Métodos de Estatística Aplicada com Python

Aula 6

Carlos Góes[1]

[1]Pós-Graduação em Ciência de Dados
Instituto de Educação Superior de Brasília

2017

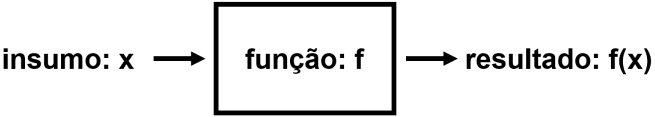
Sumário

Sumário

Funções

Definição e notação

- Funções matemáticas são como uma máquina que toma determinados insumos e entrega um resultado



Funções

Definição e notação

- Denota-se uma função que depende da variável x como uma *função de* x.
- Por exemplo:

$$
\begin{aligned}
se \qquad f(x) &\equiv x^2, \\
então: & \\
f(-2) &= (-2)^2 = 4 \\
f(10) &= (10)^2 = 10 \\
f(-100) &= (-100)^2 = 10000 \\
etc.. &
\end{aligned} \tag{1}
$$

Funções

Definição e notação

- É razoavelmente intuitivo ir do conceito de uma função matemática para uma função computacional, definindo:

```
def f(x):
    return x ** 2
```

ou

```
f = lambda x: x ** 2
```

podemos ver os resultados

```
print(f(4), f(-10), f(100))
```

Funções

Definição e notação

- É sempre possível combinar funções.
- Por exemplo:

$$
\begin{aligned}
se \qquad f(x) &\equiv x^2, \\
e: \qquad g(x) &\equiv \sqrt{x} \\
então, \quad quanto \quad seria: & \\
f(g(x))? &
\end{aligned} \tag{2}
$$

Funções

Definição e notação

- É sempre possível combinar funções.
- Por exemplo:

$$
\begin{aligned}
se \qquad f(x) &\equiv x^2, \\
e: \qquad g(x) &\equiv \sqrt{x} \\
então: & \\
f(g(x)) &= f(\sqrt{x}) = (\sqrt{x})^2 = x
\end{aligned} \tag{3}
$$

Funções

Definição e notação

- Testando no python:

```
f = lambda x: x ** 2
g = lambda x: np.sqrt(x)
```

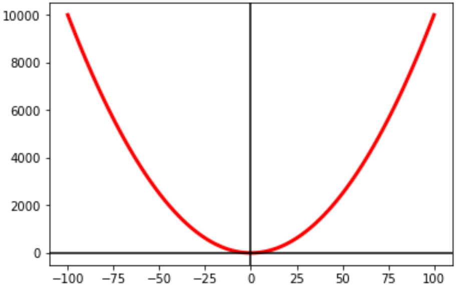
- Podemos ver os resultados

```
print(f(4), g(4), f(g(4)))
```

Funções

Representações algébrica e gráfica

- Funções univariadas têm sempre representações algébricas e gráficas.
- Essas duas representações são equivalentes.
- Por exemplo, a função $f(x) = x^2$ tem a seguinte representação gráfica:



- O que ela significa?

Funções

Representações algébrica e gráfica

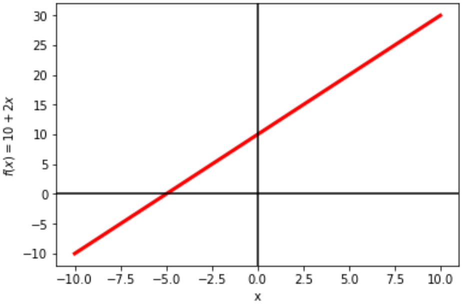
- Que pode ser reproduzida no Python

```
f = lambda x: x ** 2

x = np.linspace(-100,100,101)
y = f(x)

plt.xlabel('x')
plt.ylabel(r'$f(x) = x^2$')
plt.plot(x,y, color='red', linewidth=3)
plt.axhline(0, color='black')
plt.axvline(0, color='black')
plt.show()
```

Funções

Representações algébrica e gráfica

- O mesmo vale para outras funções.
- Por exemplo, a função $f(x) = 10 + 2x$ tem a seguinte representação gráfica:



- Em qual número a linha cruza o eixo vertical?
- E o eixo horizontal?

Funções

Representações algébrica e gráfica

- Que pode ser reproduzida no Python

```
f = lambda x: 10 + 2x

x = np.linspace(-10,10,1000)
y = f(x)

plt.xlabel('x')
plt.ylabel(r'$f(x) = 10 + 2x$')
plt.plot(x,y, color='red', linewidth=1)
plt.axhline(0, color='black')
plt.axvline(0, color='black')
plt.show()
```

Funções

Derivadas: intuição

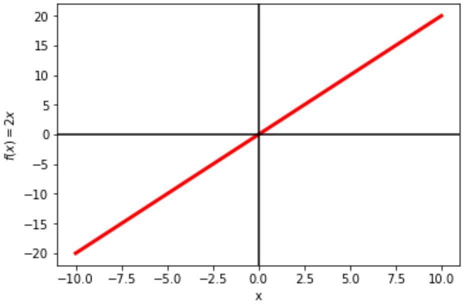
- Um conceito importante para compreender funções em estatística (em especial em relações associativas) é o de derivadas.
- Esse não é um curso de cálculo, mas é importante que saibamos a intuição.
- Em poucas palavras, a derivada de uma função é a sensibilidade do resultado de uma função à mudança no insumo daquela função.
- A notação matemática é a seguinte:

$$\frac{dy}{dx} = \frac{quanto \quad y \quad muda}{quando \quad x \quad muda} \tag{4}$$

Funções

Derivadas: intuição

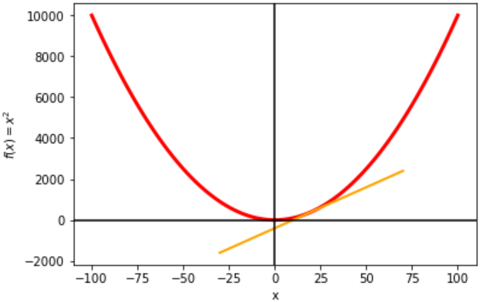
- Graficamente, podemos observar as mesmas relações.
- A derivada está relacionada à mudança relativa entre as variáveis.



Funções

Derivadas: intuição

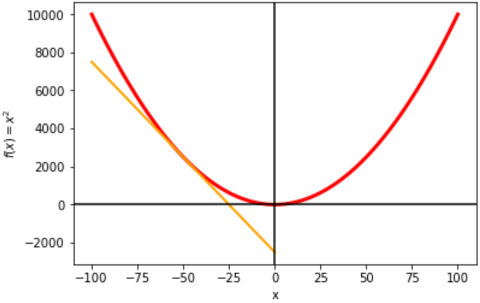
- Quando sua função não é uma variação constante, a derivada é a tangente da função em determinado ponto
- Quando a curva tem inclinação positiva (negativa), a derivada também é positiva (negativa).



Funções

Derivadas: intuição

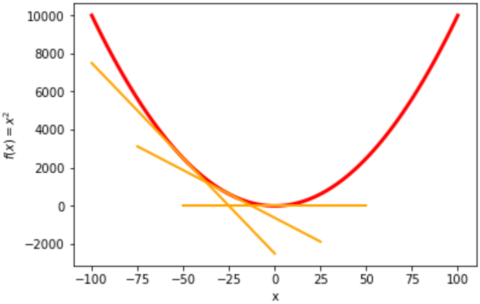
- Quando sua função não é uma variação constante, a derivada é a tangente da função em determinado ponto
- Quando a curva tem inclinação positiva (negativa), a derivada também é positiva (negativa).



Funções

Derivadas: intuição

- Dependendo da inclinação da curva de tangente, a derivada é maior ou menor.



Sumário

Probabilidade: noções e inferência

Definição e notação

- A probabilidade de um evento denota a frequência relativa de longo prazo de um evento
- Ou seja, nossa melhor resposta para qual é a chance de um evento ocorrer
- A probabilidade deve ser vista num espectro que vai de zero (0% de chance de ocorrer) a um (100% de chance de ocorrer).

Probabilidade: noções e inferência

Definição e notação

- Algumas vezes, sabemos a probabilidade de um evento por construção.
- Em um dado não viciado, a probabilidade do resultado ser qualquer um dos lados é igual: uma em seis, ou 16,6%.
- Em notação matemática:

$$
\begin{aligned}
Pr(1) &= Pr(2) = Pr(3) = Pr(4) = Pr(5) = Pr(6) && (5)\\
Pr(1) &= \frac{1}{6} = 0,1666 && (6)
\end{aligned}
$$

Probabilidade: noções e inferência

Definição e notação

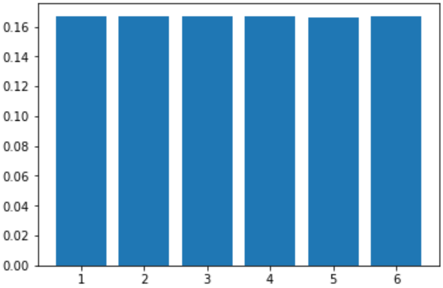
- Como sabemos isso?

```
x = [random.randint(1,6) for i in range(0,1000000)]

plt.hist(x,
    bins=[0.5+i for i in range(0,7)],
    normed=True,
    rwidth=0.8)
plt.show()
```

Probabilidade: noções e inferência

Definição e notação

- Como sabemos isso?



Probabilidade: noções e inferência

Probabilidade simples

- Para eventos mutuamente excludentes:
  - a probabilidade de A não ocorrer é igual:

$$Pr(\bar{A}) = 1 - Pr(A) \tag{7}$$

  - Ex: qual a probabilidade de o dado não tirar um?

$$Pr(\bar{1}) = 1 - Pr(A) = 1 - \frac{1}{6} = \frac{5}{6} \tag{8}$$

  - a probabilidade de A ou B ocorrerem, se A e B forem independentes, é igual a probabilidade de A ocorrer mais a probabilidade de B ocorrer

$$Pr(A \cup B) = Pr(A) + Pr(B) \tag{9}$$

  - Ex: qual a probabilidade de o dado tirar um ou seis?

$$Pr(1 \cup 6) = Pr(1) + Pr(6) = \frac{1}{6} + \frac{1}{6} = \frac{2}{6} \tag{10}$$

Probabilidade: noções e inferência

Probabilidade simples

- Para eventos independentes:
  - a probabilidade de A e B ocorrerem é a multiplicação de suas probabilidades marginais:

$$Pr(A \cap B) = Pr(A) * Pr(B) \tag{11}$$

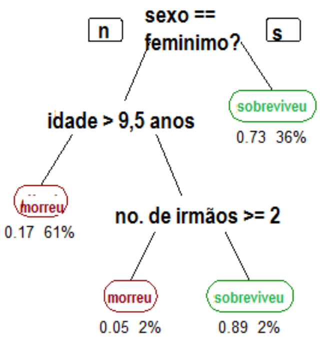
  - Ex: qual a probabilidade de um dado tirar seis duas vezes seguidas?

$$Pr(6 \cap 6) = Pr(6) * Pr(6) = \frac{1}{6} * \frac{1}{6} = \frac{1}{36} \tag{12}$$

Probabilidade: noções e inferência

Relevância

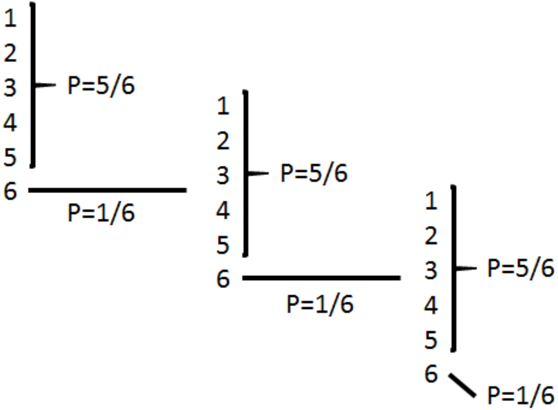
- Por que isso importa?
- Porque muitas vezes, ao escrever algoritmos de classificação, nós utilizamos essas noções de probabilidade marginal.



Probabilidade: noções e inferência

Relevância

- O que é diretamente relacionado à noção de probabilidade



Probabilidade: noções e inferência

Probabilidade condicional

- Por enquanto, falamos de eventos independentes
- Mas o que muda quando incluímos outras informações não independentes?
- Por exemplo, nós sabemos que há 1,3 bilhões de católicos no mundo, de tal modo que a probabilidade de um humano aleatório ser católico é:

$$P(Católico) = \frac{1,3 bilhões}{7 bilhões} = 18,6\%$$

Probabilidade: noções e inferência

Probabilidade condicional

- Mas a probabilidade de você ter determinada religião é independente de onde você nasceu?
- Você acha que a probabilidade de alguém ser católico é a mesma no Brasil e na Arábia Saudita?
- Qual é a probabilidade de alguém ser católico, sabendo que ele é brasileiro?

$$P(Católico|Brasileiro) = \frac{126 milhões}{210 milhões} = 60\%$$

Probabilidade: noções e inferência

Probabilidade condicional

- Isso é equivalente a:

$$P(Católico|Brasileiro) = \frac{P(Católico \cap Brasileiro)}{P(Brasileiro)} = \frac{\frac{126mi}{7bi}}{\frac{210mi}{7bi}} = \frac{126mi}{210mi}$$

- Generalizando:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{13}$$

Probabilidade: noções e inferência

Probabilidade condicional

- O que significa que podemos também fazer o caminho inverso

$$P(A \cap B) = P(A|B)P(B) \tag{14}$$

- Ou seja, se soubersmos a probabilidade de alguém ser Católico, dado que alguém e brasileiro; e soubermos a probabilidade de alguém ser brasileiro, podemos derivar a probabilidade de alguém ser católico e brasileiro!

$$\begin{aligned} P(Católico \cap Brasileiro) &= P(Católico|Brasileiro)P(Brasileiro) \\ &= \frac{126mi}{210mi} \cdot \frac{210mi}{7bi} = \frac{126mi}{7bi} \\ &= 1,8\% \end{aligned}$$

Probabilidade: noções e inferência

Teorema de Bayes

- O Teoremas de Bayes se constroi sobre essas noções de probabilidade condicional - e nos ajuda a chegar a inferências probabilísticas quando só temos informação parcial
- Dado que:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

- E:

$$\begin{aligned} P(A \cap B) &= P(A|B)P(B) \\ P(B \cap A) &= P(B|A)P(A) \\ P(A \cap B) &= P(B \cap A) \end{aligned}$$

Probabilidade: noções e inferência

Teorema de Bayes

- Podemos concluir que:

$$P(A|B)P(B) = P(B|A)P(A)$$

- E, portanto:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Probabilidade: noções e inferência

Teorema de Bayes

- O que isso significa?
- Exemplo prático:


INTERNACIONAL

Qual a probabilidade de um muçulmano ser terrorista?

Por Carlos Góes @goescarlos · Em 14/11/2015

Like 4.6K | Tweetar | G+

Probabilidade: noções e inferência

Teorema de Bayes

- Pelo Teorema de Bayes:

$$P(terrorista|muçulmano) = \frac{P(muçulmano|terrorista)P(terrorista)}{p(muçulmano)}$$

- Ou seja: A probabilidade de um muçulmano ser terrorista = (probabilidade de um terrorista ser muçulmano * a probabilidade de um ser humano ser terrorista) / a probabilidade de um ser humano ser muçulmano.

Probabilidade: noções e inferência

Teorema de Bayes

- Como chegar nesses termos?
- Digamos que 75% de todos os terroristas são muçulmanos – historicamente é bem menos, mas tudo bem, erremos pra cima.
- Há cerca de 400 casos de terrorismo registrados por ano. Se cada um deles foi cometido por cinco pessoas diferentes, a expectativa de vida média dos terroristas fosse de 80 anos e nenhum deles morrer no atentado, haveria 400 * 5 * 80 = 160 mil terroristas no mundo. A probabilidade de alguém aleatório ser terrorista seria igual a 160.000 / 7.000.000.000 = 0.002%. Seguramente esse número exagera bastante pra cima.
- Cerca de 25% da população mundial é muçulmana. Então a probabilidade de um humano selecionado aleatoriamente ser muçulmano é de 0.25.

Probabilidade: noções e inferência

Teorema de Bayes

- Portanto:

$$\begin{aligned} P(terrorista|mu\c{c}ulmano) &= \frac{P(mu\c{c}ulmano|terrorista)P(terrorista)}{p(mu\c{c}ulmano)} \\ &= \frac{75\% \cdot 0,002\%}{25\%} \\ &= 0,006\% \end{aligned}$$

- Conclusão: Provavelmente esse número exagera pra cima. Mas, mesmo que ele fosse certo, isso significaria que a gente estaria julgando 99,994% dos muçulmanos pelos atos dos 0,006%.

Sumário

Probabilidade: funções e distribuições

Funções de probabilidade

- Se nós organizarmos todo o universo amostral (todas as possibilidades de resultado em nosso experimento) e as probabilidades referentes a ele, teremos uma função de probabilidade.
- Como chegamos numa função de probabilidade?

Probabilidade: funções e distribuições

Funções de probabilidade

- Imagine que nosso experimento seja jogar uma moeda (com dois lados: cara - C - e coroa - K) duas vezes seguidas.
- Quais são as possibilidades de resultado, o nosso espaço amostral (S)?
- $S = \{CC, CK, KC, KK\}$
- Qual a probabilidade referente a cada um desses resultados?

x	2 caras	cara + coroa	2 coroas
p(x)	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

Probabilidade: funções e distribuições

Funções de probabilidade

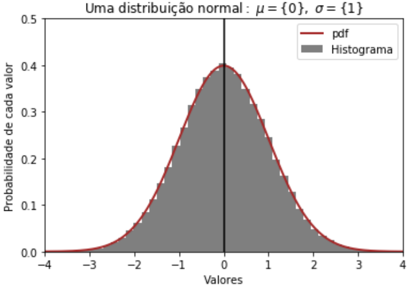
- Do mesmo modo que probabilidades sempre variam entre 0 e 1, a soma de todos os resultados possíveis (e mutuamente excludentes) em uma função de probabilidade tem de ser igual a um.
- Nesse caso:

$$p(2 \quad caras) + p(cara \quad + \quad coroa) + p(2 \quad coroas) = 1$$

Probabilidade: funções e distribuições

A distribuição normal

- Ao repetir um mesmo experimento diversas vezes, podemos observar empiricamente qual é a probabilidade de cada resultado, no longo prazo
- E podemos também aproximar os nossos resultados empíricos com uma função contínua, que resulta na probabilidade de cada possível valor infinitesimal



Probabilidade: funções e distribuições

A distribuição normal

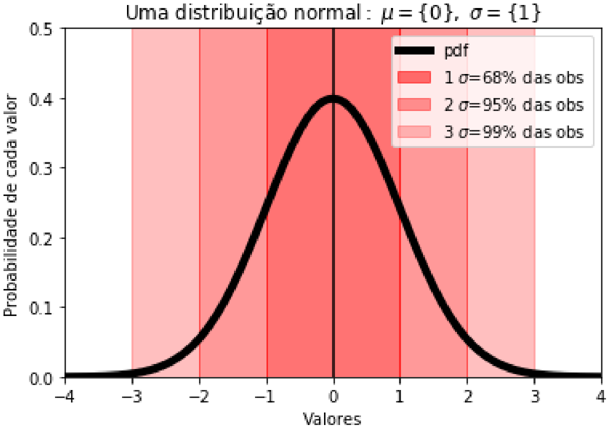
- A distribuição normal define-se, matematicamente, da seguinte forma:

$$\mathcal{N}(\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- O mais importante, contudo, é entender a regularidade quanto a proporção probabilística em cada intervalo seu.
  - 68% das distribuições estão entre $\pm$ 1 desvio padrão da média
  - 95% das distribuições estão entre $\pm$ 2 desvios padrões da média
  - 99% das distribuições estão entre $\pm$ 3 desvios padrões da média

Probabilidade: funções e distribuições

A distribuição normal



Probabilidade: funções e distribuições

A distribuição normal

```
import numpy as np
import matplotlib.pyplot as plt
import scipy
import pandas as pd

mu = 0
sigma = 1
draws = 100000
bins = 50

x = np.random.normal(mu,sigma,draws)
pdf = scipy.stats.norm.pdf(x)
df = pd.DataFrame(data=[x,pdf], index=['x','pdf']).T
df = df.sort_values('x')
```

Probabilidade: funções e distribuições

A distribuição normal

```
fig = plt.figure()

plt.plot('x', 'pdf', data=df,
         color='black', linewidth=5)

# areas sombreadas
plt.axvspan(xmin=-1,xmax=1,color='red',alpha=.55, label=r'1 $\
plt.axvspan(xmin=-2,xmax=-1,color='red',alpha=.4, label=r'2 $\
plt.axvspan(xmin=1,xmax=2,color='red',alpha=.4)
plt.axvspan(xmin=-3,xmax=-2,color='red',alpha=.25, label=r'3 $
plt.axvspan(xmin=2,xmax=3,color='red',alpha=.25)

# linha preta
plt.axvline(x=0, color='black')
```

Probabilidade: funções e distribuições

A distribuição normal

```
plt.legend(loc=1)
plt.xlabel('Valores')
plt.ylabel('Probabilidade de cada valor')
plt.title(r'$\mathrm{Uma\ distribuição\ normal:}\
\mu=\{0\},\ \sigma=\{1\}$')
plt.axis([-4, 4, 0, 0.5])  # eixos
plt.grid(False)
plt.show()
```

Probabilidade: funções e distribuições

Escore-z

- No nosso exemplo, a média é zero e o desvio padrão é um, de tal modo que o valor absoluto de X denota qual a distância, em desvios padrões, de X da média.
- Na realidade, muito raramente encontraremos observações assim

Probabilidade: funções e distribuições

Escore-z

- Felizmente, nós podemos padronizar as nossas observações, fazendo-as com que elas pareçam o nosso exemplo
- Fazemos isso subtraindo a média (μ) e dividindo pelo desvio padrão (σ) da distribuição, de tal modo que a nossa amostra passe a medir, em desvios padrões, a distância de cada observação da média
- Essa medida é chamada de escore-z:

$$z = \frac{X - \mu}{\sigma} \tag{15}$$

Probabilidade: funções e distribuições

Escore-z

- Por exemplo, se sabemos que a média das alturas das mulheres é de $1,627$ metros e o desvio padrão é de $0,62$ metros, qual a distância da média, em desvios padrões, de uma mulher que tem $1,50m$?

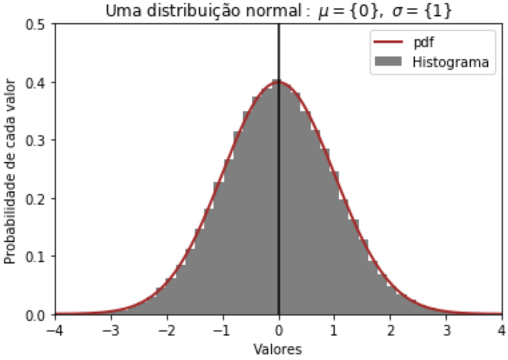
$$z = \frac{1,50 - 1,627}{0,62} = -0,20 \tag{16}$$

- Isso significa que uma mulher de $1,50m$ de altura está $0,2\sigma$ abaixo da média.
- Com base nessa medida, podemos, com a ajuda de tabelas disponíveis em livros de estatística, derivar quantos porcentos das mulheres estão abaixo ou acima daquela distribuição.

Probabilidade: funções e distribuições

Funções de densidade

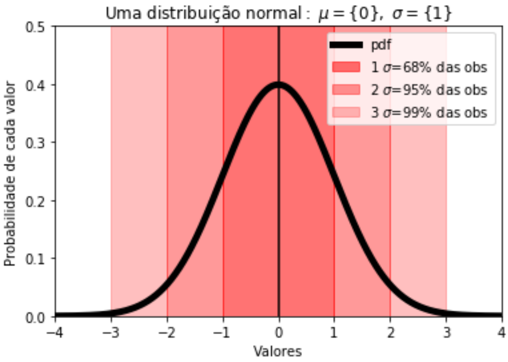
- Vocês já estão acostumados a ver funções de densidade da seguinte maneira, seja por meio de histogramas ou, como agora, por meio de uma função contínua para qualquer variável aleatória:



Probabilidade: funções e distribuições

Funções de densidade

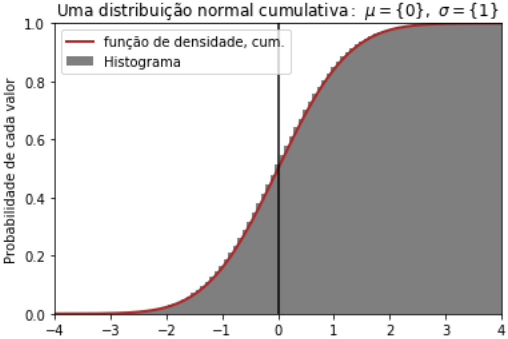
- E vocês também já entenderam que para cada número de desvios padrões ao redor da média, existe uma porcentagem do total da probabilidade do espaço amostral:



Probabilidade: funções e distribuições

Funções de densidade

- Mas existe uma outra maneira de se observar essas distribuições, não pela probabilidade de cada X, mas pela probabilidade acumulada até X.
- Uma distribuição cumulativa denota a porcentagem da distribuição que está abaixo de determinada variável aleatória X



Probabilidade: funções e distribuições

Funções de densidade

```
import numpy as np
import matplotlib.pyplot as plt
import scipy
import pandas as pd

mu = 0
sigma = 1
draws = 100000
bins = 50

x = np.random.normal(mu,sigma,draws)
cdf = scipy.stats.norm.cdf(x)
df = pd.DataFrame(data=[x,cdf], index=['x','cdf']).T
df = df.sort_values('x')
```

Probabilidade: funções e distribuições

A distribuição normal

```
hist3 = plt.figure()

plt.hist(x,
         bins=bins,
         normed=True,
         facecolor='black',
         cumulative = True,
         alpha=0.5,
         data=df,
         label='Histograma' )

plt.plot('x', 'cdf', data=df,
         color='brown', linewidth=2, label='função de densidad
```

Probabilidade: funções e distribuições

A distribuição normal

```
# black line
plt.axvline(x=0, color='black')

plt.legend(loc="upper left")
plt.xlabel('Valores')
plt.ylabel('Probabilidade de cada valor')
plt.title(r'$\mathrm{Uma\ distribuição\ normal\ cumulativa:}\
plt.axis([-4, 4, 0, 1])  # set range of axes
plt.grid(False)  # add grid
plt.show()  # plot chart
```

Probabilidade: funções e distribuições

Funções de densidade

- Compreendendo o conceito de distribuição cumulativa, podemos utilizar cortes específicos de uma distribuição normal para entender o percentual probabilístico da distribuição abaixo de determinada variável aleatória X
- Tradicionalmente, faz-se isso calculando o escore-z e depois consultando tabelas que vêm ao final de livros de estatística.

Probabilidade: funções e distribuições

Funções de densidade

- No Python, podemos ver qual a porcentagem da distribuição abaixo de determinada variável aleatória X (ou determinado corte de desvio padrão, caso $\mu = 0, \sigma = 1$) com a ajuda do scipy:

```
import scipy
#loc = média; scale = desvio padrão
scipy.stats.norm.cdf(0.2, loc=0, scale=1)
```

- Para calcular quantos porcento da amostra estão entre -1 e $+1$ desvios padrão da média, como fazemos (já sabemos que é 68%)?

```
scipy.stats.norm.cdf(1, 0, 1)
- scipy.stats.norm.cdf(-1, 0, 1)
```

Probabilidade: funções e distribuições

Escore-z

- E podemos também fazer o inverso: dizer para o Python qual é o porcentagem de corte que queremos, e ele retorna o valor:

```
scipy.stats.norm.ppf(0.5, loc=0, scale=1)
```