ROBUST STATISTICS

Peter J. Huber

A volume in the Wiley Series in Probability and Mathematical Statistics:
Ralph A. Bradley, J. Stuart Hunter, David G. Kendall, and Geoffrey S. Watson—Advisory Editors

Robust Statistics

PETER J. HUBER

Professor of Statistics
Harvard University
Cambridge, Massachusetts

John Wiley & Sons

New York • Chichester • Brisbane • Toronto

Library of Congress Cataloging in Publication Data:

Huber, Peter J
Robust statistics.

(Wiley series in probability and mathematical statistics)
"A Wiley-Interscience publication."
Includes index.
1. Robust statistics. I. Title.

QA276.H785 519.5 80-18627
ISBN 0-471-41805-6

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

WILEY SERIES IN PROBABILITY AND MATHEMATICAL STATISTICS

Probability and Mathematical Statistics

ADLER • The Geometry of Random Fields
ANDERSON • The Statistical Analysis of Time Series
ANDERSON • An Introduction to Multivariate Statistical Analysis
ARAUJO and GINE • The Central Limit Theorem for Real and Banach Valued Random Variables
ARNOLD • The Theory of Linear Models and Multivariate Analysis
BARLOW, BARTHOLOMEW, BREMNER, and BRUNK • Statistical Inference Under Order Restrictions
BARNETT • Comparative Statistical Inference
BHATTACHARYYA and JOHNSON • Statistical Concepts and Methods
BILLINGSLEY • Probability and Measure
CASSEL, SARNDAL, and WRETMAN • Foundations of Inference in Survey Sampling
DE FINETTI • Theory of Probability, Volumes I and II
DOOB • Stochastic Processes
FELLER • An Introduction to Probability Theory and Its Applications, Volume I, *Third Edition,* Revised
FELLER • An Introduction to Probability Theory and Its Applications, Volume II, *Second Edition*
FULLER • Introduction to Statistical Time Series
GRENANDER • Abstract Inference
HANNAN • Multiple Time Series
HANSEN, HURWITZ, and MADOW • Sample Survey Methods and Theory, Volumes I and II
HARDING and KENDALL • Stochastic Geometry
HOEL • Introduction to Mathematical Statistics, *Fourth Edition*
HUBER • Robust Statistics
IOSIFESCU • Finite Markov Processes and Applications
ISAACSON and MADSEN • Markov Chains
KAGAN, LINNIK, and RAO • Characterization Problems in Mathematical Statistics
KENDALL and HARDING • Stochastic Analysis
LAHA and ROHATGI • Probability Theory
LARSON • Introduction to Probability Theory and Statistical Inference, *Second Edition*
LARSON • Introduction to the Theory of Statistics
LEHMANN • Testing Statistical Hypotheses
MATHERON • Random Sets and Integral Geometry
MATTHES, KERSTAN, and MECKE • Infinitely Divisible Point Processes
PARZEN • Modern Probability Theory and Its Applications
PURI and SEN • Nonparametric Methods in Multivariate Analysis
RANDLES and WOLFE • Introduction to the Theory of Nonparametric Statistics
RAO • Linear Statistical Inference and Its Applications, *Second Edition*
ROHATGI • An Introduction to Probability Theory and Mathematical Statistics
SCHEFFE • The Analysis of Variance
SEBER • Linear Regression Analysis
SERFLING • Approximation Theorems of Mathematical Statistics
TJUR • Probability Based on Radon Measures

Probability and Mathematical Statistics (Continued)

WILLIAMS • Diffusions, Markov Processes, and Martingales, Volume I: Foundations
ZACKS • Theory of Statistical Inference

Applied Probability and Statistics

ANDERSON, AUQUIER, HAUCK, OAKES, VANDAELE, and WEISBERG • Statistical Methods for Comparative Studies
ARTHANARI and DODGE • Mathematical Programming in Statistics
BAILEY • The Elements of Stochastic Processes with Applications to the Natural Sciences
BAILEY • Mathematics, Statistics and Systems for Health
BARNETT and LEWIS • Outliers in Statistical Data
BARTHOLOMEW • Stochastic Models for Social Processes, *Second Edition*
BARTHOLOMEW and FORBES • Statistical Techniques for Manpower Planning
BECK and ARNOLD • Parameter Estimation in Engineering and Science
BELSLEY, KUH, and WELSCH • Regression Diagnostics: Identifying Influential Data and Sources of Collinearity
BENNETT and FRANKLIN • Statistical Analysis in Chemistry and the Chemical Industry
BHAT • Elements of Applied Stochastic Processes
BLOOMFIELD • Fourier Analysis of Time Series: An Introduction
BOX • R. A. Fisher, The Life of a Scientist
BOX and DRAPER • Evolutionary Operation: A Statistical Method for Process Improvement
BOX, HUNTER, and HUNTER • Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building
BROWN and HOLLANDER • Statistics: A Biomedical Introduction
BROWNLEE • Statistical Theory and Methodology in Science and Engineering, *Second Edition*
BURY • Statistical Models in Applied Science
CHAMBERS • Computational Methods for Data Analysis
CHATTERJEE and PRICE • Regression Analysis by Example
CHERNOFF and MOSES • Elementary Decision Theory
CHOW • Analysis and Control of Dynamic Economic Systems
CLELLAND, BROWN, and deCANI • Basic Statistics with Business Applications, *Second Edition*
COCHRAN • Sampling Techniques, *Third Edition*
COCHRAN and COX • Experimental Designs, *Second Edition*
CONOVER • Practical Nonparametric Statistics, *Second Edition*
CORNELL • Experiments with Mixtures: Designs, Models and The Analysis of Mixture Data
COX • Planning of Experiments
DANIEL • Biostatistics: A Foundation for Analysis in the Health Sciences, *Second Edition*
DANIEL • Applications of Statistics to Industrial Experimentation
DANIEL and WOOD • Fitting Equations to Data: Computer Analysis of Multifactor Data, *Second Edition*
DAVID • Order Statistics, *Second Edition*
DEMING • Sample Design in Business Research
DODGE and ROMIG • Sampling Inspection Tables, *Second Edition*
DRAPER and SMITH • Applied Regression Analysis, *Second Edition*
DUNN • Basic Statistics: A Primer for the Biomedical Sciences, *Second Edition*
DUNN and CLARK • Applied Statistics: Analysis of Variance and Regression
ELANDT-JOHNSON • Probability Models and Statistical Methods in Genetics
ELANDT-JOHNSON and JOHNSON • Survival Models and Data Analysis

continued on back

Applied Probability and Statistics (Continued)

FLEISS • Statistical Methods for Rates and Proportions, *Second Edition*
GALAMBOS • The Asymptotic Theory of Extreme Order Statistics
GIBBONS, OLKIN, and SOBEL • Selecting and Ordering Populations: A New Statistical Methodology
GNANADESIKAN • Methods for Statistical Data Analysis of Multivariate Observations
GOLDBERGER • Econometric Theory
GOLDSTEIN and DILLON • Discrete Discriminant Analysis
GROSS and CLARK • Survival Distributions: Reliability Applications in the Biomedical Sciences
GROSS and HARRIS • Fundamentals of Queueing Theory
GUPTA and PANCHAPAKESAN • Multiple Decision Procedures: Theory and Methodology of Selecting and Ranking Populations
GUTTMAN, WILKS, and HUNTER • Introductory Engineering Statistics, *Second Edition*
HAHN and SHAPIRO • Statistical Models in Engineering
HALD • Statistical Tables and Formulas
HALD • Statistical Theory with Engineering Applications
HARTIGAN • Clustering Algorithms
HILDEBRAND, LAING, and ROSENTHAL • Prediction Analysis of Cross Classifications
HOEL • Elementary Statistics, *Fourth Edition*
HOLLANDER and WOLFE • Nonparametric Statistical Methods
JAGERS • Branching Processes with Biological Applications
JESSEN • Statistical Survey Techniques
JOHNSON and KOTZ • Distributions in Statistics
Discrete Distributions
Continuous Univariate Distributions—1
Continuous Univariate Distributions—2
Continuous Multivariate Distributions
JOHNSON and KOTZ • Urn Models and Their Application: An Approach to Modern Discrete Probability Theory
JOHNSON and LEONE • Statistics and Experimental Design in Engineering and the Physical Sciences, Volumes I and II, *Second Edition*
JUDGE, GRIFFITHS, HILL and LEE • The Theory and Practice of Econometrics
KALBFLEISCH and PRENTICE • The Statistical Analysis of Failure Time Data
KEENEY and RAIFFA • Decisions with Multiple Objectives
LANCASTER • An Introduction to Medical Statistics
LEAMER • Specification Searches: Ad Hoc Inference with Nonexperimental Data
McNEIL • Interactive Data Analysis
MANN, SCHAFER and SINGPURWALLA • Methods for Statistical Analysis of Reliability and Life Data
MEYER • Data Analysis for Scientists and Engineers
MILLER, EFRON, BROWN, and MOSES • Biostatistics Casebook
OTNES and ENOCHSON • Applied Time Series Analysis: Volume I, Basic Techniques
OTNES and ENOCHSON • Digital Time Series Analysis
POLLOCK • The Algebra of Econometrics
PRENTER • Splines and Variational Methods
RAO and MITRA • Generalized Inverse of Matrices and Its Applications
RIPLEY • Spatial Statistics
SCHUSS • Theory and Applications of Stochastic Differential Equations
SEAL • Survival Probabilities: The Goal of Risk Theory
SEARLE • Linear Models
SPRINGER • The Algebra of Random Variables
UPTON • The Analysis of Cross-Tabulated Data
WEISBERG • Applied Linear Regression
WHITTLE • Optimization Under Constraints

Applied Probability and Statistics (Continued)

WILLIAMS • A Sampler on Sampling

WONNACOTT and WONNACOTT • Econometrics, *Second Edition*

WONNACOTT and WONNACOTT • Introductory Statistics, *Third Edition*

WONNACOTT and WONNACOTT • Introductory Statistics for Business and Economics, *Second Edition*

WONNACOTT and WONNACOTT • Regression: A Second Course in Statistics

ZELLNER • An Introduction to Bayesian Inference in Econometrics

Tracts on Probability and Statistics

BARNDORFF-NIELSEN • Information and Exponential Families in Statistical Theory

BHATTACHARYA and RAO • Normal Approximation and Asymptotic Expansions

BIBBY and TOUTENBERG • Prediction and Improved Estimation in Linear Models

BILLINGSLEY • Convergence of Probability Measures

JARDINE and SIBSON • Mathematical Taxonomy

KELLY • Reversibility and Stochastic Networks

KINGMAN • Regenerative Phenomena

Preface

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. 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. Yashizoe, who then sat in my seminar course and provided many helpful comments.

P. J. HUBER

Cambridge, Massachusetts
July 1980

Contents

1 GENERALITIES 1

1.1 Why Robust Procedures? 1
1.2 What Should a Robust Procedure Achieve? 5
1.3 Qualitative Robustness, 7
1.4 Quantitative Robustness, 10
1.5 Infinitesimal Aspects, 13
1.6 Optimal Robustness, 16
1.7 Computation of Robust Estimates, 17

2 THE WEAK TOPOLOGY AND ITS METRIZATION 20

2.1 General Remarks, 20
2.2 The Weak Topology, 20
2.3 Lévy and Prohorov Metrics, 25
2.4 The Bounded Lipschitz Metric, 29
2.5 Fréchet and Gâteaux Derivatives, 34
2.6 Hampel's Theorem, 40

3 THE BASIC TYPES OF ESTIMATES 43

3.1 General Remarks, 43
3.2 Maximum Likelihood Type Estimates (M-Estimates), 43
3.3 Linear Combinations of Order Statistics (L-Estimates), 55
3.4 Estimates Derived from Rank Tests (R-Estimates), 61
3.5 Asymptotically Efficient M-, L-, and R-Estimates, 68

4 ASYMPTOTIC MINIMAX THEORY FOR ESTIMATING A LOCATION PARAMETER 73

4.1 General Remarks, 73
4.2 Minimax Bias, 74
4.3 Minimax Variance: Preliminaries, 76
4.4 Distributions Minimizing Fisher Information, 77

4.5 Determination of F_0 by Variational Methods, 82
4.6 Asymptotically Minimax *M*-Estimates, 94
4.7 On the Minimax Property for *L*- and *R*-Estimates, 97
4.8 Descending *M*-Estimates, 100
4.9 Questions of Asymmetric Contamination, 104

5 SCALE ESTIMATES 107

5.1 General Remarks, 107
5.2 *M*-Estimates of Scale, 109
5.3 *L*-Estimates of Scale, 110
5.4 *R*-Estimates of Scale, 114
5.5 Asymptotically Efficient Scale Estimates, 116
5.6 Distributions Minimizing Fisher Information for Scale, 118
5.7 Minimax Properties, 122

6 MULTIPARAMETER PROBLEMS, IN PARTICULAR JOINT ESTIMATION OF LOCATION AND SCALE 127

6.1 General Remarks, 127
6.2 Consistency of *M*-Estimates, 127
6.3 Asymptotic Normality of *M*-Estimates, 132
6.4 Simultaneous *M*-Estimates of Location and Scale, 135
6.5 *M*-Estimates with Preliminary Estimates of Scale, 140
6.6 Quantitative Robustness Properties of Simultaneous Estimates for Location and Scale, 141
6.7 The Computation of *M*-Estimates, 146
6.8 Studentizing, 148

7 REGRESSION 153

7.1 General Remarks, 153
7.2 The Classical Linear Least Squares Case, 155
7.3 Robustizing the Least Squares Approach, 162
7.4 Asymptotics of Robust Regression Estimates, 164
7.5 Conjectures and Empirical Results, 170
7.6 Asymptotic Covariances and Their Estimation, 172
7.7 Concomitant Scale Estimates, 175
7.8 Computation of Regression *M*-Estimates, 179
7.9 Moderate Leverage Points, 192
7.10 Analysis of Variance, 195

8 ROBUST COVARIANCE AND CORRELATION MATRICES 199

8.1 General Remarks, 199
8.2 Estimation of Matrix Elements through Robust Variances, 202
8.3 Estimation of Matrix Elements through Robust Correlation, 204
8.4 An Affinely Invariant Approach, 211
8.5 Estimates Determined by Implicit Equations, 213
8.6 Existence and Uniqueness of Solutions, 215
8.7 Influence Functions and Qualitative Robustness, 223
8.8 Consistency and Asymptotic Normality, 226
8.9 Breakdown Point, 227
8.10 Least Informative Distributions, 229
8.11 Some Notes on Computation, 237

9 ROBUSTNESS OF DESIGN 243

9.1 General Remarks, 243
9.2 Minimax Global Fit, 243
9.3 Minimax Slope, 251

10 EXACT FINITE SAMPLE RESULTS 253

10.1 General Remarks, 253
10.2 Lower and Upper Probabilities and Capacities, 254
10.3 Robust Tests, 264
10.4 Sequential Tests, 273
10.5 The Neyman-Pearson Lemma for 2-Alternating Capacities, 275
10.6 Estimates Derived from Tests, 278
10.7 Minimax Interval Estimates, 282

11 MISCELLANEOUS TOPICS 286

11.1 Hampel's Extremal Problem, 286
11.2 Shrinking Neighborhoods, 290

REFERENCES 294

INDEX 301

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. During the past decades one has become increasingly aware that some of the most common statistical procedures (in particular, those optimized for an underlying normal distribution) are excessively 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 have 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 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 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 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 $\sqrt{2/\pi}\ \sigma \approx 0.80\sigma$. 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:

$$\mathrm{ARE}(\varepsilon)=\lim_{n\to\infty}\frac{\operatorname{var}(s_n)/(Es_n)^2}{\operatorname{var}(d_n)/(Ed_n)^2}$$

$$=\frac{\left[\dfrac{3(1+80\varepsilon)}{(1+8\varepsilon)^2}-1\right]\Big/4}{\dfrac{\pi(1+8\varepsilon)}{2(1+2\varepsilon)^2}-1}.$$

The results are summarized in the Exhibit 1.1.1.

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.

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

To avoid misunderstandings, we should hasten to emphasize what is *not* implied here. First, the above does not imply that we advocate the use of the mean absolute deviation (there are still better estimates of scale). Second, some people have argued that the example is unrealistic insofar as the "bad" observations will stick out as outliers, so any conscientious statistician will do something about them before calculating the mean square error. This is beside the point; outlier rejection followed by the mean square error might very well beat the performance of the mean absolute error, but we are concerned here with the behavior of the *unmodified* classical estimates.

The example clearly has to do with longtailedness: lengthening the tails of the underlying distribution explodes the variance of s_n (d_n is much less affected). Shortening the tails, on the other hand, produces quite negligible effects on the distributions of the estimates. (It may impair the absolute efficiency by decreasing the asymptotic Cramér-Rao bound, but the latter is so unstable under small changes of the distribution that this effect cannot be taken very seriously.)

The sensitivity of classical procedures to longtailedness is typical and not limited to this example. As a consequence "distributionally robust" and "outlier resistant," although conceptually distinct, are practically synonymous notions. Any reasonable, formal or informal, procedure for rejecting outliers will prevent the worst.

We might therefore ask whether robust procedures are needed at all; perhaps a two-step approach would suffice:

(1) First clean the data by applying some rule for outlier rejection.

(2) Then use classical estimation and testing procedures on the remainder.

Would these steps do the same job in a simpler way?

Unfortunately they will not, for the following reasons:

(1) It is rarely possible to separate the two steps cleanly; for instance, in multiparameter regression problems outliers are difficult to recognize unless we have reliable, robust estimates for the parameters.

(2) Even if the original batch of observations consists of normal observations interspersed with some gross errors, the cleaned data will not be normal (there will be statistical errors of both kinds, false rejections and false retentions), and the situation is even worse when the original batch derives from a genuine nonnormal distribution, instead of from a gross-error framework. Therefore the classical normal theory is not applicable to cleaned samples, and the

actual performance of such a two-step procedure may be more difficult to work out than that of a straight robust procedure.

(3) It is an empirical fact that the best rejection procedures do not quite reach the performance of the best robust procedures. The latter apparently are superior because they can make a smooth transition between full acceptance and full rejection of an observation.

1.2 WHAT SHOULD A ROBUST PROCEDURE ACHIEVE?

We are adopting what might be called an "applied parametric viewpoint": we have a parametric model, which hopefully is a good approximation to the true underlying situation, but we cannot and do not assume that it is exactly correct. Therefore any statistical procedure should possess the following desirable features:

(1) It should have a reasonably good (optimal or nearly optimal) efficiency at the assumed model.

(2) It should be robust in the sense that small deviations from the model assumptions should impair the performance only slightly, that is, the latter (described, say, in terms of the asymptotic variance of an estimate, or of the level and power of a test) should be close to the nominal value calculated at the model.

(3) Somewhat larger deviations from the model should not cause a catastrophe.

If asymptotic performance criteria are used, some care is needed. In particular, the convergence should be uniform over a neighborhood of the model, or there should be at least a one-sided uniform bound, because otherwise we cannot guarantee robustness for any finite n, no matter how large n is. This point has often been overlooked in the past.

It should be emphasized once more that the occurrence of gross errors in a small fraction of the observations is to be regarded as a small deviation, and that, in view of the extreme sensitivity of some classical procedures, a primary goal of robust procedures is to safeguard against gross errors.

The literature contains many other explicit and implicit goals for robust procedures, for example, high asymptotic *relative efficiency* (relative to some classical reference procedures), or high *absolute efficiency*, and this either for completely arbitrary (sufficiently smooth) underlying distributions, or for a specific parametric family.