\alpha,\beta)} p^{\alpha-1}(1-p)^{\beta-1}$$

Onde o coeficiente

$$B(\alpha,\beta) = \int_0^1 p^{\alpha-1}(1-p)^{\beta-1}dp = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

$\mathit{\Gamma}(n) = (n-1)!$ para números naturais. O valor esperado e variância são dados, respectivamente, por, $\mu = \frac{\alpha}{\alpha+\beta}$ e $\sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$.

FUNÇÃO DE DISTRIBUIÇÃO ACUMULADA DE BETA

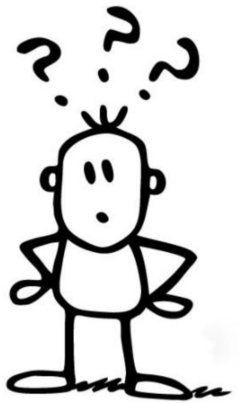
$$P(p < x; \alpha, \beta) = \frac{B(x, \alpha, \beta)}{B(\alpha, \beta)}$$

$B(x, \alpha, \beta) = \int_0^x p^{\alpha-1}(1-p)^{\beta-1} dp$ é a função gama incompleta inferior e x é o valor do limite superior que queremos acumular.

NUMERADOR DA FUNÇÃO BETA

Vamos olhar apenas para o numerador $p^{\alpha-1}(1-p)^{\beta-1}$.

Os termos no numerador – p à potência de algo multiplicado por $(1-p)$ à potência de algo - parecem familiares??? Já vimos isso antes????



Sim. Na distribuição binomial.

Binomial (PMF)	**Beta (PDF)**
$P(x;n,p) = \binom{n}{x} p^x (1-p)^{n-x}$	$f(p;\alpha,\beta) = \frac{1}{B(\alpha,\beta)} p^{\alpha-1}(1-p)^{\beta-1}$
Probabilidade entra como um **parâmetro** p	Probabilidade como uma **variável randômica** (p)

PARA QUE SERVE?

Comumente usada para resolver problemas que buscam uma probabilidade, como qual é a:

"... probabilidade de ter menos de 10% de defeitos?"

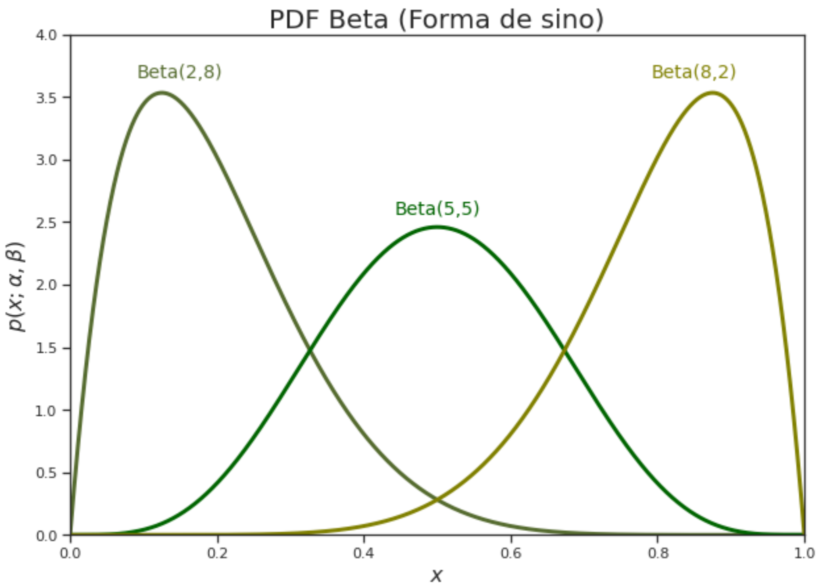
"... probabilidade de obter menos de 20% de caras em um lançamento de moeda?"

"... probabilidade de um anúncio ter uma taxa de clique (CTR) entre 30% e 40%?"

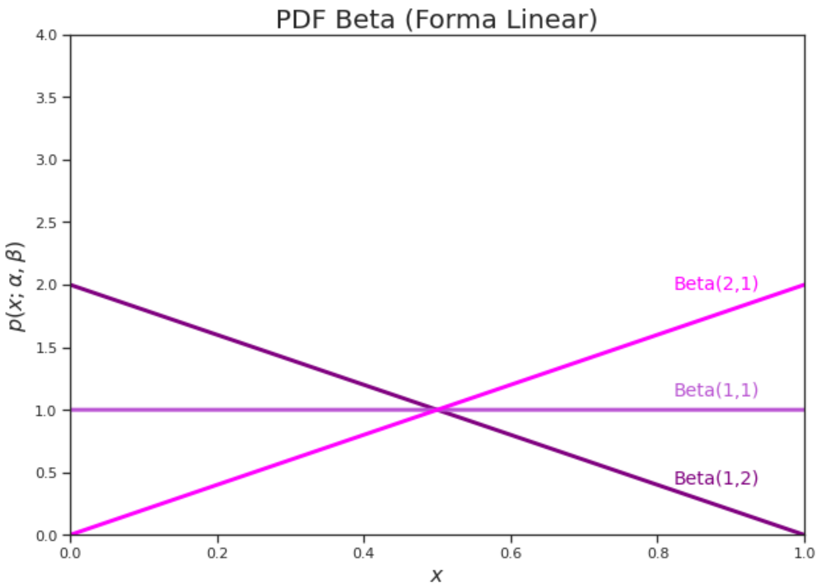
...

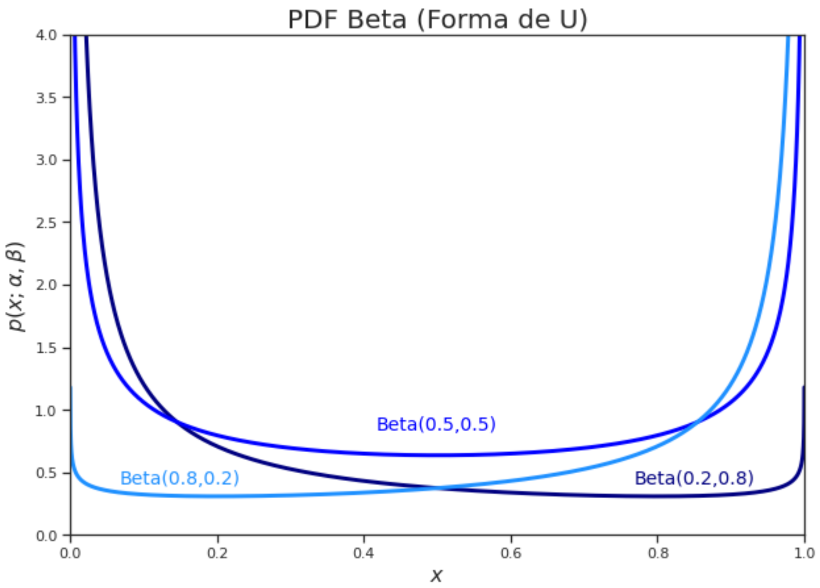
INTERPRETAÇÃO PARA α E β

Você pode pensar em $\alpha - 1$ como o número de sucessos e $\beta - 1$ como o número de falhas, assim como os termos x e $n - x$ no modelo binomial.

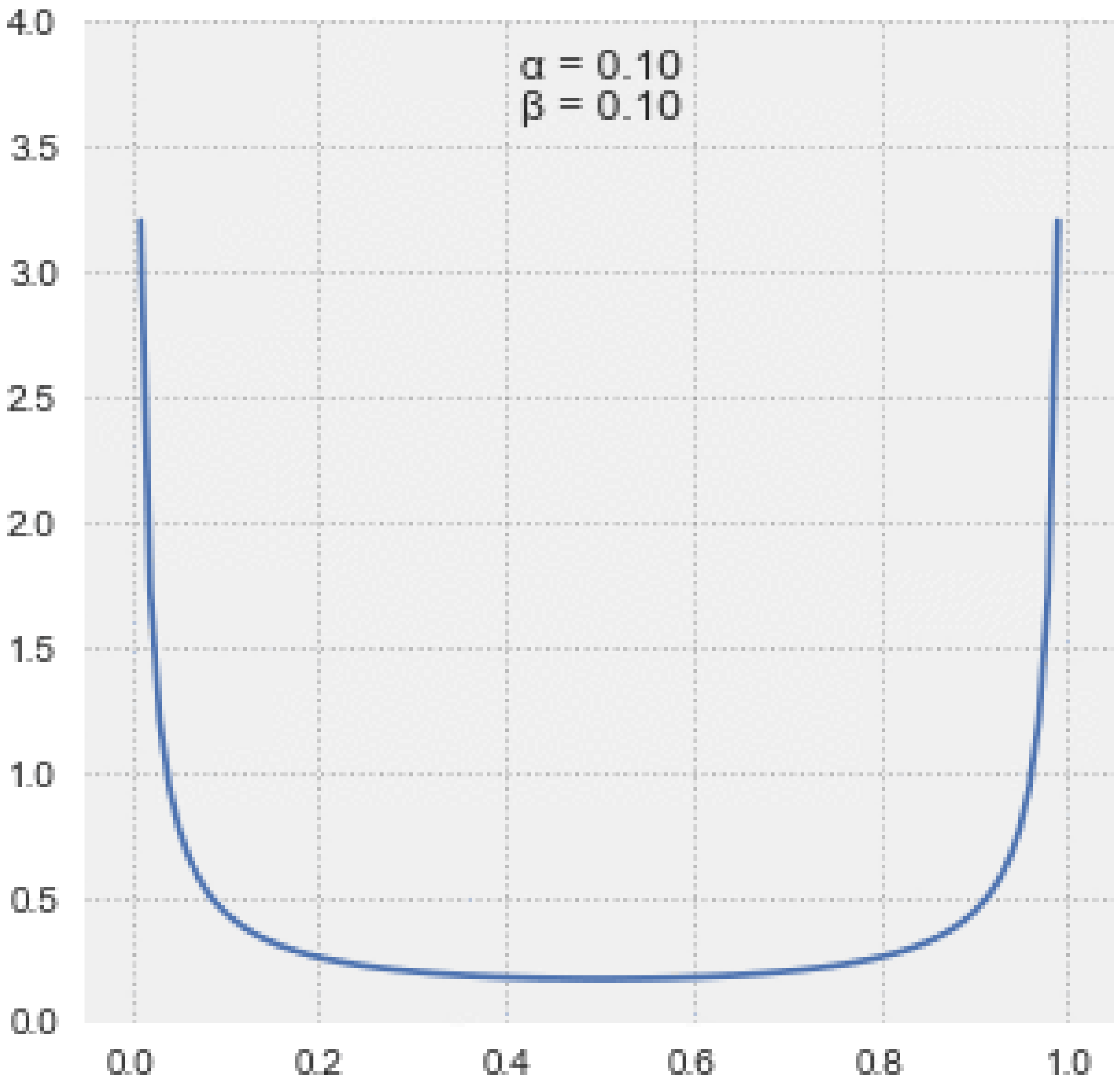


O PDF de uma distribuição beta é aproximadamente normal se $\alpha + \beta$ for grande o suficiente e α e β forem aproximadamente iguais.





$\alpha < 1, \beta < 1$


α = 0.10
β = 0.10
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0.0
0.2
0.4
0.6
0.8
1.0

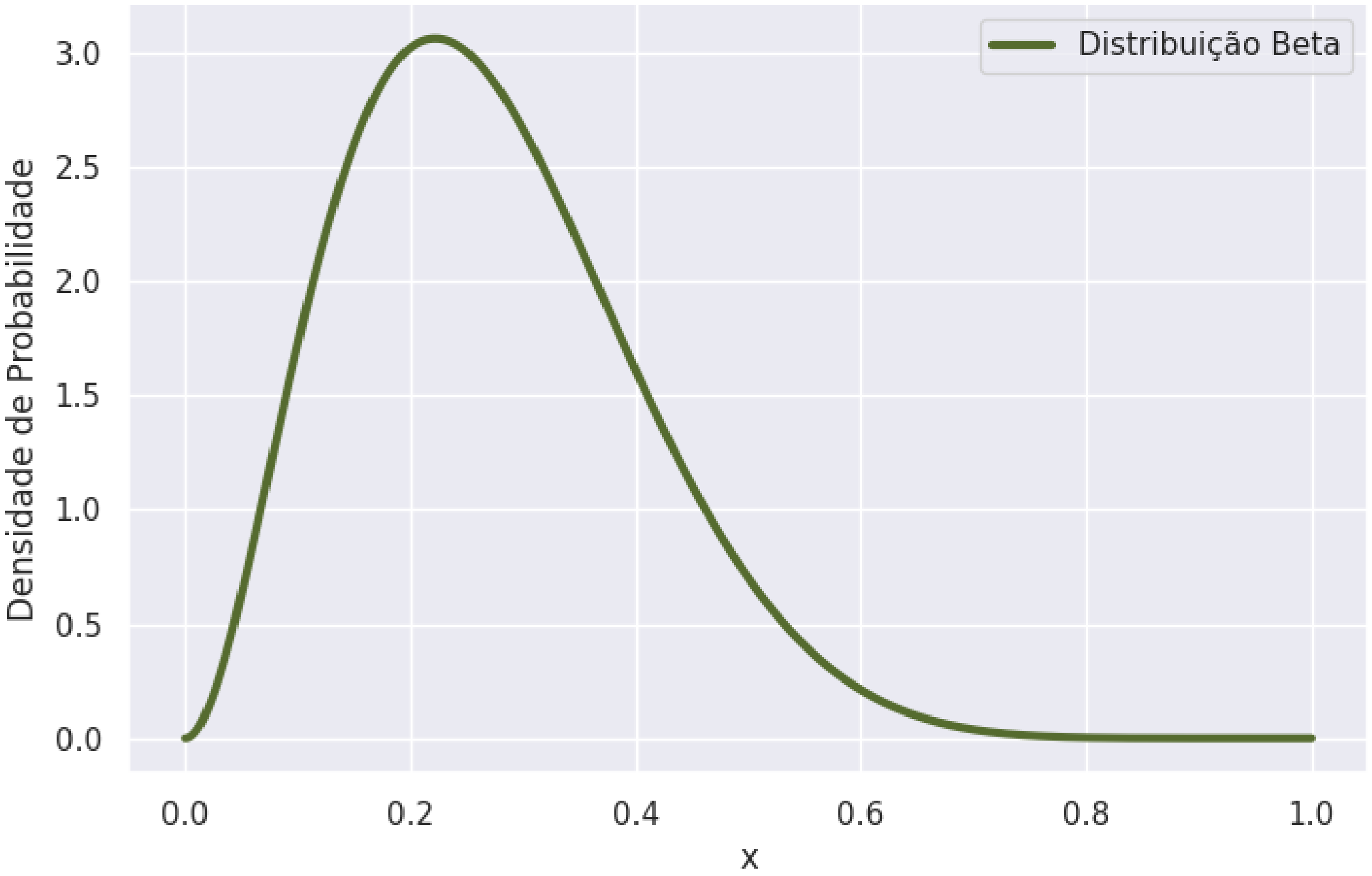
EXEMPLO

Digamos que a probabilidade de alguém entrar no seu site segue uma distribuição Beta com $\alpha = 3$ e $\beta = 8$. Qual é a probabilidade de sua taxa de sucesso ser maior que 50%?

$$P(p < x; \alpha, \beta) = \frac{B(x,\alpha,\beta)}{B(\alpha,\beta)}$$

$$B(\alpha, \beta) = \int_0^1 p^{\alpha-1}(1-p)^{\beta-1}dp = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

$$B(x, \alpha, \beta) = \int_0^x p^{\alpha-1}(1-p)^{\beta-1}dp$$



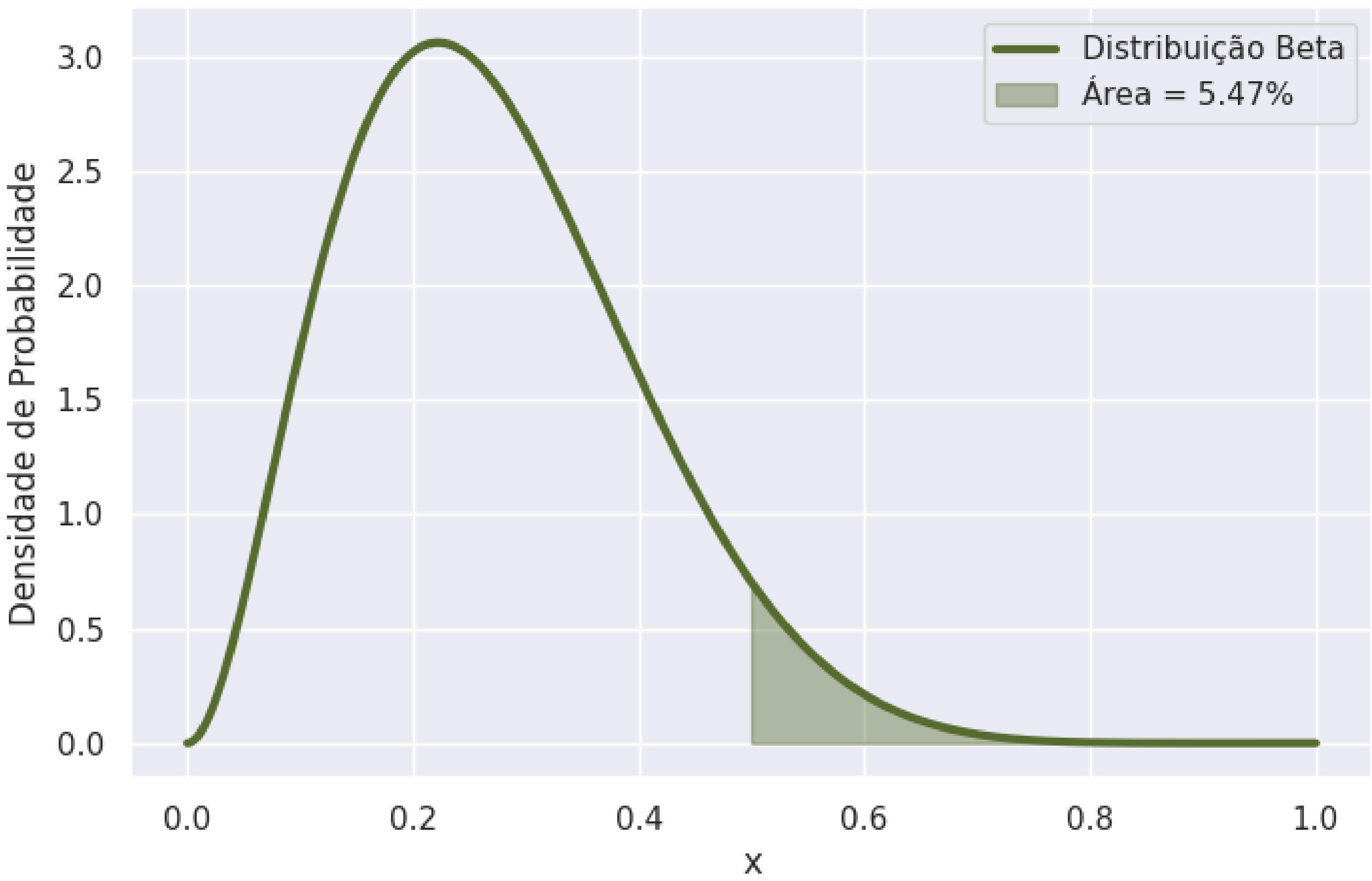
EXEMPLO

Digamos que a probabilidade de alguém entrar no seu site segue uma distribuição Beta com $\alpha = 3$ e $\beta = 8$. Qual é a probabilidade de sua taxa de sucesso ser maior que 50%?

$$P(p < x; \alpha, \beta) = \frac{B(x,\alpha,\beta)}{B(\alpha,\beta)}$$

$$B(\alpha, \beta) = \frac{\Gamma(3)\Gamma(8)}{\Gamma(11)} = \frac{2!7!}{10!} = 0{,}002778$$

$$B(x, \alpha, \beta) = \int_0^{0,5} p^2(1-p)^7 dp = 0{,}002625$$

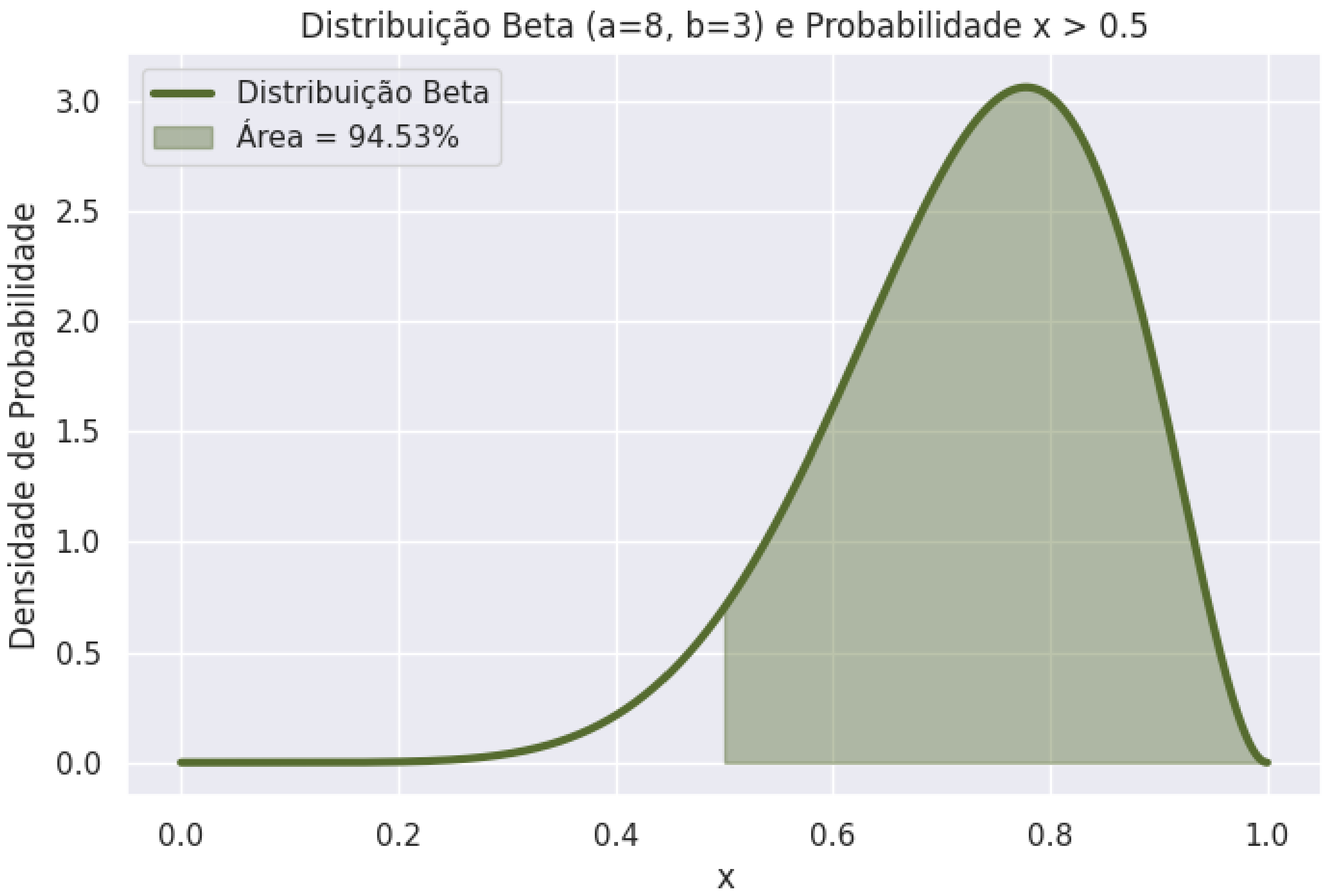


$$P(p > 0{,}5; \alpha = 3, \beta = 8) = 1 - \frac{0{,}002625}{0{,}002778} \approx 0{,}05468$$

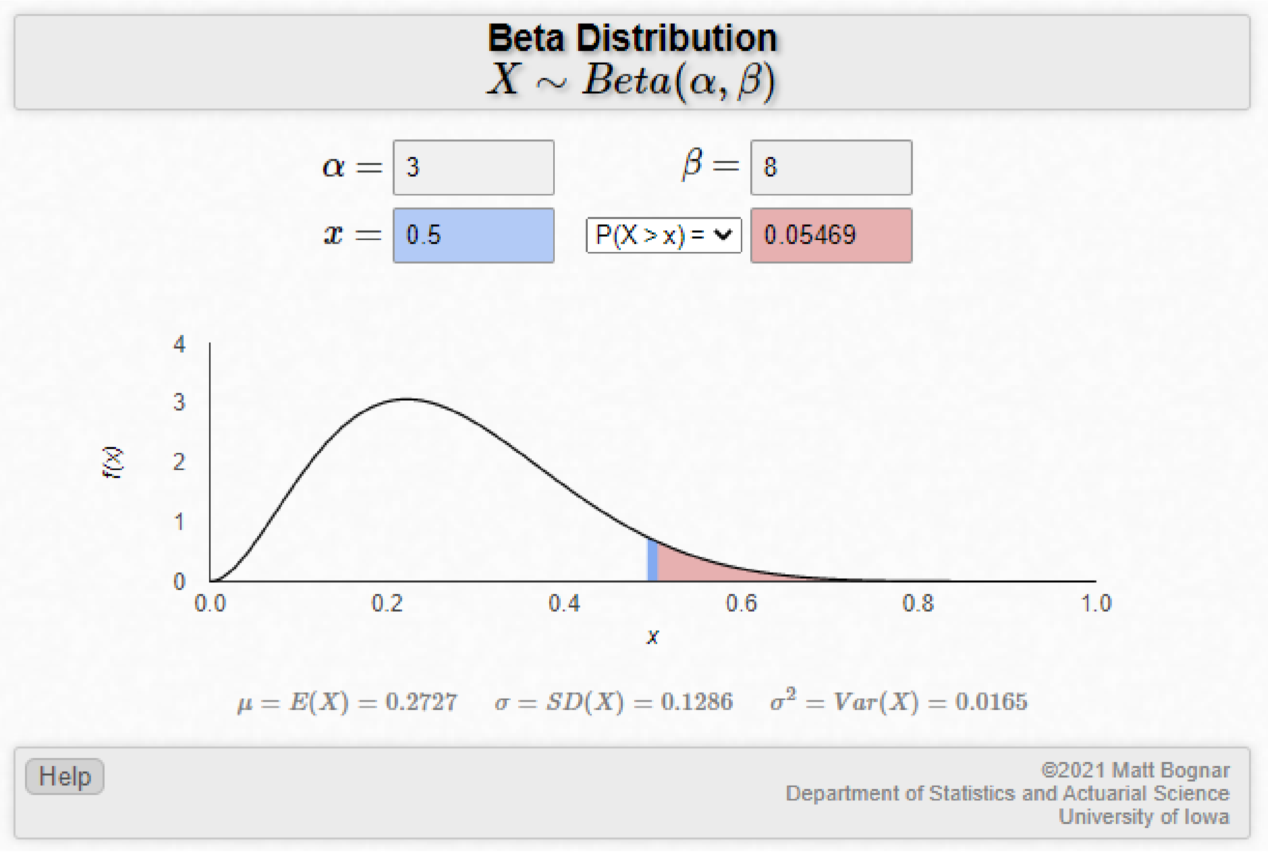
EXEMPLO

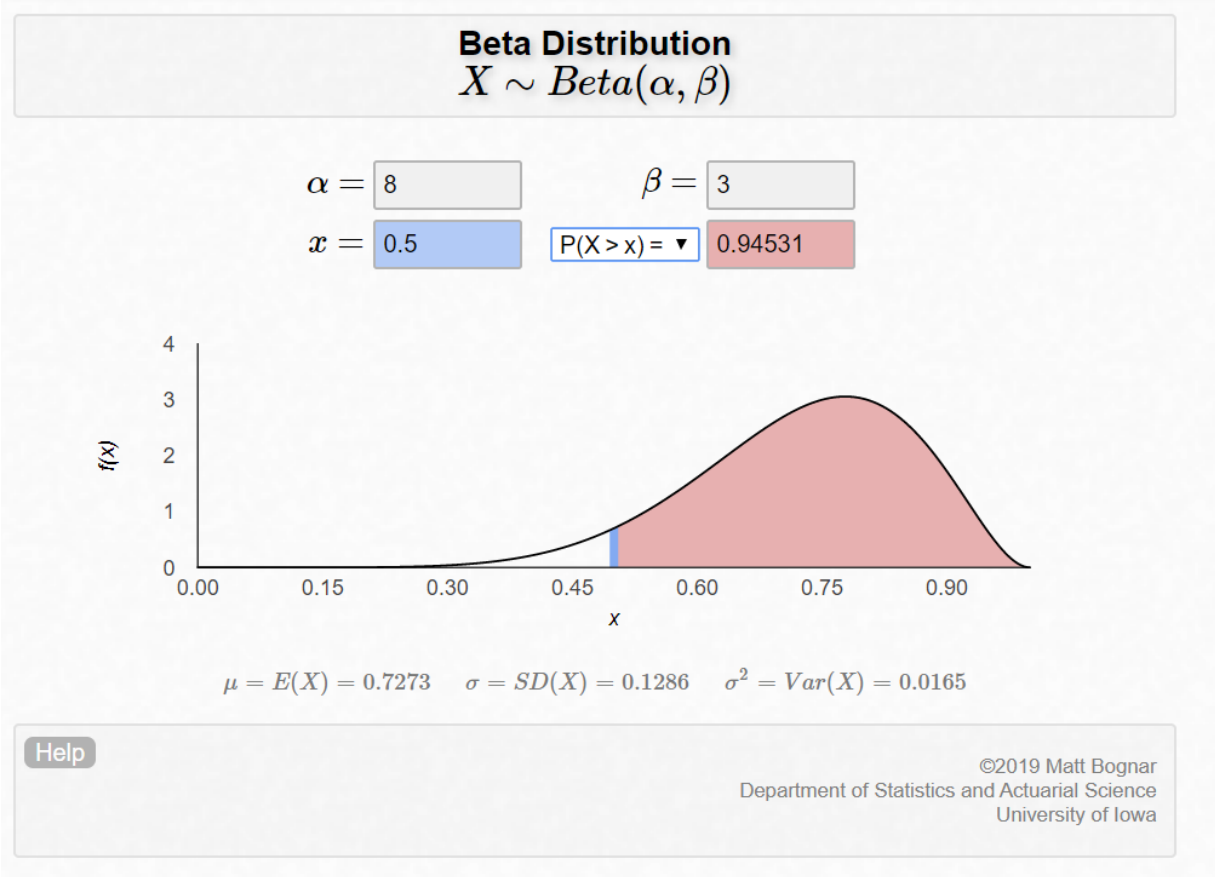
E se for ao contrário, $\alpha = 8$ e $\beta = 3$?

Invertendo os parâmetros, , a probabilidade de taxa de sucesso maior que 50% via fórmula é: $94{,}53\%$



EXEMPLO





https://homepage.divms.uiowa.edu/~mbognar/applets/beta.html

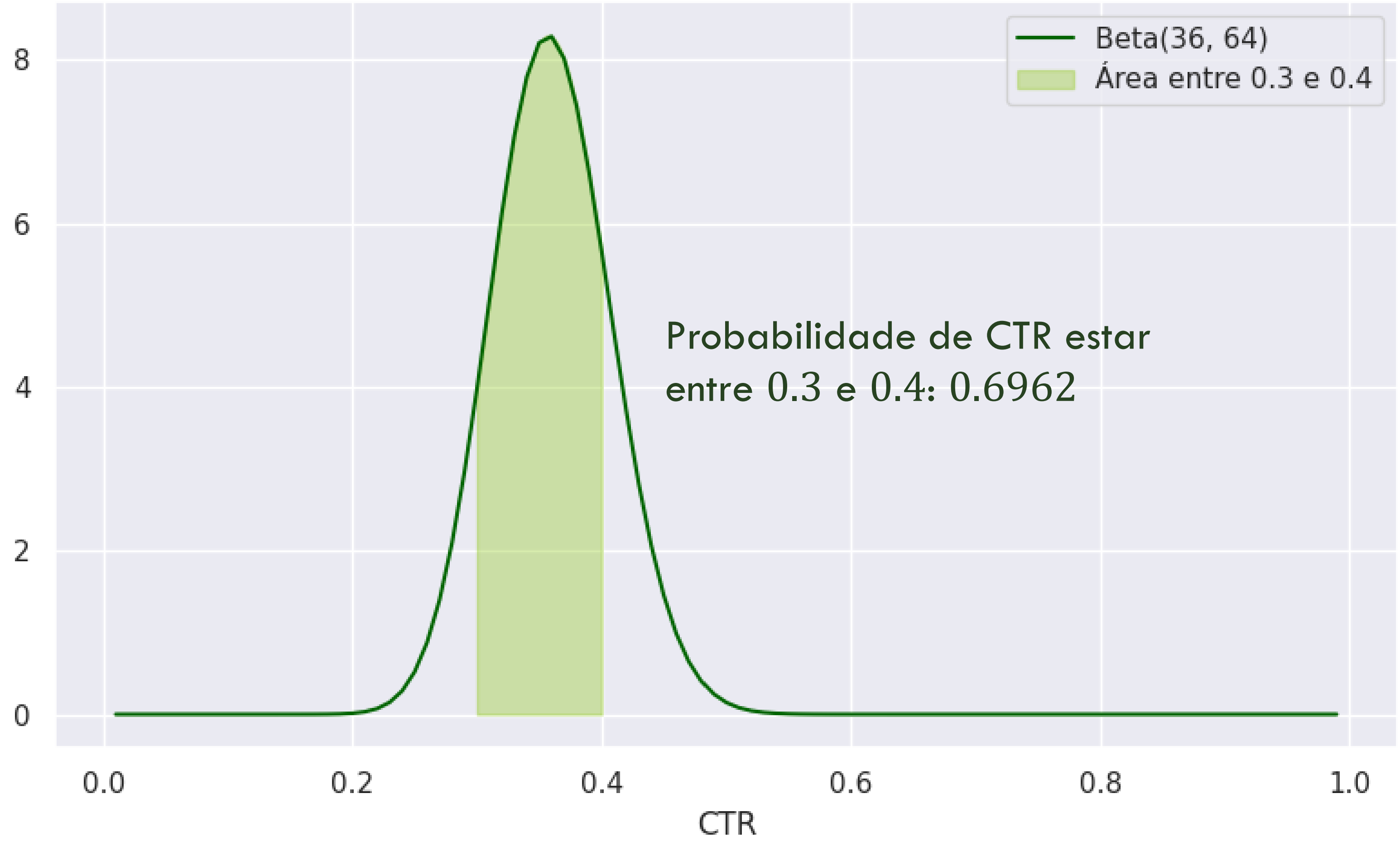
QUAL A PROBABILIDADE DE UM ANÚNCIO (COM CTR REAL DE 34,3%) TER UMA CTR ENTRE 30% E 40%?

PECE Programa de Educação Continuada
Escola Politécnica da USP

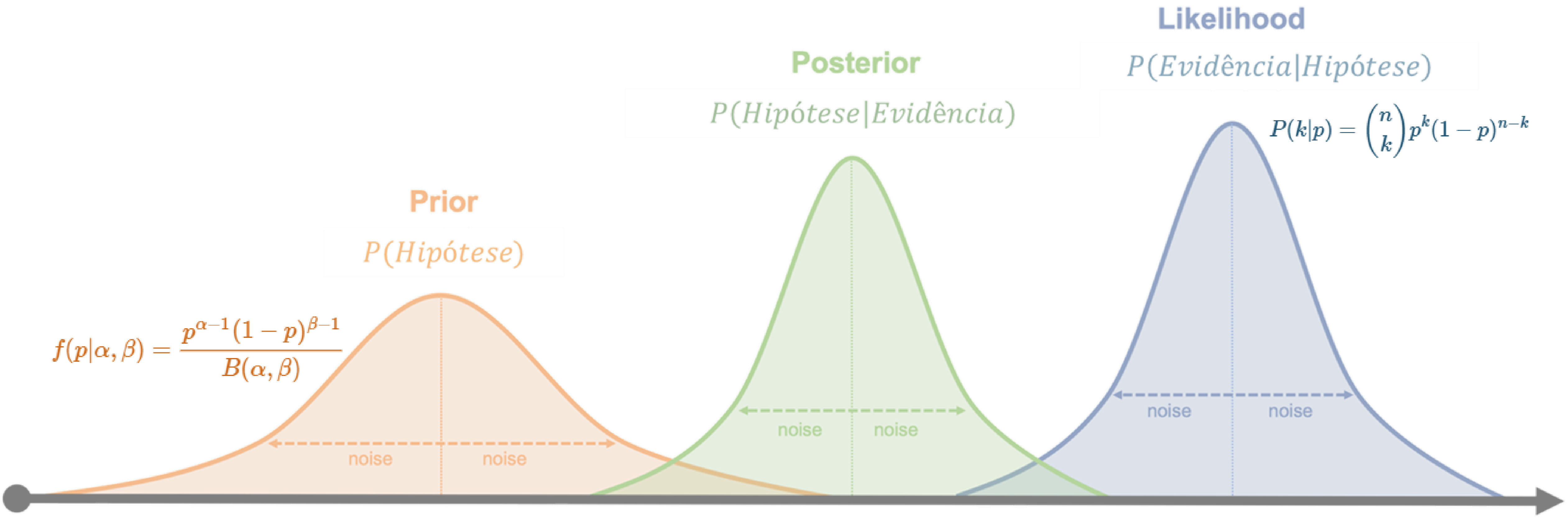
Supondo que você trabalhe em uma empresa de publicidade, e a CTR (taxa de cliques, do inglês *click-through rate*) do anúncio de uma campanha sua no mês anterior foi de 34,3%. Como o mês atual ainda não terminou, precisamos atualizar nossa **probabilidade inicial (prior)** à medida que novos dados vão chegando.

Usando CTR = 34,3% como a CTR real, vamos rodar uma simulação. Como estamos lidando com CTR, podemos representar isso com uma distribuição binomial: **1 para clique, 0 para não clique**.

Sempre que ocorre um clique, aumentamos **alpha** em 1; se não ocorre, aumentamos **beta** em 1. Isso permite atualizar continuamente a probabilidade prévia à medida que novos dados são coletados.


Beta(36, 64)
Área entre 0.3 e 0.4
Probabilidade de CTR estar
entre 0.3 e 0.4: 0.6962
0
2
4
6
8
0.0
0.2
0.4
0.6
0.8
1.0
CTR

$$P(p|k) \propto P(k|p)\,f(p|\alpha,\beta)$$



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

$$P(p|k) \propto P(k|p) f(p|\alpha, \beta)$$

$$P(p|k) \propto \binom{n}{k} p^k (1-p)^{n-k} \cdot p^{\alpha-1} (1-p)^{\beta-1}$$

$$P(p|k) \propto p^{\alpha+k-1} (1-p)^{\beta+n-k-1}$$

$$P(p|k) = \frac{p^{\alpha+k-1} (1-p)^{\beta+n-k-1}}{B(\alpha+k, \beta+n-k)}$$

Esta é uma distribuição Beta com parâmetros $\alpha + k$ e $\beta + n - k$.

ESTE É O SEU EXERCÍCIO DE DISTRIBUIÇÃO BETA E BINOMIAL

Repita o exercício anterior, construindo uma distribuição Beta a priori com base em crenças informadas.

Vamos imaginar que você foi pesquisar bastante sobre o comportamento de cliques em campanhas publicitárias e montou uma crença inicial. A partir disso, você achou que seria razoável assumir que cerca de 55% das pessoas clicariam. Nesse caso, seria razoável sugerir que:

$$\mathbb{E}(p) = \frac{\alpha}{\alpha + \beta} = 0{,}55$$

Agora, temos apenas um único parâmetro a definir (por exemplo, β), já que α está definido em função de β.

Outro aspecto da distribuição que podemos considerar é a probabilidade de p estar abaixo de um determinado valor. Para fins ilustrativos, imagine que você percebeu que seria improvável que a taxa real de cliques fosse inferior a 40%. Mais precisamente, essa chance seria de 20%. Isto é:

$$P(p < 0.4) = 0.2$$

Como a função de distribuição acumulada (CDF) da distribuição Beta é parametrizada por α e β, você pode usar as informações acima e achar os hiperparâmetros de sua distribuição priori.

SUA TAREFA (OPCIONAL)

Completar o notebook IAD_001_2025_L09.ipynb

As questões sem resposta estão em verde.

NEVER GIVE UP

ACABOU...

Exercício