NONPARAMETRIC BERNSTEIN-VON MISES
PHENOMENON: A TUNING PRIOR PERSPECTIVE

BY ZUOFENG SHANG* AND GUANG CHENG†

Department of Statistics, Purdue University
November 14, 2014

Statistical inference on infinite-dimensional parameters in Bayesian framework is investigated. The main contribution of this paper is to demonstrate that nonparametric Bernstein-von Mises theorem can be established in a very general class of nonparametric regression models under a novel tuning prior (indexed by a non-random hyper-parameter). Surprisingly, this type of prior connects two important classes of statistical methods: nonparametric Bayes and smoothing spline at a fundamental level. The intrinsic connection with smoothing spline greatly facilitates both theoretical analysis and practical implementation for nonparametric Bayesian inference. For example, we can apply smoothing spline theory in our Bayesian setting, and can also employ generalized cross validation to select a proper tuning prior, under which credible regions/intervals are frequentist valid. A collection of probabilistic tools such as Cameron-Martin theorem [7] and Gaussian correlation inequality [30] are employed in this paper.

1. Introduction. A common practice to quantify Bayesian uncertainty is to construct credible sets (or regions) that cover a large fraction of posterior mass. The frequentist validity of Bayesian credible sets is often called as Bernstein-von Mises (BvM) phenomenon. In nonparametric settings, BvM theorem was initially found not to hold by [13, 18]. An essential reason for this failure is due to the lack of flexibility in the assigned priors. In this paper, we introduce a novel tuning prior under which BvM theorem can be established in a very general class of nonparametric regression models. More interestingly, this type of prior connects two important classes of statistical methods: nonparametric Bayes and smoothing spline at a fundamental level.

*Visiting Assistant Professor
†Corresponding Author. Associate Professor, Research Sponsored by NSF CAREER Award DMS-1151692, DMS-1418042, Simons Foundation 305266. Guang Cheng is on sabbatical at Princeton while this work is finalized; he would like to thank the Princeton ORFE department for its hospitality and support.

AMS 2000 subject classifications: Primary 62C10 Secondary 62G15, 62G08
Keywords and phrases: Bayesian inference, Gaussian sequence model, nonparametric Bernstein-von Mises Theorem, smoothing spline
This coupling phenomenon is practically useful in the sense that a proper tuning prior can be selected through generalized cross validation.

We begin with Gaussian sequence models that can be used as a platform to investigate more complicated statistical problems; see [3, 35]. Recently, a series of (adaptive) credible sets have been proposed for their mean sequences in the literature; see [29, 26, 39, 40, 10]. In this paper, we consider a novel tuning prior with a (non-random) hyper-parameter that can be used to tune the assigned prior. Under tuning priors, we demonstrate that when the tuning parameter is properly selected, the frequentist coverage of an asymptotic credible set approaches one irrespective of the credibility level, called as weak BvM phenomenon. On the other hand, if the tuning parameter is poorly chosen, there always exist uncountably many true parameters that are excluded by the constructed credible set asymptotically. Hence, we observe a phase transition phenomenon, which is caused by the fact that the posterior standard deviation is always strictly greater or smaller than the deviation between the truth and the posterior mode. Also see Figure 1 for a graphical illustration. The issue of overly conservativeness of credible sets can be easily resolved by invoking a weaker topology (inspired by [10]). This leads to the so-called strong BvM phenomenon. In the end, we construct adaptive credible sets by substituting the unknown model regularity with a consistent estimate (with a mild convergence rate). With such a simple replacement, the adaptive sets can still preserve appealing features as those built upon the known regularity. In particular, their radii do not involve the annoying “blow-up” constant as in [40]; see [33] for more discussions.

The major contribution of this paper is to establish BvM theorem in a very general class of nonparametric regression models by borrowing the theoretical insights obtained from the simple Gaussian sequence models. For example, the assigned Gaussian process (GP) prior also involves a tuning hyper-parameter (this idea is rooted in [48]). As far as we know, our BvM theorem is the first nonparametric one that holds under total variation distance. Such a result is crucial in constructing credible regions for the regression function under both strong and weak topology, and credible intervals for its linear functionals, e.g., its local value. In particular, their constructions can be concretized by ODE techniques, and their frequentist behaviors are carefully investigated. In contrast, the BvM theorem developed by [11] is in terms of a weaker metric, i.e., bounded Lipschitz metric, and crucially relies on the efficiency of nonparametric estimate ([5]). Other related BvM theorems include [14, 15, 9, 4, 36, 12] for semiparametric models or functionals of the density. Our second contribution is to obtain posterior contraction rates under a type of Sobolev norm as an intermediate step in develop-
ing our BvM theorem. This Sobolev norm is stronger than the commonly used norms in deriving contraction rates such as Kullback-Leibler norm and supremum norm ([20, 21, 43, 44, 45, 46]). The uniformly consistent test arguments ([21]) together with Cameron-Martin theorem [7] and Gaussian correlation inequality [30] are applied to derive this new result.

Interestingly, a straightforward application of Hájek’s Lemma ([22]) illustrates that the posterior distribution under the assigned GP prior turns out to be closely related to the penalized likelihood in the smoothing spline literature. The intrinsic connection with smoothing spline greatly facilitates both theoretical analysis and practical implementation for nonparametric Bayesian inference. For example, we can apply smoothing spline theory in our Bayesian setting, and we can employ the generalized cross validation (GCV) method to select a proper tuning hyper-parameter in the assigned GP prior under which credible regions/intervals are frequentist valid. This novel methodology is crucially different from the empirical Bayes approach, and strongly supported by our simulation results.

The rest of this article is organized as follows. Section 2 focuses on Gaussian sequence models under two types of tuning priors, and conducts frequentist evaluation of the resulting (adaptive) credible sets. In Section 3, we present a general class of nonparametric regression models including Gaussian regression and logistic regression, and the assigned GP priors. The duality between nonparametric Bayes models and smoothing spline models are also discussed. Section 4 includes results on posterior contraction rates under a type of Sobolev norms. Section 5.1 includes the main result of the article: nonparametric BvM theorem. Sections 5.2 – 5.4 discuss two important applications of BvM theorem: construction of credible region of the regression function and credible interval of a general class of linear functionals. Their frequentist validity is also discussed. Section 6 includes simulation study. All the technical details are postponed to the online supplementary.

2. Gaussian Sequence Model. In this section, we consider Gaussian sequence models, and generalize the Gaussian priors in [18] by introducing tuning parameters. In this new Bayesian framework, we demonstrate that when the tuning parameter is properly chosen, the credible region possesses more satisfactory frequentist coverage properties, and can be further improved based on a weaker topology; also see [10]. A phase transition phenomenon of credible regions is also observed. A set of adaptive credible regions are constructed in the end. The theoretical insights obtained from this simple model will be naturally carried over to a very general class of nonparametric regression models considered in Section 3.
Consider the following Gaussian sequence model:

\begin{equation}
Y_i = \theta_i + \frac{1}{\sqrt{n}}\epsilon_i, \quad i = 1, 2, \ldots,
\end{equation}

where \( \theta = \{\theta_i\}_{i=1}^{\infty} \in \ell_2(q_0) \equiv \{\{\theta_i\}_{i=1}^{\infty} : \sum_{i=1}^{\infty} \theta_i^2 < q_0\} \) for some constant \( q_0 \geq 0 \). In particular, denote \( \ell_2 = \ell_2(0) \) as the space of real square-summable sequences. Here, \( \epsilon_i \)’s are iid standard Gaussian random variables. Suppose \( \theta_0 = \{\theta_i^0\}_{i=1}^{\infty} \in \ell_2(q_0) \) as the “true” parameter generating \( Y_i \)’s.

2.1. Tuning Prior I. We assign the following Gaussian priors on the mean sequence:

\textbf{P1.} \( \theta_i \sim N(0, \tau_i^{-2}) \), where \( \tau_i^2 = n\lambda_i^{2p}, \lambda = n^{-\alpha_0}, \alpha_0 > 0 \) and \( p \) satisfies \( p > q_0 + 1/2 \geq 1/2 \), and call them as tuning priors indexed by \( \lambda \). This leads to a more flexible Bayesian framework. In particular, when \( n\lambda = 1 \) (equivalently, \( \alpha_0 = 1 \) and \( q_0 = 0 \), the above setup reduces to the one considered in [18].

It is easy to see that the posterior for \( \theta_i \) is also Gaussian:

\[ \theta_i | Y_i \sim N\left( \hat{\theta}_{ni}, v_{ni}^2 \right), \quad i = 1, 2, \ldots, \]

where \( \hat{\theta}_{ni} = Y_i/(1 + \lambda_i^{2p}) \) and \( v_{ni}^2 = \{n(1 + \lambda_i^{2p})\}^{-1} \). Thus, given the data, we can rewrite \( \theta_i \) as \( \theta_i = \hat{\theta}_{ni} + v_{ni}\eta_i \), where \( \eta_1, \eta_2, \ldots \sim iid N(0, 1) \). Denote \( \hat{\theta} \) as the posterior mode \( \{\hat{\theta}_{ni}\}_{i=1}^{\infty} \), and let \( h = \lambda^{-1/(2p)} \). Similar to Theorem 1 of [18], it can be shown that \( n\sqrt{h}\|\theta - \hat{\theta}\|_2^2 = h^{-1/2}E_n + D_n^{1/2}Z_n \), where

\begin{equation}
D_n = \sum_{i=1}^{\infty} \frac{2h}{(1 + \lambda_i^{2p})^2}, \quad E_n = \sum_{i=1}^{\infty} \frac{h}{1 + \lambda_i^{2p}}, \quad \text{and} \quad Z_n = D_n^{-1/2} \sum_{i=1}^{\infty} \frac{\sqrt{h}}{1 + \lambda_i^{2p}}(\eta_i^2 - 1).
\end{equation}

Since \( Z_n \overset{d}{\to} N(0, 1) \), we have the following \((1 - \alpha)\) asymptotic credible region

\begin{equation}
C_n(\alpha) = \left\{ \theta = \{\theta_i\}_{i=1}^{\infty} \in \ell_2(q_0) : \|\theta - \hat{\theta}\|_2 \leq \sqrt{\frac{E_n + \sqrt{D_n}z_{\alpha}}{nh}} \right\},
\end{equation}

where \( z_{\alpha} = \Phi^{-1}(1 - \alpha) \) and \( \Phi \) is the c.d.f. of \( N(0, 1) \). It is easy to see that \( D_n \approx 2 \int_{0}^{\infty} \frac{1}{(1 + x^{2p})^2} dx \) and \( E_n \approx \int_{0}^{\infty} \frac{1}{1 + x^{2p}} dx \), where \( \alpha_n \approx \beta_n \) means \( \lim_{n \to \infty} a_n/b_n = 1 \). Hence, the radius of \( C_n(\alpha) \) contracts at the rate \( n^{-2p-\alpha_0} \).
The frequentist coverage probabilities of $C_n(\alpha)$ are summarized in the following Theorem 2.1, where a phase transition phenomenon is observed. Also see Figure 1 for a graphical illustration.

**Theorem 2.1.** Under model (2.1) and Prior $P_1$, we have

1. If $\alpha_0 \geq \frac{2p}{2q_0 + 1}$, then for any $\theta_0 \in \ell_2(q_0)$, with probability approaching one, $\theta_0 \in C_n(\alpha)$.

2. If $0 < \alpha_0 < \frac{2p}{2q_0 + 1}$, then there exist uncountably many $\theta_0 \in \ell_2(q_0)$ such that for each such $\theta_0$, with probability approaching one, $\theta_0 \notin C_n(\alpha)$.

![Figure 1](image-url)  

**Fig 1.** A phase transition phenomenon in Gaussian sequence model. $\hat{\theta}$ represents the posterior mode used as the center of the credible set (2.3). Tuning prior is indexed by $\alpha_0$. The left circle illustrates a credible set that covers the truth $\theta_0$ given a properly selected tuning parameter. The right circle represents a negative case when the tuning parameter is poorly selected. Specifically, when $\alpha_0$ decreases across the boundary $2p/(2q_0 + 1)$, the radius (in terms of $\ell^2$-norm) shrinks at a faster rate than the rate $\hat{\theta}$ approaches $\theta_0$.

Obviously, Part (1) of Theorem 2.1 discovers a valid approach in constructing credible sets by properly tuning $\lambda$. On the other hand, Part (2) of Theorem 2.1 generalizes the negative results of [18] where $\alpha_0 = 1$ and $q_0 = 0$. We further notice that the fastest contraction rate of a valid credible set, i.e., $n^{-q_0/(2q_0 + 1)}$, turns out be minimax optimal according to [10] when $\lambda \asymp n^{-2p/(2q_0 + 1)}$. It can even be shown that the posterior mode satisfies $||\hat{\theta} - \theta_0||_2 = O_P(n^{-q_0/(2q_0 + 1)})$ given the same choice of $\lambda$, where the rate
was shown by [49] to be optimal in the case $q_0 > 1/2$. Note that, without using the tuning parameter approach, one cannot get this optimal rate of convergence for $p > q_0 + 1/2$; see [49]. Another interesting observation is the phase transition of the coverage property at $\alpha_0 = 2p/(2q_0 + 1)$. We note that in more general models as considered in Section 3, calculations are so complex that phase transition is hard to detect, although we conjecture its existence. A comprehensive study on this is beyond the scope of this paper.

**Remark 2.1.** Theorem 2.1 revisits Theorem 4.2 in [25] from a different perspective. Specifically, our assumption $p > q_0 + 1/2$ implies that $2p > q_0$. Under $2p > q_0$, Theorem 4.2 of [25] shows that the probability that the credible region covers the truth tends to either zero (along a sequence of truths) or one. Actually, [25] further show that if $2p \leq q_0$ and $\alpha_0 = 2p/(2q_0 + 1)$, then there exists a sequence of truths along which the coverage probability tends to any level in $[0, 1)$. Here we restrict to the case $p > q_0 + 1/2$ to ease the technical arguments. Indeed, our results later in this section benefit from this restriction. Although our Theorem 2.1 can be viewed as a special case of [25], our new tuning prior formulation would greatly benefit the theoretical studies on the nonparametric regression models; see Section 3.

Even under the flexible tuning priors, Theorem 2.1 still cannot exactly match the credibility level with the frequentist coverage. The main reason is that the $L^2$-topology, upon which $C_n(\alpha)$ is built, is too strong to support the frequentist-matching property. This motivates us to design a weaker norm as follows; also see [10]. For any $\theta = \{\theta_i\}_{i=1}^{\infty} \in \ell_2$, define $\|\theta\|_1^\# = \sum_{i=1}^{\infty} d_i \theta_i^2$, for a given positive sequence $d_i$ satisfying $d_i \asymp 1/(\log i)^{\tau}$ for some $\tau > 1$. Clearly, $\|\cdot\|_1^\#$ defines a norm weaker than the usual $\ell_2$-norm. Define

\begin{equation}
C_n^\#(\alpha) = \left\{ \theta = \{\theta_i\}_{i=1}^{\infty} \in \ell_2(q_0) : \|\theta - \hat{\theta}\|_1^\# \leq \sqrt{c_\alpha/n} \right\},
\end{equation}

where $c_\alpha$ is the upper $(1 - \alpha)$-quantile of the distribution of the random variable $\sum_{i=1}^{\infty} d_i \eta_i^2$, that is, $P(\sum_{i=1}^{\infty} d_i \eta_i^2 \leq c_\alpha) = 1 - \alpha$. The value of $c_\alpha$ can be easily simulated, and not sensitive to the choice of $\tau$ (we select $\tau = 2$ in our simulations). It is easy to see that $P(\theta \in C_n^\#(\alpha) | \{Y_i\}_{i=1}^{\infty}) = P(n \sum_{i=1}^{\infty} d_i v_i^2 \eta_i^2 \leq c_\alpha) = P(\sum_{i=1}^{\infty} d_i v_i^2 \eta_i^2 \leq c_\alpha) = 1 - \alpha$, where the final limit follows by the fact that $\sum_{i=1}^{\infty} d_i v_i^2 \eta_i^2$ converges in distribution to $\sum_{i=1}^{\infty} d_i \eta_i^2$ as $n \to \infty$ (by moment generating function approach). That is, $C_n^\#(\alpha)$ asymptotically possesses $(1 - \alpha)$ posterior coverage.

Frequentist coverage properties of $C_n^\#(\alpha)$ are summarized below.
THEOREM 2.2. Under model (2.1) and Prior P1, we have:

(1). If \(\alpha_0 \geq 2p/(2q_0 + 1)\), then for any \(\theta_0 \in \ell_2(q_0)\), with probability approaching \(1 - \alpha\), \(\theta_0 \in C_n^\dagger(\alpha)\).

(2). If \(0 < \alpha_0 < 2p/(2q_0 + 1)\), then there exist uncountably many \(\theta_0 \in \ell_2(q_0)\) such that for each such \(\theta_0\), with probability approaching one, \(\theta_0 \notin C_n^\dagger(\alpha)\).

Given the same choice of \(\alpha_0\) as in Theorem 2.1, Part (1) of Theorem 2.2 guarantees the frequentist-matching property of \(C_n^\dagger(\alpha)\) under a weaker norm. We also remark that \(C_n^\dagger(\alpha)\) only relies on \((\hat{\theta}, c_\alpha)\) that are both easy to simulate; see Section 6. In contrast, the credible region constructed by [10] may require calculation of certain nontrivial quantities; see (14) therein.

In the end, we consider a set of adaptive credible sets that do not rely on the unknown regularity \(q_0\); see Section S.3 for more details. The main idea is to replace \(q_0\) by a proper estimator \(\hat{q}\) satisfying \(\hat{q} - q_0 = o_P(1/\log n)\); see Proposition S.1 and Remark S.1 for the construction of \(\hat{q}\). The \((1 - \alpha)\) adaptive credible sets given in (S.2) and (S.3) are constructed by replacing \((p, \lambda, h)\) in (2.3) and (2.4) by \(p = \hat{q} + 1, \lambda = n^{-2p)/(2\hat{q}+1)\) and \(h = n^{-1/(2\hat{q}+1)}\); see Theorem S.1 for its theoretical justification. Note that the radii of (S.2) and (S.3) do not involve any unknown constants. Hence, we claim that (S.2) and (S.3) are more practically feasible to construct than the region proposed in [40] with a “blow-up” constant; see [33] for more discussions.

2.2. Tuning Prior II. In this section, we assign a more subtle Gaussian prior P2 that includes an additional parameter \(q\), and derive parallel results to Section 2.1. A similar type of prior will be placed on the Fourier coefficients of the nonparametric regression functions considered in Section 3.

The new tuning prior (motivated from [48]) is of the following form:

**P2.** \(\theta_i \sim N(0, \tau_{si}^{-2})\), where \(\tau_{si}^2 = \epsilon^2 + n\lambda_1^2 q, \lambda = n^{-\alpha_0}, \alpha_0 > 0\) and \(p, q\) satisfy \(1/2 < q \leq q_0 < 2q\) and \(q + 1/2 < p < 2q + 1/2\).

By direct calculations, the posterior distribution of \(\theta_i\) is

\[
\theta_i | Y_i \sim N(\hat{\theta}_{si}, v_{si}^2), \quad i = 1, 2, \ldots,
\]

where \(\hat{\theta}_{si} = Y_i/(1 + \lambda_1^2 q + n^{-1}i^2 p)\) and \(v_{si}^2 = (n(1 + \lambda_1^2 q + n^{-1}i^2 p))^{-1}\). Denote \(\tilde{\theta}_s = \{\hat{\theta}_{si}\}_{i=1}^\infty\) as the posterior mode of \(\theta\), and let \(\alpha_1 = \min\{1/(2p), \alpha_0/(2q)\}\).

By similar analysis as in Section 2.1, we have

\[
n^{1-\alpha_1} ||\theta - \tilde{\theta}_s||_2^2 = E_{sn} + n^{-\alpha_1/2}D_{sn}^{1/2}Z_{sn},
\]
where
\[
D_\infty = \sum_{i=1}^{\infty} \frac{2n^{-\alpha_1}}{(1 + \lambda i^{2q} + n^{-1/2p})^2},
\]
(2.5)
\[
E_\infty = \sum_{i=1}^{\infty} \frac{n^{-\alpha_1}}{1 + \lambda i^{2q} + n^{-1/2p}},
\]
(2.6)
\[
Z_\infty = D_\infty^{-1/2} \sum_{i=1}^{\infty} \frac{n^{-\alpha_1/2} (\eta_i^2 - 1)}{1 + \lambda i^{2q} + n^{-1/2p}}
\]
(2.7)

and \(\eta_i \sim iid N(0, 1)\) independent of \(Y_i\). Considering Lemma S.2 that \(Z_\infty \overset{d}{\to} N(0, 1)\), we construct an asymptotic \((1 - \alpha)\) credible region
\[
C_\alpha^n = \left\{ \theta = \{\theta_i\}_{i=1}^{\infty} \in \ell_2(q_0) : \|\theta - \hat{\theta}_n\|_2 \leq \sqrt{\frac{E_\infty + \sqrt{D_\infty n^{-\alpha_1} z_\alpha}}{n^{1-\alpha_1}}} \right\},
\]
(2.8)
where \(E_\infty\) and \(D_\infty\) converge to some constants specified in Lemma S.1. For example, when \(p = 3, q = 2, \alpha_0 = 2/3\), we have \(D_\infty \approx 1.467\) and \(E_\infty \approx 0.919\). The radius of \(C_\alpha^n\) decreases at rate \(n^{-1/2}\).

Our next result demonstrates that the frequentist behavior of the credible set \(C_\alpha^n\) at any level \(\alpha \in (0, 1)\) depends on the relation between the property of the true parameter, i.e., \(q_0\), and the prior specifications, i.e., \((p, q, \alpha_0)\). Again, to obtain a frequentist-matching credible region, we need to employ a weaker norm than \(\ell_2\)-norm as follows:
\[
C^\dagger_\alpha^n = \left\{ \theta = \{\theta_i\}_{i=1}^{\infty} \in \ell_2(q) : \|\theta - \hat{\theta}_n\|_1 \leq \sqrt{c_\alpha/n} \right\},
\]
(2.9)

and assume an additional technical condition \(4q > 2q_0 + 1\).

**Theorem 2.3.** Suppose model (2.1) holds with Prior \(P2\).

(1). For \(C_\alpha^n\), the following results hold:

(a) If \(p - 1/2 \leq q_0 < 2q\) and \(\alpha_0 \geq 2q/(2q_0 + 1)\), then for any \(\theta_0 \in \ell_2(q_0)\), with probability approaching one, \(\theta_0 \in C_\alpha^n\).

(b) If \(q \leq q_0 < p - 1/2\) or \(0 < \alpha_0 < 2q/(2q_0 + 1)\), then there are uncountably many \(\theta_0 \in \ell_2(q_0)\) such that for each such \(\theta_0\), with probability approaching one, \(\theta_0 \notin C_\alpha^n\).

(2). Furthermore, suppose \(4q > 2q_0 + 1\). For \(C^\dagger_\alpha^n\), the following results hold:
(a) If \( p - 1/2 \leq q_0 < 2q \) and \( \alpha_0 \geq 2q/(2q_0 + 1) \), then for any \( \theta_0 \in \ell_2(q_0) \), with probability approaching \( 1 - \alpha \), \( \theta_0 \in C^\dagger_{e_n}(\alpha) \).

(b) If \( q \leq q_0 < p - 1/2 \) or \( 0 < \alpha_0 < 2q/(2q_0 + 1) \), then there are uncountably many \( \theta_0 \in \ell_2(q_0) \) such that for each such \( \theta_0 \), with probability approaching one, \( \theta_0 / \in C^\dagger_{e_n}(\alpha) \).

Besides frequentist validity, another important consequence of Theorem 2.3 is that \( C^\dagger_{e_n}(\alpha) \) achieves the optimal contraction rate in the sense of [25] when \( \lambda \) is carefully chosen. Specifically, \( C^\dagger_{e_n}(\alpha) \) possesses the contraction rate \( n^{-q_0/(2q_0 + 1)} \) when \( p - 1/2 \leq q_0 < 2q \) and \( \alpha_0 = 2q/(2q_0 + 1) \). Moreover, to yield a valid region, we only need the weakest possible condition on the true regularity, i.e., \( q_0 \geq p - 1/2 \). Adaptive credible regions under P2 are constructed in Section S.3 of the supplement document.

3. Nonparametric Regression Model. In this section, we present a general class of nonparametric regression models to which nonparametric Bayesian method is applied. Specifically, we place a type of prior P2 on the Fourier coefficients of the unknown regression function, and further prove that the resulting posterior distribution turns out to be closely related to the penalized likelihood in the smoothing spline literature; see (3.8). This is the first main result of this paper.

Let \( p_f(y|x) \) be the conditional probability density function of \( Y \) given \( X = x \) under an infinite dimensional parameter \( f \):

\[
Y|f, X = x \sim p_f(y|x),
\]

where \((y, x) \in \mathcal{Y} \times \mathbb{I} \) and \( \mathbb{I} := [0, 1] \). We assume that there exists a “true” parameter \( f_0 \) under which the sample is drawn, and that \( f_0 \) belongs to an \( m \)-th order Sobolev space:

\[
H^m(\mathbb{I}) = \{ f \in L^2(\mathbb{I}) | f^{(j)} \text{are absolutely continuous for } j = 0, 1, \ldots, m - 1, \text{ and } f^{(m)} \in L^2(\mathbb{I}) \}.
\]

Throughout the paper, we let \( m > 1/2 \) such that \( H^m(\mathbb{I}) \) is a reproducing kernel Hilbert space (RKHS).

We impose some regularity assumptions on \( \ell(y; f(x)) \equiv \log(p_f(y|x)) \). Denote the first-, second- and third-order derivatives of \( \ell(y; a) \) w.r.t. \( a \) by \( \hat{\ell}_a(y; a) \), \( \hat{\ell}_a(y; a) \), and \( \hat{\ell}_a(y; a) \), respectively.

**Assumption A1.** (a) \( \ell(y; a) \) is three times continuously differentiable and strictly concave w.r.t. \( a \). There exists positive constants \( C_0 \) and \( C_1 \)
\(s.t.
\)

\[
E_{f_0} \left\{ \exp(\sup_{a \in \mathbb{R}} |\tilde{\ell}_a(Y; a)|/C_0) \right\} \leq C_1, \text{ a.s.,}
\]

\[
E_{f_0} \left\{ \exp(\sup_{a \in \mathbb{R}} |\dot{\ell}_a(Y; a)|/C_0) \right\} \leq C_1, \text{ a.s.}
\]

(b) There exists a constant \(C_2 > 0\) such that

\[
C_2^{-1} \leq B(X) = -E_{f_0}\{\ell_a(Y; f_0(X))|X\} \leq C_2, \text{ a.s.}
\]

(c) \(\epsilon \equiv \ell_a(Y; f_0(X))\) satisfies \(E_{f_0}\{e|X\} = 0, E_{f_0}\{e^2|X\} = B(X), \text{ a.s.}\), and there exists a constant \(M\) such that \(E_{f_0}\{e^4|X\} \leq M, \text{ a.s.}\).

We present three common examples that satisfy Assumption A1.

**Example 3.1.** (Gaussian Model) In Gaussian regression, \(Y = f(X) + \epsilon\) with \(\epsilon|X \sim N(0, 1)\). It is easy to see that \(p_f(y|x) = (\sqrt{2\pi})^{-1} \exp(-(y - f(x))^2/2)\), and hence \(\ell(y; a) = -\log \sqrt{2\pi} - (y - a)^2/2\). We can easily verify Assumption A1 with \(B(X) = 1\).

**Example 3.2.** (Logistic Model) In logistic regression, where \(P_f(Y = 1|X) = 1 - P_f(Y = 0|X) = \exp(f(X))/(1+\exp(f(X)))\), it is easy to see that \(\ell(y; a) = ya - \log(1 + \exp(a))\) and \(B(X) = \exp(f_0(X))/(1 + \exp(f_0(X)))^2\). Assumption A1 (a) follows from the boundedness of \(\ell_a\) and \(\tilde{\ell}_a\). Assumption A1 (b) follows from the fact that \(B(X)\) is almost surely bounded away from zero and infinity since \(f_0\) is supported in a compact interval. Assumption A1 (c) can be verified by direct calculation.

**Example 3.3.** (Exponential Family) Suppose \((Y, X)\) follows one-parameter exponential family \(Y|f, X \sim \exp \{Yf(X) + A(y) - G(f(X))\}\), where \(A(-)\) and \(G(-)\) are known. It is easy to see that \(\ell(y; a) = ya + A(y) - G(a)\), and hence, \(\tilde{\ell}_a(y; a) = y - \hat{G}(a), \dot{\ell}_a(y; a) = -\hat{G}(a)\) and \(\ddot{\ell}_a(y; a) = -\hat{G}(a)\). We assume that \(G\) has bounded \(j\)th-order derivatives on \(\mathbb{R}\) for \(j = 2, 3, 4,\) and for any real number \(a\) belonging to the range of \(f_0\), \(\hat{G}(a) \geq \delta\) for some constant \(\delta > 0\). Clearly, both \(\tilde{\ell}_a\) and \(\dot{\ell}_a\) are bounded, and hence Assumption A1 (a) holds. Furthermore, \(B(X) = \hat{G}(f_0(X))\) satisfies Assumption A1 (b). Since \(\epsilon = Y - \hat{G}(f_0(X))\), it is easy to see that \(E_{f_0}\{e|X\} = E_{f_0}\{Y|X\} - \hat{G}(f_0(X)) = 0\) and \(E_{f_0}\{e^2|X\} = Var_{f_0}\{Y|X\} = \hat{G}(f_0(X)) = B(X)\) (see [31]). Furthermore, \(E_{f_0}\{e^4|X\} = G^{(4)}(f_0(X)) + 4\hat{G}(f_0(X))\hat{G}(f_0(X)) + 6\hat{G}(f_0(X))\hat{G}(f_0(X))^2 + 3\hat{G}(f_0(X))^2 + \hat{G}(f_0(X))^4\) which is almost surely bounded by a constant. Therefore, Assumption A1 (c) holds.
Consider the smoothing spline estimate \( \hat{f}_{n,\lambda} \):\
\[
\hat{f}_{n,\lambda} = \arg \max_{f \in H^m(\mathbb{I})} \ell_{n,\lambda}(f)
\]
(3.4)\
\[
(3.4) = \arg \max_{f \in H^m(\mathbb{I})} \left\{ \frac{1}{n} \sum_{i=1}^{n} [\ell(Y_i; f(X_i)) + \log \pi(X_i)] - (\lambda/2)J(f, f) \right\},
\]
where \( J(f, \tilde{f}) = \int_{\mathbb{I}} f^{(m)}(x)\tilde{f}^{(m)}(x)dx \) for any \( f, \tilde{f} \in H^m(\mathbb{I}) \), \( \pi(x) \) denotes the marginal density of \( X \) satisfying \( 0 < \inf_{x \in \mathbb{I}} \pi(x) \leq \sup_{x \in \mathbb{I}} \pi(x) < \infty \), and \( \lambda \to 0 \) as \( n \to \infty \). Define \( V(f, \tilde{f}) = E_{f_0} \{B(X)f(X)\tilde{f}(X)\} \) and \( \langle f, \tilde{f} \rangle = V(f, \tilde{f}) + \lambda J(f, \tilde{f}) \). It follows from [38] that \( \langle \cdot, \cdot \rangle \) well defines an inner product in \( H^m(\mathbb{I}) \). Let \( \| \cdot \| \) be the corresponding norm, and \( K \) be the corresponding reproducing kernel function. Let \( W_\lambda \) be a self-adjoint bounded linear operator from \( H^m(\mathbb{I}) \) to itself satisfying \( \langle W_\lambda f, \tilde{f} \rangle = \lambda J(f, \tilde{f}) \). For simplicity, we also denote \( V(f) = V(f, f) \) and \( J(f) = J(f, f) \).

**Assumption A2.** There exists a sequence of eigenfunctions \( \varphi_\nu \in H^m(\mathbb{I}) \) satisfying \( \sup_{\nu \in \mathbb{N}} \| \varphi_\nu \|_{\sup} < \infty \), and a nondecreasing sequence of eigenvalues \( \rho_\nu \approx \nu^{2m} \) such that

\[
(3.5) \quad V(\varphi_\mu, \varphi_\nu) = \delta_{\mu\nu}, \quad J(\varphi_\mu, \varphi_\nu) = \rho_\mu \delta_{\mu\nu}, \quad \mu, \nu \in \mathbb{N},
\]
where \( \delta_{\mu\nu} \) is the Kronecker’s delta. In particular, any \( f \in H^m(\mathbb{I}) \) admits a Fourier expansion \( f = \sum_\nu V(f, \varphi_\nu)\varphi_\nu \) with convergence in the \( \| \cdot \| \)-norm. In particular, the first \( m \) eigenvalues are all zero, and \( \rho_\nu > 0 \) for \( \mu > m \).

We refer to [38, Proposition 2.2] about the validity of Assumption A2. In particular, Assumption A2 yields an equivalent series representation of \( \| f \|^2 \), \( K \) and \( W_\lambda \) as follows.

**Proposition 3.1.** For any \( f \in H^m(\mathbb{I}) \) and \( z \in \mathbb{I} \), we have \( \| f \|^2 = \sum_\nu |V(f, \varphi_\nu)|^2(1+\lambda \rho_\nu), \) \( K_z(\cdot) = \sum_\nu \frac{\varphi_\nu(z)}{1+\lambda \rho_\nu} \varphi_\nu(\cdot), \) and \( W_\lambda \varphi_\nu(\cdot) = \frac{\lambda \rho_\nu}{1+\lambda \rho_\nu} \varphi_\nu(\cdot) \) under Assumption A2.

We next assign a Gaussian process (GP) prior to \( f \), and point out its relation with prior \( \mathbf{P}2 \). Let \( \{v_\nu\}_{\nu=1}^{\infty} \) be a sequence of independent random variables satisfying \( v_\nu \sim \pi_\nu \) for \( \nu = 1, \ldots, m \), where \( \pi_1, \ldots, \pi_m \) are some probability distribution functions, and \( v_\nu \sim N(0, \rho_\nu^{-(1+\beta/(2m)}) \) for \( \nu > m \). Here, the prior parameter \( \beta \) is used to measure the regularity difference between the prior and the parameter space \( H^m(\mathbb{I}) \). The random variables
\( v_\nu \)'s are assumed to be independent of the samples. The Gaussian process (GP) prior on \( f \) is given as follows:

\[
G_\lambda(t) = \sum_{\nu=1}^{\infty} w_\nu \varphi_\nu(t),
\]

where \( w_\nu = v_\nu \) for \( \nu = 1, \ldots, m \), and \( w_\nu = (1 + n\lambda\rho_\nu^{-\beta/(2m)})^{-1/2}v_\nu \) for \( \nu > m \). For simplicity, we assume that \( \pi_j \) is \( N(0, \sigma_j^2) \) for \( j = 1, \ldots, m \) in the above GP prior, where \( \sigma_j^2 \)'s are fixed constants. Our results may also apply to a general class of \( \pi_1, \ldots, \pi_m \). When \( \nu > m \), it is easy to see that \( w_\nu \sim N(0, (\rho_\nu^{1+\beta/(2m)} + n\lambda\rho_\nu)^{-1}) \), and the inverse prior variance satisfies \( \rho_\nu^{1+\beta/(2m)} + n\lambda\rho_\nu \asymp \nu^{2m+\beta} + n\lambda\nu^{2m} \). This is exactly a type of Prior P2. Clearly, the sample path of \( G_\lambda \) belongs to \( H_m(\mathbb{I}) \) for any \( \beta \in (1, 2m + 1) \) a.s. due to the trivial observation \( E\{J(G_\lambda, G_\lambda)\} = \sum_{\nu>m} \rho_\nu/(\rho_\nu^{1+\beta/(2m)} + n\lambda\rho_\nu) < \infty \). Note that \( \beta \) and \( m \) jointly determine the smoothness of RKHS induced by \( G_\lambda \); see [44]. Let \( \Pi_\lambda \) be the probability measure induced by \( G_\lambda \), i.e., \( \Pi_\lambda(B) = P(G_\lambda \in B) \).

The posterior distribution of \( f \) can be written as

\[
P(f|D_n) \propto \exp\left(\sum_{i=1}^{n} \ell(Y_i; f(X_i))\right)d\Pi_\lambda(f),
\]

where \( D_n \equiv \{Z_1, \ldots, Z_n\} \) denotes a full sample set, and \( Z_i = (Y_i, X_i) \), \( i = 1, \ldots, n \) are iid copies of \( Z = (Y, X) \). By a key Lemma 3.2 below, we can link the above posterior distribution with the penalized likelihood \( \ell_{n,\lambda} \)

\[
P(f \in S|D_n) = \frac{\int_S \exp(n\ell_{n,\lambda}(f))d\Pi(f)}{\int_{H^m(\mathbb{I})} \exp(n\ell_{n,\lambda}(f))d\Pi(f)},
\]

for any \( \Pi \)-measurable subset \( S \subseteq H^m(\mathbb{I}) \), where \( \Pi \) is the probability measures induced by the following GP \( G \):

\[
G(t) = \sum_{\nu=1}^{\infty} v_\nu \varphi_\nu(t).
\]

Similarly, we can check that the sample path of \( G \) also belongs to \( H^m(\mathbb{I}) \) for any \( \beta \in (1, 2m + 1) \) a.s.. We remark that the important duality between (3.7) and (3.8) holds universally irrespective of the model assumptions A1.

The following lemma is established based on Hájek [22], and demonstrates the relationship between \( \Pi \) and \( \Pi_\lambda \).
LEMMA 3.2. With \( f \in H^m(I) \), we have the following Radon-Nikodym derivative of \( \Pi_\lambda \) with respect to \( \Pi \): 

\[
\frac{d\Pi_\lambda}{d\Pi}(f) = \prod_{\nu=m+1}^{\infty} \left( 1 + n\lambda \rho_\nu^{-\beta/(2m)} \right)^{1/2} \cdot \exp \left( -\frac{n\lambda}{2} J(f,f) \right).
\]

4. Rate of Posterior Contraction under Sobolev Norm. In this section, we derive posterior contraction rates under a type of Sobolev norm given the GP prior (3.6). The Sobolev type norm is stronger than the commonly used norms such as Kullback-Leibler norm and supremum norm. Our result in this section is an intermediate step for developing the general non-parametric BvM theorem, and also of independent interest.

We first assume the following posterior consistency condition.

Assumption A3. For any \( b > 0 \), as \( n \to \infty \), \( P(\|f - f_0\|_{\text{sup}} > b | D_n) \to 0 \) in \( P_{f_0} \)-probability.

A similar assumption was also used in [19] for deriving posterior contraction rates in terms of supremum norm. We point out that this assumption can be removed in two situations: (1) exponential family models such as Gaussian regression model; (2) logistic regression with null space being removed.

We first employ the uniformly consistent test arguments ([21]) to obtain an initial rate of posterior contraction. The Cameron-Martin theorem ([7]) together with a Gaussian correlation inequality ([30]) are also used in the proof of Proposition 4.1. With a bit abuse of notation, we define \( h = \lambda^{1/(2m)} \).

Proposition 4.1. (An initial rate of posterior contraction) Suppose Assumptions A1–A3 hold, and \( f_0(\cdot) = \sum_{\nu=1}^{\infty} f_0^\nu \varphi_\nu(\cdot) \) satisfies Condition (S): \( \sum_{\nu=1}^{\infty} |f_0^\nu|^2 \rho_\nu^{1+\frac{\beta}{2m}} < \infty \). Then for any \( \delta_n \to 0 \) satisfying \( \delta_n \geq r_n := (nh)^{-1/2} + \lambda^{1/2} \) and \( \delta_n = o(h^{1/2}) \), we have \( P(f \in H^m(I) : \|f - f_0\| > M\delta_n | D_n) \to 0 \) in \( P_{f_0} \)-probability for sufficiently large \( M \).

We comment on the implications of Condition (S). By Assumption A2, i.e., \( \rho_\nu \propto \nu^{2m} \), we have \( \sum_{\nu=1}^{\infty} |f_0^\nu|^2 \nu^{2m+\beta-1} < \infty \). This indicates

\[
\{f_0^\nu\} \in \ell_2(m + (\beta - 1)/2).
\]

Meanwhile, the GP prior (3.6) implies \( f_\nu \sim N(0, (\nu^{2m+\beta} + n\lambda \nu^{2m}))^{-1} \) for \( \nu > m \). Then, using the notation in Section 2.2, we have \( p = m + \beta/2, q = m, q_0 = p - 1/2 \). In particular, \( q_0 = p - 1/2 \) satisfies the weakest possible condition on the regularity of \( f_0 \) for obtaining a valid credible set (see Parts (1a) and (2a) of Theorem 2.3).
The initial contraction rate in Theorem 4.1 is sub-optimal, but helps us obtain the optimal one in Theorem 4.2 below. Define the optimal rate $\tilde{r}_n = (nh)^{-1/2} + h^{m + \frac{\beta - 1}{2}}$. This rate optimality follows from the smoothness of the true function in Condition $(S)$: $f_0 \in H^{m+(\beta-1)/2}(I)$. Let $a_n$ be defined in Lemma S.4 of online supplementary.

**Theorem 4.2.** (An optimal rate of posterior contraction) Suppose Assumptions A1–A3 hold, and $f_0 = \sum_{\nu=1}^{\infty} f_{0\nu} \varphi_{\nu}$ satisfies Condition $(S)$. Furthermore, the following set of rate conditions hold (Condition $(R)$):

(i). $h = o(1)$, $\log h^{-1} = O(\log n)$ and $nh^2 \to \infty$

(ii). $r_n = o(h^{1/2})$

(iii). $a_n \log n = o(\tilde{r}_n)$

(iv). $b_{n1} := n^{-1/2} \tilde{r}_n h^{-\frac{8m-1}{4m}} (\log \log n)^{1/2} \log n + n^{-1/2} \tilde{r}_n \log n = o(1)$

(v). $b_{n2} := n^{-1/2} h^{-2} \log n (\log \log n)^{1/2} = o(1)$.

(vi). $b_{n3} := n^{-3/2} h^{-\frac{8m-1}{4m}} (\log n)(\log \log n)^{1/2} + r_n^3 h^{-1/2} \log n = o(\tilde{r}_n^2)$

(vii). $b_{n4} := n^{-1/2} h^{-\frac{8m-1}{4m}} r_n^2 (\log n)(\log \log n)^{1/2} = o(\tilde{r}_n^2)$.

Then for any $\varepsilon, \delta > 0$, there exist positive constants $M = M(\varepsilon, \delta)$ and $N = N(\varepsilon, \delta)$ such that for any $n \geq N$, with $P_{f_0}$-probability less than $\varepsilon$,

$$P(f : \|f - f_0\| \geq M\tilde{r}_n|D_n) > \delta.$$  

Except for the probabilistic tools used in the proof of Proposition 4.1, we also apply the functional Bahadur representation and a set of empirical processes tools recently developed by [38] in the proof of Theorem 4.2.

**Remark 4.1.** It can be shown that, when $m > 3/2$ and $1 < \beta < m + \frac{1}{2}$, Condition $(R)$ holds for $h \asymp h_s^* \equiv n^{-\frac{1}{m+\beta}}$. This leads to the contraction rate $n^{-\frac{2m+\beta-1}{2(2m+\beta)}}$. According to Condition $(S)$, we know that $f_0$ belongs to the Sobolev space of higher order $H^{m+(\beta-1)/2}(I)$. Hence, the above contraction rate turns out to be optimal according to [45].

**5. Nonparametric Bernstein-von Mises Theorem.** In this section, we show that BvM theorem holds in the general nonparametric regression setup established in Section 3. To the best of our knowledge, our BvM theorem is the first nonparametric one based on total variation distance. Such a result is crucial in constructing credible regions under both strong and weak topology, and credible intervals for linear functionals, as illustrated in Sections 5.2–5.4. In contrast, the BvM theorems established by [10, 11] are in terms of a weaker metric: bounded Lipschitz metric. Similar as Section 2, the
tuning hyper-parameter $\lambda$ in the GP prior $G_\lambda$ controls the frequentist performance of a credible region/interval. In the end of Section 5.1, we discuss how to employ generalized cross validation (GCV) to select a proper tuning prior, i.e., $\lambda$, by taking advantage of the duality between nonparametric Bayesian and smoothing spline models.

5.1. **Main Theorem.** In this section, we prove a general nonparametric BvM theorem which says that the posterior measure can be well approximated by a probability measure $P_0$ defined in (5.1) below:

$$P_0(B) = \frac{\int_B \exp\left(-\frac{n}{2} \| f - \hat{f}_{n,\lambda}\|^2 \right) d\Pi(f)}{\int_{H^m(\| \cdot \|)} \exp\left(-\frac{n}{2} \| f - \hat{f}_{n,\lambda}\|^2 \right) d\Pi(f)}$$

for any $B \in \mathcal{B}$, where $\mathcal{B}$ is a collection of $\Pi$-measurable sets in $H^m(\| \cdot \|)$, i.e., $\mathcal{B}$ is a $\sigma$-algebra with respect to $\Pi$. In spite of the complicated appearance, $P_0$ turns out to be a Gaussian measure on $H^m(\| \cdot \|)$ induced by a Gaussian process $W$ which has a closed form (5.3). This Gaussianity facilitates the applications of our BvM theorem, which will be seen in the subsequent sections.

**Theorem 5.1.** (Nonparametric BvM theorem) Suppose Assumptions A1–A3 holds, and $f_0 = \sum_{\nu=1}^{\infty} f^0_\nu \varphi_\nu$ satisfies Condition (S). Furthermore, let Condition (R) be satisfied and the $b_{n1}, b_{n2}$ therein further satisfy, as $n \to \infty$, $m^2(b_{n1} + b_{n2}) = o(1)$. Then we have, as $n \to \infty$,

$$\sup_{B \in \mathcal{B}} |P(B|D_n) - P_0(B)| = o_{P_0}(1).$$

We sketch the proof of Theorem 5.1 here. According to Theorem 4.2, the posterior mass is mostly concentrated on an $M_{\tilde{r}_n}$-ball of $f_0$, denoted as $\mathbb{B}_{M_{\tilde{r}_n}}(f_0)$. Hence, for any $B \in \mathcal{B}$, we decompose $P(B|D_n) = P(B \cap \mathbb{B}_{M_{\tilde{r}_n}}(f_0)|D_n) + P(B \cap \mathbb{B}^{c}_{M_{\tilde{r}_n}}(f_0)|D_n)$. Implied by Theorem 4.2, the second term is uniformly negligible for all $B \in \mathcal{B}$. By applying Taylor expansion to the penalized likelihood in (3.8) (in terms of Fréchet derivatives), we can show that the first term is asymptotically close to $P_0(B)$ uniformly for $B \in \mathcal{B}$ based on the empirical processes techniques.

The conditions in Theorem 5.1 are not restrictive. Indeed, by direct calculations, we can verify that $h \asymp h^* \equiv n^{-\frac{2m+\beta}{2m+\beta+1}}$ satisfies Condition (R) and $m^2(b_{n1} + b_{n2}) = o(1)$ when $m > 1 + \frac{\sqrt{7}}{2} \approx 1.866$ and $1 < \beta < m + \frac{1}{2}$. Therefore, (5.2) holds under $h \asymp h^*$. Interestingly, the optimal posterior contraction rate is simultaneously obtained in this case; see Remark 4.1.

We next show that $P_0$ is equivalent to a Gaussian measure $\Pi_W$, induced by a Gaussian process $W$ defined in (5.3). As will be seen later, the procedures
for Bayesian inference in Sections 5.2 – 5.4 can be easily performed in terms of $W$. Define

\[
\begin{align*}
  a_\nu &= \begin{cases} 
    n(n + \sigma_\nu^{-2})^{-1}, & \nu = 1, \ldots, m, \\
    n(1 + \lambda \rho_\nu)(n(1 + \lambda \rho_\nu) + \rho_\nu^{1+\beta/(2m)})^{-1}, & \nu > m,
  \end{cases} \\
  \xi_\nu &= \begin{cases} 
    (1 + n \sigma_\nu^2)^{-1/2} v_\nu, & \nu = 1, \ldots, m, \\
    (1 + n \rho_\nu(1+\beta/(2m))(1 + \lambda \rho_\nu))^{-1/2} v_\nu, & \nu > m,
  \end{cases}
\end{align*}
\]

where recall that $\sigma_\nu^2$'s are prior variances for $v_\nu$'s assumed in (3.6). Let

\[
W(\cdot) = \sum_{\nu=1}^{\infty} (\xi_\nu + a_\nu \hat{f}_\nu) \varphi_\nu(\cdot),
\]

where $\hat{f}_\nu = V(f_{n,\lambda}, \varphi_\nu)$ for any $\nu \geq 1$. Define $\tilde{f}_{n,\lambda}(\cdot) = \sum_{\nu=1}^{\infty} a_\nu \hat{f}_\nu \varphi_\nu(\cdot)$. Clearly, $W = \tilde{f}_{n,\lambda} + \sum_\nu \xi_\nu \varphi_\nu$ is a Gaussian process with mean function $\tilde{f}_{n,\lambda}$. Note that $\tilde{f}_{n,\lambda} \neq f_{n,\lambda}$, where $\tilde{f}_{n,\lambda}$ is the smoothing spline estimate in (3.4).

Given the data $D_n$, let $\Pi_W$ be the probability measure induced by $W$. Theorem 5.2 below essentially says that $P_0 = \Pi_W$. Together with Theorem 5.1, we note that the posterior distribution $P(\cdot|D_n)$ and $\Pi_W(\cdot)$ are asymptotically close under the total variation distance. This implies the mean function $\tilde{f}_{n,\lambda}$ of $W$ is approximately the posterior mode of $P(\cdot|D_n)$. Therefore, $\tilde{f}_{n,\lambda}$ will be used as the center of the credible region to be constructed.

**Theorem 5.2.** With $f \in H^m(I)$, the Radon-Nikodym derivative of $\Pi_W$ with respect to $\Pi$ is

\[
\frac{d\Pi_W}{d\Pi}(f) = \frac{\exp(-\frac{n}{2} \|f - \tilde{f}_{n,\lambda}\|^2)}{\int_{H^m(I)} \exp(-\frac{n}{2} \|f - \tilde{f}_{n,\lambda}\|^2) d\Pi(f)}.
\]

Theorem 5.2 implies that

\[
\frac{dP_0}{d\Pi}(f) = \frac{d\Pi_W}{d\Pi}(f).
\]

In the end of this section, we point out an important consequence of the duality between nonparametric Bayesian models and smoothing spline models established here. Specifically, we can employ the well developed GCV method to select a proper tuning prior, i.e., the value of $\lambda$ (equivalently, $h$), in practice. Let $h_{GCV}$ be the GCV-selected tuning parameter based on the penalized likelihood function, and $h_{GCV}$ is known to achieve the approximate optimal rate $n^{-1/(2m+1)}$ (see [47]). As illustrated in Sections 5.2 – 5.4,
h should be chosen in the order of $n^{-\frac{1}{2m+\beta}}$ for obtaining nice frequentist properties of credible regions/ intervals. Hence, we set $h$ as $h_{\text{GCV}}^{(2m+1)/(2m+\beta)}$ when tuning the assigned GP prior. Numerical evidence in Section 6 strongly supports this novel method.

5.2. Credible Region in Strong Topology. As the first application of our main result, we consider the construction of an asymptotic credible region for $f$ in terms of $\| \cdot \|$-norm and study its frequentist property in this section. Throughout Sections 5.2 – 5.4, we suppose that $f_0$ satisfies Condition (S), and set $h \approx h^* = n^{-\frac{1}{2m+\beta}}$ for simplicity. It will be shown that such a selection of $h$ yields credible regions with radius of optimal order $n^{-\frac{2m+\beta}{2m+\beta+1}}$.

We first re-write $W$ defined in (5.3) as $\tilde{f}_n,\lambda + Z$, where $Z(\cdot) = \sum_{\nu=1}^{\infty} \xi_\nu \varphi_\nu(\cdot)$, and then study the asymptotic behavior of $Z$. Assume for some constant $c_0 > 0$, $\rho_\nu / (c_0 \nu)^{2m} \to 1$ as $\nu \to \infty$, and define

$$c_\nu = \begin{cases} (n + \sigma_\nu^{-2})^{-1}, & \nu = 1, \ldots, m, \\ (n(1 + \lambda \rho_\nu) + \rho_\nu^{1+\beta/(2m)})^{-1}, & \nu > m. \end{cases}$$

For $j = 1, 2$, let

$$\zeta_j = c_0 h \sum_{\nu=1}^{\infty} (nc_\nu)^2 (1 + \lambda \rho_\nu), \quad \tau_j^2 = 2c_0 h \sum_{\nu=1}^{\infty} (nc_\nu)^2 (1 + \lambda \rho_\nu)^2.$$  

Lemma 5.3. As $n \to \infty$, $(n \sqrt{c_0 h} \| Z \|^2 - (c_0 h)^{-1/2} \zeta_1) / \tau_1 \xrightarrow{d} N(0, 1)$.

We construct an asymptotic credible region with credibility level $(1 - \alpha)$:

$$R_n(\alpha) = \left\{ f \in H^m(I) : \| f - \tilde{f}_{n,\lambda} \| \leq \sqrt{\frac{\zeta_1 + (c_0 h)^{1/2} \tau_1 \alpha}{c_0 n h}} \right\}.$$  

The above $R_n(\alpha)$ is very similar as the credible region $C_{sn}(\alpha)$ (defined in (2.8)) for the Gaussian sequence model under prior $\textbf{P2}$. In particular, we note that $\zeta_1$ and $\tau_1^2$ in $R_n(\alpha)$ play similar roles as $E_{sn}$ and $D_{sn}$ in $C_{sn}(\alpha)$. On the other hand, some direct calculations based on Lemma S.1 indicate that $\zeta_1$ ($\tau_1^2$) is slightly larger than $E_{sn}$ ($D_{sn}$). This is mainly due to the different norms used in constructing $C_{sn}(\alpha)$ and $R_n(\alpha)$. The posterior coverage of
$R_n(\alpha)$ follows from Theorem 5.1 and Lemma 5.3:

$$P(R_n(\alpha)|D_n) = P_0(R_n(\alpha)) + o_{p_{f_0}}(1) = P(W \in R_n(\alpha)) + o_{p_{f_0}}(1)$$

$$= P \left( \|W - \tilde{f}_{n,\lambda}\| \leq \sqrt{\frac{\zeta_{1} + (c_0h)^{1/2}\tau_{1}\alpha}{c_0nh}} \right) + o_{p_{f_0}}(1)$$

$$= P \left( \|Z\| \leq \sqrt{\frac{\zeta_{1} + (c_0h)^{1/2}\tau_{1}\alpha}{c_0nh}} \right) + o_{p_{f_0}}(1)$$

$$= P(\|n\sqrt{c_0h}\|Z\|^2 - (c_0h)^{-1/2}\zeta_{1})/\tau_{1} \leq \alpha) + o_{p_{f_0}}(1)$$

$$= 1 - \alpha + o_{p_{f_0}}(1).$$

We next examine the frequentist property of $R_n(\alpha)$ by first studying the asymptotic property of its center, i.e., $\tilde{f}_{n,\lambda}$, in Lemma 5.4.

**Lemma 5.4.** Let Assumptions A1–A3 be satisfied, and $h \asymp h^* = n^{-1/(2m+\beta)}$ for $m > 1 + \sqrt{3}/2$ and $1 < \beta < m + 1/2$. Then, as $n \to \infty$,

$$c_0nh\|\tilde{f}_{n,\lambda} - f_0\|^2 = \zeta_2 + c_0nhW(n) + o_{p_{f_0}}(1),$$

where $W(n)$ is a random variable satisfying $\frac{1}{2} c_0^{-1} n^{-1/2} W(n) \overset{d}{\to} N(0,1)$.

**Lemma 5.5.** $\tau_1 \geq \tau_2$ and $\lim_{n \to \infty}(\zeta_1 - \zeta_2) > 0$.

By Lemma 5.5, $\zeta_1 - \zeta_2 > c'$ for some positive constant $c' > 0$. By Lemma 5.4, we have $c_0nhW(n) = o_{p_{f_0}}(1)$, and with $P_{f_0}$ probability approaching one, $c_0nh\|\tilde{f} - f_0\|^2 = \zeta_2 + c_0nhW(n) + o_{p_{f_0}}(1) \leq \zeta_1 + (c_0h)^{1/2}\tau_{1}\alpha$, which is exactly the event $f_0 \in R_n(\alpha)$. In view of the above discussions, Theorem 5.6 shows that, if $f_0$ satisfies Condition (S) and $h \asymp h^*$, then $R_n(\alpha)$ is valid in the frequentist sense, and possesses the optimal contraction rate $n^{-\frac{2m+\beta-1}{2(2m+\beta)}}$.

**Theorem 5.6.** Under the conditions of Lemma 5.4, we have for any $\alpha \in (0,1)$, $\lim_{n \to \infty} P_{f_0}(f_0 \in R_n(\alpha)) = 1$.

5.3. **Credible Region in Weak Topology.** In this section, we will construct an asymptotic credible region using a *weaker* topology such that the truth can be covered with probability approaching exactly $1 - \alpha$. This echoes with the use of a weaker norm in Section 2. As far as we know, this is the first frequentist-matching Bayesian credible sets in the nonparametric regression
models. We apply the general BvM theorem in Section 5.1 and employ a strong approximation result ([41]) in this section.

For any \( f \in H^m(\mathbb{I}) \) with \( f(\cdot) = \sum_{\nu=1}^{\infty} f_{\nu} \varphi_{\nu}(\cdot) \), define \( \|f\|_{2^\dagger} = \sum_{\nu=1}^{\infty} d_{\nu} f_{\nu}^2 \), where \( d_{\nu} \) is a given positive sequence satisfying \( d_{\nu} \approx \nu^{-1} (\log 2\nu)^{-\tau} \) for a constant \( \tau > 1 \). For instance, one can choose \( d_1 = \ldots = d_m = 1 \) and \( d_{\nu} = \rho_{\nu}^{-1/(2m)} (\log(\rho_{\nu} + 1))^{-\tau} \) for \( \nu > m \). Define

\[
R_n^\dagger(\alpha) = \left\{ f \in H^m(\mathbb{I}) : \|f - \tilde{f}_{n,\lambda}\|_{2^\dagger} \leq \sqrt{c_\alpha/n} \right\},
\]

where \( c_\alpha \) is defined in Section 2.1. We first check that \( R_n^\dagger(\alpha) \) indeed covers \( (1 - \alpha) \) posterior mass. Similar analysis on the posterior coverage of \( C_n^\dagger(\alpha) \) in Section 2.1 implies that \( n \sum_{\nu} d_{\nu} \xi_{\nu}^2 \) weakly converges to \( \sum_{\nu} d_{\nu} \eta_{\nu}^2 \), where \( \eta_1, \eta_2, \ldots \overset{iid}{\sim} N(0, 1) \). It then follows by Theorem 5.1 that

\[
P(R_n^\dagger(\alpha)|D_n) = P_0(R_n^\dagger(\alpha)) + o_{P_0} \left( 1 \right)
= P_n(W - \tilde{f}_{n,\lambda}^2 \leq c_\alpha) + o_{P_0} \left( 1 \right)
= P_n(\sum_{\nu} d_{\nu} \xi_{\nu}^2 \leq c_\alpha) + o_{P_0} \left( 1 \right) \to 1 - \alpha,
\]

in \( P_{f_0} \)-probability.

To study the frequentist coverage of the credible set \( R_n^\dagger(\alpha) \), we need to assume that the basis functions \( \varphi_{\nu} \) and the real sequence \( \rho_{\nu} \) come from the following ordinary differential system:

\[
(-1)^m \varphi_{\nu}^{(2m)}(\cdot) = \rho_{\nu} B(\cdot) \pi(\cdot) \varphi_{\nu}(\cdot), \quad \varphi_{\nu}^{(j)}(0) = \varphi_{\nu}^{(j)}(1) = 0, \quad j = m, \ldots, 2m - 1,
\]

where recall that \( \pi(\cdot) \) is the marginal density of the covariate \( X \) and \( B(\cdot) \) is the function defined in Assumption A1.

Theorem 5.7 below proves that the credible region \( R_n^\dagger(\alpha) \) constructed under weak topology perfectly matches with the frequentist coverage.

**Theorem 5.7.** Let \( (\varphi_{\nu}, \rho_{\nu}) \) be the eigenvalues and eigenfunctions of the ODE system (5.7). Suppose that the conditions of Lemma 5.4 hold, and there exists a positive constant \( C_1 \) such that \( E_{f_0} \{ \exp(\|e|/C_1) \} < \infty \). Then we have for any \( \alpha \in (0, 1) \), \( \lim_{n \to \infty} P_{f_0}^n (f_0 \in R_n^\dagger(\alpha)) = 1 - \alpha \).

It should be noted that the scope of Theorem 5.7 covers a variety of nonparametric models including Examples 3.1 – 3.3. Again, we can choose a proper tuning prior in \( C_n^\dagger(\alpha) \) via the generalized cross validation method.
5.4. Linear Functionals on the Regression Function. In practice, it is also of great interest to consider a linear functional of the regression function \( f \).

Typical examples include (i) evaluation functional: \( F_z(f) = f(z) \), where \( z \in I \) is a fixed number; (ii) integral functional: \( F_\omega(f) = \int_0^1 f(z)\omega(z)dz \), where \( \omega(\cdot) \) is a known deterministic integrable function such as an indicator function. The major goal of this section is to construct asymptotic credible intervals for a general class of linear functionals. In spirit, some of the results in this section are similar as those implied by parametric BvM theorem in the sense that the constructed credible intervals may contract at the parametric rate \( n^{-1/2} \). We note that linear functionals of densities were considered in [36] but without any frequentist justification.

Let \( F : H^m(I) \mapsto \mathbb{R} \) be a linear \( \Pi \)-measurable functional, i.e., \( F(af + bg) = aF(f) + bF(g) \) for any \( a,b \in \mathbb{R} \) and \( f,g \in H^m(I) \). We assume that \( \sup_{\nu \geq 1} |F(\varphi_\nu)| < \infty \). This assumption holds for both \( F_z \) and \( F_\omega \) above due to the uniform boundedness of \( \varphi_\nu \). In addition, we assume that there exists a constant \( \kappa > 0 \) such that for any \( f \in H^m(I) \),

\[
|F(f)| \leq \kappa h^{-1/2} \|f\|.
\]

(5.8)

Lemma 5.8 below (given in [38]) implies that both \( F_z \) and \( F_\omega \) satisfy (5.8).

**Lemma 5.8.** There exists a universal constant \( c > 0 \) s.t. for any \( f \in H^m(I) \), \( \|f\|_{\sup} \leq ch^{-1/2} \|f\| \).

Define the following Bayesian credible interval for \( F(f_0) \)

\[
CI_{n,\lambda}(\alpha) : F(\tilde{f}_{n,\lambda}) \pm z_{\alpha/2} \frac{t_n}{\sqrt{n}},
\]

(5.9)

where

\[
t_n^2 = \sum_{\nu=1}^m \frac{F(\varphi_\nu)^2}{1 + n^{-1} \sigma_\nu^2} + \sum_{\nu>m} \frac{F(\varphi_\nu)^2}{1 + \lambda \rho_\nu + n^{-1} \rho_\nu^{1+\beta/(2m)}}.
\]

If \( t_n \) converges to a positive constant, then \( CI_{n,\lambda}(\alpha) \) contracts at the parametric rate \( n^{-1/2} \); see (5.17) and (5.18).

The posterior coverage of \( CI_{n,\lambda}(\alpha) \) can be easily checked via Theorems 5.1 and 5.2 as follows. It follows by Theorem 5.1 that

\[
|P(F(f) \in CI_{n,\lambda}(\alpha)|D_n) - P_0(F^{-1}(CI_{n,\lambda}(\alpha)))| \leq \sup_{B \in B} |P(B|D_n) - P_0(B)| \rightarrow 0,
\]

in \( P_{f_0} \)-probability. On the other hand, it follows by Theorem 5.2 that

\[
P_0(F^{-1}(CI_{n,\lambda}(\alpha))) = P(W \in F^{-1}(CI_{n,\lambda}(\alpha)))
\]

\[
= P(F(W) \in CI_{n,\lambda}(\alpha))
\]

\[
= P(-z_{\alpha/2} t_n/\sqrt{n} \leq F(Z) \leq z_{\alpha/2} t_n/\sqrt{n}) = 1 - \alpha,
\]
where the last equality follows from the fact that $F(Z) = \sum_{\nu=1}^{\infty} \xi_\nu F(\varphi_\nu) \sim N(0, t^2_n/n)$ given that $\xi_\nu$'s are independent zero-mean Gaussian variables.

Theorem 5.9 below shows that $CI_{n,\lambda}$ covers the true value $F(f_0)$ with probability asymptotically at least $1 - \alpha$ for a general class of functionals $F$ satisfying (5.10). Under an additional condition on $F$, i.e., $0 < \sum_{\nu=1}^{\infty} F(\varphi_\nu)^2 < \infty$, we can even prove that the coverage is asymptotically exact.

**Theorem 5.9.** Let the assumptions in Theorem 5.1 be satisfied. Suppose $m > 1 + \sqrt{3}/2$, $1 < \beta < m + 1/2$, and $f_0$ satisfies Condition (S$^\dagger$):

\[ \sum_{\nu=1}^{\infty} |f_0^0|^2 \rho_\nu^{1+\beta/2m} < \infty. \]

Furthermore, there exists a constant $r \in [0, 1]$ such that for $j = 1, 2$

\[ \liminf_{n \to \infty} h^r \left( \sum_{\nu=1}^{m} \frac{F(\varphi_\nu)^2}{(1 + n^{-1} \sigma_\nu^2)^j} + \sum_{\nu>m} \frac{F(\varphi_\nu)^2}{(1 + \lambda \rho_\nu + (\lambda \rho_\nu)^{1+\beta/(2m)})^j} \right) > 0. \]

Then

\[ \liminf_{n \to \infty} P_{f_0}^n (F(f_0) \in CI_{n,\lambda}(\alpha)) \geq 1 - \alpha. \]

In particular, if $0 < \sum_{\nu=1}^{\infty} F(\varphi_\nu)^2 < \infty$ and (5.10) holds for $r = 0$, then $CI_{n,\lambda}(\alpha)$ asymptotically achieves $(1 - \alpha)$ frequentist coverage, that is,

\[ \lim_{n \to \infty} P_{f_0}^n (F(f_0) \in CI_{n,\lambda}(\alpha)) = 1 - \alpha. \]

Remark that Condition (S$^\dagger$) is slightly stronger than Condition (S).

We next verify that Condition (5.10) holds when $F$ is the evaluation functional or integral function, and model (3.1) is Gaussian, i.e., the setting of Example 3.1. The proof relies on a nice closed form of $\varphi_\nu$ and a careful analysis of the trigonometric functions.

**Proposition 5.10.** Suppose $m = 2$, $X \sim Unif[0,1]$, and $Y|f, X \sim N(f(X), 1)$.

(i) If $F = F_z$ for any $z \in (0, 1)$, then (5.10) holds for $r = 1$;

(ii) If $F = F_\omega$ for any $\omega \in L^2(\mathbb{I})\setminus\{0\}$, then $0 < \sum_{\nu=1}^{\infty} F(\varphi_\nu)^2 < \infty$ and (5.10) holds for $r = 0$.

By carefully examining the proof of Theorem 5.9, we find that when $F = F_z$, the inequality (5.11) is actually strict. However, when $F = F_\omega$, $CI_{n,\lambda}(\alpha)$ covers the truth with probability approaching $1 - \alpha$. Therefore, we observe a subtle difference between these two types of functional, which can still be empirically detected in the simulations; see Section 6.
In the end, we give a concrete example of $CI_{n,\lambda}(\alpha)$. Under the setup of Proposition 5.10, it follows from [38] that $(\varphi_\nu, \rho_\nu)$ in Assumption A2 can be chosen as the eigensystem of the following uniform free beam problem:

\begin{equation}
\varphi_\nu^{(i)}(\cdot) = \rho_\nu \varphi_\nu(\cdot), \quad \varphi_\nu^{(j)}(0) = \varphi_\nu^{(j)}(1) = 0, \quad j = 2, 3.
\end{equation}

The eigenvalues satisfy $\lim_{\nu \to \infty} \rho_\nu / (\pi \nu)^4 = 1$; see [24, Problem 3.10]. The normalized solutions to (5.13) are

\begin{equation}
\varphi_1(z) = 1, \quad \varphi_2(z) = \sqrt{3}(2z - 1),
\end{equation}

\begin{equation}
\varphi_{2k+1}(z) = \frac{\sin(\gamma_{2k+1}(z - 1/2))}{\sin(\gamma_{2k+1}/2)} + \frac{\sinh(\gamma_{2k+1}(z - 1/2))}{\sinh(\gamma_{2k+1}/2)}, \quad k \geq 1.
\end{equation}

\begin{equation}
\varphi_{2k+2}(z) = \frac{\cos(\gamma_{2k+2}(z - 1/2))}{\cos(\gamma_{2k+2}/2)} + \frac{\cosh(\gamma_{2k+2}(z - 1/2))}{\cosh(\gamma_{2k+2}/2)}, \quad k \geq 1,
\end{equation}

where $\gamma_\nu = \rho_\nu^{1/4}$ satisfying $\cos(\gamma_\nu) \cosh(\gamma_\nu) = 1$; see [6, page 295–296].

For $F = F_z$ with $z \in (0, 1)$, we have

\begin{equation}
CI_{n,\lambda} : f_{n,\lambda}(z) \pm z_{\alpha/2} t_n \sqrt{n},
\end{equation}

where $t_n^2 = \sum_{\nu=1}^{2} \frac{\varphi_\nu(z)^2}{1 + n^{-1} \sigma_\nu^2} + \sum_{\nu>2} \frac{\varphi_\nu(z)^2}{1 + \lambda \rho_\nu + n^{-1} \rho_\nu^{1+3/4}}$, with $\rho_\nu \approx (\pi \nu)^4$ and $\varphi_\nu$ given in (5.14) – (5.16). Here, $t_n^2 = 1/h$. For $F = F_\omega$ with $\omega \in L^2(\mathbb{I}) \setminus \{0\}$, we have

\begin{equation}
CI_{n,\lambda} : \int_0^1 f_{n,\lambda}(z) \omega(z)dz \pm z_{\alpha/2} t_n \sqrt{n},
\end{equation}

where $t_n^2 = \sum_{\nu=1}^{2} \frac{F_\omega(\varphi_\nu)^2}{1 + n^{-1} \sigma_\nu^2} + \sum_{\nu>2} \frac{F_\omega(\varphi_\nu)^2}{1 + \lambda \rho_\nu + n^{-1} \rho_\nu^{1+3/4}} \to \sum_{\nu=1}^{\infty} F_\omega(\varphi_\nu)^2 = \int_0^1 \omega(z)^2dz > 0$, as $n \to \infty$. The last equality follows from the trivial fact $F_\omega(\varphi_\nu) = \int_0^1 \omega(z) \varphi_\nu(z)dz$ is the $\nu$-th Fourier coefficient of $\omega$. Theorem 5.9 implies that the frequentist coverage of (5.17) is strictly larger than $1 - \alpha$, while (5.18) has an exact frequentist coverage probability.

To use (5.17) and (5.18), we need explicit expressions of $(\rho_\nu, \varphi_\nu)$. However, the explicit expressions of the eigenvalues $\rho_\nu$ are not available, and hence we cannot directly use formulas (5.14) – (5.16) to find the eigen-functions $\varphi_\nu$. To circumvent this problem, in numerical study we use an equivalent kernel approach. To be more specific, it follows by [38] that the reproducing
kernel function \( K \) with respect to the inner product \( \langle \cdot, \cdot \rangle \) satisfies
\[
K(s, t) = \sum_{\nu=1}^{\infty} \varphi_{\nu}(s) \varphi_{\nu}(t) / (1 + \lambda_{\nu}).
\]
On the other hand, it follows by [32, page 184–185] that in the setting of Proposition 5.10 the kernel function \( K(s, t) \) yields an approximation
\[
K(s, t) \approx \frac{1}{2\sqrt{2h}} \exp \left( -\frac{|s - t|}{\sqrt{2h}} \right) \left( \sin \left( \frac{|s - t|}{\sqrt{2h}} \right) + \cos \left( \frac{s - t}{\sqrt{2h}} \right) \right) + \frac{1}{2\sqrt{2h}} \left( \Phi \left( \frac{s + t}{\sqrt{2h}}, \frac{s - t}{\sqrt{2h}} \right) + \Phi \left( \frac{2 - s - t}{\sqrt{2h}}, \frac{t - s}{\sqrt{2h}} \right) \right),
\]
where \( \Phi(u, v) = \exp(-u)(\cos(u) - \sin(u) + 2 \cos(v)) \). Then \((\rho_{\nu}, \varphi_{\nu})\) can be found by performing a spectral decomposition on the above function.

6. Simulations.

6.1. Gaussian Sequence Model. In this section, performance of the adaptive credible regions (ACR) and modified adaptive credible regions (MACR) constructed by Priors \( \text{P1} \) and \( \text{P2} \) is demonstrated through a simulation study. These adaptive regions are constructed in Section S.3 of the supplementary documents. For simplicity, we use ACR1, MACR1, ACR2 and MACR2 to denote the regions constructed by formulas (S.2), (S.3), (S.6) and (S.7) respectively. Suppose that the samples \( Y_i \) are generated from (2.1) with true parameters being
\[
\theta_{i}^{0} = (i \log i)^{-3/2}, \; i = 2, 3, \ldots
\]
Thus, the true regularity is \( q_0 = 1 \). The model errors \( \epsilon_i \) are iid independent normal random variables. Results are based on 10,000 independent experiments. In each experiment, \((1 - \alpha)\) ACR1, MACR1, ACR2 and MACR2 were constructed. We examined the coverage proportions (CP) of these regions, i.e., the proportions of the regions that cover the true model parameter \( \theta^0 \). To fully investigate how the coverage proportions relate to the credible levels, we tested a set of different values of the credibility level \( 1 - \alpha \).

Results are summarized in Figures 3 and 4. It can be seen that the coverage proportions of both ACR1 and ACR2 become larger than the credible level \( 1 - \alpha \) as \( n \) becomes sufficiently large. Meanwhile, we observe that ACR2 yields a bit smaller CP than ACR1. This is because ACR2 has smaller radius hence smaller volume. The smaller volume may be viewed as an advantage since this means that the credible region is more “concentrated.”

The fact that ACR2 has smaller volume is not magic. This is because we have used the same \( \hat{q} \) in both formulas (S.2) and (S.6). From a numerical
examination it can be seen that $\hat{E}_n > \hat{E}_n$ and $\hat{D}_n > \hat{D}_n$ (see Figure 2), and hence, $C^{\text{adapt}}_n(\alpha)$ has smaller radius than $C^{\text{adapt}}_n(\alpha)$.

It can also be observed that both MACR1 and MACR2 yield coverage proportions very close to the nominal level $1 - \alpha$ for a variety of values $\alpha$. This means that, under both priors $\mathbf{P1}$ and $\mathbf{P2}$, MACR matches the frequentist coverage better than ACR.

6.2. Smoothing Spline Model. In this section, we investigate the frequentist coverage probabilities of the credible region (5.5) and modified credible region (5.6), and credible intervals for two functionals (5.17) and (5.18). Suppose

$$Y_i = f_0(X_i) + \epsilon_i, \ i = 1, 2, \ldots, n,$$

where $X_i$ are iid uniform over $[0, 1]$, and $\epsilon_i$ are iid standard normal random variables independent of $X_i$. The true regression function $f_0$ was chosen as $f_0(x) = 3\beta_{30,17}(x) + 2\beta_{3,11}(x)$, where $\beta_{a,b}$ is the probability density function for $\text{Beta}(a,b)$. Figure 5 displays the true function $f_0$, from which it can be seen that $f_0$ has both peaks and trouts.
To examine the coverage property of the credible regions, we chose \( n \) ranging from 20 to 2000. For each \( n \), 1,000 independent trials were conducted. From each trial, a credible region (CR) and a modified credible region (MCR) were constructed. Proportions of the CR and MCR covering \( f_0 \) were calculated, and were displayed against the sample sizes. Results are summarized in Figure 6. It can be seen that for different \( 1 - \alpha \), i.e., the credibility levels, the coverage proportions (CP) of CR are greater than \( 1 - \alpha \) when \( n \) is large enough. They even tend to one for large sample sizes. However, the CP of the MCR tends to exactly \( 1 - \alpha \) when \( n \) increases. Thus, the numerical results confirm our theory developed in Sections 5.2 and 5.3.

To examine the coverage property of credible intervals, we chose \( n = 2^5, 2^7, 2^8, 2^9 \) to demonstrate the trend of coverage along with increasing
sample sizes. For pointwise functional, we considered $F = F_z$ for 15 evenly-spaced $z$ points in $[0, 1]$. For each $z$, a credible interval based on (5.17) was constructed. We then calculated the coverage probability of this interval based on 1,000 independent experiments, that is, the empirical proportion of the intervals (among the 1,000 intervals) that cover the true value $f_0(z)$. Figure 7 summarizes the results for different credibility levels $\alpha$, where coverage probabilities are plotted against the corresponding points $z$. It can be seen that the coverage probability of the pointwise intervals is a bit larger than $1 - \alpha$ for all $\alpha$ and $n$ being considered. This is consistent with Proposition 5.10 (i), except for the points near the right peak of $f_0$. Indeed, at those points near the right peak, under-coverage has been observed. This is a common phenomenon in the frequentist literature: the peak and trouts
may affect the coverage property of the pointwise interval; see [34, 38].

For integral functional, we considered $F = F_{\omega z_0}$ for $\omega_{z_0}(z) = I(0 \leq z \leq z_0)$ with 15 evenly-spaced $z_0$ points in $[0, 1]$. We evaluated the coverage probability at each $z_0$ based on 1,000 experiments. Figure 8 summarizes the results for different credibility levels $\alpha$, where coverage probabilities are plotted against the corresponding points $z_0$. It can be seen that, as $n$ increases, the coverage probability of the integral intervals tends to $1 - \alpha$ for all $\alpha$. This phenomenon is consistent with our theory, i.e., Proposition 5.10 (ii).

Acknowledgements. We thank PhD student Meimei Liu at Purdue for help with the simulation study.

REFERENCES

[1] Birkhoff, D. (1908). Boundary Value and Expansion Problems of Ordinary Linear Differential Equations. *Transactions of the American Mathematical Society*, 9, 373–395.

[2] Belitser, E. and Ghoshal, S. (2003). Adaptive Bayesian inference on the mean of an infinite dimensional normal distribution. *Annals of Statistics*, 31, 536–559.

[3] Brown, L. D. and Low, M. G. (1996). Asymptotic equivalence of nonparametric regression and white noise. *Annals of Statistics*, 24, 2384–2398.

[4] Bickel, P. J. and Kleijn, B. J. K. (2012). The semiparametric Bernstein-von Mises theorem. *Annals of Statistics*, 40, 206–237.

[5] Bickel, P. J. and Ritov, Y. (2003). Nonparametric estimators which can be “plugged-in”. *Annals of Statistics*, 31, 1033–1053.
Fig 6. Coverage proportion (CP) of CR and MCR, constructed by (5.5) and (5.6) respectively, based on various sample sizes and credibility levels. The dotted red line indicates the position of the $1 - \alpha$ credibility level.

[6] Courant, R. and Hilbert, D. (1937). Methods of Mathematical Physics, Vol. 1. Interscience Publishers, Inc., New York.

[7] Cameron, R. H. and Martin, W. T. (1944). Transformations of Wiener Integrals under Translations. Annals of Mathematics, 45, 386–396.

[8] Castillo, I. (2014). On Bayesian supremum norm contraction rates. Annals of Statistics, 42, 2058–2091.

[9] Castillo, I. (2012). A semiparametric Berstein-von Mises theorem for Gaussian process priors. Prob. Theory Rel. Fields, 152, 53–99.

[10] Castillo, I. and Nickl, R. (2013). Nonparametric Bernstein-von Mises theorem in Gaussian white noise. Annals of Statistics, 41, 1999–2028.

[11] Castillo, I. and Nickl, R. (2014). On the Bernstein-von Mises phenomenon for nonparametric Bayes procedures. Annals of Statistics, 42, 1941–1969.

[12] Castillo, I. and Rousseau, J. (2014). A Bernstein-von Mises theorem for smooth functionals in semiparametric models. Preprint.
Fig 7. Coverage proportion (CP) of the credible interval for $F_z(f_0)$ versus $z$. The dotted red line indicates the position of the $1 - \alpha$ credibility level.

[13] Cox, D. (1993). An analysis of Bayesian inference for nonparametric regression. *Annals of Statistics*, 21, 903–923.

[14] Cheng, G. and Kosorok, M. R. (2008). General frequentist properties of the posterior profile distribution. *Annals of Statistics*, 36, 1819–1853.

[15] Cheng, G. and Kosorok, M. R. (2009). The Penalized Profile Sampler. *Journal of Multivariate Analysis*, 100, 345–362.

[16] Cheng, G. and Shang, Z. (2013). Joint asymptotics for semi-nonparametric regression models with partially linear structure. http://arxiv.org/abs/1311.2628.

[17] de Jong, P. (1987). A central limit theorem for generalized quadratic forms. *Probability Theory & Related Fields*, 75, 261–277.

[18] Freedman, D. (1999). On the Bernstein-von Mises theorem with infinite-dimensional parameters. *Annals of Statistics*, 27, 1119-1140.

[19] Giné, E. and Nickl, R. (2011). Rates of contraction for posterior distributions in $L^r$-metrics, $1 \leq r \leq \infty$. *Annals of Statistics*, 39, 2795–3443.
Fig 8. Coverage proportion (CP) of the credible interval for $F_{\omega_{z_0}}(f_0)$ versus $z_0$. The dotted red line indicates the position of the $1 - \alpha$ credibility level.

[20] Ghosal, S., Ghosh, J. K. and Ramamoorthi, R. V. (1999). Posterior consistency of Dirichlet mixtures in density estimation. *Annals of Statistics*, 27, 143–158.
[21] Ghosal, S., Ghosh, J. K. and van der Vaart, A. W. (2000). Convergence rates of posterior distributions. *Annals of Statistics*, 28, 500–531.
[22] Hájek, J. (1962). On linear statistical problems in stochastic processes. *Czechoslovak Mathematical Journal*, 12, 404–444.
[23] Hoffmann-Jørgensen, Shepp, L. A. and Dudley, R. M. (1979). On the lower tail of Gaussian seminorms. *Annals of Probability*, 7, 193–384.
[24] Johnstone, I. M. (2013). *Gaussian Estimation: Sequence and Wavelet Models*.
[25] Knapik, B. T., van der Vaart, A. W. and van Zanten, J. H. (2011). Bayesian inverse problems with Gaussian priors. *Annals of Statistics*, 39, 2626–2657.
[26] Knapik, B. T., Szabó, B. T., van der Vaart, A. W. and van Zanten, J. H. (2013). Bayes procedures for adaptive inference in inverse problems for the white noise model. Preprint.
[27] Kosorok, M. R. (2008). Introduction to Empirical Processes and Semiparametric Inference. Springer: New York.

[28] Kuelbs, J., Li, W. V. and Linde, W. (1994). The Gaussian measure of shifted balls. Probability Theory and Related Fields, 98, 143–162.

[29] Leahu, H. (2011). On the Bernstein-von Mises phenomenon in the Gaussian white noise model. Electronic Journal of Statistics, 5, 373-404.

[30] Li, W. V. (1999). A Gaussian correlation inequality and its applications to small ball probabilities. Electronic Communications in Probability, 4, 111–118.

[31] Morris, C. N. (1982). Natural exponential families with quadratic variance functions. Annals of Statistics, 10, 65–80.

[32] Messer, K. and Goldstein, L. (1993). A new class of kernels for nonparametric curve estimation. Annals of Statistics, 21, 179–195.

[33] Nickl, R. (2014). Discussion to “Frequentist coverage of adaptive nonparametric Bayesian credible sets. Annals of Statistics, to appear.

[34] Nychka, D. (1988). Bayesian confidence intervals for smoothing splines. Journal of the American Statistical Association, 83, 1134–1143.

[35] Nussbaum, M. (1996) Asymptotic equivalence of density estimation and Gaussian white noise. Annals of Statistics, 24, 2399–2430.

[36] Rivoirard, V. and Rousseau, J. (2012). Bernstein-von Mises theorem for linear functionals of the density. Annals of Statistics, 40, 1489–1523.

[37] Shang, Z. (2010). Convergence rate and Bahadur type representation of general smoothing spline M-Estimates. Electronic Journal of Statistics, 4, 1411–1442.

[38] Shang, Z. and Cheng, G. (2013). Local and global asymptotic inference in smoothing spline models. Annals of Statistics, 41, 2608–2638.

[39] Szabó, B., van der Vaart, A. W. and van Zanten, H. (2014). Empirical Bayes scaling of Gaussian priors in the white noise model. Electronic Journal of Statistics, 7, 991-1018.

[40] Szabó, B., van der Vaart, A. W. and van Zanten, H. (2014). Frequentist coverage of adaptive nonparametric Bayesian credible sets. Annals of Statistics.

[41] Tusnady, G. (1977). A remark on the approximation of the sample distribution function in the multidimensional case. Periodica Mathematica Hungarica, 8, 53–55.

[42] Utreras, F. I. (1988). Boundary effects on convergence rates for Tikhonov regularization. Journal of Approximation Theory, 54, 235–249.

[43] van der Vaart, A. W. and van Zanten, J. H. (2007). Bayesian inference with rescaled Gaussian process priors. Electronic Journal of Statistics, 1, 433–448.

[44] van der Vaart, A. W. and van Zanten, J. H. (2008). Reproducing kernel Hilbert spaces of Gaussian priors. IMS Collections. Publishing the Limits of Contemporary Statistics: Contributions in Honor of Jayanta K. Ghosh. 3, 200–222.

[45] van der Vaart, A. W. and van Zanten, J. H. (2008). Rates of contraction of posterior distributions based on Gaussian process priors. Annals of Statistics, 36, 1031–1508.

[46] van der Vaart, A. W. and van Zanten, J. H. (2008). Adaptive Bayesian estimation using a Gaussian random field with inverse Gamma bandwidth. Annals of Statistics, 37, 2655–2675.

[47] Wahba, G. (1985). A comparison of GCV and GML for choosing the smoothing parameter in the generalized spline smoothing problem. Annals of Statistics, 13, 1251–1638.

[48] Wahba, G. (1990). Spline Models for Observational Data. SIAM, Philadelphia.
[49] Zhao, L. H. (2000). Bayesian aspects of some nonparametric problems. *Annals of Statistics*, **28**, 532–552.
Supplementary document to

NONPARAMETRIC BERNSTEIN-VON MISES PHENOMENON: A TUNING PRIOR PERSPECTIVE

BY ZUOFENG SHANG$ and GUANG CHENG$

Department of Statistics
Purdue University

This supplementary document contains the proofs of the main results and additional materials not included in the main manuscript.

We organize this document as follows:

• Section S.1 contains proofs in Section 2.1, i.e., proofs of Theorems 2.1 and 2.2.
• Section S.2 contains proofs in Section 2.2, i.e., proofs of Theorem 2.3 and several technical lemmas.
• In Section S.3, we construct adaptive credible regions under priors $P_1$ and $P_2$, and explore their frequentist coverage properties.
• Section S.4 includes some preliminary results in Section 3, e.g., the expression of Radon-Nikodym derivative of $\Pi_\lambda$ with respect to $\Pi$, together with several limiting results.
• Section S.5 includes results on posterior contraction rates under a type of Sobolev norm, as given in Section 4.
• Section S.6 contains the proof of the nonparametric BvM theorem, i.e., Theorem 5.1 presented in Section 5.1, together with the specification of the probability measure $P_0$.
• Section S.7 includes the proofs of the results in Section 5.2, in particular, the coverage property of the credible regions under a type of Sobolev norm.
• Section S.8 includes the proofs of the results in Section 5.3, in particular, the coverage property of the credible regions under a weaker norm.

$^1$Visiting Assistant Professor
$^5$Corresponding Author. Associate Professor, Research Sponsored by NSF CAREER Award DMS-1151692, DMS-1418042, Simons Foundation 305266. Guang Cheng is on sabbatical at Princeton while this work is finalized; he would like to thank the Princeton ORFE department for its hospitality and support.
• Section S.9 includes the proofs of the results in Section 5.4, in particular, the coverage property of the linear functional of the regression function. Both pointwise functional and integral functional are addressed.

S.1. Proofs in Section 2.1.

Proof of Theorem 2.1. For simplicity, denote $\gamma = 2p/(2q + 1)$.

Proof of Part (1). If $\alpha_0 \geq \gamma$, implying that $h^{-2q_0+1} = n^{\alpha_0/\gamma} \geq n$, then for any $\theta_0 \in \ell_2(q_0)$, by dominated convergence theorem, as $n \to \infty$,

$$n \sum_{i=1}^{\infty} \frac{\lambda^2 i^{4p}}{(1 + \lambda i^{2p})^2} |\theta^0_i|^2 = nh^2 \sum_{i=1}^{\infty} \frac{(ih)^{4p-2q_0}}{(1 + (ih)^2)^2} |\theta^0_i|^2 i^{2q_0} = o(nh^2p) = o(h^{-1}),$$

which leads to that

$$n \sum_{i=1}^{\infty} (\hat{\theta}_i - \theta^0_i)^2 = n \sum_{i=1}^{\infty} \frac{(-\lambda i^{2p} \theta^0_i + n^{-1/2} \epsilon_i)^2}{(1 + \lambda i^{2p})^2} = o_P(h^{-1}) + \sum_{i=1}^{\infty} \frac{1}{(1 + \lambda i^{2p})^2} + \sum_{i=1}^{\infty} \frac{\epsilon_i^2 - 1}{(1 + \lambda i^{2p})^2}.$$ 

It follows by direct calculations that

$$E\{| \sum_{i=1}^{\infty} \frac{\epsilon_i^2 - 1}{(1 + \lambda i^{2p})^2} |^2 \} = \sum_{i=1}^{\infty} \frac{2}{(1 + \lambda i^{2p})^4} = O(h^{-1}),$$

therefore, $\sum_{i=1}^{\infty} \frac{\epsilon_i^2 - 1}{(1 + \lambda i^{2p})^2} = O_P(h^{-1/2})$. This implies that

$$nh \|\theta_0 - \hat{\theta}\|^2_2 = o_P(1) + F_n + h \sum_{i=1}^{\infty} \frac{\epsilon_i^2 - 1}{(1 + \lambda i^{2p})^2} = F_n + o_P(1),$$

where $F_n = \sum_{i=1}^{\infty} \frac{h}{(1 + \lambda i^{2p})^2}$. It is easy to see that as $h \to 0$, $E_n \to \int_0^\infty \frac{1}{1 + x^{2p}} dx$, and $F_n \to \int_0^\infty \frac{1}{(1 + x^{2p})^2} dx$. So $\inf_{h>0}(E_n - F_n) > 0$. This implies that, with probability approaching one, $\theta_0 \in C_n(\alpha)$.

Proof of Part (2). If $0 < \alpha_0 < \gamma$, then $\alpha_0/(2p) < 1/(2q_0 + 1) \leq 1$ and $2q_0 < 2p/\alpha_0 - 1$. Let $t$ be a constant such that $2q_0 < t < 2p/\alpha_0 - 1$ and $t < 4p - 1$, and let $\theta^0_i = i^{-(1+t)/2}$. It is easy to see that $\sum_i |\theta^0_i|^2 i^{2q_0} = \sum_i i^{1-t+2q_0} < \infty$, and hence, $\theta_0 = \{\theta^0_i\}_{i=1}^{\infty} \in \ell_2(q_0)$. Then

$$n \sum_{i=1}^{\infty} \frac{\lambda^2 i^{4p}}{(1 + \lambda i^{2p})^2} |\theta^0_i|^2 = nh^{t+1} \sum_{i=1}^{\infty} \frac{(ih)^{4p-t-1}}{(1 + (ih)^{2p})^2} \approx nh^t \int_0^\infty \frac{x^{4p-t-1}}{(1 + x^{2p})^2} dx.$$
In the meantime,
\[ E\left\{ \sum_{i=1}^{\infty} \frac{(ih)^{2p} \theta_0^2 \epsilon_i^2}{(1 + (ih)^{2p})^2} \right\} = \sum_{i=1}^{\infty} \frac{(ih)^{4p} \theta_0^2 \epsilon_i^2}{(1 + (ih)^{2p})^4} \]
\[ = \sum_{i=1}^{\infty} \frac{(ih)^{4p} i^{-(t+1)}}{(1 + (ih)^{2p})^4} \approx h^t \int_{0}^{\infty} \frac{x^{4p-t-1}}{(1 + x^{2p})^4} dx, \]

implying that
\[ \sum_{i=1}^{\infty} \frac{(ih)^{2p} \theta_0^2 \epsilon_i^2}{(1 + (ih)^{2p})^2} = O_P(h^{t/2}), \]
and
\[ E\left\{ \sum_{i=1}^{\infty} \frac{\epsilon_i^2}{(1 + (ih)^{2p})^2} \right\} = \sum_{i} \frac{1}{(1 + (ih)^{2p})^2} \approx h^{-1}, \]

which further imply that \( \sum_{i=1}^{\infty} \frac{\epsilon_i^2}{(1 + (ih)^{2p})^2} = O_P(h^{-1}) \). Therefore,
\[ nh\|\theta_0 - \hat{\theta}\|^2 = nh \sum_{i=1}^{\infty} \frac{(ih)^{4p} \theta_0^2 \epsilon_i^2}{(1 + (ih)^{2p})^2} - 2n^{-1/2} \epsilon_i (ih)^{2p} \theta_0^2 + n^{-1} \epsilon_i^2 \]
\[ = nh \sum_{i=1}^{\infty} \frac{\lambda^2 i^{4p} \epsilon_i^2}{(1 + \lambda i^{2p})^2} + O_P(\sqrt{n}h^{1+t/2} + 1) \]
\[ \approx nh^{t+1}(1 + o_P(1)) = n^{1-\alpha_0(t+1)/2p}(1 + o_P(1)). \]

Since \( 1 - \alpha_0(t+1)/2p > 0 \), the above derivations imply that, with probability approaching one, \( nh\|\theta_0 - \hat{\theta}\|^2 > E_n + \sqrt{D_n h^t \alpha} \), or equivalently, \( \theta_0 \notin C_n(\alpha) \).

**Proof of Theorem 2.2.** Proof of Part (1). Note that
\[ n\|\theta_0 - \hat{\theta}\|^2 \]
\[ = n \sum_{i=1}^{\infty} d_i \frac{(- \lambda i^{2p} \theta_0^2 + n^{-1/2} \epsilon_i)^2}{(1 + \lambda i^{2p})^2} \]
\[ = n \sum_{i=1}^{\infty} d_i (ih)^{4p} \theta_0^2 \epsilon_i^2 (1 + (ih)^{2p})^2 - 2n \sum_{i=1}^{\infty} \frac{d_i (ih)^{2p} \theta_0 \epsilon_i}{(1 + (ih)^{2p})^2} + \sum_{i=1}^{\infty} \frac{d_i \epsilon_i^2}{(1 + (ih)^{2p})^2} \]

It follows by conditions on \( \theta_0 \) and dominated convergence theorem that
\[ n \sum_{i=1}^{\infty} d_i (ih)^{4p} \theta_0^2 \epsilon_i^2 (1 + (ih)^{2p})^2 \lesssim nh^{2q_0+1} \sum_{i=1}^{\infty} \frac{(ih)^{4p-2q_0-1} \theta_0^2 i^{2q_0}}{(1 + (ih)^{2p})^2} = o_nh^{2q_0+1} = o(1), \]
Define $\theta_1$ we have $n$ Therefore, with probability approaching one, 

On the other side, 

Proof of Part (2). Now suppose $0 < \alpha_0 < 2p/(2q_0+1)$, and hence, $\alpha_0(2q_0+1)/(2p)<1$. Choose $t > 0$ with $\alpha_0(2q_0+1+t)/(2p)<1$ and $t < 4p-2q_0-2$. Define $\theta^0_i = i^{-(2q_0+1+t)/2}$ for $i = 1, 2, \ldots$. It is easy to see that $\theta_0 = \{\theta^0_i\}_{i=1}^\infty \in \ell_2(q_0)$. By choice of $d_i$ we have 

On the other side, 

which implies $\sqrt{n} \sum_{i=1}^\infty \frac{d_i (ih)^{2p} |\theta^0_i| |\epsilon_i|}{(1 + (ih)^{2p})^2} = O_P(\sqrt{nh^{2q_0+1+t}(\log n)^{-\tau}})$. Since 

we have 

Therefore, with probability approaching one, $n\|\theta_0 - \hat{\theta}\|_1^2 > c_\alpha$, i.e., $\theta_0 \notin C_n^\dagger(\alpha)$. □
S.2. Proofs in Section 2.2. The following Lemma S.1 shows that both $E_{s_n}$ and $D_{s_n}$ are asymptotically positive constants. The limiting distribution of $Z_{s_n}$ is standard normal as specified in the Lemma S.2.

**Lemma S.1.** For any constant $\delta \geq 1$, we have

\[
 l_{\alpha_0} = \lim_{n \to \infty} \sum_{i=1}^{\infty} \frac{n^{-\alpha_1}}{(1 + \iota^2 q + n^{-1} i^{2p})^\delta} = \begin{cases} 
 \int_0^\infty \frac{1}{(1 + x^{2p})^\delta} dx, & \alpha_0 > q/p \\
 \int_0^\infty \frac{1}{(1 + x^2)^\delta} dx, & 0 < \alpha_0 < q/p \\
 \int_0^\infty \frac{1}{(1 + x^{2q} + x^{2p})^\delta} dx, & \alpha_0 = q/p 
\end{cases}
\]

**Lemma S.2.** As $n \to \infty$, $Z_{s_n} \overset{d}{\to} N(0, 1)$.

**Proof of Lemma S.1.** If $\alpha_0 > q/p$, then $\alpha_1 = 1/(2p)$, and hence

\[
 \sum_{i=1}^{\infty} \frac{n^{-\alpha_1}}{(1 + \iota^2 q + n^{-1} i^{2p})^\delta} = \sum_{i=1}^{\infty} \frac{n^{-1/(2p)}}{(1 + \iota^{n^{-1/(2p)}})^{2q n^{-\alpha_0} + q/p + (\iota^{-1/(2p)})^2}^\delta} \\
 \leq \sum_{i=1}^{\infty} \int_{i-1}^{i} \frac{n^{-1/(2p)}}{\left(1 + x^{-1/(2p)} \right)^{2q n^{-\alpha_0} + q/p + (x^{-1/(2p)})^2}^\delta} dx \\
 = \int_0^\infty \frac{1}{(1 + x^{2q n^{-\alpha_0} + q/p + x^{2p})^\delta} dx \overset{as \ n \to \infty}{\to} \int_0^\infty \frac{1}{(1 + x^{2p})^\delta} dx,
\]

where the last limit follows by dominated convergence theorem. Similarly, we have

\[
 \sum_{i=1}^{\infty} \frac{n^{-1/(2p)}}{(1 + \iota^{n^{-1/(2p)}})^{2q n^{-\alpha_0} + q/p + (\iota^{-1/(2p)})^2}^\delta} \\
 \geq \sum_{i=1}^{\infty} \int_{i}^{i+1} \frac{n^{-1/(2p)}}{\left(1 + x^{-1/(2p)} \right)^{2q n^{-\alpha_0} + q/p + (x^{-1/(2p)})^2}^\delta} dx \\
 = \int_{n^{-1/(2p)}}^\infty \frac{1}{(1 + x^{2q n^{-\alpha_0} + q/p + x^{2p})^\delta} dx \overset{as \ n \to \infty}{\to} \int_0^\infty \frac{1}{(1 + x^{2p})^\delta} dx.
\]

This proves the result for $\alpha_0 > q/p$. The proofs in the other cases are similar and omitted.

**Proof of Lemma S.2.** The logarithm of the moment generating func-
Meanwhile, by direct examinations we have

\[
\log(E\{\exp(tZ_n)\}) = \sum_{i=1}^{\infty} \log(E\{\exp(tD_{sn}^{-1/2} n^{-\alpha_1/2}(\eta_i^2 - 1)/1 + \lambda_i^{2q} + n^{-1}i^{2p})\})
\]

\[
= \sum_{i=1}^{\infty} \left(-\frac{1}{2} \log(1 - \frac{2tD_{sn}^{-1/2} n^{-\alpha_1/2}}{1 + \lambda_i^{2q} + n^{-1}i^{2p}}) - \frac{tD_{sn}^{-1/2} n^{-\alpha_1/2}}{1 + \lambda_i^{2q} + n^{-1}i^{2p}}\right)
\]

\[
= \sum_{i=1}^{\infty} \left(-\frac{1}{2} \left(\frac{2tD_{sn}^{-1/2} n^{-\alpha_1/2}}{1 + \lambda_i^{2q} + n^{-1}i^{2p}} - \frac{1}{2} \left(\frac{2tD_{sn}^{-1/2} n^{-\alpha_1/2}}{1 + \lambda_i^{2q} + n^{-1}i^{2p}}\right)^2\right)
\]

\[
+ O((\frac{2tD_{sn}^{-1/2} n^{-\alpha_1/2}}{1 + \lambda_i^{2q} + n^{-1}i^{2p}})^3) - \frac{tD_{sn}^{-1/2} n^{-\alpha_1/2}}{1 + \lambda_i^{2q} + n^{-1}i^{2p}}\right)
\]

\[
= \frac{t^2}{2} + O\left(\sum_{i=1}^{\infty} (\frac{2tD_{sn}^{-1/2} n^{-\alpha_1/2}}{1 + \lambda_i^{2q} + n^{-1}i^{2p}})^3\right).
\]

By Lemma S.1,

\[
\sum_{i=1}^{\infty} (\frac{2tD_{sn}^{-1/2} n^{-\alpha_1/2}}{1 + \lambda_i^{2q} + n^{-1}i^{2p}})^3 = (2tD_{sn}^{-1/2})^3 n^{-\alpha_1/2} \sum_{i=1}^{\infty} \frac{n^{-\alpha_1}}{(1 + \lambda_i^{2q} + n^{-1}i^{2p})^3} \to 0,
\]

therefore, \(E\{\exp(tZ_n)\} \to \exp(t^2/2)\), which proves the desired result. 

**Proof of Theorem 2.3.** Before formal proofs, we notice a simple fact:

for any \(i \geq 1\),

\[
\lambda_i^{2q} = (n^{-\alpha_0/(2q)} i)^{2q} \leq (n^{-\alpha_1} i)^{2q}, \quad n^{-1}i^{2p} = (n^{-1/(2p)} i)^{2p} \leq (n^{-\alpha_1} i)^{2p}.
\]

Meanwhile, by direct examinations we have

\[
\|\hat{\theta}_n - \theta_0\|^2
\]

\[
= \sum_{i=1}^{\infty} \frac{(-\lambda_i^{2q} + n^{-1}i^{2p})\theta_i^0 + \epsilon_i/\sqrt{n})^2}{(1 + \lambda_i^{2q} + n^{-1}i^{2p})^2}
\]

\[
= \sum_{i=1}^{\infty} \frac{\lambda_i^{2q} + n^{-1}i^{2p})^2|\theta_i^0|^2}{(1 + \lambda_i^{2q} + n^{-1}i^{2p})^2} - 2n^{-1/2} \sum_{i=1}^{\infty} \frac{\lambda_i^{2q} + n^{-1}i^{2p})\theta_i^0\epsilon_i}{(1 + \lambda_i^{2q} + n^{-1}i^{2p})^2}
\]

\[
+ n^{-1} \sum_{i=1}^{\infty} \frac{\epsilon_i^2}{(1 + \lambda_i^{2q} + n^{-1}i^{2p})^2}.
\]
Define the three terms on the right side of the above equation to be $T_1, T_2, T_3$. It follows immediately from Lemma S.1 that $E\{T_3\} = O(n^{-1+\alpha_1})$, implying that $T_3 = O_P(n^{-1+\alpha_1})$.

Proof of Part (1a). If $p - 1/2 \leq q_0 < 2q$ and $\alpha_0 \geq 2q/(2q_0 + 1)$, then we have $(2q_0 + 1)\alpha_1 \geq 1$ since both $1/(2p) \geq 1/(2q_0 + 1)$ and $\alpha_0/(2q) \geq 1/(2q_0 + 1)$ are true. Therefore, by (S.1), that the function $x \mapsto x/(1 + x)$ is increasing on $x \geq 0$, and dominated convergence theorem, we get that

\[
\begin{align*}
&\quad \quad n^{1-\alpha_1} T_1 \\
&= n^{1-\alpha_1} \sum_{i=1}^{\infty} \left( \frac{\lambda_i^{2q} + n^{-1} \alpha_i^{2p}}{1 + \lambda_i^{2q} + n^{-1} \alpha_i^{2p}} \right)^2 |\theta_i^0|^2 \\
&\leq n^{1-\alpha_1} \sum_{i=1}^{\infty} \left( \frac{(n^{-\alpha_1} \alpha_i)^{2q} + 2(n^{-\alpha_1} \alpha_i)^{2p}}{1 + (n^{-\alpha_1} \alpha_i)^{2q} + (n^{-\alpha_1} \alpha_i)^{2p}} \right)^2 |\theta_i^0|^2 \\
&= n^{1-\alpha_1 - 2\alpha_1 q_0} \sum_{i=1}^{\infty} (n^{-\alpha_1} \alpha_i)^{4q - 2q_0} \left( \frac{1 + (n^{-\alpha_1} \alpha_i)^{2(p-q)}}{1 + (n^{-\alpha_1} \alpha_i)^{2q} + (n^{-\alpha_1} \alpha_i)^{2p}} \right)^2 |\theta_i^0|^2 \lambda_i^{2q_0} \\
&= o(n^{1-\alpha_1 - 2\alpha_1 q_0}).
\end{align*}
\]

We also note that

\[
E\{T_2^2\} = 4n^{-1} \sum_{i=1}^{\infty} \left( \frac{\lambda_i^{2q} + n^{-1} \alpha_i^{2p}}{1 + \lambda_i^{2q} + n^{-1} \alpha_i^{2p}} \right)^2 |\theta_i^0|^2 \leq 4n^{-1} T_1,
\]

therefore, $T_2 = o_P(n^{-1/2} T_1^{1/2})$. Since $1/2 - \alpha_1 - \alpha_1 q_0 < (1 - \alpha_1 - 2\alpha_1 q_0)/2 \leq 0$, we get $n^{1-\alpha_1} T_2 = o_P(n^{1/2 - \alpha_1 - \alpha_1 q_0})$. To handle $T_3$, note that

\[
\begin{align*}
n^{1-\alpha_1} T_3 &= n^{-\alpha_1} \sum_{i=1}^{\infty} \left( \frac{\epsilon_i^2 - 1}{1 + \lambda_i^{2q} + n^{-1} \alpha_i^{2p}} \right)^2 + \sum_{i=1}^{\infty} \frac{n^{-\alpha_1}}{(1 + \lambda_i^{2q} + n^{-1} \alpha_i^{2p})^2},
\end{align*}
\]

and by Lemma S.1,

\[
E\left\{ \sum_{i=1}^{\infty} \left( \frac{\epsilon_i^2 - 1}{1 + \lambda_i^{2q} + n^{-1} \alpha_i^{2p}} \right)^2 \right\} \leq \sum_{i=1}^{\infty} \frac{2}{(1 + \lambda_i^{2q} + n^{-1} \alpha_i^{2p})^4} = O(n^{\alpha_1}),
\]

which implies $n^{1-\alpha_1} T_3 = O_P(n^{-\alpha_1/2}) + \sum_{i=1}^{\infty} \frac{n^{-\alpha_1}}{(1 + \lambda_i^{2q} + n^{-1} \alpha_i^{2p})^2}$. Combining the above, we get that

\[
n^{1-\alpha_1} \|\theta_0 - \hat{\theta}_n\|_2^2 = n^{1-\alpha_1} (T_1 + T_2 + T_3) = \sum_{i=1}^{\infty} \frac{n^{-\alpha_1}}{(1 + \lambda_i^{2q} + n^{-1} \alpha_i^{2p})^2} + o_P(1).
\]
Since the leading term in the above equation is asymptotically strictly smaller than $E_{*n}$, we have, with probability approaching one, $\theta_0 \in C_{*n}(\alpha)$.

**Proof of Part (1b).** If $q \leq q_0 < p - 1/2$ or $0 < \alpha_0 < 2q/(2q_0 + 1)$, then $\alpha_1 < 1/(2q_0 + 1)$ since either $1/(2p) < 1/(2q_0 + 1)$ or $\alpha_0/(2q) < 1/(2q_0 + 1)$. This implies that $1/\alpha_1 - 2q_0 > 1$. Since $4q - 2q_0 > 0$, we can choose an arbitrary real number $t$ such that $1 < t < \min\{2, 1/\alpha_1 - 2q_0, 4q - 2q_0 + 1\}$. Let $|\theta_i^0|^2 = i^{-2q_0-t}$ for $i = 1, 2, \ldots$, so $\theta_0 \in \ell_2(q_0)$. Observe that $\lambda_i^{2q} + n^{-1}i^{2p} \geq (n^{-\alpha_1} i^{2q})$ for $i \geq n^{\alpha_1}$ which leads to

$$
n^{1-\alpha_1}T_1 \geq n^{1-\alpha_1} \sum_{i \geq n^{\alpha_1}} \left( \frac{(n^{-\alpha_1} i^{2q})^2}{1 + (n^{-\alpha_1} i^{2q})^2} \right) i^{-2q_0-t} = n^{1-\alpha_1}(2q_0 + t) \cdot n^{-\alpha_1} \sum_{i \geq n^{\alpha_1}} \frac{(n^{-\alpha_1} i^{2q})^2}{1 + (n^{-\alpha_1} i^{2q})^2} \leq n^{1-\alpha_1}(2q_0 + t) \sum_{i = 1}^{\infty} \int_1^{\infty} x^{-2q_0-t} \left( \frac{1}{1 + x^{2q}} \right)^2 dx.
$$

Note the integral on the right of the last equation is finite since $4q - 2q_0 - t > -1$ and $2q_0 + t > 1$.

On the other hand,

$$
E\{T_2^2\} = 4n^{-1} \sum_{i = 1}^{\infty} \frac{(\lambda_i^{2q} + n^{-1}i^{2p})^2}{1 + \lambda_i^{2q} + n^{-1}i^{2p})^4} i^{-2q_0-t} \leq 4n^{-1} \sum_{i = 1}^{\infty} \left( \frac{(n^{-\alpha_1} i^{2q})^2 + (n^{-\alpha_1} i^{2q})^2}{1 + (n^{-\alpha_1} i^{2q})^2 + (n^{-\alpha_1} i^{2q})^2} \right) i^{-2q_0-t} = 4n^{-1-\alpha_1}(2q_0 + t) + \alpha_1 \cdot n^{-\alpha_1} \sum_{i = 1}^{\infty} \frac{(n^{-\alpha_1} i^{2q})^2}{1 + (n^{-\alpha_1} i^{2q})^2 + (n^{-\alpha_1} i^{2q})^2} \int_0^{\infty} x^{4q-2q_0-t} \left( \frac{1}{1 + x^{2q} + x^{2p})^2} \right) dx,
$$

where the integral in the last equation finitely exists because $4q - 2q_0 - t > -1$, $p > q$ and $2q_0 + t > 1$. Therefore, $n^{1-\alpha_1}T_2 = o_P(n^{(1-\alpha_1 - \alpha_1(2q_0 + t))/2})$.

Since $\alpha_1 < 1/(2q_0 + 1)$ and $1 < t < 2$, it can be seen that $\alpha_1(2q_0 + t - 1) < 1$. This implies that $(1 - \alpha_1 - \alpha_1(2q_0 + t))/2 < 1 - \alpha_1(2q_0 + t)$, and hence, $n^{1-\alpha_1}T_2 = o_P(n^{1-\alpha_1}T_1)$.

The above analysis of $T_1, T_2$ leads to that

$$
n^{1-\alpha_1}||\hat{\theta}_* - \theta_0||_2^2 \geq n^{1-\alpha_1}T_1(1 + o_P(1)).$$
Since $\alpha_1(2q_0 + t) < 1$, as $n \to \infty$, $n^{1-\alpha_1}T_n \to \infty$. So, with probability approaching one, $\theta_0 \notin C_{sn}(\alpha)$, which completes the proof for $C_{sn}(\alpha)$.

The proof for $C_{\alpha n}^+(\alpha)$ is similar. So we only briefly sketch the proof.

**Proof of Part (2a).** Similar to the proof of Part (1a), it can be shown by using $4q > 2q_0 + 1$ that

$$n \sum_{i=1}^{\infty} d_i \left( \frac{\lambda_i 2q + n^{-1}i^2p}{1 + \lambda_i 2q + n^{-1}i^2p} \right)^2 = o(n^{1-\alpha_1(2q_0+1)}) = o(1)$$

and

$$\sqrt{n} \sum_{i=1}^{\infty} d_i \left( \frac{\lambda_i 2q + n^{-1}i^2p}{1 + \lambda_i 2q + n^{-1}i^2p} \right)^2 \theta_i^0 \epsilon_i = O_P(1).$$

And hence, $n\|\theta_0 - \hat{\theta}_s\|_T^2 = \sum_{i=1}^{\infty} \left( \frac{d_i \epsilon_i^2}{(1 + \lambda_i 2q + n^{-1}i^2p)^2} \right) + o_P(1) \to 1 - \alpha$.

**Proof of Part (2b).** If $p - 1/2 < q_0 < 2q$ and $\alpha_0 \geq 2q/(2q_0 + 1)$, then we have $\alpha_1(2q_0 + 1) < 1$; see proof of Part (1b). For any $t > 1$, let $|\theta_i^0|^2 = i^{-2q_0+1}(\log 2i)^{-t}$. It is easy to see that $\theta_0 \in \ell_2(q_0)$. Following the proof of Part (1b) and using $4q > 2q_0 + 1$, it can be shown that

$$n \sum_{i=1}^{\infty} d_i \left( \frac{\lambda_i 2q + n^{-1}i^2p}{1 + \lambda_i 2q + n^{-1}i^2p} \right)^2 |\theta_i^0|^2 \geq n^{1-\alpha_1(2q_0+1)}(\log n)^{-\tau-t},$$

$$\sqrt{n} \sum_{i=1}^{\infty} d_i \left( \frac{\lambda_i 2q + n^{-1}i^2p}{1 + \lambda_i 2q + n^{-1}i^2p} \right)^2 \theta_i^0 \epsilon_i = O_P(n^{(1-\alpha_1(2q_0+1))/2}).$$

Therefore,

$$n\|\theta_0 - \hat{\theta}_s\|_T^2 = n \sum_{i=1}^{\infty} d_i \left( \frac{\lambda_i 2q + n^{-1}i^2p}{1 + \lambda_i 2q + n^{-1}i^2p} \right)^2 |\theta_i^0|^2 - 2\sqrt{n} \sum_{i=1}^{\infty} d_i \left( \frac{\lambda_i 2q + n^{-1}i^2p}{1 + \lambda_i 2q + n^{-1}i^2p} \right)^2 \theta_i^0 \epsilon_i + \sum_{i=1}^{\infty} \left( \frac{d_i \epsilon_i^2}{1 + \lambda_i 2q + n^{-1}i^2p} \right)^2 \geq n^{1-\alpha_1(2q_0+1)}(\log n)^{-\tau-t}(1 + o_P(1)) + O_P(1).$$

The leading term approaches infinity which implies that, with probability approaching one, $\theta_0 \notin C_{\alpha n}^+(\alpha)$. \qed
S.3. Adaptive credible regions under priors P1 and P2. In this section, we construct a set of adaptive credible regions that do not rely on the unknown regularity \( q_0 \). The main idea is to replace \( q_0 \) by a proper estimator, namely \( \hat{q} \) with \( \hat{q} \geq 0 \) a.s. This is different from the existing adaptive framework for proving contraction rates; see [2]. For simplicity, we choose \( p = \hat{q} + 1 \) in Prior P1, although other choices satisfying \( p > \hat{q} + 1/2 \) also apply in our procedure. The corresponding posterior mode \( \hat{\theta} \) becomes \( \hat{\theta}_i = Y_i/(1 + \hat{\lambda} i^{2p}) \), where for simplicity we choose \( \hat{\lambda} = n^{-2p/(2\hat{q} + 1)} \). The \((1 - \alpha)\) adaptive credible sets (parallel to (2.3) and (2.4)) are stated as

\[
C_n^{\text{adapt}}(\alpha) = \left\{ \theta = \{\theta_i\}_{i=1}^\infty \in \ell_2 : \|\theta - \hat{\theta}\|_2 \leq \sqrt{\frac{\hat{E}_n + \sqrt{\hat{D}_n h \hat{\lambda}_n}}{nh}} \right\},
\]

where \( \hat{h} = n^{-1/(2\hat{q} + 1)} \) and \( \hat{D}_n, \hat{E}_n \) are defined similar as (2.2) by replacing \( h, \lambda \) and \( q \) with their estimators.

**Theorem S.1.** Suppose that \( Y_i \)'s are generated from model (2.1) with \( \theta_0 \in \ell_2(q_0) \) for some unknown \( q_0 \geq 0 \), and the Sobolev rectangle property holds: \( \sup_{i \geq 1} |\theta_i^{0}|^2 \hat{q} + 1 < \infty \). Let \( \check{q} \) be any nonnegative-valued estimator of \( q_0 \) satisfying \( \check{q} - q_0 = o_P(1/\log n) \), and set \( p = \check{q} + 1 \) in Prior P1. Then as \( n \to \infty \), \( P(\theta \in C_n^{\text{adapt}}(\alpha)|\{Y_i\}_{i=1}^\infty) \to 1 - \alpha \) and \( P(\theta \in C_n^{\text{adapt}}(\alpha)|\{Y_i\}_{i=1}^\infty) \to 1 - \alpha \) in probability. Furthermore, with probability approaching one, \( \theta_0 \in C_n^{\text{adapt}}(\alpha) \); with probability approaching \( 1 - \alpha \), \( \theta_0 \in C_n^{\text{adapt}}(\alpha) \).

**Proof of Theorem S.1.** Define \( h_0 = n^{-1/(2q_0 + 1)} \), \( p_0 = q_0 + 1 \), \( \lambda_0 = h_0^{2p_0} \). Let \( D_n^0, E_n^0, Z_n^0 \) be defined as (2.2) with \((h, \lambda, p)\) therein replaced by \((h_0, \lambda_0, p_0)\). To simplify the notation, in the rest of the proofs we still use \( h \) and \( \lambda \) to denote \( \hat{h} \) and \( \hat{\lambda} \) respectively, where recall that \( \hat{h} = n^{-1/(2\hat{q} + 1)} \) and \( \hat{\lambda} = n^{-2p/(2\hat{q} + 1)} \) are the empirical version of the tuning parameters. By (??) we have \( n\sqrt{\hat{h}}\|\theta - \hat{\theta}\|^2 = h^{-1/2}E_n + D_n^{1/2}Z_n \), where \( D_n, E_n, Z_n \) are defined as (2.2). It can be shown that

\[
D_n - D_n^0 = o_P(1), E_n - E_n^0 = o_P(1), Z_n - Z_n^0 = o_P(1).
\]

To see this, note that \( p = \check{q} + 1 \to p_0 \) in probability and \( h \to 0 \). Hence

\[
D_n \geq 2 \sum_{i=1}^{\infty} \int_i^{i+1} \frac{h}{(1 + (xh)^{2p})^2} dx = 2 \int_0^\infty \frac{1}{(1 + x^{2p})^2} dx \to 2 \int_0^\infty \frac{1}{(1 + x^{2p_0})^2} dx.
\]
Therefore, $D_n - D^0_n = o_P(1)$ holds. The proof for $E_n - E^0_n = o_P(1)$ is similar. The proof of $Z_n - Z^0_n = o_P(1)$ is a bit nontrivial. Observe that

\[
E\{|Z_n - Z^0_n|^2\} = E\{E\{|Z_n - Z^0_n|^2|q\}\} = 2E\{\sum_{i=1}^{\infty} \frac{\sqrt{h}}{1 + \lambda_i^2p} - \frac{\sqrt{h_0}}{1 + \lambda_0^2p_0}\}^2 \leq 2E\{\sum_{i=1}^{\infty} \frac{h}{(1 + \lambda_i^2p)^2}\} - 4E\{\sum_{i=1}^{\infty} \frac{\sqrt{hh_0}}{(1 + \lambda_i^2p)(1 + \lambda_0^2p_0)}\} + 2\sum_{i=1}^{\infty} \frac{h_0}{(1 + \lambda_0^2p_0)^2}.
\]

It is easy to see that

\[
E\{\int_{h_0}^{\infty} \frac{1}{(1 + x^2p)^2}dx\} \leq E\{\sum_{i=1}^{\infty} \frac{h}{(1 + \lambda_i^2p)^2}\} \leq E\{\int_{0}^{\infty} \frac{1}{(1 + x^2p)^2}dx\}.
\]

Since $p \geq 1$, we have $\int_{0}^{\infty} \frac{1}{(1 + x^2p)^2}dx \leq 1 + \int_{1}^{\infty} 1 \frac{1}{(1 + x^2)^2}dx$. Since $\int_{0}^{\infty} \frac{1}{(1 + x^2p)^2}dx \rightarrow \int_{0}^{\infty} \frac{1}{(1 + x^2p_0)^2}dx$ in probability, it follows by bounded convergence theorem that $E\{\int_{0}^{\infty} \frac{1}{(1 + x^2p)^2}dx\} \rightarrow \int_{0}^{\infty} \frac{1}{(1 + x^2p_0)^2}dx$. Since $h \leq 1$ and $h \rightarrow 0$ in probability, by bounded convergence theorem, $E\{\int_{0}^{h} \frac{1}{(1 + x^2p)^2}dx\} \leq E\{h\} \rightarrow 0$. Therefore, $E\{\sum_{i=1}^{\infty} \frac{h}{(1 + \lambda_i^2p)^2}\} \rightarrow \int_{0}^{\infty} \frac{1}{(1 + x^2p_0)^2}dx$.

We also note that

\[
E\{\sum_{i=1}^{\infty} \frac{\sqrt{hh_0}}{(1 + \lambda_i^2p)(1 + \lambda_0^2p_0)}\} \leq E\{\int_{h_0}^{\infty} \frac{1}{(1 + x^2p(h/h_0)^{2p})(1 + x^2p_0)}dx\},
\]

and

\[
E\{\sum_{i=1}^{\infty} \frac{\sqrt{hh_0}}{(1 + \lambda_i^2p)(1 + \lambda_0^2p_0)}\} \geq E\{\int_{h_0}^{\infty} \frac{1}{(1 + x^2p(h/h_0)^{2p})(1 + x^2p_0)}dx\}.
\]

It can be easily seen that

\[
E\{\int_{h_0}^{\infty} \frac{1}{(1 + x^2p(h/h_0)^{2p})(1 + x^2p_0)}dx\} \leq E\{\sqrt{hh_0}\} \rightarrow 0.
\]

Since $q - q_0 = o_P(1/\log n)$, it is easy to show that $h/h_0 \rightarrow 1$ in probability. Then it can be seen that $\sqrt{\frac{h}{h_0}} \int_{h_0}^{\infty} \frac{1}{(1 + x^2p(h/h_0)^{2p})(1 + x^2p_0)}dx \rightarrow \int_{0}^{\infty} \frac{1}{(1 + x^2p_0)^2}dx$
in probability, and
\[
\sqrt{\frac{h}{h_0}} \int_0^\infty \frac{1}{(1 + x^2(\frac{h}{h_0})^2)(1 + x^{2p_0})} dx
\]
\[
= \int_0^\infty \sqrt{\frac{h}{h_0}} \frac{h}{(1 + \lambda x^{2p})(1 + \lambda_0 x^{2p_0})} dx
\]
\[
\leq \frac{1}{2} \int_0^\infty \frac{h}{(1 + \lambda x^{2p})} dx + \frac{1}{2} \int_0^\infty \frac{h_0}{(1 + \lambda_0 x^{2p_0})^2} dx \leq \int_0^\infty \frac{1}{(1 + x^2)^2} dx, \text{ a.s.}
\]
It follows by bounded convergence theorem that
\[
E\{\sum_{i=1}^{\infty} \sqrt{\frac{h}{h_0}(1 + \lambda x^{2p_0})} \} \rightarrow \int_0^\infty \frac{1}{(1 + x^{2p_0})^2} dx.
\]
This shows that \(E\{|Z_n - Z_n^0|^2\} \rightarrow 0\), and hence, \(Z_n - Z_n^0 = o_P(1)\). Since \(Z_n \overset{d}{\rightarrow} N(0, 1)\), we have \(Z_n \overset{d}{\rightarrow} N(0, 1)\). Therefore,
\[
P(\theta \in C_n^{\text{adapt}}(\alpha)|\{Y_i\}_{i=1}^{\infty})
\]
\[
= P\left(n \sqrt{\frac{h}{h_0}} \|\theta - \hat{\theta}\|_2 \leq h^{-1/2} E_n + \sqrt{D_n z_\alpha}|\{Y_i\}_{i=1}^{\infty}\right)
\]
\[
= P(Z_n \leq z_\alpha) \rightarrow 1 - \alpha.
\]
To the end of the proof we show that \(P(\theta_0 \in C_n^{\text{adapt}}(\alpha)) \rightarrow 1\). Define \(\hat{\theta}^0 = \{\hat{\theta}^0_{ni}\}_{i=1}^{\infty}\) with \(\hat{\theta}^0_{ni} = Y_i/(1 + \lambda_0 i^{2p_0})\). Following the proof of Theorem \(2.1\) (1), \(nh_0\|\theta_0 - \hat{\theta}^0\|_2^2 = F_n^0 + o_P(1)\), where \(F_n^0 = \sum_{i=1}^{\infty} \frac{h_0}{(1 + \lambda_0 i^{2p_0})^2}\). Since \(h/h_0 = 1 + o_P(1)\), we have \(nh\|\theta_0 - \hat{\theta}^0\|_2^2 = F_n^0 + o_P(1)\). On the other hand, for any \(x > 1\) the inequality \(|x^\alpha - 1| \leq (\log x)|x^{[\alpha]}|\) holds for any \(\alpha \in \mathbb{R}\).
Then we have

\[
\begin{align*}
  \|\hat{\theta}^0 - \hat{\theta}\|_2^2 &= \sum_{i=1}^{\infty} \mathbf{Y}_i^2 \left( \frac{1}{1 + \lambda_0 i^{2 p_0}} - \frac{1}{1 + \lambda_i^{2 p}} \right)^2 \\
  &= \sum_{i=1}^{\infty} \mathbf{Y}_i^2 \lambda_0^{2 p_0} \left( \frac{\lambda_i^{-2(p-p_0)} - 1}{(1 + \lambda_0 i^{2 p_0})^2(1 + \lambda_i^{2 p})^2} \right)^2 \\
  &= \sum_{i=1}^{\infty} \mathbf{Y}_i^2 \lambda_0^{2 p_0} \left( \frac{1}{(1 + \lambda_0 i^{2 p_0})^2(1 + \lambda_i^{2 p})^2} \right)^2 \\
  &\leq 4|\hat{\theta} - q_0|^2 \sum_{i=1}^{\infty} \mathbf{Y}_i^2 \lambda_0^{2 p_0} \left( \frac{(2q^{1+1})^{2i} \log n + \log (i\theta_0)^2(i\theta_0)^{4|\hat{\theta} - q_0|}}{(1 + \lambda_0 i^{2 p_0})^2(1 + \lambda_i^{2 p})^2} \right)^2 \\
  &= 4|\hat{\theta} - q_0|^2 n^{2q^{1+1}2^{2i+1}} \sum_{i=1}^{\infty} \mathbf{Y}_i^2 \lambda_0^{2 p_0} \left( \frac{(2q^{1+1})^{2i} \log n + \log (i\theta_0)^2(i\theta_0)^{4p_0+4|\hat{\theta} - q_0|}}{(1 + \lambda_0 i^{2 p_0})^2(1 + \lambda_i^{2 p})^2} \right)^2 \\
  \leq 8|\hat{\theta} - q_0|^2 n^{2q^{1+1}2^{2i+1}} \sum_{i=1}^{\infty} \mathbf{Y}_i^2 \lambda_0^{2 p_0} \left( \frac{(2q^{1+1})^{2i} \log n + \log (i\theta_0)^2(i\theta_0)^{4p_0+4|\hat{\theta} - q_0|}}{(1 + \lambda_0 i^{2 p_0})^2(1 + \lambda_i^{2 p})^2} \right)^2 \\
  + 8|\hat{\theta} - q_0|^2 n^{2q^{1+1}2^{2i+1}} - \sum_{i=1}^{\infty} \mathbf{Y}_i^2 \lambda_0^{2 p_0} \left( \frac{(2q^{1+1})^{2i} \log n + \log (i\theta_0)^2(i\theta_0)^{4p_0+4|\hat{\theta} - q_0|}}{(1 + \lambda_0 i^{2 p_0})^2(1 + \lambda_i^{2 p})^2} \right)^2.
\end{align*}
\]

Denote the two terms on the right side of the last inequality to be $T_1, T_2$.

Since $|\hat{\theta} - q_0| = \mathcal{O}_P(1/\log n)$, with probability approaching one, $|\hat{\theta} - q_0| \leq 1/\log n$. In what follows we assume $|\hat{\theta} - q_0| \leq 1/\log n$. To address $T_1, T_2$, we introduce an elementary inequality: for any $x > 0$,

\[(S.4) \quad \frac{x^{4p_0+4|\hat{\theta} - q_0|}}{(1 + x^{2 p})^2} \leq 1 + x^{8/\log n}.
\]

We briefly show (S.4). If $0 < x \leq 1$, then the left hand side of (S.4) is always less than $1$, so the inequality trivially holds. If $x > 1$, then

\[x^{4p_0+4|\hat{\theta} - q_0|} \leq x^{4|\hat{\theta} - q_0|+4(p_0-p)} \leq x^{8|\hat{\theta} - q_0|} < 1 + x^{8/\log n}.
\]

Using (S.4) and $n^{(2q^{1+1}2^{2i+1})} \leq \exp\left(\frac{8p}{(2q^{1+1}2^{2i+1})}\right) = \mathcal{O}_P(1)$, we have

\[T_2 = \mathcal{O}_P(1) \cdot |\hat{\theta} - q_0|^2 n^{-1} \sum_{i=1}^{\infty} \mathbf{Y}_i^2 \left( \frac{(2q^{1+1})^{2i} \log n + \log (i\theta_0)^2(i\theta_0)^{4p_0+4|\hat{\theta} - q_0|}}{(1 + (i\theta_0)^{2p_0})^2} \right)^2.
\]
Since
\begin{align*}
E\left\{ \sum_{i=1}^{\infty} \frac{2(\log n + \log(ih_0))^2((1 + (ih_0)^{8/\log n}))}{(1 + (ih_0)^{2p_0})^2} \right\} \\
= \sum_{i=1}^{\infty} \frac{(\log n + \log(ih_0))^2((1 + (ih_0)^{8/\log n}))}{(1 + (ih_0)^{2p_0})^2} \\
\leq h_0^{-1} \int_0^{\infty} \frac{(\log n + \log x)^2((1 + x^{8/\log n}))}{(1 + x^{2p_0})^2} dx \leq h_0^{-1}(\log n)^2.
\end{align*}

Therefore,
\[ T_2 = O_P((nh_0)^{-1}(\log n)^2|\widehat{q} - q_0|^2) = o_P((nh_0)^{-1}).\]

For $T_1$, using condition $\sup_{i \geq 1} |\theta^0_i|^2 < 1$ and (S.4) we have
\begin{align*}
T_1 &= O_P(1)|\widehat{q} - q_0|^2h_0^{2p_0+1} \sum_{i=1}^{\infty} \frac{(\log n + \log(ih_0))^2((ih_0)^{4p_0+4|\widehat{q} - q_0| - 2p_0 - 1})}{(1 + (ih_0)^{2p_0})^2(1 + (ih_0)^{2p_0})^2} \\
&= O_P(1)|\widehat{q} - q_0|^2h_0^{2p_0+1}h_0^{-1} \int_0^{\infty} \frac{(\log n + \log x)^2 x^{4p_0+4|\widehat{q} - q_0| - 2p_0 - 1}}{(1 + x^{2p_0})^2(1 + x^{2p_0})^2} dx \\
&= O_P((\log n)^2|\widehat{q} - q_0|^2h_0^{2p_0}) = o_P((nh_0)^{-1}).
\end{align*}

Consequently, $\|\theta^0 - \widehat{\theta}\|_1 = T_1 + T_2 = o_P((nh_0)^{-1}) = o_P((nh)^{-1})$. So we have
\[ nh\|\theta_0 - \widehat{\theta}\|_1 = nh\|\theta_0 - \widehat{\theta}\|_1^2 + o_P(1) = E_n^0 + o_P(1).\]

Since $E_n - E_n^0 = o_P(1)$ and $\lim inf_{n \to \infty}(E_n^0 - E_n^0) > 0$, with probability approaching one, $\theta_0 \in C_{n\text{adapt}}(\alpha)$.

To the end, we prove the results for $C_{n\text{adapt}}(\alpha)$. Since
\[ n\|\theta - \widehat{\theta}\|_1^2 = \sum_{i=1}^{\infty} \frac{d_i\eta_i^2}{1 + \lambda_i^2p} \to \sum_{i=1}^{\infty} d_i\eta_i^2, \]

it is easy to see that $C_{n\text{adapt}}(\alpha)$ possesses asymptotically 1 $- \alpha$ posterior coverage.

Let $\theta^0$ be defined the same as in the proof of Theorem S.1. It follows by Theorem 2.2 that $P(n\|\theta_0 - \widehat{\theta}\|_1^2 < c_\alpha) \to 1 - \alpha$. Following the proof of Theorem S.1, it can be shown that
\[ n\|\theta^0 - \widehat{\theta}\|_1^2 = O_P(\widehat{q} - q_0) = o_P(1).\]

Hence, $P(n\|\theta_0 - \widehat{\theta}\|_1^2 < c_\alpha) \to 1 - \alpha$. This completes the proof of Theorem S.1. \qed
A crucial assumption in Theorem S.1 is the existence of \( \hat{q} \) that satisfies \( \hat{q} - q_0 = o_P(1/\log n) \). Due to this proper rate, the contraction rate of \( C_n^{\text{adapt}}(\alpha) \) is of the order \( n^{-\hat{q}/(2\hat{q}+1)} \), which is asymptotically equivalent to the theoretical optimal one \( n^{-q_0/(2q_0+1)} \) under any regularity \( q_0 > 1/2 \). We next discuss the construction of a desirable \( \hat{q} \) by starting from a simple example. If it is known that the true parameter satisfies \( |\theta_i^0|^2 = (i \log i)^{-2q_0+1} \) for \( i = 2, 3, \ldots \), then we assign a prior on \( \theta_i \) as

\[
\theta_i \sim N(0, (i \log i)^{-2q_0+1}), \quad i = 2, 3, \ldots,
\]

such that the marginal distribution of \( Y_i \sim N(0, n^{-1} + (i \log i)^{-2q_0+1}) \). Let \( \phi_0 = 2q_0 + 1 \). Based on \( \{Y_i\}_{i=2}^n \), the log-likelihood function of \( \phi = 2q + 1 \) is

\[
l_n(\phi) = -\frac{1}{2} \sum_{i=2}^n \left[ \log(n^{-1} + (i \log i)^{-\phi}) + \frac{Y_i^2}{n^{-1} + (i \log i)^{-\phi}} \right].
\]

**Proposition S.1.** (Existence of \( \hat{q} \) with a desirable rate of convergence)

There exists a sequence of local maxima, denoted as \( \hat{\phi} \), of the log-likelihood function \( l_n(\phi) \) satisfying \( \hat{\phi} - \phi_0 = O_P(n^{-\frac{1}{2q_0}}(\log n)^{2q_0}) \). If we let \( \hat{q} = (\hat{\phi} - 1)_+ / 2 \), where \((a)_+\) is the positive part of a real number \( a \), then \( \hat{q} \) satisfies \( \hat{q} - q_0 = O_P(n^{-\frac{1}{2q_0}}(\log n)^{2q_0}) = o_P(1/\log n) \).

**Remark S.1.** It is possible to extend Proposition S.1 to more general situations. For example, given that \( |\theta_i^0|^2 = b_0 i^{-(2q_0+1)}(\log i)^{-a_0} \) for \( i = 2, 3, \ldots \) and unknown \( b_0 > 0, q_0 > 0, a_0 > 1 \), then one may assign priors

\[
\theta_i \sim N(0, b_0 i^{-(2q_0+1)}(\log i)^{-a_0}), \quad i = 2, 3, \ldots,
\]

which leads to log-likelihood function of \( (b, \phi, a) \) as

\[
l_n(b, \phi, a) = -\frac{1}{2} \sum_{i=2}^n \left[ \log(n^{-1} + bi^{-\phi}(\log i)^{-a}) + \frac{Y_i^2}{n^{-1} + bi^{-\phi}(\log i)^{-a}} \right].
\]

It can be shown by a multivariate extension of the proof of Proposition S.1 that there exists a sequence of local maxima of \( l_n(b, \phi, a) \), namely, \( (\hat{b}, \hat{\phi}, \hat{a}) \), such that the marginal entry \( \hat{\phi} \) satisfies \( \hat{\phi} - \phi_0 = o_P(1/\log n) \). Hence, by defining \( \hat{q} = (\hat{\phi} - 1)_+ / 2 \) we have \( \hat{q} - q_0 = o_P(1/\log n) \). The rate for \( (\hat{b}, \hat{a}) \) can also be derived but is of less interest.

**Proof of Proposition S.1.** Let \( \phi_n = n^{-\frac{1}{2q_0}}(\log n)^{2q_0} \). We will show that with probability approaching one,

\[
l_n(\phi_0 \pm \phi_n) < l_n(\phi_0),
\]
which implies that there exists a local maxima \( \hat{\phi} \) of \( l_n(\phi) \) s.t. \( \phi_0 - \phi_n < \hat{\phi} < \phi_0 + \phi_n \). We only show \( l_n(\phi_0 + \phi_n) < l_n(\phi_0) \) with probability approaching one. The proof for \( l_n(\phi_0 - \phi_n) < l_n(\phi_0) \) is exactly the same.

It is easy to show that the first-, second- and third-order derivatives of \( l_n(\phi) \) are

\[
\begin{align*}
\dot{l}_n(\phi) &= \frac{1}{2} \sum_{i=2}^{n} \left[ \frac{\log c_i}{1 + n^{-1}c_i^\phi} - \frac{Y^2_i(\log c_i)c_i^\phi}{(1 + n^{-1}c_i^\phi)^2} \right], \\
\ddot{l}_n(\phi) &= -\frac{1}{2} \sum_{i=2}^{n} \left[ \frac{(\log c_i)^2n^{-1}c_i^\phi}{(1 + n^{-1}c_i^\phi)^2} + \frac{Y^2_i(\log c_i)c_i^\phi(1 - n^{-1}c_i^\phi)}{(1 + n^{-1}c_i^\phi)^3} \right], \\
\dddot{l}_n(\phi) &= -\frac{1}{2} \sum_{i=2}^{n} \left[ \frac{(\log c_i)^3n^{-1}c_i^\phi(1 - n^{-1}c_i^\phi)}{(1 + n^{-1}c_i^\phi)^3} + \frac{(\log c_i)^2Y^2_i c_i^\phi(1 - 4n^{-1}c_i^\phi + n^{-2}c_i^{2\phi})}{(1 + n^{-1}c_i^\phi)^4} \right],
\end{align*}
\]

where \( c_i = i \log i \). By Taylor’s expansion,

\[
l_n(\phi_0 + \phi_n) - l_n(\phi_0) = \dot{l}_n(\phi_0)\phi_n + \frac{1}{2} \ddot{l}_n(\phi_0)\phi_n^2 + \frac{1}{6} \dddot{l}_n(\phi_0)\phi_n^3,
\]

for a random \( \phi_* \in [\phi_0, \phi_0 + \phi_n] \).

Noting that \( Y_i \sim N(\theta_0^i, 1/n) \), we have \( E\{Y_i^2 - c_i^{-\phi_0} - 1/n^2\} = 4c_i^{-\phi_0}/n + 2/n^2 \). By direct calculations it can be shown that

\[
\begin{align*}
E\{|\dot{l}_n(\phi_0)|^2\} &= \frac{1}{4} \sum_{i=2}^{n} c_i^{2\phi_0}(\log c_i)^2 E\{|Y_i^2 - c_i^{-\phi_0} - 1/n^2|\} \\
&= \frac{1}{4} \sum_{i=2}^{n} c_i^{2\phi_0}(\log c_i)^2 \left( \frac{4c_i^{-\phi_0}/n + 2/n^2}{(1 + n^{-1}c_i^{\phi_0})^4} \right) \\
&\leq \sum_{i=2}^{n} \frac{n^{-1}c_i^{\phi_0}(\log c_i)^2}{(1 + n^{-1}c_i^{\phi_0})^3} \leq \sum_{i=2}^{n} \frac{(\log i)^2}{(1 + n^{-1}c_i^{\phi_0})^2} \times n \frac{1}{\phi_0} (\log n)^2 \int_0^\infty \frac{1}{(1 + x^{\phi_0})^2} dx,
\end{align*}
\]

\[
\begin{align*}
-E\{\dddot{l}_n(\phi_0)\} &= \frac{1}{2} \sum_{i=2}^{n} \frac{(\log c_i)^2}{(1 + n^{-1}c_i^{\phi_0})^2} \lesssim \frac{1}{2} \sum_{i=2}^{n} \frac{(\log i)^2}{(1 + n^{-1}c_i^{\phi_0})^2} \\
&\lesssim \frac{1}{\phi_0} (\log n)^2 \int_0^\infty \frac{1}{(1 + x^{\phi_0})^2} dx,
\end{align*}
\]

\[
\begin{align*}
-E\{\dddot{l}_n(\phi_0)\} \gtrsim \frac{1}{2} \sum_{i=2}^{n} \frac{(\log i)^2(1 - \phi_0)}{(1 + n^{-1}c_i^{\phi_0})^2} \gtrsim (\log n)^2(1 - \phi_0) n \frac{1}{\phi_0} \int_0^\infty \frac{1}{(1 + x^{\phi_0})^2} dx,
\end{align*}
\]
therefore, with probability approaching one, 

\[
\lim_{n \to \infty} \mathbb{P} \left( \left| \hat{\phi}_n - \phi \right| \leq \varepsilon \right) = 1 - \frac{1}{n^\alpha}
\]

where in the last equation we have used the fact \( n^\alpha \to \infty \) uniformly for any \( \phi \). Similarly, assume that \( \hat{q} \) is a proper estimator of \( q_0 \) and let \( p = \hat{q} + 1/2 \).
The contraction rate of $\hat{C}$ counterparts. Again, Lemma S.1 gives the limiting values of $\hat{S}$ and map from $H$ satisfies

$$f \rightarrow \sum_{i} f_{i} m_{i} \rho_{i} \varphi_{i} < \infty.$$ 

Thus, for $q \rightarrow \infty$, $P(\theta \in C_{\alpha}^{\text{adapt}}(\alpha)|\{Y_{i}\}_{i=1}^{\infty}) \rightarrow 1 - \alpha$ and $P(\theta \in C_{\alpha}^{\text{adapt}}(\alpha)|\{Y_{i}\}_{i=1}^{\infty}) \rightarrow 1 - \alpha$ in probability. Furthermore, with probability approaching one, $\theta_{0} \in C_{\alpha}^{\text{adapt}}(\alpha)$; with probability approaching $1 - \alpha$, $\theta_{0} \in C_{\alpha}^{\text{adapt}}(\alpha)$.

Proof of Lemma 3.2. For any $f \in H_{m}(\mathbb{I})$, by Assumption A2, $f$ admits the unique series representation $f = \sum_{\nu=1}^{\infty} f_{\nu} \varphi_{\nu}$, where $f_{\nu} = V(f_{\nu} \varphi_{\nu})$ satisfies $\sum_{\nu} f_{\nu}^{2} \rho_{\nu} < \infty$. Therefore, $T : f \mapsto \{f_{\nu} : \nu \geq 1\}$ defines a one-to-one map from $H_{m}(\mathbb{I})$ to $\mathbb{R}_{m} \equiv \{f_{\nu} : \nu \geq 1\} \subseteq \mathbb{R}^{\infty} : \sum_{\nu=1}^{\infty} f_{\nu}^{2} \rho_{\nu} < \infty$.

Let $\Pi_{\lambda}$ and $\Pi$ be the probability measures induced by $\{w_{\nu} : \nu > m\}$ and $\{v_{\nu} : \nu > m\}$, respectively, which are both defined on $\mathbb{R}^{\infty}$. That is, for any subset $S \subseteq \mathbb{R}^{\infty}$, $\Pi(S) = \mathbb{P}(\{w_{\nu} : \nu > m\} \in S)$ and $\Pi(S) = \mathbb{P}(\{v_{\nu} : \nu > m\} \in S)$. Likewise, let $\Pi'_{\lambda}$ and $\Pi'$ be probability measures induced by $\{w_{\nu} : \nu \geq 1\}$ and $\{v_{\nu} : \nu \geq 1\}$. It is easy to see that, for any measurable $B \subseteq \mathbb{R}_{m}$,

$$\Pi_{\lambda}(T^{-1}B) = P(G_{\lambda} \in T^{-1}B) = P(\{w_{\nu} : \nu \geq 1\} \in B) = \Pi'_{\lambda}(B),$$

and

$$\Pi(T^{-1}B) = P(G \in T^{-1}B) = P(\{v_{\nu} : \nu \geq 1\} \in B) = \Pi'(B).$$

The following result can be found in Hájek [22].
PROPOSITION S.2. The Radon-Nikodym derivative of $\tilde{\Pi}_\lambda$ w.r.t. $\tilde{\Pi}$ is
\[
\frac{d\tilde{\Pi}_\lambda}{d\Pi}(\{f_\nu : \nu > m \}) = \prod_{\nu > m} \left(1 + n\lambda \rho_\nu^{-\beta/(2m)}\right)^{1/2} \exp\left(-\frac{n\lambda}{2} f_\nu^2 \rho_\nu\right)
\]
\[
= \prod_{\nu > m} \left(1 + n\lambda \rho_\nu^{-\beta/(2m)}\right)^{1/2} \cdot \exp\left(-\frac{n\lambda}{2} \sum_{\nu > m} f_\nu^2 \rho_\nu\right).
\]

Note that in Proposition S.2, $\prod_{\nu > m} \left(1 + n\lambda \rho_\nu^{-\beta/(2m)}\right)^{1/2}$ is convergent since $\rho_\nu^{-\beta/(2m)} \asymp \nu^{-\beta}$. Therefore, by Proposition S.2, we have
\[
d\Pi'_\lambda(\{f_\nu : \nu \geq 1 \})
\]
\[
= d\pi_1(f_1) \cdots d\pi_m(f_m) d\tilde{\Pi}_\lambda(\{f_\nu : \nu > m \})
\]
\[
= d\pi_1(f_1) \cdots d\pi_m(f_m) \frac{d\tilde{\Pi}_\lambda}{d\Pi}(\{f_\nu : \nu > m \}) \cdot d\Pi(\{f_\nu : \nu > m \})
\]
\[
= d\pi_1(f_1) \cdots d\pi_m(f_m) \prod_{\nu = m+1}^\infty \left(1 + n\lambda \rho_\nu^{-\beta/(2m)}\right)^{1/2}
\]
\[
\cdot \exp\left(-\frac{n\lambda}{2} \sum_{\nu = m+1}^\infty f_\nu^2 \rho_\nu\right) d\Pi(\{f_\nu : \nu > m \})
\]
\[
= \prod_{\nu = m+1}^\infty \left(1 + n\lambda \rho_\nu^{-\beta/(2m)}\right)^{1/2} \cdot \exp\left(-\frac{n\lambda}{2} \sum_{\nu = m+1}^\infty f_\nu^2 \rho_\nu\right) d\Pi'(\{f_\nu : \nu \geq 1 \}).
\]
Then for any measurable $S \subseteq H^m(\mathbb{I})$, by change of variable, we have

$$
\Pi_\lambda(S) = \Pi'_\lambda(TS) = \int_{TS} d\Pi'_\lambda(\{f_\nu : \nu \geq 1\})
= \prod_{\nu=m+1}^{\infty} \left( 1 + n\lambda \rho_\nu^{-\beta/(2m)} \right)^{1/2}
\cdot \int_{TS} \exp \left( -\frac{n\lambda}{2} \sum_{\nu=m+1}^{\infty} f_\nu^2 \rho_\nu \right) d\Pi'(\{f_\nu : \nu \geq 1\})
= \prod_{\nu=m+1}^{\infty} \left( 1 + n\lambda \rho_\nu^{-\beta/(2m)} \right)^{1/2}
\cdot \int_{TS} \exp \left( -\frac{n\lambda}{2} J(T^{-1}(\{f_\nu : \nu \geq 1\}), T^{-1}(\{f_\nu : \nu \geq 1\})) \right)
\cdot d(\Pi \circ T^{-1})(\{f_\nu : \nu \geq 1\})
= \prod_{\nu=m+1}^{\infty} \left( 1 + n\lambda \rho_\nu^{-\beta/(2m)} \right)^{1/2} \int_S \exp \left( -\frac{n\lambda}{2} J(f, f) \right) d\Pi(f).
$$

This completes the proof of the lemma.

\[ \square \]

S.5. Proofs in Section 4. Before proving Proposition 4.1, we present a preliminary lemma.

Let $\{\tilde{\psi}_\nu : \nu \geq 1\}$ be a bounded orthonormal basis of $L^2(\mathbb{I})$ under usual $L^2$ inner product. For any $b \in [0, \beta]$, define

$$
\tilde{H}_b = \left\{ \sum_{\nu=1}^{\infty} f_\nu \tilde{\psi}_\nu : \sum_{\nu=1}^{\infty} f_\nu^2 \rho_\nu^{1+b/(2m)} < \infty \right\}.
$$

Then $\tilde{H}_b$ can be viewed as a version of Sobolev space with regularity $m + b/2$. Define $\tilde{G} = \sum_{\nu=1}^{\infty} \gamma_\nu \tilde{\psi}_\nu$, a centered GP, and $\tilde{f}_0 = \sum_{\nu=1}^{\infty} f_\nu^0 \tilde{\psi}_\nu$. Define $\tilde{V}(f, g) = \langle f, g \rangle_{L^2} = \int_0^1 f(x)g(x)dx$, the usual $L^2$ inner product, $\tilde{J}(f) = \sum_{\nu=1}^{\infty} |\tilde{V}(f, \tilde{\psi}_\nu)|^2 \rho_\nu$, a functional on $\tilde{H}_0$. For simplicity, we denote $\tilde{V}(f) = \tilde{V}(f, f)$. Clearly, $\tilde{f}_0 \in \tilde{H}_\beta$. Since $\tilde{G}$ is a Gaussian process with covariance function

$$
\tilde{R}(s, t) = E\{\tilde{G}(s)\tilde{G}(t)\} = \sum_{\nu=1}^{m} \sigma_\nu^2 \tilde{\psi}_\nu(s)\tilde{\psi}_\nu(t) + \sum_{\nu>m} \rho_\nu^{-\beta/(2m)} \tilde{\psi}_\nu(s)\tilde{\psi}_\nu(t),
$$

This completes the proof of the lemma.

\[ \square \]
it follows by [44] that \( \tilde{H}_\beta \) is the RKHS of \( \tilde{G} \). For any \( \tilde{H}_b \) with \( 0 \leq b \leq \beta \), define inner product

\[
\langle \sum_{\nu=1}^{\infty} f_\nu \tilde{\varphi}_\nu, \sum_{\nu=1}^{\infty} g_\nu \tilde{\varphi}_\nu \rangle_b = \sum_{\nu=1}^{m} \sigma_\nu^{-2} f_\nu g_\nu + \sum_{\nu>m} f_\nu g_\nu \rho_\nu^{1+b/2m}.
\]

Let \( \| \cdot \|_b \) be the norm corresponding to the above inner product. The following lemma plays a similar role as quantifying the concentration function defined in [45]. The calculation of the concentration function in [44] depends on convolutions (see their proof of Theorem 4.1), while the proof of Lemma S.3 depends on a series representation.

**Lemma S.3.** Let \( d_n \) be any positive sequence. If Condition (S) holds, then there exists \( \omega \in \tilde{H}_\beta \) such that

(i). \( \tilde{V}(\omega - \tilde{f}_0) + \lambda \tilde{J}(\omega - \tilde{f}_0) \leq \frac{1}{16} \left( d_n^2 + \lambda d_n^{2m+\beta-1} \right) \), and

(ii). \( \| \omega \|_\beta^2 = O \left( d_n^{2m+\beta-1} \right) \).

**Proof of Lemma S.3.** Let \( \omega = \sum_{\nu=1}^{\infty} \omega_\nu \tilde{\varphi}_\nu \), where \( \omega_\nu = \frac{d_0^\nu}{d + (\sigma b_\nu)^{\nu/2}} \), \( \sigma = d_n^{2/(2m+\beta-1)} \), \( b_\nu = \rho_\nu^{1/(2m)} \), \( \alpha = m + (\beta - 1)/2 \), and \( d > 0 \) is a constant to be described. It is easy to see that for any \( \nu \), \( f_\nu^0 - \omega_\nu = \frac{(\sigma b_\nu)^\nu f_\nu^0}{d + (\sigma b_\nu)^\nu} \). Then

\[
\tilde{V}(\omega - \tilde{f}_0) = \sum_{\nu=1}^{\infty} (f_\nu^0 - \omega_\nu)^2 = \sum_{\nu=1}^{\infty} \frac{|f_\nu^0|^2 (\sigma b_\nu)^{2\alpha}}{(d + (\sigma b_\nu)^\alpha)^2} \leq \sigma^{2m+\beta-1} d^{-2} \sum_{\nu=1}^{\infty} |f_\nu^0|^2 \rho_\nu^{1+b/2m},
\]

and

\[
\tilde{J}(\omega - \tilde{f}_0) = \sum_{\nu=1}^{\infty} (f_\nu^0 - \omega_\nu)^2 \rho_\nu \leq \sigma^{\beta-1} \sum_{\nu=1}^{\infty} |f_\nu^0|^2 \rho_\nu^{1+b/2m} (d(\sigma b_\nu)^{-m} + (\sigma b_\nu)^{(\beta-1)/2})^{-2} \leq d^{-\frac{\beta-1}{k}} \left( \frac{2m}{\beta - 1} \right)^{-m} \left( \frac{2m}{\beta - 1} \right)^{\frac{\beta-1}{2k}} \sigma^{\beta-1} \sum_{\nu=1}^{\infty} |f_\nu^0|^2 \rho_\nu^{1+b/2m},
\]

where in the last equation \( k = m + (\beta - 1)/2 \). Therefore, we choose \( d \) as a suitably large fixed constant such that (i) holds.
To show (ii), observe that
\[
\|\omega\|_2^2 = \sum_{\nu=1}^{m} \sigma_{\nu}^{-2} \omega_{\nu}^2 + \sum_{\nu > m} \omega_{\nu}^2 \rho_{\nu}^{1+\frac{\beta}{2m}} = \sum_{\nu=1}^{m} \sigma_{\nu}^{-2} |f_{\nu}^0|^2 + \sum_{\nu > m} \frac{d^2 |f_{\nu}^0|^2 \rho_{\nu}^{1+\frac{\beta}{2m}}}{(d + (\sigma b_{\nu})^2)^2} b_{\nu} = O(\sigma^{-1}).
\]
The result follows by \(\sigma = \frac{d^2}{2(m+\beta-1)}\).

\[\]
By Cauchy-Schwarz inequality and direct examinations, we have
\[ E_{f_0} \left( J_1^2 \right) = E_{f_0} \left( \int_{B_n} \sum_{i=1}^{n} \left| R_i(f, f_0) - E_{f_0} \{ R_i(f, f_0) \} \right| d\Pi^* (f) \right)^2 \]
\[ \leq n \int_{B_n} E_{f_0} \{ R_i(f, f_0)^2 \} d\Pi^* (f). \]

For any \( f \in B_n \), by Taylor’s expansion, there exists a constant \( \tilde{C} > 0 \) s.t.
\[ E_{f_0} \{ R_i(f, f_0)^2 \} = E_{f_0} \{ (\ell(Y; f(X_i)) - \ell(Y; f_0(X_i)))^2 \} \]
\[ = E_{f_0} \{ (\tilde{\ell}_a(Y; f_0(X))) (f(X) - f_0(X)) + \tilde{\ell}_a(Y; f^*(X))(f(X) - f_0(X))^2 \} \]
\[ \leq 2 E_{f_0} \{ \varepsilon^2(X) \} + 2 E_{f_0} \{ E_{f_0} \{ \sup_{a \in \mathbb{R}} |\tilde{\ell}_a(Y; a)|^2 |X| (f(X) - f_0(X))^4 \} \}
\[ \leq \tilde{C} \delta_n^2, \]
where \( f^*(X) \) is between \( f(X) \) and \( f_0(X) \), and the last step is because \( E_{f_0} \{ \varepsilon^2 |X| \} = B(X) \) (Assumption A1(c)), \( E_{f_0} \{ \sup_{a \in \mathbb{R}} |\tilde{\ell}_a(Y; a)|^2 |X| \} \leq 2\tilde{C}^2 C_1 \), a.s., (Assumption A1 (a)). On the other hand, for any \( f \in B_n \), by Taylor’s expansion and Assumption A1(a),
\[ |E_{f_0} \{ R_i(f, f_0) \}| \leq |E_{f_0} \{ \tilde{\ell}_a(Y; f^*(X))(f(X) - f_0(X))^2 \}| \]
\[ \leq E_{f_0} \{ E_{f_0} \{ \sup_{a \in \mathbb{R}} |\tilde{\ell}_a(Y; a)| |X| (f(X) - f_0(X))^2 \} \}
\[ \leq C_0 C_1 \delta_n^2, \]
where \( f^*(X) \) is between \( f(X) \) and \( f_0(X) \). Therefore, \( J_2 \geq -C_0 C_1 n\delta_n^2 \), and by Cauchy-Schwarz inequality, \( P_{f_0} (J_1 < -\sqrt{n\delta_n^2}) \leq \tilde{C} / (n \delta_n^2) \to 0 \). So, by
\sqrt{n\delta_n^2} \leq n\delta_n^2$, we get that with probability approaching one, $I_1 \geq \exp(-(1 + C_0 C_1 n\delta_n^2 - n\lambda M_n/2)\Pi(B_n)$.

To proceed we provide a lower bound for $\Pi(B_n)$. Let $\omega \in \mathcal{H}_\beta$ satisfy (i)–(ii) of Lemma S.3 with $\delta_n$ therein replaced by $\delta_n$. It follows immediately from $\delta_n = o(1)$ and $\lambda \leq \delta_n^2$ that $\tilde{V}(\omega - \tilde{f_0}) + \lambda\tilde{J}(\omega - \tilde{f_0}) \leq \delta_n^2/4$. By Gaussian correlation inequality (see Theorem 1.1 of [30]), Cameron-Martin theorem (see [7]), and the fact $\lambda^{-1}\delta_n^2 > 1$, we have

$$\Pi(B_n) = P(V(G - f_0) \leq \delta_n^2, J(G) \leq M_n) = P(\tilde{V}(\tilde{G} - \tilde{f_0}) \leq \delta_n^2, \tilde{J}(\tilde{G}) \leq (1 + \gamma_0)^2\delta_n^2\lambda^{-1}) \geq P(\tilde{V}(\tilde{G} - \tilde{f_0}) \leq \delta_n^2, \lambda\tilde{J}(\tilde{G} - \tilde{f_0}) \leq \delta_n^2) \geq \exp(-\frac{1}{2}\|\omega\|^2_2)P(\tilde{V}(\tilde{G}) \leq \delta_n^2/4, \tilde{J}(\tilde{G}) \leq 1/4) \geq \exp(-\frac{1}{2}\|\omega\|^2_2)P(\tilde{V}(\tilde{G}) \leq \delta_n^2/8)P(\tilde{J}(\tilde{G}) \leq 1/8) \geq P(\tilde{J}(\tilde{G}) \leq 1/8) \exp(-c_1\delta_n^{-2/(2m+\beta-1)})$$

where the third last inequality follows by (4.18) of [28], and the last step follows by Lemma S.3 (ii) and Example 4.5 of [23] in which $c_1 > 0$ is a universal constant. Note in the last step $c_2 := P(\tilde{J}(\tilde{G}) \leq 1/8)$ is a universal positive constant. Since $\beta > 1$ and $\delta_n^2 \geq (nh)^{-1} + \lambda \geq n^{-2m/(2m+1)}$, we get

$$n\delta_n^{2m+\beta-1} \geq n\delta_n^{2m+\beta-1} > 1/8$$

so $n\delta_n^2 > n^{-2m+\beta-1}$. So

$$\Pi(B_n) \geq c_2 \exp(-c_1\delta_n^{-2/(2m+\beta-1)}) \geq c_2 \exp(-c_1n\delta_n^2).$$

Consequently, with $P_{f_0}$-probability approaching one, $I_1 \geq c_2 \exp(-c_3n\delta_n^2)$, where $c_3 = 1 + C_0 C_1 + c_1 + (1 + \gamma_0)^2/2$.

Let $I_2 = \int_{A_n} \prod_{i=1}^n (p_f/p_{f_0})(Z_i) \exp(-\frac{n\lambda}{2}J(f))d\Omega(f)$, where $A_n = \{ f \in H^m(\mathbb{I}) : \|f - f_0\| \geq 2M\delta_n, \|f - f_0\|_{sup} \leq \tilde{b} \}$, for some $M$ to be determined. Let $A_{n1} = \{ f \in H^m(\mathbb{I}) : V(f - f_0) \geq M^2\delta_n^2, \|f - f_0\|_{sup} \leq \tilde{b} \}$ and $A_{n2} = \{ f \in H^m(\mathbb{I}) : \lambda J(f - f_0) \geq M^2\delta_n^2, \|f - f_0\|_{sup} \leq b \}$. So $I_2 \leq I_2^1 + I_2^2$, where $I_2^1 = \int_{A_{n1}} \prod_{i=1}^n (p_f/p_{f_0})(Z_i) \exp(-\frac{n\lambda}{2}J(f))d\Omega(f)$ and $I_2^2 = \int_{A_{n2}} \prod_{i=1}^n (p_f/p_{f_0})(Z_i) \exp(-\frac{n\lambda}{2}J(f))d\Omega(f)$ for $j = 1, 2$.

Since for $f \in A_{n2}, M^2\delta_n^2 \leq \lambda J(f - f_0) \leq 2\lambda J(f) + 2\lambda J(f_0) \leq 2\lambda J(f) + 2\delta_n^2\gamma_0$, we have $\lambda J(f) \geq (M^2 - 2\gamma_0)\delta_n^2/2$, and hence, $E_{f_0}(I_2^2) \leq \exp(-(M^2 - 2\gamma_0)\delta_n^2/4)$.

To the end, we address the term $I_2^1$, for which we need to build the uniformly consistent tests. Define $\mathcal{P}_n = \{ f \in H^m(\mathbb{I}) : \|f - f_0\|_{sup} \leq b, J(f) \leq c_4 M_n, V(f - f_0) \geq M^2\delta_n^2 \}$, where $c_4 > (c_3 + 2)/(1 + \gamma_0)^2$ is a fixed constant. Let $\delta = M\delta_n$ and $D(\delta/2, \mathcal{P}_n, \|\cdot\|_{L^2})$ be the $\delta/2$-packing number
in terms of the usual $L^2$-norm. Let $M \geq 2\sqrt{c_4}(1 + \gamma_0)$. It is well known that $\log D(\delta/2, \mathcal{P}_n, \| \cdot \|_{L^2}) \leq \log D(\sqrt{c_4}(1 + \gamma_0)\delta_n, \mathcal{P}_n, \| \cdot \|_{L^2}) \asymp (\sqrt{c_4}(1 + \gamma_0)\delta_n/\sqrt{c_4M_n})^{-1/m} \leq n\delta_n^2$, where the last inequality follows by $\delta_n^2 \geq (nh)^{-1}$.

It follows by the fact that, for sufficiently small $\gamma$, $L$ in terms of the usual $L^2$-norm. Let $M \geq 2(1 + \gamma_0)$. And then, it can be seen that, as $n \to \infty$, $P(\|f - f_0\| \geq 2M\delta_n|D_n)$ converges to zero in $P_{f_0}$-probability.

Before proving Theorem 4.2, let us introduce some preliminaries. Let $\Delta f_j, \Delta f_j \in H^{m}(\mathcal{I})$ for $j = 1, 2, 3$. The Fréchet derivative of $\ell_{n, \lambda}$ can be identified as

$$D\ell_{n, \lambda}(f)\Delta f = \frac{1}{n} \sum_{i=1}^{n} \ell_{a}(Y_i; f(X_i))\langle KX_i, \Delta f \rangle - \langle W_\lambda, \Delta f \rangle \equiv \langle S_{n, \lambda}(f), \Delta f \rangle.$$

Note that $S_{n, \lambda}({\mathcal{f}}_{n, \lambda}) = 0$, and $S_{n, \lambda}(f_0)$ can be expressed as

$$S_{n, \lambda}(f_0) = \frac{1}{n} \sum_{i=1}^{n} \epsilon_i KX_i - W_\lambda f_0.$$

The Fréchet derivative of $S_{n, \lambda}$ ($D S_{n, \lambda}$) is denoted

$$DS_{n, \lambda}(f)\Delta f_1\Delta f_2(D^2S_{n, \lambda}(f)\Delta f_1\Delta f_2\Delta f_3).$$
These derivatives can be explicitly written as

\[
D^2 \ell_{n,\lambda}(f) \Delta f_1 \Delta f_2 = n^{-1} \sum_{i=1}^{n} \ell_a(Y_i; g(X_i)) \langle K_{X_i}, \Delta f_1 \rangle \langle K_{X_i}, \Delta f_2 \rangle - \langle W_\lambda \Delta f_1, \Delta f_2 \rangle,
\]

\[
D^3 \ell_{n,\lambda}(f) \Delta f_1 \Delta f_2 \Delta f_3 = n^{-1} \sum_{i=1}^{n} \ell_a(Y_i; g(X_i)) \langle K_{X_i}, \Delta f_1 \rangle \langle K_{X_i}, \Delta f_2 \rangle \langle K_{X_i}, \Delta f_3 \rangle.
\]

We next review a set of results from [38, 16]. The following result gives the rate of convergence for \( \hat{f}_{n,\lambda} \) when \( h \) satisfies the rate conditions in Theorem 4.2.

**Proposition S.3.** Suppose Assumptions A1 and A2 are satisfied. Furthermore, as \( n \to \infty \), \( h = o(1) \) and \( n^{-1/2}h^{-2}(\log n)(\log \log n)^{1/2} = o(1) \). Then \( \| \hat{f}_{n,\lambda} - f_0 \| = O_{P_{f_0}}(r_n) \).

The proof of Proposition S.3 can be found in [16].

The following lemma is also useful, whose proof can be found in [38, 37].

**Lemma S.4.** *(Functional Bahadur representation under the true model)* Suppose that Assumptions A1–A3 hold, \( h = o(1) \), and \( nh^2 \to \infty \). Recall that \( S_{n,\lambda}(f_0) \) is defined in (S.8). Then we have

\[
\| \hat{f}_{n,\lambda} - f_0 - S_{n,\lambda}(f_0) \| = O_{P_{f_0}}(a_n \log n),
\]

where

\[
a_n = n^{-1/2}r_n h^{-6m-1}/(4m) (\log \log n)^{1/2} + C_\ell h^{-1/2}r_n^2/\log n,
\]

and

\[
C_\ell = \sup_{x \in I} E_{f_0} \{ \sup_{a \in \mathbb{R}} | \ell_a(Y; a) | X = x \}.
\]

Let the true \( f_0 \) be expressed as \( f_0(\cdot) = \sum_{\nu=1}^{\infty} f_0^\nu \varphi_\nu(\cdot) \) satisfying Condition (S). Under the conditions of Lemma S.4, it can be directly verified that \( \| S_{n,\lambda}(f_0) \| = O_{P_{f_0}}(r_n) \), and therefore, \( \| \hat{f}_{n,\lambda} - f_0 \| = O_{P_{f_0}}(a_n \log n + r_n) \). To see this, by Proposition 3.1 we have

\[
E_{f_0} \{ \| \sum_{i=1}^{n} \epsilon_i K_{X_i} \|^2 \} = \frac{1}{n} E_{f_0} \{ \epsilon^2 K(X, X) \} = \frac{1}{n} \sum_{\nu=1}^{\infty} \frac{1}{1 + \lambda \rho_\nu} = O((nh)^{-1}),
\]
and
\[ \| W_\lambda f_0 \|^2 = \sum_{\nu=1}^{\infty} f_\nu^0 \frac{\lambda \rho_\nu}{1 + \lambda \rho_\nu} \varphi_\nu, \sum_{\nu=1}^{\infty} f_\nu^0 \frac{\lambda \rho_\nu}{1 + \lambda \rho_\nu} \varphi_\nu \]
\[ = \sum_{\nu=1}^{\infty} |f_\nu^0|^2 \frac{\lambda^2 \rho_\nu^2}{1 + \lambda \rho_\nu} \]
\[ = \lambda^{1+\frac{\beta-1}{2m}} \sum_{\nu=1}^{\infty} |f_\nu|^2 \rho_\nu^{\frac{\beta-1}{2m}} (\lambda \rho_\nu)^{\frac{\beta-1}{2m}} = O(h^{2m+\beta-1}), \]

where the last equation follows by \( \lambda = h^{2m}, \sup_{x \geq 0} x^{1-\frac{\beta-1}{2m}} < \infty \), and Condition (S). Then \( \| S_n, \lambda (f_0) \| = O_{f_0}(nh)^{-1/2} + h^{m+\frac{\beta-1}{2}} = O_{f_0}(\hat{r}_n) \).

To complete the proof of Theorem 4.2, we consider a function class \( G = \{ g \in H^m(I) : \| g \|_{\sup} \leq 1, J(g) \leq c^{-2h} \lambda \} \), where \( c > 0 \) is the universal constant satisfying Lemma 5.8. The following equi-continuity lemma over \( G \) is useful for further theoretical analysis. The proof can be found in [38].

**Lemma S.5.** Suppose that \( \psi_n(z; g) \) is a measurable function defined upon \( z = (y, x) \in Y \times I \) and \( g \in G \) satisfying \( \psi_n(z; 0) = 0 \) and the following Lipschitz continuity condition: for any \( i = 1, \ldots, n \) and \( f, g \in G \),
\[ |\psi_n(Z_i; f) - \psi_n(Z_i; g)| \leq c^{-1} h^{1/2} \| f - g \|_{\sup}, \text{ a.s. in } P_{f_0}-\text{probability}, \]
where \( c > 0 \) is the universal constant specified in Lemma 5.8. Then we have
\[ \lim_{n \to \infty} P_{f_0}^{\nu} \left( \sup_{g \in G} \frac{\| Z_n(g) \|}{h^{-(2m-1)/(4m)} \| g \|_{\sup}^\gamma n^{-1/2}} \leq (5 \log \log n)^{1/2} \right) = 1, \]
where \( Z_n(g) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [\psi_n(Z_i; g) K_{X_i} - E_{f_0} Z_n \{ \psi_n(Z; g) K_{X_i} \} ] \), \( E_{f_0} Z_n \) represents the expectation taken with respect to \( Z \) under \( P_{f_0} \), and \( \gamma = 1 - \frac{1}{2m} \).

**Proof of Theorem 4.2.** Let \( r_n = (nh)^{-1/2} + \lambda^{1/2} \) and denote \( \hat{f} = \hat{f}_{n, \lambda} \) for convenience. By \( \| \hat{f} - f_0 \| = O_{f_0}(\hat{r}_n) \) and Assumption A1, there exists a constant \( M_1 > 1 \) s.t. \( P_{f_0}^{\nu}(\mathcal{E}_n') \) approaches one (as \( n \to \infty \)), where
\[ \mathcal{E}_n' = \{ \| \hat{f} - f_0 \| \leq M_1 \hat{r}_n, \max_{1 \leq i \leq n} \| \tilde{\ell}_{a}(Y_i; a) \| \leq M_1 \log n, \max_{1 \leq i \leq n} \| \tilde{\ell}_{a}(Y_i; a) \| \leq M_1 \log n \}. \]

Define \( A_{i,n} = \{ |\tilde{\ell}_{a}(Y_i; f_0(X_i))| \leq M_1 \log n \} \), for \( i = 1, 2, \ldots, n \). By Assumption A1 (a), we can actually manage the above \( M_1 \) such that \( P_{f_0}(A_{i,n}') \leq
where \( Z \) satisfies Lemma 5.8. By Proposition 4.1, there exists a large \( M_0 \) s.t. \( P(f \in H^n(I) : \| f - f_0 \| \leq M_0 r_n | D_n) \) converges to one in \( P_{f_0} \)-probability. Let \( M > M_1 \) be a constant to be determined. Since

\[
P(\| f - f_0 \| > 2M\tilde{r}_n | D_n) = P(\| f - f_0 \| > M_0 r_n | D_n) + P(2M\tilde{r}_n < \| f - f_0 \| \leq M_0 r_n | D_n),
\]

where the first term on the right side converges to zero in \( P_{f_0} \)-probability, we only look at the second term. For convenience, define \( S_n = \{ f : 2M\tilde{r}_n < \| f - f_0 \| \leq M_0 r_n \} \). Since \( \tilde{r}_n \leq r_n \), on \( E_n \) and for any \( f \in S_n \), \( \| f - \hat{f} \| \leq \| f - f_0 \| + \| \hat{f} - f_0 \| \leq (M_0 + M_1)r_n \), and \( \| f - \hat{f} \| \geq \| f - f_0 \| - \| \hat{f} - f_0 \| > M\tilde{r}_n \).

Define

\[
(S.12) \quad E''_n = \{ \sup_{g \in G} \frac{\| Z^{(j)}(g) \|^2}{h^{-(2m-1)/(4m)} \| g \|^2 \sup + n - 1/2} \leq (5 \log \log n)^{1/2}, j = 1, 2 \},
\]

where \( Z^{(j)}(g) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi^{(j)}(Z_i; g) K_{X_i} \) and \( (M_1 \log n)^{-1} c^{-1} h^{1/2} I_n, \psi^{(2)}(Z; g) = (M_1 \log n)^{-1} c^{-1} h^{1/2} I_n g(X_i) I_n \).

Then Lemma S.5 says that, as \( n \to \infty \), \( P_{f_0} (E''_n) \) tends to one. Let \( E_n = E'_n \cap E''_n \) whose \( P^n_{f_0} \)-probability thus goes to one as \( n \) tends to infinity. In the rest of the proof, we will assume that \( E_n \) holds.

For any \( f \), define \( I_n(f) = \int_0^1 \int_0^1 s D S_{n, \lambda}(\hat{f} + s s'(f - \hat{f}))(f - \hat{f})d s d s' \).

Then by Taylor’s expansion we get that \( \ell_{n, \lambda}(f) - \ell_{n, \lambda}(\hat{f}) = S_{n, \lambda}(f - \hat{f}) + I_n(f) = I_n(f) \). By (3.8) we have that

\[
P(f \in S_n | D_n) = \frac{\int_{S_n} \exp(n(\ell_{n, \lambda}(f) - \ell_{n, \lambda} \hat{f})))dP(f)}{\int_{H^n(I)} \exp(n(\ell_{n, \lambda}(f) - \ell_{n, \lambda} \hat{f})))dP(f)} = \frac{\int_{S_n} \exp(nI_n(f))dP(f)}{\int_{H^n(I)} \exp(nI_n(f))dP(f)}
\]

We first derive a lower bound for \( J_1 = \int_{H^n(I)} \exp(nI_n(f))dP(f) \). Immediately,

\[
J_1 \geq \int_{f: \| f - f_0 \| \leq \tilde{r}_n} \exp(nI_n(f))dP(f).
\]

Note on \( E_n \) and when \( \| f - f_0 \| \leq \tilde{r}_n \), we have \( \| f - \hat{f} \| \leq \| f - f_0 \| + \| \hat{f} - f_0 \| \leq (M_1 + 1)\tilde{r}_n \). Let \( c > 0 \) be the constant specified in Lemma 5.8, \( d_n = c(M_1 + 1)h^{-1/2} \tilde{r}_n \) and \( g = d_n^{-1}(f - \hat{f}) \). Immediately, it can be shown that \( \| g \| \sup \leq ch^{-1/2} \| g \| = ch^{-1/2} d_n^{-1} \| f - \hat{f} \| \leq ch^{-1/2} d_n^{-1} (M_1 + 1) \tilde{r}_n = 1 \), and \( J(g) \leq \lambda^{-1} \| g \|^2 = \lambda^{-1} d_n^{-2} \| f - \hat{f} \|^2 \leq \lambda^{-1} d_n^{-2} (M_1 + 1)^2 \tilde{r}_n^2 = c^{-2} h^{-1} \). This implies \( g \in \mathcal{G} \).
Then we can rewrite $I_n(f)$ as

\begin{align*}
I_n(f) &= \int_0^1 \int_0^1 s[DS_{n,\lambda}(\hat{f} + ss'(f - \hat{f})) - DS_{n,\lambda}(f_0)](f - \hat{f})(f - \hat{f})dss' \\
&\quad + \frac{1}{2}[DS_{n,\lambda}(f_0) - E_{f_0}\{DS_{n,\lambda}(f_0)\}](f - \hat{f})(f - \hat{f}) - \frac{1}{2}\|f - \hat{f}\|^2 \\
&= d_n^2 \int_0^1 \int_0^1 s[DS_{n,\lambda}(\hat{f} + ss'd_n g) - DS_{n,\lambda}(f_0)]ggdss' \\
&\quad + \frac{d_n^2}{2}[DS_{n,\lambda}(f_0) - E_{f_0}\{DS_{n,\lambda}(f_0)\}][gg - \frac{1}{2}\|f - \hat{f}\|^2] \\
&\equiv T_1 + T_2 - \frac{1}{2}\|f - \hat{f}\|^2,
\end{align*}

where $T_1$ and $T_2$ denote the first and second terms in the above expression.

Since for any $s \in [0, 1]$, by Lemma 5.8 and $\|g\|_{\sup} \leq 1$, we have $\|\hat{f} + sd_n g - f_0\|_{\sup} \leq \|\hat{f} - f_0\|_{\sup} + d_n\|g\|_{\sup} \leq c(2M_1 + 1)h^{-1/2}\tilde{r}_n$. 

\[\text{SUPPLEMENT TO NONPARAMETRIC BVM THEOREM} \quad 29\]
Note that, on $\mathcal{E}_n$ and using $\|g\|^2 \leq c^{-2}h$, for any $s \in [0, 1]$,

$$
\|DS_{n,\lambda}(\hat{f} + sd_n g) - DS_{n,\lambda}(f_0)\| g g
\leq \frac{1}{n} \sum_{i=1}^{n} \| \dot{\hat{e}}(Y_i; (\hat{f} + sd_n g)(X_i)) - \dot{\hat{e}}(Y_i; f_0(X_i)) \| g(X_i)^2 \|g\|^2
\leq \frac{1}{n} \sum_{i=1}^{n} \sup_{a \in \mathbb{R}} |\dot{\hat{e}}(Y_i; a)| \cdot \| \hat{f} - f_0 + sd_n g \| \|g\|^2
\leq \frac{c M_1 (2M_1 + 1) h^{-1/2} \tilde{r}_n \log n}{n} \sum_{i=1}^{n} g(X_i)^2
= \frac{c M_1 (2M_1 + 1) h^{-1/2} \tilde{r}_n \log n}{n} \sum_{i=1}^{n} [g(X_i) K_{X_i} - E_{f_0}^X \{g(X) K_X\}], g\)
+ c M_1 (2M_1 + 1) h^{-1/2} \tilde{r}_n \log n E_{f_0}^X \{g(X)^2\}
\leq \frac{c^2 M_1 (2M_1 + 1) h^{-1} \tilde{r}_n \log n}{n} \{\sum_{i=1}^{n} [\psi_n^{(1)}(Z_i; g) K_{X_i} - E_{f_0}^X \{\psi_n^{(1)}(Z; g) K_X\}], g\)
+ c C_2 M_1 (2M_1 + 1) h^{-1/2} \tilde{r}_n \log n \|g\|^2
\leq c^2 M_1 (2M_1 + 1) h^{-1} n^{-1/2} \tilde{r}_n \log n \|Z^{(1)}(g)\| \cdot \|g\|
+ c C_2 M_1 (2M_1 + 1) h^{-1/2} \tilde{r}_n \log n \|g\|^2
\leq c^2 M_1 (2M_1 + 1) h^{-1/2} \tilde{r}_n \log n h^{-(2m-1)/(4m)} + n^{-1/2}(5 \log \log n)^{1/2} \|g\|
+ c C_2 M_1 (2M_1 + 1) h^{-1/2} \tilde{r}_n \log n \|g\|^2
\leq 2c M_1 (2M_1 + 1) n^{-1/2} \tilde{r}_n h^{-(4m-1)/(4m)}(5 \log \log n)^{1/2} \log n
+ c^{-1} C_2 M_1 (2M_1 + 1) h^{1/2} \tilde{r}_n \log n.
$$

Therefore,

$$
|T_1|
\leq C_1(c, C_2, M_1) r_n^2 (n^{-1/2} \tilde{r}_n h^{-(8m-1)/(4m)}(5 \log \log n)^{1/2} \log n + h^{-1/2} \tilde{r}_n \log n)
= C_1(c, C_2, M_1) r_n^2 b_{n,1},
$$

where $C_1(c, C_2, M_1)$ is a positive constant depending only on $c, C_2, M_1$. By assumption in the theorem, as $n \rightarrow \infty$, $T_1 = o_{P_{f_0}}(r_n^2 \tilde{r}_n^2)$.

To handle $T_2$, we use similar ideas. Note that on $\mathcal{E}_n''$ and by $\|g\|_{sup} \leq 1$, $\|g\| \leq c^{-1/2} h^{1/2}$, $C_2 P(A_n^{(2)}) \leq n^{-1/2} h^{-(2m-1)/(4m)}(5 \log \log n)^{1/2}$ and
direct calculations we have
\[
\|DS_n,\lambda(f_0) - E_{f_0}\{DS_n,\lambda(f_0)\}\|_g \\
\leq \frac{1}{n} \sum_{i=1}^{n} |\hat{e}_a(Y_i; f_0(X_i))g(X_i)I_{A_n}K_{X_i} - E_{f_0}\{\hat{e}_a(Y_i; f_0(X_i))g(X_i)I_{A_n}K_{X_i}\}| \\
\leq \|g\| + C_2 P(A_{1m}^c) \\
= n^{-1/2}(M_1 \log n)h^{-1/2}\|Z_n^{(2)}(g)\| \cdot \|g\| + C_2 P(A_{1m}^c) \\
\leq 2M_1n^{-1/2}h^{-(2m-1)/(4m)}\log n(5 \log \log n)^{1/2} + C_2 P(A_{1m}^c) \\
\leq 3M_1n^{-1/2}h^{-(2m-1)/(4m)}\log n(5 \log \log n)^{1/2},
\]
therefore,
\[
|T_2| \leq (d_n^2/2)3M_1n^{-1/2}h^{-(2m-1)/(4m)}\log n(5 \log \log n)^{1/2} \\
\leq C_2(c, M_1)\tilde{r}_n^2n^{-1/2}h^{-(6m-1)/(4m)}\log n(\log \log n)^{1/2} = C_2(c, M_1)b_{n2}\tilde{r}_n^2,
\]
where \(C_2(c, M_1) > 0\) is a constant depending only on \(c, M_1\). By assumptions in the theorem, \(T_2 = o_P(n^{1/2}(\tilde{r}_n^2))\).

Let \(n\) be sufficiently large such that \(C_1(c, C_2, M_1)b_{n1} + C_2(c, M_1)b_{n2} \leq c^2(M_1 + 1)^2/2\). By the above bounds of \(T_1\) and \(T_2\), on \(E_n\) and when \(\|f - f_0\| \leq \tilde{r}_n\),
\[
nI_n(f) \geq -n(C_1(c, C_2, M_1)b_{n1} + C_2(c, M_1)b_{n2} + c^2(M_1 + 1)^2/2)\tilde{r}_n^2 \\
\geq -c^2(M_1 + 1)^2n\tilde{r}_n^2.
\]
Therefore, \(J_1 \geq \exp(-c^2(M_1 + 1)^2n\tilde{r}_n^2)\Pi(B_n)\), where \(B_n = \{f : \|f - f_0\| \leq \tilde{r}_n\}\).

To proceed, we still need a lower bound for \(\Pi(B_n)\). We employ an idea used in the proof of Proposition 4.1. Let \(\omega \in \mathcal{H}_\beta\) satisfy the properties (i)–(ii) of Lemma S.3, with \(d_n\) therein replaced by \(\tilde{r}_n\). Note that \(\tilde{r}_n^2 \geq h^{2m + \beta - 1} = \lambda^{2m + \beta - 1}/2m\) which implies that \(\lambda \leq \tilde{r}_n^{2m + \beta - 1}/4\). It can be shown that, if \(\tilde{V}(\tilde{G} - \omega) \leq \tilde{r}_n^{2m + \beta - 1}/16\) and \(\tilde{J}(\tilde{G} - \omega) \leq \tilde{r}_n^{2m + \beta - 1}/16\), then \(\tilde{V}(\tilde{G} - \tilde{f}_0) + \lambda \tilde{J}(\tilde{G} - \tilde{f}_0) \leq (\tilde{r}_n^2 + \lambda\tilde{r}_n^{2m + \beta - 1})/4 \leq \tilde{r}_n^2\). It then follows by Cameron-Martin theorem, Gaussian
correlation inequality and the results of [23] that

\[
\Pi(B_n) = P(\|G - f_0\| \leq \bar{r}_n) \\
= P(\bar{V}(\bar{G} - \bar{f}_0) + \lambda \bar{J}(\bar{G} - \bar{f}_0) \leq \bar{r}_n^{2(\beta-1)}) \\
\geq P(\bar{V}(\bar{G} - \omega) \leq \bar{r}_n^{2}/16, \bar{J}(\bar{G} - \omega) \leq \bar{r}_n^{2m+\beta-1}/16) \\
\geq \exp(-\|\omega\|_{\beta}^2/2)P(\bar{V}(\bar{G}) \leq \bar{r}_n^{2}/32, \bar{J}(\bar{G}) \leq \bar{r}_n^{2m+\beta-1}/32) \\
\geq \exp(-c_5\bar{r}_n^{2/2m+\beta-1}),
\]

where the last inequality follows by

\[
-\log P(\bar{V}(\bar{G}) \leq \bar{r}_n^{2}/32) \asymp \bar{r}_n^{-2m+\beta-1},
\]

(see Example 4.5 of [23]), and \(c_5 > 0\) is a universal constant. Since \(n\bar{r}_n^{2m+\beta-1} \geq n(2 - \bar{r}_n^{-1/2m+\beta-1}) > 1\) which implies \(\bar{r}_n^{-1/2m+\beta-1} < n\bar{r}_n^{2m+\beta-1}\), we get that

\[
(S.13) \quad J_1 \geq \exp(-c_5n\bar{r}_n^{2m+\beta-1}).
\]

Next we derive an upper bound for \(J_2 = \int_{S_n} \exp(nI_n(f))d\Pi(f)\). The idea we employ is similar to how we handle \(J_1\), but with some technical difference. Let \(d_{sn} = c(M_0 + M_1)h^{-1/2}r_n\), and \(g_\ast = d_{sn}^{-1}(f - \hat{f})\). It is easy to see that, on \(\mathcal{E}_n\) and when \(f \in S_n\), we get that \(\|g_\ast\|^2 \leq c^{-2}h\), \(\|g_\ast\|_{\text{sup}} \leq 1\) and \(J(g_\ast) \leq c^{-2}h\lambda^{-1}\), which imply that \(g_\ast \in \mathcal{G}\). Then we can rewrite \(I_n(f)\) as

\[
I_n(f) = d_{sn}^2 \int_0^1 \int_0^1 s[DS_{n,\lambda}(\hat{f} + ss'd_{sn}g_\ast) - DS_{n,\lambda}(f_0)]g_\ast g_\ast dsds' \\
+ d_{sn}^2 [DS_{n,\lambda}(f_0) - E_{f_0}\{DS_{n,\lambda}(f_0)\}]g_\ast g_\ast - \frac{1}{2}\|f - \hat{f}\|^2 \\
= T_3 + T_4 + \frac{1}{2}\|f - \hat{f}\|^2.
\]

Similar to the analysis of \(T_1\) and \(T_2\), it can be shown that, on \(\mathcal{E}_n\) and for any \(f \in S_n\),

\[
|T_3| \leq 2c^3M_1(M_0 + 2M_1)(M_0 + M_1)^2r_n^3n^{-1/2}h^{-3m-1}\frac{\log n}{4m} + c^2M_1(M_0 + 2M_1)(M_0 + M_1)^2r_n^3h^{-1} \log n \\
\leq C_3(c, M_0, M_1)(r_n^3n^{-1/2}h^{-3m-1}\frac{\log n}{4m} + r_n^3h^{-1} \log n) \\
= C_3(c, M_0, M_1)b_{n,3},
\]
Thus, on \( E \) is the universal constant in (S.13).

Satisfying

\[
\parallel E \parallel \leq 4.2 \text{ that, there exist } M, N > 0 \text{ numbers within } (0, n) \text{ the above equation tends to zero as } n \to \infty \text{.}
\]

Therefore, on \( E \) and for any \( f \in S_n \),

\[
I_n(f) \leq C_3(c, M_0, M_1) b_{n3} + C_4(c, M_0, M_1) b_{n4} - \frac{M^2}{2} \tilde{r}_n^2 \leq (1 - M^2/2) \tilde{r}_n^2,
\]

which implies \( J_2 \leq \exp((1 - M^2/2)n \tilde{r}_n^2) \).

Consequently, on \( E_n \),

\[
P(f \in S_n | D_n) = \frac{J_2}{J_1} \leq \exp((1 + c_5 - M^2/2)n \tilde{r}_n^2).
\]

By choosing \( M > M_1 \) and \( M > \sqrt{2(1 + c_5)} \), we have that the right side in
the above equation tends to zero as \( n \to \infty \). This concludes the proof. \( \square \)

S.6. Proofs in Section 5.1.

Proof of Theorem 5.1. Let \( \delta, \varepsilon \) be arbitrary prefixed small positive numbers within \((0, 1/8)\). Let \( E_n' \) and \( E_n'' \) be defined as in (S.11) and (S.12), in which
the constant \( M_1 \) is chosen to be large so that, as \( n \) approaches infinity, both events have \( P_{f_0}(E_n') \geq 1 - \frac{\varepsilon}{4} \).
It follows by Theorem 4.2 that, there exist \( M, N > 0 \) such that, for any \( n \geq N \), \( P_{f_0}(E_n'') \geq 1 - \frac{\varepsilon}{4} \),
where \( E_n'' = \{ P(f : \| f - f_0 \| \leq M \tilde{r}_n | D_n) \geq 1 - \delta \} \). We can actually make
the above \( M \) greater than \( M_1 + \sqrt{2(c_5 + 1)} \) so that, on \( E_n' \) and for any \( f \)
satisfying \( \| f - f_0 \| \geq M \tilde{r}_n, \| f - \hat{f} \| \geq (M - M_1) \tilde{r}_n \geq \sqrt{2(c_5 + 1)} \), where \( c_5 \)
is the universal constant in (S.13).

Using the proofs of Theorem 4.2, we can show that, on \( E_n' \),

\[
\int_{H^n(1)} \exp(-\frac{n}{2} f - \hat{f})^2) d\Pi(f) \geq \int_{f: f - f_0 \leq \tilde{r}_n} \exp(-\frac{n}{2} f - \hat{f})^2) d\Pi(f) \geq \exp(-(c_5 + 1/2) n \tilde{r}_n^2),
\]

\[
\int_{f: f - f_0 \geq M \tilde{r}_n} \exp(-\frac{n}{2} f - \hat{f})^2) d\Pi(f) \leq \exp(-(c_5 + 1) n \tilde{r}_n^2).
\]

Thus, on \( E_n' \),

\[
P_0(f : \| f - f_0 \| \geq M \tilde{r}_n) \leq \exp(-n \tilde{r}_n^2/2),
\]
in which the right side tends to zero as $n \to \infty$. This means that the above $N$ can be further managed such that, for any $n \geq N$, $P_{f_0}'(E_n^*) > 1 - \varepsilon$, where $E_n^* = E_n' \cap E_n'' \cap E_n'''$ and $E_n''' = \{P(f_0 : \|f - f_0\| \leq M \bar{r}_n) \geq 1 - \delta\}$.

For arbitrary $B \in \mathcal{B}$, let $B' = B \cap \{f : \|f - f_0\| \leq M \bar{r}_n\}$. Then it is easy to see that, on $E_n^*$,

$$|P(B|D_n) - P_0(B)| \leq \delta + |P(B'|D_n) - P_0(B')|.$$

To prove the theorem, it is sufficient to give a small upper bound for $|P(B'|D_n) - P_0(B')|$, and this upper bound does not depend on $B$.

Observe that

$$n(\ell_{n,\lambda}(f) - \ell_{n,\lambda}(\hat{f})) + \frac{n}{2}\|f - \hat{f}\|^2 = n(T_1 + T_2),$$

where

$$T_1 = \int_0^1 \int_0^1 s[D\ell_{n,\lambda}(\hat{f} + s\ell'f - \hat{f})) - D\ell_{n,\lambda}(f_0)](f - \hat{f})(f - \hat{f})dsds',$$

$$T_2 = \frac{1}{2}[D\ell_{n,\lambda}(f_0) - E_{f_0}\{D\ell_{n,\lambda}(f_0)\}](f - \hat{f})(f - \hat{f}).$$

Based on the proof of Theorem 4.2, it can be shown that, on $E_n^*$ and for any $f$ satisfying $\|f - f_0\| \leq M \bar{r}_n$, $n|T_1| \leq C_5(c, M_1, M)\eta_n^2\beta_1$ and $n|T_2| \leq C_6(c, M_1, M)\eta_n^2\beta_2$, where $C_j(c, M_1, M)$, $j = 5, 6$, are constants depending only on $c, M_1, M$, and recall that $c$ satisfies Lemma 5.8. By assumptions, the above upper bounds for $n|T_1|$ and $n|T_2|$ are both of order $o(1)$, as $n \to \infty$.

Define

$$J_{n1} = \int_{H^m(1)} \exp(n(\ell_{n,\lambda}(f) - \ell_{n,\lambda}(\hat{f})))d\Pi(f)$$

and

$$J_{n2} = \int_{H^m(1)} \exp(-\frac{n}{2}\|f - \hat{f}\|^2)d\Pi(f).$$

Also define $J_{n1}' = \int_{f : \|f - f_0\| \leq M \bar{r}_n} \exp(n(\ell_{n,\lambda}(f) - \ell_{n,\lambda}(\hat{f})))d\Pi(f)$ and $J_{n2}' = \int_{f : \|f - f_0\| \leq M \bar{r}_n} \exp(-\frac{n}{2}\|f - \hat{f}\|^2)d\Pi(f)$. Then, on $E_n^*$, we have

$$0 \leq \frac{J_{nj} - J_{nj}'}{J_{nj}} \leq \delta, \text{ for } j = 1, 2,$$
which implies that

\[ J_{nj} \leq \frac{1}{1 - \delta} \tilde{J}_{nj}, \text{ for } j = 1, 2. \]

Let \( R_n(f) = n(\ell_{n, \lambda}(f) - \ell_{n, \lambda}(\hat{f})) + \frac{n}{2} \| f - \hat{f} \|^2. \) On \( \mathcal{E}_n^* \) and for any \( f \) satisfying \( \| f - f_0 \| \leq M \tilde{r}_n, |R_n(f)| \leq C_5(c, M_1, M) n \tilde{r}_n^2 b_{n1} + C_6(c, M_1, M) n \tilde{r}_n^2 b_{n2} \equiv t_n(c, M_1, M). \) By assumption, as \( n \to \infty, t_n(c, M_1, M) \to 0. \) It can then be seen that, on \( \mathcal{E}_n^* \),

\[
|P(B' | D_n) - P_0(B')| = \left| \int_{B'} \exp\left(\frac{n(\ell_{n, \lambda}(f) - \ell_{n, \lambda}(\hat{f}))}{J_{n1}} - \frac{\frac{n}{2} \| f - \hat{f} \|^2}{J_{n2}} \right) \frac{d\Pi(f)}{J_{n2}} \right| \\
\leq \int_{B'} \exp\left(\frac{n}{2} \| f - \hat{f} \|^2 \right) \left| \frac{\exp(R_n(f))}{J_{n1}} - \frac{1}{J_{n2}} \right| d\Pi(f) \\
\leq \int_{f: \| f - f_0 \| \leq M \tilde{r}_n} \exp\left(\frac{n}{2} \| f - \hat{f} \|^2 \right) \left| \frac{\exp(R_n(f)) J_{n2} - J_{n1}}{J_{n1} J_{n2}} \right| d\Pi(f) \\
\leq \exp\left( t_n(c, M_1, M) \right) \left| \frac{J_{n2} - J_{n1}}{J_{n1} J_{n2}} \right| J_{n2} + \frac{2 t_n(c, M_1, M)}{J_{n2}} J_{n2} \\
\leq \exp\left( t_n(c, M_1, M) \right) \left| \frac{J_{n2} - J_{n1}}{J_{n1}} \right| + 2 t_n(c, M_1, M). \\
\]

(S.15)

It can also be shown that, on \( \mathcal{E}_n^* \),

\[
|\tilde{J}_{n2} - \tilde{J}_{n1}| \leq \int_{f: \| f - f_0 \| \leq M \tilde{r}_n} \exp\left(\frac{n}{2} \| f - \hat{f} \|^2 \right) \left| \exp(R_n(f)) - 1 \right| d\Pi(f) \\
\leq 2 t_n(c, M_1, M) \int_{f: \| f - f_0 \| \leq M \tilde{r}_n} \exp\left(\frac{n}{2} \| f - \hat{f} \|^2 \right) d\Pi(f) \\
= 2 t_n(c, M_1, M) \tilde{J}_{n2},
\]

which implies

\[
\frac{1}{1 + 2 t_n(c, M_1, M)} \leq \tilde{J}_{n2} \tilde{J}_{n1} \leq \frac{1}{1 - 2 t_n(c, M_1, M)}.
\]

Combined with (S.14) we have that

\[
\frac{1 - \delta}{1 + 2 t_n(c, M_1, M)} \leq (1 - \delta) \frac{\tilde{J}_{n2}}{J_{n1}} \leq \frac{J_{n2}}{J_{n1}} \leq \frac{J_{n2}}{J_{n1}} \leq \frac{1}{1 - \delta} \frac{J_{n2}}{J_{n1}} \leq \frac{1}{(1 - \delta)(1 - 2 t_n(c, M_1, M))},
\]
which leads to
\[
\frac{2t_n(c, M_1, M) + \delta}{1 + 2t_n(c, M_1, M)} \leq \frac{J_{n2}}{J_{n1}} - 1 \leq \frac{2t_n(c, M_1, M) + \delta}{(1 - \delta)(1 - 2t_n(c, M_1, M))},
\]
therefore,
\[
\left| \frac{J_{n2}}{J_{n1}} - 1 \right| \leq \frac{2t_n(c, M_1, M) + \delta}{(1 - \delta)(1 - 2t_n(c, M_1, M))} \to \frac{\delta}{1 - \delta} \leq 2\delta.
\]

It follows by (S.15) that as \( n \) is large, on \( \mathcal{E}_n^* \) and any \( B \in \mathcal{B}, |P(B'|D_n) - P_0(B')| \leq 3\delta \). This shows the desired conclusion (5.2).

**Proof of Theorem 5.2.** Let \( T : f \mapsto \{f_\nu : \nu \geq 1\} \) be the one-to-one map from \( H^m(\mathbb{I}) \) to \( \mathcal{R}_m \), as defined in Lemma 3.2. Let \( \Pi'_W \) and \( \Pi' \) be the probability measures induced by \( \{\xi_\nu + a_\nu \hat{f}_\nu : \nu \geq 1\} \) and \( \{v_\nu : \nu \geq 1\} \). Then \( d\Pi'_W / d\Pi' \) equals \( \lim_{N \to \infty} p_1_N(f_1, \ldots, f_N) / p_2_N(f_1, \ldots, f_N) \), where \( p_1_N \) and \( p_2_N \) are the probability densities under \( f_\nu \sim \xi_\nu + a_\nu \hat{f}_\nu \) and \( f_\nu \sim v_\nu \), \( \nu = 1, \ldots, N \), respectively. A direct evaluation leads to that
\[
\frac{d\Pi'_W}{d\Pi'}(\{f_\nu : \nu \geq 1\}) = C_{n,\lambda} \exp\left(-\frac{n}{2} \sum_{\nu=1}^{\infty} (f_\nu - \hat{f}_\nu)^2 (1 + \lambda \rho_\nu)\right),
\]
where
\[
C_{n,\lambda} = \prod_{\nu=1}^{m} \left(1 + n\sigma_\nu^2\right)^{1/2} \prod_{\nu>m} \left(1 + n(1 + \lambda \rho_\nu) \rho_\nu^{-1/2} \right)^{1/2} \exp\left(\frac{1}{2} \sum_{\nu=1}^{\infty} \hat{f}_\nu^2 b_\nu\right),
\]
and \( b_\nu = n\sigma_\nu^2 / (n + \sigma_\nu^2) \) for \( \nu = 1, \ldots, m \), and \( b_\nu = n(1 + \lambda \rho_\nu) \rho_\nu^{1+\beta/(2m)} / (n(1 + \lambda \rho_\nu) + \rho_\nu^{1+\beta/(2m)}) \) for \( \nu > m \). Since \( \sum_\nu \hat{f}_\nu^2 \rho_\nu < \infty \) and \( \beta > 1 \), it is not hard to see that \( C_{n,\lambda} \) is a finite constant.

By similar arguments in the proof of Lemma 3.2, we have for any \( S \subseteq \mathcal{R}_m \), \( \Pi_W(T^{-1}B) = \Pi'_W(B) \), and \( \Pi(T^{-1}B) = \Pi'(B) \). By change of variable, for any \( \Pi \)-measurable \( S \subseteq H^m(\mathbb{I}) \),
\[
\Pi_W(S) = \Pi_W'(TS)
\]
\[
= \int_{TS} d\Pi_W'(\{f_\nu : \nu \geq 1\})
\]
\[
= C_{n,\lambda} \int_{TS} \exp\left(-\frac{n}{2} \sum_{\nu=1}^{n} (f_\nu - \hat{f}_\nu)^2 (1 + \lambda \rho_\nu)\right) d\Pi'(f_\nu : \nu \geq 1)
\]
\[
= C_{n,\lambda} \int_{S} \exp\left(-\frac{n}{2} \|f - \hat{f}_{n,\lambda}\|_2^2\right) d\Pi(f).
\]
In particular, let \( S = H^m(1) \) in the above equations, we get that
\[
C_{n, \lambda} = \left( \int_{H^m(1)} \exp\left(-\frac{n}{2} \| f - \hat{f}_{n, \lambda} \|^2 \right) d\Pi(f) \right)^{-1}.
\]
This proves the desired result. \( \square \)

S.7. Proofs in Section 5.2.

Proof of Lemma 5.3. It is easy to see that
\[
n \sqrt{c_0 h \| Z \|^2} = n \sqrt{c_0 h} \sum_{\nu=1}^{\infty} c_\nu \eta_\nu^2 (1 + \lambda \rho_\nu)
\]
\[
= n \sqrt{c_0 h} \sum_{\nu=1}^{\infty} c_\nu (1 + \lambda \rho_\nu)(\eta_\nu^2 - 1) + n \sqrt{c_0 h} \sum_{\nu=1}^{\infty} c_\nu (1 + \lambda \rho_\nu),
\]
where \( \eta_\nu \) are iid standard normal random variables independent of the data \( D_n \). Therefore, the logarithm of the moment generating function of \( U_n = (n \sqrt{c_0 h \| Z \|^2} - n \sqrt{c_0 h} \sum_{\nu=1}^{\infty} c_\nu (1 + \lambda \rho_\nu)) / \tau_1 \) is
\[
\log(E\{\exp(tU_n)\}) = \log(E\{\exp(t \tau_1^{-1} n \sqrt{c_0 h} \sum_{\nu=1}^{\infty} c_\nu (1 + \lambda \rho_\nu)(\eta_\nu^2 - 1))\})
\]
\[
= \sum_{\nu=1}^{\infty} \log(E\{\exp(t \tau_1^{-1} n \sqrt{c_0 h} c_\nu (1 + \lambda \rho_\nu)(\eta_\nu^2 - 1))\})
\]
\[
= - \sum_{\nu=1}^{\infty} \frac{1}{2} \log(1 - 2t \tau_1^{-1} n \sqrt{c_0 h} c_\nu (1 + \lambda \rho_\nu)) + t \tau_1^{-1} n \sqrt{c_0 h} c_\nu (1 + \lambda \rho_\nu)
\]
\[
= t^2 c_0 n^2 h \sum_{\nu=1}^{\infty} c_\nu^2 (1 + \lambda \rho_\nu)^2 / \tau_1^2 + O(t^2 n^3 c_0^3 h^{3/2} \sum_{\nu=1}^{\infty} c_\nu^3 (1 + \lambda \rho_\nu)^3 / \tau_1^3)
\]
\[
\to t^2 / 2,
\]
where the limit in the last equation follows by (5.4) and the fact, as \( n \to \infty \), \( \tau_1 \sim 1 \) and \( n^3 h^{3/2} \sum_{\nu=1}^{\infty} c_\nu^3 (1 + \lambda \rho_\nu)^3 / \tau_1^3 = O(h^{1/2}) = o(1) \). Consequently, the result of the lemma holds. \( \square \)

Proof of Lemma 5.4. The proof relies on the following kernel function:
\[
R(x, x') = \sum_{\nu=1}^{\infty} c_\nu \varphi_\nu(x) \varphi_\nu(x'), \text{ for any } x, x' \in \mathbb{I},
\]
where \(c_\nu\) are defined in Section 5.2. We also use \(R_\nu(\cdot)\) to denote the function \(R(x, \cdot)\), for simplicity.

Denote \(\hat{f} = \hat{f}_{n,\lambda}, \tilde{f} = \tilde{f}_{n,\lambda}\), and \(R_n = \hat{f} - f_0 - S_{n,\lambda}(f_0)\). For any \(\nu \geq 1\),

\[
\begin{align*}
\hat{f}_\nu &= V(\hat{f}, \varphi_\nu) = V(R_n, \varphi_\nu) + V(f_0, \varphi_\nu) + V(S_{n,\lambda}(f_0), \varphi_\nu) \\
&= V(R_n, \varphi_\nu) + f_0^0 + \frac{1}{n} \sum_{i=1}^n \epsilon_i \varphi_\nu(X_i) + f_0^0 \frac{\lambda \rho_\nu}{1 + \lambda \rho_\nu}.
\end{align*}
\]

Therefore, by \(a_\nu/(1 + \lambda \rho_\nu) = nc_\nu\), we get that

\[
\begin{align*}
\tilde{f} - f_0 \\
&= \sum_\nu (a_\nu \hat{f}_\nu - f_0^0) \varphi_\nu \\
&= \sum_\nu a_\nu V(R_n, \varphi_\nu) \varphi_\nu + \sum_\nu (a_\nu - 1) f_0^0 \varphi_\nu + \frac{1}{n} \sum_{i=1}^n \epsilon_i \sum_\nu a_\nu \varphi_\nu(X_i) \varphi_\nu \\
&\quad + \sum_\nu f_0^0 a_\nu \frac{\lambda \rho_\nu}{1 + \lambda \rho_\nu} \varphi_\nu \\
&= \sum_\nu a_\nu V(R_n, \varphi_\nu) \varphi_\nu + \sum_\nu (a_\nu - 1) f_0^0 \varphi_\nu + \sum_{i=1}^n \epsilon_i R_{X_i} + \sum_\nu f_0^0 a_\nu \frac{\lambda \rho_\nu}{1 + \lambda \rho_\nu} \varphi_\nu,
\end{align*}
\]

(S.16)

where \(R_{X_i} = \sum_\nu c_\nu \varphi_\nu(X_i) \varphi_\nu\), for \(i = 1, \ldots, n\). Denote the above four terms by \(I_1, I_2, I_3, I_4\). It follows by Lemma S.4 and \(a_\nu^2 \leq 1\) that

\[
\begin{align*}
n \|I_1\|^2 &= n \sum_\nu a_\nu^2 |V(R_n, \varphi_\nu)|^2 (1 + \lambda \rho_\nu) \leq n \|R_n\|^2 \\
&= O_{P_0}(n(a_n \log n)^2) = o_{P_0}(n_i^2 h_{n,i}^2) = o_{P_0}(h_{n,i}-1).
\end{align*}
\]

Since \(|a_\nu - 1| = O(n^{-1})\) for \(\nu = 1, \ldots, m\), it is easy to see that \(n \sum_{\nu=1}^m (a_\nu - 1)^2 |f_\nu^0|^2 (1 + \lambda \rho_\nu) = O(n^{-1})\). By Dominated convergence theorem and direct
calculations it follows that
\[
\begin{align*}
&\quad n \sum_{\nu>m} (a_\nu - 1)^2 |f_\nu^0|^2 (1 + \lambda \rho_\nu) \\
&= n \sum_{\nu>m} \left( \frac{n(1 + \lambda \rho_\nu)}{n(1 + \lambda \rho_\nu) + \rho_\nu^{1+\beta/(2m)}} - 1 \right)^2 |f_\nu^0|^2 (1 + \lambda \rho_\nu) \\
&\quad \times \left( \frac{(\lambda \rho_\nu)^{2+\beta/m}}{(1 + \lambda \rho_\nu + (\lambda \rho_\nu)^{1+\beta/(2m)})^2} |f_\nu^0|^2 \right) \\
&\quad + n\lambda \sum_{\nu>m} \left( \frac{(\lambda \rho_\nu)^{1+\beta/(2m)} + (\lambda \rho_\nu)^{2+\beta/(2m)}}{(1 + \lambda \rho_\nu + (\lambda \rho_\nu)^{1+\beta/(2m)})^2} |f_\nu^0|^2 \rho_\nu^{1+(\beta-1)/(2m)} \right).
\end{align*}
\]

Let \( g_\lambda(\nu) = \frac{(\lambda \rho_\nu)^{1+\beta/(2m)} + (\lambda \rho_\nu)^{2+\beta/(2m)}}{(1 + \lambda \rho_\nu + (\lambda \rho_\nu)^{1+\beta/(2m)})^2} \), for \( \nu > m \). Clearly, for all \( \lambda \) and \( \nu \),
\[
0 < g_\lambda(\nu) \leq \sup_{x>0} \frac{x^{1+\beta/(2m)} + x^{2+\beta/(2m)}}{(1 + x + x^{1+\beta/(2m)})^2} < \infty,
\]
and \( \lim_{\lambda \to 0} g_\lambda(\nu) = 0 \) for all \( \nu \). Since \( \sum_{\nu} |f_\nu^0|^2 \rho_\nu^{1+(\beta-1)/(2m)} < \infty \), it follows by dominated convergence theorem that, as \( \lambda \to 0 \), \( \sum_{\nu>m} g_\lambda(\nu)|f_\nu^0|^2 \rho_\nu^{1+(\beta-1)/(2m)} \to 0 \). Therefore, by \( n\lambda^{1+\beta-1/(2m)} \asymp h^{-1} \), we get that, as \( n \to \infty \),
\[
\begin{align*}
n\|I_2\|^2 &= n \sum_{\nu=1}^m (a_\nu - 1)^2 |f_\nu^0|^2 (1 + \lambda \rho_\nu) + n \sum_{\nu>m} (a_\nu - 1)^2 |f_\nu^0|^2 (1 + \lambda \rho_\nu) \\
&= O(n^{-1}) + o(h^{-1}) = o(h^{-1}).
\end{align*}
\]

By dominated convergence theorem and direct examinations it can be shown that
\[
\begin{align*}
n\|I_4\|^2 &= n \sum_{\nu} |f_\nu^0|_2^2 \frac{(\lambda \rho_\nu)^2}{1 + \lambda \rho_\nu} \leq n \sum_{\nu} |f_\nu^0|^2 \frac{(\lambda \rho_\nu)^2}{1 + \lambda \rho_\nu} \\
&= n\lambda^{1+(\beta-1)/(2m)} \sum_{\nu} |f_\nu^0|^2 \rho_\nu^{1+(\beta-1)/(2m)} \frac{(\lambda \rho_\nu)^{1-(\beta-1)/(2m)}}{1 + \lambda \rho_\nu} = o(h^{-1}),
\end{align*}
\]
where the last equation follows by \( 0 < 1 - (\beta - 1)/(2m) < 1 \), implying \( \sup_{x>0} \frac{x^{1-(\beta-1)/(2m)}}{1+x} \asymp \infty \), and dominated convergence theorem.
Next, let us handle $I_3$. Note that
\[
\| \sum_{i=1}^{n} \epsilon_i R_{X_i} \|_2^2 = \sum_{i=1}^{n} \epsilon_i^2 \| R_{X_i} \|_2^2 + 2 \sum_{i<j} \epsilon_i \epsilon_j \langle R_{X_i}, R_{X_j} \rangle \equiv \sum_{i=1}^{n} \epsilon_i^2 \| R_{X_i} \|_2^2 + W(n),
\]
where $W(n) = \sum_{i<j} W_{ij}$ and $W_{ij} = 2 \epsilon_i \epsilon_j \langle R_{X_i}, R_{X_j} \rangle$. For $1 \leq i, j \leq n$, it can be shown that
\[
\langle R_{X_i}, R_{X_j} \rangle = \sum_{\nu} c_{\nu}^2 (1 + \lambda \rho_{\nu}) \varphi_{\nu}(X_i) \varphi_{\nu}(X_j).
\]
Note that the weighted quadratic form $W(n)$ is clean in the sense of [17]. Let $\sigma(n)^2 = E_f \{ W(n)^2 \}$ and $G_I, G_{II}, G_{IV}$ be defined as
\[
G_I = \sum_{i<j} E_f \{ W_{ij}^4 \},
\]
\[
G_{II} = \sum_{i<j<k} (E_f \{ W_{ij}^2 W_{ik}^2 \} + E_f \{ W_{ij}^2 W_{jk}^2 \} + E_f \{ W_{ki}^2 W_{kj}^2 \}), \quad \text{and}
\]
\[
G_{IV} = \sum_{i<j<k<l} (E_f \{ W_{ij} W_{ik} W_{jk} W_{kl} \} + E_f \{ W_{ij} W_{il} W_{jk} W_{kl} \} + E_f \{ W_{ik} W_{il} W_{jk} W_{jl} \}).
\]
Since $\varphi_{\nu}$ are uniformly bounded, it can be shown that, uniformly for $x \in \mathbb{I}$,
\[
\| R_x \|_2^2 = \sum_{\nu} c_{\nu}^2 (1 + \lambda \rho_{\nu}) | \varphi_{\nu}(x) |^2 \leq \sum_{\nu=1}^{m} (n + \sigma_{\nu}^{-2})^{-2} + n^{-2} \sum_{\nu>m} \frac{1 + \lambda \rho_{\nu}}{1 + \lambda \rho_{\nu} + (\lambda \rho_{\nu})^{1+\beta/(2m)}} \lesssim n^{-2} h^{-1}.
\]
Therefore, $G_I \lesssim n^2 E_f \{ \epsilon_i^4 \langle R_{X_i}, R_{X_i} \rangle^4 \} = O(n^2 (n^{-2} h^{-1})^4) = O(n^{-6} h^{-4})$.

In the meantime, $G_{II} \lesssim n^3 E_f \{ W_{ij}^2 W_{ik}^2 \} \leq n^3 E_f \{ W_{ij}^4 \}^{1/2} E_f \{ W_{ik}^4 \}^{1/2} = O(n^3 (n^{-4} h^{-2})^2) = O(n^{-5} h^{-4})$. To address $G_{IV}$, note for pairwise distinct $i, j, k, l$,
\[
E_f \{ W_{ij} W_{ik} W_{il} W_{jk} \} = 2^4 E_f \{ \epsilon_i^2 \epsilon_j^2 \epsilon_k^2 \epsilon_l^2 \langle R_{X_i}, R_{X_j} \rangle \langle R_{X_i}, R_{X_k} \rangle \langle R_{X_i}, R_{X_l} \rangle \langle R_{X_j}, R_{X_k} \rangle \langle R_{X_j}, R_{X_l} \rangle \langle R_{X_k}, R_{X_l} \rangle \}
\]
\[
= 16 \sum_{\nu} c_{\nu}^4 (1 + \lambda \rho_{\nu})^4 = 16 \sum_{\nu=1}^{m} (n + \sigma_{\nu}^{-2})^{-8} + \sum_{\nu>m} \frac{(1 + \lambda \rho_{\nu})^4}{n(1 + \lambda \rho_{\nu} + (\lambda \rho_{\nu})^{1+\beta/(2m)})^8} \lesssim n^{-8} h^{-1},
\]
so \( G_{IV} = O(n^4)E_{\tilde{f}_0}\{W_{ij}W_{ik}W_{lj}W_{lk}\} = O(n^{-4}h^{-1}) \). By direct calculations it can be shown that

\[
\sigma(n)^2 = \binom{n}{2} E_{f_0}\{W^2_{12}\} \\
= 4 \binom{n}{2} E_{f_0} E_{\tilde{f}_0}\{\varepsilon_1^2 \varepsilon_2^2 (\sum_{\nu} c_\nu^2 (1 + \lambda \rho_\nu \varphi_\nu(X_1) \varphi_\nu(X_2))^2)\} \\
= 2n(n-1) \sum_{\nu} c_\nu^4 (1 + \lambda \rho_\nu)^2 \\
= 2n(n-1) \sum_{\nu} \frac{(n + \sigma_\nu^{-2})^{-4} + n^{-4} \sum_{\nu > m} (1 + \lambda \rho_\nu)^2}{(1 + \lambda \rho_\nu + n^{-1} \rho_\nu^{1+\beta/(2m)})^4} \\
\approx c_0 \frac{1}{2} n^{-2} h^{-1},
\]

where \( \alpha_n \approx \beta_n \) means \( \lim_{n \to \infty} a_n/b_n = 1 \). Since \( 2m + \beta > 2 \), we have \( nh^2 \sim n^{1-2/(2m+\beta)} \to \infty \) as \( n \to \infty \). Therefore, \( G_I, G_{II}, G_{IV} \) are all of order \( o(\sigma(n)^4) \). By Proposition 3.2 of [17], as \( n \) approaches \( \infty \), \( c_0^{1/2} \tau_2^{-1} nh^{1/2} W(n) \to N(0, 1) \).

To the end of the proof, we address \( \sum_{i=1}^n \varepsilon_i^2 \| R_{X_i} \|^2 \). Note that

\[
\text{Var}_{\tilde{f}_0}\left( \sum_{i=1}^n \varepsilon_i^2 \| R_{X_i} \|^2 \right) = n E_{f_0}\{\varepsilon^4 \| R_X \|^4\} - n E_{\tilde{f}_0}\{\varepsilon^2 \| R_X \|^2\}^2 \\
\leq n E_{f_0}\{\varepsilon^4 \| R_X \|^4\} \preceq n^{-3} h^{-2},
\]

and

\[
n E_{\tilde{f}_0}\{\varepsilon^2 \| R_X \|^2\} = n \sum_{\nu} c_\nu^2 (1 + \lambda \rho_\nu) = (nhc_0)^{-1} c_0 h \sum_{\nu} (nc_\nu)^2 (1 + \lambda \rho_\nu) = (nhc_0)^{-1} \zeta_2.
\]

Therefore, \( \sum_{i=1}^n \varepsilon_i^2 \| R_{X_i} \|^2 = (nhc_0)^{-1} \zeta_2 + O_{P_{\tilde{f}_0}}(n^{-3/2} h^{-1}) \). By the above analysis of \( I_1, I_2, I_3, I_4 \), we get that

\[
n \| \tilde{f} - f_0 \|^2 = n \| I_3 \|^2 + O_{P_{\tilde{f}_0}}(h^{-1}) \\
= n \sum_{i=1}^n \varepsilon_i^2 \| R_{X_i} \|^2 + n W(n) + O_{P_{\tilde{f}_0}}(h^{-1}) \\
= (hc_0)^{-1} \zeta_2 + n W(n) + O_{P_{\tilde{f}_0}}(h^{-1}) + O_{P_{\tilde{f}_0}}(n^{-1/2} h^{-1}) \\
= (hc_0)^{-1} \zeta_2 + n W(n) + O_{P_{\tilde{f}_0}}(h^{-1}).
\]

This proves Lemma 5.4. \( \square \)
Proof of Lemma 5.5. By straightforward calculations we get that

\[
\zeta_j = c_0 h \sum_{\nu=1}^{m} \left( \frac{n}{n + \sigma_\nu^2} \right)^j + c_0 h \sum_{\nu > m} \left( \frac{n}{n(1 + \lambda \rho_\nu + \rho_\nu^{1+\beta/(2m)})} \right)^j (1 + \lambda \rho_\nu).
\]

Since the first term in the right of the above equation tends to zero as \(n \to \infty\), we only focus on the second term. Note that

\[
c_0 h \sum_{\nu > m} \left( \frac{n}{n(1 + \lambda \rho_\nu + \rho_\nu^{1+\beta/(2m)})} \right)^j (1 + \lambda \rho_\nu)
\]

\[
- c_0 h \sum_{\nu > m} \left( \frac{n}{n(1 + \lambda \rho_\nu + \rho_\nu^{1+\beta/(2m)})} \right)^2 (1 + \lambda \rho_\nu)
\]

\[
= c_0 h \sum_{\nu > m} \frac{(\lambda \rho_\nu + n^{-1} \rho_\nu^{1+\beta/(2m)}) (1 + \lambda \rho_\nu)}{(1 + \lambda \rho_\nu + n^{-1} \rho_\nu^{1+\beta/(2m)})^2}
\]

\[
\approx \int_0^\infty \frac{(x^{2m} + x^{2m+\beta})(1 + x^{2m})}{(1 + x^{2m} + x^{2m+\beta})^2} \ dx > 0.
\]

This completes the proof for the \(\zeta_j\) part. The proof for the \(\tau_j^2\) part is similar, which is ignored.

\(\square\)

S.8. Proofs in Section 5.3.

Proof of Theorem 5.7. We start from the decomposition (S.16) in the proof of Lemma 5.4. Let \(I_1, I_2, I_3, I_4\) be defined the same as in (S.16).

By \(m > 1 + \sqrt{3}/2\) and \(1 < \beta < m + 1/2\), it can be shown that \(r_n = (nh)^{-1/2} + h^m = O(h^m)\), and hence \(n(a_n \log n)^2 = o(1)\). Therefore,

\[
n \|I_1\|^2 \lesssim n \sum_\nu d_\nu |V(R_n, \varphi_\nu)|^2
\]

\[
\lesssim n \sum_\nu |V(R_n, \varphi_\nu)|^2 \leq n \|R_n\|^2 = O_{P_{f_0}}(n(a_n \log n)^2) = o_{P_{f_0}}(1).
\]

By \(|a_\nu - 1| = O(1/n)\) and the choice of \(d_\nu\), we get

\[
n \|I_2\|^2 = n \sum_\nu d_\nu(a_\nu - 1)^2 |f_\nu|^2
\]

\[
= n \sum_{\nu=1}^{m} d_\nu(a_\nu - 1)^2 |f_\nu|^2 + n \sum_{\nu > m} d_\nu(a_\nu - 1)^2 |f_\nu|^2
\]

\[
= O(1/n) + n \sum_{\nu > m} d_\nu \frac{\rho_\nu^{2+\beta/m}}{(n(1 + \lambda \rho_\nu + \rho_\nu^{1+\beta/(2m)})^2) |f_\nu|^2}.
\]
Using $\lambda \asymp n^{-2m/(2m+\beta)}$ and dominated convergence theorem, it can be shown that

\[
\sum_{\nu>m} d_{\nu} \frac{\rho_{\nu}^{2+\beta/m}}{(n + \lambda \rho_{\nu} + \rho_{\nu}^{1+\beta/(2m)})^2} |f_{\nu}^0|^2 \lesssim n^{-1} \sum_{\nu>m} \rho_{\nu}^{2+\beta/m} \frac{(n + \lambda \rho_{\nu} + \rho_{\nu}^{1+\beta/(2m)})^2}{(1 + \lambda \rho_{\nu} + \rho_{\nu}^{1+\beta/(2m)})^2} |f_{\nu}^0|^2 \rho_{\nu}^{1+(\beta-1)/(2m)} = o(1).
\]

Therefore, $n\|I_2\|_1^2 = o(1)$. Similarly, we have

\[
n\|I_3\|_1^2 = \sum_{\nu>m} d_{\nu} |f_{\nu}^0|^2 \frac{a_{\nu}^2(\lambda \rho_{\nu})^2}{(1 + \lambda \rho_{\nu})^2} \lesssim n \sum_{\nu>m} d_{\nu} \frac{(\lambda \rho_{\nu})^2}{(1 + \lambda \rho_{\nu})^2} |f_{\nu}^0|^2 \lesssim n \sum_{\nu>m} \rho_{\nu}^{-1/(2m)} \frac{(\lambda \rho_{\nu})^2}{(1 + \lambda \rho_{\nu})^2} \rho_{\nu}^{-1-\beta/(2m)} |f_{\nu}^0|^2 \rho_{\nu}^{1+(\beta-1)/(2m)} \lesssim \sum_{\nu} \frac{(\lambda \rho_{\nu})^{1-\beta/(2m)}}{(1 + \lambda \rho_{\nu})^2} |f_{\nu}^0|^2 \rho_{\nu}^{1+(\beta-1)/(2m)} = o(1).
\]

Next we handle $I_3$. Define

\[
T(x, y) = \sum_{\nu=1}^{m} \frac{\varphi_{\nu}(x) \varphi_{\nu}(y)}{1 + n^{-1} \sigma_{\nu}^2} + \sum_{\nu>m} \frac{\varphi_{\nu}(x) \varphi_{\nu}(y)}{1 + \lambda \rho_{\nu} + n^{-1} \rho_{\nu}^{1+\beta/(2m)}}.
\]

It follows by Theorem 5 of [42] and proof of Proposition 2.2 in [38] that for any $j = 0, 1, 2, \ldots, 2m$, $\sup_{\nu} \|\varphi_{\nu}^{(j)}\|_{\sup} \rho_{\nu}^{-j/(2m)} < \infty$. And hence, for any $x, y \in \mathbb{I}$,

\[
\left| \frac{\partial}{\partial x} T(x, y) \right| \lesssim O(1) + \sum_{\nu>m} \frac{|\varphi_{\nu}(x) \varphi_{\nu}(y)|}{1 + (h \nu)^{2m} + (h \nu)^{2m+\beta}} \lesssim O(1) + \sum_{\nu>m} \frac{\nu}{1 + (h \nu)^{2m} + (h \nu)^{2m+\beta}} \lesssim h^{-2} \int_0^\infty \frac{s}{1 + s^{2m} + s^{2m+\beta}} ds.
\]
which implies sup$_{x,y \in \mathbb{I}} |\frac{\partial}{\partial x} T(x, y)| = O(h^{-2})$.

For simplicity, define $T_x(y) = T(x, y)$ for any $x, y \in \mathbb{I}$. Let $R_X$ be defined as in the proof of Lemma 5.4. By direct examination, it can be seen that

$$R_X = \sum_{\nu=1}^{m} \frac{\phi_\nu(X) \phi_\nu}{n + \sigma_\nu^2} + \sum_{\nu > m} \frac{\phi_\nu(X) \phi_\nu}{n(1 + \lambda \rho_\nu) + \rho_\nu^{1+\beta/(2m)}} = T_X/n.$$  

So $I_3 = n^{-1} \sum_{i=1}^{n} \epsilon_i T_X(i)$ which implies $n||I_3||^2 = n^{-1} \sum_{i=1}^{n} \epsilon_i T_X(i)^2$.

Since $E_{\mathcal{F}_0} \{\exp(|\epsilon|/C_1)\} < \infty$, we can choose a constant $C > C_1$ such that $P^n_{\mathcal{F}_0}(\mathcal{E}^n) \rightarrow 1$, where $\mathcal{E}^n = \{ \max_{1 \leq i \leq n} |\epsilon_i| \leq b_n \equiv C \log n \}$. We can even choose the above $C$ to be properly large so that the following rate condition holds:

$$(S.17) \quad h^{-1}n^{1/2} \exp(-b_n/(2C_1)) = o(1), \quad h^{-2} \exp(-b_n/(2C_1)) = o(1).$$

Define

$$H_n(z) = n^{-1/2} \sum_{i=1}^{n} \epsilon_i T_X(i), \quad H^b_n(z) = n^{-1/2} \sum_{i=1}^{n} \epsilon_i I(|\epsilon_i| \leq b_n) T_X(i).$$

Write

$$H_n = H_n - H^b_n - E_{\mathcal{F}_0} \{ H_n - H^b_n \} + H^b_n - E_{\mathcal{F}_0} \{ H^b_n \}.$$  

Clearly, on $\mathcal{E}_n$, $H_n = H^b_n$, and hence,

$$|H_n(z) - H^b_n(z)| = |E_{\mathcal{F}_0} \{ H_n(z) - H^b_n(z) \}|$$

$$= |E_{\mathcal{F}_0} \{ H_n(z) - H^b_n(z) \}|$$

$$= n^{1/2} |E_{\mathcal{F}_0} \{ \epsilon I(|\epsilon| > b_n) T_X(z) \}|$$

$$\lesssim n^{1/2} h^{-1} \exp(-b_n/(2C_1)) = o(1),$$

where the last $o(1)$-term follows by (S.17) and is free of the argument $z$. Thus,

$$(S.18) \quad \sup_{z \in \mathbb{I}} |H_n(z) - H^b_n(z) - E_{\mathcal{F}_0} \{ H_n(z) - H^b_n(z) \}| = o_{P^n_{\mathcal{F}_0}}(1).$$

Define $\mathcal{R}_n = H^b_n - E_{\mathcal{F}_0} \{ H^b_n \}$ and $Z_n(e, x) = n^{1/2}(P_n(e, x) - P(e, x))$, where $P_n(e, x)$ is the empirical distribution of $(e, X)$ and $P(e, x)$ is the population distribution of $(e, X)$ under $P^n_{\mathcal{F}_0}$-probability. It follows by Theorem 1 of [41] that

$$(S.19) \quad \sup_{e \in \mathbb{E}, x \in \mathbb{I}} |Z_n(e, x) - W(t(e, x))| = O_{P^n_{\mathcal{F}_0}}(n^{-1/2}(\log n)^2),$$  

where the last $o(1)$-term follows by (S.17) and is free of the argument $z$. Thus,
where $W(\cdot, \cdot)$ is Brownian bridge indexed on $\mathbb{I}^2$, $t(e, x) = (F_1(x), F_2(e|x))$, $F_1$ is the marginal distribution of $X$ and $F_2$ is the conditional distribution of $\epsilon$ given $X$ both under $P^n_{f_0}$-probability. It can be seen that

$$R_n(z) = \int_0^{1} \int_{-b_n}^{b_n} eT_x(z) dZ_n(e, x).$$

Define

$$R^0_n(z) = \int_0^{1} \int_{-b_n}^{b_n} eT_x(z) dW(t(e, x)).$$

Write

$$dU_n(x) = \int_{-b_n}^{b_n} edZ_n(e, x), \quad dU^0_n(x) = \int_{-b_n}^{b_n} edW(t(e, x)).$$

It follows from integration by parts where all quadratic variation terms are zero that

$$U_n(x) = \int_{-b_n}^{b_n} edZ_n(e, x) = Z_n(e, x)e|_{e=-b_n} - \int_{-b_n}^{b_n} Z_n(e, x)de,$$

$$U^0_n(x) = \int_{-b_n}^{b_n} edW(t(e, x)) = W(t(e, x))e|_{e=-b_n} - \int_{-b_n}^{b_n} W(t(e, x))de,$$

and hence, it follows by (S.19) that

$$\sup_{x \in \mathbb{I}} |U_n(x) - U^0_n(x)| = O_{P^n_{f_0}}(b_n n^{-1/2} (\log n)^2).$$

It follows from integration by parts again and $\sup_{x,y \in \mathbb{I}} |\frac{\partial}{\partial x} T(x, y)| = O(h^{-2})$ that

$$R_n(z) = \int_0^{1} T_x(z) dU_n(x) = U_n(x)T(x, z)|_{z=0}^{1} - \int_0^{1} U_n(x) \frac{\partial}{\partial x} T(x, z) dx,$$

$$R^0_n(z) = \int_0^{1} T_x(z) dU^0_n(x) = U^0_n(x)T(x, z)|_{z=0}^{1} - \int_0^{1} U^0_n(x) \frac{\partial}{\partial x} T(x, z) dx,$$

and hence,

(S.20) $\sup_{z \in \mathbb{I}} |R_n(z) - R^0_n(z)| = O_{P^n_{f_0}}(h^{-2} b_n n^{-1/2} (\log n)^2) = o_{P^n_{f_0}}(1),$

where the last equality follows by $2m + \beta > 4$ and hence

$$h^{-2} n^{-1/2} b_n (\log n)^2 = O(n^{-1/2+2/(2m+\beta)} (\log n)^3) = o(1).$$
Next we handle the term $\mathcal{R}_0^n$. Write $W(s,t) = B(s,t) - stB(1,1)$, where $B(s,t)$ is standard Brownian motion indexed on $\mathbb{I}^2$. Define
\[
\bar{\mathcal{R}}_0^n(z) = \int_0^1 \int_{-b_n}^{b_n} eT_x(z) dB(t(e,x)).
\]
Let $F(e,x) = F_1(x)F_2(e,x)$ be the joint distribution of $(\epsilon, X)$. It is easy to see that
\[
|\bar{\mathcal{R}}_0^n(z) - \mathcal{R}_0^n(z)| = |B(1,1)| \cdot \left| \int_0^1 \int_{-b_n}^{b_n} eT_x(z) dF(e,x) \right| = |B(1,1)| \cdot |E_{f_0}\{\epsilon I(\epsilon \leq b_n)T_X(z)\}| = |B(1,1)| \cdot |E_{f_0}\{\epsilon I(\epsilon > b_n)T_X(z)\}| = O_{P_n}(h^{-1}\exp(-b_n/(2C_1))) = o_{P_n}(1),
\]
where the last equality follows by (S.17). Therefore, we have shown that
\[
(S.21) \quad \sup_{z \in I}|\bar{\mathcal{R}}_0^n(z) - \mathcal{R}_0^n(z)| = o_{P_n}(1).
\]
By (S.18), (S.20) and (S.21) that
\[
(S.22) \quad n\|I_3\|_\frac{2}{1} = \|H_n\|_\frac{2}{1} = \|\bar{\mathcal{R}}_0^n\|_\frac{2}{1} + o_{P_n}(1).
\]
Define
\[
\bar{\mathcal{R}}(z) = \int_0^1 \int_{-\infty}^{\infty} eT_x(z) dB(t(e,x)).
\]
Let $\Delta(z) = \bar{\mathcal{R}}(z) - \bar{\mathcal{R}}_0^n(z)$. Then
\[
\Delta(z) = \int_0^1 \int_{|\epsilon| > b_n} eT_x(z) dB(t(e,x)).
\]
For each $z$, $\Delta(z)$ is a zero-mean Gaussian random variable with variance
\[
E_{f_0}\{\Delta(z)^2\} = \int_0^1 \int_{|\epsilon| > b_n} e^2T_x(z)^2 dF(e,x) \lesssim h^{-2}E_{f_0}\{e^2I(|\epsilon| > b_n)\} = O(h^{-2}\exp(-b_n/(2C_1))) = o(1),
\]
where the last $o(1)$-term follows from (S.17) and is free of the argument $z$. Therefore,
\[
E_{f_0}\{\|\bar{\mathcal{R}} - \bar{\mathcal{R}}_0^n\|_\frac{2}{1}^2\} \lesssim E_{f_0}\{\|\Delta\|_{L^2}^2\} = \int_0^1 E_{f_0}\{\Delta(z)^2\} dz = o(1),
\]
implying that \( \|\bar{R} - \bar{R}_n\|_\perp = o_{P^n}(1) \). Therefore, it follows by (S.22) that

(S.23) \[ n\|I_3\|^2 = \|\bar{R}\|^2 + o_{P^n}(1). \]

It follows from the definition of \( T(\cdot, \cdot) \) that

\[
\|\bar{R}\|^2 = \sum_{\nu=1}^m \frac{d\nu\tilde{\eta}_\nu^2}{(1 + n^{-1}\sigma^2_\nu)^2} + \sum_{\nu>m} \frac{d\nu\tilde{\eta}_\nu^2}{(1 + \lambda\rho_\nu + n^{-1}\rho_\nu^{1+\beta/(2m)})^2} \overset{d}{\rightarrow} \sum_{\nu=1}^\infty d\nu\tilde{\eta}_\nu^2,
\]

where \( \tilde{\eta}_\nu = \int_{0}^{1} t^\infty e\varphi\nu(x)dB(t,e,x) \). It is easy to see that for any \( \nu, \mu \),

\[
E_{f_0}\{\tilde{\eta}_\nu\tilde{\eta}_\mu\} = E_{f_0}\{\epsilon^2\varphi\nu(X)\varphi\mu(X)\} = E_{f_0}\{B(X)\varphi\nu(X)\varphi\mu(X)\} = V(\varphi\nu, \varphi\mu) = \delta_{\nu\mu},
\]

that is, \( \tilde{\eta}_\nu \) are iid standard normal random variables. Combined with the above analysis of terms \( I_1, I_2, I_3, I_4 \), we have shown that as \( n \rightarrow \infty \),

\[ n\|f_0 - \tilde{f}_{n,\lambda}\|^2 \overset{d}{\rightarrow} \sum_{\nu=1}^\infty d\nu\tilde{\eta}_\nu^2. \]

This implies that as \( n \rightarrow \infty \),

\[ P^n_{f_0}(f_0 \in R^n(\alpha)) = P^n_{f_0}(n\|f_0 - \tilde{f}_{n,\lambda}\|^2 \leq c_\alpha) \rightarrow 1 - \alpha. \]

The proof is completed. \( \square \)

S.9. Proofs in Section 5.4.

Proof of Theorem 5.9. Let \( I_1, I_2, I_3, I_4 \) be the decomposition terms in (S.16). Then \( \|\bar{f} - f_0 - I_3\|^2 \leq 3(\|I_1\|^2 + \|I_2\|^2 + \|I_4\|^2) \). We first handle the norms of \( I_1, I_2, I_4 \). Since \( m > 1 + \sqrt{3}/2 \) and \( 1 < \beta < m + 1/2 \), it can be shown that \( n(a_n \log n)^{1/2} = o(1) \). It follows the derivations in the proof of Lemma 5.4 that, as \( n \rightarrow \infty \), \( n\|I_1\|^2 = O_{P^n}(n(a_n \log n)^2) = o_{P^n}(1) \),

\[
n\|I_2\|^2 \lesssim nh^{2m+\beta} \sum_{\nu} \frac{(hv)^{2m+\beta} + (hv)^{4m+\beta}}{(1 + (hv)^{2m})^2} f^0 v^{2m+\beta} = o(1),
\]

\[
n\|I_4\|^2 \lesssim nh^{2m+\beta} \sum_{\nu} \frac{(hv)^{2m-\beta}}{1 + (hv)^{2m}} |f^0 v|^{2m+\beta} = o(1).
\]

This means that \( \|\bar{f} - f_0 - I_3\| = o_{P^n}(n^{-1/2}) \). It follows by (5.8) that \( |F(\bar{f} - f_0 - I_3)| \leq \kappa h^{-1/2}\|\bar{f} - f_0 - I_3\| = o_{P^n}(h^{-1/2}n^{-1/2}) \).
Before proceeding further, we analyze \( F(I_3) \) which equals \( \sum_{i=1}^{n} \epsilon_i F(R_{X_i}) \). Since both \( \varphi_\nu \) and \( F(\varphi_\nu) \) are uniformly bounded, we have

\[
E_{f_0}\{\epsilon^2|nF(R_X)|^2\} = \sum_{\nu} n^2 \epsilon^2 F(\varphi_\nu)^2 \\
= \sum_{\nu=1}^{m} \frac{F(\varphi_\nu)^2}{(1 + n^{-1}\sigma_\nu^{-2})^2} + \sum_{\nu>m} \frac{F(\varphi_\nu)^2}{(1 + \lambda \rho_\nu + n^{-1}\rho_\nu^{4/(2m)})^2} \\
\lesssim h^{-1},
\]

\[
|nF(R_X)| = \left| \sum_{\nu=1}^{m} \frac{1}{1 + n^{-1}\sigma_\nu^{-2}} \varphi_\nu(X) F(\varphi_\nu) \right| + \sum_{\nu>m} \frac{1}{1 + \lambda \rho_\nu + n^{-1}\rho_\nu^{4/(2m)}} \varphi_\nu(X) F(\varphi_\nu) \lesssim h^{-1}.
\]

Let \( s_n^2 = \text{Var}_{f_0}\{\sum_{i=1}^{n} \epsilon_i nF(R_{X_i})\} \). It is easy to see that

\[
s_n^2 = n \left( \sum_{\nu=1}^{m} \frac{F(\varphi_\nu)^2}{(1 + n^{-1}\sigma_\nu^{-2})^2} + \sum_{\nu>m} \frac{F(\varphi_\nu)^2}{(1 + \lambda \rho_\nu + n^{-1}\rho_\nu^{4/(2m)})^2} \right).
\]

Clearly, \( s_n^2 = O(nh^{-1}) \). By (5.10), \( s_n^2 \gtrsim nh^{-r} \).

Next we verify Lindeberg’s condition: for any \( \delta > 0 \),

\[
s_n^{-2} \sum_{i=1}^{n} E_{f_0}\{\epsilon_i^2|nF(R_X_i)|^2 I(|\epsilon_i nF(R_{X_i})| \geq \delta s_n)\} \\
= n s_n^{-2} E_{f_0}\{\epsilon^2|nF(R_X)|^2 I(|\epsilon nF(R_X)| \geq \delta s_n)\} \\
\leq n s_n^{-2} E_{f_0}\{\epsilon^4|nF(R_X)|^4\}^{1/2} P(|\epsilon nF(R_X)| \geq \delta s_n)^{1/2} \\
\leq n s_n^{-2} E_{f_0}\{\epsilon^4|nF(R_X)|^4\}^{1/2} ((\delta s_n)^{-4} E_{f_0}\{\epsilon^4|nF(R_X)|^4\})^{1/2} \\
= n s_n^{-2} (\delta s_n)^{-2} E_{f_0}\{\epsilon^4|nF(R_X)|^4\} \\
\lesssim n s_n^{-4} h^{-3} \lesssim n^{-1} h^{2r-3} = n^{-\frac{2m+\beta+2r-3}{2m+\beta}} = o(1).
\]

This implies that Lindeberg’s condition holds. By central limit theorem, \( nF(I_3)/s_n = \sum_{i=1}^{n} \epsilon_i nF(R_{X_i})/s_n \overset{d}{\to} N(0,1) \).

Observe that \( F(f_0) \in CI_{n,\lambda}(\alpha) \) is equivalent to \( |F(\tilde{f} - f_0)| \leq z_{\alpha/2} t_n / \sqrt{n} \).

In the meantime,

\[
\frac{n}{s_n} F(\tilde{f} - f_0) = \frac{n}{s_n} F(\tilde{f} - f_0 - I_3) + \frac{n}{s_n} F(I_3) = \frac{n}{s_n} F(I_3) + o_{P_{f_0}}(1) \overset{d}{\to} N(0,1).
\]
Since $t_n^2 \geq s_n^2/n$, we get that
\[
P_{f_0}(F(f_0) \in CI_{n,\lambda}(\alpha)) = P_{f_0}(|\tilde{f} - f_0| \leq z_{\alpha/2} t_n/\sqrt{n}) = P_{f_0}(\frac{n}{s_n}|\tilde{f} - f_0| \leq \frac{z_{\alpha/2} \sqrt{n} t_n}{s_n}) \geq P_{f_0}(\frac{n}{s_n}|\tilde{f} - f_0| \leq z_{\alpha/2}) \to 1 - \alpha, \text{ as } n \to \infty.
\]

In particular, we notice that when $\sum \nu F(\varphi_\nu)^2 < \infty$ and (5.10) holds for $r = 0$, it follows by dominated convergence theorem that both $s_n^2/n$ and $t_n^2/n$ converge to $\sum \nu F(\varphi_\nu)^2$. This implies that $\sqrt{n \alpha / s_n} \to 1$, and hence, $P_{f_0}(F(f_0) \in CI_{n,\lambda}(\alpha)) \to 1 - \alpha$. This completes the proof of the theorem. \hfill \Box

**Proof of Proposition 5.10.** Proof of (i). By direct examinations, it can be shown that when $\nu \geq 3$ is odd, $\cos(x) \cosh(x) = 1$ has a unique solution in $((\nu+1/2)\pi, (\nu+1)\pi)$, that is, $\gamma_\nu \in ((\nu+1/2)\pi, (\nu+1)\pi)$; when $\nu \geq 3$ is even, $\cos(x) \cosh(x) = 1$ has a unique solution in $(\nu\pi, (\nu+1/2)\pi)$, that is, $\gamma_\nu \in (\nu\pi, (\nu+1/2)\pi)$. Consequently, for any $k \geq 1$, $0 < \gamma_{2k+2} - \gamma_{2k+1} < \pi$.

Let $\delta_0$ be constant such that $0 < \delta_0 < \pi/2 - \pi|z - 1/2|$, and $d_0 = \min\{\sin^2(\delta_0), \cos^2(\delta_0 + \pi|z - 1/2|)\}$. Clearly, $d_0 > 0$ is a constant. It is easy to see that when $k \to \infty$,
\[
\frac{\sinh(\gamma_{2k+1}(z - 1/2))}{\sinh(\gamma_{2k+1}/2)} \to 0, \text{ and } \frac{\cosh(\gamma_{2k+2}(z - 1/2))}{\cosh(\gamma_{2k+2}/2)} \to 0.
\]

Then for arbitrarily small $\varepsilon \in (0, d_0/8)$, there exists $N$ s.t. for any $k \geq N$,
\[
\varphi_{2k+1}(z)^2 \geq \frac{1}{2} \sin^2(\gamma_{2k+1}(z-1/2)) - \varepsilon, \text{ and } \varphi_{2k+2}(z)^2 \geq \frac{1}{2} \cos^2(\gamma_{2k+2}(z-1/2)) - \varepsilon.
\]

Let $\phi_k' = (\gamma_{2k+2} - \gamma_{2k+1})(z - 1/2)$. Then $|\phi_k'| \leq \pi|z - 1/2| < \pi/2$. There exists an integer $l_k$ s.t. $\gamma_{2k+1}(z - 1/2) = \phi_k + l_k\pi$, where $\phi_k \in [0, \pi)$. Then
\[
\sin^2(\gamma_{2k+1}(z - 1/2)) = \sin^2(\phi_k), \text{ and } \cos^2(\gamma_{2k+2}(z - 1/2)) = \cos^2(\phi_k + \phi_k').
\]

If $0 \leq \phi_k \leq \delta_0$, then it can be seen that
\[-\pi|z - 1/2| \leq \phi_k' \leq \phi_k + \phi_k' \leq \delta_0 + \phi_k' \leq \delta_0 + \pi|z - 1/2|.
\]

Therefore, $\cos^2(\phi_k + \phi_k') \geq \cos^2(\delta_0 + \pi|z - 1/2|)$. If $\delta_0 < \phi_k < \pi - \delta_0$, then $\sin^2(\phi_k) \geq \sin^2(\delta_0)$. If $\pi - \delta \leq \phi_k < \pi$, then it can be seen that
\[
\pi - \delta_0 - \pi|z - 1/2| \leq \phi_k + \phi_k' \leq \phi_k + \pi|z - 1/2|.
\]
Therefore, \( \cos^2(\phi_k + \phi'_k) \geq \cos^2(\delta_0 + \pi|z - 1/2|) \). Consequently, for any \( k \geq N \),

\[
\varphi_{2k+1}(z)^2 + \varphi_{2k+2}(z)^2 \geq \frac{1}{2} \left( \sin^2(\gamma_{2k+1}(z - 1/2)) + \cos^2(\gamma_{2k+2}(z - 1/2)) \right) - 2\varepsilon
\]

\[
\geq \frac{1}{2} \min\{\sin^2(\delta_0), \cos^2(\delta_0 + \pi|z - 1/2|)\} - 2\varepsilon \geq d_0/4.
\]

Then we have

\[
\sum_{\nu>2} h \varphi_\nu(z)^2 (1 + \lambda \rho_\nu + (\lambda \rho_\nu)^{1+\beta/4}) \]

\[
= \sum_{k \geq 1} h \varphi_{2k+1}(z)^2 + \sum_{k \geq 1} h \varphi_{2k+2}(z)^2 (1 + \lambda \rho_{2k+2} + (\lambda \rho_{2k+2})^{1+\beta/4}) \]

\[
\geq \sum_{k \geq N} h \varphi_{2k+1}(z)^2 (1 + \lambda \rho_{2k+2} + (\lambda \rho_{2k+2})^{1+\beta/4}) \]

\[
\geq \sum_{k \geq N} \frac{\varphi_{2k+2}(z)^2}{(1 + \lambda \rho_{2k+2} + (\lambda \rho_{2k+2})^{1+\beta/4})} \]

\[
\geq \frac{h}{N} \int_N^{\infty} \frac{1}{(1 + (\pi h x)^4 + (\pi h x)^{4+\beta})} \, dx \]

\[
= \frac{1}{\pi} \int_{\pi Nh}^{\infty} \frac{1}{(1 + x^4 + x^{4+\beta})} \, dx \xrightarrow{n \to \infty} \frac{1}{\pi} \int_{0}^{\infty} \frac{1}{(1 + x^4 + x^{4+\beta})} \, dx > 0.
\]

This shows that condition (5.10) holds for \( r = 1 \).

**Proof of (ii).** Write \( \omega = \sum_\nu \omega_\nu \varphi_\nu \) where \( \omega_\nu \) is a square-summable real sequence. Then \( F_\omega(\varphi_\nu) = \int_0^1 \omega(z) \varphi_\nu(z) \, dz = \omega_\nu \). Therefore, \( \sum_\nu F_\omega(\varphi_\nu)^2 = \sum_\nu \omega_\nu^2 < \infty \). Meanwhile, since \( \omega \neq 0 \), \( \sum_\nu F_\omega(\varphi_\nu)^2 > 0 \). Consequently, for \( j = 1, 2 \), it follows by dominated convergence theorem that as \( n \to \infty \),

\[
\sum_{\nu=1}^m \frac{F_\omega(\varphi_\nu)^2}{(1 + n^{-1} \sigma_\nu^2)} + \sum_{\nu>m} \frac{F_\omega(\varphi_\nu)^2}{(1 + \lambda \rho_\nu + (\lambda \rho_\nu)^{1+\beta/(2m)})} \to \sum_{\nu=1}^\infty F_\omega(\varphi_\nu)^2 > 0.
\]

Hence (5.10) holds for \( r = 0 \).