Comment

Arnold Zellner

1. INTRODUCTION

In this stimulating and important paper, Bayes factors, posterior probabilities and P-values are considered in relation to the problem of using data to evaluate precise or sharp null hypotheses, for example $\theta = 0$ or $\theta = 1.0$. This is a very basic problem encountered in all areas of science and thus the fact that the authors, along with Jeffreys (1967) and others, conclude that widely used P-values are unsatisfactory is noteworthy. This conclusion has important implications not only for textbook treatments of the theory of testing, but also for applied scientific work.

The authors explain Jeffreys' approach to testing and show that P-values diverge markedly from posterior probabilities associated with sharp null hypotheses. They also derive lower bounds for Bayes factors and posterior probabilities and provide some advice in answer to the question, "What should be done?" Although some of the points that the authors raise have appeared in the literature, it is doubtful that they have been expressed as clearly and forcefully as in the present paper. However, as might be expected in such a controversial area, where are some points that deserve further discussion. See, e.g., Jeffreys (1967, Chapters V to VII), Edwards, Lindman and Savage (1963), Jaynes (1984) and Zellner (1971, 1980, 1984) for earlier considerations of testing issues and computation of Bayes factors for a number of problems. After taking up some general points, I shall turn to technical points and then provide some concluding remarks.

2. GENERAL POINTS

Point 1. *H_1: $\theta = \theta_0$ versus H_2: $|\theta - \theta_0| < \varepsilon$, $\varepsilon > 0$, Given.* For years I have stated that we should be able to test either H_1 or H_2 or both. To say that H_2 is "realistic" or "true" is to make an unwarranted general *a priori* statement about the "real world." To be scientific, one can compute a Bayes factor for H_1 versus H_2, as Jeffreys (1967, page 367) suggests. He also states, "I think, however, that it is both impossible and undesirable [to replace H_1 with H_2]" (page 367). I won't review Jeffreys's arguments here since they are readily available. It does seem relevant to remark that in $s = .5gt^2$ and in $E = mc^2$, the powers of t and c are predicted by physical theory to be exactly equal to 2. Also the coefficient of t^2 in the former relation is exactly $.5g$ and of c^2 in the latter exactly m. Many other examples of sharp or precise hypotheses can be given and thus it is incorrect to exclude such hypotheses *a priori* or term them "unrealistic" and important to be able to test them well as Berger and Delampady indicate.

Point 2. *Laplace's versus Jeffreys' Approaches to Testing.* Jaynes (1980) raised this point, which is equivalent to Rejoinder 3: Just Use Confidence Intervals of Berger and Delampady. As pointed out in Jeffreys (1967), Zellner (1971, 1980) and Berger and Delampady, the background prior information is different when there is a suggested value θ_0 for θ. However, in Zellner and Siow (1979) and Zellner (1984) it is shown that consideration of three hypotheses, H_1: $\theta = 0$, H_2: $\theta > 0$, and H_3: $\theta < 0$ with prior probabilities, π_1, π_2 and π_3, respectively, and truncated Jeffreys' Cauchy priors under H_2 and H_3 leads to a Bayes factor for H_2 versus H_3 that can be exactly equal to the Laplacian or diffuse prior credible region results. Also, consideration of all three hypotheses together yields a synthesis of the considerations in the present paper and those in Cassella and Berger (1987).

Point 3. Berger and Delampady err in calling Jeffreys', my and some others' Bayesian testing procedures "mechanical" or "automatic" or "default" or "conventional" or "objective." Jeffreys (1967, page 252) explains that in testing there may be very little previous information or a great deal. If there is a great deal of prior information, Jeffreys (1967, page 252) and others would use an appropriate prior distribution to represent it. Although Jeffreys mainly analyzed the situation of "little previous information" in his book, this does not imply at all that he would use these procedures when there is a great deal of previous information. I have expressed similar views (Zellner, 1980, 1984). However, it is useful, as I believe that Berger and Delampady recognize, to have testing results for the case of little previous information "on the shelf" to be used *when appropriate*. Further, the priors that Jeffreys used for the normal mean problem and others can be given different location and scale parameters without much difficulty as Jeffreys (1967),

---

Arnold Zellner is H. G. B. Alexander Distinguished Service Professor of Economics and Statistics, Graduate School of Business, University of Chicago, Chicago, Illinois 60637.