Bayesian model specification: towards a *Theory of Applied Statistics*

David Draper

Department of Applied Mathematics and Statistics, University of California, Santa Cruz, USA

Avon Local Group, Royal Statistical Society: 24 Nov 2011

Abstract

In this talk I'll examine some issues in the foundations of applied probability and applied statistics, a topic that has surprising relevance to day-to-day work in those fields.

Probability: From the 1650s (Fermat, Pascal) through the 18th century (Bayes, Laplace) to the period 1860–1930 (Venn, Boole, von Mises), three different ideas about how to think about uncertainty quantification — *classical*, *Bayesian*, and *frequentist* probability — were put forward in an intuitive way, but no one ever tried to prove a theorem of the form {given these premises, there's only one sensible way to quantify uncertainty} until the mathematician Kolmogorov (1933), the statistician and actuary de Finetti (1937), and the physicist RT Cox (1946). Kolmogorov — following and rigorizing Venn, Boole, and von Mises — created a repeated-sampling (frequentist) foundation based on set theory; this is excellent, as far as it goes, but many types of uncertainty cannot (uniquely, comfortably) be fit into Kolmogorov's framework. de Finetti developed a Bayesian foundation based on estimating betting odds about true/false *propositions*; this is more general than Kolmogorov, but betting odds are not fundamental to science. Cox — following and rigorizing Laplace — built up a Bayesian foundation (starting with principles, developing them into axioms, and finally proving a theorem) based on numerically summarizing the information content of propositions; this is both fundamental to science and as general as you can get, and it's the story that makes the most sense to me. The foundations of a version of probability flexible enough for an extremely wide range of statistical applications seem secure.

Statistics: Given an unknown θ (this could be almost anything, but think of a vector in $\Re^k$ for concreteness), a data set D relevant to decreasing your uncertainty about θ (this could also be almost anything, but think of a vector in $\Re^n$ for concreteness), and a set $\mathcal{B}$ of propositions detailing your background assumptions and judgments describing {how the world works as far as θ, D and future data D^* are concerned}, in the Laplace-Cox approach (almost) everything you'd want to be able to do is covered by a set of three equations: one for *inference* about θ, one for *prediction* of D^*, and one for making a *decision* in spite of your uncertainty about θ. However, to implement this program you have to specify

- two things for inference and prediction: two probability distributions, $p(D|\theta\,\mathcal{B})$ and $p(\theta|\mathcal{B})$, summarizing the information about θ internal and external to D, respectively; and
- two more things for decision: a set $\mathcal{A}$ of possible actions and a utility function $U(a, \theta)$ quantifying the costs and benefits arising from the choice of action a if the unknown were in fact θ;

and Cox's Theorem is (almost entirely) silent about how to perform these specification tasks. At present we have no progression, from principles through axioms to theorems, that characterizes optimal Bayesian model specification; instead we have an ad-hoc collection of methods, some of which seem more or less sensible. Thus the foundations of applied statistics do not seem to me to be secure; fixing this would yield a *Theory of Applied Statistics*, which we need and do not yet have. In this talk I'll explore the extent to which four principles (Calibration, Modeling-As-Decision, Prediction, and Decision-Versus-Inference) constitute progress toward this goal.