The Annals of Applied Probability
2022, Vol. 32, No. 2, 1459–1498
https://doi.org/10.1214/21-AAP1716

BERRY–ESSEEN BOUNDS FOR CHERNOFF-TYPE NONSTANDARD ASYMPTOTICS IN ISOTONIC REGRESSION

BY QIYANG HAN[1,a] AND KENGO KATO[2,b]

[1]*Department of Statistics, Rutgers University,* [a]*qh85@stat.rutgers.edu*
[2]*Department of Statistics and Data Science, Cornell University,* [b]*kk976@cornell.edu*

A Chernoff-type distribution is a nonnormal distribution defined by the slope at zero of the greatest convex minorant of a two-sided Brownian motion with a polynomial drift. While a Chernoff-type distribution is known to appear as the distributional limit in many nonregular statistical estimation problems, the accuracy of Chernoff-type approximations has remained largely unknown. In the present paper, we tackle this problem and derive Berry–Esseen bounds for Chernoff-type limit distributions in the canonical nonregular statistical estimation problem of isotonic (or monotone) regression. The derived Berry–Esseen bounds match those of the oracle local average estimator with optimal bandwidth in each scenario of possibly different Chernoff-type asymptotics, up to multiplicative logarithmic factors. Our method of proof differs from standard techniques on Berry–Esseen bounds, and relies on new localization techniques in isotonic regression and an anti-concentration inequality for the supremum of a Brownian motion with a Lipschitz drift.

1. Introduction.

1.1. *Overview.* Nonregular statistical estimation problems constitute a class of estimation problems for which natural estimators converge at a rate different from (often slower than) the parametric rate with nonnormal limit distributions. Such nonregular estimation problems appear in a variety of statistical problems (cf. [58]). An important example of nonnormal limit is a Chernoff-type distribution defined by the slope at zero of the greatest convex minorant of a two-sided Brownian motion with a polynomial drift [20, 47]. Asymptotic theory for Chernoff-type limiting distributions has been well developed so far; however, the accuracy of such Chernoff-type approximations has remained largely unknown, which poses a fundamental question regarding the accuracy of statistical inference in nonregular estimation problems. Indeed, the complicated nature of the Chernoff-type limit makes the problem of establishing rates of convergence for its distributional approximation substantially challenging from a probabilistic point of view.

In the present paper, we tackle this problem and derive Berry–Esseen bounds for Chernoff-type approximations in the canonical example of monotone or isotonic regression. Estimation and inference using regression models under monotonicity constraints has a long history in statistics, as they arise as a natural constraint in diverse application fields from economics, genetics, and to medicine [65–67, 79]. Historical remarks and further references in statistical inference under monotonicity constraints can be found in [47, 78].

Formally, consider the nonparametric regression model

$$Y_i = f_0(X_i) + \xi_i, \quad i = 1, \ldots, n, \tag{1.1}$$

Received December 2019; revised May 2021.
MSC2020 subject classifications. 60E15, 62G05.
Key words and phrases. Berry–Esseen bound, Chernoff's distribution, nonstandard asymptotics, empirical process, anti-concentration.

where $X_1, \dots, X_n \in [0, 1]$ are either fixed or random covariates and $\xi_1, \dots, \xi_n$ are i.i.d. error variables with mean zero and variance $\sigma^2 > 0$ (and are independent of $X_1, \dots, X_n$ if random). By isotonic regression, we assume that f_0 is nondecreasing, that is, $f_0 \in \mathcal{F}_\uparrow \equiv \{f : [0, 1] \to \mathbb{R} : f \text{ is nondecreasing}\}$, and consider the isotonic least squares estimator (LSE):

$$\widehat{f}_n \equiv \underset{f \in \mathcal{F}_\uparrow}{\arg\min} \sum_{i=1}^{n} (Y_i - f(X_i))^2. \tag{1.2}$$

The isotonic LSE constitutes a representative and rich example of nonregular asymptotics. Suppose that $X_1, \dots, X_n$ are globally equally spaced on $[0, 1]$ (i.e., $X_i = i/n$ for $i = 1, \dots, n$) and f_0 is smooth enough at x_0 with a first nonvanishing derivative of order α (α can be $\alpha = \infty$, in which case f_0 is flat). Then, α is an odd integer with $f_0^{(\alpha)}(x_0) > 0$ if α is finite (cf. [52]). Let $c_\alpha \equiv (f_0^{(\alpha)}(x_0)/(\alpha + 1)!)^{1/(2\alpha+1)}$ if $\alpha < \infty$ and $c_\infty \equiv 1$ if $\alpha = \infty$, we have

$$(n/\sigma^2)^{\alpha/(2\alpha+1)} (\widehat{f}_n(x_0) - f_0(x_0)) \xrightarrow{d} c_\alpha \mathbb{D}_\alpha. \tag{1.3}$$

Here $\mathbb{D}_\alpha$ is the slope at zero of the greatest convex minorant of $t \mapsto \mathbb{B}(t) + t^{\alpha+1}$ for $\alpha < \infty$ where $\mathbb{B}$ is a standard two-sided Brownian motion, and $\mathbb{D}_\infty$ is defined in Theorem 2.2 ahead. The canonical case is the $\alpha = 1$ case, where the isotonic LSE has the cube-root $n^{-1/3}$ rate and the limit theorem (1.3) was first proved by [12]. The distribution of $\mathbb{D}_1$ is called the Chernoff distribution, and can be also described as twice the argmax of $t \mapsto \mathbb{B}(t) - t^2$. We shall call the distribution of general $\mathbb{D}_\alpha$ a Chernoff-type distribution. These Chernoff-type distributions are non-Gaussian and fairly complicated. For $\alpha = 1$, the detailed analytical properties of the Chernoff distribution $\mathbb{D}_1$ are investigated in the seminal work of [43]; see also [3, 48].

Limit theorems akin to (1.3) with Chernoff-type limiting distributions appear in a wide range of nonparametric statistical models; see for example, [1, 2, 4, 13, 42, 44, 49, 53, 54, 70, 71, 77, 86, 87], for an incomplete list. Further developments on limit theorems for global loss functions and the law of iterated logarithm can be found in [31–33, 46, 55, 61].

The limit theorem in (1.3) showcases the intrinsic complexity of the nonstandard asymptotics with Chernoff-type distributions in the isotonic regression model (1.1), at least from two different angles: (i) The rate of convergence of the LSE $\widehat{f}_n$, that is, $(n/\sigma^2)^{-\alpha/(2\alpha+1)}$ can *adapt* to the local smoothness level α of the regression function f_0 at x_0; (ii) The limiting distributions $\{\mathbb{D}_\alpha\}$ are different across α's but with certain commonality in terms of being a nonlinear and nonsmooth functional of a Brownian motion with a drift (except for the case $\alpha = \infty$).

The main result of the present paper derives Berry–Esseen bounds for the limit theorem (1.3) in a unified setting. Specifically, we prove that if the error distribution is subexponential and f_0 is smooth enough at x_0 with a first nonvanishing derivative of order α, and a second nonvanishing derivative of order α^*, then

$$\begin{aligned} &\sup_{t \in \mathbb{R}} \left| \mathbb{P}\big((n/\sigma^2)^{\alpha/(2\alpha+1)} (\widehat{f}_n(x_0) - f_0(x_0)) \le t\big) - \mathbb{P}(c_\alpha \mathbb{D}_\alpha \le t) \right| \\ &\quad \lesssim \begin{cases} \big(n^{-\frac{\alpha^*-\alpha}{2\alpha+1}} \vee n^{-\frac{\alpha}{2\alpha+1}}\big) \cdot \operatorname{polylog}(n) & \text{if } \alpha < \infty, \\ n^{-1/2} \cdot \operatorname{polylog}(n) & \text{if } \alpha = \infty \end{cases} \end{aligned} \tag{1.4}$$

up to constants independent of n. In the canonical case of $\alpha = 1$, the bound in (1.4) is of order $n^{-1/3}$ up to logarithmic factors. Another interesting case is the $\alpha = \infty$ case, where the bound achieves nearly the parametric rate $n^{-1/2}$.

The rates given in the Berry–Esseen bounds (1.4) are natural from an oracle perspective. It is useful to recall that the LSE $\widehat{f}_n$ has a well-known representation via the max–min formula (cf. [76]): for $x_0 \in (0,1)$,

$$\widehat{f}_n(x_0) = \max_{u \le x_0} \min_{v \ge x_0} \frac{\sum_{i:u \le X_i \le v} Y_i}{|i : u \le X_i \le v|} \equiv \max_{u \le x_0} \min_{v \ge x_0} \bar{Y}|_{[u,v]} = \bar{Y}|_{[u^*,v^*]}. \tag{1.5}$$

Here $\bar{Y}_A$ is the average of $\{Y_i : i \in A\}$ as defined formally in (1.6) ahead. One can therefore view $\widehat{f}_n(x_0)$ as a local average estimator over the sample in a data-driven *random interval* $[u^*, v^*]$ around x_0. Heuristically, the isotonic LSE automatically learns the bias induced by the first nonvanishing derivative, in the sense that the data-driven bandwidth $|v^* - u^*|$ is of the optimal order $O_P(n^{-1/(2\alpha+1)})$ as that of an oracle local average estimator. Such oracle behavior gives rise to the rate of convergence $O_P(n^{-\alpha/(2\alpha+1)})$ in the limit theorem (1.3). The second nonvanishing derivative of order α^* then quantifies the rate of convergence for the remaining bias in the standardized statistic $n^{-\alpha/(2\alpha+1)}(\widehat{f}_n(x_0) - f_0(x_0))$, yielding the first term $n^{(\alpha^*-\alpha)/(2\alpha+1)}$ in (1.4). On the other hand, the "effective sample" for the isotonic LSE is of order $n_e \equiv n \cdot n^{-1/(2\alpha+1)} = n^{2\alpha/(2\alpha+1)}$, and therefore the speed for the noise $\bar{\xi}|_{[u^*,v^*]}$ to converge in distribution is of order $(n_e)^{-1/2} = n^{-\alpha/(2\alpha+1)}$. This yields the second term in (1.4). These heuristic interpretations on the Berry–Esseen bounds (1.4) also indicate that the adaptation of the isotonic LSE occurs not only at the level of the rate of convergence of $\widehat{f}_n$, but also at the level of the speed of this distributional approximation.

The proof of the Berry–Esseen bounds (1.4) is highly nontrivial reflecting the complexity of the limit theorem (1.3), and our proof strategies differ substantially from existing techniques on Berry–Esseen bounds (see a literature review below). Importantly, in contrast to regular M-estimation problems, the isotonic LSE does not admit an asymptotic linear expansion, nor can be approximated by a simple statistic for which existing techniques on Berry–Esseen bounds are applicable. Our method of proof to establish (1.4) builds on localization techniques in isotonic regression and an *anti-concentration* inequality (Theorem 3.1) for the supremum of a Brownian motion with a Lipschitz drift on a compact interval including the origin. Informally, localization shows that (i) $|n^{\alpha/(2\alpha+1)}(\widehat{f}_n(x_0) - f_0(x_0))| \le O(\sqrt{\log n})$ and (ii) $n^{1/(2\alpha+1)} \max\{|x_0 - u^*|, |v^* - x_0|\} \le O(\sqrt{\log n})$ with high enough probability. The former (i) enables us to restrict the range of t in (1.4) to $|t| \le O(\sqrt{\log n})$, while the latter (ii) enables us to restrict the range of (u, v) in the max–min formula (1.5) to $O(n^{-1/(2\alpha+1)})$ neighborhoods of x_0 up to logarithmic factors. Such localization makes possible the application of the anti-concentration inequality that quantifies the rates of convergence of the bias and the noise to the limit, which are shown to be of the same order as the desired rate in the Berry–Essen bound (1.4), up to multiplicative logarithmic factors. The prescribed proof techniques can be extended to further Chernoff-type limiting distributions in isotonic regression, allowing both interior and boundary points x_0 (cf. [62]); and both fixed and random design covariates.

As discussed before, a key technical ingredient of our proof for the Berry–Esseen bounds is an explicit anti-concentration inequality for the supremum of a standard Brownian motion with a Lipschitz drift, $T = \sup_{t \in [0,1]}(\mathbb{B}(t) + P(t))$, which is of independent interest. The anti-concentration inequality quantifies the modulus of continuity of the distribution function of a random variable, and we need an explicit quantitative anti-concentration inequality of the form $\sup_{u \in \mathbb{R}} \mathbb{P}(|T - u| \le \varepsilon) \lesssim \varepsilon$ up to logarithmic factors to derive the desired Berry–Esseen bounds. The difficulty lies in the fact that the variance of the drifted Brownian motion can be arbitrarily close to zero, so that existing results such as [24], Lemma 2.2, are not applicable, at least directly (in addition, it is highly nontrivial to obtain a density formula for T in this generality). To circumvent this problem, we use a carefully designed blocking argument; see the proof of Theorem 3.1.

The literature on Berry–Essen bounds is broad. Berry–Esseen bounds for the classical central limit theorem (CLT) and its various generalizations to multivariate, high-dimensional, and dependent settings can be found in for example, [5–9, 14, 16, 19, 21, 25, 27, 28, 34–36, 40, 41, 57, 63, 68, 72–75], just to name a few. The techniques developed in those references can not be applied to our problem since the isotonic LSE does not admit an asymptotic linear expansion (and thus has a nonnormal limit). Stein's method [18, 82] is known to be a powerful method to derive rates of convergence of distribution approximations. Recent contributions (e.g., [17, 80]) showcase the possibility of using Stein's method for deriving Berry–Esseen bounds with nonnormal limits that admit explicit and easy-to-handle densities; however, it seems unclear if the complicated Chernoff distribution is within the reach of such methods. To the best of our knowledge, this is the first paper that derives Berry–Esseen bounds for an important class of Chernoff-type limit distributions.

The rest of the paper is organized as follows. In Section 2, we first consider the problem of accuracy of distributional approximation in isotonic regression from an oracle perspective, and then derive the main Berry–Esseen bounds for the isotonic LSE in a unified setup. In Section 3, we prove the key technical result of anti-concentration inequality, and in Section 4, we develop the localization techniques in isotonic regression. Building on the techniques developed in Sections 3 and 4, we prove the main Berry–Esseen bounds in Section 5. In Section 6, we conclude the paper and outline a few open questions. The Appendix contains proofs of some auxiliary results and technical tools used in the proofs.

1.2. *Notation.* For $\varepsilon > 0$ and a subset $\mathcal{F}$ of a normed space with norm $\|\cdot\|$, let $\mathcal{N}(\varepsilon, \mathcal{F}, \|\cdot\|)$ denote the ε-covering number of $\mathcal{F}$; see page 83 of [85] for more details. For the regression model (1.1), for any $A \subset [0, 1]$, define

$$\bar{Y}|_A \equiv \frac{1}{n_A} \sum_{i:X_i \in A} Y_i, \ \bar{f}_0|_A \equiv \frac{1}{n_A} \sum_{i:X_i \in A} f_0(X_i), \ \bar{\xi}|_A \equiv \frac{1}{n_A} \sum_{i:X_i \in A} \xi_i, \tag{1.6}$$

where $n_A \equiv |\{i : X_i \in A\}|$ and $0/0 = 0$ by convention. For two real numbers a, b, $a \vee b \equiv \max\{a, b\}$ and $a \wedge b \equiv \min\{a, b\}$. The notation C_x will denote a generic constant that depends only on x, whose numeric value may change from line to line unless otherwise specified. The notation $a \lesssim_x b$ and $a \gtrsim_x b$ mean $a \le C_x b$ and $a \ge C_x b$ respectively, and $a \asymp_x b$ means $a \lesssim_x b$ and $a \gtrsim_x b$ [$a \lesssim b$ means $a \le Cb$ for some absolute constant C]. The notation $\stackrel{d}{\to}$ is reserved for convergence in distribution.

2. Main results.

2.1. *Assumptions.* We first consider local smoothness assumptions on the regression function f_0 at x_0. We consider both interior ($x_0 \in (0, 1)$) and boundary ($x_0 = 0$) points.

ASSUMPTION A. Let $x_0 \in [0, 1)$ be a fixed point of interest. Let $\alpha, \alpha^* \in \mathbb{Z}_{\ge 1} \cup \{\infty\}$, $\alpha + 1 \le \alpha^* \le \infty$ be such that $f_0^{(\alpha)}(x_0) \neq 0$ and $f_0^{(\alpha^*)}(x_0) \neq 0$ if $\alpha, \alpha^* \neq \infty$, and the Taylor expansion

$$\begin{aligned} f_0(x) = f_0(x_0) &+ \frac{f_0^{(\alpha)}(x_0)}{\alpha!}(x - x_0)^\alpha \mathbf{1}_{\alpha<\infty} \\ &+ \frac{f_0^{(\alpha^*)}(x_0)}{\alpha^*!}(x - x_0)^{\alpha^*} \mathbf{1}_{\alpha^*<\infty} + R\big((x - x_0)^{\alpha^*} \mathbf{1}_{\alpha^*<\infty}\big) \end{aligned}$$

holds for all $x \in [0, 1]$ for some function $R : \mathbb{R} \to \mathbb{R}$ such that $R(0) = 0$ and $R(\varepsilon) = o(\varepsilon)$ as $\varepsilon \to 0$.

If $x_0 = 0$, then the derivatives are understood as one-side limits. Assumption A essentially says that f_0 has a first nonvanishing derivative at x_0 of order α, and a second one of order α^*. If $x_0 \in (0, 1)$, by [52], Lemma 1, α must be an odd integer, and $f_0^{(\alpha)}(x_0) > 0$ under Assumption A. If $x_0 = 0$, α need not be an odd integer, but $f_0^{(\alpha)}(x_0) > 0$. We do not consider $x_0 = 1$ as the situation is similar to $x_0 = 0$.

The following are some examples satisfying Assumption A.

(i) $f_0(x) = x$. Then $\alpha = 1$ and $\alpha^* = \infty$ at $x_0 = 1/2$.
(ii) $f_0(x) = e^x$. Then $\alpha = 1$ and $\alpha^* = 2$ at $x_0 = 1/2$.
(iii) $f_0(x) = (x - 1/2)^3$. Then $\alpha = 3$ and $\alpha^* = \infty$ at $x_0 = 1/2$.
(iv) $f_0(x) = (x - 1/2)^3 + (x - 1/2)^5$. Then $\alpha = 3$ and $\alpha^* = 5$ at $x_0 = 1/2$.
(v) $f_0(x) = x^2 + x^4$. Then $\alpha = 2$ and $\alpha^* = 4$ at $x_0 = 0$.

When $x_0 = 0$, we consider limit distribution theory at $x_n = n^{-\rho}$ where $\rho \in (0, 1)$. Namely, we estimate $f_0(0)$ by $\widehat{f}_n(x_n)$. For notational convenience, let

$$x^* = x_0 \mathbf{1}_{x_0 \in (0,1)} + x_n \mathbf{1}_{x_0 = 0} = x_0 \mathbf{1}_{x_0 \in (0,1)} + n^{-\rho} \mathbf{1}_{x_0 = 0}. \tag{2.1}$$

Next we state assumptions on the design points.

ASSUMPTION B. Suppose that the design points $\{X_i\}_{i=1}^n$ satisfy either of the following conditions.

- *(Fixed design)* $X_1, \ldots, X_n \in [0, 1]$ are deterministic, and there exists some $\Lambda_0 > 0$ such that for some $\delta_0 > 0$, the design points restricted to $I_0 \equiv [0 \vee (x_0 - \delta_0), 1 \wedge (x_0 + \delta_0)]$, $\{X_i : X_i \in I_0, i = 1, \ldots, n\}$, are equally spaced with distance $1/(\Lambda_0 n)$.[1]

  In the case $\alpha = \infty$, or $x_0 = 0$ and $\rho \in [1/(2\alpha + 1), 1)$, we assume that the design points are globally equally spaced on $[0, 1]$ (i.e., $X_i = i/n$, $i = 1, \ldots, n$, so $\Lambda_0 = 1$).
- *(Random design)* $X_1, \ldots, X_n$ are i.i.d. with law P on $[0, 1]$, and P admits a Lebesgue density π that is continuous around x_0 and is bounded and bounded away from 0 on $[0, 1]$. Further assume that for some $1 \le \beta \le \infty$,

  $$\pi(x) - \pi(x_0) = \frac{\pi^{(\beta)}(x_0)}{\beta!}(x - x_0)^\beta \mathbf{1}_{\beta < \infty} + R_\pi\big((x - x_0)^\beta \mathbf{1}_{\beta < \infty}\big)$$

  holds for all $x \in [0, 1]$ for some function $R_\pi : \mathbb{R} \to \mathbb{R}$ such that $R_\pi(0) = 0$ and $R_\pi(\varepsilon) = o(\varepsilon)$ as $\varepsilon \to 0$. Let $\Lambda_0 = \pi(x_0)$.

  In the case $\alpha = \infty$ or $x_0 = 0$, $\rho \in [1/(2\alpha + 1), 1)$, we assume that P is the uniform distribution on $[0, 1]$.

The canonical case is the globally equally spaced fixed design with $X_i = i/n$, $i = 1, \ldots, n$ (so $\Lambda_0 = 1$). Furthermore, we have made more specific assumptions on the designs of the covariates when $\alpha = \infty$, or $x_0 = 0$ and $\rho \in [1/(2\alpha + 1), 1)$ due to the nonlocal nature of the limit distribution theory in such scenarios. This helps us to develop unified Berry–Esseen bounds for the isotonic LSE.

2.2. *Oracle considerations.* To gain some insights into what should be expected for a Berry–Esseen bound for the nonstandard limit theorem (1.3), we shall first look at the prob-

[1] In other words, $X_1 \le X_2 \le \cdots \le X_{n-1} \le X_n$ and $X_{i+1} - X_i = X_i - X_{i-1} = 1/(\Lambda_0 n)$ whenever $X_{i-1}, X_i, X_{i+1} \in I_0$.

lem from an oracle perspective. Suppose that Assumption A holds, and the regularity of f_0 at x_0 is known. Consider the local average estimator

$$\bar{f}_n(x^*) \equiv \bar{f}_n(x^*; r_n, h) = \bar{Y}|_{[x^*-h_1r_n, x^*+h_2r_n]} \tag{2.2}$$

with a tuning parameter $r_n > 0$ and constants $h_1, h_2 > 0$. The isotonic least squares estimator $\widehat{f}_n$, defined via the max–min formula (1.5), can be viewed as a local average estimator (2.2) with automatic data-driven choices of the tuning parameters h_1, h_2, r_n.

An oracle local average estimator $\bar{f}_n$ knows the regularity of f_0 at x_0 and chooses the bandwidth r_n of the following optimal order:

$$r_n \equiv \begin{cases} n^{-1/(2\alpha+1)} & \text{if } x_0 \in (0,1), \\ n^{-(1-2\rho(\alpha-1))/3} & \text{if } x_0 = 0 \text{ and } \rho \in (0, 1/(2\alpha+1)), \\ n^{-\rho} & \text{if } x_0 = 0 \text{ and } \rho \in [1/(2\alpha+1), 1), \end{cases} \tag{2.3}$$

and hence the local rate of convergence of the oracle estimator is given by

$$\omega_n^{-1} \equiv (nr_n)^{1/2} = \begin{cases} n^{\alpha/(2\alpha+1)} & \text{if } x_0 \in (0,1), \\ n^{(1+\rho(\alpha-1))/3} & \text{if } x_0 = 0 \text{ and } \rho \in (0, 1/(2\alpha+1)), \\ n^{(1-\rho)/2} & \text{if } x_0 = 0 \text{ and } \rho \in [1/(2\alpha+1), 1). \end{cases} \tag{2.4}$$

For instance, in the canonical case where $\alpha = 1$ and $x_0 \in (0,1)$, then $r_n = \omega_n = n^{-1/3}$. To describe the limiting distribution of the oracle estimator, further define

$$Q(h) \equiv \begin{cases} \dfrac{f_0^{(\alpha)}(x_0)}{(\alpha+1)!} \cdot h^{\alpha+1}\mathbf{1}_{\alpha<\infty} & \text{if } x_0 \in (0,1), \\ \dfrac{f_0^{(\alpha)}(0)}{2(\alpha-1)!} \cdot h^2 & \text{if } x_0 = 0 \text{ and } \rho \in (0, 1/(2\alpha+1)), \\ \displaystyle\sum_{\ell=1}^{\alpha} \frac{f_0^{(\alpha)}(0)}{(\alpha-\ell)!(\ell+1)!} \cdot h^{\ell+1} & \text{if } x_0 = 0 \text{ and } \rho = 1/(2\alpha+1), \\ 0 & \text{if } x_0 = 0 \text{ and } \rho \in (1/(2\alpha+1), 1), \end{cases} \tag{2.5}$$

and

$$\mathbb{B}_{\sigma,\Lambda_0,Q}(h_1,h_2) \equiv (\sigma/\Lambda_0^{1/2}) \cdot \frac{\mathbb{B}(h_2) - \mathbb{B}(-h_1)}{h_1+h_2} + \frac{Q(h_2) - Q(-h_1)}{h_1+h_2}, \tag{2.6}$$

where $\mathbb{B}$ is a standard two-sided Brownian motion starting from 0.

PROPOSITION 2.1 (Berry–Esseen bounds: Oracle considerations). *Let ξ_i's be i.i.d. errors with finite third moment and $\mathbb{E}\xi_1^2 = \sigma^2$. Suppose Assumptions A and B hold. Then with ω_n^{-1} defined in (2.4) and $\mathbb{B}_{\sigma,\Lambda_0,Q}$ defined in (2.6), the local average estimator $\bar{f}_n$ defined in (2.2) with oracle bandwidth r_n defined in (2.3) satisfies*

$$\sup_{t\in\mathbb{R}} \left| \mathbb{P}\big(\omega_n^{-1}(\bar{f}_n(x^*; r_n, h) - f_0(x^*)) \le t\big) - \mathbb{P}\big(\mathbb{B}_{\sigma,\Lambda_0,Q}(h_1,h_2) \le t\big)\right| \le K \cdot \mathcal{B}_n.$$

The constant $K > 0$ does not depend on n, and with $\mathbf{1}^r$ denoting the indicator for the random design case,

$$
\mathcal{B}_n \equiv \begin{cases} \max\{n^{-\frac{\alpha}{2\alpha+1}} (\log n)^{\mathbf{1}_{\alpha<\infty}\cdot\mathbf{1}^r}, \\ \quad n^{-\frac{\alpha^*-\alpha}{2\alpha+1}} \mathbf{1}_{\alpha^*<\infty}, \\ \quad n^{-\frac{\beta}{2\alpha+1}} \mathbf{1}_{\alpha\vee\beta<\infty}\mathbf{1}^r\} & \text{if } x_0 \in (0,1), \\ \max\{n^{-(1-(2\alpha+1)\rho)/3}, \\ \quad n^{-\rho(\alpha^*-\alpha)}\mathbf{1}_{\alpha^*<\infty}\} & \text{if } x_0 = 0 \text{ and } \rho \in \big(0, 1/(2\alpha+1)\big), \\ \max\{n^{-\frac{\alpha}{2\alpha+1}} (\log n)^{\mathbf{1}_{\alpha<\infty}\cdot\mathbf{1}^r}, \\ \quad n^{-\frac{\alpha^*-\alpha}{2\alpha+1}} \mathbf{1}_{\alpha^*<\infty}, \\ \quad n^{-\frac{\beta}{2\alpha+1}} \mathbf{1}_{\alpha\vee\beta<\infty}\mathbf{1}^r\} & \text{if } x_0 = 0 \text{ and } \rho = 1/(2\alpha+1), \\ \max\{n^{-(1-\rho)/2}(\log n)^{\mathbf{1}_{\alpha<\infty}\cdot\mathbf{1}^r}, \\ \quad n^{-((2\alpha+1)\rho-1)/2}\} & \text{if } x_0 = 0 \text{ and } \rho \in \big(1/(2\alpha+1), 1\big). \end{cases} \tag{2.7}
$$

Furthermore, the above Berry–Esseen bound cannot be improved in general, except for the logarithmic factors in the random design case.

In general, the rate $\mathcal{B}_n$ above is determined by the order of the leading term in the remainders of (2.2) after centering and normalization at the rate ω_n^{-1}. In particular, different terms in the rate $\mathcal{B}_n$ come from different sources in different scenarios:

- For $x_0 \in (0,1)$, $n^{-\frac{\alpha}{2\alpha+1}}$ is the rate for the noise to approximate its Gaussian limit, while $n^{-\frac{\alpha^*-\alpha}{2\alpha+1}}$ is the rate induced by the second nonvanishing derivative of f_0 of order α^* at x_0.
- For $x_0 = 0$ and $\rho \in (0, 1/(2\alpha+1))$, $n^{-(1-(2\alpha+1)\rho)/3}$ is the rate induced by the second order bias (since in this case the first order bias contributes to the limiting distribution), while $n^{-\rho(\alpha^*-\alpha)}$ is the rate induced by the second nonvanishing derivative of f_0 of order α^* at 0. The rate for the noise to approximate its Gaussian limit is dominated by the maximum of the two rates.
- For $x_0 = 0$ and $\rho = 1/(2\alpha+1)$, $n^{-\frac{\alpha}{2\alpha+1}}$ is the rate for the noise to approximate its Gaussian limit, while $n^{-\frac{\alpha^*-\alpha}{2\alpha+1}}$ is the rate induced by the second nonvanishing derivative of f_0 of order α^* at x_0.
- For $x_0 = 0$ and $\rho \in (1/(2\alpha+1), 1)$, $n^{-(1-\rho)/2}$ is the rate for the noise to approximate its Gaussian limit, while $n^{-((2\alpha+1)\rho-1)/2}$ is the rate induced by the first nonvanishing derivative of f_0 of order α^* at x_0 (since in this case $Q \equiv 0$).
- The rates involving β come from the regularity of the design density in the random design setting. They appear when $x_0 \in (0,1)$ or $x_0 = 0$, $\rho = 1/(2\alpha+1)$.

In the next subsection we will show that the isotonic least squares estimator $\widehat{f}_n$ converges to the limiting Chernoff distribution at a rate no slower than the oracle rate $\mathcal{B}_n$, up to logarithmic factors.

PROOF OF PROPOSITION 2.1. First consider the fixed design case with the additional assumption that $x^* \in \{X_i\}$. Applying Lemma 4.1 below in Section 4 with any fixed positive real number $\tau_n \geq h_1 \vee h_2$, for $x_0 \in (0,1)$,

$$
\begin{aligned}
&\bar{f}_0|_{[x_0-h_1r_n, x_0+h_2r_n]} - f_0(x_0) \\
&\quad = \frac{f_0^{(\alpha)}(x_0)}{(\alpha+1)!} \cdot \frac{h_2^{\alpha+1} - h_1^{\alpha+1}}{h_1+h_2} \cdot r_n^\alpha \mathbf{1}_{\alpha<\infty} + O\big(r_n^{\alpha^*}\mathbf{1}_{\alpha^*<\infty} \vee r_n^\alpha (nr_n)^{-1}\mathbf{1}_{\alpha<\infty}\big).
\end{aligned}
$$

For $x_0 = 0$,

$$
\begin{aligned}
&\bar{f}_0|_{[x_n-h_1r_n, x_n+h_2r_n]} - f_0(0)\\
&\quad = f_0^{(\alpha)}(0)\sum_{\ell=1}^{\alpha}\frac{1}{(\alpha-\ell)!(\ell+1)!}\cdot\frac{h_2^{\ell+1}-(-h_1)^{\ell+1}}{h_1+h_2}\cdot x_n^{\alpha-\ell}r_n^{\ell}\mathbf{1}_{\alpha<\infty}\\
&\qquad + O\Big(\max_{1\le\ell\le\alpha^*} x_n^{\alpha^*-\ell}r_n^{\ell}\mathbf{1}_{\alpha^*<\infty}\vee\max_{1\le\ell\le\alpha} x_n^{\alpha-\ell}r_n^{\ell}(nr_n)^{-1}\mathbf{1}_{\alpha<\infty}\Big)\\
&\quad = \begin{cases}
\frac{f_0^{(\alpha)}(0)}{2(\alpha-1)!}\cdot\frac{h_2^2-h_1^2}{h_1+h_2}\cdot x_n^{\alpha-1}r_n\mathbf{1}_{\alpha<\infty} \\
\quad + O\big(x_n^{\alpha-2}r_n^2\mathbf{1}_{\alpha<\infty}\vee x_n^{\alpha^*-1}r_n\mathbf{1}_{\alpha^*<\infty} \\
\quad \vee x_n^{\alpha-1}r_n(nr_n)^{-1}\mathbf{1}_{\alpha<\infty}\big) & \text{if } x_n \gg r_n,\\
f_0^{(\alpha)}(0)\Big(\sum_{\ell=1}^{\alpha}\frac{1}{(\alpha-\ell)!(\ell+1)!}\cdot\frac{h_2^{\ell+1}-(-h_1)^{\ell+1}}{h_1+h_2}\Big)\cdot r_n^{\alpha}\mathbf{1}_{\alpha<\infty}\\
\quad + O\big(r_n^{\alpha^*}\mathbf{1}_{\alpha^*<\infty}\vee r_n^{\alpha}(nr_n)^{-1}\mathbf{1}_{\alpha<\infty}\big) & \text{if } x_n = r_n.
\end{cases}
\end{aligned}
$$

Let $W_n \equiv \sqrt{nr_n}\cdot(\bar{\xi}|_{[x^*-h_1r_n, x^*+h_2r_n]})$,

$$Z_h \equiv \big(\sigma/\Lambda_0^{1/2}\big)\cdot\frac{\mathbb{B}(h_2)-\mathbb{B}(-h_1)}{h_1+h_2}, \quad \text{and} \quad \mu \equiv \frac{Q(h_2)-Q(-h_1)}{h_1+h_2}.$$

Note that $\mathbb{B}_{\sigma,\Lambda_0,Q}(h_1,h_2) \overset{d}{=} Z_h + \mu \sim \mathcal{N}(\mu, \sigma^2/(\Lambda_0(h_1+h_2)))$. Further, let

$$
\mathcal{R}_n^f \equiv \begin{cases}
\Big[\sqrt{nr_n^{2\alpha^*+1}}\mathbf{1}_{\alpha^*<\infty}\vee r_n^{\alpha}(nr_n)^{-1/2}\mathbf{1}_{\alpha<\infty}\Big]\\
\quad \text{if } x_0\in(0,1),\\
\Big[x_n^{\alpha-2}\sqrt{nr_n^5}\mathbf{1}_{\alpha<\infty}\vee x_n^{\alpha^*-1}\sqrt{nr_n^3}\mathbf{1}_{\alpha^*<\infty}\\
\qquad \vee x_n^{\alpha-1}r_n(nr_n)^{-1/2}\mathbf{1}_{\alpha<\infty}\Big]\\
\quad \text{if } x_0=0 \text{ and } \rho\in\big(0,1/(2\alpha+1)\big),\\
\Big[\sqrt{nr_n^{2\alpha^*+1}}\mathbf{1}_{\alpha^*<\infty}\vee r_n^{\alpha}(nr_n)^{-1/2}\mathbf{1}_{\alpha<\infty}\\
\qquad \vee\sqrt{nr_n^{2\alpha+1}}\mathbf{1}_{\rho>1/(2\alpha+1),\alpha<\infty}\Big]\\
\quad \text{if } x_0=0 \text{ and } \rho\in[1/(2\alpha+1),1).
\end{cases} \tag{2.8}
$$

Then, uniformly in $t\in\mathbb{R}$,

$$
\begin{aligned}
&\mathbb{P}\big(\sqrt{nr_n}\big(\bar{f}_n(x^*;r_n,h)-f_0(x^*)\big)\le t\big)\\
&\quad = \mathbb{P}\big(\sqrt{nr_n}(\bar{\xi}|_{[x^*-h_1r_n,x^*+h_2r_n]})+\sqrt{nr_n}\big(\bar{f}_0|_{[x^*-h_1r_n,x^*+h_2r_n]}-f_0(x_0)\big)\le t\big)\\
&\quad = \mathbb{P}\big(W_n+\mu+O\big(\mathcal{R}_n^f\big)\le t\big) = \mathbb{P}\big(Z_h+\mu+O\big(\mathcal{R}_n^f\big)\le t\big)+O\big((nr_n)^{-1/2}\big)\\
&\quad = \mathbb{P}(Z_h+\mu\le t)+O\big(\mathcal{R}_n^f\vee(nr_n)^{-1/2}\big).
\end{aligned}
$$

The second-to-last line follows from the classical Berry–Esseen bound, and the last line follows from the anti-concentration of a standard normal random variable: it holds that $\sup_{t\in\mathbb{R}}\mathbb{P}(|Z-t|\le\varepsilon)\le\varepsilon\sqrt{2/\pi}$ where $Z\sim\mathcal{N}(0,1)$. The remainder term cannot be improved in general by the sharpness of the Berry–Esseen bound for the central limit theorem; cf. [51]. Calculations show that $\mathcal{R}_n^f\vee(nr_n)^{-1/2}=\mathcal{B}_n$ in the fixed design case with $x^*\in\{X_i\}$. For x^* in general position, using Remark 4.2, the error bound is of order at most $(\mathcal{R}_n^f\vee(\omega_n^{-1}\cdot n^{-1}))\vee(nr_n)^{-1/2}=\mathcal{R}_n^f\vee(nr_n)^{-1/2}=\mathcal{B}_n$.

For the random design case, let

$$\mathcal{R}_n^r \equiv \begin{cases} \Big[\sqrt{nr_n^{2\alpha^*+1}}\mathbf{1}_{\alpha^*<\infty} \vee \sqrt{nr_n^{2(\alpha+\beta)+1}}\mathbf{1}_{\alpha\vee\beta<\infty} \\ \quad \vee\sqrt{r_n^{2\alpha}\log n \vee \frac{\log^2 n}{nr_n}}\cdot \mathbf{1}_{\alpha<\infty}\Big] \\ \quad \text{if } x_0 \in (0,1), \\ \Big[x_n^{\alpha-2}\sqrt{nr_n^5}\mathbf{1}_{\alpha<\infty} \vee x_n^{\alpha^*-1}\sqrt{nr_n^3}\mathbf{1}_{\alpha^*<\infty} \\ \quad \vee x_n^{\alpha-1}\sqrt{nr_n^{2\beta+3}}\mathbf{1}_{\alpha\vee\beta<\infty} \\ \quad \vee x_n^{\alpha-1}\sqrt{r_n^2\log n \vee \frac{\log^2 n}{nr_n}}\cdot \mathbf{1}_{\alpha<\infty}\Big] \\ \quad \text{if } x_0 = 0 \text{ and } \rho \in \big(0, 1/(2\alpha+1)\big), \\ \Big[\sqrt{nr_n^{2\alpha^*+1}}\mathbf{1}_{\alpha^*<\infty} \vee \sqrt{nr_n^{2(\alpha+\beta)+1}}\mathbf{1}_{\alpha\vee\beta<\infty} \\ \quad \vee\sqrt{r_n^{2\alpha}\log n \vee \frac{\log^2 n}{nr_n}}\cdot \mathbf{1}_{\alpha<\infty} \\ \quad \vee\sqrt{nr_n^{2\alpha+1}}\mathbf{1}_{\rho>1/(2\alpha+1),\alpha<\infty}\Big] \\ \quad \text{if } x_0 = 0 \text{ and } \rho \in [1/(2\alpha+1), 1). \end{cases} \tag{2.9}$$

Tedious and patient calculations show that $\mathcal{R}_n^r \vee (nr_n)^{-1/2} = \mathcal{B}_n$ in the random design case. For terms involving $\log n$, the bounds cannot be improved by considering $\alpha = \infty$. □

2.3. *Berry–Esseen bounds*. Some further definitions for H_1, H_2 are in Table 1.

Now we present the main results of this paper, that is, Berry–Esseen bounds for (1.3) and its generalizations in isotonic regression.

THEOREM 2.2 (Berry–Esseen bounds for isotonic LSE). *Let ξ_i's be i.i.d. mean-zero subexponential errors, that is, $\mathbb{E}\xi_1 = 0$ and $\mathbb{E}e^{\theta\xi_1} < \infty$ for all θ in a neighborhood of the origin. Let $\sigma^2 \equiv \mathbb{E}\xi_1^2$. Suppose Assumptions A and B hold, and $\rho \in (0, 1/(2\alpha+1)] \cup [2/3, 1)$. Then with ω_n^{-1} defined in (2.4), $\mathbb{B}_{\sigma,\Lambda_0,Q}$ defined in (2.6) and H_1, H_2 defined in Table 1, the isotonic least squares estimator $\widehat{f}_n$ defined in (1.5) satisfies*

$$\sup_{t\in\mathbb{R}}\Big|\mathbb{P}\big(\omega_n^{-1}(\widehat{f}_n(x^*) - f_0(x^*)) \le t\big) - \mathbb{P}\Big(\sup_{h_1\in H_1}\inf_{h_2\in H_2}\mathbb{B}_{\sigma,\Lambda_0,Q}(h_1,h_2) \le t\Big)\Big| \le K\cdot\mathcal{B}_n(\log n)^{\zeta_{\alpha,\alpha^*,\beta}}.$$

TABLE 1
Definitions of H_1, H_2

	H_1	H_2
$\alpha < \infty$	$\begin{cases}(0,\infty) & x_0\in(0,1),\\ (0,\infty) & x_0=0,\rho\in(0,1/(2\alpha+1)),\\ (0,1] & x_0=0,\rho\in[1/(2\alpha+1),1)\end{cases}$	$[0,\infty)$
$\alpha = \infty$	$\begin{cases}(0,x_0] & x_0\in(0,1),\\ (0,1] & x_0=0\end{cases}$	$\begin{cases}[0,1-x_0] & x_0\in(0,1),\\ [0,\infty) & x_0=0\end{cases}$

The constant $K > 0$ does not depend on n, $\mathcal{B}_n$ is defined in (2.7) in the statement of Proposition 2.1, and $\zeta_{\alpha,\alpha^*,\beta} > 0$ is a constant depending only on α, α^*, β.

PROOF. See Section 5. □

It is possible to track the numerical value of $\zeta_{\alpha,\alpha^*,\beta}$ in the proofs, but its value may not be optimal. For brevity, we omit the numerical value of $\zeta_{\alpha,\alpha^*,\beta}$ in the statement of the theorem.

REMARK 2.3 (Limit distributions). The limiting distribution in Theorem 2.2 is written in a compact and unified form which may not be familiar in the literature. We will recover the more familiar forms using the following switching relation: Let H be an (open or closed) interval contained in $\mathbb{R}$, and LCM_H (resp. GCM_H) be the least concave majorant (resp. greatest convex minorant) operator on H and $\mathrm{LCM}_H(\cdot)'$ (resp. $\mathrm{GCM}_H(\cdot)'$) be its left derivative. Then for any $F : H \to \mathbb{R}$, and $t, a \in \mathbb{R}$, we have (cf. [47], Lemma 3.2)

$$\begin{aligned}
\mathrm{LCM}_H(F)'(t) \ge a \quad &\Leftrightarrow \quad \mathrm{GCM}_H(-F)'(t) \le -a \\
&\Leftrightarrow \quad \operatorname*{arg\,min}_{u\in H}\{-F(u) + au\} \ge t \\
&\Leftrightarrow \quad \operatorname*{arg\,max}_{u\in H}\{F(u) - au\} \ge t.
\end{aligned}$$

If there are multiple maxima (resp. minima) in the map $u \mapsto F(u) - au$ (resp. $u \mapsto -F(u) + au$), then the argmax (resp. argmin) is defined to be the location of the first maximum (resp. minimum).

- Let $x_0 \in (0,1)$, $\alpha < \infty$. Then $Q(h) = \frac{f_0^{(\alpha)}(x_0)}{(\alpha+1)!} h^{\alpha+1}$, $H_1 = (0,\infty)$, $H_2 = [0,\infty)$, and we have

$$\begin{aligned}
&\sup_{h_1\in(0,\infty)} \inf_{h_2\in[0,\infty)} \left[\frac{\mathbb{B}(h_2) - \mathbb{B}(-h_1)}{h_1 + h_2} + \frac{Q(h_2) - Q(-h_1)}{h_1 + h_2}\right] \le t \\
&\Leftrightarrow \quad \forall h_1 \in (-\infty, 0), \exists h_2 \in [0,\infty), \\
&\qquad -\mathbb{B}(h_2) - Q(h_2) + th_2 \ge -\mathbb{B}(h_1) - Q(h_1) + th_1 \\
&\Leftrightarrow \quad \operatorname*{arg\,max}_{u\in\mathbb{R}}\big(-\mathbb{B}(u) - Q(u) - (-t)u\big) \ge 0 \\
&\Leftrightarrow \quad \mathrm{LCM}_{\mathbb{R}}\big(-\mathbb{B}(u) - Q(u)\big)'(0) \ge -t \\
&\Leftrightarrow \quad \mathrm{GCM}_{\mathbb{R}}\big(\mathbb{B}(u) + Q(u)\big)'(0) \le t \\
&\overset{d}{\Leftrightarrow} \quad \left(\frac{f_0^{(\alpha)}(x_0)}{(\alpha+1)!}\right)^{1/(2\alpha+1)} \cdot \mathbb{D}_\alpha \le t,
\end{aligned}$$

  where $\mathbb{D}_\alpha$ is the slope at zero of the greatest convex minorant of $t \mapsto \mathbb{B}(t) + t^{\alpha+1}$, and the last equivalence in distribution follows from a standard Brownian scaling argument. In particular, for $\alpha = 1$, we have

$$\mathbb{D}_1 \overset{d}{=} 2 \cdot \operatorname*{arg\,max}_{h\in\mathbb{R}}\{\mathbb{B}(h) - h^2\},$$

  where the argmax on the right hand side is a.s. uniquely defined by [58], Lemma 2.6. See [47], Problem 3.12. The case for $x_0 = 0$, $\rho \in (0, 1/(2\alpha+1))$ is similar as $H_1 = (0,\infty)$, $H_2 = [0,\infty)$ as above.

- Let $x_0 = 0$, $\rho \in (1/(2\alpha+1), 1)$. Then $Q(h) = 0$, $H_1 = (0,1]$, $H_2 = [0,\infty)$, and we have

$$\begin{aligned}
&\sup_{h_1\in(0,1]}\inf_{h_2\in[0,\infty)}\left[\frac{\mathbb{B}(h_2)-\mathbb{B}(-h_1)}{h_1+h_2}\right] \le t \\
&\quad\Leftrightarrow\quad \forall h_1 \in [-1,0), \exists h_2 \in [0,\infty), \quad -\mathbb{B}(h_2) + th_2 \ge -\mathbb{B}(h_1) + th_1 \\
&\quad\Leftrightarrow\quad \operatorname*{arg\,max}_{u\in[-1,\infty)} \bigl(-\mathbb{B}(u) - (-t)u\bigr) \ge 0 \\
&\quad\Leftrightarrow\quad \mathrm{LCM}_{[-1,\infty)}\bigl(-\mathbb{B}(u)\bigr)'(0) \ge -t \\
&\quad\Leftrightarrow\quad \mathrm{GCM}_{[-1,\infty)}\bigl(\mathbb{B}(u)\bigr)'(0) \le t,
\end{aligned}$$

  which takes a similar form as the limiting distribution found in [62], Theorem 3.1-(i) (up to a shift and a recentering of the Brownian motion).

- Let $x_0 = 0$, $\rho = 1/(2\alpha+1)$. Then $Q(h) = f_0^{(\alpha)}(0)\sum_{\ell=1}^{\alpha}\frac{h^{\ell+1}}{(\alpha-\ell)!(\ell+1)!} = \frac{f_0^{(\alpha)}(0)}{(\alpha+1)!}((1+h)^{\alpha+1} - 1 - (\alpha+1)h)$, $H_1 = (0,1]$, $H_2 = [0,\infty)$, and we have

$$\begin{aligned}
&\sup_{h_1\in(0,1]}\inf_{h_2\in[0,\infty)}\left[\frac{\mathbb{B}(h_2)-\mathbb{B}(-h_1)}{h_1+h_2} + \frac{Q(h_2)-Q(-h_1)}{h_1+h_2}\right] \le t \\
&\quad\Leftrightarrow\quad \forall h_1 \in [-1,0), \exists h_2 \in [0,\infty), \\
&\qquad\quad -\mathbb{B}(h_2) - Q(h_2) + th_2 \ge -\mathbb{B}(h_1) - Q(h_1) + th_1 \\
&\quad\Leftrightarrow\quad \operatorname*{arg\,max}_{u\in[-1,\infty)} \left(-\mathbb{B}(u) - \frac{f_0^{(\alpha)}(0)}{(\alpha+1)!}(1+u)^{\alpha+1} - \left(-\frac{f_0^{(\alpha)}(0)}{\alpha!} - t\right)u\right) \ge 0 \\
&\quad\Leftrightarrow\quad \mathrm{LCM}_{[-1,\infty)}\left(-\mathbb{B}(u) - \frac{f_0^{(\alpha)}(0)}{(\alpha+1)!}(1+u)^{\alpha+1}\right)'(0) \ge -\frac{f_0^{(\alpha)}(0)}{\alpha!} - t \\
&\quad\Leftrightarrow\quad \mathrm{GCM}_{[-1,\infty)}\left(\mathbb{B}(u) + \frac{f_0^{(\alpha)}(0)}{(\alpha+1)!}(1+u)^{\alpha+1}\right)'(0) \le \frac{f_0^{(\alpha)}(0)}{\alpha!} + t,
\end{aligned}$$

  which resembles the limiting distribution found in [62], Theorem 3.1-(ii) (again up to a shift and a recentering of the Brownian motion).

The Berry–Esseen bound in Theorem 2.2 matches the oracle rate in Proposition 2.1 up to multiplicative logarithmic factors, and the normal distribution therein is replaced by the generalized Chernoff distribution. In this sense, the isotonic least squares estimator $\widehat{f}_n$ mimics the behavior of the oracle local average estimator in Proposition 2.1 in terms of the speed of distributional approximation to the limiting random variable.

Theorem 2.2 immediately yields the following Berry–Esseen bound in a canonical setting for isotonic regression.

COROLLARY 2.4 (Berry–Esseen bound for canonical case). *Let $x_0 \in (0,1)$ and ξ_i's be as in Theorem 2.2. Suppose Assumption A holds with $\alpha = 1$, $\alpha^* \ge 2$, that is, f_0 is locally C^2 at x_0 with $f_0'(x_0) > 0$, and that $\{X_i : i = 1,\ldots,n\}$ are globally equally spaced design points on* $[0,1]$ *or i.i.d.* Unif$[0,1]$ *random variables independent of ξ_i's. Then*

$$\begin{aligned}
&\sup_{t\in\mathbb{R}}\Bigl|\mathbb{P}\bigl((n/\sigma^2)^{1/3}\bigl(\widehat{f}_n(x_0) - f_0(x_0)\bigr) \le t\bigr) \\
&\qquad - \mathbb{P}\bigl((f_0'(x_0)/2)^{1/3}\cdot\mathbb{D}_1 \le t\bigr)\Bigr| \le K\cdot n^{-1/3}(\log n)^{\zeta_{1,\alpha^*,\infty}}.
\end{aligned}$$

The constant $K > 0$ does not depend on n.

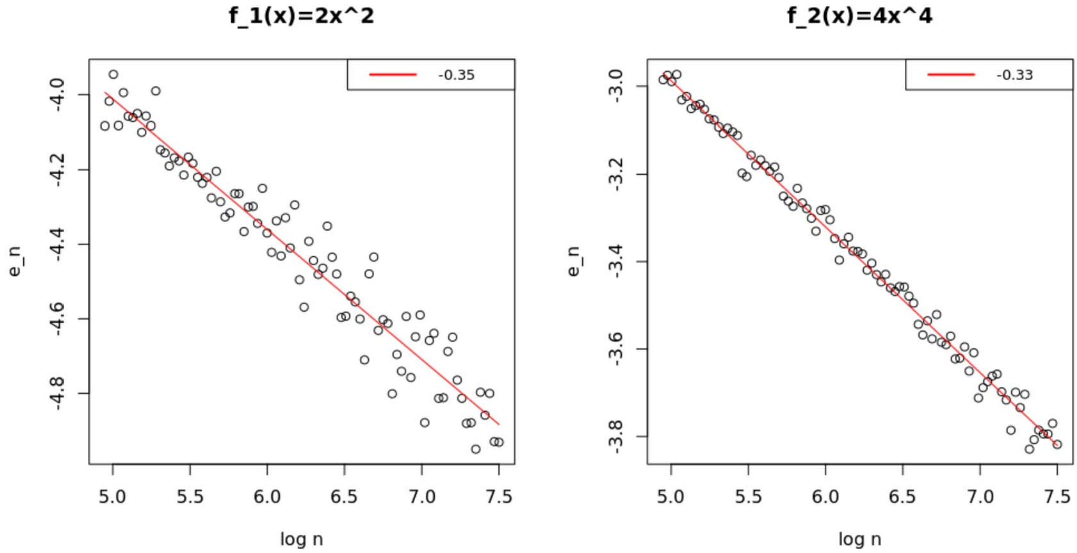

FIG. 1. $E_n \equiv \max_{t\in\{\ell/5:1\le\ell\le10\}}|\mathbb{P}^*(n^{1/3}(\widehat{f}_n(1/2)-f_i(1/2))\le t)-\mathbb{P}(\mathbb{D}_1\le t)|$ *and* $e_n \equiv \log E_n$, *where* $\mathbb{P}^*$ *denotes the empirical average based on* 5×10^5 *simulations. The number in the legend of the figure indicates the slope for linear regression fit of* $(\log n, e_n)$.

PROOF. Apply Theorem 2.2 with $\Lambda_0 = 1$ and $\alpha = 1$ with arbitrary α^*. Here $\beta = \infty$ in the random design case. $\square$

REMARK 2.5 (Simulation experiment). We present a simulation result (cf. Figure 1) in support of the $n^{-1/3}$ rate (modulo logarithmic factors) in the Berry–Esseen bound in Corollary 2.4. In this simulation we consider $f_1(x) = 2x^2$ and $f_2(x) = 4x^4$, and the fixed design as in Corollary 2.4. We use i.i.d. Rademacher errors, that is, $\mathbb{P}(\xi_i = \pm1) = 1/2$. The choice of error distribution is motivated by the fact that the worst-case Berry–Esseen bound for the central limit theorem of sample mean is attained by the Rademacher mean. Under this setup, we have

$$n^{1/3}\big(\widehat{f}_n(1/2)-f_i(1/2)\big) \xrightarrow{d} \mathbb{D}_1, \quad i=1,2.$$

By (limiting) symmetric considerations, we only compute the values of $\mathbb{P}(n^{1/3}(\widehat{f}_n(1/2) - f_i(1/2)) \le t)$ for $t \in \{\ell/5 : 1 \le \ell \le 10\}$ based on 5×10^5 simulations. The values of $\{\mathbb{P}(\mathbb{D}_1 \le t) : t \in \{\ell/5 : 1 \le \ell \le 10\}\}$ are taken from [50] (note that our $\mathbb{D}_1 = 2Z$ in their notation). The simulations provide overwhelming evidence that the Berry–Esseen bound in Corollary 2.4 is sharp modulo logarithmic factors.

Another interesting consequence of Theorem 2.2 is the following: If f_0 is flat (i.e., equals a constant), then a parametric rate (up to logarithmic factors) in the Berry–Esseen bound is possible. We formalize this result as follows.

COROLLARY 2.6 (Berry–Esseen bound for constant function). *Let* $x_0 \in (0,1)$ *and* ξ_i*'s be as in Theorem 2.2. Suppose* $f_0 \equiv c$ *for some constant* $c \in \mathbb{R}$, *and that* $\{X_i : i = 1,\dots,n\}$ *are globally equally spaced design points on* $[0,1]$ *or i.i.d.* Unif$[0,1]$ *random variables independent of* ξ_i*'s. Then*

$$\begin{aligned}\sup_{t\in\mathbb{R}}\Big|&\mathbb{P}\big((n/\sigma^2)^{1/2}(\widehat{f}_n(x_0)-f_0(x_0))\le t\big)\\ &-\mathbb{P}\Big(\sup_{h_1\in(0,x_0]}\inf_{h_2\in[0,1-x_0]}\mathbb{B}_{\sigma,1,0}(h_1,h_2)\le t\Big)\Big| \le K\cdot n^{-1/2}(\log n)^{\zeta_{\infty,\infty,\infty}}.\end{aligned}$$

The constant $K > 0$ *does not depend on* n.

PROOF. Apply Theorem 2.2 with $\Lambda_0 = 1$ and $\alpha = \infty$ (so $\alpha^* = \infty$). Here $\beta = \infty$ in the random design. $\square$

REMARK 2.7 (Boundary case). When $x_0 = 0$, the range of ρ in Theorem 2.2 is restricted to $(0, 1/(2\alpha + 1)] \cup [2/3, 1)$. The main reason for this restriction is an abrupt phase transition in the limit distribution theory. For instance, consider $f_0 \equiv 0$ (i.e. $\alpha = \infty$) with noise level $\sigma = 1$. If $x_0 \in (0, 1)$, $\sqrt{n}\widehat{f}_n(x_0)$ converges in distribution to

$$Y_0 \equiv \sup_{h_1 \in (0, x_0]} \inf_{h_2 \in [0, 1-x_0]} \mathbb{B}_{1,1,0}(h_1, h_2),$$

with a Berry–Esseen bound on the order of $O(n^{-1/2})$ up to logarithmic factors. However, as soon as $x_n \to 0$, $\sqrt{nx_n}\widehat{f}_n(x_n)$ converges in distribution to a completely different limiting random variable

$$Y_1 \equiv \sup_{h_1 \in (0, 1]} \inf_{h_2 \in [0, \infty)} \mathbb{B}_{1,1,0}(h_1, h_2),$$

in the sense that $Y_1 \le 0$ a.s. It is therefore natural to expect that for x_n converging slowly enough, a near $O((nx_n)^{-1/2})$ rate cannot be attained in the Berry–Esseen bound due to the inherent difference between Y_0 and Y_1. Our Theorem 2.2 here guarantees a near $O((nx_n)^{-1/2})$ rate for the Berry–Esseen bound when $x_n = n^{-\rho}$ converges fast enough with $\rho \in [2/3, 1)$.

2.4. *Proof sketch.* In this subsection, we give a sketch of proof for Theorem 2.2 in the canonical case (1.4), where $X_i = i/n$, $i = 1, \ldots, n$ are globally equally spaced fixed design points on $[0, 1]$, f_0 is locally C^2 at $x_0 \in (0, 1)$ with $f_0'(x_0) > 0$, and the errors ξ_i's are i.i.d. mean zero with $\mathbb{E}e^{\theta\xi_1} < \infty$ for θ in a neighborhood of the origin. For simplicity of discussion, we assume that $\mathbb{E}\xi_1^2 = 1$. We reparametrize the max–min formula (1.5) by

$$\begin{aligned} \widehat{f}_n(x_0) &= \max_{h_1 > 0} \min_{h_2 \ge 0} \bar{Y}|_{[x_0 - h_1 n^{-1/3}, x_0 + h_2 n^{-1/3}]} \\ &\equiv \bar{Y}|_{[x_0 - h_1^* n^{-1/3}, x_0 + h_2^* n^{-1/3}]}. \end{aligned} \tag{2.10}$$

The first step in the proof of (1.4) is to localize the isotonic LSE $\widehat{f}_n$ in the sense that for some slowly growing sequences $\{t_n\}$, $\{\tau_n\}$:

- $|n^{1/3}(\widehat{f}_n(x_0) - f_0(x_0))| \le t_n$ and
- $|h_1^*| \vee |h_2^*| \le \tau_n$

hold with overwhelming probability. In fact, we may take t_n, τ_n on the order of $\sqrt{\log n}$ for this purpose; see Lemmas 4.4 and 4.6 ahead.

Next, note that by the Kolmós–Major–Tusnády strong embedding theorem (see Lemma 5.1 ahead), with overwhelming probability,

$$\begin{aligned} \bar{\xi}|_{[x_0 - h_1 n^{-1/3}, x_0 + h_2 n^{-1/3}]} &\approx \frac{\sum_{X_i \in [x_0 - h_1 n^{-1/3}, x_0 + h_2 n^{-1/3}]} \xi_i}{(h_1 + h_2) n^{2/3}} \\ &\approx \frac{\mathbb{B}(h_2 n^{2/3}) + \mathbb{B}(-h_1 n^{2/3})}{(h_1 + h_2) n^{2/3}} \stackrel{d}{=} n^{-1/3} \cdot \frac{\mathbb{B}(h_2) + \mathbb{B}(-h_1)}{h_1 + h_2}, \end{aligned}$$

and by a calculation of the bias via Taylor expansion (see Lemma 4.1 ahead),

$$\bar{f}_0|_{[x_0 - h_1 n^{-1/3}, x_0 + h_2 n^{-1/3}]} - f_0(x_0) \approx n^{-1/3} \left[\frac{f_0'(x_0)}{2} \cdot \frac{h_2^2 - h_1^2}{h_1 + h_2} + R_n \right],$$

where R_n is roughly of order $n^{-1/3}$. Now using the alternative max–min formula (2.10), with $\gamma_0 \equiv f_0'(x_0)/2$, uniformly in $|t| \le t_n$,

$$\begin{aligned}&\mathbb{P}\big(n^{1/3}\big(\widehat{f}_n(x_0) - f_0(x_0)\big) \le t\big)\\ &\quad\approx \mathbb{P}\Big(\max_{0<h_1\le\tau_n}\min_{0\le h_2\le\tau_n}\big(\mathbb{B}(h_2)+\mathbb{B}(-h_1)+\gamma_0\big(h_2^2-h_1^2\big)-t(h_1+h_2)\big) \le \widetilde{O}(R_n)\Big),\end{aligned}$$

where $\widetilde{O}(R_n)$ stands for a term of order R_n up to poly-logarithmic factors. Let $T_{n,1} \equiv \max_{0<h_1\le\tau_n}(\mathbb{B}(-h_1) - \gamma_0 h_1^2 - th_1)$, $T_{n,2} \equiv \min_{0\le h_2\le\tau_n}(\mathbb{B}(h_2)+\gamma_0 h_2^2 - th_2)$, and $\mathcal{L}_i(\varepsilon) \equiv \sup_{u\in\mathbb{R}} \mathbb{P}(|T_{n,i} - u| \le \varepsilon)$. Note that $T_{n,1}$ and $T_{n,2}$ are independent. Then the above display equals

$$\begin{aligned}&\mathbb{P}\big(T_{n,1}+T_{n,2} \le \widetilde{O}(R_n)\big)\\ &\quad\le \mathbb{P}(T_{n,1}+T_{n,2}\le 0) + \min_{i=1,2}\mathcal{L}_i\big(\widetilde{O}(R_n)\big)\\ &\quad= \mathbb{P}\Big(\max_{0<h_1\le\tau_n}\min_{0\le h_2\le\tau_n}\Big(\frac{\mathbb{B}(h_2)+\mathbb{B}(-h_1)}{h_1+h_2} + \gamma_0\cdot\frac{h_2^2-h_1^2}{h_1+h_2}\Big)\le t\Big) + \min_{i=1,2}\mathcal{L}_i\big(\widetilde{O}(R_n)\big)\\ &\quad\approx \mathbb{P}\Big(\max_{h_1>0}\min_{h_2\ge 0}\Big(\frac{\mathbb{B}(h_2)+\mathbb{B}(-h_1)}{h_1+h_2} + \gamma_0\cdot\frac{h_2^2-h_1^2}{h_1+h_2}\Big)\le t\Big) + \min_{i=1,2}\mathcal{L}_i\big(\widetilde{O}(R_n)\big).\end{aligned}$$

The last approximation follows from a similar localization property as in the first step for the isotonic LSE. The first term in the above display is exactly the desired quantity

$$\mathbb{P}\big(\big(f_0'(x_0)/2\big)^{1/3}\cdot\mathbb{D}_1 \le t\big),$$

so it remains to derive a sharp control of

$$\min_{i=1,2}\mathcal{L}_i\big(\widetilde{O}(R_n)\big).$$

This is the *anti-concentration* problem that will be studied in the next Section 3. In particular, Theorem 3.1 below shows that $\min_{i=1,2}\mathcal{L}_i(\widetilde{O}(R_n)) = \widetilde{O}(R_n) = \widetilde{O}(n^{-1/3})$, by noting that $t_n \asymp \sqrt{\log n}$ and $\tau_n \asymp \sqrt{\log n}$ in the localization step (see also Remark 3.4 below). This completes the proof of (1.4) in the regime $|t| \le t_n$. The regime $|t| > t_n$ is already handled by the localization property of the isotonic LSE $\widehat{f}_n$ in the first step.

3. Anti-concentration.

3.1. *The anti-concentration problem.* As discussed in Section 2.4, the proof of our main Berry–Esseen bounds in the canonical case builds on the *anti-concentration* of the random variable $T_n \equiv \sup_{0\le h\le\tau_n}(\mathbb{B}(h)+bh^2+th)$ for certain $\tau_n \uparrow \infty$, that is, an estimate of $\mathcal{L}_{T_n}(\varepsilon) \equiv \sup_{u\in\mathbb{R}}\mathbb{P}(|T_n - u| \le \varepsilon)$, with certain uniformity in t. We note that [45] and [56] derive analytical expressions of the density function of $\sup_{h\ge 0}(\mathbb{B}(h) - \gamma h^2)$ for $\gamma > 0$, but their results are not applicable to our problem since we need anti-concentration bounds on the supremum of a Brownian motion with a linear-quadratic drift on a compact interval. In addition, the proof for the general case in Theorem 2.2 requires, as one of the key technical results, uniform anti-concentration bounds on the supremum of a Brownian motion with a general polynomial drift. Theorem 3.1 below derives such anti-concentration bounds in a more general context for Brownian motion with a Lipschitz drift.

THEOREM 3.1 (Anti-concentration of sup of BM plus a Lipschitz drift). *Let $\mathbb{B}$ be a standard Brownian motion starting from* 0. *Let $P : [0,1] \to \mathbb{R}$ be b-Lipschitz in that $|P(h_1) - P(h_2)| \le b|h_1 - h_2|$ for all $h_1, h_2 \in [0,1]$, and*

$$T \equiv \sup_{0\le h\le 1}\big(\mathbb{B}(h)+P(h)\big). \tag{3.1}$$

Then the following anti-concentration holds: there exists some absolute constant $K > 0$ such that for any $\varepsilon > 0$,

$$\sup_{u\in\mathbb{R}} \mathbb{P}(|T-u| \le \varepsilon) \le K\varepsilon \mathscr{L}_{\bar{b}}(\varepsilon), \tag{3.2}$$

where $\mathscr{L}_{\bar{b}}(\varepsilon) \equiv \bar{b}\log_+(\bar{b}/\varepsilon)(1 \vee \bar{b}\varepsilon\log_+^{-1}(1/\varepsilon))(\bar{b} \vee \log_+(\bar{b}/\varepsilon))$. Here $\log_+(\cdot) \equiv 1 \vee \log(\cdot)$ and $\bar{b} \equiv 1 \vee b$.

PROOF. See the next subsection. □

REMARK 3.2. From log-concavity of Gaussian measures, the distribution of T is absolutely continuous on (r_0, ∞), where r_0 is the left end point of the support of T; see, for example, [29], Theorem 11.1. This also shows that the density of T is bounded on (r, ∞) for any $r > r_0$. This theorem, however, does not guarantee global boundedness of the density of T, and thus does not lead to a quantitative anti-concentration inequality of the form (3.2) (since the variance of the process $h \mapsto \mathbb{B}(h) + P(h)$ attains zero at $h = 0$, [29], Proposition 11.4, is also not applicable). Indeed, as we will discuss in Remark 3.4 ahead, in our application, we need to know how the drift term $P(\cdot)$ quantitatively affects the anti-concentration inequality, and such quantitative information does not follow from [29], Theorem 11.1, or its proof.

REMARK 3.3 (Case with uniformly bounded coefficients). If $\bar{b} \lesssim 1$, then (3.2) in Theorem 3.1 reduces to

$$\sup_{u\in\mathbb{R}} \mathbb{P}(|T-u| \le \varepsilon) \le K\varepsilon \log_+^2(1/\varepsilon).$$

The above bound holds for any Lipschitz function P. If $P = 0$, then by the reflection principle for a Brownian motion, $T = \sup_{0\le h\le 1} \mathbb{B}(h) \overset{d}{=} |Z|$ for $Z \sim \mathcal{N}(0, 1)$, so that the logarithmic factor in the above display can be removed.

REMARK 3.4 (Suprema over slowly expanding intervals). In the proof of Theorem 2.2, we will need anti-concentration for random variables of the form $T_n = \sup_{0\le h\le\tau_n}(\mathbb{B}(h) + \sum_{\ell=1}^{\alpha} b'_\ell h^{\ell+1} - th)$, where $\tau_n \uparrow \infty$ is some slowly growing sequence, and $\bar{b}' \equiv 1 \vee \max_{1\le\ell\le\alpha} b'_\ell$ (typically) does not grow with n. Note that

$$T_n \overset{d}{=} \tau_n^{1/2} \sup_{0\le h'\le 1}\left(\mathbb{B}(h') + \sum_{\ell=1}^{\alpha} b'_\ell \tau_n^{\ell+1/2}(h')^{\ell+1} - t\tau_n^{1/2}h'\right) \equiv \tau_n^{1/2}\cdot T'_n.$$

Hence uniformly in $|t| \le t_n$, where t_n is potentially a slowly growing sequence, we have by Theorem 3.1

$$\begin{aligned}\sup_{u\in\mathbb{R}} \mathbb{P}(|T_n-u| \le \varepsilon) &= \sup_{u\in\mathbb{R}} \mathbb{P}(|T'_n-u| \le \varepsilon/\tau_n^{1/2}) \\ &\le K_{\bar{b}'}\cdot\varepsilon(\tau_n^\alpha \vee t_n)\log_+\left(\frac{\tau_n(\tau_n^\alpha \vee t_n)}{\varepsilon}\right) \\ &\quad\times\left(1 \vee \frac{(\tau_n^\alpha \vee t_n)\varepsilon}{\log_+(\tau_n^{1/2}/\varepsilon)}\right)\left(\tau_n^{1/2}(\tau_n^\alpha \vee t_n) \vee \log_+\left(\frac{\tau_n(\tau_n^\alpha \vee t_n)}{\varepsilon}\right)\right).\end{aligned}$$

For the canonical case $\alpha = 1$, we will take $\tau_n \asymp t_n \asymp \sqrt{\log n}$ as described in Section 2.4, and $\varepsilon = \varepsilon_n$ such that $\log_+(1/\varepsilon_n) \asymp \log n$. Then the above bound reduces to

$$\sup_{u\in\mathbb{R}} \mathbb{P}(|T_n-u| \le \varepsilon_n) \le K_{\bar{b}'}\cdot\varepsilon_n\cdot\log^{5/2} n.$$

For a general α, we will typically take $\tau_n^\alpha \asymp t_n \asymp \sqrt{\log n}$, so the above bound still holds.

REMARK 3.5 (Comparison with small ball problem). The anti-concentration problem considered in Theorem 3.1 is qualitatively different from the small ball problem, cf. [64]. For instance, [64], Theorem 3.1, shows that as $\varepsilon \downarrow 0$,

$$\mathbb{P}\Big(\sup_{0\le h\le 1}\big|\mathbb{B}(h)+P(h)\big|\le\varepsilon\Big)\sim e^{-\|P'\|_{L_2}^2/2}\mathbb{P}\Big(\sup_{0\le h\le 1}\big|\mathbb{B}(h)\big|\le\varepsilon\Big).$$

Using the well-known fact that $\log\mathbb{P}(\sup_{0\le h\le 1}|\mathbb{B}(h)|\le\varepsilon)\sim-(\pi^2/8)\varepsilon^{-2}$ (cf. [64], Theorem 6.3), we have

$$\varepsilon^2\log\mathbb{P}\Big(\sup_{0\le h\le 1}\big|\mathbb{B}(h)+P(h)\big|\le\varepsilon\Big)\sim-\pi^2/8,$$

as $\varepsilon\downarrow 0$, an estimate exhibiting a completely different behavior compared with the anti-concentration bound in Theorem 3.1.

REMARK 3.6 (Anti-concentration inequalities). The anti-concentration inequalities that are in similar in nature to Theorem 3.1 play a pivotal role in establishing Berry–Esseen bounds for central limit theorems on the class of convex sets in the multivariate setting [5] and on hyperrectangles in the high-dimensional setting [22, 23, 25]. In the latter problem, one main ingredient is the anti-concentration for the maximum of jointly Gaussian random variables with uniformly positive variance; cf. Nazarov's inequality [26, 69].

3.2. *Proof of Theorem* 3.1. The proof of Theorem 3.1 relies on several technical results. One is the anti-concentration lemma ([24], Lemma 2.2) for the supremum of a noncentered Gaussian process with *uniformly positive variance*.

LEMMA 3.7. *Let $\{X(t):t\in T\}$ be a possibly noncentered tight Gaussian random variable in $\ell^\infty(T)$. Let $\underline{\sigma}^2\equiv\inf_{t\in T}\mathrm{Var}(X(t))$. Let $d:T\times T\to\mathbb{R}_{\ge 0}$ be a pseudometric defined by $d^2(s,t)\equiv\mathbb{E}(X(s)-X(t))^2$ for $s,t\in T$. Then for any $\varepsilon>0$,*

$$\begin{aligned}&\sup_{u\in\mathbb{R}}\mathbb{P}\Big(\Big|\sup_{t\in T}X(t)-u\Big|\le\varepsilon\Big)\\&\quad\le\inf_{\delta,r>0}\Big[\frac{2}{\underline{\sigma}}\big(\varepsilon+\varpi_X(\delta)+r\delta\big)\big(\sqrt{2\log\mathcal{N}(\delta,T,d)}+2\big)+e^{-r^2/2}\Big],\end{aligned}$$

where $\varpi_X(\delta)\equiv\mathbb{E}\sup_{s,t\in T,d(s,t)\le\delta}|X(s)-X(t)|$.

Unfortunately, we can not directly apply the above anti-concentration bound to our problem since the supremum in (3.1) necessarily involves the Brownian motion at small times, whose variance can be arbitrarily close to zero. In the proof below we will use a carefully designed blocking argument to compensate the large estimate due to small variance incurred by Lemma 3.7, with small estimate for the anti-concentration of the supremum of a Brownian motion with a linear drift. To this end, we will use the following lemma, the proof of which can be found in the Appendix.

LEMMA 3.8 (Density of sup of BM with linear drift). *Let $\mathbb{B}$ be a standard Brownian motion starting from* 0, *and $\mu\in\mathbb{R}$. Let $M_\mu\equiv\sup_{0\le h\le 1}(\mathbb{B}(h)+\mu h)\equiv\sup_{0\le h\le 1}\mathbb{B}_\mu(h)$. Then the Lebesgue density of M_μ, denoted by p_{M_μ}, is given by*

$$p_{M_\mu}(y)=\big[2\varphi(y-\mu)-2\mu e^{2\mu y}\big(1-\Phi(y+\mu)\big)\big]\mathbf{1}_{y\ge 0},\tag{3.3}$$

where $\varphi(\cdot)=(2\pi)^{-1/2}e^{-(\cdot)^2/2}$ and $\Phi(\cdot)$ are the probability density function and cumulative distribution function of the standard normal distribution, respectively. Consequently, $\|p_{M_\mu}\|_\infty\lesssim(\mu\vee 1)$.

PROOF OF THEOREM 3.1. Let $N \equiv \max\{\lfloor K_1 \bar{b}^2 \varepsilon^{-2} \log_+(\bar{b}/\varepsilon)\rfloor + 1, 4\}$ for some constant $K_1 > 0$ to be chosen later. Let $h_\ell \equiv \ell/N$ for $1 \leq \ell \leq N$. Assume without loss of generality $\log_2(N/2+1) \in \mathbb{N}$. Let $L(h) \equiv \mathbb{B}(h) + P(h)$. For $1 \leq j \leq \log_2(N/2+1)$, let $\Omega_j \equiv \{\arg\max_{1\leq \ell \leq N}(\mathbb{B}(h_\ell) + P(h_\ell)) \in \{2^{j-1}, \ldots, 2^j - 1\}\}$. Then

$$
\begin{aligned}
&\mathbb{P}(u-\varepsilon \leq T \leq u+\varepsilon)\\
&\quad \leq \mathbb{P}\Big(\max_{1\leq \ell\leq N/2} L(h_\ell) \in [u-2\varepsilon, u+2\varepsilon]\Big) + \mathbb{P}\Big(\sup_{\substack{h_1,h_2\in[0,1],\\ |h_1-h_2|\leq 1/N}} \big|L(h_1)-L(h_2)\big| > \varepsilon\Big)\\
&\qquad + \mathbb{P}\Big(\sup_{h\in[1/2,1]} L(h) \in [u-\varepsilon, u+\varepsilon]\Big) \equiv (I) + (II) + (III).
\end{aligned}
$$

We first handle relatively easy terms (II) and (III).

For (II), note that for some absolute constant $K_2 > 1$, we may choose $K_1 > 800$ large enough such that

$$
\begin{aligned}
\mathbb{E}\sup_{|h_1-h_2|\leq 1/N}\big|L(h_1)-L(h_2)\big| &\leq \mathbb{E}\sup_{|h_1-h_2|\leq 1/N}\big|\mathbb{B}(h_1)-\mathbb{B}(h_2)\big| + \big|P(h_1)-P(h_2)\big|\\
&\leq K_2\big[\sqrt{\log N/N} + \bar{b}/N\big] \leq \varepsilon/10.
\end{aligned}
$$

The first inequality in the above display uses entropy integral (cf. Lemma B.1) to evaluate the expected supremum. Since

$$
\sup_{|h_1-h_2|\leq 1/N} \operatorname{Var}\big(L(h_1)-L(h_2)\big) \leq 1/N \leq \frac{\varepsilon^2}{\bar{b}^2 K_1 \log_+(\bar{b}/\varepsilon)} \leq \frac{\varepsilon^2}{K_1 \log_+(1/\varepsilon)},
$$

it follows by the Gaussian concentration (cf. Lemma B.2) that

$$
\begin{aligned}
(II) &\leq \mathbb{P}\Big(\sup_{|h_1-h_2|\leq 1/N}\big|\mathbb{B}(h_1)-\mathbb{B}(h_2)\big|\\
&\qquad - \mathbb{E}\sup_{|h_1-h_2|\leq 1/N}\big|\mathbb{B}(h_1)-\mathbb{B}(h_2)\big| > \varepsilon - \bar{b}/N - \varepsilon/10\Big)\\
&\leq \mathbb{P}\Big(\sup_{|h_1-h_2|\leq 1/N}\big|\mathbb{B}(h_1)-\mathbb{B}(h_2)\big| - \mathbb{E}\sup_{|h_1-h_2|\leq 1/N}\big|\mathbb{B}(h_1)-\mathbb{B}(h_2)\big| > \varepsilon/2\Big)\\
&\leq \exp\Big(-\frac{\varepsilon^2/4}{2\varepsilon^2/K_1\log_+(1/\varepsilon)}\Big) = \exp\big(-(K_1/8)\log_+(1/\varepsilon)\big) \leq \varepsilon^{100},
\end{aligned}
$$

by choosing $K_1 > 800$.

For (III), as the minimum standard deviation of $L(h)$ for $h \in [1/2, 1]$ is strictly bounded from below by $1/\sqrt{2}$, we may use the anti-concentration inequality for noncentered Gaussian process as in Lemma 3.7:

$$
\begin{aligned}
(III) \lesssim \inf_{r,\delta>0}\Big[\Big(\varepsilon + \mathbb{E}\sup_{\substack{1/2\leq h_i\leq 1: i=1,2\\ d(h_1,h_2)\leq\delta}}\big|L(h_1)-L(h_2)\big| + r\delta\Big)&\\
\times \big(1 \vee \sqrt{\log \mathcal{N}(\delta, [1/2,1], d)}\big) + e^{-r^2/2}\Big]&,
\end{aligned}
$$

where

$$
\begin{aligned}
d^2(h_1,h_2) = \mathbb{E}\big(L(h_1)-L(h_2)\big)^2 &= \mathbb{E}\big(\mathbb{B}(h_1)-\mathbb{B}(h_2)\big)^2 + \big(P(h_1)-P(h_2)\big)^2\\
&\leq |h_1-h_2| + \bar{b}^2(h_1-h_2)^2 \leq K_3\bar{b}^2 \cdot |h_1-h_2|.
\end{aligned}
$$

Hence by Lemma B.1,

$$\mathbb{E}\sup_{\substack{1/2\leq h_i\leq 1:i=1,2\\ d(h_1,h_2)\leq\delta}}\big|L(h_1)-L(h_2)\big|\lesssim\big[\delta\sqrt{\log(1/\delta)}+\bar{b}\delta^2\big],$$

$$\log\mathcal{N}\big(\delta,[1/2,1],d\big)\leq\log\mathcal{N}\bigg(\frac{\delta^2}{K_3\bar{b}^2},[1/2,1],|\cdot|\bigg)\lesssim\log_+\bigg(\frac{\bar{b}}{\delta}\bigg).$$

Collecting the estimates, by choosing $r\equiv 2\log_+^{1/2}(1/\varepsilon)$ and $\delta\equiv\varepsilon/\sqrt{\log_+(1/\varepsilon)}$, we arrive at

$$\begin{aligned}(III)&\lesssim\inf_{r,\delta>0}\big[\big(\varepsilon+\delta\sqrt{\log(1/\delta)}+\bar{b}\delta^2+r\delta\big)\cdot\log_+^{1/2}(\bar{b}/\delta)+e^{-r^2/2}\big]\\&\lesssim\varepsilon\big(1\vee\bar{b}\varepsilon\log_+^{-1}(1/\varepsilon)\big)\log_+^{1/2}(\bar{b}/\varepsilon).\end{aligned}$$

Finally we handle the most difficult term (I). For each $1\leq j\leq\log_2(N/2+1)$, let $h_{\ell_j^*}\in\{h_\ell:2^{j-1}\leq\ell<2^j\}$ be defined by $L(h_{\ell_j^*})=\max_{2^{j-1}\leq\ell<2^j}L(h_\ell)$. By blocking through the events $\{\Omega_j:1\leq j\leq\log_2(N/2+1)\}$, we have

$$\begin{aligned}(I)&\leq\sum_{j=1}^{\log_2(N/2+1)}\mathbb{P}\big(L(h_{\ell_j^*})\in[u-2\varepsilon,u+2\varepsilon],L(h_k)\leq u+2\varepsilon,2^j<\forall k\leq N\big)\\&\leq\sum_{j=1}^{\log_2(N/2+1)}\mathbb{P}\big(L(h_{\ell_j^*})\in[u-2\varepsilon,u+2\varepsilon],L(h_k)-L(h_{\ell_j^*})\leq 4\varepsilon,2^j<\forall k\leq N\big)\\&\leq\sum_{j=1}^{\log_2(N/2+1)}\mathbb{P}\Big(\max_{2^{j-1}\leq\ell<2^j}L(h_\ell)\in[u-2\varepsilon,u+2\varepsilon],\\&\qquad L(h_k)-L(h_{2^j})\leq 4\varepsilon+\max_{2^{j-1}\leq\ell<2^j}\big|L(h_{2^j})-L(h_\ell)\big|,2^j<\forall k\leq N\Big).\end{aligned}$$

It is not hard to show that, using similar arguments above by calculating the first moment via the entropy integral (cf. Lemma B.1) and Gaussian concentration (cf. Lemma B.2), for some large constant $K_4=K_4(m)>0$,

$$\mathbb{P}\bigg(\max_{2^{j-1}\leq\ell<2^j}\big|L(h_{2^j})-L(h_\ell)\big|>K_4\bigg[\sqrt{\frac{2^j}{N}\log_+\bigg(\frac{N}{\varepsilon 2^j}\bigg)}+\frac{\bar{b}2^j}{N}\bigg]\bigg)\leq\varepsilon^{100}.$$

Hence, we may continue bounding (I) as follows:

$$\begin{aligned}(I)&\leq\sum_{j=1}^{\log_2(N/2+1)}\mathbb{P}\bigg(\max_{2^{j-1}\leq\ell<2^j}L(h_\ell)\in[u-2\varepsilon,u+2\varepsilon],\\&\qquad L(h_k)-L(h_{2^j})\leq 4\varepsilon+K_4\bigg[\sqrt{\frac{2^j}{N}\log_+\bigg(\frac{N}{\varepsilon 2^j}\bigg)}+\frac{\bar{b}2^j}{N}\bigg],2^j<\forall k\leq N\bigg)\\&\qquad+\log_2(N/2+1)\cdot\varepsilon^{100}\\&\leq\sum_{j=1}^{\log_2(N/2+1)}\bigg[\mathbb{P}\Big(\max_{2^{j-1}\leq\ell<2^j}L(h_\ell)\in[u-2\varepsilon,u+2\varepsilon]\Big)\\&\qquad\times\mathbb{P}\bigg(\mathbb{B}(h_k-h_{2^j})\leq 4\varepsilon+\bar{b}(h_k-h_{2^j})\end{aligned}$$

$$
\begin{aligned}
&\quad + K_4\Big[\sqrt{\frac{2^j}{N}\log_+\Big(\frac{N}{\varepsilon 2^j}\Big)} + \frac{\bar{b}2^j}{N}\Big], 2^j < \forall k \le N\Big)\Big] + \log_2(N/2+1)\cdot \varepsilon^{100} \\
&\le \sum_{j=1}^{\log_2(N/2+1)} \Big[\mathbb{P}\Big(\max_{2^{j-1}\le \ell < 2^j} L(h_\ell) \in [u-2\varepsilon, u+2\varepsilon]\Big) \\
&\quad \times \mathbb{P}\Big(\sup_{0\le h\le 1-2^j/N} \big(\mathbb{B}(h) - \bar{b}h\big) \le K_5\Big[\varepsilon + \sqrt{\frac{2^j}{N}\log_+\Big(\frac{N}{\varepsilon 2^j}\Big)} + \frac{\bar{b}2^j}{N}\Big]\Big)\Big] \\
&\quad + 2\log_2(N/2+1)\cdot \varepsilon^{100} \\
&\equiv \sum_{j=1}^{\log_2(N/2+1)} \mathfrak{p}_{j,1}\cdot \mathfrak{p}_{j,2} + 2\log_2(N/2+1)\cdot \varepsilon^{100}.
\end{aligned}
$$

In the last inequality we have expanded the supremum from the discrete set $\{1/N, 2/N, \ldots, 1-2^j/N\}$ to $0 \le h \le 1-2^j/N$ at the cost of a larger constant K_5 and a larger residual probability estimate.

Now following similar calculations as in the derivation of (III) using Lemma 3.7, we have

$$
\mathfrak{p}_{j,1} \lesssim \frac{\varepsilon(1\vee \bar{b}\varepsilon \log_+^{-1}(1/\varepsilon))\log_+^{1/2}(\bar{b}/\varepsilon)}{\sqrt{2^j/N}}.
$$

On the other hand, as the supremum in $\mathfrak{p}_{j,2}$ can be restricted to $[0, 1/4]$ (by noting that $\min_{1\le j\le \log_2(N/2+1)}(1-2^j/N) \ge 1-(N/2+1)/N = 1/2 - 1/N \ge 1/4$ for $N \ge 4$) and is always nonnegative, we have

$$
\begin{aligned}
\mathfrak{p}_{j,2} &\le \mathbb{P}\Big(\sup_{0\le h\le 1/4}\big(\mathbb{B}(h) - \bar{b}h\big) \in \Big[0, K_5\Big\{\varepsilon + \sqrt{\frac{2^j}{N}\log_+\Big(\frac{N}{\varepsilon 2^j}\Big)} + \frac{\bar{b}2^j}{N}\Big\}\Big]\Big) \\
&= \mathbb{P}\Big(\sup_{0\le h\le 1}\big(\mathbb{B}(h) - (\bar{b}/2)h\big) \in \Big[0, 2K_5\Big\{\varepsilon + \sqrt{\frac{2^j}{N}\log_+\Big(\frac{N}{\varepsilon 2^j}\Big)} + \frac{\bar{b}2^j}{N}\Big\}\Big]\Big).
\end{aligned}
$$

By Lemma 3.8, the density of $\sup_{0\le h\le 1}(\mathbb{B}(h) - (\bar{b}/2)h)$ is bounded by $\bar{b}$ up to a constant depending only on m, that is, $\|p_{\sup_{0\le h\le 1}(\mathbb{B}(h)-(\bar{b}/2)h)}\|_\infty \lesssim \bar{b}$, and hence

$$
\mathfrak{p}_{j,2} \lesssim \bar{b}\Big\{\varepsilon + \sqrt{\frac{2^j}{N}\log_+\Big(\frac{N}{\varepsilon 2^j}\Big)} + \frac{\bar{b}2^j}{N}\Big\}.
$$

Collecting the estimates, it follows that with

$$
L_b(\varepsilon) \equiv \varepsilon\big(1\vee \bar{b}\varepsilon \log_+^{-1}(1/\varepsilon)\big)\log_+^{1/2}(\bar{b}/\varepsilon),
$$

we have

$$
\begin{aligned}
(I) &\lesssim \sum_{j=1}^{\log_2(N/2+1)} \frac{L_b(\varepsilon)}{\sqrt{2^j/N}}\cdot \bar{b}\Big\{\varepsilon + \sqrt{\frac{2^j}{N}\log_+\Big(\frac{N}{\varepsilon 2^j}\Big)} + \frac{\bar{b}2^j}{N}\Big\} + 2\log_2(N/2+1)\cdot \varepsilon^{100} \\
&\lesssim \bar{b}\varepsilon L_b(\varepsilon) \sum_{j=1}^{\log_2(N/2+1)} \frac{1}{\sqrt{2^j/N}} + \bar{b}L_b(\varepsilon)\sum_{j=1}^{\log_2(N/2+1)} \log_+^{1/2}(N/\varepsilon 2^j) \\
&\quad + \bar{b}^2 L_b(\varepsilon) \sum_{j=1}^{\log_2(N/2+1)} (2^j/N)^{1/2} + \log_2(N/2+1)\cdot \varepsilon^{100}
\end{aligned}
$$

$$
\begin{aligned}
&\lesssim \bar{b}\sqrt{N}\varepsilon L_b(\varepsilon) + \bar{b}L_b(\varepsilon)\log_+^{3/2}(N/\varepsilon) + \bar{b}^2 L_b(\varepsilon) + \log_2(N/2)\cdot \varepsilon^{100}\\
&\lesssim \bar{b}L_b(\varepsilon)\log_+^{1/2}(\bar{b}/\varepsilon)\big(\bar{b}\vee \log_+(\bar{b}/\varepsilon)\big) + \log_+(\bar{b}/\varepsilon)\cdot\varepsilon^{100}\\
&\lesssim \bar{b}\varepsilon\log_+(\bar{b}/\varepsilon)\big(1\vee \bar{b}\varepsilon\log_+^{-1}(1/\varepsilon)\big)\big(\bar{b}\vee\log_+(\bar{b}/\varepsilon)\big).
\end{aligned}
$$

The calculation above uses that $N \asymp \bar{b}^2\varepsilon^{-2}\log_+(\bar{b}/\varepsilon)$, as chosen in the beginning of the proof. □

4. Localization.

4.1. *Preliminary estimates.* We make a few definitions:

- Let r_n be defined in (2.3) and $\omega_n = (nr_n)^{-1/2}$ be as in (2.4).
- Let $h_1^*, h_2^* > 0$ be random variables defined by $\widehat{f}_n(x_0) \equiv \bar{Y}|_{[x_0-h_1^* r_n, x_0+h_2^* r_n]}$.
- Let $\Omega_n \equiv \{h_1^* \vee h_2^* \le \tau_n\}$ for some $\tau_n > 0$ to be specified below.
- Let $\widetilde{h}_1, \widetilde{h}_2 > 0$ be random variables defined by

$$
\sup_{h_1\in H_1}\inf_{h_2\in H_2}\mathbb{B}_{\sigma,\Lambda_0,Q}(h_1,h_2) \equiv \mathbb{B}_{\sigma,\Lambda_0,Q}(\widetilde{h}_1,\widetilde{h}_2), \tag{4.1}
$$

  where H_1, H_2 are defined in Table 1. Note that $\widetilde{h}_1, \widetilde{h}_2$ are a.s. well defined (cf. Lemma 4.5).
- Let $\widetilde{\Omega}_n \equiv \{\widetilde{h}_1 \vee \widetilde{h}_2 \le \tau_n\}$ for some $\tau_n > 0$ to be specified below.
- For some $t_n > 0$ to be specified below, let $\mathcal{E}_n \equiv \{|\omega_n^{-1}(\widehat{f}_n(x_0) - f_0(x_0))| \le t_n\}$ and $\widetilde{\mathcal{E}}_n \equiv \{|\sup_{h_1\in H_1}\inf_{h_2\in H_2}\mathbb{B}_{\sigma,\Lambda_0,Q}(h_1,h_2)| \le t_n\}$.

For simplicity of notation, we assume $\sigma = 1$ throughout the proof.

The following lemma explicitly calculates the order of the bias.

LEMMA 4.1 (Bias calculation). *Suppose Assumptions A and B hold. In the fixed design setting, further assume that $x^* \in \{X_i\}_{i=1}^n$. Let $r_n \downarrow 0$ for $\alpha < \infty$. Then for $\tau_n \ge 1$ such that $r_n\tau_n^b \downarrow 0$ for any $b > 0$, in both fixed and random designs, the following holds with probability at least $1 - O(n^{-11})$, uniformly in $h_1, h_2 \le \tau_n$:*

1. *If $x_0 \in (0,1)$,*

$$
\begin{aligned}
&(nr_n)^{-1}\sum_{x_0-h_1r_n\le X_i\le x_0+h_2r_n}\big(f_0(X_i) - f_0(x_0)\big)\\
&\quad= \frac{f_0^{(\alpha)}(x_0)}{(\alpha+1)!}\cdot\big(h_2^{\alpha+1} - h_1^{\alpha+1}\big)\cdot\Lambda_0 r_n^\alpha \mathbf{1}_{\alpha<\infty}\\
&\qquad+\begin{cases} O\big(\tau_n^{\alpha^*+1}r_n^{\alpha^*}\mathbf{1}_{\alpha^*<\infty}\vee \tau_n^\alpha r_n^\alpha (nr_n)^{-1}\mathbf{1}_{\alpha<\infty}\big) & \textit{fixed design},\\ O\Bigg(\tau_n^{\alpha^*+1}r_n^{\alpha^*}\mathbf{1}_{\alpha^*<\infty}\vee\tau_n^{\alpha+\beta+1}r_n^{\alpha+\beta}\mathbf{1}_{\alpha\vee\beta<\infty} & \\ \quad\vee\sqrt{\tau_n^{2\alpha+1}r_n^{2\alpha}\dfrac{\log n}{nr_n}\vee\Big(\dfrac{\log n}{nr_n}\Big)^2}\cdot\mathbf{1}_{\alpha<\infty}\Bigg) & \textit{random design}.\end{cases}
\end{aligned}
$$

2. *If $x_0 = 0$, $x_n \downarrow 0$,*

$$
\begin{aligned}
&(nr_n)^{-1}\sum_{x_n-h_1r_n\le X_i\le x_n+h_2r_n}\big(f_0(X_i) - f_0(x_n)\big)\\
&\quad=\sum_{\ell=1}^{\alpha}\frac{f_0^{(\alpha)}(0)}{(\alpha-\ell)!(\ell+1)!}\cdot\big(h_2^{\ell+1} - (-h_1)^{\ell+1}\big)\cdot\Lambda_0 x_n^{\alpha-\ell}r_n^\ell\mathbf{1}_{\alpha<\infty}
\end{aligned}
$$

$$
+\begin{cases} O\Big(\max_{1\le\ell\le\alpha^*}\tau_n^{\ell+1}x_n^{\alpha^*-\ell}r_n^{\ell}\mathbf{1}_{\alpha^*<\infty} \\ \quad\vee\max_{1\le\ell\le\alpha}\tau_n^{\ell}x_n^{\alpha-\ell}r_n^{\ell}(nr_n)^{-1}\mathbf{1}_{\alpha<\infty}\Big) & \textit{fixed design}, \\ O\Big(\max_{1\le\ell\le\alpha^*}\tau_n^{\ell+1}x_n^{\alpha^*-\ell}r_n^{\ell}\mathbf{1}_{\alpha^*<\infty} \\ \quad\vee\max_{1\le\ell\le\alpha}\Big\{\tau_n^{\ell+\beta+1}x_n^{\alpha-\ell}r_n^{\ell+\beta}\mathbf{1}_{\alpha\vee\beta<\infty} \\ \quad\vee x_n^{\alpha-\ell}\sqrt{\tau_n^{2\ell+1}r_n^{2\ell}\frac{\log n}{nr_n}\vee\Big(\frac{\log n}{nr_n}\Big)^2}\cdot\mathbf{1}_{\alpha<\infty}\Big)\Big\} & \textit{random design}. \end{cases}
$$

REMARK 4.2. In the fixed design setting, the assumption $x^*\in\{X_i\}$ can be removed with an additional term of order at most $O(\tau_n/n)$. See the comments on this point in the proof of Lemma 4.1 in the Appendix.

The following lemma gives exponential bounds for the supremum of a weighted partial sum process.

LEMMA 4.3. *Suppose ξ_i's are i.i.d. mean-zero subexponential random variables. Then for both fixed and random design cases, there exists some constant $K>0$ such that for $t\ge 1$,*

$$
\begin{aligned}
&\mathbb{P}\Big(\sup_{h\ge 0}|\bar{\xi}|_{[x^*-hr_n,x^*+r_n]}|>t\omega_n\Big)\vee\mathbb{P}\Big(\sup_{h\ge 0}|\bar{\xi}|_{[x^*-r_n,x^*+hr_n]}|>t\omega_n\Big)\\
&\quad\le K\big(e^{-\{t^2\wedge(nr_n)^{1/2}t\}/K}+n^{-11}\big).
\end{aligned}
$$

Proofs for the above lemmas can be found in the Appendix.

4.2. *Localization.* Recall the events $\mathcal{E}_n$ and $\widetilde{\mathcal{E}}_n$ defined in Section 4.1. The following lemma shows that each of these events has probability $1-O(n^{-11})$ for $t_n\asymp\sqrt{\log n}$.

LEMMA 4.4. *Suppose the conditions in Theorem 2.2 hold. For $t_n=K\sqrt{\log n}$ with some large $K>0$, we have $\mathbb{P}(\mathcal{E}_n^c)\vee\mathbb{P}(\widetilde{\mathcal{E}}_n^c)\le O(n^{-11})$.*

PROOF. First consider $x_0\in(0,1)$. Note that by the max–min formula and monotonicity of f_0,

$$
\begin{aligned}
\widehat{f}_n(x_0)-f_0(x_0)&\le\big(\bar{f}_0|_{[x_0-h_1^*r_n,x_0+r_n]}-f_0(x_0)\big)+\bar{\xi}|_{[x_0-h_1^*r_n,x_0+r_n]}\\
&\le\big(\bar{f}_0|_{[x_0,x_0+r_n]}-f_0(x_0)\big)+\sup_{h\ge 0}|\bar{\xi}|_{[x_0-hr_n,x_0+r_n]}|.
\end{aligned}\tag{4.2}
$$

By Lemma 4.1, in both fixed and random design cases, for n large enough, with probability at least $1-O(n^{-11})$,

$$
\bar{f}_0|_{[x_0,x_0+r_n]}-f_0(x_0)=O\big(r_n^{\alpha}\mathbf{1}_{\alpha<\infty}\big).
$$

On the other hand, by Lemma 4.3 we have for some constant $K>0$, in both fixed and random design cases,

$$
\mathbb{P}\Big(\sup_{h>0}|\bar{\xi}|_{[x_0-hr_n,x_0+r_n]}|>K\omega_n\sqrt{\log n}\Big)\le O\big(n^{-11}\big)
$$

holds for n large enough. Hence with probability at least $1-O(n^{-11})$, $(\omega_n^{-1}(\widehat{f}_n(x_0)-f_0(x_0)))_+\le K_1\sqrt{\log n}$. The other direction can be argued similarly. This proves $\mathbb{P}(\mathcal{E}_n^c)\le$

$O(n^{-11})$. The analogous claim also holds for its limit version by using [52], Lemma 5. We omit the details.

Next suppose $x_0 = 0$ and $\rho \in (0, 1)$. Using (4.2) and Lemma 4.1, we have with the same probability estimate, it holds that

$$\widehat{f}_n(x_n) - f_0(x_n) \le O\Big(\max_{1\le\ell\le\alpha} x_n^{\alpha-\ell} r_n^{\ell} \mathbf{1}_{\alpha<\infty} \vee \omega_n\sqrt{\log n}\Big).$$

The reverse direction is similar. $\square$

Next we will show that each of the events Ω_n and $\widetilde{\Omega}_n$ defined in Section 4.1 has probability $1 - O(n^{-11})$ for some slowly growing sequence τ_n.

LEMMA 4.5. *The random variables $\widetilde{h}_1$ and $\widetilde{h}_2$ in (4.1) are a.s. well defined.*

The proof of the above technical lemma can be found in the Appendix.

LEMMA 4.6. *Suppose the conditions in Theorem 2.2 hold. For $\alpha < \infty$ and*

$$\tau_n \equiv K \cdot \begin{cases} (\log n)^{1/2\alpha} & \text{if } x_0 \in (0,1), \\ \log^{1/2} n & \text{if } x_0 = 0 \text{ and } \rho \in \big(0, 1/(2\alpha+1)\big), \\ (\log n)^{1/2\alpha} & \text{if } x_0 = 0 \text{ and } \rho = 1/(2\alpha+1) \end{cases}$$

with a sufficiently large $K > 0$, we have $\mathbb{P}(\Omega_n^c) \vee \mathbb{P}(\widetilde{\Omega}_n^c) \le O(n^{-11})$.

PROOF. First consider $x_0 \in (0, 1)$. Let $K_1 > 0$ be the constant in Lemma 4.4. Consider the event $\{h_2^* \ge \tau_n\}$. On this event, by max–min formula, we have

$$\begin{aligned} \widehat{f}_n(x_0) - f_0(x_0) &\ge \bar{f}_0|_{[x_0-r_n, x_0+h_2^* r_n]} - f_0(x_0) + \bar{\xi}|_{[x_0-r_n, x_0+h_2^* r_n]} \\ &\ge \big(\bar{f}_0|_{[x_0-r_n, x_0+\tau_n r_n]} - f_0(x_0)\big) - \sup_{h\ge 0}\big|\bar{\xi}|_{[x_0-r_n, x_0+h r_n]}\big|. \end{aligned}$$

The bias term is easy to compute: by Lemma 4.1, in both fixed and random design cases, with probability at least $1 - O(n^{-11})$,

$$\bar{f}_0|_{[x_0-r_n, x_0+\tau_n r_n]} - f_0(x_0) = O\bigg(\frac{(\tau_n^{\alpha+1} - 1)}{\tau_n + 1} r_n^{\alpha}\bigg) \ge c_0 \tau_n^{\alpha} r_n^{\alpha}$$

holds for some constant $c_0 = c_0(\alpha, f_0, x_0) > 0$ and n large enough. On the other hand, by using again Lemma 4.3, we conclude that with probability at least $1 - O(n^{-11})$, $\sup_{h\ge 0}|\bar{\xi}|_{[x_0-r_n, x_0+hr_n]}| \le K_2\omega_n\sqrt{\log n}$. We choose $\tau_n \equiv ((K_1 + K_2)\sqrt{\log n}/c_0)^{1/\alpha}$. Combining the above estimates, on the intersection of $\{h_2^* \ge \tau_n\}$ and an event with probability at least $1 - O(n^{-11})$, we have

$$\omega_n^{-1}\big(\widehat{f}_n(x_0) - f_0(x_0)\big) \ge c_0\tau_n^{\alpha} - K_2\sqrt{\log n} \ge K_1\sqrt{\log n},$$

which occurs with probability at most $O(n^{-11})$ by Lemma 4.4. Hence $\mathbb{P}(h_2^* \ge \tau_n) \le O(n^{-11})$ for n large enough. Similar considerations apply to h_1^*, and the limit versions. Details are omitted.

Next consider $x_0 = 0$ with $\rho \in (0, 1/(2\alpha + 1)]$. Using Lemma 4.1, we have

$$\begin{aligned} \bar{f}_0|_{[x^*-r_n, x^*+\tau_n r_n]} - f_0(x^*) &= O\bigg(\sum_{\ell=1}^{\alpha} \frac{\tau_n^{\ell+1} - (-1)^{\ell+1}}{\tau_n + 1} \cdot x_n^{\alpha-\ell} r_n^{\ell}\bigg) \\ &\ge c_1 \max_{1\le\ell\le\alpha} x_n^{\alpha-\ell}\tau_n^{\ell} r_n^{\ell} = c_1 \begin{cases} x_n^{\alpha-1}\tau_n r_n & \rho \in \big(0, 1/(2\alpha+1)\big), \\ \tau_n^{\alpha} r_n^{\alpha} & \rho = 1/(2\alpha+1) \end{cases} \end{aligned}$$

for some $c_1 = c_1(\alpha, f_0)$, which holds in both fixed and random design settings with probability at least $1 - O(n^{-11})$. The claim now follows from similar arguments above. $\square$

5. Proof of Theorem 2.2.

5.1. *Proof for the fixed design.* In addition to the anti-concentration inequality and localization, the Kolmós–Major–Tusnády (KMT) strong embedding theorem [59, 60] will play an important role. The formulation below is taken from [15].

LEMMA 5.1 (KMT strong embedding). *Let $\xi_1, \ldots, \xi_n$ be i.i.d. mean-zero, unit variance, and subexponential random variables, that is, $\mathbb{E}\xi_1 = 0$, $\mathbb{E}\xi_1^2 = 1$, and $\mathbb{E}e^{\theta\xi_1} < \infty$ for all θ in a neighborhood of the origin. Then for each n, a version of $(S_k \equiv \sum_{i=1}^k \xi_i)_{1\le k\le n}$ and a standard Brownian motion $(\mathbb{B}_n(t))_{0\le t\le n}$ can be constructed on the same probability space such that for all $x \ge 0$,*

$$\mathbb{P}\Big(\max_{1\le k\le n} \big|S_k - \mathbb{B}_n(k)\big| \ge C\log n + x\Big) \le K\exp(-x/K).$$

Here the constants $C, K > 0$ depend on the distribution of ξ_1 only.

PROOF OF THEOREM 2.2: $x_0 \in (0,1)$ OR $x_0 = 0$, $0 < \rho \le 1/(2\alpha+1)$, $\alpha < \infty$. As in the proof of Proposition 2.1, we first work with the extra condition that $x^* \in \{X_i\}$. Then for any $|t| \le t_n$, by max–min formula we have

$$\begin{aligned}
&\mathbb{P}\big(\omega_n^{-1}\big(\widehat{f}_n(x^*) - f_0(x^*)\big) \le t\big)\\
&\quad = \mathbb{P}\Big(\max_{h_1>0}\min_{h_2\ge 0} \omega_n^{-1}\big(\bar{\xi}|_{[x^*-h_1r_n, x^*+h_2r_n]} + \bar{f}_0|_{[x^*-h_1r_n, x^*+h_2r_n]} - f_0(x^*)\big) \le t\Big)\\
&\quad \le \mathbb{P}\Bigg(\max_{\substack{0<h_1\le\tau_n\\ h_1\in H_1}} \min_{\substack{0\le h_2\le\tau_n\\ h_2\in H_2}} \Bigg[\omega_n \sum_{x^*-h_1r_n\le X_i\le x^*+h_2r_n} \xi_i\\
&\qquad + \big(\lfloor h_1r_n\cdot n\rfloor + \lfloor h_2r_n\cdot n\rfloor + 1\big)\omega_n\big(\bar{f}_0|_{[x^*-h_1r_n, x^*+h_2r_n]} - f_0(x^*)\big)\\
&\qquad - \big(\lfloor h_1r_n\cdot n\rfloor + \lfloor h_2r_n\cdot n\rfloor + 1\big)\omega_n^2 t\Bigg] \le 0\Bigg) + \mathbb{P}(\Omega_n^c)\\
&\quad \le \mathbb{P}\Bigg(\max_{\substack{0<h_1\le\tau_n\\ h_1\in H_1}} \min_{\substack{0\le h_2\le\tau_n\\ h_2\in H_2}} \Bigg[\omega_n \sum_{x^*-h_1r_n\le X_i\le x^*+h_2r_n} \xi_i + Q(h_2) - Q(-h_1)\\
&\qquad - t(h_1+h_2)\Bigg] \le O\big(\omega_n^2 t_n \vee \mathcal{R}_n^f\cdot \tau_n^{(\alpha^*+1)\mathbf{1}_{\alpha^*<\infty} + \alpha\mathbf{1}_{\alpha<\infty}}\big)\Bigg) + \mathbb{P}(\Omega_n^c).
\end{aligned}$$

Here $\mathcal{R}_n^f$ is defined in (2.8). The inequality in the last line of the above display follows since by Lemma 4.1,

$$\begin{aligned}
&\big(\lfloor h_1r_n\cdot n\rfloor + \lfloor h_2r_n\cdot n\rfloor + 1\big)\omega_n\big(\bar{f}_0|_{[x^*-h_1r_n, x^*+h_2r_n]} - f_0(x^*)\big)\\
&\quad = (nr_n)^{1/2}\cdot(nr_n)^{-1} \sum_{x^*-h_1r_n\le X_i\le x^*+h_2r_n} \big(f_0(X_i) - f_0(x^*)\big)\\
&\quad = Q(h_2) - Q(-h_1) + O\big(\mathcal{R}_n^f\cdot\tau_n^{(\alpha^*+1)\mathbf{1}_{\alpha^*<\infty} + \alpha\mathbf{1}_{\alpha<\infty}}\big),
\end{aligned}$$

and $(\lfloor h_1r_n\cdot n\rfloor + \lfloor h_2r_n\cdot n\rfloor + 1)\omega_n^2 t = t(h_1+h_2) + O(\omega_n^2 t_n)$. By the KMT strong embedding (cf. Lemma 5.1), there exist independent Brownian motions $\mathbb{B}_n$, $\mathbb{B}_n'$ such that for some constant $C_0 > 0$ that does not depend on n, with probability $1 - O(n^{-11})$, uniformly in $h_1, h_2 \geq 0$,

$$\Big|\sum_{x^* \leq X_i \leq x^* + h_2 r_n} \xi_i - \mathbb{B}_n\big(\lfloor h_2 r_n \cdot n \rfloor + 1\big)\Big| \vee \Big|\sum_{x^* - h_1 r_n \leq X_i < x^*} \xi_i - \mathbb{B}_n'\big(\lfloor h_1 r_n \cdot n \rfloor\big)\Big| \leq C_0 \log n.$$

This means that, with

$$\mathscr{R}_n^f \equiv \max\big\{\mathcal{R}_n^f \cdot \tau_n^{(\alpha^*+1)\mathbf{1}_{\alpha^*<\infty} \vee \alpha \mathbf{1}_{\alpha<\infty}}, \omega_n^2\big(t_n \vee \sqrt{\log n}\big), \omega_n \log n\big\}, \tag{5.1}$$

we have

$$\begin{aligned}
&\mathbb{P}\big(\omega_n^{-1}\big(\widehat{f}_n(x^*) - f_0(x^*)\big) \leq t\big) \\
&\quad \leq \mathbb{P}\Big(\max_{\substack{0<h_1\leq\tau_n \\ h_1\in H_1}} \min_{\substack{0\leq h_2\leq\tau_n \\ h_2\in H_2}} \big[\omega_n \mathbb{B}\big(\lfloor h_2 r_n \cdot n\rfloor + 1\big) - \omega_n \mathbb{B}\big(-\lfloor h_1 r_n \cdot n\rfloor\big) \\
&\qquad + Q(h_2) - Q(-h_1) - t(h_1 + h_2)\big] \leq K_0 \mathscr{R}_n^f\Big) + \mathbb{P}\big(\Omega_n^c\big) + O\big(n^{-11}\big) \\
&\quad \leq \mathbb{P}\Big(\max_{\substack{0<h_1\leq\tau_n \\ h_1\in H_1}} \min_{\substack{0\leq h_2\leq\tau_n \\ h_2\in H_2}} \big[\big(\mathbb{B}(h_2) - \mathbb{B}(-h_1)\big) \\
&\qquad + Q(h_2) - Q(-h_1) - t(h_1 + h_2)\big] \leq K_1 \mathscr{R}_n^f\Big) + \mathbb{P}\big(\Omega_n^c\big) + O\big(n^{-11}\big).
\end{aligned} \tag{5.2}$$

The last inequality follows since by Lemma B.1,

$$\mathbb{E} \sup_{\substack{0\leq h_i\leq\tau_n, i=1,2 \\ |h_1-h_2|\leq\omega_n^2}} \big|\mathbb{B}(h_1) - \mathbb{B}(h_2)\big| \lesssim \omega_n \sqrt{\log n}, \qquad \sup_{\substack{0\leq h_i\leq\tau_n, i=1,2 \\ |h_1-h_2|\leq\omega_n^2}} \mathrm{Var}\big(\mathbb{B}(h_1) - \mathbb{B}(h_2)\big) \leq \omega_n^2$$

and hence by the Gaussian concentration (cf. Lemma B.2), we have for a large enough constant $C_1 > 0$,

$$\mathbb{P}\Big(\sup_{\substack{0\leq h_i\leq\tau_n, i=1,2 \\ |h_1-h_2|\leq\omega_n^2}} \big|\mathbb{B}(h_1) - \mathbb{B}(h_2)\big| > C_1 \omega_n \sqrt{\log n}\Big) \leq e^{-C_1^2 \omega_n^2 \log n / 8\omega_n^2} \leq O\big(n^{-11}\big).$$

Let

$$\begin{aligned}
T_{n,1} &\equiv \max_{0\leq h\leq\tau_n, h\in H_1} \big[-\mathbb{B}(-h) - Q(-h) - th\big], \\
T_{n,2} &\equiv \min_{0\leq h\leq\tau_n, h\in H_2} \big[\mathbb{B}(h) + Q(h) - th\big] = -\max_{0\leq h\leq\tau_n, h\in H_2} \big[-\mathbb{B}(h) - Q(h) + th\big].
\end{aligned}$$

By (5.2), we have

$$\mathbb{P}\big(\omega_n^{-1}\big(\widehat{f}_n(x^*) - f_0(x^*)\big) \leq t\big) \leq \mathbb{P}\big(T_{n,1} + T_{n,2} \leq K_1 \mathscr{R}_n^f\big) + \mathbb{P}\big(\Omega_n^c\big) + O\big(n^{-11}\big).$$

Now apply the anti-concentration Theorem 3.1 with the following choices of $(t_n, \tau_n, \bar{b}, \varepsilon)$:

- $x_0 \in (0, 1)$: $t_n \asymp \sqrt{\log n}$, $\tau_n \asymp (\log n)^{1/2\alpha}$, $\bar{b} \asymp \tau_n^{\alpha+1/2} \asymp \log^{(\alpha+1/2)/2\alpha} n$, $\varepsilon \asymp \mathscr{R}_n^f / \sqrt{\tau_n}$,
- $x_0 = 0$, $\rho \in (0, 1/(2\alpha + 1))$: $t_n, \tau_n \asymp \sqrt{\log n}$, $\bar{b} \asymp \tau_n^{3/2}$, $\varepsilon \asymp \mathscr{R}_n^f / \sqrt{\tau_n}$,
- $x_0 = 0$, $\rho = 1/(2\alpha + 1)$: $t_n \asymp \sqrt{\log n}$, $\tau_n \asymp (\log n)^{1/2\alpha}$, $\bar{b} \asymp \tau_n^{\alpha+1/2} \asymp \log^{(\alpha+1/2)/2\alpha} n$, $\varepsilon \asymp \mathscr{R}_n^f / \sqrt{\tau_n}$.

Then, in view of Remark 3.4, we see that for any $|t| \leq t_n$,

$$
\begin{aligned}
&\mathbb{P}\big(\omega_n^{-1}\big(\widehat{f}_n(x^*) - f_0(x^*)\big) \leq t\big) \\
&\quad = \mathbb{P}(T_{n,1} + T_{n,2} \leq 0) + K_2 \mathscr{R}_n^f \log^{5/2} n + \mathbb{P}\big(\Omega_n^c\big) + O\big(n^{-11}\big) \\
&\quad \leq \mathbb{P}\Big(\max_{\substack{0 < h_1 \leq \tau_n \\ h_1 \in H_1}} \min_{\substack{0 \leq h_2 \leq \tau_n \\ h_2 \in H_2}} \big[\mathbb{B}(h_2) - \mathbb{B}(-h_1) + Q(h_2) - Q(-h_1) - t(h_1 + h_2)\big] \leq 0\Big) \\
&\qquad + K_2 \mathscr{R}_n^f \log^{5/2} n + \mathbb{P}\big(\Omega_n^c\big) + O\big(n^{-11}\big) \\
&\quad \leq \mathbb{P}\Big(\max_{\substack{0 < h_1 \leq \tau_n \\ h_1 \in H_1}} \min_{\substack{0 \leq h_2 \leq \tau_n \\ h_2 \in H_2}} \frac{\mathbb{B}(h_2) - \mathbb{B}(-h_1) + Q(h_2) - Q(-h_1)}{h_1 + h_2} \leq t\Big) \\
&\qquad + K_2 \mathscr{R}_n^f \log^{5/2} n + \mathbb{P}\big(\Omega_n^c\big) + O\big(n^{-11}\big) \\
&\quad \leq \mathbb{P}\Big(\sup_{h_1 \in H_1} \inf_{h_2 \in H_2} \mathbb{B}_{\sigma, \Lambda_0, Q}(h_1, h_2) \leq t\Big) + K_2 \mathscr{R}_n^f \log^{5/2} n + \mathbb{P}\big(\Omega_n^c\big) + \mathbb{P}\big(\widetilde{\Omega}_n^c\big) + O\big(n^{-11}\big).
\end{aligned}
$$

Recalling the definitions of $\mathcal{E}_n$ and $\widetilde{\mathcal{E}}_n$ and arguing the reverse direction similarly, we have

$$
\begin{aligned}
&\sup_{t \in \mathbb{R}} \Big| \mathbb{P}\big(\omega_n^{-1}\big(\widehat{f}_n(x^*) - f_0(x^*)\big) \leq t\big) - \mathbb{P}\Big(\sup_{h_1 \in H_1} \inf_{h_2 \in H_2} \mathbb{B}_{\sigma, \Lambda_0, Q}(h_1, h_2) \leq t\Big)\Big| \\
&\quad \leq K_2 \mathscr{R}_n^f \log^{5/2} n + \mathbb{P}\big(\Omega_n^c\big) + \mathbb{P}\big(\widetilde{\Omega}_n^c\big) + \mathbb{P}\big(\mathcal{E}_n^c\big) + \mathbb{P}\big(\widetilde{\mathcal{E}}_n^c\big) + O\big(n^{-11}\big).
\end{aligned}
$$

The claim of the theorem under $x^* \in \{X_i\}$ now follows from Lemmas 4.4 and 4.6. For x^* in general position, by Remark 4.2, in the definition (5.1) of $\mathscr{R}_n^f$, the quantity $\mathcal{R}_n^f$ need be replaced by $\mathcal{R}_n^f \vee (\omega_n^{-1} \cdot (\tau_n/n))$. The contribution of the additional maximum can be assimilated into the $\omega_n \log n$ term in (5.1), so the above display remains valid. $\square$

PROOF OF THEOREM 2.2: $x_0 = 0$, $1/(2\alpha + 1) < \rho < 1$, $\alpha < \infty$.. We only consider the case $x^* \in \{X_i\}$. In this regime, Lemma 4.6 does not apply so we do not have exponential localization in $h_i^*, \widetilde{h}_i$, $i = 1, 2$. However, Lemma 4.4 still applies, and we do have sub-Gaussian localization of the statistics $\omega_n^{-1}(\widehat{f}_n(x_n) - f_0(x_n))$ and the limiting distribution

$$
\sup_{h_1 \in (0,1]} \inf_{h_2 \in [0,\infty)} \mathbb{B}_{1,1,0}(h_1, h_2) = \sup_{h_1 \in (0,1]} \inf_{h_2 \in [0,\infty)} \frac{\mathbb{B}(h_2) - \mathbb{B}(-h_1)}{h_1 + h_2}.
$$

Hence, for any $|t| \leq t_n$, repeating the arguments in the previous proof, with $\bar{T}_{n,1} \equiv \max_{h \in [0,1]}(-\mathbb{B}(-h) - th)$ and $\bar{T}_{n,2} \equiv \min_{h \in [0,(1-x_n)/x_n)}(\mathbb{B}(h_2) - th_2)$,

$$
\begin{aligned}
&\mathbb{P}\big(\omega_n^{-1}\big(\widehat{f}_n(x_n) - f_0(x_n)\big) \leq t\big) \\
&\quad \leq \mathbb{P}\Big(\max_{h_1 \in (0,1]} \min_{h_2 \in [0,(1-x_n)/x_n)} \big[\mathbb{B}(h_2) - \mathbb{B}(-h_1) - t(h_1 + h_2)\big] \leq K_1 \mathscr{R}_n^f\Big) + O\big(n^{-11}\big) \\
&\quad = \mathbb{P}\big(\bar{T}_{n,1} \leq -\bar{T}_{n,2} + K_1 \mathscr{R}_n^f\big) + O\big(n^{-11}\big) \qquad (5.3) \\
&\quad \leq \mathbb{P}(\bar{T}_{n,1} + \bar{T}_{n,2} \leq 0) + K_2 \mathscr{R}_n^f \log^{5/2} n + O\big(n^{-11}\big) \\
&\quad = \mathbb{P}\Big(\sup_{h_1 \in (0,1]} \inf_{h_2 \in [0,(1-x_n)/x_n)} \mathbb{B}_{1,1,0}(h_1, h_2) \leq t\Big) + K_2 \mathscr{R}_n^f \log^{5/2} n + O\big(n^{-11}\big),
\end{aligned}
$$

where in the first inequality we used independence between $\bar{T}_{n,1}$ and $\bar{T}_{n,2}$, and anti-concentration Theorem 3.1 for $\bar{T}_{n,1}$ with $\bar{b} \asymp \sqrt{\log n}$. Let $\mathcal{E}_{n,2} \equiv \{h_2^* \geq (1 - x_n)/x_n\}$. Then on $\mathcal{E}_{n,2}$,

$$\sup_{h_1\in(0,1]}\inf_{h_2\in[0,\infty)}\mathbb{B}_{1,1,0}(h_1,h_2)\geq \inf_{h_2\geq(1-x_n)/x_n}\frac{\mathbb{B}(h_2)-\mathbb{B}(-1)}{h_2+1}$$
$$\geq -\sup_{h_2\geq n^\rho/2}\left|\frac{\mathbb{B}(h_2)}{h_2}\right|-\frac{|\mathbb{B}(-1)|}{n^\rho/2+1}\stackrel{d}{=}-\frac{Y_1}{(n^\rho/2)^{1/2}}-\frac{Y_2}{n^\rho/2+1},$$

where $Y_1\equiv\sup_{h\geq1}|\mathbb{B}(h)/h|$ and $Y_2\equiv|\mathbb{B}(-1)|$ are nonnegative and have sub-Gaussian tails. Hence on the intersection of $\mathcal{E}_{n,2}$ and an event with probability at least $1-O(n^{-11})$, we have

$$\sup_{h_1\in(0,1]}\inf_{h_2\in[0,\infty)}\mathbb{B}_{1,1,0}(h_1,h_2)\geq -O\big(n^{-\rho/2}\sqrt{\log n}\big).$$

By Lemma A.1, $\mathbb{P}(\mathcal{E}_{n,2})\leq O(n^{-\rho/4}\log^{1/2}n)$. Combined with (5.3), this means that

$$\begin{aligned}&\mathbb{P}\big(\omega_n^{-1}\big(\widehat{f}_n(x_n)-f_0(x_n)\big)\leq t\big)\\&\quad\leq\mathbb{P}\Big(\sup_{h_1\in(0,1]}\inf_{h_2\in[0,\infty)}\mathbb{B}_{1,1,0}(h_1,h_2)\leq t\Big)+K_2\mathscr{R}_n^f\log^{5/2}n+O\big(n^{-\rho/4}\log^{1/2}n\big).\end{aligned}$$

The inequality above can be reversed (note here that from (5.3) one may directly enlarge the range of inf to $h_2\in[0,\infty)$; but this argument does not work for the reversed direction). The claim now follows as the last term can be assimilated when $\rho\in[2/3,1)$. □

PROOF OF THEOREM 2.2: $\alpha=\infty$.. We only consider the case $x^*\in\{X_i\}$. First consider $x_0\in(0,1)$. This case follows quite straightforwardly: with $T_{n,1}\equiv\max_{h\in[0,x_0]}(-\mathbb{B}(-h)-th)$ and $T_{n,2}\equiv\min_{h\in[0,1-x_0]}(\mathbb{B}(h_2)-th_2)$, for any $|t|\leq t_n$, $t_n\asymp\sqrt{\log n}$, we have

$$\begin{aligned}&\mathbb{P}\big(\omega_n^{-1}\big(\widehat{f}_n(x_0)-f_0(x_0)\big)\leq t\big)\\&\quad\leq\mathbb{P}\Big(\max_{h_1\in(0,x_0]}\min_{h_2\in[0,1-x_0]}\big[\mathbb{B}(h_2)-\mathbb{B}(-h_1)-t(h_1+h_2)\big]\leq K_1\mathscr{R}_n^f\Big)+O\big(n^{-11}\big)\\&\quad=\mathbb{P}\big(T_{n,1}+T_{n,2}\leq K_1\mathscr{R}_n^f\big)+O\big(n^{-11}\big)\\&\quad\leq\mathbb{P}(T_{n,1}+T_{n,2}\leq 0)+K_2\mathscr{R}_n^f\log^{5/2}n+O\big(n^{-11}\big)\\&\quad=\mathbb{P}\Big(\sup_{h_1\in H_1}\inf_{h_2\in H_2}\mathbb{B}_{1,1,0}(h_1,h_2)\leq t\Big)+K_2\mathscr{R}_n^f\log^{5/2}n+O\big(n^{-11}\big).\end{aligned}$$

Next consider $x_0=0$. By similar arguments as in the previous proof for the case $\alpha<\infty$, $x_0=0$, $1/(2\alpha+1)<\rho<1$, we have

$$\begin{aligned}&\mathbb{P}\big(\omega_n^{-1}\big(\widehat{f}_n(x_n)-f_0(x_n)\big)\leq t\big)\\&\quad=\mathbb{P}\Big(\sup_{h_1\in(0,1]}\inf_{h_2\in[0,(1-x_n)/x_n)}\mathbb{B}_{1,1,0}(h_1,h_2)\leq t\Big)+K_2\mathscr{R}_n^f\log^{5/2}n+O\big(n^{-11}\big)\\&\quad\leq\mathbb{P}\Big(\sup_{h_1\in(0,1]}\inf_{h_2\in[0,\infty)}\mathbb{B}_{1,1,0}(h_1,h_2)\leq t\Big)+K_2\mathscr{R}_n^f\log^{5/2}n+O\big(n^{-\rho/4}\log^{1/2}n\big),\end{aligned}$$

the last term of which can be assimilated for $\rho\in[2/3,1)$. □

5.2. *Proof for the random design.*

PROOF OF THEOREM 2.2: RANDOM DESIGN. The proof strategy is broadly similar to the fixed design case, but differs quite substantially in technical details due to the randomness of $\{X_i\}$.

First consider the case $x_0 \in (0,1)$ or $x_0 = 0$, $0 < \rho \le 1/(2\alpha+1)$, $\alpha < \infty$. Note that

$$
\begin{aligned}
&\mathbb{P}\big(\omega_n^{-1}\big(\widehat{f}_n(x^*) - f_0(x^*)\big) \le t\big)\\
&\le \mathbb{P}\bigg(\max_{0<h_1\le\tau_n}\min_{0\le h_2\le\tau_n}\bigg[\omega_n \sum_{x^*-h_1r_n \le X_i \le x^*+h_2r_n} \xi_i \\
&\quad + (n\mathbb{P}_n\mathbf{1}_{[x^*-h_1r_n, x^*+h_2r_n]})\omega_n\big(\bar{f}_0|_{[x^*-h_1r_n,x^*+h_2r_n]} - f_0(x^*)\big)\\
&\quad - (\omega_n^2 n\mathbb{P}_n\mathbf{1}_{[x^*-h_1r_n,x^*+h_2r_n]})t\bigg] \le 0\bigg) + \mathbb{P}(\Omega_n^c).
\end{aligned}
\tag{5.4}
$$

By Lemma 4.1, with probability at least $1 - O(n^{-11})$,

$$
\begin{aligned}
&(n\mathbb{P}_n\mathbf{1}_{[x^*-h_1r_n,x^*+h_2r_n]})\omega_n\big(\bar{f}_0|_{[x^*-h_1r_n,x^*+h_2r_n]} - f_0(x^*)\big)\\
&\quad = Q(h_2) - Q(-h_1) + O\big(\mathcal{R}_n^r \cdot \tau_n^{\zeta^r}\big).
\end{aligned}
\tag{5.5}
$$

Here $\mathcal{R}_n^r$ is defined in (2.9) and

$$
\zeta^r \equiv \zeta^r_{\alpha,\alpha^*,\beta} \equiv (\alpha^*+1)\mathbf{1}_{\alpha^*<\infty} \vee (\alpha+\beta+1)\mathbf{1}_{\alpha\vee\beta<\infty} \vee (\alpha+1/2).
\tag{5.6}
$$

Combining (5.4)–(5.5), we have

$$
\begin{aligned}
&\mathbb{P}\big(\omega_n^{-1}\big(\widehat{f}_n(x^*) - f_0(x^*)\big) \le t\big)\\
&\le \mathbb{P}\bigg(\max_{0<h_1\le\tau_n}\min_{0\le h_2\le\tau_n}\bigg[\omega_n \sum_{x^*-h_1r_n \le X_i \le x^*+h_2r_n} \xi_i + Q(h_2) - Q(-h_1)\\
&\quad - (\omega_n^2 n\mathbb{P}_n\mathbf{1}_{[x^*-h_1r_n,x^*+h_2r_n]})t\bigg] \le O\big(\mathcal{R}_n^r \cdot \tau_n^{\zeta^r}\big)\bigg) + \mathbb{P}(\Omega_n^c) + O\big(n^{-11}\big).
\end{aligned}
\tag{5.7}
$$

By the KMT strong embedding, conditionally on $\{X_i\}$'s, there exist independent Brownian motions $\mathbb{B}_n$, $\mathbb{B}_n'$ (which may depend on $\{X_i\}$) such that for some constant $C_0 > 0$ that does not depend on n or $\{X_i\}$,

$$
\begin{aligned}
&\mathbb{P}\bigg(\sup_{h_2>0}\bigg|\sum_{x^*\le X_i\le x^*+h_2r_n}\xi_i - \mathbb{B}_n(n\mathbb{P}_n\mathbf{1}_{[x^*,x^*+h_2r_n]})\bigg|\\
&\qquad \vee \sup_{h_1>0}\bigg|\sum_{x^*-h_1r_n\le X_i<x^*}\xi_i - \mathbb{B}_n'(n\mathbb{P}_n\mathbf{1}_{[x^*-h_1r_n,x^*)})\bigg| \le C_0\log n \,|\{X_i\}\bigg)\\
&\ge 1 - O\big(n^{-11}\big).
\end{aligned}
$$

We do not compare directly $\mathbb{B}_n(n\mathbb{P}_n\mathbf{1}_{[x^*,x^*+h_2r_n]})$ with $\mathbb{B}(h_2nr_n)$ as in the fixed design case, as the standard deviation of $n\mathbb{P}_n\mathbf{1}_{[x^*,x^*+h_2r_n]}$ is of order $\sqrt{nr_n} = \omega_n^{-1}$ and therefore the comparison of Brownian motions leads to suboptimal error bounds. We use a different reparametrization idea as follows. Let $h_{1,n} \equiv \omega_n^2 n\mathbb{P}_n\mathbf{1}_{[x^*-h_1r_n,x^*)}$ and $h_{2,n} \equiv \omega_n^2 n\mathbb{P}_n\mathbf{1}_{[x^*,x^*+h_2r_n]}$. Let

$$
\mathcal{E}_{n,1} \equiv \Big\{\sup_{\substack{0\le h_i\le\tau_n,\\ i=1,2}} \big|(h_{1,n}+h_{2,n}) - (h_1+h_2)\big| \le C_1\omega_n^2\sqrt{n\tau_n r_n\log n} = C_1\omega_n\sqrt{\tau_n\log n}\Big\}.
$$

Then for $C_1 > 0$ large enough, $\mathbb{P}(\mathcal{E}_{n,1}^c) \le O(n^{-11})$. Let $\tau_{1,n} \equiv \omega_n^2 n\mathbb{P}_n\mathbf{1}_{[x^*-\tau_nr_n,x^*)}$ and $\tau_{2,n} \equiv \omega_n^2 n\mathbb{P}_n\mathbf{1}_{[x^*,x^*+\tau_nr_n]}$. On the event $\mathcal{E}_{n,1}$, we have $\tau_{1,n} \ge \tau_n - C_1\omega_n\sqrt{\tau_n\log n} \ge \tau_n/2$ and $\tau_{2,n} \le$

$\tau_n + C_1\omega_n\sqrt{\tau_n \log n} \le 2\tau_n$ for n large enough. Therefore, by (5.7), we have

$$
\begin{aligned}
&\mathbb{P}\big(\omega_n^{-1}\big(\widehat{f}_n(x^*) - f_0(x^*)\big) \le t\big) \\
&\quad \le \mathbb{P}\Big(\max_{\substack{0<h_{1,n}\le\tau_{n,1},\\ h_{1,n}\in\omega_n^2\mathbb{Z}}} \min_{\substack{0\le h_{2,n}\le\tau_{n,2},\\ h_{2,n}\in\omega_n^2\mathbb{Z}}} \big[\mathbb{B}(h_{2,n}) - \mathbb{B}(-h_{1,n}) + Q(h_{2,n}) - Q(-h_{1,n}) \\
&\qquad - t(h_{1,n}+h_{2,n})\big] \le O\big(\mathcal{R}_n^r \cdot \tau_n^{\zeta^r} \vee \omega_n(\tau_n \log n)^{1/2}\tau_n^{\alpha} \vee \omega_n \log n\big), \mathcal{E}_{n,1}\Big) \\
&\qquad + \mathbb{P}\big(\Omega_n^c\big) + O\big(n^{-11}\big) \\
&\quad \le \mathbb{P}\Big(\max_{\substack{0<h_{1,n}\le\tau_n/2,\\ h_{1,n}\in\omega_n^2\mathbb{Z}}} \min_{\substack{0\le h_{2,n}\le 2\tau_n,\\ h_{2,n}\in\omega_n^2\mathbb{Z}}} \big[\mathbb{B}(h_{2,n}) - \mathbb{B}(-h_{1,n}) + Q(h_{2,n}) - Q(-h_{1,n}) \\
&\qquad - t(h_{1,n}+h_{2,n})\big] \le O\big(\mathcal{R}_n^r \cdot \tau_n^{\zeta^r} \vee \omega_n(\tau_n \log n)^{1/2}\tau_n^{\alpha}\big)\Big) + \mathbb{P}\big(\Omega_n^c\big) + O\big(n^{-11}\big).
\end{aligned}
$$

The discretization effect in the above max–min formula can be handled in the O term up to a further probability estimate on the order of $O(n^{-11})$ (for Brownian motion), so we obtain

$$
\begin{aligned}
&\mathbb{P}\big(\omega_n^{-1}\big(\widehat{f}_n(x^*) - f_0(x^*)\big) \le t\big) \\
&\quad \le \mathbb{P}\Big(\max_{0<h_{1,n}\le\tau_n/2}\ \min_{0\le h_{2,n}\le 2\tau_n} \big[\mathbb{B}(h_{2,n}) - \mathbb{B}(-h_{1,n}) + Q(h_{2,n}) - Q(-h_{1,n}) \\
&\qquad - t(h_{1,n}+h_{2,n})\big] \le O\big(\mathcal{R}_n^r \cdot \tau_n^{\zeta^r} \vee \omega_n(\tau_n \log n)^{1/2}\tau_n^{\alpha}\big)\Big) + \mathbb{P}\big(\Omega_n^c\big) + O\big(n^{-11}\big).
\end{aligned}
$$

Now we proceed to argue as in the fixed design case, except for $\mathscr{R}_n^f$ defined in (5.1) is now replaced by

$$
\mathscr{R}_n^r \equiv \mathcal{R}_n^r \cdot \tau_n^{\zeta^r} \vee \omega_n(\tau_n \log n)^{1/2}\tau_n^{\alpha},
$$

where ζ^r is defined in (5.6). This completes the proof of the case $x_0 \in (0,1)$ or $x_0 = 0$, $0 < \rho \le 1/(2\alpha+1)$, $\alpha < \infty$.

For the remaining cases, we only consider $x_0 = 0$, $1/(2\alpha+1) < \rho < 1$, $\alpha < \infty$ as other cases are simpler. As $Q = 0$, we no longer need to work on the event $\mathcal{E}_{n,1}$. Let $\tau_{1,n}^* \equiv \omega_n^2 n\mathbb{P}_n\mathbf{1}_{[0,x^*)}$ and $\tau_{2,n}^* \equiv \omega_n^2 n\mathbb{P}_n\mathbf{1}_{[x^*,1]}$. Then using Bernstein's inequality, it is easy to see that with probability at least $1 - O(n^{-11})$, $\tau_{1,n}^* \ge 1 - O(\omega_n\sqrt{\log n})$ and $\tau_{2,n}^* \le 2n^{\rho}$ for n large enough. Hence, for $|t| \le t_n$,

$$
\begin{aligned}
&\mathbb{P}\big(\omega_n^{-1}\big(\widehat{f}_n(x^*) - f_0(x^*)\big) \le t\big) \\
&\quad \le \mathbb{P}\Big(\max_{0<h_1\le 1}\ \min_{0\le h_2\le(1-x_n)/x_n} \Big[\omega_n \sum_{x^*-h_1r_n\le X_i\le x^*+h_2r_n} \xi_i \\
&\qquad - \big(\omega_n^2 n\mathbb{P}_n\mathbf{1}_{[x^*-h_1r_n,x^*+h_2r_n]}\big)t\Big] \le O\big(\mathcal{R}_n^r \cdot \tau_n^{\zeta^r}\big)\Big) \\
&\quad \le \mathbb{P}\Big(\max_{\substack{0<h_{1,n}\le\tau_{n,1}^*,\\ h_1\in\omega_n^2\mathbb{Z}}} \min_{\substack{0\le h_{2,n}\le\tau_{n,2}^*,\\ h_{2,n}\in\omega_n^2\mathbb{Z}}} \big[\mathbb{B}(h_{2,n}) - \mathbb{B}(-h_{1,n}) - t(h_{1,n}+h_{2,n})\big] \\
&\quad \le O\big(\mathcal{R}_n^r \cdot \tau_n^{\zeta^r} \vee \omega_n \log n\big)\Big) \\
&\quad \le \mathbb{P}\Big(\max_{0<h_{1,n}\le 1-O(\omega_n\sqrt{\log n})}\ \min_{0\le h_{2,n}\le 2n^{\rho}} \big[\mathbb{B}(h_{2,n}) - \mathbb{B}(-h_{1,n}) - t(h_{1,n}+h_{2,n})\big] \\
&\quad \le O\big(\mathcal{R}_n^r \cdot \tau_n^{\zeta^r} \vee \omega_n \log n\big)\Big) + O\big(n^{-11}\big)
\end{aligned}
$$

$$\leq \mathbb{P}\Big(\max_{0<h_{1,n}\leq 1}\min_{0\leq h_{2,n}\leq 2n^{\rho}}\big[\mathbb{B}(h_{2,n})-\mathbb{B}(-h_{1,n})-t(h_{1,n}+h_{2,n})\big]$$

$$\leq O\big(\mathcal{R}_n^r\cdot\tau_n^{\zeta^r}\vee\omega_n(\log n\vee t_n\sqrt{\log n})\big)\Big)+O\big(n^{-11}\big).$$

Here the last line follows by noting that for $|t|\leq t_n$, with probability at least $1-O(n^{-11})$,

$$\max_{0<h_{1,n}\leq 1-O(\omega_n\sqrt{\log n})}\big(\mathbb{B}(-h_{1,n})-th_{1,n}\big)$$

$$\stackrel{d}{=}\big(1-O(\omega_n\sqrt{\log n})\big)^{1/2}\max_{0<h_{1,n}\leq 1}\big(\mathbb{B}(-h_{1,n})-\big(1-O(\omega_n\sqrt{\log n})\big)^{1/2}th_{1,n}\big)$$

$$=\big(1-O(\omega_n\sqrt{\log n})\big)^{1/2}\max_{0<h_{1,n}\leq 1}\big(\mathbb{B}(-h_{1,n})-th_{1,n}\big)+O\big(\omega_n\sqrt{\log n}\cdot t_n\big)$$

$$=\max_{0<h_{1,n}\leq 1}\big(\mathbb{B}(-h_{1,n})-th_{1,n}\big)+O\big(\omega_n\sqrt{\log n}\big)\cdot O\big(\sqrt{\log n}\big)+O\big(\omega_n\sqrt{\log n}\cdot t_n\big).$$

Now we may proceed as in the fixed design case to conclude. $\square$

6. Concluding remarks and open questions. In this paper we developed a new approach of proving Berry–Esseen bounds for Chernoff-type nonstandard limit theorems in the isotonic regression model, by combining problem-specific localization techniques and an anti-concentration inequality for the supremum of a Brownian motion with a Lipschitz drift. The scope of the techniques applies to various known (or near-known) Chernoff-type nonstandard asymptotics in isotonic regression allowing (i) general local smoothness conditions on the regression function, (ii) limit theorems both for interior points and points approaching the boundary, and (iii) both fixed and random design covariates.

Below we sketch two further open questions.

QUESTION 1. *Prove a matching lower bound for the cube-root rate* (*in the canonical case* $\alpha=1$) *in the Berry–Esseen bound* (1.4).

As demonstrated in the simulation (Figure 1), the oracle perspective (cf. Proposition 2.1) is quite informative in that the cube-root rate in (1.4) cannot be improved when the errors are i.i.d. Rademacher random variables. [51] used Stein's method to prove a lower bound of order $n^{-1/2}$, for the Berry–Esseen bound for the central limit theorem for the sample mean in the worst-case scenario. Unfortunately, the least squares estimator (1.2) is a highly nonlinear and nonsmooth functional of the samples in the isotonic regression model (1.1), and therefore the connection between the Stein's method and the Berry–Esseen bound for the nonstandard limit theorem (1.4) remains largely unknown. New techniques seem necessary for proving a lower bound for (1.4).

QUESTION 2. *Prove a Berry–Esseen bound for the nonstandard limit theorem for the block estimator of a multi-dimensional isotonic regression function* (*cf.* [52]).

Recently [52] established a nonstandard limit theorem for the so-called block estimator $\widehat{f}_n$ (cf. [37]) for a d-dimensional isotonic regression function f_0 on $[0,1]^d$ (i.e., $f_0(x)\leq f_0(y)$ if $x_k\leq y_k$, $1\leq k\leq d$). In particular, suppose $x_0\in(0,1)^d$ and $\partial_k f_0(x_0)>0$ for $1\leq k\leq d$, the errors ξ_i's are i.i.d. mean-zero with variance σ^2, and the design points $\{X_i\}$ are of a fixed balanced design (see [52] for a precise definition) or a random design with uniform distribution on $[0,1]^d$. Then [52] showed that

$$\big(n/\sigma^2\big)^{1/(d+2)}\big(\widehat{f}_n(x_0)-f_0(x_0)\big)\stackrel{d}{\to}\left(\prod_{k=1}^d\big(\partial_k f_0(x_0)/2\big)\right)^{1/(d+2)}\cdot\mathbb{D}_{(1,\ldots,1)},$$

where $\mathbb{D}_{(1,\dots,1)}$ is a fairly complicated random variable generalizing the Chernoff distribution $\mathbb{D}_1$; a detailed description can be found in [52]. We believe the techniques developed in this paper will be useful in establishing a Berry–Essen bound for the above limit theorem. However, the anti-concentration problem associated with $\mathbb{D}_{(1,\dots,1)}$ in the multi-dimensional regression setting seems substantially more challenging than the univariate problem studied in this paper.

APPENDIX A: PROOF OF AUXILIARY LEMMAS

A.1. Proof of Lemma 3.8.

PROOF OF LEMMA 3.8. By [10], formula (1.1.4), page 197, noting that $\mathrm{Ercf}(z) = 2(1-\Phi(\sqrt{2}z))$, we have for any $y \geq 0$,

$$\begin{aligned}\mathbb{P}(M_\mu \geq y) &= \frac{1}{2}\mathrm{Ercf}\Big(\frac{y-\mu}{\sqrt{2}}\Big) + \frac{1}{2}e^{2\mu y}\cdot \mathrm{Ercf}\Big(\frac{y+\mu}{\sqrt{2}}\Big) \\ &= 1 - \Phi(y-\mu) + e^{2\mu y}\big(1-\Phi(y+\mu)\big).\end{aligned}$$

Differentiating the above display with respect to y yields (3.3), upon using $e^{2\mu y}\varphi(y+\mu) = \varphi(y-\mu)$. Alternatively, (3.3) can be derived using [81], formula (1.9),

$$\mathbb{P}\big(M_\mu \in \mathrm{d}y, \mathbb{B}_\mu(1) \in \mathrm{d}x\big) = (2\pi)^{-1/2} 2(2y-x)e^{-\frac{(2y-x)^2}{2}} e^{\mu x - \frac{\mu^2}{2}} \mathbf{1}_{y \geq 0, y \geq x}\, \mathrm{d}x\, \mathrm{d}y,$$

which follows from the change of variables (or Cameron–Martin) formula for Gaussian measures (cf. [39], Theorem 2.6.13). Hence for $y \geq 0$, $p_{M_\mu}(y)$ can be evaluated by integrating out x:

$$\begin{aligned} p_{M_\mu}(y) &= \frac{2}{\sqrt{2\pi}} \int_{-\infty}^{y} (2y-x) e^{-\frac{(2y-x)^2}{2}} e^{\mu x - \frac{\mu^2}{2}}\, \mathrm{d}x \\ &= \frac{2}{\sqrt{2\pi}} \int_{y}^{\infty} t e^{-\frac{t^2}{2}} e^{\mu(2y-t) - \frac{\mu^2}{2}}\, \mathrm{d}t \\ &= \frac{2e^{2\mu y}}{\sqrt{2\pi}} \int_{y}^{\infty} t e^{-\frac{(t+\mu)^2}{2}}\, \mathrm{d}t = \frac{2e^{2\mu y}}{\sqrt{2\pi}} \Big[\int_{y+\mu}^{\infty} v e^{-\frac{v^2}{2}}\, \mathrm{d}v - \mu \int_{y+\mu}^{\infty} e^{-\frac{v^2}{2}}\, \mathrm{d}v\Big] \\ &= \frac{2e^{2\mu y}}{\sqrt{2\pi}} \big[e^{-\frac{(y+\mu)^2}{2}} - \sqrt{2\pi}\mu\big(1-\Phi(y+\mu)\big)\big], \end{aligned}$$

which agrees with (3.3). Since p_{M_μ} is discontinuous at 0, $p_{M_\mu}(0)$ is understood as the right limit: $p_{M_\mu}(0) \equiv \lim_{y\to 0+} p_{M_\mu}(y)$. Finally, note that for $y+\mu \leq 1$, $e^{2\mu y}(1-\Phi(y+\mu)) \leq e^{2\mu(1-\mu)}(1-\Phi(1))\mathbf{1}_{\mu \geq 0} + \mathbf{1}_{\mu<0} \leq e^{1/2}$, while for $y+\mu > 1$, $e^{2\mu y}(1-\Phi(y+\mu)) \leq (1/\sqrt{2\pi})e^{2\mu y - (y+\mu)^2/2} = (1/\sqrt{2\pi})e^{-(y-\mu)^2/2} \leq 1/\sqrt{2\pi}$. This implies that $\|p_{M_\mu}\|_\infty \lesssim (\mu \vee 1)$. $\square$

A.2. Proof of Lemma 4.1.

PROOF OF LEMMA 4.1. $\alpha = \infty$ is the trivial case, so we only consider $\alpha < \infty$. First consider fixed design with $x_0 \in (0,1)$. Then for $x_0 \in \{X_i\}$,

$$\begin{aligned} &\sum_{x_0 - h_1 r_n \leq X_i \leq x_0 + h_2 r_n} \big(f_0(X_i) - f_0(x_0)\big) \\ &= \sum_{x_0 - h_1 r_n \leq X_i \leq x_0 + h_2 r_n} \Big[\frac{f_0^{(\alpha)}(x_0)}{\alpha!}(X_i - x_0)^\alpha + \big(1+o(1)\big)\frac{f_0^{(\alpha^*)}(x_0)}{\alpha^*!}(X_i-x_0)^{\alpha^*}\mathbf{1}_{\alpha^*<\infty}\Big] \end{aligned}$$

$$
\begin{aligned}
&= \sum_{-h_1 r_n\cdot\Lambda_0 n\le m\le h_2 r_n\cdot\Lambda_0 n}\Big[\frac{f_0^{(\alpha)}(x_0)}{\alpha!}\Big(\frac{m}{\Lambda_0 n}\Big)^{\alpha}+\big(1+o(1)\big)\frac{f_0^{(\alpha^*)}(x_0)}{\alpha^*!}\Big(\frac{m}{\Lambda_0 n}\Big)^{\alpha^*}\mathbf{1}_{\alpha^*<\infty}\Big]\\
&= \frac{f_0^{(\alpha)}(x_0)}{(\alpha+1)!}\Lambda_0^{-\alpha}n^{-\alpha}\big[\lfloor h_2 r_n\cdot\Lambda_0 n\rfloor^{\alpha+1}+(-1)^{\alpha}\lfloor h_1 r_n\cdot\Lambda_0 n\rfloor^{\alpha+1}\\
&\quad + O\big(((h_1\vee h_2)r_n\Lambda_0 n)^{\alpha}\big)\big]+n^{-\alpha^*}O\big(((h_1\vee h_2)r_n\Lambda_0 n)^{\alpha^*+1}\mathbf{1}_{\alpha^*<\infty}\big)\\
&= nr_n\Big[\frac{f_0^{(\alpha)}(x_0)}{(\alpha+1)!}\cdot\big(h_2^{\alpha+1}-(-h_1)^{\alpha+1}\big)\cdot\Lambda_0 r_n^{\alpha}+O\big(\tau_n^{\alpha}r_n^{\alpha}(nr_n)^{-1}\vee\tau_n^{\alpha^*+1}r_n^{\alpha^*}\mathbf{1}_{\alpha^*<\infty}\big)\Big].
\end{aligned}
$$

For general x_0 (not necessarily a design point), $X_i-x_0=m/(\Lambda_0 n)+O(n^{-1})$ for integers m in the range $-h_1 r_n\cdot\Lambda_0 n\le m\le h_2 r_n\cdot\Lambda_0 n$, so an extra term of order at most $O(nr_n\cdot\tau_n/n)$ would be contributed in the above summation.

For fixed design with $x_0=0$, we work with $x_n\in\{X_i\}$ (the other case can be handled in similar way as above). Then

$$
\begin{aligned}
&\sum_{x_n-h_1 r_n\le X_i\le x_n+h_2 r_n}\big(f_0(X_i)-f_0(x_n)\big)\\
&= \sum_{-h_1 r_n\cdot\Lambda_0 n\le m\le h_2 r_n\cdot\Lambda_0 n}\Big[\sum_{\ell=1}^{\alpha^*-1}\frac{f_0^{(\ell)}(x_n)}{\ell!}\Big(\frac{m}{\Lambda_0 n}\Big)^{\ell}+\big(1+o(1)\big)\frac{f_0^{(\alpha^*)}(x_n)}{\alpha^*!}\Big(\frac{m}{\Lambda_0 n}\Big)^{\alpha^*}\mathbf{1}_{\alpha^*<\infty}\Big]\\
&= \sum_{\ell=1}^{\alpha}\frac{f_0^{(\alpha)}(0)}{(\alpha-\ell)!\ell!}\cdot\big(x_n^{\alpha-\ell}+O\big(x_n^{\alpha^*-\ell}\mathbf{1}_{\alpha^*<\infty}\big)\big)\Lambda_0^{-\ell}n^{-\ell}\sum_{-h_1 r_n\cdot\Lambda_0 n\le m\le h_2 r_n\cdot\Lambda_0 n}m^{\ell}\\
&\quad + f_0^{(\alpha^*)}(0)\mathbf{1}_{\alpha^*<\infty}\sum_{\ell=\alpha+1}^{\alpha^*}\frac{1+o(1)}{(\alpha^*-\ell)!\ell!}\cdot x_n^{\alpha^*-\ell}\Lambda_0^{-\ell}n^{-\ell}\sum_{-h_1 r_n\cdot\Lambda_0 n\le m\le h_2 r_n\cdot\Lambda_0 n}m^{\ell}\\
&= \sum_{\ell=1}^{\alpha}\frac{f_0^{(\alpha)}(0)}{(\alpha-\ell)!(\ell+1)!}\cdot\big(x_n^{\alpha-\ell}+O\big(x_n^{\alpha^*-\ell}\mathbf{1}_{\alpha^*<\infty}\big)\big)\Lambda_0^{-\ell}n^{-\ell}\\
&\quad\times\big\{\lfloor h_2 r_n\cdot\Lambda_0 n\rfloor^{\ell+1}-\big(-\lfloor h_1 r_n\cdot\Lambda_0 n\rfloor\big)^{\ell+1}+O\big(((h_1\vee h_2)r_n n)^{\ell}\big)\big\}\\
&\quad + O\Big(\max_{\alpha+1\le\ell\le\alpha^*}x_n^{\alpha^*-\ell}n^{-\ell}(h_1\vee h_2)^{\ell+1}(r_n n)^{\ell+1}\mathbf{1}_{\alpha^*<\infty}\Big)\\
&= nr_n\Big[\sum_{\ell=1}^{\alpha}\frac{f_0^{(\alpha)}(0)}{(\alpha-\ell)!(\ell+1)!}\cdot\big(h_2^{\ell+1}-(-h_1)^{\ell+1}\big)\cdot\Lambda_0 x_n^{\alpha-\ell}r_n^{\ell}\\
&\quad + O\Big(\max_{1\le\ell\le\alpha^*}\tau_n^{\ell+1}x_n^{\alpha^*-\ell}r_n^{\ell}\mathbf{1}_{\alpha^*<\infty}\vee\max_{1\le\ell\le\alpha}\tau_n^{\ell}x_n^{\alpha-\ell}r_n^{\ell}(nr_n)^{-1}\Big)\Big].
\end{aligned}
$$

Next consider random design with $x_0\in(0,1)$. It is easy to see by Lemma B.4 that for any $\ell\ge 1$,

$$
\begin{aligned}
&\mathbb{E}\sup_{0\le h_i\le\tau_n,i=1,2}\big|n(\mathbb{P}_n-P)\big((X-x_0)^{\ell}\mathbf{1}_{[x_0-h_1 r_n,x_0+h_2 r_n]}\big)\big|\lesssim\sqrt{n\cdot(\tau_n r_n)^{2\ell+1}\log n},\\
&\sup_{0\le h_i\le\tau_n,i=1,2}\mathrm{Var}\big((X-x_0)^{\ell}\mathbf{1}_{[x_0-h_1 r_n,x_0+h_2 r_n]}\big)\lesssim(\tau_n r_n)^{2\ell+1}.
\end{aligned}
$$

By Talagrand's concentration inequality (cf. Lemma B.3), there exists some constant $K > 0$ such that for any $x \geq 0$,

$$\mathbb{P}\Big(K^{-1} \sup_{0 \leq h_i \leq \tau_n, i=1,2} \big|n(\mathbb{P}_n - P)\big((X - x_0)^\ell \mathbf{1}_{[x_0 - h_1 r_n, x_0 + h_2 r_n]}\big)\big| \geq \sqrt{n(\tau_n r_n)^{2\ell+1}(\log n + x)} + x\Big) \leq e^{-x}.$$

Hence with probability at least $1 - O(n^{-11})$, it holds that uniformly in $h_1, h_2 \leq \tau_n$

$$\begin{aligned}
&\sum_{x_0 - h_1 r_n \leq X_i \leq x_0 + h_2 r_n} \big(f_0(X_i) - f_0(x_0)\big) \\
&= \frac{f_0^{(\alpha)}(x_0)}{\alpha!} n\mathbb{P}_n(X - x_0)^\alpha \mathbf{1}_{[x_0 - h_1 r_n, x_0 + h_2 r_n]} \\
&\quad + \big(1 + o(1)\big)\mathbf{1}_{\alpha^* < \infty} \frac{f_0^{(\alpha^*)}(x_0)}{\alpha^*!} n\mathbb{P}_n(X - x_0)^{\alpha^*} \mathbf{1}_{[x_0 - h_1 r_n, x_0 + h_2 r_n]} \\
&= \frac{f_0^{(\alpha)}(x_0)}{\alpha!}\big[nP(X - x_0)^\alpha \mathbf{1}_{[x_0 - h_1 r_n, x_0 + h_2 r_n]} + O\big(\sqrt{n(\tau_n r_n)^{2\alpha+1} \log n} \vee \log n\big)\big] \\
&\quad + \mathbf{1}_{\alpha^* < \infty} \cdot O\big[nP(X - x_0)^{\alpha^*} \mathbf{1}_{[x_0 - h_1 r_n, x_0 + h_2 r_n]} + O\big(\sqrt{n(\tau_n r_n)^{2\alpha^*+1} \log n} \vee \log n\big)\big] \\
&= nr_n\bigg[\frac{f_0^{(\alpha)}(x_0)}{(\alpha+1)!}\big(h_2^{\alpha+1} - (-h_1)^{\alpha+1}\big) \cdot \Lambda_0 r_n^\alpha \\
&\quad + O\bigg(\tau_n^{\alpha^*+1} r_n^{\alpha^*} \mathbf{1}_{\alpha^* < \infty} \vee \tau_n^{\alpha+\beta+1} r_n^{\alpha+\beta} \mathbf{1}_{\beta < \infty} \vee \sqrt{\tau_n^{2\alpha+1} r_n^{2\alpha} \frac{\log n}{nr_n}} \vee \frac{\log n}{nr_n}\bigg)\bigg].
\end{aligned}$$

Here we used that for all $\ell \geq 1$,

$$\begin{aligned}
&P(X - x_0)^\ell \mathbf{1}_{[x_0 - h_1 r_n, x_0 + h_2 r_n]} \\
&\quad = \int_{x_0}^{x_0 + h_2 r_n} (x - x_0)^\ell \pi(x)\,\mathrm{d}x + \int_{x_0 - h_1 r_n}^{x_0} (x - x_0)^\ell \pi(x)\,\mathrm{d}x \\
&\quad = \frac{\Lambda_0}{\ell+1}\big(h_2^{\ell+1} - (-h_1)^{\ell+1}\big) r_n^{\ell+1} + O\big(\big((h_1 \vee h_2) r_n\big)^{\ell+\beta+1} \mathbf{1}_{\beta < \infty}\big).
\end{aligned}$$

For random design with $x_0 = 0$, with probability at least $1 - O(n^{-11})$, we have uniformly in $h_1, h_2 \leq \tau_n$,

$$\begin{aligned}
&\sum_{x_n - h_1 r_n \leq X_i \leq x_n + h_2 r_n} \big(f_0(X_i) - f_0(x_n)\big) \\
&\quad = \sum_{\ell=1}^{\alpha^*-1} \frac{f_0^{(\ell)}(x_n)}{\ell!} \cdot n\mathbb{P}_n(X - x_n)^\ell \mathbf{1}_{[x_n - h_1 r_n, x_n + h_2 r_n]} \\
&\qquad + \big(1 + o(1)\big)\mathbf{1}_{\alpha^* < \infty} \frac{f_0^{(\alpha^*)}(x_n)}{\alpha^*!} \cdot n\mathbb{P}_n(X - x_n)^{\alpha^*} \mathbf{1}_{[x_n - h_1 r_n, x_n + h_2 r_n]} \\
&\quad = \sum_{\ell=1}^{\alpha} \frac{f_0^{(\alpha)}(0)}{(\alpha-\ell)!\ell!}\big(x_n^{\alpha-\ell} + O\big(x_n^{\alpha^*-\ell} \mathbf{1}_{\alpha^* < \infty}\big)\big) \\
&\qquad \times \big[nP(X - x_0)^\ell \mathbf{1}_{[x_0 - h_1 r_n, x_0 + h_2 r_n]} + O\big(\sqrt{n(\tau_n r_n)^{2\ell+1} \log n} \vee \log n\big)\big]
\end{aligned}$$

$$
\begin{aligned}
&+ f_0^{(\alpha^*)}(0)\mathbf{1}_{\alpha^*<\infty}\sum_{\ell=\alpha+1}^{\alpha^*}\frac{1+o(1)}{(\alpha^*-\ell)!\ell!}x_n^{\alpha^*-\ell}\\
&\times\big[nP(X-x_0)^\ell\mathbf{1}_{[x_0-h_1r_n,x_0+h_2r_n]}+O\big(\sqrt{n(\tau_nr_n)^{2\ell+1}\log n}\vee\log n\big)\big]\\
={}&\sum_{\ell=1}^{\alpha}\frac{f_0^{(\alpha)}(0)}{(\alpha-\ell)!(\ell+1)!}\Lambda_0\big(h_2^{\ell+1}-(-h_1)^{\ell+1}\big)x_n^{\alpha-\ell}nr_n^{\ell+1}\\
&+O\Big(\max_{1\le\ell\le\alpha}x_n^{\alpha-\ell}\big\{n(\tau_nr_n)^{\ell+\beta+1}\mathbf{1}_{\beta<\infty}\vee\sqrt{n(\tau_nr_n)^{2\ell+1}\log n}\vee\log n\big\}\Big)\\
&+O\Big(\max_{1\le\ell\le\alpha^*}x_n^{\alpha^*-\ell}\mathbf{1}_{\alpha^*<\infty}\big\{n(\tau_nr_n)^{\ell+1}\vee\sqrt{n(\tau_nr_n)^{2\ell+1}\log n}\vee\log n\big\}\Big)\\
={}&nr_n\Bigg[\sum_{\ell=1}^{\alpha}\frac{f_0^{(\alpha)}(0)}{(\alpha-\ell)!(\ell+1)!}\big(h_2^{\ell+1}-(-h_1)^{\ell+1}\big)\cdot\Lambda_0x_n^{\alpha-\ell}r_n^\ell\\
&+O\Bigg(\max_{1\le\ell\le\alpha}x_n^{\alpha-\ell}\Bigg\{\tau_n^{\ell+\beta+1}r_n^{\ell+\beta}\mathbf{1}_{\beta<\infty}\vee\sqrt{\tau_n^{2\ell+1}r_n^{2\ell}\frac{\log n}{nr_n}}\vee\frac{\log n}{nr_n}\Bigg\}\Bigg)\\
&+O\Bigg(\max_{1\le\ell\le\alpha^*}x_n^{\alpha^*-\ell}\mathbf{1}_{\alpha^*<\infty}\Bigg\{\tau_n^{\ell+1}r_n^\ell\vee\sqrt{\tau_n^{2\ell+1}r_n^{2\ell}\frac{\log n}{nr_n}}\vee\frac{\log n}{nr_n}\Bigg\}\Bigg)\Bigg]\\
={}&nr_n\Bigg[\sum_{\ell=1}^{\alpha}\frac{f_0^{(\alpha)}(0)}{(\alpha-\ell)!(\ell+1)!}\big(h_2^{\ell+1}-(-h_1)^{\ell+1}\big)\cdot\Lambda_0x_n^{\alpha-\ell}r_n^\ell\\
&+O\Bigg(\max_{1\le\ell\le\alpha^*}\tau_n^{\ell+1}x_n^{\alpha^*-\ell}r_n^\ell\mathbf{1}_{\alpha^*<\infty}\vee\max_{1\le\ell\le\alpha}\tau_n^{\ell+\beta+1}x_n^{\alpha-\ell}r_n^{\ell+\beta}\mathbf{1}_{\beta<\infty}\\
&\vee\max_{1\le\ell\le\alpha^*}\big(x_n^{\alpha-\ell}\mathbf{1}_{1\le\ell\le\alpha}+x_n^{\alpha^*-\ell}\mathbf{1}_{\alpha^*<\infty}\big)\Bigg\{\sqrt{\tau_n^{2\ell+1}r_n^{2\ell}\frac{\log n}{nr_n}}\vee\frac{\log n}{nr_n}\Bigg\}\Bigg)\Bigg],
\end{aligned}
$$

as desired. □

A.3. Proof of Lemma 4.3.

PROOF OF LEMMA 4.3. Let $\widetilde{x}_n\equiv x^*+r_n$. First consider a fixed design. Let ℓ_n be the smallest integer such that $\widetilde{x}_n-2^{\ell_n}r_n<0$. Then

$$
\begin{aligned}
&\mathbb{P}\Big(\sup_{h\ge0}|\bar{\xi}|_{[x^*-hr_n,x^*+r_n]}|>t\omega_n\Big)\\
&\quad=\mathbb{P}\Big(\sup_{h\ge1}|\bar{\xi}|_{[\widetilde{x}_n-hr_n,\widetilde{x}_n]}|>t\omega_n\Big)\\
&\quad\le\sum_{\ell=0}^{\ell_n}\mathbb{P}\Big(\sup_{2^\ell\le h<2^{\ell+1}}|\bar{\xi}|_{[\widetilde{x}_n-hr_n,\widetilde{x}_n]}|>t\omega_n\Big)\\
&\quad\le\sum_{\ell=0}^{\ell_n}\mathbb{P}\Bigg(\sup_{2^\ell\le h<2^{\ell+1}}\Bigg|\sum_{i=1}^n\xi_i\mathbf{1}_{X_i\in[\widetilde{x}_n-hr_n,\widetilde{x}_n]}\Bigg|>t\omega_n\lfloor2^\ell r_n\cdot\Lambda_0n\rfloor\Bigg).
\end{aligned}
$$

By Lévy's maximal inequality (cf. [30], Theorem 1.1.5), each probability in the above summation can be bounded, up to an absolute constant, by

$$\mathbb{P}\bigg(\bigg|\sum_{i=1}^{n}\xi_i \mathbf{1}_{X_i\in[\tilde{x}_n-2^{\ell+1}r_n,\tilde{x}_n]}\bigg| > K_1\cdot\sqrt{\lfloor 2^{\ell+1}r_n\cdot\Lambda_0 n\rfloor}\cdot(2^\ell r_n n)^{1/2}t\omega_n\bigg)$$
$$\leq K_2\big[\exp(-2^\ell t^2/K_2)+\exp\big(-(nr_n)^{1/2}2^\ell t/K_2\big)\big].$$

Here we used the following facts: (i) for centered subexponential random variables $\xi_1,\ldots,\xi_m$, $\mathbb{P}(|\sum_{i=1}^m \xi_i| > \sqrt{m}u) \leq Ke^{-(u^2\wedge\sqrt{m}u)/K}$ holds for $u\geq 0$ (cf. [39], Proposition 3.1.8), and (ii) $(nr_n)^{1/2} = \omega_n^{-1}$. The claim for the fixed design case now follows by summing up the probabilities.

For the random design case, without loss of generality we work with P being the uniform distribution on [0, 1]. First note that by applying (essentially) [52], Lemma 10, with $L_n \equiv nr_n/\log n$, with probability at least $1-O(n^{-11})$, we have uniformly in ℓ,

$$\frac{\mathbb{P}_n\mathbf{1}_{[\tilde{x}_n-2^\ell r_n,\tilde{x}_n]}}{P\mathbf{1}_{[\tilde{x}_n-2^\ell r_n,\tilde{x}_n]}} = 1+O\big(L_n^{-1/2}\big) = 1+O(\omega_n\sqrt{\log n}).$$

Equivalently, the event

$$\mathcal{E}_n \equiv \big\{\mathbb{P}_n\mathbf{1}_{[\tilde{x}_n-2^\ell r_n,\tilde{x}_n]} = 2^\ell r_n\big(1+O(\omega_n\sqrt{\log n})\big) : \ell\geq 1\big\}$$

satisfies $\mathbb{P}(\mathcal{E}_n^c) = O(n^{-11})$. Hence, up to an additive term of order $O(n^{-11})$, we only need to control

$$\sum_{\ell=0}^{\ell_n}\mathbb{P}\bigg(\bigg\{\sup_{2^\ell\leq h<2^{\ell+1}}\bigg|\sum_{i=1}^n \xi_i\mathbf{1}_{X_i\in[\tilde{x}_n-hr_n,\tilde{x}_n]}\bigg| > t\omega_n n\mathbb{P}_n\mathbf{1}_{[\tilde{x}_n-2^\ell r_n,\tilde{x}_n]}\bigg\}\cap\mathcal{E}_n\bigg)$$
$$\lesssim \sum_{\ell=0}^{\ell_n}\mathbb{E}\bigg[\mathbb{P}\bigg(\bigg|\sum_{i=1}^n \xi_i\mathbf{1}_{X_i\in[\tilde{x}_n-2^{\ell+1}r_n,\tilde{x}_n]}\bigg| \gtrsim t\omega_n\cdot\mathbb{P}_n\mathbf{1}_{[\tilde{x}_n-2^\ell r_n,\tilde{x}_n]}|\{X_i\}\bigg)\mathbf{1}_{\mathcal{E}_n}\bigg]$$
$$= \sum_{\ell=0}^{\ell_n}\mathbb{E}\bigg[\mathbb{P}\bigg(\bigg|\sum_{i=1}^n \xi_i\mathbf{1}_{X_i\in[\tilde{x}_n-2^{\ell+1}r_n,\tilde{x}_n]}\bigg|$$
$$\gtrsim \sqrt{n\mathbb{P}_n\mathbf{1}_{[\tilde{x}_n-2^{\ell+1}r_n,\tilde{x}_n]}}\cdot t\omega_n\cdot\frac{n\mathbb{P}_n\mathbf{1}_{[\tilde{x}_n-2^\ell r_n,\tilde{x}_n]}}{\sqrt{n\mathbb{P}_n\mathbf{1}_{[\tilde{x}_n-2^{\ell+1}r_n,\tilde{x}_n]}}}|\{X_i\}\bigg)\mathbf{1}_{\mathcal{E}_n}\bigg]$$
$$\leq K_3\sum_{\ell=0}^{\ell_n}\mathbb{E}\exp\bigg(-K_3^{-1}\min\bigg\{t^2 n\omega_n^2\frac{(\mathbb{P}_n\mathbf{1}_{[\tilde{x}_n-2^\ell r_n,\tilde{x}_n]})^2}{\mathbb{P}_n\mathbf{1}_{[\tilde{x}_n-2^{\ell+1}r_n,\tilde{x}_n]}}, tn\omega_n\mathbb{P}_n\mathbf{1}_{[\tilde{x}_n-2^\ell r_n,\tilde{x}_n]}\bigg\}\mathbf{1}_{\mathcal{E}_n}\bigg)$$
$$\lesssim \sum_{\ell=0}^{\ell_n}\exp\big(-K_4^{-1}\min\{t^2 n\omega_n^2 2^\ell r_n\big(1+O(\omega_n\sqrt{\log n})\big), tn\omega_n 2^\ell r_n\big(1+O(\omega_n\sqrt{\log n})\big)\}\big)$$
$$= \sum_{\ell=0}^{\ell_n}\exp\big(-K_5^{-1}\min\{2^\ell t^2, (nr_n)^{1/2}2^\ell t\}\big).$$

The claim now follows by summing up the probabilities. □

A.4. Proof of Lemma 4.5. The proof of Lemma 4.5 relies on the following technical lemma, which will be also used in the proof of Theorem 2.2 in Section 5 for the case with $x_0 = 0$, $1/(2\alpha+1) < \rho < 1$, $\alpha < \infty$.

LEMMA A.1. *There exists some constant $K > 0$ such that for any $\varepsilon > 0$,*

$$\mathbb{P}\Big(\sup_{h_1\in(0,1]}\inf_{h_2\in[0,\infty)}\frac{\mathbb{B}(h_2)-\mathbb{B}(-h_1)}{h_1+h_2}\geq -\varepsilon\Big)\leq K\varepsilon^{1/2}\log_+^{1/4}(1/\varepsilon).$$

PROOF. Let $M_\varepsilon = \max\{1, \varepsilon^{-1}\log_+^{1/2}(1/\varepsilon)\}$. Note that the probability in question can be bounded by

$$\begin{aligned}
&\mathbb{P}\big(\exists h_1 \in (0,1], \forall h_2 \in [0, M_\varepsilon], \mathbb{B}(h_2)+\varepsilon h_2 \geq \mathbb{B}(-h_1)-\varepsilon h_1\big)\\
&\quad\leq \mathbb{P}\big(\exists h_1 \in (0,1], \forall h_2 \in [0, M_\varepsilon], \mathbb{B}(h_2) \geq \mathbb{B}(-h_1)-(M_\varepsilon+1)\varepsilon\big)\\
&\quad= \mathbb{P}\Big(\inf_{h_2\in[0,M_\varepsilon]}\mathbb{B}(h_2)\geq \inf_{h_1\in(0,1]}\mathbb{B}(-h_1)-(M_\varepsilon+1)\varepsilon\Big).
\end{aligned}$$

By the reflection principle of a Brownian motion, we have

$$\Big(\inf_{h_2\in[0,M_\varepsilon]}\mathbb{B}(h_2), \inf_{h_1\in(0,1]}\mathbb{B}(-h_1)\Big)\stackrel{d}{=}\big(-\sqrt{M_\varepsilon}\cdot|Z_1|, -|Z_2|\big),$$

where Z_1, Z_2 are independent standard normal random variables. Hence the above display further equals

$$\begin{aligned}
&\mathbb{P}\big(-\sqrt{M_\varepsilon}|Z_1| \geq -|Z_2|-(M_\varepsilon+1)\varepsilon\big)\\
&\quad\leq \mathbb{P}\big(|Z_1|\leq \big(200\log_+(1/\varepsilon)/M_\varepsilon\big)^{1/2}+2\sqrt{M_\varepsilon}\cdot\varepsilon\big)+\mathbb{P}\big(|Z_2|\geq \big(200\log_+(1/\varepsilon)\big)^{1/2}\big)\\
&\quad\leq K\varepsilon^{1/2}\log_+^{1/4}(1/\varepsilon)+O(\varepsilon^{100}),
\end{aligned}$$

as desired. □

PROOF OF LEMMA 4.5. We only consider the case for $\widetilde{h}_2$ with $H_2 = [0,\infty)$, and for notational simplicity we set $\sigma = 1$ and $\Lambda_0 = 1$. Geometrically, $\widetilde{h}_2$ is the first touch point of $\mathbb{B}_{1,1,Q}$ and its global LCM on H_2, so it is well defined on the event $\bigcup_{n=1}^\infty\bigcap_{M=n}^\infty \mathcal{E}_M$, where $\mathcal{E}_M \equiv \{\sup_{h_1\in H_1}\inf_{h_2\in H_2}\mathbb{B}_{1,1,Q}(h_1,h_2) = \sup_{h_1\in H_1}\inf_{h_2\in[0,M]}\mathbb{B}_{1,1,Q}(h_1,h_2)\}$.

First consider the cases $x_0 \in (0,1)$ or $x_0 = 0$ with $\rho \in (0, 1/(2\alpha+1)]$. In this case Q is a nonvanishing polynomial of degree at least 2. Then on the event $\mathcal{E}_M^c$,

$$\begin{aligned}
&\sup_{h_1\in H_1}\inf_{h_2\in H_2}\mathbb{B}_{1,1,Q}(h_1,h_2)\\
&\quad= \sup_{h_1\in H_1}\inf_{h_2>M}\mathbb{B}_{1,1,Q}(h_1,h_2)\\
&\quad\geq \inf_{h_2>M}\mathbb{B}_{1,1,Q}(1,h_2) = \inf_{h_2>M}\frac{\mathbb{B}(h_2)-\mathbb{B}(-1)+Q(h_2)-Q(-1)}{1+h_2}\\
&\quad\geq O(M)-\sup_{h>M}\frac{|\mathbb{B}(h)|}{1+h}-\frac{|\mathbb{B}(-1)|}{1+M}\stackrel{d}{=}O(M)-\frac{Y_1}{M^{1/2}}-\frac{Y_2}{M+1},
\end{aligned}$$

where $Y_1 \equiv \sup_{h>1}|\mathbb{B}(h)/h|$ and $Y_2 \equiv |\mathbb{B}(-1)|$ have sub-Gaussian tails. Hence for M large, on the intersection of $\mathcal{E}_M^c$ and an event with probability at least $1-Ke^{-M^2/K}$,

$$\sup_{h_1\in H_1}\inf_{h_2\in H_2}\mathbb{B}_{1,1,Q}(h_1,h_2)\geq O(M)-\sqrt{M}-\frac{M}{M+1}\geq O(M).$$

Since the random variable $\sup_{h_1\in H_1}\inf_{h_2\in H_2}\mathbb{B}_{1,1,Q}(h_1,h_2)$ has sub-Gaussian tails (using a similar proof to that of Lemma 4.4 above), we see that $\mathbb{P}(\mathcal{E}_M^c)\leq Ke^{-M^2/K}$. Using Borel–Cantelli yields that $\mathbb{P}(\bigcup_{n=1}^\infty\bigcap_{M=n}^\infty\mathcal{E}_M) = 1$.

Next consider the case $x_0 = 0$ with $\rho \in (1/(2\alpha+1), 1)$. In this case $Q \equiv 0$. Then on the event $\mathcal{E}_M^c$,

$$\sup_{h_1 \in (0,1]} \inf_{h_2 \in [0,\infty)} \mathbb{B}_{1,1,0}(h_1, h_2) \geq \inf_{h_2 > M} \frac{\mathbb{B}(h_2) - \mathbb{B}(-1)}{h_2 + 1} \stackrel{d}{=} -\frac{Y_1}{M^{1/2}} - \frac{Y_2}{M+1},$$

where Y_1, Y_2 are defined as above. This means that on the intersection of $\mathcal{E}_M^c$ and an event with probability at least $1 - M^{-100}$,

$$\sup_{h_1 \in (0,1]} \inf_{h_2 \in [0,\infty)} \mathbb{B}_{1,1,0}(h_1, h_2) \geq -K(\log M/M)^{1/2},$$

which occurs with probability at most $O(M^{-1/4} \log^{1/2} M)$ according to Lemma A.1. Hence $P(\mathcal{E}_M^c) \leq O(M^{-1/4} \log^{1/2} M)$. Summing M through a geometric sequence and using Borel–Cantelli yields the claim. $\square$

APPENDIX B: TECHNICAL TOOLS

This appendix collects some technical tools used in the proofs. The following Dudley's entropy integral bound can be found in [39], Theorem 2.3.7.

LEMMA B.1 (Entropy integral bound). *Let (T, d) be a pseudometric space, and $(X_t)_{t \in T}$ be a separable sub-Gaussian process such that $X_{t_0} = 0$ for some $t_0 \in T$. Then*

$$\mathbb{E} \sup_{t \in T} |X_t| \leq C \int_0^{\operatorname{diam}(T)} \sqrt{\log \mathcal{N}(\varepsilon, T, d)} \, \mathrm{d}\varepsilon,$$

where $C > 0$ is a universal constant.

The following Gaussian concentration inequality can be found in [39], Theorem 2.5.8.

LEMMA B.2 (Gaussian concentration inequality). *Let (T, d) be a pseudometric space, and $(X_t)_{t \in T}$ be a separable mean-zero Gaussian process with $\sup_{t \in T} |X_t| < \infty$ a.s. Then, with $\sigma^2 \equiv \sup_{t \in T} \operatorname{Var}(X_t)$, for any $u > 0$,*

$$\mathbb{P}\Big(\Big|\sup_{t \in T} |X_t| - \mathbb{E} \sup_{t \in T} |X_t|\Big| > u\Big) \leq 2e^{-u^2/(2\sigma^2)}.$$

Talagrand's concentration inequality [83] for the empirical process in the form given by Bousquet [11], is recorded as follows, cf. [39], Theorem 3.3.9.

LEMMA B.3 (Talagrand's concentration inequality). *Let $\mathcal{F}$ be a countable class of real-valued measurable functions such that $\sup_{f \in \mathcal{F}} \|f\|_\infty \leq b$ and $X_1, \ldots, X_n$ be i.i.d. random variables with law P. Then there exists some absolute constant $K > 1$ such that*

$$\mathbb{P}\Big(K^{-1} \sup_{f \in \mathcal{F}} \big|n(\mathbb{P}_n - P)f\big| \geq \mathbb{E} \sup_{f \in \mathcal{F}} \big|n(\mathbb{P}_n - P)f\big| + \sqrt{n\sigma^2 x} + bx\Big) \leq e^{-x},$$

where $\sigma^2 \equiv \sup_{f \in \mathcal{F}} \operatorname{Var}_P f$ and $\mathbb{P}_n$ denotes the empirical distribution of $X_1, \ldots, X_n$.

Talagrand's inequality is coupled with the following local maximal inequality for the empirical process due to [38, 84]. Denote the uniform entropy integral by

$$J(\delta, \mathcal{F}, L_2) \equiv \int_0^\delta \sup_Q \sqrt{1 + \log \mathcal{N}\big(\varepsilon \|F\|_{Q,2}, \mathcal{F}, L_2(Q)\big)} \, \mathrm{d}\varepsilon,$$

where the supremum is taken over all finitely discrete probability measures.

LEMMA B.4 (Local maximal inequality). *Let $\mathcal{F}$ be a countable class of real-valued measurable functions such that* $\sup_{f\in\mathcal{F}}\|f\|_\infty \le 1$, *and* $X_1,\dots,X_n$ *be i.i.d. random variables with law* P. *Then with* $\mathcal{F}(\delta)\equiv\{f\in\mathcal{F}: Pf^2<\delta^2\}$,

$$\mathbb{E}\sup_{f\in\mathcal{F}(\delta)}\big|(\mathbb{P}_n-P)(f)\big|\lesssim n^{-1/2}J(\delta,\mathcal{F},L_2)\bigg(1+\frac{J(\delta,\mathcal{F},L_2)}{\sqrt{n}\delta^2\|F\|_{P,2}}\bigg)\|F\|_{P,2}.$$

Acknowledgments. The authors would like to thank Jon Wellner for pointing out several references. They also would like to thank the anonymous referees, an Associate Editor and the Editor for their constructive comments that improve the quality of this paper.

Funding. Q. Han was supported by NSF Grant DMS-1916221. K. Kato was supported by NSF Grants DMS-1952306 and DMS-2014636.

REFERENCES

[1] ANEVSKI, D. and HÖSSJER, O. (2006). A general asymptotic scheme for inference under order restrictions. *Ann. Statist.* **34** 1874–1930. MR2283721 https://doi.org/10.1214/009053606000000443

[2] ANEVSKI, D. and SOULIER, P. (2011). Monotone spectral density estimation. *Ann. Statist.* **39** 418–438. MR2797852 https://doi.org/10.1214/10-AOS804

[3] BALABDAOUI, F. and WELLNER, J. A. (2014). Chernoff's density is log-concave. *Bernoulli* **20** 231–244. MR3160580 https://doi.org/10.3150/12-BEJ483

[4] BANERJEE, M. and MCKEAGUE, I. W. (2007). Confidence sets for split points in decision trees. *Ann. Statist.* **35** 543–574. MR2336859 https://doi.org/10.1214/009053606000001415

[5] BENTKUS, V. (2003). On the dependence of the Berry–Esseen bound on dimension. *J. Statist. Plann. Inference* **113** 385–402. MR1965117 https://doi.org/10.1016/S0378-3758(02)00094-0

[6] BENTKUS, V., GÖTZE, F. and TIKHOMIROV, A. (1997). Berry–Esseen bounds for statistics of weakly dependent samples. *Bernoulli* **3** 329–349. MR1468309 https://doi.org/10.2307/3318596

[7] BENTKUS, V. Y. (1986). Dependence of the Berry–Esseen estimate on the dimension. *Liet. Mat. Rink.* **26** 205–210. MR0862741

[8] BOLTHAUSEN, E. (1982). The Berry–Esseén theorem for strongly mixing Harris recurrent Markov chains. *Z. Wahrsch. Verw. Gebiete* **60** 283–289. MR0664418 https://doi.org/10.1007/BF00535716

[9] BOLTHAUSEN, E. (1982). Exact convergence rates in some martingale central limit theorems. *Ann. Probab.* **10** 672–688. MR0659537

[10] BORODIN, A. N. and SALMINEN, P. (1996). *Handbook of Brownian Motion—Facts and Formulae. Probability and Its Applications*. Birkhäuser, Basel. MR1477407 https://doi.org/10.1007/978-3-0348-7652-0

[11] BOUSQUET, O. (2003). Concentration inequalities for sub-additive functions using the entropy method. In *Stochastic Inequalities and Applications. Progress in Probability* **56** 213–247. Birkhäuser, Basel. MR2073435

[12] BRUNK, H. D. (1970). Estimation of isotonic regression. In *Nonparametric Techniques in Statistical Inference* (*Proc. Sympos., Indiana Univ., Bloomington, Ind.*, 1969) 177–197. Cambridge Univ. Press, London. MR0277070

[13] BÜHLMANN, P. and YU, B. (2002). Analyzing bagging. *Ann. Statist.* **30** 927–961. MR1926165 https://doi.org/10.1214/aos/1031689014

[14] CHATTERJEE, S. (2006). A generalization of the Lindeberg principle. *Ann. Probab.* **34** 2061–2076. MR2294976 https://doi.org/10.1214/009117906000000575

[15] CHATTERJEE, S. (2012). A new approach to strong embeddings. *Probab. Theory Related Fields* **152** 231–264. MR2875758 https://doi.org/10.1007/s00440-010-0321-8

[16] CHATTERJEE, S. and MECKES, E. (2008). Multivariate normal approximation using exchangeable pairs. *ALEA Lat. Am. J. Probab. Math. Stat.* **4** 257–283. MR2453473

[17] CHATTERJEE, S. and SHAO, Q.-M. (2011). Nonnormal approximation by Stein's method of exchangeable pairs with application to the Curie–Weiss model. *Ann. Appl. Probab.* **21** 464–483. MR2807964 https://doi.org/10.1214/10-AAP712

[18] CHEN, L. H. Y., GOLDSTEIN, L. and SHAO, Q.-M. (2011). *Normal Approximation by Stein's Method. Probability and Its Applications* (*New York*). Springer, Heidelberg. MR2732624 https://doi.org/10.1007/978-3-642-15007-4

[19] CHEN, L. H. Y. and SHAO, Q.-M. (2004). Normal approximation under local dependence. *Ann. Probab.* **32** 1985–2028. MR2073183 https://doi.org/10.1214/009117904000000450

[20] CHERNOFF, H. (1964). Estimation of the mode. *Ann. Inst. Statist. Math.* **16** 31–41. MR0172382 https://doi.org/10.1007/BF02868560

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

[22] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2014). Gaussian approximation of suprema of empirical processes. *Ann. Statist.* **42** 1564–1597. MR3262461 https://doi.org/10.1214/14-AOS1230

[23] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2015). Comparison and anti-concentration bounds for maxima of Gaussian random vectors. *Probab. Theory Related Fields* **162** 47–70. MR3350040 https://doi.org/10.1007/s00440-014-0565-9

[24] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2016). Empirical and multiplier bootstraps for suprema of empirical processes of increasing complexity, and related Gaussian couplings. *Stochastic Process. Appl.* **126** 3632–3651. MR3565470 https://doi.org/10.1016/j.spa.2016.04.009

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

[26] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2017). Detailed proof of Nazarov's inequality. Preprint. Available at arXiv:1711.10696.

[27] CHERNOZHUKOV, V., CHETVERIKOV, D., KATO, K. and KOIKE, Y. (2019). Improved central limit theorem and bootstrap approximations in high dimensions. Preprint. Available at arXiv:1912.10529.

[28] CHERNOZHUKOV, V., CHETVERIKOV, D. and KOIKE, Y. (2020). Nearly optimal central limit theorem and bootstrap approximations in high dimensions. Preprint. Available at arXiv:2012.09513.

[29] DAVYDOV, Y. A., LIFSHITS, M. A. and SMORODINA, N. V. (1998). *Local Properties of Distributions of Stochastic Functionals*. *Translations of Mathematical Monographs* **173**. Amer. Math. Soc., Providence, RI. Translated from the 1995 Russian original by V. E. Nazaĭkinskiĭ and M. A. Shishkova. MR1604537 https://doi.org/10.1090/mmono/173

[30] DE LA PEÑA, V. H. and GINÉ, E. (1999). *Decoupling*: *From Dependence to Independence*. *Probability and Its Applications* (*New York*). Springer, New York. Randomly stopped processes. U-statistics and processes. Martingales and beyond. MR1666908 https://doi.org/10.1007/978-1-4612-0537-1

[31] DÜMBGEN, L., WELLNER, J. A. and WOLFF, M. (2016). A law of the iterated logarithm for Grenander's estimator. *Stochastic Process. Appl.* **126** 3854–3864. MR3565482 https://doi.org/10.1016/j.spa.2016.04.012

[32] DUROT, C. (2007). On the $\mathbb{L}_p$-error of monotonicity constrained estimators. *Ann. Statist.* **35** 1080–1104. MR2341699 https://doi.org/10.1214/009053606000001497

[33] DUROT, C., KULIKOV, V. N. and LOPUHAÄ, H. P. (2012). The limit distribution of the L_∞-error of Grenander-type estimators. *Ann. Statist.* **40** 1578–1608. MR3015036 https://doi.org/10.1214/12-AOS1015

[34] FANG, X. and KOIKE, Y. (2020). Large-dimensional central limit theorem with fourth-moment error bounds on convex sets and balls. Preprint. Available at arXiv:2009.00339.

[35] FANG, X. and KOIKE, Y. (2021). High-dimensional central limit theorems by Stein's method. *Ann. Appl. Probab.* To appear. Available at arXiv:2001.10917.

[36] FANG, X. and KOIKE, Y. (2021). New error bounds in multivariate normal approximations via exchangeable pairs with applications to Wishart matrices and fourth moment theorems. *Ann. Appl. Probab.* To appear. Available at arXiv:2004.02101.

[37] FOKIANOS, K., LEUCHT, A. and NEUMANN, M. H. (2020). On integrated L^1 convergence rate of an isotonic regression estimator for multivariate observations. *IEEE Trans. Inf. Theory* **66** 6389–6402. MR4173546 https://doi.org/10.1109/TIT.2020.3013390

[38] GINÉ, E. and KOLTCHINSKII, V. (2006). Concentration inequalities and asymptotic results for ratio type empirical processes. *Ann. Probab.* **34** 1143–1216. MR2243881 https://doi.org/10.1214/009117906000000070

[39] GINÉ, E. and NICKL, R. (2016). *Mathematical Foundations of Infinite-Dimensional Statistical Models*. *Cambridge Series in Statistical and Probabilistic Mathematics* **40**. Cambridge Univ. Press, New York. MR3588285 https://doi.org/10.1017/CBO9781107337862

[40] GOLDSTEIN, L. and RINOTT, Y. (1996). Multivariate normal approximations by Stein's method and size bias couplings. *J. Appl. Probab.* **33** 1–17. MR1371949 https://doi.org/10.1017/s0021900200103675

[41] GÖTZE, F. (1991). On the rate of convergence in the multivariate CLT. *Ann. Probab.* **19** 724–739. MR1106283

[42] GRENANDER, U. (1956). On the theory of mortality measurement. II. *Skand. Aktuarietidskr.* **39** 125–153 (1957). MR0093415 https://doi.org/10.1080/03461238.1956.10414944

[43] GROENEBOOM, P. (1989). Brownian motion with a parabolic drift and Airy functions. *Probab. Theory Related Fields* **81** 79–109. MR0981568 https://doi.org/10.1007/BF00343738

[44] GROENEBOOM, P. (1996). Lectures on inverse problems. In *Lectures on Probability Theory and Statistics* (*Saint-Flour*, 1994). *Lecture Notes in Math*. **1648** 67–164. Springer, Berlin. MR1600884 https://doi.org/10.1007/BFb0095675

[45] GROENEBOOM, P. (2010). The maximum of Brownian motion minus a parabola. *Electron. J. Probab.* **15** 1930–1937. MR2738343 https://doi.org/10.1214/EJP.v15-826

[46] GROENEBOOM, P., HOOGHIEMSTRA, G. and LOPUHAÄ, H. P. (1999). Asymptotic normality of the L_1 error of the Grenander estimator. *Ann. Statist.* **27** 1316–1347. MR1740109 https://doi.org/10.1214/aos/1017938928

[47] GROENEBOOM, P. and JONGBLOED, G. (2014). *Nonparametric Estimation Under Shape Constraints*: *Estimators*, *Algorithms and Asymptotics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **38**. Cambridge Univ. Press, New York. MR3445293 https://doi.org/10.1017/CBO9781139020893

[48] GROENEBOOM, P., LALLEY, S. and TEMME, N. (2015). Chernoff's distribution and differential equations of parabolic and Airy type. *J. Math. Anal. Appl.* **423** 1804–1824. MR3278229 https://doi.org/10.1016/j.jmaa.2014.10.051

[49] GROENEBOOM, P. and WELLNER, J. A. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation*. *DMV Seminar* **19**. Birkhäuser, Basel. MR1180321 https://doi.org/10.1007/978-3-0348-8621-5

[50] GROENEBOOM, P. and WELLNER, J. A. (2001). Computing Chernoff's distribution. *J. Comput. Graph. Statist.* **10** 388–400. MR1939706 https://doi.org/10.1198/10618600152627997

[51] HALL, P. and BARBOUR, A. D. (1984). Reversing the Berry–Esseen inequality. *Proc. Amer. Math. Soc.* **90** 107–110. MR0722426 https://doi.org/10.2307/2044679

[52] HAN, Q. and ZHANG, C.-H. (2020). Limit distribution theory for block estimators in multiple isotonic regression. *Ann. Statist.* **48** 3251–3282. MR4185808 https://doi.org/10.1214/19-AOS1928

[53] HUANG, J. and WELLNER, J. A. (1995). Estimation of a monotone density or monotone hazard under random censoring. *Scand. J. Stat.* **22** 3–33. MR1334065

[54] HUANG, Y. and ZHANG, C.-H. (1994). Estimating a monotone density from censored observations. *Ann. Statist.* **22** 1256–1274. MR1311975 https://doi.org/10.1214/aos/1176325628

[55] JANKOWSKI, H. (2014). Convergence of linear functionals of the Grenander estimator under misspecification. *Ann. Statist.* **42** 625–653. MR3210981 https://doi.org/10.1214/13-AOS1196

[56] JANSON, S., LOUCHARD, G. and MARTIN-LÖF, A. (2010). The maximum of Brownian motion with parabolic drift. *Electron. J. Probab.* **15** 1893–1929. MR2738342 https://doi.org/10.1214/EJP.v15-830

[57] JIRAK, M. (2016). Berry–Esseen theorems under weak dependence. *Ann. Probab.* **44** 2024–2063. MR3502600 https://doi.org/10.1214/15-AOP1017

[58] KIM, J. and POLLARD, D. (1990). Cube root asymptotics. *Ann. Statist.* **18** 191–219. MR1041391 https://doi.org/10.1214/aos/1176347498

[59] KOMLÓS, J., MAJOR, P. and TUSNÁDY, G. (1975). An approximation of partial sums of independent RV's and the sample DF. I. *Z. Wahrsch. Verw. Gebiete* **32** 111–131. MR0375412 https://doi.org/10.1007/BF00533093

[60] KOMLÓS, J., MAJOR, P. and TUSNÁDY, G. (1976). An approximation of partial sums of independent RV's, and the sample DF. II. *Z. Wahrsch. Verw. Gebiete* **34** 33–58. MR0402883 https://doi.org/10.1007/BF00532688

[61] KULIKOV, V. N. and LOPUHAÄ, H. P. (2005). Asymptotic normality of the L_k-error of the Grenander estimator. *Ann. Statist.* **33** 2228–2255. MR2211085 https://doi.org/10.1214/009053605000000462

[62] KULIKOV, V. N. and LOPUHAÄ, H. P. (2006). The behavior of the NPMLE of a decreasing density near the boundaries of the support. *Ann. Statist.* **34** 742–768. MR2283391 https://doi.org/10.1214/009053606000000100

[63] LAHIRI, S. N. and SUN, S. (2009). A Berry–Esseen theorem for sample quantiles under weak dependence. *Ann. Appl. Probab.* **19** 108–126. MR2498673 https://doi.org/10.1214/08-AAP533

[64] LI, W. V. and SHAO, Q.-M. (2001). Gaussian processes: Inequalities, small ball probabilities and applications. In *Stochastic Processes*: *Theory and Methods*. *Handbook of Statist*. **19** 533–597. North-Holland, Amsterdam. MR1861734 https://doi.org/10.1016/S0169-7161(01)19019-X

[65] LIN, D., SHKEDY, Z., YEKUTIELI, D., AMARATUNGA, D. and BIJNENS, L. (2012). *Modeling Dose-Response Microarray Data in Early Drug Development Experiments Using R*: *Order-Restricted Analysis of Microarray Data*. Springer, Berlin.

[66] LUSS, R., ROSSET, S. and SHAHAR, M. (2012). Efficient regularized isotonic regression with application to gene-gene interaction search. *Ann. Appl. Stat.* **6** 253–283. MR2951537 https://doi.org/10.1214/11-AOAS504

[67] MATZKIN, R. L. (1991). Semiparametric estimation of monotone and concave utility functions for polychotomous choice models. *Econometrica* **59** 1315–1327. MR1133036 https://doi.org/10.2307/2938369

[68] MECKES, E. (2009). On Stein's method for multivariate normal approximation. In *High Dimensional Probability V*: *The Luminy Volume*. *Inst*. *Math*. *Stat*. (*IMS*) *Collect*. **5** 153–178. IMS, Beachwood, OH. MR2797946 https://doi.org/10.1214/09-IMSCOLL511

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

[70] PRAKASA RAO, B. L. S. (1969). Estimation of a unimodal density. *Sankhyā Ser. A* **31** 23–36. MR0267677

[71] PRAKASA RAO, B. L. S. (1970). Estimation for distributions with monotone failure rate. *Ann*. *Math*. *Stat*. **41** 507–519. MR0260133 https://doi.org/10.1214/aoms/1177697091

[72] REINERT, G. and RÖLLIN, A. (2009). Multivariate normal approximation with Stein's method of exchangeable pairs under a general linearity condition. *Ann*. *Probab*. **37** 2150–2173. MR2573554 https://doi.org/10.1214/09-AOP467

[73] RINOTT, Y. and ROTAR, V. (1996). A multivariate CLT for local dependence with $n^{-1/2}\log n$ rate and applications to multivariate graph related statistics. *J*. *Multivariate Anal*. **56** 333–350. MR1379533 https://doi.org/10.1006/jmva.1996.0017

[74] RINOTT, Y. and ROTAR, V. (1997). On coupling constructions and rates in the CLT for dependent summands with applications to the antivoter model and weighted U-statistics. *Ann*. *Appl*. *Probab*. **7** 1080–1105. MR1484798 https://doi.org/10.1214/aoap/1043862425

[75] RIO, E. (1996). Sur le théorème de Berry–Esseen pour les suites faiblement dépendantes. *Probab*. *Theory Related Fields* **104** 255–282. MR1373378 https://doi.org/10.1007/BF01247840

[76] ROBERTSON, T., WRIGHT, F. T. and DYKSTRA, R. L. (1988). *Order Restricted Statistical Inference*. *Wiley Series in Probability and Mathematical Statistics*: *Probability and Mathematical Statistics*. Wiley, Chichester. MR0961262

[77] ROUSSEEUW, P. J. (1984). Least median of squares regression. *J*. *Amer*. *Statist*. *Assoc*. **79** 871–880. MR0770281

[78] SAMWORTH, R. J. and SEN, B. (2018). Editorial: Special issue on "Nonparametric inference under shape constraints". *Statist*. *Sci*. **33** 469–472. MR3881203 https://doi.org/10.1214/18-STS673

[79] SCHELL, M. J. and SINGH, B. (1997). The reduced monotonic regression method. *J*. *Amer*. *Statist*. *Assoc*. **92** 128–135.

[80] SHAO, Q.-M. and ZHANG, Z.-S. (2019). Berry–Esseen bounds of normal and nonnormal approximation for unbounded exchangeable pairs. *Ann*. *Probab*. **47** 61–108. MR3909966 https://doi.org/10.1214/18-AOP1255

[81] SHEPP, L. A. (1979). The joint density of the maximum and its location for a Wiener process with drift. *J*. *Appl*. *Probab*. **16** 423–427. MR0531776 https://doi.org/10.2307/3212910

[82] STEIN, C. (1986). *Approximate Computation of Expectations*. *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **7**. IMS, Hayward, CA. MR0882007

[83] TALAGRAND, M. (1996). New concentration inequalities in product spaces. *Invent*. *Math*. **126** 505–563. MR1419006 https://doi.org/10.1007/s002220050108

[84] VAN DER VAART, A. and WELLNER, J. A. (2011). A local maximal inequality under uniform entropy. *Electron*. *J*. *Stat*. **5** 192–203. MR2792551 https://doi.org/10.1214/11-EJS605

[85] VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes*: *With Applications to Statistics*. *Springer Series in Statistics*. Springer, New York. MR1385671 https://doi.org/10.1007/978-1-4757-2545-2

[86] VAN ES, B., JONGBLOED, G. and VAN ZUIJLEN, M. (1998). Isotonic inverse estimators for nonparametric deconvolution. *Ann*. *Statist*. **26** 2395–2406. MR1700237 https://doi.org/10.1214/aos/1024691476

[87] WELLNER, J. A. and ZHANG, Y. (2000). Two estimators of the mean of a counting process with panel count data. *Ann*. *Statist*. **28** 779–814. MR1792787 https://doi.org/10.1214/aos/1015951998