The Annals of Probability
1994, Vol. 22, No. 4, 1679–1706

OPTIMUM BOUNDS FOR THE DISTRIBUTIONS OF MARTINGALES IN BANACH SPACES[1]

BY IOSIF PINELIS

Michigan Technological University

A general device is proposed, which provides for extension of exponential inequalities for sums of independent real-valued random variables to those for martingales in the 2-smooth Banach spaces. This is used to obtain optimum bounds of the Rosenthal–Burkholder and Chung types on moments of the martingales in 2-smooth Banach spaces. In turn, it leads to best-order bounds on moments of sums of independent random vectors in any separable Banach spaces. Although the emphasis is put on infinite-dimensional martingales, most of the results seem to be new even for one-dimensional martingales. Moreover, the bounds on moments of the Rosenthal–Burkholder type seem to be to a certain extent new even for sums of independent real-valued random variables. Analogous inequalities for (one-dimensional) supermartingales are given.

1. Introduction. For a separable Banach space $(\mathcal{X}, \|\cdot\|)$, let $\mathcal{S}(\mathcal{X})$ denote the class of all sequences $f = (f_j) = (f_0, f_1, \ldots)$ of Bochner-integrable random vectors in $\mathcal{X}$, with $f_0 \equiv 0$, defined on a probability space $(\Omega, F, \mathbf{P})$ and adapted to a nondecreasing sequence $(F_j) = (F_0, F_1, \ldots)$ of sub-σ-fields of F. Here, $(\Omega, F, \mathbf{P})$ and (F_j) are considered attributes of f and may be different for different $f \in \mathcal{S}(\mathcal{X})$.

For $f \in \mathcal{S}(\mathcal{X})$, put $f^* = \sup\{\|f_j\|: j = 0, 1, \ldots\}$, $d_0 = d_0(f) \equiv 0$, $d_j = d_j(f) = f_j - f_{j-1}$, $j = 1, 2, \ldots$, $S_p = S_p(f) = (\sum_{j=1}^{\infty} \|d_j\|^p)^{1/p}$, $p > 0$, $s_2 = s_2(f) = (\sum_{j=1}^{\infty} \mathbf{E}_{j-1}\|d_j\|^2)^{1/2}$, where $\mathbf{E}_{j-1}$ stands for the conditional expectation given F_{j-1}.

Let $\mathcal{M}(\mathcal{X})$ denote the class of all sequences $(f_j) \in \mathcal{S}(\mathcal{X})$ that are martingales and let $\mathcal{M}_{\text{ind}}(\mathcal{X})$ denote the class of all sequences $(f_j) \in S(\mathcal{X})$ having independent increments d_j.

For any two nonnegative expressions $\mathcal{E}_1$ and $\mathcal{E}_2$, let us write $\mathcal{E}_1 \preceq \mathcal{E}_2$ (or, equivalently, $\mathcal{E}_2 \succeq \mathcal{E}_1$) if $\mathcal{E}_1 \le A\mathcal{E}_2$, and $\mathcal{E}_1 \asymp \mathcal{E}_2$ if $\mathcal{E}_1 \preceq \mathcal{E}_2 \preceq \mathcal{E}_1$. Here, A denotes a positive absolute constant.

We assume that $\inf \emptyset = \infty$, $\sup \emptyset = 0$, $\sum_{j\in\emptyset} u_j = 0$ and $\prod_{j\in\emptyset} u_j = 1$.

In Section 2, some preliminary results on 2-smooth Banach spaces and on martingales in such spaces are given.

In Section 3, a device is suggested, which provides for the extension of exponential inequalities for sums of independent real-valued random variables to

Received November 1992.
[1]Research partially supported by NSF Grant DMS-89-21369.
AMS 1991 *subject classifications*. 60E15, 60B12, 60G42, 60G50, 60F10.
Key words and phrases. Distribution inequalities, exponential inequalities, bounds on moments, martingales in Banach spaces, 2-smooth Banach spaces, sums of independent random variables.

those for martingales in 2-smooth Banach spaces. In particular, by that means an exponential inequality for martingales in 2-smooth spaces, optimal in terms of $\|d^*\|_\infty$ and $\|s_2\|_\infty$, is obtained, which is a generalization of an inequality of Bennett (1962) and Hoeffding (1963).

In Section 4, using methods of Burkholder (1973) and results of Section 3, we obtain optimal (to the above-defined relation $\asymp$) upper bounds of the Rosenthal (1970) and Burkholder (1973) type on moments of martingales in 2-smooth Banach spaces, that is, optimal in terms of $\|d^*\|_p$ and $\|s_2\|_p$, for $p \geq 2$.

In Section 5, via the martingale decomposition method of Yurinskii (1974), we apply the inequalities of Section 4 to obtain bounds of the Rosenthal–Burkholder type on $\|\|f_n\| - \mathbf{E}\|f_n\|\|_p$, $p \geq 2$, for an arbitrary separable Banach space $\mathcal{X}$, but only for $f \in \mathcal{M}_{\text{ind}}(\mathcal{X})$.

In Section 6, we show that the inequalities derived in Sections 4 and 5 are optimal in the terms used.

In Section 7, we obtain bounds on $\|\|f_n\|\|_p, p \geq 2, f \in \mathcal{M}(\mathcal{X})$, which are optimal in terms of n and $\|S_p\|_p$. We refer to them as bounds of the Chung type.

In Section 8, inequalities for supermartingales (of course, in $\mathcal{X} = \mathbf{R}$) similar to those in Sections 3 and 4 and certain refinements for real-valued martingales are presented.

A substantial part of the results was announced in Pinelis (1992).

2. Preliminaries: 2-smooth Banach spaces and a reduction of martingales. Let us call a Banach space $(\mathcal{X}, \|\cdot\|)(2, D)$-smooth, where $D > 0$, if for all $x \in \mathcal{X}$ and $y \in \mathcal{X}$,

$$\|x+y\|^2 + \|x-y\|^2 \leq 2\|x\|^2 + 2D^2\|y\|^2, \tag{2.1}$$

and 2-smooth if it is $(2, D)$-smooth for some $D > 0$. For any 2-smooth space $(\mathcal{X}, \|\cdot\|)$, let $D(\mathcal{X})$ denote the smallest $D > 0$ such that $(\mathcal{X}, \|\cdot\|)$ is $(2, D)$-smooth. Putting $x = 0$ in (2.1), one observes that actually $D(\mathcal{X}) \geq 1$ as long as $\mathcal{X} \neq \{0\}$, which will be assumed to be the case.

The importance of 2-smooth spaces was elucidated in the paper by Pisier (1975): they play the same role with respect to vector martingales as spaces of type 2 do with respect to the sums of independent random vectors.

The definition assumed in this paper is slightly different from that given by Pisier [which required only that (2.1) hold for an equivalent norm], because, in the subsequent account, we would like to follow the dependence of certain constants on D, the constant of the 2-smoothness.

It is easily seen that the condition

$$\left(\|x\|^2\right)''(v, v) \leq 2D^2\|v\|^2 \qquad \forall\, x \in \mathcal{X}\ \forall\, v \in \mathcal{X}, \tag{2.2}$$

is sufficient for the $(2, D)$-smoothness, where $(\|x\|^2)''(v, v)$ stands for the second directional derivative of the function $x \to \|x\|^2$ in the direction v.

By way of illustration, we give the following proposition.

PROPOSITION 2.1. *For any $p \geq 2$ and any measure space $(T, \mathcal{A}, \nu)$, $L^p := L^p(T, \mathcal{A}, \nu)$ is $(2, \sqrt{p-1})$-smooth.*

PROOF. For $\|x\| := (\int_T |x|^p d\nu)^{1/p}$, one has

$$
\begin{aligned}
\tfrac{1}{2}\left(\|x\|^2\right)''(v,v) &= (p-1)\|x\|^{2-p}\int_T |x|^{p-2}v^2\,d\nu \\
&\quad -(p-2)\|x\|^{2-2p}\left(\int_T |x|^{p-2}xv\,d\nu\right)^2 \qquad (2.3)\\
&\leq (p-1)\|v\|^2 - 0
\end{aligned}
$$

if $x \in \mathcal{X}\setminus\{0\}$, $v \in \mathcal{X}$ (in view of Hölder's inequality) and $\frac{1}{2}(\|x\|^2)''(v,v) = \|v\|^2$ if $x = 0$. Thus, (2.2) is checked. □

In particular, it is obvious and well known that if $\mathcal{X}$ is a Hilbert space, then it is $(2,1)$-smooth.

REMARK. If $L^p = L^p(T,\mathcal{A},\nu)$ is at least two-dimensional, that is, if there exist $T_1 \in \mathcal{A}$ and $T_2 \in \mathcal{A}$ such that $T_1 \cap T_2 = \emptyset$, $0 < \nu(T_1) < \infty$, $0 < \nu(T_2) < \infty$, then L^p is *not* $(2,D)$-smooth if $0 < D < \sqrt{p-1}$, so that Proposition 2.1 gives the best bound. Indeed, put $x \equiv \nu(T_2)^{1/p}$ on T_1, $x \equiv \nu(T_1)^{1/p}$ on T_2, $v \equiv x$ on T_1, $v \equiv -x$ on T_2 and $x \equiv v \equiv 0$ on $(T\setminus T_1)\setminus T_2$. Then $x \neq 0$ and (2.3) turns into an equality.

Condition (2.2) is not only sufficient but also necessary for $(2,D)$-smoothness if the derivatives are understood in a generalized sense. To state this remark rigorously, let us give a definition somewhat extending the notion of 2-smoothness.

For any Banach space $(\mathcal{X}, \|\cdot\|)$, we call a function $\Psi: \mathcal{X} \to [0,\infty)$ $(2,D)$-smooth, $D > 0$, if it satisfies the conditions:

$$
\begin{aligned}
\Psi(0) &= 0,\\
|\Psi(x+v) - \Psi(x)| &\leq \|v\|,\\
\Psi^2(x+v) - 2\Psi^2(x) + \Psi^2(x-v) &\leq 2D^2\|v\|^2
\end{aligned}
$$

for all $x \in \mathcal{X}$, $v \in \mathcal{X}$.

Evidently, a Banach space $(\mathcal{X}, \|\cdot\|)$ is $(2,D)$-smooth if and only if its norm function is $(2,D)$-smooth.

The results stated in the subsequent sections for norms of martingales in 2-smooth spaces can be extended to those for $(2,D)$-smooth functions of martingales in any Banach spaces.

For any $(2,D)$-smooth function Ψ on a *finite-dimensional* Banach space $(X, \|\cdot\|)$ and $\varepsilon > 0$, define

$$\Psi_\varepsilon(x) = \sqrt{\int_{\mathcal{X}} \Psi^2(x - \varepsilon y)\gamma(dy)},$$

where γ is, say, a zero-mean Gaussian measure on $\mathcal{X}$ with support$(\gamma) = \mathcal{X}$.

LEMMA 2.2. *If Ψ is a $(2,D)$-smooth function on a finite-dimensional Banach space $(\mathcal{X}, \|\cdot\|)$, then for all $\varepsilon > 0$, Ψ_ε has Fréchet derivatives $\Psi'_\varepsilon(x), \Psi''_\varepsilon(x), \dots$ of any order, and the directional derivatives in any direction $v \in \mathcal{X}$ satisfy the inequalities*

$$|\Psi'_\varepsilon(x)(v)| \le \|v\|, \qquad \left(\Psi^2_\varepsilon\right)''(x)(v,v) \le 2D^2\|v\|^2 \tag{2.4}$$

for all $x \in \mathcal{X}$. Moreover, for each $x \in \mathcal{X}$, $\Psi_\varepsilon(x) \to \Psi(x)$ as $\varepsilon \downarrow 0$. [*In this generalized sense, sufficient condition* (2.2) *is also necessary for a Banach space* $(\mathcal{X}, \|\cdot\|)$ *to be* $(2,D)$*-smooth. Note that* (2.1) *may be considered locally—for any two-dimensional subspace containing x and y.*]

PROOF. Among the statements of the lemma, only the first of the inequalities (2.4) is comparatively nontrivial. Observe that

$$\begin{aligned} \left|\left(\Psi^2_\varepsilon\right)'(x)(v)\right| &\le \limsup_{t\to 0} \frac{1}{t} \int_X |\Psi(x+tv-\varepsilon y) - \Psi(x-\varepsilon y)| \\ &\qquad\qquad \times \left(\Psi(x+tv-\varepsilon y) + \Psi(x-\varepsilon y)\right)\gamma(dy) \\ &\le \|v\| \int_X 2\Psi(x-\varepsilon y)\gamma(dy). \end{aligned}$$

One can assume that $\Psi_\varepsilon(x) \ne 0$ for all $x \in \mathcal{X}$. Now,

$$\begin{aligned} |(\Psi_\varepsilon)'(x)(v)| &= \left|\frac{\left(\Psi^2_\varepsilon\right)'(x)(v)}{2\Psi_\varepsilon(x)}\right| \\ &\le \frac{\|v\| \int_X 2\Psi(x-\varepsilon y)\gamma(dy)}{2\sqrt{\int_X \Psi^2(x-\varepsilon y)\gamma(dy)}} \\ &\le \|v\| \end{aligned}$$

by the Schwarz inequality. □

REMARK. Instead of using the convolution approximation, one might as well understand the directional derivatives as, say, the right-hand upper derivatives, and then again, (2.2) would be necessary for (2.1).

We also need the following folklorish lemma.

LEMMA 2.3. *Let $(f_j)_{j=0}^\infty \in \mathcal{M}(\mathcal{X})$ be a martingale in a separable Banach space $(\mathcal{X}, \|\cdot\|)$ relative to a filtration $(F_j)_{j=0}^\infty$. Then for any $\varepsilon > 0$, there exists a martingale $(f_{j,\varepsilon})_{j=0}^\infty \in \mathcal{M}(\mathcal{X})$ relative to a filtration $(F_{j,\varepsilon})_{j=0}^\infty$ such that $\forall\, j = 1, 2, \dots$:*

(2.5) *$f_{j,\varepsilon}$ is a random variable having only a finite number of values,*

$$f_{j,\varepsilon} \to f_j \text{ in probability as } \varepsilon \downarrow 0, \tag{2.6}$$

$$\mathbf{E}g(f_{j,\varepsilon}) \le \mathbf{E}g(f_j), \tag{2.7}$$

$$\left\|\mathbf{E}\left(g(f_{j,\varepsilon}) \mid F_{j-1,\varepsilon}\right)\right\|_\infty \le \left\|\mathbf{E}\left(g(f_j) \mid F_{j-1}\right)\right\|_\infty, \tag{2.8}$$

where g is any nonnegative convex real function on $\mathcal{X}$.

PROOF. Consider the approximation $f_{j,\varepsilon} := \mathbf{E}(f_j \mid F_{j,\varepsilon})$, where $F_{j,\varepsilon}$ is the σ-field generated by all the events of the form $\{f_i \in B_{k,\varepsilon}\}$, $i = 0, 1, \dots, j$, $k = 1, \dots, k(j,\varepsilon)$, where $(\{B_{k,\varepsilon}: k = 1, \dots, k(j,\varepsilon)\})_{j=0}^{\infty}$ is an increasing sequence of sets of balls in $\mathcal{X}$ of radius ε such that $k(0,\varepsilon) < k(1,\varepsilon) < k(2,\varepsilon) < \cdots$ and $\mathbf{P}(\bigcap_{i=1}^{j} \bigcup_{k=1}^{k(j,\varepsilon)} \{f_i \in B_{k,\varepsilon}\}) \geq 1 - \varepsilon$, $j = 1, 2, \dots$ (the existence of such a sequence of sets is guaranteed by the tightness of any probability measure on a separable Banach space).

Then (2.5) and (2.6) are satisfied. The Jensen type inequality

$$g(f_{j,\varepsilon}) \leq \mathbf{E}(g(f_j) \mid F_{j,\varepsilon})$$

implies (2.7) and

$$\|\mathbf{E}(g(f_{j,\varepsilon}) \mid F_{j-1,\varepsilon})\|_\infty \leq \|\mathbf{E}(g(f_j) \mid F_{j-1,\varepsilon})\|_\infty \leq \|\mathbf{E}(g(f_j) \mid F_{j-1})\|_\infty,$$

so that (2.8) is also true. □

REMARK 2.4. Taking into account Lemma 2.3, the standard construction $(f_{j \wedge n})_{j=0}^{\infty}$ with large n and Lemma 2.2, we may and shall, without loss of generality, restrict the proofs of all subsequent results for martingales in 2-smooth spaces only to the stopped martingales $(0, f_1, \dots, f_n, f_n, f_n, \dots)$ with each of the f_j's having only a finite number of values in a finite-dimensional Banach space $(\mathcal{X}, \|\cdot\|)$ satisfying condition (2.2). For such martingales, we put $f_\infty = f_n$.

PROPOSITION 2.5. *If $f \in \mathcal{M}(\mathcal{X})$ and $\mathcal{X}$ is a $(2,D)$-smooth separable Banach space, then*

$$\|f_n\|_2 \leq D\|s_2\|_2 = D\|S_2\|_2.$$

PROOF. For $j = 1, 2, \dots$, put

$$g(t) = \mathbf{E}\|f_{j-1} + td_j\|^2.$$

In view of Remark 2.4, it is possible to reverse the order of the integration and differentiation to obtain $g'(0) = 0$ and $g''(t) \leq 2D^2\|d_j\|_2^2$. Hence,

$$\begin{aligned}\mathbf{E}\|f_j\|^2 - \mathbf{E}\|f_{j-1}\|^2 &= g(1) - g(0) = \int_0^1 g''(t)(1-t)\,dt \\ &\leq D^2\|d_j\|_2^2, \qquad j = 1, 2, \dots.\end{aligned}$$

It remains to sum these inequalities. □

The upper bounds provided by Theorems 15.1 and 21.1 of Burkholder (1973) can be immediately extended to martingales in 2-smooth separable Banach spaces. Let us state this:

THEOREM 2.6. *If $f \in \mathcal{M}(\mathcal{X})$, $\mathcal{X}$ is a 2-smooth separable Banach space, the function g: $[0,\infty) \to [0,\infty)$ is nondecreasing, $g(0) = 0$, and $g(2u) \le c_g g(u), u \ge 0$, for some c_g, then*

$$\mathbf{E}g(f^*) \le c_{2.9}\big[\mathbf{E}g(Ds_2) + \mathbf{E}g(d^*)\big]. \tag{2.9}$$

If, moreover, g is convex, then

$$\mathbf{E}g(f^*) \le c_{2.10}\mathbf{E}g(DS_2). \tag{2.10}$$

Here $c_{2.9}$ and $c_{2.10}$ depend only on c_g.

PROOF. The proof repeats that in Burkholder (1973) with the following exceptions: (a) use $\|\cdot\|$ instead of $|\cdot|$ and (b) use Proposition 2.5 instead of the identities (in the notation therein): $\|h\|_2 = \|S(h)\|_2$ and $\|h\|_2 = \|s(h)\|_2$. □

It is well known [see (11.1) in Burkholder (1991)] that for $g(t) = |t|^p, (Ap)^p$ is optimum for $c_{2.10}$ at least if $\mathcal{X}$ is a Hilbert space. As to optimum bounds like (2.10) for conditionally symmetric martingales in 2-smooth spaces, see Section 4 and, in particular, Remark 4.4 therein, concerning conditionally symmetric martingales in Hilbert spaces.

3. Exponential bounds on tail probabilities for the martingales in 2-smooth spaces.

THEOREM 3.1. *Suppose that $f \in \mathcal{M}(\mathcal{X})$, $\mathcal{X}$ is a $(2,D)$-smooth separable Banach space, and $\lambda > 0$ is such that $\mathbf{E}e^{\lambda\|d_j\|} < \infty$ for $j = 1,2,\dots$. Then for all $r \ge 0$,*

$$\begin{aligned}\mathbf{P}(f^* \ge r) &\le 2\exp(-\lambda r)\left\|\prod_{j=1}^{\infty}\Big[1 + D^2\mathbf{E}_{j-1}\Big(\exp\big(\lambda\|d_j\|\big) - 1 - \lambda\|d_j\|\Big)\Big]\right\|_{\infty}\\ &\le 2\exp\left\{-\lambda r + D^2\left\|\sum_{j=1}^{\infty}\mathbf{E}_{j-1}\Big(\exp(\lambda\|d_j\|) - 1 - \lambda\|d_j\|\Big)\right\|_{\infty}\right\}.\end{aligned}$$

We shall obtain this theorem as a particular case of the following result for $(2,D)$-smooth functions defined in Section 2.

THEOREM 3.2. *Suppose that $f \in \mathcal{M}(\mathcal{X})$, $\mathcal{X}$ is any separable Banach space, $\lambda > 0$ is such that $\mathbf{E}e^{\lambda\|d_j\|} < \infty$ for $j = 1,2,\dots$, and the function Ψ is $(2,D)$-smooth. Then for all $r \ge 0$,*

$$\mathbf{P}\Big(\sup_j \Psi(f_j) \ge r\Big) \le 2\exp(-\lambda r)\left\|\prod_{j=1}^{\infty}(1 + e_j)\right\|_{\infty} \tag{3.1}$$

$$\le 2\exp\left\{-\lambda r + \left\|\sum_{j=1}^{\infty} e_j\right\|_{\infty}\right\}, \tag{3.2}$$

where

$$e_j = D_*^2 \mathbf{E}_{j-1}\big(e^{\lambda\|d_j\|} - 1 - \lambda\|d_j\|\big), \qquad D_* = 1 \vee D.$$

PROOF. Put $u(t) = u_{x,v}(t) = \Psi(x + tv)$ for any x, v in $\mathcal{X}$. By Lemma 2.2 and Remark 2.4, we may assume that u is differentiable, $|u'(t)| \le \|v\|$ and $(u^2)''(t) \le 2D^2\|v\|^2$. Hence, when $u''u > 0$, one has $(\cosh u)'' = u'^2 \cosh u + u'' \sinh u \le (u'^2 + u''u)\cosh u = \frac{1}{2}(u^2)'' \cosh u \le D^2\|v\|^2 \cosh u$ and, otherwise, $(\cosh u)'' \le u'^2 \cosh u \le \|v\|^2 \cosh u$. In any case,

$$(\cosh u)'' \le D_*^2 \|v\|^2 \cosh u. \tag{3.3}$$

Consider now $\varphi(t) := \mathbf{E}_{j-1}\cosh(\lambda\Psi(f_{j-1} + td_j))$, $|t| \le 1$. In view of (3.3) and Remark 2.4,

$$\begin{aligned}\varphi''(t) &\le D_*^2\lambda^2\mathbf{E}_{j-1}\|d_j\|^2\cosh\big(\lambda\Psi(f_{j-1} + td_j)\big)\\ &\le D_*^2\lambda^2\mathbf{E}_{j-1}\|d_j\|^2 e^{\lambda t\|d_j\|}\cosh\big(\lambda\Psi(f_{j-1})\big), \quad |t| < 1\end{aligned}$$

However, $\varphi'(0) = 0$ because (f_j) is a martingale and, therefore,

$$\mathbf{E}_{j-1}\cosh\big(\lambda\Psi(f_j)\big) = \varphi(1) = \varphi(0) + \int_0^1 (1-t)\varphi''(t)\,dt \le (1 + e_j)\cosh\big(\lambda\Psi(f_{j-1})\big).$$

Thus, putting $G_0 = 1$ and $G_j = \cosh(\lambda\Psi(f_j))/\prod_{i=1}^{j}(1+e_i)$, $j = 1, 2, \dots$, one has a positive supermartingale. Hence, if $\tau := \inf\{j\colon \Psi(f_j) \ge r\}$, then $\mathbf{E}G_\tau \le \mathbf{E}G_0 = 1$ and so

$$\mathbf{P}\Big(\sup_j \Psi(f_j) \ge r\Big) \le \mathbf{P}\bigg(G_\tau \ge \cosh(\lambda r)\Big/\Big\|\prod_{j=1}^{\infty}(1 + e_j)\Big\|_\infty\bigg).$$

Now (3.1) follows from Chebyshev's inequality and $\cosh u > e^u/2$; (3.2) is elementary. □

REMARK. For $f \in \mathcal{M}_{\text{ind}}(\mathbf{R})$, that is, for sums of independent zero-mean real-valued random variables, the following is used as a starting point when proving exponential inequalities:

$$\begin{aligned}\mathbf{P}(f_n \ge r) &\le \exp(-\lambda r)\prod_{j=1}^{n}\Big[1 + \mathbf{E}\big(\exp(\lambda d_j) - 1 - \lambda d_j\big)\Big]\\ &\le \exp\bigg\{-\lambda r + \sum_{j=1}^{n}\mathbf{E}\big(\exp(\lambda d_j) - 1 - \lambda d_j\big)\bigg\}.\end{aligned}$$

Thus, Theorem 3.1 provides a similar starting point for $f \in \mathcal{M}(\mathcal{X})$, $\mathcal{X}$ being 2-smooth. [In this sense, it is analogous to the results of Pinelis and Sakhanenko (1985) for sums of independent random vectors.] A general method of obtaining exact exponential inequalities for sums of independent real-valued random variables is proposed in Pinelis and Utev (1989). So, for martingales in 2-smooth spaces, these two devices taken together produce analogues of exact "independent real-valued" exponential bounds. For instance, this remark easily leads to the following analogues of classical results of Bernstein [see, e.g., Bennett (1962)] and Bennet (1962) and Hoeffding (1963) [cf. Theorems 9 and 3, respectively, in Pinelis and Utev (1989)].

THEOREM 3.3. *Suppose that $f \in \mathcal{M}(\mathcal{X})$, $\mathcal{X}$ is a $(2, D)$-smooth separable Banach space and*

$$\left\| \sum_{j=1}^{\infty} \mathbf{E}_{j-1} \|d_j\|^m \right\|_{\infty} \le m! \Gamma^{m-2} B^2 / (2D^2)$$

for some $\Gamma > 0$, $B > 0$ and $m = 2, 3, \ldots$. Then for all $r \ge 0$,

$$\mathbf{P}(f^* \ge r) \le 2 \exp\left(- \frac{r^2}{B^2 + B\sqrt{B^2 + 2\Gamma r}} \right).$$

PROOF. Under the conditions given,

$$D^2 \left\| \sum_{j=1}^{\infty} \mathbf{E}_{j-1} \left(e^{\lambda \|d_j\|} - 1 - \lambda \|d_j\| \right) \right\|_{\infty}$$

$$\le \frac{1}{2} \sum_{m=2}^{\infty} \lambda^m \Gamma^{m-2} B^2 = \frac{B^2 \lambda^2}{2(1 - \lambda\Gamma)}, \qquad 0 \le \lambda < \frac{1}{\Gamma}.$$

Now Theorem 3.1 yields

$$\mathbf{P}(f^* \ge r) \le 2 \exp\left\{ -\lambda r + \frac{B^2 \lambda^2}{2(1 - \lambda \Gamma)} \right\}.$$

It remains to minimize the r.h.s. in λ. □

THEOREM 3.4. *Suppose that $f \in \mathcal{M}(\mathcal{X})$, $\mathcal{X}$ is a $(2, D)$-smooth separable Banach space, and $\|d^*\|_\infty \le a$, $\|s_2\|_\infty \le b/D$ for some $a > 0$, $b > 0$. Then for all $r \ge 0$,*

$$\mathbf{P}(f^* \ge r) \le 2 \exp\left[\frac{r}{a} - \left(\frac{r}{a} + \frac{b^2}{a^2} \right) \ln\left(1 + \frac{ra}{b^2} \right) \right] \tag{3.4}$$

$$\le 2 \left(\frac{eb^2}{ra} \right)^{r/a}. \tag{3.5}$$

PROOF. Because the function $g(u) := u^{-2}(e^u - 1 - u)$ for $u \ne 0$, $g(0) := \frac{1}{2}$ is increasing in $u \in \mathbf{R}$,

$$\mathbf{E}_{j-1} \left(e^{\lambda \|d_j\|} - 1 - \lambda \|d_j\| \right) \le \frac{e^{\lambda a} - 1 - \lambda a}{a^2} \mathbf{E}_{j-1} \|d_j\|^2.$$

Now Theorem 3.1 yields

$$\mathbf{P}(f^* \geq r) \leq 2\exp\left\{-\lambda r + \frac{\exp(\lambda a) - 1 - \lambda a}{a^2} b^2\right\}$$

and the minimization in λ gives (3.4). Inequality (3.5) is trivial. □

In the special case $\mathcal{X} = L^2$, a bound similar to (3.4), but somewhat weaker, was proved by Kallenberg and Sztencel (1991); their method seems to be confined only to Hilbert spaces.

Theorem 3.4 was proved in Pinelis (1992) for $\mathcal{X} = L^p$, $p \geq 2$. A version for general 2-smooth spaces was given therein too, but with another, greater constant in place of D.

THEOREM 3.5. *Suppose that* $f \in \mathcal{M}(\mathcal{X})$, $\mathcal{X}$ *is a* $(2, D)$*-smooth separable Banach space, and* $\sum_{j=1}^{\infty} \|d_j\|_{\infty}^2 \leq b_*^2$ *for some* $b_* > 0$. *Then for all* $r \geq 0$,

$$\mathbf{P}(f^* \geq r) \leq 2\exp\left\{-\frac{r^2}{2D^2 b_*^2}\right\}.$$

PROOF. The proof is the same as that of Theorem 3 in Pinelis (1992a) except that, in view of (3.3), one can use D^2 instead of B therein. □

Theorem 3.5 can be improved in the special case of conditionally symmetric martingales.

THEOREM 3.6. *Suppose that* $\mathcal{X}$ *is a* $(2, D)$*-smooth separable Banach space,* $f \in \mathcal{M}(\mathcal{X})$, $\|S_2(f)\|_{\infty} \leq b$ *for some* $b > 0$, *and the increments* d_j *are* F_{j-1}*-conditionally symmetrically distributed,* $j = 1, 2, \ldots$. *Then for all* $r \geq 0$,

$$\mathbf{P}(f^* \geq r) \leq 2\exp\left\{-\frac{r^2}{2D^2 b^2}\right\}.$$

PROOF. Being conditionally symmetric, (f_j) is also a martingale relative to the sequence (G_j), where G_j is the σ-field generated by F_j and $\|d_{j+1}\|$; see, for example, Lemma 10.2 in Burkholder (1991). Now, the proof can be concluded as that of Theorem 3 in Pinelis (1992): only the conditional expectations given G_j's are taken instead of those given F_j's, and D^2, $\|d_j\|^2$ are used in place of B, b_n^2 therein, respectively. □

In the case when $\mathcal{X} = \mathbf{R}$ and d_j's are simple functions, Theorem 3.6 was given in Hitczenko (1990b).

An analogous result for sums of independent random vectors in arbitrary separable Banach spaces is given in Pinelis (1990).

4. A spectrum of Rosenthal–Burkholder type bounds on moments of martingales in 2-smooth spaces.

THEOREM 4.1. *If* $f \in \mathcal{M}(\mathcal{X})$, $\mathcal{X}$ *is a* $(2, D)$*-smooth separable Banach space,* $p \geq 2$, $1 \leq c \leq p$, *then*

$$\|f^*\|_p \preceq c\|d^*\|_p + \sqrt{c}e^{p/c}D\|s_2\|_p. \tag{4.1}$$

In particular,

$$\|f^*\|_p \preceq p\|d^*\|_p + \sqrt{p}D\|s_2\|_p, \tag{4.2}$$

$$\|f^*\|_p \preceq \frac{p}{\ln p}\big(\|d^*\|_p + D\|s_2\|_p\big), \tag{4.3}$$

$$\|f^*\|_p \preceq \alpha\|d^*\|_p + e^{p/\alpha}D\|s_2\|_p, \qquad 1 \leq \alpha \leq \frac{p}{\ln(ep)}. \tag{4.4}$$

We need the following lemma.

LEMMA 4.2. *Suppose that* $f \in \mathcal{M}(\mathcal{X})$, *the increments* d_j *are* F_{j-1}*-conditionally symmetrically distributed,* $j = 1, 2, \ldots$, $\lambda > 0$, $\delta_1 > 0$, $\delta_2 > 0$, $\beta - 1 - \delta_2 > 0$. *Then*

$$\mathbf{P}(f^* > \beta\lambda,\ w^* \leq \lambda) \leq \varepsilon\mathbf{P}(f^* > \lambda),$$

where

$$w^* = \left(\frac{d^*}{\delta_2}\right) \vee \left(\frac{Ds_2}{\delta_1}\right),$$

$$\varepsilon = 2\left(\frac{e}{N}\frac{\delta_1^2}{\delta_2^2}\right)^N, \qquad N = \frac{\beta - 1 - \delta_2}{\delta_2}. \tag{4.5}$$

PROOF. The proof is based on the method of Burkholder (1973). Put $\overline{d}_j = d_j I\{\|d_j\| \leq \delta_2\lambda\}$, $\overline{f}_j = \sum_{i=0}^{j}\overline{d}_i$ and $h_j = \overline{f}_{(j\wedge\tau\wedge\nu)\vee\mu} - \overline{f}_\mu$, where $\mu = \inf\{j\colon \|\overline{f}_j\| > \lambda\}$, $\nu = \inf\{j\colon \|\overline{f}_j\| > \beta\lambda\}$, $\tau = \inf\{j\colon \overline{s}_{j+1} > \lambda\}$, and $\overline{s}_{j+1} = \sqrt{\sum_{i=1}^{j+1}\mathbf{E}_{i-1}\|\overline{d}_i\|^2}$, $j = 0, 1, \ldots$. Then $\mathbf{E}_{j-1}\overline{d}_j = 0$, $h_j - h_{j-1} = \overline{d}_j I\{\mu < j \leq \tau \wedge \nu\}$, $j = 1, 2, \ldots$, and so, (h_j) is a martingale in $\mathcal{X}$ conditionally on F_μ (the σ-field consisting of all $\Omega_0 \in F$ such that $\Omega_0 \cap \{\mu = j\} \in F_j$ for all j). In addition,

$$\begin{aligned}\mathbf{P}(f^* > \beta\lambda,\ w^* \leq \lambda) &= \mathbf{P}(\overline{f}^* > \beta\lambda,\ w^* \leq \lambda) \leq \mathbf{P}\big(h^* > (\beta - 1 - \delta_2)\lambda\big)\\ &= \mathbf{E}\mathbf{P}\big(h^* > (\beta - 1 - \delta_2)\lambda \,|\, F_\mu\big) I\{\mu < \infty\}\\ &\leq \varepsilon\mathbf{P}(\mu < \infty) = \varepsilon\mathbf{P}(f^* > \lambda).\end{aligned}$$

Here we have applied (3.5) with $r = (\beta - 1 - \delta_2)\lambda$, $a = \delta_2\lambda$ and $b = \delta_1\lambda$. □

PROOF OF THEOREM 4.1. The argument in Hitczenko (1990a) shows that we need to consider only the following two cases: (1) the increments d_j are

F_{j-1}-conditionally symmetrically distributed and (2) $f \in \mathcal{M}_{\mathrm{ind}}(\mathcal{X})$. However, via the standard symmetrization formula $\widetilde{X} = X - X'$, where X, X' are independent copies, one can easily reduce case 2 to case 1. Thus, one can use Lemma 4.2.

Applying now Lemma 7.1 in Burkholder (1973) with $\Phi(\lambda) = \lambda^p$, $\gamma = \beta^p$ and $\delta = \eta = 1$, one has

$$\|f^*\|_p \le 2\beta\left(\frac{\|d^*\|_p}{\delta_2} + \frac{D\|s_2\|_p}{\delta_1}\right) \quad \text{if } \beta^p\varepsilon \le \frac{1}{2}, \tag{4.6}$$

where ε is given by (4.5). Choose now, for any $c \in [1, p]$,

$$\beta = 1 + e^{-p/c} + \frac{1}{c}, \qquad \delta = \frac{1}{10c}, \qquad \delta_1 = \frac{1}{10\sqrt{c}e^{p/c}}.$$

Then $\beta < 3$,

$$\begin{aligned} N &= \frac{\beta - 1 - \delta_2}{\delta_2} = 9 + 10ce^{-p/c} > 9, \\ \beta^p &= \left[1 + (N+1)\delta_2\right]^p < e^{2pN\delta_2}, \\ (\beta^p\varepsilon)^{1/N} &< 2^{1/9}e^{2p\,\delta_2}\frac{e}{10ce^{-p/c}}\frac{c}{e^{2p/c}} < \frac{1}{2}, \end{aligned}$$

so that (4.6) implies

$$\|f^*\|_p \le 60\left(c\|d^*\|_p + \sqrt{c}e^{p/c}D\|s_2\|_p\right).$$

Thus, (4.1) is proved.

Let c_p stand for the unique solution to the equation $\sqrt{c_p} = e^{p/c_p}$. Then, $c_p \sim 2p/\ln p$ as $p \to \infty$. Hence, putting $c = p$ and then $c = c_p$ in (4.1), one comes to (4.2) and then to (4.3), respectively.

The function $g(c) := \sqrt{c}e^{p/c}$ decreases on $[1,p]$, $g(1) = e^p$ and $g(p) = e\sqrt{p} < ep$. Hence, for each $\alpha \in [1, p/\ln(ep)]$, there exists $z_\alpha \in [1,p]$, the unique solution to the equation $g(z_\alpha) = e^{p/\alpha}$. In addition, $g(2\alpha) = e^{p/\alpha}\sqrt{2\alpha}e^{-p/(2\alpha)} \le e^{p/\alpha}\sqrt{2\alpha_p}e^{-p/(2\alpha_p)} < g(z_\alpha)$, where $\alpha_p := p/\ln(ep)$. Thus, $z_\alpha < 2\alpha$. Now we see that (4.1) with $c = z_\alpha$ yields (4.4). □

For $\mathcal{X} = \mathbf{R}$, inequalities (4.3) and (4.2) were proved in Hitczenko (1990a) and (1991), respectively.

A spectrum of bounds on moments of martingales in Hilbert spaces with bounded second conditional moments was found in Pinelis (1980). It is essentially equivalent to (4.1) at least in the case of independent increments d_j (see Remark 6.8), but has a much more cumbersome expression.

The infimum in c of the r.h.s. of (4.1), evaluated in Section 6, turns out to be an upper bound on $\|f^*\|_p$ optimal in terms of $\|d^*\|_p$, $\|s_2\|_p$, the optimum choice of c depending, obviously, on $\|d^*\|_p/\|s_2\|_p$. In addition, for each $c \in [1, p]$, the "individual" bound $c\|d^*\|_p + \sqrt{c}e^{p/c}\|s_2\|_p$ is optimal for a certain corresponding value of $\|d^*\|_p/\|s_2\|_p$. In particular, all the bounds in (4.2)–(4.4) are optimal. The issue of optimality is treated rigorously in Section 6.

THEOREM 4.3. *If $\mathcal{X}$ is a $(2,D)$-smooth separable Banach space, $f \in \mathcal{M}(\mathcal{X})$ and the increments d_j are F_{j-1}-conditionally symmetrically distributed, then*

$$\|f^*\|_p \preceq \sqrt{p} D \|S_2\|_p, \qquad p \geq 1.$$

PROOF. Consider f as a martingale relative to the sequence of σ-fields (G_j) defined in the proof of Theorem 3.6. Then $s_2 = S_2$. Reasoning as in the proof of Lemma 4.2 but using Theorem 3.6 instead of Theorem 3.4, we see that

$$\mathbf{P}(f^* > \beta\lambda,\ DS_2 \leq \delta\lambda) \leq \varepsilon \mathbf{P}(f^* > \lambda), \tag{4.7}$$

$\lambda > 0$, $\delta > 0$, and $\beta - 1 - \delta > 0$, where $\varepsilon = \exp[-\delta^{-2}(\beta - 1 - \delta)^2/2]$. It remains to choose, say, $\beta = 2$, $\delta = (0.1)p^{-1/2}$ and apply Lemma 7.1 of Burkholder (1973). □

REMARK 4.4. For conditionally symmetric martingales in Hilbert spaces, the exact constant A_p in the inequality

$$\|f_n\|_p \leq A_p \|S_2\|_p, \qquad p \geq 3, \tag{4.8}$$

extending real-case results of Davis (1976), was found by Wang (1991). For any real martingale with independent symmetrically distributed increments, it follows from the results of Whittle (1960) and Haagerup (1982) that (4.8) takes place with the exact, in this "independent increments" case, constant $A_p = \|\xi\|_p$, where $\xi \sim N(0,1)$, $p \geq 2$. Because $\|\xi\|_p \asymp \sqrt{p}$, the bound in Theorem 4.3 is optimal (to $\asymp$).

REMARK. Bounds given in Theorem 3.1, 4.1 and 4.3 are only possible in *2-smooth* Banach spaces, even if we need a bound like those in Theorem 4.1 for just one particular p. Indeed, all the results mentioned here imply $\|f^*\|_p \leq CD\|S_2\|_p$ for at least one particular $p \geq 2$, some $C > 0$ and all Walsh–Paley martingales in $\mathcal{X}$ because for those martingales, $s_2 = S_2 \geq d^*$. Thus, one has (4.7) with $\varepsilon = (CD\delta/(\beta - 1 - \delta))^p$, and so, $\|f^*\|_2 \geq C_1 D \|S_2\|_2$ for some $C_1 > 0$. It remains now to recall the characterization of 2-smooth Banach spaces given by Pisier (1975).

5. Applications: Bounds on central moments of the norm of the sum of independent random vectors in arbitrary Banach spaces.

THEOREM 5.1. *If $f \in \mathcal{M}_{\mathrm{ind}}(\mathcal{X})$, $(\mathcal{X}, \|\cdot\|)$ is any separable Banach space, $p \geq 2$, $1 \leq c \leq p$, and x is any nonrandom vector in $\mathcal{X}$, then for all $n = 1, 2, \ldots,$*

$$\big\| \|f_n + x\| - \mathbf{E}\|f_n + x\| \big\|_p \preceq c\|d^*\|_p + \sqrt{c} e^{p/c} \|S_2\|_2. \tag{5.1}$$

In particular,

$$\big\| \|f_n + x\| - \mathbf{E}\|f_n + x\| \big\|_p \preceq p\|d^*\|_p + \sqrt{p} \|S_2\|_2, \tag{5.2}$$

$$\big\| \|f_n + x\| - \mathbf{E}\|f_n + x\| \big\|_p \preceq \frac{p}{\ln p} \big(\|d^*\|_p + \|S_2\|_2\big), \tag{5.3}$$

$$\big\| \|f_n + x\| - \mathbf{E}\|f_n + x\| \big\|_p \preceq \alpha\|d^*\|_p + e^{p/\alpha} \|S_2\|_2, \tag{5.4}$$

where $1 \leq \alpha \leq p/\ln(ep)$.

PROOF. Use Theorem 4.1 and a modification (needed here only for $x \neq 0$) of the method of Yurinskii (1974), as in Pinelis and Sakhanenko (1985). □

Hoffmann-Jørgensen (1974) found the following extension of Rosenthal's inequality for $f \in \mathcal{M}_{\text{ind}}(\mathcal{X})$, $p \geq 1$:

$$\| \| f_n \| \|_p \leq c(p)\big(\|d^*\|_p + \| \| f_n \| \|_1\big) \tag{5.5}$$

with $c(p)$ depending only on p [it can be seen that the best choice of parameters in the method of Hoffmann-Jørgensen gives (5.5) with $c(p) \asymp p$].

For $f \in \mathcal{M}_{\text{ind}}(\mathcal{X})$, Pinelis (1978) proved that

$$\| \| f_n \| \|_p \preceq c_1(p) \| \| f_n \| \|_1 + c_2(p) \|S_p\|_p + c_3(p) \|S_2\|_2,$$

which is also a generalization of the Rosenthal (1970) inequality; the method can actually yield $c_1(p) = 1$, $c_2(p) = p$ and $c_3(p) = \sqrt{p}$.

An inequality, implying (5.1), was obtained in Pinelis (1980) (see Remark 6.8 below).

De Acosta (1981) proved a version of (5.3) with $x = 0$ and with an implicit factor $c(p)$ instead of $p/\ln p$.

Using an isoperimetric technique, Talagrand (1989) proved the following version of (5.5):

$$\| \| f_n \| \|_p \preceq \frac{p}{\ln(2p)}\big(\|d^*\|_p + \| \| f_n \| \|_1\big) \tag{5.6}$$

for $f \in \mathcal{M}_{\text{ind}}(\mathcal{X})$, $p \geq 1$, which was also proved in Kwapień and Szulga (1991) by a different method. Inequality (5.6) may be compared with (5.1) and, in particular, with (5.3). If $\|S_2\|_2 \gg \|d^*\|_p + \| \| f_n \| \|_1$, then (5.3) may lose to (5.6) in certain cases. If, however, $\mathcal{X}$ is, for example, of cotype 2, then (5.3) is often no worse then (5.6). Moreover, say, in the typical case of increments with the same or almost the same distribution, (5.2) is significantly better. On the other hand, if there are heavy distribution tails, that is, if $\|d^*\|_p$ is much greater than both $\|S_2\|_2$ and $\| \| f_n \| \|_1$, then (5.4) with $\alpha \asymp 1$ does better than (5.6). Other advantages of bounds (5.1)–(5.4) are that they are applicable to the sums of non-zero-mean random vectors (owing to the presence of x) and better reflect the concentration phenomenon of the distribution of the sum of independent random vectors.

Modifications of the method of Yurinskii (1974) for $f \in \mathcal{M}_{\text{ind}}(\mathcal{X})$, allowing reduction of the problem of upper bounds on the l.h.s. of (5.1) for any separable Banach space $\mathcal{X}$ to that of upper bounds on $\|f_n\|_p$ for $f \in \mathcal{M}_{\text{ind}}(\mathbf{R})$, were proposed in Berger (1991) (for $x = 0$) and in Pinelis (1994) (with the best constant, for any $x \in \mathcal{X}$). Actually, instead of the power moment function $u \mapsto |u|^p$, one can use any convex function there.

A straightforward application of Theorem 4.1 yields the following bounds in the case of sums of independent zero-mean random variables, stated here mainly for further reference.

THEOREM 5.2. *Suppose that $f \in \mathcal{M}_{\text{ind}}(\mathcal{X})$, $\mathcal{X}$ is a Hilbert space, $p \geq 2$, $1 \leq c \leq p$. Then*

$$\| \| f_n \| \|_p \preceq c\|S_p\|_p + \sqrt{c}e^{p/c}\|S_2\|_2. \tag{5.7}$$

In particular,

$$\| \| f_n \| \|_p \preceq p\|S_p\|_p + \sqrt{p}\|S_2\|_2, \tag{5.8}$$

$$\| \| f_n \| \|_p \preceq \frac{p}{\ln p}\left(\|S_p\|_p + \|S_2\|_2\right), \tag{5.9}$$

$$\| \| f_n \| \|_p \preceq \alpha\|S_p\|_p + e^{p/\alpha}\|S_2\|_2, \qquad 1 \leq \alpha \leq \frac{p}{\ln(ep)}. \tag{5.10}$$

As was said in Section 4, the results of Pinelis (1980) imply (5.1) and (5.7). It was also explained in that paper how to elicit bounds like (5.8) and (5.10) (for $\alpha = 4$). Nevertheless, it is not so obvious how to deduce a general inequality like (5.7) from the spectrum of bounds in Pinelis (1980) (again, we refer to Remark 6.8). From this point of view, even in the classical case of sums of independent real-valued zero-mean random variables, (5.7) is apparently new.

An inequality similar to, say, (5.9), but with 2^p instead of $p/\ln p$, was probably first found by Rosenthal (1970) for $f \in \mathcal{M}_{\text{ind}}(\mathbf{R})$. Rosenthal also obtained a lower bound, which differs from the upper one by at most a factor depending only on p.

Rosén (1970) proved a result for $f \in \mathcal{M}_{\text{ind}}(\mathbf{R})$, which implies the upper Rosenthal bound for $p = 2, 4, 6, \dots$ [this implication was unnoticed; to demonstrate it, one can put, in the notation of Rosén, $\lambda_\nu(p) = (\mathbf{E}X_\nu^{2p}/\mathbf{E}X_\nu^2)^{1/(2p-2)}$, $\rho_\nu(p) = \mathbf{E}X_\nu^2/(\lambda_\nu(p)^2)$]. Moreover, using some ideas of Dharmadhikari and Jogdeo (1969), it is possible to deduce the upper bound by Rosenthal for all real $p \geq 2$ from the Rosén's result.

A method like that just described was used in the student diploma work of Pinelis (1974) to prove an upper bound of the Rosenthal type for $f \in \mathcal{M}_{\text{ind}}(\mathbf{R})$ [via the Marcinkiewicz and Zygmund (1937) inequality, the lower Rosenthal bound was also obtained therein]. Although the constants in Pinelis (1974) were implicit, the method could yield (5.9). Regrettably, the results of Rosenthal (1970) and Burkholder (1973) long remained unknown to the author, and so, to him, the problem of the constants was not among those considered most urgent.

Inequality (5.9) for $f \in \mathcal{M}_{\text{ind}}(\mathbf{R})$ and for sums of exchangeable random variables, with the proof that $p/\ln p$ is optimal in (5.9), was first given by Johnson, Schechtman and Zinn (1985).

A. I. Sakhanenko, a referee of the above-mentioned diploma work, suggested another approach, giving in effect (5.8), again for $f \in \mathcal{M}_{\text{ind}}(\mathbf{R})$ [see Nagaev and Pinelis (1977)].

An inequality, similar to (5.10) with $\alpha = 4$, was found by Sazonov (1974).

All bounds in Theorems 4.1, 5.1 and 5.2 are optimum. We shall prove the optimality in the next section, using ideas from Pinelis and Utev (1984), where, in particular, for any $p \geq 2$,

$$\sup\left\{\|f_n\|_p : f \in \mathcal{M}_{\text{ind}}(\mathbf{R}),\ \|S_2\|_2 \text{ and } \|S_p\|_p \text{ fixed}\right\}$$

was computed up to $\asymp$, which, for instance, implies all inequalities (5.7)–(5.10) for $f \in \mathcal{M}_{\mathrm{ind}}(\mathbf{R})$. Also, it was noted in Pinelis and Utev (1984) that bounds like (5.8) and (5.10) (for $\alpha \asymp 1$) represent in a certain sense the two extreme bounds in the spectrum of all optimal bounds on moments, the optimum value of a "spectrum parameter" depending on $\|S_p\|_p/\|S_2\|_2$.

6. Optimality of the bounds on moments. Let us consider the following upper bounds for any $a_2 > 0$, $a_p > 0$, $p > 2$:

$$\begin{aligned}
B_p &:= B_p(a_p, a_2) \\
&:= \sup\{\|f^*\|_p : f \in \mathcal{M}(\mathcal{X}),\ \mathcal{X} \text{ is any 2-smooth space}, \\
&\qquad\qquad \|d^*\|_p = a_p,\ D(\mathcal{X})\|s_2\|_p = a_2\}, \\
B^{\mathcal{X}}_{p,\mathrm{ind}} &:= B^{\mathcal{X}}_{p,\mathrm{ind}}(a_p, a_2) \\
&:= \sup\{\|\|f_n + x\| - \mathbf{E}\|f_n + x\|\|_p : f \in \mathcal{M}_{\mathrm{ind}}(\mathcal{X}),\ x \in \mathcal{X}, \\
&\qquad\qquad \|d^*\|_p = a_p,\ \|S_2\|_2 = a_2,\ n = 1, 2, \dots\}, \\
B_{p,\mathrm{ind},0} &:= B_{p,\mathrm{ind},0}(a_p, a_2) \\
&:= \sup\{\|f_n\|_p : f \in \mathcal{M}_{\mathrm{ind}}(\mathbf{R}),\ \|d^*\|_p = a_p,\ \|S_2\|_2 = a_2,\ n = 1, 2, \dots\},
\end{aligned}$$

and their analogues $B_{p,S} = B_{p,S}(a_p, a_2)$, $B^{\mathcal{X}}_{p,\mathrm{ind},S} = B^{\mathcal{X}}_{p,\mathrm{ind},S}(a_p, a_2)$, and $B_{p,\mathrm{ind},0,S} = B_{p,\mathrm{ind},0,S}(a_p, a_2)$ obtained by replacing the equality $\|d^*\|_p = a_p$ in the above definitions by $\|S_p\|_p = a_p$.

We shall show that all the introduced bounds are $\asymp$-equivalent to each of the following:

$$\begin{aligned}
B^*_p &:= B^*_p(a_p, a_2) := a_p + \sqrt{p}a_2 + \frac{pa_p}{\ln\left(2 + (a_p/a_2)\sqrt{p}\right)}, \\
\widehat{B}_p &:= \widehat{B}_p(a_p, a_2) := \min\{ca_p + \sqrt{c}e^{p/c}a_2 : 1 \le c \le p\}, \\
\check{B}_p &:= \check{B}_p(a_p, a_2) := \max\{(p\alpha + 1)^{1-\alpha/2}a_p^{1-\alpha}a_2^{\alpha} : 0 \le \alpha \le 1\}.
\end{aligned}$$

Theorem 6.1 below principally means that for any pair (a_p, a_2) of the values of the characteristics used in the bounds (4.1), (5.1) and (5.7), there exists a value of the "spectrum parameter" c providing an optimum bound. Roughly speaking, it means that spectra (4.1), (5.1), and (5.7) are rich enough, so Theorem 6.1 may be called the "spectrum completeness theorem." It also means that it is not essential in this context which of the two pairs of values are fixed: $\|d^*\|_p$ and $\|s_2\|_p$, or $\|S_p\|_p$ and $\|s_2\|_p$ [this is not as obvious as it might seem; although it is true, at least for $f \in \mathcal{M}_{\mathrm{ind}}(\mathbf{R})$, that $\|S_p\|_p \preceq \|d^*\|_p + \|s_2\|_p$, this is not sufficient for obtaining, say, (4.2) from $\|f^*\|_p \preceq p\|d^*\|_p + \sqrt{p}\|s_2\|_p$ even for $f \in \mathcal{M}_{\mathrm{ind}}(\mathbf{R})$].

THEOREM 6.1. *For all $p > 2$, $a_2 > 0$, $a_p > 0$, and all separable Banach spaces $(\mathcal{X}, \|\cdot\|)$,*

$$B_p \asymp B^{\mathcal{X}}_{p,\mathrm{ind}} \asymp B_{p,\mathrm{ind},0} \asymp B_{p,S} \asymp B^{\mathcal{X}}_{p,\mathrm{ind},S} \asymp B_{p,\mathrm{ind},0,S} \asymp \check{B}_p \asymp \hat{B}_p \asymp B^*_p.$$

The proof is comparatively long and will be given later in this section.

The "spectrum" $ca_p + \sqrt{c}e^{p/c}a_2$, $1 \le c \le p$, turns out to be not only "complete" but also "minimal" in the sense that for each $c \in [1,p]$, there exist $a_p > 0$ and $a_2 > 0$ such that the "individual" bound $ca_p + \sqrt{c}e^{p/c}a_2$ is the best possible. Let us give the rigorous statement.

THEOREM 6.2. *For any* $c \in [1, p]$, $p > 2$,

$$\sup\left\{\frac{B_p(a_p, a_2)}{ca_p + \sqrt{c}e^{p/c}a_2}: a_p > 0,\ a_2 > 0\right\} \asymp 1.$$

Here any of the other five bounds $B^{\mathcal{X}}_{p,\,\mathrm{ind}}$, $B_{p,\,\mathrm{ind},\,0}$, $B_{p,S}$, $B^{\mathcal{X}}_{p,\,\mathrm{ind},\,S}$, *or* $B_{p,\,\mathrm{ind},\,0,\,S}$ *may be used in place of* B_p. *In particular, all bounds* (4.2)–(4.4), (5.2)–(5.4) *and* (5.8)–(5.10) *are optimal.*

The proof will be given after that of Theorem 6.1.

The following proposition might seem analogous to Theorem 6.2, but it is less important because for an "individual" value of α, $(p\alpha + 1)^{1-\alpha/2}a_p^{1-\alpha}a_2^{\alpha}$ does not represent an upper bound on the moments. Actually, by Theorem 6.1, $B_p(a_p, a_2) \succeq (p\alpha + 1)^{1-\alpha/2}a_p^{1-\alpha}a_2^{\alpha}$.

PROPOSITION 6.3. *For any* $\alpha \in [0, 1]$, $p > 2$,

$$\inf\left\{\frac{B_p(a_p, a_2)}{(p\alpha + 1)^{1-\alpha/2}a_p^{1-\alpha}a_2^{\alpha}}: a_p > 0,\ a_2 > 0\right\} \asymp 1.$$

Here, any of the other five bounds $B^{\mathcal{X}}_{p,\,\mathrm{ind}}$, $B_{p,\,\mathrm{ind},\,0}$, $B_{p,S}$, $B^{\mathcal{X}}_{p,\,\mathrm{ind},\,S}$ *or* $B_{p,\,\mathrm{ind},\,0,\,S}$ *may be used in place of* B_p.

The proof will be given after that of Theorem 6.2.

REMARK. It is easy to see that Doob's inequality

$$\|f\|_p \le \frac{p}{p-1}\sup_n \|\|f_n\|\|_p, \qquad p > 1,\ f \in \mathcal{M}(\mathcal{X})$$

[see also (1.4) in Burkholder (1973)], remains true for all separable Banach spaces $\mathcal{X}$. Therefore, one could replace $\|f^*\|_p$ in the definition of B_p by $\sup_n \|\|f_n\|\|_p$, and statements 6.1–6.3 would hold. This remark can be also deduced from the proof of these statements.

Assume in what follows up to the end of this section, without loss of generality, that $a_p = 1$ and set $a = a_2$ for brevity.

The proof of Theorem 6.1 is based on the following series of lemmas.

The next lemma was prompted by a remark from the referee. Using it has made the proof of Theorem 6.1 more direct and transparent.

LEMMA 6.4. *Let Y be a random variable, having the Poisson distribution with a parameter λ. Then, for sufficiently large p,*

$$\|Y-\lambda\|_p \succeq \frac{p}{\ln(p/\lambda)} I\{e^{-p/2} \le \lambda \le e^{-3}p\} + \sqrt{p\lambda}\, I\{\lambda \ge 1\} + \lambda^{1/p} I\left\{0 < \lambda \le \frac{1}{2}\right\}.$$

PROOF. Consider first the case $\lambda \in [e^{-p/2}, e^{-3}p]$. Then $p \ge [1 \vee (e\lambda)] \ln(p/\lambda)$. Put $n = \lambda + \theta p/\ln(p/\lambda)$. We may choose $\theta \in [1,2]$ so that n is integer. Then, in addition, $p/\ln(p/\lambda) \le n \le 3p/\ln(p/\lambda)$ and, hence, $\lambda/n \le 1/e < 1$. Using now $n! \le n^n$ and $p/\lambda \ge e^3$, one obtains

$$\begin{aligned}
\ln\|Y-\lambda\|_p &\ge \frac{1}{p}\ln\left(\frac{|n-\lambda|^p \lambda^n}{n!} e^{-\lambda}\right) \\
&\ge \ln\frac{p}{\ln(p/\lambda)} + \frac{3}{\ln(p/\lambda)} \ln\left(\frac{\ln(p/\lambda)}{3(p/\lambda)}\right) - \frac{1}{p/\lambda} \\
&\ge \ln\frac{p}{e^4 \ln(p/\lambda)}.
\end{aligned}$$

Next, consider the case $\lambda \ge 1$. Using induction on n, one can see that the Poisson distribution with the parameter $\lambda \ge 1$ majorizes the normal distribution in the sense that

$$\frac{\lambda^n}{n!}\exp(-\lambda) \succeq \frac{1}{\sqrt{\lambda}}\exp\left(-\frac{(n-\lambda)^2}{\lambda}\right), \qquad n \ge \lambda.$$

To complete the consideration of this case, it remains to recall that if ξ is $N(0,\sigma^2)$, then $\|\xi\|_p \asymp \sqrt{p}\sigma$.

Finally, in the case $0 < \lambda \le \frac{1}{2}$, one has $\|Y-\lambda\|_p \ge |1-\lambda|\lambda^{1/p}e^{-\lambda/p} \succeq \lambda^{1/p}$. □

LEMMA 6.5. *For all $p \ge 2$ and $a > 0$, one has $B_{p,\mathrm{ind},0}(1,a) \succeq B_p^*(1,a)$.*

PROOF. Let us say that a real-valued random variable ξ is $T(u,q)$ if $\mathbf{P}(\xi = u) = \mathbf{P}(\xi = -u) = q/2$ and $\mathbf{P}(\xi = 0) = 1-q$, $u > 0$. Results in Pinelis and Utev (1984) and Utev (1985) suggest that sums of independent $T(u,q)$ copies constitute the extremal case with respect to the upper Rosenthal bound. Here, we exploit this idea. The summands in the expression

$$B_p^*(1,a) = 1 + \sqrt{p}a + \frac{p}{\ln(2+\sqrt{p}/a)}$$

correspond, respectively, to (i) the "quasidegenerate" case, when q is small, (ii) the "quasinormal" case, when q is large (i.e., comparable to 1), and (iii) the intermediate "quasi-Poisson" case.

The next construction is used in order to satisfy the conditions $\|d^*\|_p = 1$ and $\|S_2\|_2 = a$. Put $g_n(t) = t^{-p/2}[1 - (1 - t/n)^n]$, $g_\infty(t) = t^{-p/2}(1 - e^{-t})$, $t > 0$, $n > 0$. Then $g_n(t)$ decreases in n to $g_\infty(t)$, $g_n(t)$ decreases in $t \in (0, n]$ from $g_n(0+) = \infty$ to $g_n(n) = n^{-p/2}$, and $g_\infty(t)$ decreases in $t \in (0, \infty)$ from $g(0+) = \infty$ to $g(\infty) = 0$. Hence, for any $n \in (a^2, \infty]$, there exists a unique solution $t_n \in (0, n)$ to the equation $g_n(t_n) = a^{-p}$, and $t_n \downarrow t_\infty$ as $n \uparrow \infty$. Put

$$\lambda = t_\infty, \qquad \kappa = \frac{a}{\sqrt{\lambda}}. \tag{6.1}$$

Consider the martingale $f = (f_j) \in \mathcal{M}_{\text{ind}}(\mathbf{R})$ whose first n increments are $T(u_n, q_n)$, where $q_n := t_n/n$, $u_n := a/\sqrt{t_n}$, and the other increments are 0. Then, for the so constructed f, one has $\|d^*\|_p = 1$ and $\|S_2\|_2 = a$. In addition, $\mathbf{E}e^{itf_n} = [1 + q_n(\cos tu_n - 1)]^n \to \mathbf{E}e^{itZ}$, $t \in \mathbf{R}$, $n \to \infty$, where Z is a (symmetrized Poisson) random variable with the characteristic function $\mathbf{E}e^{itZ} = \exp[(\cos \kappa t - 1)\lambda]$. Hence, by the analogue of the Fatou lemma for convergence in distribution [see, e.g., Theorem 5.3 in Billingsley (1968)],

$$\liminf_{n \to \infty} \|f_n\|_p \geq \|Z\|_p. \tag{6.2}$$

In what follows up to the end of the proof of the lemma, we assume that p is large enough (otherwise, the lemma follows from the lower Rosenthal bound).

By (6.1), $\lambda = (1 - e^{-\lambda})^{2/p} a^2$. This and the inequalities $1 - e^{-\lambda} \geq \lambda/2$ for $\lambda \in (0, 1]$, $(1 - e^{-\lambda})^{2/p} \geq \frac{1}{2}$ for $\lambda > 1$, $1 - e^{-\lambda} < 1$ for $\lambda > 0$ and $1 - e^{-\lambda} \leq \lambda$ imply

$$\tfrac{1}{2}\left[a^2 \wedge a^{2p/(p-2)}\right] \leq \lambda \leq a^2 \wedge a^{2p/(p-2)}. \tag{6.3}$$

To prove the lemma, it suffices to show that for some n,

$$\|f_n\|_p \succeq F(a,p) := 1 \vee (\sqrt{p}a) \vee \frac{p}{7\ln(2 + \sqrt{p}/a)}. \tag{6.4}$$

There may be three cases.

Case 1. $F(a, p) = 1$. Then $a < 1$. In this case, (6.4) (as well as the condition $n > a^2$, needed for the construction of f) is satisfied if $n = 1$. (Another way to treat this case is using Lemma 6.4, as we do in the two other cases.)

Case 2. $F(a, p) = \sqrt{p}a$. Then $a \succeq \sqrt{p}$, and (6.3) yields $\lambda \geq 1$. Using now Lemma 6.4 and (6.1), one has $\|Z\|_p \geq \kappa\|Y - \lambda\|_p \succeq \sqrt{p}a$. In this case, it remains to use (6.2).

Case 3. $F(a, p) = p/[7 \ln(2 + \sqrt{p}/a)]$. Then $e^{-p/6} \leq a \leq \sqrt{p/e^3}$ and, in view of (6.3), $e^{-p/2} \leq \lambda \leq p/e^3$. Because in this case $a^{1/p} \asymp 1$, (6.3) implies $\lambda \asymp a^2 \asymp a^{2p/(p-2)}$. In addition, (6.1) and (6.3) yield $\kappa \geq 1$. Using now Lemma 6.4, one has

$$\|Z\|_p \geq \kappa\|Y - \lambda\|_p \geq \|Y - \lambda\|_p \succeq \frac{p}{\ln(p/\lambda)} \asymp \frac{p}{\ln(\sqrt{p}/a)} \asymp \frac{p}{\ln(2 + \sqrt{p}/a)}.$$

Thus, in any case, (6.4) is true. □

LEMMA 6.5S. *For all $p \geq 2$ and $a > 0$, one has $B_{p,\,\mathrm{ind},\,0,\,S}(1,a) \succeq B_p^*(1,a)$.*

PROOF. The proof is quite similar to that of Lemma 6.5, and even slightly easier, because the conditions $\|S_p\|_p = 1$ and $\|S_2\|_2 = a$ are easier to satisfy, as compared to $\|d^*\|_p = 1$ and $\|S_2\|_2 = a$. In addition, here (6.3) is replaced by $\lambda = a^{2p/(p-2)}$. Finally, in the case $F(a,p) = p/[7\ln(2+\sqrt{p}/a)]$, one should use $\kappa = a^{-2/(p-2)} \succeq 1$, in place of $\kappa \geq 1$. □

LEMMA 6.6. *For all $p \geq 2$ and $a > 0$, one has $\widehat{B}_p(1,a) \preceq \check{B}_p(1,a)$.*

PROOF. Consider the functions

$$g_p(\alpha) := (p\alpha+1)^{1-\alpha/2}a^{\alpha}, \qquad 0 \leq \alpha \leq 1, \tag{6.5}$$

$$q_p(\alpha) := \sqrt{p\alpha+1}\exp\left\{\frac{1}{2}\frac{p\alpha-2p}{p\alpha+1}\right\}, \qquad 0 \leq \alpha \leq 1. \tag{6.6}$$

Then, by the definition of $\check{B}_p$,

$$\check{B}_p(1,a) = \max\{g_p(\alpha)\colon 0 \leq \alpha \leq 1\}. \tag{6.7}$$

In addition, q_p is continuous and increasing and maps [0, 1] onto the segment

$$I_p := \left[e^{-p}, \sqrt{p+1}\exp\left\{-\frac{1}{2}\frac{p}{p+1}\right\}\right]. \tag{6.8}$$

An essential relation between g_p and q_p is

$$g_p'(\alpha) = g_p(\alpha)\ln\frac{a}{q_p(\alpha)}. \tag{6.9}$$

If we have $a \in I_p$, take $\alpha_p = q_p^{-1}(a)$. Then $\alpha_p \in [0,1]$, $a = q_p(\alpha_p)$ and, in view of (6.5) and (6.6),

$$\begin{aligned} p\alpha_p + 1 &= g_p(\alpha_p)\exp\left\{-\frac{\alpha_p}{2}\frac{p\alpha_p-2p}{p\alpha_p+1}\right\} \asymp g_p(\alpha_p),\\ a\sqrt{p\alpha_p+1}\exp\left\{\frac{p}{(p\alpha_p+1)}\right\} &= g_p(\alpha_p)\exp\left\{\frac{p\alpha_p(3-\alpha_p)}{2(p\alpha_p+1)}\right\} \asymp g_p(\alpha_p). \end{aligned} \tag{6.10}$$

This implies, with $c_p := p\alpha_p + 1$, that

$$c_p + \sqrt{c_p}e^{p/c_p}a \asymp g_p(\alpha_p) \tag{6.11}$$

(if $a \in I_p$). Obviously, $c_p \in [1, p+1]$. Hence, in view of (6.7), if $a \in I_p$ and $c_p \leq p$, the lemma is true. If, however, $c_p \in (p, p+1]$, then

$$p + \sqrt{p}e^{p/p}a \leq c_p + \sqrt{c_p}e^{p/c_p}ea \asymp g_p(\alpha_p)$$

by (6.11). Thus, the lemma is true whenever $a \in I_p$.

Consider now the two cases when $a \notin I_p$.

First, suppose that

$$a < \min I_p = e^{-p}. \tag{6.12}$$

Then $a < q_p(\alpha)$ for all $\alpha \in [0,1]$, and (6.9) implies $g_p' < 0$ on $[0,1]$. Hence, $\check{B}_p(1,a) = g_p(0) = 1$. On the other hand, $\widehat{B}_p(1,a) \le 1 + e^p a \le 2$ by (6.12). Thus, the lemma is true in the case (6.12), too.

Finally, let

$$a > \max I_p\big(>\sqrt{p}/e\big). \tag{6.13}$$

Then $a > q_p(\alpha)$ for all $\alpha \in [0,1]$, and (6.6) implies $\check{B}_p(1,a) = g_p(1) \ge \sqrt{p}a$, whereas $\widehat{B}_p(1,a) \le p + \sqrt{p}a = (p/a + \sqrt{p}e)a \le 2e\sqrt{p}a$, in view of (6.13). □

LEMMA 6.7. *For all $p \ge 2$ and $a > 0$, one has $\check{B}_p(1,a) \preceq B_p^*(1,a)$.*

PROOF. Consider first the case $a \in I_p$, where a and I_p are defined in (6.5) and (6.8), respectively. Then (6.10) implies

$$\check{B}_p(1,a) = g_p(\alpha_p) \asymp c_p, \tag{6.14}$$

where g_p is given by (6.5), $\alpha_p = q_p^{-1}(a)$, and $c_p = p\alpha_p + 1$. Putting $z = (p + \frac{1}{2})/c_p$, one has $z > \frac{1}{2}$. We can rewrite the equation $\alpha_p = q_p^{-1}(a)$ as $\sqrt{z}e^z = \sqrt{e(p+\frac{1}{2})/a}$. Because $e^z/2 < \sqrt{z}e^z < e^{2z}$ when $z > \frac{1}{2}$, we deduce $z \asymp \ln(2 + \sqrt{p}/a)$,

$$c_p \asymp \frac{p}{\ln\left(2 + \sqrt{p}/a\right)}. \tag{6.15}$$

Now (6.14) and (6.15) imply $\check{B}_p \preceq B_p^*$ if $a \in I_p$.

In the cases (6.12) and (6.13), $\check{B}_p(1,a)$ is equal to $g_p(0) = 1$ and $g_p(1) \asymp \sqrt{p}a$, respectively. We see that in any case, $\check{B}_p \preceq B_p^*$. □

PROOF OF THEOREM 6.1. Observe that

$$B_p^* \preceq B_{p,\mathrm{ind},0} \preceq B_p \preceq \widehat{B}_p \preceq \check{B}_p \preceq B_p^*. \tag{6.16}$$

Indeed, the first relation in this chain is Lemma 6.5, the second is trivial, the third follows from (4.1), the fourth is Lemma 6.6 and the last is Lemma 6.7.

Note that for any separable Banach space $\mathcal{X}$,

$$B_{p,\mathrm{ind},0} \preceq B_{p,\mathrm{ind}}^{\mathcal{X}} \tag{6.17}$$

because $\mathbf{R}$ may be isometrically embedded into $\mathcal{X}$ and, for $f \in \mathcal{M}_{\mathrm{ind}}(\mathbf{R})$,

$$\|\,|f_n + x| - \mathbf{E}|f_n + x|\,\|_p \to \|f_n\|_p \quad \text{as } x \to \infty.$$

It follows from Theorem 5.1 that

$$B^{\mathcal{X}}_{p,\,\mathrm{ind}} \preceq \widehat{B}_p. \tag{6.18}$$

Now (6.16)–(6.18) yield

$$B^*_p \asymp B_{p,\,\mathrm{ind},\,0} \asymp B_p \asymp \widehat{B}_p \asymp \check{B}_p \asymp B^{\mathcal{X}}_{p,\,\mathrm{ind}}.$$

Analogously, using Lemma 6.5S in place of Lemma 6.5, one obtains

$$B^*_p \asymp B_{p,\,\mathrm{ind},\,0,\,S} \asymp B_{p,\,S} \asymp \widehat{B}_p \asymp \check{B}_p \asymp B^{\mathcal{X}}_{p,\,\mathrm{ind},\,S}. \qquad \square$$

PROOF OF THEOREM 6.2. For any given $c \in [1,p]$, set $a = \sqrt{c}e^{-p/c}$. Then $c = \sqrt{c}e^{p/c}a$. Put $\alpha = (c-1)/p$, so that $c = p\alpha + 1$, $0 \le \alpha \le 1$. Then

$$\begin{aligned}\check{B}_p(1,a) \ge g_p(\alpha) = ce^{-p\alpha/c} &\ge e^{-1}c\\ &= (2e)^{-1}\big(c + \sqrt{c}e^{p/c}a\big),\end{aligned}$$

where g_p is defined by (6.5). It remains to apply Theorem 6.1 and also to recall that inequalities (4.2)–(4.4), (5.2)–(5.4) and (5.8)–(5.10) were obtained by choosing particular values of c. $\square$

PROOF OF PROPOSITION 6.3. For any given $\alpha \in [0,1]$, set

$$a = q_p(\alpha), \tag{6.19}$$

where $q_p(\alpha)$ is defined by (6.6). Then, putting $c = p\alpha + 1$, we see that $1 \le c \le p+1$ and, in view of (6.5) and (6.19),

$$g_p(\alpha) = c\exp\left\{\frac{p\alpha}{2}\frac{\alpha-2}{p\alpha+1}\right\} \asymp c \asymp c + \sqrt{c}\exp(p/c)a$$

because a consequence of (6.19) and (6.6) is

$$\sqrt{c}\exp\left(p/c\right)a = c\exp\left\{\frac{p\alpha}{2(p\alpha+1)}\right\} \asymp c.$$

Thus, if the above-defined $c \le p$, we see that

$$\widehat{B}_p \preceq g_p(\alpha). \tag{6.20}$$

If, however, $c \in (p, p+1]$, then

$$g_p(\alpha) \asymp c + \sqrt{c}e^{p/c}a \ge p + \sqrt{p}e^{p/(p+1)}a \succeq B_p,$$

so that (6.20) holds. Now it remains to apply Theorem 6.1. $\square$

REMARK 6.8. In Pinelis (1980), a bound on moments of martingales in Hilbert spaces with bounded conditional second moments was obtained. [In fact, as was stated in Pinelis (1980), the bound given there holds for Banach spaces of the class $D(A_1,A_2)$ introduced by Zolotarev (1977); because it is possible to reduce (2.1) to (2.2) via Lemma 2.2, the class $D(A_1,A_2)$ is in effect the same as the class of 2-smooth spaces.] In particular, the corollary in Pinelis (1980) implies $B^{\mathcal{X}}_{p,\,\mathrm{ind},\,S}(1,a) \preceq \check{B}_p(1,a)$. To verify this implication (which requires some effort), choose in that corollary $a_j = a^{-p/(p-2)}$ for all j,

$$p(s) = \frac{e^3 y_s^s}{s a^2 y_{s-2}^{s-2}}, \qquad q(s) = 10^s y_s^{s-2} \quad \text{for } s \ge 2, \text{ where } y_s := \check{B}_s(1,a).$$

7. Chung type bounds on moments. Consider (for a fixed natural number n)

$$B_p^{\mathrm{Ch}}(a_p,n) := \sup\Big\{\big\|\, \|f_n\|\,\big\|_p\colon (0,f_1,\dots,f_n,f_n,f_n,\dots) \in \mathcal{M}(\mathcal{X}),$$
$$\mathcal{X} \text{ is any 2-smooth space}, D(\mathcal{X})\|S_p\|_p = a_p\Big\}.$$

THEOREM 7.1. *For all $p > 2$, $a_p > 0$, $n = 1,2,\dots$, one has*

$$B_p^{\mathrm{Ch}}(a_p,n) \asymp \sqrt{p \wedge n}\; n^{(p-2)/2p} a_p.$$

PROOF. Let $(0,f_1,\dots,f_n,f_n,f_n,\dots) \in \mathcal{M}(\mathcal{X})$. Using Hölder's inequality two times, we see that $\|s_2\|_p \le n^{(p-2)/(2p)}\|S_p\|_p$. If $p \le 3n^{(p-2)/p}$, then $p\|d^*\|_p \le p\|S_p\|_p \preceq p^{1/2} n^{(p-2)/2p}\|S_p\|_p$, and hence, (4.2) and $p \le 3n$ yield

$$B_p^{\mathrm{Ch}}(a_p,n) \preceq \sqrt{p \wedge n}\, n^{(p-2)/2p} a_p. \tag{7.1}$$

If now $p > 3n^{(p-2)/p}$, then $p > 3$, $n < p^{p/(p-2)} \preceq p$. The inequalities of Minkowski and Hölder give $\|\|f_n\|\|_p \le \sum_{j=1}^n \|\|d_j\|\|_p \le n^{(p-1)/p}\|S_p\|_p$. This and $p \succeq n$ show that (7.1) holds. It remains to prove that

$$B_p^{\mathrm{Ch}}(a_p,n) \succeq \sqrt{p \wedge n}\; n^{(p-2)/2p} a_p. \tag{7.2}$$

Let d_j be independent, $\mathbf{P}(d_j = \pm u) = 1/2$, $u > 0$, $j = 1,\dots,n$. Then the multinomial formula yields

$$\|f_n\|_{2m} = \left(\sum_{j=1}^{m \wedge n} \binom{n}{j} j!\,\Gamma_{j,m}\right)^{1/2m} u \asymp \left(\sum_{j=1}^{m \wedge n} \binom{n}{j} j^{2m}\right)^{1/2m} u. \tag{7.3}$$

Here, we put

$$\Gamma_{j,m} = (2m)! \sum_{r=1}^{j} \sum \prod_{k=1}^{r} \frac{1}{j_k!\,\big((2m_k)!\big)^{j_k}},$$

the inner summation being over all positive integers $m_1 > \cdots > m_r$ and $j_1, \dots, j_r$ such that $m_1 j_1 + \cdots + m_r j_r = m$, $j_1 + \cdots + j_r = j$, and use the estimate $\Gamma_{j,m}^{1/(2m)} \asymp j^{1-j/(2m)}$ [cf. (9) in Pinelis and Utev (1984)].

Let m be the integer part of $p/2$. If $m \le n$, then (7.3) shows that

$$\|f_n\|_p \ge \|f_n\|_{2m} \succeq \binom{n}{m}^{1/2} mu \ge \left(\frac{n}{m}\right)^{1/2} mu = \sqrt{m} n^{(p-2)/2p} \|S_p\|_p,$$

which implies (7.2) in the case $m \le n$. If, finally, $m > n$, then it follows from (7.3) that

$$\|f_n\|_p \ge \|f_n\|_{2m} \succeq \binom{n}{m}^{1/2m} nu = n^{(p-1)/p} \|S_p\|_p,$$

so that (7.2) holds. □

Chung [(1951), pages 348–349] showed that in the case $f \in \mathcal{M}_{\text{ind}}(\mathbf{R})$, the inequality of Marcinkiewicz and Zygmund [(1937), page 87] implies an estimate like (7.1) but with some $C_0(p)$ depending only on p, instead of $\sqrt{p \wedge n}$. As was pointed out in Dharmadhikari, Fabian and Jogdeo (1968), an analogous result for $f \in \mathcal{M}(\mathbf{R})$ is implied by the generalization of the Marcinkiewicz–Zygmund inequality obtained by Burkholder [(1966), Theorem 9; see also Burkholder (1973), Theorems 3.2 and 15.1]. One can see that Theorem 15.1 in Burkholder (1973) in fact gives $C_0(p) = p$; a constant of the same order p is given in Theorem 3.2 of Burkholder (1973).

The direct proof due to Dharmadhikari, Fabian and Jogdeo (1968) yields $C_0(p) = p2^p$. For $f \in \mathcal{M}_{\text{ind}}(\mathbf{R})$, Dharmadhikari and Jogdeo (1969) obtained $C_0(p) = p/\ln p$.

For $f \in \mathcal{M}_{\text{ind}}(\mathbf{R})$, the result of Whittle (1960) implies the Marcinkiewicz–Zygmund inequality with the best constant and, along with the above-mentioned remark of Chung [see also Rosén (1970)], leads to (7.1) but with $\sqrt{p}$ instead of $\sqrt{p \wedge n}$; so, for $n > p$, it gives the optimum.

What has been said is a reason for the referring to (7.1) as an optimum bound on moments of the Chung type.

REMARK. Bounds of the Chung type on central moments of the norm of the sum of independent random vectors in any separable Banach space can be easily derived from Theorem 7.1 (cf. Theorem 5.1).

8. One-sided bounds for the distributions of real-valued (super) martingales. Let $\mathcal{M}_-$ stand for the set of all real-valued supermartingales $f \in \mathcal{S}(\mathbf{R})$. For $f \in \mathcal{M}_-$, put $f_+^* = \sup_j f_j$ and $d_+^* = \sup_j d_j$.

THEOREM 8.1. *If $f \in \mathcal{M}_-$ and $\lambda > 0$ are such that $\mathbf{E}e^{\lambda d_j} < \infty$ for each j, then for all $r \geq 0$,*

$$\mathbf{P}(f_+^* \geq r) \leq \exp(-\lambda r)\left\|\prod_{j=1}^{\infty}(1+e_j)\right\|_\infty \leq \exp\left\{-\lambda r + \left\|\sum_{j=1}^{\infty} e_j\right\|_\infty\right\},$$

where $e_j := \mathbf{E}_{j-1}(e^{\lambda d_j} - 1 - \lambda d_j)$.

PROOF. The proof follows from the trivial remark that the sequence $G_0 := 1$, $G_j := e^{\lambda f_j}\prod_{i=1}^{j}(1+e_i)^{-1}$, $j = 1, 2, \dots$, is a positive supermartingale (cf. the end of the proof of Theorem 3.2). □

THEOREM 8.2. *Suppose that $f \in \mathcal{M}_-$, $\|d_+^*\|_\infty \leq a$ and $\|s_2\|_\infty \leq b$ for some $a > 0$, $b > 0$. Then for all $r \geq 0$,*

$$\mathbf{P}(f_+^* \geq r) \leq \exp\left[\frac{r}{a} - \left(\frac{r}{a} + \frac{b^2}{a^2}\right)\ln\left(1 + \frac{ra}{b^2}\right)\right] \leq \left(\frac{eb^2}{ra}\right)^{r/a}.$$

PROOF. The proof is quite similar to that of Theorem 3.4, but uses Theorem 8.1 in place of Theorem 3.1. □

THEOREM 8.3. *If $f \in \mathcal{M}_-$, $p \geq 2$, $1 \leq c \leq p$, then*

$$\|f_+^*\|_p \preceq c\|d_+^*\|_p + \sqrt{c}e^{p/c}\|s_2\|_p.$$

We use the following lemma.

LEMMA 8.4 (cf. Lemma 4.2). *If $\lambda > 0$, $\delta_1 > 0$, $\delta_2 > 0$, $\beta - 1 - \delta_2 > 0$, and $f \in \mathcal{M}_-$, then*

$$\mathbf{P}(f_+^* > \beta\lambda,\ w_+^* \leq \lambda) \leq \varepsilon\mathbf{P}(f_+^* > \lambda),$$

where

$$w_+^* = \left(\frac{d_+^*}{\delta_2}\right) \vee \left(\frac{s_2}{\delta_1}\right),$$

$$\varepsilon = \left(\frac{e}{N}\frac{\delta_1^2}{\delta_2^2}\right)^N, \qquad N = \frac{\beta - 1 - \delta_2}{\delta_2}.$$

PROOF. Put $\bar{d}_j = d_j I\{d_j \leq \delta_2\lambda\}$, $\bar{f}_j = \bar{d}_0 + \cdots + \bar{d}_j$, $\bar{s}_{2,j+1} = (\sum_{i=1}^{j+1}\mathbf{E}_{i-1}\bar{d}_i^2)^{1/2}$, $\bar{h}_j = \bar{f}_{(j\wedge\tau\wedge\nu)\vee\mu} - \bar{f}_\mu$, $j = 0, 1, \dots$, where $\mu = \inf\{j: \bar{f}_j > \lambda\}$, $\nu = \inf\{j: \bar{f}_j > \beta\lambda\}$, and $\tau = \inf\{j: \bar{s}_{2,j+1} > \delta_1\lambda\}$. Then $(\bar{h}_j)$ is a supermartingale conditionally on F_μ, and

$$\begin{aligned}\mathbf{P}(f_+^* > \beta\lambda,\, w_+^* \le \lambda) &= \mathbf{P}(\bar{f}_+^* > \beta\lambda,\, w_+^* \le \lambda)\\ &\le \mathbf{P}(\bar{h}_+^* > (\beta - 1 - \delta_2)\lambda)\\ &= \mathbf{E}\mathbf{P}(h_+^* > (\beta - 1 - \delta_2)\lambda \mid F_\mu) I\{\mu < \infty\}\\ &= \varepsilon\mathbf{P}(\bar{f}_+^* > \lambda) \le \varepsilon\mathbf{P}(f_+^* > \lambda).\end{aligned}$$

Here, we put $\bar{f}_+^* = \sup_j \bar{f}_j$, $\bar{h}_+^* = \sup_j \bar{h}_j$ and took into account Theorem 8.2. □

PROOF OF THEOREM 8.3. The proof is similar to that of Theorem 4.1, but simpler. Here, we do not need to symmetrize. Instead, we can apply Lemma 7.1 of Burkholder (1973) directly to f_+^*. □

The following is a refinement of Theorem 8.1.

THEOREM 8.5. *Let $f \in \mathcal{M}(\mathbf{R})$, $d_j = u_j - \mathbf{E}_{j-1}u_j$, $\lambda > 0$, and $\mathbf{E}e^{\lambda u_j} < \infty$, $j = 1, 2, \ldots$. Then for all $r \ge 0$,*

$$\mathbf{P}(f_+^* \ge r) \le \exp\left\{-\lambda r + \left\|\sum_{j=1}^{\infty} e_j\right\|_\infty\right\},$$

where $e_j = \mathbf{E}_{j-1}(e^{\lambda u_j} - 1 - \lambda u_j)$. If in addition, $\mathbf{E}_{j-1}u_j \ge 0$ for all $j = 1, 2, \ldots$, then, moreover,

$$\mathbf{P}(f_+^* \ge r) \le e^{-\lambda r}\left\|\prod_{j=1}^{\infty}(1 + e_j)\right\|_\infty.$$

PROOF. The proof is analogous to that in Pinelis and Sakhanenko (1985) but simpler. The elementary inequalities

$$\begin{aligned}(1 + a + b)e^{-b} &\le e^a, &\quad a \in \mathbf{R},\, b \in \mathbf{R},\\ (1 + a + b)e^{-b} &\le 1 + a, &\quad a \ge 0,\, b \ge 0,\end{aligned}$$

imply, respectively, that

$$\mathbf{E}_{j-1}e^{\lambda d_j} \le e^{e_j}$$

and, if $\mathbf{E}_{j-1}u_j \ge 0$,

$$\mathbf{E}_{j-1}e^{\lambda d_j} \le 1 + e_j,$$

$j = 1, 2, \ldots$, if one chooses $a = e_j$, $b = \lambda\mathbf{E}_{j-1}u_j$. Hence,

$$\exp\left\{\lambda f_j - \sum_{i=1}^{j} e_j\right\}, \qquad j = 0, 1, 2, \ldots,$$

is a supermartingale and, if $\mathbf{E}_{j-1}u_j \geq 0 \ \forall j$, so is

$$e^{\lambda f_j} \prod_{i=1}^{j} (1 + e_i)^{-1}, \qquad j = 0, 1, 2, \ldots.$$

It remains to use reasoning like that at the end of the proof of Theorem 3.2. □

REMARK. Martingales like those in Theorem 8.5 may arise, for example, as a result of truncating and subsequent centering of the increments of other martingales or any other adapted sequences. The aim of Theorem 8.5 is to provide for the best constants in exponential inequalities for martingales, which cannot be reached, for example, via the straightforward estimate $|d_j| \leq |u_j| + |\mathbf{E}_{j-1}u_j|$.

As an illustration, let us give the following corollaries to Theorem 8.5, which are refinements of Theorems 3.3 and 3.4, respectively, for the particular case $\mathcal{X} = \mathbf{R}$.

THEOREM 8.6. *Suppose that* $f \in \mathcal{M}(\mathbf{R})$, $d_j = u_j - \mathbf{E}_{j-1}u_j, j = 1, 2, \ldots,$ *and*

$$\left\| \sum_{j=1}^{\infty} \mathbf{E}_{j-1} |u_j|^m \right\|_{\infty} \leq m!\, \Gamma^{m-2} B^2/2$$

for some $\Gamma > 0, B > 0$, *and* $m = 2, 3, \ldots.$ *Then for all* $r \geq 0$,

$$\mathbf{P}(f_+^* \geq r) \leq \exp\left(-\frac{r^2}{B^2 + B\sqrt{B^2 + 2\Gamma r}} \right).$$

PROOF. The proof is almost literally the same as that of Theorem 3.3, but rests upon Theorem 8.5 instead of Theorem 3.1. □

THEOREM 8.7. *Suppose that* $f \in \mathcal{M}(\mathbf{R})$, $d_j = u_j - \mathbf{E}_{j-1}u_j, j = 1, 2, \ldots,$ *and* $\|u^*\|_\infty \leq a$, $\|\sum_{j=1}^{\infty} \mathbf{E}_{j-1} u_j^2\|_\infty \leq b^2$ *for some* $a > 0, b > 0$. *Then for all* $r \geq 0$,

$$\begin{aligned} \mathbf{P}(f_+^* \geq r) &\leq \exp\left[\frac{r}{a} - \left(\frac{r}{a} + \frac{b^2}{a^2} \right) \ln\left(1 + \frac{ra}{b^2} \right) \right] \\ &\leq \left(\frac{eb^2}{ra} \right)^{r/a}. \end{aligned}$$

PROOF. The proof differs from that of Theorem 3.4 only in that we use Theorem 8.5 in place of Theorem 3.1. □

If, instead of Theorem 8.5, we had used the "naive" estimate $|d_j| \leq |u_j| + |\mathbf{E}_{j-1}u_j|$ and, say, Theorem 3.1, we would have hardly been able to obtain inequalities better than ones like those in Theorems 8.6 and 8.7, but with $2\Gamma, 2B$,

$2a$, and $2b$ in place of Γ, B, a and b, respectively. The gain provided, for example, by Theorem 8.6 is quite significant. If, say,

$$\exp\left(-\frac{r^2}{(2B)^2 + 2B\sqrt{(2B)^2 + 2(2\Gamma)r}}\right) = 10^{-1},$$

then the bound given in Theorem 8.6 varies from $10^{-2\sqrt{2}} < (1.5) \times 10^{-3}$ (when Γr is much greater than B^2) to 10^{-4} (when B^2 is much greater than Γr).

Acknowledgments. I am grateful to Donald Burkholder and Walter Phillip for their support and stimulating conversations and to Pawel Hitczenko, who acquainted me with his latest results referred to here. It is a pleasure for me to thank William Johnson and Joel Zinn, organizers of the workshop "Linear Analysis and Probability" in the Department of Mathematics, Texas A&M University, for the great opportunity to discuss the results of this paper. I am pleased to thank the Associate Editor and the referee for their consideration.

REFERENCES

BENNETT, G. (1962). Probability inequalities for the sum of independent random variables. *J. Amer. Statist. Assoc.* **57** 33–45.

BERGER, E. (1991). Majorization, exponential inequalities and almost sure behavior of vector-valued random variables. *Ann. Probab.* **19** 1206–1226.

BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

BURKHOLDER, D. L. (1966). Martingale transforms. *Ann Math. Statist.* **37** 1494–1504.

BURKHOLDER, D. L. (1973). Distribution function inequalities for martingales. *Ann. Probab.* **1** 19–42.

BURKHOLDER, D. L. (1991). Explorations in martingale theory and its applications. *Ecde d'Eté de Probabilités de Saint-Flour XIX. Lecture Notes in Math.* **1464** 1–66. Springer, Berlin.

CHUNG, K. L. (1951). The strong law of large numbers. *Proc. Second Berkeley Symp. Math. Statist. Probab.* 341–352. Univ. California Press, Berkley.

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

DE ACOSTA, A. (1981). Inequalities for B-valued random vectors with applications to the strong law of large numbers. *Ann. Probab.* **9** 157–161.

DHARMADHIKARI, S. W., FABIAN, V. and JOGDEO, K. (1968). Bounds on the moments of martingales. *Ann. Math. Statist.* **39** 1719–1723.

DHARMADHIKARI, S. W. and JOGDEO, K. (1969). Bounds on moments of certain random variables. *Ann Math. Statist.* **40** 1506–1508.

HAAGERUP, U. (1982). The best constants in the Khintchine inequality. *Studia Math.* **70** 232–283.

HITCZENKO, P. (1990a). Best constants in martingale version of Rosenthal's inequality. *Ann. Probab.* **18** 1656–1668.

HITCZENKO, P. (1990b). Upper bounds for the L_p-norms of martingales. *Probab. Theory Related Fields* **86** 225–237.

HITCZENKO, P. (1991). Private Communication.

HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30.

HOFFMANN-JØRGENSEN, J. (1974). Sums of independent Banach space valued random variables. *Studia Math.* **52** 159–186.

JOHNSON, W. B., SCHECHTMAN, G. and ZINN, J. (1985). Best constants in moment inequalities for linear combination of independent and exchangeable random variables. *Ann. Probab.* **13** 234–253.

KALLENBERG, O. and SZTENCEL, R. (1991). Some dimension-free features of vector-valued martingales. *Probab. Theory Related Fields* **88** 215–247.

KWAPIEŃ, S. and SZULGA, J. (1991). Hypercontraction methods in moment inequalities for series of independent random variables in normed spaces. *Ann. Probab.* **19** 369–379.

MARCINKIEWICZ, J. and ZYGMUND, A. (1937). Sur les fonctions independantes. *Fund. Math.* **89** 60–90.

NAGAEV, S. V. and PINELIS, I. F. (1977). Some inequalities for the distributions of sums of independent random variables. *Theory Probab. Appl.* **22** 248–256.

PINELIS, I. F. (1974). On some inequalities for the distributions of sums of independent random variables. Diploma work, Novosibirsk State Univ. (in Russian).

PINELIS, I. F. (1978). Distributions of sums of independent random variables with values in a Banach space. *Theory Probab. Appl.* **23** 608–615.

PINELIS, I. F. (1980). Estimates of moments of infinite-dimensional martingales. *Math. Notes* **27** 459–462.

PINELIS, I. F. (1990). Inequalities for the distributions of sums of independent random vectors and their applications to the density estimation. *Theory Probab. Appl.* **35** 500–512.

PINELIS, I. (1992). An approach to inequalities for the distributions of infinite-dimensional martingales. In *Probability in Banach Spaces. Proceedings of the Eighth International Conference* **8** 128–134. Birkhäuser, Boston.

PINELIS, I. (1994). On a distribution inequality for sums of independent random vectors. *Statist. Probab. Lett.* **19** 97–99.

PINELIS, I. F. and SAKHANENKO, A. I. (1985). Remarks on inequalities for large deviation probabilities. *Theory Probab. Appl.* **30** 143–148.

PINELIS, I. F. and UTEV. S. A. (1984). Estimates of the moments of sums of independent random variables. *Theory Probab. Appl.* **29** 574–577.

PINELIS, I. F. and UTEV, S. A. (1989). Exact exponential bounds for sums of independent random variables. *Theory Probab. Appl.* **34** 340–346.

PISIER, G. (1975). Martingales with values in uniformly convex spaces. *Israel J. Math.* **20** 326–350.

ROSÉN, B. (1970). On bounds on the central moments of even order of a sum of independent random variables. *Ann. Math. Statist.* **41** 1074–1077.

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.

SAZONOV, V. V. (1974). On the estimation of moments of sums of independent random variables. *Theory Probab. Appl.* **19** 371–374.

TALAGRAND, M. (1989). Isoperimetry and integrability of the sum of independent Banach-space valued random variables. *Ann. Probab.* **17** 1546–1570.

UTEV, S. A. (1985). Extremal problems in moment inequalities. In *Limit Theorems in Probability Theory. Proceedings of the Institute of Mathematics* **5** 56–75. Nauka, Novosibirsk (in Russian).

WANG, M. (1991). Sharp square function inequalities for conditionally symmetric martingales. *Trans. Amer. Math. Soc.* **328** 393–419.

WHITTLE, P. (1960). Bounds for the moments of linear and quadratic forms in independent random variables. *Theory Probab. Appl.* **5** 302–305.

YURINKSII, V. V. (1974). Exponential bounds for large deviations. *Theory Probab. Appl.* **19** 154–155.

ZOLOTAREV, V. M. (1977). Ideal metrics in the problem of approximating distributions of sums of independent random variables. *Theory Probab. Appl.* **22** 433–449.

DEPARTMENT OF MATHEMATICAL SCIENCES
MICHIGAN TECHNOLOGICAL UNIVERSITY
1400 TOWNSEND DRIVE
HOUGHTON, MICHIGAN 49931