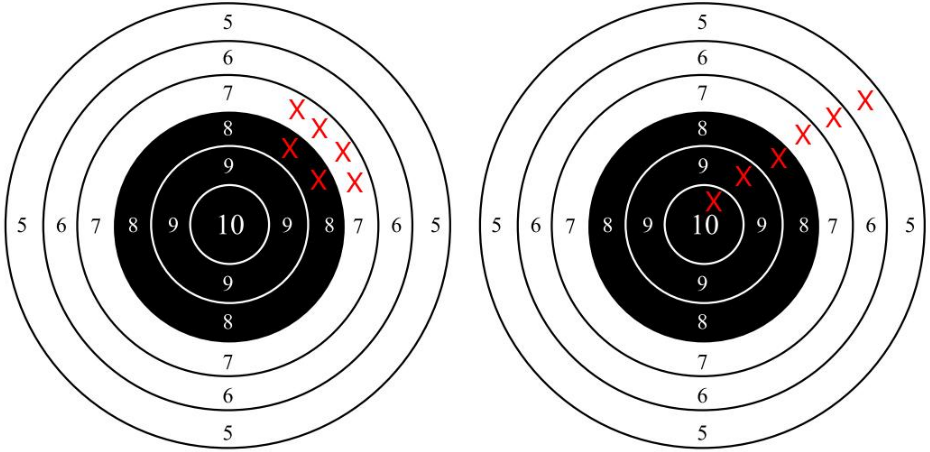
Métodos Quantitativos
Aula 02

Revisão de Estatística

Roberto Massi de Oliveira
Alex Borges Vieira

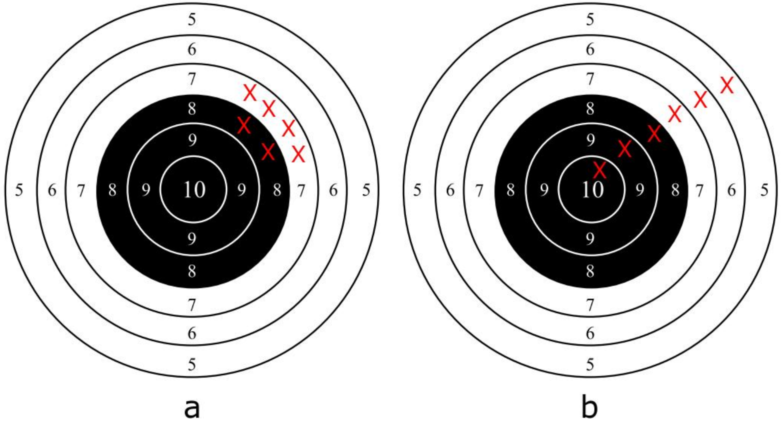
Medidas de Localização

- Média, mediana, moda
- Resumem um conjunto de dados em um único valor
- Uso isolado pode omitir informações relevantes
- Ex. 01: Qual atirador é mais preciso: **a** ou **b**?



a b

Medidas de Localização



- Método 1: observação da figura
    - Conclusão:
      O atirador **a** é mais preciso, pois acertou a maioria dos tiros próximos entre si e do centro

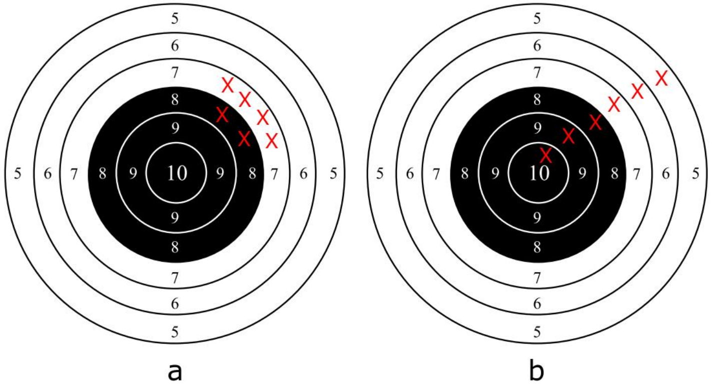
- Método 2: observação das médias

```python
import numpy as np

a = [8, 8, 7, 7, 7, 7]
mean_a = np.mean(a)

b = [10, 9, 8, 7, 6, 5]
mean_b = np.mean(b)
```

    - mean_a = 7,33
    - mean_b = 7,50
    - Conclusão: O atirador **b** é mais preciso, pois, em média, possui maior pontuação
    - Como evitar esse problema? Usar medidas de localização junto com **medidas de dispersão**.

Medidas de Dispersão

- Variância, desvio padrão
- Determinam a variabilidade, ou dispersão, dos dados de uma dada amostra
- No Ex. 01, se, além da média, tivéssemos algum número que representasse a variação de distância entre os tiros, poderíamos concluir que o atirador **a** é mais preciso que o **b**.



Média Aritmética Simples

- Medida de localização de centro de amostra
- Pode resumir um conjunto de dados com um único número
- Numa amostra **x** com **n** elementos, é dada por:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

```python
import numpy as np

x = [20000, 14000, 2000, 800]
mean_x = np.mean(x)
```

Média Aritmética Simples

- Ex. 02: Suponha que, numa empresa, os salários são atribuídos aos cargos conforme especificado na tabela:

Cargo	Salário
Presidente	R$20.000,00
Vice-presidente	R$14.000,00
Engenheiros	R$2.000,00
Estagiários	R$800,00

- A média salarial é de R$ 9.200,00. Esse dado passa a falsa impressão de que os funcionários são bem remunerados.

Média Ponderada

- É uma extensão da média simples, que utiliza pesos para as informações do conjunto de dados

- Numa amostra **x** de tamanho **n** com **k** pesos, é dada por:

$$\bar{x} = \frac{\sum_{i=1}^{k} k_i x_i}{n}$$

```
import numpy as np

x = [20000, 14000, 2000, 800]
ni = [1, 1, 12, 6]
average_x = np.average(x, weights = ni)
```

Média Ponderada

- Ex. 03: Continuando o Ex. 02, se acrescentarmos a quantidade de funcionários que ocupam cada cargo, temos:

Cargo	Quantidade	Salário
Presidente	1	R$20.000,00
Vice-presidente	1	R$14.000,00
Engenheiros	12	R$2.000,00
Estagiários	6	R$800,00

- Nesse caso, a média salarial ponderada é de R$ 3.140,00. Então, temos noção um pouco mais realista do que a obtida através da média simples

Mediana

- Outra localização de centro de distribuição de dados
- Separa a amostra em dois subgrupos: os maiores e os menores
- Ex. 04: Numa amostra x = [1, 5, 4, 2, 3] qual é a mediana?
    - Primeiro, ordena-se a amostra: x = [1, 2, 3, 4, 5]
    - O valor central é a mediana: median_x = 3

```python
import numpy as np

x = [1, 5, 4, 2, 3]
median_x = np.median(x)
```

- Se o número de elementos na amostra é par, a mediana é igual à média dos dois números centrais da mesma
    - Qual a mediana da amostra x = [1, 2, 3, 4, 5, 6]?

Moda

- É o valor mais frequente em uma amostra
- Usada para identificar tendências
- Ex. 05: Numa amostra x = [1, 1, 1, 2, 2, 3], qual a moda?
    - O número 1 aparece 3 vezes
    - O número 2 aparece 2 vezes
    - O número 3 aparece 1 vez
    - Então, o número 1 é a moda.

```python
import numpy as np

x = [1, 1, 1, 2, 2, 3]
counts = np.bincount(x)
moda_x = np.argmax(counts)
```

```python
import numpy as np

x = [3, 2, 2, 1, 1, 1]
counts = np.bincount(x)
print(counts)
```

```
[0 3 2 1]
```

```python
moda_x = np.argmax(counts)
print(np.argmax(counts))
```

```
1
```

Variância

- Mede a dispersão estatística de uma amostra
- Mostra o quão longe os valores de uma amostra estão em relação à média
- Em uma amostra **x** de tamanho **n**, a variância é dada por:

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$

https://pt.khanacademy.org/math/ap-statistics/summarizing-quantitative-data-ap/more-standard-deviation/v/another-simulation-giving-evidence-that-n-1-gives-us-an-unbiased-estimate-of-variance

```python
import numpy as np

x = [10, 9, 8, 7, 6, 5]
var_x = np.var(x)
```

Desvio Padrão

- Outra medida de dispersão, calculada através da raiz quadrada da variância
- Menos sensível a *outliers* que a variância
- Comumente usado em conjunto com a média para expressar a confiabilidade

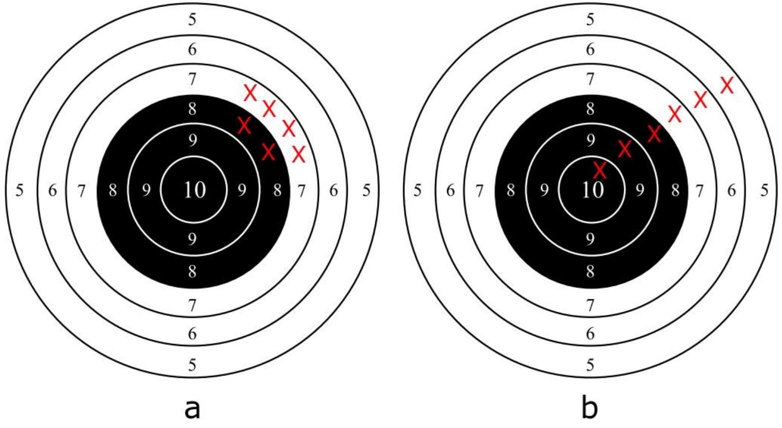
$$\mathbf{s} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

```python
import numpy as np

x = [10, 9, 8, 7, 6, 5]
std_x = np.std(x)
```

Desvio Padrão

- Ex. 06: Lembre-se que, no Ex. 01, calculamos as médias mean_a = 7,33 e mean_b = 7,50 e chegamos à conclusão errônea de que o atirador **b** é mais preciso. Como resolver esse problema?
    - Vamos calcular o desvio padrão de ambos.

```python
import numpy as np

a = [8, 8, 7, 7, 7, 7]
std_a = np.std(a)

b = [10, 9, 8, 7, 6, 5]
std_b = np.std(b)
```



    - std_a = 0,47
    - std_b = 1,71
    - Conclusão: Com pequena diferença entre as médias, os resultados do atirador **a** possuem um desvio padrão muito menor que os do **b**. Portanto, o atirador **a** é mais preciso.

Parênteses: Outlier

- Ex. 07: Como um outlier afeta as estatísticas amostrais?

```
import numpy as np
import matplotlib.pyplot as plt

sample1 = [1, 2, 2, 2, 5, 6, 7, 8, 10, 10]
sample2 = [1, 2, 2, 2, 5, 6, 7, 8, 10, 100]
```

```
print('SAMPLE 1')
print('-------------------')
print('média =', np.mean(sample1))
print('mediana =', np.median(sample1))
print('moda =', np.argmax(np.bincount(sample1)))
print('variância =', round(float(np.var(sample1)),2))
print('desvio padrão =', round(float(np.std(sample1)),2))
print('-------------------')
```

```
print('SAMPLE 2')
print('-------------------')
print('média =', np.mean(sample2))
print('mediana =', np.median(sample2))
print('moda =', np.argmax(np.bincount(sample2)))
print('variância =', round(float(np.var(sample2)),2))
print('desvio padrão =', round(float(np.std(sample2)),2))
print('-------------------')
```

```
SAMPLE 1
-------------------
média = 5.3
mediana = 5.5
moda = 2
variância = 10.61
desvio padrão = 3.26
-------------------
```

```
SAMPLE 2
-------------------
média = 14.3
mediana = 5.5
moda = 2
variância = 824.21
desvio padrão = 28.71
-------------------
```

Coeficiente de Variação

- C.V. é o desvio padrão expresso como uma porcentagem média
- Útil para comparar a variação de conjuntos de observações que diferem na média ou são medidos em grandezas diferentes (unidades diferentes)
  - CVs menores passam a ideia de estabilidade, solidez, homogeneidade
  - Ex.: Se o CV da pulsação humana CVp = 0,127 e o CV do ácido úrico CVa = 0,232, conclui-se que a pulsação é mais estável que o ácido úrico.

$$CV = \frac{s}{\bar{x}}$$

Coeficiente de Correlação

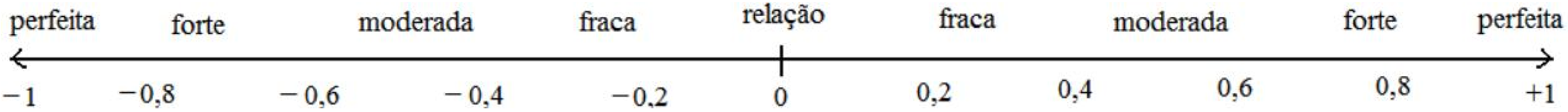
- Quando dispomos das amostras **x** e **y** de dados bivariados (e.g., peso e altura de um grupo de indivíduos), o coeficiente de correlação é dado por:

$$r_{xy} = \frac{Cov(X,Y)}{S_x S_y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

```
import numpy as np

x = [1.6, 1.7, 1.8, 1.9]
y = [60, 70, 80, 90]
xy = [x, y]

r = np.corrcoef(xy)
```

Coeficiente de Correlação

- Varia de -1,0 a 1,0.



- Quando r > 0, à medida que **x** cresce também cresce **y** (em média)

- Quando r < 0, à medida que **x** cresce, **y** decresce (em média)

Coeficiente de Correlação

- Ex. 08: Um aluno, com bastante dificuldade numa dada disciplina, foi estudando cada vez mais para melhorar suas notas a cada prova e evitar a reprovação. A tabela abaixo resume o número de horas estudadas por dia antes de cada prova realizada e a nota tirada nas mesmas. Qual o coeficiente de correlação entre as horas estudadas por dia e as notas?

Estudo/Dia	1 h	2 h	3 h
Nota	3,0	7,0	9,0

```python
import numpy as np

t = [1, 2, 3]
n = [3, 7, 10]
tn = [t, n]

r = np.corrcoef(tn)
```

r =

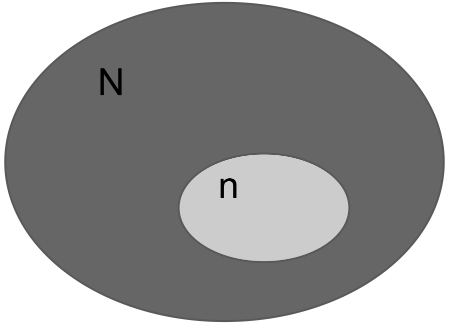
r(t,t)	r(t,n)
r(n,t)	r(n,n)

r =

1	0,997
0,997	1

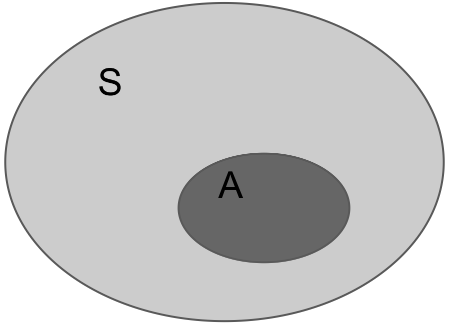
Teoria de Conjuntos

- População (ou universo): todos os **N** membros de uma classe ou grupo
  - Ex.: todos os processos executados em uma máquina
- Amostra: uma parte da população, denotada por **n**
  - Ex.: todos os processos executados em uma máquina em 12/12/2012



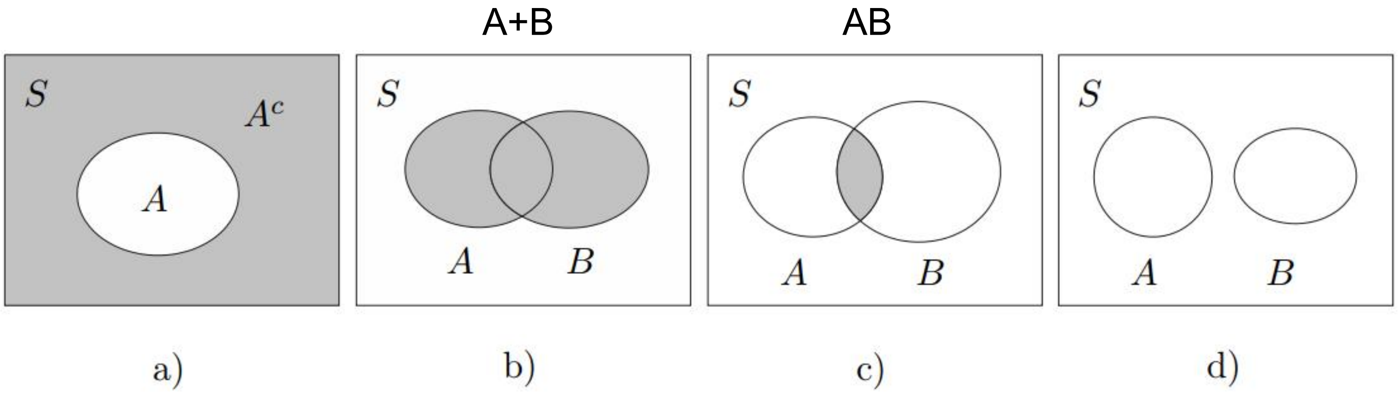
Teoria de Conjuntos

- Espaço amostral: conjunto de possibilidades de resultados, denotado por **S**
  - Ex: S = {1, 2, 3, 4, 5, 6} é o espaço amostral do lançamento de um dado de 6 faces
- Evento: Conjunto de resultados (subconjunto do espaço amostral) aos quais são associados probabilidades. Denotado por letras maiúsculas
  - Ex.: A = {2, 4, 6} é o evento no qual os resultados do lançamento do dado são pares



Teoria de Conjuntos

- Tipos de eventos (diagramas de Venn):



a) complemento, b) união, c) interseção de eventos, e d) eventos disjuntos

Definições de Probabilidade

- A probabilidade **P(A)** de um evento **A**, onde n_A é o número de ocorrências de **A** e n é tamanho do espaço amostral, é dada por:

$$P(A) = \frac{n_A}{n}$$

- Se **A** e **B** eventos mutuamente exclusivos, ou disjuntos ($\mathrm{A} \cap \mathrm{B} = \emptyset$, eventos que não podem ocorrer ao mesmo tempo; e.g., resultados do lançamento de uma moeda), então:

$$P(A+B) = P(A) + P(B)$$

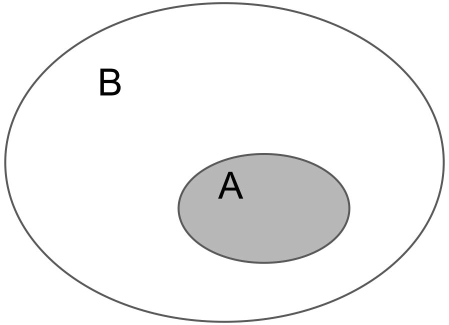
- Senão:

$$P(A+B) = P(A) + P(B) - P(AB)$$

Definições de Probabilidade

- Algumas outras propriedades do cálculo de probabilidade:
  - $0 \leq P(A) \leq 1$
  - $P(S)=1$
  - Se $A \subset B$, então $P(A) \leq P(B)$

  $P(\emptyset)=0$ e $P(\bar{A})=1-P(A)$, sendo $\emptyset$ o vazio e $\bar{A}$ o complementar de A



Definições de Probabilidade

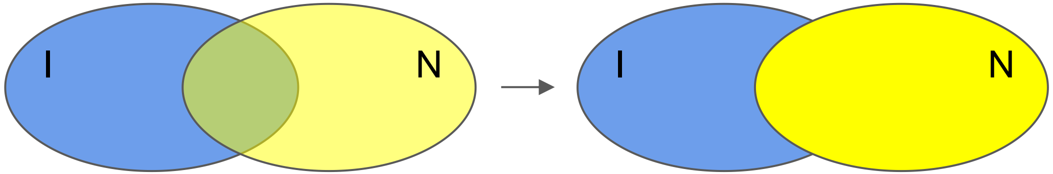
- Ex. 9: Na fila de processos de um computador, há 5 processos I/O bound, 3 processos CPU bound e 4 processos memory bound. Supondo que a resolução da fila de processos se dá de forma totalmente aleatória e que os tipos de processos são disjuntos entre si, qual a probabilidade do primeiro processo a ser atendido seja I/O bound ou CPU bound?
    - Do enunciado:
        - Número total de processos: (n = 5 + 3 + 4 = 12)
        - Tipos disjuntos: e.g., não existem processos ao mesmo tempo I/O bound e CPU bound
        - Sendo n_I o número de processos I/O bound e n_C o número de processos CPU bound

$$P(I \cup C) = P(I + C) = P(I) + P(C) = \frac{n_I}{n} + \frac{n_C}{n} = \frac{5+3}{12} = 0{,}67$$

Definições de Probabilidade

- Ex. 10: Suponha um grupo de 100 pessoas, no qual algumas têm psicose (P), enquanto outras têm neurose (N), sendo algumas idosas (I), enquanto outras são adolescentes (A). A tabela seguinte dá a classificação das referidas pessoas. Escolhendo-se ao acaso uma pessoa desse grupo, qual será a probabilidade dessa ser idosa ou ter alguma neurose?

	P	**N**	**Total**
A	1	29	30
I	2	68	70
Total	3	97	100



$$P(I \cup N) = P(I) + P(N) - P(I \cap N) = \frac{70+97-68}{100} = \frac{99}{100}$$

Probabilidade Condicional

- Probabilidade de ocorrência de um evento **B**, dado que um evento **A** ocorra
    - Ex.: Duas chances para responder uma questão com 5 opções de resposta. A probabilidade de acerto da segunda tentativa é influenciada pela primeira.

- Denotada por P(A|B) e obtida por:

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

- Teorema de Bayes:

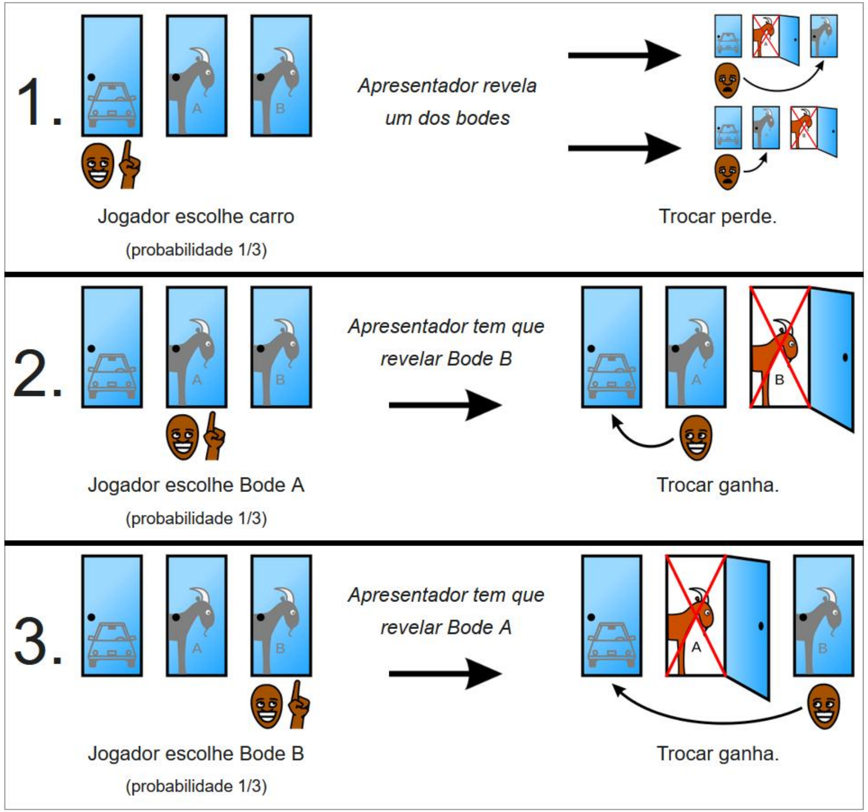
$$P(AB) = P(A|B)P(B) = P(B|A)P(A)$$

Probabilidade Condicional

- Ex. 11:

Probabilidade Condicional

- Ex. 11:



O jogador tem uma chance igual de inicialmente selecionar o carro, Bode A, ou Bode B. A troca resulta em uma vitória 2/3 das vezes.

https://pt.wikipedia.org/wiki/Problema_de_Monty_Hall

Probabilidade Condicional

- Ex. 12: Calcule a probabilidade de obter **soma 8** no lançamento de dois dados em que o resultado do lançamento foram dois números **ímpares**.
  - Seja A = Obter soma 8 e B = Obter dois números ímpares

Probabilidade Condicional

- Ex. 12: Calcule a probabilidade de obter **soma 8** no lançamento de dois dados em que o resultado do lançamento foram dois números **ímpares**.
    - Seja A = Obter soma 8 e B = Obter dois números ímpares
    - $P(A \cap B)$ é a probabilidade de se obter apenas números ímpares que somam 8 no lançamento de dois dados. As únicas 2 combinações possíveis são: {3,5} e {5,3}

Probabilidade Condicional

- Ex. 12: Calcule a probabilidade de obter **soma 8** no lançamento de dois dados em que o resultado do lançamento foram dois números **ímpares**.
    - Seja A = Obter soma 8 e B = Obter dois números ímpares
    - $P(A \cap B)$ é a probabilidade de se obter apenas números ímpares que somam 8 no lançamento de dois dados. As únicas 2 combinações possíveis são: {3,5} e {5,3}
    - Total de combinações possíveis no lançamento de 2 dados é 36 ({1,1}, {1,2}, {1,3}, …, {6,6})

Probabilidade Condicional

- Ex. 12: Calcule a probabilidade de obter **soma 8** no lançamento de dois dados em que o resultado do lançamento foram dois números **ímpares**.
  - Seja A = Obter soma 8 e B = Obter dois números ímpares
  - P(A∩B) é a probabilidade de se obter apenas números ímpares que somam 8 no lançamento de dois dados. As únicas 2 combinações possíveis são: {3,5} e {5,3}
  - Total de combinações possíveis no lançamento de 2 dados é 36 ({1,1}, {1,2}, {1,3}, …, {6,6})
  - Portanto, P(A∩B) = 2/36

Probabilidade Condicional

- Ex. 12: Calcule a probabilidade de obter **soma 8** no lançamento de dois dados em que o resultado do lançamento foram dois números **ímpares**.
  - Seja A = Obter soma 8 e B = Obter dois números ímpares
  - P(A∩B) é a probabilidade de se obter apenas números ímpares que somam 8 no lançamento de dois dados. As únicas 2 combinações possíveis são: {3,5} e {5,3}
  - Total de combinações possíveis no lançamento de 2 dados é 36 ({1,1}, {1,2}, {1,3}, …, {6,6})
  - Portanto, P(A∩B) = 2/36
  - P(B) é a probabilidade de obter somente números ímpares no lançamento de dois dados. As únicas combinações dentro das 36 possíveis são: {1,1}; {1,3}; {1,5}; {3,1}; {3,3}; {3,5}; {5,1}; {5,3} e {5,5}

Probabilidade Condicional

- Ex. 12: Calcule a probabilidade de obter **soma 8** no lançamento de dois dados em que o resultado do lançamento foram dois números **ímpares**.
  - Seja A = Obter soma 8 e B = Obter dois números ímpares
  - P(A∩B) é a probabilidade de se obter apenas números ímpares que somam 8 no lançamento de dois dados. As únicas 2 combinações possíveis são: {3,5} e {5,3}
  - Total de combinações possíveis no lançamento de 2 dados é 36 ({1,1}, {1,2}, {1,3}, …, {6,6})
  - Portanto, P(A∩B) = 2/36
  - P(B) é a probabilidade de obter somente números ímpares no lançamento de dois dados. As únicas combinações dentro das 36 possíveis são: {1,1}; {1,3}; {1,5}; {3,1}; {3,3}; {3,5}; {5,1}; {5,3} e {5,5}
  - Portanto, P(B) = 9/36

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

Probabilidade Condicional

- Ex. 12: Calcule a probabilidade de obter **soma 8** no lançamento de dois dados em que o resultado do lançamento foram dois números **ímpares**.
  - Seja A = Obter soma 8 e B = Obter dois números ímpares
  - P(A∩B) é a probabilidade de se obter apenas números ímpares que somam 8 no lançamento de dois dados. As únicas 2 combinações possíveis são: {3,5} e {5,3}
  - Total de combinações possíveis no lançamento de 2 dados é 36 ({1,1}, {1,2}, {1,3}, …, {6,6})
  - Portanto, P(A∩B) = 2/36
  - P(B) é a probabilidade de obter somente números ímpares no lançamento de dois dados. As únicas combinações dentro das 36 possíveis são: {1,1}; {1,3}; {1,5}; {3,1}; {3,3}; {3,5}; {5,1}; {5,3} e {5,5}
  - Portanto, P(B) = 9/36
  - $P(A|B) = \frac{P(A \cap B)}{P(B)} = (2/36) / (9/36) = 2/9$

Eventos Independentes

- A ocorrência do evento **A** não afeta a ocorrência do evento **B**
  - Isso não quer dizer que eles não tenham interseções
  - Eventos mutuamente exclusivos NÃO SÃO INDEPENDENTES
  - Coeficiente de correlação é zero, pois cov(A,B) = 0
  - Ex: Resultados de 2 lançamentos de moeda
- Se **A** e **B** são independentes, então:

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

- Consequentemente, pela regra de Bayes:

$$P(A \cap B) = P(AB) = P(A)P(B)$$

Eventos Independentes

- Ex. 13: Uma moeda é lançada duas vezes. Calcule a probabilidade de obtermos cara no segundo lançamento sabendo que obtivemos cara no primeiro lançamento.
  - Considerando que "**c**" representa cara e "**k**" coroa, temos:
    - **S = {cc, ck, kc, kk}**
    - **n = 4**

Eventos Independentes

- Ex. 13: Uma moeda é lançada duas vezes. Calcule a probabilidade de obtermos cara no segundo lançamento sabendo que obtivemos cara no primeiro lançamento.
  - Considerando que "**c**" representa cara e "**k**" coroa, temos:
    - **S = {cc, ck, kc, kk}**
    - **n = 4**
  - Evento A = {obtenção de **c** no 2º lançamento}, Evento B = {obtenção de c no 1º lançamento}
    - **A = {cc, kc}, B = {cc, ck}, AB = {cc}**

Eventos Independentes

- Ex. 13: Uma moeda é lançada duas vezes. Calcule a probabilidade de obtermos cara no segundo lançamento sabendo que obtivemos cara no primeiro lançamento.
  - Considerando que "**c**" representa cara e "**k**" coroa, temos:
    - **S = {cc, ck, kc, kk}**
    - **n = 4**
  - Evento A = {obtenção de **c** no 2º lançamento}, Evento B = {obtenção de c no 1º lançamento}
    - **A = {cc, kc}, B = {cc, ck}, AB = {cc}**
  - Portanto: **P(A) = P(B) = 2/4 = 1/2, P(AB) = 1/4**

Eventos Independentes

- Ex. 13: Uma moeda é lançada duas vezes. Calcule a probabilidade de obtermos cara no segundo lançamento sabendo que obtivemos cara no primeiro lançamento.
  - Considerando que "**c**" representa cara e "**k**" coroa, temos:
    - **S = {cc, ck, kc, kk}**
    - **n = 4**
  - Evento A = {obtenção de **c** no 2º lançamento}, Evento B = {obtenção de c no 1º lançamento}
    - **A = {cc, kc}, B = {cc, ck}, AB = {cc}**
  - Portanto: **P(A) = P(B) = 2/4 = 1/2, P(AB) = 1/4**
  - Sabemos que **P(A|B) = P(AB)/P(B) = (1/4) / (2/4) = 1/2**

Eventos Independentes

- Ex. 13: Uma moeda é lançada duas vezes. Calcule a probabilidade de obtermos cara no segundo lançamento sabendo que obtivemos cara no primeiro lançamento.
    - Considerando que "**c**" representa cara e "**k**" coroa, temos:
        - **S = {cc, ck, kc, kk}**
        - **n = 4**
    - Evento A = {obtenção de **c** no 2º lançamento}, Evento B = {obtenção de c no 1º lançamento}
        - **A = {cc, kc}, B = {cc, ck}, AB = {cc}**
    - Portanto: **P(A) = P(B) = 2/4 = 1/2, P(AB) = 1/4**
    - Sabemos que **P(A|B) = P(AB)/P(B) = (1/4) / (2/4) = 1/2**
    - Observem que **P(A|B) = P(A)**

Eventos Independentes

- Ex. 14: Lança-se uma moeda 3 vezes. Sejam os eventos:
  A: ocorrem três caras ou três coroas; **B**: ocorrem ao menos duas caras;
  C: ocorrem no máximo duas caras.

  Dos pares **(A,B)**, **(A,C)** e **(B,C)** quais são independentes?
  - Considerando que "**c**" representa cara e "**k**" coroa, temos:
    - **S = {ccc, kkk, cck, ckc, kcc, ckk, kck, kkc}, n = 8**

Eventos Independentes

- Ex. 14: Lança-se uma moeda 3 vezes. Sejam os eventos:
  A: ocorrem três caras ou três coroas; **B**: ocorrem ao menos duas caras;
  C: ocorrem no máximo duas caras.

  Dos pares **(A,B)**, **(A,C)** e **(B,C)** quais são independentes?
  - Considerando que "**c**" representa cara e "**k**" coroa, temos:
    - **S = {ccc, kkk, cck, ckc, kcc, ckk, kck, kkc}, n = 8**
  - **A = {ccc, kkk}, B = {cck, ckc, kcc, ccc}, C = {kkk, ckk, kck, kkc, kcc, ckc, cck}**

Eventos Independentes

- Ex. 14: Lança-se uma moeda 3 vezes. Sejam os eventos:
  A: ocorrem três caras ou três coroas; **B**: ocorrem ao menos duas caras;
  C: ocorrem no máximo duas caras.

  Dos pares **(A,B)**, **(A,C)** e **(B,C)** quais são independentes?
  - Considerando que "**c**" representa cara e "**k**" coroa, temos:
    - **S = {ccc, kkk, cck, ckc, kcc, ckk, kck, kkc}, n = 8**
  - **A = {ccc, kkk}, B = {cck, ckc, kcc, ccc}, C = {kkk, ckk, kck, kkc, kcc, ckc, cck}**
  - **AB = {ccc}, AC = {ccc}, BC = {cck, ckc, kcc}**

Eventos Independentes

- Ex. 14: Lança-se uma moeda 3 vezes. Sejam os eventos:
  A: ocorrem três caras ou três coroas; **B**: ocorrem ao menos duas caras;
  C: ocorrem no máximo duas caras.

  Dos pares **(A,B)**, **(A,C)** e **(B,C)** quais são independentes?
  - Considerando que "**c**" representa cara e "**k**" coroa, temos:
    - **S = {ccc, kkk, cck, ckc, kcc, ckk, kck, kkc}, n = 8**
  - **A = {ccc, kkk}, B = {cck, ckc, kcc, ccc}, C = {kkk, ckk, kck, kkc, kcc, ckc, cck}**
  - **AB = {ccc}, AC = {kkk}, BC = {cck, ckc, kcc}**
  - **P(A) = 2/8 = 1/4, P(B) = 4/8 = 1/2, P(C) = 7/8, P(AB) = 1/8, P(AC) = 1/8, P(BC) = 3/8**

Eventos Independentes

- Ex. 14: Lança-se uma moeda 3 vezes. Sejam os eventos:
  A: ocorrem três caras ou três coroas; **B**: ocorrem ao menos duas caras;
  C: ocorrem no máximo duas caras.

  Dos pares **(A,B)**, **(A,C)** e **(B,C)** quais são independentes?
    - Considerando que "**c**" representa cara e "**k**" coroa, temos:
        - **S = {ccc, kkk, cck, ckc, kcc, ckk, kck, kkc}, n = 8**
    - **A = {ccc, kkk}, B = {cck, ckc, kcc, ccc}, C = {kkk, ckk, kck, kkc, kcc, ckc, cck}**
    - **AB = {ccc}, AC = {ccc}, BC = {cck, ckc, kcc}**
    - **P(A) = 2/8 = 1/4, P(B) = 4/8 = 1/2, P(C) = 7/8, P(AB) = 1/8, P(AC) = 1/8, P(BC) = 3/8**

Como provaremos independência? Pela regra de Bayes:

$$P(A \cap B) = P(AB) = P(A)P(B)$$

Eventos Independentes

- Ex. 14: Lança-se uma moeda 3 vezes. Sejam os eventos:
  A: ocorrem três caras ou três coroas; **B**: ocorrem ao menos duas caras;
  C: ocorrem no máximo duas caras.

  Dos pares **(A,B)**, **(A,C)** e **(B,C)** quais são independentes?
  - Considerando que "**c**" representa cara e "**k**" coroa, temos:
    - **S = {ccc, kkk, cck, ckc, kcc, ckk, kck, kkc}, n = 8**
  - **A = {ccc, kkk}, B = {cck, ckc, kcc, ccc}, C = {kkk, ckk, kck, kkc, kcc, ckc, cck}**
  - **AB = {ccc}, AC = {ccc}, BC = {cck, ckc, kcc}**
  - **P(A) = 2/8 = 1/4, P(B) = 4/8 = 1/2, P(C) = 7/8, P(AB) = 1/8, P(AC) = 1/8, P(BC) = 3/8**
  - **P(A)P(B) = 1/4 * 1/2 = 1/8 = P(AB). Portanto, A e B são independentes**

Eventos Independentes

- Ex. 14: Lança-se uma moeda 3 vezes. Sejam os eventos:
  A: ocorrem três caras ou três coroas; **B**: ocorrem ao menos duas caras;
  C: ocorrem no máximo duas caras.

  Dos pares **(A,B)**, **(A,C)** e **(B,C)** quais são independentes?
  - Considerando que "**c**" representa cara e "**k**" coroa, temos:
    - **S = {ccc, kkk, cck, ckc, kcc, ckk, kck, kkc}, n = 8**
  - **A = {ccc, kkk}, B = {cck, ckc, kcc, ccc}, C = {kkk, ckk, kck, kkc, kcc, ckc, cck}**
  - **AB = {ccc}, AC = {ccc}, BC = {cck, ckc, kcc}**
  - **P(A) = 2/8 = 1/4, P(B) = 4/8 = 1/2, P(C) = 7/8, P(AB) = 1/8, P(AC) = 1/8, P(BC) = 3/8**
  - **P(A)P(B) = 1/4 * 1/2 = 1/8 = P(AB). Portanto, A e B são independentes**
  - **P(A)P(C) = 1/4 * 7/8 = 7/32 != P(AC). Portanto, A e C não são independentes**

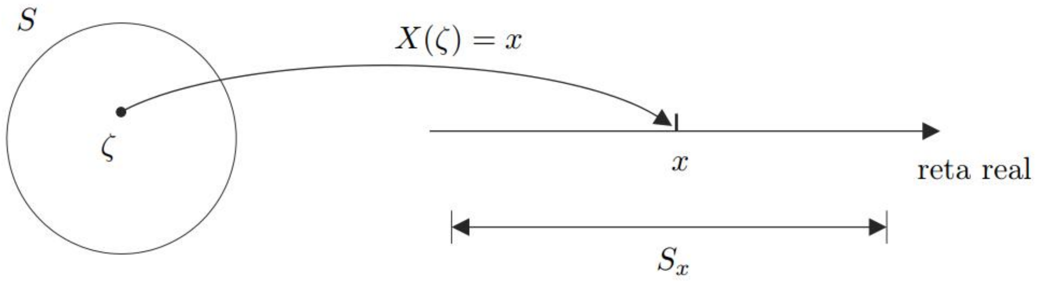
Eventos Independentes

- Ex. 14: Lança-se uma moeda 3 vezes. Sejam os eventos:
  A: ocorrem três caras ou três coroas; **B**: ocorrem ao menos duas caras;
  C: ocorrem no máximo duas caras.

  Dos pares **(A,B)**, **(A,C)** e **(B,C)** quais são independentes?
  - Considerando que "**c**" representa cara e "**k**" coroa, temos:
    - **S = {ccc, kkk, cck, ckc, kcc, ckk, kck, kkc}, n = 8**
  - **A = {ccc, kkk}, B = {cck, ckc, kcc, ccc}, C = {kkk, ckk, kck, kkc, kcc, ckc, cck}**
  - **AB = {ccc}, AC = {ccc}, BC = {cck, ckc, kcc}**
  - **P(A) = 2/8 = 1/4, P(B) = 4/8 = 1/2, P(C) = 7/8, P(AB) = 1/8, P(AC) = 1/8, P(BC) = 3/8**
  - **P(A)P(B) = 1/4 * 1/2 = 1/8 = P(AB). Portanto, A e B são independentes**
  - **P(A)P(C) = 1/4 * 7/8 = 7/32 != P(AC). Portanto, A e C não são independentes**
  - **P(B)P(C) = 1/2 * 7/8 = 7/16 != P(BC). Portanto, B e C não são independentes**

Variável Aleatória

- Uma v.a. **X** é uma função que associa um número real $X(\zeta)$ a cada resultado ζ no espaço amostral S de um experimento aleatório
- Podemos ver $X(\cdot)$ como uma função que mapeia os pontos amostrais $\zeta_1, \zeta_2, \ldots, \zeta_m$ em números reais $x_1, x_2, \ldots, x_n$



Variáveis Aleatórias

- Ex. 15 (exemplo 2.1 da apostila Ynoguti):

Especifique o espaço amostral de um experimento que consiste em jogar uma moeda 3 vezes.

Solução. O espaço amostral para este experimento é

$$S = \{CCC, CCK, CKC, CKK, KCC, KCK, KKC, KKK\},$$

onde C corresponde a “cara"e K corresponde a “coroa".

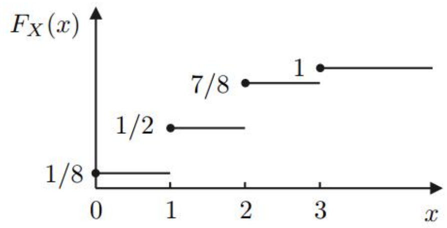
Seja X o número de caras em três jogadas da moeda. X associa a cada resultado ζ em S um número do conjunto $S_X = 0, 1, 2, 3$. A tabela abaixo lista os oito resultados de S e os valores de X correspondentes.

ζ	CCC	CCK	CKC	KCC	CKK	KCK	KKC	KKK
$X(\zeta)$	3	2	2	2	1	1	1	0

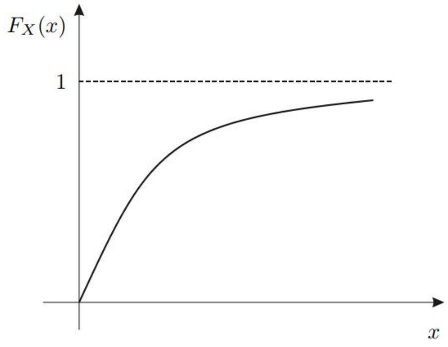
X é então uma v.a. que toma valores no conjunto $S_X = 0, 1, 2, 3$.

Tipos de Variáveis Aleatórias

- Discretas:
    - Tomam valores de um conjunto finito
    - Aparecem geralmente em aplicações que envolvem contagens
    - Gráficos que as representam geralmente são degraus
    - São comumente representadas por **CDFs** e **PMFs**
    - Cálculos geralmente envolvem somatórios



- Contínuas:
    - Tomam valores de um conjunto infinito
    - Gráficos que as representam geralmente são curvas contínuas
    - São comumente representadas por **CDFs** e **PDFs**
    - Cálculos geralmente envolvem integrais



Função de Distribuição Cumulativa

- Conhecida como CDF

- Válida para v.a. discreta ou contínua
- Mapeia um valor para uma probabilidade cujo resultado é menor ou igual a x_a

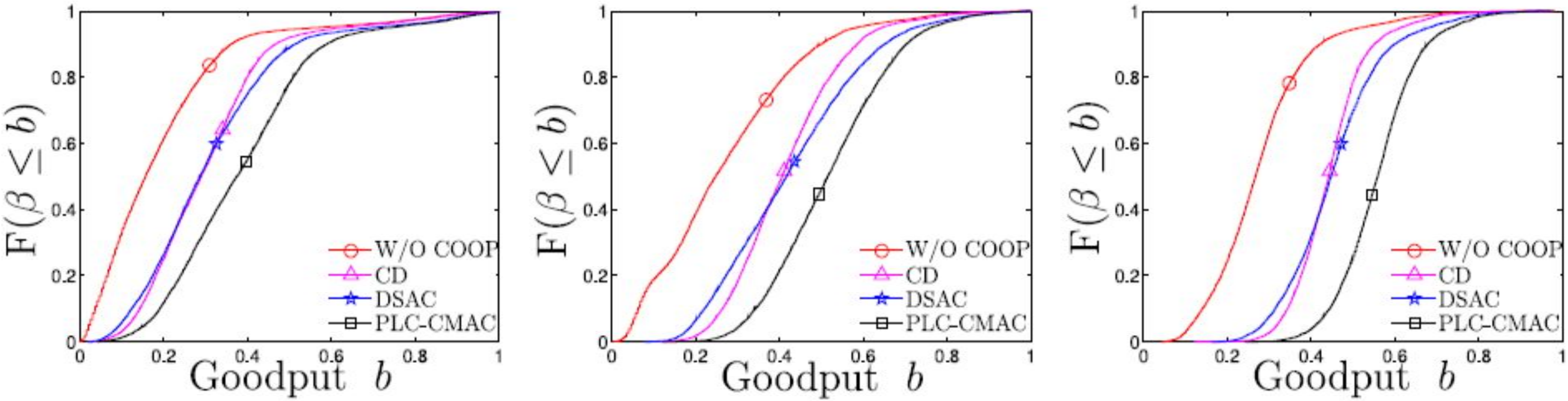
$$F_{x_a}(x_a) = P(X \leq x_a)$$

- Curvas crescentes

- Se $P(X \leq x_a) = a$, esse resultado é chamado de a - percentil ou quantil

Função de Distribuição Cumulativa

- Ex. 16:



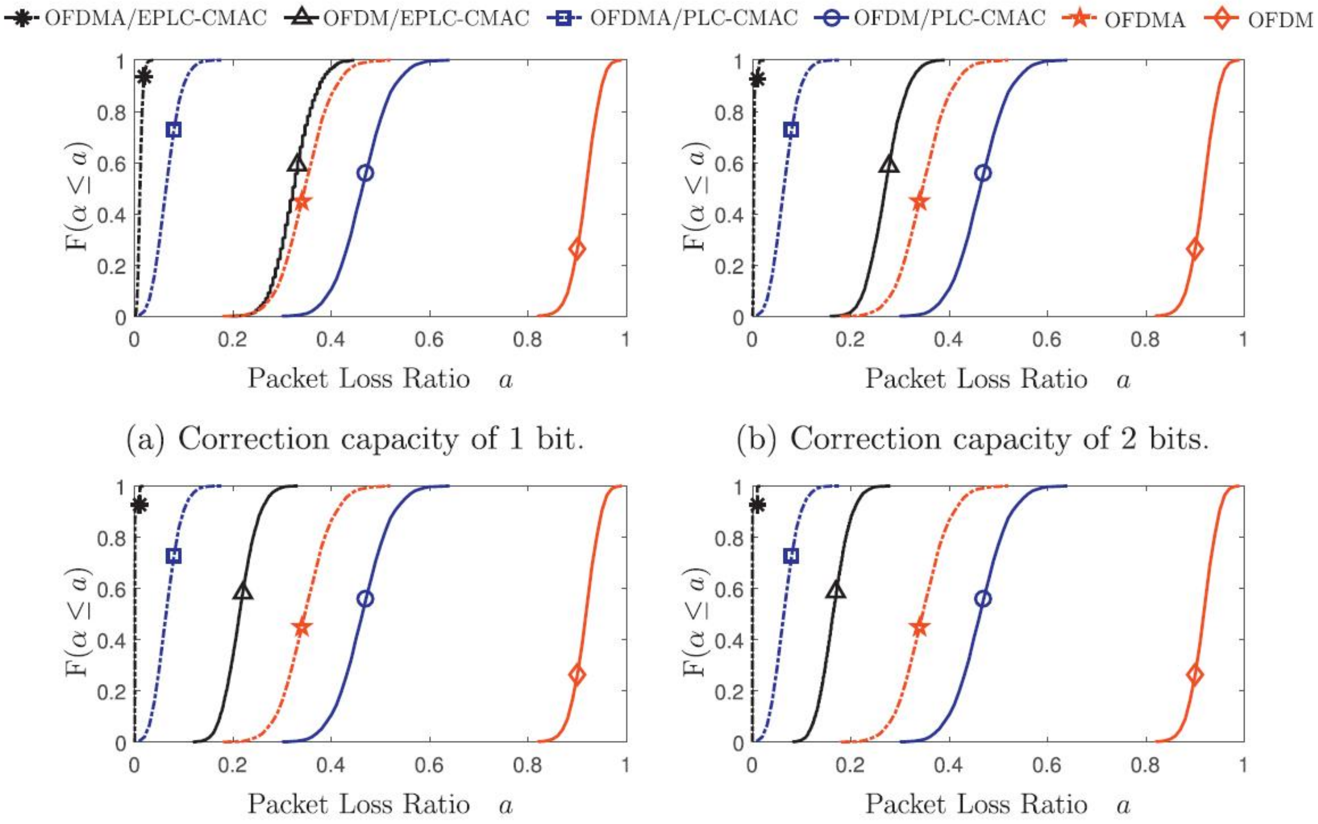
(a)Cases #1 and #3 in 1.7-30 MHz.

(b)Cases #1 and #3 in 1.7-50 MHz.

(c)Cases #1 and #3 in 1.7-100 MHz.

Função de Distribuição Cumulativa

- Ex. 17:

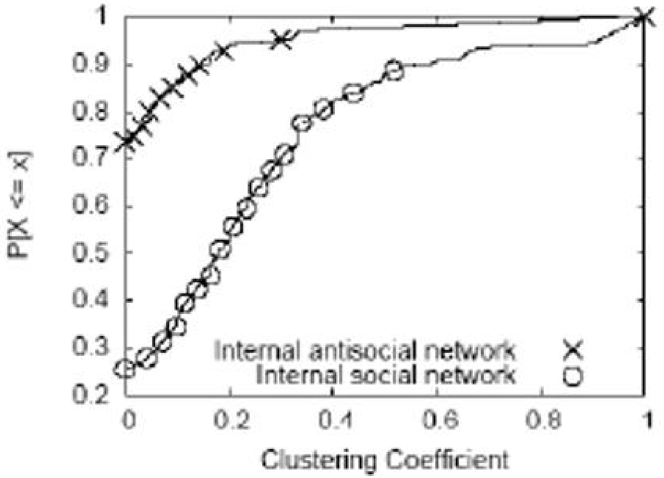


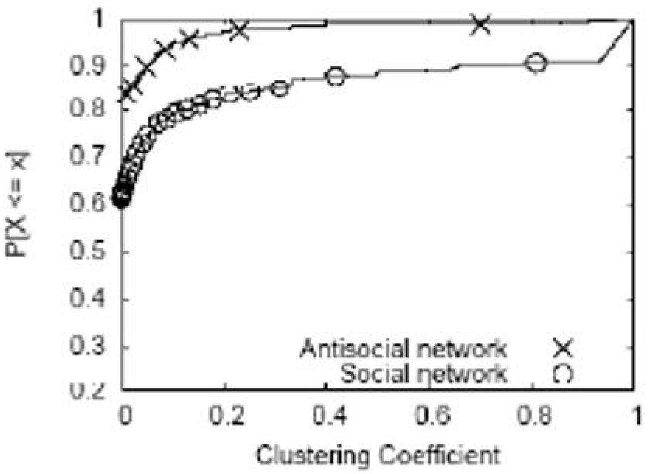
(a) Correction capacity of 1 bit.

(b) Correction capacity of 2 bits.

Função de Distribuição Cumulativa

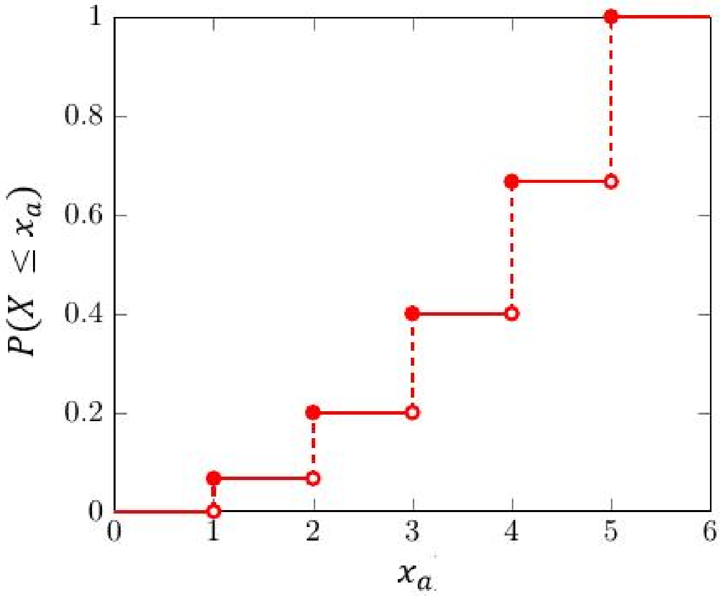
- Ex. 18: *Clustering coefficient* é a probabilidade na qual dois vizinhos de um dado nó de uma rede são, também, vizinhos entre si. Vamos encontrar alguns percentis e entender como comparar as curvas. Essa v.a. é discreta ou contínua?





Função de Distribuição Cumulativa

- Exemplo de uma CDF de v.a. discreta



Função Densidade de Probabilidade

- Conhecida como PDF
- Válida para v.a. contínua
- Derivada da CDF contínua

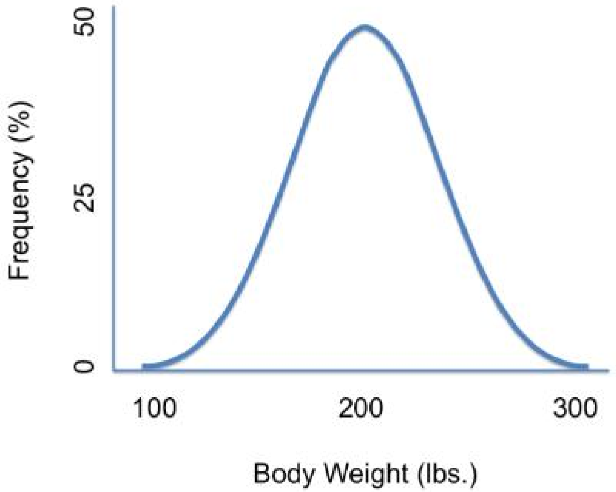
$$f(x) = \frac{dF(x)}{dx}, \; f(x) \geq 0$$

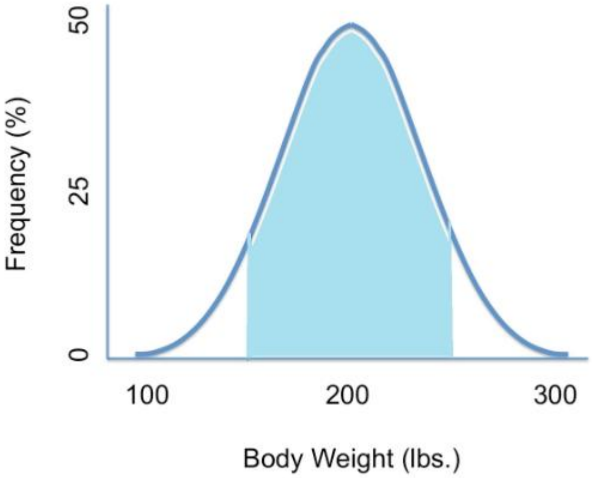
- Útil para determinar intervalos de probabilidades

$$P(x_1 < x \leq x_2) = F(x_2) - F(x_1)$$
$$= \int_{x_1}^{x_2} f(x)dx$$

Função Densidade de Probabilidade

- Ex. 19: Como faríamos para calcular a probabilidade de pessoas pesarem entre 150 lbs e 250 lbs? (curiosidade: 1 lbs ~ 0,45 kg)
  - Primeiro, temos que colher a amostra, obter a curva da PDF ou derivar a curva da CDF
  - A probabilidade de um intervalo é dada pela área desse intervalo sob a curva (integral)



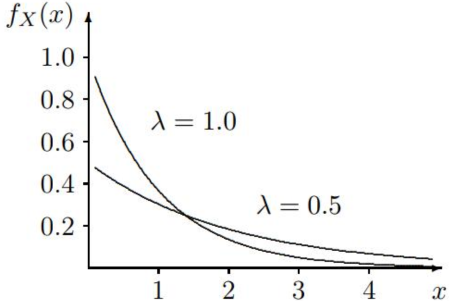


Função Densidade de Probabilidade

- Ex. 20: Vejamos a PDF e a CDF da distribuição exponencial

PDF

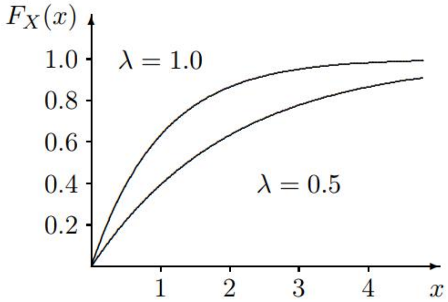
$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \text{ e } \lambda > 0 \\ 0 & \text{caso contrário} \end{cases}$$



$$F_X(x) = \int_0^x \lambda e^{-\lambda y} dy = \lambda \frac{e^{-\lambda y}}{-\lambda}\bigg|_0^x = -1(e^{-\lambda x} - e^0) \Rightarrow$$

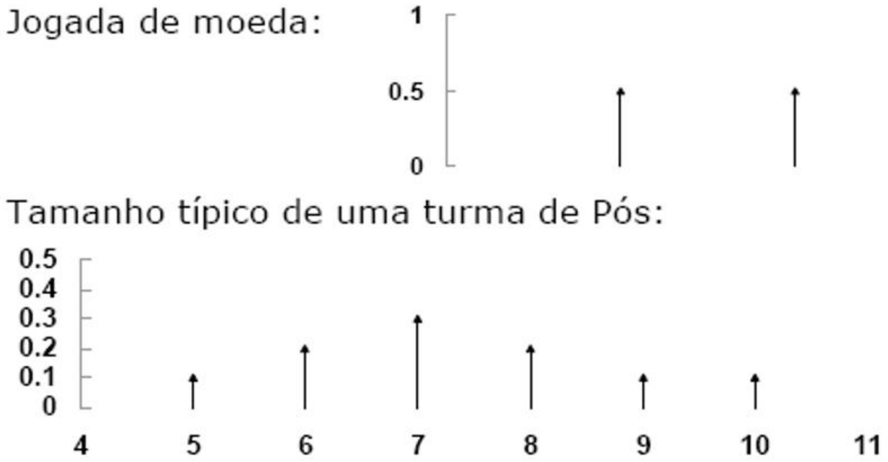
CDF

$$F_X(x) = \begin{cases} 1 - e^{-\lambda x} & x \geq 0, \lambda > 0 \\ 0 & \text{caso contrário} \end{cases}$$



Função Massa de Probabilidade

- Conhecida como PMF
- Válida para v.a. discreta
- Associa cada valor de uma v.a. (eixo x) a uma probabilidade (eixo y)
- Geralmente representada por um gráfico de barras ou similares



AAG01: Tarefa em Dupla (parte 1)

- Implementar funções que plotam CDFs e PMFs a partir de **amostras discretas e contínuas** passadas como parâmetro de entrada
- Regras:
  1. A tarefa deverá ser realizada **obrigatoriamente no Jupyter** com os uso de:
     a. Blocos de Markdowns com comandos de LaTex
     b. Blocos de código
     c. Blocos com resultados gráficos no ambiente do Jupyter
     d. Gráficos gerados devem ser interpretados e explicados em Markdown
  2. Não são permitidos comandos prontos além dos comandos básicos do python
  3. Parâmetros de entrada devem ser cuidadosamente escolhidos e especificados para cada tipo de curva (e.g., tempos de processamento, pontuação esportiva, resultados de artigos)
  4. Os formatos de entrega devem ser .pdf e .ipynb (código fonte+markdowns)
  5. Caso usem arquivos de amostras, favor anexá-los também
  6. Data de entrega: começo da próxima aula

AAG01: Tarefa em Dupla (parte 2)

- Repetir a parte 1 com os mesmos parâmetros de entrada, mas usando funções prontas de bibliotecas python para curvas empíricas
- **ATENÇÃO: Diferente da parte 1, plotar também a PDF (além da CDF e PMF) com as funções prontas das bibliotecas**