



Discussion on prior-based Bayes information criterion

Brunero Liseo

To cite this article: Brunero Liseo (2019) Discussion on prior-based Bayes information criterion, *Statistical Theory and Related Fields*, 3:1, 22-23, DOI: [10.1080/24754269.2019.1583630](https://doi.org/10.1080/24754269.2019.1583630)

To link to this article: <https://doi.org/10.1080/24754269.2019.1583630>



Published online: 04 Mar 2019.



Submit your article to this journal [↗](#)



Article views: 21



View related articles [↗](#)



View Crossmark data [↗](#)

Discussion on prior-based Bayes information criterion

Brunero Liseo

Dipartimento MEMOTEF, Sapienza Università di Roma, Rome, Italy

ARTICLE HISTORY Received 12 February 2019; Revised 12 February 2019; Accepted 13 February 2019

1. Introduction

It is a pleasure to discuss this important contribution to the methodology of objective Bayesian model selection.

It is not the first time that Jim Berger and his collaborators consider a very popular statistical tool (see, e.g., Bayarri & Berger, 2000; Berger & Sellke, 1987), very much used in applications, and included in standard statistical software, and criticise it at its deep roots and, that is crucial, propose a valid alternative.

Bayesian model selection methods suffer from the frustrating fact that improper priors cannot be used. As a consequence, most objective Bayes approaches have not experienced a broad development in practice. What is absolutely lacking is a sort of “reference”, or conventional, *off-the-shelf* answer, easy to compute, at least for standard and most popular statistical models.

This vacant place has been occupied, increasingly in the last decades, by the Bayesian Information Criterion, aka *BIC* (Schwarz, 1978). As illustrated in the paper, *BIC* may be considered an asymptotic approximation to a function of the marginal distribution of the observed data. However, its behaviour can be safely considered good, only under very restrictive assumptions, which basically confine ourselves to the standard, non-hierarchical, *i.i.d.* case, with n observation of fixed dimension p . Also, the asymptotic approximation practically ignores the contribution of the prior distribution on the parameters of the model. On the one hand, that makes *BIC*, notwithstanding the name, a really non-Bayesian proposal; on the other hand, it is easy to compute; it is often automatically computed by statistical software and then it is very popular among classical statisticians. In a word, we could say it can be considered the best non-Bayesian method for model comparison. However, it is today over-used even in situations where the underlying asymptotic expansion is not valid, or at least, should be carefully considered. The Authors provide a list of specific problems that can arise with a blind use of *BIC* in typical applications.

The new proposal is named prior-based Bayes information criterion (*PBIC*) in order to stress its main characteristic, that is, it can be considered a real Bayesian

criterion, since, in this case, an approximation of the likelihood is integrated with respect to a suitable proper ‘objective prior’, in order to obtain a valid objective Bayesian estimate of the marginal density of the observed data.

Apart from some technical details, the crucial steps in the derivation of the *PBIC* are essentially the following. For each model at hand, after splitting the entire parameter vector as $\theta = (\theta_1, \theta_2)$, where θ_2 is the set of parameters common to all models under comparison, and after a reparametrisation step which transforms θ_1 into $\xi = O\theta_1$, where the components of ξ are somehow orthogonalised, the Authors make the crucial choices, namely

- the choice of a uniform flat prior over θ_2 ;
- the choice of a suitable Cauchy-like prior for each component of ξ , within each model;
- the determination of the scale of the prior for component of ξ , within each model, in terms of their respective *effective sample sizes*.

The first choice is not discussed by the Authors: I believe this is a reasonable choice as long as the components of θ_2 are approximately location parameters. I do not see compelling reasons, other than computational, to use a uniform prior for non-location common parameters. This poses a sort of constraint on what kind of reparametrisation should be used.

The proposal of the objective prior for θ_1 is certainly one of the most interesting issues in the paper. It is amazing that the proposed scale of Normal priors is so similar to a Cauchy prior although it is not clear why the mixing prior on λ is a Beta(0.5, 1). Should any Beta(a , 1) work as well? Probably not, because we would lose the closed form expression. However, how to justify the shift from a Gamma mixing prior with mean 1 and variance 2, to a Beta prior with mean 1/3 and variance 4/45 is not clear. From another perspective, rewriting the variance of the Gaussian density in (9) as

$$\psi = \frac{\lambda}{\frac{d+b}{2} - d\lambda},$$

the prior on the mixing quantity ψ becomes

$$\pi(\psi) \propto \frac{1}{\psi^{1/2} \left(1 + \frac{d+b}{2}\psi\right)^{3/2}}$$

with support $(0, 2/(b-d))$. This depends on d , i.e. the effective sample size, but in any case, the mixing prior is dramatically different from the usual Gamma.

The other important issue discussed in the proposal is the determination of the effective sample size, already introduced, at least for linear models, in Bayarri, Berger, Pericchi (2014). The proposal is absolutely reasonable and convincing, and I have no suggestions to improve it. Did the Authors compare the behaviour of *TESS* and the new proposal in the linear models case?

Finally, I wonder whether it is possible, in some particular models, that the effective sample size, as defined in §3.2, could depend on the unknown parameters.

As a final remark, it is important to stress that the proposed modification of *BIC*, does not require independent observations, as long as the approximation in (7) can be used. This fact may broaden the use of such approximation significantly.

Disclosure statement

No potential conflict of interest was reported by the author.

References

- Bayarri, M. J., & Berger, J. O. (2000). *p* values for composite null models (with discussion). *Journal of the American Statistical Association*, 95, 1127–1142.
- Bayarri, M. J., Berger, J. O., & Pericchi, L. R. (2014). The effective sample size. *Econometric Reviews*, 3, 197–217.
- Berger, J. O., & Sellke, T. (1987). Testing a point null hypothesis: The irreconcilability of *p*-values and evidence. *Journal of the American Statistical Association*, 82, 112–122.
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6, 461–464.