Aula sobre Probabilidade

É difícil fazer ciência de dados sem algum tipo de entendimento de probabilidade e sua matemática.

Para nossos propósitos, você deve pensar em probabilidade como ***uma forma de quantificar a incerteza associada a eventos escolhidos de um universo de eventos***. Em vez de obter informações técnicas sobre o significado desses termos, pense no ato de jogar um dado. O **universo** consiste em todos os resultados possíveis. E qualquer subconjunto desses resultados é um **evento**; por exemplo, **"parar no número um"** ou **"parar em um número par"**.

No que diz respeito a **notação**, escrevemos **P(E)** para significar **"a probabilidade do evento E."**. Usamos a teoria das probabilidades para construir e avaliar modelos. A verdade é que nós vamos usar a teoria da probabilidade em todo o lugar.

Também podemos definir probabilidade nos seguintes termos:
A repetição de um experimento, mesmo sob condições semelhantes, poderá levar a resultados (eventos) diferentes. Mas se o experimento for repetido um número "suficientemente grande" de vezes haverá uma regularidade nestes resultados que permitirá calcular a sua probabilidade de ocorrência

Definição Clássica

Se um experimento aleatório puder resultar em n diferentes e igualmente prováveis resultados, e nEi destes resultados referem-se ao evento Ei, então a probabilidade do evento Ei ocorrer será:

$$P(Ei) = \frac{n_{Ei}}{n}$$

O problema reside em calcular o número total de resultados possíveis e o número de resultados associados ao evento de interesse. Isso pode ser feito usando técnicas de análise combinatória ou por considerações teóricas ("bom senso").

Exemplo

Seja o seguinte Experimento Aleatório: lançamento de um dado não viciado e observação da face voltada para cima. Calcular as probabilidades de ocorrência dos seguintes eventos:

- Face 1
- Face par
- Face menor ou igual a 2

O Espaço Amostral deste experimento será: s = {1, 2, 3, 4, 5, 6}. Sendo assim há um total de 6 resultados possíveis, resultando em n = 6. Basta então definir quantos resultados estão associados a cada evento para que seja possível calcular suas probabilidades pela definição clássica.

O evento “face 1” tem apenas um resultado associado: { 1 }.
Então nEi = 1, e a probabilidade de ocorrer a face 1 será:
P(Ei) = nEi/n = 1/6

O evento “face par” tem três resultados associados: {2, 4, 6}.
Então nEi = 3, e a probabilidade de ocorrer face par será:
P(Ei) = nEi/n = 3/6

O evento “face menor ou igual a 2” tem dois resultados associados: {1, 2}.
Então nEi = 2, e a probabilidade de ocorrência de face menor ou igual a 2 será:
P(Ei) = nEi/n = 2/6 = 1/3

Experimento Aleatório

Experimento Aleatório é um processo de obtenção de um resultado ou medida que apresenta as seguintes características:

- Não se pode afirmar, ANTES de realizar o experimento, qual será o resultado de uma realização, mas é possível determinar o conjunto de resultados possíveis.
- Quando é realizado um grande número de vezes (replicado) apresentará uma REGULARIDADE que permitirá construir um modelo probabilístico para analisar o experimento.

São experimentos aleatórios: lançamento de um dado não viciado e observação da face voltada para cima; cruzar espécimes de ervilha e observar os fenótipos dos descendentes.

Espaço Amostral

Espaço Amostral é o conjunto de TODOS os resultados possíveis de um experimento aleatório. “PARA CADA EXPERIMENTO ALEATÓRIO HAVERÁ UM ESPAÇO AMOSTRAL ÚNICO ASSOCIADO A ELE “.

Exemplo

Definir os espaços amostrais dos experimentos abaixo:

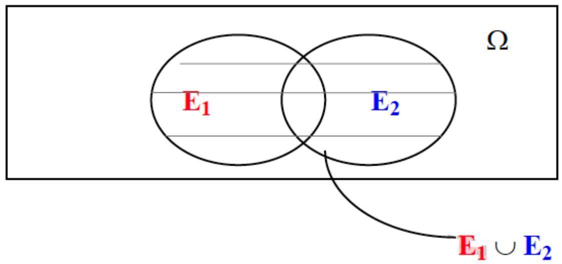
1. Lançar uma moeda e observar a face voltada para cima.
   Os dois únicos resultados possíveis são cara e coroa: s = {Cara, Coroa}.
2. Altura de homens adultos.
   De uma forma genérica poderíamos definir indivíduo adulto como tendo mais de 1,40m de altura: s = {Altura > 1,40m}
3. Observar o número de meninos em famílias de 5 filhos.
   Cada família pode ter no mínimo 0 e no máximo 5 meninos: s = {0, 1, 2, 3, 4, 5}

Evento

Evento é QUALQUER subconjunto do espaço amostral. Um evento pode conter um ou mais resultados, se pelo menos um dos resultados ocorrer o evento ocorre! Geralmente há interesse em calcular a probabilidade de que um determinado evento venha a ocorrer, e este evento pode ser definido de forma verbal, precisando ser “traduzido” para as definições da Teoria de Conjuntos, que veremos a seguir.

Seja o Experimento Aleatório lançamento de um dado não viciado e observação da face voltada para cima: o seu espaço amostral será S = {1, 2, 3, 4, 5, 6}. Definindo três eventos: E1 = {2, 4, 6}, E2 = {3, 4, 5, 6} e E3 = {1, 3} serão apresentadas as definições de Evento União, Evento Intersecção, Eventos Mutuamente Exclusivos e Evento Complementar.

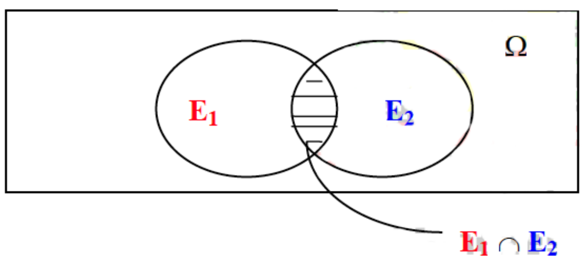
Evento União de E1 com E2 (E1 U E2): evento que ocorre se E1 OU E2 OU ambos ocorrem.



$E_1 \cup E_2 = \{2, 3, 4, 5, 6\}$

Composto por todos os resultados que pertencem a um **ou** ao outro, **ou** a ambos.

Evento Intersecção de E1 com E2 (E1 n E2) : evento que ocorre se E1 E E2 ocorrem SIMULTANEAMENTE.



SIMULTANEAMENTE

$E_1 \cap E_2 = \{4, 6\}$

Composto por todos os resultados que pertencem a ambos.

Eventos Mutuamente Exclusivos (M.E.): são eventos que NÃO PODEM OCORRER SIMULTANEAMENTE, não apresentando elementos em comum (sua intersecção é o conjunto vazio).
Dentre os três eventos definidos acima, observamos que os eventos E1 e E3 não têm elementos em comum:
E3 = {1, 3} E1 = {2, 4, 6} E1 n E3 = vazio => E1 e E3 são mutuamente exclusivos

Evento Complementar de um evento qualquer é formado por todos os resultados do espaço amostral que NÃO PERTENCEM ao evento. A união de um evento e seu complementar formará o próprio Espaço Amostral, e a intersecção de um evento e seu complementar é o conjunto vazio.

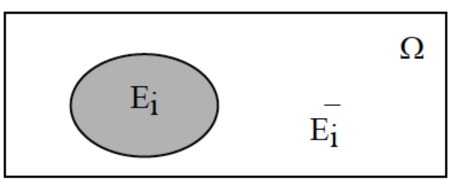


Figura 3 - Evento complementar

$E_i \cup \bar{E}_i = \Omega \qquad E_i \cap \bar{E}_i = \varnothing$

$E_1 = \{2, 4, 6\} \qquad \bar{E}_1 = \{1, 3, 5\}$

$E_2 = \{3, 4, 5, 6\} \qquad \bar{E}_2 = \{1, 2\}$

Dependência e Independência

Grosso modo, dizemos que dois eventos **E** e **F** são **dependentes** se saber algo sobre se o evento **E** acontece nos dá informações sobre a ocorrência do evento **F** (e vice-versa). Caso contrário, eles são **independentes**.

Por exemplo, considere um cenário com uma sacola na qual há 5 bolas sendo 3 verdes e 2 amarelas e dois eventos tirar uma bola verde, ou seja tirar uma bola verde duas vezes seguida. Esses podem são considerados independentes ou dependentes e o que determina como serão considerados é o que se faz com a bola tirada da sacola.
Se, após o primeiro evento, devolvermos a bola na sacola então os eventos são ***independentes*** caso contrário eles são considerados ***dependentes***

Matematicamente, dizemos que dois eventos E e F são **independentes** se a probabilidade de que ambos ocorram seja o produto das probabilidades que cada um deles acontece:

$$P(E, F) = P(E)P(F)$$

No exemplo acima, considerando que devolvemos a bola na sacola, a probabilidade da "primeira bola ser verde" é 3/5, e a probabilidade de "ambas serem verde" é 9/25, mas a probabilidade de "primeira ser amarela e ambas serem verde" é 0.

Eventos Independentes

Dois ou mais eventos são independentes quando a ocorrência de um dos eventos não influencia a probabilidade de ocorrência dos outros. Se dois eventos A e B são independentes então a probabilidade de A ocorrer dado que B ocorreu é igual à própria probabilidade de ocorrência de A, e a probabilidade de B ocorrer dado que B ocorreu é igual à própria probabilidade de ocorrência de B.

Se A e B são independentes então:

- P(A|B) = P(A) e P(B|A) = P(B)
- P(AnB) = P(A) X P(B|A) = P(A) X P(B)
- P(AnB) = P(B) X P(A|B) = P(B) X P(A)

AS EXPRESSÕES ACIMA SÃO VÁLIDAS SE E SOMENTE SE OS EVENTOS A E B FOREM INDEPENDENTES!

Em situações práticas dois eventos são independentes quando a ocorrência de um deles não modifica o Espaço Amostral do Experimento Aleatório.

Probabilidade Condicional

Muitas vezes há interesse de calcular a probabilidade de ocorrência de um evento B qualquer, sabendo (ou supondo) que um outro evento A ocorreu previamente. Em outras palavras queremos calcular a probabilidade de ocorrência de B CONDICIONADA à ocorrência prévia de A, simbolizada por P(B | A) - lê-se probabilidade de B dado A - e a sua expressão será:

$$P(B \mid A) = \frac{P(B \cap A)}{P(A)} \quad \text{para } P(A) > 0$$

A probabilidade de ocorrência de B **condicionada à ocorrência de A** será igual à probabilidade da intersecção entre B e A, dividida pela probabilidade de ocorrência de A (o evento que já ocorreu).

Exercício

Dado dois lançamentos sequenciais de uma moeda honesta (justa), calcule a probabilidade de sair cara em ambos lançamentos sabendo que o primeiro lançamento é cara.
Escreva um código em Python que comprove o resultado.

Dica: use a função abaixo para criar o efeito aleatorio do lançamento de uma moeda

In [ ]:
```
import random
import matplotlib.pyplot as plt
import math

def random_coin():
    return random.choice(["cara", "coroa"])

ambas_cara = 0
primeira_cara = 0

random.seed(0)
for _ in range(10000):
    primeira = random_coin()
    segunda = random_coin()
    if primeira == "cara":
        primeira_cara += 1
    if primeira == "cara" and segunda == "cara":
        ambas_cara += 1

print("P(ambas cara | primeira cara):", round(ambas_cara / primeira_ca
ra, 1))
```

Exercício

Calcule a probabilidade de, em um lançamento de dois dados honestos, a soma das faces ser 6 sabendo-se que ambas as faces apresentam valores par.

Nesse caso, a evento A é a soma dos dados dar 6, enquanto o evento B é que os dois apresentem um resultado par, certo?

Solução:
Os possíveis resultados para as faces são 36 (seis opções para o primeiro dado x seis opções para o segundo).
As seguintes combinações somam 6: {1,5}, {2;4}, {3;3}, {4;2} e {5;1}. Ou seja, 5/36. Esse é P(A)

Já as possibilidades de os dois darem resultado par são: {2,2}, {2,4}, {2,6}, {4,2}, {4,4}, {4,6}, {6,2}, {6,4}, {6,6}. No fim das contas, são 9/36 chances. Esse é P(B).

Agora, quais as opções que atendem aos dois requisitos? Somente {2,4} e {4,2}, certo? São 2/36. Esse resultado é P(A∩B). Colocando isso na fórmula, temos:

P (A|B) = P(A∩B)/P(B)

P (A|B) = (2/36)/(9/36)

P (A|B) = (2/36).(36/9)

P (A|B) = (2/36).(36/9)

P (A|B) = 2/9

Então o resultado da probabilidade condicional para essa questão é 2/9 de chances ou 22,22% de chance.
</div>

Teorema de Bayes

A ideia principal do Teorema de Bayes é inverter as probabilidades condicionais.
Por exemplo, queremos saber a probabilidade P(X|Y) mas conhecemos apenas a probabilidade de P(Y|X).

Antes, vamos reescrever a nossa fórmula da probabilidade condicional para utilizarmos de uma forma melhor. Vamos passar o denominador multiplicando para o outro lado da equação:

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

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

$$e\ também$$

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

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

$$Note\ que\ P(B \cap A) = P(A \cap B)$$

$$Então\ podemos\ concluir\ que:$$

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

Variação da regra da probabilidade condicional.

Dado que conhecemos P(A|B), queremos descobrir P(B|A), para isso vamos usar o que aprendemos com a variação da regra da probabilidade condicional e montar a tão famosa fórmula do Teorema de Bayes:

$$Sabendo\ que:\ P(A \cap B) = P(B|A) \cdot P(A)$$

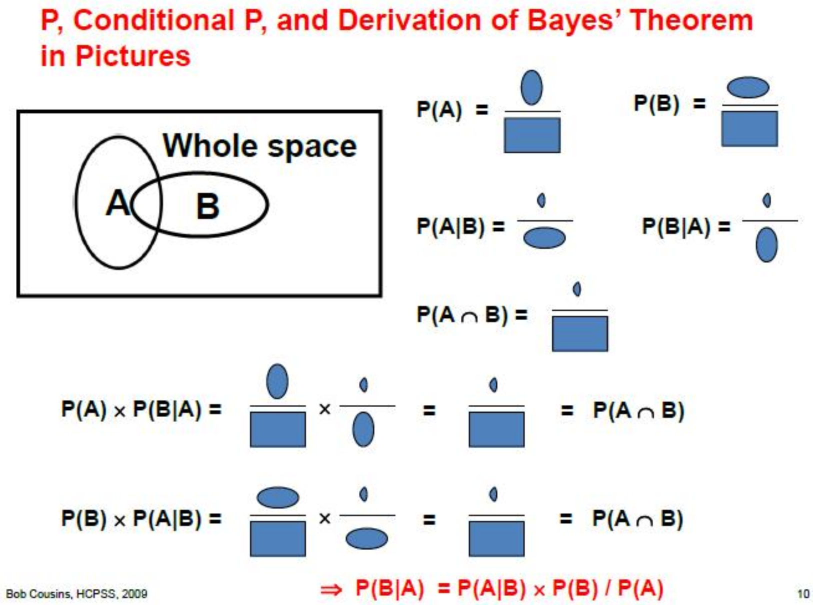
$$Podemos\ fazer:$$

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

$$E\ assim\ obtemos\ o\ \mathbf{Teorema\ de\ Bayes:}$$

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

A figura abaixo retrata o teorema de bayes em imagens, de modo a ilustrar a ideia por trás do teorema



Exemplo

Imagine uma certa doença que afeta 1 em cada 10.000 pessoas. E imagine que exista um teste para essa doença cujo resultado é correto ("doente" se você tiver a doença, "sem doença", se você não a tiver) em 99% do tempo.

Qual a probabilidade de alguém estar doente em um teste positivo?

Vamos usar **T** para o evento "o teste é positivo" e **D** para o evento "tem a doença".
Então, o teorema de Bayes diz que a probabilidade de você ter a doença, condicionada a testes positivos, é:

$$P(D \mid T) = P(T \mid D)P(D) / [P(T \mid D)P(D) + P(T \mid \neg D)P(\neg D)]$$

Sabemos que:

- P(T|D) = 0,99 - Possibilidade de alguém com a doença testar positivo
- P(D) = 0,0001 - Possibilidade de qualquer pessoa ter a doença
- P(T|Não D) = 0,01 - Possibilidade de alguém sem a doença testar positivo
- P(Não D) = 0,9999 - Possibilidade de qualquer pessoa NÃO ter a doença

Se você substituir esses números no Teorema de Bayes, você encontrará:

$$P(D \mid T) = 0.98\ \%$$

Ou seja, menos de 1% das pessoas que testam positivo realmente têm a doença.

Exemplo

Imagine que um casal tem dois filhos. Qual a probabilidade dos dois filhos serem meninos dado que um deles é menino?
Para calcular essa probabilidade, precisamos definir alguns eventos e probabilidades. Vamos definir os eventos:

- A: dois filhos meninos (evento desejado)
- B: um dos filhos é um menino (evento dado)

Solução: Para calcular essa probabilidade, precisamos definir alguns eventos e probabilidades. Vamos definir os eventos:

A: dois filhos meninos (evento desejado)
B: um dos filhos é um menino (evento dado)

Definidos os eventos, vamos definir algumas das probabilidades que precisamos para o cálculo:
P(A): probabilidade de que os dois filhos sejam meninos
P(B): probabilidade de que um filho seja um menino

Com cálculos simples, chegamos à conclusão de que a probabilidade de que dois filhos sejam meninos é ¼.
Assumindo que a probabilidade de que uma criança seja menino seja ½, então a probabilidade de que pelo menos um dos filhos do casal seja um menino é ¾.

Podemos concluir também que P(B|A), ou seja, a probabilidade de que um dos filhos seja menino dado que os dois são meninos é 1.

Sendo assim, temos:

- P(A) = 1/4
- P(B) = 3/4
- P(B|A) = 1

Logo, aplicando o Teorema de Bayes:

$$P(A|B) = \frac{1 \times \frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}$$

```
In [ ]:
```