Formulation of the problem

Let $\xi_1,\dots,\xi_n$ be a sequence of i.i.d. random variables with some distribution μ on a measurable space $(X,\mathcal{X})$.
Let a class of functions $\mathcal{F}$ be given on the space $(X,\mathcal{X})$ with some nice properties, such that all elements $f\in\mathcal{F}$ have the properties $\int f(x)\mu(\,dx)=0$, $\sup_{x\in X}|f(x)|\le 1$, and $\int f^2(x)\mu(\,dx)\le\sigma^2$ with some $0<\sigma^2\le 1$.

We also assume that $\mathcal{F}$ is a class of functions with polynomially increasing covering numbers with exponent L and parameter D. (I recall its definition later.)

Define the normalized sums $S_n(f)=\frac{1}{\sqrt n}\sum_{j=1}^n f(\xi_j)$ for all $f\in\mathcal{F}$. Our goal is to give a good estimate on the tail distribution

$$P\left(\sup_{f\in\mathcal{F}} S_n(f)>x\right)\quad \text{for all numbers } x>0$$

of the supremum of these sums. This may depend on σ^2, L and D.

A Gaussian version of the problem

The sums $S_n(f) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} f(\xi_j)$ are asymptotically Gaussian. Hence it is natural to look at a Gaussian version of this problem, where we estimate the supremum of Gaussian random variables η_t, $t \in T$.
We want to understand what kind of methods and results this Gaussian version suggests.
This Gaussian problem can be solved with the help of the so-called chaining argument. The estimate depends on the metric $d_2(\cdot,\cdot)$, defined as $d_2(s,t) = \left[E(\eta_s - \eta_t)^2\right]^{1/2}$, $s,t \in T$.
A similar result holds if the random variables η_t, $t \in T$, may be non-Gaussian, but they satisfy the Gaussian type inequality

$$P(|\eta_t - \eta_s| > u) \le C_1 e^{-C_2 u^2/d_2(s,t)^2} \quad \text{for all } s,t \in T \text{ and } u > 0 \tag{1}$$

with some constants $C_1 > 0$ and $C_2 > 0$.
The question arises whether such a result holds in our problem.

We can formulate our problem in that form as its Gaussian counterpart. By some classical estimates the inequality (1) holds under some restrictions (if $u > 0$ is not too large, σ^2 is not too small). These conditions cannot be omitted.
As a consequence, we can give a good estimate in our problem only under some additional conditions. We can handle the case when σ^2 is small only if we impose new conditions and find new methods.

The new condition: $\mathcal{F}$ is a class of functions with polynomially increasing covering numbers.
The method of proof applied in earlier papers: Symmetrization argument.

The notion of class of functions with polynomially increasing covering numbers is a natural version of the Vapnik–Červonenkis classes if we are working with classes of functions instead of classes of sets.

I give its definition in two steps.

First step of the definition.

Definition of uniform covering numbers with respect to L_1-norm. 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_1-norm is $\sup_\nu \mathcal{N}(\varepsilon, \mathcal{F}, L_1(\nu))$, where the supremum is taken for all probability measures ν on the space $(X, \mathcal{X})$, and $\mathcal{N}(\varepsilon, \mathcal{F}, L_1(\nu))$ is the smallest integer m for which there exist some functions $f_j \in \mathcal{F}$, $1 \le j \le m$, such that $\min_{1 \le j \le m} \int |f - f_j| \, d\nu \le \varepsilon$ for all $f \in \mathcal{F}$.

Second step of the definition.

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

In the symmetrization argument we exploit that under some not too restrictive conditions $\sup_{f\in\mathcal{F}} \sum_{j=1}^n f(\xi_j)$ has a similar tail distibution as its randomized version $\sup_{f\in\mathcal{F}} \sum_{j=1}^n \varepsilon_j f(\xi_j)$, where $\varepsilon_1,\dots,\varepsilon_n$ are i.i.d. random variables, independent also of the random variables ξ_j, and $P(\varepsilon_1=1)=P(\varepsilon_1=-1)=\frac{1}{2}$.
The symmetrized version can be better handled than the original expression, and this is exploited in the symmetrization argument.

Nevertheless, this technique gives a weak estimate if $Ef(\xi_1)^2\le\sigma^2$ for all $f\in\mathcal{F}$ with a very small σ^2, since in this case the 'cancellation effect' of the symmetrization does not work.

Our goal is to give a sharp estimate for all $0\le\sigma^2<1$. It turned out that the main result of this paper the Main theorem formulated below plays a crucial role in achieving this goal.

Main theorem. *Let $\mathcal{F}$ be a finite or countable class of functions on a measurable space $(X, \mathcal{X})$ which has polynomially increasing covering numbers with some parameter $D \geq 1$ and exponent $L \geq 1$, and $\sup_{x \in X} |f(x)| \leq 1$ for all $f \in \mathcal{F}$. Let $\xi_1, \dots, \xi_n$, $n \geq 2$, be a sequence of i.i.d. random variables with values in the space $(X, \mathcal{X})$ with a distribution μ, and assume that the inequality $\int |f(x)| \mu(\,dx) \leq \rho$ holds for all $f \in \mathcal{F}$ with a number $0 < \rho \leq n^{-200}$. Put $\bar{S}_n(f) = \bar{S}_n(f)(\xi_1, \dots, \xi_n) = \sum_{j=1}^n f(\xi_j)$ for all $f \in \mathcal{F}$. The inequality*

$$P\left(\sup_{f \in \mathcal{F}} |\bar{S}_n(f)| \geq u\right) \leq D\rho^{Cu} \quad \textit{for all } u > 41L \tag{2}$$

holds with some universal constant $1 > C > 0$. We can choose e.g. $C = \frac{1}{50}$.

The following details may be interesting in this result:
1.) We assumed that $\int |f(x)|\mu(dx) \le \rho$ for $f \in \mathcal{F}$ with $0 < \rho \le n^{-200}$, i.e. ρ is very small.
2.) We did not need the condition $\int f(x)\mu(dx) = 0$ for $f \in \mathcal{F}$.
3.) Formula (2) holds for $u > 41L$, i.e. even at relatively low levels.

There are simple examples that show that the Main theorem is a sharp estimate.

Why is the above result interesting for us?

It is important for us not as a new estimate, but as a new tool to investigate our problem.

Since the condition $0 < \rho \le n^{-200}$ means that ρ is a polynomially decreasing function of the sample size n we can work out a better estimate in our problem with the help of a Vapnik–Červonenkis type method than the symmetrization argument.

In my subsequent paper Sharp tail distribution estimates for the supremum of a class of sums of i.i.d. random variables I give a fairly complete solution of the problem mentioned at the start of this talk with the help of Main Theorem.

Here I could give a very concise explanation of this paper. A more detailed version of a (possible) talk about the motivations and content of this paper can be found on my homepage at the address

http://www.renyi.hu/~major/talks/positive.html

On the subsequent paper Sharp tail distribution estimates for the supremum of a class of sums of i.i.d. random variables at the address

http://www.renyi.hu/~major/talks/supremum.html