UNIVERSIDADE FEDERAL DE JUIZ DE FORA
INSTITUTO DE CIÊNCIAS EXATAS
GRADUAÇÃO DE ESTATÍSTICA

Pedro Henrique Corrêa de Almeida

Misturas Finitas de Especialistas de Modelos de Regressão para Dados Complexos

Juiz de Fora

2025

Pedro Henrique Corrêa de Almeida

Misturas Finitas de Especialistas de Modelos de Regressão para Dados Complexos

Monografia submetida ao corpo docente do Instituto de Ciências Exatas da UFJF, com parte integrante dos requisitos necessários para obtenção do grau de bacharel em estatística.

Orientador: Profa. Dr. Camila Borelli Zeller

Juiz de Fora

2025

Ficha catalográfica elaborada através do Modelo Latex do CDC da UFJF
com os dados fornecidos pelo(a) autor(a)

Corrêa de Almeida, Pedro Henrique.
Misturas Finitas de Especialistas de Modelos de Regressão para Dados Complexos / Pedro Henrique Corrêa de Almeida. – 2025.
55 f. : il.

Orientador: Camila Borelli Zeller
Conclusão de Curso (graduação) – Universidade Federal de Juiz de Fora, Instituto de Ciências Exatas. Graduação de Estatística, 2025.

1. algoritmo EM. 2. caudas pesadas. 3. clusterização. 4. distribuições assimétricas. 5. misturas de especialistas. 6. modelos de regressão. 7. misturas finitas. I. Zeller, Camila Borelli, orient. II. Misturas Finitas de Especialistas de Modelos de Regressão para Dados Complexos.

Pedro Henrique Corrêa de Almeida

Misturas Finitas de Especialistas de Modelos de Regressão para Dados Complexos

Monografia submetida ao corpo docente do Instituto de Ciências Exatas da UFJF, com parte integrante dos requisitos necessários para obtenção do grau de bacharel em estatística.

Aprovada em (dia) de (mês) de (ano)

BANCA EXAMINADORA

Profa. Dr. Camila Borelli Zeller - Orientador
Universidade Federal de Juiz de Fora

Prof. Dr. Lupércio França Bessegato
Universidade Federal de Juiz de Fora

Prof. Dr. Clécio da Silva Ferreira
Universidade Federal de Juiz de Fora

AGRADECIMENTOS

Primeiramente agradeço a toda minha família que sempre me apoiou em tudo que fiz. Em especial aos meus pais Marcia e Vagner, que sempre me estimularam em toda a minha vida me dando todo o suporte mais do que o necessário. Aos meus irmãos Alice e João, que sempre estiveram ao meu lado me ajudando nessa trajetória. Agredeço a minha namorada Laura, que está sempre comigo me dando forças, me incentivando e me apoiando emocionalmente.

Agradeço a Universidade Federal de Juiz de Fora por me proporcionar viver esses anos de muito aprendizado. A todos os professores com quem tive aula, principalmente os professores do departamento de estatística. Em especial a Dr. Camila Zeller pelas suas aulas excepcionais, por me orientar na iniciação científica e neste trabalho, e por todos os aprendizados que tive com ela. Agradeço ao Dr. Lupércio Bessegato por todos os ensinamentos, risadas, conselhos e ajudas como professor e coordenador, e ao Dr. Clécio por todas suas aulas, seus conselhos e seu empenho em organizar enriquecedores encontros com profissionais e pesquisadores.

Agradeço a todos meus colegas e amigos que fiz dentro da faculdade, especialmente ao Gustavo, Arthur, Camila, Joysce e Natasha, por toda a união que tivemos durante esses anos, sempre nos ajudando e agregando um ao outro.

RESUMO

Os modelos de regressão tradicionais partem da premissa de que todos os indivíduos apresentam a mesma relação entre as variáveis explicativa e resposta, o que, no entanto, não se aplica a populações heterogêneas. Nesse contexto, as misturas finitas de modelos de regressão têm como objetivo identificar grupos não observáveis que se comportam de maneira distinta, permitindo a abordagem da heterogeneidade. Este trabalho explora o campo das misturas finitas de regressão, com ênfase em distribuições assimétricas e/ou de caudas pesadas, o que possibilita a construção de modelos robustos para dados complexo. A inovação proposta neste estudo reside na inclusão de informações que aprimorem a classificação e auxiliem na interpretação dos grupos identificados, denominadas misturas de especialistas. Os modelos propostos foram representados hierarquicamente, juntamente com a função de verossimilhança correspondente, permitindo a aplicação do Algoritmo EM (Expectativa-Maximização) na obtenção dos estimadores de máxima verossimilhança. Estudos de simulação foram conduzidos com o intuito de avaliar a eficácia do algoritmo na recuperação dos parâmetros. Por fim, o modelo proposto foi aplicado a dois conjuntos de dados reais, com o objetivo de demonstrar sua aplicação prática e a interpretação dos resultados obtidos.

Palavras-chave: algoritmo EM; caudas pesadas; clusterização; distribuições assimétricas; misturas de especialistas; modelos de regressão; misturas finitas.

ABSTRACT

Traditional regression models start from the premise that all individuals have the same relationship between the explanatory and response variables, but this does not apply to heterogeneous populations. In this context, finite mixtures of regression models aim to identify unobservable groups that behave differently, allowing heterogeneity to be addressed. This paper explores the field of finite regression mixtures, with an emphasis on asymmetric and/or heavy-tailed distributions, which make it possible to build robust models for complex data. The innovation proposed in this study lies in the inclusion of information that improves classification and helps interpret the groups identified, known as expert mixtures. The proposed models were represented hierarchically, along with the corresponding likelihood function, allowing the EM (Expectation-Maximization) algorithm to be applied to obtain the maximum likelihood estimators. Simulation studies were carried out to assess the effectiveness of the algorithm in recovering the parameters. Finally, the proposed model was applied to two sets of real data in order to demonstrate its practical application and the interpretation of the results obtained.

Keywords: EM algorithm; heavy tails; clustering; asymmetric distributions; expert mixtures; regression models; finite mixtures.

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

BIC	Critério de Informação Bayesiano
CML	Máxima verossimilhança condicional
ECM	Expectativa-maximização condicional
ECME	Expectativa-maximização condicional alternado
EM	Expectativa-maximização
HN	Half-normal
MF	Misturas finitas de densidades
MoE	Misturas de especialistas
MEN	Mistura de escala normal
MESN	Mistura de escala skew-normal
SN	Skew-normal
ST	Skew-t

SUMÁRIO

1 INTRODUÇÃO

Ao longo dos anos, surgiram diversas técnicas de clusterização, que podem ser classificadas em diferentes tipos. Conforme descrito por Hastie, Tibshirani e Friedman (2013), essas técnicas podem ser divididas em três categorias: algoritmos combinatórios, modelos paramétricos e algoritmos, "bump hunters". Os algoritmos combinatórios operam diretamente sobre os dados, sem se basear em um modelo subjacente. Exemplos comuns são o K-means e os métodos hierárquicos. Já a clusterização baseada em modelos paramétricos parte do pressuposto de que os dados são uma amostra de uma população descrita por um modelo paramétrico, como é o caso dos modelos de misturas finitas de densidades, que são o foco deste trabalho. Por outro lado, os modelos "bump hunters"utilizam abordagens não paramétricas, sendo mais intuitivos na detecção de clusters, pois esses modelos exploram os dados para identificar regiões densas que podem indicar a presença de clusters.

No contexto de análise de agrupamento, a heterogeneidade não observada refere-se à existência de subgrupos ou subpopulações dentro de um conjunto de dados que não podem ser identificados diretamente a partir das variáveis observadas. Em outras palavras, os dados parecem vir de uma única distribuição, mas na realidade são compostos por várias distribuições subjacentes, cada uma representando um grupo diferente, com características distintas. Esses subgrupos podem ter padrões próprios de comportamento (como diferentes médias, variâncias ou formas de distribuição), mas como não são explicitamente observáveis nas variáveis de entrada, o desafio é identificar e modelar essa estrutura oculta. A multimodalidade é um dos sinais de que há heterogeneidade não observada, pois diferentes picos na distribuição dos dados indicam a presença de múltiplas subpopulações.

Os modelos de misturas finitas são frequentemente usados para lidar com esse tipo de situação, permitindo modelar os dados como uma combinação de várias distribuições, cada uma correspondendo a um subgrupo. Assim, é possível capturar a complexidade e a variabilidade dos dados, mesmo quando os subgrupos não são diretamente observáveis. McLachlan e Basford (1988) desempenharam um papel fundamental ao introduzir, de maneira abrangente, os modelos de misturas finitas. Eles discutiram as motivações e aplicações desses modelos para abordar o problema da heterogeneidade não observada. Já Veaux (1989) avançou essa abordagem ao combinar modelos de misturas finitas com modelos de regressão linear, assumindo normalidade, oferecendo assim uma ferramenta poderosa para a análise de dados no contexto de regressão.

No contexto da análise de regressão, diversas pesquisas têm se dedicado ao desenvolvimento de modelos paramétricos flexíveis que sejam capazes de acomodar a multimodalidade e desvios da suposição de normalidade. Esses modelos oferecem uma abordagem mais robusta para lidar com a complexidade dos dados, permitindo uma melhor adequação quando as distribuições normais não são suficientes para capturar todas as características

presentes como assimetria e/ou caudas pesadas. Os trabalhos de Shoham (2002) e McLachlan e Peel (2000) investigam as misturas finitas *t-Student*, que adicionam o parâmetro para lidar com caudas pesadas na modelagem. Azzalini (1985) introduziu a distribuição *skew-normal*, capaz de modelar a assimetria nos dados por meio do parâmetro λ, e essa distribuição foi aplicada em misturas finitas por Lin, Lee e Yen (2007). Posteriormente, Lin, Lee e Hsieh (2007) propuseram as misturas finitas de modelos de regressão *skew-t*, que conseguem capturar simultaneamente assimetria e caudas pesadas.

Estudos recentes têm explorado amplamente as aplicações e extensões de modelos de mistura. Prates, Lachos e Cabral (2013) introduzem o pacote *mixsmsn* para o ajuste de modelos de mistura com distribuições assimétricas no software R Core Team (2024). Lee e McLachlan (2014) investigam modelos que lidam com dados de caudas pesadas e assimetrias, com enfoque nas distribuições skew-t. Alamichel e Smith (2023) apresentam modelos bayesianos de misturas finitas com prioris assimétricas, enquanto Williams e Tanaka (2020) se concentram em modelos de mistura robustos, particularmente para dados de alta dimensionalidade. Morrison e Walters (2022) examinam a aplicação de modelos de misturas finitas no campo da econometria, e Liu e O'Brien (2021) propõem modelos de mistura esparsos voltados para a identificação dos componentes mais relevantes. Dávila, Cabral e Zeller (2018) exploram os casos univariado e multivariado de misturas finitas assimétricas aplicados no contexto de regressão. Adicionalmente, Xiang et al. (2024) apresentaram uma revisão abrangente sobre métodos e aplicações de modelos de mistura paramétricos, semi-paramétricos e novas direções nesta área.

Neste trabalho, o foco será o estudo das misturas finitas de especialistas, uma classe de modelos que tem ganhado destaque recentemente por aprimorar a classificação de grupos e, ao mesmo tempo, oferecer maior interpretabilidade em relação aos modelos tradicionais de misturas finitas. Jacobs et al. (1991) introduziram essa classe, que se baseia em uma estrutura composta por dois elementos principais: os *gatings* (portas) e os *experts* (especialistas). Os *gatings* representam as proporções de cada grupo, enquanto os *experts* correspondem às componentes da mistura, ou seja, os modelos que estimam a densidade de cada observação. Este trabalho representa a aplicação e extensão das técnicas aprendidas durante o curso de graduação em Estatística na Universidade Federal de Juiz de Fora (UFJF). Durante o desenvolvimento da monografia, foram aplicados e aprofundados os conhecimentos adquiridos ao longo da graduação, complementados pela experiência obtida por meio da participação em projetos de iniciação científica.

1.1 OBJETIVO

Este trabalho tem como objetivo estudar e aplicar os modelos de misturas finitas de especialistas, no contexto de modelos de regressão, que são uma extensão dos modelos tradicionais de misturas finitas de densidades. Esses modelos foram estudados por Jacobs

et al. (1991), por exemplo, e se destacam por oferecer uma solução flexível para lidar com a heterogeneidade dos dados reais, assumindo que tanto os pesos das misturas quanto as distribuições dos componentes correspondentes aos modelos de regressão dependem de covariáveis.

Este trabalho utiliza como base as distribuições da classe misturas escala *skew-normal*, propostas por Branco e Dey (2001). Essa escolha se deve à necessidade de lidar com problemas de assimetria e/ou caudas pesadas, comuns em dados. Assim, os modelos de misturas de especialistas baseados nessa classe de distribuições surgem como uma ferramenta eficaz para a modelagem de dados assimétricos/simétricos e heterogêneos provenientes de diversas áreas como econometria, biologia, genética, engenharia, marketing, biometria, dentre outras.

Os modelos de misturas finitas de especialistas de regressão, que utilizam distribuições assimétricas, são particularmente valiosos porque proporcionam uma modelagem flexível capaz de capturar a complexidade dos dados. Isso inclui características como multimodalidade, assimetria, caudas pesadas e heterogeneidade não observada. Além disso, esses modelos oferecem uma vantagem importante: a interpretabilidade, permitindo que os resultados sejam mais facilmente compreendidos. Esses modelos estabelecem uma relação clara entre as características dos indivíduos (também chamadas de covariáveis) e o grupo ao qual pertencem, o que é muito útil na classificação não supervisionada. Finalmente, neste trabalho serão consideradas questões relacionadas à clusterização (classificação não supervisionada) de observações, no sentido que cada indivíduo na população pertence a um de G grupos, mas não sabemos especificar a qual grupo o indivíduo pertence. Em uma análise de agrupamento, é comum que este seja implementado em variáveis de interesse sem referência às informações de covariáveis concomitantes sobre os indivíduos (ou objetos) que estão sendo agrupados. O uso de uma abordagem baseada em modelos de misturas finitas para agrupamento permite que qualquer incerteza seja contabilizada em uma estrutura probabilística, e adicionalmente, no contexto de especialistas permite a inclusão de informações de covariáveis.

1.2 ORGANIZAÇÂO

Este trabalho está organizado em cinco capítulos. No Capítulo 2, são apresentados os conceitos fundamentais utilizados ao longo do estudo, incluindo as misturas finitas de densidades, conforme Pearson (1894), suas extensões para o contexto de modelos de regressão, a classe de distribuições misturas de escala *skew-normal* (BRANCO; DEY, 2001), o algoritmo EM conforme Dempster, Laird e Rubin (1977), e o critério de informação Bayesiano (BIC).

O Capítulo 3 aborda os resultados relacionados aos modelos de misturas finitas de densidades, incluindo algumas propriedades desses modelos, a estrutura de dados

incompletos e a construção do algoritmo EM nesse cenário. O quarto capítulo pode ser considerado como o objetivo principal desse trabalho. O Capítulo 4 inicia-se com uma discussão sobre resultados referentes aos modelos de misturas finitas de especialistas de modelos de regressão, onde são apresentadas definições e propriedades acerca desses modelos, bem como a estrutura de dados incompletos e a construção do algoritmo EM nesse contexto.

Os Capítulos 5 e 6 apresentam, respectivamente, os estudos de simulação e as aplicações práticas dos modelos discutidos no Capítulo 4, com o objetivo de mostrar os resultados alcançados, além de discutir os desafios encontrados e possíveis melhorias. O Capítulo 7 conclui o trabalho, resumindo os objetivos alcançados e sugerindo os próximos passos. Por fim, o Capítulo 8 comenta sobre os aspectos computacionais deste trabalho.

2 CONCEITOS INICIAIS

Este capítulo tem como objetivo introduzir os conceitos fundamentais que embasam os modelos de misturas de especialistas, foco principal deste trabalho. Serão apresentados e definidos, de maneira clara e precisa, as misturas finitas de densidades, as distribuições assimétricas Misturas de Escala *skew-normal*, o algoritmo EM e o critério de informação Bayesiano (BIC), que constituem a base teórica para o desenvolvimento dos modelos explorados.

2.1 MISTURAS FINITAS DE DENSIDADES

Nesta seção, introduzimos os modelos de misturas finitas de densidades, que são uma ferramenta poderosa para representar estruturas subjacentes em distribuições complexas. Esses modelos são amplamente aplicados em diversas áreas da Estatística, como análise de agrupamento, análise discriminante e análise de regressão.

Para construir um modelo de misturas finitas , é necessário definir antecipadamente o número de grupos (componentes), representado por G, e escolher uma função de densidade para cada grupo, denotada por $g(.)$.

Os modelos de misturas finitas foram inicialmente propostos por Pearson (1894), com o objetivo de estudar subpopulações de caranguejos. Contudo, as limitações computacionais da época dificultaram seu uso. Somente com o desenvolvimento do algoritmo EM, apresentado por Dempster, Laird e Rubin (1977), esses modelos se tornaram mais acessíveis, pois o algoritmo simplificou o cálculo dos estimadores de máxima verossimilhança em modelos complexos, como as misturas finitas de densidade.

Neste trabalho, utilizaremos as misturas finitas de densidades para a estimação de modelos de regressão. A seguir, discutiremos as propriedades desses modelos que possibilitam a estimação por máxima verossimilhança.

Definição 2.1.1. A variável aleatória Y segue um modelo de misturas finitas de densidades com G-componentes se sua função de densidade é dada por

$$f(y) = \sum_{j=1}^{G} \pi_j g_j(y), \pi_j > 0 \text{ e } \sum_{j=1}^{G} \pi_j = 1, \tag{2.1}$$

onde $f(.)$ é a função de misturas de densidades, π_1, ..., π_G são as proporções, parâmetros da distribuição, e g_1, ..., g_G são as componetes das misturas, que dependem da suposição de distribuição adotada no modelo. A depender da distribuição específica, diferentes parâmetros serão estimados para cada componente.

2.1.1 Representação Estocástica

Visando a facilitação do processo de estimação, vamos definir a representação estocástica do modelo de misturas finitas de densidades, dado que essa abordagem será essencial para a obtenção dos parâmetros de máxima verossimilhança.

Como é comum na estrutura do algoritmo EM para misturas finitas de distribuições, introduzimos o vetor latente $\mathbf{Z}_i = (Z_{i1}, \ldots, Z_{iG})^\top \sim \text{Multinomial}(1|\pi_1, \ldots, \pi_G)$, *para* $i = 1, \ldots, n$ onde a coordenada j é uma variável aleatória binária definida como:

$$Z_{ij} = \begin{cases} 1, \text{se o i-ésimo indivíduo é pertencente ao j-ésimo componente} \\ 0, \text{caso contrário} \end{cases} \tag{2.2}$$

Dessa forma, $P(Z_{ij} = 1) = \pi_j$.

Proposição 2.1.1. *Uma variável aleatória Y segue um modelo de misturas finitas de densidades se sua representação estocástica é dada por*

$$Y|Z_j = 1 \sim D_j, \tag{2.3}$$

onde D_j é a distribuição da componente j.

2.2 MISTURAS FINITAS DE MODELOS DE REGRESSÃO

Na seção anterior, discutimos os modelos de misturas finitas de densidades. No entanto, em muitos contextos práticos, é desejável incorporar o efeito de covariáveis no modelo, permitindo que as distribuições variem conforme variáveis explicativas. Com essa motivação, é possível estender os modelos de misturas finitas para incluir componentes de regressão, nos quais a locação da distribuição da variável resposta Y é dependente de covariáveis x's. Essa extensão permite o uso de modelos de regressão linear em cada componente da mistura, enriquecendo a modelagem ao capturar as relações entre Y e x's.

Com base nessa ideia, podemos estender os modelos de misturas finitas para incluir componentes de regressão, onde a locação da distribuição da variável resposta Y é condicionada pelas covariáveis x's. Isso nos permite utilizar modelos de regressão linear para cada componente da mistura, o que enriquece a modelagem ao capturar relações entre Y e x's.

Definição 2.2.1. Uma variável aleatória Y segue um modelo de misturas finitas de regressão com G-componentes se sua função de densidade é dada por:

$$f(y) = \sum_{j=1}^{G} \pi_j g_j(y|\boldsymbol{X}\boldsymbol{\beta}_j, \boldsymbol{\theta}_j), \text{ onde } \pi_j > 0 \text{ e } \sum_{j=1}^{G} \pi_j = 1, \tag{2.4}$$

onde $\boldsymbol{\beta_j}$ são os coeficientes da regressão, $\boldsymbol{X}$ é a matriz de covariáveis e $\boldsymbol{\theta}_j$ representa o conjunto dos demais parâmetros, a depender da distribuição específica.

2.3 DISTRIBUIÇÕES DE MISTURA ESCALA *skew-normal*

Nesta seção, apresentamos a classe de distribuições assimétricas de misturas de escala *skew-normal* (MESN), proposta por Branco e Dey (2001). Essa família de distribuições é uma extensão da classe de distribuições simétricas de misturas de escala normal (MEN), originalmente discutida por Andrews e Mallows (1974). A classe MESN combina a capacidade de modelar caudas pesadas com a de capturar assimetrias, oferecendo uma abordagem mais flexível para a modelagem de dados complexos.

Definição 2.3.1. Seja Y uma variável aleatória com função de densidade de probabilidade dada por

$$\psi(y|\mu,\sigma^2,\lambda,\nu) = 2\int_0^\infty \phi(y|\mu,u\sigma^2)\Phi(u^{-1/2}A)dH(u|\boldsymbol{\nu}), y\in\mathbb{R},\ \mu\in\mathbb{R},\ \sigma>0. \tag{2.5}$$

Dizemos que a variável aleatória Y segue uma distribuição na classe de misturas de escala skew-normal, com parâmetro de locação $\mu \in R$, escala $\sigma^2 > 0$ e assimetria $\lambda \in R$. Aqui, $A = \lambda(y - \sigma\mu)$, U é um fator de escala (variável aleatória positiva) com função de distribuição acumulada $H(.|\boldsymbol{\nu})$. Utilizamos a notação $Y \sim \text{MESN}(\mu, \sigma^2, \lambda; H)$ para nos referirmos às misturas de escala *skew-normal*.

2.3.1 Casos particulares

A classe de misturas de escala *skew-normal* abrange várias distribuições que são reconhecidas como casos particulares. Entre elas, destacam-se:

1. *skew-normal*, quando U é um variável degenerada em 1.
2. *skew-t*, quando $U \sim Gama(\frac{\nu}{2}, \frac{\nu}{2})$.
3. *skew*-slash, quando $U \sim Beta(\nu, 1)$.
4. *skew-normal* contaminada, quando U é uma variável aletória discreta, que assume o valor γ com probabilidade ν, e o valor 1 com probablidade $1 - \nu$.

No entanto, apenas as distribuições *skew-normal* e *skew-t* foram utilizadas neste trabalho, nesse sentido vamos introduzi-las.

Definição 2.3.2. Seja $Y \sim SN(\mu, \sigma^2, \lambda)$ uma variável aleatória com função de densidade de probabilidade dada por

$$g(y|\mu,\sigma^2,\lambda) = 2\varphi(y|\mu,\sigma^2)\Phi\left(\lambda\frac{y-\mu}{\sigma}\right),\ y\in\mathbb{R},\ \mu\in\mathbb{R},\ \sigma>0,\ \lambda\in\mathbb{R}, \tag{2.6}$$

onde μ é o parâmetro de locação, σ^2 é o parâmetro de escala, e λ é o parâmetro de assimetria. Valores positivos de λ indicam assimetria positiva, enquanto valores negativos

indicam assimetria negativa. Além disso, quanto maior o valor absoluto de λ, mais acentuada será a assimetria. Adicionalmente, quando $\lambda = 0$, a distribuição skew-normal se reduz à distribuição normal simétrica clássica, ou seja, $Y \sim \mathrm{N}(\mu, \sigma^2)$.

Definição 2.3.3. Seja $Y \sim ST(\mu, \sigma^2, \lambda, \nu)$ uma variável aleatória com função de densidade de probabilidade dada por

$$g(y|\mu, \sigma^2, \lambda, \nu) = 2t(y|\mu, \sigma^2, \nu)\, T\left(\sqrt{\frac{\nu+1}{\nu+d}}A\right), \quad y \in \mathbb{R},\ \mu \in \mathbb{R},\ \sigma > 0,\ \nu > 0, \tag{2.7}$$

onde $t(.)$ e $T(.)$ são as funções de densidade e de distribuição acumulada da *t-Student*, e $d = \frac{(y-\mu)^2}{\sigma^2}$ representa a distância de Mahalanobis. Quanto aos parâmetros, μ é o parâmetro de locação, σ^2 é o parâmetro de escala, e λ é o parâmetro de assimetria, assim como na distribuição *skew-normal*. O parâmetro ν controla o peso das caudas, sendo que valores baixos de ν produzem caudas mais pesadas, enquanto valores altos aproximam a distribuição de uma *skew-normal*. Quando $\lambda = 0$, a distribuição *skew-t* se reduz à distribuição *t-Student* simétrica clássica, ou seja, $Y \sim t(\mu, \sigma^2, \nu)$.

2.3.2 Representação Estocástica

A fim de utilizar o Algoritmo EM na estimação dos parâmetros de uma distribuição da classe MESN, vamos definir sua representação estocástica; Dávila, Cabral e Zeller (2018).

Definição 2.3.4. Uma variável aleatória Y é considerada pertencente à família de misturas de escala *skew-normal* se puder ser expressa como

$$Y = \mu + U^{-1/2}Z,\ \mu \in \mathbb{R}, \tag{2.8}$$

onde $Z \sim SN(0, \sigma^2, \lambda)$ e U é um fator de escala independente de Z, tal que Z segue uma distribuição $\mathrm{SN}(0, \sigma^2, \lambda)$, e U é uma variável aleatória positiva, independente de Z, com função de distribuição $H(\cdot|\nu)$.

Note que $Y|U = u \sim \mathrm{SN}(0, u^{-1}\sigma^2, \lambda)$ o que implica na distribuição marginal de Y definida em 2.3.3. Utilizando as propriedades da distribuição skew-normal, conforme discutidas por Azzalini e Valle (1996), segue a seguinte proposição.

Proposição 2.3.1. *Seja $Y \sim MESN(\mu, \sigma^2, \lambda; H)$, logo*

$$Y \overset{d}{=} \mu + \Delta T + u^{-1/2}\Gamma^{1/2}T_0, \tag{2.9}$$

onde $T = U^{-1/2}|T_0|, T_0 \sim N(0,1), \delta = \sqrt{1+\lambda^2}, \Delta = \sigma\delta$ e $\Gamma = \sigma^2(1-\delta^2)$.

Essa representação permite relacionar a estrutura de variância e a assimetria da distribuição skew-normal com as variáveis aleatórias subjacentes, possibilitando a utilização de métodos de estimação, como o algoritmo EM.

A seguir, será derivada a esperança do modelo $MESN$ a partir da representação estocástica apresentada em 2.8.

Proposição 2.3.2. *Seja* $Y \sim MESN(\mu, \sigma^2, \lambda; H)$*, logo*

$$\mathbb{E}[Y] = \mu + \sqrt{\frac{2}{\pi}}\mathbb{E}[U^{1/2}]\Delta, \tag{2.10}$$

com Δ *e* U *definido anteriormente.*

Assim, uma vez que $\mathbb{E}[Y] \neq \mu$, o modelo não é centrado.

2.4 ALGORITMO EM

O algoritmo EM (Expectation-Maximization) é um método iterativo utilizado para encontrar estimativas de máxima verossimilhança de parâmetros em modelos estatísticos, especialmente quando os dados estão incompletos ou possuem variáveis latentes, como é o caso das misturas finitas.

O algoritmo EM consiste em duas etapas principais que se repetem até a convergência:

- Etapa E (Expectation): Calcular o valor esperado da função de log-verossimilhança dos dados completos, condicionada aos dados observados e às estimativas atuais dos parâmetros, ou seja, $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) = E[l_c(\boldsymbol{\theta}|\boldsymbol{y_c})|\boldsymbol{y_{obs}}, \hat{\boldsymbol{\theta}}]$, onde $l_c(\boldsymbol{\theta}|\boldsymbol{y_c})$ é a log-verossimilhança dos dados completos, $\boldsymbol{y_{obs}}$ são os dados observados e $\boldsymbol{\theta}$ é o vetor de parâmetros.
- Etapa M (Maximization): Maximizar a função Q obtida na etapa E em relação aos parâmetros do modelo.

Nesse contexto, seja $\hat{\theta}^{(k)}$ a estimativa de θ na iteração k, onde $\hat{\theta}^{(0)}$ representa o valor inicial necessário para iniciar o algoritmo.

1. Etapa E: Considera $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}^{(k)}$ e calcula-se a função Q.
2. Etapa M: Encontra $\boldsymbol{\theta}^{(k+1)}$ que maximiza Q.

As etapas E e M são repetidas até o critério de parada ser satisfeito. Existem diversas formas de definir um critério de parada, por exemplo baseado na norma da diferença relativa entre o vetor de parâmetros $\left|\frac{\hat{\boldsymbol{\theta}}^{(k+1)} - \hat{\boldsymbol{\theta}}^{(k)}}{\hat{\boldsymbol{\theta}}^{(k)}}\right|$, ou baseado na norma da diferença relativa entre a log-verossimilhança dos dados observados $\left|\frac{\hat{l}^{k+1} - \hat{l}^{k}}{\hat{l}^{k}}\right|$, onde $\hat{l}^k = l(\hat{\boldsymbol{\theta}}^k|\boldsymbol{y_{obs}})$.

Neste trabalho, será utilizado o critério de parada baseado na convergência da log-verossimilhança dos dados observados. Assim, as etapas serão repetidas até que a seguinte condição seja satisfeita:

$$\left| \frac{\hat{l}^{k+1} - \hat{l}^{k}}{\hat{l}^{k}} \right| < \epsilon. \tag{2.11}$$

2.4.1 Extensões

O algoritmo EM é uma ferramenta poderosa para a estimação de parâmetros em modelos estatísticos com dados incompletos ou variáveis latentes. Suas variantes, como ECM e ECME, expandem sua aplicabilidade e eficiência, tornando-o uma escolha frequente em diversas áreas de pesquisa e aplicação prática.

Meng e Rubin (1993) introduziram uma extensão do algoritmo EM, chamada ECM (Expectation Conditional Maximization). No ECM, a etapa M é substituída pela etapa CM (Maximização Condicional), na qual o vetor $\boldsymbol{\theta}$ é dividido em dois subvetores, $\boldsymbol{\theta}_1 \in \Theta_1$ e $\boldsymbol{\theta}_2 \in \Theta_2$, onde $\Theta = \Theta_1 \cup \Theta_2$, e modificam a etapa M da seguinte maneira:

- **ECM** $\begin{cases} \text{Etapa CM-1}: \quad \widehat{\boldsymbol{\theta}}_1^{(k+1)} = \arg\max_{\boldsymbol{\theta}_1} Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}_2^{(k)}), \\ \text{Etapa CM-2}: \quad \widehat{\boldsymbol{\theta}}_2^{(k+1)} = \arg\max_{\boldsymbol{\theta}_2} Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}_1^{(k+1)}). \end{cases}$

Liu e Rubin (1994) propuseram o algoritmo ECME, que se baseia no ECM, mas adiciona na segunda etapa de CM, a maximização da log-verossimilhança dos dados observados, como descrito a seguir.

- **ECME** $\begin{cases} \text{Etapa CM}: \quad \widehat{\boldsymbol{\theta}}_1^{(k+1)} = \arg\max_{\boldsymbol{\theta}_1} Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}_2^{(k)}), \\ \text{Etapa CML}: \quad \widehat{\boldsymbol{\theta}}_2^{(k+1)} = \arg\max_{\boldsymbol{\theta}_2} l(\boldsymbol{\theta}|y_{\text{obs}}, \widehat{\boldsymbol{\theta}}_1^{(k+1)}). \end{cases}$

Observe que nesta segunda etapa do passo CM geralmente é necessário utilizar métodos numéricos, como o método de Newton-Raphson ou o método simplex de Nelder e Mead (1965), que pode ser implementado na função optim do software R Core Team (2024).

2.5 Bayesian Information Criterion (BIC)

Na prática da modelagem baseada em agrupamento, o número de componentes G é tipicamente desconhecido, e sua seleção é frequentemente necessária. Para abordar essa questão, é possível ajustar modelos de mistura para uma gama de valores de G e, em seguida, escolher o valor de G associado ao melhor resultado com base no critério adotado.

Neste trabalho, adotamos o Critério de Informação Bayesiana (BIC; Schwarz (1978)) para selecionar o número de componentes em modelos de mistura. A forma do BIC é dada por

$$\text{BIC} = -2\ell(\widehat{\boldsymbol{\theta}}|\boldsymbol{y_{obs}}) + k \log n,$$

onde $\ell(\widehat{\boldsymbol{\theta}})$ é a log-verossimilhança dos dados observados avaliada nos parâmetros estimados e k é o número de parâmetros no modelo. Assim, modelos com menores valores de BIC são preferidos; consulte Depraetere e Vandebroek (2014) para uma revisão no contexto de modelos de mistura gaussiana. Estudos recentes de Zeller et al. (2019), Mazza e Punzo (2020) e Mirfarah, Naderi e Chen (2021) demonstraram a eficácia do BIC na determinação do número de componentes em modelos de misturas finitas.

3 MISTURAS FINITAS DE DENSIDADES MESN

Neste capítulo, apresentamos a integração das estruturas descritas nas Seções 2.1 e 2.3, culminando na formulação dos Modelos de Misturas Finitas de Densidades MESN (MF-MESN). Essa abordagem une a flexibilidade das distribuições MESN com a capacidade das misturas finitas de capturar heterogeneidade, permitindo modelagens estatísticas robustas e adaptáveis a dados complexos.

Os modelos MF-MESN têm atraído crescente interesse em aplicações práticas devido à sua capacidade de capturar características frequentemente presentes em dados reais, como assimetria e curtose. Além disso, essa classe de modelos oferece uma solução flexível para descrever padrões variados, sendo empregada em estudos recentes, como Doe e Smith (2023) e Smith e Taylor (2023).

3.1 DEFINIÇÃO

A partir da junção das Definições 2.1.1 e 2.3.3, defini-se o modelo MF-MESN.

Definição 3.1.1. Uma variável aleatória Y segue um modelo de misturas finitas de densidades MESN, com G-componentes, se sua função de densidade é dada por:

$$f(y_i) = \sum_{j=1}^{G} \pi_j \psi(y_i|\boldsymbol{\theta}_j),\ \pi_j > 0 \text{ e } \sum_{j=1}^{G} \pi_j = 1, \tag{3.1}$$

em que $\psi(y|\boldsymbol{\theta}_j)$ é a função de densidade de probabilidade da classe $MESN(\mu, \sigma^2, \lambda; H)$, definida na Definição 2.3.1.

Ao combinar duas estruturas paramétricas, as misturas finitas e a classe MESN, o modelo resultante incorpora parâmetros de ambas as classes. Assim, denotamos $Y_i|Z_{ij} = 1 \sim$ MF-MESN$(\mu_j, \sigma_j^2, \lambda_j, H)$, onde $\left\{\mu_j, \sigma_j^2, \lambda_j, H\right\}$ corresponde aos parâmetros da classe de distribuição MESN, e o vetor $\boldsymbol{\pi} = (\pi_1, ..., \pi_{G-1})$, de tamanho $G-1$ corresponde às proporções das componentes da mistura, em que $\pi_G = 1 - \sum_{j=1}^{G-1} \pi_j$.

3.2 REPRESENTAÇÃO ESTOCÁSTICA

Como definido na Seção 3.1, os modelos MF-MESN resultam da combinação das estruturas introduzidas no Capítulo 2. Dessa forma, a representação estocástica é derivada das representações na Definição 2.1.1 e Proposição 2.3.1, resultando nas seguintes Proposições.

Proposição 3.2.1. *Seja $Y \sim$ MF-MESN, dado na Definição 3.1. Este modelo pode ser escrito hierarquicamente como:*

$$Y_i|u_i, t_i, Z_{ij} = 1 \sim N(\mu_j + \Delta_j t_i, u_i \Gamma_j^{1/2}), \tag{3.2}$$

$$T_i|u_i, Z_{ij} = 1 \sim HN(0, u_i), \tag{3.3}$$

$$U_i|Z_{ij} = 1 \sim H(.|\boldsymbol{\nu}), \tag{3.4}$$

$$Z_{ij} \sim Multinomial(1|\boldsymbol{\pi}), \tag{3.5}$$

onde $\Delta_j = \sigma_j\delta_j$, $\delta_j = \frac{\lambda_j}{\sqrt{1+\lambda_j^2}}$ *e* $\Gamma_j = \sigma_j^2(1-\delta_j^2)$.

Com base no modelo hierárquico, conclui-se que a representação estocástica de Y pode ser obtida seguindo o mesmo raciocínio apresentado na Proposição 2.3.1.

Proposição 3.2.2. *Seja* $Y \sim$ *MF-MESN, dado na Definição 3.1, logo*

$$Y|Z_j = 1 \stackrel{d}{=} \mu_j + \Delta_j T + u^{-1/2}\Gamma_j^{1/2}T_0 \tag{3.6}$$

onde $T = u^{-1/2}|T_0|$, $T_0 \sim N(0,1)$, $\delta_j = \frac{\lambda_j}{\sqrt{(1+\lambda_j^2)}}$, $\Delta_j = \sigma_j\delta_j$ *e* $\Gamma_j = \sigma_j^2(1-\delta_j^2)$

3.3 ESTIMAÇÃO

Conforme discutido na Seção 2.4, a representação estocástica desempenha um papel fundamental na aplicação do Algoritmo EM para obter os estimadores de máxima verossimilhança. Dessa forma, a partir da Proposição 2.3.1, temos que a função de log-verossimilhança dos dados completos, $\boldsymbol{y_c} = (\boldsymbol{y}, \boldsymbol{u}, \boldsymbol{t}, \boldsymbol{z})$, onde $\boldsymbol{y} = (y_1, ..., y_n)$, $\boldsymbol{u} = (u_1, ..., u_n)$, $\boldsymbol{t} = (t_1, ..., t_n)$, $\boldsymbol{z} = (z_1, ..., z_n)$ e o vetor de parâmetros $\boldsymbol{\theta}_j = (\mu_j, \sigma_j^2, \lambda_j, \nu_j, \pi_j)$ é dada por:

$$l_c(\boldsymbol{\theta}) = C + \sum_{i=1}^{n}\sum_{j=1}^{G} z_{ij}[log\ \pi_j - \frac{1}{2}log\ \Gamma_j^2 - \frac{1}{2\Gamma_j^2}u_i(y_i - \mu_j - \Delta_j t_i)^2 + log(h(u_i|\nu_j))], \tag{3.7}$$

em que C é uma constante que representa a parte independente de parâmetros, dado que não é necessária para estimação.

Portanto, $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)})$ é dada por:

$$Q = C + \sum_{i=1}^{n}\sum_{j=1}^{G} \hat{z_{ij}}^{(k)}$$
$$\left\{log\ \pi_j - \frac{1}{2}log\ \Gamma_j^2 - \frac{1}{2\Gamma_j^2}\left(\Delta_j^2\widehat{ut_{ij}^2}^{(k)} + (y_i - \mu_j)^2\hat{u_{ij}}^{(k)} - 2\Delta_j\widehat{ut_{ij}}^{(k)}(y_i - \mu_j)\right)\right\}, \text{tal que}$$

$$\hat{z_{ij}} = E[Z_{ij}|\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}] = \frac{\hat{\pi}_j\psi(y_i|\hat{\boldsymbol{\theta}}_j)}{\sum_{j=1}^{G}\hat{\pi}_j\psi(y_i|\hat{\boldsymbol{\theta}}_j)}, \tag{3.8}$$

$$\widehat{z_{ij}u_i} = E[Z_{ij}U_i|\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}], \tag{3.9}$$

$$\widehat{z_{ij}u_it_i} = E[Z_{ij}U_iT_i|\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}] = \hat{z_{ij}}(\hat{u_i}\hat{\mu_{T_i}} + \hat{M_{T_i}}\hat{\eta_i}), \tag{3.10}$$

$$\widehat{z_{ij}u_it_i^2} = E[z_{ij}U_iT_i^2|\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}] = \hat{z_{ij}}(\hat{u}_i\hat{\mu_{T_i}^2} + \hat{M_{T_i}^2} + \hat{M}_{T_i}\hat{\mu_{T_i}}\hat{\eta}_i). \tag{3.11}$$

onde $\hat{\eta}_i = E\left[U_i^{1/2}W_\phi\left(\frac{U_i^{1/2}\hat{\mu_{T_i}}}{\hat{M_{T_i}}}\right)|\hat{\boldsymbol{\theta}}, y_i\right]$, $W_\phi(.) = \frac{\phi(.)}{\Phi(.)}$, $M_{T_i}^2 = \frac{\Gamma}{(\Gamma+\Delta^2)}$ e $\mu_{T_i} = M_{T_i}^2\frac{\Delta^2}{\Gamma}(y_i - \mu_j)$.

Aplicando o algoritmo ECME, a etapa E consiste em calcular os valores esperados $\widehat{z_{ij}u_i}$ $\widehat{z_{ij}u_it_i}$ e $\widehat{z_{ij}u_it_i^2}$, seguido pela etapa CM, que consiste na maximização de $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})$. Por fim, a etapa CML atualiza os parâmetros ν_j maximizando a função *log-verossimilhança* dos dados observados.

Etapa E: Considerando $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}^{(k)}$, obtém-se $\widehat{z_{ij}u_i}$, $\widehat{z_{ij}u_it_i}$ e $\widehat{z_{ij}u_it_i^2}$.

Etapa CM: Atualiza $\hat{\boldsymbol{\theta}}^{(k+1)}$ maximizando $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)})$ a partir das seguintes expressões explícitas:

$$\widehat{\mu_j}^{(k+1)} = \frac{\sum_{i=1}^n z_{ij}^{(k)}y_i - \widehat{z_{ij}u_it_i}^{(k)}\widehat{\Delta_j}^{(k)}}{\sum_{i=1}^n z_{ij}^{(k)(k)}}, \tag{3.12}$$

$$\begin{aligned}\widehat{\Gamma_j}^{(k+1)} = \frac{1}{n}\sum_{i=1}^n \Big[& z_{ij}^{(k)}(y_i - \widehat{\mu_j}^{(k+1)})^2 - 2\widehat{z_{ij}u_it_i}^{(k)}\widehat{\Delta_j}^{(k)}(y_i - \widehat{\mu_j}^{(k+1)}) \\ & + \widehat{z_{ij}u_it_i^2}^{(k)}\widehat{\Delta_j}^{2(k)}\Big],\end{aligned} \tag{3.13}$$

$$\widehat{\Delta_j}^{(k+1)} = \frac{\sum_{i=1}^n \widehat{z_{ij}u_it_i}^{(k)}(y_i - \widehat{\mu_j}^{(k+1)})}{\sum_{i=1}^n \widehat{z_{ij}u_it_i^2}^{(k)}}, \tag{3.14}$$

$$\widehat{\sigma_j}^{(k+1)} = \widehat{\Gamma_j}^{(k+1)} + \widehat{\Delta_j}^{2(k+1)}, \tag{3.15}$$

$$\widehat{\lambda_j}^{(k+1)} = \frac{\widehat{\Delta_j^{(k+1)}}}{\sqrt{\widehat{\Gamma_j}^{(k+1)}}}. \tag{3.16}$$

Etapa CML: Atualizar os parâmetros $\nu_j^{(k+1)}$ maximizando a função de *log-verossimilhança* marginal

$$\hat{\nu_j}^{(k+1)} = \arg\max_{\nu_j}\sum_{i=1}^n \log\psi(y_i|\mu_j^{(k+1)}, \sigma_j^{(k+1)}, \lambda_j^{(k+1)}, \nu_j), \tag{3.17}$$

onde $\psi(y;\theta)$ é a densidade da MESN, dada em (3.7).

Com as etapas definidas, o algoritmo é iniciado a partir de um chute inicial para $\boldsymbol{\theta}$. Mais detalhes sobre o valor inicial serão dados na Seção 4.4. Note que, quando $\lambda_j = 0$ (ou $\Delta_j = 0$), as equações da etapa M se reduzem às equações obtidas assumindo distribuições MEN.

As iterações são repetidas até que uma regra de convergência adequada seja satisfeita; veja detalhes na Seção 2.4.

4 MISTURAS FINITAS DE ESPECIALISTAS

Neste capítulo é apresentada a busca por inovação por meio do estudo misturas finitas de especialistas (*MoE*), uma extensão dos modelos de misturas que tem como objetivo a melhoria na separação dos grupos e da interpretabilidade. Jacobs et al. (1991) define essa classe de modelos baseado em uma estrutura composta pelos *gatings* (portas) e *experts* (especialistas). Mengersen, Robert e Titterington (2011) defende que a inclusão de covariáveis por meio de mistura de especialistas pode fornecer resultados diferentes e muitas vezes com uma estrutura mais clara devido ao uso das covariáveis em duas fontes de informação.

Os *gatings* são a representação das proporções de cada grupo, a ideia é que essa proporção está diretamente relacionada com as covariáveis. Enquanto os *experts* representam as componentes da mistura, ou seja, os modelos que estimam a densidade de cada observação.

A grande diferença com as misturas finitas de densidades é que a proporção ($\boldsymbol{\pi}$) dos grupos não é mais fixa, esta varia para cada indivíduo. Dessa forma, o obetivo é estimar essa proporção baseada nas características destes indivíduos.

Nesse sentido, existem diversas maneiras de ajustar os *gatings*. Neste trabalho, vamos utilizar a regressão logística multinomial nesse ajuste, para isso define-se os coeficientes das proporções $\boldsymbol{\alpha} = (\boldsymbol{\alpha_1}, \boldsymbol{\alpha_2}, ..., \boldsymbol{\alpha_{G-1}})$, em que $\boldsymbol{\alpha}_j = (\alpha_{j0}, \alpha_{j1}, ..., \alpha_{jm-1})$ e um subgrupo das covariáveis definidas como $\boldsymbol{R} = (\boldsymbol{r_1}, \boldsymbol{r_2}, ..., \boldsymbol{r_n})$, onde $\boldsymbol{r}_i = (r_{i0}, r_{i1}, ..., r_{im-1})$.

Definição 4.0.1. A variável aletória Y segue o modelo de mistura de especialistas de sua função de densidade é dada por:

$$f(y_i) = \sum_{j=1}^{G} \pi_j(\boldsymbol{r}_i|\boldsymbol{\alpha}_j)\ g(y_i|\boldsymbol{\theta}_j), \pi_j(\boldsymbol{r}_i|\boldsymbol{\alpha}_j) > 0 \text{ e } \sum_{j=1}^{G} \pi_j(\boldsymbol{r}_i|\boldsymbol{\alpha}_j) = 1, \tag{4.1}$$

onde $\pi_j(\boldsymbol{r}_i|\boldsymbol{\alpha}_j) = \frac{e^{\boldsymbol{\alpha}_j^\top \boldsymbol{r}_i}}{1+\sum_{l=1}^{G-1} e^{\boldsymbol{\alpha}_l^\top \boldsymbol{r}_i}}$ e $\boldsymbol{r}_i$ é a linha i da matriz $\boldsymbol{R}$ (covariáveis).

4.1 MISTURAS FINITAS DE ESPECIALISTAS DE REGRESSÂO

Da mesma forma que para as misturas finitas de densidades, podemos utilizar as misturas finitas de especialistas para regressão. Dessa forma, consideramos que, além dos *gatings*, os *experts* dependem de covariáveis $\boldsymbol{X} = (\boldsymbol{x_1}, \boldsymbol{x_2}, ..., \boldsymbol{x_n})$, onde $\boldsymbol{x_i} = (x_{i0}, x_{i1}, ..., x_{iq-1})$. Muitas vezes $\boldsymbol{X} = \boldsymbol{R}$, no entanto não existe essa restrição e vamos trabalhar como duas matrizes diferentes de covariáveis. Para isso, define-se os coeficientes da regressão como $\boldsymbol{\beta} = (\beta_0, \beta_1, ..., \beta_{q-1})$.

Com isso, estabelecemos que tanto a probabilidade de uma observação pertencer a um grupo quanto a variável resposta Y dependem das covariáveis, resultando nas duas matrizes $\boldsymbol{R}$ e $\boldsymbol{X}$, além dos dois vetores de coeficientes $\boldsymbol{\alpha}$ e $\boldsymbol{\beta}$.

Definição 4.1.1. Uma variável aleatória Y segue um modelo de misturas finitas de especialistas de regressão se sua função de densidade de Y é dada por:

$$f(y_i) = \sum_{j=1}^{G} \pi_j(\boldsymbol{r}_i|\boldsymbol{\alpha}_j) g(y_i|\boldsymbol{x}_i^\top \boldsymbol{\beta}_j, \boldsymbol{\theta}_j), \pi_j(\boldsymbol{r}_i|\boldsymbol{\alpha}_j) > 0 \text{ e } \sum_{j=1}^{G} \pi_j(\boldsymbol{r}_i|\boldsymbol{\alpha}_j) = 1. \tag{4.2}$$

Nota-se a semelhança entre os modelos de misturas finitas de regressão e os modelos de especialistas de regressão, de tal forma que as misturas finitas são um caso particular das misturas de especialistas de regressão.

Na Figura 1, conseguimos mostrar esse paralelo entre os métodos de misturas finitas de densidades, misturas finitas de regressão e misturas finitas de especialistas de regressão, onde $\boldsymbol{X}$ é a matriz de covariáveis $n \times q$, $\boldsymbol{Z}$ é a matriz indicadora do grupo $n \times g$ e $\boldsymbol{Y}$ é o vetor referente a variável resposta de interesse $n \times 1$.

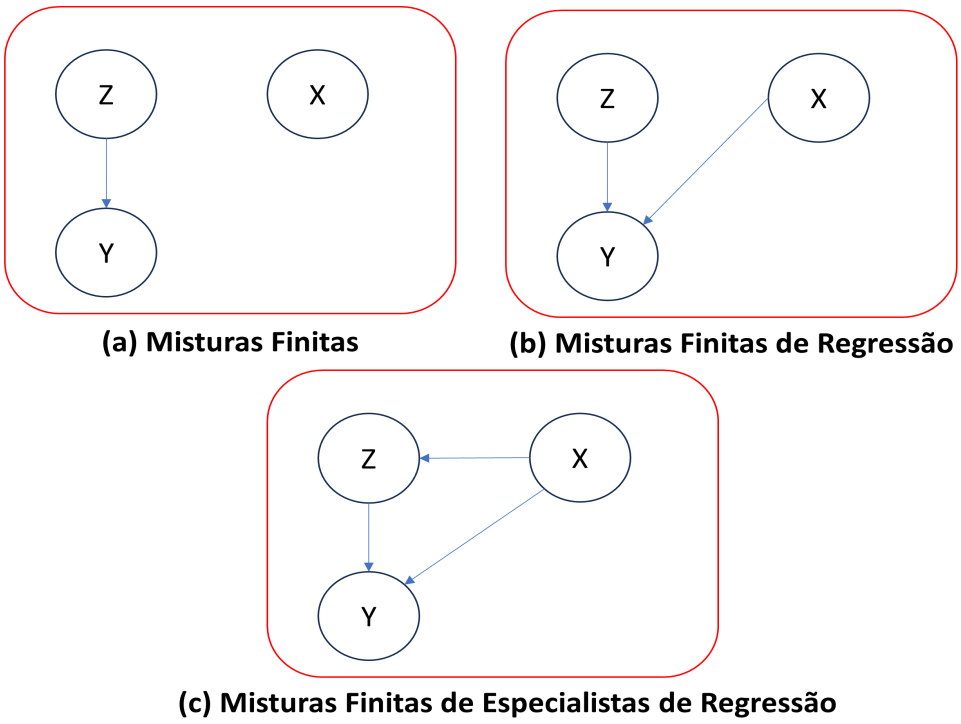


Figura 1 – Comparação entre extensões do modelo de Misturas Finitas de densidades

Fonte: Elaboração Própria.

A partir da Figura 1 vemos como cada modelo se relaciona com as variáveis, em que o modelo de Misturas Finitas de Especialistas de Regressão (c) é uma extensão do modelo de Misturas Finitas de Regressão (b), enquanto este é uma extensão do modelo de Misturas Finitas (a). Dessa forma, a principal diferença entre os modelos está na maneira

como os componentes são relacionados e como a matriz de covaráveis $\boldsymbol{X}$ influencia as demais variáveis, $\boldsymbol{Y}$ e $\boldsymbol{Z}$. O modelo de Misturas Finitas de Especialistas de Regressão (c) se destaca por conseguir capturar simultaneamente as relações entre $\boldsymbol{X}$, $\boldsymbol{Y}$ e $\boldsymbol{Z}$ dentro de uma única estrutura.

4.2 ESTIMAÇÃO

Como o modelo de misturas finitas de especialistas sofre alteração apenas na modelagem das proporções em relação ao modelo apresentado no Capítulo 3, a representação estocástica na Proposição 3.2.2 se mantém. A mudança será apenas na definição do vetor latente $\boldsymbol{Z}$, que neste caso depende de $\boldsymbol{\alpha}$ e $\boldsymbol{R}$, dessa forma:

$$\boldsymbol{Z_i} = (Z_{i1}, ..., Z_{iG}) \sim Multinomial\left(1 \mid \pi(\boldsymbol{r}_i|\boldsymbol{\alpha_1}), ..., \pi(\boldsymbol{r}_i|\boldsymbol{\alpha_{G-1}}), 1 - \sum_{j=1}^{G-1} \pi(\boldsymbol{r}_i|\boldsymbol{\alpha_j})\right), \quad (4.3)$$

Além disso, no caso da regressão considera-se $\boldsymbol{\mu} = \boldsymbol{X\beta}$, dessa forma a representação estocástica do modelo é dada por:

$$Y_i|Z_{ij} = 1 \overset{d}{=} \boldsymbol{x_i\beta_j} + \Delta_j T + u^{-1/2}\Gamma_j^{1/2}T_0. \quad (4.4)$$

Assim, a função de log-verossimilhança dos dados completos é dada por:

$$l_c(\boldsymbol{\theta}) = C + \sum_{i=1}^{n}\sum_{j=1}^{G} z_{ij}[log\ \pi_j(\boldsymbol{r}_i|\boldsymbol{\alpha}_j) - \frac{1}{2}log\ \Gamma_j - \frac{1}{2\Gamma_j}u_i(y_i - \boldsymbol{x_i\beta_j} - \Delta_j t_i)^2 + log(h(u_i|\nu_j))] \quad (4.5)$$

Consequentemente, define-se a função $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)}) = Q$ tal que:

$$Q = C + \sum_{i=1}^{n}\sum_{j=1}^{G} \hat{z_{ij}}^{(k)} \left\{log\ \pi_j(\boldsymbol{r}_i|\boldsymbol{\alpha}_j) - \frac{1}{2}log\ \Gamma_j^2 - \frac{1}{2\Gamma_j^2}\left(\Delta_j^2 \widehat{ut_{ij}^2}^{(k)} + (y_i - \boldsymbol{x}_i^\top\boldsymbol{\beta}_j)^2 \hat{u_{ij}}^{(k)} - 2\Delta_j \widehat{ut_{ij}}^{(k)}\left(y_i - \boldsymbol{x}_i^\top\boldsymbol{\beta}_j\right)\right)\right\}.$$

Aplicando o algoritmo ECME, assim como na Seção 3.3, temos:

Etapa E: Considerando $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}^{(\boldsymbol{k})}$, obtém-se $\widehat{z_{ij}u_i}$ (3.9), $\widehat{z_{ij}u_it_i}$ (3.10) e $\widehat{z_{ij}u_it_i^2}$ (3.11). No entanto, como estamos no contexto de regressão temos $\mu_{T_i} = M_{T_i}^2\frac{\Delta^2}{\Gamma}(y_i - \boldsymbol{x}_i\boldsymbol{\beta}_j)$.

Etapa CM: Atualizar $\hat{\boldsymbol{\theta}}^{(\boldsymbol{k+1})}$ maximizando $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)})$, utilizando as expressões na Seção 3.3 para os estimadores $\widehat{\sigma_j}^{(k+1)}$ (3.15) e $\widehat{\lambda_j}^{(k+1)}$ (3.16), além das seguintes expressões para $\widehat{\boldsymbol{\beta_j}}^{(k+1)}$, $\hat{\boldsymbol{\alpha}}_j^{(k+1)}$, $\widehat{\Gamma_j}^{(k+1)}$ e $\widehat{\Delta_j}^{(k+1)}$:

$$\widehat{\boldsymbol{\beta_j}}^{(k+1)} = \left(\sum_{i=1}^{n} \widehat{z_{ij}u_i}^{(k)} \boldsymbol{x_i}\boldsymbol{x_i}^{\top}\right)^{-1} \sum_{i=1}^{n} (\widehat{z_{ij}u_i}^{(k)} y_i - \widehat{z_{ij}u_i t_i}^{(k)} \widehat{\Delta_j}^{(k)}) \boldsymbol{x_i}, \tag{4.6}$$

$$\hat{\boldsymbol{\alpha}}_j^{(k+1)} = 4\left(\sum_{i=1}^{n} \boldsymbol{r}_i \boldsymbol{r}_i^{\top}\right)^{-1} \left(\sum_{i=1}^{n} \hat{z}_{ij}\left[1 - \pi_j\left(\boldsymbol{r}_i|\hat{\boldsymbol{\alpha}}_j^{(k)}\right)\right]\boldsymbol{r}_i\right) + \hat{\boldsymbol{\alpha}}_j^{(k)}, \tag{4.7}$$

$$\begin{aligned}\widehat{\Gamma_j}^{(k+1)} = \frac{1}{n}\sum_{i=1}^{n}\Big[& z_{ij}^{(k)}(y_i - \widehat{\boldsymbol{x}_i\boldsymbol{\beta}_j})^{(k+1)})^2 - 2\widehat{z_{ij}u_i t_i}^{(k)} \widehat{\Delta_j}^{(k)}(y_i - \widehat{\boldsymbol{x}_i\boldsymbol{\beta}_j})^{(k+1)}) \\ & + \widehat{z_{ij}u_i t_i^2}^{(k)} \widehat{\Delta_j}^{2(k)}\Big],\end{aligned} \tag{4.8}$$

$$\widehat{\Delta_j}^{(k+1)} = \frac{\sum_{i=1}^{n} \widehat{z_{ij}u_i t_i}^{(k)}(y_i - \widehat{\boldsymbol{x}_i\boldsymbol{\beta}_j})^{(k+1)})}{\sum_{i=1}^{n} \widehat{z_{ij}u_i t_i^2}^{(k)}}, \tag{4.9}$$

$$\tag{4.10}$$

Etapa CML: Atualizar os parâmetros $\nu_j^{(k+1)}$ maximizando a função de *log-verossimilhança* marginal assim como em (3.17).

4.3 CLASSIFICAÇÃO

Após a estimação, as classificações das obervações são obtidas a partir dos valores esperados das variáveis latentes $\boldsymbol{Z_i}$, ou seja, dos valores $\widehat{z_{ij}}$ (3.8), usando as maiores probabilidades para encontrar a componente correspondente. Para isso, são obtidos os índices j que maximizam $\widehat{z_{ij}}$, ou seja, o $\arg\max_j \ \widehat{z_{ij}}$ para cada observação i, com isso a observação é alocada no grupo j.

4.4 VALORES INICIAIS

A escolha dos valores iniciais é fundamental para a estimação pelo Algoritmo EM, uma vez que quanto mais distantes dos valores reais mais demorado e impreciso é o processo de estimação. Assim, valores iniciais mais precisos evitam máximos locais e aceleram o processo de estimação.

No contexto de misturas finitas, o chute inicial necessita separar os dados a partir de um método de clusterização e, com os grupos formados, os parâmetros são estimados. Neste trabalho, o método de *k-means* foi utilizado na clusterização, com isso o chute inicial do algoritmo do tipo EM proposto foram obtidos da seguinte forma:

1. Os dados são particionados em G grupos a partir do *k-means*. (BASSO et al., 2010).

2. Para os valores de $\boldsymbol{\alpha}_j$, Mirfarah, Naderi e Chen (2021) descreve duas estratégias, utilizaremos a mais simples, onde $\hat{\boldsymbol{\alpha}}_j^{(0)} = 0$.

3. Para cada grupo j, $\hat{\boldsymbol{\beta}^{(0)}}_j$ é estimado pelo método de mínimos quadrados.

4. Os resíduos de cada grupo são utilizados para estimar σ_j a partir da soma dos quadrados, e λ_j a partir do coeficiente de assimetria.

5. Por fim, para estimar ν_j é utilizado método de L-BFGS-B, descrito em Byrd et al. (1995), na maximização da função de *log-verossimilhança* marginal.

Todo o processo do Algoritmo EM foi implementado no software R Core Team (2024), usando a função *k-means* para clusterização, a função *optim* nas otimizações numéricas e a função *skewness* do pacote *moments* no valor inicial de λ_j.

4.5 PARCIMONIOSIDADE

Um dos grandes desafios no ajuste de modelos complexos é encontrar um equilíbrio entre adicionar informações e aumentar o número de parâmetros. A principal preocupação ao aumentar a complexidade do modelo é o risco de sobreajuste (overfitting), que pode comprometer a capacidade de generalização e a interpretação do modelo para o problema de estudo.

No contexto de modelos de misturas, um problema significativo é o aumento da quantidade de parâmetros à medida que novas componentes são adicionadas. Cada nova componente incrementa o número de parâmetros de forma aritmética, com uma razão k_1 (número de parâmetros quando $G = 1$).

Para abordar essa questão, várias estratégias buscam aumentar a parcimônia dos modelos ao adicionar componentes. Uma abordagem comum é estimar um mesmo parâmetro para todos os grupos, normalmente um parâmetro que não afete a interpretação e a relação entre as variáveis $\boldsymbol{X}$, $\boldsymbol{Z}$ e $\boldsymbol{Y}$.

Neste trabalho, exploraremos duas dessas alternativas para aumentar a parcimônia dos modelos. A primeira alternativa considera a igualdade do parâmetro ν nas distribuições *t-Student* e *skew-t*, assumindo que $\nu_1 = \nu_2 = \cdots = \nu_G$. A segunda alternativa, utilizada em Zeller, Cabral e Lachos (2016), aplica a mesma abordagem para o parâmetro de escala Γ, definido na Proposição 2.3.1, assumindo que $\Gamma_1 = \Gamma_2 = \cdots = \Gamma_G$. Nesse contexto o modelo se torna mais simples uma vez que possui menos parâmetros para estimar, podendo ainda diferenciar os grupos por meio dos parâmetros σ_j e λ_j, porém restringindo seus possíveis valores ao limitar Γ_j. Além disso no contexto das distribuições simétricas $\Gamma = \sigma^2$, logo $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_G^2$.

5 ESTUDOS DE SIMULAÇÃO

Neste capítulo, um estudo de simulação foi projetado com intuito de investigar o desempenho e as mudanças que ocorrem nas estimativas de máxima verossimilhança dos parâmetros no contexto das misturas finitas de especialistas de regressão sob a distribuição *skew-t*, com G = 2, ao variar os tamanhos amostrais, n, em 100, 200, 500 e 1000. De modo que, para cada valor escolhido de n, são geradas artificialmente 500 amostras através da representação estocástica apresentada na Proposição 3.2.2. Abaixo se apresentam as configurações assumidas para os parâmetros das misturas finitas de especialistas:

$$\boldsymbol{\beta_1} = (\beta_{01}, \beta_{11}, \beta_{21}, \beta_{31}) = (0, -1, -2, -3), \boldsymbol{\beta_2} = (\beta_{02}, \beta_{12}, \beta_{22}, \beta_{32}) = (-1, 1, 2, 3),$$

$$\boldsymbol{\Gamma} = (\Gamma_1, \Gamma_2) = (1.5, 1.5), \boldsymbol{\Delta} = (\Delta_1, \Delta_2) = (3, -3),$$
$$\boldsymbol{\lambda} = (\lambda_1, \lambda_2) = (\frac{3}{\sqrt{1.5}}, \frac{-3}{\sqrt{1.5}}), \boldsymbol{\sigma} = (\sigma_1, \sigma_2) = (1, \sqrt{2}),$$
$$\nu_1 = \nu_2 = 3, \boldsymbol{\alpha} = (\alpha_{01}, \alpha_{11}, \alpha_{21}) = (0.7, 1, 2), \text{ tal que}$$

$$\boldsymbol{x}^\top = (1, \boldsymbol{x_1}, \boldsymbol{x_2}, \boldsymbol{x_3}) \sim (1, Unif(1, 5), Unif(-2, 2), Unif(1, 4)),$$
$$\boldsymbol{r}^\top = (1, \boldsymbol{r_1}, \boldsymbol{r_2}) \sim (1, Unif(-2, 1), Unif(-1, 1)),$$

em que β_{ij} é o coeficiente β_i do grupo j, em que $i = 0$ indica o intercepto, α_{ij} é o coeficiente α_i do grupo j, enquanto para os demais parâmetros θ_j é o parâmetro θ do grupo j, $\forall\ \theta \in (\Gamma, \Delta, \lambda, \sigma, \nu)$.

Além disso, com intenção de descrever e entender melhor o comportamento dessas estimativas, o estudo de simulação foi dividido em quatro cenários, onde essa divisão é realizada de acordo com o princípio da parcimônia, conforme discutido na Seção 4.4. Para cada amostra gerada artificialmente sob as configurações descritas acima, ou seja, as amostras foram geradas apenas uma vez, ajustamos as misturas finitas de especialistas de regressão nos seguintes cenários (afetando apenas a estimação):

- Cenário 1: Sem restrições, todos os parâmetros são estimados livremente.
- Cenário 2: Restrição em Γ, em que estima-se $\Gamma = \Gamma_1 = \Gamma_2$.
- Cenário 3: Restrição em ν, em que estima-se $\nu = \nu_1 = \nu_2$.
- Cenário 4: Restrição em ν e Γ, em que estima-se $\Gamma = \Gamma_1 = \Gamma_2$ e $\nu = \nu_1 = \nu_2$.

Nesse sentido, como nas configurações dos parâmetros $\Gamma_1 = \Gamma_2$ e $\nu_1 = \nu_2$, consideramos que o cenário 4 é aquele verdadeiro, uma vez que equivale ao cenário que foram geradas as amostras.

Para obtenção dos resultados apresentados neste trabalho, utilizou-se do software R Core Team (2024) na versão 4.4.0 instalado em um notebook com processador Intel Core i5-1135G7 @ 2.40GHz 2.42 GHz, 8 GB de memoria RAM DDR4, sistema operacional de 64 bits - Windows 11 Home. Ressalta-se que, para o procedimento de estimação por máxima verossimilhança, via algoritmo EM, no modelo de misturas finitas de especialistas sob a distribuição ST foi programado no R e usou-se (2.4) como critério de parada do algoritmo EM com $\epsilon = 10\text{-}4$.

5.1 RESULTADOS – CENÁRIO 1

Observou-se que o algoritmo EM proposto foi capaz de recuperar os verdadeiros valores dos parâmetros. Quando o tamanho amostral aumenta, nota-se que as estimativas dos parâmetros de interesse se aproximam dos seus respectivos valores verdadeiros com redução da variabilidade. Desta maneira, nas Figuras 2, 3 e 4 a seguir, encontram-se os *boxplots* das estimativas para cada parâmetro do modelo estudado, com G = 2, de acordo com as configurações mostradas anteriormente sob o cenário 1 ajustado. É importante salientar que a construção dos *boxplots* foi realizada excluindo os outliers com o intuito de aprimorar a visualização dos resultados. Adicionalmente, na Figura 4 percebe-se que a recuperação dos valores dos parâmetros de escala, assimetria e curtose, principalmente, para n pequeno (neste estudo, n=100) é instável, ou seja, há muita variabilidade presente.

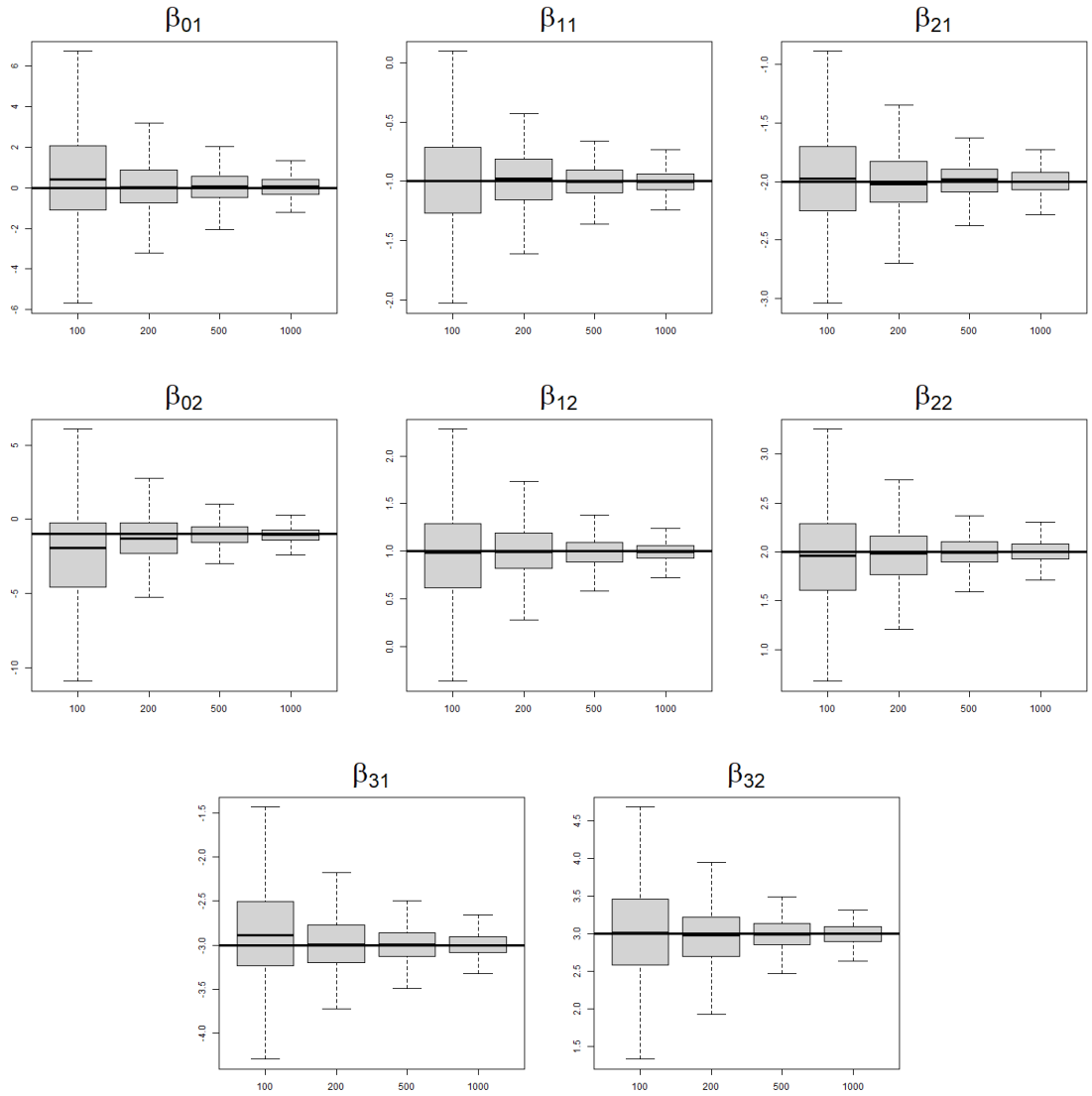


Figura 2 – *Boxplots* das estimativas dos coeficientes $\boldsymbol{\beta}$. Linha horizontal preta indica o valor real do parâmetro.

Fonte: Elaboração Própria.

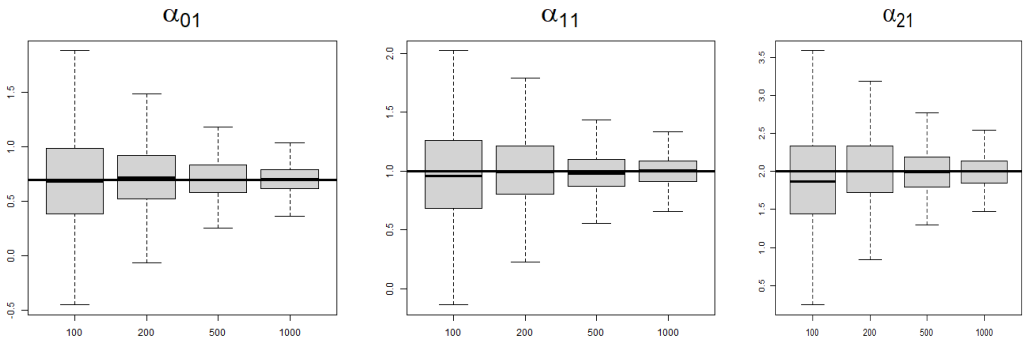


Figura 3 – *Boxplots* das estimativas dos coeficientes $\boldsymbol{\alpha}$. Linha horizontal preta indica o valor real do parâmetro.

Fonte: Elaboração Própria.

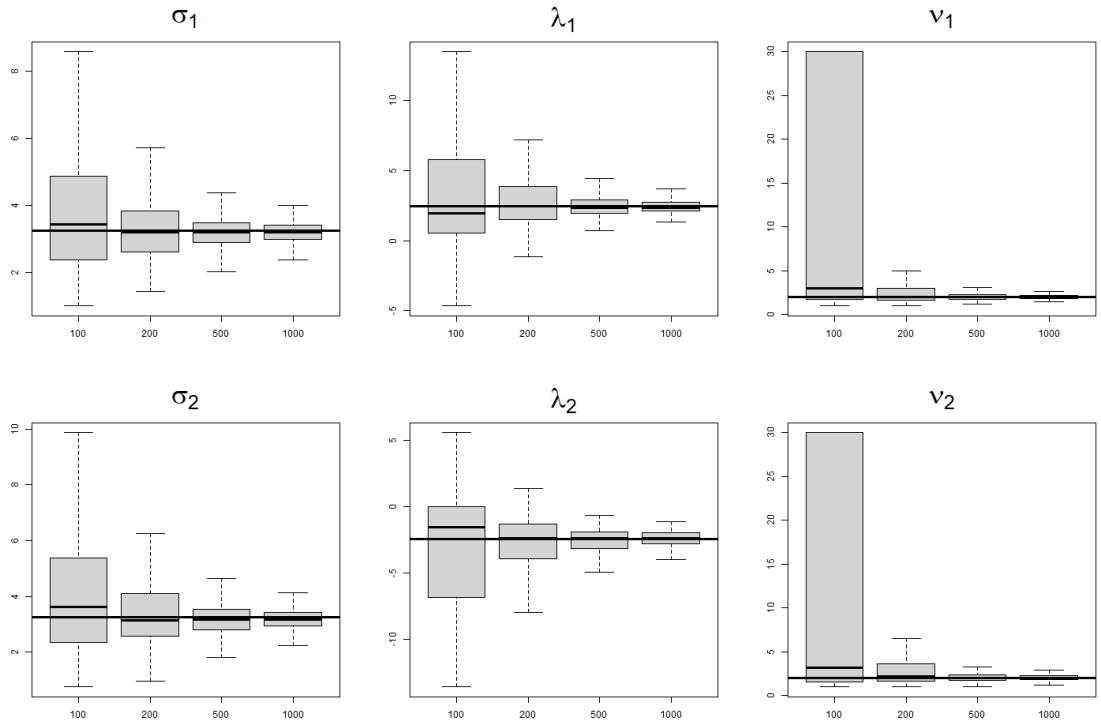


Figura 4 – *Boxplots* das estimativas de σ^2, λ e ν. Linha horizontal preta indica o valor real do parâmetro.

Fonte: Elaboração Própria.

Finalmente, os resultados para os demais cenários se encontram no Apêndice A, a partir dos boxplots dos cenários 2, 3 e 4 nota-se um comportamento similar ao cenário 1, em que o modelo foi capaz de recuperar os verdadeiros valores dos parâmetros. Observou-se uma melhora significativa ao aumentar o tamanho amostral com redução na variabilidade. Além disso, como ocorreu para o cenário 1, a recuperação dos valores dos parâmetros de escala, assimetria e curtose, principalmente, para n=100 é instável, devido ao fato de termos poucas informações para uma quantidade grande de parâmetros.

Um fator que influencia essa performance ruim para tamanhos amostrais pequenos diz respeito a quantidade de observações por parâmetro, usualmente temos que um número mínimo ideal está entre 10 e 20 observações, no entanto, considerando que precisamos estimar 17 parâmetros, teríamos que ter no mínimo de 170 a 340 observações para uma boa recuperação dos valores verdadeiros. Nesse sentido, na Seção 5.2 uma comparação entre os cenários é realizada a partir do desvio mediano absoluto das estimativas.

5.2 Comparação entre cenários - Desvio Mediano Absoluto

Considerando que o cenário 4 reflete o cenário real da simulação, procederemos com a análise de sua eficácia prática ao impor as restrições. O objetivo é verificar se, de fato, há uma melhoria na estimação dos parâmetros, ou seja, se conseguimos um

desempenho superior na recuperação dos valores dos parâmetros, levando em conta tanto o viés mediano quanto a variabilidade. Nesse sentido, para realizar a comparação entre os diferentes cenários, foi calculado o Desvio Mediano Absoluto (MAD), medida robusta a valores discrepantes, encontrada em Hoaglin, Mosteller e Tukey (2000).

$$MAD = Mediana\left|\hat{\theta} - \theta\right| \tag{5.1}$$

Nas Figuras 5 e 6, observa-se que não há diferenças significativas no desempenho entre os cenários analisados, quando considerados os parâmetros de locação (ou seja, os coeficientes de regressão). É notável que o MAD (Desvio Mediano Absoluto) é maior para os estimadores dos coeficientes associados aos interceptos, o que era esperado, dado que estamos lidando com uma distribuição assimétrica. Como mencionado na Proposição 2.3.2, temos que $\mathbb{E}[Y] = \mu + \sqrt{\frac{2}{\pi}}\mathbb{E}[U^{1/2}]\Delta$, ou seja, não estamos considerando um modelo centrado.

Na Figura 7, observa-se uma diferença significativa entre os cenários. Para o parâmetro σ^2, os cenários 2 e 4, que impõem restrição na escala, apresentaram os melhores resultados. Quanto ao parâmetro ν, os cenários 3 e 4, que estabelecem ν igual entre os grupos, foram os que obtiveram os melhores desempenhos. Por fim, em relação ao parâmetro λ, nota-se que ele foi o mais afetado em amostras pequenas. Assim como para σ^2, os cenários 2 e 4 mostraram-se superiores aos demais.

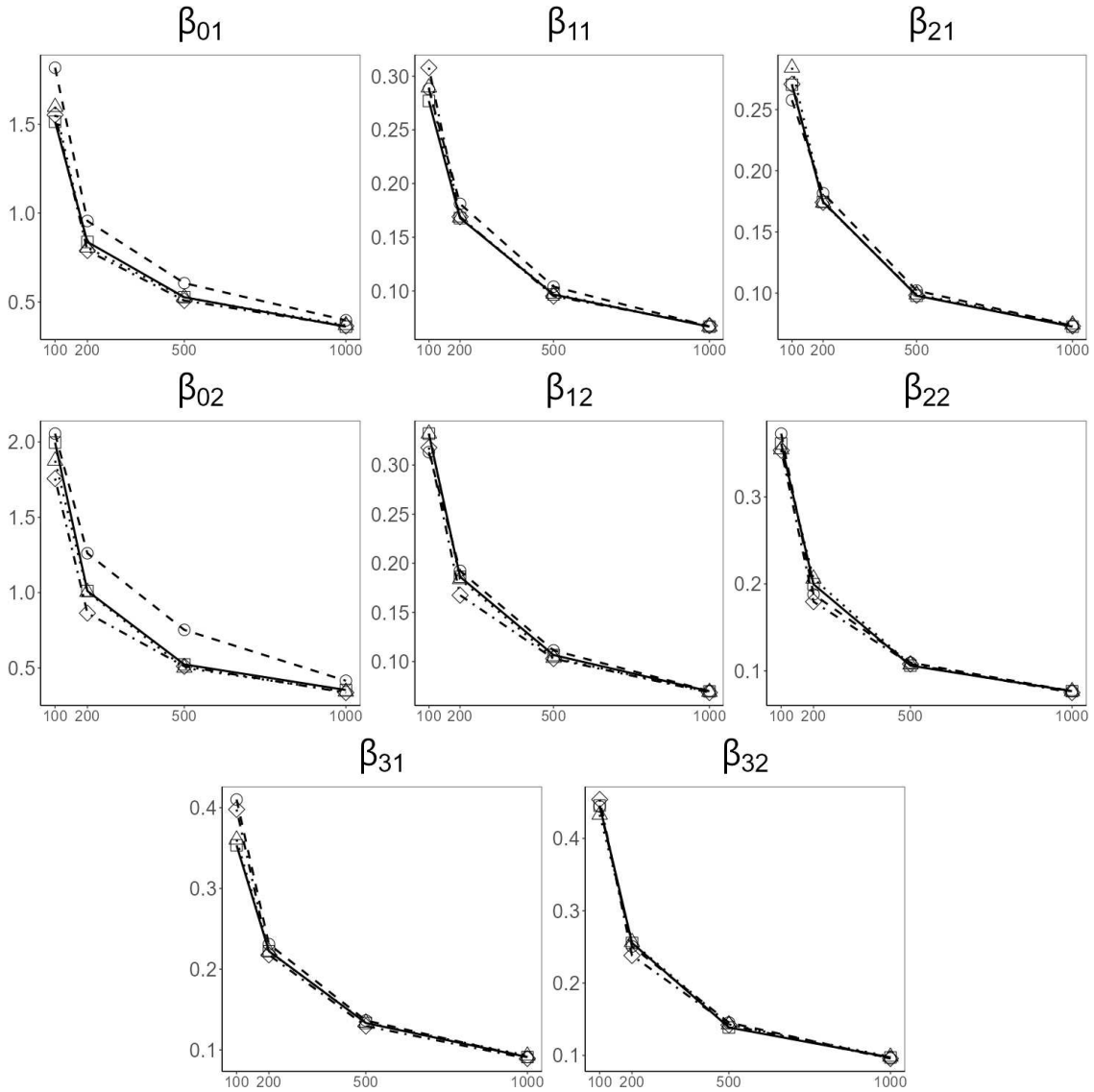


Figura 5 – Desvio mediano absoluto das estimativas dos coeficientes $\boldsymbol{\beta}$. Cenário 1: quadrado com linha sólida; 2: círculo com linha tracejada; 3: triângulo com linha pontilhada; 4: losango com linha ponto-traço.

Fonte: Elaboração Própria.

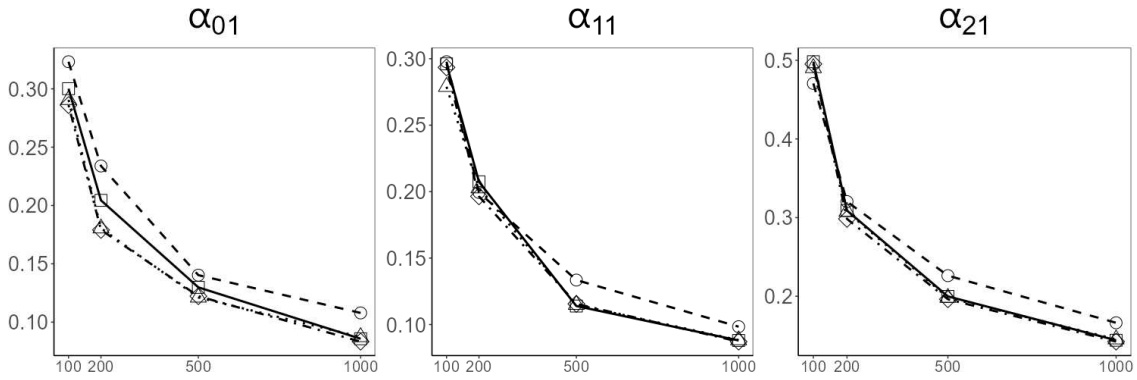


Figura 6 – Desvio mediano absoluto das estimativas dos coeficientes $\boldsymbol{\alpha}$.Cenário 1: quadrado com linha sólida; 2: círculo com linha tracejada; 3: triângulo com linha pontilhada; 4: losango com linha ponto-traço.

Fonte: Elaboração Própria.

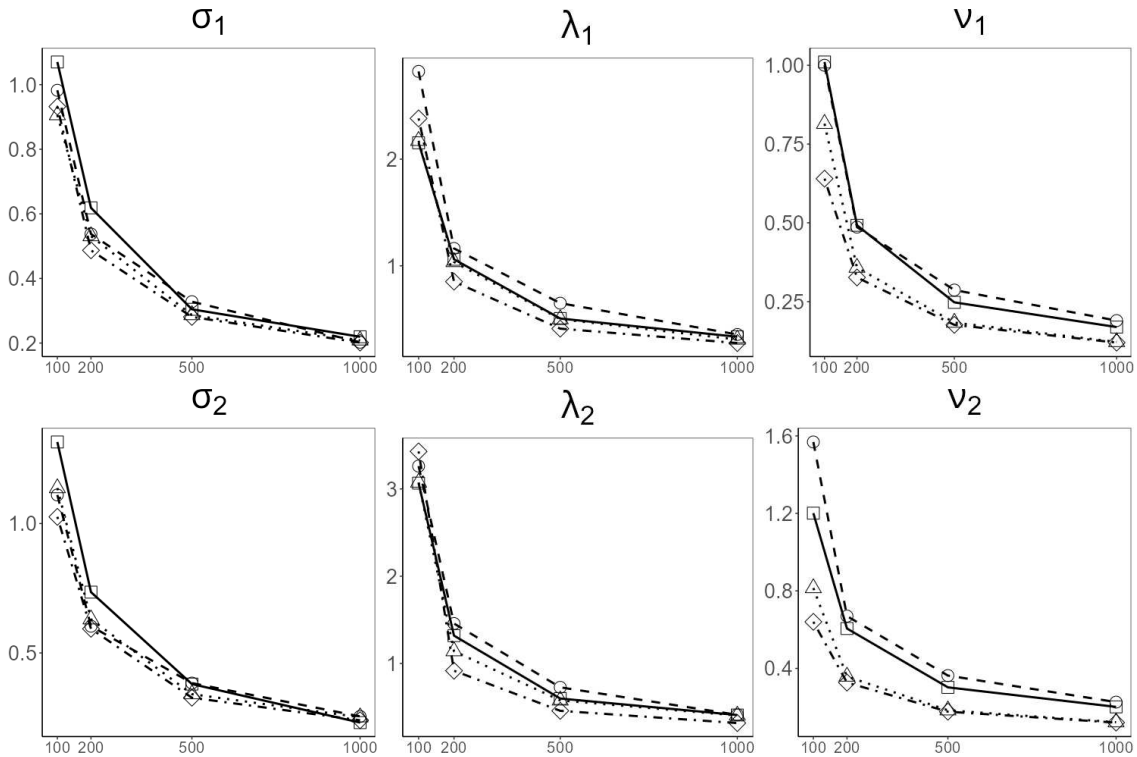


Figura 7 – Desvio mediano absoluto das estimativas de σ^2, λ e ν. Cenário 1: quadrado com linha sólida; 2: círculo com linha tracejada; 3: triângulo com linha pontilhada; 4: losango com linha ponto-traço.

Fonte: Elaboração Própria.

5.3 BIC

Fixando G=2, nesta seção, utilizamos o critério de informação bayesiano (BIC) para avaliar qual modelo, ou seja, qual cenário de acordo com o princípio da parcimônia, se ajusta melhor aos dados. A Tabela 1 apresenta a frequência com que cada cenário se

ajustou de forma mais adequada aos dados, considerando um total de 500 replicações. A partir da análise da Tabela 1, observa-se que o modelo simulado, correspondente ao cenário 4, foi o mais frequentemente selecionado em todos os tamanhos amostrais. Além disso, nota-se que o desempenho do BIC se torna mais eficiente à medida que o tamanho amostral aumenta.

n	Cenário 1	Cenário 2	Cenário 3	Cenário 4
100	28	26	109	337
200	9	8	55	428
500	3	9	9	479
1000	3	10	6	481

Tabela 1 – Quantidade de amostras com menor BIC por cenário.

Fonte: Elaboração Própria.

6 APLICAÇÕES

Neste capítulo, duas aplicações em dados reais serão apresentadas com o propósito de ilustrar os modelos de misturas finitas de especialistas e os resultados inferenciais aqui estudados. Nesse trabalho, foram considerados os seguintes conjuntos de dados: *Palmer Penguins* e *Etanol* que se destacam por terem sido utilizados em ilustrações numéricas por um grande número de programas estatísticos e autores no contexto de misturas finitas e/ou agrupamentos. Assim, pode-se analisar a flexibilidade dos modelos de misturas de especialistas (MoE) comparando-os com outros modelos já conhecidos na literatura, por exemplo, os modelos de misturas finitas (MF), ambos sob as distribuições normal, *t-Student*, skew-normal (SN) e skew-t (ST).

6.1 *PALMER PENGUINS*

O conjunto de dados, disponível no pacote palmerpenguins do software R Core Team (2024), foi extraído do estudo de Horst, Hill e Gorman (2020) e contém 344 observações. Essas observações registram as medidas do comprimento e da profundidade do bico e da nadadeira de pinguins adultos das espécies Adélie, Chinstrap e Gentoo. Os pinguins foram observados em ilhas do Arquipélago de Palmer, localizado nas proximidades da Estação Palmer, na Antártica. Os dados foram coletados e disponibilizados pela Dra. Kristen Gorman, em colaboração com o Programa de Pesquisa Ecológica de Longo Prazo (LTER) da Estação Palmer. Neste estudo, a variável resposta escolhida foi o comprimento do bico, medido em milímetros, enquanto a variável explicativa foi o comprimento da nadadeira, também em milímetros, denotada como x_1. Além disso, no contexto de especialistas, decidimos incluir informações adicionais sobre as covariáveis, como o comprimento da nadadeira (x_1) e a ilha em que as medições foram feitas (Dream, Torgensen e Biscoe), representada por x_2. Sabemos que há três espécies diferentes de pinguins (Adelie, Gentoo e Chinstrap) no conjunto de dados, o que levou à análise desses dados dentro de um modelo de misturas finitas , com G = 3 componentes. Foram exploradas diferentes opções de distribuições, tanto simétricas quanto assimétricas, além de variações nas restrições (Seção 4.5), com o objetivo de identificar modelos mais parcimoniosos. A avaliação da qualidade dos modelos ajustados foi feita com base nos valores do Critério de Informação Bayesiana (BIC).

	BIC			
	Sem Restrição		Γ com restrição	
Modelo	MF	MoE	MF	MoE
Normal	1933.318	1828.126	1926.916	1822.532
t-Student	1929.565	1820.105	1920.525	1801.039
skew-normal	1950.312	1850.583	1944.571	1826.379
skew-t	1959.248	1858.389	1952.851	1837.193

	BIC			
	ν com restrição		Γ e ν com restrição	
Modelo	MF	MoE	MF	MoE
t-Student	1919.425	1812.620	1910.691	**1794.110**
skew-t	1950.515	1854.378	1941.872	1829.938

Tabela 2 – BIC Pinguins. MoE: Mixture of *experts* (com especialistas); MF: Misturas Finitas (sem especialistas)

.

Na Tabela 2, observe que um valor baixo de BIC está associado a um modelo melhor ajustado. O modelo de três componentes misturas de especialistas sob a distribuição *t-Student*, com restrição em Γ e ν, fornece um ajuste melhor para os dados.

Parâmetro	Estimativa	Erro Padrão	valor-p	Significância
β_{01}	37.955	14.458	0.004	**
β_{02}	-18.072	5.534	0.000	***
β_{13}	-2.466	12.142	0.420	
$\beta_{11}^{(nadadeira)}$	0.001	0.079	0.494	
$\beta_{12}^{(nadadeira)}$	0.301	0.026	0.000	***
$\beta_{23}^{(nadadeira)}$	0.262	0.061	0.000	***
σ	2.176	0.117		
ν	9.915			

Tabela 3 – Estimativas de β,σ e ν para o modelo selecionado na aplicação *Palmer Penguins*.

A Tabela 3 apresenta as estimativas por máxima verossimilhança dos parâmetros nos modelos de misturas finitas de especialistas com G = 3 componentes, sob a distribuição *t-Student*, e seus respectivos erros padrão aproximados usando a matriz hessiana (H) obtida através da função *hessian* do pacote *numDeriv* do R, que utiliza o método desenvolvido em Richardson (1910), ou seja, $EP = \sqrt{diag(-H^{-1})}$. Com base nas estimativas dos parâmetros, conforme observado na Tabela 3, verificou-se que o comprimento da nadadeira foi significativo apenas para os grupos 2 (espécie Gentoo) e 3 (espécie Chinstrap). Em ambos os grupos, o valor foi positivo, o que sugere uma relação positiva entre o comprimento do bico e o comprimento da nadadeira nesses grupos.

A Tabela 4 apresenta as estimativas dos parâmetros por máxima verossimilhança na parte de especialistas, nos modelos de misturas finitas com $G = 3$ componentes, sob a distribuição *t-Student*, juntamente com seus respectivos erros padrão. Observa-se que a estimativa do coeficiente associado ao comprimento da nadadeira foi negativa e significativa, a um nível de 5%, no grupo 1, quando comparado ao grupo 3 (grupo de referência, correspondente à espécie Chinstrap). Isso indica que, quanto maior o comprimento da nadadeira, menor a probabilidade de um indivíduo ser classificado no grupo 1. Adicionalmente, em relação à variável x_2 (ilha), observa-se que apenas a estimativa

Parâmetro	Estimativa	*Odds Ratio*	Erro Padrão	valor-p	Significância
$\alpha_{11}^{(nadadeira)}$	-0.204	0.815	0.021	0.000	***
$\alpha_{12}^{(nadadeira)}$	0.048	1.049	0.102	0.320	
$\alpha_{21}^{(Dream)}$	-3.456	0.032	1.481	0.010	**
$\alpha_{22}^{(Dream)}$	-5.864	0.003	1.518	0.000	***
$\alpha_{31}^{(Torgersen)}$	-0.069	0.933	1.796	0.485	
$\alpha_{32}^{(Torgersen)}$	-0.864	0.421	1.865	0.322	

Tabela 4 – Estimativas de α para o modelo selecionado na aplicação *Palmer Penguins*.

do coeficiente associado à ilha Dream foi significativa a um nível de 5%. Isso indica que, ao pertencer a essa ilha, a probabilidade de um pinguim ser classificado nos grupos 1 e 2 diminui em comparação com o grupo 3. Essas relações serão analisadas graficamente e, para comparar as vantagens dos modelos de misturas finitas de especialistas de regressão, realizaremos uma análise com o mesmo modelo, mas sem considerar a parte dos especialistas. Além disso, as retas de regressão serão comparadas, considerando a divisão por espécie previamente conhecida.

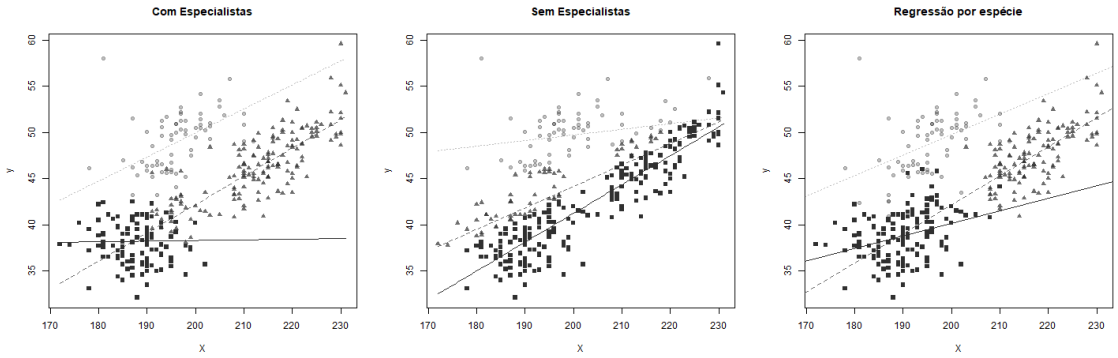


Figura 8 – Retas de regressão por grupo. Grupo 1(espécie *Adelie*): quadrado com linha sólida; 2(*Gentoo*): triângulo com linha tracejada; 3(*Chinstrap*): círculo com linha pontilhada.

Fonte: Elaboração Própria.

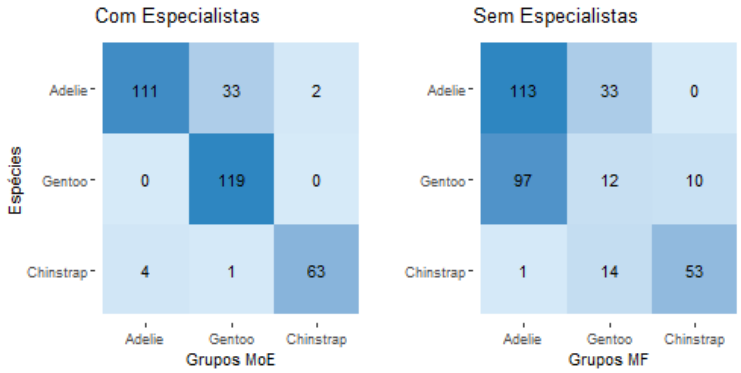


Figura 9 – Matrizes de confusão entre os grupos encontrados e as espécies reais.

Fonte: Elaboração Própria.

A partir da Figura 8, é possível observar claramente o comportamento dos três grupos no contexto dos modelos de misturas finitas de especialistas de regressão. No caso do grupo 1, o coeficiente linear não foi significativo, enquanto para os outros grupos, ele foi. Em relação aos grupos 2 e 3, destaca-se a diferença nas estimativas dos interceptos, sendo o grupo 3 associado a um valor maior, o que indica a presença de bicos maiores nesse grupo para um mesmo tamanho de nadadeira.

Além disso, pela Figura 9, ao compararmos o ajuste dos modelos de misturas finitas de regressão (sem especialistas, ou seja, sem considerar informações de covariáveis nas proporções dos grupos), fica evidente o impacto da introdução dessas novas informações na classificação dos grupos. Considerando a métrica de classificação como a acurácia (percentual de acertos em relação ao total de observações), temos que o caso com especialistas apresentou 88% de acurácia contra 54% do modelo sem especialistas. Isso possibilita uma interpretação mais simples e uma classificação mais eficiente, especialmente quando comparado ao gráfico das três espécies (gráfico à direita).

6.2 *ETANOL*

Nesta aplicação, foram utilizados dados de um experimento industrial com o objetivo de estudar o escapamento de um motor experimental de um cilindro que usa etanol como combustível, contendo 88 observações, obtidos em Brinkman (1981) e citado em Ruppert, Wand e Carroll (2003). A variável resposta, denotada por NOx, é a concentração de óxido de nitrogênio somada à concentração de dióxido de nitrogênio, normalizada pelo trabalho do motor, que são gases poluentes gerados pela queima de combustíveis em motores de combustão interna. O conjunto possui duas covariáveis, que são E, uma medida de riqueza das misturas de ar e combustível, enquanto C é a taxa de compressão do motor. Nesse contexto, a covariável da regressão $\mathbf{X} = (x_1)$ é dada por C, enquanto $\boldsymbol{R} = (x_1, r_1)$, em que r_1 é dado por E.

	Sem restrições							
	G = 1		G = 2		G = 3		G = 4	
Modelo	MF	MoE	MF	MoE	MF	MoE	MF	MoE
Normal	283.176	283.176	262.811	271.636	264.050	91.690	267.015	107.302
t-Student	289.449	289.449	272.319	118.139	277.977	118.112	285.514	125.608
skew-normal	287.654	287.601	271.572	116.285	277.478	98.953	284.935	120.618
skew-t	293.600	293.600	281.263	121.280	291.195	111.250	305.823	143.733

	Restrição em Γ							
	G = 1		G = 2		G = 3		G = 4	
Modelo	MF	MoE	MF	MoE	MF	MoE	MF	MoE
Normal	283.176	283.176	244.711	119.797	259.781	**87.637**	261.035	89.914
t-Student	289.449	289.449	254.440	125.591	273.655	99.770	278.940	110.915
skew-normal	287.654	287.601	253.668	108.597	273.213	95.604	278.949	109.878
skew-t	293.600	293.600	253.478	117.308	281.791	103.301	294.603	125.271

	Restrição em ν							
	G = 1		G = 2		G = 3		G = 4	
Modelo	MF	MoE	MF	MoE	MF	MoE	MF	MoE
t-Student	289.4491	289.4490	267.8269	276.7773	269.0220	109.2970	272.0889	112.2271
skew-t	293.6200	293.6244	276.8054	117.8975	282.2715	108.8416	290.0659	120.2119

	Restrição em Γ e ν							
	G = 1		G = 2		G = 3		G = 4	
Modelo	MF	MoE	MF	MoE	MF	MoE	MF	MoE
t-Student	289.449	289.449	249.953	124.267	264.706	92.001	265.893	95.209
skew-t	293.620	293.624	250.253	112.438	272.838	94.665	281.378	126.254

Tabela 5 – BIC Etanol. MoE: Mixture of *experts* (com especialistas); MF: Misturas Finitas (sem especialistas). Valor em negrito indica o menor BIC.

Na Tabela 5, uma vez que um valor baixo de BIC está associado a um modelo melhor ajustado. O modelo de três componentes misturas de especialistas sob a distribuição Normal, com restrição na escala, fornece um ajuste melhor para os dados.

Parâmetro	Estimativa	Erro Padrão	valor-p	Significância
β_{01}	13.273	0.357	0.000	***
β_{02}	-3.460	0.482	0.000	***
β_{13}	2.260	0.949	0.009	**
$\beta_{11}^{(C)}$	-10.497	0.317	0.000	***
$\beta_{12}^{(C)}$	7.057	0.704	0.000	***
$\beta_{23}^{(C)}$	1.516	1.084	0.081	.
σ	0.257	0.020		

Tabela 6 – Estimativas de β,σ e ν para o modelo selecionado na aplicação *Etanol*.

A Tabela 6 apresenta as estimativas por máxima verossimilhança dos parâmetros nos modelos de misturas finitas de especialistas com G = 3 componentes, sob a distribuição

Normal, e seus respectivos erros padrão aproximados usando a matriz hessiana (H) obtida através da função *hessian* do pacote *numDeriv* do R, que utiliza o método desenvolvido em Richardson (1910), ou seja, $EP = \sqrt{diag(-H^{-1})}$. Com base nas estimativas dos parâmetros, conforme observado na Tabela 6, verificou-se que a taxa de compressão do motor foi significativa, a um nível de 10%, em todos os grupos, apresentando valores distintos entre os grupos. No grupo 1 a relação foi fortemente negativa, no 2 foi fortemente positiva e no 3 foi fracamente positiva.

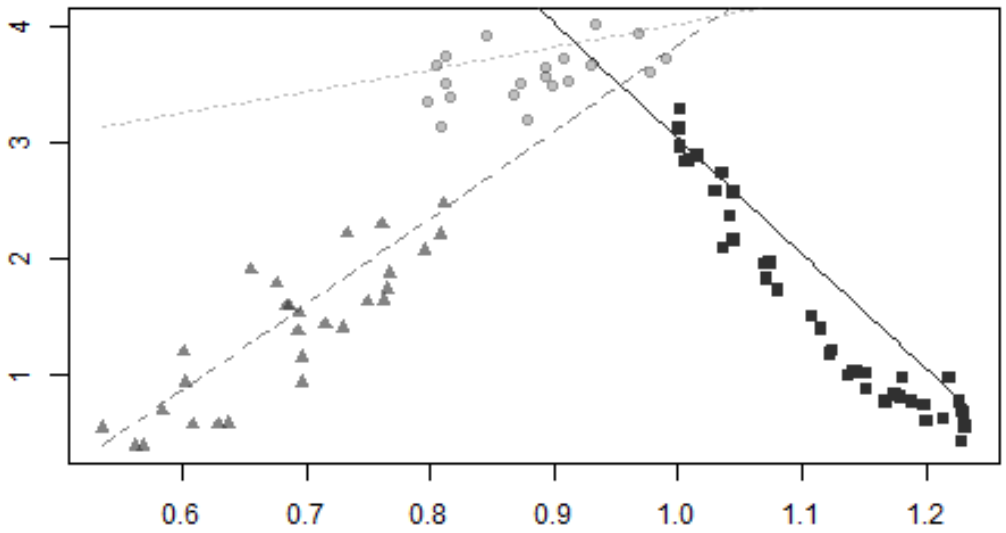


Figura 10 – Retas de regressão por grupo. Grupo 1: quadrado com linha sólida, 2: triângulo com linha tracejada, 3: círculo com linha pontilhada.

Fonte: Elaboração Própria.

A partir da Figura 10, é possível observar claramente o comportamento dos três grupos no contexto dos modelos de misturas finitas de especialistas de regressão. No caso do grupo 1, o coeficiente linear foi negativo, exibindo uma relação negativa entre a taxa de compressão e os gases emitidos. No grupo 2, o coefiente foi positivo mostrando uma forte relação entre a taxa de compressão e os gases emitidos. Por fim, o grupo 3 apresenta uma relação positiva fraca, apresentando um intercepto mais alto.

A Tabela 7 apresenta as estimativas dos parâmetros por máxima verossimilhança na parte de especialistas, nos modelos de misturas finitas com $G = 3$ componentes, sob a distribuição *Normal*, juntamente com seus respectivos erros padrão. Nota-se que a estimativa do coeficiente associado a taxa de compressão foi significativa a um nível de 5% apenas para o grupo 1 em relação ao grupo 3, em que, mantendo E constante, quanto maior

Parâmetro	Estimativa	*Odds Ratio*	Erro Padrão	valor-p	Significância
$\alpha_{11}^{(C)}$	-0.284	0.753	0.192	0.069	.
$\alpha_{12}^{(C)}$	-0.323	0.724	0.164	0.024	*
$\alpha_{21}^{(E)}$	32.836	1.82×10^{14}	12.452	0.004	**
$\alpha_{22}^{(E)}$	-29.343	0	8.939	0.000	***

Tabela 7 – Estimativas de α para o modelo selecionado na aplicação *Etanol*.

a taxa de compressão menor a probabilidade de estar no grupo 1 em relação ao grupo 3. No caso da variável E, a qualidade da mistura ar e combustível, nota-se que as estimativas foram significativas, a um nível de 5%, apresentando uma *Odds Ratio* muito alta do grupo 1 em relação ao grupo 2, e uma *Odds Ratio* nula do grupo 2 em relação ao grupo 3. Logo, esse resultado indica que, quanto maior a qualidade da mistura, probabilidade de ser do grupo 1 aumenta, enquanto quanto menor a qualidade probabilidade de ser do grupo 2 aumenta, por fim, um valor intermediário de qualidade, aumenta a probabilidade de ser do grupo 3. A fim de verificar essa relação, a Figura 11 mostra os histogramas da qualidade da mistura para cada grupo separadamente.

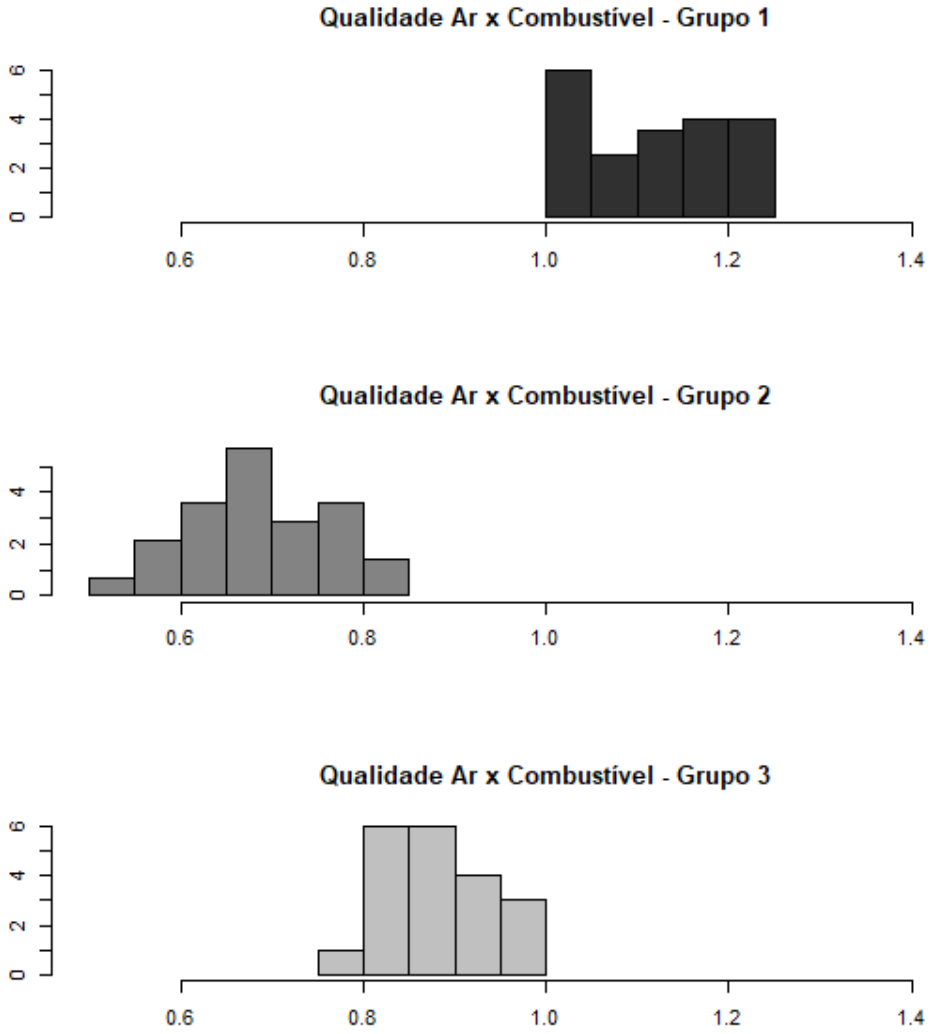


Figura 11 – Histogramas da qualidade da mistura ar e combustível para cada grupo.

Fonte: Elaboração Própria.

A partir da Figura 11 fica evidente a relação encontrada nos coeficientes $\boldsymbol{\alpha}$, em que, o grupo 1 apresenta os valores com maior qualidade, o grupo 2 apresenta os piores valores de qualidade, enquanto o grupo 3 apresenta valores intermediários de qualidade da mistura ar e combustível.

7 CONCLUSÃO

O foco deste trabalho foi apresentar as misturas finitas de especialistas de modelos de regressão robustos, ou seja, que são capazes de lidar com problemas de heterogeneidade não observada, dados assimétricos e/ou com caudas pesadas. Esta classe de modelo ainda está consolidando sua relevância e esperamos que se mostre útil em pesquisas futuras. Acreditamos que, por meio deste trabalho, seja possível despertar o interesse de pesquisadores e estudantes, contribuindo para o avanço do campo.

Nesse contexto, algoritmos do tipo EM foram obtidos, a partir de uma representação estocástica dos modelos. A busca por maior parcimônia foi discutida. O uso do critério de informação BIC foi utilizado na seleção de modelos. Finalmente, estudos de simulação foram realizados a fim de verificar o poder de recuperação dos parâmetros, enquanto, dois conjuntos de dados reais foram utilizados com objetivo de mostrar a aplicação destes modelos.

Como próximos passos, podemos citar a possibilidade de estudar outras distribuições da família mistura escala *skew-normal*, como a *skew-normal contaminada* e *skew-slash*, enquanto a extensão para respostas multivariadas e com a presença de censura possibilita a aplicação em contextos mais complexos. A avaliação por meio de validação cruzada permite avaliar melhor o desempenho do modelo. Além disso, a inclusão de técnicas de seleção de variáveis automáticas, como Regressão Lasso pode ser muito útil e interessante

8 ASPECTOS COMPUTACIONAIS

Todos os algoritmos realizados neste trabalho foram implementados no sofware R com auxílio dos pacotes: *sn*, desenvolvido em Azzalini (2022), para calcular as funções de densidades das distribuições assimétricas, e *Matrix* no cálculo de matrizes esparsas; Bates et al. (2010). Além disso, nos estudos de simulação o pacote *mixsmsn* desenvolvido em Prates, Lachos e Cabral (2013), foi utilizado na geração de valores aleatórios. Todos os códigos utilizados, algoritmos, estudos de simulação e aplicações podem ser encontrados no repositório do *GitHub* que pode ser acessado por MixRegEM .

Em relação à utilização do algoritmo é importante comentar que os problemas vistos nas simulações com os parâmetros λ e ν podem ser minimizados ao diminuir a quantidade máxima de iterações, uma vez que nos testes realizados foi observado que o parâmetro λ tem a tendência de ir para ∞ ou $-\infty$ para tamanhos amostrais pequenos, enquanto ν tende a seu valor máximo de busca que foi configurado para 30. Além disso, vale ressaltar que os algoritmos presentes possuem outras opções de ajustes que não foram apresentadas neste trabalho e que continuam em constante desenvolvimento.

REFERÊNCIAS

ALAMICHEL, P.; SMITH, J. M. Informed bayesian finite mixture models with asymmetric dirichlet priors. *Journal of Applied Statistics*, v. 57, p. 1023–1041, 2023.

ANDREWS, D. F.; MALLOWS, C. L. Scale mixtures of normal distributions. *Journal of the Royal Statistical Society: Series B (Methodological)*, Wiley Online Library, v. 36, n. 1, p. 99–102, 1974.

AZZALINI, A. sn: The skew-normal and related distributions such as the skew-t and the sun. *R package version*, v. 2, n. 0, 2022.

AZZALINI, A.; VALLE, A. D. The multivariate skew-normal distribution. *Biometrika*, Oxford University Press, v. 83, n. 4, p. 715–726, 1996.

BASSO, R. M. et al. Robust mixture modeling based on scale mixtures of skew-normal distributions. *Computational Statistics & Data Analysis*, Elsevier, v. 54, n. 12, p. 2926–2941, 2010.

BATES, D. et al. Matrix: Sparse and dense matrix classes and methods. *R package version 0.999375-43, URL http://cran. r-project. org/package= Matrix*, 2010.

BRANCO, M. D.; DEY, D. K. A general class of multivariate skew-elliptical distributions. *Journal of Multivariate Analysis*, Elsevier, v. 79, n. 1, p. 99–113, 2001.

BRINKMAN, N. D. Ethanol fuel—single—cylinder engine study of efficiency and exhaust emissions. *SAE transactions*, JSTOR, p. 1410–1424, 1981.

BYRD, R. H. et al. A limited memory algorithm for bound constrained optimization. *SIAM Journal on scientific computing*, SIAM, v. 16, n. 5, p. 1190–1208, 1995.

DÁVILA, V. H. L.; CABRAL, C. R. B.; ZELLER, C. B. *Finite mixture of skewed distributions*. [S.l.]: Springer, 2018.

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.

DEPRAETERE, N.; VANDEBROEK, M. Order selection in finite mixtures of linear regressions. *Statistical Papers*, v. 55, p. 871–911, 2014.

DOE, J.; SMITH, A. Applications of skew-normal finite mixture models in finance. *Journal of Financial Analysis*, 2023. Examines the use of skew-normal finite mixture models for modeling asymmetries and extreme events in financial returns. Disponível em: <https://example.com/finance2023>.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer New York, 2013. (Springer Series in Statistics). ISBN 9781489905185. Disponível em: <https://books.google.com.br/books?id=MuNmawEACAAJ>.

HOAGLIN, D. C.; MOSTELLER, F.; TUKEY, J. W. *Understanding robust and exploratory data analysis*. [S.l.]: John Wiley & Sons, 2000.

JACOBS, R. A. et al. Adaptive mixtures of local experts. *Neural computation*, MIT Press, v. 3, n. 1, p. 79–87, 1991.

LEE, S.; MCLACHLAN, G. J. Finite mixtures of multivariate skew t-distributions: some recent and new results. *Statistics and Computing*, Springer, v. 24, p. 181–202, 2014.

LIN, T. I.; LEE, J. C.; HSIEH, W. J. Robust mixture modeling using the skew t distribution. *Statistics and computing*, Springer, v. 17, p. 81–92, 2007.

LIN, T. I.; LEE, J. C.; YEN, S. Y. Finite mixture modelling using the skew normal distribution. *Statistica Sinica*, JSTOR, p. 909–927, 2007.

LIU, C.; RUBIN, D. B. The ecme algorithm: a simple extension of em and ecm with faster monotone convergence. *Biometrika*, Oxford University Press, v. 81, n. 4, p. 633–648, 1994.

LIU, F.; O'BRIEN, T. Sparse finite mixture models in high-dimensional data contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 43, n. 11, p. 3001–3014, 2021.

MAZZA, A.; PUNZO, A. Mixtures of multivariate contaminated normal regression models. *Statistical Papers*, v. 61, p. 787–822, 2020.

MCLACHLAN, G. J.; BASFORD, K. E. *Mixture models: Inference and applications to clustering*. [S.l.]: M. Dekker New York, 1988. v. 38.

MCLACHLAN, G. J.; PEEL, D. *Finite Mixture Models*. New York: Wiley-Interscience, 2000. (Wiley Series in Probability and Statistics).

MENG, X.-L.; RUBIN, D. B. Maximum likelihood estimation via the ecm algorithm: A general framework. *Biometrika*, Oxford University Press, v. 80, n. 2, p. 267–278, 1993.

MENGERSEN, K. L.; ROBERT, C.; TITTERINGTON, M. *Mixtures: estimation and applications*. [S.l.]: John Wiley & Sons, 2011.

MIRFARAH, E.; NADERI, M.; CHEN, D.-G. Finite mixture of regression models for censored data based on scale mixtures of normal distributions. *Computational Statistics Data Analysis*, v. 158, p. 107–182, 2021.

MORRISON, C.; WALTERS, P. Applications of finite mixture models in econometrics with covariates. *Econometric Reviews*, v. 41, n. 3, p. 295–315, 2022.

NELDER, J. A.; MEAD, R. A simplex method for function minimization. *The computer journal*, The British Computer Society, v. 7, n. 4, p. 308–313, 1965.

PEARSON, K. Contributions to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society of London. A*, JSTOR, v. 185, p. 71–110, 1894.

PRATES, M. O.; LACHOS, V. H.; CABRAL, C. R. B. mixsmsn: Fitting finite mixture of scale mixture of skew-normal distributions. *Journal of Statistical Software*, v. 54, p. 1–20, 2013.

R Core Team. *R: A Language and Environment for Statistical mputing*. Vienna, Austria, 2024. Disponível em: <https://www.R-oject.org/>.

RICHARDSON, L. The approximate arithmetical solution by finite differences of physical problems involving differential equations. *Phil. Trans. Roy. Sec., London. A*, v. 210, 1910.

RUPPERT, D.; WAND, M. P.; CARROLL, R. J. *Semiparametric regression*. [S.l.]: Cambridge university press, 2003.

SCHWARZ, G. Estimating the dimension of a model. *The annals of statistics*, JSTOR, p. 461–464, 1978.

SHOHAM, S. Robust clustering by deterministic agglomeration em of mixtures of multivariate t-distributions. *Pattern Recognition*, Elsevier, v. 35, n. 5, p. 1127–1142, 2002.

SMITH, A.; TAYLOR, B. Application of skew-normal finite mixture models in biomedical data analysis. *Journal of Biomedical Statistics*, 2023. Explora a aplicação em séries temporais de sinais vitais e exames clínicos. Disponível em: <https://example.com/biomedical2023>.

VEAUX, R. D. D. Finite mixtures of elliptically contoured distributions. *Journal of the American Statistical Association*, Taylor & Francis, v. 84, n. 408, p. 1116–1123, 1989.

WILLIAMS, S. J.; TANAKA, H. Robust finite mixture models for cluster analysis in complex data. *Computational Statistics Data Analysis*, v. 149, p. 106923, 2020.

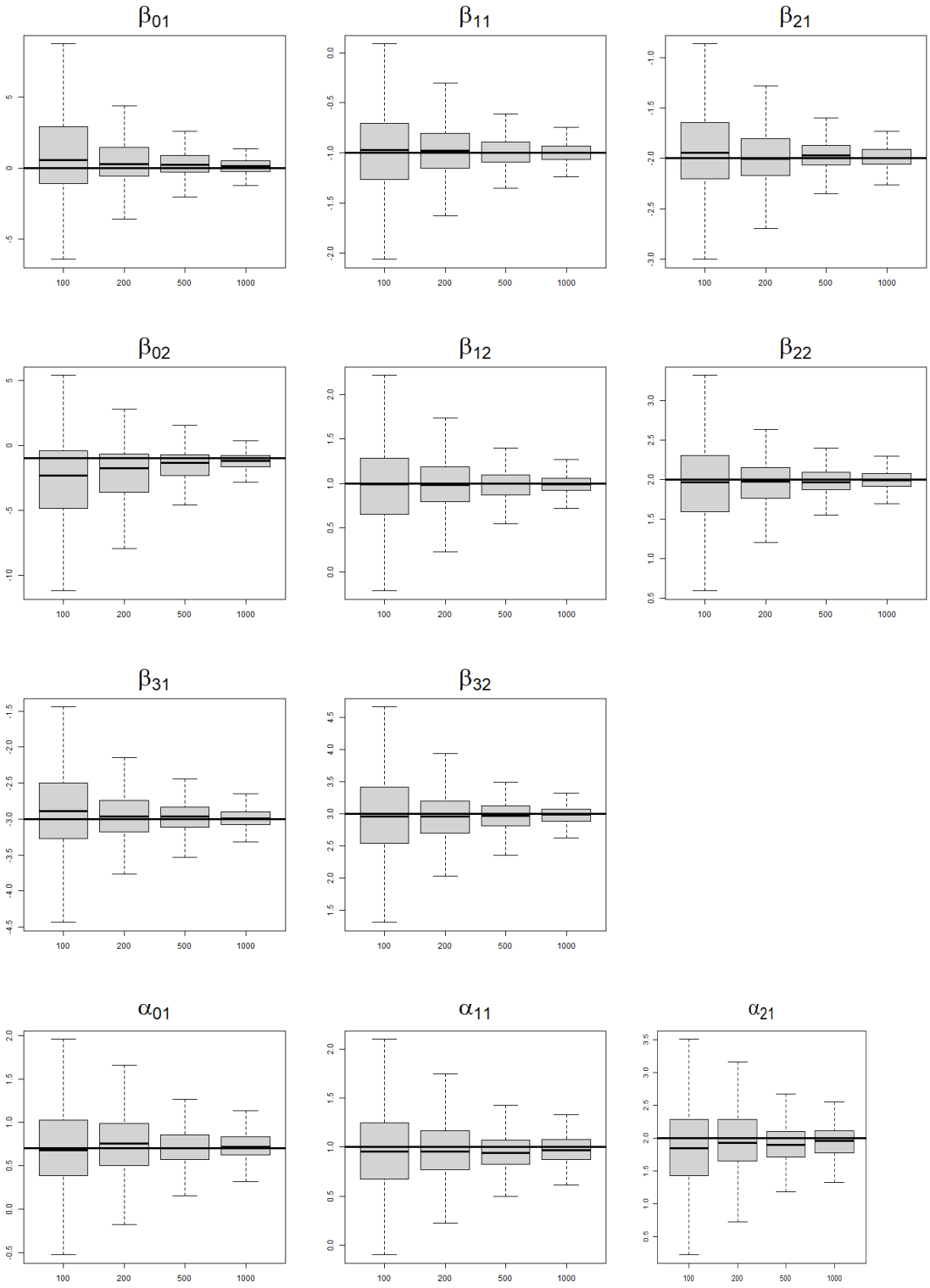
XIANG, Y. et al. Multifaceted cancer alleviation by cowpea mosaic virus in a bioprinted ovarian cancer peritoneal spheroid model. *Biomaterials*, Elsevier, p. 122663, 2024.

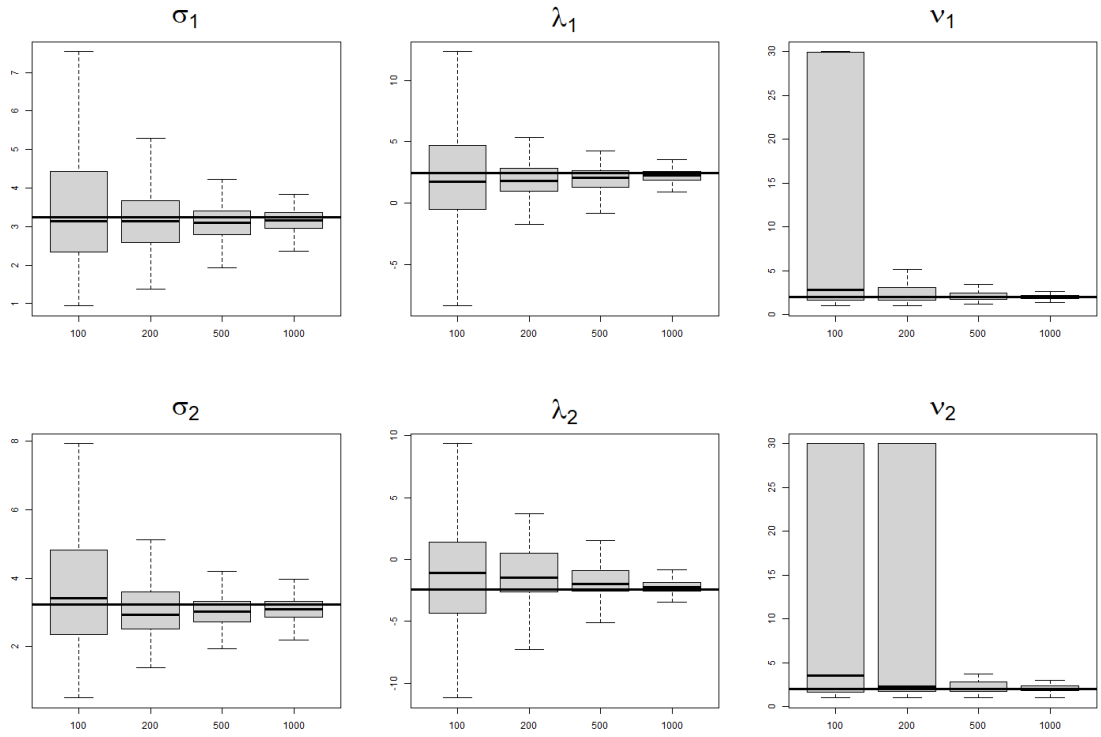
ZELLER, C. B.; CABRAL, C. R.; LACHOS, V. H. Robust mixture regression modeling based on scale mixtures of skew-normal distributions. *Test*, Springer, v. 25, p. 375–396, 2016.

ZELLER, C. B. et al. Finite mixture of regression models for censored data based on scale mixtures of normal distributions. *Advances in Data Analysis and Classification*, v. 6, p. 787–822, 2019.

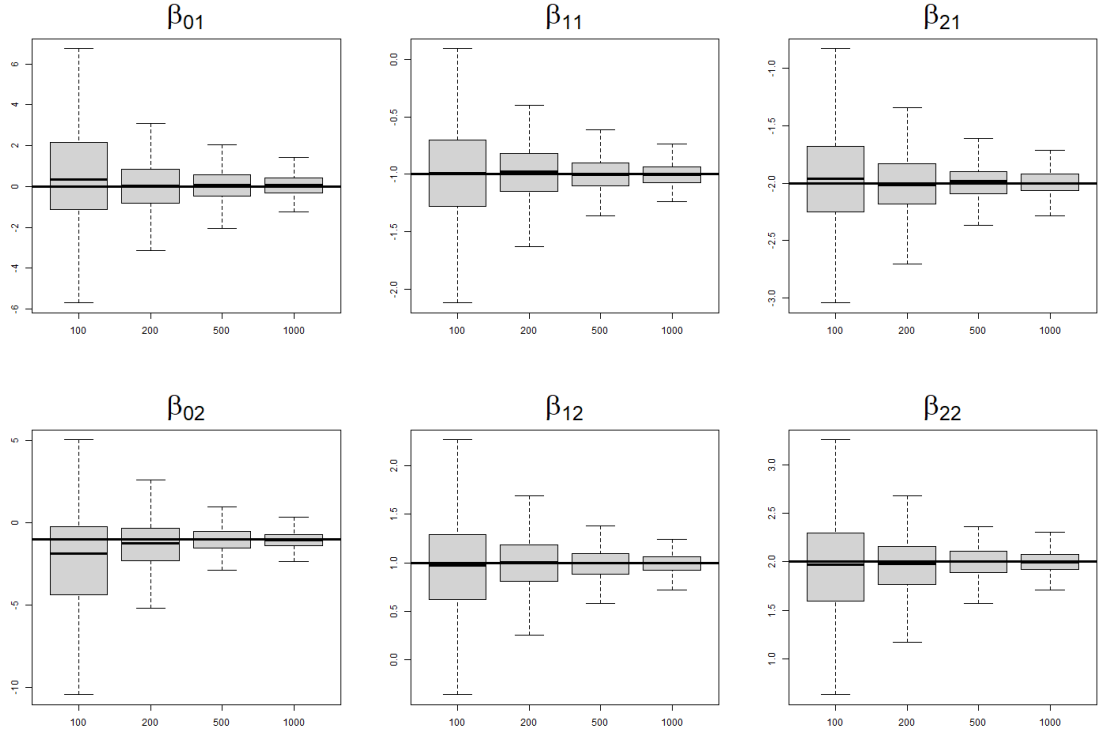
APÊNDICE A – *Boxplots* dos estudos de simulação

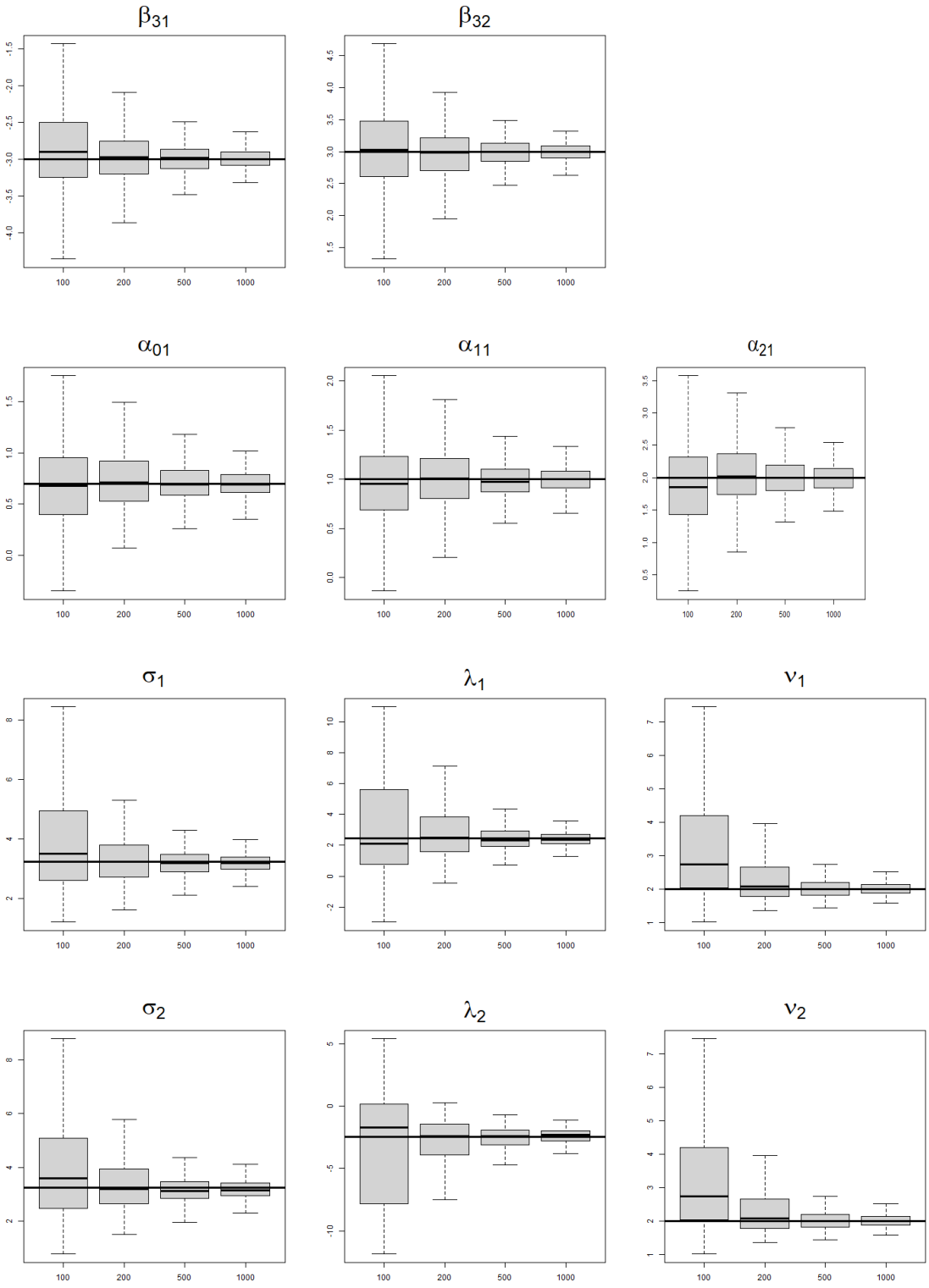
.1 Cenário 2



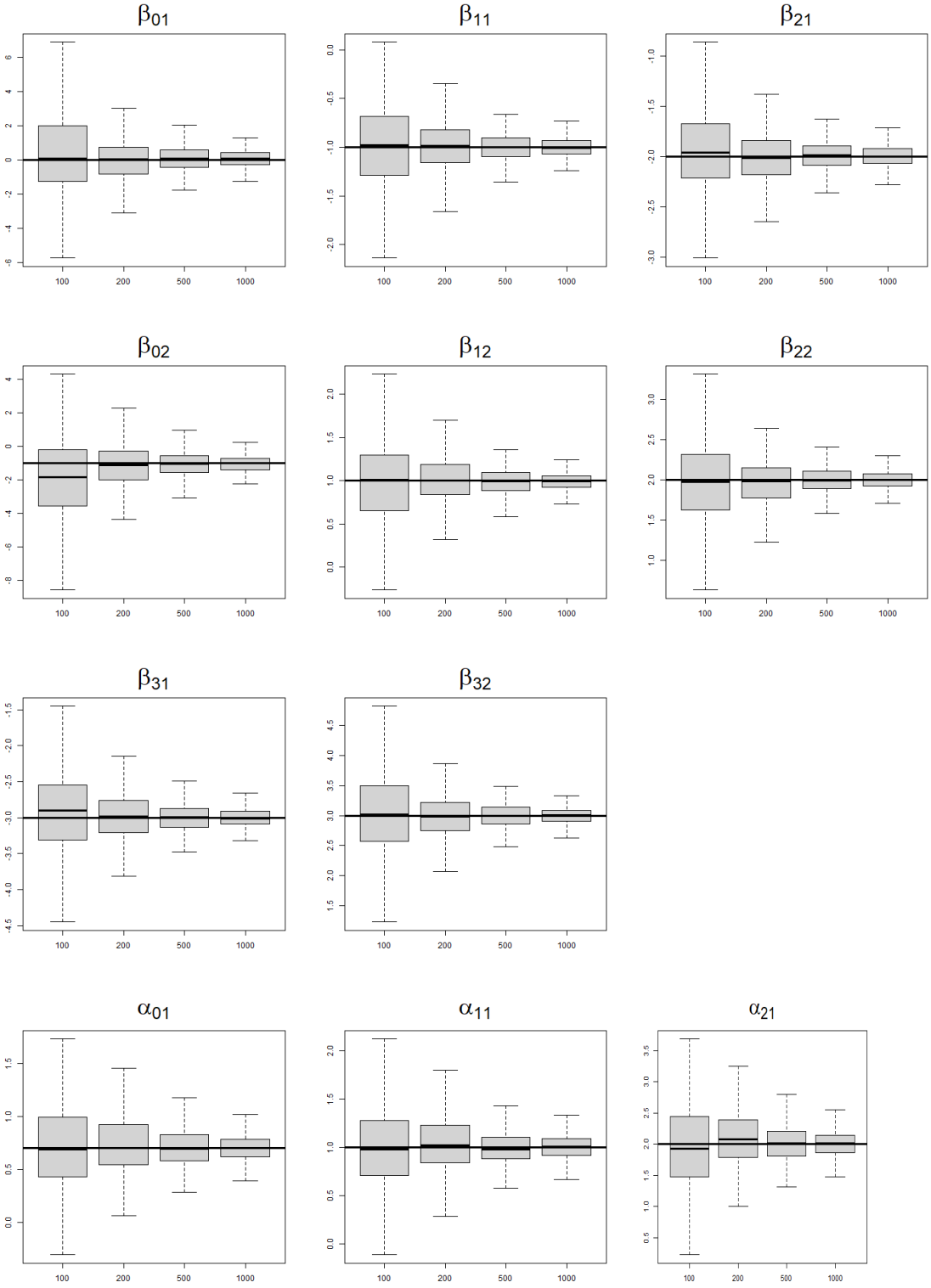

σ1
λ1
ν1
σ2
λ2
ν2
100
200
500
1000

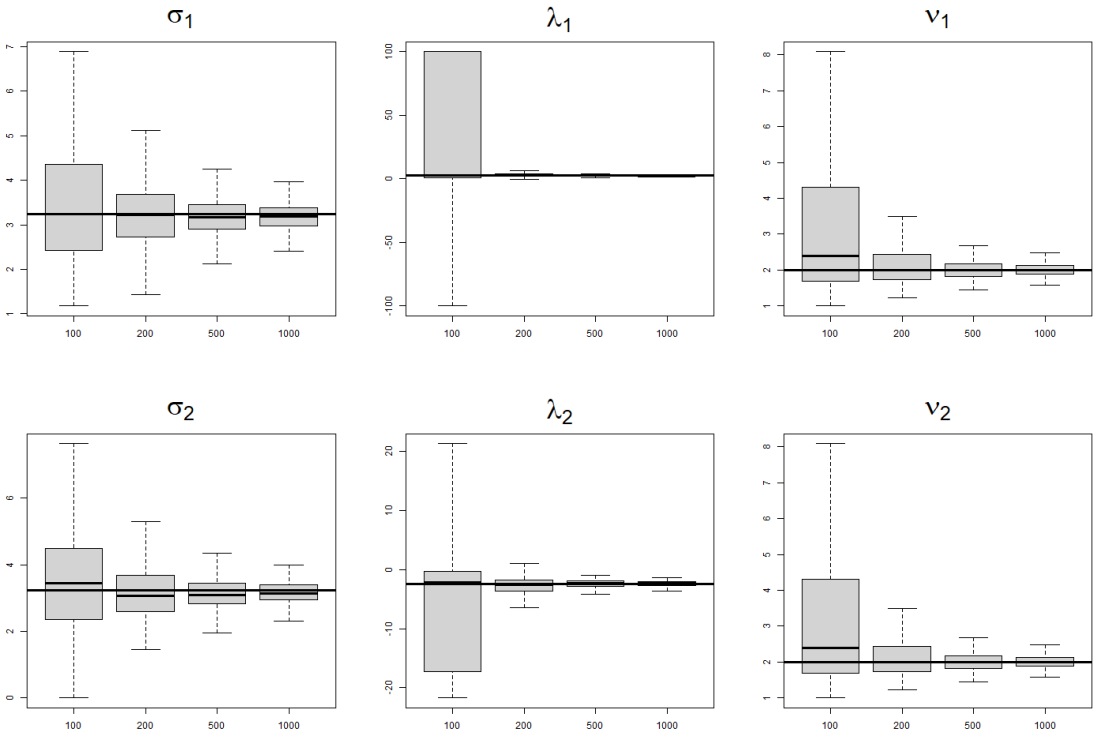
.2 Cenário 3


β01
β11
β21
β02
β12
β22
100
200
500
1000


β31
β32
α01
α11
α21
σ1
λ1
ν1
σ2
λ2
ν2
100
200
500
1000

.3 Cenário 4




σ1
λ1
ν1
σ2
λ2
ν2
100
200
500
1000