

Short Book Reviews

Editor: Simo Puntanen

Information and Complexity in Statistical Modeling

Jorma Rissanen

Springer, 2007, viii + 142 pages, US\$ 44.95, hardcover

ISBN: 978-0-387-36610-4

Table of contents

- | | |
|--------------------------------|-------------------------------------|
| 1. Introduction | 5. Stochastic complexity |
| PART I: Information and coding | 6. Structure function |
| 2. Shannon–Wiener information | 7. Optimally distinguishable models |
| 3. Coding of random processes | 8. The MDL principle |
| PART II: Statistical modelling | 9. Applications |
| 4. Kolmogorov complexity | |

Readership: Graduate students and researchers in statistics, computer science and engineering, anyone interested in statistical modelling.

This book presents a personal introduction to statistical modelling based on the principle that the objective of modelling is to extract learnable information from data with suggested classes of probability models. It grew from lectures to doctoral students at Tampere University of Technology and retains much of the economical style of a lecture series. The ideas are emphasized and proofs of theorems are given when they are short, while referring to the literature for the details of more lengthy proofs. The author does not hesitate to give important statements and relationships without proofs or with sketches, with the understanding that such proofs under reasonable conditions can be given.

The intuitive and fundamental concepts of complexity, learnable information and noise are formalized, which provides a firm information-theoretic foundation for statistical modelling. Kolmogorov's structure function in the algorithmic theory of complexity is extended to statistical models. Given the models in a suggested class, the stochastic complexity is defined as the shortest code length with which the data can be encoded. This amounts to the MDL (Minimum Description Length) principle. Although no actual coding is needed in statistical applications, the fundamental one-to-one correspondence with code length and probability measurements should be realized.

The MDL theory provides a broad principle rather than tailored criteria to be applied blindly. It is not always possible to find easily computed closed form solutions to implement it but one may derive even several criteria from the same principle. Therefore, this fascinating volume offers an excellent source of important statistical research problems calling for solution.

Erkki P. Liski
Department of Mathematics, Statistics and Philosophy
FI-33014 University of Tampere, Finland
E-mail: erkki.liski@uta.fi