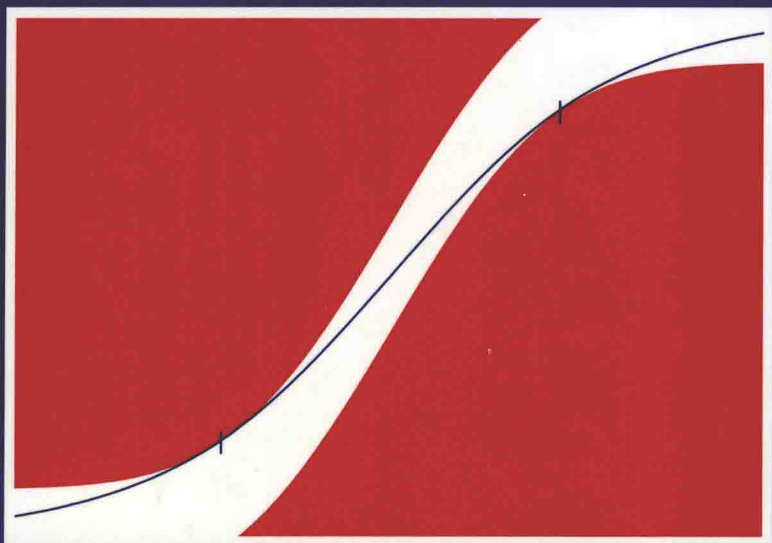
WILEY SERIES IN PROBABILITY AND STATISTICS

Robust Statistics

Second Edition



Peter J. Huber
Elvezio M. Ronchetti

WILEY

ROBUST STATISTICS

Second Edition

Peter J. Huber
Professor of Statistics, retired
Klosters, Switzerland

Elvezio M. Ronchetti
Professor of Statistics
University of Geneva, Switzerland

常州大学图书馆
藏书章

WILEY

A JOHN WILEY & SONS, INC., PUBLICATION

Copyright © 2009 by John Wiley & Sons, Inc. All rights reserved.

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.
Published simultaneously in Canada.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4470, or on the web at www.copyright.com. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at http://www.wiley.com/go/permission.

Limit of Liability/Disclaimer of Warranty: While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Neither the publisher nor author shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

For general information on our other products and services or for technical support, please contact our Customer Care Department within the United States at (800) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic format. For information about Wiley products, visit our web site at www.wiley.com.

Library of Congress Cataloging-in-Publication Data:

Huber, Peter J.
Robust statistics, second edition / Peter J. Huber, Elvezio Ronchetti.
p. cm.
Includes bibliographical references and index.
ISBN 978-0-470-12990-6 (cloth)
1. Robust statistics. I. Ronchetti, Elvezio. II. Title.
QA276.H785 2009
519.5—dc22 2008033283

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

To the memory of
John W. Tukey

PREFACE

When Wiley asked me to undertake a revision of *Robust Statistics* for a second edition, I was at first very reluctant to do so. My own interests had begun to shift toward data analysis in the late 1970s, and I had ceased to work in robustness shortly after the publication of the first edition. Not only was I now out of contact with the forefront of current work, but I also disagreed with some of the directions that the latter had taken and was not overly keen to enter into polemics. Back in the 1960s, robustness theory had been created to correct the instability problems of the "optimal" procedures of classical mathematical statistics. At that time, in order to make robustness acceptable within the paradigms then prevalent in statistics, it had been indispensable to create optimally robust (i.e., minimax) alternative procedures. Ironically, by the 1980s, "optimal" robustness began to run into analogous instability problems. In particular, while a high breakdown point clearly is desirable, the (still) fashionable strife for the highest possible breakdown point in my opinion is misguided: it is not only overly pessimistic, but, even worse, it disregards the central stability aspect of robustness.

But an update clearly was necessary. After the closure date of the first edition, there had been important developments not only with regard to the breakdown point, on which I have added a chapter, but also in the areas of infinitesimal robustness, robust tests, and small sample asymptotics. In many places, it would suffice to

此为试读，需要完整PDF请访问：www.ertongbook.com

update bibliographical references, so the manuscript of the second edition could be based on a re-keyed version of the first. Other aspects deserved a more extended discussion. I was fortunate to persuade Elvezio Ronchetti, who had been one of the prime researchers working in the two last mentioned areas (robust tests and small sample asymptotics), to collaborate and add the corresponding Chapters 13 and 14. Also, I extended the discussion of regression, and I decided to add a chapter on Bayesian robustness—even though, or perhaps because, I am not a Bayesian (or only rarely so). Among other minor changes, since most readers of the first edition had appreciated the *General Remarks* at the beginning of the chapters, I have expanded some of them and also elsewhere devoted more space to an informal discussion of motivations.

The new edition still has no pretensions of being encyclopedic. Like the first, it is centered on a robustness concept based on minimax asymptotic variance and on M-estimation, complemented by some exact finite sample results. Much of the material of the first edition is just as valid as it was in 1980. Deliberately, such parts were left intact, except that bibliographical references had to be added. Also, I hope that my own perspective has improved with an increased temporal and professional distance. Although this improved perspective has not affected the mathematical facts, it has sometimes sharpened their interpretation.

Special thanks go to Amy Hendrickson for her patient help with the Wiley LaTeX-macros and the various quirks of TeX.

PETER J. HUBER

Klosters
November 2008

PREFACE TO THE FIRST EDITION

The present monograph is the first systematic, book-length exposition of robust statistics. The technical term "robust" was coined only in 1953 (by G. E. P. Box), and the subject matter acquired recognition as a legitimate topic for investigation only in the mid-sixties, but it certainly never was a revolutionary new concept. Among the leading scientists of the late nineteenth and early twentieth century, there were several practicing statisticians (to name but a few: the astronomer S. Newcomb, the astrophysicist A.S. Eddington, and the geophysicist H. Jeffreys), who had a perfectly clear, operational understanding of the idea; they knew the dangers of long-tailed error distributions, they proposed probability models for gross errors, and they even invented excellent robust alternatives to the standard estimates, which were rediscovered only recently. But for a long time theoretical statisticians tended to shun the subject as being inexact and "dirty." My 1964 paper may have helped to dispel such prejudices. Amusingly (and disturbingly), it seems that lately a kind of bandwagon effect has evolved, that the pendulum has swung to the other extreme, and that "robust" has now become a magic word, which is invoked in order to add respectability.

This book gives a solid foundation in robustness to both the theoretical and the applied statistician. The treatment is theoretical, but the stress is on concepts, rather than on mathematical completeness. The level of presentation is deliberately uneven: in some chapters simple cases are treated with mathematical rigor; in others the results obtained in the simple cases are transferred by analogy to more complicated situations (like multiparameter regression and covariance matrix estimation), where proofs are not always available (or are available only under unrealistically severe assumptions). Also selected numerical algorithms for computing robust estimates are described and, where possible, convergence proofs are given.

Chapter 1 gives a general introduction and overview; it is a must for every reader. Chapter 2 contains an account of the formal mathematical background behind qualitative and quantitative robustness, which can be skipped (or skimmed) if the reader is willing to accept certain results on faith. Chapter 3 introduces and discusses the three basic types of estimates (M-, L-, and R-estimates), and Chapter 4 treats the asymptotic minimax theory for location estimates; both chapters again are musts. The remaining chapters branch out in different directions and are fairly independent and self-contained; they can be read or taught in more or less any order.

The book does not contain exercises—I found it hard to invent a sufficient number of problems in this area that were neither trivial nor too hard—so it does not satisfy some of the formal criteria for a textbook. Nevertheless I have successfully used various stages of the manuscript as such in graduate courses.

The book also has no pretensions of being encyclopedic. I wanted to cover only those aspects and tools that I personally considered to be the most important ones. Some omissions and gaps are simply due to the fact that I currently lack time to fill them in, but do not want to procrastinate any longer (the first draft for this book goes back to 1972). Others are intentional. For instance, adaptive estimates were excluded because I would now prefer to classify them with nonparametric rather than with robust statistics, under the heading of nonparametric efficient estimation. The so-called Bayesian approach to robustness confounds the subject with admissible estimation in an *ad hoc* parametric supermodel, and still lacks reliable guidelines on how to select the supermodel and the prior so that we end up with something robust. The coverage of L- and R-estimates was cut back from earlier plans because they do not generalize well and get awkward to compute and to handle in multiparameter situations.

A large part of the final draft was written when I was visiting Harvard University in the fall of 1977; my thanks go to the students, in particular to P. Rosenbaum and Y. Yoshizoe, who then sat in my seminar course and provided many helpful comments.

PETER J. HUBER

Cambridge, Massachusetts
July 1980

CONTENTS

此为试读，需要完整PDF请访问：www.ertongbook.com

CHAPTER 1

GENERALITIES

1.1 WHY ROBUST PROCEDURES?

Statistical inferences are based only in part upon the observations. An equally important base is formed by prior assumptions about the underlying situation. Even in the simplest cases, there are explicit or implicit assumptions about randomness and independence, about distributional models, perhaps prior distributions for some unknown parameters, and so on.

These assumptions are not supposed to be exactly true—they are mathematically convenient rationalizations of an often fuzzy knowledge or belief. As in every other branch of applied mathematics, such rationalizations or simplifications are vital, and one justifies their use by appealing to a vague continuity or stability principle: a minor error in the mathematical model should cause only a small error in the final conclusions.

Unfortunately, this does not always hold. Since the middle of the 20th century, one has become increasingly aware that some of the most common statistical procedures (in particular, those optimized for an underlying normal distribution) are excessively

Copyright © 2009 John Wiley & Sons, Inc.

sensitive to seemingly minor deviations from the assumptions, and a plethora of alternative "robust" procedures have been proposed.

The word "robust" is loaded with many—sometimes inconsistent—connotations. We use it in a relatively narrow sense: for our purposes, *robustness signifies insensitivity to small deviations from the assumptions.*

Primarily, we are concerned with *distributional robustness*: the shape of the true underlying distribution deviates slightly from the assumed model (usually the Gaussian law). This is both the most important case and the best understood one. Much less is known about what happens when the other standard assumptions of statistics are not quite satisfied and about the appropriate safeguards in these other cases.

The following example, due to Tukey (1960), shows the dramatic lack of distributional robustness of some of the classical procedures.

■ EXAMPLE 1.1

Assume that we have a large, randomly mixed batch of n "good" and "bad" observations x_i of the same quantity μ. Each single observation with probability $1-\varepsilon$ is a "good" one, with probability ε a "bad" one, where ε is a small number. In the former case x_i is $\mathcal{N}(\mu, \sigma^2)$, in the latter $\mathcal{N}(\mu, 9\sigma^2)$. In other words all observations are normally distributed with the same mean, but the errors of some are increased by a factor of 3.

Equivalently, we could say that the x_i are independent, identically distributed with the common underlying distribution

$$F(x) = (1-\varepsilon)\Phi\left(\frac{x-\mu}{\sigma}\right) + \varepsilon\Phi\left(\frac{x-\mu}{3\sigma}\right), \tag{1.1}$$

where

$$\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-y^2/2}\,dy \tag{1.2}$$

is the standard normal cumulative.

Two time-honored measures of scatter are the mean absolute deviation

$$d_n = \frac{1}{n}\sum |x_i - \bar{x}| \tag{1.3}$$

and the root mean square deviation

$$s_n = \left[\frac{1}{n}\sum (x_i - \bar{x})^2\right]^{1/2}. \tag{1.4}$$

There was a dispute between Eddington (1914, p. 147) and Fisher (1920, footnote on p. 762) about the relative merits of d_n and s_n. Eddington had advocated the use of the former: "This is contrary to the advice of most textbooks; but it can be shown to be true." Fisher seemingly settled the matter by pointing out that for identically distributed normal observations s_n is about 12% more efficient than d_n.

Of course, the two statistics measure different characteristics of the error distribution. For instance, if the errors are exactly normal, s_n converges to σ, while d_n converges to $\sqrt{2/\pi}\,\sigma \cong 0.80\sigma$. So we must be precise about how their performances are to be compared; we use the asymptotic relative efficiency (ARE) of d_n relative to s_n, defined as follows:

$$\text{ARE}(\varepsilon) = \lim_{n\to\infty} \frac{\text{var}(s_n)/(Es_n)^2}{\text{var}(d_n)/(Ed_n)^2} = \frac{\dfrac{1}{4}\left[\dfrac{3(1+80\varepsilon)}{(1+8\varepsilon)^2} - 1\right]}{\dfrac{\pi(1+8\varepsilon)}{2(1+2\varepsilon)^2} - 1}. \tag{1.5}$$

The results are summarized in Exhibit 1.1.

ε	ARE(ε)
0	0.876
0.001	0.948
0.002	1.016
0.005	1.198
0.01	1.439
0.02	1.752
0.05	2.035
0.10	1.903
0.15	1.689
0.25	1.371
0.5	1.017
1.0	0.876

Exhibit 1.1 Asymptotic efficiency of mean absolute deviation relative to root mean square deviation. From Huber (1977b), with permission of the publisher.

The result is disquieting: just 2 bad observations in 1000 suffice to offset the 12% advantage of the mean square error, and ARE(ε) reaches a maximum value greater than 2 at about $\varepsilon = 0.05$. This is particularly unfortunate since in the physical sciences typical "good data" samples appear to be well modeled by an error law of the form (1.1) with ε in the range between 0.01 and 0.1. (This does not imply that these samples contain between 1% and 10% gross errors, although this is very often true; the above law (1.1) may just be a convenient description of a slightly longer-tailed than normal distribution.) Thus it becomes painfully clear that the naturally occurring deviations from the idealized model are large enough to render meaningless the traditional asymptotic optimality theory: in practice, we should certainly prefer d_n to s_n, since it is better for all ε between 0.002 and 0.5.

此为试读，需要完整PDF请访问：www.ertongbook.com