Statistics and Probability Letters 103 (2015) 100–104

Contents lists available at ScienceDirect

Statistics and Probability Letters

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

Adaptive Bayesian inference in the Gaussian sequence model using exponential-variance priors

Debdeep Pati [a,*], Anirban Bhattacharya [b]

[a] Department of Statistics, Florida State University, Tallahassee, FL, United States

[b] Department of Statistics, Texas A & M University, College Station, TX, United States

ARTICLE INFO

Article history:
Received 7 January 2015
Received in revised form 1 April 2015
Accepted 11 April 2015
Available online 27 April 2015

Keywords:
Adaptive
Bayesian
Gaussian process
Posterior contraction
Sequence model

ABSTRACT

We revisit the problem of estimating the mean of an infinite dimensional normal distribution in a Bayesian paradigm. Of particular interest is obtaining adaptive estimation procedures so that the posterior distribution attains optimal rate of convergence without the knowledge of the true smoothness of the underlying parameter of interest. Belitser & Ghosal (2003) studied a class of power-variance priors and obtained adaptive posterior convergence rates assuming that the underlying smoothness lies inside a countable set on which the prior is specified. In this article, we propose a different class of exponential-variance priors, which leads to optimal rate of posterior convergence (up to a logarithmic factor) adaptively over all the smoothness levels in the positive real line. Our proposal draws a close parallel with signal estimation in a white noise model using rescaled Gaussian process prior with squared exponential covariance kernel.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

We consider the infinite Gaussian sequence model

$$X_i = \theta_i + \epsilon_i, \quad \epsilon_i \sim \mathrm{N}(0, \sigma^2), \ i = 1, 2, \ldots, \tag{1}$$

where the parameter of interest is the infinite-vector $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots) \in \ell_2$, i.e., $\sum_{i=1}^{\infty} \theta_i^2 < \infty$. Model (1) has received widespread attention since it encapsulates many intrinsic conceptual issues associated with non-parametric estimation. In particular, model (1) is equivalent to the canonical signal-in-white-noise model

$$X(t) = \int_0^t f(s)ds + \sigma W(t), \quad t \in [0, 1], \tag{2}$$

where $f \in L_2[0, 1]$ is the unknown function to be estimated based on noisy measurements $X(t), t \in [0, 1]$, and $W(\cdot)$ is a standard Wiener process. The equivalence between (1) and (2) can be established by considering an orthonormal basis $\{\psi_i\}$ of $L_2[0, 1]$ (for example, the Fourier basis) with the standard inner product $\langle g, h\rangle = \int_{t=0}^{1} g(t)h(t)dt$, and setting $X_i = \langle X, \psi_i\rangle, \theta_i = \langle f, \psi_i\rangle$ and $\epsilon_i = \langle W, \psi_i\rangle$; see, for example, Chapter 1 of Tsybakov (2008). With the calibration $\sigma = 1/\sqrt{n}$ in (1), Pinsker (1980) established the minimax quadratic risk of estimating $\boldsymbol{\theta}$ over Sobolev ellipsoids of the

* Corresponding author.
E-mail addresses: debdeep@stat.fsu.edu (D. Pati), anirbanb@stat.tamu.edu (A. Bhattacharya).

http://dx.doi.org/10.1016/j.spl.2015.04.012
0167-7152/© 2015 Elsevier B.V. All rights reserved.

form $\Theta_\beta(B) = \{\boldsymbol{\theta} : \sum_{i=1}^{\infty} i^{2\beta}\theta_i^2 \leq B\}$ as

$$\inf_{\hat{\boldsymbol{\theta}}} \sup_{\boldsymbol{\theta}\in\Theta_\beta(M)} \mathbb{E}_{\boldsymbol{\theta}} \left\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right\|_2^2 \asymp n^{-2\beta/(2\beta+1)}, \tag{3}$$

where $\mathbb{E}_{\boldsymbol{\theta}}$ denotes an expectation with respect to the distribution of $\mathbf{X} = (X_1, X_2, \ldots)$ given $\boldsymbol{\theta}$. Note that the minimax risk is the same as in the case of estimating a β-smooth function (Stone, 1982). Adaptive estimation strategies without the knowledge of the "smoothness"-parameter β were first developed by Efroimovich and Pinsker (1984). We refer the reader to the monograph by Johnstone (unpublished) which provides an excellent background and introduction to the infinite sequence model and reviews minimax and adaptive estimation in this context.

In a Bayesian framework, Zhao (2000) considered independent Gaussian priors on the entries of $\boldsymbol{\theta}$ of the form $\theta_i \sim \mathrm{N}(0, \tau_i^2)$ with $\tau_i^2 = \tau_i^2(\beta) = i^{-(2\beta+1)}$. The resulting posterior mean was shown to attain the minimax rate (3) when the true $\boldsymbol{\theta}_0 \in \Theta_\beta$. For the same prior, Belitser and Ghosal (2003) established that the posterior contraction rate (Ghosal et al., 2000) coincides with the minimax rate, i.e., the posterior probability assigned to an ℓ_2 neighborhood of the true parameter having radius a constant multiple of the minimax rate converges to one almost surely.

An unappealing aspect of the above prior is that it requires the knowledge of the true smoothness parameter β. The main contribution of Belitser and Ghosal (2003) was to develop an adaptive Bayesian procedure which attains the minimax rate without knowledge of the smoothness parameter. Bayesian procedures offer a natural prescription for adaptation by introducing one or more additional level of hierarchy in defining the prior. Rather than choosing a fixed level of the smoothness β, Belitser and Ghosal (2003) considered a discrete prior on β and showed that the resulting hierarchical procedure adapts to any smoothness level in the prior support. Choosing the discrete set to be a dense subset of the continuum (e.g., the set of rationals), it was additionally shown that one could adapt to any $\beta > 0$, though the posterior contraction rate overshot the minimax rate by an arbitrarily small positive power of n in this case.

In this paper, we propose a class of *exponential-variance* priors on $\boldsymbol{\theta}$ which is fundamentally different from the power-variance priors studied in the afore-mentioned literature (Zhao, 2000; Belitser and Ghosal, 2003). The proposed class of priors is indexed by a positive parameter a, which plays a similar role as an inverse-bandwidth parameter in nonparametric kernel estimation (Tsybakov, 2008). We first consider a non-adaptive scenario where an optimal choice of the parameter a given the knowledge of β is discussed. We next show that for a large class of prior distributions on a, the posterior achieves the minimax rate (up to a logarithmic term) for $\boldsymbol{\theta}_0 \in \Theta_\beta$ for any $\beta > 0$. Since the prior distribution does not require knowledge of β, the procedure is fully adaptive. Finally, we provide a heuristic argument to relate our prior with the rescaled Gaussian process priors developed by van der Vaart and van Zanten (2007, 2009) for nonparametric function estimation. This connection may be helpful in extending the results proven in this paper for the infinite sequence model to Gaussian process regression and related settings.

2. Preliminaries

Let $\ell_2 = \{\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots) : \sum_{i=1}^{\infty}\theta_i^2 < \infty\}$ denote the space of square-summable sequences. We shall write $\|\cdot\|$ for the ℓ_2 norm throughout the paper, so that for any $\boldsymbol{\theta} \in \ell_2$, $\|\boldsymbol{\theta}\|^2 = \sum_{i=1}^{\infty}\theta_i^2$. Let $\Theta_\beta = \{\boldsymbol{\theta} \in \ell_2 : \sum_{i=1} i^{2\beta}\theta_i^2 < \infty\}$ denote the Sobolev space of infinite dimensional vectors with "smoothness" $\beta > 0$, and define the Sobolev norm $\|\theta\|_\beta = (\sum_{i=1} i^{2\beta}\theta_i^2)^{1/2}$. Finally, let $\Theta_\beta(B)$ denote a Sobolev-ball of radius $\sqrt{B}$ defined as $\{\boldsymbol{\theta} \in \Theta_\beta : \|\theta\|_\beta^2 < B\}$.

Let $\mathbf{X}$ denote the infinite-dimensional random vector $\mathbf{X} = (X_1, X_2, \ldots)$ distributed as (1); we shall use $P_{\boldsymbol{\theta}}$ to denote the distribution of the $\mathbf{X}$. The notations $\mathbb{E}_\theta g(X)/\mathbb{V}_\theta g(X)$ are used to denote the expectation/ variance of $g(X)$ with respect to the distribution P_θ of $\mathbf{X}$. Define for $\epsilon > 0$ and $\boldsymbol{\theta}_0 \in \ell_2$, the Kullback–Leibler (KL) neighborhood of $P_{\boldsymbol{\theta}_0}$ as

$$\mathcal{K}(P_{\boldsymbol{\theta}_0}; \epsilon) = \left\{\boldsymbol{\theta} : \int P_{\boldsymbol{\theta}_0} \log(P_{\boldsymbol{\theta}_0}/P_{\boldsymbol{\theta}}) < \epsilon^2, \ \int P_{\boldsymbol{\theta}_0}[\log(P_{\boldsymbol{\theta}_0}/P_{\boldsymbol{\theta}})]^2 < \epsilon^2\right\}.$$

Let $L_2[0, 1]$ denote the space of square integrable functions on $[0, 1]$. To distinguish from the ℓ_2 norm, let $\|\cdot\|_2$ denote the L_2 norm on $[0, 1]$ with respect to the Lebesgue measure. Throughout $C, C', C_1, C_2, \ldots$ are generically used to denote positive constants whose values might change from one line to another, but are independent from everything else. $\lesssim$ / $\gtrsim$ denote inequalities up to a constant multiple. $a \asymp b$ when we have both $a \lesssim b$ and $a \gtrsim b$. Let $\phi(t) = (2\pi)^{-1/2}\exp(-t^2/2)$ denote the standard normal density and let $\phi_\sigma(t) = (1/\sigma)\phi(t/\sigma)$.

3. Prior specification and main results

In this Section, we propose a class of exponential-variance priors on ℓ_2 and state our results on posterior concentration using such priors. Proofs of all results are deferred to a supplemental document (see Appendix A).

Consider the infinite sequence model (1) with $\sigma = 1/\sqrt{n}$. We define a class of exponential-variance priors on $\boldsymbol{\theta}$ as follows:

$$\theta_i \mid a \sim \mathrm{N}(0, \tau_i^2(a)), \quad \tau_i^2(a) = \frac{1}{a}e^{-i/a}, \ i = 1, 2, \ldots. \tag{4}$$

Download English Version:

https://daneshyari.com/en/article/1151438

Download Persian Version:

https://daneshyari.com/article/1151438

Daneshyari.com