Monographs on Statistics and Applied Probability 120

# Statistical Inference

## The Minimum Distance Approach

**Ayanendranath Basu**
Indian Statistical Institute
Kolkata, India

**Hiroyuki Shioya**
Muroran Institute of Technology
Muroran, Japan

**Chanseok Park**
Clemson University
Clemson, South Carolina, USA

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

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



Printed in the United States of America on acid-free paper
10 9 8 7 6 5 4 3 2 1

International Standard Book Number: 978-1-4200-9965-2 (Hardback)



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

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

*To Manjula, Srabashi, and Padmini* – AB –

*To my wife, my father, and mother* – HS –

*To Late Professor Byung Ho Lee* – CP –

# *Contents*

# *Preface*

In many ways, estimation by an appropriate minimum distance method is one of the most natural ideas in statistics. A parametric model imposes a certain structure on the class of probability distributions that may be used to describe real life data generated from a process under study. There hardly appears to be a better way to deal with such a problem than to choose the parametric model that minimizes an appropriately defined distance between the data and the model.

The issue is an important and complex one. There are many different ways of constructing an appropriate "distance" between the "data" and the "model." One could, for example, construct a distance between the empirical distribution function and the model distribution function by a suitable measure of distance. Alternatively, one could minimize the distance between the estimated data density (obtained, if necessary, by using a nonparametric smoothing technique such as kernel density estimation) and the parametric model density. And when the particular nature of the distances has been settled (based on distribution functions, based on densities, etc.), there may be innumerable options for the distance to be used within the particular type of distances. So the scope of study referred to by "Minimum Distance Estimation" is literally huge.

Statistics is a modern science. In the early part of its history, minimum distance estimation was not a research topic of significant interest compared to some other topics. There may be several reasons for this. The neat theoretical development of the idea of maximum likelihood and its superior performance under model conditions meant that any other competing procedure would have had to make a real case for itself before being proposed as a viable alternative to maximum likelihood. Until other considerations such as robustness over appropriate neighborhoods of the parametric model came along, there was hardly any reason to venture outside the fold of maximum likelihood, particularly given the computational simplicity of the maximum likelihood method in most common parametric models, which minimum distance methods in general do not share.

The growth of the area of research covered by the present book can be attributed to several factors. Two of them require special mention. The first one is the growth of computing power. As in all other areas of science, research in statistical science got a major boost with the advent of computers. Previously intractable problems became numerically accessible. Approximate methods could be applied with enhanced degrees of precision. Computational complexity of the procedure became a matter of minor concern, rather than the major deciding factor. This made the construction of distances and the computation of the estimators computationally feasible. The second major reason is the emergence of the area of robust statistical inference. It was no longer sufficient to have a technique which was optimal under model conditions but had weak robustness properties. Several minimum distance techniques have natural robustness properties under model misspecifications. Thus, the computational advances and the practical requirements converged to facilitate the growth of research in minimum distance methods.

Among the class of minimum distance methods we have focused, in this book, on density-based minimum distance methods. Carrying this specialization further, our emphasis, within the class of density-based distances has been on the chi-square type distances. Counting from Beran's path breaking 1977 paper, this area has seen a major spurt of research activity during the last three decades. In fact, the general development of the chi-square type distances began in the 1960s with Csiszár (1963) and Ali and Silvey (1966), but the robustness angle in this area probably surfaced with Beran. The procedures within the class of "$\phi$-divergences" or "disparities" are popular because many of them combine strong robustness features with full asymptotic model efficiency.

There is no single book which tries to provide a comprehensive documentation of the development of this theory over the last 30 years or so. Our primary intention here has been to fill in this gap. Our development has mainly focused on the problem for independently and identically distributed data. But we have tried to be as comprehensive as possible in this regard in establishing the basic structure of this inference procedure so that the reader is sufficiently prepared to grasp the applications of this technique to more specialized scenarios. We have discussed the estimation and hypothesis testing problems for both discrete and continuous models, extensively described the robustness properties of the minimum distance methods, discussed the inlier problem and its possible solutions, described weighted likelihood estimators and considered several other related topics. We trust that this book will be a useful resource for any researcher who takes up density-based minimum distance estimation in the future.

Apart from minimum distance estimation based on chi-square type distances, on which we have spent the major part of this book, we have briefly looked at three other topics. These may be described as (i) minimum distance estimation based on the density power divergence; (ii) some recent developments on goodness-of-fit tests based on disparities and their modifications, and (iii) a discussion of the applications of these minimum distance methods in information theory and engineering. We believe that the last item will make the book useful to scientists outside the mainstream statistics area.

In this connection it is appropriate to mention some closely related books that are available in the literature. The book by Pardo (2006) gives an excellent description of minimum $\phi$-divergence procedures and is a natural resource for this area. However, Pardo deals almost exclusively, although thoroughly, with discrete models. In our book we have also provided an extensive description of continuous models. Besides, the robustness angle is a driving theme of our book, unlike Pardo's case.

Our discussion of the multinomial goodness-of-fit testing problem has been highly influenced by the classic by Read and Cressie (1988). However, we have made every effort not to be repetitive, and only described such topics not covered extensively by Cressie and Read (or extended their findings beyond the power divergence family). Unlike the minimum distance inference case where we have tried to be comprehensive, in the goodness-of-fit testing problem we have been deliberately selective.

We have also kept the description to a level where it will be easily accessible to students who have been exposed to first-year graduate courses in statistics. Our presentation, although sufficiently technical, does not assume a measure theoretic background for the reader and, except in Chapter 11, the rare references to measures do not arrest the flow of the book. The book can very well serve as the text for a one-semester graduate course in minimum distance methods.

We take this opportunity to acknowledge the help we have received from many colleagues, teachers, and students while completing the book. We should begin by acknowledging our intellectual debt to Professor Bruce G. Lindsay, two of the three authors of the current book being his Ph.D. advisees. Discussions with Professors Leandro Pardo, Marianthi Markatou and Claudio Agostinelli have been very helpful. Discussion with Professor Subir Bhandari has helped to make many of our mathematical derivations more rigorous. Many other colleagues, too innumerable to mention here, have helped us by drawing our attention to related works. We also thank Professor Wen-Tao Huang, who was instrumental in bringing this group of authors together.

Special thanks must be given to Dr. Abhijit Mandal; his assistance in working out many of the examples in the book and constructing the figures has been invaluable. Dr. Rohit Patra and Professor Biman Chakraborty also deserve thanks in this connection.

Finally, we wish to thank all our friends and family members who stood by us during the sometimes difficult phase of manuscript writing.

Ayanendranath Basu
*Indian Statistical Institute*
*India*

Hiroyuki Shioya
*Muroran Instiute of Technology*
*Japan*

Chanseok Park
*Clemson University*
*USA*

# *Acknowledgments*

Some of the numerical examples and figures represented here are from articles copyrighted to different journals or organizations. They have been reproduced with the permission of the appropriate authorities. A list is presented below, and their assistance in permitting these reproductions is gratefully acknowledged.

The simulated results in Example 5.5 have been reproduced from *Sankhya*, Series B, Volume 64, Basu, A. (author), Outlier resistant minimum divergence methods in discrete parametric models, pp. 128–140 (2002), with kind permission from *Sankhya*.

The simulated results in Example 6.1, together with Figures 6.1, 6.2 and Table 6.1, are reproduced from *Statistica Sinica*, Volume 8, Basu, A. and Basu, S. (authors), Penalized minimum disparity methods for multinomial models, pp. 841–860 (1998), with kind permission from *Statistica Sinica*.

The real data example in Section 9.4.1, together with Figures 9.2 and 9.3 have been reproduced from *The Annals of the Institute of Statistical Mathematics*, Volume 58, Basu, S., Basu, A. and Jones, M. C. (authors), Robust and efficient estimation for censored survival data, pp. 341–355 (2006), with kind permission from the *Institute of Statistical Mathematics*.