International Journal of Approximate Reasoning 66 (2015) 53–72

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Inference after checking multiple Bayesian models for data conflict and applications to mitigating the influence of rejected priors

David R. Bickel

Ottawa Institute of Systems Biology, Department of Biochemistry, Microbiology, and Immunology, Department of Mathematics and Statistics, University of Ottawa, 451 Smyth Road, Ottawa, Ontario, K1H 8M5, Canada

ARTICLE INFO

Article history:
Received 7 January 2015
Received in revised form 29 July 2015
Accepted 30 July 2015
Available online 3 August 2015

Keywords:
Imprecise probability
Model assessment
Model checking
Model criticism
Prior-data conflict
Robust Bayesian decision theory

ABSTRACT

Two major approaches have developed within Bayesian statistics to address uncertainty in the prior distribution and in the rest of the model. First, methods of model checking, including those assessing prior-data conflict, determine whether the posterior resulting from the model is adequate for purposes of inference and estimation or other decision-making. A potential drawback of this approach is that it provides little guidance for inference in the event that the model is found to be inadequate, that is, in conflict with the data. Second, the robust Bayes approach determines the sensitivity of inferences and decisions to the prior distribution and other model assumptions. This approach includes rules for making decisions on the basis of a set of posterior distributions corresponding to the set of reasonable model assumptions. Drawbacks of the second approach include the inability to criticize the set of models and the lack of guidance for specifying such a set.
Those two approaches to model uncertainty are combined into a two-stage procedure in order to overcome each of their limitations. The first stage checks each model within a large class of models to assess which models are in conflict with the data and which are adequate for purposes of data analysis. The resulting set of adequate models is then used in the second stage either for summarizing a combined posterior such as a maximum-entropy posterior or for inference according to decision rules of the robust Bayes approach and of imprecise probability more generally. This proposed framework is illustrated by the application of a class of hierarchical models to a simple data set.

1. Introduction

In Bayesian statistics, a model, including some prior distribution, is considered inadequate for inference purposes if it is assessed to conflict with the observed data. Otherwise, the model is considered adequate, and inference proceeds according to Bayes's theorem. If a loss function is specified, the Bayes rule requires the minimization of the posterior expected loss to determine the result of a hypothesis test, parameter estimation, or other action taken on the basis of the model passing the check for data conflict [2, pp. 6–7].

While this approach guards against excessive reliance on a single model, it faces threats from two fronts, one corresponding to passing the model check and the other to failing it. The fact that a model passes a reasonable check is insufficient

E-mail address: dbickel@uottawa.ca.

http://dx.doi.org/10.1016/j.ijar.2015.07.012

to justify its use in inference and decision-making. That is seen from the fact that such assessments are conservative, each only checking for certain deviations from the data. A lack of evidence for a model's conflict with data does not warrant the sole use of that model. In fact, it is usually the case that many models would pass the same check for adequate agreement with the data. In general, there is no single best explanation of the data [88, §4.2].

However, if the model reflects the beliefs of a real or hypothetical agent, problems arise not when the model passes the assessment but rather when it fails. Indeed, when a model is found to be inadequate, there is no general prescription for how to proceed with data analysis [77]. The usual advice is to return to model-formulation, perhaps with a more careful elicitation of an expert's prior opinion. According to Lad [76, §6.6.4], while this process of learning may involve seeking understanding from an expert whose model passed the check, the formulation of a new model cannot be specified by any algorithm. Indeed, there is little algorithmic guidance at this point other than to modify or relax assumptions made by the original model. Caution is needed here since moving the model in the direction of the data introduces some bias due to again using the same data for going from the prior distribution under the new model to the posterior distribution. Even without consciously using the data twice, a mechanical iteration of the modeling-assessment routine until a model is found that passes the check eventually leads to some model in better agreement with the data than previous models. Cross validation and other methods that split the data into training and testing sets attenuate such reuse of the data but often at the expense of the ability to draw useful conclusions from a subset of the original data. Moreover, such methods lack the theoretical foundations that undergird coherent Bayesian statistics.

While it is not always practical to attempt to eliminate all double-use of the data, a more principled approach can keep it under control while at the same time avoiding undue reliance on the first model that passes the check. This is achieved in two stages:

1. Divide a broad class of possible models into a set of adequate models, those passing the assessment of data agreement, and a set of inadequate models, those failing the check. (This blurs the distinction between model checking and comparative model selection, a distinction that was already somewhat blurry [28, §2.4.2]; [62, §1.1].)
2. On the basis of the set of adequate models, report posterior probabilities or perform inference, estimation, and other actions by combining distributions or by using one of the algorithms that generalizes the Bayes rule that corresponds to the loss function.

In the spirit of model assessment, this solution does not require a prior distribution over the possible models. This contrasts with the *M-closed* model selection methods of traditional Bayesian model selection and Bayesian model averaging [13, §6.1.2] in that each procedure of this type relies on a specified prior distribution across the models considered (e.g., [80]) or on a set of such priors [34].

This paper develops the desired solution as follows. Section 2 will consider sets of models that are sufficiently close to the data according to the chosen criterion of model checking. In noted respects, this approach differs from a previous approach leading to sets of models generated by a non-Bayesian model selection criterion [27, pp. 169–171]. The proposed framework (Section 2.1) will be concretely presented in terms of the Bayes factor (Section 2.2.1) and the integrated likelihood ratio (Section 2.2.2) as illustrative measures of model adequacy. To make the approach clear, examples of sets of adequate models under various criteria and assumptions are provided (Section 2.3).

Data analysis may stop with Stage 1, reporting the set of posterior probabilities or probability distributions corresponding to the set of adequate models. Alternatively, data analysis may proceed to Stage 2s reduction of each set to a single posterior distribution and/or to decisions such as hypothesis tests and point estimates using methods originally developed for the use of sets of models that do not depend on the data (Section 3). In the former case, the resulting posterior distribution may be summarized by posterior probabilities, credible sets, etc. For the latter case, the robust Bayes literature has methods of determining optimal actions on the basis of sets of prior distributions in some neighborhood of the initial prior distribution (e.g., [14]). Likewise, the economics, operations research, and imprecise probability literatures offer decision rules for sets of models corresponding to partially ordered preferences (e.g., [56,47]), whereas the imprecise probability literature features decision rules for intervals of probability induced by differences between the prices of buying and selling a gamble (e.g., [105]). Since the sets of models used rarely depend on the data at hand, the sets are in principle beyond criticism in the light of new information. For example, sets of prior distributions are rarely determined empirically but are often updated into posterior distributions by Bayes's theorem for application of the decision rules that are designed for sets of probability distributions, as will be seen in Example 5. Thus, applying those decision rules to the results of Bayesian model assessment extends their domain and may broaden their appeal to the mainstream statistics community.

Section 4 applies the proposed approach to a modern method of testing multiple hypotheses, as illustrated by a well studied data set in Section 5. Finally, Appendix A covers alternative criteria of model assessment for Stage 1 and alternative inference rules for Stage 2, Appendix B offers connections to possibility theory, and Appendix C softens the threshold between adequate and inadequate models.