OPTIMAL ESTIMATION OF PARAMETERS

This book presents a comprehensive and consistent theory of estimation. The framework described leads naturally to the maximum capacity estimator as a generalization of the maximum likelihood estimator. This approach allows the optimal estimation of real-valued parameters, their number and intervals, as well as providing common ground for explaining the power of these estimators.

Beginning with a review of coding and the key properties of information, the author goes on to discuss the techniques of estimation, and develops the generalized maximum capacity estimator, based on a new form of Shannon's mutual information and channel capacity. Applications of this powerful technique in hypothesis testing and denoising are described in detail.

Offering an original and thought-provoking perspective on estimation theory, Jorma Rissanen's book is of interest to graduate students and researchers in the fields of information theory, probability and statistics, econometrics, and finance.

JORMA RISSANEN was a member of research staff in the IBM Almaden Research Center from 1965 to 2001, and is currently Professor Emeritus at Tampere University of Technology, Finland. Among his main results are the introduction of the MDL principle for statistics, the invention of arithmetic coding, and the introduction of variable-length Markov chains with the associated Algorithm Context. He has received many awards, including the 2007 Kolmogorov Medal from the University of London's CLRC, and the 2009 Shannon Award from the Information Theory Society. He received two Outstanding Innovation Awards from IBM, one in 1980 and the other in 1988, and a Corporate Award in 1991.

The minimum description length (MDL) principle is a very universal principle of statistical modeling in estimation, prediction, testing, and coding. Jorma Rissanen, the pioneer of the MDL principle, evolves a new theory to reach the most general and complete notion, which he calls the complete MDL principle. In this book the author derives it by introducing the key notion of maximum capacity. The most fundamental methods of estimation such as maximum likelihood estimation and the MDL estimation are naturally derived as the maximum capacity estimators, and their optimality is justified within a unifying theoretical framework. Through the book, readers can revisit the meaning of estimation from the author's very original viewpoint, and will enjoy the most advanced version of the MDL principle.

Kenji Yamanishi, *The University of Tokyo*

In this splendid new book, Jorma Rissanen, the originator of the minimum description length (MDL) Principle, puts forward a comprehensive theory of estimation which differs in several ways from the standard Bayesian and frequentist approaches. During the development of MDL over the last 30 years, it gradually emerged that MDL could be viewed, informally, as a maximum *probability* principle that directly extends Fisher's classical maximum *likelihood* method to allow for estimation of a model's structural properties. Yet providing a formal link between MDL and maximum probability remained elusive until the arrival of this book. By making the connection mathematically precise, Rissanen now ties up the loose ends of MDL theory and at the same time develops a beautiful, unified, entirely original and fully coherent theory of estimation, which includes hypothesis testing as a special case.

Peter Grünwald, *Centrum voor Wiskunde en Informatica, The Netherlands*

OPTIMAL ESTIMATION OF PARAMETERS

JORMA RISSANEN
Tampere University of Technology
Helsinki Institute for Information Technology

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town,
Singapore, São Paulo, Delhi, Mexico City

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9781107004740

First published 2012

Printed in the United Kingdom at the University Press, Cambridge

A catalog record for this publication is available from the British Library

Library of Congress Cataloging in Publication data
Rissanen, Jorma.
Optimal estimation of parameters / Jorma Rissanen, Tampere University of Technology,
Helsinki Institute for Information Technology.
pages cm
ISBN 978-1-107-00474-0 (hardback)
1. Estimation theory. I. Title.
QA276.8.R57 2012
519.5′44 – dc23 2012007503

ISBN 978-1-107-00474-0 Hardback

Contents

Preface

I have a long lasting interest in estimation, which started with attempts to control industrial processes. It did not take long to realize that the control part is easy if you knew the behavior of the process you want to control, which meant that the real problem is estimation. When I was asked by the Information Theory Society to give the 2009 Shannon Lecture I thought of giving a coherent survey of estimation theory. However, during the year given to prepare the talk I found that it was not possible, because there was no coherent theory of estimation. There was a collection of facts and results but they were isolated with little to connect them. To my surprise this applied even to the works of some of the greatest names in statistics, such as Fisher, Cramér, and Rao, which I had been familiar with for decades, but which I had never questioned until now that I was more or less forced to do so. As an example, the famous maximum likelihood estimator due to Fisher [12] had virtually no formal justification. The celebrated Cramér–Rao inequality gives it a non-asymptotic justification only for special models and for more general parametric models only an asymptotic justification. Clearly, no workable theory should be founded on asymptotic behavior. About the value of asymptotics, we quote Keynes' famous quip that "asymptotically we all shall be dead."

In trying to prepare the talk I was faced with a large jigsaw puzzle in which the pieces did not quite fit, and some crucial pieces were completely missing. After a considerable struggle I was able to make the pieces fit but to do so I had to alter virtually all of them and ignore the means and concepts introduced by the masters, who of course did not have access to information and coding theory. Their followers, however, did, but since this theory did not immediately solve their problems it was regarded as irrelevant and ignored. About the only concept that survived was the maximum likelihood estimator, and even that was as a special case of a more powerful *maximum capacity* estimator, which permits estimation of the number of parameters as well as intervals. The result is a theory of estimation which covers all the aspects of estimation that I can think of including, even hypothesis testing, which in this treatment is founded on estimation, and is quite different from the usual.