Statistical Science
2014, Vol. 29, No. 1, 91–94
DOI: 10.1214/13-STS460

Discussion of Big Bayes Stories and BayesBag

Peter Bühlmann

1. INTRODUCTORY REMARKS

I congratulate all the authors for their insightful papers with wide-ranging contributions. The articles demonstrate the power and elegance of the Bayesian inference paradigm. In particular, it allows to incorporate prior knowledge as well as hierarchical model building in a convincing way. Regarding the latter, the contribution by Raftery, Alkema and German is a very fascinating piece, as it addresses a set of problems of great public interest and presents predictions for the world populations and other interesting quantities with uncertainty regions. Their approach is based on a hierarchical model, taking various characteristics into account (e.g., fertility projections). It would have been very difficult to come up with a "better" solution which would be as clear in terms of interpretation (in contrast to a "black-box machine") and which would provide (model-based) uncertainties for the predictions into the future.

2. UNCERTAINTY, STABILITY AND BAGGING THE POSTERIOR

Many of the papers quantify in one or another form various notions of uncertainties. In the Bayesian framework, this is usually based on the posterior distribution. An old "debate" is how much the results are sensitive to the choice of the prior, and I believe that some reasonable sensitivity analysis can lead to much insight. The sensitivity with respect to "perturbed data" though is not easily captured by the Bayesian framework. In the context of prediction, Leo Breiman (Breiman, 1996a, 1996b) has pointed to issues of stability with respect to perturbations of the data, Bousquet and Elisseeff (2002) provide some mathematical connections to prediction performance while Meinshausen and Bühlmann (2010) present some theory and methodology for controlling the frequentist error of expected false positives.

Peter Bühlmann is Professor, Seminar for Statistics, ETH Zürich, CH-8092 Zürich, Switzerland (e-mail: buhlmann@stat.math.ethz.ch).

As an example, the (frequentist) Lasso (Tibshirani, 1996) is very unstable for estimating the unknown parameters in a linear model, in particular, if the correlation among the covariates is high (for two highly correlated variables where at least one of them has a substantially large regression coefficient, the Lasso selects either one or the other in an unstable fashion). Thus, the MAP for a Gaussian linear model with a Double-Exponential prior for the regression coefficients is unstable. The posterior distribution is probably more stable but, presumably, it is still "rather" sensitive with respect to perturbation of the data: if the data would look a bit different, the posterior might be "rather" different. The situation becomes more exposed to stability problems when using spike and slab priors (Mitchell and Beauchamp, 1988), due to increased sparsity.

We can stabilize the posterior distribution by using a bootstrap and aggregation scheme, in the spirit of bagging (Breiman, 1996b). In a nutshell, denote by $\mathcal{D}^*$ a bootstrap- or subsample of the data $\mathcal{D}$. The posterior of the random parameters θ given the data $\mathcal{D}$ has c.d.f. $F(\cdot|\mathcal{D})$, and we can stabilize this using

$$F_{\mathrm{BayesBag}}(\cdot|\mathcal{D}) = \mathbb{E}^*[F(\cdot|\mathcal{D}^*)],$$

where $\mathbb{E}^*$ is with respect to the bootstrap- or subsampling scheme. We call it the *BayesBag* estimator. It can be approximated by averaging over B posterior computations for bootstrap- or subsamples, which might be a rather demanding task (although say $B = 10$ would already stabilize to a certain extent). Note that when conditioning on the data, the posterior $F(\cdot|\mathcal{D})$ is a fixed c.d.f., but when taking the view point that the data could change, it is useful to consider randomized perturbed versions $F(\cdot|\mathcal{D}^*)$ which are to be aggregated.

The following simple and rather stable example shows that such a bagging scheme outputs a larger uncertainty which is perhaps more appropriate.

LOCATION MODEL WITH CONJUGATE GAUSSIAN PRIOR. Consider the model

$$\theta \sim \mathcal{N}(0, \tau^2),$$
$$\text{conditional on } \theta\text{:} \quad X_1, \ldots, X_n \text{ i.i.d.} \sim \mathcal{N}(\theta, \sigma^2).$$