World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:17, No:04, 2023

Confidence Envelopes for Parametric Model Selection Inference and Post-Model Selection Inference

Authors : I. M. L. Nadeesha Jayaweera, Adao Alex Trindade

Abstract : In choosing a candidate model in likelihood-based modeling via an information criterion, the practitioner is often faced with the difficult task of deciding just how far up the ranked list to look. Motivated by this pragmatic necessity, we construct an uncertainty band for a generalized (model selection) information criterion (GIC), defined as a criterion for which the limit in probability is identical to that of the normalized log-likelihood. This includes common special cases such as AIC & BIC. The method starts from the asymptotic normality of the GIC for the joint distribution of the candidate models in an independent and identically distributed (IID) data framework and proceeds by deriving the (asymptotically) exact distribution of the minimum. The calculation of an upper quantile for its distribution then involves the computation of multivariate Gaussian integrals, which is amenable to efficient implementation via the R package "mvtnorm". The performance of the methodology is tested on simulated data by checking the coverage probability of nominal upper quantiles and compared to the bootstrap. Both methods give coverages close to nominal for large samples, but the bootstrap is two orders of magnitude slower. The methodology is subsequently extended to two other commonly used model structures: regression and time series. In the regression case, we derive the corresponding asymptotically exact distribution of the minimum GIC invoking Lindeberg-Feller type conditions for triangular arrays and are thus able to similarly calculate upper quantiles for its distribution via multivariate Gaussian integration. The bootstrap once again provides a default competing procedure, and we find that similar comparison performance metrics hold as for the IID case. The time series case is complicated by far more intricate asymptotic regime for the joint distribution of the model GIC statistics. Under a Gaussian likelihood, the default in most packages, one needs to derive the limiting distribution of a normalized quadratic form for a realization from a stationary series. Under conditions on the process satisfied by ARMA models, a multivariate normal limit is once again achieved. The bootstrap can, however, be employed for its computation, whence we are once again in the multivariate Gaussian integration paradigm for upper quantile evaluation. Comparisons of this bootstrap-aided semi-exact method with the full-blown bootstrap once again reveal a similar performance but faster computation speeds. One of the most difficult problems in contemporary statistical methodological research is to be able to account for the extra variability introduced by model selection uncertainty, the so-called post-model selection inference (PMSI). We explore ways in which the GIC uncertainty band can be inverted to make inferences on the parameters. This is being attempted in the IID case by pivoting the CDF of the asymptotically exact distribution of the minimum GIC. For inference one parameter at a time and a small number of candidate models, this works well, whence the attained PMSI confidence intervals are wider than the MLE-based Wald, as expected.

Keywords : model selection inference, generalized information criteria, post model selection, Asymptotic Theory
Conference Title : ICASDA 2023 : International Conference on Applied Statistics and Data Analytics
Conference Location : Boston, United States
Conference Dates : April 17-18, 2023