Ann Inst Stat Math (2009) 61:835–859
DOI 10.1007/s10463-008-0168-2

Asymptotic properties of posterior distributions in nonparametric regression with non-Gaussian errors

Taeryon Choi

Received: 10 January 2006 / Revised: 2 March 2007 / Published online: 21 February 2008

Abstract We investigate the asymptotic behavior of posterior distributions in nonparametric regression problems when the distribution of noise structure of the regression model is assumed to be non-Gaussian but symmetric such as the Laplace distribution. Given prior distributions for the unknown regression function and the scale parameter of noise distribution, we show that the posterior distribution concentrates around the true values of parameters. Following the approach by Choi and Schervish (*Journal of Multivariate Analysis*, 98, 1969–1987, 2007) and extending their results, we prove consistency of the posterior distribution of the parameters for the nonparametric regression when errors are symmetric non-Gaussian with suitable assumptions.

Keywords Posterior consistency · Uniformly consistent tests · Kullback-Leibler divergence · Hellinger metric · Prior positivity · Symmetric density

1 Introduction

This paper presents asymptotic results of posterior distributions in nonparametric regression problems when the noise is assumed to have a symmetric non-Gaussian distribution such as the Laplace distribution. Specifically, in this paper, we verify almost sure consistency of posterior distributions in nonparametric regression problems with symmetric non Gaussian errors when suitable prior distributions are given on both the regression function and a scale parameter of noise distribution.

It is often the case that a regression model with Gaussian noises may not provide reasonable estimates to fit the data if the data contains outliers, due to the light tails

T. Choi (✉)
Department of Statistics, Korea University, Anam-dong, Sungbuk-gu, Seoul 136-701, South Korea
e-mail: trchoi@gmail.com

of the Gaussian noise distribution. A way to overcome this problem is to use a robust regression model based on a heavy-tailed noise distribution such as the Laplace or the Student-t distribution. It is also well known that least square estimators under non-Gaussian errors also perform well and they are robust to non-Gaussian errors. For example, a regression model using the Laplace noise distribution is frequently used for such a case (see, e.g. Kotz et al. 2001 and references therein).

Nonparametric regression has been one of the most active research areas in modern statistical inference, from the methodological development to the theoretical validation. The seminal work by Stone (1977) initiated the issue of consistent estimation of nonparametric regression problems, investigating strong consistency with weak conditions imposed on the underlying distribution. So far, much effort has been given to the theoretical justification of nonparametric regression problems such as consistency, optimal rate of convergence, in particular, from a frequentist perspective. Bayesian approach to nonparametric regression problems provides an alternative statistical framework and needs to be justified in terms of asymptotic points of view, introducing the concept of posterior consistency and establishing it. Posterior consistency and the question about the rate of convergence of posterior distribution in nonparametric regression problems have been mainly studied under Gaussian noise distribution (e.g. Shen and Wasserman 2001; Huang 2004; Choi and Schervish 2007) and further efforts are expected to be taken under the general noise distribution.

Specifically, a Bayesian approach in the nonparametric problem using a prior on the regression function and specifying a Gaussian error distribution has been shown to be consistent, based on the concept of almost sure posterior consistency in Choi and Schervish (2007). However, in contrast to the case where we specify the error as Gaussian, little attention has been paid to asymptotic behavior of Bayesian regression models with non-Gaussian error. Therefore, studying the asymptotic behavior of posterior distribution of nonparametric regression model with non-Gaussian error is an important and interesting problem which we pursue here. In order to answer this problem, we establish posterior consistency of nonparametric regression model with non-Gaussian errors. That is, we justify that the posterior distribution concentrates around true values of the regression function and the scale parameter of the noise distribution.

The rest of the paper is organized as follows. In Sect. 2, we describe the Bayesian approach to nonparametric regression model and introduce the concept of posterior consistency. In Sect. 3, we provide main results by establishing posterior consistency of nonparametric regression models with non-Gaussian errors. In Sect. 4, we verify these main results. In Sect. 5, we examine supplementary results that are worth further consideration. Finally, we make concluding remarks with discussion in Sect. 6.

2 Problem description

2.1 Posterior consistency

In the Bayesian approach to statistical inference, posterior distribution summarizes information regarding unknown parameters, combined with likelihood and the prior distribution. As the sample size increases, the posterior distribution is expected to concentrate around the true value of parameter, known as the concept of posterior consistency. Posterior consistency can be thought of as a theoretical justification of the Bayesian procedure. A systematic discussion of posterior consistency can be found in Ghosh and Ramamoorthi (2003). The technical description of posterior consistency will be also presented in the next section. An important result related to posterior consistency is the Bayesian robustness, i.e., the robustness of the posterior inference with respect to the choice of prior, which means that given a sufficient number of observations, the posterior distribution should be insensitive to different choices of prior distributions as long as each of prior distributions satisfies appropriate conditions.

When the parameter space is finite dimensional, posterior consistency can be achieved easily under fairly general conditions when the true value of the parameter is in the support of the prior. However, study of asymptotic properties of a Bayesian method such as posterior consistency is much more difficult in the nonparametric context where the problem involves infinite-dimensional parameters, compared to parametric Bayesian method. When it comes to the case of infinite-dimensional parameters, it is not sufficient to imply consistency at the true value of parameter, merely having positive probability in all of suitable neighborhoods (see e.g. Freedman 1963; Diaconis and Freedman 1986). Thus, posterior consistency may fail to hold although natural priors are used to put the postive mass in every topological neighborhood of the true value of parameter. For the posterior to be consistent, it is necessary the true value of parameter need to be separated from the complements of such neighborhoods, which can be formalized as an existence of uniformly consistent test (Ghosh and Ramamoorthi, 2003, Definition 4.4.2). In fact, these two conditions have been already considered in the theorem of Schwartz (1965) for posterior consistency, one for the prior probability and the other for uniformly consistency. Recently, there have been several extensions of Schwartz' Theorem and many results giving general conditions under which features of posterior distribution are consistent in infinite-dimensional spaces, particularly for semiparametric/nonparametric regression problems such as Amewou-Atisso et al. (2003), Choudhuri et al. (2004), Ghosal and Roy (2006) and Choi and Schervish (2007).

2.2 Bayesian nonparametric regression with non-gaussian errors

Let us consider a regression model with an additive noise term when the distribution of noise is assumed to be symmetric non-Gaussian such as the Laplace (double exponential) distribution. We observe a response Y corresponding to a covariate value $X = x$ in a bounded interval $[0, 1]$. Here, two possibilities for the covariate X are considered; either it is random or fixed. In the first case, we regard the covariates as random samples from a probability distribution function. That is, let P_{θ_0} be the joint distribution of a random vector (X, Y) satisfying $Y = \eta(X) + \epsilon$. We consider random samples of $(X_1, Y_1), \ldots, (X_n, Y_n)$ that have the same distribution as (X, Y). In the second case, we regard the covariates as fixed ahead of time. That is, we consider fixed design points, $x_1, \ldots, x_n$, as covariate values and their corresponding response observations $Y_1, \ldots, Y_n$ satisfying $Y_i = \eta(x_i) + \epsilon_i,\ i = 1, \ldots, n$.

In both cases, the noise distribution ϵ is assumed to be symmetric around zero with density $\phi\left(\cdot/\sigma\right)/\sigma$ and, equivalently, the conditional p.d.f. of Y given $X = x$, $p(Y|x)$ is expressed as

$$p(y|x) = \frac{1}{\sigma}\phi\left(\frac{y-\eta(x)}{\sigma}\right), \tag{1}$$

where σ is the unknown scale parameter of noise, the unknown regression function, η is assumed to be a continuously differentiable function on [0, 1] and $\phi(\cdot)$ is a density function that satisfies

$$|\log\phi(z) - \log\phi(z')| \leq c|z - z'| \tag{2}$$

for some fixed $c > 0$ and for all $z, z' \in \mathbb{R}$, and ϕ admits a moment generating function $m(t) = \int \exp(ty)\phi(y)\mathrm{d}y$ in some interval $(-T, T)$ with $T > 0$,

$$m(t) = 1 + at^2 + o(t^2) \tag{3}$$

as $t \to 0$ for some $a > 0$, and $\int |z|^2\phi(z)\mathrm{d}z < \infty$. Note that these conditions are satisfied by commonly used probability densities, symmetric around zero such as the Laplace density in addition to the Gaussian density. Under the regression setup of (1), the parameter is the pair of the unknown scale parameter of noise, σ and the unknown regression function, $\eta(x)$.

Bayesian approaches to this problem will begin with specifying prior distributions on a given class of regression functions and the unknown σ. We assign prior probability distributions, Π_η and Π_ν for η and σ, respectively. When the covariate X is treated as a random variable, the probability distribution of X, Q, will be considered. However, since our interest is the posterior for η and σ and the distribution of X will be integrated out, we can assume, without loss of generality, that Q is a known distribution.

Several ways of putting a prior for $\eta(x)$ have been proposed in the statistical literature and typical examples include orthogonal basis representation, a spline series prior and a Gaussian process prior. See, e.g. Choi (2005) for a general survey of these three methods and some key references on the subject. However, the choice of nonparametric prior needs to be made carefully so that it reflects the underlying assumption about the true regression function. Suppose that the true regression function is completely unknown with only a few assumptions about the smoothness such as the continuity or differentiability. In this case, if one chooses a parametric function of known functional form rather than nonparametric form, by putting a prior one to the parametric function, we might not be able to achieve posterior consistency. Note that for consistency, the support of the prior under consideration need to be big enough to include the true regression function. In other words, the prior under consideration should ensure the positive probability on every neighborhood of true regression function in order to achieve posterior consistency. In this regard, assigning a prior based on known functional form to the unknown true regression function cannot be validated in terms of consistency, and much attention needs to be taken to consider a nonparametric prior for the unknown function. For instance, when we utilize Gaussian process as a prior of the regression function, suitable covariance functions should be considered (see, e.g. Ghosal and Roy 2006; Tokdar and Ghosh 2007; Choi and Schervish 2007). For a prior with orthogonal basis representation and a spline series prior, appropriate conditions for orthogonal basis and corresponding splines need to be considered, with respect to the true function and its smoothness. (e.g. Shen and Wasserman 2001; Choi and Schervish 2007; Ghosal and van der Vaart 2007).

For the implementation of Bayesian nonparametric regression and computation form a given set of data, we follow the fundamental Bayesian formalism: That is, the implementation will be based on the posterior distribution, combined with likelihood and prior distribution. When the scale parameter σ is assumed to be known and the observed data are $(Y_1, x_1), \dots, (Y_n, x_n)$, with $x_1, \dots, x_n$ either fixed or sampled from a probability distribution, the posterior distribution of $\{\eta(x_i)\}_{i=1}^n$ can be calculated in the following way.

Let $\boldsymbol{\eta}$ be a vector of values, evaluated at the n points $x_1, \dots, x_n$, $\boldsymbol{\eta} = (\eta(x_1), \eta(x_2), \dots, \eta(x_n))^T$. Thus, the posterior distribution of $\boldsymbol{\eta}$, is proportional to the product of the likelihood, joint distribution of n observations, $Y_1, \dots, Y_n$ and the prior distribution of $\boldsymbol{\eta}$:

$$\Pi(\boldsymbol{\eta}|\mathbf{Y}, \boldsymbol{x}, \sigma^2) \propto P(\mathbf{Y}|\boldsymbol{\eta}, \boldsymbol{x}, \sigma^2)\Pi(\boldsymbol{\eta}, \sigma^2),$$

where $\mathbf{Y} = (Y_1, \dots, Y_n)$ and $\boldsymbol{x} = (x_1, \dots, x_n)$. After identifying posterior distribution, we can approximate posterior distribution based on computer-intensive methods such as Gibbs samplings or Markov Chain Monte Carlo (MCMC) methods. In principle, with MCMC methods, Bayesian inference is performed and the posterior distribution of unknown parameters are obtained numerically. From methodological and computational point of views, several techniques and tools have been proposed and developed, while theoretical and asymptotic studies still leave much to be desired. For example, when the noise distribution is assumed to be known and the Gaussian process prior is used, computational methods have been developed such as in Neal (1996, 1997) and Paciorek (2003). In addition, Neal (1997) also considered Gaussian process regression by using a Student's t-distributed noise.

However, since the noise distribution is assumed to be non-Gaussian, the estimated regression curve is not always anaytically tractable although it is computationally feasibile as stated above. Hence, one needs theoretical validation of the method by verifying consistency properties indirectly. As stated previously, it is known that the dimension of the parameter space plays a role in determining posterior consistency. When the parameter space is finite dimensional, posterior consistency can be achieved easily under fairly general conditions. Hence, consistency of posteriors for nonparametric regression problems is involved with infinite-dimensional parameters is a much more challenging problem than in the finite-dimensional case. The similar nonparametric Bayesian regression setup has been considered under the Gaussian noise distribution and its posterior consistency has been well investigated by Choi and Schervish (2007). In this paper, we consider the posterior distribution of nonparametric Bayesain Laplace regression problem and study its asymptotic behavior of posterior distribution. We follow the same approach as Choi and Schervish (2007) to establish posterior consistency.

3 Main results

Let A be a appropriate neighborhood of the true parameter value $\theta_0 \equiv (\eta_0, \sigma_0)$ and $Z^n = (Z_1, \ldots, Z_n)$, $\{Z_i\}_{i=1}^n = \{(X_i, Y_i)\}_{i=1}^n$ be the data. Then, given the prior, Π, the posterior probability of A is written as

$$\Pi(A|Z^n) = \frac{\int_A \prod_{i=1}^n f(y_i - \eta(x_i))\mathrm{d}\Pi(\theta)}{\int \prod_{i=1}^n f(y_i - \eta(x_i))\mathrm{d}\Pi(\theta)} \tag{4}$$

Let Θ be the product space of $\mathcal{F}$ and $\mathbb{R}^+$, where $\mathcal{F}$ is the set of Borel measurable functions defined on [0, 1] and $\mathbb{R}^+$ is the positive real line. For now, assume that we have chosen a topology on Θ.

For each neighborhood $\mathcal{N}$ of the true regression function η_0, the true noise variance σ and each sample size n, we compute the posterior probability as in (4), $p_{n,\mathcal{N}}(Z^n) = \Pi(\{\theta \in \mathcal{N}\}|Z^n)$, as a function of the data. To say that the posterior distribution of θ is *almost surely consistent* means that, for every neighborhood $\mathcal{N}$, $\lim_{n\to\infty} p_{n,\mathcal{N}} = 1$ $a.s.$ with respect to the joint distribution of the infinite sequence of data values. Similarly, in-probability consistency means that for all $\mathcal{N}$ $p_{n,\mathcal{N}}$ converges to 1 in probability.

To make these definitions precise, we must specify the topology on Θ, in particular on $\mathcal{F}$. This topology can be chosen independently of whether one wishes to consider almost sure consistency or in-probability consistency of the posterior. For this purpose, we use a popular choice of topology on $\mathcal{F}$, L^1 topology related to a probability measure Q on the domain [0, 1] of the regression functions. The $L^1(Q)$ distance between two functions η_1 and η_2 is $\|\eta_1 - \eta_2\|_1 = \int_0^1 |\eta_1 - \eta_2|\mathrm{d}Q$. In addition, we use a Hellinger metric for joint densities f for $Z = (X, Y)$ with respect to a product measure $\xi = Q \times \lambda$, where λ is a Lebesgue measure, namely $f(x, y) = \phi([y - \eta(x)]/\sigma)/\sigma$. The Hellinger distance between two densities f_1 and f_2 is $\left\{\int \left[\sqrt{f_1(x, y)} - \sqrt{f_2(x, y)}\right]^2 \mathrm{d}\xi\right\}^{1/2}$. These metrics were considered for looking at posterior consistency under normal noise distribution by Choi and Schervish (2007). Another frequently used neighborhood is the weak neighborhood of the true probability measure of P_0 with the true joint density of X and Y, f_0. We say $p_{n,N}$ is weakly consistent at P_0 if posterior distribution $p_{n,N}$, achieves almost sure consistency when N is based on every weak neighborhood of P_0. Note that when we consider the joint distribution of (X, Y), the distribution Q of X is assumed to be a known form, thus this distribution is canceled out in the expression for the posterior distribution of η.

Before stating main results, we must make two assumptions, one about the suitable prior distributions of the regression function and the scale parameter of the noise distribution and the other about the rate at which the covariate values fill out the interval [0, 1].

The first assumption about prior distributions is sufficient for the true density of the data, i.e., either conditional density of Y when the covariate values are fixed ahead of time or the joint density of Y and X when the covariate are sampled from a known distribution Q, to be in the Kullback-Leibler support. (See Ghosh and Ramamoorthi 2003, Definition 4.4) and Choi and Schervish 2007, Theorem 1)). The second assumption ensures that most of the prior probability is given to the appropriate sieve, which grows to the parameter space, as the sample size increases.

Suppose that a suitable prior, $\Pi = \Pi_\eta \times \Pi_\sigma$ assigned to $\theta \equiv (\eta, \sigma)$, satisfies the following two assumptions:

Assumption P.1 Let $\epsilon > 0$. Define

$$B_\epsilon = \left\{ (\eta, \sigma) \ : \ \sup_{x\in[0,1]} |\eta(x) - \eta_0(x)| < \epsilon \ , \left| \frac{\sigma}{\sigma_0} - 1 \right| < \epsilon \right\} \tag{5}$$

and assume that $\Pi(B_\epsilon) > 0$ for all $\epsilon > 0$.

Assumption P.2 Define

$$E_n = \left\{ \eta \ : \ \sup_{x\in[0,1]} |\eta(x)| < n^{3/4} \ , \quad \sup_{x\in[0,1]} |\eta'(x)| < n^{3/4} \right\}, \tag{6}$$

and assume that there exists a constant $\beta > 0$ such that $\Pi_\eta \left\{ E_n^C \right\} \leq \exp(-n\beta)$.

For the assumption about covariate values, we consider two versions of the assumption, depending on the nature of the sequence $X_1, X_2, \ldots$ of convariates. First, we consider the case that the covariates are all fixed values, designed ahead of time. The following assumption is about how fast those fixed covariate values fill out the interval [0, 1].

Assumption D.1 Let $0 = x_0 < x_1 \leq x_2 \leq \cdots \leq x_n < x_{n+1} = 1$ be the design points on [0, 1] and let $S_i = x_{i+1} - x_i$, $i = 0, \ldots, n$ denote the spacings between them. There is a constant $0 < K_1 < 1$ such that the $\max_{0\leq i\leq n} S_i < 1/(K_1 n)$.

Now, we provide a result about posterior consistency for fixed covariates, in which the data $\{Y_n\}_{n=1}^\infty$ are assumed to be conditionally independent with a symmetric conditional density $\phi([y-\eta(x)]/\sigma)/\sigma$ given η, σ and the covariates. To investigate posterior consistency with nonrandom covariates, we apply Theorem 1 of Choi and Schervish (2007) by making $p_i(z; \theta)$ equal to $f_i(z; \theta_0)$ as $\phi([y_i - \eta(x)]/\sigma)/\sigma$ and by assuming D.1. In this case, the unknown parameters are the regression function and the scale parameter of the noise distribution. We prove it in the next section, by verifying sufficient conditions on existence of tests as in Theorem 1 of Choi and Schervish (2007), stated in the appendix. We assume the true regression function η_0 is continuously differentiable.

Theorem 1 *Suppose that the values of the covariate in* [0, 1] *arise from a fixed design satisfying the Assumption* D.1. *Assume that the prior, Π, satisfies Assumptions* P.1 *and* P.2. *Let P_0 be the conditional distribution of $\{Y_n\}_{n=1}^\infty$ given $\{x_n\}_{n=1}^\infty$ assuming that $\theta_0 = (\eta_0, \sigma_0)$ is the true value of parameter. Let $\epsilon > 0$ and define A_ϵ,*

$$A_\epsilon = \left\{ (\eta, \sigma) : \int |\eta(x) - \eta_0(x)| \mathrm{d}x < \epsilon, \left| \frac{\sigma}{\sigma_0} - 1 \right| < \epsilon \right\}.$$

Then, for every $\epsilon > 0$, $\Pi \left(A_\epsilon^C \mid Y_1, \ldots, Y_n, x_1, \ldots, x_n \right) \to 0$ a.s. $[P_0]$.

When we deal with random covariates, the case where the covariates are sampled from a probability distribution Q, we have a few theorems based on the joint density $f(x, y) = p(y|x)q(x)$ with respect to $\xi = Q \times \lambda$, where $p(y|x)$ is the conditional density of y given x, $q(x)$ is the marginal density of x, λ is the Lebesgue measure as stated earlier. In this case, the probability distribution of Q is assumed to be known. For example, the X_i's are assumed to be independent $U(0, 1)$ random variables. Then this distribution is canceled out of the expression (4) for the posterior of $\theta \equiv (\eta, \sigma)$. In addition, when Q is unknown, and thus a prior on Q is assigned, if we assume the indepedence priors for η, σ and Q, the posterior distribution of (η, σ) would not change regardless of a prior on Q, by the marginalization. Note that in this regard, the distribution of X would not be a matter of concern.

First, we consider weak consistency of posterior distribution. The weak consistency of posterior distribution is defined as follows: (see also Tokdar and Ghosh 2007).

Definition 1 (*weak consistency*) Suppose that random samples $X_1, \ldots, X_n$ are from a density f_0 that belongs to certain space of densities $\mathcal{F}$. Let Π be a prior distribution on $\mathcal{F}$. A prior Π on $\mathcal{F}$ is said to achieve weak posterior consistency at f_0 it for any weak neighborhood U of f_0, $\Pi(U|X_1, \ldots, X_n) \rightarrow 1$ almost surely under P_{f_0}.

The Assumption P.1 is sufficient to achieve weak consistency at the true joint density $f_0(x, y)$ so that f_0 is in the Kullback-Leibler support as explained in Schwartz (1965), Ghosh and Ramamoorthi (2003) and Tokdar and Ghosh (2007). We provide the theorem regarding weak consistency, which implies that posterior probability for any weak neighborhood of the true density converges to 1 almost surely under the true probability measure.

Theorem 2 *Suppose that the value of the covariate in* $[0, 1]$ *sampled from a probability distribution* Q *and prior distribution,* Π *satisfies the Assumption* P.1. *Let* P_0 *be the joint distribution of* $\{(X_n, Y_n)\}_{n=1}^{\infty}$ *assuming that* (η_0, σ_0) *is the true value of parameter and the true join density of* (X, Y) *is* $f_0(x, y) = p(y|x, \eta_0, \sigma_0)q(x)$. *Then the posterior is weakly consistent at* $f_0(x, y)$. *That is, for any weak neighborhood* U *of* f_0,

$$\Pi(U^C|(Y_1, X_1), \ldots, (Y_n, X_n)) \rightarrow 0 \ \ a.s. \text{ under } P_0.$$

Second, we state the Hellinger consistency of posterior distribution. In other words, we show that the posterior probability of every Hellinger neighborhood of the true value of parameter converges to 1 almost surely with respect to the true probability measure. In addition, In addition, we provide consistency of posterior distribution based on L^1 metric. For these purposes, we consider only the case in which the support of the prior distribution contains only uniformly bounded regression functions as stated in the Assumption P.3 below. Similar assumptions about the uniform boundedness of regression function can be found in Shen and Wasserman (2001) and Huang (2004). In this case, the probability distribution of Q is assumed to be known. For example, the X_i's are assumed to be independent $U(0, 1)$ random variables. Then this distribution is canceled out of the expression for the posterior of $\theta \equiv (\eta, \sigma)$.

Assumption P.3: Let Π'_η be the prior for η satisfying P.1 and P.2. Let $\Omega = \left\{\eta : \sup_{x\in[0,1]} |\eta(x)| < M\right\}$ with $M > \sup_{x\in[0,1]} |\eta_0(x)|$. Assume that $\Pi_\eta(\cdot) = \Pi'_\eta(\cdot \cap \Omega)/\Pi'_\eta(\Omega)$ with $\Pi'_\eta(\Omega) > 0$.

Theorem 3 *Suppose that the values of the covariate in* $[0, 1]$ *are sampled from a probability distribution* Q. *Suppose that assumptions P.1–P.3 hold. Let* P_0 *be the joint distribution of* $\{(X_n, Y_n)\}_{n=1}^{\infty}$ *assuming that* (η_0, σ_0) *is the true value of parameter and the true joint density of* (X, Y) *is* $f_0(x, y) = p(y|x, \eta_0, \sigma_0)q(x)$.

1. *Let* $\epsilon > 0$ *and define* B_ϵ,

$$B_\epsilon = \left\{(\eta, \sigma) : \left\{\int (\sqrt{f} - \sqrt{f_0})^2 \mathrm{d}\xi\right\}^{1/2} < \epsilon\right\}.$$

Then for every $\epsilon > 0$, $\Pi\left(B_\epsilon^C \mid (X_1, Y_1), \ldots, (X_n, Y_n)\right) \to 0$ *a.s.* $[P_0]$.

2. *Furthermore, we define* C_ϵ,

$$C_\epsilon \left\{(\eta, \sigma) : \int |\eta - \eta_0| \mathrm{d}Q < \epsilon, \; \left|\frac{\sigma}{\sigma_0} - 1\right| < \epsilon\right\}$$

Then for every $\epsilon > 0$, $\Pi\left(C_\epsilon^C \mid (X_1, Y_1), \ldots, (X_n, Y_n)\right) \to 0$ *a.s.* $[P_0]$.

Two typical examples of prior distributions for regression function—one about Gaussian process prior and the other about an orthogonal expansion of the regression function—are considered in Choi and Schervish (2007), and we can also make use of those two typical priors for regression function here. Note that these two priors satisfy the Assumptions P.1 and P.2 similarly as discussed in Choi and Schervish (2007). We give the detailed proof in the next section.

4 The proofs of main results

This section contains the proofs of the main consistency results. We state several theorems with different conditions on the covariate (nonrandom and random covariates) and different topologies (L^1, weak neighborhoods and Hellinger). The proofs of these results all rely on Theorem A.1 in the appendix (Theorem 1 in Choi and Schervish 2007), and thereby have many steps in common.

4.1 Proof of Theorem 1

First, we consider posterior consistency based on L_1 distance when covariate values arise in a fixed design, known ahead of time. This section contains the proof of condition (A1) of Theorem A.1, which virtually the same for all of the main theorems. In addition, we show how to construct uniformly consistent tests, described in (A2) of Theorem A.1. This is done by piecing together finitely many tests, one for each element of a covering of the sieve by L^∞ balls. We generalize the result by considering the noise distribution has a symmetric (about 0) density function with an appropriate condition, which includes the case of Laplace distribution.

4.1.1 Prior positivity condition: (A1) *of Theorem* A.1

Here, we state and prove those results that allow us to verify condition (A1) of Theorem A.1.

Lemma 1 *Let $\theta = (\eta, \sigma)$ and $\theta_0 = (\eta, \sigma_0)$. Suppose that the conditional density of Y given $X = x$ is given by $p_\theta(y|x) = \frac{1}{\sigma}\phi\left(\frac{y-\eta(x)}{\sigma}\right)$, where the density ϕ satisfies* (2) *and* (3).

Define $\Lambda(\theta_0, \theta) = \log \frac{p_{\theta_0}(y_i|x_i)}{p_\theta(y_i|x_i)}$, $K_i(\theta_0, \theta) = \mathrm{E}_{\theta_0}(\Lambda(\theta_0, \theta)$ and $V_i(\theta_0, \theta) = \mathrm{Var}_{\theta_0}(\Lambda(\theta_0, \theta))$. For each $\delta > 0$, consider B_δ as defined in (5).

Then, for every $\epsilon > 0$, there exists $\delta > 0$ such that $\forall\theta \in B_\delta$, $K_i(\theta_0, \theta) < \epsilon$ for all i and $\sum_{i=1}^{\infty} \frac{V_i(\theta_0,\theta)}{i^2} < \infty$, $\forall\theta \in B_\delta$.

Proof We first calculate Kullback-Leibler divergence between two conditional densities of Y given $X = x$ when (2) holds,

$$\begin{aligned}
&K(p_{\theta_0}(y|x), p_\theta(y|x))\\
&\quad= \int \log \frac{p_{\theta_0}(y|x)}{p_\theta(y|x)} p_{\theta_0}(y|x)\mathrm{d}y\\
&\quad\le \log\frac{\sigma}{\sigma_0} + c\int \left|\frac{y-\eta_0(x)}{\sigma_0} - \frac{y-\eta(x)}{\sigma}\right| p_{\theta_0}(y|x)\mathrm{d}y\\
&\quad\le \log\frac{\sigma}{\sigma_0} + c\left|\frac{\sigma_0}{\sigma} - 1\right| \int |y|\phi(y)\mathrm{d}y + \frac{c}{\sigma}\sup_{x\in[0,1]} |\eta(x) - \eta_0(x)|.
\end{aligned}$$

Note that the last expression can be chosen arbitrarily small if $\sup_{x\in[0,1]} |\eta(x)-\eta_0(x)|$ and $|\sigma/\sigma_0 - 1|$ are sufficiently small. Therefore, for every $\epsilon > 0$, there exists $\delta > 0$ such that $\forall\theta \in B_\delta$, $K_i(\theta_0, \theta) < \epsilon$ for all i.

Second, we calculate the variance of $V_{i,n} = \mathrm{Var}_{\theta_0}(\Lambda(\theta_0, \theta))$ and show that the variance is uniformly bounded for all i. By the conditions (2) and (3), we have

$$\begin{aligned}
\mathrm{Var}_{\theta_0}(\Lambda(\theta_0, \theta)) &\le \int \left\{\log\phi\left(\frac{y-\eta_0(x)}{\sigma_0}\right) - \log\phi\left(\frac{y-\eta(x)}{\sigma}\right)\right\}^2 p_{\theta_0}(y|x)\mathrm{d}y\\
&\le c^2\int \left|\frac{y-\eta_0(x)}{\sigma_0} - \frac{y-\eta(x)}{\sigma}\right|^2 p_{\theta_0}(y|x)\mathrm{d}y\\
&\le c^2\left\{\left(\frac{\sigma_0}{\sigma} - 1\right)^2 + 2\sup_{x\in[0,1]} |\eta(x) - \eta_0(x)|\right\} \int |y|^2\phi(y)\mathrm{d}y\\
&\qquad + \frac{c^2}{\sigma}\sup_{x\in[0,1]} |\eta(x) - \eta_0(x)|^2 < \infty, \quad \forall\theta \in B_\delta.
\end{aligned}$$

□

Since we assume suitable prior distributions, Π_η and Π_ν, are assigned to satisfy two Assumptions, P.1 and P.2, it is easy to see that conditions (A1) and (iii) of (A2) in Theorem A.1 in the appendix hold under Laplace regression setup as well as other types additive regression with symmetric error distributions that satisfy (2) and (3). Thus, what remains for completing the proof is to show there exist test functions that meet conditions (A2), which will be presented in the next section.

4.1.2 Existence of uniformly consistent tests: (A2) *of Theorem* A.1

To verify (A2) of Theorem 1, we consider the sieve (6) that has been specified in the Assumption, P.2., and then construct a test for each element of the sieve. The nth test is constructed by combining a collection of tests, one for each of finitely many elements of the sieve, which come from a covering the sieve (6). The construction of tests are similar to the construction of tests under the normal noise done in Choi and Schervish (2007) by following the idea of Chernoff bounds (Chernoff 1952). The main difference from the normal case is the construction of test based on random noises from either Laplace distribution or a symmetric distribution with suitable conditions that will be given afterwards.

For this purpose, we consider three possible cases for the alternative hypotheses depending on the configuration of η and σ. As in the Gaussian error case of Choi and Schervish (2007), this is also done by piecing together finitely many tests, one for each element of a covering of the sieve by L^∞ balls. After tests have been constructed, the remaining steps are exact same as those in the Gaussian error case and we omit those steps and refer to previous results by Choi and Schervish (2007).

We construct those test functions in Lemmas 2–5.

The following relatively straightforward result is useful in the construction of tests.

Proposition 1 (a) *For every random variable X with unimodal distribution symmetric around 0 and every $c \in \mathbb{R}$,*

$$\Pr(|X| \le x) \ge \Pr(|X + c| \le x).$$

This is the special case of Anderson's theorem (Anderson 1955).
(b) *Let $X_1, \ldots, X_n$ be i.i.d random variables satisfying* (a). *Let $b_i \in \mathbb{R}$, $i = 1, \ldots, n$. Then, $\forall c \in \mathbb{R}$,*

$$\Pr\left(\sum_{i=1}^n |X_i| \le c\right) \le \Pr\left(\sum_{i=1}^n |X_i + b_i| \le c\right).$$

Proof (a)

$$\begin{aligned}
&\Pr\{|X| \le x\} - \Pr\{|X + c| \le x\}\\
&\quad = F(x) - F(-x) - F(x-c) + F(-x-c)\\
&\quad = \begin{cases} \{F(x) - F(x-c)\} - \{F(x+c) - F(x)\} \ge 0, & c \ge 0\\ \{F(x) - F(x+c)\} - \{F(x-c) - F(x)\} \ge 0, & c < 0 \end{cases}
\end{aligned}$$

(b) We argue by induction and use the law of total probability.

$$\begin{aligned}
\Pr\left\{\sum_{i=1}^{n}|X_i|\le c\right\} &= \mathrm{E}\left(\Pr\left\{|X_1|+\sum_{i=2}^{n}|X_i|\le c\,\middle|\,|X_2|=x_2,\ldots,|X_n|=x_n\right\}\right)\\
&\le \mathrm{E}\left(\Pr\left\{|X_1+b_1|\le c-\sum_{i=2}^{n}x_i\,\middle|\,|X_2|=x_2,\ldots,|X_n|=x_n\right\}\right)\\
&= \Pr\left\{|X_1+b_1|+\sum_{i=2}^{n}|X_i|\le c\right\}
\end{aligned}$$

Using the same argument as above but conditioning on $(|X_1+b_1|, |X_3|, \ldots, |X_n|)$, we get

$$\Pr\left\{\sum_{i=1}^{n}|X_i|\le c\right\}\le \Pr\left\{|X_1+b_1|+|X_2+b_2|+\sum_{i=3}^{n}|X_i|\le c\right\}$$

Similarly, by conditioning on $(|X_1+b_1|,\ldots,|X_{i-1}+b_{i-1}|,|X_{i+1}|,\ldots,|X_n|)$, $i=3,\ldots,n$, we reach the final result of part (b). $\square$

Lemma 2 *Let η_1 be a continuous function on $[0,1]$ and define $\eta_{ij}=\eta_i(x_j)$ for $i=0,1$ and $j=1,\ldots,n$. Let $\epsilon>0$, and let $r>0$. Let $c_n=n^{\tau_1}$ for $\alpha_1/2<\tau_1<1/2$ and $1/2<\alpha_1<1$. Let $b_j=1$ if $\eta_{1j}\ge\eta_{0j}$ and -1 otherwise. Let $\Psi_{1n}[\eta_1,\epsilon]$ be the indicator of the set A_1, where A_1 is defined as*

$$A_1=\left\{\sum_{j=1}^{n}b_j\left(\frac{Y_j-\eta_{0j}}{\sigma_0}\right)>2c_n\sqrt{n}\right\}.$$

Suppose that the conditional density of Y given $X=x$ is given by $p_\theta(y|x)=\frac{1}{\sigma}\phi\left(\frac{y-\eta(x)}{\sigma}\right)$, where the density ϕ admits (3).

Then there exists a constant $B_3>0$ such that for all η_1 that satisfy

$$\sum_{j=1}^{n}|\eta_{1j}-\eta_{0j}|>rn, \tag{7}$$

$\mathrm{E}_{P_0}(\Psi_{1n}[\eta_1,\epsilon])<\exp\left(-B_3c_n^2\right)$. *Also, there exist constants C_4 and C_5 such that for all sufficiently large n and all η satisfying $\|\eta-\eta_1\|_\infty<r/4$ and for all $\sigma\le\sigma_0(1+\epsilon)$,*

$$\mathrm{E}_P(1-\Psi_{1n}[\eta_1,\epsilon])\le C_4\exp(-nC_5),$$

where P is the joint distribution of $\{Y_n\}_{n=1}^{\infty}$ assuming that $\theta=(\eta,\sigma)$.

Proof (1) Type I error:
For all $0 < t < 1$, by the Markov inequality,

$$\begin{aligned}
\mathrm{E}_{P_0}(\Psi_{1n}[\eta_1, \epsilon]) &= P_0 \left\{ \sum_{j=1}^{n} b_j \left(\frac{Y_j - \eta_{0j}}{\sigma_0} \right) > 2c_n\sqrt{n} \right\} \\
&\le \exp\bigl(-t \cdot 2c_n\sqrt{n}\bigr)\, m(t)^n \\
&= \exp\left(\left[-2t\frac{c_n}{\sqrt{n}} + \log\left\{1 + at^2 + o(t^2)\right\} \right] \cdot n \right) \\
&\le \exp\left(\left[-2t\frac{c_n}{\sqrt{n}} + at^2 + o(t^2) \right] \cdot n \right)
\end{aligned}$$

Take $t = \dfrac{c_n}{a\sqrt{n}}$. Then, $\mathrm{E}_{P_0}(\Psi_{1n}) = \exp\left\{ \left[-2\dfrac{c_n^2}{an} + \dfrac{c_n^2}{an} + o\left(\dfrac{c_n^2}{a^2 n}\right) \right] \cdot n \right\} \le \exp\left\{ -\dfrac{c_n^2}{a}[1 - o(1)] \right\}$. This holds for sufficiently large n.

(2) Type II error:
As in the Type II error calculation for the Gaussian case in Choi and Schervish (2007), first, assume that n is large enough so that $c_n/\sqrt{n} < r/(4\sigma_0)$. Let $\eta_{*j} = \eta(x_j)$ for $j = 1, \ldots, n$. Since $\sigma \le (1+\epsilon)\sigma_0$, then for all $-1 < t < 0$,

$$\begin{aligned}
&\mathrm{E}_P(1 - \Psi_n[\eta_1, \epsilon]) \\
&\quad \le \mathrm{E}_P(1 - \Psi_{1n}) = P \left\{ \sum_{j=1}^{n} b_j \left(\frac{Y_j - \eta_{0j}}{\sigma_0} \right) \le 2c_n\sqrt{n} \right\} \\
&\quad = P \left\{ \sum_{j=1}^{n} b_j \left(\frac{Y_j - \eta_{*j}}{\sigma_1} \right) + \sum_{j=1}^{n} b_j \left(\frac{\eta_{*j} - \eta_{1j}}{\sigma} \right) \right. \\
&\qquad \left. + \sum_{j=1}^{n} \left| \frac{\eta_{1j} - \eta_{0j}}{\sigma} \right| \le 2c_n\sqrt{n}\frac{\sigma_0}{\sigma} \right\} \\
&\quad \le P \left\{ \sum_{j=1}^{n} b_j \left(\frac{Y_j - \eta_{*j}}{\sigma} \right) \le \frac{rn}{4\sigma} - \frac{rn}{\sigma} + 2c_n\sqrt{n}\frac{\sigma_0}{\sigma} \right\} \\
&\quad \le P \left\{ \sum_{j=1}^{n} b_j \left(\frac{Y_j - \eta_{*j}}{\sigma} \right) \le \frac{-rn}{4\sigma_0(1+\epsilon)} \right\} \\
&\quad = P \left\{ t\sum_{j=1}^{n} b_j \left(\frac{Y_j - \eta_{*j}}{\sigma} \right) \ge t\frac{-rn}{4\sigma_0(1+\epsilon)} \right\} \\
&\quad \le \exp\left(-t\frac{-rn}{4\sigma_0(1+\epsilon)} \right) (1 + at^2 + o(t^2))^n
\end{aligned}$$

$$\leq \exp\left(n\left[\frac{tr}{4\sigma_0(1+\epsilon)} + at^2 + o(t^2)\right]\right)$$
$$= \exp\left(nt\left[\frac{r}{4\sigma_0(1+\epsilon)} + at + o(t)\right]\right)$$

Then, there exists t^* such that $-T < t^* < 0$ and $-t^*\left(\frac{r}{4\sigma_0(1+\epsilon)} + at^* + o(t^*)\right) = C^* > 0$. Thus, the last expression above is less than $\exp(-nC^*)$. □

As stated in Lemma 2, the test construction depends on the existence of the quantity in (7) under L^1 toplogy of the regression function. Lemma 3, asserts its existence under the assumption D.1. for the fixed covariates.

Lemma 3 *For each integer n, let A_n be the set of all continuously differentiable functions η such that $\|\eta\|_\infty < n^{3/4}$ and $\|\eta'\|_\infty < n^{3/4}$. Then for each $\epsilon > 0$ there exist an integer N and $r > 0$ such that, for all $n \geq N$ and all $\eta \in A_n$ such that $\int_0^1 |\eta(x) - \eta_0(x)|dx > \epsilon$, $\sum_{i=1}^n |\eta(x_i) - \eta_0(x_i)| \geq rn$.*

Proof It follows from Lemmas 3 and 7 of Choi and Schervish (2007). □

Lemma 4 *Suppose that the conditional density of Y given $X = x$ is given by $p_\theta(y|x) = \frac{1}{\sigma}\phi\left(\frac{y-\eta(x)}{\sigma}\right)$, where the density $\phi(z)$ admits that in some interval $(-T, T)$ with $T > 0$,*

$$\int \exp(t|y|)\phi(y)\mathrm{d}y = 1 + bt + o(t)$$

as $t \to 0$ for some $b > 0$. Let η_1 be a continuous function on $[0, 1]$ and define $\eta_{ij} = \eta_i(x_j)$ for $i = 0, 1$ and $j = 1, \ldots, n$. Let Ψ_{2n} be the indicator of the set A_2, where

$$A_2 = \left\{\sum_{j=1}^n \left|\frac{Y_j - \eta_{0j}}{\sigma_0}\right| > nb\sqrt{1+\epsilon}\right\},$$

Then there exists a constant C_6 such that for all η_1 $\mathrm{E}_{P_0}(\Psi_{2n}) < \exp(-nC_6)$. Also, there exist constants C_7 such that for all sufficiently large n and all η satisfying $\|\eta - \eta_1\|_\infty < r/4$ and for all $\sigma > \sigma_0(1+\epsilon)$,

$$\mathrm{E}_P(1 - \Psi_{2n}[\epsilon]) \leq \exp(-nC_7),$$

where P is the joint distribution of $\{Y_n\}_{n=1}^\infty$ assuming that $\theta = (\eta, \sigma)$.

Proof (1) Type I error:

For sufficiently small $0 < t_1 < T$, by the Markov inequality, we have

$$\begin{aligned}
\mathrm{E}_{P_0}(\Psi_{2n}) &= P_0\left(\sum_{j=1}^{n}\left|\frac{Y_j-\eta_{0j}}{\sigma_0}\right| > nb\sqrt{1+\epsilon}\right)\\
&\le \exp\Big\{-n\left(b\sqrt{1+\epsilon})t_1\right)+n\left(bt_1+o(t_1)\right)\Big\}\\
&= \exp\Big\{-nt_1b\left(\sqrt{1+\epsilon}-1-o(t_1)\right)\Big\}.
\end{aligned}$$

Then, there exists $0 < t_1^* < T$ such that $t_1^* \to 0$ and $t_1^* b(\sqrt{1+\epsilon}-1-o(t_1^*)) = C_6 > 0$. Therefore, $\mathrm{E}_{P_0}(\Psi_n) \le \exp(-C_6 n)$.

(2) Type II error:

Since $\sigma > (1+\epsilon)\sigma_0$, for all $-T < t < 0$,

$$\begin{aligned}
\mathrm{E}_P(1-\Psi_{2n}) &= P\left\{\sum_{j=1}^{n}\left|\frac{Y_j-\eta_{0j}}{\sigma_0}\right| \le nb\sqrt{1+\epsilon}\right\}\\
&= P\left\{\sum_{j=1}^{n}\left|\left(\frac{Y_j-\eta_{*j}}{\sigma}\right)+\left(\frac{\eta_{*j}-\eta_{0j}}{\sigma}\right)\right| \le n\frac{\sigma_0}{\sigma}b\sqrt{1+\epsilon}\right\}\\
&\le P\left\{\sum_{j=1}^{n}\left|\frac{Y_j-\eta_{*j}}{\sigma}\right| \le n\frac{\sigma_0}{\sigma}b\sqrt{1+\epsilon}\right\}\\
&\le \exp\left\{-\frac{nbt}{1+\sqrt{\epsilon^*}}+n(bt+o(t))\right\}\\
&= \exp\left\{nbt\left(-\frac{1}{1+\sqrt{\epsilon^*}}+1+o(t)\right)\right\}. \qquad (8)
\end{aligned}$$

Thus, there exists a constant C_7 such that $\mathrm{E}_P(1-\Psi_{2n}) \le \exp(-C_7 n)$. □

Lemma 5 *Suppose that the conditional density of* Y *given* $X = x$ *is given by* $p_\theta(y|x) = \frac{1}{\sigma}\phi\left(\frac{y-\eta(x)}{\sigma}\right)$, *where the density* $\phi(z)$ *admits that in some interval* $(-T, T)$ *with* $T > 0$,

$$\int \exp(t|y|)\phi(y)\mathrm{d}y = 1 + bt + o(t)$$

as $t \to 0$ *for some* $b > 0$. *Let* η_1 *be a continuous function on* $[0, 1]$, *and define* $\eta_{ij} = \eta_i(x_j)$ *for* $i = 0, 1$ *and* $j = 1, \ldots, n$. *Let* $\epsilon > 0$, *and* $0 < r < 4\sigma_0\sqrt{\epsilon-\epsilon^2}$. *Let* $\Psi_{3n}[\eta_1, \epsilon]$ *be the indicator of the set* A_3, *where*

$$A_3 = \left\{\sum_{j=1}^{n}\left|\frac{Y_j-\eta_{1j}}{\sigma_0}\right| < nb\sqrt{1-\epsilon^2}\right\},$$

Then there exists a constant C_8 such that for all η_1 that satisfy

$$\sum_{j=1}^{n} \|\eta_{1j} - \eta_{0j}\| < rn$$

$E_{P_0}(\Psi_{3n}[\eta_1, \epsilon]) < \exp(-nC_8)$ *Also, there exist a constants C_9 such that for all sufficiently large n and all η and σ satisfying $\|\eta - \eta_1\|_\infty < r/4$ and $\sigma < \sigma_0(1-\epsilon)$,*

$$\mathrm{E}_P(1 - \Psi_{3n}[\eta_1, \epsilon]) \leq \exp(-nC_9),$$

where P is the joint distribution of $\{Y_n\}_{n=1}^{\infty}$ assuming that $\theta = (\eta, \sigma)$.

Proof (1) Type I error:
For all $t < 0$, by the Markov inequality,

$$\begin{aligned}
\mathrm{E}_{P_0}(\Psi_{3n[\eta_1,\epsilon]}) &= P_0\left(\sum_{j=1}^{n} \left|\frac{Y_j - \eta_{1j}}{\sigma_0}\right| < nb\sqrt{1-\epsilon^2}\right) \\
&\leq \Pr\left(\sum_{j=1}^{n} \left|\frac{Y_j - \eta_{0j}}{\sigma_0}\right| < nb\sqrt{1-\epsilon^2}\right), \quad \text{by Proposition 1} \\
&\leq \exp\Big(-nb(\sqrt{1-\epsilon^2})t + n(bt + o(t))\Big) \\
&= \exp\Big\{ntb\left(-\sqrt{1-\epsilon^2} + 1 + o(1)\right)\Big\}
\end{aligned}$$

Then, it is clear that there exists a constant $C_9 > 0$ such that $\mathrm{E}_{P_0}(\Psi_{3n[\eta_1,\epsilon]}) \leq \exp(-C_9 n)$.

(2) Type II error:
For all sufficiently small t such that $t \in (0, T]$,

$$\begin{aligned}
&\mathrm{E}_P(1 - \Psi_{3n[\eta_1,\epsilon]}[\eta_1, \epsilon]) \\
&\quad = P\left\{nb\sqrt{1-\epsilon^2} \leq \sum_{j=1}^{n} \left|\frac{Y_j - \eta_{1j}}{\sigma_0}\right|\right\} \\
&\quad = P\left\{n\frac{\sigma_0}{\sigma}b\sqrt{1-\epsilon^2} \leq \sum_{j=1}^{n} \left|\frac{Y_j - \eta_{*j}}{\sigma} + \frac{\eta_{*j} - \eta_{1j}}{\sigma}\right|\right\} \\
&\quad \leq \Pr\left(n\frac{\sigma_0}{\sigma}b\sqrt{1-\epsilon^2} - \sum_{j=1}^{n} \left|\frac{\eta_{*j} - \eta_{1j}}{\sigma}\right| \leq \sum_{j=1}^{n} \left|\frac{Y_j - \eta_{*j}}{\sigma}\right|\right) \\
&\quad \leq \exp\left(n(bt + o(t)) - nbt\sqrt{1-\epsilon^2}\frac{\sigma_0}{\sigma} + t\sum_{j=1}^{n} \left|\frac{\eta_{*j} - \eta_{1j}}{\sigma}\right|\right)
\end{aligned}$$

If $\|\eta_1 - \eta\|_\infty < r/4$, then $|\eta_{1j} - \eta_{*j}| < r/4, \quad \forall j = 1, 2, \ldots$. Therefore,

$$\exp\left(n(bt + o(t)) - nbt\sqrt{1-\epsilon^2}\frac{\sigma_0}{\sigma} + t\sum_{j=1}^{n}\left|\frac{\eta_{*j} - \eta_{1j}}{\sigma}\right|\right) \leq \exp\left\{n\frac{bt\sigma_0}{\sigma}\left[(1+o(1))(1-\epsilon) + \frac{r}{4\sigma_0} - \sqrt{1-\epsilon^2}\right]\right\} \tag{9}$$

Because $r < 4\sigma_0(\epsilon - \epsilon^2)$, (9) is less than

$$\exp\left\{nbt(1-\epsilon)\frac{\sigma_0}{\sigma}\left[1 + o(1) + \epsilon - \frac{\sqrt{1+\epsilon}}{\sqrt{1-\epsilon}}\right]\right\}$$

Note that $\epsilon - \frac{\sqrt{1+\epsilon}}{\sqrt{1-\epsilon}} < -1$ for all $\epsilon > 0$. Thus, there exits a constant C_9 such that

$$\mathrm{E}_P(\Psi_{3n[\eta_1,\epsilon]}) \leq \exp\{-C_9 n\}\,.$$

□

In order to apply those test functions constructed in Lemmas 2–5 to the cases of testing for $(\eta, \sigma) = (\eta_0, \sigma_0)$ against $(\eta_1, \sigma_1) \in L_\epsilon$, where

$$L_\epsilon = \left\{(\eta, \sigma)\ :\ \int |\eta - \eta_0| \mathrm{d}Q > \epsilon, \left|\frac{\sigma}{\sigma_0} - 1\right| > \epsilon\right\},$$

we need to verify that if $(\eta_1, \sigma_1) \in L_\epsilon$, (7) holds, which is a key requirement for constructing test in Lemma 2. Note that test functions constructed in Lemmas 4 and 5 are based on the discrepancy in terms of σ rather than η. This verification has been studied for d-dimension nonrandom covariate with suitable assumptions and d-dimensional random covariates in Choi and Schervish (2007), independently of the error distribution. We also provide such results in the next section. On the other hand, those results have also been specialized as the result by Choi (2007) for the case of one dimensional fixed covariate under weaker assumption that the uniform bound of the derivatives is not necessary. Finally, the test functions are considered here for specific choice of (η_1, σ_1) and this dependence will be removed by utilizing the same sieve as considered in Choi and Schervish (2007) because of the Assumption D.1. Therefore, the existence of tests is proven. Hence, the proof of Theorem 1 is complete.

4.2 Proof of Theorem 2

As discussed in Schwartz (1965), Ghosh and Ramamoorthi (2003) and Tokdar and Ghosh (2007), for weak consistency, it is sufficient to show a Kullback-Leibler support condition on Π and f_0, $\forall\epsilon > 0$, $\Pi(f\ :\ K(f_0, f) < \epsilon) > 0$, where f is the joint density of (X, Y), $f = p(y|x, \eta, \sigma)q(x)$ and f_0 is the true joint density of (X, Y),

$f_0 = p(y|x, \eta_0, \sigma_0)q(x)$. In Sect. 4.1.1, we have already shown the Kullback-Leibler support condition for conditional density, $p(y|x, \eta_0, \sigma_0)$. Since the Kullback-Leibler divergence between two joint densities is the integration of the Kullback-Leibler divergence between two conditional densitites and the marginal density of X is assumed to be known, by the Assumption, P.1., it is shown that the Kullback-Leibler support condition holds. Hence, weak posterior consistency is achieved.

4.3 Proof of Theorem 3

First, assuming the noise distribution is the Laplace distribution, we calculate the Hellinger distance between two density functions, $d_H(f, f_0)$. To simplify the calculation, we consider the quantity $h(f, f_0)$ defined as

$$h(f, f_0) = \frac{1}{2} d_H^2(f, f_0) = 1 - \int \sqrt{f f_0} \mathrm{d}\mu$$

and $h(f, f_0)$ is calculated as follows.

$$\begin{aligned} h(f, f_0) &= 1 - \frac{1}{\sqrt{4\sigma\sigma_0}} \int\int \exp\left\{-\frac{1}{2\sigma}|y - \eta(x)| - \frac{1}{2\sigma_0}|y - \eta_0(x)|\right\} \mathrm{d}y\mathrm{d}Q \\ &\leq 1 - \int\int \exp\left\{-\left(\frac{1}{2\sigma} + \frac{1}{2\sigma_0}\right)\left|y - \left(\frac{\eta(x) + \eta_0(x)}{2}\right)\right|\right\} \\ &\quad \times \frac{1}{\sqrt{4\sigma\sigma_0}} \times \exp\left\{-\left(\frac{1}{4\sigma} + \frac{1}{4\sigma_0}\right)|\eta(x) - \eta_0(x)|\right\} \mathrm{d}y\mathrm{d}Q \\ &\leq 1 - \int \left[\frac{1}{\sqrt{4\sigma\sigma_0}} \times \exp\left\{-\left(\frac{1}{4\sigma} + \frac{1}{4\sigma_0}\right)|\eta(x) - \eta_0(x)|\right\}\right. \\ &\quad \left. \times \left(\frac{1}{4\sigma} + \frac{1}{4\sigma_0}\right)^{-1}\right] \mathrm{d}Q \end{aligned} \tag{10}$$

The integral in (10) is of the form $\int c_1 \exp(-c_2|\eta(x) - \eta_0(x)|)\mathrm{d}Q(x)$, where c_1 can be made arbitrarily close to 1 by choosing $|\sigma/\sigma_0 - 1|$ small enough and c_2 is bounded when σ is close to σ_0. Similarly to (a), it follows that for each ϵ there exists a δ such that (10) will be less than ϵ whenever $|\sigma/\sigma_0 - 1| < \delta$ and $d_Q(\eta, \eta_0) < \delta$, where

$$d_Q(\eta, \eta_0) = \inf\{\epsilon \ : \ Q(\{x : |\eta(x) - \eta_0(x)| > \epsilon\}) < \epsilon\}.$$

Thus, it suffices to show that the posterior is consistent in terms of joint neighborhood based on d_Q metric of the regression function when the noise distribution is assumed to be the Laplace distribution.

In addition, when we assume other noise distribution than the Laplace distribution, that satisfies (2) and 3), it is shown that posterior consistency in terms of d_Q metric, called *the* d_Q *consistency* is sufficient to achieve Hellinger consistency due to the following inequality between Kullback-Leibler divergence and the Hellinger metric:

$$\frac{H^4(f, f_0)}{4} \leq K(f, f_0),$$

which are from Corollary 1.2.1 and Proposition 1.2.2 of Ghosh and Ramamoorthi (2003). Also, note that from the conditions (2) and (3), we have

$$K(f, f_0) \leq \log \frac{\sigma}{\sigma_0} + c \left| \frac{\sigma_0}{\sigma} - 1 \right| \int |y|\phi(y)\mathrm{d}y + \frac{c}{\sigma} \int |\eta(x) - \eta_0(x)|\mathrm{d}x. \quad (11)$$

Let $\delta > 0$, Then, the integral in (11) can be written as

$$\begin{aligned}
&\int |\eta(x) - \eta_0(x)|\mathrm{d}x \\
&= \int_{|\eta - \eta_0(x)| \leq \delta} |\eta(x) - \eta_0(x)|\mathrm{d}Q(x) + \int_{|\eta(x) - \eta_0(x)| > \delta} |\eta(x) - \eta_0(x)|\mathrm{d}Q(x) \\
&\leq \delta + 2MQ(x : |\eta(x) - \eta_0(x)| > \delta),
\end{aligned}$$

where M is the uniform bound stated in the Assumption P.3.

Therefore, it follows that for each $\epsilon > 0$, there exists a $\delta > 0$ such that $H(f, f_0)$ is less that ϵ whenever $|\sigma/\sigma_0 - 1| < \delta$ and $Q(x : |\eta(x) - \eta_0(x)| > \delta) < \delta$, which means that the d_Q consistency is sufficient to show the Hellinger consistency.

To prove the d_Q consistency, we make use of the same techniques as in the proof of L^1 consistency under fixed covariates of Theorem 1 and need to modify such techniques to the case of random covariates. Specifically, when we deal with the case of random covariates, we first condition on the observed values of the covariate. Then, test functions are constructed the same as in the previous lemmas in Sect. 4.1, by understanding all probability statements regarding the test constructions in the previous lemmas as conditional on the covariate values $X_1 = x_1, \ldots, X_n = x_n$. Thus, Lemma Thus, we only need to show that the quantity in (7) occurs all but finitely often with probability 1 as in the Lemma 6.

Lemma 6 *Assume the covariate values are sampled from a probability distribution Q and the prior satisfies the Assumption* P.3. *Let η be a function such that $\int_0^1 |\eta(x) - \eta_0(x)|\mathrm{d}Q(x) > \epsilon$. Let $0 < r < \epsilon^2$, and define*

$$A_n = \left\{ \sum_{i=1}^{n} |\eta(X_i) - \eta_0(X_i)| \geq rn \right\}.$$

Then there exists a constant $C_{11} > 0$ such that $\Pr(A_n^C) \leq \exp(-C_{11}n)$ for all n and A_n occurs all but finitely often with probability 1. *The same C_{11} works for all η such that $\int_0^1 |\eta(x) - \eta_0(x)|\mathrm{d}Q(x) > \epsilon$.*

Proof Note that the regression functions are uniformly bounded by the Assumption P.3. Thus, the proof is made by the Bernstein's inequality. □

Now, Lemma 7 provides posterior consistency based on the d_Q metric.

Lemma 7 *Suppose that the values of the covariate in* $[0, 1]$ *are sampled from a probability distribution* Q. *Suppose that Assumptions* P.1. *and* P.2. *hold. Let* P_0 *be the joint distribution of* $\{(X_n, Y_n)\}_{n=1}^{\infty}$ *assuming that* (η_0, σ_0) *is the true value of parameter and the true joint density of* (X, Y) *is* $f_0(x, y) = p(y|x, \eta_0, \sigma_0)q(x)$. *Then for every* $\epsilon > 0$,

$$\Pi\left\{(\eta, \sigma) : \left|\frac{\sigma}{\sigma_0} - 1\right| < \epsilon, d_Q(\eta, \eta_0) < \epsilon \left|\{(Y_i, X_i)\}_{i=1}^{n}\right.\right\} \to 1, \; a.s. \, [P_0]. \quad (12)$$

Proof The prior positivity condition, (A1) in Theorem A.1, has been already verified in Theorem 1, which holds regardless of the feature of covariates. The existence of uniformly consistent tests, the condition of (A2), are shown from Lemma 6 and Lemmas 2–5 in the proof of Theorem 1. Hence, the proof is complete. □

To show L^1 consistency, first we need to notice the convergence of random sequences in terms of d_Q metric is equivalent to the in-probability convergence of random sequences (Choi, 2005, Lemma 3.2.1). Since we have the d_Q consistency from Lemma 7 and the regression functions are assumed to be uniformly bounded in the Assumption P.3, L^1 consistency is achieved.

5 Supplementary results

In the previous section, we provided posterior consistency results for the regression function and the scale parameter under various situations. In this section, we examine additional issues that are worth further consideration.

5.1 Multi-dimensional covariates

First of all, we consider the case of multi-dimensional covariates, i.e. $\boldsymbol{x} = (x_1, \ldots, x_d) \in [0, 1]^d$ for $d \geq 2$. Up to this point, we assumed that the covariate is one dimensional, but much concern also lies in multi-dimensional framework. Particularly, in terms of posterior consistency, the results of Ghosal and Roy (2006), Tokdar and Ghosh (2007) and Choi and Schervish (2007) are demostrated in the higher dimensions as well under a certain topology of multidimensional probability functions or densities.

In general, the main difficulty in dealing with multidimensional regression function lies in the so called phenomenon, the "curse of dimensionality", and this problem also affects the posterior consistency and makes the results less promising. As a result, stronger assumptions on design points and the regression functions are required, or different topologies in the parameter space can be considered. For example, Ghosal and Roy (2006) treated L^1 consistency of a probability function in one dimension with an assumption about the fixed design points, while in higher dimensions, they considered the consistency with respect to the empirical measure of the design points, which was also used in (Ghosal and van der Vaart, 2007, Sect. 7.7) in the regression problems for the rate of convergence. Choi and Schervish (2007) considered posterior consistency of nonparametric regression problems with Gaussian errors with multi-dimensional covariates with stronger assumptions on the regression function that those for the case of a one dimensional regression function, but still the equivalent condition about the covariates to that for one dimensional covariate. We follow the same condition as in Choi and Schervish (2007), stated as follows:

Assumption P.2d Define

$$E_{n,d} = \left\{ \eta \ : \ \sup_{\boldsymbol{x} \in [0,1]^d} |\eta(\boldsymbol{x})| < n^{3/4}\ , \quad \sup_{\boldsymbol{x} \in [0,1]^d} \left| \frac{\partial}{\partial x_i} \eta(\boldsymbol{x}) \right| < n^{3/4},\ i = 1, \ldots, d \right\},$$

and assume that there exists a constant $\beta > 0$ such that $\Pi_\eta \left\{ E^C_{n,d} \right\} \leq \exp(-n\beta)$.

Assumption D.1d For each hypercube H in $[0, 1]^d$, let $\lambda(H)$ be its Lebesgue measure. Suppose that there exists a constant K_d, $0 < K_d \leq 1$ such that whenever $\lambda(H) \geq \frac{1}{K_d n}$, H contains at least one design point.

Note that the assumption D.1d is a natural generalization of the Assumption D.1. into d-dimensional covariate.

Assumption P.3d Let Π'_η be the prior for η satisfying P.1. for d-dimensional function and P.2d. Let $\Omega = \left\{ \eta : \sup_{\boldsymbol{x} \in [0,1]} \left| \frac{\partial}{\partial x_j} \eta(\boldsymbol{x}) \right| < V,\ j = 1, \ldots, d \right\}$ with $V > \sup_{\boldsymbol{x} \in [0,1]} \left| \frac{\partial}{\partial x_j} \eta_0(\boldsymbol{x}) \right|$, $j = 1, \ldots, d$. Assume that $\Pi_\eta(\cdot) = \Pi'_\eta(\cdot \cap \Omega)/\Pi'_\eta(\Omega)$ with $\Pi'_\eta(\Omega) > 0$.

When we dealt with one dimensional regression function, the assumption P.3d was unnecessary to achieve L^1 consistency based on fixed covariates. However, this condition is required for d-dimensional fixed covariates, in order to obtain (eq:test1). Notice that this is irrespective of the feature of noise distribution, only depending on the nature of covariates (see, Choi and Schervish 2007).

Therefore, based on these modified assumptions, it is easy to achieve the same results of posterior consistency that have been shown in the previous sections. Note that under such assumptions, every calculation and verification can be done with the exact same manner as in the one dimensional case.

5.2 Bayes estimates: predictive approach

Another issue that we can think about is the consistency of Bayes estimates, or existence of consistent Bayes estimator from the predictive point of view. Generally speaking, it is known that Bayes predictive estimates inherit the convergence property of the posterior (Ghosh and Ramamoorthi 2003, Proposition 4.2.1). Here we consider this issue in detail. First, we focus on the case where we are interested in estimating the conditional density, $p(y|x)$, assuming $Q(x)$ is known. Then the most frequently used Bayes estimator is the predictive density of f_0, obtained by

$$\hat{p}_n(y|x) = \int p(y|x, \theta) \mathrm{d}\Pi(\theta | \{(Y_i, X_i)\}_{i=1}^n). \tag{13}$$

Let $\eta_0(x) = \mathrm{E}_{p_0}[Y|X = x] = \int y p_0(y|x)\mathrm{d}y$ be the true regression function and denote $\hat{\eta}_n(x) = \mathrm{E}_{\hat{p}_n}[Y|X = x]$ to be the predictive regression function, or a Bayes estimate of the regression function. Also, we consider the predictiive probability, $\hat{P}_n^Y(A|X) = \mathrm{E}_{\hat{p}_n} I_A(Y)|X)$.

The existence of consistent (predictive) Bayes estimators can be shown in the following theorem.

Theorem 4 *Suppose that $\{(Y_i, X_i)\}_{i=1}^n$ is an i.i.d. random sample from density f_0 and all of assumptions in Theorem 3 hold, so that the posterior distribution is consistent with respect to Hellinger metric as in Theorem 3. Then, (predictive) Bayes estimates are consistent in that*

1. $d_H(p_0, \hat{p}_n) \to 0,\ in\ P_0^n\ probability$, *where $d_H(p, p_0)$ is the Hellinger metric.*
2. $\int \left(\hat{P}_n^Y(A|x) - P_0^Y(A|x)\right)^2 \mathrm{d}Q(x) \to 0,\ in\ P_0^n\ probability.$
3. *Assuming the scale parameter is known as σ_0, we have*

$$\int \left(\hat{\eta}_n(x) - \eta_0(x)\right)^2 \mathrm{d}Q(x) \to 0, \quad in\ P_0^n\ probability.$$

Proof Similar results and proofs can be found in Ge and Jiang (2006), who considered the consistency of Bayes estimates under logistic regression structures.

1. Let $\epsilon > 0$ and define $A_\epsilon = \{f\ :\ d_H(p, p_0) \le \epsilon\}$. Using Jensen's inequality, we have

$$d_H^2(p_0, \hat{p}_n) \le \epsilon^2 + 4\Pi\left(A_\epsilon^C|\{(Y_i, X_i)\}_{i=1}^n\right).$$

   This result was proved in Ge and Jiang (2006), and a similar result was also proved in Shen and Wasserman (2001) for the rate of convergence.
2. Using Hölder's inequality, we have

$$\begin{aligned}
&\int \left(\hat{P}_n^Y(A|x) - P_0^Y(A|x)\right)^2 \mathrm{d}Q(x)\\
&= \int \left[\int I_A(y)(\hat{p}_n - p_0)\right]^2 \mathrm{d}Q(x)\\
&= \int \left[I_A(y)\left(\sqrt{\hat{p}_n} + \sqrt{p_0}\right)\left(\sqrt{\hat{p}_n} - \sqrt{p_0}\right)\mathrm{d}y\right]^2 \mathrm{d}Q(x)\\
&\le \int \left[\int I_A(y)^2\left(\sqrt{\hat{p}_n} + \sqrt{p_0}\right)^2 \mathrm{d}y\right]\left[\int \left(\sqrt{\hat{p}_n} - \sqrt{p_0}\right)\mathrm{d}y\right]^2 \mathrm{d}Q(x)
\end{aligned}$$

   Note that

$$\int I_A(Y)^2\left(\sqrt{\hat{p}_n} + \sqrt{p_0}\right)^2 \mathrm{d}y \le 2\int I_A(Y)^2(\hat{p}_n + p_0)\mathrm{d}y < \infty,$$

   since $\int I_A(Y)^2\hat{p}_n\mathrm{d}y \le \int\int p(y|x)\mathrm{d}y\mathrm{d}\Pi(\theta|\{(Y_i, X_i)\}_{i=1}^n) = 1$ due to the Fubini's theorem. Therefore, (i) implies (ii).

3. Similarly in (ii), we use Hölder's inequality to get

$$
\begin{aligned}
&\int \left(\hat{\eta}_n - \eta_0\right)^2 \mathrm{d}Q(x) \\
&\quad = \int \left[\int y(\hat{p}_n - p_0)\right]^2 \mathrm{d}Q(x) \\
&\quad \le \int \left[\int y^2 \left(\sqrt{\hat{p}_n} + \sqrt{p_0}\right)^2 \mathrm{d}y\right] \left[\int \left(\sqrt{\hat{p}_n} - \sqrt{\hat{p}_0}\right) \mathrm{d}y\right]^2 \mathrm{d}Q(x)
\end{aligned}
$$

In addition, using Fubini's theorem, we have

$$
\begin{aligned}
&\int y^2 \left(\sqrt{\hat{p}_n} + \sqrt{p_0}\right)^2 \mathrm{d}y \\
&\quad \le 2\int y^2(\hat{p}_n + p_0)\mathrm{d}y \\
&\quad = 2\sigma_0^2 \int \frac{(y - \eta + \eta)^2}{\sigma_0^2} p(y|x)\mathrm{d}y\mathrm{d}\Pi(\theta|\{(Y_i, X_i)\}_{i=1}^n) \\
&\qquad + 2\sigma_0^2 \int \frac{(y - \eta_0 + \eta_0)^2}{\sigma_0^2} p_0(y|x)\mathrm{d}y \\
&\quad = 2\sigma_0^2 + 2\eta^2 + 2\sigma_0^2 + 2\eta_0^2 < \infty,
\end{aligned}
$$

which are bounded for all $\eta \in \Omega$ as defined in the Assumption P.3. Therefore, (i) implies (iii), too. □

6 Conclusions

In this paper, we have studied asymptotic properties of posterior distributions of nonparametric regression problems when the noise distribution is assumed to be symmetric non-Gaussian with suitable conditions that include the case of the Laplace distribution. For posterior consistency, we could verify two sufficient conditions of Theorem A.1. in the appendix, under non-Gaussian noise distribution and achieve almost sure consistency of posterior distributions. In order to construct the uniformly consistent tests, we could use the similar test functions to the normal regression problem considered in Choi and Schervish (2007) and calculate the suitable type I and type II errors for those test functions under the symmetric non-Gaussian distribution. In our approach, the noise is assumed to be symmetric with a specific form of density that satisfies certain conditions. It would be worth while to consider more general cases when the distribution of noise is unknown but still can be taken to be a random symmetric density function. Then a prior distribution for this unknown noise distribution needs to be specified, and this involves another nonparametric Bayesian inference for the unknown noise distribution as well as the unknown regression function. An alternative approach to handling this unknown error distribution is to fix the error distribution as a known distribution, although the true unknown error distribution is different from the distribution that has been fixed. This is commonly referred as misspecification of error distribution. It would be interesting to further investigate the effect of misspecified error distribution when the error distribution is misspecified to be Laplace whereas the true distribution is unknown, which has been studied in Kleijn and van der Vaart (2006). Finally, another open issue that are worth further consideration but have not been studied in this paper is to see if our formulation can be extended to the conditional median in non-symmetric case. To achieve posterior consistency under these structures, we see the main difficulty lies in constructing the uniformly consistent tests, which depends on the formulation of the model structure. Future work should be directed at alternative approaches to constructing uniformly consistent tests and posterior consistency, using similar techniques to those in Barron (1989), Barron et al. (1999) and Walker (2004).

Acknowledgments We are grateful to an anonymous referee and Associate Editor for their careful reading of the paper and constructive comments. We also thank Mark J. Schervish for helpful advice and discussions in the first draft of the paper.

References

Amewou-Atisso, M., Ghosal, S., Ghosh, J. K., Ramamoorthi, R. V. (2003). Posterior consistency for semiparametric regression problems. *Bernoulli*, *9*, 291–312.

Anderson, T. W. (1955). The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities. *Proceedings of the American Mathematical Society*, *6*, 170–176.

Barron, A. R. (1989). Uniformly powerful goodness of fit tests. *The Annals of Statistics*, *17*, 107–124.

Barron, A. R., Schervish, M. J., Wasserman, L. (1999). The consistency of posterior distributions in nonparametric problems. *The Annals of Statistics*, *27*, 536–561.

Chernoff, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on a sum of observations. *The Annals of Mathematical Statistics*, *23*, 493–507.

Choi, T. (2005). Posterior consistency in nonparametric regression problems under Gaussian process priors. Ph.D. Thesis, Carnegie Mellon University, Pittsburgh.

Choi, T. (2007). Alternative posterior consistency results in nonparametric binary regression using Gaussian process priors. *Journal of Statistical Planning and Inference*, *137*, 2975–2983.

Choi, T., Schervish, M. J. (2007). On posterior consistency in nonparametric regression problems. *Journal of Multivariate Analysis*, *98*, 1969–1987.

Choudhuri, N., Ghosal, S., Roy, A. (2004). Bayesian estimation of the spectral density of a time series. *Journal of the American Statistical Association*, *99*, 1050–1059.

Diaconis, P., Freedman, D. (1986). On inconsistent Bayes estimates. *The Annals of Statistics*, *14*, 68–87.

Freedman, D. (1963). On the asymptotic distribution of Bayes estimates in the discrete case I. *The Annals of Mathematical Statistics*, *34*, 1386–1403.

Ge, Y., Jiang, W. (2006). On consistency of Bayesian inference with mixtures of logistic regression. *Neural Computation*, *18*, 224–243.

Ghosal, S., Roy, A. (2006). Posterior consistency of Gaussian process prior for nonparametric binary regression. *The Annals of Statistics*, *34*, 2413–2429.

Ghosal, S., van der Vaart, A.W. Convergence rates of posterior distributions for noniid observations. *The Annals of Statistics*, *35*, 192–223.

Ghosh, J. K., Ramamoorthi, R. V. (2003). *Bayesian nonparametrics*. New York: Springer.

Huang, T. M. (2004). Convergence rates for posterior distributions and adaptive estimation. *The Annals of Statistics*, *32*, 1556–1593.

Kleijn, B. J. K., van der Vaart, A. W. (2006). Misspecification in infinite-dimensional Bayesian statistics. *The Annals of Statistics*, *34*, 837–877.

Kotz, S., Kozubowski, T. J., Podgorski, K. (2001). Laplace distributions and generalizations: a revisit with applications to communications, economics, engineering and finance. Birkhäuser: Springer.

Neal, R. M. (1996). Bayesian learning for neural networks. In: *Lecture notes in Statistics* (vol. 118). New York: Springer.

Neal, R. M. (1997). Monte Carlo implementation of Gaussian process models for Bayesian regression and classification. Technical Report, 9702, Department of Statistics, University of Toronto.

Paciorek, C. J. (2003). Nonstationary Gaussian processes for regression and spatial modelling. Ph.D. Thesis. Carnegie Mellon University, Pittsburgh.

Schwartz, L. (1965). On Bayes procedures. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, *4*, 10–26.

Shen, X., Wasserman, L. (2001). Rates of convergence of posterior distributions. *The Annals of Statistics*, *29*, 687–714.

Stone, C. J. (1977). Consistent nonparametric regression. *The Annals of Statistics*, *5*, 595–620.

Tokdar, S., Ghosh, J. K. (2007). Posterior consistency of Gaussian process priors in density estimation. *Journal of Statistical Planning and Inference Inference*, *137*, 34–42.

Walker, S. (2004). New approach to Bayesian consistency. *The Annals of Statistics*, *32*, 2028–2043.