ESTATÍSTICA

Michelle Hanne Soares de Andrade
michellehanne.andrade@gmail.com

Inferência Bayesiana

Inferência Bayesiana

A teoria Bayesiana foi desenvolvida por Thomas Bayes em meados do século XVIII, o qual propôs uma **teoria subjetiva de probabilidade**, baseada principalmente em um conhecimento a **priori** em relação às incertezas envolvidas no estudo.

Inferência Bayesiana

A **inferência Bayesiana** é o processo de encontrar um **modelo de probabilidade** para um **conjunto de dados** e resumir o resultado por **uma distribuição de probabilidade** sobre os **parâmetros** do **modelo** e sobre **quantidades não observadas,** que poderão servir de **predição** para novas observações (GELMAN ET AL., 2003)

Inferência Bayesiana

A metodologia Bayesiana

- Consiste de informações referentes aos dados amostrais **(função de verossimilhança*),** de um conhecimento prévio a respeito dos **parâmetros (distribuição a priori)** e, a partir destas duas informações, do cálculo de uma **distribuição a posteriori dos parâmetros**, na qual todas as decisões e inferências são realizadas.

*O princípio de verossimilhança sustenta que toda informação dada pela amostra ou pela experiência está contida na função de verossimilhança.

Inferência Bayesiana - Aplicações

- Previsão de Investimento e Bolsa de Valores
- Fraude em cartões de crédito
- Detectar aviões num radar
- Classificação de texto (Análise de Sentimento)
- Detecção de Ataques em Redes de Computadores
- Detecção de *Spammers* em Redes Sociais
- Reconhecimento de padrão

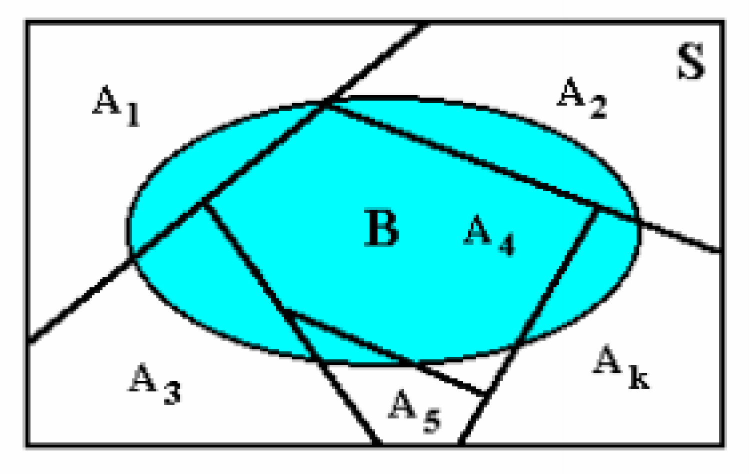
Teorema de Bayes

Definição: Seja S um espaço amostral $e\ A_1\ , A_2, \dots , A_k,\quad k$ eventos. Diz-se que $A_1\ , A_2, \dots , A_k$ formam uma partição de S se:

$$A_i \neq \varnothing,\ i = 1,\ 2,\ \dots,\ k$$

$$\bigcup_{i=1}^{k} A_i = S,$$

$$A_i \cap A_j = \varnothing,\ i \neq j$$



$A_1,\ A_2,\ A_3,\ A_4,\ A_5,\ \dots,\ A_k$ formam uma partição de S.

Teorema de Bayes

Seja B um evento qualquer de S, onde:

$$B = (B \cap A_1) \cup (B \cap A_2) \cup \ldots \cup (B \cap A_k)$$

$$P(B) = \sum_{j=1}^{k} P(A_j).P\left(B / A_j \right) \text{, } j = 1, 2, \ldots, k \qquad (*)$$

$$P(B \cap A_i) = P(A_i).P\left(B / A_i \right), \qquad (**)$$

$$P\left(A_i / B \right) = \frac{P(B \cap A_i)}{P(B)}, \qquad (***)$$

Teorema de Bayes

substituindo as equações (✱) e (✱✱) na equação (✱✱✱) temos:

$$P\left(A_i/B\right) = \frac{P(A_i).P\left(B/A_i\right)}{\sum_{j=1}^{k} P(A_j).P\left(B/A_j\right)}, j = 1, 2, ..., k.$$

Teorema de Bayes

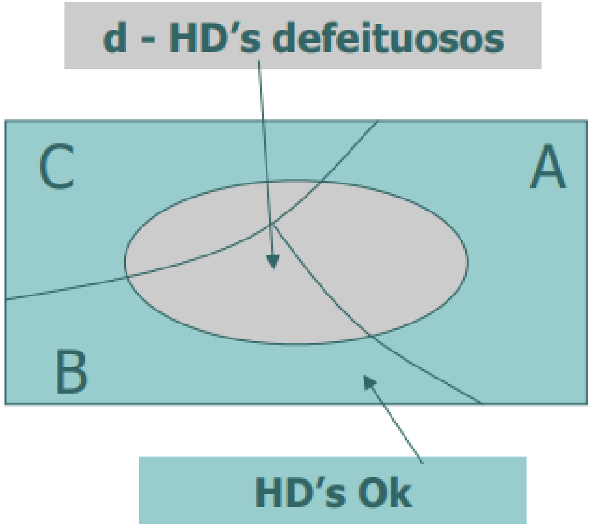
$$P(A_i \mid B) = \frac{P(A_i)\,P(B \mid A_i)}{\sum P(A_i)\,P(B \mid A_i)}$$

Teorema de Bayes – Exemplo 1

- Chamando *B* o evento "fabricado em B" e *d* o evento "HD defeituoso", podemos escrever:
- Uma peça defeituosa pode provir de qualquer uma das 3 fábricas (e só de uma!). Logo, eventos mutuamente excludentes.
- Portanto:

$$P(d) = P(A)P(d|A) + P(B)P(d|B) + P(C)P(d|C)$$

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



Teorema de Bayes – Exemplo 1

- Assim, de acordo com os valores fornecidos, temos que:

$$P(\,d\,) = (\,0{,}40 \times 0{,}02\,) + (\,0{,}35 \times 0{,}01\,) + (\,0{,}25 \times 0{,}03\,) = 0{,}019$$

- E, portanto,

$$P(B\,|\,d) = \frac{(0{,}35\,x\,0{,}01)}{(0{,}40\,x\,0{,}02) + (0{,}35\,x\,0{,}01) + (0{,}25\,x\,0{,}03)} = 0{,}184 = 18{,}4\%$$

Teorema de Bayes – Exemplo – Método Alternativo

- Construa uma Tabela de Probabilidades

	A	B	C	Totais
Bom	0,392	0,3465	0,2425	0,981
Defeito	0,008	0,0035	0,0075	0,019
Totais	0,40	0,35	0,25	1

$$P(B|\text{defeito})=0,0035/0,019=0,184$$

Teorema de Bayes – Exemplo 2

- Recomenda se que, a partir dos **40 anos, as mulheres façam mamografias anuais**. Nessa idade, **1% (0,01) das mulheres são portadoras de um tumor assintomático de mama**, ou seja, **99% (0,99) não tem câncer**. Sabe-se que a mamografia apresenta resultado positivo em **80% das mulheres com câncer de mama**, mas esse mesmo resultado ocorre também com **9,6% das mulheres sem o câncer**, isto é, a mulher pode ter o resultado positivo mesmo sem ter propriamente câncer (falso positivo).

Teorema de Bayes – Exemplo 2

- *Imagine agora que você chega em casa e encontra sua tia aos prantos, desesperada,* ***porque fez uma mamografia de rotina e o resultado foi positivo****!* ***Qual a probabilidade de ela ter um câncer de mama?***

Teorema de Bayes – Exemplo 2

Resolução: Em primeiro lugar, **sua tia tem o câncer de mama (CA)** ou não **(não-CA).** Essas alternativas, mutuamente excludentes (concorda?), podem ser colocadas em uma tabela, como abaixo:

	TEM CÂNCER	NÃO TEM CÂNCER
Probabilidade a priori	0,01	0,99

Esta é chamada **probabilidade a priori – ter câncer ou não ter**. Como em média **1% das mulheres** de 40 anos têm um tumor de mama, a **probabilidade a priori de sua tia ter um câncer é de 1% (0,01)** e de não **ter é de 99% (0,99).**

Teorema de Bayes – Exemplo 2

Agora vamos incorporar o resultado da mamografia.

- Se o câncer de mama está presente, a probabilidade condicional de a mamografia **ser positiva é 0,80 (80%),** e se não está presente **é de 0,096 (9,6%) (concorda?).**
- Multiplicando a **probabilidade a priori pela chance de a mulher ter um câncer de mama, sob condicional**, obtemos **a probabilidade conjunta:** esse ponto de vista, um teste médico funciona como um 'modificador de opinião', **atualizando uma hipótese inicial (probabilidade a priori) para gerar outra (probabilidade a posteriori).**

Teorema de Bayes – Exemplo 2

	TEM CÂNCER	NÃO TEM CÂNCER
Probabilidade a priori	0,01	0,99
Probabilidade condicional	0,8	0,096
Probabilidade Conjunta	0,01 x 0,8 = 0,008	0,99 x 0,096 = 0,0095

- Observe que a soma das **probabilidades a priori é** 1, mas isso não acontece com as **probabilidades conjuntas**.
- Para fazer com que essa segunda **soma se torne 1**, é preciso usar uma **normalização**. Para isto dividimos cada uma das probabilidades conjuntas pela **soma das duas (0,0175).** Chegamos assim à **chamada probabilidade a posteriori.**

Teorema de Bayes – Exemplo 2

	TEM CÂNCER	NÃO TEM CÂNCER
Probabilidade a priori	0,01	0,99
Probabilidade condicional	0,8	0,096
Probabilidade Conjunta	0,01 x 0,8 = 0,008	0,99 x 0,096 = 0,0095

- Observe que a soma das **probabilidades a priori é** 1, mas isso não acontece com as **probabilidades conjuntas**.
- Para fazer com que essa segunda **soma se torne 1**, é preciso usar uma **normalização**. Para isto dividimos cada uma das probabilidades conjuntas pela **soma das duas (0,0175).** Chegamos assim à **chamada probabilidade a posteriori.**

Teorema de Bayes – Exemplo 2

	TEM CÂNCER	NÃO TEM CÂNCER
Probabilidade a priori	0,01	0,99
Probabilidade condicional	0,8	0,096
Probabilidade Conjunta	0,01 x 0,8 = 0,008	0,99 x 0,096 = 0,0095

Normalização	(0,008 + 0,0095 = 0,0175)	
Probabilidade a posteriori	0,008/0,0175 = 0,46	0,0095/0,0175 = 0,54

Teorema de Bayes – Exemplo 2

- Portanto, o raciocínio Bayesiano nos levou, de modo muito simples, a concluir que a probabilidade a posteriori (ou seja, após o teste) de sua tia não ter um câncer de mama é de 0,54 (54%) e você pode tranquilizá-la de que a situação não é inevitável.

Probabilidade da interseção de dois eventos

Probabilidade da Interseção de dois Eventos

A probabilidade da interseção de dois eventos ou probabilidade de **eventos sucessivos** determina a chance, a possibilidade, de dois eventos ocorrerem simultânea ou sucessivamente.

Probabilidade da Interseção de dois Eventos

- Sejam A e B dois eventos de um espaço amostral S. A probabilidade de $A \cap B$ é dada por:

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

Onde:

$p(A \cap B)$ → é a probabilidade da ocorrência simultânea de A e B

$p(A)$ → é a probabilidade de ocorrer o evento A

$p(B|A)$ → é a probabilidade de ocorrer o evento B sabendo da ocorrência de A (probabilidade condicional)

Eventos Independentes

Se os eventos **A** e **B** forem independentes (ou seja, se a ocorrência de um não interferir na probabilidade de ocorrer outro), a fórmula para o cálculo da probabilidade da intersecção será dada por:

$$p(A \cap B) = p(A) \cdot p(B)$$

Quando o fato de ter ocorrido o evento B não alterar a probabilidade de ocorrer o evento A, ou seja, quando A e B forem eventos independentes.

Eventos Independentes - Exemplo

- Uma moeda e um dado são lançados simultaneamente. Qual a probabilidade de ocorrer coroa e número primo?

Solução: Primeiro, vamos determinar o espaço amostral S, que é o conjunto com todos os possíveis resultados. Para melhor compreensão, iremos denominar cara de C e coroa de K. Assim,

S = {(C, 1); (C; 2); (C, 3); (C, 4); (C, 5); (C, 6); (K; 1), (K, 2); (K, 3); (K, 4); (K, 5); (K, 6)} logo n(S) = 12

Eventos Independentes - Exemplo

Vamos descrever os eventos A e B.

A: ocorrer coroa

B: ocorrer número primo

É fácil ver que esses dois eventos são independentes, um pode ocorrer sem a interferência do outro. Dessa forma, para resolução, utilizaremos a fórmula:

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

p(A) = ½, pois no lançamento de uma moeda há metade de chance de sair cara e metade de sair coroa.

p(B) = 3/6 = ½, pois dos 6 possíveis resultados no lançamento de um dado, três deles são números primos.

Logo, $P(A \cap B) = 1/2 * 1/2 = 1/4$

Interseção de dois Eventos- Exemplo

- Uma urna contém 10 etiquetas identificadas pelas letras A, B, C, D, ..., I, J. Duas delas são retiradas ao acaso, sucessivamente. Qual a probabilidade de saírem duas vogais, se a extração é feita sem reposição?

Solução: Vamos determinar os dois eventos envolvidos.

Evento A: sair uma vogal
Evento B: sair uma vogal

Interseção de dois Eventos- Exemplo

O fato de não haver reposição das etiquetas indica que pois não haverá a mesma quantidade de etiquetas após a ocorrência de um deles. Dessa forma, utilizaremos a expressão:

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

Vamos então calcular p(B) e p(A|B).

p(B)= 3/10, pois, das dez letras, apenas 3 são vogais.

p(A|B)= 2/9, pois, se B ocorreu, restaram 9 letras e, dessas, apenas 2 são vogais.

Logo,

$$\mathbf{P(A \cap B) = 2/9 \cdot 3/10 = 6/90 = 1/15}$$

Teorema da Probabilidade Condicional

A probabilidade condicional também é uma probabilidade ($P(.|B)$, para B um subconjunto fixo de Ω), ou seja a probabilidade condicional satisfaz os três axiomas de probabilidade.

$$\mathbb{P}(\Omega|B) = \frac{\mathbb{P}(\Omega \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B)}{\mathbb{P}(B)} = 1$$

E que diz:

$$\mathbb{P}(\emptyset|B) = \frac{\mathbb{P}(\emptyset \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(\emptyset)}{\mathbb{P}(B)} = \frac{0}{\mathbb{P}(B)} = 0$$

E que demonstra o Primeiro Axioma.

Teorema da Probabilidade Condicional

O segundo axioma diz que $0 \leq \mathbb{P}(A|B) \leq 1$, para qualquer $A \subset \Omega$. Observe que $\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$, e como $A \cap B \subset B$. Temos que por P4 que $0 \leq \mathbb{P}(A \cap B) \leq \mathbb{P}(B)$, o que implica que $0 \leq \mathbb{P}(A|B) \leq 1$.

P4. Se A$\subset$ B, então $\mathbb{P}(A) \leq \mathbb{P}(B)$.

De fato, temos que se $A \subset B$ então $B = A \cup (B - A)$, sendo que esta união é disjunta.

Teorema da Probabilidade Condicional

- O terceiro e último axioma diz que para qualquer sequência de eventos mutuamente exclusivos A1, A2.... , temos que

$$\mathbb{P}\Big(\bigcup_{n=1}^{\infty} A_n \Big| B\Big) = \sum_{n=1}^{\infty} \mathbb{P}(A_n|B).$$

- Logo, a probabilidade condicional satisfaz todos os axiomas da probabilidade, o que implica que a probabilidade condicional também é uma probabilidade.

Árvore de Probabilidades Condicionais

- Suponha que uma pessoa está participando de um programa de televisão e lhe é fornecida a possibilidade de escolher entre portas. Atrás de uma das portas existe um carro e atrás das demais não existe prêmio algum. O participante escolhe uma porta, digamos a porta e o apresentador abre outra porta, digamos a porta , revelando que não há nada atrás dela e então oferece ao participante a oportunidade de trocar de porta. O que é mais vantajoso, trocar ou não a porta escolhida?

Árvore de Probabilidades Condicionais

- Este é um problema clássico, conhecido como paradoxo de Monty Hall.
- A resposta intuitiva é não mudar de porta. O apresentador teria nos ajudado, já que nossas chances subiram de 1/3 para 1/2, mas realmente não faria diferença trocar ou não de porta uma vez que ambas teriam as mesmas chances de possuírem o prêmio. No entanto, esta resposta está errada, pois a porta que o apresentador abre depende da porta que o concorrente escolher inicialmente.

- Na verdade, é mais vantajoso trocar de porta e, ao fazê-lo a chance do participante ganhar o carro é de 2/3.

Árvore de Probabilidades Condicionais

Solução 1: Podemos resolver utilizando a Descrição do Problema

Estratégia 1, onde o participante seleciona uma porta e, se lhe é fornecida a oportunidade de trocar de porta, ele recusa.

Estratégia 2, na qual o participante sempre troca a porta escolhida.

- Utilizando a estratégia 1, o participante ganhará o carro com **probabilidade 1/3**, já que em 1/3 das vezes a porta que ele escolhe terá o carro com o prêmio.

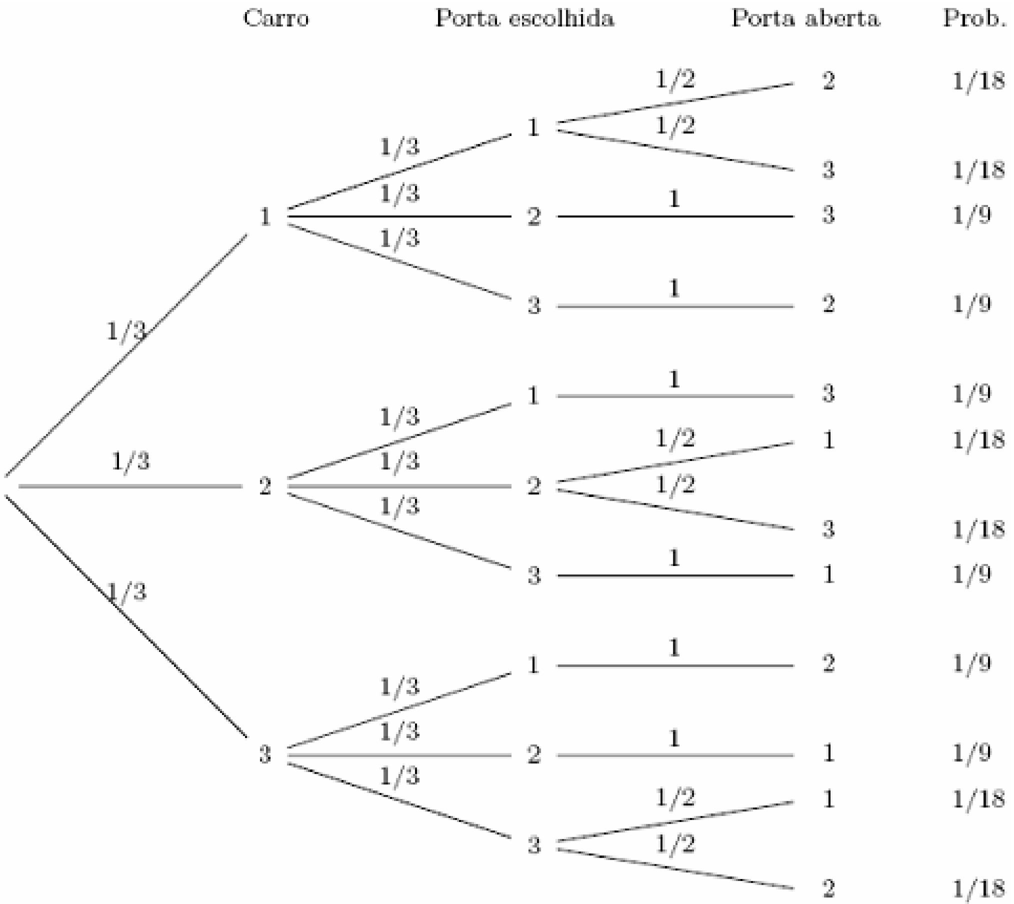
Árvore de Probabilidades Condicionais

- Utilizando a estratégia 2, o participante somente ganhará o carro se, a princípio escolhe uma porta que não contém o carro como prêmio, o que ocorre em 2/3 das vezes, ou seja, a probabilidade de ganhar com a **estratégia 2 é de 2/3 e, portanto, duas vezes maior do que utilizando a estratégia 1.**

Árvore de Probabilidades Condicionais

- **Solução 2:** Podemos resolver este problema utilizando os conceitos **de probabilidade condicional.**
- Podemos analisar o problema através de um diagrama de árvore. Assumimos que se o apresentador pode escolher entre as portas (ou seja, o participante escolheu a porta com o carro), então ele escolhe cada porta com probabilidade 1/2. Temos a Árvore Resultante.

Árvore de Probabilidades Condicionais



Exemplo 4 – Teorema Bayesiano

- Suponha que a ocorrência de chuva (ou não) dependa de das condições do tempo no dia imediatamente anterior. **Admitamos que se chova hoje, choverá amanhã com probabilidade de 0,7** e que **se não chove hoje, então choverá amanhã com probabilidade de 0,4.** Sabendo que choveu hoje, calcule a probabilidade de chover depois de amanhã.

Exemplo 4 – Teorema Bayesiano

- Solução:

Consideremos nosso espaço amostral $\Omega = \{(\omega_1, \omega_2, \omega_3) : \omega_i \in \{\text{chover, não chover}\}\}$. Seja o evento $A_1 = \{\text{chover hoje}(0)$, $A_2 = \{\text{chover amanhã}\}$ e $A_3 = \{\text{chover depois de amanhã}\}$. Queremos encontrar $\mathbb{P}(A_3|A_1)$, mas

$$\mathbb{P}(A_3|A_1) = \mathbb{P}(\Omega \cap A_3|A_1)$$

$$= \mathbb{P}(A_3 \cap (A_2 \cup A_2^c)|A_1)$$

$$= \mathbb{P}(A_3 \cap A_2|A_1) + \mathbb{P}(A_3 \cap A_2^c|A_1)$$

Exemplo 4 – Teorema Bayesiano

- Solução:

$$= \frac{\mathbb{P}({}_1 \cap A_2 \cup A_3)}{\mathbb{P}(A_1)} + \frac{\mathbb{P}({}_1 \cap A_2^C \cup A_3)}{\mathbb{P}(A_1)}$$

$$= \frac{\mathbb{P}(A_3|A_1 \cap A_2) \cdot \mathbb{P}(A_2|A_1) \cdot \mathbb{P}(A_1)}{\mathbb{P}(A_1)} + \frac{\mathbb{P}(A_3|A_1 \cap A_2^C) \cdot \mathbb{P}(A_2^C|A_1) \cdot \mathbb{P}(A_1)}{\mathbb{P}(A_1)}$$

$$= \mathbb{P}(A_3|A_1 \cap A_2) \cdot \mathbb{P}(A_2|A_1) + \mathbb{P}(A_3|A_1 \cap A_2^C) \cdot \mathbb{P}(A_2^C|A_1)$$

$$= \mathbb{P}(A_3|A_2) \cdot \mathbb{P}(A_2|A_1) + \mathbb{P}(A_3|A_2^C) \cdot \mathbb{P}(A_2^C|A_1)$$

$$= 0,7 \cdot 0,7 + 0,4 \cdot 0,3$$

$$= 0,61$$

Ou seja, sabendo que choveu hoje, a probabilidade de chover depois de amanhã é de 61%.