

Kevin Mota da Costa

Explorando a Cosmologia com Machine Learning: Redes Neurais e Processos Gaussianos na Análise de Dados Cósmicos

Vitória/ES
Dezembro/2023

Kevin Mota da Costa

Explorando a Cosmologia com Machine Learning: Redes Neurais e Processos Gaussianos na Análise de Dados Cómicos

Monografia apresentada ao curso de Física da Universidade Federal do Espírito Santo - campus Goiabeiras - como requisito essencial para a obtenção do grau de Bacharelado em Física

Universidade Federal do Espírito Santo - UFES
Centro de Ciências Exatas - CCE
Departamento de Física - DFIS

Orientador: Alan Miguel Velazques Toribio

Vitória/ES
Dezembro/2023

Kevin Mota da Costa

Explorando a Cosmologia com Machine Learning: Redes Neurais e Processos Gaussianos na Análise de Dados Cómicos

Monografia apresentada ao curso de Física da Universidade Federal do Espírito Santo - campus Goiabeiras - como requisito essencial para a obtenção do grau de Bacharelado em Física

Aprovado em:

Comissão examinadora:

Prof. Dr. Alan Miguel Velazques Toríbio
Orientador

Prof. Dr. Júlio César Fábris
Universidade Federal do Espírito Santo

Prof. Dr. Luiz Filipe Guimarães
Universidade Federal do Espírito Santo

Vitória/ES

Agradecimentos

Quero expressar minha profunda gratidão a todas as pessoas que contribuíram para a realização deste trabalho. Em especial, dedico palavras de agradecimento a

Meu namorado, **Emanuel Barbosa de Araújo**, por seu constante apoio, paciência e incentivo ao longo desta jornada acadêmica. Sua presença foi fundamental para superar desafios e celebrar conquistas.

Minha querida amiga de curso, **Mila Leite Garcia Reis**, pela amizade sincera, colaboração e compartilhamento de experiências. Sua presença tornou esta jornada mais leve e enriquecedora.

Meu orientador, **Alan Miguel Velasquez Toríbio**, pela orientação precisa, conselhos valiosos e dedicação incansável. Sua expertise e apoio foram essenciais para o desenvolvimento deste trabalho.

À minha família, que sempre esteve ao meu lado, oferecendo amor incondicional, apoio emocional e encorajamento. Vocês são a base sólida que sustenta minhas realizações, e cada conquista é compartilhada com vocês. Muito obrigado por serem a fonte inesgotável de força e inspiração em minha vida.

Agradeço a todos que, de alguma forma, contribuíram para o sucesso deste projeto. Cada palavra de incentivo, cada momento compartilhado e cada desafio superado moldaram este caminho de aprendizado e crescimento.

Muito obrigado!

Kevin Mota da Costa.

”Mas lembrar-se com saudade é como se despedir de novo.”, Clarice Lispector.

Resumo

Neste estudo, exploramos abordagens estatísticas e de inteligência artificial para a análise de dados. No contexto da inteligência artificial, enfatizamos a utilização de redes neurais e do método dos processos gaussianos para reconstruir dados cosmológicos. Através dos resultados obtidos, demonstramos a aplicabilidade e eficácia dessas técnicas na reconstituição de dados observacionais. Além disso, empregamos o método das cadeias de Markov Monte Carlo para estabelecer vínculos observacionais sobre os parâmetros de densidade de curvatura, densidade de matéria do universo e a constante de Hubble.

Palavras-chave: Inteligência artificial, Redes neurais, Processos gaussianos, cosmologia.

Listas de Figuras

2.1	Histograma de 10^6 amostras de uma distribuição gaussiana com média 0 e desvio padrão 0,01 para $\Delta x_i = 0,01$. Fonte: Criado pelo autor.	11
2.2	Histograma de 10^6 amostras de uma distribuição gaussiana com média 0 e desvio padrão 0,01 para $\Delta x_i = 0,005$. Fonte: Criado pelo autor.	11
2.3	Visualização da correspondência entre um histograma normalizado e uma pdf.	12
2.4	Visualização da correlação entre duas variáveis aleatórias	18
3.1	Representação das principais estruturas do cérebro humano. Fonte: Brasil Escola [9].	26
3.2	Diagrama de um perceptron genérico. Fonte: Criado pelo autor.	28
4.1	Função de média para o processo gaussiano. Fonte: Criado pelo autor.	43
4.2	Função de covariância para o processo gaussiano. Fonte: Criado pelo autor.	44
4.3	Distribuição à priori para \mathbf{x}^* . Fonte: Criado pelo autor.	44
4.4	Distribuição à posteriori para \mathbf{x}^* . Fonte: Criado pelo autor.	45
4.5	Previsão para \mathbf{x}^* . Fonte: Criado pelo autor.	45
4.6	Comparação das regressões pelo método dos processos gaussianos para diferentes hiperparâmetros.	47
5.1	Triângulo definido por três galáxias em um universo expandindo uniformemente. Fonte: Criado pelo autor.	54
5.2	Ilustração de uma pessoa segurando uma lanterna dentro de um foguete acelerado no espaço, vista de um observador externo. Fonte: Criado pelo autor.	63
5.3	Ilustração de uma pessoa segurando uma lanterna dentro de uma cabine sob influência de um campo gravitacional uniforme, vista de um observador externo. Fonte: Criado pelo autor.	64
6.1	Dados de supernova da amostra Pantheon. Fonte: Colaboração Pan-STARRS [22].	80
6.2	Dados de quasares. Fonte: Lusso et. al. [28].	81

6.3	Dados de galáxias HII. Fonte: Cao et. al. [4].	82
6.4	Dados do parâmetro de Hubble. Fonte: Marek et. al. [13].	83
7.1	Reconstrução dos dados de Galáxias HII obtida através do método dos processos gaussianos. Fonte: Autor próprio.	87
7.2	Reconstrução dos dados de parâmetro de Hubble obtida através do método dos processos gaussianos. Fonte: Autor próprio. [13].	88
7.3	Reconstrução dos dados de supernovas Ia obtida através do método dos processos gaussianos. Fonte: Autor próprio.	89
7.4	Reconstrução dos dados de quasares obtida através do método dos processos gaussianos. Fonte: Autor próprio.	90
7.5	Reconstrução dos dados de Galáxias HII obtida através das redes neurais. Fonte: Autor próprio.	91
7.6	Reconstrução dos dados de parâmetro de Hubble obtida através das redes neurais. Fonte: Autor próprio. [13].	92
7.7	Reconstrução dos dados de supernovas Ia obtida através das redes neurais. Fonte: Autor próprio.	93
7.8	Reconstrução dos dados de quasares obtida através das redes neurais. Fonte: Autor próprio.	94
7.9	Comparação entre os resultados obtidos para a inferência estatística dos dados do parâmetro de Hubble para os dados reconstruídos e os dados obtidos diretamente da natureza. Fonte: Autor próprio.	95

Lista de Tabelas

3.1	Saídas do perceptron para a operação lógica OR	28
3.2	Saídas do perceptron para a operação lógica AND	29
3.3	Resultados do teste da Rede neural com a função de ativação ReLU.	34
3.4	Resultados do teste da Rede neural com a função de ativação Gelu.	34
7.1	Comparação entre os resultados obtidos para a inferência estatística dos dados do parâmetro de Hubble para os dados reconstruídos e os dados obtidos diretamente da natureza	95

Conteúdo

1	Introdução	1
2	Probabilidade e estatística para a análise de dados	4
2.1	Os axiomas de Kolmogorov	5
2.2	Interpretação bayesiana da teoria das probabilidades	7
2.3	Distribuições de probabilidade	9
2.3.1	Marginalização e pdf conjunta	12
2.4	Medidas de tendência central: quantis e valores esperados	14
2.4.1	Valores esperados	15
2.5	Verossimilhança	19
2.5.1	Qui-quadrado	20
2.6	Inferência estatística	21
2.6.1	Matriz de Fisher	21
2.6.2	Inferência de máxima verossimilhança	22
2.6.3	Cadeias de Markov Monte Carlo	23
3	Redes neurais	26
3.1	O perceptron	28
3.2	A primavera da inteligência artificial: redes neurais modernas	30
3.3	Função de ativação	32
3.4	Solucionando um problema curioso usando redes neurais	33
4	Método dos processos gaussianos	36
4.1	A distribuição gaussiana	37
4.1.1	Propriedades da distribuição gaussiana: distribuição conjunta, marginal e condicional	38
4.1.2	A matriz de covariância e o vetor de média	38
4.2	Os processos gaussianos	39
4.3	Realizando previsões com o método dos processos gaussianos	40
4.3.1	Likelihood para o Método dos Processos Gaussianos	41
4.4	Usando o método dos processos gaussianos para regressão bayesiana	42

4.4.1	A função de covariância e a função de média	45
4.4.2	Processos gaussianos para a previsão das derivadas	48
5	O modelo cosmológico padrão	51
5.1	Métrica	51
5.2	Princípio cosmológico	53
5.3	Lei de Hubble-Lemaître	53
5.4	A métrica de Minkowski	56
5.5	Curvatura	57
5.6	A métrica de Friedmann-Lemaître-Robertson-Walker	59
5.6.1	Distância própria e o redshift cosmológico	60
5.7	A Teoria da Relatividade Geral de Einstein	63
5.7.1	O tensor de Einstein	65
5.7.2	O Tensor de energia-momento	65
5.7.3	A Constante Cosmológica	66
5.8	As equações de Friedmann	66
5.9	Cosmologia observacional: módulo de distância e distância de luminosidade	71
5.9.1	Distância de Luminosidade	72
5.9.2	Módulo de distância	73
5.10	Cosmografia	74
6	Vínculos observacionais sobre o modelo cosmológico padrão	78
6.1	Quasares, supernovas, galáxias HII e relógios cósmicos	79
6.1.1	Supernovas	79
6.1.2	Quasares	80
6.1.3	Galáxias HII	81
6.1.4	Oscilações acústicas da matéria bariônica e os cronômetros cósmicos	82
6.1.5	Método dos processos gaussianos e redes neurais: aplicação na linguagem de programação Python	83
6.1.6	MCMC: aplicação na linguagem de programação python	84
7	Discussão	86
8	Conclusão	97
9	Anexos	99
9.1	Algoritmos	99
9.1.1	Código em python para a exemplificação de uma rede neural	99
9.1.2	Código em python para a exemplificação do método dos processos gaussianos para regressão	103

1 Introdução

A cosmologia representa a teoria física que aborda o universo em sua totalidade. Trata-se de uma teoria robusta, essencial para o progresso científico e tecnológico da humanidade. No âmbito da cosmologia, são explorados diversos aspectos relacionados à evolução e composição do universo. Apesar de ser uma teoria sólida que possibilitou a compreensão de vários fenômenos, a cosmologia é uma das áreas da física permeada por grandes mistérios.

Esses mistérios estão intrinsecamente ligados a uma série de descobertas que nos conduziram a uma diversidade de fenômenos ainda intrigantes para a mente dos físicos até os dias atuais. Com o avanço de diversas técnicas observacionais, fomos capazes de desvendar uma gama de fenômenos que desafiam nosso entendimento sobre o universo.

Esses enigmas estão associados a uma variedade de eventos observados ao longo dos últimos anos, como a acelerada expansão do universo, cuja explicação remete à enigmática energia escura, as anisotropias na radiação cósmica de fundo, a natureza da matéria escura e a tensão na constante de Hubble. Essas problemáticas se entrelaçaram com as diversas técnicas observacionais, proporcionando uma quantidade imensa de dados. Estes dados, por sua vez, instigam questionamentos sobre vários aspectos que presumíamos já terem sido resolvidos.

Nesse contexto, onde a chamada cosmologia de precisão está cada vez mais sendo realizada, é imperativo obtermos uma compreensão profunda da natureza por trás dos dados coletados. Especialmente nas últimas décadas, diversas técnicas estatísticas foram desenvolvidas para lidar com o considerável volume de observações. Essas técnicas desempenham um papel crucial não apenas na habilidade de lidar com os dados, no sentido de compreender até que ponto são precisos e confiáveis, mas também são essenciais para converter essas observações em insights sobre o comportamento da natureza ou para realizar previsões acerca dos fenômenos físicos.

Nas últimas décadas, especialmente a partir dos anos 50, testemunhamos o surgimento das primeiras técnicas de aprendizado de máquina. Os computadores, ao se tornarem ferramentas fundamentais para o desenvolvimento da sociedade, abriram caminho para a implementação efetiva dessas técnicas. Com o avanço exponencial na capacidade dos computadores, as técnicas de aprendizado de máquina puderam finalmente ser aplicadas, revelando seu poder e, principalmente, elevando-as a um status jamais visto.

Essas técnicas promoveram uma revolução em diversas áreas da sociedade, e a cos-

mologia não ficou à margem desse progresso. A combinação das técnicas de machine learning com o aprimoramento dos computadores e os desafios observacionais marcou o início de uma nova era na cosmologia. Agora, podemos utilizar o potencial das técnicas de aprendizado de máquina, aliadas ao considerável poder computacional e às técnicas estatísticas, para tentar compreender os fenômenos cosmológicos de maneira mais profunda e abrangente.

Nesse contexto, a presente monografia concentra-se, no Capítulo 1, no estudo da teoria da probabilidade e estatística para a análise de dados. Nesta seção, desenvolvemos o ferramental necessário para as técnicas empregadas tanto na inferência estatística quanto no aprendizado de máquina. Em particular, abordamos conceitos fundamentais da estatística que se mostram crucialmente relevantes, especialmente para a compreensão da teoria subjacente à técnica de machine learning conhecida como Método dos Processos Gaussianos.

Dedicamos atenção especial também ao chamado Método das Cadeias de Markov Monte Carlo, uma das técnicas estatísticas mais significativas no âmbito da inferência estatística. Este capítulo se propõe a fornecer uma base sólida, delineando os fundamentos essenciais que servirão como alicerce para a compreensão das abordagens estatísticas adotadas posteriormente na análise dos dados.

No Capítulo 3, iniciamos nosso primeiro contato com técnicas de machine learning ao discutir o conceito de redes neurais. Esta técnica, desenvolvida nos anos 50 com a intenção de modular matematicamente o cérebro humano, destaca-se como uma poderosa ferramenta presente em diversas tecnologias contemporâneas. Responsável por inúmeros produtos de inteligência artificial disponíveis no mercado atual, as redes neurais desempenham um papel significativo no avanço da capacidade computacional e na solução de problemas complexos.

Posteriormente, no Capítulo 4, exploramos o conceito do método dos processos gaussianos, uma técnica robusta para previsão e reconstrução de dados cosmológicos. Profundamente fundamentada em conceitos estatísticos, notadamente o teorema do limite central e a distribuição gaussiana, esta técnica demonstra uma notável capacidade de prever fenômenos cosmológicos.

No Capítulo 5, abordamos o modelo cosmológico padrão, onde examinamos os conceitos básicos por trás da cosmologia contemporânea. Exploramos temas relacionados à cosmologia observacional, discutimos brevemente a teoria da relatividade geral de Einstein e definimos grandezas fundamentais para o entendimento da estrutura e evolução do universo. Destacamos, em particular, o conceito da constante de Hubble e do parâmetro de densidade, a noção de curvatura e a expansão do universo. Este capítulo visa fornecer uma base sólida para a compreensão do panorama cosmológico atual.

No Capítulo 6, exploramos os vínculos observacionais relacionados ao modelo cosmológico padrão. Nesta seção, abordamos diversos fenômenos que podem ser observa-

dos para realizar inferência estatística e, consequentemente, compreender características fundamentais do nosso universo. Em especial, discutimos ferramentas computacionais, especialmente bibliotecas da linguagem de programação Python, necessárias para a implementação das diversas técnicas apresentadas ao longo do trabalho.

Além disso, são discutidas características específicas de objetos celestes, como quasares, supernovas e galáxias HII. Exploramos também o comportamento da expansão cósmica ao longo do tempo, especialmente por meio dos chamados cronômetros cósmicos. Esta análise visa oferecer insights valiosos sobre a evolução do universo e proporcionar uma compreensão mais aprofundada dos dados observacionais.

2 Probabilidade e estatística para a análise de dados

Desde os primórdios das civilizações antigas, o pensamento humano tem sido intrinsecamente ligado aos conceitos de probabilidade e estatísticas. Quando nossos antepassados começaram a contar rebanhos, analisar colheitas e até mesmo lançar dados em jogos de azar, eles, de maneira inadvertida, exploraram noções fundamentais de probabilidade. A incerteza inerente à vida cotidiana impulsionou-os a desenvolver métodos para entender padrões, fazer previsões e tomar decisões informadas, criando assim as bases para a teoria probabilística [47].

Ao longo dos tempos e em diversas civilizações antigas, a concepção de probabilidade e estatísticas era muitas vezes implícita e fundamentada na observação empírica. Nas sociedades mesopotâmicas, por exemplo, agricultores observavam padrões climáticos e comportamentos naturais para tomar decisões sobre plantio e colheita, uma forma primitiva de aplicação da probabilidade. No Egito antigo, registros detalhados eram mantidos para avaliar padrões climáticos, inundações do rio Nilo e produtividade agrícola, proporcionando uma compreensão rudimentar das estatísticas [27].

Foi na Grécia Antiga, entretanto, que pensadores como Pitágoras e Euclides exploraram relações numéricas e padrões geométricos, contribuindo indiretamente para o desenvolvimento das ideias probabilísticas. Na China antiga, alguns dos primeiros métodos estatísticos registrados surgiram, especialmente no contexto de censo populacional e avaliação de recursos para fins administrativos [27]. A ideia moderna de probabilidade e estatística como disciplinas formais começou a se desenvolver na Europa durante os séculos XVII e XVIII, com a introdução de conceitos matemáticos formais para analisar dados e incertezas. Assim, ao longo dos séculos e através das civilizações antigas, a humanidade gradualmente progrediu de observações empíricas para métodos estatísticos sistemáticos, moldando o curso da probabilidade e estatística como as conhecemos hoje.

Neste contexto evolutivo, emerge um marco significativo na história da probabilidade: o trabalho revolucionário de Andrei Kolmogorov. No século XX, Kolmogorov formalizou a teoria da probabilidade, proporcionando uma base matemática sólida para abordar eventos aleatórios. Suas contribuições não apenas consolidaram os fundamentos da probabilidade, mas também abriram caminho para inúmeras aplicações em diversos campos

do conhecimento [34].

2.1 Os axiomas de Kolmogorov

A teoria da probabilidade, conforme formulada por Andrey Kolmogorov em seu trabalho seminal "Foundations of the Theory of Probability" em 1933 [23], é fundamentada em três axiomas cruciais. Os axiomas de Kolmogorov constituem os princípios fundamentais da teoria da probabilidade, delineando um formalismo rigoroso para modelar experimentos probabilísticos. Em primeiro lugar, define-se o espaço amostral (S) como o conjunto completo e abrangente de todos os resultados possíveis de um experimento probabilístico. Cada resultado individual dentro de S é denominado ponto amostral. Os eventos, por sua vez, são definidos como subconjuntos do espaço amostral ($A \subseteq S$). O conjunto vazio (\emptyset) representa um evento impossível, enquanto o espaço amostral completo (S) denota um evento certo [23].

Kolmogorov introduz a medida de probabilidade ($P(A)$), uma função que associa a cada evento A um número real no intervalo de 0 a 1. Esta medida representa a probabilidade de o evento ocorrer, onde $P(A) = 1$ indica que o evento A certamente ocorrerá, enquanto $P(A) = 0$ implica que o evento é impossível de ocorrer. A partir dessa definição formal, emerge um axioma adicional crucial. Considerando dois eventos A_1 e A_2 pertencentes ao espaço amostral S , e assumindo que A_1 e A_2 são mutuamente exclusivos (ou seja, $P(A_1 \cap A_2) = \emptyset$), então a probabilidade da união dos eventos é dada pela soma das probabilidades individuais dos eventos [23], ou seja

$$P(A_1 \cup A_2) = P(A_1) + P(A_2). \quad (2.1)$$

A partir dos axiomas de Kolmogorov, é possível obter alguns conceitos fundamentais para a teoria da probabilidade:

- Para qualquer evento A , temos $S = A + A^c$ onde A^c é o conjunto complementar de A .
- Para qualquer evento A temos por definição $S = A \cup A^c$, portanto

$$P(A \cup A^c) = 1. \quad (2.2)$$

- Do axioma da probabilidade da união visto que (por definição) $A \cap A^c = \emptyset$, então $P(A \cup A^c) = P(A) + P(A^c)$, o que implica em

$$P(A^c) = 1 - P(A). \quad (2.3)$$

- Considerando dois eventos A_1 e A_2 , então

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2). \quad (2.4)$$

Uma definição extremamente importante na teoria das probabilidades é a definição de probabilidade condicional: dado um evento B com $P(B) > 0$ então a probabilidade condicional de ocorrência de um evento A dado a ocorrência de um evento B , denotada por $P(A|B)$ é definida por

$$P(A|B) = \frac{P(A \cap B)}{P(B)}. \quad (2.5)$$

Uma consequência da definição da probabilidade condicional é o denominado teorema da probabilidade total, que é definido como se segue: dado um conjunto de eventos A_1, A_2, \dots, A_n mutuamente exclusivos entre sí, isto é $A_i \cap A_j = \emptyset \forall (i, j)$, então para um evento X qualquer em S

$$P(X) = \sum_n P(X \cap A_i), \quad (2.6)$$

portanto

$$P(X) = \sum_n P(A_i)P(X|A_i). \quad (2.7)$$

Uma definição importante é a definição de independência: dois eventos A_1 e A_2 são independentes se

$$P(A \cap B) = P(A)P(B), \quad (2.8)$$

portanto, dada a definição de probabilidade condicional, se dois eventos são independentes então

$$P(A|B) = P(A) \quad (2.9)$$

e

$$P(B|A) = P(B). \quad (2.10)$$

Com base nos axiomas de Kolmogorov, a teoria da probabilidade oferece uma estrutura matemática precisa para avaliar eventos aleatórios. A definição de eventos e seus complementos, juntamente com a probabilidade condicional, proporciona uma maneira rigorosa de entender relações probabilísticas entre eventos inter-relacionados. O teorema da probabilidade total simplifica o cálculo de probabilidades em cenários complexos, enquanto a noção de independência entre eventos é fundamental para análises estatísticas

precisas. Assim, esses conceitos não apenas fundamentam a teoria probabilística, mas também têm amplas aplicações em diversos contextos científicos e práticos [6].

Não obstante, entre as definições estabelecidas por Kolmogorov, uma das essenciais é o conceito de variável aleatória. Uma variável aleatória, denotada por $X : S \rightarrow \mathbb{R}$, funciona como uma função que associa cada elemento do espaço amostral S a um número real. Em outras palavras, para cada resultado s em S , a variável aleatória X atribui um valor real $X(s)$. Essa definição é crucial, pois oferece um método formal de representar e analisar eventos aleatórios [23].

Ao atribuir valores numéricos aos resultados de experimentos, a variável aleatória permite a aplicação de métodos matemáticos para compreender e prever padrões em dados incertos. Isso se revela fundamental em diversas áreas, desde a ciência até a economia, possibilitando a tomada de decisões informadas com base na probabilidade e na estatística [6]. Em síntese, a teoria da probabilidade de Kolmogorov, fundamentada em seus axiomas essenciais, fornece a estrutura fundamental para entender e quantificar a incerteza nos eventos aleatórios. Ao estabelecer o espaço amostral, definir eventos como subconjuntos desse espaço e introduzir a medida de probabilidade, Kolmogorov estabeleceu as bases para uma compreensão matematicamente rigorosa da aleatoriedade. A introdução da variável aleatória como uma ferramenta para atribuir valores numéricos aos resultados de experimentos amplia o escopo de aplicação da teoria, permitindo análises detalhadas e previsões em uma variedade de campos [23].

2.2 Interpretação bayesiana da teoria das probabilidades

Baseado nos axiomas de Kolmogorov, surgem duas interpretações fundamentais da probabilidade: a interpretação frequentista e a interpretação bayesiana. Na interpretação frequentista, a probabilidade de um evento é vista como a frequência relativa desse evento em um número infinito de repetições idênticas de um experimento. Em outras palavras, é a proporção de vezes que o evento ocorre quando o experimento é repetido um grande número de vezes. Por outro lado, na interpretação bayesiana, a probabilidade é vista como uma medida subjetiva da crença ou confiança na ocorrência de um evento, incorporando tanto a informação inicial (conhecimento prévio) quanto as evidências obtidas dos dados observados [23].

Na interpretação frequentista da probabilidade, se consideramos um experimento aleatório que pode ser repetido sob as mesmas condições, se n representa o número de repetições desse experimento e n_A denota o número de vezes que o evento A ocorre nessas repetições, a frequência relativa de A em n experimentos é dada por $\frac{n_A}{n}$. A interpretação frequentista postula que à medida que n se aproxima do infinito, essa frequência relativa converge

para um valor constante, representado por p [6]. Em outras palavras, a probabilidade de A , denotada por $P(A)$, é igual a esse valor constante p , o qual é interpretado como a proporção limite de ocorrências de A em um número infinitamente grande de repetições do mesmo experimento, representada por

$$P(A) = \lim_{n \rightarrow \infty} \frac{n_A}{n} = p. \quad (2.11)$$

Na interpretação bayesiana da probabilidade, a probabilidade de um evento A é tratada como uma medida de incerteza ou crença sobre a ocorrência desse evento. Ao contrário da interpretação frequentista, que se baseia em repetições de um experimento para estimar a probabilidade, a interpretação bayesiana considera a probabilidade como uma expressão da nossa própria incerteza sobre um evento, antes e depois de observarmos evidências [23].

Na interpretação bayesiana, a probabilidade de um evento é expressa em termos de uma distribuição de probabilidade chamada distribuição a posteriori. A distribuição a posteriori representa a probabilidade de um evento A ocorrer, dado um conjunto de evidências ou dados observados D é denotada por $P(A|D)$ e é calculada usando o teorema de Bayes [23], que relaciona a probabilidade a priori $P(A)$ (a nossa crença sobre A antes de observar os dados) com a verossimilhança $P(D|A)$ (a probabilidade de observar os dados D dado que A é verdadeiro) e a probabilidade marginal $P(D)$ (a probabilidade de observar os dados D independentemente de A) da seguinte forma:

$$P(A|D) = \frac{P(D|A)P(A)}{P(D)}. \quad (2.12)$$

Nesta equação:

- $P(A|D)$ é a probabilidade a posteriori de A após observarmos os dados D .
- $P(D|A)$ é a verossimilhança de observar os dados D dado que A é verdadeiro.
- $P(A)$ é a probabilidade a priori de A , a nossa crença sobre A antes de observarmos os dados.
- $P(D)$ é a probabilidade marginal dos dados D , que é usada para normalizar a distribuição a posteriori e garantir que a soma das probabilidades seja igual a 1 [52].

O teorema de Bayes se baseia no teorema da probabilidade total e na definição de probabilidade condicional, se (para um evento A e um evento D) tomarmos a Equação 2.5, temos

$$P(A \cap D) = P(D)P(A|D) \quad (2.13)$$

e

$$P(D \cap A) = P(A)P(D|A), \quad (2.14)$$

porém, por definição,

$$P(D \cap A) = P(A \cap D), \quad (2.15)$$

o que implica na Equação 2.12.

Na interpretação bayesiana, a probabilidade é vista como uma medida contínua e flexível que pode ser atualizada à medida que novas evidências são obtidas, permitindo assim uma modelagem mais precisa e adaptativa das incertezas em várias situações do mundo real [52].

Em resumo, as interpretações frequentista e bayesiana da probabilidade representam abordagens matematicamente sólidas e fundamentais para quantificar eventos aleatórios. A interpretação frequentista baseia-se na frequência relativa de um determinado evento, enquanto a interpretação bayesiana trata a probabilidade como uma medida subjetiva de crença, incorporando informações a priori e dados observados. Ambas desempenham papéis essenciais em campos como estatística e aprendizado de máquina, fornecendo ferramentas cruciais para análise de dados e tomada de decisões bem fundamentadas.

2.3 Distribuições de probabilidade

Refletindo sobre as interpretações da probabilidade discutidas anteriormente e as implicações derivadas, agora nos voltamos para outra ferramenta essencial: as distribuições de probabilidade. Essas representações matemáticas fundamentais desempenham um papel crucial na compreensão e quantificação da incerteza em torno de eventos aleatórios. Seja lançando um dado, medindo a temperatura em um determinado dia ou calculando a probabilidade de um evento em um jogo de cartas, as distribuições de probabilidade oferecem uma estrutura sólida para entender como diferentes resultados podem se desdobrar. Elas nos permitem atribuir probabilidades a cada resultado possível, capacitando-nos a fazer previsões informadas e tomar decisões embasadas em dados confiáveis. Ao examinar minuciosamente as características essenciais dessas distribuições, estamos preparando o terreno para uma compreensão aprofundada das análises que virão, proporcionando uma base vital para a compreensão de eventos aleatórios e a manipulação eficaz da incerteza [6].

Uma distribuição de probabilidade é uma função matemática que atribui probabilidades a diferentes resultados possíveis de um experimento aleatório. Essa função descreve a incerteza associada aos resultados e é fundamental na teoria da probabilidade e na estatística. Formalmente, para um conjunto de eventos x_1, x_2, \dots, x_n associados a uma variável aleatória X , uma distribuição de probabilidade atribui uma probabilidade $P(X)$ a cada evento x_i associado a X , satisfazendo os axiomas de Kolmogorov: a probabilidade de cada evento é um número real não negativo ($P(x_i) \geq 0$) e a soma das probabilidades de

todos os eventos possíveis é igual a 1 ($\sum_{i=1}^n P(x_i) = 1$). As distribuições de probabilidade podem assumir diversas formas e a compreensão profunda dessas distribuições é vital para análises estatísticas precisas, inferências robustas e tomadas de decisão fundamentadas em dados reais [6].

Para variáveis aleatórias contínuas, a abordagem é ligeiramente diferente. Em vez de atribuir probabilidades a eventos específicos, como fazemos com variáveis discretas, focamos nos intervalos. Esta característica é encapsulada pela função de densidade de probabilidade (pdf), representada por $f(x)$. A pdf descreve a probabilidade de que a variável aleatória X caia dentro de um intervalo infinitesimal $[x, x + dx]$. Em termos matemáticos, isso é expresso como $P(X \in [x, x + dx]) = f(x) dx$, onde dx é um diferencial infinitesimal [6]. A pdf é normalizada de acordo com os axiomas de Kolmogorov, garantindo que a integral da função $f(x)$ sobre todo o espaço amostral S seja igual a 1:

$$\int_S f(x) dx = 1. \quad (2.16)$$

Quando se trata de representar graficamente a distribuição de variáveis aleatórias contínuas, o histograma é uma ferramenta essencial. Diferentemente das variáveis discretas, onde os dados podem ser plotados como pontos discretos em um gráfico, variáveis contínuas podem assumir infinitos valores. O histograma resolve essa complexidade dividindo o espectro contínuo em compartimentos chamados de bins. Cada bin representa um intervalo específico e a altura de cada barra no histograma denota a frequência ou densidade de probabilidade das observações dentro desse intervalo. Assim, enquanto a função de densidade de probabilidade fornece uma descrição matemática precisa das probabilidades associadas a variáveis contínuas, o histograma oferece uma representação visual que facilita a compreensão da distribuição dos dados. Ao analisar um histograma, podemos identificar padrões, outliers e outras características importantes dos dados, desempenhando um papel crucial em análises estatísticas avançadas e na tomada de decisões [6].

De forma rigorosa, ao lidar com um conjunto de observações $\{x_1, x_2, \dots, x_n\}$, é possível criar um histograma da seguinte maneira: o eixo horizontal (x) do histograma é dividido em m intervalos, cada um de tamanho Δx_i , onde $i = 1, \dots, m$. A frequência n_i das ocorrências de x dentro do subintervalo i é representada no eixo vertical (y), e a área sob o histograma é dada por $\sum_{i=1}^m n_i \cdot \Delta x_i$. Para obter um histograma normalizado, é possível dividir cada n_i por $n \cdot \Delta x_i$, onde $n = \sum_{i=1}^m n_i$, a Figura 2.1 representa o histograma para 10^6 pontos obtidos a partir de uma pdf gaussiana com média 0 e desvio padrão 0,01 com bins de tamanho 0,01 e a Figura 2.1 representa os mesmos dados para bins de tamanho 0,005 [6].

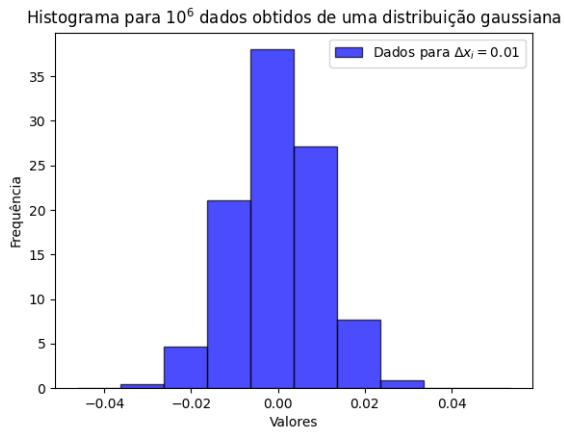


Figura 2.1: Histograma de 10^6 amostras de uma distribuição gaussiana com média 0 e desvio padrão 0,01 para $\Delta x_i = 0,01$. Fonte: Criado pelo autor.

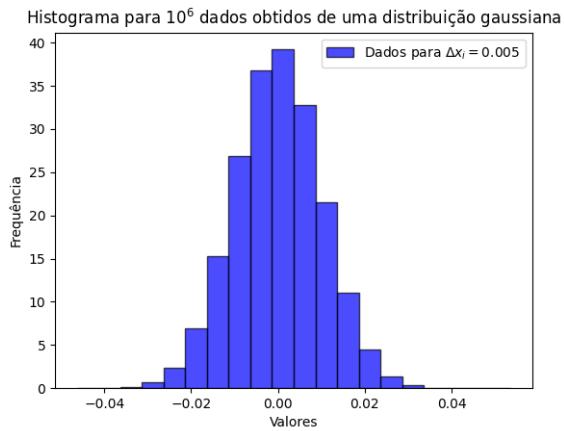
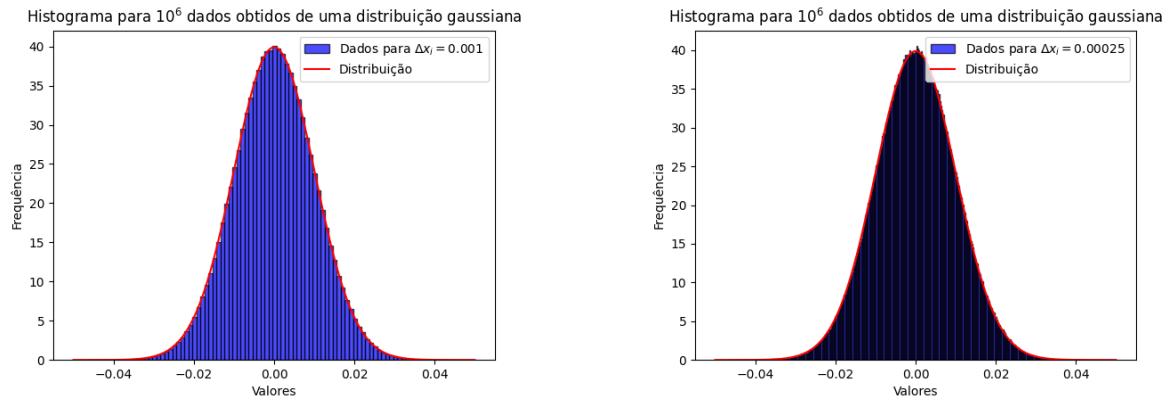


Figura 2.2: Histograma de 10^6 amostras de uma distribuição gaussiana com média 0 e desvio padrão 0,01 para $\Delta x_i = 0,005$. Fonte: Criado pelo autor.

A pdf pode ser representada como um histograma normalizado onde $\lim_{\Delta x_i \rightarrow 0}$, isso fica evidente na Figura 2.3.



(a) Histograma de 10^6 amostras de uma distribuição gaussiana com média 0 e desvio padrão 0,01 para $\Delta x_i = 0,0005$. Fonte: Criado pelo autor.

(b) Histograma de 10^6 amostras de uma distribuição gaussiana com média 0 e desvio padrão 0,01 para $\Delta x_i = 0,000025$. Fonte: Criado pelo autor.

Figura 2.3: Visualização da correspondência entre um histograma normalizado e uma pdf.

Na teoria das probabilidades, uma definição fundamental é a da função de distribuição acumulada $F(x)$, também conhecida como distribuição acumulada. Essa função representa a probabilidade de uma variável aleatória assumir um valor menor ou igual a x . Para variáveis aleatórias contínuas, a distribuição acumulada é definida matematicamente como

$$F(x) = \int_{-\infty}^x f(x') dx', \quad (2.17)$$

onde $f(x)$ é a função de densidade de probabilidade (pdf) da variável aleatória. A pdf pode, por sua vez, ser obtida através da derivada da distribuição acumulada:

$$f(x) = \frac{\partial F(x)}{\partial x}. \quad (2.18)$$

Compreender a função de distribuição acumulada $F(x)$ e sua relação direta com a função de densidade de probabilidade é essencial. Esta conexão permite calcular percentis, medianas e quartis em conjuntos de dados, sendo fundamental para interpretações estatísticas precisas e análises de dados avançadas [6].

2.3.1 Marginalização e pdf conjunta

Na natureza, é comum depararmo-nos com eventos que envolvem múltiplas variáveis. Um exemplo notável ocorre ao estudarmos estrelas, onde lidamos com dados de luminosidade, cor, redshift, entre outros parâmetros. Nestes cenários, surgem as distribuições de probabilidade conjuntas. Suponhamos a observação de duas variáveis aleatórias, x e y , e definimos o evento A como "observar um valor para a variável aleatória x dentro do intervalo $[x, x + dx]$, e um valor qualquer para a variável aleatória y ", e o evento B como

”observar um valor para a variável aleatória y dentro do intervalo $[y, y + dy]$, e um valor qualquer para a variável aleatória x ”. A distribuição conjunta de x e y , denotada por $f(x, y)$, representa a probabilidade de observar a variável aleatória x dentro do intervalo $[x, x + dx]$ e simultaneamente observar a variável aleatória y dentro do intervalo $[y, y + dy]$. Matematicamente, isso é expresso como:

$$P(A \cap B) = f(x, y) dx dy, \quad (2.19)$$

nesse caso, a condição de normalização é representada por

$$\iint_S f(x, y) dx dy = 1. \quad (2.20)$$

Considere o conhecimento de uma distribuição conjunta de probabilidade $f(x, y)$ para duas variáveis aleatórias x e y . Nesse contexto, é possível derivar a distribuição de probabilidade da variável x (ou y) utilizando a seguinte relação:

$$f_x(x) = \int_{-\infty}^{\infty} f(x, y) dy \quad (2.21)$$

ou

$$f_y(y) = \int_{-\infty}^{\infty} f(x, y) dx, \quad (2.22)$$

onde essas expressões representam as distribuições marginais de probabilidade da variável aleatória x e y , respectivamente [6].

Com a definição da pdf conjunta, torna-se possível revisitar o conceito de probabilidade condicional. No contexto em questão, a distribuição condicional da ocorrência do evento B dado que o evento A ocorreu é obtida pela seguinte relação:

$$P(B|A) = \frac{f(x, y) dy}{f_x(x)} \quad (2.23)$$

Dessa relação, surge o conceito de pdf condicional de y dado x , que é definida como:

$$h(y|x) = \frac{f(x, y)}{\int f(x, y') dy'}. \quad (2.24)$$

De maneira análoga, define-se a pdf condicional de x dado y :

$$g(x|y) = \frac{f(x, y)}{\int f(x', y) dx'}. \quad (2.25)$$

Combinando a Equação 2.25 com a Equação 2.24, é possível obter a relação:

$$g(x|y) = \frac{h(y|x)f_x(x)}{f_y(y)}. \quad (2.26)$$

Essa é uma versão do Teorema de Bayes para pdfs contínuas, e por meio dessa equação, podem ser estabelecidas relações importantes entre as pdfs condicionais e marginais:

$$f_x(x) = \int_{-\infty}^{\infty} g(x|y)f_y(y) dy \quad (2.27)$$

e

$$f_y(y) = \int_{-\infty}^{\infty} h(y|x)f_x(x) dx. \quad (2.28)$$

A partir das discussões deste capítulo, emerge naturalmente o conceito de independência entre os eventos A e B : se A e B são eventos independentes, então a pdf conjunta $f(x, y)$ pode ser expressa como o produto das pdfs marginais:

$$f(x, y) = f_x(x) \cdot f_y(y). \quad (2.29)$$

Para eventos independentes, podemos observar, com base na Equação 2.23 e na Equação 2.29, que a função de densidade de probabilidade condicional $g(x|y)$ não depende dos valores de y , e da mesma forma, a função de densidade de probabilidade condicional $h(y|x)$ não depende dos valores assumidos pela variável x [6].

Nesta seção, exploramos a complexidade das distribuições de probabilidade conjuntas ao lidar com múltiplas variáveis, como exemplificado pelo estudo das estrelas, que envolve parâmetros como luminosidade, cor e redshift. Ao definir eventos A e B relacionados às variáveis aleatórias x e y , respectivamente, e introduzir a distribuição conjunta $f(x, y)$, estabelecemos as bases para entender as interações entre essas variáveis. Através das distribuições marginais $f_x(x)$ e $f_y(y)$, derivadas de $f(x, y)$, exploramos as probabilidades individuais de x e y . Além disso, ao aplicar o Teorema de Bayes para pdfs contínuas, relacionamos $f(x, y)$ com as pdfs condicionais $g(x|y)$ e $h(y|x)$, estabelecendo conexões cruciais entre eventos e variáveis. A independência entre eventos A e B , representada pela Equação 2.29, revela que as pdfs condicionais são independentes dos valores assumidos pelas outras variáveis, destacando a riqueza e a sutileza das distribuições de probabilidade conjuntas. Estas análises são fundamentais para desvendar padrões complexos na natureza, oferecendo insights valiosos para diversas áreas da ciência e da pesquisa.

2.4 Medidas de tendência central: quantis e valores esperados

No campo da estatística, as medidas de tendência central, como média, mediana e moda, são pilares essenciais para entender a distribuição dos dados. Da mesma forma, os valores esperados oferecem uma visão precisa das expectativas em torno desses dados. Além disso, os quantis desempenham um papel crucial na análise estatística, fornecendo in-

formações sobre a dispersão dos dados e identificando pontos de corte importantes na distribuição. Este capítulo é uma exploração cuidadosa e objetiva desses conceitos fundamentais, destacando suas aplicações e importância na análise estatística. Ao compreender essas ferramentas essenciais, somos capacitados a interpretar dados de maneira significativa e informada, permitindo-nos explorar não apenas a tendência central dos dados, mas também a sua variabilidade e distribuição em profundidade.

Os quantis de ordem a , denotados por x_a , estão intrinsecamente ligados à definição da distribuição acumulada e são definidos como os valores da variável aleatória x para os quais a função de distribuição acumulada $F(x_a)$ é igual a a , onde $0 \leq a \leq 1$. Em outras palavras, o quantil pode ser representado como a função inversa da distribuição acumulada, matematicamente definido por

$$x_a = F^{-1}(a), \quad (2.30)$$

Um exemplo particularmente significativo de quantil é $x_{1/2}$, referido como a mediana, que representa a medida típica da variável aleatória. Isso ocorre porque há uma probabilidade igual de a variável aleatória assumir valores tanto maiores quanto menores do que $x_{1/2}$ [6].

Entre as medidas de tendência central, a moda destaca-se como uma das mais relevantes. Ela representa o valor de uma variável aleatória x para o qual a pdf atinge seu máximo, indicando assim o valor mais provável de ocorrer na distribuição de dados [6].

2.4.1 Valores esperados

No contexto da análise de dados, é imperativo aprofundar nosso entendimento sobre os valores esperados e suas implicações estatísticas. Dessa forma, é fundamental explorar o conceito de valor esperado e suas aplicações práticas na análise estatística. Nesse processo, vamos examinar cuidadosamente medidas essenciais como média, variância, desvio padrão, covariância e correlação, compreendendo sua importância fundamental na interpretação precisa dos dados.

Ao compreendermos esses conceitos de maneira clara e abrangente, estaremos preparados para realizar análises estatísticas significativas em uma variedade de situações do mundo real. A familiaridade com esses princípios nos capacita a interpretar dados de forma mais precisa, proporcionando uma base sólida para análises estatísticas mais avançadas e informadas.

O valor esperado μ de uma variável aleatória x , com densidade de probabilidade $f(x)$, é definido por

$$\mathbb{E}[x] = \int_{-\infty}^{\infty} xf(x) dx. \quad (2.31)$$

É crucial notar que o valor esperado não é uma função de x , dependendo unicamente da forma da densidade de probabilidade $f(x)$. Se uma pdf concentra a probabilidade em uma região específica, então o valor esperado ($\mathbb{E}(x)$) representa o valor mais provável que se obterá ao medir a variável x . No entanto, se a pdf distribuir a probabilidade entre dois picos separados, o valor esperado será um número situado entre esses picos, sendo pouco provável que seja observado diretamente [6].

Dentre os valores esperados de interesse, destaca-se o enésimo momento de x , denotado por μ_n' , definido como

$$\mathbb{E}[x^n] = \int_{-\infty}^{\infty} x^n f(x) dx, \quad (2.32)$$

cujo caso especial μ_1' é o valor esperado de x , se destaca também o enésimo momento central de x

$$\mathbb{E}[(x - \mu)^n] = \int_{-\infty}^{\infty} (x - \mu)^n f(x) dx, \quad (2.33)$$

destaca-se, especialmente, o segundo momento central, denominado variância de x , que pode ser representado por

$$\mathbb{V}[x] = \mathbb{E}[x^2] - \mu^2. \quad (2.34)$$

A variância é a medida da dispersão dos valores de x em torno da média e a partir da variância se define o desvio padrão, representado por

$$\sigma = \sqrt{\mathbb{V}[x]}, \quad (2.35)$$

o desvio padrão é particularmente útil por ser representado nas mesmas unidades de x .

A covariância entre duas variáveis aleatórias x e y , é outro valor esperado de extrema importância, a covariância representa uma relação entre duas variáveis aleatórias x e y e é definida por

$$\text{Cov}(x, y) = \mathbb{E}[(x - \mu_x)(y - \mu_y)], \quad (2.36)$$

a covariância é, por definição, simétrica ($\text{Cov}(x, y) = \text{Cov}(y, x)$) e pode ser representada por

$$\text{Cov}(x, y) = \mathbb{E}[xy] - \mu_x \mu_y, \quad (2.37)$$

onde $\mu_x = \mathbb{E}[x]$ e $\mu_y = \mathbb{E}[y]$. A covariância é uma medida que quantifica como duas variáveis estão linearmente relacionadas; se positiva, indica uma relação positiva, e se negativa, indica uma relação negativa [48].

A partir da definição da covariância, emerge o conceito de matriz de covariância, que é uma matriz simétrica que descreve a variabilidade conjunta das variáveis aleatórias envolvidas. Formalmente, para um vetor de variáveis aleatórias multivariadas $X = [X_1, X_2, \dots, X_p]$, a matriz de covariância Σ é definida por

$$\Sigma = \begin{bmatrix} \text{Cov}(X_1, X_1) & \text{Cov}(X_1, X_2) & \dots & \text{Cov}(X_1, X_p) \\ \text{Cov}(X_2, X_1) & \text{Cov}(X_2, X_2) & \dots & \text{Cov}(X_2, X_p) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(X_p, X_1) & \text{Cov}(X_p, X_2) & \dots & \text{Cov}(X_p, X_p) \end{bmatrix}, \quad (2.38)$$

onde $\text{Cov}(X_i, X_j)$ é a covariância entre as variáveis X_i e X_j [31].

A diagonal principal da matriz de covariância contém as variâncias individuais das variáveis ($\text{Var}(X_i)$), e os elementos fora da diagonal principal contêm as covariâncias entre pares de variáveis. A covariância é uma medida que quantifica como duas variáveis estão linearmente relacionadas; se positiva, indica uma relação positiva, e se negativa, indica uma relação negativa. A matriz de covariância é uma ferramenta importante na análise multivariada, pois fornece informações sobre as relações e a dispersão conjunta das variáveis aleatórias [48].

Com a introdução dos valores esperados se define o coeficiente de correlação de Pearson (ou simplesmente coeficiente de correlação), que representa uma grandeza adimensional do nível de correlação entre duas variáveis aleatórias, o coeficiente de correlação é definido por

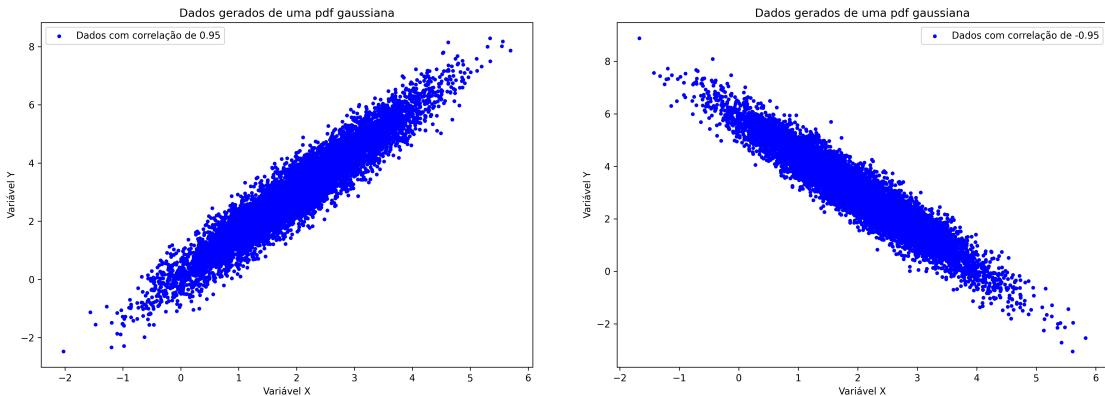
$$\rho_{xy} = \frac{\text{Cov}(x, y)}{\sigma_x \sigma_y}, \quad (2.39)$$

onde $-1 \leq \rho_{xy} \leq 1$ [6].

O coeficiente de correlação pode ser interpretado da seguinte maneira: se os valores de x são maiores que a média (μ_x) e isso aumenta a probabilidade de que y também seja maior que a média (μ_y), ou se os valores de x são menores que μ_x e isso aumenta a probabilidade de que y seja menor que μ_y , então o valor do coeficiente de covariância V_{xy} é maior que zero. Nesse caso, x e y são considerados positivamente correlacionados, o que significa que o coeficiente de correlação ρ_{xy} é maior que zero. Da mesma forma, se $x > \mu_x$ aumenta a probabilidade de $y < \mu_y$ e $x < \mu_x$ aumenta a probabilidade de $y > \mu_y$, então V_{xy} é menor que zero. Nessa situação, x e y são considerados negativamente correlacionados, indicando que ρ_{xy} é menor que zero [6].

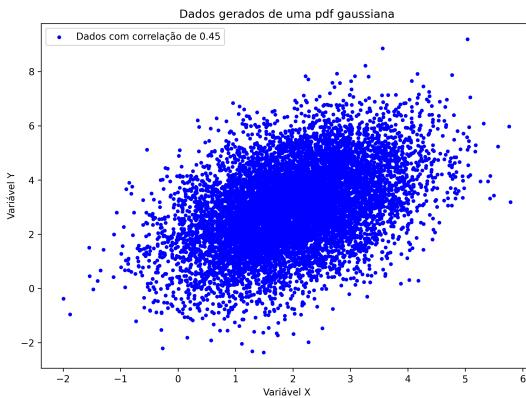
O coeficiente de correlação pode ser visualizado na Figura 2.4 por meio do gráfico de dispersão, onde os pontos que estão mais próximos de uma linha reta ascendente indicam uma forte correlação positiva, enquanto os pontos próximos de uma linha reta descendente indicam uma forte correlação negativa. Quando os pontos estão dispersos sem seguir um padrão claro, isso sugere uma correlação fraca ou inexistente entre as variáveis x e y . Em

outras palavras, a disposição dos pontos no gráfico de dispersão pode ilustrar visualmente o valor e a direção do coeficiente de correlação entre as variáveis x e y [6].

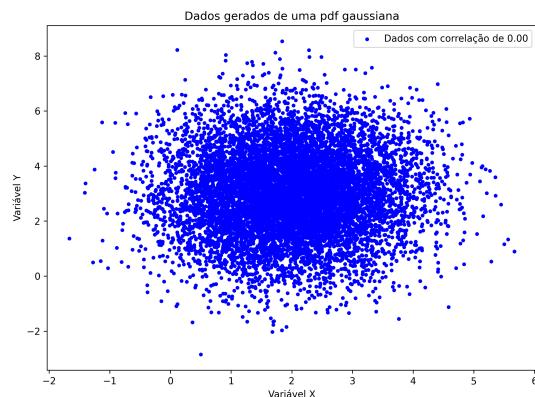


(a) Variáveis aleatórias fortemente correlacionadas positivamente. Fonte: Criado pelo autor.

(b) Variáveis aleatórias fortemente correlacionadas negativamente. Fonte: Criado pelo autor.



(c) Variáveis aleatórias fracamente correlacionadas positivamente. Fonte: Criado pelo autor.



(d) Variáveis aleatórias não correlacionadas. Fonte: Criado pelo autor.

Figura 2.4: Visualização da correlação entre duas variáveis aleatórias

Nesta seção, exploramos os pilares essenciais da estatística, desde as medidas centrais até os detalhes sutis dos valores esperados. Descobrimos como os dados se agrupam em torno de médias, medianas e modas, e entendemos a importância dos quantis para definir pontos-chave na distribuição. Aprofundamos nosso conhecimento nos valores esperados, desvendando a dispersão dos dados através da variância e do desvio padrão. Descobrimos as relações complexas entre variáveis por meio da covariância e do coeficiente de correlação. Com essas ferramentas em mãos, agora podemos transformar dados em insights claros e decisões bem fundamentadas, capacitados para enfrentar desafios complexos no mundo dos dados.

2.5 Verossimilhança

Na Seção 2.2, exploramos a interpretação bayesiana da probabilidade, concentrando-nos na verossimilhança, que representa a probabilidade dos dados observados sob uma determinada hipótese. Contudo, a verossimilhança é mais do que uma simples métrica de ajuste de dados; ela é um pilar fundamental da inferência estatística.

A verossimilhança desempenha um papel central em campos como a inteligência artificial, onde é crucial em algoritmos de aprendizado de máquina para avaliar diferentes modelos em relação aos dados observados. Em cosmologia, a verossimilhança é essencial para ajustar modelos teóricos às observações astronômicas, permitindo que os cientistas compreendam a estrutura e a evolução do universo. Além disso, na análise de dados, a verossimilhança é empregada em métodos como a máxima verossimilhança para estimar parâmetros desconhecidos e modelar fenômenos complexos, garantindo que as conclusões sejam fundamentadas em dados sólidos e interpretações estatisticamente robustas [12].

O conceito moderno de verossimilhança representa a plausibilidade dos dados observados X sob um modelo estatístico específico ou um conjunto de parâmetros θ . Suponha que temos um conjunto de dados observados X e um modelo estatístico com parâmetros θ . A função de verossimilhança, denotada por $L(\theta|X)$, não é uma probabilidade, mas sim uma medida relativa de quão bem o modelo se ajusta aos dados observados sob os parâmetros θ . A verossimilhança é calculada como a função de densidade de probabilidade dos dados, não normalizada, e é expressa como

$$L(\theta|X) = f(X|\theta), \quad (2.40)$$

onde $f(X|\theta)$ é a função de densidade de probabilidade dos dados X dado que os parâmetros do modelo são θ [12]. Em muitos casos, especialmente em contextos de modelagem estatística, assume-se que os dados são independentes e identicamente distribuídos (i.i.d.), simplificando a função de verossimilhança para o produto das densidades de probabilidade dos dados individuais

$$L(\theta|X) = \prod_{i=1}^n f(x_i|\theta), \quad (2.41)$$

onde x_i são os pontos de dados individuais no conjunto de dados X [12].

A verossimilhança desempenha um papel crucial na comparação de modelos estatísticos, na tomada de decisões baseada em dados e em diversas outras aplicações estatísticas e científicas, tornando-se um conceito fundamental na teoria estatística moderna [50].

Na Inferência de Máxima Verossimilhança (MLE), a verossimilhança é central. A técnica MLE estima os parâmetros de um modelo estatístico maximizando a função de verossimilhança dos dados observados. Em outras palavras, MLE encontra os valores dos

parâmetros que tornam os dados observados mais prováveis sob o modelo especificado. Esta técnica é amplamente utilizada devido à sua eficiência e propriedades estatísticas desejáveis, sendo a base para muitos métodos de estimação em ciência e engenharia [50].

Nas Cadeias de Markov Monte Carlo (MCMC), a verossimilhança é essencial. MCMC é uma técnica computacional poderosa para amostrar distribuições de probabilidade complexas. No contexto da verossimilhança, MCMC é frequentemente utilizado em análises bayesianas para explorar a distribuição a posteriori dos parâmetros do modelo. Os algoritmos MCMC, como o Metropolis-Hastings, guiam-se pela verossimilhança para gerar amostras representativas da distribuição a posteriori, permitindo inferências detalhadas em problemas complexos.

Nos Testes de Razão de Verossimilhança (LRT), a verossimilhança é central para a comparação de modelos. LRT compara a verossimilhança dos dados sob diferentes modelos estatísticos [53]. Ao calcular a razão das verossimilhanças, os pesquisadores podem determinar quão bem cada modelo se ajusta aos dados. Esse teste é fundamental para selecionar o modelo que melhor descreve os dados observados, sendo uma ferramenta essencial em campos como análise de dados, cosmologia e inteligência artificial [25, 53].

Assim, a verossimilhança desempenha um papel crítico em técnicas como MLE, MCMC e LRT. Ela não apenas fornece uma base teórica sólida para esses métodos, mas também se torna uma ferramenta fundamental para a inferência estatística, modelagem complexa e tomada de decisões baseada em dados em uma variedade de áreas do conhecimento [12].

2.5.1 Qui-quadrado

Nesta seção, exemplificaremos o conceito de verossimilhança utilizando a estatística do qui-quadrado para avaliar a adequação de um modelo aos dados observados. Suponhamos que tenhamos um conjunto de dados y e um modelo teórico $f(x, \theta)$ com parâmetros θ . Nossa objetivo central é determinar a precisão do ajuste do modelo aos dados e estimar os parâmetros que otimizam a verossimilhança dos dados sob o modelo [50].

A estatística do qui-quadrado compara as frequências observadas y com as frequências esperadas $f(x, \theta)$ em diferentes categorias ou intervalos. A estatística do qui-quadrado é calculada como:

$$\chi^2 = (y - f(x, \theta))^T \Sigma^{-1} (y - f(x, \theta)), \quad (2.42)$$

onde Σ representa a matriz de covariância dos dados observados [50].

A função de verossimilhança associada ao qui-quadrado é definida como:

$$L(\theta|x, y) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2}\chi^2\right), \quad (2.43)$$

aqui, n denota o número de observações nos dados [50].

A interpretação da estatística do qui-quadrado no contexto da verossimilhança é a seguinte: uma estatística χ^2 baixa indica que os dados observados são coerentes com as expectativas do modelo ($y \approx f(x, \theta)$), o que resulta em uma alta probabilidade $L(\theta|y)$. Por outro lado, um valor alto para χ^2 sugere uma disparidade significativa entre os dados observados e as expectativas do modelo, resultando em uma baixa probabilidade de verossimilhança [50].

Ao empregar técnicas como a Inferência de Máxima Verossimilhança (MLE), obtemos os valores de θ que maximizam a função de verossimilhança, tornando os dados observados mais prováveis sob o modelo especificado, dada a estatística do qui-quadrado [50].

Este contexto ressalta não apenas a importância fundamental da verossimilhança na interpretação da estatística do qui-quadrado, mas também enfatiza como essa estatística é vital para avaliar a qualidade do ajuste do modelo aos dados observados [50].

2.6 Inferência estatística

A inferência estatística é um ramo formal e essencial da estatística que lida com a tomada de decisões sobre populações com base em informações limitadas extraídas de amostras representativas. Sua base matemática sólida permite a aplicação de métodos estatísticos para inferir características, padrões e relações em dados observados. No contexto da análise de dados, a inferência estatística desempenha um papel crítico, permitindo que pesquisadores avaliem hipóteses, estabeleçam relações causa-efeito e realizem previsões. Em campos como a cosmologia, onde os dados astronômicos são complexos e multidimensionais, técnicas de inferência estatística, incluindo a maximização da verossimilhança e a análise bayesiana, são cruciais. Elas não apenas ajudam a ajustar modelos teóricos às observações, mas também permitem estimativas precisas de parâmetros cosmológicos, proporcionando insights profundos sobre a natureza e a evolução do universo. Ao longo deste capítulo, faremos uma exploração detalhada dos princípios fundamentais da inferência estatística, incluindo tópicos como a matriz de Fisher e a verossimilhança, juntamente com outros conceitos essenciais. Esta análise aprofundada visa a solidificar nossa compreensão teórica, aplicável não apenas em análise de dados, mas também em contextos de inteligência artificial e cosmologia [6].

2.6.1 Matriz de Fisher

A matriz de Fisher desempenha um papel crucial na inferência estatística, especialmente em contextos como a Inferência de Máxima Verossimilhança (MLE) e em estudos de incerteza paramétrica. No caso do MLE, a matriz de Fisher, também conhecida como matriz de informação observada, é uma medida da precisão das estimativas dos parâmetros. Ela é definida como o negativo do segundo derivado da função log-verossimilhança em relação

aos parâmetros, avaliado nos valores estimados dos parâmetros [53]. Matematicamente, para um vetor de parâmetros θ , a matriz de Fisher $I(\theta)$ é dada por

$$I(\theta) = -\mathbb{E} \left[\frac{\partial^2 \log L(\theta)}{\partial \theta \partial \theta^T} \right], \quad (2.44)$$

onde $L(\theta)$ é a função de verossimilhança e \mathbb{E} representa o valor esperado [53]. Esta matriz é fundamental porque, sob certas condições, a distribuição assintótica dos estimadores de MLE se aproxima de uma distribuição normal multivariada centrada nos valores verdadeiros dos parâmetros, com matriz de covariância dada pela inversa da matriz de Fisher:

$$\text{Cov}(\hat{\theta}_{MLE}) \approx [I(\theta)]^{-1}. \quad (2.45)$$

Assim, a matriz de Fisher é crucial para calcular intervalos de confiança e realizar testes de hipóteses no contexto de MLE, pois fornece informações sobre a dispersão dos estimadores [53].

A matriz de Fisher desempenha um papel fundamental na análise de dados e inteligência artificial, oferecendo uma ferramenta essencial para avaliar a qualidade dos modelos estatísticos [53]. Ela fornece uma medida crucial da precisão das estimativas e da confiabilidade das relações entre variáveis em um conjunto de dados. Na análise estatística, é amplamente utilizada para testar a significância global de um conjunto de variáveis explicativas em um modelo de regressão. Em inteligência artificial, especialmente em algoritmos de aprendizado supervisionado, a matriz de Fisher auxilia na avaliação da eficácia dos modelos preditivos, permitindo ajustes precisos para melhorar a precisão das previsões. Assim, ao fornecer uma estrutura matemática sólida para a validação e melhoria contínua dos modelos estatísticos e algoritmos de IA, a matriz de Fisher desempenha um papel crucial no avanço da análise de dados e na construção de sistemas de inteligência artificial mais robustos e confiáveis [53].

2.6.2 Inferência de máxima verossimilhança

A Inferência de Máxima Verossimilhança (MLE) é uma técnica estatística amplamente utilizada para estimar os parâmetros de um modelo probabilístico. Seja X uma variável aleatória representando os dados observados e θ um vetor de parâmetros desconhecidos do modelo estatístico [6]. A função de verossimilhança $L(\theta|X)$ é definida como a densidade de probabilidade conjunta dos dados X , considerada como uma função dos parâmetros θ

$$L(\theta|X) = f(X|\theta), \quad (2.46)$$

onde $f(X|\theta)$ representa a função de densidade de probabilidade dos dados X dado que os parâmetros do modelo são θ [6]. Se os dados são independentes e identicamente

distribuídos (i.i.d.), a função de verossimilhança para n observações é dada por:

$$L(\theta|X) = \prod_{i=1}^n f(x_i|\theta) \quad (2.47)$$

onde x_i são os pontos de dados individuais no conjunto de dados X .

O objetivo do MLE é encontrar os valores dos parâmetros θ que maximizam a função de verossimilhança. Formalmente, os estimadores de MLE, $\hat{\theta}_{MLE}$, são definidos como

$$\hat{\theta}_{MLE} = \arg \max_{\theta} L(\theta|X). \quad (2.48)$$

Esses estimadores são escolhidos para tornar os dados observados X mais prováveis sob o modelo especificado, representando a estimativa pontual dos parâmetros que melhor se ajustam aos dados observados de acordo com o modelo probabilístico escolhido [6].

2.6.3 Cadeias de Markov Monte Carlo

As Cadeias de Markov Monte Carlo (do inglês, MCMC) são uma técnica probabilística e estatística que possui raízes profundas na necessidade de abordar problemas complexos que desafiam a resolução analítica direta. Sua origem remonta à década de 1940, durante o Projeto Manhattan, quando cientistas como Stanislaw Ulam e John von Neumann enfrentaram complexidades inerentes aos cálculos envolvidos na construção da bomba atômica. Em busca de soluções numéricas, eles conceberam uma abordagem baseada em amostragem aleatória que, ao longo do tempo, evoluiu para as cadeias de Markov Monte Carlo que conhecemos hoje. Com fundamentos matemáticos sólidos, este método utiliza a geração de números aleatórios para aproximar soluções para problemas que envolvem integração multidimensional, simulações estocásticas e avaliações de probabilidades. Desde então, a técnica das MCMC tem se expandido exponencialmente, com contribuições de matemáticos, estatísticos e cientistas da computação, tornando-se um dos pilares da computação científica e da análise estatística [30].

Além das suas raízes na física nuclear, o método das MCMC expandiu-se para revolucionar campos complexos como cosmologia, inteligência artificial e análise de dados. Na cosmologia, cientistas utilizam técnicas de MCMC para simular a evolução do universo, entender a distribuição de matéria escura e prever a formação de estruturas galácticas em larga escala. Na inteligência artificial, algoritmos baseados em MCMC são essenciais para técnicas de aprendizado por reforço, permitindo que sistemas de inteligência artificial tomem decisões em ambientes dinâmicos e incertos. Além disso, na análise de dados, especialmente em estudos epidemiológicos e pesquisas de mercado, as MCMC são empregadas para analisar incertezas em modelos estatísticos complexos, proporcionando insights valiosos para a tomada de decisões informadas. Com o avanço da computação de alto desempenho e o surgimento de algoritmos mais sofisticados, as MCMC continuam

a evoluir, moldando o panorama da pesquisa científica e tecnológica, e revelando-se uma ferramenta inestimável na exploração do universo e na análise inteligente de dados em diversas disciplinas [46].

Cadeias de Markov

O método de Cadeias de Markov Monte Carlo é um método que une dois conceitos fundamentais: as denominadas Cadeias de Markov e o Método de Monte Carlo, onde as Cadeias de Markov são uma classe dos denominados processos estocásticos [39].

Os processos estocásticos são uma família de variáveis aleatórias $\{X(t) : t \in T\}$ definidas em um espaço de probabilidade comum, onde T é um conjunto de índices. Cada variável $X(t)$ é uma função mensurável que mapeia do espaço de amostragem para os números reais. A coleção é estocástica, indicando que as variáveis estão sujeitas a incerteza, e a indexação por t denota a dependência temporal ou espacial entre as variáveis. Estes processos fornecem uma estrutura matemática rigorosa para lidar com a aleatoriedade em sistemas dinâmicos e experimentos observacionais [41].

Uma Cadeia de Markov é um processo estocástico no qual a probabilidade de transição de um estado para o próximo depende apenas do estado atual, e não dos estados anteriores. Formalmente, seja X_t uma variável aleatória que representa o estado do sistema no tempo t , a cadeia de Markov deve satisfazer a propriedade

$$P(X_{t+1} = x | X_t = x_t, X_{t-1} = x_{t-1}, \dots, X_0 = x_0) = P(X_{t+1} = x | X_t = x_t). \quad (2.49)$$

Isso implica que a probabilidade de transição de um estado para outro só depende do estado atual X_t , não importa como o sistema chegou a esse estado. As probabilidades de transição podem ser representadas por uma matriz de transição P , onde P_{ij} é a probabilidade de transição de i para j [39].

O método de Monte Carlo

O Método de Monte Carlo é uma abordagem estatística e computacional que se baseia na utilização de amostragem aleatória para estimar valores ou realizar cálculos complexos em processos estocásticos. Considerando um processo estocástico definido por uma variável aleatória X com distribuição de probabilidade $p(x)$, o método gera N amostras independentes x_1, x_2, \dots, x_N de X [39]. Essas amostras são então utilizadas para estimar quantidades de interesse, como o valor esperado $\mathbb{E}[f(X)]$ de uma função $f(X)$, através da média das avaliações da função para cada amostra

$$\hat{\mathbb{E}}[f(X)] = \frac{1}{N} \sum_{i=1}^N f(x_i), \quad (2.50)$$

além disso o método permite uma estimativa para a variância de $f(X)$, através da Equação 2.34, onde

$$\mathbb{V}[f(X)] = \frac{1}{N} \sum_{i=1}^N [f(x_i)]^2 - \frac{1}{N^2} \left(\sum_{i=1}^N f(x_i) \right)^2. \quad (2.51)$$

A utilização do método das Cadeias de Markov para a amostragem aleatória utilizada no método de Monte Carlo define o denominado Método das Cadeias de Markov Monte Carlo (do inglês MCMC) [39].

As MCMC representam uma abordagem essencial na análise estatística, permitindo a solução de problemas complexos de maneira eficaz. Ao combinar princípios das Cadeias de Markov e do Método de Monte Carlo, as MCMC oferecem uma maneira poderosa de explorar espaços de parâmetros, realizar inferências estatísticas e fazer previsões precisas. Essa técnica, fundamentada na aleatoriedade controlada, tem se tornado uma ferramenta central na análise de dados, proporcionando insights valiosos em diversas áreas do conhecimento. O papel fundamental das MCMC na análise estatística continua a crescer, consolidando seu lugar como uma metodologia indispensável para pesquisadores e cientistas de dados [46].

3 Redes neurais

O cérebro é uma das estruturas mais extraordinárias do organismo humano, representando uma complexidade única. Este órgão magnífico, residindo dentro da caixa craniana, é composto por aproximadamente 86 bilhões de neurônios interligados por trilhões de conexões sinápticas, formando uma rede de comunicação neural [15].

Além de suas funções perceptivas primárias, o cérebro é a sede das capacidades cognitivas superiores, possibilitando a compreensão de conceitos abstratos, o processamento emocional e a aquisição de conhecimento. Ele também desempenha um papel crucial na consolidação de memórias e na coordenação de uma vasta gama de atividades motoras, desde as mais básicas, como respiração e reflexos, até tarefas mais elaboradas, como a execução de tarefas criativas e a resolução de desafios matemáticos [19].

Sua operação, embora central para nossa existência, continua a ser um dos maiores enigmas da ciência, estimulando pesquisadores a desvendar os mistérios desse órgão fascinante que é o epicentro de nossa mente e consciência.

A rede de comunicação neural forma circuitos responsáveis por funções essenciais, como o pensamento, a memória, o movimento e o controle das emoções. Estes circuitos são coordenados por quatro lobos cerebrais principais, frontal, parietal, temporal e occipital:

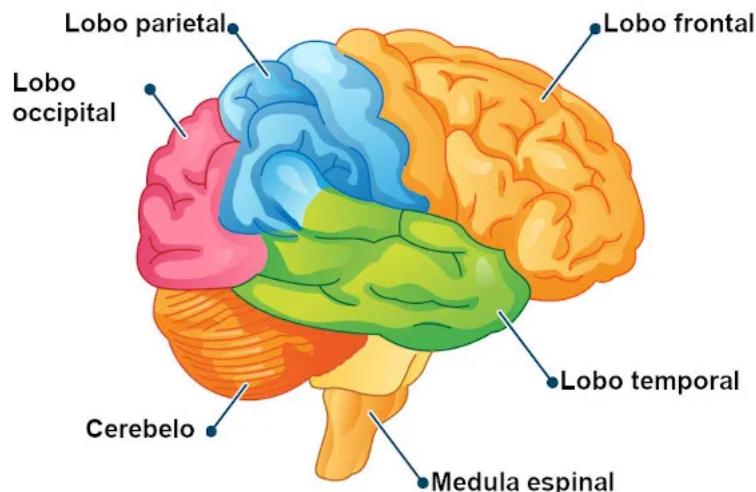


Figura 3.1: Representação das principais estruturas do cérebro humano. Fonte: Brasil Escola [9].

Além disso, áreas especializadas, como o córtex cerebral, que constitui a camada externa do cérebro, estão envolvidas em processos cognitivos superiores, enquanto o sistema límbico, composto por estruturas como a amígdala e o hipocampo, desempenha um papel crucial no processamento das emoções e das memórias [56].

Certamente, ao longo da história da humanidade, a complexidade e as capacidades do cérebro humano têm sido objeto de profundo fascínio. À medida que nossa compreensão do funcionamento cerebral e suas realizações notáveis avança, naturalmente surge o desejo de tentar modelar matematicamente o cérebro humano, o que se evidencia na década de 1940 com as primeiras iniciativas nessa direção.

Uma das tentativas pioneiras de modelar o cérebro matematicamente remonta ao trabalho do neurocientista Warren McCulloch e do lógico Walter Pitts em 1943 [29]. Eles propuseram um modelo conhecido como "Neurônio de McCulloch-Pitts" ou "Máquina de Turing Neuronal", que buscava representar as operações lógicas do cérebro através de um sistema simplificado de unidades computacionais.

Na década de 1940, já havia um conhecimento em evolução sobre o funcionamento do cérebro. No século XVII, René Descartes propôs que o cérebro era a sede da mente [8]. Durante o século XIX, Franz Joseph Gall introduziu a frenologia, uma teoria que afirmava que características da personalidade e habilidades mentais poderiam ser mapeadas através da forma do crânio, embora tenha sido posteriormente desacreditada devido à falta de evidências científicas sólidas para apoiá-la [16].

Camillo Golgi e Santiago Ramón y Cajal, no final do século XIX, estabeleceram a teoria do neurônio usando técnicas de coloração [17]. Em 1936, Otto Loewi e Henry Dale receberam o Prêmio Nobel por suas descobertas sobre neurotransmissão [26].

Assim, quando McCulloch e Pitts desenvolveram o seu modelo, acreditava-se que o cérebro funcionava principalmente como um sistema elétrico e químico. Os neurônios, as células fundamentais do sistema nervoso, eram vistos como unidades discretas que transmitiam informações por meio de sinais elétricos ao longo de suas fibras, conhecidas como axônios. As sinapses, as conexões entre neurônios, eram percebidas como pontos de contato onde ocorriam transmissões químicas de sinais entre células nervosas. Resumidamente, o processamento das informações no cérebro era visto de maneira bastante simplista, como um circuito elétrico que respondia a estímulos sensoriais e disparava respostas motoras.

O modelo McCulloch-Pitts, conhecido como neurônios artificiais, consistia em unidades binárias que podiam estar ativadas (1) ou desativadas (0) e eram interconectadas por conexões ponderadas que representavam a força das sinapses. A ideia central era que esses neurônios podiam realizar operações lógicas, como AND, OR e NOT, permitindo a construção de circuitos lógicos para processamento de informações. Embora tenha sido uma contribuição importante para a compreensão inicial das redes neurais, o modelo era simplificado em comparação com as redes neurais modernas [29].

3.1 O perceptron

Após a criação do modelo McCulloch-Pitts, o conceito de neurônios artificiais foi amadurecendo e foram publicados diversos trabalhos baseados na definição dos neurônios artificiais, como o trabalho de Weaver [55] e o trabalho de Rosenblatt [40].

Weaver discutiu a ideia de redes de neurônios artificiais como um modelo potencial para simular o processamento de informações em sistemas biológicos, incluindo o cérebro, já Rosenblatt introduziu o conceito de perceptron, que é uma unidade de processamento neural artificial.

O perceptron funciona recebendo entradas ponderadas, onde cada entrada é multiplicada por um peso associado. Esses valores ponderados são somados e passados por uma função de ativação binária, geralmente uma função degrau. A função de ativação determina se o perceptron dispara (produz uma saída de 1) ou não (produz uma saída de 0), com base em um limiar. O perceptron (representado pela Figura 3.2) foi projetado para aprender por meio de treinamento supervisionado, ajustando os pesos das entradas de acordo com o erro entre a saída do perceptron e a saída desejada durante o treinamento [40].

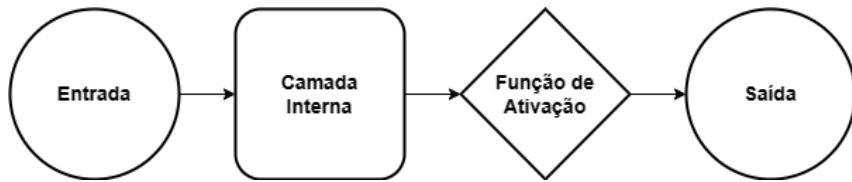


Figura 3.2: Diagrama de um perceptron genérico. Fonte: Criado pelo autor.

Se for tomada a função de ativação

$$f(t) = \begin{cases} 0, & \text{se } t < 0 \\ 1, & \text{se } t \geq 0 \end{cases}, \quad (3.1)$$

junto com um valor de bias de $b = -1$ e um vetor de peso de $w = (1; 1)$ o perceptron esquemátizado na Figura 3.2 pode representar a operação lógica OR, representada na tabela

Tabela 3.1: Saídas do perceptron para a operação lógica OR

x	$w \cdot x$	$w \cdot x + b$	Saída
(0;0)	0	-1	0
(0;1)	1	0	1
(1;0)	1	0	1
(1;1)	1	1	1

Se for tomada a função de ativação

$$f(t) = \begin{cases} 1, & \text{se } t < 0 \\ 0, & \text{se } t \geq 0 \end{cases}, \quad (3.2)$$

um valor de bias de $b = 1$ e um vetor de peso de $w = (-1; -1)$, então o perceptron esquematizado na Figura 3.2 pode representar a operação lógica AND, representado na tabela

Tabela 3.2: Saídas do perceptron para a operação lógica AND

x	$w \cdot x$	$w \cdot x + b$	Saída
(0;0)	0	1	0
(0;1)	-1	0	0
(1;0)	-1	0	0
(1;1)	-2	-1	1

Embora o perceptron tenha sido um marco importante no desenvolvimento das redes neurais, ele possui limitações significativas. Uma dessas limitações essenciais é a capacidade limitada de discernir entre diferentes grupos de dados quando esses grupos não podem ser facilmente separados por critérios simples [40].

Mesmo com o uso de múltiplas camadas de perceptrons interconectados (rede neural de uma camada), as redes construídas apenas com perceptrons simples têm limitações na representação de funções complexas, como aquelas encontradas em tarefas de visão computacional, processamento de linguagem natural e outras aplicações avançadas de aprendizado de máquina.

Após os perceptrons, houve um período de relativo declínio no interesse em redes neurais artificiais, conhecido como o "inverno da inteligência artificial", que durou até o final da década de 1980 e início da década de 1990 [49].

Durante inverno da inteligência artificial, as expectativas iniciais de que a IA poderia emular a inteligência humana em breve encontraram obstáculos significativos. Esse período de desaceleração notável foi influenciado por limitações tecnológicas da época, incluindo o perceptron, que também evidenciou suas próprias limitações.

Os computadores disponíveis careciam do poder de processamento e da capacidade de armazenamento necessários para lidar eficazmente com problemas complexos de IA. Consequentemente, o financiamento público e privado para pesquisa em IA diminuiu, levando a uma estagnação na IA com progressos notavelmente lentos e desafios técnicos substanciais que precisavam ser superados antes que a IA pudesse atingir sua atual maturidade e relevância [49].

3.2 A primavera da inteligência artificial: redes neurais modernas

O fim do ”inverno da inteligência artificial” começou a se delinear no final da década de 1980 e início da década de 1990, à medida que várias mudanças significativas ocorreram no cenário tecnológico e de pesquisa. Primeiramente, os avanços na capacidade de processamento de computadores e na capacidade de armazenamento permitiram que os pesquisadores explorassem modelos de IA mais complexos e exigentes computacionalmente. Além disso, novos algoritmos e abordagens, como redes neurais profundas e algoritmos de aprendizado de máquina mais sofisticados, começaram a ser desenvolvidos e demonstraram ser eficazes em uma variedade de tarefas [49].

Com o renascimento da área de inteligência artificial, as redes neurais modernas representam uma evolução substancial em relação ao perceptron original. Uma das mudanças cruciais é a adoção de funções de ativação não-lineares, que permitem a modelagem de relações complexas nos dados. Além disso, o desenvolvimento do algoritmo de retropropagação (backpropagation) revolucionou o treinamento de redes neurais modernas, tornando possível o aprendizado eficiente em várias camadas ocultas, o que era inatingível para o perceptron. Essas redes, com sua profundidade e capacidade de representação não-linear, são aplicadas com sucesso em uma ampla variedade de tarefas de inteligência artificial, desempenhando um papel fundamental em avanços recentes na área [18].

Uma rede neural artificial pode ser formalmente definida como uma função matemática composta por várias camadas de neurônios interconectados [18, 33]. Essa função recebe um vetor de entrada \mathbf{X} e produz uma saída \mathbf{Y} por meio de uma sequência de transformações matemáticas.

Suponha que tenhamos L camadas na rede neural, incluindo a camada de entrada e a camada de saída. Cada camada l possui N_l neurônios, onde $l = 1, 2, \dots, L$. Os neurônios em uma camada l são conectados a todos os neurônios da camada $l - 1$ por meio de pesos sinápticos.

Denotemos as saídas dos neurônios na camada l como um vetor $\mathbf{A}^{(l)} = [a_1^{(l)}, a_2^{(l)}, \dots, a_{N_l}^{(l)}]$. O vetor de entrada \mathbf{X} é considerado como $\mathbf{A}^{(0)}$, a camada de entrada.

Para calcular as saídas dos neurônios em cada camada, aplicamos a seguinte fórmula para $l = 1, 2, \dots, L$:

$$\mathbf{Z}^{(l)} = \mathbf{W}^{(l)} \mathbf{A}^{(l-1)} + \mathbf{b}^{(l)} \quad (3.3)$$

e

$$\mathbf{A}^{(l)} = f(\mathbf{Z}^{(l)}). \quad (3.4)$$

Onde:

- $\mathbf{W}^{(l)}$ é uma matriz de pesos de dimensão $N_l \times N_{l-1}$, que contém os pesos sinápticos entre a camada l e a camada $l - 1$ [18].
- $\mathbf{b}^{(l)}$ é um vetor de bias de dimensão N_l , que representa os valores de deslocamento (bias) adicionados aos neurônios na camada l [18].
- $f^{(l)}$ é uma função de ativação aplicada elemento a elemento a $\mathbf{Z}^{(l)}$, que introduz não linearidade na rede neural. Essa função é escolhida de acordo com a tarefa em questão [33].

A saída da rede neural é dada pelo vetor $\mathbf{Y} = \mathbf{A}^{(L)}$, que é a saída da camada de saída.

Durante o treinamento da rede neural artificial, um algoritmo de otimização é usado para ajustar os pesos $\mathbf{W}^{(l)}$ e os biases $\mathbf{b}^{(l)}$ a fim de minimizar uma função de perda específica, através do procedimento denominado algoritmo de retropropagação [18]. Podemos entender os pesos como análogos às sensibilidades do neurônio às diferentes entradas, determinando a influência relativa de cada entrada na saída do neurônio. O viés pode ser comparado à sensibilidade geral do neurônio, permitindo que ele ajuste seu limiar de ativação. Por fim, a função de ativação, representada como f , pode ser interpretada como o padrão de ativação do neurônio, definindo se o neurônio será ativado ou não com base na soma ponderada das entradas e do viés [24].

O treinamento de redes neurais é um pilar fundamental na área da aprendizagem profunda, capacitando esses modelos complexos a aprender com dados e executar tarefas específicas com precisão. Esse procedimento envolve diversas etapas essenciais [18].

Tudo começa com a propagação direta, na qual os dados de entrada são processados pela rede, camada por camada, desde a camada de entrada até a camada de saída. Em cada camada, as ativações são calculadas mediante o uso dos pesos e vieses da rede naquele instante. Em seguida, uma função de perda, também chamada de função de custo, é calculada para avaliar o quanto bem a rede está desempenhando a tarefa específica. A escolha da função de perda depende do tipo de problema a ser resolvido, seja ele de classificação, regressão ou outro [18].

A fase crítica é a retropropagação, um processo matemático que calcula os gradientes da função de perda em relação aos pesos e vieses da rede. Isso é feito aplicando a regra da cadeia para propagar o gradiente da camada de saída de volta à camada de entrada. Com os gradientes calculados, um algoritmo de otimização, como o Gradiente Descendente, entra em ação para ajustar os pesos e vieses da rede. O objetivo é minimizar a função de perda, e esse processo ocorre de forma iterativa, movendo os parâmetros na direção oposta ao gradiente [18].

Esse ciclo de ações, que inclui o feedforward, o cálculo da função de perda, a retropropagação e a atualização dos pesos, é repetido várias vezes, conhecidas como épocas, até que a função de perda alcance um valor mínimo ou até que ocorra uma convergência satisfatória.

3.3 Função de ativação

Compreender a teoria matemática por trás das funções de ativação é fundamental para desvendar o papel central que desempenham nas redes neurais. Essas funções são os alicerces que determinam como os neurônios artificiais processam informações. Em sua essência, as funções de ativação têm o poder de inserir não linearidade nas operações realizadas pelos neurônios. Em vez de seguir uma relação linear rígida entre entrada e saída, as funções de ativação possibilitam que os neurônios representem relações complexas e não lineares nos dados [10].

Um aspecto crucial das funções de ativação é sua habilidade de converter uma ampla gama de valores de entrada em um intervalo limitado de saídas. Essa compressão de informações é essencial para que as redes neurais representem eficientemente os padrões e características significativas dos dados, tornando-as adequadas para tarefas de aprendizado de máquina e reconhecimento de padrões [24].

Além disso, a escolha da função de ativação desempenha um papel crítico no treinamento da rede neural. Funções de ativação mais antigas, que eram mais suscetíveis ao problema do gradiente desaparecendo, tornavam o treinamento de redes profundas um desafio. No entanto, funções de ativação mais recentes, que superam esses problemas, têm se mostrado mais adequadas para redes neurais profundas, destacando a importância da seleção apropriada dessas funções para o sucesso do aprendizado de máquina [10].

Dentre as diversas funções de ativação, se destaca a ReLU (Rectified Linear Unit), que é uma inovação significativa em redes neurais profundas, desenvolvida ao longo de anos com contribuições de pesquisadores notáveis, como Geoffrey Hinton [32]. A ReLU é definida através da equação

$$f(x) = \begin{cases} x & , \text{se } x > 0 \\ 0 & , \text{se } x \leq 0 \end{cases} \quad (3.5)$$

e uma de suas principais vantagens é a capacidade de mitigar o desafio crítico do desaparecimento de gradientes em redes profundas [10]. Esse problema de gradiente ocorre quando, durante o treinamento, os gradientes que guiam os ajustes dos pesos nas camadas mais profundas da rede se tornam tão pequenos que essencialmente não contribuem para as atualizações dos pesos, levando à estagnação do aprendizado.

A função de ativação GELU (Gaussian Error Linear Unit), definida pela equação

$$f(x) = \frac{x}{2} \left(1 + \operatorname{erf} \left(\frac{x}{\sqrt{2}} \right) \right), \quad (3.6)$$

foi concebida como uma alternativa suave à função ReLU, a qual pode ocasionalmente enfrentar o problema dos "neurônios mortos", onde a saída permanece constantemente em zero. "Neurônios mortos" são um fenômeno em redes neurais em que o neurônio não é

ativado para nenhuma faixa específica de entradas, tornando-o ineficaz na aprendizagem e na propagação de informações através da rede [21].

Compreender a teoria matemática por trás das funções de ativação revela sua importância fundamental nas redes neurais. Essas funções são os alicerces que moldam o processamento de informações pelos neurônios artificiais, permitindo a representação de relações complexas nos dados. Além disso, sua capacidade de comprimir informações é essencial para o sucesso em tarefas de aprendizado de máquina e reconhecimento de padrões. A evolução das funções de ativação, exemplificada pela ReLU e GELU, destaca a necessidade contínua de pesquisa e inovação para superar desafios como o gradiente desaparecendo e os "neurônios mortos", evidenciando a importância da seleção apropriada dessas funções para o treinamento eficaz de redes neurais profundas.

3.4 Solucionando um problema curioso usando redes neurais

Com o intuito de exemplificar o funcionamento de uma rede neural na prática, definiremos uma função teste f , que recebe duas variáveis (x, y) , onde o domínio da variável x é o conjunto contendo as strings 'sim' e 'não' e o domínio da variável y é o conjunto dos números reais superiores ou iguais a 0. Já a imagem das funções e o conjunto contendo as strings 'você é legal' e 'você não é legal'. Esse domínio é representado pela equação

$$\begin{aligned} f : \{\text{sim; não}\} \times \mathbb{R}^+ &\rightarrow \{\text{você é legal; você não é legal}\} \\ (x, y) &\mapsto f(x, y). \end{aligned} \tag{3.7}$$

A função é definida pela equação

$$f(x, y) = \begin{cases} \text{você é legal} & \text{se } x = \text{sim e } y \leq 20 \\ \text{você não é legal} & \text{se } x = \text{sim e } y > 20 \\ \text{você é legal} & \text{se } x = \text{não e } y > 20 \\ \text{você não é legal} & \text{se } x = \text{não e } y \leq 20 \end{cases} \tag{3.8}$$

e representa a seguinte situação abstrata: se uma pessoa gostar de chocolate ($x = \text{sim}$) e tiver 20 anos ou menos ($y \leq 20$) ela é considerada legal, se a pessoa gostar de chocolate ($x = \text{sim}$) e tiver 20 anos ou mais ($y \geq 20$) ela é considerada legal.

Para o treinamento da rede neural será definido um conjunto 1000 pontos de teste e um conjunto de 100 pontos para treinar a rede, desses 100 pontos serão definidas uma porcentagem de pontos com o resultado incorreto.

O teste será realizado para uma rede neural composta por 3 camadas internas e será aplicada apenas uma única função de ativação para cada camada e o teste ocorrerá para

duas funções de ativação distintas, uma com a função de ativação ReLU e outra com a função de ativação Gelu.

O experimento foi realizado utilizando a biblioteca *torch* da linguagem de programação Python, na versão 3.8. O código correspondente encontra-se disponível no Anexo 9.1.1. Os resultados para a rede com a função de ativação ReLU estão apresentados na tabela

Tabela 3.3: Resultados do teste da Rede neural com a função de ativação ReLU.

–	Camadas com 64 neurônios	Camadas com 128 neurônios
Erro de treinamento	Precisão	Precisão
10%	74.77%	71.39%
20%	74.75%	71.81%
30%	50.58%	72.54%
40%	65.88%	56.71%
50%	51.15%	67.11%

e o resultado para a função de ativação Gelu se encontra na tabela

Tabela 3.4: Resultados do teste da Rede neural com a função de ativação Gelu.

–	Camadas com 64 neurônios	Camadas com 128 neurônios
Erro de treinamento	Precisão	Precisão
10%	71.60%	91.15%
20%	67.59%	67.18%
30%	66.16%	59.13%
40%	51.32%	57.65%
50%	48.74%	47.45%

Os resultados obtidos nos testes com as funções de ativação ReLU e Gelu revelam insights valiosos em relação ao desempenho da rede neural em diferentes cenários. Notavelmente, a função de ativação Gelu demonstrou uma capacidade robusta de lidar com dados de treinamento contaminados, apresentando precisões notáveis mesmo quando uma porcentagem significativa de dados estava incorreta. Por outro lado, a função ReLU mostrou uma sensibilidade maior à qualidade dos dados de treinamento, com sua precisão tendendo a declinar à medida que a porcentagem de dados incorretos aumentava. É importante salientar que esses resultados devem ser interpretados com cautela, pois os testes foram conduzidos principalmente para fins de visualização e não representam um rigoroso estudo estatístico, considerando a necessidade de controlar variáveis e considerar múltiplos conjuntos de dados.

Este capítulo nos levou a uma fascinante jornada pela complexidade do cérebro humano e sua representação nas redes neurais artificiais. Desde os primeiros modelos de neurônios artificiais de McCulloch-Pitts até as redes neurais profundas modernas, testemunhamos uma evolução notável na compreensão e modelagem do processamento neural.

Descobrimos que o cérebro humano, com sua capacidade extraordinária, continua a ser um enigma para a ciência, estimulando a busca incessante por sua compreensão. Além disso, nossos testes com diferentes funções de ativação em redes neurais nos mostraram que a escolha da arquitetura e da função de ativação desempenha um papel crucial no desempenho da rede, especialmente em cenários com dados de treinamento incorretos.

4 Método dos processos gaussianos

Nos domínios do aprendizado de máquina e estatísticas, os Processos Gaussianos (Gaussian Processes, GPs) emergem como uma ferramenta essencial para capturar relações complexas em dados. Originados da teoria das probabilidades, esses processos representam uma extensão poderosa das distribuições normais para espaços de dimensões infinitas, proporcionando uma abordagem probabilística robusta para a modelagem de funções. A característica central dos GPs reside na interpretação das funções como variáveis aleatórias, permitindo não apenas previsões pontuais, mas uma distribuição completa de possíveis resultados. Este capítulo investiga os alicerces teóricos dos Processos Gaussianos, explorando sua trajetória histórica e suas diversas aplicações em contextos variados [36].

O método foi originado na década de 1950 e representa um marco fundamental na estatística. Seu desenvolvimento inicial foi impulsionado por estudos pioneiros de matemáticos e estatísticos que buscavam entender a correlação espacial entre dados geográficos. A aplicação prática desses conceitos ganhou força nas décadas seguintes, à medida que cientistas e engenheiros começaram a utilizar o método para estimar valores em locais não amostrados em campos como geociências e engenharia [38]. Ao longo dos anos, com avanços tecnológicos, o método dos processos gaussianos evoluiu, integrando-se com técnicas modernas de análise de dados e estatísticas. Sua capacidade de fornecer estimativas precisas em espaços onde os dados são limitados ou ausentes continua a solidificar sua importância em várias disciplinas científicas, destacando-se como uma ferramenta essencial para a modelagem e compreensão de diversos tipos de dados [3].

O método dos processos gaussianos se fundamenta no Teorema do Limite Central na teoria das probabilidades. Esse teorema estabelece que a distribuição das médias de um grande número de variáveis aleatórias independentes e identicamente distribuídas se aproxima de uma distribuição normal à medida que o tamanho da amostra aumenta. No contexto dos processos gaussianos, que envolvem observações em diferentes locais ou pontos no espaço, essa propriedade é crucial. Utilizando essa base teórica, o método dos processos gaussianos estima valores em locais não observados ao pressupor que as observações seguem uma distribuição gaussiana multivariada [7].

4.1 A distribuição gaussiana

A distribuição gaussiana, amplamente conhecida como distribuição normal, representa um conceito fundamental na teoria das probabilidades. Esta distribuição foi inicialmente formalizada por Carl Friedrich Gauss no início do século XIX, consolidando-se como um dos pilares essenciais na análise estatística de dados. A distribuição normal exibe uma forma de sino simétrica em torno da média, que representa o valor mais provável em um conjunto de dados, enquanto a dispersão ao redor desta média é quantificada pelo desvio padrão [48].

Um princípio central na estatística, o teorema central do limite, estabelece que à medida que o tamanho da amostra aumenta, a distribuição das médias das amostras de qualquer população tende a se aproximar de uma distribuição normal. Esse fenômeno é observado independentemente da forma da distribuição original, conferindo à distribuição normal uma relevância teórica e prática notável. Esta propriedade permite modelar aproximadamente muitos fenômenos naturais e sociais [7].

A distribuição normal desempenha um papel crucial em diversas disciplinas científicas, incluindo física, biologia, economia e engenharia, devido à sua capacidade de descrever a variabilidade e aleatoriedade em uma ampla gama de fenômenos. Além disso, serve como base para técnicas estatísticas avançadas, como intervalos de confiança, testes de hipóteses e regressão linear. Sua importância incontestável na análise de dados e na inferência estatística a torna uma ferramenta indispensável para pesquisadores e cientistas em todo o mundo [48].

A distribuição normal multivariada é caracterizada por um vetor de médias ($\boldsymbol{\mu}$) e uma matriz de covariância ($\boldsymbol{\Sigma}$). O vetor de médias representa os valores médios das variáveis no espaço multivariado, indicando o centro da distribuição ao longo de cada eixo. A matriz de covariância descreve como as variáveis se relacionam entre si, capturando tanto a dispersão individual de cada variável quanto as correlações entre elas [31].

A função de densidade de probabilidade da distribuição normal multivariada é dada por:

$$f(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x}-\boldsymbol{\mu})}. \quad (4.1)$$

Nesta equação, \mathbf{x} representa um vetor de variáveis aleatórias, n é o número de variáveis no vetor \mathbf{x} , $|\boldsymbol{\Sigma}|$ representa o determinante da matriz de covariância $\boldsymbol{\Sigma}$, $(\mathbf{x} - \boldsymbol{\mu})^T$ denota a transposta do vetor $(\mathbf{x} - \boldsymbol{\mu})$, e $\boldsymbol{\Sigma}^{-1}$ é a matriz inversa de $\boldsymbol{\Sigma}$ [48].

Seja \mathbf{x} uma variável aleatória com média $\boldsymbol{\mu}$ e variância $\boldsymbol{\Sigma}$, onde \mathbf{x} segue uma distribuição normal. Esta relação é denotada como

$$\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}). \quad (4.2)$$

Nesta notação, \mathcal{N} representa a distribuição normal, $\boldsymbol{\mu}$ é a média da distribuição e $\boldsymbol{\Sigma}$ é a matriz de covariância [31].

4.1.1 Propriedades da distribuição gaussiana: distribuição conjunta, marginal e condicional

Sejam \mathbf{X} e \mathbf{Y} duas variáveis aleatórias distribuídas normalmente, onde

$$\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}_X, \mathbf{A}) \quad (4.3)$$

e

$$\mathbf{Y} \sim \mathcal{N}(\boldsymbol{\mu}_Y, \mathbf{B}), \quad (4.4)$$

então a distribuição conjunta de \mathbf{X} e \mathbf{Y} será definida pela distribuição gaussiana

$$\begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu}_{\mathbf{X}} \\ \boldsymbol{\mu}_{\mathbf{Y}} \end{bmatrix}, \begin{bmatrix} \mathbf{A} & \mathbf{C} \\ \mathbf{C}^T & \mathbf{B} \end{bmatrix}\right), \quad (4.5)$$

e a distribuição condicional $\mathbf{X}|\mathbf{Y}$ será

$$\mathbf{X}|\mathbf{Y} \sim \mathcal{N}(\boldsymbol{\mu}_{\mathbf{X}} + \mathbf{C}\mathbf{B}^{-1}(\mathbf{Y} - \boldsymbol{\mu}_{\mathbf{Y}}), \mathbf{A} - \mathbf{C}\mathbf{B}^{-1}\mathbf{C}^T). \quad (4.6)$$

Se

$$\begin{bmatrix} \mathbf{A} & \mathbf{C} \\ \mathbf{C}^T & \mathbf{B} \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{A}} & \tilde{\mathbf{C}} \\ \tilde{\mathbf{C}}^T & \tilde{\mathbf{B}} \end{bmatrix}^{-1}, \quad (4.7)$$

então

$$\mathbf{X}|\mathbf{Y} \sim \mathcal{N}(\boldsymbol{\mu}_{\mathbf{X}} - \tilde{\mathbf{A}}^{-1}\tilde{\mathbf{C}}(\mathbf{Y} - \boldsymbol{\mu}_{\mathbf{Y}}), \tilde{\mathbf{A}}^{-1}). \quad (4.8)$$

As equações 4.3 e 4.4 são denominadas distribuições marginais de \mathbf{X} e \mathbf{Y} , respectivamente [36].

4.1.2 A matriz de covariância e o vetor de média

A matriz de covariância em uma distribuição normal multivariada é uma matriz simétrica que descreve a variabilidade conjunta das variáveis aleatórias envolvidas. O vetor de médias é um vetor que contém os valores esperados de cada variável no conjunto de dados, fornecendo uma medida da tendência central. Cada componente μ_i indica o valor médio da variável correspondente X_i . Compreender o vetor de médias é essencial, pois fornece informações instantâneas sobre o comportamento médio das variáveis e é crucial na análise da posição relativa das variáveis no espaço multivariado, facilitando a identificação de padrões ou desvios significativos [48].

Em suma, a distribuição normal é amplamente utilizada em estatística e aprendizado de máquina para modelar conjuntos de dados multivariados. Sua forma característica de sino implica que a maioria dos pontos de dados se concentra perto da região central da distribuição, com uma dispersão que é controlada pela matriz de covariância. A normal multivariada é fundamental para diversas aplicações, incluindo análise de dados, reconhecimento de padrões, e previsão, onde é essencial entender as relações complexas entre múltiplas variáveis aleatórias [31].

4.2 Os processos gaussianos

Os processos Gaussianos são processos estocásticos dos quais qualquer combinação linear finita de variáveis aleatórias no processo segue uma distribuição normal, isto é, um processo estocástico $X(t)$ é considerado Gaussiano se, para qualquer conjunto finito de instantes de tempo t_1, t_2, \dots, t_n e quaisquer constantes a_1, a_2, \dots, a_n , a variável aleatória $Y = a_1X(t_1) + a_2X(t_2) + \dots + a_nX(t_n)$ é uma variável aleatória gaussiana. Isso pode ser denotado por

$$Y = \sum_{i=1}^n a_i X(t_i) \sim \mathcal{N}(\mu_Y, \sigma_Y^2), \quad (4.9)$$

onde μ_Y representa o valor esperado da variável aleatória Y , σ_Y representa a variância da variável aleatória Y , e o processo gaussiano pode ser representado completamente por μ_Y e σ_Y [36, 41].

Os processos gaussianos podem ser representados no ponto de vista dos espaços das funções \mathbf{f} de uma variável \mathbf{x} , onde

$$\mathbf{f}(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')), \quad (4.10)$$

onde

$$m(\mathbf{x}) = \mathbb{E}[\mathbf{f}(\mathbf{x})] \quad (4.11)$$

representa o valor esperado da função \mathbf{f} no ponto \mathbf{x} e

$$k(\mathbf{x}, \mathbf{x}') = \mathbb{E}[(\mathbf{f}(\mathbf{x}) - m(\mathbf{x}))(\mathbf{f}(\mathbf{x}') - m(\mathbf{x}'))] \quad (4.12)$$

representa a covariância entre dois pontos \mathbf{x} e \mathbf{x}' .

A função de covariância e a função de média desempenham papéis fundamentais ao representar o conhecimento a priori sobre o comportamento do fenômeno investigado. Não obstante, o método dos processos gaussianos é distintivamente bayesiano, uma vez que se baseia na atualização iterativa das estimativas à luz dos dados observados, incorporando

assim o conhecimento prévio para formular previsões precisas e confiáveis sobre o fenômeno em estudo [36].

4.3 Realizando previsões com o método dos processos gaussianos

A aplicação do método dos processos gaussianos para previsões concentra-se na integração do conhecimento proveniente do conjunto de dados de treino na função a ser prevista [36].

Considere um conjunto contendo n observações $\{(x_i, y_i, \Delta y_i) | i = 1, 2, \dots, n\}$, se assumirmos que os dados observados se comportem com uma versão ruidosa de uma função $f(x)$, onde $y = f(x) + \epsilon$, e ϵ são ruidos independentes distribuídos a partir de uma gaussiana de variância σ_n^2 , então a distribuição conjunta entre os dados de treino y e a função prevista \mathbf{f}^* dado o conhecimento à priori será

$$\begin{bmatrix} y \\ \mathbf{f}^* \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} m(X) \\ m(X^*) \end{bmatrix}, \begin{bmatrix} K(X, X) + \sigma_n^2 I & K(X, X^*) \\ K(X^*, X) & K(X^*, X^*) \end{bmatrix} \right). \quad (4.13)$$

Seguindo essa metodologia, o procedimento envolve a geração de funções a partir da distribuição a priori. Essas funções são então filtradas, rejeitando aquelas que não se ajustam aos dados de treinamento. Calcula-se a média e o desvio padrão das funções selecionadas para obter o valor previsto, contudo essa estratégia é computacionalmente ineficiente [36].

Entretanto, dado que o método implica uma distribuição gaussiana, ao empregar as identidades derivadas na Seção 4.1.1, obtemos

$$\mathbf{f}^* | X, y, X^* \sim \mathcal{N}(\bar{\mathbf{f}}^*, \text{var}(\mathbf{f}^*)), \quad (4.14)$$

onde o valor médio da função \mathbf{f}^* é

$$\bar{\mathbf{f}}^* = m(X^*) + K(X^*, X)[K(X, X) + \sigma_n^2 I]^{-1}(y - m(X)) \quad (4.15)$$

e a variância da função \mathbf{f}^* é

$$\text{var}(\mathbf{f}^*) = K(X^*, X^*) - K(X^*, X)[K(X, X) + \sigma_n^2 I]^{-1}K(X, X^*). \quad (4.16)$$

Nesta seção, investigamos a aplicação do método dos processos gaussianos para previsões, com um foco na incorporação do conhecimento derivado do conjunto de dados de treino na função a ser prevista. Ao pressupormos que os dados observados são manifestações ruidosas de uma função subjacente, somos capazes de modelar a distribuição conjunta entre os dados de treino e a função prevista. Inicialmente, esse método apresentava desafios computacionais significativos, no entanto, ao aproveitarmos as propriedades

da distribuição gaussiana, conseguimos simplificar o processo de previsão [35].

4.3.1 Likelihood para o Método dos Processos Gaussianos

Uma alternativa útil à definição de uma função de média fixa $m(x)$ é a escolha de uma base $\mathbf{h}(x)$ de funções cujos coeficientes β são definidos a partir do conjunto de treino. Nessa perspectiva o processo gaussiano $g(x)$ gerado é

$$g(x) = f(x) + \mathbf{h}(x)^T \beta, \quad (4.17)$$

onde

$$f(x) \sim \mathcal{GP}(0, k(x, x')). \quad (4.18)$$

Essa formulação expressa que os dados estão próximos de um modelo linear global, com os resíduos sendo modelados por um processo gaussiano [36].

Se tomada uma distribuição à priori gaussiana para β , onde $\beta \sim \mathcal{N}(\mathbf{b}, \mathbf{B})$, então $g(x)$ pode ser representado por

$$g(x) \sim \mathcal{GP}(\mathbf{h}(x)^T b, k(x, x') + \mathbf{h}(x)^T \mathbf{B} \mathbf{h}(x')), \quad (4.19)$$

análogamente ao cálculo desenvolvido na Seção 4.3 se obtém a previsão

$$\bar{\mathbf{g}}(x^*) = H^{*T} \bar{\beta} + K^{*T} K_y^{-1} (y - H^T \bar{\beta}) = \bar{\mathbf{f}}^* + R^T \bar{\beta}, \quad (4.20)$$

com um erro

$$\text{var}(\mathbf{g}^*) = \text{cov}(\mathbf{f}^*) + R^T (\mathbf{B}^{-1} + H K_y^{-1} H^T)^{-1} R. \quad (4.21)$$

Nessa equação H representa a matriz da função $\mathbf{h}(x)$ aplicada sobre os dados de treinamento, H^* representa a matriz da função $\mathbf{h}(x)$ aplicada sobre os dados de previsão,

$$\bar{\beta} = (\mathbf{B}^{-1} + H K_y^{-1} H^T)^{-1} (H K_y^{-1} y + \mathbf{B}^{-1} \mathbf{b}), \quad (4.22)$$

$$R = H^* - H K_y^{-1} K^* \quad (4.23)$$

e

$$K_y = K + \sigma_n^2 I. \quad (4.24)$$

Nesse contexto, $\bar{\beta}$ representa a média dos parâmetros do modelo linear global. Essa média é calculada usando previsões dos dados de treinamento e do processo gaussiano a priori, além da previsão do modelo para o resíduo. A função de covariância resultante

é a soma da função de covariância do modelo de média nula com uma contribuição não negativa [36].

Com base nos resultados obtidos e seguindo a metodologia descrita na seção 4.1.1, se compararmos a Equação 4.19 com a Equação 4.5 e a Equação 4.3 (ou a Equação 4.4), então a likelihood marginal para o processo gaussiano com uma função de média arbitrária é definida como segue

$$\log p(y|X, \mathbf{b}, \mathbf{B}) = -\frac{1}{2}(H^T \mathbf{b} - \mathbf{y})^T (K_y + H^T \mathbf{B} H)^{-1} (H^T \mathbf{b} - \mathbf{y}) - \frac{1}{2} \log |K_y + H^T \mathbf{B} H| - \frac{n}{2} \log(2\pi). \quad (4.25)$$

Através dessa estratégia, o processo gaussiano $g(x)$ resultante representa uma combinação engenhosa entre um modelo linear global e um componente não linear, proporcionando uma representação mais precisa dos dados observados. Além disso, derivamos a likelihood marginal, uma medida fundamental da adequação do modelo aos dados, considerando tanto a estrutura do modelo quanto a incerteza nos parâmetros. Esse resultado é crucial para avaliar a confiabilidade das previsões do modelo.

4.4 Usando o método dos processos gaussianos para regressão bayesiana

Com o objetivo de demonstrar o funcionamento do método dos processos gaussianos, aplicaremos o método para prever uma função $\mathbf{f}(\mathbf{x})$ em um ponto \mathbf{x}^* , com base no conjunto fictício de pontos observacionais $(\mathbf{x}, \mathbf{y}, \mathbf{y}_{erro})$. Este experimento foi conduzido utilizando a biblioteca *numpy* na linguagem de programação Python, versão 3.8. O código correspondente encontra-se disponível no Anexo 9.1.2.

A função utilizada para o teste será a função definida por

$$f(x) = \frac{\sin x}{x}, \quad (4.26)$$

\mathbf{x} será o vetor

$$\mathbf{x} = (1.86755799, -0.97727788, 3.76405235, 2.40015721, -1.02126202, 0.2408932), \quad (4.27)$$

e \mathbf{y} será o vetor

$$\mathbf{y} = (0.51205289, 0.8482521, -0.15489564, 0.28137626, 0.83501376, 0.99035643), \quad (4.28)$$

que foi gerado por

$$\mathbf{y} = f(\mathbf{x}) + \mathbf{z}, \quad (4.29)$$

onde

$$\mathbf{z} \sim \mathcal{N}(0, \text{std}(f(x))/4). \quad (4.30)$$

O vetor \mathbf{y}_{erro} é o vetor

$$\mathbf{y}_{\text{erro}} = (0.09275247, 0.17442101, 0.25037341, 0.26832998, 0.00243041, 0.11319406), \quad (4.31)$$

onde cada componente de \mathbf{y}_{erro} foi gerado pela distribuição

$$\mathcal{N}(0, \text{std}(\mathbf{y})). \quad (4.32)$$

O processo gaussiano gerado foi definido pela função de média

$$m(x) = e^{-\frac{x^2}{l}}, \quad (4.33)$$

que é representada graficamente pela figura:



Figura 4.1: Função de média para o processo gaussiano. Fonte: Criado pelo autor.

A função de covariância adotada foi

$$k(x, x') = e^{-\frac{\|x - x'\|^2}{2\sigma^2}}, \quad (4.34)$$

onde a norma utilizada é a norma euclidiana e a função de covariância é representada graficamente pela figura:

Com a função de covariância (tomando $\sigma = 1.3$), a função de média (tomando $\ell = 2$) e

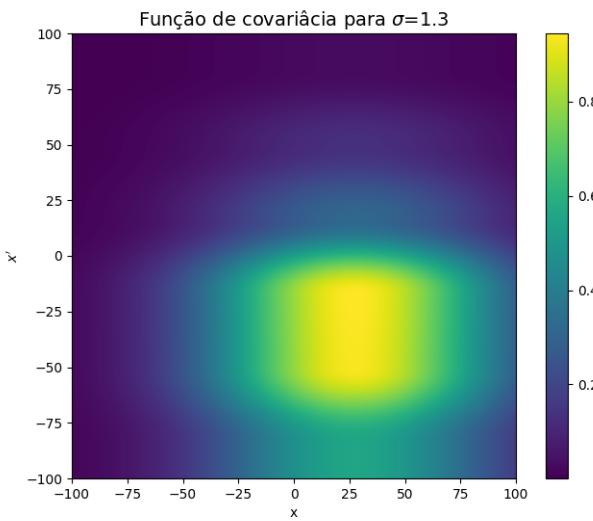


Figura 4.2: Função de covariância para o processo gaussiano. Fonte: Criado pelo autor.

os pontos de treinamento disponíveis, o processo gaussiano pode ser conduzido. Seguindo a metodologia descrita na Seção 4.3, faremos a previsão \mathbf{f}^* para os valores \mathbf{x}^* . Aqui, \mathbf{x}^* representa um vetor de 500 pontos igualmente espaçados distribuídos entre os valores mínimo e máximo de \mathbf{x} .

A distribuição à priori dos dados previstos é a distribuição representada pela Equação 4.10, ao aplicar essa Equação para os dados de treinamento se obtém o gráfico da seguinte figura:

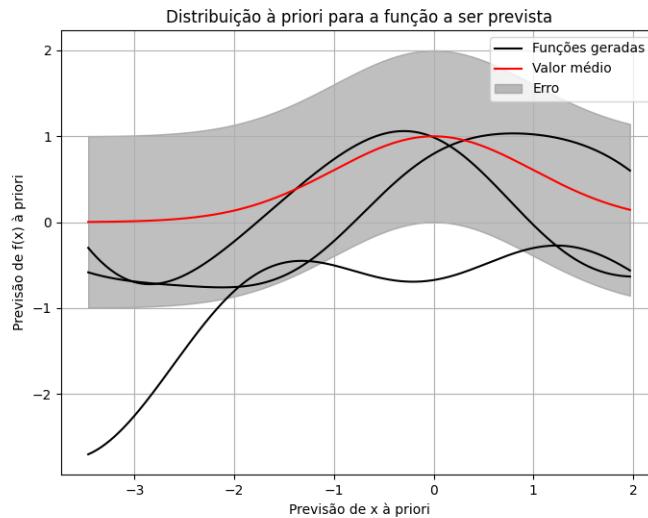


Figura 4.3: Distribuição à priori para \mathbf{x}^* . Fonte: Criado pelo autor.

A distribuição à posteriori para \mathbf{x}^* , obtida através da Equação 4.14, gera as funções representadas na seguinte figura:

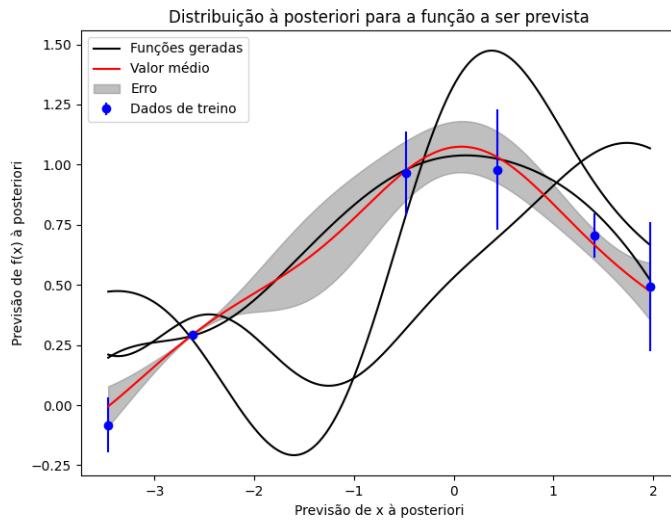


Figura 4.4: Distribuição à posteriori para \mathbf{x}^* . Fonte: Criado pelo autor.

Dessa metodologia, se obtém a previsão de $f(\mathbf{x}^*)$

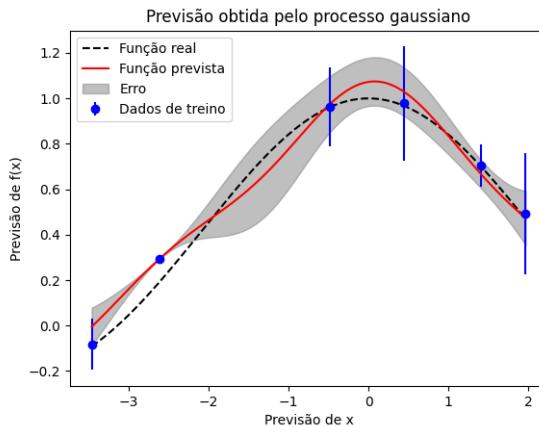


Figura 4.5: Previsão para \mathbf{x}^* . Fonte: Criado pelo autor.

Neste capítulo, aplicamos o método dos processos gaussianos para prever a função $f(x) = \frac{\sin x}{x}$ em um ponto específico \mathbf{x}^* , utilizando dados fictícios. Os resultados obtidos ilustram a eficácia do método, oferecendo uma previsão precisa e confiável para o valor desejado. Isso ressalta não apenas a robustez da abordagem, mas também sua capacidade de produzir previsões confiáveis em contextos complexos.

4.4.1 A função de covariância e a função de média

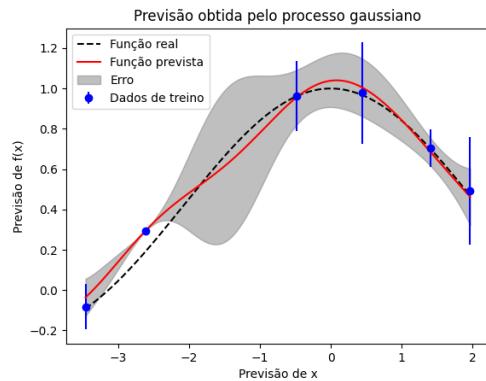
No exemplo da Seção 4.4, aplicamos o método dos Processos Gaussianos para fazer previsões sobre a variável \mathbf{x}^* com base nos dados de entrada. Para isso, selecionamos a função de covariância representada pela Equação 4.34 e a função de média representada

pela Equação 4.33. Neste contexto, exploraremos a fundo dois elementos essenciais no Método dos Processos Gaussianos: a função de média e a função de covariância. A função de média estabelece a base para as previsões, incorporando conhecimentos prévios sobre o comportamento dos dados, conforme discutido no capítulo anterior. Ela oferece uma estimativa inicial da tendência subjacente nos dados, sendo frequentemente escolhida com base nas intuições do domínio do problema [36].

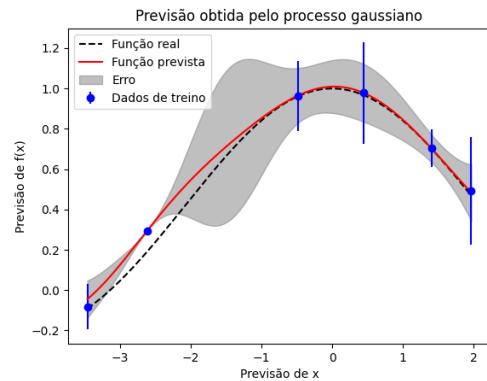
Por outro lado, a função de covariância desempenha um papel crucial ao capturar a relação e a dependência entre diferentes pontos no espaço de dados. A função de covariância permite que o processo gaussiano identifique padrões complexos e não lineares, sendo fundamental para a precisão das previsões [36]. No nosso estudo de caso, a escolha da função de covariância foi vital para analisar as relações intrincadas nos dados, enquanto a função de média proporcionou uma compreensão inicial da estrutura subjacente. Juntos, esses elementos constituem a base do nosso modelo de previsão, permitindo-nos explorar e entender as nuances dos dados de forma mais precisa.

Tanto a função de média quanto a função de covariância foram formalmente definidas com base em expressões matemáticas que incorporam parâmetros específicos. No contexto do exemplo apresentado na Seção 4.4, o parâmetro σ foi associado à função de covariância, enquanto o parâmetro ℓ foi associado à função de média. Estes parâmetros, que estão intrinsecamente ligados às funções de média e de covariância, são conhecidos como hiperparâmetros do modelo. É importante destacar que esses hiperparâmetros desempenham um papel fundamental no método dos processos gaussianos, influenciando diretamente as propriedades estatísticas e o comportamento do modelo. Sua escolha adequada é crucial para a precisão e confiabilidade das estimativas geradas pelo método dos processos gaussianos [36].

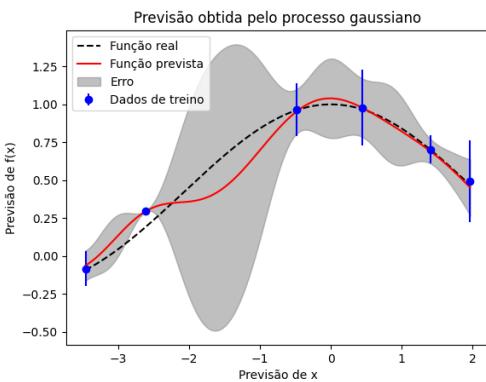
Para uma dada função de covariância e uma dada função de média, cada conjunto de hiperparâmetros $\{\theta_1, \theta_2\}$ define um modelo de processo gaussianos $\mathcal{GP}(m(\mathbf{x}|\theta_1), k(\mathbf{x}, \mathbf{x}'|\theta_2))$, a figura a seguir representa o resultado para o exemplo abordado na Seção 4.4 para diferentes conjuntos de hiperparâmetros $\{\sigma, \ell\}$:



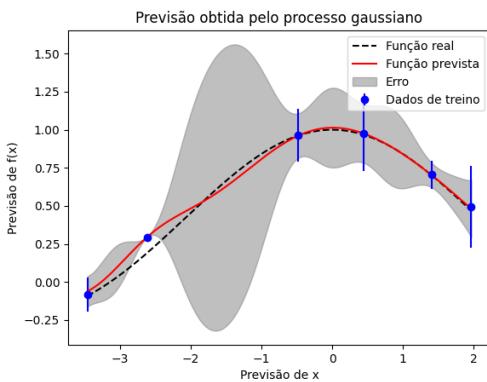
(a) Processo gaussiano para $\sigma = 3/2$ e $\ell = 6/5$, com $\chi^2 = 20489$. Fonte: Criado pelo autor.



(b) Processo gaussiano para $\sigma = 3/2$ e $\ell = 11/5$, com $\chi^2 = 20523$. Fonte: Criado pelo autor.



(c) Processo gaussiano para $\sigma = 5/2$ e $\ell = 6/5$, com $\chi^2 = 9265$. Fonte: Criado pelo autor.



(d) Processo gaussiano para $\sigma = 5/2$ e $\ell = 11/5$, com $\chi^2 = 9302$. Fonte: Criado pelo autor.

Figura 4.6: Comparação das regressões pelo método dos processos gaussianos para diferentes hiperparâmetros.

Ao analisar os gráficos apresentados na Figura 4.6, torna-se evidente que os processos gaussianos gerados por diferentes conjuntos de hiperparâmetros resultam em curvas distintas. Notavelmente, o valor do teste estatístico χ^2 varia significativamente, variando de ordens de magnitude entre 9265 e 20523. Este resultado sublinha a extrema sensibilidade do processo gaussiano ao se definir um conjunto de hiperparâmetros. Um conjunto inadequado de hiperparâmetros pode ter um impacto totalmente adverso nos resultados obtidos, enquanto a seleção cuidadosa de um conjunto de hiperparâmetros apropriado pode conduzir a previsões excepcionais [20].

Existem várias abordagens para o processo de seleção dos hiperparâmetros apropriados para um modelo, também conhecido como otimização de hiperparâmetros, tais como o método de Monte Carlo, entre outros. Em uma dessas técnicas, o objetivo é maximizar a verossimilhança. Notavelmente, o método dos processos gaussianos possui uma relação analítica para a verossimilhança, conforme descrito pela Equação 4.25. Devido a essa

propriedade, essa técnica se torna altamente valiosa. Ela é implementada em diversos algoritmos sofisticados que utilizam o método dos processos gaussianos, destacando-se como uma ferramenta crucial na busca pelos melhores hiperparâmetros para os modelos [55].

Em suma, a aplicação do Método dos Processos Gaussianos para prever variáveis com base em dados de entrada envolve a cuidadosa escolha da função de média e da função de covariância. Enquanto a função de média estabelece uma estimativa inicial da tendência nos dados, a função de covariância captura as relações complexas e não lineares entre diferentes pontos no espaço de dados. A escolha dos hiperparâmetros associados a essas funções é crítica, pois influencia diretamente a precisão das previsões. Como ilustrado em nosso exemplo, diferentes conjuntos de hiperparâmetros podem levar a resultados drasticamente diferentes, destacando a sensibilidade do Método dos Processos Gaussianos a esses valores. A seleção cuidadosa dos hiperparâmetros, muitas vezes realizada por meio de técnicas de otimização como a maximização da verossimilhança, é fundamental para garantir a confiabilidade e a precisão das estimativas geradas pelo método. Portanto, a escolha adequada dos hiperparâmetros é um passo essencial na aplicação bem-sucedida dos Processos Gaussianos para problemas de previsão, garantindo que o modelo capture com precisão as complexidades dos dados [20, 36].

4.4.2 Processos gaussianos para a previsão das derivadas

O método dos processos gaussianos também prevê a derivada da função analisada, e essa previsão é realizada a partir de um outro processo gaussiano, onde a nova função de covariância é obtida a partir da diferenciação da função de covariância original, onde

$$\text{cov} \left(f_i, \frac{\partial f_j}{\partial x_j} \right) = \frac{\partial k(x_i, x_j)}{\partial x_j} \quad (4.35)$$

e

$$\text{cov} \left(\frac{\partial f_i}{\partial x_i}, \frac{\partial f_j}{\partial x_j} \right) = \frac{\partial^2 k(x_i, x_j)}{\partial x_i \partial x_j}. \quad (4.36)$$

Portanto, dada uma função $\mathbf{f}(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'))$, o processo gaussianos para suas derivadas será

$$\mathbf{f}'(\mathbf{x}) \sim \mathcal{GP} \left(m'(\mathbf{x}), \frac{\partial^2 k(\mathbf{x}, \mathbf{x}')}{\partial x \partial x'} \right), \quad (4.37)$$

$$\mathbf{f}''(\mathbf{x}) \sim \mathcal{GP} \left(m''(\mathbf{x}), \frac{\partial^4 k(\mathbf{x}, \mathbf{x}')}{\partial x^2 \partial x'^2} \right), \quad (4.38)$$

etc [44].

A distribuição conjunta entre \mathbf{y} e $\mathbf{f}^{* \prime}$ será

$$\begin{bmatrix} y \\ \mathbf{f}^{*'} \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} m(X) \\ m'(X^*) \end{bmatrix}, \begin{bmatrix} K(X, X) + \sigma_n^2 I & K'(X, X^*) \\ K'(X^*, X) & K''(X^*, X^*) \end{bmatrix} \right), \quad (4.39)$$

Onde

$$[K'(X, X^*)]_{ij} = \frac{\partial k(x_i, x_j^*)}{\partial x_j^*} \quad (4.40)$$

e

$$[K''(X^*, X^*)]_{ij} = \frac{\partial^2 k(x_i^*, x_j^*)}{\partial x_i^* \partial x_j^*}. \quad (4.41)$$

A distribuição condicional de $\mathbf{f}^{*'}$ será

$$\mathbf{f}^{*'} | \mathbf{X}^*, \mathbf{X}, y \sim \mathcal{N}(\bar{\mathbf{f}}^{*'}, \text{var}(\mathbf{f}^{*'})), \quad (4.42)$$

onde

$$\bar{\mathbf{f}}^{*'} = m'(X^*) + K'(X^*, X)[K(X, X) + \sigma_n^2 I]^{-1}(y - m(X)) \quad (4.43)$$

e

$$\text{var}(\mathbf{f}^{*'}) = K''(X^*, X^*) - K'(X^*, X)[K(X, X) + \sigma_n^2 I]^{-1}K'(X, X^*). \quad (4.44)$$

Em muitos casos práticos, é essencial estudar funções dependentes de \mathbf{f} e suas derivadas. Nesses contextos, é valioso estabelecer uma relação analítica para as covariâncias entre \mathbf{f} e suas derivadas. A covariância é calculada mediante a seguinte relação:

$$\text{cov} \left(f^{*(i)}, f^{*(j)} \right) = k^{(i,j)}(x^*, x^*) - K^{(i)}(x^*, X)[K(X, X) + \sigma_n^2 I]^{-1}K^{(j)}(X, x^*), \quad (4.45)$$

onde $k^{(i,j)}(x^*, x^*)$ representa a i-ésima derivada de $k(x^*, x^*)$ com relação à primeira variável e a j-ésima derivada com relação à segunda variável. Essa relação é útil pois permite definir a distribuição de uma função $g^* = f(x^*)$ para cada ponto x^* :

$$\begin{bmatrix} \mathbf{f}^* \\ \mathbf{f}^{*'} \\ \vdots \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} m(X^*) \\ m'(X^*) \\ \vdots \end{bmatrix}, \begin{bmatrix} \text{var}(\mathbf{f}^*) & \text{cov}(\mathbf{f}^*, \mathbf{f}^{*'}) & \dots \\ \text{cov}(\mathbf{f}^{*'}, \mathbf{f}^*) & \text{var}(\mathbf{f}^{*'}) & \dots \\ \vdots & \vdots & \ddots \end{bmatrix} \right), \quad (4.46)$$

a partir desta distribuição, é possível gerar g^* por meio de um método estatístico, tal como o método de Monte Carlo.

Neste capítulo, exploramos o poder do método dos processos gaussianos para prever não apenas a função principal, mas também suas derivadas. A importância crucial da covariância entre a função e suas derivadas foi destacada, sendo um resultado de

extrema relevância para aplicações práticas. Demonstramos que o método dos processos gaussianos não apenas prevê a derivada da função analisada, mas o faz através de um processo gaussiano distinto. Essa previsão é fundamentada na nova função de covariância, obtida por meio da diferenciação da função de covariância original. Assim, estabelecemos as relações fundamentais entre a função principal, suas derivadas de primeira e segunda ordem, e as covariâncias correspondentes. Mostramos que, dadas uma função $\mathbf{f}(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'))$, as derivadas $\mathbf{f}'(\mathbf{x})$ e $\mathbf{f}''(\mathbf{x})$ seguem processos gaussianos com médias $m'(\mathbf{x})$ e $m''(\mathbf{x})$ respectivamente, e covariâncias determinadas pelas equações derivadas das funções de covariância originais. Além disso, investigamos a distribuição conjunta entre os dados observados \mathbf{y} e as derivadas \mathbf{f}'^* , proporcionando uma estrutura matemática sólida para análises preditivas. Em situações práticas, onde é essencial entender tanto a função quanto suas derivadas, esta análise analítica das covariâncias se mostra inestimável. A formulação apresentada fornece um quadro claro e rigoroso para estudos futuros, permitindo não apenas prever a função e suas derivadas, mas também quantificar a incerteza associada a essas previsões por meio de métodos estatísticos robustos, como o método de Monte Carlo. Assim, estabelecemos um fundamento sólido para a aplicação eficaz dos processos gaussianos em uma variedade de domínios científicos e práticos [44, 36].

Nos domínios do aprendizado de máquina e estatísticas, os Processos Gaussianos (GPs) se destacam como uma ferramenta essencial para capturar relações complexas em dados. Originados na teoria das probabilidades, esses processos representam uma extensão poderosa das distribuições normais para espaços de dimensões infinitas, permitindo uma modelagem probabilística robusta de funções [35]. A característica distintiva dos GPs está na interpretação das funções como variáveis aleatórias, possibilitando não apenas previsões pontuais, mas também uma distribuição completa de possíveis resultados. Ao integrar conhecimento prévio com dados observados, os GPs oferecem uma abordagem flexível para lidar com padrões não lineares e incertezas nos dados, tornando-os ideais para diversas aplicações, desde previsões meteorológicas até a análise de séries temporais complexas [35].

Em conclusão, os Processos Gaussianos desempenham um papel central na modelagem estatística e de aprendizado de máquina, proporcionando uma maneira elegante de lidar com complexidade e incerteza nos dados. Sua capacidade de fornecer previsões probabilísticas, incorporando tanto conhecimento prévio quanto observações, os torna uma escolha preferencial em muitos cenários do mundo real. Ao oferecer uma compreensão mais profunda dos dados e de suas relações subjacentes, os Processos Gaussianos continuam a ser uma ferramenta indispensável para pesquisadores e profissionais, permitindo insights valiosos e decisões informadas em uma variedade de disciplinas. Sua flexibilidade e poder analítico solidificam sua posição como uma técnica essencial no arsenal dos cientistas de dados modernos [35].

5 O modelo cosmológico padrão

A cosmologia é a ciência que estuda a origem, estrutura, evolução e eventual destino do universo. Ela busca compreender as leis fundamentais que regem o espaço, o tempo, a matéria e a energia em escala cósmica. Ao longo da história, diversas culturas desenvolveram mitos e narrativas para explicar a existência do universo, mas foi com o avanço do método científico que a cosmologia se transformou em uma disciplina rigorosa. No século XX, a teoria da relatividade geral de Einstein e a observação da expansão do universo abririram novos horizontes para a compreensão do Cosmos.

Atualmente, a cosmologia moderna se baseia em modelos teóricos complexos, dados observacionais precisos e simulações computacionais avançadas. Um dos pilares é o modelo do Big Bang, que descreve a origem do universo a partir de um estado inicial extremamente quente e denso. Além disso, a cosmologia explora conceitos como a matéria escura, uma forma de matéria invisível que compõe a maior parte do universo, e a energia escura, uma misteriosa força responsável pela aceleração da expansão cósmica. A compreensão desses elementos desafia os limites do conhecimento humano, levando os cosmólogos a investigar questões fundamentais sobre a natureza da realidade e nosso lugar no universo.

Na cosmologia contemporânea, a compreensão do universo é moldada pela fusão de teorias complexas com um vasto volume de dados observacionais. A utilização de técnicas avançadas de análise de dados, incluindo algoritmos de machine learning, desempenha um papel crucial. Telescópios e experimentos coletam enormes quantidades de informações sobre galáxias distantes, radiação cósmica de fundo e outros fenômenos cósmicos. Por meio da análise de dados, padrões complexos e correlações são identificados, revelando insights cruciais sobre a estrutura e evolução do universo. Algoritmos de machine learning ajudam a processar esses dados em grande escala, identificar padrões sutis e até mesmo prever eventos cósmicos. Essa abordagem não apenas valida teorias existentes, como a do Big Bang, mas também abre caminho para a descoberta de novos fenômenos e conceitos, impulsionando assim nossa compreensão da natureza fundamental do universo.

5.1 Métrica

Na física, o ato de medir é central para nossa compreensão do mundo ao nosso redor. Quando consideramos a mensuração de distâncias, por exemplo, nos deparamos com a

necessidade de definir como essas medidas são realizadas no contexto do espaço em que vivemos. É aqui que entra o conceito de métrica. A métrica, que se torna especialmente relevante na teoria da relatividade de Einstein, é essencialmente uma regra que nos diz como calcular distâncias e intervalos de tempo de forma precisa. Ela não apenas simplifica nosso entendimento da realidade mensurável, mas também desempenha um papel vital na formulação matemática de teorias físicas, permitindo-nos relacionar conceitos abstratos com as observações concretas que fazemos no mundo físico [42].

O intervalo, por outro lado, é uma importante medida que descreve a separação entre dois pontos em um espaço qualquer. Esse conceito fundamental engloba tanto as dimensões espaciais quanto outras propriedades. A métrica, por sua vez, é uma ferramenta matemática crucial que define as regras para calcular esse intervalo de maneira precisa. Em outras palavras, a métrica estabelece como as distâncias e separações são medidas dentro de um dado espaço. Ao utilizar a métrica apropriada, podemos quantificar a relação entre pontos no espaço, permitindo uma análise mais profunda das características geométricas do ambiente em consideração. Assim, a métrica desempenha um papel central ao definir e compreender o conceito de intervalo em diversos contextos físicos e matemáticos [42].

Nesse contexto, o intervalo ds (dado uma métrica específica $g_{\mu\nu}$) entre dois pontos infinitesimais dx^μ e dx^ν no espaço é definido por

$$ds^2 = \sum_{\mu,\nu} g_{\mu\nu} dx^\mu dx^\nu. \quad (5.1)$$

Em particular a definição de intervalo engloba o espaço euclidiano tridimensional (onde é válida a mecânica newtoniana), cuja métrica usual é a métrica de Euclides, definida por:

$$g_{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad (5.2)$$

e para a métrica de Euclides o intervalo obtido é

$$ds^2 = (dx^1)^2 + (dx^2)^2 + (dx^3)^2, \quad (5.3)$$

que representa a distância entre dois pontos no espaço euclidiano [11].

A métrica desempenha um papel central na física, fornecendo as regras precisas para medir distâncias e intervalos de tempo no espaço. Ela simplifica nosso entendimento da realidade mensurável e é fundamental para a formulação matemática das teorias físicas. Ao calcular o intervalo entre pontos no espaço usando uma métrica específica, conseguimos conectar conceitos abstratos com observações concretas, fechando assim a lacuna entre a teoria e a experiência empírica [11].

5.2 Princípio cosmológico

O Princípio Cosmológico é um conceito fundamental na cosmologia. Ele estabelece que o universo é espacialmente homogêneo e isotrópico em larga escala, o que significa que ele tem a mesma densidade e aparência em qualquer direção que observemos. Esta característica básica simplifica as complexidades do cosmos e serve como base para diversas teorias cosmológicas, facilitando não apenas nossos modelos, mas também tendo implicações profundas [42].

A homogeneidade no universo implica que a densidade de matéria e energia é a mesma em todos os lugares em larga escala, eliminando a existência de regiões especiais com maior ou menor densidade. Por outro lado, a isotropia refere-se à uniformidade das propriedades físicas em todas as direções observadas. Em outras palavras, não importa para onde olhamos no espaço, as leis da física são consistentes [42].

Esses princípios básicos, homogeneidade e isotropia, formam a base do Princípio Cosmológico e são essenciais para nossa compreensão atual da estrutura e evolução do universo em larga escala. Ao aceitar essa premissa, podemos entender melhor a evolução do universo, desde suas fases iniciais até as estruturas cósmicas complexas que observamos hoje [42].

5.3 Lei de Hubble-Lemaître

Na década de 1920, os astrônomos Edwin Hubble e Georges Lemaître fizeram uma descoberta notável ao observar a luz das galáxias distantes. Eles perceberam um fenômeno chamado "desvio para o vermelho", onde a luz emitida pelas galáxias se estica para comprimentos de onda mais longos à medida que as galáxias se afastam da Terra. Em termos simples, quando uma galáxia se move para longe de nós, a luz que vemos dela muda para a parte vermelha do espectro eletromagnético [42]. Quanto maior o desvio para o vermelho, maior é a velocidade de afastamento da galáxia. Esse fenômeno foi fundamental para a formulação da Lei de Hubble-Lemaître, que estabelece uma relação matemática entre o desvio para o vermelho das galáxias e suas distâncias, a Lei de Hubble-Lemaître é definida matematicamente através da relação

$$z = \frac{H_0}{c}r, \quad (5.4)$$

onde z é o desvio para o vermelho, H_0 é uma constante denominada constante de Hubble, r é a distância à galáxia e c é a velocidade da luz [42].

O desvio para o vermelho z , conhecido como redshift em inglês, é um fenômeno em que a luz proveniente de objetos astronômicos distantes é esticada para comprimentos de onda mais longos devido à expansão do universo. Quanto mais afastado um objeto está,

maior é seu redshift [42]. Esse efeito é crucial na cosmologia, fornecendo dados essenciais para calcular distâncias cósmicas e entender a expansão do universo e é definido através da relação

$$z = \frac{\lambda_{obs} - \lambda_{em}}{\lambda_{em}} \quad (5.5)$$

onde λ_{obs} é o comprimento de onda observado de um objeto que emite um comprimento de onda λ_{em} . Com o uso da definição do desvio para o vermelho (mais comumente chamado de redshift, do inglês) a Lei de Hubble-Lemaître toma a forma

$$\nu = H_0 r, \quad (5.6)$$

onde

$$\nu = cz \quad (5.7)$$

é a velocidade de afastamento ou de aproximação do objeto emissor, quando o objeto se aproxima (redshift negativo) o termo comumente utilizado é o blueshift [42].

A lei de Hubble-Lemaître surge naturalmente em um universo que possui uma expansão (ou contração) homogênea e isotrópica. Suponha três galáxias g_1 , g_2 e g_3 com respectivos vetores de posição \vec{r}_1 , \vec{r}_2 e \vec{r}_3 para um dado sistema de coordenadas, como exposto na Figura 5.1 [42].

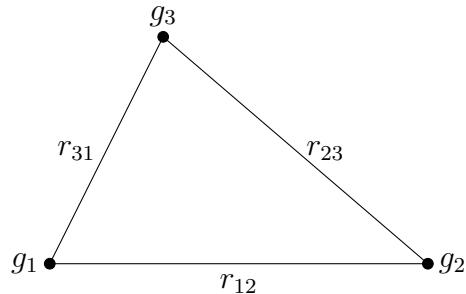


Figura 5.1: Triângulo definido por três galáxias em um universo expandindo uniformemente. Fonte: Criado pelo autor.

Na Figura 5.1 as distâncias r_{12} , r_{23} e r_{31} são definidas, respectivamente, por

$$r_{12} = |\vec{r}_1 - \vec{r}_2|, \quad (5.8)$$

$$r_{23} = |\vec{r}_2 - \vec{r}_3|, \quad (5.9)$$

e

$$r_{31} = |\vec{r}_3 - \vec{r}_1|. \quad (5.10)$$

Uma expansão homogênea e istrópica implica que a forma do triângulo é preservada durante o afastamento (ou aproximação) das galáxias, isso requer uma lei de expansão da forma

$$r_{12}(t) = a(t)r_{12}(t_0), \quad (5.11)$$

$$r_{23}(t) = a(t)r_{23}(t_0) \quad (5.12)$$

e

$$r_{31}(t) = a(t)r_{31}(t_0). \quad (5.13)$$

Na Equação 5.11, Equação 5.12 e Equação 5.13 a função $a(t)$ é o denominado fator de escala do universo, que possui um valor

$$a_0 = a(t_0) \quad (5.14)$$

no tempo presente e, como exigido pelo princípio cosmológico, independe da posição e direção espacial do objeto. O fator de escala é uma função que indica como a expansão do universo evolui no tempo [42].

Em um tempo t qualquer, um observador na galáxia g_1 enxergará as galáxias g_2 e g_3 se afastando à uma velocidade

$$v_{12}(t) = \frac{dr_{12}}{dt}, \quad (5.15)$$

todavia, com o uso da Equação 5.11

$$\frac{dr_{12}}{dt} = \frac{d(a(t)r_{12}(t_0))}{dt}, \quad (5.16)$$

o que implica em

$$\frac{dr_{12}}{dt} = \frac{da(t)}{dt}r_{12}(t_0), \quad (5.17)$$

portanto

$$\nu_{12}(t) = H(t)r_{12}, \quad (5.18)$$

onde

$$H(t) = \frac{\dot{a}(t)}{a(t)} \quad (5.19)$$

é o fator de Hubble e a Equação 5.18 é justamente a Equação 5.6 para um tempo t arbitrário [42].

Em resumo, o desvio para o vermelho em galáxias distantes é causado pela expansão do universo. Quanto mais distante uma galáxia está, maior é seu desvio para o vermelho, indicando que está se afastando de nós a uma velocidade maior. A constante de Hubble quantifica essa taxa de expansão. Essa relação entre desvio para o vermelho e distância é fundamental para medir distâncias cósmicas e compreender a dinâmica da expansão do universo, proporcionando informações valiosas sobre sua evolução [42].

5.4 A métrica de Minkowski

Na Seção 5.1 foi abordado o conceito de métrica, no entanto o exemplo abordado foi a métrica euclidiana que é válida para um espaço tridimensional newtoniano. Todavia, a visão newtoniana da mecânica se mostrou incoerente para a explicação de fenômenos físicos em alguns contextos, sendo substituída por uma teoria mais fundamental: a teoria da relatividade restrita de Einstein [42]. Na teoria da relatividade restrita o espaço e o tempo são componentes de uma mesma entidade quadri-dimensional denominada espaço-tempo e a métrica que define o intervalo espaço-temporal ds entre dois eventos (t, r, θ, ϕ) e $(t + dt, r + dr, \theta + d\theta, \phi + d\phi)$ é a denominada métrica de Minkowski, definida por

$$g_{\mu\nu} = \begin{pmatrix} -c^2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & r^2 & 0 \\ 0 & 0 & 0 & r^2 \sin^2 \theta \end{pmatrix}. \quad (5.20)$$

O intervalo obtido a partir da métrica de Minkowski é

$$ds^2 = -c^2 dt^2 + dr^2 + r^2(d\theta^2 + \sin^2 \theta d\phi^2). \quad (5.21)$$

A teoria da relatividade restrita fundamenta-se em dois postulados essenciais. O primeiro postulado afirma que as leis da física são idênticas em todos os referenciais inerciais, ou seja, em sistemas de coordenadas onde um observador está em repouso ou se move com velocidade constante em relação a outros objetos ao seu redor. O segundo postulado estabelece que a velocidade de propagação da luz no vácuo é uma constante universal, denotada por c , e é a mesma em todos os referenciais inerciais [37].

Com a aceitação dos postulados da teoria da relatividade restrita, surgiram implicações cruciais para a física moderna. Fenômenos como a dilatação do tempo, a contração do comprimento e a equivalência entre massa e energia redefiniram nossa compreensão do espaço e do tempo. Além disso, a teoria da relatividade restrita é fundamental para tecnologias como o GPS, demonstrando sua aplicação prática. Essas descobertas transformaram não apenas a teoria física, mas também nosso modo de ver o mundo, moldando a pesquisa científica e a tecnologia contemporânea. [37].

Embora a métrica de Minkowski seja extremamente útil para diversos fenômenos de interesse (física de partículas, física nuclear, dentre outros) ela possui um ”problema” fundamental: ela é útil apenas em contextos onde é válida a teoria da relatividade restrita de Einstein e a relatividade restrita lida apenas em contextos nos quais ocorre a ausência de gravidade, nessa problemática surge a necessidade de uma teoria física que trabalhe com fenômenos mais gerais, onde há a presença da atração gravitacional, essa teoria é a denominada Teoria da Relatividade Geral [42].

Na teoria da relatividade geral a gravidade está intimamente ligada à denominada curvatura espacial, sendo uma consequência da curvatura do espaço-tempo. portanto antes de iniciar as discussões sobre a teoria da relatividade geral é necessário discutir o conceito de curvatura [42].

5.5 Curvatura

Como abordado na Seção 5.4 para conceituar a gravidade no contexto einsteiniano, é necessário discutir o conceito de curvatura espacial. O conceito de curvatura espacial foi um dos grandes desafios encontrados por Einstein para a formalização de sua teoria de gravitação, no entanto o conceito de curvatura é fundamental para a formulação de uma equação de campo que permita conectar a curvatura com a densidade de massa-energia: na relatividade geral, ou restrita, o conceito de massa está conectado ao conceito de energia através da equação

$$E^2 = m^2 c^4 + p^2 c^2, \quad (5.22)$$

onde E define a energia de um objeto de massa de repouso m e momento p [37].

Para compreender o conceito de curvatura, é essencial introduzir a ideia de geodésica: uma geodésica é uma trajetória que representa o caminho mais curto ou eficiente entre dois pontos em uma superfície ou variedade, como uma superfície esférica, uma variedade curva ou até mesmo o espaço-tempo na teoria da relatividade geral [14].

Formalmente, uma geodésica pode ser definida como uma curva parametrizada $\gamma(t)$ em uma variedade M que satisfaz a equação geodésica:

$$\nabla_{\dot{\gamma}(t)} \dot{\gamma}(t) = 0 \quad (5.23)$$

Aqui, ∇ representa o operador nabla ou operador de conexão compatível com a métrica na variedade M , $\dot{\gamma}(t)$ é o vetor tangente à curva $\gamma(t)$ e $\nabla_{\dot{\gamma}(t)}$ denota a derivada covariante ao longo do vetor tangente. Em outras palavras, uma geodésica é uma curva cuja aceleração é zero quando medida com a conexão compatível com a métrica na variedade [14].

Com o conceito de geodésica inserido, vamos estudar o caso mais simples de espaço: um espaço euclidiano bidimensional. Para o espaço euclidiano bidimensional, se construído

um triângulo ligando três pontos com geodésicas (no caso do espaço euclidiano a geodésica é uma reta) os ângulos entre os vértices do triângulo obedecem a relação

$$\alpha + \beta + \gamma = \pi, \quad (5.24)$$

onde os ângulos são medidos em radianos [42]. Nesse espaço, o Teorema de Pitágoras é válido, portanto a distância $d\ell$ entre os pontos (x, y) e $(x+dx, y+dy)$ obedece a relação

$$d\ell^2 = dx^2 + dy^2. \quad (5.25)$$

Se escolhidas coordenadas polares, a distância $d\ell$ entre os pontos (r, θ) e $(r+dr, \theta+d\theta)$ obedecerá a relação

$$d\ell^2 = dr^2 + r^2 d\theta^2. \quad (5.26)$$

Se agora considerarmos um espaço bidimensional representado pela superfície de uma esfera e se conectarmos três pontos através de geodésicas (construirmos um triângulo), os ângulos dos vértices do triângulo obedecerá a relação

$$\alpha + \beta + \gamma = \pi + \frac{A}{R^2}, \quad (5.27)$$

onde A é a área do triângulo e R é o raio da esfera [42]. Todos os espaços onde

$$\alpha + \beta + \gamma > \pi \quad (5.28)$$

são denominados espaços com curvatura positiva. Na superfície de uma esfera um sistema de coordenadas pode ser definido se escolhidos dois pontos opostos, cuja geodésica que os liga é o meridiano principal [42]. Se r for a distância ao polo norte e θ for o ângulo azimutal (medido com relação ao meridiano principal) então a distância $d\ell$ entre os pontos r, θ e $r+dr, \theta+d\theta$ obedecerá a relação

$$d\ell^2 = dr^2 + R^2 \sin^2\left(\frac{r}{R}\right) d\theta^2. \quad (5.29)$$

Adicionalmente ao espaço plano e ao espaço com curvatura positiva, existe o espaço com curvatura negativa. Um exemplo de espaço de curvatura é um hiperbolóide, contudo o matemático Hilbert provou que diferentemente dos espaços de curvatura nula ou de curvatura positiva, é impossível descrever um espaço bidimensional de curvatura negativa constante em um espaço euclidiano tridimensional [42]. Portanto não é possível a visualização de uma figura que represente o espaço de curvatura negativa constante, contudo de maneira análoga ao espaço de curvatura positiva e ao espaço plano as características de um espaço de curvatura negativa constante podem ser descritas matematicamente [42].

Se construído um triângulo através da conexão de três pontos através de geodésicas,

a soma dos ângulos dos vértices desse triângulo obedecerá a relação

$$\alpha + \beta + \gamma = \pi - \frac{A}{R^2}, \quad (5.30)$$

onde R é o raio de curvatura da superfície e A é a área do triângulo.

Para o espaço de curvatura negativa constante, é possível definir um sistema de coordenadas em sua superfície, nesse sistema de coordenadas deve se escolher um ponto como o polo e a partir do polo definir uma geodésica que percorre a superfície definindo um meridiano principal [42]. Se r for a distância até o polo e θ for o ângulo medido com relação ao meridiano principal, então a distância $d\ell$ entre um ponto (r, θ) e $(r + dr, \theta + d\theta)$ obedecerá a relação

$$d\ell^2 = dr^2 + R^2 \sinh^2 \left(\frac{r}{R} \right) d\theta^2. \quad (5.31)$$

Os resultados obtidos para espaços bidimensionais podem ser extendidos diretamente para um espaço tridimensional, onde o intervalo (para as três possíveis curvaturas) entre dois eventos (r, θ, ϕ) e $(r + dr, \theta + d\theta, \phi + d\phi)$ é obtido através da relação

$$d\ell^2 = dr^2 + S_k(r)^2(d\theta^2 + \sin^2 \theta d\phi^2), \quad (5.32)$$

nesse caso

$$S_k(r) = \begin{cases} R \sin \left(\frac{r}{R} \right), & \text{se } k = 1 \\ r, & \text{se } k = 0 \\ R \sinh \left(\frac{r}{R} \right), & \text{se } k = -1 \end{cases}. \quad (5.33)$$

Em física e geometria, o conceito de curvatura espacial é essencial para compreendermos a forma do espaço ao nosso redor. Em um espaço euclidiano bidimensional, as geodésicas são linhas retas, enquanto em uma superfície esférica, as geodésicas são arcos de círculo, revelando uma curvatura positiva constante [42]. Por outro lado, em espaços com curvatura negativa constante, como a superfície de um hiperboloide, as geodésicas se estendem infinitamente. A curvatura define como linhas retas se comportam nesses espaços e influencia a geometria local e global, sendo um aspecto fundamental na compreensão da estrutura do universo e dos objetos que o habitam [42].

5.6 A métrica de Friedmann-Lemaître-Robertson-Walker

Desde o início deste capítulo, temos explorado conceitos essenciais para o estudo da cosmologia. Em particular, focamos na métrica e suas propriedades cruciais. Além disso, examinamos o princípio cosmológico, fundamental para os modelos cosmológicos modernos, e investigamos a Lei observacional de Hubble, demonstrada por Lemaître, que confirma a

expansão contínua do universo. Nesse contexto, abordamos o conceito de espaço-tempo e curvatura, levantando uma pergunta central: qual métrica é adequada para modelar um universo com expansão espacial homogênea e isotrópica em todos os pontos de sua evolução? Esta seção se dedica a explorar detalhadamente essa questão [42].

Em 1930, os físicos Howard Robertson e Arthur Walker derivaram independentemente essa métrica. A relação obtida foi expressa como:

$$g_{\mu\nu} = \begin{pmatrix} -c^2 & 0 & 0 & 0 \\ 0 & \frac{a(t)^2}{1-kr^2} & 0 & 0 \\ 0 & 0 & a(t)^2 S_k(r)^2 & 0 \\ 0 & 0 & 0 & a(t)^2 S_k(r)^2 \sin^2 \theta \end{pmatrix}, \quad (5.34)$$

onde $S_k(r)$ é a função definida pela Equação 5.33, com $R = R_0$ [42]. Na métrica de Friedmann-Lemaître-Robertson-Walker (FLRW), a componente espacial representa a métrica de um espaço com curvatura uniforme, com raio de curvatura R_0 . O fator de escala $a(t)$ descreve a expansão espacial homogênea e isotrópica ao longo do tempo, enquanto t representa o tempo próprio cosmológico. Este é o tempo medido por um observador que percebe a expansão uniforme e homogênea do universo ao seu redor. As coordenadas espaciais (r, θ, ϕ) são conhecidas como coordenadas comóveis, e em um universo que se expande de forma homogênea e isotrópica, essas coordenadas permanecem constantes ao longo do tempo [42].

Nesta seção, exploramos os fundamentos essenciais da cosmologia, incluindo a métrica de Friedmann-Lemaître-Robertson-Walker (FLRW) derivada por Robertson e Walker em 1930. Esta métrica descreve um universo com expansão espacial homogênea e isotrópica. O fator de escala $a(t)$ representa a expansão ao longo do tempo, enquanto as coordenadas comóveis (r, θ, ϕ) permanecem constantes durante a expansão uniforme do universo. Esta formulação matemática é crucial para nossa compreensão da estrutura do universo em larga escala e tem sido fundamental para investigações cosmológicas profundas [42].

5.6.1 Distância própria e o redshift cosmológico

Conforme destacado na Seção 5.1, a medição desempenha um papel fundamental na física, especialmente na cosmologia. Na cosmologia, é crucial obter medidas precisas das distâncias entre objetos cósmicos. Essas medições fornecem informações essenciais sobre a evolução do universo, a formação de estruturas e outros fenômenos relacionados. Nesse contexto, esta seção explora dois aspectos intimamente interligados na cosmologia: o conceito de distância própria e o redshift cosmológico [42].

Considere uma galáxia suficientemente distante de nós de tal modo que podemos ignorar as perturbações locais e considerar um universo descrito pela métrica FRLW. No contexto apresentado, é interessante discutir quão distante a galáxia realmente está de

nós. Em um universo em expansão, a distância entre os objetos está aumentando com o passar do tempo. Portanto, se quisermos definir a distância espacial entre dois objetos, é necessário também definir o tempo no qual estamos calculando essa distância [42]. Suponha um observador na origem de um sistema de coordenadas observando uma galáxia em uma coordenada comóvel (r, θ, ϕ) . A distância própria $d_p(t)$ entre os objetos é descrita pelo comprimento da geodésica que liga os pontos quando o fator de escala é fixado em um determinado tempo t [42].

A distância espacial entre os objetos, obtida a partir da métrica FRLW, e descrita pela relação

$$ds^2 = a(t)^2(dr^2 + S_k(r)^2d\theta^2 + S_k(r)^2\sin^2\theta d\phi^2), \quad (5.35)$$

no entanto, ao longo da geodésica espacial os angulos θ e ϕ são constantes, portanto o intervalo é representado por

$$ds = a(t)dr. \quad (5.36)$$

A distância própria é a distância obtida ao se integrar o esse intervalo sobre a coordenada radial comóvel

$$d_p(t) = \int_0^r a(t)dr \quad (5.37)$$

o que implica em

$$d_p(t) = a(t)r. \quad (5.38)$$

Dessa maneira, a taxa de variação da distância própria com relação ao tempo é calculada a partir de

$$\dot{d}_p(t) = \dot{a}(t)r, \quad (5.39)$$

essa relação retorna novamente a Lei de Hubble-Lemaître

$$v_p(t) = H(t)d_p(t). \quad (5.40)$$

É importante notar que a Lei de Hubble obtida através da Equação 5.18 tem um significado ligeiramente diferente da Equação 5.40, neste caso a variação da distância entre dois objetos separados está associada diretamente com a expansão do espaço-tempo: da mesma maneira que os objetos se afastam, o raio de curvatura R do universo aumenta na mesma taxa $R(t) = a(t)R_0$ [42].

Embora a distância própria de um objeto tenha relevância física significativa, ao observarmos um objeto no céu, não podemos determinar sua distância diretamente. Em vez

disso, medimos o desvio para o vermelho na luz emitida pelo objeto. Apesar de o desvio para o vermelho não fornecer uma medida direta da distância até o objeto, ele indica o fator de escala do objeto no momento t_e em que a luz foi emitida [42].

Para entender a relação entre o desvio para o vermelho e a distância, consideremos um pulso de luz emitido por uma galáxia em um tempo t_e e observado em um tempo t_0 [42]. Durante esse intervalo, a luz percorre uma geodésica nula de acordo com a relatividade restrita, onde $ds = 0$. Dado que a luz percorre uma geodésica espacial as direções θ e ϕ são constantes, resultado na relação

$$c \frac{dt}{a(t)} = dr. \quad (5.41)$$

Suponhamos que a galáxia emite um pulso luminoso com comprimento de onda λ_e . A distância percorrida pelo pulso da galáxia emissora até nós é dada por:

$$c \int_{t_e}^{t_0} \frac{dt}{a(t)} = \int_0^r dr. \quad (5.42)$$

Analogamente, para um pulso posterior emitido em um tempo $t_e + \frac{\lambda_e}{c}$, obtemos:

$$c \int_{t_e + \frac{\lambda_e}{c}}^{t_0 + \frac{\lambda_0}{c}} \frac{dt}{a(t)} = \int_0^r dr. \quad (5.43)$$

Comparando as equações acima, chegamos a:

$$c \int_{t_e}^{t_0} \frac{dt}{a(t)} = \int_{t_e + \frac{\lambda_e}{c}}^{t_0 + \frac{\lambda_0}{c}} \frac{dt}{a(t)}. \quad (5.44)$$

Dado que a escala de tempo para a expansão do universo é aproximadamente 32 vezes maior em ordem de grandeza do que o tempo entre dois feixes luminosos consecutivos, podemos aproximar o fator de escala constante na equação acima [42]. Assim, obtemos:

$$\frac{\lambda_e}{a(t_e)} = \frac{\lambda_0}{a(t_0)}. \quad (5.45)$$

Usando a definição de desvio para o vermelho a partir da Equação 5.5, chegamos a

$$1 + z = \frac{a_0}{a(t)}, \quad (5.46)$$

onde a_0 refere-se ao fator de escala observado no presente e $a(t)$ refere-se ao fator de escala no momento da emissão do feixe luminoso [42].

Em resumo, compreender o universo envolve desafios significativos, com a medição precisa das distâncias cósmicas desempenhando um papel crucial. Ao explorar conceitos como distância própria e desvio para o vermelho, ganhamos insights profundos sobre a expansão do espaço-tempo [42]. O desvio para o vermelho, indicando o fator de escala do

universo em diferentes momentos, é essencial para nossa compreensão da evolução cósmica [42].

5.7 A Teoria da Relatividade Geral de Einstein

Nas seções anteriores, foram apresentados os conceitos físicos e matemáticos essenciais que constituem a base para a compreensão da teoria da relatividade geral de Albert Einstein. Esta teoria, proposta em 1915, representa um marco fundamental na física moderna. Ao descrever a gravidade como uma curvatura do espaço-tempo, a relatividade geral desafia nossa intuição cotidiana sobre o universo. Neste capítulo, aprofundaremos nosso conhecimento sobre os princípios centrais desta teoria, explorando suas implicações e aplicações em nosso entendimento atual do Cosmos [42].

Um dos fundamentos essenciais da teoria da relatividade geral é o princípio da equivalência, que estabelece que em qualquer ponto do espaço-tempo, é impossível discernir localmente entre um campo gravitacional uniforme e uma aceleração uniforme. Em termos simples, um observador em queda livre devido à gravidade não consegue, através de experimentos realizados apenas dentro de seu sistema de referência local, distinguir essa situação de um estado sem gravidade, mas com uma aceleração constante na direção oposta. Esse princípio implica que, em pequenas regiões do espaço-tempo, as leis da física são idênticas para um observador em queda livre pela gravidade e para um observador acelerado na direção oposta em um ambiente sem gravidade [42].

Com o princípio da equivalência em mente, é oportuno discutir um experimento mental. Imagine uma pessoa em um foguete que está acelerando para cima a uma taxa constante. Agora, imagine que essa pessoa está segurando uma lanterna. Se ela ligar a lanterna, um observador de fora perceberá que a luz está curvada para baixo, como expresso na Figura 5.2 [42].

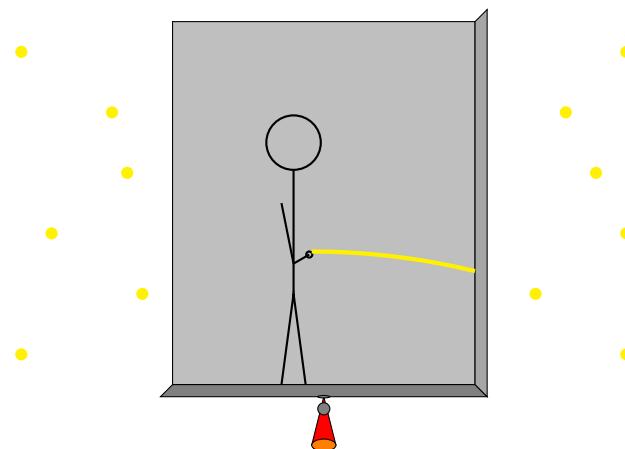


Figura 5.2: Ilustração de uma pessoa segurando uma lanterna dentro de um foguete acelerado no espaço, vista de um observador externo. Fonte: Criado pelo autor.

No entanto, conforme discutido pelo princípio da equivalência, essa situação é indistinguível para uma pessoa dentro do foguete em comparação com uma pessoa segurando uma lanterna em uma cabine sob a influência de um campo gravitacional uniforme. Essa condição pode ser visualizada imaginando-se uma pessoa em uma caixa na superfície de um planeta, como ilustrado na Figura 5.3 [42].

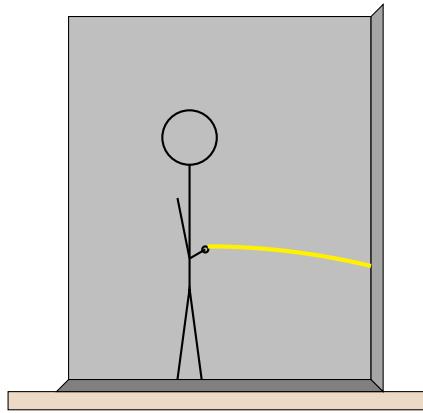


Figura 5.3: Ilustração de uma pessoa segurando uma lanterna dentro de uma cabine sob influência de um campo gravitacional uniforme, vista de um observador externo. Fonte: Criado pelo autor.

Dado que o experimento ilustrado na Figura 5.2 e na Figura 5.3 são indistinguíveis de acordo com o princípio da equivalência, somos levados à conclusão de que a presença de um campo gravitacional curva o feixe luminoso, mesmo que a luz não tenha massa [42].

Além disso, um dos princípios fundamentais da ótica é o chamado Princípio de Fermat, que estipula que a luz percorre o caminho que maximiza ou, na maioria dos casos, minimiza o tempo de percurso. Dessa maneira, no experimento realizado, o feixe luminoso deve seguir o caminho mais curto entre a lanterna e a parede da caixa e do foguete. Esse caminho mais curto é, em um espaço euclidiano, uma linha reta. No entanto, como o percurso da luz na Figura 5.3 não é uma linha reta, Einstein concluiu que o espaço não é euclidiano, mas sim curvo [42].

Na teoria de Einstein, a presença de massa-energia modifica as características do espaço-tempo, curvando-o, e os objetos seguem trajetórias geodésicas nesse espaço-tempo curvo [42]. Isso pode ser expresso pela equação de campo de Einstein:

$$G_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu} - \Lambda g_{\mu\nu}, \quad (5.47)$$

Na presente equação, G denota a constante da gravitação universal de Newton, $T_{\mu\nu}$ representa o Tensor de energia-momento, que reflete a energia e o momento de um objeto, $G_{\mu\nu}$ é o Tensor de Einstein, que descreve a curvatura do espaço-tempo, Λ é uma constante conhecida como Constante cosmológica e $g_{\mu\nu}$ é a métrica associada ao espaço-tempo [42].

5.7.1 O tensor de Einstein

O Tensor de Einstein desempenha um papel central nas equações de campo de Einstein e está intrinsecamente ligado à distribuição de massa e energia no espaço-tempo, encapsulando informações cruciais sobre a curvatura do espaço-tempo por meio de uma métrica associada, representada como $g_{\mu\nu}$, sendo representado por

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu}. \quad (5.48)$$

Nessa equação $R_{\mu\nu}$ é o denominado Tensor de Ricci, e R é o denominado Escalar de Ricci, que é simplesmente a contração $R = g^{\mu\nu}R_{\mu\nu}$ do Tensor de Ricci com a inversa da métrica, representada por $g^{\mu\nu}$, onde é levada em consideração a notação de Einstein¹ [2].

O tensor de Ricci é calculado a partir dos símbolos de Christoffel, que por sua vez estão relacionados à métrica do espaço-tempo [2]. Os símbolos de Christoffel representam a conexão entre as derivadas covariantes de um campo tensorial, levando em consideração a curvatura do espaço-tempo. Eles são dados por

$$\Gamma_{\mu\nu}^\rho = \frac{1}{2}g^{\rho\sigma}(\partial_\mu g_{\nu\sigma} + \partial_\nu g_{\mu\sigma} - \partial_\sigma g_{\mu\nu}), \quad (5.49)$$

onde $g^{\rho\sigma}$ é a inversa da métrica $g_{\mu\nu}$, e ∂_μ representa a derivada parcial em relação à coordenada μ [2]. Com os símbolos de Christoffel em mãos, podemos calcular o tensor de Ricci usando a equação

$$R_{\mu\nu} = \partial_\rho \Gamma_{\nu\mu}^\rho - \partial_\nu \Gamma_{\rho\mu}^\rho + \Gamma_{\rho\sigma}^\rho \Gamma_{\nu\mu}^\sigma - \Gamma_{\nu\sigma}^\rho \Gamma_{\rho\mu}^\sigma. \quad (5.50)$$

5.7.2 O Tensor de energia-momento

O tensor de energia-momento é uma estrutura matemática crucial na física teórica que descreve a distribuição de energia, momento e pressão em um sistema físico. Trata-se de um tensor de segunda ordem, cujas componentes representam a densidade de energia, o fluxo de momento e a pressão do sistema [1].

Uma das equações fundamentais que envolve o tensor de energia-momento é a equação de conservação, expressa pela relação matemática:

$$\partial_\mu T^{\mu\nu} = 0. \quad (5.51)$$

Essa equação de conservação reflete o princípio fundamental da conservação da energia e do momento, que estabelece que a variação local na quantidade de energia e momento em um sistema é igual à taxa com que essa energia e momento são transferidos através das fronteiras do sistema [1].

¹Na notação de Einstein, a repetição de dois índices em um tensor implica uma soma sobre esses índices, dispensando a necessidade de usar o símbolo de somatório.

O tensor de energia-momento é uma ferramenta essencial em diversas áreas da física. Sua presença em equações-chave da física teórica destaca sua importância fundamental na descrição matemática da dinâmica dos sistemas físicos [1].

5.7.3 A Constante Cosmológica

Um dos maiores mistérios da cosmologia atual reside na constante cosmológica introduzida por em sua equação de campo (Equação 5.47). Inicialmente proposta como um termo para manter o universo em um estado estático, a constante cosmológica foi posteriormente abandonada por Einstein quando as observações indicaram a expansão do universo. No entanto, ela ressurgiu como um enigma fundamental quando a expansão cósmica acelerada foi descoberta. A constante cosmológica, agora associada à energia escura, uma forma de energia que permeia o espaço vazio e é responsável pela aceleração da expansão do universo, desafia nossas concepções sobre o universo, representando um dos desafios mais intrigantes na física teórica contemporânea [42].

Em um universo onde apenas a gravidade é considerada, todos os objetos seriam inevitavelmente atraídos uns pelos outros, resultando em um colapso cósmico inevitável. Para contornar essa conclusão, Albert Einstein introduziu a constante cosmológica (Λ) em suas equações de campo, criando um termo que se opõe à atração gravitacional. Esse termo foi inicialmente proposto para manter um universo estático e eternamente equilibrado, onde a atração gravitacional dos objetos seria contrabalançada pela repulsão da constante cosmológica. Embora a concepção de um universo estático tenha sido abandonada com a descoberta da expansão cósmica, a constante cosmológica permaneceu como um componente nas equações de Einstein, agora associada à energia escura, que é responsável por impulsionar a aceleração da expansão do universo [42].

A constante cosmológica apresenta um desafio técnico significativo na física contemporânea. Sua associação a um fluido com pressão negativa contradiz as propriedades conhecidas da matéria e da energia. Além disso, não temos uma compreensão clara de suas propriedades físicas ou sua origem, o que cria uma lacuna problemática em nosso conhecimento. A dificuldade técnica reside na incapacidade de incorporar a constante cosmológica de maneira consistente em nossos modelos teóricos existentes, evidenciando assim a necessidade de avanços teóricos e experimentais para elucidar essa questão e garantir a coesão nos modelos apresentados [42].

5.8 As equações de Friedmann

A Relatividade Geral é uma teoria fundamental que serve como alicerce para a compreensão de uma variedade de sistemas físicos. Agora, munidos dessa teoria, podemos utilizá-la para derivar as leis que regem o universo. Dentro do contexto de um universo

em expansão homogênea e isotrópica, governado pela Relatividade Geral de Einstein, essas leis que modelam a dinâmica cósmica são conhecidas como Equações de Friedmann. Elas representam as primeiras soluções analíticas das equações de Einstein para esse cenário específico, descrevendo o comportamento do universo em larga escala.

Para iniciar o estudo do universo, é essencial estabelecer uma definição precisa do objeto que estamos investigando. Nesse contexto, uma abordagem eficaz é considerar o universo como um fluido perfeito composto por galáxias, que são tratadas como as "partículas constituintes" do universo. Essa abordagem é justificável porque nosso objetivo é trabalhar matematicamente com um objeto que represente o universo como um todo, tornando as propriedades detalhadas de componentes individuais, como átomos e partículas subatômicas, menos relevantes. É crucial observar as características interessantes desse fluido perfeito, que o tornam compatível com o princípio cosmológico e também um modelo adequado para o estudo da cosmologia.

Um fluido perfeito é um conceito teórico na física que descreve um fluido idealizado, onde a viscosidade e a dissipação de energia são negligenciadas [1]. Nesse contexto ideal, o fluido é considerado incompressível, o que significa que sua densidade permanece constante independentemente da pressão aplicada [1]. Além disso, um fluido perfeito é não viscoso, o que implica que não há resistência ao movimento interno das camadas do fluido, e é irrotacional, o que significa que não forma redemoinhos ou turbulências em seu movimento [1]. No modelo de fluido perfeito, o tensor energia-momento pode ser expresso como:

$$T^{\mu\nu} = \left(\rho + \frac{p}{c^2} \right) u^\mu u^\nu + p g^{\mu\nu}, \quad (5.52)$$

onde ρ representa a densidade de massa do fluido, p é a pressão, u^μ representa a quadrivelocidade do fluido em um dado referencial. É comum adotar o referencial de um observador em repouso em relação à radiação cósmica de fundo [2].

A equação de Friedmann é derivada ao considerar um fluido perfeito, conforme definido pela Equação 5.52, e ao utilizar a métrica FLRW, conforme descrita na Equação 5.34, para representar um universo com expansão homogênea e isotrópica na equação de campo de Einstein, indicada pela Equação 5.47 [2]. Esta solução foi pioneiramente obtida por Friedmann em 1922 e ficou conhecida como Equação de Friedmann, representada por

$$H^2(t) = \frac{8\pi G}{3c^2} \rho(t) - \frac{kc^2}{R_0^2 a^2(t)} + \frac{\Lambda}{3}. \quad (5.53)$$

Além da Equação de Friedmann, há outra equação fundamental na cosmologia que é derivada de princípios termodinâmicos [42]. A primeira lei da termodinâmica estabelece que a energia total de um sistema termodinâmico é conservada e pode ser expressa matematicamente como:

$$\delta Q = dE + dW. \quad (5.54)$$

Nesta expressão, δQ representa a quantidade de calor fornecida ao sistema, onde E denota a energia interna do sistema e W é o trabalho realizado pelo sistema. Se o processo ocorre a uma pressão constante, a equação é modificada para

$$\delta Q = dE + pdV, \quad (5.55)$$

onde V representa o volume ocupado pelo sistema e p representa sua pressão [42].

Ao considerarmos a expansão adiabática do universo, durante um intervalo de tempo infinitesimal dt , podemos estabelecer $\delta Q = 0$, resultando na equação

$$\dot{E} + p\dot{V} = 0. \quad (5.56)$$

Considere uma esfera com raio comóvel r_e expandindo-se junto com a expansão do universo. O raio próprio da esfera nesse contexto é dado por $R_e(t) = r_e a(t)$ [42]. Consequentemente, o volume próprio ocupado pela esfera é dado pela equação

$$V(t) = \frac{4\pi}{3} r_e^3 a^3(t). \quad (5.57)$$

Se a esfera possui uma densidade de energia $\rho(t)$, sua energia interna é representada por

$$E(t) = V(t)\rho(t). \quad (5.58)$$

Portanto, para essa esfera, a Equação 5.56 assume a forma

$$\dot{\rho} + 3\frac{\dot{a}}{a}(\rho + p) = 0, \quad (5.59)$$

que é conhecida como a Equação do Fluido [42].

Se multiplicarmos a Equação de Friedmann por $a^2(t)$ e derivarmos em relação ao tempo, chegamos à seguinte equação:

$$\frac{\ddot{a}}{a} = \frac{\Lambda}{3} + \frac{4\pi G}{3c^2} \left(2p + \dot{p}\frac{a}{\dot{a}} \right), \quad (5.60)$$

Ao substituirmos a Equação 5.59 na Equação 5.60, obtemos a chamada Equação da Aceleração, que é expressa como:

$$\frac{\ddot{a}}{a} = \frac{\Lambda}{3} - \frac{4\pi G}{3c^2} (\rho + 3p). \quad (5.61)$$

As Equações 5.61, 5.59 e 5.53 são conhecidas como as Equações de Friedmann [42]. Elas descrevem a dinâmica de um universo homogêneo e isotrópico, conforme previsto pela teoria da relatividade geral. No entanto, entre essas equações, a Equação da Aceleração é uma expressão linearmente dependente da Equação do Fluido e da Equação de Friedmann. Portanto, temos um sistema de duas equações independentes e três variáveis ($\rho(t)$, $a(t)$

e $p(t)$) [42]. Para entender completamente a dinâmica do universo, dado um conjunto específico de condições iniciais, precisamos introduzir uma equação adicional conhecida como equação de estado. Esta equação de estado, denotada por $p = p(\rho)$, descreve como a pressão de uma substância está relacionada com sua densidade. Para muitos fluidos importantes na cosmologia, a equação de estado assume a forma:

$$p(\rho) = w\rho, \quad (5.62)$$

onde w é uma constante adimensional.

Hoje sabemos que o universo é composto de diversos componentes, dentre eles a radiação, a matéria. Além disso o modelo cosmológico introduz uma nova componente, a então denominada energia escura, que é responsável pela expansão acelerada do universo e esta relacionada à constante cosmológica [42]. Nesse caso adotamos, por simplicidade, que a densidade de energia total do universo é representada pela soma das contribuições de cada uma de suas componentes

$$\rho = \sum \rho_i. \quad (5.63)$$

Analogamente, a pressão total é tomada como a soma das pressões individuais das componentes do universo

$$P = \sum w_i \rho_i, \quad (5.64)$$

portanto, ao se assumir a aditividade da contribuição de cada componente do universo para a pressão e a densidade total, e se assumir uma equação de estado como a Equação 5.62 temos uma solução da equação do fluido para cada componente i do universo

$$\rho_i(a) = \rho_{i0} a^{-3(1+w_i)}. \quad (5.65)$$

Na física estatística, é possível determinar os valores de w_i para as principais componentes do universo, o que nos permite conhecer a equação de estado das componentes relevantes [42]. Para a radiação, $w = \frac{1}{3}$, resultando na equação de estado

$$\rho_r(a) = \rho_{r0} a^{-4}, \quad (5.66)$$

e para a matéria, $w = 0$, levando à equação de estado

$$\rho_m(a) = \rho_{m0} a^{-3}. \quad (5.67)$$

Outra grandeza de interesse cosmológico é a denominada densidade crítica de energia do universo, que é definida como a densidade de um universo espacialmente plano [42].

Nesse caso, a Equação de Friedmann toma a forma

$$H^2(t) = \frac{8\pi G}{3c^2} \left(\rho(t) + \frac{\Lambda c^2}{8\pi G} \right). \quad (5.68)$$

Se definirmos um parâmetro de densidade para a constante cosmológica da forma

$$\rho_\Lambda = \frac{\Lambda c^2}{8\pi G} \quad (5.69)$$

a Equação 5.68 toma a forma

$$H^2(t) = \frac{8\pi G}{3c^2} \rho(t), \quad (5.70)$$

onde o parâmetro de densidade leva em consideração a densidade de energia escura.

Podemos então definir, para um dado parâmetro de Hubble, o parâmetro de densidade crítica de energia do universo

$$\rho_c(t) = \frac{3c^2}{8\pi G} H^2(t), \quad (5.71)$$

se a densidade de energia do universo for maior do que a densidade de energia crítica o universo possui curvatura positiva, se for menor do que a densidade crítica, o universo possui curvatura negativa [42].

Uma vez que o parâmetro de densidade de energia crítica foi estabelecido, podemos manipular a Equação de Friedmann para obter a seguinte expressão:

$$1 = \frac{\rho(t)}{\rho_c(t)} - \frac{kc^2}{R_0^2 a^2(t) H^2(t)}, \quad (5.72)$$

A partir desta equação, definimos dois parâmetros: o parâmetro de densidade, dado por

$$\Omega(t) = \frac{\rho(t)}{\rho_c(t)}, \quad (5.73)$$

e o parâmetro de densidade de curvatura, dado por

$$\Omega_k(t) = -\frac{kc^2}{R_0^2 H^2(t) a^2(t)} \quad (5.74)$$

ou

$$\Omega_k(t) = \frac{\Omega_{k0} H_0^2}{H^2(t) a^2(t)}, \quad (5.75)$$

onde

$$\Omega_{k0} = -\frac{kc^2}{R_0^2 H_0^2}. \quad (5.76)$$

Assim, há uma relação entre os parâmetros de densidade para cada componente do uni-

verso:

$$1 = \Omega_k(t) + \Omega_m(t) + \Omega_\Lambda + \Omega_r(t). \quad (5.77)$$

A Equação de Friedmann pode então ser reescrita em termos desses parâmetros de densidade, assumindo a forma:

$$H(z) = H_0 \sqrt{\Omega_{r0}(1+z)^4 + \Omega_{m0}(1+z)^3 + \Omega_{k0}(1+z)^2 + \Omega_\Lambda}. \quad (5.78)$$

Com essas equações, estabelecemos as bases teóricas essenciais para a compreensão da evolução do universo. As Equações de Friedmann, derivadas a partir dos princípios da Relatividade Geral de Einstein, descrevem de forma precisa e matemática a dinâmica do nosso cosmos em expansão. A introdução do parâmetro de densidade crítica e dos parâmetros de densidade para diferentes componentes do universo oferece uma estrutura sistemática para analisar a geometria e o conteúdo energético do espaço-tempo [42].

5.9 Cosmologia observacional: módulo de distância e distância de luminosidade

A cosmologia observacional é um campo científico dedicado à observação sistemática e à interpretação de dados astronômicos para entender a estrutura, origem, evolução e eventual destino do universo. Utilizando observações precisas de objetos celestes distantes, como galáxias, quasares e a radiação cósmica de fundo em micro-ondas, os cosmólogos analisam padrões e tendências para formular modelos que expliquem o comportamento do universo em escalas cósmicas. Essas observações fornecem insights fundamentais sobre a expansão do universo, a distribuição de matéria escura, a formação de estruturas cósmicas e a natureza da energia escura, desempenhando um papel crucial na construção de nossa compreensão atual do Cosmos [42, 45], .

A cosmologia observacional tem suas raízes profundamente ligadas à observação do céu desde os tempos antigos. Tribos e civilizações antigas, como os babilônios e os maias, foram observadores habilidosos do cosmos, estudando os movimentos das estrelas e planetas para criar calendários precisos e prever eventos astronômicos. A cosmologia, de certa forma, sempre foi uma ciência da observação [45]. No entanto, o campo da cosmologia observacional moderna começou a ganhar forma no século XX, especialmente após as descobertas de Edwin Hubble na década de 1920, quando ele observou que as galáxias distantes estavam se afastandoumas das outras [42]. Isso proporcionou a primeira evidência observational direta da expansão do universo. Desde então, avanços tecnológicos significativos, como o desenvolvimento de telescópios espaciais e instrumentos de observação sofisticados, permitiram aos cosmologistas coletar dados cada vez mais precisos sobre a

radiação cósmica de fundo em micro-ondas, a distribuição de galáxias e a luz de objetos extremamente distantes. Estas observações têm sido essenciais para refinar e testar teorias cosmológicas, levando a uma compreensão mais profunda da evolução do universo ao longo do tempo [42].

5.9.1 Distância de Luminosidade

Na Seção 5.6.1, conseguimos definir o conceito de distância própria, que representa uma medida cosmológica de distância. No entanto, é importante notar que a distância própria atual $d_p(t_0)$ de um objeto não é diretamente mensurável. Para realizar essa medição, seria necessário congelar a expansão do universo em um momento específico t_0 e então efetuar a medida, uma tarefa extremamente complexa [42].

Dado que a cosmologia é uma disciplina observational, para calcular a distância até uma galáxia, por exemplo, é essencial encontrar maneiras de determinar essa distância com base em características observáveis da galáxia, dada a impossibilidade prática de medir a distância própria diretamente [42].

Uma abordagem para associar uma característica observational à distância de um objeto cósmico é conhecida como o método da vela padrão. Nesse método, a luminosidade (L) de um objeto é previamente conhecida. Ao medir o fluxo luminoso (f) do objeto, é possível calcular sua distância de luminosidade utilizando a seguinte relação:

$$d_{l,\text{euclidiano}} = \left(\frac{L}{4\pi f} \right)^{1/2}. \quad (5.79)$$

A função $d_{l,\text{euclidiano}}$ é chamada de "distância" porque possui dimensão de distância. Essa distância representaria a distância própria até uma vela padrão se o universo fosse estático e euclidiano [42].

Suponha um observador em um universo descrito pela métrica FLRW, se o observador estiver na origem de um sistema de coordenadas e observa (em um tempo t_0) uma vela padrão com coordenadas comóveis (r, θ, ϕ), cuja luz foi emitida em um tempo t_e então os fotons estão dispostos sobre uma superfície esférica de superfície própria

$$A_p(t_0) = 4\pi S_k(r)^2. \quad (5.80)$$

Outro efeito que a expansão do universo causa nos fôtons é a alteração de sua energia do tempo t_e para o tempo t_0 . Se um fôton é emitido com uma energia

$$E_e = \frac{hc}{\lambda_e}, \quad (5.81)$$

quando o fator de escala era $a(t_e)$, então no tempo atual o comprimento de onda do fôton será esticado para um valor

$$\lambda_0 = \frac{1}{a(t_e)} \lambda_e = (1+z) \lambda_e, \quad (5.82)$$

resultando em uma energia

$$E_0 = \frac{E_e}{1+z}. \quad (5.83)$$

Além disso, se dois fótons forem emitidos na mesma direção, com um intervalo de tempo δt_e entre eles, a distância própria entre os fótons será $c\delta t_e$ [42]. Quando um observador os receber, a distância própria entre eles será $c\delta t_e(1+z)$, e eles serão detectados com um intervalo de tempo separado por $\delta t_0 = (1+z)\delta t_e$ [42].

Assim, a combinação de todos esses efeitos mencionados resulta, em um universo curvo e em expansão, em uma relação entre o fluxo medido f e a luminosidade intrínseca do objeto L , expressa pela equação

$$f = \frac{L}{4\pi S_k(r)^2(1+z)^2}, \quad (5.84)$$

onde a distância de luminosidade d_l é definida como

$$d_l = S_k(r)(1+z). \quad (5.85)$$

O conceito de distância de luminosidade desempenha um papel crucial na cosmologia, permitindo-nos entender como a luz é afetada pela expansão do universo. Esta medida, que leva em conta tanto a curvatura do espaço-tempo quanto os efeitos da expansão, é fundamental para nossa compreensão das observações astronômicas.

5.9.2 Módulo de distância

No ano 137 A.E.C, o astrônomo Hiparco desenvolveu uma escala de classificação para as estrelas, com base em sua luminosidade aparente vista a olho nu. Ele organizou as estrelas em uma escala que variava de 1 (as mais brilhantes) a 6 (as menos brilhantes). Essa abordagem proporcionou uma maneira sistemática de descrever e catalogar as estrelas com base na intensidade de sua luminosidade percebida. Essa classificação desempenhou um papel fundamental na astronomia da época e serviu como alicerce para estudos subsequentes [51].

Muitos séculos mais tarde, em 1850, o astrônomo Pogson fez uma contribuição notável para a astronomia ao introduzir a escala de magnitude. Essa escala representou um marco importante na história da astronomia, estabelecendo uma relação fundamental entre a magnitude aparente, a magnitude absoluta e a distância de um objeto em parsecs [51].

Atualmente, a magnitude aparente bolométrica de um objeto luminoso é determinada com base em seu fluxo bolométrico pela fórmula:

$$m = -\frac{5}{2} \log \left(\frac{f}{f_x} \right), \quad (5.86)$$

onde $f_x = 2,53 \times 10^{-8} \text{ W m}^{-2}$ [42].

A magnitude absoluta bolométrica de um objeto luminoso é definida como a magnitude aparente que o objeto teria a uma distância de luminosidade de 10 pc. Assim, um objeto com luminosidade L tem uma magnitude absoluta bolométrica dada por:

$$M = -\frac{5}{2} \log \left(\frac{L}{L_x} \right), \quad (5.87)$$

onde $L_x = 78.7 L_\odot$ e L_\odot representa a luminosidade solar [42].

Dado que os astrônomos frequentemente utilizam magnitudes aparentes e absolutas ao trabalhar com fluxo e luminosidade, é conveniente expressar a distância de luminosidade em termos do módulo de distância, definido como:

$$\mu = m - M, \quad (5.88)$$

portanto, a relação entre o módulo de distância (μ) e a distância de luminosidade (d_l) é dada pela equação:

$$\mu = 5 \log \left(\frac{d_l}{1 \text{ Mpc}} \right) + 25. \quad (5.89)$$

Em resumo, o estudo das distâncias cósmicas é um desafio fundamental na cosmologia observacional [42]. Embora a distância própria de um objeto não seja diretamente mensurável, métodos como o da vela padrão oferecem uma maneira indireta de determiná-la com base em características observáveis dos objetos, levando em conta não apenas a luminosidade intrínseca, mas também os efeitos da expansão do universo e a curvatura do espaço-tempo. A introdução do módulo de distância, derivado das magnitudes aparente e absoluta, fornece uma forma conveniente de expressar as distâncias de luminosidade em termos observacionais, permitindo aos astrônomos compreenderem as propriedades fundamentais do cosmos em escalas vastas [42].

5.10 Cosmografia

A despeito da robustez da Relatividade Geral, que consegue resolver uma variedade de problemas, ela apresenta algumas questões fundamentais, como discutido anteriormente, sendo uma das mais significativas a nossa falta de compreensão sobre a natureza da energia escura [42]. Diante desse cenário, os cosmólogos estão explorando alternativas para modelar o universo, incluindo teorias alternativas de gravitação, entre outras abordagens. Uma dessas alternativas é conhecida como cosmografia. Dentro da cosmografia, busca-se

analisar as características do universo sem depender de uma teoria específica de gravitação, considerando apenas uma métrica e suas implicações observacionais [54].

A abordagem cosmográfica envolve a análise das propriedades locais ligadas à métrica, com ênfase no fator de escala cosmológico. Inicialmente, essa abordagem se concentra na definição dos parâmetros cosmográficos, que são os termos usados para expandir o fator de escala em uma série de Taylor [54]. A expansão em série de Taylor do fator de escala em torno do tempo presente (t_0) é expressa pela equação

$$a(t) = a(t_0) + \frac{da}{dt} \Big|_{t_0} (t - t_0) + \frac{1}{2} \frac{d^2a}{dt^2} \Big|_{t_0} (t - t_0)^2 + \frac{1}{6} \frac{d^3a}{dt^3} \Big|_{t_0} (t - t_0)^3 + \frac{1}{24} \frac{d^4a}{dt^4} \Big|_{t_0} (t - t_0)^4 + \mathcal{O}(t^5), \quad (5.90)$$

definimos, então, os parâmetros cosmográficos: o parâmetro de Hubble

$$H(t) = \frac{1}{a} \frac{da}{dt} \Big|_t, \quad (5.91)$$

o parâmetro de desaceleração

$$q(t) = -\frac{1}{a} \frac{d^2a}{dt^2} \left(\frac{1}{a} \frac{da}{dt} \right)^{-2} \Big|_t, \quad (5.92)$$

o parâmetro jerk

$$j(t) = \frac{1}{a} \frac{d^3a}{dt^3} \left(\frac{1}{a} \frac{da}{dt} \right)^{-3} \Big|_t \quad (5.93)$$

e o parâmetro snap

$$s(t) = \frac{1}{a} \frac{d^4a}{dt^4} \left(\frac{1}{a} \frac{da}{dt} \right)^{-4} \Big|_t. \quad (5.94)$$

Dessa maneira, a expansão em série de Taylor do fator de escala escrita em termos dos parâmetros cosmográficos é expressa por

$$a(t) = a_0 \left\{ 1 + H_0(t - t_0) - \frac{1}{2} q_0 H_0^2 (t - t_0)^2 + \frac{1}{6} j_0 H_0^3 (t - t_0)^3 + \frac{1}{24} s_0 H_0^4 (t - t_0)^4 \right\}. \quad (5.95)$$

Além de utilizar a aproximação em séries para o fator de escala, é relevante reformular a função $S_k(r)$, descrita pela Equação 5.33, em termos do parâmetro de densidade de curvatura [54]. Para isso, podemos recorrer às séries de Taylor até a terceira ordem das funções seno e cosseno hiperbólico, que são

$$\sin x = x - \frac{x^3}{6} \quad (5.96)$$

e

$$\sinh x = x + \frac{x^3}{6}. \quad (5.97)$$

Assim, a função $S_k(r)$ pode ser reescrita da seguinte maneira:

$$S_k(r) = r \left(1 + \frac{r^2 H_0^2 \Omega_{k0}}{6c^2} \right), \quad (5.98)$$

portanto a distância de luminosidade é reescrita através da relação

$$d_l = (1+z)r \left(1 + \frac{r^2 H_0^2 \Omega_{k0}}{6c^2} \right). \quad (5.99)$$

Agora, consideremos um fóton que percorreu uma distância física D durante um intervalo de tempo $\frac{D}{c}$. Nesse caso, sua coordenada comóvel r pode ser expressa como:

$$r = \int_{t_0 - \frac{D}{c}}^{t_0} \frac{cdt}{a(t)} = D \left\{ 1 + \frac{1}{2} \frac{H_0 D}{c} + \left(\frac{2+q_0}{6} \right) \left(\frac{H_0 D}{c} \right)^2 + \left[\frac{6(1+q_0) + j_0}{24} \right] \left(\frac{H_0 D}{c} \right)^3 \right\}. \quad (5.100)$$

Essa distância física está associada a um redshift

$$1+z = \frac{a_0}{a(t_0 - \frac{D}{c})}, \quad (5.101)$$

com a definição da expansão em série de Taylor do fator de escala, se obtém

$$\frac{a_0}{a(t_0 - \frac{D}{c})} = 1 + \frac{H_0 D}{c} + \frac{2+q_0}{2} \frac{H_0^2 D^2}{c^2} + \frac{6(1+q_0) + j_0}{6} \frac{H_0^3 D^3}{c^3} + \frac{24 - s_0 + 8j_0 + 36q_0 + 6q_0^2}{24} \frac{H_0^4 D^4}{c^4}, \quad (5.102)$$

representando um redshift de

$$z = \frac{H_0 D}{c} + \frac{2+q_0}{2} \frac{H_0^2 D^2}{c^2} + \frac{6(1+q_0) + j_0}{6} \frac{H_0^3 D^3}{c^3} + \frac{24 - s_0 + 8j_0 + 36q_0 + 6q_0^2}{24} \frac{H_0^4 D^4}{c^4}. \quad (5.103)$$

Ao inverter essa série para obter D em função de z , obtemos

$$D = \frac{cz}{H_0} \left\{ 1 - \left(1 + \frac{q_0}{2} \right) z + \left(1 + q_0 + \frac{q_0^2}{2} - \frac{j_0}{6} \right) z^2 - \left[1 + \frac{3}{2} q_0 (1+q_0) + \frac{5}{8} q_0^3 - \frac{1}{2} j_0 - \frac{5}{12} q_0 j_0 - \frac{s_0}{24} \right] z^3 \right\}, \quad (5.104)$$

consequentemente, a distância de luminosidade é expressa a partir de

$$d_l(z; H_0, q_0, j_0, s_0, \Omega_{k0}) = \frac{cz}{H_0} \left\{ 1 + \frac{1}{2} (1 - q_0) z - \frac{1}{6} (1 - q_0 - 3q_0^2 + j_0 + \Omega_{k0}) z^2 + \frac{1}{24} [2 - 2q_0 - 15q_0^2 - 15q_0^3 + 5j_0 + 10q_0 j_0 + s_0 + 2\Omega_{k0}(1 + 3q_0)] z^3 \right\}. \quad (5.105)$$

É crucial observar que a expansão em Série de Taylor para a distância de luminosidade, representada pela Equação 5.105, enfrenta desafios de convergência quando lidamos com redshifts da ordem de $z = 1$ [5]. Nesse contexto, é pertinente explorar uma abordagem alternativa, na qual, em vez de adotarmos diretamente o redshift, realizamos a expansão em Série de Taylor para a nova variável $y = \frac{z}{z+1}$ [5]. A série resultante é expressa por:

$$d_l(y; H_0, q_0, j_0, \Omega_{k0}) = \frac{cy}{H_0} \left\{ 1 + \frac{1}{2} (3 - q_0) y + \frac{1}{6} (11 - 5q_0 + 3q_0^2 - j_0 + \Omega_{k0}) y^2 \right\}. \quad (5.106)$$

Essa mudança de variável é extremamente relevante, pois permite trabalhar com redshifts superiores a 1, o que corresponde a $y = \frac{1}{2}$. Notavelmente, a série não diverge no período que vai desde o big bang até o momento atual, conforme $\lim_{z \rightarrow \infty} y = 1$ [5].

Neste capítulo, exploramos a abordagem cosmográfica para entender o universo sem depender de uma teoria específica de gravitação. Derivamos uma expressão para a distância de luminosidade em termos do redshift usando parâmetros cosmográficos. Embora essa abordagem ofereça uma descrição precisa da expansão cósmica, suas limitações em altas ordens de aproximação exigem cautela [54]. A cosmografia continua sendo uma ferramenta útil para cosmólogos, fornecendo insights sobre a natureza do universo, apesar das correções necessárias em níveis avançados de precisão.

6 Vínculos observacionais sobre o modelo cosmológico padrão

Nos capítulos anteriores, exploramos uma variedade de conceitos relacionados aos campos da cosmologia, análise de dados e inteligência artificial. No âmbito da análise de dados, abordamos os princípios de probabilidade e estatística para realizar inferências sobre modelos estatísticos, incluindo previsões de parâmetros e testes de modelos. Na cosmologia, discutimos aspectos fundamentais, especialmente os observacionais, como o módulo de distância, o parâmetro de Hubble, distância de luminosidade, redshift, parâmetro de densidade de curvatura e de matéria, além da aproximação cosmográfica.

Na área de inteligência artificial, exploramos duas técnicas poderosas de aprendizado de máquina que desempenham papéis significativos no desenvolvimento tecnológico e na evolução do conhecimento na sociedade atual. Primeiro, examinamos o método dos processos gaussianos, uma abordagem semiparamétrica que utiliza princípios estatísticos fundamentais, como o teorema do limite central da probabilidade, conceitos de distribuição de funções e o teorema de Bayes para fazer previsões em uma variedade de tipos de dados. Este método é especialmente valioso ao lidar com dados observacionais que apresentam erros, sejam eles sistemáticos ou estatísticos, em suas medições, sendo aplicável tanto a técnicas de regressão quanto de classificação.

Além disso, discutimos o conceito de redes neurais, uma abordagem poderosa inspirada na estrutura do cérebro humano. As redes neurais se mostraram ferramentas sofisticadas para prever diversos fenômenos, resolver problemas de classificação e regressão, e são fundamentais para modelos robustos em áreas como tecnologia, economia e cosmologia.

Nesta seção, concentraremos nossos esforços em aplicar esses conceitos para obter insights sobre várias características cosmológicas. Especificamente, utilizaremos dados de quasares, supernovas e do parâmetro de Hubble. Implementaremos o método de Monte Carlo para estabelecer vínculos observacionais sobre os parâmetros cosmológicos de interesse, demonstrando assim a aplicação prática dos conhecimentos discutidos.

6.1 Quasares, supernovas, galáxias HII e relógios cósmicos

Nesta seção, focamos em quatro fontes cruciais de dados na cosmologia: supernovas do Tipo Ia, quasares e as oscilações acústicas da matéria bariônica, juntamente com os cronômetros cósmicos e Galáxias HII. Esses fenômenos astronômicos não são apenas fascinantes, mas também fornecem dados essenciais para nossa compreensão do universo. As supernovas do Tipo Ia, devido à sua padronização luminosa, são fundamentais para medir distâncias cósmicas, revelando pistas sobre a expansão do universo. Quasares, como núcleos ativos de galáxias distantes, oferecem dados valiosos sobre o universo em altos redshifts. Além disso, as oscilações acústicas da matéria bariônica e os cronômetros cósmicos proporcionam métodos precisos para medir o parâmetro de Hubble, oferecendo insights detalhados sobre a expansão do universo em diferentes épocas cósmicas. Adicionalmente, as Galáxias HII oferecem informações valiosas sobre características observacionais relevantes para regiões cósmicas extremamente distantes, relacionando-se a mecanismos físicos amplamente estabelecidos nos fenômenos estudados na Terra.

6.1.1 Supernovas

Supernovas do tipo Ia representam eventos estelares de grande magnitude e luminosidade, marcando o término da jornada de certas estrelas no universo. Essas explosões estelares têm origem em sistemas binários, nos quais uma anã branca, uma estrela compacta e densa, acumula matéria de sua estrela parceira até atingir uma massa crítica. Quando essa massa crítica é atingida, a anã branca não consegue mais suportar a pressão gravitacional, resultando em um colapso que desencadeia uma espetacular explosão, liberando uma quantidade extraordinária de energia. A importância das supernovas do Tipo Ia na cosmologia reside no fato de que o mecanismo de geração dessas supernovas ocorre de maneira consistente, conferindo a elas uma magnitude absoluta uniforme. Assim, elas desempenham o papel de "velas padrão", fornecendo uma referência crucial para medições cosmológicas.

Dentre os diversos projetos que fornecem dados sobre supernovas, se destaca a amostra Pantheon, um catálogo abrangente que reúne 1048 supernovas do Tipo Ia, cuidadosamente verificadas para confirmar sua classificação e desvio para o vermelho. Em contraste com outros levantamentos cosmológicos, o Pantheon se destaca por sua atenção particular à faixa desafiadora além de redshift da ordem de 2, anteriormente menos explorada. Originadas de várias fontes, como Pan-STARRS, Sloan Digital Sky Survey (SDSS), Supernova Legacy Survey (SNLS) e Carnegie Supernova Project (CSP), essas supernovas passaram por uma meticulosa verificação, proporcionando dados confiáveis para estudos cosmológicos. O catálogo Pantheon é uma contribuição crucial para a astrofísica, oferecendo uma base sólida para pesquisas sobre cosmologia e supernovas, ampliando nosso

entendimento da dinâmica do universo em escalas cosmológicas [43]. Nesse trabalho, serão utilizados os dados do Pantheon, representados pela Figura 6.1.

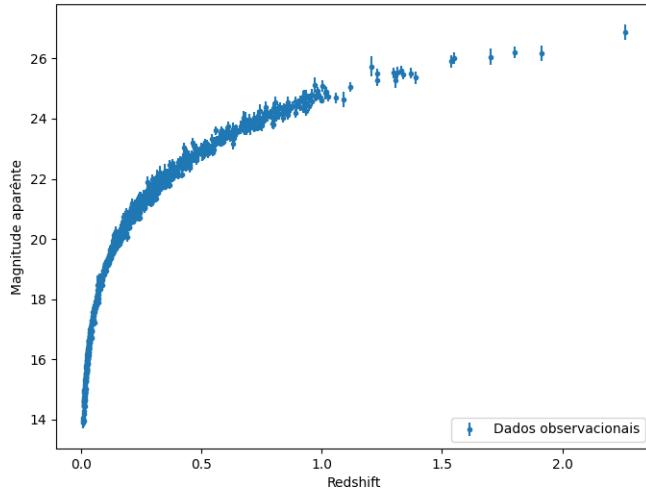


Figura 6.1: Dados de supernova da amostra Pantheon. Fonte: Colaboração Pan-STARRS [22].

6.1.2 Quasares

Quasares, como as supernovas, são objetos astronômicos extremamente intrigantes, mas por razões diferentes. Eles representam os núcleos ativos de galáxias distantes, onde enormes quantidades de energia são liberadas devido à atividade intensa em torno de um buraco negro supermassivo central. Essa atividade é alimentada pela matéria que é atraída pelo buraco negro, formando um disco de acreção extremamente quente e brilhante.

A obtenção do módulo de distância para quasares é um processo complexo devido à sua extrema distância da Terra. Diferentemente das supernovas do Tipo Ia, que são usadas como "velas-padrão", os quasares não possuem uma relação tão direta entre sua luminosidade intrínseca e outras características observáveis. No entanto, os astrônomos desenvolveram diversas técnicas para estimar as distâncias das estrelas com base em observações realizadas em diferentes regiões do espectro eletromagnético. Uma dessas abordagens envolve o estudo do desvio para o vermelho e do fluxo luminoso. Em particular, os astrofísicos criaram várias relações empíricas para determinar a distância de luminosidade dos quasares. Por exemplo, Lusso et al. (2020) apresenta a seguinte relação em seu trabalho:

$$\ln D_l = \frac{[\log F_X - \beta - \gamma(F_{UV} + 27.5)]}{2(\gamma - 1)} - \frac{1}{2} \ln 4\pi + 28.5, \quad (6.1)$$

onde D_l é a distância de luminosidade, F_X é o fluxo luminoso em raios X, F_{UV} é o fluxo luminoso em ultravioleta, e γ e β são parâmetros livres do modelo.

Nesse trabalho, utilizaremos os dados de quasares obtidos a partir da metodologia de Lusso et al. Representado pela Figura 7.4.

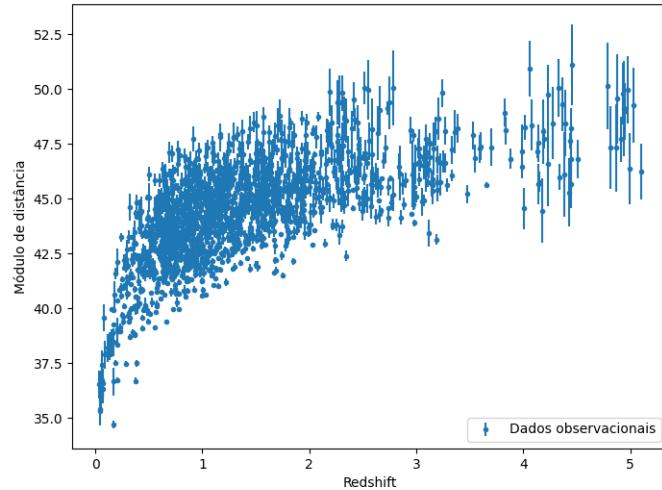


Figura 6.2: Dados de quasares. Fonte: Lusso et. al. [28].

6.1.3 Galáxias HII

Galáxias "starburst" HII representam ambientes cósmicos excepcionalmente prolíficos em termos de formação estelar. Caracterizadas por uma intensa atividade de ionização do hidrogênio, essas galáxias testemunham a influência preponderante de estrelas massivas, cuja radiação ultravioleta provoca a criação de regiões HII resplandecentes. Esse fenômeno é indicativo de uma fase transitória de notável efervescência estelar, resultando em uma taxa de formação estelar substancialmente elevada em comparação com galáxias mais estáveis. O estudo aprofundado desses sistemas oferece perspectivas cruciais sobre os processos físicos e dinâmicos que governam a evolução galáctica, desempenhando um papel significativo na compreensão das intrincadas interações entre gás interestelar, poeira e formação estelar em escalas cósmicas [4].

Para as galáxias HII, os astrofísicos encontraram relações empíricas entre a luminosidade L observada, das linhas espectrais da Série de Balmer, do hidrogênio duplamente iônizado com a velocidade de dispersão do gás (representada por σ)

$$\log L = \beta \log \sigma + \gamma, \quad (6.2)$$

portanto, com o fluxo f medido, se obtém uma relação para o módulo de distância dessas regiões:

$$\mu = \frac{5}{2} \log L - \frac{5}{2} \log f - 100.2. \quad (6.3)$$

Neste estudo, empregaremos os dados provenientes da pesquisa conduzida por Cao et al. [4], na qual foram determinados os valores de $\beta = 5.022 \pm 0.058$ e $\gamma = 33.268 \pm 0.083$. Esses parâmetros resultaram nos dados apresentados na Figura 6.3.

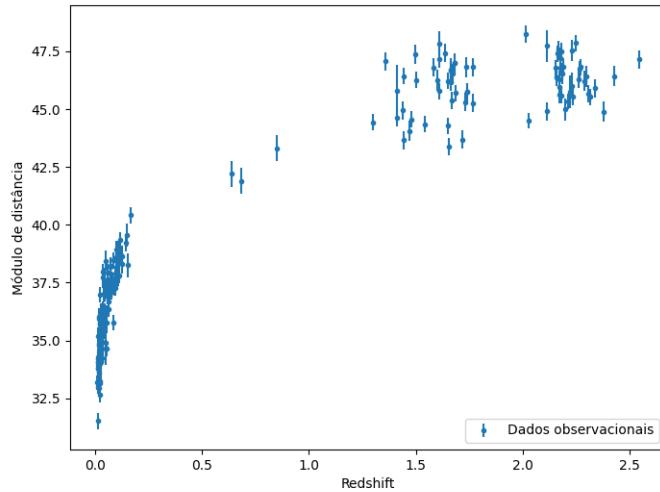


Figura 6.3: Dados de galáxias HII. Fonte: Cao et. al. [4].

6.1.4 Oscilações acústicas da matéria bariônica e os cronômetros cósmicos

A medição precisa do parâmetro de Hubble é uma questão fundamental na cosmologia, permitindo-nos entender a taxa de expansão do universo ao longo do tempo. Para alcançar essa precisão, os astrônomos utilizam técnicas sofisticadas, como as oscilações acústicas da matéria bariônica (BAO). As BAO são padrões regulares na distribuição de matéria cósmica que se originaram das flutuações no plasma primordial do universo. Esses padrões deixam uma marca nas posições das galáxias e quasares, possibilitando aos cientistas calcular distâncias cósmicas precisas. Ao combinar essas informações com a velocidade das galáxias, obtida a partir do desvio para o vermelho cosmológico, os astrônomos podem mappear a expansão do universo ao longo do tempo. Esse método oferece uma visão detalhada da expansão cósmica e contribui para nossa compreensão das forças que impulsionam o universo em sua trajetória evolutiva.

Outra abordagem essencial para medir o parâmetro de Hubble é por meio dos Cronômetros Cósmicos (CC), que são galáxias que atuam como "cronômetros naturais". Ao observar diferentes estágios de evolução em diversas galáxias, os cientistas podem calcular o tempo necessário para essas mudanças ocorrerem. Dividindo a distância dessas galáxias pela variação de velocidade ao longo do tempo, é possível determinar a taxa de expansão do universo em momentos específicos, fornecendo dados do parâmetro de hubble ao longo do desvio para o vermelho. Nesse trabalho, se utilizaram os dados obtidos do trabalho de

Marek et al. [57], intitulado ”WHAT ARE THE $\Omega m h^2$ (z_1, z_2) AND Ωm (z_1, z_2) DIAGNOSTICS TELLING US IN LIGHT OF $H(z)$ DATA?”, os dados estão apresentados na Figura 6.4.

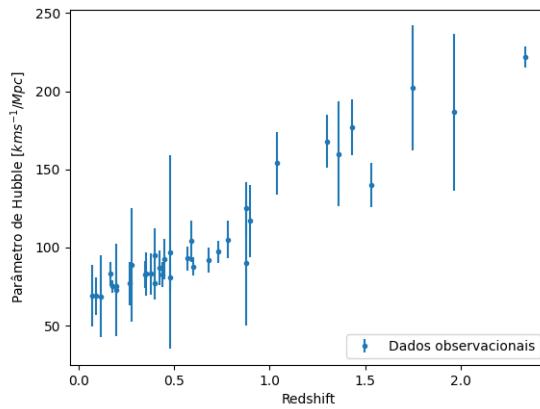


Figura 6.4: Dados do parâmetro de Hubble. Fonte: Marek et. al. [13].

Em resumo, quasares, supernovas do Tipo Ia, oscilações acústicas da matéria bariônica e cronômetros cósmicos são fenômenos astronômicos cruciais para a cosmologia. As supernovas do Tipo Ia representam explosões estelares resultantes do colapso de anãs brancas, liberando quantidades impressionantes de energia. Os quasares são núcleos ativos de galáxias distantes, alimentados por atividade intensa ao redor de buracos negros supermassivos. As oscilações acústicas da matéria bariônica são padrões na distribuição de matéria cósmica, proporcionando dados valiosos sobre a estrutura do universo. Os cronômetros cósmicos, por sua vez, oferecem insights sobre a expansão do universo ao longo do tempo. Essas características distintas tornam esses fenômenos fontes essenciais de dados na cosmologia contemporânea.

6.1.5 Método dos processos gaussianos e redes neurais: aplicação na linguagem de programação Python

Após o avanço das redes neurais e do método dos processos gaussianos, houve um significativo progresso teórico nessas técnicas. Nesse cenário, surgiram várias abordagens para implementar modelos de redes neurais e processos gaussianos. Entre essas abordagens, existem bibliotecas desenvolvidas em Python que oferecem métodos robustos para a implementação dessas técnicas de machine learning.

A biblioteca GaPP, desenvolvida por Marina Seikel em agosto de 2013, é uma ferramenta excepcionalmente robusta para a implementação do método dos processos gaussianos, oferecendo uma variedade de funcionalidades poderosas. Com o GaPP, é possível calcular a matriz de covariância, reconstruir funções complexas e suas derivadas até a terceira ordem, bem como calcular as covariâncias entre a função reconstruída e suas

respectivas derivadas. Esta capacidade proporciona uma visão abrangente do comportamento das funções a partir de dados observacionais, permitindo aos pesquisadores e cientistas de dados explorar profundamente as complexas relações nos dados. Ao possibilitar a compreensão detalhada de padrões e tendências subjacentes aos dados observados, o GaPP se destaca como uma ferramenta fundamental para análises avançadas em áreas como física, estatística e aprendizado de máquina, oferecendo insights valiosos para pesquisas e descobertas científicas.

A biblioteca Pytorch é uma abordagem inovadora para a criação de arquiteturas de Redes Neurais Artificiais (ANNs) para reconstruir funções a partir de dados observacionais, sem fazer suposições a priori sobre a distribuição dos dados. A Pytorch representa uma ferramenta poderosa para a reconstrução de parâmetros cosmológicos, destacando-se pela sua abordagem totalmente orientada por dados. Essa metodologia, exemplificada na reconstrução diversas grandezas, oferece resultados precisos e consistentes com as observações diretas. Ao superar possíveis limitações associadas a métodos tradicionais, essas redes neurais emergem como uma alternativa promissora para a estimativa de parâmetros cosmológicos, contribuindo para avanços significativos na pesquisa cosmológica.

6.1.6 MCMC: aplicação na linguagem de programação python

Dentre as diversas bibliotecas em Python destinadas à aplicação das MCMC para a inferência estatística, o emcee destaca-se. Desenvolvida por Daniel Foreman-Mackey, esta biblioteca baseia-se no trabalho de Goodman & Weare (2010).

Ao contrário das técnicas tradicionais, o emcee utiliza a técnica de amostragem invariantemente afin. Este método inicia o processo varrendo o espaço dos parâmetros em busca das regiões de maior probabilidade. Em seguida, inicia os "caminhantes" em uma esfera N-dimensional ao redor de um ponto próximo ao máximo de probabilidade, permitindo uma exploração mais dinâmica e eficaz do espaço paramétrico. Isso resulta em um menor tempo de processamento, concentrando os cálculos computacionais apenas nas áreas mais propensas a fornecer os parâmetros desejados.

Além disso, o emcee se destaca pela eficácia do paralelismo em suas implementações das MCMC. A capacidade de explorar espaços paramétricos de maneira eficiente é potencializada pelo paralelismo, permitindo que múltiplos caminhantes evoluam simultaneamente e independentemente. Isso resulta em uma utilização otimizada dos recursos computacionais disponíveis, acelerando consideravelmente a convergência do método MCMC.

Em síntese, a biblioteca emcee para Markov Chain Monte Carlo (MCMC) em Python emerge como uma ferramenta eficaz para a inferência estatística. Sua abordagem inovadora de amostragem, aliada à implementação eficiente do paralelismo, acelera a convergência do método, tornando-a uma escolha valiosa para análises estatísticas avançadas. O emcee destaca-se pela capacidade de explorar de forma dinâmica o espaço paramétrico,

resultando em um processamento mais rápido e eficaz.

7 Discussão

No Capítulo 6, abordamos a metodologia que empregamos para aplicar técnicas de machine learning na reconstrução de dados e na obtenção de vínculos sobre parâmetros de modelos. Nesse contexto, exploramos a aplicação do método dos processos gaussianos, utilizando a biblioteca GaPP, e das redes neurais, por meio da biblioteca ReFANN, para reconstruir os dados relacionados aos parâmetros do Hubble, dados de quasares, supernovas e galáxias HII.

Para a reconstrução dos dados a partir do método dos processos gaussianos, utilizamos a função de media constante nula

$$\mu(x) = 0, \quad (7.1)$$

e a função de covariância denominada Matern 5/2, definida a partir da relação

$$k(x, x') = \frac{2^{1-\frac{5}{2}}}{\Gamma(\frac{5}{2})} \left(\frac{|x - x'| \sqrt{2 \cdot \frac{5}{2}}}{\ell} \right)^{\frac{5}{2}} K_{\frac{5}{2}} \left(\frac{|x - x'| \sqrt{2 \cdot \frac{5}{2}}}{\ell} \right), \quad (7.2)$$

onde K é a função de Bessel Modificada e Γ é a função Gamma.

As redes neurais apresentam estruturas altamente versáteis, permitindo a modelagem de uma variedade de fenômenos com facilidade. Diante dessa flexibilidade, há diversas configurações possíveis para redes neurais. Portanto, optarei por empregar diferentes configurações de redes neurais para analisar os dados em questão.

Para analisar os dados relativos às galáxias H2, quasares e supernovas, empregou-se uma rede neural composta por uma camada de entrada contendo N neurônios, relacionados aos dados de entrada (redshift). Adicionalmente, foram incorporadas três camadas internas com 8, 64 e 8 neurônios, respectivamente, e uma camada de saída com $2N$ neurônios. Esta última camada está associada aos dados da magnitude grandeza dependente do redshift e ao erro associado a essa grandeza. Para os dados de parâmetro de Hubble, a estrutura da rede neural criada foi definida por uma camada de entrada, contendo N neurônios, relacionados aos dados de entrada (redshift). Além disso, foram incorporadas mais três camadas internas, compostas por 6, 32 e 6 neurônios, respectivamente. Adicionalmente, incluiu-se uma camada de saída com $2N$ neurônios, os quais

estão associados aos dados de parâmetro de Hubble e à incerteza do referido parâmetro.

Quanto à função de perda, optou-se (para ambos os modelos) pela Mean Squared Error (MSE), que é definida pela distância euclidiana entre os dados gerados pela rede neural e os dados de treinamento.

A representação dos dados das galáxias HII, do parâmetro de Hubble, das Supernovas e dos quasares, obtidos por meio do método dos processos gaussianos, está delineada nas Figuras 7.1, 7.2, 7.3 e 7.4, respectivamente.

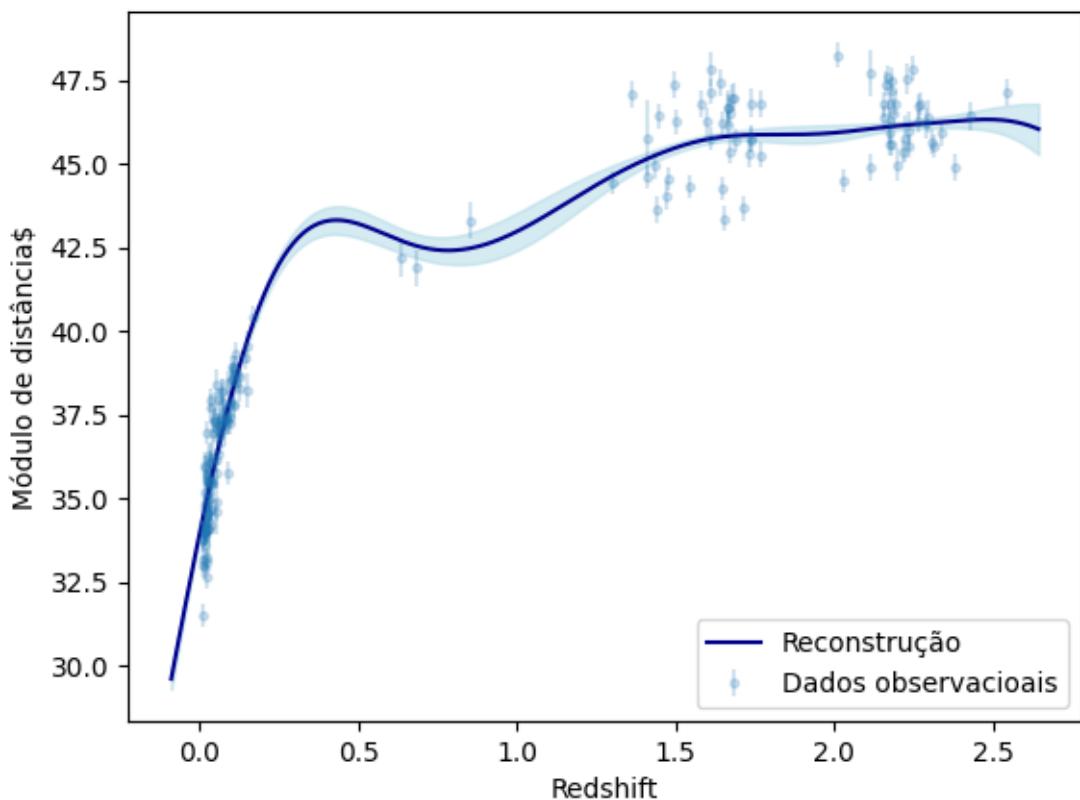


Figura 7.1: Reconstrução dos dados de Galáxias HII obtida através do método dos processos gaussianos. Fonte: Autor próprio.

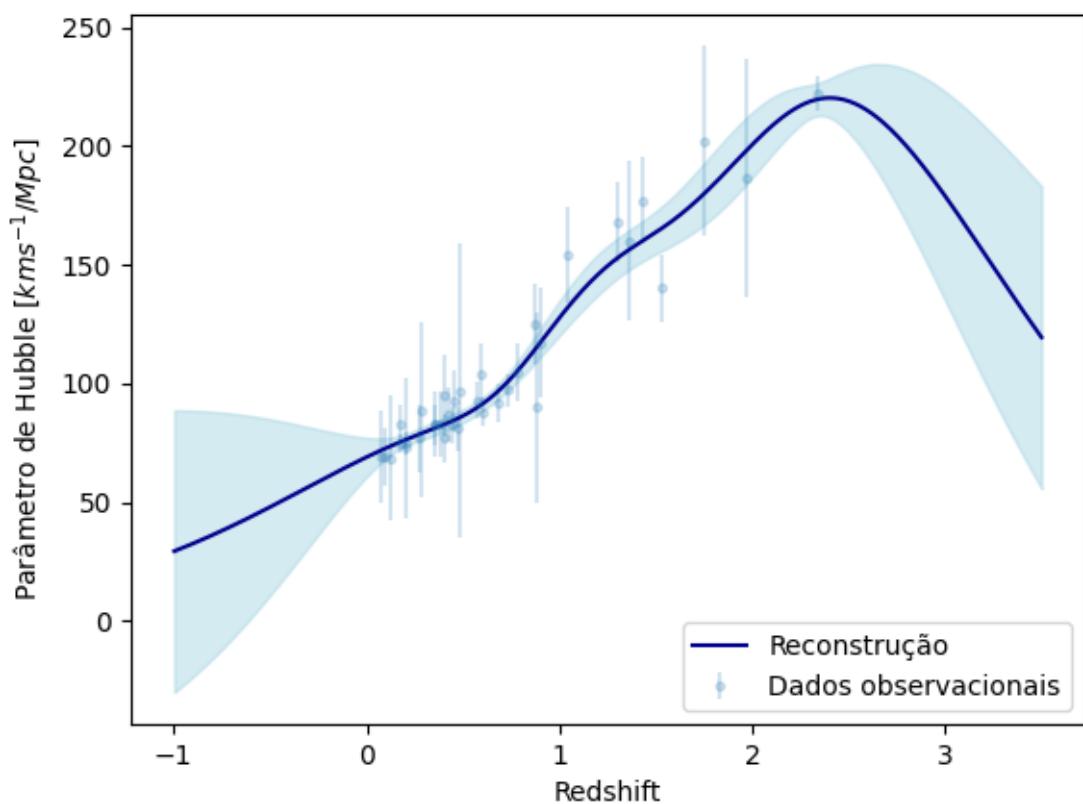


Figura 7.2: Reconstrução dos dados de parâmetro de Hubble obtida através do método dos processos gaussianos. Fonte: Autor próprio. [13].

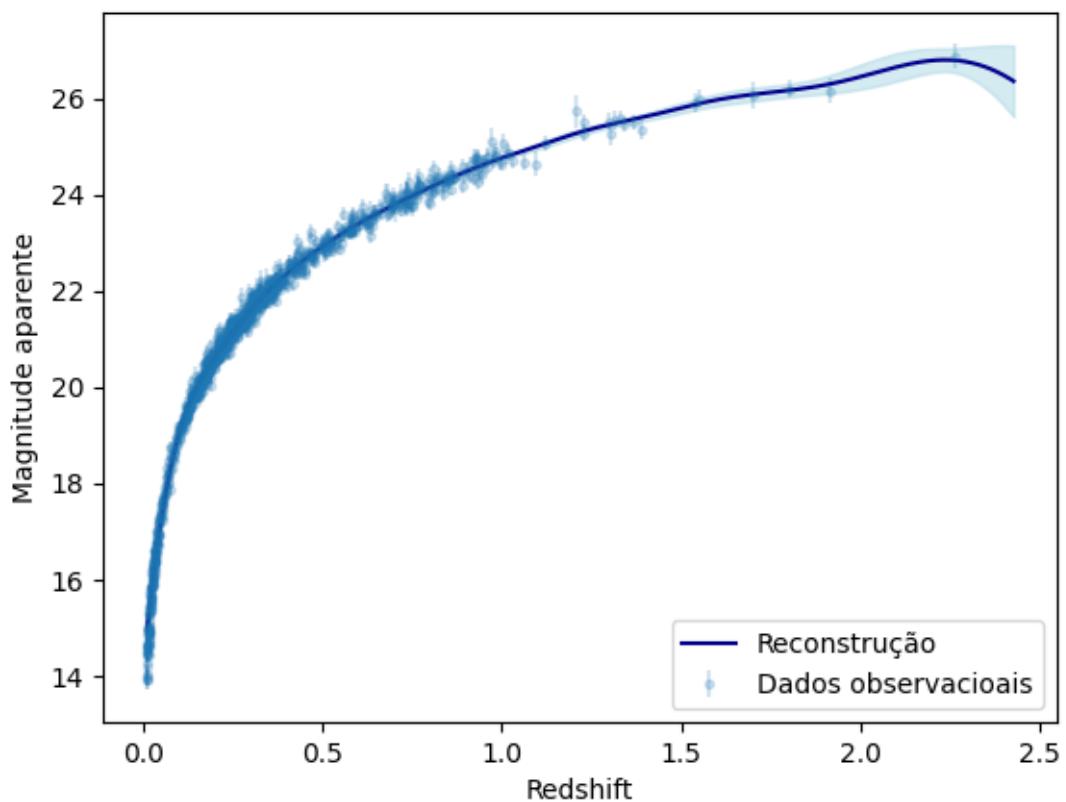


Figura 7.3: Reconstrução dos dados de supernovas Ia obtida através do método dos processos gaussianos. Fonte: Autor próprio.

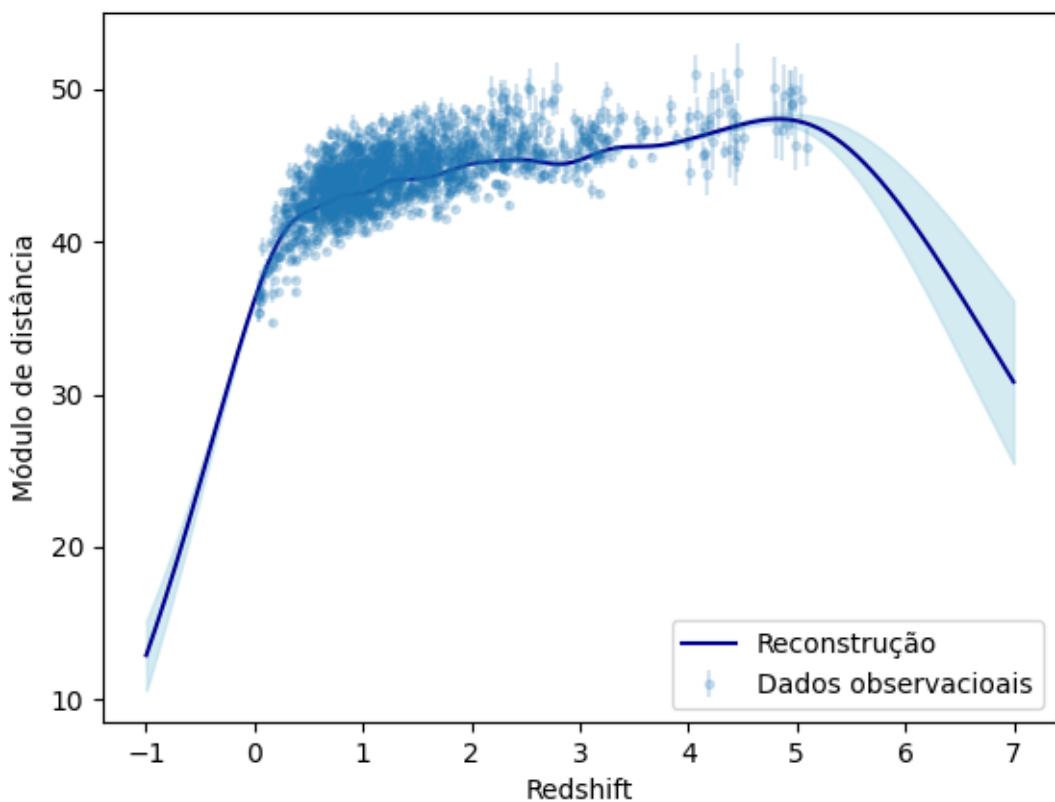


Figura 7.4: Reconstrução dos dados de quasares obtida através do método dos processos gaussianos. Fonte: Autor próprio.

No que diz respeito à reconstrução realizada pelas redes neurais, os resultados referentes aos dados das galáxias HII, do parâmetro de Hubble, das Supernovas e dos quasares são apresentados nas Figuras 7.5, 7.6, 7.7 e 7.8, respectivamente.

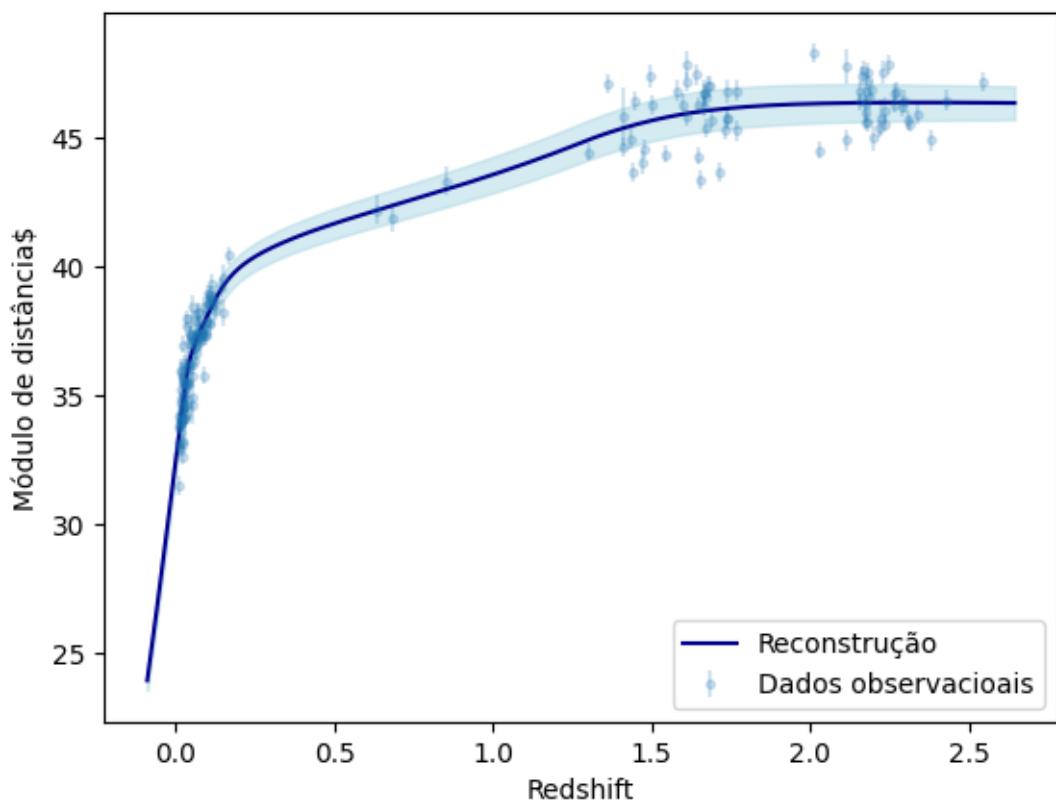


Figura 7.5: Reconstrução dos dados de Galáxias HII obtida através das redes neurais.
Fonte: Autor próprio.

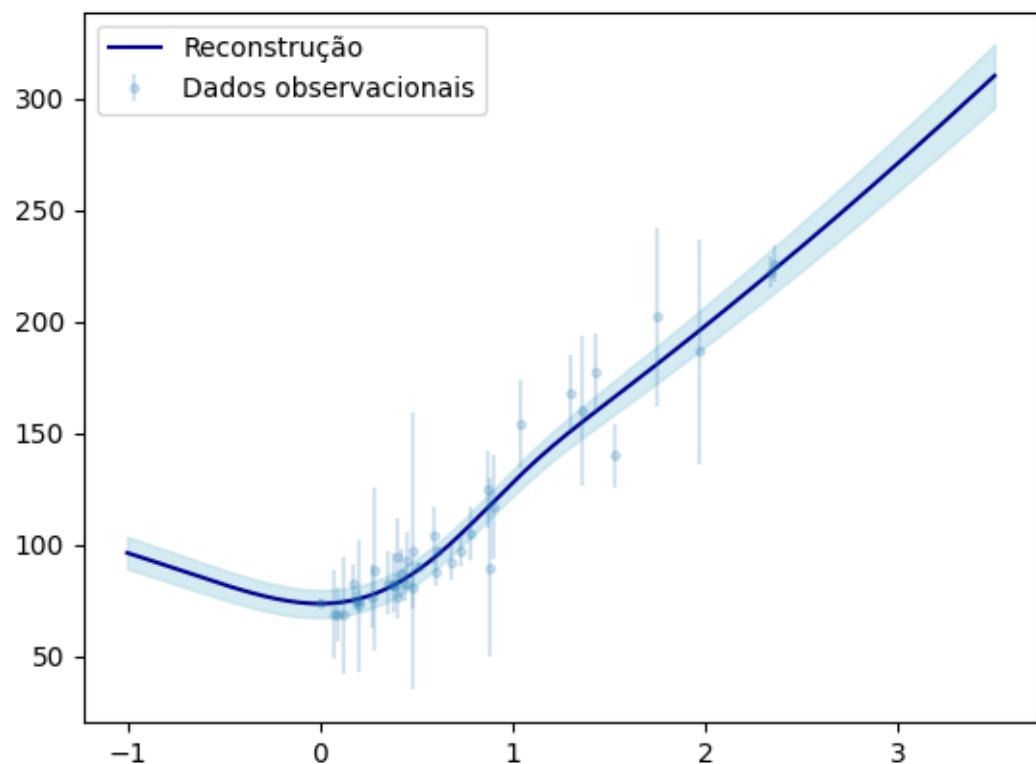


Figura 7.6: Reconstrução dos dados de parâmetro de Hubble obtida através das redes neurais. Fonte: Autor próprio. [13].

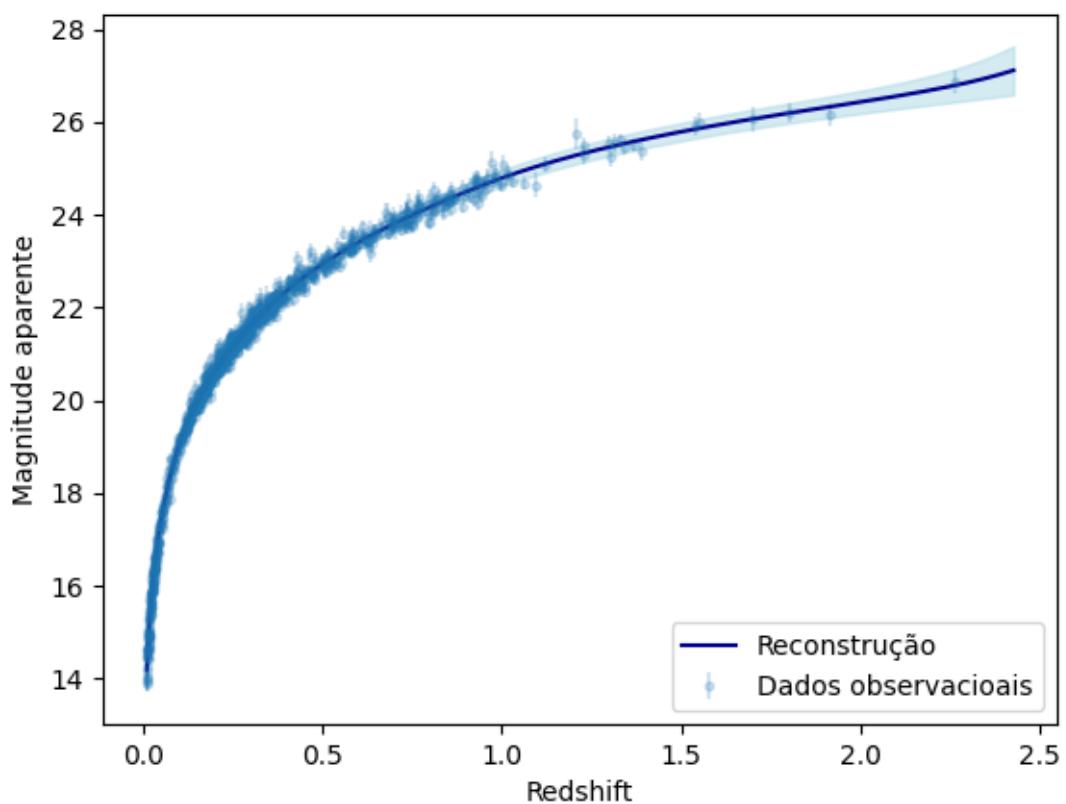


Figura 7.7: Reconstrução dos dados de supernovas Ia obtida através das redes neurais.
Fonte: Autor próprio.

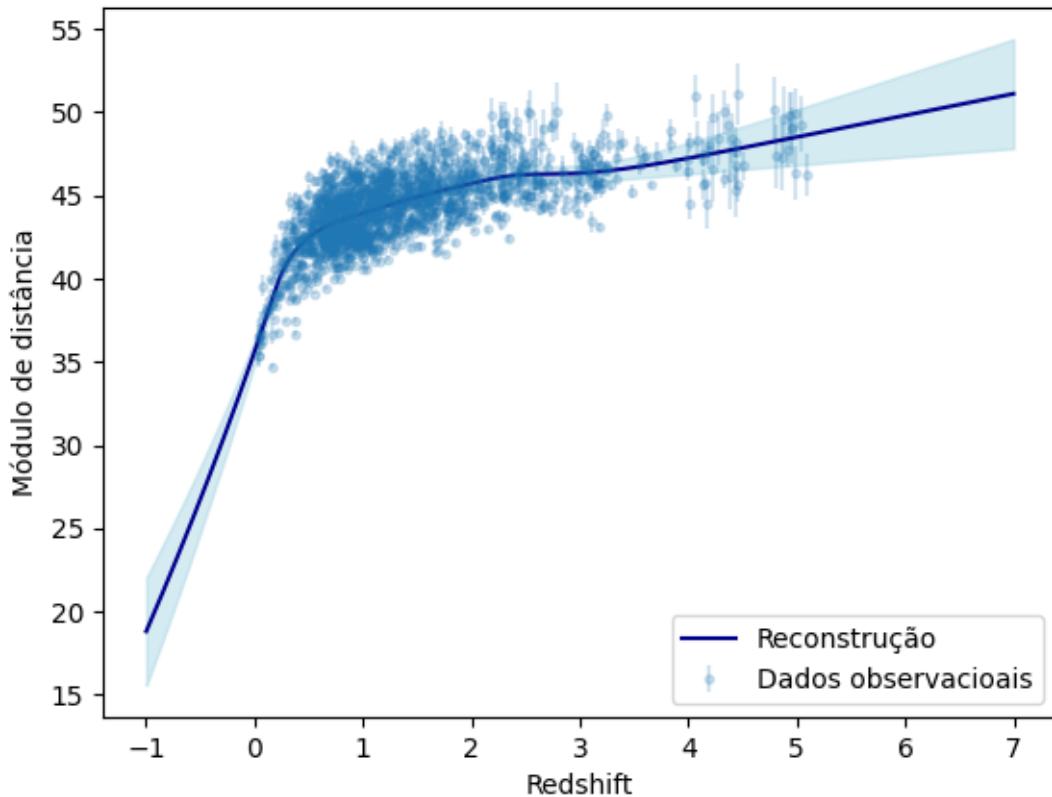


Figura 7.8: Reconstrução dos dados de quasares obtida através das redes neurais. Fonte: Autor próprio.

Os resultados destacam a eficácia dos métodos de aprendizado de máquina na reconstrução de grandezas observacionais, o que é particularmente relevante diante do cenário atual de grande volume de dados. Em tal contexto, os métodos de aprendizado de máquina desempenham um papel crucial ao proporcionar significado físico às grandezas obtidas da natureza, especialmente em situações em que a obtenção de dados de qualidade é desafiadora ou quando os dados são escassos.

É nesse cenário que técnicas como o método dos processos gaussianos e as redes neurais se mostram pertinentes. Isso se torna evidente ao aplicarmos essas técnicas aos dados gerados, permitindo a obtenção de vínculos observacionais relacionados a parâmetros cosmológicos. Dentro desse contexto, destaca-se o interesse em demonstrar a robustez dos métodos de aprendizado de máquina abordados ao longo deste estudo, particularmente quando aplicados aos dados de parâmetros do Hubble.

Nesse contexto, conduzimos uma comparação entre os vínculos observacionais obtidos diretamente da natureza, conforme apresentado no trabalho de Faroq et al. [13], e os dados reconstruídos tanto pelo método dos processos gaussianos quanto pelas redes neurais. Os resultados dessas comparações estão representados na figura 7.9.

Os valores vinculados para ambos os dados estão dispostos na Tabela 7.1.

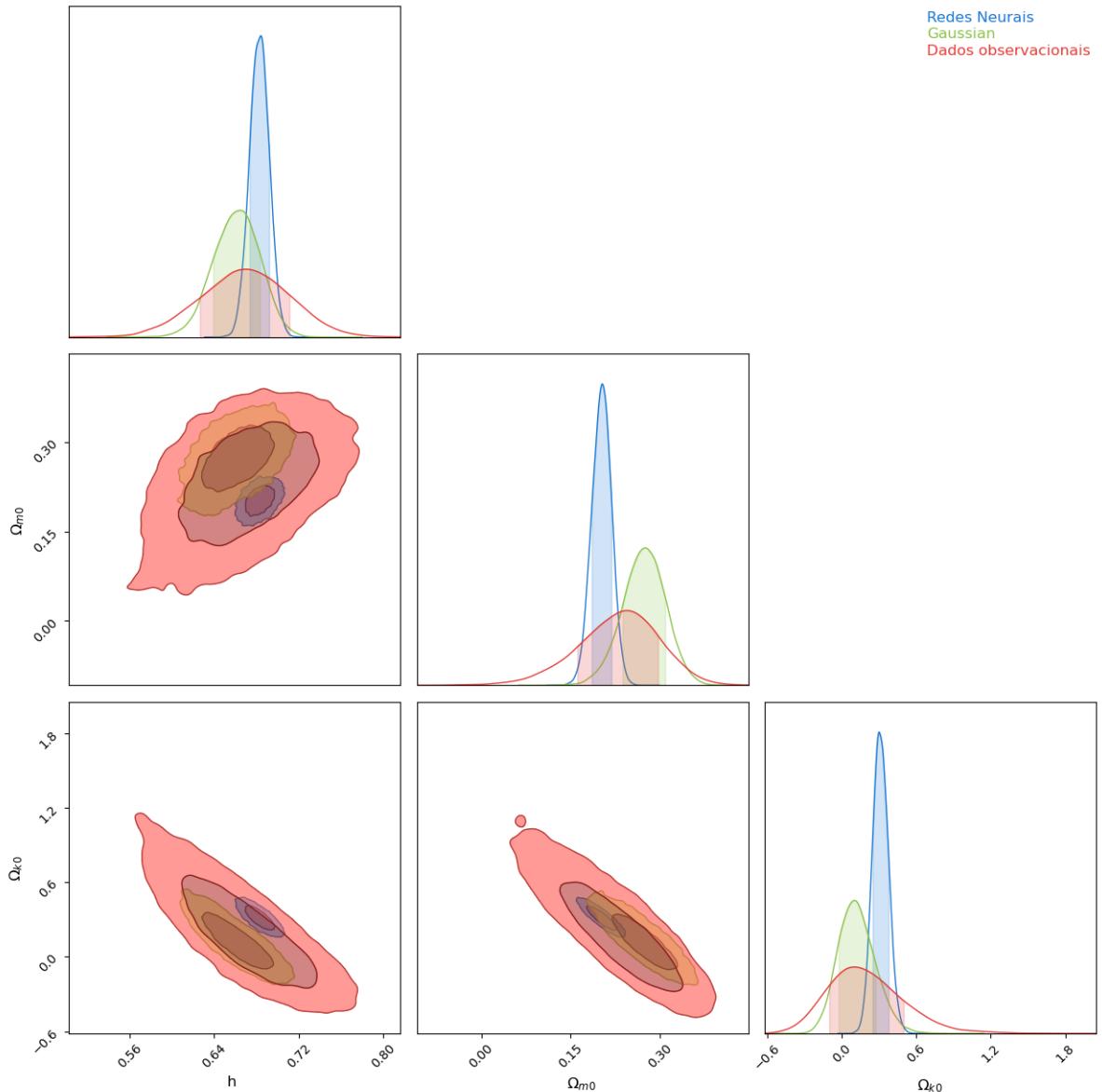


Figura 7.9: Comparação entre os resultados obtidos para a inferência estatística dos dados do parâmetro de Hubble para os dados reconstruídos e os dados obtidos diretamente da natureza. Fonte: Autor próprio.

Tabela 7.1: Comparação entre os resultados obtidos para a inferência estatística dos dados do parâmetro de Hubble para os dados reconstruídos e os dados obtidos diretamente da natureza

Dados de H	$H_0/100$	Ω_{m0}	Ω_{k0}
Redes Neurais	$(682.8 \pm 9.2) \times 10^{-3}$	0.202 ± 0.016	0.313 ± 0.064
Processos gaussianos	0.661 ± 0.022	0.273 ± 0.036	0.12 ± 0.15
Dados observacionais	0.669 ± 0.042	0.229 ± 0.068	0.20 ± 0.30

Na análise conduzida, os processos gaussianos e as redes neurais demonstraram eficácia na previsão e reconstrução dos dados observacionais, exibindo uma precisão adequada em comparação com as evidências observacionais. Embora os resultados referentes ao vínculo do parâmetro de densidade de curvatura obtidos pelas redes neurais apresentem uma ligeira disparidade em relação aos resultados provenientes do método dos processos gaussianos e dos dados observacionais, os vínculos para o parâmetro de Hubble e o parâmetro de densidade de matéria são considerados adequados. Isso ressalta o potencial significativo dos métodos de aprendizado de máquina em aplicações cosmológicas, fornecendo uma ferramenta poderosa para desvendar os mistérios que permeiam a natureza. No entanto, é crucial exercer cautela ao empregar técnicas de aprendizado de máquina, dada a diversidade de estruturas possíveis e os desafios associados à busca por vínculos consistentes com os dados observacionais.

8 Conclusão

Através da abordagem detalhada nos capítulos anteriores, exploramos conceitos fundamentais de cosmologia, análise de dados e inteligência artificial. Com ênfase em métodos estatísticos, discutimos a aplicação de técnicas avançadas de aprendizado de máquina, especificamente o método dos processos gaussianos e redes neurais, para reconstruir dados relacionados a fenômenos astronômicos cruciais, como quasares, supernovas e galáxias HII.

Os resultados obtidos por meio do método dos processos gaussianos e redes neurais revelaram a eficácia dessas técnicas na reconstrução de dados observacionais, destacando sua importância em lidar com a complexidade e o volume de informações gerados por fenômenos cósmicos. As figuras apresentadas ilustram a reconstrução bem-sucedida dos dados, fornecendo uma visão visual da precisão alcançada por essas abordagens.

Ao comparar os resultados obtidos com dados observacionais reais, notamos a consistência significativa alcançada pelo método dos processos gaussianos, particularmente nos dados do parâmetro de Hubble. No entanto, as redes neurais, embora promissoras, mostraram resultados distintos para o parâmetro de densidade de curvatura, sugerindo desafios potenciais em sua aplicação específica aos dados cosmológicos em questão.

A comparação direta entre os resultados obtidos e dados observacionais reais, conforme apresentado na Figura 7.9, destaca a capacidade robusta do método dos processos gaussianos em reproduzir vínculos observacionais. Os valores vinculados para o parâmetro de Hubble e outros parâmetros cosmológicos, apresentados na Tabela 7.1, reforçam a eficácia do método dos processos gaussianos em proporcionar resultados coerentes e comparáveis aos dados da natureza.

Embora as redes neurais tenham apresentado (para o parâmetro de curvatura) resultados divergentes na comparação direta, é essencial reconhecer o potencial dessas técnicas em aplicações cosmológicas. As redes neurais oferecem flexibilidade e versatilidade, mas os desafios associados à escolha de configurações apropriadas e à interpretação dos resultados destacam a importância de uma abordagem criteriosa ao empregar técnicas de aprendizado de máquina.

Em última análise, este estudo destaca a relevância crítica das técnicas de aprendizado de máquina, como o método dos processos gaussianos, na reconstrução de dados cosmológicos. Essas abordagens não apenas facilitam a interpretação de fenômenos as-

tronômicos complexos, mas também oferecem uma ferramenta valiosa para superar desafios associados à obtenção de dados de qualidade em escalas cósmicas. No entanto, a cautela é aconselhada ao empregar redes neurais, dada a necessidade de abordagens mais refinadas e uma compreensão profunda das configurações ideais para aplicações específicas.

Em suma, os métodos de aprendizado de máquina, como o método dos processos gaussianos, desempenham um papel crucial na extração de informações significativas a partir de observações astronômicas. Este estudo contribui para a compreensão do potencial dessas técnicas na cosmologia e ressalta a importância contínua de avanços metodológicos para enfrentar os desafios complexos inerentes à interpretação de dados cósmicos.

9 Anexos

9.1 Algoritmos

9.1.1 Código em python para a exemplificação de uma rede neural

```
1 print("""  
2 -----  
3 Trabalho de Conclusão de Curso (TCC)  
4 -----  
5 Título: Explorando a Cosmologia com Machine Learning:  
6 Redes Neurais e Processos Gaussianos na Análise  
7 de Dados Cósmicos  
8 Autor: Kevin Mota da Costa  
9 Orientador: Alan Miguel Velazques Toribio  
10 Instituição: Universidade Federal do Espírito Santo - UFES  
11 Data de Conclusão: Dezembro/2023  
12 -----  
13  
14 Este código implementa um modelo de rede neural para  
15 realizar previsões com base em dados sintéticos. O  
16 experimento é conduzido em várias etapas. Primeiro,  
17 dados de treinamento são gerados, com uma parte  
18 deles contendo erros simulados para refletir a com-  
19 -plexidade dos dados do mundo real.  
20  
21 A arquitetura da rede neural é definida, composta por  
22 camadas lineares com uma única função de ativação.  
23 O modelo é treinado usando o otimizador SGD  
24 (Stochastic Gradient Descent) e a função de perda  
25 CrossEntropyLoss. O treinamento é realizado itera-  
26 tivamente para melhorar a precisão do modelo.  
27
```

```
28 Após o treinamento, o modelo é avaliado usando dados
29 de teste aleatórios. O objetivo é criar um modelo ca-
30 paz de aprender e representar a relação complexa entre
31 os dados de entrada e as categorias 'você é chato' ou
32 'você é legal', com o intuito de fazer previsões preci-
33 sas mesmo diante de dados ruidosos e ambíguos.

34
35 O resultado obtido foi:
36
37 """
38
39
40
41 total_teste = 1000 # Número total de testes
42 train_pontos = 100 # Número de pontos de treinamento
43 train_pct_erro = 0 # Percentual de erro nos dados de treinamento
44
45 import random
46 import torch
47 import torch.nn as nn
48 import torch.optim as optim
49 import torch.nn.functional as F
50
51 # Função para gerar dados sintéticos e suas categorias correspondentes
52 def funcao(x):
53     x, y = x
54     if x == 1:
55         if y >= 20:
56             return 'você é chato'
57         else:
58             return 'você é legal'
59     elif x == 0:
60         if y <= 20:
61             return 'você é chato'
62         else:
63             return 'você é legal'
64
65 # Função para gerar erros sintéticos nos dados
66 def funcao_oposta(x):
67     x, y = x
68     if x == 1:
```

```
69     if y < 20:
70         return 'você é chato'
71     else:
72         return 'você é legal'
73 elif x == 0:
74     if y > 20:
75         return 'você é chato'
76     else:
77         return 'você é legal'
78
79 # Função para transformar dados de entrada
80 def mascara(x):
81     x, y = x
82     if x == 'sim':
83         x = 1
84     if x == 'não':
85         x = 0
86     return x, y
87
88 # Loop para diferentes percentuais de erro nos dados de treinamento
89 for train_pct_erro in (10, 20, 30, 40, 50):
90
91
92     train = []
93     # Gerando dados de treinamento com erros sintéticos
94     for i in range(int(train_pct_erro * trainPontos / 100)):
95         valor = random.choice(['sim', 'não'])
96         numero = random.randint(0, 40)
97         train.append(tuple([(valor, numero), funcao_oposta(mascara((valor,
98                           numero))))]))
99
100    # Gerando dados de treinamento sem erros
101    for i in range(int((100 - train_pct_erro) * trainPontos / 100)):
102        valor = random.choice(['sim', 'não'])
103        numero = random.randint(0, 40)
104        train.append(tuple([(valor, numero), funcao(mascara((valor,
105                           numero))))]))
106
107    # Definindo a arquitetura do modelo da rede neural
108    class MyModel(nn.Module):
109        def __init__(self):
```

```
108         super(MyModel, self).__init__()
109         self.fc1 = nn.Linear(2, 64)
110         self.fc2 = nn.Linear(64, 64)
111         self.fc3 = nn.Linear(64, 64)
112         self.fc4 = nn.Linear(64, 64)
113         self.fc5 = nn.Linear(64, 2)
114
115     def forward(self, x):
116         x = F.relu(self.fc1(x))
117         x = F.relu(self.fc2(x))
118         x = F.relu(self.fc3(x))
119         x = F.relu(self.fc4(x))
120         x = self.fc5(x)
121
122         return x
123
124
125     porc = 0
126     for i in range(10):
127         model = MyModel()
128         criterion = nn.CrossEntropyLoss()
129         optimizer = optim.SGD(model.parameters(), lr=0.01)
130
131         training_data = [(torch.tensor(mascara(input_data[0]),
132                                         dtype=torch.float32),
133                           torch.tensor(1 if input_data[1] == 'você é legal'
134                                         else 0, dtype=torch.long))
135                         for input_data in train]
136
137         epochs = 100
138         for epoch in range(epochs):
139             total_loss = 0.0
140             for inputs, labels in training_data:
141                 optimizer.zero_grad()
142                 outputs = model(inputs)
143                 loss = criterion(outputs.view(1, -1), labels.view(1))
144                 loss.backward()
145                 optimizer.step()
146                 total_loss += loss.item()
147
148         acertos = 0
149         # Avaliando o modelo com dados de teste
150         for i in range(total_teste):
```

```

147     valor = random.choice(['sim', 'não'])
148     numero = random.randint(0, 40)
149     input_example = (valor, numero)
150     input_tensor = torch.tensor(mascara(input_example),
151                                 dtype=torch.float32)
152     with torch.no_grad():
153         output = model(input_tensor)
154         predicted_class = torch.argmax(output)
155         if predicted_class == 1:
156             prediction = 'você é legal'
157         else:
158             prediction = 'você é chato'
159         if prediction == funcao(mascara(input_example)):
160             acertos += 1
161
162
163     # Calculando e imprimindo a média de acertos para 10 execuções
164     print(f'Erro de {train_pct_error:.2f}%, Acurácia média: {porc / 10:.2f}%')

```

9.1.2 Código em python para a exemplificação do método dos processos gaussianos para regressão

```

1 print("""
2 -----
3     Trabalho de Conclusão de Curso (TCC)
4 -----
5 Título: Explorando a Cosmologia com Machine Learning:
6     Redes Neurais e Processos Gaussianos na Análise
7     de Dados Cósmicos
8 Autor: Kevin Mota da Costa
9 Orientador: Alan Miguel Velazques Toribio
10 Instituição: Universidade Federal do Espírito Santo - UFES
11 Data de Conclusão: Dezembro/2023
12 -----
13
14 Esse código implementa um processo de previsão probabilística
15 utilizando Regressão Gaussiana (Gaussian Process Regression).
16
17 Inicialmente, uma função é definida e, em seguida, são gerados

```

```
18 dados de entrada aleatórios e os valores correspondentes de saída,  
19 adicionando algum ruído aos dados para simular situações do mundo real.  
20  
21 O código prossegue calculando a função de média, que representa a  
22 estimativa central para os dados, utilizando uma função exponencial.  
23 Em seguida, é definido um kernel gaussiano para calcular a covariância  
24 entre os pontos de dados. O kernel gaussiano é uma medida de similaridade  
25 entre os pontos de dados e é usado para modelar as correlações nos dados  
26 de entrada.  
27  
28 O processo de previsão inicia-se calculando a matriz de covariância entre  
29 os dados de treino e os dados de teste usando o kernel gaussiano. Em seguida,  
30 são calculados os coeficientes para a previsão com base na matriz de covariância  
31 e nos dados de treino observados.  
32  
33 O código gera múltiplas funções gaussianas para representar diferentes cenários  
34 possíveis de previsão à priori e à posteriori. Além disso, são produzidos gráficos  
35 para visualizar as funções geradas, os dados de treino, a função real e as previsões,  
36 ]incluindo as incertezas associadas.  
37 """)  
38 import numpy as np  
39 import matplotlib.pyplot as plt  
40  
41 np.random.seed(50)  
42  
43 # Definindo a função f(x)  
44 def f(x):  
45     return np.sin(x) / x  
46  
47 # Exemplo de uso da função  
48 x = np.hstack((np.random.normal(2, 1, 2), np.random.normal(-2, 1, 2)))  
49 x = np.hstack((np.random.normal(0, 1, 2), x))  
50 y = f(x)  
51 sigma_y = np.square(np.random.normal(0, np.std(y), x.shape))  
52 y = y + np.square(np.random.normal(0, np.std(y) / 4, x.shape))  
53 x_star = np.linspace(min(x), max(x), 500)  
54  
55 # Função de média  
56 def mean(x):  
57     l = 2  
58     return np.exp(-(x**2) / l)
```

```
59
60 # Plot da função de média
61 plt.plot(x_star, mean(x_star), color='red')
62 plt.title(r'Função de média para  $\ell=2$ ', fontsize=14)
63 plt.xlabel('x')
64 plt.ylabel('m(x)')
65 plt.show()
66
67 # Função que define o kernel gaussiano
68 def gaussian_kernel(x, x_prime, sigma=1.3):
69     return np.exp(-np.linalg.norm(x - x_prime)**2 / (2 * sigma**2))
70
71 # Crie uma grade de valores de x_star e y_star
72 X, Y = np.meshgrid(x_star, x)
73
74 # Calcule os valores do kernel gaussiano para todos os pontos da grade
75 Z = np.zeros_like(X)
76 for i in range(Z.shape[0]):
77     for j in range(Z.shape[1]):
78         Z[i, j] = gaussian_kernel(np.array([X[i, j], Y[i, j]]), np.array([0,
79                           0]))
80
81 # Crie o heatmap
82 plt.figure(figsize=(8, 6))
83 plt.imshow(Z, extent=(-100, 100, -100, 100), origin='lower', cmap='viridis')
84 plt.colorbar()
85 plt.xlabel(r'x')
86 plt.ylabel(r' $x'$ )
87 plt.title(r'Função de covariância para  $\sigma=2$ ', fontsize=14)
88 plt.show()
89
90 # Função que calcula a matriz de covariância entre dois vetores com base no
91 # kernel definido
92 def calculate_covariance_matrix(x1, x2, kernel_func):
93     num_x1 = len(x1)
94     num_x2 = len(x2)
95     covariance_matrix = np.zeros((num_x1, num_x2))
96     for i in range(num_x1):
97         for j in range(num_x2):
98             covariance_matrix[i, j] = kernel_func(x1[i], x2[j])
99
100    return covariance_matrix
```

```
98  
99 # Calculando a matriz de covariância  
100 m_x = np.array([mean(i) for i in x])  
101 m_star = np.array([mean(i) for i in x_star])  
102 sigma = np.diag(sigma_y)  
103 k_starstar = calculate_covariance_matrix(x_star, x_star, gaussian_kernel)  
104 k_xx = calculate_covariance_matrix(x, x, gaussian_kernel)  
105 k_x_star = calculate_covariance_matrix(x, x_star, gaussian_kernel)  
106 k_star_x = np.transpose(k_x_star)  
107  
108 # Calculando os coeficientes para a previsão  
109 arg = np.dot(k_star_x, np.linalg.inv(k_xx + sigma))  
110  
111 num_funcoes = 3  
112  
113 # Vetor de média  
114 mean = m_star  
115  
116 # Matriz de covariância  
117 cov = k_starstar  
118 std = np.diag(cov)  
119  
120 # Gerando as funções gaussianas  
121 np.random.seed(0)  
122 funcoes_gaussianas = np.random.multivariate_normal(mean, cov, num_funcoes)  
123  
124 # Plot das funções gaussianas  
125 plt.figure(figsize=(8, 6))  
126 for i in range(num_funcoes - 1):  
    plt.plot(x_star, funcoes_gaussianas[i], c='k')  
127 plt.plot(x_star, funcoes_gaussianas[i + 1], c='k', label='Funções geradas')  
128 plt.plot(x_star, mean, label='Valor médio', c='red')  
129 plt.fill_between(x_star, mean + std, mean - std, color='gray', alpha=0.5,  
    label='Erro')  
130 plt.xlabel('Previsão de x à priori')  
131 plt.ylabel('Previsão de f(x) à priori')  
132 plt.title('Distribuição à priori para a função a ser prevista')  
133 plt.legend()  
134 plt.grid(True)  
135 plt.show()  
136  
137
```

```
138 # Calculando a previsão à posteriori
139 cov_prev = k_starstar - np.dot(arg, k_x_star)
140 std_prev = np.diag(cov_prev)
141 mean_prev = m_star + np.dot(arg, (y - m_x))
142
143 # Gerando as funções gaussianas para a previsão à posteriori
144 funcoes_gaussianas = np.random.multivariate_normal(mean_prev, cov_prev,
145   ↪ num_funcoes)
146
147 # Plot das funções gaussianas para a previsão à posteriori
148 plt.figure(figsize=(8, 6))
149 for i in range(num_funcoes - 1):
150     plt.plot(x_star, funcoes_gaussianas[i], c='k')
151     plt.plot(x_star, funcoes_gaussianas[i + 1], c='k', label='Funções geradas')
152     plt.errorbar(x, y, yerr=sigma_y, fmt='o', color='b', label='Dados de treino')
153     plt.plot(x_star, mean_prev, c='red', label='Valor médio')
154     plt.fill_between(x_star, mean_prev + std_prev, mean_prev - std_prev,
155       ↪ color='gray', alpha=0.5, label='Erro')
156     plt.legend()
157     plt.xlabel('Previsão de x à posteriori')
158     plt.ylabel('Previsão de f(x) à posteriori')
159     plt.title('Distribuição à posteriori para a função a ser prevista')
160     plt.show()
161
162 # Plot dos resultados finais
163 plt.errorbar(x, y, yerr=sigma_y, fmt='o', color='b', label='Dados de treino')
164 plt.plot(x_star, [f(i) for i in x_star], label='Função real', c='k',
165   ↪ linestyle='--')
166 plt.plot(x_star, mean_prev, label='Função prevista', c='r')
167 plt.fill_between(x_star, mean_prev + std_prev, mean_prev - std_prev,
168   ↪ color='gray', alpha=0.5, label='Erro')
169 plt.legend()
170 plt.xlabel('Previsão de x')
171 plt.ylabel('Previsão de f(x)')
172 plt.title('Previsão obtida pelo processo gaussiano')
173 plt.show()
```

Bibliografia

- [1] Nils Andersson and Gregory L. Comer. Relativistic fluid dynamics: physics for many different scales. *Living Reviews in Relativity*, 24(1):3, 2021.
- [2] Joan Arnau Romeu. Derivation of friedman equations, juny 2014. Treballs Finals de Grau de Física.
- [3] Christopher M. Bishop. A brief history of gaussian processes in machine learning. *arXiv preprint arXiv:1302.4245*, 2006.
- [4] Shulei Cao, Joseph Ryan, and Bharat Ratra. Cosmological constraints from hii starburst galaxy apparent magnitude and other cosmological measurements. *MNRAS*, 497:3191–3203, 2020. arXiv:2005.12617 [astro-ph.CO] (or arXiv:2005.12617v3 [astro-ph.CO] for this version).
- [5] Celine Cattoen and Matt Visser. The hubble series: Convergence properties and redshift variables. *Classical and Quantum Gravity*, 24:5985–5998, 2007. arXiv:0710.1887 [gr-qc].
- [6] Glen Cowan. *Statistical Data Analysis*. Oxford Science Publications. Oxford University Press, USA, 1998.
- [7] Morris H. DeGroot and Mark J. Schervish. *Probability and Statistics*. Pearson, 2012.
- [8] René Descartes. *Meditações sobre a Filosofia Primeira*. 1641.
- [9] Vanessa Sardinha dos Santos. Cérebro. Brasil Escola, 2023. Acesso em 27 de setembro de 2023.
- [10] Shiv Ram Dubey, Satish Kumar Singh, and Bidyut Baran Chaudhuri. Activation functions in deep learning: A comprehensive survey and benchmark. *Neurocomputing*, 503:92–108, September 2022.
- [11] Kees Dullemond and Kasper Peeters. *Introduction to Tensor Calculus*. 2010.
- [12] A. W. F. Edwards. *Likelihood*. The Johns Hopkins University Press, 1972.

- [13] Omer Farooq, Foram Madiyar, Sara Crandall, and Bharat Ratra. Hubble parameter measurement constraints on the redshift of the deceleration-acceleration transition, dynamical dark energy, and space curvature. *The Astrophysical Journal*, 835(1):26, 2017. arXiv:1607.03537 [astro-ph.CO] (or arXiv:1607.03537v2 [astro-ph.CO] for this version).
- [14] M. Fecko. *Differential Geometry and Lie Groups for Physicists*. Cambridge University Press, 2006.
- [15] Mark E. Furman and Fred P. Gallo. *The Neurophysics of Human Behavior: Explorations at the Interface of the Brain, Mind, Behavior, and Information*. CRC Press, 1 edition, 2000.
- [16] Franz Joseph Gall. Anatomie et physiologie du système nerveux en général, et du cerveau en particulier, avec des observations sur la possibilité de reconnaître plusieurs dispositions intellectuelles et morales de l'homme et des animaux par la configuration de leurs têtes. *Journal of Phrenology*, 1:1–20, 1810.
- [17] Camillo Golgi and Santiago Ramón y Cajal. Sulla struttura della sostanza grigia del cervello. *Rivista Sperimentale di Freniatria e Medicina Legale delle Alienazioni Mentali*, 2:133–144, 1873.
- [18] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016.
- [19] Susan A. Greenfield. *The Human Brain. A Guided Tour*. Science Masters Series. Phoenix, 1st edition, 1998.
- [20] Seung gyu Hwang, Benjamin L’Huillier, Ryan E. Keeley, M. James Jee, and Arman Shafieloo. How to use gp: effects of the mean function and hyperparameter selection on gaussian process regression. *Journal of Cosmology and Astroparticle Physics*, 2023:014, February 2023. Published 7 February 2023.
- [21] Dan Hendrycks and Kevin Gimpel. Gaussian error linear units (gelus). *arXiv preprint arXiv:1606.08415*, 2016. Submitted on 27 Jun 2016 (v1), last revised 6 Jun 2023 (this version, v5).
- [22] David Jones and Dan Scolnic. Catalogs of cosmologically useful type ia supernovae from pan-starrs (“ps1cosmo”). *The Astrophysical Journal*, 857(1):51, 2018. Available at <http://dx.doi.org/10.17909/T95Q4X>.
- [23] Andrey N. Kolmogorov. *Foundations of the Theory of Probability*. Chelsea Publishing Company, 1950.

- [24] Johannes Lederer. Activation functions in artificial neural networks: A systematic overview. arXiv:2101.09957 [cs.LG], 2021.
- [25] Erich L. Lehmann. On likelihood ratio tests. *IMS Lecture Notes-Monograph Series*, 49:1–8, 2006. 2nd Lehmann Symposium - Optimality.
- [26] Otto Loewi and Henry Dale. The pharmacological transmission of nerve action. *The Journal of Pharmacology and Experimental Therapeutics*, 58:419–426, 1936.
- [27] Celi Espasandin Lopes and Elaine Meirelles. O desenvolvimento da probabilidade e da estatística: Estocástica nas séries iniciais. In *XVIII Encontro Regional de Professores de Matemática – LEM/IMECC/UNICAMP*, 2005.
- [28] E. Lusso, G. Risaliti, E. Nardini, G. Bargiacchi, M. Benetti, S. Bisogni, S. Capozziello, F. Civano, L. Eggleston, M. Elvis, G. Fabbiano, R. Gilli, A. Marconi, M. Paolillo, E. Piedipalumbo, F. Salvestrini, M. Signorini, and C. Vignali. Quasars as standard candles iii. validation of a new sample for cosmological studies. *Astronomy & Astrophysics*, 642:A150, 2020. Submitted on 19 Aug 2020.
- [29] Warren S McCulloch and Walter Pitts. A logical calculus of ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133, 1943.
- [30] N. Metropolis. The beginning of the monte carlo method. *Los Alamos Science Special Issue*, 15:125–130, 1987.
- [31] Kevin P. Murphy. *Machine Learning: A Probabilistic Perspective*. The MIT Press, 1 edition, 2012.
- [32] Vinod Nair and Geoffrey E. Hinton. Rectified linear units improve restricted boltzmann machines. *Department of Computer Science, University of Toronto*, 2010.
- [33] Michael A. Nielsen. Neural networks and deep learning, 2015. Accessed: 2023-06-19.
- [34] David Nualart. Kolmogorov and probability theory. *Arbor*, CLXXVIII(704):607–619, Agosto 2004.
- [35] Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes: A Probabilistic Perspective*. The MIT Press, Cambridge, MA, 2005.
- [36] Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, Cambridge, MA, 2006.
- [37] Robert Resnick. *Introduction to Special Relativity*. John Wiley & Sons, New York, 1st edition, 1970.

- [38] Jacques Rivoirard. The early history of the kriging method. *Mathematical Geology*, 26(3):239–262, 1994.
- [39] Christian P. Robert and George Casella. *Monte Carlo Statistical Methods*. Springer, 2nd edition, 2004.
- [40] Frank Rosenblatt. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6):386–408, 1958.
- [41] Sheldon M. Ross. *Stochastic processes*. Wiley, 2 edition, 1995.
- [42] Barbara Ryden. *Introduction to Cosmology*. Cambridge University Press, 2 edition, 2016.
- [43] D. M. Scolnic, D. O. Jones, A. Rest, Y. C. Pan, R. Chornock, R. J. Foley, M. E. Huber, R. Kessler, G. Narayan, A. G. Riess, S. Rodney, E. Berger, D. J. Brout, P. J. Challis, M. Drout, D. Finkbeiner, R. Lunnan, R. P. Kirshner, N. E. Sanders, E. Schlaflay, S. Smartt, C. W. Stubbs, J. Tonry, W. M. Wood-Vasey, M. Foley, J. Hand, E. Johnson, W. S. Burgett, K. C. Chambers, P. W. Draper, K. W. Hodapp, N. Kaiser, R. P. Kudritzki, E. A. Magnier, N. Metcalfe, F. Bresolin, E. Gall, R. Kotak, M. McCrum, and K. W. Smith. The complete light-curve sample of spectroscopically confirmed type ia supernovae from pan-starrs1 and cosmological constraints from the combined pantheon sample. *The Astrophysical Journal*, 859(2):101, 2018.
- [44] Marina Seikel, Chris Clarkson, and Mathew Smith. Reconstruction of dark energy and expansion dynamics using gaussian processes. *Journal of Cosmology and Astroparticle Physics*, 2012(06):036, 2012. arXiv:1204.2832 [astro-ph.CO].
- [45] Helaine Selin, editor. *Science Across Cultures: The History of Non-Western Science, Volume 1: Astronomy Across Cultures*. Springer, Dordrecht, Netherlands, 2000.
- [46] Sanjib Sharma. Markov chain monte carlo methods for bayesian data analysis in astronomy. *Annual Review of Astronomy and Astrophysics*, To Appear. 49 pages, draft version.
- [47] Stephen M. Stigler. *The History of Statistics: The Measurement of Uncertainty before 1900*. Belknap Press, 1986.
- [48] Y.L. Tong. *The Multivariate Normal Distribution*. Springer series in statistics. Springer-Verlag, 1 edition, 1990.
- [49] Amirhosein Toosi, Andrea Bottino, Babak Saboury, Eliot Siegel, and Arman Rahmim. A brief history of ai: How to prevent another winter (a critical review). *PET Clinics*, 16(4):449–469, October 2021.

- [50] Roberto Trotta. Bayesian methods in cosmology, 2017.
- [51] Albrecht Unsöld and Bernd Baschek. *New Cosmos: An Introduction to Astronomy and Astrophysics*. Springer, 5th edition, 2005.
- [52] Licia Verde. A practical guide to basic statistical techniques for data analysis in cosmology. 2008.
- [53] Licia Verde. *Statistical Methods in Cosmology*, volume 800, pages 147–177. 10 2010.
- [54] Matt Visser. Cosmography: Cosmology without the einstein equations. *General Relativity and Gravitation*, 37:1541–1548, 2005. arXiv:gr-qc/0411131.
- [55] Warren Weaver. Recent contributions to the mathematical theory of communication. *Mathematical Theory of Communication*, 1948.
- [56] Jiawei Zhang. Secrets of the brain: An introduction to the brain anatomical structure and biological function. 05 2019.
- [57] Xiaogang Zheng, Xuheng Ding, Marek Biesiada, Shuo Cao, and Zong-Hong Zhu. What are the omh2 (z_1, z_2) and om (z_1, z_2) diagnostics telling us in light of $h(z)$ data? *The Astrophysical Journal*, 825(1):17, jun 2016.