



A note on Bayes factor consistency in partial linear models

Taeryon Choi^{a,*}, Judith Rousseau^b

^a Department of Statistics, Korea University, Republic of Korea

^b CEREMADE, Université Paris Dauphine, France

ARTICLE INFO

Article history:

Received 24 December 2013

Received in revised form 15 November 2014

Accepted 28 March 2015

Available online 22 April 2015

Keywords:

Bayes factor

Consistency

Fourier series

Gaussian processes

Hellinger distance

Kullback–Leibler neighborhoods

Lack of fit testing

ABSTRACT

It has become increasingly important to understand the asymptotic behavior of the Bayes factor for model selection in general statistical models. In this paper, we discuss recent results on Bayes factor consistency in semiparametric regression problems where observations are independent but not identically distributed. Specifically, we deal with the model selection problem in the context of partial linear models in which the regression function is assumed to be the additive form of the parametric component and the nonparametric component using Gaussian process priors, and Bayes factor consistency is investigated for choosing between the parametric model and the semiparametric alternative.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Suppose we have two candidate models $\mathcal{M}_0$ and $\mathcal{M}_1$ for $y^n \in \mathcal{Y}^n$, a set of n observations from an arbitrary distribution P^n that is absolutely continuous with respect to a common measure μ^n on $\mathcal{Y}^n$. We also assume the two candidate models to have the following parameters and prior distributions, respectively: θ and $\pi_0(\theta)$, λ and $\pi_1(\lambda)$,

$$\mathcal{M}_0 = \{p_\theta^n(y^n), \theta \in \Theta, \pi_0(\theta)\}, \quad \mathcal{M}_1 = \{p_\lambda^n(y^n), \lambda \in \Lambda, \pi_1(\lambda)\}, \quad (1)$$

where $p_\theta^n(y^n)$ and $p_\lambda^n(y^n)$ denote the densities of y^n with respect to μ^n under $\mathcal{M}_0$ and $\mathcal{M}_1$, respectively.

Based on this set of observations, a common Bayesian procedure to measure the evidence in favor of $\mathcal{M}_0$ over $\mathcal{M}_1$ is to assess the Bayes factor (Jeffreys, 1961). Given the two candidate models $\mathcal{M}_0$ and $\mathcal{M}_1$, the Bayes factor, denoted B_{01} is given by the ratio of two marginal densities (Kass and Raftery, 1995),

$$B_{01} = \frac{p(y^n|\mathcal{M}_0)}{p(y^n|\mathcal{M}_1)} = \frac{m_0(y^n)}{m_1(y^n)} = \frac{\int p_\theta^n(y^n)\pi_0(\theta)d\theta}{\int p_\lambda^n(y^n)\pi_1(\lambda)d\lambda}. \quad (2)$$

Note that the large value of B_{01} is a strong evidence in support of model $\mathcal{M}_0$ (Jeffreys, 1961; Kass and Raftery, 1995). Accordingly, B_{01} is expected to converge to infinity as the sample size increases when $\mathcal{M}_0$ is the true model. This notion is called the consistency of the Bayes factor or *Bayes factor consistency*, and is formulated as follows

$$\lim_{n \rightarrow \infty} B_{01} = \begin{cases} \infty, & \text{in } P_{\theta_0}^n \text{ probability,} & \text{if } p_{\theta_0}^n \in \mathcal{M}_0 \\ 0, & \text{in } P_{\lambda_0}^n \text{ probability,} & \text{if } p_{\lambda_0}^n \in \mathcal{M}_1, \end{cases} \quad (3)$$

* Corresponding author. Tel.: +82 2 3290 2245; fax: +82 2 924 9895.

E-mail address: trchoi@gmail.com (T. Choi).

where P_{θ_0} represents the true probability measure belonging to the null model, and P_{λ_0} represents the true probability measure belonging to the alternative model. In (1)–(3), we typically have in mind that Λ is much larger than Θ , and Θ is often nested in Λ . In this paper we consider a setup where $\Theta \subset \mathbb{R}^d$ for some $0 < d < \infty$ and $\Lambda = \Theta \times \mathcal{F}$, where $\mathcal{F}$ is a function space.

Recent studies of Bayes factor consistency have focused on density estimation in the Bayesian goodness-of-fit testing problems. In goodness-of-fit testing for a parametric null ($\mathcal{M}_0$) against a nonparametric alternative ($\mathcal{M}_1$), Bayesian approaches have been investigated particularly in the context of Bayesian nonparametric testing of goodness-of-fit problems (see Tokdar et al., 2010 for a nontechnical survey). However, because of the specification of an alternative model and the computation of the Bayes factor in addition to theoretical issues such as Bayes factor consistency, Bayesian nonparametric goodness of fit testing is often regarded as a challenging inferential problem. In these regards, there have been several advances in Bayesian goodness-of-fit testing from the nonparametric Bayesian point of view. Readers are referred to Verdinelli and Wasserman (1998), Berger and Guglielmi (2001), Lenk (2003), Dass and Lee (2004), Walker et al. (2004), Ghosal et al. (2008), McVinish et al. (2009), and Tokdar et al. (2010). These results on Bayes factor consistency in density estimation problems are based on verifying sufficient conditions related to posterior consistency and posterior convergence rates. These conditions are mainly designed for the case of independent and identically distributed (i.i.d.) observations.

In particular, one of these sufficient conditions, the *Kullback–Leibler property*, in which the prior puts positive mass on all Kullback–Leibler neighborhoods of the true value of parameter, is known to be fundamental to Bayes factor consistency. That is, if a model has the Kullback–Leibler property, the Bayes factor always supports the model with the Kullback–Leibler property eventually when compared with any other model without the Kullback–Leibler property (e.g. Dass and Lee, 2004 and Walker et al., 2004). However, when two competing models have the Kullback–Leibler property, additional conditions are required for consistent model selection; Ghosal et al. (2008) and McVinish et al. (2009) investigated this issue by establishing sufficient conditions for the consistency of the Bayes factor under i.i.d. observations.

The consistency of Bayesian model selection in regression problems for independent observations has been largely studied in Gaussian linear regression models, particularly in the context of variable selection procedures (e.g. Liang et al., 2008, Casella et al., 2009, Moreno et al., 2010 and Shang and Clayton, 2011). On the other hand, Bayes factor consistency in partial linear models has not been fully studied in the literature except for a recent work by Choi et al. (2009), and we need to deal with the consistency of Bayes factor for partial linear models more intensively. In addition, it has become increasingly important to understand the asymptotic behavior of the Bayes factor for model selection in general statistical models. Thus, Bayes factor consistency needs to be investigated for statistical problems in non i.i.d. observations such as regression problems and Markov processes and time series model.

In this paper we discuss recent results on Bayes factor consistency for semiparametric regression problems where observations are independent but not identically distributed. Specifically, we deal with the model selection problem in the context of partial linear models in which the regression function is assumed to be the additive form of the parametric component and the nonparametric component using Gaussian process priors in Section 2, and Bayes factor consistency is discussed for choosing between the parametric model and the semiparametric alternative under these partial linear models in Section 3. Concluding remarks are made in Section 4.

2. The Bayesian partial linear model using Gaussian processes

A partial linear model (PLM: see, e.g., Härdle et al., 2004) is a semiparametric regression model whose mean function consists of two additive components, a linear part and a nonparametric part,

$$y_i = \beta_0 + \beta_1^T \mathbf{w}_i + f(x_i) + \sigma \epsilon_i, \quad \epsilon_i \stackrel{i.i.d.}{\sim} \mathcal{N}(0, 1), \quad (4)$$

where the mean function of the regression model in (4) has two parts: a p -dimensional parametric part with $\beta_1^T \mathbf{w}_i$, $\{\mathbf{w}_i\}_{i=1}^n \in [-1, 1]^p$, $p \geq 1$ and a nonparametric part with an unknown function $f(x_i)$, $\{x_i\}_{i=1}^n \in [0, 1]$ in the infinite dimensional parameter space. We consider both the case of fixed design and the case of random design. For the fixed design case, we assume that the design points are equally spaced. For the random design, we assume that $(\mathbf{w}_i, x_i)$ are sampled from a probability distribution.

Bayesian approaches to the partial linear models have been studied in the literature by developing different methods for estimating the nonparametric component $f(\cdot)$ (e.g. the Fourier series Lenk, 1999 and Choi et al., 2009), splines (Fahrmeier et al., 2013; Chib and Greenberg, 2010; Kyung, 2011), Gaussian processes (Choi and Woo, 2015), smoothing priors (Koop and Poirier, 2004), and wavelets (Qu, 2006; Ko et al., 2009).

Here, we study specific partial linear models using two different families of Gaussian processes, which are very common nonparametric priors for Bayesian regression functions. Specifically, the unknown nonparametric function $f(\cdot)$, is a priori modeled as a Gaussian process (GP) with zero mean function and a suitable covariance function $\mathbf{C}(\cdot, \cdot)$,

$$f(\cdot) \sim GP(\mathbf{0}, \mathbf{C}(\cdot, \cdot)), \quad (5)$$

where $\mathbf{C}(x_i, x_j) = \text{Cov}(f(x_i), f(x_j))$, $i, j = 1, \dots, n$. Briefly speaking, the Gaussian process $f(\cdot)$ in (5) can be decomposed, according to the Karhunen–Loève orthogonal expansion (see e.g. Grenander, 1981; Wahba, 1990)

$$f(x) = \sum_{k=0}^{\infty} \lambda_k \phi_k(x), \quad (6)$$

Download English Version:

<https://daneshyari.com/en/article/1148088>

Download Persian Version:

<https://daneshyari.com/article/1148088>

[Daneshyari.com](https://daneshyari.com)