Bernoulli **20**(4), 2014, 2020–2038
DOI: 10.3150/13-BEJ549

New concentration inequalities for suprema of empirical processes

JOHANNES LEDERER* and SARA VAN DE GEER**

Seminar für Statistik, ETH Zürich, Rämistrasse 101, 8092 Zürich, Switzerland.
*E-mail: *johanneslederer@mail.de; **geer@stat.math.ethz.ch*

While effective concentration inequalities for suprema of empirical processes exist under boundedness or strict tail assumptions, no comparable results have been available under considerably weaker assumptions. In this paper, we derive concentration inequalities assuming only low moments for an envelope of the empirical process. These concentration inequalities are beneficial even when the envelope is much larger than the single functions under consideration.

Keywords: chaining; concentration inequalities; deviation inequalities; empirical processes; rate of convergence

1. Introduction

Powerful concentration and deviation inequalities for suprema of empirical processes have been derived during the last 20 years. These inequalities turned out to be crucial for example, in the study of consistency and rates of convergence for many estimators. Unfortunately, the known inequalities are only valid for bounded empirical processes or under strict tail assumptions. So, this paper was prompted by the question whether useful inequalities can be obtained under considerably weaker assumptions.

Let us first set the framework, starting with a brief summary of the known results for bounded empirical processes, or more precisely, for empirical processes index by bounded functions. To this end, we consider independent and identically distributed random variables $X_1, \ldots, X_n$ and a countable function class $\mathcal{F}$ such that $\sup_{f\in\mathcal{F}} \|f\|_\infty \le 1$ and $\sup_{f\in\mathcal{F}} |\mathbb{E} f(X_1)| = 0$. The quantity of interest is denoted by $Y := \sup_{f\in\mathcal{F}} |\frac{1}{n}\sum_{i=1}^n f(X_i)|$ and the root of the maximal variance by $\sigma := \sup_{f\in\mathcal{F}} \sqrt{\mathbb{E}[f(X_1)]^2}$. Refining Rio's proof in [13] (see also [12], Chapter 5.3, for the proof techniques), Bousquet derives in [3] an exponential deviation inequality for Y. His result implies

$$\mathbb{P}\left(Y - (1+\varepsilon)\mathbb{E}Y \ge \sigma\sqrt{2x} + \left(\frac{1}{\varepsilon} + \frac{1}{3}\right)x\right) \le \mathrm{e}^{-nx} \qquad \text{for all } x, \varepsilon > 0. \tag{1}$$

For many statistical applications, it is important to have bounds like $\sigma\sqrt{2x} + (\frac{1}{\varepsilon} + \frac{1}{3})x$ and e^{-nx} that are, apart from the assumptions, completely independent of the functions f; the parameter $\varepsilon > 0$ is inserted to obtain such bounds. Exponential inequalities for bounded empirical processes similar to the one above have been found by Klein and Rio [8] and by Massart [11]. These inequalities are slightly less sharp, but additionally hold for nonidentically distributed random

1350-7265

variables and also for $-Y$. The derivations of the mentioned results rely on the entropy method (initiated by Ledoux in [9]), which provided a new approach to the results in Talagrand's seminal work [15]. For an overview of the techniques involved, we refer to the textbooks [2,10,12].

Results are also known for possibly unbounded empirical processes that have weak tails. We consider independent and identically distributed random variables $X_1, \dots, X_n$ and a function class $\mathcal{F}$ such that $\sup_{i, f\in\mathcal{F}} |\mathbb{E} f(X_i)| = 0$ and $\operatorname{card}\mathcal{F} = p$. We additionally assume that Bernstein conditions are fulfilled, that is, $\sup_{f\in\mathcal{F}} \frac{1}{n}\sum_{i=1}^n \mathbb{E}|f(X_i)|^m \le \frac{m!}{2} K^{m-2}$, $m = 2, 3, \dots$ for a constant K. Bühlmann and van de Geer then derive in [4] the following exponential deviation inequality for $Y := \sup_{f\in\mathcal{F}} |\frac{1}{n}\sum_{i=1}^n f(X_i)|$:

$$\mathbb{P}\left(Y - \sqrt{\frac{2\log(2p)}{n}} - \frac{K\log(2p)}{n} \ge Kx + \sqrt{2x}\right) \le \mathrm{e}^{-nx} \qquad \text{for all } x > 0.$$

The lower bounds $Kx + \sqrt{2x}$ and e^{-nx} are again independent of the functions f. Besides the classical results for Gaussian processes (see, e.g., [2] and the references therein), other exponential bounds for unbounded empirical processes are given by Adamczak in [1] and by van de Geer and Lederer in [17]. These authors assume weak tails with respect to suitable Orlicz norms.

But what if the empirical process is unbounded and does not fulfill the strict tail assumptions mentioned above? There is no hope to derive exponential bounds as above under considerably weaker assumptions. However, we show in the following that weak moment assumptions are sufficient to obtain useful moment type concentration inequalities. For this purpose, we consider independent, not necessarily identically distributed random variables $X_1, \dots, X_n$ and a countable function class $\mathcal{F}$ with an envelope that has pth moment at most M^p for a $M > 0$ and a $p \in [1, \infty)$. Our main result, Theorem 3.1, implies then for $Y := \sup_{f\in\mathcal{F}} |\frac{1}{n}\sum_{i=1}^n f(X_i)|$, $\sigma := \sup_{f\in\mathcal{F}} \sqrt{\mathbb{E}[f(X_1)]^2}$, $1 \le l \le p$, $(\cdot)_+ := \max\{0, \cdot\}$, $\|\cdot\|_l := (\mathbb{E}[\cdot]^l)^{1/l}$ and for all $\varepsilon > 0$

$$\left\|\left(Y - (1+\varepsilon)\mathbb{E}Y\right)_+\right\|_l \le \left(\frac{64}{\varepsilon} + 7 + \varepsilon\right)\left(\frac{l}{n}\right)^{1-l/p} M + 4\sqrt{\frac{l}{n}}\sigma$$

and

$$\left\|\left((1-\varepsilon)\mathbb{E}Y - Y\right)_+\right\|_l \le \left(\frac{86.4}{\varepsilon} + 7 - \varepsilon\right)\left(\frac{l}{n}\right)^{1-l/p} M + 4.7\sqrt{\frac{l}{n}}\sigma.$$

We argue in Section 3 that these bounds are especially useful in the common case where the envelope (measured by M) is much larger than the single functions (measured by σ). We also stress that the empirical process is present on the right-hand sides only through the quantities M and σ, which can be considered as properties of the single random variables $f(X_i)$, unlike in known maximal inequalities, which directly involve $\mathbb{E}Y$ or the entropy of the function set $\mathcal{F}$ (see, e.g., [12], Chapter 6, and [18]) at the corresponding spots. To obtain this, a parameter $\varepsilon > 0$ is required as above.

We close this section with a short outline of the paper. In Section 2, we give the basic definitions and assumptions. In Section 3, we then state and discuss the main result. This is followed by complementary bounds in Section 4. Detailed proofs are finally given in Section 5.

2. Random vectors, concentration inequalities and envelopes

We are mainly interested in the behavior of *suprema of empirical processes*

$$Y := \sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^{n} f(X_i)\right| \quad \text{or} \quad Y := \sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^{n}\big(f(X_i)-\mathbb{E}f(X_i)\big)\right| \tag{2}$$

for large n. Here, $X_1,\ldots,X_n$ are independent, not necessarily identically distributed random variables and $\mathcal{F}$ is a countable family of real, measurable functions. In the sequel, we may restrict ourselves to finitely many functions by virtue of the monotonous convergence theorem.

Random vectors generalize the notion of empirical processes. Let $\mathcal{Z}_1,\ldots,\mathcal{Z}_n$ be arbitrary probability spaces and $\{Z_i(j)\colon \mathcal{Z}_i \to \mathbb{R}, 1\le j\le N, 1\le i\le n\}$ a set of random variables. We then define the random vectors as $Z(j) := (Z_1(j),\ldots,Z_n(j))^T : \mathcal{Z}_1\times\cdots\times\mathcal{Z}_n \to \mathbb{R}^n$. For convenience, we introduce their mean as $PZ(j) := \frac{1}{n}\sum_{i=1}^n \mathbb{E}Z_i(j)$, their empirical mean as $\mathbb{P}_n Z(j) := \frac{1}{n}\sum_{i=1}^n Z_i(j)$, and the root of their maximal second moment as $\sigma := \max_{1\le j\le N}\sqrt{\frac{1}{n}\sum_{i=1}^n \mathbb{E}Z_i(j)^2}$ (all assumed to be finite). Throughout this paper, we then consider the generalized formulation of (2)

$$Z := \max_{1\le j\le N}\big|\mathbb{P}_n Z(j)\big|. \tag{3}$$

The corresponding results for the empirical processes (2) can be found via $Z_i(j) := f_j(X_i)$ or $Z_i(j) := f_j(X_i) - \mathbb{E}f_j(X_i)$ for $\mathcal{F} = \{f_1,\ldots,f_N\}$.

The basic assumption on the random vectors is expressed using *envelopes*. First, we call $\mathcal{E} := (\mathcal{E}_1,\ldots,\mathcal{E}_n)^T : \mathcal{Z}_1\times\cdots\times\mathcal{Z}_n \to \mathbb{R}^n$ an envelope if $|Z_i(j)| \le \mathcal{E}_i$ for all $1\le j\le N$ and $1\le i\le n$. The basic assumption of this paper is then that there is a $p\in[1,\infty)$ and an $M>0$ such that

$$\mathbb{E}\mathcal{E}_i^p \le M^p \tag{4}$$

for all $1\le i\le n$. To allow for an extension to countably infinite families of functions $\mathcal{F}$ via the monotoneous convergence theorem, we assume that the constant M is independent of N. Finally, we stress that the envelope $\mathcal{E}$ is typically much larger than the single random vectors $Z(j)$, that is, $M \gg \sigma$.

Asymptotically ($n\to\infty$), the processes (2) and (3) are typically governed by the central limit theorem. We study in this paper, however, the nonasymtotic behavior (n finite) of the process (3) (and thus of (2)). For n finite, *concentration inequalities* provide bounds for the deviations in both directions from the mean or related quantities. Similarly, *deviation inequalities* provide bounds for the deviation in one direction only. We are especially interested in bounds that depend only on n, M, σ, and p. The bounds should, in particular, not depend on the functions f and therefore not on $\mathbb{E}Z$ or $\mathcal{F}$.

3. Main result

We are mainly concerned with concentration inequalities for unbounded empirical processes that only fulfill weak moment conditions. In particular, we are interested in bounds that only depend on n, M, σ, and p and incorporate empirical processes with envelopes that may be much larger than the single functions under consideration.

The following theorem is the main result of this paper.

Theorem 3.1. *For $1 \leq l \leq p$ and all $\varepsilon > 0$ it holds that*

$$\left\| \left(Z - (1+\varepsilon)\mathbb{E}Z\right)_+ \right\|_l \leq \left(\frac{64}{\varepsilon} + 7 + \varepsilon\right)\left(\frac{l}{n}\right)^{1-l/p} M + 4\sqrt{\frac{l}{n}}\sigma$$

and

$$\left\| \left((1-\varepsilon)\mathbb{E}Z - Z\right)_+ \right\|_l \leq \left(\frac{86.4}{\varepsilon} + 7 - \varepsilon\right)\left(\frac{l}{n}\right)^{1-l/p} M + 4.7\sqrt{\frac{l}{n}}\sigma.$$

As discussed in the preceding section, we state our results in terms of random vectors instead of empirical processes. The connection can be made as described. Furthermore, we note that a considerable improvement with respect to l does not seem to be possible. Slightly better constants can be obtained, however, at the price of less incisive bounds or less accessible proofs (see Remark 5.1 in the proofs section). We finally note that the expectation $\mathbb{E}Z$ can be replaced by suitable approximations. Such approximations are usually found with chaining and entropy (see, e.g., [5,18,19]) or generic chaining (see, e.g., [7,14,16]).

Let us now have a closer look at the above result. In contrast to the known results given in the introduction, the single functions may be unbounded and may only fulfill weak moment conditions. For the envelope, the moment restrictions are increasing with increasing power l, as expected.

And what about large envelopes, that is $M \gg \sigma$? Theorem 3.1 separates the part including the size of the envelope (measured by M) from the part including the size of the single random vectors (measured by σ). For $p > 2l$ and $n \gg 1$, a possibly large value of M is counterbalanced by $\frac{1}{n^{1-l/p}} \ll \frac{1}{\sqrt{n}}$ and thus, the influence of large envelopes is tempered. In particular, the term including the size of the envelope can be neglected for $n \to \infty$ if p is sufficiently large.

We conclude this section with two straightforward consequences of Theorem 3.1 and an additional remark.

Corollary 3.1. *Theorem 3.1 directly implies probability bounds via Chebyshev's inequality. Under the above assumptions, it holds for $x > 0$*

$$\mathbb{P}\left(Z \geq (1+\varepsilon)\mathbb{E}Z + x\right) \leq \min_{1 \leq l \leq p} \frac{(((64/\varepsilon) + 7 + \varepsilon)(l/n)^{1-l/p} M + 4\sqrt{l}/n\sigma)^l}{x^l}$$

and similarly

$$\mathbb{P}\left(Z \leq (1-\varepsilon)\mathbb{E}Z - x\right) \leq \min_{1 \leq l \leq p} \frac{(((86.4/\varepsilon) + 7 - \varepsilon)(l/n)^{1-l/p} M + 4.7\sqrt{l/n}\sigma)^l}{x^l}.$$

If σ, M, ε, and $p-2l$ are strictly positive constants, this implies the logarithmic rate $(\ln(n))^{-l/2}$ for $\mathbb{P}(Y-(1+\varepsilon)\mathbb{E}Y \geq \sqrt{\ln(n)/n})$. This rate can be directly compared to the corresponding polynomial rates resulting from (1) (with $x \sim \ln(n)/n$) to observe that the avoiding of the boundedness assumption causes slower rates, as expected.

Corollary 3.2. *Concrete first order bounds under the above assumptions are for example*

$$\mathbb{E}[Z-2\mathbb{E}Z]_+ \leq 72\frac{M}{n^{1-1/p}} + 4\frac{\sigma}{\sqrt{n}}$$

and

$$\mathbb{E}\left[\frac{1}{2}\mathbb{E}Z - Z\right]_+ \leq 179.3\frac{M}{n^{1-1/p}} + 4.7\frac{\sigma}{\sqrt{n}}.$$

Remark 3.1. Allowing the right-hand side in Theorem 3.1 to depend on $\mathbb{E}Z$, we can avoid the parameter ε and find for example

$$\left\|(Z-\mathbb{E}Z)_+\right\|_l \leq 10.2\left(\frac{l}{n}\right)^{1-l/p} M + \sqrt{32\left(\frac{l}{n}\right)^{1-l/p} M\mathbb{E}Z} + \sqrt{\frac{2l}{n}}\sigma$$

and a similar bound for $\|(\mathbb{E}Z-Z)_+\|_l$. For the proof, one can proceed similarly as in Remark 5.1 and Lemma 5.3. These kinds of results are, however, of less statistical importance.

4. Complementary bounds

In this section, we complement the main result Theorem 3.1 with two additional bounds. These additional bounds can be of interest if l is close to p.

The first result reads as the following theorem.

Theorem 4.1. *Assume that the random variables $Z_i(j)$ are centered. For $1 \leq l \leq p$ it holds that*

$$\left\|(Z-4\mathbb{E}Z)_+\right\|_l \leq \left(l\Gamma(l/2)\right)^{1/l}\sqrt{\frac{32}{n}}M,$$

where Γ is the usual Gamma function.

Let us compare Theorem 4.1 with Theorem 3.1. On the one hand, the above result does not possess the flexibility of the factor $(1+\varepsilon)$ and is a deviation inequality only. On the other hand, the term including the size of the envelope M is independent of p and has a different power of n in the denominator compared to the corresponding term in Theorem 3.1. Comparing these two terms in detail, we find that the bound of Theorem 3.1 may be sharper than the corresponding bound in Theorem 4.1 if $l \leq p < 2l$.

We finally give explicit deviation inequalities for Z in the case of finitely many random vectors. For $p \geq 2$, explicit bounds are found immediately by replacing $\mathbb{E}Z$ in Theorem 3.1 or

Theorem 4.1 by the upper bound $\sqrt{\frac{8\log(2N)}{n}}M$ (see [6]). Another bound is found by an approach detailed in Section 5. The bound reads as follows.

Theorem 4.2. *Assume that the random variables $Z_i(j)$ are centered. Then, for $p \geq 2$, $l \in \mathbb{N}$, and $p \geq l$,*

$$\left\| \left(Z - 2M\frac{\log(2N)}{\sqrt{n}} \right)_+ \right\|_l \leq \sqrt{\frac{35}{n}} lM.$$

This can supersede the bound in Theorem 4.1 for $\log(2N) \leq 32$.

5. Proofs

In this last section, we give detailed proofs.

5.1. Proof of Theorem 3.1

The key idea of our proofs is to introduce an appropriate truncation that depends on the envelope of the empirical process. This allows us to split the problem into two parts that can be treated separately: On the one hand, a part corresponding to a bounded empirical process that can be treated by convexity arguments and Massart's results on bounded random vectors [11]. And on the other hand, a part corresponding to an unbounded empirical process that can be treated by rather elementary means.

For ease of exposition, we present some convenient notation for the truncation first. After deriving a simple auxiliary result, we then turn to the main task of this section: We first consider the truncated part of the problem in Lemma 5.2 and then prove Lemma 5.3, a generalization of Theorem 3.1.

A basic tool used in this section is *truncation*. Before turning to the proofs, we want to give some additional notation for this tool. First, we define the *unbounded* and the *bounded part of the random vectors* as

$$\overline{Z}(j) := \left(\overline{Z}_1(j), \dots, \overline{Z}_n(j)\right)^T := \left(Z_1(j)1_{\{\mathcal{E}_1 > K\}}, \dots, Z_n(j)1_{\{\mathcal{E}_n > K\}}\right)^T,$$
$$\underline{Z}(j) := \left(\underline{Z}_1(j), \dots, \underline{Z}_n(j)\right)^T := \left(Z_1(j)1_{\{\mathcal{E}_1 \leq K\}}, \dots, Z_n(j)1_{\{\mathcal{E}_n \leq K\}}\right)^T.$$

Similarly, we define

$$\overline{\mathcal{E}} := (\overline{\mathcal{E}}_1, \dots, \overline{\mathcal{E}}_n)^T := (\mathcal{E}_1 1_{\{\mathcal{E}_1 > K\}}, \dots, \mathcal{E}_n 1_{\{\mathcal{E}_n > K\}})^T,$$
$$\underline{\mathcal{E}} := (\underline{\mathcal{E}}_1, \dots, \underline{\mathcal{E}}_n)^T := (\mathcal{E}_1 1_{\{\mathcal{E}_1 \leq K\}}, \dots, \mathcal{E}_n 1_{\{\mathcal{E}_n \leq K\}})^T.$$

To prevent an overflow of indices, the *truncation level* $K > 0$ is not included explicitly in the notation. The truncation level is, however, given at the adequate places so that there should not be any confusion. Finally, we define the maxima of the truncated random variables as

$$\overline{Z} := \max_{1\le j\le N} \left|\mathbb{P}_n \overline{Z}(j)\right| \quad \text{and} \quad \underline{Z} := \max_{1\le j\le N} \left|\mathbb{P}_n \underline{Z}(j)\right|$$

and the maximal variance of the bounded parts as

$$\underline{\sigma} := \max_{1\le j\le N} \sqrt{\frac{1}{n}\sum_{i=1}^{n} \operatorname{Var} \underline{Z}_i(j)}.$$

Now we derive a simple auxiliary lemma.

Lemma 5.1. *Under the assumptions of Theorem 3.1, it holds that $\sigma \ge \underline{\sigma}$ and*

$$\left|\mathbb{E}[\underline{Z} - Z]\right|^l \le \frac{M^p}{K^{p-l}}$$

for the truncation level $K > 0$.

Proof. The first assertion is straightforward. For the second assertion, since $||a| - |b|| \le |a-b|$ for all $a, b \in \mathbb{R}$, we observe that

$$\begin{aligned}
\left|\mathbb{E}[\underline{Z} - Z]\right| &= \left|\mathbb{E}\Big[\max_{1\le j\le N}\left|\mathbb{P}_n \underline{Z}(j)\right| - \max_{1\le j\le N}\left|\mathbb{P}_n Z(j)\right|\Big]\right| \\
&\le \mathbb{E}\Big[\max_{1\le j\le N}\left|\left|\mathbb{P}_n \underline{Z}(j)\right| - \left|\mathbb{P}_n Z(j)\right|\right|\Big] \\
&\le \mathbb{E}\Big[\max_{1\le j\le N}\left|\mathbb{P}_n\big(\underline{Z}(j) - Z(j)\big)\right|\Big] \\
&= \mathbb{E}\Big[\max_{1\le j\le N}|\mathbb{P}_n \overline{Z}|\Big] \\
&\le \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\overline{\mathcal{E}}_i\right].
\end{aligned}$$

With Hölder's and Chebyshev's inequality, we obtain for $1 \le i \le n$

$$\begin{aligned}
\mathbb{E}\overline{\mathcal{E}}_i^l &= \mathbb{E}\mathcal{E}_i^l 1_{\{\mathcal{E}_i > K\}} \\
&\le \left(\mathbb{E}\mathcal{E}_i^p\right)^{l/p} (\mathbb{E}1_{\{\mathcal{E}_i > K\}})^{1-l/p} \\
&\le \left(\mathbb{E}\mathcal{E}_i^p\right)^{l/p} \left(\frac{\mathbb{E}\mathcal{E}_i^p}{K^p}\right)^{1-l/p} \\
&\le \frac{M^p}{K^{p-l}}.
\end{aligned}$$

These two results and Jensen's inequality yield then the second assertion. □

We can now turn to the harder part of this section. We first consider bounded random vectors in Lemma 5.2. We then proof a bound for unbounded random vectors in Lemma 5.3, from which the main result can be deduced easily.

Lemma 5.2. *Let* $1 \le l \le p$, $\varepsilon > 0$ *and denote by* $K > 0$ *the truncation level. Then*,

$$\left\| \left(\underline{Z} - (1+\varepsilon)\mathbb{E}\underline{Z}\right)_+ \right\|_l \le \left(\frac{64}{\varepsilon} + 5\right)\frac{lK}{n} + \frac{4\sqrt{l}\underline{\sigma}}{\sqrt{n}}$$

and

$$\left\| \left((1-\varepsilon)\mathbb{E}\underline{Z} - \underline{Z}\right)_+ \right\|_l \le \left(\frac{86.4}{\varepsilon} + 5\right)\frac{lK}{n} + \frac{4.7\sqrt{l}\underline{\sigma}}{\sqrt{n}}.$$

Proof. The key idea is to use convexity arguments so that we can apply well-known bounds for bounded random vectors.

To begin, we set $J := (32/\varepsilon + 2.5)K$ and $I := (2(l-1)J/n + \sqrt{8(l-1)}\underline{\sigma}/\sqrt{n})^l$ and then define the function $g_l : \mathbb{R}^+ \to (1, \infty)$ as

$$g_l(x) := \mathrm{e}^{(n/(2J^2))(\sqrt{2\underline{\sigma}^2 + J(x \vee I)^{1/l}} - \sqrt{2}\underline{\sigma})^2}.$$

We used here the notation $a \vee b := \max\{a, b\}$ for $a, b \in \mathbb{R}$. The function g_l is strictly increasing, smooth, convex on the interval (I, ∞), and its inverse on $(1, \infty)$ is given by

$$g_l^{-1}(y) = \left(\frac{2J}{n}\log y + \frac{4\underline{\sigma}}{\sqrt{n}}\sqrt{\log y}\right)^l. \tag{5}$$

The straightforward derivations of these facts are omitted for the sake of brevity.

The convexity of the function g_l makes it possible to apply a result of [11]. To show this, we introduce

$$X := \left(\underline{Z} - (1+\varepsilon)\mathbb{E}\underline{Z}\right)_+^l$$

and find with Jensen's inequality and the fact that g_l is increasing

$$g_l(\mathbb{E}X) \le g_l\left(\mathbb{E}[X \vee I]\right) \le \mathbb{E}g_l(X \vee I).$$

Massart's inequality [11], Theorem 4, (13), for bounded random vectors translates then to our setting as

$$\mathbb{P}\left(n\underline{Z} \ge (1+\varepsilon)n\mathbb{E}\underline{Z} + \underline{\sigma}\sqrt{8nx} + \left(\frac{32}{\varepsilon} + 2.5\right)Kx\right) \le \mathrm{e}^{-x},$$

where $x > 0$. This is equivalent to

$$\mathbb{P}\left(\underline{Z} \ge (1+\varepsilon)\mathbb{E}\underline{Z} + \underline{\sigma}\sqrt{\frac{8x}{n}} + \frac{J}{n}x\right) \le \mathrm{e}^{-x}. \tag{6}$$

We now deduce (cf. [17])

$$
\begin{aligned}
&\mathbb{E}\big[\mathrm{e}^{(n/(2J^2))(\sqrt{2\underline{\sigma}^2+J(X\vee I)^{1/l}}-\sqrt{2}\underline{\sigma})^2}\big]\\
&\quad=\int_0^\infty \mathbb{P}\big(\mathrm{e}^{(n/(2J^2))(\sqrt{2\underline{\sigma}^2+J(X\vee I)^{1/l}}-\sqrt{2}\underline{\sigma})^2}>t\big)\,\mathrm{d}t\\
&\quad\le 1+\int_1^\infty \mathbb{P}\big(\mathrm{e}^{(n/(2J^2))(\sqrt{2\underline{\sigma}^2+J(X\vee I)^{1/l}}-\sqrt{2}\underline{\sigma})^2}>t\big)\,\mathrm{d}t\\
&\quad=1+\int_1^\infty \mathbb{P}\left(\sqrt{2\underline{\sigma}^2+J(X\vee I)^{1/l}}>\sqrt{2}\underline{\sigma}+\sqrt{\frac{2J^2}{n}\log t}\right)\mathrm{d}t\\
&\quad=1+\int_1^\infty \mathbb{P}\left(J(X\vee I)^{1/l}>4\underline{\sigma}\sqrt{\frac{J^2}{n}\log t}+\frac{2J^2}{n}\log t\right)\mathrm{d}t
\end{aligned}
$$

and note that

$$
JI^{1/l}<4\underline{\sigma}\sqrt{\frac{J^2}{n}\log t}+\frac{2J^2}{n}\log t \quad\Leftrightarrow
$$

$$
\frac{2(l-1)J}{n}+\frac{\sqrt{8(l-1)}\underline{\sigma}}{\sqrt{n}}<4\underline{\sigma}\sqrt{\frac{\log t}{n}}+\frac{2J}{n}\log t.
$$

This is fulfilled if $t\ge \mathrm{e}^{l-1}$. Hence, with Massart's inequality (6),

$$
\begin{aligned}
&\mathbb{E}\big[\mathrm{e}^{(n/(2J^2))(\sqrt{2\underline{\sigma}^2+J(X\vee I)^{1/l}}-\sqrt{2}\underline{\sigma})^2}\big]\\
&\quad\le 1+\mathrm{e}^{l-1}-1+\int_{\mathrm{e}^{l-1}}^\infty \mathbb{P}\left(X^{1/l}>4\underline{\sigma}\sqrt{\frac{\log t}{n}}+\frac{2J}{n}\log t\right)\mathrm{d}t\\
&\quad=\mathrm{e}^{l-1}+\int_{\mathrm{e}^{l-1}}^\infty \mathbb{P}\left(\underline{Z}>(1+\varepsilon)\mathbb{E}\underline{Z}+4\underline{\sigma}\sqrt{\frac{\log t}{n}}+\frac{2J}{n}\log t\right)\mathrm{d}t\\
&\quad\le \mathrm{e}^{l-1}+\int_{\mathrm{e}^{l-1}}^\infty \exp\big(-\log t^2\big)\,\mathrm{d}t<\mathrm{e}^{l}.
\end{aligned}
$$

In summary, we have

$$
g_l(\mathbb{E}X)<\mathrm{e}^l.
$$

This is now inverted using equation (5) to obtain

$$
\mathbb{E}X\le\left(\frac{2lJ}{n}+\frac{4\sqrt{l}\underline{\sigma}}{\sqrt{n}}\right)^l.
$$

This finishes the proof of the first claim. The second claim can be deduced similarly using [11], Theorem 4, (14). □

Remark 5.1. The constants in Lemma 5.2 are not optimal. First, we note that more restrictive assumptions allow one to replace Massart's inequality (6) by sharper concentration inequalities (e.g., from Klein and Rio [8] assuming centered random vectors or from Bousquet [3] assuming centered and identically distributed random vectors) and permit therefore shaper bounds. Second, instead of using such concentration inequalities, one can work with the underlying log-Laplace transforms directly. We found that this approach leads to slightly better constants but also to a less accessible proof. Let us sketch the approach:

One may first verify that for any $t > 0$ and $a \geq 0$

$$\log \mathbb{E}\big[n(\underline{Z} - \mathbb{E}\underline{Z}) - a\big]_+^l \leq \log \mathbb{E}\big[\mathrm{e}^{tn(\underline{Z}-\mathbb{E}\underline{Z})}\big] + l\log(l/t) - l - ta. \tag{7}$$

We can now use bounds for the log-Laplace transform $\log \mathbb{E}[\mathrm{e}^{tn(\underline{Z}-\mathbb{E}\underline{Z})}]$ of $n(\underline{Z} - \mathbb{E}\underline{Z})$, for example, from [11]:

$$\log \mathbb{E}\big[\mathrm{e}^{tn(\underline{Z}-\mathbb{E}\underline{Z})}\big] \leq \frac{vt^2}{1 - 2.5Kt} \qquad \text{with } v := 2n\underline{\sigma}^2 + 32Kn\mathbb{E}\underline{Z}. \tag{8}$$

The bound (8) with $t := (\sqrt{\frac{v}{l}} + 2.5K)^{-1}$ can be inserted into (7) and the result can be simplified with

$$(\sqrt{\alpha + \beta} + \gamma)\mathrm{e}^{-\alpha/\delta/(\sqrt{\alpha+\beta}+\gamma)} \leq \sqrt{\beta} + 2\gamma + \delta$$

for $\alpha, \beta, \gamma, \delta > 0$. This then leads to the bound

$$\big\|\big(\underline{Z} - (1+\varepsilon)\mathbb{E}\underline{Z}\big)_+\big\|_l \leq \left(\frac{32}{\varepsilon} + 5\right)\frac{lK}{n} + \frac{\sqrt{2l}\underline{\sigma}}{\sqrt{n}}.$$

The quantity $\|((1-\varepsilon)\mathbb{E}\underline{Z} - \underline{Z})_+\|_l$ can be bounded similarly.

We now use the above lemma to prove the following.

Lemma 5.3. *Let $1 \leq l \leq p$, $\varepsilon > 0$ and denote by $K > 0$ the truncation level. Then,*

$$\big\|\big(Z - (1+\varepsilon)\mathbb{E}\underline{Z}\big)_+\big\|_l \leq \left(\frac{64}{\varepsilon} + 5\right)\frac{lK}{n} + \frac{4\sqrt{l}\underline{\sigma}}{\sqrt{n}} + \frac{M^{p/l}}{K^{p/l-1}}$$

and

$$\big\|\big((1-\varepsilon)\mathbb{E}\underline{Z} - Z\big)_+\big\|_l \leq \left(\frac{86.4}{\varepsilon} + 5\right)\frac{lK}{n} + \frac{4.7\sqrt{l}\underline{\sigma}}{\sqrt{n}} + \frac{M^{p/l}}{K^{p/l-1}}.$$

Proof. The key idea of the proof is to separate the bounded from the unbounded quantities. We then develop bounds for $\|\overline{Z}\|_l$ via elementary means and combine this with the above results to deduce the desired bounds.

We start with the proof of the first inequality. First, we split Z in a bounded and an unbounded part

$$\begin{aligned} Z &= \max_{1\le j\le N} \left|\mathbb{P}_n Z(j)\right| \\ &= \max_{1\le j\le N} \left|\mathbb{P}_n\left(\underline{Z}(j) + \overline{Z}(j)\right)\right| \\ &\le \max_{1\le j\le N} \left(\left|\mathbb{P}_n \underline{Z}(j)\right| + \left|\mathbb{P}_n \overline{Z}(j)\right|\right) \\ &\le \underline{Z} + \overline{Z} \end{aligned}$$

and deduce with the triangle inequality that

$$\begin{aligned} &\left\|\left(Z - (1+\varepsilon)\mathbb{E}\underline{Z}\right)_+\right\|_l \\ &\quad\le \left\|\left(\underline{Z} + \overline{Z} - (1+\varepsilon)\mathbb{E}\underline{Z}\right)_+\right\|_l \\ &\quad\le \left\|\left(\underline{Z} - (1+\varepsilon)\mathbb{E}\underline{Z}\right)_+ + \overline{Z}\right\|_l \\ &\quad\le \left\|\left(\underline{Z} - (1+\varepsilon)\mathbb{E}\underline{Z}\right)_+\right\|_l + \|\overline{Z}\|_l. \end{aligned} \tag{9}$$

Now, we turn to the development of bounds for $\|\overline{Z}\|_l$. As above, with the help of Hölder's and Chebyshev's inequalities, we obtain for $1 \le i \le n$

$$\mathbb{E}\overline{\mathcal{E}}_i^l \le \frac{M^p}{K^{p-l}}$$

and therefore with the triangle inequality

$$\|\overline{Z}\|_l \le \|\mathbb{P}_n \overline{\mathcal{E}}\|_l \le \frac{M^{p/l}}{K^{p/l-1}}. \tag{10}$$

Combining inequalities (9), (10), and the bound from Lemma 5.2 gives finally

$$\left\|\left(Z - (1+\varepsilon)\mathbb{E}\underline{Z}\right)_+\right\|_l \le \left(\frac{64}{\varepsilon} + 5\right)\frac{lK}{n} + \frac{4\sqrt{l}\sigma}{\sqrt{n}} + \frac{M^{p/l}}{K^{p/l-1}}.$$

This finishes the proof of the first part of the lemma. For the second part, we note that

$$\begin{aligned} \underline{Z} &= \max_{1\le j\le N} \left|\mathbb{P}_n \underline{Z}(j)\right| \\ &= \max_{1\le j\le N} \left|\mathbb{P}_n\left(Z(j) - \overline{Z}(j)\right)\right| \\ &\le \max_{1\le j\le N} \left(\left|\mathbb{P}_n Z(j)\right| + \left|\mathbb{P}_n \overline{Z}(j)\right|\right) \\ &\le Z + \overline{Z} \end{aligned}$$

and therefore $Z \geq \underline{Z} - \overline{Z}$. Consequently,

$$\begin{aligned}
&\big\|\big((1-\varepsilon)\mathbb{E}\underline{Z} - Z\big)_+\big\|_l \\
&\quad \leq \big\|\big((1-\varepsilon)\mathbb{E}\underline{Z} - \underline{Z} + \overline{Z}\big)_+\big\|_l \\
&\quad \leq \big\|\big((1-\varepsilon)\mathbb{E}\underline{Z} - \underline{Z}\big)_+ + \overline{Z}\big\|_l \\
&\quad \leq \big\|\big((1-\varepsilon)\mathbb{E}\underline{Z} - \underline{Z}\big)_+\big\|_l + \|\overline{Z}\|_l.
\end{aligned}$$

One can then proceed as in the first part. □

Proof of Theorem 3.1. Set $K = (\frac{n}{l})^{l/p} M$ in Lemma 5.3 and use Lemma 5.1 to replace the truncated quantities by the original ones. □

5.2. Proof of Theorem 4.1

Here, we prove Theorem 4.1 with the help of symmetrization and desymmetrization.

Proof of Theorem 4.1. The trick is to use symmetrization and desymmetrization arguments so that we are able to use [11], Theorem 9, in a favorable way.

Beforehand, we define $Z_\varepsilon := \max_{1\leq j\leq N} |\frac{1}{n}\sum_{i=1}^n \varepsilon_i Z_i(j)|$ with independent Rademacher random variables ε_i. Then, we symmetrize according to [19], Lemma 2.3.6, with the function $\Phi(x) = (x - 4\mathbb{E}Z)_+^l$ to obtain

$$\mathbb{E}[Z - 4\mathbb{E}Z]_+^l \leq \mathbb{E}[2Z_\varepsilon - 4\mathbb{E}Z]_+^l$$

and we desymmetrize with the function $\Phi(x) = x$ to obtain

$$\mathbb{E}[2Z_\varepsilon - 4\mathbb{E}Z]_+^l \leq \mathbb{E}[2Z_\varepsilon - \mathbb{E}2Z_\varepsilon]_+^l.$$

Hence,

$$\mathbb{E}[Z - 4\mathbb{E}Z]_+^l \leq 2^l \mathbb{E}\mathbb{E}_\varepsilon[Z_\varepsilon - \mathbb{E}Z_\varepsilon]_+^l, \tag{11}$$

where we write here and in the following $\mathbb{E}_\varepsilon$ for the expectation and $\mathbb{P}_\varepsilon$ for the probability w.r.t. the Rademacher random variables. Next, we observe that

$$\begin{aligned}
&\mathbb{E}_\varepsilon[Z_\varepsilon - \mathbb{E}Z_\varepsilon]_+^l \\
&\quad = \int_0^\infty \mathbb{P}_\varepsilon\big((Z_\varepsilon - \mathbb{E}Z_\varepsilon)_+^l > t\big)\,\mathrm{d}t \\
&\quad = \int_0^\infty \mathbb{P}_\varepsilon\big(Z_\varepsilon > \mathbb{E}Z_\varepsilon + t^{1/l}\big)\,\mathrm{d}t \\
&\quad \leq \int_0^\infty \mathbb{P}_\varepsilon\left(\max_{1\leq j\leq N} \frac{1}{n}\sum_{i=1}^n \varepsilon_i Z_i(j) > \mathbb{E}\max_{1\leq j\leq N} \frac{1}{n}\sum_{i=1}^n \varepsilon_i Z_i(j) + t^{1/l}\right)\mathrm{d}t
\end{aligned}$$

$$+\int_0^\infty \mathbb{P}_\varepsilon\left(\max_{1\le j\le N} -\frac{1}{n}\sum_{i=1}^n \varepsilon_i Z_i(j) > \mathbb{E}\max_{1\le j\le N}\frac{1}{n}\sum_{i=1}^n \varepsilon_i Z_i(j) + t^{1/l}\right)\mathrm{d}t$$

$$\le 2\int_0^\infty \mathbb{P}_\varepsilon\left(\max_{1\le j\le N} \frac{1}{n}\sum_{i=1}^n \varepsilon_i Z_i(j) > \mathbb{E}\max_{1\le j\le N}\frac{1}{n}\sum_{i=1}^n \varepsilon_i Z_i(j) + t^{1/l}\right)\mathrm{d}t.$$

In a final step, we apply Massart's inequality [11], Theorem 9, with

$$L^2 = \max_{1\le j\le N}\sum_{i=1}^n \bigl(2\bigl|Z_i(j)\bigr|\bigr)^2 \le 4n\mathbb{P}_n\mathcal{E}^2,$$

where $\mathbb{P}_n\mathcal{E}^2 := \frac{1}{n}\sum_{i=1}^n \mathcal{E}_i^2$. This yields

$$\begin{aligned}
&2\int_0^\infty \mathbb{P}_\varepsilon\left(\max_{1\le j\le N} \frac{1}{n}\sum_{i=1}^n \varepsilon_i Z_i(j) > \mathbb{E}\max_{1\le j\le N}\frac{1}{n}\sum_{i=1}^n \varepsilon_i Z_i(j) + t^{1/l}\right)\mathrm{d}t\\
&\quad\le 2\int_0^\infty \exp\left(-\frac{nt^{2/l}}{8\mathbb{P}_n\mathcal{E}^2}\right)\mathrm{d}t\\
&\quad= 2\left(\frac{8}{n}\right)^{l/2}\bigl(\mathbb{P}_n\mathcal{E}^2\bigr)^{l/2}\int_0^\infty \exp\bigl(-t^{2/l}\bigr)\,\mathrm{d}t\\
&\quad= 2\left(\frac{8}{n}\right)^{l/2}\bigl(\mathbb{P}_n\mathcal{E}^2\bigr)^{l/2}\frac{l\Gamma(l/2)}{2}.
\end{aligned}$$

With inequality (11), this gives

$$\mathbb{E}[Z-4\mathbb{E}Z]_+^l \le 2^l l\left(\frac{8}{n}\right)^{l/2}\mathbb{E}\bigl[\mathbb{P}_n\mathcal{E}^2\bigr]^{l/2}\Gamma\left(\frac{l}{2}\right).$$

Finally, due to the triangle inequality, it holds that

$$\mathbb{E}\bigl[\mathbb{P}_n\mathcal{E}^2\bigr]^{l/2} \le \mathbb{E}[\mathbb{P}_n\mathcal{E}]^l \le M^l$$

and hence

$$\mathbb{E}[Z-4\mathbb{E}Z]_+^l \le l\Gamma\left(\frac{l}{2}\right)\left(\frac{32}{n}\right)^{l/2}M^l. \qquad \square$$

5.3. Proof of Theorem 4.2

We eventually derive Theorem 4.2 using truncation. After some auxiliary results, we derive Lemma 5.6. This lemma settles the bounded part of the problem. It is then used to proof Lemma 5.7 which is a slight generalization of the main theorem. Finally, we derive Theorem 4.2 as a simple corollary.

We begin with two auxiliary lemmas.

Lemma 5.4. *Let W be a centered random variable with values in $[-A, A]$, $A \geq 0$, such that $\mathbb{E}W^2 \leq 1$. Then,*

$$\mathbb{E}\mathrm{e}^{W/A} \leq 1 + \frac{1}{A^2}.$$

Proof. We follow well known ideas (see, e.g., [4], Chapter 14):

$$\begin{aligned}
\mathbb{E}\mathrm{e}^{W/A} &= 1 + \mathbb{E}\left[\mathrm{e}^{W/A} - 1 - \frac{W}{A}\right] \\
&\leq 1 + \mathbb{E}\left[\mathrm{e}^{|W|/A} - 1 - \frac{|W|}{A}\right] \\
&= 1 + \sum_{m=2}^{\infty} \frac{\mathbb{E}|W|^m}{m!A^m} \\
&\leq 1 + \sum_{m=2}^{\infty} \frac{A^{m-2}}{m!A^m} \\
&\leq 1 + \frac{1}{A^2}.
\end{aligned}$$

□

Lemma 5.5. *Let $C_m^n := |\{(i_1, \ldots, i_m)^T \in \{1, \ldots, n\}^m : \forall j \in \{1, \ldots, m\} \exists j' \in \{1, \ldots, m\}, j' \neq j, i_j = i_{j'}\}|$ for $m, n \in \mathbb{N}$. Then,*

$$C_m^n \leq m! \left(\frac{n}{2}\right)^{\lfloor m/2 \rfloor}.$$

Proof. The proof of this lemma is a simple counting exercise. We start with the case $m \leq 2$. One finds easily that $C_1^n = 0$ and $C_2^n = n$, which completes the case $m \leq 2$. Next, we consider the case $m > 2$. To this end, we note that $C_m^1 = 1$, $C_3^2 = 2$ and $C_m^2 \leq 2^m \leq m!$ for $m > 3$. This completes the cases $n \leq 2$. Now, we do an induction in n. So we let $n \geq 2$ and find

$$\begin{aligned}
C_m^{n+1} &= C_m^n + \frac{m(m-1)}{2!} C_{m-2}^n \\
&\quad + \frac{m(m-1)(m-2)}{3!} C_{m-3}^n + \cdots + \frac{m(m-1)\cdots 3}{(m-2)!} C_2^n + 1.
\end{aligned}$$

By induction, this yields

$$\begin{aligned}
C_m^{n+1} &\leq m! \left[\left(\frac{n}{2}\right)^{\lfloor m/2 \rfloor} + \frac{1}{2!}\left(\frac{n}{2}\right)^{\lfloor (m-2)/2 \rfloor} \right. \\
&\quad \left. + \frac{1}{3!}\left(\frac{n}{2}\right)^{\lfloor (m-3)/2 \rfloor} + \cdots + \frac{1}{(m-2)!}\left(\frac{n}{2}\right)^{\lfloor 2/2 \rfloor}\right] + 1.
\end{aligned}$$

We now assume that m is even. So,

$$
\begin{aligned}
C_m^{n+1} &\le m!\biggl[\Bigl(\frac{n}{2}\Bigr)^{m/2} + \frac{1}{2!}\Bigl(\frac{n}{2}\Bigr)^{m/2-1} + \frac{1}{3!}\Bigl(\frac{n}{2}\Bigr)^{m/2-2} + \cdots + \frac{1}{(m-2)!}\Bigl(\frac{n}{2}\Bigr)\biggr] + 1 \\
&= m!\biggl[\Bigl(\frac{n}{2}\Bigr)^{m/2} + \frac{1}{2!}\Bigl(\frac{n}{2}\Bigr)^{m/2-1} + \sum_{j=2}^{m/2-1}\Bigl(\frac{1}{(2j-1)!} + \frac{1}{(2j)!}\Bigr)\Bigl(\frac{n}{2}\Bigr)^{m/2-j}\biggr] + 1 \\
&\le m!\biggl[\Bigl(\frac{n}{2}\Bigr)^{m/2} + \frac{m}{4}\Bigl(\frac{n}{2}\Bigr)^{m/2-1} + \sum_{j=2}^{m/2-1}\binom{\frac{m}{2}}{j}\Bigl(\frac{1}{2}\Bigr)^{j}\Bigl(\frac{n}{2}\Bigr)^{m/2-j} + \Bigl(\frac{1}{2}\Bigr)^{m/2}\biggr] \\
&= m!\sum_{j=0}^{m/2}\binom{\frac{m}{2}}{j}\Bigl(\frac{1}{2}\Bigr)^{j}\Bigl(\frac{n}{2}\Bigr)^{m/2-j} = m!\Bigl(\frac{n+1}{2}\Bigr)^{\lfloor m/2\rfloor}.
\end{aligned}
$$

This completes the proof for $m > 2$ with m even. We note finally, that for odd $m > 2$ we have $C_m^n < mC_{m-1}^n \le m!(\frac{n}{2})^{\lfloor m/2\rfloor}$. $\square$

We now settle the bounded part of the problem. Bounded random variables are in particular subexponential, so one could apply results from [20], for example. But for our purposes, a direct treatment as in the following is more suitable.

Lemma 5.6. *Let $l \in \mathbb{N}$ and $p, A \ge 2$. Then, for the truncation level $K = \frac{A}{2} + \sqrt{\frac{A^2}{4} - 1}$,*

$$
\Bigl\|\Bigl(\max_{1\le j\le N}(\mathbb{P}_n - P)\underline{Z}(j) - AM\frac{\log(N)}{n}\Bigr)_+\Bigr\|_l \le \frac{M}{A} + \frac{lAM}{n}.
$$

Proof. We assume w.l.o.g. $M = 1$ and observe that

$$
\mathbb{E}\bigl[\underline{Z}_i(j) - \mathbb{E}\underline{Z}_i(j)\bigr]^2 \le \mathbb{E}\underline{Z}_i(j)^2 \le 1.
$$

Moreover, because of Hölder's and Chebyshev's inequalities and $K \ge 1$, it holds that

$$
\bigl|\underline{Z}_i(j) - \mathbb{E}\underline{Z}_i(j)\bigr| \le \bigl|\underline{Z}_i(j)\bigr| + \bigl|\mathbb{E}\underline{Z}_i(j)\bigr| \le K + \frac{1}{K} = A.
$$

These observations, the independence of the random variables and Lemma 5.4 yield then

$$
\begin{aligned}
&\mathbb{E}\mathrm{e}^{n(\mathbb{P}_n - P)\underline{Z}(j)/A} \\
&\quad = \mathbb{E}\mathrm{e}^{\sum_{i=1}^n(\underline{Z}_i(j) - \mathbb{E}\underline{Z}_i(j))/A} \\
&\quad \le \Bigl(1 + \frac{1}{A^2}\Bigr)^n.
\end{aligned}
$$

Next, one checks easily, that the map $x \mapsto \mathrm{e}^{x^{1/l}}$ is convex on the set $[(l-1)^l, \infty)$. Hence, using Jensen's inequality again, we obtain

$$
\begin{aligned}
&\left\| \left(\max_{1 \le j \le N} (\mathbb{P}_n - P)\underline{Z}(j) - A\frac{\log(N)}{n} \right)_+ \right\|_l \\
&\quad \le \frac{A}{n} \left\| \left(\max_{1 \le j \le N} n(\mathbb{P}_n - P)\underline{Z}(j)/A - \log(N) \right)_+ \vee (l-1) \right\|_l \\
&\quad \le \frac{A}{n} \log\Big(\mathbb{E}\exp\Big(\Big(\max_{1 \le j \le N} n(\mathbb{P}_n - P)\underline{Z}(j)/A - \log(N) \Big)_+ \vee (l-1) \Big) \Big) \\
&\quad = \frac{A}{n} \log\Big(\mathbb{E}\exp\Big(\Big(\max_{1 \le j \le N} n(\mathbb{P}_n - P)\underline{Z}(j)/A - \log(N) \Big) \vee (l-1) \Big) \Big) \\
&\quad \le \frac{A}{n} \log\Big(\max_{1 \le j \le N} \mathbb{E}\exp\big(n(\mathbb{P}_n - P)\underline{Z}(j)/A \big) + \mathrm{e}^{l-1} \Big) \\
&\quad \le \frac{A}{n} \log\left(\left(1 + \frac{1}{A^2} \right)^n + \mathrm{e}^{l-1} \right).
\end{aligned}
$$

We finally note that $a + b < \mathrm{e}ab$ for all $a, b \ge 1$ and find

$$
\begin{aligned}
&\left\| \left(\max_{1 \le j \le N} (\mathbb{P}_n - P)\underline{Z}(j) - A\frac{\log(N)}{n} \right)_+ \right\|_l \\
&\quad < \frac{A}{n} \log\left(\left(1 + \frac{1}{A^2} \right)^n \mathrm{e}^l \right) \\
&\quad = \frac{A}{n} \left(\log\left(1 + \frac{1}{A^2} \right)^n + \log \mathrm{e}^l \right) \\
&\quad \le \frac{1}{A} + \frac{lA}{n}.
\end{aligned}
$$

□

The results above can now be used to derive a generalization of the main problem.

Lemma 5.7. *Assume that the random variables $Z_i(j)$ are centered. Then, for $p, A \ge 2$, $l \in \mathbb{N}$, and $p \ge l$,*

$$
\left\| \left(Z - AM\frac{\log(2N)}{n} \right)_+ \right\|_l \le \left(2\left(\frac{2}{A} \right)^{p-1} + (l!)^{1/l}\sqrt{\frac{2}{n}} + \frac{1}{A} + \frac{lA}{n} \right) M.
$$

Proof. The idea is again to separate the bounded and the unbounded quantities. The part with the unbounded quantities is treated by elementary means and Lemma 5.5. For the bounded part, we use Lemma 5.6.

First, we assume w.l.o.g. that $M = 1$ and set $K = \frac{A}{2} + \sqrt{\frac{A^2}{4} - 1}$. Then, we deduce with the triangle inequality that

$$\begin{aligned}
&\left\| \left(\max_{1 \le j \le N} \mathbb{P}_n Z(j) - A \frac{\log(N)}{n} \right)_+ \right\|_l \\
&= \left\| \left(\max_{1 \le j \le N} (\mathbb{P}_n - P) Z(j) - A \frac{\log(N)}{n} \right)_+ \right\|_l \\
&\le \left\| \left(\max_{1 \le j \le N} (\mathbb{P}_n - P) \overline{Z}(j) + \max_{1 \le j \le N} (\mathbb{P}_n - P) \underline{Z}(j) - A \frac{\log(N)}{n} \right)_+ \right\|_l \qquad (12) \\
&\le \left\| \left(\max_{1 \le j \le N} (\mathbb{P}_n - P) \overline{Z}(j) \right)_+ + \left(\max_{1 \le j \le N} (\mathbb{P}_n - P) \underline{Z}(j) - A \frac{\log(N)}{n} \right)_+ \right\|_l \\
&\le \left\| \left(\max_{1 \le j \le N} (\mathbb{P}_n - P) \overline{Z}(j) \right)_+ \right\|_l + \left\| \left(\max_{1 \le j \le N} (\mathbb{P}_n - P) \underline{Z}(j) - A \frac{\log(N)}{n} \right)_+ \right\|_l .
\end{aligned}$$

So, we are able to treat the unbounded and the bounded quantities separately. We begin with the unbounded quantities. We first note that

$$\left[(\mathbb{P}_n - P) \overline{Z}(j) \right]_+^l \le \left((\mathbb{P}_n + P) \overline{\mathcal{E}} \right)^l = \left((\mathbb{P}_n - P) \overline{\mathcal{E}} + 2P\overline{\mathcal{E}} \right)^l .$$

Hence,

$$\left\| \left(\max_{1 \le j \le p} (\mathbb{P}_n - P) \overline{Z}(j) \right)_+ \right\|_l \le 2P\overline{\mathcal{E}} + \left\| (\mathbb{P}_n - P) \overline{\mathcal{E}} \right\|_l . \qquad (13)$$

Hölder's and Chebyshev's inequalities are then used to find

$$P\overline{\mathcal{E}} = \frac{1}{n} \sum_{i=1}^n \mathbb{E}\overline{\mathcal{E}}_i \le \frac{1}{n} \sum_{i=1}^n \left(\mathbb{E}\mathcal{E}_i^p \right)^{1/p} (\mathbb{E} 1_{\{\mathcal{E}_i > K\}})^{1-1/p} \le \frac{1}{K^{p-1}} . \qquad (14)$$

To bound the left over quantity, we note that for all i and $p \ge q \in \mathbb{N}$

$$\mathbb{E}[\overline{\mathcal{E}}_i - \mathbb{E}\overline{\mathcal{E}}_i]^q \le 2^q$$

so that

$$\mathbb{E}\left[(\overline{\mathcal{E}}_{i_1} - \mathbb{E}\overline{\mathcal{E}}_{i_1}) \cdots (\overline{\mathcal{E}}_{i_l} - \mathbb{E}\overline{\mathcal{E}}_{i_l}) \right] \le 2^l .$$

Moreover, it holds that

$$\mathbb{E}\left[(\overline{\mathcal{E}}_{i_1} - \mathbb{E}\overline{\mathcal{E}}_{i_1}) \cdots (\overline{\mathcal{E}}_{i_l} - \mathbb{E}\overline{\mathcal{E}}_{i_l}) \right] = 0$$

for all $i_1, \dots, i_l$ such that there is a j with $i_j \ne i_{j'}$ for all $j' \ne j$. With Lemma 5.5, we then get for $n > 1$

$$\mathbb{E}\left[(\mathbb{P}_n - P) \overline{\mathcal{E}} \right]^l \le \frac{2^l C_l^n}{n^l} \le \frac{2^l l!}{n^l} \left(\frac{n}{2} \right)^{\lfloor l/2 \rfloor} \le l! \sqrt{\frac{2}{n}}^{\,l} . \qquad (15)$$

Clearly, this also holds for $n = 1$ and $l = 1$. For $n = 1$ and $l > 1$, we note that

$$\mathbb{E}\big[(\mathbb{P}_n - P)\overline{\mathcal{E}}\big]^l \le 2^l \le l!\sqrt{2}^l,$$

so that inequality (15) holds for all n and l under consideration. Inserting then inequalities (14) and (15) in inequality (13), we obtain the result for the unbounded part

$$\Big\|\Big(\max_{1\le j\le p}(\mathbb{P}_n - P)\overline{Z}(j)\Big)_+\Big\|_l \le \frac{2}{K^{p-1}} + (l!)^{1/l}\sqrt{\frac{2}{n}}. \tag{16}$$

Next, we plug the result of Lemma 5.6 and inequality (16) in inequality (12) to derive

$$\left\|\left(\max_{1\le j\le N}\mathbb{P}_n Z(j) - A\frac{\log(N)}{n}\right)_+\right\|_l \le 2\left(\frac{2}{A}\right)^{p-1} + (l!)^{1/l}\sqrt{\frac{2}{n}} + \frac{1}{A} + \frac{lA}{n}.$$

Finally, we define $Z(j+N) := -Z(j)$ for $1 \le j \le N$. We then get

$$\begin{aligned}\left\|\left(Z - A\frac{\log(2N)}{n}\right)_+\right\|_l &= \left\|\left(\max_{1\le j\le 2N}\mathbb{P}_n Z(j) - A\frac{\log(2N)}{n}\right)_+\right\|_l \\ &\le 2\left(\frac{2}{A}\right)^{p-1} + (l!)^{1/l}\sqrt{\frac{2}{n}} + \frac{1}{A} + \frac{lA}{n}\end{aligned}$$

replacing N by $2N$ in the results above. □

Theorem 4.2 is now a simple corollary.

Proof of Theorem 4.2. Set $A = 2\sqrt{n}$ in Lemma 5.7. □

Acknowledgements

We acknowledge partial financial support as members of the German-Swiss research group FOR916 (Regularization and Qualitative Constraints) with Grant number 20PA20E-134495/1. We also thank Michaël Chichignoud and Mohamed Hebiri for their valuable comments. Finally, we thank the editor and the referees for their helpful suggestions.

References

[1] Adamczak, R. (2008). A tail inequality for suprema of unbounded empirical processes with applications to Markov chains. *Electron. J. Probab.* **13** 1000–1034. MR2424985

[2] Boucheron, S., Lugosi, G. and Massart, P. (2013). *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford: Oxford Univ. Press.

[3] Bousquet, O. (2002). A Bennett concentration inequality and its application to suprema of empirical processes. *C. R. Math. Acad. Sci. Paris* **334** 495–500. MR1890640

[4] Bühlmann, P. and van de Geer, S. (2011). *Statistics for High-dimensional Data*: *Methods*, *Theory and Applications*. *Springer Series in Statistics*. Heidelberg: Springer. MR2807761
[5] Dudley, R.M. (1967). The sizes of compact subsets of Hilbert space and continuity of Gaussian processes. *J*. *Funct*. *Anal*. **1** 290–330. MR0220340
[6] Dümbgen, L., van de Geer, S.A., Veraar, M.C. and Wellner, J.A. (2010). Nemirovski's inequalities revisited. *Amer*. *Math*. *Monthly* **117** 138–160. MR2590193
[7] Fernique, X. (1975). Regularité des trajectoires des fonctions aléatoires gaussiennes. In *École D'Été de Probabilités de Saint-Flour IV*-1974. *Lecture Notes in Math*. **480** 1–96. Berlin: Springer. MR0413238
[8] Klein, T. and Rio, E. (2005). Concentration around the mean for maxima of empirical processes. *Ann*. *Probab*. **33** 1060–1077. MR2135312
[9] Ledoux, M. (1995/97). On Talagrand's deviation inequalities for product measures. *ESAIM Probab*. *Statist*. **1** 63–87 (electronic). MR1399224
[10] Ledoux, M. and Talagrand, M. (2011). *Probability in Banach Spaces*: *Isoperimetry and Processes*. *Classics in Mathematics*. Berlin: Springer. Reprint of the 1991 edition. MR2814399
[11] Massart, P. (2000). About the constants in Talagrand's concentration inequalities for empirical processes. *Ann*. *Probab*. **28** 863–884. MR1782276
[12] Massart, P. (2007). *Concentration Inequalities and Model Selection*. *Lecture Notes in Math*. **1896**. Berlin: Springer. Lectures from the 33rd Summer School on Probability Theory held in Saint-Flour, July 6–23, 2003, With a foreword by Jean Picard. MR2319879
[13] Rio, E. (2002). Une inégalité de Bennett pour les maxima de processus empiriques. *Ann*. *Inst*. *Henri Poincaré Probab*. *Stat*. **38** 1053–1057. MR1955352
[14] Talagrand, M. (1996). Majorizing measures: The generic chaining. *Ann*. *Probab*. **24** 1049–1103. MR1411488
[15] Talagrand, M. (1996). New concentration inequalities in product spaces. *Invent*. *Math*. **126** 505–563. MR1419006
[16] Talagrand, M. (2005). *The Generic Chaining*: *Upper and Lower Bounds of Stochastic Processes*. *Springer Monographs in Mathematics*. Berlin: Springer. MR2133757
[17] van de Geer, S. and Lederer, J. (2013). The Bernstein–Orlicz norm and deviation inequalities. *Probab*. *Theory Related Fields* **157** 225–250.
[18] van der Vaart, A. and Wellner, J.A. (2011). A local maximal inequality under uniform entropy. *Electron*. *J*. *Stat*. **5** 192–203. MR2792551
[19] van der Vaart, A.W. and Wellner, J.A. (2000). *Weak Convergence and Empirical Processes*: *With Applications to Statistics*. *Springer Series in Statistics*. New York: Springer. MR1385671
[20] Viens, F.G. and Vizcarra, A.B. (2007). Supremum concentration inequality and modulus of continuity for sub-nth chaos processes. *J*. *Funct*. *Anal*. **248** 1–26. MR2329681

Received February 2012 and revised May 2013