THE BEST BOUND IN THE $L \log L$ INEQUALITY OF HARDY AND LITTLEWOOD AND ITS MARTINGALE COUNTERPART

DAVID GILAT¹

ABSTRACT. The smallest positive constant c for which the Hardy and Littlewood [8] $L \log L$ inequality

$$(1) \quad M(f) \equiv \int \frac{dx}{x} \int_0^x |f| \leq c \left(1 + \int |f| \log^+ |f| \right) \\ \equiv c(1 + L(f))$$

is valid is proved to be the unique positive solution c_0 of the equation $e^{-c} = (c-1)^2$. This settles a question raised by Dubins and Gilat (1978) [6] and, again, more recently, by D. Cox (1984) [3].

Numerically, $c_0 \doteq 1.478$. This should be compared with $c = e(e-1)^{-1} \doteq 1.582$, obtained by Doob (1953) [5] in the context of martingale theory and, since then, widely used in the probability literature. Curiously enough, Doob's coefficient is the best upper bound, but for a slightly different inequality. If only the plus sign is removed from $\log^+ |f|$ in (1), then c must be at least $e(e-1)^{-1}$ for (1), so modified, to be valid. The original inequality (1) is a normalized form of the two-parameter inequality

$$(2) \quad M(f) \leq cL(f) + d, \quad \text{for all integrable } f.$$

The set of all ordered pairs (c, d) for which (2) holds is identified as

$$\{(c, d): c > 1, d \geq 1 + e^{-c}/(c-1)\}.$$

Furthermore, for each point on the lower boundary of this set, there is a unique f (up to null sets) which attains equality in (2).

Preface. (A mathematician's apology?) The main body of this article, §§2 and 3, is self-contained and can be read quite apart from the introductory §1. Why, then, the introduction at all? Well, as is often the case, a mathematical theme can be formulated in more than one conceptual framework. In such cases the author is faced with a serious selection problem. To be specific, the subject matter of this article has roots in both classical analysis and probability theory. To leave any one of them out, was felt to be unfair to the other. On the other hand, a coherent nonrepetitive presentation of both, seemed to be beyond my expository ability. The formulation adopted is thus a compromise between a complete omission and a confused repetitiveness. While the main body of the paper is cast entirely in purely

Received by the editors April 23, 1985.

1980 *Mathematics Subject Classification*. Primary 26D15, 60G42; Secondary 39C05, 42B25.

Key words and phrases. Hardy-Littlewood maximal function, $L \log L$, L_p , martingale inequalities, Doob.

¹This work was initiated at University of California, San Diego and was concluded at Vrije Universiteit, Amsterdam. The author wishes to thank the Mathematics Departments of both institutions for their hospitality and to acknowledge the financial support of VU.

analytic terms, the introductory §1 explores the martingale connections through a brief survey of some of the relevant literature.

1. Introduction. For a Lebesgue integrable function f on the unit interval, let $A_f(a, x)$ be the average of f over the subinterval $[a, x]$; i.e.,

$$A_f(a, x) = \frac{1}{x-a} \int_a^x f(t) dt, \quad 0 \leq a < x;$$

and let $F(x)$ be the supremum of these averages when a ranges over $[0, x]$; i.e.,

$$F(x) = \sup_{0 \leq a \leq x} A_f(a, x).$$

In their fundamental paper of 1930, Hardy and Littlewood [8] obtain bounds on the absolute moments of F in terms of corresponding moments of f . It is important to note that two f 's with the same distribution (w.r.t. Lebesgue measure), hence the same moments, may give rise to differently distributed maximal average functions F , whose moments can, therefore, also be different.

Among all functions distributed like the given f , there is an essentially unique one which is nonincreasing; denote it by f^* and let F^* be the corresponding F -function (note: $F^*(x) = A_{f^*}(0, x) = (1/x) \int_0^x f^*(t) dt$).

The basic *maximal theorem* of Hardy and Littlewood [8, Theorems 5, 6] can now be stated as follows:

(HL max). *Given an $f \geq 0$, the distribution of F^* dominates the distribution of every F , in the sense that for every real y ,*

$$m\{x: F(x) > y\} \leq m\{x: F^*(x) > y\},$$

where m stands for the Lebesgue measure.

It is due to this theorem that F^* came to be known as the Hardy-Littlewood (HL, for short) *maximal function* associated with (the distribution of) f .

The present paper focuses on the moment inequalities deduced by Hardy and Littlewood from (HL max) and on similar martingale inequalities due to Doob. To facilitate this introductory discussion, their results will now be briefly summarized.

As is customary, for $p \geq 1$, $\|\cdot\|_p$ denotes the L_p -norm, while $\int |\cdot| \log^+ |\cdot|$ will be abbreviated by $L(\cdot)$.

(HL 1). *For $p > 1$, if f belongs to L_p , then so does F , and*

$$\|F\|_p \leq (p/(p-1))\|f\|_p.$$

Conversely, $F^ \in L_p$ implies $f \in L_p$.*

Note that the implication $F \in L_p \Rightarrow f \in L_p$ is generally false, unless f is nonincreasing, in which case it is trivial.

In contrast to the situation for $p > 1$, f being merely in L_1 is not enough to guarantee the same for F . Instead

(HL 2). *If $f \in L \log L$ (i.e. $L(f) < \infty$), then $F \in L_1$, and*

$$\|F\|_1 \leq c(1 + L(f)),$$

where c is some positive constant independent of f .

(Here too, the reverse implication, $F \in L_1 \Rightarrow f \in L \log L$, holds, provided f is nonincreasing, but is otherwise false.)

Completely analogous inequalities were obtained by Doob [5, p. 317] in the context of martingale theory. Doob's results will now be quoted for the sake of comparison.

(D 1 AND 2). *Let f be the last element of a martingale (defined on some probability space) and let F be the supremum, over time, of its absolute values. Then (HL 1) and (HL 2) hold.*

In other words, with an appropriate martingale interpretation of f and F , Doob's inequalities are just identical with their HL counterparts. It is thus somewhat surprising that Doob does not even mention the HL inequalities while presenting his own.

The link between the HL theory and probability was made primarily by Blackwell and Dubins [1], Burkholder [2(i)] and Gundy [7]. Blackwell and Dubins (1963) were apparently the first to clearly point out the role of the HL maximal function in martingale theory. They show that the HL maximal function associated with the (distribution of the) terminal element of a martingale dominates (in distribution) the supremum over time of that martingale. Dubins and Gilat [6] complement this result by demonstrating that for any distribution with a finite mean, there is a martingale (in continuous time) whose terminal element has the given distribution and whose time-supremum is distributed like the HL maximal function associated with that distribution. Thus a complete equivalence is established between Doob's maximal martingale inequalities and the corresponding results of Hardy and Littlewood. A related property of the HL maximal function is obtained by Meilijson and Nadas [10].

Neither Hardy and Littlewood nor Doob addresses the question of sharpness of their respective inequalities. Dubins and Gilat [6] establish the sharpness of (HL 1), hence also of (D 1). In more recent papers, Cox [3] and Cox and Kertz [4] obtain sharp ($p > 1$) moment comparisons between the last element and the maximum of a *finite* martingale sequence. As a corollary they reestablish the sharpness of (D 1) or (HL 1), deducing the additional piece of information that for no f can equality be attained in (HL 1) or (D 1). Donald Burkholder, who read an earlier version of the present manuscript, has pointed out [2(ii)] that both the sharpness and strictness of (HL 1) are proved on p. 240 of Hardy, Littlewood and Polya [9], and that the strictness of the inequality can also be deduced from a careful analysis of Doob's [5, p. 317] proof of (D 1). This was regrettably overlooked by the authors of [6].

Dubins and Gilat [6] as well as Cox [3] explicitly leave open the question of sharpness for the $L \log L$ inequality (HL 2) or (D 2). It is the purpose of the present paper to close this gap. While Hardy and Littlewood do not attempt at all to obtain numerical values for the constant c , Doob obtains his $L \log L$ inequality with $c = e(e - 1)^{-1} = 1.582$. It turns out that this value can only be reduced by a mere 7%, the smallest possible c being roughly $c_0 = 1.478$ (see Theorem 1 below).

Curiously enough, Doob's coefficient, $c_D = e(e - 1)^{-1}$, is in fact the best upper bound, but for a slightly modified form of (HL 2). If only the plus sign is removed from $\log^+ |f|$, introducing perhaps a negative contribution to $L(f)$, c must be raised from c_0 to at least c_D , in order for (HL 2), so modified, to remain valid.

The results are stated in §2 and proved in §3.

2. Results. For $f \in L_1(0, 1)$, let

$$M(f) = \int_0^1 \frac{dx}{x} \int_0^x |f(t)| dt$$

and

$$L(f) = \int_0^1 |f(x)| \log^+ |f(x)| dx.$$

THEOREM 1. (i) *The smallest c such that the inequality*

$$(1) \quad M(f) \leq c(1 + L(f))$$

is satisfied for every f in $L_1(0, 1)$ is the unique positive solution c_0 of the equation $e^{-c} = (c - 1)^2$. Numerically, $c_0 \doteq 1.478 < e/(e - 1) \doteq 1.582$. Furthermore,

(ii) *equality in (1) with $c = c_0$ is (essentially) uniquely attained by the function*

$$f_0(x) = \max\{e^{-1}x^{-1/c_0}, 1\} = e^{-1}x^{-1/c_0}I_{(0, e^{-c_0})}(x) + I_{(e^{-c_0}, 1]}(x).$$

Inequality (1), treated in Theorem 1, is a normalized form (with $d = c$) of the more general two-parameter inequality

$$(2) \quad M(f) \leq cL(f) + d \quad \text{for all } f \in L_1(0, 1).$$

Theorem 2 identifies the set of all ordered pairs (c, d) for which (2) holds.

THEOREM 2. (i) *Inequality (2) is satisfied if and only if $c > 1$ and $d \geq d(c) = 1 + e^{-c}/(c - 1)$. Furthermore,*

(ii) *for every $c > 1$ and $d = d(c)$, equality in (2) is (essentially) uniquely attained by the function*

$$f_c(x) = \max\{e^{-1}x^{-1/c}, 1\} = e^{-1}x^{-1/c}I_{(0, e^{-c})}(x) + I_{(e^{-c}, 1]}(x).$$

Here, as well as before and henceforth, “essentially” means “modulo modifications on sets of measure zero”.

Note. Theorem 1 is the diagonal form, $c = d$, of Theorem 2. For the c_0 defined in Theorem 1, f_{c_0} (in Theorem 2) is identical with f_0 (in Theorem 1).

COROLLARY 1. *Retaining the coefficient $c = e(e - 1)^{-1}$ in Doob's inequality*

$$M(f) \leq \frac{e}{e - 1}(1 + L(f)),$$

the “1” on the right-hand side can be reduced to

$$\frac{e - 1}{e} \left[1 + (e - 1) \exp \left(-\frac{e}{e - 1} \right) \right] \doteq 0.855,$$

but not any further.

Modify $L(f)$ by removing the plus sign from $\log^+ |f|$, to obtain

$$K(f) = \int_0^1 |f(x)| \log |f(x)| dx.$$

Note that, in contrast with $L(f)$, $K(f)$ can be negative, and in any case $K(f) \leq L(f)$ with equality iff $|f| \geq 1$ a.e.

THEOREM 3. (i) *The inequality*

$$(3) \quad M(f) \leq cK(f) + d$$

is satisfied for all $f \in L_1(0, 1)$ if and only if $c > 1$ and $d \geq d(c) = e^{-1}c^2(c-1)^{-1}$. Furthermore,

(ii) for every $c > 1$ and $d = d(c)$, equality in (3) is attained by the function $f^c(x) = e^{-1}x^{-1/c}$, $0 < x \leq 1$, and f^c is (essentially) the only function with this property.

COROLLARY 2. *The smallest c for which the diagonal form*

$$(3') \quad M(f) \leq c(K(f) + 1)$$

of (3) holds is Doob's coefficient $c_D = e(e-1)^{-1}$.

3. Proofs. The functions f which take part in the statement of the results do so only through their absolute values. It is therefore possible to assume, as we shall from here on, that f is nonnegative. Also, since integration is exclusively Lebesgue, it is harmless, as well as convenient, to use abbreviations such as $\int f \log(1/x)$ in lieu of the traditional $\int f(x) \log(1/x) dx$. This abbreviated notation will be henceforth adhered to without further comment. Thus, for a nonnegative f ,

$$M(f) = \int \frac{dx}{x} \int_0^x f \quad \text{and} \quad L(f) = \int f \log^+ f.$$

" $\int$ ", without specifying the limits of integration, will always mean to be considered over the entire unit interval.

Since, as has already been observed, Theorem 1 is a special case of Theorem 2, of those two, only the latter will be proved.

Initially, two elementary facts used in the proof are listed.

(4) *For a nonnegative measurable f ,*

$$M(f) = \int f \log \frac{1}{x}.$$

(PROOF. Straightforward integration by parts. For more details see Lemma 7 of [8].)

$$(5) \quad \log y \leq e^{-1}y \quad \text{for all } y > 0,$$

with equality only at $y = e$.

(PROOF. $y \rightarrow \log y$ is concave and has the line $y \rightarrow e^{-1}y$ tangent to its graph at $y = e$.)

PROOF OF THEOREM 2. It is convenient to break up the proof into two propositions, the second being the heart of the argument. The easy Proposition 1 is designed to demonstrate that with $c \leq 1$ and an arbitrary d , inequality (2) is violated for some f .

PROPOSITION 1. *For every d , no matter how large, there exists an f such that $M(f) > L(f) + d$. Thus, only c 's strictly bigger than 1 qualify for (2).*

PROOF. It clearly suffices to consider only $d > 1$. For such d let $f(x) = x^{-\alpha}$, with $0 < \alpha = \alpha(d) < 1$ to be determined later. Then, by (4),

$$L(f) = \int f \log f = \alpha \int f \log \frac{1}{x} = \alpha M(f),$$

and by a straightforward calculation

$$M(f) = (1 - \alpha)^{-2}.$$

For this particular f , we thus obtain

$$M(f) - L(f) = (1 - \alpha)M(f) = (1 - \alpha)^{-1},$$

which obviously becomes bigger than d when α is taken from $(1 - d^{-1}, 1)$.

PROPOSITION 2. *For every fixed $c > 1$, inequality (2) is satisfied for all $d \geq 1 + (c - 1)^{-1}e^{-c}$, but for no other d 's. Furthermore, the function f_c , defined in the statement of Theorem 2, is the (essentially) unique f for which*

$$M(f) = cL(f) + (1 + e^{-c}/(c - 1)).$$

PROOF. As was already noticed, we may assume f to be nonnegative. In fact, nothing will be lost if we further assume, as we shall, that $f \geq 1$. We are allowed to do so because, otherwise, replacing f by $f \vee 1$ would leave $L(f)$ unchanged without diminishing $M(f)$, so that validating (2) for $f \vee 1$, would a fortiori validate it also for f .

REMARK. By appealing to the Hardy and Littlewood maximal theorem, we could reduce f even further to be nonincreasing. Such a reduction would have considerably simplified the argument. However, the temptation to take this "easy way out" was overruled by the desire to present a self-contained proof which does not rely on any previous results. Therefore, let $f \geq 1$ (f not necessarily monotone), and for an arbitrarily fixed $c > 1$, using (4) to begin with, calculate as follows:

$$\begin{aligned} (6) \quad M(f) &= \int f \log \frac{1}{x} = c \int f \log \frac{1}{x^{1/c}} \\ &= c \int f \left(\log f + \log \frac{1}{x^{1/c} f} \right) = cL(f) + c \int f \log \frac{1}{x^{1/c} f} \\ &= cL(f) + c \int_{\{f > 1\}} f \log \frac{1}{x^{1/c} f} + \int_{\{f = 1\}} \log \frac{1}{x}. \end{aligned}$$

Letting $d(f, c) = M(f) - cL(f)$, (6) yields

$$\begin{aligned} (7) \quad d(f, c) &= c \int_{\{f > 1\}} f \log \frac{1}{x^{1/c} f} + \int_{\{f = 1\}} \log \frac{1}{x} \\ &= c \left(\int_{\{f > 1\}} f \log \frac{1}{x^{1/c} f} + \int_{\{f = 1\}} \log \frac{1}{x^{1/c}} \right). \end{aligned}$$

Consequently, $c^{-1}d(f, c)$ has the structure

$$(8) \quad \frac{d(f, c)}{c} = \int_{\{f > 1\}} g_f + \int_{\{f = 1\}} h,$$

where

$$h(x) = \log \frac{1}{x^{1/c}}$$

and

$$g_f(x) = f(x) \log \frac{1}{x^{1/c} f(x)},$$

or, equivalently, $g_f = hf - f \log f$.

Now, apply (5) at $y = (x^{1/c}f)^{-1}$ to g_f to obtain

$$(9) \quad g_f(x) \leq e^{-1}x^{-1/c},$$

with equality iff $f(x) = e^{-1}x^{-1/c}$ a.e. (9) applied to (8) yields

$$(10) \quad \frac{d(f, c)}{c} \leq \frac{1}{e} \int_{\{f > 1\}} x^{-1/c} + \int_{\{f=1\}} h,$$

with equality iff $f(x) = e^{-1}x^{-1/c}$ a.e. on $\{f > 1\}$.

Next, to further majorize the right-hand side of (10) (keeping track of conditions for equality), note again by (5) that $e^{-1}x^{-1/c} \geq h(x)$ for all x , so it would be advantageous to take the range $\{f > 1\}$ in the first integral of (10) to be as large as possible. By (10) this means taking

$$\{f > 1\} = \{x: e^{-1}x^{-1/c} > 1\} = (0, e^{-c}).$$

With this maximal range, (10) yields

$$(11) \quad \frac{d(f, c)}{c} \leq \frac{1}{e} \int_0^{e^{-c}} x^{-1/c} + \int_{e^{-c}}^1 \log \frac{1}{x^{1/c}} = \frac{1}{c} \left(1 + \frac{e^{-c}}{c-1} \right),$$

with equality iff $f(x) = e^{-1}x^{-1/c}$ a.e. on $(0, e^{-c}]$ and $f(x) = 1$ a.e. otherwise. Multiplying through by c leads to

$$(12) \quad d(f, c) \leq 1 + e^{-c}/(c-1) \quad \text{for all } f \geq 1,$$

with equality only if $f = f_c$ almost everywhere. The proofs of Proposition 2 and, hence, of Theorem 2 are now complete.

Corollary 1 is obtained from Theorem 2 by evaluating $d(c)$ at $c = e(e-1)^{-1}$.

PROOF OF THEOREM 3. By Proposition 1, no $c \leq 1$ qualifies for (2) and a fortiori for (3), because $K(f) \leq L(f)$. Fix $c > 1$ and for $f \geq 0$ proceed initially as in (6) to obtain (as in the third row of (6))

$$(13) \quad M(f) = cK(f) + c \int f \log \frac{1}{x^{1/c}f}.$$

Apply (5) with $y = (x^{1/c}f)^{-1}$ to the last term in (13) to get

$$(14) \quad M(f) \leq cK(f) + \frac{c}{e} \int x^{-1/c} = cK(f) + \frac{c^2}{e(c-1)}$$

with equality iff $f = f_c$ almost everywhere. $c > 1$ combined with (20) is precisely the claim of Theorem 3, whose proof is thus complete. Corollary 2 is now obtained by solving $d(c) = e^{-1}c^2(c-1)^{-1} = c$ for c . The solution is obviously $c = c_D = e(e-1)^{-1}$.

A comment on extremal martingales. Using, for example, a construction as in Dubins and Gilat [6], one can now obtain, for each $c > 1$, a martingale, say $\{X_t\}$, with last element, say X , for which

$$(15) \quad E \sup |X_t| = cE|X| \log^+ |X| + 1 + (c-1)^{-1}e^{-c}$$

as well as one for which

$$(16) \quad E \sup_t |X_t| = cE|X| \log |X| + c^2(c-1)^{-1}e^{-1}.$$

Such extremal martingales are by no means unique. However, what they all must have in common is the distribution of the absolute value of their terminal element as well as the distribution of the time-supremum of their absolute values. Specifically, for (15) (or (16)) to hold, the percentile function of $|X|$ must be f_c (or f^c , respectively), and the percentile function of $\sup_t |X_t|$ must be the HL maximal function associated with f_c (or f^c , respectively). On the other hand, for no martingale can the left-hand side of either (15) or (16) exceed the corresponding right-hand side.

CONCLUDING REMARK. It is instructive to compare the proofs of Theorems 2 and 3. One key idea in both is the elementary inequality $\log y \leq e^{-1}y$. In fact, this is the only substantive element in the proof of Theorem 3. Basically, this very argument is used by Hardy and Littlewood, as well as by Doob, to obtain their $L \log^+ L$ inequalities. While this argument, as shown, is strong enough to yield a sharp bound in the $L \log L$ (without the plus) case, it leaves a gap in the original $L \log^+ L$ case. Discovering the exact size of this gap required an additional idea. This idea is the main point of the present paper.

ACKNOWLEDGEMENTS. I am indebted to Wilbert Kallenberg of Vrije Universiteit, Amsterdam for an idea which led to the discovery of the crucial last row in formula (6). I would also like to thank Piet Holewijn and Don van der Vecht, both of Vrije Universiteit, for helping me see through some tedious calculations.

REFERENCES

1. D. Blackwell and L. E. Dubins, *A converse to the dominated convergence theorem*, Illinois J. Math. **7** (1963), 508–514.
2. (i) D. L. Burkholder, *Successive conditional expectation of an integrable function*, Ann. Statist. **33** (1962), 887–893.
(ii) Private communication.
3. D. C. Cox, *Some sharp martingale inequalities related to Doob's inequality*, Inequalities in Statist. and Probability, IMS Lecture Notes Monograph Series, vol. 4, 1984, pp. 1–6.
4. D. C. Cox and R. P. Kertz, *Prophet regions and sharp inequalities for absolute moments of martingales*, J. Multivariate Anal. (to appear).
5. J. L. Doob, *Stochastic processes*, Wiley, New York, 1953.
6. L. E. Dubins and D. Gilat, *On the distribution of maxima of martingales*, Proc. Amer. Math. Soc. **68** (1978), 337–338.
7. R. F. Gundy, *On the class $L \log L$, martingales, and singular integrals*, Studia Math. **33** (1969), 109–118.
8. G. H. Hardy and J. E. Littlewood, *A maximal theorem with function-theoretic applications*, Acta Math. **54** (1930), 81–116.
9. G. H. Hardy, J. E. Littlewood and G. Polya, *Inequalities* (2nd ed.), Cambridge Univ. Press, Cambridge, 1952.
10. I. Meilijson and A. Nadas, *Convex majorization with an application to the length of critical paths*, J. Appl. Probab. **16** (1979), 671–677.

SCHOOL OF MATHEMATICAL SCIENCES, TEL-AVIV UNIVERSITY, RAMAT-AVIV 69 978, TEL-AVIV, ISRAEL