

Symmetry Studies

Experimental data can often be associated with or indexed by certain symmetrically interesting structures or sets of labels that appear, for example, in the study of short symbolic sequences in molecular biology, in preference or voting data, in visual field and corneal topography arrays, or in experimental refractive optics. The symmetry studies introduced in this book describe the interplay among symmetry transformations that are characteristic of these sets of labels, the resulting algebraic decomposition of the data that are indexed by them, and the research questions that are induced by those transformations. The overall purpose is to facilitate and guide the statistical study of the structured data from both a descriptive and inferential perspective. The text combines notions of algebra and statistics and develops a systematic methodology to better explore the many different data-analytic applications of symmetry.

CAMBRIDGE SERIES IN STATISTICAL AND PROBABILISTIC MATHEMATICS

Editorial Board:

- R. Gill, *Department of Mathematics, Utrecht University*
B. D. Ripley, *Department of Statistics, University of Oxford*
S. Ross, *Epstein Department of Industrial & Systems Engineering, University of Southern California*
B. W. Silverman, *St. Peter's College, Oxford*
M. Stein, *Department of Statistics, University of Chicago*

This series of high-quality upper-division textbooks and expository monographs covers all aspects of stochastic applicable mathematics. The topics range from pure and applied statistics to probability theory, operations research, optimization, and mathematical programming. The books contain clear presentations of new developments in the field and also of the state of the art in classical methods. While emphasizing rigorous treatment of theoretical methods, the books also contain applications and discussions of new techniques made possible by advances in computational practice.

Already Published

1. *Bootstrap Methods and Their Application*, by A. C. Davison and D. V. Hinkley
2. *Markov Chains*, by J. Norris
3. *Asymptotic Statistics*, by A. W. van der Vaart
4. *Wavelet Methods for Time Series Analysis*, by Donald B. Percival and Andrew T. Walden
5. *Bayesian Methods*, by Thomas Leonard and John S. J. Hsu
6. *Empirical Processes in M-Estimation*, by Sara van de Geer
7. *Numerical Methods of Statistics*, by John F. Monahan
8. *A User's Guide to Measure Theoretic Probability*, by David Pollard
9. *The Estimation and Tracking of Frequency*, by B. G. Quinn and E. J. Hannan
10. *Data Analysis and Graphics Using R*, by John Maindonald and John Braun
11. *Statistical Models*, by A. C. Davison
12. *Semiparametric Regression*, by D. Ruppert, M. P. Wand, and R. J. Carroll
13. *Exercise in Probability*, by Loic Chaumont and Marc Yor
14. *Statistical Analysis of Stochastic Processes in Time*, by J. K. Lindsey
15. *Measure Theory and Filtering*, by Lakhdar Aggoun and Robert Elliott
16. *Essentials of Statistical Inference*, by G. A. Young and R. L. Smith
17. *Elements of Distribution Theory*, by Thomas A. Severini
18. *Statistical Mechanics of Disordered Systems*, by Anton Bovier
19. *The Coordinate-Free Approach to Linear Models*, by Michael J. Wichura

Symmetry Studies

An Introduction to the Analysis of Structured Data
in Applications

MARLOS A. G. VIANA

University of Illinois at Chicago



CAMBRIDGE
 UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom
 One Liberty Plaza, 20th Floor, New York, NY 10006, USA
 477 Williamstown Road, Port Melbourne, VIC 3207, Australia
 314-321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre, New Delhi - 110025, India
 103 Penang Road, #05-06/07, Visioncrest Commercial, Singapore 238467

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning and research at the highest international levels of excellence.

www.cambridge.org

Information on this title: www.cambridge.org/9780521841030

© Marlos A. G. Viana 2008

This publication is in copyright. Subject to statutory exception and to the provisions of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

First published 2008

A catalogue record for this publication is available from the British Library

Library of Congress Cataloging in Publication data

Viana, Marlos A. G.

Symmetry studies : an introduction to the analysis of structured data in applications / Marlos Viana.

p. cm. — (Cambridge series in statistical and probabilistic mathematics)

Includes bibliographical references and index.

ISBN-13: 978-0-521-84103-0 (hardback)

ISBN-10: 0-521-84103-8 (hardback)

1. Symmetry (Mathematics) 2. Statistics – Data processing.

3. Probabilities – Data processing. 4. Data structures (Computer science)

I. Title. II. Series.

QA174.7.S96V53 2008

519.5–dc22 2008002983

ISBN 978-0-521-84103-0 Hardback

Cambridge University Press has no responsibility for the persistence or accuracy of URLs for external or third-party internet websites referred to in this publication, and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.

Contents

<i>Preface</i>	<i>page ix</i>
1. Symmetry, Classification, and the Analysis of Structured Data	1
1.1 Introduction	1
1.2 Symmetry and Classification	2
1.3 Data Indexed by Symmetries	3
1.4 Symmetry and Data Reduction	5
1.5 Statistical Aspects	8
1.6 Algebraic Aspects	9
1.7 Structured Data	11
1.8 Partitions	15
1.9 Summary	23
Further Reading	23
Exercises	25
Appendix A	29
2. Sorting the Labels: Group Actions and Orbits	30
2.1 Introduction	30
2.2 Permutations	30
2.3 Groups and Homomorphisms	34
2.4 Group Actions	47
2.5 Actions on Mappings	49
2.6 Genotypic Classification	53
2.7 Actions on Binary Trees and Branches	55
2.8 Counting Orbits in Linkage Analysis	56
2.9 Burnside's Lemma	58
Further Reading	59
Exercises	59
3. Connecting Symmetries and Data: Linear Representations	63
3.1 Introduction	63
3.2 Representations	63
3.3 Unitary Representations	66

3.4	Regular Representations and Group Algebras	68
3.5	Tensor Representations	69
3.6	Matrices with Group Symmetry	70
3.7	Reducibility	74
3.8	Schur's Lemma	79
3.9	Constructing the Irreducible Representations of S_n	82
	Further Reading	87
	Exercises	88
4.	Data Reduction and Inference: The Canonical Projections and Their Invariants	91
4.1	Introduction	91
4.2	Characters of a Linear Representation	91
4.3	Orthogonality Relations for Characters	92
4.4	The Canonical Projections	101
4.5	The Standard Decomposition	112
4.6	Harmonic Analysis	117
	Exercises	122
5.	Examples and Techniques	128
5.1	Introduction	128
5.2	Analysis of Variance	128
5.3	Classical Structures and Their Analysis of Variance	130
5.4	The Analysis of Variance for Dihedral Data	134
5.5	Cyclic Reduction of Binary Sequences	134
5.6	Dihedral Reduction of Binary Sequences	136
5.7	Projections in the Space of Scalar-Valued Functions	138
5.8	Decompositions in the Dual Space	139
5.9	Canonical Decompositions of Entropy Data	148
5.10	A Two-Way Cyclic Decomposition	153
5.11	Data Structures Induced by Point Groups	154
5.12	Data Indexed by Homogeneous Polynomials	156
5.13	Likelihood Decompositions	158
	Further Reading	160
	Exercises	161
6.	Symmetry Studies of Short Symbolic Sequences	165
6.1	Introduction	165
6.2	Symmetry Studies of Four Sequences in Length of 3	165
6.3	Reductions by Position Symmetry	166
6.4	Reductions by Symbol Symmetry	171
6.5	Dihedral Studies	173
	Further Reading	175
	Exercises	175
	Appendix A	177
	Appendix B	178

Contents

vii

7.	Symmetry Studies of Curvature Data	180
	7.1 Introduction	180
	7.2 Keratometry	181
	7.3 Astigmatic and Stigmatic Constraints	181
	7.4 Ranking Permutations	182
	7.5 Classification of Astigmatic Mappings	185
	7.6 Astigmatic Covariance Structures	190
	7.7 Dihedral Fourier Analysis	190
	Further Reading	194
	Exercises	195
	Appendix A	198
	Appendix B	199
	Appendix C	200
8.	Symmetry Studies of Planar Chirality	201
	8.1 Introduction	201
	8.2 Characterizing Rotations and Reversals	203
	8.3 Canonical Classification of Handedness in Elementary Images	205
	8.4 Handedness of Short Symbolic Sequences	211
	Further Reading	213
	Exercises	214
	Appendix A	217
	Appendix B	218
	Appendix C	220
	<i>Appendix A: Computing Algorithms</i>	221
	<i>Appendix B: Glossary of Selected Symbols, Notations, and Terms</i>	225
	<i>Bibliography</i>	227
	<i>Index</i>	233

Preface

This text is an introduction to data-analytic applications of symmetry principles and arguments or symmetry studies. Its motivation comes from a variety of disciplines in which these principles continue to play a significant role in describing natural phenomena, and from the goal of methodologically applying them to classification, description, and analysis of data. The product of the methodology presented here is a broader class of data-analytic tools derived from well-established and theoretically related areas in algebra and statistics, such as group representations and analysis of variance.

The principles discussed in the text reflect many defining aspects of symmetry, a Greek conception dating from the Hellenic Era and part of a class of terms and forms of expression that designated harmony, rhythm, balance, stability, good proportions, and evenness of structure. Early Greek art and architecture often capture the outstanding dualism intrinsic to the original notion of symmetry – that of retaining the static uniqueness of one's being and at the same time promoting its dynamical multipresent realizations. This dualism is only apparently hidden in the 12th-century Athenian detail shown in the front cover. I invite you to recognize the presence of these pleasant concepts in the methodology to be introduced in the coming pages.

The text is divided as follows. Chapter 1 gives a complete overview of the methodology, including an introduction to the concepts of data indexed by symmetries, finite groups, group actions, orbits and classification, and representations in the data space. At the same time it outlines the step-by-step connection between the algebra and statistical inference, in the context of analysis of variance. It also emphasizes the fact that the same symmetry arguments useful here for classification and data reduction are part of the common language of the chemist, the geneticist, and the physicist.

Chapters 2, 3, and 4 cover the algebraic background, presented along with characteristic data-analytic applications. Readers with a basic course in linear or abstract algebra will have the required preparedness to follow these chapters, with focus on the applications and motivation to complete the proposed exercises. Their familiarity with the principles of analysis of variance will make more evident, broader, and attractive the applications introduced in these chapters.

Chapter 5 describes a number of prototypic applications, ranging from classical examples in experimental designs to specific symmetry studies of data indexed by objects with dihedral and cyclic symmetries. The more detailed applications, described in Chapters 6, 7, and 8, are symmetry studies for data indexed by short symbolic sequences, corneal curvature, and the study of handedness in simple planar images.

The chapters include briefly annotated suggestions for further reading and complementary exercises, followed by a chapter with basic computing algorithms in ©Maple and a glossary of main symbols and notations. The text also includes a selected number of short biographical citations, abstracted from very valuable sources, including The MacTutor History of Mathematics Archives at the University of St. Andrews and The Nobel Lectures (Elsevier, Amsterdam). The reader is invited to send notice of errors of any form to the author at viana@uic.edu.

This text evolved from lecture notes and talks prepared for short courses and special sessions held at several institutions since 2003, including the Instituto Nacional de Matemática Pura e Aplicada (IMPA) in my hometown (Rio de Janeiro), the EURANDOM, the Greek Statistical Institute, the International Commission for Optics, the Sociedad Chilena de Estadística, the Universidad de Antofagasta, University of Connecticut, Universidad de Costa Rica, University of Cyprus, Eindhoven University of Technology, Indiana University, University of Piraeus, Universidade Federal do Rio de Janeiro, Universidade de São Paulo, Universidad Simón Bolívar, and St. Petersburg State (ICMO).

Throughout this project, I benefited from the dedicated suggestions, conversations, enthusiasm, and inspiration of many colleagues, to whom I owe my gratitude. In particular, to Steen Andersson, Henrik Aratyn, Arjeh Cohen, Persi Diaconis, Alessandro Di Bucchianico, Joe Glaz, Markos Koutras, Vasudevan Lakshminarayanan, Gérard Letac, Peter McCullagh, Ingram Olkin, Takis Papaioannou, Carlos de B. Pereira, Michael Perlman, Donald Richards, Stephen Smith, Peter van de Ven, and Henry Wynn. This text also reflects the constant enthusiasm of my students at the Honors College who attended our weekly “Symmetry in Science and Applications” seminar.

I am thankful to D. Azar, J. Bauman, and L. Kaufman for their timely facilitation of the institutional support for both research and teaching, essential to the completion of this text.

My special thanks go to my editor, Lauren Cowles with Cambridge University Press, for her professional guidance, encouragement, and constant attention to the project.

This work is dedicated to my wife, Grace, and our children, Alice, Andrew, and Alex, who together gave me the strength to pursue it over the years.

Marlos A. G. Viana
Chicago, April 23, 2008.