Statistical Science
2007, Vol. 22, No. 3, 363–367
DOI: 10.1214/07-STS235REJ
Main article DOI: 10.1214/07-STS235

Rejoinder: Bayesian Checking of the Second Levels of Hierarchical Models

M. J. Bayarri and M. E. Castellanos

We would like to thank the discussants for the valuable insights and for commenting on important aspects of model checking that we did not touch in our paper. Our goal was modest (but crucial): to select an appropriate distribution with which to judge the compatibility of the data with a hypothesized (hierarchical) model, when the test statistic is not ancillary and an improper prior is used for the hyperparameters. Since it is important to emphasize that this is by no means the only aspect of model checking, the discussants' complementary contributions and comments are all most welcome. The specific technical contributions of Evans and Johnson are also appreciated, since their developments in this area were not mentioned in our review.

Several discussants have highlighted the importance of graphical displays in model checking. We will not comment on this because we entirely agree. We similarly agree with most of the discussants' other comments, although in this rejoinder we mainly concentrate on disagreements. Our comments are organized around the main topics that arise in the discussions. We keep the same notation and terminology used in the paper (although it does conflict with the notation used by some of the discussants).

ROLE OF PRIOR PREDICTIVE DISTRIBUTIONS WHEN MODEL UNCERTAINTY IS PRESENT

Bayesian analyses, when model uncertainty is present (model choice, model averaging), are based on the prior predictive distributions for the different models under consideration. Model checking is a quick-and-dirty shortcut to bypass model choice, and "pure" Bayesian reasoning indicates that all relevant information lies in the (prior) predictive distribution $m(\mathbf{x})$ for the entertained model.

M. J. Bayarri is Professor, Department of Statistics and Operation Research, University of Valencia, Burjassot, (Valencia), 46100 Spain (e-mail: susie.bayarri@uv.es). M. E. Castellanos is Associate Professor, Department of Statistics and Operation Research, Rey Juan Carlos University, Móstoles, (Madrid), 28933 Spain (e-mail: maria.castellanos@urjc.es).

As Evans points out, objective Bayes methodology should be guided by proper Bayes methodology, so objective Bayes model checking should also be based on the prior predictive distribution. The difficulty, however, is that only some aspects of this distribution can be utilized when the prior distribution is improper. Bayarri and Berger (1997, 1999, 2000) argue that the relevant aspect to consider for model checking is a conditional (prior) predictive distribution $m(\mathbf{x} \mid u)$, where $U = U(\mathbf{X})$ is an appropriate conditioning statistic such that the posterior $\pi(\boldsymbol{\theta} \mid u)$ is proper. Model checks (measures of surprise) computed with this distribution (such as p-values or relative surprise) are called *conditional predictive* measures.

If we use a statistic T to measure departure and use U for conditioning, the relevant distribution for model checking is then $m(t \mid u)$. Evans' prescription can be put in this framework with T ancillary and U sufficient (caution: Evans' notation switches the roles of T and U). Larsen and Lu's (from now on L&L) prescription for checking group i is also of this form with $T = T(\mathbf{X}_i)$ and $U = \mathbf{X}_{(-i)}$. The complete theory of Johnson (not sketched in his discussion) relies on the whole prior predictive. Hence, all these methods produce legitimate Bayesian measures of surprise. The posterior predictive distribution cannot be expressed in this way (it would produce a trivial, degenerate distribution).

Bayarri and Berger (1997, 1999) explore several choices of U and recommend use of the *conditional* MLE of $\boldsymbol{\theta}$, that is, the MLE computed in the conditional distribution $f(\mathbf{x} \mid t, \boldsymbol{\theta})$. The resulting measures of surprise (or model checks) were shown to basically coincide with the partial posterior measures; indeed, the conditional predictive distribution for that choice of U and the partial posterior predictive distribution are asymptotically equivalent (Robins, 1999; Robins, van der Vaart and Ventura, 2000).

We have concentrated on partial posterior measures because they are basically indistinguishable from the conditional predictive ones and they are easier to compute, but their Bayesian justification comes from the conditional predictive reasoning. We should perhaps have reiterated this in the paper.