

## MIT Open Access Articles

*Empirical and multiplier bootstraps for suprema of empirical processes of increasing complexity, and related Gaussian couplings*

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

**Citation:** Chernozhukov, Victor, Denis Chetverikov, and Kengo Kato. "Empirical and Multiplier Bootstraps for Suprema of Empirical Processes of Increasing Complexity, and Related Gaussian Couplings." *Stochastic Processes and Their Applications* 126, no. 12 (December 2016): 3632–3651.

**As Published:** <http://dx.doi.org/10.1016/J.SPA.2016.04.009>

**Publisher:** Elsevier BV

**Persistent URL:** <http://hdl.handle.net/1721.1/119502>

**Version:** Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

**Terms of use:** Creative Commons Attribution-NonCommercial-NoDerivs License



# Empirical and multiplier bootstraps for suprema of empirical processes of increasing complexity, and related Gaussian couplings

Victor Chernozhukov<sup>a</sup>, Denis Chetverikov<sup>b</sup>, Kengo Kato<sup>c</sup>

<sup>a</sup>*Department of Economics and Center for Statistics, MIT, 50 Memorial Drive, Cambridge, MA 02142, USA.*

<sup>b</sup>*Department of Economics, UCLA, Bunche Hall, 8283, 315 Portola Plaza, Los Angeles, CA 90095, USA.*

<sup>c</sup>*Graduate School of Economics, University of Tokyo, 7-3-1 Hongo Bunkyo-ku, Tokyo 113-0033, Japan.*

---

## Abstract

We derive strong approximations to the supremum of the non-centered empirical process indexed by a possibly unbounded VC-type class of functions by the suprema of the Gaussian and bootstrap processes. The bounds of these approximations are non-asymptotic, which allows us to work with classes of functions whose complexity increases with the sample size. The construction of couplings is not of the Hungarian type and is instead based on the Slepian-Stein methods and Gaussian comparison inequalities. The increasing complexity of classes of functions and non-centrality of the processes make the results useful for applications in modern nonparametric statistics (Giné and Nickl [14]), in particular allowing us to study the power properties of nonparametric tests using Gaussian and bootstrap approximations.

*Keywords:* coupling, empirical process, multiplier bootstrap process, empirical bootstrap process, Gaussian approximation, supremum

*2010 MSC:* 60F17, 62E17, 62G20

---

*Email addresses:* [vchern@mit.edu](mailto:vchern@mit.edu) (Victor Chernozhukov),  
[chetverikov@econ.ucla.edu](mailto:chetverikov@econ.ucla.edu) (Denis Chetverikov), [kkato@e.u-tokyo.ac.jp](mailto:kkato@e.u-tokyo.ac.jp) (Kengo Kato)

## 1. Introduction

Let  $(S, \mathcal{S})$  be a measurable space, and let  $X, X_1, \dots, X_n$  be a sequence of i.i.d. random variables taking values in  $(S, \mathcal{S})$  with a common distribution  $P$ . We assume that  $S$  is a separable metric space and  $\mathcal{S}$  is its Borel  $\sigma$ -field. Let  $\mathcal{F}$  be a class of measurable functions  $f : S \rightarrow \mathbb{R}$  with a measurable envelope  $F : S \rightarrow \mathbb{R}$  satisfying  $F(x) \geq \sup_{f \in \mathcal{F}} |f(x)|$  for all  $x \in S$ . Define the empirical process indexed by  $\mathcal{F}$ :

$$\mathbb{G}_n f = \frac{1}{\sqrt{n}} \sum_{i=1}^n (f(X_i) - Pf), \quad f \in \mathcal{F},$$

where  $Pf = \int f dP = \mathbb{E}[f(X)]$ . Let  $e_1, \dots, e_n$  be independent standard Gaussian random variables independent of  $X_1^n := \{X_1, \dots, X_n\}$ . Define the multiplier bootstrap process indexed by  $\mathcal{F}$ :

$$\mathbb{G}_n^e f = \frac{1}{\sqrt{n}} \sum_{i=1}^n e_i (f(X_i) - P_n f), \quad f \in \mathcal{F}, \quad (1)$$

where  $P_n$  is the empirical measure with respect to  $X_1, \dots, X_n$ ; that is,  $P_n f = n^{-1} \sum_{i=1}^n f(X_i)$  for  $f \in \mathcal{F}$ . Let  $N_1, \dots, N_n$  be a sequence of random variables multinomially distributed with parameters  $n$  and (probabilities)  $1/n, \dots, 1/n$  that are independent of  $X_1^n$ . Define the empirical bootstrap process indexed by  $\mathcal{F}$ :

$$\mathbb{G}_n^* f = \frac{1}{\sqrt{n}} \sum_{i=1}^n (N_i - 1) f(X_i), \quad f \in \mathcal{F}.$$

Suppose that  $\mathcal{F} \subset \mathcal{L}^2(P)$  is a VC type class of functions (the definition of VC type classes is recalled in Section 2) with  $\sup_{f \in \mathcal{F}} |Pf| < \infty$ . Then  $\mathcal{F}$  is totally bounded with respect to the semimetric

$$e_P(f, g) = \sqrt{P(f - g)^2}, \quad f, g \in \mathcal{F},$$

and there exists a centered Gaussian process  $G_P$  indexed by  $\mathcal{F}$  with uniformly  $e_P$ -continuous sample paths and covariance function

$$\mathbb{E}[G_P(f)G_P(g)] = \text{Cov}(f(X), g(X)), \quad f, g \in \mathcal{F}. \quad (2)$$

In this paper, for a given functional  $B : \mathcal{F} \rightarrow \mathbb{R}$ , we are interested in constructing couplings for

$$Z = \sup_{f \in \mathcal{F}} (B(f) + \mathbb{G}_n f) \quad \text{and} \quad \tilde{Z} \stackrel{d}{=} \sup_{f \in \mathcal{F}} (B(f) + G_P f), \quad (3)$$

$$Z^e = \sup_{f \in \mathcal{F}} (B(f) + \mathbb{G}_n^e f) \quad \text{and} \quad \tilde{Z}^e \stackrel{d|X_1^n}{=} \sup_{f \in \mathcal{F}} (B(f) + G_P f), \quad (4)$$

$$Z^* = \sup_{f \in \mathcal{F}} (B(f) + \mathbb{G}_n^* f) \quad \text{and} \quad \tilde{Z}^* \stackrel{d|X_1^n}{=} \sup_{f \in \mathcal{F}} (B(f) + G_P f), \quad (5)$$

such that the random variables appearing in each line are close to each other with high probability. The notation  $\stackrel{d}{=}$  means equality in distribution, and  $\stackrel{d|X_1^n}{=}$  means equality in conditional distribution given  $X_1^n = \{X_1, \dots, X_n\}$ . Here we suppose that the probability space is such that

$$(\Omega, \mathcal{A}, \mathbb{P}) = (S^n, \mathcal{S}^n, P^n) \times (T, \mathcal{T}, Q) \times ([0, 1], \mathcal{B}([0, 1]), \lambda)$$

where  $X_1, \dots, X_n$  are the coordinate projections of  $(S^n, \mathcal{S}^n, P^n)$ , random variables  $e_1, \dots, e_n$  (or  $N_1, \dots, N_n$ ) depend on the “second” coordinate only, and  $([0, 1], \mathcal{B}([0, 1]), \lambda)$  is the Lebesgue probability space on  $[0, 1]$ , that is,  $\mathcal{B}([0, 1])$  is the Borel  $\sigma$ -field on  $[0, 1]$  and  $\lambda$  is the Lebesgue measure on  $[0, 1]$ . The last augmentation of the probability space enables us to generate a uniform random variable on  $[0, 1]$  independent of  $X_1, \dots, X_n$  and  $e_1, \dots, e_n$  (or  $N_1, \dots, N_n$ ). We also implicitly assume here that the functional  $B$  and the class  $\mathcal{F}$  are “nice” enough so that measurability problems do not arise; see Section 2 for explicit assumptions.

Our coupling constructions are based on the Slepian-Stein methods and Gaussian comparison inequalities and built on the ideas in [3, 4, 5, 6, 7, 8, 9]. We emphasize that the construction of couplings in this paper is non-asymptotic, and so the class of functions  $\mathcal{F} = \mathcal{F}_n$  may depend on  $n$ , and its complexity may grow as the sample size increases. This feature of the couplings is especially important in modern nonparametric statistics [14]; see [6] and [7] for examples of applications.

We also emphasize that our couplings are not of the Hungarian type, and so are different from those obtained in e.g. [16] and [23], among many others. In particular, in contrast to e.g. [23], our couplings do not depend on the maximal total variation in  $\mathcal{F}$ . Instead, the couplings only depend on VC properties of the class of functions  $\mathcal{F}$  as well as on certain moments of the

functions in  $\mathcal{F}$  and the envelope  $F$ . This feature of the construction leads to a different range of possible applications in comparison with Hungarian couplings; see the detailed discussion in [6].

Gaussian and bootstrap approximations of the supremum of a *non-centered* empirical process have many potential applications. For example, these approximations can be used to derive non-asymptotic bounds on the errors in multivariate CLT. Specifically, let  $S = \mathbb{R}^p$ , and let  $A$  be a closed convex set in  $S$ . For  $\mathcal{V}^{p-1} = \{v \in \mathbb{R}^p : \|v\| = 1\}$ , let  $V_A : \mathcal{V}^{p-1} \rightarrow \mathbb{R}$  be the support function of  $A$  defined by  $V_A(v) = \sup_{x \in A} v^T x$ . Then  $x \in A$  if and only if  $\sup_{v \in \mathcal{V}^{p-1}} (v^T x - V_A(v)) \leq 0$ . Therefore, our results can be used to approximate

$$\mathbb{P}\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i \in A\right) = \mathbb{P}\left(\sup_{v \in \mathcal{V}^{p-1}} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n v^T X_i - V_A(v)\right) \leq 0\right). \quad (6)$$

Here, the dimension  $p = p_n$  of the sample space  $S = \mathbb{R}^p$  can depend on the sample size  $n$  and increase as  $n$  grows. Importantly, if the set  $A$  is such that the set  $\mathcal{V}^{p-1}$  on the right-hand side of (6) can be reduced to a sufficiently small subset of  $\mathcal{V}^{p-1}$ , the Gaussian approximation becomes possible even if  $p$  is larger or much larger than  $n$ ; see [5] and [9] for examples. More broadly, one can use our results for distributional approximation of general convex functionals on  $\mathbb{R}^p$  where the probability measure on  $\mathbb{R}^p$  is given by the distribution of a normalized sum of i.i.d. random vectors; see Section 11 of [10] where it is demonstrated that such functionals can be represented as suprema of non-centered empirical processes.

Another possible application is to study power properties of nonparametric tests where under the null, the statistic can be approximated by  $\sup_{f \in \mathcal{F}} \mathbb{G}_n f$ , and under the alternative, the statistic can be approximated by  $\sup_{f \in \mathcal{F}} (B(f) + \mathbb{G}_n f)$ , the functional  $B$  representing deviations from the null hypothesis. Finally, non-centered empirical processes are useful for multi-scale testing where one combines many statistics corresponding to different scales into one test using scale-dependent critical value for each statistic; see [13] where such tests were used for qualitative hypotheses testing.

This paper builds upon but differs from our previous papers [5, 6, 7, 8, 9]. In particular, this paper establishes, in the infinite dimensional setting, formal results on the multiplier and empirical bootstraps when the envelope  $F$  may be unbounded. In addition, this paper allows to approximate the supremum of a possibly non-centered empirical process. These settings are not covered in our previous papers [5, 6, 7, 8, 9] and are new.

The organization of this paper is as follows. In the next section, we present our main coupling theorems. In Section 3, we derive two auxiliary theorems that deal with maxima of high-dimensional random vectors. All the proofs are deferred to Sections 4 and 5. For convenience of the reader, we cite some additional results that are useful in our derivations in Section 6.

### 1.1. Notation

We use standard notation from the empirical process literature. For any probability measure  $Q$  on a measurable space  $(S, \mathcal{S})$ , we use the notation  $Qf = \int f dQ$ . For  $p \geq 1$ , we use  $\mathcal{L}^p(Q)$  to denote the space of all measurable functions  $f : S \rightarrow \mathbb{R}$  such that  $\|f\|_{Q,p} = (Q|f|^p)^{1/p} < \infty$ . We define the (semi)metric  $e_Q$  on  $\mathcal{L}^2(Q)$  by  $e_Q(f, g) = \|f - g\|_{Q,2}$ ,  $f, g \in \mathcal{L}^2(Q)$ .

For  $\varepsilon > 0$ , an  $\varepsilon$ -net of a (semi)metric space  $(T, d)$  is a subset  $T_\varepsilon$  of  $T$  such that for every  $t \in T$  there exists a point  $t_\varepsilon \in T_\varepsilon$  with  $d(t, t_\varepsilon) < \varepsilon$ . The  $\varepsilon$ -covering number  $N(T, d, \varepsilon)$  of  $T$  is the infimum of the cardinality of  $\varepsilon$ -nets of  $T$ , that is,  $N(T, d, \varepsilon) = \inf\{\text{Card}(T_\varepsilon) : T_\varepsilon \text{ is an } \varepsilon\text{-net of } T\}$ . For a subset  $A$  of a semimetric space  $(T, d)$ , we use  $A^\delta$  to denote the  $\delta$ -enlargement of  $A$ , that is,  $A^\delta = \{x \in T : d(x, A) \leq \delta\}$  where  $d(x, A) = \inf_{y \in A} d(x, y)$ . We also use the notation  $\|\cdot\|_T = \sup_{t \in T} \|\cdot\|$ .

For a function  $g : \mathbb{R} \rightarrow \mathbb{R}$ , we write  $\|g\|_\infty = \sup_{x \in \mathbb{R}} |g(x)|$ , and assuming that  $g$  is differentiable, we use  $g'$  to denote the derivative of  $g$ . We denote by  $C^k(\mathbb{R})$  the space of  $k$ -times continuously differentiable functions on  $\mathbb{R}$ . For  $a, b \in \mathbb{R}$ , we use the notation  $a \vee b = \max\{a, b\}$ .

## 2. Main results

In this section, we construct couplings between random variables in (3), (4), and (5) when  $\mathcal{F}$  is a VC type class of functions. Recall the definition:

**Definition 2.1** (VC type class). Let  $\mathcal{F}$  be a class of measurable functions on a measurable space  $(S, \mathcal{S})$ , to which a measurable envelope  $F$  is attached. We say that  $\mathcal{F}$  is VC type with envelope  $F$  if there are constants  $A, v > 0$  such that  $\sup_Q N(\mathcal{F}, e_Q, \varepsilon \|F\|_{Q,2}) \leq (A/\varepsilon)^v$  for all  $0 < \varepsilon \leq 1$ , where the supremum is taken over all finitely discrete probability measures on  $(S, \mathcal{S})$ .

Let  $B : \mathcal{F} \rightarrow \mathbb{R}$  be a given functional, and for  $\eta > 0$ , let  $N_B(\eta)$  be the minimal integer  $N$  such that there exist  $f_1, \dots, f_N \in \mathcal{F}$  with the property that for every  $f \in \mathcal{F}$ , there exists  $1 \leq j \leq N$  with  $|B(f) - B(f_j)| < \eta$ . We make the following assumptions.

- (A) There exists a countable subset  $\mathcal{G}$  of  $\mathcal{F}$  such that for any  $f \in \mathcal{F}$ , there exists a sequence  $g_m \in \mathcal{G}$  with  $g_m \rightarrow f$  pointwise and  $B(g_m) \rightarrow B(f)$ .
- (B) The class of functions  $\mathcal{F}$  is VC type with a measurable envelope  $F$  and constants  $A \geq e$  and  $v \geq 1$ .
- (C) There exist constants  $b \geq \sigma > 0$  and  $q \in [4, \infty)$  such that  $\sup_{f \in \mathcal{F}} P|f|^k \leq \sigma^2 b^{k-2}$  for  $k = 2, 3, 4$ , and  $\|F\|_{P,q} \leq b$ .

Assumptions (B) and (C) guarantee that  $\mathcal{F}$  is totally bounded with respect to the semimetric  $e_P$ , and there exists a centered Gaussian process  $G_P$  indexed by  $\mathcal{F}$  with uniformly  $e_P$ -continuous sample paths and covariance function given in (2).

Pick any  $\eta > 0$  and put

$$K_n = K_n(v, A, b, \sigma, B, \eta) = \log N_B(\eta) + v(\log n \vee \log(Ab/\sigma)).$$

The following theorem provides a coupling for  $Z$  and  $\tilde{Z}$ .

**Theorem 2.1** (Coupling for the supremum of the empirical process). *Suppose that assumptions (A)–(C) are satisfied, and in addition suppose that  $K_n^3 \leq n$ . Let  $Z = \sup_{f \in \mathcal{F}} (B(f) + \mathbb{G}_n f)$ . Then for every  $\gamma \in (0, 1)$ , there exists a random variable  $\tilde{Z} \stackrel{d}{=} \sup_{f \in \mathcal{F}} (B(f) + G_P f)$  such that*

$$\mathbb{P}\{|Z - \tilde{Z}| > C_1(\eta + \delta_n^{(1)})\} \leq C_2(\gamma + n^{-1})$$

where  $C_1, C_2$  are positive constants that depend only on  $q$ , and

$$\delta_n^{(1)} = \delta_n^{(1)}(v, A, b, \sigma, q, B, \eta, \gamma) = \frac{bK_n}{\gamma^{1/q} n^{1/2-1/q}} + \frac{(b\sigma^2 K_n^2)^{1/3}}{\gamma^{1/3} n^{1/6}}. \quad (7)$$

The result in Theorem 2.1 is new because it allows for non-centered processes. In addition, even in the case of centered processes, i.e. when  $B \equiv 0$ , the bound here improves slightly on our previous result given in Corollary 2.2 of [6].

**Remark 2.1** (Comparison with Beck's [1] lower bounds). Suppose that  $\mathcal{F}$  is the class of indicators of closed balls in  $\mathbb{R}^d$ , and  $X_1, X_2, \dots$  are i.i.d. uniform random variables on  $[0, 1]^d$ . Then [18] proved, via KMT constructions, that there exist versions  $B_n$  of  $G_P$  such that

$$\|\mathbb{G}_n - B_n\|_{\mathcal{F}} := \sup_{f \in \mathcal{F}} |\mathbb{G}_n f - B_n f| = O\{n^{-1/(2d)}(\log n)^{3/2}\} \quad a.s., \quad (8)$$

and up to a possible power of  $\log n$ , this rate is best possible when  $d \geq 2$  (Beck [1], Theorem 2).

We shall apply Theorem 2.1 to this class of functions. In this example,  $B \equiv 0$ , and so  $N_B(\eta) \equiv 1$ . Since the class of closed balls in  $\mathbb{R}^d$  is a VC class with index  $d+2$  [see 11], assumption (B) is satisfied with  $F \equiv 1$ ,  $v = cd$  with some universal constant  $c$ , and  $A$  being some universal constant. In addition, assumption (C) is satisfied with  $\sigma = b = 1$  and arbitrary  $q \in [4, \infty)$ , so there is a universal constant  $c'$  such that

$$\delta_n^{(1)} \leq c' \{ \gamma^{-1/q} d n^{-1/2+1/q} \log n + \gamma^{-1/3} d^{2/3} n^{-1/6} (\log n)^{2/3} \}.$$

If we take  $\gamma = \gamma_n \rightarrow 0$  sufficiently slowly, say  $\gamma_n = (\log n)^{-1/2}$ , then for  $Z_n = \sup_{f \in \mathcal{F}} \mathbb{G}_n f$ , Theorem 2.1 implies that there exists a sequence  $\tilde{Z}_n$  of random variables with  $\tilde{Z}_n \stackrel{d}{=} \sup_{f \in \mathcal{F}} G_P f$  such that

$$|Z_n - \tilde{Z}_n| = o_{\mathbb{P}} \{ d n^{-1/2+1/q} (\log n)^{1+1/(2q)} + d^{2/3} n^{-1/6} (\log n)^{5/6} \}. \quad (9)$$

This holds even when  $d = d_n \rightarrow \infty$  as long as  $d \log n = o(n^{1/3})$  (which guarantees  $K_n^3 \leq n$ ), and the right-hand side on (9) is  $o_{\mathbb{P}}(1)$  if  $d(\log n)^{5/4} = O(n^{1/4})$  by setting  $q$  large enough. It is then clear that, although Theorem 2.1 is only applicable to the supremum, and there is a difference in the mode of convergence, the rate of approximation of our coupling in (9) is better than that implied by (8) when  $d$  is large.  $\blacksquare$

Next we provide a coupling for  $Z^e$  and  $\tilde{Z}^e$ .

**Theorem 2.2** (Coupling for the supremum of the multiplier bootstrap process). *Suppose that assumptions (A)–(C) are satisfied, and in addition suppose that  $K_n \leq n$ . Let  $Z^e = \sup_{f \in \mathcal{F}} (B(f) + \mathbb{G}_n^e f)$ . Then for every  $\gamma \in (0, 1)$ , there exists a random variable  $\tilde{Z}^e \stackrel{d|X_1^n}{=} \sup_{f \in \mathcal{F}} (B(f) + G_P f)$  such that*

$$\mathbb{P}\{|Z^e - \tilde{Z}^e| > C_3(\eta + \delta_n^{(2)})\} \leq C_4(\gamma + n^{-1}),$$

where  $C_3, C_4$  are positive constants that depend only on  $q$ , and

$$\delta_n^{(2)} = \delta_n^{(2)}(v, A, b, \sigma, q, B, \eta, \gamma) = \frac{bK_n}{\gamma^{1+1/q} n^{1/2-1/q}} + \frac{(b\sigma K_n^{3/2})^{1/2}}{\gamma^{1+1/q} n^{1/4}}.$$

Finally, we provide a coupling for  $Z^*$  and  $\tilde{Z}^*$ .



**Theorem 2.3** (Coupling for the supremum of the empirical bootstrap process). *Suppose that assumptions (A)–(C) are satisfied, and in addition suppose that  $K_n^3 \leq n$ . Let  $Z^* = \sup_{f \in \mathcal{F}} (B(f) + \mathbb{G}_n^* f)$ . Then for every  $\gamma \in (0, 1)$ , there exists a random variable  $\tilde{Z}^* \stackrel{d|X_1^n}{=} \sup_{f \in \mathcal{F}} (B(f) + G_P f)$  such that*

$$\mathbb{P}\{|Z^* - \tilde{Z}^*| > C_5(\eta + \delta_n^{(3)})\} \leq C_6(\gamma + n^{-1}),$$

where  $C_5, C_6$  are positive constants that depend only on  $q$ , and

$$\begin{aligned} \delta_n^{(3)} &= \delta_n^{(3)}(v, A, b, \sigma, q, B, \eta, \gamma) \\ &= \frac{bK_n}{\gamma^{1+1/q}n^{1/2-1/q}} + \frac{(b\sigma^2 K_n^2)^{1/3}}{\gamma^{1/3}n^{1/6}} + \frac{(b\sigma K_n^{3/2})^{1/2}}{\gamma^{1+1/q}n^{1/4}}. \end{aligned}$$

**Remark 2.2.** By Markov's inequality, the following inequality is directly deduced from Theorem 2.2: under the conditions of Theorem 2.2, for every  $\alpha \in (0, 1)$ , with probability at least  $1 - \alpha$ , we have

$$\mathbb{P}\{|Z^e - \tilde{Z}^e| > C_3(\eta + \delta_n^{(2)}) \mid X_1^n\} \leq \alpha^{-1}C_4(\gamma + n^{-1}).$$

Likewise, the following inequality is directly deduced from Theorem 2.3: under the conditions of Theorem 2.3, for every  $\alpha \in (0, 1)$ , with probability at least  $1 - \alpha$ , we have

$$\mathbb{P}\{|Z^* - \tilde{Z}^*| > C_5(\eta + \delta_n^{(3)}) \mid X_1^n\} \leq \alpha^{-1}C_6(\gamma + n^{-1}).$$

■

**Remark 2.3.** In applications to statistics, it is often more useful to have bounds on the Kolmogorov distance for the following pairs of distribution functions:  $\mathbb{P}(Z \leq \cdot)$  and  $\mathbb{P}(\tilde{Z} \leq \cdot)$ ;  $\mathbb{P}(Z^e \leq \cdot \mid X_1^n)$  and  $\mathbb{P}(\tilde{Z} \leq \cdot)$ ; and  $\mathbb{P}(Z^* \leq \cdot \mid X_1^n)$  and  $\mathbb{P}(\tilde{Z} \leq \cdot)$ . Once such bounds are obtained, we will have a bound on, say, the Kolmogorov distance between  $\mathbb{P}(Z \leq \cdot)$  and  $\mathbb{P}(Z^e \leq \cdot \mid X_1^n)$ . By the following simple lemma, we see that to obtain such bounds from the coupling inequalities stated in Theorems 2.1–2.3, we need an *anti-concentration* inequality for  $\tilde{Z}$ , that is, an inequality bounding  $\sup_{t \in \mathbb{R}} \mathbb{P}(|\tilde{Z} - t| \leq \varepsilon)$  for  $\varepsilon > 0$ .

**Lemma 2.1.** *Let  $V, W$  be real-valued random variables such that  $\mathbb{P}(|V - W| > r_1) \leq r_2$  for some constants  $r_1, r_2 > 0$ . Then we have*

$$\sup_{t \in \mathbb{R}} |\mathbb{P}(V \leq t) - \mathbb{P}(W \leq t)| \leq \sup_{t \in \mathbb{R}} \mathbb{P}(|W - t| \leq r_1) + r_2.$$

The proof of this lemma is immediate and hence omitted. In the case where  $B(\cdot) \equiv 0$ , a useful anti-concentration inequality for  $\tilde{Z}$  is found in Lemma A.1 in [6], which essentially follows from Theorem 3 in [8]. Lemma A.1 in [6] does not cover, however, non-centered Gaussian processes. Therefore, here we provide a new anti-concentration inequality that can be applied to non-centered Gaussian processes. The proof of the lemma can be found in Section 4.

**Lemma 2.2.** *Let  $T$  be a non-empty set, and let  $\ell^\infty(T)$  be the set of all bounded functions on  $T$  endowed with the sup-norm. Let  $X(t), t \in T$  be a possibly non-centered tight Gaussian random element in  $\ell^\infty(T)$  such that  $\underline{\sigma}^2 := \inf_{t \in T} \text{Var}(X(t)) > 0$ . Define  $d(s, t) := \sqrt{\mathbb{E}[(X(t) - X(s))^2]}$ ,  $s, t \in T$ , and for  $\delta > 0$ , define  $\phi(\delta) := \mathbb{E}[\sup_{(s,t) \in T_\delta} |X(t) - X(s)|]$ , where  $T_\delta = \{(s, t) : d(s, t) \leq \delta\}$ . Then for every  $\varepsilon > 0$ ,*

$$\begin{aligned} & \sup_{x \in \mathbb{R}} \mathbb{P}(|\sup_{t \in T} X(t) - x| \leq \varepsilon) \\ & \leq \inf_{\delta, r > 0} \left\{ 2(1/\underline{\sigma})(\varepsilon + \phi(\delta) + r\delta)(\sqrt{2 \log N(T, d, \delta)} + 2) + e^{-r^2/2} \right\}. \end{aligned}$$

■

### 3. Auxiliary results for discretized processes

This section states two auxiliary results for “discretized” processes that will be used to prove the theorems stated in Section 2.

**Theorem 3.1.** *Let  $X_1, \dots, X_n$  be independent random vectors in  $\mathbb{R}^p$  ( $p \geq 2$ ) with finite absolute third moments, that is,  $\mathbb{E}[|X_{ij}|^3] < \infty$  for all  $1 \leq i \leq n$  and  $1 \leq j \leq p$ . Define  $\mu_i = \mathbb{E}[X_i]$  and  $\tilde{X}_i = X_i - \mu_i$ ,  $1 \leq i \leq n$ , and consider the statistic  $Z = \max_{1 \leq j \leq p} n^{-1/2} \sum_{i=1}^n X_{ij}$ . Let  $Y_1, \dots, Y_n$  be independent random vectors in  $\mathbb{R}^p$  with  $Y_i \sim N(\mu_i, \mathbb{E}[\tilde{X}_i \tilde{X}_i^T])$ , and define  $\tilde{Y}_i = Y_i - \mu_i$ ,  $1 \leq i \leq n$  and  $\tilde{Z} = \max_{1 \leq j \leq p} n^{-1/2} \sum_{i=1}^n Y_{ij}$ . Then for every  $\delta > 0$  and every Borel subset  $A$  of  $\mathbb{R}$ , we have*

$$\mathbb{P}(Z \in A) \leq \mathbb{P}(\tilde{Z} \in A^{C_7 \delta}) + \frac{C_8 \log^2 p}{\delta^3 \sqrt{n}} \cdot \{L_n + M_{n,X}(\delta) + M_{n,Y}(\delta)\},$$

where  $C_7, C_8$  are universal positive constants, and

$$\begin{aligned} L_n &= \max_{1 \leq j \leq p} \frac{1}{n} \sum_{i=1}^n \mathbb{E} \left[ |\tilde{X}_{ij}|^3 \right], \\ M_{n,X}(\delta) &= \frac{1}{n} \sum_{i=1}^n \mathbb{E} \left[ \max_{1 \leq j \leq p} |\tilde{X}_{ij}|^3 \cdot \mathbb{1} \left\{ \max_{1 \leq j \leq p} |\tilde{X}_{ij}| > \delta \sqrt{n} / \log p \right\} \right], \\ M_{n,Y}(\delta) &= \frac{1}{n} \sum_{i=1}^n \mathbb{E} \left[ \max_{1 \leq j \leq p} |\tilde{Y}_{ij}|^3 \cdot \mathbb{1} \left\{ \max_{1 \leq j \leq p} |\tilde{Y}_{ij}| > \delta \sqrt{n} / \log p \right\} \right]. \end{aligned}$$

**Theorem 3.2.** *Let  $X = (X_1, \dots, X_p)^T$  and  $Y = (Y_1, \dots, Y_p)^T$  be random vectors in  $\mathbb{R}^p$  ( $p \geq 2$ ) with  $X \sim N(\mu, \Sigma^X)$  and  $Y \sim N(\mu, \Sigma^Y)$ . Let  $\Delta = \max_{1 \leq j, k \leq p} |\Sigma_{jk}^X - \Sigma_{jk}^Y|$ , where  $\Sigma_{jk}^X$  and  $\Sigma_{jk}^Y$  denote the  $(j, k)$ -th elements of  $\Sigma^X$  and  $\Sigma^Y$ , respectively. Define  $Z = \max_{1 \leq j \leq p} X_j$  and  $\tilde{Z} = \max_{1 \leq j \leq p} Y_j$ . Then for every  $\delta > 0$  and every Borel subset  $A$  of  $\mathbb{R}$ ,*

$$\mathbb{P}(Z \in A) \leq \mathbb{P}(\tilde{Z} \in A^\delta) + C_9 \delta^{-1} \sqrt{\Delta \log p},$$

where  $C_9 > 0$  is a universal constant.

#### 4. Proofs for Section 2

Recall the definition of  $K_n$ :

$$K_n = K_n(v, A, b, \sigma, B, \eta) = \log N_B(\eta) + v(\log n \vee \log(Ab/\sigma)).$$

##### 4.1. Proof of Theorem 2.1

The proof relies on the following form of Strassen's theorem.

**Lemma 4.1** (Strassen's theorem). *Let  $\mu$  and  $\nu$  be Borel probability measures on  $\mathbb{R}$ . Let  $\varepsilon > 0$  and  $\delta > 0$ . Suppose that  $\mu(A) \leq \nu(A^\delta) + \varepsilon$  for every Borel subset  $A$  of  $\mathbb{R}$ . Let  $V$  be a random variable with distribution  $\mu$ . Then there is a random variable  $W$  with distribution  $\nu$  such that  $\mathbb{P}(|V - W| > \delta) \leq \varepsilon$ .*

*Proof of Lemma 4.1.* See Lemma 4.1 in [6]. ■

*Proof of Theorem 2.1.* By Strassen's theorem, it is sufficient to prove that for every Borel subset  $A$  of  $\mathbb{R}$ ,

$$\mathbb{P}(Z \in A) \leq \mathbb{P}\{\tilde{Z} \in A^{C_1(\eta + \delta_n^{(1)})}\} + C_2(\gamma + n^{-1}), \quad (10)$$

where  $\tilde{Z} = \sup_{f \in \mathcal{F}} (B(f) + G_P f)$ . The rest of the proof is divided into several steps. In the following,  $C$  denotes a positive constant that depends only on  $q$ ; the value of  $C$  may change from place to place.

**Step 1.** The first step is to “discretize” the empirical and Gaussian processes. To this end, take

$$\varepsilon = \sigma / (bn^{1/2}), \quad N = 2 \cdot N(\mathcal{F}, e_P, \varepsilon b) \cdot N_B(\eta).$$

Since  $N(\mathcal{F}, e_P, \varepsilon b) \leq (4A/\varepsilon)^v$  by approximation of  $P$  by a finitely discrete probability measure and assumption (B), we have  $\log N \leq CK_n$ . By definition, there exist  $f_1, \dots, f_N \in \mathcal{F}$  such that for every  $f \in \mathcal{F}$ , there exists  $1 \leq j \leq N$  with  $e_P(f, f_j) < \varepsilon b$  and  $|B(f) - B(f_j)| < \eta$ . Note that under the present assumption, the Gaussian process  $G_P$  can be extended to the linear hull of  $\mathcal{F}$  in such a way that  $G_P$  has linear sample paths [see 12, Theorem 3.1]. Hence letting  $\mathcal{F}_\varepsilon := \{f - g : f, g \in \mathcal{F}, e_P(f, g) < \varepsilon b\}$ , we conclude that

$$\begin{aligned} 0 &\leq \sup_{f \in \mathcal{F}} (B(f) + \mathbb{G}_n f) - \max_{1 \leq j \leq N} (B(f_j) + \mathbb{G}_n f_j) \leq \eta + \|\mathbb{G}_n\|_{\mathcal{F}_\varepsilon}, \\ 0 &\leq \sup_{f \in \mathcal{F}} (B(f) + G_P f) - \max_{1 \leq j \leq N} (B(f_j) + G_P f_j) \leq \eta + \|G_P\|_{\mathcal{F}_\varepsilon}. \end{aligned}$$

**Step 2.** Here we wish to show that

$$\mathbb{P}\{\|G_P\|_{\mathcal{F}_\varepsilon} > C\sqrt{\sigma^2 K_n/n}\} \leq 2n^{-1}. \quad (11)$$

This follows from the Borell-Sudakov-Tsirel'son inequality [see 25, Proposition A.2.1] complemented with Dudley's maximal inequality for Gaussian processes [see 25, Corollary 2.2.8].

First, by the Borell-Sudakov-Tsirel'son inequality, we have

$$\mathbb{P}\{\|G_P\|_{\mathcal{F}_\varepsilon} > \mathbb{E}[\|G_P\|_{\mathcal{F}_\varepsilon}] + \varepsilon b \sqrt{2 \log n}\} \leq 2n^{-1}.$$

Second, by Dudley's maximal inequality together with the fact that  $N(\mathcal{F}_\varepsilon, e_P, \tau) \leq N^2(\mathcal{F}, e_P, \tau/2) \leq (8Ab/\tau)^{2v}$ , we have

$$\mathbb{E}[\|G_P\|_{\mathcal{F}_\varepsilon}] \leq C\varepsilon b \sqrt{v \log(8Ab/\varepsilon)} \leq C\sqrt{\sigma^2 K_n/n}.$$

Combining these inequalities, together with the fact that  $\log n \leq K_n$ , leads to the desired inequality.

**Step 3.** We wish to show that

$$\mathbb{P} \left\{ \|\mathbb{G}_n\|_{\mathcal{F}_\varepsilon} > CbK_n/(\gamma^{1/q}n^{1/2-1/q}) \right\} \leq \gamma. \quad (12)$$

Applying Lemma 6.1 with  $\alpha = \gamma^{-1/q}$  and  $t = \gamma^{-2/q}$  to  $\mathcal{F}_\varepsilon$ , we have with probability at least  $1 - \gamma$ ,

$$\|\mathbb{G}_n\|_{\mathcal{F}_\varepsilon} \leq C\{\gamma^{-1/q}\mathbb{E}[\|\mathbb{G}_n\|_{\mathcal{F}_\varepsilon}] + (\sigma_\varepsilon + n^{-1/2}\|M_\varepsilon\|_q)\gamma^{-1/q} + n^{-1/2}\|M_\varepsilon\|_2\gamma^{-1/q}\},$$

where  $\sigma_\varepsilon := \sup_{f \in \mathcal{F}_\varepsilon} (Pf^2)^{1/2} \leq \varepsilon b = \sigma/n^{1/2}$  and  $M_\varepsilon := 2 \max_{1 \leq i \leq n} F(X_i)$ . Here  $\|M_\varepsilon\|_2 \leq \|M_\varepsilon\|_q \leq 2n^{1/q}b$ . In addition, by Lemma 6.2, we have

$$\mathbb{E}[\|\mathbb{G}_n\|_{\mathcal{F}_\varepsilon}] \leq C\{\sigma(K_n/n)^{1/2} + bK_n/n^{1/2-1/q}\} \leq CbK_n/n^{1/2-1/q}.$$

Combining these inequalities leads to (12).

**Step 4.** Let  $Z^\varepsilon = \max_{1 \leq j \leq N} (B(f_j) + \mathbb{G}_n f_j)$  and  $\tilde{Z}^\varepsilon = \max_{1 \leq j \leq N} (B(f_j) + G_P(f_j))$ . Here we apply Theorem 3.1 to show that whenever

$$\delta \geq 2c\sigma n^{-1/2}(\log N)^{3/2} \cdot (\log n) \quad (13)$$

for some universal constant  $c > 0$ , we have for every Borel subset  $A$  of  $\mathbb{R}$ ,

$$\mathbb{P}(Z^\varepsilon \in A) \leq \mathbb{P}(\tilde{Z}^\varepsilon \in A^{C\tau\delta}) + C \left( \frac{b\sigma^2 K_n^2}{\delta^3 \sqrt{n}} + \frac{b^q K_n^q}{\delta^q n^{q/2-1}} + \frac{1}{n} \right). \quad (14)$$

Let  $\tilde{X}_i = (f_j(X_i) - Pf_j)_{1 \leq j \leq N}$ ,  $1 \leq i \leq n$ , and let  $\tilde{Y} = (G_P f_j)_{1 \leq j \leq N}$ . Then as  $\tilde{X}_1, \dots, \tilde{X}_n$  are i.i.d.,

$$\begin{aligned} L_n &= \max_{1 \leq j \leq N} \mathbb{E}[|\tilde{X}_{1j}|^3] = \sup_{f \in \mathcal{F}} \mathbb{E}[|f(X) - Pf|^3] \leq 8 \sup_{f \in \mathcal{F}} P|f|^3 \leq 8\sigma^2 b, \\ M_{n,X}(\delta) &= \mathbb{E} \left[ \max_{1 \leq j \leq N} |\tilde{X}_{1j}|^3 \cdot \mathbf{1} \left\{ \max_{1 \leq j \leq N} |\tilde{X}_{1j}| > \delta \sqrt{n} / \log N \right\} \right] \\ &\leq \frac{\log^{q-3} N}{(\delta \sqrt{n})^{q-3}} \mathbb{E} \left[ \max_{1 \leq j \leq N} |\tilde{X}_{1j}|^q \right] \leq \frac{2^q b^q \log^{q-3} N}{(\delta \sqrt{n})^{q-3}}. \end{aligned}$$

To bound  $M_{n,Y}(\delta)$ , let  $\|\cdot\|_{\psi_1}$  denote the Orlicz norm associated with the Young modulus  $\psi_1(x) = e^x - 1$ , that is,  $\|\xi\|_{\psi_1} = \inf\{u > 0 : \mathbb{E}[\psi_1(|\xi|/u)] \leq 1\}$ . Then it is routine to verify that there exists a universal constant  $c > 0$  such

that  $\|\max_{1 \leq j \leq N} |\tilde{Y}_j|\|_{\psi_1} \leq c\sigma\sqrt{\log N}$ . Hence, by Markov's inequality, for every  $x > 0$ ,

$$\mathbb{P}\left(\max_{1 \leq j \leq N} |\tilde{Y}_j| > x\right) \leq 2 \exp\left(-\frac{x}{c\sigma\sqrt{\log N}}\right).$$

Therefore, by Lemma 6.6, whenever  $\delta \geq 2c\sigma n^{-1/2}(\log^{3/2} N) \cdot (\log n)$ ,

$$\begin{aligned} M_{n,Y}(\delta) &= \mathbb{E}\left[\max_{1 \leq j \leq N} |\tilde{Y}_j|^3 \cdot \mathbb{1}\left\{\max_{1 \leq j \leq N} |\tilde{Y}_j| > \delta\sqrt{n}/\log N\right\}\right] \\ &\leq 12(\delta\sqrt{n}/\log N + c\sigma\sqrt{\log N})^3 \exp\left(-\frac{\delta\sqrt{n}}{c\sigma\log^{3/2} N}\right) \\ &\leq Cn^{-2}(\delta\sqrt{n}/\log N)^3. \end{aligned}$$

Application of Theorem 3.1 with these bounds, together with the bound  $\log N \leq CK_n$ , leads to (14).

**Step 5.** In the previous step, take

$$\delta = C' \left\{ \frac{(b\sigma^2 K_n^2)^{1/3}}{\gamma^{1/3} n^{1/6}} + \frac{bK_n}{\gamma^{1/q} n^{1/2-1/q}} \right\},$$

where  $C' > 0$  is a large enough but universal constant. It is easy to check that for this choice of  $\delta$ , (13) holds under the condition  $K_n^3 \leq n$ . Indeed, since  $q \geq 4$ ,  $b \geq \sigma$ ,  $\log n \leq K_n$  and  $\log N \leq CK_n$ , we have  $2c\sigma n^{-1/2}(\log^{3/2} N) \cdot (\log n) \leq C'\sigma K_n^{3/2}/n^{4/9} \leq C'b^{1/3}\sigma^{2/3}K_n^{2/3}/(\gamma^{1/3}n^{1/6}) \leq \delta$ . Therefore, by Step 3, we have for every Borel subset  $A$  of  $\mathbb{R}$ ,

$$\mathbb{P}(Z^\varepsilon \in A) \leq \mathbb{P}(\tilde{Z}^\varepsilon \in A^{C'\delta}) + C(\gamma + n^{-1}).$$

The desired inequality (10) thus follows from combining Steps 1-5.  $\blacksquare$

#### 4.2. Proof of Theorem 2.2

The proof of Theorem 2.2 relies on a conditional version of Strassen's theorem due to [20].

**Lemma 4.2.** *Let  $V$  be a real-valued random variable defined on a probability space  $(\Omega, \mathcal{A}, \mathbb{P})$ , and let  $\mathcal{C}$  be a countably generated sub  $\sigma$ -field of  $\mathcal{A}$ . Assume that there exists a uniform random variable on  $[0, 1]$  independent of  $\mathcal{C} \vee \sigma(V)$ .*

Let  $G(\cdot | \mathcal{C})$  be a regular conditional distribution on the Borel  $\sigma$ -field of  $\mathbb{R}$  given  $\mathcal{C}$ , and suppose that for some  $\delta > 0$  and  $\varepsilon > 0$ ,

$$\mathbb{E} \left[ \sup_A \{ \mathbb{P}(V \in A | \mathcal{C}) - G(A^\delta | \mathcal{C}) \} \right] \leq \varepsilon,$$

where  $\sup_A$  is taken over all Borel subsets  $A$  of  $\mathbb{R}$ . Then there exists a random variable  $W$  such that the conditional distribution of  $W$  given  $\mathcal{C}$  coincides with  $G(\cdot | \mathcal{C})$ , and moreover  $\mathbb{P}(|V - W| > \delta) \leq \varepsilon$ .

*Proof.* See Theorem 4 in [20]. ■

*Proof of Theorem 2.2.* Here  $C$  denotes a positive constant that depends only on  $q$ ; the value of  $C$  may change from place to place. In addition, to ease the notation, we write  $a \lesssim b$  if  $a \leq Cb$ . By Lemma 4.2, since  $\sigma(X_1^n)$  is countably generated by the construction of the probability space (in particular, recall that we have assumed that  $S$  is a separable metric space), it is sufficient to find an event  $E \in \sigma(X_1^n)$  such that  $\mathbb{P}(E) \geq 1 - \gamma - n^{-1}$ , and on this event, the inequality

$$\mathbb{P}(Z^e \in A | X_1^n) \leq \mathbb{P}\{\tilde{Z} \in A^{C(\eta + \delta_n^{(2)})}\} + C(\gamma + n^{-1}) \quad (15)$$

holds for every Borel subset  $A$  of  $\mathbb{R}$ , where  $\tilde{Z} = \sup_{f \in \mathcal{F}} (B(f) + G_P f)$ .

We first specify such an event, and then show that on this event, (15) holds for every Borel subset  $A$  of  $\mathbb{R}$ . Applying Lemma 6.1 with  $\alpha = \gamma^{-1/q}$  and  $t = (\gamma/2)^{-2/q}$  to  $\mathcal{F}$ , we have with probability at least  $1 - \gamma/2$ ,

$$\|\mathbb{G}_n\|_{\mathcal{F}} \lesssim \gamma^{-1/q} \mathbb{E}[\|\mathbb{G}\|_{\mathcal{F}}] + (\sigma + n^{-1/2} \|M\|_q) \gamma^{-1/q} + n^{-1/2} \|M\|_2 \gamma^{-1/q},$$

where  $M := \max_{1 \leq i \leq n} F(X_i)$  satisfies  $\|M\|_2 \leq \|M\|_q = (\mathbb{E}[|M|^q])^{1/q} \leq n^{1/q} b$ . In addition, by Lemma 6.2,

$$\mathbb{E}[\|\mathbb{G}_n\|_{\mathcal{F}}] \lesssim \sigma K_n^{1/2} + \|M\|_2 K_n n^{-1/2} \leq \sigma K_n^{1/2} + b K_n n^{-1/2+1/q}.$$

Hence with probability at least  $1 - \gamma/2$ ,

$$\|\mathbb{G}_n\|_{\mathcal{F}} \lesssim \sigma K_n^{1/2} / \gamma^{1/q} + b K_n / (\gamma^{1/q} n^{1/2-1/q}). \quad (16)$$

Moreover, applying Lemma 6.1 again with  $\alpha = \gamma^{-2/q}$  and  $t = (\gamma/2)^{-4/q}$  to the class  $\mathcal{F} \cdot \mathcal{F} := \{f \cdot g : f, g \in \mathcal{F}\}$ , we have with probability at least  $1 - \gamma/2$ ,

$$\|\mathbb{G}\|_{\mathcal{F} \cdot \mathcal{F}} \lesssim \gamma^{-2/q} \mathbb{E}[\|\mathbb{G}_n\|_{\mathcal{F} \cdot \mathcal{F}}] + (\bar{\sigma} + n^{-1/2} \|M^2\|_{q/2}) \gamma^{-2/q} + n^{-1/2} \|M^2\|_2 \gamma^{-2/q},$$

where  $\bar{\sigma}^2 := \sup_{f \in \mathcal{F}, \mathcal{F}} P f^2 \leq \sup_{f \in \mathcal{F}} P f^4 \leq b^2 \sigma^2$ . In addition,  $\|M^2\|_2 \leq \|M^2\|_{q/2} = (\mathbb{E}[|M|^q])^{2/q} \leq n^{2/q} b^2$ , and as shown in the proof of Corollary 2.2 in [6],  $\mathbb{E}[\|\mathbb{G}_n\|_{\mathcal{F}, \mathcal{F}}] \lesssim b\sigma K_n^{1/2} + b^2 K_n n^{-1/2+2/q}$ . Hence with probability at least  $1 - \gamma/2$ ,

$$\|\mathbb{G}_n\|_{\mathcal{F}, \mathcal{F}} \lesssim b\sigma K_n^{1/2} / \gamma^{2/q} + b^2 K_n / (\gamma^{2/q} n^{1/2-2/q}). \quad (17)$$

Finally, by Markov's inequality, with probability at least  $1 - n^{-1}$ ,

$$\|F\|_{P_n, 2} \leq n^{1/2} \|F\|_{P, 2}. \quad (18)$$

Define  $E$  as the intersection of the events in (16), (17), and (18). Then  $E \in \sigma(X_1^n)$  and  $\mathbb{P}(E) \geq 1 - \gamma - n^{-1}$ . The rest of the proof, which is divided into several steps, is devoted to proving (15) for each fixed  $X_1, \dots, X_n$  satisfying (16)–(18).

In the following, we use the notation introduced in Step 1 of the proof of Theorem 2.1. Then

$$0 \leq \sup_{f \in \mathcal{F}} (B(f) + \mathbb{G}_n^e f) - \max_{1 \leq j \leq N} (B(f_j) + \mathbb{G}_n^e f_j) \leq \eta + \|\mathbb{G}_n^e\|_{\mathcal{F}_\varepsilon}, \quad (19)$$

$$0 \leq \sup_{f \in \mathcal{F}} (B(f) + G_P f) - \max_{1 \leq j \leq N} (B(f_j) + G_P f_j) \leq \eta + \|G_P\|_{\mathcal{F}_\varepsilon}. \quad (20)$$

**Step 1.** By Step 2 of the proof of Theorem 2.1, we have

$$\mathbb{P}(\|G_P\|_{\mathcal{F}_\varepsilon} > C \sqrt{\sigma^2 K_n / n}) \leq 2n^{-1}.$$

**Step 2.** Here we wish to show that on the event  $E$ ,

$$\mathbb{P} \left\{ \|\mathbb{G}_n^e\|_{\mathcal{F}_\varepsilon} > C \left\{ (b\sigma K_n^{3/2})^{1/2} / (\gamma^{1/q} n^{1/4}) + bK_n / (\gamma^{1/q} n^{1/2-1/q}) \right\} \mid X_1^n \right\} \leq 2n^{-1}. \quad (21)$$

Fix any  $X_1, \dots, X_n$  satisfying (16)–(18). Let us write  $(\mathcal{F} - \mathcal{F})^2 := \{(f - g)^2 : f, g \in \mathcal{F}\}$ . Then observe that

$$\begin{aligned} \sigma_n^2 &:= \sup_{f \in \mathcal{F}_\varepsilon} P_n f^2 \leq \sup_{f \in \mathcal{F}_\varepsilon} \mathbb{E}[f(X)^2] + n^{-1/2} \|\mathbb{G}_n\|_{(\mathcal{F} - \mathcal{F})^2} \\ &\lesssim (\varepsilon b)^2 + n^{-1/2} \|\mathbb{G}_n\|_{\mathcal{F}, \mathcal{F}} \lesssim \sigma^2 / n + b\sigma K_n^{1/2} / (\gamma^{2/q} n^{1/2}) + b^2 K_n / (\gamma^{2/q} n^{1-2/q}) \\ &\lesssim b\sigma K_n^{1/2} / (\gamma^{2/q} n^{1/2}) + b^2 K_n / (\gamma^{2/q} n^{1-2/q}), \end{aligned}$$

where in the second line, we used the inequality  $\|\mathbb{G}_n\|_{(\mathcal{F} - \mathcal{F})^2} = \sup_{f, g \in \mathcal{F}} |\mathbb{G}_n(f - g)^2| \leq 4\|\mathbb{G}_n\|_{\mathcal{F}, \mathcal{F}}$ . Now, note that conditional on  $X_1^n$ ,  $\mathbb{G}_n^e$  is a centered Gaussian process, and  $\mathbb{E}[(\mathbb{G}_n^e f)^2 \mid X_1^n] \leq P_n f^2 \leq \sigma_n^2$  for all  $f \in \mathcal{F}_\varepsilon$ . Hence by the Borell-Sudakov-Tsirel'son inequality [see 25, Proposition A.2.1],

$$\mathbb{P} \left\{ \|\mathbb{G}_n^e\|_{\mathcal{F}_\varepsilon} > \mathbb{E}[\|\mathbb{G}_n^e\|_{\mathcal{F}_\varepsilon} \mid X_1^n] + \sigma_n \sqrt{2 \log n} \mid X_1^n \right\} \leq 2n^{-1}.$$



To bound  $\mathbb{E}[\|\mathbb{G}_n^e\|_{\mathcal{F}_\varepsilon} \mid X_1^n]$ , observe that

$$\|\mathbb{G}_n^e\|_{\mathcal{F}_\varepsilon} \leq \sup_{f \in \mathcal{F}_\varepsilon} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n e_i f(X_i) \right| + \sup_{f \in \mathcal{F}_\varepsilon} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n e_i \cdot P_n f \right| =: I + II.$$

By Dudley's maximal inequality [see 25, Corollary 2.2.8], together with the fact that  $N(\mathcal{F}_\varepsilon, e_{P_n}, 2\tau\|F\|_{P_n,2}) \leq N^2(\mathcal{F}, e_{P_n}, \tau\|F\|_{P_n,2}) \leq (A/\tau)^{2v}$ , we have

$$\begin{aligned} \mathbb{E}[I \mid X_1^n] &\lesssim \int_0^{\sigma_n \vee (\sigma/n^{1/2})} \sqrt{1 + \log N(\mathcal{F}_\varepsilon, e_{P_n}, \tau)} d\tau \\ &\lesssim (\sigma_n \vee (\sigma/n^{1/2})) \sqrt{v \log(2n^{1/2} A \|F\|_{P_n,2} / \sigma)} \lesssim (\sigma_n \vee (\sigma/n^{1/2})) K_n^{1/2}. \end{aligned}$$

Meanwhile, since  $\|P_n\|_{\mathcal{F}_\varepsilon} \leq \sigma_n$  by Jensen's inequality, we have

$$\mathbb{E}[II \mid X_1^n] \leq \|P_n\|_{\mathcal{F}_\varepsilon} \cdot \mathbb{E} \left[ \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n e_i \right| \right] \lesssim \sigma_n.$$

Combining these inequalities leads to (21).

**Step 3.** Let  $Z^{e,\varepsilon} = \max_{1 \leq j \leq N} (B(f_j) + \mathbb{G}_n^e f_j)$  and  $\tilde{Z}^\varepsilon = \max_{1 \leq j \leq N} (B(f_j) + G_P f_j)$ . We wish to show that on the event  $E$ , the inequality

$$\mathbb{P}(Z^{e,\varepsilon} \in A \mid X_1^n) \leq \mathbb{P}(\tilde{Z}^\varepsilon \in A^\delta) + \frac{C}{\delta} \left\{ \frac{(b\sigma K_n^{3/2})^{1/2}}{\gamma^{1/q} n^{1/4}} + \frac{bK_n}{\gamma^{1/q} n^{1/2-1/q}} \right\}$$

holds for every  $\delta > 0$  and every Borel subset  $A$  of  $\mathbb{R}$ . Let

$$\Delta := \max_{1 \leq j, k \leq N} |\{P_n(f_j f_k) - (P_n f_j)(P_n f_k)\} - \{P(f_j f_k) - (P f_j)(P f_k)\}|,$$

and observe that

$$\begin{aligned} |P_n(f_j f_k) - P(f_j f_k)| &\leq n^{-1/2} \|\mathbb{G}_n\|_{\mathcal{F} \cdot \mathcal{F}}, \\ |(P_n f_j)(P_n f_k) - (P f_j)(P f_k)| &\lesssim n^{-1} \|\mathbb{G}_n\|_{\mathcal{F}} \cdot \|\mathbb{G}_n\|_{\mathcal{F}} + \sigma n^{-1/2} \|\mathbb{G}_n\|_{\mathcal{F}}. \end{aligned}$$

Hence as  $K_n \leq n$ , it is not difficult to check that on the event  $E$ ,

$$\Delta \lesssim b\sigma K_n^{1/2} / (\gamma^{2/q} n^{1/2}) + b^2 K_n / (\gamma^{2/q} n^{1-2/q}).$$

The assertion of this step now follows from Theorem 3.2 (recall  $\log N \lesssim K_n$ ).

**Step 4.** Take

$$\delta = \delta_n^{(2)} = \frac{(b\sigma K_n^{3/2})^{1/2}}{\gamma^{1+1/q}n^{1/4}} + \frac{bK_n}{\gamma^{1+1/q}n^{1/2-1/q}}.$$

Then the desired inequality (15) (with suitable  $C_3, C_4$ ) follows from combining (19), (20), Steps 1,2, and 3 with this choice of  $\delta$ .  $\blacksquare$

#### 4.3. Proof of Theorem 2.3

Here  $C$  denotes a positive constant depending only on  $q$ ;  $C$  may change from place to place. In addition, to ease the notation, we write  $a \lesssim b$  if  $a \leq Cb$ . In the proof below, we find an event  $E \in \sigma(X_1^n)$  such that  $\mathbb{P}(E) \geq 1 - \gamma - n^{-1}$ , and on this event, the inequality

$$\mathbb{P}(Z^* \in A \mid X_1^n) \leq \mathbb{P}\{Z^e \in A^{C(\eta+\delta_n^{(3)})} \mid X_1^n\} + C(\gamma + n^{-1}) \quad (22)$$

holds for every Borel subset  $A$  of  $\mathbb{R}$  where  $Z^e = \sup_{f \in \mathcal{F}} (B(f) + \mathbb{G}_n^e f)$ . Combining this inequality with (15), which is established in the proof of Theorem 2.2 (and which holds on a possibly different event  $E' \in \sigma(X_1^n)$  satisfying  $\mathbb{P}(E') \geq 1 - \gamma - n^{-1}$ ), the proof is completed by applying Lemma 4.2.

We first specify the event  $E$ . We use the same notation as introduced in Step 1 of the proof of Theorem 2.1. Then

$$0 \leq \sup_{f \in \mathcal{F}} (B(f) + \mathbb{G}_n^* f) - \max_{1 \leq j \leq N} (B(f_j) - \mathbb{G}_n^* f_j) \leq \eta + \|\mathbb{G}_n^*\|_{\mathcal{F}_\varepsilon}, \quad (23)$$

$$0 \leq \sup_{f \in \mathcal{F}} (B(f) + \mathbb{G}_n^e f) - \max_{1 \leq j \leq N} (B(f_j) - \mathbb{G}_n^e f_j) \leq \eta + \|\mathbb{G}_n^e\|_{\mathcal{F}_\varepsilon}. \quad (24)$$

In addition, as in the proof of Theorem 2.2, with probability at least  $1 - \gamma/4$ ,

$$\|\mathbb{G}_n\|_{\mathcal{F}} \lesssim \sigma K_n^{1/2}/\gamma^{1/q} + bK_n/(\gamma^{1/q}n^{1/2-1/q}); \quad (25)$$

with probability at least  $1 - \gamma/4$ ,

$$\|\mathbb{G}_n\|_{\mathcal{F}, \mathcal{F}} \lesssim b\sigma K_n^{1/2}/(\gamma^{2/q}) + b^2 K_n/(\gamma^{2/q}n^{1/2-2/q}), \quad (26)$$

and with probability at least  $1 - n^{-1}$ ,

$$\|F\|_{P_{n,2}} \leq n^{1/2}\|F\|_{P,2}. \quad (27)$$

Here  $\mathcal{F} \cdot \mathcal{F} = \{f \cdot g : f, g \in \mathcal{F}\}$ . Moreover, by the triangle inequality,

$$\max_{1 \leq j \leq N} \sum_{i=1}^n |f_j(X_i) - P_n f_j|^3 \lesssim \max_{1 \leq j \leq N} \sum_{i=1}^n |f_j(X_i)|^3$$

and applying Lemma 6.5, we have with probability at least  $1 - \gamma/4$ ,

$$\max_{1 \leq j \leq N} \sum_{i=1}^n |f_j(X_i)|^3 \lesssim \mathbb{E} \left[ \max_{1 \leq j \leq N} \sum_{i=1}^n |f_j(X_i)|^3 \right] + \gamma^{-3/q} \|M^3\|_{q/3},$$

where  $M := \max_{1 \leq i \leq n} \max_{1 \leq j \leq N} |f_j(X_i)| \leq \max_{1 \leq i \leq n} F(X_i)$  is such that  $\|M^3\|_{q/3} \lesssim n^{3/q} b^3$ . In addition, by Lemma 6.4,

$$\mathbb{E} \left[ \max_{1 \leq j \leq N} \sum_{i=1}^n |f_j(X_i)|^3 \right] \lesssim n\sigma^2 b + \mathbb{E}[M^3] \log N \lesssim n\sigma^2 b + n^{3/q} b^3 K_n.$$

Therefore, with probability at least  $1 - \gamma/4$ ,

$$\max_{1 \leq j \leq N} \sum_{i=1}^n |f_j(X_i) - P_n f_j|^3 / n \lesssim \sigma^2 b + b^3 K_n / (\gamma^{3/q} n^{1-3/q}). \quad (28)$$

Finally, by Markov's inequality, with probability at least  $1 - \gamma/4$ ,

$$\max_{1 \leq i \leq n} \max_{1 \leq j \leq N} |f_j(X_i) - P_n f_j| \lesssim \max_{1 \leq i \leq n} F(X_i) \lesssim \gamma^{-1/q} n^{1/q} b. \quad (29)$$

Define  $E$  as the intersection of the events in (25)-(29). Then  $E \in \sigma(X_1^n)$  and  $\mathbb{P}(E) \geq 1 - \gamma - n^{-1}$ . In the rest of the proof, which is divided into several steps, we prove (22) for each fixed  $X_1, \dots, X_n$  satisfying (25)-(29).

**Step 1.** By Step 2 in the proof of Theorem 2.2, on the event  $E$ ,

$$\mathbb{P} \left\{ \|\mathbb{G}_n^e\|_{\mathcal{F}_\varepsilon} > C \{ (b\sigma K_n^{3/2})^{1/2} / (\gamma^{1/q} n^{1/4}) + bK_n / (\gamma^{1/q} n^{1/2-1/q}) \} \mid X_1^n \right\} \leq 2n^{-1}.$$

**Step 2.** Here we wish to show that on the event  $E$ ,

$$\mathbb{P} \left\{ \|\mathbb{G}_n^*\|_{\mathcal{F}_\varepsilon} > C \{ (b\sigma K_n^{3/2})^{1/2} / (\gamma^{1/q} n^{1/4}) + bK_n / (\gamma^{1/q} n^{1/2-1/q}) \} \mid X_1^n \right\} \leq n^{-1}. \quad (30)$$

Note that conditional on  $X_1^n$ ,  $\mathbb{G}_n^*$  is the empirical process associated with  $n$  i.i.d. observations from the empirical distribution  $P_n$ . When restricted to

the domain  $\{X_1, \dots, X_n\}$ , the class of functions  $\mathcal{F}$  has a constant envelope  $\max_{1 \leq i \leq n} F(X_i) \lesssim \gamma^{-1/q} n^{1/q} b$ . Moreover, by the same arguments as those used in Step 2 of the proof of Theorem 2.2,

$$\sigma_n^2 := \sup_{f \in \mathcal{F}_\varepsilon} P_n f^2 \lesssim b \sigma K_n^{1/2} / (\gamma^{2/q} n^{1/2}) + b^2 K_n / (\gamma^{2/q} n^{1-2/q}).$$

Hence the inequality (30) follows from application of Talagrand's inequality (Lemma 6.3) with  $t = \log n$ .

**Step 3.** Let  $Z^{*,\varepsilon} = \max_{1 \leq j \leq N} (B(f_j) + \mathbb{G}_n^* f_j)$  and  $Z^{e,\varepsilon} = \max_{1 \leq j \leq N} (B(f_j) + \mathbb{G}_n^e f_j)$ . Here we apply Theorem 3.1 to show that whenever

$$\delta \geq C \left( \frac{b \log N}{\gamma^{1/q} n^{1/2-1/q}} + \frac{\sigma (\log^{3/2} N) \cdot (\log n)}{n^{1/2}} \right) \quad (31)$$

for some sufficiently large  $C > 0$ , on the event  $E$ , the inequality

$$\mathbb{P}(Z^{*,\varepsilon} \in A \mid X_1^n) \leq \mathbb{P}(Z^{e,\varepsilon} \in A^{C\delta} \mid X_1^n) + C \left( \frac{b \sigma^2 K_n^2}{\delta^3 n^{1/2}} + \frac{b^3 K_n^3}{\delta^3 \gamma^{3/q} n^{3/2-3/q}} + \frac{1}{n} \right)$$

holds for every  $\delta > 0$  and every Borel subset  $A$  of  $\mathbb{R}$ . Let  $\tilde{X}_i = (f_j(X_i) - P_n f_j)_{1 \leq j \leq N}$ ,  $1 \leq i \leq n$ , and let  $\tilde{Y} = (\mathbb{G}_n^e f_j)_{1 \leq j \leq N}$ . Then

$$L_n = \max_{1 \leq j \leq N} \sum_{i=1}^n |\tilde{X}_{ij}|^3 / n \lesssim \sigma^2 b + b^3 K_n / (\gamma^{3/q} n^{1-3/q}),$$

$$M_{n,X}(\delta) = n^{-1} \sum_{i=1}^n \max_{1 \leq j \leq N} |\tilde{X}_{ij}|^3 \cdot \mathbb{1} \left\{ \max_{1 \leq j \leq N} |\tilde{X}_{ij}| > \delta \sqrt{n} / \log N \right\} = 0.$$

The last equality follows from (29) since  $\delta \sqrt{n} / \log N \geq C \gamma^{-1/q} n^{1/q} b$ . Moreover,  $\mathbb{E}[\tilde{Y}_j^2] \leq P_n f_j^2 \leq \sigma^2 + n^{-1/2} \|\mathbb{G}_n\|_{\mathcal{F},\mathcal{F}}$  for all  $1 \leq j \leq N$ , and so by the same argument as that used in Step 4 of the proof of Theorem 2.1, we have

$$M_{n,Y}(\delta) = \mathbb{E} \left[ \max_{1 \leq j \leq N} |\tilde{Y}_j|^3 \cdot \mathbb{1} \left\{ \max_{1 \leq j \leq N} |\tilde{Y}_j| > \delta \sqrt{n} / \log N \right\} \mid X_1^n \right]$$

$$\lesssim n^{-2} (\delta \sqrt{n} / \log N)^3,$$

since  $\delta \geq C(\sigma^2 + n^{-1/2} \|\mathbb{G}_n\|_{\mathcal{F},\mathcal{F}})^{1/2} n^{-1/2} (\log^{3/2} N) \cdot (\log n)$  for sufficiently large  $C$ . The assertion of this step then follows from Theorem 3.1.

**Step 4.** In the previous step, take

$$\delta = C' \left\{ \frac{(b\sigma^2 K_n^2)^{1/3}}{\gamma^{1/3} n^{1/6}} + \frac{bK_n}{\gamma^{1/3+1/q} n^{1/2-1/q}} \right\}$$

where  $C' > 0$  is a large constant that can be chosen to depend only on  $q$ . It is not difficult to check that for this choice of  $\delta$ , (31) holds under the condition  $K_n^3 \leq n$ . The desired inequality (22) then follows from combining (23), (24), Steps 1, 2, and 3 with this choice of  $\delta$ .  $\blacksquare$

#### 4.4. Proof of Lemma 2.2

We begin with proving the following lemma.

**Lemma 4.3.** *Let  $X = (X_1, \dots, X_p)^T$  be a possibly non-centered Gaussian random vector with  $\sigma_j^2 := \text{Var}(X_j) > 0, 1 \leq j \leq p$ . Then for every  $\varepsilon > 0$ ,*

$$\sup_{t \in \mathbb{R}} \mathbb{P}(|\max_{1 \leq j \leq p} X_j - t| \leq \varepsilon) \leq \frac{2\varepsilon}{\underline{\sigma}} (\sqrt{2 \log p} + 2),$$

where  $\underline{\sigma} = \min_{1 \leq j \leq p} \sigma_j$ .

The lemma follows from the following result due essentially to Nazarov [21]; see also [17].

**Lemma 4.4** (Nazarov's inequality). *Let  $W$  be a standard Gaussian random vector in  $\mathbb{R}^m$ , that is,  $W \sim N(0, I)$ . Let  $A \subset \mathbb{R}^m$  be the intersection of  $p$  half-spaces (a half-space in  $\mathbb{R}^m$  is the set of form  $\{w \in \mathbb{R}^m : \alpha^T w \leq t\}$  for some  $\alpha \in \mathbb{R}^m$  with  $\|\alpha\| = 1$  and  $t \in \mathbb{R}$ ). Then*

$$\lim_{\delta \downarrow 0} \frac{1}{\delta} \mathbb{P}(W \in A^\delta \setminus A) \leq \sqrt{2 \log p} + 2,$$

where the limit on the left-hand side exists.

*Proof of Lemma 4.3.* It is clear that the distribution of  $\max_{1 \leq j \leq p} X_j$  is absolutely continuous, so let  $f(\cdot)$  denote its density. Let  $W$  be a standard Gaussian random vector in  $\mathbb{R}^p$ , and let  $\mu = (\mu_1, \dots, \mu_p)^T = \mathbb{E}[X]$  and  $\Sigma = \mathbb{E}[(X - \mu)(X - \mu)^T]$ . Then  $X \stackrel{d}{=} \Sigma^{1/2}W + \mu$ , so that denoting by  $\sigma_j a_j^T$  (where  $a_j \in \mathbb{R}^p$  with  $\|a_j\| = 1$ ) the  $j$ -th row of  $\Sigma^{1/2}$ , we obtain

$$\max_{1 \leq j \leq p} (\Sigma^{1/2}W + \mu)_j \leq t \Leftrightarrow a_j^T W \leq (t - \mu_j)/\sigma_j, 1 \leq j \leq p.$$

Let  $A_t = \{w \in \mathbb{R}^p : a_j^T w \leq (t - \mu_j)/\sigma_j, 1 \leq \forall j \leq p\}$  for  $t \in \mathbb{R}$ , and observe that

$$f(t) = \lim_{\varepsilon \downarrow 0} \frac{1}{\varepsilon} \mathbb{P}(W \in A_{t+\varepsilon} \setminus A_t) \quad \text{a.e. } t \in \mathbb{R}.$$

Moreover, since

$$A_{t+\varepsilon} \subset \{w \in \mathbb{R}^p : a_j^T w \leq (t - \mu_j)/\sigma_j + \varepsilon/\underline{\sigma}, 1 \leq \forall j \leq p\},$$

we have by Lemma 4.4,

$$\frac{1}{\varepsilon} \mathbb{P}(W \in A_{t+\varepsilon} \setminus A_t) \leq \frac{1}{\varepsilon} \mathbb{P}(W \in A_t^{\varepsilon/\underline{\sigma}} \setminus A_t) + o(1) \leq \frac{1}{\underline{\sigma}} (\sqrt{2 \log p} + 2) + o(1), \quad \varepsilon \downarrow 0,$$

which leads to  $f(t) \leq (1/\underline{\sigma})(\sqrt{2 \log p} + 2)$  a.e. ■

We are now in position to prove Lemma 2.2.

*Proof of Lemma 2.2.* Pick any  $\delta > 0$ , and let  $\{t_1, \dots, t_N\}$  be a  $\delta$ -net of  $(T, d)$  with  $N = N(T, d, \delta)$ . Then

$$\left| \sup_{t \in T} X(t) - \max_{1 \leq j \leq N} X(t_j) \right| \leq \sup_{(s,t) \in T_\delta} |X(t) - X(s)| =: \zeta,$$

so that for every  $x \in \mathbb{R}, \varepsilon, r' > 0$ ,

$$\mathbb{P}(\left| \sup_{t \in T} X(t) - x \right| \leq \varepsilon) \leq \mathbb{P}(\left| \sup_{1 \leq j \leq N} X(t_j) - x \right| \leq \varepsilon + r') + \mathbb{P}(\zeta > r').$$

By Lemma 4.3, the first term on the right-hand side is bounded by

$$2(1/\underline{\sigma})(\varepsilon + r')(\sqrt{2 \log N} + 2).$$

On the other hand, by the Borell-Sudakov-Tsirel'son inequality, for every  $r > 0$ ,

$$\mathbb{P}(\zeta > \mathbb{E}[\zeta] + r\delta) \leq e^{-r^2/2}.$$

By taking  $r' = \mathbb{E}[\zeta] + r\delta = \phi(\delta) + r\delta$ , we obtain the desired conclusion. ■

## 5. Proofs for Section 3

We begin with proving the following lemma.

**Lemma 5.1.** *Let  $\delta > 0$ . For every Borel subset  $A$  of  $\mathbb{R}$ , there exists a smooth function  $g : \mathbb{R} \rightarrow \mathbb{R}$  such that  $\|g'\|_\infty \leq \delta^{-1}$ ,  $\|g''\|_\infty \leq K\delta^{-2}$ ,  $\|g'''\|_\infty \leq K\delta^{-3}$ , where  $K$  is an absolute constant, and  $1_A(t) \leq g(t) \leq 1_{A^{3\delta}}(t)$  for all  $t \in \mathbb{R}$ .*

*Proof.* The proof is essentially similar to that of Lemma 18 in Chapter 10 of [22] with the exception that we employ a compactly supported smoother. Let  $\rho$  denote the Euclidean distance on  $\mathbb{R}$ , and consider the function  $h(t) = (1 - \rho(t, A^\delta)/\delta)_+$ . Observe that  $h$  is a bounded Lipschitz function with Lipschitz constant  $\delta$ . Let  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  be the function defined by  $\varphi(t) = C \exp(1/(t^2 - 1))$  for  $|t| \leq 1$  and  $\varphi(t) = 0$  for  $|t| > 1$ , where the constant  $C$  is chosen in such a way that  $\int_{\mathbb{R}} \varphi(t) dt = 1$ . Note that  $\varphi$  is infinitely differentiable with support  $[-1, 1]$ . Define  $g : \mathbb{R} \rightarrow \mathbb{R}$  by

$$g(t) = \int_{\mathbb{R}} h(t + \delta z) \varphi(z) dz = \delta^{-1} \int_{\mathbb{R}} h(y) \varphi(\delta^{-1}(y - t)) dy.$$

Then it is routine to verify that  $g$  is infinitely differentiable and  $\|g'\|_\infty \leq \delta^{-1}$ ,  $\|g''\|_\infty \leq K\delta^{-2}$ ,  $\|g'''\|_\infty \leq K\delta^{-3}$ . In addition, for  $t \in A$ ,  $h(t + \delta z) = 1$  if  $|z| \leq 1$ , and  $\varphi(z) = 0$  if  $|z| > 1$ . Hence  $1_A(t) \leq g(t)$ . Meanwhile, for  $t \notin A^{3\delta}$ ,  $h(t + \delta z) = 0$  if  $|z| \leq 1$ , and  $\varphi(z) = 0$  if  $|z| > 1$ . Hence  $g(t) \leq 1_{A^{3\delta}}(t)$ . ■

*Proof of Theorem 3.1.* Here we write  $a \lesssim b$  if there exists a universal constant  $C > 0$  such that  $a \leq Cb$ . Fix  $\delta > 0$ , and let  $\beta = \delta^{-1} \log p$ . Since  $p \geq 2$ , we have  $1/\delta \lesssim \beta$ . Let  $A$  be a Borel subset of  $\mathbb{R}$ . Letting  $e_\beta = \beta^{-1} \log p (= \delta)$  and using Lemma 5.1, we can construct a smooth function  $g : \mathbb{R} \rightarrow \mathbb{R}$  such that  $\|g'\|_\infty \leq \delta^{-1}$ ,  $\|g''\|_\infty \leq K\delta^{-2}$ ,  $\|g'''\|_\infty \leq K\delta^{-3}$  for some absolute constant  $K > 0$ , and  $1_{A^{e_\beta}}(t) \leq g(t) \leq 1_{A^{e_\beta + 3\delta}}(t)$  for all  $t \in \mathbb{R}$ . In addition, let  $\bar{\mu} = \sum_{i=1}^n \mu_i$  and consider the function  $F_\beta : \mathbb{R}^p \rightarrow \mathbb{R}$  defined by  $F_\beta(x) = \beta^{-1} \log(\sum_{j=1}^p e^{\beta(x_j + \bar{\mu}_j)})$ ,  $x \in \mathbb{R}^p$ . Then it is seen that  $\max_{1 \leq j \leq p} x_j \leq F_\beta(x - \bar{\mu}) \leq \max_{1 \leq j \leq p} x_j + e_\beta$  for all  $x \in \mathbb{R}^p$ . Hence

$$\mathbb{P}(Z \in A) \leq \mathbb{P}\left(F_\beta\left(n^{-1/2} \sum_{i=1}^n \tilde{X}_i\right) \in A^{e_\beta}\right). \quad (32)$$

Next, let  $m = g \circ F_\beta$ . Then, as in the proof of Lemma 5.1 in [9] (see also

[3, 4]), there exist functions  $U_{jkl} : \mathbb{R}^p \rightarrow \mathbb{R}$ ,  $1 \leq j, k, l \leq p$  such that

$$\begin{aligned} |\partial_j \partial_k \partial_l m(x)| &\leq U_{jkl}(x), \quad \forall x \in \mathbb{R}^p, \\ \sum_{j,k,l=1}^p U_{jkl}(x) &\lesssim (\delta^{-3} + \beta \delta^{-2} + \beta^2 \delta^{-1}) \lesssim \beta^2 \delta^{-1}, \quad \forall x \in \mathbb{R}^p, \\ U_{jkl}(x) &\lesssim U_{jkl}(x+y) \lesssim U_{jkl}(x), \quad \forall x, y \in \mathbb{R}^p \text{ with } \max_{1 \leq j \leq p} |y_j| \leq \beta^{-1}. \end{aligned}$$

Hence proceeding as in Step 1 of the proof of Lemma 5.1 in [9] and observing that the term  $\int_0^1 \omega(t) \mathbb{E}[h(Z^{(n)}, 6)] dt$  in that paper is trivially bounded by a universal constant, one can show that for some universal constant  $c > 0$ ,

$$\begin{aligned} &\left| \mathbb{E} \left[ m \left( n^{-1/2} \sum_{i=1}^n \tilde{X}_i \right) \right] - \mathbb{E} \left[ m \left( n^{-1/2} \sum_{i=1}^n \tilde{Y}_i \right) \right] \right| \\ &\lesssim \frac{\log^2 p}{\delta^3 \sqrt{n}} \cdot \{L_n + M_{n,X}(c\delta) + M_{n,Y}(c\delta)\} =: I, \end{aligned}$$

which implies that for some universal constant  $C$ ,

$$\begin{aligned} \mathbb{P} \left( F_\beta \left( n^{-1/2} \sum_{i=1}^n \tilde{X}_i \right) \in A^{e_\beta} \right) &\leq \mathbb{E} \left[ m \left( n^{-1/2} \sum_{i=1}^n \tilde{X}_i \right) \right] \\ &\leq \mathbb{E} \left[ m \left( n^{-1/2} \sum_{i=1}^n \tilde{Y}_i \right) \right] + CI \leq \mathbb{P} \left\{ F_\beta \left( n^{-1/2} \sum_{i=1}^n \tilde{Y}_i \right) \in A^{e_\beta + 3\delta} \right\} + CI \\ &\leq \mathbb{P} \left( \tilde{Z} \in A^{2e_\beta + 3\delta} \right) + CI. \end{aligned}$$

Combining this inequality with (32) leads to the conclusion of the theorem.

■

*Proof of Theorem 3.2.* Since  $p \geq 2$ , the assertion is trivial if  $\Delta/\delta^2 > 1$ . Therefore, throughout the proof, we will assume that  $\Delta/\delta^2 \leq 1$ . Let  $\beta > 0$ , and define  $F_\beta : \mathbb{R}^p \rightarrow \mathbb{R}$  by  $F_\beta(x) = \beta^{-1} \log(\sum_{j=1}^p e^{\beta(x_j + \mu_j)})$  where  $x = (x_1, \dots, x_p)^T$  and  $\mu = (\mu_1, \dots, \mu_p)^T$ . As in the proof of Theorem 3.1, it can be shown that for every  $g \in C^2(\mathbb{R})$ , the function  $m = g \circ F_\beta$  satisfies the inequality

$$\sum_{j,k=1}^p |\partial_j \partial_k m(x)| \leq \|g''\|_\infty + 2\|g'\|_\infty \beta$$

for all  $x \in \mathbb{R}^p$ . Hence using the same arguments as those used in the proof of Theorem 1 and Comment 1 in [8] with  $X$  and  $Y$  replaced by  $X - \mu$  and  $Y - \mu$ , respectively, we have

$$\left| \mathbb{E} \left[ g \left( \max_{1 \leq j \leq p} X_j \right) \right] - \mathbb{E} \left[ g \left( \max_{1 \leq j \leq p} Y_j \right) \right] \right| \leq \|g''\|_\infty \Delta/2 + 2\|g'\|_\infty \sqrt{2\Delta \log p}.$$



Now, take any Borel subset  $A$  of  $\mathbb{R}$ . By Lemma 5.1, we can construct a function  $g \in C^2(\mathbb{R})$  such that  $\|g'\|_\infty \leq \delta^{-1}$  and  $\|g''\|_\infty \leq K\delta^{-2}$  for some absolute constant  $K$ , and  $1_A(t) \leq g(t) \leq 1_{A^{3\delta}}(t)$  for all  $t \in \mathbb{R}$ . For this  $g$  and some absolute constant  $C$ , we have

$$\begin{aligned} \mathbb{P}\left(\max_{1 \leq j \leq p} X_j \in A\right) &\leq \mathbb{E}\left[g\left(\max_{1 \leq j \leq p} X_j\right)\right] \\ &\leq \mathbb{E}\left[g\left(\max_{1 \leq j \leq p} Y_j\right)\right] + C(\Delta\delta^{-2} + \delta^{-1}\sqrt{\Delta\log p}) \\ &\leq \mathbb{P}\left(\max_{1 \leq j \leq p} Y_j \in A^{3\delta}\right) + C(\Delta\delta^{-2} + \delta^{-1}\sqrt{\Delta\log p}) \\ &\leq \mathbb{P}\left(\max_{1 \leq j \leq p} Y_j \in A^{3\delta}\right) + C\sqrt{(\Delta/\delta^2)\log p} \end{aligned}$$

where the last line follows from the fact that  $\Delta/\delta^2 \leq 1$  and  $p \geq 2$ . The conclusion of the theorem follows from replacing  $\delta$  by  $\delta/3$ .  $\blacksquare$

## 6. Some technical tools

**Lemma 6.1.** *Let  $X_1, \dots, X_n$  be i.i.d. random variables taking values in a measurable space  $(S, \mathcal{S})$  with common distribution  $P$ . Let  $\mathcal{F}$  be a pointwise measurable class of functions  $f : S \rightarrow \mathbb{R}$ , to which a measurable envelope  $F$  is attached. Consider the empirical process  $\mathbb{G}_n f = n^{-1/2} \sum_{i=1}^n (f(X_i) - Pf)$ ,  $f \in \mathcal{F}$ . Let  $\sigma^2 > 0$  be a constant such that  $\sup_{f \in \mathcal{F}} Pf^2 \leq \sigma^2 \leq \|F\|_{P,2}^2$ . Let  $M = \max_{1 \leq i \leq n} F(X_i)$ . Suppose that  $F \in \mathcal{L}^q(P)$  for some  $q \geq 2$ . Then for every  $t \geq 1$ , with probability  $> 1 - t^{-q/2}$ ,*

$$\begin{aligned} \|\mathbb{G}_n\|_{\mathcal{F}} &\leq (1 + \alpha)\mathbb{E}\|\mathbb{G}_n\|_{\mathcal{F}} + K_q \left\{ (\sigma + n^{-1/2}\|M\|_q)\sqrt{t} \right. \\ &\quad \left. + \alpha^{-1}n^{-1/2}\|M\|_2 t \right\}, \quad \forall \alpha > 0, \end{aligned}$$

where  $K_q > 0$  is a constant that depends only  $q$ .

*Proof.* The lemma is essentially due to [2], Theorem 12. See Theorem 5.1 in [6] for the version stated here.  $\blacksquare$

**Lemma 6.2.** *Consider the setting of Lemma 6.1. In addition, suppose that there exist constants  $A \geq e$  and  $v \geq 1$  such that  $\sup_Q N(\mathcal{F}, e_Q, \varepsilon\|F\|_{Q,2}) \leq (A/\varepsilon)^v$ ,  $0 < \varepsilon \leq 1$ . Then*

$$\mathbb{E}\|\mathbb{G}_n\|_{\mathcal{F}} \leq K \left\{ \sqrt{v\sigma^2 \log\left(\frac{A\|F\|_{P,2}}{\sigma}\right)} + \frac{v\|M\|_2}{\sqrt{n}} \log\left(\frac{A\|F\|_{P,2}}{\sigma}\right) \right\},$$

where  $K$  is an absolute constant.

*Proof.* See Corollary 5.1 in [6]. ■

**Lemma 6.3** (Talagrand's inequality). *Consider the setting of Lemma 6.2, but suppose now that the envelope  $F$  is bounded by a constant  $b > 0$ , and let  $\sigma^2 > 0$  be a constant such that  $\sup_{f \in \mathcal{F}} Pf^2 \leq \sigma^2 \leq b^2$ . If  $b^2 v \log(Ab/\sigma) \leq n\sigma^2$ , then for every  $0 < t \leq n\sigma^2/b^2$ ,*

$$\mathbb{P} \left\{ \|\mathbb{G}_n\|_{\mathcal{F}} > K\sigma \sqrt{t \vee (v \log(Ab/\sigma))} \right\} \leq e^{-t},$$

where  $K$  is an absolute constant.

*Proof.* This form of Talagrand's inequality is taken from Theorem B.1 in [7]; the original references go back to [24], [19], and [15]. ■

**Lemma 6.4.** *Let  $X_1, \dots, X_n$  be independent random vectors in  $\mathbb{R}^p$  with  $p \geq 2$  such that  $X_{ij} \geq 0$  for all  $i = 1, \dots, n$  and  $j = 1, \dots, p$ . Define  $Z := \max_{1 \leq j \leq p} \sum_{i=1}^n X_{ij}$  and  $M := \max_{1 \leq i \leq n} \max_{1 \leq j \leq p} X_{ij}$ . Then*

$$\mathbb{E}[Z] \leq K \left( \max_{1 \leq j \leq p} \mathbb{E}[\sum_{i=1}^n X_{ij}] + \mathbb{E}[M] \log p \right),$$

where  $K$  is an absolute constant.

*Proof.* See Lemma 9 in [8]. ■

**Lemma 6.5.** *Assume the setting of Lemma 6.4. Then for every  $\eta > 0$ ,  $s \geq 1$  and  $t > 0$ ,*

$$\mathbb{P}(Z \geq (1 + \eta)\mathbb{E}[Z] + t) \leq K\mathbb{E}[M^s]/t^s,$$

where  $K = K(\eta, s)$  is a constant that depends only on  $\eta, s$ .

*Proof.* See Lemma A.5 in [9]. ■

**Lemma 6.6.** *Let  $\xi$  be a nonnegative random variable such that  $\mathbb{P}(\xi > x) \leq Ae^{-x/B}$  for all  $x > 0$  and for some constants  $A, B > 0$ . Then for every  $t > 0$ ,  $\mathbb{E}[\xi^3 1\{\xi > t\}] \leq 6A(t + B)^3 e^{-t/B}$ .*

*Proof.* See Lemma A.8 in [9]. ■

## Acknowledgments

We would like thank an anonymous referee for valuable comments that helped improve upon the quality of the paper. K. Kato is supported by the Grant-in-Aid for Scientific Research (C) (15K03392) from the Japan Society for the Promotion of Science. V. Chernozhukov is supported by a grant from the National Science Foundation.

## References

### References

- [1] Beck, J. (1985). Lower bounds on the approximation of the multivariate empirical process. *Z. Wahrsch. verw. Gebiete.* **70** 289-306.
- [2] Boucheron, S., Bousquet, O., Lugosi, G., and Massart, P. (2005). Moment inequalities for functions of independent random variables. *Ann. Probab.* **33** 514-560.
- [3] Chatterjee, S. (2005). A simple invariance theorem. arXiv:math/0508213.
- [4] Chatterjee, S. (2005). An error bound in the Sudakov-Fernique inequality. arXiv:math/0510424.
- [5] 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.
- [6] Chernozhukov, V., Chetverikov, D. and Kato, K. (2014). Gaussian approximation of suprema of empirical processes. *Ann. Statist.* **42** 1564-1597.
- [7] Chernozhukov, V., Chetverikov, D. and Kato, K. (2014). Anti-concentration and honest, adaptive confidence bands. *Ann. Statist.* **42** 1787-1818.
- [8] Chernozhukov, V., Chetverikov, D. and Kato, K. (2014). Comparison and anti-concentration bounds for maxima of Gaussian random vectors. *Probab. Theory Related Fields*, to appear. Available at arXiv:1301.4807v3.

- [9] Chernozhukov, V., Chetverikov, D. and Kato, K. (2014). Central limit theorems and bootstrap in high dimensions. Available at arXiv:1412.3661v2.
- [10] Davydov, Y., Lifshits, M., and Smorodina, N. (1998). *Local Properties of Distributions of Stochastic Functions*. Transaction of Mathematical Monographs, Vol. 173. American Mathematical Society.
- [11] Dudley, R.M. (1979). Balls in  $\mathbb{R}^k$  do not cut all subsets of  $k + 2$  points. *Adv. Math.* **31** 306-308.
- [12] Dudley, R.M. (1999). *Uniform Central Limit Theorems*. Cambridge University Press.
- [13] Dumbgen, L. and Spokoiny, V. (2001). Multiscale testing of qualitative hypotheses. *Ann. Statist.* **29**, 124-152.
- [14] Giné, E. and Nickl, R. (2015). *Mathematical Foundations of Infinite-Dimensional Statistical Models*. Cambridge University Press, to appear. Available at the authors' websites.
- [15] Giné, E. and Guillou, A. (2001). A law of the iterated logarithm for kernel density estimators in the presence of censoring. *Ann. Inst. Henri Poincaré Probab. Stat.* **37** 503-522.
- [16] Koltchinskii, V. (1994). Komlós-Major-Tusnády approximation for the general empirica process and Haar expansions of classes of functions. *J. Theoret. Probab.* **7**, 73-118.
- [17] Klivans, A.R., O'Donnell, R. and Servedio, R. (2008). Learning geometric concepts via Gaussian surface area. In: *Proceedings of the 49th Foundations of Computer Science (FOCS)*.
- [18] Massart, P. (1989). Strong approximation for multivariate empirical and related processes, via KMT constructions. *Ann. Probab.* **17** 266-291.
- [19] Massart, P. (2000). About the constants in Talagrand's concentration inequalities for empirical processes. *Ann. Probab.* **28**, 863-884.
- [20] Monrad, D. and Philipp, W. (1991). Nearby variables with nearby conditional laws and a strong approximation theorem for Hilbert space valued martingales. *Probab. Theory Related Fields.* **88** 381-404.

- [21] 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 (2001-2002)*, p. 169-187. Lecture Notes in Math., Vol. 1807, Springer.
- [22] Pollard, D. (2002). *A User's Guide to Measure Theoretic Probability*. Cambridge University Press.
- [23] Rio, E. (1994). Local invariance principles and their application to density estimation. *Probab. Theory Related Fields* **98** 21-45.
- [24] Talagrand, M. (1996). New concentration inequalities in product spaces. *Invent. Math.* **126** 505-563.
- [25] van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer.