

UNIVERSIDADE DE SÃO PAULO

Instituto de Ciências Matemáticas e de Computação

**Misturas de regressões t de Student assimétricas com
número de componentes desconhecido: uma aplicação do
*Telescoping Sampler***

Marcus Gabriel da Silva e Silva

Dissertação de Mestrado do Programa Interinstitucional de
Pós-Graduação em Estatística (PIPGEs)

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: _____

Marcus Gabriel da Silva e Silva

Misturas de regressões t de Student assimétricas com
número de componentes desconhecido: uma aplicação do
Telescoping Sampler

Dissertação apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP e ao Departamento de Estatística – DEs-UFSCar, como parte dos requisitos para obtenção do título de Mestre em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística. *VERSÃO REVISADA*

Área de Concentração: Estatística

Orientador: Prof. Dr. Luis Aparecido Milan

Co-orientadora: Prof. Dr. Daiane Aparecida Zuanetti

USP – São Carlos
Fevereiro de 2025

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados inseridos pelo(a) autor(a)

S586m Silva, Marcus Gabriel da Silva e
Misturas de regressões t de Student assimétricas
com número de componentes desconhecido: uma
aplicação do Telescoping Sampler / Marcus Gabriel da
Silva e Silva; orientador Luis Aparecido Milan;
coorientador Daiane Aparecida Zuanetti. -- São
Carlos, 2025.
77 p.

Dissertação (Mestrado - Programa
Interinstitucional de Pós-graduação em Estatística) --
Instituto de Ciências Matemáticas e de Computação,
Universidade de São Paulo, 2025.

1. Análise de agrupamento. 2. Modelo de
regressão robusto. 3. MCMC. I. Milan, Luis
Aparecido, orient. II. Zuanetti, Daiane Aparecida ,
coorient. III. Título.

Marcus Gabriel da Silva e Silva

Mixture of skew-t regressions with an unknown number of components: an application of the Telescoping Sampler

Dissertation submitted to the Institute of Mathematics and Computer Science – ICMC-USP and to the Department of Statistics – DEs-UFSCar – in accordance with the requirements of the Statistics Interagency Graduate Program, for the degree of Master in Statistics. *FINAL VERSION*

Concentration Area: Statistics

Advisor: Prof. Dr. Luis Aparecido Milan

Co-advisor: Prof. Dr. Daiane Aparecida Zuanetti

USP – São Carlos
February 2025



UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Ciências Exatas e de Tecnologia
Programa Interinstitucional de Pós-Graduação em Estatística

Folha de Aprovação

Defesa de Dissertação de Mestrado do candidato Marcus Gabriel da Silva e Silva, realizada em 17/02/2025.

Comissão Julgadora:

Prof. Dr. Luis Aparecido Milan (UFSCar)

Prof. Dr. Erlandson Ferreira Saraiva (UFMS)

Prof. Dr. Celso Rômulo Barbosa Cabral (UFAM)

O Relatório de Defesa assinado pelos membros da Comissão Julgadora encontra-se arquivado junto ao Programa Interinstitucional de Pós-Graduação em Estatística.

AGRADECIMENTOS

Agradeço aos professores do PIPGEs pela contribuição com minha formação. Agradeço aos meus orientadores, Luiz Milan e Daiane Zuanetti, pelas sugestões e ensinamentos, além da ajuda durante a revisão deste texto. Ainda agradeço a professora Daiane pela ajuda na correção de códigos, sua paciência e atenção a detalhes fizeram a diferença.

Também agradeço a minha família. Minha mãe, Kézia Souza, por sempre ter se preocupado com minha formação. Minha avó, Vera Lúcia, pelo carinho e cuidado e ao meu irmão, Samuel Asaff, meu primeiro aluno, pelo apoio e admiração. Agradeço à minha noiva, Karollyna Beatriz, minha companheira nessa jornada acadêmica, pelo incentivo a continuar em minha carreira.

Agradeço aos colegas do PIPGEs, Natan Hilário e Samara Rilda, pela companhia e pela ajuda durante minha chegada à São Carlos, me ajustar ao novo ambiente teria sido bem mais demorado sem eles.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

RESUMO

SILVA, M. G. S. **Misturas de regressões t de Student assimétricas com número de componentes desconhecido: uma aplicação do *Telescoping Sampler***. 2025. 77 p. Dissertação (Mestrado em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2025.

Modelos de mistura de distribuições são adequados em situações em que existe heterogeneidade não observável na população. Comumente, o número de componentes presentes na mistura não é conhecido e uma forma de determiná-lo é fazendo o uso de medidas de seleção de modelos. Em um contexto bayesiano, a estimação dos parâmetros de cada componente em conjunto do número de componentes é possível e alguns algoritmos para tal foram propostos, sendo o *Telescoping Sampler* (TS) uma nova alternativa para a estimação simultânea de número de componentes e parâmetros em um contexto Bayesiano. Na análise de regressão, usualmente é feita a suposição de normalidade dos erros de observação, estendemos essa suposição considerando erros distribuídos segundo uma mistura de t de Student assimétricas, desta forma comportando dados com subgrupos latentes, assimetria e presença de *outliers*. No que segue, apresentamos o TS e o modelo de misturas de regressões t de Student assimétricas, investigamos a estimação de parâmetros do modelo de misturas de regressões utilizando dados simulados e ajustamos o modelo proposto em um conjunto de dados composto de salários de jogadores de baseball e medidas de sua performance durante o jogo. O algoritmo TS foi capaz de recuperar o número de componentes e parâmetros fixados para a simulação e vemos que as estimativas dos coeficientes de regressão são consistentes.

Palavras-chave: Análise de agrupamento, Modelo de regressão robusto, MCMC.

ABSTRACT

SILVA, M. G. S. **Mixture of skew-t regressions with an unknown number of components: an application of the Telescoping Sampler**. 2025. 77 p. Dissertação (Mestrado em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2025.

Mixture models are suitable in situations where there is unobservable heterogeneity in the population. Commonly, the number of components present in the mixture is not known, one way to determine it is by using model selection criteria. In a Bayesian context, the estimation of the parameters of each component alongside the number of components is possible, several algorithms have been proposed for this end. The *Telescoping Sampler* (TS) is a new alternative for the simultaneous estimation of the number of components and parameters in a Bayesian context. In regression analysis, the assumption of normality of observation errors is usually made. We extend this assumption by considering errors distributed according to a mixture of Skew-t distributions, thus comprising data with latent subgroups, skewness and the presence of *outliers*. In what follows, we present the TS and the mixture of Skew-t regressions, we investigate the parameter estimation of the regression mixture model using simulated data and fit the proposed model to a dataset comprising of baseball player salaries and measures of their in-game performance. The TS algorithm was able to recover the number of components and parameters fixed for the simulation and we see that the estimates of the regression coefficients in the regression mixture model are consistent.

Keywords: Clustering analysis, Robust regression model, MCMC.

LISTA DE ILUSTRAÇÕES

Figura 1 – Função densidade da distribuição ST para $\lambda = 0, 2, 5, 8$, $\sigma^2 = 1$, $\nu = 3$ e $\mu = -\frac{\lambda}{\sqrt{1+\lambda^2}} \frac{\sqrt{3/\pi}}{\Gamma(3/2)}$.	30
Figura 2 – Distribuição <i>a priori</i> para o número de grupos em um MFM, $\pi(G_+ \gamma = 1)$, quando $G \sim \text{Uniforme}\{1, \dots, 30\}$.	34
Figura 3 – Distribuições <i>a priori</i> de G e G_+ , considerando $G - 1 \sim \text{BNB}(1, 4, 3)$, $G - 1 \sim \text{Geométrica}(0, 1)$ e $G \sim \text{Uniforme}\{1, \dots, 30\}$, para um MFM e MFM generalizado, $n = 82$.	35
Figura 4 – Distribuição <i>a priori</i> dos grupos em um SFM.	36
Figura 5 – Média <i>a posteriori</i> de λ menos seu valor fixado, reta pontilhada marca zero. Nas linhas temos $\lambda = 0$ (linha 1), 2 (linha 2) e 5 (linha 3), e nas colunas $\sigma^2 = 1$ (coluna 1), 4 (coluna 2) e 9 (coluna 3).	52
Figura 6 – Intervalos de credibilidade para λ no painel, reta pontilhada marca o valor fixado de λ e a cor do intervalo é vermelha caso o mesmo não contenha o valor fixado. Nas linhas temos $\lambda = 0$ (linha 1), 2 (linha 2) e 5 (linha 3), e nas colunas $\sigma^2 = 1$ (coluna 1), 4 (coluna 2) e 9 (coluna 3).	53
Figura 7 – Diagramas de dispersão para o primeiro conjunto de dados obtido com $n = 100$ e $n = 1000$.	54
Figura 8 – Gráfico quantil-quantil dos resíduos quantílicos. A reta pontilhada é a reta $y = x$ e os segmentos representam os intervalos de credibilidade (95%) de cada resíduo, em cinza se contém a reta pontilhada e em vermelho caso contrário.	57
Figura 9 – Média ergódica por réplica para os tamanhos amostrais $n = 100$ e $n = 1000$ nos dados simulados.	75
Figura 10 – Autocorrelação da cadeia por réplica para os tamanhos amostrais $n = 100$ e $n = 1000$ nos dados simulados.	76
Figura 11 – Média ergódica e autocorrelação da cadeia por réplica, respectivamente, para os dados dos salários de jogadores de baseball.	76

LISTA DE ALGORITMOS

Algoritmo 1 – Algoritmo de Metropolis-Hastings.	43
Algoritmo 2 – Amostrador de Gibbs	44
Algoritmo 3 – <i>Telescoping Sampler</i> para o modelo MRST.	46

LISTA DE TABELAS

Tabela 1 – Estimativas <i>a posteriori</i> para $n = 100$	55
Tabela 2 – Estimativas <i>a posteriori</i> para $n = 1000$	56
Tabela 3 – Estimativas pela média e mediana <i>a posteriori</i> de cada parâmetro do modelo além de um intervalo de credibilidade (IC) de nível 95%.	58

SUMÁRIO

1	INTRODUÇÃO	21
2	METODOLOGIA	25
2.1	Distribuições de probabilidade	25
2.1.1	<i>Multinomial e categórica</i>	25
2.1.2	<i>Dirichlet</i>	26
2.1.3	<i>Normal assimétrica</i>	27
2.1.4	<i>T de Student assimétrica</i>	28
2.2	Modelos de misturas finitas	29
2.2.1	<i>Modelo de misturas de regressões t de Student assimétricas</i>	36
2.3	Problemas de identificabilidade	39
2.4	Técnicas MCMC e procedimento de estimação	42
2.4.1	<i>Algoritmo de Metropolis-Hastings</i>	42
2.4.2	<i>Amostrador de Gibbs</i>	43
2.4.3	<i>Telescoping Sampler</i>	44
2.5	Análise residual	48
3	RESULTADOS	51
3.1	Modelo simplificado	51
3.2	Modelo completo	53
3.2.1	<i>Dados simulados</i>	53
3.2.2	<i>Salários de jogadores de baseball</i>	56
4	DISCUSSÃO	59
	REFERÊNCIAS	61
APÊNDICE A	CONDICIONAIS COMPLETAS PARA O MRST	65
APÊNDICE B	DISTRIBUIÇÕES CONDICIONAIS COMPLETAS PARA O TELESCOPING SAMPLER	71
APÊNDICE C	VERIFICAÇÕES DE CONVERGÊNCIA DAS CADEIAS MCMC	75

INTRODUÇÃO

Uma das principais distribuições de probabilidades utilizadas na modelagem estatística é a distribuição normal. Um dos motivos se deve ao fato de diversos resultados probabilísticos com base na distribuição normal serem conhecidos, o que por muitas vezes facilita o processo inferencial. Porém, é possível que os dados apresentem características não comportadas pela distribuição normal, por exemplo, assimetria ou presença de grande número de observações distantes da média. Uma opção para a modelagem de dados com presença de assimetria e *outliers* é a distribuição t de Student assimétrica (ST, do inglês *Skew-t*), uma versão assimétrica da distribuição t de Student. A distribuição ST faz parte da família de distribuições inicialmente introduzidas por [Branco e Dey \(2001\)](#), a distribuição ST também foi estudada posteriormente por [Azzalini e Capitanio \(2003\)](#), que apresentam algumas propriedades da distribuição.

Em modelos de regressão, usualmente é feita a suposição que os erros de observação são normalmente distribuídos, mas podemos flexibilizar esta suposição impondo distribuição ST aos erros, desta forma comportando a presença de assimetria e *outliers* nos dados. [Branco e Dey \(2002\)](#) e [Sahu, Dey e Branco \(2003\)](#) descrevem o modelo de regressão t Student assimétrico em um contexto Bayesiano e [Azzalini e Capitanio \(2003\)](#) o apresentam sob uma ótica clássica, [Zhou \(2005\)](#) estende o modelo para comportar efeitos mistos, [Zeller, Lachos e Vilca-Labra \(2011\)](#) elaboram um modelo de regressão cuja distribuição dos erros de observação está na família de misturas de escala da normal assimétrica (SMSN, do inglês *Scale Mixture of Skew-Normal*), família de distribuições que inclui a ST, [Cancho et al. \(2011\)](#) estende o modelo considerando regressão não linear.

Uma extensão dos modelos de regressão com erros normais considerando misturas finitas de distribuições foi inicialmente proposta por [Quandt e Ramsey \(1978\)](#), extensões desse modelo considerando erros na família SMSN são abordados por [Zeller, Cabral e Lachos \(2016\)](#), [Massuia et al. \(2017\)](#) e [Cabral, Souza e Leão \(2022\)](#). Para uma revisão sobre misturas finitas de distribuições na família SMSN veja [Dávila, Cabral e Zeller \(2018\)](#). Misturas finitas de

distribuições são úteis para a modelagem de dados provenientes de uma população que apresenta heterogeneidade não observável.

Em uma abordagem clássica, a estimação dos parâmetros do modelo de misturas de regressão, seja ele com erros normais ou SMSN, é usualmente realizada pelo algoritmo EM (DEMPSTER; LAIRD; RUBIN, 1977) com número de componentes fixo, seu valor é determinado posteriormente utilizando medidas de seleção de modelos, veja McLachlan e Krishnan (2007) e Zeller, Cabral e Lachos (2016). Em uma abordagem Bayesiana, um procedimento similar já foi realizado por Nascimento (2017) para modelos de misturas de regressão com erros SMSN, com técnicas de Monte Carlo em Cadeias de Markov (MCMC, do inglês *Markov Chain Monte Carlo*) para a estimação e medidas de seleção de modelos para a determinação do número de componentes. Porém, neste contexto, a estimação simultânea do número de componentes e parâmetros é possível.

Em uma abordagem Bayesiana, o algoritmo *Reversible Jump* é uma das opções para a estimação simultânea de parâmetros e seleção de modelos. Introduzido por Green (1995), com este algoritmo podemos atribuir uma distribuição *a priori* ao número de componentes e então obter uma Cadeia de Markov cuja a distribuição estacionária é a *a posteriori* conjunta do número de componentes e parâmetros do modelo. Aplicações do *Reversible Jump* a misturas de regressão com erros normais são apresentadas por Liu *et al.* (2015) e Sabillón, Cotrim e Zuanetti (2023). Restrito aos modelos de mistura, temos os modelos de mistura por processo de Dirichlet (ANTONIAK, 1974; ESCOBAR, 1994), baseados no processo de Dirichlet (FERGUSON, 1973), nesta classe de modelos, é feita a distinção entre o número de componentes na população e o número de componentes presentes na amostra, ao qual chamaremos de grupos. Os modelos de mistura por processo de Dirichlet apresentam uma quantidade infinita de parâmetros, de certa maneira, podemos dizer que o número de componentes na população é fixado em $+\infty$. Aplicações dos modelos de misturas por processo de Dirichlet aos modelos de regressão com erros normais podem ser encontradas em West e Escobar (1994), Müller, Erkanli e West (1996). Hannah, Blei e Powell (2011) estendem para modelos lineares generalizados.

Uma nova abordagem para a estimação simultânea do número de componentes e parâmetros em modelos de misturas em um contexto Bayesiano foi recentemente introduzida por Frühwirth-Schnatter, Malsiner-Walli e Grün (2021), os modelos de misturas de misturas finitas (MFM, do inglês *Mixture of Finite Mixtures*) generalizados. Para esta classe de modelos, um amostrador do tipo MCMC foi construído, chamado *Telescoping Sampler*. Com os MFM generalizados é possível atribuir uma distribuição *a priori* para o número de componentes, e como nos modelos de mistura por processo de Dirichlet, é feita distinção entre o número de grupos e componentes, porém o número de componentes não é fixo.

Neste trabalho, estudamos modelos de misturas finitas de regressões SMSN em uma abordagem Bayesiana, como feito por Nascimento (2017), com foco na distribuição ST, porém, levaremos em consideração a estimação simultânea do número de componentes, grupos e

parâmetros por meio do *Telescoping Sampler*.

O restante do trabalho está organizado da seguinte maneira. No Capítulo 2 apresentamos a teoria necessária para o desenvolvimento deste trabalho, apresentamos distribuições de probabilidade, modelos de mistura, modelos de mistura de regressão, problemas de identificabilidade em modelos de mistura, algoritmos inferenciais e análise residual. No Capítulo 3 apresentamos resultados da aplicação das técnicas discutidas no Capítulo 2 a conjuntos de dados simulados e a um conjunto de dados referente a salários de jogadores de baseball. No Capítulo 4 reunimos nossas conclusões e pretensões futuras.

METODOLOGIA

2.1 Distribuições de probabilidade

Nesta seção, apresentamos distribuições de probabilidade, resultados e propriedades relevantes para este trabalho.

2.1.1 Multinomial e categórica

Dizemos que o vetor aleatório discreto $\mathbf{S} = (S_1, S_2, \dots, S_G)^\top$ segue distribuição multinomial com parâmetros n e $\mathbf{p} = (p_1, p_2, \dots, p_G)^\top$, denotamos isto por $\mathbf{S} \sim \text{Multinomial}(n; \mathbf{p})$, se sua função de probabilidade é dada por

$$\mathbb{P}(\mathbf{S} = \mathbf{s} | n, \mathbf{p}) = \frac{n!}{s_1! s_2! \dots s_G!} \prod_{j=1}^G p_j^{s_j}, \quad (2.1)$$

em que \mathbf{p} é um vetor de probabilidades, isto é, $\sum_{j=1}^G p_j = 1, p_j \in [0, 1]$ e $\mathbf{s} = (s_1, s_2, \dots, s_G)^\top$. O suporte desta distribuição é o conjunto

$$\left\{ (s_1, s_2, \dots, s_G) \in \mathbb{Z}^G \mid \sum_{j=1}^G s_j = n, s_j \geq 0 \forall j \geq 1 \right\}.$$

A distribuição multinomial modela o número de ocorrências de cada resultado possível de um experimento com G resultados replicado n vezes de forma independente. Por exemplo, considere replicar 10 vezes o lançamento de um dado com faces 1, 2, 3, 4, 5, 6 equiprováveis. Definindo o vetor aleatório $\mathbf{S} = (S_1, S_2, S_3, S_4, S_5, S_6)^\top$, com S_j sendo o número de ocorrências da face j , $j = 1, 2, 3, 4, 5, 6$. A probabilidade de serem observadas duas vezes as faces 1, 2, 3, 4 e uma vez

as demais é de

$$\mathbb{P} \left(\mathbf{S} = \begin{pmatrix} 2 \\ 2 \\ 2 \\ 2 \\ 1 \\ 1 \end{pmatrix} \middle| 10, \begin{pmatrix} 1/6 \\ 1/6 \\ 1/6 \\ 1/6 \\ 1/6 \\ 1/6 \end{pmatrix} \right) = \frac{10!}{2!2!2!2!} (1/6)^2 (1/6)^2 (1/6)^2 (1/6)^2 (1/6)^1 (1/6)^1 \approx 0,004.$$

Um caso particular da distribuição multinomial importante para este trabalho é quando $n = 1$. Seja $\mathbf{S} \sim \text{Multinomial}(1; \mathbf{p})$, a função de probabilidade de \mathbf{S} é dada por

$$\mathbb{P}(\mathbf{S} = \mathbf{s} | 1, \mathbf{p}) = p_j,$$

em que j é tal que $s_j = 1, j = 1, 2, \dots, G$. Neste caso particular, podemos definir a variável aleatória $Z = j \Leftrightarrow S_j = 1, j = 1, 2, \dots, G$, note que

$$\mathbb{P}(Z = j | \mathbf{p}) = \mathbb{P}(\mathbf{S} = \mathbf{s} | 1, \mathbf{p}) = p_j, j = 1, 2, \dots, G$$

se $\mathbf{s} = (s_1, s_2, \dots, s_G)$ é tal que $s_j = 1$. Dizemos que a variável aleatória Z segue distribuição categórica, escrevemos $Z \sim \text{Cat}(\mathbf{p})$.

2.1.2 Dirichlet

Dizemos que o vetor aleatório $\mathbf{X} = (X_1, X_2, \dots, X_G)^\top$ segue uma distribuição Dirichlet com vetor de parâmetros $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \dots, \gamma_G)^\top$, denotado por $\mathbf{X} \sim \text{Dirichlet}(\boldsymbol{\gamma})$, se sua função densidade é dada por

$$f(\mathbf{x} | \boldsymbol{\gamma}) = \frac{\Gamma\left(\sum_{j=1}^G \gamma_j\right)}{\prod_{j=1}^G \Gamma(\gamma_j)} \prod_{j=1}^G x_j^{\gamma_j-1}, \quad (2.2)$$

em que $\mathbf{x} = (x_1, x_2, \dots, x_G)^\top$. O suporte desta distribuição é o conjunto

$$\left\{ x_j \in [0, 1] \forall j \geq 1 \mid \sum_{j=1}^G x_j = 1 \right\}. \quad (2.3)$$

O valor esperado e variância da j -ésima componente de um vetor $\mathbf{X} \sim \text{Dirichlet}(\boldsymbol{\gamma})$ são dados, respectivamente, por:

$$\mathbb{E}[X_j] = \frac{\gamma_j}{\gamma_0} \text{ e } \mathbb{V}[X_j] = \frac{\gamma_j(\gamma_0 - \gamma_j)}{\gamma_0^2(1 + \gamma_0)}, \quad (2.4)$$

com $\gamma_0 = \sum_{j=1}^G \gamma_j$.

Um caso particular da distribuição de Dirichlet é a distribuição de Dirichlet simétrica que ocorre quando $\boldsymbol{\gamma} = (\gamma, \gamma, \dots, \gamma)^\top$. Denotamos um vetor aleatório $\mathbf{X} = (X_1, X_2, \dots, X_G)^\top$ que segue uma distribuição Dirichlet simétrica de parâmetro γ por $\mathbf{X} \sim \text{DirichSim}_G(\gamma)$.

A distribuição de Dirichlet é a distribuição *a priori* natural para vetores de probabilidade por conta das características de seu suporte, mostradas em (2.3). Outra característica notável da distribuição Dirichlet é sua conjugação quando utilizada como distribuição *a priori* em um modelo multinomial.

Por exemplo, se temos o resultado de n realizações independentes de um experimento que possui G possíveis resultados, podemos supor que cada resultado possui probabilidade de ocorrência $p_j, j = 1, 2, \dots, G$ e modelamos o número de ocorrências de cada resultado pela distribuição multinomial, cuja função de probabilidade é dada por (2.1). Seja $\mathbf{Y} = (Y_1, Y_2, \dots, Y_G)^\top$ o vetor aleatório das possíveis contagens de cada resultado do experimento, $\mathbf{p} = (p_1, p_2, \dots, p_G)^\top$ as probabilidades de cada resultado e $\mathbf{y} = (y_1, y_2, \dots, y_G)^\top$ o resultado observado de n replicações deste experimento. Supomos $\mathbf{Y} \sim \text{Multinomial}(n, \mathbf{p})$ e tomamos $\mathbf{p} \sim \text{Dirichlet}(\boldsymbol{\gamma})$. Desta forma, é possível mostrar que, a distribuição *a posteriori* é conjugada e é dada por $\mathbf{p}|\mathbf{y} \sim \text{Dirichlet}(\mathbf{y} + \boldsymbol{\gamma})$.

2.1.3 Normal assimétrica

Conforme definido por Azzalini (1985), dizemos que uma variável aleatória Y tem distribuição normal assimétrica com parâmetros de localização $\mu \in \mathbb{R}$, de dispersão $\sigma^2 > 0$ e de forma $\lambda \in \mathbb{R}$, quando a densidade de Y é dada por

$$\text{SN}(y|\mu, \sigma^2, \lambda) = 2\phi(y|\mu; \sigma^2)\Phi\left(\lambda\frac{y-\mu}{\sigma}\right), \quad (2.5)$$

em que $\phi(\cdot|\mu, \sigma^2)$ denota a função densidade da distribuição normal de média μ e variância σ^2 , $\Phi(\cdot)$ denota a função distribuição da normal padrão. Usamos a notação $Y \sim \text{SN}(\mu, \sigma^2, \lambda)$, note que, caso $\lambda = 0$, temos que $Y \sim \text{N}(\mu, \sigma^2)$.

De acordo com Henze (1986), se temos duas variáveis aleatórias independentes $T_1 \sim \text{N}(0, 1)$ e $T_2 \sim \text{N}_{(0, \infty)}(0, 1)$, ou seja, T_2 segue distribuição normal padrão truncada no intervalo $(0, +\infty)$. Sejam $\mu, \lambda \in \mathbb{R}$ e $\sigma > 0$, tome $\delta = \lambda/\sqrt{1 + \lambda^2}$ e faça

$$Y = \mu + \sigma(1 - \delta^2)^{1/2}T_1 + \sigma\delta T_2. \quad (2.6)$$

Temos que $Y \sim \text{SN}(\mu, \sigma^2, \lambda)$. Ou seja, uma variável aleatória normal assimétrica Y pode ser escrita como uma combinação linear das variáveis aleatórias T_1 e T_2 .

Seja $Y \sim \text{SN}(\mu, \sigma^2, \lambda)$, $a > 0$ e $b \in \mathbb{R}$, por (2.6), $aY + b$ tem a mesma distribuição que $a\mu + b + a\sigma(1 - \delta^2)^{1/2}T_1 + a\sigma\delta T_2$, que está na mesma forma do resultado (2.6), então temos que

$$aY + b \sim \text{SN}(a\mu + b, a^2\sigma^2, \lambda). \quad (2.7)$$

Em (2.6) note que, dado a realização da variável aleatória $T_2 = t_2$, temos uma transformação linear de uma variável aleatória normal, ou seja, $Y|T_2 = t_2 \sim \text{N}(\mu + \sigma\delta t_2, \sigma^2(1 - \delta^2))$ visto que $T_1 \sim \text{N}(0, 1)$. Obtemos assim a seguinte representação hierárquica de uma variável

aleatória normal assimétrica:

$$Y|T_2 = t_2 \sim N(\mu + \sigma\delta t_2, \sigma^2(1 - \delta^2)); \quad T_2 \sim N_{(0,+\infty)}(0, 1). \quad (2.8)$$

Relembre-se do seguinte resultado probabilístico: sejam X e Y duas variáveis aleatórias, é possível calcular, respectivamente, o valor esperado e variância da variável aleatória X da seguinte maneira

$$\begin{aligned} \mathbb{E}(X) &= \mathbb{E}(\mathbb{E}(X|Y)) \\ \mathbb{V}(X) &= \mathbb{V}(\mathbb{E}(X|Y)) + \mathbb{E}(\mathbb{V}(X|Y)). \end{aligned} \quad (2.9)$$

Para uma revisão de teoria da probabilidade veja [Ross \(2009\)](#). Utilizando (2.9) em conjunto de (2.8) obtemos o valor esperado e variância de uma variável aleatória normal assimétrica.

$$\begin{aligned} \mathbb{E}(Y) &= \mu + \sigma\delta\sqrt{\frac{2}{\pi}}, \\ \mathbb{V}(Y) &= \sigma^2 \left[1 - \frac{2}{\pi}\delta^2 \right]. \end{aligned} \quad (2.10)$$

2.1.4 T de Student assimétrica

Segundo [Azzalini e Capitanio \(2003\)](#), definimos a distribuição t de Student assimétrica (ST, do inglês *Skew-t*) por meio da transformação

$$Y = \mu + U^{-1/2}X, \quad (2.11)$$

com $\mu \in \mathbb{R}$, $X \sim \text{SN}(0, \sigma^2, \lambda)$ e $U \sim \text{Gama}(v/2, v/2)$ independentes. Utilizamos a notação $Y \sim \text{ST}(\mu, \sigma^2, \lambda, v)$.

Utilizando (2.6) podemos reescrever a variável aleatória X em (2.11). Então se $U \sim \text{Gama}(v/2, v/2)$, $T_1 \sim N(0, 1)$ e $T_2 \sim N_{(0, \infty)}(0, 1)$ são variáveis aleatórias independentes, $\mu, \lambda \in \mathbb{R}$ e $\sigma > 0$, tome $\delta = \lambda/\sqrt{1 + \lambda^2}$ e faça

$$Y = \mu + \sigma(1 - \delta^2)^{1/2}U^{-1/2}T_1 + \sigma\delta U^{-1/2}T_2, \quad (2.12)$$

temos que $Y \sim \text{ST}(\mu, \sigma^2, \lambda, v)$. Fixando $U = u$ em (2.12), note que $T = u^{-1/2}T_2 \sim N_{(0, \infty)}(0, u^{-1})$, portanto $Y|T = t, U = u \sim N(\mu + \Delta t, u^{-1}\tau^2)$, com $\tau^2 = \sigma^2(1 - \delta^2)$ e $\Delta = \sigma\delta$. Assim, podemos ver que uma variável aleatória $Y \sim \text{ST}(\mu, \sigma^2, \lambda, v)$ admite representação hierárquica

$$\begin{aligned} U &\sim \text{Gama}(v/2, v/2); \\ T|U = u &\sim N_{(0, \infty)}(0, u^{-1}); \\ Y|T = t, U = u &\sim N(\mu + \Delta t, u^{-1}\tau^2). \end{aligned} \quad (2.13)$$

Além disso, de forma análoga a (2.7), se temos $Y \sim \text{ST}(\mu, \sigma^2, \lambda, v)$, $a > 0$ e $b \in \mathbb{R}$, por (2.12), $aY + b$ tem a mesma distribuição que $a\mu + b + a\sigma(1 - \delta^2)^{1/2}U^{-1/2}T_1 + a\sigma\delta U^{-1/2}T_2$, que está na mesma forma do resultado (2.12), então temos que

$$aY + b \sim \text{ST}(a\mu + b, a^2\sigma^2, \lambda, v). \quad (2.14)$$

Como demonstrado em [Azzalini e Capitanio \(2003\)](#), a função densidade de Y é dada por

$$\text{St}(y|\mu, \sigma^2, \lambda, \nu) = 2t(y|\mu, \sigma^2, \nu)T_{\nu+1} \left(\lambda \frac{y-\mu}{\sigma} \sqrt{\frac{\nu+1}{\left(\frac{y-\mu}{\sigma}\right)^2 + \nu}} \right), \quad (2.15)$$

em que $t(\cdot|\mu, \sigma^2, \nu)$ denota a função densidade da distribuição t de Student com parâmetro de localização μ , de escala σ^2 e ν graus de liberdade, $T_\nu(\cdot)$ denota a função distribuição da t de Student padrão, ou seja, $\mu = 0$ e $\sigma^2 = 1$, com ν graus de liberdade. Como comentado pelo autor, esta densidade coincide com a densidade da t de Student assimétrica apresentada em [Branco e Dey \(2001\)](#).

É possível utilizar o resultado (2.9) em (2.11), visto que temos o valor esperado e variância de uma variável aleatória normal assimétrica em (2.10), obtendo assim o valor esperado e variância de uma variável aleatória $Y \sim \text{ST}(\mu, \sigma^2, \lambda, \nu)$ que são dados respectivamente por

$$\begin{aligned} \mathbb{E}(Y) &= \mu + \sigma \delta \sqrt{\frac{2}{\pi}} \mathbb{E}(U^{-1/2}), \\ \mathbb{V}(Y) &= \sigma^2 \left[\mathbb{E}(U^{-1}) - \frac{2}{\pi} \delta^2 \mathbb{E}^2(U^{-1/2}) \right]. \end{aligned}$$

Com $\mathbb{E}(U^{-1}) = \frac{\nu}{\nu-2}$, se $\nu > 2$, divergindo caso contrário, e $\mathbb{E}(U^{-1/2}) = \sqrt{\nu/2} \frac{\Gamma(\nu/2-1/2)}{\Gamma(\nu/2)}$, se $\nu > 1$, divergindo caso contrário. Temos assim que, se $Y \sim \text{ST}(\mu, \sigma^2, \lambda, \nu)$,

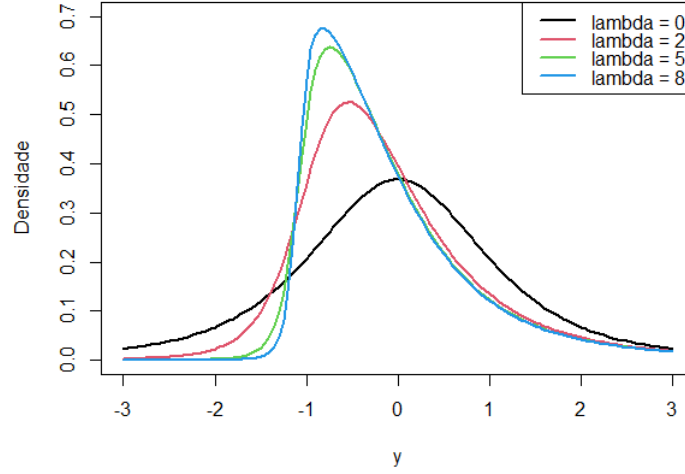
$$\begin{aligned} \mathbb{E}(Y) &= \mu + \sigma \delta \sqrt{\frac{\nu \Gamma(\frac{\nu-1}{2})}{\pi \Gamma(\nu/2)}}, \text{ se } \nu > 1; \\ \mathbb{V}(Y) &= \sigma^2 \left[\frac{\nu}{\nu-2} - \frac{\nu}{\pi} \delta^2 \left(\frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\nu/2)} \right)^2 \right], \text{ se } \nu > 2. \end{aligned} \quad (2.16)$$

A Figura 1 mostra a função densidade da distribuição ST para vários valores de λ , com $\nu = 3$, $\sigma^2 = 1$ e $\mu = -\frac{\lambda}{\sqrt{1+\lambda^2}} \frac{\sqrt{3/\pi}}{\Gamma(3/2)}$, desta forma o valor esperado é zero em todos os casos. É possível observar que o formato da função densidade não é muito afetado por valores elevados de λ , pois não há muita diferença no formato para $\lambda = 5$ e $\lambda = 8$, o que pode implicar em dificuldades em sua estimação posteriormente.

2.2 Modelos de misturas finitas

Definimos uma mistura finita de distribuições definindo uma variável de alocação $Z|\mathbf{p} \sim \text{Cat}(\mathbf{p})$, com $(p_1, p_2, \dots, p_G)^\top$. A variável de alocação Z assume valores no conjunto $\{1, 2, \dots, G\}$, $G \in \mathbb{N}$ com $\mathbb{P}(Z = j|\mathbf{p}) = p_j$, $j = 1, 2, \dots, G$. Fazemos $Y|Z = j, \boldsymbol{\theta}_j \sim F_j(\cdot|\boldsymbol{\theta}_j)$,

Figura 1 – Função densidade da distribuição ST para $\lambda = 0, 2, 5, 8$, $\sigma^2 = 1$, $\nu = 3$ e $\mu = -\frac{\lambda}{\sqrt{1+\lambda^2}} \frac{\sqrt{3/\pi}}{\Gamma(3/2)}$.



Fonte: Elaborada pelo autor.

note que a função distribuição marginal de Y é dada por

$$\begin{aligned} F(y|\boldsymbol{\theta}, \mathbf{p}) &= \sum_{j=1}^G \mathbb{P}(Y \leq y|Z = j)\mathbb{P}(Z = j) \\ &= \sum_{j=1}^G p_j F_j(y|\boldsymbol{\theta}_j). \end{aligned} \quad (2.17)$$

Dizemos que Y é uma mistura finita de G componentes, em que cada componente possui distribuição $F_1(\cdot|\boldsymbol{\theta}_1), F_2(\cdot|\boldsymbol{\theta}_2), \dots, F_G(\cdot|\boldsymbol{\theta}_G)$, $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_G)^\top$ e $\mathbf{p} = (p_1, p_2, \dots, p_G)^\top$.

O modelo de misturas finitas pode ser definido de forma hierárquica. Considere um vetor de observações independentes $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)^\top$ e fixe um número de componentes G . O modelo, na perspectiva Bayesiana, é dado por:

$$\begin{aligned} \mathbf{p} &\sim \text{Dirichlet}(\boldsymbol{\gamma}); \\ \boldsymbol{\theta}_j &\stackrel{\text{ind}}{\sim} \pi(\cdot|\boldsymbol{\varphi}), j = 1, 2, \dots, G; \\ Z_i|\mathbf{p} &\stackrel{\text{ind}}{\sim} \text{Cat}(\mathbf{p}), i = 1, 2, \dots, n; \\ Y_i|Z_i = j, \boldsymbol{\theta}_j &\stackrel{\text{ind}}{\sim} F(\cdot|\boldsymbol{\theta}_j), i = 1, 2, \dots, n, \end{aligned} \quad (2.18)$$

em que $\pi(\cdot|\boldsymbol{\varphi})$ denota uma distribuição *a priori* com hiperparâmetro $\boldsymbol{\varphi}$. Note que supomos a mesma família de distribuições F para todas as componentes, com diferenças apenas em seus parâmetros.

A escolha da distribuição *a priori* Dirichlet para \mathbf{p} deve-se ao fato de \mathbf{p} ser um vetor de probabilidades, a escolha *a priori* usual para o vetor $\boldsymbol{\gamma}$ é fixar cada um de seus elementos em uma constante γ , ou seja, fixar $\boldsymbol{\gamma} = (\gamma, \gamma, \dots, \gamma)^\top$ e por consequência utilizar uma distribuição *a priori*

Dirichlet simétrica de parâmetro γ . Esta especificação é bastante recorrente na literatura por conta de sua função densidade ser invariante a permutação nos rótulos das componentes, o mesmo comportamento é observado na função de verossimilhança dos modelos de mistura, mostrada em (2.19), fazendo esta suposição ser natural (MILLER; HARRISON, 2018). A suposição de independência e distribuição *a priori* comum entre parâmetros de componentes distintas visa facilitar a implementação de algoritmos de estimação, o que será abordado na Seção 2.4.

Utilizando a função densidade de probabilidade obtida a partir do resultado (2.17) e a suposição de independência entre observações, vemos que a função de verossimilhança deste modelo é dada por:

$$L(\mathbf{y}|\boldsymbol{\theta}) = \prod_{i=1}^n \sum_{j=1}^G p_j f(y_i|\boldsymbol{\theta}_j), \quad (2.19)$$

em que $f(\cdot|\boldsymbol{\theta}_j)$ denota a função densidade (ou massa) de probabilidade de $Y_i|Z_i = j, \boldsymbol{\theta}_j, \boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_G, \mathbf{p})^\top$ e $\mathbf{p} = (p_1, p_2, \dots, p_G)^\top$.

Também é possível obter a verossimilhança aumentada, em que supomos a observação da variável latente Z_i , isto ajudará no processo de estimação discutido na Seção 2.4. Podemos observar que a conjunta é $\pi(y_i, z_i) = p_j f(y_i|\boldsymbol{\theta}_j)$, então a verossimilhança aumentada é dada por:

$$\begin{aligned} L(\mathbf{y}, \mathbf{z}|\boldsymbol{\theta}) &= \prod_{j=1}^G \prod_{i=1}^n p_j^{\mathbb{1}(z_i=j)} f(y_i|\boldsymbol{\theta}_j)^{\mathbb{1}(z_i=j)} \\ &= \prod_{j=1}^G p_j^{m_j} \prod_{i=1}^n f(y_i|\boldsymbol{\theta}_j)^{\mathbb{1}(z_i=j)}, \end{aligned} \quad (2.20)$$

em que $\mathbb{1}(z_i = j)$ denota a função indicadora do evento $z_i = j$ e $m_j = \sum_{i=1}^n \mathbb{1}(z_i = j)$, note que o número de observações na componente j é dado por m_j .

Uma consideração importante neste modelo é a determinação do número de componentes por não ser uma informação usualmente conhecida. A estimação do número de componentes pode ser realizada do ponto de vista de seleção de modelos, ajustando um modelo com G fixado em diversos valores e escolhendo aquele com melhor valor de alguma medida de seleção de modelos.

O procedimento de seleção de modelos também pode ser realizado atribuindo uma distribuição *a priori* para G e estimando seu valor a partir da distribuição *a posteriori* obtida, note que ao fazer isto o espaço paramétrico do modelo se torna aleatório, o que dificulta o processo de estimação. Esta pode ser considerada uma abordagem mais natural do ponto de vista Bayesiano, visto que podemos quantificar a incerteza com relação ao número de componentes através da distribuição *a priori*. Além disso, podemos obter a probabilidade *a posteriori* de cada possível valor de G , ou seja, não precisamos estimar todos os possíveis modelos, que pode ser

computacionalmente muito custoso. Considere então o seguinte modelo:

$$\begin{aligned}
G &\sim \pi_G(\cdot); \\
\mathbf{p}|G &\sim \text{DirichSim}_G(\gamma); \\
\boldsymbol{\theta}_j|G &\stackrel{\text{ind}}{\sim} \pi(\cdot|\boldsymbol{\varphi}), j = 1, 2, \dots, G; \\
Z_i|G, \mathbf{p} &\stackrel{\text{ind}}{\sim} \text{Cat}(\mathbf{p}), i = 1, 2, \dots, n; \\
Y_i|G, Z_i = j, \boldsymbol{\theta}_j &\stackrel{\text{ind}}{\sim} F(\cdot|\boldsymbol{\theta}_j), i = 1, 2, \dots, n,
\end{aligned} \tag{2.21}$$

com G assumindo valores em \mathbb{N} . O modelo de mistura com distribuição *a priori* em G é chamado de modelo de mistura de misturas finitas (MFM, do inglês *Mixture of Finite Mixtures*) por [Miller e Harrison \(2018\)](#), adotaremos essa nomenclatura.

O processo de estimação dos modelos MFM é mais desafiador devido à aleatoriedade da dimensão do espaço paramétrico, que é introduzida ao associarmos uma distribuição *a priori* ao número de componentes. O algoritmo *Reversible Jump* ([RICHARDSON; GREEN, 1997](#)) é aplicável nestes modelos, porém sua utilização requer a especificação de funções de transição que podem impactar o tempo de convergência do algoritmo. Uma alternativa ao *Reversible Jump*, que não faz uso das funções de transição, são os modelos de mistura por processo de Dirichlet ([ANTONIAK, 1974](#); [ESCOBAR, 1994](#)), porém, por serem baseadas no processo de Dirichlet ([FERGUSON, 1973](#)) estão na classe de modelos de misturas infinitas.

Em seu trabalho, [Miller e Harrison \(2018\)](#) encontraram correspondências entre o MFM e o modelo de misturas por processo de Dirichlet, como representação em urnas de Pólya, *stick-breaking* e como medida aleatória. Para tal, consideram a partição \mathcal{C} induzida pelas variáveis de alocação $\mathbf{Z} = (Z_1, Z_2, \dots, Z_n)$, ou seja, $\mathcal{C} = \{\mathcal{C}_j : |\mathcal{C}_j| > 0\}$, com $\mathcal{C}_j = \{i \in \{1, 2, \dots, n\} : Z_i = j\}$ e com $|\mathcal{C}_j|$ denotando o número de elementos no conjunto. É importante notar que o número de partes na partição $G_+ = |\mathcal{C}|$ pode não coincidir com o número de componentes na mistura G . Para evidenciar essa distinção, chamaremos G_+ de número de grupos (cada parte na partição sendo um grupo) e G de número de componentes. Os autores mencionam que fixar γ constante independente de G é necessário para a obtenção de uma forma simples para a distribuição de \mathcal{C} porém pode ser considerado restritivo.

Uma extensão dos MFM foi introduzida por [Frühwirth-Schnatter, Malsiner-Walli e Grün \(2021\)](#) permitindo diferentes valores para o hiperparâmetro da distribuição *a priori* de \mathbf{p} , a extensão foi chamada de modelo MFM generalizado pelos autores. Seja $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)^\top$ vetor de observações independentes, o modelo é dado por:

$$\begin{aligned}
G &\sim \pi_G(\cdot); \\
\mathbf{p}|G &\sim \text{DirichSim}_G(\gamma_G); \\
\boldsymbol{\theta}_j|G &\stackrel{\text{ind}}{\sim} \pi(\cdot|\boldsymbol{\varphi}), j = 1, 2, \dots, G;^1 \\
Z_i|G, \mathbf{p} &\stackrel{\text{ind}}{\sim} \text{Cat}(\mathbf{p}), i = 1, 2, \dots, n; \\
Y_i|G, Z_i = j, \boldsymbol{\theta}_j &\stackrel{\text{ind}}{\sim} F(\cdot|\boldsymbol{\theta}_j), i = 1, 2, \dots, n.
\end{aligned} \tag{2.22}$$

A distribuição *a priori* de \mathbf{p} depende do número de componentes através de uma sequência de números positivos $\boldsymbol{\gamma} = \{\gamma_G, G \geq 1\}$. O modelo MFM é um caso particular, ocorrendo quando o hiperparâmetro da distribuição *a priori* Dirichlet simétrica permanece constante para todo G .

A função de verossimilhança e verossimilhança aumentada por Z são semelhantes as apresentadas em (2.19) e (2.20), respectivamente, bastando apenas a inclusão do novo parâmetro G . A função de verossimilhança é dada por

$$L(\mathbf{y}|\boldsymbol{\theta}, G) = \prod_{i=1}^n \sum_{j=1}^G p_j f(y_i|\boldsymbol{\theta}_j) \quad (2.23)$$

e a verossimilhança aumentada por Z é dada por

$$L(\mathbf{y}, \mathbf{z}|\boldsymbol{\theta}, G) = \prod_{j=1}^G p_j^{m_j} \prod_{i=1}^n f(y_i|\boldsymbol{\theta}_j)^{\mathbb{1}(z_i=j)}. \quad (2.24)$$

Note que a dimensão dos vetores de parâmetros depende do número de componentes, que é aleatório.

Como feito por [Miller e Harrison \(2018\)](#), os modelos MFM generalizados levam em consideração a partição \mathcal{C} induzida pelas variáveis de alocação $\mathbf{Z} = (Z_1, Z_2, \dots, Z_n)$, há distinção entre o número de componentes com pelo menos uma observação e o número de componentes. Definimos por G_+ o número de grupos na amostra, ou seja,

$$G_+ = |\mathcal{C}| = \sum_{j=1}^G \mathbb{1}(m_j > 0), \quad m_j = |\mathcal{C}_j| = \sum_{i=1}^n \mathbb{1}(z_i = j),$$

a priori, tanto G_+ quanto G são aleatórios.

Note que os grupos na amostra são invariantes a permutações nos rótulos dos vetores de alocação, o mesmo agrupamento é obtido caso permutemos os rótulos das componentes. [Frühwirth-Schnatter, Malsiner-Walli e Grün \(2021\)](#) obtêm a distribuição *a priori* da partição aleatória \mathcal{C} condicionada em G a partir da distribuição de $\mathbf{Z}|G$ considerando permutações de \mathbf{Z} que levam ao mesmo conjunto de grupos na amostra. A distribuição *a priori* condicional da partição é dada por

$$\pi(\mathcal{C}|G, \boldsymbol{\gamma}_G) = \frac{V_{n, G_+}^{G, \boldsymbol{\gamma}_G}}{\Gamma^{G_+}(\boldsymbol{\gamma}_G)} \prod_{j=1}^{G_+} \Gamma(m_j + \gamma_G), \quad V_{n, G_+}^{G, \boldsymbol{\gamma}_G} = \frac{G! \Gamma(G \boldsymbol{\gamma}_G)}{(G - G_+)! \Gamma(n + G \boldsymbol{\gamma}_G)}. \quad (2.25)$$

Também obtêm a distribuição *a priori* da partição aleatória \mathcal{C}

$$\pi(\mathcal{C}|\boldsymbol{\gamma}) = \sum_{G=G_+}^{\infty} \pi(G) \pi(\mathcal{C}|G, \boldsymbol{\gamma}_G), \quad (2.26)$$

¹ A versão do MFM generalizado apresentada por [Frühwirth-Schnatter, Malsiner-Walli e Grün \(2021\)](#) inclui uma distribuição *a priori* para o hiperparâmetro $\boldsymbol{\varphi}$;

com $\pi(G)$ própria. Observe que a distribuição *a priori* da partição aleatória depende da sequência $\boldsymbol{\gamma} = \{\gamma_G, G \geq 1\}$. Além disso, também obtém a distribuição *a priori* dos tamanhos de cada grupo, definida como

$$\pi(m_1, m_2, \dots, m_{G_+} | \boldsymbol{\gamma}) = \frac{n!}{G_+!} \sum_{G=G_+}^{\infty} \pi(G) \frac{V_{n, G_+}^{G, \boldsymbol{\gamma}_G}}{\Gamma^{G_+}(\boldsymbol{\gamma}_G)} \prod_{j=1}^{G_+} \frac{\Gamma(m_j + \gamma_G)}{\Gamma(m_j + 1)}. \quad (2.27)$$

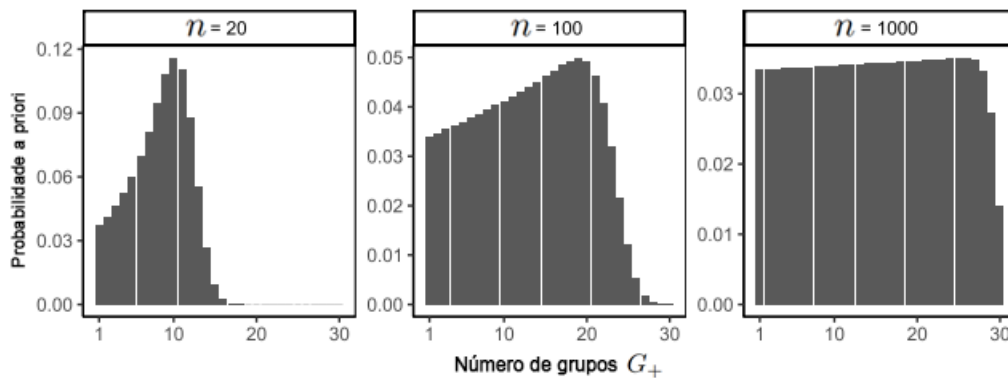
A partir desta distribuição *a priori*, obtém a distribuição *a priori* do número de grupos, somando (2.27) sob o conjunto formado pelos tamanhos de grupos válidos, ou seja, $\mathcal{V} = \{m_1, m_2, \dots, m_{G_+} > 0 : m_1 + m_2 + \dots + m_{G_+} = n\}$, dada por

$$\pi(G_+ | \boldsymbol{\gamma}) = \frac{n!}{G_+!} \sum_{G=G_+}^{\infty} \pi(G) \frac{V_{n, G_+}^{G, \boldsymbol{\gamma}_G}}{\Gamma^{G_+}(\boldsymbol{\gamma}_G)} C_{n, G_+}^{G, \boldsymbol{\gamma}_G}, \quad C_{n, G_+}^{G, \boldsymbol{\gamma}_G} = \sum_{\mathcal{V}} \prod_{j=1}^{G_+} \frac{\Gamma(m_j + \gamma_G)}{\Gamma(m_j + 1)}. \quad (2.28)$$

Os autores apresentam um algoritmo para o cálculo de $C_{n, G_+}^{G, \boldsymbol{\gamma}_G}$.

Um ponto de discussão importante é a distribuição *a priori* para o número de componentes, $\pi(G)$, visto que induz uma distribuição *a priori* para G_+ . Richardson e Green (1997) utilizam $G \sim \text{Uniforme}\{1, \dots, G_{\max}\}$ para um MFM com $\gamma = 1$, porém, como mostrado por Frühwirth-Schnatter, Malsiner-Walli e Grün (2021), neste caso, a distribuição *a priori* de G_+ pode ser informativa dependendo do tamanho amostral. Este comportamento pode ser observado na Figura 2, em que temos $\pi(G_+ | \boldsymbol{\gamma})$ para um MFM, ou seja, a sequência $\boldsymbol{\gamma}$ permanece constante com relação ao número de componentes, com valor γ e fixamos $n = 20, 100, 1000$ e $\gamma = 1$.

Figura 2 – Distribuição *a priori* para o número de grupos em um MFM, $\pi(G_+ | \gamma = 1)$, quando $G \sim \text{Uniforme}\{1, \dots, 30\}$.



Fonte: Adaptada de Frühwirth-Schnatter, Malsiner-Walli e Grün (2021).

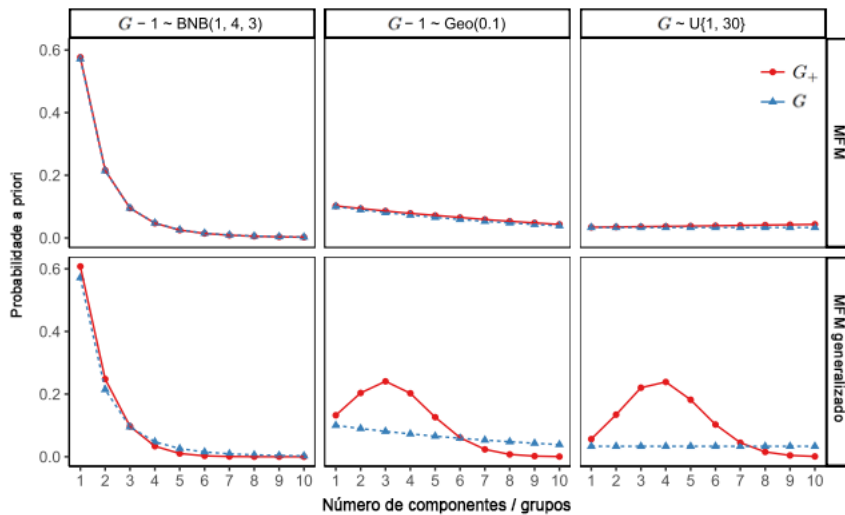
Por recomendação dos autores, utilizaremos uma distribuição *a priori* beta-binomial-negativa, $G - 1 \sim \text{BNB}(a_\lambda, a_\pi, b_\pi)$, uma generalização das distribuições de Poisson, geométrica e binomial-negativa. Dessa maneira, a função de probabilidade *a priori* de G é

$$\pi(G) = \frac{\Gamma(a_\lambda + G - 1)B(a_\lambda + a_\pi, G - 1 + b_\pi)}{\Gamma(a_\lambda)\Gamma(G)B(a_\pi, b_\pi)}, \quad (2.29)$$

em que $B(\cdot, \cdot)$ denota a função beta. Os autores sugerem fixar $a_\lambda = 1, a_\pi = 4$ e $b_\pi = 3$, desta forma obtendo uma distribuição *a priori* para G_+ com valores concentrados em torno de 4 componentes, porém ainda permitindo valores mais elevados, além disso seu comportamento permanece similar entre os MFM e MFM generalizados.

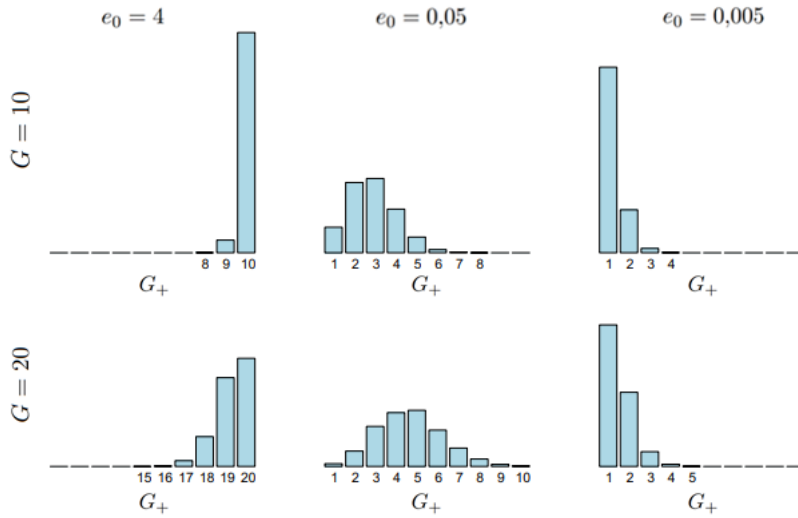
A Figura 3 mostra $\pi(G)$ e $\pi(G_+|\boldsymbol{\gamma})$ para três considerações *a priori* para G , a saber, $G - 1 \sim \text{BNB}(1, 4, 3)$, $G - 1 \sim \text{Geométrica}(0, 1)$ como em Miller e Harrison (2018) e $G \sim \text{Uniforme}\{1, \dots, 30\}$ como em Richardson e Green (1997), no modelo MFM e no MFM generalizado. Para o MFM, fixamos $\gamma = 1$, para o MFM generalizado, escolhemos $\boldsymbol{\gamma} = \{1/G, G \geq 1\}$, em ambos os casos, $n = 82$. Dentre as opções consideradas, utilizar $G - 1 \sim \text{BNB}(1, 4, 3)$ em um MFM generalizado implica em assumir distribuições *a priori* semelhantes para G e G_+ e, como mostrado em Frühwirth-Schnatter, Malsiner-Walli e Grün (2021, p. 1299-1300), utilizar as outras opções de distribuição *a priori* pode implicar em estimativas enviesadas para o número de grupos e componentes.

Figura 3 – Distribuições *a priori* de G e G_+ , considerando $G - 1 \sim \text{BNB}(1, 4, 3)$, $G - 1 \sim \text{Geométrica}(0, 1)$ e $G \sim \text{Uniforme}\{1, \dots, 30\}$, para um MFM e MFM generalizado, $n = 82$.



Fonte: Adaptada de Frühwirth-Schnatter, Malsiner-Walli e Grün (2021).

A escolha da sequência $\boldsymbol{\gamma} = \{1/G, G \geq 1\}$ foi motivada por Malsiner-Walli, Frühwirth-Schnatter e Grün (2016), que introduzem os modelos de Misturas Finitas Esparsas (SFM, do inglês *Sparse Finite Mixtures*). Nesta abordagem, G é fixado em um número que sobrestima G_+ e uma distribuição *a priori* Dirichlet simétrica de parâmetro e_0 é atribuída aos pesos das componentes. Os autores comentam que um valor pequeno para e_0 é necessário para que o comportamento desejado de componentes em “excesso” serem esvaziados ocorra. Se esse comportamento ocorre, é possível então tratar G_+ como uma variável aleatória com valores menores que G , porém, como mostra a Figura 4, e_0 deve ser inversamente proporcional a G para que o comportamento desejado ocorra, o que motiva a escolha da sequência $\boldsymbol{\gamma} = \{\gamma/G, G \geq 1\}$.

Figura 4 – Distribuição *a priori* dos grupos em um SFM.

Fonte: Adaptada de Frühwirth-Schnatter e Malsiner-Walli (2019).

Neste trabalho adotaremos a sequência $\boldsymbol{\gamma} = \{\gamma/G, G \geq 1\}$. Frühwirth-Schnatter, Malsiner-Walli e Grün (2021) mostram que MFM generalizado utilizando esta sequência converge para um modelo de mistura por processo de Dirichlet com parâmetro de concentração $\gamma > 0$ caso $\pi(G)$ concentre sua massa em $+\infty$. Os autores sugerem incluir uma distribuição F de Snedecor com 6 graus de liberdade no numerador e 3 no denominador ao hiperparâmetro γ , de forma a comportar dados com muitas componentes vazias (γ próximo de zero) e dados com poucas componentes vazias (γ distante de zero). Na Seção 2.4, apresentamos o *Telescoping Sampler*, algoritmo inferencial proposto pelos autores.

2.2.1 Modelo de misturas de regressões *t* de Student assimétricas

O foco deste trabalho são os modelos de mistura de regressão, obtidos ao assumir uma estrutura particular a $F(\cdot|\boldsymbol{\theta}_j)$ apresentada nos modelos de mistura discutidos anteriormente. Nosso interesse principal são os modelos de Misturas de Regressões ST (MRST).

O modelo MRST, para um dado número de componentes G , pode ser definido da seguinte maneira: considere um vetor de observações independentes $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)^\top$, em que, condicionada a observação de uma variável de alocação latente Z_i , a resposta do i -ésimo indivíduo Y_i , depende do vetor de covariáveis \mathbf{x}_i conforme

$$Y_i = \mathbf{x}_i^\top \boldsymbol{\beta}_j + \varepsilon_{ij}, \quad (2.30)$$

em que $\varepsilon_{ij}|Z_i = j \sim \text{ST}(b\Delta_j, \sigma_j^2, \lambda_j, \nu)$ com

$$b = -\sqrt{\frac{\nu \Gamma(\frac{\nu-1}{2})}{\pi \Gamma(\nu/2)}}, \Delta_j = \sigma_j \delta_j, \text{ e } \delta_j = \frac{\lambda_j}{\sqrt{1 + \lambda_j^2}}.$$

O valor do parâmetro de localização da distribuição ST foi escolhido com base em (2.16) de forma que o valor esperado dos erros de observação na j -ésima componente ε_{ij} seja zero para todo i , assim temos $\mathbb{E}(Y_i|Z_i = j) = \mathbf{x}_i^\top \boldsymbol{\beta}_j$. É importante notar que $\mathbb{E}(Y_i|Z_i = j)$ não existe caso $0 \leq \nu \leq 1$.

Note que mantemos ν igual entre as componentes, isto foi feito de forma a reduzir a complexidade computacional do problema, o mesmo procedimento foi realizado em Cabral, Souza e Leão (2022). O vetor $\boldsymbol{\beta}_j = (\beta_{0j}, \beta_{1j}, \dots, \beta_{kj})^\top$ contém os coeficientes de regressão da j -ésima componente. O vetor $\mathbf{x}_i^\top = (1, x_{i1}, x_{i2}, \dots, x_{ik})$ contém os valores de k variáveis regressoras. Denotamos por $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)^\top$ a matriz de planejamento.

Por (2.14) vemos que (2.30) implica que

$$\begin{aligned} Z_i | \mathbf{p} &\stackrel{\text{ind}}{\sim} \text{Cat}(\mathbf{p}) \\ Y_i | Z_i = j, \boldsymbol{\theta}'_j, \mathbf{x}_i &\stackrel{\text{ind}}{\sim} \text{ST}(\mathbf{x}_i^\top \boldsymbol{\beta}_j + b\Delta_j, \sigma_j^2, \lambda_j, \nu), \end{aligned} \quad (2.31)$$

em que $\boldsymbol{\theta}'_j = (\boldsymbol{\beta}_j^\top, \sigma_j^2, \lambda_j, \nu)^\top$. Por (2.13), vemos que (2.31) pode ser escrito como:

$$\begin{aligned} Z_i | \mathbf{p} &\stackrel{\text{ind}}{\sim} \text{Cat}(\mathbf{p}) \\ U_i | \nu &\sim \text{Gama}(\nu/2, \nu/2) \\ T_i | U_i = u_i &\stackrel{\text{ind}}{\sim} N_{(0, \infty)}(0, u_i^{-1}) \\ Y_i | Z_i = j, T_i = t_i, U_i = u_i, \boldsymbol{\theta}'_j, \mathbf{x}_i &\stackrel{\text{ind}}{\sim} N(\mathbf{x}_i^\top \boldsymbol{\beta}_j + (b + t_i)\Delta_j, u_i^{-1} \tau_j^2), \end{aligned} \quad (2.32)$$

em que $N_{(0, \infty)}(a, b^2)$ denota a distribuição normal truncada no intervalo $(0, +\infty)$, com média antes do truncamento a e variância antes do truncamento b^2 . Reparametrizaremos o modelo apresentado em (2.31) com relação as quantidades Δ_j e τ_j^2 que aparecem na representação hierárquica do modelo, dada em (2.32). A transformação e transformação inversa são dadas, respectivamente, por

$$\begin{cases} \Delta_j = \sigma_j \delta_j \\ \tau_j^2 = \sigma_j^2 (1 - \delta_j^2) \end{cases} \quad \text{e} \quad \begin{cases} \lambda_j = \Delta_j / \tau_j \\ \sigma_j^2 = \tau_j^2 + \Delta_j^2, \end{cases} \quad (2.33)$$

com $\delta_j = \lambda_j / \sqrt{1 + \lambda_j^2}$, esta reparametrização também foi utilizada por Nascimento (2017) e Cabral, Souza e Leão (2022). A reparametrização do modelo apresentado em (2.31) é

$$Y_i | Z_i = j, \boldsymbol{\theta}_j, \mathbf{x}_i \sim \text{ST}(\mathbf{x}_i^\top \boldsymbol{\beta}_j + b\Delta_j, \tau_j^2 + \Delta_j^2, \Delta_j / \tau_j, \nu), \quad (2.34)$$

em que $\boldsymbol{\theta}_j = (\boldsymbol{\beta}_j^\top, \tau_j^2, \Delta_j)$ são os parâmetros específicos de cada componente, observe que $\boldsymbol{\beta}_j \in \mathbb{R}^{k+1}$, $\tau_j^2 \in \mathbb{R}_+$ e $\Delta_j \in \mathbb{R}$. Obtemos as seguintes funções de verossimilhança e verossimilhança aumentada, respectivamente,

$$L(\mathbf{y} | G, \boldsymbol{\theta}, \mathbf{X}) = \prod_{i=1}^n \sum_{j=1}^G p_j \text{St}(y_i | \mathbf{x}_i^\top \boldsymbol{\beta}_j + b\Delta_j, \tau_j^2 + \Delta_j^2, \Delta_j / \tau_j, \nu), \quad (2.35)$$

$$L(\mathbf{y}, \mathbf{z} | G, \boldsymbol{\theta}, \mathbf{X}) = \prod_{j=1}^G p_j^{m_j} \prod_{i=1}^n \text{St}(y_i | \mathbf{x}_i^\top \boldsymbol{\beta}_j + b\Delta_j, \tau_j^2 + \Delta_j^2, \Delta_j / \tau_j, \nu)^{\mathbb{1}(z_i=j)}, \quad (2.36)$$

em que $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\tau}^2, \boldsymbol{\Delta}, \nu, \boldsymbol{p}) \in \mathbb{R}^{G(k+1)} \times \mathbb{R}_+^G \times \mathbb{R}^G \times \mathbb{R}_+ \times \Xi_G$, com $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \dots, \boldsymbol{\beta}_G^\top)$, $\boldsymbol{\tau}^2 = (\tau_1^2, \dots, \tau_G^2)$, $\boldsymbol{\Delta} = (\Delta_1, \dots, \Delta_G)$ e com $\Xi_G = \{p_j \in [0, 1] \forall j \geq 1 \mid \sum_{j=1}^G p_j = 1\}$. A reparametrização (2.33) é útil por facilitar a obtenção de distribuições condicionais completas das quais a simulação de observações é simples para a maioria dos parâmetros do modelo, como mostraremos posteriormente. Levaremos em consideração os parâmetros reparametrizados $\boldsymbol{\theta}$, a estimação dos parâmetros originais se dará utilizando a transformação inversa mostrada em (2.33).

Para $\boldsymbol{\theta}$, consideramos $\boldsymbol{\beta}, \boldsymbol{\tau}^2, \boldsymbol{\Delta}$ e ν independentes *a priori* e também que os parâmetros específicos de cada componente são independentes *a priori*, ou seja, $\boldsymbol{\theta}_j \perp \boldsymbol{\theta}_k, j \neq k$. As distribuições *a priori* são

$$\begin{aligned} \boldsymbol{\beta}_j &\sim N_{k+1}(\mathbf{0}, c^2 \mathbb{I}_{k+1}); \\ \Delta_j &\sim N(\eta, \omega^2); \\ \tau_j^2 &\sim \text{InvGama}(r, s); \\ \nu | \boldsymbol{\alpha} &\sim \text{ExpLoc}(\boldsymbol{\psi}, \boldsymbol{\alpha}), \boldsymbol{\alpha} \sim \text{Unif}(0.02, 0.5), \end{aligned} \tag{2.37}$$

em que $N_k(\boldsymbol{a}, \boldsymbol{B})$ denota a distribuição normal k variada, com vetor de médias \boldsymbol{a} e matriz de covariância \boldsymbol{B} , \mathbb{I}_k denota a matriz identidade de ordem k . $\text{ExpLoc}(\boldsymbol{\psi}, \boldsymbol{\alpha})$ denota a distribuição exponencial de localização, se $X \sim \exp(\boldsymbol{\alpha})$, $Y = X + \boldsymbol{\psi} \sim \text{ExpLoc}(\boldsymbol{\psi}, \boldsymbol{\alpha})$. A distribuição *a priori* de ν foi motivada por [Garay et al. \(2015\)](#) em que foram estudadas diversas opções de distribuições *a priori* para ν sendo observado que considerar $\text{ExpLoc}(0, \boldsymbol{\alpha}), \boldsymbol{\alpha} \sim \text{Unif}(0.02, 0.5)$ resulta na melhor recuperação dos valores dos parâmetros, a inclusão do parâmetro de localização se deve ao fato da não existência do valor esperado da distribuição ST caso $0 \leq \nu \leq 1$. As demais especificações *a priori* são similares as utilizadas por [Nascimento \(2017\)](#) e [Cabral, Souza e Leão \(2022\)](#) e foram escolhidas de forma a facilitar a obtenção de distribuições condicionais completas. Assim temos,

$$\pi(\boldsymbol{\theta}_j | G) = \pi(\boldsymbol{\beta}_j | G) \pi(\Delta_j | G) \pi(\tau_j^2 | G) \tag{2.38}$$

$$\pi(\boldsymbol{\theta} | G) = \left(\prod_{j=1}^G \pi(\boldsymbol{\theta}_j | G) \right) \pi(\nu) \pi(\boldsymbol{p} | G). \tag{2.39}$$

Utilizando as considerações *a priori* dadas em (2.37) podemos encontrar as seguintes

condicionais completas, o Apêndice A mostra detalhes sobre sua obtenção.

$$\boldsymbol{\beta}_j | \boldsymbol{\beta}_{-j}, \mathbf{y}, \mathbf{z}, \mathbf{t}, \mathbf{u}, \dots \sim N_{k+1}(\boldsymbol{\Sigma}_j^{-1} \mathbf{w}_j, \tau_j^2 \boldsymbol{\Sigma}_j^{-1}) \quad (2.40)$$

$$\tau_j^2 | \boldsymbol{\tau}_{-j}, \mathbf{y}, \mathbf{z}, \mathbf{t}, \mathbf{u}, \dots \sim \text{InvGama} \left(\frac{m_j}{2} + r, \frac{S_{3j}}{2} + s \right) \quad (2.41)$$

$$\Delta_j | \boldsymbol{\Delta}_{-j}, \mathbf{y}, \mathbf{z}, \mathbf{t}, \mathbf{u}, \dots \sim N \left(\frac{\omega^2 S_{1j} + \eta \tau_j^2}{\omega^2 S_{2j} + \tau_j^2}, \frac{\omega^2 \tau_j^2}{\omega^2 S_{2j} + \tau_j^2} \right) \quad (2.42)$$

$$\alpha | \mathbf{y}, \dots \sim \text{Gama}_{(0.02, 0.5)}(2, \nu - \psi) \quad (2.43)$$

$$Z_i | \mathbf{y}, \mathbf{z}_{-i}, \dots \sim \text{Cat}(\boldsymbol{\rho}) \quad (2.44)$$

$$U_i | \mathbf{u}_{-i}, \mathbf{y}, \mathbf{z}, \mathbf{t}, \dots \sim \text{Gama} \left(\frac{\nu}{2} + 1, \frac{(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_j - (b + t_j) \Delta_j)^2}{2 \tau_j^2} + \frac{t_i^2 + \nu}{2} \right), \quad (2.45)$$

$$T_i | \mathbf{t}_{-i}, \mathbf{y}, \mathbf{z}, \mathbf{u}, \dots \sim N_{(0, \infty)} \left(\frac{(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_j - b \Delta_j) \Delta_j}{\Delta_j^2 + \tau_j^2}, \frac{\tau_j^2}{u_i (\Delta_j^2 + \tau_j^2)} \right) \quad (2.46)$$

em que ... denota os demais parâmetros do modelo, $\boldsymbol{\tau}_{-j}^2$ denota o vetor $\boldsymbol{\tau}^2$ excluindo o j -ésimo elemento, o mesmo vale para $\boldsymbol{\beta}_{-j}$, $\boldsymbol{\Delta}_{-j}$, \mathbf{z}_{-i} , \mathbf{t}_{-i} e \mathbf{u}_{-i} , $\mathbf{w}_j = \mathbf{X}_j^\top \mathbf{U}_j (\mathbf{y}_j - (b + t_j) \Delta_j)$ e $\boldsymbol{\Sigma}_j = \mathbf{X}_j^\top \mathbf{U}_j \mathbf{X}_j + \tau_j^2 c^{-2} \mathbb{I}_{k+1}$, com $\mathbf{U}_j = \text{diag}(\mathbf{u}_j)$, \mathbf{X}_j é a matriz composta das i -ésimas linhas da matriz de planejamento \mathbf{X} tais que $z_i = j$, \mathbf{y}_j , \mathbf{t}_j , \mathbf{u}_j denotando o vetor composto pelos i -ésimos elementos dos respectivos vetores \mathbf{y} , \mathbf{t} , \mathbf{u} tais que $z_i = j$, $S_{1j} = (b + t_j)^\top \mathbf{U}_j (\mathbf{y}_j - \mathbf{X}_j \boldsymbol{\beta}_j)$, $S_{2j} = (b + t_j)^\top \mathbf{U}_j (b + t_j)$ e $S_{3j} = (\mathbf{y}_j - \mathbf{X}_j \boldsymbol{\beta}_j - (b + t_j) \Delta_j)^\top \mathbf{U}_j (\mathbf{y}_j - \mathbf{X}_j \boldsymbol{\beta}_j - (b + t_j) \Delta_j)$ e por fim, $\boldsymbol{\rho} = (\rho_1, \rho_2, \dots, \rho_G)^\top$ com $\rho_j = p_j \text{St}(y_i | \mathbf{x}_i^\top \boldsymbol{\beta}_j + b \Delta_j, \sigma_j^2, \lambda_j, \nu) / \sum_{l=1}^G p_l \text{St}(y_i | \mathbf{x}_i^\top \boldsymbol{\beta}_l + b \Delta_l, \sigma_l^2, \lambda_l, \nu)$, para $j = 1, 2, \dots, G$. A distribuição condicional completa para o parâmetro ν possui função densidade tal que

$$\pi(\nu | \mathbf{y}, \dots) \propto e^{-\alpha(\nu - \psi)} \mathbb{1}(\nu > \psi) \prod_{i=1}^n \sum_{j=1}^G p_j \text{St}(y_i | \mathbf{x}_i^\top \boldsymbol{\beta}_j + b \Delta_j, \sigma_j^2, \lambda_j, \nu), \quad (2.47)$$

que não é proporcional a função densidade de uma distribuição simples de obter amostras.

É interessante notar que o vetor de médias da distribuição condicional completa do parâmetro $\boldsymbol{\beta}_j$ coincide com o estimador da regressão por mínimos quadrados ponderados por \mathbf{u}_j se consideramos como vetor de respostas $(\mathbf{y}_j - (b + t_j) \Delta_j)$ e fazendo $c \rightarrow +\infty$. Sob as mesmas condições, a matriz de covariâncias da distribuição condicional completa do parâmetro $\boldsymbol{\beta}_j$ coincide com a matriz de covariância do estimador da regressão por mínimos quadrados ponderados.

2.3 Problemas de identificabilidade

Modelos de mistura em geral apresentam problemas de identificabilidade que podem impactar o processo de estimação dos parâmetros. Segundo [Frühwirth-Schnatter \(2006, p. 15\)](#), uma família de distribuições paramétrica, definida sobre o espaço amostral \mathcal{Y} e indexada pelo

parâmetro $\boldsymbol{\theta} \in \Theta$, é dita identificável se dois parâmetros $\boldsymbol{\theta} \in \Theta$ e $\boldsymbol{\theta}^* \in \Theta$ definem a mesma lei de probabilidade em \mathcal{Y} se, e somente se, $\boldsymbol{\theta} = \boldsymbol{\theta}^*$. Colocando em termos das respectivas funções densidade $f(\mathbf{y}|\boldsymbol{\theta})$ e $f(\mathbf{y}|\boldsymbol{\theta}^*)$ temos que se

$$f(\mathbf{y}|\boldsymbol{\theta}) = f(\mathbf{y}|\boldsymbol{\theta}^*) \text{ para quase todo } \mathbf{y} \in \mathcal{Y} \Rightarrow \boldsymbol{\theta} = \boldsymbol{\theta}^*$$

a família de distribuições é dita ser identificável. Uma família de distribuições paramétrica é dita ser não identificável se para $\boldsymbol{\theta}$ e $\boldsymbol{\theta}^*$ distintos, as funções densidade $f(\mathbf{y}|\boldsymbol{\theta})$ e $f(\mathbf{y}|\boldsymbol{\theta}^*)$ são idênticas para quase todo $\mathbf{y} \in \mathcal{Y}$.

Uma das causas da não identificabilidade das misturas finitas de distribuições são sua invariância à permutação dos rótulos das componentes. Por exemplo, considere uma mistura de duas componentes ST: por (2.17) a função densidade é dada por $f(\mathbf{y}|\boldsymbol{\theta}) = p_1 \text{St}(\mathbf{y}|\mu_1, \sigma_1^2, \lambda_1, \nu_1) + p_2 \text{St}(\mathbf{y}|\mu_2, \sigma_2^2, \lambda_2, \nu_2)$, com $\boldsymbol{\theta}_j = (\mu_j, \sigma_j^2, \lambda_j, \nu_j) \in \Theta = \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R} \times \mathbb{R}_+$ e $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, p_1, p_2) \in \Theta_2 = \Theta^2 \times \Xi_2$, em que $\Xi_2 = \{p_j \in [0, 1] \forall j \geq 1 | p_1 + p_2 = 1\}$. Seja $\boldsymbol{\theta} \in \Theta_2$ um parâmetro arbitrário com $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$, definimos $\boldsymbol{\theta}^* = (\boldsymbol{\theta}_2, \boldsymbol{\theta}_1, p_2, p_1)$ trocando os rótulos das componentes no parâmetro $\boldsymbol{\theta}$, observe que $\boldsymbol{\theta} \neq \boldsymbol{\theta}^*$, porém,

$$\begin{aligned} f(\mathbf{y}|\boldsymbol{\theta}) &= p_1 \text{St}(\mathbf{y}|\mu_1, \sigma_1^2, \lambda_1, \nu_1) + p_2 \text{St}(\mathbf{y}|\mu_2, \sigma_2^2, \lambda_2, \nu_2) \\ &= p_2 \text{St}(\mathbf{y}|\mu_2, \sigma_2^2, \lambda_2, \nu_2) + p_1 \text{St}(\mathbf{y}|\mu_1, \sigma_1^2, \lambda_1, \nu_1) = f(\mathbf{y}|\boldsymbol{\theta}^*), \end{aligned}$$

ou seja, uma mistura de duas componentes ST não é identificável.

Para o caso de uma mistura de G componentes ST, a função densidade é invariante a permutações nos rótulos das componentes. Suponha que κ_G é uma permutação arbitrária dos rótulos das componentes, ou seja, $\kappa_g : \{1, 2, \dots, G\} \rightarrow \{1, 2, \dots, G\}$, $g = 1, 2, \dots, G!$. Se tomamos um parâmetro arbitrário $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_G, p_1, p_2, \dots, p_G) \in \Theta_G = \Theta^G \times \Xi_G$, podemos definir $\boldsymbol{\theta}^* = (\boldsymbol{\theta}_{\kappa_g(1)}, \boldsymbol{\theta}_{\kappa_g(2)}, \dots, \boldsymbol{\theta}_{\kappa_g(G)}, p_{\kappa_g(1)}, p_{\kappa_g(2)}, \dots, p_{\kappa_g(G)}) \in \Theta_G = \Theta^G \times \Xi_G$ de forma que

$$\begin{aligned} f(\mathbf{y}|\boldsymbol{\theta}) &= p_1 \text{St}(\mathbf{y}|\boldsymbol{\theta}_1) + \dots + p_G \text{St}(\mathbf{y}|\boldsymbol{\theta}_G) \\ &= p_{\kappa_g(1)} \text{St}(\mathbf{y}|\boldsymbol{\theta}_{\kappa_g(1)}) + \dots + p_{\kappa_g(G)} \text{St}(\mathbf{y}|\boldsymbol{\theta}_{\kappa_g(G)}) = f(\mathbf{y}|\boldsymbol{\theta}^*). \end{aligned} \tag{2.48}$$

O problema de identificabilidade causado pela invariância à permutações nos rótulos das componentes é comumente chamado de *label switching*. Como mencionado por Frühwirth-Schnatter (2006, p. 16), o *label switching* não é um problema de identificabilidade severo por conta dos valores dos parâmetros estarem relacionados entre si por meio de permutações, sendo assim possível solucionar o problema durante o processo de estimação utilizando permutações adequadas nos rótulos das componentes, como mostraremos na Seção 2.4.

Outra causa de problemas de identificabilidade em misturas de distribuições é a inclusão de componentes redundantes. Por exemplo, uma mistura de duas componentes ST, com parâmetro $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, p_1, p_2) \in \Theta_2$, pode ser escrita como uma mistura de três componentes ST com $p_3 = 0$,

$$\begin{aligned} f(\mathbf{y}|\boldsymbol{\theta}) &= p_1 \text{St}(\mathbf{y}|\boldsymbol{\theta}_1) + p_2 \text{St}(\mathbf{y}|\boldsymbol{\theta}_2) \\ &= p_1 \text{St}(\mathbf{y}|\boldsymbol{\theta}_1) + p_2 \text{St}(\mathbf{y}|\boldsymbol{\theta}_2) + 0 \cdot \text{St}(\mathbf{y}|\boldsymbol{\theta}_3). \end{aligned}$$

A mistura de três componentes, com parâmetro $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3, p_1, p_2, 0) \in \Theta_3$, não é identificável, pois qualquer $\boldsymbol{\theta}_3$ leva ao mesmo valor da função densidade. Contudo, note que, se a mistura de duas componentes apresenta apenas problema de *label switching*, como $\Theta_2 \subset \Theta_3$ é possível ainda solucionar o problema de identificabilidade durante o processo de estimação desconsiderando a componente com peso nulo, motivações práticas para tal serão apresentadas na Seção 2.4.

Ainda podemos incluir componentes redundantes ao assumir que duas ou mais componentes possam ser representadas por apenas uma componente. Por exemplo, se em uma mistura de três componentes ST temos

$$f(y|\boldsymbol{\theta}) = p_1 \text{St}(y|\boldsymbol{\theta}_1) + (p_2 - p_3) \text{St}(y|\boldsymbol{\theta}_2) + p_3 \text{St}(y|\boldsymbol{\theta}_2),$$

vemos que a mistura de três componentes ST com parâmetro $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_2, p_1, p_2 - p_3, p_3)$ não é identificável, pois para qualquer $p_3 \in [0, p_2]$ obtemos o mesmo valor da função densidade. Como mencionado na Seção 2.2, o modelo MFM generalizado que consideraremos neste trabalho dá baixa probabilidade *a priori* para o evento de duas ou mais componentes poderem ser representadas por apenas uma em favor de componentes com peso nulo no caso de presença de componentes redundantes.

Os modelos de mistura de regressões também apresentam problemas de identificabilidade por *label switching* e por inclusão de componentes redundantes, porém esses modelos adicionam outra possível fonte de problemas de identificabilidade. Segundo Hennig (2000), um modelo de misturas de regressões não é identificável se o número G de componentes na mistura for igual ou superior ao número de hiperplanos k -dimensionais distintos gerados pelas covariáveis.

Por exemplo, considere um modelo MRST com duas componentes com $n = 2$ e $k = 1$, a função de verossimilhança deste modelo, como mostrado em (2.35), é dada por

$$L(\mathbf{y}|\boldsymbol{\theta}, \mathbf{X}) = \prod_{i=1}^2 \sum_{j=1}^2 p_j \text{St}(y_i | \mathbf{x}_i^\top \boldsymbol{\beta}_j + b\Delta_j, \tau_j^2 + \Delta_j^2, \Delta_j/\tau_j, \mathbf{v}),$$

considere $\mathbf{x}_1^\top = (1, 0)$, $\mathbf{x}_2^\top = (1, 1)$ e os seguintes parâmetros $\boldsymbol{\theta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top, (\tau^2, \tau^2), (\Delta, \Delta), \mathbf{v}, (1/2, 1/2))$ e $\boldsymbol{\theta}^* = (\boldsymbol{\beta}_{1*}^\top, \boldsymbol{\beta}_{2*}^\top, (\tau^2, \tau^2), (\Delta, \Delta), \mathbf{v}, (1/2, 1/2))$. Se $\boldsymbol{\beta}_1^\top = (1, 0)$, $\boldsymbol{\beta}_2^\top = (2, 0)$, $\boldsymbol{\beta}_{1*}^\top =$

$(1, 1)$, $\boldsymbol{\beta}_{2*}^\top = (2, -1)$, com $\Delta, \tau^2 > 0, \nu > 0$ arbitrários, observe que

$$\begin{aligned} L(\mathbf{y}|\boldsymbol{\theta}, \mathbf{X}) &= \left(\frac{1}{2} \text{St}(y_1|\mathbf{x}_1^\top \boldsymbol{\beta}_1 + b\Delta, \tau^2 + \Delta^2, \Delta/\tau, \nu) + \frac{1}{2} \text{St}(y_1|\mathbf{x}_1^\top \boldsymbol{\beta}_2 + b\Delta, \tau^2 + \Delta^2, \Delta/\tau, \nu) \right) \\ &\quad \cdot \left(\frac{1}{2} \text{St}(y_2|\mathbf{x}_2^\top \boldsymbol{\beta}_1 + b\Delta, \tau^2 + \Delta^2, \Delta/\tau, \nu) + \frac{1}{2} \text{St}(y_2|\mathbf{x}_2^\top \boldsymbol{\beta}_2 + b\Delta, \tau^2 + \Delta^2, \Delta/\tau, \nu) \right) \\ &= \left(\frac{1}{2} \text{St}(y_1|1 + b\Delta, \tau^2 + \Delta^2, \Delta/\tau, \nu) + \frac{1}{2} \text{St}(y_1|2 + b\Delta, \tau^2 + \Delta^2, \Delta/\tau, \nu) \right) \\ &\quad \cdot \left(\frac{1}{2} \text{St}(y_2|1 + b\Delta, \tau^2 + \Delta^2, \Delta/\tau, \nu) + \frac{1}{2} \text{St}(y_2|2 + b\Delta, \tau^2 + \Delta^2, \Delta/\tau, \nu) \right) \\ &= \left(\frac{1}{2} \text{St}(y_1|\mathbf{x}_1^\top \boldsymbol{\beta}_{1*} + b\Delta, \tau^2 + \Delta^2, \Delta/\tau, \nu) + \frac{1}{2} \text{St}(y_1|\mathbf{x}_1^\top \boldsymbol{\beta}_{2*} + b\Delta, \tau^2 + \Delta^2, \Delta/\tau, \nu) \right) \\ &\quad \cdot \left(\frac{1}{2} \text{St}(y_2|\mathbf{x}_2^\top \boldsymbol{\beta}_{1*} + b\Delta, \tau^2 + \Delta^2, \Delta/\tau, \nu) + \frac{1}{2} \text{St}(y_2|\mathbf{x}_2^\top \boldsymbol{\beta}_{2*} + b\Delta, \tau^2 + \Delta^2, \Delta/\tau, \nu) \right) = L(\mathbf{y}|\boldsymbol{\theta}^*, \mathbf{X}). \end{aligned}$$

Note que no exemplo temos apenas dois hiperplanos unidimensionais (pontos) distintos gerados pelas covariáveis, 0 e 1, o que coincide com o número de componentes no modelo. Este problema de identificabilidade não seria observado caso incluíssemos uma nova observação com \mathbf{x}_3 distinto de \mathbf{x}_1 e \mathbf{x}_2 , restando apenas a possível presença de *label switching* e inclusão de componentes redundantes. Relembre que as covariáveis correspondem ao segundo elemento em diante dos vetores \mathbf{x}_i .

Para mais detalhes sobre problemas de identificabilidade em modelos de mistura e modelos de mistura de regressões veja [Frühwirth-Schnatter \(2006\)](#). Em seguida discutimos métodos de Monte Carlo em Cadeias de Markov (MCMC) frequentemente utilizados em inferência Bayesiana para a aproximação de integrais complexas.

2.4 Técnicas MCMC e procedimento de estimação

Nesta seção apresentamos algumas técnicas MCMC. As técnicas MCMC são úteis para a aproximação de integrais complexas quando as técnicas usuais de Monte Carlo não podem ser facilmente aplicadas devido ao passo de amostragem ser complexo. Em geral, temos uma densidade objetivo $\pi(\cdot)$ da qual a simulação de valores não é trivial, obtemos uma realização de uma Cadeia de Markov cuja distribuição estacionária é $\pi(\cdot)$, descartamos uma certa quantidade dessas realizações e então tratamos os valores restantes como amostra independente e identicamente distribuída de $\pi(\cdot)$ e aproximamos as quantidades de interesse. Para uma revisão sobre técnicas MCMC veja [Brooks et al. \(2011\)](#).

2.4.1 Algoritmo de Metropolis-Hastings

O algoritmo de Metropolis-Hastings (MH), proposto por [Metropolis et al. \(1953\)](#) e então estendido por [Hastings \(1970\)](#), pode ser utilizado para obtenção de amostras com densidade $\pi(\cdot)$ cuja simulação de valores é complexa. O algoritmo consiste em gerar observações propostas \mathbf{x}^*

com densidade $q(\cdot|\cdot)$ a partir da qual a simulação é simples, o valor gerado \mathbf{x}^* é então aceito como na distribuição de interesse com alguma probabilidade. No Algoritmo 1, apresentamos com detalhes o algoritmo.

Algoritmo 1 – Algoritmo de Metropolis-Hastings.

Entrada: Número de iterações T , densidade objetivo $\pi(\cdot)$, densidade proposta $q(\cdot|\cdot)$, valor inicial $\mathbf{x}^{(0)}$.

Saída: Cadeia de Markov $\{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(T)}\}$ com distribuição estacionária $\pi(\cdot)$.

- 1: **para** $t = 1$ até T **faça**
- 2: Gere uma proposta \mathbf{x}^* com densidade $q(\cdot|\mathbf{x}^{(t-1)})$
- 3: Calcule a probabilidade de aceitação

$$\alpha = \min \left\{ 1, \frac{\pi(\mathbf{x}^*)q(\mathbf{x}^{(t-1)}|\mathbf{x}^*)}{\pi(\mathbf{x}^{(t-1)})q(\mathbf{x}^*|\mathbf{x}^{(t-1)})} \right\}$$

- 4: Gere u de uma Uniforme(0, 1)
 - 5: **se** $u < \alpha$ **então**
 - 6: Aceite a proposta fazendo $\mathbf{x}^{(t)} \leftarrow \mathbf{x}^*$
 - 7: **senão**
 - 8: Rejeite a proposta fazendo $\mathbf{x}^{(t)} \leftarrow \mathbf{x}^{(t-1)}$
 - 9: **fim se**
 - 10: **fim para**
-

O Algoritmo 1 será utilizado para simular valores a partir da condicional completa (2.47) posteriormente.

2.4.2 Amostrador de Gibbs

O Amostrador de Gibbs (GS, do inglês *Gibbs sampler*), proposto por Geman e Geman (1984), pode ser utilizado para obtenção de amostras com densidade $\pi(\cdot)$ cuja simulação de valores é complexa. Diferentemente do algoritmo MH, a densidade objetivo deve ser multivariada e não há passo de rejeição. Os valores seguintes da cadeia são simulados das distribuições condicionais completas. Suponha que $\mathbf{X} = (X_1, \dots, X_D)^\top \sim \pi(\cdot)$, e que $\pi(\cdot)$ é uma distribuição a partir da qual a simulação de amostras é complicada, porém, considere que é simples simular das condicionais completas, ou seja, $X_d|X_1, \dots, X_{d-1}, X_{d+1}, \dots, X_D$, $d = 1, 2, \dots, D$. O algoritmo consiste da simulação de valores a partir das distribuições condicionais completas, no Algoritmo 2 descrevemos os passos em detalhes. Cada passo do GS é um caso particular do algoritmo MH em que a distribuição proposta em cada passo é a condicional completa de cada variável e, por conta disto, a taxa de aceitação é sempre 1. O Algoritmo 2 será utilizado no Capítulo 3 para simular valores da distribuição *a posteriori* do modelo MRST utilizando as distribuições condicionais completas apresentadas nas expressões (2.40) a (2.47).

Algoritmo 2 – Amostrador de Gibbs

Entrada: Número de iterações T , densidade objetivo $\pi(\cdot)$, condicionais completas, valores iniciais $x_1^{(0)}, x_2^{(0)}, \dots, x_D^{(0)}$

Saída: Cadeia de Markov $\{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(T)}\}$ com distribuição estacionária $\pi(\cdot)$.

- 1: **para** $t = 1$ até T **faça**
- 2: **para** $d = 1$ até D **faça**
- 3: Amostre $x_d^{(t)}$ com densidade $\pi(\cdot | x_1^{(t)}, \dots, x_{d-1}^{(t)}, x_{d+1}^{(t-1)}, \dots, x_D^{(t-1)})$
- 4: **fim para**
- 5: **fim para**

2.4.3 Telescoping Sampler

O *Telescoping Sampler* (TS), é um algoritmo inferencial proposto por Frühwirth-Schnatter, Malsiner-Walli e Grün (2021) para inferência Bayesiana em modelos MFM generalizados. Como o GS, o TS utiliza as distribuições condicionais completas do modelo. Em seguida apresentamos as condicionais completas de um modelo MFM generalizado, com sequência $\boldsymbol{\gamma} = \{\gamma/G, G \geq 1\}$, dadas por

$$Z_i | \mathbf{Z}_{-i}, \mathbf{p}, \boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_G, G \sim \text{Cat}(\boldsymbol{\rho}) \quad (2.49)$$

$$\pi(\boldsymbol{\theta}_j | G, \mathcal{C}, \mathbf{y}, \boldsymbol{\theta}_{-j}) \propto \pi(\boldsymbol{\theta}_j) \prod_{i: z_i=j} f(y_i | \boldsymbol{\theta}_j) \quad (2.50)$$

$$\pi(G | \mathcal{C}, \boldsymbol{\gamma}) \propto \pi_G(G) \frac{\gamma^{G+G}!}{G^{G_+} (G - G_+)!} \prod_{j=1}^{G_+} \frac{\Gamma(m_j + \gamma/G)}{\Gamma(1 + \gamma/G)}, G \geq G_+ \quad (2.51)$$

$$\pi(\boldsymbol{\gamma} | \mathcal{C}, G) \propto \pi(\boldsymbol{\gamma}) \frac{\gamma^{G+\Gamma(\boldsymbol{\gamma})}}{\Gamma(n + \boldsymbol{\gamma})} \prod_{j=1}^{G_+} \frac{\Gamma(m_j + \gamma/G)}{\Gamma(1 + \gamma/G)} \quad (2.52)$$

$$\mathbf{p} | \mathcal{C}, \boldsymbol{\gamma}, G \sim \text{Dirichlet}(\gamma/G + m_1, \dots, \gamma/G + m_G), \quad (2.53)$$

em que $\boldsymbol{\rho} = (\rho_1, \dots, \rho_G)^\top$, $\rho_j = p_j f(y_i | \boldsymbol{\theta}_j) / \sum_{l=1}^G p_l f(y_i | \boldsymbol{\theta}_l)$, $j = 1, 2, \dots, G$. Detalhes sobre a obtenção das distribuições condicionais completas (2.51) e (2.52) estão disponíveis no Apêndice B, a obtenção das demais condicionais completas podem ser encontradas no Apêndice A para o caso do modelo MRST.

Estamos interessados em utilizar o TS para a estimação dos parâmetros do modelo MRST além da determinação do número de componentes, ou seja, estamos interessados na estimação

do seguinte modelo hierárquico, obtido ao combinar as expressões (2.21), (2.32) e (2.37)

$$\begin{aligned}
G &\sim \pi_G(\cdot); \\
\gamma &\sim \mathcal{F}(6, 3); \\
\mathbf{p}|G &\sim \text{DirichSim}_G(\gamma/G); \\
\boldsymbol{\theta}_j|G &\stackrel{\text{ind}}{\sim} \pi(\cdot), j = 1, 2, \dots, G; \\
\mathbf{v}|\boldsymbol{\alpha} &\sim \text{ExpLoc}(\boldsymbol{\psi}, \boldsymbol{\alpha}), \boldsymbol{\alpha} \sim \text{Unif}(0.02, 0.5); \\
Z_i|G, \mathbf{p} &\stackrel{\text{ind}}{\sim} \text{Cat}(\mathbf{p}), i = 1, 2, \dots, n; \\
U_i|\mathbf{v} &\stackrel{\text{ind}}{\sim} \text{Gama}(\mathbf{v}/2, \mathbf{v}/2), i = 1, 2, \dots, n; \\
T_i|U_i = u_i &\stackrel{\text{ind}}{\sim} \text{N}_{(0, \infty)}(0, u_i^{-1}), i = 1, 2, \dots, n; \\
Y_i|G, Z_i = j, U_i = u_i, T_i = t_i, \boldsymbol{\theta}_j, \mathbf{x}_i &\stackrel{\text{ind}}{\sim} \text{N}(\mathbf{x}_i^\top \boldsymbol{\beta}_j + (b + t_i)\Delta_j, u_i^{-1} \boldsymbol{\tau}_j^2), i = 1, 2, \dots, n,
\end{aligned} \tag{2.54}$$

em que $\mathcal{F}(6, 3)$ denota a distribuição F de Snedecor com 6 graus de liberdade no numerador e 3 no denominador, utilizada por recomendação dos autores, e com $\pi(\cdot)$ como dado em (2.38). A distribuição *a posteriori* deste modelo, $\pi(G, \boldsymbol{\theta}|\mathbf{y})$, não pode ser facilmente obtida, o que pode ser verificado ao combinar a função de verossimilhança do modelo MRST em (2.35) a densidade *a priori* dos parâmetros do modelo em (2.39) e as especificações *a priori* de G e γ mostradas em (2.54). A consideração das variáveis latentes \mathbf{Z} , \mathbf{U} e \mathbf{T} no modelo permite a existência das distribuições condicionais completas mostradas nas expressões (2.40) a (2.47), porém, a distribuição *a posteriori* $\pi(G, \gamma, \boldsymbol{\theta}, \boldsymbol{\alpha}, \mathbf{z}, \mathbf{u}, \mathbf{t}|\mathbf{y})$ também é de difícil obtenção.

No Algoritmo 3 apresentamos o *Telescoping Sampler* com as devidas modificações necessárias para a obtenção de uma Cadeia de Markov cuja distribuição estacionária é $\pi(G, \gamma, \boldsymbol{\theta}, \boldsymbol{\alpha}, \mathbf{z}, \mathbf{u}, \mathbf{t}|\mathbf{y})$. Diferentemente dos autores, escolhemos atualizar os parâmetros das componentes preenchidas antes de atualizar a partição \mathcal{C} pois dessa maneira observamos que o algoritmo atinge convergência mais rapidamente. Também alteramos procedimento de simulação dos parâmetros específicos das componentes, originalmente feito utilizando a distribuição condicional completa conjunta mostrada na forma geral em (2.50), neste caso dada por

$$\pi(\boldsymbol{\theta}_j|G, \mathcal{C}, \mathbf{y}, \boldsymbol{\theta}_{-j}) \propto \pi(\boldsymbol{\theta}_j) \prod_{i:z_i=j} \text{St}(y_i|\mathbf{x}_i^\top \boldsymbol{\beta}_j + b\Delta_j, \boldsymbol{\tau}_j^2 + \Delta_j^2, \Delta_j/\boldsymbol{\tau}_j, \mathbf{v}),$$

para a simulação a partir das condicionais completas mostradas em (2.40) a (2.42). Além disso, incluímos a atualização de variáveis latentes além das variáveis de alocação e incluímos a atualização de um parâmetro comum às componentes.

Algoritmo 3 – *Telescoping Sampler* para o modelo MRST.

Entrada: Número de iterações Q , vetor de respostas \mathbf{y} , matriz de planejamento \mathbf{X} , valores iniciais $G^{(0)}$, $G_+^{(0)}$, $\gamma^{(0)}$, $\alpha^{(0)}$, $\mathbf{p}^{(0)}$, $\boldsymbol{\theta}_1^{(0)}, \dots, \boldsymbol{\theta}_{G^{(0)}}^{(0)}$, $\mathbf{v}^{(0)}$, $\mathbf{Z}^{(0)}$, $\mathbf{U}^{(0)}$, $\mathbf{T}^{(0)}$ e valores dos hiperparâmetros.

Saída: Cadeia de Markov com a distribuição *a posteriori* do modelo como distribuição estacionária.

1: **para** $q = 1$ até Q **faça**

1. Atualize os parâmetros das componentes preenchidas, para $j = 1, 2, \dots, G_+^{(q)}$, amostre $\boldsymbol{\theta}_j^{(q)} = (\boldsymbol{\beta}_j^{(q)\top}, \tau_j^{2(q)}, \Delta_j^{(q)})$ da seguinte maneira:

a) amostre $\boldsymbol{\beta}_j^{(q)}$ a partir da condicional completa (2.40), fixando $\tau_j^2 = \tau_j^{2(q-1)}$, $\Delta_j = \Delta_j^{(q-1)}$, $\mathbf{v} = \mathbf{v}^{(q-1)}$, $\mathbf{z} = \mathbf{Z}^{(q-1)}$, $\mathbf{u} = \mathbf{U}^{(q-1)}$, $\mathbf{t} = \mathbf{T}^{(q-1)}$;

b) amostre $\tau_j^{2(q)}$ a partir da condicional completa (2.41), fixando $\boldsymbol{\beta}_j = \boldsymbol{\beta}_j^{(q)}$, $\Delta_j = \Delta_j^{(q-1)}$, $\mathbf{v} = \mathbf{v}^{(q-1)}$, $\mathbf{z} = \mathbf{Z}^{(q-1)}$, $\mathbf{u} = \mathbf{U}^{(q-1)}$, $\mathbf{t} = \mathbf{T}^{(q-1)}$;

c) amostre $\Delta_j^{(q)}$ a partir da condicional completa (2.42), fixando $\boldsymbol{\beta}_j = \boldsymbol{\beta}_j^{(q)}$, $\tau_j^2 = \tau_j^{2(q)}$, $\mathbf{v} = \mathbf{v}^{(q-1)}$, $\mathbf{z} = \mathbf{Z}^{(q-1)}$, $\mathbf{u} = \mathbf{U}^{(q-1)}$, $\mathbf{t} = \mathbf{T}^{(q-1)}$.

2. Atualize a partição \mathcal{C} :

a) Amostre $Z_i^{(q)}$, para $i = 1, 2, \dots, n$, da condicional completa (2.44), fixando $\mathbf{p} = \mathbf{p}^{(q-1)}$, $\boldsymbol{\beta}_j = \boldsymbol{\beta}_j^{(q)}$, $\tau_j^2 = \tau_j^{2(q)}$, $\Delta_j = \Delta_j^{(q)}$, $\mathbf{v} = \mathbf{v}^{(q-1)}$ e $G = G^{(q-1)}$;

b) Calcule $m_j = \sum_{i=1}^n \mathbb{1}(Z_i^{(q)} = j)$, para $j = 1, 2, \dots, G^{(q-1)}$, determine o número de grupos na amostra $G_+^{(q)} = \sum_{j=1}^{G^{(q-1)}} \mathbb{1}\{m_j > 0\}$ e troque os rótulos $j = 1, 2, \dots, G^{(q-1)}$ de forma que a primeiras $G_+^{(q)}$ componentes estejam preenchidas.

3. Atualize as variáveis latentes;

a) amostre $T_i^{(q)}$, para $i = 1, 2, \dots, n$, da condicional completa (2.46), fixando $\boldsymbol{\beta}_j = \boldsymbol{\beta}_j^{(q)}$, $\tau_j^2 = \tau_j^{2(q)}$, $\Delta_j = \Delta_j^{(q)}$, $\mathbf{v} = \mathbf{v}^{(q-1)}$ e $u_i = U_i^{(q-1)}$;

b) amostre $U_i^{(q)}$, para $i = 1, 2, \dots, n$, a partir da condicional completa (2.45), fixando $\boldsymbol{\beta}_j = \boldsymbol{\beta}_j^{(q)}$, $\tau_j^2 = \tau_j^{2(q)}$, $\Delta_j = \Delta_j^{(q)}$, $\mathbf{v} = \mathbf{v}^{(q-1)}$ e $t_i = T_i^{(q)}$.

4. Amostre $\alpha^{(q)}$ a partir da condicional completa (2.43), fixando $\mathbf{v} = \mathbf{v}^{(q-1)}$;

5. Dado \mathcal{C} , atualize G e γ ;

a) Amostre $G^{(q)}$ da condicional completa (2.51), fixando $\gamma = \gamma^{(q-1)}$;

b) Amostre $\gamma^{(q)}$ utilizando um passo de MH com proposta $\log(\gamma^*) \sim N(\log(\gamma^{(q-1)}), s_\gamma^2)$ para amostrar da condicional completa (2.52), fixando $G = G^{(q)}$.

continua

Perceba que a mudança no espaço paramétrico devido à aleatoriedade do número de componentes ocorrem no segundo e quinto passo do Algoritmo 3. No passo 5. novas componentes vazias são criadas as quais poderão ser preenchidas ao passo 2. ser executado na próxima iteração

continuação do Algoritmo 3

6. Adicione $G^{(q)} - G_+^{(q)}$ componentes vazias, atualize \mathbf{p} e \mathbf{v} .
 - a) Se $G^{(q)} > G_+^{(q)}$, amostre $\boldsymbol{\theta}_j^{(q)}$ a partir da distribuição *a priori* definida em (2.37), para $j = G_+^{(q)} + 1, \dots, G^{(q)}$;
 - b) dado \mathcal{C} , amostre $\mathbf{p}^{(q)}$ da condicional completa (2.53), fixando $\gamma = \gamma^{(q)}$ e $G = G^{(q)}$;
 - c) amostre $\mathbf{v}^{(q)}$ utilizando um passo de MH com proposta $\log(\mathbf{v}^* - \boldsymbol{\psi}) \sim \text{N}(\log(\mathbf{v}^{(q-1)} - \boldsymbol{\psi}), s_v^2)$ para amostrar da condicional completa (2.47), fixando $\boldsymbol{\alpha} = \boldsymbol{\alpha}^{(q)}$, $\mathbf{p}^{(q)}$ e $\boldsymbol{\theta}_j = \boldsymbol{\theta}_j^{(q)}$, para $j = 1, \dots, G^{(q)}$.

fim para

do algoritmo. O passo 2., além de poder preencher componentes vazias, caso existam, possibilita que componentes existentes sejam esvaziadas caso nenhuma observação seja alocada à pelo menos uma componente.

No passo 3. do Algoritmo 3, a simulação de valores a partir da condicional completa (2.46) foi realizada utilizando o algoritmo de Robert (1995), devido a dificuldade de simulação de amostras de uma distribuição normal truncada quando o ponto de truncamento está muito distante da média antes do truncamento. Além disso, No passo 5. do Algoritmo 3, a simulação da condicional completa (2.51) é feita na prática simulando valores de uma distribuição categórica tomando valores $\{G_+, \dots, G_{max}\}$, com G_{max} fixo e com probabilidades proporcionais as mostradas em (2.51).

Nos passos 5. e 6. do Algoritmo 3, o “passo de MH” refere-se a aplicação do Algoritmo 1. A probabilidade de aceitação de um valor proposto γ^* no passo 5. é dada por

$$\min \left\{ 1, \frac{\gamma^*}{\gamma^{(q-1)}} \frac{\pi_\gamma(\gamma^* | \mathcal{C}, G)}{\pi_\gamma(\gamma^{(q-1)} | \mathcal{C}, G)} \right\},$$

em que π_γ denota a função densidade da condicional completa de γ a menos de constantes apresentada em (2.52), enquanto a probabilidade de aceitação de um valor proposto \mathbf{v}^* no passo 6. é dada por

$$\min \left\{ 1, \frac{\mathbf{v}^* - \boldsymbol{\psi}}{\mathbf{v}^{(q-1)} - \boldsymbol{\psi}} \frac{\pi_v(\mathbf{v}^* | \mathbf{y}, \dots)}{\pi_v(\mathbf{v}^{(q-1)} | \mathbf{y}, \dots)} \right\},$$

em que π_v denota a função densidade da condicional completa de \mathbf{v} a menos de constantes apresentada em (2.47).

Ao final da execução do Algoritmo 3, obtemos uma Cadeia de Markov cuja distribuição estacionária é $\pi(G, \gamma, \boldsymbol{\theta}, \boldsymbol{\alpha}, \mathbf{z}, \mathbf{u}, \mathbf{t} | \mathbf{y})$. Porém, os problemas de identificabilidade apresentados na Seção 2.3 podem implicar em problemas na estimação dos parâmetros que será realizada utilizando esta Cadeia de Markov.

A solução para este problema será realizada utilizando o algoritmo ECR (PAPASTAMOULIS; ILIOPOULOS, 2010). Entretanto, como o algoritmo foi elaborado apenas para modelos de mistura com G conhecido, estimaremos primeiramente o número de grupos pelo

máximo *a posteriori* e então manteremos apenas as realizações da Cadeia de Markov tais que o número de grupos coincide com o máximo *a posteriori*, como feito por [Frühwirth-Schnatter, Malsiner-Walli e Grün \(2021\)](#).

A realização deste procedimento ignora a possível presença de componentes vazios por conta do funcionamento do algoritmo ECR, a solução do *label switching* leva em consideração apenas informações de componentes com pelo menos uma observação, mas, como vemos no Algoritmo 3, os parâmetros correspondentes a componentes vazias são gerados a partir da distribuição *a priori*, ou seja, nenhuma informação é obtida após a observação dos dados. Por conta disto, recomendamos o uso do TS para a estimação do número de componentes e grupos em um modelo de misturas, porém, caso as estimativas do número de grupos e número de componentes não coincidam, os parâmetros correspondentes as componentes vazias devem ser ignorados. A discrepância entre essas duas quantidades indica apenas evidência de componentes não observadas durante a coleta dos dados.

Em seguida, apresentamos um método para a verificação do ajuste do modelo aos dados.

2.5 Análise residual

Utilizaremos os resíduos quantílicos para a análise residual, similares aos resíduos apresentados por [Dunn e Smyth \(1996\)](#). Relembre do seguinte resultado, se $X \sim F(\cdot)$, então $F(X) \sim \text{Uniforme}(0, 1)$ e é possível mostrar que $\Phi^{-1}(F(X)) \sim N(0, 1)$, com $\Phi^{-1}(\cdot)$ denotando a função quantil da distribuição normal padrão. Podemos então utilizar este resultado para definir o resíduo quantílico da i -ésima observação, visto que, sob o modelo MRST $Y_i|Z_i = j \sim \text{ST}(\mathbf{x}_i^\top \boldsymbol{\beta}_j + b\Delta_j, \sigma_j^2, \lambda_j, \nu)$, então vemos que a função distribuição para a i -ésima observação é

$$\begin{aligned} F(y_i|G, \boldsymbol{\theta}, \mathbf{X}) &= \int_{-\infty}^{y_i} \sum_{j=1}^G p_j \text{St}(s|\mathbf{x}_i^\top \boldsymbol{\beta}_j + b\Delta_j, \sigma_j^2, \lambda_j, \nu) ds \\ &= \sum_{j=1}^G p_j \int_{-\infty}^{y_i} \text{St}(s|\mathbf{x}_i^\top \boldsymbol{\beta}_j + b\Delta_j, \sigma_j^2, \lambda_j, \nu) ds \\ &= \sum_{j=1}^G p_j \text{ST}(y_i|\mathbf{x}_i^\top \boldsymbol{\beta}_j + b\Delta_j, \sigma_j^2, \lambda_j, \nu), \end{aligned}$$

em que $\text{ST}(\cdot|\mu, \sigma^2, \lambda, \nu)$ denota a função distribuição ST com parâmetro de localização μ , de dispersão σ^2 , de forma λ e ν graus de liberdade. Note que a função distribuição é invariante a permutações dos rótulos das componentes.

O resíduo quantílico da i -ésima observação é dado por

$$e_i = \Phi^{-1} \left(\sum_{j=1}^G p_j \text{ST}(y_i|\mathbf{x}_i^\top \boldsymbol{\beta}_j + b\Delta_j, \sigma_j^2, \lambda_j, \nu) \right). \quad (2.55)$$

Por construção $e_i \sim N(0, 1)$, desvios da normalidade indicam que a mistura de distribuições ST assumida não é adequada para a modelagem da observação y_i . Organizando os resíduos em

ordem crescente, $e_{(1)}, e_{(2)}, \dots, e_{(n)}$, esperamos que $e_{(i)}$ esteja próximo do quantil $(i - 0,5)/n$ da distribuição normal, ou seja, próximo de $\Phi^{-1}((i - 0,5)/n)$. Podemos então elaborar um gráfico com os $\Phi^{-1}((i - 0,5)/n)$ no eixo das abcissas e $e_{(i)}$ no eixo das ordenadas. Se o modelo é adequado esperamos que os pontos estejam próximos da reta $y = x$. Ver [Chambers \(2018\)](#) para mais detalhes.

Como a função distribuição é invariante a permutações dos rótulos das componentes, podemos utilizar uma amostra MCMC antes da realização do processo de correção de *label switching* para obter uma amostra da distribuição de cada $e_{(i)}$, calculando (2.55) através da substituição dos parâmetros por suas respectivas amostras MCMC, a função distribuição ST está implementada no pacote `sn` ([AZZALINI, 2023](#)) do software R ([R Core Team, 2022](#)). Temos então uma amostra da distribuição *a posteriori* de $e_{(i)}$, procedemos com a estimação pontual de $e_{(i)}$ pela média *a posteriori* e construímos um intervalo de credibilidade para $e_{(i)}$ com os quantis amostrais 2,5% e 97,5%.

RESULTADOS

3.1 Modelo simplificado

Considere o seguinte modelo $Y_1, \dots, Y_n \stackrel{i.i.d.}{\sim} \text{ST}(0, \sigma^2, \lambda, 3)$, de forma similar ao que foi feito em (2.32). Utilizaremos a representação em variáveis latentes, apresentada em (2.13), o modelo escrito de forma hierárquica é dado por

$$\begin{aligned} Y_i | T_i = t_i, U_i = u_i &\sim \text{N}(\Delta t_i, u_i^{-1} \tau^2); \\ T_i | U_i = u_i &\sim \text{N}_{(0, \infty)}(0, u_i^{-1}); \\ U_i &\sim \text{Gama}(3/2, 3/2), \end{aligned} \quad (3.1)$$

com $\tau^2 = \sigma^2(1 - \delta^2)$, $\Delta = \sigma\delta$ e $\delta = \lambda/\sqrt{1 + \lambda^2}$. Se consideramos independência *a priori* entre τ^2 e Δ e as seguintes distribuições *a priori*: $\tau^2 \sim \text{InvGama}(r, s)$ e $\Delta \sim \text{N}(\eta, \omega^2)$, analogamente a (2.37), obtemos condicionais completas similares as dadas nas expressões (2.40) a (2.47), as distribuições condicionais completas para o modelo simplificado são dadas por

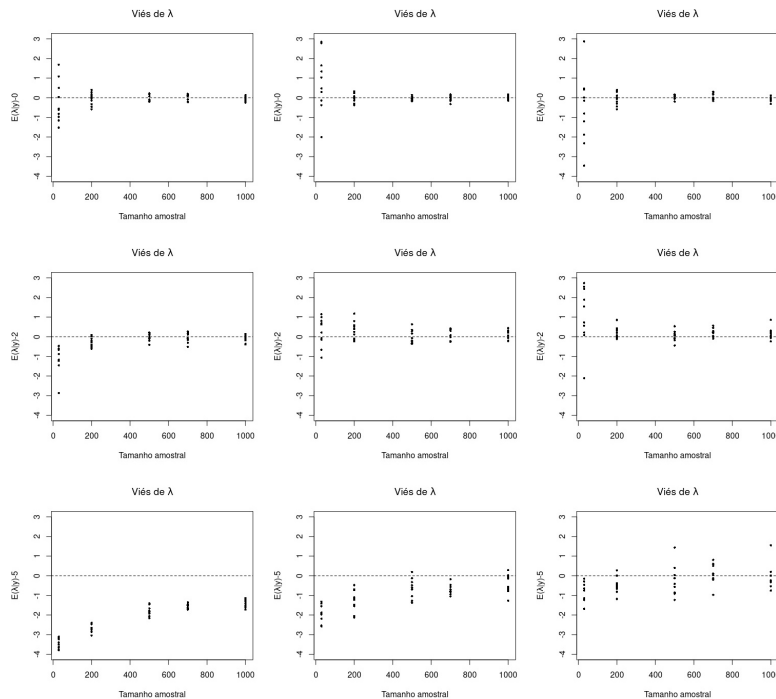
$$\begin{aligned} \tau^2 | \Delta, \mathbf{y}, \mathbf{t}, \mathbf{u} &\sim \text{InvGama}\left(\frac{n}{2} + r, \frac{\sum_{i=1}^n u_i (y_i - \Delta t_i)^2}{2} + s\right) \\ \Delta | \tau^2, \mathbf{y}, \mathbf{t}, \mathbf{u} &\sim \text{N}\left(\frac{\omega^2 \sum_{i=1}^n u_i t_i y_i + \eta \tau_j^2}{\omega^2 \sum_{i=1}^n u_i t_i^2 + \tau_j^2}, \frac{\omega^2 \tau_j^2}{\omega^2 \sum_{i=1}^n u_i t_i^2 + \tau_j^2}\right). \end{aligned} \quad (3.2)$$

Utilizando as condicionais completas (3.2), é possível aplicar o Algoritmo 2 para simular valores da distribuição *a posteriori* de $\Delta, \tau^2 | y_1, \dots, y_n$. Desta maneira, temos $\mathbb{E}(\Delta) = 0$, $\text{Var}(\Delta) = 100$, $\mathbb{E}(\tau^2) = 1$ e $\text{Var}(\tau^2) = 10$, o que configura uma distribuição *a priori* pouco informativa. A representação hierárquica (2.13) foi utilizada para gerar dez conjuntos de dados distintos compostos de amostras independentes do modelo de tamanho $n = 30, 200, 500, 700, 1000$, fixamos $\sigma^2 = 1, 4, 9$ e $\lambda = 0, 2, 5$.

O amostrador de Gibbs foi implementado utilizando as condicionais completas mostradas em (3.2) e obtemos um conjunto de 10 amostras MCMC para τ^2 e Δ em cada cenário. Após isto,

utilizamos as transformações inversas (2.33) para obter amostras da distribuição *a posteriori* de σ^2 e λ . Na Figura 5, apresentamos estimativas de λ pela média *a posteriori* subtraindo seu respectivo valor fixado. Na Figura 6 apresentamos um gráfico com intervalos de credibilidade de nível 95% para λ .

Figura 5 – Média *a posteriori* de λ menos seu valor fixado, reta pontilhada marca zero. Nas linhas temos $\lambda = 0$ (linha 1), 2 (linha 2) e 5 (linha 3), e nas colunas $\sigma^2 = 1$ (coluna 1), 4 (coluna 2) e 9 (coluna 3).

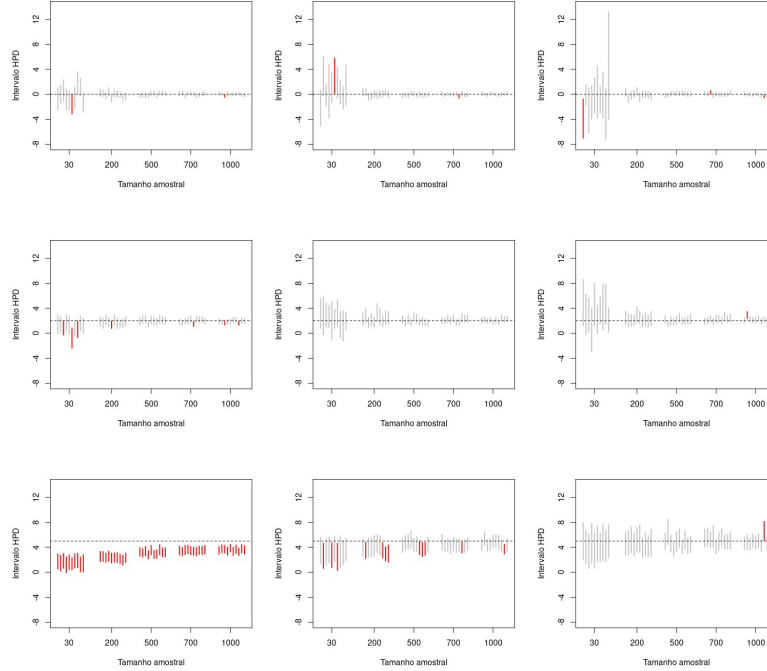


Fonte: Elaborada pelo autor.

Na Figuras 5 e 6 vemos que há uma tendência de subestimação de λ conforme o valor fixado aumenta, essa tendência aparenta ser corrigida com um σ^2 fixado em valores mais elevados. Notamos também que a medida que temos maiores tamanhos amostrais, a variabilidade da distribuição *a posteriori* tende a diminuir, com seus intervalos de credibilidade contendo o valor fixado. Além disso, observe que um valor baixo de σ^2 implica que um tamanho amostral mais elevado será necessário para obtermos estimativas mais precisas. Isto está de acordo com o que foi observado na Figura 1, o que indica que o parâmetro λ é de difícil estimação.

Em um contexto frequentista, Sartori (2006) mostra que estimativas por máxima verossimilhança de λ podem ser infinitas com probabilidade positiva. Em uma abordagem Bayesiana, Branco, Genton e Liseo (2013) propõem o uso da distribuição *a priori* de Jeffreys para λ e demonstram que as estimativas pelo máximo *a posteriori* apresentam melhor comportamento do que as estimativas de máxima verossimilhança, porém, a aplicação desta metodologia é difícil em nossa abordagem pois não leva em consideração a reparametrização τ^2 , Δ , crucial para a obtenção de distribuições condicionais completas a partir das quais a simulação de valores

Figura 6 – Intervalos de credibilidade para λ no painel, reta pontilhada marca o valor fixado de λ e a cor do intervalo é vermelha caso o mesmo não contenha o valor fixado. Nas linhas temos $\lambda = 0$ (linha 1), 2 (linha 2) e 5 (linha 3), e nas colunas $\sigma^2 = 1$ (coluna 1), 4 (coluna 2) e 9 (coluna 3).



Fonte: Elaborada pelo autor.

é simples, não havendo a necessidade da aplicação do algoritmo de Metropolis-Hastings na atualização destes dois parâmetros.

3.2 Modelo completo

Nesta seção apresentamos resultados do ajuste do modelo MRST utilizando o Algoritmo 3 em conjuntos de dados sintéticos e a um conjunto de dados referente a salários de jogadores de baseball.

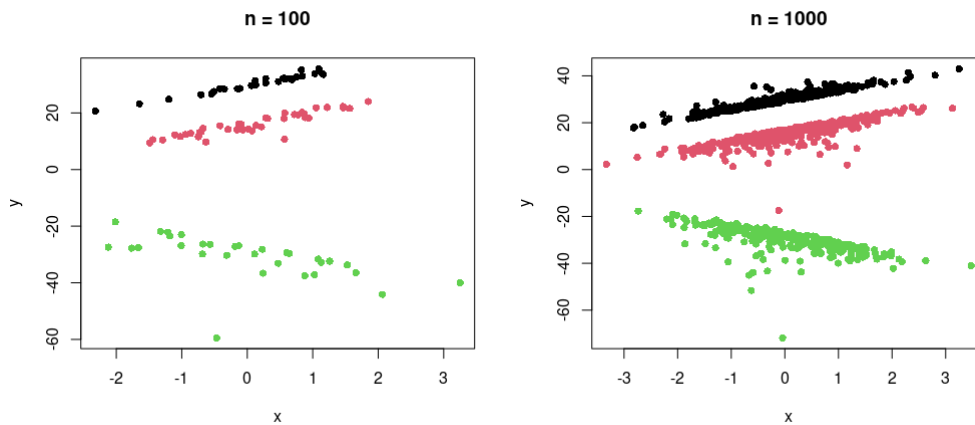
3.2.1 Dados simulados

Os dados foram gerados utilizando a representação hierárquica do modelo MRST mostrada em (2.32) com $G = 3$. A matriz de planejamento \mathbf{X} foi construída como uma matriz com uns na primeira coluna e a segunda como n amostras de uma distribuição normal padrão, ou seja, temos $k = 1$. Consideramos dois tamanhos amostrais, $n = 100$ e $n = 1000$. Os parâmetros foram fixados nos seguintes valores:

$$\mathbf{p} = \begin{bmatrix} 0,3 \\ 0,4 \\ 0,3 \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} 30 & 15 & -30 \\ 4 & 4 & -4 \end{bmatrix}, \quad \boldsymbol{\sigma}^2 = \begin{bmatrix} 1 \\ 4 \\ 9 \end{bmatrix}, \quad \boldsymbol{\lambda} = \begin{bmatrix} 10 \\ -8 \\ -8 \end{bmatrix} \text{ e } \nu = 3.$$

O valor do parâmetro \mathbf{p} foi escolhido de forma a obtermos dados com aproximadamente a mesma quantidade de observações por componente, $\boldsymbol{\beta}$ foi escolhido de forma a obtermos componentes cujas retas de regressão estejam relativamente distantes, os valores de σ^2 foram escolhidos de forma arbitrária e $\boldsymbol{\lambda}$ e ν foram determinados tais que fossem observados assimetria e *outliers*, respectivamente. Para cada tamanho amostral considerado, foram geradas dez amostras diferentes do modelo MRST utilizando os mesmos valores dos parâmetros. A Figura 7 mostra diagramas de dispersão da primeira amostra obtida para $n = 100$ e $n = 1000$, respectivamente.

Figura 7 – Diagramas de dispersão para o primeiro conjunto de dados obtido com $n = 100$ e $n = 1000$.



Fonte: Elaborada pelo autor.

Ajustamos o modelo hierárquico mostrado em (2.54), com $G - 1 \sim \text{BNB}(1, 4, 3)$ com função de probabilidade dada em (2.29) e com $G_{max} = 100$. Em cada conjunto de dados gerado, obtendo um conjunto de amostras da respectiva distribuição *a posteriori* por meio do Algoritmo 3. Os valores dos hiperparâmetros das distribuições *a priori* apresentadas em (2.37) foram fixados em

$$\begin{aligned}
 c &= 10 \\
 \eta &= 0; \omega = 10 \\
 r &= 2, 1; s = 1, 1 \\
 \psi &= 1, 1.
 \end{aligned} \tag{3.3}$$

Como discutido anteriormente, se $\nu \in [0, 1]$, o valor esperado da i -ésima observação não existe. Este problema pode ser resolvido ao tomarmos $\psi = 1$, visto que isto implica em uma probabilidade *a priori* zero para o intervalo, o valor $\psi = 1, 1$ foi escolhido devido instabilidade numérica observada quando ν está próximo de 1.

Em cada tamanho amostral, os valores iniciais necessários para a execução do algoritmo em cada conjunto de dados foram obtidos através das respectivas distribuições *a priori*, as exceções são $G^{(0)} = G_+^{(0)} = 10$, $\mathbf{p}^{(0)} = (1/G^{(0)}, \dots, 1/G^{(0)})^\top$ e com $\mathbf{Z}^{(0)}$ sendo uma amostra da distribuição $\text{Cat}(\mathbf{p}^{(0)})$. Fixamos o número de iterações do algoritmo TS em $T = 300.000$, descar-

tando as primeiras $T/2$, após isto consideramos uma a cada 50 observações para amostra MCMC de forma a obter observações distribuídas o mais próximo possível da respectiva distribuição *a posteriori*, com baixa autocorrelação. Obtemos, então, $Q = 3.000$ amostras MCMC.

Após a verificação de convergência das amostras obtidas em cada conjunto de dados, solucionamos o *label switching* e realizamos o procedimento de estimação. A Tabela 1 apresenta, para conjunto de dados com $n = 100$, as médias das estimativas dos parâmetros acompanhadas do seu respectivo valor fixado, viés e Erro Quadrático Médio (EQM). As estimativas dos parâmetros são dadas pela média *a posteriori*, com exceção de G e G_+ , que são estimados pela moda *a posteriori*. O “parâmetro” Z apresenta, na coluna referente à média, a proporção média de classificações que coincidem com o vetor de alocações que gerou os dados e em EQM a variância dessas proporções. A Tabela 2 mostra as informações referentes aos conjuntos de dados com $n = 1000$.

Tabela 1 – Estimativas *a posteriori* para $n = 100$.

Parâmetro	Fixado	Média	Viés	EQM
G	3	3	0	0
G_+	3	3	0	0
p	$\begin{bmatrix} 0,3 \\ 0,4 \\ 0,3 \end{bmatrix}$	$\begin{bmatrix} 0,278 \\ 0,401 \\ 0,319 \end{bmatrix}$	$\begin{bmatrix} -0,022 \\ 0,001 \\ 0,019 \end{bmatrix}$	$\begin{bmatrix} 0,001 \\ 0,002 \\ 0,003 \end{bmatrix}$
β	$\begin{bmatrix} 30 & 15 & -30 \\ 4 & 4 & -4 \end{bmatrix}$	$\begin{bmatrix} 29,958 & 14,824 & -30,146 \\ 4,024 & 4,008 & -3,954 \end{bmatrix}$	$\begin{bmatrix} -0,042 & -0,176 & -0,146 \\ 0,024 & 0,008 & 0,046 \end{bmatrix}$	$\begin{bmatrix} 0,058 & 0,613 & 0,353 \\ 0,018 & 0,018 & 0,036 \end{bmatrix}$
σ^2	$\begin{bmatrix} 1 \\ 4 \\ 9 \end{bmatrix}$	$\begin{bmatrix} 1,015 \\ 5,061 \\ 11,697 \end{bmatrix}$	$\begin{bmatrix} 0,015 \\ 1,061 \\ 2,697 \end{bmatrix}$	$\begin{bmatrix} 0,295 \\ 11,918 \\ 48,835 \end{bmatrix}$
λ	$\begin{bmatrix} 10 \\ -8 \\ -8 \end{bmatrix}$	$\begin{bmatrix} 1,169 \\ -3,256 \\ -4,507 \end{bmatrix}$	$\begin{bmatrix} -8,831 \\ 4,744 \\ 3,493 \end{bmatrix}$	$\begin{bmatrix} 78,164 \\ 23,808 \\ 15,342 \end{bmatrix}$
v	3	5,068	2,068	12,016
Z	-	1	-	0

Fonte: Elaborada pelo autor.

É possível notar nas Tabelas 1 e 2 as estimativas consistentes dos pesos das componentes e coeficientes de regressão, ambos apresentando baixo viés e EQM. As estimativas dos demais parâmetros tem seu viés e EQM reduzidos ao considerarmos um tamanho amostral maior. Porém, mesmo com $n = 1000$ as estimativas de λ ainda não são satisfatórias, o que pode ser explicado pelo número de observações alocado em cada componente, pois vimos na Seção 3.1 que há uma tendência de subestimação dos valores de λ a depender do tamanho amostral e dos valores fixados de σ^2 e λ . O tamanho amostral aqui considerado pode ainda não ser suficiente para uma boa estimação deste parâmetro. Uma possível solução para melhorar a estimação deste parâmetro seria o uso da distribuição *a priori* de Jeffreys para λ , como feito por Branco, Genton e Liseo (2013), o que exigiria grandes alterações do Algoritmo 3. Entretanto, em uma análise de

Tabela 2 – Estimativas *a posteriori* para $n = 1000$.

Parâmetro	Fixado	Média	Viés	EQM
G	3	3	0	0
G_+	3	3	0	0
p	$\begin{bmatrix} 0,3 \\ 0,4 \\ 0,3 \end{bmatrix}$	$\begin{bmatrix} 0,303 \\ 0,399 \\ 0,298 \end{bmatrix}$	$\begin{bmatrix} 0,003 \\ -0,001 \\ -0,002 \end{bmatrix}$	$\begin{bmatrix} 1,031 \times 10^{-4} \\ 3,883 \times 10^{-5} \\ 7,850 \times 10^{-5} \end{bmatrix}$
β	$\begin{bmatrix} 30 & 15 & -30 \\ 4 & 4 & -4 \end{bmatrix}$	$\begin{bmatrix} 29,975 & 15,016 & -30,049 \\ 4,015 & 4,033 & -3,979 \end{bmatrix}$	$\begin{bmatrix} -0,025 & 0,016 & -0,049 \\ 0,015 & 0,033 & 0,021 \end{bmatrix}$	$\begin{bmatrix} 0,008 & 0,016 & 0,061 \\ 0,001 & 0,003 & 0,009 \end{bmatrix}$
σ^2	$\begin{bmatrix} 1 \\ 4 \\ 9 \end{bmatrix}$	$\begin{bmatrix} 0,832 \\ 3,724 \\ 9,010 \end{bmatrix}$	$\begin{bmatrix} -0,168 \\ -0,276 \\ 0,010 \end{bmatrix}$	$\begin{bmatrix} 0,050 \\ 0,495 \\ 1,435 \end{bmatrix}$
λ	$\begin{bmatrix} 10 \\ -8 \\ -8 \end{bmatrix}$	$\begin{bmatrix} 2,918 \\ -5,068 \\ -6,111 \end{bmatrix}$	$\begin{bmatrix} -7,082 \\ 2,932 \\ 1,899 \end{bmatrix}$	$\begin{bmatrix} 50,237 \\ 9,071 \\ 3,899 \end{bmatrix}$
ν	3	2,993	-0,007	0,048
Z	-	1	-	0

Fonte: Elaborada pelo autor.

regressão, o principal interesse são os coeficientes de regressão que, no caso estudado, foram muito bem estimados.

3.2.2 Salários de jogadores de baseball

Também utilizamos o modelo MRST na análise do conjunto de dados de salários de jogadores de baseball, excluindo arremessadores, no ano de 1992 (WATNIK, 1998). O interesse original era verificar o impacto da performance dos jogadores em seu salário. O conjunto de dados contém, além dos salários de 337 jogadores por mil dólares, 12 medidas de performance e 4 indicadores da habilidade do jogador trocar de equipe. As medidas de performance são *batting average* (x_1), *on-base percentage* (x_2), *runs* (x_3), *hits* (x_4), *doubles* (x_5), *triples* (x_6), *home runs* (x_7), *runs batted in* (x_8), *walks* (x_9), *strikeouts* (x_{10}), *stolen bases* (x_{11}) e *errors* (x_{12}). Os indicadores de habilidade de troca de equipe são *free agency eligibility* (x_{13}), *free agent in 1991/2* (x_{14}), *arbitration eligibility* (x_{15}) e *arbitration in 1991/2* (x_{16}).

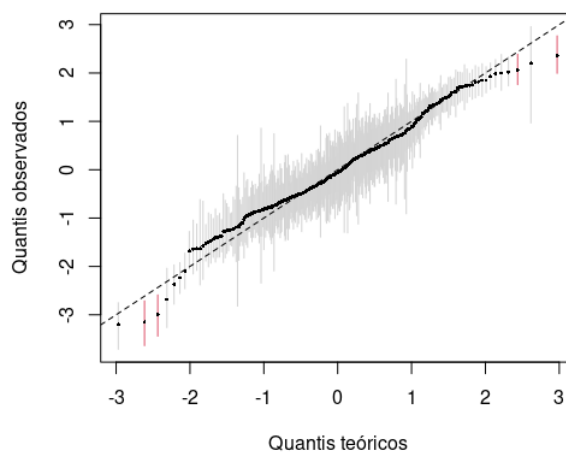
Este conjunto de dados foi anteriormente analisado por Sabillón, Cotrim e Zuanetti (2023) utilizando misturas de regressões com erros normais. Como realizado por eles, padronizamos as variáveis x_1 a x_{12} e consideramos interações entre as variáveis x_{13} a x_{16} e x_1 , x_2 , x_7 e x_8 , obtemos assim um conjunto de 32 covariáveis. Consideramos como variável resposta o logaritmo natural do salário.

Foram realizadas dez replicações do Algoritmo 3, utilizando as mesmas especificações do estudo de simulação, ou seja, com $G - 1 \sim \text{BNB}(1, 4, 3)$, $G_{max} = 100$, com hiperparâmetros como dado em (3.3) e $T = 300.000$ iterações, descartando as primeiras $T/2$ e então considerando uma a cada 50 observações para a amostra MCMC. Os valores iniciais de cada replicação

também foram obtidos da mesma forma do estudo de simulação, fixamos $G^{(0)} = G_+^{(0)} = 10$, $\mathbf{p}^{(0)} = (1/G^{(0)}, \dots, 1/G^{(0)})^\top$ e com $\mathbf{Z}^{(0)}$ sendo uma amostra da distribuição $\text{Cat}(\mathbf{p}^{(0)})$, os valores iniciais dos demais parâmetros foram fixados como amostras de suas respectivas distribuições *a priori*, ou seja, em cada replicação temos diferentes pontos iniciais. Após verificação da convergência de cada uma das replicações, as amostras obtidas foram combinadas em uma única amostra de tamanho $Q = 30.000$.

O ajuste do modelo aos dados foi verificado utilizando o gráfico quantil-quantil apresentado na Figura 8, podemos observar que, em geral, os resíduos quantílicos estão próximos dos

Figura 8 – Gráfico quantil-quantil dos resíduos quantílicos. A reta pontilhada é a reta $y = x$ e os segmentos representam os intervalos de credibilidade (95%) de cada resíduo, em cinza se contém a reta pontilhada e em vermelho caso contrário.



Fonte: Elaborada pelo autor.

quantis da distribuição normal e com o intervalo de credibilidade dos resíduos o contendo. Apenas quatro exceções foram observadas e, assim, consideraremos os resíduos aproximadamente normais. Verificada a adequação do modelo, prosseguiremos com a estimação de parâmetros após a solução do *label switching*.

A Tabela 3 apresenta estimativas dos parâmetros pela média e mediana *a posteriori*, bem como intervalos de credibilidade dados pelos quantis *a posteriori* 2,5% e 97,5%. Apresentamos as estimativas dos coeficientes de regressão cujo intervalo de credibilidade não contém zero ou caso contenha zero próximo a um de seus limites.

Vemos na Tabela 3 que o número de componentes e grupos foi estimado em 1, diferentemente de Sabillón, Cotrim e Zuanetti (2023) que estimaram 2 componentes. Os autores mencionam que a segunda componente é composta apenas de alguns jogadores atípicos, ou seja, a segunda componente pode ter sido eliminada devido a inclusão do parâmetro ν que controla a probabilidade das caudas da distribuição e permite mais observações atípicas do que a distribuição normal. O baixo valor estimado corrobora com essa conclusão. Observamos que a inclusão do parâmetro de forma λ_1 pode não ser necessária, visto que seu intervalo de

Tabela 3 – Estimativas pela média e mediana *a posteriori* de cada parâmetro do modelo além de um intervalo de credibilidade (IC) de nível 95%.

Parâmetro	Média	Mediana	IC (95%)
G	1*	1	(1;6)
G_+	1*	1	(1;1)
$\beta_{0,1}$	5,561	5,561	(5,561; 5,660)
$\beta_{4,1}$	0,312	0,311	(0,063; 0,566)
$\beta_{8,1}$	0,258	0,256	(-0,047; 0,568)
$\beta_{9,1}$	0,104	0,104	(-0,040; 0,247)
$\beta_{10,1}$	-0,119	-0,118	(-0,228; -0,012)
$\beta_{11,1}$	0,054	0,054	(-0,018; 0,125)
$\beta_{13,1}$	1,779	1,778	(1,615; 1,942)
$\beta_{14,1}$	-0,271	-0,271	(-0,481; -0,066)
$\beta_{15,1}$	1,487	1,488	(1,326; 1,646)
$\beta_{17,1}$	-0,166	-0,165	(-0,355; 0,017)
$\beta_{18,1}$	0,291	0,293	(-0,028; 0,600)
σ_1^2	0,159	0,156	(0,114; 0,220)
λ_1	0,344	0,348	(-0,224; 0,897)
ν	3,890	3,717	(2,491; 6,251)

* Moda *a posteriori*

Fonte: Elaborada pelo autor.

credibilidade contém o valor zero, ou seja, talvez a distribuição t de Student seja adequada para a modelagem destes dados. Outra observação importante é a inclusão de outros coeficientes correspondentes a medidas de desempenho, em [Sabillón, Cotrim e Zuanetti \(2023\)](#) os apenas os coeficientes $\beta_{4,1}$, $\beta_{11,1}$, $\beta_{13,1}$, $\beta_{14,1}$ e $\beta_{15,1}$ são considerados relevantes pelos autores.

DISCUSSÃO

Neste trabalho, estudamos os modelos misturas de regressão com erros t de Student assimétricos através do *Telescoping Sampler*. Com o uso deste algoritmo podemos estimar simultaneamente o número de componentes na mistura e os parâmetros do modelo. Notamos que a estimação dos coeficientes de regressão do modelo não é afetada pela presença de assimetria e observações discrepantes, apresentando baixo erro quadrático médio nos cenários de simulação considerados. Também notamos a dificuldade em estimar o parâmetro de forma, que necessita de uma quantidade elevada de amostras para ser bem estimado. Além disso, ajustamos o modelo a um conjunto de dados referentes a salários de jogadores de baseball, observamos que houve uma redução no número de componentes estimado quando comparamos a um ajuste anteriormente realizado utilizando um modelo de misturas de regressões normais. Com a aplicação do modelo MRST, também pudemos detectar covariáveis que impactam no logaritmo natural do salário diferentes das observadas no modelo com erros normais.

Futuramente é interessante buscar melhorar as estimativas do parâmetro de forma, visto que discussões nesse sentido já existem na literatura, mesmo que sejam de mais difícil implementação. Desejamos estender o modelo para considerar erros de regressão na família de misturas de escala da normal assimétrica, visto que muitos dos resultados aqui apresentados também se aplicam neste contexto mais geral.

REFERÊNCIAS

- ANTONIAK, C. E. Mixtures of dirichlet processes with applications to bayesian nonparametric problems. **The annals of statistics**, JSTOR, p. 1152–1174, 1974. Citado nas páginas 22 e 32.
- AZZALINI, A. A class of distributions which includes the normal ones. **Scandinavian Journal of Statistics**, v. 12, p. 171–178, 1985. Citado na página 27.
- AZZALINI, A.; CAPITANIO, A. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, v. 65, n. 2, p. 367–389, 2003. Disponível em: <<https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/1467-9868.00391>>. Citado nas páginas 21, 28 e 29.
- AZZALINI, A. A. **The R package sn: The skew-normal and related distributions such as the skew-t and the SUN (version 2.1.1)**. Università degli Studi di Padova, Italia, 2023. Home page: <<http://azzalini.stat.unipd.it/SN/>>. Disponível em: <<https://cran.r-project.org/package=sn>>. Citado na página 49.
- BRANCO, M. D.; DEY, D. K. A general class of multivariate skew-elliptical distributions. **Journal of Multivariate Analysis**, v. 79, n. 1, p. 99–113, 2001. ISSN 0047-259X. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0047259X00919602>>. Citado nas páginas 21 e 29.
- _____. Regression model under skew elliptical error distribution. **Journal of Mathematical Sciences**, v. 1, p. 151–169, 2002. Citado na página 21.
- BRANCO, M. D.; GENTON, M. G.; LISEO, B. Objective bayesian analysis of skew-t distributions. **Scandinavian Journal of Statistics**, Wiley Online Library, v. 40, n. 1, p. 63–85, 2013. Citado nas páginas 52 e 55.
- BROOKS, S.; GELMAN, A.; JONES, G.; MENG, X.-L. **Handbook of Markov chain monte carlo**. [S.l.]: CRC press, 2011. Citado na página 42.
- CABRAL, C. R. B.; SOUZA, N. L. de; LEÃO, J. Bayesian measurement error models using finite mixtures of scale mixtures of skew-normal distributions. **Journal of Statistical Computation and Simulation**, Taylor & Francis, v. 92, n. 3, p. 623–644, 2022. Citado nas páginas 21, 37 e 38.
- CANCHO, V. G.; DEY, D. K.; LACHOS, V. H.; ANDRADE, M. G. Bayesian nonlinear regression models with scale mixtures of skew-normal distributions: Estimation and case influence diagnostics. **Computational Statistics & Data Analysis**, v. 55, n. 1, p. 588–602, 2011. ISSN 0167-9473. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0167947310002446>>. Citado na página 21.
- CHAMBERS, J. M. **Graphical methods for data analysis**. [S.l.]: CRC Press, 2018. Citado na página 49.

- DÁVILA, V. H. L.; CABRAL, C. R. B.; ZELLER, C. B. **Finite mixture of skewed distributions**. 1. ed. [S.l.]: Springer, 2018. Citado na página 21.
- DEMPSTER, A. P.; LAIRD, N. M.; RUBIN, D. B. Maximum likelihood from incomplete data via the em algorithm. **Journal of the royal statistical society: series B (methodological)**, Wiley Online Library, v. 39, n. 1, p. 1–22, 1977. Citado na página 22.
- DUNN, P. K.; SMYTH, G. K. Randomized quantile residuals. **Journal of Computational and graphical statistics**, Taylor & Francis, v. 5, n. 3, p. 236–244, 1996. Citado na página 48.
- ESCOBAR, M. D. Estimating normal means with a dirichlet process prior. **Journal of the American Statistical Association**, Taylor & Francis, v. 89, n. 425, p. 268–277, 1994. Citado nas páginas 22 e 32.
- FERGUSON, T. S. A bayesian analysis of some nonparametric problems. **The annals of statistics**, JSTOR, p. 209–230, 1973. Citado nas páginas 22 e 32.
- FRÜHWIRTH-SCHNATTER, S. **Finite mixture and Markov switching models**. [S.l.]: Springer, 2006. Citado nas páginas 39, 40 e 42.
- FRÜHWIRTH-SCHNATTER, S.; MALSINER-WALLI, G. From here to infinity: sparse finite versus dirichlet process mixtures in model-based clustering. **Advances in data analysis and classification**, Springer, v. 13, p. 33–64, 2019. Citado na página 36.
- FRÜHWIRTH-SCHNATTER, S.; MALSINER-WALLI, G.; GRÜN, B. Generalized mixtures of finite mixtures and telescoping sampling. **Bayesian Analysis**, International Society for Bayesian Analysis, v. 16, n. 4, p. 1279–1307, 2021. Citado nas páginas 22, 32, 33, 34, 35, 36, 44, 48 e 71.
- GARAY, A. M.; BOLFARINE, H.; LACHOS, V. H.; CABRAL, C. R. Bayesian analysis of censored linear regression models with scale mixtures of normal distributions. **Journal of Applied Statistics**, Taylor & Francis, v. 42, n. 12, p. 2694–2714, 2015. Citado na página 38.
- GEMAN, S.; GEMAN, D. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. **IEEE Transactions on pattern analysis and machine intelligence**, IEEE, v. 6, n. 6, p. 721–741, 1984. Citado na página 43.
- GREEN, P. J. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. **Biometrika**, v. 82, n. 4, p. 711–732, 1995. ISSN 0006-3444. Disponível em: <<https://doi.org/10.1093/biomet/82.4.711>>. Citado na página 22.
- HANNAH, L. A.; BLEI, D. M.; POWELL, W. B. Dirichlet process mixtures of generalized linear models. **Journal of Machine Learning Research**, v. 12, n. 6, 2011. Citado na página 22.
- HASTINGS, W. K. Monte Carlo sampling methods using Markov chains and their applications. **Biometrika**, v. 57, n. 1, p. 97–109, 04 1970. ISSN 0006-3444. Disponível em: <<https://doi.org/10.1093/biomet/57.1.97>>. Citado na página 42.
- HENNIG, C. Identifiability of models for clusterwise linear regression. **Journal of classification**, v. 17, n. 2, 2000. Citado na página 41.
- HENZE, N. A probabilistic representation of the skew-normal distribution. **Scandinavian Journal of Statistics**, v. 13, p. 271–275, 1986. Citado na página 27.

LIU, W.; ZHANG, B.; ZHANG, Z.; TAO, J.; BRANSCUM, A. J. Model selection in finite mixture of regression models: a bayesian approach with innovative weighted g priors and reversible jump markov chain monte carlo implementation. **Journal of Statistical Computation and Simulation**, Taylor & Francis, v. 85, n. 12, p. 2456–2478, 2015. Citado na página 22.

MALSINER-WALLI, G.; FRÜHWIRTH-SCHNATTER, S.; GRÜN, B. Model-based clustering based on sparse finite gaussian mixtures. **Statistics and computing**, Springer, v. 26, n. 1, p. 303–324, 2016. Citado na página 35.

MASSUIA, M. B.; GARAY, A. M.; CABRAL, C. R.; LACHOS, V. Bayesian analysis of censored linear regression models with scale mixtures of skew-normal distributions. **Statistics and its Interface**, International Press of Boston, v. 10, n. 3, p. 425–439, 2017. Citado na página 21.

MCLACHLAN, G. J.; KRISHNAN, T. **The EM algorithm and extensions**. [S.l.]: John Wiley & Sons, 2007. Citado na página 22.

METROPOLIS, N.; ROSENBLUTH, A. W.; ROSENBLUTH, M. N.; TELLER, A. H.; TELLER, E. Equation of State Calculations by Fast Computing Machines. **The Journal of Chemical Physics**, v. 21, n. 6, p. 1087–1092, 06 1953. ISSN 0021-9606. Disponível em: <<https://doi.org/10.1063/1.1699114>>. Citado na página 42.

MILLER, J. W.; HARRISON, M. T. Mixture models with a prior on the number of components. **Journal of the American Statistical Association**, Taylor & Francis, v. 113, n. 521, p. 340–356, 2018. Disponível em: <<https://doi.org/10.1080/01621459.2016.1255636>>. Citado nas páginas 31, 32, 33 e 35.

MÜLLER, P.; ERKANLI, A.; WEST, M. Bayesian curve fitting using multivariate normal mixtures. **Biometrika**, Oxford University Press, v. 83, n. 1, p. 67–79, 1996. Citado na página 22.

NASCIMENTO, M. G. L. **Abordagem Bayesiana em Misturas Finitas de Distribuições Assimétricas**. Dissertação (Mestrado) — Universidade Federal do Rio de Janeiro, 2017. Citado nas páginas 22, 37 e 38.

PAPASTAMOULIS, P.; ILIOPOULOS, G. An artificial allocations based solution to the label switching problem in bayesian analysis of mixtures of distributions. **Journal of Computational and Graphical Statistics**, Taylor & Francis, v. 19, n. 2, p. 313–331, 2010. Citado na página 47.

QUANDT, R. E.; RAMSEY, J. B. Estimating mixtures of normal distributions and switching regressions. **Journal of the American statistical Association**, Taylor & Francis, v. 73, n. 364, p. 730–738, 1978. Citado na página 21.

R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2022. Disponível em: <<https://www.R-project.org/>>. Citado na página 49.

RICHARDSON, S.; GREEN, P. J. On bayesian analysis of mixtures with an unknown number of components (with discussion). **Journal of the Royal Statistical Society Series B: Statistical Methodology**, Oxford University Press, v. 59, n. 4, p. 731–792, 1997. Citado nas páginas 32, 34 e 35.

ROBERT, C. P. Simulation of truncated normal variables. **Statistics and Computing**, Springer Science and Business Media LLC, v. 5, n. 2, p. 121–125, jun 1995. Disponível em: <<https://doi.org/10.1007%2Fbf00143942>>. Citado na página 47.

ROSS, S. **Probabilidade: um curso moderno com aplicações**. [S.l.]: Bookman Editora, 2009. Citado na página 28.

SABILLÓN, G. A.; COTRIM, L. G. F.; ZUANETTI, D. A. A data-driven reversible jump for estimating a finite mixture of regression models. **TEST**, Springer, v. 32, n. 1, p. 350–369, 2023. Citado nas páginas 22, 56, 57 e 58.

SAHU, S. K.; DEY, D. K.; BRANCO, M. D. A new class of multivariate skew distributions with applications to bayesian regression models. **Canadian Journal of Statistics**, Wiley Online Library, v. 31, n. 2, p. 129–150, 2003. Citado na página 21.

SARTORI, N. Bias prevention of maximum likelihood estimates for scalar skew normal and skew t distributions. **Journal of Statistical Planning and Inference**, v. 136, n. 12, p. 4259–4275, 2006. ISSN 0378-3758. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0378375805002405>>. Citado na página 52.

WATNIK, M. R. Pay for play: Are baseball salaries based on performance? **Journal of Statistics Education**, Taylor & Francis, v. 6, n. 2, 1998. Citado na página 56.

WEST, M.; ESCOBAR, M. D. Hierarchical priors and mixture models, with application in regression and density estimation. **In Aspects of Uncertainty: A Tribute to DV Lindley**, Institute of Statistics and Decision Sciences, Duke University, p. 363–386, 1994. Citado na página 22.

ZELLER, C. B.; CABRAL, C. R.; LACHOS, V. H. Robust mixture regression modeling based on scale mixtures of skew-normal distributions. **Test**, Springer, v. 25, p. 375–396, 2016. Citado nas páginas 21 e 22.

ZELLER, C. B.; LACHOS, V. H.; VILCA-LABRA, F. E. Local influence analysis for regression models with scale mixtures of skew-normal distributions. **Journal of Applied Statistics**, Taylor & Francis, v. 38, n. 2, p. 343–368, 2011. Citado na página 21.

ZHOU, T. **Linear mixed models with non-normal distributions**. [S.l.]: University of Illinois at Urbana-Champaign, 2005. Citado na página 21.

CONDICIONAIS COMPLETAS PARA O MRST

Alguns detalhes sobre a notação utilizada ao longo das derivações são dados aqui. Em um abuso de notação, denotamos por $\pi(\cdot)$ a função densidade (ou massa) de probabilidade da variável contida em seu argumento, $\pi(\cdot|\cdot)$ denota a função densidade (ou massa) de probabilidade da variável contida no primeiro argumento condicionada ao valor da variável contida em seu segundo argumento. Se temos o vetor $\mathbf{z} = (z_1, z_2, \dots, z_n)^\top$, \mathbf{z}_{-i} denota o vetor \mathbf{z} excluindo o i -ésimo elemento, ou seja, $\mathbf{z}_{-i} = (z_1, z_2, \dots, z_{i-1}, z_{i+1}, \dots, z_n)^\top$, por I_j denotamos o conjunto de índices $\{i: z_i = j\}$, por $\text{Nt}(\cdot|a, b^2)$ denotamos a função densidade da distribuição normal truncada em zero de média a e variância b antes do truncamento e por $\text{Gm}(\cdot|a, b)$ denotamos a função densidade da distribuição gama de parâmetro de forma a e de taxa b . Todas as distribuições aqui mostradas foram obtidas condicionadas ao valor de G .

Utilizamos ... para denotar os demais parâmetros do modelo não explicitamente escritos, por exemplo, $\pi(\mathbf{p}|\mathbf{y}, \mathbf{z}, \dots) \equiv \pi(\mathbf{p}|\mathbf{y}, \mathbf{z}, G, \boldsymbol{\beta}, \boldsymbol{\tau}^2, \boldsymbol{\Delta}, \mathbf{v}, \alpha)$. A omissão das variáveis latentes \mathbf{t} , \mathbf{u} ou \mathbf{z} indica que estamos marginalizando com relação a estas variáveis.

Condicional completa de \mathbf{p}

$$\begin{aligned}
 \pi(\mathbf{p}|\mathbf{y}, \mathbf{z}, \dots) &\propto \pi(\mathbf{y}, \mathbf{z}, \boldsymbol{\theta}) = \pi(\mathbf{y}, \mathbf{z}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}) \\
 &\propto \pi(\mathbf{y}, \mathbf{z}|\boldsymbol{\theta})\pi(\mathbf{p}) \equiv L(\mathbf{y}, \mathbf{z}|G, \boldsymbol{\theta}, \mathbf{X})\pi(\mathbf{p}) \\
 &= \pi(\mathbf{p}) \prod_{j=1}^G p_j^{m_j} \prod_{i=1}^n [\text{St}(y_i|\mathbf{x}_i^\top \boldsymbol{\beta}_j + b\Delta_j, \sigma_j^2, \lambda_j, \mathbf{v})] \mathbb{1}_{(z_i=j)} \\
 &\propto \prod_{j=1}^G p_j^{\gamma_G-1} \prod_{j=1}^G p_j^{m_j} = \prod_{j=1}^G p_j^{m_j+\gamma_G-1}.
 \end{aligned}$$

Note que $\pi(\mathbf{p}|\mathbf{y}, \mathbf{z}, \dots)$ é proporcional ao núcleo de uma distribuição Dirichlet($m_1 + \gamma_G, \dots, m_G + \gamma_G$), como mostrado em (2.53).

Condicionais completa de β_j

$$\begin{aligned}
\pi(\beta_j | \beta_{-j}, \mathbf{y}, \mathbf{z}, \mathbf{t}, \mathbf{u}, \dots) &\propto \pi(\mathbf{y}, \mathbf{z}, \mathbf{t}, \mathbf{u}, \boldsymbol{\theta}) = \pi(\mathbf{y}, \mathbf{z}, \mathbf{t}, \mathbf{u} | \boldsymbol{\theta}) \pi(\boldsymbol{\theta}) \\
&\propto \pi(\mathbf{y}, \mathbf{z}, \mathbf{t}, \mathbf{u} | \boldsymbol{\theta}) \pi(\beta_j) \equiv L(\mathbf{y}, \mathbf{z}, \mathbf{t}, \mathbf{u} | G, \boldsymbol{\theta}, \mathbf{X}) \pi(\beta_j) \\
&\propto \pi(\beta_j) \prod_{I_j} \phi(y_i | \mathbf{x}_i^\top \beta_j + \Delta_j(b + t_i), u_i^{-1} \tau_j^2) \text{Nt}(t_i | 0, u_i^{-1}) \text{Gm}(u_i | \nu/2, \nu/2) \\
&\propto \pi(\beta_j) \prod_{I_j} \exp \left\{ -\frac{u_i}{2\tau_j^2} \left(y_i - \mathbf{x}_i^\top \beta_j - (b + t_i) \Delta_j \right)^2 \right\} \\
&= \pi(\beta_j) \exp \left\{ -\frac{1}{2\tau_j^2} \sum_{I_j} u_i \left(y_i - \mathbf{x}_i^\top \beta_j - (b + t_i) \Delta_j \right)^2 \right\} \\
&= \pi(\beta_j) \exp \left\{ -\frac{1}{2\tau_j^2} [\mathbf{y}_j - (\mathbf{X}_j \beta_j + (b + \mathbf{t}_j) \Delta_j)]^\top U_j [\mathbf{y}_j - (\mathbf{X}_j \beta_j + (b + \mathbf{t}_j) \Delta_j)] \right\} \\
&\propto \exp \left\{ -\frac{1}{2c^2} \beta_j^\top \beta_j \right\} \exp \left\{ -\frac{1}{2\tau_j^2} \left[\beta_j^\top \mathbf{X}_j^\top U_j \mathbf{X}_j \beta_j - 2\beta_j^\top \mathbf{X}_j^\top U_j (\mathbf{y}_j - (b + \mathbf{t}_j) \Delta_j) \right] \right\} \\
&= \exp \left\{ -\frac{1}{2c^2 \tau_j^2} \left[c^2 \beta_j^\top \mathbf{X}_j^\top U_j \mathbf{X}_j \beta_j - 2c^2 \beta_j^\top \mathbf{X}_j^\top U_j (\mathbf{y}_j - (b + \mathbf{t}_j) \Delta_j) + \tau_j^2 \beta_j^\top \beta_j \right] \right\} \\
&= \exp \left\{ -\frac{1}{2} \left[\beta_j^\top \left(\frac{c^2 \mathbf{X}_j^\top U_j \mathbf{X}_j + \tau_j^2 \mathbb{I}_{k+1}}{c^2 \tau_j^2} \right) \beta_j - 2\beta_j^\top \frac{\mathbf{X}_j^\top U_j (\mathbf{y}_j - (b + \mathbf{t}_j) \Delta_j)}{\tau_j^2} \right] \right\}.
\end{aligned}$$

Veja que $\pi(\beta_j | \beta_{-j}, \mathbf{y}, \mathbf{z}, \mathbf{t}, \mathbf{u}, \dots)$ é semelhante ao núcleo de uma normal multivariada, com inversa da matriz de variâncias e covariâncias $\Sigma^{-1} = \frac{c^2 \mathbf{X}_j^\top U_j \mathbf{X}_j + \tau_j^2 \mathbb{I}_{k+1}}{c^2 \tau_j^2}$ e com o produto entre a inversa da matriz de covariâncias e vetor de médias dado por $\Sigma^{-1} \boldsymbol{\mu} = \frac{\mathbf{X}_j^\top U_j (\mathbf{y}_j - (b + \mathbf{t}_j) \Delta_j)}{\tau_j^2}$, o vetor de médias $\boldsymbol{\mu}$ é então dado por $\left(\mathbf{X}_j^\top U_j \mathbf{X}_j + \tau_j^2 / c^2 \mathbb{I}_{k+1} \right)^{-1} \mathbf{X}_j^\top U_j (\mathbf{y}_j - (b + \mathbf{t}_j) \Delta_j)$. Se tomamos $\mathbf{w}_j = \mathbf{X}_j^\top U_j (\mathbf{y}_j - (b + \mathbf{t}_j) \Delta_j)$ e $\Sigma_j = \left(\mathbf{X}_j^\top U_j \mathbf{X}_j + \tau_j^2 / c^2 \mathbb{I}_{k+1} \right)^{-1}$, obtemos o resultado (2.40).

Note que, a matriz de covariâncias desta condicional completa pode não existir caso $\mathbf{X}_j^\top \mathbf{X}_j$ não seja positiva definida, visto que $c, \tau_j^2 > 0$ e \mathbb{I}_{k+1}, U_j são positivas definidas.

Condicionais completa de τ_j^2

$$\begin{aligned}
\pi(\tau_j^2 | \tau_{-j}, \mathbf{y}, \mathbf{z}, \mathbf{t}, \mathbf{u}, \dots) &\propto \pi(\mathbf{y}, \mathbf{z}, \mathbf{t}, \mathbf{u}, \boldsymbol{\theta}) = \pi(\mathbf{y}, \mathbf{z}, \mathbf{t}, \mathbf{u} | \boldsymbol{\theta}) \pi(\boldsymbol{\theta}) \\
&\propto \pi(\mathbf{y}, \mathbf{z}, \mathbf{t}, \mathbf{u} | \boldsymbol{\theta}) \pi(\tau_j^2) \equiv L(\mathbf{y}, \mathbf{z}, \mathbf{t}, \mathbf{u} | G, \boldsymbol{\theta}, \mathbf{X}) \pi(\tau_j^2) \\
&\propto \pi(\tau_j^2) \prod_{i=1}^n [\phi(y_i | \mathbf{x}_i^\top \beta_j + \Delta_j(b + t_i), u_i^{-1} \tau_j^2) \text{Nt}(t_i | 0, u_i^{-1}) \text{Gm}(u_i | \nu/2, \nu/2)]^{\mathbb{1}(z_i=j)} \\
&\propto \pi(\tau_j^2) \prod_{I_j} \phi(y_i | \mathbf{x}_i^\top \beta_j + \Delta_j(b + t_i), u_i^{-1} \tau_j^2) \text{Nt}(t_i | 0, u_i^{-1}) \text{Gm}(u_i | \nu/2, \nu/2)
\end{aligned}$$

$$\begin{aligned}
\pi(\tau_j^2 | \boldsymbol{\tau}_{-j}, \mathbf{y}, \mathbf{z}, \mathbf{t}, \mathbf{u}, \dots) &= \pi(\tau_j^2) \prod_{I_j} (2\pi u_i^{-1} \tau_j^2)^{-1/2} \exp \left\{ -\frac{u_i}{2\tau_j^2} \left(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_j - (b+t_i)\Delta_j \right)^2 \right\} \\
&\propto (\tau_j^2)^{-r-1} e^{-s/\tau_j^2} (\tau_j^2)^{-m_j/2} \exp \left\{ -\left[\frac{1}{2} \sum_{I_j} u_i \left(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_j - (b+t_i)\Delta_j \right)^2 \right] \frac{1}{\tau_j^2} \right\} \\
&= (\tau_j^2)^{-(m_j/2+r)-1} \exp \left\{ -\left[\frac{1}{2} \sum_{I_j} u_i \left(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_j - (b+t_i)\Delta_j \right)^2 + s \right] \frac{1}{\tau_j^2} \right\} \\
&= (\tau_j^2)^{-(m_j/2+r)-1} \exp \left\{ -\left[\frac{S_{3j}}{2} + s \right] \frac{1}{\tau_j^2} \right\}.
\end{aligned}$$

Note que $S_{3j} = \sum_{I_j} u_i (y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_j - (b+t_i)\Delta_j)^2 = (\mathbf{y}_j - \mathbf{X}_j \boldsymbol{\beta}_j - (b+\mathbf{t}_j)\Delta_j)^\top \mathbf{U}_j (\mathbf{y}_j - \mathbf{X}_j \boldsymbol{\beta}_j - (b+\mathbf{t}_j)\Delta_j)$. Observe que $\pi(\tau_j^2 | \boldsymbol{\tau}_{-j}, \mathbf{y}, \mathbf{z}, \mathbf{t}, \mathbf{u}, \dots)$ é proporcional ao núcleo de uma distribuição gama inversa, obtemos assim o resultado (2.41).

Condicional completa de Δ_j

$$\begin{aligned}
\pi(\Delta_j | \boldsymbol{\Delta}_{-j}, \mathbf{y}, \mathbf{z}, \mathbf{t}, \mathbf{u}, \dots) &\propto \pi(\mathbf{y}, \mathbf{z}, \mathbf{t}, \mathbf{u}, \boldsymbol{\theta}) = \pi(\mathbf{y}, \mathbf{z}, \mathbf{t}, \mathbf{u} | \boldsymbol{\theta}) \pi(\boldsymbol{\theta}) \\
&\propto \pi(\mathbf{y}, \mathbf{z}, \mathbf{t}, \mathbf{u} | \boldsymbol{\theta}) \pi(\Delta_j) \equiv L(\mathbf{y}, \mathbf{z}, \mathbf{t}, \mathbf{u} | G, \boldsymbol{\theta}, \mathbf{X}) \pi(\Delta_j) \\
&\propto \pi(\Delta_j) \prod_{I_j} \phi(y_i | \mathbf{x}_i^\top \boldsymbol{\beta}_j + \Delta_j(b+t_i), u_i^{-1} \tau_j^2) \text{Nt}(t_i | 0, u_i^{-1}) \text{Gm}(u_i | \nu/2, \nu/2) \\
&\propto \pi(\Delta_j) \prod_{I_j} \exp \left\{ -\frac{u_i}{2\tau_j^2} \left(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_j - (b+t_i)\Delta_j \right)^2 \right\} \\
&= \pi(\Delta_j) \exp \left\{ -\frac{1}{2\tau_j^2} \sum_{I_j} u_i \left(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_j - (b+t_i)\Delta_j \right)^2 \right\} \\
&\propto \pi(\Delta_j) \exp \left\{ -\frac{1}{2\tau_j^2} \sum_{I_j} u_i \left[(b+t_i)^2 \Delta_j^2 - 2(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_j)(b+t_i)\Delta_j \right] \right\} \\
&\propto \exp \left\{ -\frac{1}{2\omega^2} (\Delta_j^2 - 2\Delta_j \eta) \right\} \exp \left\{ -\frac{1}{2\tau_j^2} [S_{2j} \Delta_j^2 - 2S_{1j} \Delta_j] \right\} \\
&= \exp \left\{ -\frac{1}{2\omega^2 \tau_j^2} [(\tau_j^2 + \omega^2 S_{2j}) \Delta_j^2 - 2(\eta \tau_j^2 + \omega^2 S_{1j}) \Delta_j] \right\} \\
&= \exp \left\{ -\frac{1}{2 \frac{\omega^2 \tau_j^2}{\tau_j^2 + \omega^2 S_{2j}}} \left[\Delta_j^2 - 2 \frac{\eta \tau_j^2 + \omega^2 S_{1j}}{\tau_j^2 + \omega^2 S_{2j}} \Delta_j \right] \right\}.
\end{aligned}$$

Note que $S_{1j} = \sum_{I_j} u_i (y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_j)(b+t_i) = (b+\mathbf{t}_j)^\top \mathbf{U}_j (\mathbf{y}_j - \mathbf{X}_j \boldsymbol{\beta}_j)$ e $S_{2j} = \sum_{I_j} u_i (b+t_i)^2 = (b+\mathbf{t}_j)^\top \mathbf{U}_j (b+\mathbf{t}_j)$. Observe que $\pi(\Delta_j | \boldsymbol{\Delta}_{-j}, \mathbf{y}, \mathbf{z}, \mathbf{t}, \mathbf{u}, \dots)$ é proporcional ao núcleo de uma distribuição normal, temos assim o resultado mostrado em (2.42).

Condicional completa de v

$$\begin{aligned}
\pi(v|\mathbf{y}, \dots) &\propto \pi(\mathbf{y}, \boldsymbol{\theta}) = \pi(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}) \\
&\propto \pi(\mathbf{y}|\boldsymbol{\theta})\pi(v|\alpha)\pi(\alpha) \equiv L(\mathbf{y}|G, \boldsymbol{\theta}, \mathbf{X})\pi(v|\alpha)\pi(\alpha) \\
&\propto L(\mathbf{y}|G, \boldsymbol{\theta}, \mathbf{X})\pi(v|\alpha) \\
&= e^{-\alpha(v-\psi)}\mathbb{1}(v > \psi) \prod_{i=1}^n \sum_{j=1}^G p_j \text{St}(y_i|\mathbf{x}_i^\top \boldsymbol{\beta}_j + b\Delta_j, \sigma_j^2, \lambda_j, v).
\end{aligned}$$

Assim obtemos o resultado (2.47).

Condicional completa de α

$$\begin{aligned}
\pi(\alpha|\mathbf{y}, \dots) &\propto \pi(\mathbf{y}, \boldsymbol{\theta}) = \pi(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}) \\
&\propto \pi(\mathbf{y}|\boldsymbol{\theta})\pi(v|\alpha)\pi(\alpha) \equiv L(\mathbf{y}|G, \boldsymbol{\theta}, \mathbf{X})\pi(v|\alpha)\pi(\alpha) \\
&\propto \pi(v|\alpha)\pi(\alpha) = \alpha e^{-\alpha(v-\psi)}\mathbb{1}(0,02 < \alpha < 0,5).
\end{aligned}$$

Note que $\pi(\alpha|\mathbf{y}, \dots)$ é proporcional ao núcleo de uma distribuição gama de parâmetros 2 e $v - \psi$ truncada no intervalo $(0,02 < \alpha < 0,5)$, nos levando ao resultado (2.43).

Condicional completa de Z_i

$$\begin{aligned}
\pi(Z_i = j|\mathbf{z}_{-i}, \mathbf{y}, \dots) &\propto \pi(\mathbf{y}, \mathbf{z}, \boldsymbol{\theta}) = \pi(\mathbf{y}, \mathbf{z}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}) \\
&\propto \pi(\mathbf{y}, \mathbf{z}|\boldsymbol{\theta}) \equiv L(\mathbf{y}, \mathbf{z}|G, \boldsymbol{\theta}, \mathbf{X}) \\
&= \prod_{j=1}^G p_j^{m_j} \prod_{i=1}^n [\text{St}(y_i|\mathbf{x}_i^\top \boldsymbol{\beta}_j + b\Delta_j, \sigma_j^2, \lambda_j, v)]^{\mathbb{1}(Z_i=j)} \\
&= \prod_{j=1}^G \prod_{i=1}^n [\mathbb{P}(Z_i = j) \text{St}(y_i|\mathbf{x}_i^\top \boldsymbol{\beta}_j + b\Delta_j, \sigma_j^2, \lambda_j, v)]^{\mathbb{1}(Z_i=j)} \\
&\propto \mathbb{P}(Z_i = j) \text{St}(y_i|\mathbf{x}_i^\top \boldsymbol{\beta}_j + b\Delta_j, \sigma_j^2, \lambda_j, v) \\
&= p_j \text{St}(y_i|\mathbf{x}_i^\top \boldsymbol{\beta}_j + b\Delta_j, \sigma_j^2, \lambda_j, v).
\end{aligned}$$

Ou seja, a função de probabilidade condicional de Z_i é dada por

$$\mathbb{P}(Z_i = j|\mathbf{z}_{-i}, \mathbf{y}, \dots) = \frac{p_j \text{St}(y_i|\mathbf{x}_i^\top \boldsymbol{\beta}_j + b\Delta_j, \sigma_j^2, \lambda_j, v)}{\sum_{l=1}^G p_l \text{St}(y_i|\mathbf{x}_i^\top \boldsymbol{\beta}_l + b\Delta_l, \sigma_l^2, \lambda_l, v)} = \rho_j,$$

para $j = 1, 2, \dots, G$. Obtemos assim o resultado (2.44).

Condicional completa de T_i

$$\begin{aligned}
\pi(t_i | \mathbf{t}_{-i}, \mathbf{y}, \mathbf{z}, \mathbf{u}, \dots) &\propto \pi(\mathbf{u}, \mathbf{y}, \mathbf{z}, \mathbf{t}, \boldsymbol{\theta}) = \pi(\mathbf{u}, \mathbf{y}, \mathbf{z}, \mathbf{t} | \boldsymbol{\theta}) \pi(\boldsymbol{\theta}) \\
&\propto \pi(\mathbf{u}, \mathbf{y}, \mathbf{z}, \mathbf{t} | \boldsymbol{\theta}) \equiv L(\mathbf{y}, \mathbf{z}, \mathbf{t}, \mathbf{u} | G, \boldsymbol{\theta}, \mathbf{X}) \\
&= \prod_{j=1}^G p_j^{m_j} \prod_{i=1}^n [\phi(y_i | \mathbf{x}_i^\top \boldsymbol{\beta}_j + \Delta_j(b+t_i), u_i^{-1} \tau_j^2) \text{Nt}(t_i | 0, u_i^{-1}) \text{Gm}(u_i | \nu/2, \nu/2)]^{\mathbb{1}(z_i=j)} \\
&\propto \phi(y_i | \mathbf{x}_i^\top \boldsymbol{\beta}_j + \Delta_j(b+t_i), u_i^{-1} \tau_j^2) \text{Nt}(t_i | 0, u_i^{-1}) \\
&\propto \exp \left\{ -\frac{u_i}{2\tau_j^2} (y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_j - (b+t_i)\Delta_j)^2 \right\} \exp \left\{ -\frac{u_i}{2} t_i^2 \right\} \mathbb{1}(t_i > 0) \\
&= \exp \left\{ -\frac{u_i}{2\tau_j^2} (y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_j - (b+t_i)\Delta_j)^2 - \frac{u_i}{2} t_i^2 \right\} \mathbb{1}(t_i > 0) \\
&\propto \exp \left\{ -\frac{u_i}{2\tau_j^2} \left((b+t_i)^2 \Delta_j^2 - 2(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_j)(b+t_i)\Delta_j \right) - \frac{u_i}{2} t_i^2 \right\} \mathbb{1}(t_i > 0) \\
&\propto \exp \left\{ -\frac{u_i}{2\tau_j^2} \left(\Delta_j^2 t_i^2 + 2bt_i \Delta_j^2 - 2(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_j)\Delta_j t_i \right) - \frac{u_i}{2} t_i^2 \right\} \mathbb{1}(t_i > 0) \\
&= \exp \left\{ -\frac{u_i}{2\tau_j^2} \left(\Delta_j^2 t_i^2 + 2bt_i \Delta_j^2 - 2(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_j)\Delta_j t_i + \tau_j^2 t_i^2 \right) \right\} \mathbb{1}(t_i > 0) \\
&= \exp \left\{ -\frac{u_i}{2\tau_j^2} \left((\Delta_j^2 + \tau_j^2) t_i^2 - 2(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_j - b\Delta_j)\Delta_j t_i \right) \right\} \mathbb{1}(t_i > 0) \\
&= \exp \left\{ -\frac{1}{2 \frac{\tau_j^2}{u_i(\Delta_j^2 + \tau_j^2)}} \left(t_i^2 - 2 \frac{(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_j - b\Delta_j)\Delta_j}{\Delta_j^2 + \tau_j^2} t_i \right) \right\} \mathbb{1}(t_i > 0).
\end{aligned}$$

Note que $\pi(t_i | \mathbf{t}_{-i}, \mathbf{y}, \mathbf{z}, \mathbf{u}, \dots)$ é proporcional ao núcleo de uma distribuição normal truncada em $(0, \infty)$, temos assim o resultado apresentado em (2.46).

Condicional completa de U_i

$$\begin{aligned}
\pi(u_i | \mathbf{u}_{-i}, \mathbf{y}, \mathbf{z}, \mathbf{t}, \dots) &\propto \pi(\mathbf{u}, \mathbf{y}, \mathbf{z}, \mathbf{t}, \boldsymbol{\theta}) = \pi(\mathbf{u}, \mathbf{y}, \mathbf{z}, \mathbf{t} | \boldsymbol{\theta}) \pi(\boldsymbol{\theta}) \\
&\propto \pi(\mathbf{u}, \mathbf{y}, \mathbf{z}, \mathbf{t} | \boldsymbol{\theta}) \equiv L(\mathbf{y}, \mathbf{z}, \mathbf{t}, \mathbf{u} | G, \boldsymbol{\theta}, \mathbf{X}) \\
&= \prod_{j=1}^G p_j^{m_j} \prod_{i=1}^n [\phi(y_i | \mathbf{x}_i^\top \boldsymbol{\beta}_j + \Delta_j(b+t_i), u_i^{-1} \tau_j^2) \text{Nt}(t_i | 0, u_i^{-1}) \text{Gm}(u_i | \nu/2, \nu/2)]^{\mathbb{1}(z_i=j)} \\
&\propto \phi(y_i | \mathbf{x}_i^\top \boldsymbol{\beta}_j + \Delta_j(b+t_i), u_i^{-1} \tau_j^2) \text{Nt}(t_i | 0, u_i^{-1}) \text{Gm}(u_i | \nu/2, \nu/2) \\
&\propto u_i^{1/2} \exp \left\{ -\frac{u_i}{2\tau_j^2} (y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_j - (b+t_i)\Delta_j)^2 \right\} u_i^{1/2} e^{-\frac{u_i}{2} t_i^2} u_i^{\nu/2-1} e^{-\frac{\nu}{2} u_i} \\
&\propto u_i^{\nu/2} \exp \left\{ -\left(\frac{(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_j - (b+t_i)\Delta_j)^2}{2\tau_j^2} + \frac{t_i^2}{2} + \frac{\nu}{2} \right) u_i \right\}.
\end{aligned}$$

Note que $\pi(u_i | \mathbf{u}_{-i}, \mathbf{y}, \mathbf{z}, \mathbf{t}, \dots)$ é proporcional ao núcleo de uma distribuição gama, levando assim ao resultado apresentado em (2.45).

DISTRIBUIÇÕES CONDICIONAIS COMPLETAS PARA O TELESCOPING SAMPLER

Apresentamos detalhes sobre a obtenção das distribuições condicionais completas para um modelo MFM generalizado. Estes detalhes estão disponíveis nos materiais suplementares de [Frühwirth-Schnatter, Malsiner-Walli e Grün \(2021\)](#), alguns comentários do autor foram adicionados. No que segue, ainda utilizamos as mesmas notações apresentadas no primeiro parágrafo do Apêndice A, além disso, por simplicidade, todas as integrais seguintes se referem a integrais no domínio da variável sendo integrada, por exemplo, se temos $x \in [0, 1]$

$$\int f(x)dx \equiv \int_0^1 f(x)dx,$$

para alguma função integrável f .

Inicialmente consideraremos o modelo MFM com uma sequência qualquer $\boldsymbol{\gamma} = \{\gamma_G, G \geq 1\}$ para a *a priori* de \mathbf{p} , lembre-se que temos

$$\begin{aligned} G &\sim \pi_G(\cdot); \\ \mathbf{p}|G &\sim \text{DirichSim}_g(\boldsymbol{\gamma}_G); \\ \boldsymbol{\theta}_j|G &\stackrel{\text{ind}}{\sim} \pi(\cdot|\boldsymbol{\varphi}), j = 1, 2, \dots, G; \\ Z_i|G, \mathbf{p} &\stackrel{\text{ind}}{\sim} \text{Cat}(\mathbf{p}), i = 1, 2, \dots, n; \\ Y_i|G, Z_i = j, \boldsymbol{\theta}_j &\stackrel{\text{ind}}{\sim} F(\cdot|\boldsymbol{\theta}_j), i = 1, 2, \dots, n. \end{aligned}$$

Primeiramente determinamos a distribuição *a priori* da partição aleatória \mathcal{C} , para isto, encontramos a distribuição conjunta das variáveis de alocação $\mathbf{Z} = (Z_1, \dots, Z_n)^\top$ condicionadas

ao número de componentes e aos pesos \mathbf{p} que é dada por

$$\pi(\mathbf{Z}|G, \mathbf{p}) = \prod_{i=1}^n \mathbb{P}(Z_i = z_i | G, \mathbf{p}) = \prod_{i=1}^n p_{z_i} = \prod_{j=1}^G p_j^{m_j},$$

com $m_j = \sum_{i=1}^n \mathbb{1}(z_i = j)$. Encontramos então a distribuição de \mathbf{Z} condicionada apenas ao número de componentes

$$\begin{aligned} \pi(\mathbf{Z}|G, \boldsymbol{\gamma}) &= \int \pi(\mathbf{Z}, \mathbf{p}|G, \boldsymbol{\gamma}_G) d\mathbf{p} = \int \pi(\mathbf{Z}|G, \mathbf{p}) \pi(\mathbf{p}|G, \boldsymbol{\gamma}_G) d\mathbf{p} \\ &= \int \left(\prod_{j=1}^G p_j^{m_j} \right) \left(\frac{\Gamma(G\boldsymbol{\gamma}_G)}{\Gamma^G(\boldsymbol{\gamma}_G)} \prod_{j=1}^G p_j^{\boldsymbol{\gamma}_G - 1} \right) d\mathbf{p} \\ &= \frac{\Gamma(G\boldsymbol{\gamma}_G)}{\Gamma^G(\boldsymbol{\gamma}_G)} \int \prod_{j=1}^G p_j^{m_j + \boldsymbol{\gamma}_G - 1} d\mathbf{p} \\ &= \frac{\Gamma(G\boldsymbol{\gamma}_G)}{\Gamma^G(\boldsymbol{\gamma}_G)} \frac{\prod_{j=1}^G \Gamma(m_j + \boldsymbol{\gamma}_G)}{\Gamma\left(\sum_{j=1}^G (m_j + \boldsymbol{\gamma}_G)\right)} \\ &= \frac{\Gamma(G\boldsymbol{\gamma}_G)}{\Gamma(n + G\boldsymbol{\gamma}_G)} \prod_{j=1}^G \frac{\Gamma(m_j + \boldsymbol{\gamma}_G)}{\Gamma(\boldsymbol{\gamma}_G)}, \end{aligned} \tag{B.1}$$

a solução da integral é obtida utilizando (2.2). Note que (B.1) é invariante a mudança nos rótulos das componentes.

Se definirmos G_+ como o número de componentes preenchidas, ou seja, $G_+ = \sum_{j=1}^G \mathbb{1}\{m_j > 0\}$ e trocamos os rótulos de \mathbf{Z} de modo que as componentes preenchidas sejam as primeiras, \mathbf{Z} define uma partição $\mathcal{C} = \{\mathcal{C}_1, \mathcal{C}_2, \dots, \mathcal{C}_{G_+}\}$ dos índices $\{1, 2, \dots, n\}$ com $\mathcal{C}_j = \{i \in \{1, \dots, n\} : Z_i = j\}$. Podemos encontrar $\pi(\mathcal{C}|G, \boldsymbol{\gamma})$ multiplicando (B.1) pelo número de vetores \mathbf{Z} que levam às mesmas partições.

Temos $\binom{G}{G_+}$ formas de escolhermos G_+ rótulos, após fazermos esta escolha ainda temos $G_+!$ permutações possíveis de tais rótulos que levam a mesma partição. Existem então

$$\binom{G}{G_+} G_+! = \frac{G!}{G_+!(G - G_+)!} G_+! = \frac{G!}{(G - G_+)!}$$

vetores \mathbf{Z} que levam a mesma partição \mathcal{C} . Multiplicando este valor por (B.1) temos

$$\begin{aligned} \pi(\mathcal{C}|G, \boldsymbol{\gamma}) &= \frac{G!}{(G - G_+)!} \frac{\Gamma(G\boldsymbol{\gamma}_G)}{\Gamma(n + G\boldsymbol{\gamma}_G)} \prod_{j=1}^G \frac{\Gamma(m_j + \boldsymbol{\gamma}_G)}{\Gamma(\boldsymbol{\gamma}_G)} \\ &= \frac{G!}{(G - G_+)!} \frac{\Gamma(G\boldsymbol{\gamma}_G)}{\Gamma(n + G\boldsymbol{\gamma}_G)} \prod_{j=1}^{G_+} \frac{\Gamma(m_j + \boldsymbol{\gamma}_G)}{\Gamma(\boldsymbol{\gamma}_G)} \prod_{j=G_++1}^G \frac{\Gamma(m_j + \boldsymbol{\gamma}_G)}{\Gamma(\boldsymbol{\gamma}_G)} \\ &= \frac{G!}{(G - G_+)!} \frac{\Gamma(G\boldsymbol{\gamma}_G)}{\Gamma(n + G\boldsymbol{\gamma}_G)} \prod_{j=1}^{G_+} \frac{\Gamma(m_j + \boldsymbol{\gamma}_G)}{\Gamma(\boldsymbol{\gamma}_G)} \\ &= \frac{G!}{(G - G_+)!} \frac{\Gamma(G\boldsymbol{\gamma}_G)}{\Gamma(n + G\boldsymbol{\gamma}_G)} \prod_{j=1}^{G_+} \frac{\Gamma(m_j + \boldsymbol{\gamma}_G)}{\Gamma(\boldsymbol{\gamma}_G)}, G \geq G_+ \end{aligned} \tag{B.2}$$

note que $m_j = 0$ para $j \geq G_+ + 1$. Obtemos então a condicional completa (2.51) ao escolhermos a sequência $\boldsymbol{\gamma} = \{\gamma/G, G \geq 1\}$ e ao observarmos que

$$\begin{aligned}
\pi(G|\mathcal{C}, \boldsymbol{\gamma}) &\propto \pi_G(G)\pi(\mathcal{C}|G, \boldsymbol{\gamma}) \\
&= \pi_G(G) \frac{G!}{(G - G_+)!} \frac{\Gamma(\boldsymbol{\gamma})}{\Gamma(n + \boldsymbol{\gamma})} \prod_{j=1}^{G_+} \frac{\Gamma(m_j + \boldsymbol{\gamma}/G)}{\Gamma(\boldsymbol{\gamma}/G)} \\
&\propto \pi_G(G) \frac{G!}{(G - G_+)!} \prod_{j=1}^{G_+} \frac{\Gamma(m_j + \boldsymbol{\gamma}/G)}{\Gamma(\boldsymbol{\gamma}/G)} \\
&\propto \pi_G(G) \frac{G!}{(G - G_+)!} \prod_{j=1}^{G_+} \frac{\Gamma(m_j + \boldsymbol{\gamma}/G)}{G/\boldsymbol{\gamma}\Gamma(1 + \boldsymbol{\gamma}/G)} \\
&= \pi_G(G) \frac{\boldsymbol{\gamma}^{G_+} G!}{G^{G_+} (G - G_+)!} \prod_{j=1}^{G_+} \frac{\Gamma(m_j + \boldsymbol{\gamma}/G)}{\Gamma(1 + \boldsymbol{\gamma}/G)}.
\end{aligned}$$

Se uma distribuição a priori $\pi(\boldsymbol{\gamma})$ for atribuída ao parâmetro $\boldsymbol{\gamma}$ da sequência $\boldsymbol{\gamma}$, sua condicional completa pode também ser encontrada por meio de (B.2)

$$\begin{aligned}
\pi(\boldsymbol{\gamma}|\mathcal{C}, G) &\propto \pi(\boldsymbol{\gamma})\pi(\mathcal{C}|G, \boldsymbol{\gamma}) \\
&= \pi(\boldsymbol{\gamma}) \frac{G!}{(G - G_+)!} \frac{\Gamma(\boldsymbol{\gamma})}{\Gamma(n + \boldsymbol{\gamma})} \prod_{j=1}^{G_+} \frac{\Gamma(m_j + \boldsymbol{\gamma}/G)}{\Gamma(\boldsymbol{\gamma}/G)} \\
&\propto \pi(\boldsymbol{\gamma}) \frac{\Gamma(\boldsymbol{\gamma})}{\Gamma(n + \boldsymbol{\gamma})} \prod_{j=1}^{G_+} \frac{\Gamma(m_j + \boldsymbol{\gamma}/G)}{\Gamma(\boldsymbol{\gamma}/G)} \\
&\propto \pi(\boldsymbol{\gamma}) \frac{\boldsymbol{\gamma}^{G_+} \Gamma(\boldsymbol{\gamma})}{\Gamma(n + \boldsymbol{\gamma})} \prod_{j=1}^{G_+} \frac{\Gamma(m_j + \boldsymbol{\gamma}/G)}{\Gamma(1 + \boldsymbol{\gamma}/G)},
\end{aligned}$$

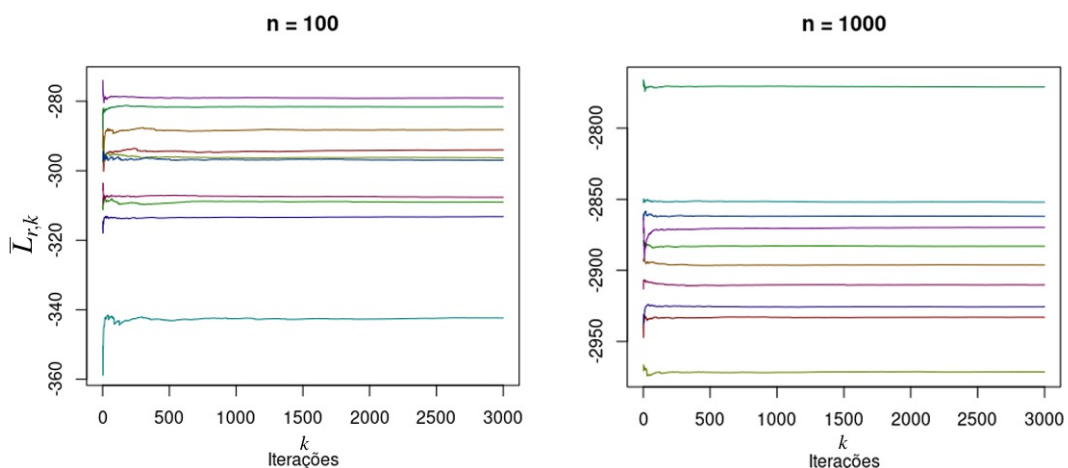
assim obtemos a condicional completa (2.52). As demais distribuições utilizadas foram apresentadas no Apêndice A.

VERIFICAÇÕES DE CONVERGÊNCIA DAS CADEIAS MCMC

A verificação de convergência das cadeias foi realizada utilizando o logaritmo da função de verossimilhança do modelo MRST apresentada em (2.36). Para a r -ésima replicação do algoritmo e q -ésima amostra MCMC calculamos $L_{r,q} = \log(L(\mathbf{y}_r, \mathbf{z}_r^{(q)} | G_r^{(q)}, \boldsymbol{\theta}_r^{(q)}, \mathbf{X}_r))$, em que \mathbf{y}_r e \mathbf{X}_r representam o vetor de respostas e matriz de planejamento considerados na r -ésima replicação do algoritmo, respectivamente, e $\mathbf{z}_r^{(q)}$, $G_r^{(q)}$ e $\boldsymbol{\theta}_r^{(q)}$ representam o q -ésimo valor da cadeia de Markov dos vetores de alocação, número de componentes e parâmetros do modelo na r -ésima replicação do algoritmo.

A Figura 9 mostra a média ergódica por réplica $\bar{L}_{r,k} = k^{-1} \sum_{q=1}^k L_{r,q}$ para os dois tamanhos amostrais considerados no estudo de simulação do Capítulo 3, diferentes cores representam diferentes replicações. Observamos um comportamento convergente em todas as replicações

Figura 9 – Média ergódica por réplica para os tamanhos amostrais $n = 100$ e $n = 1000$ nos dados simulados.

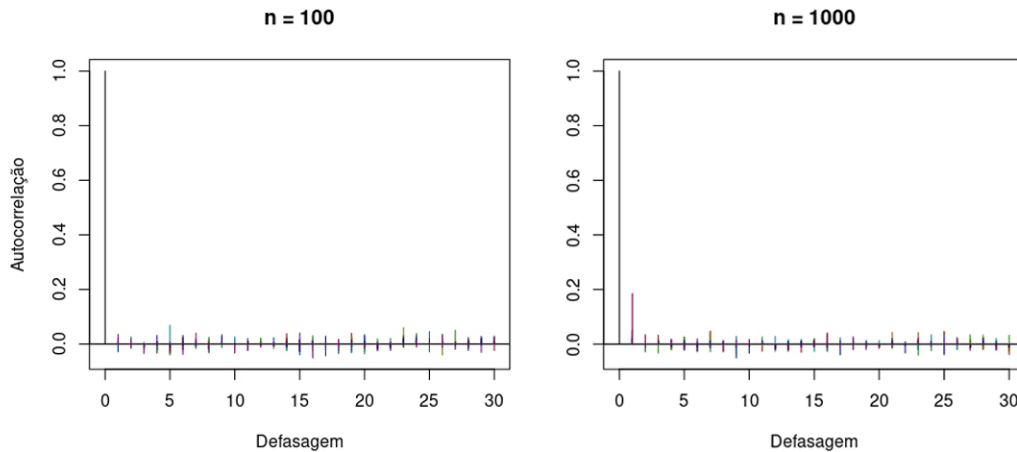


Fonte: Elaborada pelo autor.

consideradas, visto que em poucas iterações é possível observar um comportamento constante na média ergódica.

A Figura 10 mostra a autocorrelação da cadeia $\{L_{r,1}, \dots, L_{r,3000}\}$ para os dois tamanhos amostrais considerados no estudo de simulação do Capítulo 3, diferentes cores representam diferentes replicações. Observamos baixa autocorrelação nas cadeias em todas as replicações.

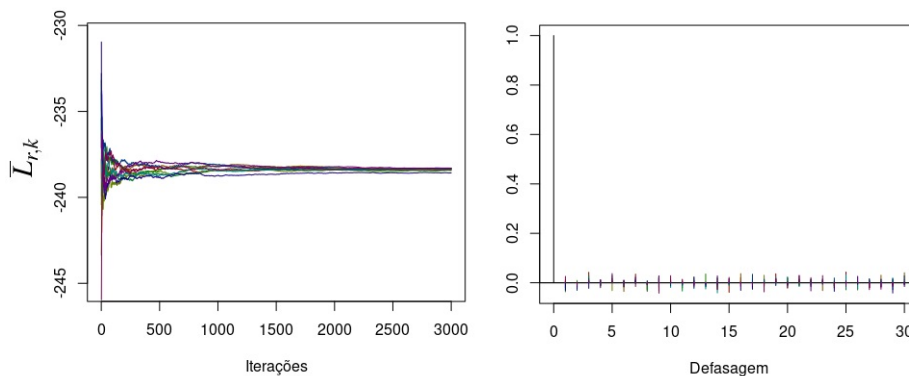
Figura 10 – Autocorrelação da cadeia por réplica para os tamanhos amostrais $n = 100$ e $n = 1000$ nos dados simulados.



Fonte: Elaborada pelo autor.

A Figura 11 mostra a média ergódica por réplica e a autocorrelação da cadeia de Markov $\{L_{r,1}, \dots, L_{r,3000}\}$ na aplicação do modelo aos dados de salários de jogadores de baseball apresentada no Capítulo 3, diferentes cores representam diferentes replicações. Relembre-se que em cada réplica do algoritmo utilizamos o mesmo conjunto de dados. Observamos um

Figura 11 – Média ergódica e autocorrelação da cadeia por réplica, respectivamente, para os dados dos salários de jogadores de baseball.



Fonte: Elaborada pelo autor.

comportamento convergente em todas as replicações consideradas, visto que em poucas iterações é possível observar um comportamento constante na média ergódica. Note ainda que todas as

médias ergódicas convergem para o mesmo valor. Observamos baixa autocorrelação nas cadeias em todas as replicações.

