High-dimensional CLT: Improvements, Non-uniform Extensions and Large Deviations

ARUN KUMAR KUCHIBHOTLA, SOMABHA MUKHERJEE, and DEBAPRATIM BANERJEE

Department of Statistics, The Wharton School
University of Pennsylvania
3730 Walnut Street,
Jon M. Huntsman Hall
Philadelphia, PA 19104, USA.
E-mail: arunku@upenn.edu, somabha@upenn.edu, dban@upenn.edu

Central limit theorems (CLTs) for high-dimensional random vectors with dimension possibly growing with the sample size have received a lot of attention in the recent times. Chernozhukov et al. (2017b) proved a Berry–Esseen type result for high-dimensional averages for the class of sparsely convex sets including hyperrectangles as a special case and they proved that the rate of convergence can be upper bounded by $n^{-1/6}$ up to a polynomial factor of $\log p$ (where $n$ represents the sample size and $p$ denotes the dimension). Convergence to zero of the bound requires $\log^7 p = o(n)$. We improve upon their result, for hyperrectangles, which only requires $\log^4 p = o(n)$ (in the best case). This improvement is made possible by a sharper dimension-free anti-concentration inequality for Gaussian process on a compact metric space. In addition, we prove two non-uniform variants of the high-dimensional CLT based on the large deviation and non-uniform CLT results for random variables in a Banach space by Bentkus, Rackauskas, and Paulauskas. We apply our results in the context of post-selection inference in linear regression and of empirical processes.

Keywords: Orlicz norms; Nonuniform CLT; Cramér type large deviation; Anti-concentration; Post-selection Inference; Empirical Processes.

1. Introduction

In modern statistical applications like high dimensional estimation and multiple hypothesis testing problems (Belloni et al., 2018), the dimension of the data is often much larger than the sample size. As can be expected from the classical asymptotic theory, the central limit theorem plays a pivotal role for inference. In this paper, we prove three variants of the high-dimensional central limit theorem. The setting we use is as follows. Consider independent mean zero random vectors $X_1, \ldots, X_n \in \mathbb{R}^p$ with covariance matrices $\Sigma_i := \mathbb{E}[X_iX_i^\top] \in$
Here $p$ is allowed to be larger than $n$. Define the scaled average
\[ S_n := \frac{X_1 + \cdots + X_n}{\sqrt{n}} \in \mathbb{R}^p. \]

Let $Y_i$ (for $1 \leq i \leq n$) represent a multivariate Gaussian random vector with mean zero and variance-covariance matrix $\Sigma_i$. Define the corresponding scaled average as
\[ U_{n,0} := \frac{Y_1 + \cdots + Y_n}{\sqrt{n}} \in \mathbb{R}^p. \]

The problem of central limit theorem is the comparison of the probabilities $\mathbb{P}(S_n \in A)$ and $\mathbb{P}(U_{n,0} \in A)$ for $A \subseteq \mathbb{R}^p$. When $p$ is fixed (or only grows at most sublinearly in $n$) which we refer to as multivariate setting, classical results show the rate $O(p^{7/4}/n^{1/2})$; see Sazonov (1981), Sazonov and Ulyanov (1982) and Bentkus (2004). The case where $p$ is allowed to grow faster than $n$, which we refer to as high-dimensional setting, has received significant interest in the recent times. The series of papers Chernozhukov et al. (2013, 2015, 2017b) have studied this problem extensively under general conditions on the random vectors when the sets $A$ are sparsely convex sets and in particular hyperrectangles. The main result of Chernozhukov et al. (2017b) bounds the difference $|\mathbb{P}(S_n \in A) - \mathbb{P}(U_{n,0} \in A)|$ uniformly over $A \in \mathcal{A}^c$ as a function of $n$ and $p$. Here $\mathcal{A}^c$ is the class of all hyperrectangles. Proposition 2.1 of Chernozhukov et al. (2017b) implies that
\[ \sup_{A \in \mathcal{A}^c} |\mathbb{P}(S_n \in A) - \mathbb{P}(U_{n,0} \in A)| \leq C \left( \frac{\log^7(pn)}{n} \right)^{1/6}, \] 
under certain exponential tail assumption and a constant $C$ depending on the distribution of the random vectors $X_i$, $1 \leq i \leq n$. In their earlier papers, Chernozhukov et al. (2013, 2015) a special sub-class of sets $\mathcal{A}^m \subset \mathcal{A}^c$ were considered, where $\mathcal{A}^m$ is the class of all sets $A$ of the form $A = \{x \in \mathbb{R}^p : x(j) \leq a \text{ for all } 1 \leq j \leq p\}$. Here and throughout, we use the notation $x(j)$ for a vector $x \in \mathbb{R}^p$ to represent the $j$-th coordinate of $x$.

From the bound (1) we require $\log^7(pn) = o(n)$ for the difference of probabilities to converge to zero uniformly. One of the lingering questions in high-dimensional CLT is the “correct” exponent of $\log(pn)$ that needs to be $o(n)$ for convergence to zero. By proving a dimension-free anti-concentration inequality and comparing the proof techniques from Chernozhukov et al. (2017b), Paulauskas and Račkauskas (1989), Bentkus et al. (2000), we reduce the requirement to $\mu^6 \log^4(ep) = o(n)$ when the sets $A$ are restricted to $l_\infty$ balls in $\mathbb{R}^p$. Here $\mu$ represents the median of $\|U_{n,0}\|_\infty$. In the best case where $\mu \asymp 1$ the requirement becomes $\log^4(ep) = o(n)$ and in the worst case where $\mu \asymp \sqrt{\log(ep)}$ the requirement becomes $\log^7(ep) = o(n)$ coinciding with the requirement from (1).

Since the dependence, $n^{-1/6}$, on the sample size in bound (1) is larger than dependence $n^{-1/2}$ that appears in multivariate Berry–Esseen bounds, the result (1) does not provide
useful information when the probability $P(U_{n,0} \in A)$ is smaller. This leads naturally to the question of non-uniform version of (1). In particular, an interesting question is to find quantitative upper bound on

$$\left| \frac{P(S_n \in A^c)}{P(U_{n,0} \in A^c)} - 1 \right|,$$

as a function of $p$ and $n$. In this paper we consider a special class of sets of the form

$$A = \{ x \in \mathbb{R}^p : -a \leq x(j) \leq a \text{ for all } 1 \leq j \leq p \},$$

and find an upper bound on (2). Note that sets of the form (3) are $l_\infty$ balls. Another variant of non-uniform CLT is to consider how the difference $|P(\|S_n\|_\infty \leq r) - P(\|U_{n,0}\|_\infty \leq r)|$ scales with $r$ for $r \geq 0$. To this end we prove upper bounds on

$$\sup_{r \geq 0} r^m |P(\|S_n\|_\infty \leq r) - P(\|Y\|_\infty \leq r)|, \text{ for } m \geq 0. \tag{4}$$

The first problem (2) is well-studied in the classical multivariate setting under the name “Cramér-type large deviation”. We refer to the encyclopedic work Saulis and Statulevičius (1991) for a review of the extensive literature on Cramér-type large deviation for sums of independent random variables along with extensions to multivariate random vectors. The classical result for univariate ($p = 1$) i.i.d. random variables is of the form

$$\frac{P(S_n > r)}{P(U_{n,0} > r)} = \exp \left( \frac{Cr^3}{6\Sigma_1^{3/2}n^{1/2}} \right) \left[ 1 + C \left( \frac{r + 1}{\sqrt{n}} \right) \right], \tag{5}$$

for $0 \leq r = O(n^{1/6})$ and $C$ a constant depending on the distribution of $X_1$. See Theorem 5.23 and Section 5.8 of Petrov (1995) for details. The second problem (4) is usually referred to as a “non-uniform CLT”. A result of this kind is also useful in proving convergence of moments. The classical result for univariate iid random variables of type (4) is given by

$$|P(S_n \leq r) - P(U_{n,0} \leq r)| \leq C(m)(1 + |r|)^{-m} \left( \frac{\mathbb{E}[|X_1|^3]}{\Sigma_1^{3/2}n^{1/2}} + \frac{\mathbb{E}[|X_1|^m]}{\Sigma_1^{m/2}n^{m-2/2}} \right), \tag{6}$$

for all $r \in \mathbb{R}, m \geq 3$ and for some constant $C(m)$ depending only on $m$. See Theorem 5.15 and Section 5.5 of Petrov (1995) for details and other results in this direction. Also, see Sazonov (1981) for multivariate setting. The classical results (5) and (6) provide rates that scale like $n^{-1/2}$ (as a function of $n$) in the non-uniform versions of CLT as does the classical Berry–Esseen bound. In lines with the Berry–Esseen type result (1) in high-dimensional setting, we derive rates in large deviation and non-uniform CLT with a scaling of order $n^{-1/6}$ as a function of the sample size $n$.\(^1\)

\(^1\)Koike (2019b) and Chernozhukov et al. (2019) recently proved an improvement in (1) with a scaling of order $n^{-1/4}$. The proof techniques are far more complicated and will not be compared with in this paper.
It is well-known (Bentkus, 1985) that the rate \( n^{-1/6} \) is optimal in the central limit theorem for Banach spaces and the space \((\mathbb{R}^p, \|\cdot\|_{\infty})\) with \( p \) diverging behaves like an infinite-dimensional space. In this respect it is of particular interest to look back at the rich literature on the CLTs for Banach space valued random variables. These old and well-known large deviation and non-uniform CLTs for Banach space play a central role in the derivation of ours presented here. The basic setting for these results is as follows: Suppose \( X_1, \ldots, X_n \) are i.i.d. random variables taking values in a Banach space \( B \) such that \( \mathbb{E}[X_1] = 0 \) and \( Y \) is a mean zero \( B \)-valued Gaussian random variable with the same covariance (operator) as \( X_1 \). The problem as before is the study of closeness of the distributions of \( \|Y\| \) and \( \|S_n\| \) where \( \|\cdot\| \) is a Banach space norm. Several results on this problem are available in Bentkus et al. (2000), Paulauskas and Račkauskas (2012). For a historical account of these results, see Paulauskas and Račkauskas (2012, p.142). The papers Bentkus (1987), Bentkus and Račkauskas (1990) and Paulauskas and Račkauskas (1991) are of particular interest to us since they provide bounds on (2) and (4) for Banach space valued random variables. In Bentkus (1987) and Bentkus and Račkauskas (1990) the problem of the convergence of ratio \( \mathbb{P}(\|S_n\| > r) / \mathbb{P}(\|Y\| > r) \) to 1 was considered and it was proved that

\[
| \mathbb{P}(\|S_n\| > r) / \mathbb{P}(\|Y\| > r) - 1 | \leq M_2(r + 1)n^{-1/6},
\]

for \( 0 \leq r \leq -1 + M_1n^{1/6} \) where \( M_1 \) and \( M_2 \) are constants depending on the distribution of \( X_1 \). The non-uniform version of central limit theorem is also available for the Banach spaces from Paulauskas and Račkauskas (1991). Understanding the implication of these results for the high-dimensional case is one contribution of our paper.

1.1. Our contributions

As described in the introduction, we study uniform and non-uniform variants of high dimensional central limit theorem. In the process we prove a sharper version of anti-concentration inequality for centered Gaussians. We assume that \( X_1, \ldots, X_n \in \mathbb{R}^p \) are independent random vectors with mean 0 and covariance matrices \( \Sigma_i = \mathbb{E}[X_iX_i^\top] \in \mathbb{R}^{p \times p} \). Let \( Y_1, \ldots, Y_n \) be centered Gaussian random vectors in \( \mathbb{R}^p \) satisfying \( \mathbb{E}[Y_iY_i^\top] = \Sigma_i \) for \( 1 \leq i \leq n \). Our results are intended for the case where \( \log(ep) \) grows sublinearly in the sample size \( n \), although we do not explicitly make an assumption that \( \log(ep) \) or \( p \) grows with \( n \). Define for \( m \geq 0 \)

\[
\Delta_n^{(m)} := \sup_{r \geq 0} r^m | \mathbb{P}(\|S_n\|_{\infty} \leq r) - \mathbb{P}(\|U_{n,0}\|_{\infty} \leq r) |.
\]

(i) One of the main ingredients of central limit theorems (both uniform and non-uniform versions) is an anti-concentration inequality which bounds \( \mathbb{P}(r - \varepsilon \leq \|U_{n,0}\|_{\infty} \leq r + \varepsilon) \)
over all \( r \geq 0 \) and \( \varepsilon > 0 \). We prove that for any \( \varepsilon > 0, m \geq 0 \)

\[
\sup_{r \geq 0} r^m \mathbb{P}(r - \varepsilon \leq \|U_{n,0}\|_\infty \leq r + \varepsilon) \leq \Phi_{AC,m} \varepsilon,
\]

where \( \Phi_{AC,m} = O(\mu^{m+1}) \) (assuming \( \sigma_{\text{max}} \) and \( \sigma_{\text{min}}^{-1} \) are of constant order) and \( \mu \) denotes the median of \( \|U_{n,0}\|_\infty \). Quantities \( \sigma_{\text{max}}^2 \) and \( \sigma_{\text{min}}^2 \) are given by the maximum and minimum variances of \( U_{n,0}(j), 1 \leq j \leq p \). This provides a refinement of Lemma 3.1 of Bentkus and Račkauskas (1990) with exact constants. Chernozhukov et al. (2017b) (based on the result of Nazarov (2003)) prove the above result for \( m = 0 \) case with \( \Phi_{AC,0} = C \sqrt{\log(ep)} \) and since \( \mu \) is at most of the order \( \sqrt{\log(ep)} \), our result is sharper.

(ii) We compare the modern proof technique of Chernozhukov et al. (2017b) and the classical proofs from the Banach space CLT literature Paulauskas and Račkauskas (2012), Bentkus et al. (2000); see Section 6 for details. Based on this we improve upon the proof of Chernozhukov et al. (2017b) to get better rates in high-dimensional CLT.

(iii) If \( X_i \) are sub-Weibull of order \( \alpha \), i.e., \( \|X_i\|_{\psi_\alpha} \leq K_\psi < \infty \) for all \( 1 \leq i \leq n, 1 \leq j \leq p \), then

\[
\Delta_n^{(0)} \leq \Theta_{\alpha} \Phi_{AC,0} K_\psi \left( \frac{\log^4(ep)}{n} \right)^{1/6} + \Theta_{\alpha} \Phi_{AC,0} K_\psi \left( \frac{\log(ep)}{n^{1/2}} \right)^{1+1/\alpha}(\log(n))^{1/\alpha}.
\]

for some constant \( \Theta_{\alpha} \) depending only on \( \alpha \). Proposition 2.1 of Chernozhukov et al. (2017b) for \( \alpha = 1 \) proves that \( \Delta_n^{(0)} \leq C(\log^7(pm)/n)^{1/6} \) which requires \( \log(pm) = o(n^{1/7}) \). In contrast if \( \Phi_{AC,0} = O(1) \) then our result only requires \( \log(ep) = o(n^{1/4}) \) which is the weakest requirement till date.

(iv) Under the same assumptions in (iii), we have for \( m \geq 0 \)

\[
\Delta_n^{(m)} \leq \Theta_{\alpha} \Phi_{AC,m} \left( \frac{\log^4(epm)}{n} \right)^{1/6} + \begin{cases} 0, & \text{if } \alpha > 1, \\ \Theta_{\alpha,m}((\log(epm)^{5+3/\alpha}/n)^{(m+1)/3}, & \text{if } \alpha \in (0,1].
\end{cases}
\]

More generally, for any function \( \phi(\cdot) \) satisfying \( \phi(x + y) \leq \phi(x)\phi(y) \) our methods can be used to obtain bounds for

\[
\Delta_n^{(\phi)} := \sup_{r \geq 0} \phi(r) |\mathbb{P}(\|S_n\|_{\infty} \leq r) - \mathbb{P}(\|U_{n,0}\|_{\infty} \leq r)|.
\]

These are analogues to the results of Paulauskas and Račkauskas (1991).

(v) Finally, we derive a Cramér-type large deviation in the high dimensional CLT setting. We assume that \( X_1, \ldots, X_n \) are independent and identically distributed and that \( \mathbb{E}[\exp \{H \|X_1\|_{\infty}\}] < \infty \) for some \( H > 0 \). Under these assumptions, we prove that

\[
|\mathbb{P}(\|S_n\|_{\infty} > r)/\mathbb{P}(\|Y\|_{\infty} > r) - 1| \leq M_1(r + 1)r^{-1/6},
\]
for any \( r \leq -1 + M_2 n^{1/6} \). Here \( M_1 \) and \( M_2 \) are constants depending on the distributions of \( X_1 \) and \( Y \) which can be bounded by polynomials of \( \log p \), under certain tail assumptions on the coordinates of \( X_1 \). The proof is motivated by the techniques of Bentkus (1987) and is modified for the high dimensional set up. The constants \( M_1 \) and \( M_2 \) are also made explicit in Theorem 5.1.

1.2. Organization of the paper

The paper is organized as follows. In Section 2, we define some useful notations. Section 3 is dedicated to our main results. In this section, we state the anti-concentration inequalities and prove refined uniform as well as non-uniform central limit theorems. In Section 4 we present applications of our results to post-selection inference where the anti-concentration constant \( \Phi_{AC,0} \) can be of order much smaller than \( \sqrt{\log(ep)} \) and to bounding the expectation of suprema of empirical processes over a (possibly infinite) weak VC-major function class. In Section 5, we prove a Cramér-type large deviation based on the results of Bentkus (1987). In Section 6, we present an outline of our proof of CLTs with detailed discussion on differences of proofs from other works. Finally, we conclude with a summary and future directions in Section 7. Proofs of all the results are given in the supplementary material.

2. Preliminaries

2.1. Notation and Setting

As discussed earlier, we shall consider independent random vectors \( X_1, \ldots, X_n \in \mathbb{R}^p \) with mean zero and covariance matrices \( \Sigma_i, 1 \leq i \leq n \). Let \( Y_1, \ldots, Y_n \in \mathbb{R}^p \) denote Gaussian random vectors with mean and covariance matching that of \( X_i, 1 \leq i \leq n \). The \( l_{\infty} \) norm on \( \mathbb{R}^p \) is denoted by \( \| \cdot \| \). By writing \( a \lesssim b \), we mean that for some constant \( C \), \( a \leq C b \). We also use the following notation throughout the paper.

\[
S_n := n^{-1/2} (X_1 + X_2 + \ldots + X_n),
\]
\[
U_{n,k} := n^{-1/2} (X_1 + \ldots + X_k + Y_{k+1} + \ldots + Y_n) \quad \text{for} \quad 0 \leq k \leq n.
\]

Note that \( U_{n,n} = S_n \) and hence proving the closeness (in distribution) of \( U_{n,k} \) and \( U_{n,0} \) for all \( k \) ensures closeness of \( S_n \) and \( U_{n,0} \). In this regard, define for \( m \geq 0 \)

\[
\delta_{n,m} := \sup_{r \geq 0} \max_{1 \leq k \leq n} r^m |\mathbb{P}(\|U_{n,k}\| \leq r) - \mathbb{P}(\|U_{n,0}\| \leq r)|. \tag{7}
\]

Also define for \( 1 \leq i \leq n \) the signed measure \( \zeta_i \) by

\[
\zeta_i(A) := \mathbb{P}(X_i \in A) - \mathbb{P}(Y_i \in A) \quad \text{for} \quad A \subseteq \mathbb{R}^p.
\]
Based on this signed measure, $|\zeta_i|$ denotes the variation of measure $\zeta_i$. It is clear that

$$\int d\zeta_i(x) = \int x(j)d\zeta_i(x) = \int x(j)x(k)d\zeta_i(x) = 0 \quad \text{for} \quad 1 \leq i \leq n, 1 \leq j, k \leq p.$$ (8)

Define the “weak” third pseudo-moment as

$$L_n := \frac{1}{n} \sum_{i=1}^{n} \max_{1 \leq j \leq p} \int |x(j)|^3 |\zeta_i|(dx),$$ (9)

and the truncated “strong” second pseudo-moment as

$$M_n(\phi) := \frac{1}{n} \sum_{i=1}^{n} \int \|x\|^2 1\{\|x\| \geq n^{1/2}\phi/\log(ep)} |\zeta_i|(dx) \quad \text{for} \quad \phi > 0.$$ (10)

These are called pseudo-moments since they are defined with respect to the variation measure and becomes zero if the distributions of $X_i$’s and $Y_i$’s are the same. Most of the results in classical multivariate setting (of Sazonov and Ulyanov (1982)) and in Banach spaces (of Paulauskas and Račkauskas (1991)) are derived in terms of pseudo-moments. We will present our results also in terms of the pseudo-moments defined above.

Quantity $M_n(\phi)$ defined above is similar to the one defined in Chernozhukov et al. (2017b) except that they have $\|x\|^3$ instead of $\|x\|^2$. This subtlety allows us to derive better rates when random vectors only have $(2 + \tau)$-moments.

Further, set $\mu_i := \text{median}(\|Y_i\|)$ and $\sigma_i^2 := \max_{1 \leq j \leq p} \text{Var}(Y_i(j))$ for $1 \leq i \leq n$. Define the weighted “weak” third moment as

$$\bar{L}_{n,0} := \frac{1}{n} \sum_{i=1}^{n} (\mu_i + \sigma_i) \max_{1 \leq j \leq p} \int |x(j)|^3 |\zeta_i|(dx),$$ (11)

and for $m > 0$,

$$\bar{L}_{n,m} := \frac{1}{n} \sum_{i=1}^{n} \left[ \mu_i^{m+1} + \sigma_i^{m+1}(m+1)/(me)^{(m+1)/2} \right] \max_{1 \leq j \leq p} \int |x(j)|^3 |\zeta_i|(dx).$$

It is clear that $\bar{L}_{n,m}$ is at most of order $(\sqrt{\log(ep)})^{m+1}$ for $m \geq 0$ (assuming $\sigma_i$’s are all of order 1). More precisely, we have

$$\bar{L}_{n,m} \leq \Theta_m L_n \left( \max_{1 \leq i \leq n} \sigma_i \right)^{m+1} (\log(ep))^{(m+1)/2}, \quad \text{for all} \quad m \geq 0.$$ (12)

Here $\Theta_m$ is a constant depending only on $m \geq 0$ scaling like $m^{m/2}$. 
2.2. Tail of Gaussian Processes and Anti-concentration Inequality

In this section, we prove various inequalities regarding the distribution function of the maximum of a Gaussian process on a compact metric space. These inequalities will lead to the sharper versions of anti-concentration inequalities and is crucial for the large deviation result derived later. The following result is a refinement of Lemma 3.1 in Bentkus and Račkauskas (1990) and make use of a result from Giné (1976).

**Theorem 2.1.** Let $Y$ be a sample continuous centered Gaussian process on a compact metric space $S$ such that $\sigma^2_{\min} \leq \mathbb{E}[Y^2(s)] \leq \sigma^2_{\max}$ for all $s \in S$. Let $\mu$ denote the median of $\|Y\| := \sup_{s \in S} |Y(s)|$. Then the following are true:

1. For all $r \geq 0$,
   \[
   \mathbb{P}(\|Y\| > r) \geq \frac{1}{6} \exp \left( -\frac{r^2}{\sigma^2_{\max}} \right).
   \]

2. For all $r, \varepsilon \geq 0$,
   \[
   \mathbb{P}(\|Y\| > r - \varepsilon) \leq 20 \exp \left( \Phi_4(r + 1) \varepsilon \right) \mathbb{P}(\|Y\| > r),
   \]
   where $\Phi_4$ is given by
   \[
   \Phi_4 := 1 + \frac{56(\mu + 1.5\sigma_{\max})(\mu + 4.1\sigma_{\max})}{\sigma^2_{\max}\sigma^2_{\min}} + \frac{32\pi(2.6\sigma_{\min} + \mu)^2}{\sigma^2_{\min}} \left( \sigma^2_{\min} + 32\sigma_{\min}\mu + 12\mu^2 \right).\]

3. For all $r \geq 0$ and $\varepsilon > 0$,
   \[
   \mathbb{P}(r - \varepsilon \leq \|Y\| \leq r + \varepsilon) \leq \Phi_2(\varepsilon) (r + 1) \mathbb{P}(\|Y\| > r - \varepsilon),
   \]
   where $\Phi_2$ is given by
   \[
   \Phi_2 := \max \left\{ \frac{51(\mu + 4.1\sigma_{\max})}{\sigma^2_{\min}}, \frac{32\pi(\mu + 2.6\sigma_{\min})^2}{\sigma^4_{\min}} \right\}
   \]

Observe that the set $\{1, \ldots, p\}$ is compact and discrete. Hence the result can be used in the high dimension case. Theorem 2.1 is dimension-free and the dependence on the “complexity” of $S$ appears only through the median of $\|Y\|$. The following anti-concentration inequalities for $\|Y\|$ can be derived as immediate corollaries of Theorem 2.1.

**Theorem 2.2.** Fix $m \geq 0$. Under the assumptions of Theorem 2.1, we have for all $\varepsilon \geq 0$

\[
\sup_{r \geq 0} r^m \mathbb{P}(r - \varepsilon \leq \|Y\| \leq r + \varepsilon) \leq \Phi_{AC,m}(\varepsilon),
\]

where $\Phi_{AC,m} := \Theta_m \sigma^{-4}_{\min}((\mu + \sigma_{\max})^{m+1}\sigma^2_{\min} + (1 + \sigma_{\max})^2\sigma^2_{\max}^{m+2})$, with $\Theta_m$ representing a constant that depends only on $m$. 


The constant $\Phi_{AC,m}$ obtained in Theorem 2.2, for any $m \geq 0$, is expected to be optimal since if the coordinates of $Y$ are independent then the discussion following Corollary 2.7 of Giné (1976) and Example 2 of Chernozhukov et al. (2013) imply that the density of $\|Y\|$ at points of order $\sqrt{\log \rho}$ is lower bounded in rate by $\sqrt{\log \rho}$. Hence in this case for any $m \geq 0$ as $\varepsilon \rightarrow 0$ the rate is lower bounded by $\mu^{m+1}$. Lemma B.2 of Lopes et al. (2018) provides optimal lower bounds of order $\sqrt{\log \rho}$ on the median of $\|Y\|$ when the coordinates of $Y$ are mostly negatively correlated with correlations bounded away from 1.

**Remark 2.1** (Comparison with Chernozhukov et al. (2015)) Theorem 3 of Chernozhukov et al. (2015) implies the anti-concentration for $\|Y\|$ and provides a dimension-free bound depending on the mean of $\|Y\|$ only under the additional assumption of $\sigma_{\max} = \sigma_{\min}$; Lemma 3.1 of Ledoux and Talagrand (2011) implies that the mean and median of $\|Y\|$ are of the same order. For the general case, their bound has an additional term of $\log(1/\varepsilon)$ which makes their bound weaker than that in Theorem 2.2. In terms of the proof technique, we note that the techniques of both works are the same for $r \leq 3(\mu + \sigma_{\max})$. For the case $r \geq 3(\mu + \sigma_{\max})$, Chernozhukov et al. (2015) apply the Gaussian concentration inequality that leads to an extra $\log(1/\varepsilon)$ factor while we use inequality (13) that leads to the sharper version. Theorem 2.2 is the first result on dimension-free anti-concentration inequality for $\|Y\|$ and hence our result readily applies to Gaussian processes on (infinite dimensional) compact metric spaces. The results of Nazarov (2003) imply an anti-concentration inequality that explicitly depends on the dimension as $\sqrt{\log(ep)}$. Nazarov’s result as proved in Chernozhukov et al. (2017a) cannot lead to a rate better than $\sqrt{\log(ep)}$ since the covariance structure of $Y$ is completely ignored in the proof. ⋆

### 2.3. Smooth Approximation of the Maximum

The starting point of almost any Berry-Esseen type result is a smoothing inequality. In our scenario, we need a smooth approximation of $\mathbb{1}\{\|S_n\| \leq r\}$ where, recall, $\| \cdot \|$ denotes the $l_\infty$-norm. Theorem 1 of Bentkus (1990) provides an infinitely differentiable (in $S_n$) approximation of this indicator with sharp bounds on the derivatives. However to get better rates of convergence in the central limit theorem a certain stability property of the derivatives is needed. This property is exactly the reason why Chernozhukov et al. (2017b) were able to get better rates than those implied by Banach space CLTs; see Banerjee et al. (2018) for implications of Banach space CLTs. The smooth approximation (taken from Chernozhukov et al. (2013)), along with its properties, is summarized in the following result. Define the “softmax” function as $F_\beta(z) := \beta^{-1} \log \sum_{j=1}^{2p} \exp(\beta z(j))$ for $z \in \mathbb{R}^{2p}$. Also define $g_0(t) := 30 \mathbb{1}\{0 \leq t \leq 1\} \int_1^t s^2(1-s)^2 ds$. 

$$\text{g}_0(t) := 30 \mathbb{1}\{0 \leq t \leq 1\} \int_1^t s^2(1-s)^2 ds.$$
Lemma 2.1. Fix $r \geq 0, \varepsilon > 0$ and set $\beta = 2 \log(2p)/\varepsilon$. Define the function $\varphi : \mathbb{R}^p \to \mathbb{R}$ as

$$
\varphi(x) = \varphi_{r,\varepsilon}(x) = g_0 \left( \frac{2(F_\beta(z_x - r\mathbf{1}_{2p}) - \varepsilon/2)}{\varepsilon} \right),
$$

where $z_x = (x^\top : -x^\top)^\top$ and $\mathbf{1}_{2p}$ is the vector of 1’s of dimension $2p$. This function $\varphi(\cdot)$ satisfies the following properties.

1. It “approximates” the indicator of the $l_\infty$-ball, that is,

$$
\varphi(x) = \begin{cases} 
1, & \text{if } \|x\| \leq r, \\
0, & \text{if } \|x\| > r + \varepsilon,
\end{cases}
$$
or equivalently $\mathbb{1}\{\|x\| \leq r\} \leq \varphi(x) \leq \mathbb{1}\{\|x\| \leq r + \varepsilon\}$.

2. There exists functions $D_j(\cdot), D_{jk}(\cdot)$ and $D_{jkl}(\cdot)$ for $1 \leq j, k, l \leq p$ as well as constant $C_0 > 0$ such that $|\partial_j \varphi(x)| \leq D_j(x), |\partial_{jk} \varphi(x)| \leq D_{jk}(x), |\partial_{jkl} \varphi(x)| \leq D_{jkl}(x)$, where $\partial_j, \partial_{jk}, \partial_{jkl}$ denote the partial derivatives of $\varphi$ with respect to the indices in subscript and for all $x \in \mathbb{R}^p$,

$$
\sum_{j=1}^p D_j(x) \leq C_0 \varepsilon^{-1}, \quad \sum_{j,k=1}^p D_{jk}(x) \leq C_0 \log(ep) \varepsilon^{-2}, \quad \sum_{j,k,l=1}^p D_{jkl}(x) \leq C_0 \log^2(ep) \varepsilon^{-3}.
$$

3. The functions $D_j, D_{jk}, D_{jkl}$ also satisfy a ratio stability property: there exists universal constant $C > 0$ such that for all $x, w \in \mathbb{R}^p$,

$$
e^{-C \log(ep)\|w\|_\infty/\varepsilon} \leq \frac{D_j(x+w)}{D_j(x)}, \quad \frac{D_{jk}(x+w)}{D_{jk}(x)}, \quad \frac{D_{jkl}(x+w)}{D_{jkl}(x)} \leq e^{-C \log(ep)\|w\|_\infty/\varepsilon}. \tag{14}
$$

The smooth approximation result above is the bottleneck in obtaining better dependence on $\log(ep)$ in CLTs. The $\log^4(ep)$ dependence in the uniform and non-uniform CLTs presented in Subsection 1.1 comes only from the $\log(ep)$ factors in the bounds of derivatives and stability property of smooth approximation.

The approximating functions in Lemma 2.1 are constructed to work for any high-dimensional distribution; the difference between $F_\beta(z)$ and $\max_j z(j)$ is at most $\log(2p)/\beta$ over all vectors $z$ and this can be smaller if $z$ comes from a specific distribution. This universality can be seen clearly in the construction of Bentkus (1990) who defines the $\varepsilon$ approximation of $\|\cdot\|$ based on $f_\varepsilon(x) = \mathbb{E}[\|x + \varepsilon \eta\|], \text{ where } \eta \sim N_p(0, I_p)$. However one can replace $\eta$ by other random vectors taking into account the dependence of $U_{n,0}$. A specific choice that we conjecture works is $f_\varepsilon(x) = \mathbb{E}[\|x + \varepsilon U_{n,0}\|]$. Since $\|\cdot\|$ is Lipschitz, we get that $|\|x\| - f_\varepsilon(x)| \leq \varepsilon \mathbb{E}[\|U_{n,0}\|] = \varepsilon O(\mu)$. Since $U_{n,0}$ and $S_n$ share the same dependence structure, we only need to bound the derivatives of $f_\varepsilon$ at $U_{n,j}$ which would lead to better rates in CLT using the proofs here; see Section 6 for details.
3. Main Results

We are now ready to state the main results of this paper. The proofs of all the results in this section are given in the supplementary material. The main steps of these proofs are presented, for readers’ convenience, in Section 6. Recall the notation $\delta_{n,m}$ from (7) in Section 2.1. Also recall that $M_n(\cdot)$ and the quantity $\bar{L}_{n,0}$ are defined in (10) and (11), respectively. The quantity $L_n$ (in (9)) denotes the “weak” third pseudo-moment and if $L_n = 0$ then $\delta_{n,0} = \delta_{n,m} = 0$ for any $m \geq 0$. For this reason, we assume $L_n > 0$. Let $\Phi_{AC,m}$ denote the anti-concentration constant in Theorem 2.2 for random vector $U_{n,0}$.

3.1. Uniform CLT

**Theorem 3.1.** For independent random vectors $X_1, \ldots, X_n \in \mathbb{R}^p$, we have

$$\delta_{n,0} \leq 4\Phi_{AC,0}\varepsilon_n + \frac{2C_0 \log(ep)M_n(\varepsilon_n)}{\varepsilon_n^2} + \frac{\log^{1/3}(ep)\bar{L}_{n,0}}{n^{1/3}L_n^{4/3}(2e5\varepsilon C_0)^{1/3}},$$

where $\varepsilon_n := (2e2^5C_0 \log^2(ep)L_n)^{1/3}/n^{1/6}$.

Theorem 3.1 is qualitatively the same as Theorem 2.1 of Chernozhukov et al. (2017b) for $l_\infty$-balls. More importantly, note that if $\Phi_{AC,0} \asymp \sqrt{\log(ep)}$ then Theorem 3.1 has a dominating term of order $(\log(ep)/n)^{1/6}$ and hence the result above is as good as Theorem 2.1 of Chernozhukov et al. (2017b). The last term in the bound of $\delta_{n,0}$ is of lower order compared to the first term. From inequality (12), we obtain that

$$\frac{\log^{1/3}(ep)\bar{L}_{n,0}}{n^{1/3}L_n^{4/3}(2e5\varepsilon C_0)^{1/3}} = O \left( \frac{(\log(ep))^{5/6}}{(nL_n)^{1/3}} \right) = O \left( \frac{(\log(ep))^{5/2}}{nL_n} \right)^{1/3},$$

which is dominated by the first term which is at least of order $(\log^4(ep)/n)^{1/6}$. Further the quantity $M_n(\varepsilon_n)$ (defined in (10)) is exactly what appears in the classic Lindeberg condition.

**Rates under $(2 + \tau)$-moments:** Recall that $M_n(\phi)$ only involves truncated second moment instead of third moment used in Chernozhukov et al. (2017b). This subtle difference allows for deriving better rates when $\|X_i\|$ have $(2 + \tau)$-moments. In this case the choice of $\varepsilon_n$ in Theorem 3.1 is not large enough since $M_n(\varepsilon_n)/\varepsilon_n^2$ does not converge to zero. Even though our results involve third “moments” $L_n, \bar{L}_{n,0}$, from the proof (in particular (22) in the supplement) it can be seen that the integral in the definition of $L_n, \bar{L}_{n,0}$ can be changed to integral over a truncated set; see Appendix A and Step E in Section 6 for details. Very recently Sun (2019) considered CLT under $(2 + \tau)$-moments using Lindeberg method but no explicit rates were provided. In the following we provide details for $\tau \geq 1$ and the
calculations for \( \tau < 1 \) are provided in Appendix A. Set
\[
\nu_{2+\tau} := \frac{1}{n} \sum_{i=1}^{n} \int \|x\|^{2+\tau} |\zeta_i| (dx) \quad \Rightarrow \quad M_n(\varepsilon) \leq \frac{\nu_{2+\tau}}{(n^{1/2} \varepsilon / \log(ep))^{\tau}} = \nu_{2+\tau} \left( \frac{\log(ep)}{n^{1/2} \varepsilon} \right)^{\tau}.
\]

The proof of Theorem 3.1 actually proves a bound that holds for all \( \varepsilon > 0 \) and we now choose \( \varepsilon = \varepsilon_n := \max\{2e^{2\varepsilon} C_0 \log^2(ep) L_n / n^{1/6}, r_n \nu_{2+\tau} ((\log(ep))^{1+\tau} / n^{\tau/2})^{1/(2+\tau)} \} \) for some \( r_n \geq 1 \). This choice implies that
\[
\frac{e^{2\varepsilon} C_0 \log^2(ep) L_n}{n^{1/2} \varepsilon_n^2} \leq \frac{1}{2} \quad \text{and} \quad \frac{C_0 \log(ep) M_n(\varepsilon_n)}{\varepsilon_n^2} \leq \frac{C_0}{r_n^{2+\tau}}.
\]

Following the proof of Theorem 3.1, for any \( r_n \geq 1 \), we get
\[
\delta_{n,0} \lesssim \frac{1}{r_n^{2+\tau}} + \Phi_{AC,0} \frac{\left( \log(ep) / n^{1/6} \right)^{1/3}}{\nu_{2+\tau} (\log(ep))^{(1+\tau)/(2+\tau)}} + \Phi_{AC,0} \frac{r_n \nu_{2+\tau} (\log(ep))^{(1+\tau)/(2+\tau)}}{n^{\tau/2(2+\tau)}}.
\]

Here \( a \land b = \min\{a, b\} \) and \( a \lor b = \max\{a, b\} \). Minimizing over \( r_n \geq 1 \), we get
\[
\delta_{n,0} \lesssim \begin{cases} \left[ \Phi_{AC,0} L_n^{1/(2+\tau)} + \Phi_{AC,0} \nu_{2+\tau} \right] (2+\tau)/(3+\tau) \left( \log(ep) \right)^{(\tau+1)/(\tau+2)} n^{\tau/(6+2\tau)}, & \text{if } \tau < 1, \\ \Phi_{AC,0} \left( \log^2(ep) L_n / n^{\tau/6} \right)^{1/3} + \Phi_{AC,0} \nu_{2+\tau} (3+\tau) (\log(ep))^{(\tau+1)/(\tau+3)} n^{\tau/(6+2\tau)}, & \text{if } \tau \geq 1, \end{cases}
\]

where \( L_{n,\tau} = n^{-1} \sum_{i=1}^{n} \max_{1 \leq j \leq p} \int |x(j)|^{2+\tau} |\zeta_i| (dx) \). As mentioned before the proof for \( \tau < 1 \) is given in Appendix A. The dependence on sample size obtained above is better than the one obtained in Proposition 2.1 of Chernozhukov et al. (2017b) for all \( \tau > 0 \). In particular for \( \tau = 1 \), we got the dependence \( n^{-1/8} \) whereas Chernozhukov et al. (2017b) obtained \( n^{-1/9} \) on the sample size. We note here that following the proof of Theorem 2.1 of Bentkus et al. (2000) it is possible to get \( n^{-1/6} \) dependence on \( n \) whenever \( \tau \geq 1 \) as shown in the following proposition. This is similar to the case of multivariate CLT where the rate of convergence scales as \( n^{-1/2} \) whenever the random vectors have more than three moments; i.e., after some number of moments the rate stabilizes. The following result is proved based on Theorem 2.1 of Bentkus et al. (2000) and does not use the stability property (14).

**Proposition 3.1.** For i.i.d. random vectors \( X_1, \ldots, X_n \),
\[
\delta_{n,0} \leq 2^{1-n} + 8 \Phi_{AC,0} \nu_3 (C_0 \log^2(ep))^{1/3} n^{-1/6}.
\]

It is clear that the rate obtained in Proposition 3.1 is sharper than the one obtained in (15) for \( 1 \leq \tau \leq 1.5 \). Note that for \( \tau = 1.5, \tau/(6+2\tau) = 1/6 \) and (15) leads to \( n^{-1/6} \) rate. A notable difference between the rates is that while Theorem 3.1 decouples (sub-optimally) the main rate \( \Phi_{AC,0} (\log^2(ep))^{1/3} / n^{1/6} \) and the term depending on moments \( \nu_{2+\tau} \), Proposition 3.1 puts these two terms together which implies that Proposition 3.1
High-dimensional CLT
does not lead to the “right” requirement between \( \log(p) \) and \( n \) in case random vectors
have exponential tails. This discussion raises an important point: What is the “right way”
of proving rates in the high-dimensional CLT? The following result combines the proof
techniques of Theorem 3.1 and Proposition 3.1 in the “right way” so that the rate scales
like \( n^{-1/6} \) for all \( \tau \geq 1 \) (as Proposition 3.1) and becomes the same as (15) for \( \tau \to \infty \).

**Theorem 3.2.** For i.i.d. random vectors \( X_1, \ldots, X_n \), and for any \( \tau \geq 1 \),
\[
\delta_{n,0} \lesssim \frac{1}{2^m} + \Phi_{AC,0} \left( \frac{L_n^2 \log^4(ep)}{n} \right)^{1/6} + \Phi_{AC,0} \nu_{2+\tau} \frac{(\log(ep))^{(\tau+1)/(\tau+2)}}{n^{\tau/(4+2\tau)}},
\]
where \( \lesssim \) hides only universal constants.

The assumption of i.i.d. random vectors in Proposition 3.1 and Theorem 3.2 can be
relaxed to observations \( X_1, \ldots, X_n \) having the same covariance matrix. It is clear that
Theorem 3.2 leads to better rates than (15) for all \( \tau \geq 1 \) since \( \tau/(6+2\tau) < \tau/(4+2\tau) \).

### 3.2. Non-uniform CLT

As an extension of Theorem 3.1, we have the following non-uniform version of CLT. Since
the statement is exact with explicit constants, it is cumbersome.

**Theorem 3.3.** For independent \( \mathbb{R}^p \) random vectors \( X_1, \ldots, X_n \), for any \( m, r_{n,m} \in (0, \infty) \),
\[
\delta_{n,m} \leq 2^{2m^2/3+8m/3+1} \varepsilon_n^{m+1} \Phi_{AC,0} + (2^{2m/3+1} + 2) \Phi_{AC,m} \varepsilon_n + 2C_0 \log(ep) r_{n,m} \varepsilon_n^{-2} M_n (2^{2m/3} \varepsilon_n) + \frac{r_{n,m} \mathcal{E}_{n,m}}{L_n 2^m e^C} \left( \frac{\log(ep)}{n(2^{2m+1} e^2 C_0 L_n)} \right)^{(m+1)/3} + \sup_{r \geq r_{n,m}} \max_{0 \leq k \leq n} r_m \Pr(||U_{n,k}|| \geq r).
\]
Here \( \varepsilon_n \) is same as the one defined in Theorem 3.1.

Theorem 3.3 does not get the correct dependence on the sample size \( n \) when random
vectors \( X_1, \ldots, X_n \) only have \( 2+\tau \)-moments as the calculation for Theorem 3.1 show. We
now present an improvement for i.i.d. random vectors which can be seen as a non-uniform
extension of Theorem 3.2.

**Theorem 3.4.** Fix \( m \geq 1 \) and \( \tau \geq m \). If \( X_1, \ldots, X_n \), are i.i.d. random vectors, then
\[
\delta_{n,m} \leq 2^{m/2} \left( \nu_m 2^{-n/2} \right)^m + 2^{2+2m} \varepsilon_n^m + 2.4 \Phi_{AC,m} \varepsilon_n,
\]
where
\[
\varepsilon_n := \max \left\{ \frac{(2^{2+3m/2} C_0 e^C L_n \log^2(ep))^{1/3}}{n^{1/6}}, \nu_{2+\tau} \left( \frac{2^{3+5m/2} C_0 \log^{\tau+1}(ep)}{n^{\tau/(4+2\tau)}} \right)^{1/(2+\tau)} \right\}.
\]

Theorem 3.4 always leads to \( n^{-1/6} \) dependence on the sample size unlike Theorem 3.3.
Comments on the Proof Technique. The proofs of uniform and non-uniform CLTs in Banach space literature are based on Lindeberg method and smooth approximation; see, for example, Paulauskas and Račkauskas (1991) and Bentkus et al. (2000). Motivated by the proof technique of Chernozhukov et al. (2017b) who introduced the stability property, we combine Lindeberg method with a minor twist from the proof of Chernozhukov et al. (2017b) to prove all the above results; see Section 6 for a detailed outline of how our proofs differ from the ones in the above mentioned references.

Recently, Koike (2019a, Proposition 2.3) proved for sub-Gaussian random variables a rate of $(\log^6(ep)/n)^{1/6}$ for $\delta_{n,0}$ which could also be weaker than our result depending on $\Phi_{AC,0}$. It maybe possible that the methods in that paper could sharpen our result. We refrained from using their proof techniques for simplicity.

3.3. Corollaries for Sub-Weibull Random Vectors

In this section, we provide simplified results under the assumption that the coordinates of the random vectors $X_1, \ldots, X_n$ are sub-Weibull. For a simplification of uniform CLTs, we choose a suitable $\tau$ in (15) and Theorem 3.2. For a simplification of Theorem 3.3, we simplify $M_n(\varepsilon_n)$ and $\bar{L}_{n,m}$ under sub-Weibull tails. Similar simplification can be done from Theorem 3.1 but this leads to sub-optimal dependence on $\log(ep)$ in the second order term.

Recall that $M_n(\varepsilon_n)$ and $\bar{L}_{n,m}$ are defined in terms of the variation measures $\zeta_j$ which we bound using the sum of the measures for simplicity. We now define Orlicz norms.

**Definition 3.1.** Let $X$ be a real-valued random variable and $\psi : [0, \infty) \mapsto [0, \infty)$ be a non-decreasing function with $\psi(0) = 0$. Then, we define

$$\|X\|_\psi = \inf \{ c > 0 : \mathbb{E} \psi(|X|/c) \leq 1 \},$$

where the infimum over the empty set is taken to be $\infty$.

Usually the definition of Orlicz “norm”, $\| \cdot \|_\psi$, includes convexity assumption of $\psi$ which we did not include since we also work with non-convex $\psi$ below. It follows from Jensen’s inequality, that when $\psi$ is a non-decreasing, convex function, $\| \cdot \|_\psi$ is a norm on the set of random variables $X$ for which $\|X\|_\psi < \infty$. The commonly used Orlicz norms are derived from $\psi_\alpha(x) := \exp(x^\alpha) - 1$, for $\alpha \geq 1$, which are obviously increasing and convex. For $0 < \alpha < 1$, $\psi_\alpha$ is not convex, and $\|X\|_{\psi_\alpha}$ is not a norm, but a quasinorm. A random variable $X$ is called sub-exponential if $\|X\|_{\psi_1} < \infty$, and a random variable $X$ is called sub-Gaussian if $\|X\|_{\psi_2} < \infty$.

Recall that the $j$-coordinate of $X$ is $X(j)$ and set $\sigma_{\max}^2 := \max_{1 \leq j \leq p} \operatorname{Var}(U_{n,0}(j))$, and $\sigma_{\min}^2 := \min_{1 \leq j \leq p} \operatorname{Var}(U_{n,0}(j))$. Throughout the following corollaries, $\Theta$ stands for a univer-
Corollary 3.1 (Uniform CLT). Suppose there exists a constant $1 \leq K_p < \infty$ such that
\begin{equation}
\max_{1 \leq i \leq n} \max_{1 \leq j \leq p} \|X_i(j)\|_{\psi_{\alpha}} \leq K_p,
\end{equation}
for some $0 < \alpha \leq 2$. If \( \log(n^{3/2}/(K_p\Theta_n\Phi_{AC,0}(\log(ep))^{2+1/\alpha})) \geq 4 \) then
\begin{equation}
\delta_{n,0} \leq \Theta_n\Phi_{AC,0}(L_n\log^2(ep))^{1/3} + \Theta_n K_p\Phi_{AC,0}(\log(ep))^{1+1/\alpha}n^{-1/2} \log^{1/\alpha} n \left( \frac{\log(ep)}{n} \right).
\end{equation}
Instead if \( \alpha \log(n/\log(ep)) \geq 3 \) and \( X_1, \ldots, X_n \) share the same covariance matrix, then
\begin{equation}
\delta_{n,0} \leq \Theta_n\Phi_{AC,0}(L_n\log^2(ep))^{1/3} + \Theta_n K_p\Phi_{AC,0}(\log(ep))^{1+1/\alpha}n^{-1/2} \log^{1/\alpha} n \left( \frac{\log(ep)}{n} \right).
\end{equation}

Proof. Note that assumption (16) implies that \( \nu_{2+\tau} \leq K_p\Theta_n(2+\tau)^{1/\alpha}(\log(ep))^{1/\alpha} \) for all \( \tau \geq 1 \). Then (17) follows from (15) by taking \( \tau + 3 = \log(n^{3/2}/(K_p\Theta_n\Phi_{AC,0}(\log(ep))^{2+1/\alpha})) \) which is possible for \( \tau \geq 1 \) since the right hand side is assumed to be larger than 4. Further (18) follows from Theorem 3.2 by taking \( \tau + 2 = \alpha \log(n/\log(ep)) \) (which is possible for \( \tau \geq 1 \) since \( \alpha \log(n/\log(ep)) \geq 3 \)).

Simplifying a little further, the bounds on \( \delta_{n,0} \) can be written as
\begin{equation}
\delta_{n,0} \leq \Theta_n\Phi_{AC,0}(L_n\log^2(ep))^{1/3} + K_p\Phi_{AC,0}(\log(ep))^{1+1/\alpha}n^{-1/2} \log^{1/\alpha} n \left( \frac{\log(ep)}{n} \right).
\end{equation}

The following corollary is obtained by controlling \( M_n(\varepsilon_n) \) and \( \tilde{L}_{n,m} \) in Theorem 3.3 for sub-Weibull random vectors. By choosing an appropriate of \( \tau \) in Theorem 3.4 we get a much simpler form under the assumption that \( X_1, \ldots, X_n \) share the same covariance structure.

Corollary 3.2 (Non-uniform CLT). Fix \( m \geq 0 \). If \( n \geq \Theta K_p^3 L_n^{-1}(2e \log(ep))^{1+3/\alpha} \), then we have (i) for \( 1 < \alpha \leq 2 \),
\begin{equation}
\delta_{n,m} \leq \Theta_m\Phi_{AC,m}(L_n^2\log^4(ep))^{1/6} + \Theta_m K_p^m\sigma_{\max}^{m+1}L_n \left( \frac{\log(ep)}{n} \right)^{(m+1)/3} + \Theta_m K_p^{m+2}L_n^{-2/3} \sigma_{\max}^{2/3},
\end{equation}

and (ii) for \( 0 < \alpha \leq 1 \),
\begin{equation}
\delta_{n,m} \leq \Theta_m\Phi_{AC,m}(L_n^2\log^4(ep))^{1/6} + \Theta_m K_p^m\sigma_{\max}^{m+1}L_n \left( \frac{\log(ep)}{n} \right)^{(m+1)/3} + \Theta_m K_p^{m+2}L_n^{-2/3} \sigma_{\max}^{2/3}.
\end{equation}
The implication from (19) can be simply written as
\[ \delta_{n,m} \leq 2^{m/2} \left( \frac{\nu_m}{2^{m/2}} \right)^m + 2^{2+2m} \varepsilon_n^m + 2.4 \Phi_{AC,m} \varepsilon_n, \]
where
\[ \varepsilon_n := \max \left\{ \frac{(2^{2+3m/2} C_0 e L_n \log^2(ep))^{1/3}}{n^{1/6}}, \Theta_{\alpha,m} \frac{(\log(ep))^{1+1/\alpha}}{n^{1/2}} \log^{1/\alpha} \left( \frac{2 C_0 n}{\log(ep)} \right) \right\}. \]

Inequality (19) follows by taking \( 2 + \tau = \alpha \log(2^{3+5m/2} C_0 n/\log(ep)) \) in \( \varepsilon_n \) of Theorem 3.4.

Remark 3.1 (Even more simplified rates) The bounds in Corollary 3.2 are finite sample and show explicit dependence on \( L_n, K_p, \sigma_{\text{max}} \) and other distributional constants. If \( \max\{K_p, \sigma_{\text{max}}, L_n\} \equiv O(1) \) then the bounds in Corollary 3.2 can simply be written as
\[ \delta_{n,m} \lesssim \Phi_{AC,m} \frac{(L_n \log^2(ep))^{1/3}}{n^{1/6}} + \begin{cases} 0, & \text{if } \alpha > 1, \\ \left( (\log(ep))^{5/4+3/\alpha} / n^{(m+1)/3} \right), & \text{if } \alpha \in (0, 1]. \end{cases} \]

The implication from (19) can be simply written as
\[ \delta_{n,m} \lesssim \left( \frac{\nu_m}{2^{m/2}} \right)^m + \Phi_{AC,m} \frac{(L_n \log^2(ep))^{1/3}}{n^{1/6}} + \Phi_{AC,m} \frac{(\log(ep))^{1+1/\alpha}}{n^{1/2}} \log^{1/\alpha} \left( \frac{2 C_0 n}{\log(ep)} \right). \]

Remark 3.2 (Convergence of Moments) Theorems 3.1 and 3.3 are useful in proving convergence of \( m \)-th moment of \( \|S_n\| \) to that of \( \|U_{n,0}\| \) at an \( n^{-1/6} \) rate (up to factors depending on \( \log(ep) \)). To prove an explicit bound, note that for \( m \geq 1 \):
\[ |E[\|S_n\|^m] - E[\|U_{n,0}\|^m]| = \int_0^\infty m r^{m-1} \left( P(\|S_n\| \geq r) - P(\|U_{n,0}\| \geq r) \right) dr \leq m \int_1^\infty \delta_{n,0} dr + \int_1^\infty \frac{m}{r^{1+\beta}} \delta_{n,m+\beta} dr \leq m \delta_{n,0} + \frac{m}{\beta} \delta_{n,m+\beta}, \]
for any \( \beta \geq 0 \). The second term in the right hand side above can be bounded using Theorem 3.3 or 3.4. Taking the main terms in the bounds for \( \delta_{n,0}, \delta_{n,m+\beta} \) and using \( \Phi_{AC,m} \lesssim \mu^{m+1} \) from Theorem 2.2, we can write the bound on moment difference as
\[ |E[\|S_n\|^m] - E[\|U_{n,0}\|^m]| \lesssim m \mu \frac{(L_n \log^2(ep))^{1/3}}{n^{1/6}} + \frac{m}{\beta} \mu^{m+1} \frac{(L_n \log^2(ep))^{1/3}}{n^{1/6}}. \]

Note that we assumed here \( \sigma_{\min}^{-1} = O(1) \). Choosing \( \beta = 1 / \log(\mu) \), we get
\[ |E[\|S_n\|^m] - E[\|U_{n,0}\|^m]| \lesssim m \mu \frac{(L_n \log^2(ep))^{1/3}}{n^{1/6}} + m \log(\mu) \mu^{m+1} \frac{(L_n \log^2(ep))^{1/3}}{n^{1/6}}. \]
Since the median and $(\mathbb{E}[\|U_{n,0}\|^\ell])^{1/\ell}, \ell \geq 1$ are of the same order, we get that for $m \geq 1$
\[
\mathbb{E}[\|S_n\|^m] = \left(1 + O\left(\frac{\mu \log(\mu)(L_n \log^2(ep))^{1/3}}{n^{1/6}}\right)\right) \mathbb{E}[\|U_{n,0}\|^m].
\]
Hence our results imply that the moments of $\|S_n\|$ match the moments of $\|U_{n,0}\|$ up to a lower order term if $\mu \log(\mu)(L_n \log^2(ep))^{1/3} = o(n^{1/6})$.

4. Applications

In this section, we present two applications. The first is to post-selection inference that shows the impact of dimension-free anti-concentration constant. The second is to maximal inequalities for empirical processes that shows the importance of non-uniform CLT.

4.1. Many Approximate Means and Post-selection Inference

In this section, we use the “many approximate means” (MAM) framework of Belloni et al. (2018) for post-selection inference (PoSI). In this PoSI problem, we show scenarios where $\Phi_{AC,0}$ (or $\mu$) grows almost like a constant.

The MAM framework is as follows. Suppose we have a parameter $\theta_0 = (\theta_0(1), \ldots, \theta_0(p))^\top$ and an estimator $\hat{\theta} = (\hat{\theta}(1), \ldots, \hat{\theta}(p))^\top$ of parameter $\theta_0$ that has an approximate linear form:
\[
n^{1/2}(\hat{\theta} - \theta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi(Z_i) + R_n,
\] (20)
where $\psi(\cdot) = (\psi_1(\cdot), \ldots, \psi_p(\cdot))^\top \in \mathbb{R}^p$ and $R_n = (R_n(1), \ldots, R_n(p))^\top \in \mathbb{R}^p$. Here $Z_1, \ldots, Z_n$ are independent random variables based on which $\hat{\theta}$ is constructed. The function $\psi_j(\cdot), 1 \leq j \leq p$ represents the influence function for estimator $\hat{\theta}(j)$. An estimator $\hat{\theta}$ satisfying (20) is called asymptotically linear and most (of the commonly used) $M$-estimators satisfy this assumption. Based on the linear approximation (20) and the anti-concentration result Theorem 2.2, we have the following result. Define
\[
\tilde{\Delta}_n(r) := \left|\mathbb{P}(\sqrt{n} \|\hat{\theta} - \theta_0\| \leq r) - \mathbb{P}(\|Y^\psi\| \leq r)\right|
\]
where $Y^\psi \sim N_p(0, n^{-1} \sum_{i=1}^n \mathbb{E}[\psi(Z_i)\psi^\top(Z_i)])$. Let $\Phi_{AC,m}^\psi$ denote the anti-concentration constant from Theorem 2.2 for $Y^\psi$. Set $S_n^\psi := n^{-1/2} \sum_{i=1}^n \psi(Z_i)$.

Proposition 4.1. For any $\delta > 0$ and $m \geq 0$, we have
\[
r^m \tilde{\Delta}_n(r) \leq 2 \left[\left(\frac{5}{4}\right)^m \Delta_{n,m}^\psi + 5^m \Phi_{AC,0}^\psi \delta^{m+1}\right] + \Phi_{AC,m}^\psi \delta + r^m \mathbb{P}(\|R_n\| > \delta),
\]
where $\Delta_{n,m} := \sup_{r \geq 0} r^m |\mathbb{P}(\|S_n^\psi\| \leq r) - \mathbb{P}(\|Y^\psi\| \leq r)|$. 

High-dimensional CLT
This result is a slight generalization of Theorem 2.1 of Belloni et al. (2018) for \( l_{\infty} \)-balls. To derive a proper non-uniform CLT result from Proposition 4.1, one needs to choose \( \delta \) depending on \( r \) and apriori bound moments of \( \|R_n\| \). For the case \( m = 0 \) (which we focus on from now), the proof of Proposition 4.1 implies \( \Delta_n(r) \leq \Delta_{n,m}^\psi + \Phi_{AC,0}^\psi \delta + \mathbb{P}(\|R_n\| > \delta) \). Hence \( \Delta_n(r) \) converges to zero if \( \|R_n\| = o_p(r_n) \) such that \( r_n \Phi_{AC,0} = o(1) \) as \( n, p \to \infty \).

We now describe the framework of post-selection inference. Suppose \( Z_i := (X_i^\top, Y_i)^\top \in \mathbb{R}^d \times \mathbb{R} \) for \( 1 \leq i \leq n \) represent regression data with \( d \)-dimensional covariates \( X_i \). For any subset \( M \subseteq \{1, 2, \ldots, d\} \), let \( X_{i,M} \) denote a subvector of \( X_i \) with indices in \( M \). Based on a loss function \( \ell(\cdot, \cdot) \) define the regression “slope” estimator \( \hat{\beta}_M \) as

\[
\hat{\beta}_M := \arg\min_{\theta \in \mathbb{R}^{|M|}} \frac{1}{n} \sum_{i=1}^n \ell(X_{i,M}^\top \theta, Y_i).
\]

The target for the estimator \( \hat{\beta}_M \) is given by \( \beta_M := \arg\min_{\theta \in \mathbb{R}^{|M|}} n^{-1} \sum_{i=1}^n \mathbb{E}[\ell(X_{i,M}^\top \theta, Y_i)] \).

Some examples of loss functions are related to linear regression \( \ell(u, v) = (u - v)^2 / 2 \), logistic regression \( \ell(u, v) = uv - \log(1 + e^u) \), Poisson regression \( \ell(u, v) = uv - \exp(u) \).

As is often done in practical data analysis, suppose we choose a model \( \hat{M} \) based on the data \( \{Z_i : 1 \leq i \leq n\} \). The problem of post-selection inference refers to the statistical inference for the (random) target \( \beta_{\hat{M}} \). In particular, PoSI problem refers to construction of confidence regions \( \{\hat{R}_M : M \subseteq \{1, 2, \ldots, d\}\} \) (depending on \( \alpha \in [0, 1] \)) such that

\[
\liminf_{n \to \infty} \mathbb{P}\left( \beta_{\hat{M}} \in \hat{R}_M \right) \geq 1 - \alpha,
\]

(21)

holds for any randomly selected model \( \hat{M} \). It was proved in Kuchibhotla et al. (2018b, Theorem 3.1) that the PoSI problem (21) is equivalent to the simultaneous inference problem

\[
\liminf_{n \to \infty} \mathbb{P}\left( \bigcap_{M} \{\beta_M \in \hat{R}_M\} \right) \geq 1 - \alpha,
\]

where the intersection is taken over all \( M \subseteq \{1, 2, \ldots, d\} \). A straightforward construction of such simultaneous confidence regions can be based on finding quantiles of the statistic

\[
\max_{M} \max_{1 \leq j \leq |M|} \frac{\sqrt{n}(\hat{\beta}_M(j) - \beta_M(j))}{\hat{\sigma}_M(j)},
\]

(22)

where \( v(j) \), for a vector \( v \), represents the \( j \)-th coordinate of \( v \) and \( \hat{\sigma}_M(j) \) represents an estimator of the standard deviation of \( \sqrt{n}(\hat{\beta}_M(j) - \beta_M(j)) \). In order to apply Proposition 4.1 for finding quantiles of the statistic (22), we need a linear approximation result such as (20). In case of linear regression \( (\ell(u, v) = (u - v)^2 / 2) \), it was proved in Kuchibhotla et al. (2018a) that for any \( 1 \leq k \leq d \),

\[
\max_{|M| \leq k} \left\| \hat{\beta}_M - \beta_M + \frac{1}{n} \sum_{i=1}^n \Omega_M^{-1} X_{i,M}(Y_i - X_{i,M}^\top \beta_M) \right\|_2 = O_p\left( \frac{k \log(ed/k)}{n} \right),
\]
where $\Omega_M := n^{-1} \sum_{i=1}^n \mathbb{E}[X_{i,M} X_{i,M}^\top]$. Extensions for general $\ell$ are available in that paper. This result is proved under certain tail assumptions on the observations and holds both for random and fixed covariates. For linear regression with fixed covariates, we have

$$\hat{\beta}_M - \beta_M = \left( \frac{1}{n} \sum_{i=1}^n x_{i,M} x_{i,M}^\top \right)^{-1} \frac{1}{n} \sum_{i=1}^n x_{i,M} (Y_i - \mathbb{E}[Y_i]) = \frac{1}{n} \sum_{i=1}^n \Omega_M^{-1} x_{i,M} (Y_i - \mathbb{E}[Y_i]),$$

where we write $x_i$ to note fixed covariates and the variance of $n^{1/2}(\hat{\beta}_M - \beta_M)$ is given by

$$\text{Var} \left( \sqrt{n}(\hat{\beta}_M - \beta_M) \right) := \frac{1}{n} \sum_{i=1}^n \left( \Omega_M^{-1} x_{i,M} \right) \left( \Omega_M^{-1} x_{i,M} \right)^\top \text{Var}(Y_i). \quad (23)$$

Set $\psi_{j,M}(x_i, Y_i) = (Y_i - \mathbb{E}[Y_i]) (\Omega_M^{-1} x_{i,M}) (j)/\sigma_M(j)$, with $\sigma_M(j)$ representing the $j$-th diagonal element of the variance matrix (23). Define

$$\Delta_{\text{PoSI}} := \sup_{r \geq 0} \left\{ \mathbb{P} \left( \max_{|M| \leq k, 1 \leq j \leq |M|} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_{j,M}(x_i, Y_i) \right| \leq r \right) \right\} - \mathbb{P} \left( \max_{|M| \leq k, 1 \leq j \leq |M|} |G_{j,M}| \leq r \right),$$

where $(G_{j,M})_{j,M}$ has a multivariate normal distribution such that

$$\text{Cov}(G_{j,M}, G_{j',M'}) = \text{Cov} \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_{j,M}(x_i, Y_i), \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_{j',M'}(x_i, Y_i) \right).$$

By taking the maximum over $|M| \leq k$ in $\Delta_{\text{PoSI}}$, we are restricting the final selected model to have cardinality at most $k$. Without loss of generality, we can take $G_{j,M} = n^{-1/2} \sum_{i=1}^n g_i (\Omega_M^{-1} x_{i,M}) (j)/\sigma_M(j)$, for $g_i \overset{\text{iid}}{\sim} N(0, \text{Var}(Y_i))$. To bound $\Delta_{\text{PoSI}}$ from our results, we need the median $\mu_{\text{PoSI}}$:

$$\mu_{\text{PoSI}} := \text{median} \left( \max_{|M| \leq k} \max_{1 \leq j \leq |M|} |G_{j,M}| \right).$$

Note that the Gaussian vector $(G_{j,M})_{j,M}$ has dimension $p \asymp (ed/k)^k$ and hence $\mu_{\text{PoSI}} = O(\sqrt{k \log(ed/k)})$. Theorem 3.1 implies that $\Delta_{\text{PoSI}} = O(\mu_{\text{PoSI}} (k^4 \log(e/d)/n)^{1/6})$. Using the pessimistic bound on $\mu_{\text{PoSI}}$, we get that

$$\Delta_{\text{PoSI}} = O(1) \left( k^7 \log^7(e/d)/n \right)^{1/6}, \quad (24)$$

and this requires $k \log(ed/k) = o(n^{1/7})$. In words, this means that the selected model cardinality is at most $o(n^{1/7})$. Using Theorem 3.1, we can get a better requirement on $k$ if we have better bounds for $\mu_{\text{PoSI}}$. It was proved in Berk et al. (2013, Proposition 5.5) that if the covariates are orthogonal, that is, $n^{-1} \sum_{i=1}^n x_ix_i^\top = I_d$, then $\mu_{\text{PoSI}} = O(\sqrt{\log d})$ for any
1 ≤ k ≤ d. This result was further improved by Bachoc et al. (2018, Theorem 3.3) in that if there exists a κ ∈ [0, 1) such that for all |M| ≤ k

\[(1 - \kappa)\|\theta\|_2^2 \leq \frac{1}{n} \sum_{i=1}^{n} (x_{i,M}^\top \theta)^2 \leq (1 + \kappa)\|\theta\|_2^2 \text{ for all } \theta \in \mathbb{R}^{|M|}, \tag{25}\]

holds then \( \mu_{\text{PoSI}} \leq \sqrt{2\log(2d)} + C(\kappa)\kappa\sqrt{2k\log(6d/k)} \), for a function \( C(\cdot) \) satisfying \( C(\delta) \to 1 \) as \( \delta \to 0 \). Condition (25) is called the restricted isometry property (RIP) and is (trivially) satisfied with \( \kappa = 0 \) if the covariates are orthogonal. Hence if \( \kappa\sqrt{k} \) converges to zero then also \( \mu_{\text{PoSI}} = O(\sqrt{\log d}) \). Hence under the RIP condition (25) with \( \kappa\sqrt{k} \to 0 \), we get from Theorem 3.1 that \( \Delta_{\text{PoSI}} = O(\sqrt{\log(\text{ed})} (k^4\log^4(\text{ed}/k)/n)^{1/6}) \). This result only requires \( k(\log(\text{ed}))^{7/4} = o(n^{1/4}) \) which is much weaker than that implied by (24). To get a better perspective take \( k = d \) and for this case, we obtain

\[\Delta_{\text{PoSI}} = O(1) (d^4\log^3(d)/n)^{1/6}, \tag{26}\]

which converges to zero if \( d^4 = o(n/\log^3 n) \). On the other hand, the Berry-Esseen bound (Bentkus (2004)) for the linear regression estimator on the full model \( M_{\text{full}} = \{1, 2, \ldots, d\} \) (no simultaneity involved) requires \( d^{4.5} = o(n) \) which is close to the requirement from (26).

4.2. Maximal Inequalities for Empirical Processes

In this section we consider application of our non-uniform CLT result for the case of suprema of empirical processes. Suppose \((\xi_i, X_i), 1 \leq i \leq n\) are independent random variables in a measurable space \( \mathbb{R} \times \chi \) such that \( \mathbb{E}[\xi_i|X_i] = 0 \) for \( 1 \leq i \leq n \) and let \( \mathcal{F} \) be a class of functions from \( \chi \) to \( \mathbb{R} \). We consider the problem of bounding the moments of

\[Z_n(\mathcal{F}) := \sup_{f \in \mathcal{F}} \left| n^{-1/2} \sum_{i=1}^{n} \xi_i f(X_i) \right|.\]

Maximal inequalities that bound \( \mathbb{E}[Z_n(\mathcal{F})] \) are of great importance in the study of \( M \)-estimators and empirical risk minimizers; see Baraud (2016). In the study of least squares estimator obtained by minimizing the quadratic loss over \( f \in \mathcal{F} \) the rate is essentially determined by the behavior of \( \mathbb{E}[Z_n(\mathcal{F}')] \) over subsets \( \mathcal{F}' \) of \( \mathcal{F} \); see van de Geer and Wainwright (2017, Lemma 3.1). Rates of convergence when \( \xi_i \) only has finite number of moments has received some interest recently. Even though \( \mathcal{F} \) is usually countable and infinite, our results can be used in case of (weak) VC-major classes when \( \xi_i \) only has 3 moments. The calculation in this section could be used to prove rates for multiple isotonic regression when the errors are dependent on covariates and only have three moments; see Han (2019, page 24).

A class \( \mathcal{F} \) is weak VC-major with dimension \( d \geq 1 \) if \( d \) is the smallest integer \( k \geq 1 \) such that for all \( u \in \mathbb{R} \), the class \( \mathcal{C}_u(\mathcal{F}) := \{ x \in \chi : f(x) > u \} : f \in \mathcal{F} \) is a VC-class of subsets of \( \chi \) with dimension not larger than \( k \); see van der Vaart and Wellner
(1996, Section 2.6) for details on VC classes. It can be proved (Baraud (2016, Eq. (2.5))) that for a weak VC-major class of dimension $d$, $\log |\mathcal{E}_u(x_1, \ldots, x_n)| \leq d \log(2en/d)$ where $\mathcal{E}_u(x_1, \ldots, x_n) := \{i = 1, \ldots, n : x_i \in C : C \in \mathcal{C}_u(\mathcal{F})\}$. If $f(x) \in [0,1]$ for all $x \in \chi$, $f \in \mathcal{F}$ and $\mathcal{F}$ is a weak VC-major class of dimension $d$, then Baraud (2016, Section 3.3) shows

$$Z_n(\mathcal{F}) \leq \int_0^1 \sup_{C \in \mathcal{C}_u(\mathcal{F})} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_i \mathbb{1} \{X_i \in C\} \right| \, du = \int_0^1 \sup_{a \in \mathcal{A}_u} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_i a_i \right| \, du,$$

where $\mathcal{A}_u = \{a = (\mathbb{1} \{X_1 \in C\}, \ldots, \mathbb{1} \{X_n \in C\}) \in \{0,1\}^n : C \in \mathcal{C}_u(\mathcal{F})\}$. From the assumption of weak VC-major class in $\mathcal{F}$, the supremum over $a \in \mathcal{A}_u$ which allows us to apply our uniform and non-uniform CLTs (by conditioning on $X_1, \ldots, X_n$ so that $\mathcal{A}_u$ can be treated non-random). To state the final bound, note that after conditioning on $X_1, \ldots, X_n$ the “limiting” Gaussian vector can be taken to as $(G_a)_{a \in \mathcal{A}_u}$ where $G_a := n^{-1/2} \sum_{i=1}^n g_i \sigma(X_i) a_i$, for $\sigma^2(X_i) := \mathbb{E}[\xi_i^2 | X_i]$ and $g_i \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$. Set $X_n = \{X_1, \ldots, X_n\}$. If $\sup_a \mathbb{E}[\xi_i^3 | X_i = x] < \infty$, then Remark 3.2 implies that

$$\mathbb{E}_{|X_n} \left[ \sup_{a \in \mathcal{A}_u} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_i a_i \right| \right] = \mu_u + O(1) \mu_u^2 \log(\mu_u) \frac{(L_{n,u} d^2 \log^2(2en/d))^{1/3}}{n^{1/6}},$$

where $\mu_u := \mathbb{E}_{|X_n} \left[ \sup_{a \in \mathcal{A}_u} \left| n^{-1/2} \sum_{i=1}^n g_i \sigma(X_i) a_i \right| \right]$. Here $\mathbb{E}_{|X_n}$ represents the expectation conditional on $X_n$. Taking expectations with respect to $X_1, \ldots, X_n$, we get if $d = O(1)$,

$$\mathbb{E} \left[ Z_n(\mathcal{F}) \right] \leq (1 + o(1)) \int_0^1 \mathbb{E} \left[ \sup_{C \in \mathcal{C}_u(\mathcal{F})} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n g_i \sigma(X_i) \mathbb{1} \{X_i \in C\} \right| \right] \, du.$$

Note that the expectation on the right hand side is the Gaussian mean width of sets $\mathcal{A}_u$. The right hand side can be bounded using several existing results; see, e.g., Talagrand (2014, Chapters 2,10), Han and Wellner (2019, Theorem 1).

5. Cramér-type Large Deviation

In this section, we prove a Cramér-type large deviation result for $\|S_n\|$. A version of this result appeared in Bentkus (1987, Theorem 1) for the case of Banach space valued random variables. In the following result, we make the dependences on distributional constants precise. For this result, we assume that the observations $X_1, \ldots, X_n$ are i.i.d. and write the variation measure as $\zeta$ instead of $\zeta_i$. Set $C_2 = C_0 \log(ep), C_3 = C_0 \log^2(ep)$ We assume the following condition for the result: there exists $H \in (0, \infty)$ such that $\int \exp(H \|x\|) \zeta(dx) \leq 4$. 
We also need the following quantities. Set \( B := 2(1 + \sigma_{\text{max}}^{-2})H^{-1} \), and
\[
\tilde{\Pi}_n := 4 + 12\Phi_{AC,0}(8C_3L_n\varepsilon)^{1/3} + \frac{20\Phi_{AC,0} \log(ep) \log(8C_0n)}{Hn^{1/3}} + \frac{5.1 \log(ep)}{C_0n^{4/6}},
\]
\[
\Pi := \max \left\{ \tilde{\Pi}_n, (32\Phi_4)^{4/3}, \frac{19C_3L_n\varepsilon}{\Phi_4}, (37C_3L_n\varepsilon)^{4/7}, \left( \frac{24C_2}{\Phi_4^3H^2} \right)^{4/11} \right\},
\]
\[
M := \max \left\{ 2\Pi, (112\Phi_2 + 83C_3L_n)^{4/3}, \left( \frac{48C_2}{\Phi_4^3H^2} \right)^{10/23}, 36 \left( C_3L_n\Phi_2 \right)^{2/3}, \left( \frac{124C_3L_n}{(\mu + 1)^{17/8}n^{5/16}} \right) \right\}.
\]

The quantity \( M \) determines how big the ratio (2) is relative to \((r + 1)n^{-1/6}\).

**Theorem 5.1.** *(Cramér-type large deviation)* Under the setting above, we have for \( n \geq 4 \),
\[
\frac{\mathbb{P}(\|S_n\| > r)}{\mathbb{P}(\|Y\| > r)} - 1 \leq 1.02M(r + 1)n^{-1/6},
\]
for all \( r \) such that \((r + 1)n^{-1/6} \leq \mathfrak{B}_0 \exp \left( -3M^{1/4}(\mu + 1)^{-17/16}\log(en)n^{-5/32} \right) \), where
\[
\mathfrak{B}_0 := \min \left\{ \frac{1}{3(\Phi_4^3M)^{1/3}}, \frac{1}{2(\Phi_4^3M)^{1/15}}, \frac{(\log(en))^{-1/3}}{2(\Phi_4^3M)^{1/3}}, \frac{\Pi^{1/9}(\log(en))^{-4/9}}{2(\Phi_4^3M)^{1/3}}, \frac{M^{1/8}(\log(en))^{-1/2}}{2(\Phi_4^3M)^{1/3}} \right\}
\]

The following corollary is obtained for sub-Weibull random vectors using \( \Phi_2 \leq \Theta\mu^2 \leq \Theta \log(ep) \) and \( \Phi_4 \leq \Theta\mu^4 \leq \Theta \log^2(ep) \). For examples where \( \mu \) is of smaller order than \( \sqrt{\log(ep)} \) (as in post-selection inference example; Section 4.1) better rates follow easily. It is noteworthy that the rates in the corollary does not depend on \( \alpha \in [1, 2] \).

**Corollary 5.1.** Suppose that assumption (16) holds for some \( 1 \leq \alpha \leq 2 \). Then, there exist positive constants \( \Theta_1, \Theta_2 \in (0, \infty) \) depending on \( K_p, \sigma_{\text{min}} \) and \( \sigma_{\text{max}} \), such that
\[
\left| \frac{\mathbb{P}(\|S_n\| > r)}{\mathbb{P}(\|Y\| > r)} - 1 \right| \leq \Theta_1(\log(ep))^{8/3}(r + 1)n^{-1/6},
\]
for all \( n \geq (\log(ep))^{64/15}(\log(en))^{32/5}(\mu + 1)^{-34/5} \) and \((r + 1)n^{-1/6} \leq \Theta_2(\log(ep + en))^{-28/9} \).

The corollary requires \( n \) to be slightly larger than \( \log^4(ep) \) which becomes relaxed if \( \mu \geq \Theta \sqrt{\log(ep)} \). Note that \( n \geq \log^4(ep) \) is required for the bound on \( \delta_{n,0} \) to be less than 1.

6. Outline of the Proofs

The following represents the main steps in our proofs of uniform and non-uniform CLTs. The outline for the large deviation result is given at the end of this section. We also detail the differences between the proofs of Chernozhukov et al. (2017b) and Paulauskas and Račkauskas (1991), Bentkus et al. (2000). We define and control for \( r \geq 0 \),
\[
\Delta_{n,k}(r) := |\mathbb{P}(\|U_{n,k}\| \leq r) - \mathbb{P}(\|U_{n,0}\| \leq r)|.
\]
Recall that $\delta_{n,0} = \sup_{r \geq 0} \max_{1 \leq k \leq n} \Delta_{n,k}(r)$ and $\delta_{n,m} = \sup_{r \geq 0} \max_{1 \leq k \leq n} r^m \Delta_{n,k}(r)$.

**Step A: Smoothing Inequality.** The first step in all Berry-Esseen type results is a smoothing inequality which replaces the probabilities by expectations of smooth approximations of indicators. By Lemma 5.1.1 of Paulauskas and Račkauskas (2012), we have

$$\Delta_{n,k}(r) \leq \max_{\ell = 1,2} \left| \mathbb{E} [\varphi_{\ell}(U_{n,k}) - \varphi_{\ell}(U_{n,0})] \right| + \mathbb{P}(r - \varepsilon \leq \|U_{n,0}\| \leq r + \varepsilon),$$

where $\varphi_1(x) := \varphi_{r,\varepsilon}(x)$ and $\varphi_2(x) := \varphi_{r-\varepsilon,\varepsilon}(x)$ are given by Lemma 2.1. The second term in (27) is controlled by Theorem 2.2 for anti-concentration.

**Step B: Lindeberg Replacement.** Write $\varphi(\cdot)$ for both $\varphi_1(\cdot)$ and $\varphi_2(\cdot)$ and fix any $k \geq 1$

To bound the first term in (27), we use Lindeberg method.

$$|\mathbb{E}[\varphi(U_{n,k}) - \varphi(U_{n,0})]| \leq \sum_{j=1}^{k} |\mathbb{E}[\varphi(U_{n,j-1}) - \varphi(U_{n,j})]| = \sum_{j=1}^{k} \left| \mathbb{E}[\varphi(W_{n,j} + n^{-1/2}x)]\zeta_j(dx) \right|,$n

where $W_{n,j} = n^{-1/2}(X_1 + \ldots + X_{j-1} + Y_{j+1} + \ldots + Y_n)$. The last equality follows from $U_{n,j} = W_{n,j} + n^{-1/2}X_j$ and $U_{n,j-1} = W_{n,j} + n^{-1/2}Y_j$. Using the Taylor expansion of $\varphi$ (possible because of smoothness) and $\mathbb{E}[X_j] = \mathbb{E}[Y_j] = 0$, $\mathbb{E}[X_jX_j^\top] = \mathbb{E}[Y_jY_j^\top]$, we get

$$\left| \int \mathbb{E}[\varphi(W_{n,j} + n^{-1/2}x)]\zeta_j(dx) \right| = \left| \int \mathbb{E}[\text{Rem}_n(W_{n,j},x)]\zeta_j(dx) \right| \leq \int \mathbb{E}[|\text{Rem}_n(W_{n,j},x)|]\zeta_j|(dx) =: I_j,$n

where

$$\text{Rem}_n(y,x) = \varphi(y + xn^{-1/2}) - \varphi(y) - \frac{1}{\sqrt{n}} \sum_{j=1}^{p} x(j)\partial_j \varphi(y) - \frac{1}{2n} \sum_{j,k=1}^{p} x(j)x(k)\partial_{jk}\varphi(y).$$

**Step C: Splitting the Integral.** Based on Step B, it remains to bound the integral of $\text{Rem}_n(W_{n,j},x)$ with respect to $\zeta_j$. Using the mean value theorem and the bound on derivatives in Lemma 2.1, we have

$$|\text{Rem}_n(W_{n,j},x)| \leq \min \left\{ \frac{C_0 \log^2(ep)\varepsilon^{-3}\|x\|^3}{6n^{3/2}}, \frac{C_0 \log(ep)\varepsilon^{-2}\|x\|^2}{n} \right\}.$$ (28)

Since the support of $\varphi_1(\cdot), \varphi_2(\cdot)$ is contained in $\{x : \|x\| \in [r - \varepsilon, r + \varepsilon]\}$, the bound above on the remainder can be multiplied by $1\{\|W_{n,j}\| - r \leq \varepsilon + n^{-1/2}\|x\|\}$. Noting that the change point (where the minimum changes from first to second term) in the minimum is at order $n^{1/2}\varepsilon/\log(ep)$, we split the integral of remainder into two parts. Set $\mathcal{E} := \{x \in \mathbb{R}^p : \|x\| \leq n^{1/2}\varepsilon/\log(ep)\}$ and this yields

$$I_j = \int_{\mathcal{E}} \mathbb{E}[|\text{Rem}_n(W_{n,j},x)|]\zeta_j|(dx) + \int_{\mathcal{E}^c} \mathbb{E}[|\text{Rem}_n(W_{n,j},x)|]\zeta_j|(dx) =: I_j^{(1)} + I_j^{(2)}.$$
It is intuitively clear that \( I_j^{(1)} \) contributes to the main rate term in the bound and \( I_j^{(2)} \) would be a second order term since it is an integral over a tail end of the distribution.

In the classical proof of Bentkus et al. (2000), the splitting of integral is done at \( \|x\| \leq n^{1/2}\varepsilon \) and in the modern proof of Chernozhukov et al. (2017b) the splitting is done as above.

**Step D: Controlling \( I_j^{(2)} \).** Chernozhukov et al. (2017b) bound \( I_j^{(2)} \) as

\[
I_j^{(2)} \leq \int_{\mathcal{E}} \frac{C_0 \log^2(ep)\varepsilon^{-3}\|x\|^3}{6n^{3/2}}|\zeta_j|(dx).
\]

This is sub-optimal since from (28) the second term in the minimum is a better upper bound. Hence in Theorem 3.1, we bound \( I_j^{(2)} \) using

\[
I_j^{(2)} \leq \int_{\mathcal{E}} \frac{C_0 \log(ep)\varepsilon^{-2}\|x\|^2}{n}|\zeta_j|(dx). \tag{29}
\]

Summing over \( 1 \leq j \leq n \) leads to \( M_n(\varepsilon) \) in Theorem 3.1 and leads to better rates under \((2+\tau)\)-moments of \( \|X_j\| \).

In the classical proof of Bentkus et al. (2000), \( I_j^{(2)} \) is bounded using

\[
I_j^{(2)} \leq \int_{\mathcal{E}} \frac{C_0 \log(ep)\varepsilon^{-2}\|x\|^2}{n} \mathbb{P}(r - \varepsilon - n^{-1/2}\|x\| \leq \|W_{n,j}\| \leq r + \varepsilon + n^{-1/2}\|x\||)\zeta_j|(dx).
\]

Note that this bound is better than (29) since the probability is bounded by 1. Since \( W_{n,j} \) is related to \( U_{n-1,j-1}\sqrt{1 - 1/n} \), we can use the definition of \( \delta_{n-1,0} \) (or \( \delta_{n-1,m} \) for non-uniform CLT) to obtain

\[
\mathbb{P}(r - \varepsilon - n^{-1/2}\|x\| \leq \|W_{n,j}\| \leq r + \varepsilon + n^{-1/2}\|x\|)
\leq \mathbb{P}(\sqrt{n/(n-1)}(r - \varepsilon - n^{-1/2}\|x\|) \leq \|U_{n,0}\| \leq \sqrt{n/(n-1)}(r + \varepsilon + n^{-1/2}\|x\|)) + 2\delta_{n-1,0}
\leq \Phi_{AC,0}(\varepsilon + n^{-1/2}\|x\|)\sqrt{n/(n-1)} + 2\delta_{n-1,0}.
\]

Here the last inequality follows from Theorem 2.2. This is used in the proof of Theorem 3.1 to get the “right” dependence on \( \tau \) for the finite moment case. To make the recursion work, we need the anti-concentration inequality to not change with sample size and for this reason i.i.d. assumption is introduced in Proposition 3.1 and Theorem 3.2. In case of non-uniform CLT we need to split \( \mathcal{E}^c \) further in order to avoid dividing by zero when using \( \delta_{n-1,m} \).

**Step E: Controlling \( I_j^{(1)} \).** The simpler approach in bound \( I_j^{(1)} \) is in the classical proof of Bentkus et al. (2000) that uses

\[
I_j^{(1)} \leq \int_{\mathcal{E}} \frac{C_0 \log^2(ep)\varepsilon^{-3}\|x\|^3}{6n^{3/2}} \mathbb{P}(r - \varepsilon - n^{-1/2}\|x\| \leq \|W_{n,j}\| \leq r + \varepsilon + n^{-1/2}\|x\||)\zeta_j|(dx).
\]

Since \( \|x\| \leq n^{1/2}\varepsilon/\log(ep) \) for \( x \in \mathcal{E} \), the probability can be bounded by \( \mathbb{P}(r - 2\varepsilon \leq \|W_{n,j}\| \leq r + 2\varepsilon) \) (which does not involve \( x \) anymore). Now as in Step D, we can relate this
probability to $\delta_{n-1,0}$ and $\Phi_{AC,0}$. This results in $\int \|x\|^3 |\zeta_j|(dx) = \nu_3^3$ in the final bound in Proposition 3.1 and leads to a sub-optimal rate in case $X_j$ have exponential tails.

A better way to control $I_j^{(1)}$ from Chernozhukov et al. (2017b) uses

$$\left|\text{Rem}_n(W_{n,j}, x)\right|$$

$$= \frac{1}{2} \sum_{j_1,j_2,j_3=1}^p \frac{x(j_1)x(j_2)x(j_3)}{n^{3/2}} \int_0^1 (1-t)^2 D_{j_1j_2j_3}(W_{n,j} + tn^{-1/2}x) dt$$

$$\leq \sum_{j_1,j_2,j_3=1}^p \frac{|x(j_1)x(j_2)x(j_3)|}{2n^{3/2}} \int_0^1 (1-t)^2 D_{j_1j_2j_3} \left(W_{n,j} + \frac{tx}{\sqrt{n}}\right) 1 \left\{ \left\|W_{n,j}\right\| - r \leq \varepsilon + \frac{\|x\|}{\sqrt{n}} \right\} dt$$

By stability property (14) of $D_{j_1j_2j_3}$ along with $\|x\|/\sqrt{n} \leq \varepsilon / \log(ep)$ for $x \in \mathcal{E}$, we get

$$e^{-\varepsilon} D_{j_1j_2j_3}(W_{n,j}) \leq D_{j_1j_2j_3}(W_{n,j} + txn^{-1/2}) \leq e^\varepsilon D_{j_1j_2j_3}(W_{n,j}).$$

This implies

$$I_j^{(1)} \leq \frac{e^\varepsilon}{2n^{3/2}} \sum_{j_1,j_2,j_3=1}^p \int_{\mathcal{E}} |x(j_1)x(j_2)x(j_3)||\zeta_j|(dx)\mathbb{E}[D_{j_1j_2j_3}(W_{n,j}) \mathbb{1}\{\|W_{n,j}\| - r \leq 2\varepsilon\}].$$

By Hölder’s inequality,

$$\int_{\mathcal{E}} |x(j_1)x(j_2)x(j_3)||\zeta_j|(dx) \leq \max_{1 \leq j \leq p} \int_{\mathcal{E}} |x(j_1)|^3 |\zeta_j|(dx).$$

If $X_j$ has $(2 + \tau)$-moments for some $\tau \geq 1$, then the integral over $\mathcal{E}$ can be replaced by integral over $\mathbb{R}^p$ which leads to $L_n$ (when summed over $1 \leq j \leq n$). In case $X_j$ has only $(2 + \tau)$-moments for some $\tau < 1$, then using $\mathcal{E} = \{\|x\| \leq n^{1/2}\varepsilon / \log(ep)\}$, we can write

$$\int_{\mathcal{E}} |x(j_1)|^3 |\zeta_j|(dx) \leq \left( \frac{n^{1/2}\varepsilon}{\log(ep)} \right)^{1-\tau} \int |x(j_1)|^{2+\tau} |\zeta_j|(dx),$$

which would lead to a version of $L_n$ only involving “weak” $(2 + \tau)$-moment rather than the “weak” third moment. Getting back to $I_j^{(1)}$, the above discussion leads to

$$I_j^{(1)} \leq \frac{e^{\varepsilon}}{2n^{3/2}} \max_{1 \leq j_1 \leq p} \int_{\mathcal{E}} |x(j_1)|^3 |\zeta_j|(dx) \sum_{j_1,j_2,j_3=1}^p \mathbb{E}[D_{j_1j_2j_3}(W_{n,j}) \mathbb{1}\{\|W_{n,j}\| - r \leq 2\varepsilon\}].$$

There are two ways to bound the right hand side. First, the way used in the proof of Chernozhukov et al. (2017b) is to split the indicator inside the expectation by adding $Y_j \in \mathcal{E}$ and $Y_j \in \mathcal{E}^c$:

$$\mathbb{E}[D_{j_1j_2j_3}(W_{n,j}) \mathbb{1}\{\|W_{n,j}\| - r \leq 2\varepsilon\}] = \mathbb{E}[D_{j_1j_2j_3}(W_{n,j}) \mathbb{1}\{\|W_{n,j}\| - r \leq 2\varepsilon, Y_j \in \mathcal{E}\}]$$

$$+ \mathbb{E}[D_{j_1j_2j_3}(W_{n,j}) \mathbb{1}\{\|W_{n,j}\| - r \leq 2\varepsilon, Y_j \in \mathcal{E}^c\}].$$
The first term on the right hand side can be linked using the stability property (14) to $\mathbb{E}[D_{j1j2j3}(U_{n,j-1})\mathbb{1}\{||U_{n,j-1}|| - r \leq 3\varepsilon\}]$. Further using the bound on derivative this can be bounded by $C_3\varepsilon^{-3}\mathbb{P}(r - 3\varepsilon \leq ||U_{n,j-1}|| \leq r + 3\varepsilon)$ that can linked to $\delta_{n,0} + \Phi_{AC,0}\varepsilon$. The second term is, anyways, small since it involves $X_j \in \mathcal{E}^c$ which is a small probability event since $Y_j$ has Gaussian tails.

The calculations after (30) are used in the proof of Chernozhukov et al. (2017b) to get a bound in terms of $\delta_{n,0}$ and solve the inequality for $\delta_{n,0}$. This is what we followed in the proof of Theorem 3.1 but using the “right” bound on $I_j^{(2)}$.

For the proof of Theorem 3.2, we proceed from (30) by using

$$
\sum_{j_1,j_2,j_3=1}^p \mathbb{E}[D_{j_1j_2j_3}(W_{n,j})\mathbb{1}\{||W_{n,j}|| - r \leq 2\varepsilon\}] \leq C_3\varepsilon^{-3}\mathbb{P}(r - 2\varepsilon \leq ||W_{n,j}|| \leq r + 2\varepsilon),
$$

where the right hand side can be bounded in terms of $\delta_{n-1,0}$ and anti-concentration term as shown above in the control of $I_j^{(2)}$.

Combining the bounds and solving recursions (when needed) completes the proofs of all three versions of uniform CLT. By appropriate (minor) modifications mentioned in the outline above, the two versions of non-uniform CLT also follow.

**Sketch of the proof of Theorem 5.1.** The proof of Theorem 5.1 relies upon Lindeberg method and a refined induction argument. The starting point for the proof is Step B of the outline. To control $I_j$ we split it into two parts depending on whether $\{||x|| \leq n^{1/2}f_n(r)\}$ or not for some function $f_n(\cdot)$ of $r$. The case when $||x|| \geq n^{1/2}f_n(r)$ the integral is bounded using $\int \exp(H||x||)|\zeta|(dx) \leq 4$ and Markov’s inequality by:

$$
\frac{8\beta C_2\varepsilon^{-2}}{n H^2} \exp(-Hn^{1/2}f_n(r)/2) \leq \frac{8C_2\varepsilon^{-2}\beta}{n^{2+\Phi_i}H^2\Phi_0} \mathbb{P}(||Y|| > r). \tag{31}
$$

The integral corresponding to the case when $||x||$ is smaller than $n^{1/2}f_n(r)$ is bounded by

$$
\frac{C_3\varepsilon^{-3}e^{c_0(r)(log(ep))^{2}/\varepsilon}}{n^{3/2}} L_n \phi_n(r) \mathbb{P}\left( \bar{a}_n(r) \leq ||W_{n,j}|| \leq \bar{b}_n(r) \right), \tag{32}
$$

for suitably chosen $\bar{a}_n(r)$ and $\bar{b}_n(r)$. Since $W_{n,j} \overset{d}{=} ((n - 1)/n)^{1/2} U_{n-1,j-1}$, we get

$$
\Delta_n(r) \leq \mathbb{P}(r - \varepsilon \leq ||Y|| \leq r + \varepsilon) + \frac{8C_2\varepsilon^{-2}\beta}{\Phi_0 H^2 n^{1/2}+\Phi_1} \mathbb{P}(||Y|| > r) + \frac{C_3\varepsilon^{-3}L_n e^{c_0(r)(log(ep))^{2}/\varepsilon}}{6\sqrt{n}} \max_{0 \leq k \leq n-1} \mathbb{P}(a_n(r) \leq ||U_{n-1,k}|| \leq b_n(r)) \tag{33}
$$

for some $a_n(r)$ and $b_n(r)$. Our next task is to bound the right hand side of (33) in term of $\mathbb{P}(||Y|| > r)$. We now inductively use a bound on $\mathbb{P}(||U_{n-1,j}|| \geq a_n(r))$ in terms of $\mathbb{P}(||Y|| > r)$ to get a bound for (33). Finally bounding $\mathbb{P}(r - \varepsilon \leq ||Y|| \leq r + \varepsilon)$ in terms of $\mathbb{P}(||Y|| > r)$
using Theorem 2.1 and summing this with the bounds for (31), (32) we obtain the following result. For any $1 \leq k \leq n$, $|\mathbb{P}(\|U_{n,k}\| > r) / \mathbb{P}(\|Y\| > r) - 1| \leq \Pi T_{n,r}^{1/4}$, for all $r \in \mathbb{R}$ satisfying

$$T_{n,r} \leq \min \left\{ \frac{1}{16\Phi_4^4}, \frac{1}{2\Phi_4 B \log(en)}, \frac{\Pi^{1/3}}{(B \log(en) \log(ep))^{4/3}} \right\},$$

where $T_{n,r} = (r + 1)^3 n^{-1/2}$. The details are in Lemma S.6.3. Since $T_{n,r}^{1/4} = (r + 1)^{3/4} n^{-1/8}$, Lemma S.6.3 gives a large deviation with rate $n^{-1/8}$ (for all $n \geq 1$). However this rate can be modified to rate $n^{-1/6}$ by a more refined induction argument. This is done in detail in the proof of Theorem 5.1.

7. Summary and Future Directions

In this paper, we proved non-uniform central limit theorems and large deviations for scaled averages of independent high-dimensional random vectors based on dimension-free anti-concentration inequalities. We further illustrated the usefulness of these results in the context of post-selection inference for linear regression and in bounding the expectation of suprema of empirical processes. All the proofs are based on the Lindeberg method which was an integral tool in Banach space CLTs. Using the stability property introduced in Chernozhukov et al. (2017b), we obtained refinements for uniform as well as non-uniform CLTs. It should be mentioned here that we credit Bentkus (1987), Bentkus and Račkauskas (1990) for the proof of Theorem 5.1. In comparison to Chernozhukov et al. (2017b), we mention that our setting is restrictive in the sense that we consider $l_\infty$ balls while Chernozhukov et al. (2017b) consider general sparsely convex sets. Once an anti-concentration result such as Theorem 2.2, it is fairly easy to follow our proofs to extend the results which we hope to pursue in the future along with the discussion of “best” anti-concentration inequalities.

In this work, we have presented the results in the simpler setting with independent random vectors. The Lindeberg method is well-known for its robustness to independence assumptions and hence extensions to the case of dependent random vectors (in particular martingales) form an interesting future direction. Apart from the application in PoSI, other areas of interest in terms of applications are bootstrap and high-dimensional vectors with a specified group structure. Our results can be used to obtain CLTs and large deviations to empirical processes which form an interesting direction. See Norvaiša and Paulauskas (1991) and Chernozhukov et al. (2014) for some results.
Appendix A: Uniform CLT for \((2 + \tau)\)-moments with \(\tau < 1\)

Following the outline in (Step E of) Section 6 for \(\tau \leq 1\), we get for any \(\varepsilon > 0\):

\[
\delta_{n,0} \leq \frac{e^{2\varepsilon} C_3 \varepsilon^{-3} L_{n,\tau}}{2n^{3/2}} \left( \frac{n^{1/2} \varepsilon}{\log(ep)} \right)^{1-\tau} [2\Phi_{AC,0} \varepsilon + \delta_{n,0}] + \frac{C_2 M_n(\varepsilon)}{\varepsilon^2} + \Phi_{AC,0} \varepsilon
\]

\[
+ \frac{e^{2\varepsilon} C_3 \varepsilon^{-3}}{2n^{3/2}} \left( \frac{n^{1/2} \varepsilon}{\log(ep)} \right)^{1-\tau} \sum_{j=1}^{n} \max_{1 \leq j_1 \leq p} \int |x(j_1)|^{2+\tau} |\zeta_j| (dx) \mathbb{P}(\|Y_j\| > n^{1/2} \varepsilon / \log(ep)),
\]

where \(L_{n,\tau} = n^{-1} \sum_{j=1}^{n} \max_{1 \leq j_1 \leq p} \int |x(j_1)|^{2+\tau} |\zeta_j| (dx)\). It is also clear that

\[
M_n(\varepsilon) \leq \nu^{2+\tau} \left( \frac{\log(ep)}{n^{1/2} \varepsilon} \right)^{2+\tau}.
\]

Set \(\varepsilon = \varepsilon_n\) so that

\[
\frac{e^{2\varepsilon} C_3 L_{n,\tau}}{2n^{3/2}} \left( \frac{n^{1/2} \varepsilon}{\log(ep)} \right)^{1-\tau} \leq \frac{1}{2} \quad \text{and} \quad \frac{C_2 \nu^{2+\tau} \log(ep)}{n^{2+\tau} \varepsilon^{2+\tau}} \leq \frac{1}{r_n^{2+\tau}},
\]

for some \(r_n \geq 1\). This implies that

\[
\delta_{n,0} \lesssim \Phi_{AC,0} \varepsilon + \frac{1}{r_n^{2+\tau}} \lesssim \Phi_{AC,0} \frac{(L_{n,\tau} \log^2(ep))^{1/(2+\tau)}}{n^{1/2} \log(ep) (1-\tau)/(2+\tau)} + r_n \Phi_{AC,0} \nu^{2+\tau} \frac{(\log(ep))^{(\tau+1)/(\tau+2)}}{n^{\tau/(4+2\tau)}},
\]

Choosing the “optimal” order of \(r_n\) by equating the last two terms, we get

\[
\delta_{n,0} \lesssim \left[ \Phi_{AC,0} \frac{L_{n,\tau}^{1/(2+\tau)}}{n^{1/2}} + \Phi_{AC,0} \frac{(\nu^{2+\tau})^{(\tau+2)/(\tau+3)}}{n^{\tau/(6+2\tau)}} \right] \frac{(\log(ep))^{(\tau+1)/(\tau+2)}}{n^{\tau/(4+2\tau)}},
\]

The last inequality follows since \(n^{\tau/(6+2\tau)} \leq n^{1/8} \leq n^{1/2}\).

**Acknowledgment**

We would like to thank Prof. Jian Ding for comments that led to an improved presentation.

**References**

High-dimensional CLT


Dordrecht. Translated and revised from the 1989 Russian original.