



AMANDA LENZI

**ANÁLISE DE CURVAS FUNCIONAIS AGREGADAS  
COM EFEITOS ALEATÓRIOS**

CAMPINAS  
2013





**UNIVERSIDADE ESTADUAL DE CAMPINAS**  
**INSTITUTO DE MATEMÁTICA, ESTATÍSTICA**  
**E COMPUTAÇÃO CIENTÍFICA**

**AMANDA LENZI**

**Análise de Curvas Funcionais Agregadas com Efeitos Aleatórios**

**Orientadora: Profa. Dra. Nancy Lopes Garcia**

Dissertação de mestrado apresentada ao Instituto de  
Matemática, Estatística e Computação Científica da Unicamp para  
obtenção do título de Mestra em Estatística.

ESTE EXEMPLAR CORRESPONDE À VERSÃO FINAL DA DISSERTAÇÃO  
DEFENDIDA PELA ALUNA AMANDA LENZI  
E ORIENTADA PELA PROFA. DRA. NANCY LOPES GARCIA

**Assinatura do Orientador**

A handwritten signature in blue ink, reading "Nancy Lopes Garcia", written over a horizontal line.

**CAMPINAS**  
**2013**

Ficha catalográfica  
Universidade Estadual de Campinas  
Biblioteca do Instituto de Matemática, Estatística e Computação Científica  
Ana Regina Machado - CRB 8/5467

L549a Lenzi, Amanda, 1988-  
Análise de curvas funcionais agregadas com efeitos aleatórios / Amanda Lenzi. – Campinas, SP : [s.n.], 2013.

Orientador: Nancy Lopes Garcia.  
Dissertação (mestrado) – Universidade Estadual de Campinas, Instituto de Matemática, Estatística e Computação Científica.

1. Spline, Teoria do. 2. Estatística não paramétrica. 3. Energia elétrica - Consumo. 4. Estimativa de parâmetro. I. Garcia, Nancy Lopes, 1964-. II. Universidade Estadual de Campinas. Instituto de Matemática, Estatística e Computação Científica. III. Título.

Informações para Biblioteca Digital

**Título em inglês:** Analysis of aggregated functional curves with random effect

**Palavras-chave em inglês:**

Spline theory

Nonparametric statistics

Electric power consumption

Parameter estimation

**Área de concentração:** Estatística

**Titulação:** Mestra em Estatística

**Banca examinadora:**

Nancy Lopes Garcia [Orientador]

Mariana Rodrigues Motta

Hedibert Freitas Lopes

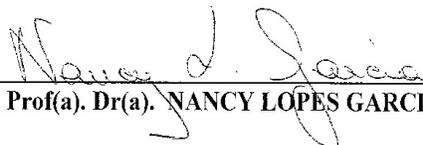
**Data de defesa:** 21-03-2013

**Programa de Pós-Graduação:** Estatística

---

**Dissertação de Mestrado defendida em 21 de março de 2013 e aprovada**

**Pela Banca Examinadora composta pelos Profs. Drs.**

  
Prof(a). Dr(a). NANCY LOPES GARCIA

  
Prof(a). Dr(a). MARIANA RODRIGUES MOTTA

  
Prof(a). Dr(a). HEDIBERT FREITAS LOPES



# Agradecimentos

Agradeço primeiramente à minha orientadora, professora Nancy Lopes Garcia. Foi um grande prazer trabalhar com uma profissional séria e objetiva que sempre me ensinou a não medir esforços para fazer o meu melhor, mas ao mesmo tempo com um coração de mãe me motivou e me apoiou nos momentos críticos. Agradeço a ela por ter me incentivado a descobrir o gosto pela pesquisa desde a iniciação científica, e por sempre acreditar na minha capacidade como pesquisadora.

Ao professor Ronaldo Dias pelo incentivo, pelas sugestões e críticas sempre construtivas, e pelas aulas inspiradoras. À professora Nancy Heckman, da UBC, com quem desenvolvi parte de meu trabalho no Canadá, por ter me proporcionado uma experiência inesquecível. Devo muito a ela pela paciência e disposição em me ensinar a fazer programas em R mais eficientes e me dar dicas de como ser clara e persuasiva na exposição dos resultados. Agradeço também à Camila Estevam, que durante a minha visita ao Canadá nunca hesitou em me ajudar e sempre me deu bons conselhos, se tornando uma grande amiga.

Aos professores Hedibert Freitas Lopes e Mariana Rodrigues Motta por aceitarem participar da banca examinadora e pelas valiosas sugestões.

Aos meus colegas de mestrado, pelos estudos em grupo e pelas experiências que compartilhamos nos últimos anos, sobretudo ao Alex Rodrigo dos Santos Sousa, por estar presente nas mais diversas situações, sempre me fazendo enxergar o lado bom das coisas. Aos amigos que fiz no Canadá, por me apoiarem e me acolherem.

Aos meus pais, Paulo e Ana, por estarem sempre ao meu lado nas dificuldades, por me apoiarem nos estudos e me incentivaram a fazer o mestrado, respeitando minhas decisões. Ao meu pai em particular por várias vezes me questionar sobre aspectos práticos do meu trabalho e pelas intermináveis revisões de texto durante o projeto.

Finalmente, agradeço a CAPES e a CBIE scholarship pelo suporte financeiro a este projeto.



# Resumo

Neste trabalho tratamos de um problema de dados funcionais agregados, isto é, combinações lineares de curvas aleatórias que não podem ser observadas individualmente. A análise de curvas de carga de transformadores elétricos em linha de distribuição é uma situação onde tais dados são encontrados. O conjunto de dados consiste de diversas curvas agregadas; assumimos que cada curva individual é a realização de um processo Gaussiano com curva média que é modelada através da expansão em bases B-splines. Além disso, assumimos que os coeficientes também são desconhecidos e temos somente uma estimativa de tais números. Nosso objetivo é utilizar os valores aproximados desses coeficientes, além das curvas agregadas, para estimar o valor real dos coeficientes, a curva típica de cada subpopulação e parâmetros de variância. Para este fim, diferentemente de trabalhos anteriores sobre o tema, propomos a utilização de um modelo de efeitos aleatórios.

**Palavras-chave:** B-splines, Curvas funcionais agregadas, Curva típica de demanda, Coeficientes aleatórios.



# Abstract

In this paper, we deal with a problem of aggregated functional data, i.e. linear combinations of random curves, which can not be seen individually. The analysis of load curves of electrical transformers in the distribution line is a situation where those data are found. The data set consists of several aggregated curves; we assume that each individual curve is the realization of a Gaussian process with the mean curve modeled through the expansion in B-splines basis. Furthermore, we assume that the coefficients are also unknown and we have only an estimate of such numbers. Our goal is to use the approximate coefficients and the aggregated curve data to estimate the true coefficients, the typical curve of each subpopulation and variance parameters. For this purpose, unlike previous works on the subject, we propose using a random effects model.

**Keywords:** B-splines, Aggregated functional curves, Typical load curve, Random coefficients.



# Sumário

Agradecimentos	vi
Resumo	ix
Abstract	xi
Lista de Figuras	xv
Lista de Tabelas	xvii
<b>1 Introdução</b>	<b>1</b>
<b>2 Modelo funcional para a demanda de energia elétrica</b>	<b>5</b>
2.1 Descrição do problema . . . . .	5
2.2 B-splines . . . . .	6
2.3 O modelo de demanda de energia com expansão em bases . . . . .	8
2.3.1 Modelo para o número reportado de consumidores em cada classe . . . . .	9
<b>3 Estimação de máxima verossimilhança</b>	<b>11</b>
3.1 A função de verossimilhança . . . . .	11
3.2 O processo de maximização . . . . .	12
<b>4 Aplicação em dados reais</b>	<b>21</b>
<b>5 Simulações</b>	<b>25</b>
5.1 Geração dos dados . . . . .	26
5.2 Análise dos dados . . . . .	29
5.3 Resultados . . . . .	29
5.3.1 Estimativas dos $\alpha_c$ 's . . . . .	29
5.3.2 Estimativas dos parâmetros de variância . . . . .	37

5.3.3 Estimativas dos $M_{ci}$ 's . . . . .	43
5.4 Conclusões . . . . .	47
<b>A Figuras e Tabelas</b>	<b>49</b>
<b>B Programa em R</b>	<b>69</b>

# Lista de Figuras

2.1	10 bases B-splines de ordem 3 em $[0,1]$ . . . . .	8
4.1	Dados do primeiro transformador com $R_{11} = 34$ e $R_{21} = 7$ . Cada curva representa a demanda total de energia para um dia da semana. . . . .	21
4.2	Tipologias estimadas dos consumidores residenciais e comerciais. A figura superior mostra as duas estimativas. A figura inferior mostra a estimativa da tipologia dos consumidores residenciais à esquerda e dos consumidores comerciais à direita. . . . .	24
5.1	Exemplo das curvas individuais simuladas comparadas com as tipologias $\alpha_1$ (curva mais escura no gráfico da esquerda) e $\alpha_2$ (curva mais escura no gráfico da direita) para o Caso 1, com 45 consumidores da classe 1 e 30 da classe 2. . . . .	28
5.2	Exemplo das curvas individuais simuladas comparadas com as tipologias $\alpha_1$ (curva mais escura no gráfico da esquerda) e $\alpha_2$ (curva mais escura no gráfico da direita) para o Caso 4, com 69 consumidores da classe 1 e 6 da classe 2. . . . .	28
5.3	Mínimo e máximo (curvas tracejadas), mediana e quartis (curvas pontilhadas) das 200 estimativas da demanda de energia para o Caso 1, para as classe $c = 1$ (coluna da esquerda) e $c = 2$ (coluna da direita) com 5 transformadores, sem replicação (linha superior) e com replicação (linha inferior). As curvas sólidas representam os valores verdadeiros. . . . .	31
5.4	Mínimo e máximo (curvas tracejadas), mediana e quartis (curvas pontilhadas) das 200 estimativas da demanda de energia para o Caso 2, para as classe $c = 1$ (coluna da esquerda) e $c = 2$ (coluna da direita) com 5 transformadores, sem replicação (linha superior) e com replicação (linha inferior). As curvas sólidas representam os valores verdadeiros. . . . .	32

5.5	Mínimo e máximo (curvas tracejadas), mediana e quartis (curvas pontilhadas) das 200 estimativas da demanda de energia para o Caso 3, para as classe $c = 1$ (coluna da esquerda) e $c = 2$ (coluna da direita) com 5 transformadores, sem replicação (linha superior) e com replicação (linha inferior). As curvas sólidas representam os valores verdadeiros. . . . .	33
5.6	Mínimo e máximo (curvas tracejadas), mediana e quartis (curvas pontilhadas) das 200 estimativas da demanda de energia para o Caso 4, para as classe $c = 1$ (coluna da esquerda) e $c = 2$ (coluna da direita) com 5 transformadores, sem replicação (linha superior) e com replicação (linha inferior). As curvas sólidas representam os valores verdadeiros. . . . .	34
5.7	Mínimo e máximo (curvas tracejadas), mediana e quartis (curvas pontilhadas) das 200 estimativas da demanda de energia para o Caso 4, para as classe $c = 1$ (coluna da esquerda) e $c = 2$ (coluna da direita) com 50 transformadores, sem replicação (linha superior) e com replicação (linha inferior). As curvas sólidas representam os valores verdadeiros. . . . .	35
5.8	Mínimo e máximo (curvas tracejadas), mediana e quartis (curvas pontilhadas) das 200 estimativas da soma ponderada das tipologias no primeiro transformador, com 66 consumidores da classe 1 e 9 consumidores da classe 2, para o Caso 4 sem replicação, com 5 transformadores. A curva sólida representa os valores verdadeiros. . . . .	36
5.9	Densidade das 200 estimativas de $\sigma^2$ para o Caso 4 sem replicação, com 5 transformadores. A linha vertical representa o valor verdadeiro. . . . .	38
5.10	Densidade das $N$ estimativas maiores do que zero de $\sigma_{\gamma,1}^2$ (coluna da esquerda) e $\sigma_{\gamma,2}^2$ (coluna da direita) para o Caso 1 com replicação, com 5 transformadores (linha superior) e 50 transformadores (linha inferior). A linha vertical representa o valor verdadeiro. . . . .	39
5.11	Densidade das $N$ estimativas maiores do que zero de $\sigma_{\gamma,1}^2$ (coluna da esquerda) e $\sigma_{\gamma,2}^2$ (coluna da direita) para o Caso 2 com replicação, com 5 transformadores (linha superior) e 50 transformadores (linha inferior). A linha vertical representa o valor verdadeiro. . . . .	40
5.12	Densidade das $N$ estimativas maiores do que zero de $\sigma_{\gamma,1}^2$ (coluna da esquerda) e $\sigma_{\gamma,2}^2$ (coluna da direita) para o Caso 3 com replicação, com 5 transformadores (linha superior) e 50 transformadores (linha inferior). A linha vertical representa o valor verdadeiro. . . . .	41

5.13 Densidade das  $N$  estimativas maiores do que zero de  $\sigma_{\gamma,1}^2$  (coluna da esquerda) e  $\sigma_{\gamma,2}^2$  (coluna da direita) para o Caso 4 com replicação, com 5 transformadores (linha superior) e 50 transformadores (linha inferior). A linha vertical representa o valor verdadeiro. . . . . 42

A.1 Densidade das 200 estimativas de  $\sigma^2$  para os Casos 1 e 2 (linha superior) e para o Caso 3 (linha inferior), sem replicação e com 5 transformadores. A linha vertical é o valor verdadeiro. . . . . 49

A.2 Densidade das 200 estimativas de  $\sigma^2$  para os Casos 1 e 2 (linha superior) e para os Casos 3 e 4 (linha inferior), com replicação e com 5 transformadores. A linha vertical é o valor verdadeiro. . . . . 50

A.3 Densidade das 200 estimativas de  $\sigma^2$  para os Casos 1 e 2 (linha superior) e para os Casos 3 e 4 (linha inferior), sem replicação e com 50 transformadores. A linha vertical é o valor verdadeiro. . . . . 51

A.4 Densidade das 200 estimativas de  $\sigma^2$  para os Casos 1 e 2 (linha superior) e para os Casos 3 e 4 (linha inferior), com replicação e com 50 transformadores. A linha vertical é o valor verdadeiro. . . . . 52

A.5 Mínimo e máximo (curvas tracejadas), mediana e quartis (curvas pontilhadas) das 200 estimativas da demanda de energia para o Caso 1, para as classe  $c = 1$  (coluna da esquerda) e  $c = 2$  (coluna da direita) com 50 transformadores, sem replicação (linha superior) e com replicação (linha inferior). As curvas sólidas representam os valores verdadeiros. . . . . 53

A.6 Mínimo e máximo (curvas tracejadas), mediana e quartis (curvas pontilhadas) das 200 estimativas da demanda de energia para o Caso 2, para as classe  $c = 1$  (coluna da esquerda) e  $c = 2$  (coluna da direita) com 50 transformadores, sem replicação (linha superior) e com replicação (linha inferior). As curvas sólidas representam os valores verdadeiros. . . . . 54

A.7 Mínimo e máximo (curvas tracejadas), mediana e quartis (curvas pontilhadas) das 200 estimativas da demanda de energia para o Caso 3, para as classe  $c = 1$  (coluna da esquerda) e  $c = 2$  (coluna da direita) com 50 transformadores, sem replicação (linha superior) e com replicação (linha inferior). As curvas sólidas representam os valores verdadeiros. . . . . 55

A.8 Densidade das  $N$  estimativas maiores do que zero de  $\sigma_{\gamma,1}^2$  (coluna da esquerda) e  $\sigma_{\gamma,2}^2$  (coluna da direita) para o Caso 1 sem replicação, com 5 transformadores (linha superior) e 50 transformadores (linha inferior). A linha vertical é o valor verdadeiro. . . . . 56

- A.9 Densidade das  $N$  estimativas maiores do que zero de  $\sigma_{\gamma,1}^2$  (coluna da esquerda) e  $\sigma_{\gamma,2}^2$  (coluna da direita) para o Caso 2 sem replicação, com 5 transformadores (linha superior) e 50 transformadores (linha inferior). A linha vertical é o valor verdadeiro. . . . . 57
- A.10 A figura superior mostra a densidade das  $N$  estimativas maiores do que zero de  $\sigma_{\gamma,2}^2$  para o Caso 3 sem replicação, com 5 transformadores. A figura inferior mostra a densidade das  $N$  estimativas maiores do que zero e de  $\sigma_{\gamma,1}^2$  (figura da esquerda) e  $\sigma_{\gamma,2}^2$  (figura da direita) para o Caso 3 sem replicação, com 50 transformadores. A linha vertical representa o valor verdadeiro. Note que 199 das 200 estimativas de  $\sigma_{\gamma,1}^2$  para o Caso 3 sem replicação, com 5 transformadores foram iguais a zero. . . . . 58
- A.11 A figura superior mostra a densidade das  $N$  estimativas maiores do que zero de  $\sigma_{\gamma,2}^2$  para o Caso 4 sem replicação, com 5 transformadores. A figura inferior mostra a densidade das  $N$  estimativas maiores do que zero e de  $\sigma_{\gamma,1}^2$  (figura da esquerda) e  $\sigma_{\gamma,2}^2$  (figura da direita) para o Caso 4 sem replicação, com 50 transformadores. A linha vertical representa o valor verdadeiro. Note que 191 das 200 estimativas de  $\sigma_{\gamma,1}^2$  para o Caso 4 sem replicação, com 5 transformadores foram iguais a zero. . . . . 59

# Lista de Tabelas

2.1	Número de consumidores em diferentes categorias: $x_{cj}$ o número de consumidores da classe $c$ que reportam ser da classe $j$ . . . . .	10
3.1	Número aproximado de consumidores nas duas classes resultante das multinomiais geradas na $b$ -ésima iteração. . . . .	18
3.2	Estimativas de $H(m_1, 75 - m_1)$ quando $R_1 = 32$ e $R_2 = 43$ . . . . .	19
4.1	Resumo das estimativas . . . . .	24
5.1	O número verdadeiro de consumidores da classe 1 e o número reportado gerado aleatoriamente para cinco transformadores. . . . .	29
5.2	Número estimado de consumidores na classe $c = 1$ (residencial) nos transformadores $i = 1, \dots, 5$ para os casos onde os $M$ 's são balanceados. Na primeira coluna da tabela estão as simulações feitas com $I = 5$ transformadores e na segunda coluna com $I = 50$ transformadores. O número verdadeiro de consumidores aparece em negrito e o número reportado aparece entre colchetes. Note-se que, no transformador 1 o número reportado de consumidores é igual ao verdadeiro e nos transformadores 4 e 5 para o caso das simulações realizadas com $I = 50$ , nenhuma estimativa foi igual ao número reportado. . . . .	45
5.3	Número estimado de consumidores na classe $c = 1$ (residencial) nos transformadores $i = 1, \dots, 5$ para os casos onde os $M$ 's são desbalanceados. Na primeira coluna da tabela estão as simulações feitas com $I = 5$ transformadores e na segunda coluna com $I = 50$ transformadores. O número verdadeiro de consumidores aparece em negrito e o número reportado aparece entre colchetes. . . . .	46

A.1	Número estimado de consumidores na classe $c = 1$ (residencial) nos transformadores $i = 6, \dots, 15$ para dados gerados com $I = 50$ nos casos onde os $M$ 's são balanceados. O número verdadeiro de consumidores aparece em negrito e o número reportado está entre colchetes. . . . .	60
A.2	Número estimado de consumidores na classe $c = 1$ (residencial) nos transformadores $i = 16, \dots, 27$ para dados gerados com $I = 50$ nos casos onde os $M$ 's são balanceados. O número verdadeiro de consumidores aparece em negrito e o número reportado está entre colchetes. . . . .	61
A.3	Número estimado de consumidores na classe $c = 1$ (residencial) nos transformadores $i = 28, \dots, 39$ para dados gerados com $I = 50$ nos casos onde os $M$ 's são balanceados. O número verdadeiro de consumidores aparece em negrito e o número reportado está entre colchetes. . . . .	62
A.4	Número estimado de consumidores na classe $c = 1$ (residencial) nos transformadores $i = 40, \dots, 50$ para dados gerados com $I = 50$ nos casos onde os $M$ 's são balanceados. O número verdadeiro de consumidores aparece em negrito e o número reportado está entre colchetes. . . . .	63
A.5	Número estimado de consumidores na classe $c = 1$ (residencial) nos transformadores $i = 6, \dots, 15$ para dados gerados com $I = 50$ nos casos onde os $M$ 's são desbalanceados. O número verdadeiro de consumidores aparece em negrito e o número reportado está entre colchetes. . . . .	64
A.6	Número estimado de consumidores na classe $c = 1$ (residencial) nos transformadores $i = 16, \dots, 27$ para dados gerados com $I = 50$ nos casos onde os $M$ 's são desbalanceados. O número verdadeiro de consumidores aparece em negrito e o número reportado está entre colchetes. . . . .	65
A.7	Número estimado de consumidores na classe $c = 1$ (residencial) nos transformadores $i = 28, \dots, 39$ para dados gerados com $I = 50$ nos casos onde os $M$ 's são desbalanceados. O número verdadeiro de consumidores aparece em negrito e o número reportado está entre colchetes. . . . .	66
A.8	Número estimado de consumidores na classe $c = 1$ (residencial) nos transformadores $i = 40, \dots, 50$ para dados gerados com $I = 50$ nos casos onde os $M$ 's são desbalanceados. O número verdadeiro de consumidores aparece em negrito e o número reportado está entre colchetes. . . . .	67

# Capítulo 1

## Introdução

A abordagem que trata o dado individual como uma função e não como um único valor em um ponto particular é chamada de análise de dados funcionais, e é bastante difundida atualmente. Uma boa introdução ao tema pode ser encontrada em [Ramsay and Silverman \(2005\)](#) e [Ferraty and Vieu \(2006\)](#). Neste sentido, dados funcionais agregados são combinações lineares de curvas aleatórias que não podem ser observadas individualmente. No presente trabalho desenvolvemos e estudamos uma metodologia para tratar o problema de estimação de dados funcionais quando não se tem disponíveis as curvas individuais que os representam, e dispõe-se apenas de dados que representam a soma dessas curvas.

Dentre as aplicações de tal metodologia podemos citar o problema do dimensionamento de transformadores para a distribuição de energia elétrica. A estratégia mais comum utilizada atualmente é a de definir a capacidade dos transformadores pelo pico de demanda total dos usuários por ele abastecidos, o que evita a ocorrência de sobrecarga. Uma alternativa mais eficiente para lidar com essa questão é combinar consumidores de diversos tipos (residenciais, comerciais, industriais, etc) para serem abastecidos por cada transformador, de tal forma a obter, em cada equipamento, uma curva de demanda agregada aproximadamente homogênea ao longo do dia.

Para que isso possa ser feito, é necessário inicialmente identificar as curvas típicas de demanda (tipologia) de cada classe de consumidores ao longo do dia, em especial, determinando os horários típicos de ocorrência de picos para cada classe. A partir do conhecimento dessas curvas, pode-se combinar quantidades apropriadas de consumidores de cada tipo, que exibem picos de demanda em horários distintos entre si, de modo a homogeneizar a curva de demanda agregada no transformador. Por exemplo, é razoável pensar que no Brasil, consumidores residenciais têm seu pico de demanda de energia entre 19:00 e 21:00 (devido ao uso de chuveiro), enquanto consumidores comerciais e industriais apresentam maior demanda entre 8:00 e 18:00

(horário comum de expediente para empresas). A estimação da curva típica de demanda de energia utilizando dados funcionais agregados foi feita pela primeira vez por [Dias, Garcia, and Martarelli \(2009\)](#) e posteriormente com uma abordagem Bayesiana por [Dias, Garcia, and Schmidt \(2011\)](#). Uma forma de determinar a tipologia seria tomar grandes amostras de demanda ao longo do dia para obter estimativas das médias e variâncias de cada uma das classes. No entanto, do ponto de vista prático, é bastante difícil e dispendioso obter amostras de consumidores individuais com o detalhamento necessário.

Em geral, os dados sobre demanda que as companhias de distribuição de eletricidade têm disponíveis restringem-se a medidas instantâneas da demanda agregada, obtidas em cada transformador ao longo do dia; ou seja, não se tem a curva de demanda individual, por consumidor, através da informação disponível nem tampouco a curva característica de cada classe de consumo (residencial, comercial, industrial, etc.) em cada transformador. Além disso, as companhias de distribuição de energia possuem informações referentes aos mercados (número de consumidores em cada classe) mas podem haver distorções no número reportado de cada tipo de usuário devido a diversos fatores, como fraudes, mudanças de categorias não informadas, problemas nos medidores, etc. Por essas razões, há grande dificuldade em calcular a demanda prevista para novas linhas com base nas informações disponíveis acerca de linhas de distribuição existentes. De modo a demonstrar sua aplicação prática, utilizamos a metodologia aqui desenvolvida para, a partir de um conjunto de dados de demanda agregada fornecidos por uma concessionária de energia elétrica, estimar o número real de consumidores em cada classe abastecidos por cada transformador da linha, bem como obter a curva de demanda esperada das classes de consumo. Com base nesses resultados, o fornecedor poderá dimensionar transformadores para novas linhas, ou redimensioná-los para linhas existentes, de modo a minimizar custos e reduzir riscos de subdimensionamento.

Em nosso trabalho assumimos que a curva esperada de demanda típica de cada classe de consumo é uma curva suave, a qual pode ser aproximada por uma função que é uma combinação linear de bases B-splines. A perturbação individual do consumidor também é modelada como uma combinação linear de bases com coeficientes aleatórios, e o número reportado de consumidores em cada categoria é considerado uma variável aleatória.

Detalhes do nosso modelo, bem como uma breve introdução à teoria de B-splines podem ser encontrados no [Capítulo 2](#). O [Capítulo 3](#) é dedicado aos métodos de estimação. No [Capítulo 4](#) aplicamos a metodologia em dados reais de demanda de energia para três transformadores. O [Capítulo 5](#) contém detalhes e resultados de um estudo de simulação. Material suplementar com gráficos e tabelas, bem como o código do programa desenvolvido em *R* utilizado nas simulações, podem ser encontrados no Apêndice.





# Capítulo 2

## Modelo funcional para a demanda de energia elétrica

### 2.1 Descrição do problema

Estamos interessados em uma metodologia que, dados apenas a curva de demanda agregada e o número de consumidores reportados por classe de consumo, permita estimar características da tipologia destes consumidores e também o número real deles em cada uma das classes.

Para isso, vamos indexar o  $i$ -ésimo transformador por  $i = 1, \dots, I$ , a classe por  $c = 1, \dots, C$  e o  $l$ -ésimo consumidor servido pelo transformador  $i$  por  $l = 1, \dots, N_i$ . Adicionalmente, a classe verdadeira do  $l$ -ésimo consumidor no transformador  $i$  será denotada por  $c_{li}$  e a classe reportada por  $r_{li}$ .

Assumimos aqui que para um indivíduo da classe  $c$ , a demanda de energia é um processo estocástico com distribuição dependente de  $c$  mas não do transformador. Supomos que  $W_{li}(t)$ , a demanda de energia no tempo  $t$  do  $l$ -ésimo consumidor no transformador  $i$  da classe  $c_{li} = c$ , é dada por

$$W_{li}(t) \Big|_{c_{li}=c} = \alpha_c(t) + \alpha_{cli}^*(t) \quad (2.1)$$

onde  $\alpha_c$  é a curva de demanda típica não aleatória na classe  $c$ ,  $\alpha_{cli}^*$  é a perturbação aleatória referente ao consumidor  $l$  específico, e consideramos que as componentes aleatórias são processos independentes.

No entanto, no  $i$ -ésimo transformador não observamos  $W_{li}$ , e sim a soma da demanda de todos os usuários abastecidos por aquele transformador, mais um erro de medição em  $n$  medidas efetuadas nos tempos  $t_1 < t_2 \dots < t_n$ . Isto é, temos o vetor de dados:  $\mathbf{Y}_i \equiv (Y_i(t_1), \dots, Y_i(t_n))'$

onde, para  $\epsilon_i(\cdot)$  e  $W_{li}(\cdot)$  independentes, com  $\epsilon_i(t) \sim N(0, \sigma^2)$ ,

$$\begin{aligned} Y_i(t) &= \sum_{l=1}^{N_i} \sum_{c=1}^C W_{li}(t) \mathbf{I}\{c_{li} = c\} + \epsilon_i(t) \\ &= \sum_{c=1}^C M_{ci} \alpha_c(t) + \sum_{l=1}^{N_i} \sum_{c=1}^C \alpha_{cli}^*(t) \mathbf{I}\{c_{li} = c\} + \epsilon_i(t), \end{aligned} \quad (2.2)$$

onde  $M_{ci}$  é o número real de consumidores da classe  $c$  no  $i$ -ésimo transformador.

Note que nosso método pode ser facilmente estendido para o caso com diferentes medidas no tempo em cada transformador, bem como diferentes variâncias para o erro. No entanto, com o objetivo de simplificar a notação, nosso desenvolvimento é feito considerando tais medidas iguais em todos os transformadores.

É importante perceber que não observamos o número real de consumidores no  $i$ -ésimo transformador, pertencentes a classe  $c$ , ou seja,  $M_{ci}$ . Na verdade, observamos apenas  $R_{ci}$ , o número reportado de consumidores no  $i$ -ésimo transformador pertencentes à classe  $c$ .

No presente trabalho, descreveremos um método não paramétrico que utiliza a representação das curvas típicas de demanda como uma combinação linear de bases B-splines. As perturbações individuais, no método aqui proposto, são modeladas como B-splines com coeficientes aleatórios. Em dados funcionais, uma abordagem que utiliza modelos com efeitos aleatórios para a representação de curvas suavizadas por splines pode ser encontrada em [Ruppert, Wand, and Carroll \(2003\)](#), [Verbyla, Cullis, Kenward, and Welham \(2002\)](#) e [Heckman, Lockhart, and Nielsen \(2009\)](#). O número reportado de consumidores em cada categoria é tratado como uma variável aleatória que depende apenas do número verdadeiro de consumidores em cada classe e das probabilidades de que consumidores de uma determinada classe reportem ser de outra.

A seguir, daremos uma introdução à técnica de expansão em bases B-splines, para mais informações sobre tal técnica ver [de Souza and Dias \(2008\)](#).

## 2.2 B-splines

Spline é um polinômio por partes cujos diferentes segmentos são conectados suavemente em certos valores chamados nós. Normalmente funções splines dão um bom ajuste aos dados, sem fortes suposições paramétricas na função de regressão.

Mais precisamente,  $\phi$  é um spline de grau  $D$  em  $[a, b]$  com nós interiores  $\tau_1, \dots, \tau_L$ , tal que  $a < \tau_1 < \dots < \tau_L < b$ , se  $\phi$  é um polinômio de grau  $D$  em cada um dos subintervalos  $[a, \tau_1], [\tau_1, \tau_2], \dots, [\tau_L, b]$  e  $\phi$  tem  $D - 1$  derivadas contínuas. Dados  $D$  e os nós, o conjunto de

todas as funções splines é um espaço de funções de dimensão finita. Em geral, a dimensão do espaço de funções, que é igual ao número de parâmetros, é dada por  $L + 1 + D$ .

Precisamos agora escolher funções bases para todas as funções splines em  $[a, b]$  de grau  $D$  com nós interiores  $\tau_1, \dots, \tau_L$ . Uma escolha comum são as bases B-splines, por serem computacionalmente rápidas e estáveis. A Figura 2.1 mostra 10 bases B-splines de ordem 3 com nós equidistantes no intervalo  $[0, 1]$ . As linhas verticais pontilhadas representam a localização dos nós internos.

A construção dessas bases consiste em encontrar  $\phi_1, \dots, \phi_{L+1+D}$  que sejam splines linearmente independentes. Abaixo, descrevemos o algoritmo de De Boor (1978), desenvolvido para calcular B-splines de forma recursiva.

**Algoritmo de De Boor (1978)** *O  $i$ -ésimo B-spline de grau  $D$  para uma sequência de nós  $\tau = \{\tau_i\}$  pode ser calculado como:*

$$B_{i,D}(t) = \frac{t - \tau_i}{\tau_{i+D} - \tau_i} B_{i,D-1}(t) + \frac{\tau_{i+D+1} - t}{\tau_{i+D+1} - \tau_{i+1}} B_{i+1,D-1}(t)$$

onde

$$B_{i,1}(t) = \begin{cases} 1, & \text{se } \tau_i < t < \tau_{i+1} \\ 0, & \text{c.c.} \end{cases}$$

Assim, seja  $K = D + L + 1$ , é possível expandir uma função  $x$  em bases B-splines de grau  $D$  com  $L$  nós interiores da seguinte forma

$$x(t) = \sum_{i=1}^K c_i B_{i,D}(t).$$

Como a ordem do B-spline em geral é clara pelo contexto, podemos no lugar de  $B_{i,D}(t)$  escrever apenas  $B_i(t)$ . Então, a estimativa suave de  $x$  calculada a partir dos dados, é dada por

$$\hat{x}(t) = \sum_{i=1}^K \hat{c}_i B_i(t).$$

Desta forma, o problema de estimação se reduz a estimar os parâmetros  $c_1, \dots, c_K$ .

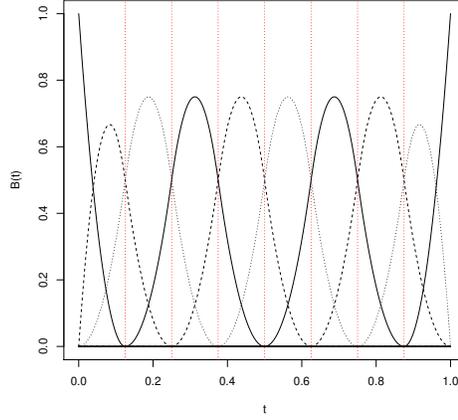


Figura 2.1: 10 bases B-splines de ordem 3 em [0,1]

## 2.3 O modelo de demanda de energia com expansão em bases

Nesta seção vamos propor uma abordagem alternativa através da representação das tipologias e da perturbação aleatória como uma expansão em bases B-splines. Tal abordagem possui boas propriedades de aproximação e diminui consideravelmente o número de parâmetros a serem estimados comparativamente à estimação de  $\alpha_c(t)$  e  $\alpha_{cli}^*(t)$  através da Expressão (2.2). Utilizando a notação de que a  $k$ -ésima componente de um vetor  $\mathbf{v}$  é  $\mathbf{v}[k]$ , modelamos  $\alpha_c(t)$  e  $\alpha_{cli}^*(t)$  com bases B-splines  $\phi_1, \dots, \phi_K$  e  $\psi_1, \dots, \psi_{K^*}$  da seguinte forma

$$\alpha_c(t) = \sum_{k=1}^K \gamma^c[k] \phi_k(t) \equiv \phi(t)' \gamma^c \quad (2.3)$$

e

$$\alpha_{cli}^*(t) = \sum_{k=1}^{K^*} \gamma^{cli}[k] \psi_k(t) \equiv \psi(t)' \gamma^{cli}. \quad (2.4)$$

Supomos que  $\gamma^c$  é um vetor desconhecido de parâmetros e os vetores  $\gamma^{cli}$  são efeitos aleatórios, normalmente distribuídos, independentes, com média zero e matriz de covariância  $\Sigma_c$ . Desta forma, os parâmetros desconhecidos para a distribuição da demanda de energia de um indivíduo da classe  $c$  são  $\gamma^c$  e  $\Sigma_c$ .

É importante ressaltar que as matrizes de bases  $\phi(t)$  e  $\psi(t)$  em (2.3) e (2.4) não dependem da classe  $c$  à qual o consumidor pertence, já que estamos utilizando o mesmo número de bases e a mesma posição dos nós para todos os tipos de consumidores. É possível alocar o número e

a posição dos nós de acordo com as classes; por exemplo, podemos colocar mais nós entre as 18 e 21 horas para a classe de consumidores residenciais por causa dos picos de demanda desses horários. No entanto, é muito importante manter as mesmas bases  $\phi(t)$  e  $\psi(t)$  em todos os transformadores já que, caso contrário, teremos apenas uma amostra para estimar diferentes coeficientes.

Note que, dado  $c_{li} = c$ , a demanda individual de energia  $W_{li}$ , dado pela Expressão (2.1), é um processo Gaussiano, ou seja, qualquer distribuição finito-dimensional  $W_{li}(t_1, \dots, t_k)$ , com  $t_1, \dots, t_k \in T$  é uma normal multivariada com média  $\alpha_c(t_1, \dots, t_k)$  e matriz de covariância dada por

$$\sigma_c(t_s, t_j) \equiv \text{cov}(W_{li}(t_s), W_{li}(t_j)) = \text{cov}(\alpha_{cli}^*(t_s), \alpha_{cli}^*(t_j)) = \psi(t_s)' \Sigma_c \psi(t_j), \text{ com } s, j = 1, \dots, k.$$

Da mesma forma, podemos ver facilmente que o vetor que representa a demanda de energia agregada observado em cada um dos transformadores tem distribuição normal multivariada com valor esperado e matriz de covariância dados por

$$\begin{aligned} E(Y_i(t)) &= \sum_{c=1}^C M_{ci} \alpha_c(t) \text{ e} \\ \text{Cov}(Y_i(s), Y_i(t)) &= \sum_{l=1}^{N_i} \sum_{c=1}^C \mathbb{I}\{c_{li} = c\} \text{cov}(\alpha_{cli}^*(s), \alpha_{cli}^*(t)) + \text{cov}(\epsilon_i(s), \epsilon_i(t)). \end{aligned}$$

Observe que a covariância calculada acima resulta da independência dos  $\alpha_{cli}^*$ 's.

Agora, definindo as matrizes de bases  $\Phi$  e  $\Psi$ , segue que

$$\mathbf{Y}_i \sim N \left( \Phi \sum_{c=1}^C M_{ci} \gamma^c, \Psi \sum_{c=1}^C M_{ci} \Sigma_c \Psi' + \sigma^2 \mathbf{I} \right). \quad (2.5)$$

### 2.3.1 Modelo para o número reportado de consumidores em cada classe

Em nosso modelo, o número reportado de consumidores em cada uma das classes é uma variável aleatória, enquanto o número verdadeiro é considerado parâmetro fixo. Desta forma, seria interessante encontrar a distribuição do número reportado de consumidores em cada classe quando o número verdadeiro é fixo.

Para isso, devemos lembrar que  $r_{li}$  é a classe reportada pelo consumidor  $l$  no transformador  $i$  e  $c_{li}$  é a classe verdadeira. O número real de consumidores na classe  $c$  no transformador  $i$  é denotado por  $M_{ci}$  e o reportado por  $R_{ci}$ . Supomos que todos os consumidores reportam ser de alguma classe, isto é, o número total de consumidores no transformador  $i$ , denotado por  $N_i$ , é

tal que:  $N_i = \sum_c R_{ci} = \sum_c M_{ci}$ .

Vamos supor também que exista uma *matriz fraude*  $\mathcal{F}$ , conhecida e que não depende do transformador, com:

$$\mathcal{F}(c, r) = P\{r_{li} = r | c_{li} = c\}, \quad r, c = 1, \dots, C,$$

igual à probabilidade de que um consumidor da classe  $c$  reporte pertencer à classe  $r$ . Vamos assumir que consumidores reportam-se independentemente de outros consumidores, o que é bastante razoável.

Para o melhor entendimento da distribuição do número de reportados, construímos a Tabela 2.1. Seja  $x_{cj}$  o número de consumidores que são da classe  $c$  mas que reportam ser da classe  $j$ , por conveniência vamos omitir o subíndice  $i$  da notação. Então, o número reportado de consumidores da classe  $j$  é dado pela soma dos números da  $j$ -ésima coluna:  $R_j = \sum_c x_{cj}$ , enquanto que a quantidade real de consumidores da classe  $c$  é a soma dos números que estão na  $c$ -ésima linha:  $M_c = \sum_j x_{cj}$ .

Para  $c = 1, \dots, C$ , seja  $\mathbf{X}_c \equiv (x_{c1}, \dots, x_{cC})'$ , o vetor com o número de consumidores que são da classe  $c$  mas que reportam ser das classes  $j = 1, \dots, C$ , encontrado na linha  $c$  da Tabela 2.1. Então, dado o valor de  $M_c$ ,  $\mathbf{X}_c$  tem distribuição multinomial( $M_c, \mathcal{F}(c, 1), \dots, \mathcal{F}(c, C)$ ). Como os consumidores reportam-se de forma independente, segue que  $\mathbf{X}_1, \dots, \mathbf{X}_C$  são independentes.

Desta forma, definimos a distribuição conjunta de  $R_1, \dots, R_C$  quando o número verdadeiro de consumidores em cada classe é dado por  $M_1, \dots, M_C$ .

Classe verdadeira \ Classe reportada	Classe reportada					Total da linha
	1	...	$j$	...	$C$	
1	$x_{11}$	...	$x_{1j}$	...	$x_{1C}$	$M_1$
⋮	⋮		⋮		⋮	⋮
$c$	$x_{c1}$	...	$x_{cj}$	...	$x_{cC}$	$M_c$
⋮	⋮		⋮		⋮	⋮
$C$	$x_{C1}$	...	$x_{Cj}$	...	$x_{CC}$	$M_C$
Total da coluna	$R_1$	...	$R_j$	...	$R_C$	$N$

Tabela 2.1: Número de consumidores em diferentes categorias:  $x_{cj}$  o número de consumidores da classe  $c$  que reportam ser da classe  $j$

# Capítulo 3

## Estimação de máxima verossimilhança

### 3.1 A função de verossimilhança

Na Seção 2.3 encontramos a distribuição do vetor que representa a demanda de energia agregada e na Seção 2.3.1 encontramos a distribuição do número reportado de consumidores em cada classe. Nesse capítulo, vamos definir a função de verossimilhança, a fim de encontrar estimativas para os parâmetros de interesse.

Os vetores de parâmetros a serem estimados são:  $\mathcal{G} = (\gamma^1, \dots, \gamma^C)'$ ,  $\mathcal{S} = \{\Sigma_1, \dots, \Sigma_C\}$ ,  $\sigma^2$  e  $\mathbf{M}_i = (M_{1i}, \dots, M_{Ci})'$ . Lembremos aqui que  $\mathcal{G}$  é o vetor de coeficientes da expansão dos efeitos fixos  $\alpha_c$ 's em bases,  $\mathcal{S}$  é um vetor com as matrizes de covariância dos coeficientes aleatórios da expansão dos efeitos aleatórios  $\alpha_{cli}^*$ 's em bases,  $\sigma^2$  é a variância do erro em cada transformador e  $\mathbf{M}_i$  é o vetor do número real de consumidores em cada classe em um determinado transformador. Além disso,  $\mathbf{M} = (\mathbf{M}_1, \dots, \mathbf{M}_I)'$  não é observado.

Para construir a verossimilhança, precisaremos dos vetores de dados  $\mathbf{Y}_i$ , dado em (2.5), e  $\mathbf{R}_i = (R_{1i}, \dots, R_{Ci})'$ , com o número reportado de consumidores, para  $i = 1, \dots, I$ .

Pela independência dos transformadores, temos que a verossimilhança é dada por

$$\mathcal{L}(\mathcal{G}, \mathcal{S}, \sigma^2, \mathbf{M}) = \prod_{i=1}^I \mathcal{L}_i(\mathcal{G}, \mathcal{S}, \sigma^2, \mathbf{M}_i)$$

onde

$$\begin{aligned} \mathcal{L}_i(\mathcal{G}, \mathcal{S}, \sigma^2, \mathbf{M}_i) &\equiv \mathcal{L}_i(\mathcal{G}, \mathcal{S}, \sigma^2, \mathbf{M}_i | \mathbf{Y}_i, \mathbf{R}_i) \\ &= f(\mathbf{Y}_i | \mathcal{G}, \mathcal{S}, \sigma^2, \mathbf{M}_i) \times P\{\mathbf{R}_i | \mathbf{M}_i\} \end{aligned} \quad (3.1)$$

em que  $P\{\mathbf{R}_i | \mathbf{M}_i\}$  é a função de probabilidade de  $\mathbf{R}_i$ , assumindo que o número verdadeiro de

consumidores em cada classe para o  $i$ -ésimo transformador são as componentes do vetor  $\mathbf{M}_i$ .

Então, queremos encontrar o conjunto de parâmetros,  $\mathcal{G}, \mathcal{S}, \sigma^2, \mathbf{M}_i$ , que maximize o logaritmo da verossimilhança, dado por

$$l(\mathcal{G}, \mathcal{S}, \sigma^2, \mathbf{M}) = \sum_{i=1}^I \log f(\mathbf{Y}_i | \mathcal{G}, \mathcal{S}, \sigma^2, \mathbf{M}_i) + \sum_{i=1}^I \log P\{\mathbf{R}_i | \mathbf{M}_i\}. \quad (3.2)$$

Segue de nosso modelo, dado em (2.5), que a primeira componente da Expressão (3.2) é simplesmente

$$\begin{aligned} \sum_{i=1}^I \log f(\mathbf{Y}_i | \mathcal{G}, \mathcal{S}, \sigma^2, \mathbf{M}_i) &\equiv -I \frac{1}{2} (n+1) \log(2\pi) - \frac{1}{2} \sum_{i=1}^I \log \left| \Psi \sum_{c=1}^C M_{ci} \Sigma_c \Psi' + \sigma^2 \mathbf{I} \right| \\ &\quad - \frac{1}{2} \sum_{i=1}^I \left( \mathbf{Y}_i - \sum_{c=1}^C M_{ci} \Phi \gamma^c \right)' \left( \Psi \sum_{c=1}^C M_{ci} \Sigma_c \Psi' + \sigma^2 \mathbf{I} \right)^{-1} \left( \mathbf{Y}_i - \sum_{c=1}^C M_{ci} \Phi \gamma^c \right). \end{aligned}$$

O cálculo da segunda parte de  $l(\mathcal{G}, \mathcal{S}, \sigma^2, \mathbf{M}_i)$  não é trivial pois a função  $P\{\mathbf{R}_i | \mathbf{M}_i\}$  não possui uma forma fechada, nem tampouco a sua derivada. Assim, essa parte da verossimilhança será discutida com mais detalhes na Seção 3.2 a seguir.

É importante lembrar que muitas vezes o conjunto de dados disponível contém medidas de demanda de energia de vários dias da semana para cada transformador. Testes feitos em nossas simulações mostram que a melhor maneira de tratar esse tipo de dado é usar os dias separadamente, como se fossem replicações.

Se considerarmos  $D$  dias em um transformador  $i$  como medidas independentes e identicamente distribuídas, teremos  $\mathbf{Y}_{i1}, \dots, \mathbf{Y}_{iD}$ , independentes e todos com a distribuição dada em (2.5). Consequentemente, a verossimilhança é dada por  $\prod_{i,d} \mathcal{L}_i(\mathcal{G}, \mathcal{S}, \sigma^2, \mathbf{M}_i | \mathbf{Y}_{id}, \mathbf{R}_i)$ , com  $\mathcal{L}_i$  definido como em (3.2).

## 3.2 O processo de maximização

Realizamos a maximização da log verossimilhança de forma iterativa, maximizando cada componente separadamente sem diminuir o valor de  $l$ . Na  $s$ -ésima iteração, atualizamos as estimativas de  $\mathcal{G}^{(s)}$ ,  $\mathcal{S}^{(s)}$ ,  $\sigma^{2(s)}$  e  $\mathbf{M}^{(s)}$  para  $\mathcal{G}^{(s+1)}$ ,  $\mathcal{S}^{(s+1)}$ ,  $\sigma^{2(s+1)}$  e  $\mathbf{M}^{(s+1)}$  de tal forma que  $l(\mathcal{G}^{(s+1)}, \mathcal{S}^{(s+1)}, \sigma^{2(s+1)}, \mathbf{M}^{(s+1)}) \geq l(\mathcal{G}^{(s)}, \mathcal{S}^{(s)}, \sigma^{2(s)}, \mathbf{M}^{(s)})$ . Inicializamos o procedimento com  $\mathbf{M}_i^{(0)} \equiv \mathbf{R}_i$  e repetimos os dois passos a seguir até a convergência.

1. Dado  $\mathbf{M}^{(s)}$ , sejam  $\mathcal{G}^{(s+1)}$ ,  $\mathcal{S}^{(s+1)}$  e  $\sigma^{2(s+1)}$  que maximizam a log verossimilhança:

$$l(\mathcal{G}, \mathcal{S}, \sigma^2, \mathbf{M}^{(s)}) = \sum_{i=1}^I \log f(\mathbf{Y}_i | \mathcal{G}, \mathcal{S}, \sigma^2, \mathbf{M}_i^{(s)}) + \sum_{i=1}^I \log P\{\mathbf{R}_i | \mathbf{M}_i^{(s)}\}.$$

2. Dados  $\mathcal{G}^{(s+1)}$ ,  $\mathcal{S}^{(s+1)}$  e  $\sigma^{2(s+1)}$ , seja  $\mathbf{M}^{(s+1)}$  que maximiza a log verossimilhança:

$$l(\mathcal{G}^{(s+1)}, \mathcal{S}^{(s+1)}, \sigma^{2(s+1)}, \mathbf{M}) = \sum_{i=1}^I \log f(\mathbf{Y}_i | \mathcal{G}^{(s+1)}, \mathcal{S}^{(s+1)}, \sigma^{2(s+1)}, \mathbf{M}_i) + \sum_{i=1}^I \log P\{\mathbf{R}_i | \mathbf{M}_i\}.$$

Vamos discutir agora, como realizamos cada um dos dois passos de maximização.

### Passo 1: Atualizando estimativas de $\mathcal{G}$ , $\mathcal{S}$ e $\sigma^2$

Utilizando o fato de que  $\mathbf{Y}_i$  tem distribuição normal multivariada dada pela Expressão (2.5), devemos minimizar

$$l_1(\mathcal{G}, \mathcal{S}, \sigma^2) \equiv \sum_{i=1}^I \log \left| \Psi \sum_{c=1}^C M_{ci} \Sigma_c \Psi' + \sigma^2 \mathbf{I} \right| \\ + \sum_{i=1}^I \left( \mathbf{Y}_i - \sum_{c=1}^C M_{ci} \Phi \gamma^c \right)' \left( \Psi \sum_{c=1}^C M_{ci} \Sigma_c \Psi' + \sigma^2 \mathbf{I} \right)^{-1} \left( \mathbf{Y}_i - \sum_{c=1}^C M_{ci} \Phi \gamma^c \right)$$

em relação a  $\mathcal{G}$ ,  $\mathcal{S}$  e  $\sigma^2$  mantendo os  $M_{ci}$ 's fixos. Realizamos a minimização de  $l_1$  novamente de forma iterativa, em três passos. A forma como escolhemos valores iniciais para  $\mathcal{S}$  e  $\sigma^2$  será discutida posteriormente.

1.a Dados  $\mathcal{S}^{(s)}$  e  $\sigma^{2(s)}$ , encontremos  $\mathcal{G}^{(s+1)}$  que minimize  $l_1$ . Esse passo não impõe dificuldades e pode ser feito explicitamente. Encontramos estimativas para  $\mathcal{G}^{(s)}$  através da seguinte forma fechada

$$\hat{\mathcal{G}} = \left( \sum_{i=1}^I \left( [\Phi M_{1i} \mathbf{I} \dots \Phi M_{Ci} \mathbf{I}]' \left( \Psi \sum_{c=1}^C M_{ci} \Sigma_c \Psi' + \sigma^2 \mathbf{I} \right)^{-1} [\Phi M_{1i} \mathbf{I} \dots \Phi M_{Ci} \mathbf{I}] \right) \right) \\ \times \sum_{i=1}^I \left( [\Phi M_{1i} \mathbf{I} \dots \Phi M_{Ci} \mathbf{I}]' \left( \Psi \sum_{c=1}^C M_{ci} \Sigma_c \Psi' + \sigma^2 \mathbf{I} \right)^{-1} \mathbf{Y}_i \right). \quad (3.3)$$

1.b Dado  $\mathcal{G}^{(s+1)}$  e  $\mathcal{S}^{(s)}$ , encontremos  $\sigma^{2(s+1)}$  que minimize  $l_1$ . Esse passo é feito minimizando  $l_1$  com respeito à  $\sigma^2$ . Note que aqui não é possível encontrar uma forma fechada para a

estimativa de  $\sigma^2$  e por essa razão é necessário o uso de um método numérico de minimização.

1.c Dado  $\mathcal{G}^{(s+1)}$  e  $\sigma^{2(s+1)}$ , encontremos  $\mathcal{S}^{(s+1)}$  que minimize  $l_1$ . Nesse passo, também não é possível encontrar uma forma fechada para a estimativa e utilizamos um método numérico de estimação. Ademais, formas simplificadas das matrizes de covariância  $\Sigma_c$ 's são utilizadas para facilitar o processo, como explicado a seguir.

Note agora que podemos minimizar  $l_1$  com a restrição de que  $\Sigma_c = \sigma_{\gamma,c}^2 \mathbf{I}$ , de tal forma que precisamos estimar no passo 3:  $\mathcal{S} = \{\sigma_{\gamma,1}^2, \dots, \sigma_{\gamma,C}^2\}$ . Com essa simplificação, estimamos um número menor de parâmetros e alguns cálculos podem ser feitos no começo do processo iterativo, o que é computacionalmente mais viável. Vamos reescrever  $l_1$  utilizando essa restrição e a decomposição espectral de  $\Psi\Psi'$  de tal forma que basta encontrar  $\sigma_{\gamma,c}^2$ 's que minimizem  $l_1$ . Primeiro, note-se que podemos escrever

$$\Psi\Psi' = P'\Lambda P$$

onde  $\Lambda$  é diagonal e  $P$  ortogonal.

Desta forma, a inversa da matriz de covariância de  $\mathbf{Y}_i$  é dada por

$$\begin{aligned} \left[ \Psi \sum_{c=1}^C M_{ci} \Sigma_c \Psi' + \sigma^2 \mathbf{I} \right]^{-1} &= \left[ \left( \sum_{c=1}^C M_{ci} \sigma_{\gamma,c}^2 \right) \Psi\Psi' + \sigma^2 \mathbf{I} \right]^{-1} \\ &= P' \left[ \left( \sum_{c=1}^C M_{ci} \sigma_{\gamma,c}^2 \right) \Lambda + \sigma^2 \mathbf{I} \right]^{-1} P \equiv P' \mathcal{D}^{-1} P. \end{aligned}$$

Seja  $\mathbf{Y}_i^* = P \left( \mathbf{Y}_i - \Phi \sum_{c=1}^C M_{ci} \gamma^c \right)$ . Então, devemos encontrar  $\mathcal{S} = \{\sigma_{\gamma,1}^2, \dots, \sigma_{\gamma,C}^2\}$  que minimize

$$\begin{aligned} l_1(\mathcal{G}, \mathcal{S}, \sigma^2) &= \sum_i \log |\mathcal{D}| + \sum_i \mathbf{Y}_i^{*2} \mathcal{D}^{-1} \mathbf{Y}_i^* \\ &= \sum_i \sum_j \log \left( \sum_{c=1}^C M_{ci} \sigma_{\gamma,c}^2 \Lambda[j, j] + \sigma^2 \right) + \sum_i \sum_j \frac{\mathbf{Y}_i^{*2}[j]}{\sum_c M_{ci} \sigma_{\gamma,c}^2 \Lambda[j, j] + \sigma^2}. \end{aligned}$$

As matrizes  $P$  e  $\Lambda$  são calculadas através das bases antes da primeira iteração e então usadas durante o processo todo. Note que, para o caso de um conjunto de dados com replicações, maximizamos a verossimilhança da mesma forma, seguindo os três passos descritos acima.

Daremos mais detalhes agora de como encontrar valores iniciais para os Passos 1.b e 1.c. Como podemos ver, não é necessário inicializar  $\mathcal{G}$  no Passo 1.a pois encontramos uma forma fechada para a sua estimativa.

Uma forma de encontrar estimativa inicial para  $\sigma^2$  no caso em que os dados não possuem replicação é utilizando a soma do quadrado dos resíduos a partir da suavização dos dados observados no  $i$ -ésimo transformador; chamemos de  $\hat{\mathbf{Y}}_i$  os dados suavizados do  $i$ -ésimo transformador utilizando uma suavizações por funções splines. Supondo que usamos o mesmo parâmetro de suavização para todos os transformadores, o qual pode ser escolhido visualmente, por exemplo, seja  $gl$  o número de graus de liberdade resultante de cada um dos ajustes. Então temos

$$\hat{\sigma}^{2(0)} = \frac{\sum_{i=1}^I \sum_{j=1}^n \|\mathbf{Y}_i[j] - \hat{\mathbf{Y}}_i[j]\|^2}{I(n - gl)}. \quad (3.4)$$

Se os dados tiverem replicação, suavizamos cada replicação do transformador  $i$ , e usamos os resíduos resultantes das  $D$  suavizações para calcular  $\hat{\sigma}^{2(0)}$ .

Para encontrar estimativas iniciais do Passo 1.c, vamos assumir que  $\Sigma_c = \sigma_{\gamma,c}^2 \mathbf{I}$ , e portanto  $\mathcal{S} = \{\sigma_{\gamma,1}^2, \dots, \sigma_{\gamma,C}^2\}$ . As estimativas iniciais dos  $\sigma_c^2$ 's são encontradas através do método dos momentos. Primeiramente, reparametrizamos o modelo como  $\sigma_{\gamma,c}^2 = s_c \sigma_{\gamma,C}^2$ , para  $c = 1, \dots, C - 1$ . Então, utilizamos o fato de que  $\mathbf{Y}_i$  tem distribuição normal multivariada com matriz de covariância  $[\sum_{c=1}^C M_{ci} \Sigma_c] \Psi \Psi' + \sigma^2 \mathbf{I}$ . Fixando  $s_C = 1$ , para o caso sem replicação, segue que

$$\sum_{i,j} \text{var}(\mathbf{Y}_i[j]) = \sigma_{\gamma,C}^2 \sum_i \left[ \sum_{c=1}^C M_{ci} s_c \right] \text{traço}(\Psi \Psi') + I \times n \times \sigma^2. \quad (3.5)$$

Utilizamos nossas estimativas iniciais de  $M_{ci}$  e  $\gamma^c$  para estimar o lado esquerdo da Equação (3.5)

$$\sum_j \widehat{\text{var}}(\mathbf{Y}_i[j]) = \|\mathbf{Y}_i - \widehat{M}_{ci}^{(0)} \Phi \hat{\gamma}^{c(0)}\|^2.$$

Para o caso onde observamos  $D$  replicações do transformador  $i$ , basta considerar as  $D$  replicações nos dois lados de (3.5). Neste caso, onde vários dias são observados, estimamos a variância de  $\mathbf{Y}_{i,d}[j]$  através da variância amostral de  $\mathbf{Y}_{i,1}[j], \dots, \mathbf{Y}_{i,D}[j]$ .

Substituindo nossas estimativas iniciais de  $M_{ci}$  e  $\sigma^2$ , assim como as variâncias amostrais encontradas na Equação (3.5) e resolvendo com respeito a  $\sigma_{\gamma,C}^2$ , obtemos  $\hat{\sigma}_{\gamma,C}^{2(0)}$  e, conseqüentemente, todas as outras estimativas iniciais:  $\hat{\sigma}_{\gamma,c}^{2(0)} \equiv s_c \hat{\sigma}_{\gamma,C}^{2(0)}$ .

Assim, finalizamos o processo iterativo do Passo 1. Agora daremos detalhes de como estimar  $\mathbf{M}_1, \dots, \mathbf{M}_I$  no Passo 2.

## Passo 2: Atualizando estimativas de $\mathbf{M}_1, \dots, \mathbf{M}_I$

Neste passo, a maximização da verossimilhança com relação a  $\mathbf{M}_1, \dots, \mathbf{M}_I$  pode ser feita separadamente para cada transformador considerando  $\mathcal{G}, \mathcal{S}, \sigma^2$  constantes. Isto é, fixamos o transformador  $i$  e procuramos  $M_{ci}$ ,  $c = 1, \dots, C$ , que maximize  $\mathcal{L}_i(\mathcal{G}, \mathcal{S}, \sigma^2, \mathbf{M}_i)$  em (3.2), para  $i = 1, \dots, I$ .

A maior dificuldade em relação a este passo está no cálculo da função  $P\{\mathbf{R}_i|\mathbf{M}_i\}$ . Esta função não possui forma analítica, e métodos convencionais de maximização não são adequados já que  $M_{ci}$  é um número inteiro e próximo de  $R_{ci}$ .

Uma alternativa é aproximar  $P\{\mathbf{R}_i|\mathbf{M}_i\}$  via simulação. A forma mais direta de simulação aqui, é gerar um número grande de variáveis multinomiais  $\mathbf{R}_i$ 's conforme descrito na Seção 2.3.1 com todos os possíveis valores que  $\mathbf{M}_i$  pode assumir; por exemplo, se o número total de consumidores no transformador é 75, devemos simular 76 possibilidades, e em seguida contar quantas vezes os reportados gerados em cada simulação foram iguais aos reportados observados. Assim, o valor de  $\mathbf{M}_i$  que maximiza a log verossimilhança será aquele que possuir a maior proporção de reportados gerados igual aos reportados observados. Além disso, esse processo teria que ser feito para cada transformador.

O modo de aproximar  $P\{\mathbf{R}_i|\mathbf{M}_i\}$  descrito acima é trabalhoso e demorado, razão pela qual criamos um método alternativo de aproximação que será descrito abaixo.

Na Seção 2.3.1 definimos a distribuição dos  $\mathbf{R}_i$ 's quando os  $\mathbf{M}_i$ 's são valores conhecidos, ou seja, a distribuição das linhas da Tabela 2.1. No entanto, seria interessante aproximarmos a função  $P\{\mathbf{R}_i|\mathbf{M}_i\}$  através de alguma distribuição que não dependa do número verdadeiro de consumidores em cada classe, já que essa informação é desconhecida nos dados observados.

Por conveniência, vamos omitir na notação o subíndice  $i$  que indica o transformador. Para  $j = 1, \dots, C$ , seja  $\tilde{\mathbf{X}}_j$  com distribuição multinomial( $r_j, p_{1j}, \dots, p_{Cj}$ ), tal que

$$p_{cj} = \frac{\mathcal{F}(c, j)}{\sum_{l=1}^C \mathcal{F}(l, j)}. \quad (3.6)$$

Então,  $\tilde{\mathbf{X}}_j$  é um vetor que divide os  $r_j$  consumidores que reportaram ser da classe  $j$ , nas classes em que eles pertencem de fato. Com isso, conseguimos definir o Teorema 1, que nos permite aproximar  $P\{\mathbf{R} = r|\mathbf{M} = m\}$  com apenas uma simulação.

**Teorema 1.** *Considere as variáveis aleatórias  $R_1, \dots, R_C$  definidas na Seção 2.3.1 e  $\tilde{\mathbf{X}}_1, \dots, \tilde{\mathbf{X}}_C$*

independentes e com distribuição definida acima, então

$$\begin{aligned} P\{\mathbf{R} = \mathbf{r} | \mathbf{M} = \mathbf{m}\} &= P\{R_1 = r_1, \dots, R_C = r_C | M_1 = m_1, \dots, M_C = m_C\} \\ &= \frac{\prod_{j=1}^C \left[ \sum_{c=1}^C \mathcal{F}(c, j) \right]^{r_j}}{\prod_{j=1}^C r_j!} \times \prod_{c=1}^C m_c! \times H(m_1, \dots, m_C) \end{aligned}$$

onde

$$H(m_1, \dots, m_C) = E \left( \mathbb{1} \left\{ \sum_{j=1}^C \tilde{\mathbf{X}}_j[c] = m_c, c = 1, \dots, C \right\} \right).$$

*Comentário:* Para aproximar essa probabilidade, aproximamos  $H$  simulando  $B$  conjuntos independentes com a distribuição de  $\{\tilde{\mathbf{X}}_1, \dots, \tilde{\mathbf{X}}_C\}$  denotados por  $\{\tilde{\mathbf{X}}_{1b}, \dots, \tilde{\mathbf{X}}_{Cb}\}$ ,  $b = 1, \dots, B$ . Ou seja,

$$\hat{H}(m_1, \dots, m_C) = \frac{1}{B} \sum_{b=1}^B \mathbb{1} \left\{ \sum_{j=1}^C \tilde{\mathbf{X}}_j[c] = m_c, c = 1, \dots, C \right\}.$$

Antes de mostrarmos a prova do Teorema 1, vamos apresentar um exemplo simples de como ele pode ser utilizado.

*Exemplo:* Para entender o uso do Teorema 1, vamos supor que existem duas classes e os dados são referentes a um transformador que abastece 75 consumidores, com número de reportados em cada classe iguais a  $r_1 = 32$  e  $r_2 = 43$ . Suponhamos ainda que a matriz fraude seja dada por

$$\mathcal{F} = \begin{bmatrix} 0.98 & 0.02 \\ 0.05 & 0.95 \end{bmatrix}. \quad (3.7)$$

O modo mais direto, porém inviável, de aproximar  $P\{R_1 = 32, R_2 = 43 | M_1 = m_1, M_2 = m_2\}$  é fazer um estudo de simulação para cada conjunto possível de valores  $(m_1, m_2)$ , gerando vetores  $\mathbf{X}_1$  e  $\mathbf{X}_2$  com distribuição dada na Seção 2.3.1 e contando a proporção de vezes que  $\mathbf{X}_1 + \mathbf{X}_2$  é igual a  $(32, 43)'$ . O Teorema nos dá uma forma para  $P\{R_1 = 32, R_2 = 43 | M_1 = m_1, M_2 = m_2\}$  capaz de usar apenas uma simulação para todos os vetores  $(m_1, m_2)$ . Simulamos  $B$  conjuntos de dados, com o  $b$ -ésimo conjunto consistindo de dois vetores de multinomiais independentes dadas por

$$\tilde{\mathbf{X}}_1^b \sim \text{multinomial} \left( 32, \frac{0.98}{1.03}, \frac{0.05}{1.03} \right)$$

e

$$\tilde{\mathbf{X}}_2^b \sim \text{multinomial} \left( 43, \frac{0.02}{0.97}, \frac{0.95}{0.97} \right).$$

Simulamos  $B$  vetores independentes da primeira distribuição multinomial dada acima, cada um deles será formado pelas componentes:  $\tilde{X}_1^b[1]$  e  $\tilde{X}_1^b[2]$ , em que  $\tilde{X}_1^b[1]$  é o número de consumidores que de fato são da classe 1, dentre aqueles que reportaram ser da classe 1 e  $\tilde{X}_1^b[2]$  é o número de consumidores que são classe 2 dentre aqueles que reportaram ser da classe 1. O mesmo procedimento é feito para a segunda multinomial, da mesma forma cada um dos vetores será dado por:  $\tilde{X}_2^b[1]$  e  $\tilde{X}_2^b[2]$ . Então somamos cada um dos  $B$  vetores das duas multinomiais, de tal forma que:  $M_1^b = \tilde{X}_1^b[1] + \tilde{X}_2^b[1]$  e  $M_2^b = \tilde{X}_1^b[2] + \tilde{X}_2^b[2]$ . Na tabela abaixo, as colunas representam estas multinomiais geradas.

		Total da linha
	$\tilde{X}_1^b[1]$ $\tilde{X}_2^b[1]$	$M_1^b$
	$\tilde{X}_1^b[2]$ $\tilde{X}_2^b[2]$	$M_2^b$
Total da coluna	32   43	75

Tabela 3.1: Número aproximado de consumidores nas duas classes resultante das multinomiais geradas na  $b$ -ésima iteração.

Desta forma, podemos aproximar  $H$  por  $\hat{H}$ , usando os valores de  $M_1^b$  e  $M_2^b$  que resultam da simulação. Por exemplo,  $\hat{H}(32, 43)$  é a proporção de conjunto de dados simulados que tiveram  $M_1^b = 34$  e de  $M_2^b = 42$ .

Realizamos esse procedimento para  $B = 100,000$ . A Tabela 3.2 mostra os resultados das aproximações de  $\hat{H}(m1, 75 - m1)$  quando  $R_1 = 32$  e  $R_2 = 43$ .

$m_1$	$\hat{H}(m_1, 75 - m_1)$
24	0
25	0
26	0.002
27	0.007
28	0.027
29	0.075
30	0.166
31	0.256
32	0.255
33	0.143
34	0.051
35	0.014
36	0.003
37	0
38	0

Tabela 3.2: Estimativas de  $H(m_1, 75 - m_1)$  quando  $R_1 = 32$  e  $R_2 = 43$ 

Abaixo, daremos a prova do Teorema 1.

*Prova do Teorema 1* : Escreva  $P\{\mathbf{R} = \mathbf{r} | \mathbf{M} = \mathbf{m}\}$  como

$$\begin{aligned}
& \sum_{x_{cj}: \sum_c x_{cj} = r_j} P\{\mathbf{X}_1 = (x_{11}, \dots, x_{1C})\} \times P\{\mathbf{X}_2 = (x_{21}, \dots, x_{2C})\} \times \dots \times P\{\mathbf{X}_C = (x_{C1}, \dots, x_{CC})\} \\
&= \sum_{\substack{x_{cj}: \sum_c x_{cj} = r_j \\ \text{e } \sum_j x_{cj} = m_c}} \left\{ \left[ \frac{m_1!}{x_{11}! \dots x_{1C}!} \mathcal{F}(1, 1)^{x_{11}} \dots \mathcal{F}(1, C)^{x_{1C}} \right] \times \dots \right. \\
&\quad \left. \times \left[ \frac{m_C!}{x_{C1}! \dots x_{CC}!} \mathcal{F}(C, 1)^{x_{C1}} \dots \mathcal{F}(C, C)^{x_{CC}} \right] \right\}.
\end{aligned}$$

Rearranjando os termos, concluímos que esta probabilidade é igual a:

$$\begin{aligned}
& \frac{\prod_1^C m_c!}{\prod_1^C r_c!} \sum_{\substack{x_{cj}: \sum_c x_{cj} = r_j \\ \text{e } \sum_j x_{cj} = m_c}} \left\{ \left[ \frac{r_1!}{x_{11}! x_{21}! \dots x_{C1}!} \mathcal{F}(1, 1)^{x_{11}} \dots \mathcal{F}(C, 1)^{x_{C1}} \right] \times \dots \right. \\
&\quad \left. \times \left[ \frac{r_C!}{x_{1C}! x_{2C}! \dots x_{CC}!} \mathcal{F}(1, C)^{x_{1C}} \dots \mathcal{F}(C, C)^{x_{CC}} \right] \right\}
\end{aligned}$$

$$\begin{aligned}
&= \frac{\prod_1^C m_c!}{\prod_1^C r_c!} \sum_{\substack{x_{cj}: \sum_c x_{cj} = r_j \\ \text{e } \sum_j x_{cj} = m_c}} \left[ \frac{r_1!}{x_{11}! x_{21}! \cdots x_{C1}!} p_{11}^{x_{11}} p_{21}^{x_{21}} \cdots p_{C1}^{x_{C1}} \right] \times \cdots \times \\
&\quad \left[ \frac{r_C!}{x_{1C}! x_{2C}! \cdots x_{CC}!} p_{1C}^{x_{1C}} p_{2C}^{x_{2C}} \cdots p_{CC}^{x_{CC}} \right] \times \prod_{j=1}^C \left[ \sum_{c=1}^C \mathcal{F}(c, j) \right]^{r_j} \\
&= \frac{\prod_{j=1}^C \left[ \sum_{c=1}^C \mathcal{F}(c, j) \right]^{r_j}}{\prod_1^C r_c!} \times \prod_1^C m_c! \times \sum_{x_{cj}: \sum_j x_{cj} = m_c} \prod_{j=1}^C P\{\tilde{\mathbf{X}}_j = (x_{1j}, \dots, x_{Cj})\} \\
&= \frac{\prod_{j=1}^C \left[ \sum_{c=1}^C \mathcal{F}(c, j) \right]^{r_j}}{\prod_1^C r_c!} \times \prod_{c=1}^C m_c! \times H(m_1, \dots, m_C).
\end{aligned}$$

Mostramos então, a partir do Teorema 1, que podemos calcular  $P\{\mathbf{R}_i | \mathbf{M}_i\}$  na log verossimilhança facilmente através da função  $H_i$ , ou seja, precisamos encontrar  $\mathbf{M}_i$  que minimize:

$$\begin{aligned}
l_i^*(\mathbf{M}_i | \mathbf{Y}_i, \mathbf{R}_i) &\equiv l_i^*(\mathbf{M}_i) \equiv \frac{1}{2} \log \left| \Psi \sum_{c=1}^C M_{ci} \Sigma_c \Psi' + \sigma^2 \mathbf{I} \right| \\
&+ \frac{1}{2} \left( \mathbf{Y}_i - \sum_{c=1}^C M_{ci} \Phi \gamma^c \right)' \left( \Psi \sum_{c=1}^C M_{ci} \Sigma_c \Psi' + \sigma^2 \mathbf{I} \right)^{-1} \left( \mathbf{Y}_i - \sum_{c=1}^C M_{ci} \Phi \gamma^c \right) \\
&\quad - \sum_{c=1}^C \log M_{ci}! - \log H_i(\mathbf{M}_i).
\end{aligned}$$

Como pode-se observar na Tabela 3.2 do exemplo dado anteriormente, não existem muitos valores de  $\mathbf{M}_i$  para serem considerados. Muitas vezes, calcular  $l_i^*(\mathbf{M}_i)$  para todos os possíveis valores de  $\mathbf{M}_i$  pode não ser prático; para esses casos, podemos calcular  $l_i^*(\mathbf{M}_i)$  apenas para uma faixa pequena de valores de  $\mathbf{M}_i$  próximos à estimativa atual  $\mathbf{M}_i^{(s)}$ .

# Capítulo 4

## Aplicação em dados reais

Os dados que utilizaremos neste capítulo foram cedidos pela Companhia Paulista de Força e Luz (CPFL). Consistem de amostras da demanda de energia diária em três transformadores, observados durante cinco dias de uma determinada semana, a cada 15 minuto. A Figura 4.1 mostra os dados coletados de um transformador, onde as curvas representam a demanda total de energia medida em kVA (kilo-volt-ampere) de segunda-feira à sexta-feira. Os três transformadores servem juntos um total de 168 consumidores, dos quais 155 reportam ser residenciais e 13 reportam ser comerciais. Portanto, temos  $C = 2$  tipos de consumidores: residencial ( $c = 1$ ) e comercial ( $c = 2$ ). A classe dos consumidores residenciais é constituída de instalações monofásicas e instalações bifásicas, mas não consideramos tal diferenciação em nossa análise.

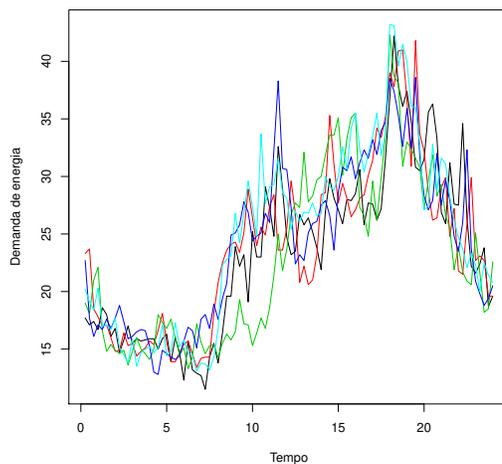


Figura 4.1: Dados do primeiro transformador com  $R_{11} = 34$  e  $R_{21} = 7$ . Cada curva representa a demanda total de energia para um dia da semana.

Assumimos que o consumidor comercial típico reporta ser residencial com probabilidade igual a 0,05, enquanto que um consumidor residencial reporta ser comercial com probabilidade igual a 0,02. A primeira probabilidade de fraude está relacionada ao fato de que o custo da energia para consumidores residenciais é menor do que para consumidores comerciais. A segunda probabilidade de fraude é menor do que a primeira, mas não é zero já que esse cenário é bastante improvável mas não impossível; imagine-se, por exemplo, alguém que tenha um comércio no mesmo local onde mora, decide fechar o comércio e encerrar as atividades mas não comunique imediatamente a companhia fornecedora de energia. Tal caso não constitui propriamente uma fraude, mas gera um registro que de fato não corresponde à realidade. Deste modo, a matriz fraude é da forma dada em (3.7).

Todo os procedimentos de maximização são feitos com o *software R*. No Passo 1.a atualizamos as estimativas de  $\gamma^1$  e  $\gamma^2$  através da Equação (3.3). No Passo 1.b, para atualizar estimativas de  $\sigma^2$ , utilizamos a função *optimize* do *R*, a qual é geralmente usada para otimizações de parâmetros unidimensionais. Para as estimativas de  $\sigma_{\gamma,1}^2$  e  $\sigma_{\gamma,2}^2$  do Passo 1.c, precisamos de um método de otimização bidimensional, para este propósito optamos aqui por utilizar a função *optim* do *R* com o método L-BFGS-B, que é uma modificação do método *quasi-Newton*. Tal função permite a especificação de limites inferior e superior para cada variável, o que é importante em nosso caso, já que valores negativos para os parâmetros da variância não fazem sentido e podem causar problemas no processo de maximização.

Utilizamos uma modificação direta do procedimento do Passo 2 da Seção 3.2 para acrescentar replicações. Neste mesmo passo, fixamos  $B = 100.000$  para aproximar os valores de  $H$  e tabelamos os resultados. Vale ressaltar que, os valores de  $\mathbf{M}_1$  e  $\mathbf{M}_2$  que não ocorrem na tabela de valores simulados não serão considerados como valores possíveis no cálculo de  $l_i^*(\mathbf{M}_i)$ . Portanto, utilizar um valor muito pequeno para  $B$  não é recomendável, já que restringe o número de valores possíveis de  $\mathbf{M}_1$  e  $\mathbf{M}_2$ . Por outro lado, valores muito grandes de  $B$  são computacionalmente demorados e geram valores de  $\mathbf{M}_1$  e  $\mathbf{M}_2$  distantes do número reportado, e portanto pouco prováveis de serem iguais ao número real de consumidores. Utilizando  $B = 100.000$ , pudemos facilmente calcular  $l_i^*(\mathbf{M}_i)$  para todos os valores plausíveis de  $\mathbf{M}_i$ , os quais já foram tabelados, e então escolher aquele que minimiza  $l_i^*$ .

Podemos analisar esses dados considerando apenas uma curva por transformador, e utilizar o método sem replicação descrito na Seção 3.2. Para obter uma única curva por transformador basta considerar cada dia da semana separadamente ou usar a média dos cinco dias. No entanto, nossos estudos de simulações mostram que utilizar os dias de um mesmo transformador simultaneamente, como replicações, produz melhores resultados. Desta forma, modelamos a demanda

de energia para cada transformador  $i$  com  $D = 5$  replicações independentes:  $\mathbf{Y}_{i1}, \dots, \mathbf{Y}_{i5}$ .

Utilizamos  $\psi$ 's assim como  $\phi$ 's, um conjunto de nove B-splines cúbicos com nós igualmente espaçados e consideramos o caso onde  $\Sigma_c = \sigma_{\gamma,c}^2 \mathbf{I}$ ,  $c = 1, 2$ . Encontramos estimativas iniciais para  $\sigma^2$ ,  $\sigma_{\gamma,1}^2$  e  $\sigma_{\gamma,2}^2$  com o método descrito na Seção 3.2.

Calculamos  $\hat{\mathbf{Y}}_{id}$  utilizando a função *smooth.spline* do *software R*, a qual ajusta um spline cúbico aos dados, escolhemos o parâmetro de suavização a olho, o qual resultou no uso de 10 graus de liberdade para ajustar cada replicação. Encontramos  $\hat{\sigma}^{2(0)} = 7,16$  como estimativa inicial de  $\sigma^2$ , através de

$$\hat{\sigma}^{2(0)} = \frac{\sum_{i=1}^3 \sum_{d=1}^5 \sum_{j=1}^{96} \|\mathbf{Y}_{id}[j] - \hat{\mathbf{Y}}_{id}[j]\|^2}{15(96 - 10)}. \quad (4.1)$$

A fim de encontrar estimativas iniciais para  $\sigma_{\gamma,1}^2$  e  $\sigma_{\gamma,2}^2$  fixamos  $\sigma_{\gamma,2}^2 = 6 \sigma_{\gamma,1}^2$ , e a partir da Equação 3.5, para o caso com replicação temos que

$$\sum_{i=1}^3 \sum_{d=1}^5 \sum_{j=1}^{96} \text{var}(\mathbf{Y}_{id}[j]) = 5 \left\{ \sigma_{\gamma,1}^2 \sum_{i=1}^3 [M_{1i} + 6M_{2i}] \text{traço}(\Psi\Psi') + 3 \times 96 \times \sigma^2 \right\} \quad (4.2)$$

Para  $i = 1, 2, 3$ , calculamos  $\text{var}(\mathbf{Y}_{id}[j])$  através da variância amostral dos cinco dias da semana. Substituímos  $\sigma^2$  por sua estimativa inicial  $\hat{\sigma}^{2(0)}$  e  $M_{1i}$  e  $M_{2i}$  pelos respectivos números reportados  $R_{1i}$  e  $R_{2i}$ .

Resolvendo a Equação (4.2) em relação a  $\sigma_{\gamma,1}^2$  resulta nas estimativas iniciais:  $\hat{\sigma}_{\gamma,1}^{2(0)} = 0,07$  e conseqüentemente,  $\hat{\sigma}_{\gamma,2}^{2(0)} = 0,45$ .

A Tabela 4.1 apresenta um resumo das estimativas encontradas: estimativa do número verdadeiro de consumidores em cada uma das classes, bem como estimativas de  $\sigma^2$ ,  $\sigma_{\gamma,1}^2$  e  $\sigma_{\gamma,2}^2$ .

A Figura 4.2, mostra as curvas de demanda típicas dos consumidores residenciais e comerciais, na figura superior podemos ver as estimativas de ambas as curvas no mesmo gráfico, enquanto que na figura inferior temos um gráfico para cada curva estimada. Como podemos ver, consumidores comerciais demandam mais energia elétrica do que consumidores residenciais durante qualquer momento do dia. Além disso, o pico de demanda de energia dos consumidores comerciais é quase seis vezes maior do que o pico dos consumidores residenciais.

Consumidores residenciais têm um pico relativamente alto por volta das 21 horas, certamente devido ao hábito brasileiro de tomar banho à noite; contudo, há um outro pico em torno do meio dia, provavelmente porque muitas pessoas costumam voltar para casa no horário de almoço.

Consumidores comerciais demandam mais energia das 5 da manhã até às 8 da noite, e seu pico acontece por volta das 18 horas. A principal diferença entre as duas classes de consumidores é que a demanda da classe comercial cresce continuamente das 5 da manhã até as 6 da tarde, enquanto a demanda dos residenciais possui algumas oscilações durante o dia.

É importante observar que não podemos quantificar a certeza dessas estimativas, já que não conhecemos seus desvios padrão.

Tabela 4.1: Resumo das estimativas

Transformador	Residencial		Comercial		
	Reportado	Estimado	Reportado	Estimado	
1	34	34	7	7	
2	47	46	3	4	
3	74	76	3	1	
		$\hat{\sigma}_{\gamma,1}^2 = 0,13$		$\hat{\sigma}_{\gamma,2}^2 = 0,25$	
$\hat{\sigma}^2 = 8,82$					

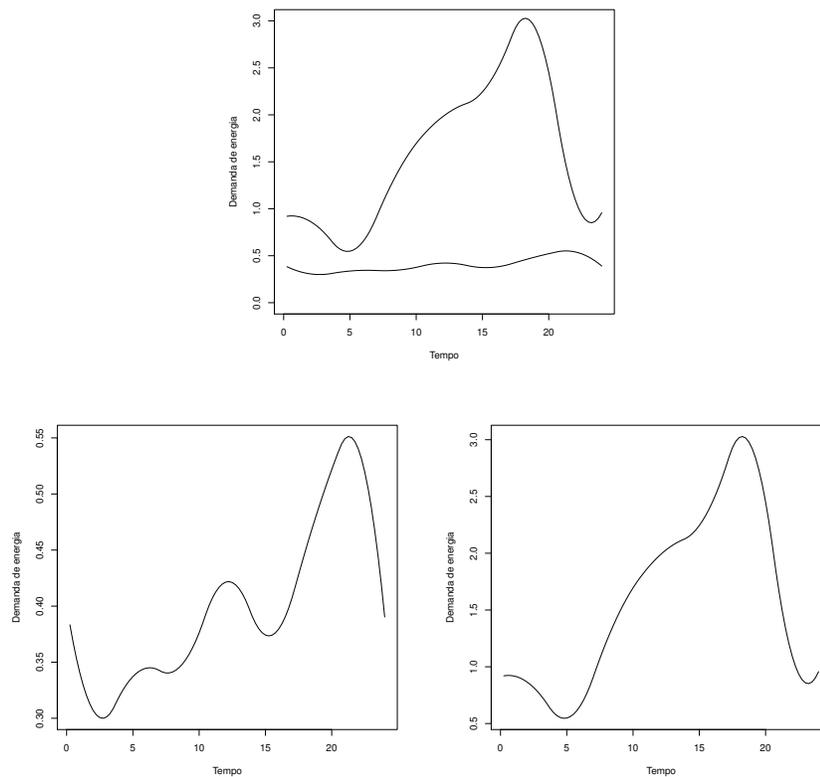


Figura 4.2: Tipologias estimadas dos consumidores residenciais e comerciais. A figura superior mostra as duas estimativas. A figura inferior mostra a estimativa da tipologia dos consumidores residenciais à esquerda e dos consumidores comerciais à direita.

# Capítulo 5

## Simulações

Primeiramente, realizamos oito simulações diferentes, gerando 200 conjuntos de dados em cada uma. Cada conjunto de dados foi obtido a partir de cinco transformadores, cada um abastecendo consumidores de duas classes: residencial ( $c = 1$ ) e comercial ( $c = 2$ ). Quatro das oito simulações possuem replicações das curvas de demanda de energia e as outras quatro não possuem.

Numa etapa posterior, para estudar o efeito do aumento do número de transformadores em nossas estimativas, fizemos o mesmo trabalho descrito acima, desta vez gerando dados a partir de 50 transformadores.

Consideramos para ambos os conjuntos de transformadores, com 5 e com 50, dois cenários distintos para as curvas de demanda típicas: uma similar ao conjunto de dados, com a curva de demanda da classe comercial muito maior do que a da classe residencial e outra com as curvas de demanda de ambas as classes na mesma escala. Também consideramos dois casos para o número de consumidores em cada classe: um similar ao conjunto de dados, com todos os transformadores contendo muito mais consumidores residenciais do que comerciais, e outro constituído de tal forma que a soma do número de consumidores residenciais em todos os transformadores fosse aproximadamente igual à soma do número de consumidores comerciais nos mesmos transformadores.

Na primeira seção apresentaremos o processo de geração dos dados e na segunda seção falaremos de algumas suposições utilizadas para a análise dos dados. Na Seção 5.3 mostraremos os principais resultados para as estimativas de cada parâmetro, e na Seção 5.4 apresentaremos algumas conclusões importantes.

## 5.1 Geração dos dados

Em todas as simulações, a demanda de energia em cada transformador foi observada em intervalos de 15 minutos, de forma a termos 96 observações por dia com  $t \in [0, 24]$ . Todos os conjuntos de dados foram gerados a partir das Equações (2.2), (2.3) e (2.4), com  $\text{Var}(\epsilon_i(t)) = 3, 5$ , para  $i = 1, \dots, I$ . Para um consumidor  $l$ , pertencente à classe  $c$  e abastecido pelo transformador  $i$ , construímos a componente  $\alpha_{cli}^*$  definida em (2.4), amostrando  $\gamma^{cli}$  de uma distribuição normal multivariada com média zero e matriz de covariância:  $\Sigma_c = \sigma_{\gamma,c}^2 \mathbf{I}$ . Para as funções bases, nós usamos  $\phi$ 's iguais a  $\psi$ 's, um conjunto de nove B-splines cúbicos com nós igualmente espaçados. Escolhemos  $\alpha_c$  e  $\sigma_{\gamma,c}^2$ ,  $c = 1, 2$ , de acordo com dois casos diferentes, conforme explicado mais adiante.

No  $i$ -ésimo transformador, para gerar o número reportado de consumidores nas duas classes usamos a matriz fraude dada em (3.7) e o fato de que o vetor de consumidores reportados na classe  $c$ ,  $X_c = (x_{c1}, x_{c2})$  tem uma distribuição multinomial com parâmetro de contagem  $M_c$  e parâmetros de probabilidade  $F(c, 1)$  e  $F(c, 2)$ . Então, para obter o vetor do número de consumidores reportados, nós amostramos de duas multinomiais  $X_1 = (x_{11}, x_{12})$  e  $X_2 = (x_{21}, x_{22})$  e somamos estas amostras, de tal forma que o número reportado de consumidores em cada classe é dado por:  $R_1 = x_{11} + x_{21}$  e  $R_2 = x_{12} + x_{22}$ . Para cada transformador, optamos por gerar o número de consumidores reportados uma vez e usá-lo em todas as simulações. Fizemos isso por duas razões principais: é computacionalmente mais rápido, já que precisamos calcular as funções  $H_i$ ,  $i = 1, \dots, I$  apenas uma vez, e também porque facilita a análise das propriedades do número verdadeiro estimado, como podemos ver nas Tabelas 5.2 e 5.3 discutidas na Seção 5.3.

Através de simulações iniciais percebemos que diferentes combinações de  $M_c$ 's e  $\alpha_c$ 's produziam diferentes resultados, portanto, conforme já mencionamos, decidimos estudar os quatro casos abaixo:

**Caso 1:** As funções  $\alpha_1$  e  $\alpha_2$  (Figura 5.1) estão na mesma escala e  $M_1$  e  $M_2$  são balanceados (Tabela 5.1).

**Caso 2:** As funções  $\alpha_1$  e  $\alpha_2$  estão na mesma escala (Figura 5.1) e  $M_1$  é muito maior do que  $M_2$  (Tabela 5.1).

**Caso 3:** A função  $\alpha_1$  é muito menor em escala (Figura 5.2) do que a função  $\alpha_2$  e estão na e  $M_1$  e  $M_2$  são balanceados (Tabela 5.1).

**Caso 4:** A função  $\alpha_1$  é muito menor em escala (Figura 5.1) do que a função  $\alpha_2$  e  $M_1$  é muito maior do que  $M_2$  (Tabela 5.1).

Note que não consideramos o caso onde  $\alpha_1$  é muito menor em escala do que  $\alpha_2$  e  $M_1$  é muito menor do que  $M_2$ , já que em simulações iniciais as estimativas de  $\alpha_1$  e  $\alpha_2$  foram bastante ruins

para esse caso.

Para relacionar esses casos com nosso conjunto de dados, considere que a classe 1 é a classe residencial e a classe 2 é a comercial. Então, esperamos que a demanda de energia de um consumidor residencial seja menor do que a de um consumidor comercial (Casos 3 e 4). Da mesma forma, tipicamente, o número de consumidores comerciais servidos pelos transformadores é menor do que o de consumidores residenciais (Casos 2 e 4).

A escolha das funções  $\alpha_1$  e  $\alpha_2$  foram baseadas em nosso conjunto de dados. Vale lembrar que, em todos os casos, nós geramos  $\alpha_1$  e  $\alpha_2$  a partir das Equações (2.3) e (2.4), com  $\phi$ 's e  $\psi$ 's um conjunto de nove bases B-spline cúbicos com nós igualmente espaçados. Para os Casos 3 e 4, usamos  $\gamma^1$  e  $\gamma^2$  iguais a  $\hat{\gamma}^1$  e  $\hat{\gamma}^2$ , estimativas estas provenientes de nosso conjunto de dados. Para os Casos 1 e 2, alteramos a escala de  $\hat{\gamma}^1$  e  $\hat{\gamma}^2$  para que todas as suas componentes estivessem entre zero e um. Por exemplo, na classe 1, seja  $a_1$  a menor componente de  $\hat{\gamma}^1$  e  $b_1$  a maior, então, temos que  $\alpha_1(t) = \phi'(t)\gamma^{*1}$ , com

$$\gamma^{*1}[k] = \frac{\hat{\gamma}^1[k] - a_1}{b_1 - a_1}. \quad (5.1)$$

Nos Casos 1 e 2, definimos  $\sigma_{\gamma,1}^2 = 0,03$  e  $\sigma_{\gamma,2}^2 = 0,06$ . Nos Casos 3 e 4, para manter a variabilidade relativa das curvas de demanda das classes, mudamos a escala de  $\sigma_{\gamma,c}^2$ , multiplicando pela constante apropriada:  $(b_c - a_c)^2, c = 1, 2$ . Em todos os casos, o valor de  $\sigma_{\gamma,2}^2$  é o dobro de  $\sigma_{\gamma,1}^2$  porque acreditamos que a variabilidade entre os consumidores comerciais é maior do que entre os residenciais.

A Figura 5.1 mostra um exemplo de dados simulados para o Caso 1 e a Figura 5.2 mostra um exemplo para o Caso 4. Em cada figura, o gráfico da esquerda mostra a demanda de energia de consumidores da classe 1 (residencial) e o gráfico da direita apresenta a demanda de energia de consumidores da classe 2 (comercial). Vale lembrar que essas curvas individuais não são observadas na prática: observamos apenas a soma das curvas dos gráficos da esquerda e da direita (demanda agregada). Alguns dos valores de demanda resultaram negativos, o que era esperado já que geramos os dados a partir de distribuições normais multivariadas com média zero, enquanto que em linhas de distribuição típicas a demanda não assume valores negativos. No entanto, devemos lembrar que demanda aqui é apenas uma analogia para o exemplo que estamos tratando e, uma vez que nosso interesse principal é verificar se as estimativas estão próximas das curvas utilizadas para gerar os dados, valores negativos não devem afetar nossas conclusões.

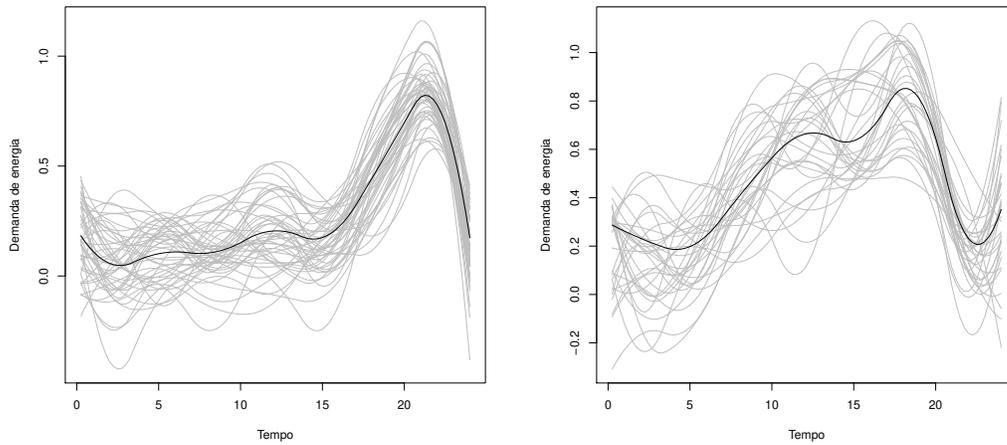


Figura 5.1: Exemplo das curvas individuais simuladas comparadas com as tipologias  $\alpha_1$  (curva mais escura no gráfico da esquerda) e  $\alpha_2$  (curva mais escura no gráfico da direita) para o Caso 1, com 45 consumidores da classe 1 e 30 da classe 2.

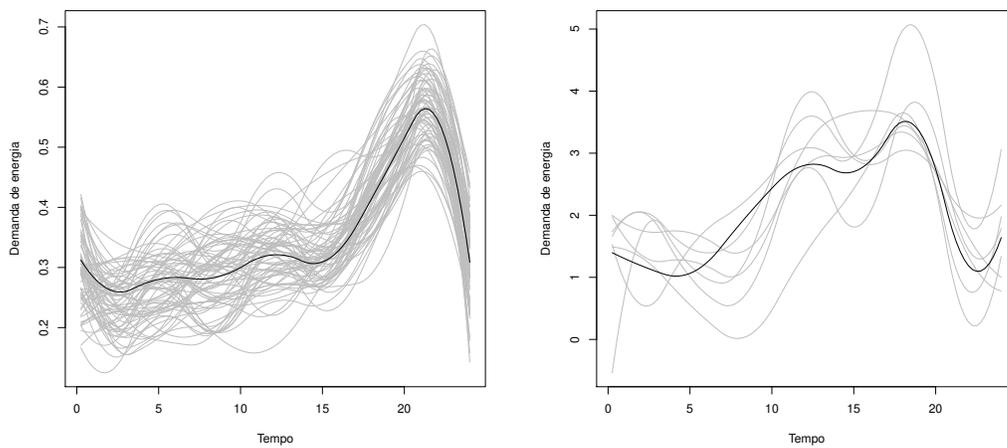


Figura 5.2: Exemplo das curvas individuais simuladas comparadas com as tipologias  $\alpha_1$  (curva mais escura no gráfico da esquerda) e  $\alpha_2$  (curva mais escura no gráfico da direita) para o Caso 4, com 69 consumidores da classe 1 e 6 da classe 2.

A Tabela 5.1 mostra o número de consumidores gerado em cada classe nas simulações com 5 transformadores. Os transformadores nos quais  $M_1$  e  $M_2$  foram gerados de forma balanceada foram utilizados para construir os Casos 1 e 3, enquanto que os transformadores com  $M_1$  muito maior que  $M_2$  foram utilizados nos Casos 2 e 4.

Transformadores	Balanceado		Desbalanceado	
	Verdadeiro	Reportado	Verdadeiro	Reportado
1	45	45	66	65
2	29	32	65	66
3	61	60	69	68
4	24	28	62	63
5	12	16	72	71

Tabela 5.1: O número verdadeiro de consumidores da classe 1 e o número reportado gerado aleatoriamente para cinco transformadores.

Nos quatro casos descritos, geramos dois tipos de conjunto de dados: um com apenas um dia observado por transformador (sem replicação) e outro com cinco dias observados por transformador (com replicação). Geramos os dias independentemente, o que é uma simplificação, já que provavelmente há relação entre a demanda de um dia e de outro.

## 5.2 Análise dos dados

Na análise, para todos os conjuntos de dados, assumimos a mesma matriz fraude  $F$  dada em (3.7) e as mesmas bases utilizadas para gerar os dados. Consideramos também o caso onde  $\Sigma_c = \sigma_{\gamma,c}^2 \mathbf{I}$ . Utilizamos os mesmos processos de maximização usados na análise do conjunto de dados, os quais foram descritos no Capítulo 4.

Vamos lembrar que a maximização do Passo 2 requer os valores da função  $H$ . Aproximamos  $H$  em cada transformador, simulando  $B = 100.000$  conjuntos independentes uma só vez e tabelamos os resultados como uma função de  $M_1$ . Pela Tabela 3.2, podemos perceber que o número de valores possíveis para  $M_1$  não é grande e está em torno de 32, desta forma é fácil calcular  $l_i^*(\mathbf{M}_i)$  para todos esses valores e então, escolher aquele que minimiza  $l_i^*$ .

## 5.3 Resultados

### 5.3.1 Estimativas dos $\alpha_c$ 's

Nas Figuras 5.3 - 5.7 os gráficos na parte superior referem-se a dados gerados sem replicação e os de baixo a dados obtidos com replicação. As curvas tracejadas mostram o máximo e mínimo das 200 estimativas de  $\alpha_1$  nos gráficos da esquerda, e  $\alpha_2$  nos gráficos da direita, os quartis e mediana aparecem em curvas pontilhas. Os verdadeiros valores de  $\alpha_1$  e  $\alpha_2$  estão representados

como uma linha sólida no mesmo gráfico.

As linhas na Figura 5.8 representam, para cada transformador do Caso 4 sem replicação, o máximo e mínimo (curvas tracejadas), quartis e mediana (curvas pontilhadas) da soma ponderada das tipologias estimada para os 200 conjuntos de dados, e a linha sólida nesses mesmos gráficos se refere à soma ponderada verdadeira. Para o transformador  $i$ , a soma ponderada estimada é dada simplesmente por

$$\hat{\mathbf{y}}_i = \hat{M}_{1i}\hat{\boldsymbol{\alpha}}_1 + \hat{M}_{2i}\hat{\boldsymbol{\alpha}}_2. \quad (5.2)$$

As Figuras de 5.3 a 5.6 referem-se aos Casos de 1 a 4 com dados de 5 transformadores. Como podemos ver pela Figura 5.3, referente ao Caso 1, as medianas de  $\boldsymbol{\alpha}_1$  e  $\boldsymbol{\alpha}_2$  estão bastante próximas da verdadeira e há pouca variabilidade nas estimativas.

Os Casos 2 e 3 apresentam cenários opostos, enquanto o Caso 2, representado na Figura 5.4, apresenta boas estimativas de  $\boldsymbol{\alpha}_1$  e estimativas não tão boas e com bastante variabilidade para  $\boldsymbol{\alpha}_2$ , o Caso 3, mostrado na Figura 5.5, tem estimativas piores e com mais variabilidade para  $\boldsymbol{\alpha}_1$  e estimativas melhores para  $\boldsymbol{\alpha}_2$ . No Caso 2, como podemos ver no gráfico da direita da Figura 5.4,  $\boldsymbol{\alpha}_2$  apresenta valores superestimados. No Caso 3,  $\boldsymbol{\alpha}_1$  é subestimado e  $\boldsymbol{\alpha}_2$  é superestimado, como mostrado na Figura 5.5.

O Caso 4 é aquele que mais traz desafios, como podemos ver pela Figura 5.6: tanto as estimativas de  $\boldsymbol{\alpha}_1$  como de  $\boldsymbol{\alpha}_2$  possuem bastante variabilidade, mesmo quando o conjunto de dados tem replicação. Vale lembrar que esse é o caso observado em nosso conjunto de dados, onde há muito mais consumidores residenciais do que comerciais e a curva de demanda de energia dos consumidores comerciais é bem maior do que dos residenciais.

A esta altura, para verificar a influência do número de transformadores nas estimativas das tipologias, fizemos o mesmo estudo para os 4 casos, agora com 50 transformadores. Como já era esperado, nos Casos 1, 2, 3 e 4 com 50 transformadores as estimativas das tipologias de ambas as classes são consideravelmente melhores do que aquelas com apenas 5 transformadores. Para exemplificar a melhora, a Figura 5.7 mostra as tipologias estimadas para o Caso 4 (justamente aquele com as piores estimativas mostradas anteriormente) com 50 transformadores. Como podemos ver, mesmo para o Caso 4, sem replicação, as estimativas possuem pouca variabilidade e estão bastante próximas da verdadeira.

De fato, as estimativas da soma ponderada das tipologias são sempre muito boas em todos os casos. A Figura 5.8 mostra essas estimativas para o primeiro transformador do Caso 4, com 5 transformadores e sem replicação, vê-se que a curva verdadeira está muito próxima da mediana e dos quartis das 200 estimativas.

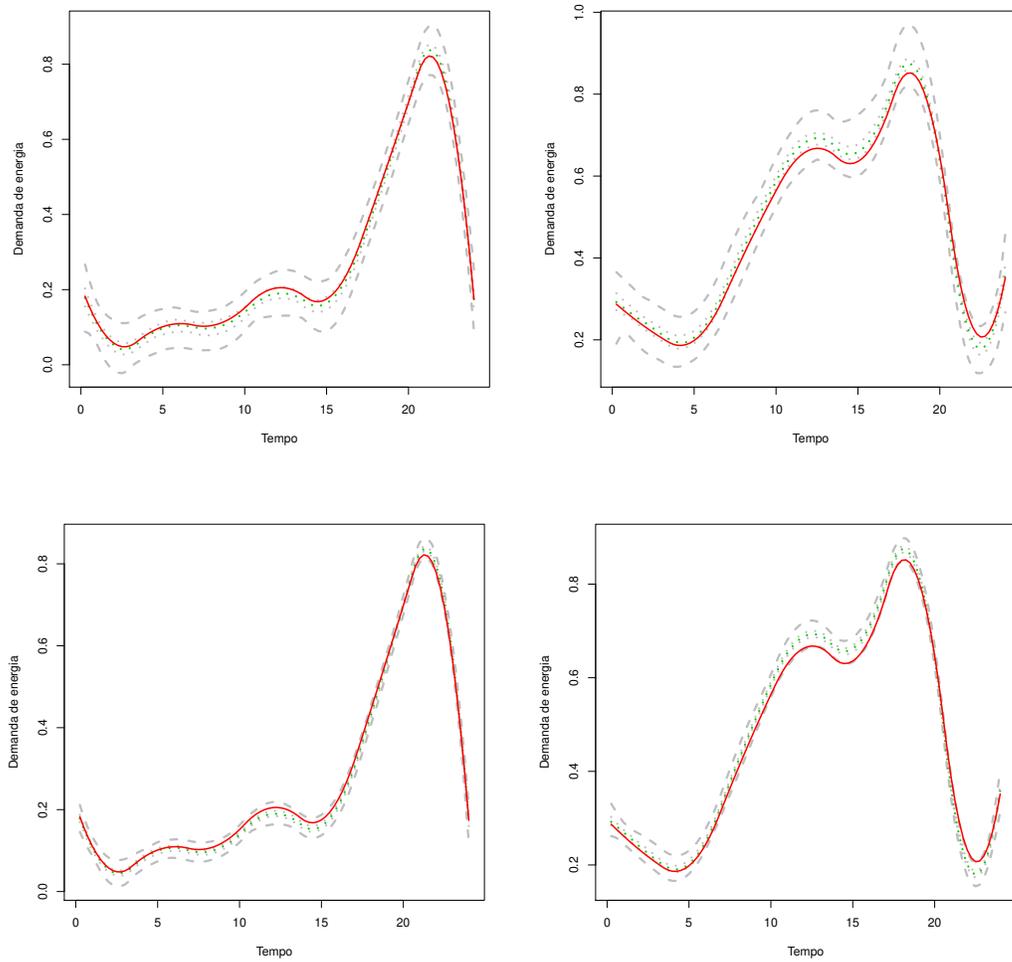


Figura 5.3: Mínimo e máximo (curvas tracejadas), mediana e quartis (curvas pontilhadas) das 200 estimativas da demanda de energia para o Caso 1, para as classe  $c = 1$  (coluna da esquerda) e  $c = 2$  (coluna da direita) com 5 transformadores, sem replicação (linha superior) e com replicação (linha inferior). As curvas sólidas representam os valores verdadeiros.

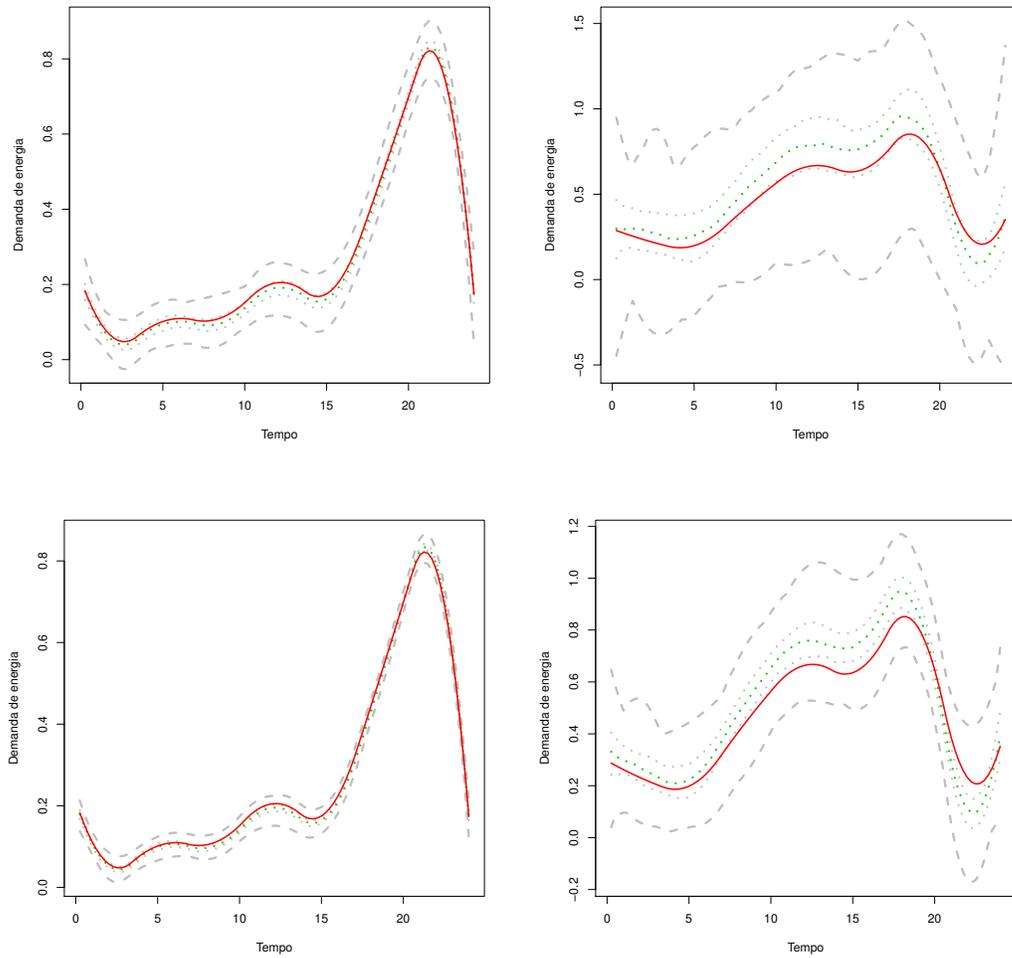


Figura 5.4: Mínimo e máximo (curvas tracejadas), mediana e quartis (curvas pontilhadas) das 200 estimativas da demanda de energia para o Caso 2, para as classe  $c = 1$  (coluna da esquerda) e  $c = 2$  (coluna da direita) com 5 transformadores, sem replicação (linha superior) e com replicação (linha inferior). As curvas sólidas representam os valores verdadeiros.

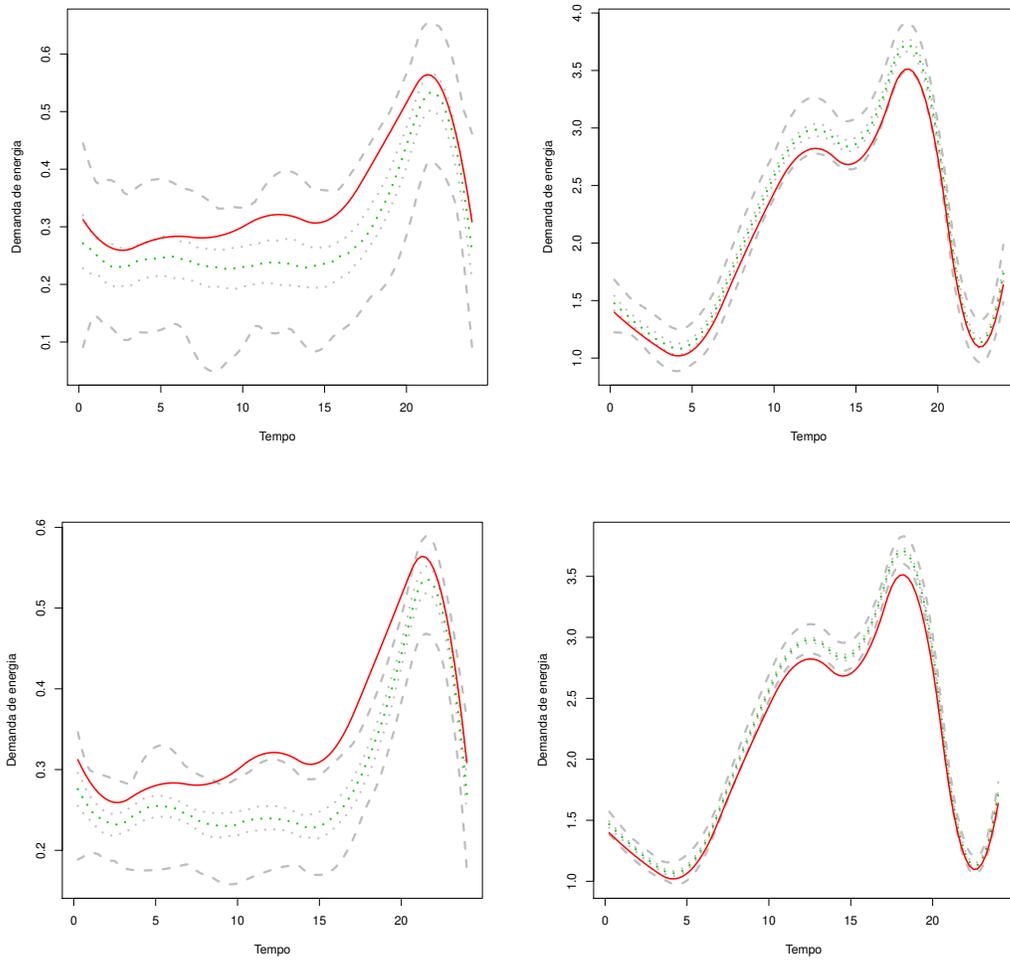


Figura 5.5: Mínimo e máximo (curvas tracejadas), mediana e quartis (curvas pontilhadas) das 200 estimativas da demanda de energia para o Caso 3, para as classe  $c = 1$  (coluna da esquerda) e  $c = 2$  (coluna da direita) com 5 transformadores, sem replicação (linha superior) e com replicação (linha inferior). As curvas sólidas representam os valores verdadeiros.

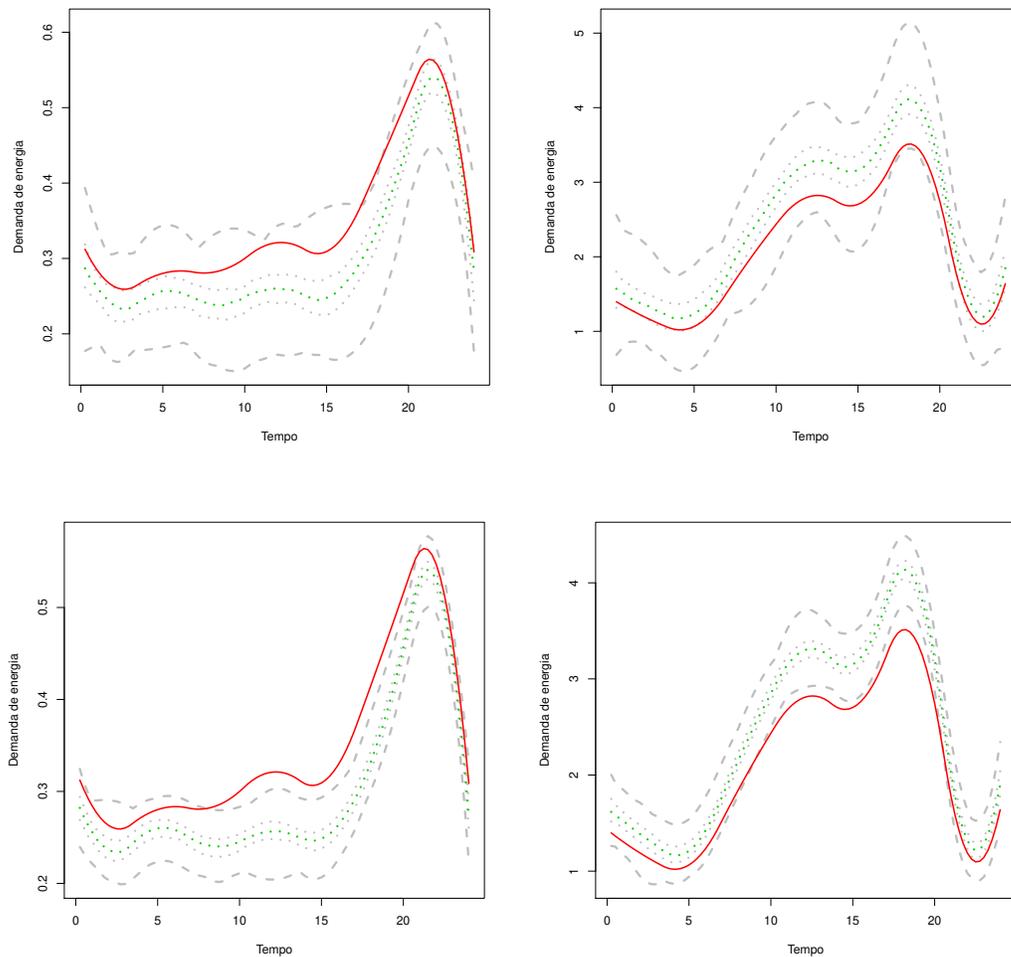


Figura 5.6: Mínimo e máximo (curvas tracejadas), mediana e quartis (curvas pontilhadas) das 200 estimativas da demanda de energia para o Caso 4, para as classe  $c = 1$  (coluna da esquerda) e  $c = 2$  (coluna da direita) com 5 transformadores, sem replicação (linha superior) e com replicação (linha inferior). As curvas sólidas representam os valores verdadeiros.

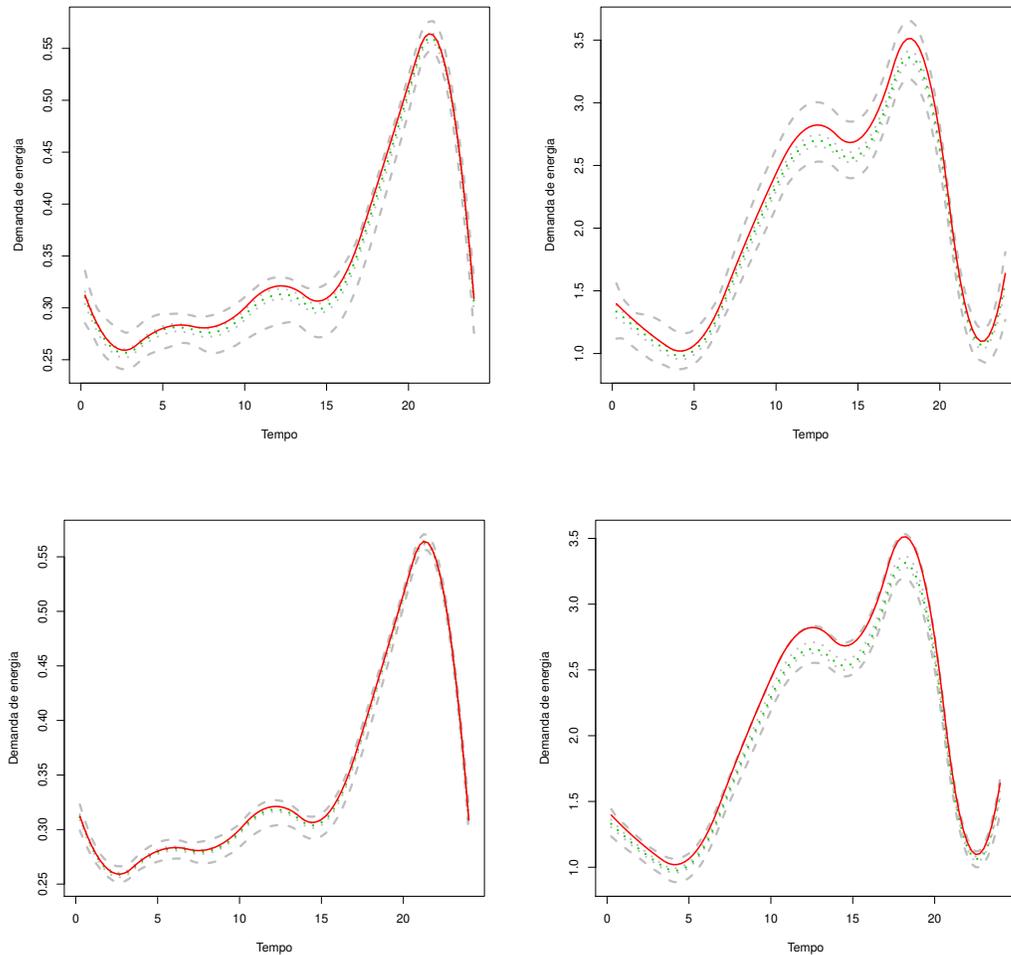


Figura 5.7: Mínimo e máximo (curvas tracejadas), mediana e quartis (curvas pontilhadas) das 200 estimativas da demanda de energia para o Caso 4, para as classe  $c = 1$  (coluna da esquerda) e  $c = 2$  (coluna da direita) com 50 transformadores, sem replicação (linha superior) e com replicação (linha inferior). As curvas sólidas representam os valores verdadeiros.

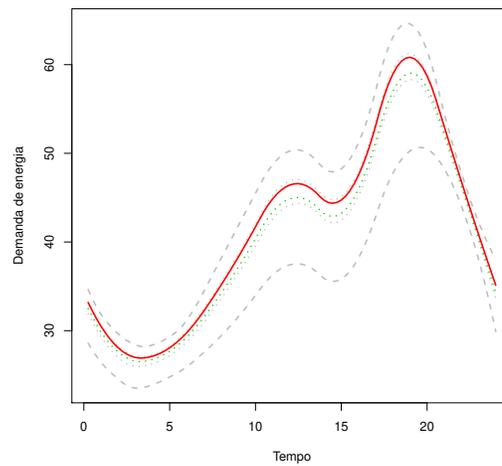


Figura 5.8: Mínimo e máximo (curvas tracejadas), mediana e quartis (curvas pontilhadas) das 200 estimativas da soma ponderada das tipologias no primeiro transformador, com 66 consumidores da classe 1 e 9 consumidores da classe 2, para o Caso 4 sem replicação, com 5 transformadores. A curva sólida representa os valores verdadeiros.

### 5.3.2 Estimativas dos parâmetros de variância

A Figura 5.9 mostra a densidade das estimativas de  $\sigma^2$  para as 200 simulações no Caso 4, com 5 transformadores. A linha vertical é o verdadeiro valor, o qual é igual a 3,5 em todos os conjuntos de dados gerados. Não houve muita diferença na estimação de  $\sigma^2$  entre os 4 casos, como podemos ver pela Figura 5.9, a estimação é boa, mesmo quando as curvas de demanda estão em escala diferente e o número de consumidores em cada classe é desbalanceado.

As Figuras de 5.10 a 5.13 apresentam a estimativa da função densidade de  $\sigma_{\gamma,1}^2$  baseada nos 200 conjuntos de dados gerados com replicação, nos gráficos da esquerda, e de  $\sigma_{\gamma,2}^2$  nos gráficos da direita, incluindo estimativas iguais a zero. A parte superior desses gráficos mostra estimativas para dados gerados a partir de 5 transformadores e a parte inferior para 50 transformadores. A linha vertical representa o verdadeiro valor utilizado para gerar os dados e o número de estimativas iguais a zero está indicado abaixo do eixo  $x$ .

As estimativas de  $\sigma_{\gamma,1}^2$  e  $\sigma_{\gamma,2}^2$  para os conjuntos de dados gerados sem replicação não ofereceram bons resultados. De fato, várias delas foram iguais a zero. Por esta razão, para tais parâmetros, vamos nos concentrar nas figuras dos casos com replicação.

A parte superior das Figuras 5.10 e 5.11 mostram que, mesmo quando observamos apenas 5 transformadores, as estimativas de  $\sigma_{\gamma,1}^2$  e  $\sigma_{\gamma,2}^2$  são razoavelmente boas se as curvas de demanda das duas classes estão na mesma escala, já que apenas alguns conjuntos de dados apresentam estimativas iguais a zero para esses cenários, e a maior densidade de  $\sigma_{\gamma,1}^2$  e  $\sigma_{\gamma,2}^2$  se encontra muito próxima dos valores verdadeiros. Nota-se pela parte inferior das mesmas figuras que quando observamos 50 transformadores nos Casos 1 e 2 as estimativas de ambos os parâmetros ficam muito boas e não há mais nenhum conjunto de dados com estimativa igual a zero.

Para os Casos 3 e 4 com 5 transformadores observados, as estimativas de  $\sigma_{\gamma,2}^2$  são boas mas temos problemas na estimação de  $\sigma_{\gamma,1}^2$ , parâmetro este relacionado com a variância da classe 1, a qual possui a demanda de energia bem menor nestes casos. Grande parte das estimativas de  $\sigma_{\gamma,1}^2$  é igual a zero, conforme podemos ver pelas Figuras 5.12 e 5.13. A parte inferior das figuras nos mostra que, quando observamos 50 transformadores ao invés de 5, as estimativas de  $\sigma_{\gamma,1}^2$  e  $\sigma_{\gamma,2}^2$  apresentam relativa melhora.

Conforme mostra a parte inferior da Figura 5.13, para o Caso 4 com 50 transformadores, o valor com maior densidade de  $\sigma_{\gamma,1}^2$  está em torno de 0,012, enquanto o valor verdadeiro é 0,005. Os valores estimados de  $\sigma_{\gamma,2}^2$  variam entre 0,6 e 0,95, sendo que aquele com a maior densidade está em torno de 0,75 e o valor usado para gerar os dados é igual a 0,84. Além disso, não há mais estimativas iguais a zero.

Para o Caso 3, a estimativa de  $\sigma_{\gamma,2}^2$  foi visivelmente boa. Por outro lado,  $\sigma_{\gamma,1}^2$  ainda possui mais da metade das estimativas iguais a zero, como vemos na parte inferior da Figura 5.12.

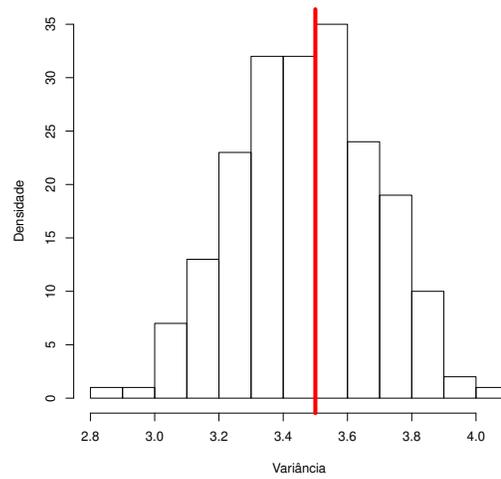


Figura 5.9: Densidade das 200 estimativas de  $\sigma^2$  para o Caso 4 sem replicação, com 5 transformadores. A linha vertical representa o valor verdadeiro.

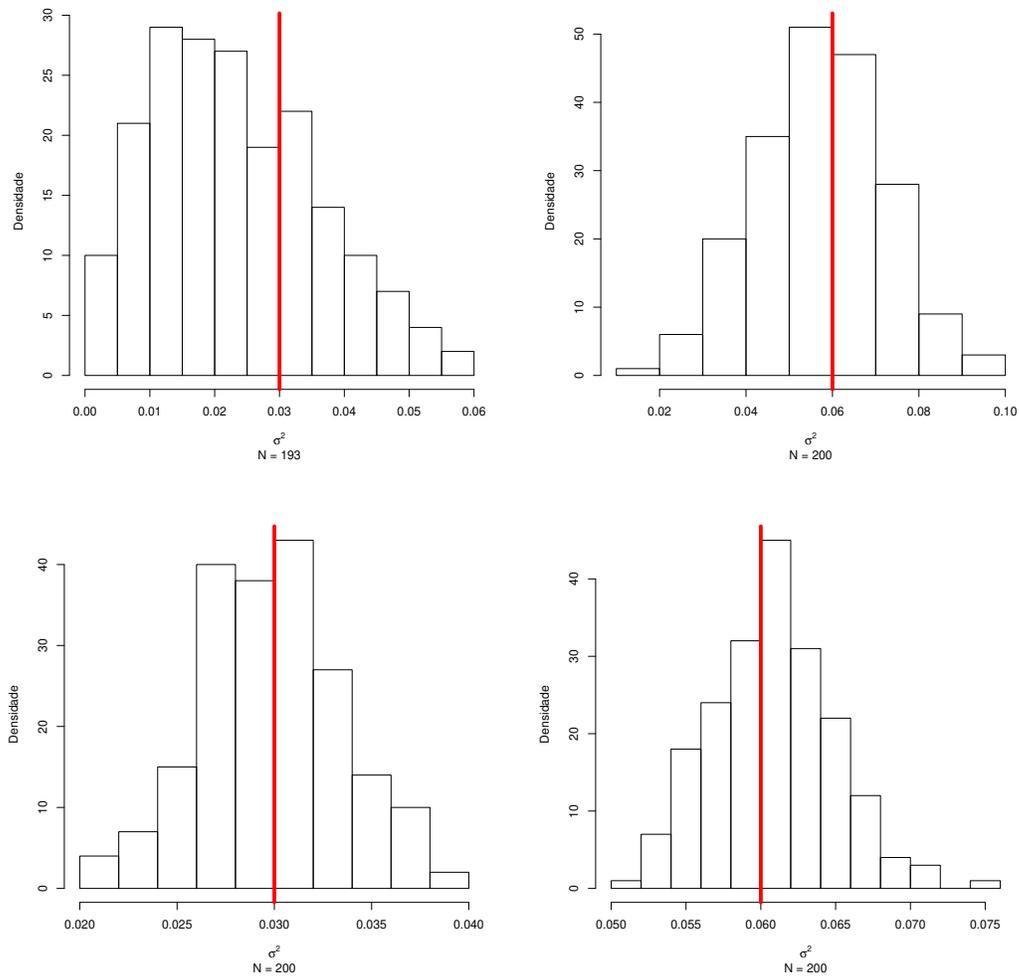


Figura 5.10: Densidade das  $N$  estimativas maiores do que zero de  $\sigma_{\gamma,1}^2$  (coluna da esquerda) e  $\sigma_{\gamma,2}^2$  (coluna da direita) para o Caso 1 com replicação, com 5 transformadores (linha superior) e 50 transformadores (linha inferior). A linha vertical representa o valor verdadeiro.

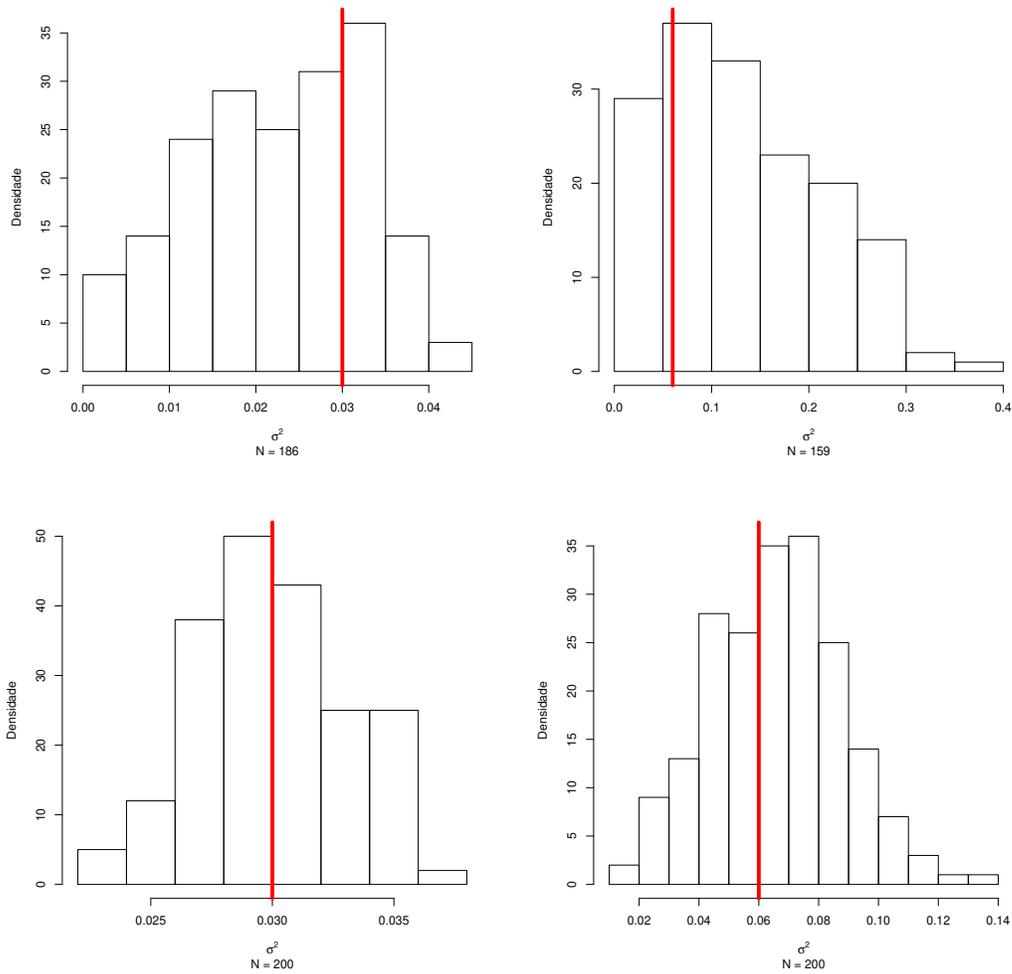


Figura 5.11: Densidade das  $N$  estimativas maiores do que zero de  $\sigma_{\gamma,1}^2$  (coluna da esquerda) e  $\sigma_{\gamma,2}^2$  (coluna da direita) para o Caso 2 com replicação, com 5 transformadores (linha superior) e 50 transformadores (linha inferior). A linha vertical representa o valor verdadeiro.

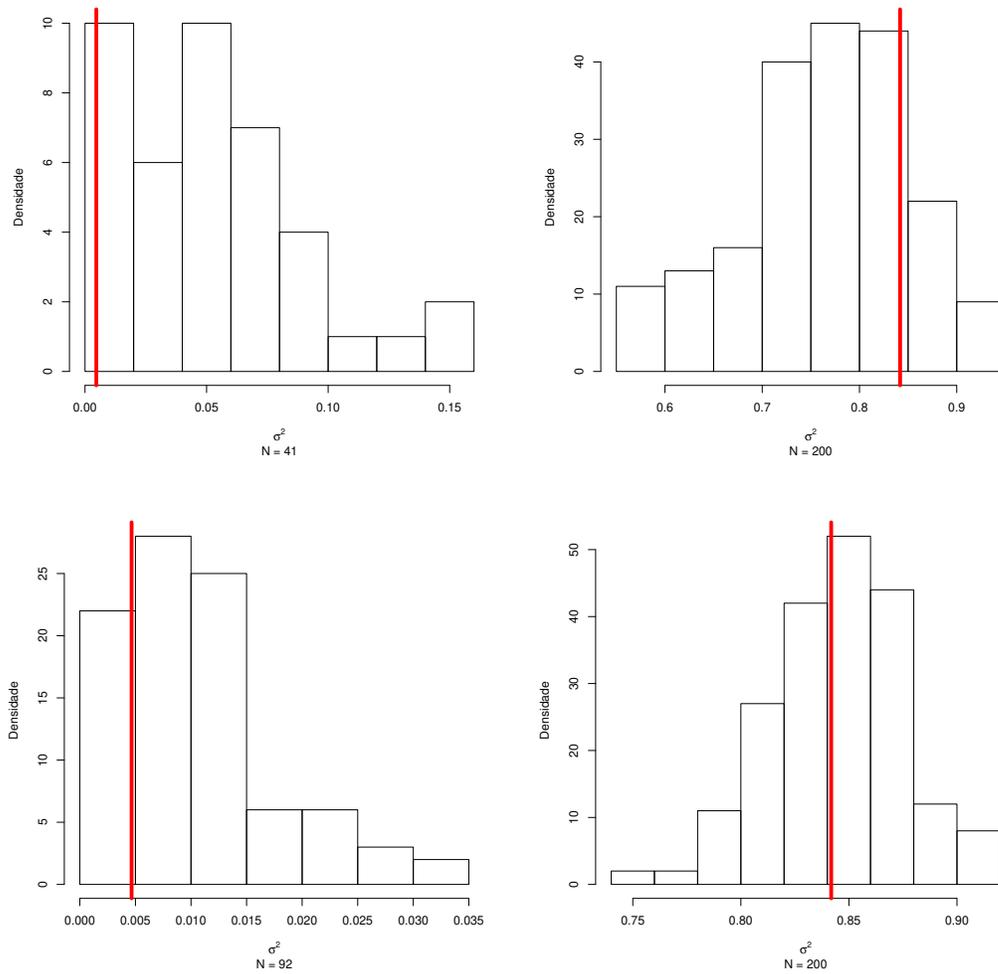


Figura 5.12: Densidade das  $N$  estimativas maiores do que zero de  $\sigma_{\gamma,1}^2$  (coluna da esquerda) e  $\sigma_{\gamma,2}^2$  (coluna da direita) para o Caso 3 com replicação, com 5 transformadores (linha superior) e 50 transformadores (linha inferior). A linha vertical representa o valor verdadeiro.

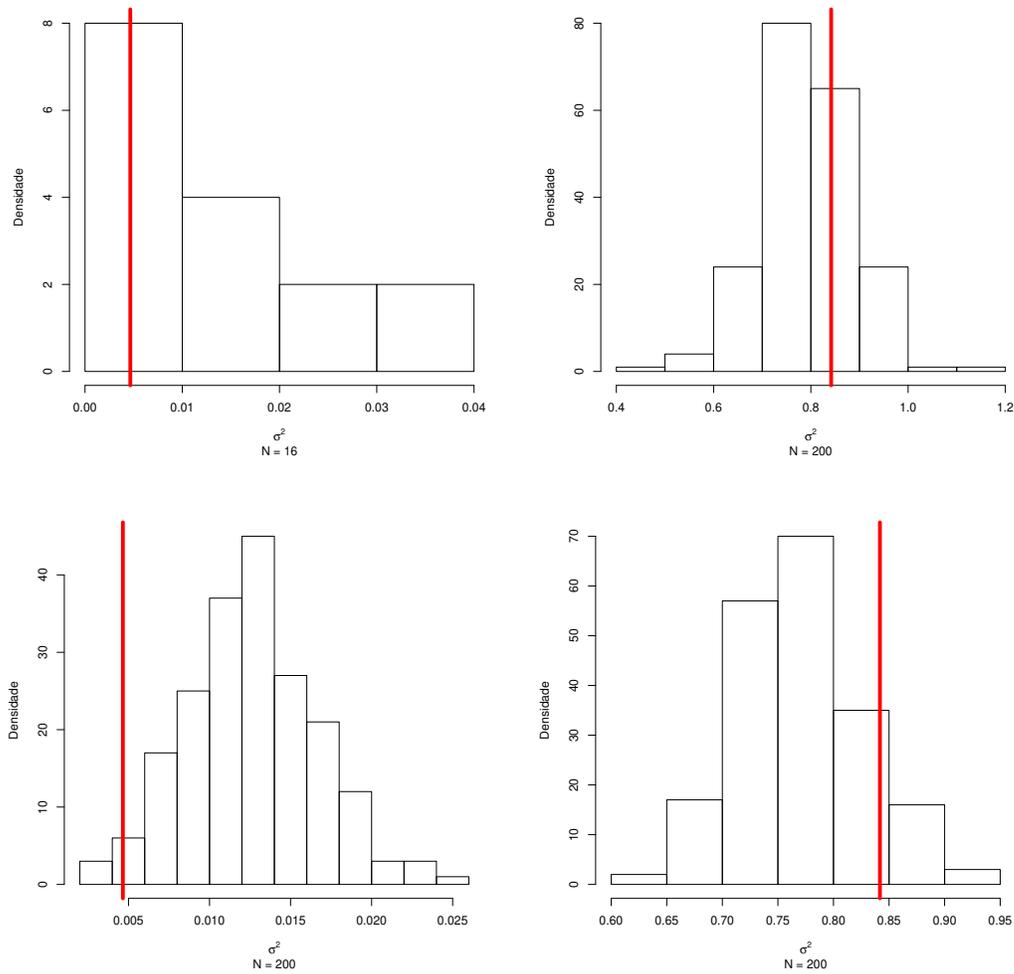


Figura 5.13: Densidade das  $N$  estimativas maiores do que zero de  $\sigma_{\gamma,1}^2$  (coluna da esquerda) e  $\sigma_{\gamma,2}^2$  (coluna da direita) para o Caso 4 com replicação, com 5 transformadores (linha superior) e 50 transformadores (linha inferior). A linha vertical representa o valor verdadeiro.

### 5.3.3 Estimativas dos $M_{ci}$ 's

As Tabelas 5.2 e 5.3 contêm informação relativa às estimativas de  $M_1$  nas simulações feitas com 5 transformadores ( $I = 5$ ) e 50 transformadores ( $I = 50$ ), de tal forma que no caso de 50 transformadores, apenas informações relativas aos 5 primeiros são apresentadas. A Tabela 5.2 refere-se aos transformadores com o número de consumidores balanceado e a Tabela 5.3 refere-se àqueles com mais consumidores residenciais do que comerciais. Para cada um dos transformadores de 1 a 5 o número verdadeiro de consumidores residenciais está em negrito e o número reportado está em entre colchetes. Nas linhas abaixo dos valores de  $\widehat{M}_1$  encontramos a frequência das estimativas de  $M_1$  nos 200 conjuntos de dados gerados para cada um dos quatro casos estudados no respectivo transformador.

Por exemplo, através da Tabela 5.2, podemos ver que, no transformador 2 para o caso onde os  $M$ 's são balanceados, a estimativa mais comum de  $M_1$  no Caso 3 sem replicação para as simulações feitas com  $I = 5$  transformadores é igual a 31, enquanto que para as simulações feitas sob as mesmas condições porém com  $I = 50$  transformadores, o valor estimado de  $M_1$  que mais aparece é igual ao valor verdadeiro, ou seja, 29.

Em geral, podemos ver que conjuntos de dados gerados com replicação levam a menor variabilidade nas estimativas de  $M_1$ , mas nem sempre possuem mais estimativas iguais ao valor verdadeiro. Adicionalmente, notamos que, via de regra, ao adicionarmos mais 45 transformadores nas simulações as estimativas melhoram substancialmente.

Abaixo, temos uma descrição das estimativas para cada transformador, com o número de consumidores gerados de forma balanceada (Casos 1 e 3 apresentados na tabela 5.2) e de forma desbalanceada (Casos 2 e 4 apresentados na tabela 5.3).

#### **Transformadores onde os $M$ 's são balanceados**

O transformador 1 possui número reportado de consumidores residenciais igual ao verdadeiro, como podemos ver pela Tabela 5.2. Podemos perceber que aqui as estimativas de  $M_1$  são melhores nos casos com replicação.

No transformador 2, três consumidores reportaram como sendo residenciais quando na verdade eram comerciais. Aqui, a maioria das estimativas de  $M_1$  nos conjuntos de dados gerados com 5 transformadores é igual 31, enquanto que, ao adicionarmos 45 transformadores, a maior parte das estimativas passa a ser igual a 29 ou 30.

O transformador 3 possui o número de consumidores que reportam ser da classe residencial menor do que o número verdadeiro de consumidores residenciais, fato que na prática tem baixa probabilidade de ocorrer. Note que este não é o comportamento mais comum com a matriz fraude que estamos usando. Nesse caso, para os conjuntos de dados gerados com  $I = 5$ , as estimativas quase não se diferem dos valores reportados, fato este que pode ser explicado pelo

cenário inesperado deste transformador. Por outro lado, quando utilizamos  $I = 50$ , nosso método parece captar esse tipo de comportamento inesperado, já que a estimativa mais comum de  $M_1$  igual passa a ser sempre igual ao valor verdadeiro.

Os transformadores 4 e 5 possuem quatro consumidores comerciais que reportaram como sendo residenciais. Em ambos os transformadores, para os conjuntos de dados gerados com  $I = 5$ , muito poucas estimativas atingem o valor verdadeiro; de fato, a maioria delas é igual a dois ou três números acima do valor verdadeiro. Quando adicionamos 45 transformadores, a maioria dos conjuntos de dados atingem o valor verdadeiro de  $M_1$  ou diferem apenas por uma unidade.

### **Transformadores onde os $M$ 's são desbalanceados**

Como podemos ver pela Tabela 5.3, as estimativas de  $M_1$  para o transformador 1 são bastante razoáveis mesmo para o caso onde  $I = 5$ : apenas um conjunto de dados no Caso 2 sem replicação obteve estimativa menor do que o número reportado, em todos os outros casos mais de 75% dos conjuntos de dados tiveram estimativas iguais ao valor verdadeiro. Aqui, adicionar 45 transformadores na geração dos dados não parece melhorar nossas estimativas, já que menos conjuntos de dados atingiram o valor verdadeiro de  $M_1$ .

Para o transformador 2, o Caso 4 apresenta estimativas melhores do que o Caso 2, tanto para dados gerados com 5 como com 50 transformadores. De fato, a maioria dos conjuntos de dados no Caso 2 possui estimativa igual ao valor reportado de  $M_1$ , enquanto no Caso 4 a maioria é igual ao valor verdadeiro.

No transformador 3, exceto no caso 4 sem replicação onde o número de estimativas iguais ao verdadeiro valor de  $M_1$  é o mesmo para dados gerados com  $I = 5$  ou  $I = 50$ , em todos os outros casos há mais estimativas iguais ao valor verdadeiro quando geramos dados com  $I = 50$  do que com  $I = 5$  transformadores.

No transformador 4, apesar do número reportado de consumidores na classe 1 ser maior do que o número verdadeiro, a maioria das estimativas é igual ao valor reportado para dados gerados com  $I = 5$ . Nos casos onde  $I = 50$ , pelo menos 25% dos conjuntos de dados sempre atingiram o valor verdadeiro. Nota-se que no Caso 4 muitos conjuntos de dados tiveram estimativas menor do que o número verdadeiro, enquanto que no Caso 2 a maioria das estimativas foi igual ao valor reportado.

O transformador 5 possui estimativas muito boas de  $M_1$  quando os dados são gerados com 50 transformadores, de fato, quando há replicação, mais de 70% das estimativas atingem o valor verdadeiro. Quando geramos dados com 5 transformadores, a grande maioria das estimativas é igual ao valor reportado, inclusive, para o Caso 4, todos os conjuntos de dados possuem estimativas iguais ao valor reportado. Esse fato, novamente, pode ter resultado de que esses

transformadores possuem o número reportado de consumidores da classe 1 menor do que o número verdadeiro, o que não é usual.

Transformador	$\widehat{M}_1$													
	$I = 5$							$I = 50$						
# 1	42	43	44	[45]	46	47	48	42	43	44	[45]	46	47	48
1 sem replicação	0	6	34	96	57	6	1	0	11	54	95	38	2	0
3 sem replicação	0	1	14	83	92	10	0	0	5	49	121	24	1	0
1 com replicação	0	0	11	180	9	0	0	0	0	38	160	2	0	0
3 com replicação	0	0	0	145	55	0	0	0	0	13	185	2	0	0
# 2	27	28	<b>29</b>	30	31	[32]	33	27	28	<b>29</b>	30	31	[32]	33
1 sem replicação	0	3	24	61	74	35	3	2	23	51	88	32	4	0
3 sem replicação	0	1	7	37	97	55	3	0	31	75	74	19	1	0
1 com replicação	0	0	0	63	135	2	0	0	0	51	137	12	0	0
3 com replicação	0	0	0	25	161	14	0	0	4	95	101	0	0	0
# 3	49	50	[60]	<b>61</b>	62	63	64	49	50	[60]	<b>61</b>	62	63	64
1 sem replicação	0	7	173	20	0	0	0	0	9	79	84	26	1	0
3 sem replicação	0	2	197	1	0	0	0	0	1	77	106	16	0	0
1 com replicação	0	0	200	0	0	0	0	0	0	83	114	3	0	0
3 com replicação	0	0	200	0	0	0	0	0	0	18	180	2	0	0
# 4	23	<b>24</b>	25	26	27	[28]	29	20	22	23	<b>24</b>	25	26	27
1 sem replicação	1	8	26	85	57	22	1	1	1	13	49	74	53	9
3 sem replicação	0	2	17	68	89	24	0	0	3	12	90	65	25	5
1 com replicação	0	0	2	144	54	0	0	0	0	0	24	137	39	0
3 com replicação	0	0	3	116	80	1	0	0	0	0	59	138	3	0
# 5	11	<b>12</b>	13	14	15	[16]	17	10	11	<b>12</b>	13	14	15	[16]
1 sem replicação	0	0	11	47	106	35	1	1	18	65	78	29	9	0
3 sem replicação	0	0	0	22	82	82	14	1	27	63	81	26	2	0
1 com replicação	0	0	2	93	105	0	0	0	0	45	153	2	0	0
3 com replicação	0	0	1	22	163	14	0	0	0	70	127	3	0	0

Tabela 5.2: Número estimado de consumidores na classe  $c = 1$  (residencial) nos transformadores  $i = 1, \dots, 5$  para os casos onde os  $M$ 's são balanceados. Na primeira coluna da tabela estão as simulações feitas com  $I = 5$  transformadores e na segunda coluna com  $I = 50$  transformadores. O número verdadeiro de consumidores aparece em negrito e o número reportado aparece entre colchetes. Note-se que, no transformador 1 o número reportado de consumidores é igual ao verdadeiro e nos transformadores 4 e 5 para o caso das simulações realizadas com  $I = 50$ , nenhuma estimativa foi igual ao número reportado.

Transformador	$\widehat{M}_1$											
	$I = 5$						$I = 50$					
# 1	64	[65]	<b>66</b>	67	68	69	64	[65]	<b>66</b>	67	68	69
2 sem replicação	1	64	90	31	4	10	4	86	84	6	0	20
4 sem replicação	0	25	159	16	0	0	7	126	65	2	0	0
2 com replicação	0	26	161	13	0	0	0	87	111	0	0	2
4 com replicação	0	2	198	0	0	0	0	146	54	0	0	0
# 2	63	64	<b>65</b>	[66]	67	68	62	63	64	<b>65</b>	[66]	67
2 sem replicação	0	5	63	109	21	2	1	1	5	41	136	15
4 sem replicação	0	0	144	56	0	0	0	0	79	116	5	0
2 com replicação	0	0	7	192	1	0	0	0	0	12	188	0
4 com replicação	0	0	170	30	0	0	0	0	55	145	0	0
# 3	66	67	[68]	<b>69</b>	70	71	66	67	[68]	<b>69</b>	70	71
2 sem replicação	1	15	105	59	20	0	0	3	75	106	16	0
4 sem replicação	0	0	111	89	0	0	0	1	109	89	1	0
2 com replicação	0	0	93	107	0	0	0	0	49	150	1	0
4 com replicação	0	0	129	71	0	0	0	0	99	101	0	0
# 4	60	61	<b>62</b>	[63]	64	65	59	60	61	<b>62</b>	[63]	64
2 sem replicação	0	0	20	133	45	2	1	1	9	67	103	19
4 sem replicação	0	0	53	145	2	0	0	6	100	88	6	0
2 com replicação	0	0	0	194	6	0	0	0	0	52	148	0
4 com replicação	0	0	5	195	0	0	0	0	111	89	0	0
# 5	69	70	[71]	<b>72</b>	73	74	69	70	[71]	<b>72</b>	73	74
2 sem replicação	0	2	147	51	0	0	0	1	59	129	11	0
4 sem replicação	0	0	200	0	0	0	0	0	76	124	0	0
2 com replicação	0	0	178	22	0	0	0	0	8	192	0	0
4 com replicação	0	0	200	0	0	0	0	0	26	174	0	0

Tabela 5.3: Número estimado de consumidores na classe  $c = 1$  (residencial) nos transformadores  $i = 1, \dots, 5$  para os casos onde os  $M$ 's são desbalanceados. Na primeira coluna da tabela estão as simulações feitas com  $I = 5$  transformadores e na segunda coluna com  $I = 50$  transformadores. O número verdadeiro de consumidores aparece em negrito e o número reportado aparece entre colchetes.

## 5.4 Conclusões

Através de nossas análises pudemos perceber que as estimativas obtidas através do método proposto para as curvas típicas de demanda são boas, especialmente quando os  $\alpha_c$ 's estão na mesma escala e o número de transformadores é grande. Pudemos perceber também que em todos os casos, a maioria das estimativas do número verdadeiro de consumidores nas classes é igual ao reportado ou igual ao verdadeiro, ou está entre eles. Em muitos casos, mesmo para dados gerados com apenas 5 transformadores, o valor mais comum foi igual ao verdadeiro ou no mínimo uma unidade deslocada do reportado na direção do verdadeiro. Via de regra, obtivemos estimativas muito boas para a demanda total de energia em cada transformador através da soma ponderada  $\hat{M}_{1i}\hat{\alpha}_1 + \hat{M}_{2i}\hat{\alpha}_2$ ,  $i = 1, \dots, I$ .

Pudemos notar também que, quando o número total de consumidores em cada classe de consumo, somados para todos os transformadores, não apresentam grandes diferenças entre si e a classe residencial apresenta demanda de energia bem menor que a classe comercial, as estimativas para sua curva típica variam mais do que as dos comerciais. Por outro lado, se existem muito mais consumidores residenciais do que comerciais e as curvas típicas de demanda estão na mesma escala então a curva típica dos consumidores residenciais é bastante viesada. Uma possível forma de melhorar a estimação das tipologias quando estamos em casos deste tipo é aumentar o número de transformadores, o que torna as estimativas mais precisas.

A variância do erro da regressão foi sempre muito bem estimada. As estimativas da variância para a perturbação aleatória individual em ambas as classes não foram tão boas, especialmente quando observamos apenas um dia por transformador e poucos transformadores. De fato, para estes casos, muitos conjuntos de dados apresentaram estimativas iguais a zero.

Como era de se esperar, as estimações nos conjuntos de dados com replicação tiveram menos variabilidade. Além disso, menos estimativas da variância relacionada às classes foram iguais a zero. Tanto as estimativas das tipologias como das variâncias melhoram consideravelmente com o aumento do número de transformadores.

Note que, apesar de nosso método se mostrar consistente através de estudos de simulação, seria interessante podermos quantificar a incerteza das nossas estimativas, através de intervalos de confiança e testes de hipótese. Para esta finalidade, em trabalhos futuros, pretendemos desenvolver um ferramental teórico das distribuições assintóticas de nossos estimadores.



# Apêndice A

## Figuras e Tabelas

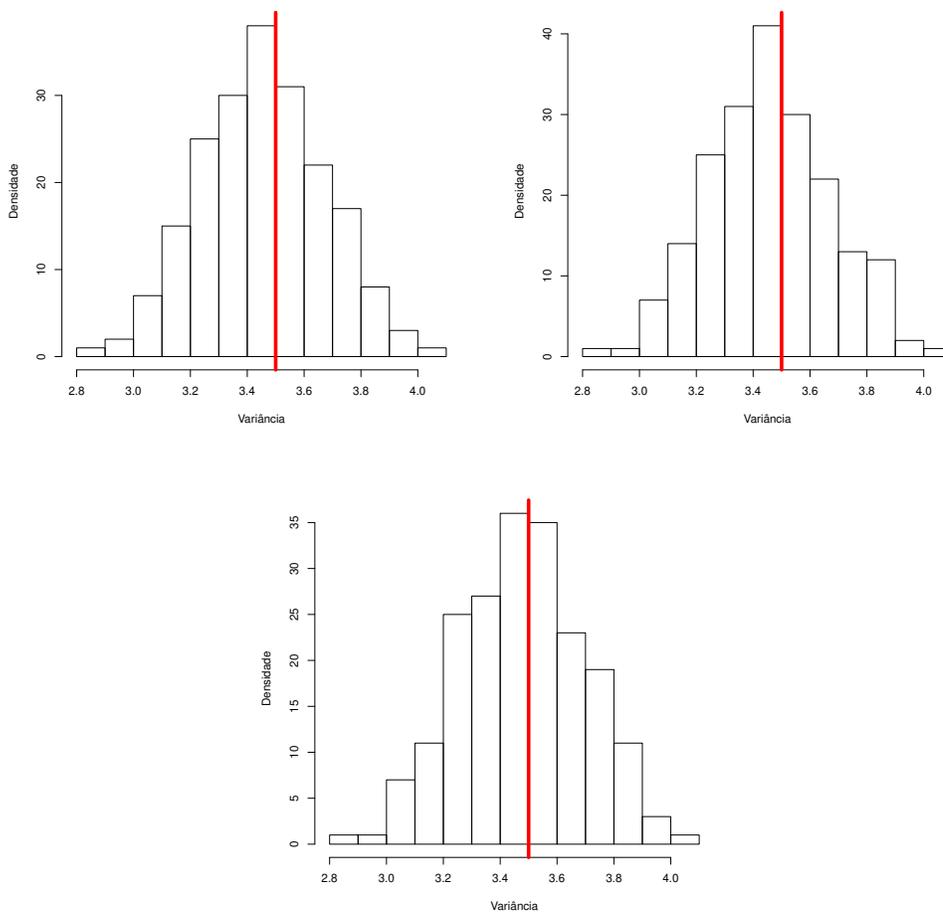


Figura A.1: Densidade das 200 estimativas de  $\sigma^2$  para os Casos 1 e 2 (linha superior) e para o Caso 3 (linha inferior), sem replicação e com 5 transformadores. A linha vertical é o valor verdadeiro.

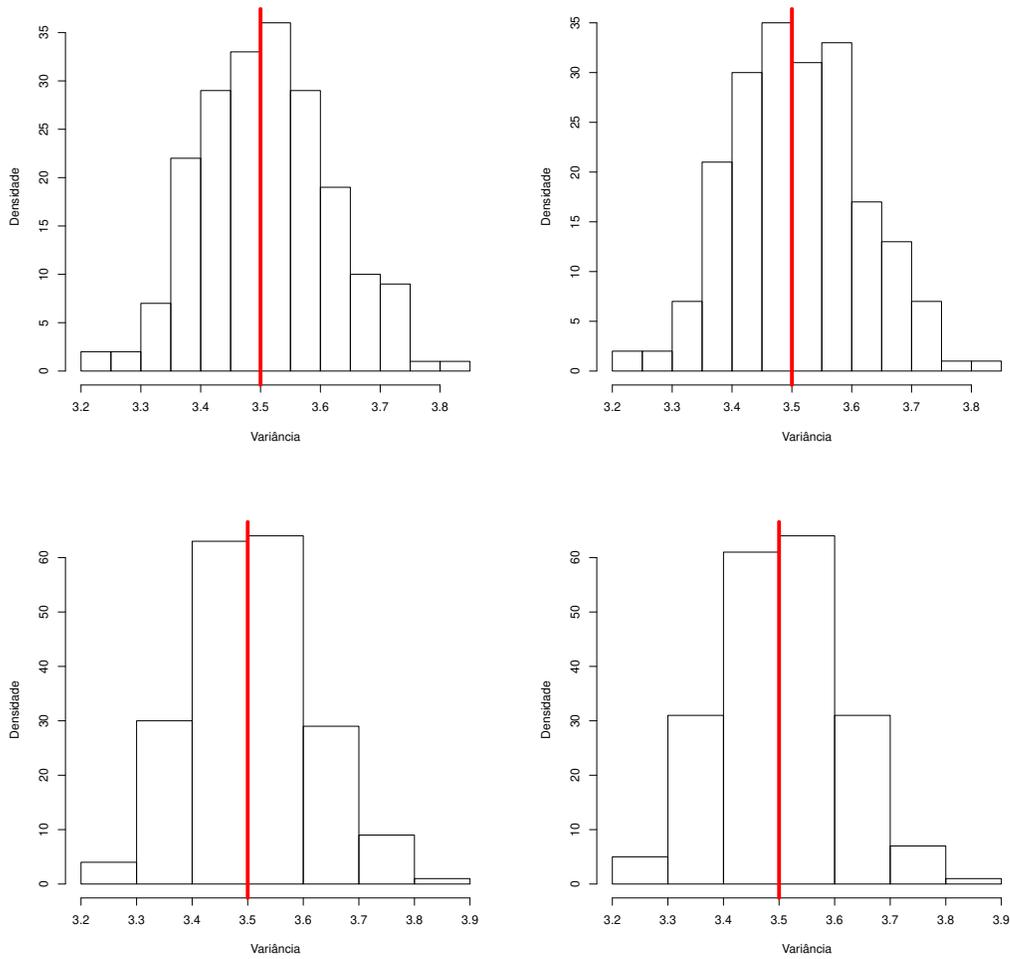


Figura A.2: Densidade das 200 estimativas de  $\sigma^2$  para os Casos 1 e 2 (linha superior) e para os Casos 3 e 4 (linha inferior), com replicação e com 5 transformadores. A linha vertical é o valor verdadeiro.

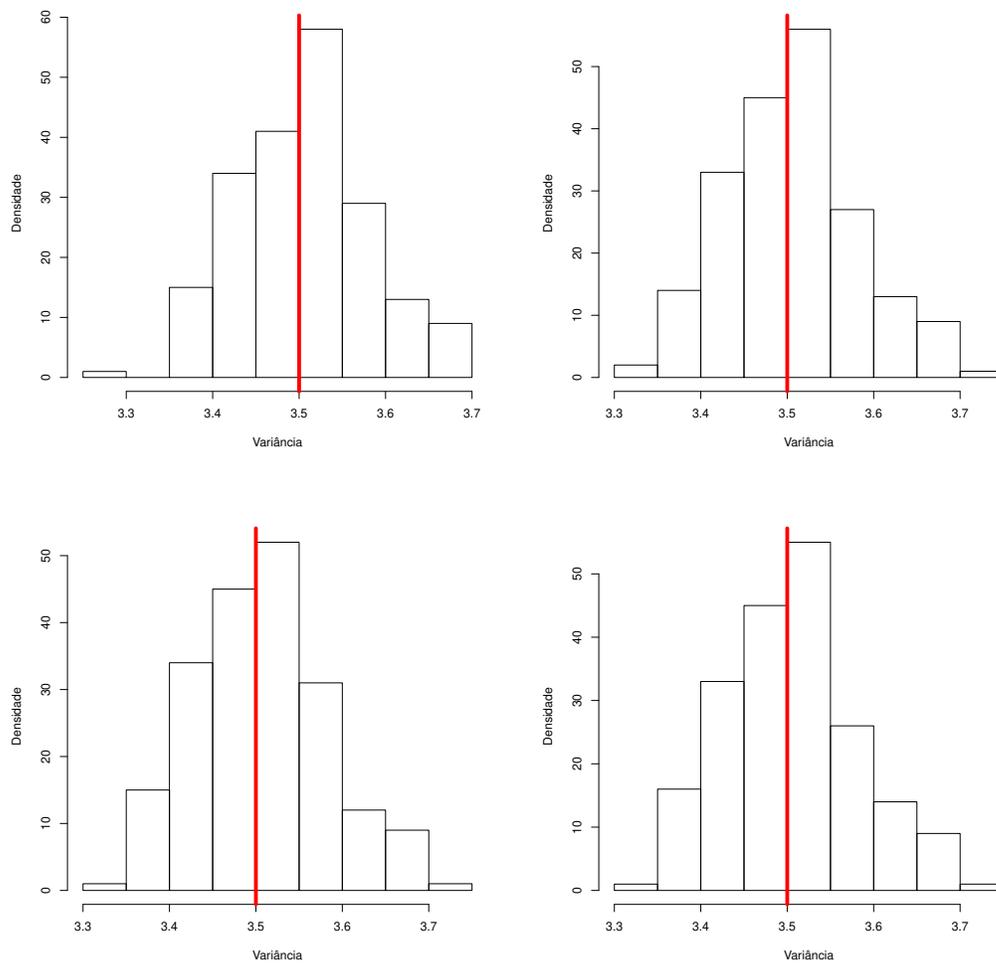


Figura A.3: Densidade das 200 estimativas de  $\sigma^2$  para os Casos 1 e 2 (linha superior) e para os Casos 3 e 4 (linha inferior), sem replicação e com 50 transformadores. A linha vertical é o valor verdadeiro.

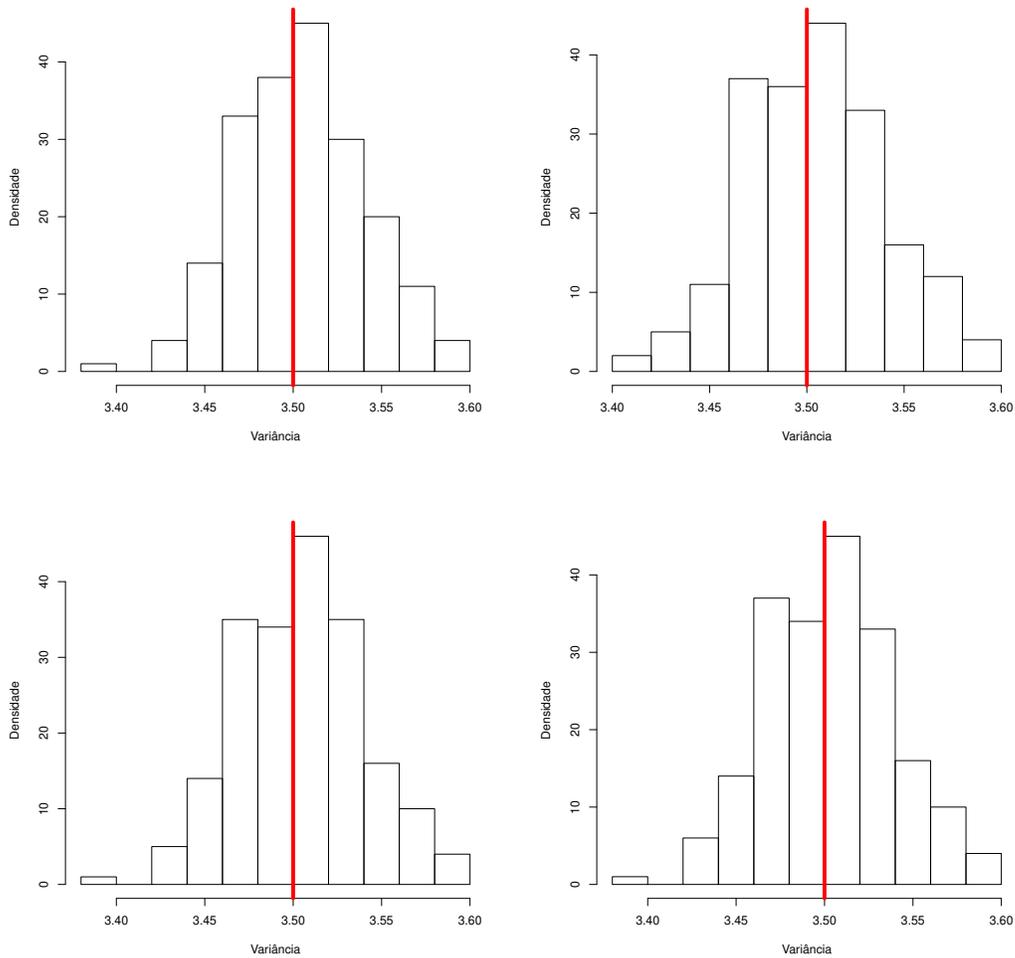


Figura A.4: Densidade das 200 estimativas de  $\sigma^2$  para os Casos 1 e 2 (linha superior) e para os Casos 3 e 4 (linha inferior), com replicação e com 50 transformadores. A linha vertical é o valor verdadeiro.

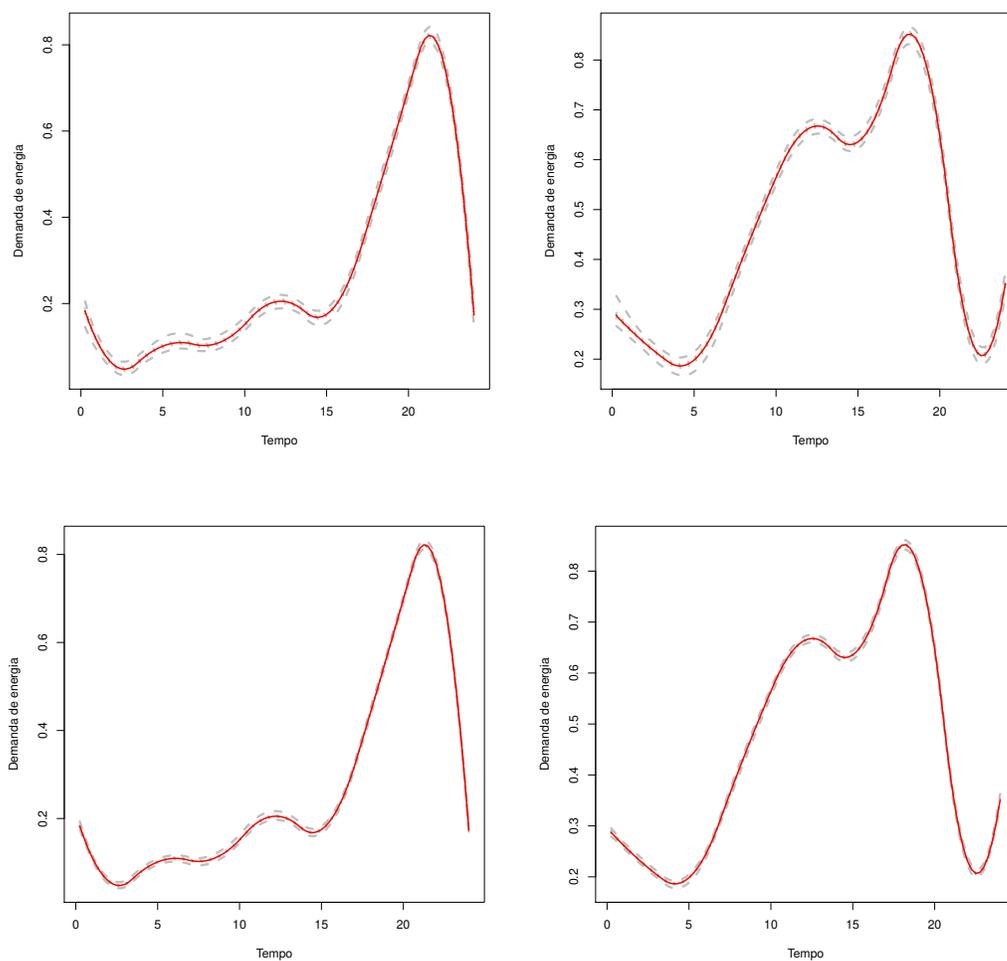


Figura A.5: Mínimo e máximo (curvas tracejadas), mediana e quartis (curvas pontilhadas) das 200 estimativas da demanda de energia para o Caso 1, para as classe  $c = 1$  (coluna da esquerda) e  $c = 2$  (coluna da direita) com 50 transformadores, sem replicação (linha superior) e com replicação (linha inferior). As curvas sólidas representam os valores verdadeiros.

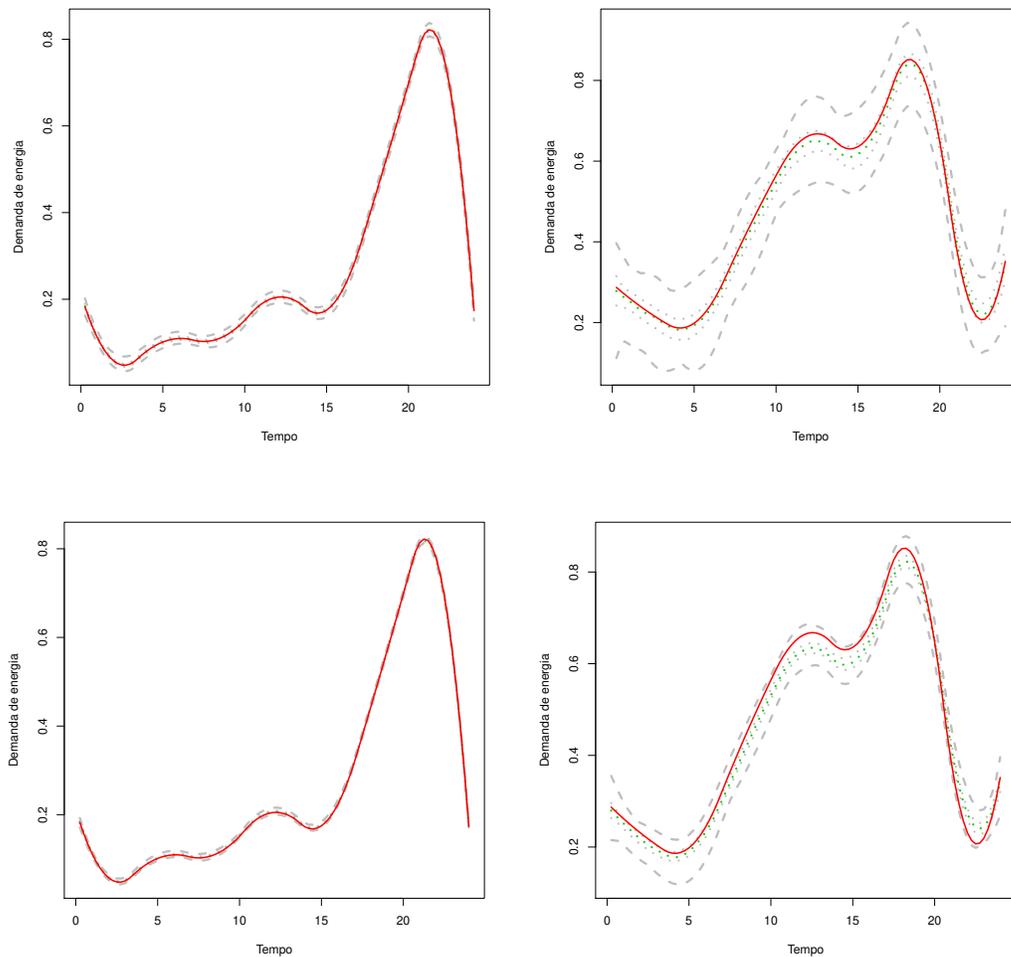


Figura A.6: Mínimo e máximo (curvas tracejadas), mediana e quartis (curvas pontilhadas) das 200 estimativas da demanda de energia para o Caso 2, para as classe  $c = 1$  (coluna da esquerda) e  $c = 2$  (coluna da direita) com 50 transformadores, sem replicação (linha superior) e com replicação (linha inferior). As curvas sólidas representam os valores verdadeiros.

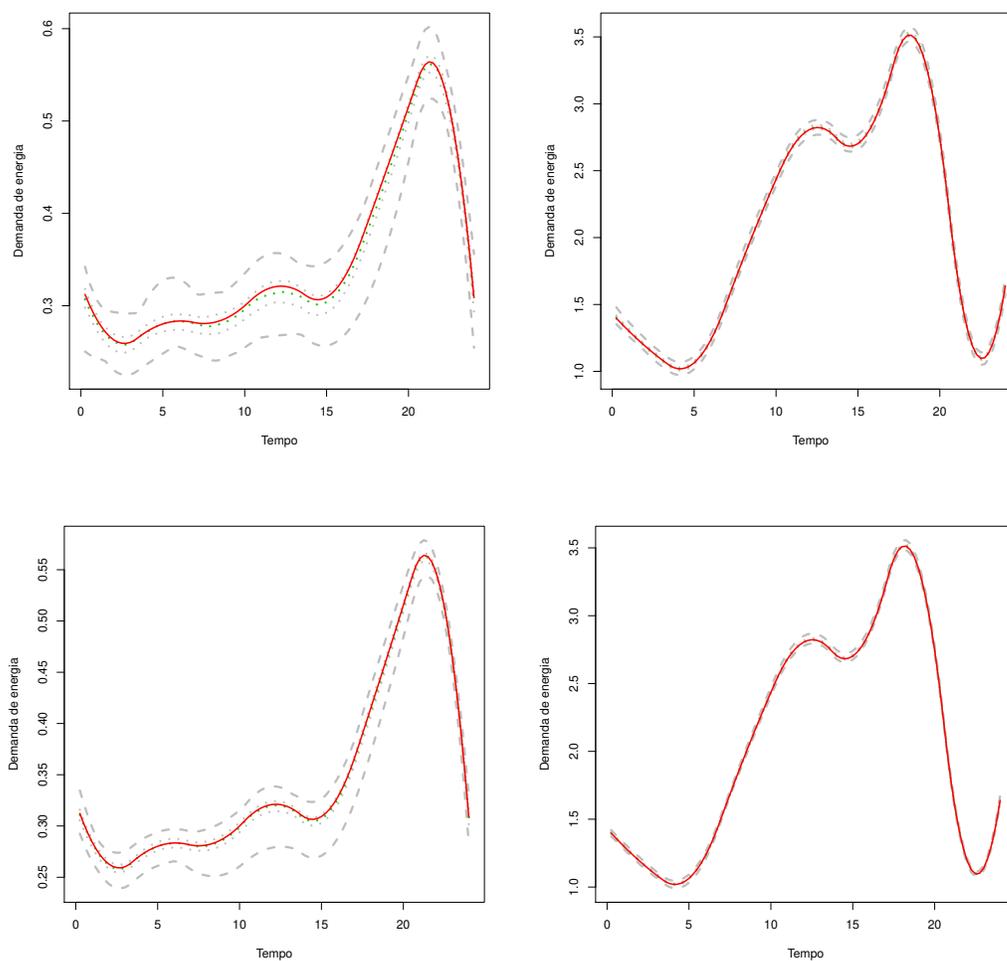


Figura A.7: Mínimo e máximo (curvas tracejadas), mediana e quartis (curvas pontilhadas) das 200 estimativas da demanda de energia para o Caso 3, para as classe  $c = 1$  (coluna da esquerda) e  $c = 2$  (coluna da direita) com 50 transformadores, sem replicação (linha superior) e com replicação (linha inferior). As curvas sólidas representam os valores verdadeiros.

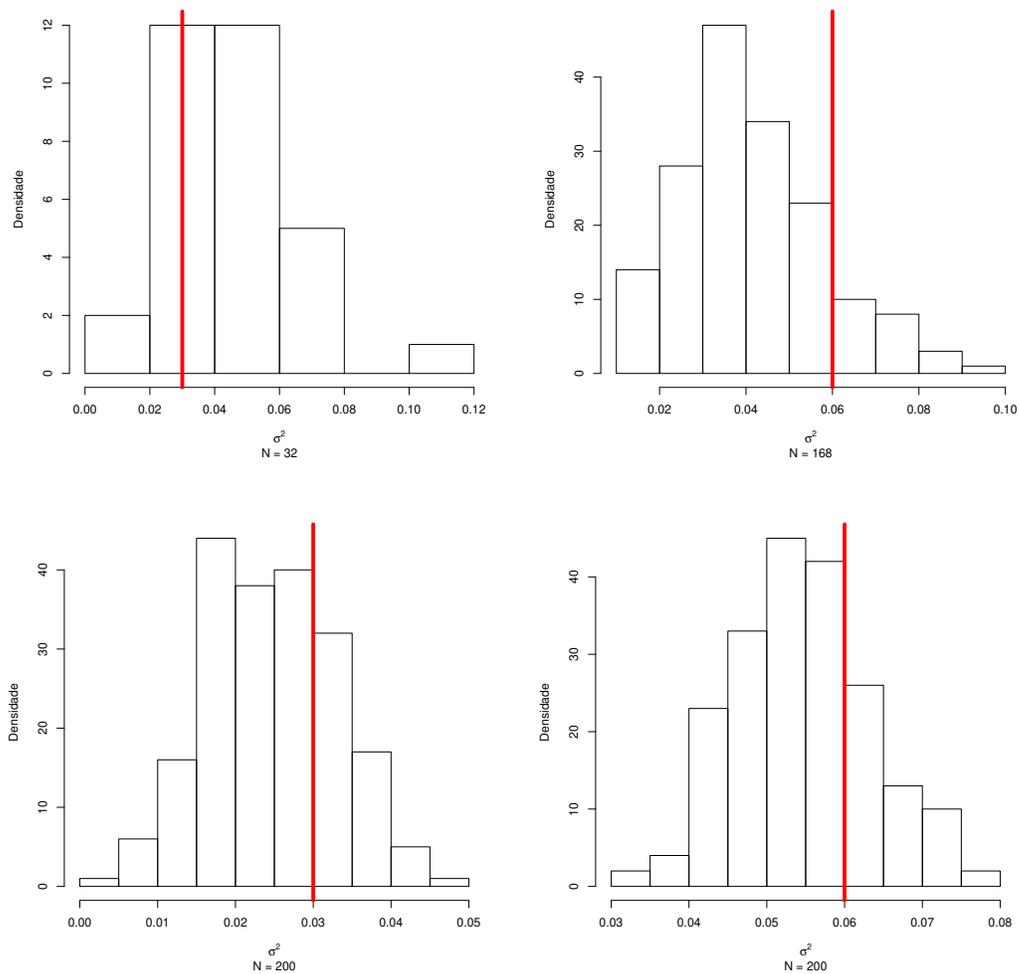


Figura A.8: Densidade das  $N$  estimativas maiores do que zero de  $\sigma_{\gamma,1}^2$  (coluna da esquerda) e  $\sigma_{\gamma,2}^2$  (coluna da direita) para o Caso 1 sem replicação, com 5 transformadores (linha superior) e 50 transformadores (linha inferior). A linha vertical é o valor verdadeiro.

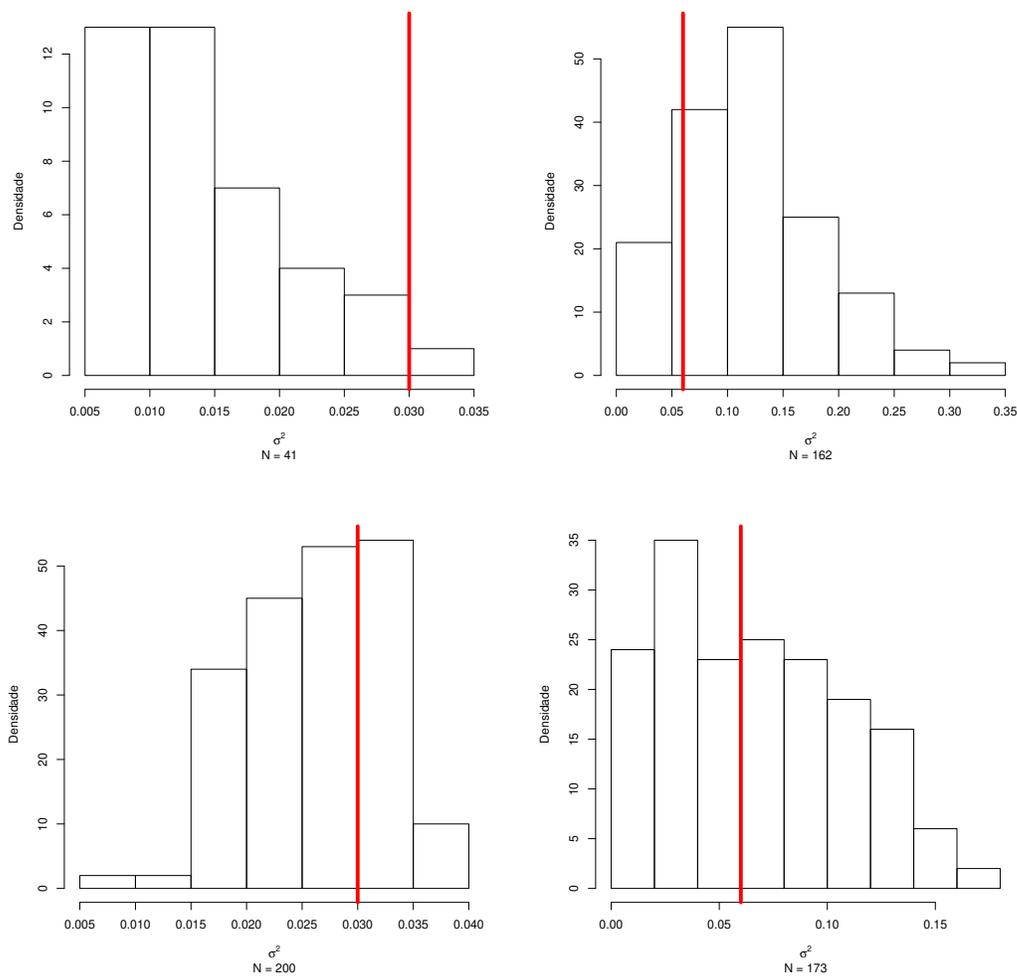


Figura A.9: Densidade das  $N$  estimativas maiores do que zero de  $\sigma_{\gamma,1}^2$  (coluna da esquerda) e  $\sigma_{\gamma,2}^2$  (coluna da direita) para o Caso 2 sem replicação, com 5 transformadores (linha superior) e 50 transformadores (linha inferior). A linha vertical é o valor verdadeiro.

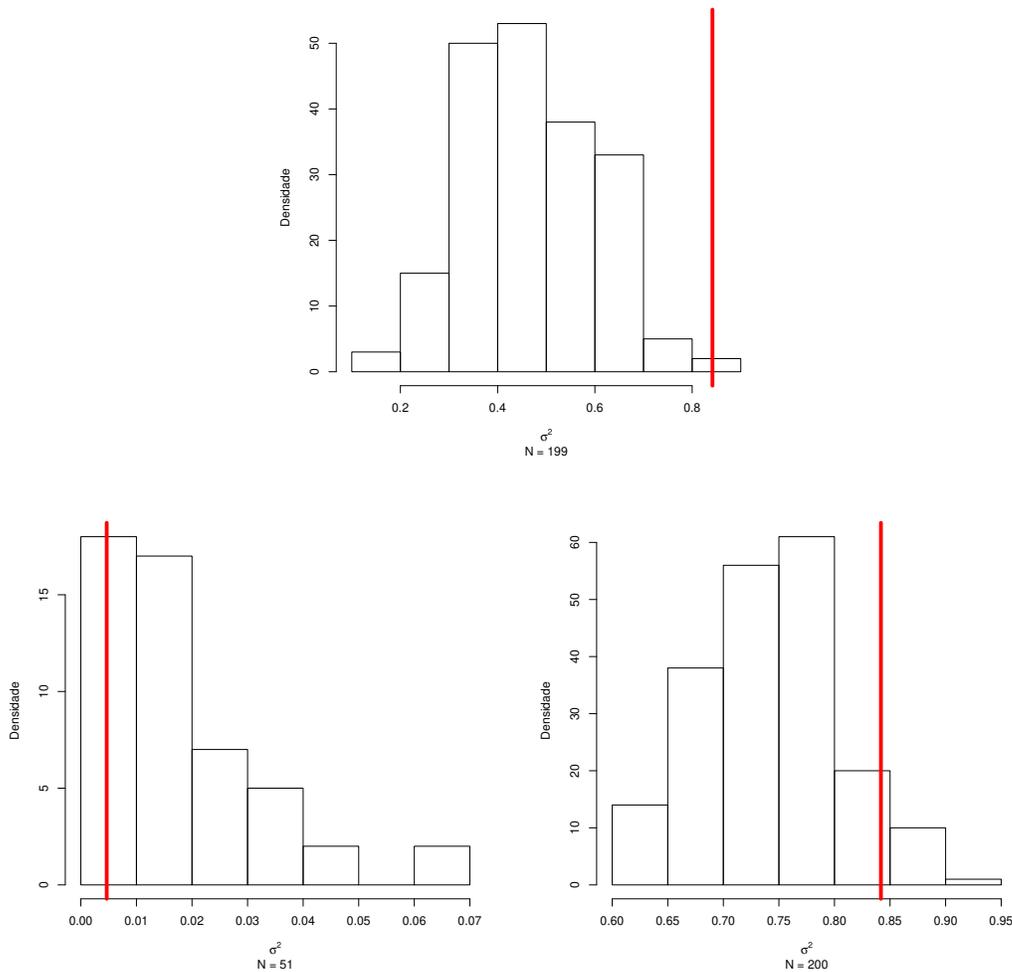


Figura A.10: A figura superior mostra a densidade das  $N$  estimativas maiores do que zero de  $\sigma_{\gamma,2}^2$  para o Caso 3 sem replicação, com 5 transformadores. A figura inferior mostra a densidade das  $N$  estimativas maiores do que zero e de  $\sigma_{\gamma,1}^2$  (figura da esquerda) e  $\sigma_{\gamma,2}^2$  (figura da direita) para o Caso 3 sem replicação, com 50 transformadores. A linha vertical representa o valor verdadeiro. Note que 199 das 200 estimativas de  $\sigma_{\gamma,1}^2$  para o Caso 3 sem replicação, com 5 transformadores foram iguais a zero.

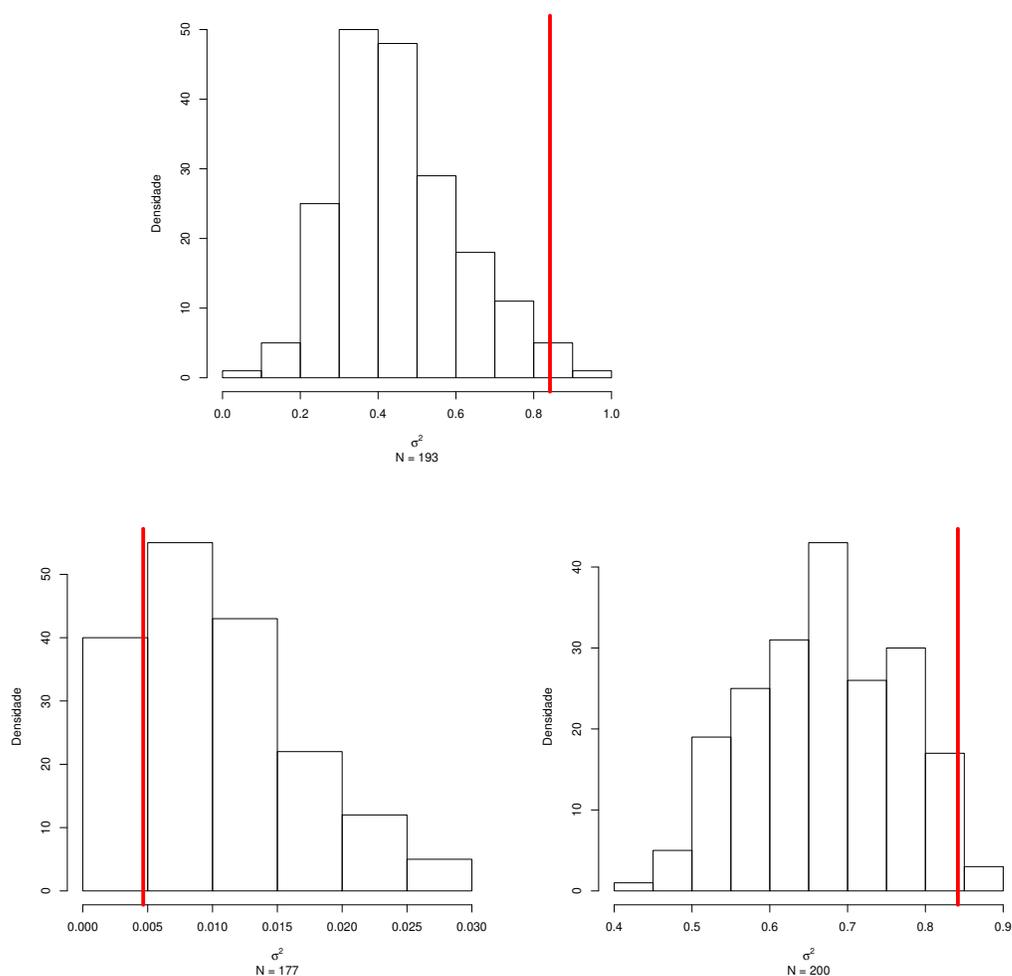


Figura A.11: A figura superior mostra a densidade das  $N$  estimativas maiores do que zero de  $\sigma_{\gamma,2}^2$  para o Caso 4 sem replicação, com 5 transformadores. A figura inferior mostra a densidade das  $N$  estimativas maiores do que zero e de  $\sigma_{\gamma,1}^2$  (figura da esquerda) e  $\sigma_{\gamma,2}^2$  (figura da direita) para o Caso 4 sem replicação, com 50 transformadores. A linha vertical representa o valor verdadeiro. Note que 191 das 200 estimativas de  $\sigma_{\gamma,1}^2$  para o Caso 4 sem replicação, com 5 transformadores foram iguais a zero.

		$\widehat{M}_1$													
		#6							#7						
		42	43	44	[45]	46	47	48	49	6	7	8	<b>9</b>	10	[11]
1 sem replicação		1	9	64	93	25	8	0	0	1	26	56	83	32	2
3 sem replicação		0	4	54	110	30	2	0	0	3	14	55	81	44	3
1 com replicação		0	0	32	167	1	0	0	0	0	0	39	155	6	0
3 com replicação		0	0	20	177	3	0	0	0	0	2	31	157	10	0
		#8							#9						
		19	20	21	<b>22</b>	23	24	25	[26]	40	[41]	42	<b>43</b>	44	45
1 sem replicação		1	2	17	45	76	43	15	1	12	77	90	18	3	0
3 sem replicação		0	3	19	67	73	36	2	0	2	23	102	59	13	1
1 com replicação		0	0	0	25	149	26	0	0	0	84	114	2	0	0
3 com replicação		0	0	0	50	145	5	0	0	0	4	156	40	0	0
		#10							#11						
		43	44	45	[46]	<b>47</b>	48	49	50	35	36	37	[38]	39	40
1 sem replicação		0	3	23	98	61	15	0	0	0	16	85	80	19	0
3 sem replicação		0	0	5	87	88	20	0	0	1	7	62	107	22	1
1 com replicação		0	0	0	158	42	0	0	0	0	0	80	120	0	0
3 com replicação		0	0	0	71	127	2	0	0	0	0	45	153	2	0
		#12							#13						
		33	34	35	36	[37]	<b>38</b>	39	40	38	39	<b>40</b>	41	[42]	43
1 sem replicação		0	0	6	51	88	50	5	0	3	26	66	83	19	3
3 sem replicação		0	0	0	19	99	66	15	1	2	23	91	72	10	2
1 com replicação		0	0	0	13	172	15	0	0	0	1	69	127	3	0
3 com replicação		0	0	0	0	132	68	0	0	0	0	124	76	0	0
		#14							#15						
		14	15	16	17	<b>18</b>	19	20	21	[61]	62	63	64	<b>65</b>	66
1 sem replicação		4	0	0	19	52	79	41	5	1	23	58	0	0	118
3 sem replicação		1	0	4	22	59	86	27	1	0	0	13	44	112	31
1 com replicação		0	0	0	0	38	149	13	0	0	45	153	2	0	0
3 com replicação		0	0	0	0	66	131	3	0	0	0	23	154	23	0

Tabela A.1: Número estimado de consumidores na classe  $c = 1$  (residencial) nos transformadores  $i = 6, \dots, 15$  para dados gerados com  $I = 50$  nos casos onde os  $M$ 's são balanceados. O número verdadeiro de consumidores aparece em negrito e o número reportado está entre colchetes.

$\widehat{M}_1$														
#16								#17						
	59	60	61	62	<b>[63]</b>	64	65	7	<b>8</b>	9	10	11	12	<b>[13]</b>
1 sem replicação	0	0	0	19	113	66	0	71	24	46	43	16	0	0
3 sem replicação	0	0	1	25	154	20	0	52	51	56	39	2	0	0
1 com replicação	0	0	0	0	155	45	0	30	49	94	27	0	0	0
3 com replicação	0	0	0	4	194	2	0	19	76	94	11	0	0	0
#18							#19							
	50	51	52	<b>[53]</b>	54	55	56	65	<b>[66]</b>	67	<b>68</b>	69	70	71
1 sem replicação	1	2	33	106	55	3	0	1	26	88	68	17	0	0
3 sem replicação	0	1	49	126	24	0	0	0	4	76	111	9	0	0
1 com replicação	0	0	6	184	10	0	0	0	1	132	67	0	0	0
3 com replicação	0	0	6	189	5	0	0	0	0	35	165	0	0	0
#20							#21							
	17	18	19	20	<b>[21]</b>	22	23	13	14	15	16	<b>[17]</b>	18	19
1 sem replicação	1	6	41	99	46	6	1	1	12	44	101	40	2	0
3 sem replicação	0	2	25	87	73	11	2	0	2	26	101	64	6	1
1 com replicação	0	0	8	180	12	0	0	0	0	23	167	10	0	0
3 com replicação	0	0	1	128	71	0	0	0	0	2	154	44	0	0
#22							#23							
	-1	0	<b>1</b>	2	<b>[3]</b>	4	5	7	8	<b>9</b>	10	11	12	<b>[13]</b>
1 sem replicação	181	0	19	0	0	0	0	0	29	59	75	33	4	0
3 sem replicação	141	16	39	4	0	0	0	0	27	70	73	26	2	2
1 com replicação	200	0	0	0	0	0	0	0	0	85	111	4	0	0
3 com replicação	195	4	1	0	0	0	0	0	1	85	113	1	0	0
#24							#25							
	3	4	5	6	<b>7</b>	8	<b>[9]</b>	38	39	40	41	<b>42</b>	<b>[43]</b>	44
1 sem replicação	0	0	80	7	87	24	2	0	0	7	33	88	64	8
3 sem replicação	0	0	42	14	102	39	3	0	0	3	40	104	45	8
1 com replicação	0	0	51	2	145	2	0	0	0	0	4	156	39	1
3 com replicação	0	0	15	24	155	6	0	0	0	0	8	170	22	0
#26							#27							
	16	17	18	19	<b>[20]</b>	21	22	1	2	<b>3</b>	4	5	6	<b>[7]</b>
1 sem replicação	1	13	52	77	50	6	1	0	0	50	133	15	2	0
3 sem replicação	0	5	11	84	86	14	0	0	0	56	119	20	5	0
1 com replicação	0	0	13	167	20	0	0	0	0	10	189	1	0	0
3 com replicação	0	0	2	134	64	0	0	0	0	23	177	0	0	0

Tabela A.2: Número estimado de consumidores na classe  $c = 1$  (residencial) nos transformadores  $i = 16, \dots, 27$  para dados gerados com  $I = 50$  nos casos onde os  $M$ 's são balanceados. O número verdadeiro de consumidores aparece em negrito e o número reportado está entre colchetes.

		$\widehat{M}_1$													
		#28						#29							
		57	58	59	<b>[60]</b>	61	62	41	42	43	44	<b>45</b>	46	<b>[47]</b>	48
1 sem replicação		0	4	27	108	53	8	0	0	2	16	60	91	28	3
3 sem replicação		0	0	47	136	17	0	0	0	1	30	91	70	8	0
1 com replicação		0	0	0	178	22	0	0	0	0	0	31	162	7	0
3 com replicação		0	0	4	191	5	0	0	0	0	1	126	73	0	0
		#30						#31							
		67	68	<b>[69]</b>	70	71	72	39	40	41	<b>42</b>	43	44	45	<b>[46]</b>
1 sem replicação		0	12	112	49	0	27	0	2	8	36	82	54	15	3
3 sem replicação		0	41	150	9	0	0	0	1	13	64	94	25	3	0
1 com replicação		0	0	124	72	0	4	0	0	0	1	75	121	3	0
3 com replicação		0	20	180	0	0	0	0	0	0	43	150	7	0	0
		#32						#33							
		55	56	<b>[57]</b>	58	59	60	24	25	26	27	<b>28</b>	29	30	<b>[31]</b>
1 sem replicação		4	37	108	45	4	2	0	0	7	23	61	76	31	2
3 sem replicação		1	37	130	29	3	0	0	0	4	24	58	83	29	2
1 com replicação		0	4	185	11	0	0	0	0	0	1	54	141	4	0
3 com replicação		0	6	184	10	0	0	0	0	0	2	108	89	1	0
		#34						#35							
		25	26	27	<b>[28]</b>	29	30	7	8	9	10	11	<b>12</b>	13	<b>[14]</b>
1 sem replicação		1	35	85	71	6	2	16	0	0	6	53	86	36	3
3 sem replicação		2	9	67	108	13	1	4	0	3	13	37	92	47	4
1 com replicação		0	5	149	46	0	0	1	0	0	0	67	123	9	0
3 com replicação		0	0	96	103	1	0	0	0	0	1	36	148	15	0
		#36						#37							
		31	32	33	<b>34</b>	35	<b>[36]</b>	14	15	16	17	18	<b>19</b>	<b>[20]</b>	21
1 sem replicação		0	4	27	77	71	21	0	1	1	30	72	68	26	2
3 sem replicação		0	6	32	90	64	8	0	0	1	17	59	95	27	1
1 com replicação		0	0	1	106	93	0	0	0	0	1	94	105	0	0
3 com replicação		0	0	1	145	53	1	0	0	0	0	52	147	1	0
		#38						#39							
		4	5	6	<b>7</b>	8	<b>[9]</b>	18	19	20	21	22	<b>23</b>	<b>[24]</b>	25
1 sem replicação		0	4	27	77	71	21	0	0	4	21	47	97	30	1
3 sem replicação		0	6	32	90	64	8	0	0	0	11	47	101	40	1
1 com replicação		0	0	1	106	93	0	0	0	0	0	64	135	1	0
3 com replicação		0	0	1	145	53	1	0	0	0	0	45	152	3	0

Tabela A.3: Número estimado de consumidores na classe  $c = 1$  (residencial) nos transformadores  $i = 28, \dots, 39$  para dados gerados com  $I = 50$  nos casos onde os  $M$ 's são balanceados. O número verdadeiro de consumidores aparece em negrito e o número reportado está entre colchetes.

$\widehat{M}_1$														
#40														
	21	22	23	24	<b>25</b>	26	[27]	#41						
	21	22	23	24	<b>25</b>	26	[27]	63	64	65	66	[67]	68	69
2 sem replicação	0	1	8	35	77	67	12	0	0	0	14	108	74	4
4 sem replicação	0	0	11	41	86	57	5	0	0	0	49	147	4	0
2 com replicação	0	0	0	9	156	35	0	0	0	0	0	162	38	0
4 com replicação	0	0	0	6	158	36	0	0	0	0	15	185	0	0
#42							#43							
	13	14	<b>15</b>	16	17	18	[19]	41	42	43	<b>44</b>	45	46	[47]
2 sem replicação	4	18	61	73	37	7	0	0	0	10	31	83	65	11
4 sem replicação	2	24	70	72	31	1	0	0	0	19	71	97	13	0
2 com replicação	0	0	39	148	13	0	0	0	0	0	7	127	65	1
4 com replicação	0	1	83	115	1	0	0	0	0	2	89	109	0	0
#44							#45							
	10	13	14	<b>15</b>	16	17	[18]	18	19	20	<b>21</b>	22	23	[24]
2 sem replicação	19	1	28	44	71	33	4	4	3	25	72	61	31	4
4 sem replicação	0	7	38	81	54	20	0	0	0	33	80	64	21	2
2 com replicação	0	0	13	91	96	0	0	0	0	1	96	102	1	0
4 com replicação	0	0	12	115	72	1	0	0	0	4	112	84	0	0
#46							#47							
	55	56	57	58	<b>[59]</b>	60	61	12	<b>13</b>	14	15	16	17	[18]
2 sem replicação	0	0	1	32	99	59	9	46	25	69	41	17	2	0
4 sem replicação	0	0	0	38	132	30	0	40	35	77	40	7	1	0
2 com replicação	0	0	0	1	175	24	0	11	9	148	32	0	0	0
4 com replicação	0	0	0	4	186	10	0	8	31	154	7	0	0	0
#48							#49							
	39	40	[41]	42	<b>43</b>	44	45	27	28	29	30	[31]	32	33
2 sem replicação	0	10	73	90	26	1	0	0	2	31	86	68	12	1
4 sem replicação	0	2	25	100	64	8	1	0	1	18	75	91	13	2
2 com replicação	0	0	94	106	0	0	0	0	0	5	122	73	0	0
4 com replicação	0	0	1	166	33	0	0	0	0	0	76	124	0	0
#50														
	56	57	58	59	<b>[60]</b>	61	62							
2 sem replicação	0	0	0	22	120	53	5							
4 sem replicação	0	0	1	34	134	31	0							
2 com replicação	0	0	0	3	175	22	0							
4 com replicação	0	0	0	1	190	9	0							

Tabela A.4: Número estimado de consumidores na classe  $c = 1$  (residencial) nos transformadores  $i = 40, \dots, 50$  para dados gerados com  $I = 50$  nos casos onde os  $M$ 's são balanceados. O número verdadeiro de consumidores aparece em negrito e o número reportado está entre colchetes.

		$\widehat{M}_1$											
		#6						#7					
		65	66	67	[68]	<b>69</b>	70	58	[59]	60	<b>61</b>	62	63
2 sem replicação	0	0	3	75	94	28	7	63	87	33	0	10	
4 sem replicação	0	0	1	102	95	2	1	49	118	32	0	0	
2 com replicação	0	0	0	35	161	4	0	71	125	4	0	0	
4 com replicação	0	0	0	91	109	0	0	36	152	12	0	0	
		#8						#9					
		62	63	[64]	65	66	67	64	[65]	66	67	68	<b>69</b>
2 sem replicação	3	24	125	43	4	1	0	2	13	92	93	0	
4 sem replicação	8	105	84	3	0	0	0	0	0	4	196	0	
2 com replicação	0	0	194	6	0	0	0	0	1	140	59	0	
4 com replicação	0	118	82	0	0	0	0	0	0	0	200	0	
		#10						#11					
		67	68	[69]	<b>70</b>	71	72	64	65	[66]	67	<b>68</b>	69
2 sem replicação	0	1	67	110	0	22	0	1	29	103	61	6	
4 sem replicação	0	0	101	99	0	0	0	0	7	122	68	3	
2 com replicação	0	0	29	169	0	2	0	0	4	181	15	0	
4 com replicação	0	0	74	126	0	0	0	0	0	148	52	0	
		#12						#13					
		64	65	[66]	67	<b>68</b>	69	65	66	[67]	<b>68</b>	69	70
2 sem replicação	0	0	31	114	15	40	0	1	72	97	29	1	
4 sem replicação	0	0	8	126	65	1	0	0	114	86	0	0	
2 com replicação	0	0	3	179	6	12	0	0	42	154	4	0	
4 com replicação	0	0	1	154	45	0	0	0	115	85	0	0	
		#14						#15					
		65	66	67	[68]	69	70	68	[69]	70	71	72	<b>73</b>
2 sem replicação	0	0	11	120	65	4	0	0	12	0	188	0	
4 sem replicação	0	2	75	113	10	0	0	0	0	0	200	0	
2 com replicação	0	0	0	154	46	0	0	0	1	0	200	0	
4 com replicação	0	0	46	154	0	0	0	0	0	0	200	0	

Tabela A.5: Número estimado de consumidores na classe  $c = 1$  (residencial) nos transformadores  $i = 6, \dots, 15$  para dados gerados com  $I = 50$  nos casos onde os  $M$ 's são desbalanceados. O número verdadeiro de consumidores aparece em negrito e o número reportado está entre colchetes.

$\widehat{M}_1$												
#16						#17						
	69	70	71	72	<b>[73]</b>	74	58	59	60	<b>[61]</b>	62	63
2 sem replicação	0	0	0	0	114	86	0	4	33	129	31	3
4 sem replicação	0	0	0	26	174	0	1	7	115	76	1	0
2 com replicação	0	0	0	0	94	106	0	0	7	193	0	0
4 com replicação	0	0	0	1	199	0	0	0	149	51	0	0
#18						#19						
	67	68	69	<b>[70]</b>	<b>71</b>	72	69	70	<b>[71]</b>	72	73	<b>74</b>
2 sem replicação	0	0	2	58	126	14	0	0	4	37	159	0
4 sem replicação	0	0	0	84	115	1	0	0	0	0	200	0
2 com replicação	0	0	0	10	190	0	0	0	0	6	194	0
4 com replicação	0	0	0	50	150	0	0	0	0	0	200	0
#20						#21						
	60	61	<b>[62]</b>	63	<b>64</b>	65	58	<b>[59]</b>	60	61	62	<b>63</b>
2 sem replicação	0	3	60	81	50	6	0	3	49	89	59	0
4 sem replicação	0	0	21	128	50	1	0	0	3	55	142	0
2 com replicação	0	0	26	162	12	0	0	0	59	137	4	0
4 com replicação	0	0	7	172	21	0	0	0	0	57	143	0
#22						#23						
	57	58	59	<b>[60]</b>	61	62	59	<b>[60]</b>	61	<b>62</b>	63	64
2 sem replicação	0	5	37	130	27	1	3	48	104	41	0	4
4 sem replicação	0	18	119	60	3	0	0	46	121	31	2	0
2 com replicação	0	0	16	182	2	0	0	41	158	1	0	0
4 com replicação	0	1	155	44	0	0	0	23	168	9	0	0
#24						#25						
	56	<b>[57]</b>	58	59	60	<b>61</b>	60	61	62	<b>[63]</b>	64	65
2 sem replicação	0	10	70	19	101	0	0	3	34	121	34	8
4 sem replicação	0	0	6	46	147	1	1	9	111	78	1	0
2 com replicação	0	1	117	18	64	0	0	0	1	193	6	0
4 com replicação	0	0	0	50	150	0	0	0	133	67	0	0
#26						#27						
	69	70	71	<b>[72]</b>	73	74	66	67	<b>[68]</b>	<b>69</b>	70	71
2 sem replicação	0	0	1	102	96	1	0	3	76	85	0	36
4 sem replicação	0	0	33	166	1	0	0	1	108	91	0	0
2 com replicação	0	0	0	62	138	0	0	0	20	180	0	0
4 com replicação	0	0	1	199	0	0	0	0	82	118	0	0

Tabela A.6: Número estimado de consumidores na classe  $c = 1$  (residencial) nos transformadores  $i = 16, \dots, 27$  para dados gerados com  $I = 50$  nos casos onde os  $M$ 's são desbalanceados. O número verdadeiro de consumidores aparece em negrito e o número reportado está entre colchetes.

		$\widehat{M}_1$											
		#28						#29					
		69	70	71	72	[73]	<b>74</b>	65	66	67	[ <b>68</b> ]	69	70
2 sem replicação		0	0	0	47	4	149	0	1	14	129	53	3
4 sem replicação		0	0	0	51	148	1	0	2	80	110	8	0
2 com replicação		0	0	0	2	0	198	0	0	0	144	56	0
4 com replicação		0	0	0	5	195	0	0	0	46	154	0	0
		#30						#31					
		67	68	69	[70]	71	<b>72</b>	63	64	[65]	<b>66</b>	67	68
2 sem replicação		0	0	0	28	111	61	0	8	83	82	23	4
4 sem replicação		0	0	0	0	101	99	0	7	126	65	2	0
2 com replicação		0	0	0	0	159	41	0	0	81	118	1	0
4 com replicação		0	0	0	0	60	140	0	0	142	58	0	0
		#32						#33					
		61	62	63	[64]	<b>65</b>	66	59	[60]	61	<b>62</b>	63	64
2 sem replicação		0	0	8	77	90	25	6	63	81	37	11	2
4 sem replicação		0	0	13	119	66	2	2	31	132	33	2	0
2 com replicação		0	0	0	101	99	0	0	57	140	3	0	0
4 com replicação		0	0	0	162	38	0	0	25	163	12	0	0
		#34						#35					
		64	65	[66]	<b>67</b>	68	69	61	62	63	[ <b>64</b> ]	65	66
2 sem replicação		0	4	79	92	2	23	0	0	20	135	40	5
4 sem replicação		0	2	128	70	0	0	0	8	105	86	1	0
2 com replicação		0	0	69	131	0	0	0	0	0	186	14	0
4 com replicação		0	0	135	65	0	0	0	0	112	88	0	0
		#36						#37					
		63	[64]	<b>65</b>	66	67	68	70	71	72	73	[ <b>74</b> ]	75
2 sem replicação		9	80	90	16	0	5	0	0	0	0	95	105
4 sem replicação		7	134	58	1	0	0	0	0	0	9	191	0
2 com replicação		0	90	109	1	0	0	0	0	0	0	61	139
4 com replicação		0	154	46	0	0	0	0	0	0	0	200	0
		#38						#39					
		61	62	<b>63</b>	[64]	65	69	67	68	[ <b>69</b> ]	70	71	72
2 sem replicação		0	14	53	121	11	1	0	5	126	58	3	8
4 sem replicação		10	95	93	2	0	0	0	73	127	0	0	0
2 com replicação		0	0	27	173	0	0	0	0	146	54	0	0
4 com replicação		0	92	107	1	0	0	0	22	178	0	0	0

Tabela A.7: Número estimado de consumidores na classe  $c = 1$  (residencial) nos transformadores  $i = 28, \dots, 39$  para dados gerados com  $I = 50$  nos casos onde os  $M$ 's são desbalanceados. O número verdadeiro de consumidores aparece em negrito e o número reportado está entre colchetes.

$\widehat{M}_1$												
#40						#41						
	61	62	[63]	64	65	66	62	63	[64]	65	66	67
2 sem replicação	4	29	115	46	5	1	1	29	119	45	5	1
4 sem replicação	13	104	80	3	0	0	16	101	82	1	0	0
2 com replicação	0	3	191	6	0	0	0	0	187	13	0	0
4 com replicação	0	136	64	0	0	0	0	121	79	0	0	0
#42						#43						
	69	70	71	[72]	73	74	59	60	61	<b>62</b>	[63]	64
2 sem replicação	0	0	1	92	105	2	0	1	10	66	106	17
4 sem replicação	0	0	39	157	4	0	0	14	101	78	7	0
2 com replicação	0	0	0	67	133	0	0	0	0	62	138	0
4 com replicação	0	0	3	197	0	0	0	0	105	95	0	0
#44						#45						
	64	[65]	66	67	68	<b>69</b>	62	63	[64]	65	<b>66</b>	67
2 sem replicação	0	0	33	80	87	0	0	0	43	84	48	25
4 sem replicação	0	0	0	12	188	0	0	0	15	136	48	1
2 com replicação	0	0	3	159	38	0	0	0	14	161	25	0
4 com replicação	0	0	0	0	200	0	0	0	0	160	40	0
#46						#47						
	59	[60]	61	62	63	<b>64</b>	57	58	59	[60]	61	62
2 sem replicação	0	2	59	54	85	0	0	7	50	112	30	1
4 sem replicação	0	0	0	32	168	0	0	19	103	78	0	0
2 com replicação	0	0	54	123	23	0	0	0	18	182	0	0
4 com replicação	0	0	0	28	172	0	0	3	150	47	0	0
#48						#49						
	57	58	59	[60]	<b>61</b>	62	61	62	63	[64]	<b>65</b>	66
2 sem replicação	0	0	13	100	71	16	0	0	6	93	80	21
4 sem replicação	0	0	25	135	40	0	0	0	11	122	66	1
2 com replicação	0	0	0	169	31	0	0	0	0	103	97	0
4 com replicação	0	0	1	183	16	0	0	0	0	160	40	0
#50												
	68	69	70	71	[72]	73						
2 sem replicação	0	0	0	2	103	95						
4 sem replicação	0	0	0	45	152	3						
2 com replicação	0	0	0	0	49	151						
4 com replicação	0	0	0	3	197	0						

Tabela A.8: Número estimado de consumidores na classe  $c = 1$  (residencial) nos transformadores  $i = 40, \dots, 50$  para dados gerados com  $I = 50$  nos casos onde os  $M$ 's são desbalanceados. O número verdadeiro de consumidores aparece em negrito e o número reportado está entre colchetes.



# Apêndice B

## Programa em R

```
# Bibliotecas utilizadas
library(MASS)
library(fda)
#
no = 96                                ##### Número de observações
F = matrix(c(0.98, 0.05, 0.02, 0.95),2,2)  ### Matriz fraude
p = matrix(c((F[1,1]/(F[1,1]+F[2,1])), (F[2,1]/(F[1,1]+F[2,1])),
(F[1,2]/(F[1,2]+F[2,2])), (F[2,2]/(F[1,2]+F[2,2]))),2,2)

#####
#### Simulando o número real de consumidores em cada transformador ####
#####
#### 5 transformadores - m's são balanceados (Casos 1 e 3)
ntrafos=5
m = matrix(c(0), ntrafos, 2)
set.seed(3)
m[,1] = sample(0:75, ntrafos, replace=TRUE )
m[,2] = 75 - m[, 1]
save(m, file="m.Rdata")
##### 5 transformadores - m's são desbalanceados (Casos 2 e 4)
ntrafos=5
m = matrix(c(0), ntrafos, 2)
set.seed(3)
m[,1] = sample(60:75, ntrafos, replace=TRUE )
```

```

m[,2] = 75 - m[, 1]
save(m, file="m.Rdata")
#### 50 transformadores - m's são balanceados (Casos 1 e 3)
ntrafos=50
m = matrix(c(0), ntrafos, 2)
set.seed(3)
m[,1] = sample(0:75, ntrafos, replace=TRUE )
m[,2] = 75 - m[, 1]
save(m, file="m.Rdata")
#### 50 transformadores - m's são desbalanceados (Casos 2 e 4)
ntrafos=50
m = matrix(c(0), ntrafos, 2)
set.seed(3)
m[,1] = sample(60:75, ntrafos, replace=TRUE )
m[,2] = 75 - m[, 1]
save(m, file="m.Rdata")

nclasse1 = colSums(m)[1]
nclasse2 = colSums(m)[2]

#####
## Simulando o número reportado de consumidores a partir do número verdadeiro ##
#####
load("m.Rdata")
X1 = matrix(c(NA),ntrafos,2)
X2 = matrix(c(NA),ntrafos,2)
set.seed(2)
for (k in 1:ntrafos){
  X1[k,] = rmultinom(1, size = m[k,1], prob=c(F[1,1], F[1,2]))
  X2[k,] = rmultinom(1, size = m[k,2], prob=c(F[2,1], F[2,2]))
}
R = X1 + X2
save(R, file="R.Rdata")
ncostumer = rowSums(R)

```

```
#####
##### Valores tabulados da função H #####
#####
load("R.Rdata")
Hhat_function = function(trafo,R,p,B, ncostumer){
  C = dim(R)[2]
  X = array(NA,c(C,C,B))
  for (c in 1:C){
    X[c,,] = rmultinom(B, R[trafo,c], prob = c(p[,c]))}
  Xmat <- apply(X,c(2,3),sum)  ## C by B matrix with M1 and M2's
  Hhat.l <- data.frame( table(Xmat[1,])/B )# gives frequencies of M1's
  Hhat.l.m1 <- as.numeric(as.vector(Hhat.l$Var1))
  Hhat.l.freq <- as.numeric(as.vector(Hhat.l$Freq))
  H = -log(Hhat.l.freq)
  factorial = - ( log(factorial(Hhat.l.m1)) +
  log(factorial(ncostumer[trafo]-Hhat.l.m1)) )
  Hhatl = data.frame(Hhat.l.m1, Hhat.l.freq, factorial, H)
  return(Hhatl)
}
set.seed(1)
Hhat = lapply(1:ntrafos,Hhat_function,R=R,p=p,B=100000, ncostumer=ncostumer)
save(Hhat, file="Hhat.Rdata")

#####
##### Gerando o conjunto de dados #####
#####
load("m.Rdata")
load("gamma_dataset.Rdata") # estimativas de gamma dos dados
sgtrafo_real=3.5 # Variância do erro aleatório
nbases_fixo = 9      ## Número de bases para o efeito fixo
nbases_aleatorio = 9  ## Número de bases para o efeito aleatório
t = (1/4)*1:96

#####
##### Gerando o conjunto de bases B-spline
bspline <- create.bspline.basis(rangeval=c(0,24),norder=3,nbasis=nbases_fixo)
```

```

bf <- getbasismatrix(evalarg=t, bspline, nderiv=0)
bspline <- create.bspline.basis(rangeval=c(0,24),norder=3,nbasis=nbases_aleatorio)
ba <- getbasismatrix(evalarg=t, bspline, nderiv=0)

##### Gerando a curva de demanda típica das classes (a1 e a2)
##### e as variâncias dos efeitos aleatórios (ae1 e ae2)
r1 = ( max(gamma_average[,1]) - min(gamma_average[,1]))
r2 = ( max(gamma_average[,2]) - min(gamma_average[,2]))

#### Casos 1 e 2 (a1 e a2 estão na mesma escala)
#sg1_real=0.03
#gama1_real = (gamma_average[,1] - min(gamma_average[,1]))/r1
#sg2_real=0.06
#gama2_real = (gamma_average[,2] - min(gamma_average[,2]))/r2

#### Cases 3 and 4 (a1 e a2 não estão na mesma escala)
sg1_real = 0.03 * r1^2
gama1_real = gamma_average[,1]
sg2_real = 0.06 * r2^2
gama2_real = gamma_average[,2]

a1 = bf %*% gama1_real
a2 = bf %*% gama2_real

##### Gerando o processo Gaussiano da classe 1 (a1li* )
sigma1 = ba %*% diag(sg1_real, nbases_aleatorio) %*% t(ba)
ae1 = mvrnorm(n=nclasse1, mu=matrix(c(0),1,no), Sigma= sigma1)
ae1 = t(ae1)
##### Gerando o processo Gaussiano da classe 2 (a2li* )
sigma2 = ba %*% diag(sg2_real, nbases_aleatorio) %*% t(ba)
ae2 = mvrnorm(n=nclasse2, mu=matrix(c(0),1,no), Sigma= sigma2)
ae2 = t(ae2)

##### Matriz com a curva de todos os consumidores em cada classe
w1 = matrix(c(0), no, nclasse1)

```

```

for( i in 1:nclasse1)
  w1[,i] = a1 + ae1[,i]
w2 = matrix(c(0), no, nclasse2)
for( i in 1:nclasse2)
  w2[,i] = a2 + ae2[,i]
### Simulando as curvas em cada transformador no caso sem replicação
y=matrix(c(0), no,ntrafos)
v = cbind(w1[,1:m[1,1]], w2[,1:m[1,2]])
y[, 1] = apply(v,1,sum)
for ( i in 1:(ntrafos-1)){
  v = cbind(w1[, (sum(m[1:i ,1])+1) : (sum( m[1: (i+1),1])) ],
            w2[, (sum(m[1:i ,2])+1) : (sum( m[1: (i+1),2])) ] )
  y[,i+1] = apply(v,1,sum) }

Y = matrix(NA, no, ntrafos)
for ( i in 1:ntrafos)
  Y[, i] = y[, i] + rnorm(no,0,sqrt(sgtrafo_real))

### Simulando as curvas em cada transformador no caso com replicação
data = list(NA)
nrep=5
for ( j in 1:nrep){
y = matrix(NA, no, ntrafos)
for ( i in 1:ntrafos){
  m1 = m[i,1]
  m2 = m[i,2]
  if(m1 !=0 & m1!=1 & m2!=0 & m2!=1){
    y[,i] = rowSums( cbind( m1*a1, m2*a2,
                           t(mvrnorm(n=m1, mu=matrix(c(0),1,no), Sigma= sigma1)),
                           t(mvrnorm(n=m2, mu=matrix(c(0),1,no), Sigma= sigma2)) ))
  }
  if(m1 == 0){
    y[,i] = rowSums( cbind( m1*a1, m2*a2,
                           t(mvrnorm(n=m2, mu=matrix(c(0),1,no), Sigma= sigma2))))
  }
}
}

```

```

if(m2 == 0){
  y[,i] = rowSums( cbind( m1*a1, m2*a2,
    t(mvrnorm(n=m1, mu=matrix(c(0),1,no), Sigma= sigma1))))
}
if(m1 == 1){
  y[,i] = rowSums( cbind( m1*a1, m2*a2,
    mvrnorm(n=m1, mu=matrix(c(0),1,no), Sigma= sigma1),
    t(mvrnorm(n=m2, mu=matrix(c(0),1,no), Sigma= sigma2))) )
}
if(m2 == 1){
  y[,i] = rowSums( cbind( m1*a1, m2*a2,
    t(mvrnorm(n=m1, mu=matrix(c(0),1,no), Sigma= sigma1)),
    mvrnorm(n=m2, mu=matrix(c(0),1,no), Sigma= sigma2)) )
}
}
Y = matrix(NA, no, ntrafos)
for ( i in 1:ntrafos)
  Y[, i] = y[, i] + rnorm(no,0,sqrt(sgtrafo_real))
  data[[j]] = Y
}
#####
#### Passos 1 e 2 sem replicação ####
#####
set.seed(1)
Y_all = NULL
updates=NULL
M1s = NULL
source("first.R")
source("second.R")
source("third.R")
M=R
bb=200
for (a in 1:bb){
  source("fourth_norepl.R")
  Y_all = cbind(Y_all, Y)
}

```

---

```

M=R
svd = svd(t(ba)%*%ba)
others = list(c = (1/2)*(no+1)*log(2*pi), D=svd$d, bp = ba%*%svd$v,
bb = ba%*%t(ba), ba=ba, bf=bf, Y=Y)
update_step1 = NULL
current_step1 = list(gamma = matrix(c(gama1_real, gama2_real),nbases_fixo,2),
                      sg_class = c(sg1_real, sg2_real), sgt = sgtrafo_real)
all_updates <- NULL
all_criteria <- NULL
all_M1 = NULL
dif = matrix(c(1),50,1)

while ( sum(dif) > 0){

  iteration = 0
  vc = c(0.0001, 0.0001, 0.001, 0.001, 0.001)
  criteria=rep(1000, 5)
  all_updates <- NULL
  all_criteria <- NULL

  ##### Passo 1 #####
  while( sum(criteria > vc) > 0 ) {
    ##### Passo 1.a #####
    l1trafo = function(sg_class, sgt, Y, bp, D, ba, M ) {
      lambda = rowSums(t(t(M)*sg_class))
      lambdastar = sgt/lambda
      termo1=0
      termo2=0
      for ( i in 1:ntrafos) {
        A = (1/sgt)*(diag(1,no) - bp%*(diag(1/(lambdastar[i]+D)))%*%t(bp))
        B = cbind(M[i,1] * bf, M[i,2] * bf)
        termo1 = termo1 + t(B)%*%A%*%B
        termo2 = termo2 + t(B)%*%A%*%Y[,i]
      }
      return(solve(termo1)%*%termo2)
    }
  }
}

```

```

}
out = l1trafo(sg_class = current_step1$sg_class, sgt = current_step1$sgt,
             Y=others$Y, bp = others$bp, D=others$D, ba=others$ba, M=M)

update_step1$gamma = matrix(c(out), nbases_fixo, 2)

##### Passo 1.b #####
l2trafo <- function(x,i,gamma,sg_class,c,bb,D,bp,bf,Y,M) {
  sgt = x[1]
  vec = (bf%*%gamma[,1]%*%M[,1]) + (bf%*%gamma[,2]%*%M[,2])
  mat = Y - vec
  lambda = rowSums(t(t(M)*sg_class))
  lambdastar = (sgt)/lambda
  out1 = c + log(det(lambda[i]*bb + sgt*diag(1,no))) + (1/sgt)*(t(mat[,i])
  %*% (diag(1,no) - bp%*%(diag(1/(lambdastar[i]+D))))%*%t(bp)) %*% mat[,i])
  return(out1)
}

l12 = function(x,i,gamma,sg_class,c,bb,D,bp,bf,Y,M) {
  out2 = sum(sapply(1:ntrafos,l2trafo,x=x, gamma=update_step1$gamma,
  sg_class=current_step1$sg_class,c=others$c,bb=others$bb,D=others$D,
  bp=others$bp,bf=others$bf,Y=others$Y,M=M))
  return(out2)
}

update_step1$sgt = optimize(l12, c(0, 10), gamma=update_step1$gamma,
  sg_class=current_step1$sg_class,bb=others$bb,
  D=others$D, bp=others$bp,Y=others$Y,M=M)$minimum

##### Passo 1.c #####
l3trafo<- function(x,i,gamma,sgt,c,bb,D,bp,bf,Y,M)
{
  sg_class = x
  vec = (bf%*%gamma[,1]%*%M[,1]) + (bf%*%gamma[,2]%*%M[,2])
  mat = Y - vec
  lambda = rowSums(t(t(M)*sg_class))
  lambdastar = (sgt)/lambda

```

---

```

    out1 = c+ log(det(lambda[i]*bb + sgt*diag(1,no))) + (1/sgt)*(t(mat[,i])
      %% (diag(1,no) - bp%%(diag(1/(lambdastar[i]+D)))%%t(bp))
      %% mat[,i])
    return(out1)
  }
l13 = function(x,i,gamma,sgt,c,bb,D,bp,bf,Y,M) {
  out2 = sum(sapply(1:ntrafos,l3trafo,x=x, gamma=update_step1$gamma,
    sgt=update_step1$sgt, c=others$c, bb=others$bb, D=others$D,
    bp=others$bp, bf=others$bf, Y=others$Y, M=M))
  return(out2)
}

update_step1$sg_class = optim(current_step1$sg_class,
  gamma=update_step1$gamma, sgt=update_step1$sgt,
  c=others$c, bb=others$bb, D=others$D,
  bp=others$bp, bf=others$bf,Y=others$Y, M=M,
  l13, method="L-BFGS-B",
  lower=c(0.0001, 0.0001), upper=c(10,10))$par

criteria = c(colMeans(abs(apply(update_step1$gamma,2,
  function(x,b)
  {b%%x},b=bf) -
  apply(current_step1$gamma,2,function(x,b){b%%x},b=bf))),
  Mod(update_step1$sg_class - current_step1$sg_class)/
  Mod(update_step1$sg_class),
  Mod(update_step1$sgt-current_step1$sgt)/Mod(update_step1$sgt))

print(criteria)
iteration = iteration + 1
print(iteration)
all_updates <- list(all_updates,update_step1)
all_criteria <- cbind(all_criteria,criteria)
current_step1 <- update_step1
}

```

```
##### Passo 2 #####
l2_trafo = function(l, gamma, sgt, sg_class, c, bb, bp, bf, D, Y) {
  load("Hhat.Rdata")
  Hhat2.1 = Hhat[[1]]
  var1 = Hhat2.1$Hhat.1.m1
  var2 = ncostumer[1] - var1
  var = matrix(c(var1, var2), length(var1), 2)
  vec = (bf**gamma[,1]**var1) + (bf**gamma[,2]**var2)
  mat = Y[,1] - vec
  lambda = rowSums(t(t(var)*sg_class))
  lambdastar = (sgt)/lambda
  l2 = matrix(NA, length(var1),1)
  for (i in 1:length(var1)){
    l2[i] = c + ( (1/2)*log(det(lambda[i]*bb + sgt*diag(1,no))) +
      (1/2)*(1/sgt)*(t(mat[,i]) ** (diag(1,no) -
        bp**(diag(1/(lambdastar[i]+D))))**t(bp)) ** mat[,i]) +
      Hhat2.1$factorial[i] + Hhat2.1$H[i] )
  }
  M1_trafo = var1[which(l2==min(l2))]
  return(M1_trafo)
}
M1 = sapply(1:ntrafos, l2_trafo, gamma=current_step1$gamma,
  sgt=current_step1$sgt, sg_class=current_step1$sg_class, c=others$c,
  bb=others$bb, bp=others$bp, bf=others$bf, D=others$D, Y=others$Y)
M_new=matrix(c(M1, ncostumer-M1),ntrafos,2)

all_M1 = cbind(all_M1, M1)
M_c1 = cbind(R[,1], all_M1, m[,1])
dif = Mod(M1 - M[,1])
M = M_new
}
M1s = cbind(M1s, M1)
updates[[a]] = current_step1
print(M_c1)
print(a)
```

```
}

#####
#### Passos 1 e 2 com replicação ####
#####

set.seed(1)
Y_all = NULL
updates=NULL
M1s = NULL
source("first.R")
source("second.R")
source("third.R")
M=R
bb=200
for (a in 1:bb){
  source("fourth_repl.R")
  M=R
  svd = svd(t(ba)%*%ba)
  others = list(c = (1/2)*(no+1)*log(2*pi), D=svd$d, bp = ba%*%svd$v,
  bb = ba%*%t(ba), ba=ba, bf=bf, data=data)
  update_step1 = NULL
  current_step1 = list(gamma = matrix(c(gama1_real, gama2_real),nbases_fixo,2),
                        sg_class = c(sg1_real, sg2_real), sgt = sgtrafo_real)
  all_updates <- NULL
  all_criteria <- NULL
  all_M1 = NULL
  dif = matrix(c(1),ntrafos,1)

  while ( sum(dif) > 0){

    iteration = 0
    vc = c(0.0001, 0.0001, 0.001, 0.001, 0.001)
    criteria=rep(1000, 5)
    all_updates <- NULL
    all_criteria <- NULL
```

```

##### Passo 1 #####
while( sum(criteria > vc) > 0 ) {
  ##### Passo 1.a #####
  l1trafo = function(sg_class, sgt, Y, bp, D, ba, M, data ) {
    lambda = rowSums(t(t(M)*sg_class))
    lambdastar = sgt/lambda
    termo1=0
    termo2=0

    for ( i in 1:ntrafos) {
      for ( j in 1:nrep){
        A = (1/sgt)*(diag(1,no) - bp%*(diag(1/(lambdastar[i]+D)))%*t(bp))
        B = cbind(M[i,1] * ba, M[i,2] * ba)
        termo1 = termo1 + t(B)%*A%*B
        termo2 = termo2 + t(B)%*A%*data[[j]][,i]
      }
    }
    return(solve(termo1)%*termo2)
  }
  out = l1trafo(sg_class = current_step1$sg_class, sgt = current_step1$sgt,
    data=others$data, bp = others$bp, D=others$D, ba=others$ba, M=M)
  update_step1$gamma = matrix(c(out), nbases_fixo, 2)

  ##### Passo 1.b #####
  l2trafo <- function(x,i,gamma,sg_class,c,bb,D,bp,bf,data,M) {
    sgt = x[1]
    out1=0
    for ( j in 1:nrep){
      vec = (bf%*gamma[,1]%*M[,1]) + (bf%*gamma[,2]%*M[,2])
      mat = data[[j]] - vec
      lambda = rowSums(t(t(M)*sg_class))
      lambdastar = (sgt)/lambda

      out1 = out1 + c + log(det(lambda[i]*bb + sgt*diag(1,no))) +
        (1/sgt)*(t(mat[,i]) %*% (diag(1,no) -

```

```

        bp%*(diag(1/(lambdastar[i]+D)))%*t(bp)) %*% mat[,i])
    }
    return(out1)
}
l12 = function(x,i,gamma,sg_class,c,bb,D,bp,bf,Y,data,M) {
    out2 = sum(sapply(1:ntrafos,l2trafo,x=x, gamma=update_step1$gamma,
        sg_class=current_step1$sg_class,c=others$c,bb=others$bb,D=others$D,
        bp=others$bp,bf=others$bf,data=others$data,M=M))
    return(out2)
}
update_step1$sgt = optimize(l12, c(0, 10), gamma=update_step1$gamma,
    sg_class=current_step1$sg_class,bb=others$bb,
    D=others$D, bp=others$bp,data=others$data,M=M)$minimum

##### Passo 1.c #####
l3trafo<- function(x,i,gamma,sgt,c,bb,D,bp,bf,data,M)
{
    sg_class = x
    out1=0
    for ( j in 1:nrep){
        vec = (bf%*%gamma[,1]%*%M[,1]) + (bf%*%gamma[,2]%*%M[,2])
        mat = data[[j]] - vec
        lambda = rowSums(t(t(M)*sg_class))
        lambdastar = (sgt)/lambda

        out1 = out1+ c+ log(det(lambda[i]*bb + sgt*diag(1,no))) +
            (1/sgt)*(t(mat[,i]) %*% (diag(1,no) -
            bp%*(diag(1/(lambdastar[i]+D)))%*t(bp)) %*% mat[,i])
    }
    return(out1)
}
l13 = function(x,i,gamma,sgt,c,bb,D,bp,bf,data,M) {
    out2 = sum(sapply(1:ntrafos,l3trafo,x=x, gamma=update_step1$gamma,
        sgt=update_step1$sgt, c=others$c, bb=others$bb, D=others$D,
        bp=others$bp, bf=others$bf, data=others$data, M=M))

```

```

    return(out2)
  }

  update_step1$sg_class = optim(current_step1$sg_class,
    gamma=update_step1$gamma, sgt=update_step1$sgt, c=others$c,
    bb=others$bb, D=others$D, bp=others$bp, bf=others$bf,
    data=others$data, M=M, l13, method="L-BFGS-B",
    lower=c(0.0001, 0.0001), upper=c(10,10))$par

  criteria = c( colMeans(abs(apply(update_step1$gamma,2,
    function(x,b){b*%x},b=bf) - apply(current_step1$gamma,2,
    function(x,b){b*%x},b=bf))),
    Mod(update_step1$sg_class - current_step1$sg_class)/
    Mod(update_step1$sg_class),
    Mod(update_step1$sgt - current_step1$sgt)/Mod(update_step1$sgt) )

  print(criteria)
  iteration = iteration + 1
  print(iteration)
  all_updates <- list(all_updates,update_step1)
  all_criteria <- cbind(all_criteria,criteria)
  current_step1 <- update_step1
}

##### Passo 2 #####
l2_trafo = function(l, gamma, sgt, sg_class, c, bb, bp, bf, D, data) {
  load("Hhat.Rdata")
  Hhat2.1 = Hhat[[1]]
  var1 = Hhat2.1$Hhat.1.m1
  var2 = ncostumer[1] - var1
  var = matrix(c(var1, var2), length(var1), 2)
  vec = (bf*%gamma[,1]*%var1) + (bf*%gamma[,2]*%var2)
  lambda = rowSums(t(t(var)*sg_class))
  lambdastar = (sgt)/lambda
  l2 = matrix(c(0), length(var1),1)
  for (i in 1:length(var1)){

```

---

```

    for( j in 1:nrep){
      mat = data[[j]][,1] - vec
      l2[i] = l2[i] + c + ( (1/2)*log(det(lambda[i]*bb + sgt*diag(1,no)))
        + (1/2)*(1/sgt)*(t(mat[,i]) %*% (diag(1,no) -
          bp%*%(diag(1/(lambdastar[i]+D)))%*%t(bp)) %*% mat[,i]) +
          Hhat2.1$factorial[i] + Hhat2.1$H[i] )
    }
  }
  M1_trafo = var1[which(l2==min(l2))]
  return(M1_trafo)
}

M1 = sapply(1:ntrafos, l2_trafo, gamma=current_step1$gamma,
  sgt=current_step1$sgt, sg_class=current_step1$sg_class, c=others$c,
  bb=others$bb, bp=others$bp, bf=others$bf, D=others$D, data=others$data)
M_new = matrix(c(M1, (ncostumer - M1)), ntrafos,2)
all_M1 = cbind(all_M1, M1)
M_c1 = cbind(R[,1], all_M1, m[,1])
dif = Mod(M1 - M[,1])
M = M_new
}

M1s = cbind(M1s, M1)
updates[[a]] = current_step1
print(M_c1)
print(a)
}

```



# Referências Bibliográficas

- de Souza, C. and Dias, R. (2008), “Testes de hipóteses para dados funcionais baseados em distâncias: um estudo usando splines,” . 6
- Dias, R., Garcia, N., and Martarelli, A. (2009), “Non-parametric estimation for aggregated functional data for electric load monitoring,” *Environmetrics*, 20, 111–130. 2
- Dias, R., Garcia, N., and Schmidt, A. (2011), “A hierarchical model for aggregated functional data,” *arXiv preprint arXiv:1102.2773*. 2
- Ferraty, F. and Vieu, P. (2006), *Nonparametric functional data analysis: theory and practice*, Springer. 1
- Heckman, N., Lockhart, R., and Nielsen, J. (2009), “Penalized Regression, Mixed Effects Models and Appropriate Modelling,” . 6
- Ramsay, J. O. and Silverman, B. W. (2005), *Functional data analysis*, Wiley Online Library. 1
- Ruppert, D., Wand, M. P., and Carroll, R. J. (2003), *Semiparametric regression*, vol. 12, Cambridge University Press. 6
- Verbyla, A., Cullis, B., Kenward, M., and Welham, S. (2002), “The analysis of designed experiments and longitudinal data by using smoothing splines,” *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 48, 269–311. 6