Introdução à Probabilidade

Silvia Shimakura

silvia.shimakura@ufpr.br

Probabilidade

- **O que é probabilidade?**

  Medida que quantifica a incerteza frente a um acontecimento futuro

- **Como quantificar incerteza?**

  - **Definição Clássica**
  - **Definição Frequentista**

Problema 1

- Experimento 1: Lançamento de um dado balanceado

- Espaço amostral: conjunto de todos os resultados possíveis

$$E=\{1, 2, 3, 4, 5, 6\}$$

- Evento A: face par

$$A=\{2, 4, 6\}$$

Tipos especiais de eventos

- Evento complementar de A: elementos de E que não estão em A

E={1, 2, 3, 4, 5, 6}   A={2, 4, 6}

$$\bar{A} = \{1, 3, 5\}$$

- Evento interseção: elementos estão em A e em B

A={2, 4, 6}   B={1, 3} $A \cap B = \emptyset$

- Evento união: elementos que estão em A ou B

$A \cup B = \{1, 2, 3, 4, 6\}$

Tipos especiais de eventos

- Eventos mutuamente exclusivos: não existem elementos comuns em A e B

$$A=\{2, 4, 6\} \quad B=\{1, 3\} \quad A \cap B = \emptyset$$

Definição Clássica de Probabilidade

- Se os elementos de E forem equiprováveis a probabilidade de um evento A ocorrer:

$$P(A) = \frac{\text{número de elementos em } A}{\text{número de elementos em } E}$$

- Experimento 1: E={1,2,3,4,5,6} → 6 elementos (equiprováveis)

- Evento A: A={2,4,6} → P(A)=3/6

Definição Clássica de Probabilidade

- Evento complementar de A:

$$\bar{A} = \{1, 3, 5\} \Rightarrow P(\bar{A}) = 3/6$$

- Evento interseção:

$$A \cap B = \emptyset \rightarrow P(A \cap B) = 0$$

- Evento união:

$$A \cup B = \{1, 2, 3, 4, 6\} \Rightarrow P(A \cup B) = 5/6$$

Propriedades de Probabilidade

- $0 \leq P(A) \leq 1$, para qualquer evento A
- $P(E)=1$, em que E é o espaço amostral
- $P(\bar{A}) = 1 - P(A)$
- Para dois eventos A e B quaisquer, a probabilidade de que A ou B ocorra:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

- Se A e B são mutuamente exclusivos, a probabilidade de que A ou B ocorra é a soma das probabilidades.

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

Probabilidade condicional

- $P(A|B)$=P(A ocorrer dado que B ocorreu)
- Para A e B quaisquer:

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

- Para A e B independentes: $P(A|B)=P(A)$

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

Exemplo: Lançamento de dois dados não viciados

- Espaço amostral:

  E={(1,1), (1,2), …, (1,6), (2,1), (2,2), …, (2,6), …, (6,6)}

  36 elementos → equiprováveis com prob 1/36 cada

- Evento A: soma dos dados é 8

- A=?

- P(A)=?

Exemplo: Lançamento de dois dados não viciados

- Espaço amostral: 36 elementos com prob 1/36 cada
- Evento A: soma dos dados é 8
- A={(2, 6), (6, 2), (3, 5), (5, 3), (4, 4)}
- P(A)=?

Exemplo: Lançamento de dois dados não viciados

- Espaço amostral: 36 elementos com prob 1/36 cada
- Evento A: soma dos dados é 8
- A={(2, 6), (6, 2), (3, 5), (5, 3), (4, 4)}
- P(A)=5/36

Exemplo: Lançamento de dois dados não viciados

- Espaço amostral: 36 elementos com prob 1/36 cada
- Evento A: soma dos dados é 8
- A={(2, 6), (6, 2), (3, 5), (5, 3), (4, 4)}
- P(A)=5/36

Se soubermos que o resultado no primeiro dado é 3, qual será a probabilidade da soma dos dois dados ser 8?

Exemplo (cont.)

- Evento A:

  $A=\{(2, 6), (6, 2), (3, 5), (5, 3), (4, 4)\} \rightarrow P(A)=5/36$

- Evento B: primeiro dado é 3

  $B=\{(3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6)\} \rightarrow P(B)=6/36$

- Evento interseção: **$A \cap B=\{(3,5)\} \rightarrow P(A \cap B)=1/36$**

- $P(A|B)=?$
- A e B são independentes?
- A e B são mutuamente exclusivos?

Exemplo (cont.)

- Evento A:

  $A=\{(2, 6), (6, 2), (3, 5), (5, 3), (4, 4)\} \rightarrow P(A)=5/36$

- Evento B: primeiro dado é 3

  $B=\{(3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6)\} \rightarrow P(B)=6/36$

- Evento interseção: **$A \cap B=\{(3,5)\} \rightarrow P(A \cap B)=1/36$**

- $P(A|B)=$ **$P(A \cap B)/P(B)=(1/36)/(6/36)=1/6$**
- A e B são independentes?
- A e B são mutuamente exclusivos?

Problema 2

- Experimento 3: Lançamento de uma moeda

- Espaço amostral: E={Cara, Coroa}

- Evento C: Cara

  C={Cara}

- Elementos de E são equiprováveis?

- P(C)=?

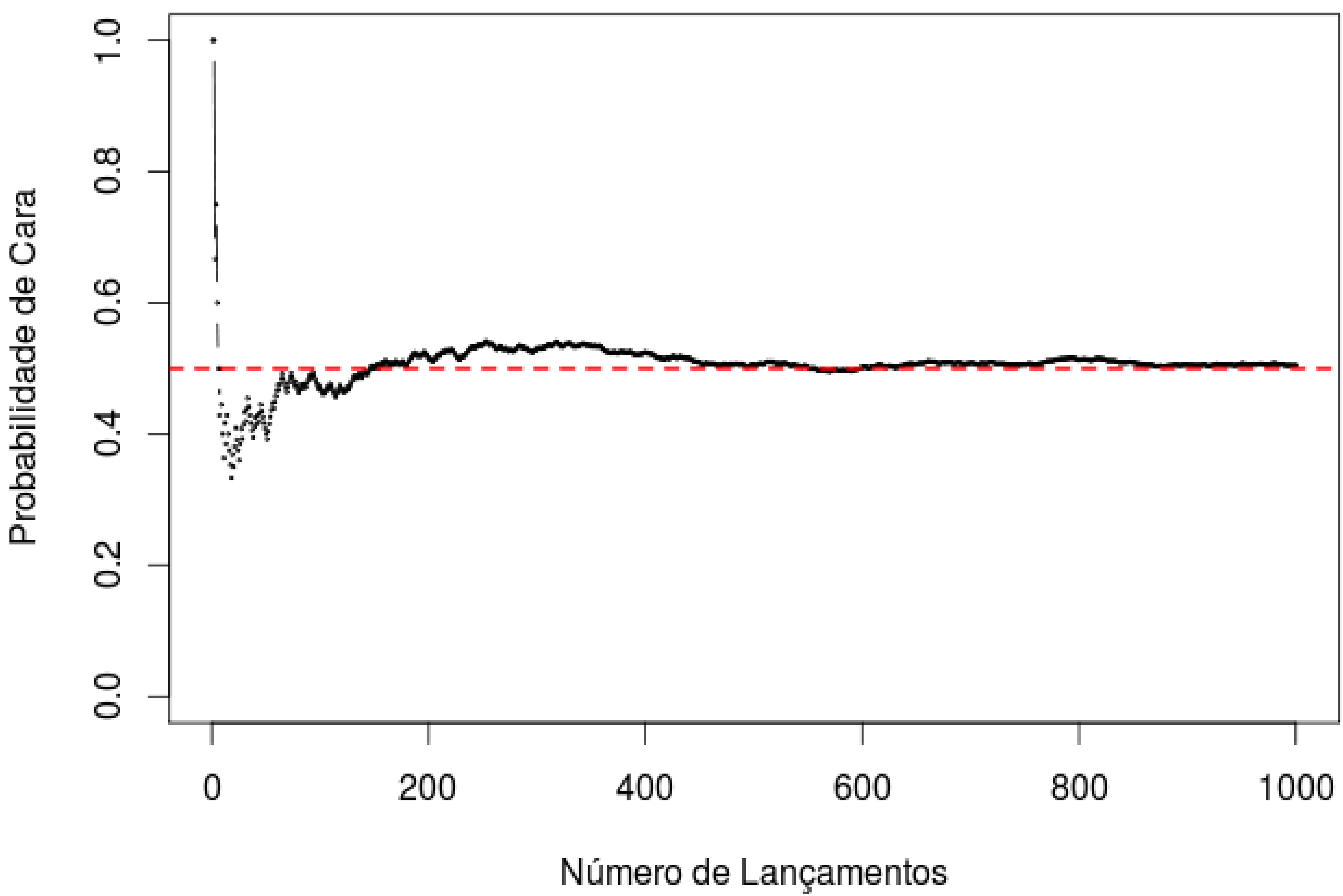
Visão frequentista de probabilidade

- Probabilidade: frequência relativa de ocorrência do evento para um grande número de sorteios

Frequência relativa

- C: Cara O: Coroa

Resultado	C	C	C	O	C	O	O	O	O	O	O	C
Frequência acumulada de Caras	1	2	3	3	4	4	4	4	4	4	4	5
Número de lançamentos	1	2	3	4	5	6	7	8	9	10	11	12
Freq. relativa de caras	1/1	2/2	3/3	3/4	4/5	4/6	4/7	4/8	4/9	4/10	4/11	5/12
%	100	100	100	75	80	67	57	50	44	40	36	42



Probabilidade de Cara
0.0
0.2
0.4
0.6
0.8
1.0
0
200
400
600
800
1000
Número de Lançamentos

Exemplo: Distribuição de peso e pressão arterial

Pressão arterial	Peso			
	Excesso	Normal	Deficiente	Total
Elevada	0,10	0,08	0,02	0,2
Normal	0,15	0,45	0,20	0,8
Total	0,25	0,53	0,22	1

- Probabilidade de uma pessoa escolhida ao acaso ter pressão elevada?

- Probabilidade de uma pessoa escolhida ao acaso ter pressão elevada e excesso de peso?

- Sabendo que a pessoa tem excesso de peso, qual a probabilidade de uma pessoa escolhida ao acaso ter pressão elevada?

Exemplo: Distribuição de peso e pressão arterial (cont.)

- Peso em excesso e pressão arterial normal são eventos mutuamente exclusivos?

- Pressão arterial e peso são independentes?

Pressão arterial	Peso			
	Excesso	Normal	Deficiente	Total
Elevada	0,10	0,08	0,02	0,2
Normal	0,15	0,45	0,20	0,8
Total	0,25	0,53	0,22	1

Exemplo: Qualidade de teste diagnóstico

Num estudo sobre a qualidade do teste ergométrico, Wrinner et al. (1979) compararam os resultados obtidos entre indivíduos com (D) e sem ($\bar{D}$) doença coronariana.

- T_p :teste positivo T_n :teste negativo

- T_p : mais de 1mm de depressão ou elevação do segmento ST, por pelo menos 0,08s, em comparação com paciente em repouso.
- **D** e $\bar{D}$: angiografia (teste padrão ouro).

Exemplo: Qualidade de teste diagnóstico

Doença coronariana	Teste Ergométrico		
	T+	T-	Total
D+	815	208	1023
D-	115	327	442
Total	930	535	1465

Temos interesse em conhecer as probabilidades de acerto do teste:

- Probabilidade de teste positivo num paciente doente
- Probabilidade de teste negativo num paciente não doente

Exemplo: Qualidade de teste diagnóstico

Doença coronariana	Teste Ergométrico		
	T+	T-	Total
D+	815	208	1023
D-	115	327	442
Total	930	535	1465

Temos interesse em conhecer as probabilidades de acerto do teste:

- Probabilidade de teste positivo num paciente doente
  $P(T_p|D)$ = sensibilidade = s
- Probabilidade de teste negativo num paciente não doente
  $P(T_n|\bar{D})$ = especificidade = e

Exemplo: Qualidade de teste diagnóstico

Doença coronariana	Teste Ergométrico		
	T+	T-	Total
D+	815	208	1023
D-	115	327	442
Total	930	535	1465

Temos interesse em conhecer as probabilidades de acerto do teste:

- Probabilidade de teste positivo num paciente doente
  s = 815/1023 = 0,80
- Probabilidade de teste negativo num paciente não doente
  e = 327/442 = 0,74

Exemplo: Qualidade de teste diagnóstico (cont.)

Do ponto de vista do médico que precisa fechar o diagnóstico além da sensibilidade e especificidade do teste é importante conhecer as seguintes probabilidades:

- Probabilidade de que uma pessoa com resultado positivo no teste realmente tenha a doença
- Probabilidade de que uma pessoa com resultado negativo no teste realmente não tenha a doença

Exemplo: Qualidade de teste diagnóstico (cont.)

Do ponto de vista do médico que precisa fechar o diagnóstico além da sensibilidade e especificidade do teste é importante conhecer as seguintes probabilidades:

- Probabilidade de que uma pessoa com resultado positivo no teste realmente tenha a doença

$$P\left(D \middle| T_p\right) = VPP$$

- Probabilidade de que uma pessoa com resultado negativo no teste realmente não tenha a doença

$$P\left(\bar{D} \middle| T_n\right) = VPN$$

Exemplo: Qualidade de teste diagnóstico

Doença coronariana	Teste Ergométrico		
	T+	T-	Total
D+	815	208	1023
D-	115	327	442
Total	930	535	1465

Podemos obter VPP e VPN direto da tabela acima?

$$VPP = P\left(D \middle| T_p\right) =$$

$$VPN = P\left(\bar{D} \middle| T_n\right) =$$

Teorema de Bayes

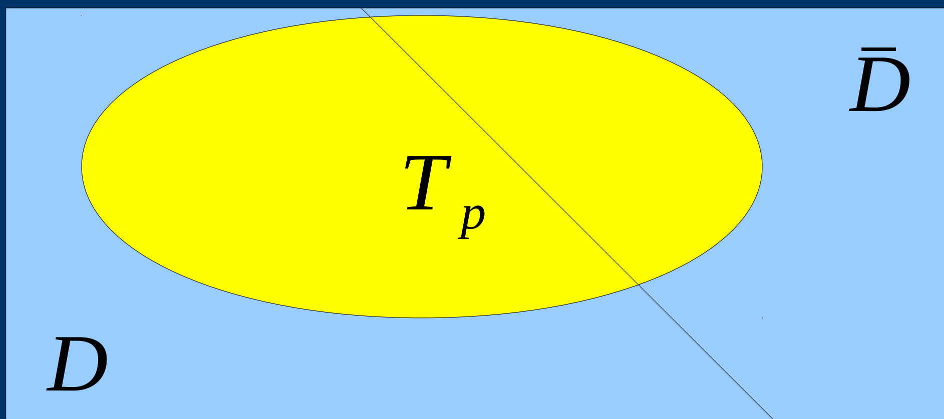
- D e $\bar{D}$ são eventos mutuamente exclusivos e exaustivos:

$$P\left(D \cup \bar{D}\right) = P\left(D\right) + P\left(\bar{D}\right) = 1$$

- Pelo Teorema de Bayes:

$$\boxed{P\left(D|T_p\right) = \frac{P\left(D\right) P\left(T_p|D\right)}{P\left(D\right) P\left(T_p|D\right) + P\left(\bar{D}\right) P\left(T_p|\bar{D}\right)}}$$

Exemplo: Qualidade de teste diagnóstico

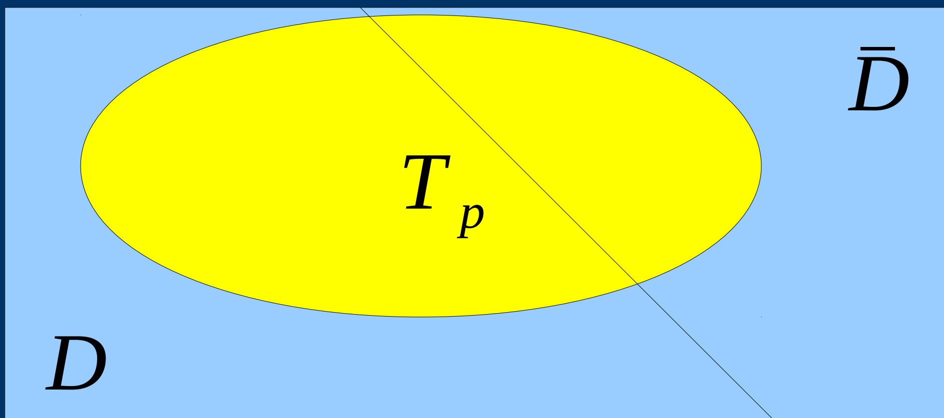


- $D \cap \bar{D} = \emptyset$
- $D \cup \bar{D} = E$
- Conhecidos: P(D)

$$P(T_p | D) = s \quad P(T_n | \bar{D}) = e$$

$$VPP = P(D | T_p) = \frac{P(D \cap T_p)}{P(T_p)}$$

Exemplo: Qualidade de teste diagnóstico



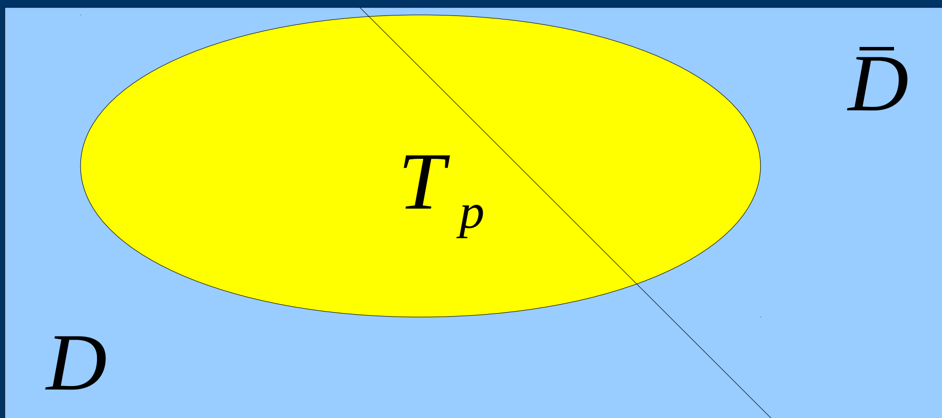
- $D \cap \bar{D} = \emptyset$
- $D \cup \bar{D} = E$
- Conhecidos: P(D)

$$P(T_p|D) = s \quad P(T_n|\bar{D}) = e$$

$$P(T_p|D) = \frac{P(T_p \cap D)}{P(D)} \Rightarrow P(T_p \cap D) = P(T_p|D)\,P(D)$$

$$VPP = P(D|T_p) = \frac{P(D \cap T_p)}{P(T_p)}$$

Exemplo: Qualidade de teste diagnóstico



- $D \cap \bar{D} = \emptyset$
- $D \cup \bar{D} = E$
- Conhecidos: P(D)

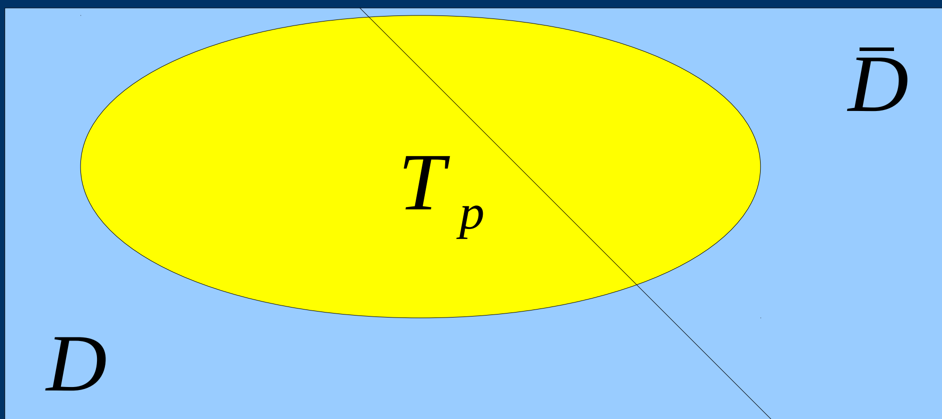
$$P(T_p | D) = s \quad P(T_n | \bar{D}) = e$$

$$P(T_p | D) = \frac{P(T_p \cap D)}{P(D)} \Rightarrow P(T_p \cap D) = P(T_p | D) P(D)$$

$$= s \times p$$

$$VPP = P(D | T_p) = \frac{P(D \cap T_p)}{P(T_p)}$$

Exemplo: Qualidade de teste diagnóstico



- $D \cap \bar{D} = \emptyset$
- $D \cup \bar{D} = E$
- Conhecidos: P(D)

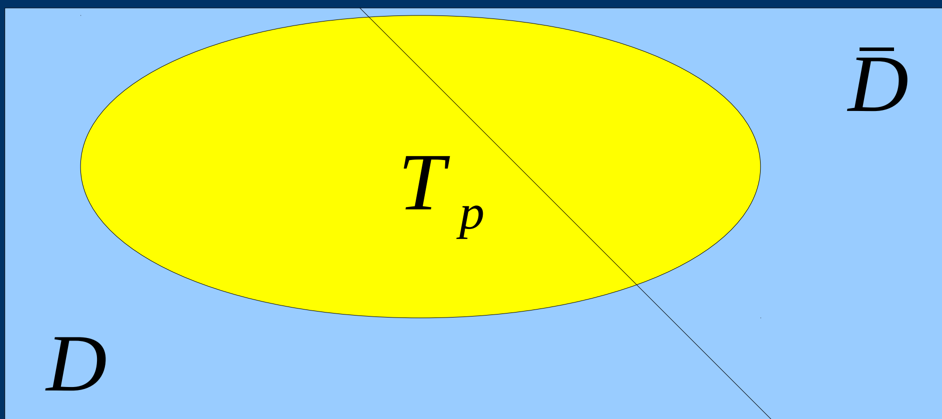
$$P(T_p | D) = s \quad P(T_n | \bar{D}) = e$$

$$VPP = P(D | T_p) = \frac{P(D \cap T_p)}{P(T_p)}$$

$$P(T_p | D) = \frac{P(T_p \cap D)}{P(D)} \Rightarrow P(T_p \cap D) = P(T_p | D) P(D)$$
$$= s \times p$$

$$P(T_p) = P(T_p \cap D) + P(T_p \cap \bar{D})$$

Exemplo: Qualidade de teste diagnóstico



- $D \cap \bar{D} = \emptyset$
- $D \cup \bar{D} = E$
- Conhecidos: P(D)

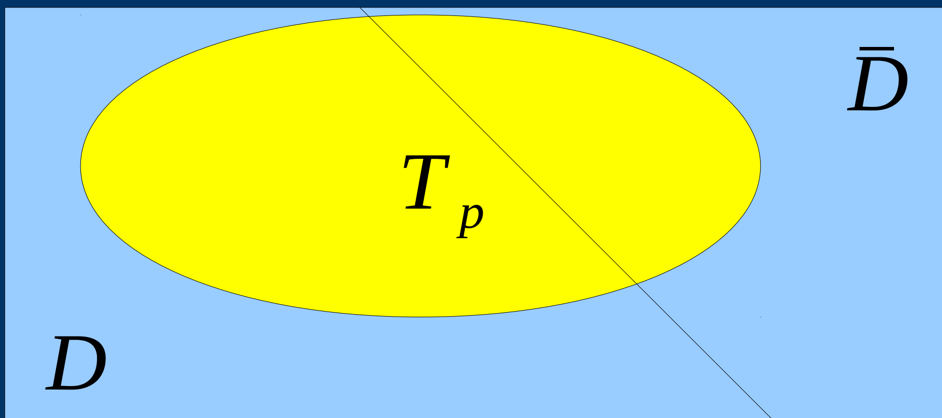
$$P(T_p|D) = s \quad P(T_n|\bar{D}) = e$$

$$VPP = P(D|T_p) = \frac{P(D \cap T_p)}{P(T_p)}$$

$$P(T_p|D) = \frac{P(T_p \cap D)}{P(D)} \Rightarrow P(T_p \cap D) = P(T_p|D)\,P(D)$$
$$= s \times p$$

$$P(T_p) = P(T_p \cap D) + P(T_p \cap \bar{D})$$
$$= P(T_p|D)\,P(D) + P(T_p|\bar{D})\,P(\bar{D})$$

Exemplo: Qualidade de teste diagnóstico



- $D \cap \bar{D} = \emptyset$
- $D \cup \bar{D} = E$
- Conhecidos: P(D)

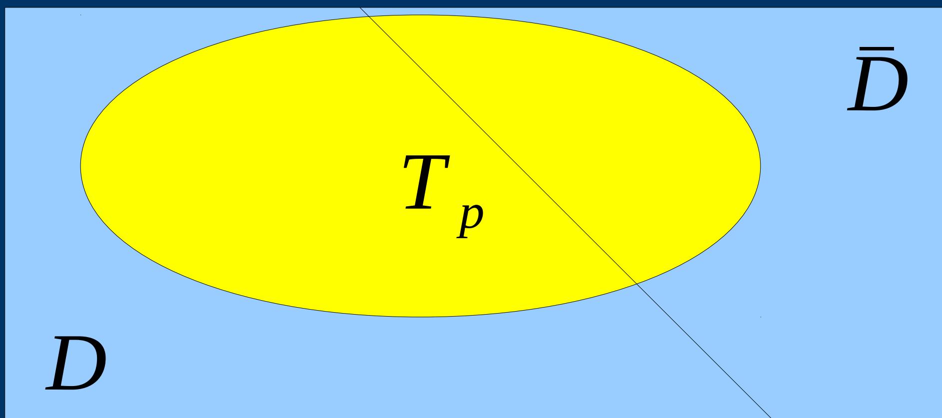
$$P(T_p | D) = s \qquad P(T_n | \bar{D}) = e$$

$$VPP = P(D | T_p) = \frac{P(D \cap T_p)}{P(T_p)}$$

$$P(T_p | D) = \frac{P(T_p \cap D)}{P(D)} \Rightarrow P(T_p \cap D) = P(T_p | D) P(D)$$
$$= s \times p$$

$$\begin{aligned} P(T_p) &= P(T_p \cap D) + P(T_p \cap \bar{D}) \\ &= P(T_p | D) P(D) + P(T_p | \bar{D}) P(\bar{D}) \\ &= s \times p + (1 - e) \times (1 - p) \end{aligned}$$

Exemplo: Qualidade de teste diagnóstico



- $D \cap \bar{D} = \emptyset$
- $D \cup \bar{D} = E$
- Conhecidos: P(D)

$$P\left(T_p \middle| D\right) = s \qquad P\left(T_n \middle| \bar{D}\right) = e$$

$$VPP = P\left(D \middle| T_p\right) = \frac{P\left(D \cap T_p\right)}{P\left(T_p\right)} = \frac{s\ p}{s\ p + (1-e)\,(1-p)}$$

Exemplo: Qualidade de teste diagnóstico

Doença coronariana	Teste Ergométrico		
	T+	T-	Total
D+	815	208	1023
D-	115	327	442
Total	930	535	1465

Podemos obter VPP e VPN direto da tabela acima?

$$VPP = P\left(D \middle| T_p\right) =$$

$$VPN = P\left(\bar{D} \middle| T_n\right) =$$

$$p = P\left(D+\right) = \frac{1023}{1465} = 0{,}70\,!\,!\,!$$

Exemplo: Qualidade de teste diagnóstico – tabela fictícia

Doença coronariana	Teste Ergométrico			
	T+	T-	Total	
D+	469	117	586	s=0,8
D-	229	650	879	e=0,74
Total	698	767	1465	p=0,4

Podemos obter VPP e VPN direto da tabela acima?

$$VPP = P\left(D \middle| T_p\right) = 469/698 = 0,67$$

$$VPN = P\left(\bar{D} \middle| T_n\right) = 650/767 = 0,85$$