# The problem we are interested in 

Let $\xi_1,\dots,\xi_n$ be a sequence of i.i.d. random variables with some distribution $\mu$ on a measurable space $(X,\mathcal{X})$.
Let a class of functions $\mathcal{F}$ consisting of countably many functions be given on the space $(X,\mathcal{X})$ with the properties $\int f(x)\mu(\,dx)=0$, $\sup_{x\in X}|f(x)|\le 1$ and $\int f(x)^2\mu(\,dx)\le\sigma^2$ with some $0<\sigma\le 1$ for all elements $f\in\mathcal{F}$.
Let $\mathcal{F}$ be a class of functions with polynomially increasing covering numbers with exponent $L\ge 1$ and parameter $D\ge 1$. (I recall the definition of this notion later.)
Define the normalized sums $S_n(f)=\frac{1}{\sqrt{n}}\sum_{j=1}^n f(\xi_j)$ for all $f\in\mathcal{F}$, and give a good estimate on the tail distribution

$$P\left(\sup_{f\in\mathcal{F}} S_n(f)>v\right)\quad\text{for all numbers } v>0 \tag{1}$$

of the supremum of these sums. This estimate may depend on $\sigma$, $L$ and $D$.

An important result, called the concentration inequality, has the consequence that there is a number $v_0$ such that for $v < v_0$ the probability in formula (1) is almost 1, while for $v > v_0$, it begins to decrease rapidly. An important part of our problem is to give a good estimate on this numbern $v_0$.

There is a solution for a natural version of our problem when we estimate the supremum of Gaussian random variables. This is done by means of a method called the chaining argument.
A similar estimate can be proved under some restrictions in our model with the help of the so-called symmetrization argument. Our goal is to prove a good estimate in in the general case, which holds without these restrictions.

To discuss this problem first I recall the definition of classes of functions with polynomially incerasing covering numbers together with their exponent $L$ and parameter $D$. I do this in two steps on the next page.

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 $\varepsilon$, $\varepsilon > 0$, 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 $\nu$ 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.

# A useful example



First I present a simple example which may help to understand how to estimate the concentration point $v_0$.

Example. Take a sequence of independent, uniformly distributed random variables $\xi_1, \dots, \xi_n$ on the unit interval $[0, 1]$, fix a number $0 \le \sigma^2 \le 1$, and define a class of functions $\mathcal{F}_\sigma$ and $\bar{\mathcal{F}}_\sigma$ as set of 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 $k = k(\sigma) = [\frac{1}{\sigma^2}]$, 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 $f_j(x) = f_j(x|\sigma) = \bar{f}_j(x) - \sigma^2$, $1 \le j \le n$. Give a good estimate on $P_n(v) = P(\sup_j S_n(f_j) > v)$.

$\mathcal{F}$ satisfies the conditions in our problem. It is a class of functions with polynomially increasing covering numbers with some exponent $L$ and parameter $D$ which do not depend on $\sigma^2$, and the parameter $\sigma^2$ introduced in the model is an upper bound for all $\int f_j(x)^2 \mu(\,dx)$.

Our first problem: Find a good lower bound for the numbers $v_0$ for which $P_n(v) \ll 1$ only for $v > v_0$. The next result gives a solution.

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 $0 \le \sigma \le 1$ the inequality

$$P_n(\hat{u}(\sigma)) = P\left(\sup_{f\in\mathcal{F}_\sigma} |S_n(f)| \ge \hat{u}(\sigma)\right) \ge 1-\delta,$$

holds with

1.) $\hat{u}(\sigma) = \frac{\bar{C}}{\sqrt{n}}$ if $\sigma^2 \le n^{-400}$,

2.) $\hat{u}(\sigma) = \frac{\bar{C}}{\sqrt{n}} \frac{\log n}{\log(\frac{\log n}{n\sigma^2})}$ if $n^{-400} < \sigma^2 \le \frac{\log n}{8n}$, and

3.) $\hat{u}(\sigma) = \bar{C}\sigma \log^{1/2} \frac{2}{\sigma}$ if $\frac{\log n}{8n} \le \sigma^2 \le 1$.

This result says that $v_0 \ge \hat{u}(\sigma)$ for the numbers $v_0$ with the above demanded property. Our results will show that this estimate is sharp. The above example helps us to understand what kind of result we can expect in the general case.

Next I explain the picture behind this result.

In case 3.) of this example $\sigma^2$ is relatively large, and we get such an estimate for the best choice of $\hat{u}(\sigma)$ as in the Gaussian case.

In case 2.) we get a right choice for $\hat{u}(\sigma)$ not by a Gaussian but by a Poissonian approximation of our model.

In case 1.) $\sigma^2$ is very small. We can make a trivial estimate by exploiting that given an arbitrary partition of our space, some element of this partitition contains a given sample point. In this case this fact yields the right estimate for $\hat{u}(\sigma)$.

Next I formulate the Theorem, and its Extension, the main results of my paper. They say that in the general case we have the estimate that the above Example and some classical estimates on the tail distribution of sums of independent random variables suggest.

**Theorem.** *Let a sequence of i.i.d. random variables* $\xi_1, \dots, \xi_n$, $n \ge 2$, *with values in* $(X, \mathcal{X})$ *with some distribution* $\mu$ *and a countable class of functions* $\mathcal{F}$ *on the same space* $(X, \mathcal{X})$ *with polynomially increasing covering numbers with exponent* $L \ge 1$ *and parameter* $D \ge 1$ *be given. Let the functions* $f \in \mathcal{F}$ *satisfy the relations* $\sup_{x \in X} |f(x)| \le 1$, $\int f(x)\mu(\,dx) = 0$, *and* $\int f^2(x)\mu(\,dx) \le \sigma^2$ *with some number* $0 \le \sigma^2 \le 1$ *for all* $f \in \mathcal{F}$. *The normalized sums* $S_n(f)$, $f \in \mathcal{F}$, *satisfy the inequality*

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

*with some universal constants* $C_j > 0$, $1 \le j \le 5$, *if one of the following conditions is satisfied.*

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

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

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

**Extension of the Theorem.** *Let us consider, similarly to the Theorem, a sequence of i.i.d. random variables* $\xi_1, \dots, \xi_n$, $n \ge 2$, *with values in a space* $(X, \mathcal{X})$ *with some distribution* $\mu$ *which satisfies the conditions of the Theorem. In the case* $\frac{\log n}{8n} < \sigma^2 \le 1$ *the supremum of the normalized sums* $S_n(f)$, $f \in \mathcal{F}$, *satisfies the inequality*

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

*with appropriate (universal) constants* $\alpha > 0$, $C > 0$ *and* $C_6 > 0$ *if* $\sqrt{n}\sigma^2 \ge v \ge \bar{u}(\sigma)$, *where* $\bar{u}(\sigma)$ *is defined as* $\bar{u}(\sigma) = C_6\sigma(L^{3/4}\log^{1/2}\frac{2}{\sigma} + (\log D)^{1/2})$.

In cases 1.) and 2.) Theorem gives a good estimate for $v \ge u(\sigma)$ with an $u(\sigma)$ suggested by the Example. It is a (non-Gaussian) estimate suggested by Bennett's inequality. In case 3.) the Theorem and its Extension together give a good estimate. It holds for $v \ge \bar{u}(\sigma)$, as it is suggested by the Example. The Extension gives a good Gaussian estimate if $\bar{u}(\sigma) \le v \le \sqrt{n}\sigma^2$. Over this level we have a weaker estimate formulated in the Theorem.

The main ideas of the proof: I can simplify our problem with the help of my paper Sharp estimate on the supremum of a class of sums of small i.i.d. random variables by exploiting that we are working with a class of functions with polynomially increasing covering numbers. This enables us to reduce our problem to the case when we have finally many random sums $S_n(f)$ with some nice properties. The reduced problem can be solved by means of classical methods, like the Chaining argument and good estimates for sums of i.i.d. random variables.

The main point of the proof is that we can separate the regular and irregular contributions to the supremum we are investigating. The regular part can be well investigated by classical tools like the chaining argument. The real novelty in my research was to find a new method for the estimation of the irregular effects. This demanded the application of new arguments.

I put a more detailed version of this talk to the address

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