

Ysak Anykueury Claudino da Silva

**Estatística Bayesiana: Estudo dos Parâmetros
Cosmológicos e sua Relevância na Compreensão
do Universo**

Cuité - PB

30 de junho de 2023

Ysak Anykueury Claudino da Silva

**Estatística Bayesiana: Estudo dos Parâmetros
Cosmológicos e sua Relevância na Compreensão
do Universo**

Cuité - PB

30 de junho de 2023

S586e Silva, Ysak Anykueury Claudino da.

Estatística Bayesiana: estudo dos parâmetros cosmológicos e sua relevância na compreensão do universo. / Ysak Anykueury Claudino da Silva. - Cuité, 2023.

49 f.: il. color.

Trabalho de Conclusão de Curso (Licenciatura em Física) - Universidade Federal de Campina Grande, Centro de Educação e Saúde, 2023.

"Orientação: Prof. Joseclécio Dutra Dantas".

Referências.

1. Cosmologia. 2. Estatística Bayesiana. 3. Universo - origem. 4. Universo - estrutura. 5. observações astronômicas. 6. Dados cosmológicos. I. Dantas, Joseclécio Dutra. II. Título.

CDU 523.11(043)

Ysak Anykueury Claudino da Silva

Estatística Bayesiana: Estudo dos Parâmetros Cosmológicos e sua Relevância na Compreensão do Universo

Trabalho de Conclusão de Curso apresentado ao Curso de Licenciatura em Física como requisito para obtenção do título de Licenciado em Física da Universidade Federal de Campina Grande (UFCG), Centro de Educação e Saúde (CES) - *Campus Cuité*.

Trabalho aprovado. Cuité - PB, 30 de junho de 2023:

Joseclécio Dutra Dantas
Orientador

Nilton Ferreira Frazão
Examinador Interno

Heron Neves de Freitas
Examinador Interno

Djardiel da Silva Gomes
Examinador Externo - Suplente
Universidade Federal da Paraíba

Cuité - PB
30 de junho de 2023

Dedico este trabalho de forma especial à memória amorosa da minha querida avó. Sua presença e influência positiva em minha vida foram verdadeiramente transformadoras. Ao longo dos anos, ela sempre me incentivou a buscar o conhecimento, perseguir meus sonhos e enfrentar os desafios com coragem e determinação. Embora sua ausência seja profundamente sentida, seu legado de bondade e força continua a inspirar-me todos os dias. Dedico este trabalho a você, minha querida avó, como uma forma de honrar sua memória e expressar minha gratidão eterna. Você estará para sempre em meu coração.

Agradecimentos

Gostaria de aproveitar este espaço para expressar meus sinceros agradecimentos a todas as pessoas que contribuíram para a realização deste trabalho e para a conclusão do meu TCC.

Em primeiro lugar, gostaria de aproveitar este momento para expressar meus sinceros agradecimentos a Deus, por estar sempre presente em minha vida, me dando força, sabedoria e orientação em todas as etapas deste trabalho. Sua graça e proteção foram fundamentais para minha jornada acadêmica e conclusão deste TCC.

Gostaria de expressar minha gratidão aos meus familiares, em especial a minha mãe Ana Lúcia, aos meus avós Josuel e Jovenila (em memória), a minha maninha Janaysa e minha sobrinha Ana Clara que mesmo distantes sempre me ajudaram, aos meus tios na pessoa de José Milton (tio Dedé) e família e as minhas tias nas pessoas de Ducileide, Elizelda e Mariquinha, ao meu padrasto Riva (vulgo papai), aos meus primos na pessoa de Emanuel Claudino, as minhas primas na pessoa de Cyngrind e Esdla, não esquecendo da minha sogrinha Nena Salgados, a toda minha família EJC, e todos que estiveram ao meu lado durante toda essa jornada. Seu apoio emocional, incentivo e compreensão foram essenciais para superar os desafios e obstáculos encontrados ao longo do caminho. Vocês foram uma fonte constante de motivação e força.

Além disso, gostaria de agradecer imensamente a minha noiva Gabrielly, por todo o amor, apoio e compreensão ao longo dessa caminhada. Sua paciência, incentivo constante e apoio incondicional foram um verdadeiro suporte emocional durante os momentos de desafio. Sou grato por ter você ao meu lado, compartilhando não apenas as alegrias, mas também as dificuldades desta jornada acadêmica.

Sou grato também a todos os meus amigos do Busão Cuité, na pessoa de Vanderlei Mariano, Fernando Emanuel, ao meu bonde da fofoca kkkkkk aqui tem gente viu, Maelly, Lucila, Janinha, Weverton, Jefferson, Yasmim, Liliane, Lyvia, Janiel, Daiane, Keithy, Eliane, Fabrine; já peço perdão se esqueci de algum nome, ao meu grande amigo seminarista Emanuel e ao padre Henrique, vocês estão em meu coração por me ajudarem nessa difícil caminhada, agradeço pelo companheirismo, pelos momentos compartilhados e pelas lembranças que criamos juntos. Vocês sempre estiveram ao meu lado, oferecendo apoio, encorajamento e palavras de sabedoria, o que me ajudou a superar desafios e alcançar meus objetivos.

A minha Equipe Farmácia São Luiz (Equipe FSL), na pessoa de meus patrões Ramalho e Rayelle, que foram muito compreensíveis comigo e nunca se negaram em me ajudar com minhas necessidades para com o curso, e as minhas atuais companheiras de

trabalho Daiane, Raquel e Ana. Que nossa equipe continue a prosperar, conquistando novos desafios e alcançando resultados ainda mais impressionantes.

Gostaria de agradecer ao meu orientador, Joseclécio Dutra, pela orientação e apoio ao longo de todo o processo. Sua expertise, paciência e disponibilidade foram fundamentais para o desenvolvimento deste trabalho. Agradeço também pela confiança depositada em mim e pelas valiosas orientações que foram essenciais para a qualidade deste projeto.

Agradeço aos membros da banca examinadora por dedicarem seu tempo e conhecimento para avaliar este trabalho e fornecer contribuições valiosas para o seu aprimoramento. Suas sugestões e críticas foram de extrema importância para o enriquecimento do estudo.

Não posso deixar de agradecer aos professores no nome de Nilton Frazão (O professor inspiração aaaaauuuuuu!), Luciano Martins e Kiara Tatianny, aos servidores(as) nas pessoas de Laura, Schirlene, Jardel e seu Vital e aos colegas de curso que compartilharam seus conhecimentos e experiências ao longo da graduação, em especial Ronayde Emanuel, que está nessa batalha comigo desde o primeiro dia de curso, a Igor Guimarães que chegou no meio do caminho mas deu uma força enorme, em disciplinas e nos projetos de pesquisa, sempre me motivando a participar das coisas, Genezio Pereira, Maria Suêd, Eloiza Queiroz, Jayanne Dantas, Natanael Costa, entre outras pessoas, vocês são muito especiais para mim. Cada interação, discussão em e troca de ideias contribuíram para o meu crescimento acadêmico e pessoal.

Ao grupo do Bodão, em especial aos que sempre colaboraram comigo diariamente Damião, Isaac Bodão, Luis, Ruam e Nallyson, que nossa relação continue se fortalecendo e que possamos compartilhar muitos momentos especiais juntos.

Expresso minha profunda gratidão a todos os mencionados e a todos aqueles que, de alguma forma, contribuíram para a realização deste TCC. Vocês fizeram parte dessa conquista e sou imensamente grato por todo o apoio e incentivo.

Muito obrigado(a)!

*“Mas, os que esperam no Senhor
renovam suas forças, criam asas
como águias. Correm e não se
afadigam, podem andar que não
se cansam.”
(Bíblia Sagrada, Isaías 40, 31)*

Resumo

A análise de dados na cosmologia é uma área de estudo que busca compreender a origem, evolução e estrutura do universo por meio da análise de observações astronômicas e dados cosmológicos. Nesse contexto, esse TCC visa apresentar uma visão geral dos principais aspectos dessa análise, fornecendo uma base sólida para a compreensão dos fundamentos e técnicas utilizadas. A partir de revisões bibliográficas, discutimos métodos estatísticos empregados para essa análise de dados, incluindo técnicas de ajuste de modelos, estimação de parâmetros e inferência estatística. Fizemos um breve estudo sobre parâmetros cosmológicos, que desempenham um papel crucial na compreensão da evolução e estrutura do universo, fornecendo informações sobre sua idade, taxa de expansão e composição. Através da aplicação de técnicas estatísticas avançadas, como o MCMC, investigamos a relação entre os parâmetros cosmológicos e a estrutura do universo. Essa abordagem estatística nos permite estimar e interpretar os valores dos parâmetros, fornecendo uma base sólida para a compreensão da expansão cósmica, da composição da matéria e energia escuras, e da formação das estruturas no universo. Ao combinar a estatística bayesiana com a análise de dados cosmológicos, avançamos em nossa compreensão do universo e abrimos portas para novas descobertas sobre nossa origem e destino cósmicos.

Palavras-chaves: Cosmologia. Análise de dados. Estatística Bayesiana.

Abstract

Data analysis in cosmology is an area of study that seeks to understand the origin, evolution and structure of the universe through the analysis of astronomical observations and cosmological data. In this context, this TCC aims to present an overview of the main aspects of this analysis, providing a solid basis for understanding the fundamentals and techniques used. Based on bibliographic reviews, we discuss statistical methods used for this data analysis, including model adjustment techniques, parameter estimation and statistical inference. We did a brief study on cosmological parameters, which play a crucial role in understanding the evolution and structure of the universe, providing information about its age, expansion rate and composition. Through the application of advanced statistical techniques, such as the MCMC, we investigate the relationship between cosmological parameters and the structure of the universe. This statistical approach allows us to estimate and interpret parameter values, providing a solid foundation for understanding cosmic expansion, the composition of dark matter and energy, and the formation of structures in the universe. By combining Bayesian statistics with cosmological data analysis, we advance our understanding of the universe and open the door to new discoveries about our cosmic origin and destiny.

Key-words: Cosmology. Data analysis. Bayesian statistics.

Lista de ilustrações

Figura 1 – Telescópio Hubble. Fonte: (GARNER, 2022).	15
Figura 2 – Satélite COBE. Fonte: (GRISWOLD, 2010).	16
Figura 3 – Satélite WMAP. Fonte: (MONAGHAN, 2021).	17
Figura 4 – Telescópio Espacial Planck. Fonte: (GREICIUS, 2012).	18
Figura 5 – Telescópio James Webb. Fonte: (GARNER, 2009).	19
Figura 6 – Esquema clássico X Esquema bayesiano. Fonte: (PAULINO et al., 2018).	22
Figura 7 – Esquema interpretativo de distribuições inferencialmente relevantes. Fonte:(PAULINO et al., 2018)	28
Figura 8 – Aplicação da metodologia MCMC a um modelo linear gaussiano	34
Figura 9 – Distribuições marginais de dois parâmetros obtidos	35
Figura 10 – Linhas de Comando em Python.	39
Figura 11 – Linhas de Comando em Python.	39
Figura 12 – Linhas de Comando em Python.	40
Figura 13 – Linhas de Comando em Python.	40
Figura 14 – Linhas de Comando em Python.	41
Figura 15 – Linhas de Comando em Python.	42
Figura 16 – Linhas de Comando em Python.	42
Figura 17 – Linhas de Comando em Python.	42
Figura 18 – Linhas de Comando em Python.	43
Figura 19 – Linhas de Comando em Python.	43
Figura 20 – Linhas de Comando em Python.	43
Figura 21 – Linhas de Comando em Python.	44
Figura 22 – Linhas de Comando em Python.	44
Figura 23 – Linhas de Comando em Python.	45
Figura 24 – Linhas de Comando em Python.	45

Sumário

	Introdução	12
1	INFERÊNCIA ESTATÍSTICA	14
1.1	A importância da análise de dados na cosmologia	14
1.2	O Desafio Essencial da Estatística	20
1.3	Os paradigmas clássico e bayesiano	21
1.3.1	Paradigma clássico	21
1.3.2	Paradigma bayesiano.	22
2	ESTATÍSTICA BAYESIANA	24
2.1	Teorema de Bayes	24
2.1.1	Distribuição <i>a posteriori</i>	25
2.1.2	Distribuição <i>a priori</i>	25
2.1.3	Evidência	26
2.1.4	Verossimilhança	26
2.2	Inferência Bayesiana	27
3	ESTIMATIVA DE PARÂMETROS	29
3.1	Máxima Verossimilhança	29
3.2	Método χ^2	31
3.3	Método MCMC	33
4	A CONSTANTE DE HUBBLE E A DENSIDADE RELATIVA DE MATÉRIA	36
4.1	Equações Cosmológicas	36
4.2	Estimativa dos parâmetros H_0 e Ω_0	38
	Conclusão	46
	REFERÊNCIAS	47

Introdução

A Teoria do Big Bang representa um paradigma de grande sucesso. Este sucesso é sustentado por um número considerável de pilares observacionais confiáveis. O diagrama de Hubble, a partir do qual se verifica a expansão do Universo; a abundância de elementos leves, que concordam com as previsões da Nucleossíntese do Big Bang (NBB); a temperatura e as anisotropias da radiação cósmica de fundo (RCF), são exemplos desses pilares (RIESS et al., 2018) (DODELSON; SCHMIDT, 2021).

Em 1929, E. Hubble observou o desvio para o vermelho na luz emitida por galáxias distantes. A lei de Hubble enuncia que a velocidade de recessão das galáxias é proporcional a suas distâncias. Isto representa uma evidência observacional importante da expansão do Universo (HASHIMOTO et al., 2018). A ideia de expansão sugere a George Gamow (GAMOW, 1946), Ralph Alpher e Robert Herman (ALPHER; BETHE; GAMOW, 1948; ALPHER; HERMAN, 1950) que, em determinada fase, a radiação desacoplou da matéria e se propagou pelo Universo. Esta radiação, a RCF, foi prevista em 1940 e detectada cerca de duas décadas depois, em 1965, pelos radioastrônomos Arno Penzias e Robert Wilson (PENZIAS; WILSON, 1965). A temperatura da RCF, em torno de 2,265 K, é uma das medidas mais precisas na Cosmologia atual (ROSSMANITH, 2013; FIXSEN, 2009). A uniformidade observada na temperatura da RCF vinda de diversas direções do céu, concorda com o Princípio Cosmológico que, na Teoria do Big Bang, assume a homogeneidade e isotropia do Universo em grande escala. Essas descobertas/observações fazem da cosmologia hoje uma área bem fundamentada.

A cosmologia atual é alimentada por uma grande quantidade de dados observacionais coletados em experimentos e instrumentos que envolvem colaborações internacionais importantes. Dentre os instrumentos, podemos citar os mais conhecidos, em ordem cronológica de início de funcionamento. O *Cosmic Background Explorer* (COBE) foi lançado pela agência norte americana *National Aeronautics and Space Administration* (NASA) em 1990. As medidas foram refinadas pelo *Wilkinson Microwave Anisotropy Probe* (WMAP), lançado pela NASA em 2001. Os dados mais recentes vêm do satélite *Planck*, uma colaboração da NASA e da *European Space Agency* (ESA), lançado em 2009 (EVANS, 2015; PETTINARI, 2016).

Os dados coletados nessas observações têm contribuído para que a comunidade científica possa entender melhor o Universo, porque os dados confirmam previsões, como a homogeneidade e isotropia em grande escala, mas também alimentam novos desafios, novos problemas a serem solucionados, como é o caso do problema do horizonte, que diz respeito justamente à observação da homogeneidade e isotropia (BAUMANN; MCALISTER,

2015; DIMOPOULOS, 2021). Trabalhos recentes importantes, fundamentados nos dados da missão Planck, apresentam valores de parâmetros cosmológicos extraídos da análise do espectro de potência da RCF (DURRER, 2020; GIOVANNINI, 2008; DODELSON; SCHMIDT, 2021). Observações de supernovas do tipo Ia fornecem as medidas mais precisas de distâncias e da expansão do Universo (RIESS et al., 2018; REST et al., 2014; BETOULE et al., 2014; CAMPBELL et al., 2013; SUZUKI et al., 2012; SULLIVAN et al., 2011), apontando para a existência de um conteúdo exótico desconhecido - a energia escura -, ou para a necessidade de uma modificação da teoria da gravitação. Novas descobertas impulsionam novas ideias, novas proposições, novos desafios.

Mas esses novos dados precisam ser interpretados de maneira cuidadosa. E essa interpretação exige um tratamento estatístico que atenda às necessidades de validação de modelos a partir dos dados observacionais e do conhecimento teórico prévio. É nesse contexto que a estatística bayesiana tem contribuído: no tratamento de dados cosmológicos e refinamento de parâmetros (MARCH, 2013; PAULINO et al., 2018; ANDREON; WEAVER, 2015; HOBSON et al., 2010). Nesse sentido, este trabalho visa entender o processo de tratamento de dados cosmológicos através da estatística bayesiana.

1 Inferência Estatística

A cosmologia é a ciência que estuda a origem, a estrutura e a evolução do universo como um todo, e a análise de dados desempenha um papel crucial nessa investigação, permitindo investigar e compreender as propriedades e a evolução do universo.

1.1 A importância da análise de dados na cosmologia

Em 24 de abril de 1990, o Telescópio Hubble, que é mostrado na Figura 1, foi lançado para a órbita baixa da Terra para revolucionar a cosmologia. O telescópio Hubble é uma das ferramentas mais importantes da astronomia moderna e suas descobertas foram essenciais para a compreensão da cosmologia. Destas descobertas podemos citar as principais: Buracos negros supermassivos - essa descoberta nos permite fazer estudos sobre a relação de buracos negros e galáxias; A compreensão da natureza da matéria escura - Observou principalmente os efeitos que essa matéria causa na luz, fornecendo assim evidência da existência dessa matéria ainda não observada diretamente; Observações de galáxias distantes - detectou galáxias em formação a bilhões de anos-luz de distância, e essa descoberta é extremamente importante para a determinação da taxa de expansão do universo. Como sabemos hoje, o universo está em uma expansão acelerada. Foi possível medir essa taxa de expansão, a qual chamamos de constante de Hubble, parâmetro importante também para estimar a idade do universo.

Em 7 de março de 2021, o telescópio espacial Hubble entrou no “modo de segurança” devido a um erro de software, por volta das 4 horas da manhã. Esta foi a primeira vez que o Hubble entrou nesse modo desde 2018, quando ocorreu um incidente semelhante. Em 2018, o telescópio também entrou no modo de segurança devido a uma falha em um de seus giroscópios. É importante destacar que o modo de segurança é ativado automaticamente quando ocorrem eventos inesperados ou problemas técnicos que podem afetar a operação adequada do telescópio. Essa medida de proteção é implementada para evitar danos adicionais e permitir que os engenheiros e técnicos da missão investiguem e resolvam o problema. Após entrar no modo de segurança, a equipe responsável pelo Hubble trabalhou para resolver o erro de software e restaurar o telescópio às suas operações normais. O Hubble tem sido alvo de manutenção e atualizações ao longo dos anos para garantir seu bom funcionamento e maximizar seu tempo de vida útil. A NASA estima que somente em meados de 2030 o Hubble voltará à Terra, encerrando as atividades. (TEIXEIRA et al., 2022)

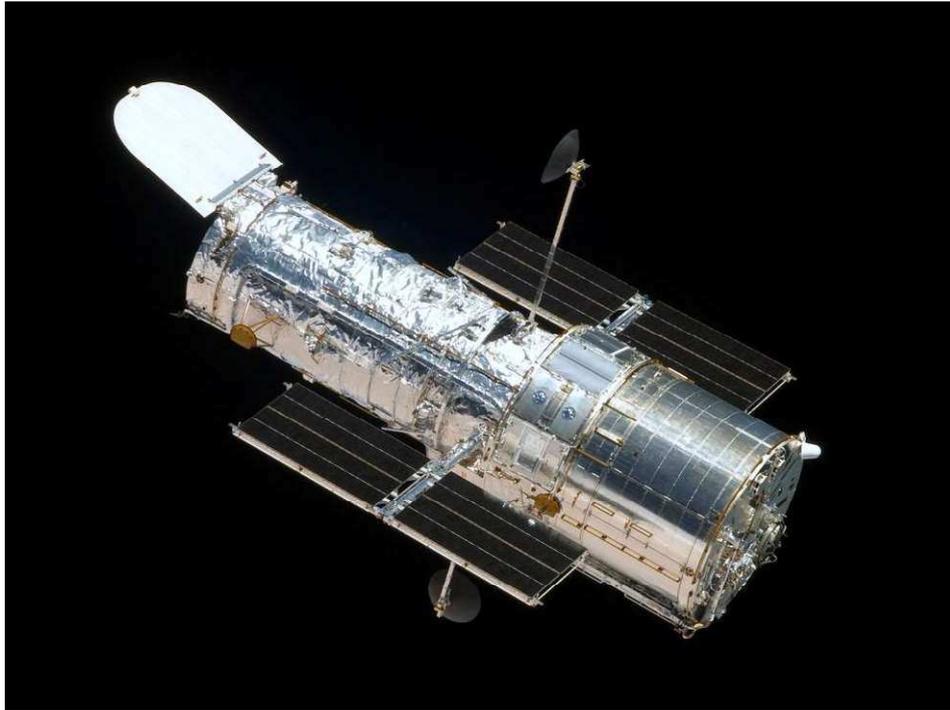


Figura 1 – Telescópio Hubble. Fonte: (GARNER, 2022).

Na Figura 2, temos o Explorer 66 que foi lançado em 1989, ou como é mais conhecido o satélite *Cosmic Background Explorer* (COBE). Esse foi o primeiro satélite dedicado à cosmologia, e foi responsável por algumas das descobertas mais importantes da cosmologia moderna. Sua principal descoberta foi a da CMB, sigla em inglês para Fundo Cósmico de Microondas (*Cosmic Background Explorer*), a qual denominados Radiação Cósmica de Fundo. O COBE analisou essa radiação remanescente de tempos primordiais da evolução. Essa descoberta teve um papel fundamental para o entendimento do universo, no que diz respeito a: fortalecimento da teoria do Big Bang; estimativa da idade do universo; observação de anisotropias na CMB; estimativa da densidade de matéria no universo. Esta última é de grande importância para que os cosmólogos estimem a taxa de expansão do universo. Ou seja, O COBE apontou evidências em favor da teoria do Big Bang e ajudou a revelar informações importantes sobre o universo primordial. O COBE encerrou oficialmente sua missão em 23 de dezembro de 1993, após operar por mais de quatro anos. Durante esse período, o satélite coletou dados cruciais sobre a radiação cósmica de fundo, que têm sido usados por pesquisadores em cosmologia para aprimorar nossa compreensão do universo em seus estágios iniciais. (MARQUES, 2012)



Figura 2 – Satélite COBE. Fonte: (GRISWOLD, 2010).

Após as descobertas do COBE, foi lançado, em 2001, o satélite *Wilkinson Microwave Anisotropy Probe* (WMAP) - ver Figura 3 -, que foi projetado para medir a radiação cósmica de fundo em micro-ondas com maior nível de detalhamento que o COBE. Com as observações realizadas pelo WMAP das anisotropias e da radiação cósmica de fundo, temos o indicativo de que só conhecemos aproximadamente 4% da matéria que é composta por bárions e léptons; o resto dessa densidade é aproximadamente 74% composta por energia escura e 22% por matéria escura. Outras descobertas importantes que o WMAP nos trouxe foram: confirmação da teoria da inflação cósmica e detecção de flutuações de temperatura na CMB. Teorias que explicam a CMB ser uniforme em todas as direções, o universo parecer tão plano, e detectou também flutuações de temperatura na CMB, que são justamente essas flutuações que deram origem às galáxias. A missão do WMAP foi extremamente bem-sucedida, fornecendo um dos mapas mais detalhados e precisos da radiação cósmica de fundo. Ele operou por mais de nove anos e foi desligado em 20 de agosto de 2010, encerrando oficialmente sua missão. Os dados coletados pelo WMAP têm sido fundamentais para pesquisas em cosmologia e continuam sendo utilizados até os dias de hoje para aprimorar nosso conhecimento sobre a história e a natureza do universo. (VIELVA et al., 2004)

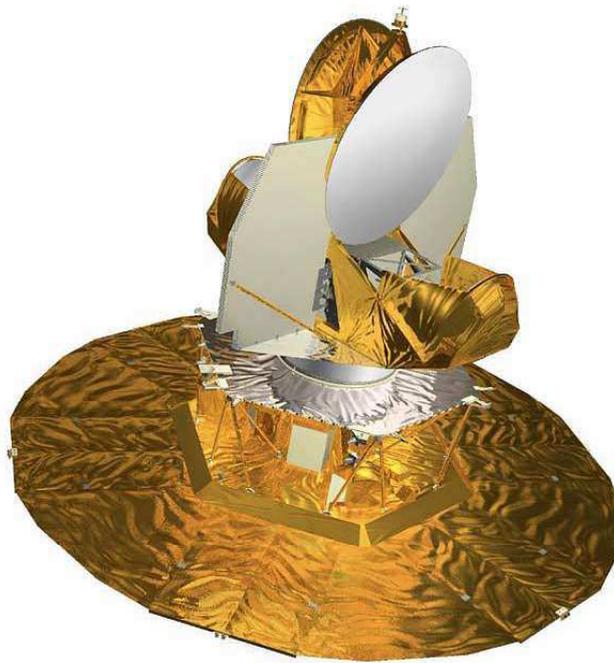


Figura 3 – Satélite WMAP. Fonte: (MONAGHAN, 2021).

A missão espacial Planck foi uma missão da Agência Espacial Europeia (ESA) dedicada ao estudo da radiação cósmica de fundo em micro-ondas (CMB). O Telescópio Espacial Planck, que é mostrado na Figura 4, foi lançado em maio de 2009, teve como objetivo mapear com alta precisão a radiação remanescente do Big Bang, que preenche todo o universo observável. A missão Planck utilizou um telescópio espacial equipado com instrumentos altamente sensíveis, capazes de medir a radiação cósmica de fundo em micro-ondas com extrema precisão. O objetivo principal da missão era mapear as flutuações na temperatura dessa radiação em todo o céu, com uma resolução sem precedentes. Os dados coletados pela missão Planck foram fundamentais para aprimorar nosso conhecimento sobre os principais parâmetros cosmológicos, como a taxa de expansão do universo, a densidade de matéria e energia escuras, a idade do universo e a distribuição das flutuações primordiais. A missão espacial Planck foi uma importante empreitada científica que nos forneceu dados detalhados e precisos sobre a radiação cósmica de fundo em micro-ondas, possibilitando avanços significativos na compreensão da cosmologia e do universo em que vivemos. (MELO, 2022)

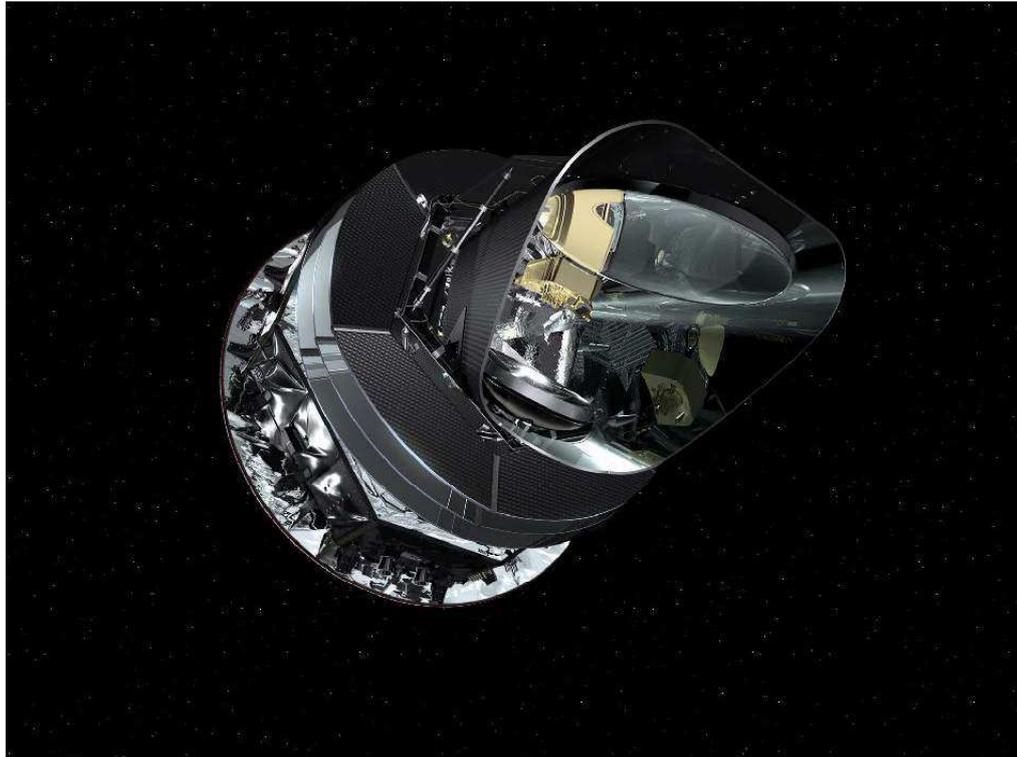


Figura 4 – Telescópio Espacial Planck. Fonte: (GREICIUS, 2012).

Em uma missão planejada pela NASA em conjunto com outras agências internacionais, foi desenvolvido o Telescópio Espacial James Webb (JWST), que é mostrado na Figura 5, com o objetivo de aprofundar ainda mais nossa compreensão do universo. Ele observa o universo em comprimentos de onda infravermelha, permitindo que ele revele coisas que o Telescópio Hubble não pode. Após vários testes foi possível ocorrer o lançamento em 25 de dezembro de 2021. As expectativas de novas descobertas e explicações de alguns problemas sobre o nosso cosmos são gigantes. (COELHO, 2022)

Em seus primeiros anos já esperamos ele cumprir atividades como: observações de galáxias primordiais - ele pode ser capaz de observar as primeiras galáxias do universo, transferindo informações sobre a formação das estrelas e galáxias após o Big Bang; estudo de exoplanetas - irá examinar a atmosfera de exoplanetas, trazendo informações sobre sua composição e condições climáticas; estudo de estrelas e formações estelares - o JWST estudará estrelas e suas formações, nos permitindo entender melhor como as estrelas se formam e evoluem. Também esperamos estudar buracos negros e matéria escura, pois o Telescópio Espacial pode detectar a luz infravermelha emitida por buracos negros e matéria escura, permitindo aos cosmólogos estudá-los em mais detalhes. As descobertas do Telescópio Espacial James Webb têm o potencial de fornecer informações inovadoras sobre muitos aspectos do universo. (COELHO, 2022)

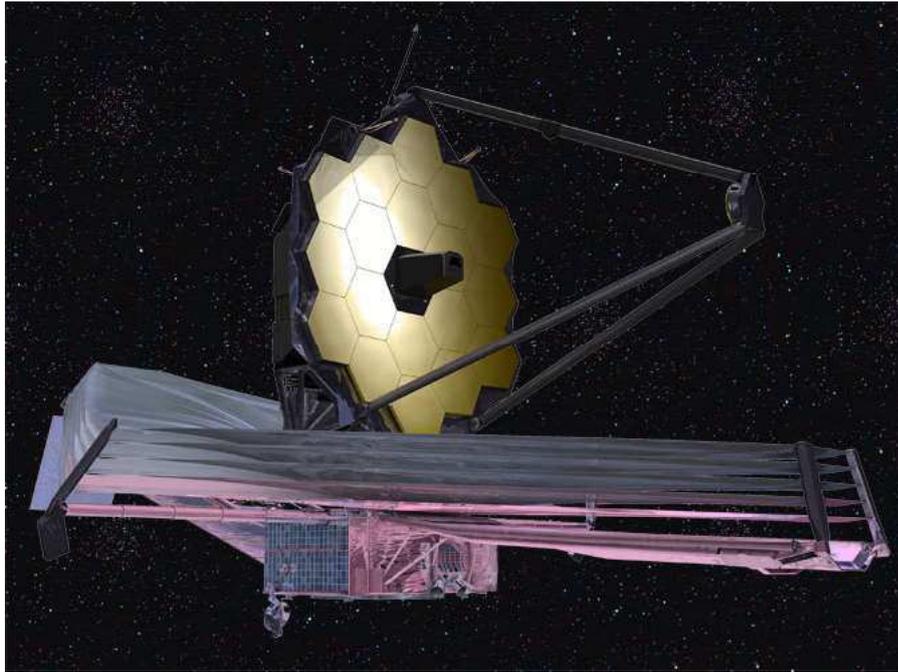


Figura 5 – Telescópio James Webb. Fonte: (GARNER, 2009).

A cosmologia é o estudo do universo como um todo, incluindo sua origem, evolução e estrutura. A análise de dados é crucial na cosmologia para entender as propriedades e comportamentos do universo, através da observação de fenômenos como a radiação cósmica de fundo, as lentes gravitacionais, pois o universo é muito vasto e complexo.

Na cosmologia precisamos dessas análises. Para interpretar observações usamos técnicas avançadas da estatística para sintetizar informações dos dados, como estimativas de parâmetros cosmológicos, como a idade do universo, a taxa de expansão, a densidade de matéria e energia escura, como radiação eletromagnética, ondas gravitacionais e partículas cósmicas. A análise de dados também é importante para testar e refinar as teorias na cosmologia e distinguir entre diferentes modelos cosmológicos, a fim de entender melhor a natureza e a evolução do universo.

A análise de dados na cosmologia é uma área em constante evolução, surgindo cada vez mais novas técnicas de observação e instrumentação. Na cosmologia tem contribuído significativamente para avançar nosso conhecimento sobre a origem e evolução do universo, bem como testar e refinar as teorias cosmológicas existentes.

A análise de dados é fundamental na cosmologia, porque a cosmologia é uma ciência observacional. A partir dessa análise podemos extrair valiosas informações sobre o universo. Diante de todas essas perspectivas, faz-se necessário fazer um estudo estatístico de dados. A estatística é uma ferramenta muito importante para analisarmos os dados obtidos por esses instrumentos cosmológicos, pois ela ajuda a lidar com a incerteza dos dados, facilitando a construção de modelos mais precisos e a obtenção de estimativas

confiáveis dos parâmetros cosmológicos.

1.2 O Desafio Essencial da Estatística

Estatística é um campo do estudo centrado na produção de metodologia para coleta, organização, descrição, análise e interpretação de dados bem como na obtenção de conclusões válidas e na tomada de decisões razoáveis baseadas em tais análises (SILVA; FERNANDES; ALMEIDA, 2015).

Para O'Hagan (1994), “O problema fundamental para qual o estudo da estatística se dirige é o da inferência. Alguns dados são observados e desejamos fazer afirmações, inferências, sobre uma ou mais características desconhecidas do sistema físico que deu origem a esses dados.”

Podemos afirmar que a *inferência* estatística, por sua vez, é a parte da metodologia da Ciência que tem por objetivo a coleta, redução, análise e modelagem dos dados, a partir do que, finalmente, faz-se a inferência para uma população da qual os dados (a amostra) foram obtidos. Um aspecto importante da modelagem dos dados é fazer previsões, a partir das quais se pode tomar decisões (MORETTIN, 2017).

Podemos dizer que, depois de proceder a uma análise de observações passadas, o propósito do estatístico é fazer inferências ou predições acerca de novos fenômenos da mesma natureza. É importante descrever alguns ramos/termos da estatística - o que será feito a seguir. A *Estatística Descritiva* é uma área da estatística que utiliza diversas técnicas para descrever e resumir um conjunto de dados. A *Inferência Estatística* é um ramo da estatística que visa fazer inferências e tirar conclusões sobre uma população com base em um conjunto de dados representativo, conhecido como amostra. A *Estatística Matemática* utiliza a teoria da probabilidade na análise estatística. Ao contrário das técnicas de coleta de dados estatísticos, a estatística matemática se concentra no desenvolvimento e na aplicação de métodos probabilísticos para análise, modelagem e interpretação de dados. A *Estatística Computacional*, também conhecida como *Computação Estatística*, é um campo que abrange a interseção entre a estatística e a ciência da computação. É uma área especializada que combina os princípios da estatística com as ferramentas e técnicas da ciência da computação.

A Estatística Bayesiana é uma abordagem para a análise de dados que permite a incorporação de informações prévias. Tem sido amplamente utilizada para inferência em modelos cosmológicos, como a determinação de parâmetros cosmológicos. Uma das principais vantagens dessa abordagem é trazer a combinação de informações prévias e dados observados. É uma ferramenta poderosa para a análise de dados na cosmologia.

1.3 Os paradigmas clássico e bayesiano

Na primeira metade deste século, a inferência clássica teve um importante papel com seus precursores: Karl Pearson¹, Ronald A. Fisher² e Jerzy Neyman³. O principal objetivo dessa inferência é determinar que generalizações sobre a população se podem fazer a partir da amostra que ela foi recolhida.

Os procedimentos clássicos são baseados em uma interpretação frequentista da probabilidade. Esse é o princípio adotado para que tenhamos a interpretação inferencial. Além disso, uma das limitações dessa teoria é não levar em conta o conhecimento a priori sobre os parâmetros ou o modelo. Temos dois pensamentos importantes a respeito dessas teorias estatísticas. Segundo Jaynes (1996),

Os métodos frequentistas tradicionais que usam apenas distribuições de amostragem são utilizáveis e úteis em muitos problemas particularmente simples e idealizados; mas eles representam os casos especiais mais proscritos da teoria da probabilidade, porque pressupõem condições que dificilmente são atendidas em problemas reais. Essa abordagem é bastante inadequada para as necessidades atuais da ciência (JAYNES, 1996).

Logo, faz-se necessário estabelecer uma distinção importante entre o paradigma *clássico* e a abordagem que será discutida neste trabalho, o paradigma *bayesiano*.

1.3.1 Paradigma clássico

O que fundamenta esse paradigma é o entendimento de que o principal objetivo da inferência estatística é a determinação de generalizações a respeito da população a partir da amostra que ela foi recolhida. A interpretação clássica define a probabilidade de um acontecimento como o quociente entre o número de casos favoráveis e o número de casos possíveis, supondo todos os casos igualmente possíveis. A principal limitação da interpretação frequentista é só poder aplicar-se em situações que possam repetir-se indefinidamente nas mesmas circunstâncias. (PAULINO et al., 2018)

¹ Karl Pearson foi um estatístico britânico que fez importantes contribuições para o desenvolvimento da estatística e da teoria das probabilidades.

² Ronald A. Fisher, conhecido como R.A. Fisher, foi um estatístico e biólogo britânico que fez contribuições significativas para a estatística e a genética. Ele é considerado um dos principais estatísticos do século XX.

³ Jerzy Neyman foi um estatístico polonês-americano que desempenhou um papel fundamental no desenvolvimento da teoria estatística moderna. Ele é conhecido por suas contribuições para o campo da inferência estatística, em particular, o conceito de teste de hipóteses.

1.3.2 Paradigma bayesiano.

Diferentemente da concepção anterior, a estatística bayesiana defende a ideia de uma probabilidade como grau de credibilidade. A atitude de um indivíduo ao apostar na veracidade de uma proposição está associada com o grau de credibilidade que esse indivíduo lhe atribui.

Ambas tem diferenças fundamentais na maneira como lidam com a incerteza e a inferência. As principais contraposições entre os dois paradigmas são: interpretação da probabilidade; testes de hipóteses e atualização bayesiana; estimativas pontuais e distribuições de probabilidade.

Uma rápida comparação entre os procedimentos clássicos e procedimentos bayesianos pode ser vista na Figura 6. Podemos observar a importância do Teorema de Bayes no estudo da cosmologia, pois ele é amplamente utilizado para a inferência de parâmetros cosmológicos a partir de observações, levando em conta informações a priori e atualizando probabilidades a posteriori. No próximo capítulo faremos a dedução detalhada do Teorema de Bayes.

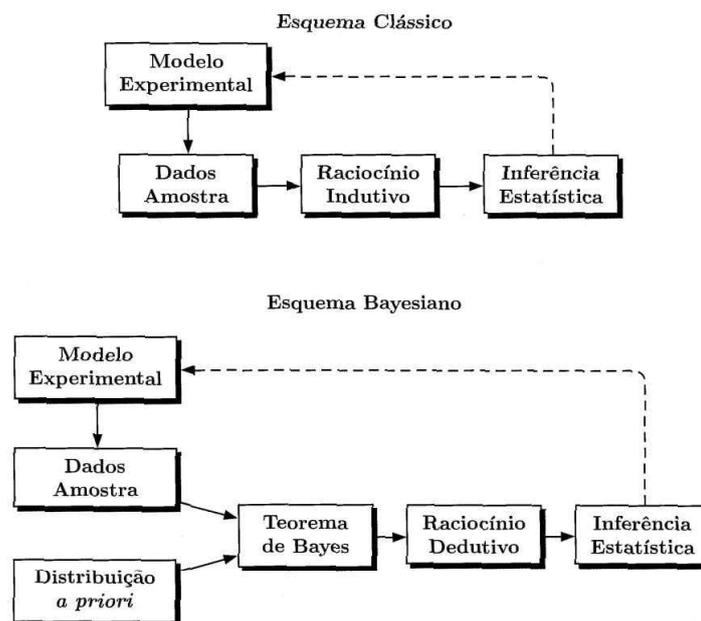


Figura 6 – Esquema clássico X Esquema bayesiano. Fonte: (PAULINO et al., 2018).

O esquema clássico e o esquema bayesiano são abordagens diferentes na análise estatística. Vamos destacar as principais diferenças entre eles como é mostrado na figura 6, iniciando pela informação prévia, que irá passar pelo teorema de Bayes, uma grande diferença também existe em seus raciocínios pois o raciocínio indutivo se baseia em evidências específicas para tirar conclusões gerais, enquanto o raciocínio dedutivo parte de premissas gerais para chegar a conclusões específicas. Essas são algumas das principais

diferenças entre o esquema clássico e o esquema bayesiano. Cada abordagem possui suas vantagens e desvantagens, e a escolha entre elas depende do contexto do problema, das informações disponíveis e das preferências do pesquisador.

2 Estatística Bayesiana

A Estatística Bayesiana é uma abordagem estatística que se baseia no Teorema de Bayes e na Inferência Bayesiana. Essa metodologia é amplamente utilizada para modelar e analisar dados, incorporando informações prévias e atualizando-as com base nas observações.

2.1 Teorema de Bayes

Thomas Bayes (1701-1761) foi um matemático, estatístico e clérigo britânico, conhecido por ter desenvolvido o Teorema de Bayes. Ele nasceu em Londres e estudou teologia no St. John's College, em Cambridge. O seu trabalho foi publicado só em 1763, após sua morte e deu origem ao nome estatísticas Bayesianas.

Em muitas situações práticas, é comum que o fenômeno aleatório com o qual estamos lidando possa ser decomposto em várias etapas distintas. Nesse contexto, a informação sobre o que ocorreu em uma etapa específica pode ter influência sobre as probabilidades de ocorrência das etapas subsequentes. Nesses casos, quando adquirimos informações adicionais, podemos atualizar as probabilidades de interesse por meio de recálculos. Essas probabilidades atualizadas são conhecidas como probabilidades condicionais.

Suponha que tenhamos dois eventos A e B , a probabilidade condicional de A dado que B ocorreu é expressa por $P(A|B)$ e pode ser escrita na forma:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad P(B) > 0. \quad (2.1)$$

Dado um outro acontecimento B qualquer, com $P(B) > 0$, a decomposição de B na união de conjuntos disjuntos é da forma:

$$B = \cup_i (A_i \cap B). \quad (2.2)$$

A função P é aditiva e atende à definição de probabilidade condicionada. Isso nos permite escrever:

$$P(B) = \sum_i P(A_i \cap B) = \sum_i P(B|A_i)P(A_i), \quad (2.3)$$

de onde observamos que:

$$P(A_i \cap B) = P(B|A_i)P(A_i) = P(A_i|B)P(B), \quad (2.4)$$

Daqui obtemos o *Teorema de Bayes*:

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)} = \frac{P(B|A_i)P(A_i)}{\sum_i P(B|A_i)P(A_i)}. \quad (2.5)$$

Podemos interpretar essa esta expressão da seguinte forma:

1. “Sejam $A_i = 1, 2, \dots, n$ antecedentes, causas, hipóteses ou estados a que o investigador atribui graus de credibilidade. A estes parâmetros associamos as probabilidades *a priori* $P(A_i)$, $i = 1, 2, \dots, n$. Esta representa uma distribuição de probabilidades que tem natureza subjetiva. É, por isso, denominada probabilidade *a priori*.” (PAULINO et al., 2018)
2. “Depois da informação adicional - o que significa saber de um acontecimento B (ex.: observação dos dados), o investigador revê suas probabilidades *a priori* através da *fórmula de Bayes* e passa a atribuir aos A_i , $i = 1, 2, \dots, n$ as probabilidades *a posteriori* $P(A_i|B)$, $i = 1, 2, \dots, m$.” (PAULINO et al., 2018)

Quando o investigador está na completa ignorância em relação aos A_i , ou seja, desconhece qualquer informação prévia, a proposta de Laplace (princípio da razão insuficiente ou critério de Bayes-Laplace) consiste em atribuir igual probabilidade aos A_i , ou seja, $P(A_i) = 1/m$, a que denominamos distribuição uniforme ou plana. Nesta situação, a equação (2.5) toma a forma:

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_i P(B|A_i)P(A_i)} = \frac{P(B|A_i)P(A_i)}{P(A_i) \sum_i P(B|A_i)},$$

resultando em

$$P(A_i|B) = \frac{P(B|A_i)}{\sum_i P(B|A_i)}. \quad (2.6)$$

A seguir, consideramos importante definir cada um dos termos presentes na *fórmula de Bayes*, a equação (2.5) completa.

2.1.1 Distribuição *a posteriori*

$P(A_i|B)$ é a distribuição *a posteriori*: ela é uma distribuição de probabilidade do parâmetro de interesse, ao analisar os dados observados e a combinação da distribuição *a priori* com a função de verossimilhança dos dados, ou seja, é a probabilidade condicional do evento A_i ocorrer, dado que o evento B já ocorreu.

2.1.2 Distribuição *a priori*

$P(A_i)$ é a distribuição *a priori*: essa é a distribuição de probabilidade que exprime o conhecimento prévio sobre um parâmetro ou conjunto de parâmetros de um modelo estatístico antes mesmo de observar algum dado, ou seja, é a probabilidade a do evento A_i ocorrer, antes de se observar o evento B .

2.1.3 Evidência

$\sum_i P(B|A_i)$ é a marginalização: é o processo usado para a obtenção da distribuição de probabilidade de uma variável, após a integração das distribuições de probabilidade de outras variáveis. Por exemplo, na cosmologia ela é usada na obtenção da distribuição de probabilidade dos parâmetros cosmológicos.

2.1.4 Verossimilhança

$P(B|A_i)$ é a função de verossimilhança: essa função descreve a probabilidade dos dados observados, ou seja, a função de probabilidade é uma função dos parâmetros de um modelo estatístico que permite inferir sobre o seu valor a partir de um conjunto de observações. Temos dois casos que iremos tratar separadamente agora: casos discretos e casos contínuos.

Caso discreto

Se as variáveis aleatórias $X_i, i = 1, 2, \dots, n$ são i.i.d. (independentes e identicamente distribuídas) com função de probabilidade $f(x|\theta)$, $A = (X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$ é um acontecimento com probabilidade $\prod_i f(x_i|\theta)$. Se fixarmos A e fazendo variar θ em Θ , obtemos a função de verossimilhança:

$$L(\theta) = L(\theta|A) = \prod_i f(x_i|\theta). \quad (2.7)$$

Cada θ exprime a verossimilhança ou a plausibilidade que lhe é atribuída quando se sabe que o acontecimento A se realizou.

A verossimilhança não é uma probabilidade, com isso, somente a razão de verossimilhanças tem significado.

$$\frac{L(\theta)}{L(\theta^*)} = \frac{\prod_i f(x_i|\theta)}{\prod_i f(x_i|\theta^*)}. \quad (2.8)$$

Mede o peso da evidência ou plausibilidade de θ contra θ^* decorrente da observação da amostra. Por esta razão, a função de verossimilhança é definida a menos de um fator constante (independente de θ) positivo,

$$L(\theta) = K \prod_i f(x_i|\theta), \quad \theta \in \Theta. \quad (2.9)$$

Caso contínuo

Se as variáveis aleatórias $X_i, i = 1, 2, \dots, n$ são i.i.d. (independentes e identicamente distribuídas) com função densidade de probabilidade $f(x|\theta)$. Para preservarmos a transição de probabilidade a verossimilhança, é necessário determinar um limite aproximado com acréscimos dx_i , arbitrariamente pequenos:

$$A^* = (x_1 < X_1 < x_1 + dx_1, x_2 < X_2 < x_2 + dx_2, \dots, x_n < X_n < x_n + dx_n), \quad (2.10)$$

É um acontecimento que tem probabilidade elementar, logo:

$$L(\theta) = \prod_i f(x_i|\theta)dx_i = \left[\prod_i dx_i \right] \prod_i f(x_i|\theta). \quad (2.11)$$

Como o termo entre parênteses pode absorver-se na constante, pois não depende de θ , temos:

$$L(\theta) = K \prod_i f(x_i|\theta). \quad (2.12)$$

Podemos escrever o Teorema de Bayes para densidades:

$$h(\theta|x) = \frac{f(x|\theta)h(\theta)}{\int_{\Theta} f(x|\theta)h(\theta)d\theta}, \quad \theta \in \Theta. \quad (2.13)$$

Como no denominador não depende de θ a expressão pode ser escrita:

$$h(\theta|x) \propto f(x|\theta)h(\theta), \quad \theta \in \Theta. \quad (2.14)$$

A função de verossimilhança tem importante papel na fórmula de Bayes. Ela representa o meio através do qual os dados x transformam o conhecimento a priori sobre θ . Dito de outra forma, a função de verossimilhança mede a probabilidade de o modelo estudado reproduzir os dados. Neste sentido, esta função relaciona o conhecimento *a priori* com o conhecimento *a posteriori*.

2.2 Inferência Bayesiana

De acordo com [Paulino et al. \(2018\)](#), podemos distinguir dois objetivos relacionados aos procedimentos Bayesianos:

I – “Realizar inferências sobre o parâmetro não observável θ . As inferências clássicas são baseadas em probabilidades associadas com as diferentes amostras, x , que poderiam ocorrer para algum valor fixo, mas desconhecido, do parâmetro θ ” ([PAULINO et al., 2018](#)).

II – “Muitos Bayesianos consideram que a inferência não tem de se restringir a proposições sobre parâmetros não observáveis. Se as inferências paramétricas possuem inconvenientes na medida em que os valores dos parâmetros poucas vezes são conhecidos e, portanto, as conclusões a que tais inferências conduzem raramente podem ser confrontadas pela realidade” ([PAULINO et al., 2018](#)).

As inferências Bayesianas dependem de probabilidades subjetivas ou crenças *a posteriori* que estão associadas a diferentes valores do parâmetro θ , levando em conta o valor específico de x observado. O valor de x é mantido fixo, enquanto a variação de θ é considerada.

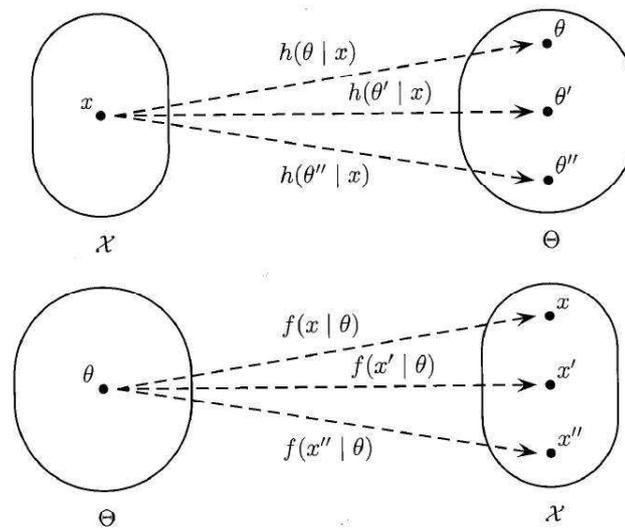


Figura 7 – Esquema interpretativo de distribuições inferencialmente relevantes.
 Fonte:(PAULINO et al., 2018)

Na figura 7, temos um esquema interpretativo de distribuições inferencialmente relevantes que refere-se a uma abordagem que permite a interpretação e análise das distribuições de probabilidade no contexto da inferência estatística. Essas distribuições são importantes porque fornecem informações sobre os parâmetros ou características desconhecidas de uma população com base nos dados observados.

Para O'Hagan (1994), enquanto os Bayesianos podem emitir enunciados probabilísticos sobre os parâmetros, que consideram como variáveis aleatórias, isso não é possível com os clássicos. Consequentemente as probabilidades para os clássicos dizem respeito aos dados e não ao parâmetro, embora depois sejam reformuladas para que aparentemente digam respeito ao parâmetro. A função de verossimilhança desempenha um papel crucial ao transmitir a informação contida na amostra.

3 Estimativa de parâmetros

Este capítulo tem como objetivo discutir algumas formas de estimar parâmetros em modelos estatísticos. A estimativa de parâmetros é um dos principais objetivos da inferência estatística, e é fundamental para a tomada de decisões baseadas em dados. Aqui serão apresentados alguns dos métodos mais comuns de estimar parâmetros, incluindo a estimativa por Máxima Verossimilhança, Método χ^2 e a Técnica MCMC. Além disso, serão discutidos alguns aspectos importantes relacionados à qualidade das estimativas.

3.1 Máxima Verossimilhança

Por conveniência, vamos reescrever o Teorema de Bayes na forma de probabilidade condicional, relacionando a probabilidade *a posteriori*, a probabilidade *a priori* e a verossimilhança dos dados observados:

$$L(\theta_j; x_i) = \frac{f(x_i; \theta_j)p(\theta_j)}{g(x_i)}, \quad (3.1)$$

onde $p(\theta_j)$ é chamado de probabilidade *a priori* para os parâmetros θ_j que estamos estudando, e x_i representa os dados de que se tem conhecimento, enquanto $g(x_i)$ é a função de densidade de probabilidade que fornece a probabilidade relativa de que a variável aleatória assumira valores em um determinado intervalo. A função final $L(\theta_j; x_i)$ é denominada *a posteriori*, $f(x_i; \theta_j)$ é a verossimilhança e geralmente é representada por L . *A posteriori* possui as informações que estamos procurando: nos fornecem informações sobre a incerteza associada a eventos ou parâmetros desconhecidos, com base em dados observados e conhecimentos prévios.

Como $L(\theta_j; x_i)$ é uma função distribuição de probabilidades para θ_j , ela deve ser normalizada à unidade. Neste caso, devemos ter:

$$\int L(\theta_j; x_i) d^n \theta_j = 1 = \frac{\int f(x_i; \theta_j)p(\theta_j) d^n \theta_j}{g(x_i)}. \quad (3.2)$$

Daí resulta que:

$$\int f(x_i; \theta_j)p(\theta_j) d^n \theta_j = g(x_i). \quad (3.3)$$

A integral do lado esquerdo de (3.3) é conhecida como *evidência* e, às vezes, a mesma denominação é dada a $g(x_i)$. A função $g(x_i)$ não depende dos parâmetros θ_i , portanto, não ajuda a estimá-los e, do ponto de vista de $L(\theta_j)$, é apenas um fator de normalização. *A priori* $p(\theta_j)$, por sua vez, geralmente é desconhecida. Porém, muitas das vezes podemos obter informações dos dados disponíveis, como os resultados de experimentos anteriores.

Conhecendo a posteriori $L(\theta_j)$, precisamos determinar o máximo dessa função, para obter os estimadores de verossimilhança máxima $\hat{\theta}_i$, ou seja, os valores θ_i que localizam o máximo da função de distribuição a posteriori. Para isso, fazemos:

$$\frac{\partial L(\theta_i)}{\partial \theta_i} = 0, \quad i = 1, \dots, n. \quad (3.4)$$

Normalmente o denominador $g(x_i)$ é descartado, porque é independente do parâmetro θ . Neste caso, a normalização de $L(\theta_j)$ precisa ser recalculada. A constante de normalização é:

$$N = \int L(\theta_j) d^n \theta_j, \quad (3.5)$$

de forma que $\int (L(\theta_j)/N) d^n \theta_j = 1$. Tendo a posteriori normalizada $L(\theta_j)/N$, ou também como é chamada a verossimilhança podemos derivar as regiões de confiança. As regiões $R(\alpha)$ são definidas como as áreas delimitadas pela constante $L(\theta_i)$, para as quais:

$$\int_{R(\alpha)} L(\theta_i) d^n \theta = \alpha. \quad (3.6)$$

$R(\alpha)$ é a região para a qual a integral acima é calculada com $0 < \alpha < 1$. Para encontrar R , estima-se:

$$\int \hat{L}(L_i) d^n \theta = \alpha_i. \quad (3.7)$$

Se $L > L_i$, temos $\hat{L}(L_i) = L$ e 0 em qualquer outro lugar diferente disso. Por tentativas e erros, chegamos aos α_i mais utilizados. Na cosmologia, os valores preferidos são: $\alpha = 0,683$; $0,954$; $0,997$. O valor L_i que corresponde a α_i é a divisão que fazemos em L para encontrar a região $R(\alpha_i)$. Quando esse valor representa o estimador de máxima verossimilhança, é dito que a verossimilhança é maximizada em relação a esse parâmetro, em vez de ser marginalizada.

Como essa é a primeira vez que estamos lidando com métodos de máxima verossimilhança, vamos usar um exemplo simples, como a verossimilhança gaussiana $f(x_i; \mu, \sigma^2)$, dada por:

$$f(x_i; \mu, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp \left[-\frac{1}{2} \sum_i^n \frac{(x_i - \mu)^2}{\sigma^2} \right], \quad (3.8)$$

em que x_i é o parâmetro que estamos estimando, μ é o valor médio conhecido desse parâmetro e σ é o desvio padrão. Fazendo uso da equação (3.4), podemos determinar o estimador de máxima verossimilhança de x_i , ou seja, o valor $\hat{\mu}$ de x_i que localiza o máximo de f . Temos que:

$$\frac{\partial f}{\partial x_i} = (2\pi\sigma^2)^{-n/2} \exp \left[-\frac{1}{2} \sum_i^n \frac{(x_i - \mu)^2}{\sigma^2} \right] \left[-\frac{1}{\sigma^2} \sum_i^n (x_i - \mu) \right] = 0.$$

Isto leva a:

$$n\hat{\mu} - \sum_i^n x_i = 0.$$

Logo o estimador de máxima verossimilhança de μ é dado por:

$$\hat{\mu} = \frac{1}{n} \sum_i^n x_i. \quad (3.9)$$

Assim, analogamente temos que o estimador de variância de máxima verossimilhança, que é um método utilizado para estimar a variância de uma distribuição de probabilidade com base nos dados observados. Ele busca encontrar o valor da variância que maximiza a função de verossimilhança, que é uma medida da probabilidade dos dados observados ocorrerem sob determinados parâmetros:

$$\sigma^2 = \frac{1}{n} \sum_i^n (x_i - \mu)^2. \quad (3.10)$$

Note que esse resultado difere do resultado padrão, segundo o qual o estimador de variância amostral tem $(n - 1)$ no denominador em vez de n . Nesse caso, o estimador de máxima verossimilhança é enviesado, o que significa que o seu valor esperado não é igual ao verdadeiro valor. De fato, os estimadores de máxima verossimilhança não são necessariamente imparciais, embora sob algumas condições gerais, eles sejam assintoticamente imparciais, ou seja, para n tendendo a infinito.

3.2 Método χ^2

Para realizar o ajuste de parâmetros usando a estatística bayesiana, o objetivo é obter a configuração paramétrica de maior probabilidade dada as observações dos dados. Isso implica em maximizar a expressão $P(\theta|\{Y_i\})$. Em um cenário geral, assumimos que o conjunto de N dados medidos, A , segue uma distribuição de probabilidade multivariada normal, o que resulta na função de verossimilhança.

$$L(\vec{\theta}) = \frac{\exp\left(-\frac{1}{2}\chi_A^2(\vec{\theta})\right)}{\left(\frac{2\pi}{N}\right)^{\frac{1}{2}} |\det C|^{\frac{1}{2}}}. \quad (3.11)$$

A função χ^2 é definida como:

$$\chi_A^2(\vec{\theta}) = \sum_{i=1}^N \sum_{j=1}^N \left(y_{\text{teo}}(z_i, \vec{\theta}) - y_{\text{obs},i}\right) C_{ij}^{-1} \left(y_{\text{teo}}(z_j, \vec{\theta}) - y_{\text{obs},j}\right), \quad (3.12)$$

onde $y_{\text{obs},i}$ é a grandeza y observada no *redshift* $y_{\text{teo}}(z_i, \vec{\theta})$ é a expressão funcional do modelo teórico aplicado ao i -ésimo valor de *redshift*, z_i e C_{ij}^{-1} são os elementos da matriz de covariância entre os valores de $y_{\text{obs},i}$.

Em casos em que os dados são não correlacionados, a matriz de covariância é diagonal, com elementos $C_{ii}^{-1} = \frac{1}{\sigma_i^2}$, onde σ_i é o desvio padrão. Nesse cenário, a função χ_A^2 assume sua forma usual:

$$\chi_A^2(\vec{\theta}) = \sum_{i=1}^N \frac{\left(y_{\text{teo}}(z_i, \vec{\theta}) - y_{\text{obs},i}\right)^2}{\sigma_i^2}. \quad (3.13)$$

Em cosmologia, ao contrário de outras áreas do conhecimento, os dados observacionais raramente são independentes. Isso ocorre principalmente devido aos procedimentos de determinação de distâncias cosmológicas por meio de calibrações sucessivas, o que pode implicar em uma certa correlação entre os dados obtidos.

É comum utilizar informações adicionais em ajustes de modelos cosmológicos. Enquanto alguns parâmetros do modelo têm prior uniforme, podemos utilizar priors gaussianos em outros parâmetros cujos valores foram obtidos por meio de experimentos não correlacionados. Para esses casos, o prior individual de um parâmetro geral θ_k é dado por:

$$P_k(\theta_k) = c_1 \exp\left(-\frac{1}{2} \left(\frac{\theta_k - \theta_{k,\text{obs}}}{\sigma_{\theta_{ki}}}\right)^2\right), \quad (3.14)$$

onde c_1 é uma constante de normalização, $\theta_{k,\text{obs}}$ é o valor do parâmetro θ_k obtido através de algum outro procedimento experimental com uma distribuição normal, com desvio padrão $\sigma_{\theta_{ki}}$. Em um modelo de K parâmetros, onde L desses K parâmetros possuem priors gaussianas e os restantes possuem priors uniformes, a função de densidade de probabilidade *à priori* $P(\vec{\theta})$ é dada por:

$$P(\vec{\theta}) = c_2 \prod_{i=1}^L P_k(\theta_k) = c_2 \exp\left(-\frac{1}{2} \sum_{k=1}^L \frac{(\theta_k - \theta_{k,\text{obs}})^2}{\sigma_{\theta_{ki}}^2}\right), \quad (3.15)$$

e c_2 é uma constante que engloba as priors uniformes e as constantes de normalização das priors gaussianas.

A função de densidade de probabilidade *a posteriori* dos parâmetros do modelo, levando em consideração as restrições observacionais, será calculada:

$$P(\vec{\theta}|\{Y_i\}) = N \exp\left(-\frac{1}{2} \chi_A^2(\vec{\theta}) - \frac{1}{2} \sum_{k=1}^L \frac{(\theta_k - \theta_{k,\text{obs}})^2}{\sigma_{\theta_{ki}}^2}\right). \quad (3.16)$$

Assim, a estimativa dos parâmetros é obtida ao encontrar o valor máximo da função mencionada (3.16). Além disso, a partir dessa função, é possível calcular os intervalos de confiança, a correlação entre os parâmetros ajustados e as estimativas individuais obtidas por meio da marginalização dos parâmetros não desejados.

Até o momento, consideramos um conjunto de dados A , composto por N medições da grandeza y em função do *redshift*. No entanto, temos a liberdade de adicionar quantos conjuntos de dados desejarmos, contendo diferentes grandezas. Por exemplo, se desejarmos incluir na análise um conjunto independente de dados, chamado A_2 , com N_2 medições de uma grandeza geral u , basta multiplicar a verossimilhança original apresentada em (3.11) por um termo proporcional a $\exp^{-\frac{1}{2} \chi_{A_2}^2}$, onde $\chi_{A_2}^2$ é análogo à expressão (3.12), substituindo y por u . Dessa forma, uma expressão para a probabilidade de um vetor de parâmetros específico, considerando os conjuntos de dados mencionados, será análoga à

equação (3.16).

$$P(\vec{\theta}|\{Y_i\}) = N \exp\left(-\frac{1}{2}\chi_A^2(\vec{\theta}) - \frac{1}{2}\chi_{A_2}^2(\vec{\theta}) - \frac{1}{2}\sum_{k=1}^L \frac{(\theta_k - \theta_{k,\text{obs}})^2}{\sigma_{\theta_{ki}}^2}\right). \quad (3.17)$$

Considerando M conjuntos de dados independentes, onde cada conjunto contém suas próprias medições e grandezas, podemos generalizar a análise.

$$P(\vec{\theta}|\{Y_i\}) = N \exp\left(-\frac{1}{2}\sum_{i=1}^M \chi_{A_i}^2(\vec{\theta}) - \frac{1}{2}\sum_{k=1}^L \frac{(\theta_k - \theta_{k,\text{obs}})^2}{\sigma_{\theta_{ki}}^2}\right). \quad (3.18)$$

Um detalhe final a ser destacado é a prática comum de definir a função de verossimilhança $L(\vec{\theta}|\{Y_i\})$ no processo de otimização de probabilidades, de acordo com:

$$L(\vec{\theta}|\{Y_i\}) = -\ln(P(\vec{\theta}|\{Y_i\})) + \ln(N). \quad (3.19)$$

Dessa forma, o ajuste é realizado através da minimização da função:

$$L(\vec{\theta}|\{Y_i\}) = \frac{1}{2}\sum_{i=1}^M \chi_{A_i}^2\{Y_i|\vec{\theta}\} + \frac{1}{2}\sum_{k=1}^L \frac{(\theta_k - \theta_{k,\text{obs}})^2}{\sigma_{\theta_{ki}}^2}. \quad (3.20)$$

Em relação a M que é o número de conjuntos de dados, onde Y_i representa a grandeza do i -ésimo conjunto e L é o número de parâmetros com prioris gaussianas, é importante ressaltar que esse procedimento não deve gerar diferenças no valor otimizado do vetor de parâmetros $\vec{\theta}_k$.

3.3 Método MCMC

Em muitos casos, calcular a distribuição *a posteriori* de forma analítica é extremamente difícil. Portanto, é necessário recorrer a ferramentas numéricas para realizar a tarefa de estimativa. Dentre as diversas opções disponíveis, a Cadeia de Markov Monte Carlo (MCMC) com o Algoritmo Metropolis-Hastings (MHA) é uma das principais escolhas. Iremos estudar essa técnica no restante deste capítulo.

As cadeias de Markov e o método Monte Carlo são duas ferramentas fundamentais da estatística computacional e são amplamente utilizados na análise de dados em física e cosmologia. Segundo Barbu (2020), temos, de maneira geral, duas categorias de métodos de Monte Carlo:

1. Monte Carlo Sequencial - método estocástico que preserva e propaga uma população de exemplos por amostragem sequencial e repesagem de importância, normalmente em um espaço de estado de baixa dimensão.
2. Cadeia de Markov Monte Carlo - método estocástico que simula uma cadeia de Markov para explorar o espaço de estado com uma probabilidade estacionária projetada para convergir para uma determinada probabilidade alvo.

Nos métodos MCMC (Cadeias de Markov Monte Carlo), uma sequência de pontos consecutivos, chamada de cadeia, é gerada no espaço de parâmetros, sendo a densidade dos pontos proporcional *a posteriori* e convergindo para uma distribuição estável. Vamos tratar sobre o algoritmo Metropolis-Hastings e aplicá-lo a um modelo linear Gaussiano. No algoritmo Metropolis-Hastings, um ponto inicial θ_s é escolhido aleatoriamente dentro de uma região especificada do espaço de parâmetros, e *a posteriori* não normalizada é avaliada nesse ponto, chamado de P_s . Em seguida, um salto é realizado de acordo com uma densidade proposta para um ponto candidato, θ_c , e *a posteriori* é avaliada nesse ponto, chamado de P_c . Uma decisão é tomada para mover-se para o ponto candidato e adicioná-lo à cadeia, ou permanecer no ponto inicial, adicionando-o novamente à cadeia e escolhendo um novo ponto candidato.

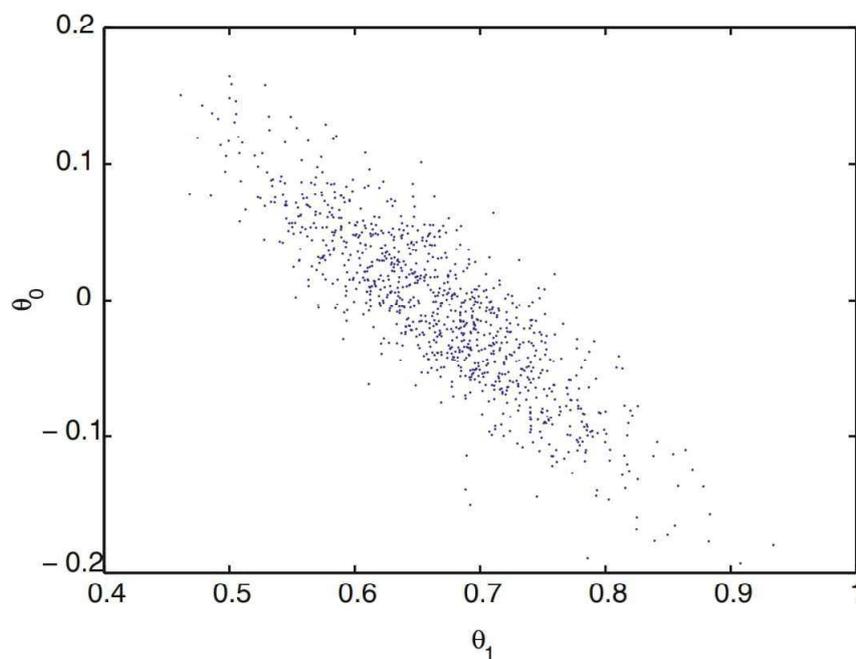


Figura 8 – Aplicação da metodologia MCMC a um modelo linear gaussiano

. Fonte: (MARCH, 2013).

A seleção do ponto candidato é determinada pela densidade da proposta, que deve ser escolhida com cuidado. Quando obtemos a distribuição posterior conjunta por meio dos métodos MCMC e é representada pela densidade de pontos no espaço de parâmetros que formam a “cadeia”. As Figuras 8 e 9, são algumas situações ilustrativas da estimativa de dois parâmetros quaisquer, com a ideia de mostrar como é montado o conjunto de pontos no espaço de parâmetros e as distribuições marginais.

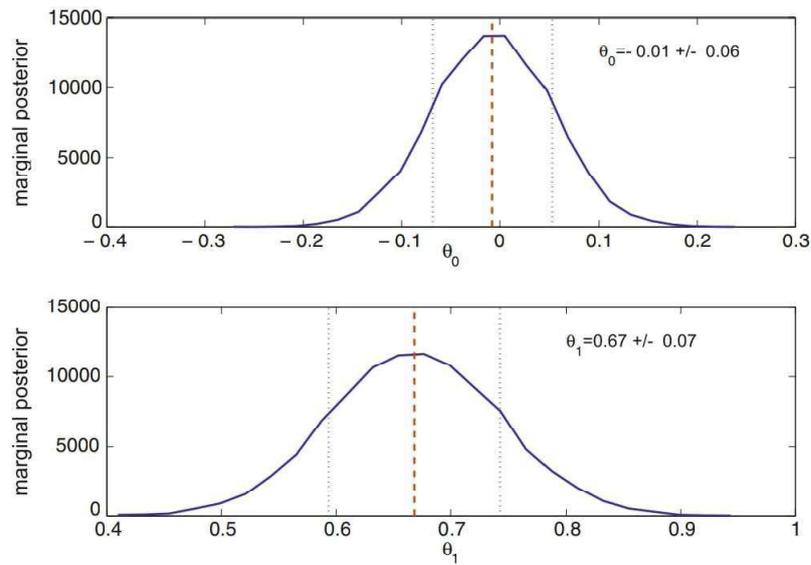


Figura 9 – Distribuições marginais de dois parâmetros obtidos

. Fonte: (MARCH, 2013).

A técnica MCMC (Markov Chain Monte Carlo) está diretamente relacionada com a estatística bayesiana. a técnica MCMC é uma ferramenta computacional fundamental na estatística bayesiana, permitindo a obtenção de aproximações das distribuições posteriores e a realização de inferências probabilísticas sobre os parâmetros desconhecidos em modelos estatísticos bayesianos. No capítulo seguinte, Vamos discutir como usar os dados observacionais para estimar o parâmetros tais como a constante de Hubble e a densidade relativa de matéria.

4 A constante de Hubble e a densidade relativa de matéria

4.1 Equações Cosmológicas

O objetivo desta seção é apresentar as principais equações do modelo Λ CDM (Lambda Cold Dark Matter) que é um dos pilares fundamentais da cosmologia moderna. Ele descreve a estrutura e a evolução do universo com base em duas componentes principais: a constante cosmológica Λ (Lambda) e a matéria escura fria (CDM - Cold Dark Matter). (PERIVOLAROPOULOS; SKARA, 2022) Fornecemos uma notação clara, embora não deduzamos todas as equações neste momento.

A métrica de Friedmann-Lemaître-Robertson-Walker (FLRW) é uma descrição matemática de um Universo que é homogêneo, isotrópico e em expansão. Pode ser expressa como:

$$ds^2 \equiv g_{\mu\nu} dx^\mu dx^\nu = c^2 dt^2 - a(t)^2 \left(\frac{dr^2}{1 - \kappa r^2} \right) + r^2 d\theta^2 + r^2 \sin^2(\theta) d\phi^2, \quad (4.1)$$

em que $g_{\mu\nu}$ representa as componentes da métrica, $a(t)$ é o fator de escala, uma função que depende apenas do tempo e descreve como as distâncias evoluem ao longo do tempo em um Universo em expansão.

Se a métrica fornece o contexto no qual os eventos ocorrem, é responsabilidade do tensor energia-momento descrever a distribuição e fluxo de energia e momento associados ao conteúdo material e energético que preenche o espaço-tempo. Na cosmologia, é comum utilizar o tensor energia-momento de um fluido perfeito para modelar o substrato material em escalas amplas,

$$T_{\mu\nu} = (\rho_T + p_T) \frac{u_\mu u_\nu}{c^2} - p_T g_{\mu\nu}. \quad (4.2)$$

Nessa formulação, $\rho_T = \sum \rho_i$ e $p_T = \sum p_i$ representam a densidade total de energia e a pressão total do fluido, respectivamente, onde a soma é realizada sobre os diferentes componentes do fluido. A quadrivelocidade do fluido, u^μ , é a velocidade medida por um observador comóvel. É importante ressaltar que algumas referências utilizam a letra ρ para denotar a densidade de massa da matéria e ε para denotar a densidade de energia. Essas quantidades estão relacionadas pela famosa expressão $E = mc^2$, ou seja, $\varepsilon = \rho c^2$. Além disso, é comum adotarmos unidades naturais e usar a convenção $c = 1$, o que pode causar confusão ao comparar as unidades de cada grandeza.

As equações de campo de Einstein desempenham um papel fundamental ao relacionar o conteúdo material e energético com a geometria do espaço-tempo. Ao longo da história, essas equações passaram por modificações em sua forma. Inicialmente, para obter

uma solução para um Universo estático, Einstein introduziu a constante cosmológica. No entanto, evidências da expansão do Universo levaram ao abandono dessa constante em anos posteriores. No entanto, estudos recentes, como os baseados em dados de supernovas do tipo I_a , sugerem que essa constante possa ter um valor não nulo. Por essa razão, as equações de campo são escritas com a inclusão do termo cosmológico.

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = \frac{8\pi G}{c^4}T_{\mu\nu} + \Lambda g_{\mu\nu}, \quad (4.3)$$

em que G é a constante gravitacional de Newton, $R_{\mu\nu}$ representa o tensor de Ricci e R é o escalar de curvatura de Ricci. Essas quantidades estão diretamente relacionadas à geometria do espaço-tempo, pois são calculadas a partir do tensor métrico e suas derivadas. A constante cosmológica Λ também desempenha um papel nessas equações, possuindo dimensões de (comprimento)⁻².

Substituindo as equações (4.1) e (4.2) nas equações de campo de Einstein (4.3), chegamos às equações de Friedmann.

$$H^2 = \frac{8\pi G}{3c^2}\rho_T - \frac{\kappa c^2}{a^2}, \quad (4.4)$$

$$2\frac{\ddot{a}}{a} + \frac{\dot{a}^2}{a^2} + \frac{\kappa c^2}{a^2} = -\frac{8\pi G}{c^2}p_T + \Lambda c^2. \quad (4.5)$$

Ao definirmos o parâmetro de Hubble como $H \equiv \frac{\dot{a}}{a}$, que é uma função do tempo que mede a taxa de expansão do Universo, podemos expressar as equações em termos dele. O valor dessa função no presente, H_0 , é conhecido como constante de Hubble. Além disso, introduzimos a densidade de energia relacionada à constante cosmológica, $\rho_\Lambda = \frac{\rho_\Lambda}{8\pi G}$, incorporando essa componente ao somatório da densidade de energia total ρ_T .

Pela expressão (4.4), podemos observar que, no caso em que o Universo não possui curvatura espacial ($k = 0$), a densidade assume um valor definido como “densidade crítica”, $\rho_c = \frac{3c^2 H^2}{8\pi G}$. Essa definição tem um significado simples: se a densidade de energia total que medimos no Universo for igual à densidade crítica, então temos um Universo espacialmente plano. A razão entre a densidade de energia de uma determinada componente do Universo e a densidade crítica é chamada de parâmetro de densidade, $\omega_i \equiv \frac{\rho_i}{\rho_c}$, e mede a contribuição energética relativa dessa componente. Portanto, utilizando essas definições, a primeira equação de Friedmann pode ser expressa como:

$$H^2 = H_0^2 \left[\Omega_{r0} \left(\frac{a_0}{a} \right)^4 + \Omega_{m0} \left(\frac{a_0}{a} \right)^3 + \Omega_{\kappa 0} \left(\frac{a_0}{a} \right)^2 + \Omega_{\Lambda 0} \right], \quad (4.6)$$

em que Ω_{r0} , Ω_{m0} e $\Omega_{\Lambda 0}$ representam os parâmetros de densidade da radiação, matéria e constante cosmológica, respectivamente, avaliados no presente (hoje). Além disso, $\Omega_{\kappa 0} \equiv \frac{-\kappa c^2}{(a_0 H_0)^2}$ é o parâmetro de densidade associado à curvatura. Podemos estabelecer uma relação entre esses parâmetros ao substituímos os valores atuais ($H = H_0$ e $a = a_0$) na expressão

acima.

$$1 = \Omega_{\kappa 0} + \sum_i \Omega_{i0}. \quad (4.7)$$

O modelo Λ CDM consiste em matéria (bariônica + escura fria) sem pressão, a presença da constante cosmológica e possivelmente curvatura. A radiação, embora desempenhe um papel importante na evolução do Universo jovem (até alguns milhares de anos após o Big Bang), possui uma densidade muito baixa nos dias atuais e pode ser negligenciada ao lidar com dados em baixos deslocamentos para o vermelho (redshifts). Portanto, podemos expressar o parâmetro de densidade da curvatura como $\Omega_{\kappa 0} = 1 - \Omega_{m0} - \Omega_{\Lambda 0}$. Além disso, utilizando a relação entre o fator de escala e o deslocamento para o vermelho (redshift), $(1 + z = \frac{a_0}{a})$, podemos escrever a equação de Friedmann na forma final:

$$H^2 = H_0^2 \left[\Omega_{m0}(1+z)^3 + (1 - \Omega_{\Lambda 0} - \Omega_{m0})(1+z)^2 + \Omega_{\Lambda 0} \right]. \quad (4.8)$$

A partir das equações de Friedmann (4.4) e (4.5), é possível obter uma equação que descreve a aceleração da expansão do Universo.

$$\frac{\ddot{a}}{a} = \frac{\Lambda c^2}{3} - \frac{4\pi G}{3c^2}(\rho_T + 3p_T). \quad (4.9)$$

Podemos observar a importância da constante cosmológica λ na dinâmica cosmológica por meio da equação acima. Dependendo do valor dessa quantidade, o Universo pode estar em expansão acelerada ou desacelerada. É importante destacar que, para um Universo composto apenas por matéria, as densidades de energia (ρ_T) e pressão (p_T) são sempre positivas. Portanto, na ausência da constante cosmológica, a expansão é sempre desacelerada ($\ddot{a} < 0$).

A constante de Hubble H_0 é um parâmetro fundamental em qualquer modelo cosmológico de interesse. Sua determinação desempenha um papel crucial, pois fornece informações sobre a taxa de expansão do Universo atual, conhecida como fluxo de Hubble. Além disso, diversos dados observacionais são calibrados com base em um valor de H_0 obtido de forma independente dos modelos cosmológicos, destacando a sua importância. Hubble verificou a expansão cósmica com a medição das distâncias e das velocidades de afastamento galácticas.

4.2 Estimativa dos parâmetros H_0 e Ω_0

Python é uma linguagem de programação amplamente utilizada e conhecida pela sua simplicidade e legibilidade. Uma das formas mais eficientes de utilizar o Python é através da linha de comando, onde é possível executar comandos e interagir diretamente com o interpretador Python.

Na Figura 10 estamos instalando a biblioteca EMCEE, que é uma biblioteca Python que implementa o método MCMC com invariância afim, proposto por Goodman e Weare.



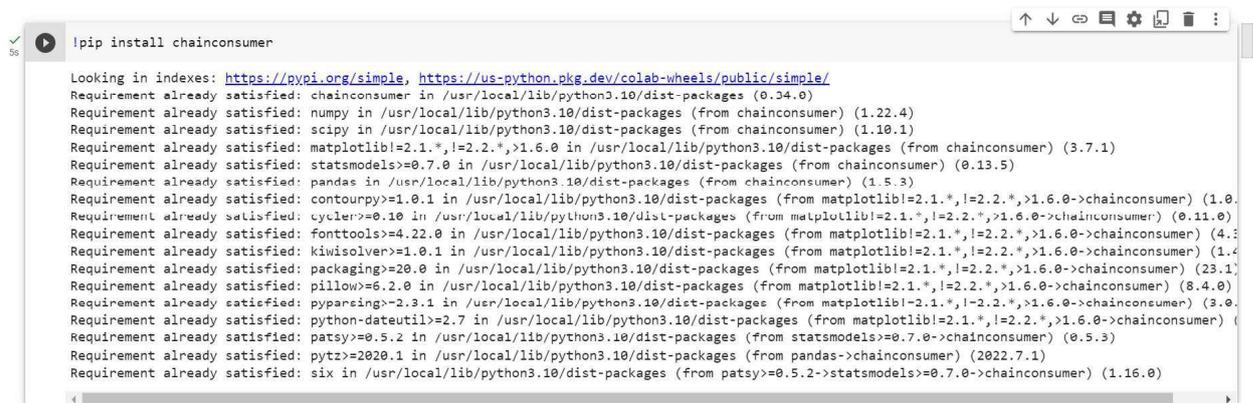
```
# instalando a biblioteca emcee
!pip install emcee

Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Collecting emcee
  Downloading emcee-3.1.4-py2.py3-none-any.whl (46 kB)
----- 46.2/46.2 kB 2.4 MB/s eta 0:00:00
Requirement already satisfied: numpy in /usr/local/lib/python3.10/dist-packages (from emcee) (1.22.4)
Installing collected packages: emcee
Successfully installed emcee-3.1.4
```

Figura 10 – Linhas de Comando em Python.

Após instalarmos essa biblioteca, usamos o comando “!pip install chainconsumer”, na Figura 11. Isso irá baixar e instalar a biblioteca “chainconsumer” e suas dependências necessárias no ambiente Python em que você está trabalhando. Segundo (Hinton, 2016),

ChainConsumer é um pacote python projetado para fazer uma coisa - consumir a saída de cadeias de processos de Monte Carlo como MCMC, que foi um dos métodos trabalhados no capítulo anterior. O ChainConsumer pode utilizar essas cadeias para produzir plotagens da superfície posterior inferidas das distribuições da cadeia, para plotar as cadeias como caminhadas (para verificar a mistura e a convergência) e para gerar resumos de parâmetros na forma de tabelas LaTeX. Além de tudo isso, se você tiver vários modelos (cadeias), poderá carregá-los todos e realizar algumas comparações de modelos usando métricas AIC, BIC ou DIC (Hinton, 2016).



```
!pip install chainconsumer

Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Requirement already satisfied: chainconsumer in /usr/local/lib/python3.10/dist-packages (0.34.0)
Requirement already satisfied: numpy in /usr/local/lib/python3.10/dist-packages (from chainconsumer) (1.22.4)
Requirement already satisfied: scipy in /usr/local/lib/python3.10/dist-packages (from chainconsumer) (1.10.1)
Requirement already satisfied: matplotlib>=2.1.*,!>=2.2.*,>1.6.0 in /usr/local/lib/python3.10/dist-packages (from chainconsumer) (3.7.1)
Requirement already satisfied: statsmodels>=0.7.0 in /usr/local/lib/python3.10/dist-packages (from chainconsumer) (0.13.5)
Requirement already satisfied: pandas in /usr/local/lib/python3.10/dist-packages (from chainconsumer) (1.5.3)
Requirement already satisfied: contourpy>=1.0.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib>=2.1.*,!>=2.2.*,>1.6.0->chainconsumer) (1.0.7)
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.10/dist-packages (from matplotlib>=2.1.*,!>=2.2.*,>1.6.0->chainconsumer) (0.11.0)
Requirement already satisfied: fonttools>=4.22.0 in /usr/local/lib/python3.10/dist-packages (from matplotlib>=2.1.*,!>=2.2.*,>1.6.0->chainconsumer) (4.22.0)
Requirement already satisfied: kiwisolver>=1.0.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib>=2.1.*,!>=2.2.*,>1.6.0->chainconsumer) (1.4.5)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.10/dist-packages (from matplotlib>=2.1.*,!>=2.2.*,>1.6.0->chainconsumer) (23.1)
Requirement already satisfied: pillow>=6.2.0 in /usr/local/lib/python3.10/dist-packages (from matplotlib>=2.1.*,!>=2.2.*,>1.6.0->chainconsumer) (8.4.0)
Requirement already satisfied: pyparsing>=2.3.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib>=2.1.*,!>=2.2.*,>1.6.0->chainconsumer) (3.0.9)
Requirement already satisfied: python-dateutil>=2.7 in /usr/local/lib/python3.10/dist-packages (from matplotlib>=2.1.*,!>=2.2.*,>1.6.0->chainconsumer) (2.8.2)
Requirement already satisfied: patsy>=0.5.2 in /usr/local/lib/python3.10/dist-packages (from statsmodels>=0.7.0->chainconsumer) (0.5.3)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.10/dist-packages (from pandas->chainconsumer) (2022.7.1)
Requirement already satisfied: six in /usr/local/lib/python3.10/dist-packages (from patsy>=0.5.2->statsmodels>=0.7.0->chainconsumer) (1.16.0)
```

Figura 11 – Linhas de Comando em Python.

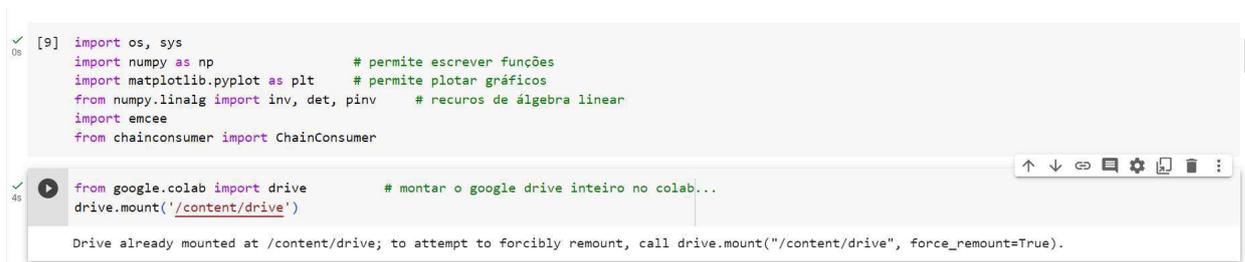
Na Figura 12, vemos o módulo OS e o módulo sys. O módulo OS em Python é uma biblioteca padrão que oferece funcionalidades para interagir com o sistema operacional.

Ele é parte dos módulos utilitários padrão do Python e proporciona uma forma portátil de utilizar recursos específicos do sistema operacional. O módulo `OS`, em conjunto com o módulo `os.path`, disponibiliza uma ampla variedade de funções para manipulação de arquivos e interação com o sistema de arquivos.

O módulo `sys` em Python é uma biblioteca padrão que permite acessar variáveis e funções relacionadas ao interpretador Python. Ele fornece acesso a variáveis internas e funções que estão diretamente ligadas ao funcionamento do interpretador. O módulo `sys` está sempre disponível, permitindo aos desenvolvedores interagir com o ambiente de execução do Python de forma mais avançada e personalizada.

O submódulo `numpy.linalg` do NumPy é uma ferramenta poderosa que oferece funcionalidades para realizar operações comuns da álgebra linear. Ele é projetado especificamente para manipular sistemas lineares, calcular inversas de matrizes, encontrar autovalores e autovetores, e muito mais. Com o submódulo `numpy.linalg`, é possível realizar uma ampla gama de cálculos e transformações matemáticas relacionadas à álgebra linear de maneira eficiente e conveniente.

Na parte final da Figura 12, estamos importando meu drive com arquivos e dados para rodarmos futuramente aqui.



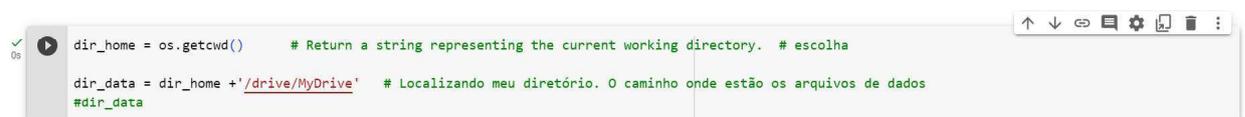
```
[9] import os, sys
import numpy as np          # permite escrever funções
import matplotlib.pyplot as plt # permite plotar gráficos
from numpy.linalg import inv, det, pinv # recursos de álgebra linear
import emcee
from chainconsumer import ChainConsumer

from google.colab import drive # montar o google drive inteiro no colab...
drive.mount('/content/drive')
```

Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).

Figura 12 – Linhas de Comando em Python.

O método `getcwd()` do módulo `os` em Python retorna uma string que representa o diretório de trabalho atual da aplicação, em inglês o “Current Working Directory” (CWD). Essa string contém o caminho completo para o diretório atual em que a aplicação está sendo executada. Assim como mostra a Figura 13.



```
dir_home = os.getcwd() # Return a string representing the current working directory. # escolha
dir_data = dir_home + '/drive/MyDrive' # Localizando meu diretório. O caminho onde estão os arquivos de dados
#dir_data
```

Figura 13 – Linhas de Comando em Python.

Definimos o modelo, que neste caso (Figura 14), os parâmetros são H_0 e Ω_m , e o parâmetro de Hubble tem termos de matéria e de constante cosmológica:

$$H(z) = H_0 \sqrt{\Omega_m (1+z)^3 + (1 - \Omega_m)}. \quad (4.10)$$

Definimos agora a função logarítmica da verossimilhança, para esse caso temos:

$$\ln L = -\frac{1}{2} \chi^2, \quad (4.11)$$

onde,

$$\chi^2 = (H_{\text{obs},i} - H_{\text{teo}}) \mathbf{C}_{ij}^{-1} (H_{\text{obs},j} - H_{\text{teo}}), \quad (4.12)$$

em que C_{ij} é a matriz de correlação. Aqui iniciamos o algoritmo de Metropolis-Hastings, que é baseado no princípio de aceitar ou rejeitar uma nova amostra com base em uma proporção de probabilidades. Ele gera uma sequência de amostras aleatórias que convergem para a distribuição de interesse. Com o método da Seção 3.1, vemos que para obter a verossimilhança máxima precisamos ter o χ^2 mínimo (método trabalhado na Seção 3.2).

A partir daqui entra a Estatística Bayesiana, onde definimos uma prior plana e outra gaussiana, iniciando esse processo na linha 14 da Figura 14.



```

def Hzmodel(params, z):
    H0, Omegam = params
    return H0*np.sqrt(Omegam*(1+z)**3+(1-Omegam))

def lnlike(p, zHz, Hz, inv_cov_matr):
    chi2=0.
    residual=Hz-Hzmodel(p, zHz)
    for i in range(0, len(zHz)):
        for j in range(0, len(zHz)):
            chi2=chi2+((residual[i])*inv_cov_matr[i,j]*(residual[j]))
    return -0.5 * chi2

def lnflatprior(p):
    H0, Omegam = p
    if (50. < H0 < 100. and 0.01 < Omegam < 0.99):
        return 0.0
    return -np.inf

def lnprior(p):
    H0, Omegam = p
    if (50. < H0 < 100. and 0.01 < Omegam < 0.99 and -3.0 < w0 < 0.0 and -5.0 < wa < 5.0):
        return -0.5 * (np.sum(((Omegam-Omegam_fid)/errOmegam_fid)**2) + np.sum(((H0-H0_fid)/errH0_fid)**2))
    return -np.inf

def lnprob(p, zHz, Hz, inv_cov_matr):
    lp = lnflatprior(p)
    return lp + lnlike(p, zHz, Hz, inv_cov_matr) if np.isfinite(lp) else -np.inf

```

Figura 14 – Linhas de Comando em Python.

Na Figura 15, usamos alguns dados obtidos pelo satélite Planck, em que H_0_fid representa a constante de Hubble, que é uma medida da taxa atual de expansão do universo. O valor H_0_fid é o valor fiducial adotado para a constante de Hubble e é expresso em unidades de quilômetros por segundo por megaparsec (km/s/Mpc). Neste caso, $H_0_fid = 67,37$ km/s/Mpc. O termo $errH_0_fid$ representa a incerteza ou erro associado ao valor fiducial da constante de Hubble. É uma estimativa da incerteza na medição da taxa de expansão do universo. Neste caso, $H_0_fid = 0.54$ km/s/Mpc. O termo $Omegam_fid$ é o

parâmetro de densidade de matéria no universo. Ele descreve a fração de densidade de matéria em relação à densidade crítica necessária para a expansão do universo. O valor fiducial `Omegam_fid` é o valor de referência adotado para Ω_m e representa a fração de densidade de matéria no universo. Neste caso, `Omegam_fid = 0.308`.

```

H0_fid = 67.37
errH0_fid = 0.54
Omegam_fid = 0.308

H0_pl = H0_fid
Omegam_pl = Omegam_fid
planck=[H0_pl, Omegam_pl]

```

Figura 15 – Linhas de Comando em Python.

Na Figura 16, estamos obtendo dados de um arquivo de texto, ao qual importamos ao drive no início do programa.

```

os.chdir(dir_data)
filename = 'HzTable_MM_BC03.dat'
z, Hz, errHz = np.genfromtxt(filename, comments='#', usecols=(0,1,2), unpack=True, delimiter=',') # obtendo dados de um arquivo de texto
ref = np.genfromtxt(filename, comments='#', usecols=(3), unpack=True, dtype=str, delimiter=',')

filename = 'data_MM20.dat'
zmod, imf, slib, sps, spsooo = np.genfromtxt(filename, comments='#', usecols=(0,1,2,3,4), unpack=True)
os.chdir(dir_home)

```

Figura 16 – Linhas de Comando em Python.

A seguir definimos matrizes de covariância, nas Figuras 17 e 18, sendo que a da Figura 17 é uma matriz de covariância diagonal, na qual fazemos interpolação com 100 pontos.

```

cov_mat_diag = np.zeros((len(z), len(z)), dtype='float64')

for i in range(len(z)):
    cov_mat_diag[i,i] = errHz[i]**2

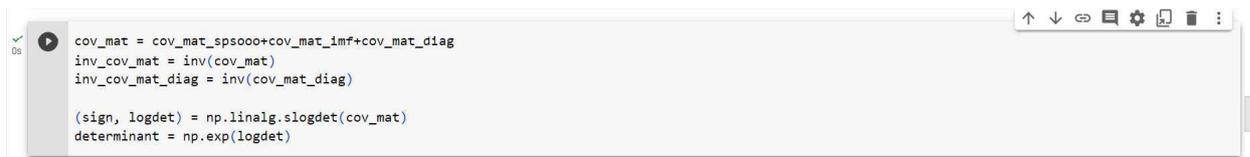
imf_intp = np.interp(z, zmod, imf)/100 # interpolando com 100 pontos
slib_intp = np.interp(z, zmod, slib)/100
sps_intp = np.interp(z, zmod, sps)/100
spsooo_intp = np.interp(z, zmod, spsooo)/100

cov_mat_imf = np.zeros((len(z), len(z)), dtype='float64')
cov_mat_slib = np.zeros((len(z), len(z)), dtype='float64')
cov_mat_sps = np.zeros((len(z), len(z)), dtype='float64')
cov_mat_spsooo = np.zeros((len(z), len(z)), dtype='float64')

for i in range(len(z)):
    for j in range(len(z)):
        cov_mat_imf[i,j] = Hz[i] * imf_intp[i] * Hz[j] * imf_intp[j]
        cov_mat_slib[i,j] = Hz[i] * slib_intp[i] * Hz[j] * slib_intp[j]
        cov_mat_sps[i,j] = Hz[i] * sps_intp[i] * Hz[j] * sps_intp[j]
        cov_mat_spsooo[i,j] = Hz[i] * spsooo_intp[i] * Hz[j] * spsooo_intp[j]

```

Figura 17 – Linhas de Comando em Python.



```

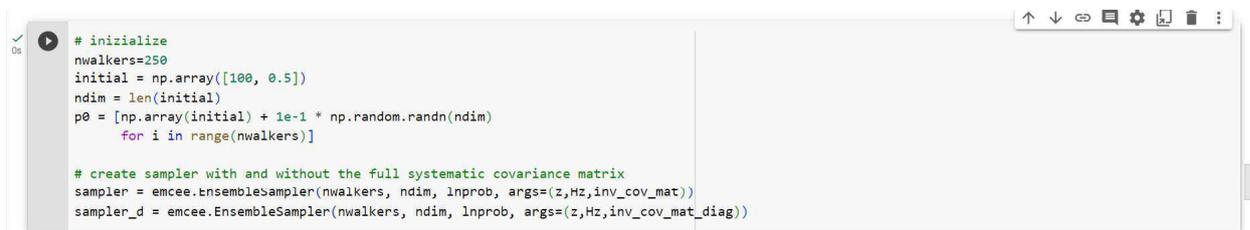
cov_mat = cov_mat_sps000+cov_mat_imf+cov_mat_diag
inv_cov_mat = inv(cov_mat)
inv_cov_mat_diag = inv(cov_mat_diag)

(sign, logdet) = np.linalg.slogdet(cov_mat)
determinant = np.exp(logdet)

```

Figura 18 – Linhas de Comando em Python.

Os comandos da Figura 19 são usados para inicializar um conjunto de pontos iniciais para os “caminhantes” em um algoritmo MCMC. E serve para criar um conjunto de pontos iniciais para os “caminhantes” no algoritmo MCMC, com base em valores iniciais definidos e adicionando um deslocamento aleatório. Essa abordagem visa explorar de forma mais eficiente o espaço de parâmetros e obter uma melhor amostragem da distribuição de interesse.



```

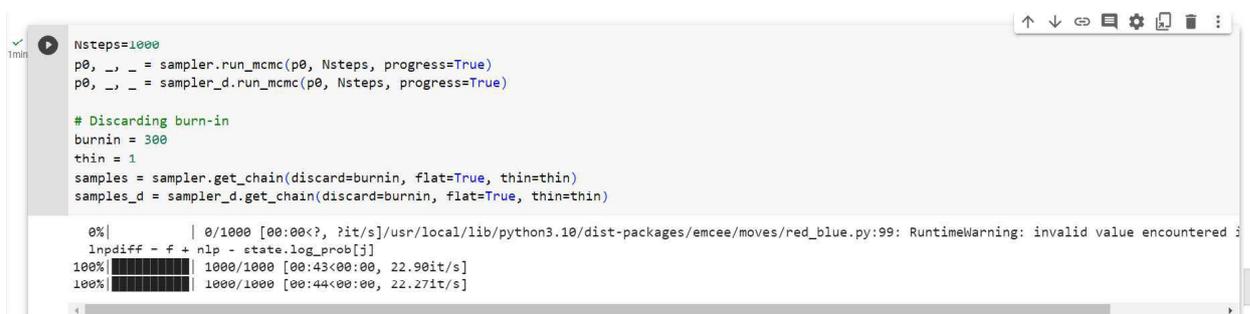
# initialize
nwalkers=250
initial = np.array([100, 0.5])
ndim = len(initial)
p0 = [np.array(initial) + 1e-1 * np.random.randn(ndim)
      for i in range(nwalkers)]

# create sampler with and without the full systematic covariance matrix
sampler = emcee.EnsembleSampler(nwalkers, ndim, lnprob, args=(z,Hz,inv_cov_mat))
sampler_d = emcee.EnsembleSampler(nwalkers, ndim, lnprob, args=(z,Hz,inv_cov_mat_diag))

```

Figura 19 – Linhas de Comando em Python.

Na Figura 20, fizemos algo parecido com o da Figura 19, o comando em questão é usado para executar o algoritmo MCMC (Método de Monte Carlo de Cadeia de Markov) em duas instâncias diferentes, `sampler` e `sampler_d`, e realizar a análise das amostras geradas. Essas etapas são comumente utilizadas para executar o algoritmo MCMC, descartar os passos de burn-in e obter as amostras finais para análise.



```

Nsteps=1000
p0, _, _ = sampler.run_mcmc(p0, Nsteps, progress=True)
p0, _, _ = sampler_d.run_mcmc(p0, Nsteps, progress=True)

# Discarding burn-in
burnin = 300
thin = 1
samples = sampler.get_chain(discard=burnin, flat=True, thin=thin)
samples_d = sampler_d.get_chain(discard=burnin, flat=True, thin=thin)

0%|          | 0/1000 [00:00<?, ?it/s] /usr/local/lib/python3.10/dist-packages/emcee/moves/red_blue.py:99: RuntimeWarning: invalid value encountered
lnpdiff - f + nlp - state.log_prob[j]
100%|██████████| 1000/1000 [00:43<00:00, 22.90it/s]
100%|██████████| 1000/1000 [00:44<00:00, 22.27it/s]

```

Figura 20 – Linhas de Comando em Python.

Na Figura 21 está sendo realizada uma análise estatística dos resultados obtidos a partir do algoritmo MCMC. Na Figura 22 o trecho de código está utilizando a biblioteca

ChainConsumer para realizar a visualização e análise das cadeias de amostras geradas pelo MCMC.



```

H0_mcmc, Omegam_mcmc = map(lambda v: (v[1], v[2]-v[1], v[1]-v[0]),
                           zip(*np.percentile(samples_d, [16, 50, 84],
                                               axis=0)))

print("MCMC result 1 sigma - without systematics:")
print("""   H0 = {0[0]:.3f} +{0[1]:.3f} -{0[2]:.3f} (planck18: {1:.3f})
         Omegam = {2[0]:.3f} +{2[1]:.3f} -{2[2]:.3f} (planck18: {3:.3f})""").format(H0_mcmc, H0_pl, Omegam_mcmc, Omegam_pl))

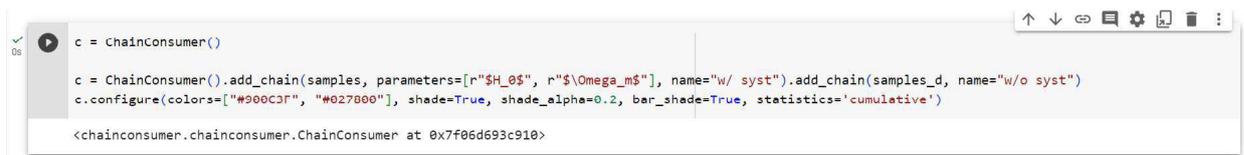
H0_mcmc, Omegam_mcmc = map(lambda v: (v[1], v[2]-v[1], v[1]-v[0]),
                           zip(*np.percentile(samples, [16, 50, 84],
                                               axis=0)))

print("MCMC result 1 sigma - with systematics:")
print("""   H0 = {0[0]:.3f} +{0[1]:.3f} -{0[2]:.3f} (planck18: {1:.3f})
         Omegam = {2[0]:.3f} +{2[1]:.3f} -{2[2]:.3f} (planck18: {3:.3f})""").format(H0_mcmc, H0_pl, Omegam_mcmc, Omegam_pl))

MCMC result 1 sigma - without systematics:
H0 = 66.139 +3.874 -3.993 (planck18: 67.370)
Omegam = 0.355 +0.112 -0.087 (planck18: 0.308)
MCMC result 1 sigma - with systematics:
H0 = 66.009 +5.530 -5.710 (planck18: 67.370)
Omegam = 0.355 +0.114 -0.089 (planck18: 0.308)

```

Figura 21 – Linhas de Comando em Python.



```

c = ChainConsumer()

c = ChainConsumer().add_chain(samples, parameters=[r"$H_0$", r"$\Omega_m$"], name="w/ syst").add_chain(samples_d, name="w/o syst")
c.configure(colors=["#990033", "#027000"], shade=True, shade_alpha=0.2, bar_shade=True, statistics='cumulative')

<chainconsumer.chainconsumer.ChainConsumer at 0x7f06d693c910>

```

Figura 22 – Linhas de Comando em Python.

Nesse código, estamos realizando a plotagem dos resultados utilizando o ChainConsumer. A gaussiana representada pela linha vermelha é uma representação com efeitos sistemáticos e a linha verde sem efeitos sistemáticos. Portanto, esse trecho de código gera um gráfico utilizando o ChainConsumer com as configurações especificadas, exibindo as cadeias de amostras e os valores de referência.

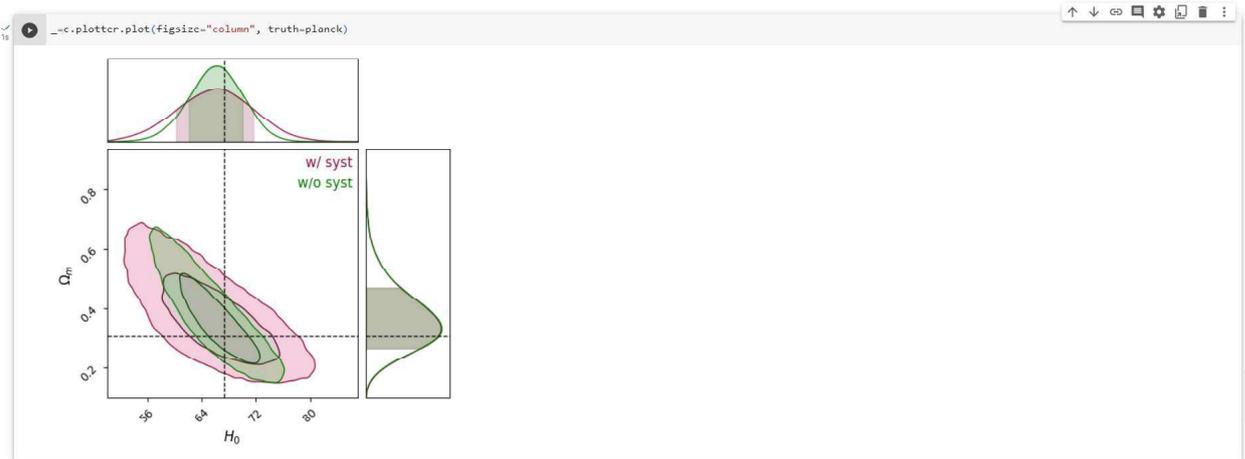


Figura 23 – Linhas de Comando em Python.

```

MCMC result 1 sigma - without systematics:
H0 = 66.132 +3.888 -4.030 (planck18: 67.370)
Omegam = 0.356 +0.113 -0.089 (planck18: 0.308)
MCMC result 1 sigma - with systematics:
H0 = 65.783 +5.577 -5.654 (planck18: 67.370)
Omegam = 0.355 +0.118 -0.089 (planck18: 0.308)

```

Figura 24 – Linhas de Comando em Python.

Em conclusão, a utilização de dados observacionais para estimar o parâmetro de Hubble, levando em conta os termos de matéria e constante cosmológica, desempenha um papel fundamental na compreensão da expansão do universo e da sua evolução. Através de análises estatísticas e modelos cosmológicos adequados, é possível obter estimativas precisas do parâmetro de Hubble e suas incertezas associadas, como mostro na figura 24. Essas estimativas nos fornecem insights valiosos sobre a expansão cósmica, as propriedades da matéria e os desafios que enfrentamos na busca por respostas sobre a origem e o destino do universo.

Conclusão

A análise de dados na cosmologia desempenha um papel fundamental na compreensão do universo e na investigação dos fenômenos cósmicos. Ao longo deste trabalho, exploramos os conceitos e métodos envolvidos nessa área fascinante. A abordagem estatística bayesiana permitiu estimar e interpretar os valores dos parâmetros com base em dados cosmológicos, considerando também informações prévias. Essa análise nos oferece uma visão mais profunda do universo, permitindo testar modelos, confrontar teoria e observações, e avançar em nosso conhecimento cósmico.

Durante nossa pesquisa, pudemos compreender a importância da coleta e análise de dados precisos e confiáveis na cosmologia. Através de técnicas estatísticas e algoritmos avançados, exploramos padrões, e propriedades dos dados cósmicos, buscando insights valiosos sobre a natureza do universo.

Nossos resultados revelaram a complexidade e a diversidade dos fenômenos cósmicos, bem como a importância de abordagens estatísticas robustas para interpretar os dados observacionais. À medida que continuamos a refinar nossos métodos estatísticos e aprimorar as técnicas de análise de dados, podemos esperar novas descobertas e avanços emocionantes no campo da cosmologia.

Este trabalho busca contribuir para o avanço da cosmologia e servir como uma fonte de inspiração para futuras pesquisas nessa área. A aplicação da estatística bayesiana em conjunção com a análise de dados cosmológicos desempenha um papel fundamental na busca por respostas que revelem a nossa origem e o nosso lugar dentro do vasto e enigmático universo. A abordagem estatística bayesiana oferece uma ferramenta poderosa para estimar e interpretar os parâmetros cosmológicos, permitindo-nos compreender a expansão cósmica, a composição da matéria e energia escuras, bem como a formação das estruturas cósmicas. Ao unir a estatística bayesiana com a análise de dados cosmológicos, podemos avançar em nosso conhecimento e desvendar os segredos que o universo guarda, revelando pistas fundamentais sobre nossa existência e o contexto em que estamos inseridos. Espera-se que essa abordagem estatística continue a impulsionar descobertas significativas e aprofundar nossa compreensão do universo em constante evolução.

Referências

- ALPHER, R. A.; BETHE, H.; GAMOW, G. The origin of chemical elements. *Physical Review*, v. 73, 4 1948. ISSN 0031-899X. Citado na página 12.
- ALPHER, R. A.; HERMAN, R. C. Theory of the origin and relative abundance distribution of the elements. *Reviews of Modern Physics*, v. 22, 4 1950. ISSN 0034-6861. Citado na página 12.
- ANDREON, S.; WEAVER, B. *Bayesian Methods for the Physical Sciences*. [S.l.]: Springer International Publishing, 2015. v. 4. ISBN 978-3-319-15286-8. Citado na página 13.
- BARBU, S.-C. Z. A. *Monte Carlo Methods*. [S.l.]: Springer, 2020. ISBN 978-981-13-2971-5. Citado na página 33.
- BAUMANN, D.; MCALISTTER, L. *Inflation and string theory*. [S.l.]: Cambridge University Press, 2015. ISBN 978-1-107-08969-3. Citado na página 13.
- BETOULE, M. et al. Improved cosmological constraints from a joint analysis of the sdss-ii and snls supernova samples. *Astronomy & Astrophysics*, v. 568, 8 2014. ISSN 0004-6361. Citado na página 13.
- CAMPBELL, H. et al. Cosmology with photometrically classified type ia supernovae from the sdss-ii supernova survey. *The Astrophysical Journal*, v. 763, 1 2013. ISSN 0004-637X. Citado na página 13.
- COELHO, J. G. O telescópio espacial james webb-uma nova era na astronomia. *Cadernos de Astronomia*, v. 3, n. 2, p. 112–121, 2022. Citado na página 18.
- DIMOPOULOS, K. *Introduction to cosmic inflation and dark energy*. [S.l.]: CRC Press, 2021. ISBN 9781351174862. Citado na página 13.
- DODELSON, S.; SCHMIDT, F. *Modern Cosmology*. [S.l.]: Academic Press, 2021. ISBN 978-0-12-815948-4. Citado 2 vezes nas páginas 12 e 13.
- DURRER, R. *The Cosmic Microwave Background*. [S.l.]: Cambridge University Press, 2020. ISBN 9781316471524. Citado na página 13.
- EVANS, R. *The Cosmic Microwave Background*. [S.l.]: Springer International Publishing, 2015. ISBN 978-3-319-09927-9. Citado na página 12.
- FIXSEN, D. J. The temperature of the cosmic microwave background. *The Astrophysical Journal*, v. 707, 12 2009. ISSN 0004-637X. Citado na página 12.
- GAMOW, G. Expanding universe and the origin of elements. *Physical Review*, v. 70, 10 1946. ISSN 0031-899X. Citado na página 12.
- GARNER, R. *James Webb Space Telescope First Flight Mirror Completes Cryogenic Testing*. 2009. Url https://www.nasa.gov/topics/technology/features/mirror_ryo.html. Citado 2 vezes nas páginas 10 e 19.

- GARNER, R. *Hubble Space Telescope*. 2022.
Urlhttps://www.nasa.gov/mission_pages/hubble/story/index.html. Citado 2 vezes nas páginas 10 e 15.
- GIOVANNINI, M. *A primer on the physics of the cosmic microwave background*. [S.l.]: World Scientific Publishing, 2008. Citado na página 13.
- GREICIUS, T. *Planck Spacecraft*. 2012.
Urlhttps://www.nasa.gov/mission_pages/planck/pia13953.html. Citado 2 vezes nas páginas 10 e 18.
- GRISWOLD, B. *COBE Images*. 2010.
Url<https://wmap.gsfc.nasa.gov/media/081000/index.html>. Citado 2 vezes nas páginas 10 e 16.
- HASHIMOTO, M. aki et al. *Big-Bang Nucleosynthesis*. [S.l.]: Springer Singapore, 2018. ISBN 978-981-13-2934-0. Citado na página 12.
- Hinton, S. R. ChainConsumer. *The Journal of Open Source Software*, v. 1, p. 00045, ago. 2016. Citado na página 39.
- HOBSON, M. P. et al. *Bayesian methods in cosmology*. [S.l.]: Cambridge University Press, 2010. ISBN 978-0-511-77015-9. Citado na página 13.
- JAYNES, E. T. *Probability Theory - The logic of science*. [S.l.: s.n.], 1996. Citado na página 21.
- MARCH, M. C. *Advanced Statistical Methods for Astrophysical Probes of Cosmology*. [S.l.]: Springer Berlin Heidelberg, 2013. ISBN 978-3-642-35059-7. Citado 3 vezes nas páginas 13, 34 e 35.
- MARQUES, T. *Radiação Cósmica de Fundo: Características e Atualidades*. [S.l.]: Caderno de Física da UEFES, 2012. Citado na página 15.
- MELO, J. R. P. d. Emissão anômala de micro-ondas na via láctea. 2022. Citado na página 17.
- MONAGHAN, H. *Wilkinson Microwave Anisotropy Probe (WMAP)*. 2021.
Url<https://www.nasa.gov/directorates/heo/scan/services/missions/universe/WMAP.html>. Citado 2 vezes nas páginas 10 e 17.
- MORETTIN, W. O. B. P. A. *Estatística básica*. [S.l.]: Saraiva, 2017. ISBN 978-85-472-2023-5. Citado na página 20.
- O'HAGAN, A. *Bayesian Inference. Kendall's Advanced Theory of Statistics*. [S.l.: s.n.], 1994. Citado 2 vezes nas páginas 20 e 28.
- PAULINO, C. D. et al. *Estatística bayesiana*. 2. ed. [S.l.]: Fundação Calouste Gulbenkian, 2018. Citado 7 vezes nas páginas 10, 13, 21, 22, 25, 27 e 28.
- PENZIAS, A. A.; WILSON, R. W. A measurement of excess antenna temperature at 4080 mc/s. *The Astrophysical Journal*, v. 142, 7 1965. ISSN 0004-637X. Citado na página 12.
- PERIVOLAROPOULOS, L.; SKARA, F. Challenges for λ cdm: An update. *New Astronomy Reviews*, Elsevier, p. 101659, 2022. Citado na página 36.

- PETTINARI, G. W. *The Intrinsic Bispectrum of the Cosmic Microwave Background*. [S.l.]: Springer International Publishing, 2016. ISBN 978-3-319-21881-6. Citado na página 12.
- REST, A. et al. Cosmological constraints from measurements of type ia supernovae discovered during the first 1.5 yr of the pan-starrs1 survey. *The Astrophysical Journal*, v. 795, 10 2014. ISSN 1538-4357. Citado na página 13.
- RIESS, A. G. et al. Milky way cepheid standards for measuring cosmic distances and application to <i>gaia</i> dr2: Implications for the hubble constant. *The Astrophysical Journal*, v. 861, 7 2018. ISSN 1538-4357. Citado 2 vezes nas páginas 12 e 13.
- ROSSMANITH, G. *Non-linear Data Analysis on the Sphere*. [S.l.]: Springer International Publishing, 2013. ISBN 978-3-319-00308-5. Citado na página 12.
- SILVA, J. L. de Castro e; FERNANDES, M. W.; ALMEIDA, R. L. F. de. *Estatística e Probabilidade*. [S.l.]: EdUECE, 2015. ISBN 978-85-7826-439-0. Citado na página 20.
- SULLIVAN, M. et al. Snls3: Constraints on dark energy combining the supernova legacy survey three-year data with other probes. *The Astrophysical Journal*, v. 737, 8 2011. ISSN 0004-637X. Citado na página 13.
- SUZUKI, N. et al. The hubble space telescope cluster supernova survey. v. improving the dark-energy constraints above $z > 1$ and building an early-type-hosted surpernova sample. *The Astrophysical Journal*, v. 746, 2 2012. ISSN 0004-637X. Citado na página 13.
- TEIXEIRA, I. M. et al. Da luneta de galileu ao telescópio espacial hubble: contributos do desenvolvimento tecnológico na divulgação da história da astronomia. *Research, Society and Development*, v. 11, n. 5, p. e26811528214–e26811528214, 2022. Citado na página 14.
- VIELVA, P. et al. Detection of non-gaussianity in the wilkinson microwave anisotropy probe first-year data using spherical wavelets. *The Astrophysical Journal*, IOP Publishing, v. 609, n. 1, p. 22, 2004. Citado na página 16.