The Annals of Probability
1988, Vol. 16, No. 2, 840–853

A BEST POSSIBLE IMPROVEMENT OF WALD'S EQUATION

BY MICHAEL J. KLASS[1]

University of California, Berkeley

Let $X_1, X_2, \dots$ be independent random elements taking values in a Banach space $(B, \|\cdot\|)$ and having partial sums $S_n = X_1 + \cdots + X_n$. Let $\alpha > 0$ and let Φ: $[0, \infty)$ be a nondecreasing continuous function such that $\Phi(0) = 0$ and $\Phi(cx) \le c^\alpha \Phi(x)$ for all $c \ge 2$, $x \ge 0$. Put $a_n^* = E \max_{1 \le k \le n} \Phi(\|S_k\|)$. Let T be any (possibly randomized) stopping time w.r.t. $\{S_n\}$. We prove that $E \max_{1 \le n \le T} \Phi(\|S_n\|) \le 20(18^\alpha) E a_T^*$. If $\{S_n\}$ is a mean-zero B-valued martingale and $\lim_{n\to\infty} E\|S_{T \wedge n}\| < \infty$, it is shown that $L \equiv \lim_{n\to\infty} E S_n I(T > n)$ always exists and $ES_T = -L$, so that $ES_T = 0$ iff $L = 0$. Let $s_n = E\|S_n\|$ and $s_n^* = E \max_{1 \le k \le n} \|S_k\|$. As a consequence of these facts it follows that if $\{X_n\}$ are independent and have mean zero, then $E\|S_T\| < \infty$ and $ES_T = 0$ whenever $Es_T^* < \infty$. In the mean-zero case $s_n^* \le 4 s_n$; and so, in fact, $Es_T < \infty$ implies $ES_T = 0$. This constitutes a best possible improvement of Wald's equation.

1. Introduction and summary. Let $X, X_1, X_2, \dots$ be any sequence of independent identically distributed (i.i.d.) random variables with partial sums $S_n = X_1 + \cdots + X_n$. Let T be any (possibly randomized) stopping time with respect to $\{X_n\}$. Wald's (1945) famous equation, extended by Blackwell (1946) to include all X-distributions with finite first moment (which we take to be zero), states that

$$ES_T = 0 \tag{1.1}$$

provided

$$ET < \infty. \tag{1.2}$$

Burkholder and Gundy (1970) and Gordon [in some unpublished work referred to in Chung (1974), page 343], refined this result, showing that whenever $EX^2 < \infty$, the weaker condition $ET^{1/2} < \infty$ implies $ES_T = 0$.

Chow, Robbins and Siegmund (1971) (CRS) then extended Burkholder and Gundy's result to variates having some moment between one and two. Specifically, they proved that if $E|X|^\alpha < \infty$ and $ET^{1/\alpha} < \infty$ for some $1 \le \alpha \le 2$, then $ES_T = 0$.

Addressing the issue in utmost generality, we ask:

PROBLEM. For any given nonconstant mean-zero X-distribution, what is the weakest condition on the tail behavior of a stopping time T which will ensure that $ES_T = 0$?

Received April 1986; revised July 1987.
[1]Supported in part by NSF Grant MCS-83-01793.
AMS 1980 *subject classifications*. Primary 60G40; secondary 60G50, 60E15.
Key words and phrases. Expectations of functions of normed partial sums, random numbers of summands, Wald's equation.

Let $a_n = E|S_n|$ and suppose T is independent of $\{X_n\}$. If we are to have $ES_T = 0$, then technically we require $E|S_T| < \infty$, in which case $Ea_T < \infty$. Hence, to obtain $ES_T = 0$ for all stopping times T having common marginal distribution, $Ea_T < \infty$ is a minimal necessary condition on that distribution. We prove that this condition is also sufficient, a result we now state formally.

THEOREM 1.1. *Let $X, X_1, X_2, \dots$ be i.i.d. mean-zero random variables. Let T be any stopping time w.r.t. $\{X_n\}$. Then*

$$ES_T = 0, \quad \textit{whenever } Ea_T < \infty. \tag{1.3}$$

The quantity a_n can be approximated directly from the X-distribution. Let $K(y) \equiv 0$ if $X \equiv 0$ a.s. Otherwise, let $K(y)$ be the unique positive real satisfying (for $y > 0$)

$$yE\left\{\left(\frac{X}{K(y)}\right)^2 \wedge \frac{|X|}{K(y)}\right\} = 1. \tag{1.4}$$

In Klass (1980) it was shown that

$$(1 - O(n^{-1/2}))0.673^+ K(n) \le E|S_n| \le 2K(n). \tag{1.5}$$

Hence, (1.3) may be re-expressed in terms of T and the X-distribution as

$$ES_T = 0, \quad \text{whenever } EK(T) < \infty. \tag{1.6}$$

Owing to the fact [proved in Klass (1973)] that $E|S_n| = o(n^{1/\alpha})$ [or alternatively $K(n) = o(n^{1/\alpha})$] whenever $E|X|^\alpha < \infty$ and $1 \le \alpha < 2$, it is clear that Theorem 1.1 strictly refines that of CRS. Moreover, these results [(1.3)–(1.6)] extend to the nonidentically distributed case and (1.3) extends to Banach space as well. Furthermore, if $\{S_n\}$ is a mean-zero martingale and T any stopping time such that $\lim_{n\to\infty} E|S_{T\wedge n}| < \infty$, then (see Theorem 4.1) $ES_T = 0$ iff $\lim_{n\to\infty} ES_n I(T > n) = 0$, a result which improves upon Theorem 2.3 of CRS.

Plainly, the sequence $\{S_{T\wedge n}\}$ is uniformly integrable whenever

$$E \max_{1\le n\le T} |S_n| < \infty. \tag{1.7}$$

In such instances,

$$ES_T = \lim_{n\to\infty} ES_{T\wedge n} = 0.$$

(Alternatively, just invoke dominated convergence.) Therefore, in the interest of obtaining $ES_T = 0$, it suffices to establish (1.7). In Section 2 a somewhat elementary proof is presented which works for i.i.d. random variables. To extend the result to random elements (and not necessarily identically distributed ones at that) requires a different approach. The key tool is the following fundamental lemma used by Burkholder and Gundy (1970) exploiting *dependence* between random variables.

LEMMA 1.2 [Burkholder and Gundy (1970) and Burkholder (1973)]. *Let U and V be nonnegative random variables. Suppose there exist positive reals β, δ, γ*

such that $\beta^{-1} - \gamma > 0$ *and*

$$P(U \geq \beta y, V \leq \delta y) \leq \gamma P(U \geq y), \quad \text{for all } y > 0. \tag{1.8}$$

Then

$$EU \leq \left(\beta^{-1} - \gamma\right)^{-1}\delta^{-1}EV. \tag{1.9}$$

Since the proof is short, we will include it here.

PROOF. For any $y > 0$,

$$\begin{aligned} P(U \geq \beta y) &= P(U \geq \beta y, V \leq \delta y) + P(U \geq \beta y, V > \delta y) \\ &\leq \gamma P(U \geq y) + P(V \geq \delta y). \end{aligned}$$

Integrating with respect to y, $E\beta^{-1}U \leq \gamma EU + \delta^{-1}EV$.
Solving for EU gives (1.9). □

Since Lemma 1.2 is to be used to derive a sufficient condition for (1.7) (i.e., to upper-bound $E\max_{1\leq n\leq T}|S_n|$), it will be no more difficult to upper-bound $E\max_{1\leq n\leq T}\Phi(\|S_n\|)$ for increasing functions of polynomial growth. In this regard we prove

THEOREM 1.3. *Fix any* $\alpha > 0$. *Let* Φ: $[0, \infty)$ *be any nondecreasing continuous function such that* $\Phi(0) = 0$ *and* $\Phi(cx) \leq c^{\alpha}\Phi(x)$ *for all* $c \geq 2$ *and* $x \geq 0$. *Let* $X_1, X_2, \ldots$ *be independent random elements taking values in a Banach space* $(B, \|\cdot\|)$ *and put* $S_n = X_1 + \cdots + X_n$. *Let* T *be any (possibly randomized) stopping time with respect to* $\{X_n\}$. *Then there exists a universal constant* $0 < C^*(\alpha) < \infty$ *depending only on* α *such that*

$$E\max_{1\leq n\leq T}\Phi(\|S_n\|) \leq C^*(\alpha)Ea_T^*, \tag{1.10}$$

where

$$a_n^* = E\max_{1\leq k\leq n}\Phi(\|S_k\|). \tag{1.11}$$

Equivalently, letting τ *be independent of* $\{X_n\}$ *and distributed as* T,

$$E\max_{1\leq n\leq T}\Phi(\|S_n\|) \leq C^*(\alpha)E\max_{1\leq n\leq \tau}\Phi(\|S_n\|). \tag{1.12}$$

In a forthcoming work it is shown that there exists a universal constant $C_*(\alpha) > 0$ such that

$$C_*(\alpha)Ea_T^* \leq E\max_{1\leq n\leq T}\Phi(\|S_n\|). \tag{1.13}$$

The following corollary, due in part to a remark of de la Peña, is an immediate consequence of (1.10).

COROLLARY 1.4. *Let* $a_\infty^* = \lim_{n\to\infty}a_n^*$, *where* $a_n^* = E\max_{l\leq k\leq n}\Phi(\|S_k\|)$. *For* $c > 0$, *let*

$$T_c = \begin{cases} \text{first } n \geq 1\colon \Phi(\|S_n\|) > ca_n^*, \\ \infty, \quad \text{if no such } n \text{ exists}. \end{cases} \tag{1.14}$$

Then for each $c \geq C^*(\alpha)$,

$$Ea^*_{T_c} = \infty, \tag{1.15}$$

whenever

$$P(T_c < \infty) = 1. \tag{1.16}$$

Combining (1.14) *and* (1.15) *it also follows that for any* $c \geq C^*(\alpha)$,

$$E \max_{1 \leq n \leq T_c} \Phi(\|S_n\|) = \infty \tag{1.17}$$

whenever (1.16) *holds.*

Employing Theorem 1.3 together with a couple of additional ideas, the paper concludes (in Section 4) with our most comprehensive extension of Wald's equation, stated below.

THEOREM 1.5. *Let* $(B, \|\cdot\|)$ *be any Banach space. Let* $X_1, X_2, \dots,$ *be independent mean-zero random elements taking values in* B, $S_n = X_1 + \cdots + X_n$ *and* $a_n = E\|S_n\|$. *Let* T *be any (possibly randomized) stopping time with respect to* $\{X_n\}$. *Then*

$$ES_T = 0, \quad \text{if } Ea_T < \infty. \tag{1.18}$$

Moreover, if

$$\lim_{n \to \infty} E\|S_{T \wedge n}\| < \infty, \tag{1.19}$$

then

$$L \equiv \lim_{n \to \infty} ES_n I(T > n) \tag{1.20}$$

and ES_T *both exist and*

$$ES_T = 0, \quad \text{iff } \lim_{n \to \infty} ES_n I(T > n) = 0 \; (\text{since } ES_T = -L). \tag{1.21}$$

Even in the real-variables context, (1.19)–(1.21) strictly improves Theorem 2.3 of Chow, Robbins and Siegmund (1971) (see Example 4.5).

2. Wald's equation on $\mathbb{R}^1$. Let $Y_1, Y_2, \dots$ be (any) independent mean-zero random variables, n any integer ≥ 1 and τ any stopping time with respect to $\{Y_j\}$. Then there exist universal constants C_1 and C_2 (independent of $\{Y_j\}$, n, τ) such that

$$E\left(\sum_{j=1}^{n} \left(Y_j^2\right)\right)^{1/2} \leq C_1 E\left|\sum_{j=1}^{n} Y_j\right| \quad \text{[Marcinkiewicz and Zygmund (1938)]} \tag{2.1}$$

and

$$E \max_{1 \leq n \leq \tau}\left|\sum_{j=1}^{n} Y_j\right| \leq C_2 E\left(\sum_{j=1}^{\tau} \left(Y_j^2\right)\right)^{1/2} \quad \text{[Davis (1970)]}. \tag{2.2}$$

These two inequalities plus that of (1.5) make it possible to proceed along pleasingly elementary lines to establish the following result. Extension of Wald's equation (in the i.i.d. $\mathbb{R}^1$ setting) is a simple consequence.

THEOREM 2.1. *Let $X, X_1, X_2, \ldots$ be i.i.d. mean-zero random variables, $S_n = X_1 + \cdots + X_n$, and T any stopping time w.r.t. $\{X_n\}$. Define $K(\cdot)$ as in* (1.4). *Then there exists a universal constant $c < \infty$ (c is independent of X and T) such that*

$$E \max_{1 \le n \le T} |S_n| \le cEK(T). \tag{2.3}$$

REMARK 2.2. Inequality (2.3) upper-bounds $E \sup_{1 \le n \le T} |S_n|$ directly in terms of T and the underlying X-distribution. Moreover, the derivation of (2.3) is facilitated by the fact that $\{2^{-n/2}K(2^n)\}$ is nondecreasing in n. Because $a_n = E|S_n|$ does not quite enjoy this property, it is easier to work with $K(n)$. Nevertheless, in view of (1.5), it is clear that the conclusion of Theorem 2.1 could have been equivalently stated as

$$E \max_{1 \le n \le T} |S_n| \le c' E a_T, \tag{2.4}$$

for some universal constant $c' < \infty$.

PROOF OF THEOREM 2.1. Notice that $K(y)$ increases and so $y^{-1}K^2(y) = E(X^2 \wedge |X|K(y))$ is nondecreasing. Hence

$$y^{-1/2}K(y) \text{ is nondecreasing.} \tag{2.5}$$

Let

$$k_T = \min\{k: 2^k > T\} \tag{2.6}$$

and put

$$T^* = 2^{k_T} - 1. \tag{2.7}$$

By the concavity of the square-root function,

$$\begin{aligned} \left(\sum_{j=1}^{T^*} X_j^2\right)^{1/2} &= \left(\sum_{k=0}^{k_T - 1} \sum_{2^k \le j < 2^{k+1}} X_j^2\right)^{1/2} \\ &\le \sum_{k=0}^{k_T - 1} \left(\sum_{2^k \le j < 2^{k+1}} X_j^2\right)^{1/2} \\ &= \sum_{k=0}^{\infty} \left(\sum_{2^k \le j < 2^{k+1}} X_j^2\right)^{1/2} I(T^* \ge 2^k). \end{aligned} \tag{2.8}$$

Taking expectations and using in turn the inequalities of Davis (1970), (2.8), Marcinkiewiez and Zygmund (1938) and Klass (1973), we have

$$
\begin{aligned}
E \max_{1 \le n \le T} |S_n| &\le C_2 E\left(\sum_{j=1}^{T} X_j^2 \right)^{1/2} && \text{[by (2.2)]} \\
&\le C_2 E\left(\sum_{j=1}^{T^*} X_j^2 \right)^{1/2} && \text{(since } T^* \ge T\text{)} \\
&\le C_2 E \sum_{k=0}^{\infty} \left(\sum_{2^k \le j < 2^{k+1}} X_j^2 \right)^{1/2} I(T^* \ge 2^k) && \text{[by (2.8)]} \\
&= C_2 E \sum_{k=0}^{\infty} I(T^* \ge 2^k) E\left(\sum_{2^k \le j < 2^{k+1}} X_j^2 \right)^{1/2} && \text{(by independence)} \\
&\le C_1 C_2 E \sum_{k=0}^{\infty} I(T^* \ge 2^k) E|S_{2^k}| && \text{[by (2.1)]} \\
&\le 2C_1 C_2 E \sum_{k=0}^{\infty} I(T^* \ge 2^k) K(2^k) && \text{[by (1.5)]} \\
&= 2C_1 C_2 E \sum_{k=0}^{k_T - 1} (2^{-k/2} K(2^k)) 2^{k/2} \\
&\le 2C_1 C_2 E 2^{-(k_T - 1)/2} K(2^{k_T - 1}) \sum_{k=0}^{k_T - 1} 2^{k/2} && \text{[by (2.5)]} \\
&\le 2C_1 C_2 E K(2^{k_T - 1})(1 - 2^{-1/2})^{-1} \\
&\le 2C_1 C_2 (1 - 2^{-1/2})^{-1} EK(T) && \text{[since } K(y) \text{ increases]}.
\end{aligned}
$$

□

By the uniform integrability argument following (1.7), (1.6) is immediate and (1.3) holds by application of (1.5). Hence, we have

COROLLARY 2.3. *Under the assumptions of Theorem* 2.1,

$$ES_T = 0, \quad \text{if } EK(T) < \infty, \tag{2.9}$$

and, equivalently,

$$ES_T = 0, \quad \text{if } Ea_T < \infty. \tag{2.10}$$

These results will be extended in the next sections.

3. Upper-bounding $E\max_{1\le n\le T}\Phi(\|S_n\|)$. In this section Theorem 2.1 is generalized. Though somewhat more involved than that of Theorem 2.1, the proof given here is not lengthy. Moreover, it has both the virtue of proceeding from first principles and that of producing explicit constants.

In what follows, let (for $\alpha > 0$)

$$F_\alpha = \{\Phi\colon [0,\infty) \to [0,\infty) \text{ such that } \Phi(0) = 0, \tag{3.1}$$
$$\Phi(\cdot) \text{ is nondecreasing and continuous,}$$
$$\text{and } \Phi(cx) \le c^\alpha \Phi(x) \text{ for all } x \ge 0,\ c \ge 2\}.$$

THEOREM 3.1. *Fix $\alpha > 0$ and $\Phi \in F_\alpha$. Let $X_1, X_2, \dots$ be independent random elements taking values in a Banach space $(B, \|\cdot\|)$. Let $S_0 = 0$ and $S_n = X_1 + \cdots + X_n$ (for $n \ge 1$). Define*

$$a_n^* \equiv E\max_{1\le j\le n}\Phi(\|S_j\|). \tag{3.2}$$

Let T be any (possibly randomized) stopping time with respect to $\{X_j\}$. Then for any $\delta > 0$ and $\beta > 3^\alpha(1+\delta)$ such that $q \equiv \beta^{-1} - 6^\alpha\delta(\beta - 3^\alpha(1+\delta))^{-1} > 0$,

$$E\max_{1\le n\le T}\Phi(\|S_n\|) \le (q\delta)^{-1}(1 + 2^{\alpha+2})Ea_T^*. \tag{3.3}$$

Moreover, by letting $\beta = 3^\alpha(1+\delta)(1 - \sqrt{6^\alpha\delta})^{-1}$ and $\delta = (4\cdot 6^\alpha)^{-1}$, it can be seen that $\min_{\beta,\delta}(q\delta)^{-1} < 20(18^\alpha)$.

PROOF. For simplicity, let

$$S^*_{(m,n]} = \max_{m\le j\le n}\|S_j - S_m\| \tag{3.4}$$

and

$$S_n^* = S^*_{(0,n]}. \tag{3.5}$$

We intend to employ Lemma 1.2. Fix $y > 0$. Let

$$T_y = \begin{cases} \text{first } 1 \le m \le T\colon \Phi(S_m^*) \ge y, & \text{if such } m \text{ exists,} \\ \infty, & \text{otherwise.}\end{cases}$$

Let $d^* = \max_{1\le j\le T}\|X_j\|$ and $n^* = \sup\{n\colon a_n^* \le \delta y\}$. Note that for any $0 \le a \le b \le c$,

$$\Phi(a+b+c) \le \Phi(3c) \le 3^\alpha\Phi(c) \le 3^\alpha(\Phi(a) + \Phi(b) + \Phi(c)).$$

Hence,

$$\{\Phi(S_T^*) \ge \beta y,\ \Phi(d^*) \vee a_T^* \le \delta y\}$$

$$\subseteq \left\{3^\alpha\left(\Phi\left(S_{T_y-1}^*\right) + \Phi\left(\|X_{T_y}\|\right) + \Phi\left(S_{(T_y,T]}^*\right)\right) \ge \beta y,\ \Phi(d^*) \le \delta y,\ T \le n^*\right\}$$

$$\subseteq \left\{3^\alpha \Phi\left(S_{(T_y,T\wedge n^*]}^*\right) \ge y(\beta - 3^\alpha(1+\delta))\right\}.$$

The bounds on $\Phi(S_{(T_y-1)}^*)$, $\Phi(\|X_{T_y}\|)$ and T derive from the construction of the stopping time T_y and the constraints on $\Phi(d^*)$ and a_T^*. By Markov's inequality

$$P(\Phi(S_T^*) \ge \beta y,\ \Phi(d^*) \vee a_T^* \le \delta y)$$
$$\le 3^\alpha y^{-1}(\beta - 3^\alpha(1+\delta))^{-1} E\Phi\left(S_{(T_y,T\wedge n^*]}^*\right).$$

We must bound the latter expectation:

$$E\Phi\left(S_{(T_y,T\wedge n^*]}^*\right)$$
$$= \sum_{j=1}^{n^*-1} E\Phi(S_{(j,T\wedge n^*]}^*) I(T_y = j,\ T > j)$$
$$\le \sum_{j=1}^{n^*-1} E\Phi(S_{(j,n^*]}^*) I(T_y = j,\ T > j)$$
$$= \sum_{j=1}^{n^*-1} E\Phi(S_{(j,n^*]}^*) P(T_y = j,\ T > j) \quad \text{(by independence)}$$
$$\le \sum_{j=1}^{n^*-1} E\Phi(2S_{n^*}^*) P(T_y = j,\ T > j)$$
$$\le 2^\alpha E\Phi(S_{n^*}^*) P(T_y < T)$$
$$\le 2^\alpha \delta y P(\Phi(S_T^*) \ge y).$$

Letting $\gamma = 6^\alpha\delta(\beta - 3^\alpha(1+\delta))^{-1}$, Lemma 1.2 implies that

$$E\Phi(S_T^*) \le \delta^{-1}(\beta^{-1} - \gamma)^{-1} E(\Phi(d^*) \vee a_T^*).$$

Since $E(\Phi(d^*) \vee a_T^*) \le E\Phi(d^*) + Ea_T^*$, to establish (3.3) we must bound $E\Phi(d^*)$. Let $n_0 = 0$ and $n_1 =$ first m: $E\Phi(S_m^*) > 0$. Having defined $n_0, \dots, n_k$ let

$$n_{k+1} = \begin{cases} \text{first } m \ge n_k\text{: } E\Phi(S_m^*) \ge 2^k E\Phi(S_{n_1}^*), \\ \infty, \quad \text{if no such } m \text{ exists.} \end{cases}$$

Note that not all of the n_k need be distinct. Let $k^* =$ last k: $n_k \le T$ and $k^{**} = \sup\{k\colon n_k < \infty\}$. Observe that

$$2^{k^*-1} E\Phi(S_{n_1}^*) \le a_T^* < 2^{k^*} E\Phi(S_{n_1}^*). \tag{3.6}$$

Hence

$$
\begin{aligned}
E\Phi(d^*) &\le E \sum_{k=1}^{k^{**}} \Phi\Big(\max_{n_k \le j < n_{k+1}} \|X_j\|\Big) I(T \ge n_k) \\
&\le \sum_{k=1}^{k^{**}} E\Phi\Big(\max_{n_k \le j < n_{k+1}} \|X_j\|\Big) P(T \ge n_k) \\
&\le \sum_{k=1}^{k^{**}} E\Phi\big(2S^*_{n_{k+1}-1}\big) P(T \ge n_k) \\
&\le 2^\alpha \sum_{k=1}^{k^{**}} E\Phi\big(S^*_{n_{k+1}-1}\big) P(T \ge n_k) \\
&\le 2^\alpha \sum_{k=1}^{k^{**}} 2^k E\Phi\big(S^*_{n_1}\big) P(T \ge n_k) \\
&= 2^\alpha E\Phi\big(S^*_{n_1}\big) E \sum_{k=1}^{k^*} 2^k \\
&\le 2^{\alpha+2} E\Big(2^{k^*-1} E\Phi\big(S^*_{n_1}\big)\Big) \\
&\le 2^{\alpha+2} E a^*_T \quad [\text{by } (3.6)].
\end{aligned}
$$

Consequently,

$$E\Phi(S^*_T) \le (q\delta)^{-1}(1 + 2^{\alpha+2}) E a^*_T. \qquad \square$$

REMARK 3.2. When $\Phi(x) = x$, the bound given in (3.3) can be somewhat improved. Notice that in this case

$$\{S^*_T \ge \beta_y,\ d^* \vee a^*_T \le a_y\} \subseteq \{S^*_{(T_y,\, T \wedge n^*]} \ge y(\beta - 1 - \delta)\}$$

and

$$
\begin{aligned}
E\big(S^*_{(j,\, n^*]}\big) &\le E \max_{j < k \le n^*} \|S_k\| \quad (\text{by Jensen's inequality}) \\
&\le E S^*_{n^*}.
\end{aligned}
$$

Hence, $P(S^*_T \ge \beta y,\ d^* \vee a^*_T) \le \delta(\beta - 1 - \delta)^{-1} P(S^*_T \ge y)$ provided $\beta > 1 + \delta$. Putting $\beta = (1 + \delta)(1 - \sqrt{\delta})^{-1}$ and $\sqrt{\delta} = 2^{-1}(\sqrt{5} - 1)$, Lemma 1.2 implies that

$$(3.7) \qquad E \max_{1 \le n \le T} \|S_n\| \le 4(1 + \sqrt{5})(\sqrt{5} - 1)^{-1}(3 - \sqrt{5})^{-2} E(d^* \vee a^*_T).$$

4. Wald's equation: Further extension. We inquire when a randomly stopped Banach-space-valued martingale has zero mean.

THEOREM 4.1. *Let $(B, \|\cdot\|)$ be any Banach space. Let $S_n = X_1 + \cdots + X_n$ be a mean-zero martingale taking values in B and let T be any (possibly* *randomized) stopping time with respect to* $\{S_n\}$. *Suppose*

$$\lim_{n\to\infty} E\|S_{T\wedge n}\| < \infty. \tag{4.1}$$

Then

$$L \equiv \lim_{n\to\infty} ES_n I(T > n) \tag{4.2}$$

and ES_T *exist and*

$$ES_T = 0, \quad \textit{iff } L = 0 \ (\textit{since } ES_T = -L). \tag{4.3}$$

PROOF. By Fatou's lemma, $E\|S_T\| < \infty$. Hence ES_T exists and its value is given by

$$\begin{aligned} ES_T &= \lim_{n\to\infty} ES_T I(T \le n) \\ &= \lim_{n\to\infty} E\big(S_{T\wedge n} - S_n I(T > n)\big) \\ &= -\lim_{n\to\infty} ES_n I(T > n) \quad (\text{since } ES_{T\wedge n} \equiv 0). \end{aligned}$$

Consequently, L exists and $ES_T = -L$. □

REMARK 4.2. In the real-valued setting, Chow, Robbins and Siegmund [(1971), Theorem 2.3] (CRS) show that $ES_T = 0$ provided $E|S_T| < \infty$ and $\liminf_{n\to\infty} E|S_n| I(T > n) = 0$. Since $E|S_{T\wedge n}|$ is nondecreasing, the CRS conditions are equivalent to the two conditions $\lim_{n\to\infty} E|S_{T\wedge n}| < \infty$ and $\liminf_{n\to\infty} E|S_n| I(T > n) = 0$. Hence, Theorem 4.1 above strictly improves the CRS result, as Example 4.5 will demonstrate.

REMARK 4.3. Even though $E\|S_{T\wedge n}\|$ may tend to infinity, $E\|S_T\|$ can be finite and ES_T can equal zero. Thus, an all-encompassing theorem identifying when the randomly stopping sum continues to have mean zero is probably impossible. However, if a bit more regularity is required, Theorem 4.1 again appears to be definitive. Dubins and Freedman (1966) showed that (for real-valued martingales) there always exists a stopping time $T_0 \le T$ such that $E|S_{T_0}| = \infty$ whenever $\lim_{n\to\infty} E|S_{T\wedge n}| = \infty$. Hence, for real-valued martingales, $\{(4.1)$ and $L = 0\}$ is necessary and sufficient for

$$\{ES_T = 0 \text{ and } E|S_{T_0}| < \infty \text{ for all } T_0 \le T\}.$$

As a trivial consequence of Theorem 4.1 we have

COROLLARY 4.4. *Under the conditions of Theorem* 4.1,

$$ES_T = 0, \quad \textit{if } E \sup_{1\le n\le T} \|S_n\| < \infty. \tag{4.4}$$

PROOF. $\|S_{T\wedge 1}\|, \|S_{T\wedge 2}\|, \dots$ is a submartingale which is L^1 bounded whenever $E\sup_{1\le n\le T}\|S_n\| < \infty$. Moreover, in this case $\lim_{n\to\infty} E\|S_n\| I(T > n) = 0$. Hence, (4.1) holds and $L = 0$, proving (4.4). Of course, a simple proof using uniform integrability or dominated convergence would have sufficed. □

EXAMPLE 4.5. Let $X_1, X_2, \ldots$ be i.i.d. with $P(X_n = 1) = P(X_n = -1) = 2^{-1}$ and let $T =$ first $n \geq 2$: $S_n = 0$, where $S_n = X_1 + \cdots + X_n$. Then $|S_{T \wedge n}|$ is a martingale (!) so $E|S_{T \wedge n}| \equiv 1$. By symmetry, $ES_n I(T > n) \equiv 0$. Thus Theorem 4.1 but not Theorem 2.3 of CRS (1971) may be invoked to conclude the obvious fact $ES_T = 0$. Nor does Corollary 4.4 have the strength to entail $ES_T = 0$, since $E\max_{1 \leq n \leq T}|S_n| = \infty$. This latter assertion follows from (1.13), using the fact that $E\max_{1 \leq k \leq n}|S_k|$ has order $n^{1/2}$, and the well-known fact that $P(T \geq n) \approx Cn^{-1/2}$. Furthermore, it is interesting to note that $E|S_{T_0}| < \infty$ for every stopping time $T_0 \leq T$ even though $E\sup_{1 \leq k \leq T}|S_k| = \infty$. (To see this observe that

$$E|S_{T_0}| \leq \liminf_{n \to \infty} E|S_{T_0 \wedge n}| = \liminf_{n \to \infty} E|S_{T \wedge n}| = 1.)$$

Applying Theorem 3.1 to Corollary 4.4 it follows that if $S_n = \sum_{j=1}^n X_j$ is a sum of independent mean-zero Banach-space-valued random elements and T is any stopping time with respect to $\{S_n\}$,

$$ES_T = 0, \quad \text{if } Ea_T^* < \infty, \tag{4.5}$$

where

$$a_n^* = E \max_{1 \leq k \leq n} \|S_n\|. \tag{4.6}$$

With the next lemma, this result can be recast in the same form as (1.3) and (2.10).

LEMMA 4.6. *Let $X_1, X_2, \ldots$ be independent mean-zero random elements taking values in a Banach space $(B, \|\cdot\|)$. For each $k \geq 1$ let $S_k = X_1 + \cdots + X_k$. Then for any $n \geq 1$,*

$$E \max_{1 \leq k \leq n} \|S_k\| \leq 4E\|S_n\|. \tag{4.7}$$

Moreover, if $B = \mathbb{R}^1$, this can be improved to read

$$E \max_{1 \leq k \leq n} |S_k| \leq 3E|S_n|. \tag{4.8}$$

PROOF. An analogue of Ottaviani's inequality is required. Notice that a conditional version of Jensen's inequality implies that

$$E\|X_{k+1} + \cdots + X_n\| \leq E\|X_1 + \cdots + X_n\|,$$

for any $1 \leq k < n$. Hence, Markov's inequality entails

$$P(\|X_{k+1} + \cdots + X_n\| \leq 2E\|X_1 + \cdots + X_n\|) \geq 2^{-1}. \tag{4.9}$$

Fix $n \geq 1$. For any $y \geq 0$ let

$$\tau_y = \begin{cases} \text{first } 1 \leq k \leq n\text{: } \|S_k\| \geq y + 2E\|S_n\|, \\ \infty, \quad \text{if no such } k \text{ exists.} \end{cases}$$

Then

$$
\begin{aligned}
P\Big(\max_{1\le k\le n}\|S_k\| \ge y + 2E\|S_n\|\Big) &= \sum_{k=1}^{n} P(\tau_y = k) \\
&\le \sum_{k=1}^{n} 2P(\tau_y = k,\ \|S_n - S_k\| \le 2E\|S_n\|) \\
&\le 2P(\tau_y < \infty,\ \|S_n\| \ge y) \\
&\le 2P(\|S_n\| \ge y).
\end{aligned}
$$

This is the required Ottaviani analogue. Hence,

$$
\begin{aligned}
E\max_{1\le k\le n}\|S_k\| &= \int_0^\infty P\Big(\max_{1\le k\le n}\|S_k\| \ge t\Big)\,dt \\
&\le 2E\|S_n\| + \int_{2E\|S_n\|}^\infty P\Big(\max_{1\le k\le n}\|S_k\| \ge t\Big)\,dt \\
&\le 2E\|S_n\| + \int_0^\infty P\Big(\max_{1\le k\le n}\|S_k\| \ge y + 2E\|S_n\|\Big)\,dy \\
&\le 2E\|S_n\| + \int_0^\infty 2P(\|S_n\| \ge y)\,dy \\
&= 4E\|S_n\|.
\end{aligned}
$$

In the real-variables context, let

$$
\tau_y = \begin{cases} \text{first } 1 \le k \le n\colon S_k \ge y + E|S_n|, \\ \infty, \quad \text{if no such } k \text{ exists.} \end{cases}
$$

Notice that

$$
\begin{aligned}
P(S_n - S_k \ge -E|S_n|) &= 1 - P(S_k - S_n > E|S_n|) \\
&\ge 1 - E(S_k - S_n)^+/E|S_n| \quad \text{(by Markov)} \\
&\ge 1 - ES_n^-/E|S_n| \quad \text{(by conditional Jensen)} \\
&= 2^{-1} \quad (\text{since } ES_n = 0 \text{ implies } E|S_n| = 2ES_n^-).
\end{aligned}
$$

Reasoning as above,

$$
\begin{aligned}
P\Big(\max_{1\le k\le n} S_k \ge y + E|S_n|\Big) &\le 2\sum_{k=1}^{n} P(\tau_y = k,\ S_n - S_k \ge -E|S_n|) \\
&\le 2P(S_n \ge y).
\end{aligned}
$$

Similarly,

$$
P\Big(\max_{1\le k\le n}(-S_k) \ge y + E|S_n|\Big) \le 2P(-S_n \ge y).
$$

These two bounds combined imply that

$$
P\Big(\max_{1\le k\le n}|S_k| \ge y + E|S_n|\Big) \le 2P(|S_n| \ge y).
$$

Therefore,

$$\begin{aligned} E \max_{1\le k\le n} |S_n| &= \int_0^{E|S_n|} P\Big(\max_{1\le k\le n} |S_k| \ge y\Big)\,dy + \int_0^{\infty} P\Big(\max_{1\le k\le n} |S_k| \ge y + E|S_n|\Big)\,dy \\ &\le E|S_n| + \int_0^{\infty} 2P(|S_n| \ge y)\,dy \\ &= 3E|S_n|. \qquad \square \end{aligned}$$

REMARK 4.7. Inequality (4.8) improves the bound $E\max_{1\le k\le n}|S_n| \le 8E|S_n|$ found in Doob (1953), Theorem 5.1, Chapter VII.

Combining Lemma 4.6, Theorem 3.1 and Corollary 4.4, the next result is immediate.

COROLLARY 4.8. *Let $X_1, X_2, \dots$ be independent mean-zero random elements taking values in a Banach space $(B, \|\cdot\|)$. Let $S_n = X_1 + \cdots + X_n$ and $a_n = E\|S_n\|$. Let T be any (possibly randomized) stopping time with respect to $\{S_n\}$ (or equivalently $\{X_n\}$). Then*

$$ES_T = 0, \quad \text{if } Ea_T < \infty. \tag{4.10}$$

REFERENCES

BLACKWELL, D. (1946). On an equation of Wald. *Ann. Math. Statist.* **17** 84–87.

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., DAVIS, B. J. and GUNDY, R. F. (1972). Integral inequalities for convex functions of operators on martingales. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **2** 223–240. Univ. California Press.

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

CHOW, Y. S., ROBBINS, H. and SIEGMUND, D. (1971). *Great Expectations. The Theory of Optimal Stopping.* Houghton Mifflin, New York.

CHUNG, K. L. (1974). *A Course in Probability Theory*, 2nd ed. Academic, New York.

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

DOEBLIN, W. (1939). Sur les sommes d'un grand nombre de variables aléatoires independantes. *Bull. Sci. Math.* **63** 23–32 and 35–64.

DOOB, J. L. (1953). *Stochastic Processes.* Wiley, New York.

DUBINS, L. E. and FREEDMAN, D. (1966). On the expected value of a stopped martingale. *Ann. Math. Statist.* **37** 1505–1509.

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

GUNDY, R. F. (1967). The martingale version of a theorm of Marcinkiewicz and Zygmund. *Ann. Math. Statist.* **38** 725–734.

KHINTCHINE, A. (1923). Über dyadische Bruch. *Math. Z.* **18** 109–116.

KLASS, M. J. (1973). Properties of optimal extended-valued stopping rules for S_n/n. *Ann. Probab.* **1** 719–757.

KLASS, M. J. (1974). On stopping rules and the expected supremum of S_n/a_n and $|S_n|/a_n$. *Ann. Probab.* **2** 889–905.

KLASS, M. J. (1976). Toward a universal law of the iterated logarithm. *Z. Wahrsch. verw. Gebiete* **36** 165–178.

KLASS, M. J. (1980). Precision bounds for the relative error in the approximation of $E|S_n|$ and extensions. *Ann. Probab.* **8** 350–367.

KLASS, M. J. (1981). A method of approximating expectations of functions of sums of independent random variables. *Ann. Probab.* **9** 413–428.

KLASS, M. J. (1985). Order of magnitude bounds for expectations of functions of Banach space sums. Preprint.

MARCINKIEWICZ, J. (1937). Quelques théorèmes sur les séries orthogonales. *Ann. Soc. Math. Polon.* **16** 84–96 (pages 307–318 of the *Collected Papers* of Marcinkiewicz).

MARCINKIEWICZ, J. and ZYGMUND, A. (1937). Sur les fonctions indépendantes. *Fund. Math.* **29** 60–90 (pages 233–259 of the *Collected Papers* of Marcinkiewicz).

MARCINKIEWICZ, J. and ZYGMUND, A. (1938). Quelques théorèmes sur les fonctions indépendantes. *Studia Math.* **7** 104–120 (pages 374–388 of the *Collected Papers* of Marcinkiewicz).

MILLAR, P. W. (1968). Martingale integrals. *Trans. Amer. Math. Soc.* **133** 145–166.

PALEY, R. E. A. C. (1932). A remarkable series of orthogonal functions. I. *Proc. London Math. Soc.* **34** 241–264.

ROBBINS, H. E. and SAMUEL, E. (1966). An extension of a lemma of Wald. *J. Appl. Probab.* **3** 272–273.

SAMUEL, E. (1967). On the existence of the expectation of randomly stopped sums. *J. Appl. Probab.* **4** 197–200.

WALD, A. (1945). Sequential tests of statistical hypotheses. *Ann. Math. Statist.* **16** 117–186.

YOR, M. (1985). Private communication.

DEPARTMENTS OF MATHEMATICS
AND STATISTICS
UNIVERSITY OF CALIFORNIA, BERKELEY
BERKELEY, CALIFORNIA 94720