Statistical Science
2007, Vol. 22, No. 3, 359–362
DOI: 10.1214/07-STS235B
Main article DOI: 10.1214/07-STS235

Comment: Bayesian Checking of the Second Level of Hierarchical Models: Cross-Validated Posterior Predictive Checks Using Discrepancy Measures

Michael D. Larsen and Lu Lu

1. INTRODUCTION

We compliment Bayarri and Castellanos (BC) on producing an interesting and insightful paper on model checking applied to the second level of hierarchical models. Distributions of test statistics (functions of the observed data not involving parameters) for judging appropriateness of hierarchical models typically involve nuisance (i.e., unknown) parameters. BC (2007) focus on ways to remove the dependency on nuisance parameters so that test statistics can be used to assess models, either through p-values or Berger's relative predictive surprise (RPS). They demonstrate shortcomings in terms of very low power of posterior predictive checks and a posterior empirical Bayesian method. They also demonstrate better performance of their partial posterior predictive (*ppp*) method over a prior empirical Bayesian method. Methods of Dey et al. (1998), O'Hagan (2003) and Marshall and Spiegelhalter (2003) also are compared.

Methods are contrasted in terms of whether they require proper prior distributions, how many measures of surprise (one per group or one total) are produced, and the degree to which data are used twice in estimation and testing. Their preferred method (*ppp*) can use improper prior distributions, which are referred to as objective, produces a single measure of surprise for each test statistic, and avoids double use of the data. For the models and statistics considered, in comparison to the alternatives presented, *ppp* has a more uniform null distribution of p-values and more power versus alternatives.

M. D. Larsen is Associate Professor, L. Lu is graduate student, Department of Statistics and Center for Survey Statistics & Methodology, Iowa State University, Snedecor Hall, Ames, Iowa 50011, USA (e-mail: larsen@iastate.edu; icyemma@iastate.edu).

In this discussion, we suggest that cross-validated posterior predictive checks using discrepancy measures hold some promise for evaluating complex models. We apply them to O'Hagan's data example, provide some comments on the paper and discuss possible future work.

2. CROSS-VALIDATED POSTERIOR PREDICTIVE CHECKS USING DISCREPANCY MEASURES

Suppose there are data for I groups: $X_i, i = 1, \dots, I$, where $X_i = (X_{ij}, j = 1, \dots, n_i)$. The unknown parameters in the first level in group i are θ_i: $f(X_i|\theta_i)$ independently. The parameters in the second level of the model are η: $\pi(\theta|\eta) = \prod_{i=1}^{I} \pi(\theta_i|\eta)$. The prior distribution on η is $\pi(\eta)$. Let $D(X, \theta, \eta)$ be a generalized discrepancy measure. If $D(X, \theta, \eta) = D(X)$, then it is a test statistic. Examples are given in the next section for the normal-normal model considered by BC (2007). Cross-validated posterior predictive model checking using a discrepancy measure is implemented as follows. Separately for each $i = 1, \dots, I$:

1. Generate M values ($m = 1, \dots, M$) from the posterior distribution of $\eta|X_{(-i)}$; call them $\eta^m_{(-i)}$, where $X_{(-i)}$ represents all the data without group i. Generating values of η will be accomplished in many cases through iterative simulation methods that will generate values of $\theta_{(-i)}$, where $\theta_{(-i)}$ is the collection of group parameters excluding group i: $f(\eta|X_{(-i)}) = \int f(\eta, \theta_{(-i)}|X_{(-i)})\, d\theta_{(-i)} \propto \int \pi(\eta)\pi(\theta_{(-i)}|\eta) f(X_{(-i)}|\theta_{(-i)})\, d\theta_{(-i)}$.
2. Generate values θ^m_i of θ_i given the hyperparameters $\eta^m_{(-i)}$ independently from $\pi(\theta_i|\eta^m_{(-i)})$, $m = 1, \dots, M$.
3. Generate replicate data X^m_i independently from $f(X_i|\theta^m_i)$, $m = 1, \dots, M$.
4. Compute the proportion of times out of M that $D(X^m_i, \theta^m_i, \eta^m_{(-i)})$ is greater than $D(x_i, \theta^m_i, \eta^m_{(-i)})$, $m = 1, \dots, M$.