Contributed Discussion

Matteo Iacopini[*,†] and Stefano Tonellato[‡]

We congratulate the authors for their excellent research which leaded to the development of a new statistical method for model comparison. The procedure is computationally fast and can be applied in a variety of settings ranging from mixture models to variable selection in regression frameworks.

Pseudo Bayesian model averaging and reference pseudo Bayesian model averaging

The authors mentioned the contribution by Li and Dunson (2016) as a possible alternative for weighting competing models. In this discussion, we present a small simulation study comparing the performance of the pseudo Bayesian model averaging (Pseudo-BMA) introduced in Section 3.4 with the performance of the reference pseudo Bayesian model averaging (Reference-Pseudo-BMA), mentioned in Section 2, and based on $\widetilde{KL}_2$, as in Li and Dunson (2016).

The data are generated from the following model: $Y_i|x_i \sim e^{-2x_i}\mathcal{N}(y|x,0.1) + (1 - e^{-2x_i})N(y|x^4,0.1),\ X_i \sim \mathcal{U}(0,1),\ i = 1,\dots,N$. We estimated $K = 5$ different linear regression models, where model M_k is defined as: $Y_i|x_i = \beta_k x^k + \varepsilon_i,\ \varepsilon_i \sim N(0,0.1)$, $\beta_k \sim N(0,10)$ and $\sigma \sim Ga(0.1,0.1)$.

Reference-Pseudo-BMA requires the preliminary estimation of the predictive density via a Bayesian nonparametric approach, which we computed by using a weight dependent Dirichlet process prior for the estimation of a fully nonparametric Bayesian density regression (Müller et al., 2013, ch. 4).

We run 100 simulations for each different value of the sample size in the grid $N \in \{5,10,20,30,40,50\}$ and for each N we computed the mean log-predictive densities of the two combination methods. The results are shown in Figure 1, which plots the posterior log-predictive density (averaged over simulations) for each value of the sample size, for the two cases. As expected, the Pseudo-BMA performs better than the Reference-Pseudo-BMA, and the difference of performance decreases with N.

For one of the previously run simulations (similar results were found in the other cases), Figure 2 shows the true conditional density $p(y|x)$ (red curve) at some fixed values of the covariate x together with the predictive densities provided by Pseudo-BMA (blue) and by Reference-Pseudo-BMA (black), respectively. The unsatisfactory approximation of the true predictive density is due to the inadequacy of the parametric

[*]Ca' Foscari University of Venice, Cannaregio 873, 30121, Venice, Italy
[†]Université Paris I – Panthéon-Sorbonne, 106-112 Boulevard de l'Hôpital, 75642 Paris Cedex 13, France, matteo.iacopini@unive.it
[‡]Ca' Foscari University of Venice, Cannaregio 873, 30121, Venice, Italy, stone@unive.it

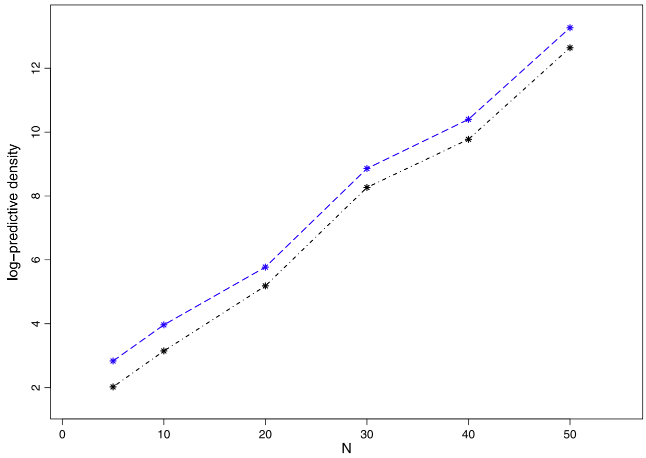

Figure 1: Posterior log-predictive density of model (a) (*blue*) and model (b) (*black*), for different values of the sample size $N \in \{5, 10, 20, 30, 40, 50\}$.

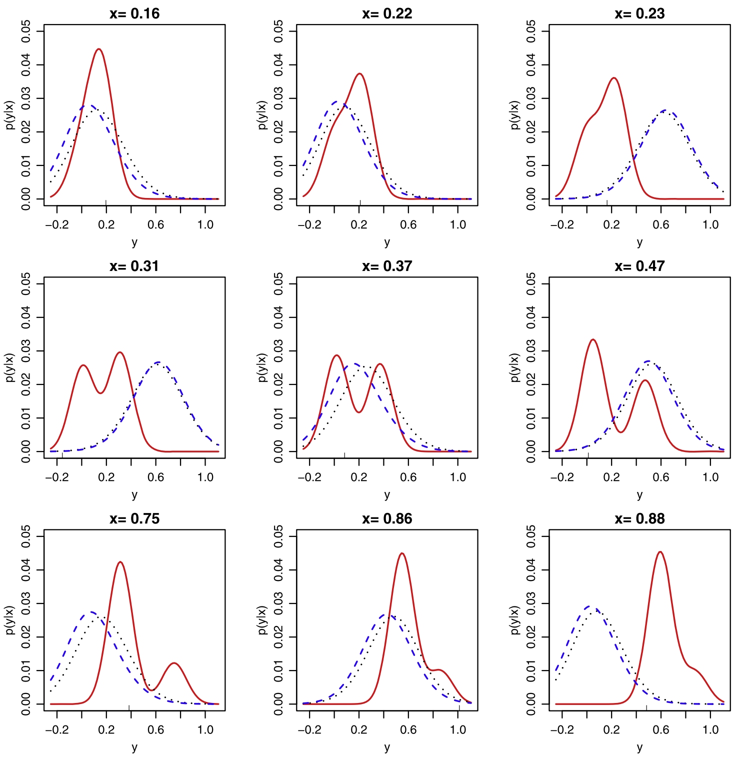

Figure 2: Conditional densities $p(y|x)$ for several values of x, $N = 20$. True function (*red*), model (b) estimate (*black*) and model (a) estimate (*blue*).

models under comparison, not to the methods used in order to produce stacking. What is interesting to notice is that despite the different weights computed according to the two schemes, the combined conditional predictive densities are rather similar for all the values of the conditioning variable x. This feature has been proved to hold also when the sample size increases and similar results (not reported here) have been obtained for different model specifications.

The main insight from this small simulation study is twofold: first, the approach proposed by the authors outperforms the alternative weighting schemes, both in fitting and in computational efficiency. Second, the scheme of Li and Dunson (2016) yields comparable results in terms of conditional density estimation. This might suggest that coupling Pseudo-BMA and Reference-Pseudo-BMA might be a successful strategy in those circomstances when leave-one-out or Pareto smoothed importance sampling leave-one-out cross-validation are suspected to produce unstable results, due to small sample size or large values of $\hat{k}$.

References

Li, M. and Dunson, D. B. (2016). "Comparing and weighting imperfect models using D-probabilities." 52, 54

Müller, P., Quintana, F. A., Jara, A., and Hanson, T. (2013). *Bayesian nonparametric data analysis*. Springer. MR3309338. doi: https://doi.org/10.1007/978-3-319-18968-0. 52