World Academy of Science, Engineering and Technology
International Journal of Mathematical and Computational Sciences
Vol:10, No:11, 2016

Performance and Limitations of Likelihood Based Information Criteria and Leave-One-Out Cross-Validation Approximation Methods

Authors : M. A. C. S. Sampath Fernando, James M. Curran, Renate Meyer

Abstract : Model assessment, in the Bayesian context, involves evaluation of the goodness-of-fit and the comparison of several alternative candidate models for predictive accuracy and improvements. In posterior predictive checks, the data simulated under the fitted model is compared with the actual data. Predictive model accuracy is estimated using information criteria such as the Akaike information criterion (AIC), the Bayesian information criterion (BIC), the Deviance information criterion (DIC), and the Watanabe-Akaike information criterion (WAIC). The goal of an information criterion is to obtain an unbiased measure of out-of-sample prediction error. Since posterior checks use the data twice; once for model estimation and once for testing, a bias correction which penalises the model complexity is incorporated in these criteria. Cross-validation (CV) is another method used for examining out-of-sample prediction accuracy. Leave-one-out cross-validation (LOO-CV) is the most computationally expensive variant among the other CV methods, as it fits as many models as the number of observations. Importance sampling (IS), truncated importance sampling (TIS) and Pareto-smoothed importance sampling (PSIS) are generally used as approximations to the exact LOO-CV and utilise the existing MCMC results avoiding expensive computational issues. The reciprocals of the predictive densities calculated over posterior draws for each observation are treated as the raw importance weights. These are in turn used to calculate the approximate LOO-CV of the observation as a weighted average of posterior densities. In IS-LOO, the raw weights are directly used. In contrast, the larger weights are replaced by their modified truncated weights in calculating TIS-LOO and PSIS-LOO. Although, information criteria and LOO-CV are unable to reflect the goodness-of-fit in absolute sense, the differences can be used to measure the relative performance of the models of interest. However, the use of these measures is only valid under specific circumstances. This study has developed 11 models using normal, log-normal, gamma, and student's t distributions to improve the PCR stutter prediction with forensic data. These models are comprised of four with profile-wide variances, four with locus specific variances, and three which are two-component mixture models. The mean stutter ratio in each model is modeled as a locus specific simple linear regression against a feature of the alleles under study known as the longest uninterrupted sequence (LUS). The use of AIC, BIC, DIC, and WAIC in model comparison has some practical limitations. Even though, IS-LOO, TIS-LOO, and PSIS-LOO are considered to be approximations of the exact LOO-CV, the study observed some drastic deviations in the results. However, there are some interesting relationships among the logarithms of pointwise predictive densities (lppd) calculated under WAIC and the LOO approximation methods. The estimated overall lppd is a relative measure that reflects the overall goodness-of-fit of the model. Parallel log-likelihood profiles for the models conditional on equal posterior variances in lppds were observed. This study illustrates the limitations of the information criteria in practical model comparison problems. In addition, the relationships among LOO-CV approximation methods and WAIC with their limitations are discussed. Finally, useful recommendations that may help in practical model comparisons with these methods are provided.

Keywords : cross-validation, importance sampling, information criteria, predictive accuracy
Conference Title : ICSA 2016 : International Conference on Statistics and Analysis
Conference Location : Singapore, Singapore
Conference Dates : November 21-22, 2016