ON CHEBYSHEV'S OTHER INEQUALITY

BY A. M. FINK and MAX JODEIT, JR.
Iowa State University and University of Minnesota

We formulate the notion of a best possible inequality. This involves finding the largest class of functions and measures for which an inequality is true. We give two examples of Chebyshev's inequality, e.g. $\int_a^b d\mu \int_a^b fg d\mu \geq \int_a^b fd\mu \int_a^b gd\mu$ for all pairs (f,g) which are increasing if and only if $\int_a^x du \geq 0$, $\int_x^b du \geq 0$ for all x. Other examples include Jensen's inequality.

1. Introduction.   Let μ be a probability measure on the real line and f and g increasing functions. Then

$$(1.1) \qquad \int_{\mathcal{R}} fg \, d\mu \geq \int_{\mathcal{R}} f d\mu \int_{\mathcal{R}} g \, d\mu$$

says that the random variables f and g are positively correlated. This is Chebyshev's 'other' inequality.

It is common to ask if an inequality is best possible. In most instances this means having the largest (or smallest) constant(s) for which the inequality holds and settling the cases of possible equality. For (1.1) equality holds if one of the functions is a constant or the measure is a point mass.

In this paper we would like to explore a different meaning of 'best possible.' In order to formulate our ideas in the context of inequality (1.1), consider a related version

$$(1.2) \qquad \int_a^b d\mu \int_a^b fg \, d\mu \geq \int_a^b f d\mu \int_a^b g \, d\mu$$

where $[a,b]$ is any real interval. It was already observed by Andreief (1883), that (1.2) holds under the hypothesis that

$$(1.3) \qquad [f(x) - f(y)][g(x) - g(y)] \geq 0 \quad \text{for all } (x,y) \in [a,b] \times [a,b],$$

and

$$(1.4) \qquad \mu \text{ is a non-negative measure.}$$

The condition (1.3) is read "f and g are similarly ordered," see Hardy, Littlewood, and Pólya ((1952), p. 43). (More history of the inequality (1.2) appears in the article by Mitrinović and Vasić (1974).) It is clear that (1.3) is satisfied if both f and g are increasing.

Our viewpoint is that the inequality (1.2) has "two variables," the pairs of functions and the measures. 'Best possible' should mean that:

(A) the inequality (1.2) holds for all similarly ordered pairs if and only if μ is a non-negative measure, and

(B) the inequality (1.2) holds for all non-negative measures if and only if f and g are similarly ordered.

We will show below that both statements are correct. This means that each class, similarly ordered pairs, and non-negative measures, is the largest class for which the inequality can be proved, given that it must hold for all elements in the other class.

Contrast this with the condition

(1.5) f,g are both increasing or both are decreasing, or one is a constant.

The set of pairs satisfying (1.5) is smaller than the set of similarly ordered pairs, so the requirement that (1.2) holds for such pairs is a less restrictive condition on the measure μ which is to satisfy it. In fact, for some signed measures μ inequality (1.2) holds for all pairs satisfying (1.5). We will derive a condtion to replace $\mu \geq 0$ for which the appropriate versions of (A) and (B) hold.

In summary, what we are looking for is a class of measures M so that the following statement is true.

"The inequality (1.2) holds for all pairs f,g satisfying (1.5) if and only if $\mu \in M$."

Thus M is the largest class of measures for which we can prove the inequality (1.2) under the conditions (1.5). It is then natural to ask for the largest class of pairs of functions for which inequality (1.2) holds for all measures in M. Our view is that 'best possible' should mean that the conditions are those in (1.5).

In the succeeding sections we formulate this question in general and give several examples of this phenomenon.

2. Function-Measure Duality.   Let F be a class of functions and M be a class of measures and $J(f) \geq 0$ be an integral inequality.

The classes F and M are said to be in duality with respect to J if (I) $J(f) \geq 0$ for all $f \in F$ if and only if the measure is in M, and (II) $J(f) \geq 0$ for all measures in M if and only if $f \in F$. This is analogous to the situation, in the theory of locally convex topological linear spaces, leading to weak and weak* topologies on $\mathcal{B}$ and $\mathcal{B}^*$. Each formulation of classes in duality requires specifying the universe of functions or measures. For us here, all functions are to be Borel measurable and all measures are regular Borel (signed) measures.

It is important to notice that if the integrals are linear in the measure, the class M being as large as possible will be a cone. If the inequality is convex in the functions, then F will also be a cone.

As a simple example, the class F_+ of non-negative functions on an interval $[a,b]$ are in duality with the class M_+ of non-negative measures with respect to the inequality

(2.1) $\int_a^b f d\mu \geq 0$.

The proof is straightforward and is omitted.

The inequality (2.1) is a point of contact with known theory. If C is a cone of functions in a Banach space X, then the measures $M \subset X^*$ for which (2.1) holds is called the conjugate cone C^*, see Kelly and Namioka (1963). If then we look at the class of functions in X for which (2.1) holds for all measures in C^*, then this class might be called $*(C^*)$. We are interested in cones C for which $*(C^*) = C$.

As a second example, let $(FI)_+$ be the class of non-negative increasing functions on $[a,b]$ and M_0 be the class of measures μ such that $\int_x^b d\mu \geq 0$ for all $x \in [a,b]$. Then $(FI)_+$ and M_0 are in duality with respect to the inequality (2.1). The proof is omitted.

3. Chebyshev's Inequality.   We return to the inequality (1.2) which is our main motivation. Observe that if μ is replaced by $-\mu$ the inequality is unchanged. For the two theorems on this inequality we will assume that μ is somewhere positive.

THEOREM 1.   *Let SO be the pairs of functions which are similarly ordered and M_+ the set of non-negative measures on $[a,b]$. Then SO and M_+ are in duality with respect to the*

inequality (1.2). Equality holds for a pair in SO and a measure in M_+ if and only if one of the functions is constant a.e. μ.

Proof. The inequality

(3.1) $\qquad \frac{1}{2}\int_a^b\int_a^b [f(x)-f(y)][g(x)-g(y)]d\mu(x)d\mu(y) \geq 0$

is obvious under the assumptions that the pair (f,g) are similarly ordered (see (1.3)) and $\mu \geq 0$. This establishes the sufficiency in both (I) and (II), since if the expression in (3.1) is expanded one gets (1.2). For the necessity, let $\mu = \delta_x + \delta_y$, $x \neq y$. Then the inequality (1.2) is exactly (1.3). Finally, to show that $\mu \geq 0$ if (1.2) holds for all similarly ordered pairs, take $f = g = X_{[c,d]}$, the indicator function of an interval $[c,d]$. Then (1.2) is $\mu[a,b]\mu[c,d] \geq (\mu[c,d])^2 \geq 0$. Thus μ is always the same sign. Since μ is somewhere positive, it is positive everywhere.

To establish the cases of equality, observe that if A and B are any sets of μ positive measure then (3.1) with equality implies that $[f(x)-f(y)][g(x)-g(y)] = 0$, $x \in A$, $y \in B$. Assume that f is not a constant a.e. μ. If further one assumes that $f^{-1}\{\alpha\}$ is always a set of μ measure zero, let A and B be disjoint positive measure sets. Then $f(x) - f(y) \neq 0$ a.e. $x \in A$, $y \in B$ and thus $g(x) \equiv g(y)$ a.e. $x \in A$, $y \in B$. This implies that g is a constant. If there is an α so that $f^{-1}(\alpha) = A$ has positive measure, then take B to be the complement of A and the above argument gives g a constant. $\qquad\square$

For our second example on Chebyshev's inequality we consider the class SM (for similarly monotone), those pairs of functions satisfying (1.5). The corresponding measures EP (for end-positive) are those measures μ such that

$$\int_a^x d\mu \geq 0 \quad \text{and} \int_x^b d\mu \geq 0 \quad \text{for all } x \in [a,b]$$

and $\int_a^b d\mu \neq 0$. This class will reappear in a different example.

THEOREM 2. *The pair SM and EP are in duality with respect to inequality (1.3). Equality holds if and only if (when f,g are right continuous)*

(+) $\qquad$ sup supp $R\,dg \leq$ inf supp $L\,df$, *and* sup supp $R\,df \leq$ inf. supp $L\,dg$.

$\qquad$ *If either pair of supports meet, their common point is a set of measure zero for at least one of the measures.*

Here sup $\phi \equiv a$, inf $\phi \equiv b$, $R(x) = \int_x^b d\mu$ *and* $L(x) = \int_a^x d\mu$.

Proof. We will show the sufficiency of (I) and (II) by writing the inequality in the form (3.1). Assume first that f is increasing and right continuous. Then there is a non-negative measure μ_1 so that for $x < y$,

$$f(y)-f(x) = \mu_1[a,y] - \mu_1[a,x] = \mu_1(x,y] = \int \chi_{(x,y]}(t)d\mu_1(t).$$

Then (3.1) can be written as (μ_2 the measure for g.)

$$\frac{1}{2}\int_a^b\int_a^b (f(y)-f(x))(g(y)-g(x))d\mu(x)d\mu(y)$$

$$= \int_a^b\int_a^b (y-x)_+^0 \int_a^b \chi_{(x,y]}(t)d\mu_1(t) \int_a^b \chi_{(x,y]}(s)d\mu_2(s)d\mu(x)d\mu(y)$$

(integrand is 0 on the diagonal)

$$= \int_a^b\int_a^b\int_a^b\int_a^b (y-x)_+^0 \chi_{(x,y]}(t) \chi_{(x,y]}(s)d\mu(x)d\mu(y)d\mu_1(t)d\mu_2(s).$$

The part of the integrand involving x is expressible as

$$\chi(y \wedge t \wedge s > x)\chi(t \vee s \leq y).$$

where χ (P(variables)) denotes the indicator functions of the set of variables for which P(variables) is true. This gives

$$\int_a^b\int_a^b\int_a^b \chi(t \vee s \leqslant y)L(y \wedge t \wedge s)d\mu(y)d\mu_1(t)d\mu_2(s)$$
$$= \int_a^b\int_a^b\int_a^b L(t \wedge s)X(t \vee s \leqslant y)d\mu(y)d\mu_1(t)d\mu_2(s)$$
$$= \int_a^b\int_a^b L(t \wedge s)R(t \vee s)d\mu_1(t)d\mu_2(s),$$

where $L(x) = \mu([a,x))$, $R(x) = \mu([x,b])$. Thus (3.1) is equivalent to

(3.2) $\int_a^b\int_a^b L(t \wedge s)R(t \vee s)d\mu_1(t)d\mu_2(s) \geqslant 0$.

This is clearly non-negative. Since any increasing function is the pointwise limit of right continuous functions, the sufficiency is shown.

Conversely, if (3.1) holds for all $(f,g) \in$ SM, then as above, $\int_a^b\int_a^b L(t \wedge s)R(t \vee s)d\mu_1(t)d\mu_2(s) \geqslant 0$ for all non-negative measures μ_1, and μ_2. Thus $L(x_1)R(x_2) \geqslant 0$ if $x_1 \leqslant x_2$. Since $L(x) + R(x) \equiv \mu(I)$ we have $L(x)R(x) \geqslant 0$. If $\mu(I) \neq 0$, $L(x)$ and $R(x)$ have the same sign as $\mu(I)$ unless one of them is 0. The case $\mu(I) = 0$ cannot occur unles $\mu = 0$, for then $0 \geqslant \int f \, d\mu \int g \, d\mu$ for all (f,g) in SM; thus taking $f = g = X_{[a,x)}$ and $f = g = Z_{[a,x]}$ lead to $\mu((x_1,x_2)) = 0$ if $x_1 < x_2$, so $\mu = 0$.

Next we need to show that if (1.2) holds for all end-positive measures μ, the pair (f,g) is in SM.

It may help the exposition to let a, b, c, etc. denote the values taken by f at x, y, z, etc. respectively, and A, B, C and so on denote the corresponding values of g. We need two observations.

First, non-negative measures are end-positive, so f and g are similarly ordered (Theorem 1); i.e., $(a-b)(A-B) \geqslant 0$ for all pairs x, y in I.

If $x < y < z$ the measure $\delta_x - \delta_y + \delta_z$ is end-positive, so by (1.3)

$$aA - bB + cC \geqslant (a-b+c)(A-B+C).$$

This is equivalent to

$$0 \geqslant (a-b)(C-B) + (c-b)(A-B).$$

Neither product can be positive—the other would be negative, all four differences would be non-zero, and similar ordering would imply that both products have the same sign (sgn $(C - B) = $ sgn$(c - b)$, etc.). Thus we have the second observation: for all triples $x < y < z$, $(a-b)(C-B) \leqslant 0$, $(c-b)(A-B) \leqslant 0$. Note that x and z are "separated."

It is enough to show that $(a - b)(c - b) \leqslant 0$, for all triples $x < y < z$, unless g is constant (i.e. $f(y)$ is between $f(x)$ and $f(z)$ if $x < y < z$). Suppose not. Then for some triple $x_0 < y_0 < z_0$, $(a_0 - b_0)(c_0 - b_0) > 0$.

The key argument is this: if $a_0 < b_0$ then by the second observation $(a_0 - b_0)(C_0 - B_0) \leqslant 0$ if $z > y_0$, so $C \geqslant B_0$. Similarly, $A \geqslant B_0$ if $x \leqslant y_0$, so $g(y_0)$ is a global minimum for g. Moreover, $A_0 = B_0 = C_0$. If not, say $A_0 > B_0$. Then $(A_0 - B_0)(a_0 - b_0) < 0$, which contradicts similar ordering. In case $a_0 > b_0$ we get that $g(y_0)$ is a global maximum for g and $A_0 = B_0 = C_0$.

Next (still supporting $a_0 < b_0$) we show $A = A_0$, $x \leqslant x_0$, and $C = C_0$, $z \geqslant z_0$. For, if $A > B_0$ for some $x < x_0$, similar ordering gives $a \geqslant b_0 > a_0$. Apply the "$a_0 > b_0$" case of the key argument to a, a_0, b_0 (the triple being $x < x_0 < y_0$). It gives $A = A_0 = B_0$, which contradicts $A > B_0$. Similarly, $C = C_0$ if $z > z_0$.

If g were not constant there would exist y, $x_0 < y < z_0$, such that $B > A_0 = B_0 = C_0$. We may suppose $x_0 < y < y_0$. Now apply the key argument, with the roles of f and g interchanged, to conclude that $a_0 = b = b_0$, which contradicts $a_0 < b_0$.

Since (f,g) and μ satisfy (1.2) if and only if $(-f,-g)$ and μ do, f is monotone unless g is constant. It follows, using similar ordering, that $(f,g) \in$ SM. $\square$

Finally, we discuss the case of equality. If f or g is constant, equality holds, and so do the conditions (+). We shall assume that neither f nor g is constant. We may assume, too, that each is non-decreasing and non-negative (a constant added to f or g adds equal quantities to both sides).

If $\int d\mu = 0$ then we know $\mu = 0$ (for then $0 \geqslant (\int f d\mu)^2$, where $f = y = \chi_{[a,x)}$, so $L(x) \equiv 0$). Thus we assume $\int d\mu \neq 0$, and we may assume $\int \delta\mu > 0$.

Note that, if $\mu \geqslant 0$, the case of equality is covered in Theorem 1.

A case of interest, which we largely ignore, is that in which $\int f g \, d\mu = 0$ and, say, $\int f \, d\mu = 0$. We shall only consider this under the foregoing assumptions, and the further assumption that f, g are right-continuous. Thus $f(x) = \nu_1[a,x]$, $g(x) = \nu_2[a,x]$, where ν_1, ν_2 are non-negative. Then

$$\int_a^b f d\mu = \int_a^b \int_a^x d\mu_1(t) d\mu(x) = \int_a^b R(t) d\nu_1(t) = 0,$$

so $R d\nu_1 = 0$, and the first condition (+) holds. Also,

$$\int_a^b f g \, d\mu = \int_a^b \int_a^x d\nu_1(t) \int_a^x d\nu_2(s) d\mu(x) = \int\int R(t \vee s) d\nu_1(t) d\nu_2(s)$$

$$= \int\int_{t \leqslant s} \int R(s) d\nu_1(t) d\nu_2(s) = \int_a^b f(s) R(s) d\nu_2(s) = 0,$$

so $f R d\nu_2 = 0$. This implies inf supp $f \geqslant$ sup supp $R d\nu_2$, which gives the second condition in (+). A moment's reflection gives that the third part of (+) also holds. We now assume that none of the integrals in the equality is zero.

Let us verify that (+) is equivalent to equality, if f, g are right-continuous, non-increasing, non-negative, not constant, $\int d\mu > 0$, and none of the integrals in the equality is zero. With $f(x) = \nu_1[a,x]$, $g(x) = \nu_2[a,x]$, we have, as in the sufficiency argument, that

$$0 = \int\int L(s \wedge t) R(s \vee t) \delta\nu_1(t) d\nu_2(s) \text{ (the limits on the integrals are } a+0 \text{ and } b)$$

$$= \int\int_{s \geqslant t} L(t) R(s) d\nu_1(t) d\nu_2(s) + \int\int_{s < t} L(s) R(t) d\nu_1(t) d\nu_2(s)$$

$$= \int_{a+0}^b \int_{a+0}^s L(t) d\nu_1(t) \cdot R(s) d\nu_2(s) + \int_{a+0}^b \int_{a+0}^{t-0} L(s) d\nu_2(s) \cdot R(t) d\nu_1(t),$$

so both terms are zero. Since the roles of ν_1 and ν_2 can be reversed, the last term is still zero if t–0 is replaced by t. Now (e.g.) $\int_{a+0}^s L(t) d\nu_1(t) \geqslant 0$ is non-decreasing and $R d\nu_2 \neq 0$, so $L d\nu_{3-i} = 0$ in $[a, c_i)$, where $c_i = \sup(\text{supp } R d\nu_i)$, $i = 1, 2$. This gives the first two parts of (+). The last part of (+) follows because if (e.g.) $R(c_2) \nu_2\{c_2\} > 0$, then $0 = \int_{a+0}^{c_2} L(t) d\nu_1(t) \geqslant L(c_2) \nu_1\{c_2\}$. Finally, if (+) holds the last iterated integrals are both 0, so equality holds.

If f and g are not right-continuous, we can write $f = f_0 + j$, $g = g_0 + k$, where f_0, g_0 are right-continuous and j, k are left-continuous jump functions. Then equality holds for each pair (f_0, g_0), (f_0, k), (j, g_0), (j, k). The conditions (+) are changed by replacing $R(t)$, $L(t)$ by $R(t+0)$, $L(t+0)$ when they appear with dj or dk.

We remark that if R and L are both positive in (a, b), and continuous, equality can only happen if one of f, g is constant.

4. Further Results and Problems.

In the foregoing, it was essential to have a suitable integral representation for the functions in the class F. Appropriate manipulations then permitted reduction to the fundamental inequality (2.1). We state two more examples of such results here, with problems we hope are of interest.

We let M_p denote the class of functions f with the representation

$$f(x) = \int_0^\infty (x-t)_+^p \, d\nu(t), \quad 0 \leqslant x \leqslant T \leqslant \infty,$$

where $x_+ = \max(0, x)$, $(x-t)_+^0$ means $\chi_{[t,\infty)}(x)$, ν is a non-negative measure, and $p \geqslant 0$.

THEOREM 3.   *If p is an integer, M_p is in duality with the set M_p^* of signed measures μ which satisfy*

$$\int_0^T (x-t)_+^p \, d\mu(x) \geq 0, \quad 0 \leq t \leq T,$$

with respect to the inequality (2.1).

The proof uses the basic spline of Curry and Schoenberg (see Schoenberg (1973), p. 3), and a representation theorem of Bernstein (1926).

Problem: If $p > 0$ is not an integer, prove this.

The difficulty is in proving the representation.

Remark.   This result, with $p = 1$, can be used to prove Theorem 108, page 89, in Hardy, Littlewood and Pólya (1952).

The second result concerns Jensen's inequality:

$$\varphi(\textstyle\int f d\mu / \int d\mu) \leq \int \varphi(f) d\mu / \int d\mu.$$

With appropriate modification of "best possible" in our sense (here we have three "variables"), this inequality is best possible for convex φ, Borel measurable functions, and non-negative measures.

If we restrict f to be monotone, we have:

THEOREM 4.   *The convex functions are in duality with the end-positive measures with respect to Jensen's inequality, when F is the class of monotone functions.*

The conditions for equality are too lengthy to be stated here.

Elsewhere we have shown that

$$\varphi(\textstyle\int_0^1 f d\sigma / \int_0^1 d\sigma) \leq [(\int_0^1 f d\sigma)^p / \int_0^1 f^p \, d\sigma] \int_0^1 w(f) d\sigma / \int_0^1 d\sigma$$

holds for all f monotone and $\varphi \in M_p$ if and only if σ is end-positive, with equality for $\varphi(x) = x^p$.

Problem.   Is this "best possible" in the sense of this paper?

REFERENCES

ANDRÉIEF, C. (1883), Note sur une relation entre les intégrales définies des produits des fonctions, *Mém. Soc. Ci. Bordeaux 3 2*, 1–14.

BERNSTEIN, S. (1926). *Lecons sur les propriétes extrémales et la meilleure approximation des fonctions analytiques d'une variable réelle*, Paris.

FINK, A. M. (1980), Pointwise inequalities and continuation of solutions of an nth order equation, *Tohoku Math. Journal 32*, 169–175.

FINK, A. M. and Jodeit, Max (1976), A generalization of the arithmetic-geometric means inequality, *Proc. Amer. Math. Soc. 61*, 255–261.

HARDY, LITTLEWOOD, PÓLYA (1952), *Inequalities,* 2nd ed. Cambridge University press, London and New York.

KELLEY, J., NAMIOKA (1963), *Linear Topological Spaces,* D. Van Nostrand, Princeton, New Jersey.

MITRINOVIĆ. D. S. and VASIĆ, P. M. (1974), History, Variations and Generalizations of the Cebysev Inequality and the Question of Some Priorities, *Publ. Elek. Fak. Univ. Beograd. No. 461–497*, 1–30.

RIESZ, F. and NAGY, B. (1955), *Functional Analysis,* Frederick Ungar Publ., New York.

SCHOENBERG, I. J. (1973), Cardinal Spline Interpolation, *Regional Conf. Ser. in Applied Math. 12* SIAM, Philadelphia, Pa.