UNIVERSIDADE FEDERAL DE SÃO CARLOS - CAMPUS SOROCABA
DEPARTAMENTO DE FÍSICA, QUÍMICA E MATEMÁTICA
LICENCIATURA EM MATEMÁTICA

Laís Lima de Souza

APRESENTANDO O TEOREMA DE BAYES POR MEIO DE APLICAÇÕES

Sorocaba
2022

UNIVERSIDADE FEDERAL DE SÃO CARLOS - CAMPUS SOROCABA

DEPARTAMENTO DE FÍSICA, QUÍMICA E MATEMÁTICA

LICENCIATURA EM MATEMÁTICA

Laís Lima de Souza

APRESENTANDO O TEOREMA DE BAYES POR MEIO DE APLICAÇÕES

Trabalho de Conclusão de Curso de Graduação apresentado como requisito parcial para a obtenção do título de licenciada em Matemática sob a orientação do Prof. Dr. Antônio Luís Venezuela.

Sorocaba

2022

FUNDAÇÃO UNIVERSIDADE FEDERAL DE SÃO CARLOS

COORDENAÇÃO DO CURSO DE LICENCIATURA EM MATEMÁTICA DE SOROCABA - CCML-So/CCTS

Rod. João Leme dos Santos km 110 - SP-264, s/n - Bairro Itinga, Sorocaba/SP, CEP 18052-780
Telefone: (15) 32298874 - http://www.ufscar.br

DP-TCC-FA nº 2/2022/CCML-So/CCTS

Graduação: Defesa Pública de Trabalho de Conclusão de Curso

Folha Aprovação (GDP-TCC-FA)

FOLHA DE APROVAÇÃO

LAIS LIMA DE SOUZA

APRESENTANDO O TEOREMA DE BAYES POR MEIO DE APLICAÇÕES

Trabalho de Conclusão de Curso

Universidade Federal de São Carlos – Campus Sorocaba

Sorocaba, 20 de abril de 2022

ASSINATURAS E CIÊNCIAS

Cargo/Função	Nome Completo
Orientador	Prof. Dr. Antonio Luis Venezuela
Membro da Banca 1	Profa. Dra. Silvia Maria Simões de Carvalho
Membro da Banca 2	Prof. Dr. Paulo Cesar Oliveira

Documento assinado eletronicamente por **Antonio Luis Venezuela**, **Docente**, em 21/04/2022, às 18:13, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

Documento assinado eletronicamente por **Paulo Cesar Oliveira**, **Docente**, em 27/04/2022, às 13:47, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

Documento assinado eletronicamente por **Silvia Maria Simoes de Carvalho**, **Docente**, em 27/04/2022, às 14:22, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

A autenticidade deste documento pode ser conferida no site https://sei.ufscar.br/autenticacao, informando o código verificador **0656147** e o código CRC **9FA098F7**.

Referência: Caso responda a este documento, indicar expressamente o Processo nº 23112.009327/2022-07

SEI nº 0656147

Modelo de Documento: Grad: Defesa TCC: Folha Aprovação, versão de 02/Agosto/2019

AGRADECIMENTOS

Primeiramente agradeço a Deus por guiar minha caminhada até aqui, por todos os momentos em que pensei em desistir e que lamentei as dificuldades desse processo árduo e mesmo assim consegui forças para continuar.

Agradeço também aos meus pais, Silvia e Cezar e meu irmão, Caio, por todos os ensinamentos, apoio e esforços que se submeteram para me ajudar e possibilitar a realização desse sonho.

Agradeço ao meu companheiro, Lucas, por toda a compreensão, carinho, paciência e motivação que me deu para seguir em frente, me oferecendo suporte em todas as situações necessárias. Bem como agradeço a todos meus amigos e amigas que tornaram minha caminhada mais leve através dos momentos bons que passamos juntos.

Não posso deixar de agradecer àqueles que foram meus principais incentivadores desde muito nova, meus professores do ensino básico, não só os de Matemática, mas também de outras disciplinas, que sempre enxergaram meu potencial e acreditaram que eu poderia cursar uma universidade pública.

Por fim, gostaria de agradecer a todos os professores que ministraram aulas para mim ao longo da graduação, que tanto me ensinaram não apenas sobre Matemática, mas a ser uma boa profissional e a acreditar que é possível mudar a educação do nosso país. Em especial destaco o orientador, Prof. Dr. Antonio Luís Venezuela, que me guiou desde o início com muita paciência, dedicação e que sempre esteve disponível para quaisquer esclarecimentos.

Deixo também um reconhecimento à Universidade Federal de São Carlos (UFSCar), campus Sorocaba, e ao conjunto de funcionários que se doam e promovem uma educação de qualidade para todos.

RESUMO

Neste trabalho, o objetivo principal é apresentar o Teorema de Bayes por meio de algumas de suas principais aplicações: a Teoria da Decisão Bayesiana, o Algoritmo de Naive Bayes e na educação básica. Apesar de no início ter existido uma grande resistência dos matemáticos em usá-lo, este teorema é considerado atualmente uma das relações mais importantes envolvendo probabilidades condicionais e isso reflete em suas aplicações que são de grande utilidade em diversos segmentos. Para cada uma das aplicações foi abordada a linguagem matemática no estudo do referido teorema, articulado com exemplos de situações em que elas foram empregadas, tomando por base uma pesquisa qualitativa na modalidade bibliográfica. Ao fim, notou-se que, apresentar o Teorema de Bayes através das aplicações, torna-o menos complexo, constituindo-se uma forma alternativa de abordagem, em termos de ensino-aprendizagem no estudo de noções probabilísticas.

Palavras-chave: Teorema de Bayes; Aplicações; Probabilidade.

ABSTRACT

The main objective in this work is to present the Bayes Theorem through some of its main applications: the Bayesian Decision Theory, the Naive Bayes Algorithm and in the basic education. Despite at the beginning there was a great resistance from mathematicians to use it, this theorem is currently considered one of the most important relations involving conditional probabilities and this reflects in its applications that are very useful in several segments. For each of the applications, the mathematical language was approached in the study of the referred theorem, articulated with examples of situations in which they were used, based on qualitative research in the bibliographic modality. In the end, it was noticed that presenting Bayes' Theorem through applications makes it less complicated, constituting an alternative form of approach, in terms of teaching-learning in the study of probabilistic notions.

Key words: Bayes Theorem; Applications; Probability.

SUMÁRIO

1 INTRODUÇÃO E OBJETIVOS

A probabilidade é um ramo da Matemática que têm se tornado cada vez mais importante para a sociedade de modo geral e por mais que não se perceba, ela é utilizada a todo momento em nosso cotidiano. Por exemplo, quando verificamos na previsão do tempo quais são as chances de chover em determinada data, consultamos as notícias para saber se o trânsito estará congestionado ou quando jogamos na loteria, todas essas situações, entre outras, utilizam a probabilidade.

O mesmo se aplica aos meios acadêmicos, científicos e empresariais, a probabilidade apoia muitos estudos e tomadas de decisões calculando as chances de que um determinado evento ocorra e oferecendo maior grau de segurança para aqueles que a utilizam.

Entre tantas ferramentas que a probabilidade oferece, uma delas é o Teorema de Bayes (T.B.), o qual é considerado atualmente uma das relações mais importantes envolvendo situações em que deseja-se estimar as chances de um determinado evento ocorrer, dado que um outro já ocorreu. Em decorrência de sua eficiência, o T.B. possui uma variedade de aplicações, sendo assim, o objetivo geral deste trabalho é apresentá-lo por meio de algumas delas.

Já em relação aos objetivos específicos, tem-se:

- Estudar a História da Probabilidade e seus principais conceitos;
- Estudar a Probabilidade Condicional;
- Apresentar a História de Thomas Bayes;
- Desenvolver teoricamente e demonstrar o T.B.;
- Apresentar aplicações da estatística bayesiana.

Para cumprir os objetivos desta pesquisa, optamos por uma perspectiva qualitativa na modalidade bibliográfica. Para Severino (2007, p.122), a pesquisa bibliográfica realiza-se pelo:

> [...] registro disponível, decorrente de pesquisas anteriores, em documentos impressos, como livros, artigos, teses etc. Utilizam-se dados de categorias teóricas já trabalhadas por outros pesquisadores e devidamente registrados. Os textos tornam-se fontes dos temas a serem pesquisados. O pesquisador trabalha a partir de contribuições dos autores dos estudos analíticos constantes dos textos.

O primeiro trabalho selecionado foi sobre a Teoria da Decisão Bayesiana na Estratégia Mercadológica (GOLDSCHMIDT, 1970). O autor utiliza o T.B. juntamente com uma árvore de decisões para apresentar possibilidades diante de um problema de negócio, como o aumento ou não do preço de um produto, por exemplo, na qual o administrador precisa tomar uma decisão. Nas conclusões, ele cita que a abordagem bayesiana se mostrou um poderoso instrumento para auxiliar os administradores em sua difícil tarefa de tomar decisões.

Na área de medicina uma aplicação importante trata-se de usar o modelo bayesiano como um caminho para a análise estatística de dados genéticos, conforme aborda Rodrigues (2010) em seu trabalho. Ela cita que esse modelo facilita a interpretação dos resultados de problemas complexos em genética e onde muitas vezes a estatística clássica acaba falhando. Suas grandes vantagens são que ele permite enfocar uma questão de interesse de modo mais direto, incorporar informações *a priori*, evitar problemas com os testes de hipóteses múltiplas e no fim solucionar problemas complexos. O lado negativo desse modelo é sua complexidade.

Partindo para uma questão mais atual, o T.B. também pode ser aplicado no monitoramento de redes sociais, que é o tema do trabalho de Figueira et al (2013). Elas utilizaram o site Causa Brasil que monitora as questões e reivindicações nas redes sociais, relacionando-o com as notícias divulgadas na Radiobrás, criando um histórico para verificar se existe alguma relação entre as notícias da mídia com as causas das redes sociais. Ou seja, elas criaram bases técnicas através do T.B. para fazer previsões da repercussão de fatos nas redes sociais com as notícias e acontecimentos. Infelizmente os resultados finais não comprovaram essa premissa, mas as autoras afirmaram que provavelmente com um maior histórico e alguns ajustes poderiam chegar em resultados mais satisfatórios.

Outra aplicação muito interessante é a do problema ou paradoxo de Monty Hall, é um jogo em que existem três portas e em apenas uma delas um carro como prêmio, o participante escolhe uma das portas, o apresentador abre outra que não tem nada e dá ao participante a escolha de trocar sua porta. Salomão (2014) aplicou-o na educação básica, pois seus aspectos lúdicos podem ser introduzidos como uma oportunidade de cativar o aluno para o estudo, além de trazer a aplicabilidade da Matemática através de atividades em forma de jogos.

Continuando na área da educação, Ferreira (2016) também leva o T.B. para os anos finais da educação básica. Sua proposta consistiu em aplicações teóricas e práticas, envolvendo toda a parte histórica acerca do teorema, o estudo dos elementos de probabilidade necessários à compreensão do mesmo e por fim, o autor levou para os alunos três problemas que poderiam ser resolvidos a partir do T.B. que são o problema de Monty Hall, já citado no parágrafo anterior, a problemática do falso-positivo e o percentual de assertividade referente à sensibilidade do resultado de um exame. Ele concluiu que essa abordagem como conteúdo complementar do estudo de probabilidade é possível e proporciona aos alunos uma gama maior de conhecimentos, além de estimular a criatividade e o raciocínio.

Por fim, partindo para tecnologia, uma aplicação que está sendo muito utilizada atualmente é a do algoritmo classificador de Naive Bayes, originado do T.B. Melo (2007) desenvolveu um trabalho abordando o problema de reconhecimento de padrões textuais aplicada ao processo de classificação automática de documentos. O classificador Naive Bayes é um algoritmo para o aprendizado indutivo com abordagem probabilística e é um dos mais utilizados em categorização de textos por ser simples, rápido e de fácil implementação. A autora destacou no final que os resultados obtidos a partir desse classificador foram muito satisfatórios, atingindo uma taxa de 95% de acerto.

A fim de organização deste trabalho, é feita uma introdução ao tema de pesquisa, bem como é apresentado o objetivo e o estado da arte do mesmo. Em seguida, no capítulo 2 a probabilidade é apresentada tanto em seu histórico quanto os principais conceitos que são necessários para alcançar o T.B. No capítulo 3 o T.B. também tem sua história retratada, bem como seus conceitos e demonstrações. Por fim, no último capítulo são abordadas a Teoria da Decisão Bayesiana, o algoritmo de Naive Bayes e o T.B. aplicado à educação básica.

2 PROBABILIDADE

Neste capítulo será apresentada uma breve história sobre a probabilidade, quando surgiu, de que forma e principalmente como foi seu desenvolvimento ao longo dos anos. De acordo com Viali (2008), diferente da História da Matemática em geral, a História da Probabilidade é bem escassa de informações e quando são encontradas, elas estão misturadas à História da Estatística, o que acaba dificultando ainda mais as pesquisas.

Além disso, também serão apresentados conceitos básicos de probabilidade que podem ser vistos como pré-requisitos para o entendimento do T.B., como experimento, espaço amostral e evento. Por fim, trata-se da probabilidade condicional, eventos independentes e as diferenças entre a estatística bayesiana e frequentista.

2.1 BREVE HISTÓRIA DA PROBABILIDADE

Este capítulo se baseia preferencialmente no trabalho de Viali (2008), o qual define a probabilidade como o ramo da Matemática que tem o objetivo de modelar fenômenos não determinísticos, isto é, aqueles fenômenos em que o "acaso" representa um papel preponderante.

Por muito tempo e pode-se dizer que até hoje, a sociedade tem uma certa dificuldade de aceitar o acaso como um fenômeno natural, que não é algo divino, mas sim que ocorre naturalmente, sem causas totalmente definidas, sendo que apenas modelos probabilísticos podem representá-los.

Viali (2008) ainda traz que abordagens matemáticas da probabilidade só surgiram de fato há cerca de 500 anos, mas antes disso já era possível detectar atividades que tendiam a este caminho com tentativas de se calcular as possibilidades de ganhar jogos de azar e também de quantificar os riscos associados a sinistros para em seguida determinar o prêmio que uma seguradora deveria pagar ao cliente lesado.

Abordando um pouco mais sobre essas primeiras manifestações probabilísticas, a primeira delas foi em um jogo chamado Tali, o qual utilizava um dado irregular, formado pelo osso de um animal. O jogo foi utilizado para a realização de experimentos, levando ao cálculo da frequência da ocorrência de cada valor, para apostas, previsões do futuro e até mesmo tomada de decisões importantes, como a divisão de heranças. Existem alguns registros que mostram a tentativa de “prever” matematicamente os resultados que poderiam ser obtidos nesses jogos.

Outra aplicação observada foi para seguros de vida, os comerciantes utilizavam probabilidade para calcular as possibilidades da ocorrência de acidentes e até mesmo da chance de sobrevivência, para então determinar o prêmio que seria pago pelo seguro. Inclusive, os estudos matemáticos surgem exatamente atendendo a esses tipos de negócios, seguradoras, que com o progresso da probabilidade passaram a ter condições de trabalhar com embasamento científico.

Depois disso, houveram muitos outros registros de matemáticos que contribuíram com o desenvolvimento da probabilidade por todo o mundo. Os italianos foram os pioneiros nos séculos XV e XVI, que apesar de não terem desenvolvido teoremas ou conceitos, trabalharam em problemas reais envolvendo comparação de frequências de ocorrências e ganhos em jogos de azar. Pacioli (1445-1517) tentou resolver um problema que anos depois, matemáticos como Pascal (1623-1662), Fermat (1607-1665), Tartaglia (1499-1557) e Cardano (1501-1576) também tentaram, porém todos sem sucesso. Cardano inclusive, foi o primeiro a introduzir técnicas de combinatória no cálculo dos casos possíveis de um evento e também a considerar a probabilidade de um evento como a razão entre o número de casos favoráveis e o número de casos possíveis. Galileo foi outro italiano que também teve suas contribuições e está ligado ao surgimento da distribuição normal.

O francês Pascal, assim como Pacioli, desenvolveu estudos relacionados à probabilidade em jogos e obteve melhores resultados. Alguns anos depois, François Viète (1540-1603) traduziu a resolução de Pascal em um cálculo literal, mais parecido com o que é conhecido hoje. Historiadores acreditam que em um desses estudos de Pascal, os quais ele compartilhava por correspondência com Fermat, foi que iniciou a Teoria das Probabilidades, pois eles conseguiram solucionar o problema que já havia sido alvo de outros matemáticos e que mais tarde se transformou em um teorema.

Pouco tempo depois, o holandês Christiaan Huygens (1629-1695) também tomou conhecimento do problema envolvendo jogos de azar, publicando então o primeiro trabalho impresso sobre cálculo de probabilidades. Já o suíço Jacques Bernoulli (1655-1705) começou um processo de sistematizar todos esses conteúdos, de forma mais generalista e em 1689 publicou um trabalho sobre séries, com um importante teorema.

O inglês Abraham de Moivre (1667-1754) trouxe em uma de suas publicações o conceito de “independência”, investigou taxas de mortalidade e os fundamentos da teoria das anuidades, além da fórmula de Stirling que foi erroneamente atribuída ao escocês James Stirling (1692-1770) e que ele utilizou para derivar a curva normal, como uma aproximação da distribuição binomial.

Nesse período, os trabalhos fundamentais foram produzidos pelo francês Pierre-Simon Laplace (1749-1827), que ampliou o campo de aplicações da probabilidade para outras áreas, pois até então ela estava voltada principalmente para o cálculo de jogos de azar.

Foi então que, depois de Laplace muitos matemáticos começaram a se dedicar e trazer maiores contribuições a este ramo da Matemática, como Gauss (1777-1855), Euler (1707-1783), Markov (1856-1922), d’Alembert (1717-1783), entre outros. A distribuição normal, por exemplo, de extrema importância para os ramos da estatística e probabilidade, é conhecida comumente como uma descoberta de Gauss, porém muitos historiadores dizem que na verdade De Moivre e Laplace já haviam deduzido-a antes.

Mais recentemente houve também o russo Andrei Andreyevich Markov (1856 -1922) e o austriaco Richard von Mises (1883-1953) que trouxeram grandes contribuições, porém foi Andrey Nikolaevich Kolmogorov (1903-1987) que ficou conhecido como o pai da probabilidade moderna, ele apresentou resultados relevantes, publicou trabalhos que de certa forma revolucionou os conceitos que havíamos até então e além disso, suas abordagens também impressionavam pela rigorosidade.

Alguns dos trabalhos de Kolmogorov foram a apresentação do teorema das “três séries” e resultados sobre desigualdades de somas parciais de variáveis aleatórias, a publicação de uma monografia sobre os Fundamentos da Teoria da Probabilidade e a axiomatização da teoria da probabilidade.

2.2 CONCEITOS BÁSICOS DA PROBABILIDADE

Este capítulo se baseia nos estudos de Bussab e Morettin (2010) e Meyer (2000), os quais dizem que a probabilidade é um campo da Matemática que analisa e calcula as possibilidades de que algo ocorra e as chances de que se obtenham determinados resultados para tal. A probabilidade, cada vez mais, tem se tornado uma ferramenta essencial para diversas áreas do conhecimento, afinal através de cálculos matemáticos ela proporciona previsões importantes para decisões a serem tomadas. Mas para falar de probabilidade, primeiramente é essencial descrever alguns dos principais conceitos que envolvem esse tema.

Os experimentos podem ser divididos entre dois tipos de modelos, determinístico ou não-determinístico. No modelo determinístico, segundo Meyer (2000) admite-se que o resultado efetivo seja determinado pelas condições sob as quais o experimento ou procedimento seja executado. No entanto, existem situações em que esse modelo não atende, então utiliza-se o modelo não-determinístico ou estocástico, que tratam-se dos experimentos aleatórios, em que as condições da experimentação determinam somente o comportamento probabilístico do resultado observável.

Segundo Bussab e Morettin (2010, p. 103), experimento aleatório é o conjunto de todos os resultados possíveis que podem ocorrer num experimento, tais como lançar uma moeda ou um dado e observar a face de cima ou então, questionar a um consumidor sobre sua preferência entre dois ou mais produtos. Nota-se que nessas duas situações, pode-se não saber o resultado final, mas os possíveis resultados são conhecidos exatamente pelas condições do experimento. E quando executado por muitas vezes, algum padrão pode ser identificado e então surgem os modelos matemáticos.

Já o espaço amostral é o conjunto formado por todos os resultados possíveis de um experimento aleatório, por exemplo, quando lançamos um dado e observamos a sua face de cima, o espaço amostral será igual a seis, visto que um dado possui seis faces e qualquer uma delas é um possível resultado.

Outro conceito importante é o de evento, que é qualquer subconjunto de um espaço amostral definido de acordo com os objetivos do experimento, sendo que qualquer resultado individual também pode ser considerado um evento. Por exemplo, ao jogar um dado existem três possibilidades de cair um número par (2, 4 ou 6), esse subconjunto é um evento.

Pode-se dizer que, seja ε um experimento e S um espaço amostral relacionado a ε, então a cada evento A pode ser calculado e obtido um número real de $P(A)$, que representa a probabilidade de A, satisfazendo as seguintes propriedades:

1. $0 \leq P(A) \leq 1$,
2. $P(S) = 1$,
3. Se A e B forem eventos mutuamente excludentes, então $P(A \cup B) = P(A) + P(B)$,
4. Se A_1, A_2, ..., A_n, forem dois a dois eventos mutuamente excludentes, então $P(\cup_{i=1}^{\infty} A_i) = P(A_1) + P(A_2) + \ldots + P(A_n)$.

Por fim, suponha que repetimos n vezes o experimento ε, e sejam A e B dois eventos associados a ε, então n_A e n_B serão o número de vezes que os eventos A e B ocorreram nas n repetições. Portanto, $f_A = n_A/n$ é denominada frequência relativa do evento A nas n repetições de ε. Por exemplo, suponhamos que está chovendo e há uma goteira em determinada casa, causada por algumas falhas no telhado, hora o pingo da goteira cai dentro de uma panela, hora cai no fogão onde está também a panela, se uma pessoa fica observando esses pingos por muito tempo e anotando seu ponto de impacto, ela irá perceber uma sequência e então será capaz de calcular sua frequência relativa dos eventos A = {pingo cai na panela} ou B = {pingo cai no fogão}. Obviamente não é possível ao certo prever onde esses pingos irão cair, pois o experimento depende de diversas variáveis, mas se ele continuasse por um longo período de tempo, a frequência relativa iria estabilizar.

Essa característica citada no exemplo trata-se de uma propriedade chamada regularidade estatística, a qual diz que se um experimento for executado um grande número de vezes, a frequência relativa da ocorrência de um determinado evento tenderá a variar cada vez menos à medida que o número de repetições aumentar.

2.3 PROBABILIDADE CONDICIONAL

De acordo com Meyer (2000), sejam A e B dois eventos associados ao experimento ε, então pode-se denotar por $P(B|A)$ a probabilidade condicional do evento B acontecer, quando A tenha ocorrido. Mas, vamos usar um exemplo para entender o que de fato é a probabilidade condicional.

Exemplo 1: Em um lote de peças fabricadas por uma empresa, 80 delas são não-defeituosas e 20 defeituosas e então vamos extrair duas dessas peças aleatoriamente, porém primeiro repondo a peça e depois sem repor. Podem ser definidos dois eventos:

A = {a primeira peça é defeituosa};

B = {a segunda peça é defeituosa}.

Se estivermos extraindo com reposição, $P(A) = P(B) = \frac{20}{100} = \frac{1}{5}$ e esse resultado é bastante simples, porém se a extração ocorrer sem reposição será um pouco mais complexo. $P(A)$ continua tendo o mesmo valor, mas para calcular $P(B)$ deve-se conhecer a composição do lote no momento em que for extraída a segunda peça, ou seja, saber se a primeira peça foi defeituosa ou não.

Pensando a respeito desse exemplo e a probabilidade condicional, tem-se então que $P(B|A) = \frac{19}{99}$, pois se A tiver ocorrido, então para a segunda extração sobrarão apenas 99 peças, sendo que 19 serão defeituosas. Ou seja, pode-se notar que sempre que calcular $P(B|A)$, está sendo calculado $P(B)$ em relação ao espaço amostral reduzido de A e não ao espaço amostral original.

Relacionando a definição inicial com o Exemplo 1, pode-se concluir que a probabilidade condicional trata-se de analisar a chance de ocorrência de um determinado evento, dado o acontecimento prévio de outro evento obviamente relacionado, o que poderá afetar as condições do espaço amostral.

Agora pode-se seguir para uma definição formal desse conceito.

Definição 1: Admitindo-se que um experimento ε tenha sido repetido n vezes, então sejam n_A, n_Be $n_{A\cap B}$ o número de vezes que os eventos A, B e $A \cap B$ tenham ocorrido em n repetições, pode-se considerar:

$$\frac{n_{A\cap B}}{n_A} = \frac{n_{A\cap B}/n}{n_A/n} = \frac{f_{A\cap B}}{f_A},$$

em que f representa a frequência, sendo assim, $\frac{n_{A\cap B}}{n_A}$ representa a frequência relativa de B naqueles resultados em que A tenha ocorrido. Se o número de repetições for alto, então $f_{A\cap B}$ será próxima de $P(A \cap B)$ e f_A será próxima de $P(A)$. Consequentemente, a relação acima nos leva a concluir que $\frac{n_{A\cap B}}{n_A}$ será próxima de $P(B|A)$, por isso é estabelecido que $P(B|A) = \frac{P(A\cap B)}{P(A)}$, se $P(A) > 0$.

É importante observar que se $A = S$, então $P(B|S) = P(B \cap S)$, tal que $P(S) = P(B)$. Desse modo, existem duas formas de se calcular $P(B|A)$:

- de forma direta, considerando a probabilidade de B em relação ao espaço amostral reduzido de A, conforme já citado no Exemplo 1;
- ou empregando a Definição 1, em que $P(A \cap B)$ e $P(A)$ são calculados em relação ao espaço amostral original S.

Teorema 1: Se $P(B) \neq 0 \neq P(A)$, então $P(A \cap B) = P(B|A)\,P(A) = P(A|B)\,P(B)$.
O Teorema 1 é chamado de teorema da multiplicação de probabilidades ou regra do produto, considerado como a consequência mais importante da definição de probabilidade condicional.

Pode-se observar também que a igualdade permanece válida se $P(A)$ ou $P(B)$ for igual a zero, se a probabilidade condicional neste caso fosse qualquer valor real arbitrário. O Teorema 1 pode ser aplicado para calcular a probabilidade da ocorrência conjunta dos eventos A e B e também permite determinar de maneira natural a probabilidade da intersecção de mais de dois eventos, ou seja, dados os eventos $A_1, A_2,..., A_n$, $P(A_1 \cap A_2 \cap ... \cap A_n) = P(A_1)P(A_2|A_1)... P(A_n|A_1 \cap A_2 \cap ... \cap A_{n-1})$ com a respectiva interpretação das probabilidades condicionais envolvidas. Essa igualdade é muito útil em experimentos que consistem de etapas dependentes.

Para aplicar a Definição 1, volta-se ao exemplo do lote de peças defeituosas e não-defeituosas, mas dessa vez se forem escolhidas aleatoriamente duas peças, sem reposição, qual será a chance de que ambas sejam defeituosas?

Sabe-se que pelo Teorema 1 visto acima, $P(A \cap B) = P(B|A)\, P(A)$ e conforme deduzido anteriormente $P(B|A) = \frac{19}{99}$ e $P(A) = \frac{1}{5}$, o que implica que $P(A \cap B) = \frac{19}{495}$.

Além de utilizar a probabilidade condicional para calcular a chance de dois eventos ocorrerem em conjunto, pode-se também utilizar esse conceito para um simples evento A.

Definição 2: Diz-se que os eventos $B_1, B_2, ..., B_k$ representam uma partição do espaço amostral S, quando:

a) $B_i \cap B_j = \emptyset$ para todo $i \neq j$, sendo $i, j = 1, 2, ..., k$,

b) $\bigcup_{i=1}^{k} B_i = S$,

c) P(B_i) > 0 para $i = 1, 2, ..., k$.

Ou seja, quando o experimento ε é realizado um, e somente um, dos eventos B_i ocorre. Vamos exemplificar para ficar mais fácil; em um jogada de um dado pode-se ter três eventos que representam uma partição do espaço amostral, sejam $B_1 = \{1, 2, 3\}$, $B_2 = \{4, 5\}$ e $B_3 = \{6\}$. Considerando A um evento qualquer referente a S e $B_1, B_2, ..., B_k$ uma partição de S, então pode-se escrever:

$A = (A \cap B_1) \cup (A \cap B_2) \cup ... \cup (A \cap B_k)$.

Alguns dos conjuntos $A \cap B_j$ podem ser vazios, mas isso não invalida a decomposição de A e todos os eventos $A \cap B_1, \dots, A \cap B_k$ são dois a dois mutuamente excludentes. Sendo assim, pode-se aplicar a propriedade da adição de eventos mutuamente excludentes e escrever:

Propriedade 1: $P(A) = P(A \cap B_1) + P(A \cap B_2) + ... + P(A \cap B_k)$.

Sabendo que se $P(A \cap B_j)$ pode ser escrito como $P(A|B_j)P(B_j)$, então tem-se o teorema da probabilidade total, enunciado a seguir.

Teorema 2: Seja B_1, B_2,..., B_k uma partição do espaço amostral S e seja A um evento qualquer em S, então $P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + ... + P(A|B_k)P(B_k)$, ou seja, $P(A) = \sum_{i=1}^{k} P(A|B_i)P(B_i)$.

O Teorema 2 pode ser muito útil porque quando precisa-se calcular $P(A)$ diretamente é complicado, mas sabendo que B_j já ocorreu, então pode ser aplicada a fórmula mencionada.

Por fim, voltando novamente ao Exemplo 1 do lote de peças defeituosas e não-defeituosas, do qual serão extraídas duas peças sem reposição. Pode-se calcular $P(B)$ da seguinte forma:

$P(B) = P(B|A)P(A) + P(B|\overline{A})P(\overline{A}) \Rightarrow P(B) = \frac{19}{99} \cdot \frac{1}{5} + \frac{20}{99} \cdot \frac{4}{5} = \frac{1}{5}$, sendo que $P(\overline{A})$ é a probabilidade da não ocorrência de A, ou seja, $\overline{A}$ é o inverso de A.

Esse resultado é idêntico ao encontrado quando calculado $P(B)$ com reposição.

2.4 EVENTOS INDEPENDENTES

Segundo Meyer (2000) e conforme visto no Subcapítulo 2.3, na probabilidade condicional deseja-se saber as chances de ocorrência de um determinado evento B dada a informação de que um outro evento A já ocorreu, porém em muitos casos saber que B ocorreu não influencia em nada a ocorrência ou não de A. Basta pensar na seguinte situação: tem-se um dado equilibrado, o qual é jogado duas vezes, considere os seguintes eventos A e B, dados por:

A = {na primeira jogada o dado mostra um número par} e

B = {na segunda jogada o dado mostra o número 5}.

Intuitivamente, pode-se concluir que os eventos A e B não possuem nenhuma relação, a primeira jogada resultar em número par não vai mudar em nada a segunda jogada. Mas vejamos outra situação com os seguintes eventos C e D, dados por:

C = {pessoa contrai AIDS} e

D = {pessoa é usuária de drogas injetáveis}.

A chance de C ocorrer dado que D já ocorreu aponta uma probabilidade muito maior. Sendo assim, é possível observar a diferença entre eventos dependentes e independentes.

Definição 3: A e B serão eventos independentes se, e somente se, $P(A \cap B) = P(A)P(B|A) = P(A)P(B)$, $P(A \cap B) = P(B)P(A|B) = P(B)P(A)$.

A Definição 3 é equivalente a dizer que A e B são independentes quando $P(B|A) = P(B)$ e $P(A|B) = P(A)$, o que acaba sendo intuitivo pois diz que A e B serão independentes se o conhecimento da ocorrência de A de nenhum modo influenciar a probabilidade da ocorrência de B. Na maioria das situações será necessário adotar a hipótese de independência de dois eventos A e B e depois empregar essa suposição para calcular $P(A \cap B)$ como igual a $P(A)P(B)$, simplificando razoavelmente os cálculos. Porém, é importante validar com cuidado as condições sob as quais o experimento é realizado para estabelecer a validade de uma suposição de independência entre os vários eventos.

3 O TEOREMA DE BAYES

Neste capítulo será iniciado o assunto que é a base para a realização deste trabalho de conclusão de curso, o Teorema de Bayes. Inicialmente, será contada a história do reverendo e matemático que criou e deu nome a esse teorema, Thomas Bayes. Infelizmente, não se encontra abundância de informações a seu respeito, mas sim um resumo de quem ele foi; nem mesmo sobre seus estudos existem informações concretas. O leitor verá neste capítulo que o T.B. não começou a ser aplicado tão recentemente, ele já vem sendo útil há bastante tempo para a humanidade.

Em seguida, será apresentado e explicado o T.B., que é uma das relações mais importantes envolvendo probabilidades condicionais, segundo Bussab e Morettin (2010, p. 116), pois ele expressa uma probabilidade condicional em termos de outras probabilidades condicionais e marginais. Para que a compreensão seja completa, também será apresentada a demonstração do teorema em sua forma simples e geral.

3.1 HISTÓRIA DE THOMAS BAYES

O criador do T.B., segundo Pena (2009), foi um reverendo presbiteriano chamado Thomas Bayes. Ele era o mais velho entre os sete filhos de seus pais, Joshua Bayes (1671-1746) e Anne Carpenter (s.d.[1]) e viveu no início do século XVIII, sendo que sua data de nascimento não é conhecida ao certo, provavelmente foi no ano de 1701 ou 1702, falecendo em 1761.

Ele nasceu e viveu na Inglaterra, porém estudou Teologia e Lógica na Universidade de Edimburgo, na Escócia, pois segundo O'Connor e Robertson (2004), seu pai foi um dos primeiros ministros não-conformistas ordenado na Inglaterra e na época não era permitido aos não-conformistas se matricularem nas universidades locais e nem mesmo nas escolas de ensino regular, por isso historiadores se dividem em duas opiniões, alguns dizem ser muito provável que

1 Sem data, informação desconhecida.

Bayes tenha sido educado em aulas particulares, enquanto outros dizem que ele recebeu uma educação liberal para o ministério.

Acredita-se também que, por todo seu conhecimento com os números, Thomas tenha estudado Matemática em algum momento. Apesar de não existirem evidências que isso tenha ocorrido na Universidade de Edimburgo, vários registros informais tanto de Thomas, quanto de colegas, mostram que ele era um bom ministro, tanto quanto matemático. Um caderno, que certamente foi escrito por ele, mostra diversas discussões sobre probabilidade, geometria, trigonometria, soluções de equações, séries e cálculo diferencial.

Ele expôs sua teoria de probabilidade, que hoje é conhecido como o Teorema de Bayes, em um documento chamado "Ensaio para resolver um problema na doutrina das chances", que foi enviado por Richard Price (1723-1791), um amigo de Bayes, poucos anos após a sua morte e publicado pela Royal Society, sociedade científica britânica. O próprio Richard reconheceu que aquele trabalho era de grande mérito.

Segundo Pena (2009), após a publicação da Royal Society, o teorema caiu no esquecimento e voltou a ser estudado pelo matemático francês Pierre-Simon de Laplace (1749-1827) anos depois, quando realmente ganhou maior destaque.

Entretanto, segundo Ferreira (2016, apud McGrayne, 2011), após a morte de Laplace, considerado como o primeiro bayesiano do mundo, o T.B. novamente declinou por aproximadamente um século, os teóricos diziam que seus cálculos eram complexos demais para compreender. Então, o T.B. só voltou a ser relembrado e utilizado na Segunda Guerra Mundial pelos britânicos com o objetivo de decifrar os códigos que os alemães e soviéticos utilizavam durante a guerra pelas linhas telefônicas. Porém, após o final da guerra, todas as provas de que a decodificação havia sido promissora foram destruídas e o T.B. mais uma vez caiu no esquecimento.

De acordo com McGrayne (2011), o T.B. foi utilizado para diversas outras finalidades, como para a definição de taxas de seguradoras, a Marinha dos Estados Unidos usou-o para procurar uma bomba perdida e localizar submarinos soviéticos, RAND Corporation avaliou a probabilidade de um acidente nuclear, entre outros.

Sempre que o T.B. caia no esquecimento, a estatística frequentista se fortalecia, porém ela só fornecia bons resultados quando uma hipótese era um caso especial de outra, mas quando hipóteses estavam competindo e tinham mudanças bruscas nos dados o frequentismo não funcionava.

Apesar da resistência no uso do teorema e das queixas sobre seus cálculos complexos, nos anos 90 quando a tecnologia começou a avançar e os computadores se tornaram mais rápidos, o T.B. ganhou muito destaque em vários campos de estudo, mostrando-se muito eficaz no desenvolvimento de pesquisas.

3.2 INTRODUZINDO O TEOREMA DE BAYES

Segundo Rossato (2016, apud Jaynes et. al., 2003), existem três pontos principais em que o T.B. se apoia para encontrar a probabilidade de determinados acontecimentos: a representação numérica, ou seja, o grau de certeza ou incerteza de um evento deve ser representado numericamente por uma função matemática, a correspondência qualitativa com o senso comum, ou seja, quando recebe-se informações sobre um evento existe uma garantia de que a probabilidade de sua ocorrência aumente e o terceiro e último ponto é a consistência, ou seja, se forem realizadas duas tentativas de formas diferentes de um cálculo, no fim as duas devem obter o mesmo valor.

Os autores Bussab e Morettin (2010, p. 116) também trazem que o T.B. pode ser apresentado de duas formas, a mais simples e a geral que será apresentada no Subcapítulo 3.3.

Teorema 3: Sejam A e B eventos independentes e suponha que suas probabilidades sejam conhecidas, então através da equação a seguir é possível calcular a probabilidade de que o evento A ocorra, dado que B já ocorreu:

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

Considerando que:

a) $P(A)$ é a probabilidade *a priori* de A, ou seja, um valor pressuposto por experiência, que ainda não foi verificado,
b) $P(B|A)$ é a probabilidade condicional de B dado que A ocorreu,
c) $P(B)$ é a probabilidade total dos eventos B ocorrerem,
d) $P(A|B)$ é a probabilidade *a posteriori* de A ocorrer dado que B ocorreu, ou seja, um valor já calculado com base em novas informações fornecidas.

Resumidamente, a probabilidade inicial é atualizada, multiplicando-a por $P(B|A)$. Pode-se observar também que $P(A|B) > P(A)$, se $P(B|A) > P(B)$.

3.2.1 Demonstração da forma simples do Teorema de Bayes

Conforme mencionado no Subcapítulo 3.2, segundo Bussab e Morettin (2010, p. 116), existem duas versões da fórmula do T.B. A forma simples, do Teorema 3, que será demonstrada a seguir, é obtida facilmente através da probabilidade condicional e ela pouco é utilizada porque a probabilidade total dos eventos ocorrerem, que está no denominador da fórmula, seja ela $P(A)$ ou $P(B)$ dependendo da situação, raramente é conhecida, geralmente depende de vários fatores, ao contrário das demais probabilidades que precisa-se para chegar ao resultado e normalmente são dadas pela situação enunciada. Porém, se em algum caso essa informação for fornecida, a forma simples pode e deve ser utilizada.

Pelo Teorema 1 da probabilidade condicional, tem-se que

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

Analogamente pode-se dizer que

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

E que, pela Definição 3, pode ser escrita como $P(A \cap B) = P(A)P(B|A)$, ou seja:

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

Tendo conhecido a equação simples do T.B., referente ao Teorema 3, é possível verificar o porquê seria tão complexo obter o valor do denominador, no caso $P(B)$, de forma direta.

Exemplo 2: Suponha-se que um exame de câncer, o qual tem uma acurácia de 90%, foi realizado em um paciente obtendo resultado positivo, deseja-se saber então qual é a probabilidade de que essa pessoa realmente tenha a doença. Historicamente, sabe-se que apenas 1% da população tem câncer. Sendo assim, tem-se as seguintes informações para a resolução da equação referente ao Teorema 3:

$P(A)$ = {ter câncer},

$P(B)$ = {resultado do exame positivo},

$P(A|B)$ = {probabilidade de ter o câncer, dado que o resultado do exame foi positivo},

$P(B|A)$ = {probabilidade do resultado do exame ser positivo, dado que o paciente tem o câncer}.

De acordo com as informações fornecidas anteriormente, $P(B|A)$ é a própria acurácia do exame, ou seja, $P(B|A) = 0,9$, enquanto que $P(A) = 0,01$, que é a chance baseada nos dados históricos. Entretanto, o valor de $P(B)$ não é possível determinar de forma direta, pois não foi mencionado pelo enunciado, já que depende de outras variáveis. Veja, a probabilidade do resultado do exame ser positivo será diferente se a pessoa tiver ou não a doença, por isso a forma simples do T.B. é raramente utilizada.

3.3 A FORMA GERAL DO TEOREMA DE BAYES

Para apresentar a forma geral do T.B., o autor matemático Meyer (2000), o qual terá seu trabalho utilizado como base neste Subcapítulo, traz a seguinte citação abaixo.

Exemplo 3: Suponha que uma determinada peça é produzida por três fábricas e sabe-se que a fábrica X produz o dobro de peças que a Y, que por sua vez produz o mesmo número de peças que a Z. Além disso, sabe-se também que 2% das peças fabricadas pelas empresas X e Y são defeituosas, já pela empresa Z são 4% de peças defeituosas. Todas as peças são colocadas em um depósito ao final da produção, independente de serem defeituosas ou não. Supondo que uma peça retirada aleatoriamente do depósito seja defeituosa, qual a probabilidade de que tenha sido produzida pela fábrica X?

Então Meyer (2000) continua explicando que, nesse caso, o que precisa-se calcular é $P(B_1|A)$, ou seja, a probabilidade condicional do evento B_1 acontecer, quando A tiver ocorrido, sendo:

A = {peça defeituosa};

B_1 = {peça produzida pela fábrica X}.

Teorema 4: Sejam os eventos B_1, B_2,..., B_k uma partição do espaço amostral S e seja A um evento associado a S, então aplica-se o Teorema 1 da probabilidade condicional e tem-se:

$$P(B_i|A) = \frac{P(A|B_i)\,P(B_i)}{\sum_{j=1}^{k} P(A|B_j)P(B_j)},\ i = 1,\ 2,\ \ldots\ ,\ k.$$

O resultado acima é conhecido como a versão geral do T.B., também conhecido como fórmula da probabilidade das "causas", pois tem-se que desde que os B_i constituam uma partição do espaço amostral, então um, e somente um, dos eventos B_i ocorrerá. Entretanto, para aplicar esse teorema os valores da $P(B_i)$ devem ser conhecidos, o que não acontece com muita frequência, limitando sua aplicabilidade e ocasionando escolhas impróprias da mesma.

Pode-se então voltar ao Exemplo 3 e aplicar o T.B:

$$P(B_1|A) = \frac{(0{,}02)\cdot(0{,}5)}{(0{,}02)\cdot(0{,}5)+(0{,}02)\cdot(0{,}25)+(0{,}04)\cdot(0{,}25)} = 0,4.$$

3.3.1 Demonstração da forma geral do Teorema de Bayes

A forma geral do T.B. nada mais é do que a sua forma simples, porém com o denominador escrito de outra maneira genérica, que representa a soma das probabilidades de determinado evento ocorrer ou não. Essa maneira genérica facilita e muito a resolução da fórmula de Bayes nas situações, pois as probabilidades se tornam conhecidas.

A partir da probabilidade condicional, que calcula as chances de ocorrência de um evento A dado que um evento B ocorreu, tem-se:

$$P(A|B) = \frac{P(A\cap B)}{P(B)},\ P(B) \neq 0.$$

Analogamente:

$$P(B|A) = \frac{P(A\cap B)}{P(A)},\ P(A) \neq 0.$$

Sabendo que uma das propriedades da união é a comutatividade, ou seja $A \cap B = B \cap A$, então:

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

$\Rightarrow P(A|B)\, P(B) = P(B|A)\, P(A) \Rightarrow$

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

A partir do Teorema 2 da probabilidade total, tem-se que:

$$P(A) = \sum P(A|B_i)P(B_i).$$

Analogamente:

$$P(B) = \sum P(B|A_i)P(A_i).$$

Substituindo então a fórmula do Teorema 2 da probabilidade total na fórmula da probabilidade condicional, obtém-se o T.B., enunciado no Teorema 4:

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

Pode-se notar que voltando ao Exemplo 2 citado no Subcapítulo 3.2.1, agora sim a partir da equação do Teorema 4 é possível calculá-la, pois o $P(B)$ é equacionado de forma que é calculada a probabilidade do evento B ocorrer nas duas situações, quando A ocorre e quando A não ocorre. Veja:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A_1)P(A_1)+P(B|A_2)P(A_2)},$$

sendo A_1 a probabilidade de realmente o paciente ter câncer e A_2 não ter. Substituindo então os valores fornecidos no Subcapítulo 3.2.1, tem-se:

$$P(A|B) = \frac{0{,}90\cdot 0{,}01}{0{,}90\cdot 0{,}01+0{,}10\cdot 0{,}99} = \frac{0{,}009}{0{,}009+0{,}099} = \frac{0{,}009}{0{,}108} = 0{,}08\overline{3},$$

ou seja, a probabilidade da pessoa ter o câncer, dado que o resultado do exame foi positivo é de 8,3%.

3.4 ESTATÍSTICA BAYESIANA E FREQUENTISTA

No ramo da Estatística existem duas grandes filosofias, a bayesiana e a frequentista, que na verdade são dois pontos de vista diferentes a respeito da probabilidade, então é comum que elas sejam comparadas em estudos, por exemplo.

Para os frequentistas, a probabilidade está fundamentalmente relacionada com a frequência de eventos repetidos, enquanto que para os bayesianos, a probabilidade está fundamentalmente relacionada com a incerteza dos eventos.

Exemplificando de forma prática, suponha que se tenha uma moeda e o experimento seja jogá-la para cima, com o evento de interesse em verificar quantas vezes a face “cara” da moeda aparece para cima após o lançamento. Depois disso definido, o experimento começa a ser executado e observada a realização do evento de interesse por diversas vezes, quantas se queira.

Na estatística clássica, o cálculo para essa probabilidade é feito a partir da divisão do evento de interesse pelo espaço amostral, ou seja, o resultado seria igual a 50% de chances no caso da moeda. Na abordagem frequentista, o experimento é realizado muitas vezes até que seja observado que a probabilidade está convergindo para um determinado valor.

Já no caso da abordagem bayesiana, utilizando o mesmo experimento e evento citados anteriormente da moeda, para calcular a probabilidade de ocorrência, é definido inicialmente um valor baseado em experiência e conhecimentos anteriores. Nesse caso, pode-se dizer que as chances de cair a face “cara” da moeda para cima em uma jogada é de 50%, intuitivamente, já que sabe-se que uma moeda tem duas faces. Em seguida, são observados alguns eventos e por fim, utilizada a equação do T.B., referente ao Teorema 4, para atualizar o valor dado inicialmente, ou seja, gerando então a probabilidade *a posteriori*.

Sendo assim, uma das grandes diferenças entre as duas abordagens é a quantidade de eventos que precisam ser realizados. Na bayesiana, esse número é bem mais baixo, já que com poucos dados ela já apresenta resultados satisfatórios.

Independente de qual seja, em ambas as abordagens existem os parâmetros desconhecidos, dados observados e modelo, que é a ligação entre os parâmetros e dados. Porém, no ponto de vista do frequentista, os parâmetros do modelo são fixos, os dados aleatórios e os modelos são inúmeros, a cada parâmetro definido vai existir um modelo diferente. Enquanto que no bayesianismo, os parâmetros são aleatórios, os dados fixos e o modelo único.

Na prática, Bayes é mais indicado em situações de resolução de problemas, pois funciona muito bem para cenários com poucos dados, permite adicionar experiências empíricas, ou seja, um valor definido de forma intuitiva e ainda assim, atualizar esse valor após a observação. Em termos de implementação, também é vantajoso porque possui um único estimador. Já em relação aos resultados, todos são em termos de probabilidade, facilitando o entendimento de qualquer pessoa que precise interpretá-los.

Segundo Oliveira (2014), uma das desvantagens da abordagem bayesiana é que na maior parte dos problemas, o cálculo da função *a posteriori* de forma analítica e numérica podem ser bastante complexos, principalmente em casos multi-dimensionais. Porém, nos últimos anos, com o avanço da tecnologia e dos computadores, surgiram métodos de simulação como Monte Carlo e Monte Carlo Via Cadeias de Markov (MCMC) que facilitam e agilizam esse trabalho, tornando totalmente viável a aplicação dessa abordagem.

4 APLICAÇÕES

O T.B. possui uma vasta gama de aplicações, tanto em seu formato original, em que existe uma situação e deseja-se calcular uma probabilidade *a posteriori*, como é o caso das aplicações na educação básica que serão mencionadas nesse capítulo, mas também é base de desenvolvimento para outras aplicações, como é o caso da Teoria da Decisão Bayesiana e algoritmo de Naive Bayes.

Na medicina, por exemplo, o T.B. pode ser utilizado como um caminho alternativo para a análise estatística de dados genéticos, já que a estatística clássica muitas vezes falha nesse processo, de acordo com Rodrigues (2010). A autora cita que em um estudo sobre a estrutura genética de 20 comunidades de um determinado organismo, na estatística clássica consideram-se os estudos prévios realizados em uma dessas comunidades apenas, ignorando as informações de todas as outras, sendo que elas poderiam ser relevantes ao estudo. E é exatamente essa possibilidade que o método Bayesiano oferece, proporcionando resultados diferentes e que condizem mais com o que é esperado.

O T.B. também foi utilizado em um estudo sobre principais fatores de risco e preditores para o aborto induzido, sendo aplicado para a determinação das probabilidades *a posteriori* e permitindo a transformação dos dados agregados em dados individuais, conforme o trabalho feito por Olinto e Moreira-Filho (2006). Os autores também mencionam que a vantagem de utilizar o T.B. ao invés de outras abordagens é o fato de considerar a probabilidade de que uma hipótese seja verdadeira frente aos dados observados.

Outro estudo de Martinez et al (2008), desenvolve um modelo estatístico Bayesiano para estimar o risco de infecção tuberculosa para estudos com perdas de seguimento e assim como os dois outros trabalhos citados na área da medicina, também compara os resultados com um outro modelo estatístico mais tradicional e destaca a importância do conhecimento do pesquisador nessa metodologia. Os autores concluem ao final do artigo que esse modelo exige o uso de um algoritmo para a simulação das distribuições *a posteriori* dos parâmetros de interesse, o que acaba causando certa complexidade, mas que isso não deve ser considerado um impeditivo já que existem programas computacionais facilitadores e que não exigem amplos conhecimentos de programação. Ou seja, o modelo Bayesiano mostrou-se uma alternativa muito eficiente para o objetivo proposto.

Já na área da economia as aplicações também são muito relevantes. Uma delas trata sobre a aplicação do modelo “Bayes-Fuzzy” para a previsão de um mercado financeiro. Os autores Silva et. al. (2009) criaram um modelo para prever variáveis do mercado financeiro e fundamentar decisões de investimento, como saber se é o momento certo para vender ou comprar uma ação. Tal modelo utiliza a lógica Fuzzy e o T.B. e tem como base o histórico de valorizações de um ativo para então prever o que deve acontecer com ele. Os resultados finais foram satisfatórios quando comparados a outro modelo, porém os autores citam que esse é um modelo simplificado, ou seja, ainda não é adequado para ser colocado em prática no mundo real.

Outro trabalho bastante interessante é o de Oliveira (2014), o qual tem o objetivo analisar riscos de crédito, desenvolvendo um modelo que incorpore a informação fornecida por um perito com a informação histórica de uma carteira de crédito. Isso justifica o uso do T.B., que permite o uso das duas fontes de informação.

Os usos do T.B. citados acima foram apenas alguns exemplos das muitas áreas que podem usufruir desta ferramenta probabilística, nos próximos Subcapítulos serão abordados em mais detalhes a Teoria da Decisão Bayesiana, o algoritmo de Naive Bayes e como aplicar o T.B. na educação básica.

4.1 TEORIA DA DECISÃO BAYESIANA

A parte teórica sobre o conteúdo abordado neste capítulo foi baseada no trabalho de Kinas e Andrade (2017).

A Teoria da Decisão Bayesiana é uma abordagem estatística fundamental em problemas de classificação ou reconhecimento de padrões, além de também ser utilizada como técnica para analisar e apoiar a tomada de decisões em clima de incerteza, assim como Goldschmidt (1970) aplicou em seu trabalho em relação a estratégia mercadológica. Sendo assim, o campo de aplicação da Teoria da Decisão Bayesiana é muito amplo, afinal em todos os ambientes e cenários é necessário lidar com tomada de decisões e com o avanço da tecnologia, o reconhecimento de padrões também se tornou uma realidade, sendo aplicado em diversos aspectos.

Para entender como funciona essa teoria é preciso primeiramente verificar os componentes existentes para modelar o problema:

- Classes $\omega = (\omega_1, \omega_2, \ldots, \omega_k)$ e vetores de padrões desconhecidos $x = (x_1, x_2, \ldots, x_k)$;
- P(ω_k): probabilidades *a priori* das i-ésimas classes, que conforme já discutida no Subcapítulo 3.2, é uma noção intuitiva;
- P($x|\omega_k$): probabilidades condicionais das classes ω_k, ou seja, é o modelo probabilístico que descreve o comportamento da variável aleatória x dado que ela pertence a classe ω_k;
- P($\omega_k|x$): probabilidades *a posteriori* das classes ω_k, que é a probabilidade de decidir pela classe ω_k dado que está sendo observado o x, ou seja, é a probabilidade *a priori* atualizada;

Com a combinação desses três componentes é possível calcular a probabilidade *a posteriori* a partir da densidade condicional e da probabilidade *a priori*, obtendo:

$$P(\omega_k|x) = \frac{P(x|\omega_k)\,P(\omega_k)}{P(x)},$$

em que $P(x)$ é uma constante normalizadora que não depende da classe ω_k, então será igual para todas as classes, ou seja:

$$P(x) = \sum_{j=1}^{k} P(x|\omega_j)\,P(\omega_j),$$

o que resulta finalmente em:

$$P(\omega_k|x) = \frac{P(x|\omega_k)\,P(\omega_k)}{\sum_{j=1}^{k} P(x|\omega_j)\,P(\omega_j)}. \qquad \text{(I)}$$

Dada a equação (I) acima, a Teoria da Decisão Bayesiana diz que:

- se $P(\omega_1|x) > P(\omega_2|x)$, então classifique $x \in \omega_1$;
- se $P(\omega_2|x) > P(\omega_1|x)$, então classifique $x \in \omega_2$;

- e por fim, se $P(\omega_1|x) = P(\omega_2|x)$, então a classificação de x será arbitrária.

Pode-se notar que isso nada mais é do que uma formalização do senso comum, pois ela ordena que a classificação seja feita de acordo com a classe mais provável.

Assim como qualquer teorema probabilístico, existem possibilidades de erros, ou seja, classificar um objeto da classe ω_1 como sendo da classe ω_2 e vice-versa, porém a Teoria da Decisão Bayesiana é excelente em relação à minimização da probabilidade de erro na classificação. Sabendo que, conforme mostra a Figura 1, R_1 é a região de decisão em que a classificação sempre será da classe ω_1, então o erro da teoria (P_E), que é representado pela área hachurada na Figura 1, pode ser calculado por:

$$P_E = P(x \in R_1,\ \omega_2) + P(x \in R_2,\ \omega_1)\ ,$$

utilizando então a definição de probabilidade condicional em que $P(A|B) = \frac{P(A,B)}{P(B)}$, temos

$$P(x \in R_1 |\ \omega_2)P(\omega_2) + P(x \in R_2 |\ \omega_1)P(\omega_1) =$$

$$= \int_{R1} P(x|\omega_2)P(\omega_2)\, dx + \int_{R2} P(x|\omega_1)P(\omega_1) dx.$$

Porém, pelo T.B. vimos que $P(\omega_j|x)$ pode ser escrita como

$\frac{P(x|\omega_j)P(\omega_j)}{P(x)} \Rightarrow P(x|\omega_j)P(\omega_j) = P(\omega_j|x)P(x)$, sendo assim

$$P_E = \int_{R1} P(\omega_2|x)P(x)\, dx + \int_{R2} P(\omega_1|x)P(x)\, dx.$$

Nota-se que

$$P(\omega_1) = \int_{-\infty}^{\infty} P(x,\ \omega_1)dx = \int_{-\infty}^{\infty} P(\omega_1|x)P(x)\, dx = \int_{R1} P(\omega_1|x)P(x)\, dx + \int_{R2} P(\omega_1|x)P(x)\, dx \Rightarrow$$

$\Rightarrow \int_{R2} P(\omega_1|x)P(x)\, dx - \int_{R1} P(\omega_1|x)P(x)\, dx$, substituindo então em P_E:

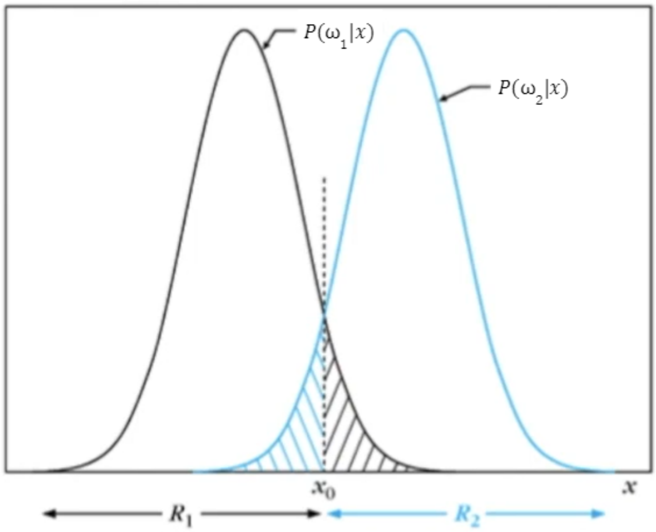
$$P_E = \int_{R1} P(\omega_2|x)P(x)\, dx + P(\omega_1) - \int_{R1} P(\omega_1|x)P(x)\, dx,$$

temos finalmente o seguinte erro:

$P_E = P(\omega_1) - \int_{R1} [P(\omega_1|x) - P(\omega_2|x)]P(x)\, dx.$

O erro então será mínimo quando $P(\omega_1|x) > P(\omega_2|x)$ em R_1, caso contrário R_2 se torna a região.

Figura 1: Regiões de decisão.



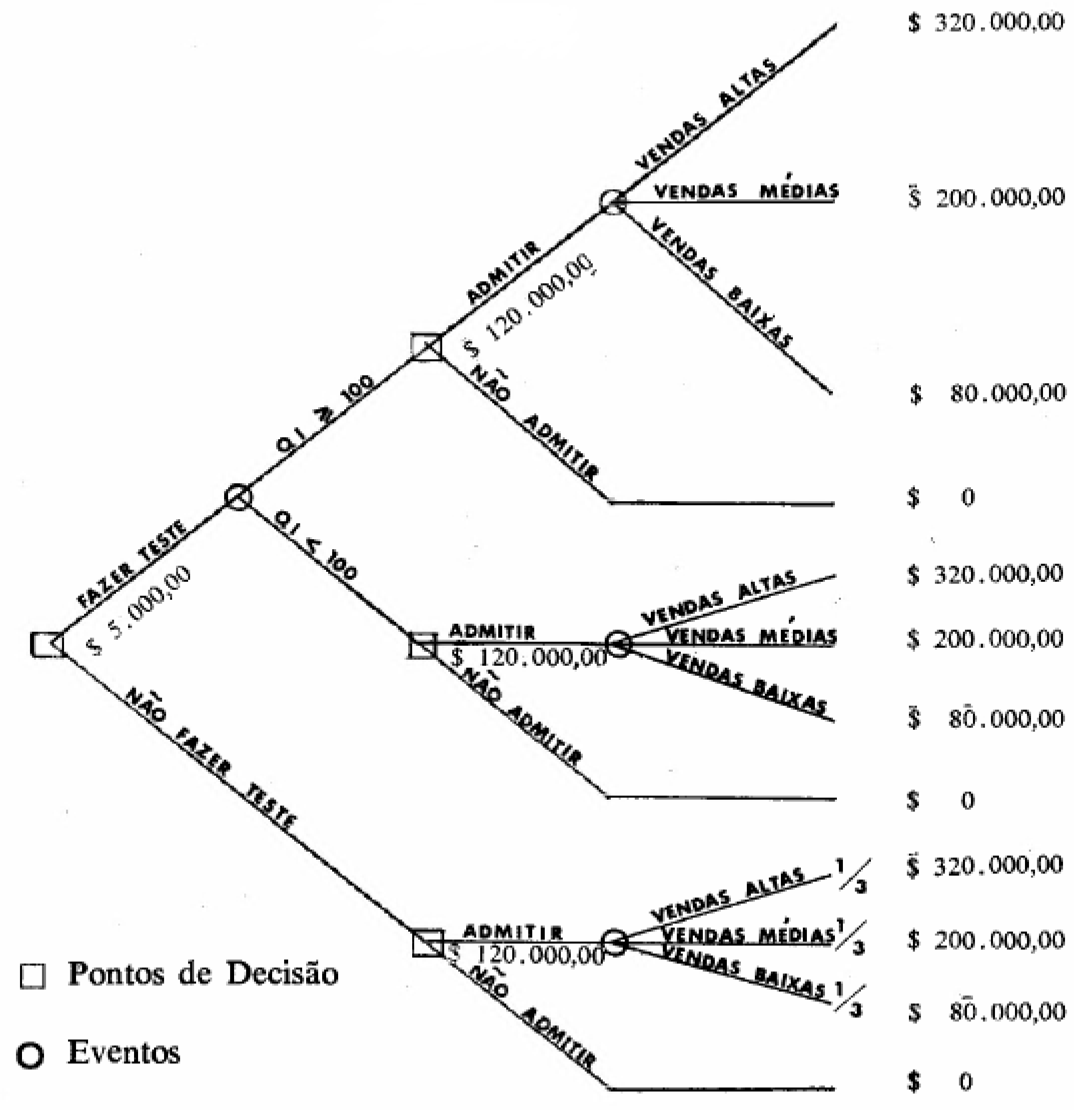
Fonte: Kinas e Andrade, 2017

Um exemplo de aplicação da Teoria da Decisão Bayesiana é o trabalho de Goldschmidt (1970), em que ele a utiliza para a análise de um problema empresarial e consequentemente uma tomada de decisão.

O problema tratado pelo autor é o seguinte: uma determinada empresa precisava preencher uma vaga de gerente de vendas em sua equipe e a diretoria da mesma se reuniu para decidir quais critérios deveriam ser considerados. Alguns membros da diretoria eram favoráveis à aplicação de um teste para o único candidato que tinham no momento para o cargo, mas outros eram contra, pois o custo disso seria muito alto. Por fim, ficou resolvido que, caso a decisão fosse a favor do teste, a admissão ou não do candidato iria depender do seu quociente de inteligência (QI). A empresa então apresentou uma série de dados históricos mostrando a relação entre número de vendas, períodos e QI dos funcionários, além de outras informações de custos, expectativa de vendas, etc.

Para solucionar esse problema por meio da Teoria da Decisão Bayesiana, o primeiro passo foi a construção de uma árvore de decisão, que na realidade não faz parte da teoria mencionada, mas de acordo com o autor, nesse caso se mostrou importante para melhor visualização de todo o contexto, conforme a Figura 2:

Figura 2: Árvore de decisão do problema mercadológico.



Fonte: Goldschmidt, 1970.

É possível notar que até mesmo os custos envolvidos e esperados em cada possibilidade também foram colocados na árvore. Porém, algumas informações ainda eram necessárias para a realização dos cálculos, como por exemplo a probabilidade das vendas serem altas com a aplicação prévia do teste dos candidatos, lembrando que ainda existe a condição de que o candidato tenha QI $\geq$ 100. Considerando as seguintes nomenclaturas:

VA - vendas altas;

VM - vendas médias.

Aplicando então a fórmula do T.B., tem-se:

$$P(VA|QI \geq 100) = \frac{P(QI{\geq}100|VA)P(VA)}{P(QI{\geq}100|VA)P(VA)+P(QI{\geq}100|VM)} \div P(VM) + P(QI \geq 100|VB)P(VB),$$

quando substituímos pelos valores informados, obtém-se:

$$P(VA|QI \geq 100) = \frac{1{\times}1/3}{(1{\times}1/3)+(0{,}60{\times}1/3)+(0{\times}1/3)} = 5/8,$$

sendo “×” a operação multiplicação.

De forma análoga:

$$P(VM|QI \geq 100) = \frac{0{,}6{\times}1/3}{(1{\times}1/3)+(0{,}60{\times}1/3)+(0{\times}1/3)} = 3/8,$$

$$P(VB|QI \geq 100) = \frac{0{\times}1/3}{(1{\times}1/3)+(0{,}60{\times}1/3)+(0{\times}1/3)} = 0.$$

Da mesma maneira é possível calcular a probabilidade de que as vendas sejam baixas, médias ou altas, dado que o candidato tenha o QI $<$ 100:

$P(VB|QI < 100) = 0;$

$P(VM|QI < 100) = 2/7;$

$P(VA|QI < 100) = 5/7.$

Por fim, o último passo consiste na avaliação de todas as alternativas obtidas a partir dos cálculos para que o decisor possa fazer a melhor escolha possível. Como ele está utilizando a Teoria da Decisão Bayesiana, o critério é escolher a alternativa que leve ao valor mais alto esperado.

O artigo em que está sendo baseada essa aplicação é do ano de 1970, por isso o padrão monetário utilizado na época era o Cruzeiro Novo (Ncr$). Levando em conta a Figura 2 e realizando os cálculos com a possibilidade de fazer o teste e admitir o candidato com QI ≥ 100, os custos podem ser calculados da seguinte forma:

$Ncr\$\,320.000 \times 5/8 + Ncr\$\,200.000 \times 3/8 + Ncr\$\,80.000 \times 0 = Ncr\$\,275.000,$ porém o custo de admissão do candidato é de Ncr$ 120.000, sendo assim restariam Ncr$ 155.000, comparando com a outra alternativa de não admitir o candidato (Ncr$ 0), o decisor deve escolher a opção de admitir o candidato. De forma análoga também foram feitos os cálculos de admitir um funcionário com QI < 100 e também de não fazer o teste, que são os outros eventos da árvore de decisão. Finalmente após todos os cálculos, a melhor alternativa encontrada foi a de não fazer o teste e admitir o candidato.

O autor finaliza o artigo concluindo que esse estudo, apesar de simples, apresenta quase todos os estudos possíveis através da aplicação da Teoria da Decisão Bayesiana. Ele também destaca outros tópicos da área empresarial que podem utilizar a mesma abordagem bayesiana, como problemas para determinar preços e problemas de pesquisa de mercado.

E é exatamente sobre essa decisão de preço em clima de incerteza que fala Motta (1997) em seu trabalho. Assim como Goldschmidt (1970), o autor fala sobre o quão complexo é para administradores estabelecer uma estratégia de estipulação de preço de um produto ou serviço, ainda mais quando existem muitas variáveis influenciando. Sendo assim, ele utilizou uma análise Bayesiana para lidar com esse tipo de problema.

Diferente de outros exemplos citados anteriormente sobre o uso do T.B., em que sua importância era a consideração de dados históricos, na área de negócios, a maior vantagem do T.B. é a possibilidade de calcular as probabilidades com muitas variáveis, pois os métodos de determinação de preço comumente utilizados levam em consideração apenas custos de fabricação, despesas de comercialização e de administração e financiamento das vendas, ignorando fatores como, variação de custos com as alterações das vendas, o papel que o preço desempenha nas decisões de compra, comportamento passado e ações prováveis dos concorrentes.

O T.B. então, contribui muito nesse processo, proporcionando aos tomadores de decisão um procedimento que lhes permitam desdobrar problemas de estipulação de preço em partes, para assim avaliar a incerteza de cada uma delas e no fim selecionar a estratégia que proporcione o maior lucro esperado.

Mudando um pouco o foco da área administrativa, a Teoria da Decisão Bayesiana também pode ser aplicada à medicina. Granato (2005) utilizou-a com o objetivo de auxiliar na avaliação das condições da laringe, que é muito importante na prevenção, diagnóstico e tratamento de quaisquer alterações vocais e laríngeas que possam ser detectadas, indicando a presença de uma doença. E até mesmo no acompanhamento da evolução do tratamento após o diagnóstico.

Para isso, foram extraídos sinais de voz diagnosticados e calculadas as probabilidades de que o paciente fosse portador de edema, fenda, nódulo, pólipo ou estivesse saudável. Em todas as etapas e combinações realizadas para os cálculos, obteve-se no mínimo 90% de acerto, embora a maioria dos resultados tenham sido de 100% de acerto. Isso mostrou que o uso da análise bayesiana é realmente uma ferramenta poderosa no auxílio aos especialistas.

Além disso, outros benefícios que podem ser citados no uso desse sistema são os baixos custos envolvidos, facilidade de uso, possibilidade de um pré-diagnóstico de patologias da laringe em estágios iniciais e a geração de uma base de dados clínica.

4.2 ALGORITMO DE NAIVE BAYES

De acordo com Melo (2007), o algoritmo de Naive Bayes possui uma abordagem completamente probabilística, afinal sua base é o T.B. e mais especificamente a Teoria da Decisão Bayesiana, abordada no Subcapítulo 4.1. É um dos classificadores mais utilizados na categorização de textos, como filtros de “spam”, mineração de emoções, separação de documentos, entre outros, isso porque ele é simples, rápido, fácil para implementar e muito eficiente, além de ser pouco afetado com a presença de ruídos e atributos irrelevantes. Por outro lado, o algoritmo não lida com interdependências e quando existem valores que são numéricos e contínuos em distribuições diferentes, é comum Naive Bayes não obter resultados tão bons quanto de outros métodos.

> Baseado na probabilidade condicional de determinadas palavras aparecerem em um documento o qual pertence a uma determinada categoria, esta técnica permite calcular as probabilidades de um novo documento pertencer a cada uma das categorias e atribuir a este as categorias de maior probabilidade. (LEWIS, RINGUETTE, 1994, apud MELO, 2007)

O nome Naive significa inocente, é utilizado por assumir que os atributos e as classes são condicionalmente independentes entre si, ou seja, que a ocorrência de um determinado evento não interfere na probabilidade de ocorrência de outro evento.

Entendendo melhor como funciona o algoritmo, basicamente ele possui documentos que são utilizados como um conjunto de treinamento, pois ele possui aprendizado indutivo, e cada um desses documentos é descrito por atributos que indicam a presença ou ausência dos termos a_1, a_2,..., a_n, então o classificador atribui a cada um dos documentos a categoria mais provável, isso é feito através de uma função f que devolve valores (categorias) pertencentes a um conjunto finito V. Ou seja, o classificador considera que a probabilidade de ocorrência de uma conjunção de atributos em um dado exemplo é igual ao produtório das probabilidades de ocorrência de cada atributo isoladamente:

$P(y_i|X) \propto P(y_i) \prod_{j=1}^{m} P(x_j|y_i)$, onde:

- $P(y_i|X)$ é a probabilidade dos dados (x) corresponderem à i-ésima classe (y_i);
- $P(y_i)$ é a probabilidade *a priori* da i-ésima classe no conjunto de dados, e
- $P(x_j|y_i)$ é a probabilidade de ocorrer o valor do j-ésimo atributo, dada a ocorrência i-ésima classe.

De maneira mais prática, suponha que deseja-se determinar se uma pessoa está ou não gripada, ou seja, um problema de classificação com duas classes. Sendo que as variáveis explicativas seriam coriza, tosse, febre, dor muscular e dor de garganta. A partir de Naive Bayes é possível realizar um produtório de cada uma das probabilidades condicionais, ou seja, qual a probabilidade de ter coriza dado que tem ou não tem gripe, vezes a probabilidade de ter tosse dado que tem ou não tem gripe e assim por diante, ficando muito mais fácil de se calcular. Depois de obter os resultados de cada classe, é selecionada a classe com a maior probabilidade, é nesse momento que entra também a Teoria da Decisão Bayesiana.

Nos parágrafos abaixo está descrito um pouco de como Melo (2007) aplicou o algoritmo de Naive Bayes como uma das abordagens do problema de reconhecimento de padrões textuais aplicada ao processo de categorização automática de documentos.

A primeira etapa foi a seleção de características, em que os atributos, nesse caso as palavras, foram avaliadas para realizar uma primeira classificação daquele texto. Para isso, são removidos sinais de pontuação e palavras que não agregam à classificação, como as preposições e artigos, também foi feita uma formatação do texto para obter uma representação estruturada, além do cálculo de relevância para identificar os termos mais importantes e também o grau de correlação entre uma palavra a um documento. Esses são apenas alguns dos processos.

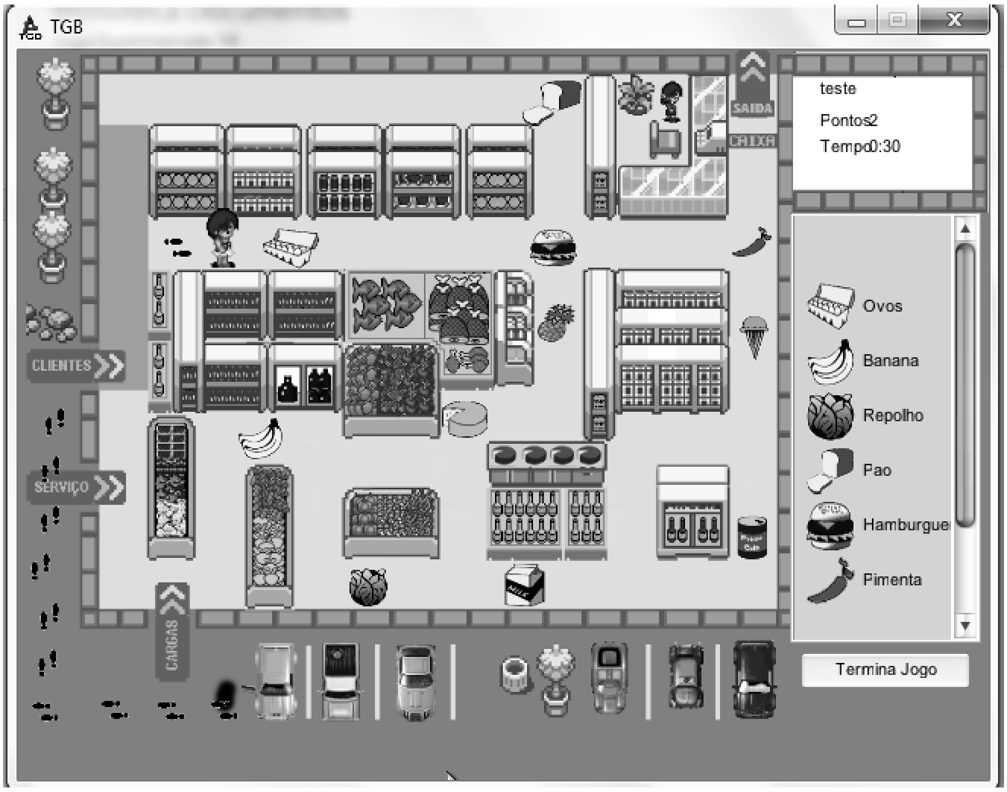
A segunda etapa já é a classificação, feita através do reconhecimento de padrões. Foram utilizados dois conjuntos de textos para essa etapa, o primeiro advindo de artigos de jornais com temas como esportes, imóveis, política, etc. Já o segundo conjunto, composto de textos com título e resumo de dissertações de mestrado e teses de doutorados de uma universidade, com temas como microeletrônica, processamento de sinais, controle, entre outros. Todos esses textos passaram então pelo processo de seleção de características e também foram divididos para treino e testes da ferramenta de aprendizado. As ferramentas utilizadas foram várias, algumas de autoria própria, outras já disponíveis no mercado.

Para os experimentos foi utilizado o algoritmo de Naive Bayes. Infelizmente todas as tabelas que apresentam os resultados obtidos foram borradas, então não é possível realizar uma interpretação dos mesmos, mas a autora conclui que essa proposta foi muito eficiente e pode contribuir para a classificação automática de textos em português. Quando comparado a outro método, Naive Bayes se destacou mais, tanto pelos resultados obtidos, quanto pela sua rapidez, atingindo algo próximo a 95% de acerto.

Outro estudo bastante interessante com o uso de Naive Bayes é na área da medicina, mais especificamente, psiquiatria. O trabalho de Bastos et al (2012) faz a utilização de um jogo e do algoritmo para auxiliar na avaliação cognitiva do Transtorno de Déficit de Atenção/Hiperatividade (TDAH) em crianças e adolescentes. Os autores ressaltam que essa ferramenta não substituiria a avaliação clínica do médico, seria apenas um facilitador do diagnóstico.

Para isso, inicialmente foi aplicado aos participantes um jogo chamado Jogo do Supermercado, o qual foi construído exatamente com o objetivo de auxiliar na avaliação do TDAH. Na Figura 3 é possível observar uma das telas do jogo. Basicamente, ele apresenta dois modos, em um deles os itens da lista de compras podem ser adquiridos em qualquer ordem, mas o jogador deve executar tudo no menor tempo possível, sem quebrar regras. Ou seja, nesse modo é observada a capacidade de planejamento. Já no outro modo, os itens da lista devem ser adquiridos na ordem exata que aparecem e também no menor tempo possível. Seu objetivo é avaliar a capacidade de execução do jogador.

Figura 3: Jogo do Supermercado



Fonte: Bastos et al, 2012.

Além do jogo, também foram aplicados questionários para a obtenção de mais informações, como a quantificação dos sintomas, estabelecimento de perfis estereotipados e um manuseamento facilitado. Tanto os dados obtidos no jogo, quanto dos questionários respondidos pelos pais das crianças e professores, foram armazenados e analisados cuidadosamente.

Sobre o algoritmo, os autores mencionam rapidamente que o classificador Naive Bayes foi utilizado durante os testes, com e sem filtros, e com combinações de atributos para obter os melhores resultados e que nos poucos casos em que isso não ocorreu, o motivo claramente era a quantidade pequena de dados, o que acabou prejudicando o aprendizado do algoritmo.

Por fim, Bastos et al (2012) falam que os resultados foram razoavelmente satisfatórios quando são comparados os resultados obtidos pelo jogo e pelos questionários, portanto a aplicação do Jogo do Supermercado pode ser considerado um instrumento eficaz na avaliação cognitiva do TDAH.

4.3 NA EDUCAÇÃO BÁSICA

O T.B. quando visto pela primeira vez pode parecer complexo, principalmente pelos pré-requisitos que são necessários de se aprender para entendê-lo por completo, entretanto ainda assim é possível aplicá-lo na educação básica. Segundo Ferreira (2016), o T.B. se adequa aos Parâmetros Curriculares Nacionais para o Ensino Médio, abrangendo as seguintes competências e habilidades definidas para a investigação e compreensão:

> Desenvolver a capacidade de questionar processos naturais e tecnológicos, **identificando regularidades**, apresentando interpretações e **prevendo evoluções**.
> Desenvolver o raciocínio e a capacidade de aprender.
> • Formular questões a partir de situações reais e compreender aquelas já enunciadas.
> • Desenvolver modelos explicativos para sistemas tecnológicos e naturais.
> • **Utilizar instrumentos de medição e de cálculo**.
> • Procurar e sistematizar informações relevantes para a compreensão da situação-problema.
> • Formular hipóteses e **prever resultados**.
> • Elaborar estratégias de enfrentamento das questões.
> • Interpretar e criticar resultados **a partir de experimentos e demonstrações**.
> • Articular o conhecimento científico e tecnológico numa perspectiva interdisciplinar.
> • Entender e aplicar métodos e procedimentos próprios das Ciências Naturais.
> • **Compreender o caráter aleatório e não determinístico dos fenômenos naturais e sociais e utilizar instrumentos adequados para medidas, determinação de amostras e cálculo de probabilidades**.
> • Fazer uso dos conhecimentos da Física, da Química e da Biologia para explicar o mundo natural e para planejar, executar e avaliar intervenções práticas.
> • Aplicar as tecnologias associadas às Ciências Naturais na escola, no trabalho e em outros contextos relevantes para sua vida. (PCN, 2000, p. 12, grifo de Ferreira, 2016)

Além disso, o T.B. é utilizado em diversas áreas de conhecimento, sendo assim pode ser aplicado como conteúdo complementar para o estudo das probabilidades, além de estimular a criatividade e raciocínio lógico.

> O conhecimento matemático deve estar englobado em todo processo de ensino aprendizagem, porque ele é um elo que trabalha a interdisciplinaridade, este fato não é só observado nas matérias exatas, com em todas as outras áreas de ensino, isso porque, quando o educando possui bom raciocínio matemático, ele terá uma visão mais ampla de todo processo de aprendizagem. A matemática na Educação Básica é de suma importância para o desenvolvimento intelectual do aluno, não apenas mais uma “matéria”, mais uma disciplina presente no currículo. (Salomão, 2014)

Na sequência serão apresentados alguns exemplos de aplicações do T.B. em sala de aula que foram encontrados na literatura.

O autor Ferreira (2016) trabalhou com o objetivo de “propor o estudo do Teorema de Bayes para as séries finais da educação básica brasileira, através do aprendizado de sua teoria e aplicações em sala de aula”. Inicialmente na parte teórica, ele apresentou o conteúdo de probabilidade que é pré requisito para a aprendizagem do T.B., como experimentos aleatórios, eventos, espaços amostrais, definições de probabilidade, probabilidade da união de eventos, probabilidade da intersecção de eventos, probabilidade condicional e teorema da probabilidade total, apenas depois partiu para o T.B. e sua história. Para facilitar a compreensão dos alunos, o autor utilizou muitos exemplos práticos. Sua conclusão foi que a aplicação desse tema como conteúdo complementar de probabilidade é possível e proporciona uma gama maior de conhecimentos, além de estimular a criatividade e raciocínio.

Salomão (2014) por sua vez, visou estudar sobre o paradoxo de Monty Hall e a matemática nele contido, contando que seus aspectos lúdicos poderiam ser traduzidos como uma oportunidade de cativar o educando para o estudo das probabilidades. Sendo assim, ele também desenvolveu aplicações do T.B. na educação básica através de atividades em formas de jogos e recursos computacionais, seu objetivo era atrair a atenção dos alunos para o ensino de probabilidade a partir de aspectos lúdicos. Assim como Ferreira (2016), ele também abordou anteriormente os conteúdos de pré-requisitos para o T.B.. Ao final, Salomão (2014) apresentou conclusões apenas sobre os resultados teóricos das atividades realizadas e sugestões para próximos trabalhos, não falou sobre a eficácia dessa abordagem para a educação.

4.3.1 Monty Hall

Um dos problemas citados foi o de Monty Hall, que era o nome de um apresentador de um famoso programa de televisão americano, chamado “Let’s Make a Deal". Nesse programa os participantes tinham à sua disposição três portas, sendo que atrás de uma delas havia um carro como prêmio e nas outras duas haviam bodes. O problema iniciava quando o participante escolhia uma das portas, o apresentador abria uma segunda porta com o bode e então questionava ao participante se ele queria trocar sua porta escolhida inicialmente pela terceira que restou. Acontece que a pessoa participante sempre achava que se não mudasse sua escolha nesse momento, sua chance de acertar e levar o prêmio aumentava de 33,33% para 50%, mas na verdade ao escolher uma das portas, quando havia três alternativas, sua chance de acertar era de 33,33% e ao trocá-la quando o apresentador restringia as opções para apenas duas portas, sua chance passaria a ser de 66,66%, pois essa era a probabilidade das portas que o participante não escolheu no início.

Porém, até chegar a essa conclusão muitos estudiosos e matemáticos tentaram entender o problema, se envolveram em polêmicas e por fim só foi possível comprovar tal hipótese quando houve o avanço da tecnologia, então com a ajuda de computadores e através do T.B., certificou-se que as probabilidades estavam corretas.

O autor Ferreira (2016), se baseou na discussão ao redor do problema de Monty Hall para levar o tema até a sala de aula. Ele adaptou a situação com copos ao invés de portas e bombons como premiação, os alunos fizeram o papel de participantes. O professor como apresentador construiu uma tabela na lousa e conforme cada aluno ia jogando, ele anotava as escolhas feitas e o resultado obtido. Isso foi feito repetidas vezes até que os alunos notassem que quem optou por trocar o primeiro copo escolhido acertou mais do que aqueles que mantiveram a escolha inicial.

Por fim, o autor trouxe duas maneiras de apresentar a solução do problema. A primeira é através da árvore de possibilidades, a qual demonstra todos os caminhos possíveis em uma combinação dos caminhos obtidos, mas que para a finalidade do presente trabalho está fora do escopo, portanto não serão abordados os detalhes.

A outra solução pode ser obtida através do T.B. Sejam os seguintes eventos:

- PA: o prêmio está em A,
- PB: o prêmio está em B,
- PC: o prêmio está em C,
- RA: o apresentador revela o copo (vazio) de A,
- RB: o apresentador revela o copo (vazio) de B,
- RC: o apresentador revela o copo (vazio) de C.

Primeiramente considera-se que o aluno escolhe o copo A e não o troca quando o professor revela que o prêmio não está no copo B. A probabilidade de interesse passa a ser $P(PA|RB)$. Calculando através do T.B. tem-se:

$$P(PA|RB) = \frac{P(RB|PA) \cdot P(PA)}{P(RB|PA) \cdot P(PA) + P(RB|PB) \cdot P(PB) + P(RB|PC) \cdot P(PC)},$$

sendo que,

$P(RB|PA) = \frac{1}{2}$, pois o professor pode escolher o copo B ou C,

$P(RB|PB) = 0$, pois o professor nunca levanta o copo onde está o prêmio,

$P(RB|PC) = 1$, como o professor nunca levanta o copo onde está o prêmio, então só resta mostrar o copo B, já que o A foi escolhido pelo aluno e no C está o prêmio.

A partir desses dados é obtido o seguinte resultado:

$$P(PA|RB) = \frac{\frac{1}{2} \cdot \frac{1}{3}}{\frac{1}{2} \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}} = \frac{1}{3} .$$

Em um segundo caso, considera-se que o aluno escolhe o copo A e quando questionado pelo professor, aceita a troca pelo copo C. Calculando novamente as probabilidades através do T.B. tem-se:

$$P(PC|RB) = \frac{P(RB|PC) \cdot P(PC)}{P(RB|PA) \cdot P(PA) + P(RB|PB) \cdot P(PB) + P(RB|PC) \cdot P(PC)} \Rightarrow$$

$$\Rightarrow P(PC|RB) = \frac{1 \cdot \frac{1}{3}}{\frac{1}{2} \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}} = \frac{2}{3} .$$

Dessa forma, é possível concluir que no segundo caso, em que o aluno trocou o copo escolhido inicialmente, a chance de acertar o copo em que estava o prêmio era maior.

O autor Salomão (2014) também trabalhou na educação básica com o T.B. utilizando o paradoxo de Monty Hall, um de seus objetivos era desvincular a Matemática como uma sequência de aplicações de regras e mostrar aos alunos que é possível aplicá-la na vida prática, além de ensinar por meio de jogos, de forma lúdica e que causa grande interesse nos alunos. Ele destaca que o professor nesse momento além de conduzir a situação e experimento, também terá as funções de ouvir o aluno, confrontar seu ponto de vista, levar o aluno a criar suas próprias estratégias e incentivar a discussão saudável em sala de aula. A preocupação dessa atividade não deve ser apenas a parte teórica em si, mas sim a experimental, oferecendo aos alunos situações explícitas de tomadas de decisões.

Salomão (2014) aplicou as atividades de Monty Hall com turmas de 9º ano do ensino fundamental e 2º ano do ensino médio. Para o 9º ano ele tomou cuidado com formalismos, pois alunos desse nível ainda se encontram em fase de desenvolvimento cognitivo, já os do 2º ano podem assimilar bem os formalismos e alguns possuem habilidades mais avançadas tanto matemáticas, quanto computacionais.

Ele realizou duas atividades diferentes baseadas no paradoxo e instrui que, caso um outro professor vá realizar a mesma atividade, entre no personagem do apresentador para tornar a situação mais divertida e consequentemente atrativa, além disso, ele também diz que o professor não deve mencionar o nome Monty Hall para evitar que algum aluno faça pesquisas no celular e estrague o jogo e também que deixe os próprios alunos perceberem e questionarem se há alguma diferença entre manter a escolha do primeiro copo ou trocá-lo.

A primeira atividade é praticamente idêntica à situação original e também feita por Ferreira (2016), ou seja, o professor troca as portas por copos e o carro de prêmio por um bombom. Um a um vai chamando os alunos a participarem do desafio. Já na segunda atividade ele pega três copos e três bolinhas, duas pretas e uma dourada, divide a turma em grupos de seis alunos e entrega um “kit” de copos e bolinhas para cada grupo. Da mesma forma, os alunos vão se revezando e fazendo o desafio. Porém, dessa vez eles recebem folhas com tabelas destinadas ao preenchimento dos resultados obtidos nos jogos, para que no final o professor consiga fazer cálculos com eles. O autor também sugere que essa atividade com as bolinhas seja feita com 4 e 5 copos, pois assim, os alunos conseguem uma amostragem maior e fica mais fácil perceber a tendência de que a troca de copo oferece maiores chances de ganhar o prêmio. Claro que a essa altura o professor já deve estimular esse tipo de discussão, caso não tenha surgido naturalmente.

Muitas outras atividades foram desenvolvidas por Salomão (2014), uma delas utilizando simulação computacional, que exige um conhecimento prévio dos alunos em relação a linguagem de programação, outras foram exemplos derivados do jogo original e mudando apenas alguns pontos para que os alunos pudessem ir percebendo a convergência para a resposta correta de que, trocar de porta quando o participante é questionado aumenta sim suas probabilidades de ganhar o prêmio. Para isso, o professor utilizou também a árvore de decisões como um importante recurso visual.

Tanto nas atividades desenvolvidas por Ferreira (2016), quanto por Salomão (2014), é possível observar que a forma como eles as conduzem coincide com a metodologia de resolução de problemas. Um dos autores clássicos com relação à essa temática é Polya, que propõe a seguinte definição:

> [...] a Resolução de Problemas corresponde a um modo de organizar o ensino o qual envolve mais que aspectos puramente metodológicos, incluindo uma postura frente ao que é ensinar e, conseqüentemente, do que significa aprender (2001, p.89).

Polya (1997) preconiza o ensino ativo da Matemática. Segundo ele, para tornar o estudante um ‘resolvedor’ de problemas é necessário seguir determinada ordem na qual o concreto se apresenta antes do abstrato, a ação e a percepção antes das palavras e conceitos e, os conceitos antes dos símbolos.

Ao professor cabe propor atividades problemáticas calcadas na realidade cotidiana dos estudantes, primeiramente baseadas no concreto. Posteriormente, cabe formalizar os conceitos e, finalmente, vem o uso da simbologia matemática.

Alguns estudos (POZO, 1998; DINIZ & SMOLE, 2001) já propõem caminhos a serem seguidos, como o uso da proposição de problemas que seriam o disparador do conhecimento e não seriam usados somente como instrumento para repetição de uma estratégia de solução que aparentemente é única.

Sendo assim, é possível notar que na aplicação do problema de Monty Hall, tanto no caso de Ferreira (2016), quanto de Salomão (2014), o único processo que se distancia da metodologia de resolução de problemas é o fato de que ambos introduziram o conteúdo teórico sobre o T.B. antes de abordar o problema em si.

Mesmo assim, é perceptível que os alunos não associaram a resolução da questão com a equação do T.B., pois a forma com que foi mencionado o problema de Monty Hall foi bastante natural, não os deu informações para simplesmente seguirem um passo a passo. Foi necessário um longo exercício, muita observação e discussão em grupos para que chegassem a uma conclusão e somente depois o professor demonstrasse a resolução via T.B.

4.3.2 Desafio do falso-positivo

A segunda proposta de ensino do T.B. do autor Ferreira (2016) aos seus alunos foi com o desafio do falso-positivo. Foi passada para eles uma situação em que 1% da população de uma cidade estava infectada por uma doença, um morador fez o teste para a mesma e recebeu resultado positivo, sabendo que teste acerta 99% das vezes, os alunos deveriam calcular a probabilidade do morador não estar realmente infectado, caso o resultado do exame seja um falso-positivo.

Quando questionados, antes de qualquer cálculo e apenas se baseando nas informações dadas pelo enunciado, os alunos palpitaram que a chance do indivíduo ter a doença era maior que 99%. Partindo então para a resolução baseada no T.B., tem-se:

D – o evento ter a doença;

N – o evento não ter a doença;

T – o evento TESTE dar positivo.

Tabela 1: Quadro de dados do desafio do falso-positivo

TIPOS DE PROBABILIDADE	TEM A DOENÇA	NÃO TEM A DOENÇA
Probabilidade *a priori*	0,01 ou 1%	0,99 ou 99%
Probabilidade Condicional (teste estar certo por ter dado positivo)	0,99 ou 99%	0,01 ou 1%

Fonte: Ferreira, 2016.

$$P(N|T) = \frac{P(T|N)\cdot P(N)}{P(T|D)\cdot P(D)+P(T|N)\cdot P(N)} = \frac{0{,}01\cdot 0{,}99}{0{,}99\cdot 0{,}01+0{,}01\cdot 0{,}99} = 0{,}5.$$

Conclui-se então que a probabilidade do morador estar realmente infectado pela doença é de 50%, resultado que segundo o autor surpreendeu os alunos, o que é muito bom por gerar curiosidade no processo de ensino aprendizagem.

4.3.3 Cálculo da assertividade

Por fim, Ferreira (2016) desenvolveu uma terceira atividade, parecida com a anterior, mas agora trata-se de uma mulher que fez uma mamografia e recebeu resultado positivo para neoplasia, ou seja, um câncer de mama. O objetivo é descobrir qual a probabilidade de que o resultado do exame esteja incorreto. Para coletar maiores dados do passado que apoiassem os cálculos, o professor também passou um texto com as seguintes informações:

> No mundo, o risco cumulativo de uma mulher, com idade de 74 anos, ter apresentado esse tipo de câncer durante a vida é de 4,62%. Risco cumulativo é a razão entre o número de casos novos de uma doença e o total da população sob risco.
> A sensibilidade (probabilidade de confirmar a presença ou ausência de patologia em um indivíduo) da mamografia é de 85%, mas em 9,6% dos casos ele acusa resultados positivos para mulheres que não possuem a doença. (Ferreira, 2016, apud Diretrizes para a Detecção Precoce do Câncer de Mama no Brasil, 2015)

Antes de iniciar os cálculos é importante deixar os alunos refletirem e darem seus palpites. Outra informação relevante que deve ser lembrada é que ter e não ter o câncer de mama são eventos mutuamente excludentes, ou seja, não podem acontecer simultaneamente em nenhuma ocasião. Tem-se então a tabela 2:

Tabela 2: Quadro de dados do cálculo da assertividade 1.

TIPOS DE PROBABILIDADE	TEM CÂNCER	NÃO TEM CÂNCER
Probabilidade *a priori*	0,0462 ou 4,62%	0,9538 ou 95,38%
Probabilidade Condicional	0,85 ou 85%	0,096 ou 9,6%
Probabilidade Conjunta	$0{,}0462 \cdot 0{,}85 = 0{,}0393$	$0{,}9538 \cdot 0{,}096 = 0{,}0916$
Probabilidade *a posteriori*	$0{,}0393/0{,}1309 = 0{,}3002$	$0{,}0916/0{,}1309 = 0{,}6998$

Fonte: Ferreira, 2016.

Os dados da primeira e segunda linhas da Tabela 2 foram retirados do texto apresentado anteriormente. A probabilidade conjunta foi obtida multiplicando a probabilidade *a priori* pela condicional. Por fim, é possível notar que a soma das probabilidades *a priori* é igual a 1, mas o mesmo não ocorre com as probabilidades conjuntas. Nesse caso, usa-se o fator de normalização que é a divisão de cada probabilidade pela soma dos valores das probabilidades conjuntas, então chega-se à soma 1 que significará 100%. Esse cálculo representa a probabilidade *a posteriori*.

Partindo para os cálculos através do T.B. tem-se:

C – o evento ter câncer;

N – o evento não ter câncer;

M – o evento exame de mamografia dar positivo.

$$P(N|M) = \frac{P(M|N)\cdot P(N)}{P(M|C)\cdot P(C)+P(M|N)\cdot P(N)} = \frac{0{,}096\cdot 0{,}9538}{0{,}85\cdot 0{,}0462+0{,}096\cdot 0{,}9538} = 0{,}6998.$$

Ou seja, a paciente tem 69,98% de chances de não ter o câncer, contra 30,02% de chances de ter.

No campo da medicina esse raciocínio bayesiano é muito utilizado para detectar falsos-positivos em exames de diversos tipos. Na atualidade pode-se citar por exemplo os exames de coronavírus. Em sala de aula tanto essa, quanto a questão anterior são muito interessantes, pois tudo que existe um contexto próximo à realidade chama mais a atenção dos alunos, estimulando a curiosidade matemática. Além disso, é uma excelente oportunidade para uma atividade interdisciplinar, inclusive em campanhas de saúde.

5 CONSIDERAÇÕES FINAIS

O presente trabalho de conclusão de curso visou apresentar o Teorema de Bayes por meio de algumas aplicações, para isso percorreu um longo caminho desde os conceitos básicos da probabilidade, até a história do surgimento do mesmo, seus conceitos e demonstrações, para alcançar o objetivo de finalmente abordar a Teoria da Decisão Bayesiana, o algoritmo de Naive Bayes e como aplicar o T.B. na educação básica.

A partir da revisão bibliográfica realizada, não foi possível identificar nenhum trabalho anterior que resumisse tais aplicações, apenas trabalhos que tratam uma ou outra aplicação de forma separada. Sendo assim, pode-se dizer que este é um tema inédito e de extrema importância, pois foi possível conhecer e comparar diferentes modos de se utilizar o T.B. em situações diversas.

A principal dificuldade foi encontrar variedade de materiais para o estudo da parte teórica das aplicações, em destaque a Teoria da Decisão Bayesiana e Naive Bayes, pois apesar de existirem vários artigos de uso das técnicas, raramente são demonstradas matematicamente. O acesso a livros infelizmente está mais restrito devido a pandemia de Coronavírus e na internet também não foi tão fácil encontrá-los, a maioria em outras línguas.

Contrapondo cada uma das aplicações apresentadas através também das pesquisas realizadas ao redor delas, foi possível verificar que a Teoria da Decisão Bayesiana, apesar de muito relevante até mesmo em nosso cotidiano de forma intuitiva, por ser uma formalização do senso comum, já não é mais tão citada em trabalhos atuais.

O algoritmo de Naive Bayes, por sua vez, está em alta, principalmente porque a inteligência artificial está sendo cada vez mais utilizada e é muito recente. Mesmo surgindo outros algoritmos com o mesmo intuito, Naive Bayes continua se mostrando necessário.

Por fim, a aplicação do T.B. na educação básica traz grande contribuição para os docentes que buscam incessantemente inovar no processo de ensino aprendizagem de Matemática, sendo um conteúdo que apesar de não ser obrigatório no currículo, pode ser abordado como um extra pelos professores, aproveitando inclusive o uso de metodologias ativas em sala de aula.

Considera-se que tanto o objetivo geral, quanto os objetivos específicos foram cumpridos. Um ponto importante observado é que as aplicações tornam o T.B. mais compreensível, pois o conceito em si do mesmo é de certa forma abrangente e bastante subjetivo, mas quando são demonstradas as maneiras em que ele pode ser aplicado, se torna algo palpável e de fácil visualização em nosso cotidiano.

Para trabalhos futuros, fica a sugestão de maior aprofundamento na parte de estatística bayesiana e frequentista, principalmente em relação aos métodos de simulação Monte Carlo e Monte Carlos Via Cadeias de Markov (MCMC). Além também de explorar outros conceitos e aplicações do T.B., como a epistemologia bayesiana.

REFERÊNCIAS

BASTOS, A. P. Z. et al. Utilização de um Jogo Sério e Naïve Bayes para Auxiliar na Avaliação Cognitiva do Transtorno de Déficit de Atenção/Hiperatividade. **Anais do 23º Simpósio Brasileiro de Informática na Educação**, Rio de Janeiro, 26-30, nov. 2012. Disponível em: <https://www.br-ie.org/pub/index.php/sbie/article/view/1804>. Acesso em: 02 mar. 2022.

BUSSAB, W. O.; MORETTIN, P. A.. **Estatística Básica**. 6 ed. São Paulo: Saraiva, 2010.

DELIBERAL, J. P.. FIGUEIRA, M. V.. **Aplicabilidade do Teorema de Bayes no Monitoramento de Redes Sociais**. XXI Mostra de Iniciação Científica, Pós-Graduação, Pesquisa e Extensão, Universidade de Caxias do Sul, 2013. Disponível em: <http://www.ucs.br/etc/conferencias/index.php/mostraucsppga/mostrappga2013/paper/viewFile/3639/1086>. Acesso em: 06 jan. 2022.

FERREIRA, A. M. A.. **Conhecendo o Teorema de Bayes**: Sugestão de aplicação ao Ensino Médio. Dissertação (Mestrado em Matemática) - Universidade Federal do Tocantins. Palmas, p. 43, 2016. Disponível em: <https://sucupira.capes.gov.br/sucupira/public/consultas/coleta/trabalhoConclusao/viewTrabalhoConclusao.jsf?popup=true&id_trabalho=4412184>. Acesso em: 05 jan. 2022.

FLORÊNCIO, P. H. B.; NETO, A. S. S.; DANTAS, M. J. P.. **Análise do problema de Monty Hall**: um enfoque bayesiano. Simpósio Acadêmico de Engenharia de Produção, Universidade Federal de Viçosa, 2014. Disponível em: <https://www.saepro.ufv.br/wp-content/uploads/2014.13.pdf>. Acesso em: 03 jan. 2022.

GOLDSCHMIDT, P. C.. A Teoria da Decisão Bayesiana na Estratégia Mercadológica. **Revista de Administração de Emprêsas**, Rio de Janeiro, 10(1): 65/77, 1970. Disponível em: <https://www.scielo.br/j/rae/a/RN3VwftGjS6fkn74fYCDgPt/?lang=pt>. Acesso em: 03 jan. 2022.

GRANATO, L. F.. **Teoria Bayesiana na avaliação das condições da laringe**. Dissertação (Graduação em Engenharia Elétrica) - Universidade de São Paulo. São Carlos, 2005. Disponível em: <https://teses.usp.br/teses/disponiveis/18/18133/tde-29022016-085426/publico/Tese_Granato_LuisF.pdf>. Acesso em: 01 mar. 2022.

KINAS, P. G; ANDRADE, H. A.. **Introdução à Análise Bayesiana (Com R)**. 1ª ed. Consultor Editorial, 4 mai. 2017.

MARTINEZ, E. Z. et al. **Modelagem Bayesiana do risco de infecção tuberculosa para estudos com perdas de seguimento**. Revista de Saúde Pública, Ribeirão

Preto, 42(6), p. 999-1004, jun. 2008. Disponível em: <https://www.scielo.br/j/rsp/a/Fvhzs6Hc44p4HpH5zBBkw7c/?lang=pt>. Acesso em: 29 dez. 2021.

MELO, L. B. S.. **Reconhecimento de Padrões Textuais para Categorização Automática de Documentos**. Dissertação (Mestrado em Ciências) - Universidade Federal do Rio de Janeiro. Rio de Janeiro, p. 74, 2007. Disponível em: <http://www.pee.ufrj.br/index.php/pt/producao-academica/dissertacoes-de-mestrado/2007-1/2007122001-2007122001/file>. Acesso em: 04 jan. 2022.

MEYER, P. L. **Probabilidade - Aplicações à Estatística**. São Paulo: Livros Técnicos e Científicos Editora (LTC), 2000.

MOTTA, J.. **Decisões de Preço em Clima de Incerteza**: Uma Contribuição da Análise Bayesiana. Revista de Administração de Empresas, São Paulo, v. 37, n. 2, p. 31-46, abr./jun 1997. Disponível em: <https://www.scielo.br/j/rae/a/dWNQVsHjjCXwLSjFxPVp5yR/?format=pdf&lang=pt>. Acesso em: 01 mar. 2022.

OLIVEIRA, A. D.. **Aplicação da Estatística Bayesiana ao Risco de Crédito**. Dissertação (Mestrado em Matemática Financeira) - Lisboa School of Economics & Management. Lisboa, 2014. Disponível em: <https://www.repository.utl.pt/handle/10400.5/7712>. Acesso em: 29 dez. 2021.

OLINTO, M. T. A.; MOREIRA-FILHO, D. C.. **Fatores de risco e preditores para o aborto induzido**: estudo de base populacional. Caderno de Saúde Pública, Rio de Janeiro, 22(2), p. 365-375, fev. 2006. Disponível em: <https://www.scielo.br/j/csp/a/jxCsd6GVqJNpkJnZ4XFtG8B/?lang=pt#:~:text=Abortos%20induzidos%20estiveram%20fortemente%20relacionados,n%C3%BAmero%20maior%20de%20m%C3%A9todos%20contraceptivos.>. Acesso em: 29 dez. 2021.

O'CONNOR, J. J.; ROBERTSON, E. F. **Thomas Bayes**. Math History, 2004. Disponível em: <https://mathshistory.st-andrews.ac.uk/Biographies/Bayes/> Acesso em: 02 out. 2021.

PALMA, D. C. A.; SANTOS, E. S.; IGNOTTI, E.. Análise dos padrões espaciais e caracterização dos suicídios no Brasil entre 1990 e 2015. **Cadernos de Saúde Pública**, 36 (4), 2020. Disponível em: <https://www.scielo.br/j/csp/a/qwjLzp7Ttv8mNMpHMzLs5xP/abstract/?lang=pt>. Acesso em: 06 jan. 2022.

PENA, S. D.. **Thomas Bayes**: O Cara!. Ciência Hoje, 2009. Disponível em: <https://cienciahoje.org.br/artigo/thomas-bayes-o-cara/>. Acesso em: 04 out. 2021.

POLYA, George. **A arte de resolver problemas**. Tradução Heitor Lisboa de Araújo. Rio de Janeiro, 2006.

POZO, J. I. (org.). **A solução de problemas**: aprender a resolver, resolver para aprender. Porto Alegre: ArtMed, 1998.

RODRIGUES, F. M.. Estatística Bayesiana aplicada à genética. **Estudos Vida e Saúde**, Goiânia, v. 37, n. 3/4, p. 309-317, mar./fev. 2010. Disponível em: <http://seer.pucgoias.edu.br/index.php/estudos/article/view/1482>. Acesso em: 03 jan. 2022.

ROSSATO, F. C. et al. **Teorema de Bayes**: Estudo e Aplicação. Salão do Conhecimento, Universidade Regional do Noroeste do Estado do Rio Grande do Sul, 2016. Disponível em: <https://publicacoeseventos.unijui.edu.br/index.php/salaoconhecimento/article/view/6858/5625>. Acesso em: 18 jan. 2022.

SALOMÃO, M. S.. **Estudo e Generalizações do Paradoxo de Monty Hall na Educação Básica**. Dissertação (Mestrado em Matemática) - Instituto de Ciências Exatas, Universidade Federal de Juiz de Fora. Juiz de Fora, p. 69, 2014. Disponível em: <https://sucupira.capes.gov.br/sucupira/public/consultas/coleta/trabalhoConclusao/viewTrabalhoConclusao.jsf?popup=true&id_trabalho=2197224>. Acesso em: 05 jan. 2022.

SEVERINO, A. J. **Metodologia do Trabalho Científico**. São Paulo, SP: Cortez, 2007

SILVA, J. M.; SICHMAN, J. S.; CUGNASCA, P. S.. **Um Estudo da Aplicação de Modelo "Bayes-Fuzzy" para a Previsão de um Mercado Financeiro**. Congresso da Sociedade Brasileira de Computação, Escola Politécnica da Universidade de São Paulo, p. 859-868, São Paulo, 2009. Disponível em: <http://csbc2009.inf.ufrgs.br/anais/pdf/enia/st05_05.pdf>. Acesso em: 04 jan. 2022.

SMOLE, K. S. & DINIZ (orgs.), M. I. **Ler, escrever e resolver problemas**. Porto Alegre: Artmed, 2001.

VIALI, L.. Algumas considerações sobre a origem da Teoria da Probabilidade. **Revista Brasileira de História da Matemática**, PUC-RS e UFRGS, v. 8, n. 16, p. 143-153, mar. 2009. Disponível em: <http://rbhm.org.br/index.php/RBHM/article/view/177/163>. Acesso em 24 out. 2021.