Statistical Science
2017, Vol. 32, No. 1, 41–43
DOI: 10.1214/16-STS607
Main article DOI: 10.1214/16-STS576
© Institute of Mathematical Statistics, 2017

Toward Automated Prior Choice

David B. Dunson

1. GENERAL THOUGHTS ON PRIOR CHOICE

There is a pressing need for more work providing general guidelines for prior choice in realistically complex Bayesian models for real world applications. I find that the rich literature on "objective Bayes" (O'Bayes) lacks useful suggestions, with too much focus on "flat" and noninformative priors, and on approaches designed to mimic "old school" (i.e., prior to the modern era of penalization) frequentist inferences. In practice, I find that it is almost always a bad idea to choose a noninformative or very high variance/diffuse prior in complex modeling settings. Such priors tend to only work well in very simple settings; for example, when the data contain ample information and the model under consideration is regular and contains a modest number of parameters.

In practice, priors that tend to have good performance in realistically complex models almost always favor some degree of shrinkage toward some notion of a low-dimensional structure. If the prior is overly vague and the data are potentially not very informative about certain model parameters, then instabilities can result computationally and Bayesian inferences can have relatively poor behavior (e.g., in a mean square prediction or estimation error sense). Although shrinkage is most famously important in high-dimensional low sample size data settings, it can lead to gains much more broadly. There is an increasingly vast literature proposing shrinkage priors that are targeted toward specific settings and do not require subjective elicitation of hyperparameters using domain knowledge. Although most of the focus (by far) has been on Gaussian linear regression and closely related modeling contexts, there is an increasing literature on more elaborate settings ranging from factor modeling of high-dimensional covariance matrices (e.g., Bhattacharya and Dunson, 2011) to analysis of many way contingency tables and high-dimensional categorical data Zhou et al., 2015.

David B. Dunson is Arts and Sciences Professor, Department of Statistical Science, Duke University, Durham, North Carolina 27708, USA (e-mail: dunson@duke.edu).

Much of my own research agenda focuses on designing new and better classes of priors for complex data and models, with a particular emphasis on high dimensional and object data settings. In our work, we often attempt to design priors that will lead to appealing frequentist properties, such as efficient rates of concentration of the posterior distribution in asymptotic regimes in which the dimension of the data increases with the sample size (refer, e.g., to Bhattacharya et al., 2015 and Zhou et al., 2015). In addition, a common theme is designing the prior in such a manner that a very small number of tuning parameters control the degree of shrinkage toward some simple structure (zero coefficient values, low rank factorization, etc.). However, often it can be complicated to choose such priors and validate their properties. Hence, it is appealing to have new prescriptive approaches that can help one to target design of new priors. Current thinking in the "pragmatic" Bayes community is that priors should be chosen to be (a) weakly informative in the sense of placing high probability on a wide range of plausible values while avoiding an overly-vague specification; (b) concentrated near some lower-dimensional structure (e.g., zero parameter values) while having heavy tails to be robust to deviations from this structure; (c) have a simple form favoring interpretation and computation. Of course, in practice it is often not clear how exactly to choose a prior having properties (a)–(c); although there are many widely used families that satisfy (a)–(c) in certain common classes of problems, it is typically not clear how to choose hyperpriors and best select from among the members of a family of priors. In addition, it is difficult to develop appropriate priors in classes of problems that have not been as widely studied; for example, outside of locally Gaussian and/or linear models.

2. PENALIZED COMPLEXITY PRIORS AND THE SIMPSON ET AL. APPROACH

The Simpson et al. article provides an important and thought-provoking contribution to the rich literature on penalized complexity (PC) priors. Many (most?) of the existing shrinkage priors in the Bayesian literature can also be said to penalize complexity in shrinking toward a simple baseline model structure. However, the