Análise Estatística de dados

Inteligência Artificial

PECE Programa de Educação Continuada
Escola Politécnica da USP

USP

AULA 09 – MODELOS DE PROBABILIDADE CONTÍNUOS

Arturo Forner-Cordero
Larissa Driemeier

PROGRAMA DO CURSO

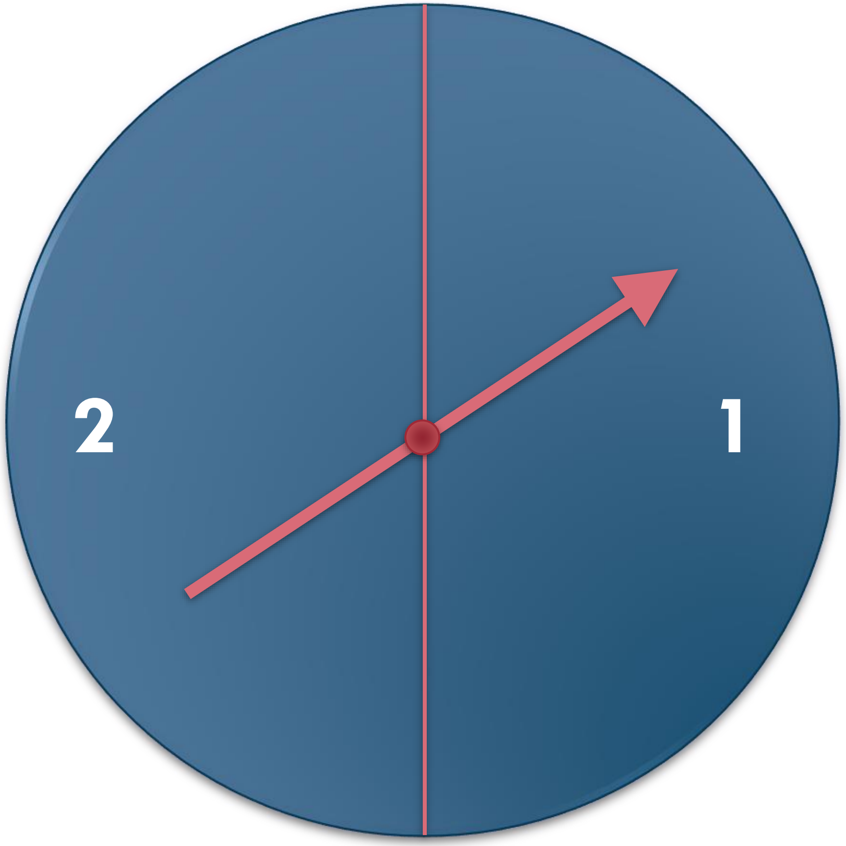
Aula	Data	Conteúdo da Aula
01	27/02	Aula Inaugural
02	05/03	Introdução ao Curso. Noções de Álgebra Linear , Geometria Analítica Parte I
03	12/03	Noções de Álgebra Linear , Geometria Analítica Parte II
04	19/03	Decomposição de valor singular (SVD)
05	26/03	Otimização: derivadas, derivadas parciais (operadores gradiente, Jacobiano, Hessiano e Laplaciano), algoritmos de gradiente
06	02/04	Variáveis independentes e não independentes. Estatística Descritiva e Indutiva. Definições de medidas de dispersão e tendência central.
07	09/04	Probabilidade e Teorema de Bayes
08	16/04	Modelos de probabilidade discretos
09	23/04	Modelos de probabilidade contínuos
10	30/04	Modelo de Markov. Modelo de Markov oculto.

IMAGINE...

Um jogo de azar é realizado da seguinte forma: fixa-se um ponteiro no centro de um círculo e divide-se o círculo em duas partes iguais, **1** e **2**.

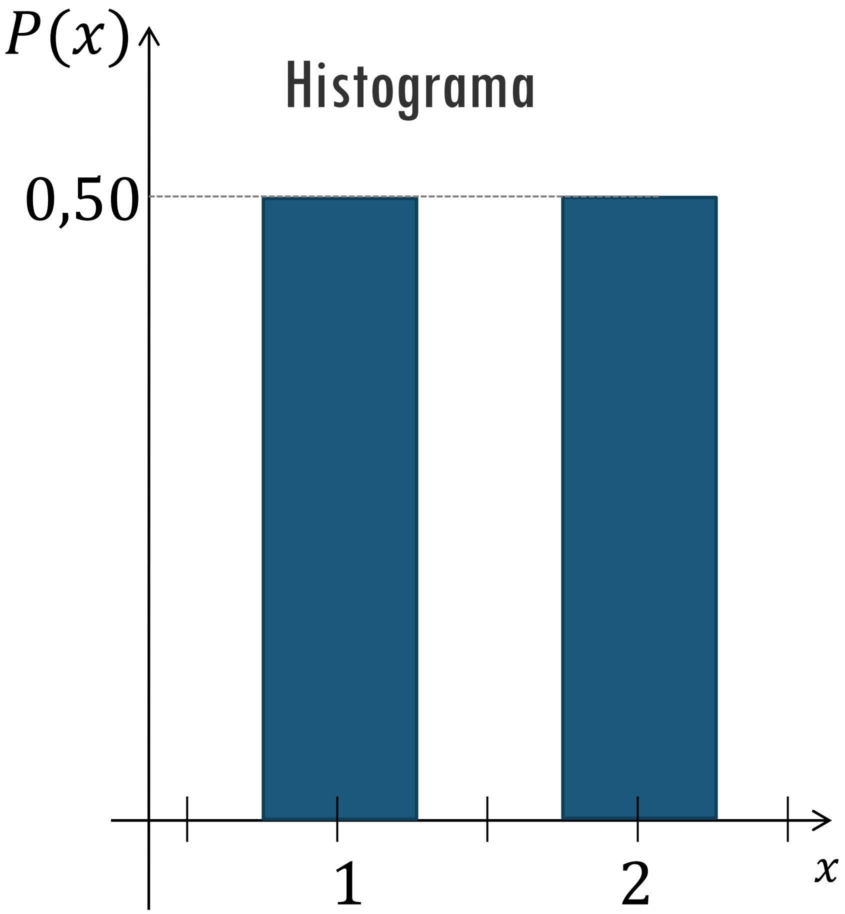
Gira-se o ponteiro e anota-se o número que o ponteiro, quando parado, está sinalizando.

- Construir a distribuição de probabilidades para o número obtido neste experimento.

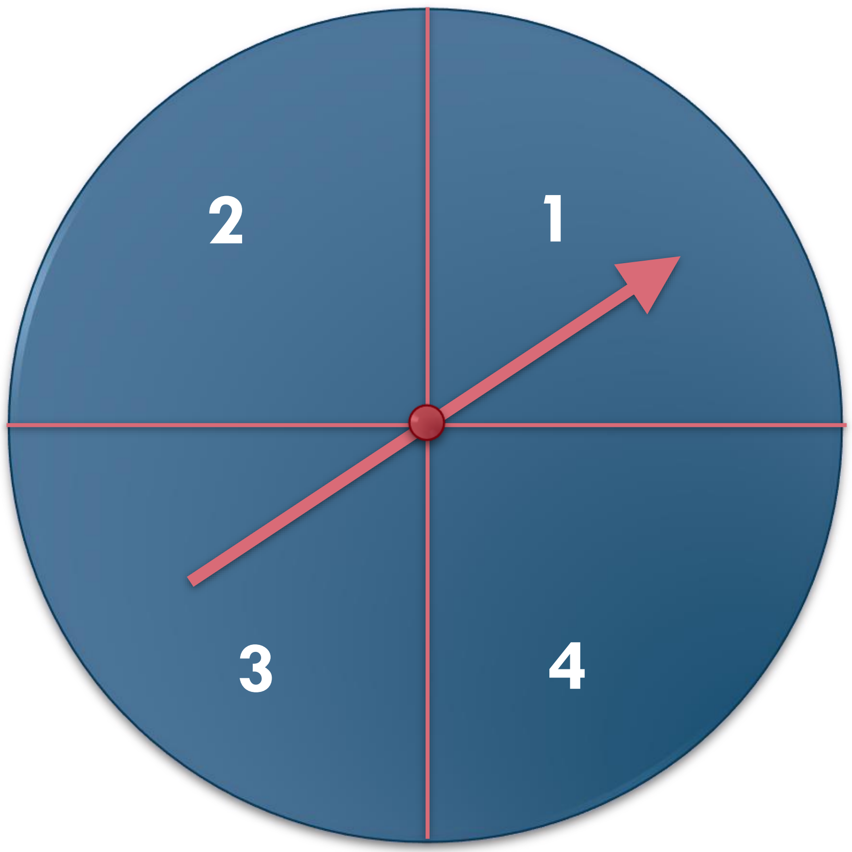


DISTRIBUIÇÃO DE PROBABILIDADES

x	P(X=x)
1	0,5
2	0,5
Total:	1,0



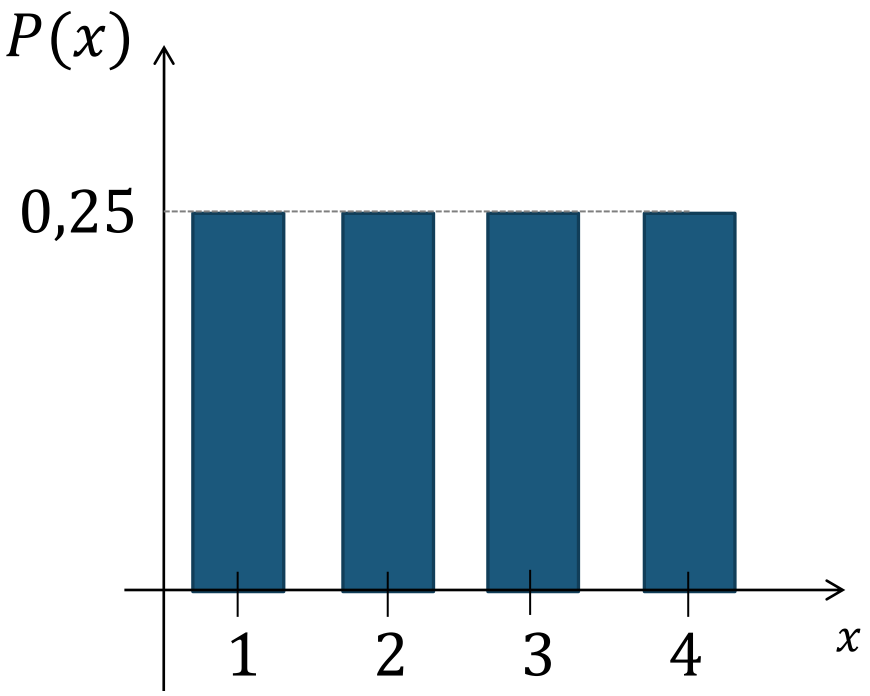
Ainda com respeito ao jogo anterior, o círculo é dividido agora em quatro partes iguais. Construir a distribuição de probabilidades para o **número que o ponteiro, quando parado, está sinalizando.**



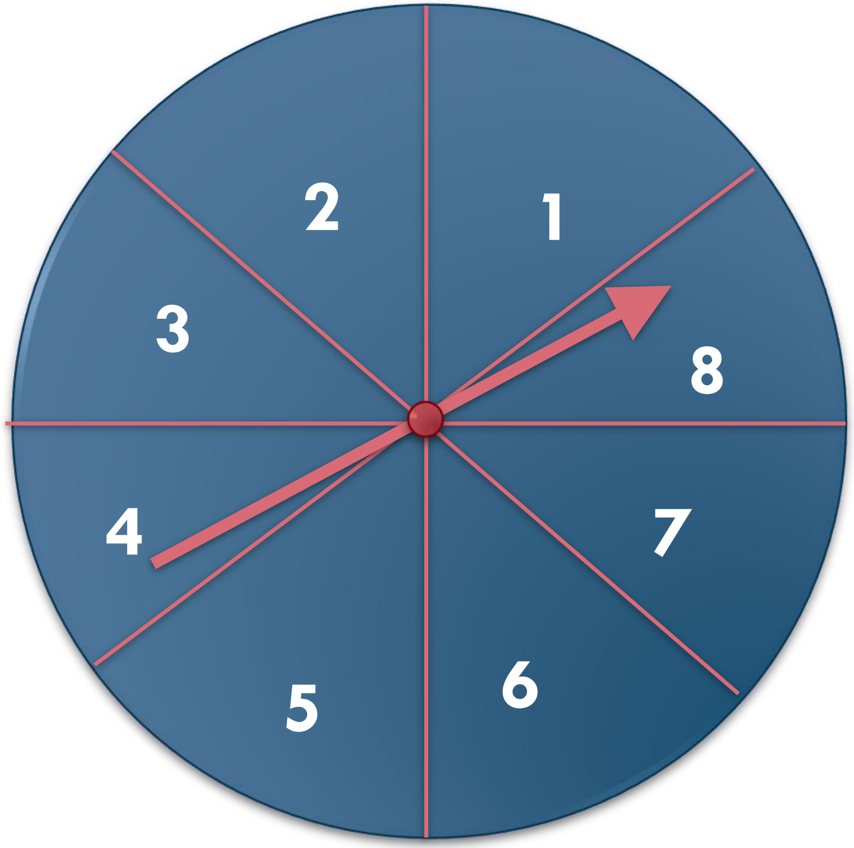
DISTRIBUIÇÃO DE PROBABILIDADES

x	P(X=x)
1	0,25
2	0,25
3	0,25
4	0,25
Total:	1,0

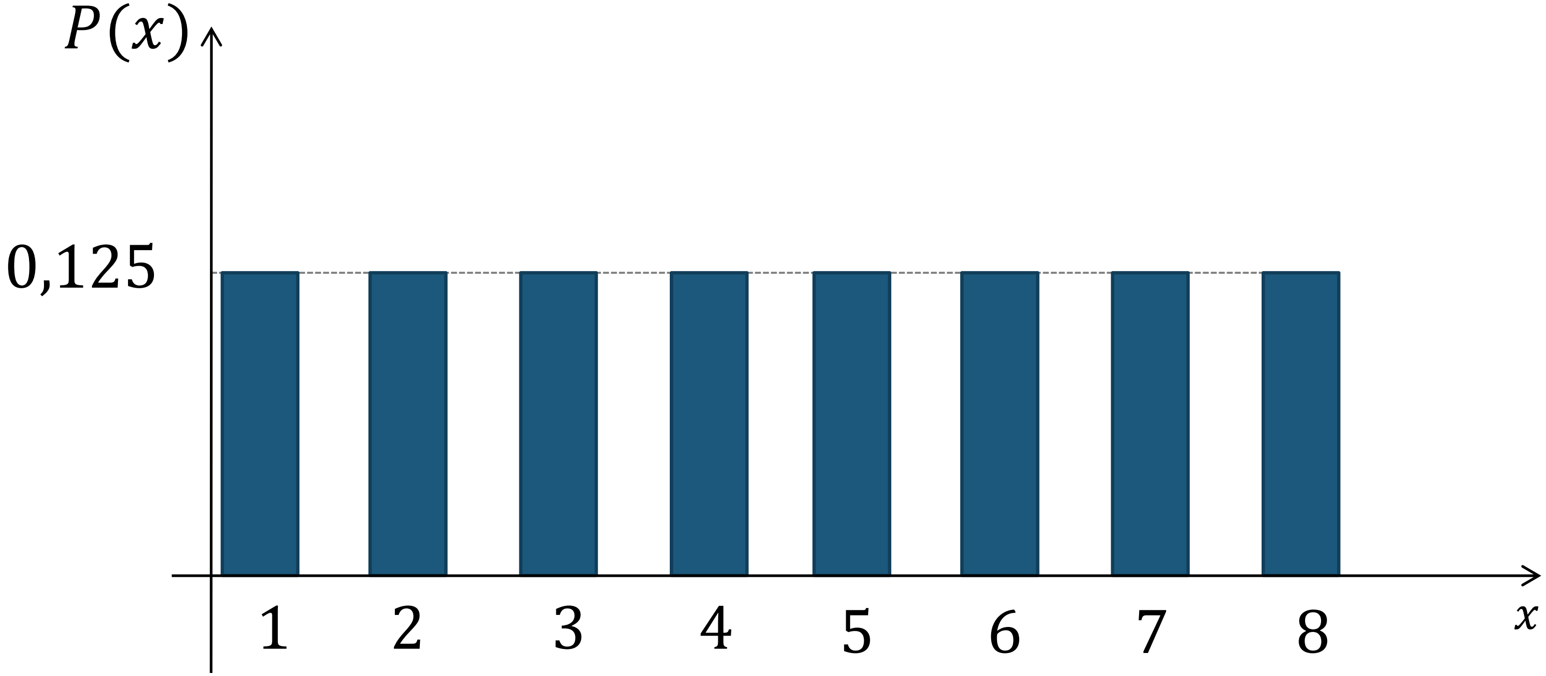
Histograma



Agora construir a distribuição de probabilidade, mas dividindo o círculo em 8 partes.

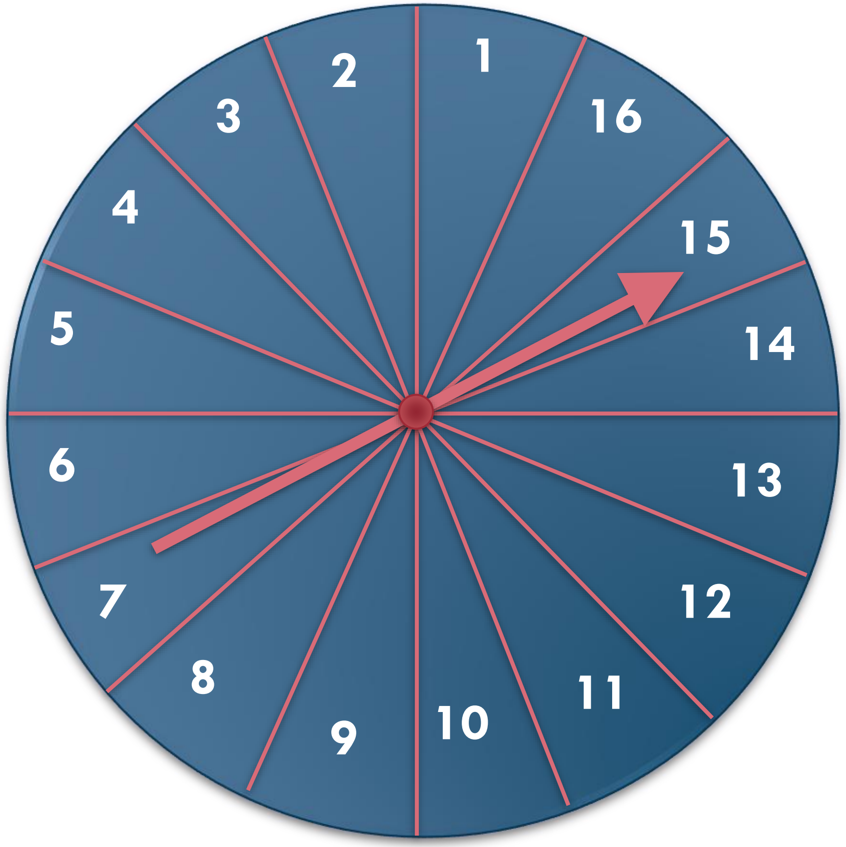


HISTOGRAMA

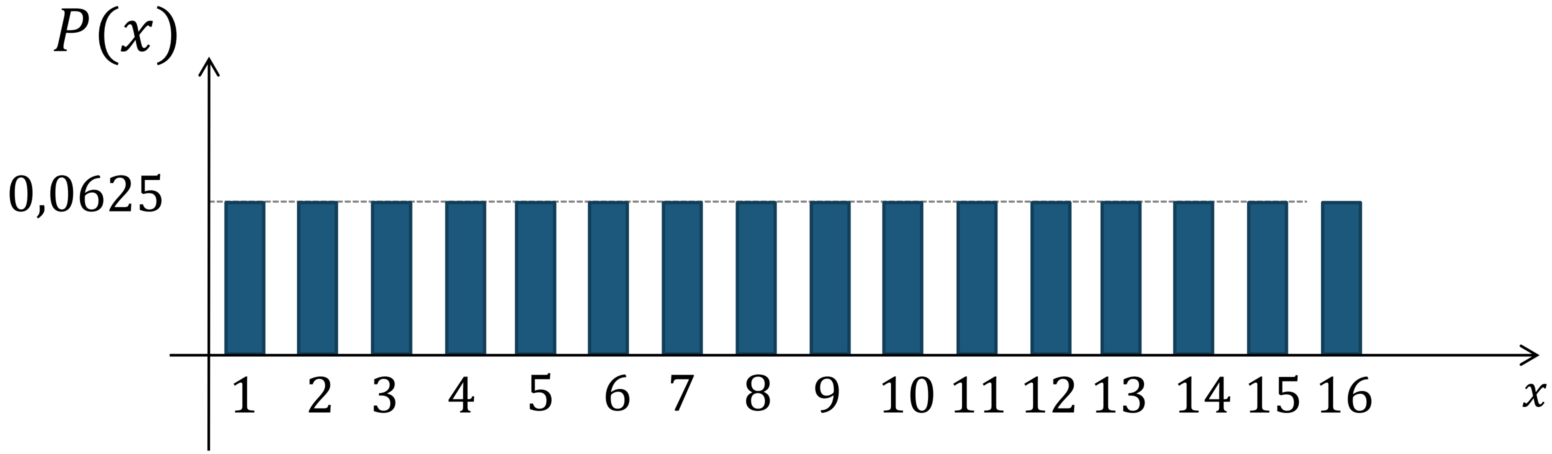


PECE Programa de Educação Continuada
Escola Politécnica da USP

Construir a distribuição de probabilidades para o número obtido neste experimento.



HISTOGRAMA



DÚVIDA...

Qual é o número máximo de setores que se consegue em um círculo?

Qual é a probabilidade dessa variável aleatória contínua assumir um determinado valor (10, por exemplo)?

A **distribuição de probabilidades**, ou modelo probabilístico, indica, para uma variável aleatória, **quais** são os **resultados** que podem ocorrer e **qual** é a **probabilidade** de cada resultado acontecer.

PROBABILIDADES...

O estudo de uma variável aleatória contínua é análogo ao das variáveis discretas. Porém, variáveis contínuas apresentam uma gama de resultados admissíveis variando continuamente em um intervalo de valores. Portanto,

- As probabilidades não podem mais ser calculadas através de equações do tipo PMF vistas na aula anterior:

$$P(X = x) = \text{FÓRMULA.}$$

- Para identificar uma distribuição contínua, existe a **função densidade de probabilidade**, que é uma equação do tipo *probability density function (PDF)*:

$$y = p(x).$$

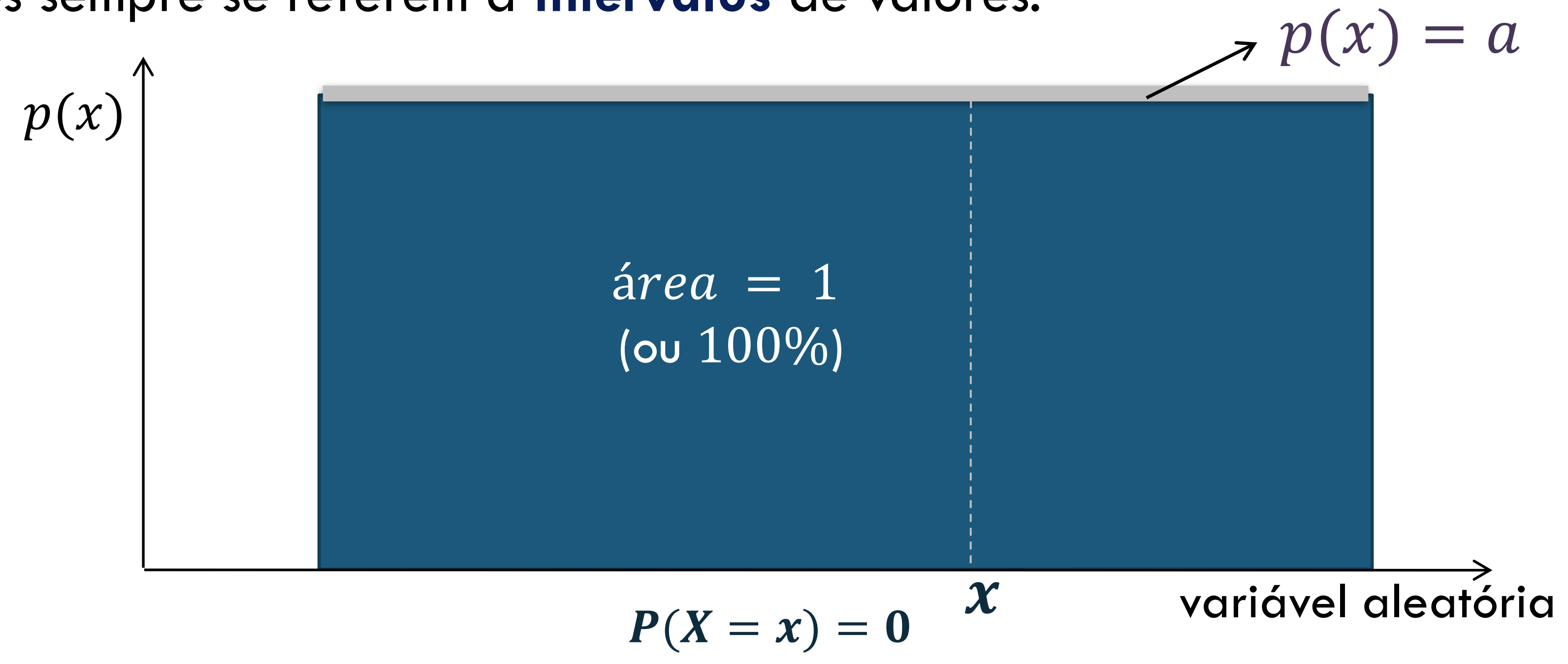
A probabilidade é calculada como a área sob a função (integrando).

FUNÇÃO DA DENSIDADE DE PROBABILIDADE

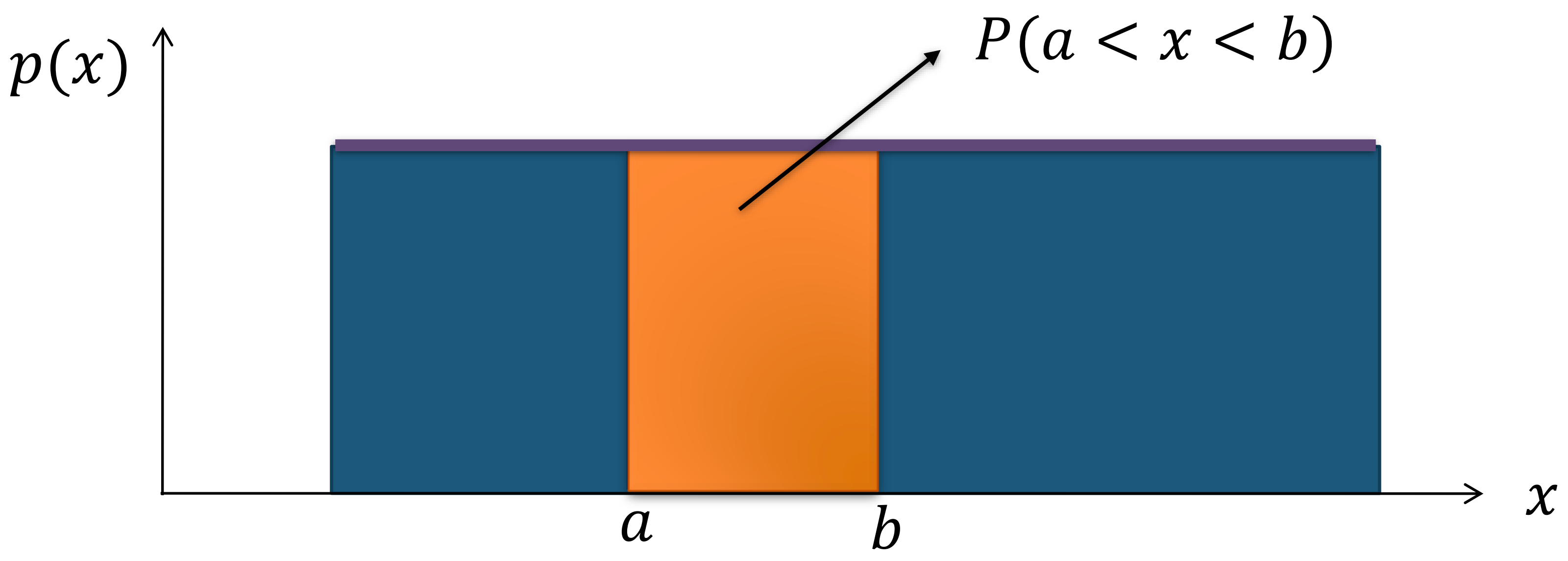
A área sob a função densidade é **1**.

A probabilidade da variável aleatória assumir um valor determinado é **zero** .

As probabilidades sempre se referem a **intervalos** de valores.

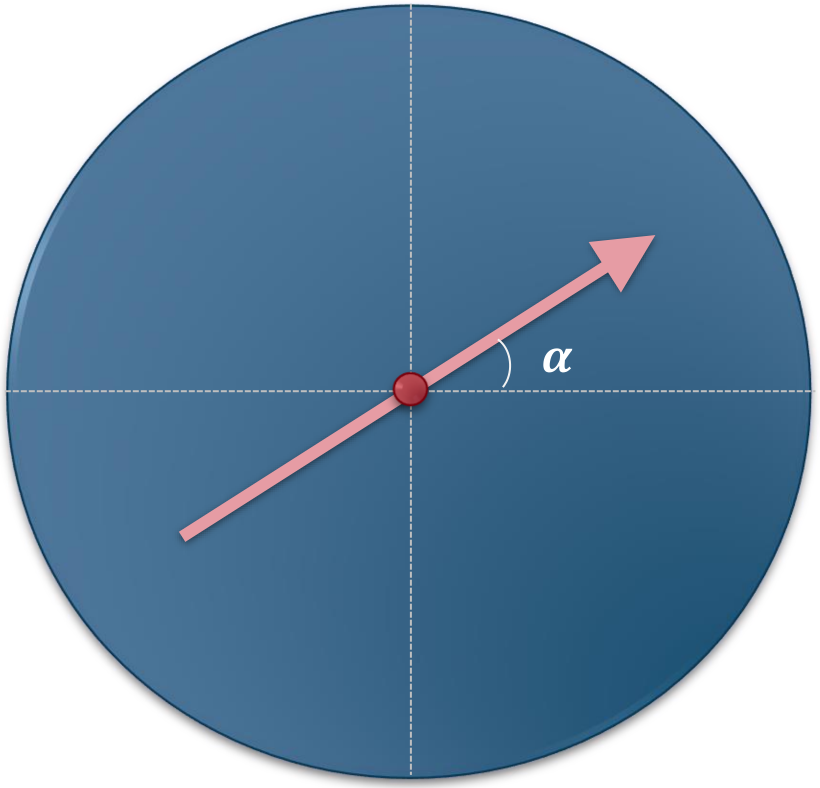


A probabilidade da variável aleatória assumir um valor em um intervalo é igual à **área** sob a função densidade naquele intervalo.

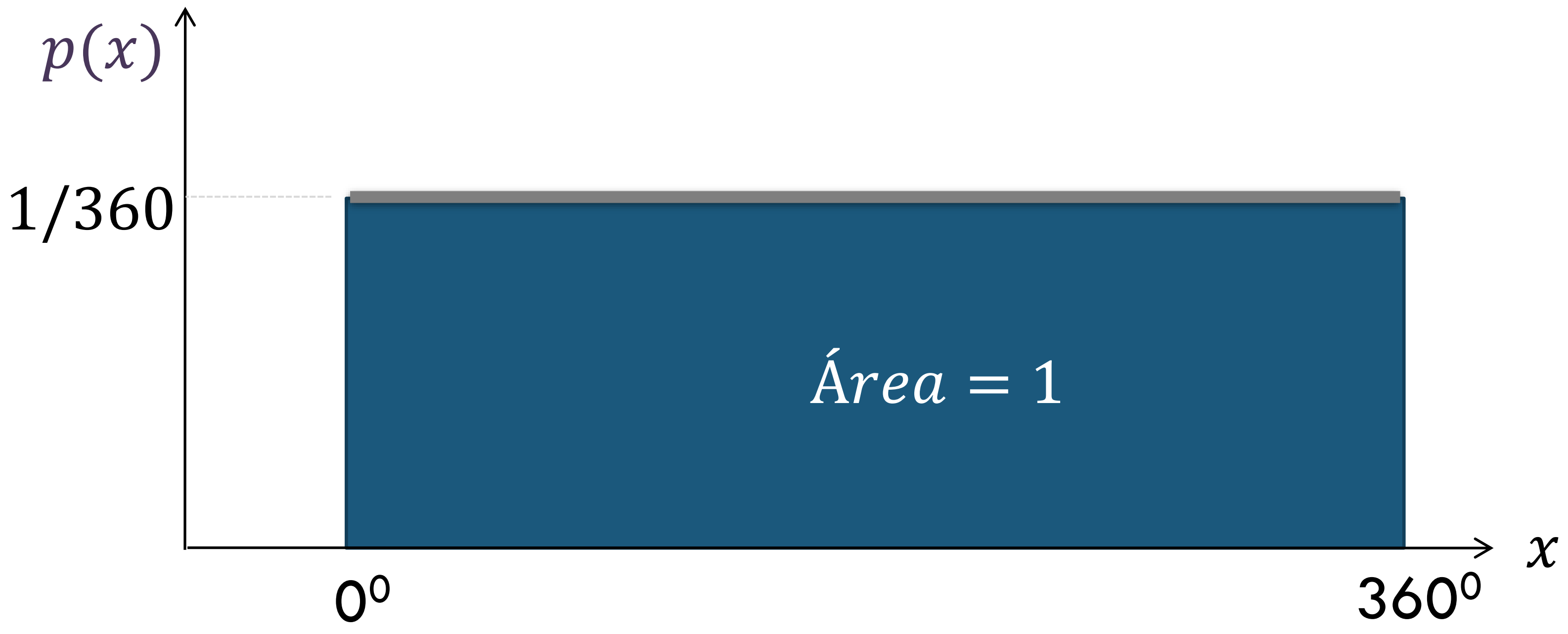


VOLTANDO AO EXERCÍCIO ANTERIOR

Fixa-se um ponteiro no centro de um círculo e anota-se o ângulo formado pelo ponteiro, quando parado, com o eixo horizontal.

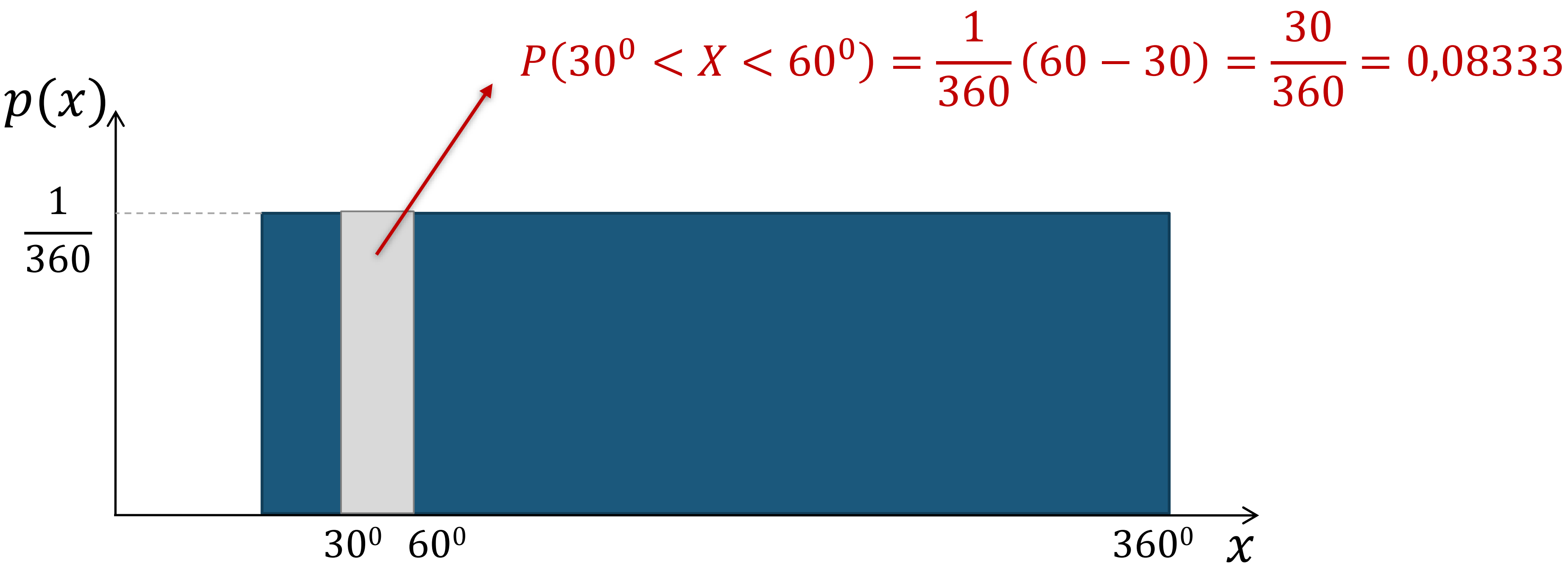


- Definir a **função densidade de probabilidades** para o ângulo α obtido.


$p(x)$
$1/360$
$Área = 1$
x
0^0
360^0

EXERCÍCIO

Qual é a probabilidade de se obter um ângulo entre 30^0 e 60^0?



PORTANTO, PARA FIXAR:

Variáveis aleatórias contínuas têm sua distribuição de probabilidade representada pela ***função densidade de probabilidade***, denominada de $p(x)$, sendo x a variável aleatória.

As probabilidades de uma variável aleatórias contínuas são definidas pela área embaixo da ***função densidade de probabilidade.***

FUNÇÃO DENSIDADE DE PROBABILIDADE

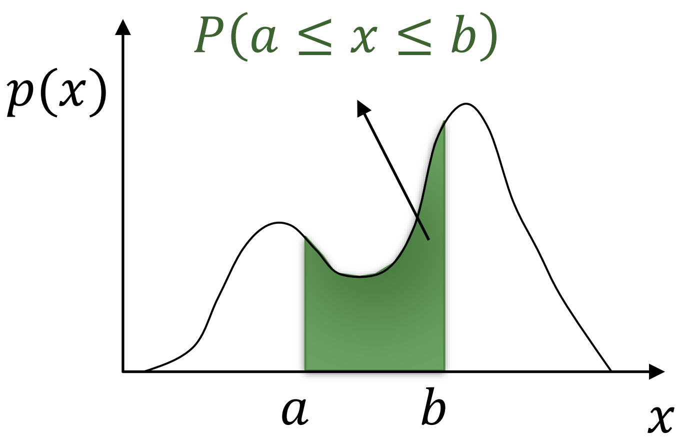
A função densidade de probabilidade ¨(em inglês *probability density function –pdf-)* DEVE apresentar as seguintes propriedades:

- A função $p(x) \geq 0$ para todos os valores admissíveis de x
- A integral da função deve assumir valor unitário no espaço amostral de x:

$$\int_x p(x)dx = 1$$

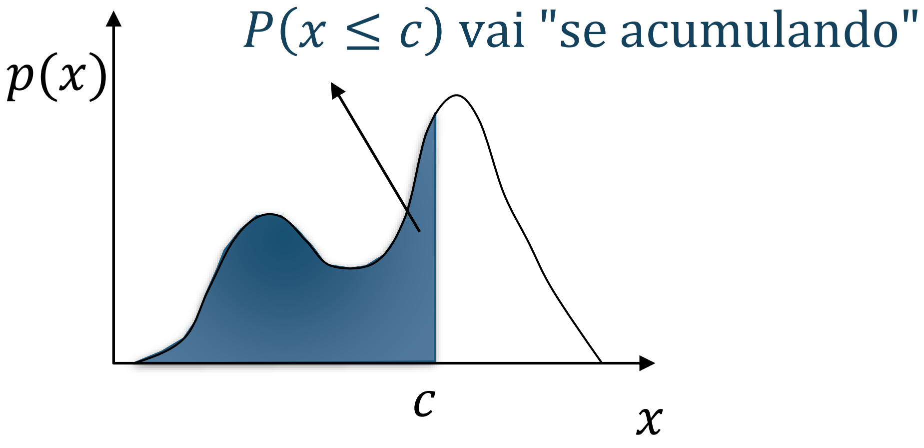
- A probabilidade de x assumir valores entre a e b é dada pela relação:

$$\int_a^b p(x)dx = P(a \leq x \leq b)$$



- A probabilidade de x ser inferior a um valor c é:

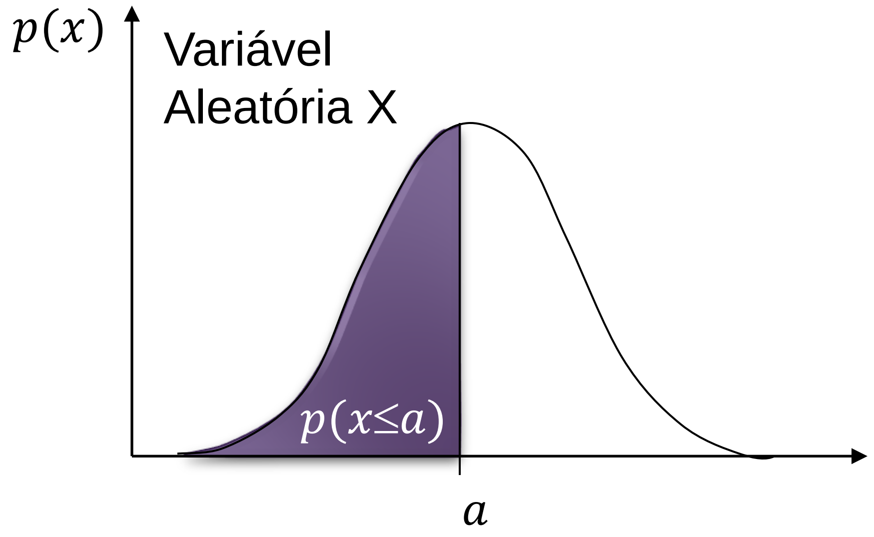
$$\int_{-\infty}^{c} p(x)dx = P(x \leq c)$$



- $P(x = x_0) = 0$, para x_0 fixo

Assim, $P(a < x < b) = P(a \leq x < b)$

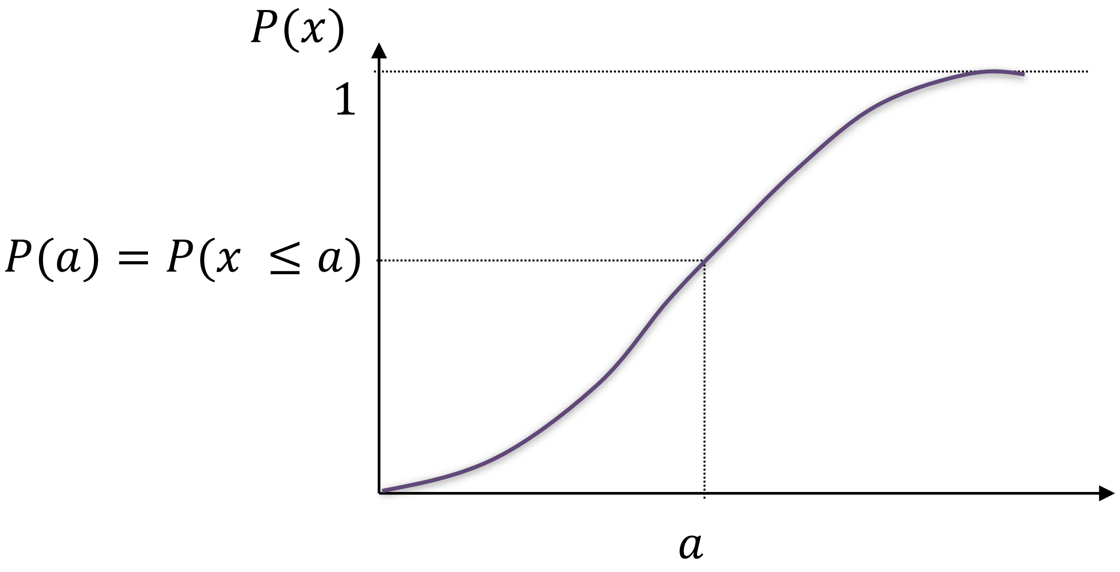
$= P(a < x \leq b) = P(a \leq x \leq b)$.



Função
Densidade de
Probabilidade

Probability
Density
Function

pdf



Função
Distribuição
Acumulada

Cumulative
Distribution
Function

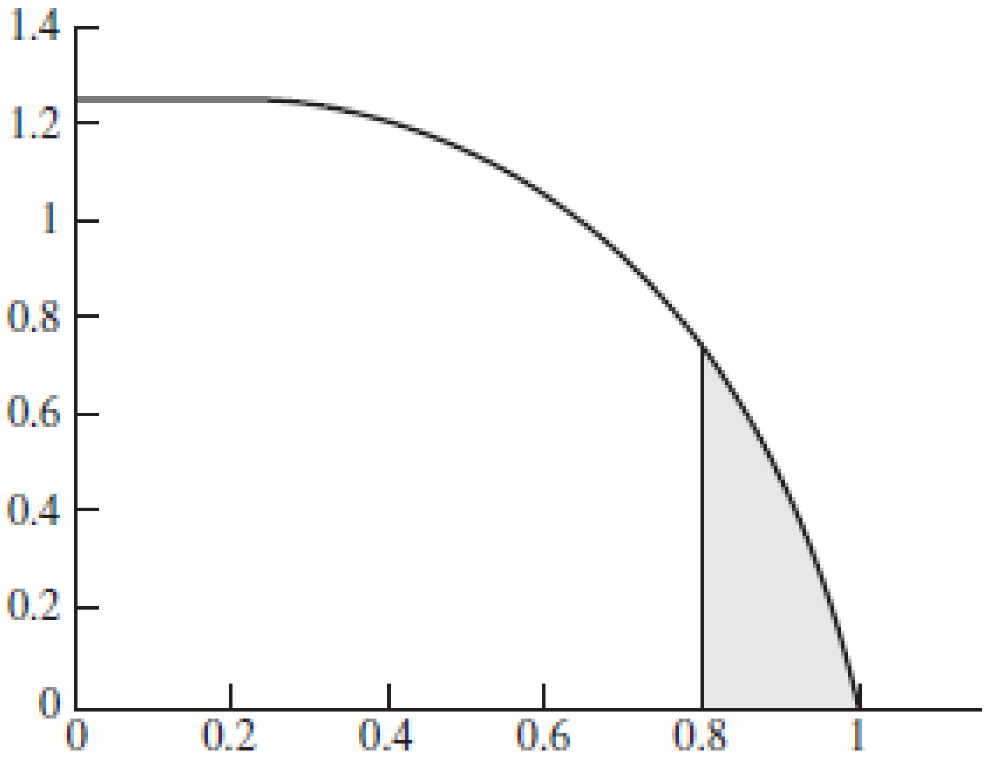
cdf

EXERCÍCIO. NOTEBOOK

Se perfura um componente de uma folha de metal e depois se insere um eixo através do furo. A folga do eixo é igual à diferença entre o raio do furo e o raio do eixo. Seja X a variável aleatória que denota a folga, em milímetros. A função de densidade de probabilidade de X é:

$$f(x) = \begin{cases} 1.25(1 - x^4) & \text{se } 0 < x < 1 \\ 0 & \text{caso contrário} \end{cases}$$

Os componentes com folgas maiores de 0.8 mm devem ser descartados. Quantos descartamos?



PECE Programa de Educação Continuada
Escola Politécnica da USP

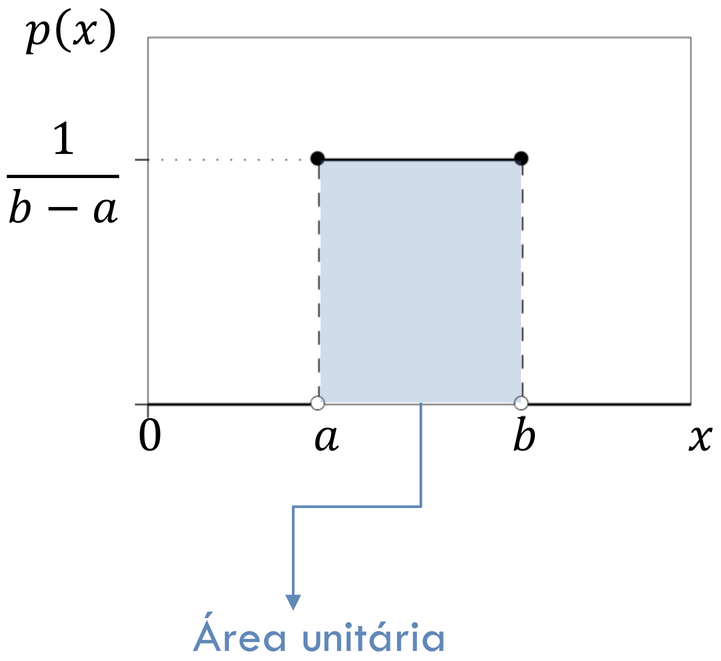
DISTRIBUIÇÃO UNIFORME

A mais simples

DISTRIBUIÇÃO UNIFORME

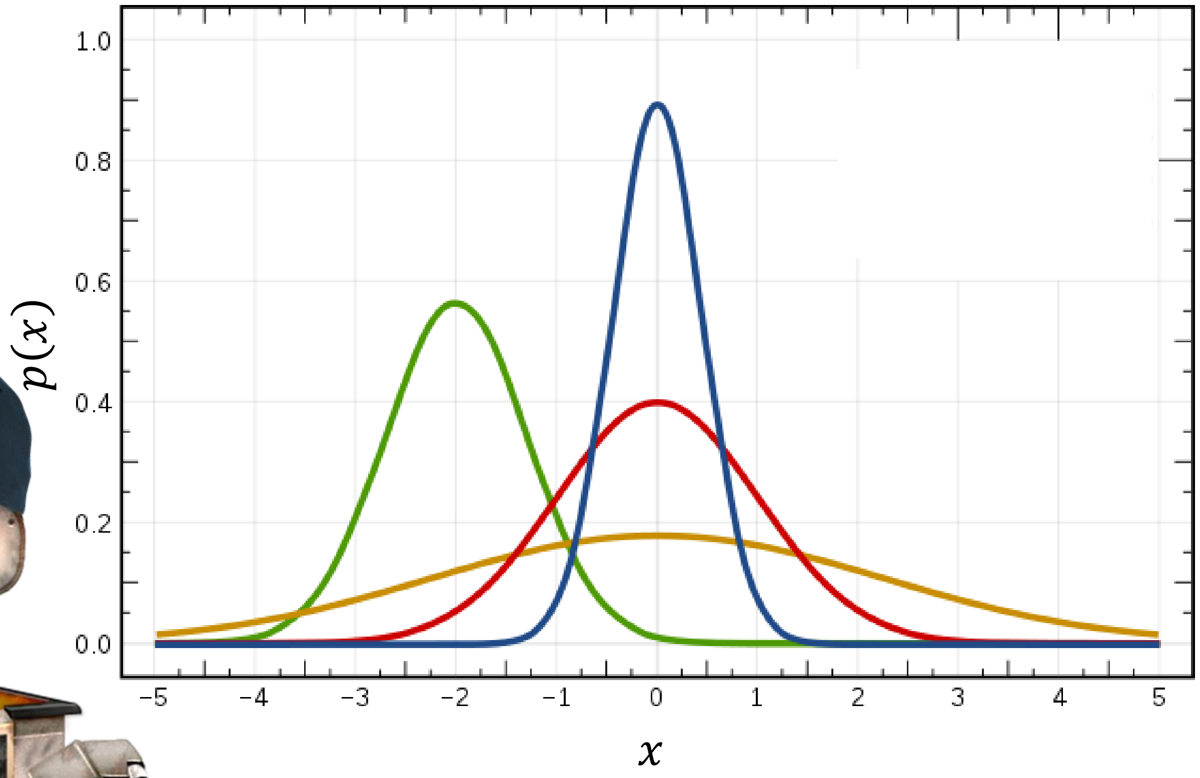
A função densidade de probabilidade (em inglês PDF, de *probability density function*) de uma distribuição uniforme $x \sim Uniforme(a, b)$ é dada por:

$$p(x) = \begin{cases} \dfrac{1}{b-a} & a \leq x \leq b \\ 0 & x < a \text{ ou } x > b \end{cases}$$



sendo x a variável aleatória.

Qualquer intervalo de números de largura igual tem uma probabilidade igual de ser observado.



DISTRIBUIÇÃO GAUSSIANA

$$N \sim (\mu, \sigma^2)$$

DISTRIBUIÇÃO NORMAL

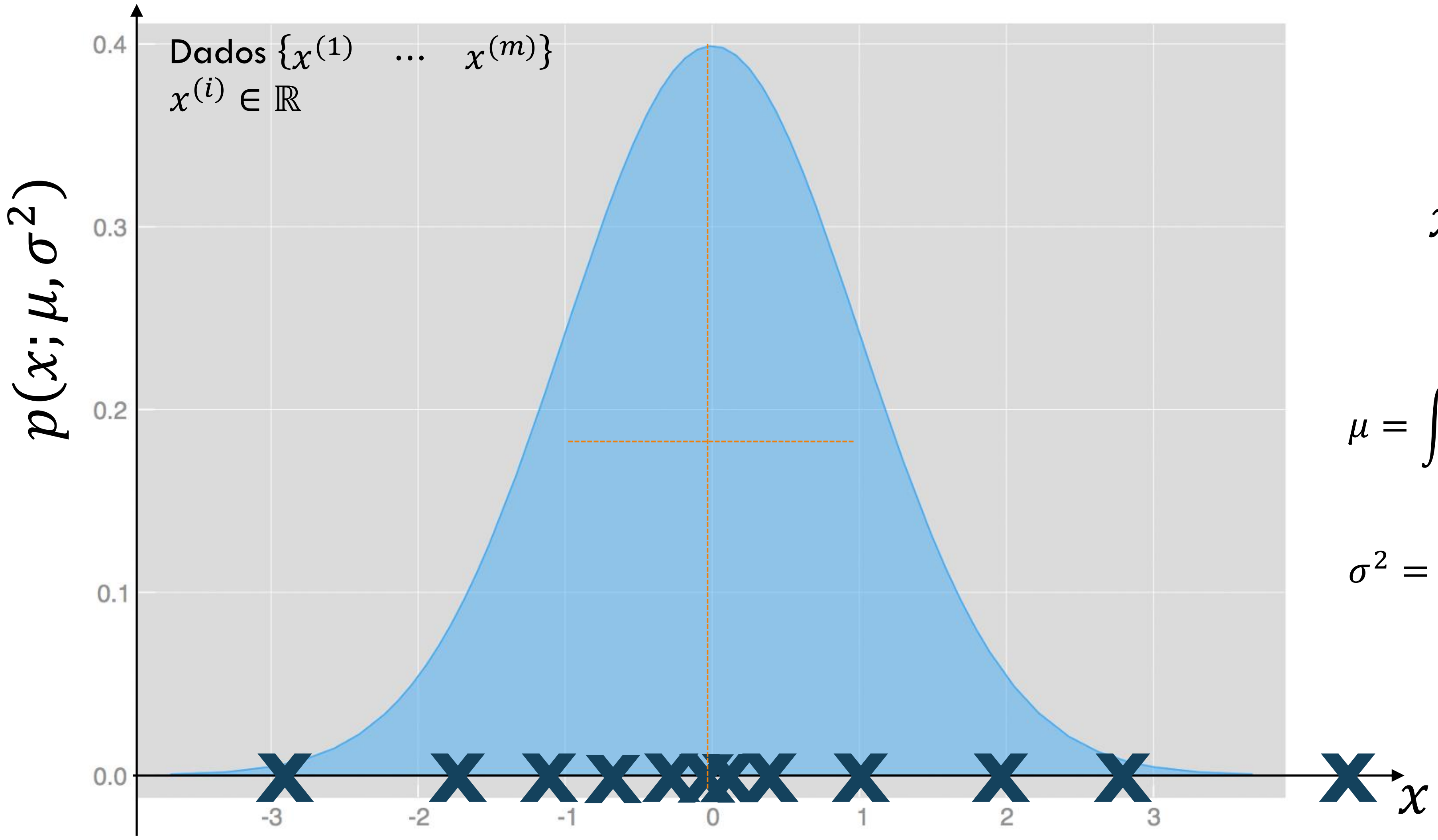
A função densidade de probabilidade de uma distribuição normal é dada por:

$$p(x;\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad -\infty \leq x \leq \infty$$

sendo x a variável aleatória.

A distribuição normal é uma distribuição caracterizada por dois parâmetros: a média da população (μ) e variância da população (σ_x), $N \sim (\mu, \sigma^2)$.

INTUIÇÃO DE GAUSS – DISTRIBUIÇÃO NORMAL

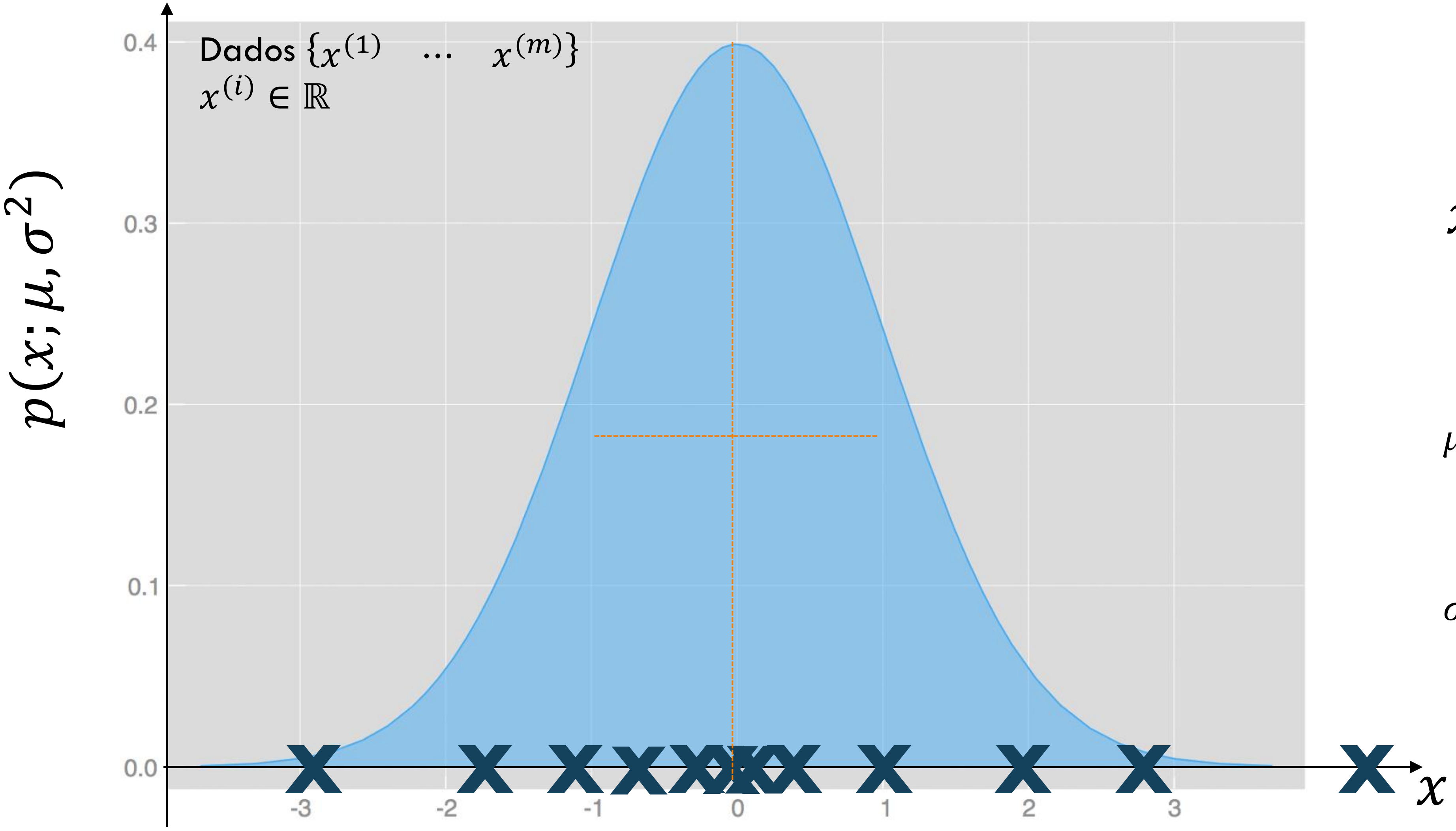


$$x^{(i)} \sim N(\mu, \sigma^2)$$

$$\mu = \int_{-\infty}^{\infty} x p(x) dx$$

$$\sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 p(x) dx$$

INTUIÇÃO DE GAUSS – ESTIMAÇÃO DE PARÂMETROS



$$x^{(i)} \sim N(\mu, \sigma^2)$$

$$\mu = \frac{1}{m}\sum_{i=1}^{m} x^{(i)}$$

$$\sigma^2 = \frac{1}{m}\sum_{i=1}^{m}\left(x^{(i)} - \mu\right)^2$$

MÁXIMA VEROSSIMILHANÇA PARA GAUSS

$$\mathcal{L}\{\mu,\sigma\} = \prod_{i=1}^{m} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2\right]$$

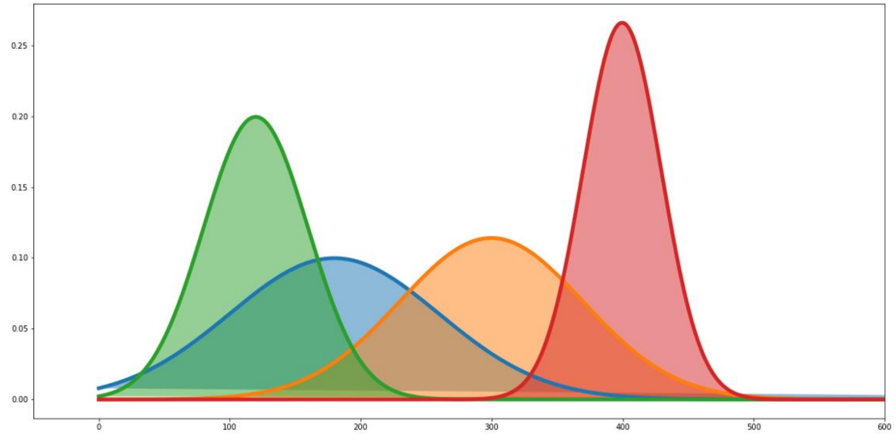
$$\mathcal{L}\{\mu,\sigma\} = (2\pi)^{-m/2}\sigma^{-m} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{m}(x_i-\mu)^2\right]$$

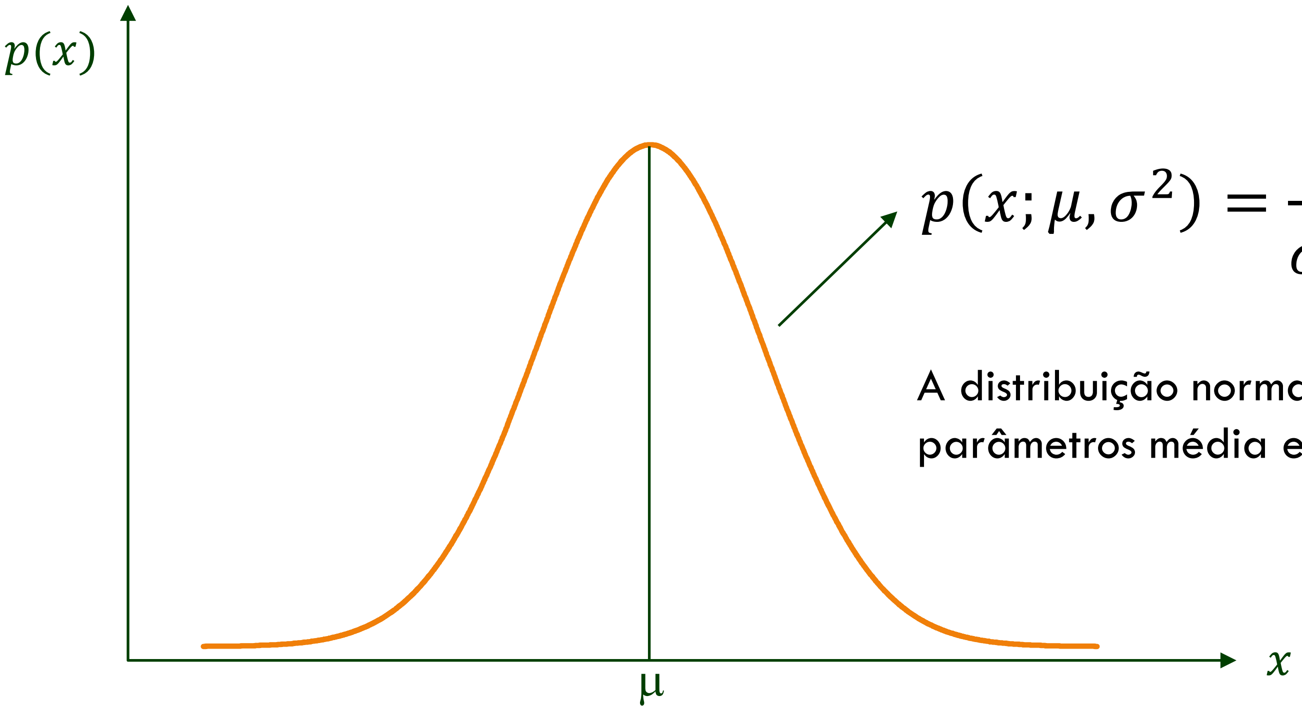
função de log-verossimilhança negativa

$$\mathbf{L}\{\mu,\sigma\} = \frac{m}{2}\ln 2\pi + m\ln\sigma + \frac{1}{2\sigma^2}\sum_{i=1}^{m}(x_i-\mu)^2$$

$$\frac{\partial \mathbf{L}\{\mu,\sigma\}}{\partial\mu} = \frac{1}{\sigma^2}\sum_{i=1}^{m}(x_i-\mu) = 0 \quad\Longrightarrow\quad \hat{\mu} = \frac{\sum_{i=1}^{m} x_i}{m}$$

$$\frac{\partial \mathbf{L}\{\mu,\sigma\}}{\partial\sigma} = \frac{m}{\sigma} - \frac{1}{\sigma^3}\sum_{i=1}^{m}(x_i-\mu)^2 = 0 \quad\Longrightarrow\quad \hat{\sigma} = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(x_i-\mu)^2}$$

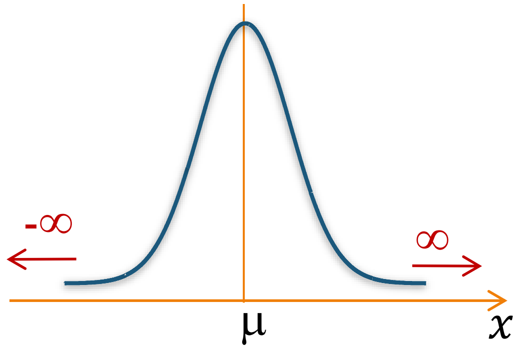


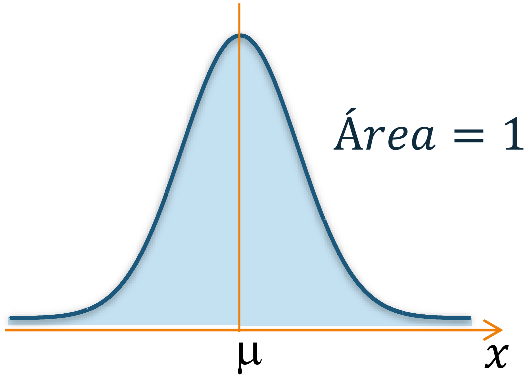


$$p(x;\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

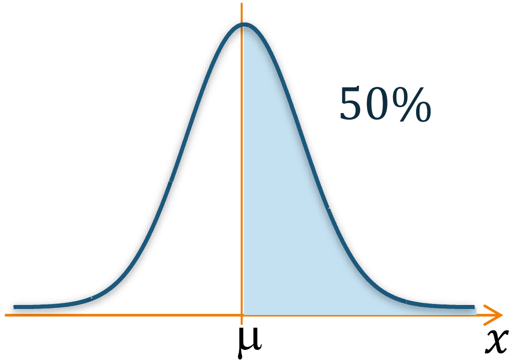
A distribuição normal é totalmente caracterizada pelos parâmetros média e variância.

- A variável aleatória pode assumir valores de -∞ a +∞.
- Área unitária.
- A distribuição normal é simétrica, centrada na própria média da população, sendo coincidentes os valores da moda, mediana e média.

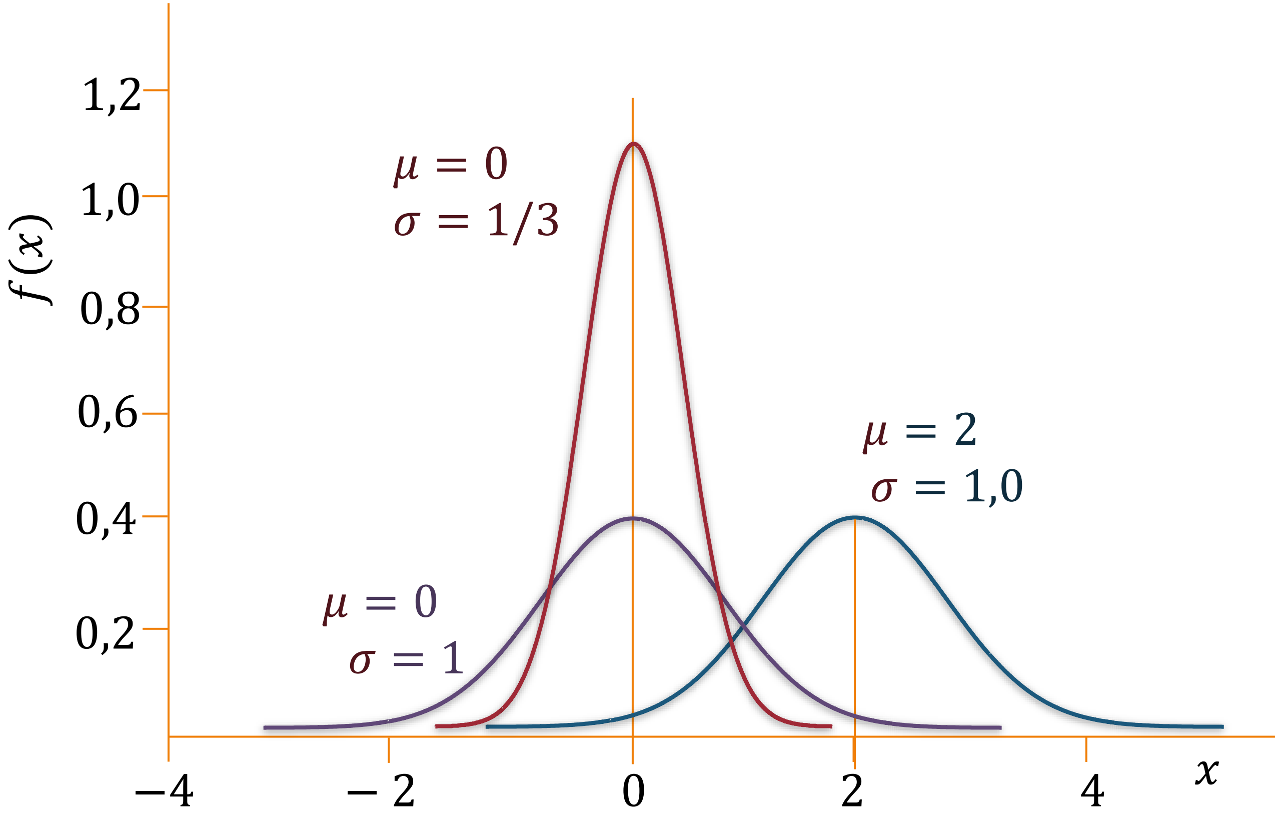




Dessa forma, tem-se que 50% da distribuição encontra-se à direita da média e os restantes 50% à esquerda desse parâmetro.



A **representação gráfica da função densidade de probabilidade da distribuição normal** é a seguinte:



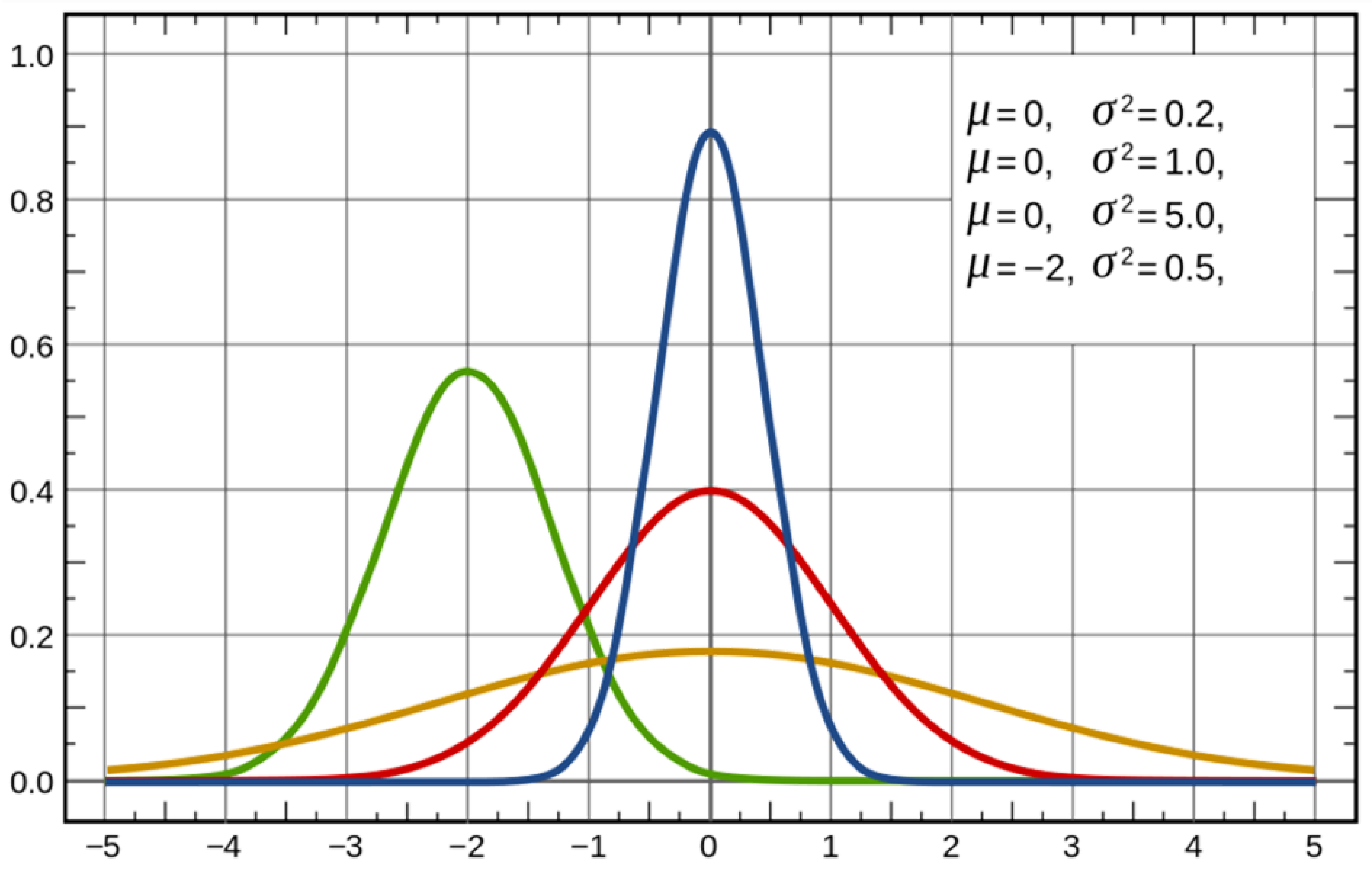
A média me da o centro da função distribuição de probabilidade e coincide com moda e mediana.

O desvio padrão (ou a variância) determina "largura" e "achatamento".

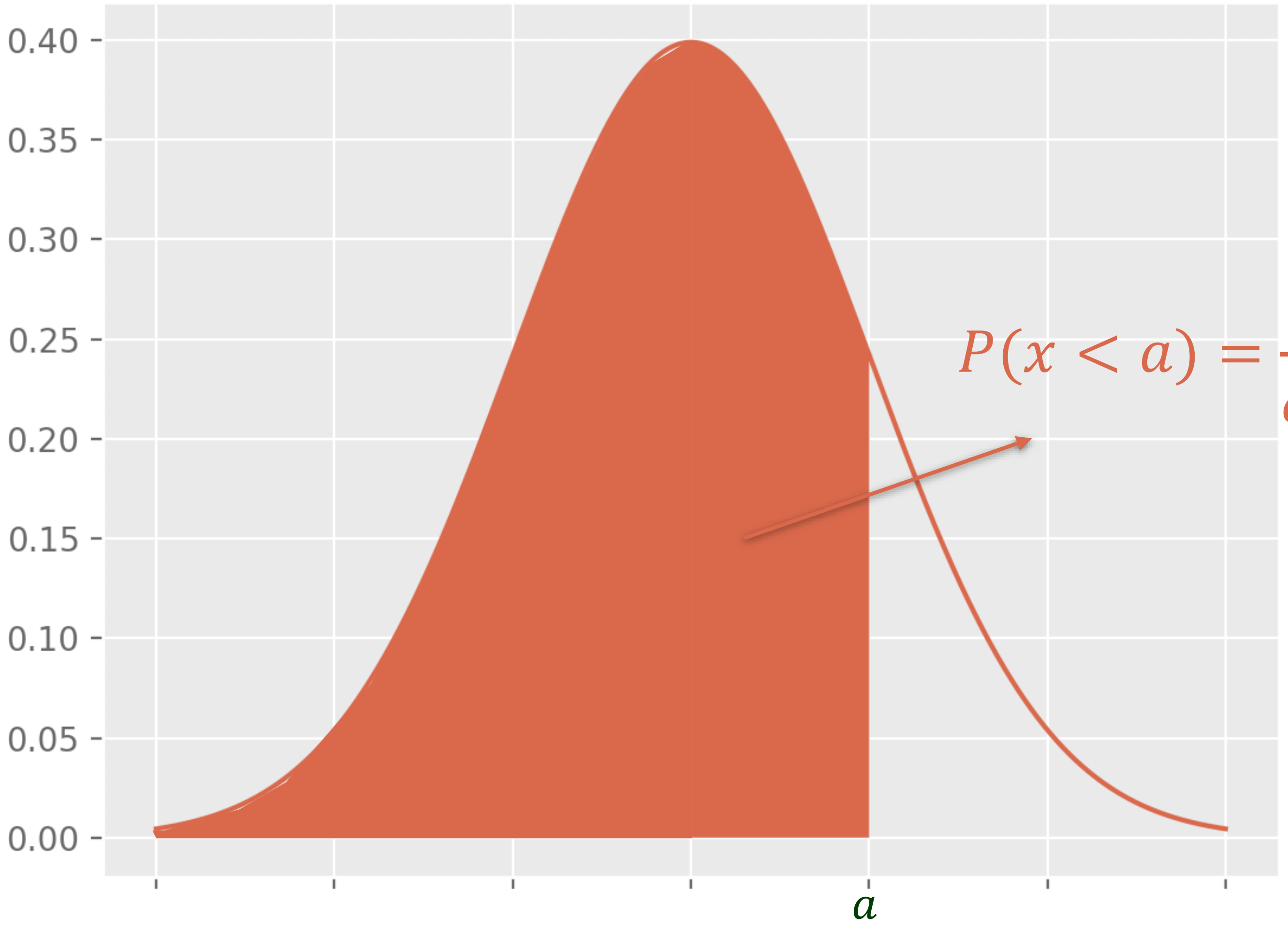
Quanto maior a variância maior será a dispersão da distribuição e mais achatada será a curva da função densidade de probabilidade.

Ao final a área total da função de densidade de probabilidades deve ser ??

COMPLETE A LEGENDA...

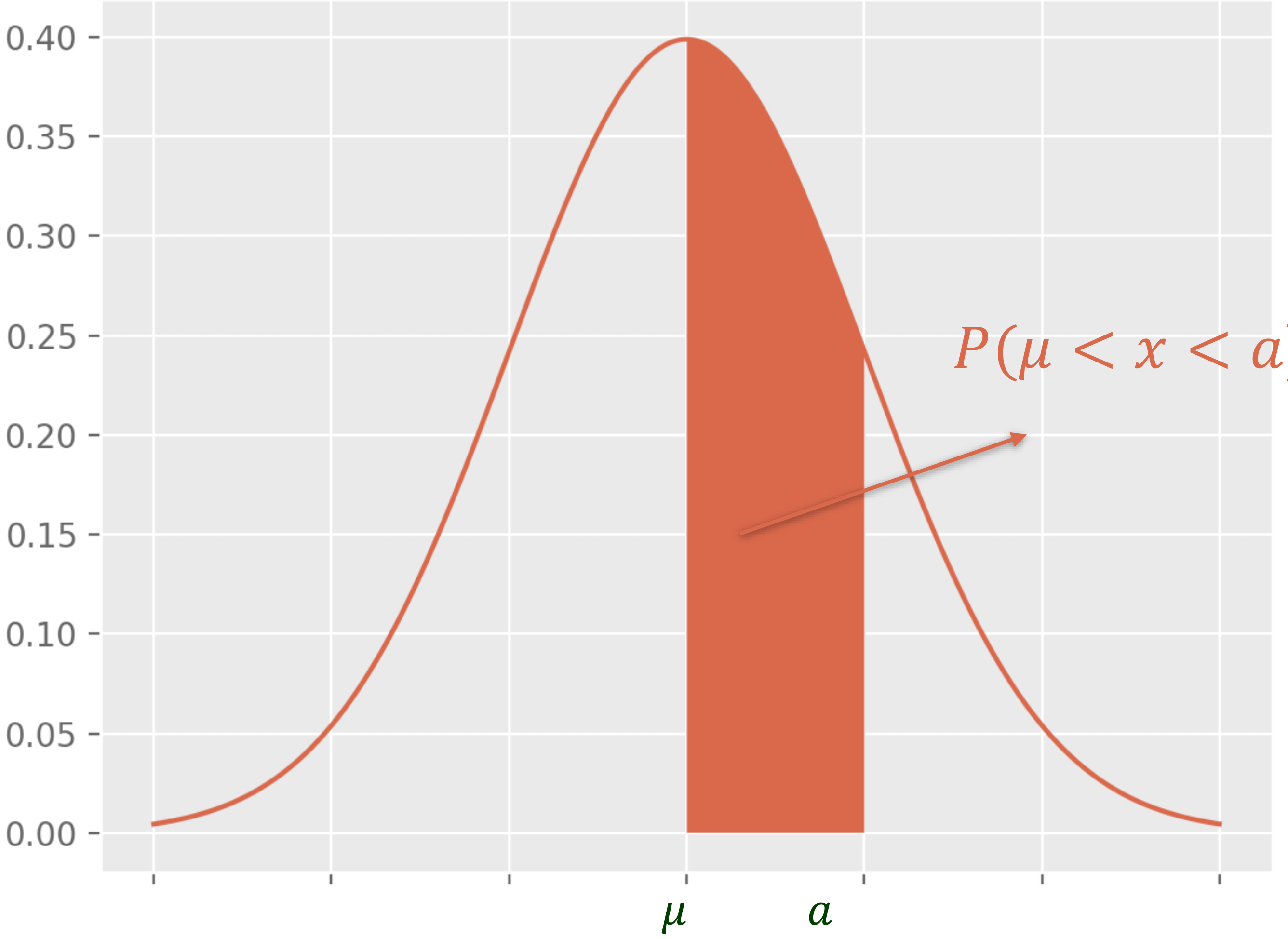


FUNÇÃO DISTRIBUIÇÃO ACUMULADA DA DISTRIBUIÇÃO NORMAL



$$P(x < a) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{a} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

CÁLCULO DE PROBABILIDADE



$$P(\mu < x < a) = \frac{1}{\sigma\sqrt{2\pi}} \int_{\mu}^{a} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

DISTRIBUIÇÃO ACUMULADA

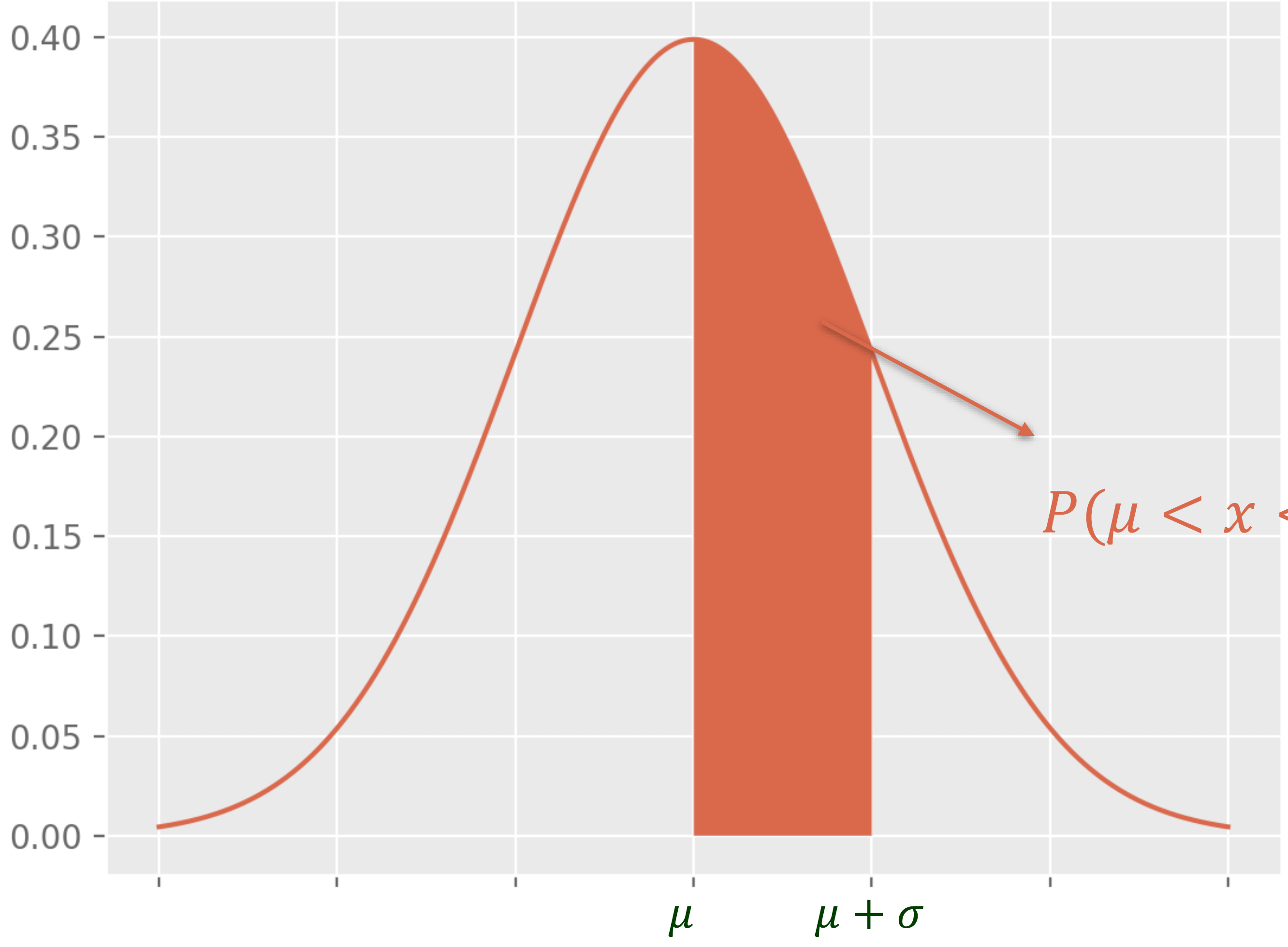
A função distribuição acumulada da distribuição normal é obtida pela execução da seguinte operação de integração:

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

O cálculo de áreas sob a curva normal é consideravelmente complexo.

Integração de $p(x)$ só com desenvolvimento em séries. Por isso, é conveniente trabalhar com valores padronizados.

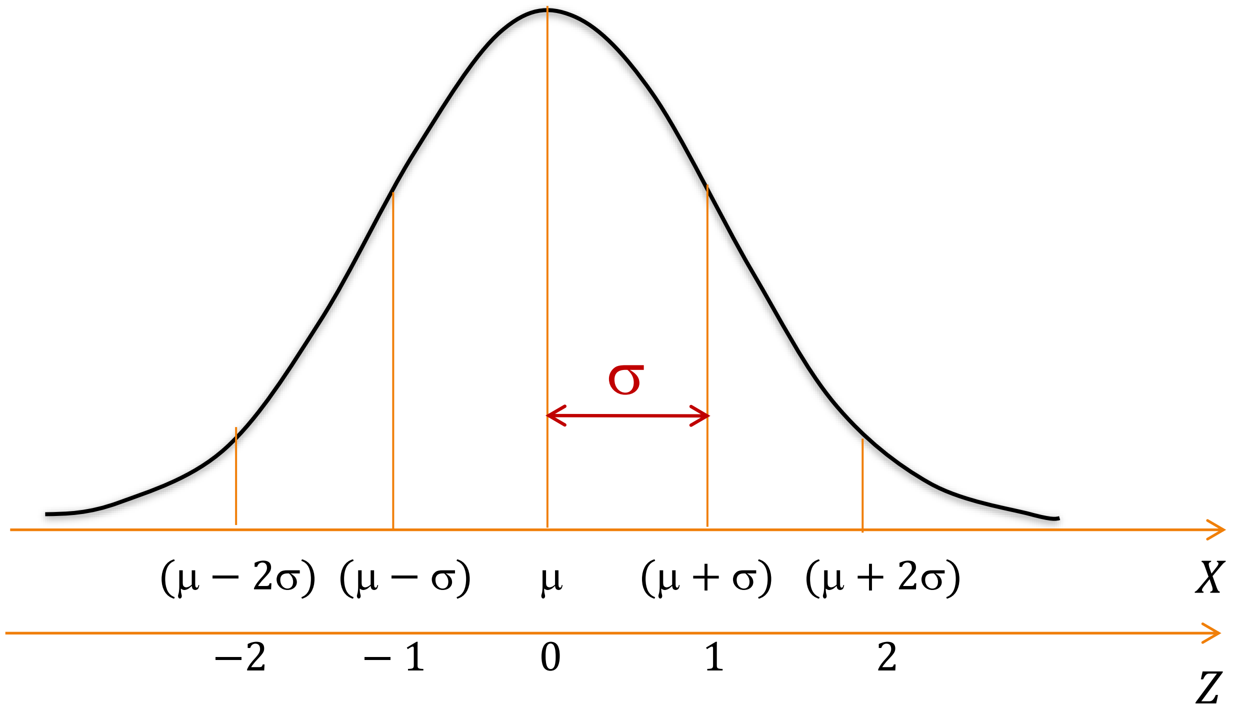
DISTÂNCIA PADRONIZADA



Distância expressa em função do número de desvios padrões (distância dividida pelo desvio padrão).

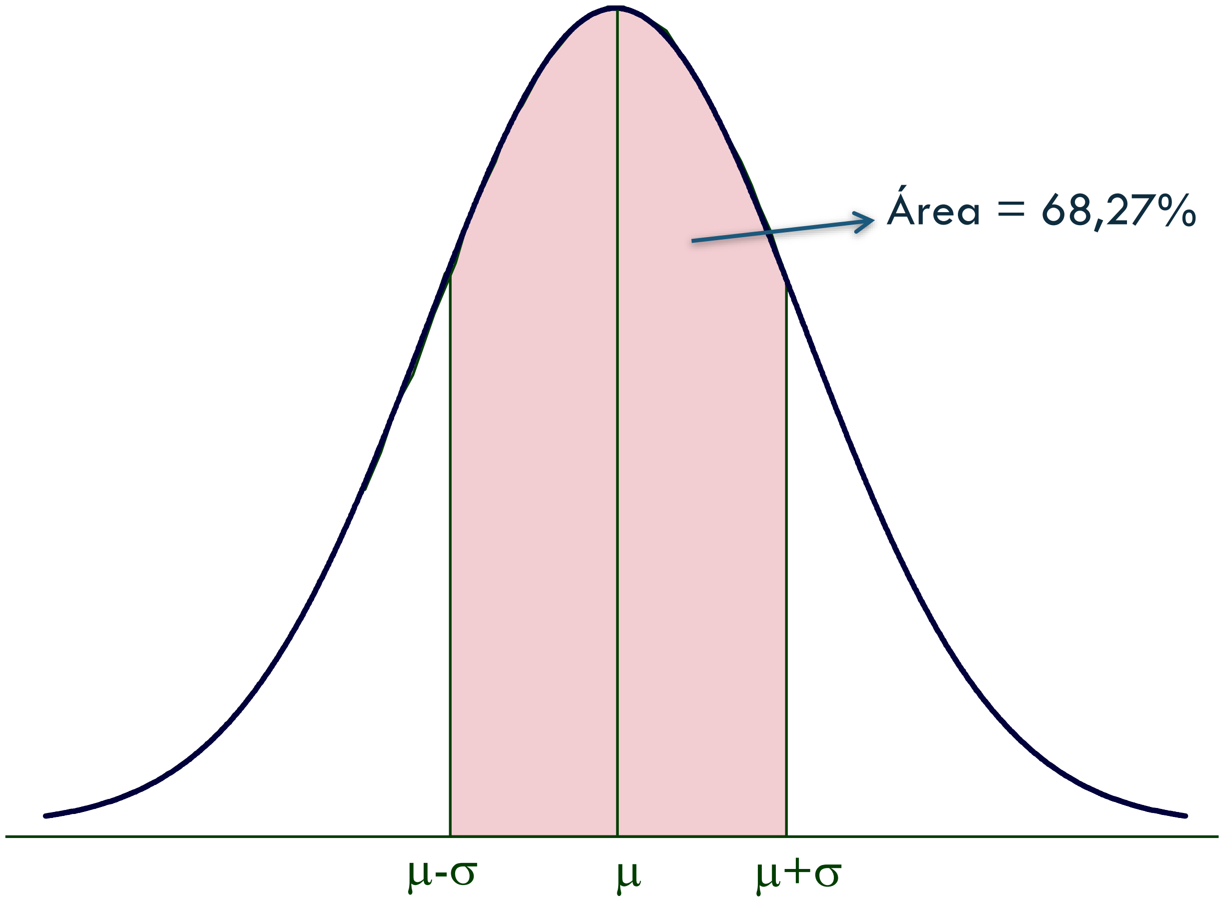
$$P(\mu < x < \mu + \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \int_{\mu}^{\mu+\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

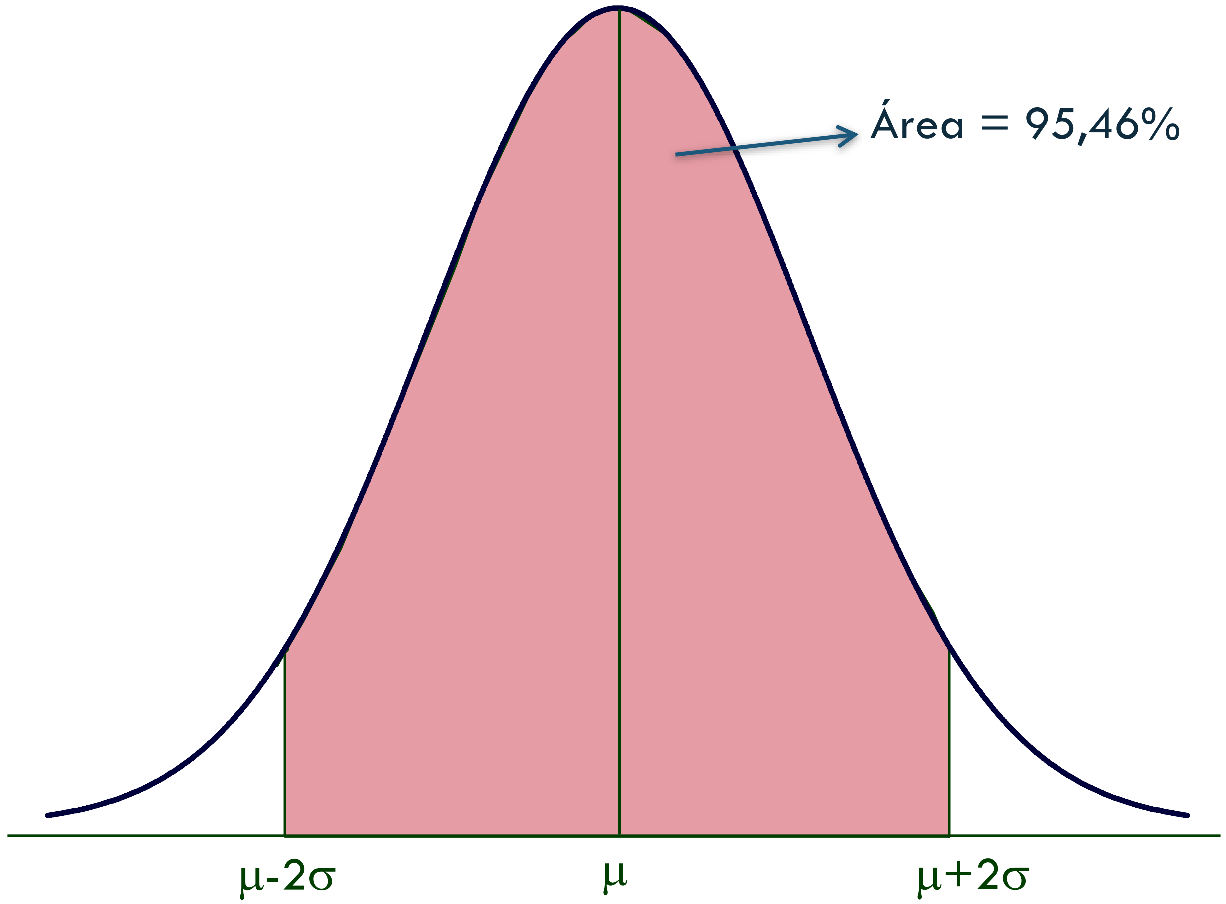
VARIÁVEL NORMAL PADRONIZADA



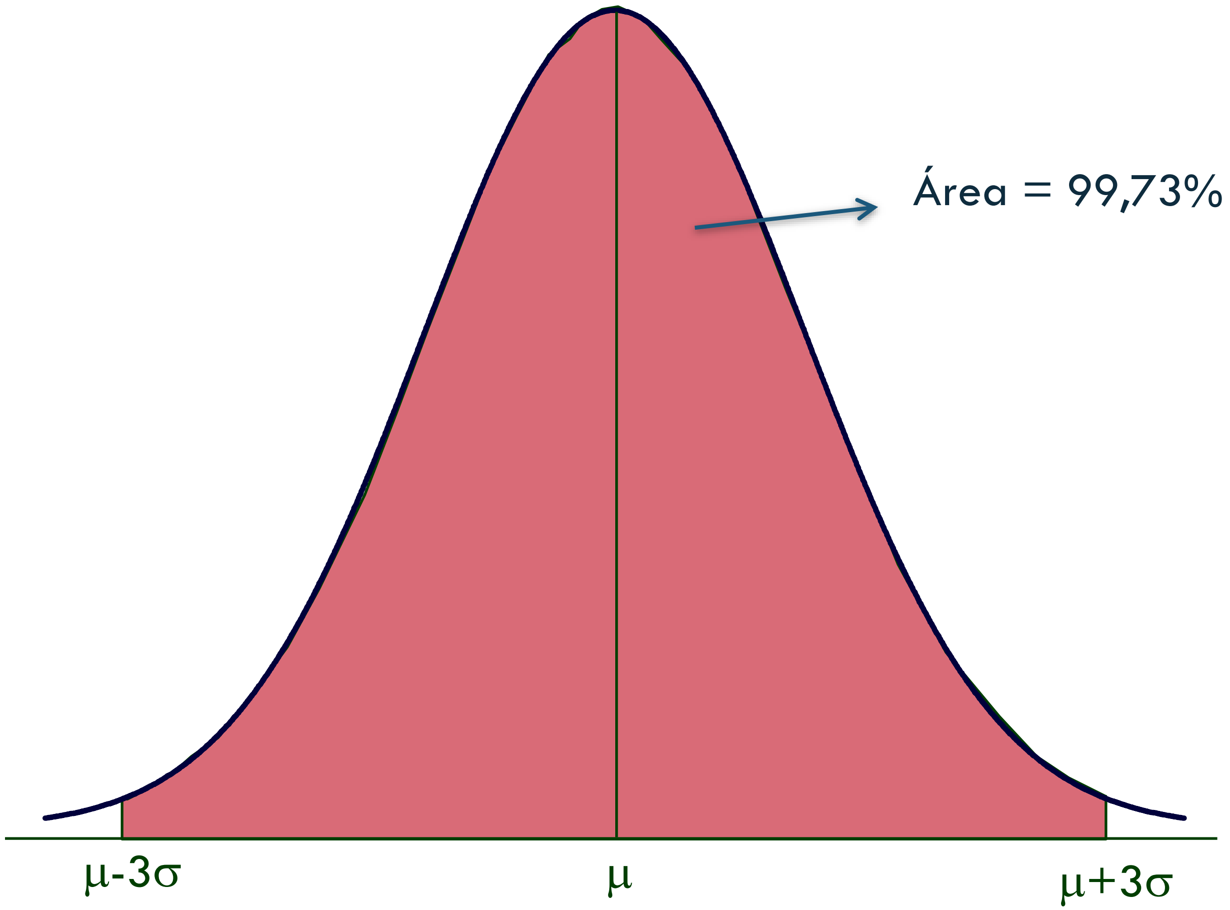
Score Z

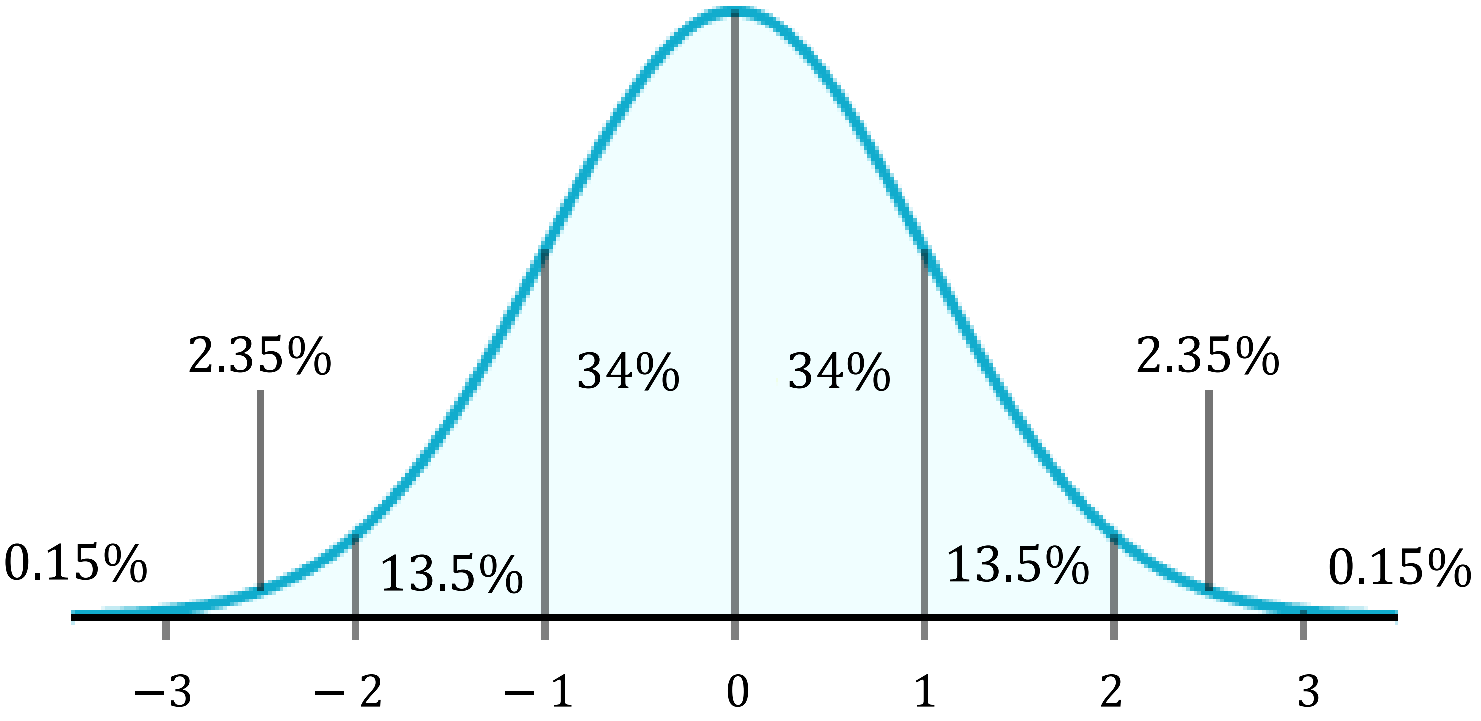
É o quanto uma medida se afasta da média em termos de Desvios Padrão (*para mais ou para menos*).


Área = 68,27%
μ-σ
μ
μ+σ



Área = 95,46%
μ-2σ
μ
μ+2σ


Área = 99,73%
μ-3σ
μ
μ+3σ

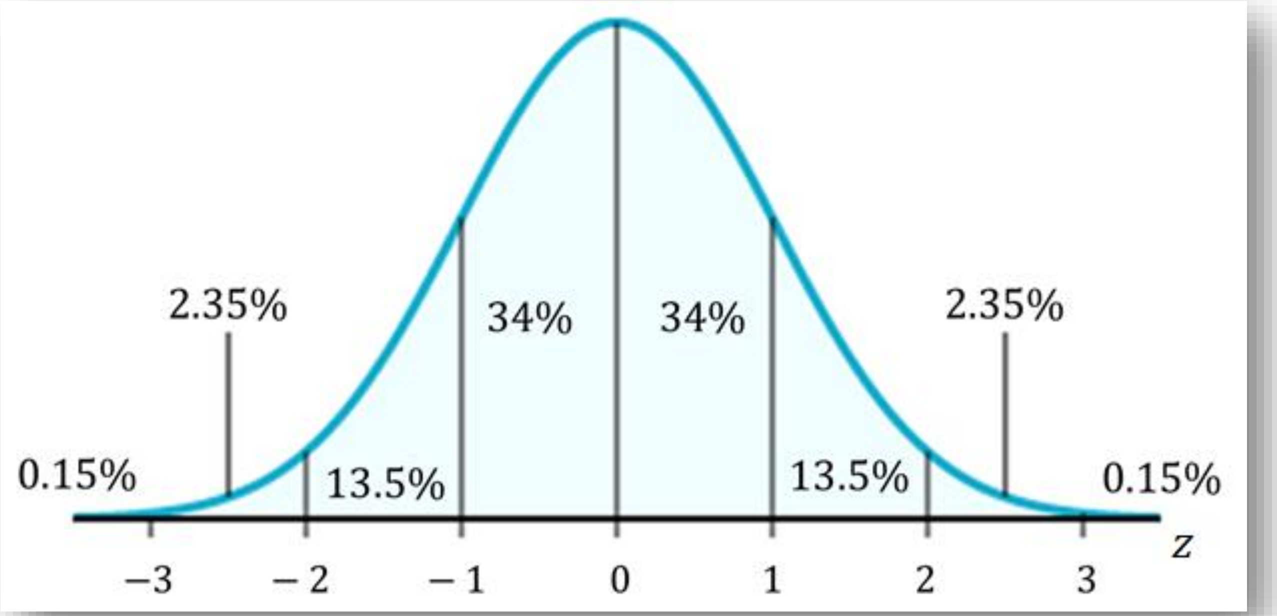


Aproximadamente 68% por cento dos dados estão dentro de 1 desvio padrão da média
Aproximadamente 95% dos dados estão dentro de 2 desvios padrão da média
Aproximadamente 99,7% dos dados estão dentro de 3 desvios-padrão da média

EXERCÍCIO

Com base nos valores normais padronizados, calcule as probabilidades:

A. $P(Z > 1)$
B. $P(Z < -1)$
C. $P(Z < 1)$
D. $P(-1 < Z < 1)$
E. $P(-2 < Z < 2)$
F. $P(-1 < Z < 3)$



NORMAL PADRONIZADA

Para padronizar uma variável normal, toma-se a média como ponto de referência e o desvio padrão como medida de afastamento,

$$\mu_z = 0$$
$$\sigma_z = 1$$

A nova variável será chamada de z,

$$z = \frac{x - \mu}{\sigma}$$

A função densidade de probabilidade da variável z é dada por,

$$\varphi(z) = \frac{1}{2\pi} e^{-\frac{1}{2}z^2}$$

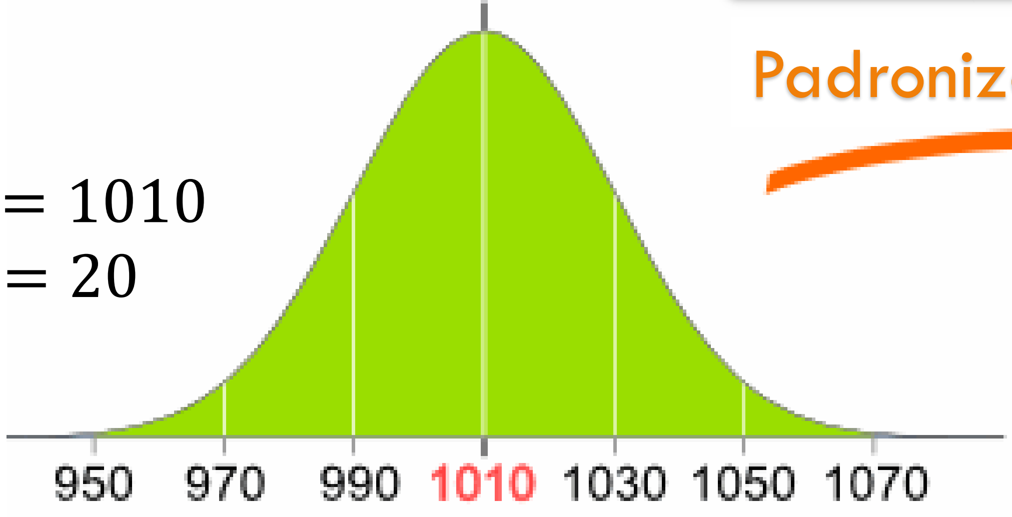
A distância entre a média e um ponto qualquer é dado em número de desvios padrões (z)

PECE Programa de Educação Continuada
Escola Politécnica da USP

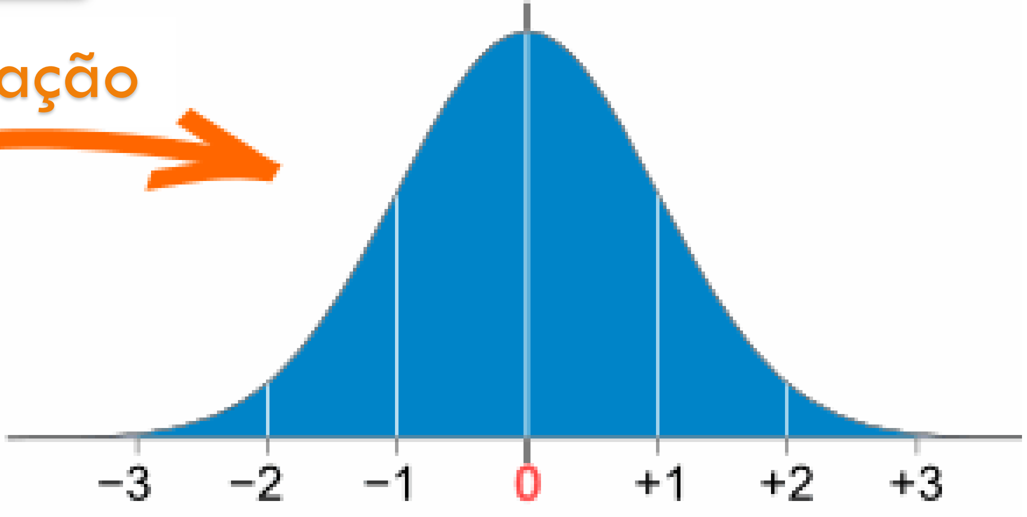
$$z = \frac{x - \mu}{\sigma}$$

Padronização

$\mu = 1010$
$\sigma = 20$



Distribuição normal



Distribuição normal padronizada

EXEMPLO

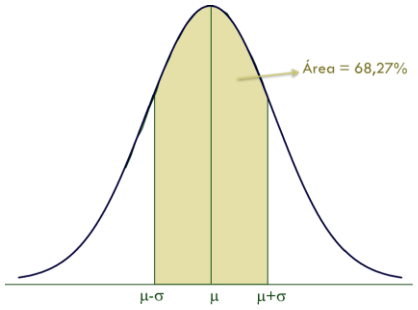
Considere que a glicemia de uma população de pessoas sadias tenha uma distribuição normal, com media 90mg e desvio padrão 5mg. Pode-se concluir que,

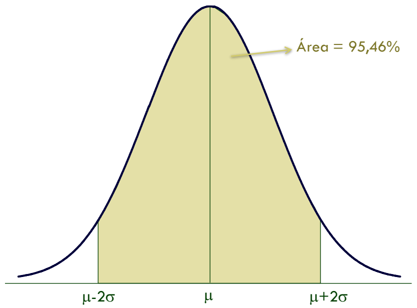
Aproximadamente $2/3$ $(\approx 68\%)$ da população de indivíduos sadios possui glicemia entre $\mu - \sigma = 90 - 5 = 85mg$ e $\mu + \sigma = 90 + 5 = 95mg$;

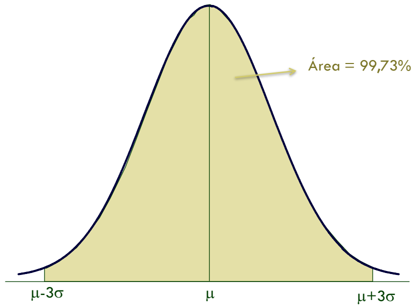
Grande parte $(\approx 95\%)$ da população de indivíduos sadios possui glicemia entre $\mu - 2\sigma = 90 - 10 = 80mg$ e $\mu + 2\sigma = 90 + 10 = 100mg$;

Praticamente toda população $(\approx 99.7\%)$ de indivíduos sadios possui glicemia entre $\mu - 3\sigma = 90 - 15 = 75mg$ e $\mu + 3\sigma = 90 + 15 = 105mg$;

A propabilidade que uma pessoa saudável tenha um valor de glicemia de jejum entre 90 e 95 é de aproximadamente 34%.







EXEMPLO

A idade da população de uma cidade é normalmente distribuída com média de 43 e desvio padrão 10. A cidade tem uma população de 5.000 habitantes. Quantos você esperaria ter entre 33 e 63 anos?

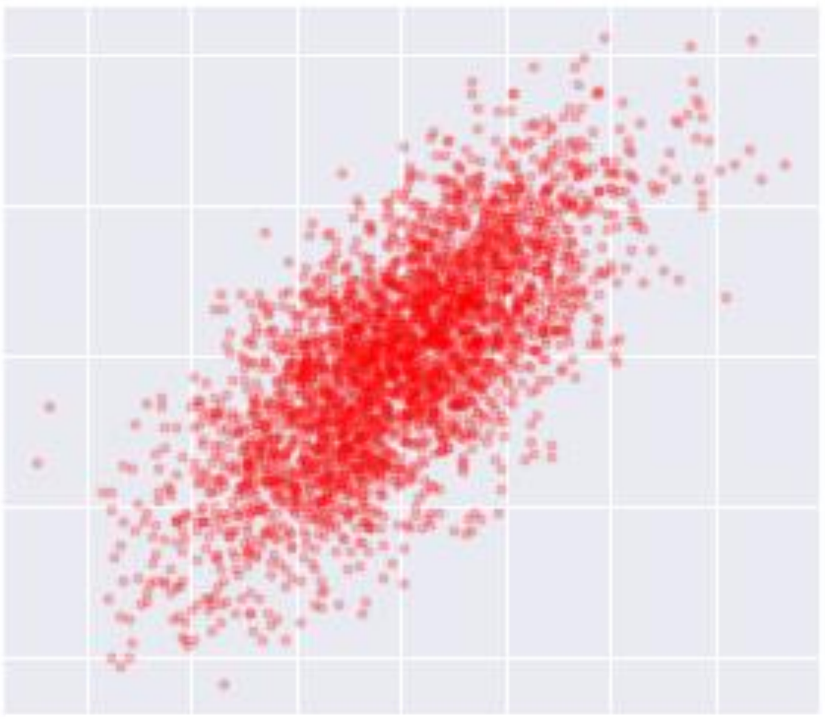
PECE Programa de Educação Continuada
Escola Politécnica da USP

DISTRIBUIÇÃO GAUSSIANA MULTIVARIADA

Vetor de dados de entrada com diversas características

PESO X ALTURA

$p(x_2; \mu_2, \sigma_2^2)$



$p(x_1; \mu_1, \sigma_1^2)$

DISTRIBUIÇÃO GAUSSIANA MULTIVARIADA

$$p(x;\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

A distribuição normal multivariada é uma generalização multidimensional da distribuição normal unidimensional que acabamos de ver. A distribuição normal multivariada é útil para analisar a relação entre n múltiplas variáveis normalmente distribuídas. A distribuição normal multivariável é dada por

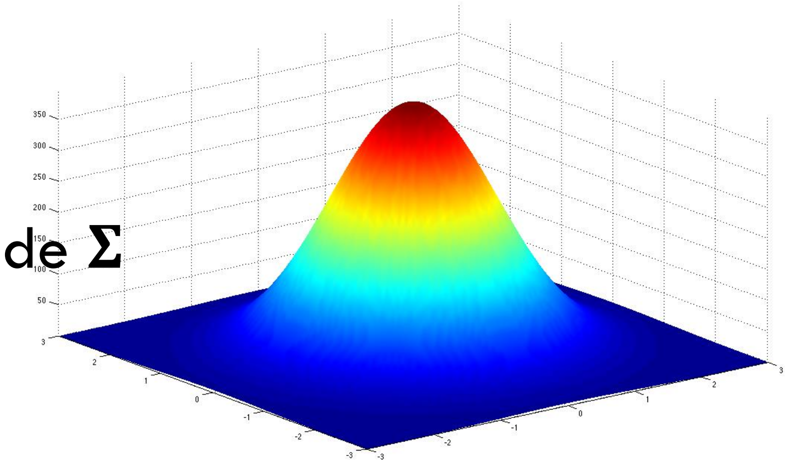
$$p(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}} e^{-\frac{1}{2}[(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})]}$$

$\boldsymbol{x}$ é um vetor de dimensão n

$\boldsymbol{\mu}$ vetor com as médias, $n \times 1$

$\boldsymbol{\Sigma}$ é a matriz de covariância, $n \times n$ e $|\boldsymbol{\Sigma}|$ é o determinante de $\boldsymbol{\Sigma}$

Dessa forma, tem-se que $\boldsymbol{x} \sim N(\boldsymbol{\mu},\boldsymbol{\Sigma})$.



MÁXIMA VEROSSIMILHANÇA PARA GAUSS MULTIVARIADO DE m OBSERVAÇÕES

função de log-verossimilhança negativa

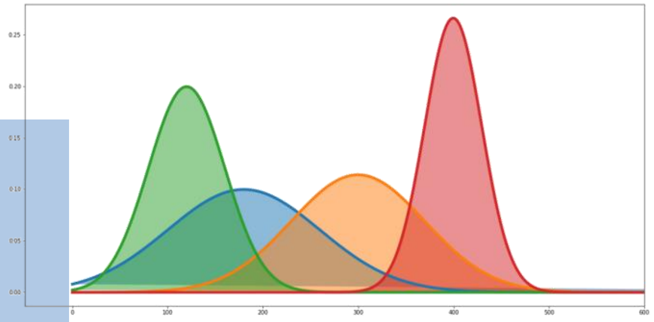
$$\mathcal{L}(\boldsymbol{\mu},\boldsymbol{\Sigma}) = \prod_{i=1}^{m} \frac{1}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}} e^{-\frac{1}{2}\left[(\boldsymbol{x}^{i}-\boldsymbol{\mu})^{T}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}^{i}-\boldsymbol{\mu})\right]}$$

$$\mathcal{L}(\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{nm/2}|\boldsymbol{\Sigma}|^{m/2}} e^{-\frac{1}{2}\sum_{i=0}^{m-1}\left[(\boldsymbol{x}^{i}-\boldsymbol{\mu})^{T}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}^{i}-\boldsymbol{\mu})\right]}$$

$$L(\boldsymbol{\mu},\boldsymbol{\Sigma}) = -\frac{m}{2}\ln|\boldsymbol{\Sigma}| - \frac{nm}{2}\ln 2\pi - \frac{1}{2}\sum_{i=1}^{m}\left[(\boldsymbol{x}^{i}-\boldsymbol{\mu})^{T}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}^{i}-\boldsymbol{\mu})\right]$$

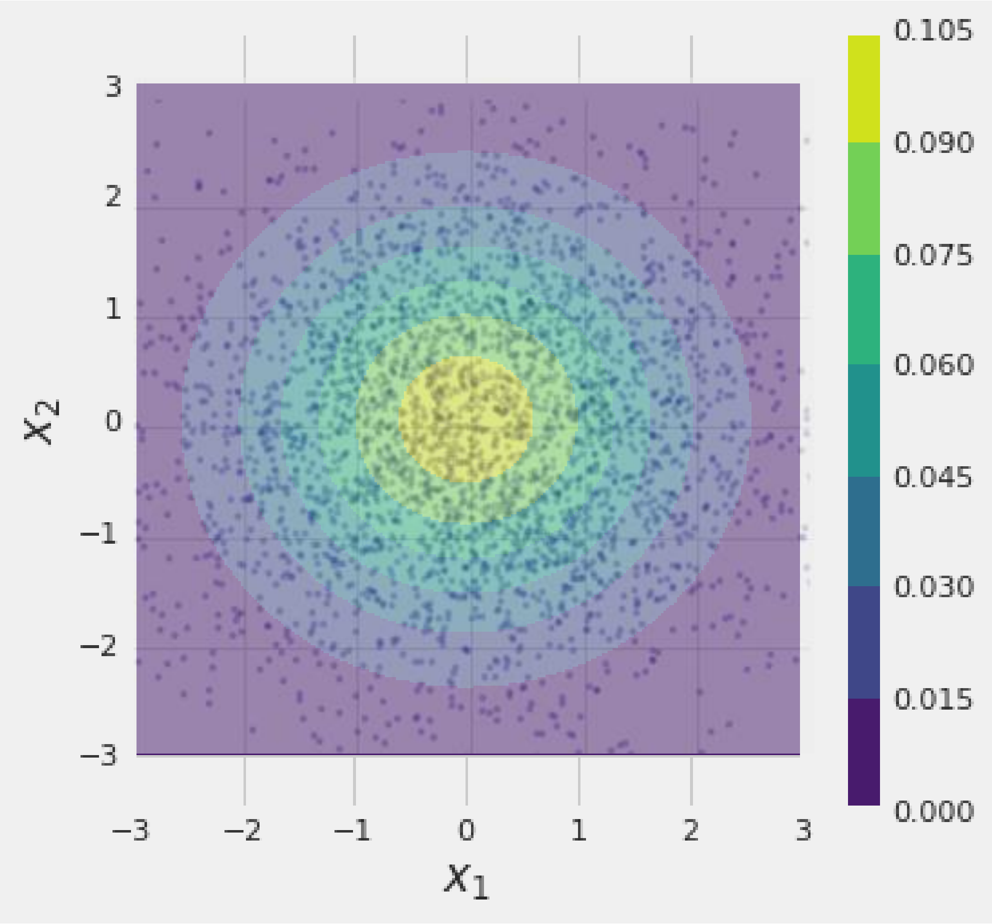
$$\frac{\partial L(\boldsymbol{\mu},\boldsymbol{\Sigma})}{\partial \boldsymbol{\mu}} = -\boldsymbol{\Sigma}^{-1}\sum_{i=1}^{m}(\boldsymbol{x}^{i}-\boldsymbol{\mu}) \quad \rightarrow \quad \widehat{\boldsymbol{\mu}} = \frac{1}{m}\sum_{i}\boldsymbol{x}^{i}$$

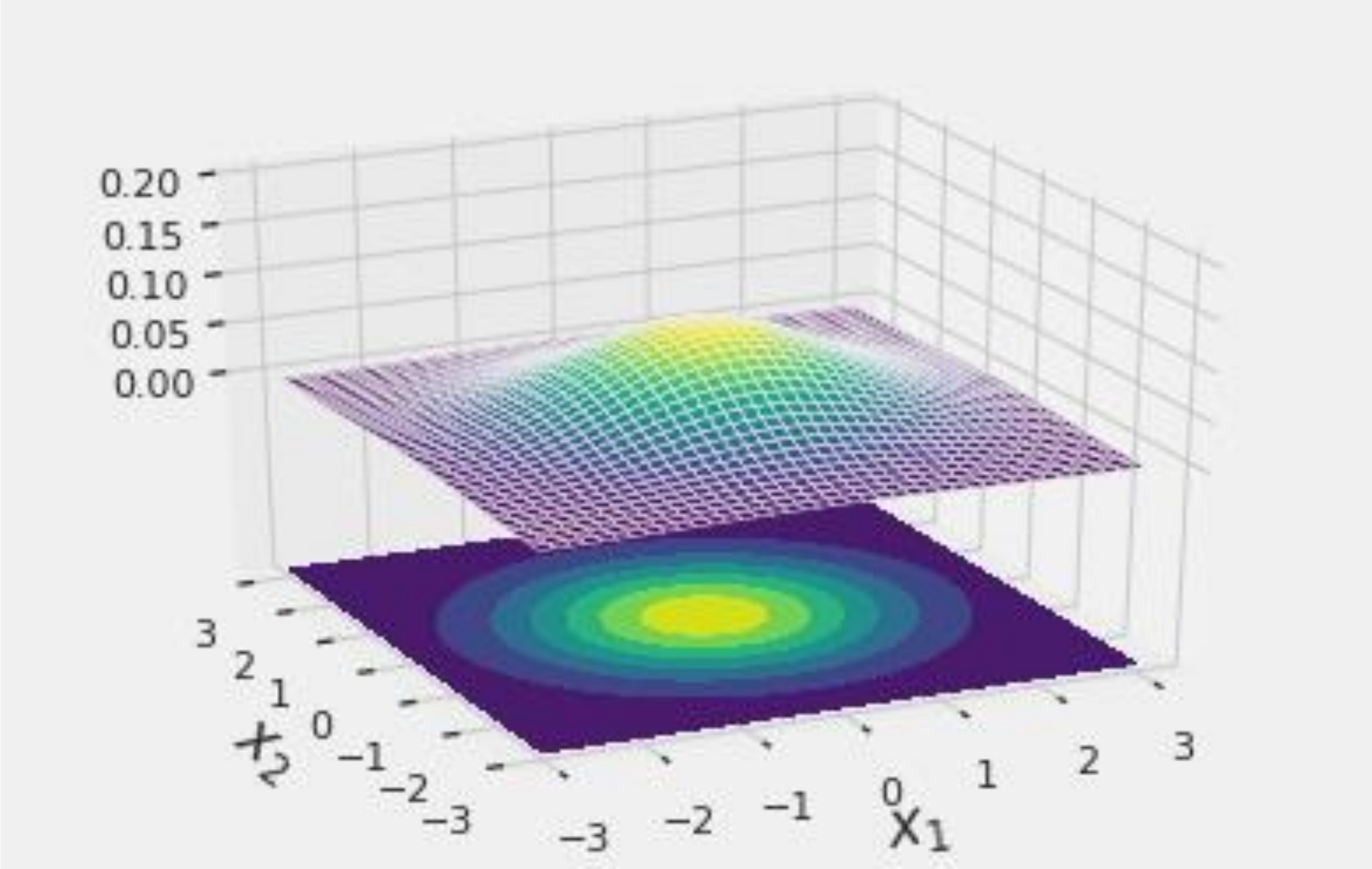
$$\frac{\partial L(\boldsymbol{\mu},\boldsymbol{\Sigma})}{\partial \boldsymbol{\Sigma}} = m - \boldsymbol{\Sigma}^{-1}\sum_{i=1}^{m}\left[(\boldsymbol{x}^{i}-\boldsymbol{\mu})^{T}(\boldsymbol{x}^{i}-\boldsymbol{\mu})\right] = \boldsymbol{0} \quad \rightarrow \quad \widehat{\boldsymbol{\Sigma}} = \frac{1}{m}\sum_{i}(\boldsymbol{x}^{i}-\boldsymbol{\mu})^{T}(\boldsymbol{x}^{i}-\boldsymbol{\mu})$$

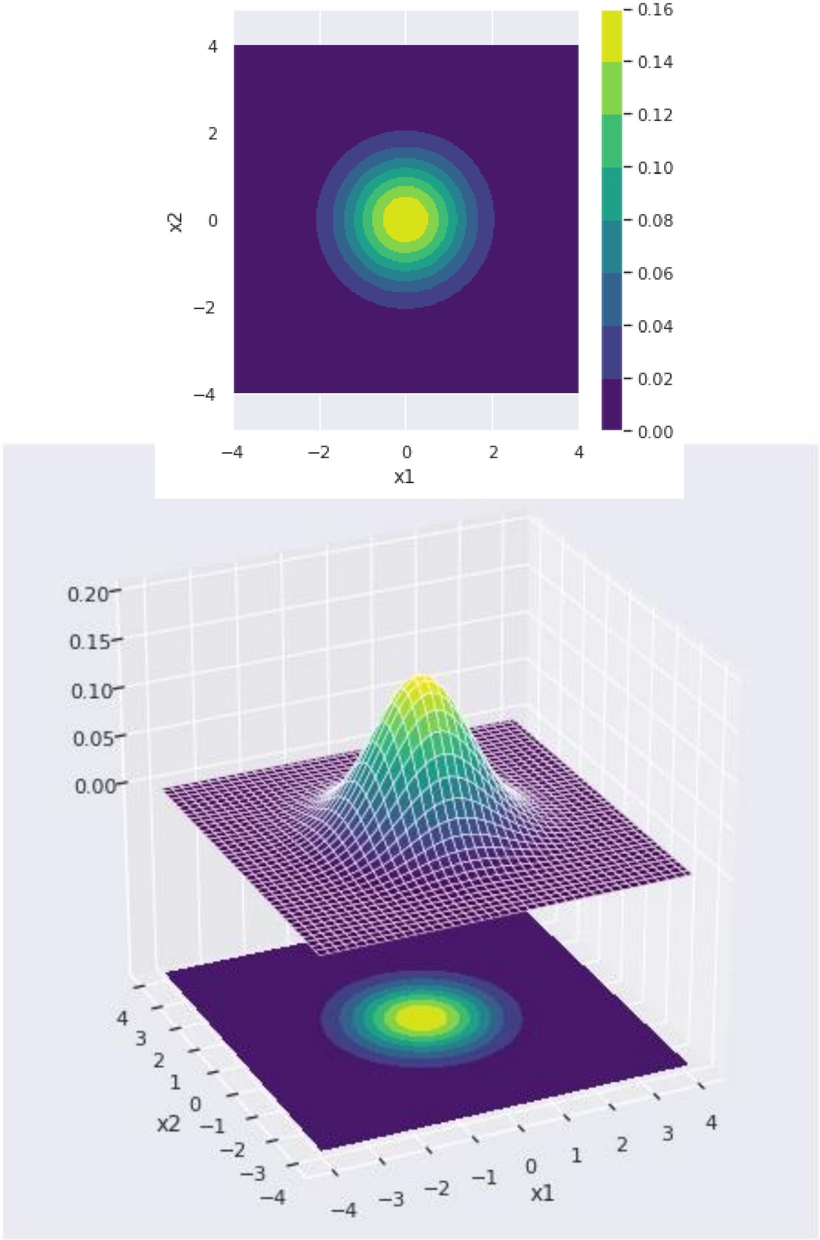


EXEMPLO

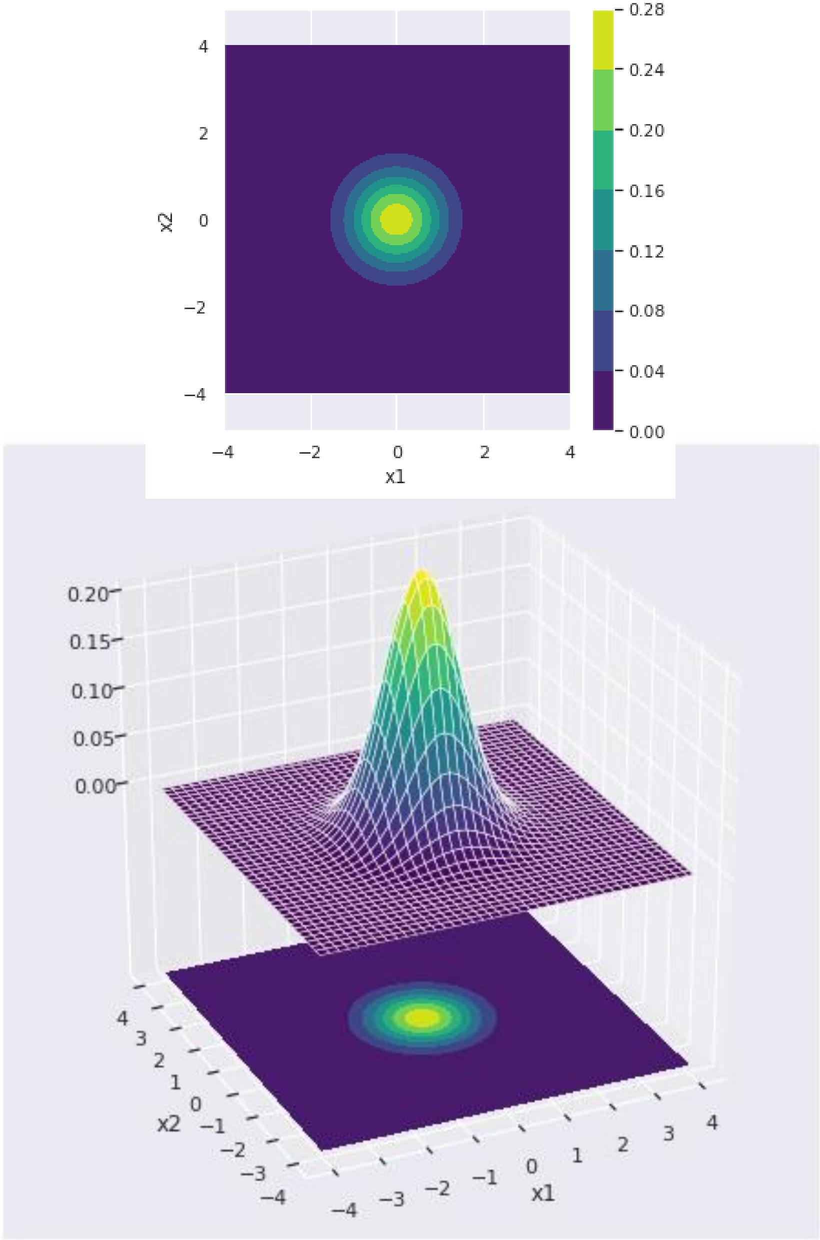
$$\boldsymbol{\mu} = \begin{pmatrix} 0. \\ 0. \end{pmatrix} \quad \boldsymbol{\Sigma} = \begin{pmatrix} 1.7 & 0. \\ 0. & 1.3 \end{pmatrix}$$



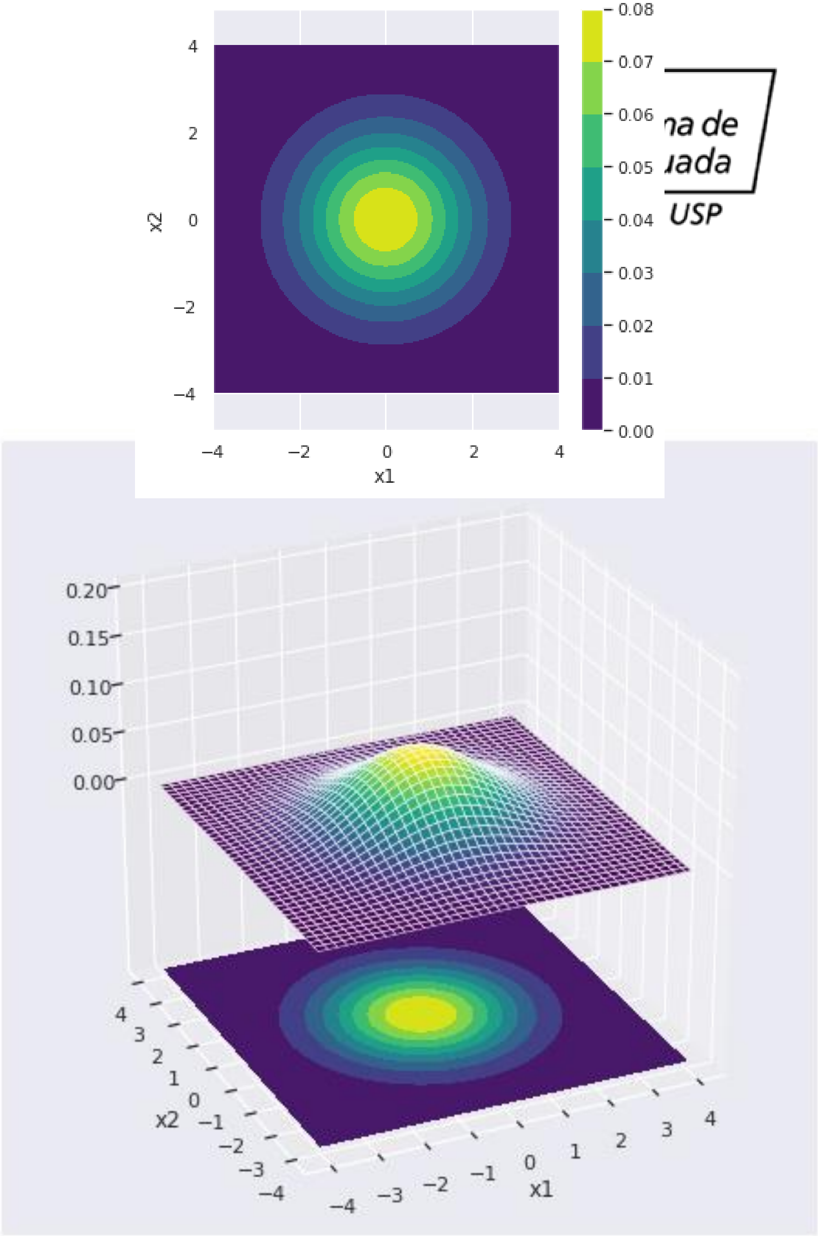




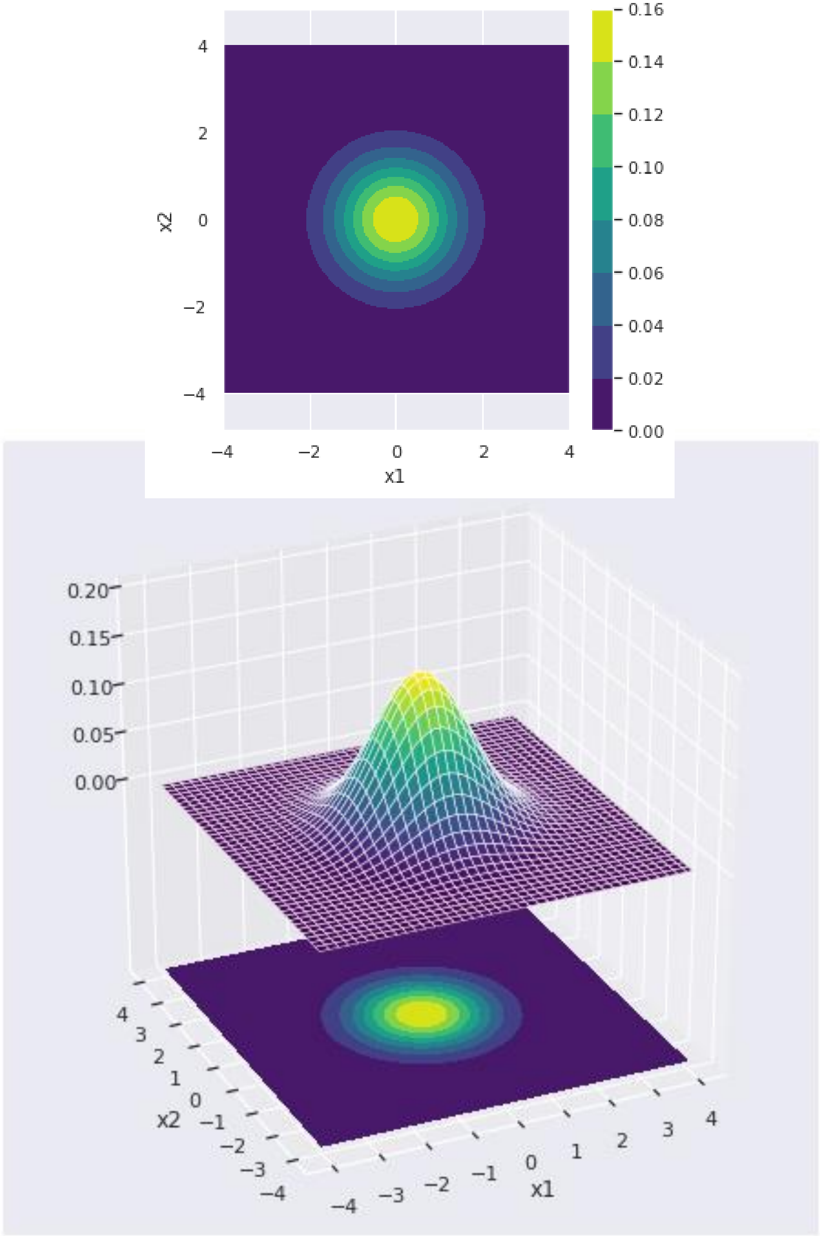
$$\boldsymbol{\mu} = \begin{pmatrix} 0. \\ 0. \end{pmatrix} \quad \boldsymbol{\Sigma} = \begin{pmatrix} 1. & 0. \\ 0. & 1. \end{pmatrix}$$



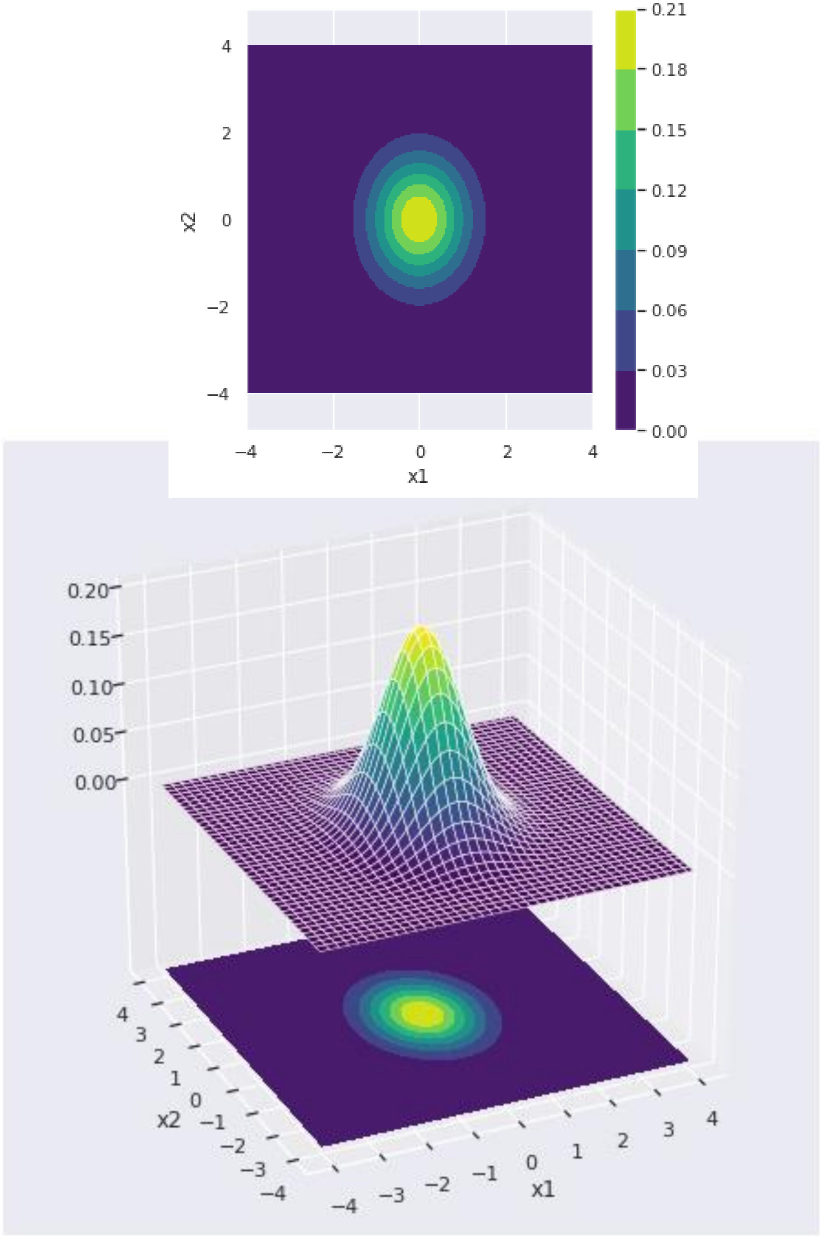
$$\boldsymbol{\mu} = \begin{pmatrix} 0. \\ 0. \end{pmatrix} \quad \boldsymbol{\Sigma} = \begin{pmatrix} 0.6 & 0. \\ 0. & 0.6 \end{pmatrix}$$



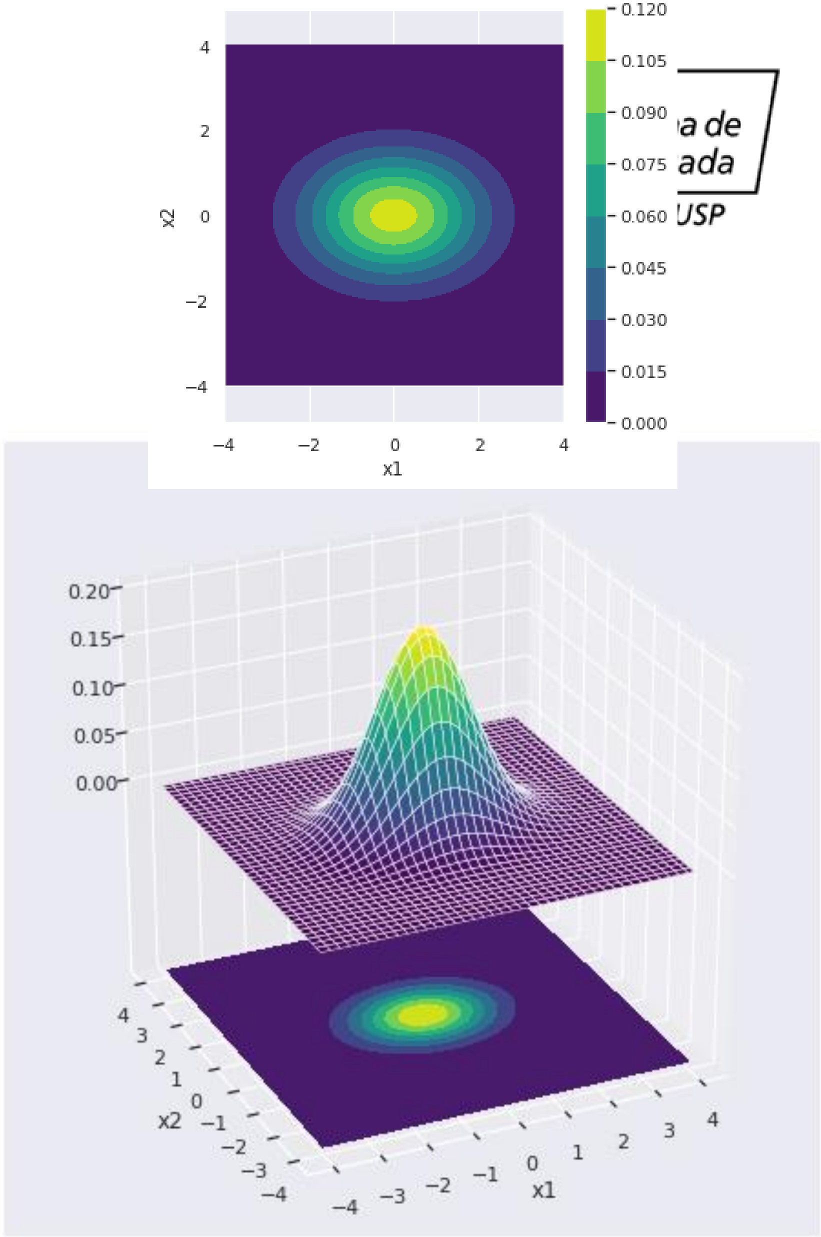
$$\boldsymbol{\mu} = \begin{pmatrix} 0. \\ 0. \end{pmatrix} \quad \boldsymbol{\Sigma} = \begin{pmatrix} 2. & 0. \\ 0. & 2. \end{pmatrix}$$



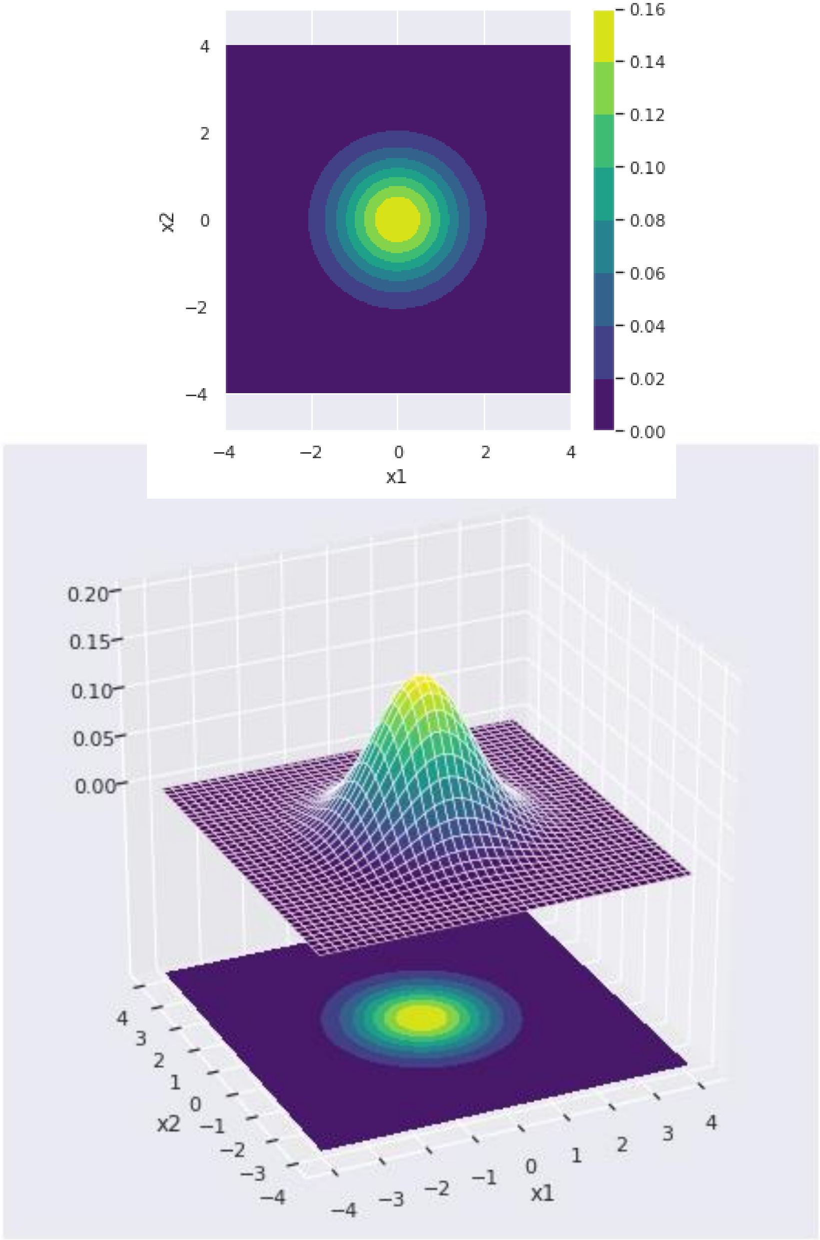
$$\boldsymbol{\mu} = \begin{pmatrix} 0. \\ 0. \end{pmatrix} \quad \boldsymbol{\Sigma} = \begin{pmatrix} 1. & 0. \\ 0. & 1. \end{pmatrix}$$



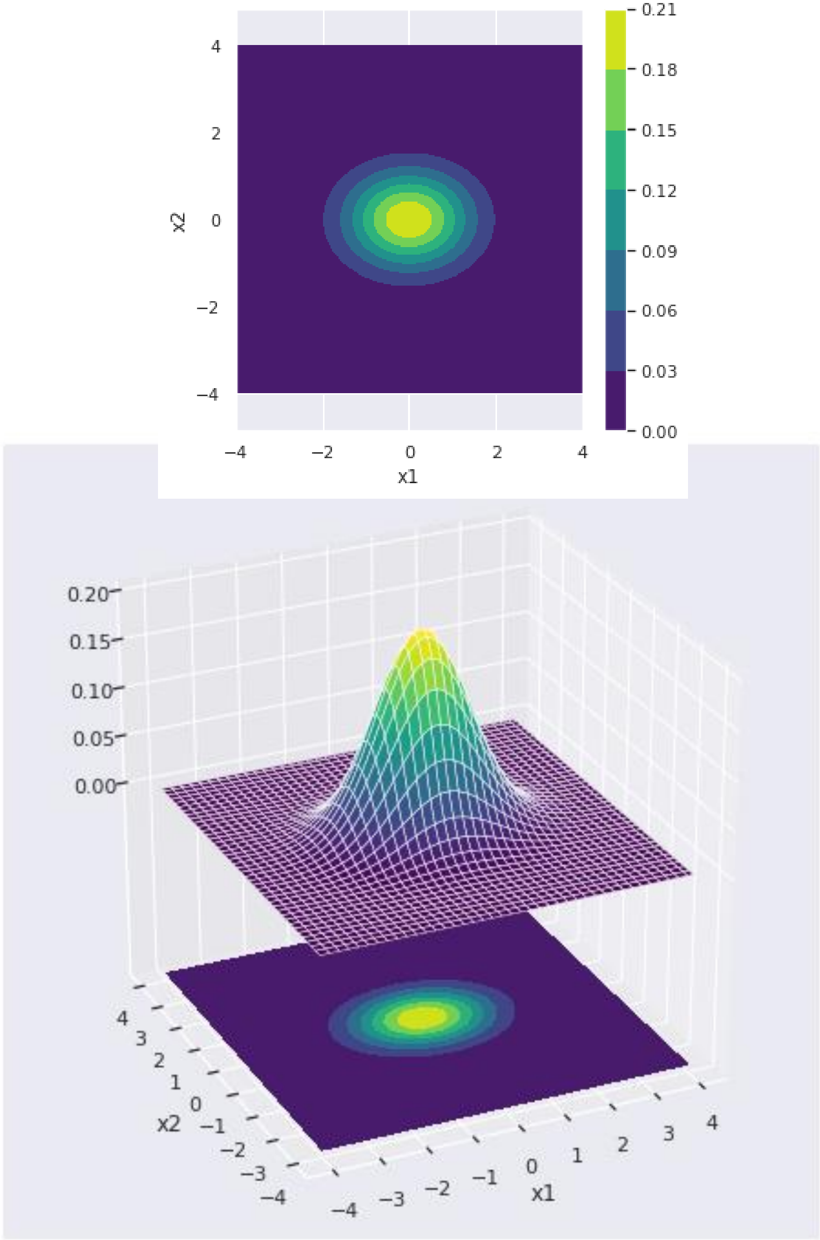
$$\boldsymbol{\mu} = \begin{pmatrix} 0. \\ 0. \end{pmatrix} \quad \boldsymbol{\Sigma} = \begin{pmatrix} 0.6 & 0. \\ 0. & 1. \end{pmatrix}$$



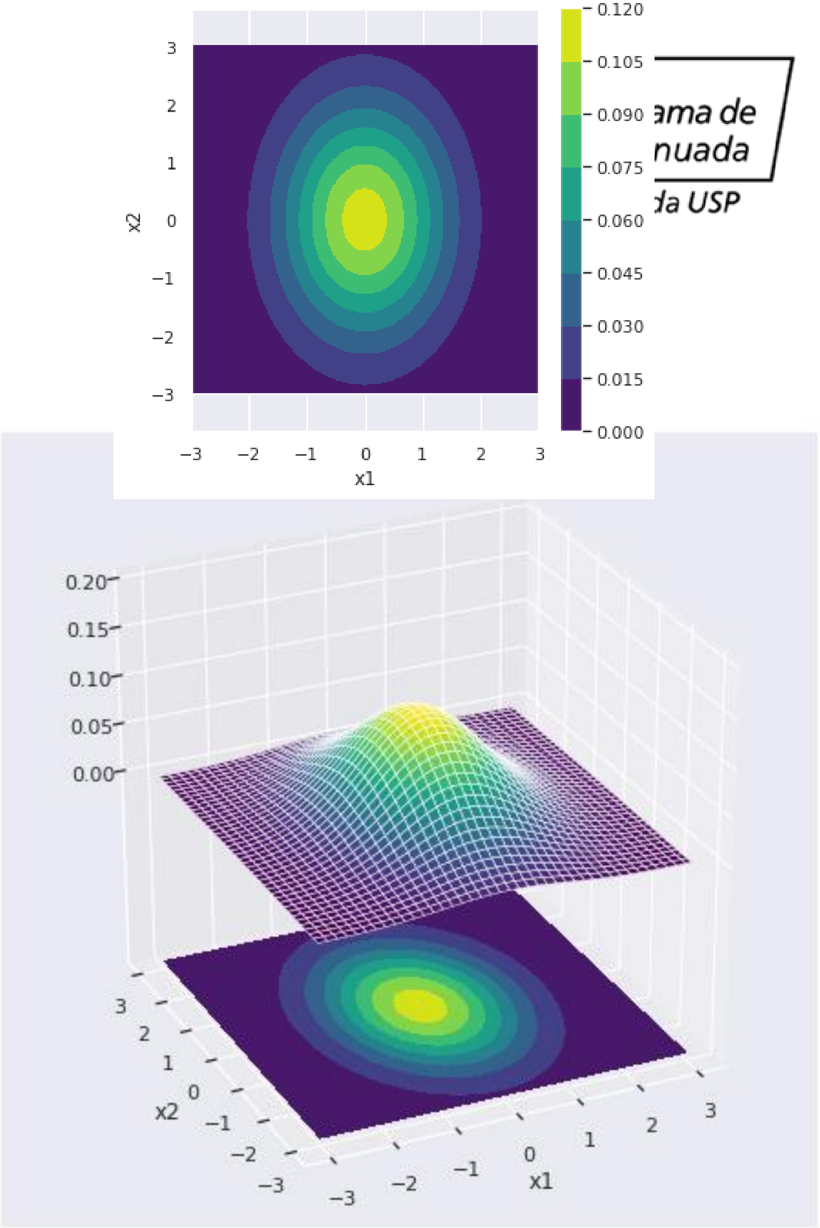
$$\boldsymbol{\mu} = \begin{pmatrix} 0. \\ 0. \end{pmatrix} \quad \boldsymbol{\Sigma} = \begin{pmatrix} 2. & 0. \\ 0. & 1. \end{pmatrix}$$



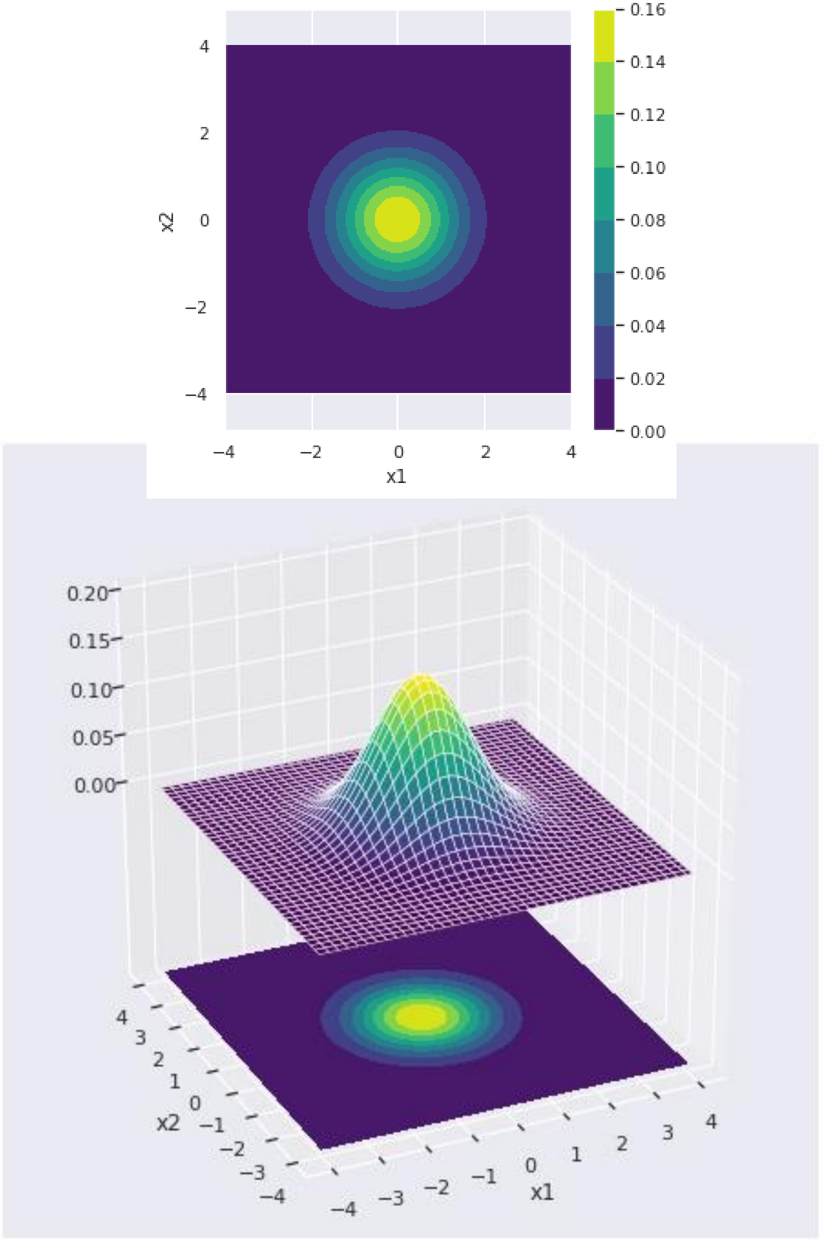
$$\boldsymbol{\mu} = \begin{pmatrix} 0. \\ 0. \end{pmatrix} \quad \boldsymbol{\Sigma} = \begin{pmatrix} 1. & 0. \\ 0. & 1. \end{pmatrix}$$



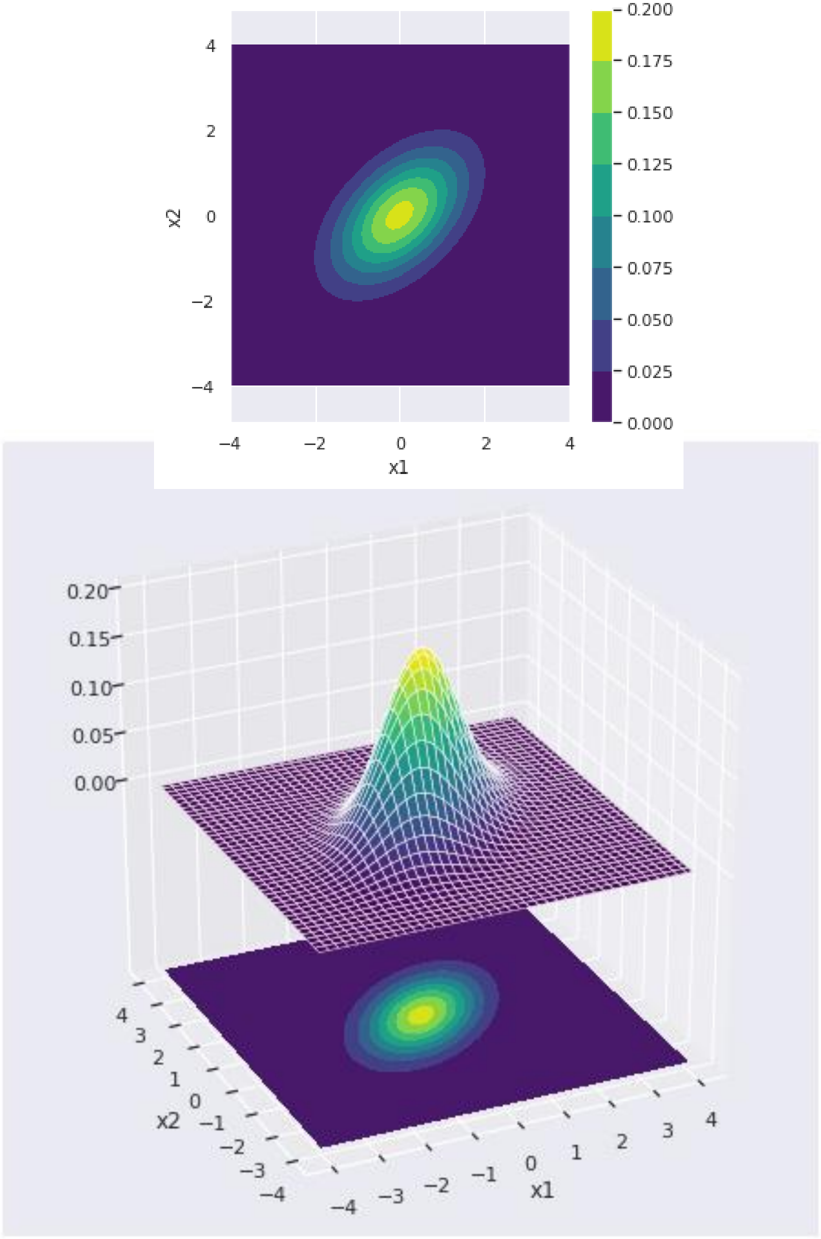
$$\boldsymbol{\mu} = \begin{pmatrix} 0. \\ 0. \end{pmatrix} \quad \boldsymbol{\Sigma} = \begin{pmatrix} 1. & 0. \\ 0. & 0.6 \end{pmatrix}$$



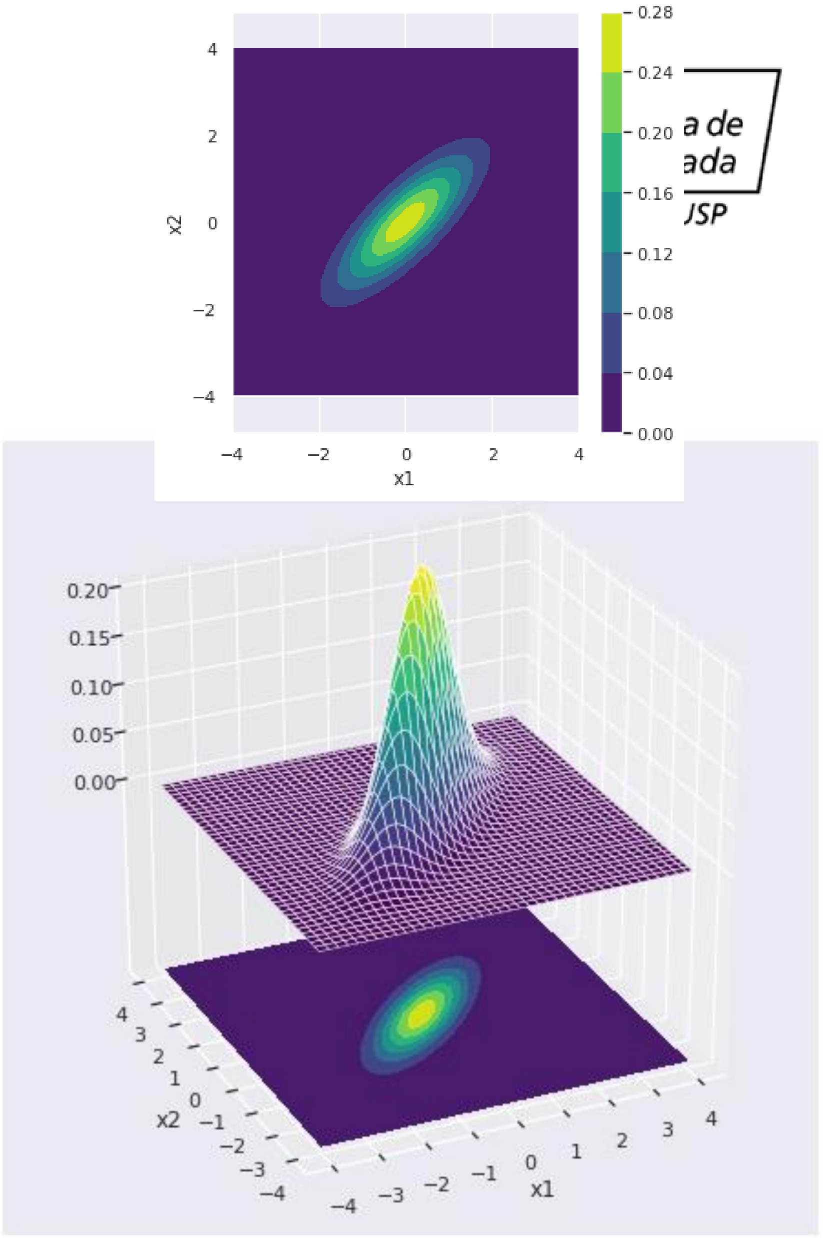
$$\boldsymbol{\mu} = \begin{pmatrix} 0. \\ 0. \end{pmatrix} \quad \boldsymbol{\Sigma} = \begin{pmatrix} 1. & 0. \\ 0. & 2. \end{pmatrix}$$



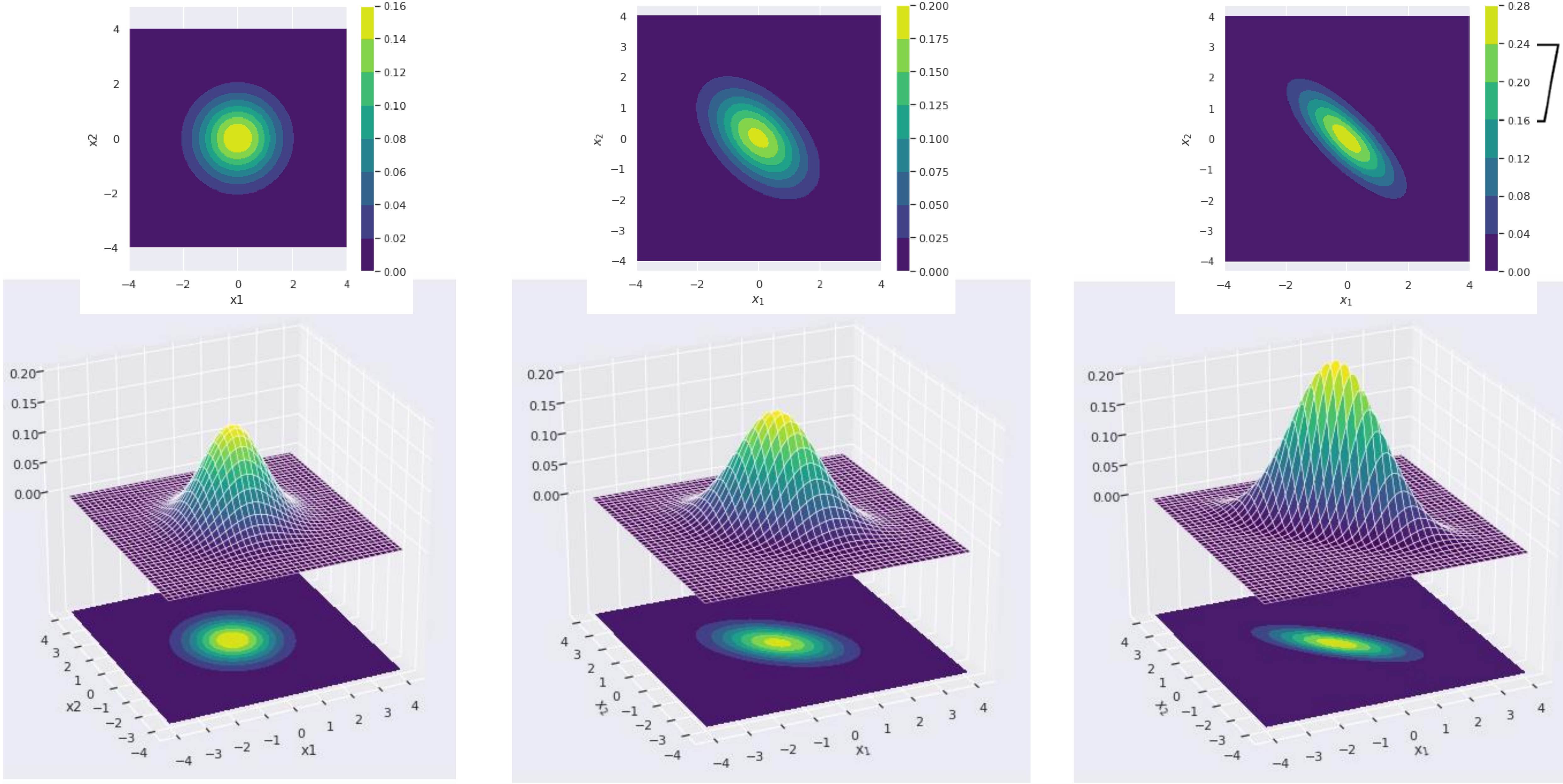
$$\boldsymbol{\mu} = \begin{pmatrix} 0. \\ 0. \end{pmatrix} \quad \boldsymbol{\Sigma} = \begin{pmatrix} 1. & 0. \\ 0. & 1. \end{pmatrix}$$



$$\boldsymbol{\mu} = \begin{pmatrix} 0. \\ 0. \end{pmatrix} \quad \boldsymbol{\Sigma} = \begin{pmatrix} 1. & 0.5 \\ 0.5 & 1. \end{pmatrix}$$



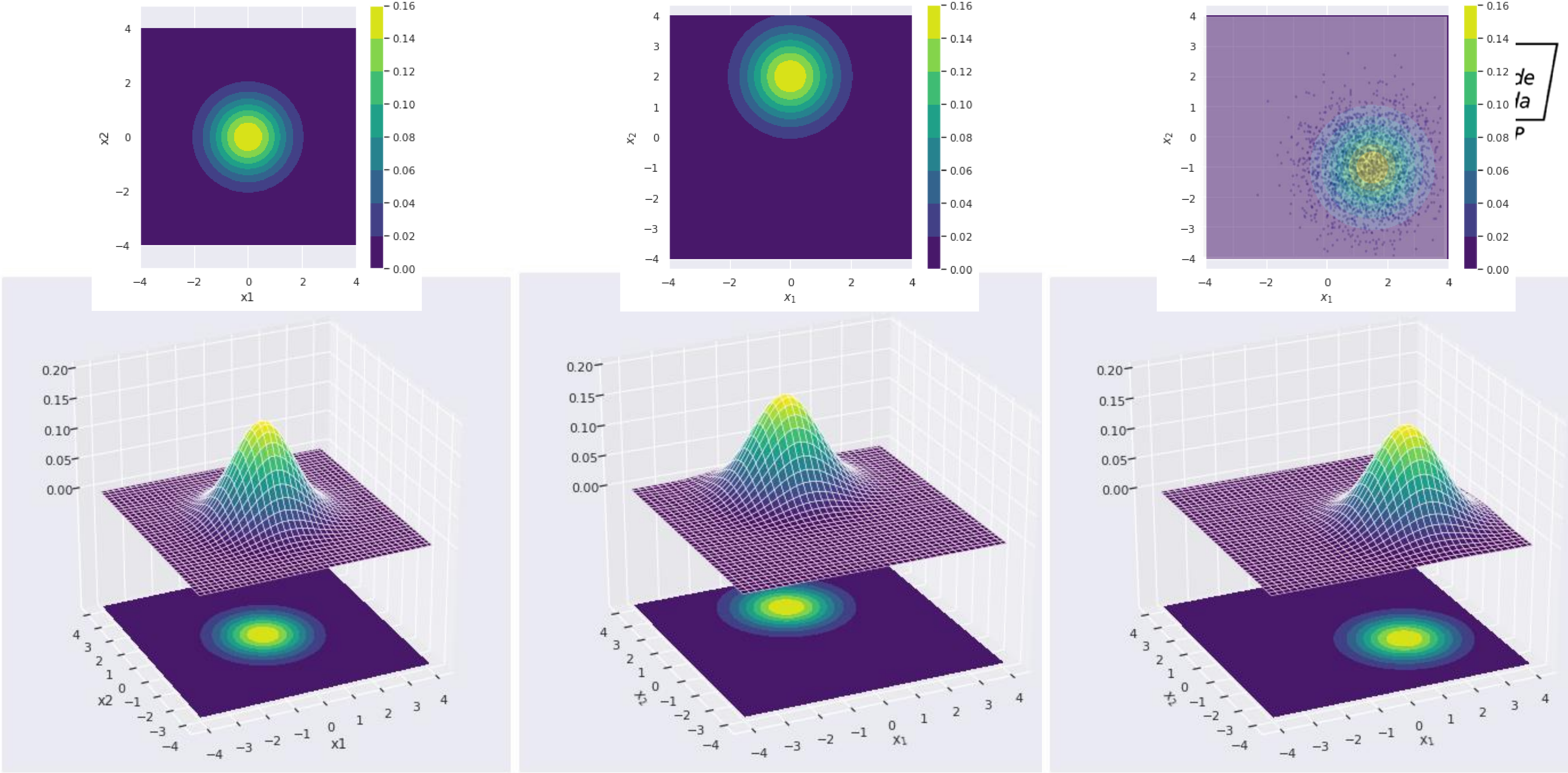
$$\boldsymbol{\mu} = \begin{pmatrix} 0. \\ 0. \end{pmatrix} \quad \boldsymbol{\Sigma} = \begin{pmatrix} 1. & 0.8 \\ 0.8 & 1. \end{pmatrix}$$



$\boldsymbol{\mu} = \begin{pmatrix} 0. \\ 0. \end{pmatrix} \quad \boldsymbol{\Sigma} = \begin{pmatrix} 1. & 0. \\ 0. & 1. \end{pmatrix}$

$\boldsymbol{\mu} = \begin{pmatrix} 0. \\ 0. \end{pmatrix} \quad \boldsymbol{\Sigma} = \begin{pmatrix} 1. & -0.5 \\ -0.5 & 1. \end{pmatrix}$

$\boldsymbol{\mu} = \begin{pmatrix} 0. \\ 0. \end{pmatrix} \quad \boldsymbol{\Sigma} = \begin{pmatrix} 1. & -0.8 \\ -0.8 & 1. \end{pmatrix}$



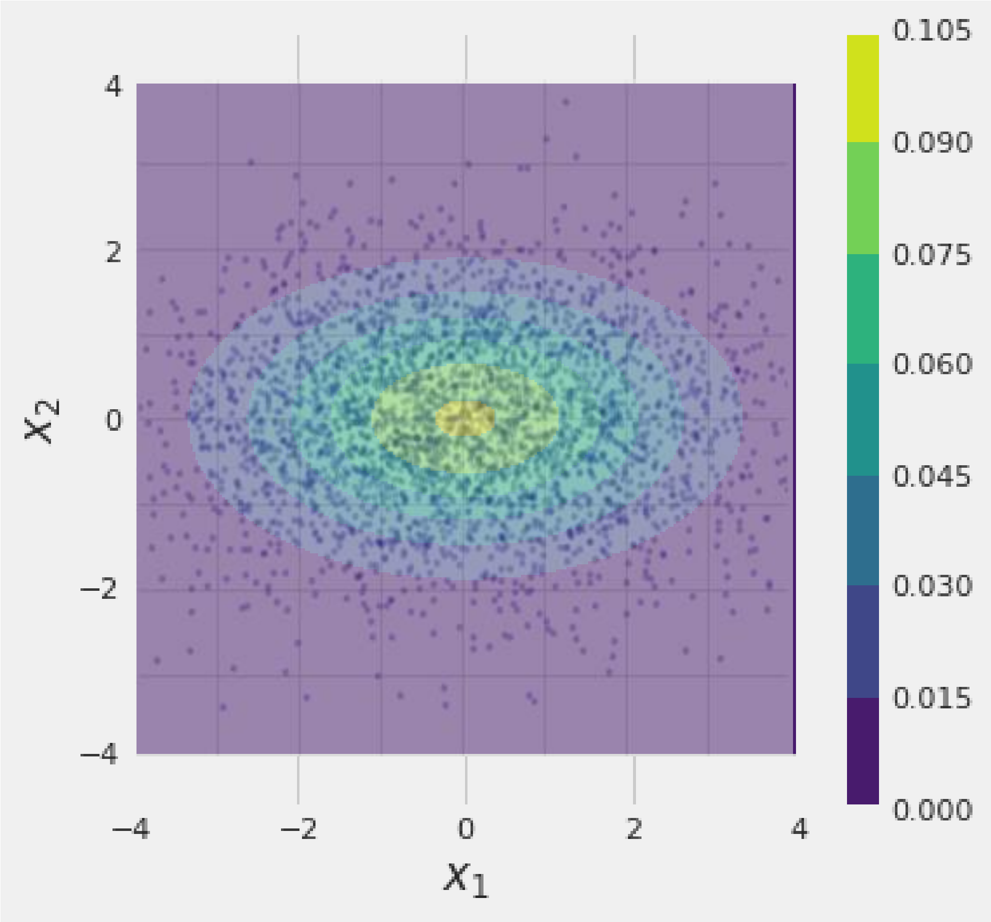
$$\boldsymbol{\mu} = \begin{pmatrix} 0. \\ 0. \end{pmatrix} \quad \boldsymbol{\Sigma} = \begin{pmatrix} 1. & 0. \\ 0. & 1. \end{pmatrix}$$

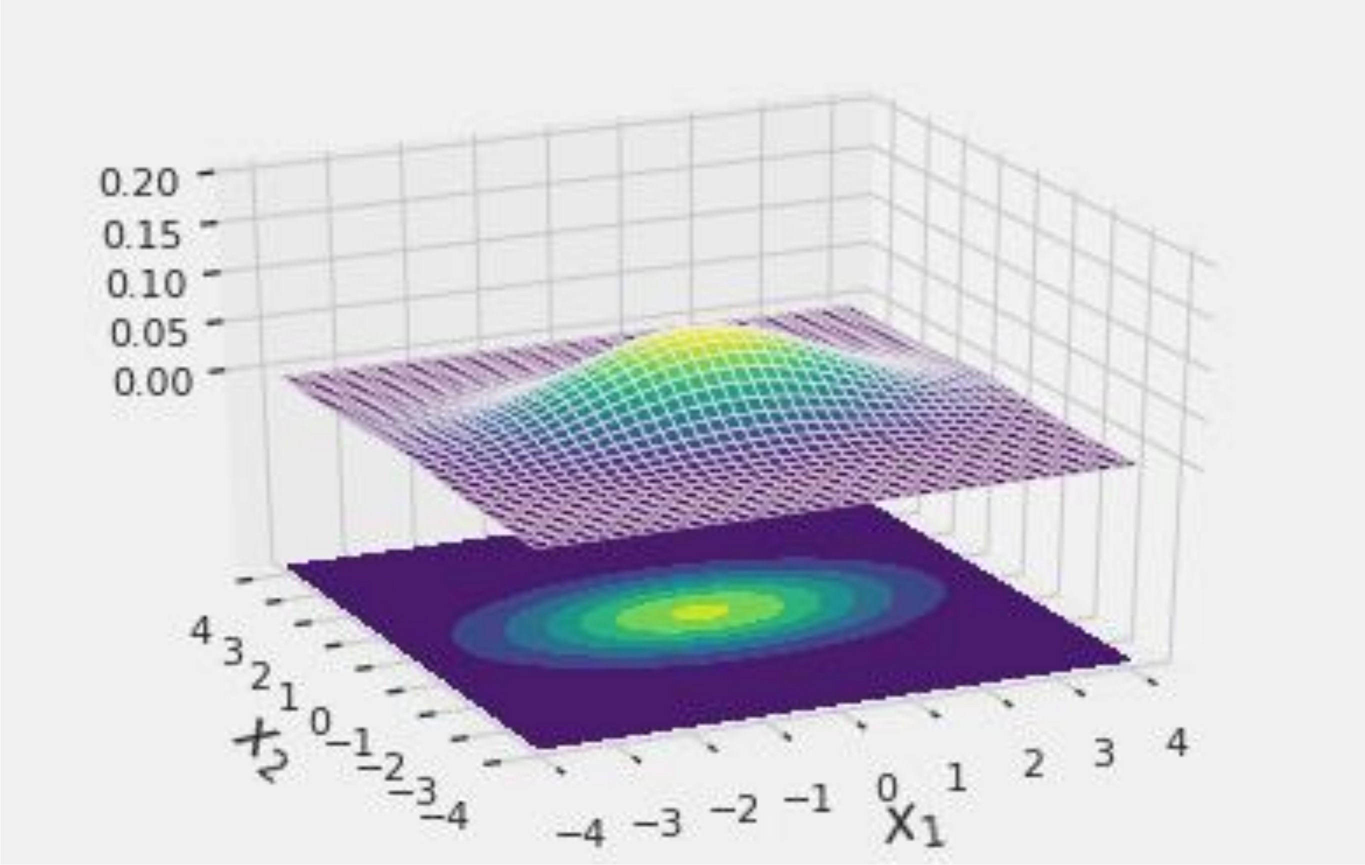
$$\boldsymbol{\mu} = \begin{pmatrix} 0. \\ 2.0 \end{pmatrix} \quad \boldsymbol{\Sigma} = \begin{pmatrix} 1. & 0. \\ 0. & 1. \end{pmatrix}$$

$$\boldsymbol{\mu} = \begin{pmatrix} 1.5 \\ -1.0 \end{pmatrix} \quad \boldsymbol{\Sigma} = \begin{pmatrix} 1. & 0. \\ 0. & 1. \end{pmatrix}$$

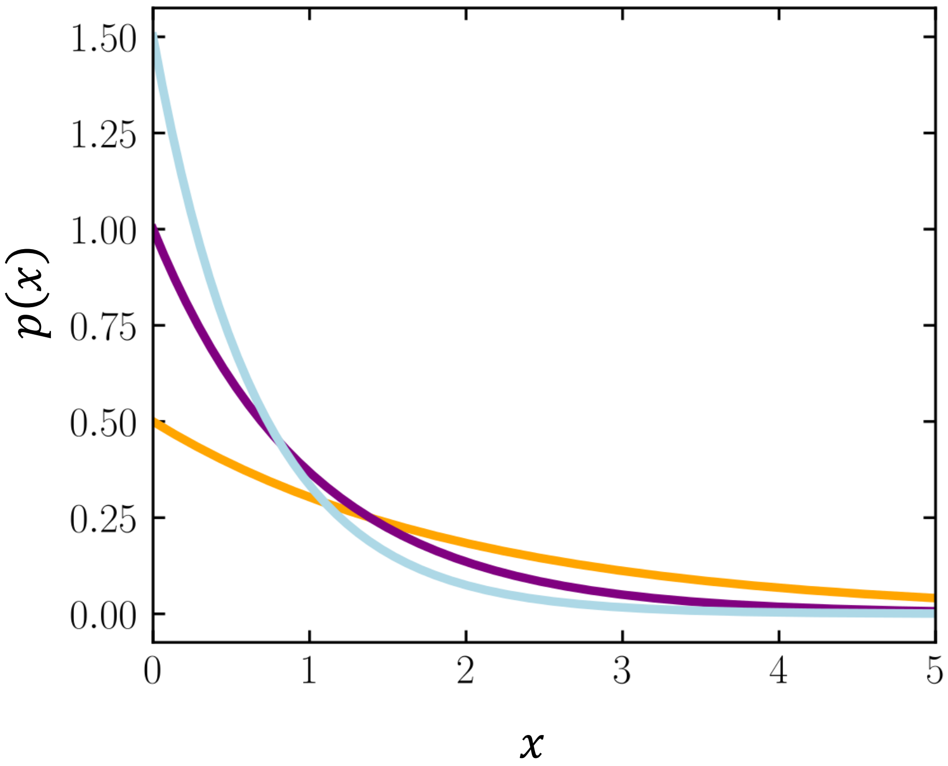
EXEMPLO

$$\boldsymbol{\mu} = \begin{pmatrix} 0. \\ 0. \end{pmatrix} \qquad \boldsymbol{\Sigma} = \begin{pmatrix} A & C \\ D & B \end{pmatrix}$$





PECE Programa de Educação Continuada
Escola Politécnica da USP



DISTRIBUIÇÃO EXPONENCIAL

$p(x;\lambda)$

DISTRIBUIÇÃO EXPONENCIAL

PECE Programa de Educação Continuada
Escola Politécnica da USP

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

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

Sendo comumente representada também por β,

$$\beta = \frac{1}{\lambda}$$

que é o intervalo (tempo ou espaço) médio entre eventos.

Frequentemente usada para calcular a quantidade de tempo até que um determinado evento ocorra.

Durante uma chuva, quantidade de tempo (a partir de agora) até que ocorra um raio, por exemplo, tem uma distribuição exponencial. Também a duração das chamadas telefônicas em minutos e o tempo de duração de uma bateria de carro em meses.

O valor do troco no bolso ou na bolsa de um conjunto de pessoas segue uma distribuição aproximadamente exponencial.

VALOR ESPERADO E VARIÂNCIA

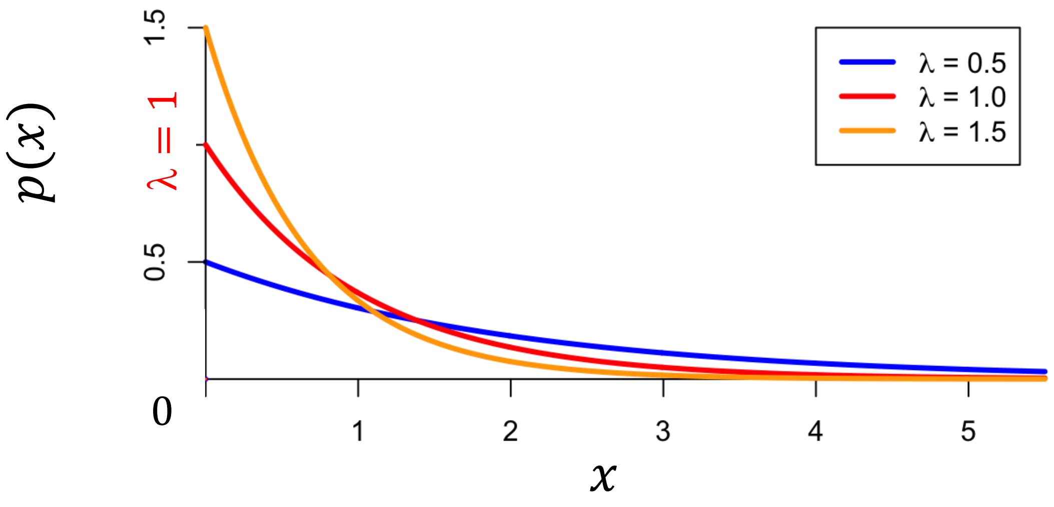
A média e variância são dados por,

$$\mu = \beta = \frac{1}{\lambda}$$

$$\sigma^2 = \beta^2 = \frac{1}{\lambda^2}$$

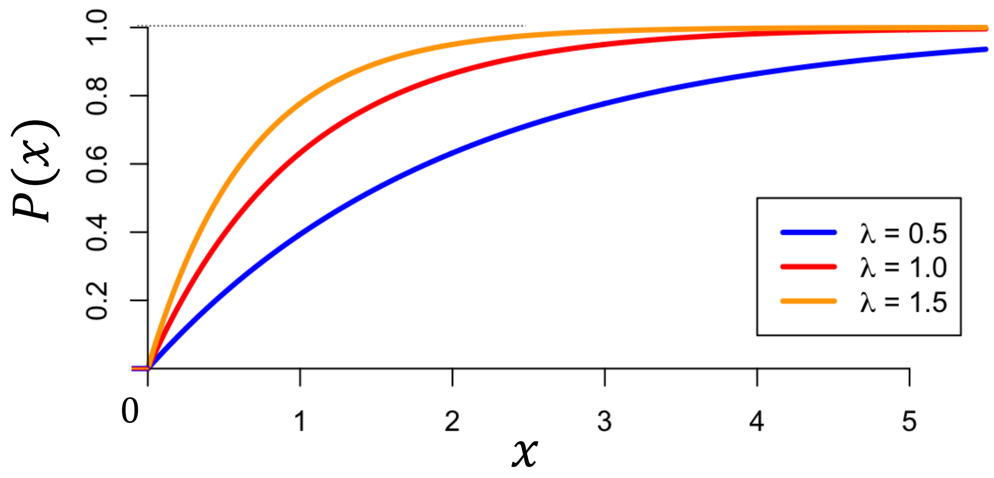
Função densidade de probabilidade da distribuição exponencial:

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



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

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

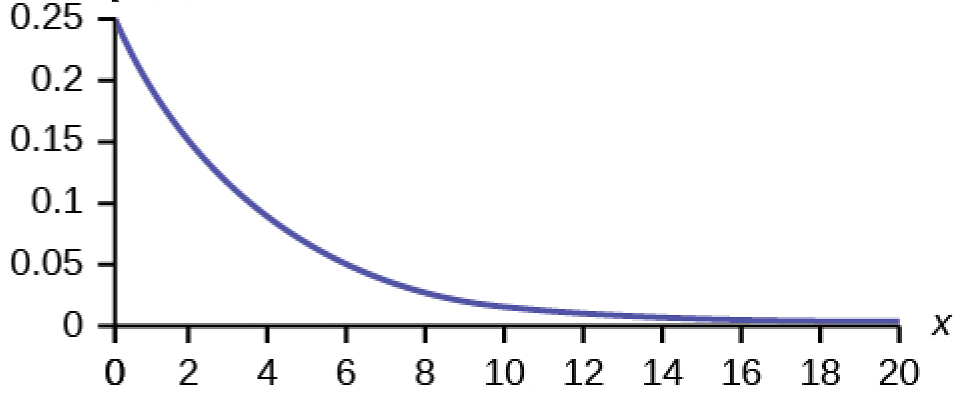


EXEMPLO

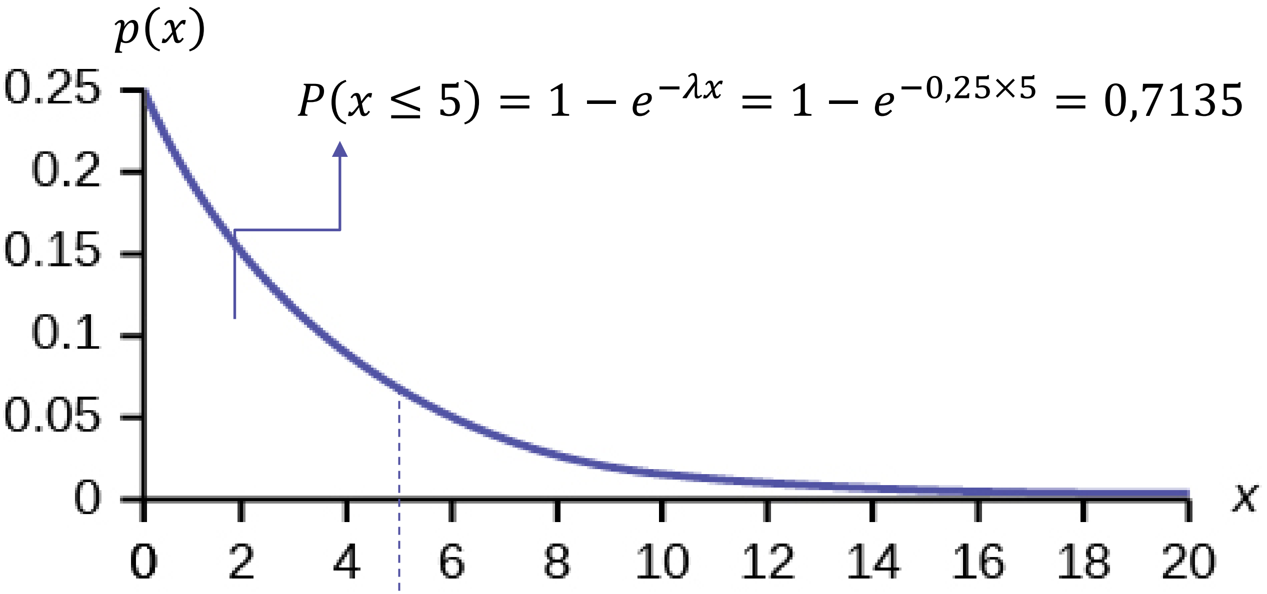
Seja x a quantidade de tempo (em minutos) que uma pessoa espera na fila de um banco. O tempo é conhecido por ter uma distribuição exponencial com a média igual a quatro minutos. Qual a probabilidade da pessoa esperar mais de 5 minutos na fila?

$$\mu = \beta = 4min \rightarrow \lambda = \frac{1}{4} = 0{,}25$$

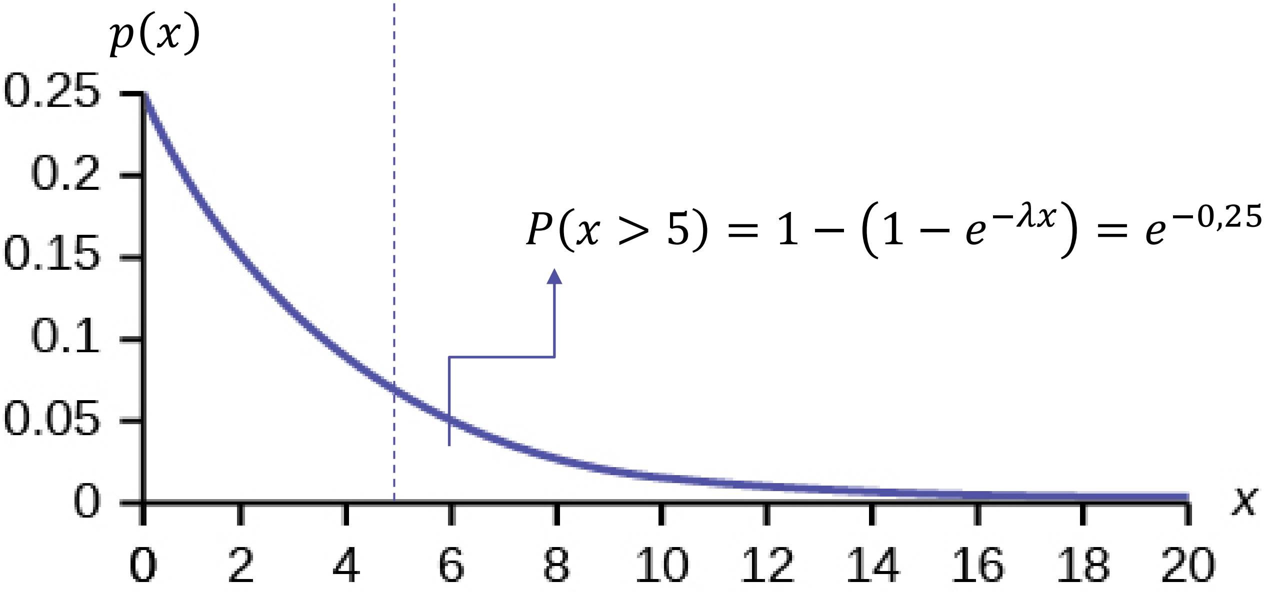
$$p(x) = 0{,}25e^{-0{,}25x}$$

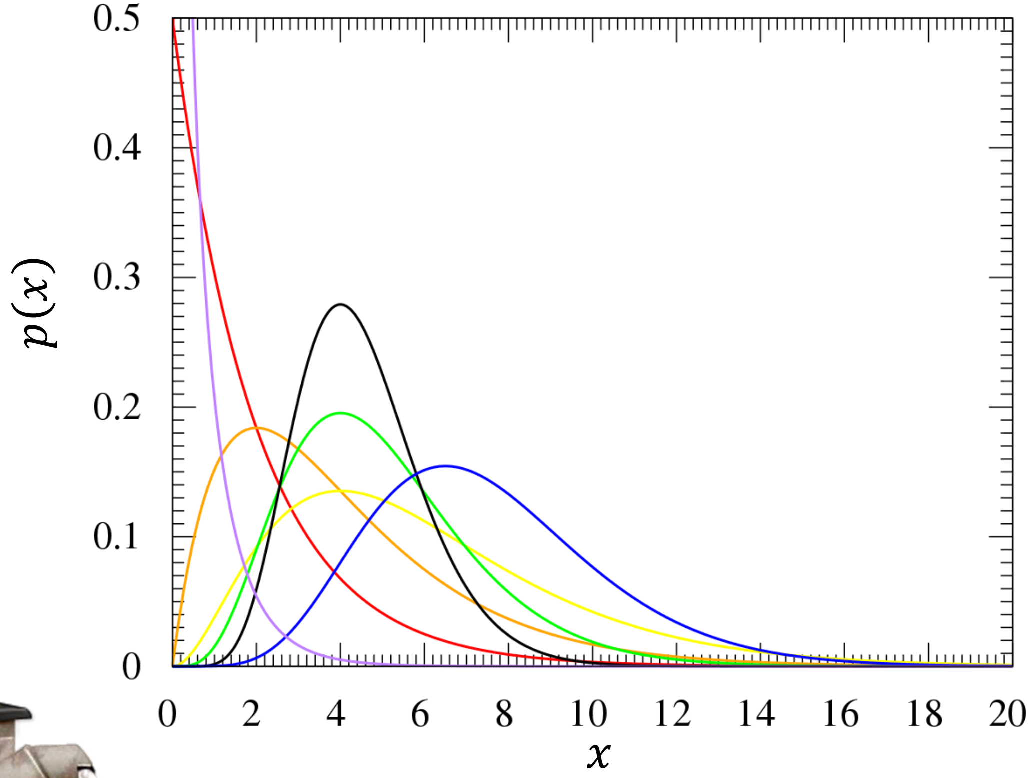


$$P(x \leq 5) = 1 - e^{-\lambda x} = 1 - e^{-0{,}25 \times 5} = 0{,}7135$$



$$P(x > 5) = 1 - \left(1 - e^{-\lambda x}\right) = e^{-0{,}25 \times 5} = \boxed{0{,}2865}$$





DISTRIBUIÇÃO GAMA

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

DISTRIBUIÇÃO GAMA

Uma variável aleatória x tem distribuição gama com parâmetros $\alpha > 0,\ \beta = \lambda > 0$ se sua função densidade de probabilidade é dada por $x \sim Gama(\alpha, \beta)$,

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

Onde, $\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1} e^{-x} dx$. Se α é um número natural, $\Gamma(\alpha) = (\alpha - 1)!$

$\beta = 1/\lambda$ e α é o número de eventos. Isto é, se a distribuição Exponencial refere-se ao espaço/tempo decorrido até a ocorrência do primeiro evento, Gama é o tempo/espaço até o α–ésimo evento.

Quando $\alpha = 1$, Gama se reduz à distribuição exponencial.

MÉDIA E VARIÂNCIA

PECE Programa de Educação Continuada
Escola Politécnica da USP

A média e variância são dados por,

$$\mu = \alpha\beta = \frac{\alpha}{\lambda}$$

$$\sigma^2 = \alpha\beta^2 = \frac{\alpha}{\lambda^2}$$

EXEMPLO

Você foi ao Bandejão e entrou em uma fila com duas pessoas à sua frente (milagres acontecem!). Um está sendo servido e o outro está esperando. Seus tempos de atendimento são variáveis aleatórias exponenciais independentes com média de 2 minutos.

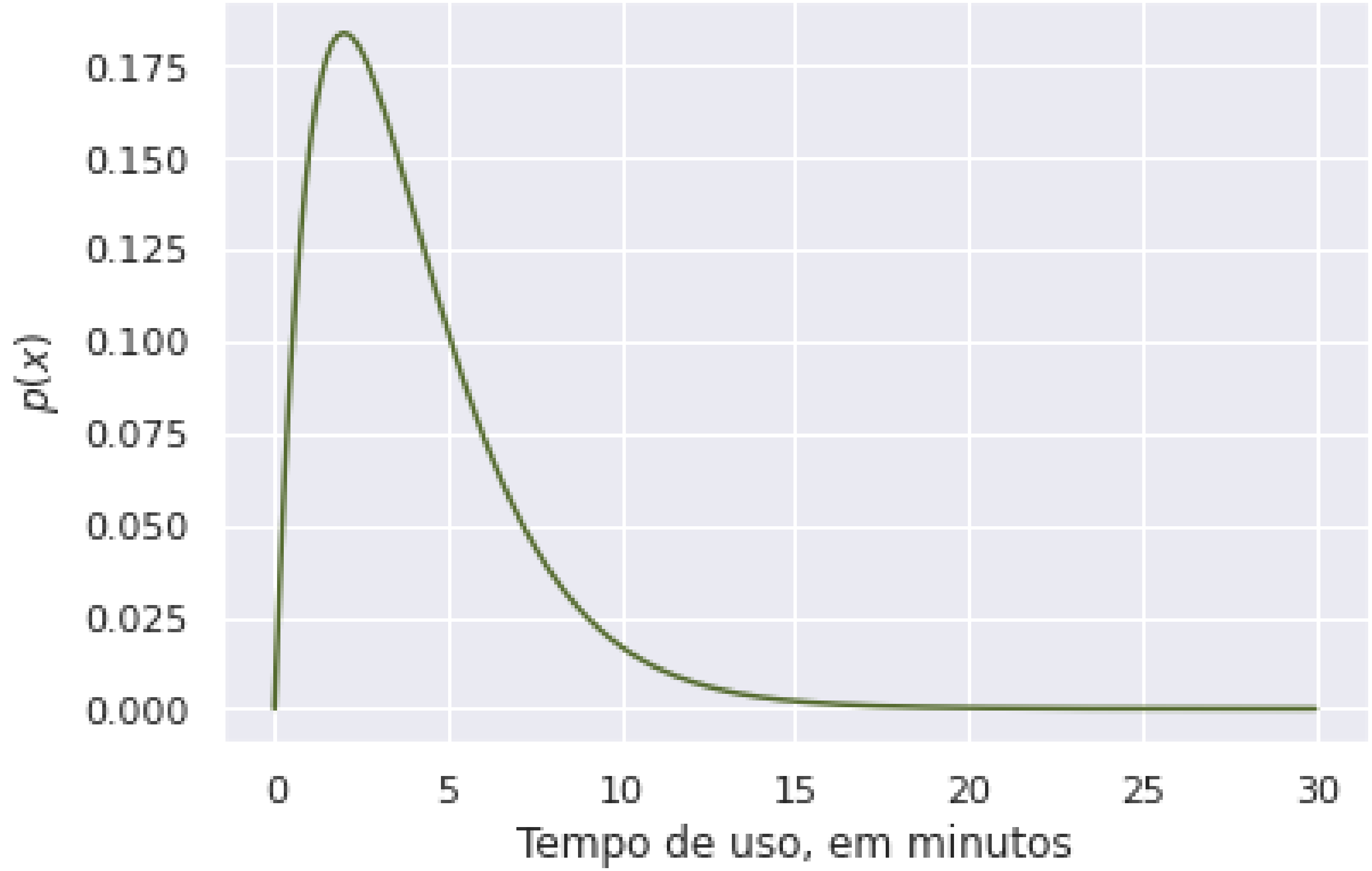
Pede-se,

1. Quais os valores de α e β?
2. Qual é a probabilidade de você esperar mais de 5 minutos na fila? (use Python)
3. Plote a distribuição Gama do problema.

$$\alpha = \beta = 2$$
$$t = 5 \text{ min}$$

```
#Item 2
prob = 1.0 - gamma.cdf(tempo, a=alpha, loc=0, scale=beta)
```

```
2. A probabilidade de
esperar mais de 5 minutos na
fila é: 28.730%
```



IMPORTANTE

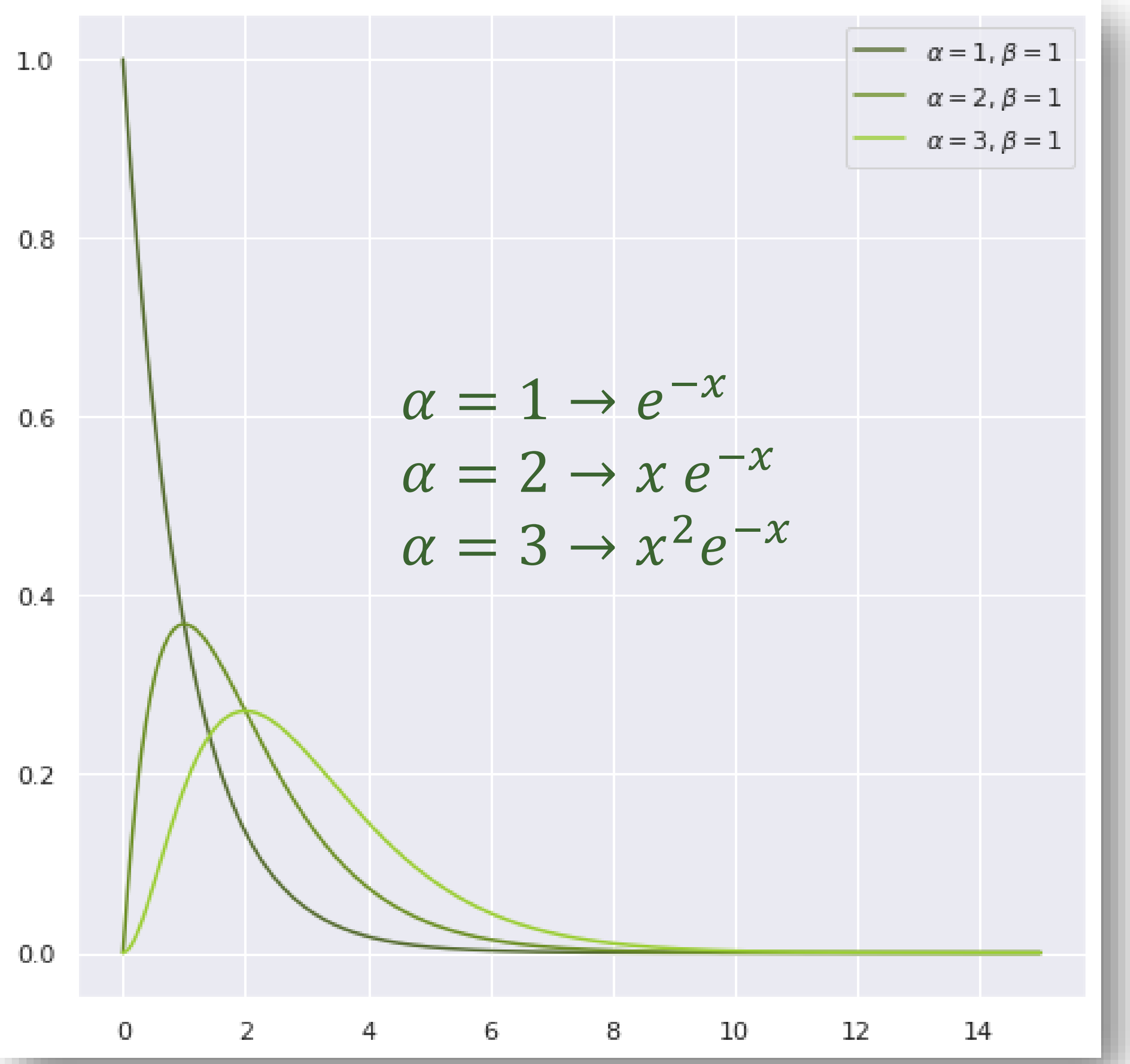
Se você quer dizer que há um fenômeno interessante que segue uma distribuição gama geral, não há nenhum. Existem casos especiais da família gama - as distribuições exponencial, qui-quadrado e Erlang - que são aproximações razoáveis de alguns fenômenos.

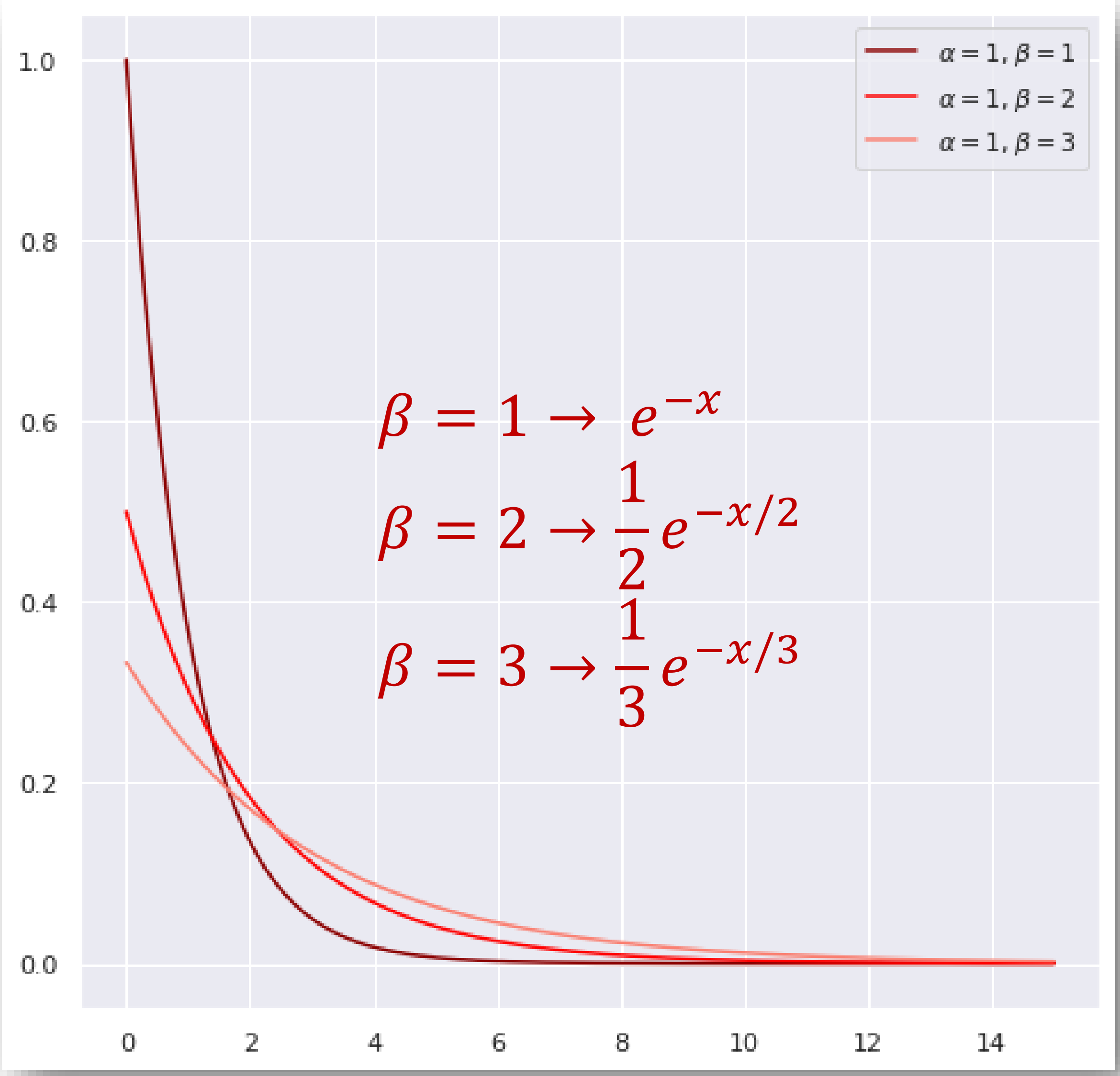
O principal uso da distribuição gama geral é na modelagem. Você pode escolher parâmetros para selecionar uma ampla faixa de distribuições em forma de sino ou exponencial aproximadamente assimétricas em [0,∞) que se aproximam de alguns dados ou fenômenos.

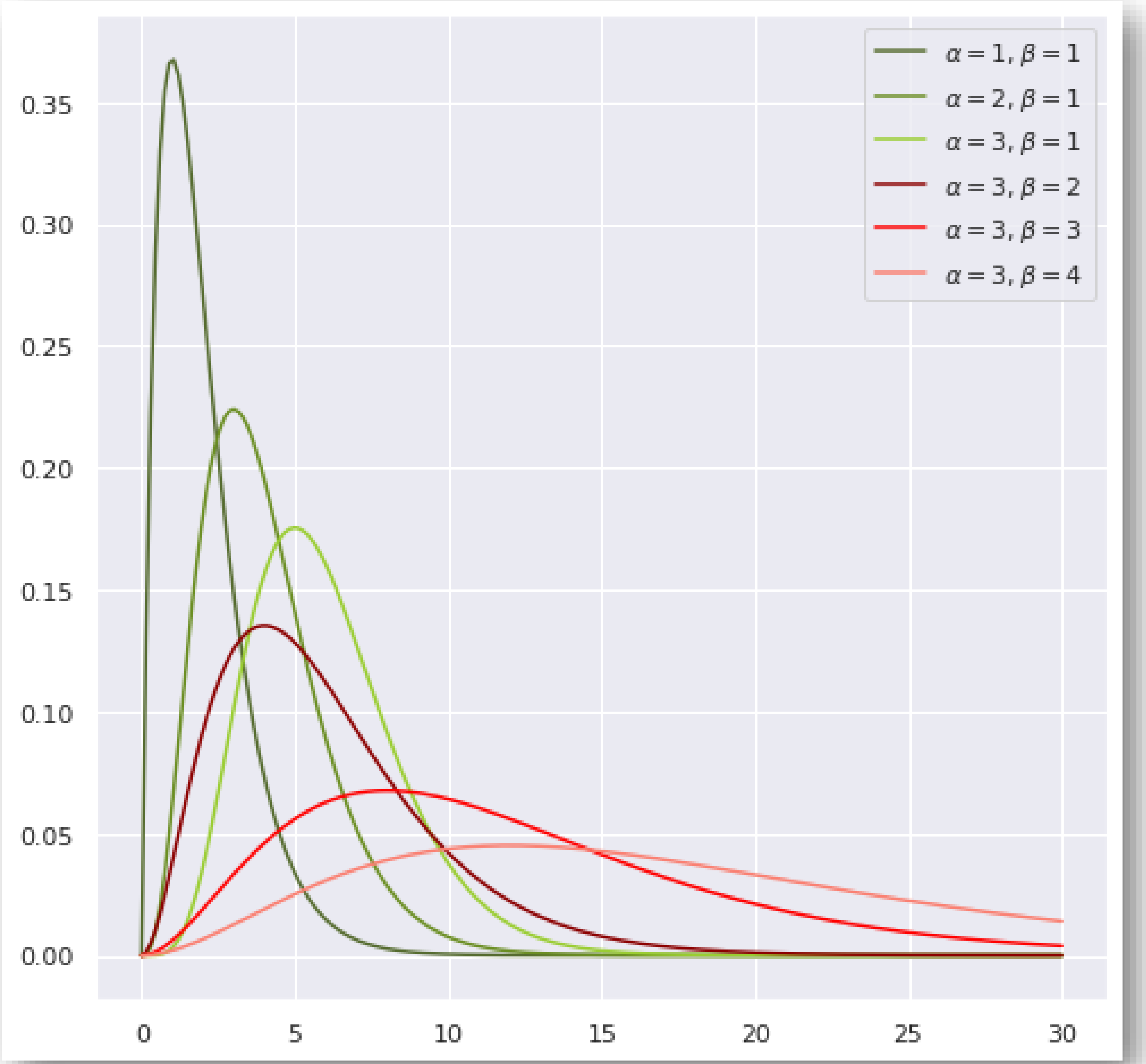
Quando α é um número inteiro, Gama se reduz à distribuição Erlang e quando $\alpha = 1$ tem-se a distribuição exponencial. Quando $\alpha = n/2$, n inteiro e $\beta = 1/2$, tem-se distribução qui-quadrado com n graus de liberdade.

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

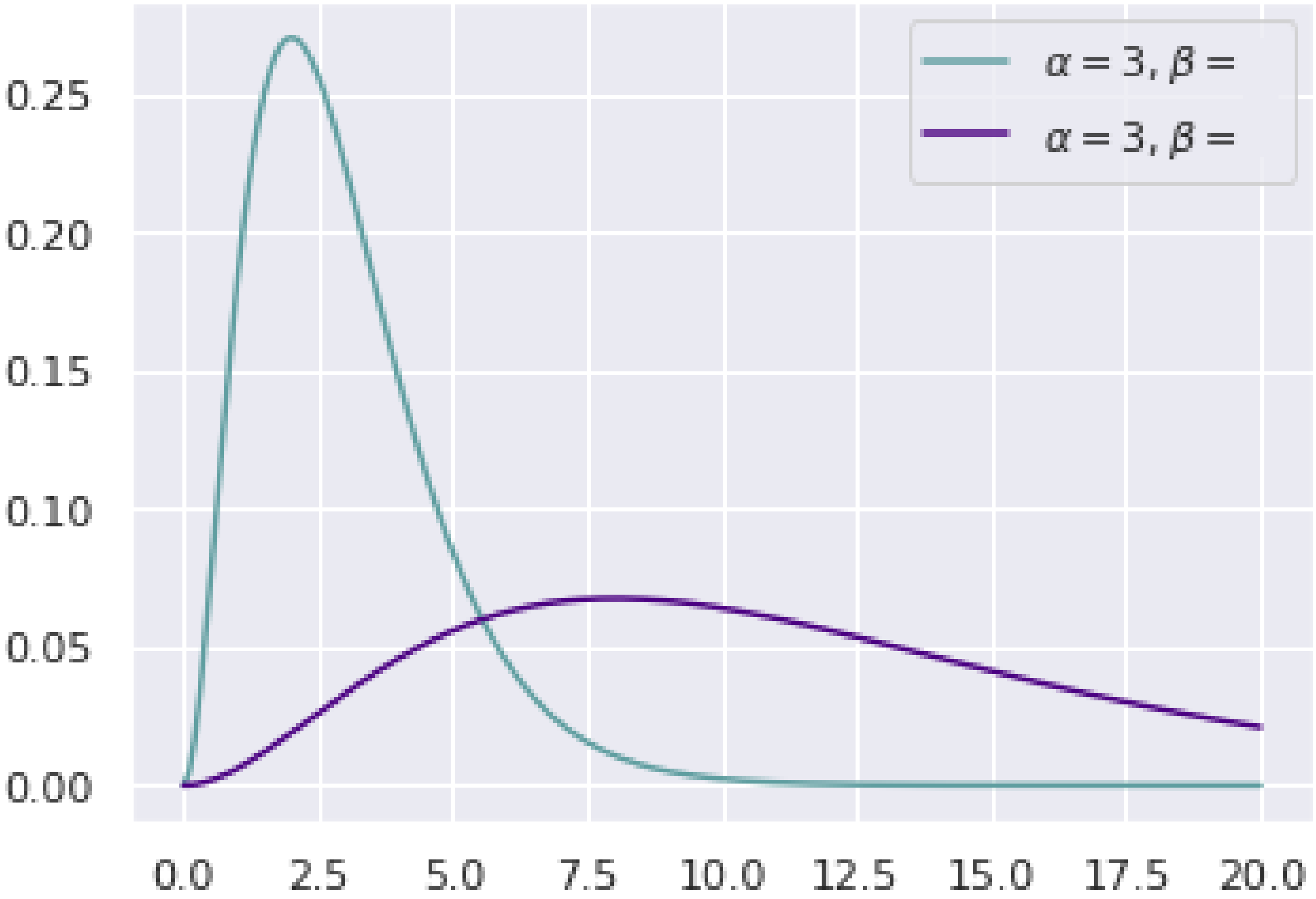
PECE Programa de Educação Continuada
Escola Politécnica da USP







PECE Programa de Educação Continuada
Escola Politécnica da USP



$\beta = 1$ e $\beta = 4$

ou

$\beta = 4$ e $\beta = 1$

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.

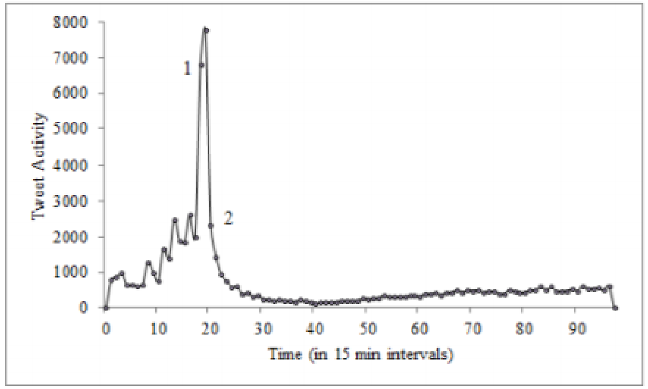
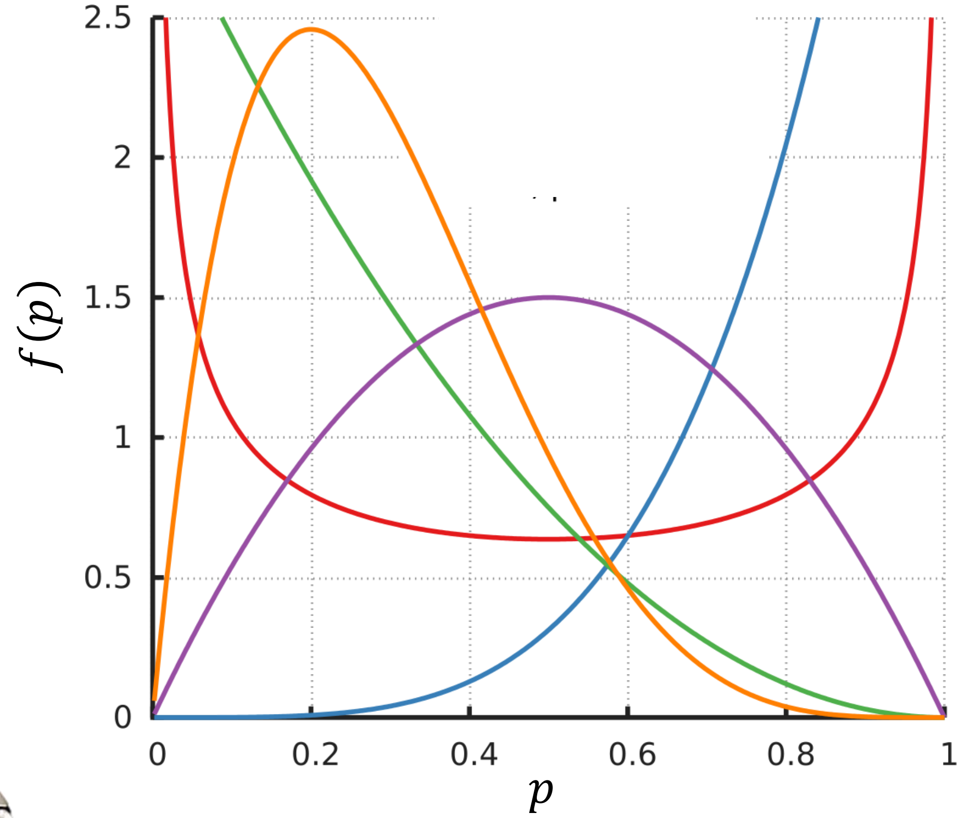


Figure 1: Frequency distribution of the "Stanley Cup" topic.



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

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

$\Gamma(n) = (n-1)!$ Para números naturas.

MÉDIA E VARIÂNCIA

A média e variância são dados por,

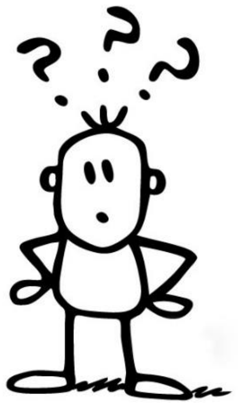
$$\mu = \frac{\alpha}{\alpha + \beta}$$

$$\sigma^2 = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta - 1)}$$

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????

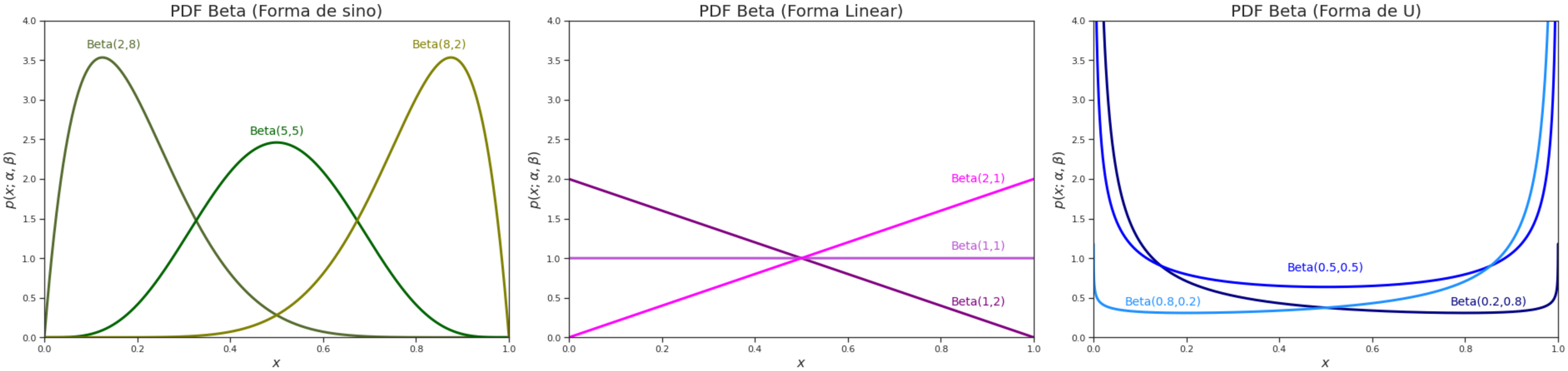


Sim. Na distribuição binomial.

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

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.



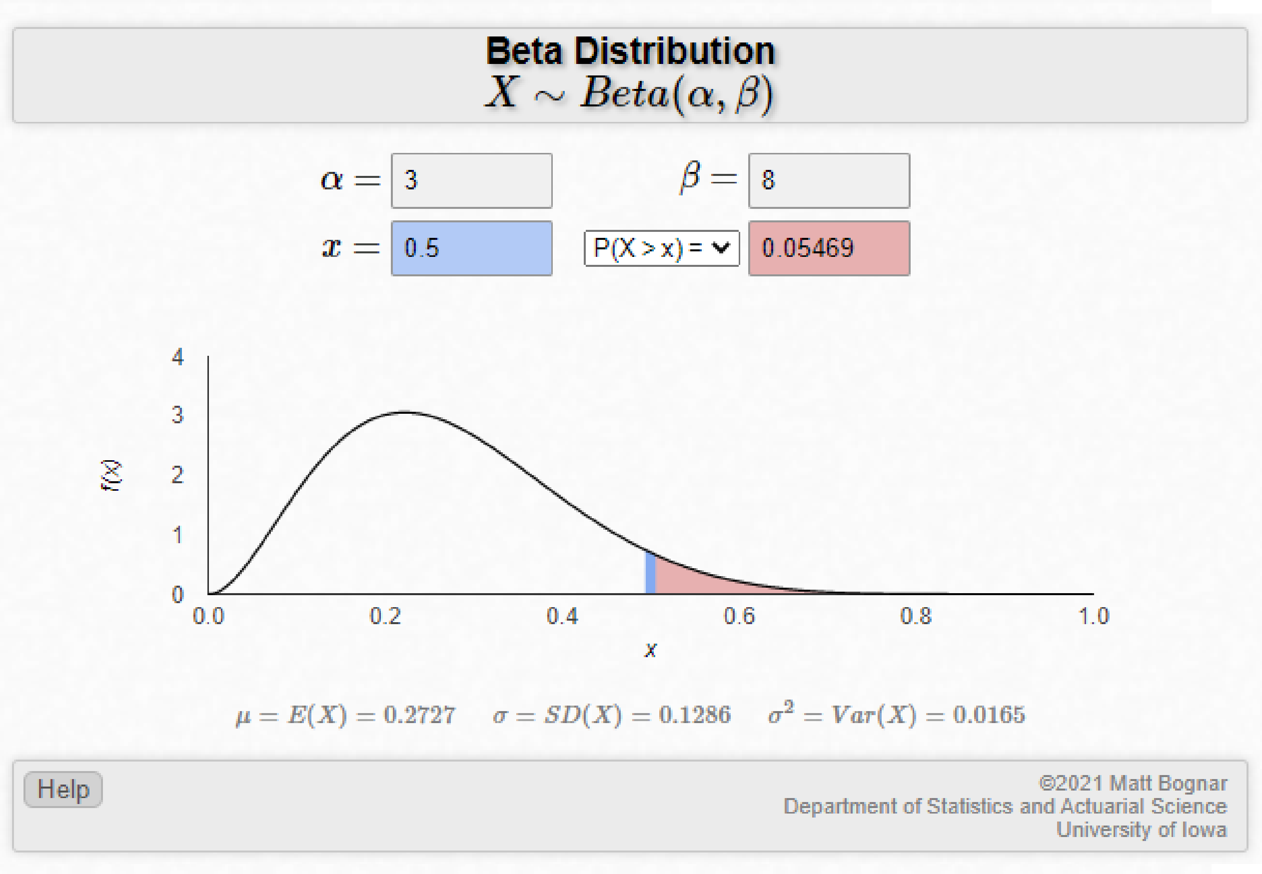
O PDF de uma distribuição beta é aproximadamente normal se $\alpha + \beta$ for grande o suficiente e α e β forem aproximadamente iguais.

$\alpha < 1, \beta < 1$

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%?

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



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

SUA TAREFA

PECE Programa de Educação Continuada
Escola Politécnica da USP

Completar o notebook IAD_001_2024_L09.ipynb

As questões sem resposta estão em verde.

Serão respondidas em uma prova online

NEVER GIVE UP

ACABOU...

Reveja a aula antes de resolver os exercícios.