Bayes factors for high and infinite dimensional models

Steve MacEachern, Ohio State University

The traditional Bayesian approach to model comparison is based on the Bayes factor—often said to be the relative support *of the data* for one model in comparison to another. In fact, the Bayes factor can be heavily influenced by the prior distributions under the two models. This causes Bayesian model comparison to break down when a prior is improper; it creates difficulties when prior information is weak and an arbitrarily overdispersed prior distribution is used; the problems compound when one of the models is of high or infinite dimension, where the prior is, in many ways, arbitrary. To reduce the arbitrariness of the Bayes factor, we propose a novel criterion called the calibrated Bayes factor. The new criterion relies on training samples to yield a partial-posterior distribution that is calibrated to a fixed level of concentration. The calibrated Bayes factor is then computed as the Bayes factor over the remaining data. We motivate the method, explain it, illustrate its benefits through simulation in simple (one-sample) and more complex settings, and use it on obesity data from the Ohio Family Health Survey.