Statistics and Probability Letters 130 (2017) 5–11

Contents lists available at ScienceDirect

Statistics and Probability Letters

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

Bayesian sieve method for piece-wise smooth regression ☆

Taihe Yi *, Zhengming Wang

College of Science, National University of Defense Technology, Changsha, 410073, Hunan, China

ARTICLE INFO

Article history:
Received 6 February 2017
Received in revised form 30 April 2017
Accepted 12 July 2017
Available online 22 July 2017

MSC:
62G08
62G10
62G20

Keywords:
Change-point
Piece-wise smooth
Sieve
Posterior contraction rate
Posterior consistency

ABSTRACT

We study the piece-wise smooth regression from a theoretical Bayesian perspective. Our results indicate that under some mild assumptions, the posterior of the regression model and the change-points locations contracts at optimal nonparametric convergence rate up to a log-factor, and the number of change-points is posterior consistent.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

The piece-wise constant model

$$y_i = \sum_{j=0}^{p_0} s_j^0 \mathbf{1}\left(\tau_j^0 < t_i \le \tau_{j+1}^0\right) + \sigma\varepsilon_i, \quad i = 1, \dots, n \tag{1}$$

has recently received great attention from statisticians for its wide applications in finance (Lavielle and Teyssière, 2007), signal processing (Hotz et al., 2013) and genetic engineering (Jeng et al., 2010), see Lian (2010), Li et al. (2016), Du et al. (2016), Pein et al. (2016) and the references therein. A natural generalization is the piece-wise smooth regression,

$$y_i = \sum_{j=0}^{p_0} s_j^0(t_i) \mathbf{1}\left(\tau_j^0 < t_i \le \tau_{j+1}^0\right) + \sigma\varepsilon_i, \quad i = 1, \dots, n, \tag{2}$$

which has far from been well studied, where p_0 is the unknown number of change-points; $\tau_0^0 = 0$, $\tau_{p_0+1}^0 = 1$, and $\left(\tau_1^0, \dots, \tau_{p_0}^0\right)$ are unknown change-points locations; $s_j^0(t) \in \mathcal{S}(j = 0, \dots, p_0)$ are unknown smooth functions; $\varepsilon_1, \dots, \varepsilon_n$ are independent standard Gaussian variables; $\sigma > 0$ is a known noise level; and $\left\{t_i = \frac{i}{n+1} : i = 1, \dots, n\right\}$ is a deterministic uniform design.

☆ This work is supported by the National Natural Science Foundation of China (61573367).

* Corresponding author.
E-mail address: yitaihe@nudt.edu.cn (T. Yi).

http://dx.doi.org/10.1016/j.spl.2017.07.005
0167-7152/© 2017 Elsevier B.V. All rights reserved.

We study model (2) from a theoretical Bayesian perspective. Let $\delta > 0$ be a known constant, $\mathcal{S}(\delta)$ denote the piece-wise smooth function space

$$\left\{ \sum_{j=0}^{p} s_j(t)\mathbf{1}\left(\tau_j < t \le \tau_{j+1}\right) : p \ge 0, \tau_0 = 0, \tau_{p+1} = 1, \tau_{j+1} - \tau_j > \delta, s_j \in \mathcal{S} \right\}.$$

Suppose the true parameter $s_0(t) := \sum_{j=0}^{p_0} s_j^0(t)\mathbf{1}\left(\tau_j^0 < t \le \tau_{j+1}^0\right) \in \mathcal{S}(\delta)$. Let $\mathcal{D}_n := \{(t_i, y_i)\}_{i=1}^n$, $P_s^{(n)}$ be the conditional distribution of $\mathcal{D}_n$ given s, Π a prior on $\mathcal{S}(\delta)$. The posterior $\Pi(\cdot|\mathcal{D}_n)$ is said to contract to s_0 with rate $\varepsilon_n \to 0$ w.r.t. the norm $\|\cdot\|_n$ if $P_{s_0}^{(n)}\Pi\left(s \in \mathcal{S}(\delta) : \|s - s_0\|_n > M_n\varepsilon_n|\mathcal{D}_n\right) \to 0$ for all $M_n \to \infty$, where $\|s\|_n := \left[\frac{1}{n}\sum_{i=1}^n s^2(t_i)\right]^{\frac{1}{2}}$ is the empirical norm. Let $\varepsilon_n \to 0$ and $\bar{\varepsilon}_n \to 0$ be positive sequences satisfying $n\varepsilon_n^2 \to \infty$ and $n\bar{\varepsilon}_n^2 \to \infty$. The "test approach" (Ghosal et al., 2000, Ghosal and van der Vaart, 2007) states that the posterior contracts to s_0 with rate $\max\{\varepsilon_n, \bar{\varepsilon}_n\}$, if

- Step (a) the prior puts sufficient mass near the truth, i.e. $-\log \Pi(B_n(s_0, \bar{\varepsilon}_n)) \lesssim n\bar{\varepsilon}_n^2$, where $B_n(s_0, \bar{\varepsilon}_n) := \{s \in \mathcal{S}(\delta) : \|s_0 - s\|_n \le \sqrt{2}\sigma\bar{\varepsilon}_n\}$ and $a_n \lesssim b_n$ means that there is a constant $C > 0$ independent with n, such that $a_n \le Cb_n$.
- Step (b) there is a measurable set $\mathcal{S}_n(\delta) \subset \mathcal{S}(\delta)$ satisfying $\log \Pi(\mathcal{S}_n^c(\delta)) \lesssim -n\bar{\varepsilon}_n^2$;
- Step (c) there are universal constants $K > 0$ and $\xi \in (0, 1)$ such that for all $s_1 \in \mathcal{S}(\delta)$ with $\|s_1 - s_0\|_n > \varepsilon > 0$, there is a test ϕ_n satisfying

$$P_{s_0}^{(n)}\phi_n \le e^{-Kn\varepsilon^2}, \quad \sup_{s \in \mathcal{S}(\delta):\|s - s_1\|_n \le \xi\varepsilon} P_s^{(n)}(1 - \phi_n) \le e^{-Kn\varepsilon^2}, \tag{3}$$

and $\log N(\varepsilon_n, \mathcal{S}_n(\delta), \|\cdot\|_n) \lesssim n\varepsilon_n^2$, where $N(\varepsilon_n, \mathcal{S}_n(\delta), \|\cdot\|_n)$ is the minimal number of ε_n-balls needed to cover $\mathcal{S}_n(\delta)$ and $\log N(\varepsilon_n, \mathcal{S}_n(\delta), \|\cdot\|_n)$ the *metric entropy* of $\mathcal{S}_n(\delta)$ w.r.t. $\|\cdot\|_n$.

LeCam (1973) states that the minimal ε_n satisfying Step (c) is the *optimal minimax convergence rate*. A prior, which is irrelevant to the metric entropy of the parameter space meanwhile the posterior contracts with the optimal rate (up to a log-factor), is said to be *rate adaptive* (Scricciolo, 2015).

The main contributions of this paper are: (1) to bound the metric entropy of the piece-wise smooth model $\mathcal{S}(\delta)$ by the metric entropy of $\mathcal{S}$; and (2) to construct a sieve prior for piece-wise smooth regression and to give the posterior contraction rates of the model as well as the change-points locations. Sieve prior has become a common choice in nonparametric Bayesian and has been well studied in different models (such as smooth regression, density estimation and white noise model) recently, see de Jonge and van Zanten (2012), Arbel et al. (2013), Shen and Ghosal (2015) and the references cited therein. Our results can be seen as a new application of the sieve prior.

The paper is organized as follows. In Section 2, we provide a metric entropy bound for the piece-wise smooth model and formulate model assumptions. In Section 3 we present the main results. Section 4 is a short discussion on possible generalizations and limitations. All of the proofs are gathered in Section 5.

2. Metric entropy of the piece-wise smooth model

Let $\delta > 0$ be a constant, $\mathcal{S}$ be a smooth function space defined on [0, 1],

$$\mathcal{S}^p(\delta) := \left\{ \sum_{j=0}^{p} s_j(t)\mathbf{1}(\tau_j < t \le \tau_{j+1}) : \tau_0 = 0, \tau_{p+1} = 1, \tau_{j+1} - \tau_j \ge \delta, s_j \in \mathcal{S} \right\}$$

denote the piece-wise smooth model with p change-points. Define $\mathcal{S}^0(\delta) := \mathcal{S}$ and $\mathcal{S}(\delta) := \bigcup_{p=0}^{p_{\max}} \mathcal{S}^p(\delta)$, where $p_{\max}$ is the integer part of δ^{-1}. Note that if $\delta \ge 1$, both $\mathcal{S}^p(\delta)$ and $\mathcal{S}(\delta)$ degenerate to $\mathcal{S}$.

Theorem 2.1. *There is a constant C, such that for all $\varepsilon > 0$,*

$$\log N\left(\varepsilon, \mathcal{S}(\delta), \|\cdot\|_2\right) \le \frac{2}{\delta}\log N\left(\frac{\delta}{2}\varepsilon, \mathcal{S}, \|\cdot\|_2\right) + \frac{2}{\delta}\log\varepsilon^{-1} + C. \tag{4}$$

Note that $\log N\left(\frac{\delta}{2}\varepsilon, \mathcal{S}, \|\cdot\|_2\right)$ and $\log\varepsilon^{-1}$ correspond to the metric entropy of the smooth functions and the change-points locations respectively. Since δ is a constant and $\mathcal{S}^p(\delta) \subset \mathcal{S}(\delta)$, Eq. (4) indicates that if $\log N(\varepsilon, \mathcal{S}, \|\cdot\|_2) \gtrsim \log\varepsilon^{-1}$, which is always true, the metric entropy of $\mathcal{S}$, $\mathcal{S}^p(\delta)$ and $\mathcal{S}(\delta)$ are of the same order. Therefore, the optimal convergence rates of model (2) with $s_0 \in \mathcal{S}$, $s_0 \in \mathcal{S}^p(\delta)$ and $s_0 \in \mathcal{S}(\delta)$ are of the same order.

A sequence of sets $\{\mathcal{S}_k\}$ is called a q_k*-dimensional sieve model* of $(\mathcal{S}, \|\cdot\|_2)$ if (a) for all $k \in \mathbb{N}$, $\mathcal{S}_k = \{s_k(t; \boldsymbol{\theta}_k) : \boldsymbol{\theta}_k \in \mathbb{R}^{q_k}\} \in \mathcal{S}$, (b) for all $k_1 < k_2$, $\mathcal{S}_{k_1} \subset \mathcal{S}_{k_2}$ and (c) $\bigcup_{k=1}^{\infty}\mathcal{S}_k$ is a dense subset of $\mathcal{S}$. For all $s \in \mathcal{S}$ and $k \in \mathbb{N}$, let $\boldsymbol{\theta}_{k,s}$ denote an element of the set $\arg\min_{\boldsymbol{\theta}_k \in \mathbb{R}^{q_k}} \|s - s_k(t; \boldsymbol{\theta}_k)\|_2$. Let $\beta > 0$ be a constant, if for all $s \in \mathcal{S}$, $\left\|s - s_k(t; \boldsymbol{\theta}_{k,s})\right\|_2 \lesssim k^{-\beta}$, we say that the *regularity* of $\mathcal{S}$ w.r.t. $\{\mathcal{S}_k\}$ is β. Generally, if the highest regularity of $\mathcal{S}$, w.r.t. all possible sieves, is β, then the optimal convergence rate of estimating a parameter $s_0 \in \mathcal{S}$ is $n^{-\beta/(2\beta+1)}$ (take the Hölder space and Sobolev space as examples).

Download English Version:

https://daneshyari.com/en/article/5129637

Download Persian Version:

https://daneshyari.com/article/5129637

Daneshyari.com