The Annals of Probability
1979, Vol. 7, No. 5, 858–863

A SHARP INEQUALITY FOR MARTINGALE TRANSFORMS[1]

BY D. L. BURKHOLDER

University of Illinois

If g is the transform of a martingale f under a predictable sequence v uniformly bounded in absolute value by 1, then

$$\lambda P(g^* \geq \lambda) \leq 2\|f\|_1, \qquad \lambda > 0,$$

and this inequality is sharp.

1. Introduction. If $d = (d_1, d_2, \cdots)$ is a martingale difference sequence and $\varepsilon_1, \varepsilon_2, \cdots$ are numbers in $\{-1, 1\}$, then

$$P(|\Sigma_{k=1}^n \varepsilon_k d_k| \geq \lambda) \leq c\|\Sigma_{k=1}^n d_k\|_1 \tag{1}$$

where c is some absolute constant. Applications of this inequality and its natural extensions abound. For example, it leads immediately, by a simple interpolation and duality argument, to

$$\|\Sigma_{k=1}^n \varepsilon_k d_k\|_p \leq c_p\|\Sigma_{k=1}^n d_k\|_p, \qquad 1 < p < \infty, \tag{2}$$

which implies that any martingale difference sequence in L^p is an unconditional basis for its closed linear span. In turn, inequality (2) gives at once, by Khintchin's inequality, the two-sided L^p inequality for the martingale square function. For further details and discussion, see [1].

Our main goal here is to give a new proof of (1), a proof that throws additional light on the inequality by yielding the best constant.

Let $(\Omega, \mathcal{A}, P)$ be a probability space and $\mathcal{A}_0, \mathcal{A}_1, \cdots$ a nondecreasing sequence of sub-σ-fields of $\mathcal{A}$. Let $f = (f_1, f_2, \cdots)$ be a martingale with difference sequence $d = (d_1, d_2, \cdots) : f_n = \Sigma_{k=1}^n d_k$ where $d_k : \Omega \to \mathbb{R}$ is integrable and $\mathcal{A}_k$–measurable with $E(d_{k+1}|\mathcal{A}_k) = 0$, $k \geq 1$. Let $v = (v_1, v_2, \cdots)$ be a predictable sequence: $v_k : \Omega \to \mathbb{R}$ is $\mathcal{A}_{k-1}$–measurable, $k \geq 1$. Then $g = (g_1, g_2, \cdots)$, defined by $g_n = \Sigma_{k=1}^n v_k d_k$, is the transform of the martingale f under v. The L^1-norm of f is $\|f\|_1 = \sup_n\|f\|_1$ and the maximal function of g is defined by $g^*(\omega) = \sup_n|g_n(\omega)|$. The following extends (1).

THEOREM 1. *Suppose that g is the transform of a martingale f under a predictable sequence v uniformly bounded in absolute value by* 1. *Then*

$$\lambda P(g^* \geq \lambda) \leq 2\|f\|_1, \qquad \lambda > 0. \tag{3}$$

Except for the constant, this is a special case of Theorem 6 of [1], which was later extended by Davis [4]. Other proofs of (3) with the number 2 replaced by some

Received June 23, 1978.

[1]This work was supported in part by a grant from the National Science Foundation.

AMS 1970 *subject classifications.* Primary 60G45, 60H05.

Key words and phrases. Martingale, martingale transform, maximal function, square function, Brownian motion, Itô integral.

larger constant may be found in Gundy [10], Neveu [14], and Rao [15]. Also, see Meyer [12] and, for the special case of Haar series, Gaposhkin [7]. Each of these proofs has its own advantages but none can yield the best constant.

Our method here is to prove first a somewhat analogous inequality for the Itô integral and then to obtain (3) by Skorohod embedding.

2. An inequality for the Itô integral. Let $B = \{B_t, 0 \leqslant t < \infty\}$ be a standard Brownian motion in $\mathbb{R}$ starting at 0. Consider the local martingale X, with continuous sample functions, defined by the Itô integral

$$X_t = X_0 + \textstyle\int_0^t \varphi\, dB, \qquad t \geqslant 0. \tag{4}$$

Here $X_0 \in \mathbb{R}$ and the nonanticipating functional $\varphi : [0, \infty) \times \Omega \to \mathbb{R}$ satisfies $P(\int_0^t \varphi^2\, ds < \infty, t \geqslant 0) = 1$. (For background on the Itô integral, see [11].) Let Y be defined similarly by

$$Y_t = Y_0 + \textstyle\int_0^t \psi\, dB, \qquad t \geqslant 0. \tag{5}$$

Using the notation $S_t(X) = (X_0^2 + \int_0^t \varphi^2\, ds)^{\frac{1}{2}}$ together with notation analogous to that introduced in Section 1, we have the following inequality between X and Y.

LEMMA 1. *If $S_t(Y) \leqslant S_t(X)$ for all $t \geqslant 0$, then*

$$\lambda P(Y^* \geqslant \lambda) \leqslant 2\|X\|_1, \qquad \lambda > 0. \tag{6}$$

The assumption holds, for example, if $|Y_0| \leqslant |X_0|$ and $|\psi| \leqslant |\varphi|$.

PROOF. Let $\|X\|_2 = \sup_{t \geqslant 0} \|X_t\|_2$ and $S(X) = S_\infty(X)$. Then

$$\|X\|_2 = \|S(X)\|_2. \tag{7}$$

Furthermore, if μ is a stopping time of B and X^μ denotes X stopped at μ ($X_{\mu \wedge t}$ is the tth term), then

$$(X^\mu)_t = X_0 + \textstyle\int_0^t \varphi I\, dB \tag{8}$$

where $I(s, \cdot)$ is the indicator of the event $\{\mu \geqslant s\}$. (Both (7) and (8) follow easily from the methods and results of Section 2.3 of [11].) So (7) also holds for X^μ, $S(X^\mu) = S_\mu(X)$, and, by the assumption of the lemma,

$$\|Y^\mu\|_2 \leqslant \|X^\mu\|_2. \tag{9}$$

We shall now prove (6) using (9). Since $|X_0| \leqslant \|X\|_1$, inequality (6) holds trivially for $\lambda \leqslant |X_0|$. Therefore, assume that $\lambda > |X_0|$ and consider the stopping time μ defined by

$$\mu(\omega) = \inf\{t : |X_t(\omega)| > \lambda\}.$$

Note that $\{X^* \leqslant \lambda\} = \{\mu = \infty\}$ and, by the sample-function continuity of X, the stopped process X^μ is uniformly bounded by λ. So, by the weak $-L^2$ inequality for the martingale maximal function and (9),

$$\begin{aligned}\lambda^2 P(Y^* > \lambda, X^* \leqslant \lambda) &\leqslant \lambda^2 P((Y^\mu)^* > \lambda)\\ &\leqslant \|Y^\mu\|_2^2 \leqslant \|X^\mu\|_2^2\\ &\leqslant \|X^\mu\|_\infty \|X^\mu\|_1 \leqslant \lambda\|X\|_1.\end{aligned}$$

Therefore,

$$\lambda P(Y^* > \lambda) \leqslant \lambda P(Y^* > \lambda, X^* \leqslant \lambda) + \lambda P(X^* > \lambda)$$
$$\leqslant \|X\|_1 + \|X\|_1,$$

which implies (6).

3. Proof of Theorem 1. Let n be a positive integer and $g_n^* = \sup_{k \leqslant n} |g_k|$. It is enough to show that

$$\lambda P(g_n^* > \lambda) \leqslant 2\|f_n\|_1. \tag{10}$$

First consider the following special case. *Let* $H_k : \mathbb{R}^k \to [-1, 1]$ *be continuous and*

$$v_k = H_k(f_0, \cdots, f_{k-1}), \qquad 1 \leqslant k \leqslant n,$$

where $f_0 = 0$. *Then* (10) *holds*. Since $(f_1, \cdots, f_n)$ is the almost everywhere limit of the sequence of martingales $(f_{j1}, \cdots, f_{jn})$ defined by $f_{jk} = E[j \wedge (-j \vee f_n)|\mathcal{Q}_k]$ and $\|f_{jn}\|_1 \to \|f_n\|_1$, it is enough to prove the special case under the additional assumption that $f_n \in L^2$. Then, by the Skorohod embedding thoerem (a convenient reference is [6]), there are integrable stopping times $\tau_1 \leqslant \cdots \leqslant \tau_n$ of a Brownian motion B, which may be assumed to be defined also on $(\Omega, \mathcal{Q}, P)$, such that $(B_{\tau_1}, \cdots, B_{\tau_n})$ has the same distribution as $(f_1 - Ef_1, \cdots, f_n - Ef_1)$. This implies that $(f_1, \cdots, f_n)$ has the same distribution as $(X_{\tau_1}, \cdots, X_{\tau_n})$ and $(g_1, \cdots, g_n)$ has the same distribution as $(Y_{\tau_1}, \cdots, Y_{\tau_n})$ where X and Y are defined by (4) and (5) with $X_0 = Ef_1$, $Y_0 = Eg_1$, $\varphi(s, \cdot)$ the indicator function $I(\tau_n \geqslant s)$, and

$$\psi(s, \cdot) = \Sigma_{k=1}^n H_k(X_{\tau_0}, \cdots, X_{\tau_{k-1}}) I(\tau_{k-1} < s \leqslant \tau_k)$$

where $\tau_0 = -1$ and $X_{-1} = 0$. The assumption of Lemma 1 is satisfied: $|Y_0| = |v_1 E\, d_1| \leqslant |Ed_1| = |X_0|$ and $|\psi| \leqslant \varphi$. Furthermore, $\|X\|_1 = \|X_{\tau_n}\|_1$ since the integrability of τ_n implies that X is L^2-bounded, hence uniformly integrable. Accordingly,

$$\lambda P(g_n^* > \lambda) = \lambda P(\sup_{k \leqslant n} |Y_{\tau_k}| > \lambda)$$
$$\leqslant \lambda P(Y^* > \lambda) \leqslant 2\|X\|_1$$
$$= 2\|X_{\tau_n}\|_1 = 2\|f_n\|_1,$$

which completes the proof of the special case.

To finish the proof of (10), we now construct a new martingale F and a transform G to which the special case applies. We may assume that $(r_0, r_1, \cdots)$ is an independent sequence on $(\Omega, \mathcal{Q}, P)$ satisfying $P(r_k = -1) = P(r_k = 1) = \frac{1}{2}$, $k \geqslant 0$, and such that $(r_0, r_1, \cdots)$ is independent of $\mathcal{Q}_\infty = \vee_{k=1}^\infty \mathcal{Q}_k$. Define the difference sequence D of F by

$$D_{3k-2} = \varepsilon r_{3k-2} v_k^+, \qquad D_{3k-1} = \varepsilon r_{3k-1} v_k^-, \qquad D_{3k} = r_0\, d_k$$

where $v_k^+ = v_k \vee 0$, $v_k^- = -(v_k \wedge 0)$, and $\varepsilon > 0$. Then F is a martingale (relative to the sequence of σ-fields generated by F) and $F_{3n} = \Sigma_{k=1}^{3n} D_k = r_0 f_n + R_n$ where

$|R_n| \leqslant \varepsilon n$. Let G be the transform of F under V defined by

$$V_{3k-2} = V_{3k-1} = 0, \qquad V_{3k} = v_k.$$

Then $G_{3n} = \Sigma_{k=1}^{3n} V_k D_k = r_0 g_n$. Since V_{3k} may be written in the form

$$V_{3k} = H\left(|\varepsilon^{-1} D_{3k-2}| - |\varepsilon^{-1} D_{3k-1}|\right),$$

where $H(x) = 1 \wedge (-1 \vee x)$, and this is a continuous function of $F_0, \cdots, F_{3k-1}$ into $[-1, 1]$, the above special case gives

$$\begin{aligned} \lambda P(g_n^* > \lambda) &= \lambda P(G_{3n}^* > \lambda) \\ &\leqslant 2\|F_{3n}\|_1 \\ &\leqslant 2\|f_n\|_1 + 2\varepsilon n. \end{aligned}$$

Now let $\varepsilon \to 0$ to obtain (10).

4. Sharpness of the above inequalities. Consider the following simple example pointed out to us by Leonard Dor. Let P be Lebesgue measure on $[0, 1)$. Let $d_1 = 1$ on $[0, 1)$, $d_2 = 1$ on $[0, \frac{1}{2})$, $d_2 = -1$ on $[\frac{1}{2}, 1)$, $d_3 = 2$ on $[0, \frac{1}{4})$, $d_3 = -2$ on $[\frac{1}{4}, \frac{1}{2})$, and $d_3 = 0$ on $[\frac{1}{2}, 1)$; these are the first three Haar functions appropriately normalized. Then $\|d_1 + d_2 + d_3\|_1 = 1$ and $|d_1 - d_2 + d_3| \equiv 2$ so that

$$2P(|d_1 - d_2 + d_3| \geqslant 2) = 2\|d_1 + d_2 + d_3\|_1.$$

This shows that the inequalities (1), with $c = 2$, and (3) are sharp.

An analogous example shows that (6) is sharp. Let $\tau_1 = \inf\{t : |B_t| = 1\}$, $\tau_2 = \inf\{t > \tau_1 : |B_t - B_{\tau_1}| = 1\}$, and $\tau_3 = \inf\{t > \tau_2 : |B_t - B_{\tau_2}| = 2\}$. Define X and Y by (4) and (5) where $X_0 = Y_0 = 0$ and

$$\varphi(s, \cdot) = I(\tau_1 \geqslant s) + I(\tau_1 < s \leqslant \tau_2) + I(B_{\tau_2} \neq 0) I(\tau_2 < s \leqslant \tau_3),$$

$$\psi(s, \cdot) = I(\tau_1 \geqslant s) - I(\tau_1 < s \leqslant \tau_2) + I(B_{\tau_2} \neq 0) I(\tau_2 < s \leqslant \tau_3).$$

Then $P(Y^* = 2) = 1$ and $\|X\|_1 = 1$ so that

$$2P(Y^* \geqslant 2) = 2\|X\|_1$$

showing that (6) is sharp.

5. Remarks. (a) The above methods also yield a smaller constant than any heretofore known in the weak $-L^1$ inequality for the martingale square function. If f is a martingale with difference sequence d and $S(f) = (\Sigma_{k=1}^{\infty} d_k^2)^{\frac{1}{2}}$, then

$$\lambda P(S(f) \geqslant \lambda) \leqslant 2\|f\|_1, \qquad \lambda > 0. \tag{11}$$

To prove (11), we let X and $\tau_1, \cdots, \tau_n$ be as in Section 3. To define Y, we let $Y_0 = 0$,

$$\tau_{n+1} = \inf\{t > \tau_n : |B_t - B_{\tau_n}| = 1\},$$

$$V_t = \left[(X_0 + B_{\tau_1 \wedge t})^2 + \Sigma_{k=2}^{n}(B_{\tau_k \wedge t} - B_{\tau_{k-1} \wedge t})^2\right]^{\frac{1}{2}},$$

and $\psi(s, \cdot) = I(\tau_n < s \leqslant \tau_{n+1}) V_{\tau_n}$. If μ is the stopping time defined in Section 2 (or any other stopping time of B), then

$$\left(B_{\tau_{n+1}\wedge\mu} - B_{\tau_n\wedge\mu}\right)^2 V_{\tau_n}^2 \leqslant V_\mu^2.$$

(If $\mu \leqslant \tau_n$, the left-hand side is 0; if $\mu > \tau_n$, then $V_\mu = V_{\tau_n}$.) Therefore, by an elementary calculation,

$$\|Y^\mu\|_2^2 = E\left(B_{\tau_{n+1}\wedge\mu} - B_{\tau_n\wedge\mu}\right)^2 V_{\tau_n}^2$$
$$\leqslant EV^2{}_\mu = \|X^\mu\|_2^2.$$

Also, note that $Y^* = V_{\tau_n}$, which has the same distribution as $S_n(f) = (\Sigma_{k=1}^n d_k^2)^{\frac{1}{2}}$, and, as in Section 3, $\|X\|_1 = \|f_n\|_1$. So using the fact that (6) follows from (9), we obtain $\lambda P(S_n(f) > \lambda) \leqslant 2\|f_n\|_1$, which gives (11).

Apart from the constant, (11) was proved in [1] and the above proof is similar to the original proof in its main concept. Other approaches may be found in [10], [14], [15], [8], and [2].

Suppose that f is a Rademacher martingale: $|d_k| \equiv a_k \in \mathbb{R}$, $k \geqslant 1$. Then $S_n(f) \equiv (\Sigma_{k=1}^n a_k^2)^{\frac{1}{2}}$ so that $\lambda P(S_n(f) > \lambda) \leqslant S_n(f)$ and, by a result of Szarek [16], $S_n(f) \leqslant 2^{\frac{1}{2}}\|f_n\|_1$. Therefore, (11) holds here with 2 replaced by $2^{\frac{1}{2}}$ and, it is easy to see, no smaller number suffices. Our guess is that $2^{\frac{1}{2}}$ is also the best constant for the class of all martingales.

(b) An analysis of the proof of Lemma 1 shows that (6) holds if Y^* is replaced by $X^* \vee Y^*$. Therefore, (3) holds if g^* is replaced by $f^* \vee g^*$ and (11) holds if $S(f)$ is replaced by $S(f) \vee f^*$.

(c) For some related examples of the interaction between discrete-time and continuous-time martingale inequalities, see [13], [3], [9], and [5].

REFERENCES

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

[2] BURKHOLDER, D. L. (1973). Distribution function inequalities for martingales. *Ann. Probability* **1** 19–42.

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

[4] DAVIS, BURGESS (1969). A comparison test for martingale inequalities. *Ann. Math. Statist.* **40** 505–508.

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

[6] DUBINS, LESTER E. (1968). On a theorem of Skorohod. *Ann. Math. Statist.* **39** 2094–2097.

[7] GAPOSHKIN, V. F. (1974). The Haar system as an unconditional basis in L_p [0, 1]. *Math. Notes* **15** 108–111.

[8] GARSIA, ADRIANO M. (1973). *Martingale Inequalities: Seminar Notes on Recent Progress*. Benjamin, Reading, Mass.

[9] GORDON, LOUIS (1972). An equivalent to the martingale square function inequality. *Ann. Math. Statist.* **43** 1927–1934.

[10] GUNDY, RICHARD F. (1968). A decomposition for L^1-bounded martingales. *Ann. Math. Statist.* **39** 134–138.

[11] MCKEAN, H. P. (1969). *Stochastic Integrals*. Academic Press, New York.

[12] MEYER, PAUL-ANDRÉ (1972). Martingales and stochastic integrals I. *Lecture Notes in Mathematics* **284**. Springer-Verlag, Berlin.

[13] MILLAR, P. WARWICK (1968). Martingale integrals. *Trans Amer. Math. Soc.* **133** 145–166.
[14] NEVEU, J. (1972). *Martingales à temps discret*. Masson et Cie, Paris.
[15] RAO, MURALI (1972). Doob decomposition and Burkholder inequalities. Séminaire de Probabilités VI (Univ. Strasbourg, 1970–71) 198–201. Lecture Notes in Mathematics **258**. Springer-Verlag, Berlin.
[16] SZAREK, S. J. (1976). On the best constants in the Khintchin inequality. *Studia Math.* **58** 197–208.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF ILLINOIS
URBANA, ILLINOIS 61801