On the construction of Bayesian based adaptive confidence sets

Botond Szabó[1], Aad van der Vaart[2], Harry van Zanten[3]

Second Bayesian Young Statisticians Meeting (BAYSM 2014)
Vienna, September 18–19, 2014

[1] Budapest University of Technology and Economics, Budapest, Hungary
bszabo@math.bme.hu

[2] Leiden University, Leiden, Netherlands
a.w.van.der.vaart@umail.leidenuniv.nl

[2] University of Amsterdam, Amsterdam, Netherlands
hvzanten@uva.nl

Abstract

We investigate the frequentist properties of Bayesian credible sets in the Gaussian white noise model, where the truth is assumed to have regularity $\beta \in [B, 2B]$. First we show that the classical adaptive Bayesian techniques, i.e. the hierarchical Bayes and the marginal likelihood empirical Bayes, do not provide trustworthy confidence sets. Then we introduce a new empirical Bayes method based on risk estimation which has optimal frequentist behaviour in the sense that it provides rate adaptive confidence sets with good coverage.
Keywords: Uncertainty quantification; credible sets; coverage; adaptation.

1 Introduction

In statistical applications confidence sets play a highly important role. However, the construction of confidence sets might be problematic do to lack of theoretical results or computational complexity, specially in nonparametric models.

Bayesian methods offer an alternative method for interval estimation. The Bayesian credible set is the set accumulating a large fraction (typically 95%) of the posterior mass. These sets can have various forms and are intended to visualize the remaining uncertainty in the estimation. Due to the heavy computational machinery for Bayesian methods the construction of credible sets is more feasible. However the frequentist interpretation of these sets are rather unclear. It is not know in general whether one can use credible sets as confidence sets or by doing so one gets missleading uncertainty quantification.

In our work we investigate the frequentist coverage of adaptive Bayesian credible sets in the Gaussian white noise model. First we show that the classical adaptive Bayesian methods, i.e. the hierarchical Bayes (HB) and the marginal likelihood empirical Bayes (MLEB) methods provide overconfident, haphazard uncertainty quantification. Then we introduce a new empirical Bayes method based on risk estimation, which has optimal frequentist properties.

2 Results

In our work we considered the Gaussian white noise model

$$X_i = \theta_{0,i} + \frac{1}{\sqrt{n}} Z_i, \quad i = 1, ...,$$

where $X = (X_1, ..)$ denotes the noisy observation, Z_i are iid standard normal random variables and $\theta_0 = (\theta_{0,1}, ...)$ is the unknown infinite dimensional parameter of interest. We assume furthermore that the true parameter θ_0 belongs to the Sobollev ball $S^\beta(M) = \{\theta \in \ell_2 : \sum_{i=1}^\infty \theta_i^2 i^{2\beta} < M\}$ for some (unknown) regularity parameter $\beta \in [B, 2B]$ (with known boundary parameter B) and (known) radius parameter $M > 0$.

In the Bayesian approach as a first step we endow θ_0 with a prior distribution

$$\Pi_\alpha(\cdot) = \prod_{i=1}^\infty N(0, i^{-1-2\alpha}),$$

where $\alpha > 0$ is the regularity hyper-parameter of the prior. The optimal choice of the hyper-parameter α highly depends on the unknown regularity parameter β, hence one has to use a data driven approach in the choice.

2.1 Classical adaptive techniques

The two, probably most well known method for choosing the hyper-parameter α are the MLEB and HB methods. In the HB method the unknown hyper-parameter is endowed first with a hyper-prior $\lambda(\alpha)$ and in the Bayesian analysis this two level hierarchical prior is applied

$$\Pi(\cdot) = \int_B^{2B} \lambda(\alpha)\Pi_\alpha(\cdot) d\alpha.$$

In contrast to this in the MLEB method the hyper-parameter is estimated by the maximum marginal likelihood estimator $\hat{\alpha}_n$ and then this estimator is plugged in into the posterior distribution:

$$\Pi_{\hat{\alpha}_n}(\cdot|X) = \Pi_\alpha(\cdot|X)\Big|_{\alpha=\hat{\alpha}_n}.$$

Natural credible sets are balls centered around the posterior mean $\hat{\theta}_n^M$ and $\hat{\theta}_n^H$ with radius $r_{n,\gamma}^M$ and $r_{n,\gamma}^H$, for the MLEB and HB methods, respectively, given as

$$\Pi_{\hat{\alpha}_n}(\theta : \|\theta - \hat{\theta}_n^M\|_2 \le r_{n,\gamma}^M | X) = 1 - \gamma \quad \text{and} \quad \Pi(\theta : \|\theta - \hat{\theta}_n^H\|_2 \le r_{n,\gamma}^H | X) = 1 - \gamma.$$

Then we introduce some additional flexibility by letting the balls to be blown up by a constant factor $L > 0$:

$$C_n^M(L) = \{\theta : \|\theta - \hat{\theta}_n^M\|_2 \le L r_{n,\gamma}^M\} \quad \text{and} \quad C_n^H(L) = \{\theta : \|\theta - \hat{\theta}_n^H\|_2 \le L r_{n,\gamma}^H\}. \qquad (1)$$

Theorem 1 (Theorem 1 and 2 of [14]) *For every $\beta \in [B, 2B]$ and $M > 0$ there exists a $\theta_0 \in S^\beta(M)$ and a subsequence $n_j \to \infty$ such that for every $L > 0$*

$$P_{\theta_0}(\theta_0 \in C_{n_j}^M(L)) \to 0 \quad \textit{and} \quad P_{\theta_0}(\theta_0 \in C_{n_j}^H(L)) \to 0.$$

The above negative results are due to the fact that the MLEB and HB methods are based on the marginal likelihood function, while to have good coverage and optimal size for the credible sets a correct bias-variance trade-off is required, which is not directly related to the likelihood.

2.2 New empirical Bayes procedure

We propose a new empirical Bayes method which corrects the weaknesses of the classical adaptive Bayesian techniques and achieves the frequentist limitations. Our estimator $\tilde{\alpha}_n$ instead of maximizing the marginal likelihood function balances out the bias and variance terms:

$$\tilde{\alpha}_n = \inf\{\alpha \geq B : \hat{B}_{n,k_n}(\alpha) \geq n^{-\alpha/(1+2\alpha)}\} \wedge 2B,$$

where $n^{-\alpha/(1+2\alpha)}$ is the variance of the posterior distribution for fixed hyper-parameter α and $\hat{B}_{n,k_n}(\alpha)$ is an estimator of the bias

$$\hat{B}^2_{n,k_n}(\alpha) = \sum_{i=1}^{n^{1/(1/2+2B)}} \frac{i^{2+4\alpha}}{(i^{1+2\alpha}+n)^2}(X_i^2 - \frac{1}{n}).$$

The construction of the credible sets $C_n^R(L)$ is done in the same way as in (1) with $\hat{\alpha}_n$ replaced by $\tilde{\alpha}_n$. The next theorem states that the so constructed credible sets have rate adaptive size and good coverage properties.

Theorem 2 (Theorem 3 of [14]) *There exists a large enough constant $L > 0$ such that*

$$\inf_{\theta_0 \in \cup_{\beta\in[B,2B]} S^\beta(M)} P_{\theta_0}(\theta_0 \in C_n^R(L)) \geq 1-\gamma.$$

Furthermore for every $\beta \in [B, 2B]$ there exists a constant $K(\beta) > 0$ such that

$$\inf_{\theta \in S^\beta(M)} P_{\theta_0}(\|C_n^R(L)\|_2 \leq K(\beta) n^{-\beta/(1+2\beta)}) \geq 1-\gamma.$$

3 Discussion

In nonparametric models the construction of confidence sets with optimal size and good coverage is not possible in general; see for instance [7], [4], [6]. However, by assuming in the Gaussian white noise model that $\beta \in [B, 2B]$ (for some fixed $B > 0$) the construction of adaptive confidence sets is possible [11], [3].

For general regularity parameter β the construction of adaptive confidence sets is not possible, some additional restriction has to be introduced. In the frequentist literature a well studied, natural assumption is the self-similarity condition, under which the construction of adaptive confidence sets is possible; see [9], [5], [2], [8]. Assuming self-similarity in [13] it was shown that the MLEB credible sets have good frequentist properties (optimal size and good coverage).

Follow up and related results on the coverage of adaptive credible sets appeared recently. In the working paper [12] the coverage properties of the rescaled Brownian motion is studied in the nonparametric regression model. Oracle type of results for credible sets were derived for the Gaussian white noise model in [1]. An adaptive nonparametric Bernstein-von Mises theorem under self-similarity constraints was given in [10].

References

[1] Belitser, E. (2014). "On coverage and oracle radial rate of DDM-credible sets under excessive bias restriction." *ArXiv e-prints 1407.5232*

[2] Bull, A.(2012). “Honest adaptive confidence bands and self-similar functions.” *Electronic Journal of Statistics*, **6**, 1490–1516.

[3] Bull, A., Nickl, R. (2013). “Adaptive confidence sets in L^2.” *Probability Theory and Related Fields*, **156**, 889–919.

[4] Cai, T.T. and Low, M.G.(2004). “An adaptation theory for nonparametric confidence intervals.” *The Annals of Statistics*, **32**(5), 1805–1840.

[5] Cai, T.T. and Low, M.G.(2010). “Confidence bands in density estimation.” *The Annals of Statistics*, **38**(2), 1122–1170.

[6] Juditsky, A. and Lambert-Lacroix, S. (2003). “On nonparametric confidence set estimation.” *Mathematical Methods of Statistics*, **19**(4), 410–428.

[7] Low, M. G.(1997). “On nonparamteric confidence intervals.” *The Annals of Statistics*, **25**(6), 2547–2554.

[8] Nickl, R., and Szabó, B (2014). “A sharp adaptive confidence ball for self-similar functions” *ArXiv e-prints 1406.3994.*

[9] Picard, D. and Tribouley, K. (2000). “Adaptive confidence interval for pointwise curve estimation.” *The Annals of Statistics*, **28**(1), 298–335.

[10] Ray, K. (2014). “Bernstein-von Mises theorems for adaptive Bayesian nonparametric procedures.” *ArXiv e-prints 1407.3397.*

[11] Robins, J. and van der Vaart, A. (2006). “Adaptive nonparametric confidence sets.” *The Annals of Statistics*, **34**(1), 229–253.

[12] Sniekers, S. and van der Vaart, A. (2014+). “Bayesian credible sets in the fixed design model for polished tail functions.” *Working paper.*

[13] Szabó, B., van der Vaart, A. and van Zanten, H. (2013). “Frequentist coverage of adaptive nonparametric Bayesian credible sets.” *ArXiv e-prints 1310.4489.*

[14] Szabó, B., van der Vaart, A. and van Zanten, H. (2014). “Honest Bayesian confidence sets for the L_2-norm.” *Journal of Statistical Planning and Inference*, http://dx.doi.org/10.1016/j.jspi.2014.06.005.