Model Selection
IMS Lecture Notes - Monograph Series (2001) Volume 38

Scales of Evidence for Model Selection: Fisher versus Jeffreys

Bradley Efron and Alan Gous

Stanford University and Cariden Technologies, Inc.

Abstract

Model selection refers to a data-based choice among competing statistical models, for example choosing between a linear or a quadratic regression function. The most popular model selection techniques are based on interpretations of p-values, using a scale originally suggested by Fisher: .05 is moderate evidence against the smaller model, .01 is strong evidence, etc. Recent Bayesian literature, going back to work by Jeffreys, suggests a quite different answer to the model selection problem. Jeffreys provided an interpretive scale for Bayes factors, a scale which can be implemented in practice by use of the BIC (Bayesian Information Criterion.) The Jeffreys scale often produces much more conservative results, especially in large samples, so for instance a .01 p-value may correspond to barely any evidence at all against the smaller model. This paper tries to reconcile the two theories by giving an interpretation of Fisher's scale in terms of Bayes factors. A general interpretation is given which works fine when checked for the one-dimensional Gaussian problem, where standard hypothesis testing is seen to coincide with a Bayesian analysis that assumes stronger (more informative) priors than those used by the BIC. This argument fails in higher dimensions, where Fisher's scale must be made more conservative in order to get a proper Bayes justification.

Bradley Efron is Max H. Stein Professor of Statistics, Department of Statistics, Sequia Hall, Stanford University, Paolo Alto, CA 94305-4065, U.S.A; email: brad@stat.Stanford.EDU. Alan Gous is Director of Research & Development, Cariden Technologies, Inc., 1733 Woodside Road, Suite 310, Redwood City, CA 94061, U.S.A.; email: alan@cariden.com.