Análise Estatística de dados

Inteligência Artificial

PECE Programa de Educação Continuada
Escola Politécnica da USP

USP

AULA 09 – MODELOS DE PROBABILIDADE CONTÍNUOS

Larissa Driemeier

PROGRAMA DO CURSO

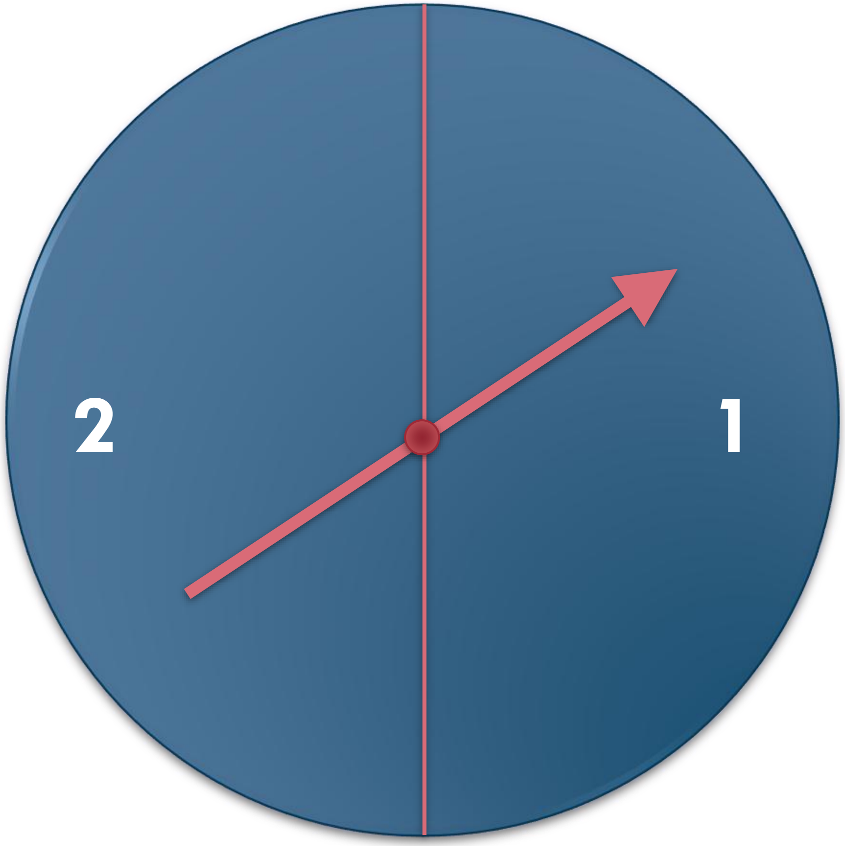
Aula	Data	Conteúdo da Aula
01	18/02	Aula Inaugural
02	25/02	Introdução ao Curso. Noções de Álgebra Linear , Geometria Analítica Parte I
03	11/03	Noções de Álgebra Linear , Geometria Analítica Parte II
04	18/03	Decomposição de valor singular (SVD)
05	25/03	Otimização: derivadas, derivadas parciais (operadores 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.

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**.

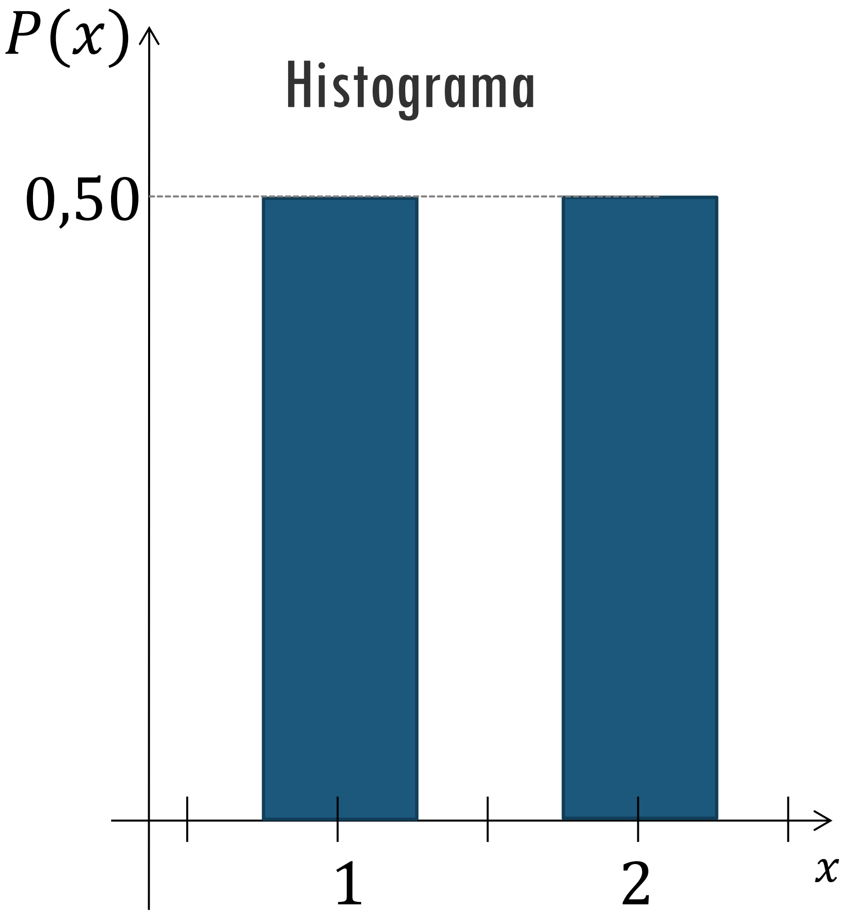
Gira-se o ponteiro e anota-se o número que o ponteiro, quando parado, está sinalizando.

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

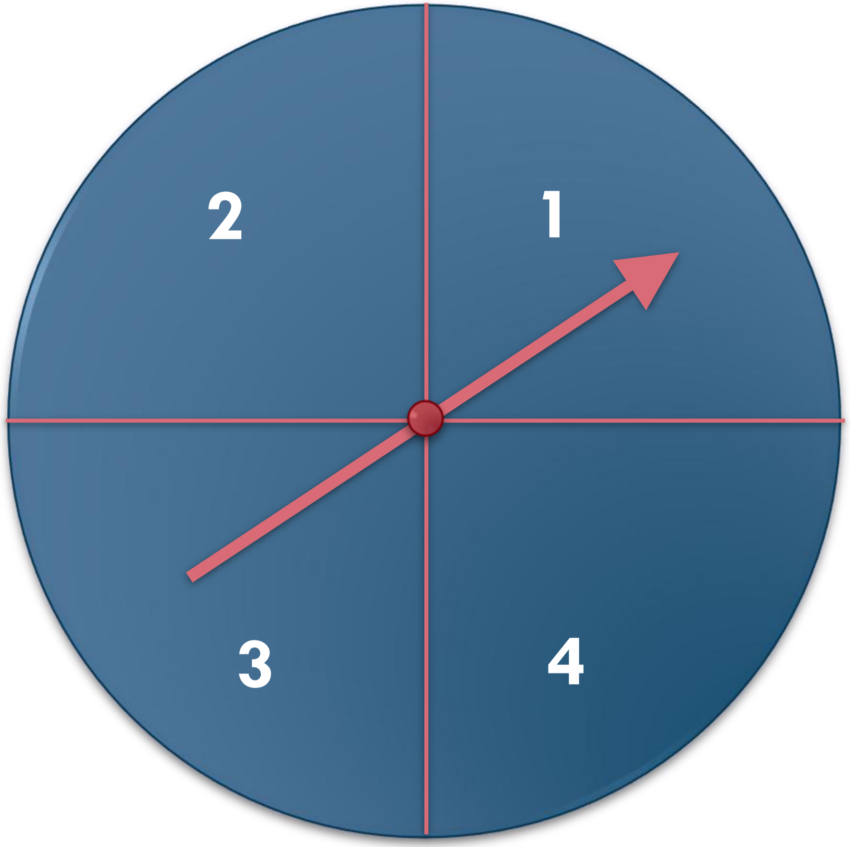


DISTRIBUIÇÃO DE PROBABILIDADES

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



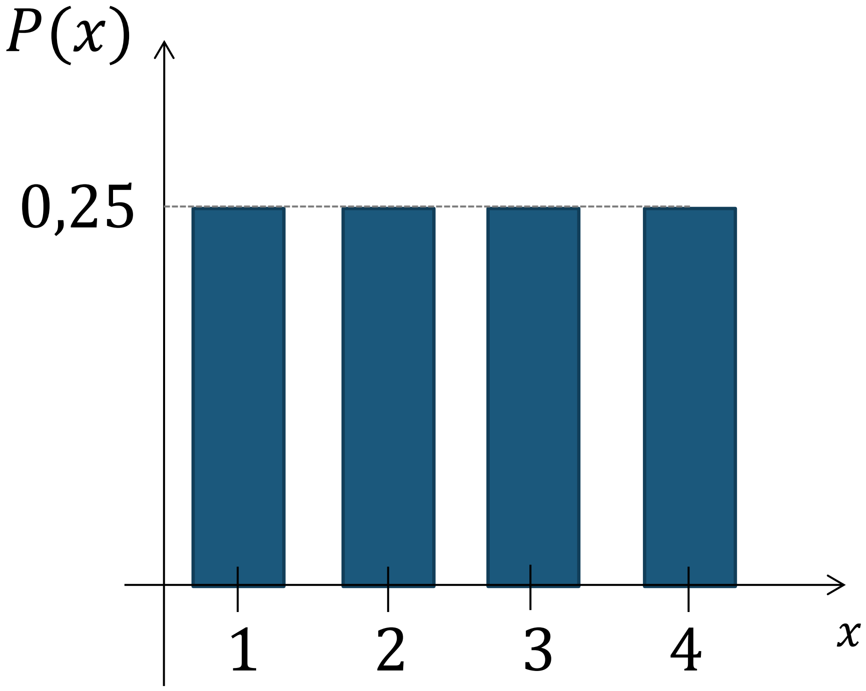
Ainda com respeito ao jogo anterior, o círculo é dividido agora em quatro partes iguais. Construir a distribuição de probabilidades para o **número que o ponteiro, quando parado, está sinalizando.**



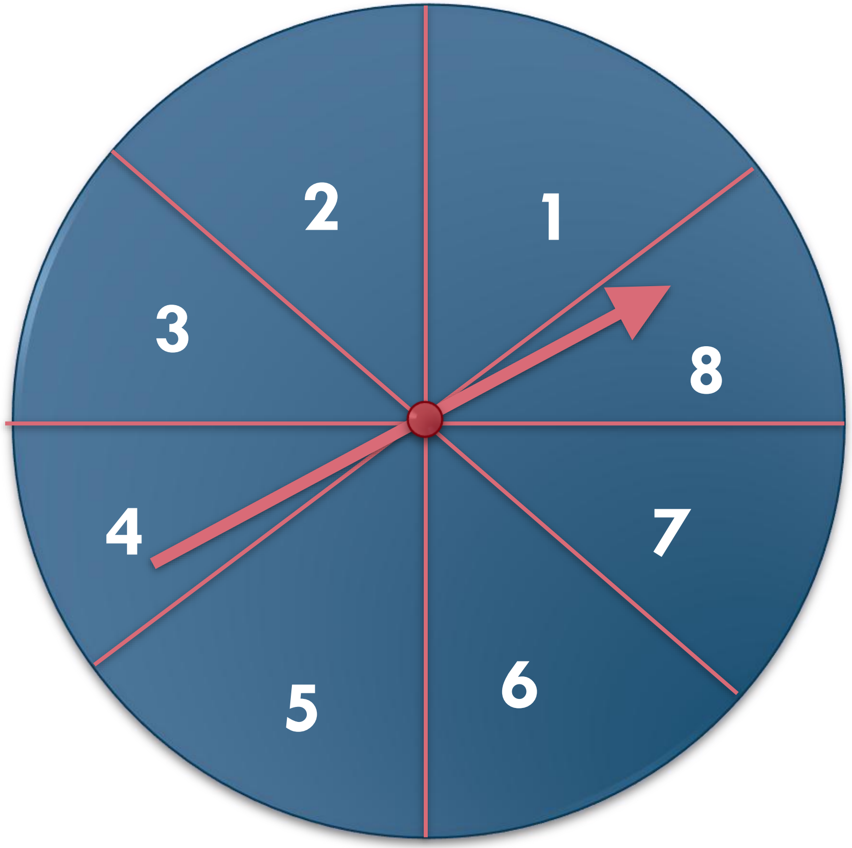
DISTRIBUIÇÃO DE PROBABILIDADES

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

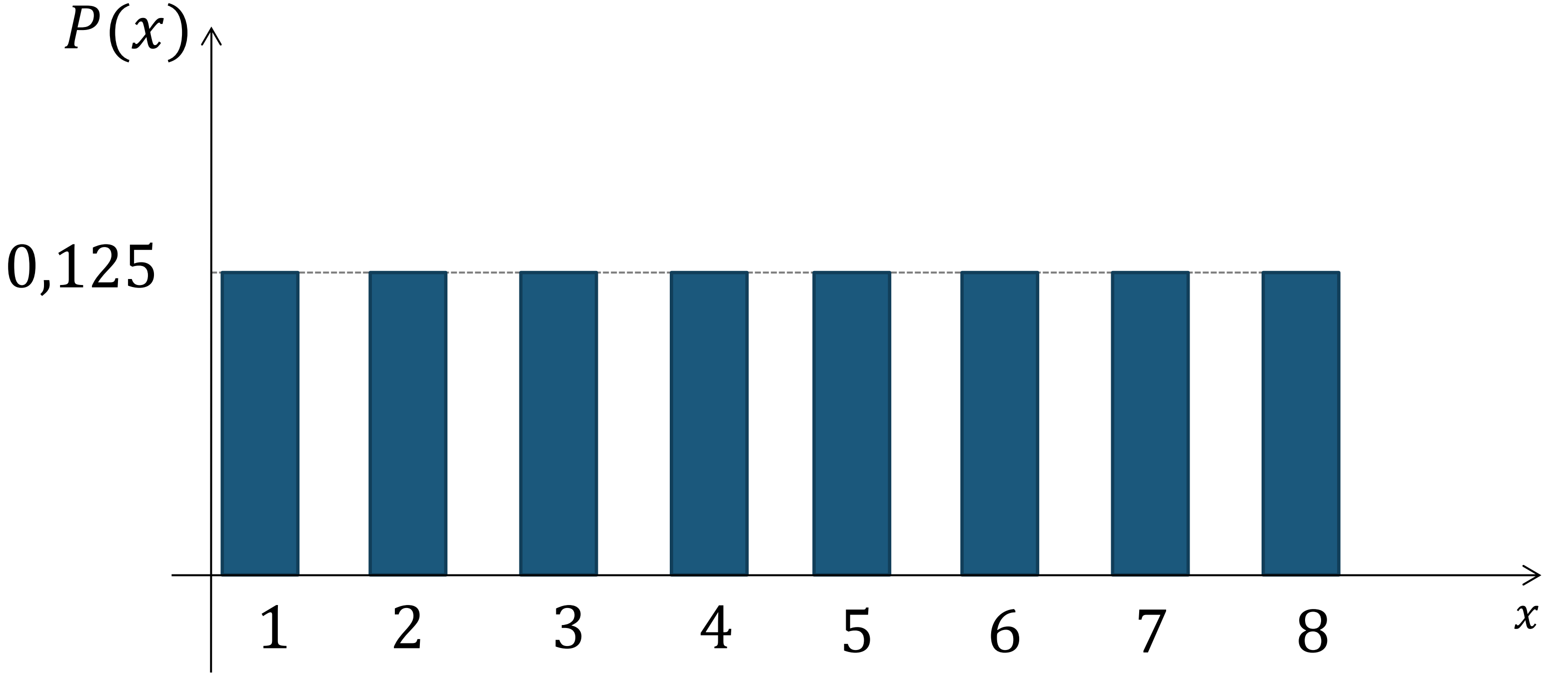
Histograma



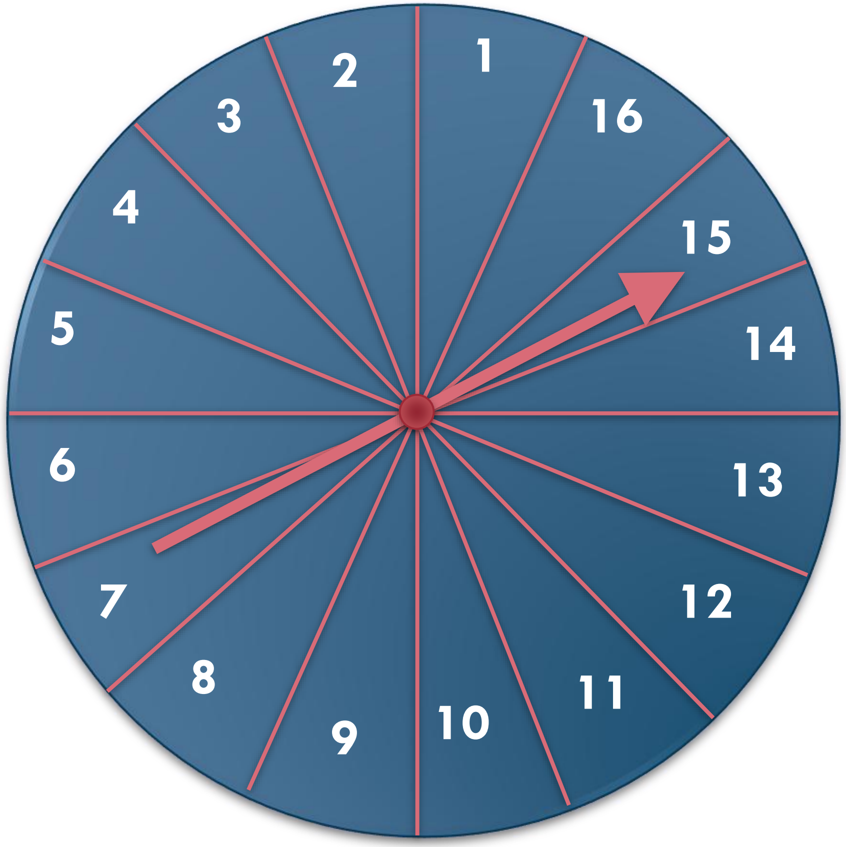
Agora construir a distribuição de probabilidade, mas dividindo o círculo em 8 partes.



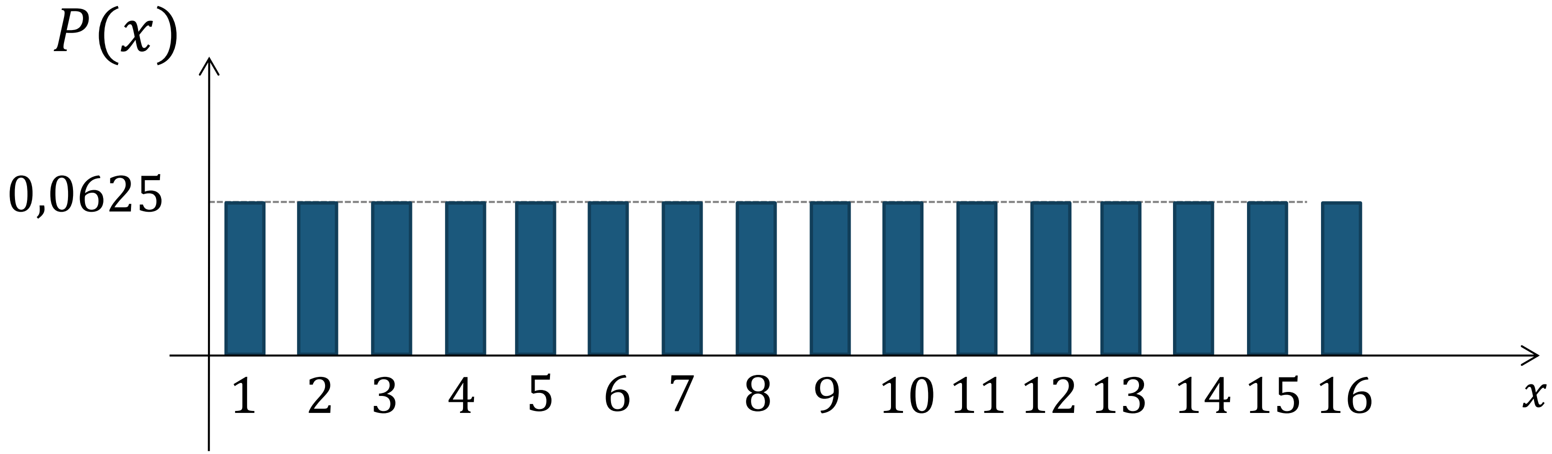
HISTOGRAMA



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)*:

$$y = p(x).$$

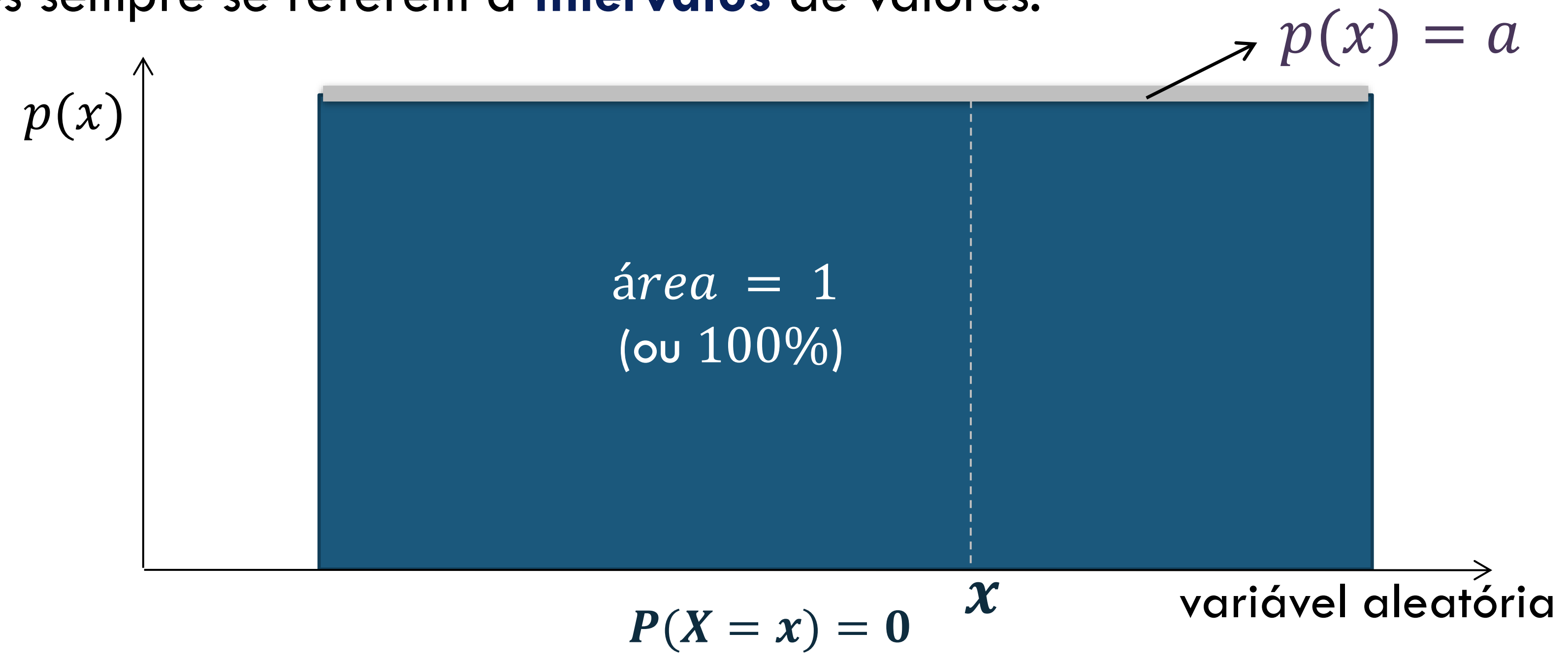
A probabilidade é calculada como a área sob a função (integrando).

FUNÇÃO DA DENSIDADE DE PROBABILIDADE

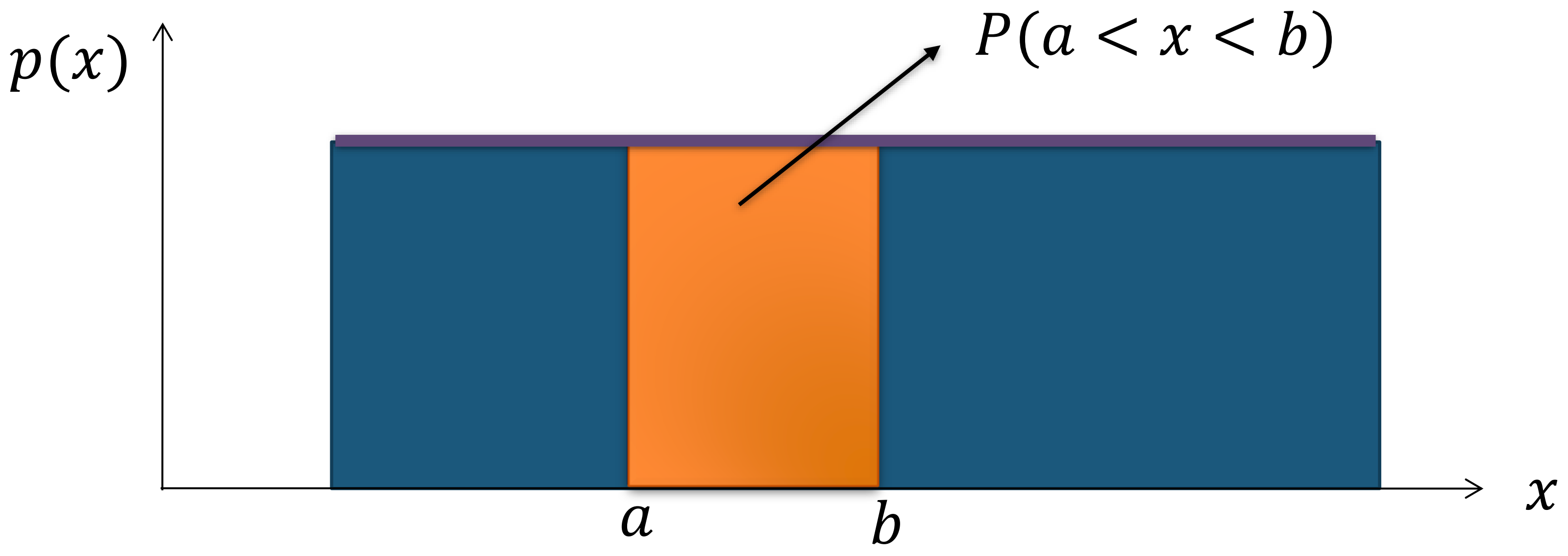
A área sob a função densidade é **1**.

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

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

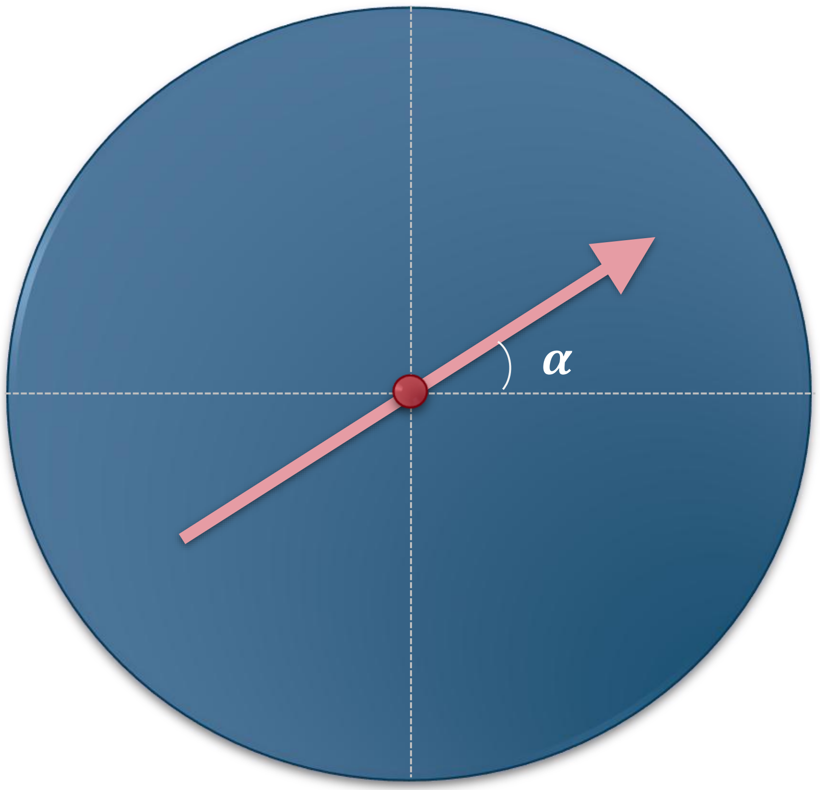


A probabilidade da variável aleatória assumir um valor em um intervalo é igual à **área** sob a função densidade naquele intervalo.

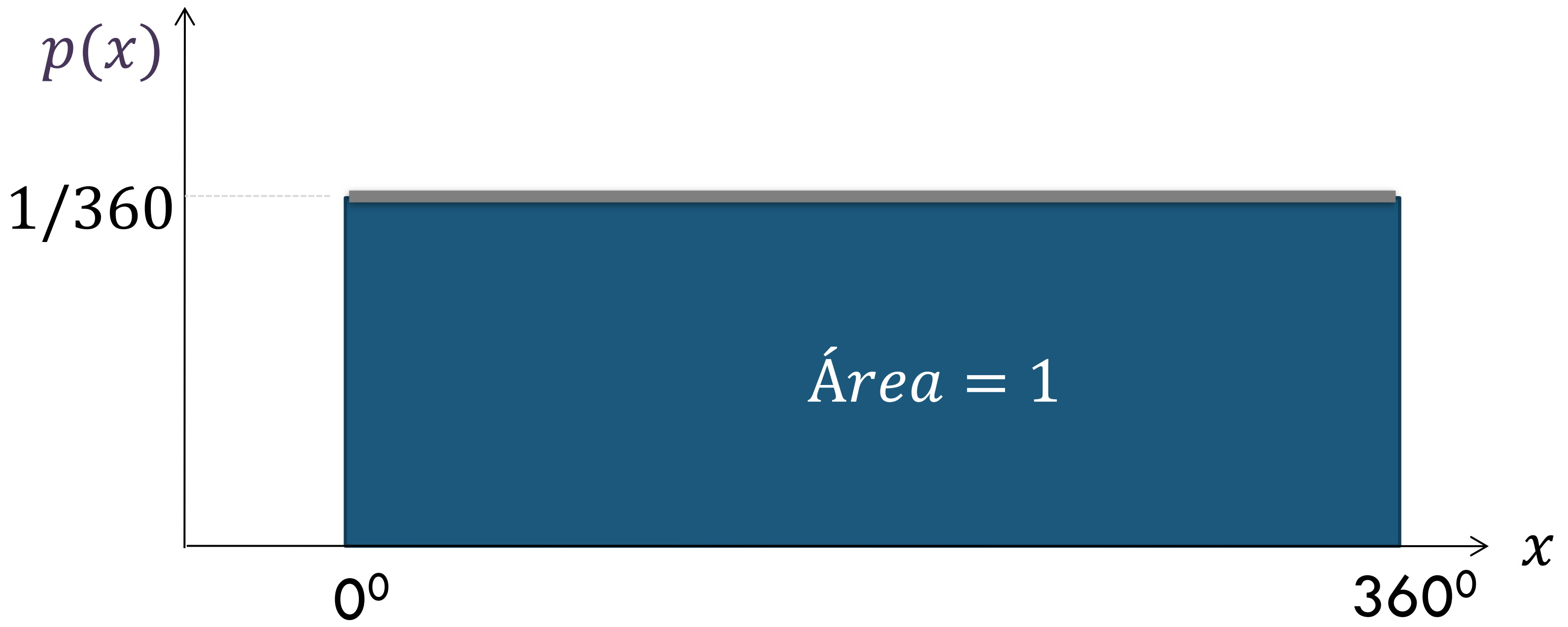


VOLTANDO AO EXERCÍCIO ANTERIOR

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.

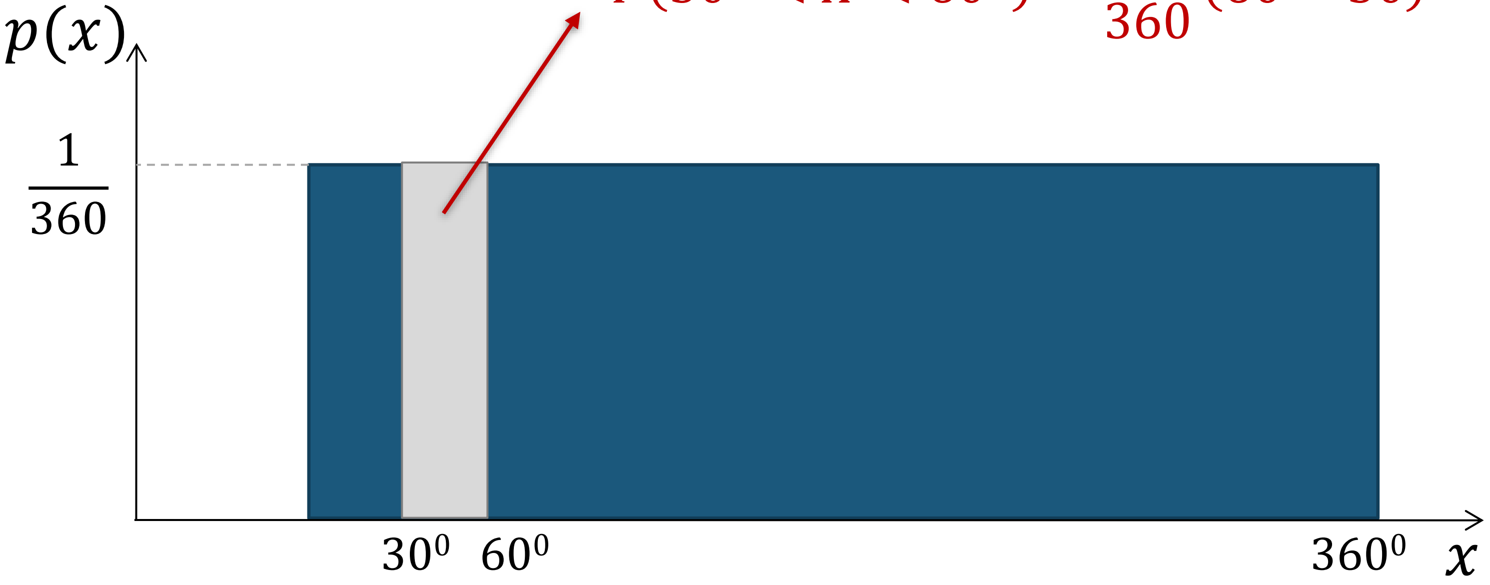


$p(x)$
$1/360$
$Área = 1$
0^0
360^0
x

EXERCÍCIO

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

$$P(30^0 < X < 60^0) = \frac{1}{360}(60 - 30) = \frac{30}{360} = 0{,}08333$$



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

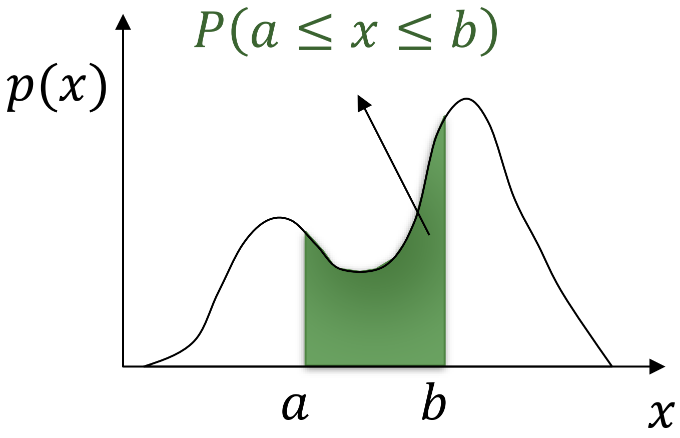
A função densidade de probabilidade ¨(em inglês *probability density function –pdf-)* DEVE apresentar as seguintes propriedades:

- 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:

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

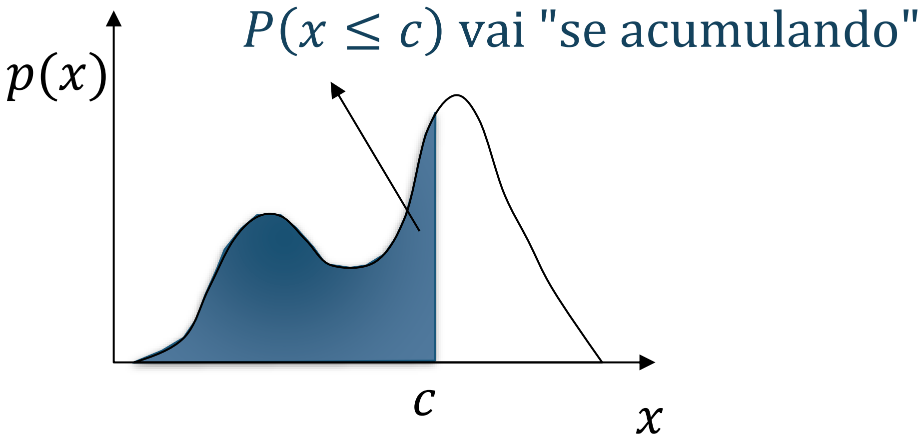
- A probabilidade de x assumir valores entre a e b é dada pela relação:

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



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

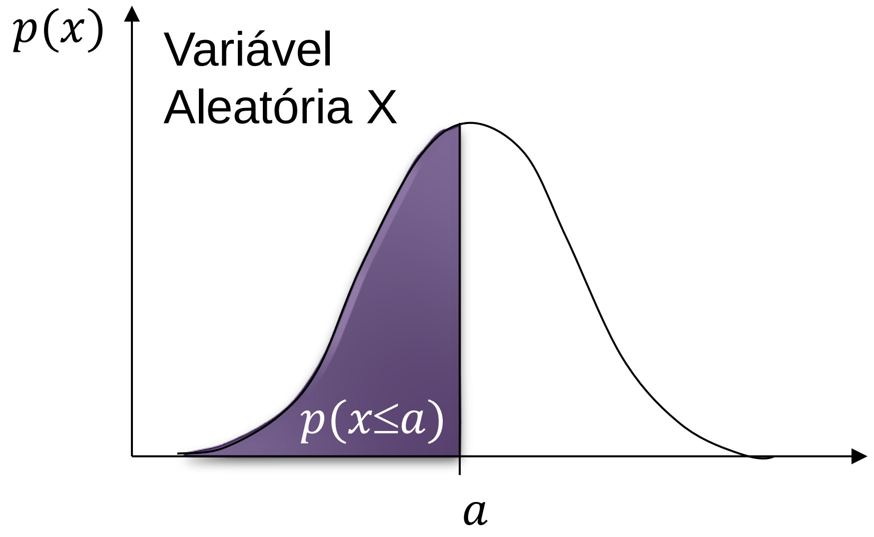
$$\int_{-\infty}^c p(x)dx = P(x \leq c)$$



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

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

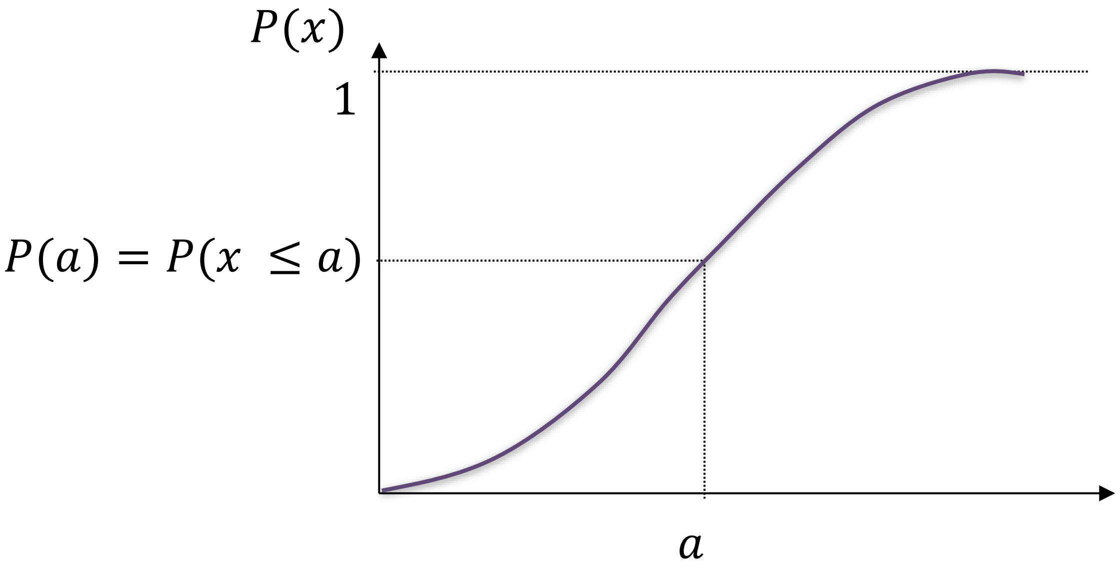
$= P(a < x \leq b) = P(a \leq x \leq b)$.



Função
Densidade de
Probabilidade

Probability
Density
Function

pdf



Função
Distribuição
Acumulada

Cumulative
Distribution
Function

cdf

EXERCÍCIO

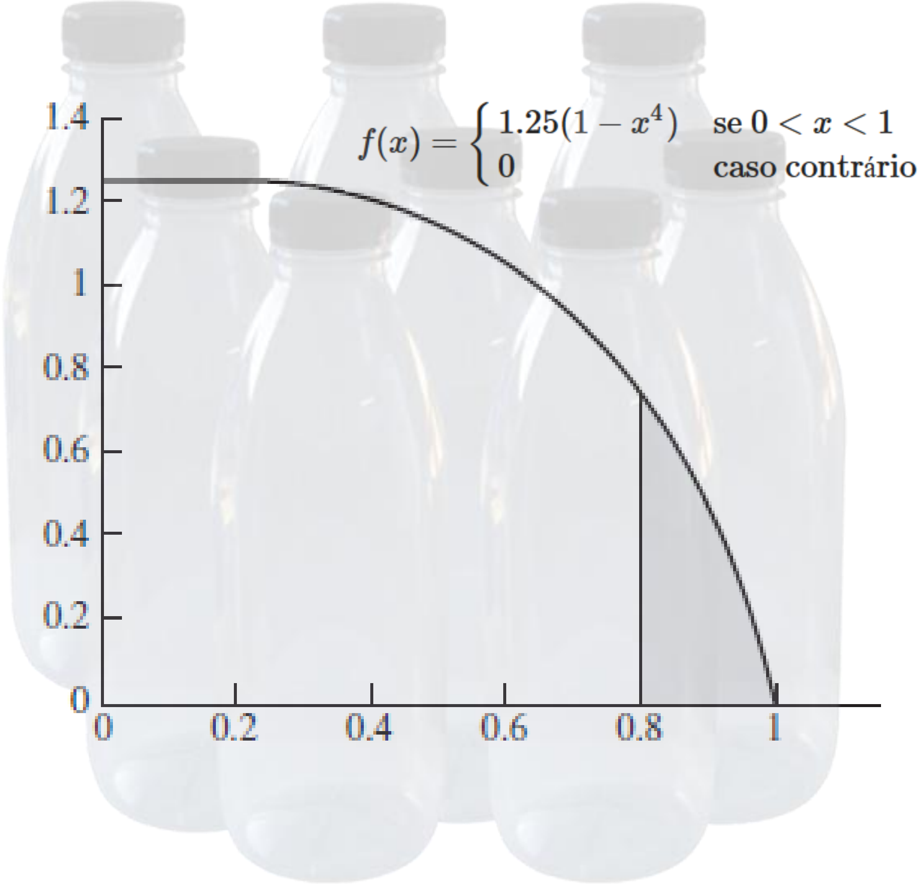
Imagine que uma fábrica produz tampas de garrafa. Cada tampa precisa encaixar perfeitamente no gargalo da garrafa. A diferença entre a tampa e o gargalo é chamada de folga – é o "*jogo*" que sobra, que varia de tampa para tampa. Essa folga é medida em milímetros.

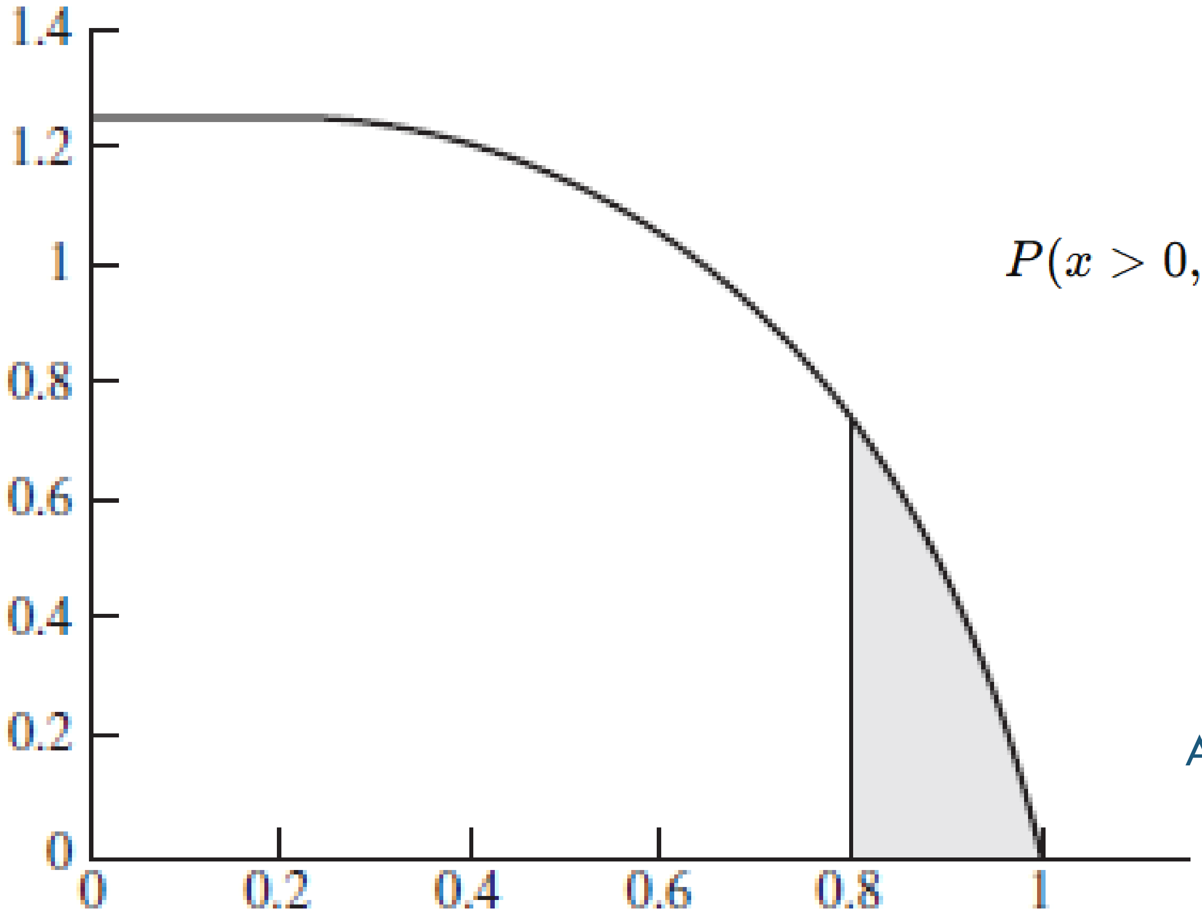
Vamos chamar essa folga de x, que pode ser descrita por uma função de densidade de probabilidade (ou seja, uma forma de modelar como os valores de folga se distribuem),

$$f(x) = \begin{cases} 1{,}25(1 - x^4) & se\ 0 < x < 1 \\ 0 & cc \end{cases}$$

A fábrica decidiu que, se a folga for maior que 0,8 mm, a tampa não serve, porque o gargalo da garrafa vai ficar frouxo e pode vazar. Então, essas tampas com folga muito grande são descartadas.

Pergunta: Sabendo como a folga x se distribui, qual a porcentagem das tampas que serão descartadas?





PECE Programa de Educação Continuada
Escola Politécnica da USP

$$P(x > 0{,}8) = \int_{0{,}8}^{1} 1{,}25(1 - x^4)\,dx$$

$$\int 1{,}25(1 - x^4)dx = 1{,}25\left(x - \frac{x^5}{5}\right)$$

Aplicando os limites de integração

$$P(x > 0{,}8) = 1{,}25\left[\left(1 - \frac{1^5}{5}\right) - \left(0{,}8 - \frac{0{,}8^5}{5}\right)\right]$$

$$P(x > 0{,}8) = 1{,}25 \cdot (0{,}8 - 0{,}734464) = 1{,}25 \cdot 0{,}065536 = 0{,}08192$$

Aproximadamente 8,2% das tampas serão descartadas, pois têm folga maior que 0,8 mm.

PECE Programa de Educação Continuada
Escola Politécnica da USP

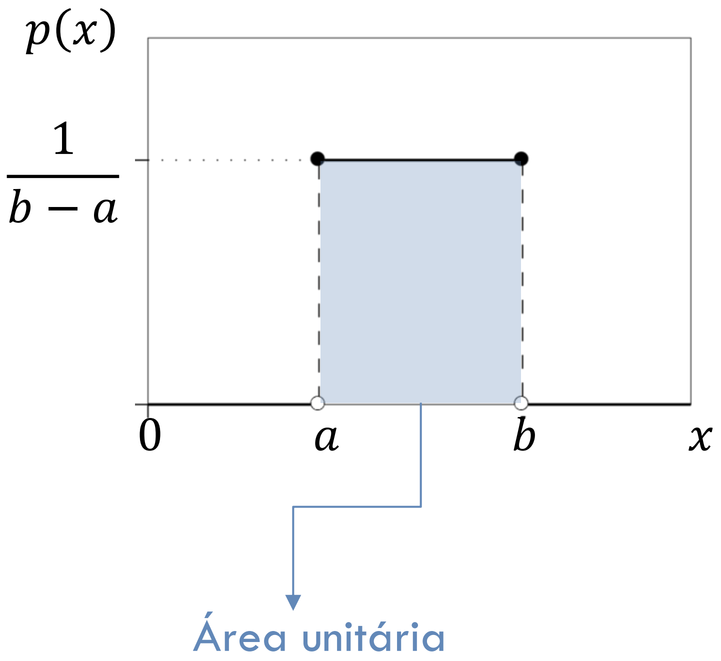
DISTRIBUIÇÃO UNIFORME

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.

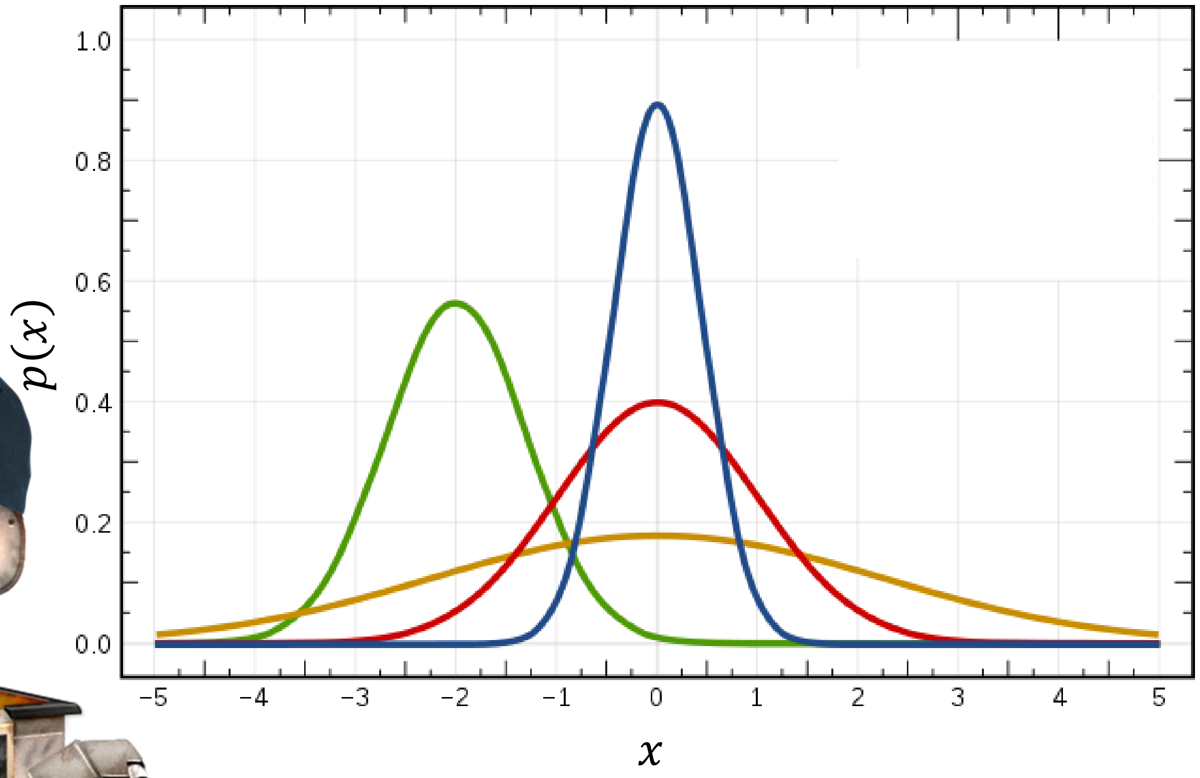
Qualquer intervalo de números de largura igual tem uma probabilidade igual de ser observado.

EXERCÍCIO DE DISTRIBUIÇÃO UNIFORME

Suponha que você esteja esperando pelo metrô. Você tem de zero a dez minutos para esperar. Sua probabilidade de ter que esperar qualquer número de minutos nesse intervalo é a mesma.

Dada a história acima, construa os seguintes gráficos:

1. Qual é a probabilidade de você esperar de 4 a 6 minutos pelo trem?
2. Qual é a probabilidade de aparecer um trem depois de 5 minutos?
3. Qual é a probabilidade de não se esperar mais que 2,5 minutos para o próximo trem?
4. Qual é a probabilidade de se esperar mais de 2 minutos por um trem, dado que já passou 1 minuto sem aparecer?
5. Qual é a probabilidade de você esperar exatamente 1,375 minutos?



DISTRIBUIÇÃO GAUSSIANA

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

DISTRIBUIÇÃO NORMAL

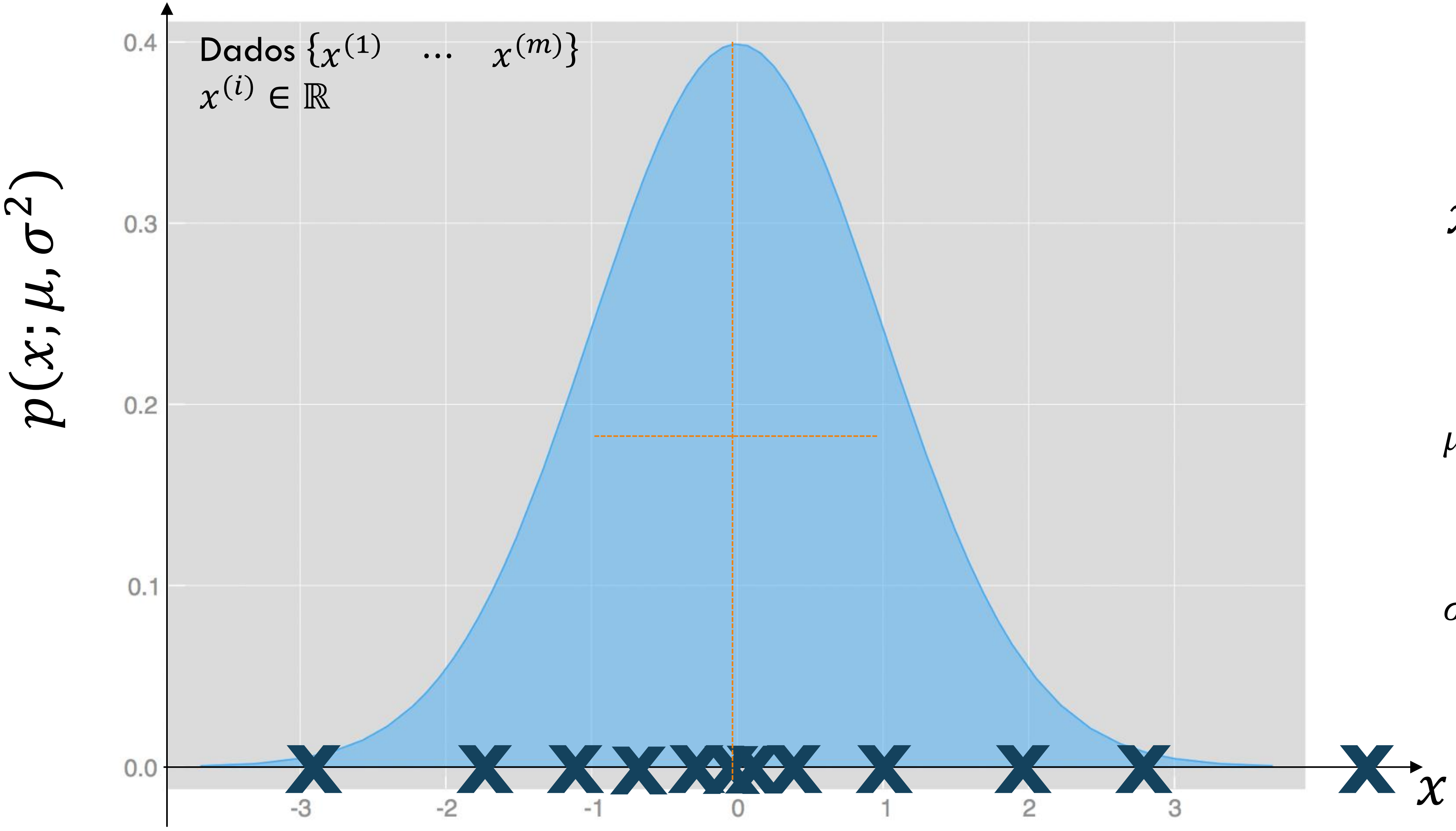
A função densidade de probabilidade de uma distribuição normal é dada por:

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

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 – 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

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

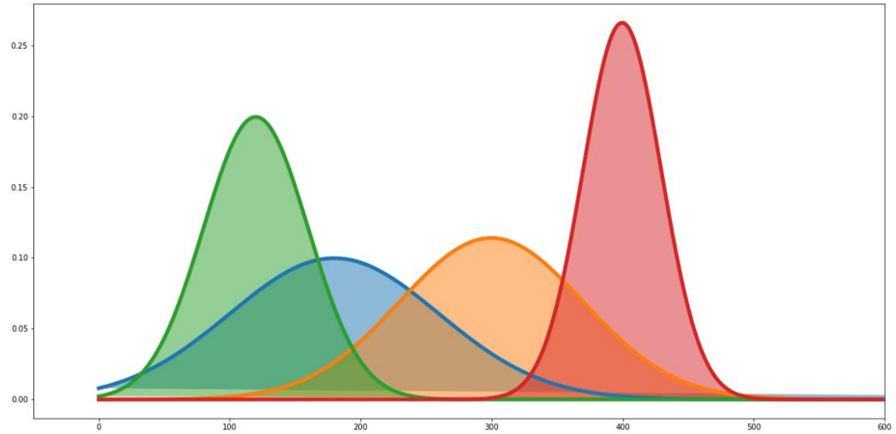
$$\mathcal{L}\{\mu,\sigma\} = (2\pi)^{-m/2}\sigma^{-m} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{m}(x_i-\mu)^2\right]$$

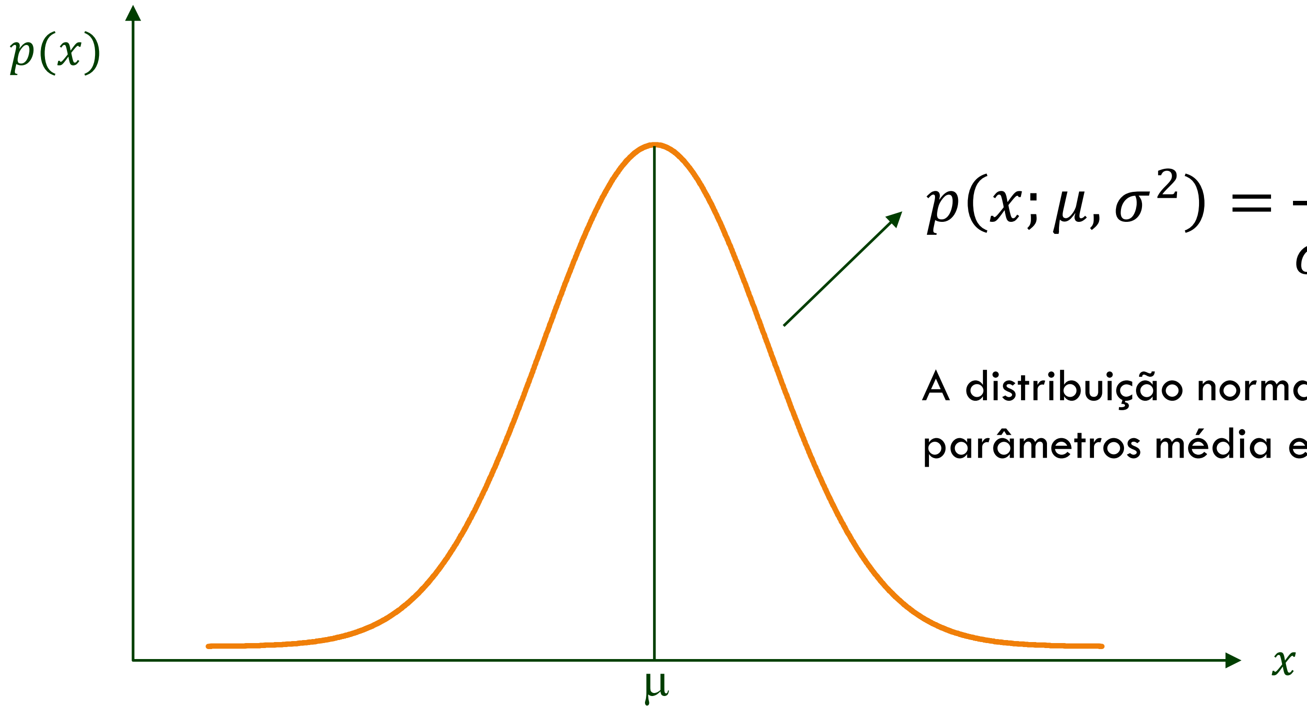
função de log-verossimilhança negativa

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

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

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

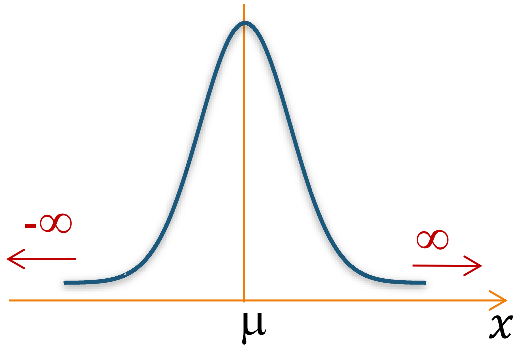


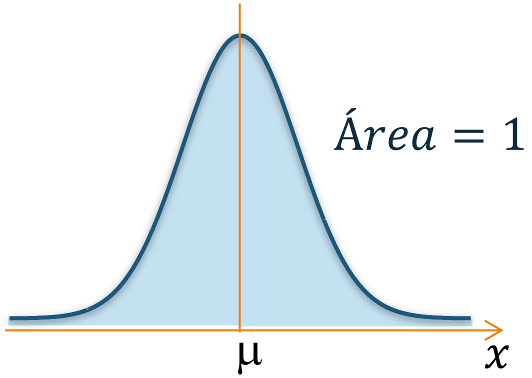


$$p(x;\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

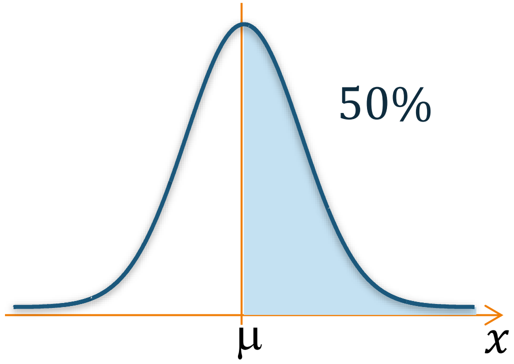
A distribuição normal é totalmente caracterizada pelos parâmetros média e variância.

- 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.

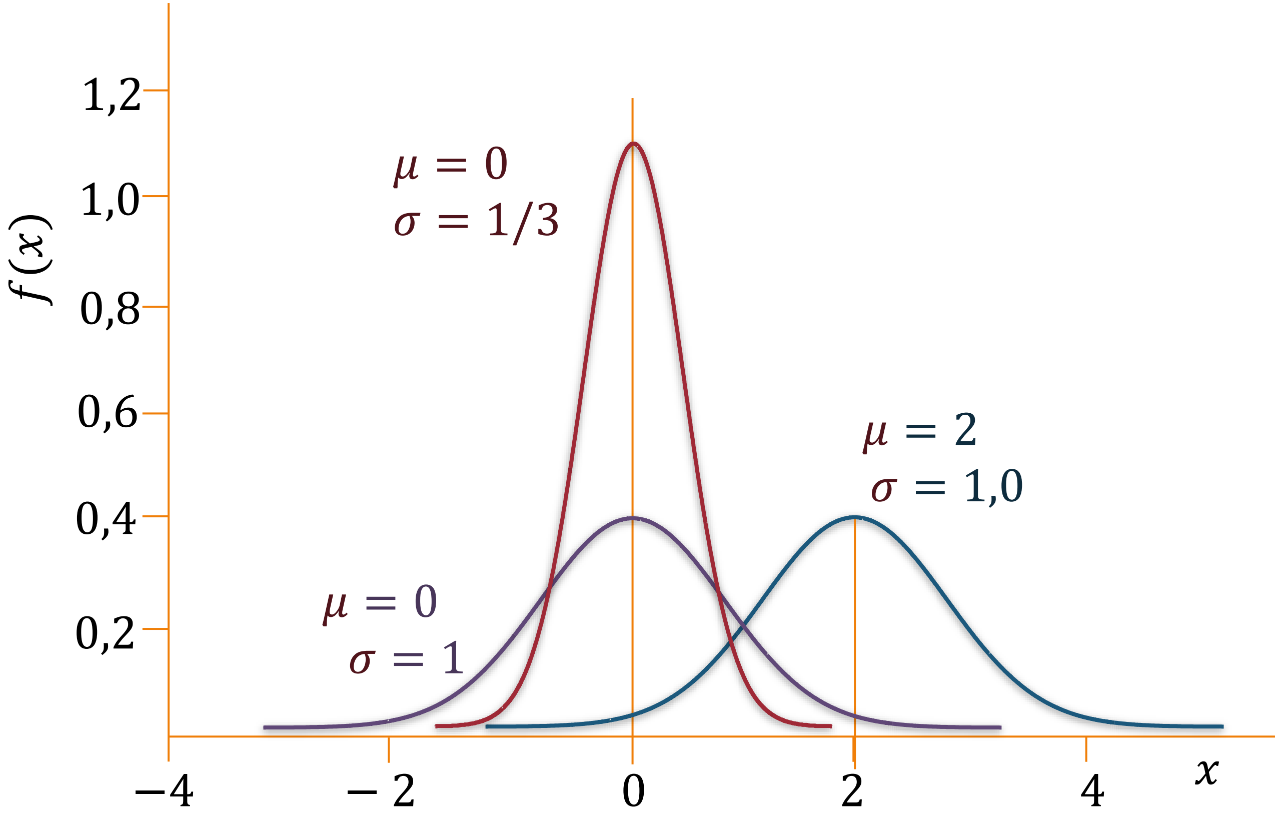




Dessa forma, tem-se que 50% da distribuição encontra-se à direita da média e os restantes 50% à esquerda desse parâmetro.



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



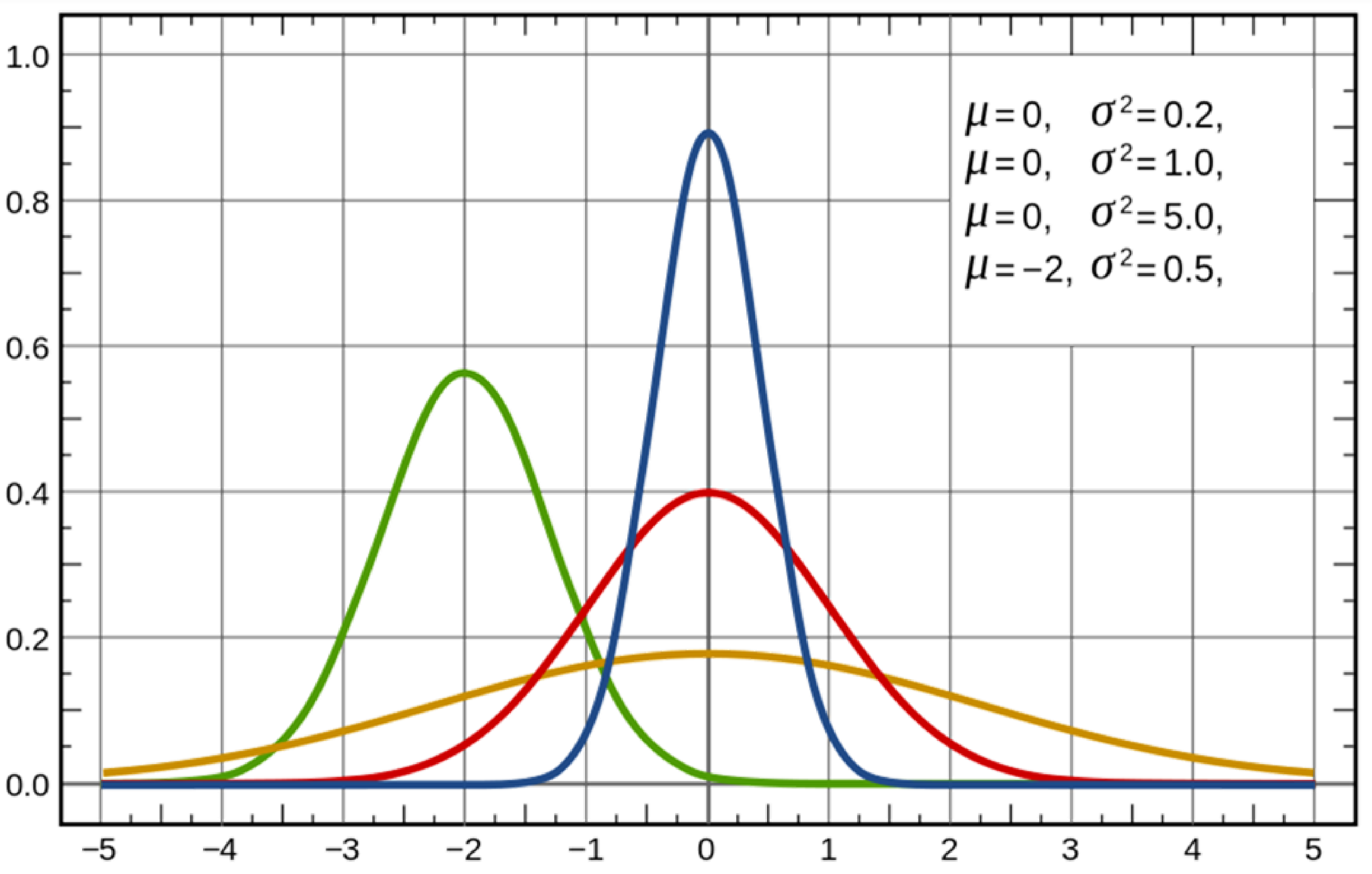
A média me da o centro da função distribuição de probabilidade e coincide com moda e mediana.

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

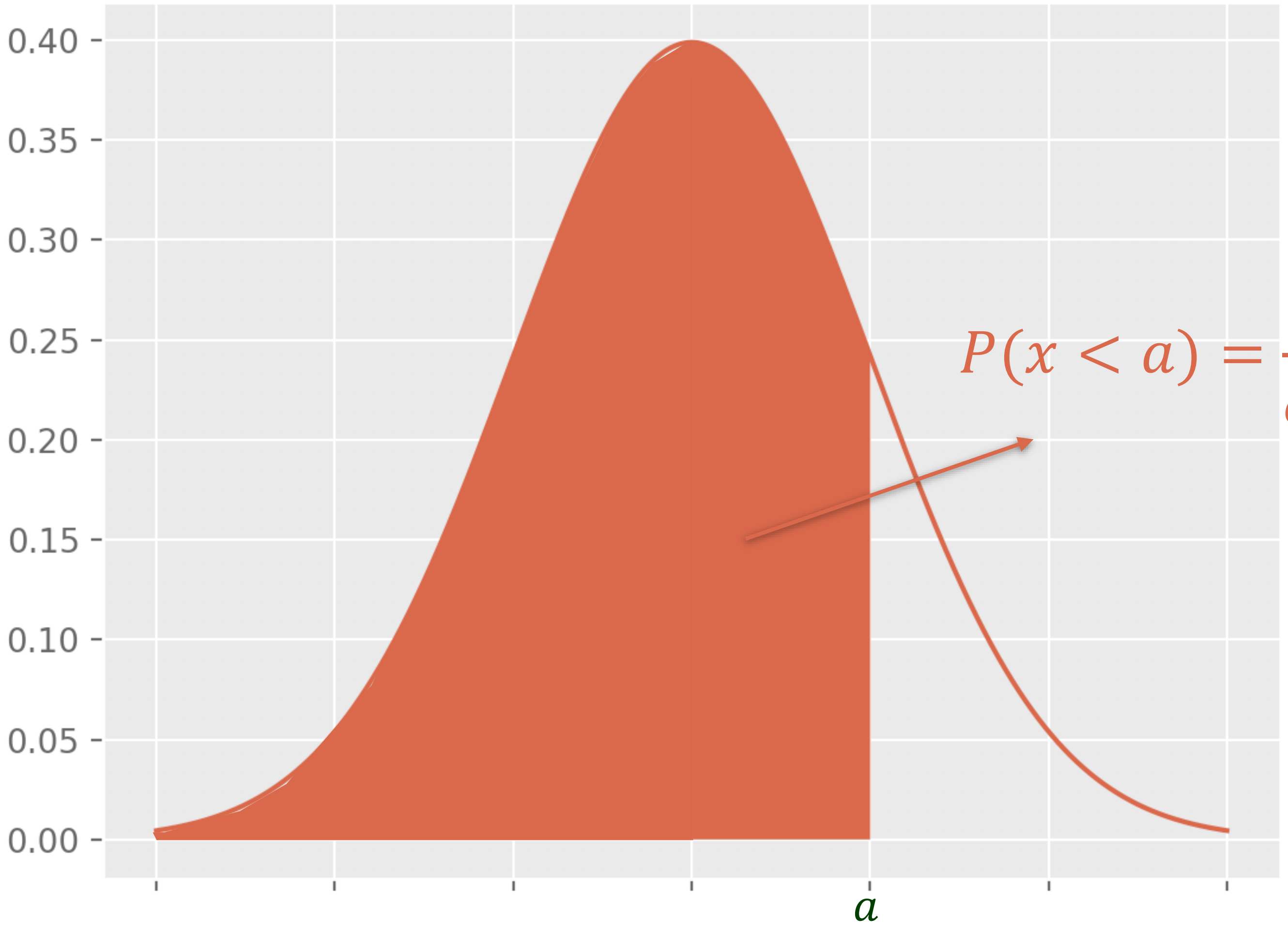
Quanto maior a variância maior será a dispersão da distribuição e mais achatada será a curva da função densidade de probabilidade.

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

COMPLETE A LEGENDA...

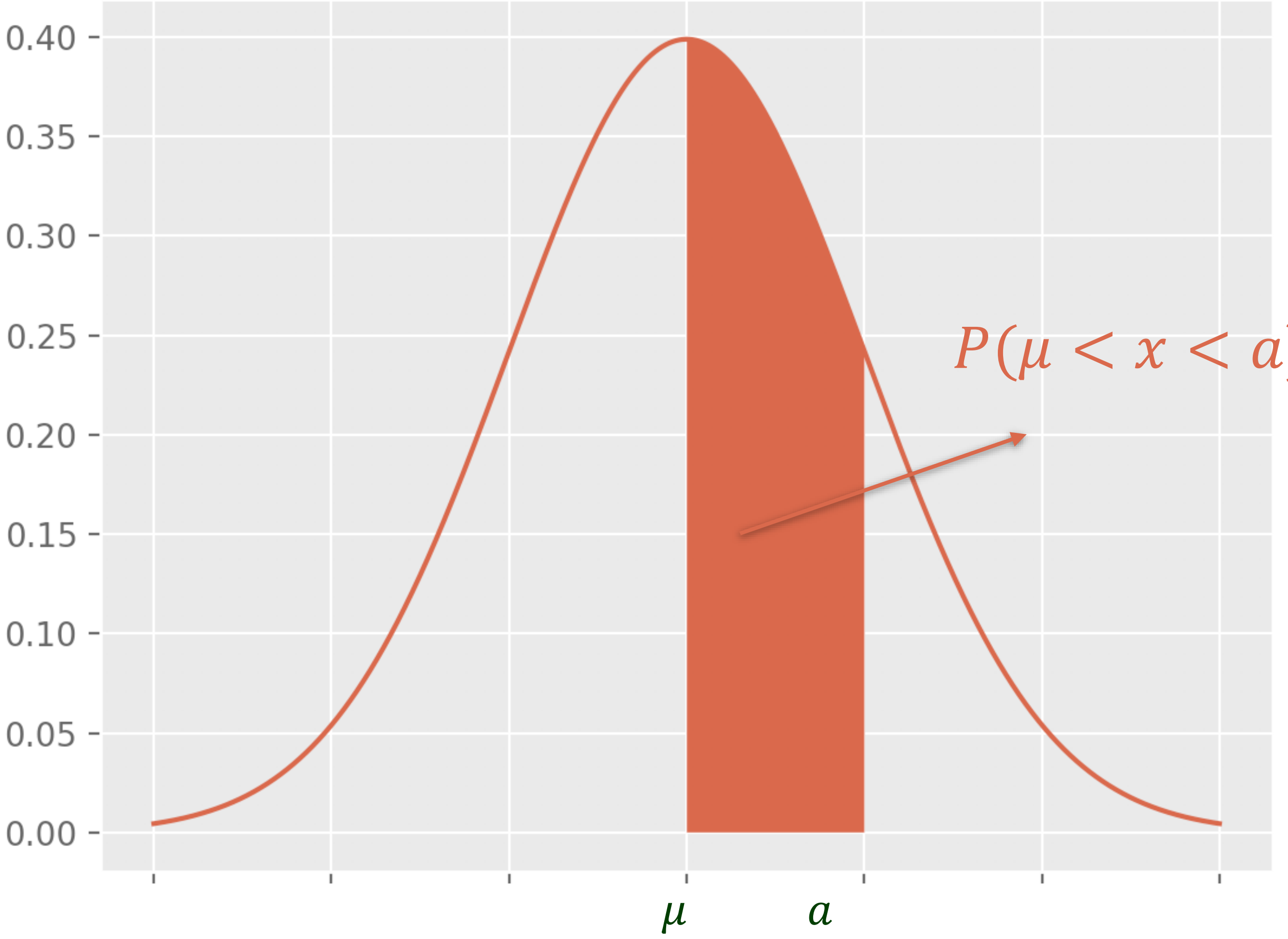


FUNÇÃO DISTRIBUIÇÃO ACUMULADA DA DISTRIBUIÇÃO NORMAL



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

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

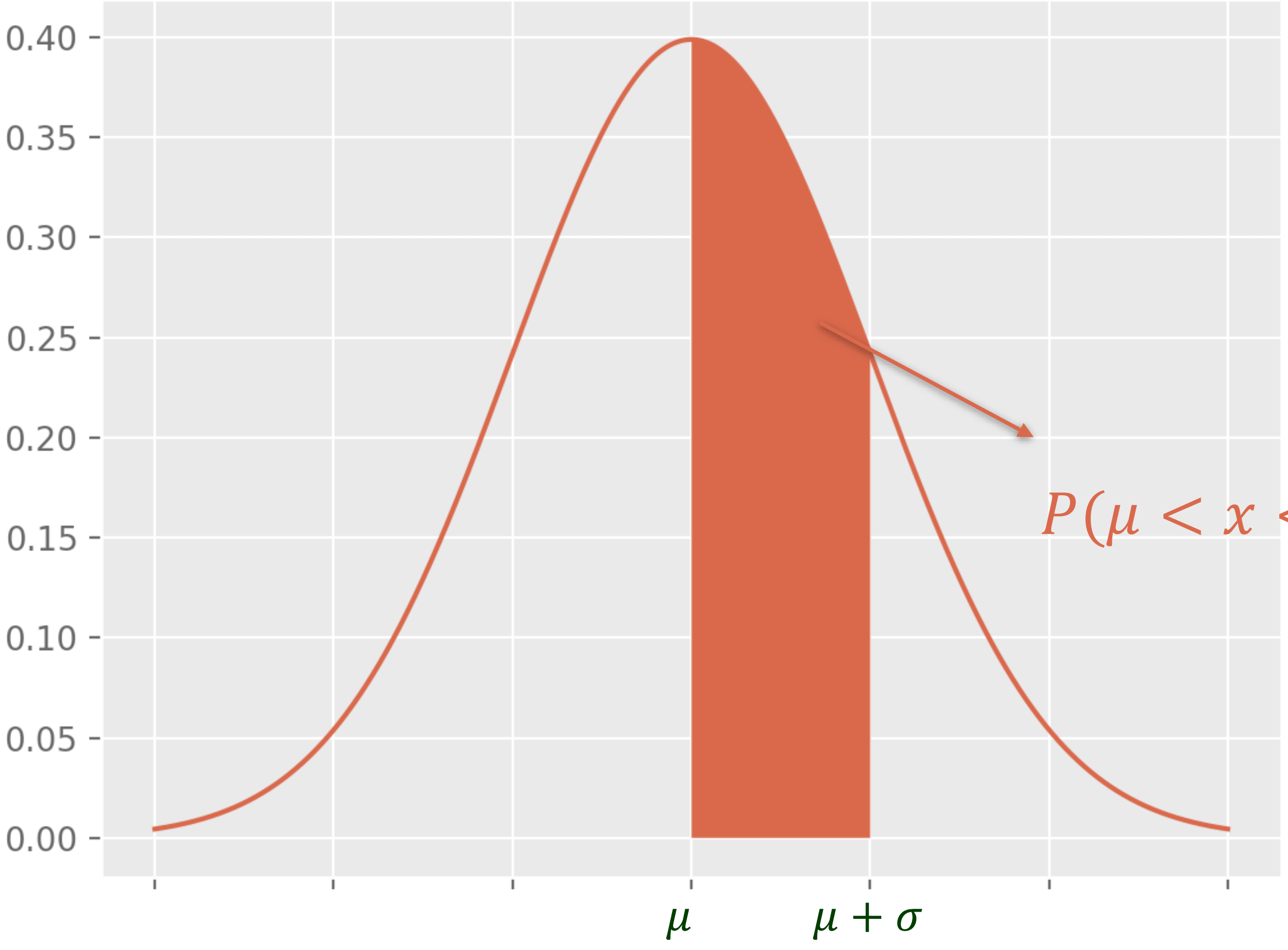
A função distribuição acumulada da distribuição normal é obtida pela execução da seguinte operação de integração:

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

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.

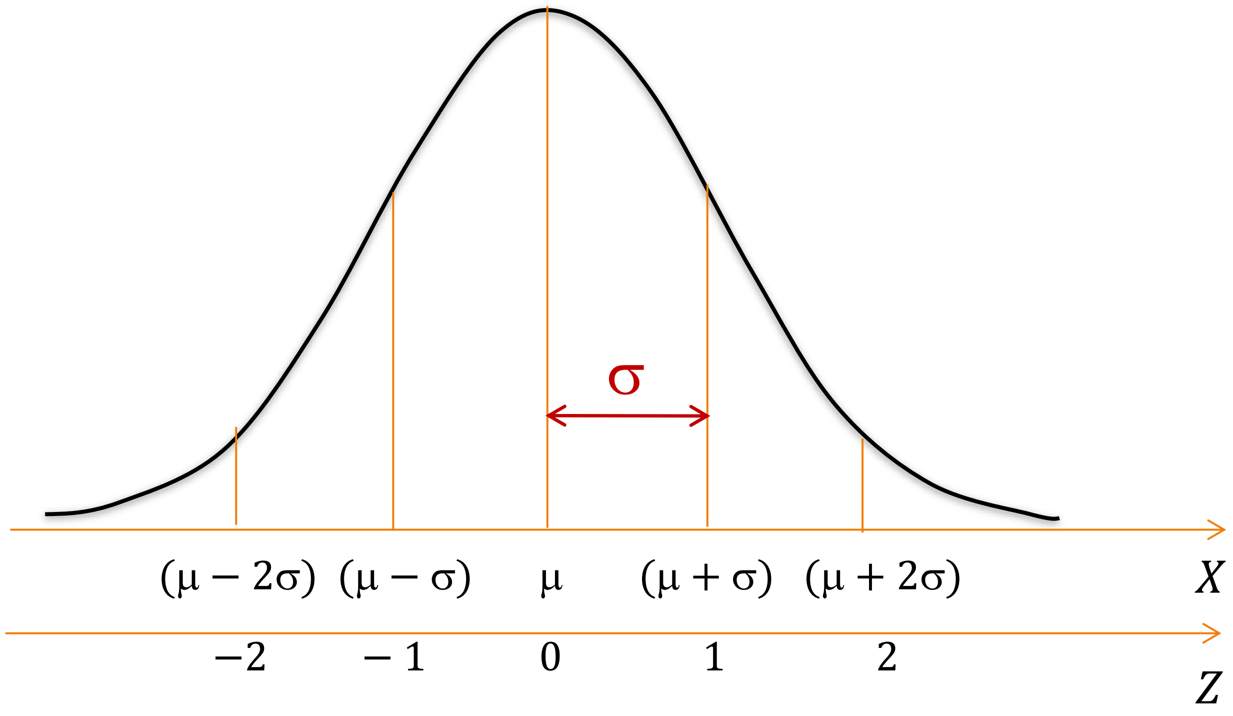
DISTÂNCIA PADRONIZADA



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

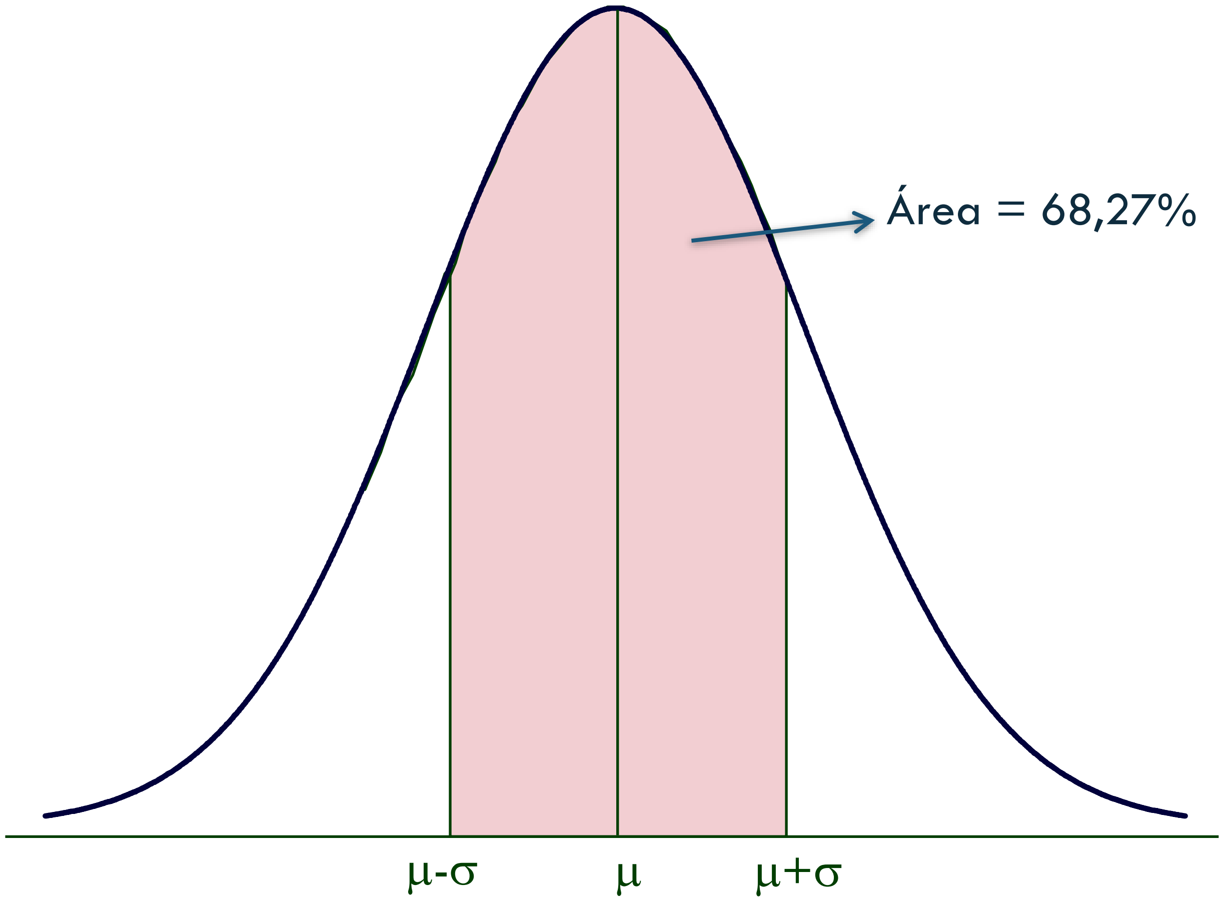
$$P(\mu < x < \mu + \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \int_{\mu}^{\mu+\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

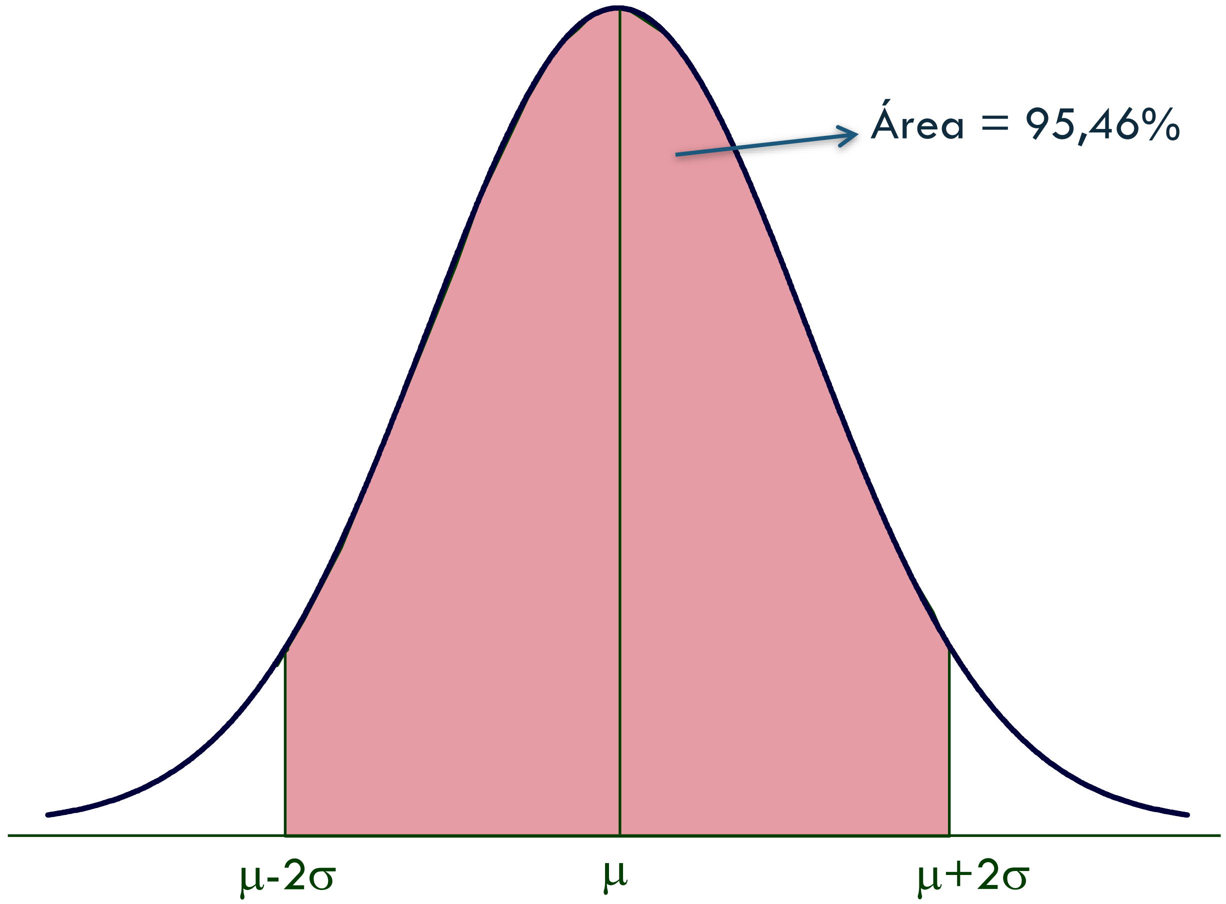
VARIÁVEL NORMAL PADRONIZADA

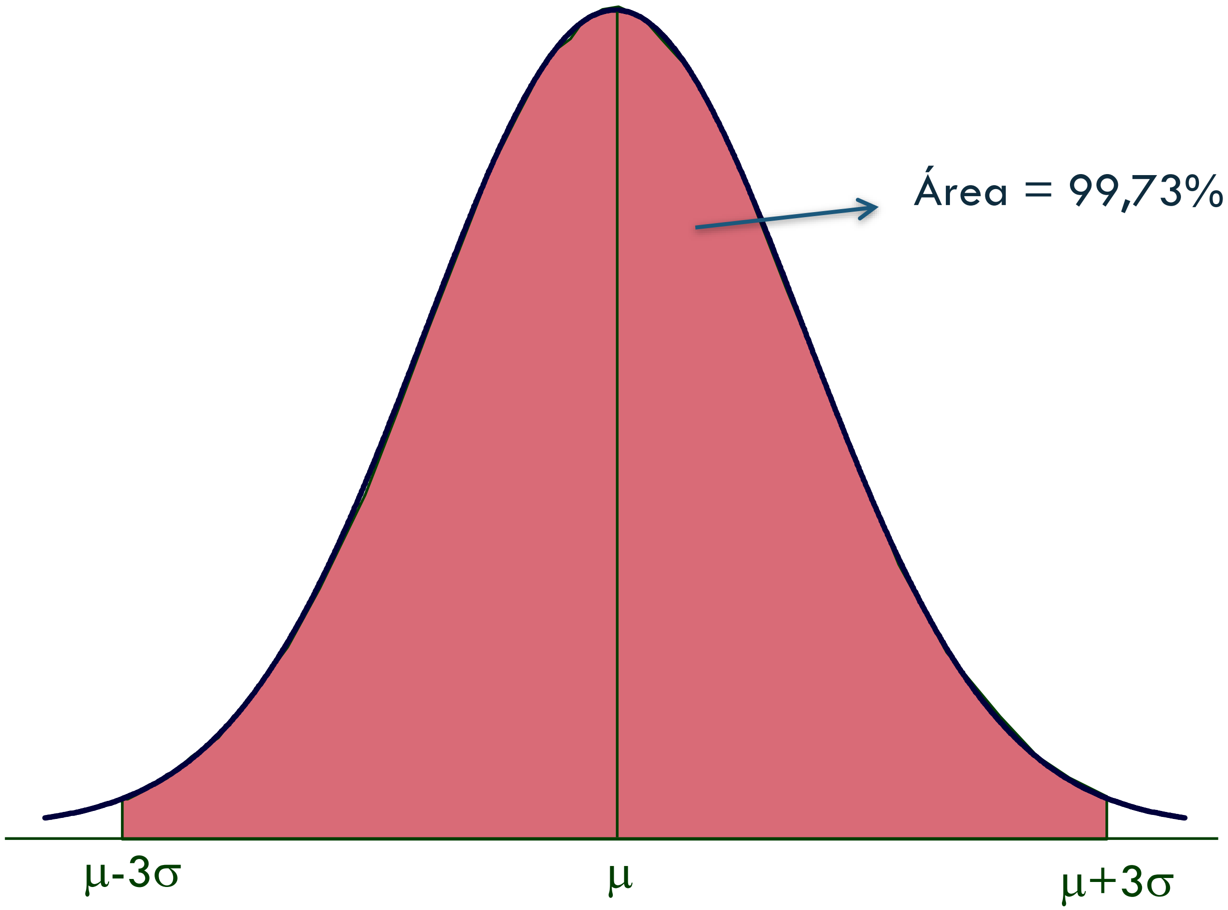


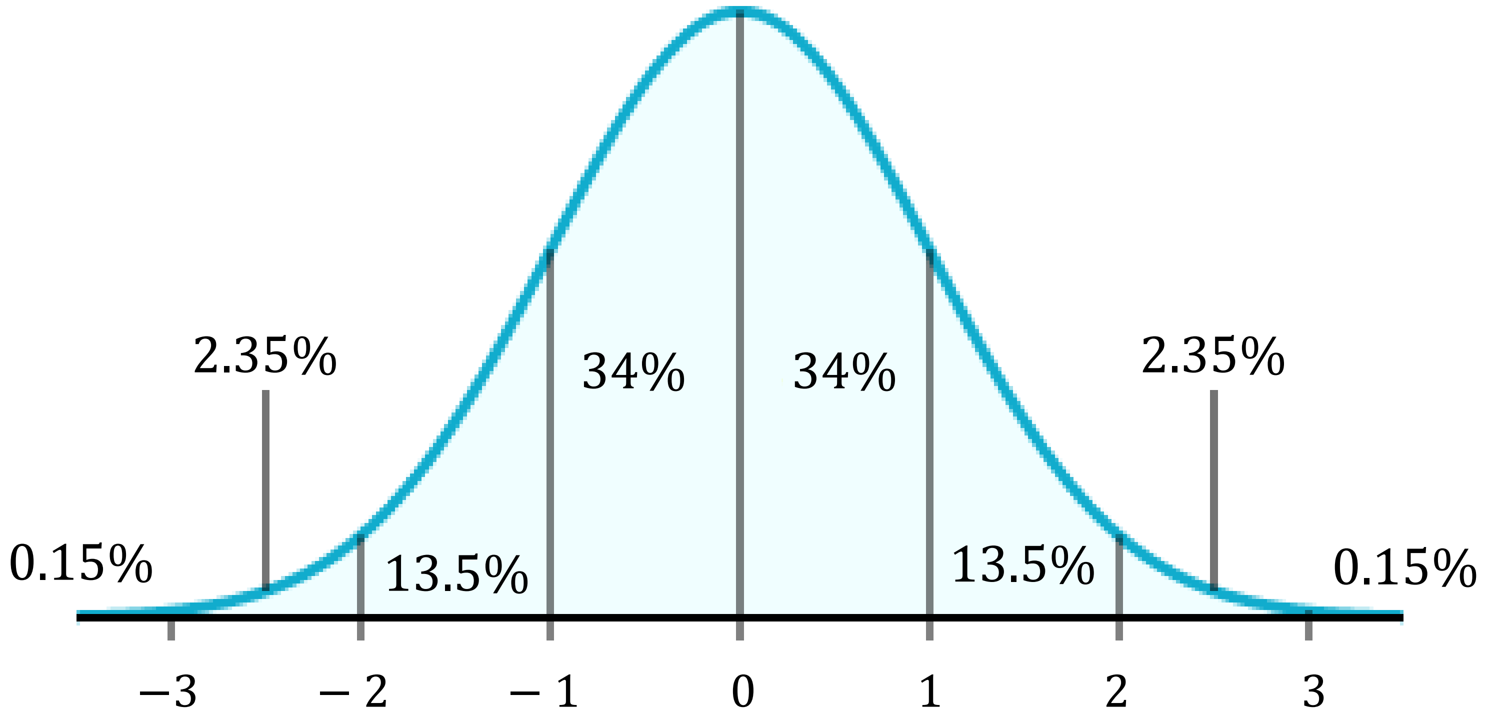
Score Z

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



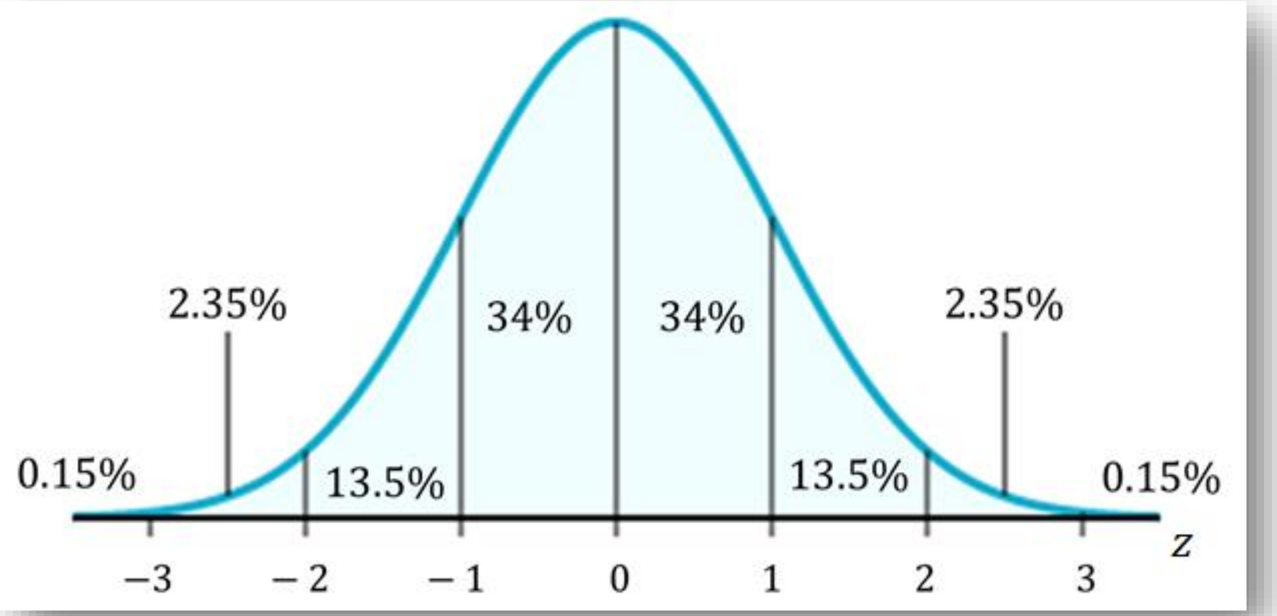
Aproximadamente 68% por cento dos dados estão dentro de 1 desvio padrão da média
Aproximadamente 95% dos dados estão dentro de 2 desvios padrão da média
Aproximadamente 99,7% dos dados estão dentro de 3 desvios-padrão da média

PECE Programa de Educação Continuada
Escola Politécnica da USP

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,

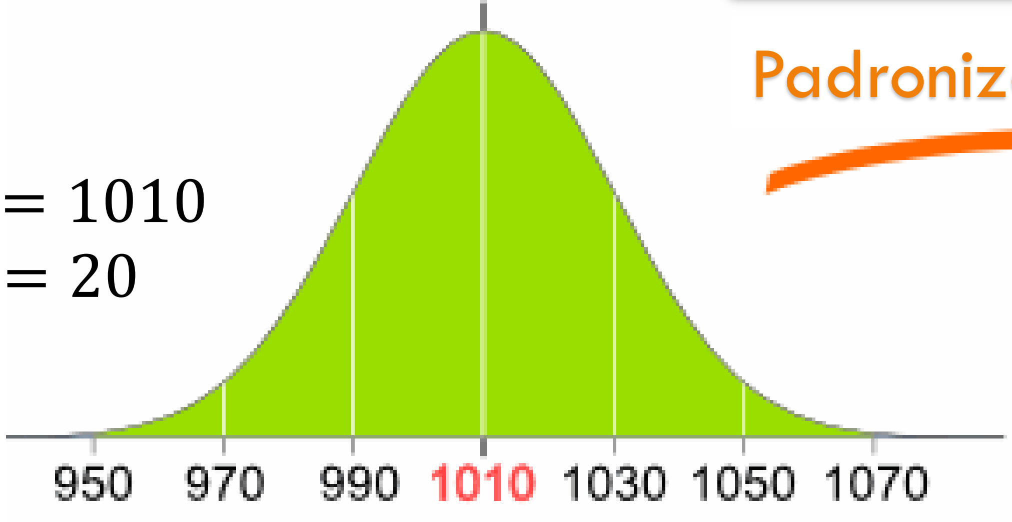
$$\varphi(z) = \frac{1}{2\pi} e^{-\frac{1}{2}z^2}$$

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

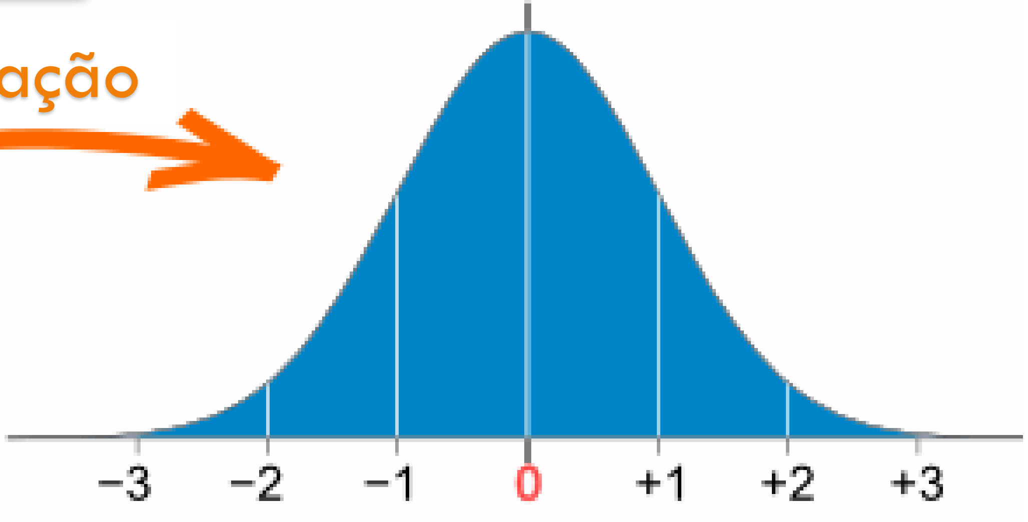
$$z = \frac{x - \mu}{\sigma}$$

Padronização

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



Distribuição normal



Distribuição normal padronizada

EXEMPLO

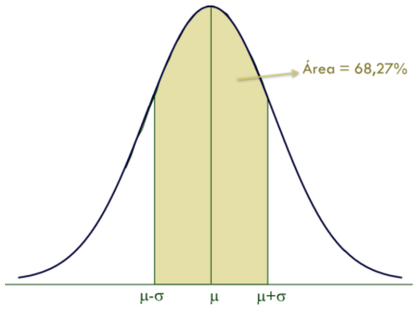
Considere que a glicemia de uma população de pessoas sadias tenha uma distribuição normal, com media 90mg e desvio padrão 5mg. Pode-se concluir que,

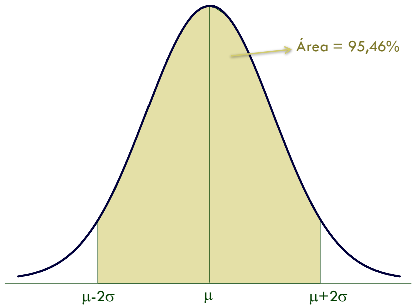
Aproximadamente $2/3$ $(\approx 68\%)$ da população de indivíduos sadios possui glicemia entre $\mu - \sigma = 90 - 5 = 85mg$ e $\mu + \sigma = 90 + 5 = 95mg$;

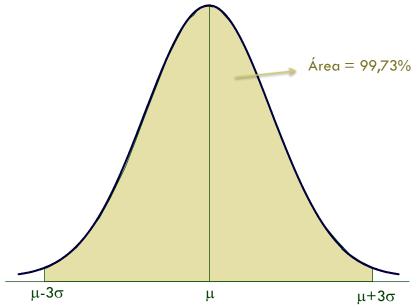
Grande parte $(\approx 95\%)$ da população de indivíduos sadios possui glicemia entre $\mu - 2\sigma = 90 - 10 = 80mg$ e $\mu + 2\sigma = 90 + 10 = 100mg$;

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

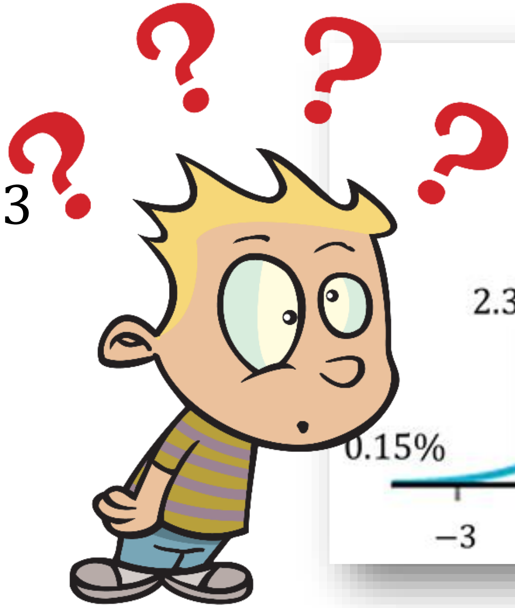
A idade da população de uma cidade é normalmente distribuída com média de 43 e desvio padrão 10. A cidade tem uma população de 5.000 habitantes. Quantos você esperaria ter entre 33 e 63 anos?

MAS, E SE...

A idade da população de uma cidade é normalmente distribuída com média de 40 e desvio padrão 10. A cidade tem uma população de 5.000 habitantes. Quantos você esperaria ter entre 33 e 63 anos?

$$z = \frac{x-\mu}{\sigma} = \frac{63-40}{10} = 2{,}3$$

$$z = \frac{x-\mu}{\sigma} = \frac{33-40}{10} = -1{,}3$$



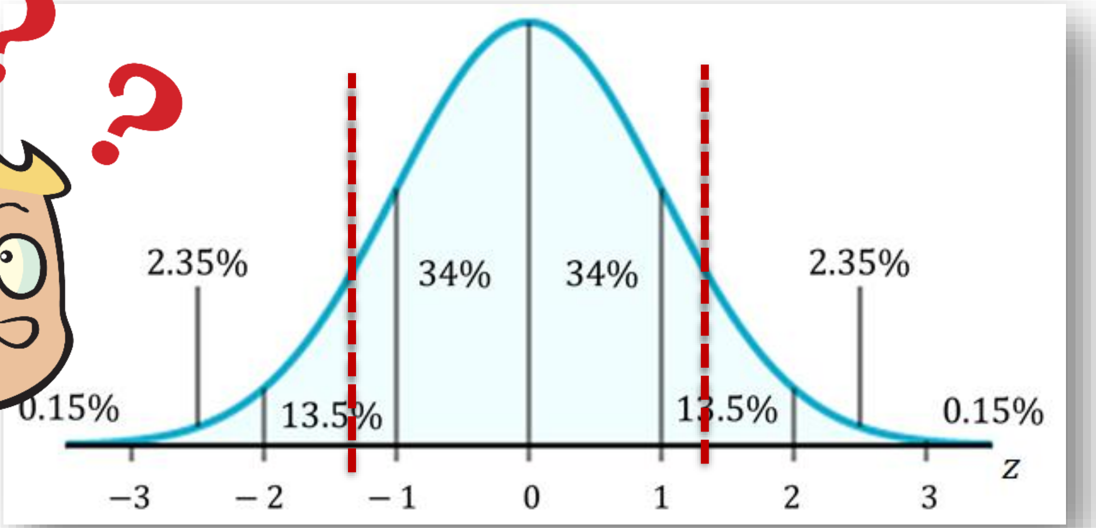


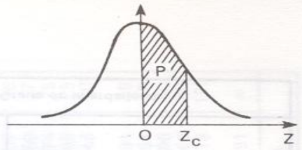
TABELA DA DISTRIBUIÇÃO NORMAL REDUZIDA

Programa de
ão Continuada
litécnica da USP

TÁBUA III

Distribuição normal reduzida: N(0;1)

Probabilidades p tais que $p = P(O < Z < Z_c)$



parte inteira e primeira decimal de Z_c	SEGUNDA DECIMAL DE Z_c										parte inteira e primeira decimal de Z_c
	0	1	2	3	4	5	6	7	8	9	
	p = 0										
0,0	00000	00399	00798	01197	01595	01994	02392	02790	03188	03586	0,0
0,1	03983	04380	04776	05172	05567	05962	06356	06749	07142	07535	0,1
0,2	07926	08317	08706	09095	09483	09871	10257	10642	11026	11409	0,2
0,3	11791	12172	12552	12930	13307	13683	14058	14431	14803	15173	0,3
0,4	15542	15910	16276	16640	17003	17364	17724	18082	18439	18793	0,4
0,5	19146	19497	19847	20194	20540	20884	21226	21566	21904	22240	0,5
0,6	22575	22907	23237	23565	23891	24215	24537	24857	25175	25490	0,6
0,7	25804	26115	26424	26730	27035	27337	27637	27935	28230	28524	0,7
0,8	28814	29103	29389	29673	29955	30234	30511	30785	31057	31327	0,8
0,9	31594	31859	32121	32381	32639	32894	33147	33398	33646	33891	0,9
1,0	34134	34375	34614	34850	35083	35314	35543	35769	35993	36214	1,0
1,1	36433	36650	36864	37076	37286	37493	37698	37900	38100	38298	1,1
1,2	38493	38686	38877	39065	39251	39435	39617	39796	39973	40147	1,2
1,3	40320	40490	40658	40824	40988	41149	41309	41466	41621	41774	1,3
1,4	41924	42073	42220	42364	42507	42647	42786	42922	43056	43189	1,4
1,5	43319	43448	43574	43699	43822	43943	44062	44179	44295	44408	1,5
1,6	44520	44630	44738	44845	44950	45053	45154	45254	45352	45449	1,6
1,7	45543	45637	45728	45818	45907	45994	46080	46164	46246	46327	1,7
1,8	46407	46485	46562	46638	46712	46784	46856	46926	46995	47062	1,8
1,9	47128	47193	47257	47320	47381	47441	47500	47558	47615	47670	1,9
2,0	47725	47778	47831	47882	47932	47982	48030	48077	48124	48169	2,0
2,1	48214	48257	48300	48341	48382	48422	48461	48500	48537	48574	2,1
2,2	48610	48645	48679	48713	48745	48778	48809	48840	48870	48899	2,2
2,3	48928	48956	48983	49010	49036	49061	49086	49111	49134	49158	2,3
2,4	49180	49202	49224	49245	49266	49286	49305	49324	49343	49361	2,4
2,5	49379	49396	49413	49430	49446	49461	49477	49492	49506	49520	2,5
2,6	49534	49547	49560	49573	49585	49598	49609	49621	49632	49643	2,6
2,7	49653	49664	49674	49683	49693	49702	49711	49720	49728	49736	2,7
2,8	49744	49752	49760	49767	49774	49781	49788	49795	49801	49807	2,8
2,9	49813	49819	49825	49831	49836	49841	49846	49851	49856	49861	2,9
3,0	49865	49869	49874	49878	49882	49886	49889	49893	49897	49900	3,0
3,1	49903	49906	49910	49913	49916	49918	49921	49924	49926	49929	3,1
3,2	49931	49934	49936	49938	49940	49942	49944	49946	49948	49950	3,2
3,3	49952	49953	49955	49957	49958	49960	49961	49962	49964	49965	3,3
3,4	49966	49968	49969	49970	49971	49972	49973	49974	49975	49976	3,4
3,5	49977	49978	49978	49979	49980	49981	49981	49982	49983	49983	3,5
3,6	49984	49985	49985	49986	49986	49987	49987	49988	49988	49989	3,6
3,7	49989	49990	49990	49990	49991	49991	49992	49992	49992	49992	3,7
3,8	49993	49993	49993	49994	49994	49994	49994	49995	49995	49995	3,8
3,9	49995	49995	49996	49996	49996	49996	49996	49996	49997	49997	3,9
4,0	49997	49997	49997	49997	49997	49997	49998	49998	49998	49998	4,0
4,5	49999	50000	50000	50000	50000	50000	50000	50000	50000	50000	4,5

parte inteira e primeira decimal de Z_c	SEGUNDA E TERCEIRA DECIMAIS DE Z_c										parte inteira e primeira decimal de Z_c
	05	15	25	35	45	55	65	75	85	95	
	p = 0,										
0,0	00199	00598	09997	01396	01795	02193	02591	02989	03387	03784	0,0
0,1	04181	04578	04974	05369	05764	06159	06553	06946	07339	07730	0,1
0,2	08121	08512	08901	09290	09677	10064	10450	10834	11218	11600	0,2
0,3	11982	12362	12741	13119	13495	13871	14244	14617	14988	15358	0,3
0,4	15726	16093	16458	16822	17184	17545	17903	18261	18500	18970	0,4
0,5	19322	19672	20021	20368	20712	21055	21396	21735	22073	22408	0,5
0,6	22741	23072	23401	23729	24054	24377	24697	25016	25333	25647	0,6
0,7	25959	26270	26577	26883	27186	27488	27786	28083	28377	28669	0,7
0,8	28959	29246	29531	29814	30094	30372	30648	30921	31192	31461	0,8
0,9	31727	31990	32252	32511	32767	33021	33273	33522	33769	34013	0,9

305

Fonte: Bussab e Morettin “Estatística Básica”, 5ª edição, Editora Saraiva, 2002.

EXEMPLO DE UTILIZAÇÃO DA TABELA

P (0 < Z < 0,45) = 0,17364

0,4

0,_5

parte inteira e primeira decimal de Z_c	SEGUNDA DECIMAL DE Z_C										parte inteira e primeira decimal de Z_c
	0	1	2	3	4	5	6	7	8	9	
	p = 0										
0,0	00000	00399	00798	01197	01595	01994	02392	02790	03188	03586	0,0
0,1	03983	04380	04776	05172	05567	05962	06356	06749	07142	07535	0,1
0,2	07926	08317	08706	09095	09483	09871	10257	10642	11026	11409	0,2
0,3	11791	12172	12552	12930	13307	13683	14058	14431	14803	15173	0,3
0,4	15542	15910	16276	16640	17003	17364	17724	18082	18439	18793	0,4
0,5	19146	19497	19847	20194	20540	20884	21226	21566	21904	22240	0,5

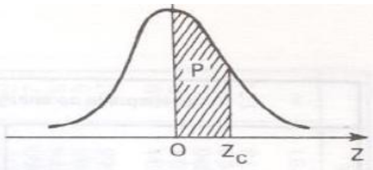
P (0 < Z < 0,45)

ACHE $P(0 \leq z \leq 2{,}3)$

TÁBUA III

Distribuição normal reduzida: N(0;1)

Probabilidades p tais que p = $P(0 < Z < Z_c)$



parte inteira e primeira decimal de Z_c	SEGUNDA DECIMAL DE Z_c										parte inteira e primeira decimal de Z_c
	0	1	2	3	4	5	6	7	8	9	
	p = 0										
0,0	00000	00399	00798	01197	01595	01994	02392	02790	03188	03586	0,0
0,1	03983	04380	04776	05172	05567	05962	06356	06749	07142	07535	0,1
0,2	07926	08317	08706	09095	09483	09871	10257	10642	11026	11409	0,2
0,3	11791	12172	12552	12930	13307	13683	14058	14431	14803	15173	0,3
0,4	15542	15910	16276	16640	17003	17364	17724	18082	18439	18793	0,4
0,5	19146	19497	19847	20194	20540	20884	21226	21566	21904	22240	0,5
0,6	22575	22907	23237	23565	23891	24215	24537	24857	25175	25490	0,6
0,7	25804	26115	26424	26730	27035	27337	27637	27935	28230	28524	0,7
0,8	28814	29103	29389	29673	29955	30234	30511	30785	31057	31327	0,8
0,9	31594	31859	32121	32381	32639	32894	33147	33398	33646	33891	0,9
1,0	34134	34375	34614	34850	35083	35314	35543	35769	35993	36214	1,0
1,1	36433	36650	36864	37076	37286	37493	37698	37900	38100	38298	1,1
1,2	38493	38686	38877	39065	39251	39435	39617	39796	39973	40147	1,2
1,3	40320	40490	40658	40824	40988	41149	41309	41466	41621	41774	1,3
1,4	41924	42073	42220	42364	42507	42647	42786	42922	43056	43189	1,4
1,5	43319	43448	43574	43699	43822	43943	44062	44179	44295	44408	1,5
1,6	44520	44630	44738	44845	44950	45053	45154	45254	45352	45449	1,6
1,7	45543	45637	45728	45818	45907	45994	46080	46164	46246	46327	1,7
1,8	46407	46485	46562	46638	46712	46784	46856	46926	46995	47062	1,8
1,9	47128	47193	47257	47320	47381	47441	47500	47558	47615	47670	1,9
2,0	47725	47778	47831	47882	47932	47982	48030	48077	48124	48169	2,0
2,1	48214	48257	48300	48341	48382	48422	48461	48500	48537	48574	2,1
2,2	48610	48645	48679	48713	48745	48778	48809	48840	48870	48899	2,2
2,3	$\boxed{48928}$	48956	48983	49010	49036	49061	49086	49111	49134	49158	2,3

$$P(0 \leq z \leq 2{,}3) = 0{,}48928$$

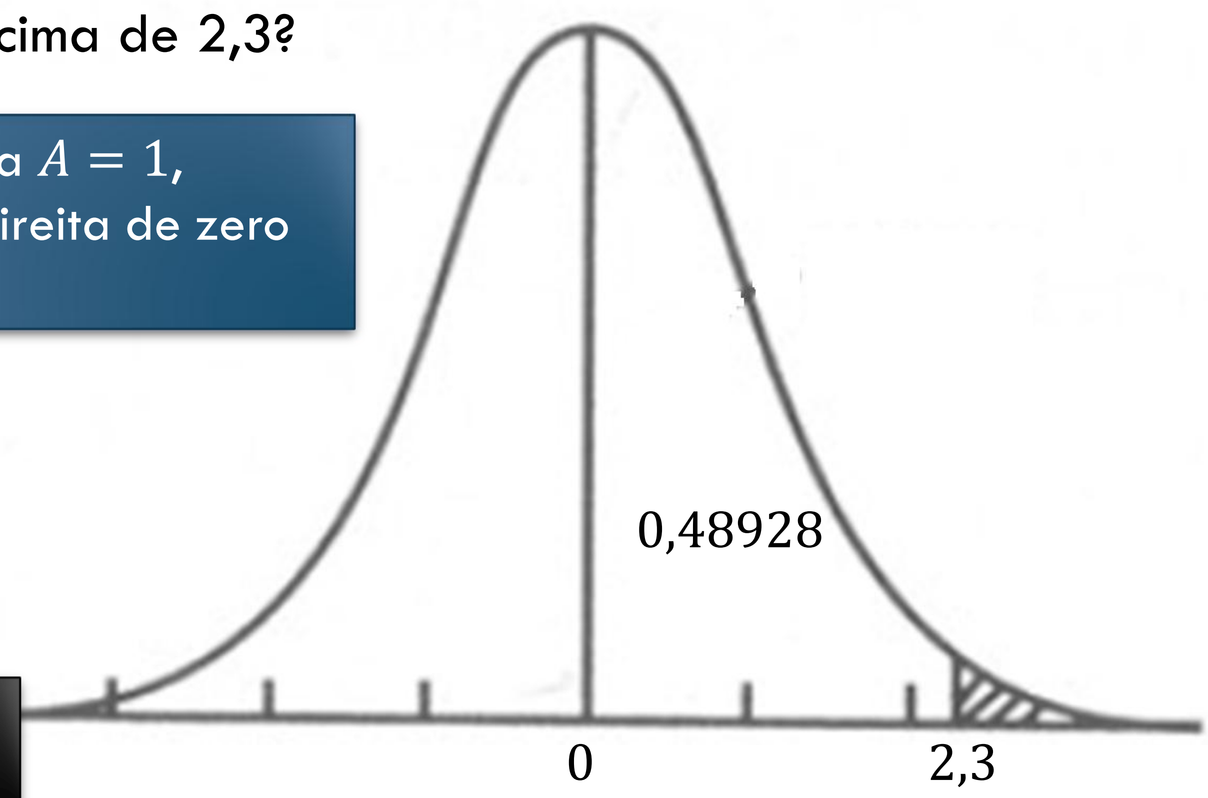
EXEMPLO

Qual a área correspondente a valores de z acima de 2,3?

A curva toda tem área $A = 1$, portanto a área da direita de zero $A/2 = 0{,}5$

Na tabela de curva normal, verifica-se que a área entre $z = 0$ e $z = 2{,}3$ é $A = 0{,}48928$

A área à direita de 2,3 é, portanto, $0{,}5 - 0{,}48928 = 0{,}01072$



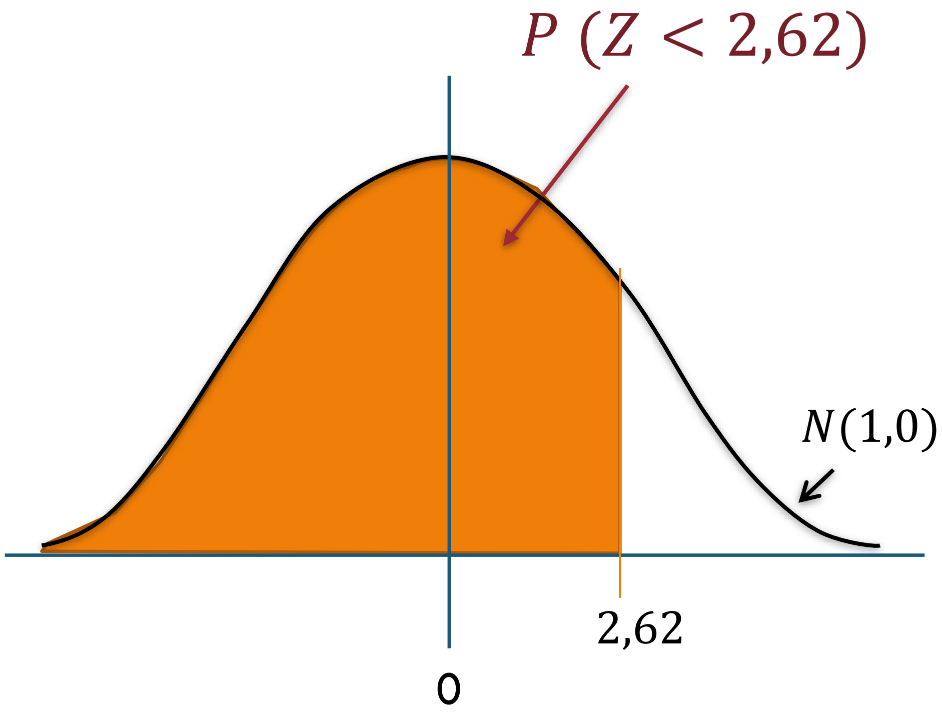
EXEMPLO

1. Considerando a distribuição normal reduzida, calcule as seguintes probabilidades:
   A. $P(Z \leq 2{,}62)$
   B. $P(Z \leq -1{,}45)$
   C. $P(Z > 1{,}45)$
   D. $P(-1{,}5 \leq Z \leq 2{,}5)$

Fonte: Krishnamoorthi, K.S. “Reliability Methods for Engineers”, 1ª edição, ASQC Quality Press, 1992.

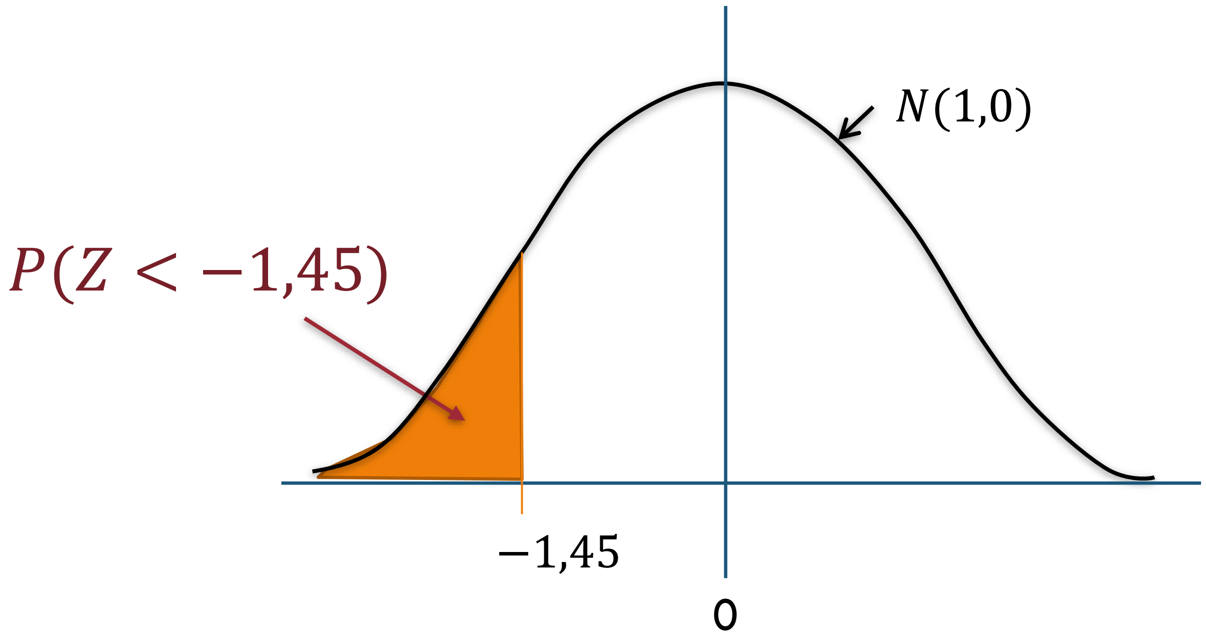
A. Da tabela da distribuição normal reduzida tem-se:

$$P(Z \leq 2{,}62) \ = \ 0{,}5 + 0{,}4956 \ = \ 0{,}9956$$



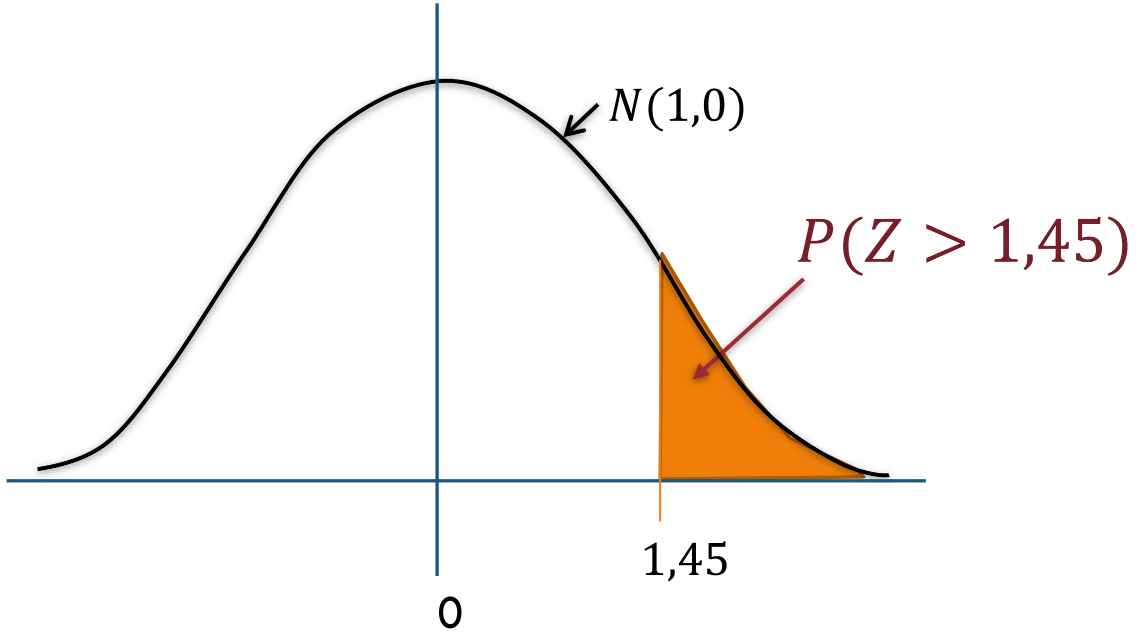
B. Da tabela da distribuição normal tem-se:

$$P(Z \leq -1{,}45) = 0{,}5 - 0{,}4265 = 0{,}0735$$



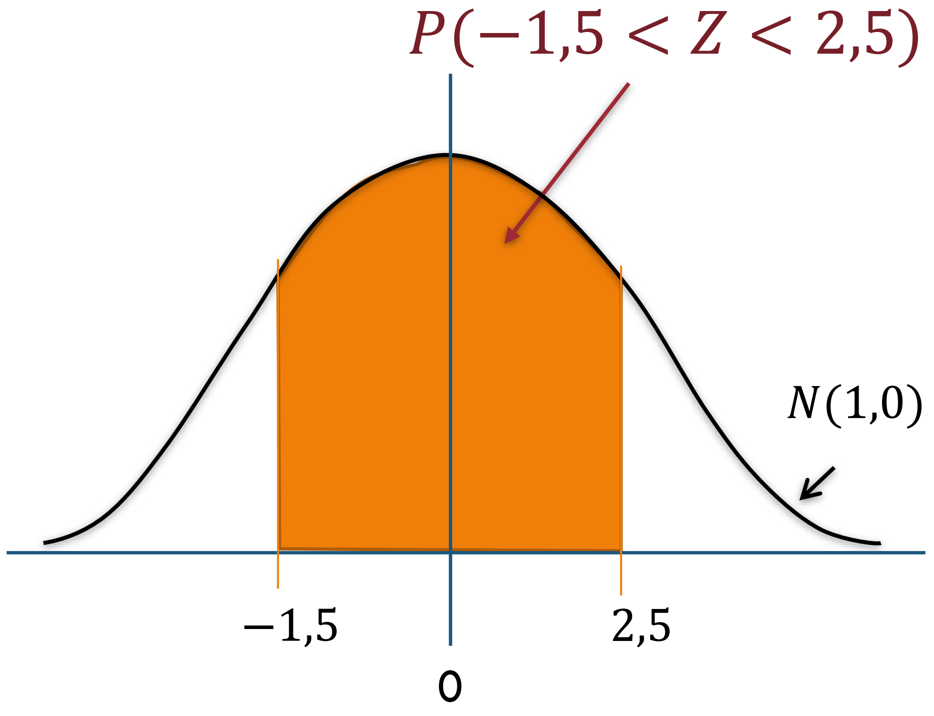
C. Através da propriedade de simetria da distribuição normal tem-se:

$$P(Z > 1{,}45) = P(Z \leq -1{,}45) = 0{,}0735$$



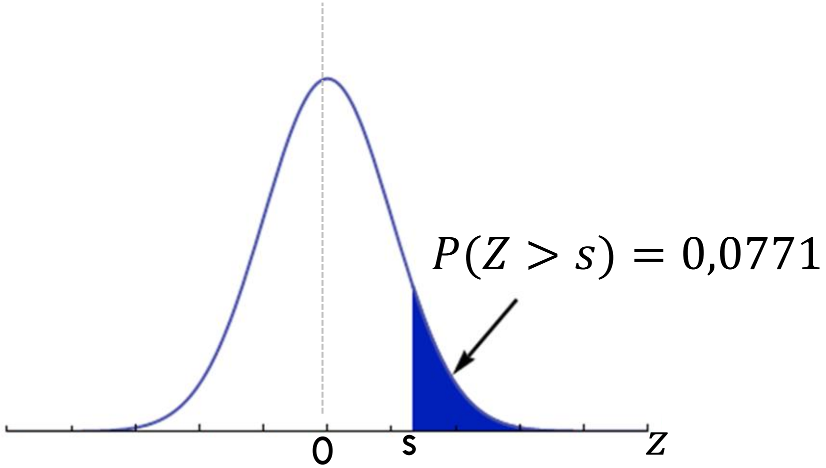
D. Da tabela da distribuição normal reduzida tem-se:

$$P(-1{,}5 \leq Z \leq 2{,}5) = P(-1{,}5 \leq Z \leq 0) + P(0 \leq Z \leq 2{,}5)$$
$$= 0{,}4332 + 0{,}4938 = 0{,}927$$



EXERCÍCIO

Dado que $P(\,Z > s) \;=\; 0{,}0771$, determine s.



Da tabela da distribuição normal reduzida:

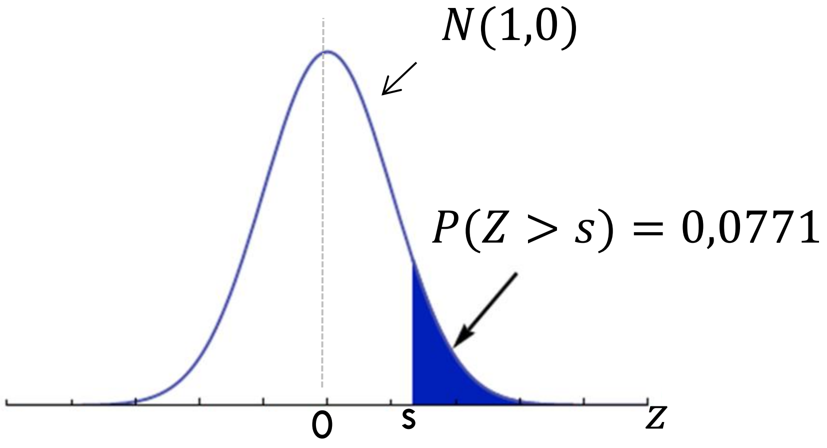
$$P(0 \leq Z \leq s) = 0{,}5 - P(Z > s) = 0{,}5 - 0{,}0771 = 0{,}4229$$

Na tabela em referência tem-se:

$$P(0 \leq Z \leq 1{,}42) = 0{,}4222$$

$$P(0 \leq Z \leq 1{,}43) = 0{,}4236$$

Através de uma interpolação linear tem-se: $\boldsymbol{s = 1{,}425}$



EXEMPLO PARA ESTUDAR

2. Selecionar, aleatoriamente, de uma certa universidade, um estudante do sexo masculino. Seja **x** o valor de sua altura, em centímetros. Admitindo que nesta universidade os estudantes têm altura média de 170 cm com desvio padrão de 10 cm,

   A. qual a probabilidade do estudante sorteado ter altura superior a 185 cm?

   B. qual a probabilidade do estudante sorteado ter menos de190 cm?

EXERCÍCIO DE DISTRIBUIÇÃO NORMAL

Supondo que temos dois jogadores, Mário e Ana, e a habilidade de jogo de cada um deles é definida, respectivamente, por $MarSkill = 12.5$ e \`$AnaSkill = 15$. Esses números parecem ser completamente arbitrários, e a escala na qual medimos a habilidade é de fato arbitrária. O que importa, no entanto, é como os valores das habilidades se comparam entre os jogadores. Atribuímos a Ana um maior valor de habilidade para indicar que ela é a jogadora mais forte.

Mas agora nos deparamos com o primeiro de nossos desafios: o jogador mais forte em um jogo de videogame nem sempre é o vencedor. Se Ana e Fred jogassem muitos jogos um contra o outro, esperaríamos que Ana vencesse mais da metade deles, mas não necessariamente vencesse todos eles. Podemos capturar a variabilidade no resultado de um jogo, introduzindo a noção de *performance* para cada jogador, que expressa o quão bem eles jogaram um jogo específico. O jogador com maior performance para um jogo específico será o vencedor desse jogo. Um jogador com um alto nível de habilidade tenderá a ter uma alta performance, mas sua performance real variará de um jogo para outro. Considere a performance de Ana como $AnaPerf$ e a performance de Mario por $MarPerf$, com variância constante e igual para os dois jogadores, $\sigma^2 = 25$.

Agora temos que definir o vencedor do jogo. Para isso, usaremos uma variável binária \`AnaVence\`, que vale \`True\` se Ana for a vencedora e \`False\` se Mário for o vencedor. O valor dessa variável é determinado por qual das duas variáveis \`AnaPerf\` e \`MarPerf\` é maior - será \`True\` se \`AnaPerf\` for maior ou \`False\` caso contrário,

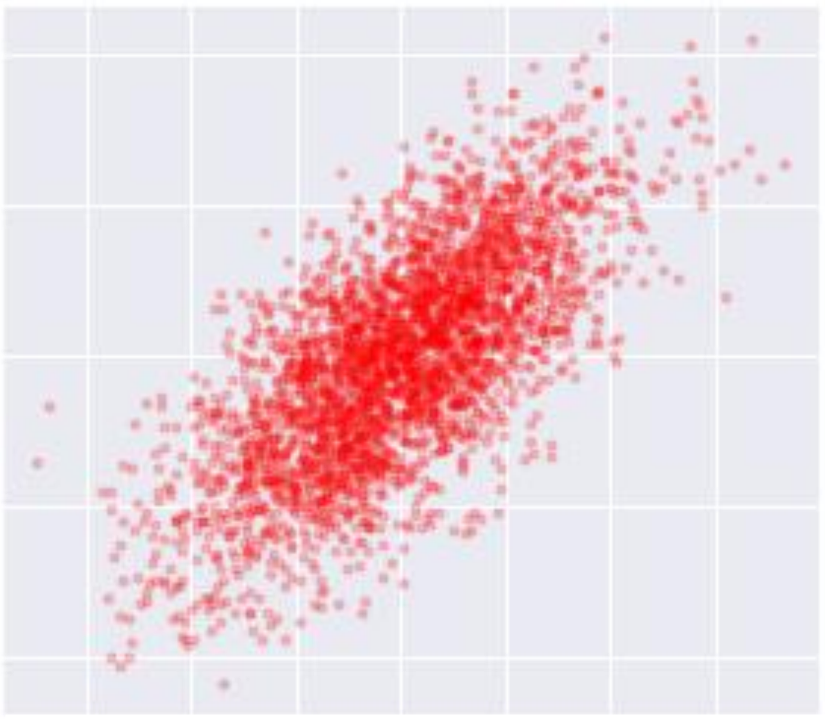
Dada a história acima, construa os seguintes gráficos:

1. Construa as curvas de distribuição do desempenho de Ana e de Mário;
2. Faça um gráfico desempenho de Ana versus desempenho de Mário e plote os valores obtidos em 1000 jogos. Coloque também uma linha dividindo a área em que a Ana da Área que o Mário vence;
3. Se repetirmos o processo de amostragem muitas vezes, como fizemos no item anterior (1000 vezes), a fração de vezes que $AnaVence$ for $True$ fornece, aproximadamente, a probabilidade de Ana ganhar um jogo. Calcule essa probabilidade aproximada.
4. O que acontece com a probabilidade do item anterior se o desempenho da Ana ou do Mário ficarem instáveis? Obs. mude a variância de cada um separadamente e analise a resposta.

DISTRIBUIÇÃO GAUSSIANA MULTIVARIADA

Vetor de dados de entrada com diversas características

PESO X ALTURA



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

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

PECE Programa de Educação Continuada
Escola Politécnica da USP

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

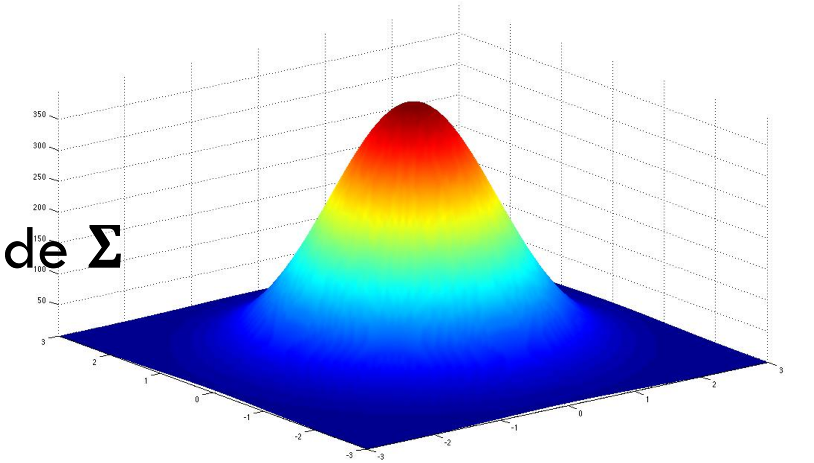
$$p(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}} e^{-\frac{1}{2}\left[(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right]}$$

$\boldsymbol{x}$ 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

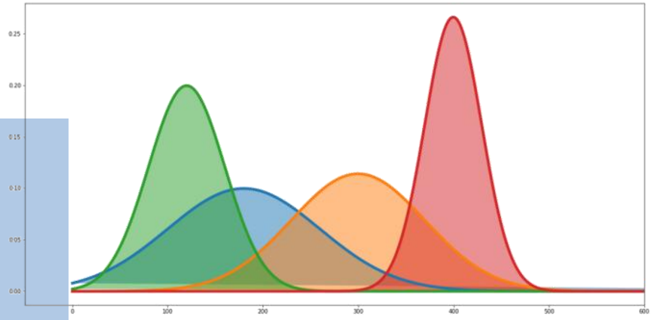
$$\mathcal{L}(\boldsymbol{\mu},\boldsymbol{\Sigma}) = \prod_{i=1}^{m} \frac{1}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}} e^{-\frac{1}{2}\left[(\boldsymbol{x}^{i}-\boldsymbol{\mu})^{T}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}^{i}-\boldsymbol{\mu})\right]}$$

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

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

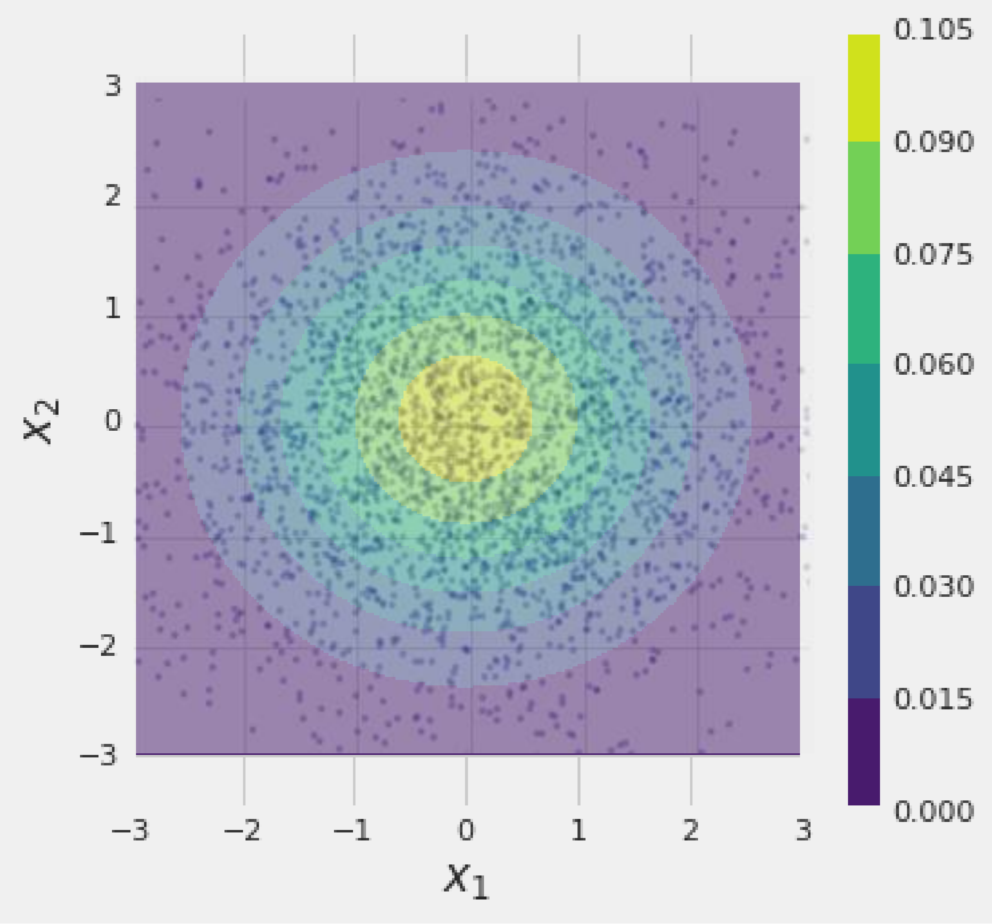
$$\frac{\partial L(\boldsymbol{\mu},\boldsymbol{\Sigma})}{\partial \boldsymbol{\mu}} = -\boldsymbol{\Sigma}^{-1}\sum_{i=1}^{m}(\boldsymbol{x}^{i}-\boldsymbol{\mu}) \quad \rightarrow \quad \hat{\boldsymbol{\mu}} = \frac{1}{m}\sum_{i}\boldsymbol{x}^{i}$$

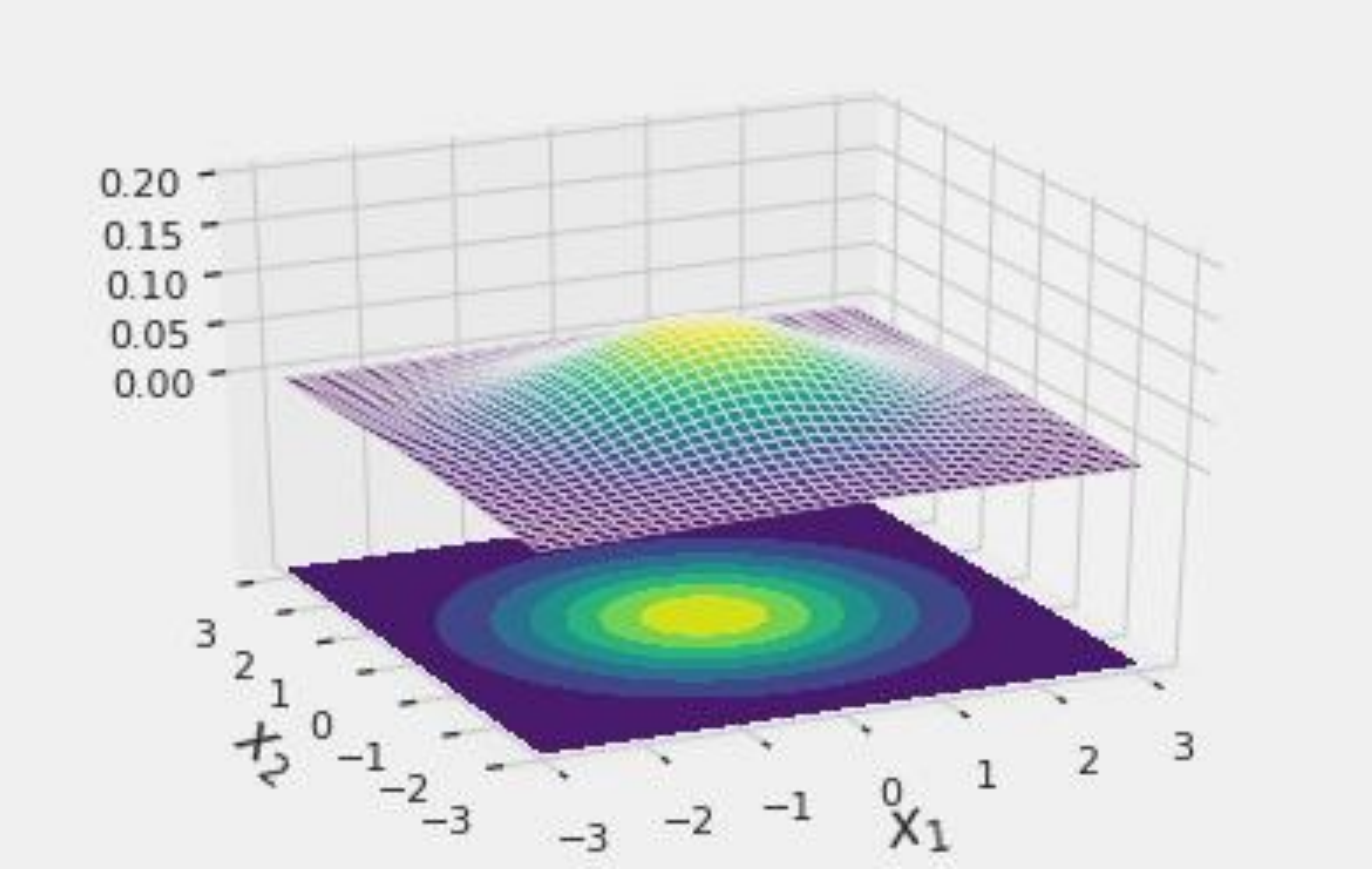
$$\frac{\partial L(\boldsymbol{\mu},\boldsymbol{\Sigma})}{\partial \boldsymbol{\Sigma}} = m - \boldsymbol{\Sigma}^{-1}\sum_{i=1}^{m}\left[(\boldsymbol{x}^{i}-\boldsymbol{\mu})^{T}(\boldsymbol{x}^{i}-\boldsymbol{\mu})\right] = \boldsymbol{0} \quad \rightarrow \quad \hat{\boldsymbol{\Sigma}} = \frac{1}{m}\sum_{i}(\boldsymbol{x}^{i}-\boldsymbol{\mu})^{T}(\boldsymbo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

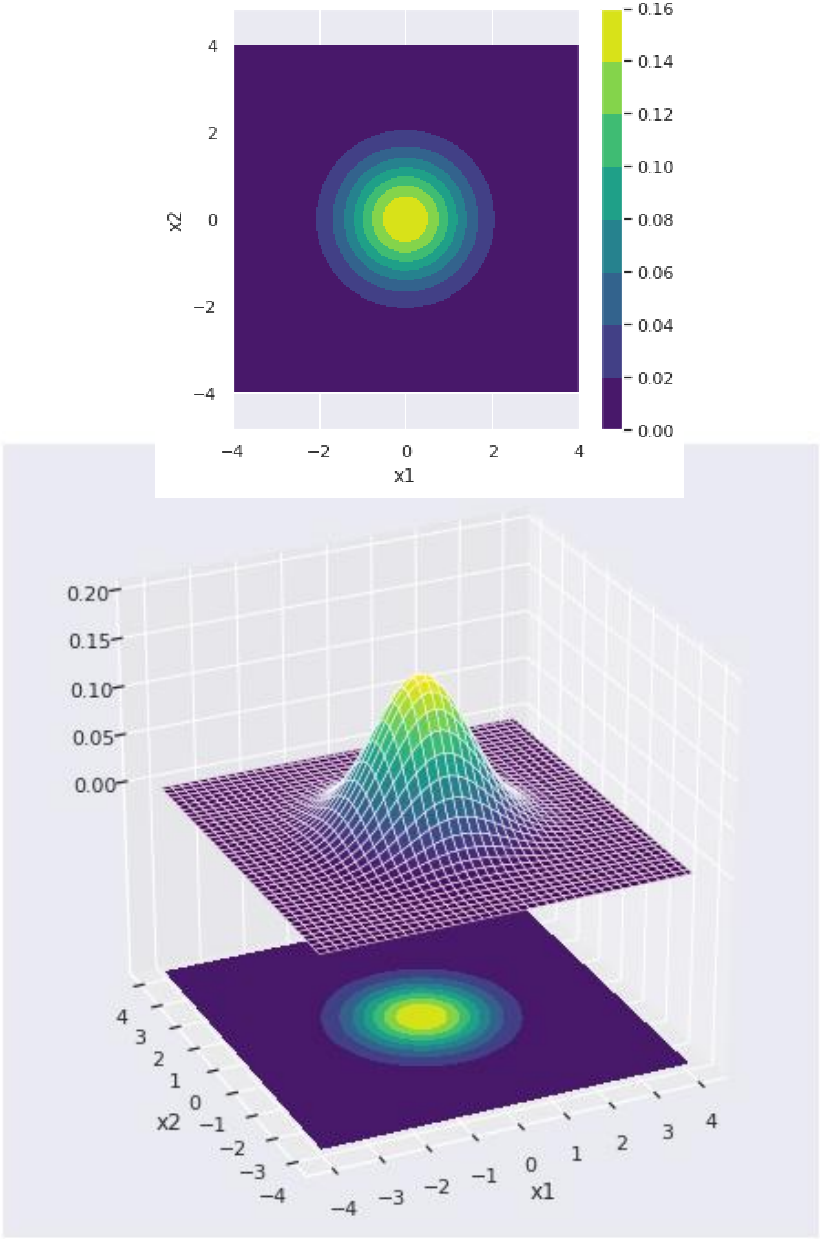


EXEMPLO

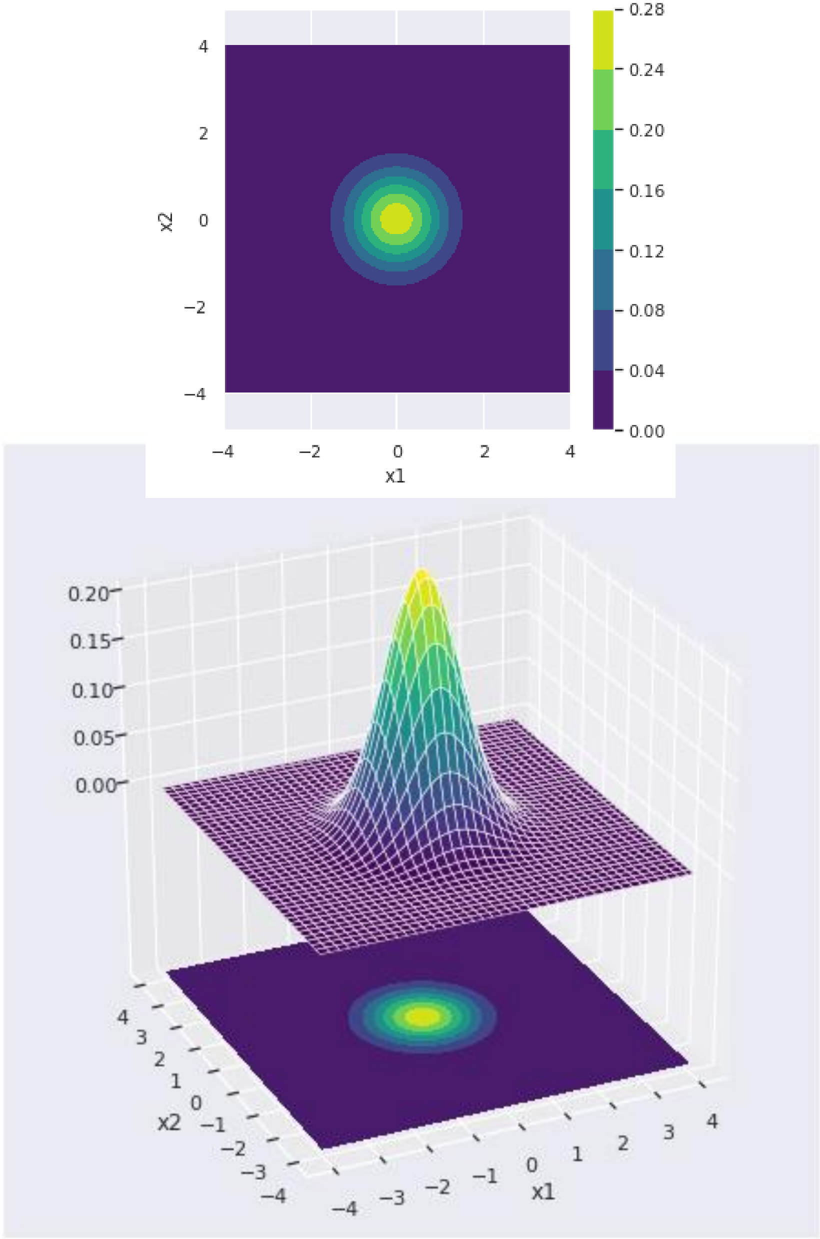
$$\boldsymbol{\mu} = \begin{pmatrix} 0. \\ 0. \end{pmatrix} \quad \boldsymbol{\Sigma} = \begin{pmatrix} 1.7 & 0. \\ 0. & 1.3 \end{pmatrix}$$



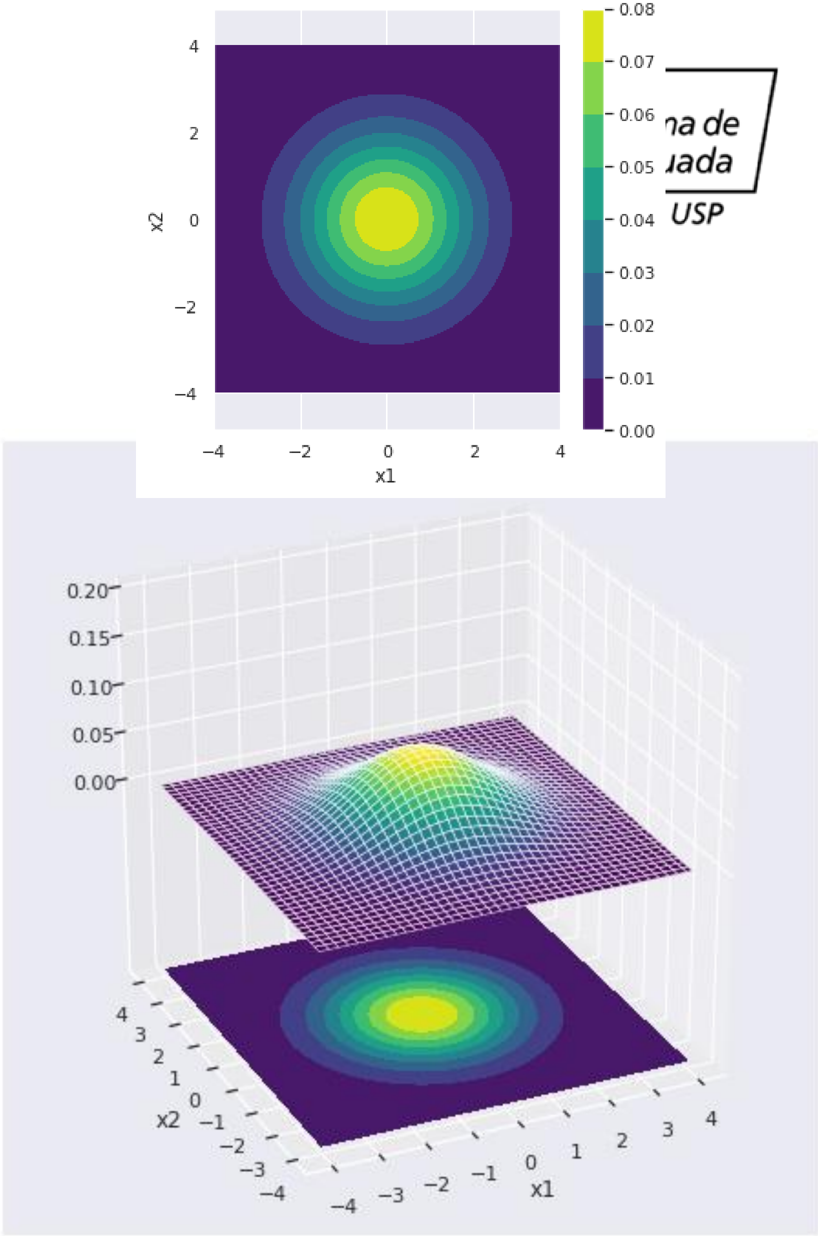




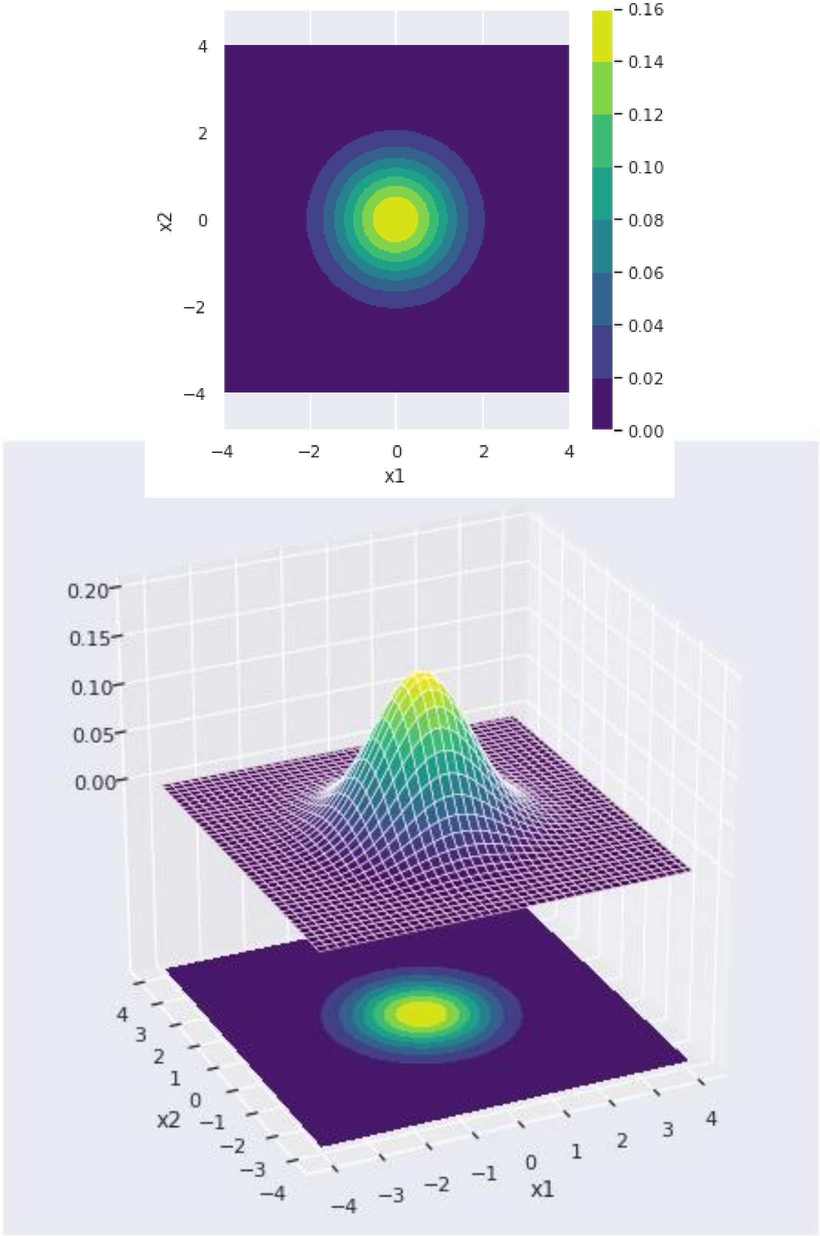
$$\boldsymbol{\mu} = \begin{pmatrix} 0. \\ 0. \end{pmatrix} \quad \boldsymbol{\Sigma} = \begin{pmatrix} 1. & 0. \\ 0. & 1. \end{pmatrix}$$



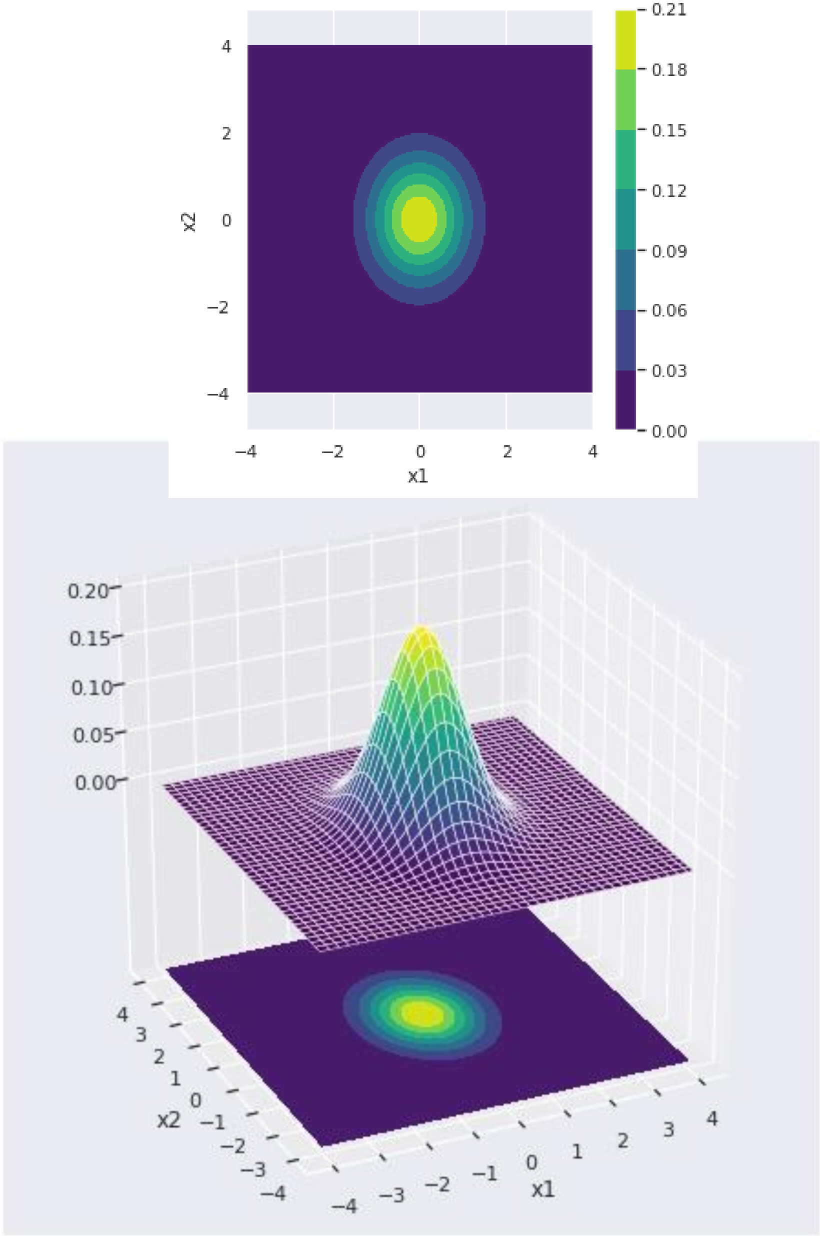
$$\boldsymbol{\mu} = \begin{pmatrix} 0. \\ 0. \end{pmatrix} \quad \boldsymbol{\Sigma} = \begin{pmatrix} 0.6 & 0. \\ 0. & 0.6 \end{pmatrix}$$



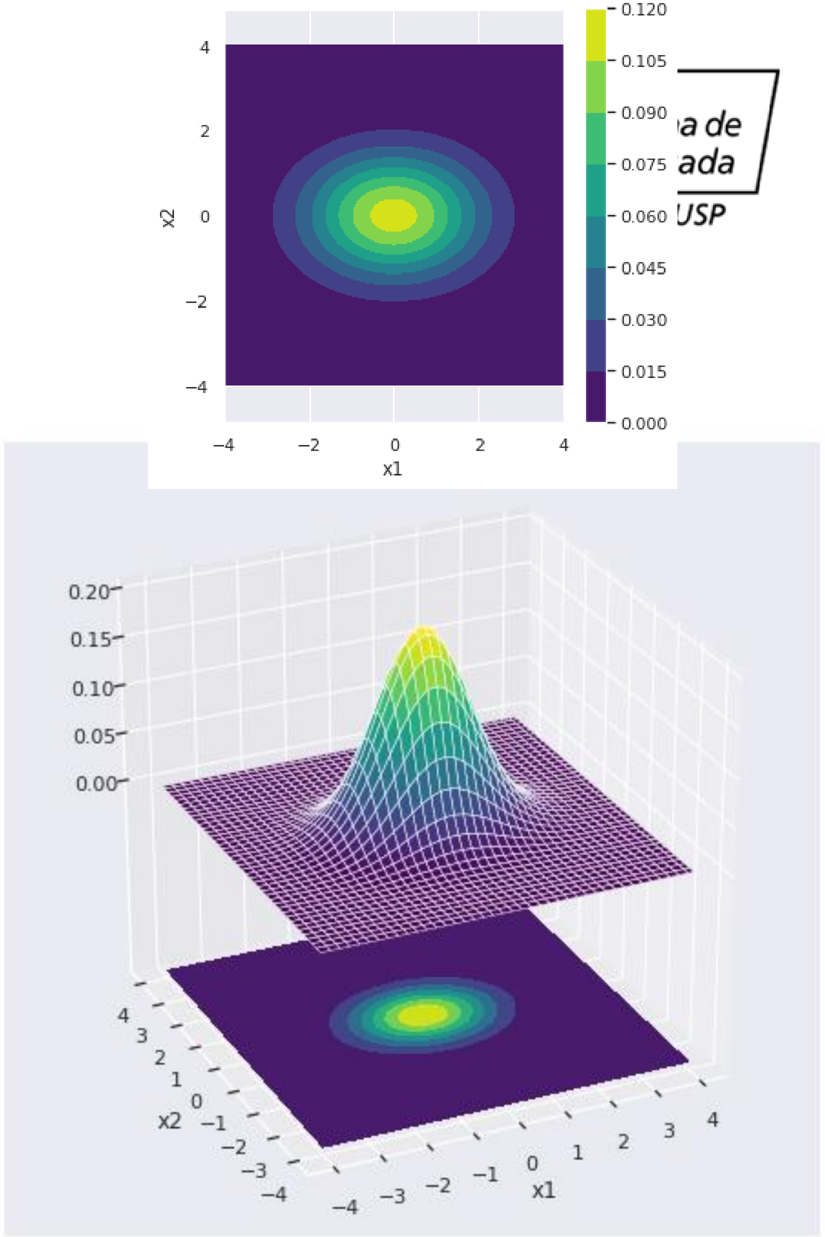
$$\boldsymbol{\mu} = \begin{pmatrix} 0. \\ 0. \end{pmatrix} \quad \boldsymbol{\Sigma} = \begin{pmatrix} 2. & 0. \\ 0. & 2. \end{pmatrix}$$



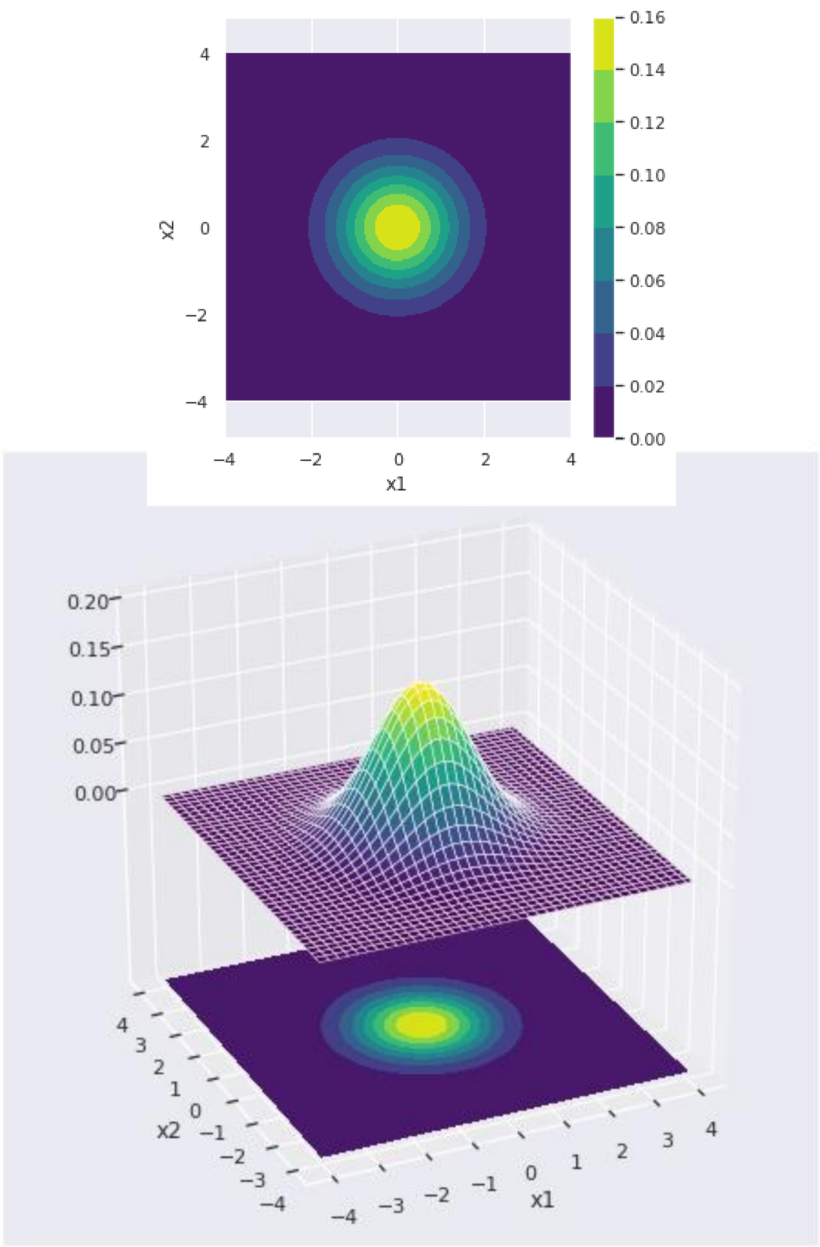
$$\boldsymbol{\mu} = \begin{pmatrix} 0. \\ 0. \end{pmatrix} \quad \boldsymbol{\Sigma} = \begin{pmatrix} 1. & 0. \\ 0. & 1. \end{pmatrix}$$



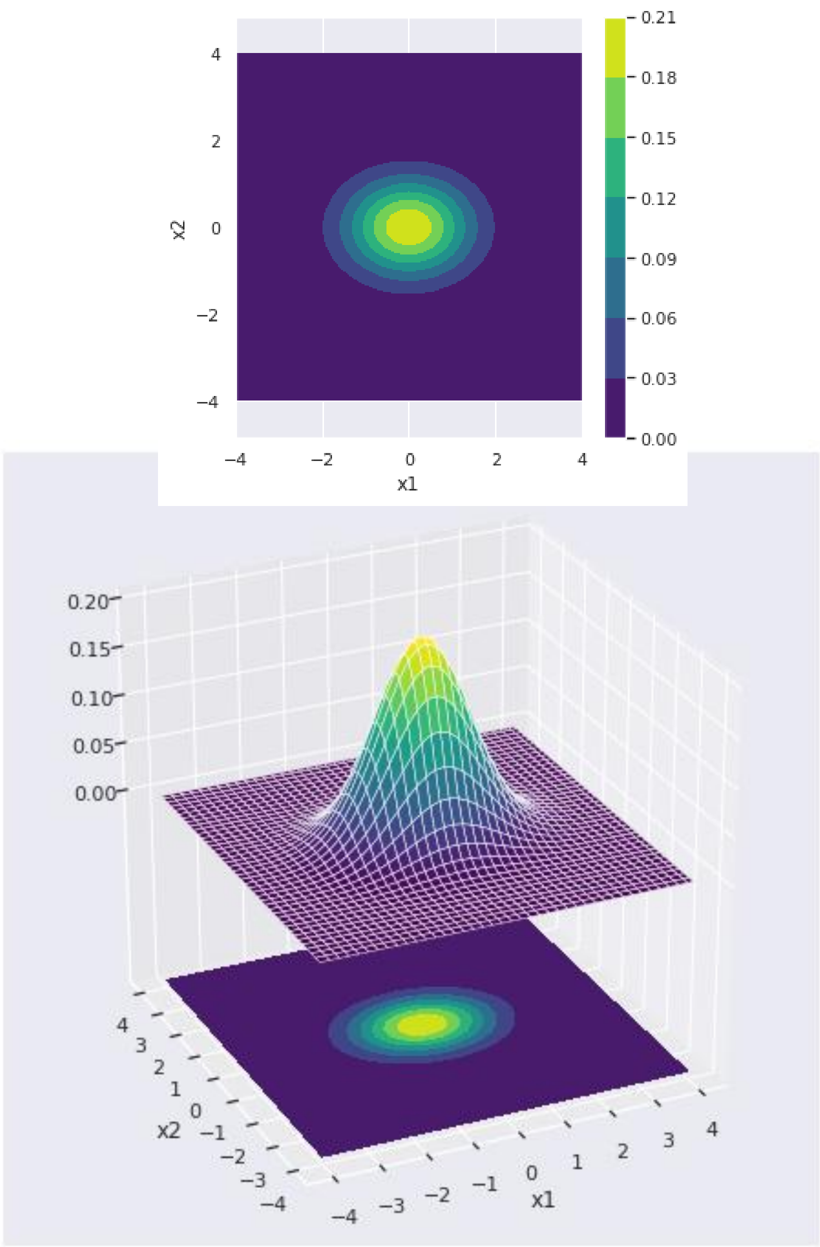
$$\boldsymbol{\mu} = \begin{pmatrix} 0. \\ 0. \end{pmatrix} \quad \boldsymbol{\Sigma} = \begin{pmatrix} 0.6 & 0. \\ 0. & 1. \end{pmatrix}$$



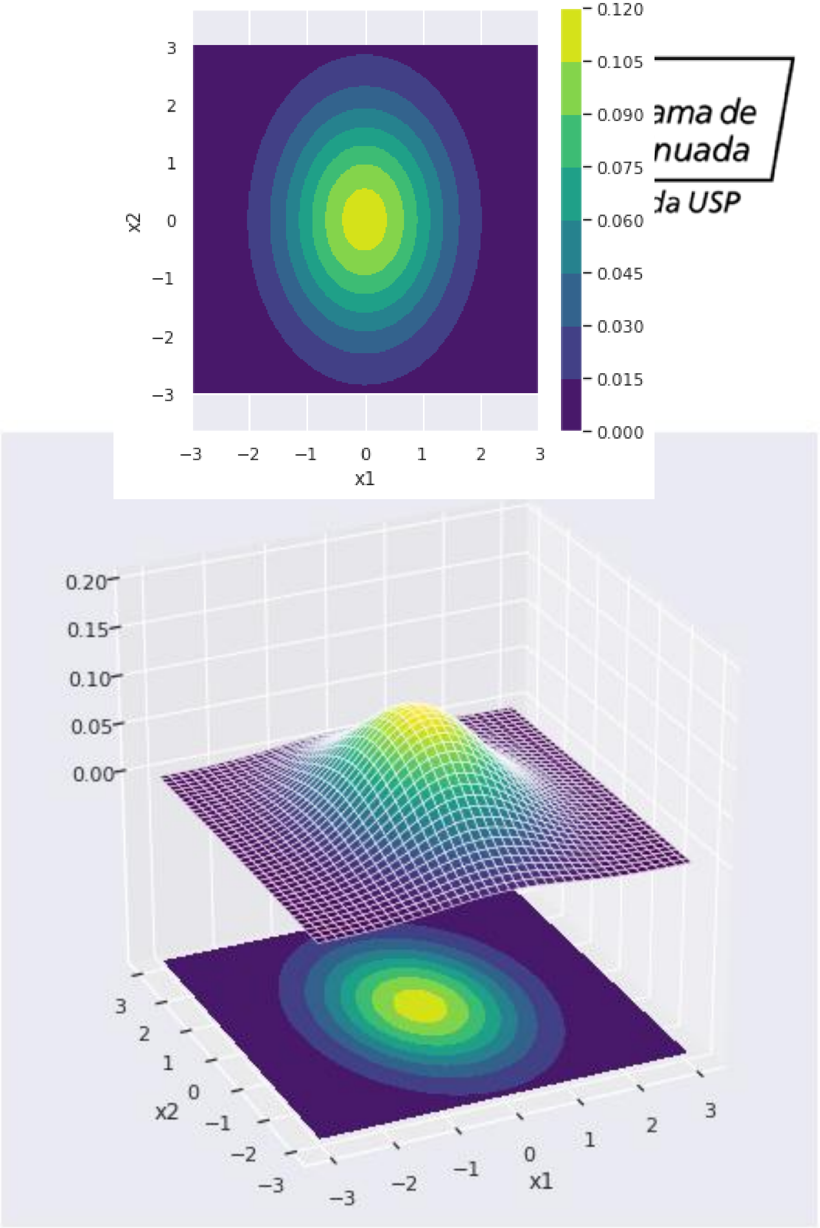
$$\boldsymbol{\mu} = \begin{pmatrix} 0. \\ 0. \end{pmatrix} \quad \boldsymbol{\Sigma} = \begin{pmatrix} 2. & 0. \\ 0. & 1. \end{pmatrix}$$



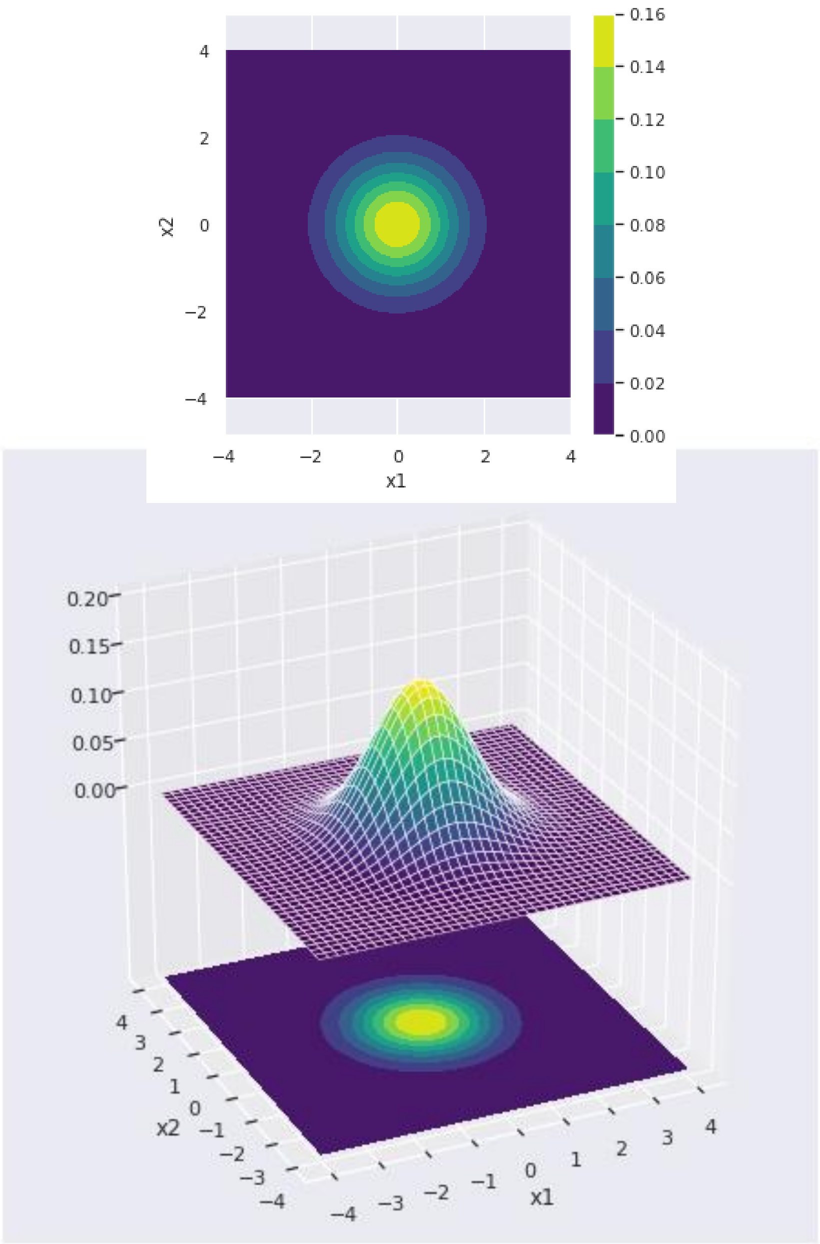
$$\boldsymbol{\mu} = \begin{pmatrix} 0. \\ 0. \end{pmatrix} \quad \boldsymbol{\Sigma} = \begin{pmatrix} 1. & 0. \\ 0. & 1. \end{pmatrix}$$



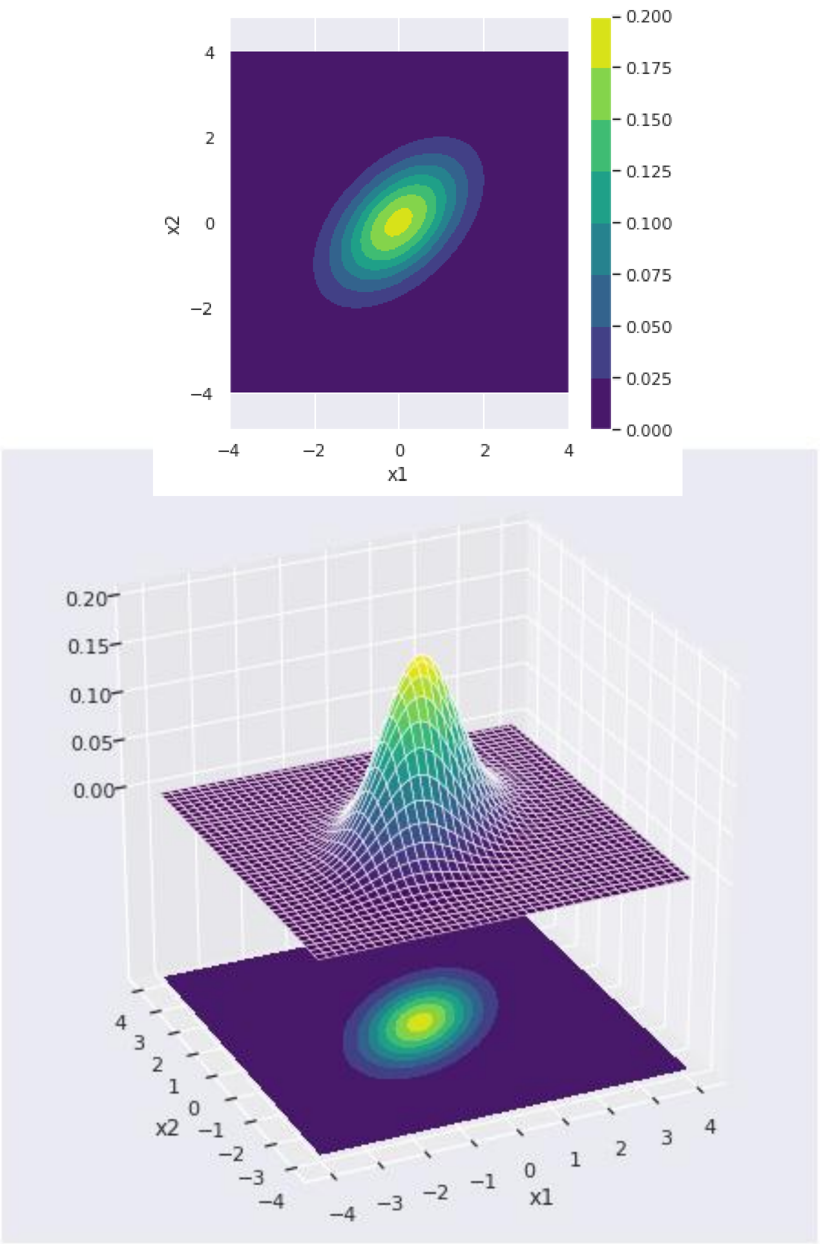
$$\boldsymbol{\mu} = \begin{pmatrix} 0. \\ 0. \end{pmatrix} \quad \boldsymbol{\Sigma} = \begin{pmatrix} 1. & 0. \\ 0. & 0.6 \end{pmatrix}$$



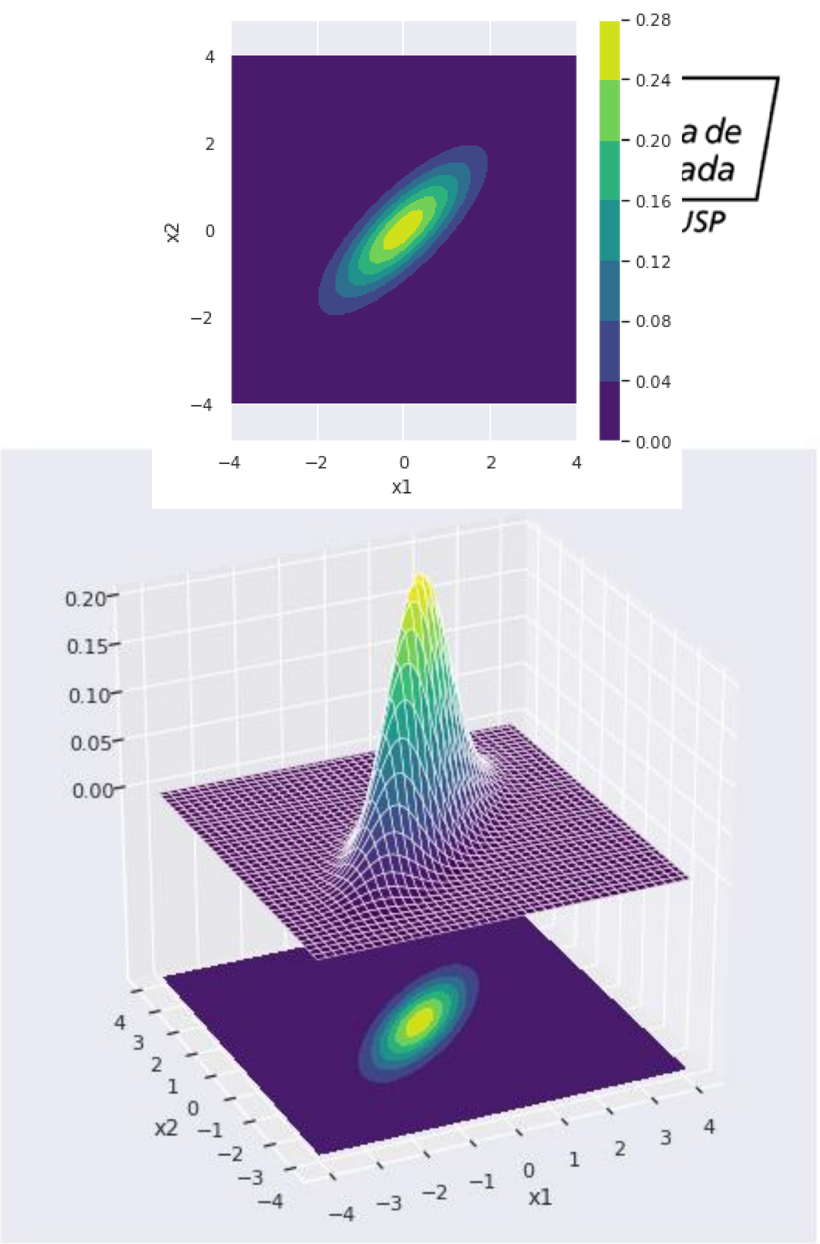
$$\boldsymbol{\mu} = \begin{pmatrix} 0. \\ 0. \end{pmatrix} \quad \boldsymbol{\Sigma} = \begin{pmatrix} 1. & 0. \\ 0. & 2. \end{pmatrix}$$



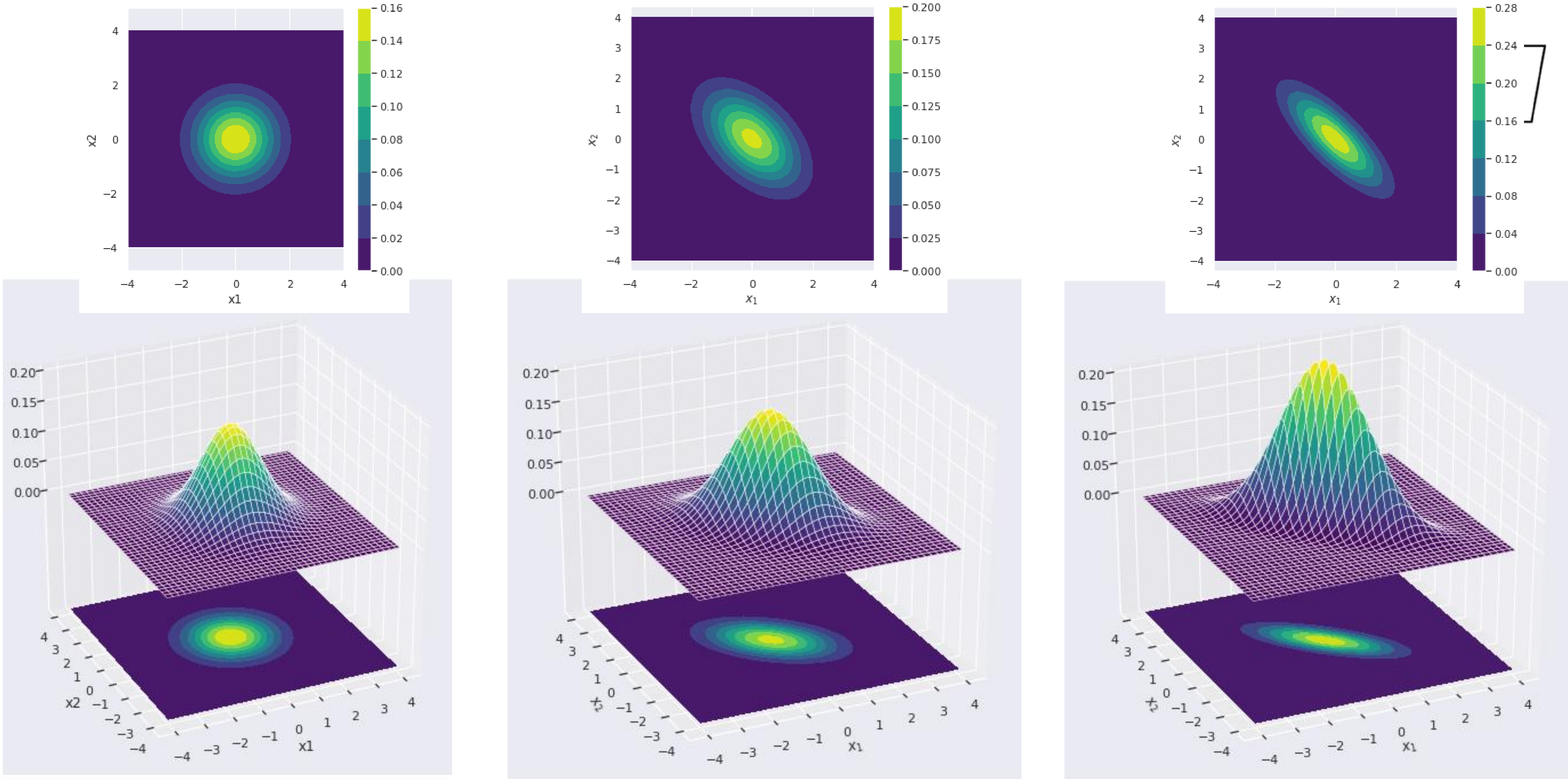
$$\boldsymbol{\mu} = \begin{pmatrix} 0. \\ 0. \end{pmatrix} \quad \boldsymbol{\Sigma} = \begin{pmatrix} 1. & 0. \\ 0. & 1. \end{pmatrix}$$



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



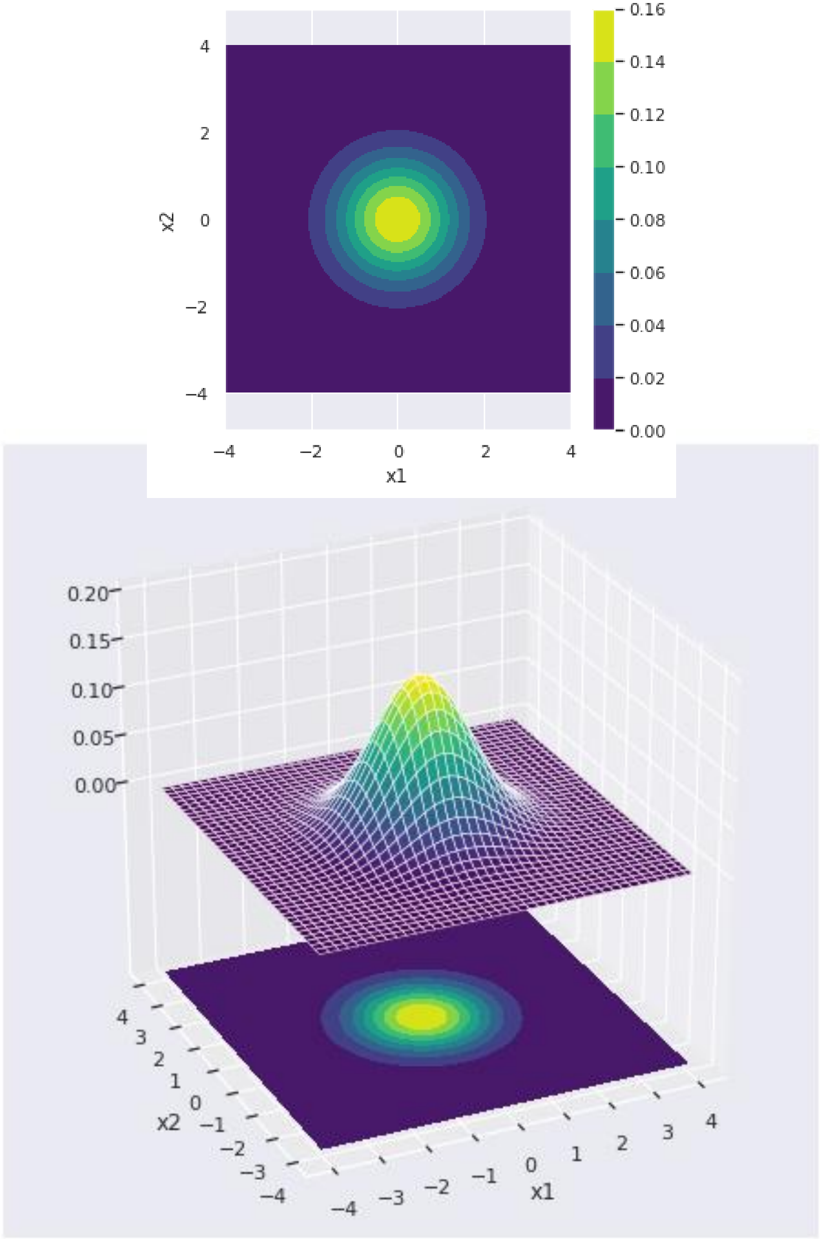
$$\boldsymbol{\mu} = \begin{pmatrix} 0. \\ 0. \end{pmatrix} \quad \boldsymbol{\Sigma} = \begin{pmatrix} 1. & 0.8 \\ 0.8 & 1. \end{pmatrix}$$



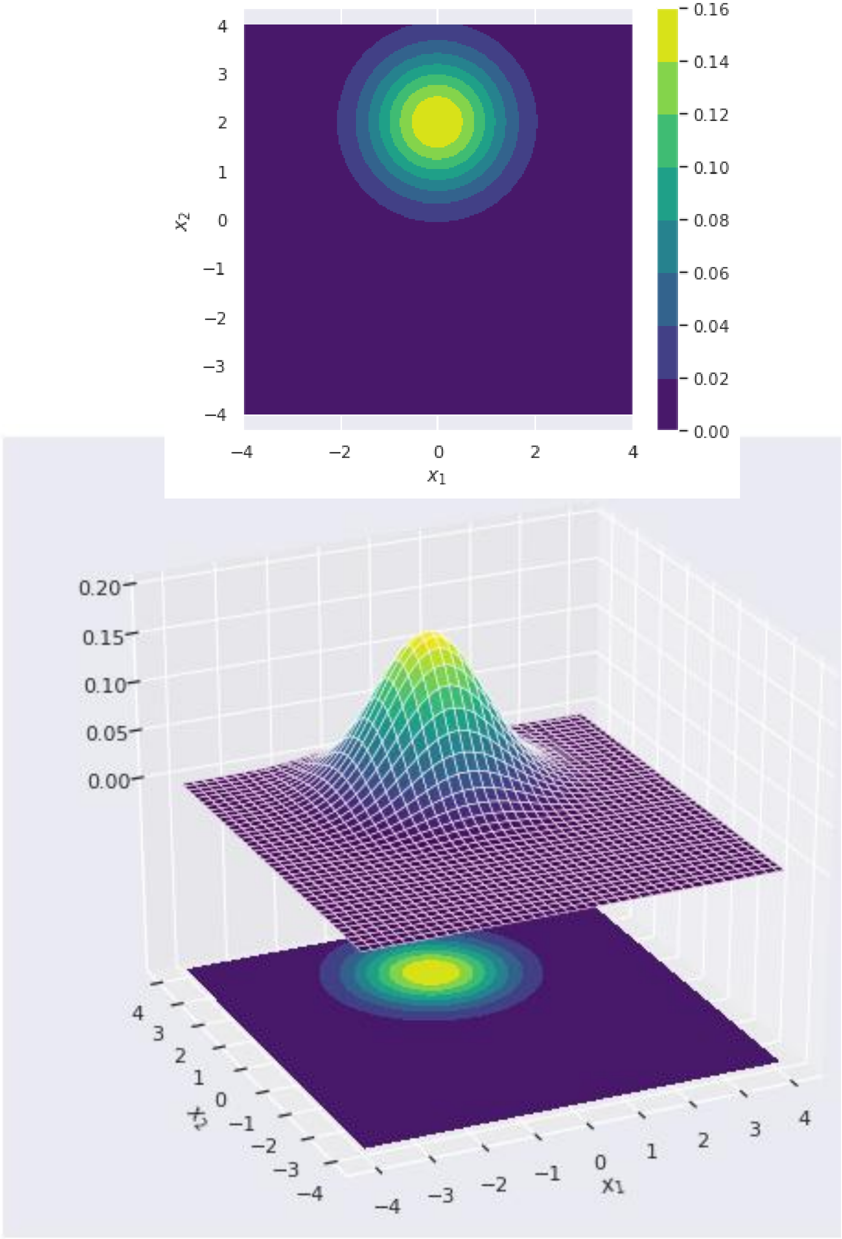
$$\boldsymbol{\mu} = \begin{pmatrix} 0. \\ 0. \end{pmatrix} \quad \boldsymbol{\Sigma} = \begin{pmatrix} 1. & 0. \\ 0. & 1. \end{pmatrix}$$

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

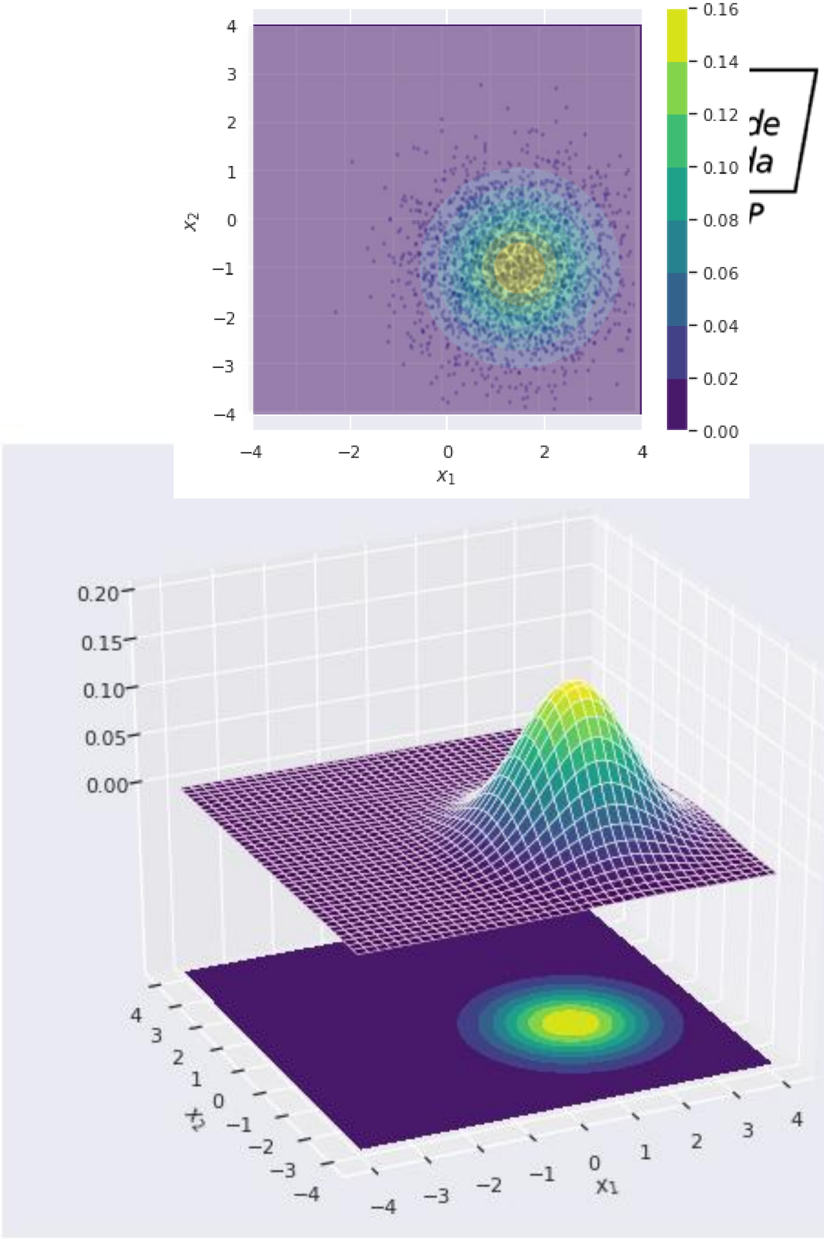
$$\boldsymbol{\mu} = \begin{pmatrix} 0. \\ 0. \end{pmatrix} \quad \boldsymbol{\Sigma} = \begin{pmatrix} 1. & -0.8 \\ -0.8 & 1. \end{pmatrix}$$



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



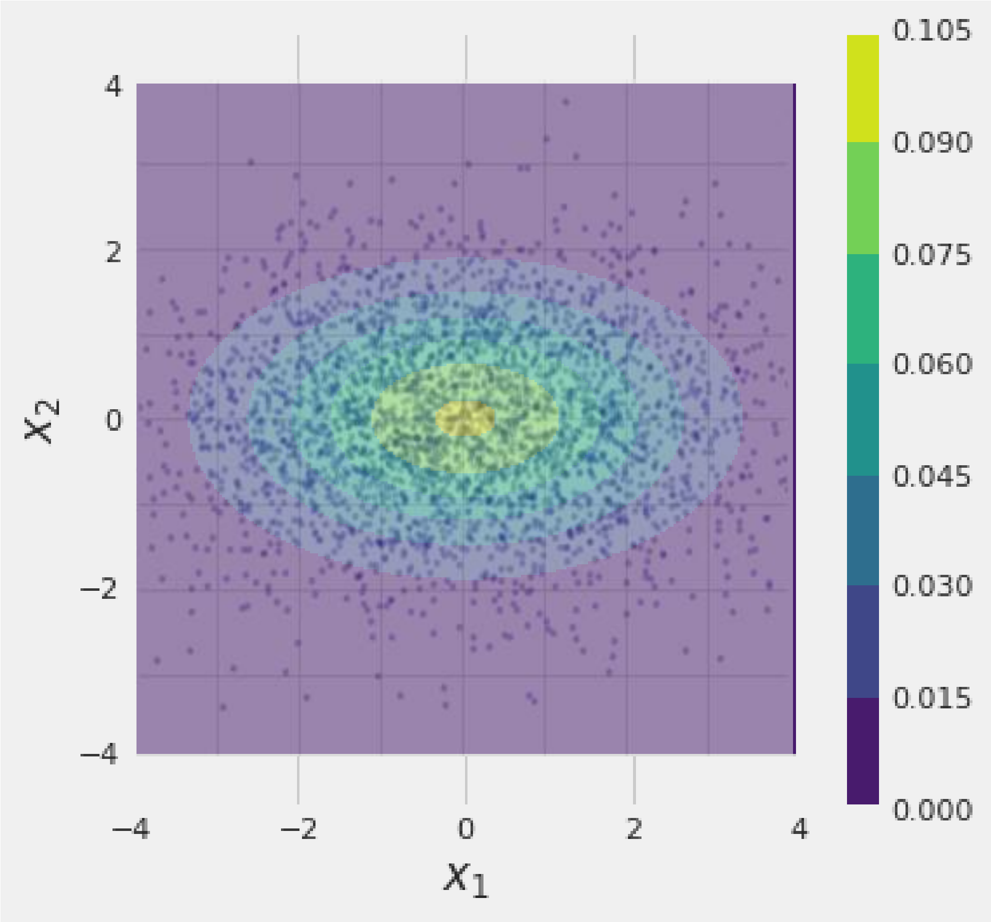
$$\boldsymbol{\mu} = \begin{pmatrix} 0. \\ 2.0 \end{pmatrix} \quad \boldsymbol{\Sigma} = \begin{pmatrix} 1. & 0. \\ 0. & 1. \end{pmatrix}$$

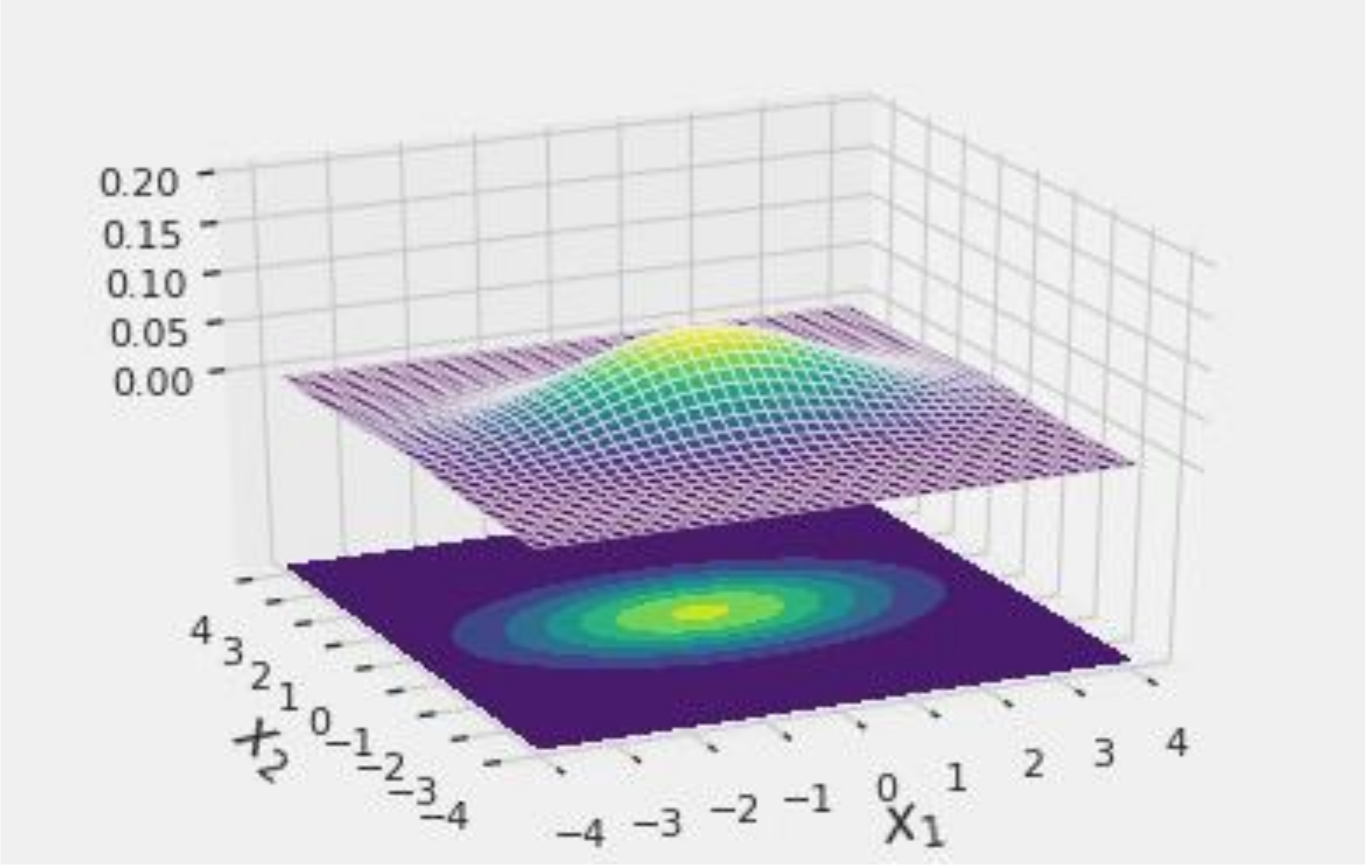


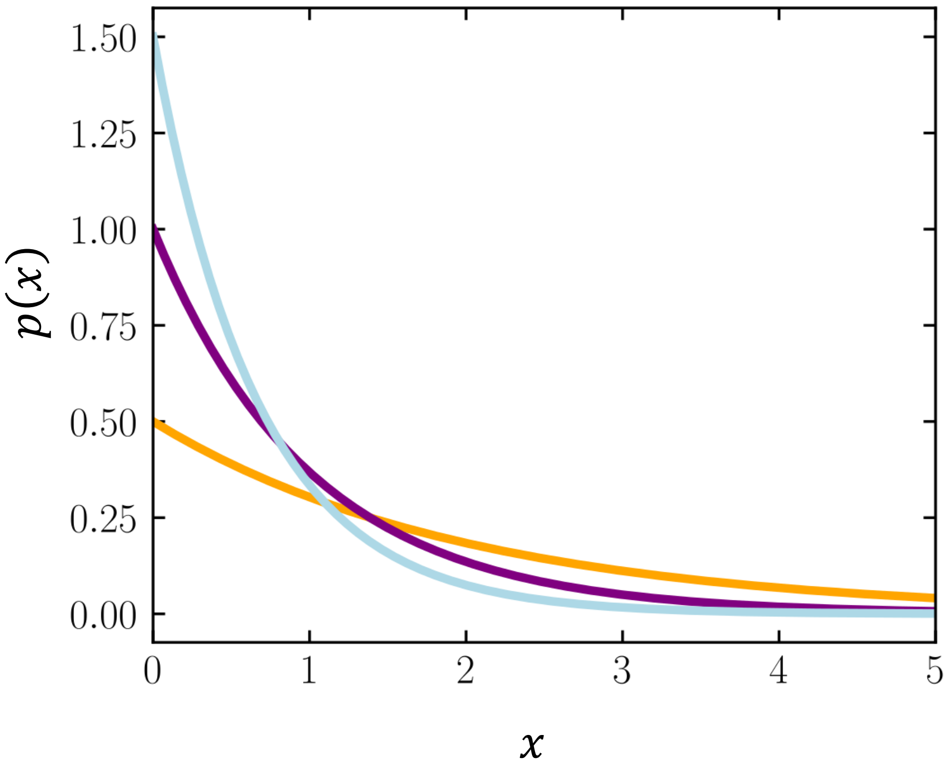
$$\boldsymbol{\mu} = \begin{pmatrix} 1.5 \\ -1.0 \end{pmatrix} \quad \boldsymbol{\Sigma} = \begin{pmatrix} 1. & 0. \\ 0. & 1. \end{pmatrix}$$

EXEMPLO

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







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

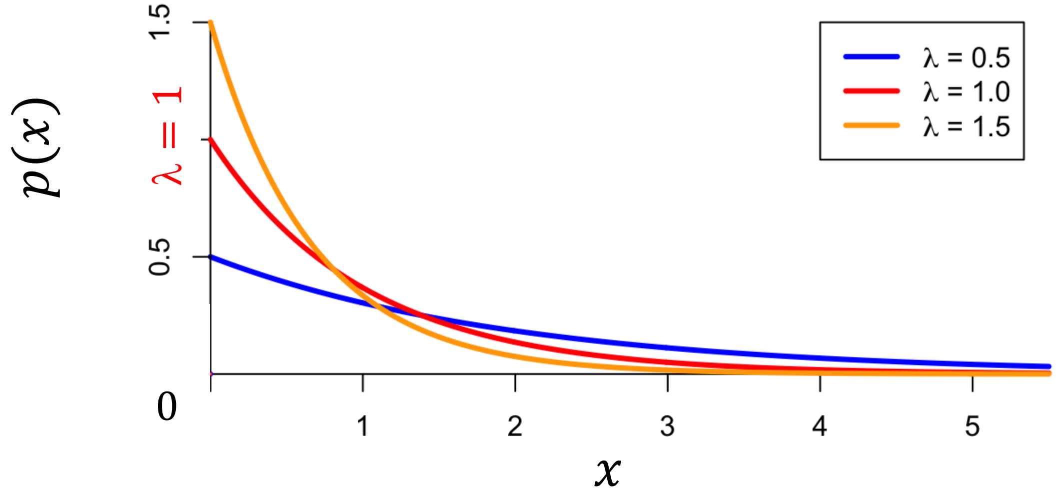
A média e variância são dados por,

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

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

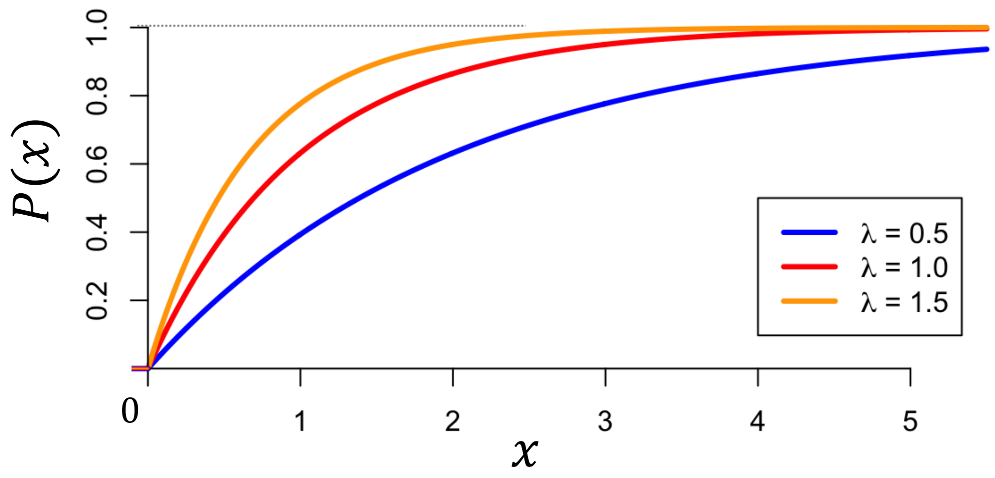
Função densidade de probabilidade da distribuição exponencial:

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



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

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

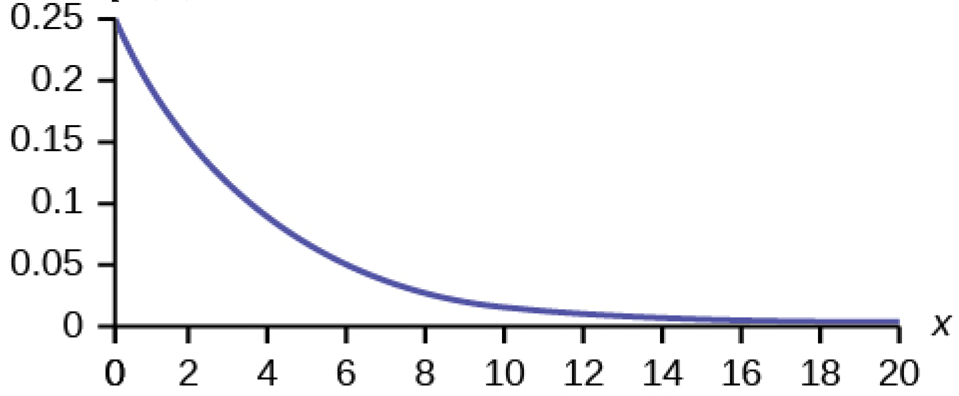


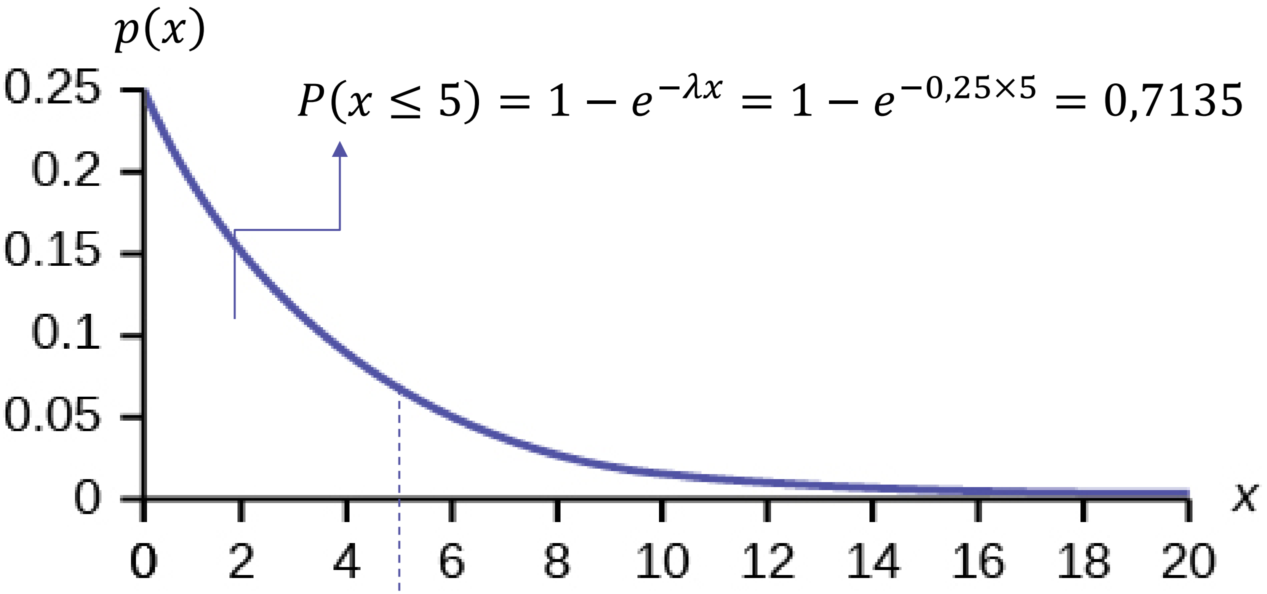
EXEMPLO

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?

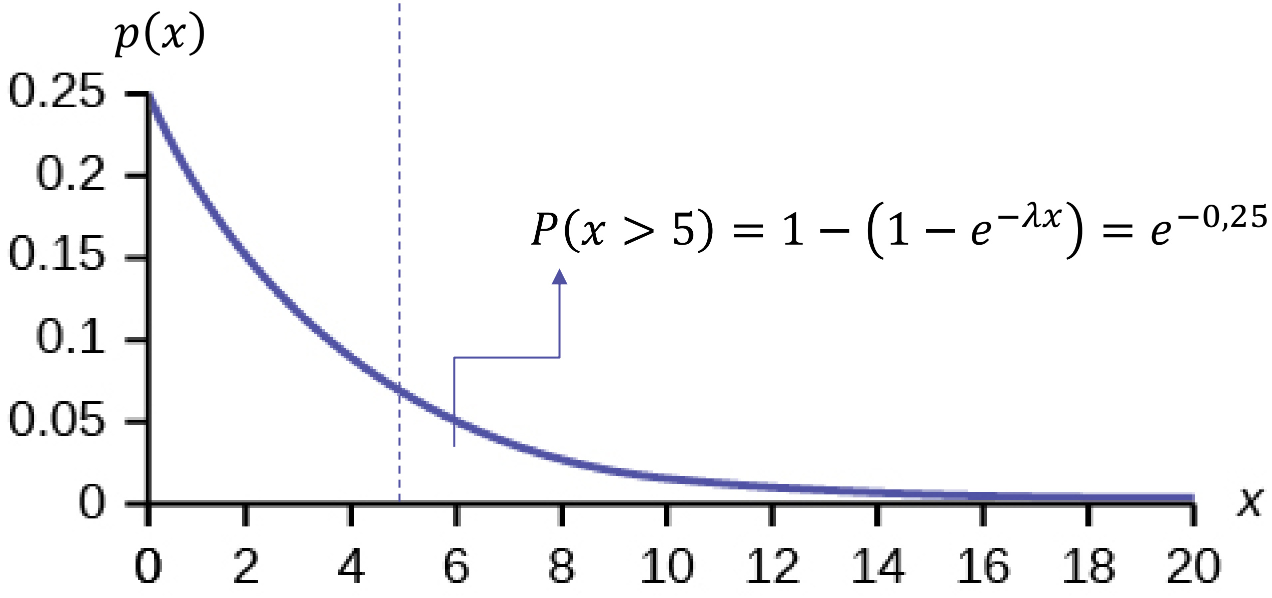
$$\mu = \beta = 4min \rightarrow \lambda = \frac{1}{4} = 0{,}25$$

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



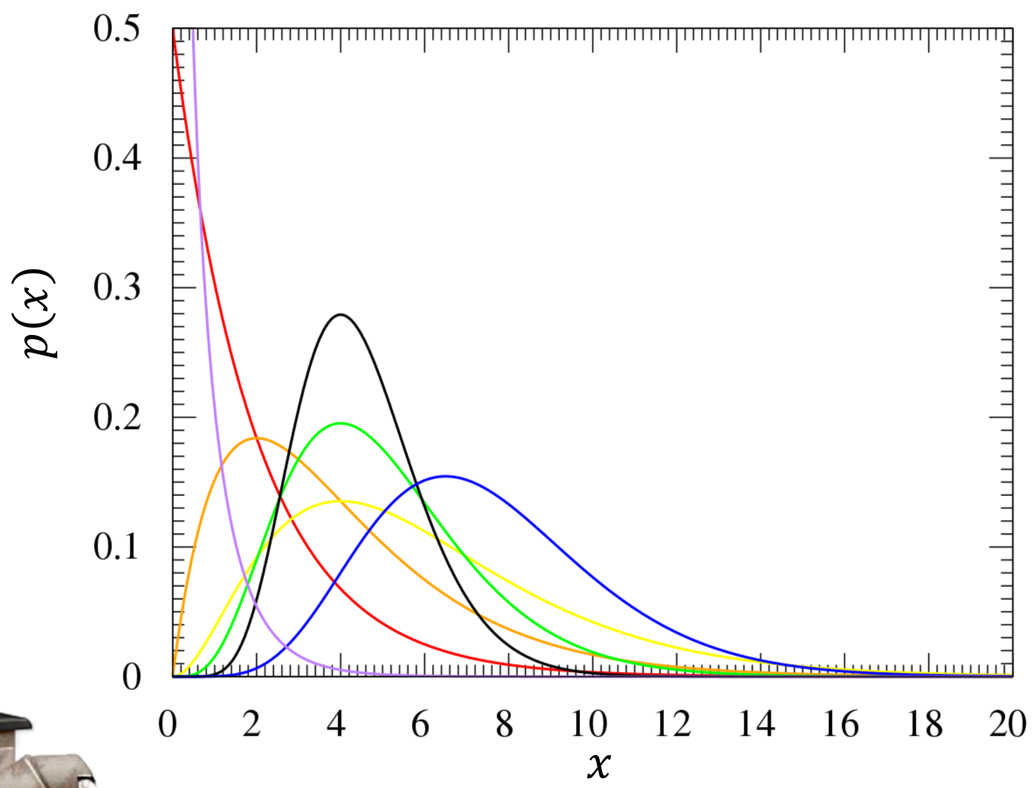


$$P(x \leq 5) = 1 - e^{-\lambda x} = 1 - e^{-0,25\times 5} = 0,7135$$



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

PECE Programa de Educação Continuada
Escola Politécnica da USP



DISTRIBUIÇÃO GAMA

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

DISTRIBUIÇÃO GAMA

A distribuição gama é uma distribuição de "tempo de espera". Suponha que os eventos ocorram de forma independente e aleatória, com um tempo médio entre eventos de β. O tempo de espera até que os eventos α ocorram é uma variável aleatória gama(α, β).

Isto é, se a distribuição Exponencial refere-se ao espaço/tempo decorrido até a ocorrência do primeiro evento, Gama é espaço/tempo até o $\alpha-$ ésimo evento.

Exemplo	α	β
Uma substância radioativa emite, em média, duas partículas alfa a cada segundo. Seja X o tempo de espera para a emissão de três partículas.		
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.		
Lâmpadas de portas de garagem duram, em média, cinco anos e são substituídas quando queimam. Seja X o tempo de duração de uma caixa com seis lâmpadas.		

DISTRIBUIÇÃO GAMA

A distribuição gama é uma distribuição de "tempo de espera". Suponha que os eventos ocorram de forma independente e aleatória, com um tempo médio entre eventos de β. O tempo de espera até que os eventos α ocorram é uma variável aleatória gama(α, β).

Isto é, se a distribuição Exponencial refere-se ao espaço/tempo decorrido até a ocorrência do primeiro evento, Gama é espaço/tempo até o $\alpha-$ ésimo evento.

Exemplo	α	β
Uma substância radioativa emite, em média, duas partículas alfa a cada segundo. Seja X o tempo de espera para a emissão de três partículas.	3	0,5
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.	5	2
Lâmpadas de portas de garagem duram, em média, cinco anos e são substituídas quando queimam. Seja X o tempo de duração de uma caixa com seis lâmpadas.	6	5

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

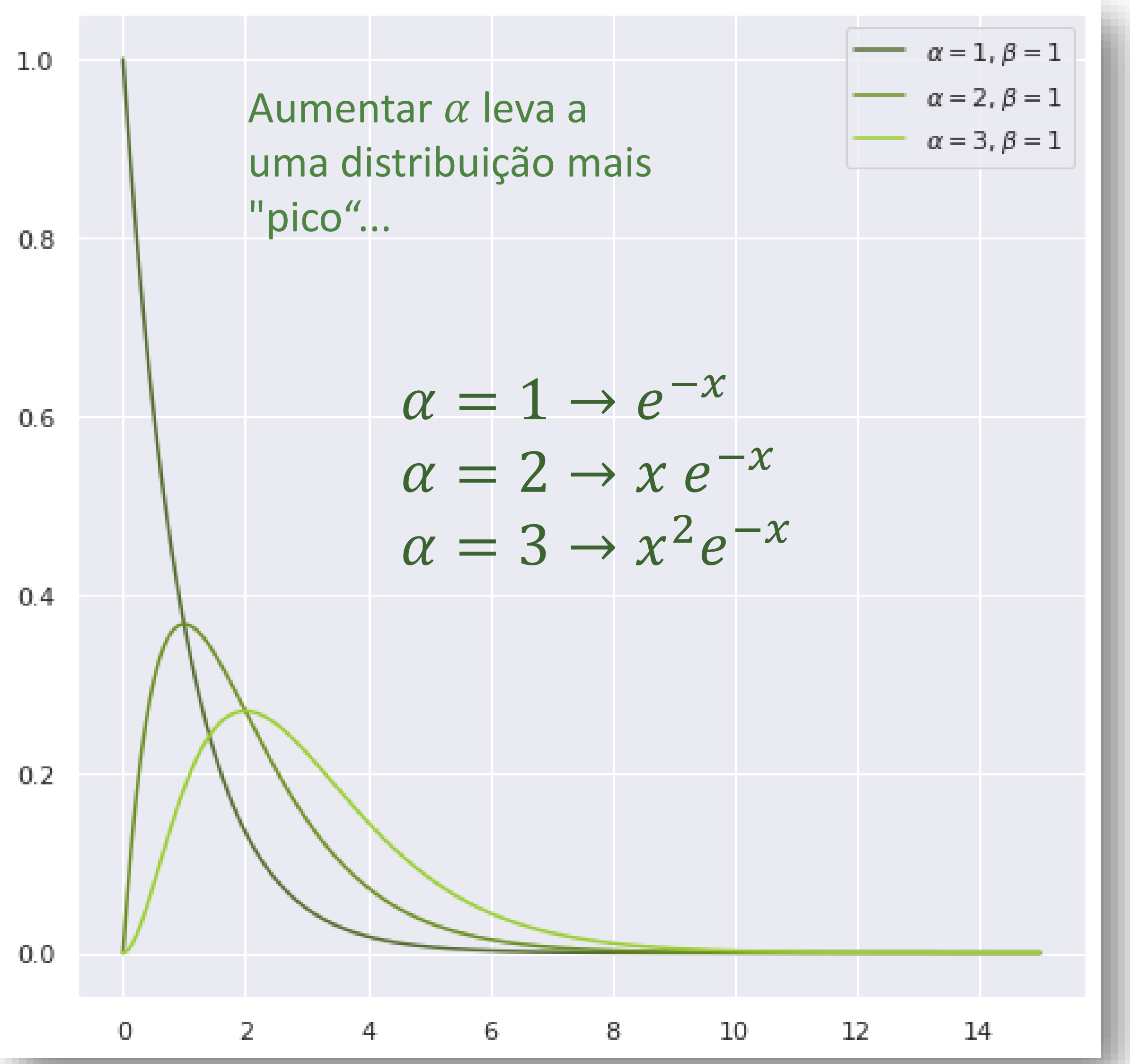
Quando $\alpha = 1$, Gama se reduz à distribuição exponencial. A média e variância são dadas, respectivamente, por, $\mu = \alpha\beta = \frac{\alpha}{\lambda}$ e $\sigma^2 = \alpha\beta^2 = \frac{\alpha}{\lambda^2}$.

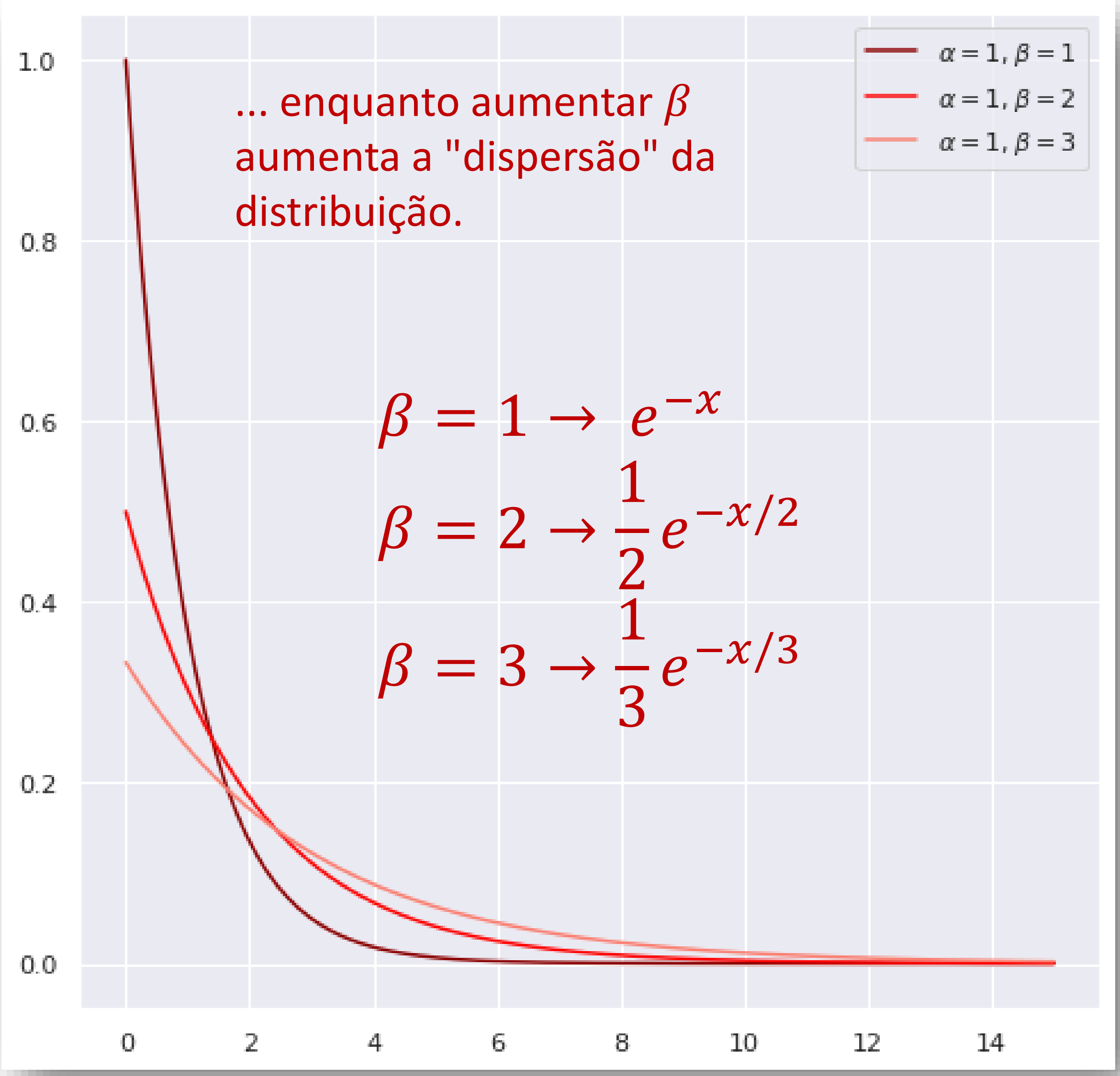
O parâmetro α é conhecido como parâmetro de forma, e o parâmetro β é chamado de parâmetro de escala.

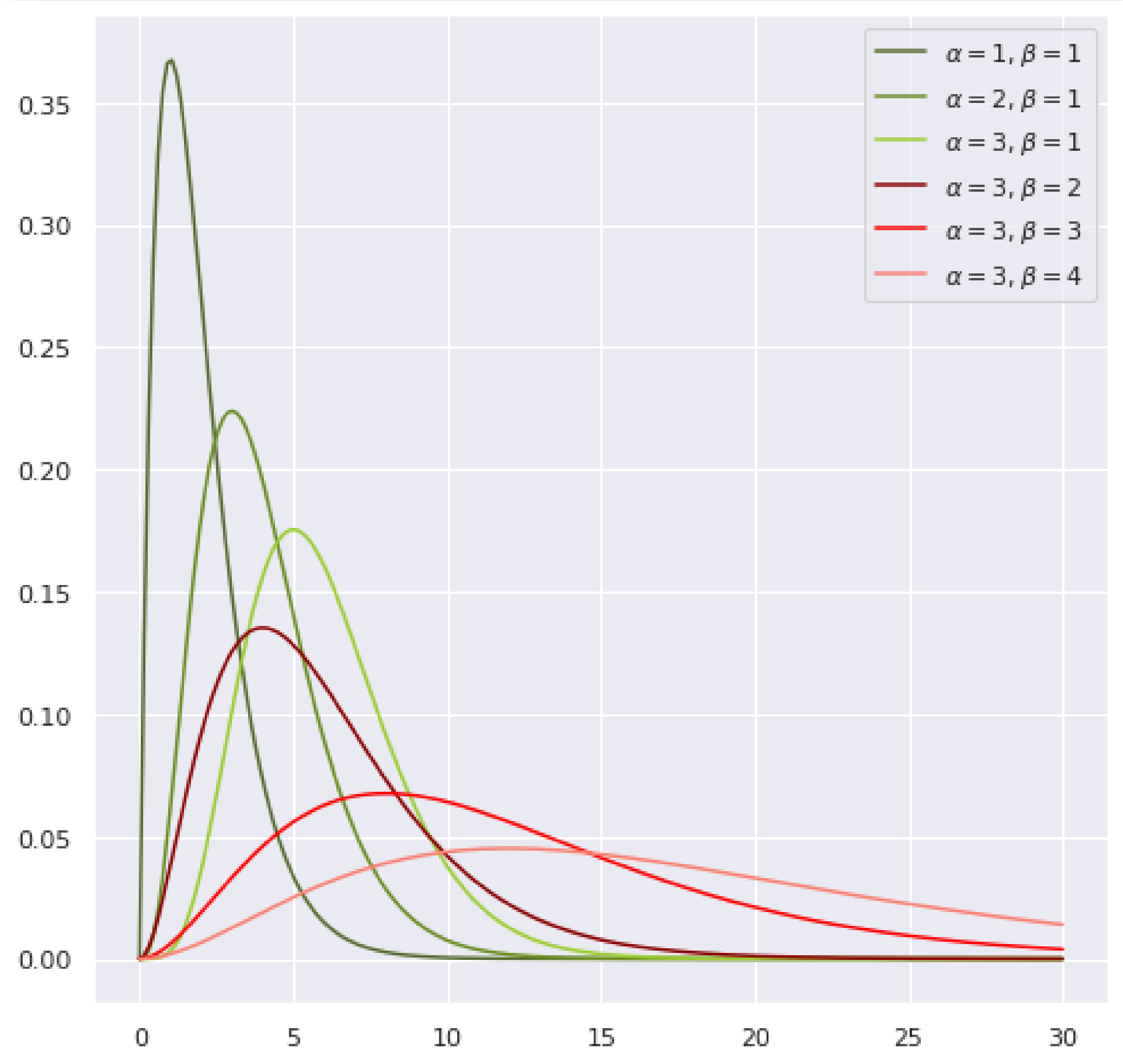
$$p(x;\alpha,\beta) = \frac{\beta^{-\alpha} x^{\alpha-1} e^{-x/\beta}}{\Gamma(\alpha)}$$

PECE Programa de Educação Continuada

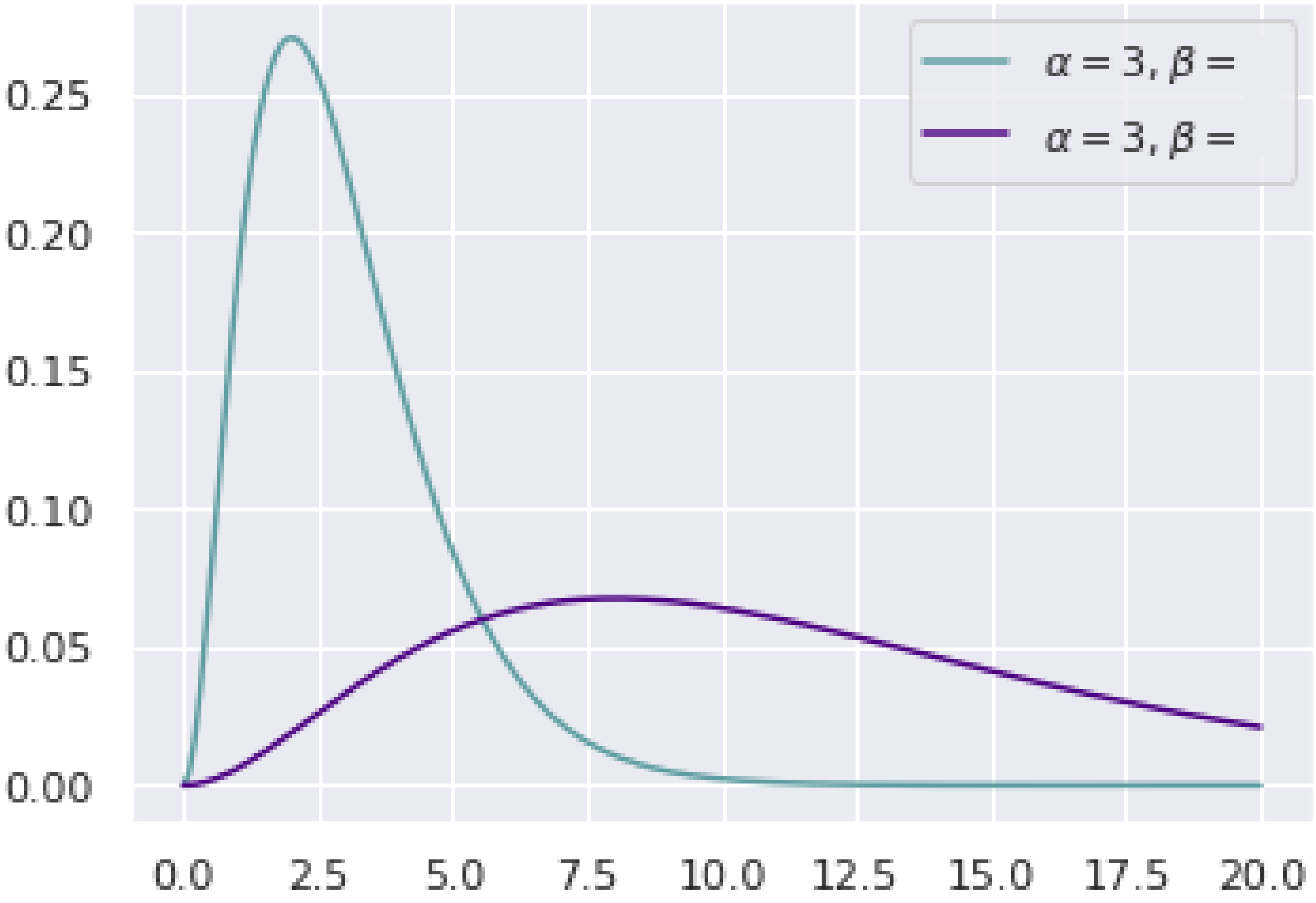
Escola Politécnica da USP







PECE Programa de Educação Continuada
Escola Politécnica da USP



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

ou

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

IMPORTANTE

A função de densidade acumulada (CDF) da Gama é:

$$P(X \leq x) = \frac{1}{\Gamma(\alpha)} \int_0^{x/\beta} t^{\alpha-1} e^{-t} dt$$

$$P(X > x) = 1 - \frac{\gamma(\alpha, x/\beta)}{\Gamma(\alpha)}$$

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.

$\alpha = 5$ e $\beta = 2$

Pede-se,

1. Qual é a probabilidade de você esperar mais de 7 minutos para o evento?

$$P(X > 7) = 1 - \frac{\gamma(5, 7/2)}{\Gamma(5)}$$

2. Plote a distribuição Gama do problema.

$\alpha = \beta = 2$

$t = 7 \text{ min}$

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

A probabilidade de esperar mais de 7 minutos até que 5 carros cheguem é aproximadamente 71,27%.

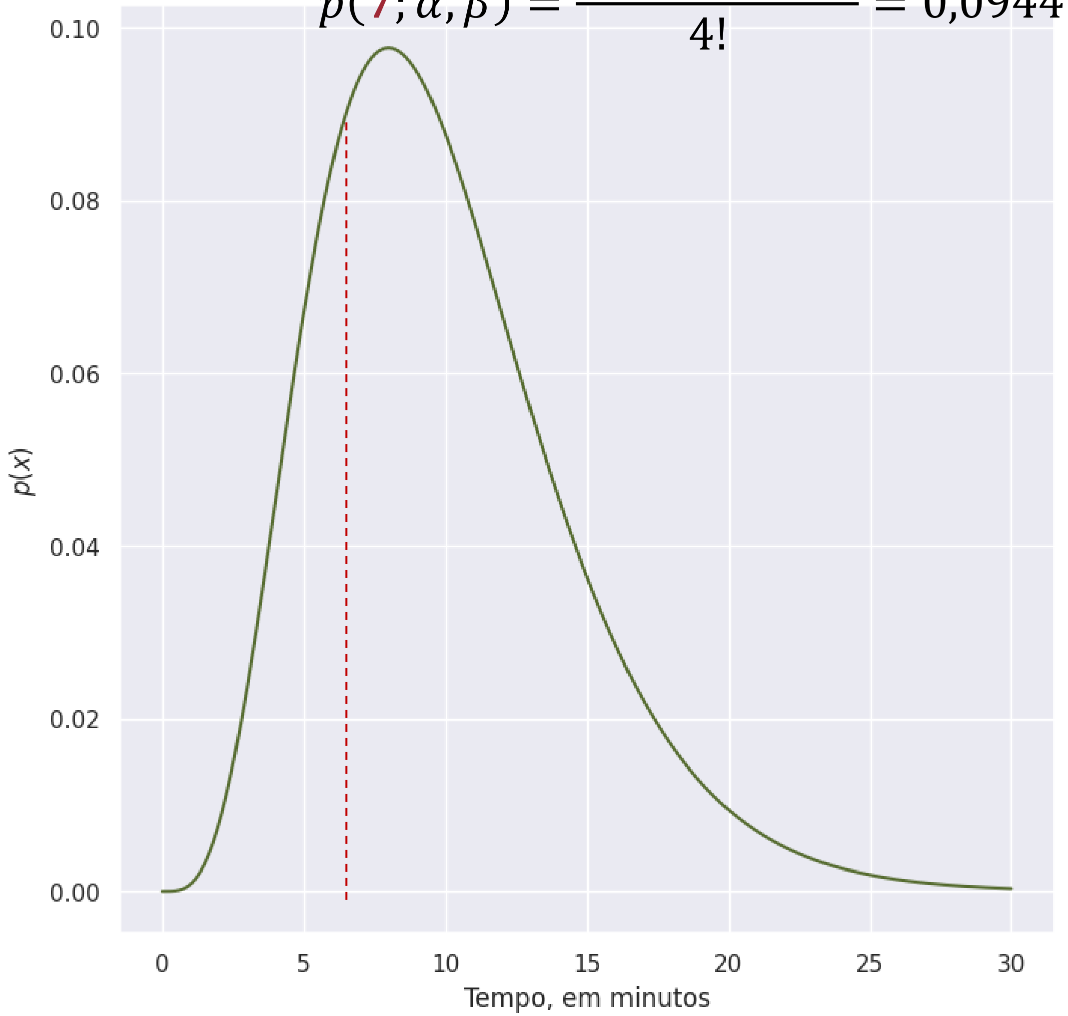
$$\gamma(5, 3.5) = \int_0^{3.5} t^4 e^{-t} dt$$

Esse valor **não dá pra integrar na mão facilmente**, mas tabelas ou calculadoras científicas dão:

$$\frac{\gamma(5, 3.5)}{\Gamma(5)} \approx 0.2873$$

$$P(X > 7) \approx 1 - 0.2873 = 0.7127$$

$$p(7; \alpha, \beta) = \frac{2^{-5} 7^{5-1} e^{-7/2}}{4!} = 0{,}0944$$



DISTRIBUIÇÃO DE GAMA VS EXPONENCIAL VS POISSON

PECE Programa de Educação Continuada
Escola Politécnica da USP

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?

PECE Programa de Educação Continuada
Escola Politécnica da USP

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.

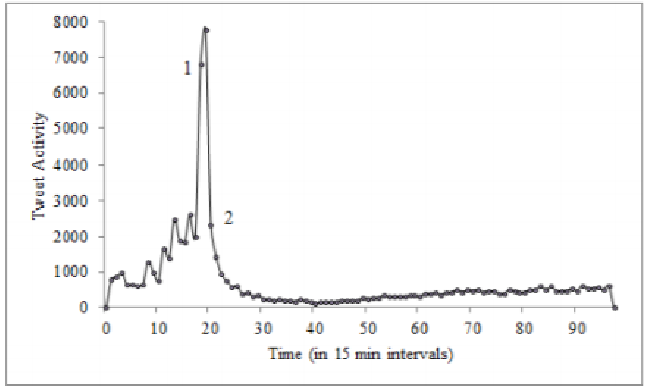
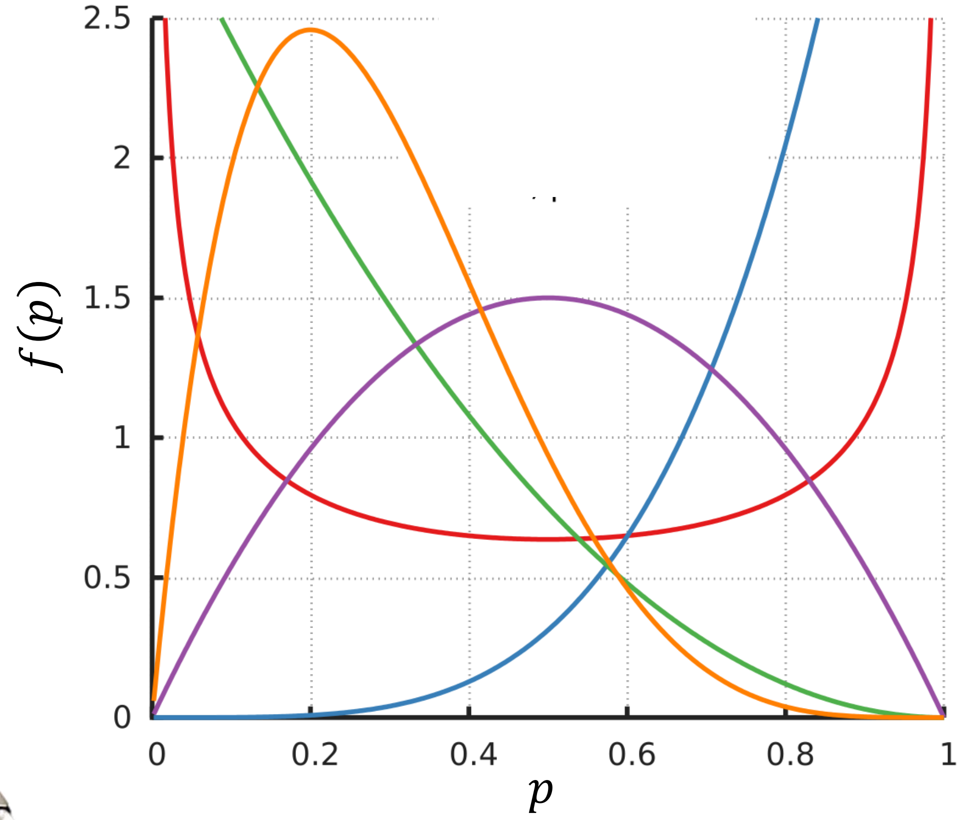


Figure 1: Frequency distribution of the "Stanley Cup" topic.



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

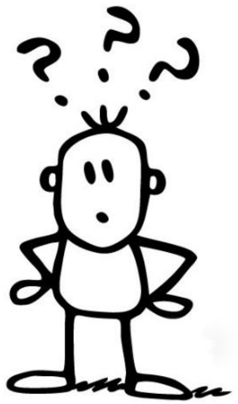
$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

$\Gamma(n) = (n-1)!$ para números naturais. A média e variância são dadas, respectivamente, por, $\mu = \frac{\alpha}{\alpha+\beta}$ e $\sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$.

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)**
$P(x;n,p) = \binom{n}{x} p^x (1-p)^{n-x}$	$f(p;\alpha,\beta) = \frac{1}{B(\alpha,\beta)} p^{\alpha-1}(1-p)^{\beta-1}$
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?"

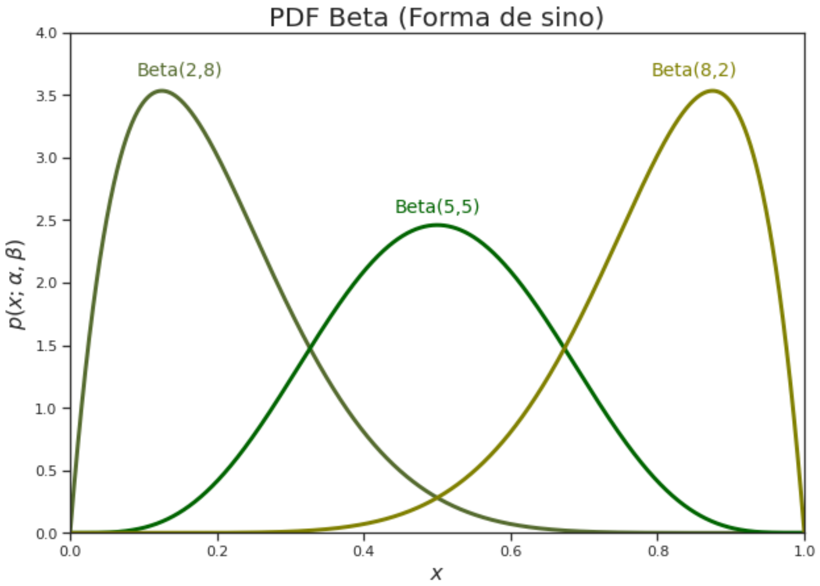
"... probabilidade de obter menos de 20% de caras em um lançamento de moeda?"

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

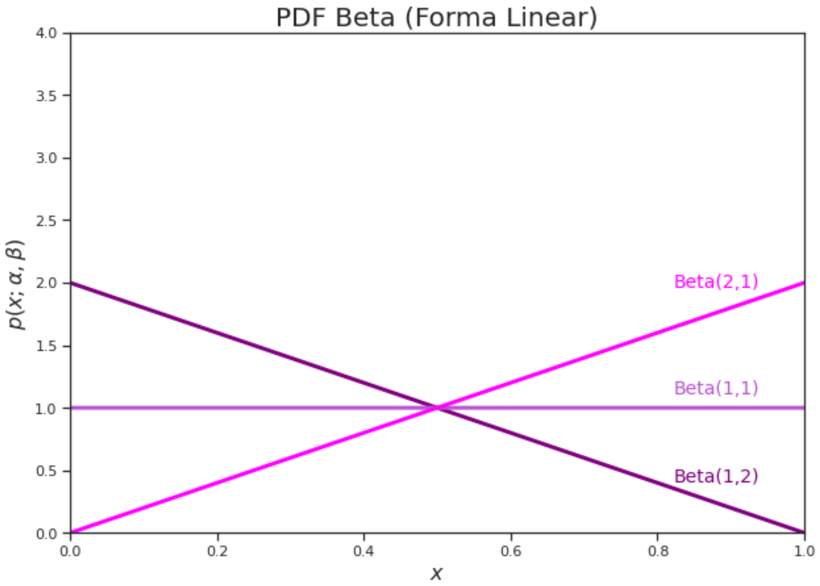
...

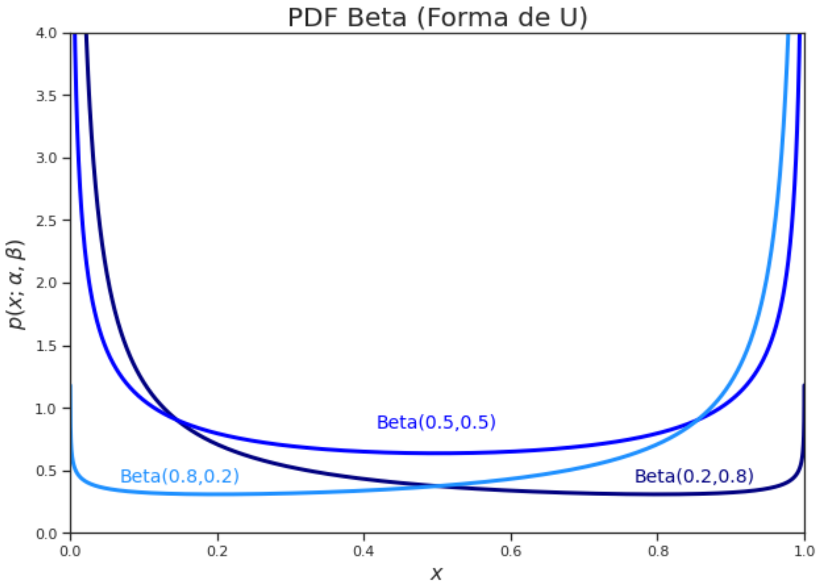
INTERPRETAÇÃO PARA α E β

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.

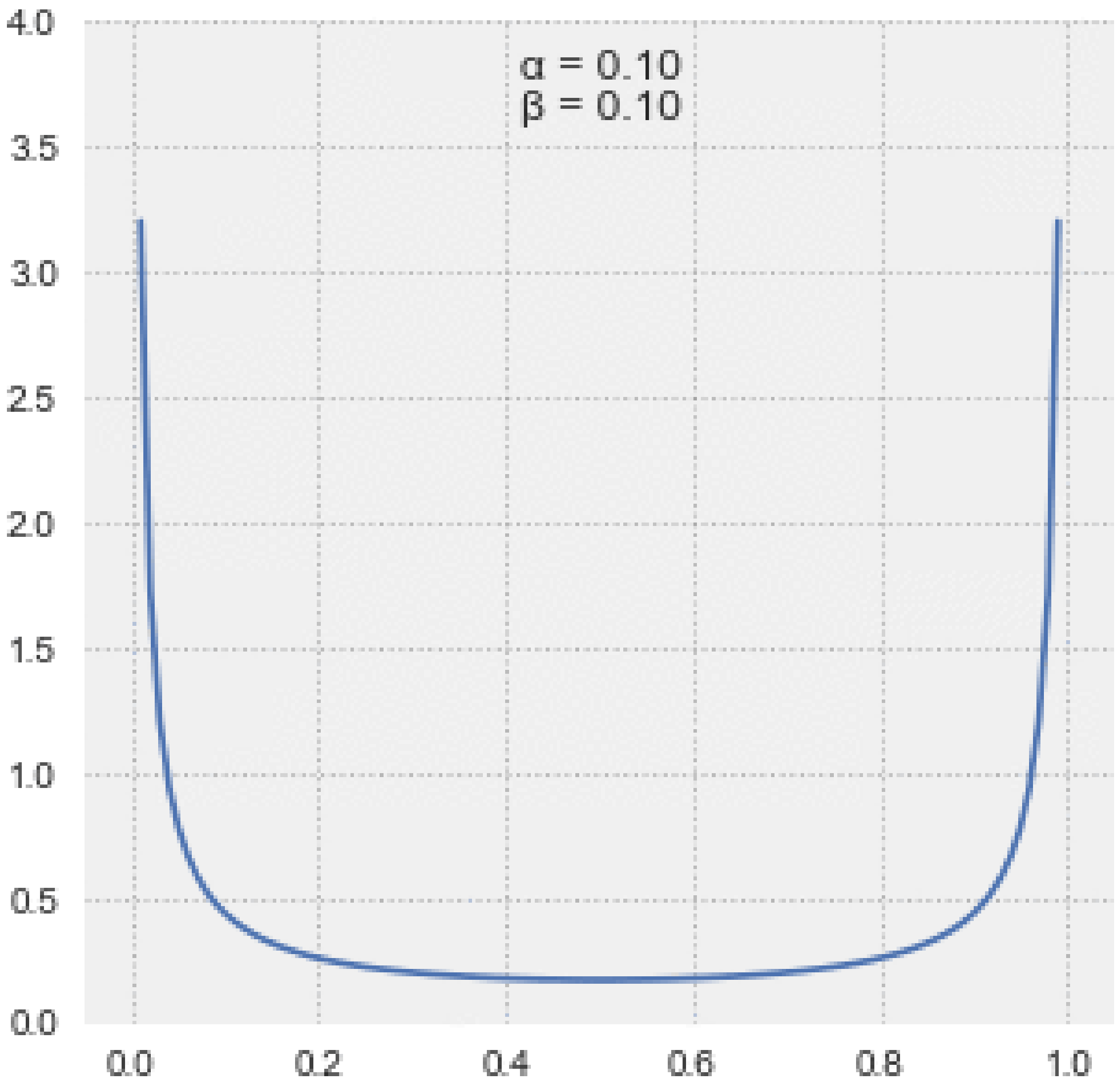


O PDF de uma distribuição beta é aproximadamente normal se $\alpha + \beta$ for grande o suficiente e α e β forem aproximadamente iguais.





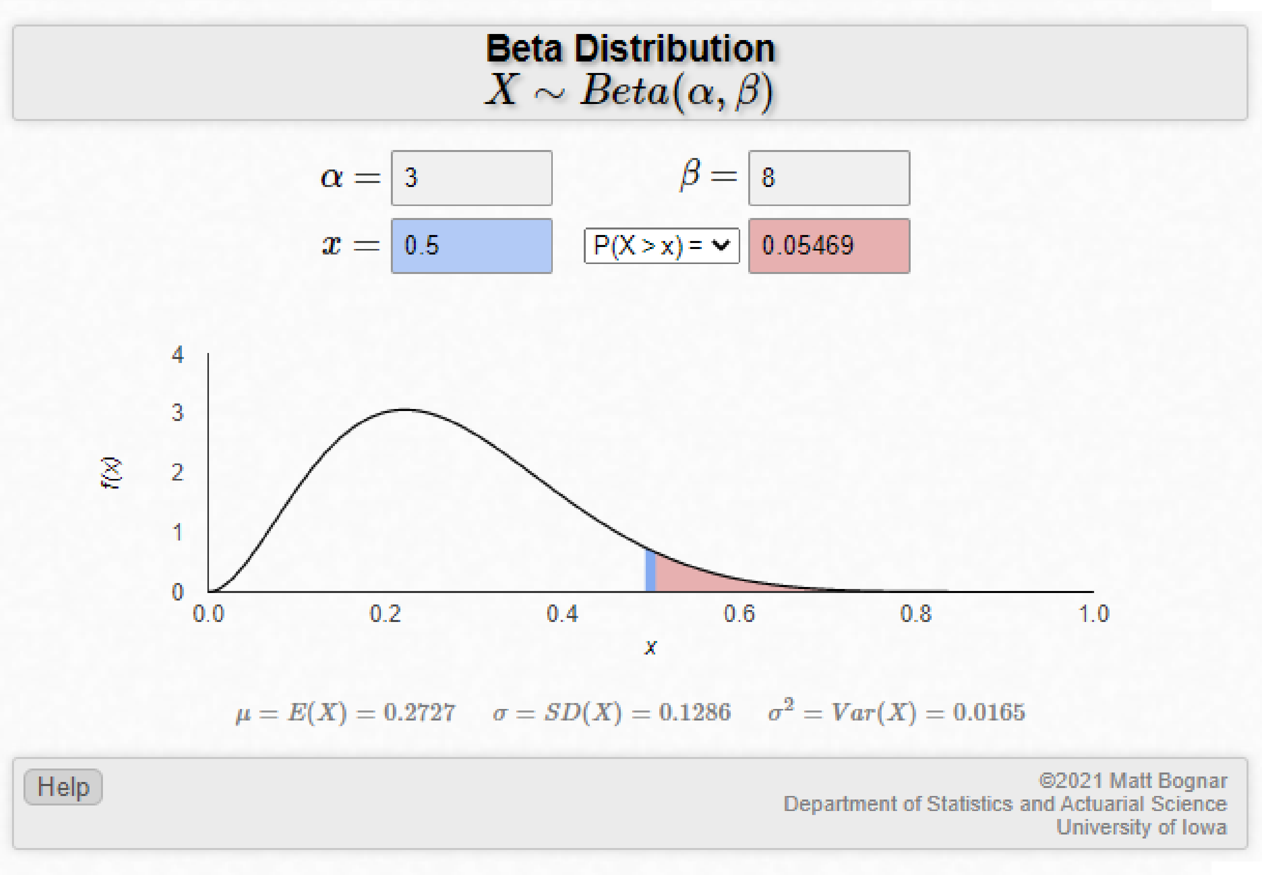
$\alpha < 1, \beta < 1$


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%?

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



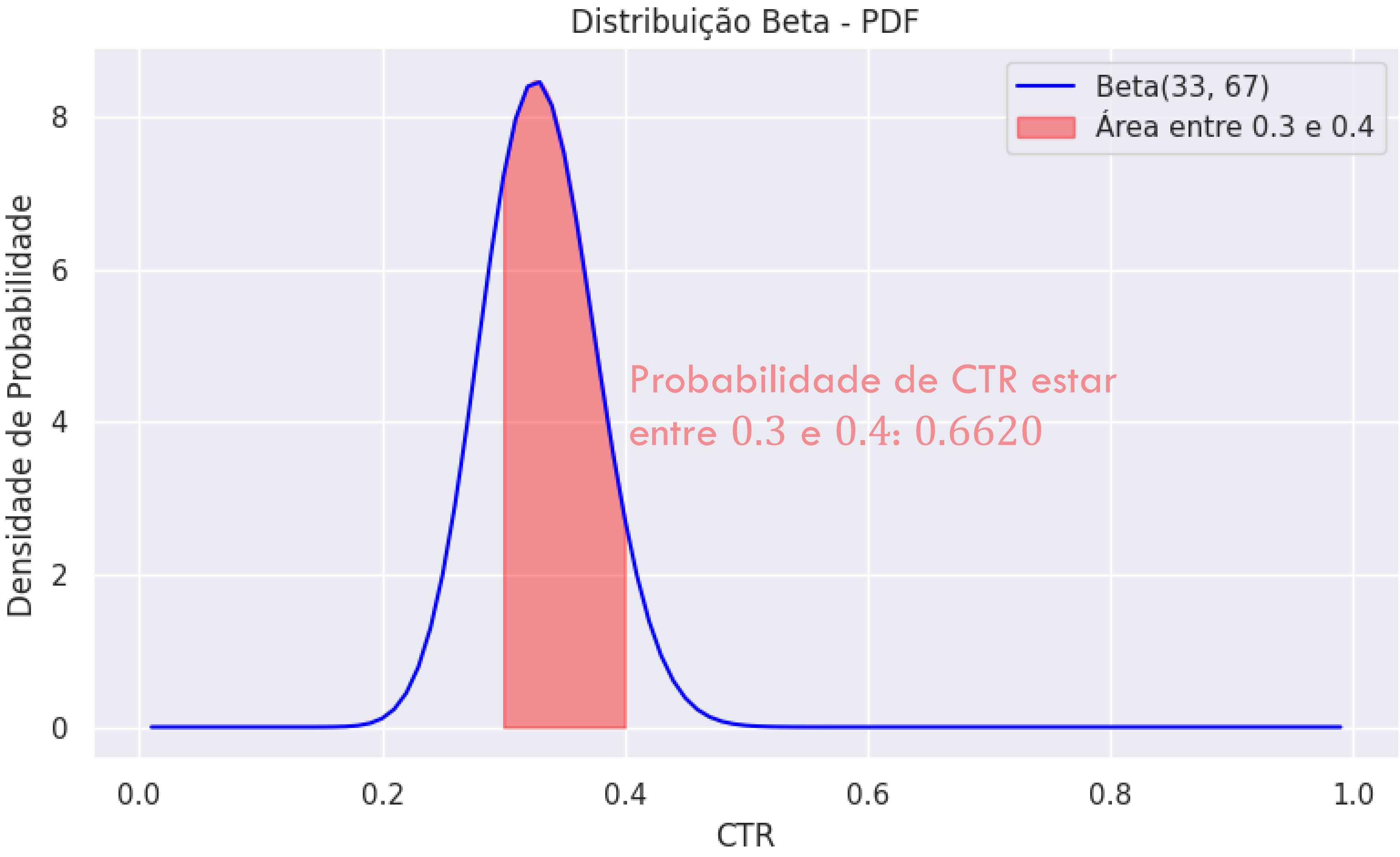
https://homepage.divms.uiowa.edu/~mbognar/applets/beta.html

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

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.

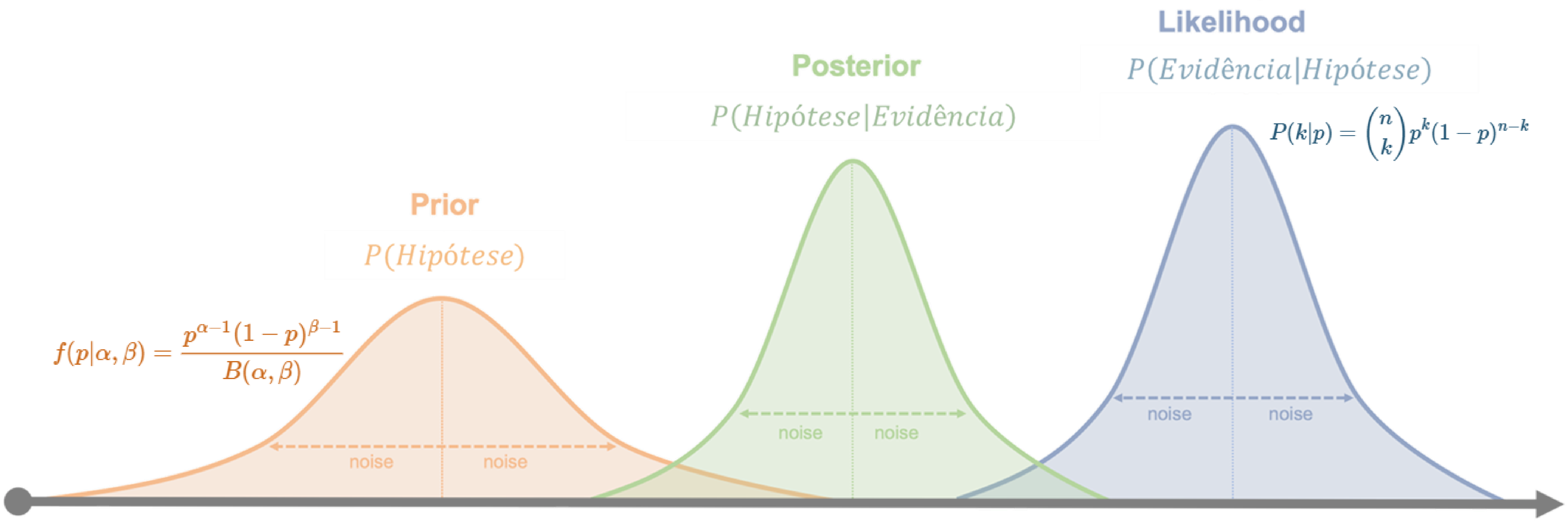
Sempre que ocorre um clique, aumentamos **alpha** em 1; se não ocorre, aumentamos **beta** em 1. Isso permite atualizar continuamente a probabilidade prévia à medida que novos dados são coletados.



Probabilidade de CTR estar entre 0.3 e 0.4: 0.6620

PECE Programa de Educação Continuada
Escola Politécnica da USP

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



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

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

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

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

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

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

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

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:

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

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

Completar o notebook IAD_001_2025_L09.ipynb

As questões sem resposta estão em verde.

NEVER GIVE UP

ACABOU...

Reveja a aula antes de resolver os exercícios.