



UNIVERSIDADE FEDERAL DO MARANHÃO

Curso de Ciência da Computação

Renan Merighi Alves de Lima

**Uma discussão sobre os classificadores
estatísticos: Bayes Simples e Redes Bayesianas**

São Luís/MA

2022

Renan Merighi Alves de Lima

Uma discussão sobre os classificadores estatísticos: Bayes Simples e Redes Bayesianas

Monografia apresentada ao curso de Ciência da Computação da Universidade Federal do Maranhão, como parte dos requisitos necessários para obtenção do grau de Bacharel em Ciência da Computação.

Curso de Ciência da Computação
Universidade Federal do Maranhão

Orientador: Prof. Dr. Ivo José da Cunha Serra

São Luís - MA

2022

Renan Merighi Alves de Lima

Uma discussão sobre os classificadores estatísticos: Bayes Simples e Redes Bayesianas

Monografia apresentada ao curso de Ciência da Computação da Universidade Federal do Maranhão, como parte dos requisitos necessários para obtenção do grau de Bacharel em Ciência da Computação.

Trabalho aprovado em São Luís - MA, 02 de Fevereiro de 2022

Prof. Dr. Ivo José da Cunha Serra
Orientador

Prof.^a Dra. Simara Vieira da Rocha
Examinadora 1

**Prof. Me. Carlos Eduardo Portela
Serra de Castro**
Examinador 2

São Luís/MA
2022

Ficha gerada por meio do SIGAA/Biblioteca com dados fornecidos pelo(a) autor(a).
Diretoria Integrada de Bibliotecas/UFMA

de Lima, Renan Merighi Alves.

Uma discussão sobre os classificadores estatísticos:
Bayes Simples e Redes Bayesianas / Renan Merighi Alves de
Lima. - 2022.

52 p.

Orientador(a): Ivo José da Cunha Serra.

Monografia (Graduação) - Curso de Ciência da
Computação, Universidade Federal do Maranhão, São Luís,
Maranhão, 2022.

1. Bayes Simples. 2. Classificadores. 3.
Probabilidade Condicional. 4. Redes Bayesianas. 5.
Teorema de Bayes. I. Serra, Ivo José da Cunha. II.
Título.

Agradecimentos

Os agradecimentos iniciais vão para o meu orientador Prof. Dr. Ivo José da Cunha Serra, por toda paciência na longa jornada que resultou na produção deste trabalho e todo o apoio fornecido na forma de conhecimento que foi necessário para superar os obstáculos deste trabalho.

A minha mãe, por todo o suporte moral indispensável nesse longo período de dificuldades.

Ao meu pai, por me nortear com toda sua sabedoria sobre como encerrar as dificuldades da vida.

Ao meu irmão, por sempre cuidar de mim desde que nasci, sendo uma fonte de inspiração para mim.

Ao meu amigo Diego Oliveira, que me ajudou extensivamente desde o início do processo de graduação, seja me motivando e ensinando nas cadeiras que tive dificuldades, assim como fornecendo suas sinceras opiniões sobre os mais diversos assuntos.

"Imitation is the sincerest form of flattery that mediocrity can pay to greatness."

(Oscar Wilde)

Resumo

Com o advento do avanço tecnológico na área de computação, tornou-se possível e viável a simulação de eventos reais em um modelo computacional, permitindo assim a aplicação de conceitos matemáticos e estatísticos com o intuito de encontrar respostas para problemas práticos. Esses modelos podem ser implementados nos mais diversos ambientes, seja no cenário de suporte ao sistema hospitalar, ou para a predição de inadimplência de um solicitante de empréstimo, por exemplo. Este trabalho tem por objetivo apresentar e discutir os classificadores Bayesianos, em específico Bayes simples e as redes Bayesianas. Dando enfoque a probabilidade condicional, onde uma condição observável afeta a probabilidade final de um experimento qualquer, e ao teorema de Bayes, utilizado extensivamente pelos classificadores Bayesianos para determinar a probabilidade final de um determinado modelo computacional. Além de breve discussão sobre as características de ambos os classificadores, como o passo-a-passo do processo matemático para a tomada de decisão durante a predição; o comportamento entre as variáveis aleatórias (atributos), seja em uma relação de dependência, onde haverá impacto entre as variáveis dependentes, ou independência, onde a classe é a única variável dependente no modelo preditivo; além de fatores de incerteza sobre os valores assumidos pelas variáveis aleatórias e as consequências que isso produz sobre o cálculo de probabilidade de ambos os classificadores.

Palavras-chave: Probabilidade Condicional. Teorema de Bayes. Bayes Simples. Redes Bayesianas. Classificadores.

Abstract

With the advent of technological advances in computing, it has become possible and feasible to simulate real events in a computational model, thus allowing the application of mathematical and statistical concepts in order to find answers to practical problems. These models can be implemented in the most diverse environments, whether in the scenario of support to the hospital system, or for the prediction of default of a loan applicant, for example. This work aims to present and discuss Bayesian classifiers, specifically simple Bayes and Bayesian networks. Focusing on conditional probability, where an observable condition affects the final probability of any experiment, and on Bayes' theorem, used extensively by Bayesian classifiers to determine the final probability of a given computational model. In addition to a brief discussion about the characteristics of both classifiers, such as the step-by-step mathematical process for decision making during prediction; the behavior between random variables (attributes), either in a dependent relationship, where there will be an impact between the dependent variables, or independent, where the class is the only dependent variable in the predictive model; in addition to uncertainty factors about the values assumed by the random variables and the consequences that this produces on the probability calculation of both classifiers.

Keywords: Conditional Probability. Bayes' theorem. Bayesian networks. Classifiers.

Lista de ilustrações

Figura 1 – União dos Eventos A e B, mutuamente exclusivos	15
Figura 2 – União dos Eventos A e B, não mutuamente exclusivos	15
Figura 3 – Intersecção dos Eventos A e B	16
Figura 4 – Complementação dos Eventos A e \bar{A}	16
Figura 5 – Complementação dos Eventos A e B	19
Figura 6 – Cinco lançamentos em sequencia de uma moeda	20
Figura 7 – Variáveis aleatórias em dois lançamentos de uma moeda não viciada . .	21
Figura 8 – Variável aleatória em dois lançamentos de uma moeda viciada	21
Figura 9 – Três urnas	26
Figura 10 – Representação geométrica da Tabela 4	29
Figura 11 – $P(\text{proprietariodecasa} = \text{nao} \text{sim})$	32
Figura 12 – $P(\text{estadocivil} = \text{casado} \text{sim})$	33
Figura 13 – $P(\text{proprietariodecasa} = \text{nao} \text{nao})$	33
Figura 14 – $P(\text{estadocivil} = \text{casado} \text{nao})$	34
Figura 15 – Exemplos de grafo acíclico e cíclico	36
Figura 16 – Exemplos de grafos direcionado e não direcionado	37
Figura 17 – Representação em grafo da relação câncer e úlcera	37

Lista de tabelas

Tabela 1 – Classificação do IMC	22
Tabela 2 – Base de dados hipotética de IMC	23
Tabela 3 – Tabela do lançamento de um dado e moeda	23
Tabela 4 – Base de dados de Inadimplentes	29
Tabela 5 – $P(\text{proprietariodecasa} = \text{nao} \text{sim})$	32
Tabela 6 – $P(\text{estadocivil} = \text{casado} \text{sim})$	32
Tabela 7 – $P(\text{proprietariodecasa} = \text{nao} \text{nao})$	33
Tabela 8 – $P(\text{estadocivil} = \text{casado} \text{nao})$	34
Tabela 9 – Probabilidade de Obesidade	38
Tabela 10 – Probabilidade de Alcoolismo	38
Tabela 11 – Probabilidade de Câncer de Estômago	39
Tabela 12 – Probabilidade de Úlcera Gástrica	39
Tabela 13 – Probabilidade de Vômito com Sangue	39
Tabela 14 – Probabilidade conjunta de $P(CE = \text{sim})$ da Figura 17	42
Tabela 15 – Probabilidade Conjunta de $P(CE = \text{sim}, UG = \text{nao}, VS = \text{sim}, Ob = \text{nao}, Al = \text{nao})$	43
Tabela 16 – Probabilidade Conjunta do Grafo do grafo 17 para $CE = \{\text{sim}\}$	44
Tabela 17 – Probabilidade Conjunta do Grafo do grafo 17 para $CE = \{\text{nao}\}$	44
Tabela 18 – Probabilidade de $P(CE = \text{sim}, UG = \text{nao}, VS = \text{sim})$	44
Tabela 19 – Probabilidade de $P(CE = \text{nao}, UG = \text{nao}, VS = \text{sim})$	45
Tabela 20 – Probabilidade Conjunta de $P(Ob = \text{nao}, Al = \text{nao}, CE = \text{sim}, UG = \text{nao}, VS = \text{sim})$	45
Tabela 21 – Probabilidade Conjunta de $P(Ob = \text{nao}, Al = \text{nao}, CE = \text{nao}, UG = \text{nao}, VS = \text{sim})$	46

Sumário

	Lista de tabelas	8
1	INTRODUÇÃO	11
1.1	Objetivos	12
1.1.1	Objetivo Geral	12
1.1.2	Objetivos Específicos	12
1.2	Organização do Trabalho	12
2	FUNDAMENTOS TEÓRICOS PROBABILÍSTICOS	14
2.1	Espaço Amostral	14
2.2	Evento	14
2.2.1	União de Eventos	14
2.2.2	Intersecção de Eventos	15
2.2.3	Eventos Complementares	16
2.3	Probabilidade do Evento	16
2.3.1	Probabilidades Iguais	17
2.3.2	Probabilidades Diferentes	18
2.3.3	Teorema da Soma	19
2.3.4	Probabilidade Complementar	19
2.4	Variável Aleatória Discreta	20
2.5	Variável Aleatória Contínua	22
2.6	Variáveis Aleatórias Independentes	23
2.7	Probabilidade Condicional	24
2.8	Teorema de Bayes	25
3	CLASSIFICADORES BAYESINANOS	28
3.1	Classificador Bayesiano Simples	28
3.1.1	Atributos Discretos	30
3.1.2	Atributos Contínuos	34
3.1.3	Aplicação do Classificador de Bayes Simples	35
3.2	Classificador de Redes de Crenças Bayesianas	36
3.3	Características do Bayes Simples e das Redes Bayesianas	46
4	CONSIDERAÇÕES FINAIS	49

REFERÊNCIAS 51

1 Introdução

O ato de predizer o futuro não é recente, há vários momentos no passado nos quais figuras históricas afirmam saber o desfecho de eventos futuros. Seja no Oráculo de Delfos, na Grécia Antiga, onde foi profetizado que Édipo mataria seu próprio pai e se casaria com sua própria mãe. Seja na participação de Joana D'arc durante o período final da Guerra dos Cem Anos na França, onde suas revelações iriam impactar no desfecho vitorioso durante o Cerco a Orléans e na Conquista de Reims. Ou mesmo no livro *Les Prophéties*, escrito por Michel de Nostredame no século XVI, o qual continha várias predições sobre o futuro.

Um outro caso que apesar da distância temporal possui relevância no mundo contemporâneo, é o das cheias do rio Nilo, responsável pela fertilidade do solo e por consequência impactando na produção agrícola da região. Dessa forma, saber se a próxima cheia seria boa ou ruim permitiria se preparar adequadamente para enfrentar tal situação. Porém esses eventos retratados ocorreram séculos ou mesmo milênios atrás, período no qual a tecnologia era limitada e pouco difundida.

Hoje em dia não temos à nossa disposição um Oráculo de Delfos, uma Joana D'arc ou mesmo Nostradamus, para que possamos perguntar como eles conseguiam prever o futuro. O ser humano, por sua vez, na incansável busca em predizer o futuro, “criou” seus próprios “profetas” na forma de fórmulas matemáticas capazes de gerar “profecias” a respeito de determinados fatos que desejam ser “adivinhados”. Desfrutamos de uma era de prosperidade marcada por um rápido crescimento tecnológico, uma ampla distribuição global dos conhecimentos e um longo período de paz. Isto proporciona oportunidades no desenvolvimento tecnológico de várias áreas, sendo uma delas a de “prever o futuro”, ou em outras palavras, a dos classificadores.

Seja no formato de árvore de decisões, que realiza a classificação embasando-se no comportamento dos atributos para tomada de decisão, seja utilizando os conceitos de regressão, que se fundamenta na observação do passado para predição do futuro, seja nas redes neurais, que simulam a forma do ser humano aprender, ou mesmo em Bayes, que buscam realizar predição apenas observando uma parcela das informações. Cada um dos classificadores é um “profeta” único tendo pontos fortes e fracos para determinadas circunstâncias, estando todos eles à disposição do ser humano.

Um exemplo foi o caso do Air France 447, que em 31 de maio de 2009, desapareceria pelos próximos 2 anos ao realizar a rota Rio de Janeiro para França. Conforme aponta o relatório final da BEA (Bureau d'Enquêtes et d'Analyses - Bureau de Investigações e Análise), o processo de recuperação da caixa preta do 447 passou por 4 fases até que os

destroços do avião fossem localizados no fundo do oceano Atlântico. De acordo com o relatório final (D'ANALYSES, 2012) e após os insucessos nas 3 primeiras fases de busca do avião, a BEA solicitou um estudo (STONE et al., 2014) que foi pautado na utilização da inferência bayesiana, justificado pela existência de grande quantidade de incertezas a respeito dos dados, além do fato do evento ser único impossibilitando a replicação do mesmo por observação de eventos parecidos. Uma semana após o estudo ser realizado utilizando inferência bayesiana, os destroços do Air France 447 foram encontrados, encerrando-se assim a penosa busca do Airbus.

Este trabalho de graduação tem por objetivo expor os conceitos e funcionamento da predição realizado pelo classificador bayesiano, através da explanação de um experimento. Para isso serão apresentados, inicialmente, os conceitos básicos de probabilidade, presentes no capítulo 02, visando proporcionar ao leitor deste trabalho um embasamento nos princípios probabilísticos, para assim prosseguir para a explicitação do processo analítico realizado pelo classificador bayesiano, presentes no capítulo 03.

1.1 Objetivos

1.1.1 Objetivo Geral

Realizar uma apresentação e discussão sobre os classificadores estatísticos dos tipos Bayes simples e as Redes de crenças Bayesianas.

1.1.2 Objetivos Específicos

- Explicitar conhecimentos estatísticos básicos, através de exemplos, necessários para o entendimento do classificador Bayesiano.
- Apresentar os classificadores dos tipos Bayes simples e Redes de crenças Bayesianas evidenciando os conceitos estatísticos nos quais são embasados.
- Apresentar uma discussão sobre a aplicabilidade de cada um dos tipos de classificadores levando em consideração a teoria a eles relacionada e exemplos ilustrativos.

1.2 Organização do Trabalho

O capítulo 01 expõe brevemente o interesse humano em prever eventos, e como o avanço tecnológico permitiu que as predições deixaram de ser misticismos e passaram a ser ciência. Também contém os objetivos gerais e específicos deste trabalho acadêmico.

O capítulo 02 apresenta e exemplifica conceitos probabilísticos básicos que posteriormente servirão como base para o entendimento do classificador bayesiano. Esses conceitos

compreendem a relação entre espaço amostral, evento e a probabilidade de ocorrência de um evento, noções de variáveis aleatórias e os seus tipos, além de probabilidade condicional e o teorema de bayes.

O capítulo 03 foca nos classificadores bayesianos, especificamente em bayes simples e redes de crenças bayesianas. Inicialmente é exposto o funcionamento matemático do bayes simples por meio de um exemplo, em seguida é exposto as redes bayesianas, também por meio de um exemplo, visando deixar claro para o leitor as diferenças na aplicabilidade de ambos os classificadores.

2 Fundamentos Teóricos Probabilísticos

Este capítulo tem como objetivo expor os conceitos introdutórios probabilísticos, que posteriormente serão utilizados para explicação do capítulo 3, referente aos classificadores Bayesianos.

O capítulo inicia com uma breve explicação a respeito dos conceitos de espaço amostral e eventos, além de abordar os diferentes tipos de probabilidade e variáveis aleatórias e finaliza apresentando sobre probabilidade condicional e o teorema de Bayes, que são o cerne do funcionamento dos classificadores Bayes simples e as redes Bayesianas.

2.1 Espaço Amostral

De acordo com Combinatória (1991) espaço amostral é constituído pelo conjunto de todas as possibilidades de um experimento. Usemos como exemplo o lançamento de uma moeda. O total de possibilidades que pode ser obtido corresponde a dois, cara ou coroa. Dessa forma o espaço amostral, normalmente identificado pela letra grega Ω , é constituído de cara e coroa, logo $\Omega = \{cara, coroa\}$.

Ainda utilizando o exemplo anterior, porém incrementando o número de lançamentos para duas moedas, o espaço amostral passa a ser: $\Omega = \{cara\ cara, cara\ coroa, coroa\ cara, coroa\ coroa\}$.

Um outro exemplo a ser tomado será o espaço amostral de um baralho de cartas. O baralho contém 4 naipes diferentes (paus, ouro, copas e espadas) e cada naipe contém 13 cartas, de Ás a Rei, totalizando 52 cartas. Portanto a remoção aleatória de uma carta do baralho representa o espaço amostral, $\Omega = \{As_{\clubsuit}, 2_{\clubsuit}, \dots, Rei_{\spadesuit}\}$.

2.2 Evento

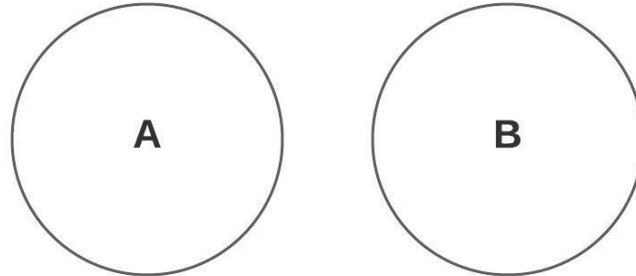
Conforme aponta Dantas (2013) um evento é um subconjunto do espaço amostral do experimento. Tomando o exemplo do lançamento de duas moedas, um evento, E_1 , seria a ocorrência de pelo menos uma coroa. Já no exemplo do baralho, um possível evento, E_2 , seria a obtenção de uma carta de naipe de espadas, ou seja, $E_2 = \{As_{\spadesuit}, 2_{\spadesuit}, \dots, Reis_{\spadesuit}\}$.

2.2.1 União de Eventos

Dado dois eventos quaisquer, A e B , a união de ambos os eventos corresponde a junção ou adição dos eventos e é representado por $A \cup B$. Segundo Falco (2016), a união

de eventos pode ocorrer entre eventos que são mutuamente exclusivos, conforme mostrado na Figura 1, ou não mutuamente exclusivos, conforme a Figura 2.

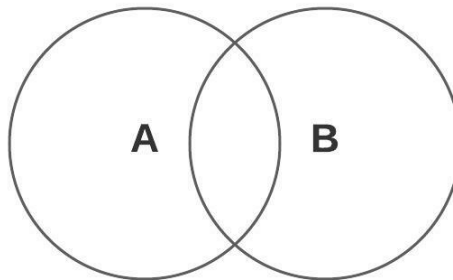
Figura 1 – União dos Eventos A e B, mutuamente exclusivos



Fonte: Elaborada pelo autor

Dado o conjunto $X_1 = \{2, 4, 6, 8, 10\}$ e o conjunto $Y_1 = \{1, 3, 5, 7, 9\}$, constata-se que não há valores em comum entre os conjuntos, logo eles são mutuamente exclusivos.

Figura 2 – União dos Eventos A e B, não mutuamente exclusivos



Fonte: Elaborada pelo autor

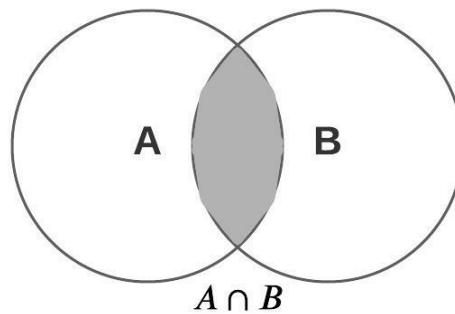
Por outro lado, o conjunto $X_2 = \{1, 2, 3, 4, 5\}$ e o conjunto $Y_2 = \{2, 3, 5, 7, 11\}$ possui valores em comum, por sua vez são não mutuamente exclusivos.

2.2.2 Intersecção de Eventos

De acordo com Ynoguti (2011), a intersecção de eventos, normalmente representado por $(A \cap B)$, condiz com o cruzamento dos eventos, conforme mostrado na Figura 3.

A intersecção de eventos é, conforme mostra a Figura 3, representada pelo símbolo \cap e corresponde, obrigatoriamente, à ocorrência de todos os elementos em comum entre os eventos.

Figura 3 – Intersecção dos Eventos A e B

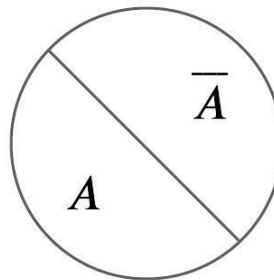


Fonte: Elaborada pelo autor

Utilizando novamente os conjuntos $X_2 = \{1, 2, 3, 4, 5\}$ e $Y_2 = \{2, 3, 5, 7, 11\}$, temos que $(X_2 \cap Y_2) = \{2, 3, 5\}$.

2.2.3 Eventos Complementares

Segundo Reis e Reis (2002), a complementação de eventos, conforme mostra a Figura 4, é representada pelo símbolo de uma barra sobre o evento, (\bar{A}) , e corresponde a não ocorrência do evento no espaço amostral que ele pertence.

Figura 4 – Complementação dos Eventos A e \bar{A} 

Fonte: Elaborada pelo autor

Seja o espaço amostral formado por todos os números naturais, e dado o conjunto X_3 formado por todos os números pares, então \bar{X}_3 corresponde a todos os números não pares, ou ímpares, do espaço amostral.

2.3 Probabilidade do Evento

Embasando-se em Novaes, Queiroz e Coutinho (2000) pode-se definir a axiomática de Kolmogorov dado um experimento qualquer E , e um espaço amostral Ω , onde:

1. A probabilidade de qualquer evento deve ser maior ou igual a zero.

$$P(A) \geq 0$$

2. A probabilidade do espaço amostral deve ser igual a 1.

$$P(\Omega) = 1$$

3. Para dois eventos disjuntos, a probabilidade da união de ambos os eventos é igual ao somatório da probabilidade dos eventos.

$$A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$$

4. No entanto, caso o espaço amostral seja não finito, o axioma acima pode ser substituído por:

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

A probabilidade de um evento $P(A)$ pode ser expressa levando em conta duas situações: a probabilidade dos eventos serem iguais ou diferentes.

2.3.1 Probabilidades Iguais

Pautado em Martins (2014), a equiprobabilidade de eventos ocorre quando cada evento de um espaço amostral apresenta mesma probabilidade de ocorrer.

Tomando como exemplo o lançamento de uma moeda, como já foi abordado nas seções 2.1 e 2.2, sabemos de antemão que o espaço amostral corresponde a dois eventos, $\Omega = \{cara, coroa\}$, assumindo que a moeda não seja viciada, e utilizando-se das definições apresentadas na seção 2.3, obtém-se a Equação 2.1.

$$P(cara) + P(coroa) = 1 \tag{2.1}$$

Como $P(cara) = P(coroa)$, tendo em vista que a moeda não é viciada, ou seja, a probabilidade são iguais em ambos os eventos, podemos assumir que $P(Y)$ corresponde a probabilidade de qualquer um dos eventos ocorrer, dessa forma:

$$\begin{aligned} 2P(Y) &= 1 \\ P(Y) &= 1/2 \end{aligned} \tag{2.2}$$

Generalizando as Equações 2.1 e 2.2, para uma probabilidade $P(Y)$, obtém-se a Equação 2.3 para eventos com mesma probabilidade:

$$P(Y) = \frac{P(\text{Quantidade de Eventos})}{P(\text{Espaco Amostral})} \quad (2.3)$$

Aplicando a Equação 2.3 para o exemplo de dois lançamentos de dados e buscando descobrir a probabilidade de se obter pelo menos uma coroa, obtém-se:

$$\begin{aligned} P(\text{Quantidade de Eventos}) &= \{\text{cara coroa}, \text{coroa cara}, \text{coroa coroa}\} = 3 \\ P(\text{Espaco Amostral}) &= \{\text{cara cara}, \text{cara coroa}, \text{coroa cara}, \text{coroa coroa}\} = 4 \\ P(Y) &= \frac{3}{4} = 75\% \end{aligned}$$

2.3.2 Probabilidades Diferentes

No mundo real, no entanto, há situações nas quais eventos não possuem a mesma probabilidade de ocorrer, como por exemplo, de um dia ser ensolarado ou chuvoso. Um outro exemplo desta desigualdade é apontado por Caruzo, Neves e Teixeira (2007), ao indagar sobre a probabilidade do somatório do lançamento de dois dados.

Seguindo o mesmo exemplo exposto na seção anterior, caso a moeda tivesse o dobro de chance de sair cara, a probabilidade deste evento seria calculada da seguinte forma:

$$\begin{aligned} P(CARA) &= 2 \times P(COROA) \\ P(CARA) + P(COROA) &= 1 \\ 2 \times P(COROA) + P(COROA) &= 1 \\ P(COROA) &= \frac{1}{3} \\ P(CARA) &= \frac{2}{3} \end{aligned}$$

Como o exemplo trata de dois lançamentos, então temos que:

$$\begin{aligned} P(CARA) &= P(\text{cara cara}) + P(\text{cara coroa}) + P(\text{coroa cara}) + P(\text{coroa coroa}) \\ P(CARA) &= \left(\frac{2}{3} \times \frac{2}{3}\right) + \left(\frac{2}{3} \times \frac{1}{3}\right) + \left(\frac{1}{3} \times \frac{2}{3}\right) \\ P(CARA) &= \frac{4}{9} + \frac{2}{9} + \frac{2}{9} \\ P(CARA) &= \frac{8}{9} \approx 88,88\% \end{aligned}$$

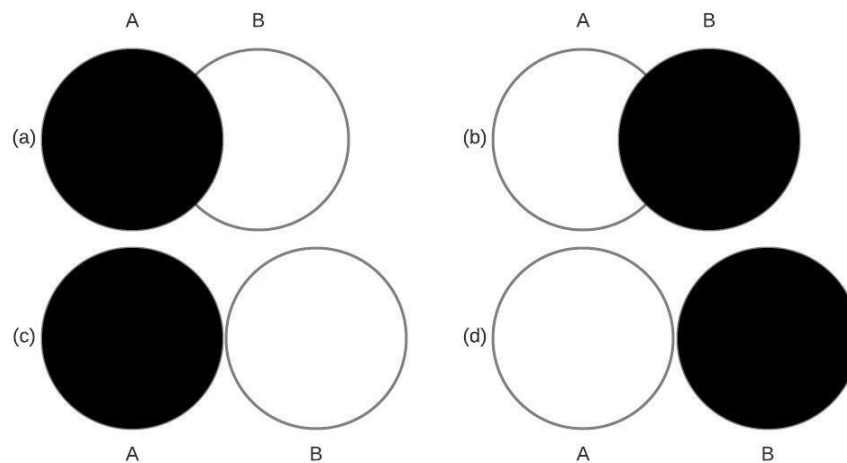
Logo, a probabilidade de se obter pelo menos uma cara no lançamento de uma moeda viciada é aproximadamente 88,88%.

2.3.3 Teorema da Soma

Em concordância com Correa (2003) tem-se que dado dois eventos, A e B, a probabilidade da união, ou teorema da soma, é expresso pela Equação 2.4

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad (2.4)$$

Figura 5 – Complementação dos Eventos A e B



Fonte: Elaborada pelo autor

É importante ressaltar que a subtração advém do fato de que no caso dos ambos os eventos interseccionarem, o somatório dos eventos possuirão duplicações devido a essa intersecção. Isso pode ser observado na Figura 5 onde (a) corresponde a todos os elementos presentes no evento A e (b) a todos os elementos presentes no evento B, ficando visível que os eventos A e B possuem elementos em comum.

No caso de eventos disjuntos, a intersecção é vazia e por sua vez respeita o terceiro axioma de Kolmogorov mencionado na seção 2.3. Isso pode ser averiguado na Figura 5 onde (c) os elementos do evento A não intersectam (d) os elementos do evento B.

2.3.4 Probabilidade Complementar

Baseado em Souza (2020) tem-se que dado um evento A, a complementação desse evento, \bar{A} , pode ser expressa por:

$$P(\bar{A}) = 1 - P(A) \quad (2.5)$$

Essa equação pode ser interpretada como, a probabilidade do evento A não acontecer. A subtração por 1 ocorre, pois $P(\Omega) = 1$, então a diferença corresponde ao conjunto de

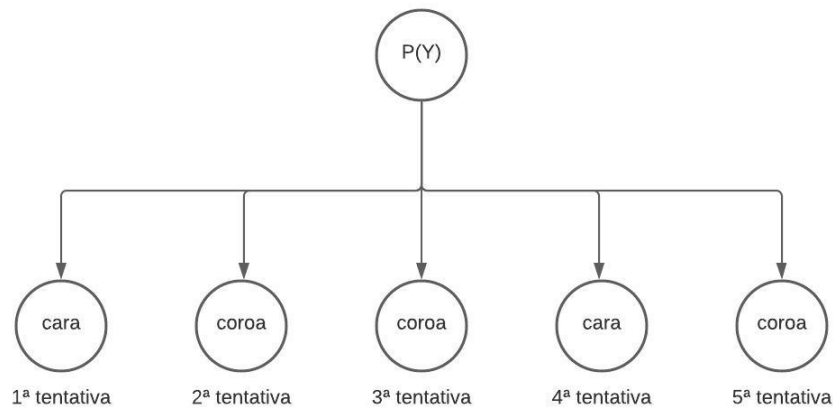
eventos que o evento A não ocorre. Isso pode ser observado na Figura 4, se tratarmos o evento B como o espaço amostral Ω .

2.4 Variável Aleatória Discreta

Em conformidade com o que aponta Portnoi (2005), valores são classificados como discretos quando é possível enumerar ou contar uma representação, por exemplo, a quantidade de vezes que uma moeda, ao ser lançada, resultou em coroa, ou o número de ocorrência do naipe de ouro ao puxar as primeiras 5 cartas do topo do baralho.

Observando a Figura 6 é possível enumerar, ou quantificar, que ao lançar cinco vezes uma mesma moeda foi obtido duas caras e três coroas.

Figura 6 – Cinco lançamentos em sequencia de uma moeda



Fonte: Elaborada pelo autor

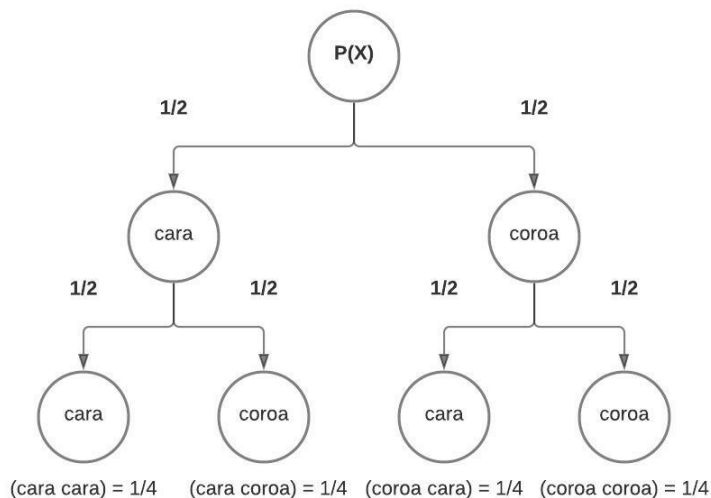
Por definição temos que $P(X = x_i)$, onde X é a variável aleatória, x_i os valores que a variável aleatória pode assumir e $P(X)$ a função da probabilidade da variável aleatória.

Utilizando a Figura 7 para melhor exemplificar, temos que X corresponde a face da moeda voltado para cima. x_0 representa cara e x_1 coroa. $P(X = x_0)$ ou $P(x_0)$ a probabilidade da face voltada para cima de uma moeda ser igual a cara.

É importante ressaltar que os axiomas descritos na seção 2.3 devem se aplicar a $P(X)$, ou seja:

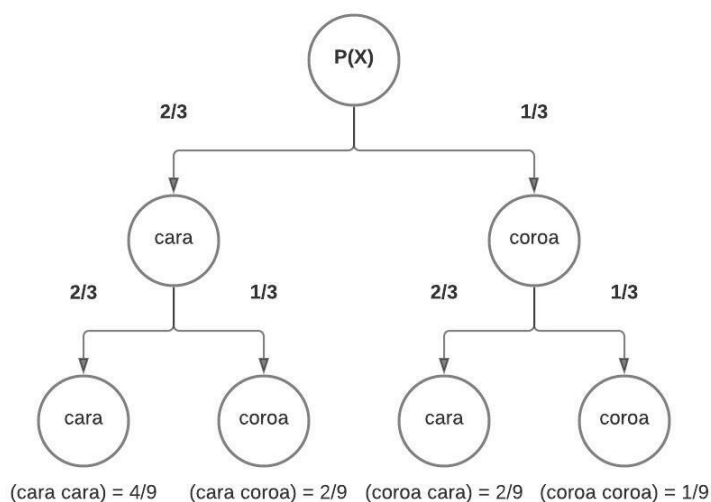
1. $P(x_i) \geq 0$, a probabilidade do valor da variável aleatória deve ser maior ou igual a 0.
2. $\sum_{i=1}^{\infty} P(x_i) = 1$, o somatório de todas as probabilidades dos valores das variáveis aleatórias deve ser igual a 1.

Figura 7 – Variáveis aleatórias em dois lançamentos de uma moeda não viciada



Fonte: Elaborada pelo autor

Figura 8 – Variável aleatória em dois lançamentos de uma moeda viciada



Fonte: Elaborada pelo autor

A probabilidade da variável aleatória é responsável por mensurar a chance de um determinado evento ocorrer, o que implica na possibilidade de realizar uma escolha mais otimizada durante um determinado experimento.

Supondo um experimento onde seja testado a confiabilidade de uma moeda, se após 100 lançamentos forem observadas as probabilidades da Figura 7, isso significa que tanto faz escolher o lado da moeda na hora do lançamento, pois as probabilidades serão iguais. No entanto, se o mesmo experimento for realizado e os resultados observados forem

iguais ao da Figura 8, então a escolha do lado passa a ser decisiva durante o lançamento, pois as chances são diferentes.

2.5 Variável Aleatória Contínua

Uma variável aleatória é considerada contínua quando não é possível realizar a "contagem" da variável aleatória devido ao fato desta variável poder assumir uma quantidade infinita de valores. Farias (2017) aborda um exemplo referente a altura dos brasileiros que, por ser um valor real, reflete a existência de uma quantidade infinita de valores possíveis de serem atribuídos a cada pessoa.

No entanto, é importante diferenciar que apesar de peso e altura serem variáveis contínuas, elas podem ser discretizadas quando combinadas e associadas a classificação do índice de massa corpórea (IMC). Isso ocorre porque os infinitos valores que as variáveis podem assumir estão agora limitados a um número finito ou contável de valores.

Utilizando as classificações do IMC apresentados na Tabela 1 sobre os valores expressos nas duas primeiras colunas da Tabela 2 e realizando o cálculo do IMC torna-se possível a verificação da característica contável da variável IMC. Nota-se que o valor numérico real do IMC é contínuo, no entanto, a classificação final é discreta, uma vez que está dentro de seis valores predeterminados que esta variável pode assumir.

Tabela 1 – Classificação do IMC

Classificação	IMC
Deficiência de Peso	<18,5
Normal	18,5 - 24,9
Excesso de Peso	25 - 29,9
Obesidade I	30 - 34,9
Obesidade II	35 - 39,9
Obesidade III	>40

Fonte: Silva (2006)

Tabela 2 – Base de dados hipotética de IMC

Peso(kg)	Altura(m)	IMC	Classificação
95	1,87	27,17	Excesso de peso
107	1,88	30,27	Obesidade de Classe 1
110	1,64	40,90	Obesidade de Classe 3
90	1,73	30,07	Obesidade de Classe 1
90	1,8	27,78	Excesso de peso
64	1,76	20,66	Peso normal
99	1,88	28,01	Excesso de peso
108	1,63	40,65	Obesidade de Classe 3
106	1,67	38,01	Obesidade de Classe 2
94	1,79	29,34	Excesso de peso

Fonte: Elaborada pelo autor

2.6 Variáveis Aleatórias Independentes

Pautando-se em Combinatória (1991), define-se que variáveis são independentes se o valor da probabilidade conjunta for igual ao produto das probabilidades marginais, ou seja:

$$P(X = i, Y = j) = P(X = i) \times P(Y = j) \quad (2.6)$$

Observando o exemplo da Tabela 3, que representa o lançamento de uma moeda e um dado ao mesmo tempo, os valores de i representam o resultado da moeda $P(X = i) = \{0, 1\}$, cara e coroa, respectivamente. Os valores de j representam o resultado do dado $P(Y = j) = \{0, 1, 2, 3, 4, 5, 6\}$.

Tabela 3 – Tabela do lançamento de um dado e moeda

i \ j	0	1	2	4	5	6	$P(X = i)$
0	1/12	1/12	1/12	1/12	1/12	1/12	1/2
1	1/12	1/12	1/12	1/12	1/12	1/12	1/2
$P(Y = j)$	1/6	1/6	1/6	1/6	1/6	1/6	1

Fonte: Elaborada pelo autor

Supondo o exemplo em que o resultado da moeda seja cara, representado por 0, e o resultado do dado seja 6, aplicando a Fórmula 2.6 temos:

$$P(X = 0, Y = 6) = 1/12 \quad (2.7)$$

$$P(X = 0) \times P(Y = 6) = 1/2 \times 1/6 = 1/12 \quad (2.8)$$

A Equação 2.7 é obtida a partir da Tabela 3 na qual foi extraído a probabilidade conjunta dos seguintes eventos: resultado cara no lançamento do dado, linha 2, e resultado 6 no lançamento do dado, coluna 7.

A Equação 2.8 é obtida pelo produto de cada probabilidade, ou seja, o resultado da probabilidade de obter cara, $1/2$, e de obter resultado 6 no lançamento do dado, $1/6$.

Como a definição de variáveis independentes se provou verdadeira no caso exemplificado, conclui-se que o resultado do lançamento de um dado não influencia no resultado do lançamento de uma moeda, conforme esperado.

2.7 Probabilidade Condicional

Iniciaremos a explicação de probabilidade condicional com base no seguinte exemplo: uma pessoa remove uma carta de um baralho e lhe informa que a carta é um número par. Qual a probabilidade de você acertar a carta que a pessoa removeu?

Vamos analisar este exemplo. Um baralho contém 52 cartas, que vão de Ás até rei e de 4 naipes diferentes, paus, ouros, copas e espadas. Como a pessoa que retirou a carta lhe informou a condição que a carta retirada é par, isso significa que o espaço amostral que antes era 52 passa a ser 20, tendo em vista que agora a carta retirada deve estar dentro do conjunto $\alpha = \{2, 4, 6, 8, 10\}$. Dessa forma, a chance de adivinhar a carta correta corresponde a $P(X) = 1/20 = 5\%$.

Estendendo o exemplo e assumindo que a pessoa também lhe informe que a carta é de cor vermelha, ou seja, ouros e copas. O novo espaço amostral possuirá 10 elementos e a probabilidade de adivinhar a carta correta passa a ser de $P(X) = 1/10 = 10\%$.

Dessa forma, a probabilidade condicional, como o nome já explicita, impõem uma condição sobre o experimento observado que por sua vez impacta na probabilidade de um determinado evento.

Pontifícia (2000) expressa essa relação condicional na formulação matemática da Equação 2.9:

$$P(A|B) = \frac{A \cap B}{P(B)} \quad (2.9)$$

1. $P(A)$, corresponde com a probabilidade de uma carta qualquer do baralho.
2. $P(B)$, corresponde com a probabilidade de a carta ser par do baralho.

3. $P(A|B)$, significa a probabilidade de o evento A ocorrer dado que o evento B já ocorreu. No exemplo descrito acima, essa probabilidade condiz com o fato de adivinhar uma carta de um baralho dado que ela é par.
4. $P(A \cap B)$, corresponde a intersecção de A e B, ou seja, a quantidade de elementos iguais em ambos os conjuntos. No nosso exemplo temos que o conjunto A corresponde a todas as cartas do baralho e o conjunto B apenas as cartas pares, dessa forma, $(A \cap B) = \{2\clubsuit, 4\clubsuit, \dots, 10\spadesuit\}$.

Substituindo na fórmula temos que:

$$\begin{aligned}
 P(A) &= \frac{1}{52} \\
 P(B) &= \frac{20}{52} \\
 P(A \cap B) &= \frac{1}{52} \\
 P(A|B) &= \frac{\frac{1}{52}}{\frac{20}{52}} = \frac{1}{20} = 5\%
 \end{aligned}$$

E na expansão do exemplo, onde foi informado que o naipe era ou de copas ou de ouros, temos o seguinte cálculo:

$$\begin{aligned}
 P(A) &= \frac{1}{52} \\
 P(B) &= \frac{10}{52} \\
 P(A \cap B) &= \frac{1}{52} \\
 P(A|B) &= \frac{\frac{1}{52}}{\frac{10}{52}} = \frac{1}{10} = 10\%
 \end{aligned}$$

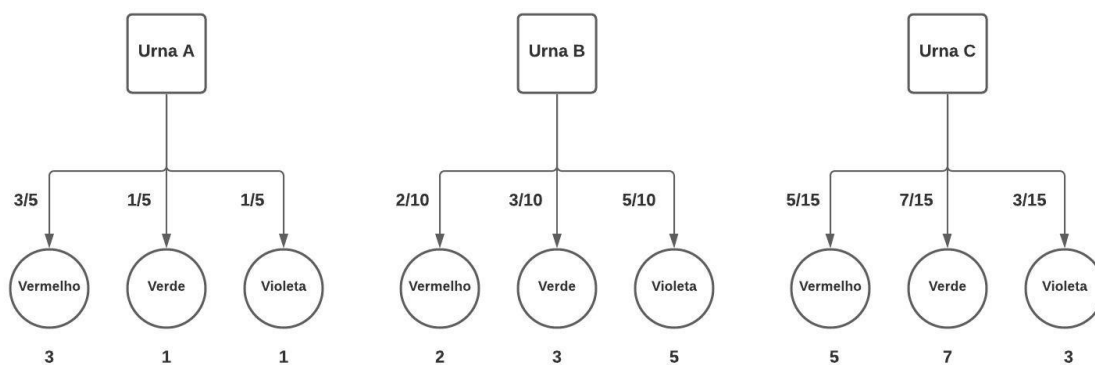
2.8 Teorema de Bayes

O teorema de bayes estabelece uma relação de conhecimento sobre o evento observado, associando um conhecimento prévio, conhecido como probabilidade priori, sobre o que deseja ser previsto, também conhecido como probabilidade posteriori. Conforme aponta Rossatto et al. (2016), essa relação pode ser expressa de acordo com a Equação 2.10.

$$P(A|B) = \frac{P(A) \times P(B|A)}{P(B)} \quad (2.10)$$

Suponha o exemplo da Figura 9, cada uma das três urnas contém bolinhas de três cores diferentes. Em cada uma das urnas o número total de bolinhas e a distribuição das cores são diferentes. Uma pessoa retira uma bola aleatoriamente e informa que a cor da mesma é verde, qual a probabilidade dela ter vindo da urna A?

Figura 9 – Três urnas



Fonte: Elaborada pelo autor

A princípio somos levados a crer que a probabilidade corresponde a $1/15$, dado que há equiprobabilidade da seleção das urnas, ou seja, todas elas tem a mesma chance de ser escolhida, e de a urna A conter apenas uma bola verde. No entanto, essa conclusão é errônea uma vez que deseja-se saber a probabilidade de a urna escolhida ter sido a A dado a cor da bola ser verde e não a probabilidade de a bola verde sair na urna A.

Conforme foi descrito no exemplo, um dos eventos já é conhecido, que é a cor da bola, e em posse dessa informação deve-se calcular a probabilidade desse evento estar associado à urna A. Matematicamente essa probabilidade fica descrita como: $P(A|B)$, sendo A a urna e B a cor da bola, em outras palavras, a probabilidade da urna ser $a_i = \{A, B, C\}$ dado que a cor da bola é $b_i = \{vermelha, verde, violeta\}$.

Utilizando a Equação 2.10, temos que descobrir os valores de $P(A)$, $P(B|A)$ e $P(B)$.

1. $P(A)$ corresponde a probabilidade de a urna selecionada ser a Urna A, trataremos essa escolha como aleatória e equiprovável, logo, $P(A) = 1/3$.
2. $P(B|A)$ corresponde a probabilidade de a bola selecionada ser da cor verde dado que a urna selecionada foi a urna A, ou seja, dentro do espaço amostral que corresponde a Urna A, qual a probabilidade de a cor verde ter sido retirada. Conforme a Figura 9, sabemos que a probabilidade de a bola verde ser selecionada na Urna A é de $1/5$.
3. $P(B)$ corresponde a probabilidade de a bola removida ser verde, em outras palavras, o somatório da probabilidade de retirar uma bola verde na urna A mais a probabilidade

de remover uma bola verde na urna B mais a probabilidade de remover uma bola verde na urna C. $P(B) = (1/3 \times 1/5) + (1/3 \times 3/10) + (1/3 \times 7/15) \approx 0,32$.

Com essas informações podemos aplicar os valores ao Teorema de Bayes:

$$P(A) = 1/3$$

$$P(B|A) = 1/5$$

$$P(B) = (1/3 \times 1/5) + (1/3 \times 3/10) + (1/3 \times 7/15) \approx 0,32$$

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

$$P(A|B) = \frac{1/3 \times 1/5}{0,32} \approx 0,2068 = 20,68\%$$

Dessa forma, a probabilidade de a bola verde ter sido removida da urna A é de aproximadamente 20,68% ao invés dos 1/15 previamente determinado intuitivamente.

3 Classificadores Bayesianos

Este capítulo tem como objetivo ilustrar a diferença entre o funcionamento do classificador bayes simples e as redes de crenças bayesianas por meio de dois exemplos. O primeiro, referente ao bayes simples, demonstra que o classificador trata os atributos como sendo condicionalmente independentes. Por outro lado, o segundo exemplo demonstra que havendo condicionalidade entre as variáveis aleatórias, há a necessidade de aplicação de um classificador que leve essa condicionalidade em consideração, motivo pelo qual é utilizado o classificador de redes bayesianas.

Este capítulo apresenta 2 seções, na primeira será apresentado um exemplo de como o Classificador Bayes Simples realiza os cálculos tanto para atributos discretos quanto contínuos. Na seção seguinte é exposto um outro exemplo, no qual há condicionalidade entre os atributos, visando demonstrar o funcionamento do classificador de redes de crenças bayesianas.

3.1 Classificador Bayesiano Simples

Conforme aponta Pang-Ning, Steinbach e Kumar (2009), o classificador bayesiano simples trata os atributos como variáveis aleatórias independentes, esta suposição facilita o funcionamento do classificador, tendo em vista que a variação dos atributos impactam apenas na predição da classe. Esta relação de independência, presente em bayes simples, pode ser observada na Fórmula 3.1.

$$P(Y|X) = \frac{P(Y) \prod_{i=1}^n P(X_i|Y)}{P(X)} \quad (3.1)$$

É importante ressaltar que o produtório presente na Equação 3.1, reflete a característica de independência dos atributos. Dado esta independência, todos os valores dos atributos, X_i , são relacionados diretamente com a classe que deseja ser predita.

Para explicar o funcionamento do classificador bayesiano simples, foi optado a utilização do exemplo apresentado por Pang-Ning, Steinbach e Kumar (2009), tendo em vista que neste exemplo os cálculos do classificador são realizados sobre variáveis discretas e contínuas. No entanto, as omissões de cálculos presentes no livro foram preenchidas neste trabalho, dado que tais omissões podem prejudicar o entendimento dos leitores menos experientes sobre o assunto.

Em síntese, o classificador irá calcular as probabilidades de cada atributo, X_i , em relação a classe, Y . Para melhor explicar o funcionamento do algoritmo de Bayes Simples, será utilizado o exemplo da Tabela 4.

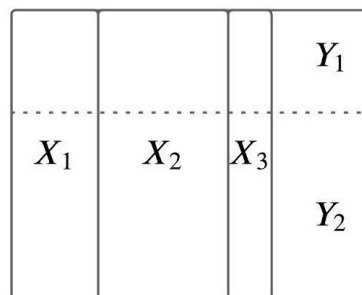
As linhas da Tabela 4 apresentam um conjunto de registros de pessoas, enquanto as colunas informam se esses indivíduos são proprietários de uma residência, o estado civil, a renda anual e se estão inadimplentes. No exemplo expresso na tabela, as informações referentes a $X = \{\text{proprietario casa, estado civil, renda anual}\}$ retratam os atributos, enquanto $Y = \{\text{solicitante emprestimo inadimplente}\}$ corresponde a classe, em outras palavras, a classe corresponde a informação que se deseja prever, enquanto os atributos são informações que ajudarão na predição.

Tabela 4 – Base de dados de Inadimplentes

Tid	Proprietário de Casa	Estado Civil	Renda Anual	Solicitante de Empréstimo Inadimplente
1	sim	solteiro	125k	não
2	não	casado	100k	não
3	não	solteiro	70k	não
4	sim	casado	120k	não
5	não	divorciado	95k	sim
6	não	casado	60k	não
7	sim	divorciado	220k	não
8	não	solteiro	85k	sim
9	não	casado	75k	não
10	não	solteiro	90k	sim

Fonte: Pang-Ning, Steinbach e Kumar (2009)

Figura 10 – Representação geométrica da Tabela 4



Fonte: Elaborada pelo autor

Na Figura 10, X_1 , X_2 e X_3 correspondem a proprietário de casa, estado civil e renda anual, respectivamente. Y_1 e Y_2 ilustram os valores que a classe, solicitante de empréstimo inadimplente, pode assumir, dessa forma, $Y_1 = \text{sim}$ e $Y_2 = \text{nao}$. A linha horizontal tracejada apenas representa uma separação das informações de acordo com o valor assumido pela classe, esta separação será utilizada pelo classificador bayesiano simples durante o processo

de cálculo das variáveis aleatórias independentes. A Figura 10 será utilizada para facilitar a visualização dos dados que são utilizados pelo classificador.

Considerando a inserção de um novo registro na tabela, cujo os atributos desse registro sejam: $P(X) = \{\text{proprietario casa} = \text{nao}, \text{estado civil} = \text{casado}, \text{renda anual} = 120k\}$. Em outras palavras, uma pessoa que não é proprietário de uma casa, que possui o estado civil casado e tem renda anual de \$120.000,00. Dado às informações já registradas na tabela, essa pessoa será classificada como tendo mais chances de ser ou não ser inadimplente?

Para obter a resposta da pergunta precisa-se adquirir as seguintes informações, $P(X|\text{nao})$ e $P(X|\text{sim})$ onde:

1. $P(X|\text{nao}) = \{\text{proprietario casa} = \text{nao}|\text{nao}, \text{estado civil} = \text{casado}|\text{nao}, \text{renda anual} = 120k|\text{nao}\}$
2. $P(X|\text{sim}) = \{\text{proprietario casa} = \text{nao}|\text{sim}, \text{estado civil} = \text{casado}|\text{sim}, \text{renda anual} = 120k|\text{sim}\}$

Para continuar os cálculos é necessário definir as seguintes questões: quais atributos são contínuos e quais são discretos?

3.1.1 Atributos Discretos

Conforme definido na seção 2.4, constata-se que os atributos referentes a Proprietário de Casa e Estado Civil, são discretos, tendo em vista que o primeiro atributo é representado por $x_1 = \{\text{sim}, \text{nao}\}$ enquanto o segundo atributo pode ser representado por $x_2 = \{\text{solteiro}, \text{casado}, \text{divorciado}\}$.

Para calcular a probabilidade nos dois casos, usamos os conceitos já apresentados na seção 2.8. Dessa forma, tem-se que:

$$\begin{aligned}
 &P(\text{proprietario de casa} = \text{nao}|\text{sim}) \\
 P(X|Y) &= \frac{P(X) \times P(Y|X)}{P(Y)} \\
 P(X) &= 7/10 \\
 P(Y|X) &= 3/7 \\
 P(Y) &= (7/10 \times 3/7 = 0,3) + (3/10 * 0/3) = 0,3 \\
 P(X|Y) &= \frac{7/10 \times 3/7}{0,3} = \frac{0,3}{0,3} = 1
 \end{aligned}$$

$$P(\text{estado civil} = \text{casado} | \text{sim})$$

$$P(X|Y) = \frac{P(X) \times P(Y|X)}{P(Y)}$$

$$P(X) = 4/10$$

$$P(Y|X) = 0/4 = 0$$

$$P(Y) = [4/10 \times 0/4] + [4/10 \times 2/4] + [2/10 \times 1/2] = 0,3$$

$$P(X|Y) = \frac{0 \times 4/10}{0,3} = 0$$

$$P(\text{proprietario de casa} = \text{nao} | \text{nao})$$

$$P(X|Y) = \frac{P(X) \times P(Y|X)}{P(Y)}$$

$$P(X) = 7/10$$

$$P(Y|X) = 4/7$$

$$P(Y) = [7/10 \times 4/7] + [3/10 \times 3/3] = 0,7$$

$$P(X|Y) = \frac{7/10 \times 4/7}{0,7} = 4/7$$

$$P(\text{estado civil} = \text{casado} | \text{nao})$$

$$P(X) = 4/10$$

$$P(Y|X) = 4/4$$

$$P(Y) = [4/10 \times 4/4] + [4/10 \times 2/4] + [2/10 \times 1/2] = 0,4 + 0,2 + 0,1 = 0,7$$

$$P(X|Y) = \frac{4/10 \times 4/4}{0,7} = 4/7$$

Uma outra forma de obter os valores calculados pelo teorema de Bayes consiste nos seguintes passos:

1. Excluir os registros que não pertencem a classe que será prevista.
2. Extrair desse novo espaço amostral os valores referentes aos atributos.
3. $P(X|Y)$ será igual ao número de casos onde o atributo observado ocorre dividido pelo espaço amostral.

Seguindo a mesma sequência que foi usada no teorema de Bayes, temos os números onde os valores marcados em negritos representam $P(X|Y)$.

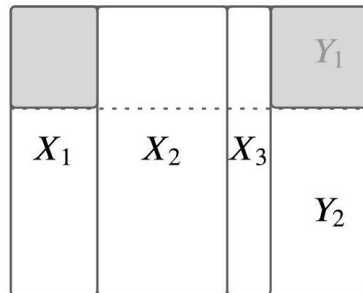
Tabela 5 – $P(\text{proprietariodecasa} = \text{nao}|\text{sim})$

Tid	Proprietário de Casa	Estado Civil	Renda Anual	Solicitante de Emprestimo Inadimplente
5	não	divorciado	95k	sim
8	não	solteiro	85k	sim
10	não	solteiro	90k	sim

Fonte: Pang-Ning, Steinbach e Kumar (2009)

Os três casos que foram isolados, conforme mostrado na Tabela 5, englobam a condição de não ser proprietário de casa e ser inadimplente, logo temos $P(X|Y) = 3/3 = 1$.

Figura 11 – $P(\text{proprietariodecasa} = \text{nao}|\text{sim})$



Fonte: Elaborada pelo autor

A Tabela 5 pode ser representada na Figura 11, onde todos os valores de X_1 estão presentes no domínio de Y_1 , ou seja, os valores de X_1 integram todo o espaço amostral acima da linha horizontal tracejada, comprovando assim que $P(X|Y) = 1$.

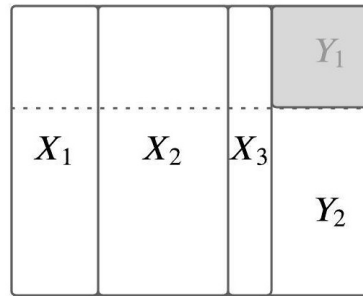
Tabela 6 – $P(\text{estadocivil} = \text{casado}|\text{sim})$

Tid	Proprietário de Casa	Estado Civil	Renda Anual	Solicitante de Emprestimo Inadimplente
5	não	divorciado	95k	sim
8	não	solteiro	85k	sim
10	não	solteiro	90k	sim

Fonte: Pang-Ning, Steinbach e Kumar (2009)

No caso onde a pessoa é casada e inadimplente, observamos que não há nenhuma registro, de acordo com a Tabela 6, onde a pessoa que é casada seja inadimplente, portanto $P(X|Y) = 0$.

Figura 12 – $P(\text{estadocivil} = \text{casado}|\text{sim})$



Fonte: Elaborada pelo autor

A Figura 12 mostra que $P(X|Y) = 0$, pois não há valores de X_2 condizente com o registro que deseja ser predito, $P(\text{estadocivil} = \text{casado})$, dado que o domínio em questão é Y_1 .

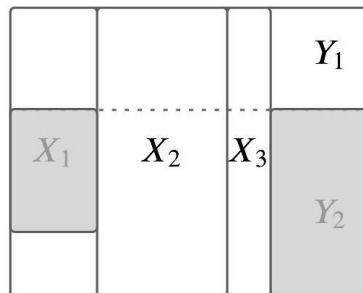
Tabela 7 – $P(\text{proprietariodecasa} = \text{nao}|\text{nao})$

Tid	Proprietário de Casa	Estado Civil	Renda Anual	Solicitante de Emprestimo Inadimplente
1	sim	solteiro	125k	não
2	não	casado	100k	não
3	não	solteiro	70k	não
4	sim	casado	120k	não
6	não	casado	60k	não
7	sim	divorciado	220k	não
9	não	casado	75k	não

Fonte: Pang-Ning, Steinbach e Kumar (2009)

Para a situação onde o registro informe que a pessoa não é proprietário de casa e não é inadimplente, segundo a Tabela 7, temos 4 situações, dessa forma $P(X|Y) = 4/7$.

Figura 13 – $P(\text{proprietariodecasa} = \text{nao}|\text{nao})$



Fonte: Elaborada pelo autor

Na Figura 13 o classificador realiza o mesmo cálculo de probabilidade sendo a única alteração no espaço amostral que o atributo será aplicado, ou seja, o classificador

verificará a probabilidade de uma pessoa não ser proprietário de casa dado que ela não é inadimplente. Diferente da Figura 11, apenas uma parcela das informações, precisamente $4/7$, é representado em X_1 , enquanto outro conjunto de valores, especificamente $3/7$, não está de acordo com o que deseja ser previsto, por essa razão o retângulo na Figura 13 está parcialmente preenchido em X_1 .

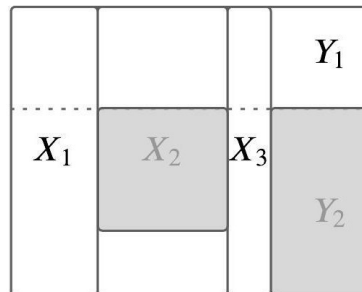
Tabela 8 – $P(\text{estadocivil} = \text{casado}|\text{nao})$

Tid	Proprietário de Casa	Estado Civil	Renda Anual	Solicitante de Emprestimo Inadimplente
1	sim	solteiro	125k	não
2	não	casado	100k	não
3	não	solteiro	70k	não
4	sim	casado	120k	não
6	não	casado	60k	não
7	sim	divorciado	220k	não
9	não	casado	75k	não

Fonte: Pang-Ning, Steinbach e Kumar (2009)

Há também 4 registros que mostram que pessoas casadas não são inadimplentes, em concordância com a Tabela 8, dessa forma $P(X|Y) = 4/7$.

Figura 14 – $P(\text{estadocivil} = \text{casado}|\text{nao})$



Fonte: Elaborada pelo autor

O mesmo procedimento aplicado em X_1 dado Y_2 ocorre em X_2 , expresso na Figura 14, apenas ocorrendo alteração no atributo que será comparado a classe, finalizando assim o cálculo da probabilidade para os atributos discretos.

3.1.2 Atributos Contínuos

Para o atributo Renda Anual não é possível realizar o mesmo processo detalhado anteriormente, pois este atributo é contínuo. No entanto, segundo Oliveira¹ e Pereira (2018), é possível realizar o tratamento deste atributo através da utilização de dois procedimentos:

1. Discretização do atributo contínuo.

2. Utilização de uma distribuição para calcular o valor da probabilidade do atributo.

Visando permanecer no escopo do deste trabalho, será utilizado a distribuição Gaussiana proposta pelos autores Pang-Ning, Steinbach e Kumar (2009), representada na Equação 3.2, para obtenção da probabilidade referente a renda anual.

$$P(X_i = x_i | Y = y_i) = \frac{1 \times \exp^{-\frac{1}{2} \times \left(\frac{x_i - \mu_{ij}}{\sigma_{ij}}\right)^2}}{\sqrt{2 \times \pi \times \sigma_{ij}^2}} \quad (3.2)$$

Onde σ^2 , representa a variância e μ , representa a média. Com o prosseguimento dos cálculos, temos que:

$$\begin{aligned} \mu &= \frac{125 + 100 + 70 + 120 + 60 + 220 + 75}{7} = 110 \\ \sigma^2 &= \frac{(125 - 110)^2 + (100 - 110)^2 + \dots + (75 - 110)^2}{7 \times 6} = 2975 \\ \sigma &\approx 54,54 \end{aligned}$$

Aplicado os valores de μ e σ na Formula 3.2 para $P(\text{renda anual} = 120 | \text{nao})$, temos:

$$P(\text{renda anual} = 120 | \text{nao}) = \frac{1 \times \exp^{-\frac{1}{2} \times \left(\frac{120 - 110}{54,54}\right)^2}}{\sqrt{2 \times \pi \times 2975}} \approx 0,0072$$

Para $P(\text{renda anual} = 120 | \text{sim})$, temos que $\mu = 90$ e $\sigma = 25$

$$P(\text{renda anual} = 120 | \text{sim}) = \frac{1 \times \exp^{-\frac{1}{2} \times \left(\frac{120 - 90}{25}\right)^2}}{\sqrt{2 \times \pi \times 25}} \approx 1,2 \times 10^{-9}$$

3.1.3 Aplicação do Classificador de Bayes Simples

Retomando o exemplo da Tabela 4, podemos aplicar a Fórmula 3.1 para predição do registro $P(X) = \{\text{proprietario casa} = \text{nao}, \text{estado civil} = \text{casado}, \text{renda anual} = 120k\}$

$$\begin{aligned} P(X | \text{sim}) &= (\text{proprietario de casa} = \text{nao} | \text{sim}) \times (\text{estado civil} = \text{casado} | \text{sim}) \\ &\quad \times (\text{renda anual} = 120 | \text{sim}) \end{aligned}$$

$$P(X | \text{sim}) = 1 \times 0 \times 1,2 \times 10^{-9} = 0$$

$$P(X|nao) = (\text{proprietario de casa} = \text{nao}|nao) \times (\text{estado civil} = \text{casado}|nao) \\ \times (\text{renda anual} = 120|nao)$$

$$P(X|nao) = 4/7 \times 4/7 \times 0,0074 = 0,0024$$

Para finalizar a predição é necessário calcular $P(nao|X)$ e $P(sim|X)$. O maior resultado indicará qual classe o registro será classificado. Como $P(Y|X) = \alpha \times P(X) \times P(X|Y)$, onde $\alpha = \frac{1}{P(X)}$, essa substituição ocorre pois o valor de $P(X)$ é constante em ambos os cálculos, então temos:

$$P(nao|X) = \alpha \times 7/10 \times 0,0024 = 0,0016\alpha$$

$$P(sim|X) = \alpha \times 3/10 \times 0 = 0$$

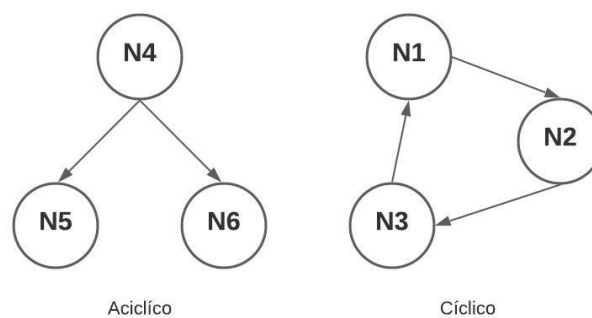
Como $P(nao|X) > P(sim|X)$ o registro será classificado como não inadimplente.

3.2 Classificador de Redes de Crenças Bayesianas

As Redes de Crenças Bayesianas, ou Redes Bayesianas, possibilita representar as relações de condicionalidade presentes entre as variáveis aleatórias, possibilitando assim que tais correlações não apenas sejam expressas como também calculadas.

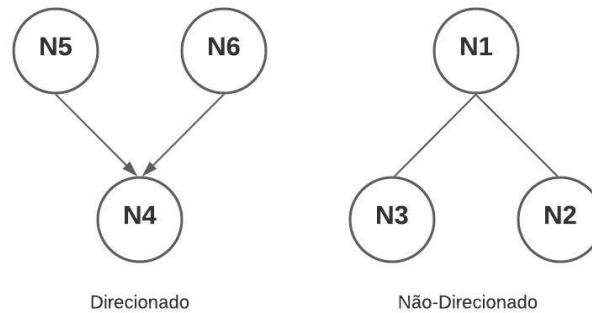
As Redes Bayesianas são representadas por um grafo acíclico direcionado. Acíclico conforme mostra a Figura 15 e direcionado conforme exibido na Figura 16.

Figura 15 – Exemplos de grafo acíclico e cíclico



Fonte: Elaborada pelo autor

Figura 16 – Exemplos de grafos direcionado e não direcionado

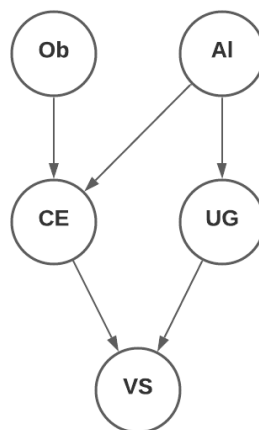


Fonte: Elaborada pelo autor

Nas figuras 15 e 16, cada nó, ou círculo, representa a variável aleatória, podendo ela ser um atributo ou uma classe. E as arestas, ou linhas, representam a relação entre as variáveis aleatórias. Dessa forma, na Figura 15, o nó N4 é pai dos nós N5 e N6, pois a aresta sai N4 e direciona-se para N5 e N6. Enquanto na Figura 16, N4 é filho de N5 e N6, pois ocorre o inverso.

Cada nó do grafo direcionado acíclico possui uma tabela de probabilidade que será utilizado nos cálculos do classificador de redes bayesianas para realizar as predições.

Figura 17 – Representação em grafo da relação câncer e úlcera



Fonte: Elaborada pelo autor

Conforme aponta Gonçalves (2008) como cada nó possui sua própria tabela de probabilidade é possível extrair dessas tabelas uma nova tabela de probabilidades relacionando todas essas informações, ou seja, uma tabela de probabilidade conjunta de todas as variáveis aleatórias. Os valores da tabela da probabilidade conjunta são obtidos através da Equação 3.3.

$$P(u) = \prod_{i=1}^n P(A_i|pa(A_i)) \quad (3.3)$$

$P(u)$ representa o produto das probabilidades condicionais de toda rede bayesiana, $P(A_i|pa(A_i))$ representa a probabilidade da variável aleatória A_i ocorrer, dado os pais dessa mesma variável.

O exemplo a seguir apresenta valores fictícios, todavia, no contexto deste trabalho, servirá para elucidação do funcionamento do classificador de redes bayesianas.

O grafo 17 é representado pelo seguintes nós: obesidade (Ob), alcoolismo (Al), câncer de estômago (CE), úlcera gástrica (UG) e vômito com sangue (VS) e todas as variáveis aleatórias possuem estados boolianos, ou seja, verdadeiro ou falso.

O exemplo da Figura 17 retrata o cenário onde um indivíduo, ao dar entrada em um hospital, realiza o preenchimento do formulário hospitalar para agilizar o processo de direcionamento do paciente e melhorar a precisão da identificação da aflição. Este formulário contém três perguntas com respostas do tipo “sim” e “não”, sendo elas:

1. Posso algum grau de obesidade?
2. Posso histórico de alcoolismo?
3. Apresentei sangue no vômito?

Tabela 9 – Probabilidade de Obesidade

Obesidade (Ob)	
sim	não
0,35	0,65

Fonte: Elaborada pelo autor

A Tabela 9 expõem as probabilidades de um indivíduo ser obeso ou não. É importante ressaltar que diferentemente do exposto no seção 2.5 onde a obesidade pode ser tratada como variável contínua, IMC, neste exemplo ela será tratada como uma variável binária.

Tabela 10 – Probabilidade de Alcoolismo

Alcoolismo (Al)	
sim	não
0,1	0,9

Fonte: Elaborada pelo autor

A Tabela 10 representa se o individuo possui problemas com alcoolismo ou não.

Tabela 11 – Probabilidade de Câncer de Estômago

Câncer de Estômago (CE)		
Pais (Ob, Al)	sim	não
nao, nao	0,05	0,95
nao, sim	0,15	0,85
sim, nao	0,25	0,75
sim, sim	0,85	0,15

Fonte: Elaborada pelo autor

A Tabela 11 representa as probabilidades do individuo apresentar câncer de estômago. Esta tabela se diferencia das Tabelas 9 e 10, pois a variável aleatória câncer de estômago é filho das variáveis obesidade e alcoolismo. Isso implica na ocorrência de oito eventos distintos decorrentes da correlação entre as variáveis câncer de estômago com obesidade e alcoolismo, em outras palavras, corresponde a todas as combinações possíveis das variáveis pais com a variável filho.

Tabela 12 – Probabilidade de Úlcera Gástrica

Úlcera Gástrica (UG)		
Pai (Al)	sim	não
nao	0,1	0,9
sim	0,8	0,2

Fonte: Elaborada pelo autor

A Tabela 12 representa a possibilidade do individuo apresentar úlcera gástrica, dado o histórico de alcoolismo. Similar a Tabela 11, úlcera gástrica é filha da variável alcoolismo e, por sua vez, resultará em uma tabela de probabilidade com quatro possibilidades, resultantes da combinação entre a variável pai, alcoolismo, e filho, úlcera gástrica.

Tabela 13 – Probabilidade de Vômito com Sangue

Vômito com Sangue (VS)		
Pais (CE, UG)	sim	não
nao, nao	0,05	0,95
nao, sim	0,65	0,35
sim, nao	0,25	0,75
sim, sim	0,85	0,15

Fonte: Elaborada pelo autor

A Tabela 13 retrata a possibilidade de o indivíduo apresentar sangue ao vomitar, estando esse sintoma relacionado a câncer de estômago e úlcera gástrica, em outras palavras, vômito com sangue é filho de câncer de estômago e úlcera gástrica.

As Tabelas 9, 10 e 13 são obtidas através do preenchimento do formulário pelo próprio paciente. Já as Tabelas 11 e 12 serão obtidas após realização e confirmação do diagnóstico do paciente.

Desta forma, pode-se interpretar o grafo modelado na Figura 17 do seguinte modo:

1. Obesidade e alcoolismo são agravadores das patologias câncer de estômago e úlcera gástrica, no entanto, enquanto obesidade afeta ambas doenças, alcoolismo só afeta úlcera gástrica.
2. Câncer de estômago e úlcera gástrica são as doenças que o médico deseja descobrir para aplicar o tratamento correto.
3. Vômito com sangue é um sintoma comum a ambas as afecções.

Com base na exposição das informações do exemplo e utilizando os conceitos apresentados na seção 3.1, para calcular a probabilidade de uma pessoa ter câncer de estômago, dado que não se sabe o histórico de obesidade ou alcoolismo, basta realizar o somatório de todas as probabilidades onde a pessoa possui câncer de estômago com todas as combinações das variáveis obesidade e alcoolismo. Esse somatório ocorre pois a ausência de informação, alcoolismo e obesidade, força a consideração de todos os casos onde o paciente apresentou câncer de estômago, uma vez que tanto obesidade e alcoolismo afetam diretamente a variável câncer de estômago.

$$\begin{aligned}
 P(CE = sim) &= \sum_i \sum_j P(CE = sim, Ob = x_i, Al = x_j) \times P(Ob = x_i) \times P(Al = x_j) \\
 P(CE = sim) &= (0,05 \times 0,65 \times 0,9) + (0,15 \times 0,65 \times 0,1) + \\
 &\quad (0,25 \times 0,35 \times 0,9) + (0,85 \times 0,35 \times 0,1) \\
 P(CE = sim) &= 0,1475
 \end{aligned}$$

Dessa forma, a probabilidade de uma pessoa contrair câncer de estômago, sem conhecimento prévio do histórico de alcoolismo e obesidade, é de 14,75%, por consequência, a chance dela não ter é de 85,25%, dado que $P(CE = nao) = 1 - P(CE = sim)$.

Por outro lado, introduzindo uma informação extra sobre o paciente, como o fato de o mesmo não ser obeso e mantendo a incerteza sobre o histórico de alcoolismo, os cálculos só levariam em consideração as combinações entre obesidade e alcoolismo na qual

a obesidade é falsa. Observando a Tabela 11 nota que apenas as linhas 3 e 4 retratam tais casos de falsidade na variável aleatória obesidade.

$$\begin{aligned}
 P(CE = sim|Ob = nao, Al = x) &= \\
 \sum_i P(CE = sim|Ob = nao, Al = i) \times P(Ob = nao) \times P(Al = i) &= \\
 0,05 \times 0,65 \times 0,9 + 0,15 \times 0,65 \times 0,1 &= \\
 0,02925 + 0,00975 &= \\
 P(CE = sim|Ob = nao, Al = x) = 0,039 = 3,9\% &
 \end{aligned}$$

Dessa forma, conclui-se que o simples fato de a pessoa não ser obesa reduz em mais de 70% a chance de a mesma contrair câncer de estômago.

No entanto, os dois exemplos descritos anteriormente, $P(CE = sim|Ob = i, Al = j)$ e $P(CE = sim|Ob = nao, Al = x)$, apesar de fazer parte de um grafo acíclico e direcionado, requerido pelas redes Bayesiana e conforme mostra a Figura 17, representam um caso de independência das variáveis aleatórias que podem ser solucionados por classificadores que requerem esse comportamento de independência, como é o caso do classificador Bayes Simples.

No entanto, se considerarmos um caso onde há dependência condicional entre as variáveis, tal como o de uma pessoa apresentar o sintoma de sangramento ao vomitar e ao efetuar um teste para saber se ela tem úlcera gástrica, o resultado obtido é negativo. Dessa forma, qual a probabilidade de o paciente ter câncer de estômago? Em outras palavras, $P(CE = sim|VS = sim, UG = nao)$.

Para solucionar esse problema será utilizada a equação 3.3 correspondente a distribuição da probabilidade conjunta. Aplicando a equação 3.3 ao grafo da Figura 17, obtemos a probabilidade conjunta do grafo, expresso na equação 3.4:

$$P(u) = P(Ob) \times P(Al) \times P(CE|Ob, Al) \times P(UG|Al) \times P(VS|CE, UG) \quad (3.4)$$

Manipulando $P(CE = sim|VS = sim, UG = nao)$ para melhor adequar a probabilidade conjunta, obtém-se:

$$\begin{aligned}
 P(CE = sim|VS = sim, UG = nao) &= \\
 \frac{P(VS = sim, UG = nao|CE = sim) \times P(CE = sim)}{P(VS = sim, UG = nao)} &= \\
 \frac{P(CE = sim, VS = sim, UG = nao)}{P(VS = sim, UG = nao)} &
 \end{aligned} \quad (3.5)$$

Dessa forma, para descobrir a probabilidade $P(CE = sim|VS = sim,UG = nao)$, temos que calcular a probabilidade conjunta de $P(CE = sim, VS = sim, UG = nao)$ que corresponde aos eventos câncer de estômago ocorrer, vômito com sangue ocorrer e úlcera gástrica não ocorrer, dividido pela probabilidade conjunta de vômito com sangue ocorrer com úlcera gástrica não ocorrer, $P(VS = sim, UG = nao)$.

Antes de demonstrar todos os valores da probabilidade conjunta conforme mostram as Tabelas 16 e 17, será explicitado como se calcula o valor de uma célula da tabela, tendo em vista que as demais células seguirão a mesma linha de cálculo.

Sendo assim, é necessário obter as seguintes informações:

- Probabilidade Conjunta do Grafo dado por: $P(Ob) \times P(Al) \times P(CE|Ob, Al) \times P(UG|Al) \times P(VS|CE, UG)$
- O valores correspondente a: $P(CE = sim, VS = sim, UG = nao)$
- O valores correspondente a: $P(VS = sim, UG = nao)$

A Tabela 14, referente ao primeiro item, $P(Ob) \times P(Al) \times P(CE|Ob, Al) \times P(UG|Al) \times P(VS|CE, UG)$, será inicialmente apresentada sem valores, visando facilitar o entendimento do processo de preenchimento das probabilidades conjuntas. Além disso, Os valores referentes a $P(CE = sim)$ e $P(CE = nao)$ serão expostos em tabelas distintas para facilitar a visualização dos procedimentos matemáticos que serão realizados posteriormente.

Tabela 14 – Probabilidade conjunta de $P(CE = sim)$ da Figura 17

Pais (Ob, Al)	UG = nao, VS = nao	UG = nao, VS = sim	UG = sim, VS = nao	UG = sim, VS = sim
não, não				
não, sim				
sim, não				
sim, sim				

Fonte: Elaborada pelo autor

A Tabela 14 relaciona as cinco variáveis aleatórias combinando todas as possibilidades de valores entre as variáveis. Como deseja saber $P(CE = sim, VS = sim, UG = nao)$ e como não foi informado $P(Ob)$ e $P(Al)$, então deve ser considerado todos os casos onde $Ob = \{sim\}$ e $Ob = \{nao\}$ e também $Al = \{sim\}$ e $Al = \{nao\}$.

A primeira célula a ser calculada será a que se refere a: $P(Ob = nao) \times P(Al = nao) \times P(CE = sim|Ob = nao, Al = nao) \times P(UG = nao|Al = nao) \times P(VS = sim|CE = sim, UG = nao)$.

- $P(Ob = nao)$: De acordo com a Tabela 9, a probabilidade de um paciente não ser obeso é de 0,65.
- $P(Al = nao)$: De acordo com a Tabela 10, a probabilidade de um paciente não ser alcoólatra é de 0,9.
- $P(CE = sim|Ob = nao, Al = nao)$: De acordo com a tabela 11, a probabilidade de um paciente ter câncer de estômago dado que ele não é obeso e nem alcoólatra é de 0,05.
- $P(UG = nao|Al = nao)$: De acordo com a Tabela 12, a probabilidade de um paciente não ter úlcera gástrica dado que ele não é alcoólatra é de 0,9.
- $P(VS = sim|CE = sim, UG = nao)$: De acordo com a Tabela 13, a probabilidade de um paciente ter vômito com sangue dado que ele possui câncer de estômago e não possui úlcera gástrica é de 0,25.

Aplicando a formula 3.3 aos cinco itens acima, temos que:

$$\begin{aligned}
 P(u) &= P(Ob = nao) \times P(Al = nao) \times P(CE = sim|Ob = nao, Al = nao) \\
 &\quad \times P(UG = nao|Al = nao) \times P(VS = sim|CE = sim, UG = nao) = \\
 &= 0,65 \times 0,9 \times 0,05 \times 0,9 \times 0,25 = 0,00658125
 \end{aligned}$$

O resultado será inserido na segunda linha e terceira coluna interna da Tabela 15, uma vez que esta posição corresponde as combinações referentes a $P(Ob = nao) \times P(Al = nao) \times P(CE = sim|Ob = nao, Al = nao) \times P(UG = nao|Al = nao)$.

Tabela 15 – Probabilidade Conjunta de $P(CE = sim, UG = nao, VS = sim, Ob = nao, Al = nao)$

Pais (Ob, Al)	UG = nao, VS = nao	UG = nao, VS = sim	UG = sim, VS = nao	UG = sim, VS = sim
não, não		0,00658125		
não, sim				
sim, não				
sim, sim				

Fonte: Elaborada pelo autor

Realizando os demais cálculos para o restante da tabela, obtemos os valores conforme mostram as Tabelas 16 e 17, representando os valores para $CE = \{sim\}$ e $CE = \{nao\}$, respectivamente.

Tabela 16 – Probabilidade Conjunta do Grafo do grafo 17 para $CE = \{sim\}$

Pais (Ob, Al)	UG = nao, VS = nao	UG = nao, VS = sim	UG = sim, VS = nao	UG = sim, VS = sim
não, não	0,01974375	0,00658125	0,00043875	0,00248625
não, sim	0,0014625	0,0004875	0,00117	0,00663
sim, não	0,05315625	0,01771875	0,00118125	0,00669375
sim, sim	0,0044625	0,0014875	0,00357	0,02023

Fonte: Elaborada pelo autor

Tabela 17 – Probabilidade Conjunta do Grafo do grafo 17 para $CE = \{nao\}$

Pais (Ob, Al)	UG = nao, VS = nao	UG = nao, VS = sim	UG = sim, VS = nao	UG = sim, VS = sim
não, não	0,47516625	0,02500875	0,01945125	0,03612375
não, sim	0,0104975	0,0005525	0,01547	0,02873
sim, não	0,20199375	0,01063125	0,00826875	0,01535625
sim, sim	0,0009975	0,0000525	0,00147	0,00273

Fonte: Elaborada pelo autor

Com as Tabelas 16 e 17 preenchidas, é possível extrair os valores para solucionar a equação 3.5.

Tabela 18 – Probabilidade de $P(CE = sim, UG = nao, VS = sim)$

Pais (Ob, Al)	UG = nao, VS = nao	UG = nao, VS = sim	UG = sim, VS = nao	UG = sim, VS = sim
não, não	0,01974375	0,00658125	0,00043875	0,00248625
não, sim	0,0014625	0,0004875	0,00117	0,00663
sim, não	0,05315625	0,01771875	0,00118125	0,00669375
sim, sim	0,0044625	0,0014875	0,00357	0,02023

Fonte: Elaborada pelo autor

Os valores em negrito da Tabela 18 representam a probabilidade condicionada de todas as combinações de $P(CE = sim, UG = nao, VS = sim)$. É realizado o somatório de todos os valores da coluna em virtude de não sabermos se o paciente tem histórico de alcoolismo ou obesidade, dessa forma:

$$P(CE = sim, UG = nao, VS = sim) = 0,00658125 + 0,0004875 \\ + 0,01771875 + 0,0014875 = 0,026275$$

Também é possível obter o valor de $P(UG = nao, VS = sim)$, conforme mostra os valores em negrito das Tabelas 18 e 19.

Tabela 19 – Probabilidade de $P(CE = nao, UG = nao, VS = sim)$

Pais (Ob, Al)	UG = nao, VS = nao	UG = nao, VS = sim	UG = sim, VS = nao	UG = sim, VS = sim
não, não	0,47516625	0,02500875	0,01945125	0,03612375
não, sim	0,0104975	0,0005525	0,01547	0,02873
sim, não	0,20199375	0,01063125	0,00826875	0,01535625
sim, sim	0,0009975	0,0000525	0,00147	0,00273

Fonte: Elaborada pelo autor

Os valores de $P(CE = nao)$ são levados em consideração nesse cálculo, pois a probabilidade de $P(UG = nao, VS = sim)$ desconsidera o valor de $P(CE)$, dessa forma:

$$P(UG = nao, VS = sim) = 0,00658125 + 0,02500875 + 0,0004875 + 0,0005525 \\ + 0,01771875 + 0,01063125 + 0,0014875 + 0,0000525 = 0,06252$$

Por fim, substituindo os valores de $P(CE = sim, VS = sim, UG = nao)$ e $P(VS = sim, UG = nao)$ na equação 3.5, obtém-se:

$$\frac{P(CE = sim, VS = sim, UG = nao)}{P(VS = sim, UG = nao)} = \\ \frac{0,026275}{0,06252} = 0,420265515 \approx 42,02\%$$

Ou seja, a probabilidade do paciente possuir câncer de estômago, dado que o mesmo não possui úlcera gástrica e apresentou o sintoma de vômito com sangue é igual a 42,02%.

A adição de uma informação extra a predição desse classificador, como o fato de a paciente não ter histórico de obesidade nem de alcoolismo, $P(Ob = nao)$ e $P(Al = nao)$, bastaria analisar os seguintes valores em negrito das Tabelas 20 e 21.

Tabela 20 – Probabilidade Conjunta de $P(Ob = nao, Al = nao, CE = sim, UG = nao, VS = sim)$

Pais (Ob, Al)	UG = nao, VS = nao	UG = nao, VS = sim	UG = sim, VS = nao	UG = sim, VS = sim
não, não	0,01974375	0,00658125	0,00043875	0,00248625
não, sim	0,0014625	0,0004875	0,00117	0,00663
sim, não	0,05315625	0,01771875	0,00118125	0,00669375
sim, sim	0,0044625	0,0014875	0,00357	0,02023

Fonte: Elaborada pelo autor

Tabela 21 – Probabilidade Conjunta de $P(Ob = nao, Al = nao, CE = nao, UG = nao, VS = sim)$

Pais (Ob, Al)	UG = nao, VS = nao	UG = nao, VS = sim	UG = sim, VS = nao	UG = sim, VS = sim
não, não	0,47516625	0,02500875	0,01945125	0,03612375
não, sim	0,0104975	0,0005525	0,01547	0,02873
sim, não	0,20199375	0,01063125	0,00826875	0,01535625
sim, sim	0,0009975	0,0000525	0,00147	0,00273

Fonte: Elaborada pelo autor

Manipulando novamente o teorema de bayes, porém inserindo as novas informações adicionadas ao experimento, tem-se:

$$\begin{aligned}
 &P(CE = sim|Ob = nao, Al = nao, VS = sim, UG = nao) = \\
 &\frac{P(Ob = nao, Al = nao, VS = sim, UG = nao|CE = sim) \times P(CE = sim)}{P(Ob = nao, Al = nao, VS = sim, UG = nao)} = \\
 &\frac{P(Ob = nao, Al = nao, CE = sim, VS = sim, UG = nao)}{P(Ob = nao, Al = nao, VS = sim, UG = nao)}
 \end{aligned}$$

$P(Ob = nao, Al = nao, CE = sim, VS = sim, UG = nao)$ foi calculado e explicado previamente na Tabela 15. $P(Ob = nao, Al = nao, VS = sim, UG = nao)$ representa a probabilidade conjunta de todas as variáveis aleatórias, exceto a variável câncer de estômago. Esse valor é obtido através do somatório dos valores em negrito das Tabelas 20 e 21.

$$\begin{aligned}
 &P(Ob = nao, Al = nao, CE = sim, VS = sim, UG = nao) = 0,00658125 \\
 &P(Ob = nao, Al = nao, VS = sim, UG = nao) = 0,00658125 + 0,02500875 = 0,03159 \\
 &\frac{P(Ob = nao, Al = nao, CE = sim, VS = sim, UG = nao)}{P(Ob = nao, Al = nao, VS = sim, UG = nao)} = \frac{0,00658125}{0,03159} \approx 20,83
 \end{aligned}$$

Dessa forma, ao inserir duas informações adicionais ao exemplo, obesidade e alcoolismo, o valor da predição do classificador alterou, uma vez que, havendo a relação de condicionalidade entre as variáveis, o acréscimo de informação afeta a predição do classificador de redes Bayesiana.

3.3 Características do Bayes Simples e das Redes Bayesianas

Considerando as descrições dos dois modelos de classificação apresentados neste capítulo, bem como os exemplos utilizados para ilustrar suas aplicações, podemos, nesta

seção, discorrer sobre algumas características relevantes a cada um deles. Iremos destacar duas características: valores faltantes e variável condicional.

Os valores faltantes correspondem a casos de incerteza em uma ou mais variáveis aleatórias, ou seja, não se sabe o valor de um dos atributos responsáveis por ajudar o classificador a prever a classe.

Já as variáveis condicionais representam a existência, ou ausência, de uma condição que afeta as variáveis aleatórias. Citando o exemplo apresentado na seção 3.1, temos que os três atributos expostos na Tabela 4, proprietário de casa, estado civil e renda anual, não afetam a probabilidade um do outro, por essa razão são tratados como variáveis aleatórias independentes.

Por outro lado, a dependência entre as variáveis aleatórias ocorre quando um atributo afeta a probabilidade de outro atributo, conforme notado na Figura 17, onde a variável alcoolismo afeta câncer de estômago que por sua vez afeta vômito com sangue.

Classificador de Bayes Simples

- Valores Faltantes
 - Pode lidar com valores que estão faltando, ignorando o valor ausente durante a classificação do modelo. Isto é evidenciado, apesar de não exposto especificamente, nas Tabelas 5, 6, 7 e 8, referentes a possibilidade de inadimplência após receber um empréstimo. Nessas 4 tabelas apenas os valores referentes ao que deseja ser predito são utilizados, reduzindo assim o espaço amostral total, dessa forma, se houvessem registros com valores faltantes na base de dados, tais registros seriam desconsiderados para fins da predição.
 - Caso o valor ausente esteja presente no registro que será classificado ao invés da base de dados, o comportamento do classificador será distinto conforme foi implementado. Duas abordagens serão sugeridas: a primeira condiz com a exclusão do atributo durante o cálculo do classificador, isso é possível por causa da natureza independente dos atributos. A segunda é similar ao que é realizado pela rede Bayesiana, onde a incerteza sobre o atributo será tratado como sendo todas as possibilidades que o atributo pode assumir. Essas duas abordagens que são sugeridas dependem de cada caso onde o classificador será aplicado, por exemplo, se a variável aleatória tiver grande relevância para o processo de predição, a total remoção deste atributo pode gerar resultados negativos. Caso contrário, sua remoção pouco afetará o resultado final do classificador.
- Variáveis Condicionais
 - Atributos correlacionados podem degradar o desempenho de classificadores Bayes simples porque a suposição de independência condicional não é mais

verdadeira para tais atributos. Isto pode ser evidenciado no exemplo da Tabela 4 onde todos os atributos são independentes, apenas se relacionando com a classe.

Classificador de Redes Bayesianas

- Valores Faltantes
 - São apropriadas para lidar com dados incompletos. Instâncias com atributos em falta podem ser manipuladas através de soma ou integração de probabilidade por todos os valores possíveis do atributo. Esta característica é evidenciada em todos os exemplos expostos na seção 3.2, tendo em vista que ao não apresentar uma informação, o classificador considerava todos os casos possíveis.
- Variáveis Condicionais
 - Permite a representação de condicionalidade entre as variáveis aleatórias, possibilitando assim uma flexibilidade na simulação de um determinado experimento onde os atributos não são independentes. O grafo representado na Figura 17 (representação em grafo da relação câncer e úlcera) demonstra a existência de dependências entre as variáveis aleatórias em diferentes níveis do grafo.

4 Considerações Finais

Este trabalho abordou dois classificadores probabilísticos, Bayes simples e redes Bayesianas. Para tal apresentou conceitos probabilísticos básicos, dando enfoque a probabilidade condicional e ao teorema de Bayes.

Apesar da similaridade matemática entre a probabilidade condicional e o teorema de Bayes, representados por $P(A|B)$, ambos abordam o mesmo problema sob dois aspectos diferentes. O primeiro trata que a condição que afeta a probabilidade ocorre antes do que deseja ser previsto, ou seja, representa a probabilidade posteriori. Já o teorema de Bayes realiza o processo inverso, assumindo “uma certeza” sobre o evento que sucede o que deseja ser previsto, representando a probabilidade priori. Uma outra forma de explicar o mesmo conceito seria na seguinte leitura de $P(A|B)$: para a probabilidade condicional, B ocorre antes de A, enquanto para o teorema de Bayes, B ocorre depois de A.

Esta abordagem trazida pelo teorema de Bayes embasa a forma de predição dos classificadores Bayesianos apresentados neste trabalho, que apesar de similares neste aspecto, apresentam diferentes características que podem ser melhor exploradas em diferentes cenários.

O classificador Bayesiano simples trata as variáveis aleatórias, ou atributos, com um caráter de independência, ou seja, o valor de uma variável aleatória não influi no valor das demais, sendo a única exceção a variável aleatória da classe, uma vez que ela é a variável que o classificador irá prever. Isso pode ser notado no exemplo apresentado na seção 3.1, onde deseja saber se uma pessoa, ao pegar um empréstimo, se tornará inadimplente ou não, dado um conjunto de informações a respeito desta pessoa, como o estado civil, proprietário de casa e renda anual.

Já o classificador de redes Bayesianas, permite criar relações de dependência entre as variáveis aleatórias, ou seja, uma variável aleatória pode impactar em outras variáveis aleatórias, não necessariamente apenas na classe. Essa relação de dependência possibilita a modelagem de sistemas que exijam tais relações, como é o caso apresentado na seção 3.2, onde o estilo de vida de uma pessoa, alcoolismo e má alimentação, são agravadores de patologias, câncer de estômago e úlcera gástrica, que por sua vez tem como um dos sintomas a presença de sangue ao vomitar.

Dessa forma, o classificador Bayesiano simples possui melhor aplicabilidade em cenários onde as variáveis aleatórias são independentes, tornando desnecessário o esforço para a criação de um grafo, tendo em vista que o classificador não precisará criar uma tabela de probabilidade conjunta conforme realizado pelo classificador de redes Bayesianas.

Por outro lado, o classificador de redes Bayesiana se torna mais adequado em situações onde é necessário expressar dependência entre os atributos, como por exemplo as relações de causa, efeito e consequência. Apesar disso, o processo se torna mais custoso tendo em vista que exigem a modelagem de um grafo para representação destas relações de dependência, conforme Figura 17, dando margem para que possíveis erros de representação afetem negativamente a predição do classificador.

Pautando-se em toda base de estudo necessário para o desenvolvimento deste trabalho, foi notado que os autores que apresentam informações referentes a classificadores, podem assumir que seus leitores, por optarem por acessar tais estudos, já possuem todo o embasamento necessário para compreensão do que será apresentado. Por outro lado, autores que adotem um caráter mais didático tendem a não abordar os aspectos de classificadores, uma vez que tal assunto possui uma complexidade natural.

Este trabalho entrega um conjunto de conhecimentos probabilísticos básicos necessários para o entendimento teórico dos classificadores Bayes simples e redes Bayesianas, além do processo matemático realizado por ambos, não tendo como finalidade afirmar a superioridade entre os dois classificadores.

Como sugestão de continuidade para trabalhos futuros, sugere-se:

- Exibição do funcionamento do classificador Bayesiano em uma simulação prática, abordando os aspectos referentes aos possíveis erros, e suas tratativas, de predição do classificador, além da possibilidade de otimizações no processo de predição.
- Abordar técnicas para modelagem das variáveis aleatórias em grafos acíclicos e direcionados, uma vez que a relação de dependência entre as variáveis impactam no resultado final do classificador.

Referências

- CARUZO, A. de C.; NEVES, K. F.; TEIXEIRA, R. R. P. A entropia e o triângulo de pascal. *Revista tecnológica*, v. 16, p. 11–25, 2007. Citado na página 18.
- COMBINATÓRIA, A. Probabilidade. *Rio de Janeiro: SBM*, 1991. Citado 2 vezes nas páginas 14 e 23.
- CORREA, S. *Probabilidade e estatística*. [S.l.]: Belo Horizonte: PUC Minas Virtual, 2003. Citado na página 19.
- DANTAS, C. A. B. *Probabilidade: Um Curso Introdutório Vol. 10*. [S.l.]: Edusp, 2013. Citado na página 14.
- D'ANALYSES, B. d'Enquêtes et. *Flight AF447 - Final Report*. 2012. 83-84 p. Acesso em: 01 de Dezembro de 2021. Disponível em: <<https://www.bea.aero/docspa/2009/f-cp090601.en/pdf/f-cp090601.en.pdf>>. Citado na página 12.
- FALCO, J. G. Estatística aplicada. 2016. Citado na página 14.
- FARIAS, A. M. L. de. Apostila de estatística ii. 2017. Citado na página 22.
- GONÇALVES, A. R. Redes bayesianas. *Campinas: Unicamp*, 2008. Citado na página 37.
- MARTINS, M. E. G. Probabilidade. 2014. Citado na página 17.
- NOVAES, D. V.; QUEIROZ, C. de; COUTINHO, S. *Estatística para educação profissional*. [S.l.]: Editora Atlas SA, 2000. Citado na página 16.
- OLIVEIRA¹, G. S. de; PEREIRA, S. do L. Influência da discretização de atributos na acurácia da classificação bayesiana. *Boletim Técnico*, v. 45, p. 33, 2018. Citado na página 34.
- PANG-NING, T.; STEINBACH, M.; KUMAR, V. Introdução ao “data mining”-mineração de dados.”. *Rio de Janeiro: Editora Ciência Moderna*, 2009. Citado 6 vezes nas páginas 28, 29, 32, 33, 34 e 35.
- PONTIFÍCIA, E. da. *PROBABILIDADE CONDICIONAL*. Tese (Doutorado) — Pontifícia Universidade Católica de São Paulo, 2000. Citado na página 24.
- PORTNOI, M. Probabilidade, variáveis aleatórias, distribuição de probabilidades e geração aleatória. *Salvador: Unifacs*, 2005. Citado na página 20.
- REIS, E.; REIS, I. Avaliação de testes diagnósticos. *Técnico do Departamento de Estatística da Universidade Federal de Minas Gerais. Disponível em: www.est.ufmg.br. Acesso em*, v. 1, 2002. Citado na página 16.
- ROSSATTO, F. C.; DAMMANN, J.; KAMPHORST, E. M.; KAMPHORST, C. H.; DONADEL, A. P. D. P. Teorema de bayes: Estudo e aplicação. *Salão do Conhecimento*, 2016. Citado na página 25.

SILVA, A. I. da. O imc e o perímetro da cintura como indicadores de risco para a saúde de árbitros de futebol do brasil. *Fitness & performance journal*, Colégio Brasileiro de Atividade Física, Saúde e Esporte, n. 4, p. 223–231, 2006. Citado na página 22.

SOUZA, R. A. de. Probabilidade: Uma visão olímpica. *Cadernos Cajuína*, v. 5, n. 1, p. 148–164, 2020. Citado na página 19.

STONE, L. D.; KELLER, C. M.; KRATZKE, T. M.; STRUMPFER, J. P. Search for the Wreckage of Air France Flight AF 447. *Statistical Science*, Institute of Mathematical Statistics, v. 29, n. 1, p. 69 – 80, 2014. Disponível em: <<https://doi.org/10.1214/13-STS420>>. Citado na página 12.

YNOGUTI, C. A. Probabilidade, estatística e processos estocásticos. *Apostilha*. Universidade de São Paulo, 2011. Citado na página 15.