DISCUSSION OF "CATCHING UP FASTER BY SWITCHING SOONER: A PREDICTIVE APPROACH TO ADAPTIVE ESTIMATION WITH AN APPLICATION TO THE AKAIKE INFORMATION CRITERION – BAYESIAN INFORMATION CRITERION DILEMMA" BY VAN ERVEN ET AL.

XIN GAO AND HANNA JANKOWSKI

We would like to congratulate the authors on an important and thought-provoking paper. Existing Bayesian model selection methods implicitly assume a fixed prior over all competing models. However, as the authors point out, there exists a trend on the posterior probabilities as the sample size increases. This implies that we can learn from the data gradually and update our prior distribution with the sample size. By taking advantage of this behaviour, the authors solve the AIC-BIC dilemma, where the proposed method is not only selection consistent but also achieves minimum risk. From our point of view, the switch distribution is more in the spirit of the BIC, where inference is based on the posterior distributions of the models, while the improvement comes from fine tuning of the prior distribution. Bayesian-based model approaches are typically selection consistent, and therefore the proposed method converges to the true model in the long run. In the finite sample domain, however, the fine tuning of the prior makes model selection more adaptive to the data, resulting in optimal performance in terms of cumulative risk. This is a very significant breakthrough in the development of adaptive model selection procedures.

We have one concern regarding the interpretation of the obtained results. Suppose that we have observed a dataset of length n, and we have a fixed estimator p based on this dataset. Then, as is shown, the predictor achieves optimal convergence rates for the cumulative risk, $R(p^*, p, n)$. For the purposes of prediction, a practitioner would be interested in $r(p^*, p, n+1)$ whereas the quantity $R(p^*, p, n)$ can only provide some information on $r(p^*, p, k)$ for $k \leq n$ (for 'most' k). Here, we have used the notation of the paper. If the data has already been collected, wouldn't the practitioner be more interested in the instantaneous risk for the future rather than the cumulative and/or instantaneous risk for the past? If the data is not ready, will the proposed method provide any guidelines on how to determine the sample size, or how to balance between the goals of prediction and model selection into their data collection process?

Date: October 25, 2011.

We would ask the authors to provide further clarification and/or insight into this problem.

Department of Mathematics and Statistics, York University
e-mail: xingao@mathstat.yorku.ca, hkj@mathstat.yorku.ca