Journal of Multivariate Analysis 150 (2016) 27–41

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Posterior convergence for Bayesian functional linear regression

Heng Lian [a], Taeryon Choi [b,*], Jie Meng [c], Seongil Jo [b]

[a] School of Mathematics and Statistics, University of New South Wales, Sydney, 2052, Australia
[b] Department of Statistics, Korea University, Republic of Korea
[c] Department of Marketing and Management, Macquarie University, North Ryde, NSW 2109, Australia

ARTICLE INFO

Article history:
Received 23 February 2015
Available online 18 May 2016

AMS 2000 subject classifications:
62G08
62F15

Keywords:
Functional regression
Minimax rate
Posterior contraction rate
Prediction risk
Reproducing kernel Hilbert space

ABSTRACT

We consider the asymptotic properties of Bayesian functional linear regression models where the response is a scalar and the predictor is a random function. Functional linear regression models have been routinely applied to many functional data analytic tasks in practice, and recent developments have been made in theory and methods. However, few works have investigated the frequentist convergence property of the posterior distribution of the Bayesian functional linear regression model. In this paper, we attempt to conduct a theoretical study to understand the posterior contraction rate in the Bayesian functional linear regression. It is shown that an appropriately chosen prior leads to the minimax rate in prediction risk.

1. Introduction

Over the last few decades, there has been considerable interest and research work in *Functional Data Analysis*, since the seminal work of Ramsay and Dalzell [35] introduced the statistical model and the tool for functional data, that is, a group of data defined by functions. Currently, it is common to collect functional data whose measurements vary smoothly with an underlying variable owing to advances in science and technology. Functional data analysis and its statistical inference have been widely studied in terms of the underlying theory and the methods used and have been increasingly used for many real applications (see, e.g., [36,37], and the references therein).

In particular, in the context of functional data analysis, we are often confronted with functional regression problems where the predictor X is a random function defined on a compact interval, say $[0, 1]$, and the response Y is a scalar random variable. For example, X can be daily temperature records at a location over a certain period to predict other climate variables, or it can be measurements of brain signals at a region in the brain over an extended time to predict disease status. Numerous applications and the recent developments in theory and methods for functional regression problems can also be found in, for instance, [37,18,17,26].

As mentioned above, Frequentist approaches to functional regression are well documented in two monographs [37,18] for both parametric and nonparametric analyses as mentioned before, and there is a sizable literature studying the asymptotic properties of functional regression from the frequentist perspective. For example, Cardot et al. [7] is an early

* Corresponding author.
E-mail address: trchoi1@gmail.com (T. Choi).

http://dx.doi.org/10.1016/j.jmva.2016.04.008

work using operator theory to study functional linear regression. Further, Cai and Hall [4] and Hall and Horowitz [25] give a more complete theory for this class of models. For functional data measured sparsely with error, Yao et al. [52] considers an estimator and its asymptotic properties, with a theory similar to classical nonparametric regression. Yuan and Cai [53] and Cai and Yuan [6] study minimax rates of convergence in the framework of the functional linear model and reproducing kernel Hilbert space.

On the other hand, the Bayesian counterparts of functional regression are comparatively rare although increasing, and contributions include methodological developments such as wavelet-based models [33,32], spline models [15], nonparametric Bayesian functional data analysis [40,34], and Gaussian process models [45,50]. However, all these works mainly focus on modeling and implementation as well as real applications to functional data analysis. Thus, to the best of our knowledge, there exist few theoretical results on posterior distribution for Bayesian functional regression. In particular, for Bayesian nonparametric problems such as the functional regression models we consider in this paper, one important issue that has attracted increasing attention over the past two decades is the frequentist validation of the Bayesian procedure. That is, we need to investigate the asymptotic behaviors of posterior distributions, whether the posterior distribution accumulates to the true value of the unknown parameter, and how fast the posterior distribution concentrates around the true value of the unknown parameter, in which one takes the frequentist view that the true value of the parameter exists (see e.g. [11,19] for nontechnical reviews).

It is known that even consistency of the posterior distribution may be illusive in nonparametric models with an infinite-dimensional parameter space (e.g., [16,27]). Accordingly, much effort has been expended on developing posterior consistency and to obtain posterior contraction rates in various Bayesian nonparametric models such as density estimation problems (e.g., [20]) and regression problems for non i.i.d. observations [3,21,13,22,47].

Although there exist many works on posterior consistency and posterior convergence rates and, multiple general theories on them have been proposed, it is increasingly important and fundamental to justify specific Bayesian methods. Further, the general results do not cover all cases of interest (e.g., posterior contraction rates in L^r distance for $1 \leq r \leq \infty$ need somewhat different approaches (see, e.g., [23,9])) to tackle the problems arising from the existing general theory. Even when the general theory can be applied, verifying the sufficient conditions for posterior consistency and contraction rates is not trivial, as this usually involves the complicated construction of appropriate tests, which varies across problems. Furthermore, the general theory typically provides posterior contraction rates in terms of Hellinger distance, which is not necessarily the natural metric implied by a specific problem. In particular, in the functional data structure, the current framework with general posterior consistency and posterior convergence rates would not be straightforward to apply, and thus the asymptotic investigation of posterior distribution in Bayesian functional data analysis problems has rarely been considered in the literature.

In this paper, we initiate the study of posterior contraction rates for a simple case, which is the Bayesian functional linear regression with a scalar response. That is, we have independent and identically distributed (i.i.d.) observations $(X_i, Y_i), i = 1, \ldots, n$, satisfying

$$Y_i = \langle X_i, \beta \rangle + \epsilon_i, \quad (1)$$

where $\langle X, \beta \rangle = \int_0^1 X(t)\beta(t)dt$, $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$, and X is a mean-zero square integrable stochastic process on $[0, 1]$. For simplicity, we assume that σ^2 is known and put a Gaussian process prior on the unknown slope function $\beta(\cdot)$. However, note that while our result can be extended to the general case with unknown σ, in which σ is assumed to be in an interval $[a, b]$ with $0 < a < b < \infty$ and a prior on σ^2 is specified, this is beyond the scope of this paper. Thus, the only unknown parameter is the slope function β, and we denote the true but unknown slope function as $\beta_0(\cdot)$.

The goal of this paper is to study the asymptotic behavior of the posterior distribution of $\beta(\cdot)$ as the sample size n increases. Specifically, under appropriate assumptions on the prior covariance operator, we show that the posterior distribution of $\beta(\cdot)$ contracts at the minimax rate if the smoothness parameter is correctly specified in the prior, and that some (non-optimal) rates in prediction risk can be obtained even if the smoothness of the prior is misspecified. Our results are comparable to previous works such as Ghosal and van der Vaart [22] and Knapik et al. [29], where the prior is "correctly" specified, and in this first attempt, we do not pursue the case wherein the Bayesian procedure automatically adapts to the smoothness of $\beta_0(\cdot)$, but leave it for the future.

The rest of the paper is organized as follows. In Section 2, we introduce some background material on functional regression, Gaussian process prior, and Bayesian contraction rates. In Section 3, the main results are presented and proved, followed by several remarks. In Section 4, empirical analysis is presented to illustrate the effect of the prior on the sensitivity of the posterior and the conditions for posterior convergence that we establish. We conclude the paper in Section 5 with a discussion.

2. Preliminaries

In this section, we present a summary of some background materials that include some of the notations and concepts used in the asymptotic analysis of this study. They are useful for discussing the main results in the next section and serve as motivations for our adoption of prediction risk in the evaluation of the posterior concentration rates. To define prediction risk, we begin with two fundamental quantities in posterior consistency and concentration rates, that is, the Kullback–Leibler divergence and the Hellinger distance between two probability distributions.