



UNIVERSIDADE  
ESTADUAL DE LONDRINA

---

CINTHIA NUNES DE LIMA

**RECONSTRUÇÃO E CALIBRAÇÃO DE RELAÇÕES  
RIQUEZA-MASSA COM COSMODC2-LSST**

---

Londrina  
2023

CINTHIA NUNES DE LIMA

**RECONSTRUÇÃO E CALIBRAÇÃO DE RELAÇÕES  
RIQUEZA-MASSA COM COSMODC2-LSST**

Qualificação apresentada ao Programa de Mestrado em Física da Universidade Estadual de Londrina como requisito parcial para obtenção do título de Mestre em Física.

Orientador: Prof. Dr. Sandro Dias Pinto Vitenti.

Londrina  
2023

Ficha de identificação da obra elaborada pelo autor, através do Programa de Geração Automática do Sistema de Bibliotecas da UEL

D278r Lima, Cinthia Nunes de.  
Reconstrução e calibração de relações riqueza-massa com CosmoDC2-LSST / Cinthia Nunes De Lima. - Londrina, 2024.  
48 f. : il.

Orientador: Sandro Dias Pinto Vitenti.  
Dissertação (Mestrado em Física) - Universidade Estadual de Londrina, Centro de Ciências Exatas, Programa de Pós-Graduação em Física, 2024.  
Inclui bibliografia.

1. Aglomerados de Galáxias - Tese. 2. Relações Riqueza-Massa - Tese. 3. Análise Estatística - Tese. 4. Reconstrução de distribuições - Tese. I. Vitenti, Sandro Dias Pinto . II. Universidade Estadual de Londrina. Centro de Ciências Exatas. Programa de Pós-Graduação em Física. III. Título.

CDU 53

CINTHIA NUNES DE LIMA

**RECONSTRUÇÃO E CALIBRAÇÃO DE RELAÇÕES  
RIQUEZA-MASSA COM COSMODC2-LSST**

Qualificação apresentada ao Programa de Mestrado em Física da Universidade Estadual de Londrina como requisito parcial para obtenção do título de Mestre em Física.

**BANCA EXAMINADORA**

---

Orientador: Prof. Dr. Sandro Dias Pinto Vitenti  
Universidade Estadual de Londrina - UEL

---

Profa. Dra. Mariana Penna Lima Vitenti  
Universidade de Brasília - UnB

---

Prof. Dr. Rodrigo Fernando Lugon Cornejo von  
Martens  
Universidade Federal da Bahia - UFBA

Londrina, 28 de fevereiro de 2024.

## AGRADECIMENTOS

Primeiramente agradeço meus pais Paulo Geovane Souza de Lima e Ana Kátia Nunes de Lima por incentivarem desde minha infância a seguir o caminho do conhecimento e da educação.

Agradeço também a todos meus amigos que me acompanharam durante essa jornada, nos dias bons e ruins, sendo compreensivos e estendendo a mão quando eu mais precisava. Em especial, aos mestrandos e doutorandos da Cosmologia que fizeram meus estudos serem mais divertidos e empolgantes. Serei eternamente grata!

Sou muito grata ao meu orientador Sandro D. P. Vitenti e sua esposa Mariana Penna Lima que acreditaram em mim e me mostraram o caminho para ser uma cientista. Muito obrigada, vocês estarão para sempre guardados num lugar especial do meu coração!

Sou grata também aos professores da banca por disponibilizarem seu tempo para a avaliação do meu trabalho e assim, auxiliarem na lapidação de seu desenvolvimento.

Obrigada à colaboração do DESC-LSST por me permitir participar de um projeto tão grande e importante quanto esse. Também agradeço à Capes pelo apoio financeiro e ao Programa de Pós-Graduação em Física (PPG-Física) e UEL por todo o suporte para minha excelente formação profissional.

Por fim agradeço ao João Antônio, pois seu apoio, incentivo e suporte foram muito importantes nessa etapa da minha vida. Obrigada por tudo que tem feito por mim!

*"She had studied the universe all her life,  
but had overlooked its clearest message: For  
small creatures such as we the vastness is  
bearable only through love."*

Contato  
Carl Sagan

*"When you follow two separate chains  
of thought, Watson, you will find some  
point of intersection which should  
approximate to  
the truth."*

Sherlock Holmes, The Disappearance of  
Lady Frances Carfax.  
**Sir Arthur Conan Doyle**

## RESUMO

LIMA, C. N. **Reconstrução e calibração de relações riqueza-massa com CosmoDC2-LSST**. 2023. 47 f. Qualificação (Mestrado em Física) – Universidade Estadual de Londrina, Londrina, 2023.

A Cosmologia observacional tem ganhado força nas últimas décadas devido aos grandes levantamentos de dados. Os aglomerados de galáxias são um dos objetos mais importantes para esses levantamentos, por serem gigantescas estruturas ligadas gravitacionalmente e assim terem características ligadas intrinsecamente a formação de estruturas em larga escala. Uma propriedade importante, amplamente utilizada como proxy de massa é a riqueza, a medida do número de galáxias membros do aglomerado. Pode-se associar a riqueza à massa de um aglomerado por meio da distribuição riqueza-massa. Neste trabalho estudou-se a calibração da distribuição riqueza-massa, utilizando parametrizações para a média e desvio padrão da distribuição, cujo o formato é supostamente gaussiano, com o objetivo de estudar o comportamento destes modelos. Para isto, utilizou-se dados do catálogo simulado cosmoDC2 para ajustar os valores dos parâmetros desses modelos através de métodos estatísticos de inferência. Além disso, através da divisão em pequenos intervalos, foi realizada uma reconstrução destas relações, que foi usada para fazer uma validação cruzada com os modelos calibrados.

**Palavras-chave:** Análise Estatística; Aglomerados de Galáxias; Relações Riqueza-Massa.

## ABSTRACT

LIMA, C. N. **Richness-mass relations calibration and reconstruction with CosmoDC2-LSST.** 2023. 47 p. Qualification (Master in Physics) – State University of Londrina, Londrina, 2023.

Observational Cosmology has gained strength in recent decades due to big data surveys. Galaxy clusters are one of the most important objects for these surveys, as they are the largest gravitationally-bound structures in the universe and, therefore, have characteristics intrinsically linked to the formation of large-scale structures. An important property, widely used as a mass proxy, is richness, a quantity that estimates the number of galaxies that are members of the cluster. It can be associated with the mass of a cluster through the richness-mass distribution. In this work, we studied the calibration of the richness-mass distribution, using parameterizations for the mean and standard deviation of the distribution, with the aim of studying the behavior of these models. To do this, we used data from the cosmoDC2 simulated catalog to infer the values of the parameters of these models through statistical inference methods. Furthermore, through division into small intervals, a reconstruction of these relations was performed, which was used to cross-validate with the calibrated models.

**Key-words:** Statistic Analysis. Galaxies Clusters. Richness-Mass Relations.

## SUMÁRIO

1	<b>INTRODUÇÃO</b> .....	08
2	<b>AGLOMERADOS DE GALÁXIAS E AS RELAÇÕES MASSA-OBSERVÁVEL</b> .....	12
2.1	<b>Formação de estruturas, função de massa e distribuição de aglomerados</b> .....	12
2.2	<b>Relações de escala</b> .....	15
2.3	<b>Distribuição Massa-Observável</b> .....	16
3	<b>MÉTODOS ESTATÍSTICOS NA INFERÊNCIA DE PARÂMETROS</b> .....	18
3.1	<b>Inferência bayesiana e método de Markov Chain Monte Carlo</b> .....	18
3.2	<b>Estimadores de máxima verossimilhança (MLE)</b> .....	20
4	<b>CALIBRAÇÃO DE RELAÇÕES RIQUEZA-MASSA</b> .....	21
4.1	<b>Relações Riqueza-Massa (<math>\lambda</math>-M)</b> .....	21
4.2	<b>Análise I: Modelo de Ascaso</b> .....	22
4.3	<b>Análise II: Modelo estendido</b> .....	25
4.4	<b>Análise III: Análises em intervalos</b> .....	27
5	<b>CONCLUSÃO</b> .....	36
	<b>REFERÊNCIAS</b> .....	37
	<b>APÊNDICES</b> .....	42
	<b>APÊNDICE A – Probabilidade</b> .....	43
	<b>ANEXOS</b> .....	45
	<b>ANEXO A – Análise III: histogramas dos 200 primeiros e 200 últimos intervalos de <math>\ln \lambda</math></b> .....	46

# 1 INTRODUÇÃO

Desde os primórdios a humanidade tem olhado para o céu e buscado explicações sobre os processos naturais que via ao seu redor. O que são os pontos brilhantes no céu? Porque há dia e noite? O céu é infinito? A curiosidade humana produziu diversas perguntas que ao longo de milênios levaram a discussões que impulsionaram a busca por conhecimento. Hoje, sabe-se que algumas dessas questões foram respondidas, algumas ainda busca-se respostas, outras talvez nunca serão esclarecidas.

A atração pelo céu levou a construção de instrumentos para melhor observá-lo, assim nasceram os telescópios. Com os primeiros telescópios foi possível ver que o planeta Terra era apenas um de vários outros e além disso, havia corpos que também orbitavam esses planetas. Assim, sistemas cada vez mais complexos seriam mais e mais desbravados.

Com o avanço dos telescópios, foi possível enxergar cada vez mais longe. Diversos sistemas astrofísicos novos e maiores foram descobertos, o que gerou debates. O quão grande é o Universo em que vivemos? Alguns pensavam que o Universo se restringia a nossa própria galáxia, outros que nossa galáxia é apenas uma de muitas outras espalhadas pelo cosmos (PIGATTO, 2005). As evidências encontradas mostraram que o segundo grupo estava correto (HUBBLE, 1929b).

Foi alimentado pela vontade de saber mais sobre esses objetos que Edwin Hubble as estudou e classificou. Além disso, ele percebeu que elas se afastavam umas das outras na mesma taxa (figura 1), conhecida como ‘constante’ de Hubble,  $H_0$  (HUBBLE, 1929a). No entanto, a taxa medida por Hubble na época estava bastante imprecisa, cerca de  $500 \text{ km s}^{-1} \text{ Mpc}^{-1}$ , devido a subestimação nos valores das distâncias. Hoje sabe-se que o parâmetro de Hubble evolui com o *redshift* (desvio para o vermelho), sendo  $H_0$  quando o *redshift* é 0 (e.g. Nishizawa, Taruya e Saito (2011), Moresco e Marulli (2017)). Diversas pesquisas vêm tentando medir precisamente  $H_0$ , mensuramentos feitos através de estrelas cefeidas e supernovas do tipo Ia levam a uma faixa de valores de  $73.2 \pm 1.4 \text{ km s}^{-1} \text{ Mpc}^{-1}$  (RIESS et al., 2021), já inferências do Planck apontam para uma faixa de  $67.4 \pm 0.5 \text{ km s}^{-1} \text{ Mpc}^{-1}$  (Planck Collaboration et al., 2020b), o que indica divergências entre diferentes tipos de medidas, veja Valentino et al. (2021).

Com os resultados de Hubble pôde-se concluir que o Universo está em expansão e a partir disso, ao longo das décadas diferentes modelos cosmológicos começaram a ser considerados. Naturalmente criou-se o interesse em esclarecer quais os mecanismos que iniciaram essa expansão, bem como entender a física do Universo primordial.

Em 1917, Einstein publicou o que é considerado por muitos o primeiro modelo cosmológico. O modelo era baseado em uma solução para suas equações de campo da

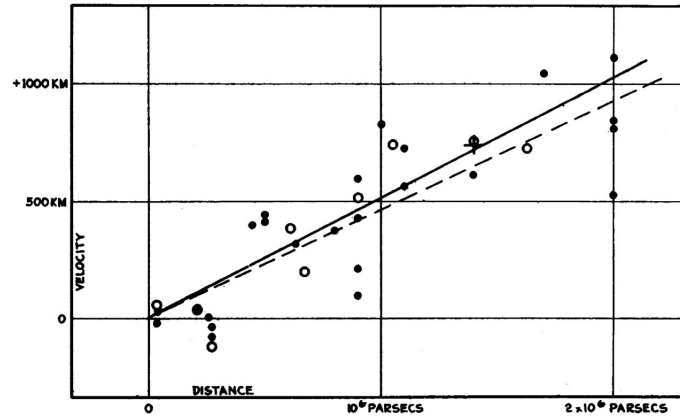


Figura 1 – Gráfico de distância versus velocidade radial de galáxias, publicado por Hubble em 1929.

Relatividade Geral (RG) (EINSTEIN, 1917). Neste foi introduzida uma constante, hoje chamada de constante cosmológica,  $\Lambda$ , para contrabalancear o efeito atrativo da gravidade, afim de modelar um Universo estático, de acordo com o que ele pensava na época.

Atualmente, o modelo mais consolidado é o Modelo Cosmológico Padrão,  $\Lambda$ CDM, que tem como princípio a homogeneidade e isotropia do espaço, em escalas maiores que 100 Mpc, além de considerar a estrutura da RG com a constante cosmológica (CARROLL; PRESS; TURNER, 1992). Medidas observacionais, como por exemplo a observação das distâncias de supernovas tipo Ia (RIESS et al., 1998), mostram uma expansão acelerada do Universo e trazem a necessidade da inclusão, no  $\Lambda$ CDM, de uma componente de energia com pressão negativa, chamada de energia escura, a qual a natureza ainda é desconhecida. Outra questão em aberto que vem da RG, são as evidências, como as curvas de rotação de galáxias espirais (RUBIN; FORD W. KENT, 1970), velocidades de dispersão de galáxias (ZWICKY, 1933) e lenteamento gravitacional fraco (MARKEVITCH et al., 2004), que apontam para um tipo de matéria desconhecida, não observável, que interage apenas de forma gravitacional, denominada matéria escura.

Diversos grandes levantamentos de dados vem trabalhando a mais de 20 anos em buscas de respostas para as questões em aberto, como (Planck Collaboration et al., 2020a), que buscou, através dos dados, informações a respeito do Universo primordial, e SDSS (*Sloan Digital Sky Survey*) (Kollmeier et al., 2019), um grande levantamento espectroscópico do céu, buscando dados de milhões de objetos.

Pesquisas observacionais de grande escala estão cada vez mais entrelaçadas ao desenvolvimento de modelos cosmológicos sólidos. Em 2013, o DES (*Dark Energy Survey*) (Dark Energy Survey Collaboration et al., 2016) entrou em operação. Uma colaboração que contou com pesquisadores de mais de 25 países, incluindo Brasil. De 2013 a 2019, foram levantados os dados de 300 milhões de galáxias e 5000 graus quadrados do céu

com 5 filtros ópticos diferentes <sup>1</sup>. Essas valiosas informações são usadas por inúmeras pesquisas hoje que investigam a natureza da energia e matéria escura através da restrições de parâmetros cosmológicos, e.g. Abbott et al. (2023).

Hoje outras grandes colaborações estão em andamento para investigar a evolução do universo, como DESC-LSST <sup>2</sup> (*Dark Energy Science Collaboration - Rubin Observatory Legacy Survey of Space and Time*) (The LSST Dark Energy Science Collaboration et al., 2018) e Euclid <sup>3</sup> (Euclid Collaboration et al., 2022). O DESC é uma colaboração internacional que pretende usar dados do telescópio LSST para mensurar com alta precisão parâmetros cosmológicos fundamentais. Atualmente, a colaboração possui o catálogo de galáxias sintéticas cosmoDC2 (KORYTOV et al., 2019) que cobre 440 graus quadrados do céu, com *redshift* até  $z=3$ , e tem o objetivo principal de servir de apoio para a ciência de precisão da colaboração.

Um dos objetos de estudo mais importantes desses levantamentos são os aglomerados de galáxias. Formados por coleções de galáxias, esses objetos são uma das maiores estruturas do Universo e são conhecidos por serem "sondas" cosmológicas, isso porque suas propriedades, principalmente sua massa e distribuição espacial, estão intrinsecamente relacionadas com a formação das estruturas em larga escala, como será discutido no próximo capítulo.

Dados de propriedades de aglomerados de galáxias podem ser utilizados para encontrar limites nos modelos cosmológicos (e.g. Abdullah et al. (2023)), mas, medidas observacionais contêm erros intrínsecos que precisam ser considerados. Além disso, a modelagem estatística deve ser robusta e coerente com o comportamento dos dados. Erros na análise estatística enviesam as estimativas e isso pode resultar em conclusões errôneas a respeito de modelos cosmológicos.

Algumas propriedades de aglomerados que são utilizadas para testar modelos cosmológicos são massa, *redshift* e riqueza. A massa não é um observável direto, para ser mensurada, frequentemente é necessário utilizar sua relação com outras propriedades (observáveis), que são denominadas *proxies*. Uma propriedade importante que pode ser utilizada como *proxy* de massa é a riqueza (e.g. Andreon (2010)), ela é a medida da quantidade de galáxias de um certo tipo, membros do aglomerado. Os critérios de medida de riqueza não são bem definidos visto que ela depende do critério de identificação do aglomerado, tendo significados distintos para diferentes referências, por exemplo, Chiu et al. (2020) usam a definição de riqueza como a sobredensidade de galáxias de sequência vermelha, já Croft et al. (1997) define a riqueza como o número ponderado de galáxias dentro de um raio de contagem em um intervalo de magnitude. Neste trabalho, considera-se riqueza a

<sup>1</sup> [www.darkenergysurvey.org/](http://www.darkenergysurvey.org/)

<sup>2</sup> [lsstdesc.org/](http://lsstdesc.org/)

<sup>3</sup> [www.esa.int/Science\\_Exploration/Space\\_Science/Euclid\\_overview](http://www.esa.int/Science_Exploration/Space_Science/Euclid_overview)

quantidade de partículas de matéria escura associadas ao halo (aglomerado) identificado via *Spherical Overdensity* (SO) da simulação de N-Corpos.

Pode-se associar a riqueza à massa de um aglomerado por meio da distribuição riqueza-massa, que é a distribuição de probabilidade das medidas de riqueza condicionada aos dados de massa e *redshift*. A gaussianidade da distribuição massa-observável é amplamente proposta, desde Lima e Hu (2005), em que motivam esse formato pela dispersão dos observáveis em relações de escala entre medidas de propriedades de aglomerados. Algumas dessas pesquisas propõe diferentes parametrizações para a média e variância da distribuição (e. g. Saro et al. (2015) e Murata et al. (2018)). Deve-se ter cuidado ao assumir a normalidade dessa distribuição, visto que há indícios de uma não-gaussianidade que deve ser considerada para observáveis que possuem grande dispersão, como os de riqueza, sendo mais perceptível para valores mais altos de massa e *redshift* dos aglomerados, veja Shaw, Holder e Dudley (2010).

Neste trabalho estudou-se a calibração da distribuição riqueza-massa utilizando o catálogo simulado cosmoDC2 assumindo o modelo de , como o de Lima e Hu (2005), e usando parametrizações para a média e desvio padrão da distribuição, com o objetivo de estudar o comportamento destes modelos. Para isto, primeiramente foram ajustados os valores para os parâmetros desses modelos através de métodos estatísticos de inferência. Depois, através da divisão em pequenos intervalos dos dados de massa e *redshift*, foi realizada uma reconstrução destas relações, que foi usada para fazer uma validação cruzada com os modelos calibrados.

Para a calibração das relações riqueza-massa foram retirados dados de riqueza, massa e *redshift* do cosmoDC2 disponíveis no NERSC<sup>4</sup>(*National Energy Research Scientific Computing Center*), um supercomputador localizado nos Estados Unidos a qual a equipe tem acesso por meio do DESC-LSST. Para auxiliar nos cálculos numéricos, foi utilizado a *Numerical Cosmology Library*<sup>5</sup> (NumCosmo), que é uma biblioteca numérica de cosmologia com diversas ferramentas que podem ser usadas para análises de modelos estatísticos e cálculos cosmológicos.

O conteúdo dessa dissertação está dividido da seguinte forma: No capítulo 2, é feita uma breve revisão sobre como a formação de estruturas no universo está relacionada à abundância de aglomerados de galáxias e como liga-se suas propriedades observáveis à massa através de relações massa-observáveis. No capítulo 3, é tratado sobre os métodos estatísticos usados na inferência de parâmetros. No capítulo 4 são descritas as análises e os resultados obtidos e no 5 a conclusão do trabalho.

---

<sup>4</sup> <https://www.nersc.gov/>

<sup>5</sup> <https://numcosmo.github.io/>

## 2 AGLOMERADOS DE GALÁXIAS E AS RELAÇÕES MASSA-OBSERVÁVEL

Os aglomerados de galáxias são grandes estruturas supermassivas formadas por matéria escura, gás no meio intergalático e galáxias. Suas propriedades podem ser ligadas à evolução das estruturas em grandes escalas através da função de massa, que nos dá o número de aglomerados em uma região. Neste capítulo é feita uma breve revisão sobre a formação de estruturas e sua ligação com a contagem de aglomerados e como associamos as propriedades observacionais dos aglomerados à massa por via de relações de escala e de distribuições massa-observáveis.

### 2.1 Formação de estruturas, função de massa e distribuição de aglomerados

Segundo o modelo  $\Lambda$ CDM, o Universo apresenta uma homogeneidade espacial em grandes escalas. Mas para escalas menores deve-se considerar a existência de variações de densidade conforme a localidade  $\vec{x}$  de um ponto no espaço. As perturbações de densidade de matéria são descritas pelo contraste de densidade  $\delta(\vec{x})$ , que é a variação da densidade de matéria em um certo ponto  $\rho(\vec{x})$  em relação a densidade média de matéria do Universo  $\rho_m$ ,

$$\delta(\vec{x}) := \frac{\rho(\vec{x}) - \rho_m}{\rho_m}. \quad (2.1)$$

A formação de estruturas ocorre em consequência da evolução do contraste de densidade de matéria. Devido à interação gravitacional a densidade em um ponto do espaço pode aumentar ao passo que em outra região ela diminui, em razão disso as flutuações de densidade acabam dando origem a estruturas cosmológicas. Essas flutuações foram originadas no Universo primordial, onde havia flutuações quânticas em um campo de densidade quântico aleatório, onde posteriormente deram origem as perturbações primordiais no campo de matéria, que são a condição inicial para a evolução de  $\delta(\vec{x})$  (DODELSON, 2003).

O valor mensurado do contraste de densidade em um ponto do espaço é uma realização de uma distribuição gaussiana (isso é assumido considerando o grande número de flutuações de densidade e o teorema do limite central), assim, o campo de densidade pode ser descrito através dos momentos da distribuição, a média e a variância. Tomando o valor esperado na equação 2.1, chega-se que  $\langle \delta(\vec{x}) \rangle = 0$ . Considerando  $\delta(\vec{x})$  para dois pontos do espaço separados por uma distância  $\vec{r} = |\vec{x} - \vec{x}'|$ , a covariância é calculada por meio de uma função de correlação de 2-pontos, em que consideramos a hipótese de isotropia estatística para que seja válida uma dependência apenas na distância de

separação entre os pontos,

$$\xi_{\delta\delta'} = \langle \delta(\vec{x})\delta^*(\vec{x}') \rangle. \quad (2.2)$$

Através do coeficiente  $\xi_{\delta\delta'}$  é possível estimar o quanto os objetos estão agrupados em uma região do espaço e assim identificar regiões de sobredensidade.

Considerando o espaço de Fourier,  $\xi_{\delta\delta'}$  pode ser calculado tomando a transformada de Fourier de  $\delta(\vec{x})$ ,

$$\delta(\vec{k}) = \int_{-\infty}^{\infty} d^3\vec{x} \delta(\vec{x}) e^{-i\vec{k}\vec{x}}. \quad (2.3)$$

Assim,

$$\begin{aligned} \langle \delta(\vec{k})\delta^*(\vec{k}') \rangle &= \left\langle \int_{-\infty}^{\infty} d^3\vec{x} d^3\vec{x}' \delta(\vec{x}) e^{-i\vec{k}\vec{x}} \delta(\vec{x}') e^{i\vec{k}'\vec{x}'} \right\rangle \\ &= \left\langle \int_{-\infty}^{\infty} d^3\vec{x} d^3\vec{x}' \delta(\vec{x})\delta(\vec{x}') e^{-i\vec{k}\vec{x}} e^{i\vec{k}'(\vec{r}+\vec{x})} \right\rangle \\ &= \int_{-\infty}^{\infty} d^3\vec{x} d^3\vec{x}' \xi_{\delta\delta'}(\vec{r}) e^{i\vec{x}(\vec{k}'-\vec{k})} e^{i\vec{k}'\vec{r}}. \end{aligned} \quad (2.4)$$

Como a delta de Dirac é  $\tilde{\delta}^3(\vec{k}' - \vec{k}) = \frac{1}{(2\pi)^3} \int_{-\infty}^{\infty} d^3\vec{x} e^{i\vec{x}(\vec{k}'-\vec{k})}$  e a transformada de Fourier de  $\xi_{\delta\delta'}$ , chamada de espectro de potência da matéria  $P(k)$ , é

$$P(\vec{k}) = \int_{-\infty}^{\infty} d^3\vec{x} \xi_{\delta\delta'} e^{i\vec{k}\vec{r}}, \quad (2.5)$$

logo pode ser observado que o coeficiente de correlação de 2-pontos do contraste de densidade no espaço de Fourier é

$$\langle \delta(\vec{k})\delta^*(\vec{k}') \rangle = (2\pi)^3 \tilde{\delta}^3(\vec{k}' - \vec{k}) P(k). \quad (2.6)$$

O contraste de densidade média filtrado é definido como

$$\delta_R(\vec{x}) = \int_{-\infty}^{\infty} d^3\vec{x}' \delta(\vec{x}') W(|x' - x|). \quad (2.7)$$

Em que  $W$  é um filtro do tipo cartola, que suaviza o campo de densidade média dentro de uma região  $R$ ,

$$W(r) = \begin{cases} \frac{1}{4/3\pi r^3}, & r \leq R \\ 0, & r > R. \end{cases} \quad (2.8)$$

Assim, observando 2.4, 2.5 e 2.9, a variância filtrada do campo de densidade média é

$$\begin{aligned} \sigma_R^2 &= \langle \delta(\vec{x})^2 \rangle = \int_{-\infty}^{\infty} d^3\vec{x}' \langle \delta(\vec{x}')^2 \rangle W^2(|x' - x|) \\ &= \int_{-\infty}^{\infty} d^3\vec{x}' W^2(|x' - x|) \int_{-\infty}^{\infty} \frac{d^3\vec{k}}{(2\pi)^3} e^{i\vec{k}\vec{x}'} P(k) \\ &= \int_{-\infty}^{\infty} \frac{d^3\vec{k}}{(2\pi)^3} P(k) \int_{-\infty}^{\infty} d^3\vec{x}' W^2(|x' - x|) e^{i\vec{k}\vec{x}'} \\ &= \int_{-\infty}^{\infty} \frac{d^3\vec{k}}{(2\pi)^3} P(k) \tilde{W}^2(|k' - k|), \end{aligned} \quad (2.9)$$

em que  $\tilde{W}$  é a função cartola no espaço de Fourier.

Então, dada a variância filtrada, equação 2.9 e a média nula, contraste de densidade filtrado tem a distribuição  $p(\delta_R)$ ,

$$p(\delta_R) = \frac{1}{\sqrt{2\pi\sigma_R^2}} \exp\left[-\frac{\delta_R^2}{2\sigma_R^2}\right]. \quad (2.10)$$

Segundo o formalismo de Press-Schechter, a formação de uma estrutura ligada está condicionada a  $\delta_R$ , em uma região de raio  $R$ , ser maior que uma densidade crítica  $\delta_{crit}$ . A fração de objetos colapsados com massa maior que  $M$  ( $M = \frac{4\pi}{3} R^3 \rho_m$ ) é dado pela função de seleção (PRESS; SCHECHTER, 1974)

$$F(> M) = \int_{\delta_{crit}}^{\infty} p(\delta_R) d\delta_R = \frac{1}{2} \left[ 1 - \operatorname{erf}\left(\frac{\delta_{crit}}{\sqrt{2}\sigma_R}\right) \right]. \quad (2.11)$$

A função de seleção está relacionada à distribuição de aglomerados em uma região através de função de massa  $n(M, z)$ , que descreve o número de aglomerados com massa  $M$  e *redshift*  $z$  em um certo volume do espaço (PENNA-LIMA, 2010). Assim,

$$\frac{dn(M, z)}{dM} = -\frac{\rho_m}{M} \frac{dF(> M)}{dM}. \quad (2.12)$$

De forma geral pode-se calcular a função de massa através de outros formalismos que descrevem o processo de colapso de estruturas gravitacionalmente ligadas através da função de multiplicidade  $f(\sigma_R(z))$ . Neste caso, a relação 2.12 pode ser escrita como (e.g. Penna-Lima (2010))

$$\frac{dn(M, z)}{dM} = -f(\sigma_R(z)) \frac{\rho_m}{M\sigma_R(z)} \frac{d\sigma_R(z)}{dM}. \quad (2.13)$$

Por exemplo, uma função de multiplicidade muito conhecida é a de Jenkins (JENKINS et al., 2001), obtida por meio de simulações da função de massa, cuja forma é dada por

$$f_J(\sigma_R) = A \exp[-|\ln(\sigma_R)^{-1} + B|^\epsilon]. \quad (2.14)$$

Em que  $A$ ,  $B_c$  e  $\epsilon$  são coeficientes ajustados através da simulação.

Algo importante a se destacar é que na equação 2.13, ao mesmo tempo que se verifica a “universalidade” da função de massa, observa-se que a cosmologia e a informação sobre o processo de evolução da formação de estruturas está implícito em  $\sigma_R(z)$ , ao passo que o formalismo de colapso em estruturas ligadas gravitacionalmente está contida na função de multiplicidade  $f(\sigma_R(z))$ , de forma que pode-se evidenciar uma ligação entre a contagem de aglomerados no céu e a formação de estruturas no Universo.

A função de massa tem uma papel fundamental na Cosmologia com aglomerados, através dela pode fazer a contagem de aglomerados de forma teórica. A densidade numérica de aglomerados em um intervalo de massa (PENNA-LIMA, 2010; LIMA; HU, 2005; SHAW; HOLDER; DUDLEY, 2010) é dada pela equação:

$$\bar{N} = \int_{M_i}^{M_{i+1}} d \ln M \frac{dn(M, z)}{d \ln M}. \quad (2.15)$$

Integrando num volume espacial, obtém-se o número médio de aglomerados com massa  $M > M_{min}$  em um intervalo de redshift  $[z_{min}, z_{max}]$  (PENNA-LIMA, 2010):

$$\bar{n}(M > M_{min}, z) = \int d\Omega \int_{z_{min}}^{z_{max}} dz \frac{D_A^2(z)}{H(z)} \int_{M_{min}}^{\infty} d \ln M \frac{dn(M, z)}{d \ln M}. \quad (2.16)$$

Em que:  $D_A(z)$  é a distância de diâmetro-angular;  $H(z)$  é o parâmetro de Hubble;  $\Omega$  é o angulo sólido,  $d\Omega = \sin \theta d\theta d\phi$ .

Pode-se notar que os primeiros termos da equação 2.16 dependem da geometria do espaço-tempo, já o último termo depende da função de massa (que carrega implicitamente informações sobre a cosmologia). Logo, fica explícito que a contagem de aglomerados em um volume do espaço depende dos modelos teóricos de Universo. Assim, como é uma quantidade que pode ser também medida observacionalmente, pode ser usada para testá-los. A distribuição teórica pode ser vinculada a observacional através da distribuição de massas observadas, que é a convolução da função de massa com a distribuição massa-observável (SHAW; HOLDER; DUDLEY, 2010):

$$\frac{dn(M, z)}{d \ln M_{obs}} = \int_0^{\infty} d \ln M \frac{dn(M, z)}{d \ln M} P(M_{obs}|M). \quad (2.17)$$

Em que  $P(M_{obs}|M)$  é a distribuição massa-observável.

Para a consolidação dos modelos teóricos se faz necessário a busca por observações que os validem. Como pode-se notar pela função de massa, a massa dos aglomerados é uma importante propriedade dos aglomerados e através de distribuições massa-observáveis é possível vincular as medidas observacionais com os modelos teóricos. Porém a massa não é medida diretamente, sua estimativa muitas vezes é feita por meio de outras quantidades observacionais, chamadas *proxies*. Nas próxima seção discutir-se-á a respeito de como pode-se relacionar essas quantidades mediante relações massa-observáveis.

## 2.2 Relações de escala

A massa dos aglomerados é uma propriedade que não pode ser medida diretamente, sendo necessários *proxies* de massa, isto é, observáveis que indicam o valor dessa quantidade. Entre esses *proxies*, podem ser destacados o decréscimo do fluxo de Sunyaev-Zel'dovich (e.g. Nagai (2006)), temperatura de raios-X (e. g. Ikebe et al. (2002)), riqueza

(e. g. Andreon (2010) ) e o efeito de cisalhamento provindo de lentes gravitacionais fracas (e. g. Hoekstra et al. (2013)). Algumas destas propriedades físicas e a massa de aglomerados podem ser correlacionadas através das relações de escala.

As relações de escala correlacionam importantes propriedades físicas dos aglomerados de galáxias. Uma relação de escala pode ser definida como uma relação de proporcionalidade entre uma quantidade e outra que a represente. No caso dos aglomerados, as relações massa-observáveis relacionam a massa, que é quantidade intrínseca, ou seja, não pode ser observada diretamente, com uma quantidade observável.

De forma geral, as relações massa-observável podem ser escritas através de leis de potência, ou seja, têm a forma

$$P = aM^b. \quad (2.18)$$

$P$  e  $M$  são o *proxy* de massa e a massa do aglomerado respectivamente e,  $a$  e  $b$  são os coeficientes da relação de escala.

Neste trabalho o foco será nos dados de massa e riqueza. Nem sempre está disponível um grande número de dados observacionais, por isso, simulações de N-Corpos são muito utilizadas para a calibração dos modelos.

Para a análise de relações riqueza-massa, neste trabalho foram utilizados dados obtidos por meio de simulações, que estão disponíveis em catálogos acessados através do NERSC. Nestas simulações são necessários algoritmos que identifiquem os halos formados e suas massas, chamados de *halo finders*. Dois dos mais utilizados são *Friends-Of-Friends* (FOF) (HUCHRA; GELLER, 1982; DAVIS et al., 1985) e *Spherical Overdensity* (SO) (LACEY; COLE, 1994). O FOF tem como princípio reconhecer partículas ligadas entre si se a distância entre elas estiver abaixo de um limiar denominado "comprimento de ligação". Já SO identifica halos analisando uma região esférica, o halo é identificado quando ao variar o tamanho da região, a densidade dentro atinge um valor maior que uma densidade de referência.

Para a modelagem dos dados são utilizados modelos estatísticos, com esses modelos e métodos estatísticos de inferência é possível estimar parâmetros cosmológicos e suas incertezas. Na próxima seção será introduzida a distribuição massa-observável, cujo formato e sua calibração (estimativa de seus parâmetros) são o objeto de estudo desse trabalho.

## 2.3 Distribuição Massa-Observável

Para estudar os aglomerados de galáxias e suas propriedades físicas é necessário medir quantidades observáveis através de instrumentos astronômicos. Essas medidas possuem erros intrínsecos, de forma que se faz necessário o uso de uma estatística apropriada para a análise desses dados.

Tomando as relações massa-observável, que podem ser escritas como

$$q_{obs} = aM^b, \quad (2.19)$$

em que  $a$  e  $b$  são parâmetros de escala,  $q_{obs}$  é um observável e  $M$  é a massa medida do aglomerado. Pode-se linearizar aplicando um logaritmo, chegando em

$$\log(q_{obs}) = \log(a) + b \log(M). \quad (2.20)$$

Considerando  $q_{obs}$  uma realização ruidosa de uma distribuição de medidas  $Q_{obs}$ , 2.21 pode ser escrito como

$$\langle \log(Q_{obs}) | M \rangle = \log(a) + b \log(M). \quad (2.21)$$

Na literatura a distribuição de  $Q_{obs}$  é proposta como uma gaussiana em  $\log(Q_{obs})$  (LIMA; HU, 2005), ou seja, a distribuição massa-observável tem a forma

$$P(\log(Q_{obs}) | M) = \frac{1}{Q_{obs} \tilde{\sigma} \sqrt{2\pi}} \exp \left[ -\frac{(\log(Q_{obs}) - \langle \log(Q_{obs}) | M \rangle)^2}{2\tilde{\sigma}^2} \right]. \quad (2.22)$$

Através da distribuição massa-observável é possível construir a função verossimilhança dos dados. Essa função nos mostra o quanto um conjunto de dados é provável considerando o conjunto de parâmetros em um determinado modelo. Assim, ela é usada por meio de métodos estatísticos na estimativa de parâmetros de modelos estatísticos e cosmológicos (e. g. (ABDULLAH et al., 2023)). No próximo capítulo é feita uma breve revisão sobre dois desses métodos, que foram utilizados no trabalho.

### 3 MÉTODOS ESTATÍSTICOS NA INFERÊNCIA DE PARÂMETROS

#### 3.1 Inferência bayesiana e método de *Markov Chain Monte Carlo*

A estatística bayesiana tem como base a interpretação de que a probabilidade expressa o grau de confiança em um evento. Este grau de confiança pode ser baseado em conhecimentos prévios sobre o evento. Assim, a inferência bayesiana consiste em medir a probabilidade de uma hipótese usando os dados e as informações anteriores (GELMAN et al., 2004).

Considere um conjunto de dados  $d = d_i (i = 1, 2, \dots)$  e uma hipótese  $\theta_0$  dada por um conjunto de parâmetros  $\omega$ ,  $\theta_0 = \omega_i (i = 1, 2, \dots)$ , que modela os dados. Pelo teorema de Bayes (Apêndice B), pode-se relacionar a verossimilhança dos nossos dados  $\mathcal{L}(d|\theta_0)$  à posterior, que é a probabilidade da hipótese  $\theta_0$  condicionada aos dados mensurados,  $P(\theta_0|d)$ . Assim,

$$P(\theta_0|d) = \frac{\mathcal{L}(d|\theta_0)P(\theta_0)}{P(d)}. \quad (3.1)$$

Em que  $P(d)$  é a probabilidade dos dados, chamada de evidência.

Uma observação importante é que a verossimilhança é uma estatística que mostra o quão verossímeis são o conjunto de dados dado uma hipótese e o prior  $P(\theta_0)$  é a informação à priori sobre o quão provável é a hipótese.

Estamos interessados na hipótese que possui a maior probabilidade, assim pode-se fornecer uma nova hipótese, calcular a probabilidade e compará-la a anterior. Fornecendo outra hipótese  $\theta_1 = \lambda_i (i = 1, 2, \dots)$ , pode-se comparar a posterior de  $\theta_0$  à de  $\theta_1$  calculando a razão entre as duas, ou seja,

$$\frac{P(\theta_1|d)}{P(\theta_0|d)} = \frac{\mathcal{L}(d|\theta_1)P(\theta_1)}{\mathcal{L}(d|\theta_0)P(\theta_0)}. \quad (3.2)$$

De forma um pouco menos simplista pode-se utilizar métodos de cadeia de Markov para estimar a distribuição posterior. Uma cadeia de Markov é um processo estocástico, formado por um conjunto de estados  $\{x_0, x_1, x_2, \dots\}$ , em que, dado um tempo  $t$ , somente o atual estado  $x_t$  afeta o estado futuro  $x_{t+1}$ , ou seja, a probabilidade de um estado  $x_{t+1}$  está condicionada ao estado anterior  $x_t$ ,  $P(x_{t+1}|x_t)$ . Ao considerar todo  $t = 0, 1, 2, \dots$ , pode se construir uma matriz de transição de probabilidades, em que as entradas são dadas por  $p_{ij} = P(x_{t+1} = j|x_t = i)$ ,  $i, j \in S = \text{espaço dos estados}$ .

As cadeias de Markov são utilizadas por meio do método de *Markov Chain Monte Carlo* (MCMC). Este método tem como resultado a amostragem de uma distribuição de

probabilidade. Uma classe de MCMC são os algoritmos de Metropolis-Hastings (HASTINGS, 1970), que tem como base as etapas a seguir:

1. Definir uma distribuição proposta  $g$ .
2. Definir uma hipótese inicial  $\theta_0$  pertencente ao domínio de  $g$ .
3. Iteração:
  - a) Gere um candidato a partir de  $g$  cujo valor seja dependente do anterior,  $\theta_{t+1}$ .
  - b) Gere um número  $u$  a partir de uma distribuição uniforme  $U(0, 1)$ .
  - c) Calcule a taxa de aceitação:

$$\alpha(\theta_t, \theta_{t+1}) = \min \left( \frac{P(\theta_{t+1}) g(\theta_t|\theta_{t+1})}{P(\theta_t) g(\theta_{t+1}|\theta_t)}, 1 \right). \quad (3.3)$$

Em que  $P(\theta)$  é distribuição alvo.

- d) Se  $u \leq \alpha(\theta_t, \theta_{t+1})$ , aceite  $\theta_{t+1}$ . Caso contrário faça  $\theta_{t+1} = \theta_t$
4. Repita a iteração até atingir a acurácia desejada.

Uma limitação que ocorre para algoritmos mais simples é que a matriz de transição depende apenas do ponto atual e não utiliza outras informações para gerar um novo ponto. Isso é contornado usando muitas cadeias paralelas combinadas. Alguns algoritmos fazem isso, como o APES (VITENTI; BARROSO, 2023) e o Affine Invariant MCMC Ensemble Sampler (GOODMAN; WEARE, 2010).

Com a posterior também é possível construir intervalos de confiança que indicam as margens de incertezas dos parâmetros. Isso pode ser feito escolhendo um parâmetro e marginalizando sobre os demais parâmetros, ou seja, dado um intervalo  $[\delta_{mín}, \delta_{máx}]$  que contém um parâmetro  $\omega_0$  da hipótese  $\theta_0$ , o nível de confiança desse parâmetro nessa região é

$$\int_{\delta_{mín}}^{\delta_{máx}} P(\omega_0|d) d\omega_0 = 1 - \alpha, \quad (3.4)$$

em que

$$P(\omega_0|d) = \int_{\theta_0} P(\theta_0|d) d\omega_1 d\omega_2 \dots \quad (3.5)$$

é a distribuição marginal de  $\omega_0$  e  $\alpha$  é um valor que indica a probabilidade fora da região de confiança.

### 3.2 Estimadores de máxima verossimilhança (MLE)

A verossimilhança  $\mathcal{L}$  é a probabilidade de obter o conjunto de dados  $x = \{x_i\}$  dado o conjunto de parâmetros  $\theta = \{w_i\}$ ,

$$\begin{aligned}\mathcal{L}(x|\theta) &= p(x_1|\theta)p(x_2|\theta)\dots \\ &= \prod_{i=1}^n p(x_i|\theta).\end{aligned}\tag{3.6}$$

Considerando que a verossimilhança domina a estimativa da probabilidade posterior (LUPTON, 1993), pode-se estimar os parâmetros de  $\theta$  que resultam na máxima probabilidade dos dados.

Problemas de maximização podem ser muito complicados de se calcular computacionalmente, assim, frequentemente utiliza-se a minimização de  $\Upsilon$ ,

$$\Upsilon = -2 \ln(\mathcal{L}(x|w)).\tag{3.7}$$

Isso é possível pois as funções  $\mathcal{L}$  e  $\ln\mathcal{L}$  são monotonicamente relacionadas e a mesma estimativa de máxima verossimilhança é obtida através de qualquer uma delas, sendo mais vantajosa a segunda por ter em seu cálculo um somatório ao invés de um produtório. Além disso, quando a verossimilhança vem de uma distribuição gaussiana, esse método se torna um problema de mínimos quadrados,  $\Upsilon = \chi^2$ ,

$$\chi^2 = \left( \sum_j^n \frac{(x_j - f_j(\theta))^2}{f_j(\theta)} \right).\tag{3.8}$$

Em que  $f_j(\theta)$  é o modelo teórico com parâmetros  $\theta$ .

Então os estimadores de máxima verossimilhança (em inglês *Maximum Likelihood Estimators*, MLE) podem ser definidos como

$$\hat{w}_i = \arg \min_{w_i \in \theta} \{-2 \ln(\mathcal{L}(x|w_i))\}.\tag{3.9}$$

Analiticamente, calcula-se a minimização através da diferenciação de  $\Upsilon$  em relação ao parâmetro que se quer estimar, ou seja,

$$\left. \frac{\partial \ln \mathcal{L}}{\partial w} \right|_{w=w_i} = 0.\tag{3.10}$$

Esta equação pode não ter uma forma fechada para a estimativa, assim, em muitos casos é necessário calcular o MLE numericamente.

## 4 CALIBRAÇÃO DE RELAÇÕES RIQUEZA-MASSA

Este capítulo contém a descrição da calibração e análises deste trabalho, que foram feitas na distribuição Riqueza-Massa. Para a calibração foram usados os dados de riqueza  $\lambda$ , *redshift*  $z$  e massa  $M$  do catálogo cosmoDC2, dos halos com densidade 200 vezes maior que a densidade crítica do Universo<sup>1</sup>, identificados via SO diretamente da simulação de N-Corpos. Considerou-se um modelo de parametrização linear, dependente do logaritmo natural da massa  $\ln M$  e  $z$ , e outro com termos quadráticos, como é discutido na próxima seção.

### 4.1 Relações Riqueza-Massa ( $\lambda$ -M)

Considerando uma distribuição massa-observável como na equação 2.22, a distribuição riqueza-massa pode ser escrita da seguinte forma:

$$P(\ln(\lambda)|M, z) = \frac{1}{\sigma_{\ln \lambda} \sqrt{2\pi}} \exp \left[ -\frac{(\ln(\lambda) - \langle \ln(\lambda)|M, z \rangle)^2}{2\sigma_{\ln \lambda}^2} \right]. \quad (4.1)$$

Segundo o modelo de Ascaso, Ascaso et al. (2016), a média  $\langle \ln \lambda | M, z \rangle$  da distribuição pode ser parametrizada considerando uma dependência no *redshift* e na massa do tipo:

$$\langle \ln \lambda | M, z \rangle = \mu_0 + \mu_1 \ln \left( \frac{M}{M_0} \right) + \mu_2 \ln \left( \frac{1+z}{1+z_0} \right). \quad (4.2)$$

De mesma forma, o desvio padrão  $\sigma_{\ln \lambda}$  pode ser parametrizado como

$$\sigma_{\ln \lambda} = \sigma_0 + \sigma_1 \ln \left( \frac{M}{M_0} \right) + \sigma_2 \ln \left( \frac{1+z}{1+z_0} \right). \quad (4.3)$$

Em que,  $\mathbf{p}_1 = \{\mu_0, \mu_1, \mu_2, \sigma_0, \sigma_1, \sigma_2\}$  é o vetor de parâmetros deste modelo de relação riqueza-massa,  $M_0$  é uma massa *pivot*,  $M_0 = 3.0 \times 10^{14} M_\odot$  e  $z_0$  é um *redshift pivot*,  $z_0 = 0.6$ .

Neste trabalho, além dos termos lineares, também considerou-se uma parametrização estendida com termos quadráticos, além de um termo cruzado entre massa e *redshift*, afim de comparar os dois tipos de parametrização e o impacto que termos de ordem superior teriam no formato da distribuição. Assim, o segundo modelo de parametrização para

<sup>1</sup> A densidade crítica do Universo  $\rho_c$  é uma quantidade definida, a partir das equações de Friedmann, como

$$\rho_c \equiv \frac{3H_0^2}{8\pi G},$$

em que  $H_0$  é a constante de Hubble e  $G$  é a constante gravitacional universal de Newton. Veja sobre em Carroll (2014), capítulo 8.

<b>Tabela 1 - Ascaso</b>			
	<b>Inicial</b>	<b>Melhor ajuste MLE</b>	<b>MCMC</b>
$\mu_0$	3.190	4.454	$4.454 \pm 0.005$
$\mu_1$	0.868	0.689	$0.873 \pm 0.001$
$\mu_2$	-0.304	1.237	$0.042 \pm 0.002$
$\sigma_0$	0.000	0.227	$1.773 \pm 0.001$
$\sigma_1$	0.000	-0.051	$-0.141 \pm 0.0007$
$\sigma_2$	0.000	-0.125	$0.142 \pm 0.001$

a média e desvio padrão avaliado é:

$$\begin{aligned} \langle \ln \lambda | M, z \rangle = & \mu_0 + \mu_{M1} \ln \left( \frac{M}{M_0} \right) + \mu_{M2} \left( \ln \left( \frac{M}{M_0} \right) \right)^2 + \\ & + \mu_{z1} \ln \left( \frac{1+z}{1+z_0} \right) + \mu_{z2} \left( \ln \left( \frac{1+z}{1+z_0} \right) \right)^2 + \\ & + \mu_{Mz} \ln \left( \frac{1+z}{1+z_0} \right) \ln \left( \frac{M}{M_0} \right) \end{aligned} \quad (4.4)$$

e

$$\begin{aligned} \sigma_{\ln \lambda} = & \sigma_0 + \sigma_{M1} \ln \left( \frac{M}{M_0} \right) + \sigma_{M2} \left( \ln \left( \frac{M}{M_0} \right) \right)^2 + \\ & + \sigma_{z1} \ln \left( \frac{1+z}{1+z_0} \right) + \sigma_{z2} \left( \ln \left( \frac{1+z}{1+z_0} \right) \right)^2 + \\ & + \sigma_{Mz} \ln \left( \frac{1+z}{1+z_0} \right) \ln \left( \frac{M}{M_0} \right). \end{aligned} \quad (4.5)$$

Em que,  $\mathbf{p}_2 = \{\mu_0, \mu_{M1}, \mu_{M2}, \mu_{z1}, \mu_{z2}, \mu_{Mz}, \sigma_0, \sigma_{M1}, \sigma_{M2}, \sigma_{z1}, \sigma_{z2}, \sigma_{Mz}\}$  é o vetor com os 12 parâmetros do modelo estendido.

## 4.2 Análise I: Modelo de Ascaso

A inferência dos parâmetros  $\mathbf{p}_1$  foi realizada utilizando os métodos descritos no capítulo 3. Tanto para os MLE quanto para o MCMC foram utilizados métodos implementados na NumCosmo. A verossimilhança (Equação 3.6) foi calculada usando a distribuição  $P(\ln(\lambda)|M, z, \mathbf{p})$  (Equação 4.1) e os momentos do modelo de parametrização (para a análise I, equações 4.2 e 4.3) :

$$\mathcal{L}(\ln \lambda | \mathbf{p}) = \prod_{i=1}^n P(\ln(\lambda_i) | M, z, \mathbf{p}). \quad (4.6)$$

Primeiramente, foram escolhidos valores iniciais para os parâmetros contidos, veja tabela 1, depois esses parâmetros foram deixados livres. Utilizando o conjunto completo de dados foi encontrado o melhor ajuste para os dois modelos, via MLE, também disponíveis na tabela 1. Depois esses valores foram utilizados como valores iniciais para o MCMC.

Para encontrar os limites nos parâmetros dos modelos, utilizou-se o algoritmo NcmFitESMCMC em conjunto com o algoritmo APES (VITENTI; BARROSO, 2023),

implementados na NumCosmo, em que o algoritmo calculou 300 cadeias, considerando o modelo de Ascaso. Para gerar as amostras da posterior além da verossimilhança (Equação 4.6) foi utilizado um prior uniforme para os parâmetros. Os resultados constam na Tabela 1 e as regiões de confiança de 68% e 95% para os parâmetros inferidos podem ser vistas no gráfico da figuras 2.

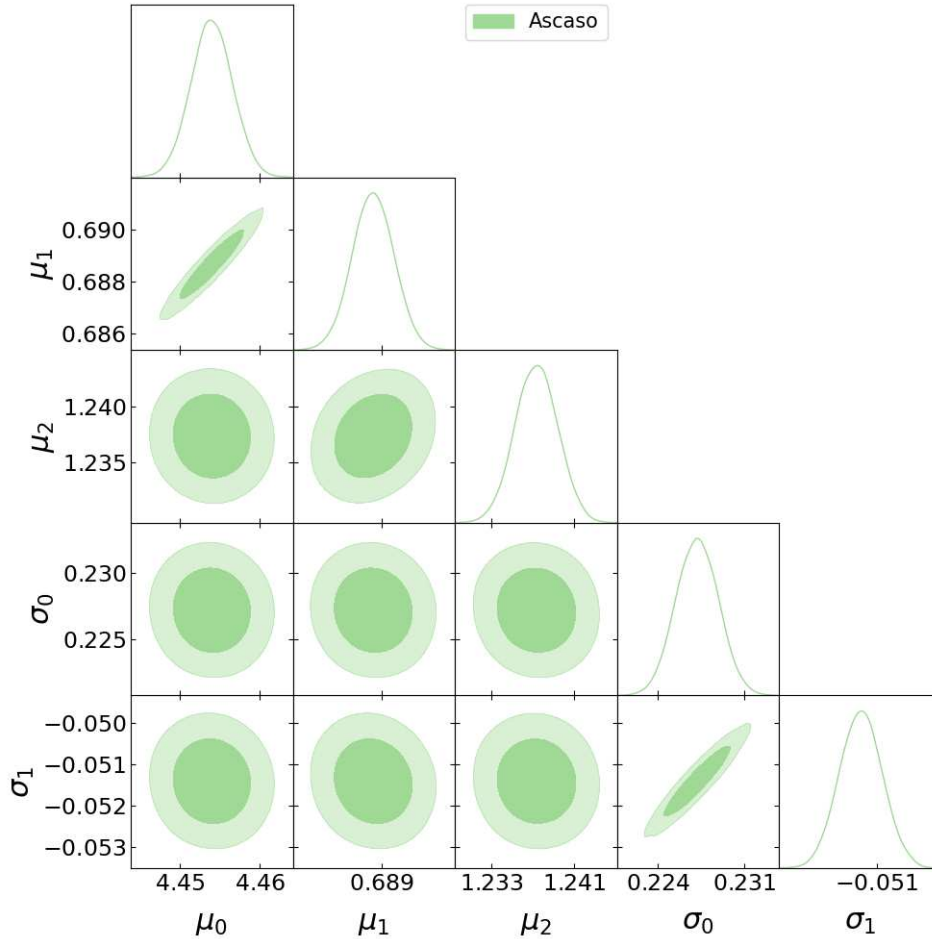


Figura 2 – Resultado para a análise por MCMC dos parâmetros do modelo de parametrização de Ascaso, com  $3 \times 10^5$  pontos. Os contornos correspondem as regiões de confiança dos parâmetros de 68% (verde mais escuro) e 95% (verde mais claro).

Na determinação de um modelo que descreva os dados, é importante analisar as diferenças entre os valores previstos pelo modelo e os dados reais. Assim, a análise dos resíduos pode conter informações a cerca da qualidade do ajuste e indicar a possível existência de um viés. Neste trabalho calculamos um vetor de resíduos **res** para cada valor de riqueza nos dados e seu correspondente valor de massa e *redshift*, fixando os valores dos parâmetros conforme o melhor ajuste da tabela 1 e ponderando pelo seu respectivo

desvio padrão:

$$\mathbf{res} = \left\{ \frac{\ln \lambda_i - \langle \ln \lambda_i | M_i, z_i \rangle}{\sigma_{\ln \lambda_i}} \right\}, i = 1, 2, 3, \dots \quad (4.7)$$

O resultado do cálculo dos resíduos para o conjunto completo dos dados e o modelo de Ascaso é mostrado nos gráficos da Figura 3. Pode-se observar que os resíduos apresentam uma grande heteroscedasticidade, diminuindo a variância, quando traçados em relação à  $\ln M$  e o mesmo, mas de forma mais leve, quando traçados em relação ao *redshift* (para altos valores de  $z$ ).

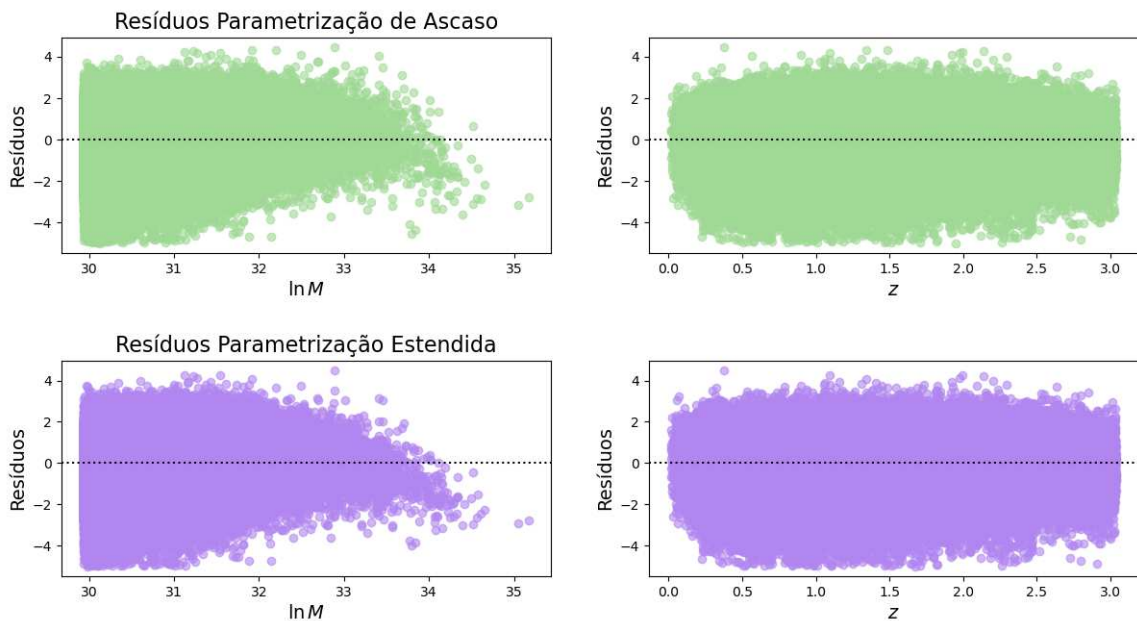


Figura 3 – Resíduos: Os dois gráficos de dispersão superiores representam os resíduos calculados levando em consideração o modelo de parametrização de Ascaso, os inferiores o da parametrização estendida. Os da esquerda representam resíduos  $\times \ln M$  e os da direita resíduos  $\times z$ , a linha preta pontilhada marca resíduo  $= 0.0$ .

Quanto a normalidade dos resíduos, ao fazer os histogramas (Figura 4), pode-se notar visualmente que a distribuição dos resíduos apresenta uma assimetria não desprezível, indicando que não é uma distribuição normal.

Para testar a normalidade dos resíduos de modo formal, utilizou-se o teste de Kolmogorov-Smirnov (KS) (KOLMOGOROV, 1933; SMIRNOV, 1939), que compara duas distribuições através da distância entre suas respectivas funções cumulativas (ou função de distribuição acumulada). Esse teste foi escolhido por ser eficaz em amostras com grandes números de dados. Assim, foi aplicado o teste KS para  $\mathbf{res}$  e para as amostras de uma distribuição normal com média e desvio padrão igual ao de  $\mathbf{res}$ . Para que a distribuição de  $\mathbf{res}$  seja igual a da normal, com uma confiança de 95%, a estatística p-valor deve ser

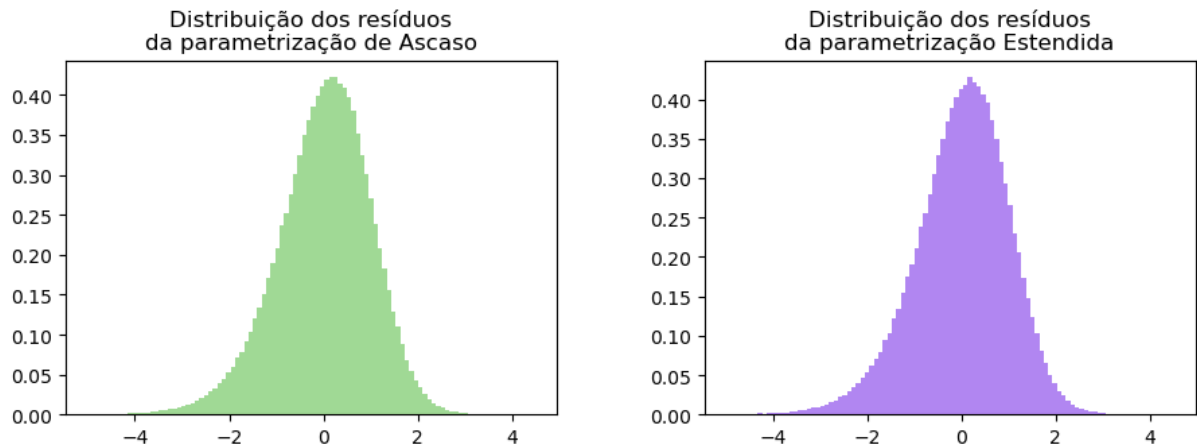


Figura 4 – Histograma dos Resíduos: pode-se notar uma assimetria na distribuição dos resíduos tanto para o modelo de Ascaso (verde), quanto o estendido (roxo).

<b>Tabela 2 - Estendido</b>			
	<b>Inicial</b>	<b>Melhor ajuste MLE</b>	<b>MCMC</b>
$\mu_0$	3.190	4.604	$4.606 \pm 0.014$
$\mu_{M1}$	0.868	0.872	$0.873 \pm 0.010$
$\mu_{M2}$	-0.304	0.042	$0.042 \pm 0.002$
$\mu_{Z1}$	0.000	1.774	$1.773 \pm 0.015$
$\mu_{Z2}$	0.000	-0.141	$-0.141 \pm 0.010$
$\mu_{MZ}$	0.000	0.142	$0.142 \pm 0.005$
$\sigma_0$	0.330	0.288	$0.291 \pm 0.011$
$\sigma_{M1}$	-0.080	0.017	$0.019 \pm 0.008$
$\sigma_{M2}$	0.000	0.015	$0.016 \pm 0.002$
$\sigma_{Z1}$	0.000	-0.017	$-0.015 \pm 0.013$
$\sigma_{Z2}$	0.000	0.008	$0.009 \pm 0.007$
$\sigma_{MZ}$	0.000	0.037	$0.038 \pm 0.004$

maior que 0.05, caso contrário essa hipótese é rejeitada. Os p-valor encontrados para o modelo de Ascaso foi 0.00, isso indica que a distribuição de **res** não é normal.

Os resultados desta análise motivaram a uma segunda análise com um modelo estendido, que considera termos quadráticos dependentes de  $\ln(M)$  e  $\ln(1+z)$ , para entender se o formato dos resíduos se deve a uma característica dos dados ou é dependente do modelo de relação riqueza-massa.

### 4.3 Análise II: Modelo estendido

Nesta análise o mesmo procedimento feito para a inferência dos parâmetros  $\mathbf{p}_1$  foi realizada para os parâmetros do modelo estendido,  $\mathbf{p}_2$ . Os valores fiduciais para o cálculo dos MLE e os resultados do ajuste podem ser vistos na tabela 2.

Para encontrar os limites nos parâmetros modelo estendido, foram calculadas 2200 cadeias no NcmFitESMCMC. Os resultados podem ser vistos na Tabela 2 e as regiões de confiança de 68% e 95% para os parâmetros inferidos podem ser vistas nos gráficos das Figura 5 .

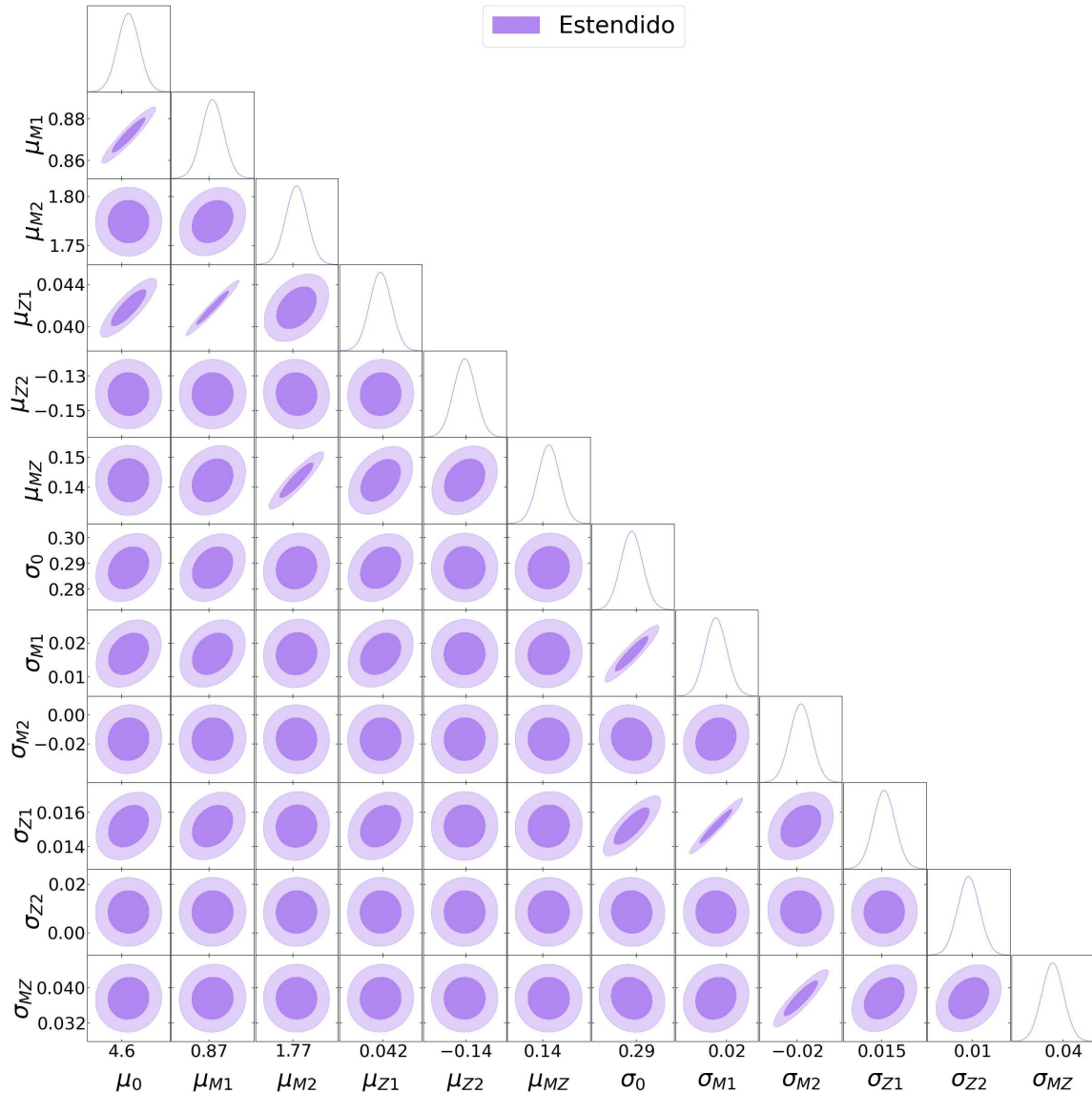


Figura 5 – Resultado para a análise por MCMC dos parâmetros do modelo de parametrização estendida, com  $2,288 \times 10^6$  pontos. Os contornos correspondem as regiões de confiança dos parâmetros de 68% (roxo mais escuro) e 95% (roxo mais claro).

Quanto à análise dos resíduos, pode-se perceber, nas figuras 3 e 4 que os resultados são bem parecidos com o do modelo de Ascaso, apresentando heterocedasticidade na dispersão dos resíduos em relação a  $\ln M$  e  $z$  e mantendo uma distribuição assimétrica de formato semelhante.

A proximidade entre o formato dos resíduos das relações de escala de Ascaso e sua versão estendida ainda não determina que esses resultados provenham de uma

característica natural dos dados. Isso motiva a uma reconstrução da média  $\langle \ln \lambda | M, z \rangle$  e variância  $\sigma_{\ln \lambda}$  usando pequenos intervalos de  $\ln M$  e  $z$  e a comparação desta reconstrução com as parametrizações já discutidas, isso será apresentado na próxima seção.

#### 4.4 Análise III: Análises em intervalos

Para fazer uma análise de validação mais profunda dos modelos analisados até aqui, o procedimento ideal seria uma comparação com valores reais dessas quantidades através de uma validação cruzada, como isso não é possível, propôs-se reconstruir o comportamento da média e desvio padrão através da divisão dados em pequenos intervalos. A cada pequeno intervalo de massa e *redshift* são calculados a média e desvio padrão de  $\ln \lambda$  e, ao fim, é obtido um conjunto de médias e desvios padrões que podem ser vinculadas a um valor aproximado de  $\ln M$  e  $z$ .

Para fazer a reconstrução, primeiramente os dados do catálogo provindo do cosmoDC2 (KORYTOV et al., 2019), que tem um *range* de massas de halos variando entre  $1.00 \times 10^{13} M_{\odot}$  e  $1.88 \times 10^{15} M_{\odot}$  e, *redshifts* de 0 a 3, foram divididos em intervalos de  $z$  com largura 0.05 e  $\ln M$  com largura 0.25, de forma que os 454475 halos do catálogo simulado foram separados em 1281 intervalos. Considera-se  $\ln \lambda_i$ ,  $\ln M_i$  e  $z_i$ , os  $i$ -ésimos conjuntos de valores de  $\ln \lambda$ ,  $\ln M$  e  $z$ , respectivamente, que estão dentro de cada intervalo. Para cada  $\ln \lambda_i$  foi feito um histograma, o resultado dos primeiros 200 e últimos 200 histogramas estão contidos na Figura 6, para ilustrar o formato das distribuições dos  $\ln \lambda_i$  (caso queira uma visualização maior, veja o Anexo A).

Observa-se pelos histogramas que a distribuição de  $\ln \lambda_i$  aparenta ter uma forma gaussiana e para averiguar normalidade foram realizados os testes KS e de Shapiro-Wilk (SHAPIRO; WILK, 1965), SW, que testa se uma amostra vem de uma distribuição normal e é recomendado para amostras com menos de 50 elementos (PIGATTO, 2005). Para intervalos com menos de 5 elementos não foram aplicados os testes pois, poucos ou nenhum elemento, não fornecem um resultado confiável. Para intervalos de 5 a 50 elementos foi aplicado o teste SW e para maiores de 50 elementos foi aplicado o teste KS. Os resultados foram que 72.55% dos  $\ln \lambda_i$  analisados pelo teste KS (206 intervalos) e 85.92 % dos  $\ln \lambda_i$  analisados pelo teste SW (499 intervalos) foram considerados normais com 95% de confiança.

A partir dos resultados dos testes KS e SW, pode-se supor uma normalidade na distribuição de  $\ln \lambda_i$  e assim considerar seu valor esperado como a média aritmética de  $\ln \lambda_i$ ,  $\mu_{\ln \lambda}$ . O conjunto de médias traçado em relação a valores de  $\ln M$  e  $z$  é uma reconstrução de  $\langle \ln \lambda | M, z \rangle$  e pode ser visto na Figura 7.

Nos gráficos da Figura 7 observa-se que  $\mu_{\ln \lambda}$  aparenta ter uma relação linear com  $\ln M$ , já quanto a  $z$ , a forma das curvas aparentam ter uma relação quadrática. Pode-se

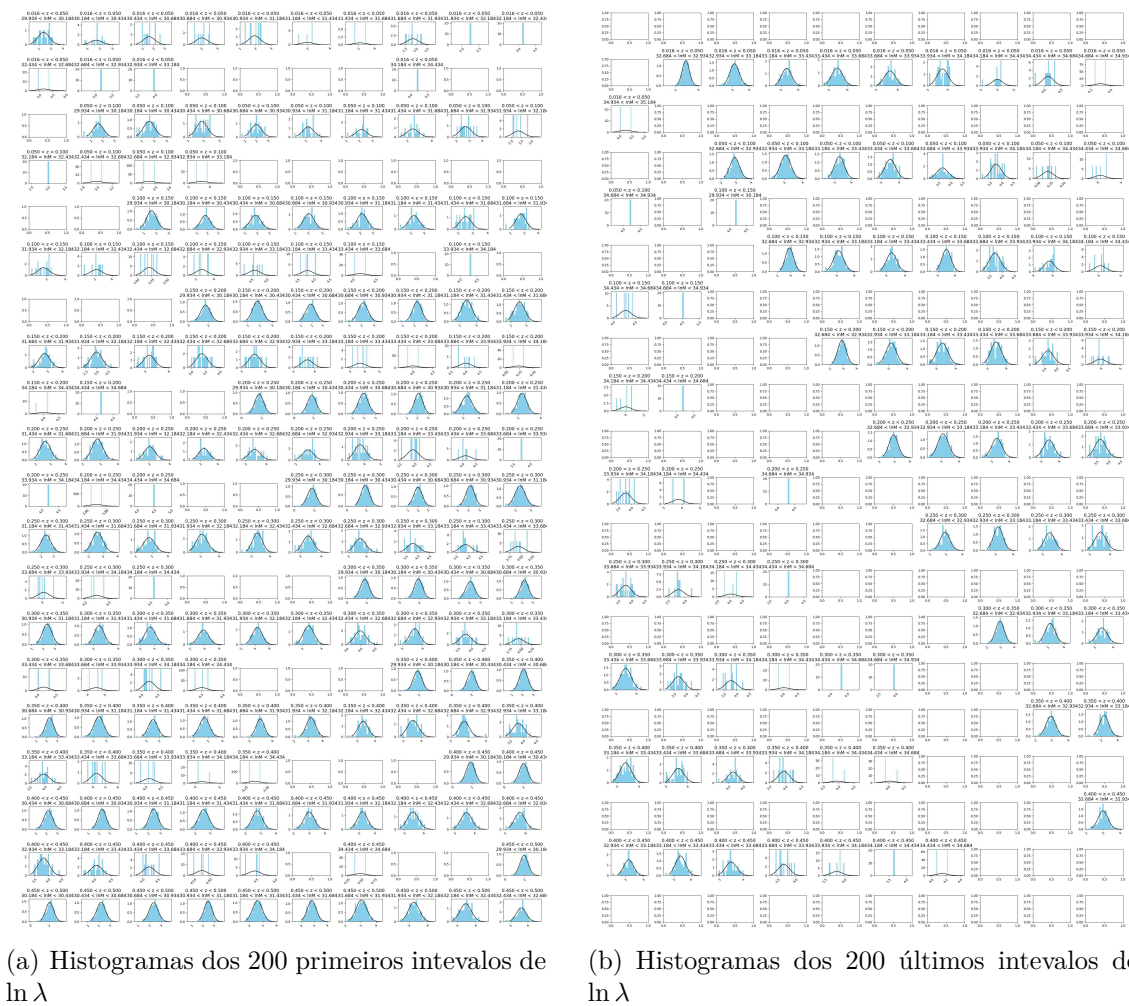


Figura 6 – Histogramas dos 200 primeiros (baixas massas e *redshifts*) e 200 últimos (altas massas e *redshifts*) intervalos de  $\ln \lambda$ : Em azul os histogramas dos conjuntos  $\ln \lambda_i$ . As linhas pretas são o traço de uma gaussiana de média e desvio padrão igual ao do respectivo  $\ln \lambda_i$ .

verificar a relação entre  $\mu_{\ln \lambda}$  e  $\ln z$  na Figura 8, nota-se que essa relação não é linear.

Usando a reconstrução da média  $\mu_{\ln \lambda}$  pode-se então validar os resultados dos modelos de relações  $\langle \ln \lambda | M, z \rangle$  e desvios padrões  $\sigma_{\ln \lambda}$ . Traçando a evolução de  $\mu_{\ln \lambda}$  e  $\langle \ln \lambda | M, z \rangle$  em relação a  $\ln M$  e  $z$ , de forma sobreposta, é possível comparar de forma visual o comportamento dessas duas quantidades. Isso é feito nos gráficos das figuras 9 e 10.

Percebe-se, pela Figura 10, que visualmente o modelo estendido da média aparenta se encaixar melhor na sobreposição com  $\mu_{\ln \lambda}$ . Para testar a qualidade de ajuste de ambos os modelos em comparação com a  $\mu_{\ln \lambda}$  foi realizada uma validação cruzada em que 30% dos dados foram separados em um conjunto de teste e 70% em um conjunto de treinamento para os modelos. Isso realizado com o algoritmo *train\_test\_split*<sup>2</sup> da biblioteca numérica

<sup>2</sup> [https://scikit-learn.org/stable/modules/generated/sklearn.model\\_selection.train\\_test\\_split.html](https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.train_test_split.html)

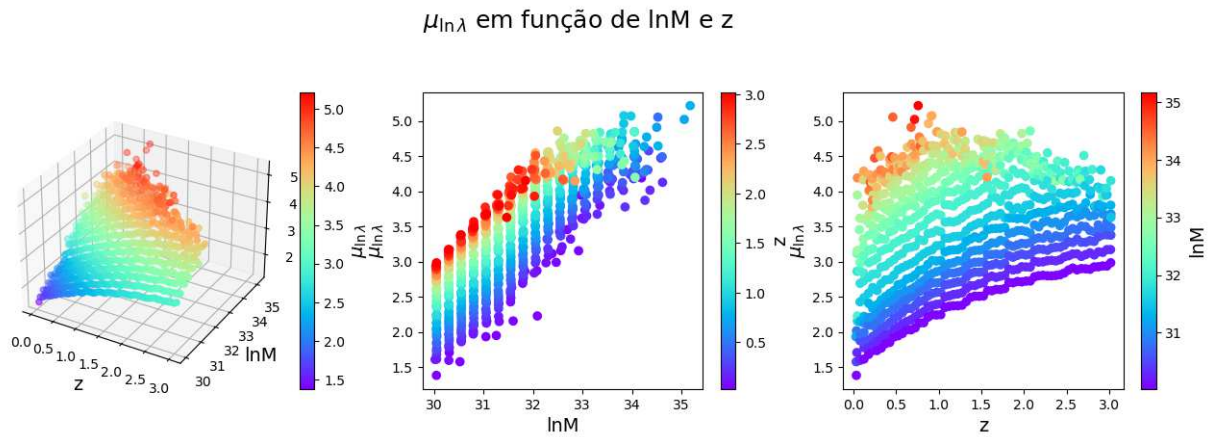


Figura 7 –  $\mu_{\ln \lambda}$  em função de  $\ln M$  e  $z$ : Os gráficos mostram as médias dos conjuntos  $\ln \lambda_i$  traçadas em relação a  $\ln M$  e  $z$ . Para fazer o gráfico de dispersão foram usadas as médias de  $\ln M$  e  $z$  em cada intervalo e relacionadas à  $\mu_{\ln \lambda}$  do respectivo intervalo. No gráfico mais à esquerda há  $\mu_{\ln \lambda}$  traçado em relação a  $\ln M$  e  $z$ , no gráfico ao centro há uma projeção  $\mu_{\ln \lambda}$  do primeiro gráfico em relação a superfície  $\ln M \times \mu_{\ln \lambda}$ , e no gráfico mais a direita uma projeção na superfície  $z \times \mu_{\ln \lambda}$ .

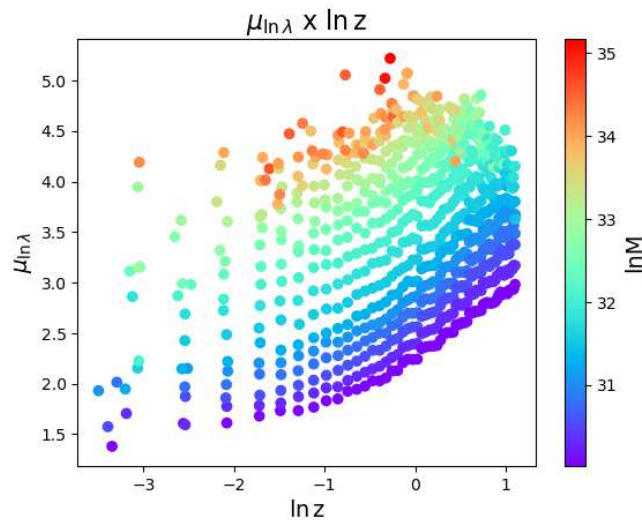


Figura 8 –  $\mu_{\ln \lambda}$  em função de  $\ln z$ .

scikit-learn, que divide os dados em subconjuntos aleatórios de treinamento e teste. Com o conjunto de treinamento e usando o método de MLE, foram ajustados os parâmetros dos modelos. Depois, com o conjunto de teste, foi calculado  $\langle \ln \lambda | M, z \rangle$  e comparado com  $\mu_{\ln \lambda}$  utilizando estatísticas como o erro relativo, o erro quadrático médio (*Mean Squared Error, MSE*) e sua raiz (*Root Mean Squared Error, RMSE*), e o erro médio absoluto percentual (*Mean Absolute Percentage Error, MAPE*).

## Modelo de Ascaso

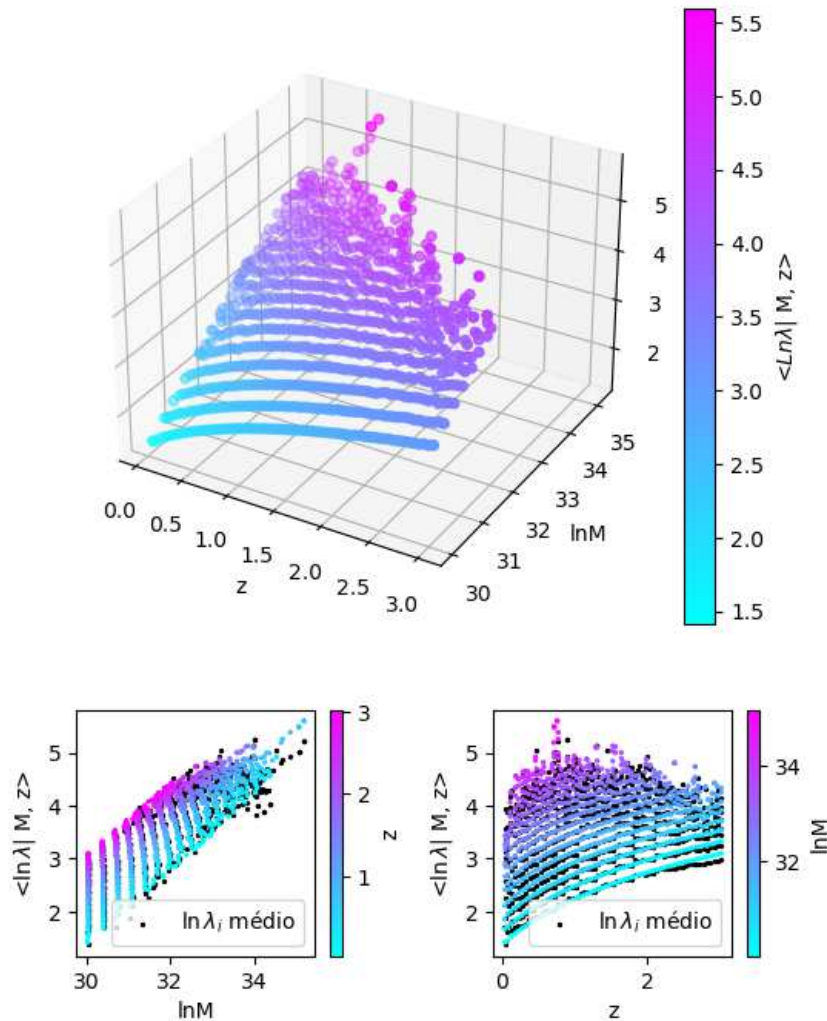


Figura 9 – Modelo de Ascaso para  $\langle \ln \lambda | M, z \rangle$  em comparação com a reconstrução através de intervalos: em escala de cores o modelo de Ascaso, em preto a reconstrução usando as médias  $\mu_{\ln \lambda}$ .

O erro relativo,  $ER$ , pode ser calculado através da equação

$$ER_{\langle \ln \lambda_i | M_i, z_i \rangle} = \left\{ \frac{\langle \ln \lambda_i | M_i, z_i \rangle - \mu_{\ln \lambda}^i}{\mu_{\ln \lambda}^i} \right\}, i = 1, 2, 3, \dots \quad (4.8)$$

Ele indica o quanto um ponto do modelo,  $\langle \ln \lambda_i | M_i, z_i \rangle$  está afastado do respectivo ponto  $\mu_{\ln \lambda}^i$ . Traçando um gráfico de dispersão de  $ER$  vs. o conjunto de predição, Figura 11, tem-se que o ideal é que os pontos estejam dispersos de forma aleatória, de forma que os erros não tenham algum tipo de correlação (homocedasticidade), mas no gráfico observa-se que para valores mais altos do conjunto de predição a variância dos erros relativos é maior.

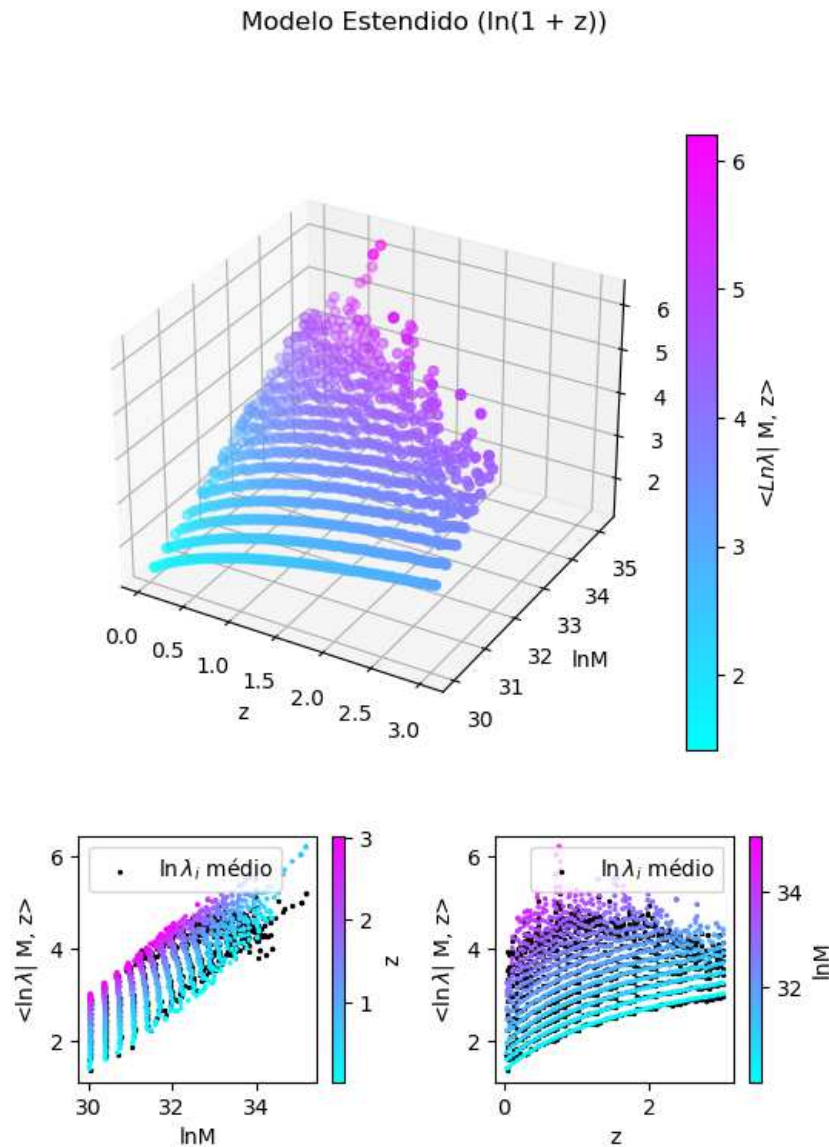


Figura 10 – Modelo estendido para  $\langle \ln \lambda | M, z \rangle$  em comparação com a reconstrução através de intervalos: em escala de cores o modelo estendido, em preto a reconstrução usando as médias  $\mu_{\ln \lambda}$ .

Na figura 12, pode-se observar os histogramas dos erros relativos e, aplicando um teste KS, o resultado é que em ambos os modelos os resíduos não apresentam normalidade, o p-valor para Ascaso é  $1.013 \times 10^{-10}$  e para o modelo estendido é  $4.111 \times 10^{-35}$ .

O erro quadrático médio é uma estatística que dá uma noção geral de quanto os pontos do modelo diferem do que se quer estimar. O cálculo para o MSE dos modelos de  $\langle \ln \lambda_i | M_i, z_i \rangle$  são feitos por

$$MSE_{\langle \ln \lambda_i | M_i, z_i \rangle} = \frac{1}{N} \sum (\mu_{\ln \lambda}^i - \langle \ln \lambda_i | M_i, z_i \rangle)^2. \quad (4.9)$$

O resultado usando o conjunto de teste é que para o modelo de Ascaso o MSE foi de 0.018 e para o modelo estendido foi de 0.040.

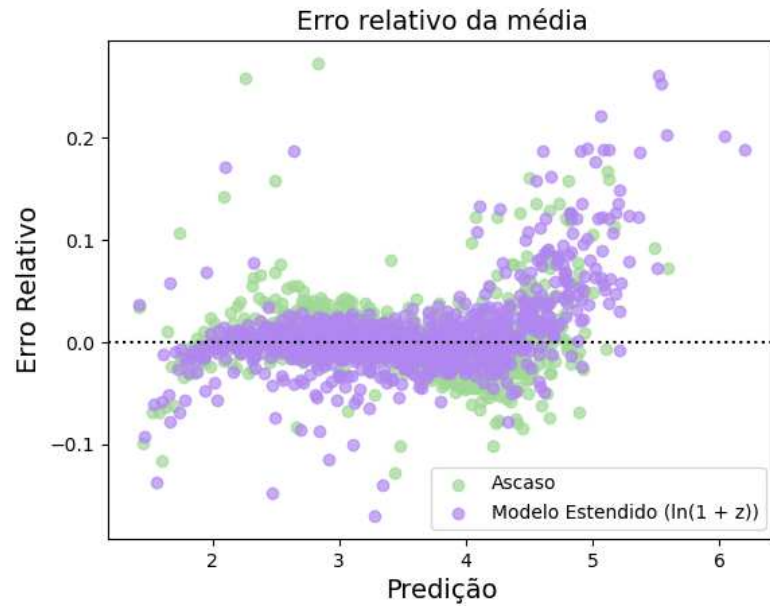


Figura 11 – Erros relativos dos modelos de  $\langle \ln \lambda_i | M_i, z_i \rangle$  calculados usando o conjunto de teste.

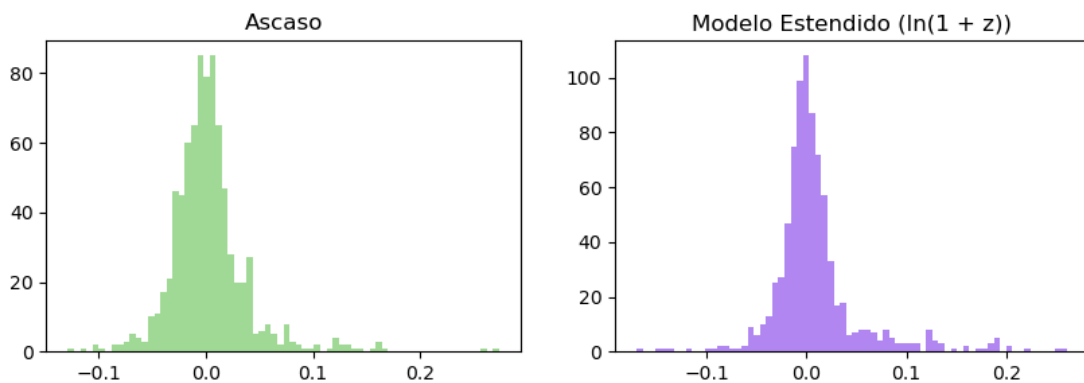


Figura 12 – Histogramas dos erros relativos dos modelos de  $\langle \ln \lambda_i | M_i, z_i \rangle$  calculados usando o conjunto de teste.

Quanto ao MAPE, essa é uma estatística de erro relativo médio que utiliza valores absolutos para impedir que os erros positivos e negativos se anulem. É bastante utilizada em áreas como economia e logística por ser uma meio intuitivo de inferir a qualidade de ajuste de um modelo de regressão e comparação com outros modelos de previsão (MYTTENAERE et al., 2016). Para o caso do ajuste de  $\langle \ln \lambda_i | M_i, z_i \rangle$ , o MAPE pode ser calculado como

$$MAPE_{\langle \ln \lambda_i | M_i, z_i \rangle} = \frac{1}{N} \sum_i^N \frac{|\langle \ln \lambda_i | M_i, z_i \rangle - \mu^i_{\ln \lambda}|}{|\mu^i_{\ln \lambda}|}. \quad (4.10)$$

Para Ascaso o erro de previsão da estimativa da média, segundo o  $MAPE_{\langle \ln \lambda_i | M_i, z_i \rangle}$  é de 2.19%, já para o modelo estendido é de 2.57%.

Além de  $\langle \ln \lambda_i | M_i, z_i \rangle$ , o modelo do desvio padrão  $\sigma_{\ln \lambda}$  também foi avaliado. Usando os intervalos divididos anteriormente, foi calculado o  $\sigma^i$ , para os dados dentro de cada intervalo  $i$ . Nas Figuras 13 e 14 pode ser observado o gráfico de dispersão que sobrepõe os valores de  $\sigma_{\ln \lambda}$ , ao longo de  $\ln M$  e  $z$ , com os valores de  $\sigma^i$ .

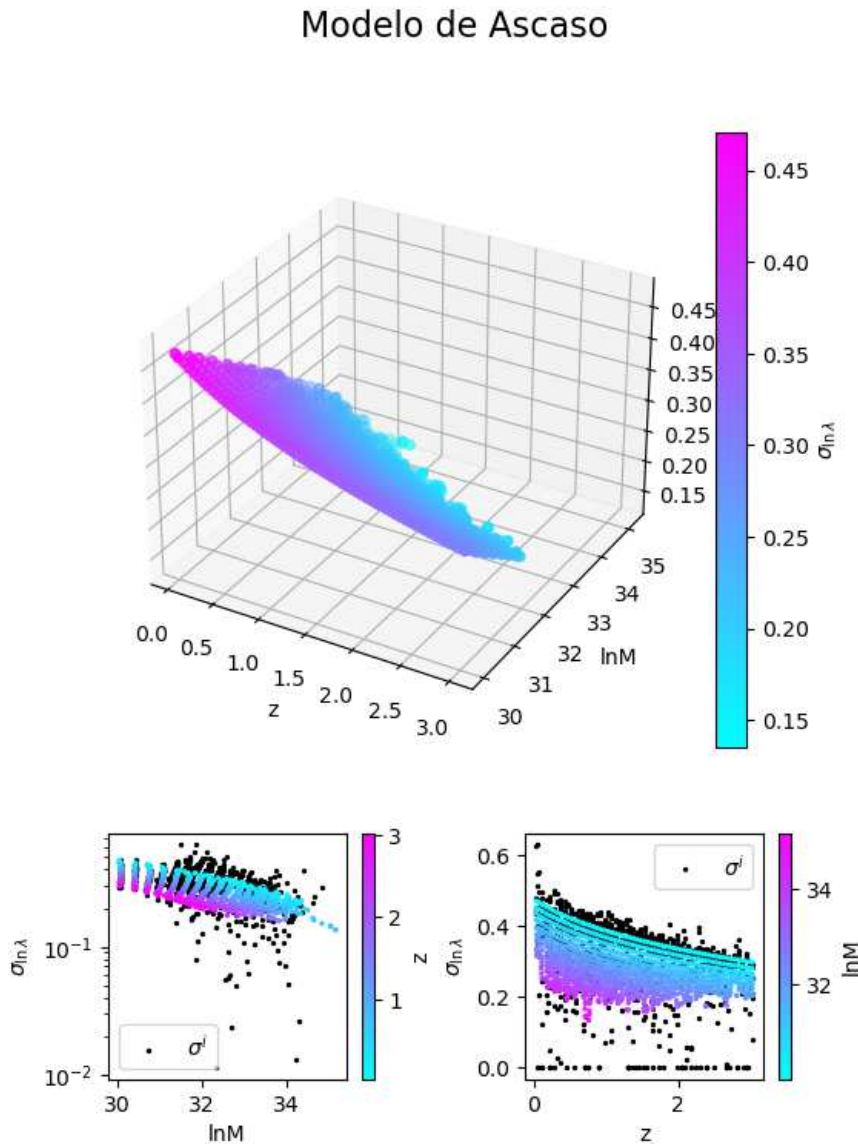


Figura 13 – Modelo de Ascaso para  $\sigma_{\ln \lambda}$  em comparação com a reconstrução do desvio padrão usando intervalos: em escala de cores o modelo de Ascaso, em preto a reconstrução usando os desvios padrões  $\sigma^i$ .

Pode se notar nas figuras que ambos os modelos  $\sigma_{\ln \lambda}$  não se aproximam bem do formato da reconstrução usando os  $\sigma^i$ , tendo uma diferença notável quando traçado em relação a  $\ln M$ . Fazendo a validação cruzada, usando os mesmos conjuntos de treinamento e teste anterior, foram calculadas as estatísticas de erros. O resultado para os erros relativos pode ser visualizado na Figura 15(a), na Figura 15(b) tem-se os mesmos resultados mas excluindo possíveis *outliers* ( $ER > 10$ ). Quanto ao  $MSE_{\sigma_{\ln \lambda}}$ , para Ascaso o resultado foi

### Modelo Estendido ( $\ln(1 + z)$ )

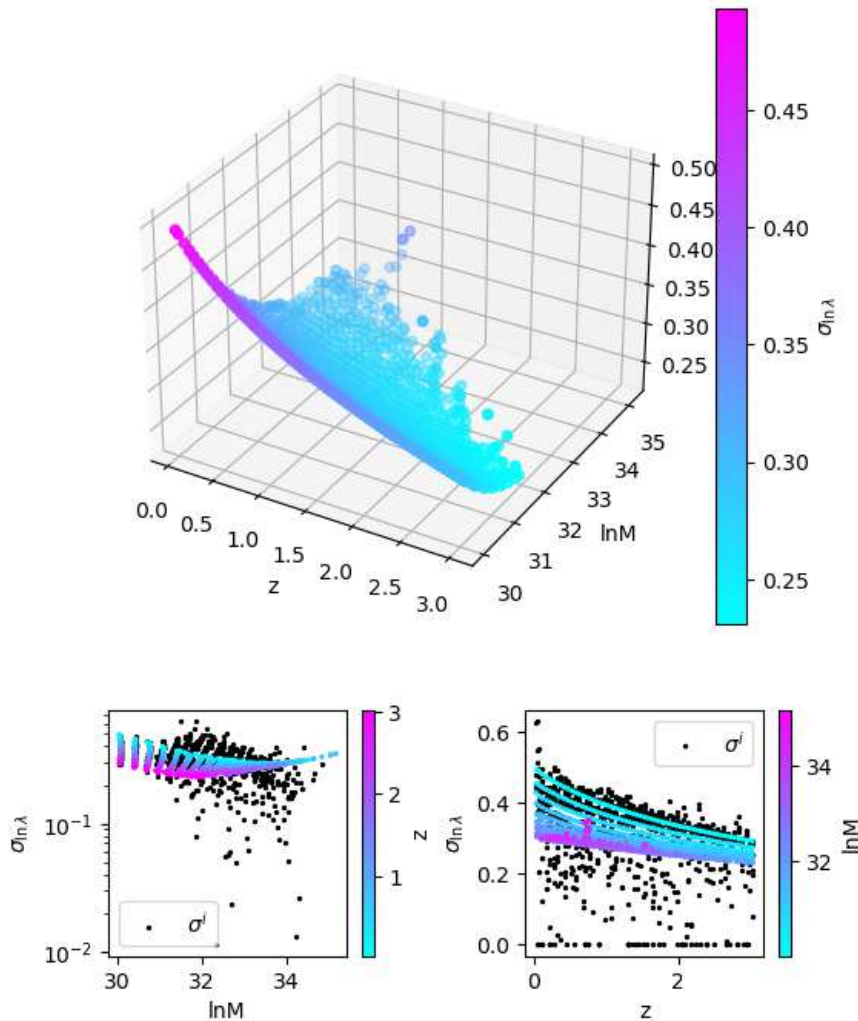
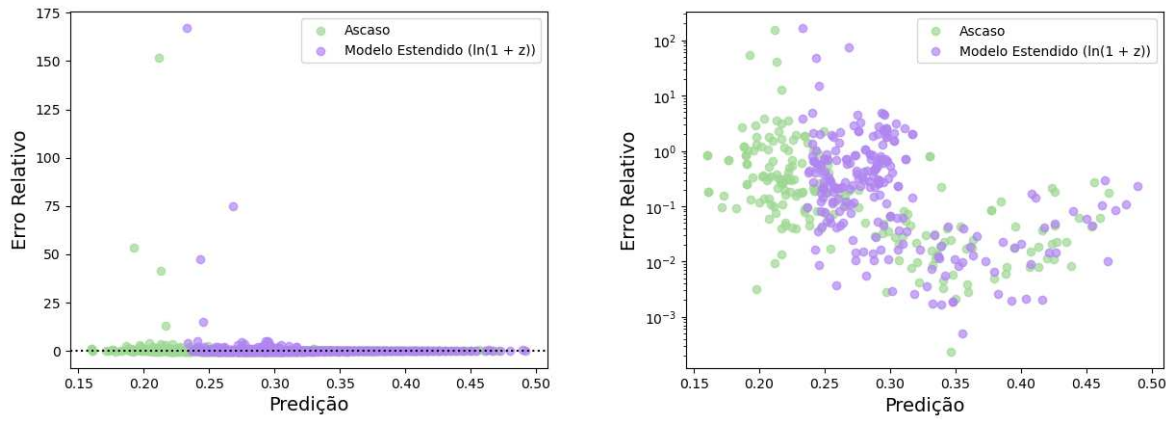


Figura 14 – Modelo estendido para  $\sigma_{\ln \lambda}$  em comparação com a reconstrução do desvio padrão usando intervalos: em escala de cores o modelo estendido, em preto a reconstrução usando os desvios padrões  $\sigma^i$ .

de 0.008 e para o modelo estendido 0.012. Para o  $MAPE_{\sigma_{\ln \lambda}}$  os resultados foram 58.35% e 70.83% respectivamente. Esses valores mostram quantitativamente o quão distante estão os valores calculados pelos modelos dos valores reconstruídos através dos  $\sigma^i$ .

Os resultados da análise do desvio padrão mostram que há uma grande discrepância entre os resultados dos modelos e o da reconstrução por intervalos. Além disso, esses resultados podem estar relacionados aos dos resíduos das análises I e II que apresentaram heteroscedasticidade, o decrescimento na variância de res ao longo  $\ln M$  e  $z$ . Isso indica que esses modelos de parametrização de  $\sigma_{\ln \lambda}$  podem não ser adequados.



(a) Erros relativos do desvio padrão considerando todos os dados do conjunto de teste.

(b) Erros relativos do desvio padrão em escala logaritmica.

Figura 15 – Erros relativos dos modelos de  $\sigma_{\ln \lambda}$  calculados usando o conjunto de teste.

## 5 CONCLUSÃO

A Cosmologia atual está fortemente interligada às análises estatísticas de dados observacionais. Por meio de distribuições massa-observáveis pode-se inferir parâmetros cosmológicos e testar os modelos teóricos. Devido a isso, as próprias distribuições devem ser testadas, a fim de avaliar o quão bem elas descrevem os dados.

Neste trabalho foram estudados dois modelos de parametrização para a distribuição Riqueza-Massa, a parametrização de Ascaso e a estendida. Inferiram-se seus parâmetros via MLE, e seus resultados foram validados usando uma comparação entre os resultados dos modelos de parametrização e das reconstruções da média e desvio padrão usando intervalos de dados de largura 0.05 para  $z$  e largura 0.25 para  $\ln M$ .

Para as calibrações de  $\langle \ln \lambda_i | M_i, z_i \rangle$ , concluí-se que os modelos de Ascaso e estendido, equações 4.2 e 4.4, possuem resultados parecidos. Quando comparados com a reconstrução da média por intervalos, visualmente (Figuras 9 e 10) tem comportamentos próximos ao de  $\mu_{\ln \lambda}$  ao longo de  $\ln M$  e  $z$ . Esse resultado é validado quantitativamente através das estatísticas de erros, para o modelo de Ascaso o RMSE é de 0.135 e o MAPE é 2.19%, já para o modelo estendido esses valores são 0.200 e 2.57% respectivamente. Esses valores indicam uma diferença relativamente baixa entre os modelos e a reconstrução.

A respeito da calibração dos modelos de desvios padrões  $\sigma_{\ln \lambda}$  nota-se facilmente pelos gráficos das figuras 13 e 14 que há uma grande discrepância entre os modelos e a reconstrução através dos dados. Esses resultados ficam ainda mais notáveis ao calcular o MAPE, para o modelo de Ascaso 58.35% e para o estendido 70.83%, indicando uma grande diferença dos comportamentos dos modelos em relação a reconstrução. Esse comportamento indica que esse modelo de parametrização não é adequado, o que pode explicar a heteroscedasticidade que foi observada na dispersão dos resíduos nas análises I e II, Figura 3.

Esses resultados motivam uma futura investigação mais detalhada a cerca da discrepância de  $\sigma_{\ln \lambda}$  e de modelos mais precisos para as relações Riqueza-Massa. Além disso, o estudo a respeito da reconstrução e calibração dessas relações é apenas um passo inicial para futuros trabalhos a respeito de reconstruções não-paramétricas de distribuições massa-observáveis.

## REFERÊNCIAS

- ABBOTT, T. et al. Joint analysis of Dark Energy Survey year 3 data and CMB lensing from SPT and Planck. III. Combined cosmological constraints. *Physical Review D*, American Physical Society (APS), v. 107, n. 2, jan 2023. Disponível em: <<https://doi.org/10.1103/PhysRevD.107.023531>>.
- ABDULLAH, M. H. et al. Constraining cosmological parameters using the cluster mass–richness relation. *The Astrophysical Journal*, The American Astronomical Society, v. 955, n. 1, p. 26, sep 2023. Disponível em: <<https://dx.doi.org/10.3847/1538-4357/ace773>>.
- ANDREON, M. A. H. S. The scaling relation between richness and mass of galaxy clusters: a Bayesian approach. *Monthly Notices of the Royal Astronomical Society*, v. 404, n. 4, p. 1922–1937, 05 2010. ISSN 0035-8711. Disponível em: <<https://doi.org/10.1111/j.1365-2966.2010.16406.x>>.
- ASCASO, B. et al. An accurate cluster selection function for the J-PAS narrow-band wide-field survey. *Monthly Notices of the Royal Astronomical Society*, v. 456, n. 4, p. 4291–4304, mar. 2016. Disponível em: <<https://ui.adsabs.harvard.edu/abs/2016MNRAS.456.4291A>>.
- CARROLL, S. M. *Spacetime and geometry. An introduction to general relativity*. Pearson new international edition. [S.l.]: Pearson, 2014.
- CARROLL, S. M.; PRESS, W. H.; TURNER, E. L. The cosmological constant. *Annual Rev. Astron. Astrophys.*, v. 30, p. 499–542, jan. 1992. Disponível em: <<https://ui.adsabs.harvard.edu/abs/1992ARA&A..30..499C>>.
- CHIU, I.-N. et al. The richness-to-mass relation of CAMIRA galaxy clusters from weak-lensing magnification in the Subaru Hyper Suprime-Cam survey. *Monthly Notices of the Royal Astronomical Society*, v. 495, n. 1, p. 428–450, 04 2020. Disponível em: <<https://doi.org/10.1093/mnras/staa1158>>.
- CROFT, R. A. C. et al. The richness dependence of galaxy cluster correlations: results from a redshift survey of rich APM clusters. *Monthly Notices of the Royal Astronomical Society*, v. 291, n. 2, p. 305–313, 10 1997. Disponível em: <<https://doi.org/10.1093/mnras/291.2.305>>.
- Dark Energy Survey Collaboration et al. The Dark Energy Survey: more than dark energy - an overview. *Monthly Notices of the Royal Astronomical Society*, v. 460, n. 2, p. 1270–1299, ago. 2016. Disponível em: <<https://ui.adsabs.harvard.edu/abs/2016MNRAS.460.1270D>>.
- DAVIS, M. et al. The evolution of large-scale structure in a universe dominated by cold dark matter. *The Astrophysical Journal*, v. 292, p. 371–394, maio 1985. Disponível em: <<https://ui.adsabs.harvard.edu/abs/1985ApJ...292..371D>>.
- DODELSON, S. *Modern Cosmology*. [S.l.]: Academic Press, 2003.

- EINSTEIN, A. Kosmologische Betrachtungen zur allgemeinen Relativitätstheorie. *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften*, p. 142–152, jan. 1917. Disponível em: <<https://ui.adsabs.harvard.edu/abs/1917SPAW.....142E>>.
- Euclid Collaboration et al. Euclid preparation. I. The Euclid Wide Survey. *Astronomy & Astrophysics*, v. 662, p. A112, jun. 2022. Disponível em: <<https://ui.adsabs.harvard.edu/abs/2022A&A...662A.112E>>.
- GELMAN, A. et al. *Bayesian Data Analysis*. 2. ed. [S.l.]: Chapman & Hall/CRC, 2004.
- GOODMAN, J.; WEARE, J. Ensemble samplers with affine invariance. *Communications in Applied Mathematics and Computational Science*, v. 5, n. 1, p. 65–80, jan. 2010.
- HASTINGS, W. K. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, [Oxford University Press, Biometrika Trust], v. 57, n. 1, p. 97–109, 1970. ISSN 00063444. Disponível em: <<http://www.jstor.org/stable/2334940>>.
- HOEKSTRA, H. et al. Masses of galaxy clusters from gravitational lensing. *Space Science Reviews*, v. 177, p. 75–118, 2013. Disponível em: <<https://doi.org/10.1007/s11214-013-9978-5>>.
- HUBBLE, E. P. A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae. *Proceedings of the National Academy of Science*, v. 15, n. 3, p. 168–173, mar. 1929. Disponível em: <<https://www.pnas.org/doi/10.1073/pnas.15.3.168>>.
- HUBBLE, E. P. A spiral nebula as a stellar system, Messier 31. *Astrophysical Journal*, v. 69, p. 103–158, mar. 1929. Disponível em: <<https://ui.adsabs.harvard.edu/abs/1929ApJ...69..103H>>.
- HUCHRA, J. P.; GELLER, M. J. Groups of Galaxies. I. Nearby groups. *The Astrophysical Journal*, v. 257, p. 423–437, jun. 1982.
- IKEBE, Y. et al. A new measurement of the x-ray temperature function of clusters of galaxies. *A&A*, v. 383, n. 3, p. 773–790, 2002. Disponível em: <<https://doi.org/10.1051/0004-6361:20011769>>.
- JENKINS, A. et al. The mass function of dark matter haloes. *Monthly Notices of the Royal Astronomical Society*, v. 321, n. 2, p. 372–384, fev. 2001.
- Kollmeier, J. et al. SDSS-V Pioneering Panoptic Spectroscopy. v. 51, p. 274, set. 2019.
- KOLMOGOROV, A. Sulla Determinazione Empirica di una Legge di Distribuzione. *Giornale dell'Istituto Italiano degli Attuari*, v. 4, p. 83–91, 1933. Disponível em: <<http://digitale.bnc.roma.sbn.it/tecadigitale/giornale/CFI0353791/1933/unico>>.
- KORYTOV, D. et al. Cosmodc2: A synthetic sky catalog for dark energy science with lsst. *The Astrophysical Journal Supplement Series*, The American Astronomical Society, v. 245, n. 2, p. 26, dec 2019. Disponível em: <<https://dx.doi.org/10.3847/1538-4365/ab510c>>.

- LACEY, C.; COLE, S. Merger Rates in Hierarchical Models of Galaxy Formation - Part Two - Comparison with N-Body Simulations. *Monthly Notices of the Royal Astronomical Society*, v. 271, p. 676, dez. 1994. Disponível em: <<https://ui.adsabs.harvard.edu/abs/1994MNRAS.271..676L>>.
- LIMA, M.; HU, W. Self-calibration of cluster dark energy studies: Observable-mass distribution. *Phys. Rev. D*, American Physical Society, v. 72, p. 043006, Aug 2005. Disponível em: <<https://link.aps.org/doi/10.1103/PhysRevD.72.043006>>.
- LUPTON, R. *Statistics in Theory and Practice*. [S.l.]: Princeton University Press - Princeton, 1993.
- MARKEVITCH, M. et al. Direct constraints on the dark matter self-interaction cross section from the merging galaxy cluster 1e 0657-56. *The Astrophysical Journal*, American Astronomical Society, v. 606, n. 2, p. 819–824, may 2004. Disponível em: <<https://doi.org/10.1086%2F383178>>.
- MORESCO, M.; MARULLI, F. Cosmological constraints from a joint analysis of cosmic growth and expansion. *Monthly Notices of the Royal Astronomical Society: Letters*, v. 471, n. 1, p. L82–L86, 07 2017. ISSN 1745-3925. Disponível em: <<https://doi.org/10.1093/mnrasl/slx112>>.
- MURATA, R. et al. Constraints on the mass–richness relation from the abundance and weak lensing of sdss clusters. *The Astrophysical Journal*, The American Astronomical Society, v. 854, n. 2, p. 120, feb 2018. Disponível em: <<https://dx.doi.org/10.3847/1538-4357/aaaab8>>.
- MYTTENAERE, A. de et al. Mean absolute percentage error for regression models. *Neurocomputing*, v. 192, p. 38–48, 2016. ISSN 0925-2312. Advances in artificial neural networks, machine learning and computational intelligence. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0925231216003325>>.
- NAGAI, D. The impact of galaxy formation on the sunyaev-zel'dovich effect of galaxy clusters. *The Astrophysical Journal*, v. 650, n. 2, p. 538, oct 2006. Disponível em: <<https://dx.doi.org/10.1086/506467>>.
- NISHIZAWA, A.; TARUYA, A.; SAITO, S. Tracing the redshift evolution of hubble parameter with gravitational-wave standard sirens. *Phys. Rev. D*, American Physical Society, v. 83, p. 084045, Apr 2011. Disponível em: <<https://link.aps.org/doi/10.1103/PhysRevD.83.084045>>.
- PENNA-LIMA, M. *Abundância de Aglomerados de Galáxias como Observável Cosmológico: Aplicações aos Levantamentos Fotométricos DES e SDSS*. Tese (Doutorado) — CBPF - Centro Brasileiro de Pesquisas Físicas, Rio de Janeiro, RJ, 2010. Disponível em: <<http://cbpfindex.cbpf.br/index.php?module=main&moduleFile=pubDetails&pubId=3534&typeId=9>>. Acesso em: 27 de junho de 2023.
- PIGATTO, L. Island Universes, Novae and Supernovae. A Great Debate of the XX Century. In: Turatto, M. et al. (Ed.). *1604-2004: Supernovae as Cosmological Lighthouses*. [s.n.], 2005. (Astronomical Society of the Pacific Conference Series, v. 342), p. 43. Disponível em: <<https://ui.adsabs.harvard.edu/abs/2005ASPC..342..43P>>.

- Planck Collaboration et al. Planck 2018 results - i. overview and the cosmological legacy of planck. *A&A*, v. 641, p. A1, 2020. Disponível em: <<https://doi.org/10.1051/0004-6361/201833880>>.
- Planck Collaboration et al. Planck 2018 results - vi. cosmological parameters. *A&A*, v. 641, p. A6, 2020. Disponível em: <<https://doi.org/10.1051/0004-6361/201833910>>.
- PRESS, W. H.; SCHECHTER, P. Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation. *The Astrophysical Journal*, v. 187, p. 425–438, fev. 1974. Disponível em: <<https://ui.adsabs.harvard.edu/abs/1974ApJ...187..425P>>.
- RIESS, A. G. et al. Cosmic distances calibrated to 1% precision with gaia EDR3 parallaxes and hubble space telescope photometry of 75 milky way cepheids confirm tension with  $\lambda$ CDM. *The Astrophysical Journal Letters*, The American Astronomical Society, v. 908, n. 1, p. L6, feb 2021. Disponível em: <<https://dx.doi.org/10.3847/2041-8213/abdbaf>>.
- RIESS, A. G. et al. Observational evidence from supernovae for an accelerating universe and a cosmological constant. *The Astronomical Journal*, American Astronomical Society, v. 116, n. 3, p. 1009–1038, sep 1998. Disponível em: <<https://doi.org/10.1086%2F300499>>.
- RUBIN, V. C.; FORD W. KENT, J. Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions. *The Astrophysical Journal*, v. 159, p. 379, fev. 1970. Disponível em: <<https://ui.adsabs.harvard.edu/abs/1970ApJ...159..379R>>.
- SARO, A. et al. Constraints on the richness–mass relation and the optical-SZE positional offset distribution for SZE-selected clusters. *Monthly Notices of the Royal Astronomical Society*, v. 454, n. 3, p. 2305–2319, 10 2015. Disponível em: <<https://doi.org/10.1093/mnras/stv2141>>.
- SHAPIRO, S. S.; WILK, M. B. An Analysis of Variance Test for Normality (Complete Samples). *Biometrika*, v. 52, p. 591–611, 1965. Disponível em: <<https://doi.org/10.2307/2333709>>.
- SHAW, L. D.; HOLDER, G. P.; DUDLEY, J. Non-Gaussian Scatter in Cluster Scaling Relations. *The Astrophysical Journal*, v. 716, n. 1, p. 281–285, jun. 2010. Disponível em: <<https://ui.adsabs.harvard.edu/abs/2010ApJ...716..281S>>.
- SMIRNOV, N. Sur les écarts de la courbe de distribution empirique. *Rec. Math. [Mat. Sbornik] N.S.*, v. 6(48), p. 3–26, 1939. Disponível em: <[https://www.mathnet.ru/php/archive.phtml?wshow=paper&jrnid=sm&paperid=5810&option\\_lang=eng](https://www.mathnet.ru/php/archive.phtml?wshow=paper&jrnid=sm&paperid=5810&option_lang=eng)>.
- The LSST Dark Energy Science Collaboration et al. The LSST Dark Energy Science Collaboration (DESC) Science Requirements Document. *arXiv e-prints*, p. arXiv:1809.01669, set. 2018. Disponível em: <<https://ui.adsabs.harvard.edu/abs/2018arXiv180901669T>>.
- VALENTINO, E. D. et al. In the realm of the hubble tension—a review of solutions. *Classical and Quantum Gravity*, IOP Publishing, v. 38, n. 15, p. 153001, jul 2021. Disponível em: <<https://dx.doi.org/10.1088/1361-6382/ac086d>>.

- VITENTI, S. D.; BARROSO, E. J. APES: approximate posterior ensemble sampler. *Monthly Notices of the Royal Astronomical Society*, v. 525, n. 1, p. 73–90, 07 2023. Disponível em: <<https://doi.org/10.1093/mnras/stad2245>>.
- ZWICKY, F. Die Rotverschiebung von extragalaktischen Nebeln. *Helvetica Physica Acta*, v. 6, p. 110–127, jan. 1933. Disponível em: <<https://ui.adsabs.harvard.edu/abs/1933AcHPh...6..110Z>>.

## Apêndices

## APÊNDICE A – PROBABILIDADE

Cálculos probabilísticos são essenciais para a análise de dados observacionais pois estes possuem uma incerteza associada a sua medição. Além disso, usando um ferramental estatístico é possível estimar propriedades intrínsecas através de observáveis de um objeto de estudo, como os aglomerados de galáxias, por exemplo.

Considerando um espaço amostral  $\Omega$  (espaço de todos os eventos possíveis), a probabilidade de se observar um evento  $X$ , resultado provindo desse espaço, é definida como

$$P(X) = \frac{n(X)}{n(\Omega)}, \quad (\text{A.1})$$

em que  $n(X)$  é o número de observações de  $X$  e  $n(\Omega)$  é o número total de eventos que podem ser medidos. A probabilidade também pode ser interpretada como o quanto de confiança se tem de que um evento  $X$  ocorra, dado um espaço amostral  $\Omega$ .

Se considerarmos dois eventos  $X$  e  $Y$ , a **probabilidade conjunta** desses dois eventos,  $P(X \cap Y)$  pode ser definida como a chance de  $X$  e  $Y$  ocorrerem simultaneamente e pode ser definida como

$$P(X, Y) = P(X \cap Y) = P(X)P(Y). \quad (\text{A.2})$$

Também pode-se calcular a probabilidade de  $X$  ou  $Y$  ocorrer, ou seja, a probabilidade desses eventos não ocorrerem juntos. Esta será dada pela junção da probabilidade de cada caso excluindo a situação em que os dois eventos ocorram simultaneamente, ou seja,

$$P(X \cup Y) = P(X) + P(Y) - P(X \cap Y). \quad (\text{A.3})$$

Caso o evento  $X$  esteja condicionado a ocorrência de  $Y$ , a **probabilidade condicional**  $P(X|Y)$  pode ser definida por

$$P(X|Y) = \frac{P(X \cap Y)}{P(Y)}. \quad (\text{A.4})$$

A probabilidade também pode ser calculada para conjuntos de variáveis contínuas dentro de um intervalo. Dado um conjunto  $w = \{w_i\}$  e outro  $v = \{v_i\}$ , pertencentes ao espaço  $\Omega$ , pode-se definir uma função densidade de probabilidade,  $p(w, v)$ , tal que a **probabilidade conjunta** de  $w$  e  $v$  em um intervalo  $\Gamma = \{a \leq w_i \leq b; c \leq v_i \leq d\} \in \Omega$  é

$$P(w, v \in \Gamma) = \int \int_{\Gamma} p(w, v) dw dv. \quad (\text{A.5})$$

As **distribuições marginais** de  $w$  e  $v$  são definidas como as densidades de probabilidade de cada variável independentemente, ou seja,

$$\begin{aligned} P(w) &= \int_v p(w, v)dv, \\ P(v) &= \int_w p(w, v)dw. \end{aligned} \tag{A.6}$$

Voltando à probabilidade condicional, a probabilidade de uma variável  $v$  condicionada a outra  $w$ ,  $P(v|w)$ , pode ser invertida, de forma conveniente, da seguinte forma:

Consideramos as probabilidades

$$P(w|v) = \frac{P(w \cap v)}{P(v)} \Rightarrow P(w \cap v) = P(w|v)P(v) \tag{A.7}$$

e

$$P(v|w) = \frac{P(v \cap w)}{P(w)} \Rightarrow P(v \cap w) = P(v|w)P(w). \tag{A.8}$$

Como  $P(w \cap v) = P(v \cap w)$ , podemos juntar A.7 e A.8 em uma única equação, a qual é denominada teorema de Bayes,

$$P(v|w) = \frac{P(w|v)P(v)}{P(w)}. \tag{A.9}$$

Aqui,  $P(v|w)$  é chamada de probabilidade posterior,  $P(w|v)$  é a verossimilhança,  $P(v)$  é a probabilidade anterior ou *prior* e  $P(w)$  é a evidência. Esta equação é muito importante na inferência de parâmetros a partir de dados observacionais, neste tipo de análise consideramos  $w$  como um conjunto de dados e  $v$  como um conjunto de parâmetros de um modelo.

## **Anexos**



