Use R!

Series Editors:
Robert Gentleman Kurt Hornik Giovanni Parmigiani

For other titles published in this series, go to
http://www.springer.com/series/6991

J.O. Ramsay · Giles Hooker · Spencer Graves

Functional Data Analysis with R and MATLAB

Springer

J.O. Ramsay
2748, Howe Street
Ottawa, ON K2B 6W9
Canada
ramsay@psych.mcgill.ca

Giles Hooker
Department of Biological Statistics
& Computational Biology
Cornell University
1186, Comstock Hall
Ithaca, NY 14853
USA
gjh27@cornell.edu

Spencer Graves
Productive Systems Engineering
751, Emerson Ct.
San Jose, CA 95126
USA
spencer.graves@prodsyse.com

Series Editors:

Robert Gentleman
Program in Computational Biology
Division of Public Health Sciences
Fred Hutchinson Cancer
Research Center
1100, Fairview Avenue, N. M2-B876
Seattle, Washington 98109
USA

Kurt Hornik
Department of Statistik
and Mathematik
Wirtschaftsuniversität
Wien Augasse 2-6
A-1090 Wien
Austria

Giovanni Parmigiani
The Sidney Kimmel
Comprehensive Cancer Center
at Johns Hopkins University
550, North Broadway
Baltimore, MD 21205-2011
USA

ISBN 978-0-387-98184-0 e-ISBN 978-0-387-98185-7
DOI 10.1007/978-0-387-98185-7
Springer Dordrecht Heidelberg London New York

Library of Congress Control Number: 2009928040

© Springer Science+Business Media, LLC 2009
All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer Science+Business Media, LLC, 233 Spring Street, New York, NY 10013, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.
The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

This contribution to the useR! series by Springer is designed to show newcomers how to do functional data analysis in the two popular languages, Matlab and R. We hope that this book will substantially reduce the time and effort required to use these techniques to gain valuable insights in a wide variety of applications.

We also hope that the practical examples in this book will make this learning process fun, interesting and memorable. We have tried to choose rich, real-world problems where the optimal analysis has yet to be performed. We have found that applying a spectrum of methods provides more insight than any single approach by itself. Experimenting with graphics and other displays of results is essential.

To support the acquisition of expertise, the "scripts" subdirectory of the companion fda package for R includes files with names like "fdarm-ch01.R", which contain commands in R to reproduce virtually all of the examples (and figures) in the book. This can be found on any computer with R and fda installed using `system.file('scripts', package='fda')`. The Matlab code is provides as part of the fda package for R. From within R, it can be found using `system.file( 'Matlab', package='fda')`. It also can obtained by downloading the `.tar.gz` version of the fda package for R from the Comprehensive R Archive Network (CRAN, `www.r-project.org`), unzipping it and looking for the `inst/Matlab` subdirectory.

The contents of a book are fixed by schedules for editing and printing. These script files are not similarly constrained. Thus, in some cases, the script files may perform a particular analysis differently from how it is described in the book. Such differences will reflect improvements in our understanding of preferred ways of performing the analysis described in the book. The web site www.functionaldata.org is a resource for ongoing developments of software, new tools and current events.

The support for two languages is perhaps a bit unusual in this series, but there are good reasons for this. Matlab is expensive for most users, but its for capacity modeling dynamical systems and other engineering applications has been critical in the development of today's fda package, especially in areas such chemical engineering where functional data are the rule rather than the exception and where Matlab is widely used. On the other hand, the extendibility of R, the easy interface with lower-

level languages, and above all its cost explain its popularity in many fields served by statisticians, students and new researchers. We hope that we can help many of our readers to appreciate the strengths of each language, so as to invest wisely later on. Secondarily, we hope that any user of either language wanting to learn the other can benefit from seeing the same analyses done in both languages.

As with most books in this useR! series, this is not the place to gain enough technical knowledge to claim expertise in functional data analysis nor to develop new tools. But we do hope that some readers will find enough of value here to want to turn to monographs on functional data analysis already published, such as Ramsay and Silverman (2005), and to even newer works.

We wish to end this preface by thanking our families, friends, students, employers, clients and others who have helped make us what we are today and thereby contributed to this book and to our earlier efforts. In particular, we wish to thank John Kimmel of Springer for organizing this series and inviting us to create this book.

James Ramsay, McGill University
Giles Hooker, Cornell University
Spencer Graves, San Jose, CA

Contents

1 Introduction to Functional Data Analysis 1
1.1 What Are Functional Data? 1
1.1.1 Data on the Growth of Girls 1
1.1.2 Data on US Manufacturing 3
1.1.3 Input/Output Data for an Oil Refinery 4
1.2 Multivariate Functional Data 5
1.2.1 Data on How Children Walk 5
1.2.2 Data on Handwriting 7
1.3 Functional Models for Nonfunctional Data 9
1.4 Some Functional Data Analyses 10
1.5 First Steps in a Functional Data Analysis 12
1.5.1 Data Representation: Smoothing and Interpolation 12
1.5.2 Data Registration or Feature Alignment 13
1.5.3 Graphing Functional Data 14
1.5.4 Plotting Pairs of Derivatives: Phase-Plane Plots 15
1.6 Exploring Variability in Functional Data 16
1.6.1 Functional Descriptive Statistics 16
1.6.2 Functional Principal Components Analysis 16
1.6.3 Functional Canonical Correlation 17
1.7 Functional Linear Models 17
1.8 Using Derivatives in Functional Data Analysis 18
1.9 Concluding Remarks 18
1.10 Some Things to Try 19

2 Essential Comparisons of the Matlab and R Languages 21
2.1 A Quick Comparison of Matlab and R Syntax 21
2.1.1 Minor Differences 21
2.1.2 Using Functions in the Two Languages 23
2.2 Singleton Index Issues 24
2.3 Classes and Objects in R and Matlab 24
2.4 More to Read 26

3 How to Specify Basis Systems for Building Functions 29
3.1 Basis Function Systems for Constructing Functions 29
3.2 Fourier Series for Periodic Data and Functions 32
3.3 Spline Series for Nonperiodic Data and Functions 33
3.3.1 Break Points and Knots 33
3.3.2 Order and Degree 34
3.3.3 Examples 34
3.3.4 B-Splines 35
3.3.5 Computational Issues Concerning the Range of t 38
3.4 Constant, Monomial and Other Bases 39
3.4.1 The Constant Basis 39
3.4.2 The Monomial Basis 39
3.4.3 Other Basis Systems 40
3.5 Methods for Functional Basis Objects 40
3.6 The Structure of the `basisfd` or `basis` Class 42
3.7 Some Things to Try 44

4 How to Build Functional Data Objects 45
4.1 Adding Coefficients to Bases to Define Functions 45
4.1.1 Coefficient Vectors, Matrices and Arrays 45
4.1.2 Labels for Functional Data Objects 46
4.2 Methods for Functional Data Objects 48
4.2.1 Illustration: Sinusoidal Coefficients 50
4.3 Smoothing Using Regression Analysis 51
4.3.1 Plotting the January Thaw 52
4.4 The Linear Differential Operator or `Lfd` Class 55
4.5 Bivariate Functional Data Objects: Functions of Two Arguments . 56
4.6 The Structure of the `fd` and `Lfd` Classes 57
4.7 Some Things to Try 57

5 Smoothing: Computing Curves from Noisy Data 59
5.1 Regression Splines: Smoothing by Regression Analysis 59
5.2 Data Smoothing with Roughness Penalties 62
5.2.1 Choosing a Roughness Penalty 62
5.2.2 The Roughness Penalty Matrix $\mathbf{R}$ 64
5.2.3 The Smoothing or "Hat" Matrix and Degrees of Freedom . 65
5.2.4 Defining Smoothing by Functional Parameter Objects 66
5.2.5 Choosing Smoothing Parameter λ 66
5.3 Case Study: The Log Precipitation Data 67
5.4 Positive, Monotone, Density and Other Constrained Functions . . . 70
5.4.1 Positive Smoothing 70
5.4.2 Monotone Smoothing 71
5.4.3 Probability Density Functions 74
5.5 Assessing the Fit to the Data 77
5.6 Details for the `fdPar` Class and `smooth.basis` Function 78

5.6.1 The fdPar class 78
5.6.2 The smooth.basis Function 80
5.7 Some Things to Try 81
5.8 More to Read 82

6 Descriptions of Functional Data 83
6.1 Some Functional Descriptive Statistics 83
6.1.1 The Bivariate Covariance Function $v(s,t)$ 84
6.2 The Residual Variance-Covariance Matrix Σ_e 87
6.3 Functional Probes ρ_ξ 87
6.4 Phase-Plane Plots of Periodic Effects 88
6.4.1 Phase-Plane Plots Show Energy Transfer 88
6.4.2 The Nondurable Goods Cycles 90
6.4.3 Phase-Plane Plotting the Growth of Girls 91
6.5 Confidence Intervals for Curves and Their Derivatives 92
6.5.1 Two Linear Mappings Defining a Probe Value 93
6.5.2 Computing Confidence Limits for Probe Values 95
6.5.3 Confidence Limits for Prince Rupert's Log Precipitation . . 95
6.6 Some Things to Try 96

7 Exploring Variation: Functional Principal and Canonical Components Analysis 99
7.1 An Overview of Functional PCA 100
7.2 PCA with Function pca.fd 102
7.2.1 PCA of the Log Precipitation Data 103
7.2.2 PCA of Log Precipitation Residuals 106
7.3 More Functional PCA Features 106
7.4 PCA of Joint X-Y Variation in Handwriting 108
7.5 Exploring Functional Covariation with Canonical Correlation Analysis 110
7.6 Details for the pca.fd and cca.fd Functions 113
7.6.1 The pca.fd Function 113
7.6.2 The cca.fd Function 114
7.7 Some Things to Try 114
7.8 More to Read 115

8 Registration: Aligning Features for Samples of Curves 117
8.1 Amplitude and Phase Variation 117
8.2 Time-Warping Functions and Registration 119
8.3 Landmark Registration with Function landmarkreg 121
8.4 Continuous Registration with Function register.fd 122
8.5 A Decomposition into Amplitude and Phase Sums of Squares 125
8.6 Registering the Chinese Handwriting Data 126
8.7 Details for Functions landmarkreg and register.fd 127
8.7.1 Function landmarkreg 127

8.7.2 Function register.fd 128
8.8 Some Things to Try 129
8.9 More to Read 130

9 Functional Linear Models for Scalar Responses 131
9.1 Functional Linear Regression with a Scalar Response 131
9.2 A Scalar Response Model for Log Annual Precipitation 132
9.3 Setting Up the Functional Linear Model 132
9.4 Three Estimates of the Regression Coefficient Predicting Annual Precipitation 133
9.4.1 Low-Dimensional Regression Coefficient Function β 134
9.4.2 Coefficient β Estimate Using a Roughness Penalty 135
9.4.3 Choosing Smoothing Parameters 138
9.4.4 Confidence Intervals 140
9.4.5 Scalar Response Models by Functional Principal Components 141
9.5 Statistical Tests 143
9.6 Some Things to Try 145
9.7 More to Read 146

10 Linear Models for Functional Responses 147
10.1 Functional Responses and an Analysis of Variance Model 147
10.1.1 Climate Region Effects on Temperature 147
10.1.2 Trends in Seabird Populations on Kodiak Island 149
10.1.3 Choosing Smoothing Parameters 153
10.2 Functional Responses with Functional Predictors: The Concurrent Model 154
10.2.1 Estimation for the Concurrent Model 155
10.2.2 Confidence Intervals for Regression Functions 157
10.2.3 Knee Angle Predicted from Hip Angle 158
10.3 Beyond the Concurrent Model 162
10.4 A Functional Linear Model for Swedish Mortality 163
10.5 Permutation Tests of Functional Hypotheses 165
10.5.1 Functional t-Tests 166
10.5.2 Functional F-Tests 168
10.6 Details for R Functions fRegress, fRegress.CV and fRegress.stderr 169
10.6.1 Function fRegress 169
10.6.2 Function fRegress.CV 173
10.6.3 Function fRegress.stderr 173
10.7 Details for Function plotbeta 174
10.8 Details for Function linmod 174
10.9 Details for Functions Fperm.fd and tperm.fd 175
10.9.1 Function Fperm.fd 175
10.9.2 Function tperm.fd 176

10.10 Some Things to Try 177
10.11 More to Read 177

11 Functional Models and Dynamics 179
11.1 Introduction to Dynamics 179
11.1.1 An Example of First-Order Dynamics 180
11.1.2 Interpreting Second-Order Linear Dynamics 181
11.1.3 Higher-Dimensional Linear Systems 183
11.2 Principal Differential Analysis for Linear Dynamics 184
11.3 Principal Differential Analysis of the Lip Data 185
11.4 PDA of the Handwriting Data 187
11.5 Registration and PDA 190
11.6 Details for `pda.fd`, `eigen.pda`, `pda.overlay` and `register.newfd` 191
11.6.1 Function `pda.fd` 191
11.6.2 Function `eigen.pda` 192
11.6.3 Function `pda.overlay` 193
11.6.4 Function `register.newfd` 193
11.7 Some Things to Try 193
11.8 More to Read 194

Symbol Table 197

References 199

Index 203