Review of 'Statistical and Inductive Inference by Minimum Message Length' by C.S. Wallace

Peter Grünwald
CWI, Amsterdam, the Netherlands

March 13, 2006

This book is about the *Minimum Message Length (MML) Principle*, an information-theoretic approach to induction, hypothesis testing, model selection and statistical inference. MML, which can be seen as a mathematically precise version of Occam's Razor, asserts that the "best" explanation of the observed data is the shortest.

The book is essentially the manuscript left behind by Professor Chris Wallace when he died on August 7, 2004. Professor Wallace was a remarkably versatile scientist: originally trained as a physicist, he has made significant contributions to various areas of computer science including computer architecture, arithmetic and simulation. Simultaneously, he is the main originator of the MML Principle, which has implications for applied and theoretical statistics as well as for the philosophy of science. MML was developed between 1968 and 2004 in a series of papers by Wallace and several coworkers such as D. Boulton, P. Freeman and D. Dowe. These papers mostly concern practical applications of MML to problems of statistics and machine learning, as well as the mathematical development of the underlying ideas. Since these papers were often quite short and technical, there was much need of an extensive introduction to, and comprehensive overview of, the MML idea. As its first contribution, this book provides such a, most welcome, introduction. Yet its aim and scope are much wider: as its second contribution, the book presents MML as a general theory of inductive inference, and there is extensive discussion of its philosophical foundations and implications. Thus, the book will appeal to a broad audience: the technical part of the book is mainly of interest to researchers in applied Machine Learning, Data Mining and Statistics who wish to learn about a non-main stream but practically successful, generally applicable statistical method. The philosophical part is interesting for researchers in these fields, as well as philosophers of science, who will enjoy Wallace's original ideas on the foundations of statistical and inductive inference.

The technical part (Chapters 2-7, and, in part, Chapter 1) describes the MML Principle in great detail. It starts with a discussion of probabilistic and statistical preliminaries, with an emphasis on Bayesian inference and decision theory. This is followed, in Chapter 2, by an introduction to Shannon information theory and algorithmic (or Kolmogorov) complexity. This chapter introduces the idea of a two-part code as an explanation of observed data. This idea is the cornerstone of "strict" MML estimation, which is introduced in Chapter 3. Strict MML (SMML) is really just a prescription of an algorithm for doing inductive inference in the form of hypothesis testing, model selection and parameter estimation. Unfortunately, in many situations this algorithm is not computationally efficient. Chapters 4 and 5 discuss several efficient approximations to it, including the quadratic approximation to the SMML estimator that has been most commonly used in practice, usually simply under the name of 'MML'. Chapter 6 gives some details of MML and SMML for well-known statistical models such as Poisson distributions, mixture models and linear regression models. Chapter 7 discusses 'structural models' such as regular grammars, probabilistic finite state machines, classification trees and nets and causal networks.

At this point, the discussion of MML 'in the small' – essentially an algorithm – ends abruptly. Instead, the principle is now applied informally and 'in the large' to grand problems of physics (Chapter 8) and natural language and philosophy of science (Chapter 9). Chapter 10 connects MML to J. Rissanen's related, and perhaps more well-known *Minimum Description Length (MDL) Principle*, as well as to earlier related work by R. Solomonoff on Kolmogorov complexity and prediction.

Both parts of the book are very well written. The first part of the book relies on quite a lot of mathematics. While invariably well-explained, its treatment does reveal Wallace's roots as a physicist: the standards of rigor would probably not be acceptable to most mathematicians.

The book is highly opinionated. I dare to predict that hardly anybody will agree with the author on all issues. To give but one example, the book takes a subjective Bayesian stance, usually interpreting prior distributions as representing prior beliefs. To some statisticians (even some 'objective' Bayesian statisticians) such a subjective approach seems untenable in practice. Yet Wallace then argues that this approach, while providing the right framework, is incomplete: it misses a well-founded method of parameter estimation and model selection that can be used in the absence of a utility function. MML provides exactly that: a technically sound, unifying treatment of parameter estimation and model selection from a Bayesian perspective. Subjective Bayesian statisticians, who would agree with the first part of the author's message, may tend to disagree with this second part! Like many Bayesians, Wallace points out serious deficiencies in Bayesian MAP (Maximum A Posteriori), mean and median estimates. But unlike most other Bayesians, he insists that estimation and model selection sometimes have to be carried out when no clear utility function is available (page 43-45).

The author explains and defends his stance very well, which makes the book interesting to read even if, like this reviewer, one often disagrees. Yet, and this is my only real criticism of the book, Wallace ignores a lot of relevant literature. He frankly admits in a disclaimer in the preface, and later at several other places in the book, that he does not always properly represent alternative methods and viewpoints. Taking my own field of expertise as an example, in the discussion of Rissanen's MDL vs. Wallace's MML principle, the emphasis is on one particular form of MDL - the so-called Normalized Maximum Likelihood method. Other forms, such as predictive MDL, are ignored, and this distorts the discussion somewhat. Similarly, there is no discussion at all of the recent game-theoretic approaches to learning and prediction, even though, just like MML, such approaches provide a bridge between computer science and statistics.

In conclusion, Wallace's writing is invariably very sharp, original, stimulating and thought-provoking. It would be a real pity if readers were scared away by the fact that this book is so highly opinionated. They would miss a lucid exposition of some profound ideas which have not received nearly as much attention as they deserve. In my opinion, this is one of the most interesting books on statistics to have appeared in the last few years.