

## ON AN INEQUALITY FOR CONVEX FUNCTIONS

H. D. BRUNK

1. **Introduction.** In an interesting note [6], E. M. Wright refers to recent mention in the literature of the inequality

$$(1) \quad \sum_{k=1}^m (-1)^{k-1} f(a_k) \geq f\left(\sum_{k=1}^m (-1)^{k-1} a_k\right),$$

where

$$(2) \quad a_1 \geq a_2 \geq \cdots \geq a_m \geq 0,$$

and points out that its validity for continuous convex  $f$  vanishing at the origin is a consequence of Theorem 108 in [5].

It is well known that if  $G(x)$  is a distribution function (monotone nondecreasing,  $G(-\infty) = 0$ ,  $G(\infty) = 1$ ) then

$$(3) \quad \int f(x) dG(x) \geq f\left(\int x dG(x)\right)$$

for all continuous convex  $f$ . The similarity in form of (1) and (3) suggests a formulation including both. Conditions on the function  $G(t)$  which are necessary and sufficient in order that

$$(4) \quad \int_{[a,b]} f[X(t)] dG(t) \geq f\left[\int_{[a,b]} X(t) dG(t)\right]$$

for all continuous convex  $f$ , where  $X(t)$  is a given (finite) nondecreasing function on a closed interval  $[a, b]$ , can be given immediately from a theorem (stated in §3) of Hardy, Little wood and Pólya. It is the purpose of this note to rephrase these conditions as a condition on a nondecreasing approximant to  $G(t)$ . The condition is that the nondecreasing approximant to  $G(t)$ , defined in §2, shall be a distribution function.

2. **A nondecreasing approximant to a function.** Let  $X(t)$  be nondecreasing on the interval  $[a, b]$ , and let  $G(t)$  be measurable with respect to the Lebesgue-Stieltjes measure induced by  $X(t)$ ; in particular  $G$  will have this property if it is Borel measurable, in particular,

---

Presented to the Society, December 28, 1955; received by the editors May 18, 1955.

again, if it is a function of bounded variation. If  $a \leq c \leq d \leq b$ , let  $\{c, d\}$  denote any of the four intervals  $[c, d]$ ,  $(c, d]$ ,  $[c, d)$ ,  $(c, d)$ , with the convention that if  $c < u < d$  then  $\{c, u\}$ ,  $\{u, d\}$  are disjoint and  $\{c, u\} \cup \{u, d\} = \{c, d\}$ . For  $t_0 \in (a, b)$ , define

$$(5) \quad G^*(t_0) = G_X^*(t_0) = \sup_u \inf_{v, t_0 \in \{u, v\}} M\{u, v\},$$

where

$$(6) \quad M\{u, v\} = \int_{\{u, v\}} G(t) dX(t) / \int_{\{u, v\}} dX(t)$$

(in (5) only intervals  $\{u, v\}$  are to be considered which contain  $t_0$  and for which  $\int_{\{u, v\}} dX(t) > 0$ ). For convenience in stating the condition given in Theorem 1, we shall define

$$(7) \quad G^*(a) = \min \left\{ G(a), \inf_v M[a, v] \right\}$$

and

$$(8) \quad G^*(b) = \max \left\{ G(b), \sup_u M\{u, b\} \right\}.$$

Note that if  $X(t)$  has a discontinuity at  $a(b)$  then (7) ((8)) coincides with (5) for  $t_0 = a(b)$ .

The function  $G^*$  has occurred in investigations of certain minimum problems suggested by the maximum likelihood estimation of ordered parameters [1; 2; 3]. Certain of its properties are listed here for easy reference.

$$(9) \quad G^*(t_0) = \inf_v \sup_{u, t_0 \in \{u, v\}} M\{u, v\}, \quad t_0 \in (a, b).$$

$$(10) \quad G^*(t) \text{ minimizes the integral } \int_{[a, b]} [G(t) - g(t)]^2 dX(t)$$

in the class of nondecreasing functions  $g(t)$ . (The function  $G^*(t)$  also minimizes, in this class of functions, the integral obtained by replacing the squared difference in the integrand by any of a class of functions sharing appropriate properties of the squared difference.)

$G^*(t)$  is constant on each of a countable set of intervals

$$(11) \quad \{u_i, v_i\}: G^*(t) = M\{u_i, v_i\} \text{ for } t \in \{u_i, v_i\}; \text{ elsewhere, except on a set of } X\text{-measure } 0, G^*(t) \equiv G(t).$$

- (12) If  $\{u, v\}$  is a maximal interval on which  $G^* \equiv c$ , then for each interval  $\{\xi, v\}$  of positive  $X$ -measure we have  $M\{\xi, v\} \leq c$ , and for each interval  $\{u, \eta\}$  of positive  $X$ -measure we have  $M\{u, \eta\} \geq c$ .

The following is an immediate consequence of (11):

$$(13) \quad \int_{[a, b]} G^*(t) dX(t) = \int_{[a, b]} G(t) dX(t).$$

The following characterization of  $G^*$  when  $X(t) \equiv t$ , and its generalization for the case of arbitrary nondecreasing  $X(t)$ , are due to W. T. Reid [2].

- (14) If  $X(t) \equiv t$ , then  $\int_a^x G^*(t) dt$  is the maximal convex function dominated by  $\int_a^x G(t) dt$ ;  $\int_x^b G^*(t) dt$  is the minimal concave function dominating  $\int_x^b G(t) dt$ .

Either (14) or (11) and (12), with Bonnet's form of the second law of the mean, may be used to prove the following property of  $G^*$ .

- (15) If  $X(t) \equiv t$ , and if  $h(x)$  is non-negative and nondecreasing, then  $\int_a^b h(t) G(t) dt \leq \int_a^b h(t) G^*(t) dt$ .

In Theorem 1 below,  $X(t)$  is a given nondecreasing function, finite, hence bounded on  $[a, b]$ . The theorem gives necessary and sufficient conditions on  $G$ , expressed in terms of  $G^*$ , in order that (4) shall hold for all continuous convex  $f$ . The integrals  $\int_{[a, b]} X dG$  and  $\int_{[a, b]} G dX$  appearing in (4) and (6) are related through the formula for integration by parts:

$$\int_I X dG + \int_I G dX = \int_I d(XG)$$

for an arbitrary interval  $I$  on which  $X$  and  $G$  are of bounded variation. In order for this to hold, and also to avoid minor difficulties involved in the determination of  $G$  at common points of discontinuity of  $X$  and  $G$ , we shall suppose that  $X$  is continuous from the left (right) and that  $G$  is continuous from the right (left).

The function  $G$  is to be of bounded variation on  $[a, b]$ , hence bounded on  $[a, b]$ .

If (4) is to hold when  $f$  is identically constant, then  $\int_{[a, b]} dG(t) = 1$ ; we may then suppose without loss of generality that  $G(b) = 1$ ,  $G(a) = 0$ .

**THEOREM 1.** *Let  $X(t)$  be nondecreasing and continuous from the left*

(right) on  $[a, b]$ . Let  $G(t)$  be of bounded variation and continuous from the right (left) on  $[a, b]$ , with  $G(a) = 0$ ,  $G(b) = 1$ . A necessary and sufficient condition that

$$(4) \quad \int_{[a,b]} f[X(t)]dG(t) \geq f \left[ \int_{[a,b]} X(t)dG(t) \right]$$

for every continuous convex  $f$  is that  $G^*$  be a distribution function on  $[a, b]$ , that is, that  $G^*(a) = 0$ ,  $G^*(b) = 1$ .

One has immediately, on applying the theorem to  $f - c\phi$  for appropriate constants  $c$ ,

$$\inf \frac{f''(x)}{\phi''(x)} \leq \frac{\int_{[a,b]} f[X(t)]dG(t) - f \left[ \int_{[a,b]} X(t)dG(t) \right]}{\int_{[a,b]} \phi[X(t)]dG(t) - \phi \left[ \int_{[a,b]} X(t)dG(t) \right]} \leq \sup \frac{f''(x)}{\phi''(x)},$$

where the sup and inf are over the range of  $X(t)$  for  $t \in [a, b]$ , if  $f''$  and  $\phi''$  exist on this range and if  $G^*(t)$  is a distribution function on  $[a, b]$ . On specializing  $\phi$ , say, to be  $x^2/2$ , one has an estimate of the difference between the left and right members of (4) in terms of  $f''$ .

Theorem 1 can be shown to be a consequence of a theorem of Hardy, Littlewood and Pólya, discussed in §3. A proof of Theorem 1 as a consequence of the validity of (4) for distribution functions  $G$  is given in §4.

**3. An inequality of Hardy, Littlewood and Pólya.** The following theorem of Hardy, Littlewood and Pólya [4] (cf. also [5, Theorem 249, p. 170]) is here restated in Stieltjes integral form. For a real number  $x$ , let  $x^+$  denote the larger of the two numbers,  $x$  and 0.

**THEOREM A** (Hardy, Littlewood, and Pólya). *Let  $G(t)$ ,  $H(t)$  be functions of bounded variation on  $[a, b]$ ,  $[c, d]$  respectively. Let  $x(t)$ ,  $Y(t)$  be bounded, Borel measurable functions on  $[a, b]$ ,  $[c, d]$  respectively. Necessary and sufficient conditions in order that*

$$(16) \quad \int_{[a,b]} f[X(t)]dG(t) \geq \int_{[c,d]} f[Y(t)]dH(t)$$

for every function  $f$  continuous and convex on a closed interval containing the ranges of  $X(t)$  and  $Y(t)$  are that

$$(17) \quad \int_{[a,b]} dG(t) = \int_{[c,d]} dH(t),$$

$$(18) \quad \int_{[a,b]} X(t) dG(t) = \int_{[c,d]} Y(t) dH(t),$$

and

$$(19) \quad \int_{[a,b]} [X(t) - u]^+ dG(t) \geq \int_{[c,d]} [Y(t) - u]^+ dH(t)$$

for all real  $u$ .

PROOF. Hardy, Littlewood, and Pólya introduce the term "angle" for the function  $f(x) = [x - u]^+$ , where  $u$  is fixed. The validity of (17), (18), and (19) is necessary in order that (16) should hold for linear functions and angles, and the sufficiency follows from the observation that an arbitrary function continuous and convex on a closed, bounded interval may there be approximated uniformly by the sum of a linear function and a finite sum of positive multiples of angles.

#### 4. Proof of Theorem 1. Set

$$(20) \quad \alpha = \int_{[a,b]} X(t) dG(t).$$

Since  $G(b) = 1$ ,  $G(a) = 0$ , we have

$$\alpha = X(a) + \int_{[a,b]} [X(t) - X(a)] dG(t).$$

Applying the formula for integration by parts, we have

$$(21) \quad \alpha = X(a) + \int_{[a,b]} [1 - G(t)] dX(t).$$

In Theorem A, set  $H(t) \equiv G(t)$ ,  $c = a$ ,  $d = b$ , and  $Y(t) \equiv \alpha$ ; then (17) and (18) are satisfied. Equation (16) becomes (4), and the right member of (19) becomes  $(\alpha - u)^+$ . Thus a necessary and sufficient condition that (4) hold for all continuous convex  $f$  is that

$$(19') \quad \int_{[a,b]} [X(t) - u]^+ dG(t) \geq (\alpha - u)^+$$

for all real  $u$ . It remains to verify that (19') is equivalent to the condition  $G^*(a) = 0$ ,  $G^*(b) = 1$ , or equivalently, in view of (7) and (8),  $\inf_v M[a, v] \geq 0$ ,  $\sup_v M\{v, b\} \leq 1$ . If

$$\sup_v M\{v, b\} = \sup_v \int_{[v,b]} G(t) dX(t) / \int_{[v,b]} dX(t) \leq 1,$$

then  $\int_{\{v, b\}} [1 - G(t)] dX(t) \geq 0$  for  $\{v, b\} \subset [a, b]$ . On integrating by parts one finds that

$$(22) \quad \int_{\{v, b\}} [X(t) - X(v \pm)] dG(t) \geq 0,$$

where

$$\begin{aligned} X(v \pm) &= X(v+) \text{ if } \{v, b\} = (v, b], \\ X(v \pm) &= X(v-) = X(v) \text{ if } \{v, b\} = [v, b]. \end{aligned}$$

Hence

$$(23) \quad \int_{[a, b]} [X(t) - u]^+ dG(t) \geq 0$$

if there exists a number  $v$  such that  $u = X(v+)$  or  $u = X(v)$ . If for some  $v$  we have  $X(v) < u < X(v+)$ , then the application of (22) to  $[v, b]$  and to  $(v, b]$  yields (23). If  $\inf_v \int_{[a, v]} G(t) dX(t) / \int_{[a, v]} dX(t) \geq 0$  then  $\int_{[a, v]} G(t) dX(t) \geq 0$  for  $[a, v] \subset [a, b]$ , and from (21) we have

$$\begin{aligned} \int_{\{v, b\}} [1 - G(t)] dX(t) &= \alpha - X(a) - \int_{[a, v]} [1 - G(t)] dX(t) \\ &= \alpha - X(a) - \int_{[a, v]} dX(t) + \int_{[a, v]} G(t) dX(t) \\ &= \alpha - X(v \pm), \end{aligned}$$

where  $X(v \pm) = X(v+)$  or  $X(v-) = X(v)$  according as  $[a, v] = [a, v]$  or  $[a, v] = [a, v)$ ; i.e., according as  $\{v, b\} = (v, b]$  or  $\{v, b\} = [v, b]$ . As above, this implies that

$$(24) \quad \int_{[a, b]} [X(t) - u]^+ dG(t) \geq \alpha - u,$$

and (23) and (24) together yield (19'). The converse, that (19') implies  $\inf_v M[a, v] \geq 0$ ,  $\sup_v M\{v, b\} \leq 1$ , similarly follows from the identity

$$\int_{\{v, b\}} [1 - G(t)] dX(t) = \int_{\{v, b\}} [X(t) - X(v \pm)] dG(t).$$

This completes the proof of Theorem 1.

The above proof exhibits Theorem 1 as a consequence of Theorem A of Hardy, Littlewood and Pólya. The necessity of the condition on  $G$  follows from the application to appropriate angles of the in-

equality. The sufficiency of the condition can also be seen as a consequence of the well known theorem to the effect that (4) is valid for continuous convex  $f$  when  $G$  is a distribution function. We indicate the proof for the case  $X(t) \equiv t$ . A change of variable then yields Theorem 1.

If  $G^*(t)$  is a distribution function, then

$$\int_{[a,b]} f(t) dG(t) \geq f \left[ \int_{[a,b]} t dG(t) \right]$$

for every continuous convex  $f$ .

We have

$$\begin{aligned} \int_{[a,b]} f(t) dG(t) &= \int_{[a,b]} [f(t) - f(a)] dG(t) + f(a) \\ &= f(a) + \int_a^b f'(t) [1 - G(t)] dt. \end{aligned}$$

By (15),

$$\int_a^b f'(t) [1 - G(t)] dt \geq \int_a^b f'(t) [1 - G^*(t)] dt,$$

hence

$$\begin{aligned} \int_a^b f(t) dG(t) &\geq \int_a^b f(t) dG^*(t) \geq f \left[ \int_a^b t dG^*(t) \right] \\ &= f \left[ a + \int_a^b [1 - G^*(t)] dt \right]. \end{aligned}$$

The last expression is equal to  $f[a + \int_a^b [1 - G(t)] dt]$  by (13), hence  $\int_{[a,b]} f(t) dG(t) \geq f[\int_{[a,b]} t dG(t)]$ .

In each of the following corollaries, Corollary 1 and Corollary 2, it is assumed that  $X(t)$  is nondecreasing and continuous from the left on  $[a, b]$ , that  $f(x)$  is convex and continuous on  $[X(a), X(b)]$ , that  $G(t)$  is of bounded variation and continuous from the right on  $[a, b]$ , and that  $G(b) = 1$ .

**COROLLARY 1.** *If  $X(a) = 0$ , if  $f(0) \leq 0$ , and if  $G^*(a) \geq 0$ ,  $G^*(b) = 1$ , then*

$$(4) \quad \int_{[a,b]} f[X(t)] dG(t) \geq f \left[ \int_{[a,b]} X(t) dG(t) \right].$$

PROOF. Set  $G_1(t) = G(t)$  for  $t > a$ ,  $G_1(a) = 0$ . Then  $G_1^*(a) = 0$ ,  $G_1^*(b) = 1$ , hence

$$\int_{[a,b]} f[X(t)] dG_1(t) \geq f \left[ \int_{[a,b]} X(t) dG_1(t) \right].$$

But  $\int_{[a,b]} X(t) dG_1(t) = \int_{[a,b]} X(t) dG(t)$ , since  $X(a) = 0$ ,  $\int_{[a,b]} f[X(t)] dG_1(t) = f(0)G(a) + \int_{[a,b]} f[X(t)] dG(t)$ , whence (4) follows.

COROLLARY 2. *If either (i)  $G(a) = 0$ , or (ii)  $X(a) = 0$ ,  $f(0) \leq 0$ , and if (iii)  $0 \leq G(t) \leq 1$  for  $a < t < b$ , then*

$$(4) \quad \int_{[a,b]} f[X(t)] dG(t) \geq f \left[ \int_{[a,b]} X(t) dG(t) \right].$$

If (i) and (iii) hold, then  $G^*$  is a distribution function. If (ii) and (iii) hold, the hypotheses of Corollary 1 are satisfied.

COROLLARY 3. *Let  $f(x)$  be continuous, convex on  $[0, b]$ , with  $f(0) \leq 0$ ; let  $b \geq a_1 \geq a_2 \geq \dots \geq a_m \geq 0$ , and let  $1 \geq h_1 \geq h_2 \geq \dots \geq h_m \geq 0$ . Then*

$$(25) \quad \sum_{k=1}^m (-1)^{k-1} h_k f(a_k) \geq f \left[ \sum_{k=1}^m (-1)^{k-1} h_k a_k \right].$$

For inequality (4) becomes (25) if  $X(t) \equiv t$ ,  $a = 0$ , and if  $G(t)$  is a step function having saltus  $(-1)^{k-1} h_k$  at  $t = a_k$ ,  $k = 1, 2, \dots, m$ , with  $G(b) = 1$ . Note that if  $h_1 = h_2 = \dots = h_m = 1$  then (25) becomes (1).

#### REFERENCES

1. M. C. Ayer, H. D. Brunk, G. M. Ewing, W. T. Reid, and Edward Silverman, *An empirical distribution function for sampling with incomplete information*, to be published in *Ann. Math. Statist.* vol. 26 (1955) pp. 641-647.
2. H. D. Brunk, G. M. Ewing, and W. T. Reid, *The minimum of a certain definite integral suggested by the maximum likelihood estimate of a distribution function*, *Bull. Amer. Math. Soc.* Abstract 60-6-684.
3. H. D. Brunk, G. M. Ewing, and W. R. Utz, *Minimizing integrals in certain classes of monotone functions*, to appear in *Pacific Journal of Mathematics*.
4. G. H. Hardy, J. E. Littlewood, and G. Pólya, *Some simple inequalities satisfied by convex functions*, *Messenger of Math.* vol. 58 (1928) pp. 145-152.
5. ———, *Inequalities*, Cambridge, 1934.
6. E. M. Wright, *An inequality for convex functions*, *Amer. Math. Monthly* vol. 61 (1954) pp. 620-622.

UNIVERSITY OF MISSOURI