Sharp tail distribution estimates for the supremum of a class of sums of i.i.d. random variables

Péter Major
Alfréd Rényi Institute of Mathematics, Hungarian Academy of Sciences

Summary. *We take a class of functions $\mathcal{F}$ with polynomially increasing covering numbers on a measurable space $(X,\mathcal{X})$ together with a sequence of i.i.d. X-valued random variables $\xi_1,\dots,\xi_n$, and give a good estimate on the tail behaviour of $\sup\limits_{f\in\mathcal{F}}\sum\limits_{j=1}^n f(\xi_j)$ if the relations $\sup\limits_{x\in X}|f(x)|\le 1$, $Ef(\xi_1)=0$ and $Ef(\xi_1)^2<\sigma^2$ hold with some $0\le\sigma\le 1$ for all $f\in\mathcal{F}$. Roughly speaking this estimate states that under some natural conditions the above supremum is not much larger than the largest element taking part in it. The proof heavily depends on the main result of paper [6]. We also present an example that shows that our results are sharp, and compare them with results of earlier papers.*

Introduction.

The main result of this paper is an estimate about the tail-distribution of the supremum of sums of i.i.d. random variables presented in Theorem 1 together with an extension of it that provides an estimate for this tail-distribution in some cases not covered in Theorem 1. At first glance these results may look rather complicated, but as I try to explain in Section 2 they yield sharp estimates under natural conditions. They express such a fact that under some natural conditions we can get an almost as good bound for the supremum of an appropriately defined class of random sums as for one term taking part in this supremum. Before presenting these results I recall the definition of uniform covering numbers and classes of functions with polynomially increasing covering numbers, since they appear in the formulation of our results. Here I define the notion of uniform covering numbers, unlike to [6], with respect to all L_p-norms, $p\ge 1$, because in some arguments I shall apply it for $p=2$ and not for $p=1$.

Definition of uniform covering numbers with respect to L_p-norms. Let a measurable space $(X,\mathcal{X})$ be given together with a class of measurable, real valued functions $\mathcal{F}$ on this space. The uniform covering number of this class of functions at level ε, $\varepsilon>0$, with respect to the L_p-norm, $1\le p<\infty$, is $\sup\limits_{\nu}\mathcal{N}(\varepsilon,\mathcal{F},L_p(\nu))$, where the supremum is taken for all probability measures ν on the space $(X,\mathcal{X})$, and $\mathcal{N}(\varepsilon,\mathcal{F},L_p(\nu))$ is the smallest integer m for which there exist some functions $f_j\in\mathcal{F}$, $1\le j\le m$, such that $\min\limits_{1\le j\le m}\int|f-f_j|^p\,d\nu\le\varepsilon^p$ for all $f\in\mathcal{F}$.

Definition of a class of functions with polynomially increasing covering numbers. We say that a class of functions $\mathcal{F}$ has polynomially increasing covering numbers with parameter D and exponent L if the inequality

$$\sup_{\nu}\mathcal{N}(\varepsilon,\mathcal{F},L_1(\nu))\le D\varepsilon^{-L} \tag{1.1}$$

holds for all $0 < \varepsilon \le 1$ with the number $\sup_\nu \mathcal{N}(\varepsilon, \mathcal{F}, L_1(\nu))$ introduced in the previous definition with parameter $p = 1$.

Theorem 1 yields the following estimate.

Theorem 1. *Let a sequence of independent, identically distributed random variables $\xi_1, \dots, \xi_n$, $n \ge 2$, be given with values in a measurable space $(X, \mathcal{X})$ with some distribution μ together with a countable class of functions $\mathcal{F}$ on the same space $(X, \mathcal{X})$ which has polynomially increasing covering numbers with parameter $D \ge 1$ and exponent $L \ge 1$. Let the class of functions $\mathcal{F}$ satisfy also the relations $\sup_{x \in X} |f(x)| \le 1$, $\int f(x)\mu(\,dx) = 0$, and $\int f^2(x)\mu(\,dx) \le \sigma^2$ with some number $0 \le \sigma^2 \le 1$ for all $f \in \mathcal{F}$. Define the normalized random sums $S_n(f) = \frac{1}{\sqrt{n}} \sum_{j=1}^n f(\xi_j)$ for all $f \in \mathcal{F}$. There are some universal constants $C_j > 0$, $1 \le j \le 5$, (such that also the inequality $C_2 < 1$ holds), for which the inequality*

$$P\left(\sup_{f \in \mathcal{F}} |S_n(f)| \ge v\right) \le C_1 e^{-C_2 \sqrt{n} v \log(v/\sqrt{n}\sigma^2)} \quad \textit{for all } v \ge u(\sigma) \tag{1.2}$$

holds if one of the following conditions is satisfied.

(a) $\sigma^2 \le \frac{1}{n^{400}}$, *and* $u(\sigma) = \frac{C_3}{\sqrt{n}}(L + \frac{\log D}{\log n})$,

(b) $\frac{1}{n^{400}} < \sigma^2 \le \frac{\log n}{8n}$, *and* $u(\sigma) = \frac{C_4}{\sqrt{n}}\left(L \frac{\log n}{\log(\frac{\log n}{n\sigma^2})} + \log D\right)$,

(c) $\frac{\log n}{8n} < \sigma^2 \le 1$, *and* $u(\sigma) = \frac{C_5}{\sqrt{n}}(n\sigma^2 + L \log n + \log D)$.

I complete the result of Theorem 1 with an extension which almost agrees with Theorem 4.1 in [5]. It yields an estimate for $P\left(\sup_{f \in \mathcal{F}} |S_n(f)| \ge v\right)$ in cases not covered in Theorem 1. In case (c) it enlarges the set of levels v for which a good estimate can be given for the probability at the left-hand side of (1.2). I discuss this result to give a more complete solution of the problem discussed in Theorem 1. Besides, it may be interesting to understand what kind of tools are applied in its proof.

Extension of Theorem 1. *Let us consider, similarly to Theorem 1, a sequence of independent, identically distributed random variables $\xi_1, \dots, \xi_n$, $n \ge 2$, with values in a measurable space $(X, \mathcal{X})$ with some distribution μ together with a countable class of functions $\mathcal{F}$ on the same space $(X, \mathcal{X})$, which has polynomially increasing covering numbers with parameter $D \ge 1$ and exponent $L \ge 1$, and such that $\sup_{x \in X} |f(x)| \le 1$, $\int f(x)\mu(\,dx) = 0$, and $\int f^2(x)\mu(\,dx) \le \sigma^2$ with some number $0 \le \sigma^2 \le 1$ for all $f \in \mathcal{F}$. The supremum of the normalized sums $S_n(f)$, $f \in \mathcal{F}$, introduced in Theorem 1 satisfies the inequality*

$$P\left(\sup_{f \in \mathcal{F}} |S_n(f)| \ge v\right) \le C \exp\left\{-\alpha \frac{v^2}{\sigma^2}\right\} \tag{1.3}$$

with appropriate (universal) constants $\alpha>0$, $C>0$ and $C_6>0$ if $\frac{\log n}{8n}<\sigma^2\le 1$, and $\sqrt n\sigma^2\ge v\ge \bar u(\sigma)$, where $\bar u(\sigma)$ is defined as

$$\bar u(\sigma)=C_6\sigma\left(L^{3/4}\log^{1/2}\frac{2}{\sigma}+(\log D)^{1/2}\right).$$

The value $\frac{\log n}{8n}$ determining the boundary between cases (b) and (c) in Theorem 1 could be replaced by $\alpha\frac{\log n}{n}$ with an arbitrary number $0<\alpha<1$. To see this one has to check that the formula defining $u(\sigma)$ in cases (b) and (c) give a value of the same order if $\sigma^2\sim\alpha\frac{\log n}{n}$ with $0<\alpha<1$. I chose the parameter $\alpha=\frac18$ because some calculations were simpler with such a choice. Let me remark that a similar statement holds for the value of boundary n^{-200} between cases (a) and (b). This could have been replaced by $n^{-\beta}$ with any $\beta>1$.

In Section 2 I try to explain why the above results are natural. I present an example which shows that Theorem 1 and its Extension are sharp. There are models satisfying the conditions of these results for which these results would not hold any longer if we replaced the functions $u(\sigma)=u(\sigma,n)$ or $\bar u(\sigma)$ by a much smaller function. More explicitly, they would become invalid if we replaced the function $u(\sigma,n)$ by a function $v(\sigma,n)$ such that $\lim\limits_{n\to\infty}\frac{v(n,\sigma)}{u(n,\sigma)}=0$, or $\bar u(\sigma)$ by a function $\bar v(\sigma)$ such that $\lim\limits_{\sigma\to0}\frac{\bar v(\sigma)}{\bar u(\sigma)}=0$. (Because of the condition $\bar u(\sigma)\le v\le\sqrt n\sigma^2$ in the Extension of Theorem 1 the value of $\bar u(\sigma)$ for small values σ is interesting only in the case of large sample size n.) In Section 3 I present the proof of Theorem 1 and its Extension. In Section 4 I discuss the content of these results in more detail. I explain the main problems and ideas behind them, and I also make a comparison with the results of earlier works.

2. Discussion on the conditions of these results.

We defined with the help of a sequence of i.i.d. random variables $\xi_1,\dots,\xi_n$ on a measurable space $(X,\mathcal X)$ and a class of functions $\mathcal F$ with polynomially increasing covering numbers on the same space $(X,\mathcal X)$ the random sums $S_n(f)$ for all $f\in\mathcal F$, and wanted to give a good estimate on the tail distribution $P_n(v)=P\left(\sup\limits_{f\in\mathcal F}|S_n(f)|>v\right)$ of the supremum of these random sums at all levels $v>0$ if the conditions $\sup\limits_{x\in X}|f(x)|\le 1$, $Ef(\xi_1)=0$ and $Ef^2(\xi_1)\le\sigma^2$ hold with some number $0\le\sigma\le1$ for all functions $f\in\mathcal F$. In particular, we were interested in the dependence of this estimate on the number σ. In this section I discuss the sharpness of our results, and present an example that indicates that the estimates given in Theorem 1 and in its Extension are sharp.

Although I gave an estimate for the supremum of a class of random sums defined with the help of a class of functions $\mathcal F$ which has polynomially increasing covering numbers with arbitrary exponent L and parameter D, I was mainly interested in the special case when these parameters L and D are bounded, more precisely when they have a bound not depending on the parameter σ. In this case the functions $u(\sigma)$ and

$\bar u(\sigma)$ in Theorem 1 and in its Extension have a simpler form. Namely, we can choose $u(\sigma)=\frac{C_3}{\sqrt n}$ in case (a), $u(\sigma)=\frac{C_4}{\sqrt n}\frac{\log n}{\log\frac{\log n}{n\sigma^2}}$ in case (b), $u(\sigma)=C_5\sqrt n\sigma^2$ in case (c), and $\bar u(\sigma)=C_6\sigma\log^{1/2}\frac2\sigma$. In this section I present an example that indicates that our results are sharp in this case. I shall call the estimates in these results sharp, because only the value of the universal constants appearing in them can be improved. I do not try to find the optimal value of these constants, but I want to present such an example where these estimates cannot be improved in any other respect. In particular, I shall show that the estimates in this example do not hold any longer if we replace the coefficients C_j in the definition of the quantities $u(\sigma)$ and $\bar u(\sigma)$ with very small positive constant, because after such a replacement the probabilities $P_n(v)$ would be very close to one at level $v=u(\sigma)$ or $v=\bar u(\sigma)$ at large sample size n. I shall consider the following example.

Example. Take a sequence of independent, uniformly distributed random variables $\xi_1,\dots,\xi_n$ on the unit interval $[0,1]$, fix a number $0\le\sigma^2\le1$, and define a class of functions $\mathcal F_\sigma$ and $\bar{\mathcal F}_\sigma$ with functions defined on the unit interval $[0,1]$ in the following way. $\mathcal F_\sigma=\{f_1,\dots,f_k\}$, and $\bar{\mathcal F}=\{\bar f_1,\dots,\bar f_k\}$ with $k=k(\sigma)=[\frac1{\sigma^2}]$, where $[\cdot]$ denotes integer part, and $\bar f_j(x)=\bar f_j(x|\sigma)=1$ if $x\in[(j-1)\sigma^2,j\sigma^2)$, $\bar f_j(x)=\bar f_j(x|\sigma)=0$ if $x\notin[(j-1)\sigma^2,j\sigma^2)$, $1\le j\le k$, and $f_j(x)=f_j(x|\sigma)=\bar f_j(x)-\sigma^2$, $1\le j\le n$.

It can be seen that $\mathcal F_\sigma$ satisfies the conditions of Theorem 1 with $Ef(\xi_j)\le\sigma^2$ for all $f\in\mathcal F_\sigma$. In particular, it has polynomially increasing covering numbers with such a parameter D and exponent L that can be bounded by numbers not depending on σ^2. This can be seen directly, but it is also a consequence of some classical results by which the indicator functions of a Vapnik–Červonenkis class of sets constitute a class of functions with polynomially increasing covering numbers. (See e.g. Theorem 5.2 in [5]). I shall show that the sequence of random variables $\xi_1,\dots,\xi_n$ and class of functions $\mathcal F$ in the above example have the following property.

An estimate on the function $P_n(v)$ in the models of the above Example. *A number $\bar C>0$ can be chosen in such a way that for all $\delta>0$ there is an index $n_0(\delta)$ such that for all sample sizes $n\ge n_0(\delta)$ and numbers $0\le\sigma\le1$ the inequality*

$$P_n(\hat u(\sigma))=P\left(\sup_{f\in\mathcal F_\sigma}|S_n(f)|\ge\hat u(\sigma)\right)\ge1-\delta, \tag{2.1}$$

holds with $\hat u(\sigma)=\frac{\bar C}{\sqrt n}$ in case (a), i.e. if $\sigma^2\le n^{-400}$, $\hat u(\sigma)=\frac{\bar C}{\sqrt n}\frac{\log n}{\log(\frac{\log n}{n\sigma^2})}$ in case (b), i.e. if $n^{-400}<\sigma^2\le\frac{\log n}{8n}$, and $\hat u(\sigma)=\bar C\sigma\log^{1/2}\frac2\sigma$ in case (c) i.e. if $\frac{\log n}{8n}\le\sigma^2\le1$.

In Theorem 1 and in its Extension we gave a good estimate for $P_n(v)$ if $v\ge u(\sigma)$ in cases (a) and (b), and $v\ge\bar u(\sigma)$ in case (c), while in formula (2.1) I claimed that there are such models satisfying the conditions of these results for which no good estimate holds for $P_n(\hat u(\sigma))$, if we define the function $\hat u(\sigma)$ by replacing the coefficients C_j by a sufficiently small constant $\bar C$ in their definition. Let me recall that here we restricted our attention to the case when the exponent L and parameter D of the polynomially increasing covering numbers of the class of functions $\mathcal{F}_\sigma$ we considered are bounded by a constant not depending on the parameter σ. Actually, in the case (c) we have to explain the estimate on $P_n(v)$ in more detail. In this case we have to compare the estimates given by Theorem 1 and its Extension.

It may happen that $\sqrt{n}\sigma^2 \ge \bar{u}(\sigma)$, and in this case the estimate (1.3) of the Extension of Theorem 1 is an empty statement. I claim that in this case we can replace the condition $v \ge u(\sigma)$ by the condition $v \ge \bar{u}(\sigma)$ in case (c) of Theorem 1 with an appropriate constant C_6 in the definition of $\bar{u}(\sigma)$, and Theorem 1 remains valid after such a modification. To show this it is enough to check that $u(\sigma)$ and $\bar{u}(\sigma)$ have the same order of magnitude in this case, i.e. there are universal constants $C' > 0$ and $C'' > 0$ such that $C'\bar{u}(\sigma) \le u(\sigma) \le C''\bar{u}(\sigma)$.

We have $\bar{u}(\sigma) \le \sqrt{n}\sigma^2 = \frac{1}{C_5}u(\sigma)$ in this case, which implies the first inequality. On the other hand, $\sigma^2 \ge \frac{\log n}{8n}$ in case (c), hence as some calculation shows $\bar{u}(\sigma) = C_6\sigma \log^{1/2}\frac{2}{\sigma} \ge \text{const.}\sqrt{n}\sigma^2$. This implies the second inequality.

If $\sqrt{n}\sigma^2 \ge \bar{u}(\sigma)$, and the estimate (1.3) is not an empty statement, then we can give a good estimate for $P_n(v)$ for all $v \ge \bar{u}(\sigma)$, i.e. also if $u(\sigma) \ge v \ge \sqrt{n}\sigma^2$, in a case which was covered neither in Theorem 1 nor in its Extension. In this case we have $C_6\sqrt{n}\sigma^2 \ge v$, and we can write the following inequality by means of relation (1.3) with the choice $v = \sqrt{n}\sigma^2$.

$$P\left(\sup_{f\in\mathcal{F}} |S_n(f)| \ge v\right) \le P\left(\sup_{f\in\mathcal{F}} |S_n(f)| \ge \sqrt{n}\sigma^2\right) \le Ce^{-\alpha n\sigma^2} \le Ce^{-\bar{\alpha}v^2/\sigma^2},$$

with some $\bar{\alpha} > 0$, i.e. relation (1.3) holds (with a possible different parameter $\bar{\alpha} > 0$) for all $u(\sigma) \ge v \ge \bar{u}(\sigma)$.

To understand the content of Theorem 1 and its Extension together with the estimate on the function $P_n(v)$ in the models of the Example formulated above it may be useful to recall a result called the concentration inequality for the supremum of sums of i.i.d. random variables. (See e.g. [11]). It states that there is a concentration point of the tail distribution of the supremum of sums of i.i.d. random variables. This concentration point has the property that the supremum is strongly concentrated in a small neighbourhood of it. I do not formulate this result in a more precise and detailed form, because we need it here only for the sake of some orientation. The problem with its application is that this result determines the concentration point only in an implicit way, as the expected value of the supremum we are investigating, and we cannot calculate it explicitly in the general case. On the other hand, the concentration inequality implies that we can get a good, non-trivial estimate for the tail distribution of the supremum of sums of i.i.d. random variables only at levels higher than their concentration point. (Otherwise we cannot give a better estimate for the tail distribution than the trivial upper bound 1.) The number $u(\sigma)$ defined in Theorem 1 is an upper bound for the concentration point in cases (a) and (b), while the number $\bar{u}(\sigma)$ defined the Extension of Theorem 1 is an upper bound for it in the case (c). On the other hand, the number $\hat{u}(\sigma)$ introduced in formula (2.1) is a lower bound for the concentration point in the models introduced in the Example. So in this case we have determined the concentration point up to a (universal) multiplying constant.

Thus the functions $u(\sigma)$ and $\bar{u}(\sigma)$ can be considered as good upper bounds on the concentration point of the supremum of the random sums $S_n(f)$, $f \in \mathcal{F}$, if the conditions of Theorem 1 and its Extension are satisfied. In formulas (1.2) and (1.3) we also gave an estimate on the function $P_n(v)$, i.e. on the tail distribution of the supremum we are investigating for $v \ge u(\sigma)$ or $v \ge \bar{u}(\sigma)$. We can say that this estimate is also sharp, we cannot get a better bound (if we disregard the value of the universal constants C_1, C_2) in formula (1.2) and C and α in formula (1.3) even if we took a single normalized sum $S_n(f)$ whose terms $f(\xi_j)$ have variance $Ef(\xi_j)^2 = \sigma^2$.

Indeed, if we disregard the value of the universal constants appearing in our estimates, then we can say that formula (1.2) yields such an estimate for the tail distribution $P_n(v)$ as Bennett's inequality yields for the tail distribution of a single term $S_n(f)$ if the terms in this normalized sum have variance σ^2. At least this is the case if we consider the estimate of Bennett's inequality at level $v \ge 2\sqrt{n}\sigma^2$. (This follows e.g. from formula (3.3) in this paper. Here we recalled Bennett's inequality, and formula (3.3) is a part of it.) On the other hand, we considered in Theorem 1 only such levels v where this condition is satisfied, since $u(\sigma) \ge 2\sqrt{n}\sigma$ in all cases of Theorem 1. Moreover, there are examples that show that inequality (3.3) is sharp, we cannot get a better estimate without some additional restrictions. (See Example 3.3 in [5]). In inequality (1.3) we gave a Gaussian type upper bound, and this is also a sharp estimate if we disregard the value of the absolute constants in it.

To complete this section we still have to show that the model introduced in the Example satisfies formula (2.1). This will be done in the following proof.

The proof of the estimate on the function $P_n(v)$ formulated about the models in the Example of this Section. In the proof of relation (2.1) we introduce the following notation. Define the empirical distribution function $F_n(x)$ of the random variables $\xi_1, \dots, \xi_n$, i.e. put

$$F_n(x) = \frac{1}{n}\{\text{the number of indices } j,\ 1 \le j \le n,\ \text{such that } \xi_j < x\}$$

for all $0 < x \le 1$, and take its normalization $G_n(x) = \sqrt{n}(F_n(x) - x)$, $0 < x \le 1$. Observe that

$$\left\{ \sup_{f \in \mathcal{F}_\sigma} |S_n(f)| \ge \hat{u}(\sigma) \right\} = \left\{ \max_{1 \le j \le k(\sigma)} |G_n(j\sigma^2) - G_n((j-1)\sigma^2)| \ge \hat{u}(\sigma) \right\}. \qquad (2.2)$$

By a classical results of probability theory, the normalized empirical distribution functions converge weakly to the Brownian bridge as $n \to \infty$. In our next considerations it will be also interesting that the modulus of continuity of a Brownian bridge, (which actually agrees with the modulus of continuity of a Wiener process) can be also calculated, (see e.g. [8]). By a similar, but simpler calculation we can estimate the probability of the event we get by replacing the normalized empirical distribution function $G_n(\cdot)$ by a Brownian bridge in the right-hand side expression of (2.2). This is actually done with the choice $\hat{u}(\sigma) = \bar{C}\sigma \log^{1/2} \frac{2}{\sigma}$ in the fourth chapter of [5] (page 27), and it is shown that this probability is almost one for large parameters n for all $\sigma > 0$ if the coefficient $\bar{C}$ of $\hat{u}(\sigma)$ is chosen sufficiently small. (Actually we have to choose $\bar{C} < \sqrt{2}$.) Let us call this estimate the Gaussian version of formula (2.1). At a heuristic level this result together with formula (2.2) and the weak convergence of the normalized empirical processes $G_n(\cdot)$ to a Brownian bridge suggests that formula (2.1) should hold with $\hat{u}(\sigma) = \bar{C}\sigma \log^{1/2} \frac{2}{\sigma}$ and a small coefficient $\bar{C} > 0$.

This heuristic argument is nevertheless misleading, since the weak convergence of the empirical processes $G_n(\cdot)$ to the Brownian bridge in itself does not allow to carry out a limiting procedure that leads to formula (2.1). On the other hand, a stronger version of the weak convergence of the normalized empirical processes (see [4]) yields a useful result in this direction. This result states that a normalized empirical process $G_n(x)$ and a Brownian bridge $B(x)$, $0 \le x \le 1$, can be constructed in such a way that $\sup\limits_{0\le x\le 1} |B(x)-G_n(x)| \le K\frac{\log n}{\sqrt{n}}$ for all $n \ge 2$ and sufficiently large $K > 0$ with probability almost 1. This result together with the Gaussian version of formula (2.1) imply the validity of formula (2.1) if $\sigma^2 \ge B\frac{\log n}{2n}$ with a sufficiently large $B > 0$. Indeed, in this case $\hat{u}(\sigma) \ge 2K\frac{\log n}{\sqrt{n}}$, hence the Gaussian version of formula of (2.1) together with the result of [4] imply that

$$P\left(\max_{1\le j\le k(\sigma)} |G_n(j\sigma^2) - G_n((j-1)\sigma^2)| \ge \frac{\hat{u}(\sigma)}{2}\right) \ge 1-\delta$$

if $\sigma^2 \ge B\frac{\log n}{n}$, and $n \ge n_0(\delta)$, hence inequality (2.1) holds in this case if we choose $\bar{C} > 0$ sufficiently small in the definition of $\hat{u}(\sigma)$. Moreover, this relation also holds for all $\sigma^2 \ge \frac{\log n}{8n}$, i.e. in the case (c) if we choose $\hat{u}(\sigma) = \bar{C}\sigma \log^{1/2} \frac{2}{\sigma}$ with a sufficiently small $\hat{C} > 0$. To see this it is enough to observe that if $\max\limits_{1\le j\le k(\sigma)} |G_n(j\sigma^2)-G_n((j-1)\sigma^2)| \le \hat{u}(\sigma)$, then for any positive integers A we have $\max\limits_{1\le j\le k(\sqrt{A}\sigma)} |G_n(j(A\sigma))-G_n((j-1)(A\sigma))| \le A\hat{u}(\sigma)$, and that the corresponding result holds if $\sigma^2 \ge B\frac{\log n}{8n}$.

In cases (a) and (b) the above Gaussian approximation argument does not work. Moreover, inequality (2.1) holds only with a different function $\hat{u}(\sigma)$ in these cases. In case (b) we shall prove formula (2.1) by means of a Poissonian approximation method described below. It can be considered as a more detailed elaboration of the argument in Example 4.3 of [5].

In this argument first we consider the following problem. Take a Poisson process $Z_n(t)$, $0 \le t \le 1$, with parameter n, (i.e. let $EZ_n(t) = nt$ for all $0 \le t \le 1$) in the interval $[0,1]$. Fix some number $0 \le \sigma^2 \le \frac{1}{7}\frac{\log n}{n}$, and define with its help the number $\hat{u}(\sigma) = \hat{u}(\sigma,n) = \frac{3}{4\sqrt{n}}\frac{\log n}{\log(\frac{\log n}{n\sigma^2})}$ and the random variables $\bar{V}_j = \bar{V}_j^{(n)}(\sigma) = Z_n(j\sigma^2)-Z_n((j-1)\sigma^2)$ for $1 \le j \le k$ with $k = k(\sigma) = [\frac{1}{\sigma^2}]$. (Here we defined $\hat{u}(\sigma)$ similarly to the quantity introduced with the same notation at the formulation of inequality (2.1) in the case (b). We only made small modifications. Namely, we considered σ^2 in the interval $[0, \frac{\log n}{7n}]$ instead of the interval $[\frac{1}{n^{200}}, \frac{\log n}{8n}]$, and we fixed the value $\bar{C} = \frac{3}{4}$ in the definition of

$\hat{u}(\sigma)$.) We shall show that for all $\delta > 0$ there is some threshold index $n_0(\delta)$ such that the inequality

$$P\left(\max_{1\le j\le k(\sigma)} \bar V_j^{(n)}(\sigma) \ge \sqrt{n}\hat{u}(\sigma,n)\right) \ge 1-\delta \quad \text{if } n\ge n_0(\delta) \tag{2.3}$$

holds for all $0\le\sigma^2\le\frac17\frac{\log n}{n}$.

To prove this inequality let us first observe that

$$\begin{aligned}P\left(\max_{1\le j\le k(\sigma)} \bar V_j^{(n)}(\sigma) \ge \sqrt{n}\hat{u}(\sigma,n)\right) &\ge P(\bar V_j^{(n)}(\sigma)=\sqrt{n}\hat{u}(\sigma,n) \text{ for some } 1\le j\le k)\\ &= 1-P(\bar V_1^{(n)}(\sigma)\ne\sqrt{n}\hat{u}(\sigma,n))^k,\end{aligned}$$

and

$$\begin{aligned}P(\bar V_1^{(n)}(\sigma)\ne\sqrt{n}\hat{u}(\sigma,n)) &= 1-P(\bar V_1^{(n)}(\sigma)=\sqrt{n}\hat{u}(\sigma,n))\\ &=1-\frac{(n\sigma^2)^{\sqrt{n}\hat{u}(\sigma,n)}}{(\sqrt{n}\hat{u}(\sigma,n))!}e^{-n\sigma^2}\le 1-\left(\frac{n\sigma^2}{\sqrt{n}\hat{u}(\sigma,n)}\right)^{\sqrt{n}\hat{u}(n,\sigma)} e^{-n\sigma^2}.\end{aligned}$$

Since we have $k=[\frac{1}{\sigma^2}]$ we can bound the left-hand side of (2.3) from below as

$$\begin{aligned}&P\left(\max_{1\le j\le k(\sigma)} \bar V_j^{(n)}(\sigma)\sqrt{n}\hat{u}(\sigma,n)\right)\\ &\qquad\ge 1-\left[1-\left(\frac{n\sigma^2}{\sqrt{n}\hat{u}(\sigma,n)}\right)^{\sqrt{n}\hat{u}(n,\sigma)} e^{-n\sigma^2}\right]^{1/\sigma^2}\ge 1-e^{-T}\end{aligned}$$

with $T=\frac{1}{\sigma^2}\left(\frac{n\sigma^2}{\sqrt{n}\hat{u}(\sigma,n)}\right)^{\sqrt{n}\hat{u}(n,\sigma)} e^{-n\sigma^2}$. Hence to prove (2.3) it is enough to show that

$$\left(\frac{n\sigma^2}{\sqrt{n}\hat{u}(\sigma,n)}\right)^{\sqrt{n}\hat{u}(n,\sigma)} \ge \sigma^2 e^{n\sigma^2}\log\frac1\delta \quad \text{if } n\ge n_0(\delta). \tag{2.4)}$$

The right-hand side of (2.4) can be bounded from above as

$$\sigma^2 e^{n\sigma^2}\log\frac1\delta=\frac{\log\frac1\delta}{n}(n\sigma^2)e^{n\sigma^2}\le\frac{\log\frac1\delta}{n}\left(\frac17\log n\right)e^{(\log n)/7}\le n^{-5/6}$$

if $n\ge n_0(\delta)$, since $n\sigma^2\le\frac17\log n$, and $\log\frac1\delta\le n^{1/100}$ for $n\ge n_0(\delta)$. Hence we prove (2.4) if we show that

$$\frac{\sqrt{n}\hat{u}(n,\sigma)}{n\sigma^2}\log\left(\frac{\sqrt{n}\hat{u}(\sigma,n)}{n\sigma^2}\right)\le\frac56\frac{\log n}{n\sigma^2}.$$

By applying the definition of $\hat{u}(n,\sigma)$ and introducing the quantity $z=\frac34\frac{\log n}{n\sigma^2}$ we can rewrite the last inequality as $\frac{z}{\log(\frac{4z}{3})}\log(\frac{z}{\log(\frac{4z}{3})})\le\frac{10}{9}z$, or since $z\ge\frac{21}{4}$ in the case we are investigating it can be rewritten as $\frac19\log\frac{4z}{3}\ge-\log\log\frac{4z}{3}-\log\frac43$ if $z\ge\frac{21}{4}$, and this relation clearly holds. Thus we proved (2.3).

We shall prove relation (2.1) in the case (b) by means of formula (2.3) for a Poisson process with parameter $\frac{99}{100}n$ instead of n and a simple coupling argument between an empirical process and a Poisson process. Namely, we make the following coupling. Let us consider a sequence of independent random variables $\xi_1,\xi_2\dots$ with uniform distribution on the unit interval $[0,1]$ together with a Poissonian random variable $\eta=\eta_n$ with parameter $\frac{99}{100}n$ independent of the random variables ξ_j, $j=1,2,\dots$, and take the first η_n terms of the random variables ξ_j, i.e. the sequence $\xi_1,\xi_2,\dots,\xi_{\eta_n}$ with the random stopping index η_n. In such a way we constructed a Poisson process with parameter $\frac{99}{100}n$, which is smaller than the (non-normalized) empirical distribution of the sequence $\xi_1,\dots,\xi_n$ in the following sense. For large parameter n with probability almost 1 all intervals $[a,b]\subset[0,1]$ contain more points from the sequence $\xi_1,\dots,\xi_n$ than from the above constructed Poisson process. This is a simple consequence of the fact that $P(\eta_n>n)\to0$ as $n\to\infty$.

The above coupling construction and formula (2.3) (with a Poisson process with parameter $\frac{99}{100}$) imply that

$$P\left(\sup_{\bar f\in\bar{\mathcal F}_\sigma}\sqrt n S_n(\bar f)\ge\sqrt{\frac{99}{100}n\hat u\left(\sigma,\frac{99}{100}n\right)}\right)\ge1-\delta\quad\text{if } n\ge n_0(\delta)$$

with the class of functions $\bar{\mathcal F}_\sigma$ introduced before the formulation (2.1) and the function $\hat u(\sigma,n)$ defined in the estimate about the models of the Example in case (b). To complete the proof of (2.1) in the case (b) it is enough to check that the above relation remains valid if the class of functions $\bar{\mathcal F}_\sigma$ is replaced by the class of functions $\mathcal F_\sigma$ and the term $\sqrt{\frac{99}{100}n\hat u(\sigma,\frac{99}{100}n)}$ is replaced by $\hat u(\sigma,n)=\frac{\bar C}{\sqrt n}\frac{\log n}{\log(\frac{\log n}{n\sigma^2})}$ with some appropriate $\bar C>0$. Since the functions $f\in\mathcal F$ are of the form $f(x)=\bar f(x)-\sigma^2$ with some $\bar f\in\mathcal F$, the identity $\sqrt nS_n(f)=\sqrt nS_n(\bar f)-n\sigma^2$ holds, and to prove the desired relation it is enough to check that

$$\sqrt{\frac{99}{100}\frac34\frac{\log n}{\log(\frac{\log n}{\frac{99}{100}n\sigma^2})}}-n\sigma^2\ge\sqrt{\frac{99}{100}\frac34\frac{\log n}{\log(\frac{\log n}{n\sigma^2})}}-n\sigma^2\ge\bar C\frac{\log n}{\log(\frac{\log n}{n\sigma^2})}$$

with some appropriate $\bar C>0$ if $8n\sigma^2\le\log n$. The first inequality clearly holds, and the second inequality is equivalent to the relation

$$\sqrt{\frac{99}{100}\frac34\frac{\frac{\log n}{n\sigma^2}}{\log(\frac{\log n}{n\sigma^2})}}\ge\alpha$$

with some $\alpha>1$. But this relation clearly holds if $8n\sigma^2\le\log n$. Thus we have proved (2.1) also in case (b).

In the case (a) the proof of (2.1) is very simple. It is enough to observe that the sample points ξ_j fall into one of the intervals $[(j-1)\sigma^2, j\sigma^2)$, $1 \le j \le k$, (we disregard the event that they fall into the last interval $[k\sigma^2, 1)$ which has negligibly small probability), hence

$$P\left(\sup_{\bar f \in \bar{\mathcal F}_\sigma} \sqrt{n} S_n(\bar f) \ge 1\right) \ge 1-\delta \quad \text{if } n \ge n_0(\delta),$$

and since σ^2 is very small for large n relation (2.1) holds in case (a) with $\bar C = 1-\varepsilon$ for any $\varepsilon > 0$.

I finish this section with some remarks on the paper [2], about whose existence I learned only after finishing this work. Theorem 4 in Section 2 of that paper contains an almost sure limit theorem on the appropriately normalized supremum of the increase of the empirical distribution functions F_n, $n = 1, 2, \dots$, of a sequence of i.i.d. random variables in small intervals if these i.i.d. random variables are uniformly distributed in the interval $[0, 1]$. Here we take the supremum of F_n for all subintervals of $[0, 1]$ whose length is smaller than a prescribed number a_n. Actually paper [2] contains a more general result, but its application about the growth of the empirical distribution functions in small intervals seems to be its most interesting application. I do not give a precise formulation of Theorem 4 in [2], I omit its rather technical conditions. In particular, I do not describe what kind of conditions the number a_n must satisfy in this theorem. I only want to make some comments about its relation to the result about the model in our Example.

If we look carefully at the result of [2], then we can understand that it gives an improved version of the statement about the properties of the model discussed in the Example of this section in case (b). It enables us to define such a function $\hat u(\sigma)$ in this case for which even the relation

$$P\left((1-\varepsilon)\hat u(\sigma) \le \sup_{f \in \mathcal F_\sigma} |S_n(f)| \le (1+\varepsilon)\hat u(\sigma)\right) \to 1$$

holds for all $\varepsilon > 0$ if $n \to \infty$, and $\sigma = \sigma(n)$ satisfies the relation $n^{-400} \le \sigma^2 \le \frac{\log n}{8n}$. This means that in this case we can determine the value of the concentration point precisely and not only up to a multiplicative constant. Actually a precise explanation of this statement demands the elaboration of some technical details, but I omit this.

Finally I remark that our approach to the problem studied in this section is essentially different from that of [2]. In that paper the results are proved by means of some deep inequalities contained in earlier results, while here I tried to explain that they can be proved by means of a good Poissonian coupling. This may explain the situation better, and this approach seems to be appropriate also for the proof of the results in [2].

3. Proof of Theorem 1 and its extension.

Proof of Theorem 1. In the case (a) inequality (1.2) is a simple consequence of Theorem 1 in [6]. We can apply this result by writing σ instead of ρ in its formulation, since $\left(\int |f(x)|\mu(\,dx)\right)^2 \le \int f^2(x)\mu(\,dx) \le \sigma^2$ by the Cauchy–Schwarz inequality. Hence under the conditions of Theorem 1 the inequality $\int |f(x)|\mu(\,dx) \le \rho$ holds for all $f \in \mathcal{F}$ with $\rho = \sigma$, and by Theorem 1 of [6]

$$P\left(\sup_{f\in\mathcal{F}} |S_n(f)| \ge v\right) \le De^{-\frac{1}{25}\sqrt{n}v\log(\sigma^{-2})} \quad \text{if } v \ge \frac{\bar{C}}{\sqrt{n}}L \text{ and } \sigma^2 \le \frac{1}{n^{400}} \tag{3.1}$$

with an appropriate $\bar{C} > 0$. (Here we apply a division by $\sqrt{n}$ in the definition of $S_n(f)$, which was not done in [6], and this causes some difference in the formulas.)

I claim that we can drop the coefficient D at the right-hand side of (3.1) if we replace the coefficient $\frac{1}{25}$ by $\frac{1}{50}$ in the exponent, we choose such a constant $\bar{C}$ in (3.1) for which $\bar{C} \ge \frac{1}{8}$, and exploit that by condition (a) $v \ge u(\sigma) \ge \frac{\bar{C}}{\sqrt{n}}(L + \frac{\log D}{\log n})$. To show this it is enough to check that $D \le e^{\sqrt{n}v\log(\sigma^{-2})/50}$ in this case. This relation holds, since $\frac{\log D}{\log n} \le 8\sqrt{n}v$, and $\log(\sigma^{-2}) \ge 400\log n$, thus $D = \exp\{\frac{1}{400}(\frac{\log D}{\log n})(400\log n)\} \le \exp\{\frac{1}{50}\sqrt{n}v\log(\sigma^{-2})\}$, as I claimed.

Next I show that formula (3.1) or its previous modification remains valid if we replace $\log(\sigma^{-2})$ by $\log(\frac{v}{\sqrt{n}\sigma^2})$ in the exponent of its right-hand side. In the proof of this statement we can restrict our attention to the case $v \le \sqrt{n}$, since otherwise the probability at the left-hand side of (3.1) equals zero. In this case the inequality $\sigma^{-2} \ge \frac{v}{\sqrt{n}\sigma^2}$ holds, and this allows the above replacement. The above modifications of formula (3.1) imply inequality (1.2) in case (a).

Remark. If we are not interested in the value of the (universal) constants in (1.2), then in the case (a) this inequality has the same strength if we replace the term $\log(v/\sqrt{n}\sigma^2)$ by $\log(\sigma^{-2})$ in it. To see this, observe that beside the inequality $\sigma^{-2} \ge \frac{v}{\sqrt{n}\sigma^2}$ (if $v \le \sqrt{n}$), the inequality $\frac{v}{\sqrt{n}\sigma^2} \ge \frac{1}{n\sigma^2} \ge \sigma^{-2+1/200}$ also holds in case (a) because of the inequalities $v \ge u(\sigma) \ge n^{-1/2}$ and $n^{-200} \ge \sigma^2$. The original form of (1.2) has the advantage that it simultaneously holds in all cases (a), (b) and (c).

The proof of Theorem 1 in cases (b) and (c). We exploit that the class of functions $\mathcal{F}$ satisfies (1.1). We use this relation with the choice $\varepsilon = n^{-400}$ and the measure μ instead of $\sup\limits_{\nu}$. We may find in such a way $m \le Dn^{400L}$ functions $f_j \in \mathcal{F}$, $1 \le j \le m$, such that $\min\limits_{1\le j\le m} \int |f_j(x) - f(x)|\mu(\,dx) \le n^{-400}$ for all $f \in \mathcal{F}$. This means that $\mathcal{F} = \bigcup\limits_{j=1}^{n} \mathcal{D}_j$ with

$$\mathcal{D}_j = \left\{f\colon\ f \in \mathcal{F},\ \int |f_j(x) - f(x)|\mu(\,dx) \le n^{-400}\right\},$$

and as a consequence

$$P\left(\sup_{f\in\mathcal{F}}|S_n(f)|\ge v\right)\le\sum_{j=1}^m P\left(|S_n(f_j)|\ge\frac v2\right)+\sum_{j=1}^m P\left(\sup_{f\in\mathcal{D}_j}|S_n(f-f_j)|\ge\frac v2\right) \tag{3.2}$$

for all $v>0$. We shall estimate both terms at the right-hand side of (3.2) if $v\ge u(\sigma)$, the first term by means of Bennett's inequality, more precisely by a consequence of it formulated below, and the second term by means of the already proved case (a) of Theorem 1. We shall apply the following version of Bennett's inequality, see [5].

Bennett's inequality. *Let $X_1,\dots,X_n$ be independent and identically distributed random variables such that $P(|X_1|\le 1)=1$, $EX_1=0$ and $EX_1^2\le\sigma^2$ with some $0\le\sigma\le 1$. Put $S_n=\frac{1}{\sqrt n}\sum\limits_{j=1}^n X_j$. Then*

$$P(S_n>v)\le\exp\left\{-n\sigma^2\left[\left(1+\frac{v}{\sqrt n\sigma^2}\right)\log\left(1+\frac{v}{\sqrt n\sigma^2}\right)-\frac{v}{\sqrt n\sigma^2}\right]\right\}$$

for all $v>0$. As a consequence, for all $\varepsilon>0$ there exists some $B=B(\varepsilon)>0$ such that

$$P\left(S_n>v\right)\le\exp\left\{-(1-\varepsilon)\sqrt n v\log\frac{v}{\sqrt n\sigma^2}\right\}\quad \text{if } v>B\sqrt n\sigma^2,$$

and there exists some positive constant $K>0$ such that

$$P\left(S_n>v\right)\le\exp\left\{-K\sqrt n v\log\frac{v}{\sqrt n\sigma^2}\right\}\quad \text{if } v>2\sqrt n\sigma^2. \tag{3.3}$$

The above result is a special case of Theorem 3.2 in [5] in the case when we restrict our attention to sums of independent and identically distributed random variables. It has a slightly different form, because in the definition of S_n we considered normalized sums (with a multiplication by $n^{-1/2}$). Here we need only the inequality formulated in (3.3) which helps to estimate the probabilities appearing in the first sum at the right-hand side of (3.2). To apply (3.3) in the estimation of these terms we have to show that if the constants C_4 and C_5 are chosen sufficiently large in Theorem 1, then $u(\sigma)>2\sqrt n\sigma^2$ in cases (b) and (c).

In case (b) it is enough to show that $\sqrt n u(\sigma)\ge C_4\frac{\log n}{\log(\frac{\log n}{n\sigma^2})}\ge 2n\sigma^2$, and even $C_4\frac{\log n}{\log(\frac{\log n}{n\sigma^2})}\ge 20n\sigma^2$, or in an equivalent form $\frac{C_4}{20}\frac{\log n}{n\sigma^2}\ge\log(\frac{\log n}{n\sigma^2})$. (Observe that $\frac{\log n}{n\sigma^2}\ge 8$, hence $\log(\frac{\log n}{n\sigma^2})>0$ in case (b).) This statement holds, since $z=\frac{\log n}{n\sigma^2}\ge 8$ in case (b), and $\frac{C_4}{20}z\ge\log z$ if $z\ge 8$, and C_4 is sufficiently large.

In case (c), clearly $u(\sigma)\ge\frac{C_5}{\sqrt n}n\sigma^2\ge 20\sqrt n\sigma^2$ for sufficiently large constant C_5. These relations together with formula (3.3) imply that in cases (b) and (c)

$$P\left(|S_n(f_j)|\ge\frac v2\right)\le 2\exp\left\{-K\sqrt n v\log\frac{v}{\sqrt n\sigma^2}\right\}\quad\text{if } v\ge u(\sigma) \tag{3.4}$$

with an appropriate $K>0$ for all $1\le j\le m$. (In formula (3.4) we have exploited that $\log(\frac{\frac v2}{\sqrt n\sigma^2})\ge\frac12\log(\frac{v}{\sqrt n\sigma^2})$ since $\frac{v}{\sqrt n\sigma^2}\ge 20$, and as a consequence $\log(\frac{v}{\sqrt n\sigma^2})\ge 2\log 2$.)

Let us define, with the help of the class of functions $\mathcal{D}_j$ the class of functions $\mathcal{D}'_j=\{h\colon\, h=\frac{f-f_j}{2},\ f\in\mathcal{D}_j\}$ for all $1\le j\le m$. It is not difficult to see that $\sup\limits_{x\in X}|h(x)|\le 1$, $\int h^2(x)\mu(\,dx)\le\int|h(x)|\mu(\,dx)\le n^{-400}$ for all $h\in\mathcal{D}'_j$, and $\mathcal{D}'_j$ is a class of functions which has polynomially increasing covering numbers with parameter D and exponent L, $1\le j\le m$. I claim that

$$P\left(\sup_{f\in\mathcal{D}_j}|S_n(f-f_j)|\ge\frac v2\right)=P\left(\sup_{h\in\mathcal{D}'_j}|S_n(h_j)|\ge\frac v4\right)$$
$$\le e^{-C_2\sqrt n v\log(vn^{195})}\quad\text{if } v\ge u(\sigma)\tag{3.5}$$

for all $1\le j\le n$ in both cases (b) and (c). We shall get this estimate by applying Theorem 1 in the already proved case (a) with the choice of parameter $\sigma^2=n^{-400}$. To apply this result we have to check that $\frac v4\ge\frac{u(\sigma)}{4}\ge u(n^{-200})=\frac{C_3}{\sqrt n}(L+\frac{\log D}{\log n})$ if the constants C_4 and C_5 are sufficiently large. These statements hold, since in case (b) $\frac{\log n}{\log\frac{\log n}{n\sigma^2}}\ge\frac{\log n}{\log(n^{399}\log n)}\ge\frac{1}{400}$, hence $u(\sigma)\ge\frac{C_4}{\sqrt n}(\frac{L}{200}+\log D)\ge 4u(n^{-200})$ if C_4 is chosen sufficiently large, and an analogous but simpler argument supplies this relation in case (c) if C_5 is chosen sufficiently large.

It is not difficult to see that the right-hand side both of (3.4) and (3.5) can be bounded from above by $C_1e^{-\bar C_2\sqrt n v\log(v/\sqrt n\sigma^2)}$ with some appropriate constants $C_1>0$ and $\bar C_2>0$. Hence relations (3.2), (3.4) and (3.5) together with the inequality $m\le Dn^{400L}$ imply that

$$P\left(\sup_{f\in\mathcal{F}}|S_n(f)|\ge v\right)\le C_1Dn^{400L}e^{-\bar C_2\sqrt n v\log(v/\sqrt n\sigma^2)}\quad\text{if } v\ge u(\sigma)\tag{3.6}$$

in both cases (b) and (c). Hence to complete the proof of Theorem 1 (with the choice $C_2=\frac{\bar C_2}{2}$) it is enough to show that

$$e^{-\frac{\bar C_2}{2}\sqrt n v\log(v/\sqrt n\sigma^2)}\le e^{-\frac{\bar C_2}{2}\sqrt n u(\sigma)\log(u(\sigma)/\sqrt n\sigma^2)}\le D^{-1}n^{-400L}\quad\text{if } v\ge u(\sigma)\tag{3.7}$$

in cases (b) and (c) if the constants C_4 and C_5 are chosen sufficiently large.

It is enough to prove the second inequality in formula (3.7), since its proof also implies that the expressions in the exponent of this formula have negative value, and they are decreasing functions for $v\ge u(\sigma)$. The second inequality in (3.7) clearly holds in case (c), since $\frac{\bar C_2}{2}\sqrt n u(\sigma)\ge 400L\log n+\log D$, and $\log\frac{u(\sigma)}{\sqrt n\sigma^2}\ge 1$ in this case. In case (b) relation (3.7) can be reduced to the inequalities $\bar C_2\sqrt n u(\sigma)\log(\frac{u(\sigma)}{\sqrt n\sigma^2})\ge 800L\log n$, and $\bar C_2\sqrt n u(\sigma)\log(\frac{\sqrt n(\sigma)}{n\sigma^2})\ge 4\log D$. To prove the second inequality observe that in case (b)

$$\bar C_2\sqrt n u(\sigma)\ge C_4\bar C_2\log D\ge 4\log D,\quad\text{and}\quad \log\left(\frac{\sqrt n u(\sigma)}{n\sigma^2}\right)\ge 1.$$

The second of these inequalities follows from the relation

$$\frac{\sqrt{n}u(\sigma)}{n\sigma^2} \ge C_4 \frac{\frac{\log n}{n\sigma^2}}{\log(\frac{\log n}{n\sigma^2})} \ge 3,$$

which holds because of the relation $\frac{\log n}{n\sigma^2} \ge 8$ in case (b).

The remaining inequality can be rewritten as $\bar{C}_2 \frac{\sqrt{n}u(\sigma)}{n\sigma^2} \log(\frac{\sqrt{n}u(\sigma)}{n\sigma^2}) \ge 800L \frac{\log n}{n\sigma^2}$. To prove it observe that because of the definition of the function $u(\sigma)$ in case (b) we can write $\bar{C}_2 \frac{\sqrt{n}u(\sigma)}{n\sigma^2} \ge 1600L \frac{\log n}{n\sigma^2} \frac{1}{\log(\frac{\log n}{n\sigma^2})}$. I also claim that $\log(\frac{\sqrt{n}u(\sigma)}{n\sigma^2}) \ge \frac{1}{2}(\frac{\log n}{n\sigma^2})$. By multiplying the last two inequalities we get the desired inequality, and this completes the proof of Theorem 1.

To prove the above formulated inequality introduce the notation $z = \frac{\log n}{n\sigma^2}$ and $\hat{u}(\sigma) = \frac{1}{\sqrt{n}} \frac{\log n}{\log(\frac{\log n}{n\sigma^2})}$. By exploiting the definition of $u(\sigma)$ in case (b) we can write with the help of this notation that $\log(\frac{\sqrt{n}u(\sigma)}{n\sigma^2}) \ge \log(\frac{\sqrt{n}\hat{u}(\sigma)}{n\sigma^2}) = \log z - \log\log z \ge \frac{1}{2}\log z = \frac{1}{2}(\frac{\log n}{n\sigma^2})$. In this calculation we have exploited that in case (b) $z \ge 8$, hence $\log z - \log\log z \ge \frac{1}{2}\log z$. Theorem 1 is proved.

The extension of Theorem 1 is a slight generalization of Theorem 4.1 in [5], and its proof is based on the same ideas. The original proof in [5] is made by means of two results, formulated in Propositions 6.1 and 6.2 of that work. Here I present a slightly improved version of Proposition 6.2 in Theorem 3.1, which is, as I show, a simple consequence of Theorem 1. Then I formulate Theorem 3.2 which is a (simplified) version of Proposition 6.1 in [5]. I show that the Extension of Theorem 1 can be proved with the help of these results by slightly modifying (and simplifying) the proof of Theorem 4.1 in [5]. In Section 4 I shall discuss the role of Theorems 3.1 and 3.2 together with the idea behind them in more detail.

First I formulate Theorem 3.1.

Theorem 3.1. *Let us have a probability measure μ on a measurable space $(X, \mathcal{X})$ together with a sequence of independent and μ distributed random variables $\xi_1, \dots, \xi_n$, $n \ge 2$, and a countable class $\mathcal{F}$ of functions $f = f(x)$ on $(X, \mathcal{X})$ which has polynomially increasing covering numbers with some parameter $D \ge 1$ and exponent $L \ge 1$. Let this class of functions $\mathcal{F}$ also satisfy the relations $\sup_{x \in X} |f(x)| \le 1$, $\int f(x)\mu(\,dx) = 0$ and $\int f^2(x)\mu(\,dx) \le \sigma^2$ for all $f \in \mathcal{F}$ with some $0 < \sigma \le 1$ that satisfies the inequality $n\sigma^2 > L\log n + \log D$. Then there exists a threshold index A_0 such that the normalized random sums $S_n(f)$, $f \in \mathcal{F}$, introduced in Theorem 1 satisfy the inequality*

$$P\left(\sup_{f \in \mathcal{F}} |S_n(f)| \ge An^{1/2}\sigma^2\right) \le e^{-An\sigma^2} \quad \text{if } A \ge A_0. \tag{3.8}$$

I show that the estimate (3.8) in Theorem 3.1 is a weakened version of formula (1.2) of Theorem 1. First I show that the probability at the left-hand side of (3.8) can be estimated by means of Theorem 1 in case (c) with the choice $v = An^{1/2}\sigma^2$ if $A \ge A_0$ with a sufficiently large threshold index $A_0 > 0$. We have to check that $n\sigma^2 \ge \frac{1}{8}\log n$, and $v \ge u(\sigma)$ (with the function $u(\sigma)$ defined in case (c) of Theorem 1) if A_0 is chosen sufficiently large. These inequalities hold, since under the conditions of Theorem 3.1 $n\sigma^2 \ge L\log n \ge \frac{1}{8}\log n$, and for $v \ge A_0 n^{1/2}\sigma^2$ we can write $v \ge \frac{A_0}{\sqrt{n}} n\sigma^2 \ge \frac{A_0}{2\sqrt{n}} n\sigma^2 + \frac{A_0}{2\sqrt{n}}(L\log n + \log D) \ge \frac{C_5}{\sqrt{n}}(n\sigma^2 + L\log n + \log D) = u(\sigma)$.

Thus we can apply formula (1.2) with $v = An^{1/2}\sigma^2$ to estimate the left-hand side of (3.8), and we get the upper bound

$$C_1 e^{-C_2\sqrt{n}v\log(v/\sqrt{n}\sigma^2)} = C_1 e^{-C_2 An\sigma^2\log A} \le e^{-An\sigma^2}$$

for $A \ge A_0$ if the (universal) constant A_0 is chosen sufficiently large. Thus we proved Theorem 3.1 which provides a slightly better estimate than Proposition 6.2 in [5].

In the proof of the Extension of Theorem 1 I shall also apply following Theorem 3.2 which is a simple modified version of Proposition 6.1 in [5].

Theorem 3.2. *Let us have a sequence of i.i.d. random variables $\xi_1,\dots,\xi_n$, $n \ge 2$, on a measurable space $(X,\mathcal{X})$ with some distribution μ and a class of functions $\mathcal{F}$ on the space $(X,\mathcal{X})$ that satisfies the inequality $\mathcal{N}(\varepsilon,\mathcal{F},L_2(\mu)) \le \bar{D}\varepsilon^{-L}$ with some numbers $\bar{D} \ge 1$ and $L \ge 1$ for all $0 < \varepsilon \le 1$. Let us also assume that this class of functions $\mathcal{F}$ also has the properties $\sup\limits_{x\in X}|f(x)| \le 1$, $\int f(x)\mu(\,dx) = 0$ and $\int f^2(x)\mu(\,dx) \le \sigma^2$ with a prescribed number $0 < \sigma \le 1$ for all $f \in \mathcal{F}$. Take the normalized sums $S_n(f) = \frac{1}{\sqrt{n}}\sum\limits_{l=1}^{n} f(\xi_l)$ for all $f \in \mathcal{F}$, and let us fix a number $\bar{A} \ge 1$.*

There exists a number $M = M(\bar{A}) > 0$ such that with these parameters $\bar{A}$ and $M = M(\bar{A}) \ge 1$ the following relations hold. For all numbers $v > 0$ such that $n\sigma^2 \ge \left(\frac{v}{\sigma}\right)^2 \ge M(L\log\frac{2}{\sigma} + \log\bar{D})$ define the number $\bar{\sigma}_0 = \bar{\sigma}_0(v) = \frac{1}{8\sqrt{n}}\frac{v}{\bar{A}\sigma}$. Then for all numbers $\bar{\sigma}_0 \le \bar{\sigma} \le \sigma$ a collection of functions $\mathcal{F}_{\bar{\sigma}} = \{f_1,\dots,f_m\} \subset \mathcal{F}$ with $m \le \bar{D}2^{2L}\bar{\sigma}^{-L}$ elements can be chosen in such a way that the union of the sets $\mathcal{D}_j = \{f\colon f \in \mathcal{F}, \int|f-f_j|^2\,d\mu \le \bar{\sigma}^2\}$, $1 \le j \le m$, cover the set of functions $\mathcal{F}$, i.e. $\bigcup\limits_{j=1}^{m}\mathcal{D}_j = \mathcal{F}$, and the normalized random sums $S_n(f)$, $f \in \mathcal{F}_{\bar{\sigma}}$, $n \ge 2$, satisfy the inequality

$$P\left(\sup_{f\in\mathcal{F}_{\bar{\sigma}}}|S_n(f)| \ge \frac{v}{\bar{A}}\right) \le 4\exp\left\{-\alpha\left(\frac{v}{10\bar{A}\sigma}\right)^2\right\} \tag{3.9}$$

with an appropriate constant α and with the previously chosen parameter $\bar{A}$. (In formula (3.9) we have assumed that the number v appearing in it satisfies the condition $n\sigma^2 \ge (\frac{v}{\sigma})^2 \ge M(L\log\frac{2}{\sigma} + \log\bar{D})$.)

Remark. Theorem 3.2 is an empty statement if the inequality $n\sigma^2 \ge (\frac{v}{\sigma})^2 \ge M(L\log\frac{2}{\sigma} + \log\bar{D})$ has no solution. This result can be considered as a consequence of Proposition 6.1 in [5], although it contains some statements which are proved but not explicitly stated in [5]. In that work inequality (3.9) is proved in the special case when $\bar\sigma = 4^k\sigma$ with some non-negative integer k, and $\bar\sigma \ge \bar\sigma_0$, and in that case the set $\mathcal{F}_{\bar\sigma}$ can be chosen with smaller cardinality $m \le \bar D\bar\sigma^{-L}$. It is not difficult to deduce Theorem 3.2 from this result. Actually Theorem 3.2 contains the result one can prove with the help of the classical chaining method under the conditions of the Extension of Theorem 1. It is a classical method which works in 'regular Gaussian' or 'almost Gaussian' models, see [5].

In the proof of the Extension of Theorem 1 let us first check that under its conditions also the conditions of Theorem 3.2 hold with $\bar D = D2^L$. In particular, all numbers v satisfying the conditions of the Extension of Theorem 1 satisfy also the condition $n\sigma^2 \ge \left(\frac{v}{\sigma}\right)^2 \ge M(L\log\frac{2}{\sigma} + \log \bar D)$ of Theorem 3.2 if we choose the constant C_6 in the definition of $\bar u(\sigma)$ (in dependence of the value $M(\bar A)$) sufficiently large. The inequality $\mathcal{N}(\varepsilon, \mathcal{F}, L_2(\mu)) \le \bar D\varepsilon^{-L}$ with $\bar D = 2^L D$ under the conditions of this Extension follows from the inequality $\mathcal{N}(\varepsilon, \mathcal{F}, L_2(\mu)) \le \mathcal{N}(\frac{\varepsilon}{2}, \mathcal{F}, L_1(\mu))$ if $\mathcal{F}$ consists of functions whose absolute value is bounded by 1. This relation is a consequence of the observation that $\int |f-g|^2\,d\mu \le 2\int |f-g|\,d\mu$ if $\sup|f(x)| \le 1$ and $\sup|g(x)| \le 1$.

We still have to check that $n\sigma^2 \ge \left(\frac{v}{\sigma}\right)^2 \ge M(L\log\frac{2}{\sigma} + \log\bar D)$ under the conditions $\bar u(\sigma) \le v \le \sqrt{n}\sigma^2$. But this is a simple consequence of the definition of $\bar u(\sigma)$ if the constant C_6 is chosen sufficiently large in it. Actually at this point we could replace the number $L^{3/4}$ by $L^{1/2}$ in the definition of $\bar u(\sigma)$.

We shall prove the Extension of Theorem 1 with the help of Theorem 3.2 with the choice of $\bar\sigma^2 = \frac{v}{\bar A\sqrt{n}}$, where $\bar A = \max(2, A_0)$ and Theorem 3.1 with $\sigma = \bar\sigma$. First we have to check that this number $\bar\sigma$ satisfies the conditions $\bar\sigma_0 = \frac{1}{8\sqrt{n}}\frac{v}{A\sigma} \le \bar\sigma \le \sigma$ (to apply Theorem 3.2) and $n\bar\sigma^2 \ge L\log n + \log D$ (to apply Theorem 3.1) (if the number $M(\bar A)$ is sufficiently big). We shall also show that $n\bar\sigma^2 \ge L\log\frac{1}{\bar\sigma} + \log(2^{3L}D)$.

The inequality $\bar\sigma_0 \le \bar\sigma \le \sigma$ can be rewritten as $\frac{v^2}{64\sqrt{n}\bar A\sigma^2} = \bar A\sqrt{n}\bar\sigma_0^2 \le v \le \bar A\sqrt{n}\sigma^2$. Both of these inequalities hold if $v \le \bar A\sqrt{n}\sigma^2$, or in an equivalent form $(\frac{v}{\sigma})^2 \le \bar A^2 n\sigma^2$. This inequality holds under the conditions of Theorem 3.2, since we chose a number $\bar A \ge 1$.

To prove the second inequality let us observe that $n\bar\sigma^2 \ge n\bar\sigma_0^2 = \frac{1}{64}\frac{v^2}{A^2\sigma^2} \ge \frac{M}{64A^2}(L\log\frac{2}{\sigma} + \log(2^L D))$. This calculation implies the desired inequality in the case $\sigma \le n^{-1/3}$ if the constant $M = M(\bar A)$ is chosen sufficiently large, since in this case $\log\frac{2}{\sigma} \ge \frac13\log n$. In the case $n^{-1/3} \le \sigma \le 1$ we exploit that in the Extension of Theorem 1 we restricted our attention to the case when the number u satisfies the more restrictive condition $v \ge \bar u(\sigma) = C_6\sigma(L^{3/4}\log^{1/2}\frac{2}{\sigma} + (\log D)^{1/2})$. In this case we can write $n\bar\sigma^2 = \frac{\sqrt{n}v}{A} \ge \frac{\sqrt{n}\bar u(\bar\sigma)}{A} \ge \frac{C_6\sqrt{n}\sigma L^{3/4}\log^{1/2}\frac{2}{\sigma}}{A} \ge L^{3/4}n^{1/6}$ if the constant C_6 is sufficiently large, and $n\bar\sigma^2 \ge n\bar\sigma_0^2 = \frac{1}{64}\frac{v^2}{A^2\sigma^2} \ge \frac{1}{64}\frac{\bar u(\sigma)^2}{A^2\sigma^2} \ge C_6 L^{3/2}\log\frac{2}{\sigma} \ge C_6L^{3/2}$. The last two inequalities imply that in the case $n^{-1/3}\sigma \le 1$ we have $n\bar\sigma^2 = (n\bar\sigma^2)^{2/3}(n\bar\sigma^2)^{1/3} \ge C_6^{1/3}Ln^{1/9} \ge 2L\log n$. On the other hand, the former results imply that $n\bar\sigma^2 \ge 2\log D$, and as a consequence the desired inequality holds also in the case $n^{-1/3} \le \sigma \le 1$. The remaining inequality $n\bar\sigma^2 \ge L\log\frac{1}{\bar\sigma} + \log(2^{3L}D)$ follows from the first estimate we proved about $n\bar\sigma^2$.

To prove the Extension of Theorem 1 let us choose with the help of Theorem 3.2 a sequence of functions $\mathcal{F}_{\bar\sigma}=\{f_1,\dots,f_m\}\subset\mathcal{F}$ and sets $\mathcal{D}_j=\{f\colon f\in\mathcal{F}, \int|f-f_j|^2\,d\mu\le\bar\sigma^2\}$, $1\le j\le m$, with $m\le\bar D2^{2L}\bar\sigma^{-L}$ elements such that $\bigcup\limits_{j=1}^m\mathcal{D}_j=\mathcal{F}$.

Since we chose a number $A\ge2$ with the above notation we can write up the inequality

$$P\left(\sup_{f\in\mathcal{F}}|S_n(f)|\ge v\right)\le P\left(\sup_{f\in\mathcal{F}_{\bar\sigma}}|S_n(f)|\ge\frac{v}{A}\right)+\sum_{j=1}^m P\left(\sup_{f\in\mathcal{D}_j}|S_n(f-f_j)|\ge\frac{v}{2}\right)$$

if $\bar u(\sigma)\le v\le\sqrt{n}\sigma^2$, and the two terms at the right-hand side of this inequality can be estimated by means of Theorems 3.1 and 3.2.

We can write

$$P\left(\sup_{f\in\mathcal{F}_{\bar\sigma}}|S_n(f)|\ge\frac{v}{\bar A}\right)\le 4e^{-\alpha v^2/100\bar A^2\sigma^2}$$

by Theorem 3.2, and

$$\sum_{j=1}^m P\left(\sup_{f\in\mathcal{D}_j}|S_n(f-f_j)|\ge\frac{v}{2}\right)\le me^{-nA\bar\sigma^2}$$

by Theorem 3.1.

On the other hand,

$$me^{-nA\bar\sigma^2}\le D2^{3L}\bar\sigma^{-L}e^{-nA\bar\sigma^2}\le e^{-nA\bar\sigma^2/2}\le e^{-nA\bar\sigma_0^2/2}=e^{-Av^2/64\bar A\sigma^2}\le e^{-v^2/64\sigma^2},$$

since

$$D2^{3L}\bar\sigma^{-L}e^{-nA\bar\sigma^2/2}\le D2^{3L}\bar\sigma^{-L}e^{-n\bar\sigma^2}\le1$$

by the inequality $n\bar\sigma^2\ge L\log\frac{1}{\bar\sigma}+\log(D2^{3L})$. The above inequalities imply that

$$P\left(\sup_{f\in\mathcal{F}}|S_n(f)|\ge v\right)\le 4e^{-\alpha v^2/100\bar A^2\sigma^2}+e^{-v^2/64\sigma^2}$$

if $\bar u(\sigma)\le v\le\sqrt{n}\sigma^2$. Thus formula (1.3) and the Extension of Theorem 1 is proved.

I finish this paper with a discussion about its methods and results.

4. A discussion about the methods and results of this paper.

The problem of this paper, the estimation of the tail-distribution of the supremum $\sup\limits_{f\in\mathcal F} S_n(f)$ of the normalized sums $S_n(f)=\frac1{\sqrt n}\sum\limits_{k=1}^n f(\xi_k)$ for a sequence of i.i.d. random variables $\xi_1,\dots,\xi_n$ and a class of functions $\mathcal F$ with some nice properties has a long history. Such a problem arises in a natural way in the study of the uniform central limit theorem for a class of normalized sums $S_n(f)$, $f\in\mathcal F$, with a nice class of functions $\mathcal F$, see [3]. An important part of such a study is to prove the 'tightness' of the class of functions $S_n(f)$, $f\in\mathcal F$, by showing first that for a subclass $\mathcal F'\subset\mathcal F$ such that $E(f-g)^2\le\delta$ with a small number δ for any pairs $f,g\in\mathcal F'$ the supremum $\sup\limits_{f\in\mathcal F'} S_n(f-g)$ with a fixed $g\in\mathcal F$ is small with probability almost 1. There are some results that give a bound on the tail-distribution of such a supremum if $\delta=\delta_n$, and $\delta_n\to0$ as $n\to\infty$. But the estimates I know about in this direction do not provide a sharp estimate if $\delta_n\to0$ very fast. My goal in this paper was to give a good estimate also in such cases, and to give a good bound on the tail distribution of $\sup\limits_{f\in\mathcal F'} S_n(f-g)$ in the general case $\delta=\delta_n$.

Let us remark that for large indices n the random variables $S_n(f)$ are asymptotically Gaussian. Hence it is natural to study first the natural Gaussian counterpart of the above problem to understand what kind of estimates hold in this modified problem, what kind of methods are useful in their study, and how they can be adapted to our original problem. The following problem can be considered as this natural Gaussian counterpart. Take a class of (jointly) Gaussian random variables η_t, $E\eta_t=0$, $t\in T$, with a (countable) parameter set T, and give a good estimate on the tail distribution of $\sup\limits_{t\in T}\eta_t$ with the help of the (pseudo) metric $\rho(s,t)$, $\rho^2(s,t)=E(\eta_s-\eta_t)^2$, $s,t\in T$. There is a good solution of this problem with the help of the so-called chaining argument. This is worked out in detail in [12], and this book contains the sharpest results in this direction. We get a good estimate if for all $\varepsilon>0$ we can find a set $\{t_1,\dots,t_M\}\subset T$ with relatively few $M=M(\varepsilon)$ elements, whose ε-neighbourhood with respect to the metric ρ covers the whole space T. The estimate depends on this function $M(\varepsilon)$. In particular, if $M(\varepsilon)\le D\varepsilon^{-L}$ with some constants $D>1$ and $L>1$, and $E\eta_t^2\le\sigma^2\le1$ for all $t\in T$ and $\varepsilon>0$, then the estimate $P\left(\sup\limits_{t\in T}\eta_t>u\right)\le De^{-(u-u(\sigma))^2/2\sigma^2}$ holds for all $u\ge u(\sigma)$ with $u(\sigma)=CL^{1/2}\sigma\log^{1/2}\frac2\sigma$, where $C>0$ is a universal constant. The book [12] contains a sharper result which provides a good estimate in the general case. It is also mentioned in this book that a similar estimate holds for an arbitrary set of random variables ζ_t, $t\in T$, if they satisfy the 'Gaussian type estimate' $P(|\zeta_t-\zeta_s|>u)\le Ce^{-\alpha u^2/\rho^2(s,t)}$ with some fixed numbers $C>0$ and $\alpha>0$ for all $s,t\in T$ and $u>0$. The question arises whether such an estimate holds also in our original problem about the supremum of normalized random sums $S_n(f)$, $f\in\mathcal F$, if they are defined with the help of a nice class of functions $\mathcal F$.

Let us assume that $\sup|f(x)|\le1$, and $ES_n(f)=0$ for all $f\in\mathcal F$ in the class of functions $\mathcal F$ we consider. Then we may try to apply the above indicated result with $T=\mathcal F$ and an appropriate metric ρ on it. Observe that $E[S_n(f)-S_n(g)]^2=\int(f-g)^2\,d\mu$

for all $f, g \in \mathcal{F}$, where μ denotes the distribution of the random variables ξ_j. This means that we have to work with the metric $\rho(f,g)$ defined as $\rho^2(f,g) = \int (f-g)^2 \, d\mu$ in this case. The question arises whether the above formulated 'Gaussian type estimate' which provides a good estimate on the tail distribution of the supremum we are interested in holds in this case.

I discussed this problem in detail in the third chapter of my book [5]. The main point is that there are some classical results, like the Bernstein or Bennett inequality that give good estimates for the tail distribution of sums of bounded i.i.d. random variables, but they provide so good 'Gaussian type estimates' that we need only at not too high levels u. There are also examples that show that in certain cases we cannot get good 'Gaussian type estimates' at high levels u. This has the consequence that the chaining argument worked out to handle the Gaussian counterpart of our problem is not good enough to solve our problem. It enables us to reduce it to the special case, when the distance $\rho(f,g)$ is very small for all $f, g \in \mathcal{F}$, but it does not give more help. (How small this distance must be that depends on the sample size n.) The study of this reduced problem demands new ideas. Moreover, to get good estimates we have to introduce some new conditions about the behaviour of the class $\mathcal{F}$, it is not enough to have good control on the metric $\rho(f,g)$, $f, g \in \mathcal{F}$, introduced above.

There are two main approaches to introduce appropriate new conditions which enable us to prove good estimates in the problem we are interested in. The first one can be found in the book of Talagrand [12]. He introduced a condition by which for all $\varepsilon > 0$ the class of functions $\mathcal{F}$ must have an ε-dense subset with relatively few elements not only with respect to the metric ρ but also with respect to the supremum norm. Theorems 1.2.7 and 2.7.2 in [12] are results in this spirit. Talagrand also proved some interesting and deep consequences of these results in Chapter 3 of [12]. There are however important problems where such an approach does not work. Such problems are e.g. the behaviour of the models considered in the Example of Section 2 or the problems considered in Section 2 of [6]. More generally such a problem appears if $\mathcal{F}$ consists of the indicator functions χ_A of different sets or if we consider their normalized versions $f_A(x) = \chi_A(x) - \mu(A)$. (We may apply such a normalization to get functions whose integral with respect to the measure μ equals zero.) In such cases all functions of $\mathcal{F}$ are far from each other in the supremum norm, and as a consequence of it $\mathcal{F}$ has no dense subset with respect to the supremum norm with relatively few elements. To overcome this difficulty a different additional condition was introduced. This new condition demands that $\mathcal{F}$ must be a class of functions with polynomially increasing covering numbers. This approach proved to be useful in several interesting cases when the method of [12] does not work. There are some works, see e.g. [3], [10], [14] where it is shown that there are many classes of functions $\mathcal{F}$ with polynomially increasing covering numbers. The proof about their existence is closely related to the theory of Vapnik–Červonenkis classes.

The original technique for proving good estimates on the tail distribution of the supremum of the random sums $S_n(f)$, $f \in \mathcal{F}$, under the condition that the class of functions $\mathcal{F}$ has polynomially increasing covering numbers was the application of the so-called symmetrization argument. This technique is applied in several works, see e.g. [3], [5], [7], [9], [10], [13], and it works in several models when the method of [12] is not applicable. I do not describe this method, I only remark that I compared it with that of Talagrand in Chapter 18 of [5] at pp. 235–237. Here I also made a comparison between the applicability of these two methods.

Nevertheless, the symmetrization argument does not provide a sharp estimate if the bound $\sigma^2 \ge \sup\limits_{f\in\mathcal{F}} Ef^2(\xi_j)$ is too small. The main goal of the present paper to give a sharp estimate also in this case. To understand our results better let us compare them with the results of some previous papers in the case when the class of functions $\mathcal{F}$ contains functions bounded by 1, and it has polynomially increasing covering numbers with bounded exponent L and parameter D, i.e. these numbers have a bound not depending on σ. Paper [14] gives the following upper bound for the value of the concentration point of the distribution of $\sup\limits_{f\in\mathcal{F}} S_n(f)$ in this case.

$$E^* \sup_{f\in\mathcal{F}} S_n(f) \le CJ(\sigma,\mathcal{F},L_2)\left(1+\frac{J(\sigma,\mathcal{F},L_2)}{\sigma^2\sqrt{n}}\right)$$

with a universal coefficient C, where

$$J(\sigma,\mathcal{F},L_2) = \sup_\nu \int_0^\sigma \sqrt{1+\log\mathcal{N}(\varepsilon,\mathcal{F},L_2(\nu))}\,d\varepsilon$$

with the uniform covering number $\sup\limits_\nu \mathcal{N}(\cdot,\cdot,\cdot)$ with respect to L_2-norms. (Here the notation E^* is applied, since the choice of a non-countable class of functions $\mathcal{F}$ is also allowed, and in this case the outer expectation E^* is applied.)

Some calculation shows that in this case $J(\sigma,\mathcal{F},L_2) \asymp \sigma\sqrt{\log\frac{2}{\sigma}}$, hence we get the upper bound const. $\left(\sigma\sqrt{\log\frac{2}{\sigma}}+\frac{\log\frac{2}{\sigma}}{\sqrt{n}}\right)$ for the value of the concentration point in this case. This yields the upper bound $C\sigma\sqrt{\log\frac{2}{\sigma}}$ if $\sigma^2 \ge \text{const.}\frac{\log n}{n}$ and $C\frac{\log\frac{2}{\sigma}}{\sqrt{n}}$ if $\sigma^2 \le \text{const.}\frac{\log n}{n}$ for the value of the concentration point. This result is sharp in the first case, (see Theorem 1 and its Extension together with the Example in Section 2). But it is not sharp in the second case. Moreover, it can be improved in the following trivial way. If $\sigma^2 \le \sigma_0^2 = \frac{\log n}{n}$, then we can apply the above estimate for σ_0^2 instead of σ^2, and this yields the upper bound $\frac{\log n}{\sqrt{n}}$ instead of the estimate $\frac{\log\frac{2}{\sigma}}{\sqrt{n}}$ for the value of the concentration point. This means that the result of [14] could not yield a better estimate if $\sigma^2 \ll \sigma_0^2$ than in the case $\sigma^2 = \sigma_0^2$.

Massart's paper [7] contains another result about the tail distribution of the supremum of $S_n(f)$, $f \in \mathcal{F}$. The proof in that paper is based on a modified version of the symmetrization argument. The result of [7] is rather complicated, but one can get an estimate for the value of the concentration point with its help. Here I shall consider the version of this result presented in Theorem 2.14.16 of the book [13]. We can get the bound for the value of the concentration point by calculating when the bound given for the tail distribution of the supremum given in this result becomes smaller than 1. Some calculation that I would omit would provide the right bound $C\sigma\sqrt{\log\frac{2}{\sigma}}$ if $\sigma^2 \ge \text{const.}n^{-1/4}$ and a much weaker bound $Cn^{-1/4}\log^{1/2}\frac{2}{\sigma}$ for the value of the concentration point in the other case. It is also worth considering the estimate that Alexander's method worked out in [1] supplies. It is based on the chaining argument, and it yields a good estimate, similarly to [14] if $\sigma^2 \ge \sigma_0^2 = \frac{\log n}{n}$, and a weak one in the other case.

Actually the proof of the result in [5] corresponding to the Extension of the Theorem 1 is based on Alexander's idea in [1], and it yields a good estimate only for $\sigma \ge \sigma_0$. The starting point in the present investigation was an attempt to find a refinement of this method which supplies a good estimate for the tail distribution of the supremum we are investigating if the number σ^2 satisfying the inequality $\sigma^2 \ge \sup\limits_{f\in\mathcal{F}} Ef^2(\xi_j)$ can be chosen in an arbitrary way. The original result was proved by means of an appropriate inductive hypothesis. To get an improved version of it we have to find a good reformulation of this inductive hypothesis that takes into account that in the case of small parameter σ^2 we have a better estimate. This lead me to the investigation of the problem in paper [6]. Then it turned out that a direct application of the results in [6] enables us to work out a different method that yields a more general result with less effort. It may be interesting to compare this method with some standard techniques applied in the study of other probabilistic problems.

In the proof of limit theorems for sums of independent random variables or in the study of some similar problems a standard method is the application of the so-called truncation. The truncated random terms show nice 'regular' behaviour, since they are bounded. This enables us to study them with the help of classical methods. The contributions omitted by truncation contain the 'irregular' part of the random variables, and they cannot be handled by standard methods. But in nice cases it can be proved that they are negligible, hence we can prove the desired results.

Here we applied a similar approach to prove our estimates with the help of the result in [6]. We took some appropriately chosen functions $f_j \in \mathcal{F}$, considered their small neighbourhoods with respect to the metric ρ defined in this section, and estimated the increase of $S_n(f)$ in these neighbourhoods. More explicitly, we chose some appropriate functions $f_j \in \mathcal{F}$ and an appropriate small number $\sigma > 0$, and we estimated the tail distribution of $\sup\limits_{f\in\mathcal{F},\,\rho(f,f_j)\le\sigma} S_n(f-f_j)$. The tail distribution of these terms could be well estimated by means of the result in [6]. They played a role similar to that part of random sums which were omitted at truncation in some analogous problems because of their large value. These terms are small by the results of [6]. On the other hand, they enable us to restrict our attention to such problems where we can make good estimations by means of some standard methods, like the application of classical estimates on the tail distribution of the single terms $S_n(f)$ or the chaining argument. In the proof of Theorem 1 and its Extension actually such an approach was followed.

If we look carefully how we could work with the help of the result of [6] and how the symmetrization argument was applied in other works, then we can see that they played a similar role. It seems to me that the result of [6] can replace the symmetrization argument in most applications, moreover it supplies a more powerful tool.

References.

[1] K. S. Alexander, Probability inequalities for empirical processes and a law of the iterated logarithm. Annals of Probability **15** no. 4 1041–1067 (1984)

[2] P. Deheuvels, J. H. J. Einmahl, D. M. Mason, F. H. Ruymgaart, The almost sure behavior of maximal and minimal multivariate k_n-spacings. J. Multivariate Anal. **24**, 155–176 (1988)

[3] R. M. Dudley, *Uniform Central Limit Theorems. Second Edition.* Cambridge University Press, Cambridge (2014)

[4] J. Komlós, P. Major, G. Tusnády, An approximation of partial sums of independent rv.'s and the sample DF. I. Z. Wahrscheinlichkeitstheorie verw. Gebiete 32, 111–131 (1975)

[5] P. Major, On the estimation of multiple random integrals and U-statistics *Lecture Notes in Mathematics* vol 2079 (Springer) Heidelberg New York Dordrecht London (2013)

[6] P. Major, Sharp estimate on the supremum of a class of partial sums of small i.i.d. random variables, Stochastic Processes and their Applications (2015), http://dx.doi.org/10.1016/j.spa.2015.07.016

[7] P. Massart, Rates of convergence in the central limit theorem for empirical processes. Annales Institut Henri Poincaré **22**, 381–423 (1986)

[8] H. P. McKean Jr., *Stochastic integrals.* Academic Press New York London (1969)

[9] D. Pollard, A central limit theorem for empirical processes. Austral Math. Soc. (Series A). 235–248 **33**, (1982)

[10] D. Pollard, *Convergence of Stochastic Processes.* (Springer, New York, (1984)

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

[12] M. Talagrand, *The generic chaining.* Springer Monographs in Mathematics. Springer–Verlag, Berlin Heidelberg New York (2005)

[13] A. W. van der Vaart, J. A. Wellner, *Weak Convergence and Empirical Processes.* Springer, New York (1996)

[14] A. W. van der Wart, J. A. Wellner, A local maximal inequality under uniform entropy. Electronic Journal of Statistics **5**, 192–203 (2011)