Journal of Statistical Planning and Inference 166 (2015) 78–86

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

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

Credible sets in the fixed design model with Brownian motion prior

Suzanne Sniekers [*,1], Aad van der Vaart [2]

Mathematical Institute, Leiden University, Niels Bohrweg 1, 2333 CA Leiden, The Netherlands

ARTICLE INFO

Article history:
Received 23 February 2014
Received in revised form 24 June 2014
Accepted 16 July 2014
Available online 24 July 2014

Keywords:
Bayesian inference
Nonparametric regression
Fixed design
Credible sets
Brownian motion prior

ABSTRACT

We consider the nonparametric regression problem, where we take fixed design points $x_i \in [0, 1]$. We apply Bayesian methods, taking scaled Brownian motion as a prior. The posterior mean is used as an estimator for the function of interest f at a given point. Bayesian credible sets are constructed using the posterior distribution, which are then studied using frequentist methods. Results on the coverage of such credible sets are obtained, which are seen to depend on the Hölder smoothness of the regression function f and the choice of scaling. An optimal scaling is derived for a given smoothness.

1. Introduction and main result

We consider estimating the regression function f in the fixed design regression problem, where we have data

$$Y_i = f(x_i) + Z_i, \quad i \in \{1, \ldots, n\}. \tag{1.1}$$

Here (x_i) is a known sequence of points in the interval $[0, 1]$, and (Z_i) is a sequence of unobservable i.i.d. standard normal random variables. We take a nonparametric Bayesian approach, using a Gaussian process prior $W = (W_t : t \in [0, 1])$ on f, and are interested in the resulting credible sets. These are sets of prescribed posterior probability, which in the Bayesian paradigm are used to quantify the remaining uncertainty of the statistical analysis. We investigate the coverage of these sets when treating them as confidence sets in the non-Bayesian setting. Specifically we focus on credible intervals for $f(x)$, the function f evaluated at a given point x, which can be derived from the marginal posterior distribution of W_x.

As a prior for f we consider the distribution of a scaled Brownian motion. Thus we are given a mean-zero Gaussian process $W = (W_t : t \in [0, 1])$ with covariance function $\mathrm{cov}(W_s, W_t) = c_n(s \wedge t)$, for given scale factors $c_n > 0$. We take this process to be independent of the sequence (Z_i). In the Bayesian setup the observations are distributed according to the model

$$Y_i = W_{x_i} + Z_i.$$

Furthermore, the posterior distribution of $f(x)$ is the conditional distribution of W_x given $Y_1, \ldots, Y_n$. In this Gaussian model the posterior distribution is also Gaussian and hence is characterized by its posterior mean $\hat{f}(x) = \mathrm{E}(W_x|Y_1, \ldots, Y_n)$ and posterior variance $\sigma_n^2 = \mathrm{var}(W_x|Y_1, \ldots, Y_n)$. The natural credible interval with level η for $f(x)$ is the central interval

$$C_\eta = \big(\hat{f}(x) - \sigma_n\zeta_\eta, \hat{f}(x) + \sigma_n\zeta_\eta\big),$$

* Corresponding author.
E-mail addresses: suzanne.sniekers@math.leidenuniv.nl, S.Sniekers@uva.nl (S. Sniekers).

[1] Research supported by Netherlands Organization for Scientific Research NWO.

[2] The research leading to these results has received funding from the European Research Council under ERC Grant Agreement 320637.

http://dx.doi.org/10.1016/j.jspi.2014.07.008
0378-3758/

where ζ_η is a standard normal quantile such that $\mathrm{P}(|Z| < \zeta_\eta) = \eta$ for $Z \sim \mathcal{N}(0, 1)$. The coverage of this interval in the frequentist setting is the probability $P_f(f(x) \in C_\eta)$, where P_f refers to the distribution of $Y_1, \ldots, Y_n$ in the original model (1.1), where a "true" f is given.

This model has been widely studied in the literature. In Kimeldorf and Wahba (1970), Kimeldorf and Wahba showed that the posterior mean is the solution to a penalized smoothing problem. Rates of contraction of the posterior distribution $W \mid \vec{Y}$ relative to the L_2-metric were obtained in van der Vaart and van Zanten (2007, 2008, 2011). The present paper studies the marginal posterior distribution $W_x \mid \vec{Y}$. The results on posterior mean and variance can be used to obtain rates of contraction for this marginal posterior. Bayesian credible sets for the function f in some infinite-dimensional space were considered in Wahba (1983), Cox (1993), Leahu (2011) and Knapik et al. (2011), but only in a heuristic discussion and simulation study, without proofs, or only for the white noise model. (Estimation of a smooth functional of f is a different problem, which may be studied using Bernstein–von Mises theorems.) The present paper extends this to pointwise credible sets in the regression model. Scaling factors c_n in the variance were introduced in van der Vaart and van Zanten (2007) with the purpose of adapting the prior to the smoothness of the underlying regression function. These authors show that the rescaled Brownian motion with $c_n = n^{(1-2\alpha)/(1+2\alpha)}$ is a suitable prior for a true function f of Hölder smoothness α, where $\alpha \in (0, 1]$. In this paper, we obtain a similar result in the marginal setting.

The prior W in this paper takes the value $W_0 = 0$ at the origin. This could be remedied by adding an independent normal variable to W, but as we consider the performance of the posterior distribution at fixed $x > 0$, this will be irrelevant in the following.

The present model permits a fairly explicit solution. We will consider the case where the design points are given by $x_i = i/n_+$ with $n_+ = n + 1/2$. The exact formulas cannot easily be extended to a more general choice of design points. Furthermore, we take the scaling factors equal to $c_n = n_+^\beta$ for some $\beta \in (-1, 1)$.

Define $C^\alpha[0, 1]$ as the space of Hölder continuous functions with exponent $\alpha \in (0, 1)$. The main result of the paper is the following theorem.

Theorem 1.1. *Define $\xi_\beta := \frac{1-\beta}{2(1+\beta)}$. The following holds for the coverage $c_\eta^f := P_f\big(f(x) \in C_\eta\big)$:*

- *If $\alpha > \xi_\beta$, we have $c_\eta^f \to P(|U| < \zeta_\eta) =: p_\eta > \eta$ for all $f \in C^\alpha[0, 1]$, where $U \sim \mathcal{N}(0, 1/2)$.*
- *If $\alpha = \xi_\beta$, then for each $p \in (0, p_\eta]$ there exists $f \in C^\alpha[0, 1]$ such that $c_\eta^f \to p$.*
- *If $\alpha < \xi_\beta$, there exists $f \in C^\alpha[0, 1]$ such that $c_\eta^f \to 0$.*

In the first of the three cases the credible interval is a conservative confidence set (i.e. $p_\eta > \eta$). Although it is wider than necessary for coverage, its width shrinks to zero at the same order of magnitude as the frequentist confidence interval based on the posterior mean, which would use the frequentist standard deviation of the posterior mean, rather than the standard deviation of the posterior distribution. This follows from the fact that p_η is strictly smaller than 1. As $\xi_\beta \downarrow 0$ as $\beta \uparrow 1$, the range of α for which this favourable conclusion holds can be made arbitrarily large by choice of β. However, we shall see that

$$\sigma_n \asymp n^{(\beta-1)/4}.$$

Therefore using a large value of β will also increase the width of the credible set, even by an order of magnitude.

In the third case the credible interval is too narrow to give positive coverage for all functions of given Hölder smoothness. The standard deviation of the posterior distribution is of smaller order than the bias of the posterior mean in this case. This is due to oversmoothing of the true function by the prior, the Bayesian way of choosing too large a bandwidth in a smoothing method.

Without scaling (i.e. $\beta = 0$) the cut-off between good and bad performance of the credible sets is at $\xi_0 = 1/2$. This can be viewed as the smoothness of Brownian motion itself. In this case, functions of smoothness bigger than $\frac{1}{2}$ yield credible sets with positive coverage, whereas functions that are rougher than Brownian motion do not.

Inspection of the proof shows that the assumption $f \in C^\alpha[0, 1]$ can be relaxed to Hölder continuity in an arbitrarily small neighbourhood of x.

In the next section, we gain insight into the posterior mean by analysing its coefficients as an L^2 projection. In the third section, we study the bias and variance of the posterior mean as a frequentist estimator, as well as the posterior variance. Combining these results, we arrive at our main theorem. Throughout, we use $A \lesssim B$ to mean $A \leq cB$ and $A \asymp B$ to mean $A \lesssim B$ and $B \lesssim A$.

2. Understanding the posterior mean

In order to be able to analyse credible sets, we will need to know more about the posterior mean $\hat{f}(x) = \mathrm{E}(W_x \mid Y_1, \ldots, Y_n)$. Since conditional expectations correspond to L^2 projections, we may write

$$\mathrm{E}(W_x \mid Y_1, \ldots, Y_n) = \sum_{i=1}^n a_i^n Y_i = Y^\top a^n,$$