WILEY SERIES IN PROBABILITY AND STATISTICS

Robustness Theory and Application

Brenton R. Clarke

WILEY

ROBUSTNESS THEORY AND APPLICATION

WILEY SERIES IN PROBABILITY AND STATISTICS
Established by *Walter A. Shewhart and Samuel S. Wilks*

The ***Wiley Series in Probability and Statistics*** is well established and authoritative. It covers many topics of current research interest in both pure and applied statistics and probability theory. Written by leading statisticians and institutions, the titles span both state-of-the-art developments in the field and classical methods.

Reflecting the wide range of current research in statistics, the series encompasses applied, methodological and theoretical statistics, ranging from applications and new techniques made possible by advances in computerized practice to rigorous treatment of theoretical approaches.

This series provides essential and invaluable reading for all statisticians, whether in academia, industry, government, or research.

A complete list of titles in this series can be found at
http://www.wiley.com/go/wsps

ROBUSTNESS THEORY AND APPLICATION

BRENTON R. CLARKE
Murdoch University

WILEY

This edition first published 2018

Registered Offices
John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, USA

Editorial Office
111 River Street, Hoboken, NJ 07030, USA

For details of our global editorial offices, customer services, and more information about Wiley products visit us at www.wiley.com.

Wiley also publishes its books in a variety of electronic formats and by print-on-demand. Some content that appears in standard print versions of this book may not be available in other formats.

Library of Congress Cataloging-in-Publication Data

Names: Clarke, Brenton R., author
Title: Robustness theory and application / by Brenton R. Clarke.
Description: Hoboken, NJ : John Wiley & Sons, 2018. | Series: Wiley series in probability and statistics | Includes bibliographical references and index. |
Identifiers: LCCN 2018007658 (print) | LCCN 2018010629 (ebook) | ISBN 9781118669501 (pdf) | ISBN 9781118669372 (epub) | ISBN 9781118669303 (cloth)
Subjects: LCSH: Robust statistics.
Classification: LCC QA276.A2 (ebook) | LCC QA276.A2 C53 2018 (print) | DDC 519.5–dc23
LC record available at https://lccn.loc.gov/2018007658

Cover Design: Wiley
Cover Image: © lvcandy/GettyImages

Set in 10/12pt Times by SPi Global, Pondicherry, India

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

To my darling wife Erica and much loved sons Andrew and Stephen

CONTENTS

FOREWORD

It could be said that the genesis of this book came out of a unit which was on robust statistics and taught by Noel Cressie in 1976 at Flinders University of South Australia. Noel's materials for the lectures were gathered from Princeton University where he had just completed his PhD. Having been introduced to M-, L-, and R-estimators I shifted to the Australian National University in 1977 to work on the staff at the Statistics Department in the Faculties. There I enrolled part time in a PhD with Professor C. R. Heathcote (affectionately known as Chip by his colleagues and family) who happened to be researching the integrated squared error method of estimation and more generally the method of minimum distance. The common link between the areas of study, robust statistics and minimum distance estimation, was that of M- estimation. Some minimum distance estimation methods can be represented by M-estimators. A typical model used in the formulation of robustness studies was the "epsilon-contaminated normal distribution." In the spirit of John W. Tukey from Princeton University the relative performance of the estimator, usually of location, was to consider it in such contaminated models. It occurred to me that one could also estimate the proportion of contamination, epsilon, in such models and when I proposed this to Chip he became enthusiastic that I should work on these mixture models for estimation in my PhD. Chip was aware that the trend for PhDs was to have a motivating set of data and to this end he introduced me to recently acquired earthquake data recordings which could be modeled with mixture modeling. A portion of a large data set was passed on to me by Professor R.S. Anderssen (known as Bob), also at the Australian National University. Bob also introduced me to the Fortran Computing Language.

My brief was to compute minimum distance estimators on the earthquake data. In the mean time, Chip introduced me to Professor Frank Hampel's PhD thesis and several references on mixture modeling. After 1 year of trying to compute variance covariance matrices for the minimum distance estimation methods and for some reason failing to get positive definite matrices as was expected, I decided to come back to M-estimation and study the theory more closely. An idea germinated that I could study the M-estimator at a distribution other than at the model parametric family and other than at a symmetric contaminating distributions. This became the inspiration for my own PhD work.

I had the good fortune to then cross paths with Peter Hall. Chip who had been burdened with the duties as Dean of the Faculty of Economics at ANU took sabbatical at Berkeley for a year and Peter became my supervisor. Peter was always so cheerful and encouraging when it came to research. He was publishing a book on the "Martingale limit theory and its application" with Chris Heyde, and he encouraged me to read books and papers on limit theorems. I thus became interested in the calculus associated with limit theorems, and asymptotic theory of M-estimators. Chip returned to ANU in 1980 and kindly advised me on the presentation of my thesis and arranged for three quality referees, one of whom was Noel Cressie!

For some reason I wanted to go overseas and see the world. This was made possible with a postdoctoral research assistant position to study time series analysis at Royal Holloway College, University of London, in the period 1980–1982. While I worked on time series, I took my free time to put together my first major publication. Huber's (1981) monograph had come out. My paper was to illustrate that for a large class of statistics that could be represented by statistical functionals, which were in fact M-estimators, it was possible to inherit both weak continuity and Fréchet differentiability. These qualities in turn provide inherent robustness of the statistics. From the time of first submission to actual publication in *The Annals of Statistics* it took approximately 2.5 years to see it come out. It was during this time of waiting that I traveled to Zürich after writing to Professor Hampel. He was keen to see my work published as it supported with rigor notions which he had put forward in a heuristic manner, vis-à-vis the influence function. Subsequently, I spent almost a year at the Swiss Federal Institute of Technology (ETH), working as a research assistant and tutoring a class on the analysis of variance class lectured by Professor Hampel.

The Conditions A and discussion that are given in Chapter 2 of this book are from that Annals of Statistics paper. To facilitate the theory of weak continuity and Fréchet differentiability, I initially had to make smoothness assumptions on the defining ψ-functions for the M-estimators. It was not until I traveled to the University of North Carolina at Chapel Hill where I picked up the newly published book by Frank H. Clarke on "Optimization and Nons- mooth Analysis" that I realized how proofs of weak continuity and Fréchet differentiability for M-estimators with nonsmooth psi-functions, or psi-functions which were bounded and continuous but had "sharp corners," could follow through. I subsequently wrote a paper from

Murdoch University where I had taken up in 1984 a newly appointed lecturing position in the then Mathematics Department. The paper was eventually published in 1986 in *Probability Theory and Related Fields*. This book brings together both these papers and a paper on what are called selection functionals.

My sojourn at Murdoch University has been one of teaching and research. I benefited from many years of teaching in service course and undergraduate mathematics and statistics units, having developed materials for a unit on "Linear Models" which later became a Wiley publication in 2008. I have also developed a unit on "Time Series Analysis" and have two PhD students write theses in that general area. These forays while time consuming have helped me understand statistics a lot better. It has to be said that to teach robust statistics properly one needs to understand the mathematics that comes with it. Essentially, my experience in robust statistics has been one coming out of mathematics departments or at least statistics groups heavily supported by mathematics departments. But from the mathematics comes understanding and eventually new ideas on how to analyze data and further appreciation of why some things work and others do not. This book is a reflection of that general summary.

In writing this book I have also alluded to or summarized many works that have been collaborative. An author with whom I have published much is Pro- fessor Tadeusz Bednarski from Poland. I met Professor Bednarski at a Robust Statistics Meeting in Oberwolfach, Germany, in 1984. He recognized the importance of Fréchet differentiability and in particular the two works mentioned earlier and we proceeded to make a number of joint works on the asymptotics of estimators. He spoke on Fréchet differentiability at the Australian Statistics Conference 2010 held in Fremantle in Western Australia. However, with the tyranny of distance and our paths diverging since then it is clear that this book could not be written collaboratively. However, I owe much to the joint research that we did as is acknowledged in the book.

I have also benefited from collaborative works with many other authors. These works have helped in the presentation of new material in the book. In 1993 I published a paper with Robin Milne and Geoffrey Yeo in the *Scandinavian Journal of Statistics*. I thank both Robin and Geoff for making me think about the asymptotic theory when there are multiple solutions to ones estimat- ing equations. There are subsequently new examples and results on asymptotic properties of roots and tests in Chapter 4 of this book that have been developed by the author. In 2004 the author published a paper with former honors student Chris Milne in the Australian & New Zealand Journal of Statistics on a small sample bias correction to Huber's Proposal 2 estimator of location and scale and followed this with a paper in 2013 at the ISI meeting in Hong Kong. Summary results are included with permission in Section 5.1.2. In 2006 I collaborated with Andreas Futschik from the University of Vienna, to study the properties of the Newton Algorithm when dealing with either M-estimation or density estimation and a new Theorem 5.1 is borne out of that work. My interest in minimum distance estimation and its applications are summarized in Chapter 6. These include references to work with Chip Heathcote and also other collaborators such as Peter McKinnon and Geoff Riley. A new theorem on the unbiased nature of the vector parameter estimator of proportions given all other location and scale parameters in a mixture of normal distributions are known is given in Theorem 10.1. In addition plots in Figures 2.1, 6.5, 6.6, 6.7, 6.8, and 6.9 are reproduced with acknowledgment from their source.

No book on robustness is complete without the study of L-estimators or estimation of linear combinations of order statistics. I have only attempted to introduce the ideas which lead on to natural extensions on to least trimmed squares and generalizations to trimmed likelihood and adaptive trimmed likeli- hood algorithms. I have found these useful for identifying outliers where there are outliers to be found, yet caution the reader to use Fréchet differentiable es- timators for robust statistical inference. The outlier detection methods depart from the general use of Cooks distance in regression estimation yet have the appealing feature that they work even when there are what are termed points of influence.

The book does not canvas robust methods in time series or robust survival analysis, though references are given. Maronna et al. (2006) book is a useful starting point for robust time series. Developments on robust survival analysis continue to accrue. The presentation of this book is not exhaustive and many areas of endeavor in robust statistics are not countenanced in this book. The book mainly is a link between many areas of research that the author has been personally involved with in some way and attempts to weave the essence of relevant theory and application.

The work would never have been possible without the introduction of Fréchet differentiability into the statistical literature by Professor C. R. Rao and Professor G. Kallianpur in their ground-breaking paper in 1955. We have much to remember the French mathematician Maurice Renáe Fréchet for as well as Sir Ronald Aylmer Fisher who helped to motivate the 1955 paper.

PREFACE

This book requires a strong mathematics background as a prerequisite in or- der to be read in its entirety. Students and researchers will then appreciate the generally easily read Chapter 1 in the Introduction. Chapters 2 and 3 require a mathematics background, but it is possible to avoid the difficult proofs requiring the knowledge of the inverse function theorem in the proofs of weak continuity and Frèchet differentiability of M-functionals, should you need to gloss over the mathematics. On the other hand, great understanding can be gleaned from paying attention to such proofs. There are references later in the book to other important theorems such as the fixed point theorem and the implicit function theorem, though these are only referred to, and keen students may chase up their statements and proofs in related papers and mathematics texts. In this book Chapter 4 is important to the understanding that there can be more than one root to one's estimating equations and gives new results in this direction. Chapters 5–9 include applications and vary from the simple applications of computing robust estimators to the asymptotic theory descriptions which are a composite of exact calculation, such as in the theory of L_2 estimation of proportions, or descriptions of asymptotic normality results that can be further gleaned by studying the research papers cited. The attempt is to bring together works on estimation theory in robust estimation. I leave it to others to consider the potentially more difficult theory of testing, albeit robust confidence intervals based on asymptotics are a starting point for such.

This book has been written in what may be the last decade of my working career. Hopefully, others may benefit from the insights that this compendium of knowledge, which covers much research into robustness that I have had a part to play with, gives.

7 December 2017

BRENTON R. CLARKE
Perth, Western Australia

ACKNOWLEDGMENTS

I wish to acknowledge my two PhD supervisors Chip Heathcote and Professor Peter G. Hall. Both passed away in 2016. I remember them for their generous guidance in motivating my PhD studies in the period 1977–1980. Also I have to thank Professor Frank R. Hampel and his colleagues and students for helping me on my way during postdoctoral training as a Wissenschaftliche Mitarbeiter at ETH in 1982–1983. Their influence is unmistakeable. I owe much to the late Professor Alex Roberston for his help in bringing me to the then Mathematics Department, now called Mathematics and Statistics at Murdoch University, and I thank all my mathematics and statistics colleagues past and present for their generosity in allowing me to teach and research while at Murdoch University.

To my collaborators mentioned in the Foreword I give my thanks. Special thanks are to Professor Tadeusz Bednarski for fostering international collaboration in mathematics and statistics pre and post Communism (in Poland) and showing that there are no international borders religious or political in the common language of mathematics. I also thank Professor Andrzej Kozek who first hosted me at the University of Wroclaw in Poland and introduced me to Professor Bednarski.

Other researchers with whom I have the pleasure of being able to work with include Thomas Davidson, Andreas Futschik, David Gamble, Robert Hammarstrand, Toby Lewis, Peter McKinnon, Chris Milne, Robin Milne, Geoffrey Yeo, and Geoff Riley just to name some. More recent collaborations are with Christine Mueller and students. Robert Hammarstrand has also contributed by working under my direction to polish some of the R-algorithms associated with this book, for which I take full responsibility.

I have to acknowledge the work with Daniel Schubert. He wrote and gained his PhD under my supervision on the area of trimmed likelihood, but after gaining a position in CSIRO had his life cut short in a motor bike accident in 2007. I remember his eccentricities and for his enthusiasm for his newly found passion of robust statistics when he was a student. I include in the suite of associated R-algorithms our contribution to the Adaptive Trimmed Likelihood Algorithm for multivariate data.

As I came nearer the publication due date in July 2016, I had the privilege of visiting the Department of Statistics at The University of Glasgow headed by Professor Adrian Bowman. In August 2016 I visited Professor Mette Langaas and colleagues in the Statistics Group in the Mathematics Department at the Norwegian University of Science and Technology (NTNU). Some of this book was inspired by these visits. The journey was also facilitated by a visit to the University of Surrey, where I was hosted by Dr. Janet Godolphin in the Department of Mathematics. Finally, I acknowledge Murdoch University for the sabbatical that was taken nominally at the University of Western Australia, for the remainder of second semester 2016 used in preparation of this book. I thank Berwin Turlach at the University of Western Australia for his administrative role in arranging this. In addition I would like to thank Professor Luke Prendergast for encouragement and comments on a penultimate version of the book. Also thanks to Professors Alan Welsh and Andrew Wood for encouragement.

It goes without saying that I owe much to my wife and children in the formulation of this book and its predecessor *Linear Models: The Theory and Application of Analysis of Variance* which was published in 2008. They are duly acknowledged. Subsequently, I dedicate both books to them.

BRENTON R. CLARKE

NOTATION

$\in$	element of
$\notin$	not an element of
$A \cap B$	intersection of sets A and B
$A \cup B$	union of sets A and or B
$A \subset B$	A is contained in B
$\tilde{R}$	observation space
R	separable metrizeable space
$\mathcal{E}$	real line
$\mathcal{E}^+$	positive real line
$\mathcal{E}^r$	Euclidean r-space
$\mathcal{G}$	space of distribution functions
$\mathcal{A}^\delta$	closed delta neighborhood of set A
Θ	parameter space
$E[X]$	expected value of the random variable X
$\phi(z)$	standard normal density function
$\Phi(z)$	cumulative normal distribution
$\mathcal{I}(\theta)$	Fisher information
$\lvert\cdot\rvert$	modulus or absolute value of
$\rightarrow_p$	convergence in probability
$\rightarrow_d$	convergence in distribution
$\rightarrow_{a.s.}$	convergence almost surely
$\Rightarrow$	converges weakly
$\Longrightarrow$	implies

$\forall$	for every
$\lVert\cdot\rVert$	for vectors and matrices it is the Euclidean norm
	for elements on a normed linear space it is the norm
$N(\mu, \sigma^2)$	univariate normal random variable with mean μ and variance σ^2
$N(\mu, \Sigma)$	multivariate normal random variable
	with mean μ and covariance matrix Σ
$\Sigma_{i=1}^{n} x_i$	$x_1 + x_2 + \dots + x_n$

ACRONYMS

ATLA	adaptive trimmed likelihood algorithm
CLT	central limit theorem
EM	expectation maximization
LLN	law of large numbers
MAD	median absolute deviation
MADN	normalized median absolute deviation
MCVM	minimum Cramèr–von Mises estimator
MLE	maximum likelihood estimator
SLLN	strong law of large numbers
TLE	trimmed likelihood estimator
WLLN	weak law of large numbers

ABOUT THE COMPANION WEBSITE

This book is accompanied by a companion website:

www.wiley.com/go/clarke/robustnesstheoryandapplication

BCS Ancillary Website contains:

- several computing programs, some in R and some in MatLab.
- The suite of routines were developed by Brenton and in some cases in conjunction with other students and collaborators of Brenton.
- No responsibility lies with Brenton or any others mentioned for the use of the routines given in endeavours outside the book.

BRENTON R. CLARKE
November 2017