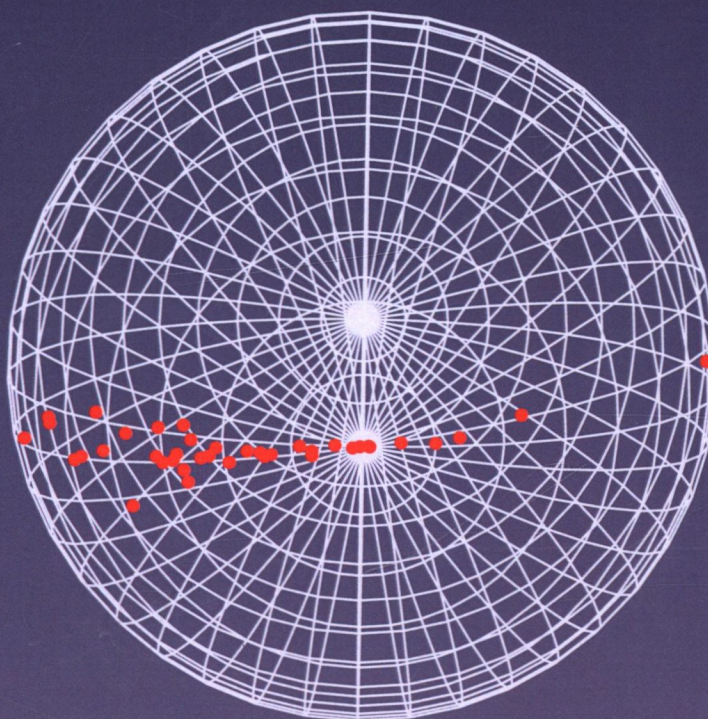


Chapman & Hall/CRC
Interdisciplinary Statistics Series

Modern Directional Statistics



Christophe Ley • Thomas Verdebout



CRC Press
Taylor & Francis Group

A CHAPMAN & HALL BOOK

**Chapman & Hall/CRC
Interdisciplinary Statistics Series**

Modern Directional Statistics

Modern Directional Statistics collects important advances in methodology and theory for directional statistics over the last two decades. It provides a detailed overview and analysis of recent results that can help both researchers and practitioners. Knowledge of multivariate statistics eases the reading but is not mandatory.

The field of directional statistics has received a lot of attention over the past two decades, due to new demands from domains such as life sciences or machine learning, to the availability of massive data sets requiring adapted statistical techniques, and to technological advances. This book covers important progress in distribution theory, high-dimensional statistics, kernel density estimation, efficient inference on directional supports, and computational and graphical methods.

Christophe Ley is professor of mathematical statistics at Ghent University. His research interests include semi-parametrically efficient inference, flexible modeling, directional statistics and the study of asymptotic approximations via Stein's Method. His achievements include the Marie-Jeanne Laurent-Duhamel prize of the Société Française de Statistique and an elected membership at the International Statistical Institute. He is associate editor for the journals *Computational Statistics & Data Analysis* and *Econometrics and Statistics*.

Thomas Verdebout is professor of mathematical statistics at Université libre de Bruxelles (ULB). His main research interests are semi-parametric statistics, high-dimensional statistics, directional statistics and rank-based procedures. He has won an annual prize of the Belgian Academy of Sciences and is an elected member of the International Statistical Institute. He is associate editor for the journals *Statistics and Probability Letters* and *Journal of Multivariate Analysis*.



CRC Press
Taylor & Francis Group
an informa business

6000 Broken Sound Parkway, NW
Suite 300, Boca Raton, FL 33487
711 Third Avenue
New York, NY 10017
2 Park Square, Milton Park
Abingdon, Oxon OX14 4RN, UK

K24840

ISBN: 978-1-4987-0664-3

90000



9 781498 706643

www.crcpress.com

Modern Directional Statistics

Ley • Verdebout
CRC Press

Chapman & Hall/CRC
Interdisciplinary Statistics Series

Modern Directional Statistics

Christophe Ley • Thomas Verdebout



CRC Press

Taylor & Francis Group

Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

A CHAPMAN & HALL BOOK

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

© 2017 by Taylor & Francis Group, LLC
CRC Press is an imprint of Taylor & Francis Group, an Informa business

No claim to original U.S. Government works

Printed on acid-free paper by Ashford Colour Press Ltd.
Version Date: 20170522

International Standard Book Number-13: 978-1-4987-0664-3 (Hardback)

This book contains information obtained from authentic and highly regarded sources. Reasonable efforts have been made to publish reliable data and information, but the author and publisher cannot assume responsibility for the validity of all materials or the consequences of their use. The authors and publishers have attempted to trace the copyright holders of all material reproduced in this publication and apologize to copyright holders if permission to publish in this form has not been obtained. If any copyright material has not been acknowledged please write and let us know so we may rectify in any future reprint.

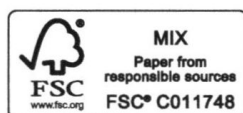
Except as permitted under U.S. Copyright Law, no part of this book may be reprinted, reproduced, transmitted, or utilized in any form by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying, microfilming, and recording, or in any information storage or retrieval system, without written permission from the publishers.

For permission to photocopy or use material electronically from this work, please access www.copyright.com (<http://www.copyright.com/>) or contact the Copyright Clearance Center, Inc. (CCC), 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400. CCC is a not-for-profit organization that provides licenses and registration for a variety of users. For organizations that have been granted a photocopy license by the CCC, a separate system of payment has been arranged.

Trademark Notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Visit the Taylor & Francis Web site at
<http://www.taylorandfrancis.com>

and the CRC Press Web site at
<http://www.crcpress.com>



Modern Directional Statistics

CHAPMAN & HALL/CRC

Interdisciplinary Statistics Series

Series editors: N. Keiding, B.J.T. Morgan, C.K. Winkle, P. van der Heijden

Published titles

AGE-PERIOD-COHORT ANALYSIS: NEW MODELS, METHODS, AND EMPIRICAL APPLICATIONS Y. Yang and K. C. Land

ANALYSIS OF CAPTURE-RECAPTURE DATA R. S. McCrea and B. J. T. Morgan

AN INVARIANT APPROACH TO STATISTICAL ANALYSIS OF SHAPES
S. Lele and J. Richtsmeier

ASTROSTATISTICS G. Babu and E. Feigelson

BAYESIAN ANALYSIS FOR POPULATION ECOLOGY R. King, B. J. T. Morgan,
O. Gimenez, and S. P. Brooks

**BAYESIAN DISEASE MAPPING: HIERARCHICAL MODELING IN SPATIAL
EPIDEMIOLOGY, SECOND EDITION** A. B. Lawson

BIOEQUIVALENCE AND STATISTICS IN CLINICAL PHARMACOLOGY
S. Patterson and B. Jones

CLINICAL TRIALS IN ONCOLOGY, THIRD EDITION S. Green, J. Benedetti, A. Smith,
and J. Crowley

CLUSTER RANDOMISED TRIALS R. J. Hayes and L. H. Moulton

CORRESPONDENCE ANALYSIS IN PRACTICE, THIRD EDITION M. Greenacre

THE DATA BOOK: COLLECTION AND MANAGEMENT OF RESEARCH DATA
M. Zozus

**DESIGN AND ANALYSIS OF QUALITY OF LIFE STUDIES IN CLINICAL TRIALS,
SECOND EDITION** D. L. Fairclough

DYNAMICAL SEARCH L. Pronzato, H. Wynn, and A. Zhigljavsky

FLEXIBLE IMPUTATION OF MISSING DATA S. van Buuren

**GENERALIZED LATENT VARIABLE MODELING: MULTILEVEL, LONGITUDI-
NAL, AND STRUCTURAL EQUATION MODELS** A. Skrondal and S. Rabe-Hesketh

GRAPHICAL ANALYSIS OF MULTI-RESPONSE DATA K. Basford and J. Tukey

**INTRODUCTION TO COMPUTATIONAL BIOLOGY: MAPS, SEQUENCES, AND
GENOMES** M. Waterman

MARKOV CHAIN MONTE CARLO IN PRACTICE W. Gilks, S. Richardson, and
D. Spiegelhalter

**MEASUREMENT ERROR AND MISCLASSIFICATION IN STATISTICS AND EPIDE-
MIOLOGY: IMPACTS AND BAYESIAN ADJUSTMENTS** P. Gustafson

Published titles

MEASUREMENT ERROR: MODELS, METHODS, AND APPLICATIONS

J. P. Buonaccorsi

MEASUREMENT ERROR: MODELS, METHODS, AND APPLICATIONS

J. P. Buonaccorsi

MENDELIAN RANDOMIZATION: METHODS FOR USING GENETIC VARIANTS IN CAUSAL ESTIMATION

S. Burgess and S. G. Thompson

META-ANALYSIS OF BINARY DATA USING PROFILE LIKELIHOOD

D. Böhning, R. Kuhnert, and S. Rattanasiri

MISSING DATA ANALYSIS IN PRACTICE

T. Raghunathan

MODERN DIRECTIONAL STATISTICS

C. Ley and T. Verdebout

POWER ANALYSIS OF TRIALS WITH MULTILEVEL DATA

M. Moerbeek and S. Teerenstra

SPATIAL POINT PATTERNS: METHODOLOGY AND APPLICATIONS WITH R

A. Baddeley, E. Rubak, and R. Turner

STATISTICAL ANALYSIS OF GENE EXPRESSION MICROARRAY DATA

T. Speed

STATISTICAL ANALYSIS OF QUESTIONNAIRES: A UNIFIED APPROACH BASED ON R AND STATA

F. Bartolucci, S. Bacci, and M. Gnaldi

STATISTICAL AND COMPUTATIONAL PHARMACOGENOMICS

R. Wu and M. Lin

STATISTICS IN MUSICOLOGY

J. Beran

STATISTICS OF MEDICAL IMAGING

T. Lei

STATISTICAL CONCEPTS AND APPLICATIONS IN CLINICAL MEDICINE

J. Aitchison, J. W. Kay, and I. J. Lauder

STATISTICAL AND PROBABILISTIC METHODS IN ACTUARIAL SCIENCE

P. J. Boland

STATISTICAL DETECTION AND SURVEILLANCE OF GEOGRAPHIC CLUSTERS

P. Rogerson and I. Yamada

STATISTICS FOR ENVIRONMENTAL BIOLOGY AND TOXICOLOGY

A. Bailer and W. Piegorsch

STATISTICS FOR FISSION TRACK ANALYSIS

R. F. Galbraith

VISUALIZING DATA PATTERNS WITH MICROMAPS

D. B. Carr and L. W. Pickle

Preface

Aim of the book

The field of directional statistics has received a lot of attention over the past two decades. This is mainly due to new demands from domains such as bioinformatics, machine learning and cosmology, but also to the availability of massive data requiring adapted statistical techniques, and to technological advances. Our goal is to provide a thorough overview of the developments that have taken place since the beginning of the 21st century, hereby building upon earlier works like the seminal and inspiring monograph by Mardia & Jupp (2000). We shall focus here solely on theoretical developments; the description of modern datasets and their analysis will be provided in the companion book *Applied Directional Statistics: Modern Methods and Case Studies* for which the authors are acting as editors.

Complementarity with existing literature

The present book has a well-defined target: reference the methodological improvements on directional statistics that have appeared after the cornerstone reference books by Mardia & Jupp (2000) and Jammalamadaka & SenGupta (2001). This has so far been only partially done by Pewsey et al. (2013) but specifically for the circular setting and with more focus on applications. The present book fills this gap and will therefore be of interest for researchers and practitioners dealing with directional data and aiming for a recent methodological flavour.

Acknowledgments

There are a number of people we wish to thank for their contributions to our book. We thank our editor, John Kimmel, for insightful exchanges and for the initial proposal to write this book. We are grateful to Arthur Pewsey for a very careful reading of the entire book, to Toshihiro Abe, Rosa Crujeiras, Yves Dominicy,

Oliver Dukes, Eduardo García-Portugués, Thomas Hamelryck, Shogo Kato, Irene Klugkist, Davy Paindaveine, Thanh Mai Pham Ngoc and Yvik Swan for their helpful comments and suggestions, to José Ameireijas-Alonso for his help with some of the graphics, and to Kim Schalbar for checking the bibliography.

Our very special thanks go to our families, to whom we dedicate this book.

Contents

1	Introduction	1
1.1	Overview	1
1.1.1	A brief introduction to directional statistics	1
1.1.2	A brief outline of the theoretical advances presented in this book	2
1.2	Directional datasets	4
1.2.1	Paleomagnetism	4
1.2.2	Political sciences	6
1.2.3	Text mining	6
1.2.4	Wildfire orientation	7
1.2.5	Life sciences and bioinformatics	8
1.3	Basics and notations	10
1.4	Plan of the book	12
2	Advances in flexible parametric distribution theory	17
2.1	Introduction	17
2.1.1	Flexible parametric modeling: an active research area on $\mathbb{R}^p$	17
2.1.2	Organization of the remainder of the chapter	18
2.2	Flexible circular distributions	18
2.2.1	Four ways to construct circular densities	18
2.2.2	The classics: von Mises, cardioid and wrapped Cauchy distributions	19
2.2.3	Beyond the classics: modern flexible circular modeling	22
2.2.4	Flexible modeling of symmetric data: the Jones–Pewsey distribution	22
2.2.5	Sine-skewing: a simple tool to skew any symmetric distribution	24

2.2.6	Skewness combined with unimodality: the scale-transforming approach	27
2.2.7	A general device for building symmetric bipolar distributions	28
2.2.8	A brief description of three other flexible models	30
2.3	Flexible spherical distributions	35
2.3.1	Classical spherical distributions	35
2.3.2	Rotationally symmetric distributions	37
2.3.3	A general method to skew rotationally symmetric distributions	40
2.4	Flexible toroidal and cylindrical distributions	42
2.4.1	Some history, motivations and goals	42
2.4.2	The bivariate von Mises distribution and its variants	43
2.4.3	Mardia–Sutton type cylindrical distributions	46
2.4.4	Johnson–Wehrly type cylindrical distributions	47
2.4.5	The copula approach	50
2.5	Further reading	52
3	Advances in kernel density estimation on directional supports	55
3.1	Introduction	55
3.1.1	Kernel density estimation on the real line	55
3.1.2	Organization of the remainder of the chapter	57
3.2	Definitions and main properties	57
3.2.1	Spherical kernel density estimation	58
3.2.2	Cylindrical kernel density estimation	60
3.3	A delicate yet crucial issue: bandwidth choice	61
3.3.1	Spherical AMISE and bandwidth selection	61
3.3.2	Rule of thumb based on the FvML distribution	63
3.3.3	A gain in generality: AMISE via mixtures of FvML densities	64
3.3.4	Three further proposals	65
3.3.5	Bandwidth selection in the cylindrical setting	66
3.4	Inferential procedures	66
3.4.1	Non-parametric goodness-of-fit test for directional data	66
3.4.2	Non-parametric independence test for cylindrical data	68
3.4.3	An overview of non-parametric regression	69
3.5	Further reading	70

4	Computational and graphical methods	73
4.1	Ordering data on the sphere: quantiles and depth functions	73
4.1.1	Ordering on $\mathbb{R}$ and $\mathbb{R}^p$, and organization of the remainder of the section	73
4.1.2	Classical depth functions on the sphere	74
4.1.3	Projected quantiles and related asymptotic results	75
4.1.4	The angular Mahalanobis depth	78
4.1.5	Statistical procedures based on projected quantiles and the angular Mahalanobis depth	78
4.2	Statistical inference under order restrictions on the circle	81
4.2.1	Isotonic regression estimation and organization of the remainder of the section	81
4.2.2	Order restrictions on the circle	82
4.2.3	Circular isotonic regression	83
4.3	Exploring data features with the CircSiZer	84
4.3.1	The SiZer, scale space theory and organization of the remainder of the section	84
4.3.2	The CircSiZer	86
4.3.3	Kernel choice based on causality: the special role of the wrapped normal	88
4.4	Computationally fast estimation for high-dimensional FvML distributions	89
4.4.1	Maximum likelihood expressions for the parameters of FvML distributions and organization of the section	89
4.4.2	Approximations for the concentration parameter from Mardia & Jupp (2000) and their limitations in high dimensions	89
4.4.3	New (high-dimensional) approximations for the concentration parameter	90
4.5	Further reading	94
5	Local asymptotic normality for directional data	97
5.1	Introduction	97
5.1.1	The LAN property on $\mathbb{R}^p$ and its deep impact on asymptotic statistics	97
5.1.2	Organization of the remainder of the chapter	98

5.2	Local asymptotic normality and optimal testing	98
5.2.1	Contiguity	98
5.2.2	Local asymptotic normality	99
5.2.3	Optimal testing in LAN experiments	102
5.2.4	LAN, semiparametric efficiency and invariance	105
5.3	LAN for directional data	109
5.3.1	The Le Cam methodology for curved experiments and associated efficient tests	110
5.3.2	LAN property for rotationally symmetric distributions	111
5.3.3	Application 1: Optimal inference based on signed-ranks	113
5.3.4	Application 2: ANOVA on spheres	115
5.3.5	Application 3: Asymptotic power of tests of concentration	118
5.4	Further reading	120
6	Recent results for tests of uniformity and symmetry	121
6.1	Introduction	121
6.1.1	Organization of the remainder of the chapter	122
6.2	Recent advances concerning the Rayleigh test of uniformity	123
6.3	Sobolev tests of uniformity	124
6.4	Uniformity tests based on random projections	126
6.5	Testing for uniformity with noisy data	127
6.6	Tests of reflective symmetry on the circle	128
6.7	Tests of rotational symmetry on hyperspheres	129
6.8	Testing for spherical location in the vicinity of the uniform distribution	131
6.9	Further reading	133
7	High-dimensional directional statistics	135
7.1	Introduction	135
7.1.1	High-dimensional techniques in $\mathbb{R}^p$	135
7.1.2	Organization of the remainder of the chapter	136
7.2	Distributions on high-dimensional spheres	137
7.3	Testing uniformity in the high-dimensional case	138
7.4	Location tests in the high-dimensional case	142
7.5	Concentration tests in the high-dimensional case	143
7.6	Principal nested spheres	146

<i>Contents</i>	xi
7.7 Further reading	148
Bibliography	149
Index	173

