Statistical Theory and Related Fields

ISSN: 2475-4269 (Print) 2475-4277 (Online) Journal homepage: https://www.tandfonline.com/loi/tstf20

Discussion on Prior-based Bayesian Information Criterion (PBIC) by M. J. Bayarri, James O. Berger, Woncheol Jang, Surajit Ray, Luis R. Pericchi, and Ingmar Visser

Ruobin Gong & Minge Xie

STATISTICAL THEORY AND RELATED FIELDS
2019, VOL. 3, NO. 1, 35–36
https://doi.org/10.1080/24754269.2019.1611146

Discussion on Prior-based Bayesian Information Criterion (PBIC) by M. J. Bayarri, James O. Berger, Woncheol Jang, Surajit Ray, Luis R. Pericchi, and Ingmar Visser

Ruobin Gong and Minge Xie

Department of Statistics, Rutgers, The State University of New Jersey, 501 Hill Center, Busch Campus, 110 Frelinghuysen Rd, Piscataway, NJ 08854

ARTICLE HISTORY Received 18 April 2019; Revised 20 April 2019; Accepted 22 April 2019

It is our pleasure to comment on this very interesting article on model selection. The Bayesian Information Criterion (BIC) is one of the most popular metrics for model selection. It is taught in every classroom on statistical modelling, and widely implemented as part of a standard routine in statistical programming languages. The fact that BIC is so popular means it is often erroneously applied to model classes that are beyond its design. In this elucidating paper, the authors unpacked a dangerous complication when one takes the classic BIC verbatim as an approximation to the marginal likelihood. An additive constant c, ignored in the derivation of BIC, is in fact model dependent. When two BICs are computed on different models and their difference or ratio used for comparison, their respective constants are also implicitly factored into the comparison.

We think the contribution of this paper is threefold. First and foremost, to address the bias induced into the comparison of classic BIC, the authors proposed the Prior-based Bayesian Information Criterion (PBIC) as a principled correction. As discussed in the paper, the original definition of BIC can be interpreted as an approximation to the marginal likelihood corresponding to unit information priors for each model, centred at the model likelihood. The proposed metrics PBIC and PBIC* achieve 'debiasing' of BIC through modifying this prior to a different, data-independent class. With the correction, model comparisons can again be made on a fair ground. Second, through the process of constructing PBIC, the authors also supplied an informative discussion around the concept of 'effective sample size'. While the precise definition is left as an open area for future research, the metrics constructed through orthogonalised parameters and their corresponding effective sample sizes exhibit intuitive appeal, strengthening the proposal's logical soundness. Thirdly, it is commendable that the proposed method relies minimally on advanced computational techniques. The ease of computation in a large part contributed to the popularity of BIC. Much like BIC, all that is needed to compute PBIC and PBIC* are the MLE and observable information. This allows for the possibility of turning PBIC and PBIC* into an off-the-shelf routine.

In reality, the value of BIC is often taken literally for model selection. Practitioners who employ PBIC and PBIC* will likely rely on them in a similar way. However, since both metrics are *prior-based* in a proper sense, what kind of uncertainty measurements do they come with when viewed as Bayesian estimates to the marginal likelihood? We thus pose the following question: how can the proposed framework of PBIC and PBIC* be augmented to accommodate uncertainty quantification on model selection? Specifically, for users who might rely on ΔPBIC or ΔPBIC* as the sole decision criterion, can we provide a sense of measurable variability associated with the respective quantities – anything from a full distributional description to an approximate interval?

Below we discuss two potential directions to pursue this question. One natural route to take is a multilevel analysis on the class of robust priors that PBIC and PBIC* call for. One may impose a hyperprior on π^R in some way, such as through its tuning parameters. Robust Bayesian prior neighbourhoods present another viable option. The density ratio neighbourhood of π^R in relation to the Cauchy(0, b) prior is a promising choice. Other neighbourhoods that lie outside of the proposed family can be explored. For example, mixture priors or contamination neighbourhoods involving the original data-dependent BIC prior can be of interest. A somewhat different route exploits the probabilistic nature of the effective sample size n_j^e. If the model observed information is data-dependent, or if the expected information is a function of un-modelled predictors (such as the linear regression example), the specification of the effective sample size will be data-dependent as well. The sampling variability of the data can thus be harnessed to reflect uncertainty in the

CONTACT Minge Xie mxie@stat.rutgers.edu

specification of the robust prior and PBIC. Doing so would rely on a certain bootstrap procedure of the observed data, or if possible, on deriving the sampling distribution of the observed information (hence effective sample size) for each given model under contemplation. The latter amounts to a parametric version of the bootstrap. That being said, either of these can be challenging if the goal is to also maintain the computational convenience that PBIC and PBIC* have to offer.

Model comparison and selection are quintessential statistical inference problems, and like all others, they should be accompanied by proper uncertainty quantification. A rich description of uncertainty can be particularly valuable if the discernment task is not confined to two models but many, such as for variable selection in large dimensional datasets. Current practices of model selection often rely on a total order created by a metric, such as BIC, which itself is a random phenomenon. Recent literature saw efforts to construct confidence sets in the model space, e.g. Ferrari and Yang (2015) and Zheng, Ferrari, and Yang (2017). We think that a thoroughly constructed, objective Bayesian-motivated criterion such as PBIC and PBIC* can be a great starting point for developing a Bayesian approach to uncertainty quantification in model selection. The tradeoff between the size of the credible set in the model space and its associated credibility may be left to the practitioners. In our view, that is more truthful to the spirit of Bayesian reasoning.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the National Science Foundation [1513483, 1737857, 1812048].

Notes on contributors

Ruobin Gong is Assistant Professor of Statistics with the Department of Statistics, Rutgers University. Her research interests lie at the foundations of uncertainty reasoning, theory and methods of statistical inference with random sets and imprecise probability, and computation and applications of Bayesian and generalized Bayesian modeling to robustness and privacy statistics. Ruobin obtained her Ph.D. in statistics from Harvard University, and B.Sc. in cognitive psychology from the University of Toronto.

Min-ge Xie is Distinguished Professor of Statistics with the Department of Statistics, Rutgers University and Director of Rutgers Office of Statistical Consulting. His main research interest lies in bridging the foundations of statistical inference and developing new statistical methodologies and theories for problems stemming from interdisciplinary research. His research interests include statistical applications in biomedical sciences, social sciences, industry, engineering, and environmental sciences. He received the B.S. degree in mathematics with a high honor from the University of Science and Technology of China (USTC) and the M.S. and Ph.D. degrees in statistics from University of Illinois at Urbana-Champaign (UIUC). He has conducted research projects sponsored by grants from the National Science Foundation (NSF), the National Institute of Health (NIH), The Department of Veterans Affairs (VA), the Federal Aviation Administration, among others.

ORCID

Ruobin Gong http://orcid.org/0000-0003-2965-9266

References

Ferrari, D., & Yang, Y. (2015). Confidence sets for model selection by f-testing. *Statistica Sinica*, *25*, 1637–1658.

Zheng, C., Ferrari, D., & Yang, Y. (2017). Model selection confidence sets by likelihood ratio testing. Preprint arXiv:1709.04342.