数理解析研究所講究録
第 1703 巻 2010 年 134-137

A Bayes factor with reasonable model selection consistency for ANOVA model

丸山 祐造
YUZO MARUYAMA*
東京大学・空間情報科学研究センター
CENTER FOR SPATIAL INFORMATION SCIENCE, THE UNIVERSITY OF TOKYO

We start with a simple one-way balanced ANalysis-Of-VAriance (ANOVA). There are two possible models. In one model, all random variables have the same mean. In the other model, random variables in each level has a different mean. Formally, the independent observations y_{ij} $(i=1,\ldots,p,\ j=1,\ldots,r,\ n=pr)$ are assumed to arise from the linear model:

$$y_{ij}=\mu+\alpha_i+\epsilon_{ij}, \quad \epsilon_{ij}\sim N(0,\sigma^2) \tag{1}$$

where μ, α_i $(i=1,\ldots,p)$ and σ^2 are unknown. We assume $\sum\alpha_i=0$ as uniqueness constraint. Clearly two models are written as follows:

$$\begin{aligned} &\mathcal{M}_1: \ \boldsymbol{\alpha}=(\alpha_1,\ldots,\alpha_p)'=\mathbf{0} \\ \text{vs } &\mathcal{M}_{A+1}: \ \boldsymbol{\alpha}\in\{\boldsymbol{a}\in\mathcal{R}^p|\boldsymbol{a}\neq\mathbf{0},\ \boldsymbol{a}'\mathbf{1}_p=0\}. \end{aligned} \tag{2}$$

In (2), A means the name of the factor and the subscript $A+1$ is from the fact that $E[y_{ij}]$ in (1) consists of the sum of the constant term and the level of the factor.

In this paper, we will consider Bayesian model selection based on Bayes factor for ANOVA problem. Model comparison, which refers to using the data in order to decide on the plausibility of two or more competing models, is a common problem in modern statistical science. In the Bayesian framework, the approach for model selection and hypothesis testing is essentially same, whereas there is a big difference in classical frequentist procedures for model selection and hypothesis testing. A natural approach is to use Bayes factor (ratio of marginal densities of two models), which is based on the posterior model probabilities (Kass and Raftery (1995)). That is the reason why we take Bayesian approach based on Bayes factor in this paper.

One of the most important topic on Bayesian model selection is consistency. Consistency means that the true model will be chosen if enough data are observed, assuming

*maruyama@csis.u-tokyo.ac.jp

that one of the competing models is true. It is well-known that BIC by Schwarz (1978) has consistency in classical (so called "$n > p$") situation. As a variant of "$p > n$" problem, which is hot in modern statistics, the consistency in the case where $p \to \infty$ and r is fixed in one-way ANOVA setup, has been considered by Stone (1979) and Berger et al. (2003). In the following, "CASE I" and "CASE II" denote the cases where

I. r goes to infinity and p is fixed,

II. p goes to infinity and r is fixed,

respectively. Under known σ^2 and CASE II, Stone (1979) showed that BIC always chooses the null model $\mathcal{M}_1$ (that is, BIC is not consistent under $\mathcal{M}_{A+1}$) even if $\boldsymbol{\alpha}'\boldsymbol{\alpha}/\{p\sigma^2\}$ is sufficiently large. This is reasonable because BIC is originally derived by the Laplace approximation under classical situation. Under known σ^2, Berger et al. (2003) proposed the Bayesian criterion called GBIC, which is derived by the Laplace approximation under CASE II. Then they showed that GBIC has model selection consistency under CASE II.

Generally, the original representation of Bayes factors or marginal densities involve integral. In the normal linear model setup, even if conjugate prior is used, hyperparameter and its prior distribution are usually introduced in order to guarantee objectivity, which is called fully Bayes method. (On the other hand, in empirical Bayes method, maximization of the conditional marginal density given hyperparameter with respect to hyperparameter is applied.) Since finding a prior of hyperparameter, which enables analytical calculation completely, is considered as extremely hard, the Laplace approximation has been applied. Needless to say, the Laplace approximation needs some assumptions, in particular, on "what goes to infinity". However, when both p and r are large (or small) in analysis of real data, the answer to the question which type of the Laplace approximation is more appropriate, is obscure. Moreover an approximated Bayes factor under one assumption does not necessarily have consistency on the other assumption, which is not good for practitioners. Therefore Bayes factor

1. without integral representation, which is however based on fully Bayes method,

2. with model selection consistency for two asymptotic situations, CASE I and II

is desirable, which we will propose in this paper. Actually, a special choice of the prior of hyperparameter, which completely enables analytical calculation of the marginal density, is the key in the paper.

Eventually the Bayes factor which we recommend is given by

$$\mathrm{BF}_{FB}[\mathcal{M}_{A+1};\mathcal{M}_1] = \frac{\Gamma(p/2)\Gamma(p(r-1)/2)}{\Gamma(1/2)\Gamma(\{pr-1\}/2)}\left(\frac{W_E}{W_T}\right)^{-p(r-1)/2+1/2} \tag{3}$$

where

$$\frac{W_E}{W_T} = \frac{\sum_{ij}(y_{ij} - \bar{y}_{i\cdot})^2}{\sum_{ij}(y_{ij} - \bar{y}_{\cdot\cdot})^2}, \ \bar{y}_{i\cdot} = \frac{\sum_j y_{ij}}{r}, \ \bar{y}_{\cdot\cdot} = \frac{\sum_{ij} y_{ij}}{pr}.$$

It is not only exactly proportional to the posterior probability of $\mathcal{M}_{A+1}$, but also a function of W_E/W_T, which is fundamental aggregated information of one-way ANOVA, from the frequentist viewpoint.

参考文献

[1] BERGER, J. O., GHOSH, J. K. & MUKHOPADHYAY, N. (2003). Approximations and consistency of Bayes factors as model dimension grows. *J. Statist. Plann. Inference* **112**, 241–258.

[2] CASELLA, G., GIRÓN, F. J., MARTÍNEZ, M. L. & MORENO, E. (2009). Consistency of Bayesian procedures for variable selection. *Ann. Statist.* **37**, 1207–1228.

[3] GEORGE, E. I. & FOSTER, D. P. (2000). Calibration and empirical Bayes variable selection. *Biometrika* **87**, 731–747.

[4] KASS, R. E. & RAFTERY, A. E. (1995). Bayes factors. *J. Amer. Statist. Assoc.* **90**, 773–795.

[5] LIANG, F., PAULO, R., MOLINA, G., CLYDE, M. A. & BERGER, J. O. (2008). Mixtures of g-priors for Bayesian variable selection. *J. Amer. Statist. Assoc.* **103**, 410–423.

[6] MARUYAMA, Y. & GEORGE, E. I. (2008). A g-prior extension for $p > n$. ArXiv:0801.4410v1 [stat.ME].

[7] MARUYAMA, Y. & STRAWDERMAN, W. E. (2005). A new class of generalized Bayes minimax ridge regression estimators. *Ann. Statist.* **33**, 1753–1770.

[8] MORENO, E., GIRÓN, F. J. & CASELLA, G. (2009). Consistency of objective Bayes tests as the model dimension increases. Available at: `http://stat.wharton.upenn.edu/statweb/Conference/OBayes09/`.

[9] SCHWARZ, G. (1978). Estimating the dimension of a model. *Ann. Statist.* **6**, 461–464.

[10] STONE, M. (1979). Comments on model selection criteria of Akaike and Schwarz. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **41**, 276–278.

[11] STRAWDERMAN, W. E. (1971). Proper Bayes minimax estimators of the multivariate normal mean. *Ann. Math. Statist.* **42**, 385–388.

[12] ZELLNER, A. (1986). On assessing prior distributions and Bayesian regression analysis with g-prior distributions. In *Bayesian inference and decision techniques*, vol. 6 of *Stud. Bayesian Econometrics Statist.* Amsterdam: North-Holland, pp. 233–243.

[13] ZELLNER, A. & SIOW, A. (1980). Posterior odds ratios for selected regression hypotheses. In *Bayesian Statistics: Proceedings of the First International Meeting held in Valencia (Spain)*, J. M. Bernardo, M. H. DeGroot, D. V. Lindley & A. F. M. Smith, eds. University of Valencia.