International Journal of Approximate Reasoning 80 (2017) 257–276

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Nonparametric adaptive Bayesian regression using priors with tractable normalizing constants and under qualitative assumptions

Khader Khadraoui

Laval University, Department of Mathematics and Statistics, Québec city G1V 0A6, Canada

ARTICLE INFO

Article history:
Received 14 December 2015
Received in revised form 26 August 2016
Accepted 23 September 2016
Available online 28 September 2016

Keywords:
Bayesian regression
Localized shape constraints
Free-knot B-spline
Reversible jump MCMC scheme

ABSTRACT

Shape constrained regression models are useful for analyzing data with specific shape responses, such as (monotone) dose-response curves, the (concave) utility functions of a risk averse decision maker, the (increasing) growth curves of children's height through time, that are particularly common in medicine, economic and epidemiological studies. This paper proposes a new adaptive Bayesian approach towards constructing prior distributions, with known normalizing constants, which enables us to take into account combinations of shape constraints and to localize each shape constraint on a given interval. Our strategy enables us to compute the simulation from the posterior distribution using a reversible jump Metropolis–Hastings scheme. The major advantages of the proposal are its flexibility achieved by adjusting local shape restrictions to detect better the high and low variability regions of the data and to facilitate the control of the regression function shape when there is no data at all in some regions. We give asymptotic results that show that our Bayesian method provides consistent function estimator from the adaptive prior. The performance of our method is investigated through a simulation study with small samples. An analysis of two real data sets is presented to illustrate the new approach.

1. Introduction

The problem of estimating functions under qualitative assumptions has a long history dating back at least to [14]. Regression under shape and smoothness constraints is of considerable interest both for theoretical and practical reasons. From a theoretical point of view, shape constraints and smoothing usually reduce the variance of the estimators. From a practical point of view, there are many situations in which it is necessary to take into account some prior knowledge on the shape of the regression function. The extensive literature on shape constrained problems has partly been motivated by specific applications but also by the fact that it has features that are shared with non-parametric function estimation.

As smoothing splines are defined as minimizers of a penalized sum of squares, they provide a natural theoretical framework to incorporate shape constraints simply by restricting the minimization over the set of constrained estimates [22,23,34]. Nevertheless, minimizing over a restricted set may lead to numerical difficulties in practice. Another strategy for monotone or convex regressions is to decompose the regression function into a tailored basis of splines as in [29] and [24]. The constraints are then taken into account simply by restricting the coefficients to be nonnegative. In the two papers just cited, the knots of the basis spline functions are fixed as is usually the case in the frequentist literature on shape restricted

E-mail address: khader.khadraoui@mat.ulaval.ca.

http://dx.doi.org/10.1016/j.ijar.2016.09.011
0888-613X/

regression. Regression splines with fixed knots also appear in the Bayesian literature on shape restricted regression. In these cases, the number of knots is typically large. A roughness penalty can be introduced from the prior distribution in order to avoid overfitting as in [5] and [1]. In [26] and Shively and Sager [33, Section 3], indicator variables are introduced in the model for selecting a subset of the basis functions. Usually, the constraint is included by restricting the spline coefficients to some subset.

Another way of selecting the number and the position of the knots is to use the Reversible Jump Markov Chain Monte Carlo (RJMCMC) algorithm introduced by [13]. Recall that a RJMCMC algorithm is a Metropolis–Hastings algorithm whose acceptance probability involves the computations of the ratio of priors and the ratio of likelihoods for two sets of parameters whose dimension may differ. In the unconstrained case, [7] and [9] get rid of the coefficients of the basis functions and run a RJMCMC algorithm for sampling from the posterior marginal distribution of knots only while [21] use a RJMCMC algorithm with all the parameters. [38] follows the former strategy but in the monotone constrained case. Contrary to unconstrained situation, the conditional posterior distribution of the data given the knots cannot be computed analytically and the likelihood ratio is not exactly known. The latter strategy [21] can also induce numerical difficulties in the computation of the prior ratio. In some convenient situations, the constraint on the coefficients is simple and the prior distribution is completely known. This is the case in [25] where the coefficients are constrained to be nonnegative. Then, a Gamma distribution for each coefficient fulfills the constraint and the prior ratio can easily be computed even for coefficient vectors with different dimensions. The same happens in [4] but with a different model. However, in these two papers, the support of the basis functions is the whole interval of the data. Then, such models do not enable us to localize the constraint or to consider a combination of constraints like being increasing on $[1, 2]$ and concave on $[3, 4]$. Such combinations of constraints have been undertaken by using B-splines with fixed knots in [2]. Free-knot B-splines for monotone regression are used in [16] and [17] and [19]. With B-splines, the constraint on coefficients is typically not as simple as forcing the coefficients to be nonnegative. Thus, it is usually included in the prior by means of truncated distributions like the truncated Gaussian distributions and the prior density is only known up to the normalizing constant which does change with the dimension of the coefficients vector. Thus, simplification of the unknown normalizing constants in the prior ratio cannot be done and the prior ratio remains typically unknown. This feature was already pointed out by [13] and it is apparently missed by [17]. Note that this problem can be avoided by running the algorithm in the unconstrained case and retaining only the samples for which the constraint is fulfilled as in [16]. Nevertheless, this strategy may behave poorly if the shape constraint is not sufficiently supported by the data.

In the present paper a new spline framework is introduced for estimating function under localized shape restrictions with unknown knot vector. This framework is focused on the construction of an adaptive prior that extends to estimating functions with all or at a least large classes of shape restrictions as well as combination of several constraints. There are two main goals of the present paper. The first is, as the dimension of the model parameters is unknown, to introduce a new construction of a truncated normal prior distribution with known normalizing constant for usual shape constraints. The idea is this prior enables us to compute the prior ratio of the acceptance probability of the RJMCMC algorithm used for sampling from the posterior distribution. The second goal is, when the regression functions are uniformly bounded and the design points are random, to prove almost sure consistency of the posterior probabilities of Kullback–Leibler and Hellinger neighborhoods of the joint density of the response and design point.

The paper is organized as follows. Section 2 introduces and discusses the concepts of B-spline and the control polygon. In Section 3, we introduce the model and the associated prior. Section 4 is devoted to the reversible jump Markov chain Monte Carlo sampling scheme to implement the method. Section 5 discusses the asymptotic properties of the posterior distribution. Section 6 presents simulation results to show the small sample properties across examples of functions as well as applications based on real world data sets. Section 7 contains a short discussion. Auxiliary results and proofs are deferred to the Appendix.

2. Notations and preliminaries

The intent in this section is to give a simple and direct development for B-splines via the recurrence relations. Let the integer k denote the order of the B-spline. Given a nondecreasing sequence of points $\{t_j \in \mathbb{R} | t_j \le t_{j+1}\}$ called knots, the B-spline function of order 1 is given by: $B_{j,1}(x) = 1$ if $t_j \le x < t_{j+1}$ and $B_{j,1}(x) = 0$ otherwise. From the first-order B-splines, we obtain higher-order B-splines by recurrence:

$$B_{j,k}(x) = \omega_{j,k}(x)\, B_{j,k-1}(x) + \Big(1 - \omega_{j+1,k}(x)\Big)\, B_{j+1,k-1}(x), \tag{2.1}$$

with

$$\omega_{j,k}(x) = \begin{cases} \frac{x - t_j}{t_{j+k-1} - t_j}, & \text{if } t_j < t_{j+k-1} \\ 0, & \text{otherwise.} \end{cases} \tag{2.2}$$

From this, we infer that $B_{j,k}$ is a piecewise polynomial of degree $< k$ which vanishes outside the interval $[t_j, t_{j+k})$. In particular, $B_{j,k}$ is just the zero function in case $t_j = t_{j+k}$. Also, by induction, $B_{j,k}$ is positive on the open interval (t_j, t_{j+k}), since both $\omega_{j,k}$ and $(1 - \omega_{j+1,k})$ are positive there. We refer the reader to de Boor [6] for a thorough presentation of the many other properties of B-splines. A spline of order k with knot sequence $t = (t_j)$ is, by definition, a linear combination