



Universidade Estadual de Campinas
Instituto de Matemática, Estatística e
Computação Científica - IMECC
Departamento de Estatística



ESTIMAÇÃO NÃO PARAMÉTRICA DA FUNÇÃO DE
COVARIÂNCIA PARA DADOS FUNCIONAIS AGREGADOS

Dissertação de Mestrado

Guilherme Vieira Nunes Ludwig

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

Coorientador: **Prof. Dr. Ronaldo Dias**

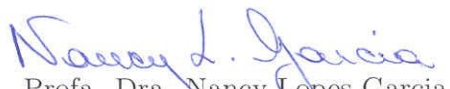
Campinas – SP

2011

Estimação não paramétrica da função de covariância para dados funcionais agregados

Este exemplar corresponde à redação final da dissertação devidamente corrigida e defendida por Guilherme Vieira Nunes Ludwig, aprovada pela Comissão Julgadora.

Campinas 20 de maio de 2011


Profa. Dra. Nancy Lopes Garcia
Orientadora


Prof. Dr. Ronaldo Dias
Coorientador

Banca Examinadora:

- Profa. Dra. Nancy Lopes Garcia (IMECC - UNICAMP) - Orientadora
- Profa. Dra. Alexandra Mello Schmidt (IM - UFRJ)
- Prof. Dr. Álvaro Rodolfo de Pierro (IMECC - UNICAMP)

Dissertação apresentada ao Instituto de Matemática, Estatística e Computação Científica da UNICAMP, como requisito parcial para obtenção do Título de Mestre em Estatística.

**FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DO IMECC DA UNICAMP**

Bibliotecária: Maria Fabiana Bezerra Müller – CRB8 / 6162

L966e	Ludwig, Guilherme Vieira Nunes Estimação não paramétrica da função de covariância para dados funcionais agregados/Guilherme Vieira Nunes Ludwig-- Campinas, [S.P. : s.n.], 2011. Orientador : Nancy Lopes Garcia; Ronaldo Dias Dissertação (mestrado) - Universidade Estadual de Campinas, Instituto de Matemática, Estatística e Computação Científica. 1.Estatística não paramétrica. 2.Função de covariância. 3.Spline, Teoria do. 4.Análise multivariada. I. Garcia, Nancy Lopes. II.Dias, Ronaldo. III. Universidade Estadual de Campinas. Instituto de Matemática, Estatística e Computação Científica. IV. Título.
-------	--

Título em inglês: Nonparametric estimation of the covariance function for aggregated functional data

Palavras-chave em inglês (Keywords): 1.Nonparametric statistics. 2.Covariance function. 3.Spline theory. 4.Multivariate analysis.

Área de concentração: Estatística Não Paramétrica

Titulação: Mestre em Estatística

Banca examinadora: Profª. Dra. Nancy Lopes Garcia (IMECC – UNICAMP)
Prof. Dr. Álvaro Rodolfo de Pierro (IMECC – UNICAMP)
Profª. Dra. Alexandra Mello Schmidt (IM - UFRJ)

Data da defesa: 20/05/2011

Programa de Pós-Graduação: Mestrado em Estatística

Dissertação de Mestrado defendida em 20 de maio de 2011 e aprovada

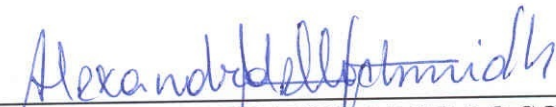
Pela Banca Examinadora composta pelos Profs. Drs.



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



Prof(a). Dr(a). ALVARO RODOLFO DE PIERRO



Prof(a). Dr(a). ALEXANDRA MELLO SCHMIDT

“Nossos historiadores, que são os mais perspicazes do orbe, inventaram um método para corrigir o acaso; é notório que as operações desse método são (em geral) fidedignas; malgrado, naturalmente, não se divulguem sem alguma dose de engano.”

Jorge Luis Borges, *A Loteria da Babilônia* (1941)

“Babilônia não é outra coisa senão um infinito jogo de acasos.”

Ibidem

Agradecimentos

Esta dissertação não teria se materializado sem o apoio de colegas, amigos e professores a quem dedico essa página. Primeiramente agradeço aos meus orientadores, a Profa. Nancy Garcia e o Prof. Ronaldo Dias. Fiquei muito feliz por trabalhar sob orientação da Profa. Nancy. Ela é uma professora e pesquisadora fantástica, e como orientadora sempre foi muito gentil e paciente, corrigindo contas e defeitos na minha exposição de ideias, oferecendo apoio e sugestões quando eu ficava meio perdido e incentivo quando as contas davam certo! Também agradeço ao meu coorientador Prof. Ronaldo, com quem eu tive diversas discussões animadas sobre a teoria de Splines (suavização, transformadas de Fourier, equações diferenciais...) e que me deu bons conselhos sobre como ser claro e persuasivo na exposição dos nossos resultados – espero ter conseguido sê-lo neste trabalho. Muito do fascínio que eu sinto pela análise de dados funcionais atualmente, e pelas possibilidades de pesquisa que ela tem a oferecer, devo a eles!

Agradeço à banca de minha defesa, composta pelo Prof. Álvaro de Pierro e pela Profa. Alexandra Schmidt. Trabalhei com o Prof. Álvaro em minha Iniciação Científica e ele também me motivou muito a seguir a carreira acadêmica. Fiquei feliz em tê-lo como membro da banca! Também agradeço muito à leitura do meu texto feita pela Profa. Alexandra, por seu importante questionamento sobre como quantificar a incerteza de nossas estimativas – que tentei sanar, ao menos nos Capítulos 2 e 4 – e finalmente pelas suas palavras encorajadoras durante a defesa!

Agradeço aos meus colegas de vida acadêmica, pelas experiências que compartilhamos nos últimos dois anos. Sobretudo meus amigos mais próximos: o Rodrigo Lambert,

que me arrastou para longe da zona de conforto intelectual, sempre me lembrando da importância da matemática, e me incentivando a conhecer trabalhos de outras áreas, além das festas e seminários na USP que foram muito legais; o José Schultz, com quem dividi uns momentos de depressão e outros de criatividade, e quase as mesmas opiniões; e o Márcio Diniz, que em todo ponto de vista é um cara mais sensato do que eu, e com quem pude várias vezes conversar sobre minhas inseguranças e ouvir uma opinião mais ajuizada. Sem falar que *ele* era o gabarito de qualquer matéria que fizéssemos juntos, o que sempre salvou minha pele quando estudávamos para qualificações e provas complicadas.

Também quero agradecer ao Prof. Miguel Abadi, com quem trabalhei no final da minha graduação em estatística, e que me influenciou a seguir em frente no mestrado! Espero algum dia fazer matemática com a elegância dele.

Agradeço à minha amiga de infância Raquel Gryszczenko, que atualmente é um dos prodígios entre os doutorandos do Depto. de História aqui da UNICAMP, pelo apoio e pela cuidadosa revisão que ela fez deste texto.

Agradeço minha família por suportar minha ausência, mesmo ainda morando com eles durante todo o período. Minha irmã e minha mãe me deram todo o apoio que podiam durante essa etapa da minha vida.

A dissertação foi concebida principalmente ao som de *Kid A* do Radiohead e *Fear of Music* do Talking Heads. É uma trilha sonora bem excêntrica pra escrever uma dissertação de mestrado, mas acho que funcionou!

Este trabalho contou com apoio financeiro do Conselho Nacional de Pesquisa e Desenvolvimento Tecnológico - CNPq, processo de número 135291/2009-7.

Resumo

O objetivo desta dissertação é desenvolver estimadores não paramétricos para a função de covariância de dados funcionais agregados, que consistem em combinações lineares de dados funcionais que não podem ser observados separadamente. Estes métodos devem ser capazes de produzir estimativas que separem a covariância típica de cada uma das subpopulações que geram os dados, e que sejam funções não negativas definidas. Sob estas restrições, foi definida uma classe de funções de covariância não estacionárias, à qual resultados da teoria de estimação de covariância de processos estacionários podem ser estendidos. Os métodos desenvolvidos foram ilustrados com a aplicação em dois problemas reais: a estimação do perfil de consumidores de energia elétrica, em função do tempo, e a estimação da transmitância de substâncias puras em espectroscopia de infravermelho, através da inspeção de misturas, em função do espectro da luz.

Abstract

The goal of this dissertation is to develop nonparametric estimators for the covariance function of aggregated functional data, which consists into linear combinations of functional data that cannot be sampled separately. Such methods must be able to produce estimates that not only separate the typical covariance of the subpopulations composing the data, but also be nonnegative definite functions. Under these restrictions, a class of nonstationary covariance functions was proposed, to which stationary processes' covariance function estimation results can be readily extended. The developed methods were illustrated with an application to two real problems: the estimation of electric energy consumers' profiles, as a function of the time of the day, and the estimation of the transmittance of pure substances in infrared spectroscopy, while inspecting mixtures of them, as a function of light spectrum.

Sumário

Lista de Figuras	xviii
Introdução	xix
1 Estimação de Tipologia para Dados Funcionais Agregados	1
1.1 Definição de Dado Funcional	1
1.2 B-Splines	3
1.3 Dados Funcionais Agregados	5
1.4 Estimação da Tipologia	7
1.5 Simulação de Curvas Agregadas	10
2 Função de Covariância de Dados Funcionais Agregados	15
2.1 Covariância de uma Função Aleatória	15
2.2 Covariância Semiparametrizada de Dados Funcionais Agregados	22
2.3 Exemplo Simulado	24
3 Estimação de Covariância pelo Produto Tensorial de Splines	33
3.1 O Produto Tensorial de Splines	33
3.2 O Produto Tensorial de Splines Suavizados	36
3.3 Transformada de Fourier do Produto Tensorial de Splines	44
3.4 Exemplo Simulado	47
4 Estimação da Covariância pela sua Representação Espectral	53

4.1	O Teorema de Karhunen-Loève	53
4.2	Aplicação em Dados Agregados	55
4.3	Exemplo Simulado	58
5	Aplicação em Dados Reais	67
5.1	Consumo de Energia Elétrica	67
5.1.1	Modelagem de Tipologia	69
5.1.2	Modelagem da Função de Covariância Típica	70
5.2	Quimiometria	76
5.2.1	Modelagem de Tipologia	78
5.2.2	Modelagem da Função de Covariância Típica	78
6	Considerações Finais	83
	Referências Bibliográficas	86
A	A Transformada de Fourier de Splines	91
A.1	Construindo B-splines	91
A.2	Propriedades da Transformada de Fourier	92
A.3	Construindo B-splines através de Convoluções	93
A.4	Transformada do B-spline de ordem m	98
B	Algoritmos	103
B.1	Sobre o Método de Nelder-Mead	103
B.2	Rotinas no GNU/R	105

Lista de Figuras

1.1	11 bases B-spline em $[0, 10]$	4
1.2	Tipologias simuladas.	11
1.3	$J = 5$ amostras de $y_i(t)$, e as curvas estimadas $\hat{y}_i(t)$	12
1.4	$J = 30$ amostras de $y_i(t)$, e as curvas estimadas $\hat{y}_i(t)$	12
1.5	Tipologias estimadas, com diferentes números de replicações.	13
2.1	$\hat{\eta}_1^2(t)$ e $\hat{\eta}_2^2(t)$ estimados com $J = 5$	26
2.2	$\hat{\eta}_1^2(t)$ e $\hat{\eta}_2^2(t)$ estimados com $J = 30$	26
2.3	$\hat{\eta}_1^2(t)$ e $\hat{\eta}_2^2(t)$ estimados com $J = 100$	27
2.4	$\hat{\eta}_1^2(t)$ e $\hat{\eta}_2^2(t)$ estimados com $J = 300$	27
2.5	Estimativas $\hat{\eta}_c^2(t)$ sobrepostas, para diferentes J	28
2.6	Estimativas $\hat{\eta}_c^2(t)$ sobrepostas, para diferentes J , com truncamento em $\hat{\mathbf{Z}}_i$	28
2.7	Estimativas $\hat{\eta}_1^2(t)$ e $\hat{\eta}_2^2(t)$ com $J = \infty$	29
2.8	Intervalo de confiança (confiabilidade de aproximadamente 95%) para $\hat{\eta}_1^2(t)$ e $\hat{\eta}_2^2(t)$ estimados com $J = 30$	30
2.9	Intervalo de confiança (confiabilidade de aproximadamente 95%) para $\hat{\eta}_1^2(t)$ e $\hat{\eta}_2^2(t)$ estimados com $J = 300$	30
2.10	Intervalo de confiança (confiabilidade de aproximadamente 95%) para ϕ_1 e ϕ_2 estimados com $J = 30$	31
2.11	Intervalo de confiança (confiabilidade de aproximadamente 95%) para ϕ_1 e ϕ_2 estimados com $J = 300$	31

3.1	A base $B_6(t)$ da Figura 1.1, e a base $B_6(s)B_6(t)$ do produto tensorial de Splines	34
3.2	(A) A função $f(t)$ e a amostra $\mathbf{y}$; (B) Estimativa com 4 nós internos; (C) Estimativa com 1 nó interno; (D) Estimativa com 30 nós internos . . .	38
3.3	A função $f(s, t)$ para o exemplo de estimação por produto tensorial . .	39
3.4	A aproximação $\hat{f}(s, t)$, com 7^2 bases	39
3.5	A aproximação $\hat{f}(s, t)$, com 18^2 bases	40
3.6	A aproximação $\hat{f}(s, t)$, com 20^2 bases e penalização de $\lambda = 6.10^{-5}$. . .	42
3.7	$\hat{\sigma}_c(s, t)$ estimado com $J = 30$ e $\hat{\lambda} = 1,28.10^{-5}$	47
3.8	$\hat{\sigma}_c(t, t)$ estimado com $J = 30$ e $\hat{\lambda} = 1,28.10^{-5}$	48
3.9	$\hat{\rho}_1(s, t)$ e $\hat{\rho}_2(s, t)$, após truncamento e suavização	49
3.10	Parte Real e Imaginária de $\hat{\mathcal{F}}_1(w_1, w_2)$	50
3.11	Parte Real e Imaginária de $\hat{\mathcal{F}}_2(w_1, w_2)$	50
3.12	$\hat{\sigma}_1(s, t)$ e $\hat{\sigma}_2(s, t)$ estimados com $J = 30$, após correção da sua Transformada de Fourier	51
3.13	As variâncias $\hat{\sigma}_1(t, t)$ e $\hat{\sigma}_2(t, t)$, estimadas com $J = 30$, após correção da sua Transformada de Fourier	51
4.1	Polinômios de Legendre $P_0(x)$ a $P_4(x)$	58
4.2	Funções de covariância de $W_1(t)$ e $W_2(t)$	59
4.3	Curvas simuladas $y_1(t)$, $y_2(t)$ e $y_3(t)$, $J = 30$	60
4.4	As três primeiras autofunções obtidas de $\hat{\mathbf{Z}}$	60
4.5	Estimativas de $\sigma_1(s, t)$ e $\sigma_2(s, t)$, $J = 30$	61
4.6	Estimativas de $\sigma_1(s, t)$ e $\sigma_2(s, t)$, $J = 100$	61
4.7	Estimativas de $\sigma_1(t, t)$ e $\sigma_2(t, t)$	62
4.8	Estimativas de $\sigma_1(t, t + 0,125)$ e $\sigma_2(t, t + 0,125)$	62
4.9	Região de confiança (aproximadamente 90% de confiabilidade) para $\sigma_c(t, t)$, com $J = 30$	64

4.10	Região de confiança (aproximadamente 90% de confiabilidade) para $\sigma_c(t, t)$, com $J = 100$	64
4.11	Região de confiança (aproximadamente 90% de confiabilidade) para $\sigma_c(t, t + 0,125)$, com $J = 30$	65
4.12	Região de confiança (aproximadamente 90% de confiabilidade) para $\sigma_c(t, t + 0,125)$, com $J = 100$	65
5.1	Curvas agregadas amostradas em certa região de Campinas.	68
5.2	As 5 amostras dos Transformadores TR07**** e TR09****, e suas estimativas.	70
5.3	Tipologias estimadas dos consumidores residenciais monofásicos e bifásicos, de uma região central de Campinas.	71
5.4	A função de variância estimada para os consumidores monofásico e bifásico, por semiparametrização.	72
5.5	As funções $\hat{\sigma}_1(t, t + 0,5)$ e $\hat{\sigma}_2(t, t + 0,5)$, dos consumidores monofásico e bifásico, respectivamente.	72
5.6	A função de variância estimada para os consumidores monofásico e bifásico, por produto tensorial.	73
5.7	As funções $\hat{\sigma}_1(t, t + 0,5)$ e $\hat{\sigma}_2(t, t + 0,5)$, dos consumidores monofásico e bifásico, respectivamente.	74
5.8	Intervalos de Confiança pontual a 90% para as estimativas de Tipologia	75
5.9	Absorbância de carne, para diferentes concentrações de umidade, gordura e proteína	77
5.10	Frequências das concentrações de Umidade, Gordura e Proteína	78
5.11	Dispersão duas a duas das concentrações de Umidade, Gordura e Proteína	78
5.12	As tipologias de absorbância para Umidade, Gordura e Proteína	79
5.13	Ordenação das concentrações de umidade	79
5.14	Os três primeiros autovetores e as autofunções normalizadas	80
5.15	As estimativas de Covariância para cada um dos constituintes	81

A.1	11 bases B-spline em $[0, 10]$, bases com nós repetidos realçadas.	96
A.2	As bases $B_{-2,4}(t)$, $B_{-1,4}(t)$ e $B_{0,4}(t)$, com $r = 1$	99
A.3	A Transformada de Fourier das bases $B_{-2,4}(t)$, $B_{-1,4}(t)$, $B_{0,4}(t)$ e $B_{1,4}(t)$, com $r = 1$; a parte real em linha sólida, a parte imaginária em linha pontilhada	101
A.4	A função $\exp\{-t^2\}$, sua transformada de Fourier e as respectivas apro- ximações por Splines	102

Introdução

A análise de dados funcionais é um ramo da estatística que estuda curvas, superfícies ou funções que variam em intervalos contínuos, tais como tempo e espaço. A área encontra motivação em problemas físicos e químicos, biológicos e financeiros. O objetivo é modelar características dos dados, como funções de valor esperado e covariância, ou analisar outras características como curvatura, taxas de mudança e derivadas.

Os dados funcionais *agregados* surgem quando é impossível observar os dados funcionais individuais diretamente. Nesta situação cada curva agregada é composta pela combinação linear de curvas de diferentes subpopulações, onde uma subpopulação é um conjunto de indivíduos que compartilham o mesmo perfil médio, chamado *tipologia*, e a mesma estrutura de dependência. Os trabalhos de Martarelli (2006) e Dias, Garcia & Martarelli (2009) abordam a questão de estimação das tipologias de consumidores de energia elétrica, considerando três subpopulações (residencial monofásico, residencial bifásico e comercial) e amostrando as curvas agregadas em diversos transformadores de Campinas, São Paulo. Saraiva (2009) considera a estimação de tipologias de absorbância/transmitância, para diferentes substâncias químicas, agregadas em misturas de diferentes concentrações, com o objetivo de determinar a tipologia de uma substância pura.

A estimação da covariância, entretanto, envolve um problema adicional: é necessário que os estimadores propostos, além de separarem a superfície de covariância típica de cada subpopulação, também produzam estimativas dentro de um espaço restrito de funções, a saber, o das funções não negativas definidas. Além disso, há evidências

da não estacionariedade dos erros nas aplicações examinadas anteriormente. Isto é, é possível observar que a autocovariância do processo em dois pontos no tempo não é meramente uma função da diferença dos pontos. Os métodos a serem propostos devem, portanto, prescindir da hipótese de estacionariedade.

O Capítulo 1 é uma breve introdução à teoria de dados funcionais ordinários e agregados, desde seus alicerces probabilísticos, aos resultados mais recentes de estimação de perfis e suas aplicações. A teoria de covariância é desenvolvida separadamente, no Capítulo 2 – no qual é também proposta uma classe mais flexível de funções de covariância do que a dos processos fracamente estacionários, e que possibilita a estimação semiparamétrica de superfícies de covariância.

Nos Capítulos 3 e 4 são propostos estimadores não paramétricos da função de covariância de dados agregados. O Capítulo 3 apresenta o produto tensorial de Splines como ferramenta para estimação e separação das superfícies de covariância típicas. Como as estimativas assim obtidas podem não ser funções de covariância, propõe-se a análise da sua Transformada de Fourier como ferramenta de “correção”, de modo que se tornem funções não negativas definidas. O Capítulo 4 apresenta o Teorema de Karhunen-Loève, que estende a teoria matricial de autovetores e autofunções para curvas aleatórias, e explica como essa ferramenta pode ser empregada para obter estimativas definidas positivas. Sob a hipótese de homogeneidade das autofunções entre os diferentes componentes das curvas agregadas, é desenvolvida uma metodologia de estimação com base neste resultado.

O Capítulo 5 exhibe a aplicação do método de estimação de covariância em dois problemas reais de curvas agregadas. O primeiro é o consumo diário de energia elétrica em Campinas, São Paulo, onde cada residência/estabelecimento representa um dado funcional, mas as curvas só podem ser amostradas nos transformadores. Neste caso observa-se apenas as curvas agregadas, e é de interesse separar e estimar curvas médias e autocovariâncias individuais. A segunda aplicação é na análise de espectroscopia de infravermelho (*Infrared Spectroscopy*), onde substâncias são submetidas a um feixe luminoso e a proporção de luminosidade transmitida (ou *transmitância*, no jargão qui-

miométrico) é anotada, em função da frequência (ou comprimento de onda) do feixe. A lei de Beer-Lambert justifica a adição das absorbâncias (definidas como o $-\log$ da transmitância) obtidas individualmente, o que permite usar a teoria de dados funcionais agregados para estimar a transmitâncias típicas e a autocovariância individual de componentes através da observação da transmitância (e autocovariância) de uma mistura.

Capítulo 1

Estimação de Tipologia para Dados Funcionais Agregados

Este capítulo reproduz algumas definições importantes à teoria de dados funcionais, desenvolvidas em maior detalhe nas referências oferecidas. Também é apresentada a ferramenta não paramétrica que foi adotada neste trabalho para a estimação dos dados funcionais: a expansão dos dados em bases do tipo *B-spline*. Além disso, resultados necessários de trabalhos anteriores em dados funcionais agregados são resumidamente enunciados. Por fim, um esquema de simulação de dados funcionais agregados é delineado, com o propósito de exemplo da aplicação da rotina de estimação de tipologia, e também para servir de base às rotinas de simulação dos capítulos seguintes.

1.1 Definição de Dado Funcional

Em Ferraty & Vieu (2006), define-se um objeto aleatório $\mathcal{X}$ como variável funcional se ela toma valores em um espaço de dimensão infinita (ou espaço funcional) $\mathcal{T}$. Uma observação x de $\mathcal{X}$ é chamada de *dado funcional*. Essa definição é bastante geral; podemos considerar apenas um caso particular, encontrado em Yaglom (1987), onde a variável funcional é simplesmente uma curva aleatória $Y(t)$, $t \in [0, T]$, e é considerada

especificada se existem funções de distribuição acumulada para cada $t_1 \in [0, T]$

$$F_{t_1}(y_1) = \mathbb{P}(Y(t_1) \leq y_1),$$

e para cada n -upla de elementos $t_1, t_2, \dots, t_n$ em $[0, T]^n$

$$F_{t_1, t_2, \dots, t_n}(y_1, y_2, \dots, y_n) = \mathbb{P}(Y(t_1) \leq y_1, \dots, Y(t_n) \leq y_n),$$

para sucessivos n finitos. Essa definição “hierárquica” é baseada na teoria de probabilidade multivariada. Note que ela não é uma extensão para o caso de dimensão infinita, o que exige mais cuidado, mas é suficiente para, por exemplo, definir os momentos de ordem finita (Yaglom 1987, p. 38). Neste caso o interesse é principalmente em $n = 2$, pois a dissertação tratará da estimação de superfícies de covariância, definidas como uma integral com respeito a $dF_{t_1, t_2}(s, t)$. Isto será visto mais detalhadamente no Capítulo 2.

O texto de Ferraty & Vieu (2006) ainda faz uma interessante distinção: Seja ϕ uma função que depende da distribuição de $\mathcal{X}$ (como, por exemplo, $\phi = \mathbb{E}(\mathcal{X})$). Um modelo para a estimação de ϕ é alguma restrição do tipo $\phi \in \mathcal{C}$. Esse modelo é chamado *paramétrico* se $\mathcal{C}$ é indexado por um número finito de elementos em $\mathbb{R}$. Caso contrário, o modelo é chamado *não paramétrico*.

Os métodos paramétricos, como por exemplo modelos lineares, não são adequados à análise de dados funcionais pois exigem hipóteses sobre a forma das curvas. A alta correlação entre as observações de dados funcionais também prejudica o uso de técnicas do tipo de regressão, causando problemas de multicolinearidade. Mas esse problema é controlado com a introdução da hipótese de *suavidade* das estimativas não paramétricas.

Há muitas ferramentas na literatura que permitem a estimação não paramétrica de dados funcionais. Por exemplo, *splines*, *wavelets* ou *smoothing kernels*. O texto de Ramsay & Silverman (2005) apresenta uma breve introdução aos dois últimos, sendo seu foco é essencialmente em splines. Uma leitura mais profunda dos aspectos matemáticos dos splines encontra-se em de Boor (2001). Também se empregará splines nos procedimentos de estimação não paramétricos descritos nesta dissertação, uma vez que eles possibilitam a separação de curvas agregadas, como será mostrado na Seção 1.4.

1.2 B-Splines

Um spline (univariado) é uma função $B : [a, b] \rightarrow \mathbb{R}$, onde o domínio $[a, b]$ está dividido em uma sequência de subintervalos dados pelos pontos $a = \tau_1 \leq \tau_2 \leq \dots \leq \tau_L = b$. Os pontos τ_ℓ são chamados *nós*. Os nós diferentes de a e b são chamados *nós internos*. Em cada subintervalo gerado pelos nós, o spline é um polinômio de ordem m , isto é, um polinômio cujo número de coeficientes necessários para determiná-lo é igual a m , e seu maior grau é igual a $m - 2$. Além disso, se $m > 1$, em cada nó interno o spline ainda é suave, com pelo menos $m - 2$ derivadas contínuas nesses pontos.

Uma propriedade dos splines que é herdada dos polinômios é que a combinação linear de splines de ordem m também é um spline de ordem m . Essa propriedade motiva a construção das bases *B-splines*, splines tais que qualquer outra função spline possa ser escrita como combinação linear deles. Dois algoritmos de construção de B-splines são descritos na Seção A do Apêndice. A Figura 1.1 mostra 11 bases com nós equidistantes em $[0, 10]$. As linhas pontilhadas representam a localização dos nós internos.

O modelo da expansão de uma variável funcional em B-splines, dado um conjunto de nós $0 = \tau_1 \leq \tau_2 \leq \dots \leq \tau_{K-m+2} = T$ no intervalo $[0, T]$, é dado por

$$Y(t) = \sum_{k=1}^K \beta_k B_k(t) + \varepsilon(t)$$

onde $\varepsilon(t)$ é algum erro de aproximação. Assume-se que $\mathbb{E}(\varepsilon(t)) = 0$. Por estimação do dado funcional, pode-se entender a estimação da função média do dado funcional, pois

$$\mathbb{E}(Y(t)) = \mathbb{E}\left(\sum_{k=1}^K \beta_k B_k(t)\right) + \mathbb{E}(\varepsilon(t)) = \sum_{k=1}^K \beta_k B_k(t).$$

O estimador do dado funcional $\hat{Y}(t)$ coincide com o estimador de seu valor esperado, ou seja,

$$\hat{Y}(t) = \sum_{k=1}^K \hat{\beta}_k B_k(t) = \widehat{\mathbb{E}(Y(t))}.$$

O problema de estimação de $\mathbb{E}(Y(t))$ reduz-se agora a estimar os K parâmetros β_k . Usualmente o estimador de β é aquele que minimiza a soma dos quadrados do erro de

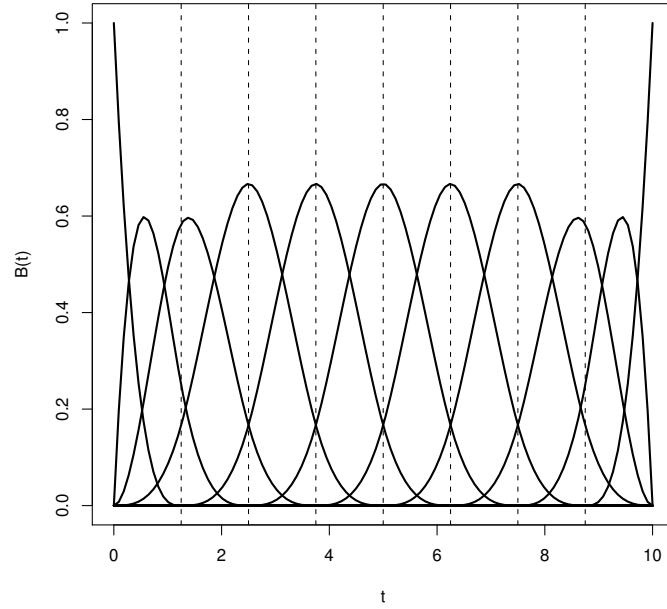


Figura 1.1: 11 bases B-spline em $[0, 10]$.

previsão nos pontos observados, isto é,

$$\hat{\boldsymbol{\beta}} = \arg \min_{\boldsymbol{\beta}} \sum_{r=1}^n \left(y(t_r) - \sum_{k=1}^K \beta_k B_k(t_r) \right)^2, \quad (1.1)$$

e que tem uma forma explícita dada por

$$\hat{\boldsymbol{\beta}} = (\mathbf{B}'\mathbf{B})^{-1} \mathbf{B}'\mathbf{y}.$$

onde $\mathbf{B}$ é a matriz $n \times K$ em que cada coluna corresponde à base B_k calculada nos pontos t_r , $r = 1, 2, \dots, n$. Mais rigorosamente, K também é uma quantidade estimável, bem como a alocação dos nós. Mais detalhes sobre a teoria de estimação de dados funcionais podem ser encontrados em Ramsay & Silverman (2002), Ramsay & Silverman (2005) ou em de Boor (2001). Os artigos de Dias (2000) e Dias & Garcia (2007) propõem ferramentas para a estimação de K . Aspectos computacionais e aplicações são detalhados em Ramsay, Hooker & Graves (2009). A Seção 3.2 retomará a análise desse problema, no caso bivariado.

B-splines desfrutam de algumas propriedades numéricas que devem ser destacadas, em comparação com outros métodos não paramétricos. Eles tem *suporte compacto*, isto é, cada base é diferente de zero (e positiva) apenas em m intervalos adjacentes. Eles são suaves (com derivadas contínuas até ordem $m - 2$), e a estimação pode ser feita por um algoritmo de mínimos quadrados generalizados. A propriedade de suporte compacto implica em uma estrutura de banda na matriz de produtos internos empregada para estimar β por mínimos quadrados, o que gera uma estabilidade computacional similar àquela encontrada em bases ortogonais (Ramsay & Silverman 2005, p. 50).

1.3 Dados Funcionais Agregados

Os dados funcionais são chamados *agregados* quando as curvas observadas são combinações lineares de curvas sobre as quais deseja-se fazer inferência, sendo impossível amostrá-las individualmente. Cada curva agregada será denotada por $Y_i(t)$, $i = 1, 2, \dots, I$. Define-se por *subpopulação* a família de curvas individuais que compartilham o mesmo perfil de curva média e covariância, distinto daquele das outras subpopulações. Cada curva agregada $Y_i(t)$ pode portanto ser escrita como

$$Y_i(t) = \sum_{c=1}^C q_{i,c} W_{i,c}(t), \quad (1.2)$$

em que $q_{i,c}$ são coeficientes da combinação linear, e a curva $W_{i,c}(t)$ representa a soma das $N_{i,c}$ curvas aleatórias individuais $W_{n_c,c}(t)$ que compõem a subpopulação c , isto é,

$$W_{i,c}(t) = \sum_{n_c=1}^{N_{i,c}} W_{n_c,c}(t).$$

Os coeficientes $q_{i,c}$ e $N_{i,c}$ serão considerados constantes e conhecidos. O perfil típico de um indivíduo $W_{n_c,c}(t)$, dado pela sua curva esperada, é chamado de *tipologia* e denotado por $\alpha_c(t)$. Isto é,

$$\mathbb{E}(W_{n_c,c}(t)) = \alpha_c(t).$$

O indivíduo $W_{n_c,c}(t)$ também possui uma função de covariância típica, denotada por

$$\text{Cov}(W_{n_c,c}(s), W_{n_c,c}(t)) = \sigma_c(s, t).$$

A esperança da curva agregada é a combinação linear das tipologias, isto é,

$$\mathbb{E}(Y_i(t)) = \sum_{c=1}^C q_{i,c} \left(\sum_{n_c=1}^{N_{i,c}} \mathbb{E}(W_{n_c,c}(t)) \right) = \sum_{c=1}^C q_{i,c} N_{i,c} \alpha_c(t).$$

De modo análogo, a função de covariância é dada por

$$\text{Cov}(Y_i(s), Y_i(t)) = \sum_{c=1}^C q_{i,c}^2 N_{i,c} \sigma_c(s, t).$$

A equação (1.2) é uma forma bastante geral para escrever as curvas agregadas. Ao longo da dissertação serão considerados dois cenários particulares: o primeiro consiste em que a curva agregada seja a combinação linear de curvas individuais (isto é, $N_{i,c} = 1$); o segundo é tal que a cada subpopulação é dado o mesmo peso unitário (ou seja, $q_{i,c} = 1$). Cada um desses casos está ligado a uma aplicação – o primeiro ao cenário de quimiometria, o segundo ao de consumo elétrico – como será mostrado no Capítulo 5.

Como as curvas individuais $W_{n_c,c}(t)$ tem função de covariância finita, então cada cenário é dado por

$$Y_i(t) = \sum_{c=1}^C q_{i,c} \alpha_c(t) + \sum_{c=1}^C q_{i,c} \varepsilon_{i,c}(t), \quad t \in [0, T], \quad (1.3)$$

para o caso de combinações lineares, ou

$$Y_i(t) = \sum_{c=1}^C N_{i,c} \alpha_c(t) + \sum_{c=1}^C \sum_{n_c=1}^{N_{i,c}} \varepsilon_{i,c,n_c}(t), \quad t \in [0, T], \quad (1.4)$$

para o caso de somas de indivíduos da mesma subpopulação. Aqui $\varepsilon_{i,c,n_c}(t)$ é um erro aleatório com média zero e função de covariância

$$\text{Cov}(\varepsilon_{i,c,n_c}(s), \varepsilon_{i,c,n_c}(t)) = \sigma_c(s, t).$$

Na circunstância de replicações dos dados, indexadas por $j = 1, 2, \dots, J$, assume-se ainda que as curvas não dependem da replicação j , isto é, são observadas de maneira independente.

1.4 Estimação da Tipologia

A separação das tipologias *agregadas* é possível durante o processo de estimação dos coeficientes $\beta_{c,k}$. Para transformar o problema de estimar $\alpha_c(t)$ em um problema de regressão linear, note primeiro que assumimos que $\alpha_c(t)$ tem uma expansão em bases B-spline dada pela forma

$$\alpha_c(t) = \sum_{k=1}^K \beta_{c,k} B_k(t).$$

Os dados são observados nos instantes $t_1, t_2, \dots, t_n$. Usando a notação matricial, temos que a j -ésima replicação do vetor de observações da curva agregada $Y_i(t)$, nos instantes $t_1, t_2, \dots, t_n$, é dado por

$$\mathbf{y}_{ij} = \begin{pmatrix} y_{ij}(t_1) \\ y_{ij}(t_2) \\ \vdots \\ y_{ij}(t_n) \end{pmatrix}_{n \times 1}$$

Para criar a matriz $\mathbf{X}$ de delineamento de regressão, serão usados apenas os coeficientes $N_{i,c}$, mas neste caso de estimação de tipologia o procedimento é idêntico com os coeficientes $q_{i,c}$ presentes. Primeiramente constrói-se uma matriz auxiliar $\mathbf{B}_i$, dada por

$$\mathbf{B}_i = \begin{pmatrix} N_{i,1}B_1(t_1) \cdots N_{i,1}B_K(t_1) \cdots N_{i,C}B_1(t_1) \cdots N_{i,C}B_K(t_1) \\ \vdots \\ N_{i,1}B_1(t_n) \cdots N_{i,1}B_K(t_n) \cdots N_{i,C}B_1(t_n) \cdots N_{i,C}B_K(t_n) \end{pmatrix}_{n \times CK} \quad (1.5)$$

A matriz obtida em (1.5) é repetida para cada replicação J , concatenando-se cópias de $\mathbf{B}_i$, uma abaixo da outra, J vezes. Isso gera uma segunda matriz auxiliar, dada por $\mathbf{B}_i^*$, na equação (1.6).

$$\mathbf{B}_i^* = \begin{pmatrix} \mathbf{B}_i \\ \vdots \\ \mathbf{B}_i \end{pmatrix}_{Jn \times CK} \quad (1.6)$$

De forma análoga, cada observação j da curva agregada $\mathbf{y}_{ij}$ será concatenada com a observação anterior, gerando vetores $\mathbf{y}_i$ de dimensão $Jn \times 1$. Finalmente escreve-se

$\mathbf{X}$ como na equação (1.7). Os vetores de observações $\mathbf{y}_i$ são igualmente concatenados em i , num único vetor $\mathbf{y}$ de dimensão $IJn \times 1$.

$$\mathbf{X} = \begin{pmatrix} \mathbf{B}_1^* \\ \vdots \\ \mathbf{B}_I^* \end{pmatrix}_{IJn \times CK} \quad (1.7)$$

Empregando a matriz $\mathbf{X}$ definida por (1.7), pode-se definir o problema de estimação de $\boldsymbol{\beta}$ como um problema de regressão linear da forma

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}.$$

Note ainda que a maneira que $\mathbf{X}$ é construída é simplesmente $\mathbf{X} = \mathbf{N} \otimes \mathbf{B}^*$, onde $\otimes$ denota o produto matricial de Kronecker e $\mathbf{B}^*$ é a matriz de delineamento de B-splines, replicada J vezes, pelo procedimento de concatenação de $\mathbf{B}$ repetida. Uma vantagem de introduzir a notação do produto de Kronecker é que ela torna evidente que $\mathbf{N}$ deve ser de posto completo. Isso significa que é preciso um número de curvas agregadas superior ou igual ao número de subpopulações ($I \geq C$) e que, além disso, os vetores $N_{\cdot,i}$ com o número de indivíduos de cada subpopulação c na curva $Y_i(t) = N'_{\cdot,i} \mathbf{W}(t)$ sejam linearmente independentes entre si.

A estimação é feita usando o método de mínimos quadrados generalizados, através de um procedimento iterativo. Cada dado agregado será tratado como vetor de dimensão $n \times 1$, dado por $\mathbf{y}_{ij}$. $\mathbf{y}$ representa todos os vetores $\mathbf{y}_{ij}$ concatenados primeiramente em j , depois em i .

Passo 1: Estima-se $\hat{\boldsymbol{\beta}}^{(0)}$ através de mínimos quadrados ordinários, isto é,

$$\hat{\boldsymbol{\beta}}^{(0)} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}$$

Note que o vetor $\hat{\boldsymbol{\beta}}^{(0)}$ é a concatenação de vetores $\hat{\boldsymbol{\beta}}_c^{(0)}$, ou seja,

$$\hat{\boldsymbol{\beta}}^{(0)} = \begin{pmatrix} \hat{\boldsymbol{\beta}}_1^{(0)} \\ \vdots \\ \hat{\boldsymbol{\beta}}_C^{(0)} \end{pmatrix}$$

Passo 2: Na notaco usual de regresso, $\hat{\mathbf{y}}^{(0)} = \mathbf{X}\hat{\boldsymbol{\beta}}^{(0)}$ e $\mathbf{r}^{(0)} = \mathbf{y} - \hat{\mathbf{y}}^{(0)}$. Note agora que $\hat{\mathbf{y}}^{(0)}$ é dado por

$$\hat{\mathbf{y}}^{(0)} = \begin{pmatrix} \hat{\mathbf{y}}_1^{(0)} \\ \vdots \\ \hat{\mathbf{y}}_1^{(0)} \\ \hat{\mathbf{y}}_2^{(0)} \\ \vdots \\ \hat{\mathbf{y}}_I^{(0)} \end{pmatrix}_{IJn \times 1}$$

ou seja, o vetor de valores preditos $\hat{\mathbf{y}}^{(0)}$ é composto por $\hat{\mathbf{y}}_i^{(0)}$ repetidos J vezes, e aí concatenados para $i = 1, 2, \dots, I$.

A estimativa da matriz de covarincia de cada curva agregada vetorizada é dada por

$$\hat{\mathbf{Z}}_i^{(0)} = \frac{1}{J} \sum_{j=1}^J \hat{\mathbf{r}}_{ij}^{(0)} \hat{\mathbf{r}}_{ij}^{(0)'}.$$

A matriz de covarincia dos dados concatenados, $\hat{\mathbf{Z}}^{(0)}$, é bloco diagonal. Primeiramente considera-se bloco diagonais com repeties da matriz $\hat{\mathbf{Z}}_i^{(0)}$, J vezes. Essa matriz $\hat{\mathbf{Z}}_i^{*(0)}$ é dada por $\hat{\mathbf{Z}}_i^{*(0)} = \mathbf{I}_{J \times J} \otimes \hat{\mathbf{Z}}_i^{(0)}$ ($\mathbf{I}_{J \times J}$ denota a matriz identidade de dimenses J e J , $\otimes$ é o produto de Kronecker) ou explicitamente por

$$\hat{\mathbf{Z}}_i^{*(0)} = \begin{pmatrix} \hat{\mathbf{Z}}_i^{(0)} & & \\ & \ddots & \\ & & \hat{\mathbf{Z}}_i^{(0)} \end{pmatrix}_{Jn \times Jn} \quad (1.8)$$

A matriz de covarincia dos dados é construda de forma similar, e dada por

$$\hat{\mathbf{Z}}^{(0)} = \begin{pmatrix} \hat{\mathbf{Z}}_1^{*(0)} & & & \\ & \hat{\mathbf{Z}}_2^{*(0)} & & \\ & & \ddots & \\ & & & \hat{\mathbf{Z}}_I^{*(0)} \end{pmatrix}_{IJn \times IJn} \quad (1.9)$$

Passo 3: Fixe $\ell = 0$. Enquanto $|\hat{\mathbf{Z}}^{(\ell)}(t_{r_1}, t_{r_2}) - \hat{\mathbf{Z}}^{(\ell-1)}(t_{r_1}, t_{r_2})| \geq \varepsilon, \forall t_{r_1}, t_{r_2} = 1, 2, \dots, IJn$,

- $\ell \leftarrow \ell + 1$;
- $\hat{\boldsymbol{\beta}}^{(\ell)} = \left(\mathbf{X}' (\mathbf{Z}^{(\ell-1)})^{-1} \mathbf{X} \right)^{-1} \mathbf{X}' (\mathbf{Z}^{(\ell-1)})^{-1} \mathbf{y}$;
- $\hat{\mathbf{Z}}_i^{(\ell)} = J^{-1} \sum_{j=1}^J \hat{\mathbf{r}}_{ij}^{(\ell)} \hat{\mathbf{r}}_{ij}^{(\ell)'};$
- $\hat{\mathbf{Z}}^{(\ell)}$ como definida em (1.8) e (1.9), salvo o índice (ℓ) ao invés de (0) .

Um problema deste método está no *Passo 2*: devido à multicolinearidade das observações, as matrizes $\hat{\mathbf{Z}}_i$ são *numericamente singulares*, possuindo muitos autovalores pequenos. Essas matrizes acabam deteriorando a qualidade da estimativa de mínimos quadrados generalizada, às vezes. Uma solução é forçar algum tipo de truncamento em $\hat{\mathbf{Z}}_i$, da forma

$$z^*(t_{r_1}, t_{r_2}) = z(t_{r_1}, t_{r_2}) \left(1 - \frac{|t_{r_2} - t_{r_1}|}{\gamma} \right) \mathbf{1} (|t_{r_2} - t_{r_1}| \leq \gamma). \quad (1.10)$$

A escolha de γ é feita empiricamente. Adotou-se $\gamma = 6 \cdot |t_{r+1} - t_r|$ para um cenário das simulações. Note que o truncamento não é suficiente para produzir matrizes que não sejam numericamente singulares sempre, mas nos casos experimentados diminuiu o número de condição da matriz sensivelmente. Esse tipo de truncamento é comum quando se estima covariância, e está presente no artigo de Hall, Fisher & Hoffman (1994, p. 2118), por exemplo.

1.5 Simulação de Curvas Agregadas

Para ilustrar os procedimentos das seções anteriores, foram simulados dados com as $C = 2$ tipologias exibidas na Figura 1.2. As curvas fictícias foram amostradas em 48 pontos, em intervalos de 0,5 unidades. O cenário corresponde ao caso em que as curvas são agregadas pela soma de $N_{i,c}$ indivíduos de uma mesma subpopulação, com coeficientes $q_{i,c} = 1$.

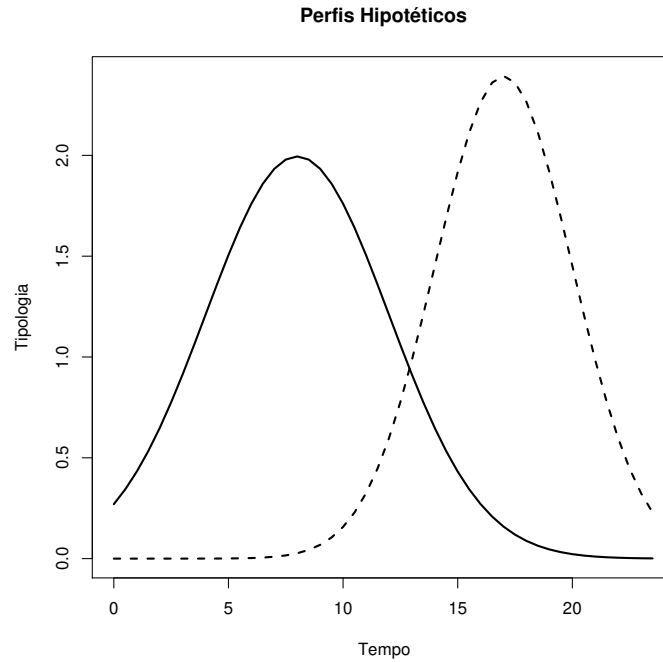


Figura 1.2: Tipologias simuladas.

Foram geradas $I = 3$ curvas agregadas, denotadas por $Y_1(t)$, $Y_2(t)$ e $Y_3(t)$. O número de indivíduos do tipo c na curva i é determinado pela matriz $\mathbf{N}_{3 \times 2}$, dada por

$$\mathbf{N}_{3 \times 2} = \begin{pmatrix} 20 & 10 \\ 15 & 15 \\ 10 & 20 \end{pmatrix}$$

A função de covariância da curva agregada é dada por

$$\text{Cov}(Y_i(s), Y_i(t)) = \sum_{c=1}^C \sum_{n_c=1}^{N_{i,c}} \sigma_c(s, t) = \sum_{c=1}^C N_{i,c} \sigma_c(s, t)$$

Neste exemplo introdutório, adotou-se uma função de covariância conhecida, da forma

$$\sigma_c(s, t) = 0,25 \exp\{-0,5|t - s|\}, \quad c = 1, 2.$$

Será demonstrado que esta função é de covariância no Capítulo 2. É possível notar que na presente forma ela implica em uma variância constante em função do tempo,

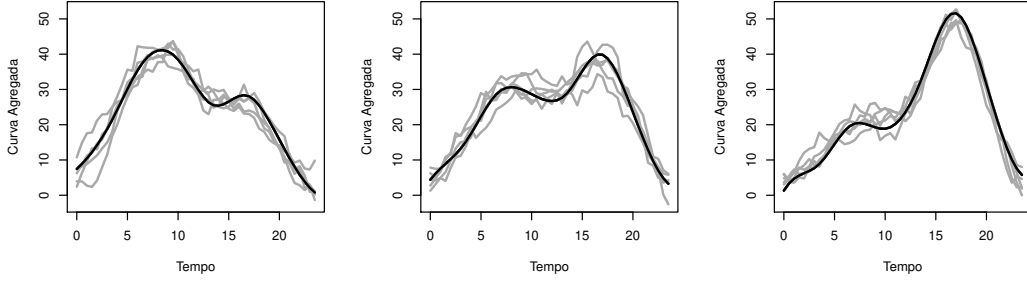


Figura 1.3: $J = 5$ amostras de $y_i(t)$, e as curvas estimadas $\hat{y}_i(t)$

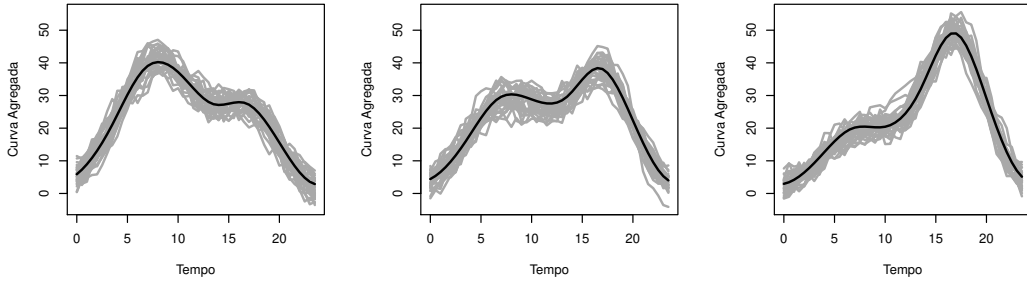


Figura 1.4: $J = 30$ amostras de $y_i(t)$, e as curvas estimadas $\hat{y}_i(t)$

pois $\sigma_c(t, t) = 0,25$. Isso é uma simplificação dos dados reais, mas servirá ao propósito de exemplo.

Após a simulação, as estimativas foram obtidas através de $K = 10$ bases em nós equispaçados. O resultado da realização de $J = 5$ e $J = 30$ replicações, com as curvas esperadas $\hat{y}_i(t)$, $i = 1, 2, 3$ sobrepostas (as linhas mais grossas, em preto), encontra-se nas Figuras 1.3 e 1.4, respectivamente.

O resultado mais interessante é, todavia, a estimativa das tipologias. A Figura 1.5 exhibe as tipologias estimadas, já separadas, sobrepostas às tipologias verdadeiras (linhas sólidas), para os dois tipos de subpopulações. As linhas tracejadas são estimativas com 5 replicações, enquanto as linhas pontilhadas com 30 replicações. É possível observar que o aumento de replicações produziu estimativas mais precisas, embora mesmo para $J = 5$ as estimativas sejam bastante similares às curvas originais, capturando suas características mais acentuadas e efetivamente separando as tipologias. O efeito do

aumento de replicações, bem como uma análise de diferentes configurações da matriz $\mathbf{N}$, é exibido em Martarelli (2006) e Dias et al. (2009).

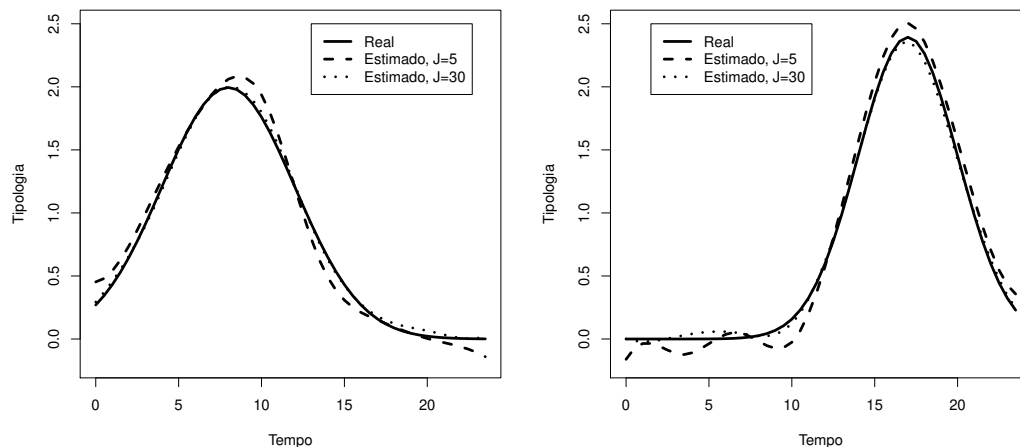


Figura 1.5: Tipologias estimadas, com diferentes números de replicações.

Como quantificar a incerteza sobre essas estimativas? Seguindo Ramsay & Silverman (2005, p. 104), o intervalo de confiança pontual para $\alpha_c(t)$ é obtido considerando, para cada t *fixado*,

$$\text{Var}(\hat{\alpha}_c(t)) = \mathbf{L}\mathbf{S}\Sigma_c\mathbf{S}'\mathbf{L}', \quad (1.11)$$

onde $\mathbf{L}$ é um vetor $1 \times K$ composto pelas bases B-spline calculadas no instante t , $\mathbf{S} = (\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'$, com $\mathbf{B}$ a matriz de desenho B-spline, e Σ_c a matriz obtida ao se observar $\sigma_c(s, t)$ nos instantes $t_1, \dots, t_n$. O intervalo é pontual no sentido que a confiança do intervalo não é conjunta. Para obter o intervalo de confiança, considera-se simplesmente para o caso de erros Gaussianos

$$IC(\alpha_c(t), (1 - \gamma)100\%) = \hat{\alpha}_c(t) \pm q_{\frac{\gamma}{2}\%} \sqrt{\frac{\text{Var}(\hat{\alpha}_c(t))}{J}}. \quad (1.12)$$

No entanto, não temos estimativas de $\sigma_c(s, t)$ ainda. Resolver essa questão torna-se uma das muitas aplicações das metodologias de estimação de covariância que serão apresentadas nos capítulos seguintes.

Capítulo 2

Função de Covariância de Dados Funcionais Agregados

O Capítulo 2 aborda mais profundamente as propriedades da covariância de uma função aleatória, agregada ou não. A definição formal de covariância é apresentada, bem como a propriedade de *estacionariedade fraca* de um processo estocástico, que depende da covariância. O Teorema de Bochner é enunciado para processos fracamente estacionários, e uma família não estacionária, semiparamétrica de funções de covariância é definida a partir dele. A estimação por mínimos quadrados de covariâncias pertencentes a essa família, para dados agregados, é apresentada e exemplificada por simulação.

2.1 Covariância de uma Função Aleatória

A função de covariância de uma função aleatória $X(t)$, definida no intervalo $\mathcal{T}$, com $\mathbb{E}(X(t)) = 0$ e distribuição caracterizada por $F_{t_1, \dots, t_n}(x_1, \dots, x_n)$ é dada por

$$\text{Cov}(X(s), X(t)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 dF_{s,t}(x_1, x_2) = \sigma(s, t). \quad (2.1)$$

Para que uma função satisfaça a definição de Covariância, isto é, possa ser representada na forma (2.1), é necessário que $\sigma(s, t)$ seja simétrica (ou seja, $\sigma(s, t) = \sigma(t, s)$) e seja um *kernel não negativo definido*. Reproduz-se aqui alguns resultados extraídos

de Bochner (1959), Stewart (1976), Loève (1977), Loève (1978) e Yaglom (1987), com demonstrações adaptadas para casos mais simples que os dos resultados originais.

Teorema 2.1.1 *Um kernel $\mathcal{K}(u, v)$, ou kernel integral, é uma função bivariada que define uma transformada integral, ou seja, um operador T na forma*

$$Tf(u) = \int_a^b \mathcal{K}(u, v) f(v) dv.$$

Um kernel é não negativo definido se para todo qualquer n , qualquer conjunto de pontos $t_1, \dots, t_n$ do conjunto $\mathcal{T}$ e qualquer conjunto de números reais $c_1, \dots, c_n$, vale

$$\sum_{j,k=1}^n \mathcal{K}(t_j, t_k) c_j c_k \geq 0, \quad (2.2)$$

Se não existem $c_1, \dots, c_n$ onde a somatória se anule exceto por $c_1 = \dots = c_n = 0$, então a função é positiva definida. Toda função $\sigma(s, t)$ dada por (2.1) é não negativa definida. O inverso também é verdadeiro: todo kernel não negativo definido em $\mathcal{T}$ é função de covariância de algum processo de média zero em $\mathcal{T}$.

Demonstração: A afirmação de que $\sigma(s, t)$ satisfaz a definição de kernel não negativo definido é direta. Aplicando a definição, temos que

$$\begin{aligned} \sum_{j,k=1}^n \sigma(t_j, t_k) c_j c_k &= \sum_{j,k=1}^n \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} c_j x_1 c_k x_2 dF_{t_j, t_k}(x_1, x_2) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \sum_{j,k=1}^n c_j x_1 c_k x_2 dF_{t_j, t_k}(x_1, x_2) \\ &= \mathbb{E} \left(\sum_{j,k=1}^n c_j X(t_j) c_k X(t_k) \right) \\ &= \mathbb{E} (c_1^2 X(t_1)^2 + 2c_1 c_2 X(t_1) X(t_2) + \dots + c_n^2 X(t_n)^2) \\ &= \mathbb{E} ((c_1 X(t_1) + \dots + c_n X(t_n))^2) \geq 0 \end{aligned}$$

A afirmação recíproca, que todo kernel não negativo definido é uma função de covariância, é demonstrável de maneira mais geral do que aquela que será apresentada, de que todo kernel não negativo definido $\mathcal{K}$ é covariância de um processo Gaussiano. Porém a forma mais fraca de demonstração é preferível por sua relativa simplicidade. A demonstração foi extraída de Loève (1978, p. 132):

Essencialmente pode-se reescrever a desigualdade (2.2) como

$$\sum_{j,k=1}^n \mathcal{K}(t_j, t_k) c_j c_k = \mathbf{c}' \mathbf{K} \mathbf{c} \geq 0, \quad (2.3)$$

onde $\mathbf{c}$ é o vetor com coordenadas $c_1, \dots, c_n$ e a matriz $\mathbf{K}$ tem entradas $(\mathbf{K})_{j,k} = \mathcal{K}(t_j, t_k)$. Como a matriz $\mathbf{K}$ é positiva semi-definida (por definição), é possível reescrever (2.3) como uma forma quadrática

$$R(\tilde{c}) = \sum_{j=1}^n \lambda_j^2 \tilde{c}_j^2, \quad \lambda_j^2 \geq 0,$$

com $\tilde{c}_j$ combinações lineares de $c_1, \dots, c_n$. Neste caso,

$$\exp\{-R(\tilde{c})/2\} = \prod_{j=1}^n \exp\{-\lambda_j^2 \tilde{c}_j^2/2\}$$

é a função característica conjunta de n variáveis aleatórias Y_j , independentes e Normais, com média 0 e variâncias $\lambda_j^2, j = 1, 2, \dots, n$. Note que a demonstração de Loève admite variáveis aleatórias degeneradas, pois nada impede que $\lambda_j = \lambda_{j+1} = \dots = \lambda_n = 0$ a partir de algum j . Consequentemente, tem-se que

$$R(\tilde{c}) = \sum_{j=1}^n \mathbb{E}(Y_j^2) \tilde{c}_j^2 = \mathbb{E} \left(\sum_{j=1}^n (Y_j \tilde{c}_j)^2 \right).$$

Mas como $Y_j \tilde{c}_j$ são combinações lineares de variáveis aleatórias Normais, então também são Normais. Repetindo a identidade de forma quadrática usada anteriormente, tem-se que

$$\sum_{j=1}^n (Y_j \tilde{c}_j)^2 = \sum_{j,k=1}^n X_j X_k c_j c_k$$

e então $\mathcal{K}(t_j, t_k) = \mathbb{E}(X_j X_k)$, o que prova que o kernel não negativo definido é uma função de covariância. ■

A média e a covariância são empregadas para classificar um processo estocástico como *fracamente estacionário*. A definição de um processo fracamente estacionário é como segue:

Definição 2.1.2 *Seja $X(t)$ um processo estocástico a tempo contínuo. O processo é fracamente estacionário se*

1. $\mathbb{E}(X(t)) = m$, uma constante real.
2. $\text{Cov}(X(s), X(t))$ é uma função somente de $|t - s|$, isto é, $\text{Cov}(X(s), X(t)) = \sigma^2 \rho(t - s)$, ρ uma função definida positiva.

Uma consequência importante é que $\text{Var}(X(t)) = \text{Cov}(X(t), X(t)) = \sigma^2 \rho(0)$, uma função constante (positiva). Note que o inverso não é verdadeiro: um processo não estacionário pode ter variância constante.

Existe uma conexão entre estacionariedade fraca e a transformada de Fourier da função de covariância, dada pelo Teorema de Bochner. Esse resultado é muito conveniente para verificar se uma função $\rho(t - s)$ é de covariância.

Teorema 2.1.3 (Bochner) *Seja $X(t)$ um processo fracamente estacionário em $\mathcal{T} = \mathbb{R}$, ou seja, $\mathbb{E}(X(t)) = m$ e $\text{Cov}(X(s), X(t)) = \sigma^2 \rho(t - s)$. Então $\mathcal{F}\{\rho\} = P$, a Transformada de Fourier de ρ , é real, contínua e positiva. O inverso também é verdadeiro, isto é, se P é uma função real, contínua e positiva, então $\rho = \mathcal{F}^{-1}\{P\}$ é uma função de covariância.*

Demonstração: A prova foi extraída de Loève (1977, p. 220). Define-se a transformada de Fourier de uma função por

$$\mathcal{F}(w) := \int_{-\infty}^{\infty} e^{-iwt} f(t) dt$$

e sua transformada inversa é dada por

$$f(t) := \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{iwt} \mathcal{F}(w) dw.$$

Suponha primeiro que a transformada de Fourier de ρ seja contínua, real e positiva. Então ela é função característica para alguma medida μ_ρ (não necessariamente de probabilidade), e $\rho(t)$ é sua densidade. Daí, aplicando a somatória em j, k da definição de kernels não negativos definidos, tem-se que

$$\begin{aligned} \sum_{j,k=1}^n P(\omega_k - \omega_j) c_j c_k &= \frac{1}{2\pi} \sum_{j,k=1}^n \int_{-\infty}^{\infty} \exp\{it(\omega_k - \omega_j)\} c_j c_k d\mu_\rho \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \left(\sum_{j,k=1}^n \exp\{it(\omega_k - \omega_j)\} c_j c_k \right) \rho(t) dt \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \left| \sum_{k=1}^n 2 \cos\{t(\omega_k)\} c_k \right|^2 \rho(t) dt \geq 0 \end{aligned}$$

onde a última desigualdade é válida pois para cada par (j, k) e (k, j) ,

$$\exp\{it(\omega_k - \omega_j)\} c_j c_k + \exp\{it(\omega_j - \omega_k)\} c_k c_j = 2 \cos\{t(\omega_k - \omega_j)\} c_j c_k,$$

e o fato do quadrado da somatória ser igual à somatória dupla em j, k é verificável expandindo os termos do quadrado. Como isso implica que ρ é uma função não negativa definida, então pelo Teorema 2.1.1 ela é função de covariância.

Para demonstrar o outro sentido do teorema, suponha agora que ρ é função de covariância. Para cada $T > 0$ e $\omega \in \mathbb{R}$ vale

$$p_T(\omega) = \frac{1}{T} \int_0^T \int_0^T \rho(v - u) e^{-i(u-v)\omega} du dv \geq 0.$$

Para mostrar que a função é real (e logo poder fazer afirmações sobre ela ser ou não maior ou igual a zero), basta tomar a identidade de Euler e obter

$$p_T(\omega) = \frac{1}{T} \int_0^T \int_0^T \rho(v - u) \cos((u - v)\omega) du dv - i \frac{1}{T} \int_0^T \int_0^T \rho(v - u) \sin((u - v)\omega) du dv.$$

A partir daí, o termo imaginário é reescrito, tomando a lei do seno da soma de ângulos, como

$$-\operatorname{Im}(p_T(\omega)) = \frac{1}{T} \int_0^T \int_0^T \rho(v - u) (\sin(u\omega) \cos(v\omega) - \cos(u\omega) \sin(v\omega)) du dv.$$

Aproveitando a simetria de ρ em torno de zero, tem-se que

$$\begin{aligned} -\operatorname{Im}(p_T(\omega)) &= \frac{1}{T} \int_0^T \int_0^T \rho(v-u) \operatorname{sen}(u\omega) \cos(v\omega) du dv \\ &\quad - \frac{1}{T} \int_0^T \int_0^T \rho(u-v) \operatorname{sen}(v\omega) \cos(u\omega) du dv, \end{aligned}$$

donde conclui-se que $\operatorname{Im}(p_T(\omega)) = 0$. A não negatividade da função vem do fato que o integrando ρ é por hipótese não negativo definido e contínuo, logo pode ser escrito como o limite de somas não negativas de Riemann.

Seja $u = v + t$, então

$$\begin{aligned} p_T(\omega) &= \frac{1}{T} \int_0^T \int_{-v}^{T-v} \rho(v-v+t) e^{-i(v+t-v)\omega} dt dv \\ &= \frac{1}{T} \int_{-T}^T \int_{\max\{0, -t\}}^{\min\{T, T-t\}} \rho(t) e^{-it\omega} dv dt \\ &= \frac{1}{T} \int_{-T}^T (\min\{T, T-t\} - \max\{0, -t\}) \rho(t) e^{-it\omega} dt \\ &= \begin{cases} \frac{1}{T} \int_{-T}^T (T-t) \rho(t) e^{-it\omega} dt & \text{se } t < 0 \\ \frac{1}{T} \int_{-T}^T (T-t) \rho(t) e^{-it\omega} dt & \text{se } t \geq 0 \end{cases} \end{aligned}$$

Com a coincidência dos dois casos, tem-se então simplesmente que

$$p_T(\omega) = \int_{-T}^T \left(1 - \frac{t}{T}\right) \rho(t) e^{-it\omega} dt \geq 0$$

Defina $g_T(t)$ como

$$g_T(t) = \left(1 - \frac{t}{T}\right) \rho(t) \mathbf{1}_{[-T, T]}(t).$$

Temos então que $g_T(t) \rightarrow \rho(t)$ quando $T \rightarrow \infty$, e

$$p_T(\omega) = \int_{-\infty}^{\infty} e^{-it\omega} g_T(t) dt \geq 0.$$

Agora, multiplica-se os dois lados da equação por

$$\frac{1}{2\pi} \left(1 - \frac{|\omega|}{\Omega}\right) e^{iu\omega}$$

e integrando em $[-\Omega, \Omega]$ com respeito a $d\omega$, tem-se

$$\int_{-\Omega}^{\Omega} \frac{1}{2\pi} \left(1 - \frac{|\omega|}{\Omega}\right) e^{iu\omega} p_T(\omega) d\omega = \int_{-\Omega}^{\Omega} \frac{1}{2\pi} \left(1 - \frac{|\omega|}{\Omega}\right) e^{iu\omega} \int_{-\infty}^{\infty} e^{-it\omega} g_T(t) dt d\omega,$$

ou simplesmente

$$\frac{1}{2\pi} \int_{-\Omega}^{\Omega} e^{iu\omega} \left(1 - \frac{|\omega|}{\Omega}\right) p_T(\omega) d\omega = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\Omega}^{\Omega} \left(1 - \frac{|\omega|}{\Omega}\right) e^{i(u-t)\omega} d\omega g_T(t) dt. \quad (2.4)$$

A integral à direita pode ser simplificada pois

$$\int_{-\Omega}^{\Omega} \left(1 - \frac{|\omega|}{\Omega}\right) e^{i(u-t)\omega} d\omega = \frac{\sin^2(2^{-1}\Omega(t-s))}{(2^{-1}\Omega(t-s))^2}.$$

O lado esquerdo da equação (2.4) é proporcional à função característica de uma variável aleatória com distribuição proporcional a $(1 - \Omega^{-1}|\omega|) p_T(\omega)$ em $[-\Omega, \Omega]$ (pois $p_T(\omega)$ é uma função não negativa). Já o termo à direita, quando $\Omega \rightarrow \infty$, converge a

$$\int_{-\infty}^{\infty} \delta(u-t) g_T(t) dt,$$

em que δ é a distribuição delta de Dirac, e logo a integral toda é simplesmente $g_T(t)$ (a convolução de uma função com o delta de Dirac é a própria função. Veja Bracewell (1986, p. 75)).

Com isso, temos que $g_T(u)$ é o limite de uma sequência de funções características simétricas em 0. Pelo teorema de continuidade de Lévy, ela é uma função característica também. Como $g_T(t) \rightarrow \rho(t)$, então $\rho(t)$ é uma função característica de alguma variável aleatória simétrica em 0. Ou seja, que sua transformada de Fourier é proporcional à densidade desta variável aleatória e, logo, real e não negativa em toda parte. ■

Com base nisso, pode-se construir uma classe de funções de covariância não estacionárias, mais interessantes para a modelagem de fenômenos reais. Tome $\rho(s, t) = e^{-\phi|t-s|}$. Essa função é claramente de correlação pois, pelo teorema de Bochner, qualquer $\rho(\tau)$ função característica de uma variável aleatória simétrica em $\tau = 0$, onde $\tau = t - s$,

é uma função não negativa definida (neste caso, $e^{-\phi|t-s|}$ é função característica de uma variável aleatória com distribuição de Cauchy, centrada no zero). O processo $X(t)$ com média zero e covariância $\sigma(s, t) = \sigma^2 e^{-\phi|t-s|}$ é um processo fracamente estacionário.

Para definir uma verossimilhança, assume-se que $X(t)$ é um Processo Gaussiano. O seguinte teorema é crucial e define a classe de processos não estacionários que serão empregados como modelo para a estimação da covariância de curvas agregadas:

Teorema 2.1.4 *Seja $\eta(t)$ uma função suave no intervalo $[0, T]$ e que é igual a zero fora dele. Então $\eta(t)X(t)$ é um Processo Gaussiano em $[0, T]$ com média zero e covariância $\eta(s)\eta(t)e^{-\phi|t-s|}$.*

Demonstração: A demonstração é imediata, pois para todo par (s, t) fixado,

$$\begin{aligned} \text{Cov}(\eta(s)X(s), \eta(t)X(t)) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \eta(s)x_1\eta(t)x_2 dF_{s,t}(x_1, x_2) \\ &= \eta(s)\eta(t) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1x_2 dF_{s,t}(x_1, x_2) \\ &= \eta(s)\eta(t) \text{Cov}(X(s), X(t)) \quad \blacksquare \end{aligned}$$

A condição de suporte compacto pode ser trocada pela que $|\eta(t)|$ seja integrável em $\mathbb{R}$. Esse modelo é não estacionário (desde que $\eta(t)$ seja não constante), e embora seja uma semiparametrização da função de covariância, $\eta(t)$ ainda pode ser estimada de maneira não paramétrica, o que dá flexibilidade ao analista. Essa forma de covariância foi proposta em Dias, Garcia & Schmidt (2011).

2.2 Covariância Semiparametrizada de Dados Funcionais Agregados

O modelo considerado nesta seção é o da soma de $N_{i,c}$ indivíduos $W_{n_c,c}(t)$, e $q_{i,c} = 1$. A adaptação para o modelo mais geral, com coeficientes $q_{i,c}$ é imediata. Em notação de

inferência, o modelo para a curva agregada Y_{ij} é

$$Y_{ij}(t) \sim PG\left(\sum_{c=1}^C N_{i,c} \alpha_c(t), \mathbf{Z}_i(s, t)\right) \quad (2.5)$$

para $i = 1, \dots, I$ indexando a i -ésima curva agregada, $j = 1, \dots, J$ a sua replicação, $N_{i,c}$ o número de curvas individuais de perfil $\alpha_c(t)$ na curva agregada i , $c = 1, \dots, C$ as diferentes subpopulações e $t \in [0, T]$. PG é a notação para um *Processo Gaussiano*. Temos ainda que

$$\mathbf{Z}_i(s, t) = \text{Cov}(Y_{ij}(s), Y_{ij}(t)) = \sum_{c=1}^C \sum_{n_c=1}^{N_{i,c}} \text{Cov}(\varepsilon_{i,c,n_c}(s), \varepsilon_{i,c,n_c}(t)) = \sum_{c=1}^C N_{i,c} \sigma_c(t, s)$$

é a função de covariância da curva agregada, com $\sigma_c(t, s)$ correspondendo à covariância de cada curva individual, e possuindo a forma proposta em Dias et al. (2011), dada por

$$\sigma_c(s, t) = \eta_c(t) \eta_c(s) e^{-\phi_c |t-s|}. \quad (2.6)$$

Na circunstância em que as curvas são agregadas em combinações lineares ao invés de representarem somas de curvas individuais, o modelo dado em (2.5) permanece inalterado (exceto pelos tamanhos da subpopulações $N_{i,c}$ serem substituídos pelos pesos $q_{i,c}$ dos componentes $W_c(t)$), mas a covariância $\mathbf{Z}_i(s, t)$ é corretamente representada por

$$\mathbf{Z}_i(s, t) = \sum_{c=1}^C q_{i,c}^2 \sigma_c(s, t).$$

Como o processo da distribuição dos erros foi definido (é um Processo Gaussiano), pode-se estimar $\eta_c(\mathbf{t})$ e ϕ_c (com os pontos t fixados) através do método de máxima verossimilhança, com função de verossimilhança dada por

$$\mathcal{L}(\boldsymbol{\eta}(\mathbf{t}), \boldsymbol{\phi} \mid \mathbf{r}_{1,1}, \dots, \mathbf{r}_{I,J}) = \prod_{i=1}^I \frac{1}{(2\pi)^{Jn/2} |\mathbf{Z}_i|^{J/2}} e^{-\frac{1}{2} \left(\sum_{j=1}^J \mathbf{r}'_{i,j} \mathbf{Z}_i^{-1} \mathbf{r}_{i,j} \right)}.$$

Não é fácil, contudo, determinar uma solução analítica para o máximo – não há forma explícita para a matriz inversa da covariância agregada, o que exige inversão

numérica e dificulta encontrar o gradiente da função de verossimilhança. Deve-se recorrer à otimização numérica do tipo de Nelder-Mead. A rotina é sensível aos pontos iniciais devido à alta multicolinearidade das matrizes $\mathbf{Z}_i$ (onde cada entrada $(\mathbf{z}_i)_{s_r, t_r}$ é a função $\mathbf{Z}_i(s, t)$ calculada em (s_r, t_r)), e ocasionalmente a estimativa é numericamente singular para determinada iteração, o que exige um truncamento do tipo utilizado para os mínimos quadrados generalizados da estimação de tipologia. A Seção B.1 do Apêndice B discorre sobre algumas ideias para obter estimativas iniciais $\mathbf{Z}_i^{(0)}$, que foram empregadas na aplicação. Contudo, decidiu-se utilizar uma alternativa mais estável, a estimação por mínimos quadrados. As estimativas não são equivalentes, mas não há mais necessidade de trabalhar com matrizes inversas, e o algoritmo ganha muito em estabilidade e velocidade.

2.3 Exemplo Simulado

As curvas simuladas seguem a mesma estrutura das simulações conduzidas na Seção 1.5, exceto pelas seguintes diferenças:

- Os cenários simulados serão $J = 5$, $J = 30$, $J = 100$, $J = 300$ e $J = \infty$. O último caso corresponde a utilizar as funções de covariâncias agregadas verdadeiras $\mathbf{Z}_i(s, t)$, e serve para avaliar quão bem o método separa as covariâncias típicas $\sigma_c(s, t)$.
- A covariância usada nas simulações será $\eta_c(s)\eta_c(t) \exp\{-\phi_c|t-s|\}$. $\eta_c(t)$ foi fixada em $\sqrt{0,5\alpha_c(t)}$, e $\phi_1 = 0,5$, $\phi_2 = 1,5$. Isto servirá para mostrar que o método é capaz de capturar características de não estacionariedade nos dados. O modelo será de contribuições individuais, dado por (1.4), e não (1.3).

As estimativas foram obtidas através de um procedimento numérico de otimização de mínimos quadrados. Isto é, a função objetivo é dada por

$$Q(\boldsymbol{\theta}, \boldsymbol{\phi}) = \sum_{i=1}^I \sum_{s=t_1}^{t_n} \sum_{t=t_1}^{t_n} \left(\mathbf{Z}_i(s, t) - \sum_{c=1}^C N_{i,c} \eta_c(s) \eta_c(t) \exp\{-\phi_c|t-s|\} \right)^2,$$

onde $\eta_c(t)$ é dado pela expansão usual em bases B-spline

$$\eta_c(t) = \sum_{k=1}^K \theta_{c,k} B_k(t)$$

e as estimativas são dadas por

$$(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\phi}}) = \arg \min_{\boldsymbol{\theta}, \boldsymbol{\phi}} Q(\boldsymbol{\theta}, \boldsymbol{\phi}).$$

Foram usadas 10 funções bases com nós internos equispaçados. A rotina numérica de otimização usou o algoritmo de Nelder-Mead, com um limite máximo de 50000 iterações.

Primeiramente, será analisada a qualidade das estimativas de $\sigma_c(t, t) = \eta_c^2(t)$. O estudo de simulações também servirá para mostrar se o truncamento proposto na fórmula (1.10) é realmente necessário para obter-se melhores estimativas.

A Figura 2.1 mostra as funções $\eta_c^2(t)$ estimadas quando o número J de replicações é igual a 5. A linha sólida representa o valor verdadeiro de $\eta_c^2(t)$, $c = 1, 2$. A linha tracejada é $\hat{\eta}_c^2(t)$ obtida utilizando a matriz $\hat{\mathbf{Z}}_i(s, t)$ completa; a linha de traços e pontos alternados é $\hat{\eta}_c^2(t)$ obtida utilizando a matriz $\hat{\mathbf{Z}}_i(s, t)$ truncada, com $\gamma = 6$, pela fórmula (1.10). As Figuras 2.2, 2.3 e 2.4 representam estimativas nas mesmas condições, mas com um número de replicações J igual a 30, 100 e 300, respectivamente.

Pode-se notar que as estimativas não diferem muito, quando há truncamento, se J é grande. O truncamento atenua mas não é capaz de evitar as oscilações quando $\hat{\eta}_c^2(t)$ é próximo de zero, presentes quando t é próximo de 24 para $\hat{\eta}_1^2(t)$, e t é próximo de zero em $\hat{\eta}_2^2(t)$, nas Figuras 2.1, 2.2 e 2.3. De fato, o fenômeno é controlado apenas quando $J = 300$ (Figura 2.4), pois as estimativas da matriz de variância-covariância $\hat{\mathbf{Z}}_i$ já estão muito próximas da verdadeira $\mathbf{Z}_i$. Porém o truncamento produz estimativas mais precisas onde a variância é grande, quando J é pequeno.

As Figuras 2.5 e 2.6 mostram as mesmas curvas, mas agora sobrepostas em função de J , de modo que fique mais claro o ganho quando o número de replicações é aumentado. A Figura 2.5 mostra as estimativas obtidas com $\hat{\mathbf{Z}}_i$ inteira, enquanto a Figura 2.6 utiliza $\hat{\mathbf{Z}}_i$ truncada.

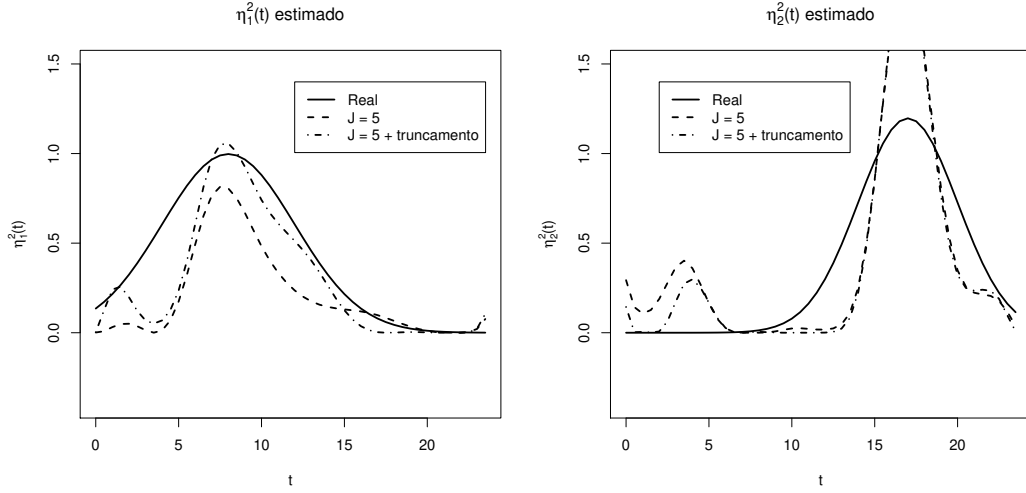


Figura 2.1: $\hat{\eta}_1^2(t)$ e $\hat{\eta}_2^2(t)$ estimados com $J = 5$

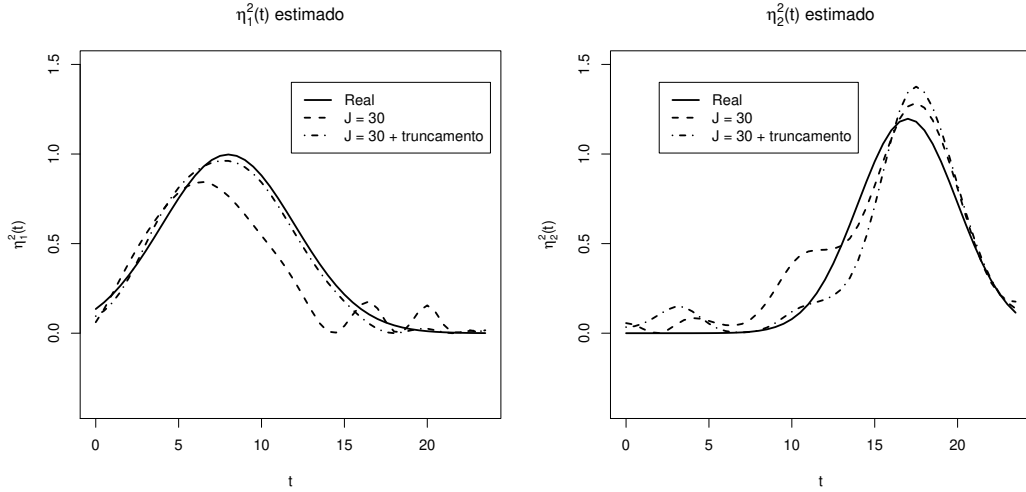


Figura 2.2: $\hat{\eta}_1^2(t)$ e $\hat{\eta}_2^2(t)$ estimados com $J = 30$

É mais fácil observar, através das Figuras 2.5 e 2.6, que quando $J = 30$, o truncamento melhora decisivamente a estimativa de $\hat{\eta}_1^2(t)$ para t entre 6 e 14, e da mesma forma para $\hat{\eta}_2^2(t)$ quando t está aproximadamente entre 8 e 14.

A Figura 2.7 mostra a estimativa obtida quando as rotinas de chute inicial e otimização são alimentadas com os valores verdadeiros de $\mathbf{Z}$. Como o estimador $\hat{\mathbf{Z}}_i$ é consistente, é razoável supor que o desempenho assintótico da rotina seja exemplifi-

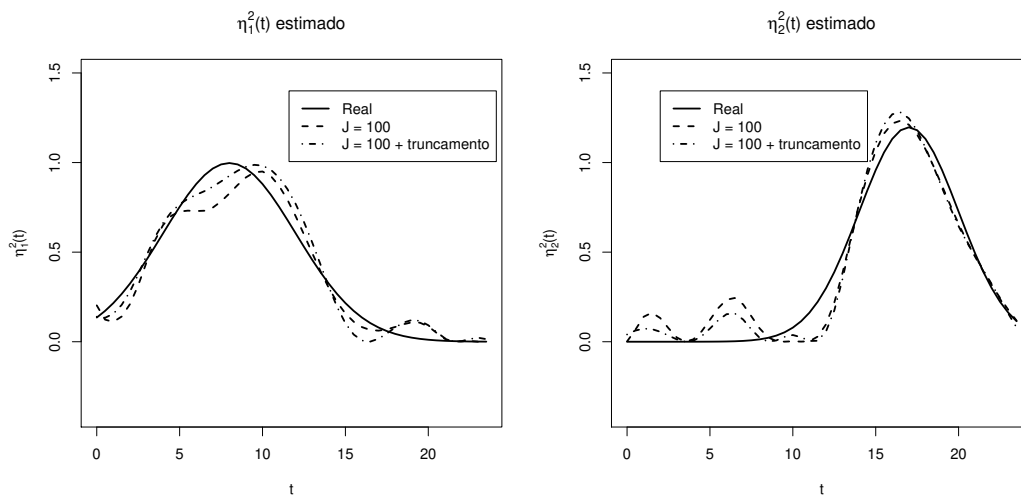


Figura 2.3: $\hat{\eta}_1^2(t)$ e $\hat{\eta}_2^2(t)$ estimados com $J = 100$

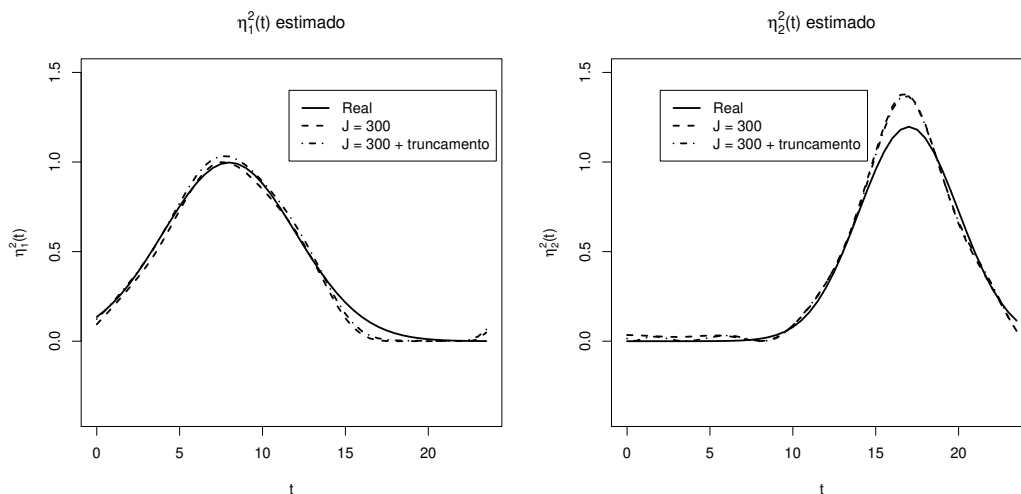


Figura 2.4: $\hat{\eta}_1^2(t)$ e $\hat{\eta}_2^2(t)$ estimados com $J = 300$

cado na Figura 2.7, ou seja, a estimativa apresentada é a melhor possível, a menos da precisão arbitrária da rotina de otimização.

Em suma, nos quatro casos observados nota-se que as estimativas ficaram menos precisas para valores da variância próximos de zero. O estimador é bastante sensível ao número J de replicações, com uma performance satisfatória apenas se $J > 100$ – muito mais que o necessário para boas estimativas de tipologia. A precisão das estimativas

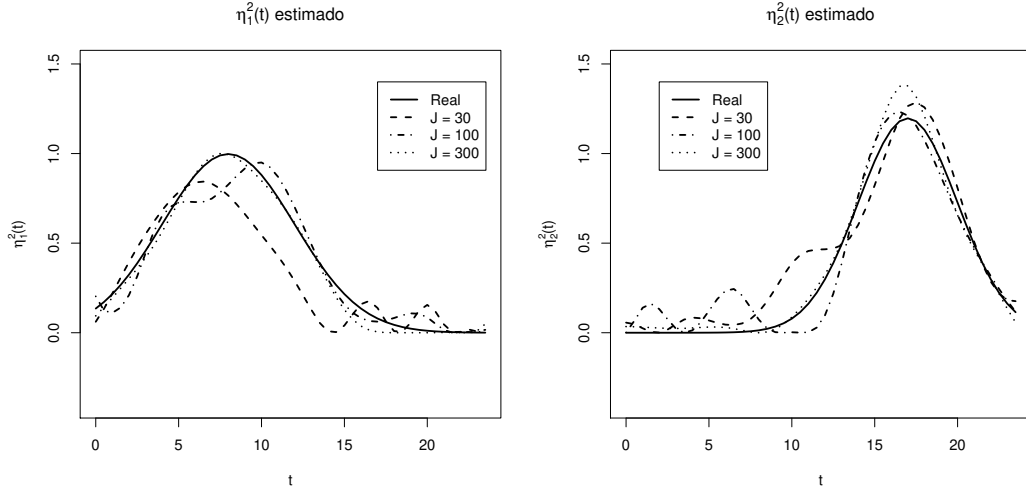


Figura 2.5: Estimativas $\hat{\eta}_c^2(t)$ sobrepostas, para diferentes J

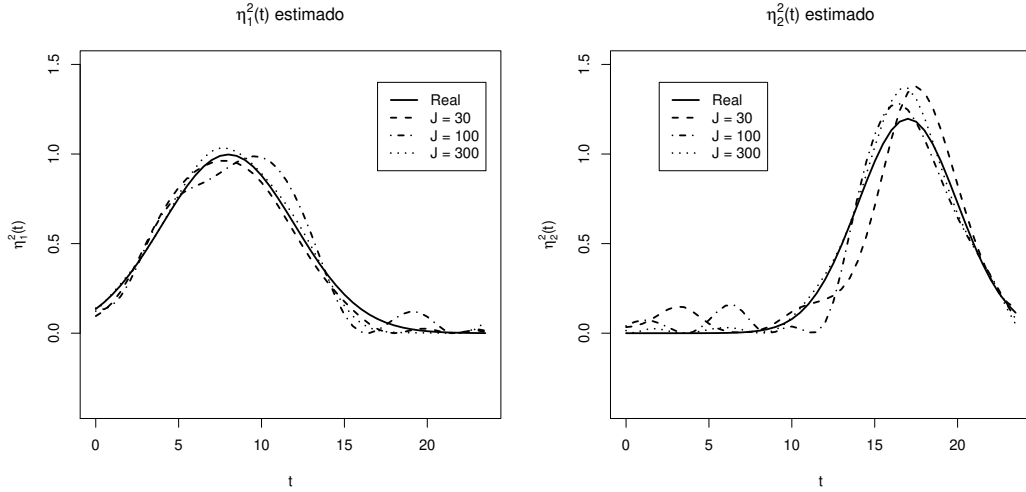


Figura 2.6: Estimativas $\hat{\eta}_c^2(t)$ sobrepostas, para diferentes J , com truncamento em $\hat{\mathbf{Z}}_i$

de *variância* é ligeiramente melhorada pelo truncamento, mas a diferença é significativa apenas se J é pequeno. A maior parte da informação sobre $\hat{\eta}_c^2(t)$ está, naturalmente, na diagonal das matrizes $\hat{\mathbf{Z}}_i$.

Sobre a correlação, a Tabela 2.1 mostra os valores estimados de ϕ , nos respectivos casos: chute inicial (dado por $\hat{\phi}^{(0)}$), com a matriz $\mathbf{Z}_i(s, t)$ completa (dado por $\hat{\phi}_c$) e com a matriz $\mathbf{Z}_i(s, t)$ truncada na forma (1.10) (dado por $\hat{\phi}'_c$).

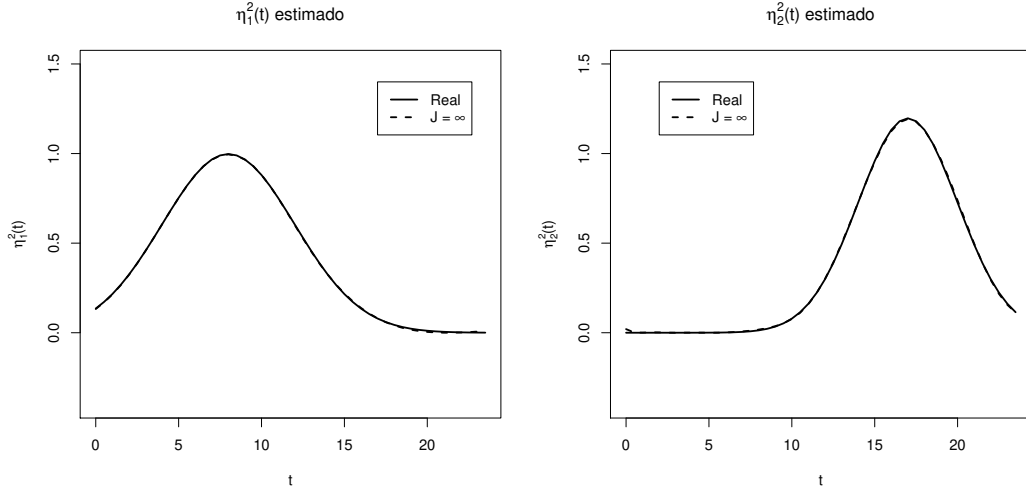


Figura 2.7: Estimativas $\hat{\eta}_1^2(t)$ e $\hat{\eta}_2^2(t)$ com $J = \infty$

Tabela 2.1: Estimativa inicial $\hat{\phi}^{(0)}$ e estimativas $\hat{\phi}_c$, obtidas com as matrizes $\mathbf{Z}$ completas, e $\hat{\phi}'_c$, com $\mathbf{Z}$ truncadas (valores verdadeiros $\phi_1 = 0,5$, $\phi_2 = 1,5$)

	$\hat{\phi}^{(0)}$	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\phi}'_1$	$\hat{\phi}'_2$
$J = 5$	1,34	0,05	1,14	0,88	1,39
$J = 30$	1,09	0,39	1,51	1,02	1,87
$J = 100$	1,05	0,48	1,62	1,02	1,92
$J = 300$	1,01	0,52	1,44	1,01	1,78
$J = \infty$	—	0,50	1,49	—	—

Nota-se aqui que o truncamento é indesejável, uma vez que interfere na velocidade do decaimento exponencial e leva a estimativas viesadas de ϕ_c . A qualidade do chute inicial também mostra-se irrelevante – o algoritmo é capaz de detectar velocidades diferentes de decaimento, mesmo com a restrição artificial que $\phi_1 = \phi_2$ empregada para obter $\hat{\phi}^{(0)}$ (veja o Apêndice B, Seção B.1). Bem como $\eta_c^2(t)$, $\phi_c(t)$ é estimado com precisão alta a partir de $J = 100$ replicações, mas desde que não haja truncamento.

Para apreciar a incerteza sobre as estimativas, foi realizado um procedimento de *bootstrapping* para construir intervalos de confiança. As replicações foram sorteadas com

reposição em j , de maneira independente para cada curva agregada i . Considerou-se o IC para η_c para cada ponto t fixado, de maneira independente dos demais – as extremidades não são funções suaves, consequentemente, devido ao erro amostral. O resultado de 300 reamostras encontra-se nas Figuras 2.8 e 2.9, para $J = 30$ e $J = 300$ replicações, respectivamente.

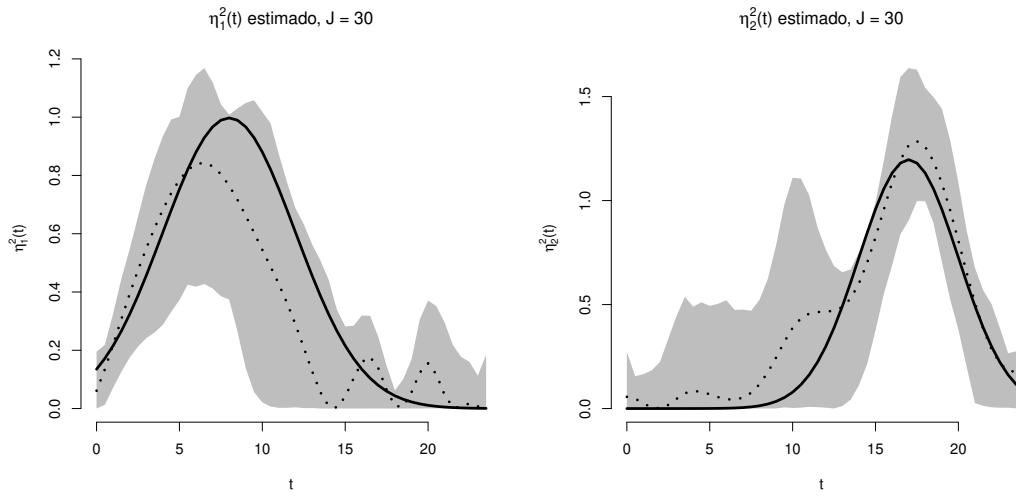


Figura 2.8: Intervalo de confiança (confiabilidade de aproximadamente 95%) para $\hat{\eta}_1^2(t)$ e $\hat{\eta}_2^2(t)$ estimados com $J = 30$

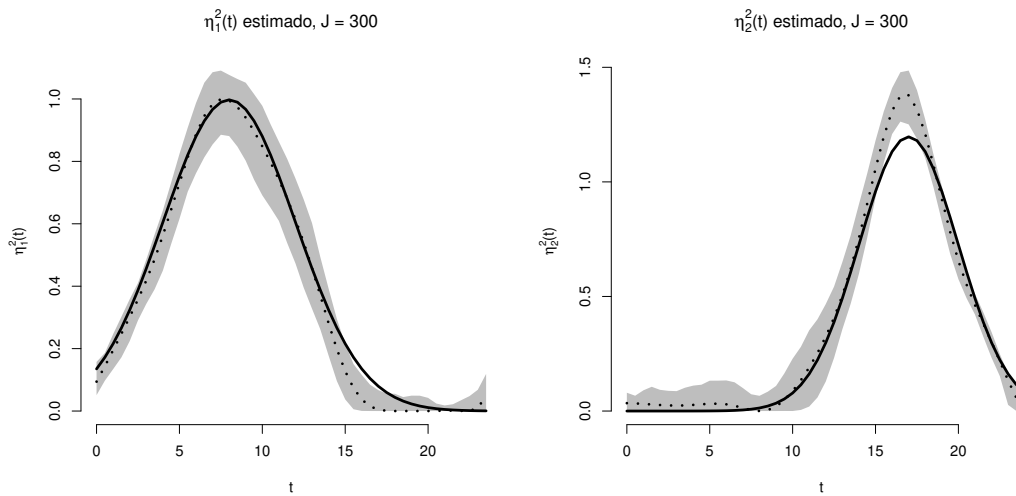


Figura 2.9: Intervalo de confiança (confiabilidade de aproximadamente 95%) para $\hat{\eta}_1^2(t)$ e $\hat{\eta}_2^2(t)$ estimados com $J = 300$

O resultado do bootstrapping de ϕ_1 e ϕ_2 é exibido nas Figuras 2.10 e 2.11, para $J = 30$ e $J = 300$ replicações, respectivamente. Os histogramas representam os dados da reamostra, as linhas sólidas os valores verdadeiros, as tracejadas representam os valores estimados e as linhas pontilhadas representam os limites que formam aproximadamente 95% de confiabilidade.

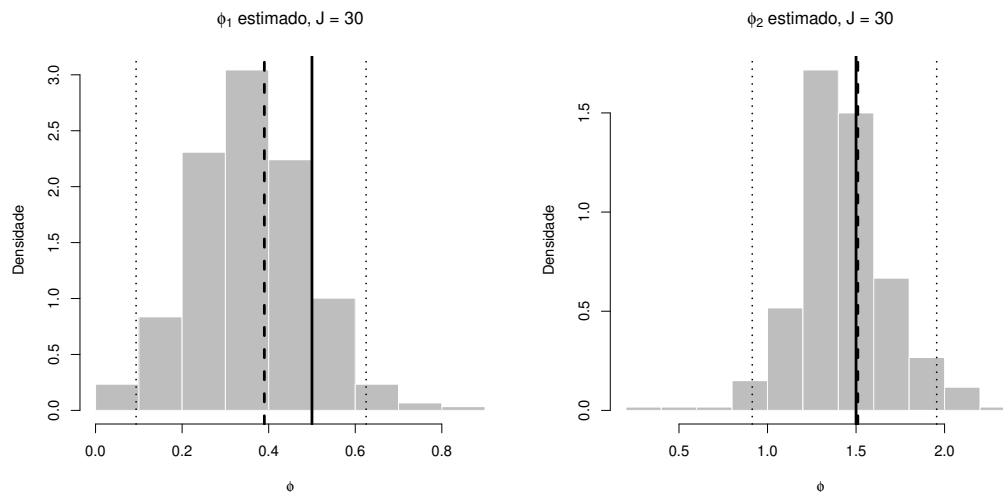


Figura 2.10: Intervalo de confiança (confiabilidade de aproximadamente 95%) para ϕ_1 e ϕ_2 estimados com $J = 30$

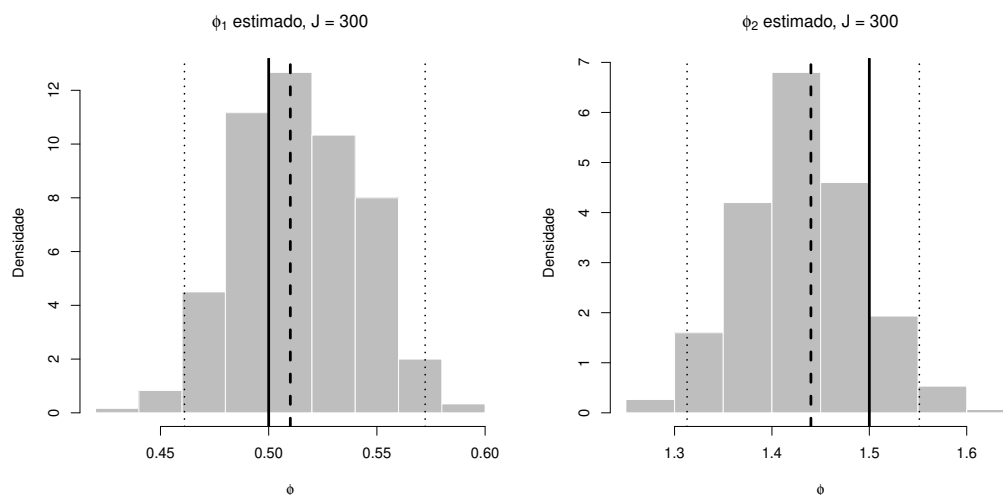


Figura 2.11: Intervalo de confiança (confiabilidade de aproximadamente 95%) para ϕ_1 e ϕ_2 estimados com $J = 300$

Capítulo 3

Estimação de Covariância pelo Produto Tensorial de Splines

Este capítulo desenvolve a implementação da metodologia de estimação proposta em Dias et al. (2009) que consiste em estimar as superfícies de covariância através do produto tensorial de Splines $S(s, t)$. Como as estimativas resultantes deste método não são necessariamente funções não negativas definidas (logo, não são funções de covariância), uma adaptação da correção proposta por Hall et al. (1994) é realizada, inspecionando a Transformada de Fourier das estimativas, de modo que elas (ou as suas respectivas superfícies de correlação) satisfaçam o Teorema de Bochner. Resultados de estudo de simulação são exibidos.

3.1 O Produto Tensorial de Splines

Seja $B(t)$ um spline unidimensional em $[0, T]$ com conjunto de nós dado por $\boldsymbol{\tau} = \{\tau_1 \leq \dots \leq \tau_L\}$. O produto tensorial de splines $S(s, t)$ é uma função $S : [0, T] \times [0, T] \rightarrow \mathbb{R}$, com nós no conjunto $\boldsymbol{\tau} \times \boldsymbol{\tau}$, onde $\times$ é o produto cartesiano, e que é representado explicitamente por

$$S(s, t) = B(s) \otimes B(t) = \sum_{k_1=1}^K \sum_{k_2=1}^K \theta_{k_1, k_2} B_{k_1}(s) B_{k_2}(t).$$

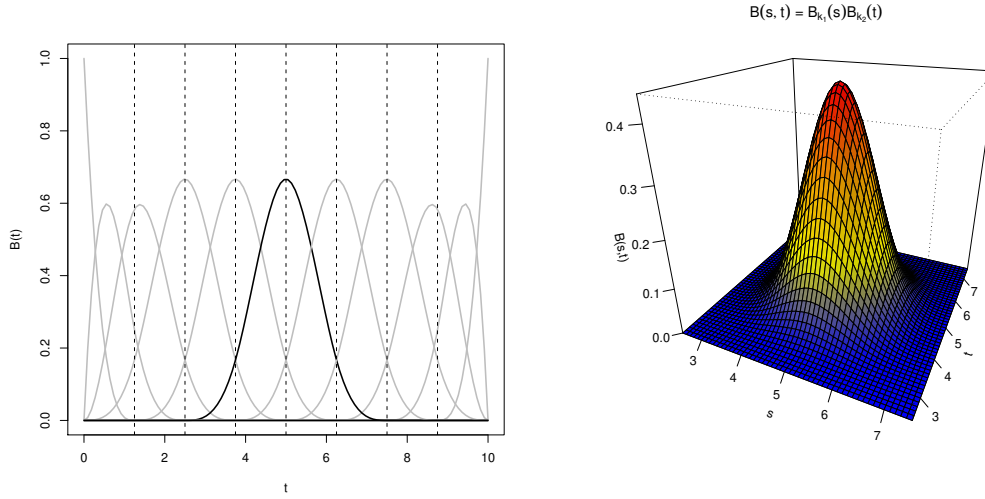


Figura 3.1: A base $B_6(t)$ da Figura 1.1, e a base $B_6(s)B_6(t)$ do produto tensorial de Splines

Uma base do tipo $B_{k_1, k_2}(s, t)$ está ilustrada na Figura 3.1. A estimativa não paramétrica de $\sigma_c(s, t)$ é obtida pelas estimativas dos coeficientes $\theta_{k_1, k_2, c}$, e tem forma

$$\hat{\sigma}_c(s, t) = \sum_{k_1=1}^K \sum_{k_2=1}^K \hat{\theta}_{k_1, k_2, c} B_{k_1}(s) B_{k_1}(t).$$

As superfícies $\hat{\sigma}_c(s, t)$ são obtidas a partir da separação das superfícies agregadas $\hat{\mathbf{Z}}_i(s, t)$, de modo análogo ao caso das curvas e tipologias. Tem-se aqui que

$$\mathbf{Z}_i(s, t) = \sum_{c=1}^C N_{i,c} q_{i,c}^2 \sigma_c(s, t).$$

Observa-se somente $\hat{\mathbf{Z}}_i(s, t)$, para $i = 1, \dots, I$, e os números $N_{i,c}$ e $q_{i,c}$ são conhecidos. Considere a vetorização da matriz $\mathbf{Z}_i(s, t)$, dada por

$$\mathbf{z}_i = \begin{pmatrix} \mathbf{Z}_i(s_1, t_1) \\ \vdots \\ \mathbf{Z}_i(s_n, t_1) \\ \mathbf{Z}_i(s_1, t_2) \\ \vdots \\ \mathbf{Z}_i(s_n, t_n) \end{pmatrix}_{n^2 \times 1}$$

A matriz de delineamento é construída considerando primeiramente as bases $B_k(t)$, $k = 1, \dots, K$, que formam uma matriz de delineamento da forma

$$\mathbf{B} = \begin{pmatrix} B_1(t_1) & B_2(t_1) & \cdots & B_K(t_1) \\ \vdots & \vdots & \vdots & \vdots \\ B_1(t_n) & B_2(t_n) & \cdots & B_K(t_n) \end{pmatrix}_{n \times K}$$

Agora, o produto de Kronecker de $\mathbf{B}$ consigo mesma gera a matriz de delineamento para o produto tensorial de splines, na forma

$$\mathbf{B} \otimes \mathbf{B} = \begin{pmatrix} B_1(s_1)B_1(t_1) & \cdots & B_1(s_1)B_K(t_1) & B_2(s_1)B_1(t_1) & \cdots & B_K(s_1)B_K(t_1) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ B_1(s_n)B_1(t_n) & \cdots & B_1(s_n)B_K(t_n) & B_2(s_n)B_1(t_n) & \cdots & B_K(s_n)B_K(t_n) \end{pmatrix}_{n^2 \times K^2}$$

Repare que cada linha da matriz $\mathbf{B} \otimes \mathbf{B}$, onde $B_{k_1}(s)B_{k_2}(t)$ é calculada nos índices s_{r_1}, t_{r_2} , para todo $k_1, k_2 = 1, \dots, K$, corresponde ao componente $\mathbf{Z}_i(s_{r_1}, t_{r_2})$ da covariância amostral vetorizada $\mathbf{z}_i$. Novamente fixa-se $q_{i,c} = 1$ para o desenvolvimento das contas. Considerando $\mathbf{W} = \mathbf{N} \otimes (\mathbf{B} \otimes \mathbf{B})$, uma matriz $In^2 \times CK^2$, o argumento para a construção é análogo ao da estrutura da matriz de delineamento da estimação de tipologia, em que os componentes agregados são escritos por $\mathbf{N} \otimes \mathbf{B}^*$. Tomando

$$\mathbf{z} = \begin{pmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \\ \vdots \\ \mathbf{z}_I \end{pmatrix}_{In^2 \times 1},$$

então as equações do modelo calculado nos pontos $s, t = t_1, t_2, \dots, t_n$, dadas por

$$\mathbf{z}_i(s, t) = \sum_{c=1}^C N_{i,c} \sigma_c(s, t) + \varepsilon(s, t), \quad i = 1, 2, \dots, I,$$

onde naturalmente ε é um erro de aproximação, podem ser escritas como

$$\mathbf{z} = \mathbf{W}\boldsymbol{\theta} + \boldsymbol{\varepsilon},$$

em que $\boldsymbol{\theta}$ é o vetor de $\boldsymbol{\theta}_c$ concatenados, de modo análogo ao vetor $\boldsymbol{\beta}$ que separa as curvas de tipologia. A estimativa de mínimos quadrados ordinários para $\boldsymbol{\theta}$ é dada por

$$\hat{\boldsymbol{\theta}} = (\mathbf{W}'\mathbf{W})^{-1} \mathbf{W}'\mathbf{z}. \quad (3.1)$$

Não há nada que assegure que, por exemplo, $\hat{\sigma}_c(t, t) \geq 0$, o que claramente viola a definição de função de variância (e esta condição é ainda menos restritiva que a de que a estimativa seja não negativa definida). O artigo de Hall et al. (1994) encontra uma solução considerando a transformada de Fourier da estimativa: como demonstrado no Teorema 2.1.3, a Transformada de Fourier da função de covariância de um processo estacionário é real e não negativa. Ao aplicar um truncamento nos valores reais negativos e remover toda a parte imaginária das transformadas das estimativas, a transformada inversa será uma função de covariância estacionária. Essa condição é, contudo, muito restritiva para as estimativas – ela implica em variância constante e decaimento homogêneo da função de correlação. As estimativas resultantes são constantes no eixo $t + s$, e decaem homogeneousmente no eixo $t - s$.

Neste trabalho, a análise da transformada de Fourier da função de correlação será empregada para comparar as estimativas com a forma de covariância oferecida no Teorema 2.1.4, o que devolve parte da flexibilidade da estimação pelo produto tensorial de splines ao procedimento de Hall et al. (1994). Para tanto, assumiremos que $\eta_c^{-1}(s)\eta_c^{-1}(t)\sigma_c(s, t)$ é uma função de covariância de um processo estacionário, de onde a metodologia de correção de Hall et al. (1994) pode ser adaptada para as estimativas $\hat{\eta}_c^{-1}(s)\hat{\eta}_c^{-1}(t)\hat{\sigma}_c(s, t)$. Antes disso, contudo, será discutida a qualidade da estimação por produto tensorial através de um exemplo, e uma alternativa – o produto tensorial de splines suavizados – será proposta.

3.2 O Produto Tensorial de Splines Suavizados

O problema de estimação de K , o número de bases que compõem um Spline, não foi abordado anteriormente em consequência da sua discussão prévia em Martarelli (2006)

e Dias et al. (2009) para as tipologias, e do fato que a estimação de $\boldsymbol{\eta}(t)$ através da semiparametrização ser essencialmente similar à estimação das tipologias $\boldsymbol{\alpha}(t)$, exceto pela inclusão dos parâmetros $\boldsymbol{\phi}$ da correlação. Neste caso, decidiu-se manter uma quantidade similar de bases.

O número de nós está relacionado com a suavidade da estimativa. No caso unidimensional, muitos nós distintos aproximam melhor a estimativa aos dados observados, mas incorporam ruído e variabilidade também. De fato, utilizar todos os pontos em que o dado funcional foi observado como nós produz uma *interpolação* dos pontos. Nesse caso, diminuir o número de nós, e consequentemente o de bases, suaviza a estimativa – mas um número muito pequeno de nós pode perder informações importantes com respeito à curvatura e flutuações locais.

Considere por exemplo a função polinomial

$$f(t) = 400t(t - 0,25)(t - 0,5)(t - 0,75)^2(t - 1),$$

amostrada em 50 pontos equidistantes do intervalo $[0, 1]$, e perturbada por um erro aleatório Normal e independente, com média zero e variância $(0,2)^2$. A Figura 3.2 exhibe $f(t)$ (linha sólida) bem como o resultado de uma amostra $\mathbf{y}$ com erros de medição, em (A). (B) representa um cenário com um número de nós internos satisfatório (4, totalizando 8 funções B-spline cúbicas). (C) representa uma estimativa suavizada demais, com apenas 1 nó interno (5 funções base), e (D) mostra o caso em que há o ajuste de ruído, próximo da interpolação, com 30 nós internos.

Selecionar o número de bases para o produto tensorial é um problema ainda mais delicado. Isso porque a superfície que desejamos estimar, $\sigma_c(s, t)$, pode ter um decaimento de correlação muito rápido, exigindo muitos nós em torno da reta $s = t$ para capturar informação sobre a função de variância. Porém, um número grande de nós gera estimativas cada vez mais próximas da interpolação nos pontos, carregando um perfil oscilatório na estimativa final, especialmente nas extremidades da função.

Para ilustrar esse problema, considere a função $f : [0, 10] \times [0, 10] \rightarrow \mathbb{R}$ dada por $f(s, t) = \cos(10^{-1}\pi(s - 5)) \cos(10^{-1}\pi(t - 5)) \exp\{-2(t - s)^2\}$, amostrada em 20^2 pontos

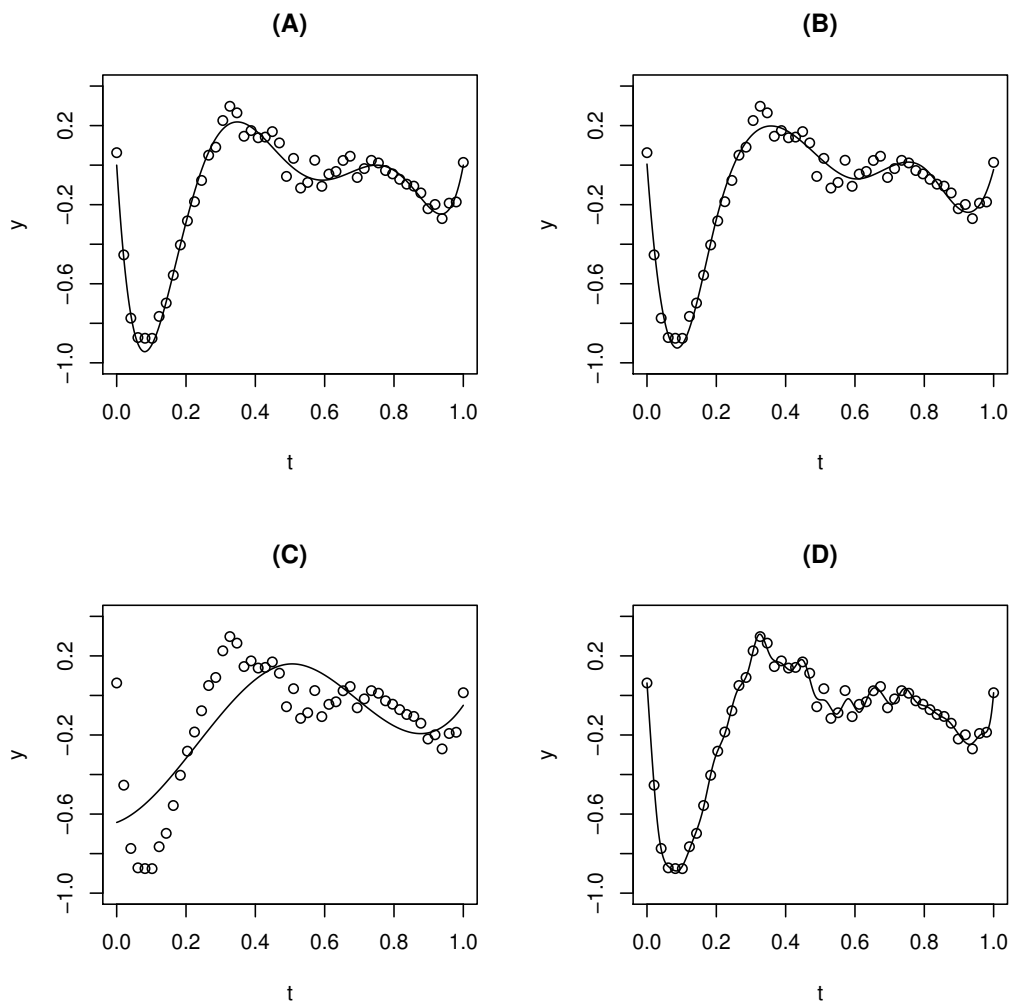


Figura 3.2: (A) A função $f(t)$ e a amostra y ; (B) Estimativa com 4 nós internos; (C) Estimativa com 1 nó interno; (D) Estimativa com 30 nós internos

equispaçados. Ela decai rapidamente para valores grandes de $|t - s|$, como mostra a Figura 3.3.

Aplicar a metodologia de ajuste pelo produto tensorial de splines produz as aproximações exibidas nas Figuras 3.4 e 3.5, com respectivamente $K = 7^2$ e $K = 18^2$ bases. Pode-se observar que a aproximação com menos bases é mais suave; contudo, ela subestima a diagonal $s = t$. Se forem empregadas mais bases, a aproximação consegue capturar melhor o decaimento rápido em $|t - s|$, mas a diagonal começa a apresentar

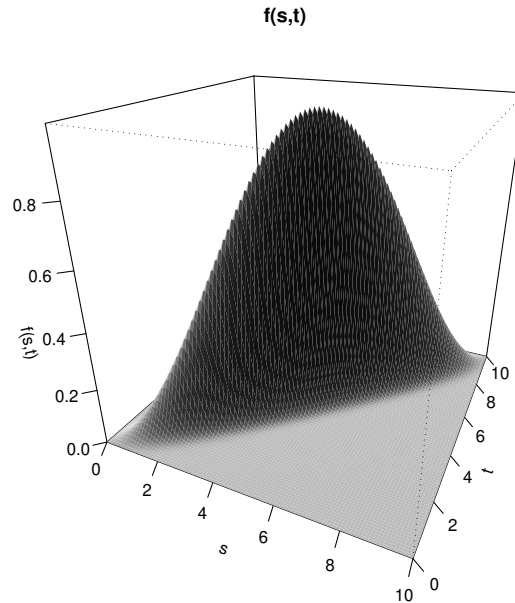


Figura 3.3: A função $f(s, t)$ para o exemplo de estimação por produto tensorial

um carácter oscilatório, sobretudo nas extremidades do intervalo.

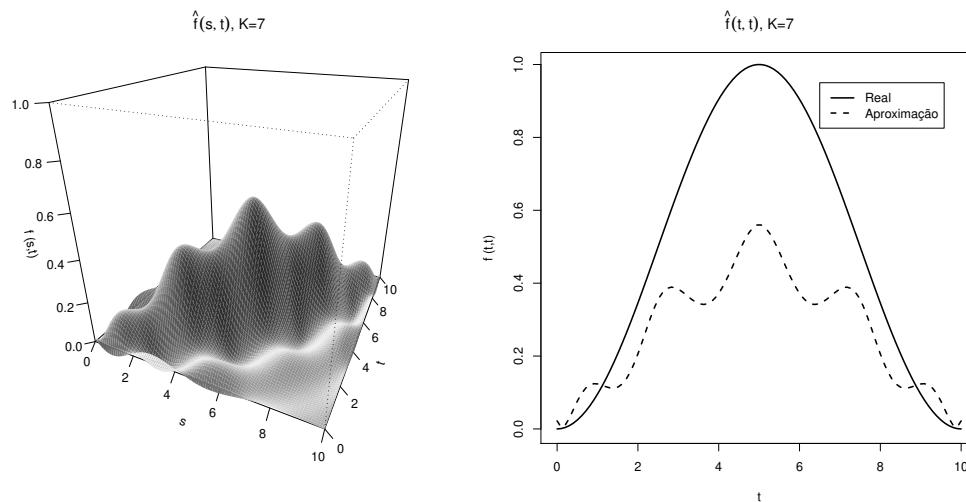


Figura 3.4: A aproximação $\hat{f}(s, t)$, com 7^2 bases

Uma solução para o problema de determinar o número de nós, tanto para dados unidimensionais quanto para superfícies, é empregar um procedimento denominado *suavização spline*. O método de suavização spline consiste primeiramente em tomar

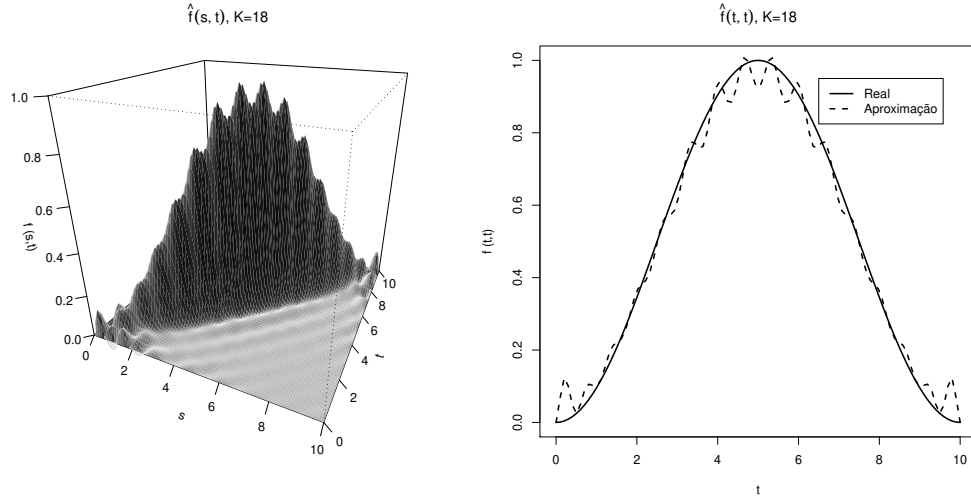


Figura 3.5: A aproximação $\hat{f}(s, t)$, com 18^2 bases

como nós todos os pontos em que o dado funcional foi observado. Mas a medida de qualidade de ajuste (por exemplo a soma de quadrados dos erros, dada por (1.1)) é penalizada por um termo que mede a não suavidade da estimativa, controlado por um parâmetro λ , de tal maneira que a interpolação seja evitada. A medida da não suavidade de um spline univariado é dada por

$$\text{PEN}_2(y(t)) := \int_0^T \left| \frac{d^2 y(t)}{dt^2} \right|^2 dt.$$

Aproximando $\text{PEN}_2(y(t))$ por $\text{PEN}_2(\hat{y}(t))$, obtemos

$$\text{PEN}_2(\hat{y}(t)) = \int_0^T \left| \frac{d^2 \hat{y}(t)}{dt^2} \right|^2 dt = \hat{\beta}' \mathbf{R} \hat{\beta},$$

onde $\mathbf{R}$ é uma matriz $K \times K$ tal que cada elemento r_{k_1, k_2} é dado pela integral

$$r_{k_1, k_2} = \int_0^T \left(\frac{d^2}{dt^2} B_{k_1}(t) \right) \left(\frac{d^2}{dt^2} B_{k_2}(t) \right) dt.$$

O critério de ajuste dos dados, incluindo a penalização, para um valor de λ fixado, é dado por

$$Q(\beta_1, \dots, \beta_K | \lambda) = \sum_{r=1}^n \left(y(t_r) - \sum_{k=1}^K \beta_k B_k(t_r) \right)^2 + \lambda \text{PEN}_2(\hat{y}(t)), \quad (3.2)$$

que em notação matricial é escrito como

$$Q(\boldsymbol{\beta} \mid \lambda) = (\mathbf{y} - \mathbf{B}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{B}\boldsymbol{\beta}) + \lambda\boldsymbol{\beta}'\mathbf{R}\boldsymbol{\beta}.$$

A equação (3.2) possui uma solução explícita para seu mínimo, dada por

$$\hat{\boldsymbol{\beta}}_\lambda = (\mathbf{B}'\mathbf{B} + \lambda\mathbf{R})^{-1} \mathbf{B}'\mathbf{y},$$

As funções spline têm uma ligação muito forte com equações da forma (3.2). De fato, splines cúbicos são a solução de equações dessa forma entre quaisquer funções suaves com primeira derivada contínua e com $\text{PEN}_2(y(t))$ finita, como demonstrado em de Boor (2001). Outros aspectos teóricos da suavização spline são discutidos em Wahba (1990).

Com a introdução do termo $\text{PEN}_2(\hat{y}(t))$, o problema de determinar a suavidade das estimativas deixa de depender de K , o número de bases, para depender de λ , o coeficiente de penalização. Quando λ é pequeno, sabe-se que a estimativa fica próxima da interpolação spline; se $\lambda \rightarrow \infty$, a estimativa produzida tende a uma regressão linear nos pontos. O objetivo é encontrar um valor de λ tal que haja um balanço entre a fidedignidade aos dados (ao minimizar o termo da soma de quadrados) e a suavidade das estimativas (ao minimizar o termo $\text{PEN}_2(\hat{y}(t))$), como discutido em Wahba (1990, p. viii) ou Ramsay & Silverman (2005, p. 94). O critério adotado neste trabalho para escolher o λ é tomar aquele que minimize a *validação cruzada generalizada*, ou **GCV**. Por ora o **GCV** não será explicitamente deduzido, mas sua fórmula é dada por

$$\text{GCV}(\lambda) = \frac{n^{-1}\text{SQE}}{(n^{-1}\text{tr}(\mathbf{I} - \mathbf{H}_\lambda))^2}, \quad (3.3)$$

onde SQE é a soma de quadrados dos resíduos do ajuste, $\text{tr}(\mathbf{I} - \mathbf{H}_\lambda)$ representa o traço da matriz $\mathbf{I} - \mathbf{H}_\lambda$, e $\mathbf{H}_\lambda$ é a matriz dada por

$$\mathbf{H}_\lambda = \mathbf{B}(\mathbf{B}'\mathbf{B} + \lambda\mathbf{R})^{-1} \mathbf{B}'.$$

O **GCV** é um procedimento equivalente a retirar um dos pontos observados da amostra e tentar prevê-lo, usando um modelo ajustado sem essa mesma observação, anotando

o erro de previsão em função de λ (este procedimento é chamado de validação cruzada *leave-one-out*). Ser equivalente significa que o λ ótimo da validação cruzada *leave-one-out* coincide com aquele que minimiza $\text{GCV}(\lambda)$. Mais sobre toda a teoria de suavização pode ser lido no Capítulo 5 de Ramsay & Silverman (2005).

Para o problema de superfícies, será aproveitada a estrutura de produto tensorial de bases e vetorização da matriz de covariância observada. Considere a fórmula para $\mathbf{R}$, onde cada entrada r_{k_1, k_2} é o produto interno $\langle B''_{k_1}(t), B''_{k_2}(t) \rangle$. Então uma maneira aproximada de medir a suavidade da superfície é empregando $\mathbf{R} \otimes \mathbf{R}$ como matriz de penalização. A matriz $\mathbf{R} \otimes \mathbf{R}$ inclui em cada uma de suas entradas $r_{k_1 k_2, k_3 k_4}$ o produto dos termos r_{k_1, k_3} e r_{k_2, k_4} da matriz de penalização por não suavidade das bases que formam o produto tensorial, ou seja, $\langle B''_{k_1}(s), B''_{k_3}(s) \rangle \cdot \langle B''_{k_2}(t), B''_{k_4}(t) \rangle$. Isso é equivalente a examinar a suavidade de $f(s_0, t)$ para cada s_0 fixado, e vice versa. Então o estimador de $\boldsymbol{\theta}$ para o modelo de produto tensorial de Splines, fixado λ , é dado por

$$\hat{\boldsymbol{\theta}} = ((\mathbf{B} \otimes \mathbf{B})'(\mathbf{B} \otimes \mathbf{B}) + \lambda \mathbf{R} \otimes \mathbf{R})^{-1} (\mathbf{B} \otimes \mathbf{B})' \mathbf{f}.$$

e o critério do GCV é empregado para obter $\hat{\lambda}$.

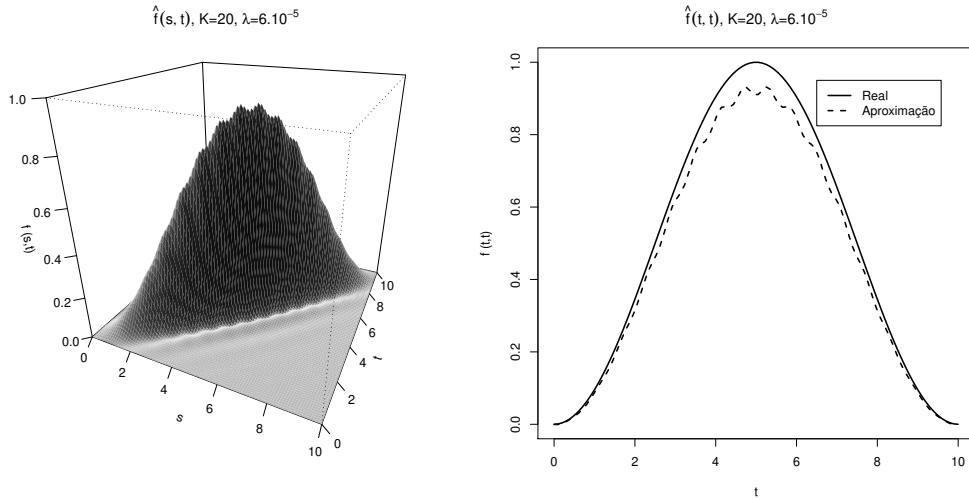


Figura 3.6: A aproximação $\hat{f}(s, t)$, com 20^2 bases e penalização de $\lambda = 6 \cdot 10^{-5}$

Retomando o exemplo anterior de aproximação da função exibida na Figura 3.3, a aproximação obtida pelo produto tensorial de splines, quando submetidos a uma

penalização por não suavidade, é exibida na Figura 3.6. Foram usadas 20^2 bases, e um parâmetro de suavização relativamente pequeno, de $\lambda = 6.10^{-5}$, estimado pelo GCV. Se comparada com as estimativas exibidas nas Figuras 3.4 e 3.5, é possível notar que as estimativas produzidas por esse método são bastante suaves e não subestimam a diagonal $f(t, t)$ da função $f(s, t)$.

Porém convém ressaltar que como $\text{PEN}_2(\hat{y}(t))$ é uma norma, não se pode afirmar que o melhor parâmetro de suavização para a superfície agregada seja ótimo para as superfícies individuais. Como a metodologia de Análise de Fourier será desenvolvida a partir dos parâmetros já estimados, a questão de melhor estimação de λ pode ser aprofundada em trabalhos futuros. A estimação de superfícies agregadas com penalização será feita com base na não suavidade de $\hat{\mathbf{Z}}$, a estimativa das superfícies de covariância agregadas. As estimativas individuais serão menos suaves em decorrência da desigualdade triangular

$$\|Z(s, t)\| \leq |N_{i,1}| \|\sigma_1(s, t)\| + \dots + |N_{i,C}| \|\sigma_C(s, t)\|,$$

em que $\|\cdot\|$ representa a norma funcional PEN_2 definida anteriormente. Neste caso, a soma de quadrados dos erros incluindo o termo de penalização é dado por

$$Q(\boldsymbol{\theta} \mid \lambda) = (\mathbf{z} - \mathbf{W}\boldsymbol{\theta})'(\mathbf{z} - \mathbf{W}\boldsymbol{\theta}) + \lambda \boldsymbol{\theta}'((\mathbf{N}'\mathbf{N}) \otimes \mathbf{R} \otimes \mathbf{R})\boldsymbol{\theta}, \quad (3.4)$$

em que $\mathbf{R} \otimes \mathbf{R}$ representa a matriz de penalização para o produto tensorial $\mathbf{B} \otimes \mathbf{B}$ escolhida anteriormente, e a escolha do produto de Kronecker da última componente por $\mathbf{N}'\mathbf{N}$ dá-se por argumento similar à construção de $\mathbf{W}$, e pela propriedade

$$(\mathbf{N} \otimes \mathbf{B} \otimes \mathbf{B})'(\mathbf{N} \otimes \mathbf{B} \otimes \mathbf{B}) = \mathbf{N}'\mathbf{N} \otimes ((\mathbf{B} \otimes \mathbf{B})'(\mathbf{B} \otimes \mathbf{B})),$$

que leva à escolha de uma penalização proporcional ao termo do produto cruzado $(\mathbf{B} \otimes \mathbf{B})'(\mathbf{B} \otimes \mathbf{B})$. O estimador $\hat{\boldsymbol{\theta}}$, fixado λ , é dado por

$$\hat{\boldsymbol{\theta}} = (\mathbf{W}'\mathbf{W} + \lambda (\mathbf{N}'\mathbf{N}) \otimes \mathbf{R} \otimes \mathbf{R})^{-1} \mathbf{W}'\mathbf{z}.$$

3.3 Transformada de Fourier do Produto Tensorial de Splines

O Teorema de Bochner é uma ferramenta para garantir que as estimativas obtidas sejam não negativas definidas. O resultado do teorema, no entanto, só vale para processos *fracamente estacionários*. Mas se considerarmos a família de funções de covariância definidas no Teorema 2.1.4, pode-se deduzir que $|\eta(t)|_c^{-1}W_c(t)$ tem covariância de um processo fracamente estacionário exceto por alguma mudança de sinal (a saber, se $\eta(t)$ não for positiva para todo t). Como a estimativa por produto tensorial de splines é capaz de separar as covariâncias individuais, a correção pelo Teorema de Bochner pode ser aplicada diretamente às estimativas.

Primeiramente, define-se a Transformada de Fourier da função de covariância por

$$\Sigma_c(w_1, w_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{i(sw_1 + tw_2)} \sigma_c(s, t) ds dt.$$

A integral pode divergir; uma condição suficiente para a sua existência é que

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |\sigma_c(s, t)| ds dt \leq \infty,$$

como discutido em Bracewell (1986). A realização dos processos examinados, contudo, está limitada a $[0, T]$, e consequentemente a função de covariância é não nula apenas em $[0, T] \times [0, T]$. Considere um processo estacionário, tal que $\sigma_c(s, t) = \rho(t - s)$ e $\mathcal{F}\{\rho\} = P$. Então a transformada de Fourier de $\rho(t - s)\mathbf{1}_{[0, T]}(s)\mathbf{1}_{[0, T]}(t)$ é dada por

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-i(sw_1 + tw_2)} \mathbf{1}_{[0, T]}(s) \mathbf{1}_{[0, T]}(t) \rho(t - s) ds dt = \int_0^T \int_0^T e^{-i(sw_1 + tw_2)} \rho(t - s) ds dt$$

Fazendo a integral por substituição em $x = t$, $y = t - s$, então $dx dy = | -1 | ds dt$ e

$$\begin{aligned} & \int_0^T \int_0^T e^{-i(sw_1 + tw_2)} \rho(t - s) ds dt \\ &= \int_0^T \int_{-T}^T e^{-i((x-y)w_1 + xw_2)} \rho(y) dy dx = \int_0^T e^{-ix(w_1 + w_2)} dx \int_{-T}^T e^{iyw_1} \rho(y) dy \quad (3.5) \\ &= e^{-i\frac{T}{2}(w_1 + w_2)} T \frac{\text{sen}(T2^{-1}(w_1 + w_2))}{(T2^{-1}(w_1 + w_2))} \left(P(w_1) - 2 \int_T^\infty \cos(w_1 y) \rho(y) dy \right) \end{aligned}$$

Como a função $P(w)$ é positiva e converge, $P(w) - 2 \int_T^\infty \cos(wy) \rho(y) dy$ corresponde simplesmente à integral de Fourier restrita a $[-T, T]$, e converge a $P(w)$ quando $T \rightarrow \infty$. Ela também é sempre real. Mas não é certo que para todo w essa componente seja positiva, se T é finito. Por exemplo, para a função $\rho(y) = \exp\{-\phi|y|\}$, sendo T constante, temos que

$$\int_{-T}^T \cos(wy) e^{-\phi|y|} dy = \frac{1}{\phi^2 + w^2} 2 \left(\phi - e^{-\phi T} (\phi \cos(Tw) - w \operatorname{sen}(Tw)) \right).$$

Observe que tomando $\phi = 1$, $T = 1$ e $w = 5$, a integral definida resulta num valor de $-0,066$, o que justifica a preocupação com a não negatividade da função. A região que deve ser não negativa corresponde à da desigualdade

$$2\phi + 2e^{-\phi T} (w \operatorname{sen}(Tw) - \phi \cos(Tw)) \geq 0.$$

Agora repare que as funções trigonométricas estão limitadas em $[-1, 1]$. Escrevendo

$$(w \operatorname{sen}(Tw) - \phi \cos(Tw)) \geq -\phi e^{\phi T},$$

pode-se concluir que o caso extremo da desigualdade é dada por

$$-w - \phi \geq -\phi e^{\phi T},$$

ou simplesmente a não negatividade deve valer para $\phi + |w| \leq \phi e^{\phi T}$. Será adotada precisamente essa desigualdade para escolher onde a positividade de $P(w_1)$ será examinada, pois $e^{-\phi|t-s|}$ corresponde ao processo investigado pelo procedimento de simulação. Mas com base nas condições de existência da transformada de Fourier, espera-se que essa região seja grande desde que a janela de inspeção, dada por T , seja grande também. Determinar a região não negativa para um caso mais geral permanece um problema em aberto. Em suma, para o processo restrito a $[0, T]$, é certo que a transformada da covariância $e^{-\phi_c|t-s|}$, dada por $\Sigma_c(w_1, w_2)$, verifica

$$e^{i\frac{T}{2}(w_1+w_2)} \left(\frac{\operatorname{sen}(2^{-1}(w_1+w_2))}{(2^{-1}(w_1+w_2))} \right)^{-1} \Sigma_c(w_1, w_2) \geq 0 \quad (3.6)$$

em um intervalo $|w_1| < e^{\phi T} - \phi$.

Explicitamente, a transformada do produto tensorial de splines é dada por

$$\begin{aligned}
\hat{\Sigma}_c(w_1, w_2) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-i(sw_1+tw_2)} \hat{\sigma}_c(s, t) ds dt \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-i(sw_1+tw_2)} \sum_{k_1=1}^K \sum_{k_2=1}^K \hat{\theta}_{c,k_1,k_2} B_{k_1}(s) B_{k_2}(t) ds dt \\
&= \sum_{k_1=1}^K \sum_{k_2=1}^K \hat{\theta}_{c,k_1,k_2} \int_{-\infty}^{\infty} e^{-isw_1} B_{k_1}(s) ds \int_{-\infty}^{\infty} e^{-itw_2} B_{k_2}(t) dt \\
&= \sum_{k_1=1}^K \sum_{k_2=1}^K \hat{\theta}_{c,k_1,k_2} \mathcal{F}_{k_1}(w_1) \mathcal{F}_{k_2}(w_2)
\end{aligned} \tag{3.7}$$

Aplicando (A.6), ou seja, considerando as transformadas unidimensionais, define-se um vetor (em função de w) onde cada componente corresponde à transformada de Fourier da k -ésima base B-spline, isto é,

$$\mathcal{F}(w) = \begin{pmatrix} \mathcal{F}_1(w) \\ \mathcal{F}_2(w) \\ \vdots \\ \mathcal{F}_K(w) \end{pmatrix}_{K \times 1}$$

Então (3.7) pode ser escrita como

$$\hat{\Sigma}_c(w_1, w_2) = \mathcal{F}(w_1)' \Theta_c \mathcal{F}(w_2),$$

onde Θ_c é uma matriz $K \times K$ com os componentes θ_{c,k_1,k_2} nas suas entradas, c fixado. A transformada vale para qualquer forma que Θ_c tenha sido estimado, seja splines ordinários ou suavização spline. O algoritmo para corrigir as estimativas obtidas pelo produto tensorial de splines consiste em

Passo 1: Estimar Θ e desagregar as covariâncias $\sigma_c(s, t)$ das C subpopulações, aplicando suavização spline se necessário.

Passo 2: Estimar $|\eta(t)|$ através de $\sqrt{\hat{\sigma}_c(t, t)}$, suavizando novamente se necessário.

Passo 3: Supondo o modelo (2.6), $\hat{\rho}_c(s, t) = |\hat{\eta}(s)|^{-1}|\hat{\eta}(t)|^{-1}\hat{\sigma}_c(s, t)$ é estimador de covariância de um processo estacionário. Logo, sua transformada de Fourier deve satisfazer (3.5). A transformada de Splines é dada por (3.7); logo, basta truncar os valores negativos da estimativa da transformada, modificada como em (3.6).

3.4 Exemplo Simulado

Considere novamente os cenários simulados nas seções 1.5 e 2.3. O artigo de Dias et al. (2009) prova a consistência da estimativa por produto tensorial, sem a correção baseada na análise da Transformada de Fourier. Será portanto examinado somente um caso, em que o número de replicações J é igual a 30, e as estimativas de $\mathbf{Z}$ não são muito boas ainda.

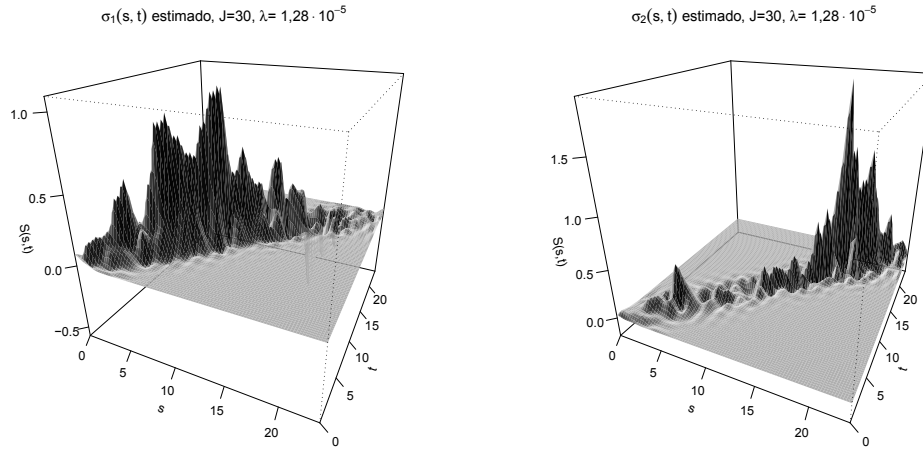


Figura 3.7: $\hat{\sigma}_c(s, t)$ estimado com $J = 30$ e $\hat{\lambda} = 1,28 \cdot 10^{-5}$

Considere as estimativas de $\hat{\sigma}_c(s, t)$ exibidas na Figura 3.7, bem como as diagonais, que representam a função variância, na Figura 3.8. Foram usadas $K = 36^2$ bases, em $L = 32^2$ nós internos equispaçados. O fato das duas funções serem negativas em alguns intervalos já é suficiente para descartar a hipótese de serem estimativas de covariância. Analisando os autovalores da matriz gerada por $\sigma_1(s, t)$ calculada nos pontos $t_1, \dots, t_{48}$, obtemos que o menor deles é $\hat{S}_{48} = -1,17$. Claramente a estimativa não é positiva

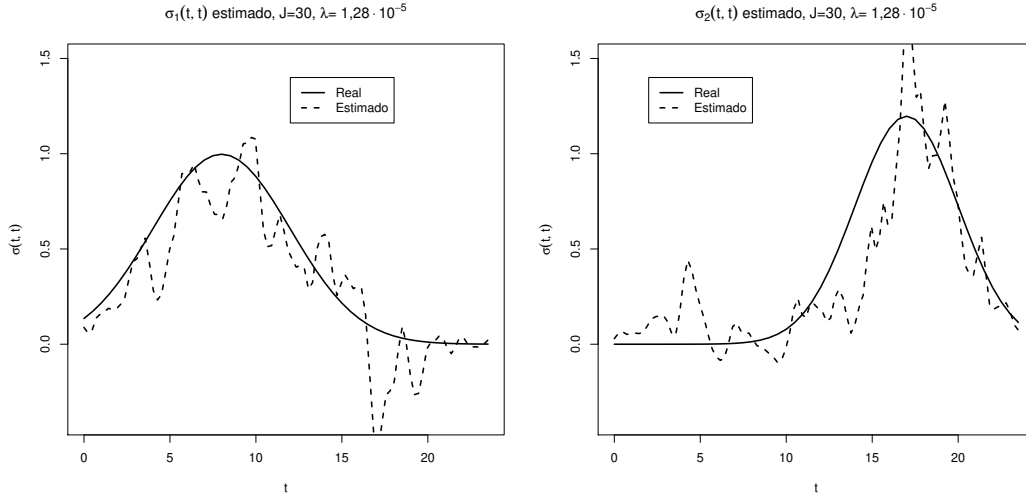


Figura 3.8: $\hat{\sigma}_c(t, t)$ estimado com $J = 30$ e $\hat{\lambda} = 1,28 \cdot 10^{-5}$

definida (sequer não negativa definida).

A proposta é assumir que os dados tenham covariância da forma $\eta_c(s)\eta_c(t)\varphi_c(t-s)$, em que φ_c seja real, simétrica e definida positiva, como postulado no Teorema de Bochner. De fato, os dados foram simulados com $\varphi_c(\tau) = \exp(-\phi_c|\tau|)$, que tem Transformada de Fourier dada por $2\phi_c(\phi_c^2 + \omega^2)^{-1}$. Mas isso não será considerado, preservando o caráter não paramétrico da metodologia. Partindo dessa hipótese, temos que $\hat{\rho}_c(s, t) = \hat{\sigma}_c(s, t)\hat{\eta}_c^{-1}(s)\hat{\eta}_c^{-1}(t)$ é função de covariância de um processo estacionário, amostrado numa janela $[0, T]$, e que conseqüentemente deve satisfazer (3.6).

A Figura 3.9 exhibe as funções resultantes do processo de padronização. $|\hat{\eta}_c(t)|$ foi estimado como a raiz quadrada do valor absoluto de $\sigma_c(t, t)$; a estimativa toda foi truncada para que $\hat{\rho}_c(s, t) \leq 1$ se $s \neq t$ e $\hat{\rho}_c(s, t) = 1$ se $s = t$. Finalmente as estimativas foram filtradas por suavização spline simples (sem tratar de modelos agregados). As regiões onde $\sigma_c(t, t)$ era negativo ou próximo de zero apresentaram muita instabilidade nesse processo, que ainda está visível na Figura 3.9 (por exemplo para valores próximos de $t = 16$ em $\sigma_1(t, t)$, e $t = 10$ em $\sigma_2(t, t)$).

A Transformada de Fourier das estimativas, já transladada por $e^{iT/2(w_1+w_2)}$, encontra-

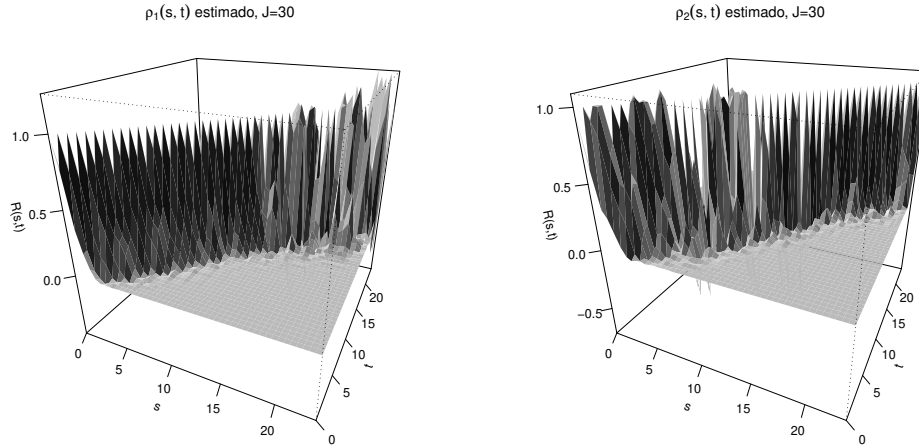


Figura 3.9: $\hat{\rho}_1(s, t)$ e $\hat{\rho}_2(s, t)$, após truncamento e suavização

se nas Figuras 3.10 e 3.11. A operação de padronização por

$$\left(\frac{\sin(2^{-1}(w_1 + w_2))}{(2^{-1}(w_1 + w_2))} \right)^{-1}$$

foi trocada por simplesmente multiplicar a transformada pelo sinal dessa função. De outra forma os zeros de $\sin(2^{-1}(w_1 + w_2)) / (2^{-1}(w_1 + w_2))$ causam problemas numéricos. O procedimento seguinte foi truncar tanto toda a parte Imaginária, quanto os valores negativos da parte Real. A função resultante foi transladada de volta a $[0, T]^2$ no domínio de tempo multiplicando-se a transformada por $e^{-iT/2(w_1 + w_2)}$.

A função inversa foi obtida numericamente, e multiplicada pelas respectivas estimativas de $|\hat{\eta}_c(t)|$, agora truncadas no zero. As estimativas resultantes encontram-se na Figura 3.12. Sua diagonal encontra-se na Figura 3.13. Inspecionar os autovalores das estimativas resultantes ainda revela autovalores negativos – porém, da ordem de 10^{-15} , o que significa que possivelmente ocorreram por causa de erros de aproximação do algoritmo de decomposição em autovalores. Um problema que limita o uso da metodologia é o fato das estimativas serem apenas não negativas definidas, ao invés de positivas definidas, o que seria uma propriedade bastante desejável, necessária para construção das matrizes inversas (matrizes não negativas definidas podem ser singulares). Mas funções

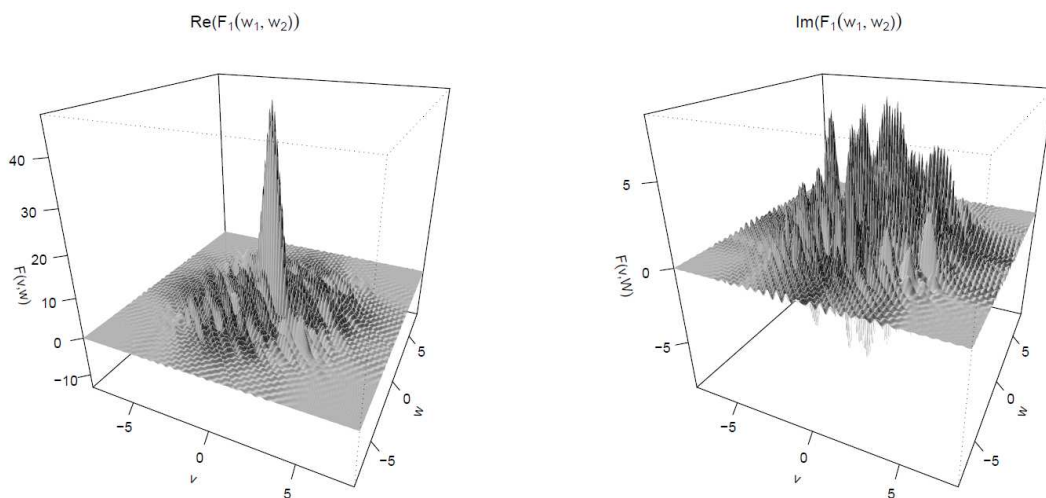


Figura 3.10: Parte Real e Imaginária de $\hat{\mathcal{F}}_1(w_1, w_2)$

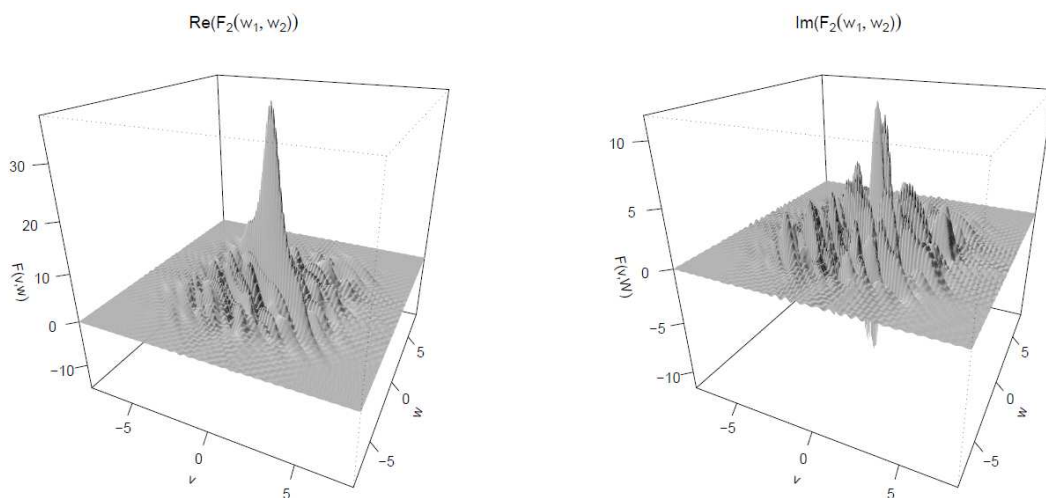


Figura 3.11: Parte Real e Imaginária de $\hat{\mathcal{F}}_2(w_1, w_2)$

assim são muito difíceis de obter-se. Note que mesmo as estimativas semiparamétricas da Seção 2.3 possuem autovalores próximos de zero (isto é, de ordem 10^{-15}).

É possível construir regiões de confiança para as estimativas, usando as fórmulas (1.11) e (1.12). Contudo, essas regiões são formalmente *erradas*: elas partem da hipótese de que o erro de aproximação entre a estimativa e o valor verdadeiro são normais – o que não é o caso quando se trata de variância – e mesmo que se aplique um procedimento de transformada de Fourier nas fronteiras da região, similar ao aplicado

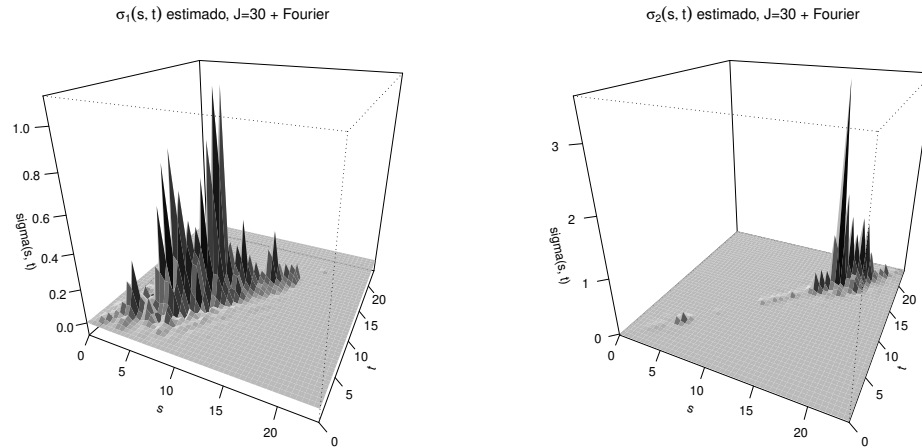


Figura 3.12: $\hat{\sigma}_1(s, t)$ e $\hat{\sigma}_2(s, t)$ estimados com $J = 30$, após correção da sua Transformada de Fourier

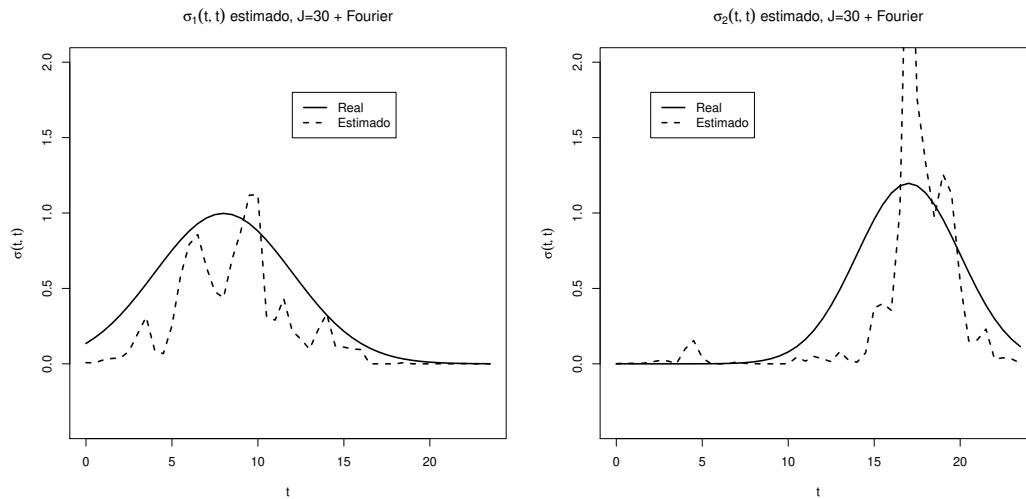


Figura 3.13: As variâncias $\hat{\sigma}_1(t, t)$ e $\hat{\sigma}_2(t, t)$, estimadas com $J = 30$, após correção da sua Transformada de Fourier

nas estimativas, nada garante que qualquer superfície dentro desses limites seja uma função positiva definida. Por força disto, não serão apresentadas as regiões de confiança para as estimativas neste Capítulo.

Capítulo 4

Estimação da Covariância pela sua Representação Espectral

A ideia do método desenvolvido neste capítulo baseia-se no artigo de Lee, Zhang & Song (2002), que realiza a estimação dos autovalores e autofunções do processo através da matriz de covariância amostral. Para tanto, os autores baseiam-se em uma representação espectral do processo fornecida pelo *teorema de Karhunen-Loève* (Yaglom 1987, p. 448). O teorema determina, entre outros resultados, uma representação da função de covariância em uma combinação linear de funções determinísticas (chamadas autofunções, pela analogia com os autovetores do caso matricial) e autovalores. Sob a restrição de homogeneidade das autofunções entre as diferentes curvas agregadas, o algoritmo pode ser adaptado para obter estimativas que separam as funções de covariância individuais a partir das matrizes de covariância agregadas.

4.1 O Teorema de Karhunen-Loève

Seja $X(t)$ um processo estocástico de média zero, definido em algum intervalo fechado (sem perda de generalidade toma-se o intervalo $[0, T]$), e que seja contínuo por média

quadrática, isto é, $\mathbb{E}(|X(t)|^2) < \infty$ e

$$\lim_{t \rightarrow s} \mathbb{E}(|X(t) - X(s)|^2) = 0.$$

Na notação usual, $\rho(s, t) = \text{Corr}(X(s), X(t))$ é a função de correlação do processo. Considere a equação integral

$$\int_0^T \rho(t, s) \phi(s) ds = \gamma \phi(t), \quad t \in [0, T] \quad (4.1)$$

A fórmula (4.1) é uma equação linear integral com um kernel simétrico e não negativo definido. Então o teorema afirma que existe um conjunto enumerável de autovalores não negativos γ_ν , $\nu = 1, 2, \dots$ associados a um conjunto de autofunções ortonormais $\phi_\nu(t)$ tais que

$$\rho(t, s) = \sum_{\nu=1}^{\infty} \gamma_\nu \phi_\nu(t) \phi_\nu(s). \quad (4.2)$$

Note que a representação impõe uma condição de ortogonalidade em $\phi_\nu(t)$, isto é,

$$\int_0^T \phi_\nu(s) \phi_\mu(t) dt = 0 \quad \text{se} \quad \mu \neq \nu,$$

$$\int_0^T \phi_\nu(s) \phi_\nu(t) dt = 1.$$

Uma consequência da representação (4.2) é que se torna possível representar o processo $X(t)$ através da equação

$$X(t) = \sum_{\nu=1}^{\infty} \gamma_\nu^{1/2} \phi_\nu(t) \eta_\nu, \quad (4.3)$$

onde η_ν são variáveis aleatórias não correlacionadas, e que não dependem de t . O inverso também é verdadeiro, isto é, processos na forma (4.3) também tem covariância na forma (4.2). A decomposição espectral de matrizes simétricas e positivas definidas corresponde ao caso vetorial da forma (4.2). Deve-se ressaltar que, bem como na análise de componentes principais, se dois autovalores forem idênticos então a decomposição (4.2) não é única. Isto é, se $\lambda_{n_1} = \lambda_{n_2}$, então é possível trocar a ordem de $\phi_{n_1}(t)$ e $\phi_{n_2}(t)$, precisamente da mesma forma que é possível permutar os autovetores da decomposição

de matrizes num caso análogo. Mas se os autovalores são distintos, então todas as autofunções são identificáveis exceto por uma mudança de sinal, como descrito em Lee et al. (2002).

Note que ainda em consequência do teorema η_ν podem ser determinados por

$$\eta_\nu = \gamma_\nu^{-1/2} \int_0^T X(t) \phi_\nu(t) dt, \quad (4.4)$$

que são variáveis aleatórias relacionadas com os componentes principais do caso matricial.

4.2 Aplicação em Dados Agregados

Ao longo do Capítulo, usaremos $N_{i,c}$ como referência; os resultados são facilmente adaptados para $q_{i,c}^2$, fazendo apenas uma substituição dos coeficientes onde for necessário. Considere a representação de Karhunen-Loève de $\sigma_c(s, t)$, dada por

$$\sigma_c(s, t) = \sum_{\nu=1}^n \gamma_{\nu,c} \phi_\nu(s) \phi_\nu(t), \quad (4.5)$$

em que $\gamma_{\nu,c}$ são os autovalores e $\phi_\nu(t)$ as autofunções associadas. A representação é truncada para um número finito de autovalores por razões sobretudo computacionais – não é possível obter mais autofunções do que o número de pontos em que a curva foi observada. O texto de Ramsay & Silverman (2005) discute o problema da representação finita. Assume-se que $\phi_\nu(t)$ não depende de c , o que permite que a variância da curva agregada seja dada por

$$\mathbf{Z}_i(s, t) = \sum_{c=1}^C N_{i,c} \sigma_c(s, t) = \sum_{\nu=1}^n \gamma_\nu^{(i)} \phi_\nu(s) \phi_\nu(t),$$

onde

$$\gamma_\nu^{(i)} = \sum_{c=1}^C N_{i,c} \gamma_{\nu,c}.$$

Além disso, a independência de $\phi_\nu(t)$ com relação à subpopulação c implica que a soma de todas as curvas agregadas também tem função de covariância dada pela

expansão

$$\mathbf{Z}(s, t) = \sum_{i=1}^I \mathbf{Z}_i(s, t) = \sum_{i=1}^I \sum_{\nu=1}^n \gamma_{\nu}^{(i)} \phi_{\nu}(s) \phi_{\nu}(t).$$

A hipótese de autofunções idênticas entre as curvas individuais é, portanto, essencial aos resultados que serão apresentados, pois a decomposição espectral da combinação linear de funções não tem forma simples no caso mais geral. Isso implica que as componentes principais de variação devem ter aproximadamente a mesma forma: nos cenários simulados nos Capítulos 2 e 3, esse método não funciona pois é incapaz de lidar com os picos da variância em pontos diferentes de $[0, T]$.

O método consiste em estimar $\phi_{\nu}(t)$ a partir dos autovetores de $\mathbf{Z}$, a matriz de covariância amostral da soma das curvas agregadas. Este procedimento foi proposto para dados funcionais simples por Lee et al. (2002). Primeiramente considere as matrizes de covariância amostral, dadas por

$$\hat{\mathbf{Z}}_i = \sum_{j=1}^J \frac{(Y_{i,j}(s_r) - \hat{Y}_i(s_r))(Y_{i,j}(t_r) - \hat{Y}_i(t_r))}{J}.$$

O estimador de $\mathbf{Z}$ é simplesmente

$$\hat{\mathbf{Z}} = \sum_{i=1}^I \hat{\mathbf{Z}}_i(t, s).$$

O algoritmo para determinar as autofunções e seus autovalores é dado a seguir:

Passo 1: Calcule

$$\hat{\gamma}_1 \geq \hat{\gamma}_2 \geq \dots \hat{\gamma}_n$$

os autovalores da matriz $\hat{\mathbf{Z}}$, bem como seus autovetores correspondentes

$$\hat{\phi}_1, \hat{\phi}_2, \dots, \hat{\phi}_n.$$

Passo 2: Suavize os autovetores usando suavização spline, brevemente introduzidos na Seção 3.2, obtendo estimativas das autofunções do processo

$$\tilde{\phi}_1(\cdot), \tilde{\phi}_2(\cdot), \dots, \tilde{\phi}_n(\cdot).$$

É preciso ressaltar que a normalização do autovetor não implica na normalização das curvas com respeito ao produto interno L^2 . Neste caso, as estimativas das autofunções devem ser normalizadas de modo que

$$\int_0^T \left(\tilde{\phi}_\nu(t) \right)^2 dt = 1, \quad \nu = 1, 2, \dots, n.$$

Passo 3: Estime os autovalores $\gamma_\nu^{(i)}$ para cada curva agregada usando o fato que

$$\gamma_\nu^{(i)} = \text{Var} \left(\eta_\nu^{(i)} \right),$$

onde, adotando a notação $\langle \cdot \rangle$ para o produto interno entre funções no intervalo fechado $[0, T]$,

$$\eta_\nu^{(i)} = \int_0^T \left(Y_{i,j}(t) - \hat{Y}_{i,j}(t) \right) \phi_j(t) dt = \left\langle Y_{i,j}(t) - \hat{Y}_{i,j}(t), \phi_j(t) \right\rangle. \quad (4.6)$$

As variáveis aleatórias $\eta_\nu^{(i)}$ são desconhecidas e $Y_{i,j}(t)$ foi somente observada nos pontos $t_1, \dots, t_n$, então deve-se usar como “proxy”, para $j = 1, \dots, J$,

$$\tilde{\eta}_{\nu,j}^{(i)} = \sum_{r=1}^n \left(Y_{i,j}(t_r) - \hat{Y}_{i,j}(t_r) \right) \tilde{\phi}_\nu(t_r) \Delta t_r,$$

em que $\Delta t_r = |t_{r+1} - t_{r-1}|/2$. Com as J replicações pode-se estimar diretamente a variância através de

$$\tilde{\gamma}_\nu^{(i)} = \frac{1}{J} \sum_{j=1}^J \left(\tilde{\eta}_{\nu,j}^{(i)} - \frac{1}{J} \sum_{j=1}^J \tilde{\eta}_{\nu,j}^{(i)} \right)^2. \quad (4.7)$$

Passo 4: Munido das estimativas dos autovalores $\tilde{\gamma}_\nu^{(1)}, \dots, \tilde{\gamma}_\nu^{(I)}$ das curvas agregadas, para cada uma das autofunções $\tilde{\phi}_\nu(t)$, é possível estimar os autovalores $\hat{\gamma}_{\nu,c}$ das curvas individuais resolvendo um sistema linear, fixado ν :

$$\begin{cases} \tilde{\gamma}_\nu^{(1)} &= N_{1,1} \hat{\gamma}_{\nu,1} + \dots + N_{1,C} \hat{\gamma}_{\nu,C} \\ &\vdots \\ \tilde{\gamma}_\nu^{(i)} &= N_{i,1} \hat{\gamma}_{\nu,1} + \dots + N_{i,C} \hat{\gamma}_{\nu,C} \\ &\vdots \\ \tilde{\gamma}_\nu^{(I)} &= N_{I,1} \hat{\gamma}_{\nu,1} + \dots + N_{I,C} \hat{\gamma}_{\nu,C} \end{cases} \quad (4.8)$$

com a restrição que $\gamma_{\nu,c} > 0$.

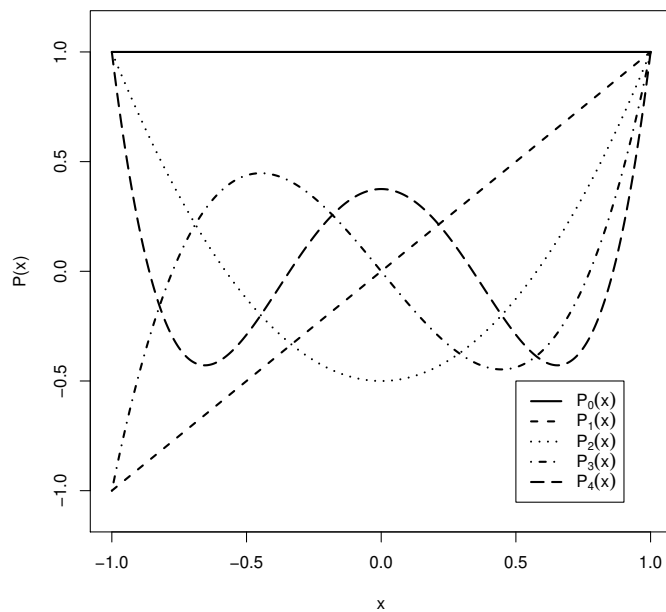


Figura 4.1: Polinômios de Legendre $P_0(x)$ a $P_4(x)$

4.3 Exemplo Simulado

Os polinômios de Legendre são uma família de funções polinomiais ortogonais em $[-1, 1]$, definidos pela fórmula

$$\begin{aligned} P_0(x) &= 1 \\ P_1(x) &= x \\ (n+1)P_{n+1}(x) &= (2n+1)xP_n(x) - nP_{n-1}(x). \end{aligned}$$

É possível verificar que o produto interno L^2 entre os polinômios é dado por

$$\int_{-1}^1 P_m(x)P_n(x)dx = \frac{2}{2n+1}\delta_{m,n},$$

em que $\delta_{m,n} = 1$ se $m = n$ e 0 caso contrário. Portanto $(n+1/2)^{1/2}P_n(x)$ são ortogonais. Os primeiros 5 polinômios de Legendre estão na Figura 4.1. Com base neles,

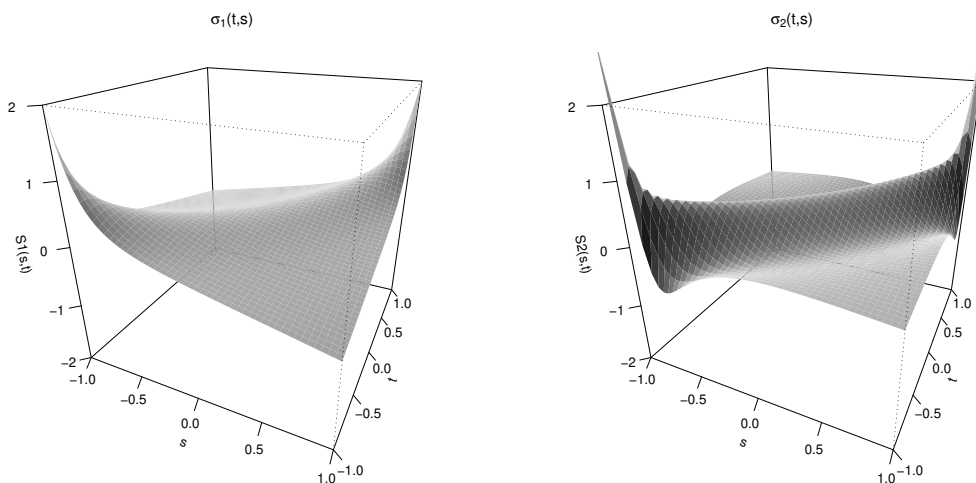


Figura 4.2: Funções de covariância de $W_1(t)$ e $W_2(t)$

define-se os processos $W_c(t)$, $c = 1, 2$ por

$$W_c(t) = \sum_{\nu=1}^n \gamma_{\nu,c}^{1/2} (\nu + 1/2)^{1/2} P_\nu(t) \eta_\nu,$$

onde η_ν são variáveis aleatórias independentes $N(0, 1)$ e os autovalores $\gamma_{\nu,c}$ são dados por

$$\begin{aligned} \gamma_{1,1} &= \gamma_{1,2} = 0 \\ \gamma_{\nu,1} &= 2^{-(\nu-1)}, \quad \nu = 2, \dots, n \\ \gamma_{\nu,2} &= (\nu - 1)^{5/2} \cdot \left(\frac{2}{5}\right)^\nu, \quad \nu = 2, \dots, n. \end{aligned}$$

As funções de covariância resultantes são mostradas na Figura 4.2, em um gráfico de perspectiva. Note que elas são não estacionárias, pois a função variância $\sigma_c(t, t)$ não é constante. As curvas geradas foram agregadas usando a mesma alocação do Capítulo 1, dada por

$$\mathbf{N}_{3 \times 2} = \begin{pmatrix} 20 & 10 \\ 15 & 15 \\ 10 & 20 \end{pmatrix}.$$

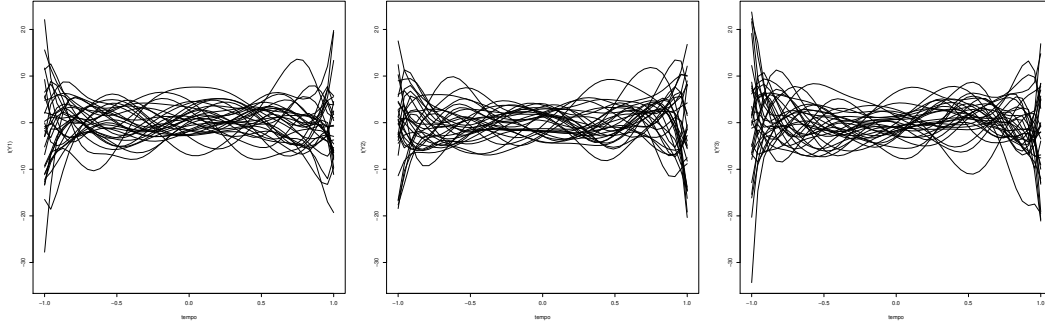


Figura 4.3: Curvas simuladas $y_1(t)$, $y_2(t)$ e $y_3(t)$, $J = 30$

A Figura 4.3 exibe 30 curvas agregadas, geradas através do modelo dado pela Equação (4.3), amostradas em $n = 48$ pontos. Como $\eta_\nu \sim N(0, 1)$, os processos são consequentemente Gaussianos. O experimento foi realizado com $J = 30$ e $J = 100$ replicações.

A Figura 4.4 exibe as estimativas das autofunções – a linha sólida representa a autofunção verdadeira, a tracejada representa a estimativa com $J = 30$ replicações, e a pontilhada com $J = 100$ – obtidas empregando splines cúbicos, com 25 bases equidistantes, e com parâmetro de suavização dado pelo valor que minimiza o GCV da estimativa. Note ainda que ao invés de $\hat{\phi}_3(t)$, para $J = 100$ replicações, e $\hat{\phi}_2(t)$, para $J = 30$ e $J = 100$ replicações, são exibidas $-\hat{\phi}_2(t)$ e $-\hat{\phi}_3(t)$, o que não afeta as estimativas de covariância, como foi discutido no final da Seção 4.1.

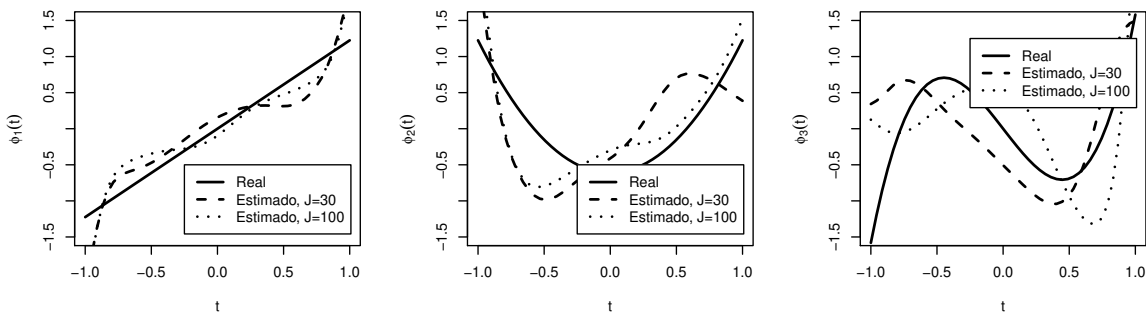


Figura 4.4: As três primeiras autofunções obtidas de $\hat{\mathbf{Z}}$

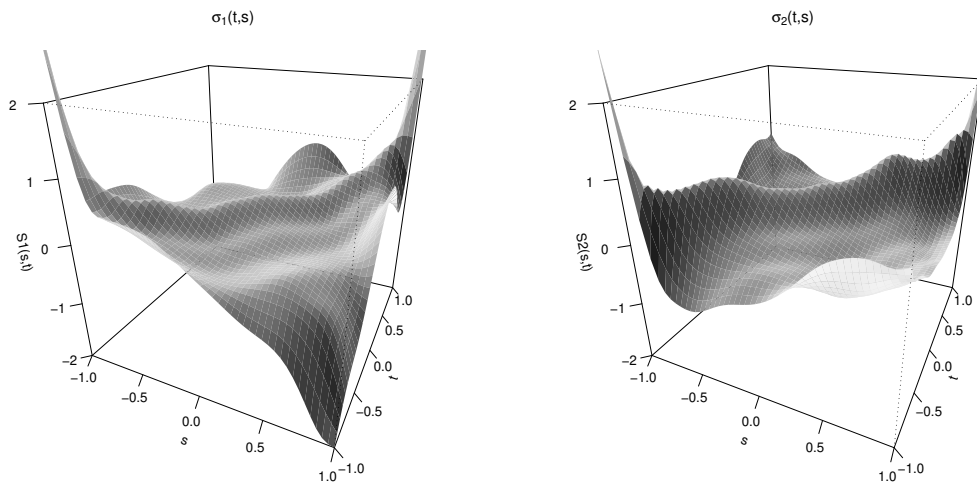


Figura 4.5: Estimativas de $\sigma_1(s, t)$ e $\sigma_2(s, t)$, $J = 30$

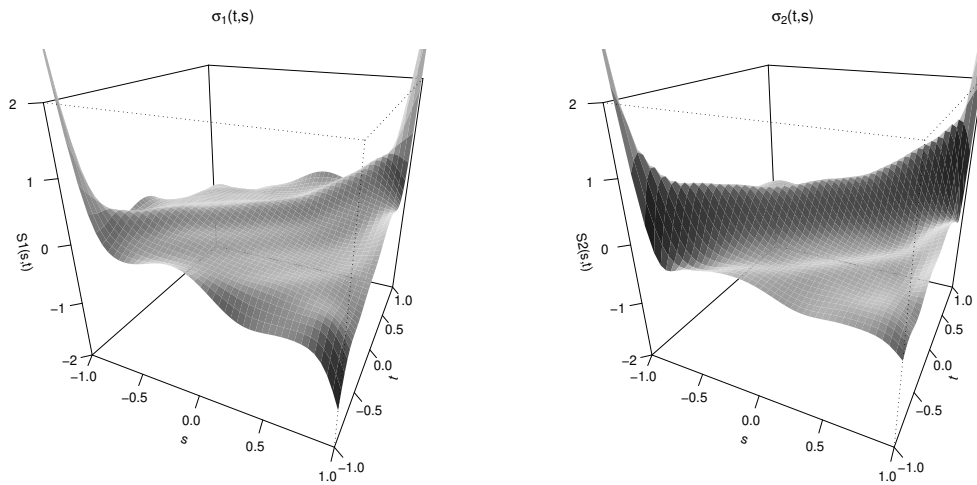


Figura 4.6: Estimativas de $\sigma_1(s, t)$ e $\sigma_2(s, t)$, $J = 100$

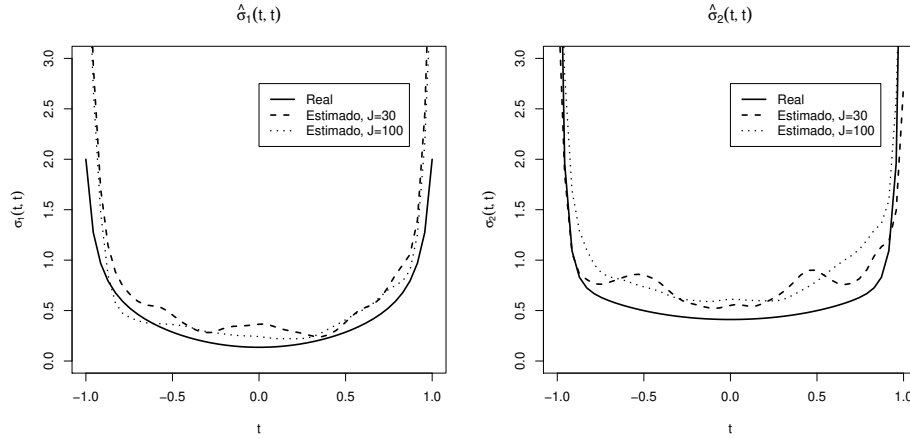


Figura 4.7: Estimativas de $\sigma_1(t, t)$ e $\sigma_2(t, t)$

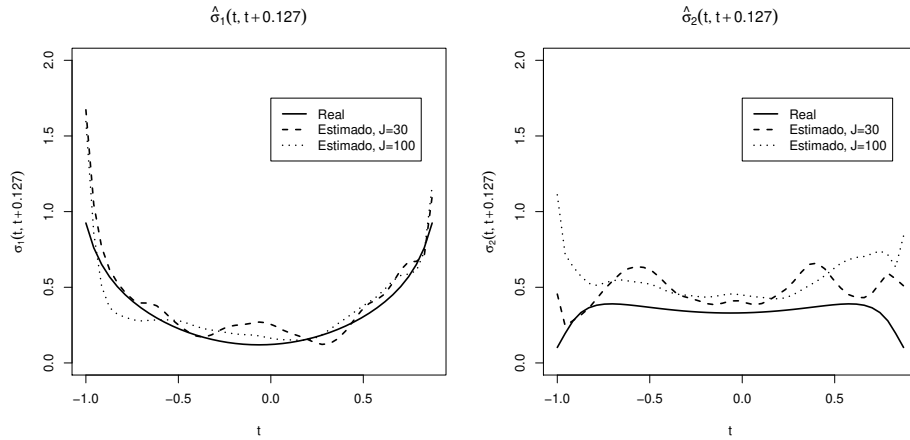


Figura 4.8: Estimativas de $\sigma_1(t, t + 0,125)$ e $\sigma_2(t, t + 0,125)$

As estimativas de $\sigma_c(s, t)$ foram feitas usando os primeiros 14 autovalores (correspondendo a 99,996% e 99,982% da variação total de $W_1(t)$ e $W_2(t)$, respectivamente). De fato, os parâmetros de suavização que minimizam o GCV são muito pequenos para cada uma das primeiras 14 autofunções, todos em torno de $2 \cdot 10^{-4}$, e muito grandes para as autofunções subsequentes (todos maiores que 10^{11}). Como quando a penalização por não suavidade é muito grande as estimativas se aproximam de uma reta, pode-se utilizar o próprio GCV para deduzir um critério de parada para o número de autofunções

a serem empregadas. Gráficos de perspectiva das estimativas resultantes são exibidos nas Figuras 4.5 e 4.6, para $J = 30$ e $J = 100$ replicações, respectivamente.

Para facilitar a comparação entre estimativas e os valores verdadeiros das funções $\sigma_c(s, t)$, as Figuras 4.7 e 4.8 exibem, respectivamente, a função variância $\sigma_c(t, t)$ e a covariância entre $W_c(t)$ e $W_c(t + 0,125)$, dada por $\sigma_c(t, t + 0,125)$.

Para avaliar a incerteza com respeito às estimativas, serão propostas regiões de confiança para as estimativas, baseadas no seguinte argumento: suponha que as estimativas das autofunções foram obtidas sem erro (ou que são conhecidas). Valendo essa hipótese, todas as variáveis $\eta_\nu^{(i)}$ são independentes, então o proxy $\tilde{\eta}_{\nu,j}^{(i)}$ tem distribuição Normal padrão. Consequentemente, para a variância amostral vale

$$\frac{J\tilde{\gamma}_\nu^{(i)}}{\gamma_\nu^{(i)}} \sim \chi_{J-1}^2.$$

É fácil ver que isso implica em uma subestimação da incerteza sobre as estimativas, pois as estimativas de ϕ_ν tem erro. Um argumento assintótico, contudo, pode ser desenvolvido, justificando a convergência de $\tilde{\phi}_\nu(t)$ para $\phi_\nu(t)$ e consequentemente a utilização do procedimento proposto.

Para obter a região de confiança para $\sigma_c(s, t)$, constrói-se os intervalos de confiança para $\gamma_\nu^{(i)}$, a $0,9^{1/n}100\%$ de confiabilidade por exemplo (de modo que a região resultante tenha aproximadamente 90% de confiabilidade), e separa-se os autovalores individuais usando as extremidades do intervalo para cada $\gamma_\nu^{(i)}$. As Figuras 4.9 e 4.10 mostram as regiões de confiança resultantes, quando $J = 30$ e $J = 100$, respectivamente. Além disso, as Figuras 4.11 e 4.12 mostram a região de confiança para $\sigma_c(t, t + 0,125)$, com $J = 30$ e $J = 100$, respectivamente.

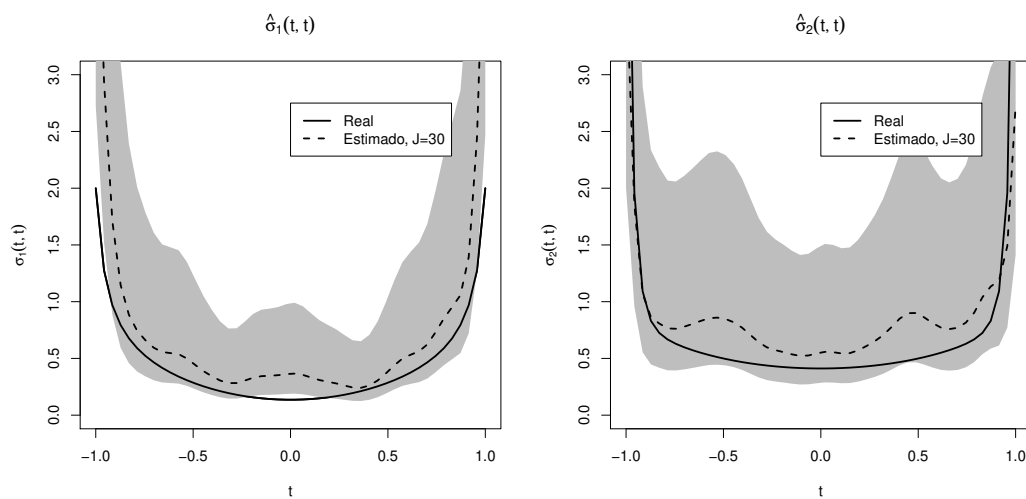


Figura 4.9: Região de confiança (aproximadamente 90% de confiabilidade) para $\sigma_c(t, t)$, com $J = 30$

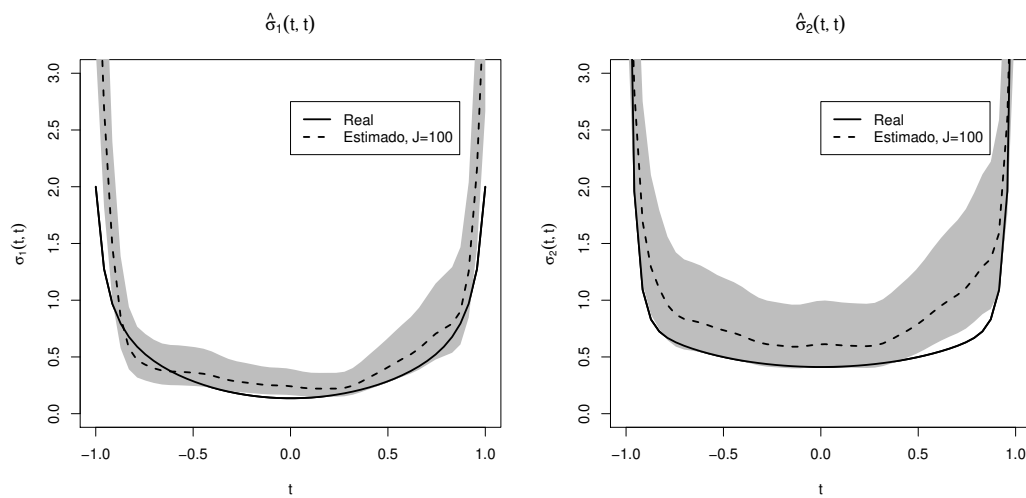


Figura 4.10: Região de confiança (aproximadamente 90% de confiabilidade) para $\sigma_c(t, t)$, com $J = 100$

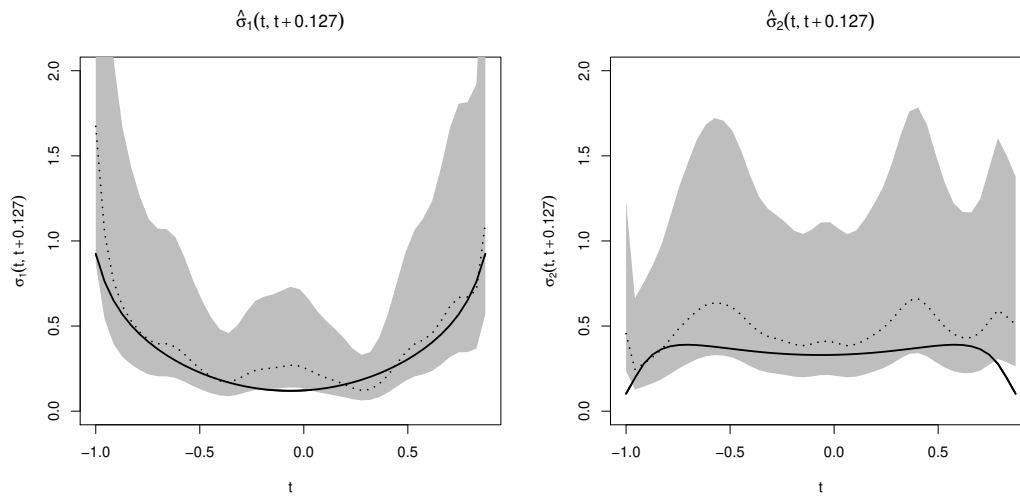


Figura 4.11: Região de confiança (aproximadamente 90% de confiabilidade) para $\sigma_c(t, t + 0,125)$, com $J = 30$

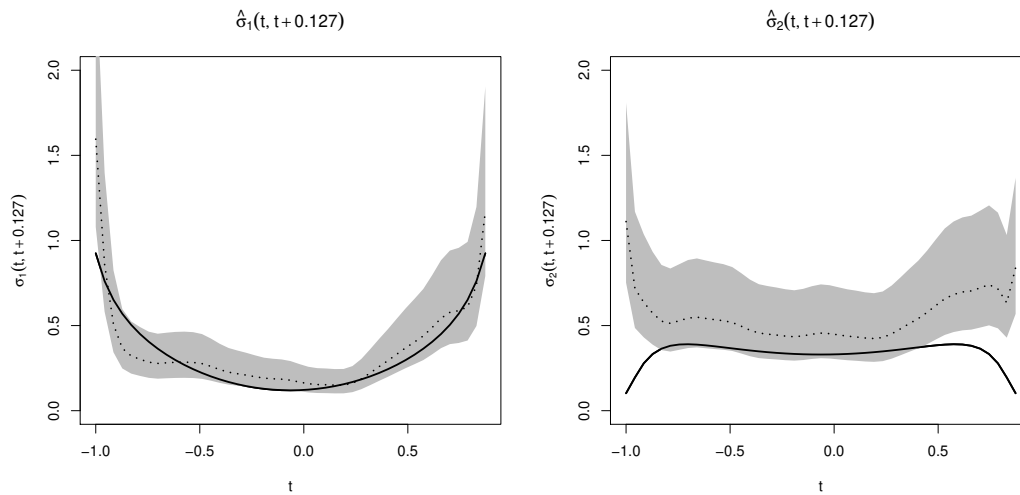


Figura 4.12: Região de confiança (aproximadamente 90% de confiabilidade) para $\sigma_c(t, t + 0,125)$, com $J = 100$

Capítulo 5

Aplicação em Dados Reais

Este capítulo aplica as metodologias desenvolvidas para a estimação de covariância em dois problemas reais: o de consumidores de energia elétrica e o de espectrometria. Neste caso, o cenário de energia elétrica serve para ilustrar o caso onde as curvas são agregadas por somas de curvas individuais, com coeficientes denotados ao longo da dissertação por $N_{i,c}$. As curvas individuais representam as residências; a curva agregada, o transformador de energia ou mesmo uma subestação. A aplicação em quimiometria serve para exemplificar curvas agregadas por combinações lineares. Neste cenário $q_{i,c}$ representa a concentração da c -ésima componente da mistura i , e conseqüentemente sua contribuição à transmitância/absorbância amostrada.

5.1 Consumo de Energia Elétrica

O problema dos dados agregados foi encontrado originalmente no consumo de energia elétrica. Um sistema perfeito de distribuição de energia deveria ter as estações, subestações e transformadores com uma demanda constante durante todo o dia, todos os dias do ano. Este é um cenário, entretanto, impossível, pois há horários de pico no consumo de energia – devido, por exemplo, ao uso de chuveiros elétricos. A fim de prevenir sobrecargas, a rede de distribuição é superdimensionada para lidar com o consumo máximo sem considerar os diferentes tipos de consumidores que fazem parte

desta rede. Um modo fácil de otimizar os recursos já existentes seria redistribuir os consumidores de tal modo que as estações e subestações tivessem um consumo mais homogêneo ao longo do dia, evitando picos de consumo. Para atingir este objetivo é necessário identificar os perfis de consumo de cada tipo de consumidor, isto é, a curva típica ou *tipologia*.

A medição do consumo diário de cada habitação é, no entanto, muito cara, senão impossível. Além disso, a variabilidade individual seria muito grande (no jargão de engenharia, com uma taxa *signal-to-noise* muito baixa). A solução encontrada pelas companhias de energia elétrica é amostrar os transformadores (*trafos*), onde são captados os sinais de todo um bairro ou vizinhança. A Figura 5.1 mostra aproximadamente trinta dias amostrados pela Companhia Elétrica de Força e Luz (CPFL) de uma subestação de Campinas. As curvas tracejadas representam os sábados, enquanto as curvas pontilhadas os domingos. A queda acentuada das dez da manhã ao meio-dia denota um *blackout*.

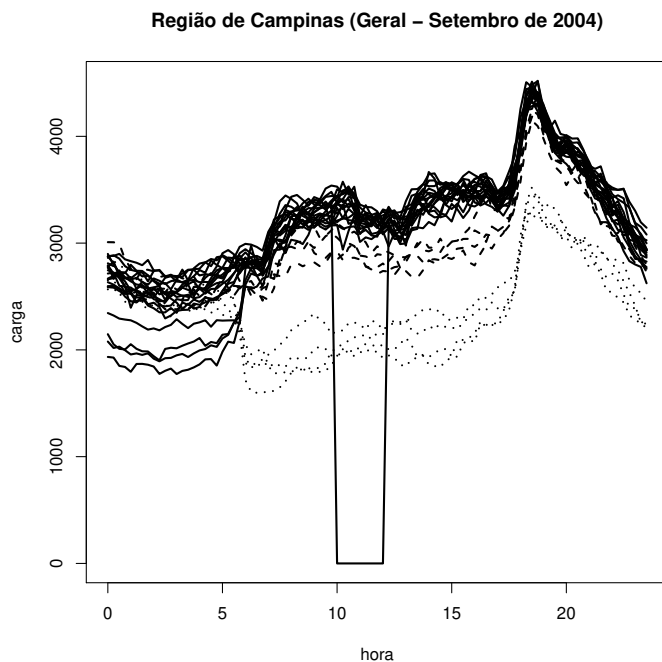


Figura 5.1: Curvas agregadas amostradas em certa região de Campinas.

Dias et al. (2009) resolvem o problema de estimação da tipologia escrevendo as curvas $y(t)$ observadas como na notação apresentada para dados agregados por soma de $N_{i,c}$ indivíduos, isto é,

$$Y_{ij}(t) = \sum_{c=1}^C \sum_{n_c=1}^{N_{i,c}} \alpha_c(t) + \sum_{c=1}^C \sum_{n_c=1}^{N_{i,c}} \varepsilon_{i,c,j,n_c}(t),$$

onde i é o i -ésimo transformador, j o j -ésimo dia útil, c o tipo de consumidor, $N_{i,c}$ o número de consumidores do tipo c no i -ésimo trafo, e $\alpha_c(t)$ sua respectiva tipologia. $\varepsilon_{i,c,j,n_c}(t)$ é um processo gaussiano de média 0, que assumimos independente de j ; contudo, a variabilidade parece mudar durante as horas do dia, o que é uma evidência de não estacionariedade dos erros. Escrevendo

$$\alpha_c(t) = \sum_{k=1}^K \beta_{c,k} B_k(t),$$

chega-se à expressão

$$Y_{i,j}(t) = \sum_{c=1}^C \sum_{k=1}^K N_{i,c} \beta_{c,k} B_k(t) + \varepsilon_{i,j}(t),$$

onde $\varepsilon_{i,j}(t)$ é a soma em n_c e i de $\varepsilon_{c,j,n_c,i}(t)$. E desde que $I \geq C$ e o vetor de consumidores em cada trafo seja linearmente independente dos outros trafos, o problema é identificável e a estimativa é obtida por quadrados mínimos generalizados. Uma alternativa Bayesiana de estimação é desenvolvida em Dias et al. (2011).

5.1.1 Modelagem de Tipologia

Serão reproduzidos aqui resultados de Dias et al. (2009). As curvas são observações de cinco dias úteis nos transformadores TR07**** e TR09**** – a numeração completa foi omitida a pedido da CPFL. As amostras foram colhidas de 15 em 15 minutos, resultando em $n = 96$ pontos amostrais. As subpopulações são consumidores residenciais monofásicos e bifásicos: TR07**** tem 87 monofásicos e 5 bifásicos, TR09**** tem 25 monofásicos e 25 bifásicos. Na aplicação do método de estimação de tipologia, foi

utilizada uma matriz banda na forma (1.10) com $\gamma = 6$; o algoritmo precisou alternar a estimação de tipologia com a estimação de covariância amostral por 32 iterações.

As estimativas das curvas agregadas, bem como as observações agregadas (em tom cinza), encontram-se na Figura 5.2. A Figura 5.3 mostra as estimativas de tipologia dos consumidores residenciais monofásicos e bifásicos, respectivamente, que habitam na região destes transformadores.

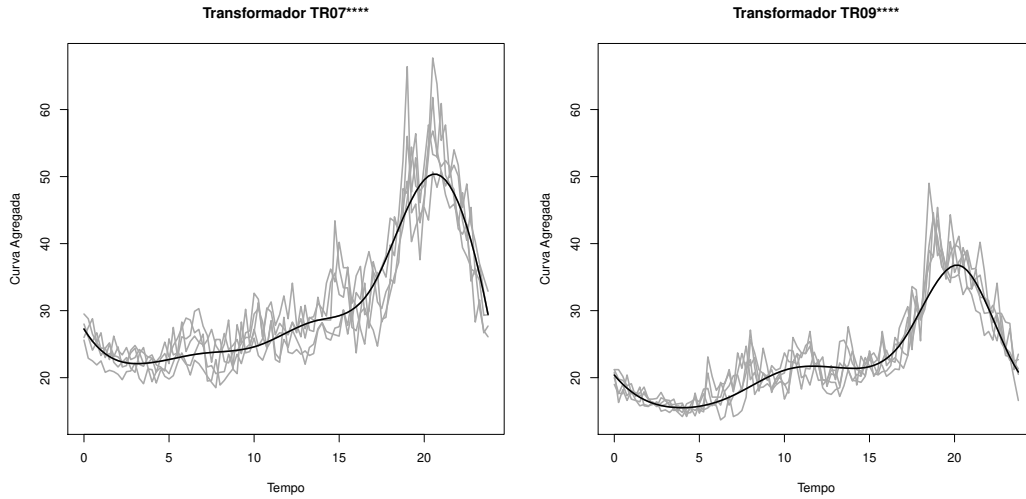


Figura 5.2: As 5 amostras dos Transformadores TR07**** e TR09****, e suas estimativas.

5.1.2 Modelagem da Função de Covariância Típica

Para a modelagem da covariância típica desses consumidores, serão empregadas e comparadas duas das metodologias desenvolvidas: a semiparametrização da covariância (cuja estimação Bayesiana foi desenvolvida em Dias et al. (2011)) e a estimação por produto tensorial de Splines. Finalmente uma aplicação da função de covariância – na construção de intervalos de confiança para as estimativas de tipologia – será exibida no final da subseção.

A Figura 5.4 mostra as estimativas de função de variância, utilizando o método de semiparametrização. As estimativas para ϕ foram de $\hat{\phi}_1 = 2,29$ e $\hat{\phi}_2 = 2,76$, respecti-

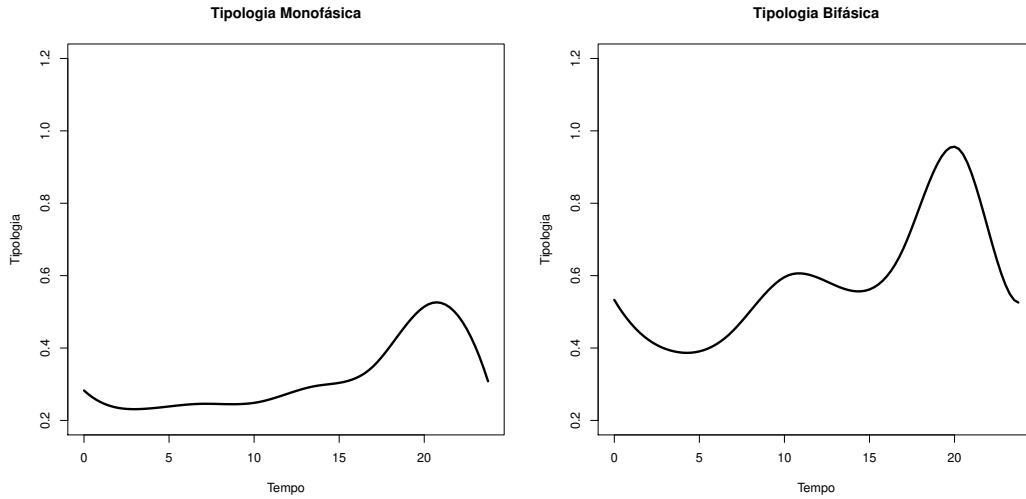


Figura 5.3: Tipologias estimadas dos consumidores residenciais monofásicos e bifásicos, de uma região central de Campinas.

vamente para os consumidores monofásico e bifásico. Isso implica que a autocorrelação fica inferior a 0,05 somente após cerca uma hora. De fato, a Figura 5.5 mostra a estimativa da função de covariância dos processos após um intervalo de meia hora, que já praticamente se extinguiu exceto nos períodos de final de tarde. Repare ainda que as estimativas de $\eta_c^2(t)$ obtidas por máxima verossimilhança são bastante similares às aquelas obtidas por métodos bayesianos no artigo de Dias et al. (2011), com o mesmo banco de dados e igual semiparametrização da estrutura de covariância.

A metodologia de produto tensorial de splines, apresentada na Seção 3.2, também foi empregada. No entanto, limitações computacionais com a alocação de memória exigiram que o número máximo de nós internos utilizado fosse de 40^2 , e não os 94^2 possíveis. Como as estimativas suavizadas com esse número de nós resultavam num GCV muito baixo, optou-se pelo produto tensorial de bases ordinárias. Convém repetir que menos bases subestimam muito a variância, como foi discutido no Capítulo 3.

A Figura 5.6 mostra a estimativa da função de variância, após o processo de transformação, truncamento das partes imaginária e reais negativas, e inversão da transformada corrigida ao domínio de tempo original. Pode-se notar que a estimativa final

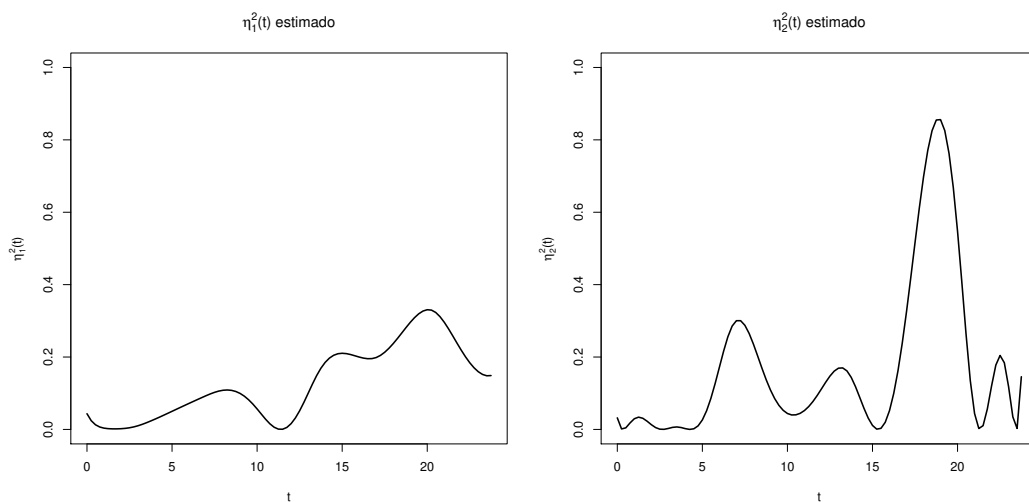


Figura 5.4: A função de variância estimada para os consumidores monofásico e bifásico, por semiparametrização.

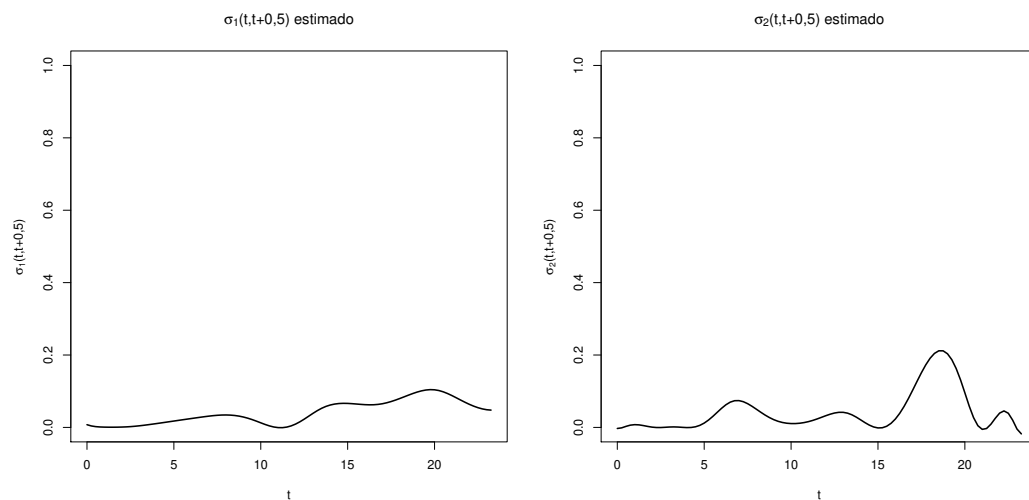


Figura 5.5: As funções $\hat{\sigma}_1(t, t + 0,5)$ e $\hat{\sigma}_2(t, t + 0,5)$, dos consumidores monofásico e bifásico, respectivamente.

concorda em forma com aquela obtida pela semiparametrização – picos no final da tarde, variância do consumidor monofásico em geral menor que a do bifásico, um segundo pico menor para consumidores bifásicos no horário da manhã – mas é muito menos suave, possivelmente sobreestimando o pico da variância do consumidor bifásico entre as 18 e

19 horas, mas em geral subestimando a variância: pelo número reduzido de nós, e pelo fato que as regiões onde a estimativa é negativa foram truncadas em 0.

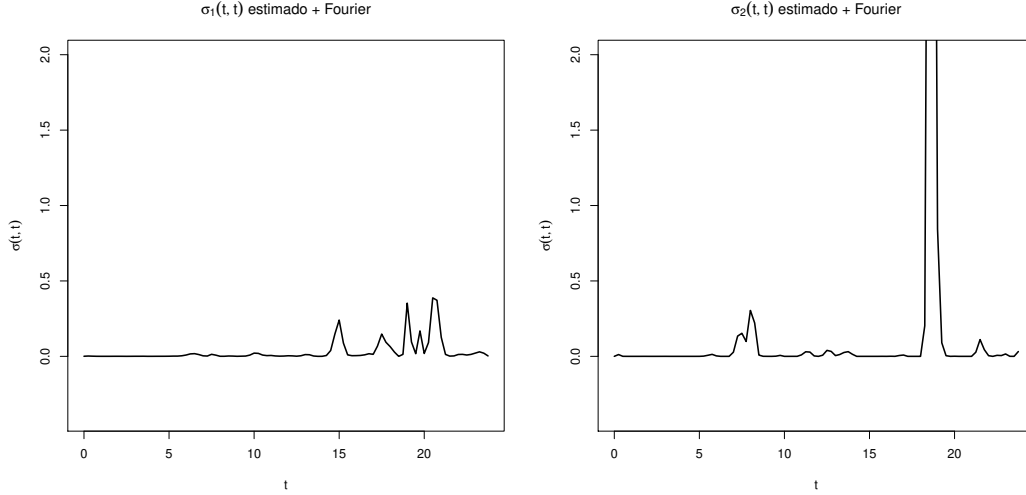


Figura 5.6: A função de variância estimada para os consumidores monofásico e bifásico, por produto tensorial.

A estimativa da covariância entre o processo no instante t e meia hora depois é exibida na Figura 5.7. Há, novamente, uma concordância com a covariância semi-parametrizada, no sentido que a estimativa é proporcional à estimativa da função de variância, mas com algum amortecimento exponencial.

As estimativas obtidas pelo método de semiparametrização, quando observadas nos pontos $s_1, \dots, s_n$ e $t_1, \dots, t_n$, formam neste caso matrizes definidas positivas. Contudo, as estimativas de produto tensorial formam matrizes meramente não negativas definidas. Como isso impossibilita o cálculo da matriz inversa, foi introduzido um pequeno viés nas estimativas, da forma

$$\Sigma_c^* = \Sigma_c + \varepsilon \mathbf{I},$$

onde Σ_c é a função $\sigma_c(s, t)$ observada nos pontos $s_1, \dots, s_n$ e $t_1, \dots, t_n$, ε uma constante e $\mathbf{I}$ a matriz identidade $n \times n$. Foi escolhido o menor ε possível que torne as estimativas positivas definidas. Esse viés aumenta a variância, mas também melhora o número de condição da matriz. Para $\sigma_1(s, t)$, o valor obtido foi $\varepsilon_1 = 0,019$, o que é um valor

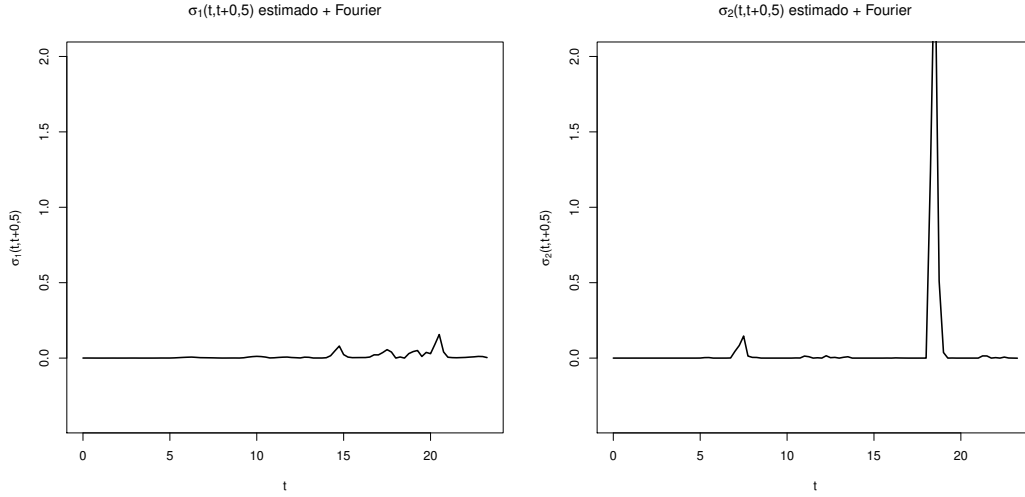


Figura 5.7: As funções $\hat{\sigma}_1(t, t + 0,5)$ e $\hat{\sigma}_2(t, t + 0,5)$, dos consumidores monofásico e bifásico, respectivamente.

razoável de viés. Para $\sigma_2(s, t)$, no entanto, $\varepsilon_2 = 3,398$ foi o menor valor que resultou numa matriz definida positiva, o que está muito longe da estimativa original. É preciso manter isso em mente ao examinar-se a qualidade do intervalo de confiança obtido, para a segunda tipologia.

O intervalo de confiança para cada tipologia é obtido através das equações (1.11) e (1.12), onde em (1.11) substituímos a matriz Σ_c desconhecida pela sua estimativa $\hat{\Sigma}_c$, com cada entrada correspondendo à função $\sigma_c(s, t)$ observada no produto cartesiano dos pontos $t_1, \dots, t_n$. Enfim, a variância de $\hat{\alpha}_c(t)$ é dada por

$$\text{Var}(\hat{\alpha}_c(t)) = \mathbf{B}(t) \mathbf{S} \hat{\Sigma}_c \mathbf{S}' \mathbf{B}'(t),$$

com $\mathbf{S} = (\mathbf{B}'\mathbf{B})^{-1} \mathbf{B}'$ e $\mathbf{B}(t)$ é um vetor com cada componente igual a k -ésima base calculada no ponto t . O intervalo é pontual no sentido que a confiança do intervalo não é conjunta. De fato, tratar o vetor como um modelo de regressão e tomar os quantis de uma Normal n variada resultou em intervalos imensos, e esses experimentos foram descartados do texto da dissertação.

Fixando uma confiança de 90%, os intervalos em torno das tipologias estimadas são exibidos na Figura 5.8. Apesar da instabilidade das estimativas de covariância, os

intervalos são muito bons – exceto pelo que envolve a estimativa de produto tensorial de $\sigma_2(s, t)$, mesmo corrigida pelo viés grande. Aumentar o viés aumenta a amplitude do intervalo, pois afinal a variância se torna maior quando ε_2 aumenta.

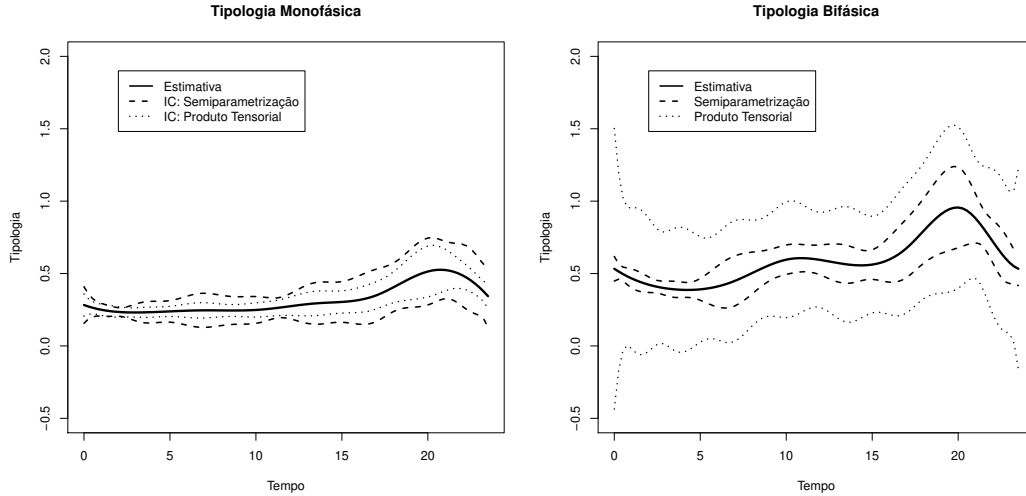


Figura 5.8: Intervalos de Confiança pontual a 90% para as estimativas de Tipologia

É importante comentar que a região de confiança baseada na estimativa de $\sigma_1(s, t)$ pelo produto tensorial parece ter área menor que a resultante das estimativas obtidas por semiparametrização. Mas isso se dá por razão do truncamento dos valores negativos da estimativa de $|\hat{\eta}_1(t)|$ durante a volta da Transformada de Fourier, e possivelmente por uma subestimação do produto tensorial de splines da função variância (já que o número de nós ficou pequeno em relação ao número de pontos amostrais).

Com esta aplicação, onde o número de replicações é muito pequeno, pode-se concluir a favor da semiparametrização para a modelagem da covariância dos consumidores de energia elétrica, uma vez que ela produziu estimativas positivas definidas, comportamento similar ao do produto tensorial em questão da forma das estimativas, e melhor estabilidade numérica.

5.2 Quimiometria

Uma segunda aplicação foi posteriormente estudada em Saraiva (2009) e Dias et al. (2011), em quimiometria. Diferentes substâncias refletem luz de maneira diferente. A *transmitância* de uma substância é a porcentagem I/I_0 de luz que não é refletida, onde I é a intensidade de luz transmitida, depois da substância ter sido inserida num feixe de luz, e I_0 é a intensidade de luz incidente, antes da substância ter sido inserida. A *absorbância* x da substância é definida por

$$x = -\log_{10} \left(\frac{I}{I_0} \right).$$

A absorbância e a transmitância são funções dos comprimentos de onda, ou frequências, do feixe de luz transmitido. A luz ultravioleta possui frequências de 1nm a 400nm, e a luz visível está entre 400nm e 750nm. As frequências de *near infrared* estão na faixa de comprimentos de onda de 750nm a 2500nm, mas o espectro infravermelho todo inclui frequências de 750nm a 10^6 nm. Quando a absorbância é medida em diferentes comprimentos de onda, o resultado é um vetor, que representa observações de uma função do comprimento de onda que caracteriza a substância.

Uma mistura é a combinação de m substâncias puras (simples ou compostas), ou simplesmente *constituintes*, com concentrações y_ℓ , $\ell = 1, 2, \dots, m$, e tal que $y_1 + \dots + y_m = 1$. A Lei de Beer-Lambert postula que para essa mistura de m constituintes, a absorbância num ponto ω_0 do comprimento de onda é dada por

$$x(\omega_0) = \sum_{\ell=1}^m y_\ell \alpha_\ell(\omega_0) + \varepsilon(\omega_0),$$

onde α_ℓ é a absorbância do ℓ -ésimo constituinte, e ε algum erro de medição. Neste caso, os experimentos de espectroscopia constituem uma aplicação da análise de dados funcionais agregados pois é possível determinar a tipologia de um constituinte observando misturas com diferentes concentrações dos mesmos. É portanto um caso de dado funcional agregado por combinação linear (Equação (1.3)), isto é, que segue o modelo

$$X_i(\omega) = \sum_{\ell=1}^m y_{i,\ell} \alpha_\ell(\omega) + \sum_{\ell=1}^m y_{i,\ell}^2 \varepsilon_{i,\ell}(\omega),$$

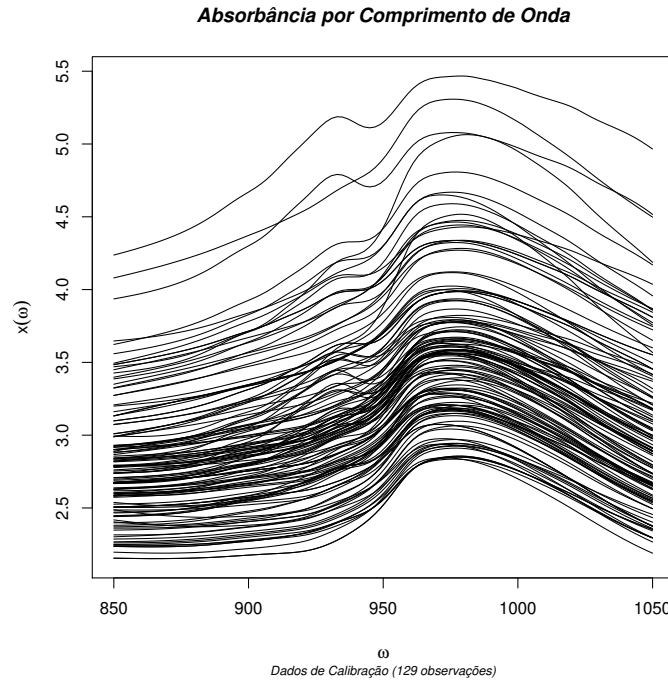


Figura 5.9: Absorbância de carne, para diferentes concentrações de umidade, gordura e proteína

onde $X_i(\omega)$ são as absorbâncias da i -ésima amostra, $y_{i,\ell}$ são as concentrações das substâncias com perfil $\alpha_\ell(\omega)$, $y_{i,1} + \dots + y_{i,m} = 1$ para todo i , e $\varepsilon_{i,\ell}(\omega)$ é um processo Gaussiano.

A Figura 5.9 exhibe os dados obtidos através do aparelho *Tecator Infratec Food and Feed Analyzer*¹, para 100 pontos equispçados entre os comprimentos de onda 850nm e 1050nm, que correspondem à absorbância de carne, com diferentes concentrações de umidade, gordura e proteína (os constituintes).

A Figura 5.10 mostra as frequências das concentrações de cada constituinte. Uma diferença deste cenário com os vistos anteriormente é que não existem replicações, mas o número de curvas agregadas observadas é grande (aqui, igual a 129). A Figura 5.11 mostra gráficos de dispersão das concentrações entre as diferentes curvas agregadas.

¹Os dados são da empresa Tecator e estão disponíveis em domínio público. Eles podem ser obtidos em <http://lib.stat.cmu.edu/datasets/tecator>.

Pode-se notar que as concentrações de umidade e gordura são quase perfeitamente complementares, isto é, maiores concentrações de umidade implicam em menores concentrações de gordura.

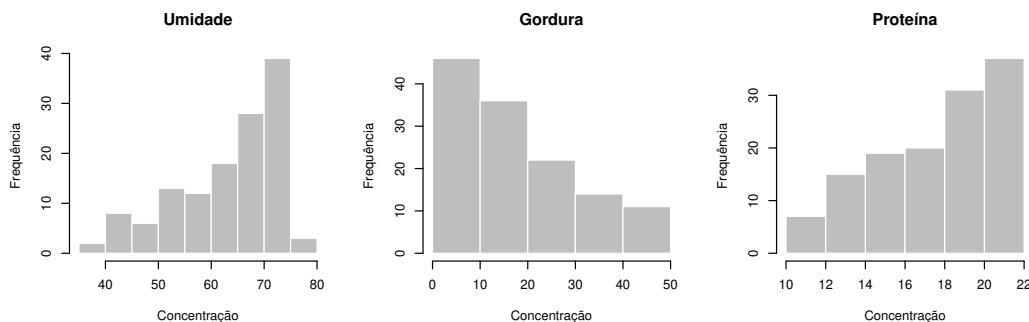


Figura 5.10: Frequências das concentrações de Umidade, Gordura e Proteína

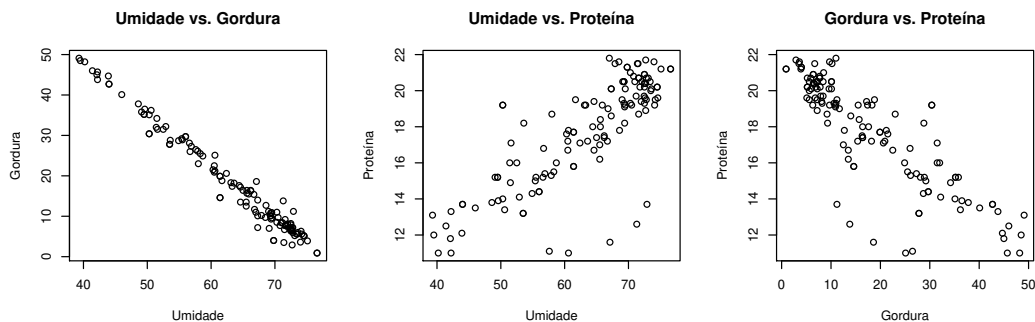


Figura 5.11: Dispersão duas a duas das concentrações de Umidade, Gordura e Proteína

5.2.1 Modelagem de Tipologia

A modelagem de tipologia foi feita com $K = 16$ bases, e as estimativas são mostradas na Figura 5.12.

5.2.2 Modelagem da Função de Covariância Típica

A covariância típica das componentes foi estimada usando o método de autofunções, descrito no Capítulo 4. Não existem, porém, replicações. Como os autovalores serão

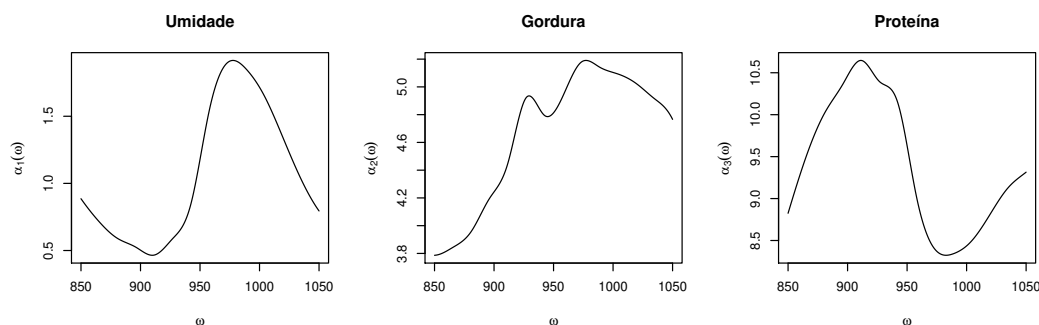


Figura 5.12: As tipologias de absorvância para Umidade, Gordura e Proteína

estimados pela variância amostral do componente η_ν , mas só existe uma observação da mesma, foi realizada a seguinte adaptação do modelo de *proxy*: Como as curvas agregadas estão “quase” ordenadas em termos de concentração dos constituintes – o banco parece ter sido amostrado em dois *clusters* com concentrações decrescentes de umidade, e consequentemente também ordenadas para as outras componentes, devido a sua alta correlação como exibida na Figura 5.11 – para a i -ésima curva foram consideradas replicações as curvas $i - 1$ e $i + 1$, que por esse agrupamento são bastante similares. Uma justificativa visual para este argumento encontra-se na Figura 5.13, que mostra a Umidade, em função da entrada na curva i no banco de dados.

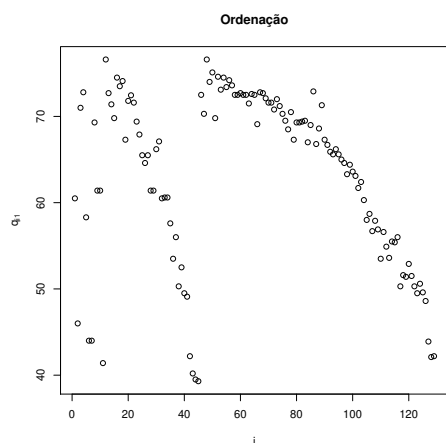


Figura 5.13: Ordenação das concentrações de umidade

As três primeiras autofunções encontram-se na Figura 5.14. A primeira função em particular chama a atenção: ela é quase uma reta horizontal. De fato, Ferraty & Vieu (2006, p. 105) executa uma análise de componentes principais deste mesmo banco de dados, mas bastante simplificada, usando somente a Gordura como variável explicativa. Nos resultados apresentados, a conclusão é que a função constante, vezes a concentração de Gordura, é a componente principal da variação dos dados. Uma inspeção casual da Figura 5.9 sugere mesmo que a principal fonte de variação nos dados é a altura. Em nosso modelo, ela é vista como componente das três variáveis explicativas. Repare ainda que a normalização do autovetor é feita a partir da norma de vetores, enquanto a estimativa da autofunção é tal que a raiz da integral de seu quadrado seja unitária, o que justifica a aparente diferença de escala no gráfico.

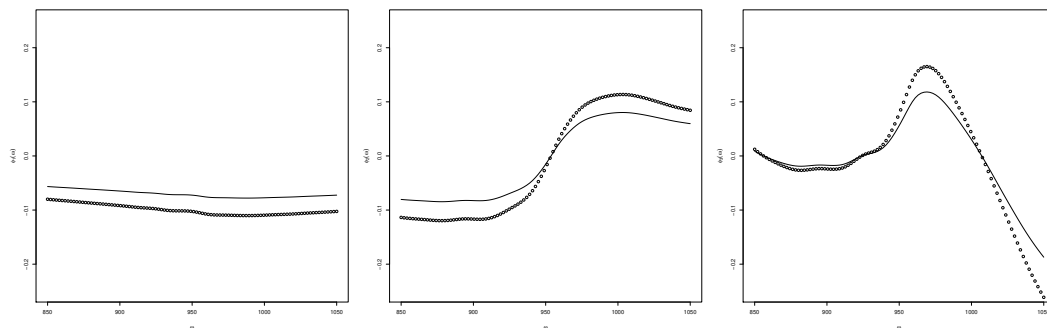


Figura 5.14: Os três primeiros autovetores e as autofunções normalizadas

Os autovalores foram separados usando o modelo $\mathbf{Q}^2$, isto é, cada componente $q_{i,c}$ foi elevado ao quadrado para construir-se as estimativas, de tal forma que os autovalores $\gamma_{\nu,c}$ foram estimados a partir de $\gamma_{\nu}^{(i)}$ (obtidos a partir do proxy de vizinhança definido anteriormente), o que permitiu separar as estimativas de covariância de cada componente, dadas em perspectiva pela Figura 5.15.

Em conclusão, o método de autovalores e autofunções funcionou muito bem neste caso, uma vez que existia uma correlação muito forte entre as covariáveis, e logo as componentes principais eram muito parecidas, já que os coeficientes variam quase conjuntamente. A restrição de ortogonalidade não é portanto severa neste caso, mesmo

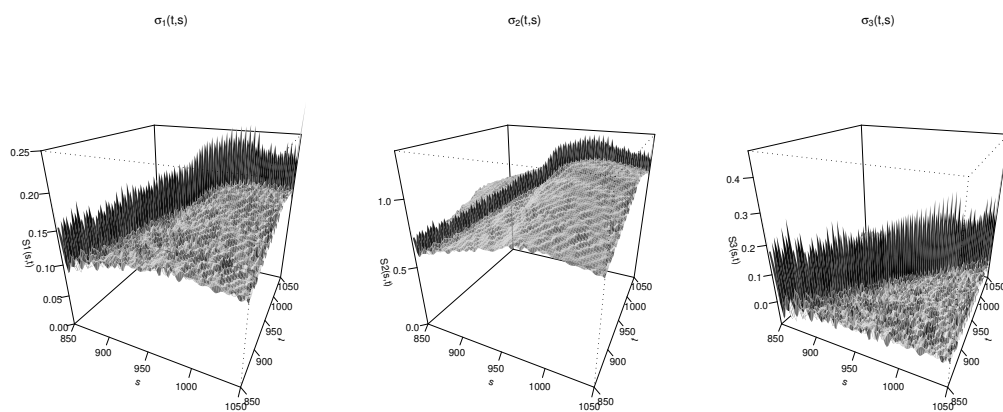


Figura 5.15: As estimativas de Covariância para cada um dos constituintes que as estimativas de tipologia tenham sido marcadamente distintas.

Capítulo 6

Considerações Finais

Neste trabalho, foram apresentados três métodos de estimação da função de covariância para dados funcionais agregados, e cada um apresentou suas vantagens e desvantagens em termos de eficiência, custo computacional, número de hipóteses exigidas etc. Os três métodos, contudo, tem uma característica em comum: são capazes de produzir estimativas que são funções não negativas definidas e separá-las quando as curvas que observamos são agregadas.

O método de semiparametrização, apresentado no Capítulo 2, baseia-se no modelo de função de covariância individual dado por

$$\sigma_c(s, t) = \eta_c(s)\eta_c(t)e^{-\phi_c|t-s|},$$

em que $\eta_c(t)$ pode ser estimada de maneira não paramétrica. A fórmula é simples mas resulta num problema de estimação quase irrestrito (basta apenas que ϕ_c seja positivo), já que para qualquer configuração de $\eta_c(t)$ a estimativa será pelo menos não negativa definida. Este é, de fato, o único método onde pode-se garantir estimativas *positivas definidas*. Para tanto, basta assegurar que $\hat{\eta}(t) \neq 0$. Estimativas positivas definidas abrem espaço a uma gama maior de aplicações que dependem da existência de uma inversa para a covariância amostral.

Por outro lado, esse método é o mais limitado em termos de hipóteses: impõe-se uma forma para a correlação do processo, $e^{-\phi_c|t-s|}$, o que implica em uma dependência ho-

mogênea no tempo, com um decaimento do tipo exponencial. Além disso, um processo de estimação baseado na verossimilhança é bastante caro computacionalmente.

O método apresentado no Capítulo 3, baseado no produto tensorial de splines, é mais versátil: toda a superfície de covariância é examinada, não se fazendo hipóteses sobre o tipo de decaimento da correlação. A função de covariância individual é dada por

$$\sigma_c(s, t) = \sum_{k_1=1}^K \sum_{k_2=1}^K \theta_{c,k_1,k_2} B_{k_1}(s) B_{k_2}(t).$$

O preço é bem caro: embora consistente, o método não é capaz de produzir estimativas não negativas definidas quando o número de replicações é pequeno. Esse método também tende a subestimar a variância quando o decaimento da correlação é muito rápido, forçando uma suavização onde ainda há muita informação. A análise da transformada de Fourier do processo serve para corrigir esse defeito, ainda que introduza uma hipótese à forma da correlação: ela é igual a função característica de uma variável aleatória contínua e simétrica no zero. Essa hipótese surge unicamente do fato que só conhecemos a transformada (analítica) de Fourier de processos dessa forma. Uma generalização do Teorema de Bochner para processos não estacionários – algo nada trivial – permitiria melhorar esse algoritmo de modo a produzir estimativas de funções de covariância mais gerais.

Por fim, a terceira metodologia, apresentada no Capítulo 4, baseia-se na representação espectral da função de covariância, dada por

$$\sigma_c(s, t) = \sum_{\nu=1}^{\infty} \gamma_{c,\nu} \phi_{\nu}(s) \phi_{\nu}(t).$$

Essa representação permite separar as covariâncias individuais quando assumimos que $\phi_{\nu}(t)$ não depende de c para nenhum ν – ou simplesmente que as funções de covariância individuais tem as mesmas autofunções. Essa hipótese é equivalente, na interpretação estatística das autofunções de um processo, a dizer que as funções de covariância tem as mesmas componentes principais de variabilidade.

As autofunções são estimadas a partir dos dados, o que dá uma grande versatilidade ao método, sendo desnecessário fazer hipóteses sobre estacionariedade, por exemplo.

Verificar que as autofunções são as mesmas, contudo, é difícil. Outra desvantagem é que nem sempre é possível obter estimativas boas de todas as autofunções: no máximo as n primeiras autofunções podem ser encontradas, e ainda assim as últimas autofunções pouco servem para explicar a variabilidade dos dados, sendo mais interessante descartá-las em favor de estimativas mais suaves.

Decidir qual metodologia é mais adequada requer uma análise descritiva cuidadosa das curvas agregadas. Muito do comportamento da covariância individual reflete-se na matriz de covariância agregada, sendo esta o ponto inicial para as análises. A metodologia de semiparametrização é uma boa escolha padrão, em termos de eficiência, mas as estimativas pelo produto tensorial com transformada corrigida podem lidar com casos em que não há decaimento exponencial da correlação: por exemplo, nos casos em que a autocorrelação é negativa. E se as componentes principais das curvas agregadas forem as mesmas, provavelmente as das covariâncias individuais também serão, o que motiva o uso do método baseado na decomposição espectral.

Referências Bibliográficas

- Bochner, S. (1959), *Lectures on Fourier Integrals*, Princeton University Press, Princeton.
- Bochner, S. & Chandrasekharan, K. S. (1949), *Fourier Transforms - Annals of Mathematics Studies*, Princeton University Press, Princeton.
- Bracewell, R. N. (1986), *The Fourier Transform and its Applications, 2nd edition*, McGraw Hill, Boston.
- de Boor, C. (2001), *A Practical Guide to Splines, Revised Edition*, Springer-Verlag, New York.
- Dias, R. (2000), ‘A note on density estimation using a proxy of the kullback-leibler distance’, *Brazilian Journal of Probability and Statistics* **13**(2), 181–192.
- Dias, R. & Garcia, N. L. (2007), ‘Consistent estimator for basis selection based on a proxy of the kullback-leibler distance’, *Journal of Econometrics* **141**(1).
- Dias, R., Garcia, N. L. & Martarelli, A. (2009), ‘Non-parametric estimation for aggregated functional data for electric load monitoring’, *Environmetrics* **20**, 111–130.
- Dias, R., Garcia, N. L. & Schmidt, A. (2011), ‘A hierarchical model for aggregated functional data’, *arXiv:1102.2773*.
- Ferraty, F. & Vieu, P. (2006), *Nonparametric Functional Data Analysis*, Springer-Verlag, New York.

- Hall, P., Fisher, N. I. & Hoffman, B. (1994), ‘On the nonparametric estimation of covariance functions’, *Annals of Statistics* **22**(4), 2115–2134.
- Lee, S.-Y., Zhang, W. & Song, X.-Y. (2002), ‘Estimating the covariance function with functional data’, *The British Journal of Mathematical and Statistical Psychology* **55**, 247–261.
- Loève, M. (1977), *Probability Theory, Volume I*, Springer-Verlag, New York.
- Loève, M. (1978), *Probability Theory, Volume II*, Springer-Verlag, New York.
- Martarelli, A. (2006), Estimação de tipologia para dados funcionais agrupados, Master’s thesis, Universidade Estadual de Campinas, Campinas.
- R Development Core Team (2009), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
- URL:** <http://www.R-project.org>
- Ramsay, J. O., Hooker, G. & Graves, S. (2009), *Functional Data Analysis with R and MATLAB*, Springer-Verlag, New York.
- Ramsay, J. O. & Silverman, B. W. (2002), *Applied Functional Data Analysis*, Springer-Verlag, New York.
- Ramsay, J. O. & Silverman, B. W. (2005), *Functional Data Analysis, Second Edition*, Springer-Verlag, New York.
- Saraiva, M. A. (2009), Análise não-paramétrica de dados funcionais: Uma aplicação à quimiometria, Master’s thesis, Universidade Estadual de Campinas, Campinas.
- Stewart, J. (1976), ‘Positive definite functions and generalizations, an historical survey’, *Rocky Mountain Journal of Mathematics* **6**(3), 409–434.
- Wahba, G. (1990), *Spline Models for Observational Data*, SIAM, Philadelphia.

- Yaglom, A. M. (1987), *Correlation Theory of Stationary and Related Random Functions*
- *Volume I*, Springer-Verlag, New York.

Apêndice A

A Transformada de Fourier de Splines

O primeiro Apêndice discute questões técnicas de implementação de B-splines – como gerar as bases, como tratar das bases nas extremidades, e como obter a Transformada de Fourier de um spline. Um exemplo é construído para demonstrar a qualidade da aproximação da Transformada de Splines à Transformada da função que gerou as observações.

A.1 Construindo B-splines

B-splines são definidos da seguinte forma: tome a sequência de nós onde eles estão definidos, isto é, $a = \tau_1 \leq \tau_2 \leq \dots \leq \tau_L = b$. Ao todo serão $K = L + m - 2$ funções base. Para a ℓ -ésima base, que envolve os nós $\{\tau_\ell, \tau_{\ell+1}, \dots, \tau_{\ell+m}\}$, o B-spline é dado por uma forma recursiva conhecida como algoritmo de Cox-de Boor:

$$\begin{aligned} B_{\ell,1}(t) &:= \begin{cases} 1, & \text{se } \tau_\ell \leq t < \tau_{\ell+1} \\ 0 & \text{caso contrário} \end{cases} \\ B_{\ell,m}(t) &:= \frac{t - \tau_\ell}{\tau_{\ell+m-1} - \tau_\ell} B_{\ell,m-1}(t) + \frac{\tau_{\ell+m} - t}{\tau_{\ell+m} - \tau_{\ell+1}} B_{\ell+1,m-1}(t). \end{aligned} \tag{A.1}$$

Em qualquer circunstância onde $\tau_\ell = \tau_{\ell+1}$, o algoritmo fixa o quociente indetermi-

nado como 0. Isso é necessário em particular para as bases que se localizam nas extremidades do intervalo; se $m > 1$, para construí-las, repetem-se $m - 1$ vezes os nós τ_1 e τ_L – o algoritmo exige esses nós “auxiliares” $\tau_{-m+2} = \dots = \tau_0 = \tau_1$ e $\tau_L = \tau_{L+1} = \dots = \tau_{L+m-1}$. Note ainda que a numeração das bases da forma aqui definida segue de Boor (2001), para conveniência do algoritmo, isto é, $\ell = -m + 2, \dots, 0, 1, 2, \dots, L - 1$. A dissertação usa no entanto a numeração de Ramsay & Silverman (2005), que indexa as bases por k ao invés de ℓ , e é tal que a primeira base ($k = 1$) é dada pela base $\ell = -m + 2$ da numeração de de Boor, $k = 2$ é dada por $-m + 3$, e assim por diante, até $\ell = L - 1$ que é dada por $k = K$.

Repare que (A.2) pode ser reescrito na forma

$$B_{\ell,m}(t) = h(B_{\ell,m-1}(t) * B_{\ell,1}(t))$$

onde $h(\cdot)$ é alguma transformação de translação e escala. Aqui $*$ é o operador de convolução, isto é,

$$B_{\ell,m-1}(t) * B_{\ell,1}(t) = \int_{-\infty}^{\infty} B_{\ell,m-1}(t-u) B_{\ell,1}(u) du$$

A.2 Propriedades da Transformada de Fourier

A Transformada de Fourier de uma função é dada por

$$\mathcal{F}(w) := \int_{-\infty}^{\infty} e^{-iwt} f(t) dt \quad (\text{A.2})$$

Existem formulações alternativas à forma (A.2) na literatura, por exemplo usando $2\pi w$ como argumento da função exponencial. Essa integral também pode não existir para qualquer $f(t)$. Uma condição suficiente é que a integral de $|f(t)|$ com respeito a t convirja. Muitos resultados aqui considerados encontram-se em Bracewell (1986), com suas respectivas demonstrações.

A transformada inversa retorna $\mathcal{F}(w)$ à função original $f(t)$, e é dada por

$$\mathcal{F}^{-1}\{\mathcal{F}\}(t) := \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{iwt} \mathcal{F}(w) dw.$$

As seguintes propriedades da transformada de Fourier podem ser demonstradas (exceto (P.4)) simplesmente utilizando uma mudança de variáveis na equação (A.2):

- (P.1) (Similaridade): Se $f(t)$ tem transformada $\mathcal{F}(w)$, então $f(at)$ tem transformada $|a|^{-1}\mathcal{F}(w/a)$, $a \in \mathbb{R}$.
- (P.2) (Aditividade): Se $f(t)$ e $g(t)$ tem transformadas $\mathcal{F}(w)$ e $\mathcal{G}(w)$, então a transformada de $af(t) + bg(t)$ é $a\mathcal{F}(w) + b\mathcal{G}(w)$, $a, b \in \mathbb{R}$.
- (P.3) (Translação): Se $f(t)$ tem transformada $\mathcal{F}(w)$, então $f(t - \mu)$ tem transformada $e^{-iw\mu}\mathcal{F}(w)$, $\mu \in \mathbb{R}$.
- (P.4) (Convolução): $\mathcal{F}(g * h)(w) = \mathcal{F}(g)(w) \cdot \mathcal{F}(h)(w)$; note que vale o inverso, isto é, $\mathcal{F}(g \cdot h)(w) = \mathcal{F}(g)(w) * \mathcal{F}(h)(w)$.

A demonstração de (P.4), bem como a da transformada inversa, encontra-se em Bochner & Chandrasekharan (1949, p. 5 e p. 10, respectivamente).

O B-spline de ordem 1 tem Transformada de Fourier dada por

$$\begin{aligned} \mathcal{F}(B_{\ell,1})(w) &= \int_{-\infty}^{\infty} e^{-itw} \mathbf{1}_{\{\tau_{\ell} \leq t < \tau_{\ell+1}\}} dt = \int_{\tau_{\ell}}^{\tau_{\ell+1}} e^{-itw} dt, \\ \int_{\tau_{\ell}}^{\tau_{\ell+1}} e^{-itw} dt &= -\frac{1}{iw} e^{-itw} \Big|_{\tau_{\ell}}^{\tau_{\ell+1}} = \frac{e^{-i\tau_{\ell}w} - e^{-i\tau_{\ell+1}w}}{iw} \end{aligned} \quad (\text{A.3})$$

A.3 Construindo B-splines através de Convoluções

O cálculo da transformada de Fourier de um Spline simplifica-se consideravelmente se for considerado um algoritmo alternativo para criar as bases $B_{\ell}(t)$, que permita empregar a propriedade (P.4). Isto é, é possível construir um B-spline em nós equispaçados através da convolução de outros B-spline de ordem menor. Então cada base $B_{\ell}(t)$ de ordem m será definida a partir de translações e dilatações de uma *blending function* $N_m(t)$.

A $N_m(t)$ blending function é a base B-spline de ordem m , definida por

$$\begin{aligned} N_1(t) &= \mathbf{1}_{[0,1]}(t) \\ N_m(t) &= N_{m-1}(t) * N_1(t) \end{aligned}$$

Por exemplo, temos que para $N_2(t)$, que

$$N_2(t) = N_1(t) * N_1(t) = \begin{cases} \int_0^t ds & \text{se } 0 \leq t < 1 \\ \int_{t-1}^1 ds & \text{se } 1 \leq t < 2 \end{cases} = \begin{cases} t & \text{se } 0 \leq t < 1 \\ 2 - t & \text{se } 1 \leq t < 2 \end{cases}$$

Repare que a função $N_1(s)$ é não nula para $s \in [0, 1]$ e logo o intervalo de integração $[t-1, t]$ é não-nulo para $t \in [0, 2]$. De maneira similar, temos que para $m = 3$, $N_2(s)$ é não nula para $s \in [0, 2]$ e aí

$$\begin{aligned} N_3(t) = N_2(t) * N_1(t) &= \begin{cases} \int_0^t N_2(s) ds & \text{se } 0 \leq t < 1 \\ \int_{t-1}^1 N_2(s) ds + \int_1^t N_2(s) ds & \text{se } 1 \leq t < 2 \\ \int_{t-1}^2 N_2(s) ds & \text{se } 2 \leq t < 3 \end{cases} \\ &= \begin{cases} \frac{t^2}{2} & \text{se } 0 \leq t < 1 \\ \frac{1}{2}(-2t^2 + 6t - 3) & \text{se } 1 \leq t < 2 \\ \frac{1}{2}(t-3)^2 & \text{se } 2 \leq t < 3 \end{cases} \end{aligned}$$

E finalmente, temos que os B-splines cúbicos (ordem $m = 4$) uniformes são dados

por $N_3(t) * N_1(t)$, ou simplesmente

$$N_4(t) = N_3(t) * N_1(t) = \begin{cases} \frac{t^3}{6} & \text{se } 0 \leq t < 1 \\ \frac{2}{3} - 2t + 2t^2 - \frac{t^3}{2} & \text{se } 1 \leq t < 2 \\ -\frac{22}{3} + 10t - 4t^2 + \frac{t^3}{2} & \text{se } 2 \leq t < 3 \\ -\frac{1}{6}(t-4)^3 & \text{se } 3 \leq t < 4 \end{cases}$$

Os splines do tipo $N_m(t)$ estão definidos nos nós $0, 1, \dots, m$. Os B-splines $B_{\ell,m}(t)$ podem ser escritos como transformações de translação e escala de $N_m(t)$. Assuma os nós $|\tau_\ell - \tau_{\ell+1}| = r$ (equidistantes); essa hipótese pode ser relaxada, mas a simplificação de contas é considerável no caso equidistante.

O resultado será mostrado apenas para os B-splines de ordem 4. Note que como $N_4(t)$ é simétrico em torno de 2, então $N_4(t + 2 - \tau_{\ell+2})$ estará centrado em $\tau_{\ell+2}$. Ele é não nulo, agora, em $t \in [\tau_{\ell+2} - 2, \tau_{\ell+2} + 2]$. Então a mudança de escala necessária é simplesmente $2/(2r) = r^{-1}$. Chega-se, portanto, à fórmula

$$B_{\ell,4}(t) = N_4(r^{-1}(t - \tau_{\ell+2}) + 2) = N_4(r^{-1}(t - \tau_\ell)).$$

Agora $N_4(r^{-1}t + 2 - r^{-1}\tau_{\ell+2})$ é não nulo em $t \in [\tau_{\ell+2} - 2r, \tau_{\ell+2} + 2r] = [\tau_\ell + 2r - 2r, \tau_\ell + 2r + 2r] = [\tau_\ell, \tau_{\ell+4}]$, exatamente como na fórmula de Cox-de Boor. Dedução similar pode ser feita para os Splines de ordem inferior, obtendo-se:

$$B_{\ell,1}(t) = N_1\left(r^{-1}\left(t - \frac{\tau_{\ell+1} + \tau_\ell}{2}\right) + \frac{1}{2}\right)$$

$$B_{\ell,2}(t) = N_2(r^{-1}(t - \tau_{\ell+1}) + 1)$$

$$B_{\ell,3}(t) = N_3\left(r^{-1}\left(t - \frac{\tau_{\ell+2} + \tau_{\ell+1}}{2}\right) + \frac{3}{2}\right)$$

E, aproveitando o fato dos nós serem equispaçados, pode-se generalizar a fórmula para Splines de qualquer ordem por

$$B_{\ell,m}(t) = N_m \left(r^{-1}(t - \tau_\ell) \right).$$

Essa fórmula, no entanto, vale para $\ell = 1, \dots, L - 4$, sendo necessário definir outra convolução para as bases que envolvem os nós repetidos em $\ell = -2, -1, 0, L - 3, L - 2, L - 1$. No caso de nós equispaçados, é possível verificar que as bases $\ell = -2, -1, 0$ são reflexões de $\ell = L - 3, L - 2, L - 1$, respectivamente, em torno de $T/2$, para Splines definidos em $[0, T]$. Serão calculadas explicitamente, portanto, as bases $B_{-2,4}(t)$, $B_{-1,4}(t)$ e $B_{0,4}(t)$ (correspondentes a $B_1(t)$, $B_2(t)$ e $B_3(t)$ na notação de Ramsay & Silverman (2005)), usando o algoritmo de Cox-de Boor. A Figura A.1 mostra as bases das extremidades, em tons de cinza, ligeiramente deformadas com relação às bases constituídas de nós únicos e equispaçados (que estão em cor mais escura).

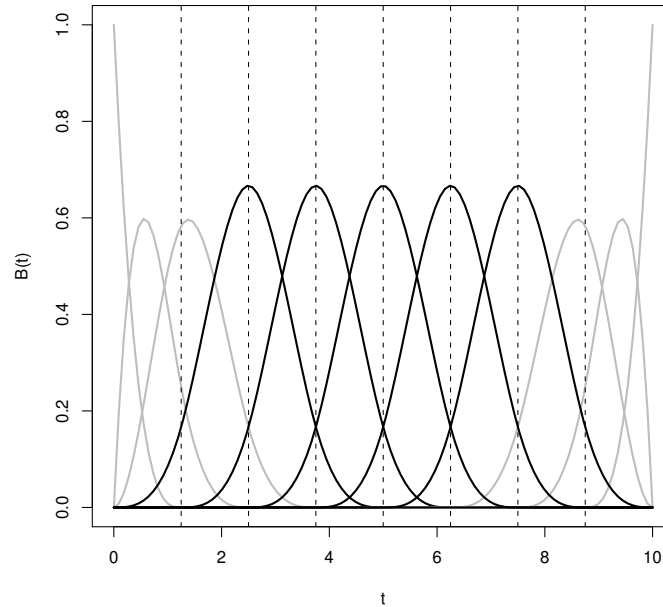


Figura A.1: 11 bases B-spline em $[0, 10]$, bases com nós repetidos realçadas.

Primeiramente, a base $B_{-2,4}(t)$, com nós $\tau_1, \tau_1, \tau_1, \tau_1, \tau_2$, e $\tau_1 = 0$ por conveniência,

é construída por

$$\begin{aligned}
B_{-2,1}(t) &= \mathbf{1}_0(t) \\
B_{-2,2}(t) &= (\tau_{-1} - \tau_{-2})^{-1}(t - \tau_{-2})B_{-2,1}(t) + (\tau_0 - \tau_{-1})^{-1}(\tau_0 - t)B_{-1,1}(t) \\
B_{-2,3}(t) &= (\tau_0 - \tau_{-2})^{-1}(t - \tau_{-2})B_{-2,2}(t) + (\tau_1 - \tau_{-1})^{-1}(\tau_1 - t)B_{-1,2}(t) \\
B_{-2,4}(t) &= (\tau_1 - \tau_{-2})^{-1}(t - \tau_{-2})B_{-2,3}(t) + (\tau_2 - \tau_{-1})^{-1}(\tau_2 - t)B_{-1,3}(t)
\end{aligned}$$

ou simplesmente

$$\begin{aligned}
B_{-2,1}(t) &= 0 \\
B_{-2,2}(t) &= 0 \\
B_{-2,3}(t) &= 0 \\
B_{-2,4}(t) &= \tau_2^{-1}(\tau_2 - t)B_{-1,3}(t)
\end{aligned}$$

Agora, a base $B_{-1,4}(t)$, com nós $\tau_1, \tau_1, \tau_1, \tau_2, \tau_3$, é construída de maneira similar, obtendo-se

$$\begin{aligned}
B_{-1,1}(t) &= 0 \\
B_{-1,2}(t) &= 0 \\
B_{-1,3}(t) &= \tau_2^{-1}(\tau_2 - t)B_{0,2}(t) \\
B_{-1,4}(t) &= \tau_2^{-1}tB_{-1,3}(t) + \tau_3^{-1}(\tau_3 - t)B_{0,3}(t)
\end{aligned}$$

E a base $B_{0,4}(t)$, com nós $\tau_1, \tau_1, \tau_2, \tau_3, \tau_4$, é dada por

$$\begin{aligned}
B_{0,1}(t) &= 0 \\
B_{0,2}(t) &= \tau_2^{-1}(\tau_2 - t)B_{1,1}(t) \\
B_{0,3}(t) &= \tau_2^{-1}tB_{0,2}(t) + \tau_3^{-1}(\tau_3 - t)B_{1,2}(t) \\
B_{0,4}(t) &= \tau_3^{-1}tB_{0,3}(t) + \tau_4^{-1}(\tau_4 - t)B_{1,3}(t)
\end{aligned}$$

A base $B_{1,4}(t)$ está nos nós $\tau_1, \tau_2, \tau_3, \tau_4, \tau_5$, não apresentando nós repetidos, portanto. Mas note que valem as seguintes relações (desde que $\tau_1 = 0$):

$$B_{1,m}(t) = N_m(r^{-1}t), \quad m = 1, 2, 3. \quad (\text{A.4})$$

Então

$$B_{-2,4}(t) = \left(1 - \frac{t}{r}\right)^3 N_1\left(\frac{t}{r}\right)$$

$$B_{-1,4}(t) = \left(2 - \frac{3t}{2r}\right) \frac{t}{r} \left(1 - \frac{t}{r}\right) N_1\left(\frac{t}{r}\right) + \left(1 - \frac{t}{2r}\right)^2 N_2\left(\frac{t}{r}\right)$$

$$B_{0,4}(t) = \frac{t}{2r} \frac{t}{r} \left(1 - \frac{t}{r}\right) N_1\left(\frac{t}{r}\right) + \frac{t}{2r} \left(1 - \frac{t}{2r}\right) N_2\left(\frac{t}{r}\right) + \left(1 - \frac{t}{3r}\right) N_3\left(\frac{t}{r}\right)$$

A Figura A.2 mostra essas bases, sendo possível verificar que a sua construção é adequada.

A.4 Transformada do B-spline de ordem m

Recorde que, pela equação (A.3) e (P.3), temos que

$$\mathcal{F}(N_1)(w) = \frac{1 - e^{-iw}}{iw} = e^{-\frac{iw}{2}} \frac{\text{sen}(w/2)}{w/2}$$

E que por (P.4), então

$$\mathcal{F}(N_m)(w) = \left(e^{-\frac{iw}{2}} \frac{\text{sen}(w/2)}{w/2}\right)^m.$$

Mas por (A.4), aplicando (P.1) e (P.3), temos que a transformada de Fourier do B-spline $B_{\ell,m}(t)$ com nós distintos e equispaçados é dada por

$$\begin{aligned} \mathcal{F}(B_{\ell,m})(w) &= r e^{-i(r^{-1}\tau_\ell)rw} \left(e^{-\frac{irw}{2}} \frac{\text{sen}(rw/2)}{rw/2}\right)^m \\ &= r e^{-i(\tau_\ell + \frac{rm}{2})w} \left(\frac{\text{sen}(rw/2)}{rw/2}\right)^m. \end{aligned}$$

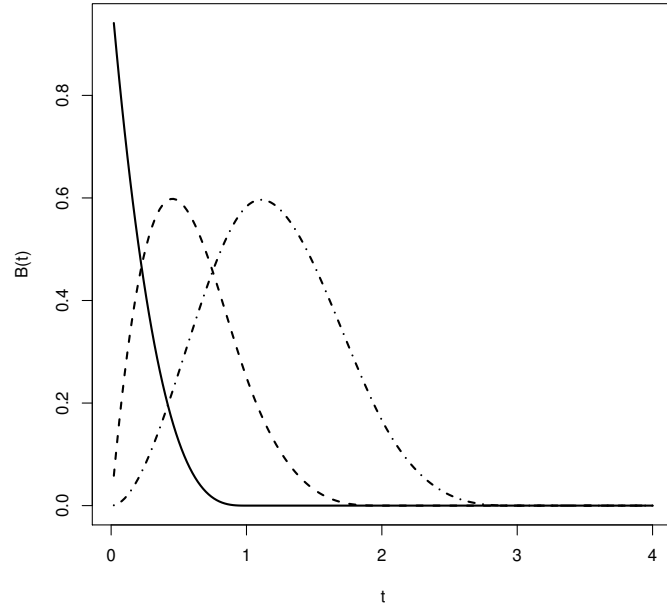


Figura A.2: As bases $B_{-2,4}(t)$, $B_{-1,4}(t)$ e $B_{0,4}(t)$, com $r = 1$

E em particular para o B-spline cúbico

$$\mathcal{F}(B_{\ell,4})(w) = r e^{-i\tau_{\ell+2}w} \left(\frac{\text{sen}(rw/2)}{rw/2} \right)^4.$$

As bases das extremidades serão calculadas explicitamente, para o caso $m = 4$.

Aplicando a definição, obtém-se

$$\begin{aligned} \mathcal{F}(B_{-2,4})(w) &= \int_{-\infty}^{\infty} e^{-iwt} \left(1 - \frac{t}{r}\right)^3 N_1\left(\frac{t}{r}\right) dt, \\ \mathcal{F}(B_{-1,4})(w) &= \int_{-\infty}^{\infty} e^{-iwt} \left[\left(2 - \frac{3t}{2r}\right) \frac{t}{r} \left(1 - \frac{t}{r}\right) N_1\left(\frac{t}{r}\right) + \left(1 - \frac{t}{2r}\right)^2 N_2\left(\frac{t}{r}\right) \right] dt, \\ \mathcal{F}(B_{0,4})(w) &= \int_{-\infty}^{\infty} e^{-iwt} \left[\frac{t}{2r} \frac{t}{r} \left(1 - \frac{t}{r}\right) N_1\left(\frac{t}{r}\right) \right. \\ &\quad \left. + \frac{t}{2r} \left(1 - \frac{t}{2r}\right) N_2\left(\frac{t}{r}\right) + \left(1 - \frac{t}{3r}\right) N_3\left(\frac{t}{r}\right) \right] dt. \end{aligned} \tag{A.5}$$

Seja $\mathcal{N}_m(w)$ a m -ésima transformada do spline natural $N_m(t)$. Usando a propriedade de que a transformada da derivada de uma função $f(t)$, exibida em Bracewell (1986, p. 117), é dada por

$$\int_{-\infty}^{\infty} e^{-iwt} f'(t) dt = iw\mathcal{F}(w),$$

e conseqüentemente a transformada de $tf(t)$ ser dada por $i\mathcal{F}'(w)$, podemos reescrever as fórmulas de (A.5) por

$$\begin{aligned}\mathcal{F}(B_{-2,4})(w) &= r(\mathcal{N}_1(rw) - 3i\mathcal{N}'_1(rw) - 3\mathcal{N}''_1(rw) + i\mathcal{N}'''_1(rw)) \\ \mathcal{F}(B_{-1,4})(w) &= r(2i\mathcal{N}'_1(rw) + 7/2\mathcal{N}''_1(rw) - 3/2i\mathcal{N}'''_1(rw) \\ &\quad + \mathcal{N}_2(rw) - i\mathcal{N}'_2(rw) - 1/4\mathcal{N}''_2(rw)) \\ \mathcal{F}(B_{0,4})(w) &= r(1/2(-\mathcal{N}''_1(rw) + i\mathcal{N}'''_1(rw)) + 1/2(i\mathcal{N}'_2(rw) + 1/2\mathcal{N}''_2(rw)) \\ &\quad + \mathcal{N}_3(rw) - 1/3i\mathcal{N}'_3(rw))\end{aligned}$$

Note que $\mathcal{N}'_1(rw)$ é uma notação compacta para $d/d\omega\mathcal{N}_1(\omega)|_{\omega=rw}$, não sendo necessário aplicar Leibniz e multiplicar o resultado por r após a diferenciação. A notação vale igualmente para as derivadas de ordem superior. Para conveniência de implementação, emprega-se a regra de Leibniz para as derivadas das bases de ordem superior a 2, a saber

$$\mathcal{N}'_m(rw) = m\mathcal{N}_{m-1}(rw)\mathcal{N}'_1(rw),$$

o que gera equações que dependem somente de $\mathcal{N}_1(rw)$ e suas derivadas:

$$\begin{aligned}\mathcal{F}(B_{-2,4})(w) &= r(\mathcal{N}_1(rw) - 3i\mathcal{N}'_1(rw) - 3\mathcal{N}''_1(rw) + i\mathcal{N}'''_1(rw)) \\ \mathcal{F}(B_{-1,4})(w) &= r(2i\mathcal{N}'_1(rw) + 7/2\mathcal{N}''_1(rw) - 3/2i\mathcal{N}'''_1(rw) + \mathcal{N}_1^2(rw) \\ &\quad - 2i\mathcal{N}'_1(rw)\mathcal{N}_1(rw) - 1/2(\mathcal{N}''_1(rw)\mathcal{N}_1(rw) + [\mathcal{N}'_1(rw)]^2)) \\ \mathcal{F}(B_{0,4})(w) &= r(1/2(-\mathcal{N}''_1(rw) + i\mathcal{N}'''_1(rw)) + i\mathcal{N}'_1(rw)\mathcal{N}_1(rw) + 1/2[\mathcal{N}'_1(rw)]^2 \\ &\quad + 1/2\mathcal{N}''_1(rw)\mathcal{N}_1(rw) + \mathcal{N}_1^3(rw) - i\mathcal{N}_1^2(rw)\mathcal{N}'_1(rw))\end{aligned}$$

As bases transformadas das extremidades, bem como a transformada da base $N_4(t)$, encontram-se na Figura A.3.

As transformadas da extremidade 0 fechada à esquerda podem ser facilmente generalizadas para as bases nas extremidades dos intervalos $[\tau_1, \tau_L]$. Basta tomar $B_{-\ell,4}(t - \tau_1)$

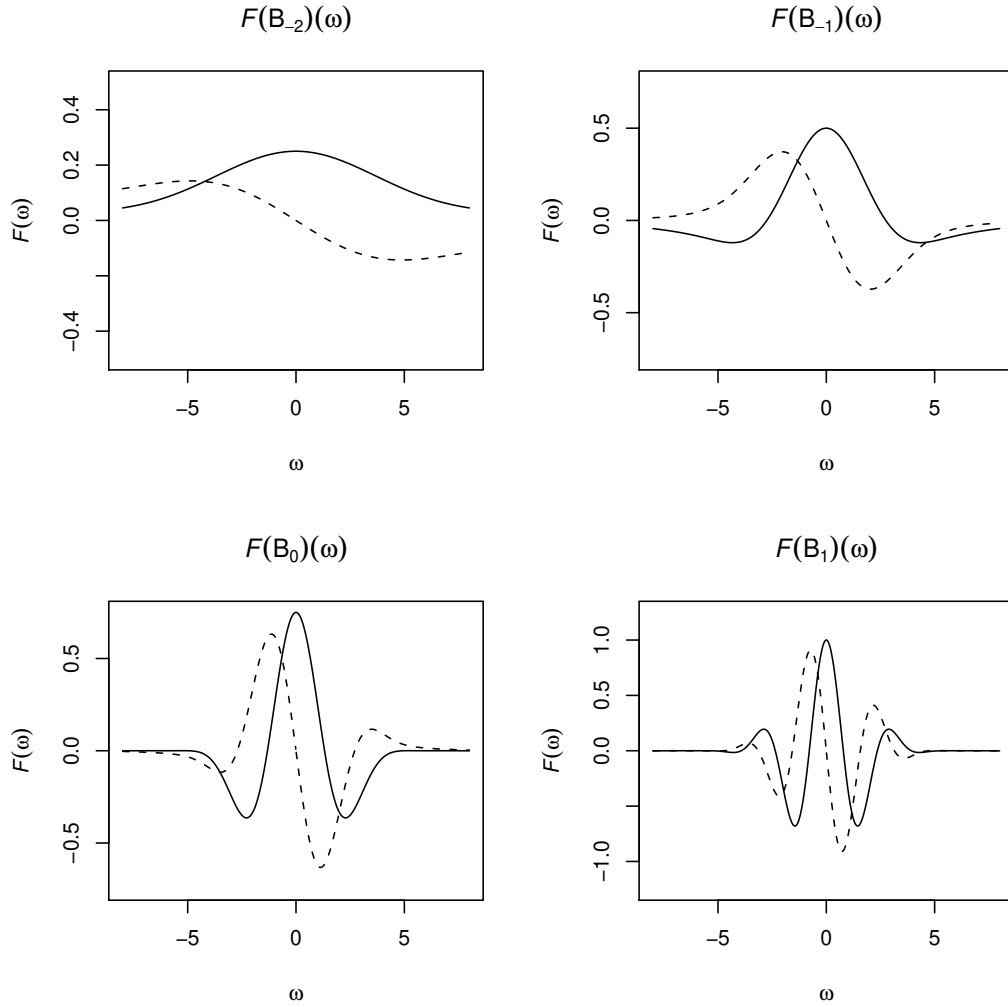


Figura A.3: A Transformada de Fourier das bases $B_{-2,4}(t)$, $B_{-1,4}(t)$, $B_{0,4}(t)$ e $B_{1,4}(t)$, com $r = 1$; a parte real em linha sólida, a parte imaginária em linha pontilhada

e $B_{-\ell,4}(-(t - \tau_L))$, para $\ell = 0, 1, 2$, e aplicar (P.1) e (P.3). Com isso, a Transformada de Fourier de um Spline de ordem 4, usando agora a notação de Ramsay & Silverman (2005), é dada explicitamente por

$$\begin{aligned}
 \mathcal{F}(\hat{y})(w) = & e^{-i\tau_1 w} (\beta_1 \mathcal{F}(B_1)(w) + \beta_2 \mathcal{F}(B_2)(w) + \beta_3 \mathcal{F}(B_3)(w)) \\
 & + e^{i\tau_{K-m+2} w} (\beta_{K-2} \mathcal{F}(B_3)(-w) + \beta_{K-1} \mathcal{F}(B_2)(-w) + \beta_K \mathcal{F}(B_1)(-w)) \quad (\text{A.6}) \\
 & + \sum_{k=4}^{K-3} e^{-i\tau_{k-1} w} \beta_k \left(\frac{\text{sen}(rw/2)}{rw/2} \right)^4
 \end{aligned}$$

Para exemplificar o uso da Transformada de Fourier de Splines, o método será aplicado em uma função suave cuja transformada de Fourier é conhecida: $\exp\{-t^2\}$, que tem transformada dada por $\sqrt{\pi} \exp\{-w^2/4\}$. Restrita ao intervalo $[-T, T]$, a função $\exp\{-t^2\} \mathbf{1}_{[-T, T]}(t)$ tem transformada finita $\sqrt{\pi} \exp\{-w^2/4\} - 2 \int_T^\infty \cos(wt) \exp(-t^2) dt$. Para $T = 4$, a diferença não é maior que $3 \cdot 10^{-8}$, e a função é real para argumentos reais.

Usando $K = 11$ bases, tem-se 7 nós internos equispçados. A Figura A.4 mostra a função $\exp\{-t^2\}$ à esquerda, com a aproximação por Spline, e a transformada de $\exp\{-t^2\}$ à direita, com as partes real e imaginária da transformada de Splines sobrepostas. A parte imaginária é fruto de erro de aproximação. Experimentos subsequentes sugerem que a aproximação da transformada de Fourier é tão boa quanto o ajuste obtido pela aproximação Spline.

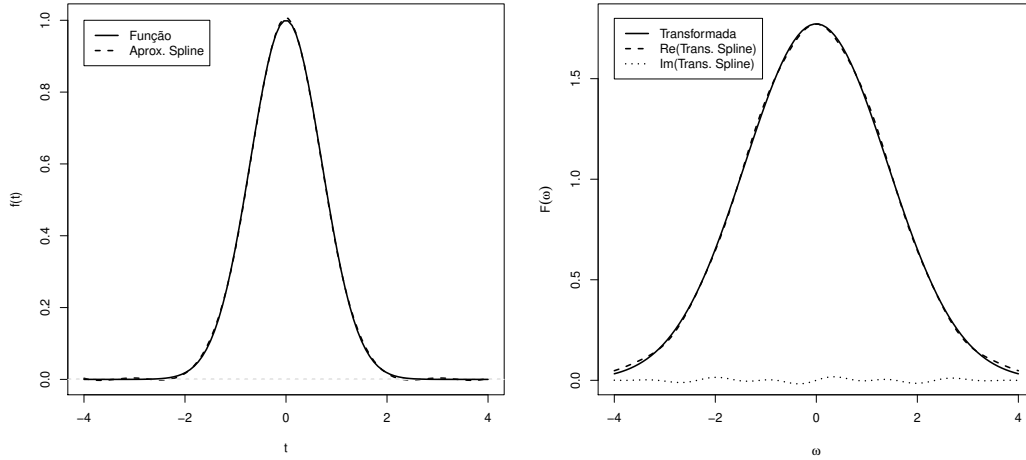


Figura A.4: A função $\exp\{-t^2\}$, sua transformada de Fourier e as respectivas aproximações por Splines

Apêndice B

Algoritmos

Este segundo Apêndice comenta sobre a otimização da verossimilhança do Capítulo 2, e também fornece alguns exemplos de rotinas utilizadas ao longo do projeto.

B.1 Sobre o Método de Nelder-Mead

O algoritmo de Nelder-Mead, disponível no pacote GNU/R na rotina `optim`, é um método numérico de otimização de funções, que não precisa do gradiente da função objetivo, e foi utilizado para obter as estimativas de máxima verossimilhança na Seção 2.3. Contudo, ao se levar em conta que são 22 parâmetros a serem estimados, e a estrutura das matrizes de covariância observadas é mal condicionada, o método exige valores iniciais que estejam razoavelmente próximos dos valores verdadeiros da função.

Como estimativa inicial da variância, foi usada a diagonal da matriz $\hat{\mathbf{Z}}$, e uma modificação do algoritmo que estima/desagrega $\alpha_c(t)$. Novamente ao usar como exemplo o modelo (1.4), é possível observar que

$$\text{diag}(\mathbf{Z}_i)(t) = \sum_{c=1}^C N_{i,c} (\eta_c(t))^2,$$

de onde pode-se obter estimativas da forma

$$\hat{\eta}_c^2(t) = \sum_{k=1}^K \theta_{c,k}^* B_k(t).$$

O valor usado para inicializar o algoritmo é a raiz quadrada de $\hat{\eta}_c^2(t)$. A transformação de raiz não pode ser usada diretamente em $\text{diag}(\mathbf{Z}_i)(t)$ pois é preciso separar as curvas, e o sistema não linear resultante dessa transformação não é fácil de se resolver. Note ainda que esse chute inicial presume que $\eta_c(t) > 0$, mas isso não é sempre verdade. Todas essas ressalvas, porém, devem ser observadas sob a perspectiva de que estas estimativas servirão apenas de chute inicial à rotina de otimização.

Quanto às estimativas iniciais de ϕ , note que o processo $X_c(t)$ (e não $\eta_c(t)X_c(t)$) é, se amostrado a tempo discreto, e sob a condição de estacionariedade fraca, um processo AR(1). Isto porque sua autocorrelação mostra um decaimento exponencial, como observado em Yaglom (1987, p. 78). Isto significa que $(X_c)_{t_r}$, definido como o processo $X_c(t)$ observado nos pontos $t_1, t_2, \dots, t_n$, é dado por

$$X_{t_r} = \rho X_{t_r-1} + \varepsilon_{t_r}, \quad \varepsilon_{t_r} \sim N(0, \sigma_\varepsilon^2).$$

A restrição de estacionariedade para a série temporal AR(1) é que $|\rho| < 1$, pois

$$X_{t_r} - \rho X_{t_r-1} = \varepsilon_{t_r} = (1 - \rho D)X_{t_r},$$

onde D é um operador de diferenciação no tempo, tal que $DX_{t_r} = X_{t_r-1}$. Note ainda que $\text{Var}(X_{t_r}) = \text{Var}(X_{t_r-1})$, por força da hipótese de estacionariedade (de onde se conclui $|D| = 1$). Obtém-se então que

$$X_{t_r} = (1 + \rho D + \rho^2 D^2 + \dots)\varepsilon_{t_r},$$

e a série converge desde que $|\rho| < 1$. A variância de X_{t_r} é simplesmente

$$\begin{aligned} \text{Var}(X_{t_r}) &= \text{Var}(\varepsilon_{t_r}) + \rho^2 \text{Var}(\varepsilon_{t_r-1}) + \rho^4 \text{Var}(\varepsilon_{t_r-2}) + \dots \\ &= (1 + \rho^2 + \rho^4 + \dots) \text{Var}(\varepsilon_{t_r}) = \frac{\sigma_\varepsilon^2}{1 - \rho^2}. \end{aligned}$$

Além disso,

$$\text{Cov}(X_{t_r}, X_{t_r+m}) = \mathbb{E}(X_{t_r} X_{t_r+m}) = \frac{\sigma_\varepsilon^2}{1 - \rho^2} \rho^{|m|} = \frac{\sigma_\varepsilon^2}{1 - \rho^2} e^{|m| \log(\rho)},$$

desde que $\rho > 0$. Como os índices (s, t) são amostrados em períodos nem sempre iguais a 1 de tempo (em geral, em τ^{-1} unidades de tempo), temos que

$$\frac{\sigma_\varepsilon^2}{1 - \rho^2} e^{|m| \log(\rho)} \approx e^{|m| \frac{|t-s|}{\tau-1} \log(\rho)} = e^{-\phi|m||t-s|} \Leftrightarrow \tau|t-s| \log(\rho) = -\phi|t-s|,$$

ou simplesmente que

$$\phi = -\tau \log(\rho).$$

Ainda que o seguinte nem sempre seja verdadeiro, adotou-se por simplicidade que valha uma invariância tal que

$$\hat{\phi} = -\tau \log(\hat{\rho}_{MQ}).$$

onde $\hat{\rho}_{MQ}$ é o estimador de mínimos quadrados de ρ . Contudo observa-se apenas os dados agregados. Quando $C = 2$, por exemplo, tem-se que

$$Z_i(t, s) = N_{i,1} \eta_1(t) \eta_1(s) e^{-\phi_1|t-s|} + N_{i,2} \eta_2(t) \eta_2(s) e^{-\phi_2|t-s|},$$

o que impossibilita uma separação e estimação simultânea de η_c e ϕ_c . Mas se introduzirmos a hipótese $\phi_c = \phi_{c'}, \forall c \neq c'$, a fórmula do exemplo reduz-se a

$$Z_i(t, s) = (N_{i,1} \eta_1(t) \eta_1(s) + N_{i,2} \eta_2(t) \eta_2(s)) e^{-\phi|t-s|},$$

e podemos empregar o inverso do teorema 2.1.4, isto é, que

$$\frac{Y_{i,j}(t) - \hat{Y}_{i,j}(t)}{\sqrt{N_{i,1} \eta_1^2(t) + N_{i,2} \eta_2^2(t)}} \sim N(0, e^{-\phi(\cdot, \cdot)}).$$

A rotina `get.phi.emv`, disponível na Seção B.2, obtém $\hat{\phi}$ a partir de dados já padronizados.

B.2 Rotinas no GNU/R

O objetivo dessa Seção é fornecer algumas rotinas mais complicadas, desenvolvidas para o pacote GNU/R, publicado por R Development Core Team (2009).

Truncamento de Matrizes

A rotina `gera.banda` é uma implementação do truncamento para as estimativas de tipologia, a Fórmula (1.10). A rotina é capaz de gerar uma banda ligeiramente mais versátil, igual a 1 para `banda1` diagonais, e que decai linearmente para zero nas `banda2` subdiagonais seguintes. Basta então multiplicar as matrizes de covariância amostral por ela, dada a estrutura de multiplicação de matrizes entrada a entrada que é implementada no R.

```
gera.banda<-function(L=n,banda1=0,banda2=6){
  banda.dens<-function(T,b1=banda1,b2=banda2){
    temp<-0
    for(i in 1:length(T)){
      if(T[i]<=-(b1+b2)) temp[i]<-0
      if({(-(b1+b2))<T[i]}&&{T[i]<(-b1)}) temp[i]<-(T[i]+b1+b2)/b2
      if({-b1<=T[i]}&&{T[i]<=b1}) temp[i]<-1
      if({b1<T[i]}&&{T[i]<b1+b2}) temp[i]<-(b1+b2-T[i])/b2
      if(T[i]>=(b1+b2))temp[i]<-0
    }
    return(temp)
  }
  temp<-diag(L)
  for(i in 1:L){
    temp[i,]<-banda.dens(1:L-i)
  }
  return(temp)
}
```

Covariância Amostral Agregada

A rotina `sq.em.j` localiza a *i*-ésima curva agregada no vetor `res`, e calcula a co-

variância amostral assumindo que **res** são dados com média zero (e.g., por hipótese, o resíduo do ajuste das curvas agregadas).

```
sq.em.j<-function(res,i,J,n){
  retorna<-matrix(0,ncol=n,nrow=n)
  for(j in 1:J){
    r1<-res[(i-1)*J*n+(j-1)*n+1:n]
    retorna<-retorna+r1%o%r1
  }
  retorna<-retorna/J
  return(retorna)
}
```

Estimador de Tipologias

A seguinte estrutura gera as estimativas de tipologia. É necessário declarar as constantes n , I , C e J (no código, armazenadas nos objetos **n**, **I**, **C** e **J**, respectivamente), bem como a matriz **N** ou a matriz **Q** (armazenada em **N**) e um parâmetro lógico **quimio**, que se verdadeiro usa uma matriz com o quadrado dos coeficientes armazenados em **N** quando for calcular a covariância. **RESP** contém as observações, no formato de vetor, concatenados em função do tempo, seguido da replicação, e por curva agregada, na maneira descrita na Seção 1.4.

```
K <- 6+4
splinebasis=create.bspline.basis(c(min(tempo),max(tempo)), K)
B <- matrix(0,nrow=n*J,ncol=K)
for(w2 in 1:J){
  B[((w2-1)*n+1):(w2*n),] <- eval.basis(splinebasis,tempo)
}
X <- kronecker(N,B)
Beta <- qr.solve(X,RESP)
```

```

residuo <- as.numeric(RESX-X%*%Beta)
hatZ <- array(0,c(n,n,I))
for(i in 1:I){
  hatZ[, ,i] <- cov.em.j(RESX,i,J,n)*gera.banda()
}
cond<-NULL
for(i in 1:I) cond[i] <- (det(hatZ[, ,i])>0)
if(all(cond)&(J>1)){
  hatZ0 <- hatZ+1
  hatZ.inv <- array(0,c(n,n,I))
  for(i in 1:I){
    hatZ.inv[, ,i] <- solve(hatZ[, ,i])
    hatZ.inv[, ,i] <- (hatZ.inv[, ,i]+t(hatZ.inv[, ,i]))/2
    # tira erro de aproximacao (matriz tem q ser simetrica!)
  }
  S.X<-matrix(0,nrow=nrow(X),ncol=ncol(X))
  # Essa matriz guarda SigmaInv%*%X
  for(j in 1:J){
    for(i in 1:I){
      S.X[((i-1)*J*n+(j-1)*n+1):((i-1)*J*n+j*n),] <-
        hatZ.inv[, ,i]%*%X[((i-1)*J*n+(j-1)*n+1):((i-1)*J*n+j*n),]
    }
  }
  W <- t(X)%*%S.X
  W.inv <- solve(W)
  W.inv <- (W.inv+t(W.inv))/2 # Elimina erros de aproximacao
  Beta0 <- W.inv%*%t(S.X)%*%as.matrix(RESX)
  Beta <- Beta0
  ajuste <- X%*%Beta

```

```

residuo0 <- residuo
residuo <- RESP-ajuste
itera <- 0
max.itera <- 50
epsilon <- 10^-3
cond2<-NULL
for(i in 1:I) cond2[i] <- any(abs(hatZ[,i]-hatZ0[,i])>epsilon)
while(any(cond2)&(itera<=max.itera)){
  # Construção da nova matriz de covariancia amostral #
  hatZ0 <- hatZ
  hatZ <- array(0,c(n,n,I))
  for(i in 1:I){
    hatZ[,i] <- sq.em.j(residuo,i,J,n)*gera.banda()
  }
  hatZ.inv <- array(0,c(n,n,I))
  for(i in 1:I){
    hatZ.inv[,i] <- solve(hatZ[,i])
    hatZ.inv[,i] <- (hatZ.inv[,i]+t(hatZ.inv[,i]))/2
  }
  S.X <- matrix(0,nrow=nrow(X),ncol=ncol(X))
  # Essa matriz guarda SigmaInv%*%X
  for(j in 1:J){
    for(i in 1:I){
      S.X[((i-1)*J*n+(j-1)*n+1):((i-1)*J*n+j*n),] <-
        hatZ.inv[,i]%*%X[((i-1)*J*n+(j-1)*n+1):((i-1)*J*n+j*n),]
    }
  } # Fim da Construção da nova matriz de covariancia amostral #
  residuo0 <- residuo
  Beta0 <- Beta

```

```

W <- t(X)%*%S.X
W.inv <- solve(W)
W.inv <- (W.inv+t(W.inv))/2 # Força simetria
Beta <- W.inv%*%t(S.X)%*%as.matrix(ESP)
ajuste <- X%*%Beta
residuo <- ESP-ajuste
cond2 <- NULL
for(i in 1:I)
  cond2[i] <- any(abs(hatZ[, ,i]-hatZ0[, ,i])>epsilon)
  itera <- itera+1
} # Fim do laço de atualização de Z.i
} # Fim do laço if(all(det(hatZ1)>0,det(hatZ2)>0,...,det(hatZI)>0))
itera <- itera-1
Beta <- as.numeric(Beta)
ajuste <- X%*%Beta
Beta.matriz <- matrix(Beta,nrow=K) # Cada coluna corresponde a pop c
Beta.agregado <- N%*%t(Beta.matriz) # cada linha corresponde a
# N[1,1]*Beta[1:K]+n[N,2]*Beta[(K+1):(2*K)]+...

```

Estimador de ϕ para um processo AR(1)

Para um objeto `serie` de tamanho $nJ \times 1$, amostrado em intervalos de tempo τ^{-1} , a rotina `get.phiqm`, aplicada no GNU/R, retorna uma estimativa $\hat{\phi}$ da intensidade do decaimento exponencial.

```

get.phiqm<-function(serie,n,J,tau){
  ret<-0
  temp<-0
  for(j in 1:J){
    temp<-serie[1:n+(j-1)*n]

```

```

    z<-temp[1:(n-1)]*temp[2:n]
    zz<-temp[1:(n-1)]**2
    rho<-sum(z)/sum(zz)
    ret[j]<--tau*log(rho)
  }
return(ret)
}

```

Cálculo do GCV

A matriz R pode ser obtida no pacote `fda` usando o comando `getbasispenalty`. O nome dos objetos da rotina são consistentes com a notação usada na dissertação e, portanto, ela é meramente a fórmula (3.3) escrita na linguagem do R. Para rodar algum algoritmo de otimização sobre a função, como por exemplo a rotina `optimize`, é conveniente fixar os parâmetros y , B , R , K com valores *default*, alterando a linha no cabeçalho para `y = RESPOSTA`, por exemplo.

```

gcv <- function(lambda, y, B, R, K){
  n <- length(y)
  H.lambda <- B%*%solve(t(B)%*%B+lambda*R)%*%t(B)
  hat.y <- H.lambda%*%y
  SQE <- sum((y-hat.y)^2)
  temp <- diag(n) - H.lambda
  tr <- (sum(diag(temp)))^2
  return(n*SQE/tr)
}

```

Cálculo da Transformada de Fourier

A matriz $Fou_{K \times \Omega}$ contém em cada uma de suas linhas a transformada de Fourier da k -ésima base, calculada em quantos pontos ω do espectro ω forem necessários.

Para se obter a Transformada de Fourier de um Spline unidimensional, com parâmetros $\beta_{K \times 1}$, basta calcular $\text{Fou}'\beta$ – o resultado é um vetor $\Omega \times 1$ com a Transformada calculada nos pontos ω do espectro. Para a transformada de Fourier do Produto Tensorial, com parâmetros $\Theta_{K \times K}$, basta calcular $\text{Fou}'\Theta\text{Fou}$. O resultado é uma matriz $\Omega \times \Omega$ com a transformada calculada nos pontos $\omega\omega'$.

```

N1 <- function(w){
  z <- exp(-1i*w/2)*(sin(w/2))/(w/2)
  return(z)
}

DN1 <- function(w){
  z <- 1i + exp(-1i*w)*(w-1i)
  return(z/(w^2))
}

DDN1 <- function(w){
  z <- -2 + 2*exp(1i*w) + w*(-2*1i + w)
  z <- -1i*exp(-1i*w)*z
  return(z/(w^3))
}

DDD1 <- function(w){
  z <- -6*1i + 6*1i*exp(1i*w) + w*(6 - w*(w - 3*1i))
  z <- exp(-1i*w)*z
  return(z/(w^4))
}

FB.ex1 <- function(w,r=1){
  n1 <- N1(r*w)
  dn1 <- DN1(r*w)
  ddn1 <- DDN1(r*w)
  dddn1 <- DDD1(r*w)

```



```

    z <- n1 - 3*1i*dn1 - 3*ddn1 + 1i*ddd1
    return(r*z)
}

FB.ex2 <- function(w,r=1){
  n1 <- N1(r*w)
  dn1 <- DN1(r*w)
  ddn1 <- DDN1(r*w)
  ddd1 <- DDDN1(r*w)
  z <- 2*1i*dn1 + (7/2)*ddn1 - (3/2)*1i*ddd1 +
      (n1)^2 - 2*1i*dn1*n1 - (1/2)*(ddn1*n1 + dn1*dn1)
  return(r*z)
}

FB.ex3 <- function(w,r=1){
  n1 <- N1(r*w)
  dn1 <- DN1(r*w)
  ddn1 <- DDN1(r*w)
  ddd1 <- DDDN1(r*w)
  z <- (1/2)*(-ddn1 + 1i*ddd1) + 1i*dn1*n1 +
      0.5*(dn1)^2 + 0.5*ddn1*n1 + (n1)^3 - 1i*(n1^2)*dn1
  return(r*z)
}

# transformada
knots <- c(0,splinebasis.ft$params,n)
r <- mean(diff(knots))
OMEGA <- 400
omega <- seq(-n/8, n/8, length.out=OMEGA)
Fou <- matrix(0,nrow=K, ncol=OMEGA)
Fou[1,] <- exp(-1i*omega*knots[1])*FB.ex1(omega,r)
Fou[2,] <- exp(-1i*omega*knots[1])*FB.ex2(omega,r)

```

```

Fou[3,] <- exp(-1i*omega*knots[1])*FB.ex3(omega,r)
for(ell in 4:(K-3)){
  Fou[ell,] <- r*exp(-1i*omega*knots[ell-1])*
    ((sin(r*omega/2)/(r*omega/2))^4)
}
Fou[K-2,] <- exp(+1i*omega*knots[K.ft-2])*FB.ex3(-omega,r)
Fou[K-1,] <- exp(+1i*omega*knots[K.ft-2])*FB.ex2(-omega,r)
Fou[K,] <- exp(+1i*omega*knots[K.ft-2])*FB.ex1(-omega,r)

```

Cálculo da expansão em Autofunções

Esse algoritmo realiza a estimação das autofunções ϕ_ν a partir da soma das matrizes de covariância amostrais agregadas.

```

RESP.p<-c(as.numeric(RESP-ajuste))
hatZ1.lzs <- sq.em.j(RESP.p,1,J,n,n)
hatZ2.lzs <- sq.em.j(RESP.p,2,J,n,n)
hatZ3.lzs <- sq.em.j(RESP.p,3,J,n,n)
hatZ.lzs <- hatZ1.lzs+hatZ2.lzs+hatZ3.lzs
Auto <- eigen(hatZ.lzs, symmetric = TRUE)
K.lzs <- n+2
splinebasis.lzs=create.bspline.basis(c(min(tempo),max(tempo)),K.lzs)
X.lzs <- eval.basis(splinebasis.lzs,tempo)
R.lzs <- getbasispenalty(splinebasis.lzs)
L <- T
lambda <- rep(0,L)
for(w in 1:L){
  show(w)
  lambda[w] <- optimize(gcv,c(0.0001,10))$minimum
# observacao: gcv foi modificada para receber os Auto

```

```

# e retornar o gcv, em funcao de lambda, do w-esimo autovetor
}
temp <- which(lambda>9.99)
L <- temp[1]-1
Eigenfunctions.rough.smooth<-matrix(0,n,L)
for(w in 1:L){
  temp.fd <- fd(coef=as.numeric(solve(t(X.lzs)%*%X.lzs+
    lambda[w]*R.lzs)%*%t(X.lzs)%*%Auto$vectors[,w]),
    basisobj=splinebasis.lzs)
  temp.r <- function(t, arg = temp.fd){
    return(as.numeric(eval.fd(temp.fd,t)^2))
  }
  norma <- integrate(temp.r, 0,T)$value
  Eigenfunctions.rough.smooth[,w] <- X.lzs%*%solve(t(X.lzs)%*%X.lzs +
    lambda[w]*R.lzs)%*%t(X.lzs)%*%Auto$vectors[,w] / sqrt(norma)
  # Eigenfunctions.rough.smooth.ns[,w] <- NS.lzs%*%solve(t(NS.lzs)%*%
    NS.lzs+lambda*R.lzs)%*%t(NS.lzs)%*%Auto$vectors[,w]
}
# cria uma matriz de proxies
proxy<-array(0,c(I,J,L))
dt <- mean(diff(tempo))
for(i in 1:I){
  for(j in 1:J){
    for(l in 1:L){
      proxy[i,j,l]<-sum(
        RESP.p[((i-1)*J*T+(j-1)*T+1):((i-1)*J*T+j*T)]*
        Eigenfunctions.rough.smooth[,l])*dt
    }
  }
}

```

```

}
Eigenvalues.rough.smooth<-array(0,c(I,L))
for(i in 1:I){
  for(l in 1:L){
    Eigenvalues.rough.smooth[i,l]<-(1-1/J)*
      var(proxy[i,,l])
  }
}
X.eta<-kronecker(N,diag(L))
Gamma <- qr.solve(X.eta,as.numeric(t(Eigenvalues.rough.smooth)))
Gamma[Gamma<=0]<-0.0001
objetivo<-function(beta,y=as.numeric(t(Eigenvalues.rough.smooth)),x=X.eta){
  return(var(y-x%*%beta))
}
Otimizacao.restrita <- constrOptim(Gamma, objetivo, NULL,
  ui=diag(C*L), ci=rep(0,C*L), control=list(trace=0,maxit=5000))
Gamma <- Otimizacao.restrita$par
Gamma.1<-Gamma[1:L]
Gamma.2<-Gamma[1:L+L]
# Variancia do Consumidor c
hat.Sigma1.lzs<-matrix(0,nrow=n,ncol=n)
hat.Sigma2.lzs<-matrix(0,nrow=n,ncol=n)
for(l in 1:L){
  hat.Sigma1.lzs<-hat.Sigma1.lzs+Gamma.1[l]*
    Eigenfunctions.rough.smooth[,l]%o%Eigenfunctions.rough.smooth[,l]
  hat.Sigma2.lzs<-hat.Sigma2.lzs+Gamma.2[l]*
    Eigenfunctions.rough.smooth[,l]%o%Eigenfunctions.rough.smooth[,l]
}

```