On frequentist coverage errors of Bayesian credible sets in moderately high dimensions

KEISUKE YANO[1] and KENGO KATO[2]

[1]*Department of Mathematical Informatics, Graduate School of Information Science and Technology, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan. E-mail: yano@mist.i.u-tokyo.ac.jp*
[2]*Department of Statistics and Data Science, Cornell University, 1194 Comstock Hall, Ithaca, NY 14853, USA. E-mail: kk976@cornell.edu*

In this paper, we study frequentist coverage errors of Bayesian credible sets for an approximately linear regression model with (moderately) high dimensional regressors, where the dimension of the regressors may increase with but is smaller than the sample size. Specifically, we consider quasi-Bayesian inference on the slope vector under the quasi-likelihood with Gaussian error distribution. Under this setup, we derive finite sample bounds on frequentist coverage errors of Bayesian credible rectangles. Derivation of those bounds builds on a novel Berry–Esseen type bound on quasi-posterior distributions and recent results on high-dimensional CLT on hyperrectangles. We use this general result to quantify coverage errors of Castillo–Nickl and L^∞-credible bands for Gaussian white noise models, linear inverse problems, and (possibly non-Gaussian) nonparametric regression models. In particular, we show that Bayesian credible bands for those nonparametric models have coverage errors decaying polynomially fast in the sample size, implying advantages of Bayesian credible bands over confidence bands based on extreme value theory.

Keywords: Castillo–Nickl band; credible rectangle; sieve prior

1. Introduction

Bayesian inference for high or nonparametric statistical models is an active research area in the recent statistics literature. Posterior distributions provide not only point estimates but also credible sets. In a classical regular statistical model with a fixed finite dimensional parameter space, it is well known that the Bernstein–von Mises (BvM) theorem holds under mild conditions and the posterior distribution can be approximated (under the total variation distance) by a normal distribution centered at an efficient estimator (e.g., MLE) and with covariance matrix identical to the inverse of the Fisher information matrix as the sample size increases. The BvM theorem implies that a Bayesian credible set is typically a valid confidence set in the frequentist sense, namely, the coverage probability of a $(1 - \alpha)$-Bayesian credible set evaluated under the true parameter value is approaching $(1 - \alpha)$ as the sample size increases; cf. [49], Chapter 10. There is also a large literature on the BvM theorem in nonparametric statistical models. Compared to the finite dimensional case, however, Bayesian uncertainty quantification is more complicated and more sensitive to prior choices in the infinite dimensional case. [19,23] find some negative results on the BvM theorem in the infinite dimensional case. [6,34,37] develop conditions under which the BvM theorem holds for Gaussian white noise models and nonparametric regression models; see also [18,25,46]. Employing weaker topologies than L^2, [9] elegantly formulate and

establish the BvM theorem for Gaussian white noise models; see also [42] for the adaptive BvM theorem for Gaussian white noise models. Subsequently, [10] establish the BvM theorem in a weighted L^∞-type norm for nonparametric regression and density estimation. There are also several papers on frequentist coverage errors of Bayesian credible sets in the L^2-norm. [36] study asymptotic frequentist coverage errors of L^2-type Bayesian credible sets based on Gaussian priors for linear inverse problems; see also [45,47] for related results. Using an empirical Bayes approach, [48] develop L^2-type Bayesian credible sets adaptive to unknown smoothness of the function of interest. We refer the reader to Chapter 7 in [30] and Chapter 12 in [27] for further references on these topics.

This paper aims at studying frequentist coverage errors of Bayesian credible rectangles in an approximately linear regression model with an increasing number of regressors. We provide finite sample bounds on frequentist coverage errors of (quasi-)Bayesian credible rectangles based on sieve priors, where the model allows both an unknown bias term and an unknown error variance, and the true distribution of the error term may not be Gaussian. Sieve priors are distributions on the slope vector whose dimension increases with the sample size. We allow sieve priors to be non-Gaussian or not to be an independent product. We employ a "quasi-Bayesian" approach with Gaussian error distributions. The resulting posterior distribution is called a "quasi-posterior."

An important application of our results is finite sample quantification of Bayesian nonparametric credible bands based on sieve priors. We derive finite sample bounds on coverage errors of Castillo–Nickl [10] and L^∞-credible bands in Gaussian white noise models, linear inverse problems, and (possibly non-Gaussian) nonparametric regression models; see Section 3.1 ahead for the definition of Castillo–Nickl credible bands. The literature on frequentist confidence bands is broad. Frequentist approaches to constructing confidence bands date back to Smirnov and Bickel–Rosenblatt [5,44]; see also [14,17,28] for more recent results. In contrast, there are relatively limited results on Bayesian uncertainty quantification based on L^∞-type norms. [29] study posterior contraction rates in the L^r-norm for $1 \le r \le \infty$, and [8] derive sharp posterior contraction rates in the L^∞-norm. [32] derive adaptive posterior contraction rates in the L^∞-norm for Gaussian white noise models and density estimation; see also [52] for adaptive posterior contraction rates. Building on their new BvM theorem, [10] develop credible bands (Castillo–Nickl bands) based on product priors that have correct frequentist coverage probabilities and at the same time shrink at (nearly) minimax optimal rates for Gaussian white noise models. [53] study conditions under which frequentist coverage probabilities of credible bands based on Gaussian series priors approach one as the sample size increases for nonparametric regression models with sub-Gaussian errors. [42] establish qualitative results on adaptive credible bands for Gaussian white noise models. Still, *quantitative results* on frequentist coverage errors of nonparametric credible bands are scarce. Our quantitative result complements the qualitative results established by [10] and [53] and contributes to the literature on Bayesian nonparametrics by developing deeper understanding on Bayesian uncertainty quantification in nonparametric models. More recently, [50] also derive a quantitative result on coverage errors of Bayesian credible bands based on Gaussian process priors. We will clarify the difference between their results and ours in Section 1.1 ahead.

Notably, our results lead to an implication that supports the use of Bayesian approaches to constructing nonparametric confidence bands. It is well known that confidence bands based on extreme value theory (such as e.g., those of [5]) perform poorly because of the slow convergence

of Gaussian maxima. In the kernel density estimation case, [31] shows that confidence bands based on extreme value theory have coverage errors decaying only at the $1/\log n$ rate (regardless of how we choose bandwidths) where n is the sample size, while those based on bootstrap have coverage errors (for the surrogate function) decaying polynomially fast in the sample size; see also [14]. Our result shows that Bayesian credible bands (for the true function in Gaussian white noise models and linear inverse problems; for the surrogate function in nonparametric regression models) have also coverage errors decaying polynomially fast in the sample size, implying an advantage of Bayesian credible bands over confidence bands based on extreme value theory; see Remarks 3.2 and 3.8 for more details. Another potentially interesting implication of our analysis of the Castillo–Nickl band is the following. In this paper, we use a sieve prior that truncates high frequency terms of the function. In a Gaussian white noise model, our results show that the coverage error for the true function of the Castillo–Nickl band decays fast in the sample size (i.e., decays at a polynomial rate in the sample size), and at the same time the L^∞-diameter converges at a minimax optimal rate as long as the cut-off level 2^J is chosen in such a way that $2^J \sim (n/\log n)^{1/(2s+1)}$ where s is the smoothness level. This implies that, as long as we confine ourselves to nonadaptive credible bands, a sieve prior would not be less favorable than a prior that models high-frequency terms of the function.

The main ingredients in the derivation of the coverage error bound in Section 2 are (i) a novel Berry–Esseen type bound for the BvM theorem for sieve priors, that is, a finite sample bound on the total variation distance between the quasi-posterior distribution based on a sieve prior and the corresponding Gaussian distribution, and (ii) recent results on high dimensional CLT on hyperrectangles [13,15]. Our Berry–Esseen type bound improves upon existing BvM-type results for sieve priors; see the discussion in Section 1.1. The high dimensional CLT is used to approximate the sampling distribution of the centering estimator by the Gaussian distribution that matches with the Gaussian distribution approximating the (normalized) posterior distribution.

In addition, importantly, derivations of coverage error bounds for nonparametric models in Section 3 are by no means trivial and require further technical arguments. Specifically, for Gaussian white noise models, we will consider both credible bands based on centering estimators with fixed cut-off dimensions and without cut-off dimensions, which require different analyses on bounding the effect of the bias to the coverage error. For linear inverse problems, we will cover both mildly and severely ill-posed cases. For nonparametric regression models, we will consider random designs and so cannot directly apply the result of Section 2 since we assume fixed designs in Section 2; hence we have to take care of the randomness of the design, and to this end, we will employ some empirical process techniques.

1.1. Literature review and contributions

For a nonparametric regression model, [50] derive finite sample bounds on frequentist coverage errors of Bayesian credible bands based on Gaussian process priors. They assume (i) Gaussian process priors, (ii) that the error term follows a sub-Gaussian distribution, and (iii) that the error variance is known. The present paper markedly differs from [50] in that (i) we work with possibly non-Gaussian priors; (ii) we allow a more flexible error distribution; and (iii) we allow the error variance to be unknown. More specifically, (i) to allow for non-Gaussian priors, we develop

novel Berry–Esseen type bounds on quasi-posterior distributions in (mildly) high dimensions. (ii) In addition, to weaken the dimensionality restriction and the moment assumption on the error distribution, we make use of high-dimensional CLT on hyperrectangles developed in [13,15]. (iii) Finally, when the error variance is unknown, the quasi-posterior contraction for the error variance impacts on the coverage error for the slope vector and so a careful analysis is required to take care of the unknown variance.

The present paper also contributes to the literature on the BvM theorem in nonparametric statistics, which is now quite broad; see [9,10,23,34,37,42] for Gaussian white noise models, [6, 25] for linear regression models with high dimensional regressors, and [50,53] for nonparametric regression models with Gaussian process priors. See [12] for high-dimensional linear regression under sparsity constraints. Note that [12] also discusses non-Gaussian error distributions. See also [7,11,24,26,39,40,43] for related results. We refer the reader to [3,16,22,35] on the BvM theorem for quasi-posterior distributions.

Importantly, our Berry–Esseen type bound improves on conditions on the critical dimension for the BvM theorem. [6,25,46] study such critical dimensions for sieve priors. First, [6] does not cover the case with an unknown error variance, while the results in [25,46] cover the case with an unknown error variance. Our result is consistent with the result of [6] when the error variance is assumed to be known. Meanwhile, our result substantially improves on the results of [25,46] for the unknown error variance case. Namely, the results of [25,46] show that the BvM theorem holds if $p^3 = o(n)$ under typical situations when the error variance is unknown, where p is the number of regressors and n is the sample size; on the other hand, our result shows that the BvM theorem holds if $p^2(\log n)^3 = o(n)$, thereby improving on the condition of [25,46]. See Remark 2.2 for more details. Our BvM-type result allows us to cover wider smoothness classes of functions when applied to the analysis of Bayesian credible bands in nonparametric models.

1.2. Organization and notation

The rest of the paper is organized as follows. In Section 2, we consider Bayesian credible rectangles for the slope vector in an approximately linear regression model and derive finite sample bounds on frequentist coverage errors of the credible rectangles. In Section 3, we discuss applications of the general result established in Section 2 to nonparametric models. Specifically, we cover Gaussian white noise models, linear inverse models, and nonparametric regression models with possibly non-Gaussian errors. In Section 4, we give a proof of the main theorem (Theorem 2.1). Proofs of the other results are given in [51].

Throughout the paper, we will obey the following notation. Let $\| \cdot \|$ denote the Euclidean norm, and let $\| \cdot \|_\infty$ denote the max or supremum norm for vectors or functions. Let $\mathcal{N}(\mu, \Sigma)$ denote the Gaussian distribution with mean vector μ and covariance matrix Σ. For $x \in \mathbb{R}$, let $x_+ = \max\{x, 0\}$. For two sequences $\{a_n\}$ and $\{b_n\}$ depending on n, we use the notation $a_n \lesssim b_n$ if $a_n \leq cb_n$ for some universal constant $c > 0$, and $a_n \sim b_n$ if $a_n \lesssim b_n$ and $b_n \lesssim a_n$. For any symmetric positive semidefinite matrices A and B, the notation $A \preceq B$ means that $B - A$ is positive semidefinite. Constants $c_1, c_2, \ldots, c$, and $\widetilde{c}_1, \widetilde{c}_2, \ldots$ do not depend on the sample size n and the dimension p. The values of $c, c_1, c_2, \ldots$ and $\widetilde{c}_1, \widetilde{c}_2, \ldots$ may be different at each appearance.

2. Bayesian credible rectangles

Consider an approximately linear regression model

$$Y = X\beta_0 + r + \varepsilon, \tag{1}$$

where $Y = (Y_1, \ldots, Y_n)^\top \in \mathbb{R}^n$ is a vector of outcome variables, X is an $n \times p$ design matrix, $\beta_0 \in \mathbb{R}^p$ is an unknown coefficient vector, $r = (r_1, \ldots, r_n)^\top \in \mathbb{R}^n$ is a deterministic (i.e., non-random) bias term, and $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^\top \in \mathbb{R}^n$ is a vector of i.i.d. error terms with mean zero and variance $0 < \sigma_0^2 < \infty$. We are primarily interested in the situation where the number of regressors p increases with the sample size n, that is, $p = p_n \to \infty$ as $n \to \infty$, but we often suppress the dependence on n for the sake of notational simplicity. In addition, we allow the error variance σ_0^2 to depend on n, i.e., $\sigma_0^2 = \sigma_{0,n}^2$, which allows us to include Gaussian white noise models in the subsequent analysis as a special case. In the general setting, the error variance σ_0^2 is also unknown. In the present paper, we work with the dense model with moderately high-dimensional regressors where β_0 need not be sparse and $p = p_n$ may increase with the sample size n but $p \leq n$. To be precise, we will maintain the assumption that the design matrix X is of full column rank, i.e., rank $X = p$. The approximately linear model (1) is flexible enough to cover various nonparametric models such as Gaussian white noise models, linear inverse problems, and nonparametric regression models, via series expansions of functions of interest in those nonparametric models; see Section 3.

We consider Bayesian inference on the slope vector β_0. To this end, we work under the quasi-likelihood with a Gaussian distribution on the error ε. Namely, we work with the *quasi*-likelihood of the form

$$(\beta, \sigma^2) \mapsto (2\pi\sigma^2)^{-n/2} e^{-\|Y - X\beta\|^2 / (2\sigma^2)}.$$

We assume independent priors on β and σ^2, i.e.,

$$\beta \sim \Pi_\beta, \qquad \sigma^2 \sim \Pi_{\sigma^2}, \qquad \beta \perp\!\!\!\perp \sigma^2, \tag{2}$$

where we assume that Π_β is absolutely continuous with density π, that is, $\Pi_\beta(d\beta) = \pi(\beta)\, d\beta$, and Π_{σ^2} is supported in $(0, \infty)$. Then the resulting quasi-posterior distribution for (β, σ^2) is

$$\Pi\big(d(\beta, \sigma^2) \mid Y\big) \propto (2\pi\sigma^2)^{-n/2} e^{-\|Y - X\beta\|^2 / (2\sigma^2)} \pi(\beta)\, d\beta\, \Pi_{\sigma^2}(d\sigma^2),$$

and the marginal quasi-posterior distribution for β is $\Pi_\beta(d\beta \mid Y) = \pi(\beta \mid Y)\, d\beta$, where

$$\pi(\beta \mid Y) = \pi(\beta) \int \frac{e^{-\|Y - X\beta\|^2 / (2\sigma^2)}}{\int e^{-\|Y - X\widetilde{\beta}\|^2 / (2\sigma^2)} \pi(\widetilde{\beta})\, d\widetilde{\beta}} \Pi_{\sigma^2}(d\sigma^2 \mid Y).$$

Here $\Pi_{\sigma^2}(d\sigma^2 \mid Y)$ denotes the marginal quasi-posterior distribution for σ^2:

$$\Pi_{\sigma^2}(d\sigma^2 \mid Y) = \frac{\int (2\pi\sigma^2)^{-n/2} e^{-\|Y - X\beta\|^2 / (2\sigma^2)} \pi(\beta)\, d\beta\, \Pi_{\sigma^2}(d\sigma^2)}{\int \int (2\pi\widetilde{\sigma}^2)^{-n/2} e^{-\|Y - X\beta\|^2 / (2\widetilde{\sigma}^2)} \pi(\beta)\, d\beta\, \Pi_{\sigma^2}(d\widetilde{\sigma}^2)}.$$

We will assume that Π_{σ^2} may be data-dependent, e.g., $\Pi_{\sigma^2} = \delta_{\widehat{\sigma}^2}$ for some estimator $\widehat{\sigma}^2$ of σ^2 (in that case, $\Pi_{\sigma^2}(\cdot \mid Y) = \delta_{\widehat{\sigma}^2}$), but Π_β is data-independent.

We will derive finite sample bounds on frequentist coverage errors of Bayesian credible rectangles for the approximately linear model (1) under a prior of the form (2). For a vector $c = (c_1, \ldots, c_p)^\top \in \mathbb{R}^p$, a positive number $R > 0$, and a positive sequence $\{w_j\}_{j=1}^p$, let $I(c, R)$ denote the hyperrectangle of the form

$$I(c, R) := \left\{ \beta = (\beta_1, \ldots, \beta_p)^\top \in \mathbb{R}^p : \frac{|\beta_j - c_j|}{w_j} \leq R, 1 \leq \forall j \leq p \right\}.$$

Let $\widehat{\beta}$ denote the OLS estimator for β_0 with $r = 0$, that is, $\widehat{\beta} = \widehat{\beta}(Y) = (X^\top X)^{-1} X^\top Y$. For given $\alpha \in (0, 1)$, we consider a $(1 - \alpha)$-credible rectangle of the form $I(\widehat{\beta}, \widehat{R}_\alpha)$, where the radius $\widehat{R}_\alpha$ is chosen in such a way that the posterior probability of the set $I(\widehat{\beta}, \widehat{R}_\alpha)$ is $1 - \alpha$, that is, $\Pi_\beta\{I(\widehat{\beta}, \widehat{R}_\alpha) \mid Y\} = 1 - \alpha$.

We assume the following conditions on the priors Π_β and Π_{σ^2}. For $R > 0$, let

$$B(R) := \left\{ \beta \in \mathbb{R}^p : \left\| X(\beta - \beta_0) \right\| \leq R\sigma_0 \right\} \quad \text{and} \quad \phi_{\Pi_\beta}(R) := 1 - \inf_{\beta, \widetilde{\beta} \in B(R)} \left\{ \frac{\pi(\widetilde{\beta})}{\pi(\beta)} \right\}, \quad (3)$$

where ϕ_{Π_β} quantifies "lack of flatness" of the prior density $\pi(\beta)$ around the true value β_0.

Condition 2.1. *There exists a positive constant C_1 such that*

$$\pi(\beta_0) \geq \sigma_0^{-p} \sqrt{\det(X^\top X)} n^{-C_1 p}.$$

Condition 2.2. *There exist nonnegative constants $\delta_1, \delta_2, \delta_3 \in [0, 1)$ such that with probability at least $1 - \delta_3$, $\Pi_{\sigma^2}(\{\sigma^2 : |\sigma^2/\sigma_0^2 - 1| > \delta_1\} \mid Y) \leq \delta_2$.*

Condition 2.3. *The inequality $\phi_{\Pi_\beta}(1/\sqrt{n}) \leq 1/2$ holds.*

Condition 2.1 assumes that the prior Π_β on β has a sufficient mass around its true value β_0. Condition 2.2 is an assumption on the marginal posterior contraction for the error variance σ^2. Condition 2.2 includes the known error variance case as a special case; if the error variance is known, then we may take $\Pi_{\sigma^2} = \delta_{\sigma_0^2}$ (Dirac delta at σ_0^2) and $\delta_1 = \delta_2 = \delta_3 = 0$. Condition 2.3 is a preliminary flatness condition on Π_β. More detailed discussions on these conditions are provided after the main theorem (Theorem 2.1).

We also assume the following conditions on the model.

Condition 2.4. *There exists a positive constant C_2 such that $\left\| X(X^\top X)^{-1} X^\top r \right\| \leq C_2 \sigma_0 \sqrt{p \log n}$.*

Condition 2.5. *There exists a positive constant C_3 such that one of the following conditions holds*:

(a) $\mathbb{E}[|\varepsilon_1/(\sigma_0 C_3)|^q] \leq 1$ *for some integer $4 \leq q < \infty$*;
(b) $\mathbb{E}[\exp\{\varepsilon_1^2/(\sigma_0 C_3)^2\}] \leq 2$.

Condition 2.4 controls the norm of the bias term. Condition 2.5 is a moment condition on the error distribution. These conditions are sufficiently weak and in particular covers all the applications we will cover.

The following theorem, which is the main result of this section, provides bounds on frequentist coverage errors of the Bayesian credible rectangle $I(\widehat{\beta}, \widehat{R}_\alpha)$ together with bounds on the "radius" $\widehat{R}_\alpha$ of $I(\widehat{\beta}, \widehat{R}_\alpha)$. In what follows, let $\overline{\lambda}$ and $\underline{\lambda}$ denote the maximum and minimum eigenvalues of the matrix $(X^\top X)^{-1}$, respectively, and let $\overline{w} := \max\{w_1, \ldots, w_p\}$ and $\underline{w} := \min\{w_1, \ldots, w_p\}$ denote the maximal and minimal weights, respectively.

Theorem 2.1 (Coverage errors of credible rectangles). *Suppose that Conditions 2.1–2.4 and either of Condition 2.5(a) or (b) hold. Then there exist positive constants c_1 and c_2 depending only on C_1, C_2, C_3 and q such that the following hold. For every $n \geq 2$, we have*

$$\left| \mathbb{P}\big(\beta_0 \in I(\widehat{\beta}, \widehat{R}_\alpha)\big) - (1 - \alpha) \right|$$
$$\leq \phi_{\Pi_\beta}(c_1 \sqrt{p \log n}) + c_1 \left(\delta_1 p \log n + \delta_2 + \delta_3 + \frac{\tau}{\sigma_0 \underline{\lambda}^{1/2}} \sqrt{\log p} + \zeta_n \right), \qquad (4)$$

where $\tau := \|(X^\top X)^{-1} X^\top r\|_\infty$ and

$$
\zeta_n =
\begin{cases}
p^{1-q/2}(\log n)^{-q/2} + \left(\dfrac{\overline{\lambda}}{\underline{\lambda}} \dfrac{p \log^7(pn)}{n} \right)^{1/6} + \left(\dfrac{\overline{\lambda}}{\underline{\lambda}} \dfrac{p \log^3(pn)}{n^{1-2/q}} \right)^{1/3} \\
\qquad\qquad\qquad\qquad\qquad \text{under Condition 2.5(a)}, \\[2ex]
n^{-c_2 p} + \left(\dfrac{\overline{\lambda}}{\underline{\lambda}} \dfrac{p \log^7(pn)}{n} \right)^{1/6} \quad \text{under Condition 2.5(b)}, \\[2ex]
n^{-c_2 p} \qquad\qquad\qquad\qquad \text{if } \varepsilon_i\text{'s are Gaussian}.
\end{cases}
$$

In addition, there exist positive constants c_3 and c_4 depending only on α and $\underline{w}$ such that the following two bounds (5) *and* (6) *hold with probability at least*

$$
\begin{cases}
1 - c_1 p^{1-q/2}(\log n)^{-q/2} - \delta_3 & \text{under Condition 2.5(a)}, \\
1 - c_1 n^{-c_2 p} - \delta_3 & \text{under Condition 2.5(b)}.
\end{cases}
$$

Provided that the right-hand side on (4) *is smaller than* $\min\{\alpha/2, (1-\alpha)/2\}$, *the diameter $\widehat{R}_\alpha$ is bounded from above as*

$$\widehat{R}_\alpha \leq c_3 \sigma_0 \overline{\lambda}^{1/2} \mathbb{E}\left[\max_{1 \leq i \leq p} |N_i / w_i| \right] \qquad (5)$$

for $N_1, \ldots, N_p \sim \mathcal{N}(0, 1)$ i.i.d., and for sufficiently large p depending only on α, the diameter $\widehat{R}_\alpha$ is bounded from below as

$$c_4 \sigma_0 \underline{\lambda}^{1/2} \overline{w}^{-1} \sqrt{\log p} \leq \widehat{R}_\alpha. \qquad (6)$$

Theorem 2.1 shows that that the frequentist coverage error of the Bayesian credible rectangle depends on the prior Π_β on β only through the lack-of-flatness function ϕ_{Π_β}. The discussions below provide a typical bound on ϕ_{Π_β}. We note that the requirement that the right-hand side on (4) is smaller than $\alpha/2$ is used to derive the upper bound on $\widehat{R}_\alpha$, while the requirement that the same quantity is smaller than $(1-\alpha)/2$ is used to derive the lower bound on $\widehat{R}_\alpha$.

2.1. Discussions on conditions

We first verify that a *locally log-Lipschitz prior* satisfies Conditions 2.1 and 2.3, providing an upper bound of ϕ_{Π_β}.

Definition 2.1. A *locally log-Lipschitz prior* is defined as a prior distribution on β such there exists $L = L_n > 0$ with

$$\left|\log\pi(\beta) - \log\pi(\beta_0)\right| \le L\|\beta - \beta_0\| \quad \text{for all } \beta \text{ with } \|\beta - \beta_0\| \le \sigma_0\overline{\lambda}^{-1/2}\sqrt{p\log n}.$$

Proposition 2.1. *For a locally log-Lipschitz prior Π_β with log-Lipschitz constant L, we have* $\phi_{\Pi_\beta}(c\sqrt{p\log n}) \le cL\sigma_0\overline{\lambda}^{-1/2}\sqrt{p\log n}$ *for any $c > 0$. Hence the prior Π_β satisfies Condition 2.3 if $\sigma_0 L\overline{\lambda}^{-1/2}/\sqrt{n} \le 1/2$.*

To provide examples of prior distributions on β that satisfy Condition 2.1, we focus on the following two subclasses of locally log-Lipschitz priors. Let $B := \|\beta_0\|$ denote the Euclidean norm of β_0.

(Isotropic prior) An *isotropic prior* is of the form $\pi(\beta) = \rho(\|\beta\|)/\int\rho(\|\beta\|)\,d\beta$ where ρ is a probability density function on $\mathbb{R}_+$ such that ρ is strictly positive and continuously differentiable on $[0, B + \sigma_0\overline{\lambda}^{-1/2}\sqrt{p\log n}]$, and such that $\int_0^\infty x^k\rho(x)\,dx \le \exp(mk\log k)$ for all $k \in \mathbb{N}$ for some positive constant m.

(Product prior) A *product prior* of log-Lipschitz priors is of the form $\pi(\beta) = \prod_{i=1}^p \pi_i(\beta_i)$ where each $\log\pi_i$ is strictly positive on $[0, B + \sigma_0\overline{\lambda}^{-1/2}\sqrt{p\log n}]$ and $\widetilde{L}$-Lipschitz for some $\widetilde{L} > 0$.

For the sake of exposition, we make the following additional condition to verify that isotropic or product priors satisfy Condition 2.1.

Condition 2.6. *There exists a positive constant c such that $\log\{\sqrt{\det(X^\top X)}/\sigma_0^p\} \le cp\log n$.*

This condition is satisfied in all the applications we will cover in Section 3. The following proposition shows that isotropic or product priors are locally log-Lipschitz priors satisfying Condition 2.1.

Proposition 2.2. *Under Condition 2.6, an isotropic prior and a product prior of log-Lipschitz priors satisfy Condition 2.1. An isotropic prior is a locally log-Lipschitz prior with locally log-*

Lipschitz constant L such that

$$L \leq c_1 B \max_{x:0 \leq x \leq B + \sigma_0 \overline{\lambda}^{1/2} \sqrt{p \log n}} \left| (\log \rho)'(x) \right|$$

for some positive constant c_1 depending only on m and c that appear in the definition of ρ and Condition 2.6. In particular, if $\pi(\beta)$ is the standard Gaussian density, then $L \leq c_1 B^2$. A product prior of log-Lipschitz priors with log-Lipschitz constant $\widetilde{L}$ is locally log-Lipschitz with $L = \widetilde{L} p^{1/2}$.

Next, we will discuss Condition 2.2. We consider following two cases:

(Plug-in) $\Pi_{\sigma^2} = \Pi_{\widehat{\sigma}_u^2}$ with $\widehat{\sigma}_u^2(Y) := \|Y - X(X^\top X)^{-1} X^\top Y\|^2/(n-p)$;
(Full-Bayes) Π_β is the standard Gaussian distribution and Π_{σ^2} is the inverse Gamma distribution $\mathrm{IG}(\mu_1, \mu_2)$ with shape parameter $\mu_1 > 1/2$ and scale parameter $\mu_2 > 1/2$.

The following two propositions yield possible choices of $\delta_1, \delta_2,$ and δ_3.

Proposition 2.3 (Plug-in). *Suppose that Condition 2.5 holds and also that $n \geq cp$ for some $c > 1$. In addition, suppose that $\delta_1 > 0$ satisfies that $\widetilde{\delta}_1 := [\delta_1 - 2\|r\|^2/\{\sigma_0^2(n-p)\} - 1/(n-p)] > 0$. Then there exist positive constants c_1 and c_2 depending only on c, C_3 and q such that*

$$\mathbb{P}\left(\left| \widehat{\sigma}_u^2/\sigma_0^2 - 1 \right| \geq \delta_1 \right) \leq \begin{cases} c_1 \max\{n^{-4/q} \delta_1^{-q/2}, n^{1-q/2} \widetilde{\delta}_1^{-q}\} & \text{under Condition 2.5(a)}, \\ c_1 \exp\left(-c_2 n \max\{\delta_1^2, \widetilde{\delta}_1^2\}\right) & \text{under Condition 2.5(b)}. \end{cases}$$

Proposition 2.4 (Full-Bayes). *Suppose that Condition 2.5 holds and also $n \geq cp$ for some $c > 1$. In addition, suppose that $\delta_1 > 0$ satisfies that $\widetilde{\delta}_1 := [\delta_1 - 2\|r\|^2/\{\sigma_0^2(n-p)\} - 1/(n-p)] > 0$. Then there exist positive constants c_1 and c_2 depending only on c, μ_1, μ_2, C_3 and q such that*

$$\Pi_{\sigma^2}\left(\sigma^2 : \left| \sigma^2/\sigma_0^2 - 1 \right| > \delta_1 \mid Y \right) \leq c_1 (n\widetilde{\delta}_1)^{-1}$$

with probability at least

$$\begin{cases} 1 - c_1 \max\{n^{-4/q} \delta_1^{-q/2}, n^{1-q/2} \widetilde{\delta}_1^{-q}\} & \text{under Condition 2.5(a)}, \\ 1 - c_1 \exp\left(-c_2 n \max\{\delta_1^2, \widetilde{\delta}_1^2\}\right) & \text{under Condition 2.5(b)}. \end{cases}$$

To better understand implications of these propositions, Table 1 summarizes possible rates of $\delta_1, \delta_2, \delta_3$ when $n \geq cp$ for some $c > 0$, $\|r\|^2/n = o(n^{-1/2})$, and σ_0^2 is independent of n.

Remark 2.1 (Comparison with [53]). Proposition 4.1 in [53] studies possible rates for δ_1 when a prior for β is Gaussian and the error distribution is sub-Gaussian. Our results in Propositions 2.3 and 2.4 are compatible with their result up to logarithmic factors under their setup.

Table 1. Possible rates of $\delta_1, \delta_2, \delta_3$ with respect to n: κ is arbitrary

Condition 2.5 and prior	δ_1	δ_2	δ_3
(a) and plug-in	$n^{-1/2+\kappa/q}$	0	$\max\{n^{-\kappa/2}, n^{1-\kappa}\}$
(a) and full Bayes	$n^{-1/2+\kappa/q}$	$n^{-1/2-\kappa/q}$	$\max\{n^{-\kappa/2}, n^{1-\kappa}\}$
(b) and plug-in	$n^{-1/2}\sqrt{\log n}$	0	n^{-1}
(b) and full Bayes	$n^{-1/2}\sqrt{\log n}$	$n^{-1/2}(\log n)^{-1/2}$	n^{-1}

2.2. Berry–Esseen type bounds on posterior distributions

Before presenting applications of the main theorem, we derive an important ingredient of the proof of Theorem 2.1, namely, the Berry–Esseen type bound on posterior distributions. For $R > 0$, let $H(R)$ be the intersection of the sets $\{Y \in \mathbb{R}^n : \|X(\widehat{\beta}(Y) - \beta_0)\| \le R\sqrt{p \log n}\sigma_0/4\}$ and $\{Y \in \mathbb{R}^n : \Pi_{\sigma^2}(|\sigma^2/\sigma_0^2 - 1| \ge \delta_1 \mid Y) \le \delta_2\}$. For two probability measures P and Q, $\|P - Q\|_{\mathrm{TV}}$ denotes the total variation between P and Q.

Proposition 2.5 (Berry–Esseen type bounds on posterior distributions). *Under Conditions 2.1–2.3, there exist positive constants c_1 and c_2 depending only on C_1, C_2, C_3 such that for every $n \ge 2$,*

$$\left\|\Pi_\beta(\cdot \mid Y) - \mathcal{N}\left(\widehat{\beta}, \sigma_0^2(X^\top X)^{-1}\right)\right\|_{\mathrm{TV}} \le \phi_{\Pi_\beta}(c_1\sqrt{p \log n}) + c_1\left(\delta_1 p \log n + \delta_2 + n^{-c_2 p}\right)$$

whenever $Y \in H(c_1)$.

Proposition 2.6. *Under Conditions 2.4 and 2.5, there exist positive constants c_1 and c_2 depending only on C_2, C_3, and q such that*

$$\mathbb{P}\left(Y \notin H(c_1)\right) \le \begin{cases} c_1 p^{1-q/2}(\log n)^{-q/2} + \delta_3 & \text{under Condition 2.5(a)}, \\ c_1 n^{-c_2 p} + \delta_3 & \text{under Condition 2.5(b)}. \end{cases}$$

Remark 2.2 (Critical dimension for the Bernstein–von Mises theorem). The previous propositions immediately lead to the critical dimension for the BvM theorem. We will compare our result with the results on the critical dimension by [6,26,46]. In this comparison, we assume a locally log-Lipschitz prior with locally log-Lipschitz constant L; that $\|\beta_0\|$ and L are independent of n; and that $\sigma_0\overline{\lambda}^{1/2} \sim n^{-1/2}$. The following are a summary of the existing results:

- [26] shows that when the error distribution has a smooth density with known scale parameter, the BvM theorem holds if $p^4 \log p = o(n)$ and some additional assumptions are verified;
- [46] shows that when the high-dimensional local asymptotic normality holds, the BvM theorem holds if $p^3 = o(n)$; see also [41];
- [6] shows that when the error distribution is Gaussian with known variance, the BvM theorem holds if $p \log n = o(n)$.

Our result (Propositions 2.1, 2.3, 2.5, and 2.6) improves on [26,46] in that

- when the error variance is assumed to be known (i.e., $\delta_1 = \delta_2 = \delta_3 = 0$), our result implies that the BvM theorem (for the quasi-posterior distribution) holds if $p \log n = o(n)$ and if the error distribution has finite fourth moment. Compared to [26], our result substantially improves on the critical dimension by employing the Gaussian likelihood even when the Gaussian specification is incorrect;
- when the error variance is unknown, our result shows that the BvM theorem holds for β if $p^2 (\log n)^3 = o(n)$ for sub-Gaussian error distributions, thereby improving on the condition of [46].

Importantly, our result covers the unknown error variance case, which makes our analysis different from [6]. In nonparametric regression, it is usually the case that the error variance is unknown, and hence it is important to consider unknown variance cases in such an application. If the error distribution is Gaussian with a known error variance, our result is consistent with [6].

3. Applications

In this section, we consider applications of the general results developed in the previous sections to quantifying coverage errors of Bayesian credible sets in Gaussian white noise models, linear inverse problems, and (possibly non-Gaussian) nonparametric regression models.

3.1. Gaussian white noise model

We first consider a Gaussian white noise model and analyze coverage errors of Castillo–Nickl credible bands. Consider a Gaussian white noise model

$$dY(t) = f_0(t)\,dt + \frac{1}{\sqrt{n}}\,dW(t), \quad t \in [0,1],$$

where dW is a canonical white noise and f_0 is an unknown function. We assume that f_0 is in the Hölder–Zygmund space $B^s_{\infty,\infty}$ with smoothness level $s > 0$. It will be convenient to define the Hölder–Zygmund space $B^s_{\infty,\infty}$ by using a wavelet basis. Let $S > s$ be an integer and fix sufficiently large $J_0 = J_0(S)$. Let $\{\phi_{J_0,k} : 0 \le k \le 2^{J_0} - 1\} \cup \{\psi_{l,k} : J_0 \le l, 0 \le k \le 2^l - 1\}$ be an S-regular Cohen–Daubechies–Vial (CDV) wavelet basis of $L^2[0,1]$. Then the Hölder–Zygmund space $B^s_{\infty,\infty}$ is defined by $B^s_{\infty,\infty} = \{f : \|f\|_{B^s_{\infty,\infty}} < \infty\}$ with

$$\|f\|_{B^s_{\infty,\infty}} := \max_{0 \le k \le 2^{J_0}-1} \left|\langle \phi_{J_0,k}, f\rangle\right| + \sup_{J_0 \le l < \infty, 0 \le k \le 2^l-1} 2^{l(s+1/2)} \left|\langle \psi_{l,k}, f\rangle\right|,$$

where $\langle \cdot, \cdot \rangle$ denotes the $L^2[0,1]$ inner product, i.e., $\langle f, g\rangle := \int_{[0,1]} f(t) g(t)\, dt$. In what follows, for the notational convention, let $\psi_{J_0-1,k} := \phi_{J_0,k}$ for $0 \le k \le 2^{J_0} - 1$.

Consider a sieve prior for f, that is, a prior deduced from a prior Π_β on $\mathbb{R}^{2^J}$ with $J \ge J_0$ via the map $(\beta_{J_0-1,0}, \beta_{J_0-1,1}, \ldots, \beta_{J-1,2^{J-1}-1}) \mapsto \sum_{(l,k) \in \mathcal{I}(J)} \psi_{l,k}(\cdot) \beta_{l,k}$, where $\mathcal{I}(J) := \{(l,k) : J_0 \le l \le J-1, 0 \le k \le 2^l - 1\} \cup \{(l,k) : l = J_0 - 1, 0 \le k \le 2^{J_0} - 1\}$.

For given $\alpha \in (0, 1)$, the $(1 - \alpha)$-Castillo–Nickl credible band based on an efficient estimator $\widehat{f}$, an admissible sequence $w = (w_1, w_2, \ldots)$, and a sieve prior Π_β is defined as

$$\mathcal{C}_w(\widehat{f}, \widehat{R}_\alpha) := \left\{ f : \sup_{(l,k) \in \mathcal{I}_\infty} \frac{|\langle f - \widehat{f}, \psi_{l,k} \rangle|}{w_l} \leq \widehat{R}_\alpha \right\},$$

where $\mathcal{I}_\infty := \{(l, k) : J_0 \leq l < \infty, 0 \leq k \leq 2^l - 1\} \cup \{(l, k) : l = J_0 - 1, 0 \leq k \leq 2^{J_0} - 1\}$, and an admissible sequence w is defined as a positive sequence such that $w_l/\sqrt{l} \uparrow \infty$ as $l \to \infty$. The radius $\widehat{R}_\alpha$ of the band is taken in such a way that $\Pi_\beta \{ \mathcal{C}_w(\sum_{(l,k) \in \mathcal{I}(J)} \langle \widehat{f}, \psi_{l,k} \rangle \psi_{l,k}, \widehat{R}_\alpha) \mid Y \} = 1 - \alpha$. Truncating a centering estimator ensures that such radius indeed exists for a sieve prior.

The following proposition derives bounds on the coverage error and the L^∞-diameter of the Castillo–Nickl credible band based on a sieve prior. In the following proposition, we use $\widehat{f}_\infty := \sum_{(l,k) \in \mathcal{I}_\infty} \psi_{l,k} \int \psi_{l,k} \, dY$ (which converges almost surely in $\mathcal{M}_0(w)$) as a centering estimator. See p. 1946 of [10] for the definition of $\mathcal{M}_0(w)$ and well-definedness of $\widehat{f}_\infty$. Let

$$u_J := \inf_{J \leq l < \infty} w_l/\sqrt{l}, \qquad v_J := \max_{J_0 - 1 \leq l \leq J - 1} w_l/\sqrt{l} \quad \text{and} \quad \overline{w}_J := \max_{J_0 - 1 \leq l \leq J - 1} w_l.$$

In addition, let $\widetilde{H} := \{Y : \sup_{J \leq l < \infty, 0 \leq k \leq 2^l - 1} |\langle f_0 - \widehat{f}_\infty, \psi_{l,k} \rangle|/w_l \leq \widehat{R}_\alpha\}$. For simplicity, we assume that $\sqrt{l} \leq w_l$ for $J_0 - 1 \leq l < \infty$ and $1 \leq (J/\overline{w}_J^2) u_J^2 \uparrow \infty$ as $J \to \infty$.

Proposition 3.1. *Under Conditions* 2.1 *and* 2.3 *for* Π_β *with* $p = 2^J$, $X = I_p$, *and* $\sigma_0 = 1/\sqrt{n}$, *there exist positive constants* c_1, c_2 *depending only on* C_1 *appearing in Condition* 2.1 *such that the following hold. For* $n \geq 2$, *we have*

$$\left| \mathbb{P}\big(f_0 \in \mathcal{C}_w(\widehat{f}_\infty, \widehat{R}_\alpha) \big) - (1 - \alpha) \right| \leq \phi_{\Pi_\beta}\big(c_1 \sqrt{2^J \log n} \big) + c_1 n^{-c_2 2^J} + \mathbb{P}(Y \notin \widetilde{H}).$$

In addition, there exist positive constants c_3, c_4 *depending only on* α *such that the following hold. Assume that the right-hand side above except* $\mathbb{P}(Y \notin \widetilde{H})$ *is smaller than* $\min\{\alpha/2, (1 - \alpha)/2\}$. *Then*

$$\mathbb{P}(Y \notin \widetilde{H}) \leq c_3 \big(\mathrm{e}^{-c_4 J (J/\overline{w}_J^2) u_J^2} + n^{-c_2 2^J} \big)$$

for sufficiently large J *depending only on* α *and* $\{w_l\}$; *and the* L^∞-*diameter of the intersection* $\mathcal{C}_w^B(\widehat{f}_\infty, \widehat{R}_\alpha) := \mathcal{C}_w(\widehat{f}_\infty, \widehat{R}_\alpha) \cap \{f : \|f\|_{B_{\infty,\infty}^s} \leq B\}$ *for any* $B > 0$ *is bounded from above as*

$$\sup_{f, g \in \mathcal{C}_w^B(\widehat{f}_\infty, \widehat{R}_\alpha)} \|f - g\|_\infty \leq c_3 \left(v_J \sqrt{\frac{2^J J}{n}} + 2^{-Js} B \right)$$

with probability at least $1 - c_1 n^{-c_2 2^J}$.

Proof sketch of Proposition 3.1. First, we transform the Gaussian white noise model into a Gaussian infinite sequence model $Y_{l,k} = \beta_{0,l,k} + \varepsilon_{l,k}, (l, k) \in \mathcal{I}_\infty$, where $\beta_{0,l,k} := \langle f_0, \psi_{l,k} \rangle$ for $(l, k) \in \mathcal{I}_\infty$, and $\varepsilon_{l,k}$ are i.i.d. $\mathcal{N}(0, 1/n)$ variables. Second, we apply Theorem 2.1.

Let $Y_\infty = \{Y_{l,k} : (l,k) \in \mathcal{I}_\infty\}$ and observe that $\mathbb{P}(Y \notin \widetilde{H}) = \mathbb{P}(Y_\infty \notin \widetilde{H}')$ with $\widetilde{H}' = \{Y_\infty : \sup_{J < l, 0 \le k \le 2^l - 1} |Y_{l,k} - \beta_{0,l,k}|/w_l \le \widehat{R}_\alpha\}$. Since

$$\mathbb{P}\big(f_0 \in \mathcal{C}_w(\widehat{f}_\infty, \widehat{R}_\alpha)\big) = \mathbb{P}\Big(\max_{(l,k) \in \mathcal{I}(J)} |\varepsilon_{l,k}/w_l| \vee \sup_{J \le l < \infty, 0 \le k \le 2^l - 1} |\varepsilon_{l,k}/w_l| \le \widehat{R}_\alpha \Big),$$

we have

$$\Big| \mathbb{P}\big(f_0 \in \mathcal{C}_w(\widehat{f}_\infty, \widehat{R}_\alpha)\big) - \mathbb{P}\Big(\max_{(l,k) \in \mathcal{I}(J)} |\varepsilon_{l,k}/w_l| \le \widehat{R}_\alpha \Big) \Big| \le \mathbb{P}\big(Y_\infty \notin \widetilde{H}'\big).$$

Then we apply Theorem 2.1 with $p = 2^J$, $Y = \{Y_{l,k} : (l,k) \in \mathcal{I}(J)\}$, $X = I_p$, $\sigma_0 = 1/\sqrt{n}$, and $r = 0$ to obtain bounds on $\mathbb{P}(\max_{(l,k) \in \mathcal{I}(J)} |\varepsilon_{l,k}/w_l| \le \widehat{R}_\alpha)$ and $\widehat{R}_\alpha$. It remains to bound $\mathbb{P}(Y_\infty \notin \widetilde{H}')$. To this end, we use the concentration inequality for the Gaussian maximum together with a high-probability lower bound on $\widehat{R}_\alpha$. The detail can be found in Appendix C.1 of [51]. □

Remark 3.1 (Coverage error rates). The finite sample bound in Proposition 3.1 leads to the following asymptotic results as $n \to \infty$. In this discussion, we assume a locally log-Lipschitz prior with locally log-Lipschitz constant $L = L_n$ and a true function f_0 with $\|f_0\|_{B^s_{\infty,\infty}} \le B$ for some $B = B_n$. Set $2^J = (n/\log n)^{1/(2s+1)}$ and set $w_l = \sqrt{l}$ for $l \le J - 1$ and $w_l = u_l \sqrt{l}$ for $l \ge J$ with $u_l \uparrow \infty$ as $l \to \infty$. Then we have

$$\big| \mathbb{P}\big(f_0 \in \mathcal{C}^B_w(\widehat{f}, \widehat{R}_\alpha)\big) - (1 - \alpha) \big| \le O\big(L_n(n/\log n)^{-s/(2s+1)}\big) \quad \text{and} \tag{7}$$

$$\sup_{f,g \in \mathcal{C}^B_w(\widehat{f}_\infty, \widehat{R}_\alpha)} \|f - g\|_\infty \le O\big(B_n(n/\log n)^{-s/(2s+1)}\big), \tag{8}$$

where the latter holds with probability at least $1 - c_1 n^{-c_2 2^J}$ (the sequence $\{w_l\}$ here depends on n, but we can apply Proposition 3.1; see Remark C.1 in [51] for the detail). In particular, for the standard Gaussian prior, the coverage error is $O(B_n^2(n/\log n)^{-s/(2s+1)})$. We note that the above asymptotic results are derived from the non-asymptotic result in Proposition 3.1 where the constants do no depend on f_0; hence the above asymptotic results hold uniformly in f_0 as long as $\|f_0\|_{B^s_{\infty,\infty}} \le B$. The same comments apply to the subsequent results.

Remark 3.2 (Comparison of coverage errors). The previous remark shows that Bayesian credible bands have coverage errors (for the true function) decaying polynomially fast in the sample size n. This rate is much faster than that of confidence bands based on Gumbel approximations (see Proposition 6.4.3 in [30]); confidence bands based on Gumbel approximations have coverage errors decaying only at the $1/\log n$ rate. In the kernel density estimation case, [31] shows that confidence bands based on Gumbel approximations have coverage errors decaying only at the $1/\log n$ rate, while bootstrap confidence bands have coverage errors decaying polynomially fast in n for the surrogate function.

Remark 3.3 (Undersmothing). In most cases, a priori bound on $\|f_0\|_{B^s_{\infty,\infty}}$ is unknown, and so $B = B_n$ should be chosen as a slowly divergent sequence, which can be thought of as a

"undersmoothing" penalty (cf. [10] Remark 5). Interestingly, however, our result shows that this undersmoothing penalty only affects the L^∞-diameter and not affect the coverage error of the band, which is a sharp contrast with standard L^∞-confidence bands for densities or regression functions.

Consider another centering estimator: $\widehat{f}_J := \sum_{(l,k)\in\mathcal{I}(J)} \psi_{l,k} \int \psi_{l,k}\, dY$. The following proposition derives bounds on the coverage error and the L^∞-diameter of the Castillo–Nickl credible band based on a sieve prior and the centering estimator $\widehat{f}_J$. We use the same notation $u_J, v_J, \overline{w}_J$ as in the previous proposition. Let

$$\widetilde{H}_2 := \left\{ Y : \sup_{J\leq l<\infty, 0\leq k\leq 2^l-1} |\langle f_0, \psi_{l,k}\rangle|/w_l \leq \widehat{R}_\alpha \right\}.$$

For simplicity, we assume $\sqrt{l} \leq w_l$ for $J_0 - 1 \leq l < \infty$.

Proposition 3.2. *Under Conditions 2.1 and 2.3 for Π_β with $p = 2^J$, $X = I$, and $\sigma_0 = 1/\sqrt{n}$, there exist positive constants c_1, c_2, c_3 depending only on C_1 appearing in Condition 2.1 and α such that the following hold. For $n \geq 2$ and for $B > 0$ satisfying $\|f_0\|_{B^s_{\infty,\infty}} \leq B$, we have*

$$\left| \mathbb{P}\big(f_0 \in \mathcal{C}_w(\widehat{f}_J, \widehat{R}_\alpha)\big) - (1-\alpha) \right| \leq \phi_{\Pi_\beta}\big(c_1\sqrt{2^J \log n}\big) + c_1 n^{-c_2 2^J} + \mathbb{P}(Y \notin \widetilde{H}_2).$$

In addition, assume that the right hand side above except $\mathbb{P}(Y \notin \widetilde{H}_2)$ is smaller than $\min\{\alpha/2, (1-\alpha)/2\}$. Then the L^∞-diameter of the intersection $\mathcal{C}^B_w(\widehat{f}_J, \widehat{R}_\alpha) := \mathcal{C}_w(\widehat{f}_J, \widehat{R}_\alpha) \cap \{f : \|f\|_{B^s_{\infty,\infty}} \leq B\}$ is bounded from above as

$$\sup_{f,g\in\mathcal{C}^B_w(\widehat{f}_J,\widehat{R}_\alpha)} \|f - g\|_\infty \leq c_3 \left(v_J \sqrt{\frac{2^J J}{n}} + 2^{-Js}B \right)$$

with probability at least $1 - c_1 n^{-c_2 2^J}$. If in addition $(\sqrt{n}\,\overline{w}_J B)/(u_J J 2^{J(s+1/2)}) \downarrow 0$ as $J \to \infty$, then $\mathbb{P}(Y \notin \widetilde{H}_2) \leq c_1 n^{-c_2 2^J}$ for sufficiently large J depending only on α, $\{w_l\}$, and B.

A proof of the proposition is given in Appendix C.2 of [51].

Remark 3.4 (Choice of the sequence w). Consider the same setting as in Remark 3.1. Then we have $(\sqrt{n}\,\overline{w}_J B)/(u_J J 2^{J(s+1/2)}) = O(B/u_J)$ and so the sequence u_l must satisfy $u_J/B_n \to \infty$ as $n \to \infty$ to ensure that $(\sqrt{n}\,\overline{w}_J B)/(u_J J 2^{J(s+1/2)}) \downarrow 0$ as $J \to \infty$. Without this exception, the same asymptotic results hold as in Remark 3.1.

3.2. Linear inverse problem

In this section, we extend the previous analysis to a linear inverse problem

$$dY(t) = K(f_0)(t)\, dt + \frac{1}{\sqrt{n}}\, dW(t), \quad t \in [0, 1],$$

where K is a known linear operator and f_0 is included in the Hölder–Zygmund space $B^s_{\infty,\infty}$ for some $s > 0$ as described in the previous section. To describe the degree of ill-posedness, we use the wavelet-vaguelette decomposition $\{\psi_{l,k}, v^{(1)}_{l,k}, v^{(2)}_{l,k}, \kappa_{l,k} : (l,k) \in \mathcal{I}_\infty\}$ of K, where $\{\psi_{l,k}\}$ is a wavelet basis (with the same notational convention used in the previous subsection), $\{v^{(1)}_{l,k}\}$ and $\{v^{(2)}_{l,k}\}$ are near-orthogonal functions, and $\{\kappa_{l,k}\}$ are quasi-singular values such that $K(\psi_{l,k}) = \kappa_{l,k} v^{(2)}_{l,k}$ for $(l,k) \in \mathcal{I}_\infty$. For details, see [1,21,33,35] and references therein. Our results cover both mildly ill-posed and severely ill-posed cases for $\{\kappa_{l,k}\}$. Say that the problem of recovering f_0 is mildly ill-posed if $\kappa_{l,k} \sim 2^{-rl}$ for some $r > 0$, and severely ill-posed if $\kappa_{l,k} \sim \mathrm{e}^{-r2^l}$ for some $r > 0$.

We consider a sieve prior induced from a prior Π_β on $\mathbb{R}^{2^J}$ with $J \geq J_0$ via expanding the function f using the wavelet basis $\{\psi_{l,k}\}$. For given $\alpha \in (0, 1)$, consider the $(1 - \alpha)$-Castillo–Nickl credible band for f based on a sieve prior Π_β and a sequence $w = (w_1, w_2, \ldots)$ such that $\min_{0 \leq k \leq 2^l - 1} \kappa_{l,k} w_l / \sqrt{l} \uparrow \infty$ as $l \to \infty$:

$$\mathcal{C}_w(\widehat{f}_\infty, \widehat{R}_\alpha) := \left\{ f : \max_{(l,k) \in \mathcal{I}_\infty} \frac{|\langle f - \widehat{f}_\infty, \psi_{l,k}\rangle|}{w_l} \leq \widehat{R}_\alpha \right\},$$

where the centering estimator is $\widehat{f}_\infty := \sum_{(l,k) \in \mathcal{I}_\infty} \psi_{l,k} \kappa^{-1}_{l,k} \int v^{(1)}_{l,k} \, dY$, which converges almost surely in $\mathcal{M}_0(w)$. See p. 13 of the supplement for well-definedness of $\widehat{f}_\infty$. In linear inverse problems, the radius $\widehat{R}_\alpha$ is chosen in such a way as $\Pi_\beta(\mathcal{C}_w(\sum_{(l,k) \in \mathcal{I}(J)} \langle \widehat{f}_\infty, \psi_{l,k}\rangle \psi_{l,k}, \widehat{R}_\alpha) \mid Y) = 1 - \alpha$, where $\Pi_\beta(\cdot \mid Y)$ is the quasi-posterior under the likelihood of the truncated indirect Gaussian sequence model: $\int v^{(1)}_{l,k} \, dY = \kappa_{l,k} \beta_{l,k} + \frac{1}{\sqrt{n}} \int v^{(1)}_{l,k} \, dW$ for $(l,k) \in \mathcal{I}(J)$. This slight modification using the quasi-posterior as well as truncating the centering estimator is required to apply the main theorem; see the proof sketch below.

The following theorem derives bounds on the coverage error of the Castillo–Nickl credible band in the linear inverse problem. We use the same notation $\overline{w}_J$ as in the previous section. Let $u_J := \inf_{J \leq l, 0 \leq k \leq 2^l - 1} \kappa_{l,k} w_l / \sqrt{l}$ and $v_J := \sup_{J_0 \leq l \leq J - 1, 0 \leq k \leq 2^l - 1} \kappa_{l,k} w_l / \sqrt{l}$. In addition, let $\overline{\kappa}_J := \max_{(l,k) \in \mathcal{I}(J)} \kappa_{l,k}$ and let $\underline{\kappa}_J := \min_{(l,k) \in \mathcal{I}(J)} \kappa_{l,k}$. Let Σ be denote the $2^J \times 2^J$ covariance matrix of $\{\int v^{(1)}_{l,k} \, dY : (l,k) \in \mathcal{I}(J)\}$. Finally, let $\widetilde{H}_3 = \{Y : \sup_{J \leq l, 0 \leq k \leq 2^l - 1} |\langle f - \widehat{f}_\infty, \psi_{l,k}\rangle|/w_l \leq \widehat{R}_\alpha\}$. For simplicity, we assume that $1 \leq \{J^{1/2}/(\overline{\kappa}_J \overline{w}_J)\} u_J \uparrow \infty$ as $J \to \infty$.

Proposition 3.3. *Under Conditions* 2.1 *and* 2.3 *for* Π_β *with* $p = 2^J$, $X = \Sigma^{-1/2} \operatorname{diag}\{\kappa_{l,k} : (l,k) \in \mathcal{I}(J)\}$, *and* $\sigma_0 = 1$, *there exist positive constants* c_1, c_2 *depending only on* C_1 *appearing in Condition* 2.1, K, *and* $\{\psi_{l,k} : (l,k) \in \mathcal{I}_\infty\}$ *such that the following hold. For* $n \geq 2$, *we have*

$$\left| \mathbb{P}\big(f_0 \in \mathcal{C}_w(\widehat{f}_\infty, \widehat{R}_\alpha)\big) - (1 - \alpha) \right| \leq \phi_{\Pi_\beta}\big(c_1 \sqrt{2^J \log n}\big) + c_1 n^{-c_2 2^J} + \mathbb{P}(Y \notin \widetilde{H}_3).$$

In addition, there exist positive constants $c_3, c_4 > 0$ *depending only on* α, K, *and* $\{\psi_{l,k} : (l,k) \in \mathcal{I}_\infty\}$ *such that the following hold. Assume that the right hand side above except* $\mathbb{P}(Y \notin \widetilde{H}_3)$ *is smaller than* $\min\{\alpha/2, (1 - \alpha)/2\}$. *Then,*

$$\mathbb{P}(Y \notin \widetilde{H}_3) \leq c_3\big(\mathrm{e}^{-c_4 J\{J/(\overline{\kappa}_J \overline{w}_J)\}^2\} u_J^2} + n^{-c_2 2^J}\big)$$

for sufficiently large J depending only on α, $\{w_l\}$, K, and $\{\psi_{l,k} : (l,k) \in \mathcal{I}_\infty\}$; and the L^∞-diameter of $\mathcal{C}_w^B(\widehat{f}_\infty, \widehat{R}_\alpha) := \mathcal{C}_w(\widehat{f}_\infty, \widehat{R}_\alpha) \cap \{f : \|f\|_{B_{\infty,\infty}^s} \le B\}$ for any $B > 0$ is bounded from above as

$$\sup_{f,g \in \mathcal{C}_w^B(\widehat{f}_\infty, \widehat{R}_\alpha)} \|f - g\|_\infty \le c_3\left(v_J\sqrt{\frac{2^J J}{\underline{\kappa}_J^2 n}} + 2^{-Js}B\right)$$

with probability at least $1 - c_1 n^{-c_2 2^J}$.

Proof sketch of Proposition 3.3. The proof is almost the same as that of Proposition 3.1, but it requires an additional analysis due to the non-orthogonality of $\{v_{l,k}^{(1)} : (l,k) \in \mathcal{I}_\infty\}$. First, we transform the indirect Gaussian white noise model into an indirect Gaussian sequence model via $\{v_{l,k}^{(1)} : (l,k) \in \mathcal{I}_\infty\}$: $\widetilde{Y}_{l,k} = \kappa_{l,k}\beta_{0,l,k} + \widetilde{\varepsilon}_{l,k}$, $(l,k) \in \mathcal{I}_\infty$, where $\beta_{0,l,k} := \langle f_0, \psi_{l,k}\rangle$ for $(l,k) \in \mathcal{I}_\infty$ and $\widetilde{\varepsilon}_{l,k}$ are (dependent) jointly Gaussian variables. Then

$$\mathbb{P}\big(f_0 \in \mathcal{C}_w(\widehat{f}_\infty, \widehat{R}_\alpha)\big) = \mathbb{P}\Big(\sup_{(l,k) \in \mathcal{I}_\infty} |\kappa_{l,k}^{-1}\widetilde{Y}_{l,k} - \beta_{0,l,k}|/w_l \le \widehat{R}_\alpha\Big).$$

Second, we apply Theorem 2.1. Let $\widetilde{Y}_\infty = \{\widetilde{Y}_{l,k} : (l,k) \in \mathcal{I}_\infty\}$ and observe that $\mathbb{P}(Y \notin \widetilde{H}_3) = \mathbb{P}(\widetilde{Y}_\infty \notin \widetilde{H}_3')$ with $\widetilde{H}_3' = \{\widetilde{Y}_\infty : \sup_{J \le l, 0 \le k \le 2^l - 1} |\kappa_{l,k}^{-1}\widetilde{Y}_{l,k} - \beta_{0,l,k}|/w_l \le \widehat{R}_\alpha\}$. Then

$$\left|\mathbb{P}\big(f_0 \in \mathcal{C}_w(\widehat{f}_\infty, \widehat{R}_\alpha)\big) - \mathbb{P}\Big(\max_{(l,k) \in \mathcal{I}(J)} |\kappa_{l,k}^{-1}\widetilde{Y}_{l,k} - \beta_{0,l,k}|/w_l \le \widehat{R}_\alpha\Big)\right| \le \mathbb{P}(\widetilde{Y}_\infty \notin \widetilde{H}_3').$$

Consider the linear regression model with $p = 2^J$, $Y = \Sigma^{-1/2}(\widetilde{Y}_{J_0-1,0}, \dots, \widetilde{Y}_{J-1,2^{J-1}-1})^\top$, $X = \Sigma^{-1/2}\text{diag}\{\kappa_{l,k} : (l,k) \in \mathcal{I}(J)\}$, $\beta_0 = (\beta_{0,J_0-1,0}, \dots, \beta_{0,J-1,2^{J-1}-1})^\top$, $r = 0$, $\sigma_0 = 1$, and $\varepsilon = \Sigma^{-1/2}(\widetilde{\varepsilon}_{J_0-1,0}, \dots, \widetilde{\varepsilon}_{J-1,2^{J-1}-1})^\top \sim \mathcal{N}(0, I_p)$. For this model, the OLS estimator is $\widehat{\beta} = (X^\top X)^{-1}X^\top Y = (\kappa_{l,k}^{-1}\widetilde{Y}_{l,k})_{(l,k) \in \mathcal{I}(J)}$, and so

$$\mathbb{P}\Big(\max_{(l,k) \in \mathcal{I}(J)} |\kappa_{l,k}^{-1}\widetilde{Y}_{l,k} - \beta_{0,l,k}|/w_l \le \widehat{R}_\alpha\Big) = \mathbb{P}\big(\beta_0 \in I(\widehat{\beta}, \widehat{R}_\alpha)\big)$$

with weights $w_{l,k} = w_l$ for $(l,k) \in \mathcal{I}(J)$. Thus we can apply Theorem 2.1 to obtain bounds on $\mathbb{P}(\max_{(l,k) \in \mathcal{I}(J)} |\kappa_{l,k}^{-1}\widetilde{Y}_{l,k} - \beta_{0,l,k}|/w_l \le \widehat{R}_\alpha)$ and $\widehat{R}_\alpha$. It remains to bound $\mathbb{P}(\widetilde{Y}_\infty \notin \widetilde{H}_3')$, which is similar to the final step of the proof of Proposition 3.3. The detail can be found in Appendix C.3 of [51]. □

Remark 3.5 (Coverage error rates in linear inverse problems). Consider a locally log-Lipschitz prior with locally log-Lipschitz constant $L = L_n$. We assume a true function f_0 with $\|f_0\|_{B_{\infty,\infty}^s} \le B$ for some $B = B_n$. Set J as follows: for a (positive) constant c with $c < 1/(2r)$,

$$2^J = \begin{cases} (n/\log n)^{1/(2s+2r+1)} & \text{in mildly ill-posed cases (Case M);} \\ c\log n & \text{in severely ill-posed cases (Case S).} \end{cases}$$

Set $w_l = (\max_{0 \le k \le 2^l - 1} \kappa_{l,k})^{-1} \sqrt{l}$ for $l \le J - 1$ and $w_l = u_l(\min_{0 \le k \le 2^l - 1} \kappa_{l,k})^{-1} \sqrt{l}$ for $l \ge J$ with $u_l \uparrow \infty$ as $l \to \infty$. Then we have

$$\left|\mathbb{P}\big(f_0 \in \mathcal{C}_w(\widehat{f}, \widehat{R}_\alpha)\big) - (1 - \alpha)\right| \le \begin{cases} O\big(L_n(n/\log n)^{-s/(2s+2r+1)}\big) & \text{in Case M,} \\ O\big(L_n(\log n)^{-s}\big) & \text{in Case S,} \end{cases} \quad \text{and}$$

$$\sup_{f, g \in \mathcal{C}_w^{B_n}(\widehat{f}_\infty, \widehat{R}_\alpha)} \|f - g\|_\infty \le \begin{cases} O\big(B_n(n/\log n)^{-s/(2s+2r+1)}\big) & \text{in Case M,} \\ O\big(B_n(\log n)^{-s}\big) & \text{in Case S,} \end{cases}$$

where the latter holds with probability at least $1 - c_1 n^{-c_2 2^J}$ (again the sequence $\{w_l\}$ here depends on n but we can apply Proposition 3.3; see Remark C.2 in [51] for the detail).

3.3. Nonparametric regression model

Finally, we consider a nonparametric regression model

$$Y_i = f_0(T_i) + \varepsilon_i, \quad i = 1, \ldots, n,$$

where $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^\top$ is the vector of i.i.d. error terms with mean zero and variance σ_0^2 and $T_1, \ldots, T_n$ are an i.i.d. sample with values in $[0, 1]$. For simplicity, we assume that ε and $\{T_i : i = 1, \ldots, n\}$ are independent, and σ_0 does not depend on n.

We consider a sieve prior for f_0. To this end, we use p basis functions $\{\psi_j^p(\cdot) : 1 \le j \le p\}$, and constrict a credible band for f of the form

$$\mathcal{C}(\widehat{f}, \widehat{R}_\alpha) = \left\{ f : \left\| \frac{f(\cdot) - \widehat{f}(\cdot)}{\|\psi^p(\cdot)\|} \right\|_\infty \le \widehat{R}_\alpha \right\},$$

where $\widehat{f}(\cdot) := \sum_{j=1}^p \psi_j^p(\cdot)\widehat{\beta}_j$ with $\widehat{\beta} := \operatorname{argmin}_\beta \sum_{i=1}^n (Y_i - \sum_{j=1}^p \psi_j^p(T_i)\beta_j)^2$, $\widehat{R}_\alpha$ is chosen in such a way that $\Pi_f\{\mathcal{C}(\widehat{f}, \widehat{R}_\alpha) \mid Y\} = 1 - \alpha$, and $\psi^p(\cdot) := (\psi_1^p(\cdot), \ldots, \psi_p^p(\cdot))^\top$. We consider a prior Π_f of f induced from a sieve prior Π_β on $\mathbb{R}^p$ via the map $(\beta_1, \ldots, \beta_p) \mapsto \sum_{j=1}^p \beta_j \psi_j^p(\cdot)$.

The setting of the nonparametric regression is different from that of Section 2 in that the regressors $T_1, \ldots, T_n$ are stochastic. Due to this additional randomness, we need an additional analysis to develop bounds on the coverage error and the L^∞-diameter of the band. To this end, we modify Conditions 2.1 and 2.3, and add conditions on the basis functions Let $\widetilde{\psi}^p(\cdot) := \psi^p(\cdot)/\|\psi^p(\cdot)\|$, $\xi_p := \sup_{t \in [0,1]} \|\psi^p(t)\|$, and $\beta_0 := \operatorname{argmin}_\beta \mathbb{E}[(f_0(T_1) - \psi^p(T_1)^\top \beta)^2]$. For $R > 0$, let

$$\widetilde{B}(R) := \big\{\beta : \|\beta - \beta_0\| \le n^{-1/2}R\big\} \quad \text{and} \quad \widetilde{\phi}_{\Pi_\beta}(R) := 1 - \inf_{\beta, \widetilde{\beta} \in \widetilde{B}(R)} \frac{\pi(\beta)}{\pi(\widetilde{\beta})}.$$

Condition 3.1. *There exists a positive constant C_1 such that $\pi(\beta_0) \ge n^{-C_1 p}$.*

Condition 3.2. *The inequality $\widetilde{\phi}_{\Pi_\beta}(1/\sqrt{n}) \le 1/2$ holds.*

Condition 3.3. *There exist strictly positive constants $\underline{b}$ and $\overline{b}$ such that the eigenvalues of the $p \times p$ matrix $(\mathbb{E}[\psi_i^p(T_1)\psi_j^p(T_1)])_{1 \leq i,j \leq p}$ are included in $[\underline{b}^2, \overline{b}^2]$.*

Condition 3.4. *There exist positive constants C_4 and C_5 such that*

$$\log \xi_p \leq C_4 \log p \quad and \quad \log \sup_{t \neq t' \in [0,1]} \frac{\|\widetilde{\psi}^p(t) - \widetilde{\psi}^p(t')\|}{|t - t'|} \leq C_5 \log p.$$

Conditions 3.1 and 3.2 are versions of Conditions 2.1 and 2.3 under stochastic regressors. Condition 3.3 is standard. Condition 3.4 is not restrictive; for example, this condition holds for Fourier series, Spline series, CDV wavelets, and local polynomial partition series; see [4] for details.

The following proposition derives bounds on the coverage error and the L^∞-diameter of $\mathcal{C}(\widehat{f}, \widehat{R}_\alpha)$. Let $\tau_2 := \sqrt{\mathbb{E}[(f_0(T_1) - \psi^p(T_1)^\top \beta_0)^2]}$, $\tau_\infty := \|f_0(\cdot) - \psi^p(\cdot)^\top \beta_0\|_\infty$, and $\tau := \||f_0(\cdot) - \psi^p(\cdot)^\top \beta_0|/\|\psi^p(\cdot)\|\|_\infty$. These parameters quantify the approximation errors by the basis functions.

Proposition 3.4. *Under Conditions 3.1–3.4 together with Conditions 2.2 and 2.5, there exist positive constants c_1, c_2 depending only on $C_1, \ldots, C_5, \underline{b}, \overline{b}$, and q appearing in these conditions such that the following hold. For $n \geq 2$ and any sufficiently small $\delta > 0$, we have*

$$\left|\mathbb{P}\left(f_0 \in \mathcal{C}(\widehat{f}, \widehat{R}_\alpha)\right) - (1 - \alpha)\right|$$
$$\leq \widetilde{\phi}_{\Pi_\beta}(c_1\sqrt{p \log n}) + \delta_2 + \delta_3 + c_1\left(n^{-2\delta} + \delta_1 p \log n + \zeta_n + \gamma_n\right), \quad (9)$$

where

$$\gamma_n := \frac{n}{\log n}\frac{\tau_2^2}{p} + \max\left\{1, \left(p\xi_p^2/n\right)^{1/2}\right\}\tau_\infty n^\delta \log p + \sqrt{n}\tau\sqrt{\log p} \quad and$$

$$\zeta_n := \begin{cases} n^\delta(\log n)^{7/6} \max\left\{\left(\frac{\xi_p^2}{n}\right)^{1/2} n^{1/q}(\log n)^{1/3}, \left(\frac{\xi_p^2}{n}\right)^{1/6}\right\} & under\ Condition\ 2.5(a), \\ n^\delta(\log n)^{7/6}\left(\xi_p^2/n\right)^{1/6} & under\ Condition\ 2.5(b). \end{cases}$$

In addition, there exists a positive constant c_3 depending only on α and $\underline{b}$ such that the following holds: *provided that the right-hand side on* (9) *is smaller than $\alpha/2$, we have*

$$\sup_{f,g \in \mathcal{C}(\widehat{f},\widehat{R}_\alpha)} \|f - g\|_\infty \leq c_3\sqrt{\xi_p^2(\log p)/n}$$

with probability at least $1 - \delta_3 - c_1\{\sqrt{n}\tau\sqrt{\log p} + n^{-c_2 p}\}$.

We note that the proof of Proposition 3.4 does not use a lower bound on $\widehat{R}_\alpha$ in Theorem 2.1 (more precisely, its version for random designs). Hence, we do not have to assume that the right-hand side on (9) is smaller than $(1 - \alpha)/2$; see the discussion after Theorem 2.1.

Remark 3.6 (Magnitudes of ξ_p, τ_2, τ_∞, and τ). For typical basis functions including Fourier series, spline series, and CDV wavelets, we have $\xi_p \lesssim \sqrt{p}$; see Section 3 in [4]. If f_0 is in the Hölder–Zygmund space with smoothness level $s > 0$, then $\tau_2 \sim \tau_\infty \sim p^{-s}$ for an S-regular CDV wavelet basis with $S > s$. For other bases and other function classes, bounds on τ_2 and τ_∞ can be found in approximation theory; see, for example, [20] and Section 3 in [4]. Finally, for the Haar wavelet basis, we have $\tau \sim \tau_\infty/\sqrt{p}$, since $\tau \leq \tau_\infty/\inf_{t\in[0,1]} \|\psi^p(t)\|$; for periodic S-regular wavelets, we also have $\tau \sim \tau_\infty/\sqrt{p}$ as shown in Appendix C.4.3 of [51].

Remark 3.7 (Coverage error rates for the true function). Consider the unknown variance case. Assume that there exists a constant $s > 1/2$ such that $\tau_2 \sim \tau_\infty \sim p^{-s}$, $\tau \sim p^{-s-1/2}$, and $\xi_p \lesssim \sqrt{p}$. Assume also that the error distribution is Gaussian (for the non-Gaussian case, add ζ_n to the bound on the coverage error). We use a locally log-Lipschitz prior with locally log-Lipschitz constant $L = L_n$ on β and use the estimator $\widehat{\sigma}^2 = \widehat{\sigma}_u^2$ as in Proposition 2.3. Take $p \sim (n/\log n)^{1/(2s+1)} b_n$ with a positive nondecreasing sequence $b_n = O(\log n)$. In this case, we have

$$\left| P\big(f_0 \in \mathcal{C}(\widehat{f}, \widehat{R}_\alpha)\big) - (1-\alpha) \right|$$
$$\leq C\left[L_n \left(\frac{n}{\log n}\right)^{-s/(2s+1)} b_n^{1/2} + \left(\frac{n}{\log n}\right)^{-(s-1/2)/(2s+1)} b_n \log n + \frac{\log n}{b_n^{s+1/2}} \right] \quad \text{and}$$

$$\sup_{f,g\in\mathcal{C}(\widehat{f},\widehat{R}_\alpha)} \|f - g\|_\infty \leq C\left(\frac{n}{\log n}\right)^{-s/(2s+1)} b_n^{1/2},$$

where the latter holds with probability at least $1 - c_1 (\log n)/b_n^{s+1/2}$, and the constant C is independent of n.

Remark 3.8 (Coverage error rates for the surrogate function). Consider coverage errors for the surrogate function $f_{0,p} := \psi^p(\cdot)^\top \beta_0$ when the error distribution is Gaussian. In this case, since $\tau_\infty = \tau_2 = \tau = 0$, we have

$$\left| \mathbb{P}(f_{0,p} \in \mathcal{C}(\widehat{f}, \widehat{R}_\alpha) - (1-\alpha) \right| \leq O\big((n/\log n)^{-(s-1/2)/(2s+1)} b_n \log n\big) \quad \text{and}$$

$$\sup_{f,g\in\mathcal{C}(\widehat{f},\widehat{R}_\alpha)} \|f - g\|_\infty \leq O\big((n/\log n)^{-s/(2s+1)} b_n^{1/2}\big),$$

where the latter holds with probability at least $1 - c_1 \exp\{-c_2 (n/\log n)^{1/(2s+1)}\}$. This shows that Bayesian credible bands have coverage errors (for the surrogate function) decaying polynomially fast in the sample size n in nonparametric regression models.

4. Proof of Theorem 2.1

4.1. Supporting lemmas

We begin with stating some supporting lemmas that will be used in the proof of Theorem 2.1. They include the high-dimensional CLT on hyperrectangles, the anti-concentration inequality for the Gaussian distribution, Anderson's lemma, and the concentration inequality for the Gaussian maximum.

The high-dimensional CLT on hyperrectangles is stated as follows: in the following lemma, let $Z_1, \ldots, Z_n$ be independent p-dimensional random vectors with mean zero. Let Z_{ij} ($i = 1, \ldots, n$, $j = 1, \ldots, p$) denote the jth coordinate of Z_i. Let $\widetilde{Z}_1, \ldots, \widetilde{Z}_n$ be independent centered p-dimensional Gaussian vectors such that each $\widetilde{Z}_i$ has the same covariance matrix as Z_i. Let $\mathcal{A}^{\text{re}}$ be the class of all closed hyperrectangles in $\mathbb{R}^p$: for any $A \in \mathcal{A}^{\text{re}}$, A is of the form $A = \{\beta \in \mathbb{R}^p : \underline{a}_i \leq \beta_i \leq \overline{a}_i, 1 \leq \forall i \leq p\}$ with $(\underline{a}_1, \ldots, \underline{a}_p)^\top \in \mathbb{R}^p$ and $(\overline{a}_1, \ldots, \overline{a}_p)^\top \in \mathbb{R}^p$. We assume the following three conditions:

H1. There exists $b > 0$ such that $n^{-1} \sum_{i=1}^n \mathbb{E}[Z_{ij}^2] \geq b$ for all $1 \leq j \leq p$;

H2. There exists a sequence $B_n \geq 1$ such that $n^{-1} \sum_{i=1}^n \mathbb{E}[|Z_{ij}|^{2+k}] \leq B_n^4$ for all $1 \leq j \leq p$ and for $k = 1, 2$;

H3. Either one of the following two conditions holds:
   (a) There exists an integer $4 \leq q < \infty$ such that $\mathbb{E}[(\max_{1 \leq j \leq p} |Z_{ij}|/B_n)^q] \leq 1$ for all $1 \leq i \leq n$;
   (b) $\mathbb{E}[\exp(|Z_{ij}|/B_n)] \leq 2$ for all $1 \leq i \leq n$ and $1 \leq j \leq p$.

Lemma 4.1 (High dimensional CLT on hyperrectangles; Proposition 2.1 in [15]). *Let*

$$\rho = \rho_n := \sup_{A \in \mathcal{A}^{\text{re}}} \left| \mathbb{P}\left(\sum_{i=1}^n Z_i / \sqrt{n} \in A \right) - \mathbb{P}\left(\sum_{i=1}^n \widetilde{Z}_i / \sqrt{n} \in A \right) \right|.$$

Under Conditions H1–H3, *there exists a positive constant* $\widetilde{c}_1$ *such that*

$$\rho \leq \begin{cases} \widetilde{c}_1 \left(\dfrac{B_n^2 \log^7(pn)}{n} \right)^{1/6} + \widetilde{c}_1 \left(\dfrac{B_n^2 \log^3(pn)}{n^{1-2/q}} \right)^{1/3} & \text{under Condition H3(a),} \\ \widetilde{c}_1 \left(\dfrac{B_n^2 \log^7(pn)}{n} \right)^{1/6} & \text{under Condition H3(b).} \end{cases}$$

The constant $\widetilde{c}_1$ *depends only on* b *appearing in Condition* H1 *and* q *appearing in Condition* H3.

Next, we state the anti-concentration inequality for the Gaussian distribution, Anderson's lemma, and the concentration inequality for the Gaussian maximum.

Lemma 4.2 (Anti-concentration inequality for the Gaussian distribution; [38]). *Let* $Z = (Z_1, \ldots, Z_p)^\top$ *be a centered Gaussian random vector in* $\mathbb{R}^p$ *with* $\sigma_j^2 := \mathbb{E}[Z_j]^2 > 0$ *for all*

$1 \leq j \leq p$. Let $\underline{\sigma} := \min\{\sigma_j\}$. *There exists a universal positive constant* $\widetilde{c}_2$ *such that for every* $z = (z_1, \ldots, z_p)^\top \in \mathbb{R}^p$ *and* $R > 0$,

$$\gamma := \gamma(R) := \mathbb{P}(Z_j \leq z_j + R \ 1 \leq \forall j \leq p) - \mathbb{P}(Z_j \leq z_j \ 1 \leq \forall j \leq p) \leq \widetilde{c}_2 \frac{R}{\underline{\sigma}} \sqrt{\log p}.$$

Lemma 4.3 (Anderson's lemma; Corollary 3 in [2]). *Let* Σ *and* $\widetilde{\Sigma}$ *be symmetric positive semidefinite* $p \times p$ *matrices, and let* $\mathcal{C}$ *be a symmetric convex set in* $\mathbb{R}^p$. *If* $\Sigma - \widetilde{\Sigma}$ *is positive semidefinite, then* $\mathbb{P}(Z \in \mathcal{C}) \leq \mathbb{P}(\widetilde{Z} \in \mathcal{C})$ *for* $Z \sim \mathcal{N}(0, \Sigma)$ *and* $\widetilde{Z} \sim \mathcal{N}(0, \widetilde{\Sigma})$,

Lemma 4.4 (Concentration inequality for the Gaussian maximum; Theorem 2.5.8. in [30]). *Let* $N_1, \ldots, N_p \sim \mathcal{N}(0, 1)$ *i.i.d. and let* $\{w_i\}_{i=1}^p$ *be a positive sequence with* $\underline{w} = \min_{1 \leq i \leq p} w_i$. *Then for every* $R > 0$,

$$\mathbb{P}\left(\left|\max_{1 \leq i \leq p}|N_i/w_i| - \mathbb{E}\left[\max_{1 \leq i \leq p}|N_i/w_i|\right]\right| \geq R\right) \leq 2\exp\left(-\underline{w}^2 R^2/2\right).$$

4.2. Proof of Theorem 2.1

We only prove the theorem under Condition 2.5(a). The proof under Condition 2.5(b) is done by replacing Lemma 4.1(a) by Lemma 4.1(b).

The proof is divided into two parts. We first derive an upper bound on the coverage error $|\mathbb{P}(\beta_0 \in I(\widehat{\beta}(Y), \widehat{R}_\alpha)) - (1 - \alpha)|$ and then bound the radius $\widehat{R}_\alpha$ of $I(\widehat{\beta}(Y), \widehat{R}_\alpha)$.

Step 1: *Upper bound on the coverage error*

We start with proving that $\widehat{R}_\alpha$ concentrates on the $(1-\alpha)$-quantile of some distribution with high probability. Let $\overline{\zeta}$ be the upper bound in Proposition 2.5. From Proposition 2.5, we have

$$\underbrace{\left|\Pi_\beta\big(I\big(\widehat{\beta}(Y), \widehat{R}_\alpha\big) \mid Y\big)\right.}_{=1-\alpha} - \mathcal{N}\big(I\big(\widehat{\beta}(Y), \widehat{R}_\alpha\big) \mid \widehat{\beta}(Y), \sigma_0^2\big(X^\top X\big)^{-1}\big)\big| \leq \overline{\zeta} \quad \text{for } Y \in H,$$

where recall that $H = \{Y : \|X(\widehat{\beta}(Y) - \beta_0)\| \leq c_1\sqrt{p \log n}\sigma_0/4\} \cap \{Y : \Pi_{\sigma^2}(|\sigma^2/\sigma_0^2 - 1| \geq \delta_1 \mid Y) \leq \delta_2\}$. Let $\widetilde{S} \sim \mathcal{N}(0, (X^\top X)^{-1})$ and let G be the distribution function of $\sigma_0 \max\{|e_{(p),i}^\top \widetilde{S}|/w_i\}$, where $e_{(p),i}$ is the p-dimensional unit vector whose ith component is 1. Now since $\mathcal{N}(I(\widehat{\beta}(Y), \widehat{R}_\alpha) \mid \widehat{\beta}(Y), \sigma_0^2(X^\top X)^{-1}) = G(\widehat{R}_\alpha)$, we have $|(1 - \alpha) - G(\widehat{R}_\alpha)| \leq \overline{\zeta}$ for $Y \in H$. This implies

$$G^{-1}(1 - \alpha - \overline{\zeta}) \leq \widehat{R}_\alpha \leq G^{-1}(1 - \alpha + \overline{\zeta}) \quad \text{for } Y \in H, \tag{10}$$

where G^{-1} denotes the quantile function of G.

Next, we will derive an upper bound on $\mathbb{P}(\beta_0 \in I(\widehat{\beta}(Y), \widehat{R}_\alpha)) - (1 - \alpha)$ (the lower bound follows similarly). Let ρ be the constant in Lemma 4.1 when $Z_j = n(X^\top X)^{-1}X_{j\cdot}\varepsilon_j$ for $j = 1, \ldots, n$, where $X_{j\cdot} = (X_{j1}, \ldots, X_{jp})^\top$ for $j = 1, \ldots, n$. For $R > 0$, let $\gamma(R)$ be the constant

in Lemma 4.2 when $Z = \sigma_0 \widetilde{S}$. Finally, let $\widetilde{r} := (X^\top X)^{-1} X^\top r$. From inequality (10) and by the definitions of ρ, G, and γ, we have

$$
\begin{aligned}
&\mathbb{P}\big(\beta_0 \in I(\widehat{\beta}(Y), \widehat{R}_\alpha)\big) - (1 - \alpha) \\
&\quad \leq \mathbb{P}\Big(\max_{1 \leq i \leq p} \big\{ |e_{(p),i}^\top (X^\top X)^{-1} X^\top (\varepsilon + r)|/w_i \big\} \leq G^{-1}(1 - \alpha + \overline{\zeta}) \Big) - (1 - \alpha) + \mathbb{P}(Y \notin H) \\
&\quad \leq \mathbb{P}\Big(\max_{1 \leq i \leq p} \big\{ |e_{(p),i}^\top (\sigma_0 \widetilde{S} + \widetilde{r})|/w_i \big\} \leq G^{-1}(1 - \alpha + \overline{\zeta}) \Big) - (1 - \alpha) + \rho + \mathbb{P}(Y \notin H) \\
&\quad \leq \gamma\big(\|\widetilde{r}\|_\infty\big) + \overline{\zeta} + \rho + \mathbb{P}(Y \notin H).
\end{aligned}
$$

Proposition 2.6 gives an upper bound on $\mathbb{P}(Y \notin H)$. From Lemmas 4.1 and 4.2, we obtain the following bounds on ρ and γ: For some $\widetilde{c}_1 > 0$ depending only on q,

$$
\rho \leq \widetilde{c}_1 \left\{ \left(\frac{p \log^7(pn)}{n} \frac{\overline{\lambda}}{\underline{\lambda}} \right)^{1/6} + \left(\frac{p \log^3(pn)}{n^{1 - 2/q}} \frac{\overline{\lambda}}{\underline{\lambda}} \right)^{1/3} \right\} \quad \text{and} \quad \gamma \leq \widetilde{c}_1 \frac{\|\widetilde{r}\|_\infty}{\sigma_0 \underline{\lambda}^{1/2}} \sqrt{\log p},
$$

which completes Step 1.

Step 2: *Upper bound on the max-diameter*

We start with deriving a high-probability upper bound on $\widehat{R}_\alpha$ using the quantile function F^{-1} of $\max_{1 \leq i \leq p} |N_i/w_i|$ for independent standard Gaussian random variables $\{N_i : i = 1, \ldots, p\}$. From Lemma 4.3, we have

$$
\mathbb{P}\Big(\max_{1 \leq i \leq p} |N_i/w_i| \leq R/(\sigma_0 \overline{\lambda}^{1/2})\Big) \leq \mathbb{P}\Big(\max_{1 \leq i \leq p} |\sigma_0 \widetilde{S}_i/w_i| \leq R\Big) \quad \text{for } R > 0.
$$

Together with inequality (10), we have

$$
\widehat{R}_\alpha \leq \sigma_0 \overline{\lambda}^{1/2} F^{-1}(1 - \alpha + \overline{\zeta}) \quad \text{for } Y \in H. \tag{11}
$$

Next, we will bound $F^{-1}(1 - \alpha + \overline{\zeta})/\mathbb{E}[\max_{1 \leq i \leq p} |N_i/w_i|]$. From Lemma 4.4, there exists $\widetilde{c}_2 > 1$ depending only on α and $\underline{w}$ such that

$$
\mathbb{P}\Big(\max_{1 \leq i \leq p} |N_i/w_i| - \mathbb{E}\Big[\max_{1 \leq i \leq p} |N_i/w_i|\Big] \geq \widetilde{c}_2 \mathbb{E}\Big[\max_{1 \leq i \leq p} |N_i/w_i|\Big]\Big) < \alpha - \alpha/2 < \alpha - \overline{\zeta}.
$$

Therefore, by the definition of F^{-1}, we have

$$
F^{-1}(1 - \alpha + \overline{\zeta}) = \inf \Big\{ R : \mathbb{P}\Big(\max_{1 \leq i \leq p} |N_i/w_i| \geq R\Big) \leq \alpha - \overline{\zeta} \Big\}
$$

$$
\leq (1 + \widetilde{c}_2) \mathbb{E}\Big[\max_{1 \leq i \leq p} |N_i/w_i|\Big].
$$

Together with (11), we obtain the desired upper bound on $\widehat{R}_\alpha$.

Step 3: *Lower bound on the max-diameter*

As in Step 2, we have

$$\sigma_0 \overline{\lambda}^{1/2} \overline{w}^{-1} \widetilde{F}^{-1}(1 - \alpha - \overline{\zeta}) \leq \widehat{R}_\alpha \quad \text{for } Y \in H \tag{12}$$

Next, we will show that $\widetilde{F}^{-1}(1 - \alpha - \overline{\zeta}) \geq \widetilde{c}_3 \sqrt{\log p}$ for some constant $\widetilde{c}_3$ depending only on α. From the Paley–Zygmund inequality, we have for $\theta \in [0, 1]$,

$$\mathbb{P}\left(\max_{1 \leq i \leq p} |N_i| \geq \theta \mathbb{E}\left[\max_{1 \leq i \leq p} |N_i|\right]\right) \geq (1 - \theta)^2 \frac{(\mathbb{E}[\max_{1 \leq i \leq p} |N_i|])^2}{\mathbb{E}[(\max_{1 \leq i \leq p} |N_i|)^2]}. \tag{13}$$

From Lemma 4.4 together with the inequality $\mathbb{E}[\max_{1 \leq i \leq p} |N_i|] \geq \sqrt{\log p}/12$, there exists a universal positive constant $\widetilde{c}_4$ such that

$$\mathbb{E}\left[\left(\max_{1 \leq i \leq p} |N_i|\right)^2\right] \leq \left(\mathbb{E}\left[\max_{1 \leq i \leq p} |N_i|\right]\right)^2 (1 + \widetilde{c}_4/\sqrt{\log p}), \tag{14}$$

where we have used use Lemma 4.4 to deduce that

$$\mathbb{E}\left[\left(\max_{1 \leq i \leq p} |N_i|\right)^2\right] \leq \left(\mathbb{E}\left[\max_{1 \leq i \leq p} |N_i|\right]\right)^2 + 4 \int_{\mathbb{E}[\max_{1 \leq i \leq p} |N_i|]}^{\infty} t e^{-(t - \mathbb{E}[\max_{1 \leq i \leq p} |N_i|])^2/2} \, dt$$

$$\leq \left(\mathbb{E}\left[\max_{1 \leq i \leq p} |N_i|\right]\right)^2 + \widetilde{c}_5\left(\mathbb{E}\left[\max_{1 \leq i \leq p} |N_i|\right] + 1\right)$$

for some universal positive constant $\widetilde{c}_5$. Let $\eta := (1 + \alpha)/2$. Take p such that $1/\{1 + \widetilde{c}_4/\sqrt{\log p}\} \geq (\eta + 1)/2$, and take $\theta_\alpha^* = 1 - \sqrt{(2\eta)/(\eta + 1)}$. Then, from inequalities (13) and (14), we have

$$\mathbb{P}\left(\max_{1 \leq i \leq p} |N_i| \geq \theta_\alpha^* \mathbb{E}\left[\max_{1 \leq i \leq p} |N_i|\right]\right) \geq \left(1 - \theta_\alpha^*\right)^2 \frac{(\mathbb{E}[\max_{1 \leq i \leq p} |N_i|])^2}{\mathbb{E}[(\max_{1 \leq i \leq p} |N_i|)^2]} \geq \eta \geq \alpha + \overline{\zeta}.$$

Thus we have

$$\widetilde{F}^{-1}(1 - \alpha - \overline{\zeta}) \geq \theta_\alpha^* \mathbb{E}\left[\max_{1 \leq i \leq p} |N_i|\right] \geq \left(\theta_\alpha^*/12\right)\sqrt{\log p}. \tag{15}$$

Together with (12), we obtain the desired lower bound on $\widehat{R}_\alpha$. □

5. Conclusion

We have studied finite sample bounds on frequentist coverage errors of Bayesian credible rectangles to approximately linear regression models with moderately high dimensional regressors. As an application, we have shown that Bayesian credible bands have coverage errors (for the true function) decaying polynomially fast in the sample size in Gaussian white noise models

and linear inverse problems; the similar results hold for the surrogate function in nonparametric regression models. This supports the use of Bayesian approaches to constructing nonparametric confidence bands.

Acknowledgements

Supplementary Material

Supplement to "On frequentist coverage errors of Bayesian credible sets in high dimensions" (DOI: 10.3150/19-BEJ1142SUPP; .pdf). The supplementary material contains the proofs omitted in the main text.

References

[1] Abramovich, F. and Silverman, B.W. (1998). Wavelet decomposition approaches to statistical inverse problems. *Biometrika* **85** 115–129. MR1627226 https://doi.org/10.1093/biomet/85.1.115

[2] Anderson, T.W. (1955). The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities. *Proc. Amer. Math. Soc.* **6** 170–176. MR0069229 https://doi.org/10.2307/2032333

[3] Atchadé, Y.A. (2017). On the contraction properties of some high-dimensional quasi-posterior distributions. *Ann. Statist.* **45** 2248–2273. MR3718168 https://doi.org/10.1214/16-AOS1526

[4] Belloni, A., Chernozhukov, V., Chetverikov, D. and Kato, K. (2015). Some new asymptotic theory for least squares series: Pointwise and uniform results. *J. Econometrics* **186** 345–366. MR3343791 https://doi.org/10.1016/j.jeconom.2015.02.014

[5] Bickel, P.J. and Rosenblatt, M. (1973). On some global measures of the deviations of density function estimates. *Ann. Statist.* **1** 1071–1095. MR0348906

[6] Bontemps, D. (2011). Bernstein–von Mises theorems for Gaussian regression with increasing number of regressors. *Ann. Statist.* **39** 2557–2584. MR2906878 https://doi.org/10.1214/11-AOS912

[7] Boucheron, S. and Gassiat, E. (2009). A Bernstein–von Mises theorem for discrete probability distributions. *Electron. J. Stat.* **3** 114–148. MR2471588 https://doi.org/10.1214/08-EJS262

[8] Castillo, I. (2014). On Bayesian supremum norm contraction rates. *Ann. Statist.* **42** 2058–2091. MR3262477 https://doi.org/10.1214/14-AOS1253

[9] Castillo, I. and Nickl, R. (2013). Nonparametric Bernstein–von Mises theorems in Gaussian white noise. *Ann. Statist.* **41** 1999–2028. MR3127856 https://doi.org/10.1214/13-AOS1133

[10] Castillo, I. and Nickl, R. (2014). On the Bernstein–von Mises phenomenon for nonparametric Bayes procedures. *Ann. Statist.* **42** 1941–1969. MR3262473 https://doi.org/10.1214/14-AOS1246

[11] Castillo, I. and Rousseau, J. (2015). A Bernstein–von Mises theorem for smooth functionals in semiparametric models. *Ann. Statist.* **43** 2353–2383. MR3405597 https://doi.org/10.1214/15-AOS1336

[12] Castillo, I., Schmidt-Hieber, J. and van der Vaart, A. (2015). Bayesian linear regression with sparse priors. *Ann. Statist.* **43** 1986–2018. MR3375874 https://doi.org/10.1214/15-AOS1334

[13] Chernozhukov, V., Chetverikov, D. and Kato, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Ann. Statist.* **41** 2786–2819. MR3161448 https://doi.org/10.1214/13-AOS1161

[14] Chernozhukov, V., Chetverikov, D. and Kato, K. (2014). Anti-concentration and honest, adaptive confidence bands. *Ann. Statist.* **42** 1787–1818. MR3262468 https://doi.org/10.1214/14-AOS1235

[15] Chernozhukov, V., Chetverikov, D. and Kato, K. (2017). Central limit theorems and bootstrap in high dimensions. *Ann. Probab.* **45** 2309–2352. MR3693963 https://doi.org/10.1214/16-AOP1113

[16] Chernozhukov, V. and Hong, H. (2003). An MCMC approach to classical estimation. *J. Econometrics* **115** 293–346. MR1984779 https://doi.org/10.1016/S0304-4076(03)00100-3

[17] Claeskens, G. and Van Keilegom, I. (2003). Bootstrap confidence bands for regression curves and their derivatives. *Ann. Statist.* **31** 1852–1884. MR2036392 https://doi.org/10.1214/aos/1074290329

[18] Clarke, B. and Ghosal, S. (2010). Reference priors for exponential families with increasing dimension. *Electron. J. Stat.* **4** 737–780. MR2678969 https://doi.org/10.1214/10-EJS569

[19] Cox, D.D. (1993). An analysis of Bayesian inference for nonparametric regression. *Ann. Statist.* **21** 903–923. MR1232525 https://doi.org/10.1214/aos/1176349157

[20] DeVore, R.A. and Lorentz, G.G. (1993). *Constructive Approximation. Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **303**. Berlin: Springer. MR1261635 https://doi.org/10.1007/978-3-662-02888-9

[21] Donoho, D.L. (1995). Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition. *Appl. Comput. Harmon. Anal.* **2** 101–126. MR1325535 https://doi.org/10.1006/acha.1995.1008

[22] Florens, J.-P. and Simoni, A. (2012). Nonparametric estimation of an instrumental regression: A quasi-Bayesian approach based on regularized posterior. *J. Econometrics* **170** 458–475. MR2970326 https://doi.org/10.1016/j.jeconom.2012.05.016

[23] Freedman, D. (1999). On the Bernstein–von Mises theorem with infinite-dimensional parameters. *Ann. Statist.* **27** 1119–1140. MR1740119 https://doi.org/10.1214/aos/1017938917

[24] Gao, C. and Zhou, H.H. (2016). Bernstein–von Mises theorems for functionals of the covariance matrix. *Electron. J. Stat.* **10** 1751–1806. MR3522660 https://doi.org/10.1214/15-EJS1048

[25] Ghosal, S. (1999). Asymptotic normality of posterior distributions in high-dimensional linear models. *Bernoulli* **5** 315–331. MR1681701 https://doi.org/10.2307/3318438

[26] Ghosal, S. (2000). Asymptotic normality of posterior distributions for exponential families when the number of parameters tends to infinity. *J. Multivariate Anal.* **74** 49–68. MR1790613 https://doi.org/10.1006/jmva.1999.1874

[27] Ghosal, S. and van der Vaart, A. (2017). *Fundamentals of Nonparametric Bayesian Inference. Cambridge Series in Statistical and Probabilistic Mathematics* **44**. Cambridge: Cambridge Univ. Press. MR3587782 https://doi.org/10.1017/9781139029834

[28] Giné, E. and Nickl, R. (2010). Confidence bands in density estimation. *Ann. Statist.* **38** 1122–1170. MR2604707 https://doi.org/10.1214/09-AOS738

[29] Giné, E. and Nickl, R. (2011). Rates of contraction for posterior distributions in L^r-metrics, $1 \le r \le \infty$. *Ann. Statist.* **39** 2883–2911. MR3012395 https://doi.org/10.1214/11-AOS924

[30] Giné, E. and Nickl, R. (2016). *Mathematical Foundations of Infinite-Dimensional Statistical Models. Cambridge Series in Statistical and Probabilistic Mathematics*. New York: Cambridge Univ. Press. MR3588285 https://doi.org/10.1017/CBO9781107337862

[31] Hall, P. (1991). On convergence rates of suprema. *Probab. Theory Related Fields* **89** 447–455. MR1118558 https://doi.org/10.1007/BF01199788

[32] Hoffmann, M., Rousseau, J. and Schmidt-Hieber, J. (2015). On adaptive posterior concentration rates. *Ann. Statist.* **43** 2259–2295. MR3396985 https://doi.org/10.1214/15-AOS1341

[33] Johnstone, I. Gaussian Estimation: Sequence and Wavelet Models. Unpublished manuscript.

[34] Johnstone, I.M. (2010). High dimensional Bernstein–von Mises: Simple examples. In *Borrowing Strength*: *Theory Powering Application – a Festschrift for Lawrence D. Brown*. *Inst. Math. Stat.* (*IMS*) *Collect.* **6** 87–98. Beachwood, OH: IMS. MR2798513

[35] Kato, K. (2013). Quasi-Bayesian analysis of nonparametric instrumental variables models. *Ann. Statist.* **41** 2359–2390. MR3127869 https://doi.org/10.1214/13-AOS1150

[36] Knapik, B.T., van der Vaart, A.W. and van Zanten, J.H. (2011). Bayesian inverse problems with Gaussian priors. *Ann. Statist.* **39** 2626–2657. MR2906881 https://doi.org/10.1214/11-AOS920

[37] Leahu, H. (2011). On the Bernstein–von Mises phenomenon in the Gaussian white noise model. *Electron. J. Stat.* **5** 373–404. MR2802048 https://doi.org/10.1214/11-EJS611

[38] Nazarov, F. (2003). On the maximal perimeter of a convex set in $\mathbb{R}^n$ with respect to a Gaussian measure. In *Geometric Aspects of Functional Analysis*. *Lecture Notes in Math.* **1807** 169–187. Berlin: Springer. MR2083397 https://doi.org/10.1007/978-3-540-36428-3_15

[39] Nickl, R. (2019). Bernstein–von Mises theorems for statistical inverse problems I: Schrödinger equation. *J. Eur. Math. Soc.* (*JEMS*). To appear.

[40] Nickl, R. and Söhl, J. (2017). Nonparametric Bayesian posterior contraction rates for discretely observed scalar diffusions. *Ann. Statist.* **45** 1664–1693. MR3670192 https://doi.org/10.1214/16-AOS1504

[41] Panov, M. and Spokoiny, V. (2015). Finite sample Bernstein–von Mises theorem for semiparametric problems. *Bayesian Anal.* **10** 665–710. MR3420819 https://doi.org/10.1214/14-BA926

[42] Ray, K. (2017). Adaptive Bernstein–von Mises theorems in Gaussian white noise. *Ann. Statist.* **45** 2511–2536. MR3737900 https://doi.org/10.1214/16-AOS1533

[43] Rivoirard, V. and Rousseau, J. (2012). Bernstein–von Mises theorem for linear functionals of the density. *Ann. Statist.* **40** 1489–1523. MR3015033 https://doi.org/10.1214/12-AOS1004

[44] Smirnov, N.V. (1950). On the construction of confidence regions for the density of distribution of random variables. *Dokl. Akad. Nauk SSSR* **74** 189–191. MR0037494

[45] Sniekers, S. and van der Vaart, A. (2015). Credible sets in the fixed design model with Brownian motion prior. *J. Statist. Plann. Inference* **166** 78–86. MR3390135 https://doi.org/10.1016/j.jspi.2014.07.008

[46] Spokoiny, V. Bernstein–von Mises Theorem for growing parameter dimension. Available at arXiv:1302.3430.

[47] Szabó, B., van der Vaart, A. and van Zanten, H. (2015). Honest Bayesian confidence sets for the L^2-norm. *J. Statist. Plann. Inference* **166** 36–51. MR3390132 https://doi.org/10.1016/j.jspi.2014.06.005

[48] Szabó, B., van der Vaart, A.W. and van Zanten, J.H. (2015). Frequentist coverage of adaptive nonparametric Bayesian credible sets. *Ann. Statist.* **43** 1391–1428. MR3357861 https://doi.org/10.1214/14-AOS1270

[49] van der Vaart, A.W. (1998). *Asymptotic Statistics. Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge: Cambridge Univ. Press. MR1652247 https://doi.org/10.1017/CBO9780511802256

[50] Yang, Y., Bhattacharya, A. and Pati, D. Frequentist coverage and sup-norm convergence rate in Gaussian process regression. Available at arXiv:1708.04753.

[51] Yano, K. and Kato, K. (2020). Supplement to "On frequentist coverage errors of Bayesian credible sets in moderately high dimensions." https://doi.org/10.3150/19-BEJ1142SUPP.

[52] Yoo, W., Rousseau, J. and Rivoirard, V. Adaptive supremum norm posterior contraction: Wavelet spike-and-slab and anisotropic Besov spaces. Available at arXiv:1708.01909.

[53] Yoo, W.W. and Ghosal, S. (2016). Supremum norm posterior contraction and credible sets for nonparametric multivariate regression. *Ann. Statist.* **44** 1069–1102. MR3485954 https://doi.org/10.1214/15-AOS1398