A simulation-intensive approach for checking hierarchical models

Dipak K. Dey and Alan E. Gelfand
Department of Statistics, University of Connecticut
Storrs, CT, 06269, USA.

Tim B. Swartz
Department of Mathematics and Statistics at Simon Fraser University
Burnaby, B.C., V5A 1S6, Canada

PanteUs K. Vlachos
Department of Statistics at Carnegie Mellon University
Pittsburgh, PA 15213, U.S.A.

Abstract

Recent computational advances have made it feasible to fit hierarchical models in a wide range of serious applications. In the process, the question of model adequacy arises. While model checking usually addresses the entire model specification, model failures can occur at each hierarchical stage. Such failures include outliers, mean structure errors, dispersion misspecification, and inappropriate exchangeabilities. We propose an approach which is entirely simulation based. Given a model specification and a dataset, we need only be able to simulate draws from the resultant posterior. By replicating a posterior of interest using data obtained under the model we can “see” the extent of variability in such a posterior. Then, we can compare the posterior obtained under the observed data with this medley of posterior replicates to ascertain whether the former is in agreement with them and accordingly, whether it is plausible that the observed data came from the proposed model. Many such comparisons can be run, each focusing on a different potential model failure. Focusing on generalized linear mixed models, we explore the questions of when hierarchical model stages are separable and checkable and illustrate the approach with both real and simulated data.

Key Words: Discrepancy measures, generalized linear mixed model, Monte Carlo tests, sampling-based model fitting, stagewise model adequacy.


1 Introduction

Recent computational advances have made it feasible to fit hierarchical models in a wide range of serious applications. Hence, the problem of model
determination, i.e., model adequacy and model selection arises. Model choice has received considerable attention; for model adequacy much less has been said. In providing the probabilistic components of a hierarchical model, we rarely believe that any of the distributions are correct. Those specifications further removed from the data are often intentionally made less precise, not because we believe them to be correct but in order to permit the data to drive the inference. However, what is true is apart from model checking. If we undertake model criticism we must examine the adequacy of what is specified and we must assume proper priors (or else the observed data could not have arisen under the model). High dimensional models, e.g., those having more parameters than data points, as well as very vaguely specified hierarchical models will be difficult to criticize. Here we are interested in that middle range of models which are not so parsimonious as to prevent useful hierarchical modeling but are not so high dimensional as to render adequacy a nonissue.

A formal Bayesian model adequacy criterion (as in Box, 1980) proposes that the marginal density of the data be evaluated at the observations. Large values support the model, small values do not. Assessment of the magnitude of this value could be facilitated by standardizing, using the maximum value or an average value of this density (Berger, 1985). However, a high dimensional density ordinate will be difficult to estimate well and hopeless to calibrate. In addition, with hierarchical models, failures, such as outliers, mean structure errors, dispersion misspecifications and inappropriate exchangeabilities, can occur at each hierarchical stage. The formal procedure does not provide feedback regarding the adequacy of the stagewise specifications. Alternative, informal approaches for model criticism are needed. We briefly review what has been proposed.

Chaloner and Brant (1988), Chaloner (1994) and Weiss (1995), focusing on outlier detection, suggest posterior-prior comparison. Their strategy is to identify random variables whose distribution, a priori, is a standard one. In particular, they choose functions of so-called realized residuals. Given the data, the posterior distribution of each such function is obtained. If it differs considerably from its associated prior, using tail area comparison, a lack of model fit is claimed. For a realized residual, an outlying observation is asserted. If the entire model specification is correct, such comparisons will be successful on average but will fail to recognize the variability in the posterior.