DFG-SNF Research Group FOR916

Statistical Regularization and Qualitative Constraints

Johannes Lederer

Sara van de Geer

New Concentration Inequalities for Suprema of Empirical Processes

Preprint FOR916 11-7

Updated Version (15-11-2011)

Preprint-Series of the Research Group FOR916

FOR916/11-7

November 2011

NEW CONCENTRATION INEQUALITIES FOR SUPREMA OF EMPIRICAL PROCESSES

By Johannes C. Lederer*, and Sara A. van de Geer

ETH Zurich

Abstract 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.

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 $X_1, ..., X_n$ and a countable function class $\mathcal{F}$ such that $\sup_{f\in\mathcal{F}} \|f\|_\infty \leq 1$ and $\sup_{f\in\mathcal{F}} |\mathbb{E}f(X_1)| = 0$. The quantity of interest is then $Y := \sup_{f\in\mathcal{F}} |\frac{1}{n}\sum_{i=1}^n f(X_i)|$. Bousquet derives in [2] with $\sigma_Y^2 := \sup_{f\in\mathcal{F}} \operatorname{Var} f(X_1)$ and $\nu := \sigma_Y^2 + 2\mathbb{E}[Y]$ the exponential deviation inequality

$$\mathbb{P}\left(Y - EY \geq \sqrt{2x\nu} + \frac{x}{3}\right) \leq e^{-nx} \quad \text{for all } x > 0.$$

Bousquet's proof is a refinement of Rio's proof in [8] and relies on the entropy method (see for example [7, Chapter 5.3]). Similar exponential bounds

*The authors acknowledge partial financial support as members of the German-Swiss Research Group FOR916 (Statistical Regularization and Qualitative Constraints) with grant number 20PA20E-134495/1.

AMS 2000 subject classifications: Primary 60E15; secondary 60F10

Keywords and phrases: Central Limit Theorem, Chaining, Concentration Inequalities, Deviation Inequalities, Empirical Processes, Rate of Convergence

for bounded empirical processes have been found by Klein and Rio [5] and by Massart [6]. These bounds are slightly less sharp, but additionally hold for not necessarily identically distributed random variables and also for -Y. We finally mention Talagrand's work [10] that probably provided the spark for the development in this field.

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

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

for Y as above. Other exponential bounds for unbounded empirical processes are given by Adamczak in [1] and by van de Geer and Lederer in [12]. These authors assume very 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 $X_1, ..., X_n$ and a countable function class $\mathcal{F}$ with an envelope that has pth moment at most M^p for a $p \in [1, \infty)$. Our main result, Theorem 3.1, implies then for the quantity of interest $Y := \sup_{f\in\mathcal{F}} |\frac{1}{n}\sum_{i=1}^n (f(X_i) - \mathbb{E}f(X_i))|$, and for $1 \le l \le p$, σ_Y as above, and $(\cdot)_+^l := (\max\{0, \cdot\})^l$

$$\mathbb{E}\left[Y - (1+\epsilon)\mathbb{E}Y\right]_+^l \le \left(\left(\frac{64}{\epsilon} + 7 + \epsilon\right)\frac{lM}{n^{1-\frac{l}{p}}} + \frac{4\sqrt{l}\sigma_Y}{\sqrt{n}}\right)^l \quad \text{for all } \epsilon > 0$$

and

$$\mathbb{E}\left[(1-\epsilon)\mathbb{E}Y - Y\right]_+^l \le \left(\left(\frac{86.4}{\epsilon} + 7 - \epsilon\right)\frac{lM}{n^{1-\frac{l}{p}}} + \frac{4.7\sqrt{l}\sigma_Y}{\sqrt{n}}\right)^l \quad \text{for } \epsilon \in (0, 1].$$

We argue in Section 3 that this result is especially useful in the common case where the envelope (measured by M) is much larger than the single functions (measured by σ_Y).

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*

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

for large n. Here, $X_1, ..., X_n$ are independent, not necessarily identically distributed random variables and $\{f : f \in \mathcal{F}\}$ 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 $\mathcal{Z}_1, ..., \mathcal{Z}_n$ be arbitrary probability spaces and $\{Z_i(j) : \mathcal{Z}_i \to \mathbb{R}, 1 \leq j \leq N, 1 \leq i \leq n\}$ a set of random variables. We then define the random vectors as $Z(j) := (Z_1(j), ..., Z_n(j))^T : \mathcal{Z}_1 \times ... \times \mathcal{Z}_n \to \mathbb{R}^n$. For convenience, we introduce their mean as $PZ(j) := \frac{1}{n} \sum_{i=1}^{n} \mathbb{E} Z_i(j)$ and their empirical mean as $\mathbb{P}_n Z(j) := \frac{1}{n} \sum_{i=1}^{n} Z_i(j)$. Throughout this paper, we then consider the generalized formulation of (1)

$$(2) \qquad Z := \max_{1 \leq j \leq N} |\mathbb{P}_n Z(j)|.$$

The corresponding results for the empirical processes (1) 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 $\mathcal{F} = \{f_1, ..., f_N\}$.

Concentration inequalities are a standard tool to characterize the behavior of the process (2) (and thus of (1)). For $n \to \infty$, the process (2) is typically governed by the central limit theorem. In contrast, concentration inequalities bound the deviation in both directions from the mean or related quantities for finite n. Similarly, *deviation inequalities* bound the deviation in one direction only. Concentration or deviation inequalities - contrarily to maximal inequalities for example - provide bounds that depend only on n and moment properties of an envelope and the single functions f (and particularly not on N).

Let us finally express the moment properties of the envelope. First, we call $\mathcal{E} := (\mathcal{E}_1, ..., \mathcal{E}_n)^T : \mathcal{Z}_1 \times ... \times \mathcal{Z}_n \to \mathbb{R}^n$ an *envelope* if $|Z_i(j)| \leq \mathcal{E}_i$ for all $1 \leq j \leq N$ and $1 \leq i \leq n$. The basic assumption of this paper is then that there is a $p \in [1, \infty)$ and an $M > 0$ such that

$$\mathbb{E}\mathcal{E}_i^p \leq M^p \tag{3}$$

for all $1 \leq i \leq n$. Typically, the envelope is much larger than the single random vectors. In these cases, we have $M \gg \sigma := \sqrt{\max_{1 \leq j \leq N} \frac{1}{n} \sum_{i=1}^n \mathbb{E} Z_i(j)^2}$.

3. Main Result. We are mainly concerned with concentration inequalities for unbounded empirical processes that only fulfill weak moment conditions. Additionally, we want to 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 $\epsilon \in \mathbb{R}^+$ it holds that*

$$\mathbb{E}\left[Z - (1+\epsilon)\mathbb{E}Z\right]_+^l \leq \left(\left(\frac{64}{\epsilon} + 7 + \epsilon \right) \frac{lM}{n^{1-\frac{l}{p}}} + \frac{4\sqrt{l}\sigma}{\sqrt{n}} \right)^l$$

and, if additionally $\epsilon \in (0, 1]$,

$$\mathbb{E}\left[(1-\epsilon)\mathbb{E}Z - Z\right]_+^l \leq \left(\left(\frac{86.4}{\epsilon} + 7 - \epsilon \right) \frac{lM}{n^{1-\frac{l}{p}}} + \frac{4.7\sqrt{l}\sigma}{\sqrt{n}} \right)^l.$$

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 relinquished slightly better constants to obtain incisive bounds, however, a considerable improvement in l does not seem to be possible. 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 for example [3], [13], [14]) or generic chaining (see for example [9], [11]).

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 $\frac{1}{n^{1-\frac{l}{p}}} \ll \frac{1}{\sqrt{n}}$ 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.

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

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

and similarly

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

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

$$\begin{aligned} \mathbb{E}\left[Z - 2\mathbb{E}Z\right]_+ &\leq 72\frac{M}{n^{1-\frac{1}{p}}} + 4\frac{\sigma}{\sqrt{n}} \\ \textit{and} \quad \mathbb{E}\left[\frac{1}{2}\mathbb{E}Z - Z\right]_+ &\leq 179.3\frac{M}{n^{1-\frac{1}{p}}} + 4.7\frac{\sigma}{\sqrt{n}}. \end{aligned}$$

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:

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

$$\mathbb{E}\left[Z - 4\mathbb{E}Z\right]_+^l \leq l\ \Gamma\left(\frac{l}{2}\right)\left(\frac{32}{n}\right)^{\frac{l}{2}} M^l,$$

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 $(1+\epsilon)$ 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 4.1 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 [4]). Another bound is found by an approach detailed in Section 5. The bound reads:

THEOREM 4.2. *Let the random variables $Z_i(j)$ be centered. Then, for $p \geq 2$, $l \in \mathbb{N}$, and $p \geq l$*

$$\mathbb{E}\left[Z - 2M\frac{\log(2N)}{\sqrt{n}}\right]_+^l \leq \left(\frac{35l^2}{n}\right)^{\frac{l}{2}} M^l.$$

This can supersede the bound in Theorem 4.1 for $2\sqrt{\log(2N)} \leq 8\sqrt{2}$.

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 [6]. 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 two simple auxiliary results, we then turn to the main task of this section: We prove Lemma 5.3, a generalization of Theorem 3.1. The main result of this paper, Theorem 3.1, is then an easy consequence.

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

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

Similarly, we define

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

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

$$\overline{Z} := \max_{1 \leq j \leq N} |\mathbb{P}_n \overline{Z}(j)| \text{ and } \underline{Z} := \max_{1 \leq j \leq N} |\mathbb{P}_n \underline{Z}(j)|.$$

Now we derive two simple auxiliary lemmas.

LEMMA 5.1. *Let $l \geq 1$, $W_i : \mathcal{Z}_i \to \mathbb{R}_0^+$ for $1 \leq i \leq n$ and $\mathbb{E}W_i^l \leq 1$. Then,*

$$\mathbb{E}\left[\mathbb{P}_n W\right]^l \leq 1$$

for the corresponding random vector W on the product space.

PROOF. By the triangle inequality we have

$$\left(\mathbb{E}\left[\mathbb{P}_n W\right]^l\right)^{\frac{1}{l}} = \frac{1}{n}\left(\mathbb{E}\left[\sum_{i=1}^n W_i\right]^l\right)^{\frac{1}{l}} \leq \frac{1}{n}\sum_{i=1}^n \left(\mathbb{E}\left[W_i^l\right]\right)^{\frac{1}{l}} \leq 1.$$

□

LEMMA 5.2. *Under the assumptions of Theorem 3.1 it holds for $K = n^{\frac{l}{p}} M$ that*

$$|\mathbb{E}[\underline{Z} - Z]| \leq \frac{M}{n^{l(1-\frac{1}{p})}}$$

and $\underline{\sigma} \leq \sigma$ for

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

PROOF. Since $||a|-|b|| \le |a-b|$ for all $a, b \in \mathbb{R}$, it holds that

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

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

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

These two results yield then the first assertion. The second assertion is straightforward. □

We can now turn to the harder part of this section. The following lemma is a generalization of Theorem 3.1. The derivation of Theorem 3.1 from this and Lemma 5.2 is then a simple task.

LEMMA 5.3. *Let $1 \le l \le p$ and $\epsilon, K \in \mathbb{R}^+$. Then,*

$$\mathbb{E}\left[Z - (1+\epsilon)\mathbb{E}\underline{Z}\right]_+^l \le \left(\left(\frac{64}{\epsilon}+5\right)\frac{lK}{n} + \frac{4\sqrt{l}\underline{\sigma}}{\sqrt{n}} + \frac{M^{\frac{p}{l}}}{K^{\frac{p}{l}-1}}\right)^l$$

and

$$\mathbb{E}\left[(1-\epsilon)\mathbb{E}\underline{Z} - Z\right]_+^l \le \left(\left(\frac{86.4}{\epsilon}+5\right)\frac{lK}{n} + \frac{4.7\sqrt{l}\underline{\sigma}}{\sqrt{n}} + \frac{M^{\frac{p}{l}}}{K^{\frac{p}{l}-1}}\right)^l.$$

PROOF. The key idea of the proof is to separate the bounded from the unbounded quantities. On the one hand, we develop bounds for $\mathbb{E}\overline{Z}_+^l$ via elementary means. On the other hand, we develop bounds for $\mathbb{E}\left[\underline{Z} - (1+\epsilon)\mathbb{E}\underline{Z}\right]_+^l$ and $\mathbb{E}\left[(1-\epsilon)\mathbb{E}\underline{Z} - \underline{Z}\right]_+^l$ via convexity arguments and [6, Theorem 4]. They can then be combined to deduce the result.
We start with the proof of the first inequality. First, we split Z in a bounded and an unbounded part

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

and deduce with the triangle inequality that

$$\begin{aligned} &\mathbb{E}\left[Z - (1+\epsilon)\mathbb{E}\underline{Z}\right]_+^l \\ \le &\mathbb{E}\left[\underline{Z} + \overline{Z} - (1+\epsilon)\mathbb{E}\underline{Z}\right]_+^l \\ \le &\mathbb{E}\left[(\underline{Z} - (1+\epsilon)\mathbb{E}\underline{Z})_+ + \overline{Z}_+\right]^l \\ \le &\left((\mathbb{E}\left[\underline{Z} - (1+\epsilon)\mathbb{E}\underline{Z}\right]_+^l)^{\frac{1}{l}} + (\mathbb{E}\overline{Z}_+^l)^{\frac{1}{l}}\right)^l. \end{aligned} \tag{4}$$

Now, we turn to the development of bounds for $\mathbb{E}\overline{Z}_+^l$. With the help of Hölder's and Chebyshev's Inequalities we obtain as above for $1 \le i \le n$

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

That is,

$$\mathbb{E}\left[\frac{\overline{\mathcal{E}}_i}{(\frac{M^p}{K^{p-l}})^{\frac{1}{l}}}\right]^l \le 1.$$

We may consequently apply Lemma 5.1 and obtain

$$\mathbb{E}\overline{Z}_+^l = \mathbb{E}\overline{Z}^l \le \mathbb{E}[\mathbb{P}_n\overline{\mathcal{E}}]^l \le \frac{M^p}{K^{p-l}}. \tag{5}$$

As a next step, we derive bounds for $\mathbb{E}\left[\underline{Z} - (1+\epsilon)\mathbb{E}\underline{Z}\right]_+^l$. To begin, we set $J := (\frac{32}{\epsilon} + 2.5)K$, $\sigma := \underline{\sigma}$ and define the function $g_l : \mathbb{R}^+ \to (1, \infty)$ as

$$g_l(x) := \exp\left(\frac{\sqrt{n}(\sqrt{2\sigma^2 + Jx^{\frac{1}{l}}} - \sqrt{2}\sigma)}{\sqrt{2}J}\right)^2.$$

Note that g_l is strictly increasing and smooth. Moreover, the first and second derivatives are

$$\begin{aligned} g_l'(x) =& \frac{2n(\sqrt{2\sigma^2 + Jx^{\frac{1}{l}}} - \sqrt{2}\sigma)}{4J^2\sqrt{2\sigma^2 + Jx^{\frac{1}{l}}}} \frac{J}{l} x^{\frac{1}{l}-1} g_l(x) \\ =& \frac{n}{2lJ}\left(1 - \frac{\sqrt{2}\sigma}{\sqrt{2\sigma^2 + Jx^{\frac{1}{l}}}}\right) x^{\frac{1}{l}-1} g_l(x) \end{aligned}$$

and

$$\begin{aligned} g_l''(x) =& \frac{n^2}{4l^2J^2}\left(1 - \frac{\sqrt{2}\sigma}{\sqrt{2\sigma^2 + Jx^{\frac{1}{l}}}}\right)^2 x^{\frac{2}{l}-2} g_l(x) \\ &+ \frac{n}{2lJ}\frac{\sqrt{2}\sigma}{2(2\sigma^2 + Jx^{\frac{1}{l}})^{\frac{3}{2}}} \frac{J}{l} x^{\frac{2}{l}-2} g_l(x) \\ &+ \frac{n}{2lJ}\left(1 - \frac{\sqrt{2}\sigma}{\sqrt{2\sigma^2 + Jx^{\frac{1}{l}}}}\right)\left(\frac{1}{l} - 1\right) x^{\frac{1}{l}-2} g_l(x) \\ \geq& \frac{nx^{\frac{2}{l}-2} g_l(x)}{4l^2J}\left(1 - \frac{\sqrt{2}\sigma}{\sqrt{2\sigma^2 + Jx^{\frac{1}{l}}}}\right)\left(\left(1 - \frac{\sqrt{2}\sigma}{\sqrt{2\sigma^2 + Jx^{\frac{1}{l}}}}\right)\frac{n}{J} + 2(1-l)x^{-\frac{1}{l}}\right). \end{aligned}$$

We now use the lower bound for the second derivative to find an interval, on which the function g_l is convex. To this end, we observe that for $\sigma > 0$

$$\left(1 - \frac{\sqrt{2}\sigma}{\sqrt{2\sigma^2 + Jx^{\frac{1}{l}}}}\right)\frac{n}{J} + 2(1-l)x^{-\frac{1}{l}} \geq 0 \tag{6}$$

is equivalent to

$$1 - \frac{1}{\sqrt{1 + \frac{J}{2\sigma^2}x^{\frac{1}{l}}}} \geq \frac{2(l-1)J}{n} x^{-\frac{1}{l}}.$$

This can be rewritten with the definition $u := \frac{J}{2\sigma^2}x^{\frac{1}{l}} > 0$ as

$$1 - \frac{1}{\sqrt{1+u}} \geq \frac{(l-1)J^2}{n\sigma^2 u}$$

and with the definition $C := \frac{(l-1)J^2}{n\sigma^2} \geq 0$ as

$$1 - \frac{1}{\sqrt{1+u}} \geq \frac{C}{u}. \tag{7}$$

We assume now, that $u \geq C$. Then,

$$\begin{aligned} & 1-\frac{1}{\sqrt{1+u}} \geq \frac{C}{u} \\ \Leftrightarrow \quad & \sqrt{1+u}\left(1-\frac{C}{u}\right) \geq 1 \\ \Leftrightarrow \quad & (1+u)(u^2-2Cu+C^2) \geq u^2 \\ \Leftrightarrow \quad & u^3-2Cu^2+(C^2-2C)u+C^2 \geq 0. \end{aligned}$$

Considering the equality

$$u^2-2Cu+C^2-2C=0$$

with roots $\{C \pm \sqrt{2C}\}$, we deduce that

$$u^3-2Cu^2+(C^2-2C)u+C^2 \geq 0$$

for all $u \geq C+\sqrt{2C}$. Consequently, for $u \geq C+\sqrt{2C}$, Inequality (7) holds true. Hence, if we postulate

$$\frac{J}{2\sigma^2}x^{\frac{1}{l}} \geq \frac{(l-1)J^2}{n\sigma^2}+\sqrt{\frac{2(l-1)J^2}{n\sigma^2}} \tag{8}$$

Equation (6) holds true. The postulate (8) is equivalent to

$$x^{\frac{1}{l}} \geq \frac{2(l-1)J}{n}+\frac{\sqrt{8(l-1)}\sigma}{\sqrt{n}}$$

and to

$$x \geq \left(\frac{2(l-1)J}{n}+\frac{\sqrt{8(l-1)}\sigma}{\sqrt{n}}\right)^l =: I.$$

Additionally, note that with this condition on x, Equation (6) is also true for $\sigma=0$. So we finally derived, since $\frac{nx^{\frac{2}{l}-2}g_l(x)}{4l^2J}(1-\frac{\sqrt{2}\sigma}{\sqrt{2\sigma^2+Jx^{\frac{1}{l}}}})$ is positive, that the function g_l is convex on the domain (I,∞).

This convexity property makes it possible to apply a result of [6]. To show this, we introduce

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

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

$$g_l(\mathbb{E}X) \leq g_l(\mathbb{E}[X \vee I]) \leq \mathbb{E}g_l(X \vee I).$$

We used here the notation $a \vee b := \max\{a,b\}$ for $a,b \in \mathbb{R}$. Massart's Inequality [6, Theorem 4, (13)] for bounded random vectors translates then to our setting as

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

where $\epsilon, x > 0$. This is equivalent to

$$\mathbb{P}(\underline{Z} \geq (1+\epsilon)\mathbb{E}\underline{Z} + \sigma\sqrt{\frac{8x}{n}} + \frac{J}{n}x) \leq e^{-x}. \tag{9}$$

We then deduce (cf. [12])

$$\begin{aligned}
&\mathbb{E}\exp\left(\frac{\sqrt{n}\left(\sqrt{2\sigma^2 + J(X\vee I)^{\frac{1}{l}}} - \sqrt{2}\sigma\right)}{\sqrt{2}J}\right)^2\\
=&\int_0^\infty \mathbb{P}\left(\exp\left(\frac{\sqrt{n}\left(\sqrt{2\sigma^2 + J(X\vee I)^{\frac{1}{l}}} - \sqrt{2}\sigma\right)}{\sqrt{2}J}\right)^2 > t\right)dt\\
\leq& 1 + \int_1^\infty \mathbb{P}\left(\exp\left(\frac{\sqrt{n}\left(\sqrt{2\sigma^2 + J(X\vee I)^{\frac{1}{l}}} - \sqrt{2}\sigma\right)}{\sqrt{2}J}\right)^2 > t\right)dt\\
=& 1 + \int_1^\infty \mathbb{P}\left(\sqrt{2\sigma^2 + J(X\vee I)^{\frac{1}{l}}} > \sqrt{2}\sigma + \sqrt{\frac{2J^2}{n}\log t}\right)dt\\
=& 1 + \int_1^\infty \mathbb{P}\left(J(X\vee I)^{\frac{1}{l}} > 4\sigma\sqrt{\frac{J^2}{n}\log t} + \frac{2J^2}{n}\log t\right)dt
\end{aligned}$$

and note that

$$\begin{aligned}
& JI^{\frac{1}{l}} < 4\sigma\sqrt{\frac{J^2}{n}\log t} + \frac{2J^2}{n}\log t\\
\Leftrightarrow\quad & \frac{2(l-1)J}{n} + \frac{\sqrt{8(l-1)}\sigma}{\sqrt{n}} < 4\sigma\sqrt{\frac{\log t}{n}} + \frac{2J}{n}\log t.
\end{aligned}$$

This is fulfilled if $t \geq e^{l-1}$. Hence, with Massart's Inequality (9),

$$\begin{aligned}
&\mathbb{E}\exp\left(\frac{\sqrt{n}\left(\sqrt{2\sigma^2+J(X\vee I)^{\frac{1}{l}}}-\sqrt{2}\sigma\right)}{\sqrt{2}J}\right)^2\\
\leq&1+e^{l-1}-1+\int_{e^{l-1}}^{\infty}\mathbb{P}\left(X^{\frac{1}{l}}>4\sigma\sqrt{\frac{\log t}{n}}+\frac{2J}{n}\log t\right)dt\\
=&e^{l-1}+\int_{e^{l-1}}^{\infty}\mathbb{P}\left(\underline{Z}>(1+\epsilon)\mathbb{E}\underline{Z}+4\sigma\sqrt{\frac{\log t}{n}}+\frac{2J}{n}\log t\right)dt\\
\leq&e^{l-1}+\int_{e^{l-1}}^{\infty}\exp(-\log t^2)dt<e^l.
\end{aligned}$$

In summary, we have

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

This is now inverted, observing that for $y\in(1,\infty)$ such that

$$\begin{aligned}
&&y=\exp\left(\frac{\sqrt{n}(\sqrt{2\sigma^2+Jx^{\frac{1}{l}}}-\sqrt{2}\sigma)}{\sqrt{2}J}\right)^2\\
\Rightarrow&&\sqrt{\log y}=\frac{\sqrt{n}(\sqrt{2\sigma^2+Jx^{\frac{1}{l}}}-\sqrt{2}\sigma)}{\sqrt{2}J}\\
\Rightarrow&&\frac{\sqrt{2}J}{\sqrt{n}}\sqrt{\log y}+\sqrt{2}\sigma=\sqrt{2\sigma^2+Jx^{\frac{1}{l}}}\\
\Rightarrow&&\frac{2J^2}{n}\log y+\frac{4\sigma J}{\sqrt{n}}\sqrt{\log y}=Jx^{\frac{1}{l}}\\
\Rightarrow&&\left(\frac{2J}{n}\log y+\frac{4\sigma}{\sqrt{n}}\sqrt{\log y}\right)^l=x.
\end{aligned}$$

This is now applied with $x=\mathbb{E}X$ to obtain

$$\mathbb{E}X\leq\left(\frac{2lJ}{n}+\frac{4\sqrt{l}\sigma}{\sqrt{n}}\right)^l. \tag{10}$$

We are now ready to collect the terms. Inequalities (4), (5) and (10) give

$$\mathbb{E}\left[Z-(1+\epsilon)\mathbb{E}\underline{Z}\right]_+^l=\left(\frac{2lJ}{n}+\frac{4\sqrt{l}\sigma}{\sqrt{n}}+\frac{M^{\frac{p}{l}}}{K^{\frac{p}{l}-1}}\right)^l.$$

Hence, since $J = (\frac{32}{\epsilon} + 2.5)K$,

$$\mathbb{E}\left[Z - (1+\epsilon)\mathbb{E}\underline{Z}\right]_+^l \leq \left(\left(\frac{64}{\epsilon} + 5\right)\frac{lK}{n} + \frac{4\sqrt{l}\sigma}{\sqrt{n}} + \frac{M^{\frac{p}{l}}}{K^{\frac{p}{l}-1}}\right)^l.$$

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

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

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

$$\begin{aligned}&\mathbb{E}\left[(1-\epsilon)\mathbb{E}\underline{Z} - Z\right]_+^l \\ \leq&\mathbb{E}\left[(1-\epsilon)\mathbb{E}\underline{Z} - \underline{Z} + \overline{Z}\right]_+^l \\ \leq&\mathbb{E}\left[((1-\epsilon)\mathbb{E}\underline{Z} - \underline{Z})_+ + \overline{Z}_+\right]^l \\ =&\left(\mathbb{E}[(1-\epsilon)\mathbb{E}\underline{Z} - \underline{Z}]_+^l)^{\frac{1}{l}} + (\mathbb{E}\overline{Z}_+^l)^{\frac{1}{l}}\right)^l.\end{aligned}$$

One can then proceed as in the first part and use [6, Theorem 4, (14)]. □

Proof of Theorem 3.1. Set $K = n^{\frac{l}{p}} M$ in Lemma 5.3 and use Lemma 5.2 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 Massart [6]. The proof here is considerably shorter than the proof of Theorem 3.1. This is because we do not need tedious convexity arguments.

Proof of Theorem 4.1. The trick is to use symmetrization and desymmetrization arguments so that we are able to use [6, Theorem 9] in a favorable way.

Beforehand, we define $Z_\epsilon := \max_{1\leq j\leq N} |\frac{1}{n}\sum_{i=1}^n \epsilon_i Z_i(j)|$ with independent Rademacher random variables ϵ_i. Then, we symmetrize according to [13, Lemma 2.3.6] with the function $\Phi(x) = (x - 4\mathbb{E}Z)_+^l$ to obtain

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

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

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

Hence,

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

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

$$\begin{aligned}
&\mathbb{E}_\epsilon\left[Z_\epsilon - \mathbb{E}Z_\epsilon\right]_+^l \\
=&\int_0^\infty \mathbb{P}_\epsilon\left(\left(Z_\epsilon - \mathbb{E}Z_\epsilon\right)_+^l > t\right) dt \\
=&\int_0^\infty \mathbb{P}_\epsilon\left(Z_\epsilon > \mathbb{E}Z_\epsilon + t^{\frac{1}{l}}\right) dt \\
\leq&\int_0^\infty \mathbb{P}_\epsilon\left(\max_{1\leq j\leq N}\frac{1}{n}\sum_{i=1}^n \epsilon_i Z_i(j) > \mathbb{E}\max_{1\leq j\leq N}\frac{1}{n}\sum_{i=1}^n \epsilon_i Z_i(j) + t^{\frac{1}{l}}\right) dt \\
&+\int_0^\infty \mathbb{P}_\epsilon\left(\max_{1\leq j\leq N} -\frac{1}{n}\sum_{i=1}^n \epsilon_i Z_i(j) > \mathbb{E}\max_{1\leq j\leq N}\frac{1}{n}\sum_{i=1}^n \epsilon_i Z_i(j) + t^{\frac{1}{l}}\right) dt \\
\leq&2\int_0^\infty \mathbb{P}_\epsilon\left(\max_{1\leq j\leq N}\frac{1}{n}\sum_{i=1}^n \epsilon_i Z_i(j) > \mathbb{E}\max_{1\leq j\leq N}\frac{1}{n}\sum_{i=1}^n \epsilon_i Z_i(j) + t^{\frac{1}{l}}\right) dt.
\end{aligned}$$

In a final step, we apply Massart's Inequality [6, Theorem 9] with

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

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

$$\begin{aligned}
&2\int_0^\infty \mathbb{P}_\epsilon\left(\max_{1\leq j\leq N}\frac{1}{n}\sum_{i=1}^n \epsilon_i Z_i(j) > \mathbb{E}\max_{1\leq j\leq N}\frac{1}{n}\sum_{i=1}^n \epsilon_i Z_i(j) + t^{\frac{1}{l}}\right) dt \\
\leq&2\int_0^\infty \exp\left(-\frac{nt^{\frac{2}{l}}}{8\mathbb{P}_n\mathcal{E}^2}\right) dt \\
=&2\left(\frac{8}{n}\right)^{\frac{l}{2}}\left(\mathbb{P}_n\mathcal{E}^2\right)^{\frac{l}{2}}\int_0^\infty \exp\left(-t^{\frac{2}{l}}\right) dt \\
=&2\left(\frac{8}{n}\right)^{\frac{l}{2}}\left(\mathbb{P}_n\mathcal{E}^2\right)^{\frac{l}{2}}\frac{l\Gamma\left(\frac{l}{2}\right)}{2}.
\end{aligned}$$

With Inequality (11) this gives

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

Finally, because of Lemma 5.1, it holds that

$$\mathbb{E}\left[\mathbb{P}_n\mathcal{E}^2\right]^{\frac{l}{2}} \leq \mathbb{E}\left[\mathbb{P}_n\mathcal{E}\right]^l \leq M^l$$

and hence

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

□

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,*

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

PROOF. We follow well known ideas (see e.g. [3, Chapter 14]):

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

□

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

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

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

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

By induction, this yields

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

We now assume that m is even. So,

$$\begin{aligned} C_m^{n+1} \leq& m! \left[\left(\frac{n}{2}\right)^{\frac{m}{2}} + \frac{1}{2!} \left(\frac{n}{2}\right)^{\frac{m}{2}-1} + \frac{1}{3!} \left(\frac{n}{2}\right)^{\frac{m}{2}-2} + ... + \frac{1}{(m-2)!} \left(\frac{n}{2}\right) \right] + 1 \\ =& m! \left[\left(\frac{n}{2}\right)^{\frac{m}{2}} + \frac{1}{2!} \left(\frac{n}{2}\right)^{\frac{m}{2}-1} + \sum_{j=2}^{\frac{m}{2}-1} \left(\frac{1}{(2j-1)!} + \frac{1}{(2j)!} \right) \left(\frac{n}{2}\right)^{\frac{m}{2}-j} \right] + 1 \\ \leq& m! \left[\left(\frac{n}{2}\right)^{\frac{m}{2}} + \frac{m}{4} \left(\frac{n}{2}\right)^{\frac{m}{2}-1} + \sum_{j=2}^{\frac{m}{2}-1} \binom{\frac{m}{2}}{j} \left(\frac{1}{2}\right)^j \left(\frac{n}{2}\right)^{\frac{m}{2}-j} + \left(\frac{1}{2}\right)^{\frac{m}{2}} \right] \\ =& m! \sum_{j=0}^{\frac{m}{2}} \binom{\frac{m}{2}}{j} \left(\frac{1}{2}\right)^j \left(\frac{n}{2}\right)^{\frac{m}{2}-j} = m! \left(\frac{n+1}{2}\right)^{\lfloor \frac{m}{2} \rfloor}. \end{aligned}$$

This completes the proof for $m > 2$ with m even. We note finally, that for odd $m > 2$ we have $C_m^n < m C_{m-1}^n \leq m! \left(\frac{n}{2}\right)^{\lfloor \frac{m}{2} \rfloor}$. □

We now settle the bounded part of the problem. Bounded random variables are in particular subexponential, so one could apply results from [15] for example. But for our purposes, a direct treament as in the following seems to be more suitable.

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

$$\mathbb{E}\left[\max_{1\leq j\leq N}(\mathbb{P}_n - P)\underline{Z}(j) - AM\frac{\log(N)}{n}\right]_+^l \leq \left(\frac{M}{A} + \frac{lAM}{n}\right)^l.$$

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

$$\mathbb{E}\left[\underline{Z}_i(j) - \mathbb{E}\underline{Z}_i(j)\right]^2 \leq \mathbb{E}\underline{Z}_i(j)^2 \leq 1.$$

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

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

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

$$\begin{aligned}
&\mathbb{E}e^{\frac{n(\mathbb{P}_n - P)\underline{Z}(j)}{A}}\\
=&\mathbb{E}e^{\frac{\sum_{i=1}^n(\underline{Z}_i(j) - \mathbb{E}\underline{Z}_i(j))}{A}}\\
\leq&\left(1 + \frac{1}{A^2}\right)^n.
\end{aligned}$$

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

$$\begin{aligned}
&\mathbb{E}\left[\max_{1\leq j\leq N}(\mathbb{P}_n - P)\underline{Z}(j) - A\frac{\log(N)}{n}\right]_+^l\\
\leq&\frac{A^l}{n^l}\mathbb{E}\left[\left(\max_{1\leq j\leq N} n(\mathbb{P}_n - P)\underline{Z}(j)/A - \log(N)\right)_+^l \vee (l-1)^l\right]\\
\leq&\frac{A^l}{n^l}\log^l\left(\mathbb{E}\exp\left(\left(\max_{1\leq j\leq N} n(\mathbb{P}_n - P)\underline{Z}(j)/A - \log(N)\right)_+ \vee (l-1)\right)\right)\\
=&\frac{A^l}{n^l}\log^l\left(\mathbb{E}\exp\left(\left(\max_{1\leq j\leq N} n(\mathbb{P}_n - P)\underline{Z}(j)/A - \log(N)\right) \vee (l-1)\right)\right)\\
\leq&\frac{A^l}{n^l}\log^l\left(\max_{1\leq j\leq N}\mathbb{E}\exp(n(\mathbb{P}_n - P)\underline{Z}(j)/A) + e^{l-1}\right)\\
\leq&\frac{A^l}{n^l}\log^l\left(\left(1 + \frac{1}{A^2}\right)^n + e^{l-1}\right).
\end{aligned}$$

We finally note that $a+b<eab$ for all $a,b\geq 1$ and find

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

□

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 $l\in\mathbb{N}$, $p\geq 2$, $p\geq l$ and $A\geq 2$,*

$$\mathbb{E}\left[Z-AM\frac{\log(2N)}{n}\right]_+^l\leq\left(2\left(\frac{2}{A}\right)^{p-1}+(l!)^{\frac{1}{l}}\sqrt{\frac{2}{n}}+\frac{1}{A}+\frac{lA}{n}\right)^l M^l.$$

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

$$\begin{aligned}&\mathbb{E}\left[\max_{1\leq j\leq N}\mathbb{P}_nZ(j)-A\frac{\log(N)}{n}\right]_+^l\\=&\mathbb{E}\left[\max_{1\leq j\leq N}(\mathbb{P}_n-P)Z(j)-A\frac{\log(N)}{n}\right]_+^l\\\leq&\mathbb{E}\left[\max_{1\leq j\leq N}(\mathbb{P}_n-P)\overline{Z}(j)+\max_{1\leq j\leq N}(\mathbb{P}_n-P)\underline{Z}(j)-A\frac{\log(N)}{n}\right]_+^l\\\leq&\mathbb{E}\left[\left[\max_{1\leq j\leq N}(\mathbb{P}_n-P)\overline{Z}(j)\right]_++\left[\max_{1\leq j\leq N}(\mathbb{P}_n-P)\underline{Z}(j)-A\frac{\log(N)}{n}\right]_+\right]^l\end{aligned}$$

(12)

$$\leq\left(\left(\mathbb{E}\left[\max_{1\leq j\leq N}(\mathbb{P}_n-P)\overline{Z}(j)\right]_+^l\right)^{\frac{1}{l}}+\left(\mathbb{E}\left[\max_{1\leq j\leq N}(\mathbb{P}_n-P)\underline{Z}(j)-A\frac{\log(N)}{n}\right]_+^l\right)^{\frac{1}{l}}\right)^l.$$

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

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

Hence,

$$\mathbb{E}\left[\max_{1\leq j\leq p}(\mathbb{P}_n - P)\overline{Z}(j)\right]_+^l \leq \left(2P\overline{\mathcal{E}} + (\mathbb{E}\left[(\mathbb{P}_n - P)\overline{\mathcal{E}}\right]^l)^{\frac{1}{l}}\right)^l. \tag{13}$$

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

$$P\overline{\mathcal{E}} = \frac{1}{n}\sum_{i=1}^n \mathbb{E}\overline{\mathcal{E}}_i \leq \frac{1}{n}\sum_{i=1}^n (\mathbb{E}\mathcal{E}_i^p)^{\frac{1}{p}}(\mathbb{E}1_{\{\mathcal{E}_i > K\}})^{1-\frac{1}{p}} \leq \frac{1}{K^{p-1}}. \tag{14}$$

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

$$\mathbb{E}\left[\overline{\mathcal{E}}_i - \mathbb{E}\overline{\mathcal{E}}_i\right]^q \leq 2^q$$

so that

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

Moreover, it holds that

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

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

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

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

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

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

$$\mathbb{E}\left[\max_{1\leq j\leq p}(\mathbb{P}_n - P)\overline{Z}(j)\right]_+^l \leq \left(\frac{2}{K^{p-1}} + (l!)^{\frac{1}{l}}\sqrt{\frac{2}{n}}\right)^l \tag{16}$$

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

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

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

$$\mathbb{E}\left[Z - A\frac{\log(2N)}{n}\right]_+^l = \mathbb{E}\left[\max_{1\le j\le 2N}\mathbb{P}_n Z(j) - A\frac{\log(2N)}{n}\right]_+^l$$
$$\le\left(2\left(\frac{2}{A}\right)^{p-1} + (l!)^{\frac{1}{l}}\sqrt{\frac{2}{n}} + \frac{1}{A} + \frac{lA}{n}\right)^l$$

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. □

Acknowledgments. We thank Michaël Chichignoud and Mohamed Hebiri for their valuable comments.

References.

[1] Adamczak, R. (2008). A tail inequality for suprema of unbounded empirical processes with applications to Markov chains. *Electron. J. Probab.* **13** no. 34, 1000–1034.
[2] Bousquet, O. (2002). A Bennett concentration inequality and its application to suprema of empirical processes. *C. R. Math. Acad. Sci. Paris* **334** 495–500.
[3] Bühlmann, P. and van de Geer, S. (2011). *Statistics for high-dimensional data. Springer Series in Statistics*. Springer, Heidelberg.
[4] 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.
[5] Klein, T. and Rio, E. (2005). Concentration around the mean for maxima of empirical processes. *Ann. Probab.* **33** 1060–1077.
[6] Massart, P. (2000). About the constants in Talagrand's concentration inequalities for empirical processes. *Ann. Probab.* **28** 863–884.
[7] Massart, P. (2007). *Concentration Inequalities and Model Selection: Ecole d'Eté de Probabilités de Saint-Flour XXXIII-2003*. Springer.
[8] Rio, E. (2002). Une inégalité de Bennett pour les maxima de processus empiriques. *Ann. Inst. H. Poincaré Probab. Statist.* **38** 1053–1057.
[9] Talagrand, M. (1996). Majorizing measures: The generic chaining. *Ann. Probab.* **24** 1049–1103.

[10] Talagrand, M. (1996). New concentration inequalities in product spaces. *Invent. Math.* **126** 505–563.

[11] Talagrand, M. (2005). *The Generic Chaining: Upper and Lower Bounds of Stochastic Processes*. Springer.

[12] van de Geer, S. A. and Lederer, J. C. (2011). The Bernstein-Orlicz norm and deviation inequalities. *preprint*.

[13] van der Vaart, A. W. and Wellner, J. A. (2000). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer.

[14] van der Vaart, A. W. and Wellner, J. A. (2011). A local maximal inequality under uniform entropy. *Electron. J. Statist.* 5 192-203.

[15] 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.

Johannes C. Lederer
Seminar für Statistik, ETH Zürich
HG G 10.2
Rämistrasse 101
8092 Zürich, Switzerland
E-mail: lederer@stat.math.ethz.ch

Sara A. van de Geer
Seminar für Statistik, ETH Zürich
HG G 24.1
Rämistrasse 101
8092 Zürich, Switzerland
E-mail: geer@stat.math.ethz.ch