Statistical Science
2011, Vol. 26, No. 2, 210–211
DOI: 10.1214/11-STS338REJ
Main article DOI: 10.1214/10-STS338

Rejoinder

Malay Ghosh

It is a pleasure to receive comments from two very distinguished statisticians who themselves have made fundamental contributions to the development of objective priors. Their comments clarify many of the ideas presented in this paper, thereby providing further insight to the selection of objective priors. I will respond individually to their comments.

BERNARDO

I agree with Professor Bernardo that prior elicitation is nearly impossible in situations which call for very complex models. What I meant to say is that with years of accumulated data (e.g., for medical practitioners), it is sometimes possible to elicit a reasonable prior for certain parameters of frequent interest (e.g., the cure probability of a particular drug). In dose-response models, it is often possible to find meaningful priors for the logistic regression parameters.

I agree with Bernardo that objective Bayesian methods are unquestionably more appealing than ad hoc frequentist methods. A classic example is the Behrens–Fisher problem. Also, he is correct in asserting that even frequentist concepts such as minimaxity, admissibility, etc. call for Bayesian tools, and objective priors can become quite handy for such situations. A point to note here, though, is that since these concepts are not primarily Bayesian, often the choice leads to quite unappealing priors. For example, for the binomial proportion, minimaxity demands a $(\sqrt{n}/2, \sqrt{n}/2)$ prior, where n is the sample size. I sincerely doubt that any practitioner will ever be interested in using such a prior.

I owe an apology to Professor Bernardo for not referring to Berger, Bernardo and Mendoza. I am also thankful to him for pointing out that in reference analysis, one does not let the sample size n go to infinity, but lets k, the conceptual number of replicates of the original experiment, go to infinity.

It was never my intention to advocate priors alternate to Jeffreys in the one parameter case. My sole objective was to point out that if one considers a general class of divergence priors, Jeffreys' prior emerges as the solution in the interior of the parameter space, but not on the boundary. This is more in the spirit of telling a complete story rather than preaching something new. For instance, in the binomial case, I do not recommend necessarily using the $\text{Beta}(1/4, 1/4)$ prior in preference to Jeffreys' $\text{Beta}(1/2, 1/2)$ prior unless there are other good reasons for using the former.

I like to point out that in the ratio of normal means problem, the probability matching criterion does not reproduce the conventional Fieller–Creasy frequentist solution. This has been exploited in a very general framework by Ghosh, Yin and Kim (2003). Also, I like to add that while reference priors have general universal appeal, often their choice is very much dependent on the ordering of parameters. This may be a daunting task, especially for very complex models. Presumably, one can salvage such situations by considering prediction rather than estimation.

SWEETING

I agree essentially with every single comment made by Professor Sweeting and indeed thank him for bringing out several important issues barely touched upon in my article. I take this opportunity to underscore a couple of the fundamental arguments that he has put forward.

The first one is the contrast between estimation and prediction. Bernardo's proper scoring rule is based on the negative of the logarithm of the prior predictive pdf, geared primarily toward parametric estimation. In contrast, the negative of the logarithm of the posterior predictive pdf is ideally suited for prediction. In many situations, it is difficult, if not impossible to pinpoint the parameter of interest. Predictive inference for unobserved but potentially observable quantities does not face this problem, and often is the most desired mode of inference. The currently popular neural nets and machine learning techniques aim solely toward prediction. A more classical example is finite population sampling where the goal is to find the predictive distribution of the unobserved given the observed.

The second important point is that often the prior can overshadow the data. The simple (albeit artificial) example put forward by Professor Sweeting amply

Malay Ghosh is Distinguished Professor, University of Florida, 223 Griffin–Floyd Hall, Gainesville, Florida 32611-8545, USA (e-mail: ghoshm@stat.ufl.edu).