Uspekhi Mat. Nauk. Vol. 50, No. 5, 1995, (3-62), (Russian).
Russian Math. Surveys Vol. 50, No. 5, 1995, (849-904), (English).
Preprint Ser. No. 24, 1994, *Math. Inst. Aarhus*
Research Report No. 304 (1995), *Dept. Theoret. Statist. Aarhus*

To the centenary of the birth of
Aleksandr Yakovlevitch Khintchine
(19.07.1894 - 18.11.1959)

The Inequalities of Khintchine and Expanding Sphere of Their Action

G. PESKIR* *and* A. N. SHIRYAEV**

Contents:

---

* Research supported by Danish Natural Science Research Council and by Danish Research Academy.

** Research supported by Department of Theoretical Statistics (University of Aarhus) and by Steklov Mathematical Institute (Russian Academy of Sciences).

AMS 1980 *subject classifications.* Primary 60E15, 60G40, 60G42, 60G50. Secondary 42C05, 42C10, 60F10, 60G46.

Key words and phrases: The inequalities of Khintchine, Burkholder-Davis-Gundy (BDG-inequalities), Rosenthal; the strong law of large numbers and its refinements, the law of iterated logarithm; the subgaussian inequalities; the systems of Haar and Walsh; the inequalities of Paley and Marcinkiewicz-Zygmund; martingale, local martingale, supermartingale, submartingale, martingale differences, martingale transforms, stopping times, the quadratic variation, the predictable quadratic variation (quadratic characteristic), Doob's decomposition; maximal inequalities; the method of symmetrization; the method of randomization; comparison inequalities for martingale transforms; the decomposition of Davis; the domination property of Lenglart; the best constants (in inequalities), Schur-convexity (-concavity), the Schur condition; the inequalities of Khintchine in exponential Orlicz spaces; the inequalities of Khintchine for (complex) Steinhaus variables; good λ-inequality; martingale extension of the subgaussian inequality. (Second edition)

Introduction

In year 1923 A. Ya. Khintchine published the paper “Über dyadische Brüche”, [74], in which he was trying to find the right rate of convergence in the strong law of large numbers of E. Borel, and proved the following statement:

Let $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots)$ be a sequence of independent and identically distributed random variables, defined on some probability space $(\Omega, \mathcal{F}, P)$ and taking values $+1$ and -1 with probability $1/2$. Then for any integer $p = 2m$, $m \geq 1$, there exists a universal constant B_p, such that for each sequence of numbers $a = (a_1, a_2, \dots)$ and every $n \geq 1$, the following inequality is valid:

$$\text{(I)} \qquad E\Big|\sum_{k=1}^{n} a_k \varepsilon_k\Big|^p \leq B_p \Big(\sum_{k=1}^{n} |a_k|^2\Big)^{p/2} .$$

This statement, formulated in terms of probability theory (in (I) the symbol E denotes the mathematical expectation with respect to the measure P) admits the following formulation, which is accustomed in the “metrical” theory of functions.

Let $r = (r_1, r_2, \dots)$ be a system of Rademacher functions, $r_k = r_k(x)$, $k \geq 1$, defined on the set $[0,1[$ with Lebesgue measure, and determined by the equalities:

$$r_1(x) = \begin{cases} +1 \ , \ 0 \leq x < \frac{1}{2} \\ -1 \ , \ \frac{1}{2} \leq x \leq 1 \ , \end{cases}$$

$$r_k(x) = r_1\big(2^{k-1}x\big) \ , \ k \geq 2 \ .$$

Then for any integer $p = 2m$, $m \geq 1$, there exists a universal constant $\mathbf{B}_p \ \big(= B_p^{1/p}\big)$, such that for each sequence of numbers $a = (a_1, a_2, \dots)$ and every $n \geq 1$, the following inequality is valid:

$$\text{(I')} \qquad \Big\|\sum_{k=1}^{n} a_k r_k\Big\|_p \leq \mathbf{B}_p \Big\|\Big(\sum_{k=1}^{n} |a_k|^2\Big)^{1/2}\Big\|_p ,$$

where $\|\cdot\|_p$ denotes the $L_p\big([0,1[\big)$-norm: $\|\xi\|_p = \big(E|\xi|^p\big)^{1/p}$.

It is well-known (and will be shown below) that the inequalities (I), (I’) remain valid for any $0 < p < \infty$. Together with the estimate from below (which also holds for all $p > 0$):

$$\text{(II)} \qquad A_p \Big(\sum_{k=1}^{n} |a_k|^2\Big)^{p/2} \leq E\Big|\sum_{k=1}^{n} a_k \varepsilon_k\Big|^p ,$$

or equivalently:

$$\text{(II')} \qquad \mathbf{A}_p \Big\|\Big(\sum_{k=1}^{n} |a_k|^2\Big)^{1/2}\Big\|_p \leq \Big\|\sum_{k=1}^{n} a_k r_k\Big\|_p$$

with a universal constant $A_p > 0$ and $\mathbf{A}_p = A_p^{1/p}$, we arrive to the inequalities:

$$\text{(III)} \qquad A_p\Big(\sum_{k=1}^{n}|a_k|^2\Big)^{p/2} \le E\Big|\sum_{k=1}^{n}a_k\varepsilon_k\Big|^p \le B_p\Big(\sum_{k=1}^{n}|a_k|^2\Big)^{p/2},$$

or equivalently:

$$\text{(III')} \qquad \mathbf{A}_p\Big\|\Big(\sum_{k=1}^{n}|a_k|^2\Big)^{1/2}\Big\|_p \le \Big\|\sum_{k=1}^{n}a_k r_k\Big\|_p \le \mathbf{B}_p\Big\|\Big(\sum_{k=1}^{n}|a_k|^2\Big)^{1/2}\Big\|_p$$

which are widely called the "*inequalities of Khintchine*".

It is the aim of the present paper to give an exposition of the basic results, which relate to these inequalities from different points of view – their proofs, applications, clarifications, values of the best constants A_p , B_p , expanding sphere of action of the inequalities (III), (III') by replacing the sequence $\big(\sum_{k=1}^{n}a_k\varepsilon_k\big)_{n\ge1}$ or the system of Rademacher functions $r=(r_k)_{k\ge1}$ with a random sequence $f=(f_n)_{n\ge1}$ or systems of functions with a more general structure. As we shall see below, the inequalities (III) extend in a natural way to random sequences $f=(f_n)_{n\ge1}$ which appear to be martingales or sequences related to them (for example, local martingales). The next natural step should be the extension to martingale or local martingale $f=(f_t)_{t\ge0}$ for the case of continuous time.

The results exposed below will concern only the case of discrete time. From a formal point of view many of them could be deduced as consequences of more general results for continuous time, but it should be remarked, that the corresponding (continuous time) theory needs to attach rather complicated notions and results from "stochastic calculus". At the same time in the case of discrete time parameter one succeeds to obtain rather simple and transparent formulations and proofs by minimal tools. This explains that from the point of view of "expanding sphere of action of the inequalities of Khintchine" the main attention is devoted to the case of discrete time, whereas the corresponding case of continuous time is much investigated in the theory of stochastic processes.

The present exposition will be given in a chronological order which, in our opinion, gives a better understanding of those motives and aims, which led A. Ya. Khintchine and subsequent researchers in connection with "Khintchine's inequalities", to their refinements, extensions, etc.

1. The strong law of large numbers of Borel and its refinements

1. To understand, why A. Ya. Khintchine needed the inequality (I), let us turn back to the history, related to the strong law of large numbers of Borel, which in the number-theoretical context might be formulated in the following way.

Let $\Omega=[0,1[$, let $\mathcal{F}$ be the family of Borel subsets, and let P be Lebesgue measure. Consider the binary decomposition $x=0\,.\,x_1x_2\,\ldots$ of a number $x\in\Omega$. In other words, let

$$\text{(1.1)} \qquad x=\sum_{k=1}^{\infty}\frac{x_k}{2^k}\ .$$

For uniqueness of such a representation we shall consider only those decompositions which contain infinite number of zeros. For example, in between the two representations

$$\frac{1}{2}=\frac{1}{2}+\frac{0}{2^2}+\frac{0}{2^3}+\ \ldots \quad\text{and}\quad \frac{1}{2}=\frac{0}{2}+\frac{1}{2^2}+\frac{1}{2^3}+\ \ldots$$

under consideration, we allow only the first one. Putting $\xi_k(x) = x_k$, $k \geq 1$, we see, that for $a_i = 0$ or 1 :

$$P\,\{\, x : \xi_1(x) = a_1 \,,\ \ldots\ ,\ \xi_n(x) = a_n \,\}$$
$$= P\left\{\, x : \tfrac{a_1}{2} + \tfrac{a_2}{2^2} + \ldots + \tfrac{a_n}{2^n} \leq x < \tfrac{a_1}{2} + \tfrac{a_2}{2^2} + \ldots + \tfrac{a_n}{2^n} + \tfrac{1}{2^n} \,\right\}$$
$$= P\left\{\, x : x \in \left[\, \tfrac{a_1}{2} + \tfrac{a_2}{2^2} + \ldots + \tfrac{a_n}{2^n} \,,\ \tfrac{a_1}{2} + \tfrac{a_2}{2^2} + \ldots + \tfrac{a_n}{2^n} + \tfrac{1}{2^n} \,\right[\, \right\} = \tfrac{1}{2^n} \ .$$

From this it follows, that $\xi = (\xi_1, \xi_2, \ \ldots\ )$ with $\xi_k = \xi_k(x)$, $k \geq 1$, is a Bernoulli sequence of independent and identically distributed random variables, taking values 0 and 1 with probability $1/2$.

Putting $r_k(x) = 1 - 2\xi_k(x)$, $k \geq 1$, we get a system $r = (r_1, r_2, \ \ldots\ )$ of Rademacher functions $r_k = r_k(x)$, $k \geq 1$. Clearly, in this way, the random quantities r_1 , r_2 , $\ldots$ form a sequence of independent random variables with $P\{\, r_k = 1 \,\} = P\{\, r_k = -1 \,\} = 1/2$, $k \geq 1$.

The system of Rademacher functions is, obviously, orthonormal, $Er_k r_l = \delta_{kl}$, but not complete, because $Er_k \cdot 1 = 0$. By "perestroika" of this system one gets two well-known systems, of Haar and Walsh, having not only the property of orthogonality but completeness as well; see below for their definitions and formulations of a "martingale" character.

The above mentioned result of Borel (1909, [13]) states, that almost every number x from $[0,1[$ is *normal* in the sense, that with probability one the fraction of zeros and units in the binary decomposition (1.1) converges to $1/2$, that is for $n \to \infty$:

(1.2)
$$\frac{1}{n}\sum_{k=1}^{n} \xi_k(x) \to \frac{1}{2} \quad (\,P\text{-a.e.}\,)\ ,$$

or, in terms of Rademacher functions:

$$\frac{1}{n}\sum_{k=1}^{n} r_k(x) \to 0 \quad (\,P\text{-a.e.}\,)\ .$$

Denoting, in accordance to the traditions of probability theory, $S_n = r_1 + r_2 + \ldots + r_n$, we find $ES_n = 0$, $DS_n = n$ and for $n \to \infty$:

(1.3)
$$\frac{S_n}{n} \to 0 \quad (\,P\text{-a.s.}\,)\ .$$

The proof of the statement (1.2), or equivalently (1.3), which is called the strong law of large numbers of Borel, is rather simple: By the Chebyshev-Markov inequality and simple estimate $ES_n^4 \leq 3n^2$, it follows, that for any $\delta > 0$:

$$P\left(\left|\frac{S_n}{n}\right| > \delta\right) \leq \frac{ES_n^4}{n^4\delta^4} \leq \frac{3}{n^2\delta^4} \ ,$$

and thus:

$$P\left(\sup_{m\geq n}\left|\frac{S_m}{m}\right| > \delta\right) \leq \sum_{m\geq n} P\left(\left|\frac{S_m}{m}\right| > \delta\right) \leq \frac{3}{\delta^4}\sum_{m\geq n}\frac{1}{m^2} \to 0 \quad , \quad n \to \infty$$

which proves the validity of the property (1.3).

2. The next step in the clarification of the rate of convergence in (1.3) has been made by Hausdorff (1913, [61]), who noted, that for each $\varepsilon > 0$:

(1.4) $$\frac{S_n}{n^{\frac{1}{2}+\varepsilon}} \to 0 \quad (P\text{-a.s.})$$

i.e. that $S_n = o\big(n^{\frac{1}{2}+\varepsilon}\big)$ with probability one.

The proof of this statement is also rather simple and is based upon the fact that $ES_n^{2r} = O(n^r)$, from which for an integer $r > 1/2\varepsilon$ we find, that:

$$P\left(\sup_{m\geq n}\left|\frac{S_m}{m^{\frac{1}{2}+\varepsilon}}\right| > \delta\right) \leq \sum_{m\geq n} P\left(\left|\frac{S_m}{m^{\frac{1}{2}+\varepsilon}}\right| > \delta\right)$$

$$\leq \frac{1}{\delta^{2r}}\sum_{m\geq n} E\left|\frac{S_m}{m^{\frac{1}{2}+\varepsilon}}\right|^{2r} \leq \frac{C}{\delta^{2r}}\sum_{m\geq n}\frac{m^r}{m^{r+2r\varepsilon}} \to 0 \quad , \quad n\to 0$$

because $2r\varepsilon > 1$ (where C is a constant depending on r).

Remark. As noted in [126] (Theorem 4.3), already in this way it was possible to obtain the following result; with probability one for $n\to\infty$:

(1.5) $$\limsup_n \frac{|S_n|}{\sqrt{n}\log n} \leq 1 \ .$$

Indeed, because:

$$Ee^{tS_n} = \left(\frac{e^t+e^{-t}}{2}\right)^n ,$$

we get:

(1.6) $$E\exp\left(n^{-1/2}S_n\right) \to e^{1/2} \quad , \quad n\to\infty \ .$$

It follows, that for any $\varepsilon > 0$:

$$P\left(\frac{S_m}{\sqrt{m}\log m} \geq 1+\varepsilon\right) = P\left(\exp\left(\frac{S_m}{\sqrt{m}}\right) \geq \exp\Big((1+\varepsilon)\log m\Big)\right)$$

$$= P\left(\exp\left(\frac{S_m}{\sqrt{m}}\right) \geq m^{1+\varepsilon}\right) \leq \frac{E\exp\left(\frac{S_m}{\sqrt{m}}\right)}{m^{1+\varepsilon}} ,$$

and so by (1.6):

$$P\left(\sup_{m\geq n}\frac{S_m}{\sqrt{m}\log m} \geq 1+\varepsilon\right) \leq \sum_{m\geq n}\frac{E\exp\left(\frac{S_m}{\sqrt{m}}\right)}{m^{1+\varepsilon}} \to 0 \quad , \quad n\to 0 \ .$$

By symmetry:

$$P\left(\sup_{m\geq n} \frac{|S_m|}{\sqrt{m}\log m} \geq 1+\varepsilon\right) \to 0 \quad , \quad n \to 0$$

which is by the freedom of choice of $\varepsilon > 0$ equivalent to the property (1.5).

Next refinements of the rate of convergence in (1.4) have been obtained by Hardy and Littlewood (1914, [59]), who showed, that with probability one:

$$|S_n| = O\big(\sqrt{n \log n}\big) \ , \tag{1.7}$$

and by Steinhaus (1922, [140]), who established, that:

$$\limsup_n \frac{|S_n|}{\sqrt{8\, n \log n}} \leq 1 \ . \tag{1.8}$$

3. In the above mentioned paper from year 1923, [74], A. Ya. Khintchine made a next step towards the establishment of the rate of convergence in (1.2), by showing that (P-a.s.):

$$\limsup_n \frac{|S_n|}{\sqrt{2\, n \log\log n}} \leq \sqrt{2} \ . \tag{1.9}$$

In other words, he was the first who found a statement in which the iterated logarithm appears, that figures, as we now know, in the final formulation of "Khintchine's law of the iterated logarithm" (1924, [78]); with probability one:

$$\limsup_n \frac{|S_n|}{\sqrt{2\, n \log\log n}} = 1 \ . \tag{1.10}$$

From the above stated historical excursion, we see, that the refinements of the rate of convergence in (1.3) were connected with finding a "good" estimate of the probability:

$$P\big(\, |S_n| \geq t \,\big)$$

where $t = t(n)$, which was provided convergence of the series $\sum_{n=1}^{\infty} P(\, |S_n| \geq t(n)\,)$, and finally (by the Borel-Cantelli lemma) gave possibility to obtain statements (1.4), (1.5), (1.7) and (1.8). For this, in the cases considered above:

$t(n) = n$ (Borel)

$t(n) = n^{\frac{1}{2}+\varepsilon}$ (Hausdorff)

$t(n) = n^{1/2} \log n$ ((1.5); Révész)

$t(n) = n^{1/2} \log^{1/2} n$ (Hardy-Littlewood, Steinhaus).

It was the aim of obtaining a "good" estimate of probability $P\big(\, |S_n| \geq t\,\big)$ or, in a little bit more general form, probability $P\big(\, |\sum_{k=1}^{n} a_k \varepsilon_k| \geq t\,\big)$, that led A. Ya. Khintchine to prove the validity of inequalities (I), and with the help of this to get the exponential (subgaussian) estimate:

$$(1.11)\qquad P\left(\left|\sum_{k=1}^{n} a_k\varepsilon_k\right| > t\right) \le C e^{-t^2/2\sum_{k=1}^{n}|a_k|^2}$$

with some absolute constant $C\ (= e\sqrt{2} = 3.84\ldots)$.

This estimate is given by A. Ya. Khintchine in year 1923, [74], when he proved the validity of a "weak" variant of the law of the iterated logarithm (1.9). In the next year (1924) in the paper [75], A. Ya. Khintchine obtained the final formulation (1.10), in which for the estimate of probability $P(\,|S_n|\ge t\,)$ he applied a direct analysis of the binomial distribution $P(S_n=k)$.

Our exposition in the sequel will be essentially concentrated around the "inequalities of Khintchine" (I) in themselves and their extensions. Concerning the "law of iterated logarithm" itself, on the way to which have been derived the "inequalities of Khintchine", we shall here notice only that the next principle fact about the possibility of extending the law of iterated logarithm to the random sequences of a more general structure has been obtained in 1929 by A. N. Kolmogorov in [78]. His result may be formulated in the following way. *Let* $\xi_1,\xi_2,\ \ldots$ *be a sequence of independent random variables with zero mean,* $\sigma_n^2 = E\xi_n^2$, $B_n = \sum_{k=1}^{n}\sigma_k^2 \to \infty$, $n\to\infty$, *and suppose that there exist constants* M_n , $n \ge 1$, *such that:*

$$(1.12)\qquad \left|\xi_n\right| \le M_n = o\left(\left(\frac{B_n}{\log\log B_n}\right)^{1/2}\right) .$$

Then

$$(1.13)\qquad \limsup_{n\to\infty} \frac{S_n}{\sqrt{2B_n\log\log B_n}} = 1 \quad (P\text{-a.s.}).$$

(In 1937 Marcinkiewicz and Zygmund established, that in (1.12) "o-small" cannot be replaced by "O-capital".) If additionally to the independence of the quantities $\xi_1,\xi_2,\ \ldots$ we suppose also identical distribution, then as shown by Hartman and Wintner in 1941 for the validity of the law of the iterated logarithm it is sufficient only that $\sigma^2 = E\xi_1^2 < \infty$.

2. The inequalities of Khintchine

1. Let $\varepsilon = (\varepsilon_1,\varepsilon_2,\ \ldots\ )$ be a sequence of independent and identically distributed random variables, defined on the probability space $(\Omega,\mathcal{F},P)$ and taking values ± 1 with probability $1/2$. Let us denote $f = (f_1,f_2,\ \ldots\ )$ with:

$$(2.1)\qquad f_n = \sum_{k=1}^{n} a_k\varepsilon_k$$

where $a = (a_1,a_2,\ \ldots\ )$ is a sequence of numbers. Put:

$$(2.2)\qquad S_n(f) = \left(\sum_{k=1}^{n}\left(\Delta f_k\right)^2\right)^{1/2}$$

where $\Delta f_k = f_k - f_{k-1}$, $f_0 = 0$. Since $\varepsilon_k = \pm 1$, then:

(2.3) $$S_n(f) = \left(\sum_{k=1}^{n} |a_k|^2 \right)^{1/2} .$$

In this notation the inequalities (III) get the following form:

(2.4) $$A_p\big(S_n(f)\big)^p \le E|f_n|^p \le B_p\big(S_n(f)\big)^p \ , \ p > 0 \ .$$

To prove the right-hand inequality in the case $p = 2m$, $m \ge 1$, A. Ya. Khintchine, [74], writes, that:

$$E|f_n|^{2m} = \sum_{k_1+\ldots+k_n=m;\ k_i\ge 0} \frac{(2m)!}{(2k_1)!\ \ldots\ (2k_n)!}\ |a_1|^{2k_1}\ldots |a_n|^{2k_n} E|\varepsilon_1|^{2k_1}\ldots E|\varepsilon_n|^{2k_n}$$
$$= \sum_{k_1+\ldots+k_n=m;\ k_i\ge 0} \frac{(2m)!}{(2k_1)!\ \ldots\ (2k_n)!}\ |a_1|^{2k_1}\ldots |a_n|^{2k_n} \ .$$

It is evident that for $m = k_1 + \ldots + k_n$, $k_i \ge 0$:

$$2^m k_1!\ \ldots\ k_n! \le (2k_1)!\ \ldots\ (2k_n)! \ .$$

Hence:

(2.5) $$E|f_n|^{2m} \le \frac{(2m)!}{2^m m!} \sum_{k_1+\ldots+k_n=m;\ k_i\ge 0} \frac{m!}{k_1!\ \ldots\ k_n!}\ |a_1|^{2k_1}\ldots |a_n|^{2k_n}$$
$$= \frac{(2m)!}{2^m m!} \left(\sum_{k=1}^{n} |a_k|^2 \right)^m = \frac{(2m)!}{2^m m!} \big(S_n(f)\big)^{2m} ,$$

which proves the right-hand inequality in (2.4) for $p=2m$ (only such a case did A. Ya. Khintchine consider) with the constant:

$$B_{2m} = \frac{(2m)!}{2^m m!} \ .$$

which exactly coincides with the moment ($E\xi^{2m}$) of order $2m$ of a standard normal random variable ξ ($E\xi = 0$, $D\xi = 1$).

From the given consideration one can also see, that the right-hand side of the inequality (2.5) is exactly equal to $E\big|\sum_{k=1}^{n} a_k\xi_k\big|^{2m}$, where $\xi_1, \ldots, \xi_n$ are independent standard normal random variables, because:

$$E\left|\sum_{k=1}^{n} a_k\xi_k\right|^{2m} = \sum_{k_1+\ldots+k_n=m;\ k_i\ge 0} \frac{(2m)!}{(2k_1)!\ \ldots\ (2k_n)!}\ |a_1|^{2k_1}\ldots |a_n|^{2k_n} E|\xi_1|^{2k_1}\ldots E|\xi_n|^{2k_n}$$
$$= \sum_{k_1+\ldots+k_n=m;\ k_i\ge 0} \frac{(2m)!}{(2k_1)!\ \ldots\ (2k_n)!}\ \frac{(2k_1)!\ \ldots\ (2k_n)!}{2^{k_1}\ldots 2^{k_n} k_1!\ldots k_n!}\ |a_1|^{2k_1}\ldots |a_n|^{2k_n}$$
$$= \frac{(2m)!}{2^m m!} \sum_{k_1+\ldots+k_n=m;\ k_i\ge 0} \frac{m!}{k_1!\ \ldots\ k_n!}\ |a_1|^{2k_1}\ldots |a_n|^{2k_n} \ .$$

In this way one can give to the inequalities of Khintchine the following form:

$$E\Big|\sum_{k=1}^{n} a_k\varepsilon_k\Big|^{2m} \le E\Big|\sum_{k=1}^{n} a_k\xi_k\Big|^{2m}. \tag{2.6}$$

2. By Stirling's formula $\big(n! = \sqrt{2\pi n}\; n^n\, e^{-n}\, e^{R_n}\,,\ 1/(12n+1) < R_n < 1/(12n) \big)$:

$$\frac{(2m)!}{2^m m!} \le D\left(\frac{2}{e}\right)^m m^m$$

with $D = \sqrt{2}$. Thus from (2.5) by use of the Chebyshev-Markov inequality we get, that for:

$$m = \left[\frac{t^2}{2S_n^2(f)}\right],$$

one has:

$$P\big(|f_n| > t\big) \le t^{-2m} E|f_n|^{2m} \le \frac{(2m)!}{2^m m!}\, t^{-2m}\big(S_n(f)\big)^{2m}$$

$$\le D\left(\frac{2m\big(S_n(f)\big)^2}{e\,t^2}\right)^m \le De^{-m} \le De^{1-t^2/2(S_n(f))^2} = De\cdot e^{-t^2/2(S_n(f))^2}.$$

In this way from the inequality (2.5) the estimate is obtained:

$$P\big(|f_n| > t\big) \le C e^{-t^2/2(S_n(f))^2} \tag{2.7}$$

with $C = De = \sqrt{2}e = 3.84\ \ldots$ which was precisely what gave to A. Ya. Khintchine a possibility to derive the statement (1.9). (Below we will see, that the inequality (2.7) holds with the constant $C = 2$.)

Being concerned with the establishment of Khintchine's inequality:

$$E|f_n|^{2m} \le B_{2m}\big(S_n(f)\big)^{2m} \tag{2.8}$$

with the constant $B_{2m} = (2m-1)!!$ it naturally appears a question about is that constant the best possible one. It is reasonable that the answer to this question depends on are we interested in that the inequality holds true with given and fixed constant for *all* $n \ge 1$, or just for *fixed* n (when the optimal constant in (2.8) depends on it).

In Section 8 below we will present such type of results. Here we shall only note, that in 1961 S. B. Stechkin [139] showed, that in (2.8), being considered for *all* $n \ge 1$, the constant $B_{2m} = (2m-1)!!$ appears indeed to be the best possible.

3. The above exposed method of A. Ya. Khintchine which gives the right-hand inequality in (2.4) "works" only for $p = 2m$, $m \ge 1$. Let us show that the validity of these inequalities for *all* $p > 0$ follows from the inequality (2.7).

In this context it might be useful to prove independently (from the proof of inequality (2.4) for $p = 2m$) the validity of inequality (2.7), and then shortly afterwards to derive from it the inequalities (2.4) for all $p > 0$.

For this, notice, that since $\cosh x \equiv \frac{(e^x + e^{-x})}{2} \le e^{x^2/2}$, then for all $\lambda > 0$:

$$
\begin{aligned}
P\{\,|f_n| \ge t\,\} &= 2P\{\,f_n \ge t\,\} = 2P\big\{\exp(\lambda f_n) \ge \exp(\lambda t)\big\} \\
&\le 2e^{-\lambda t} E e^{\lambda f_n} = 2e^{-\lambda t} \prod_{k=1}^{n} \cosh(\lambda a_k) \le 2e^{-\lambda t} \prod_{k=1}^{n} \exp\Big(\frac{\lambda^2 |a_k|^2}{2}\Big) \\
&= 2\exp\left(\frac{\lambda^2 (S_n(f))^2}{2} - \lambda t\right) .
\end{aligned}
$$

By the freedom of choice of $\lambda > 0$ hence:

$$
P\{\,|f_n| \ge t\,\} \le \min_{\lambda>0}\ 2\exp\left(\frac{\lambda^2 (S_n(f))^2}{2} - \lambda t\right) = 2\exp\left(-\frac{t^2}{2(S_n(f))^2}\right)
$$

because the minimum is attained for $\lambda = t/(S_n(f))^2$.

So, we have just proved, that the inequality (2.7) holds indeed with $C = 2$. Turn now to the derivation of "inequalities of Khintchine" (2.4) for $p > 0$.

The case $0 < p \le 2$ for the right-hand inequality in (2.4) and the case $2 \le p < \infty$ for the left-hand inequality in (2.4) are trivial, since:

$$
E|f_n|^2 = \big(S_n(f)\big)^2
$$

and the norm $\|f\|_p = \big(E|f|^p\big)^{1/p}$ is non-decreasing in $p>0$. This moreover shows that in (2.4) one may take $B_p = 1$ for $p \le 2$ and $A_p = 1$ for $p \ge 2$.

Let now $p > 2$. Then from (2.7) (with $C=2$) we find, that:

(2.9)
$$
\begin{aligned}
E\left|\frac{f_n}{S_n(f)}\right|^p &= p\int_0^\infty t^{p-1} P\left(\left|\frac{f_n}{S_n(f)}\right| > t\right)\, dt \\
&\le 2p\int_0^\infty t^{p-1} e^{-t^2/2}\, dt = p2^{p/2}\Gamma(p/2) ,
\end{aligned}
$$

which proves the right-hand inequality in (2.4) with the constant:

$$
B_p = 2^{p/2} p\, \Gamma(p/2) \ ,\ p > 2 \ .
$$

In the case $0 < p < 2$, by the Cauchy-Bunyakovskii inequality:

$$
\begin{aligned}
1 &= E\left|\frac{f_n}{S_n(f)}\right|^2 = E\left(\left|\frac{f_n}{S_n(f)}\right|^{2-p/2}\left|\frac{f_n}{S_n(f)}\right|^{p/2}\right) \le \\
&\le \left(E\left|\frac{f_n}{S_n(f)}\right|^{4-p}\right)^{1/2}\left(E\left|\frac{f_n}{S_n(f)}\right|^{p}\right)^{1/2} \le \big(B_{4-p}\big)^{1/2} E\left(\left|\frac{f_n}{S_n(f)}\right|^{p}\right)^{1/2} ,
\end{aligned}
$$

whence:

$$
B_{4-p}^{-1} \le E\left|\frac{f_n}{S_n(f)}\right|^p ,
$$

i.e. the left-hand inequality in (2.4) holds true with the constant:

$$A_p = B_{4-p}^{-1} .$$

This completes the proof of the "inequalities of Khintchine" (2.4) with the constants A_p and B_p , $p > 0$, given by:

$$B_p = \begin{cases} 1 & , \quad p \le 2 \\ 2^{p/2} p\,\Gamma(p/2) & , \quad p > 2 \end{cases} \tag{2.10}$$

$$A_p = \begin{cases} B_{4-p}^{-1} & , \quad p < 2 \\ 1 & , \quad p \ge 2 . \end{cases} \tag{2.11}$$

In Section 8 the best values for the constants A_p and B_p in the inequalities of Khintchine will be considered (compare (2.10) and (2.11) with (8.18) and (8.19)).

4. In the next two sections it will be considered a generalization of the inequalities of Khintchine (2.4) to the case of martingale sequence (f_n) . But before coming to that generalization, we shall consider one more interesting "expanding sphere of action of Khintchine's inequalities", connected with an observation that the inequality (2.5) can take form (2.6), where on the right-hand side of that inequality $\xi_1, \dots, \xi_n$ are independent normally distributed random variables with $E\xi_k = 0$ and $D\xi_k = 1$, $k \ge 1$. Having the inequality (2.6) it is natural to pose a question if it remains valid if one assumes that $\xi_1, \dots, \xi_n$ belong to the class Σ_n consisting of independent identically distributed random variables with symmetric distribution and $E\xi_k = 0$, $D\xi_k = 1$, $1 \le k \le n$.

Such a setting of question was considered by Utev [148], 1984, Pinelis [119], 1994, and the authors (T. Figiel, P. Hitczenko, W. B. Johnson, G. Schechtman, J. Zinn) in the article [41], 1994. From the results of these works it follows, that for $p = 2$ and $p \ge 3$ and for $(\xi_1, \dots, \xi_n) \in \Sigma_n$:

$$E\Big|\sum_{i=1}^n a_i\varepsilon_i\Big|^p \le E\Big|\sum_{i=1}^n a_i\xi_i\Big|^p ,$$

and, consequently:

$$E\Big|\sum_{i=1}^n a_i\varepsilon_i\Big|^p = \inf_{(\xi_1,\dots,\xi_n)\in\Sigma_n} E\Big|\sum_{i=1}^n a_i\xi_i\Big|^p ,$$

which shows a definite extremal role of Rademacher functions.

3. Martingale extensions of Khintchine's inequalities I (preliminary considerations)

1. Let us begin with two results from the classic theory of orthogonal series. It is well-known, that with the system of Rademacher functions $r = (r_1, r_2, \dots)$ one connects two remarkable orthogonal and complete systems of functions, namely of Haar $h = (h_1, h_2, \dots)$ and of Walsh $w = (w_1, w_2, \dots)$, which are defined in the following way: For $x \in [0,1]$:

$$h_1(x) = 1 , \; h_2(x) = r_1(x) , \tag{3.1}$$

$$h_n(x) = \begin{cases} 2^{j/2} r_{j+1} & \text{if } \frac{k-1}{2^j} \le x < \frac{k}{2^j} , \; n = 2^j + k , \; 1 \le k \le 2^j , \; j \ge 1 \\ 0 & \text{otherwise} , \end{cases}$$

and:

(3.2) $$w_1(x) = 1 \ ,$$

$$w_n(x) = r_{n_1}(x)\dots r_{n_k}(x) \ , \text{if } \ n = 2^{n_1}+\dots+2^{n_k}+1 \ , \text{where } \ n_1 > \dots > n_k \geq 0 \ .$$

Let the function $\varphi \in L^2\big([0,1]\big)$ or $\varphi \in L^1\big([0,1]\big)$ with $E\varphi = 0$. Denote:

(3.3) $$\varphi_n(h) = \sum_{k=1}^{n}(\varphi, h_k)h_k \quad , \quad \varphi_n(w) = \sum_{k=1}^{n}(\varphi, w_k)w_k$$

where (φ, h_k) and (φ, w_k) are coefficients of Fourier-Haar and Fourier-Walsh respectively.

In year 1932 Paley, [114], showed, that if:

$$f_n = \varphi_{2^n}(w) \ , \ n \geq 1 \ ,$$

then for all $1 < p < \infty$ and $n \geq 1$ there are such universal constants $\mathbf{A}_p$ and $\mathbf{B}_p$, that:

(3.4) $$\mathbf{A}_p \, \big\|S_n(f)\big\|_p \leq \big\|f_n\big\|_p \leq \mathbf{B}_p \, \big\|S_n(f)\big\|_p \ ,$$

where:

$$S_n(f) = \left(\sum_{k=1}^{n}\big(\Delta f_k\big)^2\right)^{1/2}, \quad \Delta f_k = f_k - f_{k-1} \ .$$

In year 1937 Marcinkiewicz, [96], noted, that the result of Paley (3.4) (for Walsh functions) follows from the validity of (3.4) for Haar functions (with $f_n = \varphi_{2^n}(h)$) and the fact (recorded also by Paley in [114] and initially proved by Walsh, [149], in year 1923) that $\varphi_{2^n}(h) = \varphi_{2^n}(w)$.

In year 1938 Marcinkiewicz and Zygmund, [97], showed, that if $\xi = (\xi_1, \xi, \ \dots)$ is a sequence of independent random variables with $E\xi_i = 0$, $i \geq 0$, then for all $p \geq 1$ one finds such universal constants $\mathbf{A}_p$ and $\mathbf{B}_p$, that again the inequality (3.4) holds with $f_n = \xi_1 + \dots + \xi_n$ and:

$$S_n(f) = \left(\sum_{k=1}^{n}\xi_k^2\right)^{1/2} \ .$$

2. Apparently, D. Burkholder and R. Gundy were the first, who realized ([14], [15], [52]) that sequences $f = (f_1, f_2, \ \dots \ )$, for which one obtains inequalities of type (3.4), both in the case of Khintchine $\big(f_n = \sum_{k=1}^{n} a_k\varepsilon_k\big)$, and in the cases of Paley $\big(f_n = \varphi_{2^n}(w)\big)$, Marcinkiewicz $\big(f_n = \varphi_{2^n}(h)\big)$, Marcinkiewicz and Zygmund $\big(f_n = \sum_{k=1}^{n}\xi_k\big)$, posses one important and remarkable property – all of them appear to be *martingales*. So, the two-sided inequalities considered above appear to be inequalities for special classes of martingales.

It was this circumstance that defined "martingale" direction of investigation of the validity of the inequalities of type (3.4), where basic and fundamental results were obtained by D. Burkholder, R. Gundy, B. Davis, with whose names are connected the so-called "BDG-inequalities" (see Section 6 below), first roots of which were the inequalities of Khintchine.

3. Let us recall needed definitions of martingales and related notions. Let $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n\geq 0}, P)$ be a filtred probability space $(\Omega, \mathcal{F}, P)$, equipped with a non-decreasing family (filtration) of

σ-algebras $(\mathcal{F}_n)_{n\geq 0}$, such that $\mathcal{F}_m \subset \mathcal{F}_n \subset \mathcal{F}$, $m \leq n$.

Definition 1. A stochastic sequence $X = (X_n)_{n\geq 0}$ of random variables $X_n = X_n(\omega)$ is called a *martingale*, if:

(i) X_n is $\mathcal{F}_n$-measurable, $n \geq 0$;

(ii) $E|X_n| < \infty$, $n \geq 0$;

(iii) the "martingale property" is satisfied: P-a.s.

$$E(X_{n+1}\,|\,\mathcal{F}_n) = X_n \ , \ n \geq 0 \ , \tag{3.5}$$

where $E(X_{n+1}\,|\,\mathcal{F}_n)$ is the conditional expectation of X_{n+1} with respect to $\mathcal{F}_n$.

Definition 2. A stochastic sequence $X = (X_n)_{n\geq 0}$ of random variables $X_n = X_n(\omega)$ is called a *local* martingale, if:

(i) X_n is $\mathcal{F}_n$-measurable, $n \geq 0$;

(ii) the conditional expectation $E(X_{n+1}\,|\,\mathcal{F}_n)$ is well-defined for all $n \geq 0$;

(iii) the "martingale property" (3.5) is satisfied.

(We say, that the conditional expectation $E(X_{n+1}\mid\mathcal{F}_n)$ is well-defined if the set $\{\,\omega\,:\,E(X^+_{n+1}\mid\mathcal{F}_n) < \infty\,\} \cup \{\,\omega\,:\,E(X^-_{n+1}\mid\mathcal{F}_n) < \infty\,\}$ coincides with Ω up to a set of P-probability zero. In this case we set $E(X_{n+1}\,|\,\mathcal{F}_n) = E(X^+_{n+1}\,|\,\mathcal{F}_n) - E(X^-_{n+1}\,|\,\mathcal{F}_n)$; for more details see [138], VII, §1.)

Definition 3. A random variable $\tau = \tau(\omega)$ with values in the set $\{0, 1, \ldots, +\infty\}$ is called a *Markov time*, if for each $n \geq 0$:

$$\{\omega\,:\,\tau \leq n\} \in \mathcal{F}_n \ .$$

If, moreover, $P\{\tau < \infty\} = 1$, then we say that $\tau = \tau(\omega)$ is a *stopping time*.

With the help of stopping times one obtains the following criterium for when a stochastic sequence forms a local martingale: A given sequence $X = (X_n)_{n\geq 0}$ of $\mathcal{F}_n$-measurable random variables X_n forms a local martingale if and only if there exists a sequence of stopping times $(\tau_k)_{k\geq 1}$, such that $\tau_k \uparrow \infty$ and for each $k \geq 1$ the "stopped" sequence $X^{(k)} = (X_{n\wedge\tau_k})_{n\geq 0}$ is a martingale (see [138], VII, §1, Theorem 1).

If $\mathcal{M}$ denotes the class of all martingales, and $\mathcal{M}_{loc}$ the class of all local martingales, then evidently:

$$\mathcal{M} \subset \mathcal{M}_{loc} \ .$$

The above given criterium makes it possible that the proof of some or the other property of local martingales is reduced (by corresponding localization) to the case of martingales. This fact clarifies why most of the results in the sequel will be formulated for martingales although they extend to the case of local martingales as well.

If in (3.5) the equality is replaced by the inequality $\geq$ ($\leq$) , then one says, that the sequence X forms a *submartingale* (*supermartingale*).

As in the case of Khintchine $\big(f_n=\sum_{k=1}^{n} a_k\varepsilon_k\big)$, as well as in the case of Marcinkiewicz and Zygmund $\big(f_n = \sum_{k=1}^{n} \xi_k\big)$, the martingale property of the sequence $f = (f_n)_{n\geq 1}$ is evident from the independence of its components and properties $E|a_k\varepsilon_k| = |a_k| < \infty$, $E|\xi_k| < \infty$ and $E(a_k\varepsilon_k) = 0$, $E(\xi_k) = 0$.

This is less evident in the case of the system of Haar, where $f_n = \varphi_n(h) = \sum_{k=1}^{n}(\varphi, h_k)h_k$. However, one could observe, that if $\mathcal{F}_n = \sigma(h_1,\dots,h_n)$, then $E(\varphi \mid \mathcal{F}_n)$ coincides (P-a.s.) with f_n :

$$f_n = E(\varphi \mid \mathcal{F}_n) \ , \ n \geq 1 \ .$$

Hence, immediately, it follows that $f = (f_n)_{n\geq 1}$ forms (with respect to the family $(\mathcal{F}_n)_{n\geq 1}$) a (Lévy) martingale, since:

$$E(f_{n+1} \mid \mathcal{F}_n) = E\big(E(\varphi \mid \mathcal{F}_{n+1}) \mid \mathcal{F}_n\big) = E(\varphi \mid \mathcal{F}_n) = f_n \ .$$

The case of the "Walsh system" reduces to the case of the "Haar system", since, as already noted above, $\varphi_{2^n}(h) = \varphi_{2^n}(w)$ and (as it is easily seen) the sequence $\big(\varphi_{2^n}(h)\big)$ is a martingale (with respect to $\big(\mathcal{F}_{2^n}\big)$) .

4. Martingale extensions of Khintchine's inequalities II (the inequalities of Burkholder: $p > 1$)

1. Let us suppose, that $f = (f_n)_{n\geq 0}$ is a martingale, defined on a filtred probability space $\big(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n\geq 0}, P\big)$ with $f_0 = 0$, $\mathcal{F}_0 = \{\emptyset, \Omega\}$. So $Ef_n = 0$, $n \geq 1$. Putting $d_0 = 0$ and:

$$d_n = f_n - f_{n-1} \ (\equiv \Delta f_n \ ) \ , \ n \geq 1 \ ,$$

we obtain a sequence $d = (d_n)_{n\geq 1}$ associated with the martingale $f = (f_n)_{n\geq 1}$, which is called a *martingale difference* sequence; d_n is $\mathcal{F}_n$-measurable, $E|d_n| < \infty$ and:

$$E(d_{n+1} \mid \mathcal{F}_n) = 0 \ .$$

Put:

$$S_n(f) = \Big(\sum_{k=1}^{n} \big(\Delta f_k\big)^2\Big)^{1/2} = \Big(\sum_{k=1}^{n} d_k^2\Big)^{1/2} \ . \tag{4.1}$$

In the theory of martingales the quantity:

$$[f]_n = \sum_{k=1}^{n} \big(\Delta f_k\big)^2 \tag{4.2}$$

is widely called the *quadratic variation* of the sequence $(f_1,\dots,f_n)$. In the square-integrable case ($E|f|^2 < \infty$, $n \geq 1$) a significant role is played by the *quadratic characteristic*:

$$\langle f \rangle_n = \sum_{k=1}^{n} E\Big((\Delta f_k)^2 \mid \mathcal{F}_{k-1}\Big) \ , \tag{4.3}$$

for which we have the following important properties:

(i) $\langle f \rangle_n$ is $\mathcal{F}_{n-1}$-measurable;

(ii) $\left(f_n^2 - \langle f \rangle_n\right)_{n\geq 1}$ is a martingale;

(iii) $\left([f]_n - \langle f \rangle_n\right)_{n\geq 1}$ is a martingale.

Property (i), which is evident from definition (4.3), is also expressed by saying that the sequence $\left(\langle f \rangle_n\right)_{n\geq 1}$ is *predictable* (i.e. $\langle f \rangle_n$ is $\mathcal{F}_{n-1}$-measurable). Property (ii) is checked directly and its equivalent formulation is the fact, that the submartingale $(f_n^2)_{n\geq 1}$ admits *Doob's decomposition*:

$$f_n^2 = \langle f \rangle_n + M_n \ , \ n \geq 1 \tag{4.4}$$

with predictable sequence $\left(\langle f \rangle_n\right)_{n\geq 1}$ and martingale $(M_n)_{n\geq 1}$. Property (iii) is established by a straightforward verification. It is clear from (4.1) and (4.2), that:

$$[f]_n = S_n^2(f) \ . \tag{4.5}$$

All of the above mentioned inequalities have the following form (compare with (3.4)):

$$A_p ES_n^p(f) \leq E|f_n|^p \leq B_p ES_n^p(f) \ , \tag{4.6}$$

herewith in the case of Khintchine they hold for $p > 0$, in the case of Marcinkiewicz-Zygmund for $p \geq 1$, and in the case of Haar and Walsh for $p > 1$.

For which values of p the result (4.6) remain valid for an *arbitrary* martingale $f = (f_n)_{n\geq 1}$? Basic results in this direction were obtained by Burkholder, [14], who found, that the inequality (4.6) holds (with some universal constants A_p and B_p) *for all* $p > 1$.

From the examples in Section 6 below it will be clear, that it is impossible to extend the validity of the left-hand inequality in (4.6) for $p = 1$. However, as established by B. Davis, [31], for $p = 1$ the proper form corresponding to the analogue of the inequality (4.6) is the following one:

$$K_1^* ES_n(f) \leq E \max_{1\leq k\leq n} |f_k| \leq L_1^* ES_n(f) \ . \tag{4.7}$$

By the well-known inequality of Doob for submartingale $\left(|f_n|^p\right)_{n\geq 1}$ (see [138], VII, §3, Theorem 1) for any $p > 1$:

$$E|f_n|^p \leq E \max_{1\leq k\leq n} |f_n|^p \leq \left(\frac{p}{p-1}\right)^p E|f_n|^p \tag{4.8}$$

so, the result of Burkholder (for $p > 1$) and the result of Davis (for $p = 1$) admit the coupling formulation: For all $p \geq 1$ there exist such universal constants K_p^* and L_p^* (not dependent neither on n, nor on the martingale f), that:

$$K_p^* ES_n^p(f) \leq E \max_{1\leq k\leq n} |f_n|^p \leq L_p^* ES_n^p(f) \ , \tag{4.9}$$

or in the equivalent forms:

$$(4.10) \qquad \mathbf{K}_p^* \big\| S_n(f) \big\|_p \le \Big\| \max_{1\le k\le n} |f_n| \Big\|_p \le \mathbf{L}_p^* \big\| S_n(f) \big\|_p$$

$$(4.11) \qquad \mathbf{K}_p^* \Big\| \sqrt{[f]_n} \Big\|_p \le \Big\| \max_{1\le k\le n} |f_n| \Big\|_p \le \mathbf{L}_p^* \Big\| \sqrt{[f]_n} \Big\|_p ,$$

where $\mathbf{K}_p^* = K_p^{*1/p}$ and $\mathbf{L}_p^* = L_p^{*1/p}$.

2. Let us turn to the basic steps in the proof of the inequality (4.6), established by Burkholder for $p > 1$. From the technical point of view the proof of Burkholder is rather complicated. The scheme of the proof proposed below differs from the traditional one, and, it seems to us, better clarifies the key role of the initial "inequalities of Khintchine". In the proof one establishes the validity of the inequality (4.6) for arbitrary martingales.

Let $\varepsilon = (\varepsilon_k)_{k\ge 1}$ be a Rademacher system. According to the inequalities of Khintchine for $p > 0$:

$$(4.12) \qquad A_p \Big(\sum_{k=1}^n |a_k|^2 \Big)^{p/2} \overset{\{1\}}{\le} E \Big| \sum_{k=1}^n a_k \varepsilon_k \Big|^p \overset{\{2\}}{\le} B_p \Big(\sum_{k=1}^n |a_k|^2 \Big)^{p/2} .$$

Suppose, in addition, that we can prove, that for any sequence of numbers $b = (b_k)_{k\ge 1}$ with $b_k = \pm 1$ and for the given martingale $f = (f_n)_{n\ge 0}$ the following inequality is valid:

$$(4.13) \qquad F_p\, E\big|f_n\big|^p \overset{\{3\}}{\le} E \Big| \sum_{k=1}^n b_k d_k \Big|^p \overset{\{4\}}{\le} G_p\, E\big|f_n\big|^p$$

with some (universal, i.e. not dependent on n and f) constants F_p and G_p .

Note, that the sequence:

$$f(b) = \big(f_n(b)\big)_{n\ge 1} ,$$

with $f_n(b) = \sum_{k=1}^n b_k d_k \ \big(= \sum_{k=1}^n b_k \Delta f_k \big)$ is also a martingale, and (4.13) can be written in the form:

$$(4.14) \qquad F_p\, E\big|f_n\big|^p \le E\big|f_n(b)\big|^p \le G_p\, E\big|f_n\big|^p ,$$

and might be viewed as a "comparison inequality" for martingales f and $f(b)$.

It will be seen from the sequel, that inequalities (4.14) are valid only for $p > 1$. (In the case $p = 1$ we will have "maximal" inequalities (6.2).)

Let us now show, how the "inequalities of Khintchine (4.12)" + "comparison inequalities (4.14)" together with the ideas of *randomization* lead (for $p > 1$) to the "inequality of Burkholder".

Having the probability space $(\Omega, \mathcal{F}, P)$ and martingale $f = \big(f_n(\omega)\big)_{n\ge 0}$ defined on it, consider a new probability space $(\Omega_\varepsilon, \mathcal{F}_\varepsilon, P_\varepsilon)$ and a sequence of Rademacher functions $\varepsilon = \big(\varepsilon_k(\omega_\varepsilon)\big)_{k\ge 0}$ defined on it.

On $\big(\Omega \times \Omega_\varepsilon, \mathcal{F} \times \mathcal{F}_\varepsilon, P \times P_\varepsilon\big)$ define random variables $\varepsilon_k(\omega, \omega_\varepsilon)$ and $d_k(\omega, \omega_\varepsilon)$, by putting:

$$\varepsilon_k(\omega, \omega_\varepsilon) = \varepsilon_k(\omega_\varepsilon) \quad \text{and} \quad d_k(\omega, \omega_\varepsilon) = d_k(\omega) .$$

It is clear, that the sequences:

$$\big(\varepsilon_k(\omega,\omega_\varepsilon)\big)_{k\geq 1} \quad \text{and} \quad \big(d_k(\omega,\omega_\varepsilon)\big)_{k\geq 1}$$

are independent with respect to the measure $P\times P_\varepsilon$. Using this fact and applying the inequalities of Khintchine with $a_k = d_k(\omega)$, $\varepsilon_k = \varepsilon_k(\omega_\varepsilon)$ and comparison inequalities with $b_k = \varepsilon_k(\omega_\varepsilon)$, $d_k = d_k(\omega)$ we find, that:

$$(4.15) \qquad A_p E\Big|\sum_{k=1}^n d_k^2(\omega)\Big|^{p/2} \overset{by\ \{1\}}{\leq} EE_\varepsilon\Big|\sum_{k=1}^n d_k(\omega)\varepsilon_k(\omega_\varepsilon)\Big|^p$$

$$= EE_\varepsilon\Big|\sum_{k=1}^n d_k(\omega,\omega_\varepsilon)\varepsilon_k(\omega,\omega_\varepsilon)\Big|^p \overset{(Fubini)}{=} E_\varepsilon E\Big|\sum_{k=1}^n \varepsilon_k(\omega_\varepsilon)d_k(\omega)\Big|^p$$

$$\overset{by\ \{4\}}{\leq} G_p E_\varepsilon E\Big|\sum_{k=1}^n d_k(\omega)\Big|^p = G_p E\Big|\sum_{k=1}^n d_k(\omega)\Big|^p$$

$$= G_p\, E\times E_\varepsilon\Big|\sum_{k=1}^n d_k(\omega,\omega_\varepsilon)\Big|^p \overset{by\ \{3\}}{\leq} \frac{G_p}{F_p}\, E\times E_\varepsilon\Big|\sum_{k=1}^n d_k(\omega,\omega_\varepsilon)\varepsilon_k(\omega,\omega_\varepsilon)\Big|^p$$

$$= \frac{G_p}{F_p}\, EE_\varepsilon\Big|\sum_{k=1}^n d_k(\omega)\varepsilon_k(\omega_\varepsilon)\Big|^p \overset{by\ \{2\}}{\leq} \frac{G_p}{F_p}\, B_p\, E\Big|\sum_{k=1}^n d_k^2(\omega)\Big|^{p/2}.$$

In this way it is showed, that:

$$\frac{A_p}{G_p}\, E\Big|\sum_{k=1}^n d_k^2(\omega)\Big|^{p/2} \leq E\Big|\sum_{k=1}^n d_k(\omega)\Big|^p \leq \frac{B_p}{F_p}\, E\Big|\sum_{k=1}^n d_k^2(\omega)\Big|^{p/2},$$

(for $p>1$, since the comparison inequalities are valid exactly for such values of p ; see Section 7 below).

3. Let us here indicate a particular case (Marcinkiewicz-Zygmund) of the validity of comparison inequalities (4.13) (for all $p\geq 1$).

Namely, suppose, that $d = (d_n)_{n\geq 1}$ is a sequence of *independent* random variables with $Ed_n = 0$. It is clear, that $f = (f_n)_{n\geq 0}$ is a martingale.

In tis case the inequalities (4.13) are easily proved, by applying the "method of symmetrization", which is as follows.

Together with $(\Omega,\mathcal{F},P)$ consider a new probability space $(\Omega',\mathcal{F}',P')$ and on it defined a sequence of independent random variables $d' = (d'_n)_{n\geq 1}$ having the same distribution as $d = (d_n)_{n\geq 1}$:

$$Law\big(d' \mid P\big) = Law\big(d \mid P\big)\ .$$

Then, defining on $\big(\Omega\times\Omega',\mathcal{F}\times\mathcal{F}',P\times P'\big)$ variables $d_k(\omega,\omega') = d_k(\omega)$ and $d'_k(\omega,\omega') = d'_k(\omega')$ and taking into account that $|x+y|^p \leq 2^{p-1}\big(|x|^p+|y|^p\big)$, $x,y\in\mathbf{R}$, $p\geq 1$, we find ($b_k = \pm 1$):

$$E\Big|\sum_{k=1}^{n} b_k d_k(\omega)\Big|^p = E\Big|\sum_{k=1}^{n} b_k d_k(\omega) - E'\big(b_k d'_k\big)\Big|^p$$

$$= E\Big|E'\sum_{k=1}^{n} b_k\big(d_k(\omega,\omega') - d'_k(\omega,\omega')\big)\Big|^p \overset{(Jensen)}{\leq} EE'\Big|\sum_{k=1}^{n} b_k\big(d_k(\omega,\omega') - d'_k(\omega,\omega')\big)\Big|^p$$

$$\overset{\{a\}}{=} E\times E'\Big|\sum_{k=1}^{n}\big(d_k(\omega,\omega') - d'_k(\omega,\omega')\big)\Big|^p \leq 2^{p-1}\left(E\Big|\sum_{k=1}^{n} d_k\Big|^p + E'\Big|\sum_{k=1}^{n} d'_k\Big|^p\right)$$

$$= 2^p\, E\Big|\sum_{k=1}^{n} d_k\Big|^p ,$$

where in {a} we use the symmetry of the distribution of the sequence $d - d'$, which leads to the fact, that:

$$Law\Big(b_1(d_1 - d'_1), b_2(d_2 - d'_2),\ \dots\ |P\times P'\Big) = Law\Big(d_1 - d'_1, d_2 - d'_2,\ \dots\ |P\times P'\Big)\ .$$

In this way the right-hand inequality in (4.13) is proved with $G_p = 2^p$. To prove the left-hand inequality, it is sufficient to observe, that if $b_k = \pm 1$, then by the above proved:

$$E|f_n|^p = E\Big|\sum_{k=1}^{n} d_k\Big|^p = E\Big|\sum_{k=1}^{n} b_k\big(b_k d_k\big)\Big|^p \leq 2^p E\Big|\sum_{k=1}^{n} b_k d_k\Big|^p ,$$

from where the left-hand inequality in (4.13) follows with $F_p = 2^{-p}$.

Unfortunately, such a simple proof of comparison inequalities (4.13) relies heavily upon the assumption of independence of the random variables d_k , $k \geq 1$, and does not "work" in the case, when random variables d_k , $k \geq 1$, form only a martingale difference sequence.

5. The maximal inequalities of Khintchine

1. In the scheme, considered by A. Ya. Khintchine, besides the inequalities (III) (for $p > 0$) the validity of the following maximal inequalities (which are not considered by Khintchine himself) can be established: For $p > 0$ there exist such universal constants A^*_p and B^*_p , that:

(5.1)
$$A^*_p\left(\sum_{k=1}^{n}|a_k|^2\right)^{p/2} \overset{\{1^*\}}{\leq} E\left(\max_{1\leq m\leq n}\Big|\sum_{k=1}^{m} a_k\varepsilon_k\Big|^p\right) \overset{\{2^*\}}{\leq} B^*_p\left(\sum_{k=1}^{n}|a_k|^2\right)^{p/2} .$$

If we denote $f_k = \sum_{i=1}^{n} a_i\varepsilon_i$ and:

$$f^*_n = \max_{1\leq k\leq n}|f_k| ,$$

then (5.1) can be written in the following form:

(5.2)
$$A^*_p\big(S_n(f)\big)^p \leq E\big(f^*_n\big)^p \leq B^*_p\big(S_n(f)\big)^p ,$$

where $S_n(f)$ is defined by (2.3).

It is clear that by the (ordinary) left-hand inequality of Khintchine in (2.4) we get the left-hand inequality in (5.2) with $A_p^* = A_p$, since $|f_n| \le f_n^*$.

For the proof of the right-hand inequality in (5.2) we shall use the well-known maximal inequality of Lévy:

(5.3) $$P\Big\{ \max_{1\le k\le n} |S_k| > t \Big\} \le 2P\{ |S_n| > t \} ,$$

which holds for sums $S_n = \xi_1 + \ldots + \xi_n$ of independent and symmetrically distributed ($P(\xi_k \in B) = P(-\xi_k \in B)$, $B \in \mathcal{B}(\mathbf{R})$) random variables (see for instance [138], IV, §4).

Putting $\xi_k = a_k \varepsilon_k$, $k = 1, \ldots, n$, and hence using the fact, that for $p \ge 2$:

$$E\big(f_n^*\big)^p = \int_0^\infty p\, t^{p-1} P\{f_n^* > t\}\; dt \le 2 \int_0^\infty p\, t^{p-1} P\{|f_n| > t\}\; dt = 2E|f_n|^p ,$$

from the right-hand inequality in (2.4) we get (for $p \ge 2$) the right-hand inequality in (5.2) with the constant $B_p^* = 2B_p$.

Concerning the case $0 < p < 2$, we should note, that for such p's by Jensen's inequality:

$$E\big(f_n^*\big)^p \le \Big(E(f_n^*)^2\Big)^{p/2} ,$$

and, so, this case is reduced to the case $p = 2$. In this:

$$B_p^* = \big(2B_p\big)^{p/2} = 2^{p/2} ,$$

since $B_2 = 1$.

2. The question about the best constants A_p^* and B_p^* in the maximal inequalities of Khintchine (5.1) (still) remains open. Some aspects in this direction may be found in the works [37], [38], [50], [51], [68], [117].

6. Martingale extensions of maximal Khintchine's inequalities (the inequalities of Davis: $p=1$; the inequalities of Burkholder-Davis: $p \ge 1$)

1. In the case $p > 1$ the inequalities of Khintchine and comparison inequalities (4.14) gave a possibility to obtain the inequalities of Burkholder (4.6) for martingales $f = (f_n)_{n\ge1}$.

As already noted above, in the case $p = 1$ the inequalities (4.6) generally fail to hold. Here is the corresponding example.

Let $d = (d_n)_{n\ge1}$ be a Bernoulli sequence of independent and identically distributed random variables, taking values ±1 with probability $1/2$. Put $f_n = d_1 + \ldots + d_n$ and:

$$\tau = \min\{ n \ge 1 : f_n = 1 \} .$$

It is well-known, that $P\{\tau < \infty\} = 1$, but $E\tau^{1/2} = \infty$. Using the martingale $f = (f_n)_{n\ge1}$ and stopping time τ we shall now construct a new martingale $g = (g_n)_{n\ge1}$, by putting:

$$g_n = f_{n\wedge\tau} = \left(\sum_{k=1}^{n\wedge\tau} d_k = \sum_{k=1}^{n} I(k \le \tau) d_k \right) .$$

Then:

(6.1)
$$E|g_n| = E(2g_n^+ - g_n) = 2Eg_n^+ \to 2 \quad , \quad n \to \infty \ ,$$

since $0 \le g_n^+ \le 1$ and $g_n^+ \to 1$ (P-a.s.). On the other hand:

$$ES_n(g) = E\sqrt{\tau \wedge n} \to \infty \quad , \quad n \to \infty \ .$$

Comparing this with (6.1) we see, that the left-hand inequality in (4.6) cannot be satisfied with a single constant A_1 , which does not depend on n .

As already remarked in Section 4, Davis, [31], have discovered, that the proper shape of the corresponding analogue of the inequality (4.6) has the form (4.7) and then for all $p \ge 1$ the maximal inequalities must have the form (4.9).

2. Let us now show, that the proof of these inequalities can be obtained (upon the same scheme as the inequalities (4.6) with $p > 1$) from:

(i) maximal inequalities of Khintchine (5.1) for the system of Rademacher functions (playing the role of "randomization" of the martingale f); and

(ii) maximal comparison inequalities for martingale transforms (compare with (4.14)):

(6.2)
$$F_p^* E(f_n^*)^p \overset{\{3^*\}}{\le} E(f_n^*(b)) \overset{\{4^*\}}{\le} G_p^* E(f_n^*)^p$$

where $f_n^* = \max_{1 \le m \le n} |f_m|$, $f_m(b) = \sum_{k=1}^m b_k d_k$ and $b = (b_i)_{i \ge 1}$ is an arbitrary sequence of numbers with $b_i = \pm 1$.

Indeed, by using the same notation as in the chain of inequalities (4.15), we find that for $p \ge 1$ ($\{1^*\}$, $\{2^*\}$, $\{3^*\}$ and $\{4^*\}$ are defined in (5.1) and (6.2)):

(6.3)
$$\begin{aligned}
A_p^* ES_n^p(f) &\overset{by\ \{1^*\}}{\le} EE_\varepsilon \left(\max_{1 \le m \le n} \Big| \sum_{k=1}^m d_k(\omega)\varepsilon_k(\omega_\varepsilon) \Big|^p \right) \\
&= EE_\varepsilon \left(\max_{1 \le m \le n} \Big| \sum_{k=1}^m d_k(\omega,\omega_\varepsilon)\varepsilon_k(\omega,\omega_\varepsilon) \Big|^p \right) \\
&\overset{(Fubini)}{\le} E_\varepsilon E \left(\max_{1 \le m \le n} \Big| \sum_{k=1}^m \varepsilon_k(\omega_\varepsilon) d_k(\omega) \Big|^p \right) \\
&\overset{by\ \{4^*\}}{\le} G_p^* \, E_\varepsilon E \left(\max_{1 \le m \le n} \Big| \sum_{k=1}^m d_k(\omega) \Big|^p \right) = G_p^* \, E(f_n^*)^p \\
&= G_p^* \, E \times E_\varepsilon \left(\max_{1 \le m \le n} \Big| \sum_{k=1}^m d_k(\omega,\omega_\varepsilon) \Big|^p \right) \\
&\overset{by\ \{3^*\}}{\le} \frac{G_p^*}{F_p^*} \, E \times E_\varepsilon \left(\max_{1 \le m \le n} \Big| \sum_{k=1}^m d_k(\omega,\omega_\varepsilon)\varepsilon_k(\omega,\omega_\varepsilon) \Big|^p \right)
\end{aligned}$$

$$= \frac{G_p^*}{F_p^*} EE_\varepsilon \Big(\max_{1 \le m \le n} \Big| \sum_{k=1}^m d_k(\omega)\varepsilon_k(\omega_\varepsilon) \Big|^p \Big)$$

$$\overset{by\ \{2^*\}}{\le} \frac{G_p^*}{F_p^*} B_p^* \, E \Big| \sum_{k=1}^m d_k^2(\omega) \Big|^{p/2} = \frac{G_p^*}{F_p^*} B_p^* \, ES_n^p(f) \ .$$

In this way we get for all $p \ge 1$:

$$K_p^* ES_n^p(f) \le E\big(f_n^*\big)^p \le L_p^* ES_n^p(f) \ ,$$

with:

$$K_p^* = \frac{A_p^*}{G_p^*} \ , \ L_p^* = \frac{B_p^*}{F_p^*} \ ,$$

where (A_p^*, B_p^*) and (G_p^*, F_p^*) are the constants from the maximal inequalities of Khintchine (5.2) and maximal comparison inequalities (6.2).

7. Comparison inequalities for martingale transforms

1. The above exposed scheme of the proof of inequalities of the form (4.6) or (5.2) clarifies the key role of the two ingredients – Khintchine's inequalities ((2.4), (5.2)) and comparison inequalities for martingale transforms (4.14), (6.2)).

Since the situation with Khintchine's inequalities is clear, let us address the question of comparison inequalities.

In the proofs given above (see (4.15) and (6.3)) we have used the martingale transform:

$$f(b) = \big(f_n(b)\big)_{n \ge 1} \ ,$$

of a rather special form:

$$f_n(b) = \sum_{k=1}^n b_k d_k \ , \tag{7.1}$$

where the numbers b_k take the two values ± 1 .

Let us now naturally formulate a question of the validity of comparison inequalities in a more general form admitting for $b = (b_k)_{k \ge 1}$ an arbitrary predictable sequence (i.e. such one, that b_k is $\mathcal{F}_{k-1}$-measurable with $|b_k| \le 1$ for all $k \ge 1$; $\mathcal{F}_0 = \{\emptyset, \Omega\}$, $b_1 =$ const.).

In this way it becomes of interest to clarify the validity of inequalities of the form:

$$E\big|f_n(b)\big|^p \le G_p E\big|f_n\big|^p \tag{7.2}$$

$$E \max_{1 \le m \le n} \big|f_m(b)\big|^p \le G_p^* \, E \max_{1 \le m \le n} \big|f_m\big|^p \ , \tag{7.3}$$

where $f(b) = \big(f_k(b)\big)_{k \ge 1}$ is a martingale transform with a predictable sequence $b = (b_k)_{k \ge 1}$, or in a more general form, the inequalities of the type:

$$E\big|g_n\big|^p \le G_p E\big|f_n\big|^p \tag{7.4}$$

$$E \max_{1\le m\le n} \big|g_m\big|^p \le G_p^* \, E \max_{1\le m\le n} \big|f_m\big|^p \ , \tag{7.5}$$

for two martingales $f=(f_n)_{n\ge1}$ and $g=(g_n)_{n\ge1}$, such that (P-a.s) for all $n\ge1$:

$$\big|\Delta g_n\big| \le \big|\Delta f_n\big| \ . \tag{7.6}$$

In the above given inequalities one assumes that G_p and G_p^* are universal (i.e. not dependent neither on the martingales, nor on n) constants. Naturally, it is of considerable interest to address the question about the best constants. (Under the "best" constant, say, G_p in (7.4) we understand that value, that if $G'_p < G_p$, then one can find a probability space $(\Omega',\mathcal{F}',P')$, filtration $(\mathcal{F}'_n)_{n\ge0}$ and martingales g' and f' such that $\big|\Delta g'_n\big| \le \big|\Delta f'_n\big|$, $n\ge1$, but $E\big|g'_n\big|^p > G'_p E\big|f'_n\big|^p$.)

2. The validity of inequalities (7.2) for $p>1$ and (7.3) for $p\ge1$ can be established, for example, by using techniques, developed in works of Burkholder and Davis.

Below we give a proof of the inequality (7.3) for the case $p=1$. The corresponding proof for $p>1$ from the point of view of idea is rather similar and can be obtained by the very same method, like the proof of Theorem 7 in §9 of Chapter 1, [89], following the scheme of the proof given below for the case $p=1$. Let us underline that the formulation and proof of Theorem 7 in §9 of Chapter 1 in [89] have been done at once for local martingales in the case of continuous time (the case of discrete time can be imbedded into it), and in this manner it is interesting as a proof which works in a rather general situation. In the case of discrete time a very interesting result about the validity (for $p\ge1$) of the inequality (7.4) under assumption (7.6) was obtained by Burkholder [20], [23]. The formulation of this will be given below in part 4 of this section.

3. So, suppose that on the filtred probability space $\big(\Omega,\mathcal{F},(\mathcal{F}_n)_{n\ge0},P\big)$ with $\mathcal{F}_0=\{\emptyset,\Omega\}$ a martingale $f=(f_n,\mathcal{F}_n)_{n\ge1}$ with $f_0=0$ is given. Let $d_n=f_n-f_{n-1}$ $(=\Delta f_n)$, such that:

$$f_n = \sum_{k=1}^{n} d_k \quad , \quad d_0 = 0 \ . \tag{7.7}$$

Let $b=(b_n,\mathcal{F}_{n-1})_{n\ge0}$ (with $\mathcal{F}_{-1}=\mathcal{F}_0$) be a predictable sequence with $|b_n|\le1$. Set:

$$f_n(b) = \sum_{k=1}^{n} b_k d_k \quad , \quad f_n^* = \max_{1\le k\le n} |f_k| \quad , \quad f_n^*(b) = \max_{1\le k\le n} |f_k(b)| \ . \tag{7.8}$$

Our aim is to prove the following comparison inequality; for any stopping time τ :

$$Ef_\tau^*(b) \le G_1^* \, Ef_\tau^* \ , \tag{7.9}$$

or in a more detailed notation:

$$E \max_{1\le n\le\tau} \left|\sum_{k=1}^{n} b_k d_k\right| \le G_1^* \, E \max_{1\le n\le\tau} \left|\sum_{k=1}^{n} d_k\right| \ , \tag{7.10}$$

where G_1^* is a universal constant (from the estimates given below it follows $G_1^*=185$).

Introduce the following notation:

$$d_n^* = \max_{1 \le k \le n} |d_k| \quad , \quad I_n = I\Big(|d_n| > 2d_{n-1}^*\Big)$$

$$g_n = \sum_{k=1}^{n} I_k d_k \quad , \quad g_n(b) = \sum_{k=1}^{n} I_k b_k d_k$$

$$\tilde{g}_n = \sum_{k=1}^{n} E\big(I_k d_k | \mathcal{F}_{k-1}\big) \quad , \quad \tilde{g}_n(b) = \sum_{k=1}^{n} E\big(I_k b_k d_k | \mathcal{F}_{k-1}\big)$$

$$f_n' = g_n - \tilde{g}_n \quad , \quad f_n'(b) = g_n(b) - \tilde{g}_n(b)$$

$$\Big(Var\, g\Big)_n = \sum_{k=1}^{n} |I_k d_k| \quad , \quad \Big(Var\, \tilde{g}\Big)_n = \sum_{k=1}^{n} \big|E\big(I_k d_k | \mathcal{F}_{k-1}\big)\big| \ .$$

At once we note, since $|b_k| \le 1$, that:

(7.11)
$$\Big(Var\, g(b)\Big)_n = \sum_{k=1}^{n} |I_k b_k d_k| \le \Big(Var\, g\Big)_n \ ,$$

and:

$$\Big(Var\, \tilde{g}(b)\Big)_n = \sum_{k=1}^{n} \big|E\big(I_k b_k d_k | \mathcal{F}_{k-1}\big)\big| = \sum_{k=1}^{n} |b_k| \big|E\big(I_k d_k | \mathcal{F}_{k-1}\big)\big| \le \Big(Var\, \tilde{g}\Big)_n \ .$$

Lemma 1. *It takes place the following inequality:*

(7.12)
$$E \max_{1 \le n \le \tau} \big|f_n'(b)\big| \le C_1\, E \max_{1 \le n \le \tau} |f_n|$$

with the constant $C_1 = 8$.

Proof. We have:

(7.13)
$$E \max_{1 \le m \le \tau \wedge n} \big|f_m'(b)\big| \le E\Big(Var f'(b)\Big)_{\tau \wedge n} \le E\Big(Var\, g(b)\Big)_{\tau \wedge n} + E\Big(Var\, \tilde{g}(b)\Big)_{\tau \wedge n}$$

$$\le E\Big(Var\, g\Big)_{\tau \wedge n} + E\Big(Var\, \tilde{g}\Big)_{\tau \wedge n} \le E \sum_{k=1}^{\tau \wedge n} |I_k d_k| + E \sum_{k=1}^{\tau \wedge n} E\big(|I_k d_k| \big| \mathcal{F}_{k-1}\big)$$

$$\overset{\{m\}}{=} 2E \sum_{k=1}^{\tau \wedge n} |I_k d_k| \le 2E \sum_{k=1}^{\tau} |I_k d_k| \ ,$$

where the equality $\{m\}$ follows from the fact that $\Big(\sum_{k \le n} \big(|I_k d_k| - E(|I_k d_k| \big| \mathcal{F}_{k-1})\big)\Big)$ is a martingale and stopping time $\tau \wedge n \le n$. From (7.13) in an evident way by passing to the limit when $n \to \infty$ we get the inequality:

(7.14)
$$E \max_{1 \le m \le \tau} \big|f_m'(b)\big| \le 2E \sum_{k=1}^{\tau} |I_k d_k| \ .$$

Let us begin to estimate the right-hand side of that inequality. Since $|d_k| \le d_k^*$ and on $\{\omega : |d_k| > 2d_{k-1}^*\}$ we have $|d_k| + 2d_{k-1}^* \le 2|d_k| \le 2d_k^*$, then $|d_k| \le 2\Delta d_k^*$. Therefore:

$$E\sum_{k=1}^{\tau} |I_k d_k| \le 2E\sum_{k=1}^{\tau} \Delta d_k^* = 2Ed_\tau^* \ . \tag{7.15}$$

But:

$$d_n^* = \max_{1\le k\le n} |d_k| = \max_{1\le k\le n} |f_k - f_{k-1}| \le 2 \max_{1\le k\le n} |f_k| = 2f_n^* \ . \tag{7.16}$$

The desired inequality (7.12) follows evidently from (7.14)-(7.16). The lemma is proved.

To formulate the next lemma denote:

$$f_n'' = f_n - f_n' \quad , \quad f_n''(b) = f_n(b) - f_n'(b) \ .$$

Note, that:

$$f_n'' = \sum_{k=1}^{n} \Big(d_k \bar{I}_k + E\big(d_k I_k | \mathcal{F}_{k-1}\big)\Big) \ ,$$

and:

$$f_n''(b) = \sum_{k=1}^{n} b_k \Big(d_k \bar{I}_k + E\big(d_k I_k | \mathcal{F}_{k-1}\big)\Big) \ .$$

Also, put:

$$[f'']_n = \sum_{k=1}^{n} \big(\Delta f_k''\big)^2 \quad , \quad [f''(b)]_n = \sum_{k=1}^{n} \big(\Delta f_k''(b)\big)^2 \ .$$

It is evident, that:

$$\big[f''(b)\big]_n \le \big[f''\big]_n$$

since $|b_k| \le 1$.

Lemma 2. *It takes place the following inequality:*

$$E \max_{1\le n\le \tau} \big|f_n''(b)\big| \le 3\bigg(E\big[f''\big]_\tau^{1/2} + 4Ed_\tau^*\bigg) \ . \tag{7.17}$$

Proof. We shall first show that for any stopping time τ :

$$\big|\Delta f_\tau''(b)\big| \le \big|\Delta f_\tau''\big| \le 4d_{\tau-1}^* \ . \tag{7.18}$$

Indeed, we have:

$$\Delta f_n''(b) = b_n\Big(\bar{I}_n d_n + E\big(I_n d_n | \mathcal{F}_{n-1}\big)\Big) \ ,$$

and so:

$$\big|\Delta f_n''(b)\big| \le \big|\Delta f_n''\big| = \big|\bar{I}_n d_n + E\big(I_n d_n | \mathcal{F}_{n-1}\big)\big| \ . \tag{7.19}$$

Since $E(d_n|\mathcal{F}_{n-1})=0$, then $E(I_n d_n|\mathcal{F}_{n-1})+E(\bar{I}_n d_n|\mathcal{F}_{n-1})=0$ and it follows:

$$\tilde{g}_n = \sum_{k=1}^{n} E(I_k d_k|\mathcal{F}_{k-1}) = -\sum_{k=1}^{n} E(\bar{I}_k d_k|\mathcal{F}_{k-1}) \ .$$

Hence from (7.19):

(7.20)
$$\big|\Delta f''_n(b)\big| \le \big|\bar{I}_n d_n\big| + \big|\Delta\tilde{g}_n\big| \le \big|d_n\big| I\big(|d_n|\le 2d^*_{n-1}\big) + \Big|E\big(d_n I\big(|d_n|\le 2d^*_{n-1}\big)\big|\mathcal{F}_{n-1}\big)\Big| \le 2d^*_{n-1}+2d^*_{n-1}=4d^*_{n-1} \ ,$$

which proves (7.18) in the case $\tau\equiv n$.

Let now τ be a predictable stopping time and let the martingale $f=(f_n,\mathcal{F}_n)$ be uniformly integrable. Then again $E(d_\tau|\mathcal{F}_{\tau-1})=0$ and the chain of inequalities (7.20) remains valid with changing n to τ .

But if τ is an arbitrary stopping time, then from (7.20) it follows, that it is sufficient only to check, that $\big|\Delta\tilde{g}_\tau\big| \le 2d^*_{\tau-1}$.

With this in goal note, that $\tilde{g}=(\tilde{g}_n)$ is a predictable sequence, and consequently, the times of its jumps are predictable. So either $\big|\Delta\tilde{g}_\tau\big|=0$ or, as already proved, $\big|\Delta\tilde{g}_\tau\big|\le 2d^*_{\tau-1}$, if f is a uniformly integrable martingale. The case of arbitrary martingale reduces to the previous one by help of the usual procedure of localization, consisting of this, that one can find such a sequence of stopping times $(\sigma_k)_{k\ge1}$, that $\sigma_k\uparrow\infty$ and the "stopped" martingales $f^{(k)}=(f_{n\wedge\sigma_k},\mathcal{F}_n)$ are already uniformly integrable (see for example [89], p.73).

The established property (7.18) permits us to claim, that the martingale $f''(b)=(f''_n(b),\mathcal{F}_n)$ is locally square-integrable, i.e. there exists such a sequence of stopping times $(\gamma_k)_{k\ge1}$, $\gamma_k\uparrow\infty$, that the "stopped" sequence $\big(f''_{n\wedge\gamma_k}(b)\big)$ forms a square-integrable martingale, i.e. such a martingale, that:

$$\sup_{n\ge1} E\Big|f''_{n\wedge\gamma_k}(b)\Big|^2 < \infty \ .$$

(For example, for localizing stopping times it is sufficient to take the following ones:

$$\gamma_k = \inf\ \big\{\ n\ge1 : |f''_n(b)|\vee 4d^*_{n-1}\ge k\ \big\}\ .)$$

The fact, that $f''(b)$ and f'' are local square-integrable martingales permits us to state that for any stopping time τ :

(7.21)
$$E\big|f''_\tau(b)\big|^2 \le E\big[f''(b)\big]_\tau \le E\big[f''\big]_\tau \le E\max_{1\le m\le\tau}\big(f''_m\big)^2 \ .$$

If we denote:

(7.22)
$$X_n = \big|f''_n(b)\big|^2 \quad , \quad Y_n = \big[f''\big]_n \ ,$$

then from (7.21) we have, in particular, that:

(7.23)
$$EX_\tau \le EY_\tau$$

for any stopping time τ .

This property is widely called the *property of* L*-domination* (of process X by process Y), or the *property of domination of Lenglart*; see [69], p.35 or [89], p.68.

According to Theorem 4 on p.68 in [89] the property of L-domination permits us to assert, that if $\Delta Y_n \le D_{n-1}$, $n \ge 1$, where $D = (D_n)_{n\ge1}$ is an increasing adapted process, then for any $\alpha > 0$, $\beta > 0$ and stopping time τ , it takes place the following inequality:

$$P\Big(\max_{1\le n\le\tau} X_n \ge \alpha\Big) \le \frac{1}{\alpha}E\Big(\Big(Y_\tau + D_\tau\Big) \wedge \beta\Big) + P\Big(Y_\tau + D_\tau \ge \beta\Big) \ . \tag{7.24}$$

Let us apply this to the case (7.22), then from (7.24) we find, that:

$$\begin{aligned} P\Big(\max_{1\le n\le\tau} \big|f''_n(b)\big| \ge \alpha\Big) &= P\Big(\max_{1\le n\le\tau} \big|f''_n(b)\big|^2 \ge \alpha^2\Big) \\ &\le \frac{1}{\alpha^2}E\Big(\Big(\big[f''\big]_\tau + D_\tau\Big) \wedge \alpha^2\Big) + P\Big(\big[f''\big]_\tau + D_\tau \ge \alpha^2\Big) \ , \end{aligned} \tag{7.25}$$

where one may take $D_k = \big(4d^*_k\big)^2$ since by the force of (7.18):

$$\Delta\big[f''\big]_k = \big(\Delta f''_k\big)^2 \le D_{k-1} \ .$$

Denote:

$$\xi = \big[f''\big]_\tau + D_\tau \ .$$

Then (7.25) takes the following form:

$$P\Big(\max_{1\le n\le\tau} \big|f''_n(b)\big| \ge \alpha\Big) \le \frac{1}{\alpha^2}E\big(\xi \wedge \alpha^2\big) + P\big(\xi \ge \alpha^2\big) \ ,$$

and so:

$$\begin{aligned} E\max_{1\le n\le\tau} \big|f''_n(b)\big| &\le \int_0^\infty \frac{1}{\alpha^2}E\big(\xi \wedge \alpha^2\big)\,d\alpha + \int_0^\infty P\big(\sqrt{\xi} \ge \alpha\big)\,d\alpha \\ &= E\int_0^{\sqrt{\xi}} \frac{1}{\alpha^2}\cdot\alpha^2\,d\alpha + E\xi\int_{\sqrt{\xi}}^\infty \frac{1}{\alpha^2}\,d\alpha + \int_0^\infty P\big(\sqrt{\xi} \ge \alpha\big)\,d\alpha \\ &= E\sqrt{\xi} + E\sqrt{\xi} + E\sqrt{\xi} = 3E\sqrt{\xi} \ . \end{aligned} \tag{7.26}$$

Note, that:

$$\sqrt{\xi} \equiv \sqrt{\big[f''\big]_\tau + D_\tau} \le \sqrt{\big[f''\big]_\tau} + \sqrt{D_\tau} \le \sqrt{\big[f''\big]_\tau} + 4d^*_\tau \ ,$$

which together with (7.26) proves the desired inequality (7.17). The lemma is proved.

Lemma 3. *It takes place the following inequality:*

$$E\big[f''\big]_\tau^{1/2} \le 51\,E\max_{1\le n\le\tau} |f_n| \ . \tag{7.27}$$

Proof. From (7.21) we have:

$$E\big[f''\big]_\tau \le E \max_{1\le n\le\tau} \big(f''_n\big)^2 \ ,$$

i.e. it takes place the property of L-domination (7.23) with:

$$X_n = \big[f''\big]_n \quad , \quad Y_n = \max_{1\le m\le n} \big(f''_m\big)^2 \ .$$

Then from (7.24) we get:

$$P\Big(\big[f''\big]_\tau \ge \alpha^2\Big) = P\Big(\max_{1\le m\le\tau} \big[f''\big]_m \ge \alpha^2\Big) \tag{7.28}$$

$$\le \frac{1}{\alpha^2} E\Bigg(\bigg(\max_{1\le m\le\tau} \big(f''_m\big)^2 + D_\tau\bigg) \wedge \alpha^2\Bigg) + P\bigg(\max_{1\le m\le\tau} \big(f''_m\big)^2 + D_\tau \ge \alpha^2\bigg) \ ,$$

where one may take $D_n = \big(4d^*_n\big)^2$, since then $\Delta\big(\max_{1\le m\le n} (f''_m)^2\big) \le \Delta f''_n \le \big(4d^*_{n-1}\big)^2 = D_{n-1}$, which implies the validity of the inequality (7.24). Let:

$$\eta = \max_{1\le m\le\tau} |f''_m| + 4d^*_{\tau-1} \ .$$

Then from (7.28):

$$P\Big(\big[f''\big]^{1/2}_\tau \ge \alpha\Big) \le \frac{1}{\alpha^2} E\big(\alpha^2 \wedge \eta^2\big) + P\{\, \eta \ge \alpha \,\} \ ,$$

and so:

$$E\big[f''\big]^{1/2}_\tau \le \int_0^\infty \frac{1}{\alpha^2} E\big(\alpha^2 \wedge \eta^2\big)\, d\alpha + E\eta$$

$$= E\bigg(\eta^2 \int_\eta^\infty \frac{1}{\alpha^2}\, d\alpha\bigg) + E\int_0^\eta \frac{1}{\alpha^2} \cdot \alpha^2\, d\alpha + E\eta = E\eta + E\eta + E\eta = 3E\eta \ .$$

Hence together with (7.12) we find, that:

$$E\big[f''\big]^{1/2}_\tau \le 3\, E \max_{1\le m\le\tau} |f''_m| + 12 E d^*_{\tau-1}$$

$$\le 3\, E \max_{1\le m\le\tau} |f_m| + 3\, E \max_{1\le m\le\tau} |f'_m| + 24\, E \max_{1\le m\le\tau} |f_m|$$

$$\le 3\, E \max_{1\le m\le\tau} |f_m| + 24\, E \max_{1\le m\le\tau} |f_m| + 24\, E \max_{1\le m\le\tau} |f_m| = 51\, E \max_{1\le m\le\tau} |f_m| \ .$$

The lemma is proved.

From Lemma 2, Lemma 3, and (7.16) we find, that:

$$E \max_{1\le m\le\tau} \big|f''_m(b)\big| \le 3\Big(E\big[f''\big]^{1/2}_\tau + 4Ed^*_\tau\Big) \tag{7.29}$$

$$\leq 3\Big(51\, E \max_{1\leq m\leq \tau} |f_m| + 8\, E \max_{1\leq m\leq \tau} |f_m|\Big) = 177\, E \max_{1\leq m\leq \tau} |f_m| \ .$$

Together with estimate (7.12) from (7.29) we find:

$$E \max_{1\leq m\leq \tau} \big|f_m(b)\big| \leq 185\, E \max_{1\leq m\leq \tau} |f_m| \ ,$$

which proves the inequality (7.9) with constant $G_1^* = 185$. (The question about the value for the best constant G_1^* still remains open.)

4. Let us formulate the above mentioned result of Burkholder (see for example [18], [19], [20]) for the case of discrete time.

Let $p > 1$, and let $f = (f_n)$ and $g = (g_n)$ be two martingales such that P-a.s. for all $n \geq 1$:

$$\big|\Delta g_n\big| \leq \big|\Delta f_n\big| \ ,$$

(see (7.4) and (7.6)). Then for all $n \geq 1$:

$$E\big|g_n\big|^p \leq G_p E\big|f_n\big|^p \ ,$$

where:

$$G_p = \big(p^* - 1\big)^p \quad , \quad p^* = \max\,(p, q) \quad , \quad 1/p + 1/q = 1 \ .$$

Moreover, the constant $G_p = \big(p^* - 1\big)^p$ is the best possible.

8. On the best constants in the Khintchine's inequalities

1. In accordance with (2.1) and (2.5) for any $n \geq 1$ and any sequence of numbers $a = (a_1, a_2, \ \dots \ )$:

(8.1) $$E\Big|\sum_{k=1}^{n} a_k \varepsilon_k\Big|^{2m} \leq E\xi^{2m}\Big(\sum_{k=1}^{n} |a_k|^2\Big)^m \ ,$$

where $\xi \sim N(0,1)$ and, consequently:

$$E\xi^{2m} = \frac{(2m)!}{2^m m!} \ .$$

Put in (8.1):

(8.2) $$a_1 = \dots = a_n = \frac{1}{\sqrt{n}} \ .$$

Then we find, that:

(8.3) $$E\Big|\frac{1}{\sqrt{n}}\sum_{k=1}^{n} \varepsilon_k\Big|^{2m} \leq E\xi^{2m} \ .$$

By the central limit theorem:

(8.4) $$\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\varepsilon_k \xrightarrow{d} \xi \ ,$$

($\xrightarrow{d}$ denotes convergence in distribution). From the inequalities of Khintchine (2.4) it follows, that for any $p > 1$:

$$\sup_{n\geq 1} E\left|\left(\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\varepsilon_k\right)^{2m}\right|^p < \infty \ ,$$

and consequently the family:

$$\left(\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\varepsilon_k\right)^{2m}_{n\geq 1}$$

is uniformly integrable. Hence from (8.4) it follows, that:

$$\lim_{n\to\infty} E\left|\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\varepsilon_k\right|^{2m} = E\xi^{2m} \ .$$

From this it follows that in the inequality (8.1), being considered for *all* $n \geq 1$ and *all* sequences $a = (a_1, a_2, \ \ldots \ )$, the constant $B_{2m} = E\xi^{2m} \ \big(= (2m-1)!! \ \big)$ is the best possible (in the sense explained in the end of part 1 of Section 7).

Rewrite the inequality (8.1) to the form:

(8.5) $$E\left|\sum_{k=1}^{n} a_{k,n}\,\varepsilon_k\right|^{2m} \leq E\xi^{2m} \ ,$$

where

$$a_{k,n} = a_k \Big/ \sqrt{\textstyle\sum_{k=1}^{n} |a_k|^2} \ ,$$

and, consequently, $\sum_{k=1}^{n} a_{k,n}^2 = 1$.

It is clear, that by symmetry of distributions of random variables ε_k it is sufficient to consider only the case of non-negative values a_k , excluding in this case, when $a_1 = \ldots = a_n = 0$.

The consideration given above shows, that the choice of values $a_k = 1/\sqrt{n}$, $1 \leq k \leq n$, has some extreme properties. Indeed, as it will be shown below, for *any* $(a_1, \ldots, a_n)$:

$$E\left|\sum_{k=1}^{n} a_{k,n}\,\varepsilon_k\right|^{2m} \leq E\left|\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\varepsilon_k\right|^{2m} \ ,$$

and, moreover, the numbers:

$$E\left|\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\varepsilon_k\right|^{2m}$$

are monotone-increasing to $E\xi^{2m}$.

The explanation of this phenomenon and its analogues in other interesting for us cases might be obtained, if we turn to the notion of "convexity in the sense of Schur", introduced by Schur in the paper [131], published as it turns out, in the same year, when the paper of Khintchine [74] with “the inequalities of Khintchine” has been appeared.

2. Let us introduce concepts and facts related to the notion of “Schur-convexity” and “the theory of majorization”. The motivation for introducing the corresponding notions relates with a natural desire to have a “right” definition of the property, that “the components of a vector x are less spread out than are the components of a vector y ”.

A precise definition of this vague intention one may obtain by introducing the concept (Hardy, Littlewood, Pólya 1929) which states, that “ x is majorized by y ” (in the notation: $x \prec y$).

Let $z = (z_1, \dots, z_n)$ be a vector in $\mathbf{R}^n$ and let $z_1^*, \dots, z_n^*$ be the components of vector z in decreasing (more precisely, non-increasing) order: $z_1^* \geq z_2^* \geq \dots \geq z_n^*$. For the two vectors x , $y \in \mathbf{R}^n$ we say, that “ x is majorized by y ” and we write $x \prec y$, if:

$$\sum_{i=1}^{k} x_i^* \leq \sum_{i=1}^{k} y_i^* ,$$

for all $k = 1, 2, \dots, n-1$ and:

$$\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i .$$

For example:

$$\Big(\frac{1}{n}, \dots, \frac{1}{n}\Big) \prec \Big(\frac{1}{n-1}, \dots, \frac{1}{n-1}, 0\Big) \prec \Big(\tfrac{1}{2}, \tfrac{1}{2}, 0, \dots, 0\Big) \prec (1, 0, \dots, 0) ; \tag{8.6}$$

$$\Big(\frac{1}{n}, \dots, \frac{1}{n}\Big) \prec (a_1, \dots, a_n) \prec (1, 0, \dots, 0) \quad \text{for} \quad a_i \geq 0 \quad \text{with} \quad \textstyle\sum_{i=1}^n a_i = 1 ; \tag{8.7}$$

$$\Big(n^{-1} \textstyle\sum_{i=1}^n x_i \, , \ \dots \ , \ n^{-1} \sum_{i=1}^n x_i\Big) \prec (x_1, \dots, x_n) \quad \text{for} \quad x_i \geq 0 . \tag{8.8}$$

Let $D \subset \mathbf{R}^n$ be a set, and let $\Phi = \Phi(x)$ be a function, defined on D with values in $\mathbf{R}$. We will say that the function $\Phi = \Phi(x)$ posses the property of “convexity in the sense of Schur” (“Schur-convex”) on D , if:

$$\Phi(x) \leq \Phi(y)$$

for x , $y \in D$ such that $x \prec y$. A function $\Phi = \Phi(x)$, $x \in D$, is said to be “Schur-concave”, if the function $-\Phi = -\Phi(x)$ is “Schur-convex”.

Note, if the function $\Phi = \Phi(x)$ is symmetric and convex then it is “Schur-convex” (this, in essence, was contained in the paper of Schur [131], 1923).

For our aims the following result (together with (8.6)-(8.8)) leads to many interesting corollaries (Schur [131], 1923; Hardy, Littlewood, Pólya [60], 1929):

Let $I \subset \mathbf{R}$ be an interval, and suppose that the function $g = g(x)$ defined on I with values in $\mathbf{R}$ is convex. Then the function:

$$\Phi(x) = \sum_{i=1}^{n} g(x_i)$$

is “Schur-convex” on I^n .

The following fundamental theorem (Schur [131], 1923; Ostrowski [113], 1952) permits us to verify the property of "convexity in the sense of Schur" in terms of a property of the first partial derivatives.

Let $I \subset \mathbf{R}$ be an interval, and let a real valued function $\Phi = \Phi(x)$, $x \in I^n$ be continuously differentiable. Then $\Phi = \Phi(x)$ is "Schur-convex" on I^n , if and only if the following two conditions are satisfied:

(i) Φ is symmetric on I^n ;

(ii) $(x_i - x_j)\Big(\frac{\partial\Phi}{\partial x_i}(x) - \frac{\partial\Phi}{\partial x_j}(x)\Big) \geq 0$ for all $x \in I^n$ and all $i \neq j$.

Under the validity of condition (i), the condition (ii) is equivalent to the condition:

(ii)' $(x_1 - x_2)\Big(\frac{\partial\Phi}{\partial x_1}(x) - \frac{\partial\Phi}{\partial x_2}(x)\Big) \geq 0$ for all $x \in I^n$.

An analogous characterization takes place for "concavity in the sense of Schur" as well, with changing the inequalities in (ii) and (ii)' to the reversed ones. Note, that condition (ii) (or (ii)') is widely called "the Schur condition".

Remark 1. The paper of Schur [131], 1923, contains a first sufficiently comprehensive study of functions $\Phi = \Phi(x)$, for which " $x \prec y \Rightarrow \Phi(x) \leq \Phi(y)$ ". In this paper one finds "the Schur condition" (on $\mathbf{R}^n_+$, which later was extended by Ostrowski [113], 1952, to $\mathbf{R}^n$). Functions satisfying that property Schur called "convex" as opposite to the functions which were "convex in the sense of Jensen". Now the last ones are called convex, and convex functions by Schur, according to Ostrowski [113], 1952, are called "Schur-convex".

Remark 2. An important analytic property in the theory of majorization was established by Hardy, Littlewood and Pólya in 1929. This property states, that $x \prec y$ if and only if $x = Py$ for some double stochastic matrix $P = [p_{ij}]$ ($p_{ij} \geq 0$, $\sum_i p_{ij} = \sum_j p_{ij} = 1$, $\forall i, j$). In fact, this property was used by Schur as definition.

Let us give two interesting examples of functions having the property of "convexity in the sense of Schur" and the property of "concavity in the sense of Schur".

Example 1. Let $p = (p_1, \ldots, p_n)$, $p_i \geq 0$, $\sum_{i=1}^n p_i = 1$. The entropy of the vector $p = (p_1, \ldots, p_n)$ is by definition the quantity:

$$H(p) = -\sum_{i=1}^{n} p_i \log p_i \ .$$

This function is (strictly) "Schur-concave". In particular:

$$H(1, 0, \ldots, 0) \leq H(p) \leq H(\tfrac{1}{n}, \ldots, \tfrac{1}{n}) \ .$$

Example 2. Put:

$$\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i \ .$$

Consider the standard deviation of the vector $x_1,\dots,x_n$:

$$\Big(\sum_{i=1}^{n}\big(x_i-\bar{x}_n\big)^2\Big)^{1/2} .$$

Then this function is “Schur-convex”.

3. Now we are in position to be ready to explain why the choice of values for $a_1,\dots,a_n$, made in (9.2) ($a_1=\ldots=a_n=1/\sqrt{n}$) posses some extreme properties.

Consider the function:

$$\varphi_{2m}(x_1,\dots,x_n)=E\Big|\sum_{k=1}^{n}\sqrt{x_k}\,\varepsilon_k\Big|^{2m} \ , \ \ m\ge 1 . \tag{8.9}$$

It turns out, that this function considered on $\mathbf{R}_+^n$ is “Schur-concave”, which is directly verified by using the “Schur-condition” (ii)’. Hence by using (8.7) we find, that:

$$\varphi_{2m}(\tfrac{1}{n},\dots,\tfrac{1}{n})\ge\varphi_{2m}(x_1,\dots,x_n) \ \ , \ \ x_i\ge 0 \ , \ \textstyle\sum_{i=1}^{n}x_i=1 . \tag{8.10}$$

This precisely means, that the choice made in (8.2) posses the extreme property, such that with this choice the left-hand side in (8.5) takes its *maximal* value.

Returning back to (8.6) we find, that the “Schur-concavity” of the function $\varphi_{2m}(x_1,\dots,x_n)$ leads to an interesting property, that:

$$E\Big|\,\frac{1}{\sqrt{n-1}}\sum_{k=1}^{n-1}\varepsilon_k\,\Big|^{2m}\le E\Big|\,\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\varepsilon_k\,\Big|^{2m}$$

Thus, consequently, additionally to the statement of the central limit theorem (8.4), we find, that the moments:

$$E\Big|\,\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\varepsilon_k\,\Big|^{2m}$$

are *monotone-increasing* to $E\xi^{2m}$.

4. Now denote:

$$\varphi_p(x_1,\dots,x_n)=E\Big|\sum_{k=1}^{n}\sqrt{x_k}\,\varepsilon_k\Big|^{p} . \tag{8.11}$$

As a matter of fact, the function $\varphi_p(x_1,\dots,x_n)$ is “Schur-concave” on $\mathbf{R}_+^n$ not only for $p=2m$, $m\ge 1$, but also for all $p\ge 3$. This was proved by a straightforward consideration (without referring to or mentioning “convexity in the sense of Schur”) in the work of P. Wittle [153], 1960. (In this work it was proved the validity of this property for $2<p<3$ as well, which is like written in [153] wrong, since one can give an example of failure of the inequality $E\big|\sum_{i=1}^{n}\sqrt{x_i}\varepsilon_i\,\big|^p\le E\big|(1/\sqrt{n})\sum_{i=1}^{n}\varepsilon_i\,\big|^p$ for those values of p and $\sum_{i=1}^{n}x_i=1$, $x_i\ge 0$.) The argumentation, based upon the concept of “convexity in the sense of Schur”, is contained in the articles of M. Eaton [39], 1970, and Komorowski [79], 1988.

In this way, the turn to "concavity in the sense of Schur" of the function $\varphi_p(x_1,\dots,x_n)$ permits us to make some special conclusions concerning the values for the best constants $\mathbf{B}_p(n)$ in the inequality:

$$\Big\|\sum_{k=1}^n a_k\varepsilon_k\Big\|_p \le \mathbf{B}_p(n)\Big(\sum_{k=1}^n |a_k|^2\Big)^{1/2}, \tag{8.12}$$

for every $n \ge 1$ and $p \in \{2\}\cup[3,\infty[$:

$$\mathbf{B}_p(n) = \max\Big\{\Big\|\sum_{k=1}^n a_k\varepsilon_k\Big\|_p : \sum_{k=1}^n |a_k|^2 = 1\Big\} = \Big(E\Big|\frac{1}{\sqrt{n}}\sum_{k=1}^n \varepsilon_k\Big|^p\Big)^{1/p}, \tag{8.13}$$

$$E\Big|\frac{1}{\sqrt{n}}\sum_{k=1}^n \varepsilon_k\Big|^p = 2^{-n}n^{-p/2}\sum_{k=0}^n C_n^k\,|n-2k|^p\ \Big(\le \frac{2^{p/2}}{\sqrt{\pi}}\,\Gamma\Big(\frac{p+1}{2}\Big)\Big)\ . \tag{8.14}$$

Note that the case $p\in]0,2[$ is trivial (with $\mathbf{B}_p(n)=1$), since $\|\sum_{k=1}^n a_k\varepsilon_k\|_p \le \|\sum_{k=1}^n a_k\varepsilon_k\|_2 = (\sum_{k=1}^n |a_k|^2)^{1/2}$, and taking $a_1=1$, $a_2=\ldots=a_n=0$, we have $\|\sum_{k=1}^n a_k\varepsilon_k\|_p = 1$.

Concerning the case $p\in]2,3[$, as far as we know about, the values for the best constants $\mathbf{B}_p(n)$ are not known. Consequently, for any $n\ge 1$:

$$\mathbf{B}_p(n) = \begin{cases} 1 & , \quad 0<p\le 2 \\ ? & , \quad 2<p<3 \\ 2^{-n/p}\ n^{-1/2}\big(\sum_{k=0}^n C_n^k\,|n-2k|^p\big)^{1/p} & , \quad 3\le p<\infty\ . \end{cases} \tag{8.15}$$

Let us turn to the estimate from below:

$$\mathbf{A}_p(n)\ \Big(\sum_{k=1}^n |a_k|^2\Big)^{1/2} \le \Big\|\sum_{k=1}^n a_k\varepsilon_k\Big\|_p\ .$$

The case $p\ge 2$ is simple: $\mathbf{A}_p(n)=1$, since $(\sum_{k=1}^n |a_k|^2)^{1/2} = \|\sum_{k=1}^n a_k\varepsilon_k\|_2 \le \|\sum_{k=1}^n a_k\varepsilon_k\|_p$, and for $a_1=1$, $a_2=\ldots=a_n=0$, the inequality turns into an equality. In the case $0<p<2$ the value for $\mathbf{A}_p(n)$, as far as we know about, is not known. Consequently, for any $n\ge 1$:

$$\mathbf{A}_p(n) = \begin{cases} ? & , \quad 0<p<2 \\ 1 & , \quad 2\le p<\infty\ . \end{cases} \tag{8.16}$$

5. Let us return to the problem about the best (universal) constants $\mathbf{A}_p$ and $\mathbf{B}_p$ in the inequalities of Khintchine:

$$\mathbf{A}_p\Big(\sum_{k=1}^n |a_k|^2\Big)^{1/2} \le \Big\|\sum_{k=1}^n a_k\varepsilon_k\Big\|_p \le \mathbf{B}_p\Big(\sum_{k=1}^n |a_k|^2\Big)^{1/2}\ , \tag{8.17}$$

being considered not just for one concrete and fixed n , but for all $n\ge 1$.

About the constants $\mathbf{B}_p$ the following statement holds true:

$$(8.18) \qquad \mathbf{B}_p = \begin{cases} 1 & , \ 0 < p \le 2 \\ \sqrt{2}\left(\frac{\Gamma\left(\frac{p+1}{2}\right)}{\sqrt{\pi}}\right)^{1/p} & , \ 2 \le p < \infty \ , \end{cases}$$

where $\sqrt{2}\,(\Gamma((p+1)/2)/\sqrt{\pi})^{1/p} = \|\xi\|_p$, $\xi \sim N(0,1)$. As already noted above, in the case $p = 2m$, $m \ge 1$:

$$\mathbf{B}_{2m} = \|\xi\|_{2m} = \Big((2m-1)!!\Big)^{1/2m} .$$

The question about the values for the (best) constants $\mathbf{B}_p$ (exposed in (8.18) has a long history. The case of optimality of $\mathbf{B}_p$ for $p = 2m$ has been established by S. B. Stechkin [139], 1961, (by a direct calculation, without invoking the central limit theorem as presented above). From the article of Wittle [153] it can be concluded that for $p \ge 3$ the best constant $\mathbf{B}_p$ in (8.17) is less or equal to $\sqrt{2}\,(\Gamma((p+1)/2)/\sqrt{\pi})^{1/p}$. The optimality of the constants $\mathbf{B}_p$ for $p \ge 3$ follows from the considerations mentioned above about "concavity in the sense of Schur" of the function $\varphi_p(x_1, \dots, x_n)$. It follows from this (as well as for $p = 2m$) that:

$$\mathbf{B}_p = \lim_{n\to\infty} \mathbf{B}_p(n) = \lim_{n\to\infty} \left(E\left| \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \varepsilon_k \right|^p \right)^{1/p}$$
$$= \Big(E|\xi|^p\Big)^{1/p} = \sqrt{2}\left(\frac{\Gamma\left(\frac{p+1}{2}\right)}{\sqrt{\pi}}\right)^{1/p} .$$

Notice, that the case $p \ge 3$ was considered (by another method) in the paper of Young [156], 1976.

Finally, the case $p \in \,]2,3[$ does not fit under the analysis based upon "Schur-concavity" and was investigated in the paper of Haagerup [57], 1982, in which the value of the constant $\mathbf{B}_p$ has been obtained by direct (rather complicated) calculations.

It is interesting to observe, that with $p \to \infty$:

$$\mathbf{B}_p \sim \sqrt{\frac{p}{e}} .$$

About the constants $\mathbf{A}_p$ the following is known:

$$(8.19) \qquad \mathbf{A}_p = \begin{cases} 2^{1/2-1/p} & , \ 0 < p \le p_0 \\ \sqrt{2}\left(\frac{\Gamma\left(\frac{p+1}{2}\right)}{\sqrt{\pi}}\right)^{1/p} & , \ p_0 \le p \le 2 \ , \\ 1 & , \ 2 \le p < \infty \ , \end{cases}$$

where p_0 is the root of the equation:

$$\Gamma\Big(\frac{p+1}{2}\Big) = \frac{\sqrt{\pi}}{2} \ , \ 1 < p < 2 \ ,$$

($p_0 = 1.84742 \ \dots$).

Note, that the case $p \ge 2$ is trivial again. The case $p = 1$ was investigated by Szarek in the paper [141], 1978. (In Hall [58], 1975, the history of this problem is described; especially that

$\mathbf{A}_1$ equals to $2^{-1/2}$ was conjectured by Littlewood.) The case $p \in]0,1[\cup]1,2[$ has been investigated by Haagerup [57] in 1982. His method (like in the case of constant $\mathbf{B}_p$ for $p \in]2,3[$) is of a computational character and it is not seen, can one explain his results by basing them upon some general concept of type of the "convexity in the sense of Schur".

9. The inequalities of Khintchine in the (exponential) Orlicz spaces

1. In the first part of this section we demonstrate how the knowledge about the best constants in the (classical) "inequalities of Khintchine" enables one to derive the analogue of such inequalities (with best constants) in the case of some spaces of Orlicz. Let us recall a few of needed facts.

A function $\psi = \psi(x)$, defined on $\mathbf{R}_+$ and with values in $\mathbf{R}_+$, is called a Young function, if it is convex, increasing, $\psi(0) = 0$ and $\psi(\infty) = \infty$. For such a Young function ψ define the Orlicz space $L^\psi(P)$ as a vector space of all random variables $X = X(\omega)$, defined on $(\Omega, \mathcal{F}, P)$ and satisfying for some $C > 0$ the property:

$$E\psi\left(\frac{|X|}{C}\right) < \infty \ . \tag{9.1}$$

In the space $L^\psi(P)$ one introduces the *Orlicz norm*:

$$\|X\|_\psi = \inf \left\{ C > 0 : E\psi\big(|X|/C\big) \le 1 \right\} , \tag{9.2}$$

with respect to which the space is becoming a Banach space.

If $\psi(x) = x^p$, $1 \le p < \infty$, then $L^\psi(P)$ coincides with the space $L^p(P)$. Another interesting example is provided by the exponential function:

$$\psi_p(x) = \exp\left(x^p\right) - 1 \quad , \quad 1 \le p < \infty \ .$$

The consideration given below, concerning the case $p = 2$, will relate the question on, how to compute the norm:

$$\left\| \frac{1}{\left(\sum\limits_{k=1}^n |a_k|^2\right)^{1/2}} \sum_{k=1}^n a_k \varepsilon_k \right\|_{\psi_2} . \tag{9.3}$$

Denote $S_n = \sum_{k=1}^n a_k \varepsilon_k$, $A_n = \sum_{k=1}^n |a_k|^2$. In accordance with (9.2), for finding the norm we are interested in, it is necessary to consider the quantity:

$$E\psi_2\left(\frac{|S_n|}{C\sqrt{A_n}}\right) = E\exp\left(\frac{|S_n|}{C\sqrt{A_n}}\right) - 1 \ ,$$

and to find such a minimal C , for which this quantity is less or equal to 1 .

By force of Taylor expansion and knowledge about the values of the best constants B_p for $p = 2m$, $m \ge 1$ in the inequalities of Khintchine, we find, that:

$$E\exp\left(\frac{|S_n|}{C\sqrt{A_n}}\right) = \sum_{m=0}^{\infty}\frac{E|S_n|^{2m}}{C^{2m}(A_n)^m m!} \le \sum_{m=0}^{\infty}\frac{B_{2m}}{C^{2m}m!} = \sum_{m=0}^{\infty}\frac{(2m-1)!!}{C^{2m}m!}$$
$$= \sum_{m=0}^{\infty}\frac{(2m-1)!!}{2^m m!}\left(\frac{2}{C^2}\right)^m = \left(1-\frac{2}{C^2}\right)^{-1/2}$$

for $\left|2/C^2\right| < 1$. Solving the equation:

$$\left(1-\frac{2}{C^2}\right)^{-1/2} = 2 \ ,$$

we find, that:

$$C = \sqrt{\frac{8}{3}} \ ,$$

and consequently we obtain the analogue (right-hand) of the inequality of Khintchine for the exponential Orlicz space $L^{\psi_2}(P)$:

$$\text{(9.4)} \qquad \Big\| \sum_{k=1}^{n} a_k \varepsilon_k \Big\|_{\psi_2} \le \sqrt{\frac{8}{3}} \left(\sum_{k=1}^{n} |a_k|^2\right)^{1/2} .$$

Being based upon the property of uniform integrability and the central limit theorem we can verify, that for $n \to \infty$:

$$\text{(9.5)} \qquad \Big\| \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \varepsilon_k \Big\|_{\psi_2} \longrightarrow \|\xi\|_{\psi_2} \ ,$$

where $\xi \sim N(0,1)$. It is easily computed, that:

$$\|\xi\|_{\psi_2} = \sqrt{8/3} \ .$$

Together with (9.4) and (9.5) it shows, that the constant $\sqrt{8/3}$ is the best possible (Peškir, [116], 1992).

2. We continue by presenting Khintchine's inequalities in the (exponential) Orlicz spaces of which the inequality (9.4) is a particular (but best known) instance. For this, two cases are to be distinguished: $1 \le p \le 2$ and $2 < p < \infty$. (The case $0 < p < 1$ is analogous to the case $1 \le p \le 2$, but after resolving a convexity problem about ψ_p on the interval containing zero (letting ψ_p to be linear on this interval, and unchanged on the rest of $\mathbf{R}_+$). We shall omit this.)

The inequality is valid:

$$\text{(9.6)} \qquad \mathbf{A}_p\left(\sum_{k=1}^{n}|a_k|^2\right)^{1/2} \le \Big\|\sum_{k=1}^{n} a_k\varepsilon_k\Big\|_{\psi_p} \le \mathbf{B}_p\left(\sum_{k=1}^{n}|a_k|^2\right)^{1/2}$$

for all $1 \le p \le 2$. (For proof see [87], p.93.) The values for the best constants $\mathbf{A}_p$ and $\mathbf{B}_p$ in (9.6) are not known (except that $\mathbf{B}_2 = \sqrt{8/3}$ as recorded in the end of part 1 above).

To obtain an extension of (9.6) to $p > 2$, we shall introduce the *Laurent norm*:

$$ (9.7) \qquad \|X\|_{\psi,\infty} = \inf \left\{ C>0 \ : \ \sup_{t>0} \Big(\psi(t) \, P\{ \, |X| > Ct \, \} \Big) \leq 1 \right\} , $$

where ψ is a Young function (given and fixed). The Laurent space $L^{\psi,\infty}(P)$ is defined in a natural way (as the Orlicz space $L^{\psi}(P)$ in part 1 above). It is an Frechet space with respect to $\| \cdot \|_{\psi,\infty}^{1/2}$ (see [116]). In particular, if $\psi(x) = x^p$ for $1 \leq p < \infty$, then we have:

$$ (9.8) \qquad \|X\|_{p,\infty} = \left(\sup_{t>0} t^p \, P\{ |X| > t \} \right)^{1/p} , $$

and the "norm" $\| \cdot \|_{p,\infty}$ is called the *weak L^p-norm.* We moreover have:

$$ (9.9) \qquad \|X\|_{p,\infty} \leq \|X\|_p \leq \left(\frac{r}{r-p} \right)^{1/p} \|X\|_{r,\infty} $$

for all $1 \leq p < r < \infty$. (In fact, (9.8) and (9.9) extend to all $r > p > 0$.)

In exactly the same way the previous definitions carry over from the underlying probability space $(\Omega, \mathcal{F}, P)$ to any measure space $(X, \mathcal{B}, \mu)$; in particular, to $(\mathbf{N}, 2^{\mathbf{N}}, \nu)$, where ν is the counting measure. In this way it is easily verified that:

$$ (9.10) \qquad \big\| (a_k)_{k \geq 1} \big\|_{p,\infty} = \left(\sup_{t>0} t^p \, card \, \{ \, k \geq 1 : |a_k| > t \, \} \right)^{1/p} = \sup_{k \geq 1} k^{1/p} \, |a_k|^* , $$

where $|a_1|^* \geq |a_2|^* \geq \ldots$ is the non-increasing rearrangement of the sequence $|a_1|$, $|a_2|$, $\ldots$. The "norm" $\| \cdot \|_{p,\infty}$ appearing in (9.10) is called the *weak l^p-norm.* We moreover have:

$$ (9.11) \qquad \|(a_k)_{k \geq 1}\|_{p,\infty} \leq \left(\sum_{k=1}^{\infty} |a_k|^p \right)^{1/p} \leq \left(\frac{p}{p-r} \right)^{1/p} \|(a_k)_{k \geq 1}\|_{r,\infty} $$

for all $1 \leq r < p < \infty$. (Again, (9.10) and (9.11) extend to all $p > r > 0$.)

The extension of (9.6) mentioned above is now formulated as follows:

$$ (9.12) \qquad \mathbf{C}_p \, \big\| (a_k)_{k \geq 1} \big\|_{q,\infty} \leq \left\| \sum_{k=1}^{n} a_k \varepsilon_k \right\|_{\psi_p} \leq \mathbf{D}_p \, \big\| (a_k)_{k \geq 1} \big\|_{q,\infty} $$

for all $2 < p < \infty$ with $1/p + 1/q = 1$. (For proof see [87], p.94.) The values for the best constants $\mathbf{C}_p$ and $\mathbf{D}_p$ in (9.12) are not known.

3. We conclude this section by pointing out to possibilities of extending Khintchine's inequalities to "higher" dimensions. Namely, in the Khintchine's inequalities ε_k's may be seen as uniformly distributed on the unit sphere $S_1 = \{-1, 1\}$ in $\mathbf{R}^1$. Here we present the analogous results for the case of the unit sphere S_2 in $\mathbf{R}^2$ $(\equiv \mathbf{C})$.

Let $\{\varphi_k\}_{k \geq 1}$ be a sequence of independent random variables uniformly distributed on $[0, 2\pi[$. Then each (complex valued) random variable $\sigma_k = e^{i\varphi_k}$, $k \geq 1$, is uniformly distributed on $S_2 = \{ \, z \in \mathbf{C} : |z| = 1 \, \}$. The sequence $\{\sigma_k\}_{k \geq 1}$ is called a *Steinhaus* sequence (σ_k's are called Steinhaus random variables). Then the *Khintchine inequalities for Steinhaus variables* are valid:

$$(9.13)\qquad \mathbf{A}_p\Big(\sum_{k=1}^n |z_k|^2\Big)^{1/2} \le \Big\| \sum_{k=1}^n z_k e^{i\varphi_k} \Big\|_p \le \mathbf{B}_p\Big(\sum_{k=1}^n |z_k|^2\Big)^{1/2}$$

for all $z_1,\dots,z_n \in \mathbf{C}$ and all $0 < p < \infty$.

The best constants $\mathbf{A}_p$ and $\mathbf{B}_p$ in (9.13) were announced by Sawa (see [130], p.125-126), (however, as far as we know, no published proof of this exists by now):

$$(9.14)\qquad \mathbf{A}_p = \begin{cases} ? & , \quad 0 < p \le p_0 \\ \Big(\Gamma\big(\frac{p+2}{2}\big)\Big)^{1/p} & , \quad p_0 \le p \le 2 \; , \\ 1 & , \quad 2 \le p < \infty \; , \end{cases}$$

where p_0 $(= 0.47562)$ is the unique root of the equation:

$$2^{p/2}\,\Gamma\Big(\frac{p+1}{2}\Big) = \sqrt{\pi}\,\Big(\Gamma\Big(\frac{p+2}{2}\Big)\Big)^2 \quad , \quad 0 < p < 2 \; ,$$

while on the other hand:

$$(9.15)\qquad \mathbf{B}_p = \begin{cases} 1 & , \quad 0 < p \le 2 \\ \Big(\Gamma\big(\frac{p+2}{2}\big)\Big)^{1/p} & , \quad 2 \le p < \infty \; . \end{cases}$$

(The proof announced by Sawa [130] is purely computational. In this context (9.14)+(9.15) should be compared with (8.18)+(8.19), and (9.22) below should be noted.)

In fact, similarly to the Khintchine's approach in (2.5), one obtains (note that $E(\sigma_k)^2 = 0$, though $E|\sigma_k|^2 = 1$):

$$(9.16)\qquad E\Big|\sum_{k=1}^n z_k\, e^{i\varphi_k}\Big|^{2p} \le \; p!\,\Big(\sum_{k=1}^n |z_k|^2\Big)^p$$

for all $z_1,\dots,z_n \in \mathbf{C}$ and all $n \ge 1$, where the constant $p!$ is the best possible (Peškir, [118], 1993). Hence, by a similar method to the one presented in part 1 above, one gets:

$$(9.17)\qquad \Big\|\sum_{k=1}^n z_k\, e^{i\varphi_k}\Big\|_{\psi_2} \le \sqrt{2}\,\Big(\sum_{k=1}^n |z_k|^2\Big)^{1/2}$$

for all $z_1,\dots,z_n \in \mathbf{C}$ and all $n \ge 1$, where the constant $\sqrt{2}$ is the best possible (Peškir, [118], 1993). Moreover, the function:

$$(9.18)\qquad \big(|z_1|,\dots,|z_n|\big) \mapsto E\Big|\sum_{k=1}^n \sqrt{|z_k|}\, e^{i\varphi_k}\Big|^{2m}$$

is shown to be Schur-concave on $\mathbf{R}_+^n$ for all $m \ge 1$ (Peškir, [118], 1993). Hence, by Taylor expansion, the function:

$$(9.19)\qquad \big(|z_1|, \ldots, |z_n|\big) \mapsto E\left(\exp\left|\sum_{k=1}^n \sqrt{|z_k|}\, e^{i\varphi_k}\right|^2\right)$$

is Schur-concave on $\mathbf{R}_+^n$ as well.

It should be noted above that by the two-dimensional central limit theorem:

$$(9.20)\qquad \frac{1}{\sqrt{n}} \sum_{k=1}^n e^{i\varphi_k} \xrightarrow{d} Z_1 + iZ_2 \ ,$$

where $Z_1 \sim N(0,1)$ and $Z_2 \sim N(0,1)$ are independent, as well as that:

$$(9.21)\qquad \big\|Z_1 + iZ_2\big\|_{\psi_2} = \sqrt{2}$$

$$(9.22)\qquad \big\|Z_1 + iZ_2\big\|_p = \left(\Gamma\Big(\frac{p+2}{2}\Big)\right)^{1/p}$$

for all $0 < p < \infty$. (Recall (9.14)-(9.17) above.)

10. On the best constants in the inequalities of Burkholder ($p > 1$) and Davis ($p = 1$)

1. The inequalities of Burkholder (see (4.6)) state that for any $p > 1$ one finds such universal constants A_p and B_p , that for any martingale $f = (f_n)_{n \ge 1}$:

$$(10.1)\qquad A_p\, ES_n^p(f) \le E|f_n|^p \le B_p\, ES_n^p(f) \ ,$$

or, equivalently:

$$(10.2)\qquad \mathbf{A}_p\, \|S_n(f)\|_p \le \|f_n\|_p \le \mathbf{B}_p\, \|S_n(f)\|_p \ ,$$

with $\mathbf{A}_p = A_p^{1/p}$, $\mathbf{B}_p = B_p^{1/p}$.

In 1966 and 1988 Burkholder showed in [14] and [20], that the constants $\mathbf{A}_p$ and $\mathbf{B}_p$ can take the values:

$$(10.3)\qquad \mathbf{A}_p = \big(p^* - 1\big)^{-1} \ , \ \mathbf{B}_p = \big(p^* - 1\big) \ ,$$

with:

$$p^* = \max\,(p, q) \ , \ \frac{1}{p} + \frac{1}{q} = 1 \ .$$

Note, since:

$$p^* - 1 = \max\,\Big(p-1\,,\,(p-1)^{-1}\Big) \ ,$$

we find, that:

$$(10.4)\qquad \|f_n\|_p \le (p-1)\, \|S_n(f)\|_p \ , \ 2 \le p < \infty$$

(10.5) $$(p-1)\,\|S_n(f)\|_p \le \|f_n\|_p \quad , \quad 1 < p \le 2 \ .$$

Moreover, the constant $p-1$ in these inequalities appear to be as the best possible one. (The inequality (10.5) for $p \ge 3$ was established by Pittenger in [122], 1979. The inequalities (10.4) and (10.5) in this form are given in [20].)

It is interesting to observe that for $p \to \infty$ in the inequalities of Khintchine (see Section 8):

$$\mathbf{B}_p \sim \sqrt{\frac{p}{e}} \ .$$

In the martingale case (as it follows from (10.4)):

$$\mathbf{B}_p \sim p \ .$$

Note, that, as far as we know about, the values of the best constants $\mathbf{B}_p$ for $1 < p < 2$ and $\mathbf{A}_p$ for $p > 2$ are not known, although there are particular results about the order for these values. For example, when $p \to \infty$:

$$\mathbf{A}_p = O\big(p^{-1/2}\big) \quad \text{([17], [20])} \ ,$$

and if $p \to 1$:

$$\mathbf{B}_p \sim 1 \quad \text{([15], [16], [20])} \ .$$

Turn to the maximal inequalities:

(10.6) $$\mathbf{A}^*_p\,\|S_n(f)\|_p \le \|f^*_n\|_p \le \mathbf{B}^*_p\,\|S_n(f)\|_p$$

with $f^*_n = \max_{1 \le k \le n} |f_k|$. According to Doob's inequality for $1 < p < \infty$:

$$\|f^*_n\|_p \le q\,\|f_n\|_p \ .$$

Hence for $p > 1$ from (10.2) we find, that for $p \ge 2$:

$$\|f^*_n\|_p \le q\,\|f_n\|_p \le q(p-1)\,\|S_n(f)\|_p = p\,\|S_n(f)\|_p \ .$$

Moreover, the constant $\mathbf{B}^*_p = p$ appears to be the best possible, [20] (p.87).

11. On some refinements of the inequalities of Khintchine (the inequalities of Rosenthal, their modifications and extensions to the martingale case)

1. In the inequalities of Khintchine (see (2.4)) the control of behaviour of the quantity $E|f_n|^p$, where $f_n = \sum_{k=1}^n a_k \varepsilon_k$ is realized by the help of the quantity $S_n(f) = (\sum_{k=1}^n |a_k|^2)^{1/2}$, which is exactly the square-root of the quadratic variation:

(11.1) $$S^2_n(f) = \sum_{k=1}^{n} \big(\Delta f_k\big)^2 \ \big(= [f]_n \big) \ .$$

In the inequalities of Burkholder and Davis (see (4.6), (4.7)) the control of behaviour of the quantity $E|f_n|^p$ for martingale $f=(f_n)_{n\geq 1}$ is also realized by the help of the quadratic variation $[f]_n = S_n^2(f)$. For example, let us note, that the inequality (4.6) can be written in the following form:

$$A_p\, E[f]_n^{p/2} \;\leq\; E|f_n|^p \;\leq\; B_p\, E[f]_n^{p/2} \;. \tag{11.2}$$

Although the quadratic variation $[f]$ appears to be quite a natural "martingale" characteristic, it is, reasonably, far from being a unique natural "candidate", controlling the behaviour of the quantity $E|f_n|^p$. For example, another important characteristic is the predictable quadratic variation (the quadratic characteristic):

$$\langle f\rangle_n = \sum_{k=1}^{n} E\Big(\big(\Delta f_k\big)^2 \big|\mathcal{F}_{k-1}\Big) \;,$$

(see Section 4), for the square-root of which we will also use the notation:

$$\tilde{S}_n(f) = \left(\sum_{k=1}^{n} E\Big(\big(\Delta f_k\big)^2 \big|\mathcal{F}_{k-1}\Big)\right)^{1/2} \;. \tag{11.3}$$

It is well-known (see for instance [69]), that this predictable characteristic plays an important, and for many cases a crucial role in the study of various properties of the square-integrable martingales.

Taking into account these circumstances, and also, that for $p = 2$:

$$E\big|f_n\big|^2 = E\big[f\big]_n = E\big\langle f\big\rangle_n \;,$$

it is natural to derive inequalities of type (4.6), (4.7), where instead of the quantity $S_n(f) = [f]_n^{1/2}$ one considers the quantity $\tilde{S}_n(f) = \big\langle f\big\rangle_n^{1/2}$, or other related characteristics.

Such inequalities do exist indeed and with this in connection it is reasonable to begin the corresponding exposition with the so-called *inequalities of Wittle* [153], 1960, and *inequalities of Rosenthal* [128], 1970, for the case when $f_n = d_1 + \ldots + d_n$, $n \geq 1$, where $d_1, d_2, \ldots$ are independent random variables, continuing then with their extension to martingale sequences $f = (f_n)_{n\geq 1}$.

2. Let $f_n = d_1 + \ldots + d_n$, $n \geq 1$, where $d_1, d_2, \ldots$ is a sequence of independent random variables, having symmetric distribution. Then for any $p \geq 1$ it is evident (under the assumption $\|d_k\|_p < \infty$, $1 \leq k \leq n$):

$$\|f_n\|_p = \Big\|\sum_{k=1}^{n} d_k\Big\|_p \leq \sum_{k=1}^{n} \|d_k\|_p \;.$$

Since:

$$\sum_{k=1}^{n} \|d_k\|_p \;\leq\; \mathbf{C}_p(n)\Big(\sum_{k=1}^{n} \|d_k\|_p^2\Big)^{1/2} ,$$

then, it follows:

$$\|f_n\|_p \le \mathbf{C}_p(n)\Big(\sum_{k=1}^n \|d_k\|_p^2\Big)^{1/2} \tag{11.4}$$

with some constant depending on n and p ($\mathbf{C}_p(n) = \sqrt{n}$, for instance; by Jensen's inequality).

In year 1960, in the paper [153], P. Wittle found out that for $p \ge 3$ the inequality (11.4) remains valid with the constant not depending on n ; we have:

$$\|f_n\|_p \le \mathbf{B}_p\Big(\sum_{k=1}^n \|d_k\|_p^2\Big)^{1/2} \tag{11.5}$$

where $\mathbf{B}_p$ is exactly the same constant as in (8.18).

It is clear, that for the case of Khintchine, where $d_k = a_k\varepsilon_k$, the Wittle inequality (11.5) contains the inequality of Khintchine.

The next by importance and similar by its character, a result of Rosenthal, [128], 1970, appeared, who under the same assumptions on independence of random variables $d_1, d_2, \ldots$ with symmetric distribution has found, that for each $p \ge 2$:

$$\begin{aligned} \max\left\{\Big(\sum_{k=1}^n \|d_k\|_2^2\Big)^{1/2}, \Big(\sum_{k=1}^n \|d_k\|_p^p\Big)^{1/p}\right\} &\le \big\|f_n\big\|_p \\ &\le \mathbf{B}_p^0 \max\left\{\Big(\sum_{k=1}^n \|d_k\|_2^2\Big)^{1/2}, \Big(\sum_{k=1}^n \|d_k\|_p^p\Big)^{1/p}\right\}. \end{aligned} \tag{11.6}$$

Note that for $p \ge 2$ we have $\|d_k\|_2^2 \le \|d_k\|_p^2$, and:

$$\Big(\sum_{k=1}^n \|d_k\|_p^p\Big)^{1/p} \le \Big(\sum_{k=1}^n \|d_k\|_p^2\Big)^{1/2}.$$

From this:

$$\max\left\{\Big(\sum_{k=1}^n \|d_k\|_2^2\Big)^{1/2}, \Big(\sum_{k=1}^n \|d_k\|_p^p\Big)^{1/p}\right\} \le \Big(\sum_{k=1}^n \|d_k\|_p^2\Big)^{1/2}$$

which shows that the right-hand inequality of Rosenthal in (11.6) could be considered as an improvement upon the inequality of Wittle (11.5).

Our considerations below will show that the inequality (11.5) is equivalent to the inequality:

$$\mathbf{A}_p^1\left\{\tilde{S}_n(f) + \Big(E(d_n^*)^p\Big)^{1/p}\right\} \le \big\|f_n\big\|_p \le \mathbf{B}_p^1\left\{\tilde{S}_n(f) + \Big(E(d_n^*)^p\Big)^{1/p}\right\} \tag{11.7}$$

with some constants $\mathbf{A}_p^1$ and $\mathbf{B}_p^1$ and $d_n^* = \max_{1\le k\le n}|d_k|$.

The corresponding extension to the case of martingales ([17], [62]) is formulated in the following way: If $p \ge 2$, then:

$$\mathbf{A}_p^1\left\{\big\|\tilde{S}_n(f)\big\|_p + \|d_n^*\|_p\right\} \le \big\|f_n\big\|_p \le \mathbf{B}_p^1\left\{\big\|\tilde{S}_n(f)\big\|_p + \|d_n^*\|_p\right\}, \tag{11.8}$$

where:

(11.9) $$\|\tilde S_n(f)\|_p = \Big(E\big(\tilde S_n(f)\big)^p\Big)^{1/p} \ , \quad \|d_n^*\|_p = \Big(E(d_n^*)^p\Big)^{1/p} \ ,$$

$$\tilde S_n(f) = \left(\sum_{k=1}^n E\Big(\big(\Delta f_k\big)^2\big|\mathcal F_{k-1}\Big)\right)^{1/2} .$$

(Concerning the corresponding inequalities for $0 < p < 2$ see part 3 d) below.)

3. Let us concentrate on the basic points in the proof of the right-hand inequality in (11.8).

a) We begin with a remark, that for a proof of the inequality of type:

(11.10) $$EX^p \le K_p^p \, EY^p$$

for two non-negative random variables X and Y (and not "very small" p), it suffices to establish the validity of the so-called "good λ-inequality":

(11.11) $$P\big(X > \beta\lambda \, , \, Y \le \delta\lambda\big) \le 2\varepsilon \, P\big(X > \lambda\big) \ , \ \lambda > 0 \ ,$$

which should be valid for some $\delta > 0$, $\beta > 1$, $\varepsilon > 0$ such that $\beta^p\varepsilon < 1/2$.

Indeed, since:

$$P\big(X > \beta\lambda\big) \le P\big(Y > \delta\lambda\big) + P\big(X > \beta\lambda \, , \, Y \le \delta\lambda\big) \ ,$$

then from (11.11) we find, that:

$$P\big(X > \beta\lambda\big) \le P\big(Y > \delta\lambda\big) + 2\varepsilon \, P\big(X > \lambda\big) \ .$$

Multiplying both sides of this inequality by $p\lambda^{p-1}$ and then integrating over $\lambda > 0$, we find:

$$E\Big(\frac{X}{\beta}\Big)^p \le E\Big(\frac{Y}{\delta}\Big)^p + 2\varepsilon \, EX^p \ .$$

If $\varepsilon\beta^p < 1/2$, then from this we find, that:

(11.12) $$EX^p \le \Big(\frac{\beta}{\delta}\Big)^p \Big(1 - 2\varepsilon\beta^p\Big)^{-1} EY^p \ .$$

(Compare the argumentation just given with the derivation in (2.91).)

Supposing that (11.11) holds with:

$$\delta = \frac{1}{3p} \log\left(\frac{p}{\log p}\right) \ , \ \beta = 1 + \delta\left(\frac{2p}{\log p} - 1\right) \ ,$$

and $\varepsilon = \varepsilon(\alpha)$, where:

(11.13) $$\varepsilon(\alpha) = e^{-\alpha\Big(\log\big(1+\alpha\big) - 1\Big)}$$

with $\alpha = \big(\beta - 1 - \delta\big)/\delta$, we find (see Lemma 3.3 in [62]), that for $p \ge e^2$ the inequality (11.12) leads to the inequality:

(11.14) $$\|X\|_p \le K \frac{p}{\log p} \|Y\|_p ,$$

where K is some constant.

b) From the exposed it follows, that for the proof of the right-hand inequality in (11.8), having the form "almost equal" to (11.14), it is necessary for the underlying X and Y first of all to establish the validity of the corresponding "good λ-inequality".

With this in goal, following [17], [62], suppose that the martingale $f=(f_n)_{n\ge 1}$ is such, that the variables $d_n = f_n - f_{n-1}$ satisfy the inequality $|d_n| \le W_n$, $n\ge 1$, where $W=(W_n)_{n\ge 1}$ is some predictable sequence. (If $|d_n| \le$ const., $n \ge 1$, then, evidently, this condition is fulfilled.)

Put:

(11.15) $$X = f^* \quad , \quad Y = \max\{\tilde{S}(f), W^*\} ,$$

where $f^* = \max_{n\ge 1} |f_n|$, $\tilde{S}(f) = \lim_{n\to\infty} \tilde{S}_n(f)$ and $W^* = \max_{n\ge 1} W_n$.

It turns out, that under the given assumptions the inequality (11.11) is valid for $\delta > 0$, $\beta > 1+\delta$ and $\varepsilon = \varepsilon(\alpha)$, defined by (11.13), and at the basis of the corresponding proof it lies a martingale extension of the subgaussian inequality.

For the proof introduce, following [17], [62], the stopping times $(\beta > 1)$:

$$\tau_1 = \inf\{\, n\ge 1 : |f_n| > \lambda \,\}$$

$$\tau_\beta = \inf\{\, n\ge 1 : |f_n| > \beta\lambda \,\}$$

$$\tau = \inf\Big\{\, n\ge 1 : \tilde{S}_{n+1}(f) > \delta\lambda \ \text{ or } \ W^*_{n+1} > \delta\lambda \,\Big\} .$$

(If the sets $\{\, n : \cdot \,\} = \emptyset$, then the corresponding stopping times are to be equal $+\infty$.) Taking into account this notation, one may write, that:

(11.16) $$\Big\{\, f^* > \beta\lambda \,,\, \tilde{S}(f)\vee W^* \le \delta\lambda \,\Big\} = \{\, \tau_\beta < \infty \,,\, \tau = \infty \,\} .$$

Let us show, that:

(11.17) $$\{\, \tau_\beta < \infty \,,\, \tau = \infty \,\} \subset \Big\{\, \big|\textstyle\sum_{k=\tau_1+1}^{\tau_\beta\wedge\tau} d_k \big| \ge (\beta - 1 - \delta)\,\lambda \Big\} .$$

We have:

$$\{\, \tau_\beta < \infty \,,\, \tau = \infty \,\} = \Big\{\, \big|\textstyle\sum_{k=1}^{\tau_\beta} d_k \big| \,,\, \tau_\beta < \infty \,,\, \tau = \infty \Big\}$$

$$= \Big\{\, \big|\textstyle\sum_{k=1}^{\tau_\beta} d_k \big| \ge \beta\lambda \,,\, \tau_\beta < \infty \,,\, \tau = \infty \,,\, \tau_1 \le \tau_\beta \Big\}$$

$$= \Big\{\, \big|\textstyle\sum_{k=1}^{\tau_1} d_k + \sum_{k=\tau_1+1}^{\tau_\beta\wedge\tau} d_k \big| \ge \beta\lambda \,,\, \tau_\beta < \infty \,,\, \tau = \infty \,,\, \tau_1 \le \tau_\beta \Big\}$$

$$\subset \Big\{\, \big|\textstyle\sum_{k=1}^{\tau_1} d_k\big| + \big|\sum_{k=\tau_1+1}^{\tau_\beta\wedge\tau} d_k \big| \ge \beta\lambda \,,\, \tau_\beta < \infty \,,\, \tau = \infty \,,\, \tau_1 \le \tau_\beta \Big\}$$

$$\subset \Big\{\, \big|\textstyle\sum_{k=\tau_1+1}^{\tau_\beta\wedge\tau} d_k \big| \ge \beta\lambda - \big|\sum_{k=1}^{\tau_1} d_k\big| \,,\, \tau_\beta < \infty \,,\, \tau = \infty \,,\, \tau_1 \le \tau_\beta \Big\}$$

$$\subset \Big\{\, \big|\textstyle\sum_{k=\tau_1+1}^{\tau_\beta\wedge\tau} d_k \big| \ge \beta\lambda - (\lambda + \lambda\delta) \Big\} ,$$

which proves (11.17).

So, in this way we get:

$$(11.18)\qquad P\Big\{\, f^* > \beta\lambda\,,\ \tilde S(f)\vee W^* \le \delta\lambda \,\Big\} \le P\Big\{\, \big|\textstyle\sum_{k=\tau_1+1}^{\tau_\beta\wedge\tau} d_k \big| \ge (\beta-1-\delta)\,\lambda \,\Big\} =$$

$$= EP\Big(\, \big|\textstyle\sum_{k=\tau_1+1}^{\tau_\beta\wedge\tau} d_k \big| \ge (\beta-1-\delta)\,\lambda \;\big|\; \mathcal F_{\tau_1}\Big)\ .$$

To get the needed inequality (11.11), which under (11.15) gets the following form:

$$(11.19)\qquad P\Big\{\, f^* > \beta\lambda\,,\ \tilde S(f)\vee W^* \le \delta\lambda \,\Big\} \le 2\,\varepsilon(\alpha)\,P(f^* > \lambda)\ ,$$

it only remains (by force of (11.18)) to show, that:

$$(11.20)\qquad EP\Big(\, \big|\textstyle\sum_{k=\tau_1+1}^{\tau_\beta\wedge\tau} d_k \big| \ge (\beta-1-\delta)\,\lambda \;\big|\; \mathcal F_{\tau_1}\Big) \le 2\,\varepsilon(\alpha)\,P(f^* > \lambda)\ .$$

For the establishment of this inequality we need the following result.

Let $f=(f_n)_{n\ge 0}$ be a martingale with $f_0\equiv 0$ such that $|d_k|\le M$ $(P\text{-}a.s.)$, $k\ge 1$, where $d_k = f_k - f_{k-1}$, $M>0$, $\langle f\rangle_n = \tilde S_n^2(f)$. Then for every $\lambda>0$ the sequence:

$$(11.21)\qquad Z_n = \exp\Big\{\lambda f_n - \psi(\lambda)\,\langle f\rangle_n\Big\}\ ,\ n\ge 1\ ,$$

where $\psi(\lambda) = \big(e^{\lambda M} - 1 - \lambda M\big)/M^2$, is a supermartingale.

Indeed, since:

$$\Delta Z_n = Z_{n-1}\Big[\Big(e^{\lambda d_n}-1\Big)\,e^{-\Delta\langle f\rangle_n\psi(\lambda)} + \Big(e^{-\Delta\langle f\rangle_n\psi(\lambda)} - 1\Big)\Big]\ ,$$

then (using that $e^{\lambda x} - 1 - \lambda x \le x^2\psi(\lambda)$ for $|x|\le M$ and $xe^{-x} + (e^{-x}-1)\le 0$ for $x\ge 0$):

$$E\big(\Delta Z_n\,|\,\mathcal F_{n-1}\big) = Z_{n-1}\Big[e^{-\Delta\langle f\rangle_n\psi(\lambda)}\,E\Big(e^{\lambda d_n}-1\,\big|\,\mathcal F_{n-1}\Big) + \Big(e^{-\Delta\langle f\rangle_n\psi(\lambda)} - 1\Big)\Big]$$

$$= Z_{n-1}\Big[e^{-\Delta\langle f\rangle_n\psi(\lambda)}\,E\Big(e^{\lambda d_n}-1-\lambda d_n\,\big|\,\mathcal F_{n-1}\Big) + \Big(e^{-\Delta\langle f\rangle_n\psi(\lambda)} - 1\Big)\Big]$$

$$\le Z_{n-1}\Big[e^{-\Delta\langle f\rangle_n\psi(\lambda)}\,E\Big(\big(\Delta f_n\big)^2\psi(\lambda)\,\big|\,\mathcal F_{n-1}\Big) + \Big(e^{-\Delta\langle f\rangle_n\psi(\lambda)} - 1\Big)\Big]$$

$$= Z_{n-1}\Big[e^{-\Delta\langle f\rangle_n\psi(\lambda)}\,\Delta\langle f\rangle_n\psi(\lambda) + \Big(e^{-\Delta\langle f\rangle_n\psi(\lambda)} - 1\Big)\Big] \le 0\ ,$$

which proves the supermartingale property of the sequence $(Z_n)_{n\ge 0}$.

From (11.21) we find, that for any finite stopping time σ and every $\lambda>0$:

$$P\big(f_\sigma \ge x\big) = P\Big(e^{\lambda f_\sigma - \psi(\lambda)\langle f\rangle_\sigma} \ge e^{\lambda x - \psi(\lambda)\langle f\rangle_\sigma}\Big)$$

$$= P\Big(e^{\lambda f_\sigma - \psi(\lambda)\langle f\rangle_\sigma} \ge e^{\lambda x - \psi(\lambda)\langle f\rangle_\sigma}\,,\ \langle f\rangle_\sigma \le K^2\Big)$$

$$+ P\Big(e^{\lambda f_\sigma - \psi(\lambda)\langle f\rangle_\sigma} \ge e^{\lambda x - \psi(\lambda)\langle f\rangle_\sigma} \,,\, \langle f\rangle_\sigma > K^2\Big) \le$$

$$\le P\Big(e^{\lambda f_\sigma - \psi(\lambda)\langle f\rangle_\sigma} \ge e^{\lambda x - \psi(\lambda)K^2}\Big) + P\Big(\langle f\rangle_\sigma > K^2\Big)$$

$$\le e^{-\lambda x + \psi(\lambda)K^2} E e^{\lambda f_\sigma - \psi(\lambda)\langle f\rangle_\sigma} + P\Big(\langle f\rangle_\sigma > K^2\Big) \le e^{-\lambda x + \psi(\lambda)K^2} + P\Big(\langle f\rangle_\sigma > K^2\Big) \,,$$

where we used that for non-negative supermartingale $(Z_n)_{n\ge 0}$ we have $EZ_\sigma \le EZ_0 = 1$.

In this way we have derived the inequality:

$$P\big(f_\sigma \ge x\big) \le e^{-\lambda x + \psi(\lambda)K^2} + P\Big(\langle f\rangle_\sigma > K^2\Big) \,, \tag{11.22}$$

from which, in particular, it follows, that if $(P\text{-}a.s.)$

$$\langle f\rangle_\sigma = \tilde{S}^2_\sigma(f) = \sum_{k=1}^{\sigma} E\big(d_k^2 \mid \mathcal{F}_{k-1}\big) \le K^2 \,,$$

then

$$P\big(f_\sigma \ge x\big) \le e^{-\lambda x + \psi(\lambda)K^2} \,. \tag{11.23}$$

It is useful to note, that since for $y \ge 0$:

$$e^y - y - 1 \le e^y + e^{-y} - 2 = 2\big(\cosh(y) - 1\big) \le y\sinh(y) \,,$$

then from (11.23):

$$P\big(f_\sigma \ge x\big) \le \exp\Big(-\lambda x + \psi(\lambda)K^2\Big) \le \exp\left(-\lambda x + \frac{\lambda M \sinh(\lambda M)}{M^2}\, K^2\right)$$

and:

$$P\big(f_\sigma \ge x\big) \le \exp\left(\inf_{\lambda>0}\Big(-\lambda x + \psi(\lambda)K^2\Big)\right) \tag{11.24}$$

$$\le \exp\left(\inf_{\lambda>0}\left(-\lambda x + \frac{\lambda M \sinh(\lambda M)}{M^2}\, K^2\right)\right) .$$

In the first inequality the infimum is attained at $\lambda = \lambda_0$, where:

$$e^{\lambda_0 M} - 1 = \frac{xM}{K^2} \,,$$

i.e.

$$\lambda_0 = \frac{1}{M}\log\left(1 + \frac{xM}{K^2}\right) .$$

Hence:

$$P\big(f_\sigma \ge x\big) \le \exp\left(-\frac{x}{M}\left[\log\left(1+\frac{xM}{K^2}\right) - 1\right]\right) \cdot \exp\left(-\Big(\frac{K}{M}\Big)^2 \log\left(1 + \frac{xM}{K^2}\right)\right) . \tag{11.25}$$

In the second inequality in (11.24) we shall not make a minimization over $\lambda > 0$, but will take $\lambda = \lambda_1$, where:

$$\lambda_1 = \frac{1}{M} \text{ arcsh}\left(\frac{xM}{2K^2}\right) .$$

Then:

$$\exp\left(-\lambda_1 x + \frac{\lambda_1 M \sinh(\lambda_1 M)}{M^2} K^2\right) = \exp\left(-\frac{\lambda_1 x}{2}\right) = \exp\left(-\frac{x}{2M} \text{ arcsh}\left(\frac{xM}{2K^2}\right)\right) .$$

In this way we get:

$$\text{(11.26)} \qquad P\big(f_\sigma \geq x\big) \leq \exp\left(-\frac{x}{M}\left[\log\left(1+\frac{xM}{K^2}\right)-1\right]-\left(\frac{K}{M}\right)^2 \log\left(1+\frac{xM}{K^2}\right)\right)$$

$$\leq \exp\left(-\frac{x}{2M} \text{ arcsh}\left(\frac{xM}{2K^2}\right)\right) .$$

For our aim it is sufficient to use the following estimate which is evident from (11.25):

$$\text{(11.27)} \qquad P\big(f_\sigma \geq x\big) \leq \exp\left(-\frac{x}{M}\left[\log\left(1+\frac{xM}{K^2}\right)-1\right]\right) .$$

(Compare with [62], [70]. The second inequality obtained in (11.26) is a natural extension of the well-known “Arcsh-inequality” of Prokhorov ([123], 1959).) From (11.27):

$$\text{(11.28)} \qquad P\big(|f_\sigma| \geq x\big) \leq 2 \exp\left(-\frac{x}{M}\left[\log\left(1+\frac{xM}{K^2}\right)-1\right]\right) .$$

Passing, if necessary, from the stopping time σ to $\sigma \wedge N$ with the limit passage when $N \to \infty$, we get, that (11.28) holds true for all stopping times $\sigma \leq \infty$, where $|f_\infty| = \lim_{n\to\infty} |f_n|$.

Remark. Let γ be a stopping time, and let $\mathcal{F}_\gamma$ be a family of sets $A \in \mathcal{F}$, such that $A \cap \{\gamma \leq n\} \in \mathcal{F}_n$ for all $n \geq 1$. It is certain that $\mathcal{F}_\gamma$ is a σ-algebra.

With insignificant modifications of the proof given above it is possible to show, that if γ_1 and γ_2 are stopping times such that $\gamma_1 \leq \gamma_2$ $(P\text{-}a.s.)$, then $\{\gamma_1 < \infty ; P\text{-}a.s.\}$:

$$\text{(11.28')} \qquad P\Big\{ |f_{\gamma_2} - f_{\gamma_1}| \geq x \,\big|\, \mathcal{F}_{\gamma_1}\Big\} \leq 2 \exp\left(-\frac{x}{M}\left[\log\left(1+\frac{xM}{K^2}\right)-1\right]\right) .$$

c) Let us turn back to the proof of the inequality (11.20). If we put $b_k = I\big(\tau_1 < k \leq \tau_\beta \wedge \tau\big)$, $k \geq 1$, and form the variables:

$$f_n(b) = \sum_{k=1}^{n} b_k d_k ,$$

then one may note, that by force of the $\mathcal{F}_{k-1}$-measurability of b_k , $k \geq 1$, the sequence

$f(b) = \big(f_n(b)\big)_{n\geq 1}$ forms a martingale. Moreover, we have:

$$\sum_{k=1}^{\infty} E\Big(\big(b_k d_k\big)^2 \mid \mathcal{F}_{k-1}\Big) \le \delta^2\lambda^2 ,$$

and $|b_k d_k| \le \delta\lambda$. Hence from the inequality (11.28') being applied to $f_n = f_n(b)$, we find (with $M = \delta\lambda$, $K = \delta^2\lambda^2$) , that:

$$\begin{aligned} &EP\Big(\,\big|\textstyle\sum_{k=\tau_1+1}^{\tau_\beta\wedge\tau} d_k \,\big| \ge (\beta-1-\delta)\,\lambda \mid \mathcal{F}_{\tau_1}\Big) \\ &\le 2\exp\Big(-\alpha\big(\log(1+\alpha)-1\big)\Big) EI\big(\tau_1<\infty\big) \\ &= 2\exp\Big(-\alpha\big(\log(1+\alpha)-1\big)\Big) EI\big(f^*>\lambda\big) \,\Big(= 2\,\varepsilon(\alpha)\,P(f^*>\lambda)\Big) , \end{aligned}$$

where $\alpha = (\beta-1-\delta)/\delta$ (and $\varepsilon(\alpha) = \exp\big(-\alpha\big(\log(1+\alpha)-1\big)\big)$) , which, as already mentioned above, proves the needed "good λ-inequality".

d) Resuming done, we get the following. If $X = f^*$, $Y = \max\big(\tilde S(f) \,,\, W^*\big)$, where $f^* = \sup_{n\ge 1}|f_n|$, $W^* = \sup_{n\ge1} W_n$, $\tilde S(f) = \sup_{n\ge 1}\tilde S(f)$, $|d_n| \le W_n$, $n\ge 1$, then (see (11.14)) for $p \ge e^2$:

$$\text{(11.29)} \qquad \big\|f^*\big\|_p \le K_1 \frac{p}{\log p} \Big\| \max\big(\tilde S(f) \,,\, W^*\big)\Big\|_p .$$

Let us now show how from this inequality it follows the inequality:

$$\text{(11.30)} \qquad \big\|f^*\big\|_p \le K_2 \frac{p}{\log p} \Big\{ \big\|\tilde S(f)\big\|_p + \big\|d^*\big\|_p \Big\} ,$$

restricting ourselves to the case when the variables d_n , $n \ge 1$, have symmetric conditional distribution $P\big(d_n \le x \mid \mathcal{F}_{n-1}\big)$, $n\ge 1$. (The general case can be reduced to this one; see [62].)

Let, as in Section 7, $I_n = I\big(|d_n| > 2d^*_{n-1}\big)$ and:

$$d'_n = d_n \bar I_n \quad , \quad d''_n = d_n I_n .$$

So, we have $d_n = d'_n + d''_n$. By force of the assumed conditional symmetry:

$$E\big(d'_n \mid \mathcal{F}_{n-1}\big) = E\big(d''_n \mid \mathcal{F}_{n-1}\big) = 0 .$$

Consequently, both of the sequences $d' = (d'_n)_{n\ge1}$ and $d'' = (d''_n)_{n\ge1}$ are martingale differences.

Put $f'_n = \sum_{k=1}^n d'_k$, $f''_n = \sum_{k=1}^n d''_k$. Estimate $\big\|(f'')^*\big\|$. With this in goal, note that, as already remarked in Section 7, on the set $\big\{|d_n| > 2d^*_{n-1}\big\}$ we have $|d''_n| \le 2\,(d^*_n - d^*_{n-1})$. Thus $\sum_{n=1}^{\infty} |d''_n| \le 2\,d^*$ and:

$$\text{(11.31)} \qquad \big\|(f'')^*\big\| = \Big\| \sup_{n\ge1} \big|\textstyle\sum_{k=1}^n d''_k\,\big| \Big\|_p \le \Big\| \textstyle\sum_{k=1}^{\infty} |d''_k| \Big\|_p \le 2\,\big\|d^*\big\|_p .$$

To estimate $\big\|(f')^*\big\| \equiv \big\|\sup_{n\ge1}\big|\sum_{k=1}^n d'_k\big|\,\big\|_p$, one may use considerations from b), taking

$f_n = f'_n$ and noticing, that $|d'_n| \leq W_n$ with $W_n = 2d^*_{n-1}$, so from (11.29) we conclude, that:

$$\|(f')^*\|_p \leq K_1 \frac{p}{\log p} \left\| \max\left(\tilde{S}(f') \,,\, W^*\right) \right\|_p \tag{11.32}$$

$$\leq K_1 \frac{p}{\log p} \left[\|\tilde{S}(f')\|_p + \|2d^*\|_p \right] \leq 2K_1 \frac{p}{\log p} \left[\|\tilde{S}(f')\|_p + \|d^*\|_p \right] .$$

From (11.31) and (11.32) we get (for $p \geq e^2$) the desired inequality (11.30).

Moreover, as it is shown in [17], the inequality (11.30) remains valid for $2 \leq p < e^2$ as well (with some universal constant on the right-hand side). So, for any $p \geq 2$ and all $n \geq 1$:

$$\|f_n\|_p \leq \mathbf{B}^1_p \left\{ \|\tilde{S}_n(f)\|_p + \|d^*_n\|_p \right\} \tag{11.33}$$

with:

$$\mathbf{B}^1_p = K \cdot \frac{p}{\log p} \ , \tag{11.34}$$

where K is some universal constant, not depending neither on n , nor on the structure of martingale f .

It is important to note in this, that from the point of view of order (when $p \to \infty$) the quantity $\mathbf{B}^1_p \approx p/\log\ p$ is the best possible, because it is already the best possible in the case of independent random variables d_n , $n \geq 1$; see [70].

Note also, that in the case $0 < p \leq 2$ the inequality (11.33) can be improved upon in the sense, that the following inequality holds:

$$\|f_n\|_p \leq \mathbf{B}^1_p \, \|\tilde{S}_n(f)\|_p$$

with:

$$\mathbf{B}^1_p = \sqrt{\frac{2}{p}} \ ,$$

where the constant is the best possible (Wang, [152], 1991).

BIBLIOGRAPHY

[1] ASSOUAD, P. (1975). Burgess-Davis inequalities in Banach spaces. *Proc. Sem. Random Series, Convex Sets, Geom. Banach Spaces (Aarhus, Denmark, 1974). Various Publ. Ser.* 24 (1-20).

[2] BANUELOS, R. (1986). Martingale transforms and related singular integrals. *Trans. Amer. Math. Soc.* 293 (547-563).

[3] BARLOW, M. T., JACKA, S. D. *and* YOR, M. (1986). Inequalities for a pair of processes stopped at a random time. *Proc. London Math. Soc. (3)*, 52 (142-172).

[4] BASSILY, N. L. (1987). A proof of the left-hand side of the Burkholder-Davis-Gundy inequality for a large class of Orlicz spaces. *Ann. Univ. Sci. Budapest Eotvos Sect. Math.* 30 (195-199).

[5] BASSILY, N. L. (1988). A new proof of the right hand side of the Burkholder-Davis-Gundy inequality. *Probab. Theory Math. Statist. Appl., Proc. 5th Pannonian Symp. (Visegrad, Hungary, 1985)* (7-21).

[6] BASSILY, N. L. (1991). On two-sided inequalities for stopped random walks in Orlicz-spaces. *Ann. Univ. Sci. Budapest Eotvos Sect. Comput.* 11 (63-73).

[7] BASSILY, N. L. (1991). Duality of the Burkholder-Davis-Gundy inequality and the generalized Fefferman-Garsia inequality. I. *Approx. Theory, Proc. Conf. (Kecskemet, Hungary, 1990). Colloq. Math. Soc. Janos Bolyai* 58 (85-96).

[8] BASSILY, N. L. (1992). Duality of the Burkholder-Davis-Gundy inequality and the generalized Fefferman-Garsia inequality. II. *Probab. Theory Appl., Essays Mem. J. Mogyorodi, Math. Appl.* 80 (33-45).

[9] BERTOIN, J. (1988). Une extension d'une inegalite de Burkholder, Davis, Gundy pour les processus a α-variation bornee et applications. *Stochastics* 24 (75-86).

[10] BOCHKARIEV, S. V. (1987). Probabilistic method of approach to certain problems of Fourier analysis. *Probab. Theory Appl., Proc. World Congress Bernoulli Soc., (Tashkent, USSR, 1986)*, Vol. 1 (567-570).

[11] BOLAND, P. J. *and* PROSCHAN, F. (1987). Schur convexity of the maximum likelihood function for the multivariate hypergeometric and multinomial distributions. *Stat. Probab. Lett.* 5 (317-322).

[12] BOLLOBAS, B. (1980). Martingale inequalities. *Math. Proc. Cambridge Philos. Soc.* 87 (377-382).

[13] BOREL, E. (1909). Les probabilites denombrables et leurs applications arithmetiques. *Rend. Circ. Mat. Palermo* 27 (247-271).

[14] BURKHOLDER, D. L. (1966). Martingale transforms. *Ann. Math. Statist.* 37 (1494-1504).

[15] BURKHOLDER, D. L. *and* GUNDY, R. F. (1970). Extrapolation and interpolation of quasi-linear operators on martingales. *Acta Math.* 124 (249-304).

[16] BURKHOLDER, D. L., DAVIS, B. *and* GUNDY, R. F. (1972). Integral inequalities for convex functions of operators on martingales. *Proc. Sixth Berkeley Symp. Math. Statist. Prob.* 2 (223-240).

[17] BURKHOLDER, D. L. (1973). Distribution function inequalities for martingales. *Ann. Probab.* 1 (19-42).

[18] BURKHOLDER, D. L. (1982). A nonlinear partial differential equation and the unconditional constant of the Haar system in L^p . *Bull. Amer. Math. Soc. (N.S.)* 7 (591-595).

[19] BURKHOLDER, D. L. (1984). Boundary value problems and sharp inequalities for martingale transforms. *Ann. Probab.* 12 (647-702).

[20] BURKHOLDER, D. L. (1988). Sharp inequalities for martingales and stochastic integrals. *Asterisque* 157-158 (75-94).

[21] BURKHOLDER, D. L. (1989). Differential subordination of harmonic functions and martingales. *Proc. Harmonic Anal. Part. Diff. Eq. (El Escorial, Spain, 1987). Lecture Notes in Math.* 1384 (1-23).

[22] BURKHOLDER, D. L. (1991). Explorations in martingale theory and its applications. *Calcul des probabilites, Ec. d'Ete (Saint-Flour, France, 1989). Lecture Notes in Math.* 1464 (1-66).

[23] BURKHOLDER, D. L. (1994). Strong differential subordination and stochastic integration. *Ann. Probab.* 22 (995-1025).

[24] CADET, V. (1994). Une extension des inegalites de Khintchine-Kahane dans les espaces de Banach. *Analyse Fonctionnelle et Harmonique 1992-1993. Publ. Math. Orsay* 92-94 (155-184).

[25] CARLEN, E. *and* KREE, P. (1991). L^p estimates on iterated stochastic integrals. *Ann. Probab.* 19 (354-368).

[26] CHAO, J-A. *and* LONG, R-L. (1992). Martingale transforms and Hardy spaces. *Probab. Theory Related Fields* 91 (399-404).

[27] CHOI, K. P. (1988). Some sharp inequalities for martingale transforms. *Trans. Am. Math. Soc.* 307 (279-300).

[28] CHOW, Y. *and* TEICHER, H. (1988). *Probability Theory. Independence, Interchangeability, Martingales.* Springer-Verlag.

[29] COTLAR, M. *and* SADOSKY, C. (1983). Vector valued inequalities of Marcinkiewicz-Zygmund and Grothendieck type for Toeplitz forms. *Harmonic Analysis, Proc. Conf. (Cortona, Italy, 1982). Lecture Notes in Math.* 992 (276-308).

[30] COX, D. C. (1982). The best constant in Burkholder's weak-L^1 inequality for the martingale square function. *Proc. Amer. Math. Soc.* 85 (427-433).

[31] DAVIS, B. (1970). On the integrability of the martingale square function. *Israel J. Math.* 8 (187-190).

[32] DAVIS, B. (1976). On the L^p norms of stochastic integrals and other martingales. *Duke Math. J.* 43 (697-704).

[33] DAVIS, B. (1978). On stopping times for n dimensional Brownian motion. *Ann. Probab.* 6 (651-659).

[34] DEBLASSIE, R. D. (1987). Stopping times of Bessel processes. *Ann. Probab.* 15 (1044-1051).

[35] DELLACHERIE, C. *and* MEYER, P-A. (1980). Probabilites et potentiel. Chapitres V a VIII: Theorie des martingales. *Publications de l'Institut de Mathematique de l'Universite de Strasbourg, XVII. Paris: Hermann XVIII.*

[36] DODSON, M. M., RYNNE, B. P. *and* VICKERS, J. A. G. (1991). Khintchine-type theorems on manifolds. *Acta Arith.* 57 (115-130).

[37] DUBINS, L. E. *and* SCHWARZ, G. (1988). A sharp inequality for sub-martingales and stopping-times. *Asterisque* 157-158 (129-145).

[38] DUBINS, L. E., SHEPP, L. A. *and* SHIRYAEV, A. N. (1993). Optimal stopping rules and maximal inequalities for Bessel processes. *Theory Probab. Appl.* 38 (English transl.) (288-330).

[39] EATON, M. L. (1970). A note on symmetric Bernoulli random variables. *Ann. Math. Statist.* 41 (1223-1226).

[40] EFRON, B. (1969). Student's t-test under symmetry conditions. *J. Amer. Statist. Assoc.* 64 (1278-1302).

[41] FIGIEL, T., HITCZENKO, P., JOHNSON, W. B., SCHECHTMAN, G. *and* ZINN, J. (1994). Extremal properties of Rademacher functions with applications to the Khintchine and

Rosenthal inequalities. *Preprint*.

[42] FRANGOS, N. E. *and* IMKELLER, P. (1987). Quadratic variation for a class of $L\log^{+}L$-bounded two-parameter martingales. *Ann. Probab.* 15 (1097-1111).

[43] FRANGOS, N. E. *and* IMKELLER, P. (1988). Some inequalities for strong martingales. *Ann. Inst. H. Poincare Probab. Statist.* 24 (395-402).

[44] GALCHUK, L. I. (1989). On the Lenglart and Burkholder-Davis-Gundy inequality for optional martingales. *Statistics and Control of Stochastic Processes, Collect. Artic. Moskva* (24-28).

[45] GARSIA, A. M. (1973). The Burgess-Davis inequalities via Fefferman's inequality. *Ark. Mat.* 11 (229-237).

[46] GIKHMAN, I. I. (1982). Two-parameter martingales. *Usp. Mat. Nauk.* 37 (3-28); *Russ. Math. Surv.* 37 (1-29).

[47] GLUSKIN, E. D., PIETSCH, A. *and* PUHL, J. (1980). A generalization of Khintchine's inequality and its application in the theory of operator ideals. *Studia Math.* 67 (149-155).

[48] GORIN, E. A. *and* FAVOROV, S. Y. (1990). Variants of the Khinchin inequality. *Issled. Teor. Funkts. Mnogikh Veshchestv. Perem.* 1990 (Russian) (52-63).

[49] GORIN, E. A. *and* FAVOROV, S. Y. (1990). Generalization of the Khinchin inequality. *Theory Probab. Appl.* 35 (English transl.) (766-771).

[50] On Wald-type optimal stopping for Brownian motion. *Math. Inst. Aarhus, Preprint Ser.* No. 10, 1994, (with S. E. Graversen), (15 pp). *J. Appl. Probab.* Vol. 34, No. 1, 1997, (66-73).

[51] GRAVERSEN, S. E *and* PESKIR, G. (1994). On Wald-type optimal stopping for Brownian motion. *Math. Inst. Aarhus, Preprint Ser.* No. 10, (15 pp). *J. Appl. Probab.* 34, 1997 (66-73).

[52] GRAVERSEN, S. E *and* PESKIR, G. (1994). Maximal inequalities for Bessel processes. *Math. Inst. Aarhus, Preprint Ser.* No. 23, (15 pp). *J. Inequal. Appl.* 2, 1998 (99-119).

[53] GUNDY, R. F. (1966). Martingale theory and pointwise convergence of certain orthogonal series. *Trans. Amer. Math. Soc.* 124 (228-248).

[54] GUNDY, R. F. (1989). Some martingale inequalities with applications to harmonic analysis. *J. Funct. Anal.* 87 (212-230).

[55] GUSHCHIN, A. A. *and* MISHURA, Y. S. (1991). The Davis inequalities and the Gundy decomposition for two-parameter strong martingales. I. *Theory Probab. Math. Stat.* 42 (English transl.) (29-37).

[56] GUSHCHIN, A. A. *and* MISHURA, Y. S. (1991). The Davis inequalities and the Gundy decomposition for two-parameter strong martingales. II. *Theory Probab. Math. Stat.* 43 (English transl.) (65-76).

[57] GUSHCHIN, A. A. *and* MISHURA, Y. S. (1992). Davis inequalities and the Gundy decomposition for two-parametric strong martingales. III. *Theory Probab. Math. Stat.* 43 (English transl.) (45-51).

[58] HAAGERUP, U. (1982). The best constants in the Khintchine inequality. *Studia Math.* 70 (231-283).

[59] HALL, R. R. (1975). On a conjecture of J. E. Littlewood. *Proc. Cambridge Philos. Soc.* 78 (443-445).

[60] HARDY, G. H. *and* LITTLEWOOD, J. E. (1914). Some problems of Diophantine

approximations. *Acta Math.* 37 (155-239).

[61] HARDY, G. H., LITTLEWOOD, J. E. *and* POLYA, G. (1929). Some simple inequalities satisfied by convex functions. *Messenger Math.* 58 (145-152).

[62] HAUSDORFF, F. (1913). *Grundzuge der Mengenlehre*. Viet, Leipzig.

[63] HITCZENKO, P. (1990). Best constants in martingale version of Rosenthal's inequality. *Ann. Probab.* 18 (1656-1668).

[64] HITCZENKO, P. (1993). Domination inequality for martingale transforms of a Rademacher sequence. *Israel J. Math.* (161-178).

[65] HITCZENKO, P. (1994). On a domination of sums of random variables by sums of conditionally independent ones. *Ann. Probab.* 22 (453-468).

[66] HOFFMANN-JØRGENSEN, J. (1994). Maximal inequalities for random elements. *Proc. Conf. Functional Anal. IV (Dubrovnik, Croatia, 1993). Various Publ. Ser.* 43 (105-158).

[67] IMKELLER, P. (1989). On inequalities for two-parameter martingales. *Math. Ann.* 284 (329-341).

[68] ISHAK, S. (1992). Remarks on a theorem of Burkholder and Gundy. *J. Theoret. Probab.* 5 (153-158).

[69] JACKA, S. D. (1991). Optimal stopping and best constants for Doob-like inequalities I: the case $p = 1$. *Ann. Probab.* 19 (1798-1821).

[70] JACOD, J. *and* SHIRYAEV, A. N. (1987). *Limit Theorems for Stochastic Processes*. Springer-Verlag Berlin Heidelberg.

[71] JOHNSON, W. B., SCHECHTMAN, G. *and* ZINN, J. (1985). Best constants in moment inequalities for linear combinations of independent and exchangeable random variables. *Ann. Probab.* 13 (234-253).

[72] KAHANE, J. P. (1968 & 1985). *Some Random Series of Functions*. D. C. Heath & Co. (first edition) & Cambridge University Press (second edition).

[73] KAIGH, W. D. *and* SORTO, M. A. (1993). Subsampling quantile estimator majorization inequalities. *Statist. Probab. Lett.* 18 (373-379).

[74] KALLENBERG, O. *and* SZTENCEL, R. (1991). Some dimension-free features of vector-valued martingales. *Probab. Theory Related Fields* 88 (215-247).

[75] KHINTCHINE, A. (1923). Uber dyadische Bruche. *Math. Z.* 18 (109-116).

[76] KHINTCHINE, A. (1924). Uber einen Satz der Wahrscheinlichkeitsrechnung. *Fund. Math.* 6 (9-20).

[77] KHOSHNEVISAN, D. (1992). Moment inequalities for functionals of the Brownian convex hull. *Ann. Probab.* 20 (627-630).

[78] KIKUCHI, M. *and* KAZAMAKI, N. (1987). Martingale inequalities for ratios of the maximal function and the square function. *Math. Rep. Toyama Univ.* 10 (181-186).

[79] KOLMOGOROV, A. N. (1929). Uber das Gesetz des iterierten Logarithmus. *Math. Ann.* 101 (126-135).

[80] KOMOROWSKI, R. (1988). On the best possible constants in the Khintchine inequality for $p \geq 3$. *Bull. London Math. Soc.* 20 (73-75).

[81] KOROLYUK, V. S. *and* BOROVSKIKH, Y. V. (1988). *Martingale Approximation.* Ed. by V. L. Girko. Kiev: Naukova Dumka.

[82] KRYLOV, N. V. (1989). Some estimates from stochastic analysis. *Statistics and Control of Stochastic Processes, Collect. Artic. Moskva* (96-100).

[83] KURITSYN, Y. G. (1984). On an inequality of A. Ya. Khinchin. *Problems of Stability of Stochastic Models, Proc. Sem. Moskva 1984* (92-94).

[84] KURITSYN, Y. G. (1986). The Khinchin inequality. *J. Sov. Math.* 35 (2363-2365).

[85] KWAPIEN, S. *and* WOYCZYNSKI, W. A. (1989). Tangent sequences of random variables: Basic inequalities and their applications. *Almost everywhere convergence, Proc. Int. Conf. (Columbus, Ohio, 1988)* (237-265).

[86] LEDOUX, M. (1981). Inegalites de Burkholder pour martingales indexees par $N \times N$. *Processus aleatoires a deux indices, Proceedings, Paris 1980, Lecture Notes in Math.* 863 (122-127).

[87] LEDOUX, M. (1985). Sur une inegalite de H. P. Rosenthal et le theoreme limite central dans les espaces de Banach. *Israel J. Math.* 50 (290-318).

[88] LEDOUX, M. *and* TALAGRAND, M. (1991). *Probability in Banach Spaces. Isoperimetry and Processes.* Springer-Verlag Berlin Heidelberg.

[89] LENGLART, E. (1980). Martingales et series stochastiques en temps discret. *Rev. Cethedec* 65 (19-92).

[90] LIPTSER, R. Sh. *and* SHIRYAEV, A. N. (1989). *Theory of Martingales.* Kluwer Acad. Publ.

[91] LIU, P. (1990). Davis inequalities of B-valued martingales and the geometrical properties of Banach spaces. *Chinese Ann. Math. Ser. A*, 11 (206-211).

[92] LONG, R. (1993). *Martingale Spaces and Inequalities.* Peking Univ. Press.

[93] LUECKING, D. H. (1993). Embedding theorems for spaces of analytic functions via Khinchine's inequality. *Michigan Math. J.* 40 (333-358).

[94] LUST-PIQUARD, F. (1986). Inegalites de Khintchine dans C_p $(1 < p < \infty)$. *C. R. Acad. Sci. Paris Ser. I*, 303 (289-292).

[95] LUST-PIQUARD, F. *and* PISIER, G. (1991). Non commutative Khintchine and Paley inequalities. *Ark. Mat.* 29 (241-260).

[96] LUXEMBURG, W. A. J. (1972). On an inequality of A. Khintchine for zero-one matrices. *J. Combin. Theory Ser. A*, 12 (289-296).

[97] MARCINKIEWICZ, J. (1937). Quelques theoremes sur les series orthogonales. *Ann. Soc. Polon. Math.* 16 (241-264).

[98] MARCINKIEWICZ, J. *and* ZYGMUND, A. (1938). Quelques theoremes sur les fonctions independantes. *Studia Math.* 7 (104-120).

[99] MARSHALL, A. W. *and* OLKIN, I. (1979). *Inequalities: Theory of Majorization and Its Applications.* Academic Press.

[100] MATSAK, I. K. *and* PLICHKO, A. N. (1988). The Khinchin inequality for k-multiple products of independent random variables. *Math. Notes* 44 (690-694).

[101] MATSAK, I. K. *and* PLICHKO, A. N. (1990). Khinchin's inequality and the asymptotic behavior of the sums $\sum \varepsilon_n x_n$ in Banach lattices. *Ukrainian Math. J.* 42 (566-570).

[102] MATUS, F. *and* TUZAR, A. (1992). Short proofs of Khintchine-type inequalities for zero-one matrices. *J. Combin. Theory Ser. A*, 59 (155-159).

[103] MCCONNELL, T. R. *and* TAQQU, M. S. (1986). Decoupling inequalities for multilinear

forms in independent symmetric random variables. *Ann. Probab.* 14 (943-954).

[104] McEACHIN, R. (1992). A sharp estimate in an operator inequality. *Proc. Amer. Math. Soc.* 115 (161-165).

[105] MILLAR, P. W. (1968). Martingale integrals. *Trans. Amer. Math. Soc.* 133 (145-166).

[106] MILMAN, V. *and* PAJOR, A. (1989). Cas limites dans des inegalites du type de Khinchine et applications geometriques. *C. R. Acad. Sci. Paris Ser. I*, 308 (91-96).

[107] MISHURA, Y. S. (1981). On properties of the quadratic variation of two-parameter strong martingales. *Theory Probab. Math. Stat.* 25 (English transl.) (99-106).

[108] MOGYORODI, J. (1981). Duality of the maximal inequality for non-negative submartingales and of the convexity inequality of Burkholder. *Math. Statist. 1st Pannonian Symp. (Bad Tatzmannsdorf, Austria, 1979). Lecture Notes in Statist.* 8 (165-173).

[109] MONAT, P. (1994). Remarques sur les inegalites de Burkholder-Gundy. *Sem. Probab. XXVIII, Lecture Notes in Math.* 1583 (92-97).

[110] NEWMAN, C. M. (1975). An extension of Khintchine's inequality. *Bull. Amer. Math. Soc.* 81 (913-915).

[111] NOVIKOV, A. A. (1971). On stopping times for a Wiener process. *Theory Probab. Appl.* 16 (English transl.) (449-456).

[112] NOVIKOV, A. A. (1971). On moment inequalities for stochastic integrals. *Theory Probab. Appl.* 16 (English transl.) (538-541).

[113] NOVIKOV, I. (1991). Martingale inequalities in rearrangement invariant function spaces. *Function Spaces, Proc. 2nd Int. Conf. (Poznan, Polland, 1989). Teubner-Texte Math.* 120 (120-127).

[114] OSTROWSKI, A. M. (1952). Sur quelques applications des fonctions convexes et concaves au sens de I. Schur. *J. Math. Pures Appl.* 31 (253-292).

[115] PALEY, R. E. A. C. (1932). A remarkable series of orthogonal functions I. *Proc. London Math. Soc.* 34 (241-264).

[116] PENA, V. H. de la (1992). Decoupling and Khintchine's inequalities for U-statistics. *Ann. Probab.* 20 (1877-1892).

[117] PESKIR, G. (1992). Best constants in Kahane-Khintchine inequalities in Orlicz spaces. *Math. Inst. Aarhus, Preprint Ser.* No. 10, (42 pp). *J. Multivariate Anal.* 45, 1993 (183-216).

[118] PESKIR, G. (1992). Maximal inequalities of Kahane-Khintchine's type in Orlicz spaces. *Math. Inst. Aarhus, Preprint Ser.* No. 33, (18 pp). *Math. Proc. Cambridge Philos. Soc.* 115, 1994 (175-190).

[119] PESKIR, G. (1993). Best constants in Kahane-Khintchine inequalities for complex Steinhaus variables. *Math. Inst. Aarhus, Preprint Ser.* No. 24, (10 pp). *Proc. Amer. Math. Soc.* 123, 1995 (3101-3111).

[120] PINELIS, I. F. (1994). Extremal probabilistic problems and Hotelling's T^2 test under a symmetry condition. *Ann. Statist.* 22 (357-368).

[121] PISIER, G. (1978). Les inegalites de Khintchine-Kahane d'apres C. Borell. *Sem. Geom. Espaces Banach, Ec. Polytech. Cent. Math. 1977-1978, Expose* 7 (1-14).

[122] PITMAN, J. (1981). A note on L_2 maximal inequalities. *Sem. Probab. XV, Univ. Strasbourg 1979/80. Lecture Notes in Math.* 850 (251-258).

[123] PITTENGER, A. O. (1979). Note on a square function inequality. *Ann. Probab.* 7 (907-308).

[124] PROKHOROV, Yu. V. (1959). An extremal problem in probability theory. *Theory Probab. Appl.* 4 (English transl.) (201-203).

[125] PURI, M. L. *and* SHEU, S. S. (1979). Realization of l_p by spaces of random variables. *Appl. Anal.* 8 (337-347).

[126] RAO, M. (1972). Doob decomposition and Burkholder inequalities. *Sem. Probab. VI, Univ. Strasbourg 1970/71. Lecture Notes in Math.* 258 (198-201).

[127] REVESZ, P. (1990). *Random Walk in Random and Non-Random Environments.* World Scientific Publ. Co.

[128] RODIN, V. A. *and* SEMYONOV, E. M. (1975). Rademacher series in symmetric spaces. *Anal. Math.* 1 (207-222).

[129] ROSENTHAL, H. P. (1970). On the subspaces of L^p $(p > 2)$ spanned by sequences of independent random variables. *Israel J. Math.* 8 (273-303).

[130] SAEKI, S. (1989). Some applications of Khintchin's inequalities. *Commutative Harmonic Analysis, Proceedings (Canton, NY, 1987). Contemp. Math.* 91 (247-271).

[131] SAWA, J. (1985). The best constant in the Khintchine inequality for complex Steinhaus variables, the case $p = 1$. *Studia Math.* 81 (107-126).

[132] SCHUR, I. (1923). Uber eine Klasse von Mittelbildungen mit Anwendungen die Determinanten. *Theory Sitzungsber. Berlin. Math. Gesellschaft* 22 (9-20).

[133] SEKIGUCHI, T. (1980). Weighted norm inequalities on the martingale theory. *Math. Rep. Toyama Univ.* 3 (37-100).

[134] SEMENOV, E. M. *and* SHNEJBERG, I. Y. (1988). Hypercontractive operators and Khinchin's inequality. *Functional Anal. Appl.* 22 (244-246).

[135] SEMENOV, E. M. *and* SHNEJBERG, I. Y. (1990). Contractive operators and Khinchin's inequality. *Siberian Math. J.* 31 (119-127).

[136] SHAO, Q. (1988). A moment inequality and its applications. *Acta Math. Sinica* 31 (736-747).

[137] SHEN, X. *and* ZHONG, L. (1990). Generalization of Marcinkiewicz-Zygmund inequalities. *Acta Sci. Natur. Univ. Pekinensis* 26 (257-265).

[138] SHEPP, L. A. (1967). A first passage problem for the Wiener process. *Ann. Math. Statist.* 38 (1912-1914).

[139] SHIRYAEV, A. N. (1984). *Probability.* Springer-Verlag New York.

[140] STECHKIN, S. B. (1961). On the best lacunary systems of functions. *Izv. Akad. Nauk SSSR, Ser. Mat.* 25 (Russian) (357-366).

[141] STEINHAUS, H. (1922). Les probabilites denombrables et leur rapport a la theorie de la mesure. *Fund. Math.* 4 (286-310).

[142] SZAREK, S. J. (1978). On the best constant in the Khintchine inequality. *Studia Math.* 58 (197-208).

[143] TAKAFUMI, M. *and* AKIHITO, U. (1985). Good lambda inequalities for the area integral and the nontangential maximal function. *Studia Math.* 83 (251-262).

[144] TOMASZEWSKI, B. (1982). Two remarks on the Khintchine-Kahane inequality. *Colloq. Math.* 46 (283-288).

[145] TUBARO, L. (1984). An estimate of Burkholder type for stochastic processes defined by

the stochastic integral. *Stochastic Anal. Appl.* 2 (187-192).

[146] ULLRICH, D. C. (1988). Khinchin's inequality and the zeroes of Bloch functions. *Duke Math. J.* 57 (519-535).

[147] ULLRICH, D. C. (1988). An extension of the Kahane-Khinchine inequality in a Banach space. *Israel J. Math.* 62 (56-62).

[148] ULLRICH, D. C. (1988). An extension of the Kahane-Khinchine inequality. *Bull. Amer. Math. Soc. (NS)* 18 (52-54).

[149] UTEV, S. A. (1984). Extremal problems in moment inequalities. *Predel'nye Theorii Veroyatnostej; Trudy Inst. Mat. (Novosibirsk)* 5 (Russian) (56-75).

[150] WALSH, J. L. (1923). A closed set of normal orthogonal functions. *Amer. J. Math.* 45 (5-24).

[151] WANG, G. (1991). Sharp maximal inequalities for conditionally symmetric martingales and Brownian motion. *Proc. Amer. Math. Soc.* 112 (579-586).

[152] WANG, G. (1991). Sharp square-function inequalities for conditionally symmetric martingales. *Trans. Amer. Math. Soc.* 328 (393-419).

[153] WANG, G. (1991). Sharp inequalities for the conditional square function of a martingale. *Ann. Probab.* 19 (1679-1688).

[154] WITTLE, P. (1960). Bounds for the moments of linear and quadratic forms in independent random variables. *Theory Probab. Appl.* 5 (302-305).

[155] XU, Y. (1991). The Marcinkiewicz-Zygmund inequality with derivatives. *Approx. Theory Appl.* 7 (100-107).

[156] XU, Y. (1991). The generalized Marcinkiewicz-Zygmund inequality for trigonometric polynomials. *J. Math. Anal. Appl.* 161 (447-456).

[157] YOUNG, R. M. G. (1976). On the best possible constants in the Khintchine inequality. *J. London Math. Soc.* 14 (496-504).

[158] ZHONG, L. SHEN, X. (1994). Weighted Marcinkiewicz-Zygmund inequalities. *Adv. in Math. (China)* 23 (66-75).

[159] ZINN, J. (1985). Comparison of martingale difference sequences. *Probability in Banach spaces V, Proc. 5th Int. Conf. (Medford, Massachusetts, 1984). Lecture Notes in Math.* 1153 (453-457).

[160] ZYGMUND, A. (1959). *Trigonometric Series*. Cambridge Univ. Press.

Goran Peskir
Department of Mathematical Sciences
University of Aarhus, Denmark
Ny Munkegade, DK-8000 Aarhus
home.imf.au.dk/goran
goran@imf.au.dk

Albert N. Shiryaev
Steklov Mathematical Institute
Gubkina str. 8
117966 Moscow
Russia
shiryaev@mi.ras.ru