Statistical Science
2009, Vol. 24, No. 2, 176–178
DOI: 10.1214/09-STS284D
Main article DOI: 10.1214/09-STS284

Bayes, Jeffreys, Prior Distributions and the Philosophy of Statistics[1]

Andrew Gelman

I actually own a copy of Harold Jeffreys's *Theory of Probability* but have only read small bits of it, most recently over a decade ago to confirm that, indeed, Jeffreys was not too proud to use a classical chi-squared p-value when he wanted to check the misfit of a model to data (Gelman, Meng and Stern, 2006). I do, however, feel that it is important to understand where our probability models come from, and I welcome the opportunity to use the present article by Robert, Chopin and Rousseau as a platform for further discussion of foundational issues.[2]

In this brief discussion I will argue the following: (1) in thinking about prior distributions, we should go beyond Jeffreys's principles and move toward weakly informative priors; (2) it is natural for those of us who work in social and computational sciences to favor complex models, contra Jeffreys's preference for simplicity; and (3) a key generalization of Jeffreys's ideas is to explicitly include model checking in the process of data analysis.

THE ROLE OF THE PRIOR DISTRIBUTION IN BAYESIAN DATA ANALYSIS

At least in the field of statistics, Jeffreys is best known for his eponymous prior distribution and, more generally, for the principle of constructing noninformative, or minimally informative, or objective, or reference prior distributions from the likelihood (see, for example, Kass and Wasserman, 1996). But it can notoriously difficult to choose among noninformative priors; and, even more importantly, seemingly noninformative distributions can sometimes have strong and undesirable implications, as I have found in my own experience (Gelman, 1996, 2006). As a result I have become a convert to the cause of *weakly informative priors*, which attempt to let the data speak while being strong enough to exclude various "unphysical" possibilities which, if not blocked, can take over a posterior distribution in settings with sparse data—a situation which is increasingly present as we continue to develop the techniques of working with complex hierarchical and nonparametric models.

HOW THE SOCIAL AND COMPUTATIONAL SCIENCES DIFFER FROM PHYSICS

Robert, Chopin and Rousseau trace the application of Ockham's razor (the preference for simpler models) from Jeffreys's discussion of the law of gravity through later work of a mathematical statistician (Jim Berger), an astronomer (Bill Jefferys) and a physicist (David MacKay). From their perspective, Ockham's razor seems unquestionably reasonable, with the only point of debate being the extent to which Bayesian inference automatically encompasses it.

My own perspective as a social scientist is completely different. I've just about never heard someone in social science object to the *inclusion* of a variable or an interaction in a model; rather, the most serious criticisms of a model involve worries that certain potentially important factors have *not* been included. In the social science problems I've seen, Ockham's razor is at best an irrelevance and at worse can lead to acceptance of models that are missing key features that the data could actually provide information on. As such, I am no fan of methods such as BIC that attempt to justify the use of simple models that do not fit observed data. Don't get me wrong—*all the time* I use simple

Andrew Gelman is Professor, Department of Statistics and Department of Political Science, Columbia University (e-mail: gelman@stat.columbia.edu; URL: http://www.stat.columbia.edu/~gelman).

[1]Discussion of "Harold Jeffreys's Theory of Probability revisited," by Christian Robert, Nicolas Chopin, and Judith Rousseau, for *Statistical Science.*

[2]On the topic of other books on the foundations of Bayesian statistics, I confess to having found Savage (1954) to be nearly unreadable, a book too much of a product of its time in its enthusiasm for game theory as a solution to all problems, an attitude which I find charming in the classic work of Luce and Raiffa (1957) but more of annoyance in a book of statistical methods. When it comes to Cold War-era foundational work on Bayesian statistics, I much prefer the work of Lindley, in his 1965 book and elsewhere.

Also, I would be disloyal to my coauthors if I did not report that, despite what is said in the second footnote in the article under discussion, there is at least one other foundational Bayesian text of 1990s vintage that continues to receive more citations than Jeffreys.