Bayes Factors

Robert E. Kass and Adrian E. Raftery*

In a 1935 paper and in his book *Theory of Probability*, Jeffreys developed a methodology for quantifying the evidence in favor of a scientific theory. The centerpiece was a number, now called the *Bayes factor*, which is the posterior odds of the null hypothesis when the prior probability on the null is one-half. Although there has been much discussion of Bayesian hypothesis testing in the context of criticism of *P*-values, less attention has been given to the Bayes factor as a practical tool of applied statistics. In this article we review and discuss the uses of Bayes factors in the context of five scientific applications in genetics, sports, ecology, sociology, and psychology.

We emphasize the following points:

- From Jeffreys' Bayesian viewpoint, the purpose of hypothesis testing is to evaluate the evidence in favor of a scientific theory.
- Bayes factors offer a way of evaluating evidence *in favor of* a null hypothesis.
- Bayes factors provide a way of incorporating external information into the evaluation of evidence about a hypothesis.
- Bayes factors are very general and do not require alternative models to be nested.
- Several techniques are available for computing Bayes factors, including asymptotic approximations that are easy to compute using the output from standard packages that maximize likelihoods.
- In "nonstandard" statistical models that do not satisfy common regularity conditions, it can be technically simpler to calculate Bayes factors than to derive non-Bayesian significance tests.
- The Schwarz criterion (or BIC) gives a rough approximation to the logarithm of the Bayes factor, which is easy to use and does not require evaluation of prior distributions.
- When one is interested in estimation or prediction, Bayes factors may be converted to weights to be attached to various models so that a composite estimate or prediction may be obtained that takes account of structural or model uncertainty.
- Algorithms have been proposed that allow model uncertainty to be taken into account when the class of models initially considered is very large.
- Bayes factors are useful for guiding an evolutionary model-building process.
- It is important, and feasible, to assess the sensitivity of conclusions to the prior distributions used.

KEY WORDS: Bayesian hypothesis tests; BIC; Importance sampling; Laplace method; Markov chain Monte Carlo; Model selection; Monte Carlo integration; Posterior model probabilities; Posterior odds; Quadrature; Schwarz criterion; Sensitivity analysis; Strength of evidence.

1. INTRODUCTION

The Bayesian approach to hypothesis testing was developed by Jeffreys (1935, 1961) as a major part of his program for scientific inference. Although Jeffreys called his methods "significance tests," apparently borrowing the term from Fisher, this is misleading, because Jeffreys's perspective and goals were quite different. Jeffreys was concerned with the comparison of predictions made by two competing scientific theories. In his approach, statistical models are introduced to represent the probability of the data according to each of the two theories, and Bayes's theorem is used to compute the posterior probability that one of the theories is correct.

Considerable attention has been given to distinctions between the two approaches (e.g., Berger and Delampady 1987, Berger and Berry 1988, and references therein). Often lost from the controversy, however, are the practical aspects of

the Bayesian methods: how conclusions may be drawn from them, how they can provide answers when non-Bayesian methods are hard to construct, what their strengths and limitations are. These concerns are the focus of this article. We will also discuss the Bayesian approach to accounting for uncertainty in the model-building process, which is closely connected to the methodology for hypothesis testing.

In Section 2 we motivate the work with several applications from the areas of genetics, sports, ecology, sociology, and psychology. These help connect hypothesis testing with model selection and introduce several problems that Bayesian methodology can solve, including the evaluation of the evidence in favor of a null hypothesis, the inclusion of other information in the weighing of evidence, the comparison of nonnested models, and accounting for uncertainty in the choice of models. In Section 3 we introduce the *Bayes factor*, which is the posterior odds of one hypothesis when the prior probabilities of the two hypotheses are equal.

Bayesian methods involve integrals and thus, often, numerical integration. Many integration techniques have been adapted to problems of Bayesian inference, including the computation of Bayes factors; this is discussed in Section 4. Bayes factors require priors on the parameters appearing in the models that represent the competing hypotheses. The choice of these priors and the extent to which Bayes factors are sensitive to this choice is discussed in Section 5.