The Annals of Mathematical Statistics
1969, Vol. 40, No. 3, 764–769

A BEST POSSIBLE KOLMOGOROFF-TYPE INEQUALITY FOR MARTINGALES AND A CHARACTERISTIC PROPERTY

By W. L. Steiger

Australian National University, Canberra

1. Introduction and summary. In this paper some Kolmogoroff-type inequalities are derived. One of them is best possible for martingales. Another holds for sums of independent random variables and for martingales in which conditional variances satisfy a certain property. But it does not hold, in general, for martingales. This elucidates an important difference between martingales and independent sums.

Let E denote expectation. Consider a sequence of random variables $X_1, \cdots, X_n$ and the following three assumptions:

(a) $E(X_1) = 0, \quad E(X_i \mid X_{i-1}, \cdots, X_1) = 0, \qquad i = 2, \cdots, n.$

(b) $|X_i| \leqq T$ almost surely, $\qquad i = 1, \cdots, n.$

(c) $E(X_1^2) \neq 0, \quad E(X_i^2 \mid X_{i-1}, \cdots, X_1) \neq 0$ almost surely, $\quad i = 2, \cdots, n.$

The first says that $\{X_i\}$ is absolutely fair, the second that $\{X_i\}$ is almost surely bounded, and the third, that conditional variances are positive.

We shall deal with the following classes of sums. $M(n)$ is the class of all martingales $\{S_i\}$ of n partial sums $S_i = X_1 + \cdots + X_i$ where $\{X_i\}$ satisfies (a). $B(n) \subset M(n)$ is the subclass in which $\{X_i\}$ satisfies (a), (b), (c), $V(n) \subset B(n)$ the subclass where, additionally, $E(S_n^2) = E(S_n^2 \mid X_{n-1}, \cdots, X_1)$, and finally $I(n) \subset B(n)$ the subclass where, in addition, $\{X_i\}$ are independent. $M(n)$ is a pneumonic for martingale, $B(n)$ for bounded martingale, $V(n)$ for martingales where variance equals conditional variance, and $I(n)$ for independent sums.

Write $\sigma_i^2 = E(X_i^2)$, $s_i^2 = \sigma_1^2 + \cdots + \sigma_i^2$, $i = 1, \cdots, n$, $C_1^2 = E(X_1^2)$, $C_i^2 = E(X_i^2 \mid X_{i-1}, \cdots, X_1)$, $i = 2, \cdots, n$, $C^2 = C_1^2 + \cdots + C_n^2$. Let $M_n = \max(S_1, \cdots, S_n)$, $p(t) = \Pr\{M_n \geqq ts_n^2\}$ and

$$r(t) = [e/(1 + tT)]^{ts_n^2/T}[1/(1 + tT)]^{s_n^2/T^2}.$$

Bennett [1] and Hoeffding [4] showed independently that

$$\sup_{I(n)} [\Pr\{S_n \geqq ts_n^2\}] \leqq r(t).$$

Steiger [7] extended this to the context of Kolmogoroff inequalities by showing that $p(t) \leqq r(t)$ for all sums $\{S_i\} \,\varepsilon\, I(n)$. In Section 3 we show that the inequality actually holds throughout the larger class, $V(n)$, and that it is best-possible there. However the inequality can be false in $B(n)$. This shows that the maxima of martingale partial sums may have larger tail probabilities than those of independent sums or of sums in $V(n)$ — *a* characteristic property.

Received 4 March 1968.

In Section 2 we prove two preliminary lemmas of independent interest. These are used in Section 3 to obtain Theorem 1, which gives a best-possible upper bound in $B(n)$ for $\Pr\{M_n \geqq tC^2\}$ which, when restricted to $V(n)$ gives $\sup_{V(n)} [p(t)] \leqq r(t)$, also best possible. These results are compared to known results of Marshall [6], Dubins and Savage [2], and Steiger [7]. Finally, in Section 4, a partial converse to Theorem 1 is given.

2. Preliminary results. Assumption (c) guarantees that for sums $\{S_i\} \varepsilon B(n)$, there is a number $b > 0$ such that $C^2 \geqq b$ almost surely. The following lemma shows how the idea used in proving the well known Bernstein Inequality adapts to the present situation, and has interest in its own right.

LEMMA 1. *Let $k > 0$ be a real number and $h > 0$ a non-decreasing, convex function. Then for all $\{S_i\} \varepsilon B(n)$*

$$\Pr\{M_n \geqq tC^2\} \leqq E[h(kS_n)]/h(ktb).$$

PROOF. The assumptions and a result of Feller ([3], p. 215) imply that $\{h(S_i)\}$ is a sub-martingale so that for $i \geqq j > k \geqq 1$,

$$E(h(S_i)|\, h(S_k), \cdots, h(S_1)) \geqq E(h(S_j)|\, h(S_k), \cdots, h(S_1)). \tag{1}$$

Take $y \geqq 0$ and define the random variable

$$\begin{aligned} L(y) &= \min(j, 1 \leqq j \leqq n\colon S_j \geqq y), \qquad \text{or} \\ &= 0 \quad \text{if} \quad S_j < y, \qquad 1 \leqq j \leqq n. \end{aligned}$$

Then $\Pr\{M_n \geqq ty\} = \sum_{j=1}^{n} \Pr\{L(ty) = j\}$, $t > 0$. It is easy to see that

$$\begin{aligned} E(h(kS_n)) &\geqq \textstyle\sum_{i=1}^{n} E[h(kS_n)|\, L(ty) = i] \Pr\{L(ty) = i\} \\ &\geqq \textstyle\sum_{i=1}^{n} E[h(kS_i)|\, L(ty) = i] \Pr\{L(ty) = i\} \\ &\geqq h(kty) \Pr\{M_n \geqq ty\} \end{aligned} \tag{2}$$

the first line following from $h > 0$, the second from (1) and the last from the definition of L and because h is non-decreasing. In particular (2) holds for $y = b$ and since $C^2 \geqq b$ a.s., the lemma is proved.

The next lemma, also of independent interest, shows how to bound the right hand side in Lemma 1, when $h = \exp$, in the best-possible way.

For all $i = 1, \cdots, n$ take $0 < d_i \leqq T$ such that $C_i^2 \leqq d_i^2$ almost surely and put $d^2 = d_1^2 + \cdots + d_n^2$. Note that by assumption (b), $C_i^2 \leqq T^2$, $i = 1, \cdots, n$.

LEMMA 2. *For $k > 0$ and $\{S_i\} \varepsilon B(n)$,*

$$E(e^{kS_n}) \leqq \textstyle\prod_{i=1}^{n} [1 + (d_i^2/T^2)(e^{kT} - kT - 1)]. \tag{3}$$

REMARK 1. (3) and the arithmetic-geometric mean inequality imply that $E(e^{kS_n})$ is no more than $[1 + d^2(e^{kT} - kT - 1)/(nT^2)]^n$ and $1 + u \leqq e^u$ then yields

$$E(e^{kS_n}) \leqq \exp((e^{kT} - kT - 1)\, d^2/T^2). \tag{4}$$

PROOF. $e^{kx} \leqq 1 + kx + ax^2$ for $k > 0$ and all $x \varepsilon [-T, T]$ if and only if

$a \geqq (e^{kx} - kx - 1)/x^2$ which implies that $e^{kx} \leqq 1 + kx + x^2(e^{kT} - kT - 1)/T^2$ since $(e^{kx} - kx - 1)/x^2$ is increasing. Put X_1 for x into this inequality and take expectations to prove the lemma for $n = 1$. To advance the induction suppose the statement is true for $n = m$. Then

$$E(e^{kS_{m+1}}) = \int_{-mT}^{mT} e^{ka} E(e^{kX_{m+1}} \mid S_m = a)\, d\Pr\{S_m \leqq a\}. \tag{5}$$

An argument similar to that used for the case $n = 1$ shows that

$$E(e^{kX_{m+1}} \mid S_m = a) = E[(1 + kX_{m+1} + X_{m+1}^2(e^{kT} - kT - 1)/T^2) \mid S_m = a]$$

and since $C_{m+1}^2 \leqq d_{m+1}^2$ almost surely, the right hand side of (5) is no more than

$$[1 + d_{m+1}^2(e^{kT} - kT - 1)/T^2] \int_{-mT}^{mT} e^{ka}\, d\Pr\{S_m \leqq a\}.$$

Using the induction hypothesis, the lemma is proved.

REMARK 2. (3) is best-possible. Take $T > 0$. Choose $d > 0$ such that $d \leqq T$. Define for each $i = 1, \cdots, n$ the random variable X_i by

$$\Pr\{X_i = T\} = d^2/(d^2 + nT^2),$$
$$\Pr\{X_i = -d^2/(nT)\} = nT^2/(d^2 + nT^2)$$

and suppose $\{X_i\}$ independent. Then $|X_i| \leqq T$, $E(X_i) = 0$, and $E(X_i^2) = d^2/n$, $i = 1, \cdots, n$, so that $\{S_i\} \varepsilon B(n)$. From the independence of $\{X_i\}$ a computation shows that for $k > 0$,

$$\begin{aligned} E(e^{kS_n}) &= \textstyle\prod_{i=1}^n E(e^{kX_i}) \\ &= e^{-kd^2/T}[nT^2/(d^2 + nT^2)]^n[1 + d^2(e^{kT}e^{kd^2/(nT)})/(nT^2)]^n \end{aligned}$$

and so $\lim_{n\to\infty} E(e^{kS_n}) = \exp((e^{kT} - kT - 1)\, d^2/T^2)$ which is (4). Thus there is no function w on the positive integers such that for each $\epsilon > 0$ and all n,

$$E(e^{kS_n}) < w(n) + \epsilon \leqq \exp((e^{kT} - kT - 1)/d^2/T^2).$$

3. A Kolmogoroff-type inequality for $B(n)$.

From Lemmas 1 and 2 it is easy to establish

THEOREM 1. *Let $\{S_i\} \varepsilon B(n)$, $0 < b \leqq C^2$ almost surely, and take $0 \leqq d^2 \leqq nT^2$ such that $C^2 \geqq d^2$ almost surely. Then for all $t > 0$,*

$$\Pr\{M_n \geqq tC^2\} \leqq e^{tb/T}[d^2/(d^2 + tbT)]^{(tbT+d^2)/T^2}. \tag{6}$$

PROOF. Use Lemma 1 with $h = \exp$ and then (4). Minimize the resulting inequality with respect to k.

REMARK 3. The right hand side of (6) decreases as b increases and hence, (6) is strongest when $b = \sup(x > 0 : x \leqq C^2 \text{ a.s.})$. With this choice of b, (6) is best-possible, based on Lemma 1, because (4) is.

Dubins and Savage ([2], see [5], Chap. 14) prove that for all $\{S_i\} \varepsilon M(n)$,

$$Q(n) = \inf_{M(n)}[\Pr\{S_k < t(C_1^2 + \cdots + C_k^2 + 1),\quad \text{all}\quad k, 1 \leqq k \leqq n\}]$$

where $Q(1) = 4t^2/(1 + 4t^2)$, $Q(n) = t^2f_{n-1}(4)/(1 + t^2f_{n-1}(4))$, $n \geqq 2$, and

$f_1(s) = (2s/(1+s))^2, f_n(s) = f_1(f_{n-1}(s))$. Note that $\lim_{n\to\infty} Q(n) = t^2/(1+t^2)$ because $f_1(1) = 1, f_1(s) < s, s > 1, f_1(s) > s, s < 1$, so that $f_n(s) \to 1$ for all $s > 0$. This result implies that

$$\text{(7)} \qquad \Pr\{M_n \geqq t(C^2+1)\} \leqq 1/(1+t^2 f_{n-1}(4)),$$

for all $\{S_i\} \varepsilon B(n)$. Table 1 compares the bound in (7) with that of (6) for selected values of b, d^2, and t.

REMARK 4. In $V(n)$, $C^2 \equiv s_n^2$ and we may take $b = d^2 = s_n^2$. (6) then becomes

$$\text{(8)} \qquad \sup_{V(n)} [p(t)] \leqq r(t).$$

Combining Remarks 2 and 3, (8) is best-possible.

Marshall [6] has shown that $\sup_{M(n)} [p(t)] = 1/(1 + t^2 s_n^2)$ and Steiger [7], that $\sup_{B(n)} [p(t)] \leqq [e/(1+tnT)]^{ts_n^2/(nT)}[1/(1+tnT)]^{s_n^2/(n^2T^2)}$. Table 2 compares (8) with these results for selected values of n, s_n^2 and t and shows how they can be improved in $V(n)$.

REMARK 5. $p(t) \leqq r(t)$ is false in $B(n)$ and thus points out a characteristic difference between the classes $I(n)$, $V(n)$, and $B(n)$, as the following example

TABLE 1

Comparison of the right hand side of (6), $T = 2$, $b = 1$ and selected values of d^2 with the right hand side of (7) for selected values of n.

t	(6) $b = 1$, $d^2 = 2$	(6) $b = 1$, $d^2 = 4$	(7) $n = 6$	(7) $n = 10$
1.5	.673	.795	.222	.248
3.0	.280	.454	.066	.076
4.5	.087	.206	.031	.035
6.0	.022	.079	.018	.020
7.5	.005	.026	.011	.013
9.0	.001	.008	.008	.009
10.5	.000	.002	.006	.007

TABLE 2

Comparison of $r(t)$ with the inequalities of Marshall and Steiger for $T = 2$ and selected values of n, s_n^2, t.

t	$r(t)$ $s_n^2 = 1$	Marshall $s_n^2 = 1$	Steiger $n = 6$, $s_n^2 = 1$	Steiger $n = 10$, $s_n^2 = 5$
1.5	.529	.308	.768	.385
3.0	.149	.100	.508	.092
4.5	.030	.047	.315	.018
6.0	.005	.027	.187	.003
7.5	.001	.018	.108	.001
9.0	.000	.012	.061	.000
10.5	.000	.001	.034	.000

shows. Take $0 < q < 1$ and define the random variables X_1 and X_2 by

$$\Pr\{X_1 = 1\} = q/(1+q)$$
$$\Pr\{X_1 = -q\} = 1/(1+q)$$
$$\Pr\{X_2 = 1 \mid X_1 = 1\} = (1+q)/(3+q)$$
$$\Pr\{X_2 = -(1+q)/2 \mid X_1 = 1\} = 2/(3+q)$$
$$\Pr\{X_2 = 1 \mid X_1 = -q\} = q(1+q)/(2+q(1+q))$$
$$\Pr\{X_2 = -q(1+q)/2 \mid X_1 = -q\} = 2/(2+q(1+q)).$$

Define $\{S_i\}$, $S_i = \sum_{j=1}^{i} X_j$, $i = 1, 2$. It is easy to check that assumptions (a), (b), (c) are satisfied so that $\{S_i\} \varepsilon B(2)$. Because $E(S_2^2) = 2q \neq E(S_2^2 \mid X_1 = 1) = (3+q)/2$, $\{S_i\} \not\varepsilon V(2)$. A sample computation reveals that when $q = .025$ and $t = 36$, $p(t) = \Pr\{\max(S_1, S_2) \geqq 1.65\} = .0083 \cdots$ while $r(t) = .0076 \cdots$ so that (8) is false in $B(n)$. Loosely speaking, the tail probabilities, $\Pr\{M_n \geqq ts_n^2\}$, for maxima of martingale sums are larger than those for sums in $V(n)$ and hence $I(n)$.

4. A further result. The proof of Theorem 1 used (4) which bounds moment generating functions of elements of $B(n)$. It is interesting that the conclusion of the Theorem, (6), itself provides a bound for moment generating functions of elements of $M(n)$ which is a partial converse to "(4) implies (6)". Specifically,

THEOREM 2. *Take $\{S_i\} \varepsilon M(n)$ and numbers b, d^2, T such that almost surely $0 < b \leqq C^2 \leqq d^2 \leqq nT^2$, If (6) holds for $\{S_i\}$ then for each $K > 0$, the moment generating function of S_n, $g(k) = E(e^{kS_n})$, exists, and*

$$g(k) \leqq e^{Wk^2}, \qquad |k| \leqq K, \tag{9}$$

where the constant W is independent of k.

PROOF. The statements

$$\begin{aligned} E(e^{kS_n}) &= \textstyle\int_{-\infty}^{\infty} e^{kx}\, d\Pr\{S_n \leqq x\} \\ &= 1 + \textstyle\int_{-\infty}^{\infty} (e^{kx} - kx - 1)\, d\Pr\{S_n \leqq x\} \\ &\leqq 1 + \tfrac{1}{2}\textstyle\int_{-\infty}^{\infty} k^2x^2e^{|k|x}\, d\Pr\{S_n \leqq x\} \end{aligned} \tag{10}$$

follow from definition, $\{S_i\} \varepsilon M(n)$, and the inequality $e^{kx} - kx - 1 \leqq (k^2x^2e^{|k|x})/2$, respectively. From (6) and $\Pr\{|S_n| \geqq td^2\} \leqq 2\Pr\{M_n \geqq tC^2\}$ we have

$$\Pr\{|S_n| \geqq u\} \leqq 2e^{ub/(d^2T)}/[1 + ubT/d^2]^{(ub/(d^2T)+d^2/T^2)}, \qquad u \geqq 0, \tag{11}$$

Integrate the last line of (10) by parts and then use (11) and $|k| \leqq K$ to see that

$$E(e^{kS_n}) \leqq 1 + 2k^2 \textstyle\int_{-\infty}^{\infty} xe^{Kx}[e^{xb/(d^2T)}/(1 + xbT/d^2)^{(xb/(d^2T)+d^2/T^2)}]\, dx. \tag{12}$$

The integral in (12) exists, independent of k and equals, say, $W/2$ which proves the theorem because $1 + k^2W \leqq e^{k^2W}$.

(4) is stronger than (9) in $B(n)$ since (4) with $|k| \leqq K$ implies (9).

Acknowledgement. I would like to thank Professor P.A.P. Moran for many helpful discussions about this work, and the referee for several useful remarks.

REFERENCES

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

[2] DUBINS, L. E. and SAVAGE, L. J. (1965). *How to Gamble if You Must.* McGraw-Hill, New York.

[3] FELLER, W. (1966). *An Introduction to Probability Theory and its Applications*, **2**. Wiley, New York.

[4] HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–29.

[5] KARLIN, S. and STUDDEN, W. J. (1966). *Tchebycheff systems: With Applications in Analysis and Statistics.* Interscience, New York.

[6] MARSHALL, A. W. (1960). A one-sided analogue of Kolmogoroff's inequality. *Ann. Math. Statist.* **31** 483–487.

[7] STEIGER, W. L. (1967). Some Kolmogoroff-type inequalities for bounded random variables. *Biometrika.* **54** 641–648.