BAYESIAN MODEL SPECIFICATION

David Draper and Milovan Krnjajic
Department of Applied Mathematics and Statistics
Baskin School of Engineering
University of California, Santa Cruz
(e-mail: draper@soe.ucsc.edu)

Abstract

A standard (data-analytic) approach to statistical model specification, practiced with equal vigor by Bayesians and non-Bayesians alike, involves the initial choice, for the structure of the model, of one or another of a variety of standard parametric families, followed by modification of this initial choice – once data begin to arrive – if the data suggest deficiencies in the original specification. In this talk (a) I will argue that this approach is formally incoherent, because it amounts to using the data both to specify the prior distribution on structure space and to update using this data-determined prior; (b) I will identify two approaches to avoiding (at least in principle, and with a fair amount of data) the incoherence in (a) – (1) Bayesian nonparametric modeling and (2) three-way out-of-sample predictive validation; (c) I will provide details on implementing (2); (d) I will argue that to make progress in coherent Bayesian model specification in complicated problems you have to either implicitly or explicitly choose a utility structure which defines, for you, when the model currently being examined is "good enough"; (e) I will argue that it is best to make this choice explicitly on the basis of real-world considerations regarding the use to which the model will be put; and (f) I will contrast model selection methods based on the log score and deviance information criteria (DIC) as two examples of (e) with utilities governed by predictive accuracy.

Key Words: Bayesian model specification, DIC, model selection as a decision problem, predictive log scoring rule, three-way out-of-sample predictive validation, Bayesian nonparametric modeling