A MODIFICATION OF SCHWARZ'S INEQUALITY WITH APPLICATIONS TO DISTRIBUTIONS[1]

By Sigeiti Moriguti

University of North Carolina and University of Tokyo

Summary. Theorem 1 provides us with a result from which we can derive the modified Schwarz inequality (1.2) or, more generally, (3.8). The formulas hold when $x(t)$ is any nondecreasing function belonging to a certain wide class, and $\bar{\varphi}(t)$ is the right-hand derivative of the "greatest convex minorant" of $\Phi(t)$. The necessary and sufficient conditions for equality to hold are also given. Applications to distribution problems in statistics are discussed in Section 4.

1. Introduction. In a previous paper [1], the author made use of Schwarz's inequality in obtaining the least upper bound of an integral of the form

$$\int_a^b x(t)\varphi(t)\,dt, \tag{1.1}$$

where $x(t)$ is a variable function and $\varphi(t)$ is a given function. However, if the domain of variation of $x(t)$ is limited to the class of nondecreasing functions, then Schwarz's inequality does not give the least upper bound unless $\varphi(t)$ itself is effectively a nondecreasing function, because the equality holds only if $x(t)$ is effectively proportional to $\varphi(t)$.

In the present paper, we shall derive a modified Schwarz inequality

$$\int_a^b x(t)\varphi(t)\,dt \leqq \left\{\int_a^b x(t)^2\,dt\right\}^{\frac{1}{2}} \left\{\int_a^b \bar{\varphi}(t)^2\,dt\right\}^{\frac{1}{2}}, \tag{1.2}$$

where $\bar{\varphi}(t)$ is a nondecreasing function closely related to $\varphi(t)$. It holds for any nondecreasing function $x(t)$ and for any function $\varphi(t)$ in a certain wide class, the equality being satisfied if $x(t) = A\bar{\varphi}(t)$ holds with nonnegative constant coefficient A almost everywhere.

The relation between $\varphi(t)$ and $\bar{\varphi}(t)$ is most simply stated by means of the concept of the "greatest convex minorant," whose definition and some of whose properties we shall state in the next section.

The inequality (1.2) can profitably be generalized to (3.8), which reduces to (1.2) if $\Phi(t)$ is absolutely continuous and $\Phi'(t) = \varphi(t)$.

2. Greatest convex minorant. See for instance ([2] p. 440) and ([3] pp. 91, 94). For a given function $\Phi(t)$ in a closed interval $[a, b]$, we consider the supremum of all convex functions dominated by $\Phi(t)$ in the whole interval. (A convex function is characterized by the fact that the middle point of any chord of its graph

Received 3/13/52, revised 9/24/52.

[1] Presented at the Blacksburg meeting of the Institute of Mathematical Statistics, March 19–21, 1952.

lies above or on the curve.) Let it be denoted by $\bar{\Phi}(t)$ and be called the *greatest convex minorant* of $\Phi(t)$ in the interval $[a, b]$. It is easy to see that $\bar{\Phi}(t)$ itself is a convex function dominated by $\Phi(t)$.

In the following we consider only the case where $\Phi(t)$ is a function of bounded variation in $[a, b]$ and continuous at both ends. Then its greatest convex minorant $\bar{\Phi}(t)$ is bounded above; hence, it is continuous (see the references). For every t in $[a, b]$, $\bar{\Phi}(t) \leqq \Phi(t)$. The equality holds certainly at the end points because of the assumed continuity of $\Phi(t)$ there. The set of values of t for which $\bar{\Phi}(t) = \min \{\Phi(t-0), \Phi(t+0)\}$ is a closed set (due to the definition of $\bar{\Phi}(t)$). Therefore its complementary set, that is the set of values of t for which $\bar{\Phi}(t) < \min \{\Phi(t-0), \Phi(t+0)\}$, is an open set and consequently either empty or the union of a denumerable number of disjunct open intervals. In each of those intervals, if any, $\bar{\Phi}(t)$ is a linear function, as follows from the definition of $\bar{\Phi}(t)$.

As a continuous convex function, $\bar{\Phi}(t)$ has everywhere (except perhaps at the end points) finite left-hand and right-hand derivatives; the left-hand derivative is not greater than the right-hand derivative at the same point. Let us denote, for the sake of definiteness, the right-hand derivative of $\bar{\Phi}(t)$ by $\bar{\varphi}(t)$. Then $\bar{\varphi}(t)$ is of course a nondecreasing function and is consequently continuous except perhaps for a denumerable set of values of t. $\bar{\varphi}(t)$ is a constant in any interval where $\bar{\Phi}(t) < \min \{\Phi(t-0), \Phi(t+0)\}$.

3. Modified Schwarz inequality.

THEOREM 1. *Let $\Phi(t)$ be a function of bounded variation in the closed interval $[a, b]$ and continuous at both ends. Then the relation*

$$\int_a^b x(t)\, d\Phi(t) \leqq \int_a^b x(t)\bar{\varphi}(t)\, dt \tag{3.1}$$

holds for any nondecreasing function $x(t)$ for which the integrals exist and are finite, where $\bar{\varphi}(t)$ is the right-hand derivative of the greatest convex minorant $\bar{\Phi}(t)$ of $\Phi(t)$. The equality in (3.1) *holds if and only if $x(t)$ is a constant in every interval where*

$$\bar{\Phi}(t) < \min \{\Phi(t-0), \Phi(t+0)\} \tag{3.2}$$

and, at every point of discontinuity, if any, of $\Phi(t)$,

$$\begin{aligned} x(t_n) &= x(t_n + 0) \quad \textit{when}\ \Phi(t_n - 0) < \Phi(t_n + 0), \\ &= x(t_n - 0) \quad \textit{nhen}\ \Phi(t_n - 0) > \Phi(t_n + 0). \end{aligned} \tag{3.3}$$

PROOF. Let us first assume that $x(t)$ is bounded in $[a, b]$ and hence of bounded variation. So is $\Phi(t)$. Therefore we can apply the formula of integration by parts (See [4] for a formula which can be adapted to our purpose with a slight modification.) and get

$$\int_a^b x(t \mp 0)\, d\Phi(t) = [x(t)\Phi(t)]_{a-0}^{b+0} - \int_a^b \Phi(t \mp 0)\, dx(t), \tag{3.4}$$

where, at every point of discontinuity, if any, of $\Phi(t)$, we take the smaller of $\Phi(t-0)$ and $\Phi(t+0)$ and decide the sign correspondingly. (We take for instance $x(t+0)$ and $\Phi(t-0)$ if $\Phi(t-0) < \Phi(t+0)$.) Similarly, as $\bar{\Phi}(t)$ is continuous, we get

$$\int_a^b x(t)\bar{\varphi}(t)\,dt = [x(t)\bar{\Phi}(t)]_{a-0}^{b+0} - \int_a^b \bar{\Phi}(t)\,dx(t). \tag{3.5}$$

Thus we have

$$\int_a^b x(t)\bar{\varphi}(t)\,dt - \int_a^b x(t \mp 0)\,d\Phi(t) = \int_a^b \{\Phi(t \mp 0) - \bar{\Phi}(t)\}\,dx(t), \tag{3.6}$$

the first term on the right-hand side of (3.4) and (3.5) cancelling because $\Phi(t) = \bar{\Phi}(t)$ at both ends.

Now, as stated in the previous section, $\{\Phi(t \mp 0) - \bar{\Phi}(t)\}$ in (3.6) vanishes except perhaps on a set consisting of a denumerable number of open intervals, where it is positive. Therefore, (3.6) is nonnegative and vanishes if and only if $x(t)$ is constant in every such interval.

The difference between the left-hand member of (3.1) and the second term of the left-hand member of (3.6) can come only from the contributions at the points of discontinuity of $\Phi(t)$. Thus

$$\begin{aligned} &\int_a^b x(t \mp 0)\,d\Phi(t) - \int_a^b x(t)\,d\Phi(t) \\ &\qquad = \textstyle\sum_n \{x(t_n \mp 0) - x(t_n)\}\{\Phi(t_n + 0) - \Phi(t_n - 0)\}, \end{aligned} \tag{3.7}$$

summation taken over all points of discontinuity of $\Phi(t)$. Since $x(t_n - 0) \leqq x(t_n) \leqq x(t_n + 0)$, and in view of the above-mentioned choice as to the double sign, it is clear that each term is positive unless $x(t_n) = x(t_n \pm 0)$. Hence, (3.7) is nonnegative and vanishes if and only if (3.3) holds at every t_n.

Adding (3.6) and (3.7), we get an equation expressing the difference between the two members of (3.1) as the sum of the two nonnegative quantities. Hence (3.1) holds. In order that the equality in (3.1) hold, it is necessary and sufficient that both (3.6) and (3.7) vanish. Thus the theorem is proved for bounded $x(t)$.

When $x(t)$ is not bounded in $[a, b]$, we can still derive (3.6) by taking the limit of a sequence of similar formulas for narrower intervals. The rest of the proof does not need any change.

COROLLARY. *Under the same assumptions, the relation*

$$\int_a^b x(t)\,d\Phi(t) \leqq \left\{\int_a^b x(t)^2\,dt\right\}^{\frac{1}{2}} \left\{\int_a^b \bar{\varphi}(t)^2\,dt\right\}^{\frac{1}{2}} \tag{3.8}$$

holds for any nondecreasing function $x(t)$ such that the integrals including $x(t)$ exist and are finite. If $\bar{\varphi}(t)^2$ is summable in (a, b) and if $\bar{\varphi}(t)$ is not identically equal to zero in (a, b), then necessary and sufficient conditions for the equality in (3.8) *to hold are that $x(t) = A\bar{\varphi}(t)$ almost everywhere (with nonnegative constant coefficient A) and that $x(t)$ satisfy* (3.3) *at every point of discontinuity, if any, of $\Phi(t)$.*

PROOF. The ordinary Schwarz inequality applied to the right-hand member of (3.1) leads us to the inequality (3.8).

We note that (3.8) may be called a modified Schwarz inequality.

As a particular case of (3.1), the inequality

$$\int_a^b x(t)\varphi(t)\,dt \geqq \frac{1}{b-a}\int_a^b x(t)\,dt \int_a^b \varphi(t)\,dt \tag{3.9}$$

can be derived for any nondecreasing functions $x(t)$ and $\varphi(t)$, both summable in (a, b). This is Tchebycheff's inequality (See [3] p. 168, Theorem 236). Incidentally, a change of variables leads us from (3.9) to a formula which implies the one on p. 601 of [11].

THEOREM 2. *Let $\Phi(t)$ be a function of bounded variation in $[a, b]$ and continuous at both ends. Let moreover its value be the same at both ends. Then for any nondecreasing function $x(t)$ belonging to $L_2(a, b)$ and summable with respect to Φ,*

$$\int_a^b x(t)\,d\Phi(t) \leqq \left[\int_a^b \{x(t) - C\}^2\,dt\right]^{\frac{1}{2}}\left\{\int_a^b \bar{\varphi}(t)^2\,dt\right\}^{\frac{1}{2}}, \tag{3.10}$$

where C is any constant and $\bar{\varphi}(t)$ is the right-hand derivative of the greatest convex minorant $\bar{\Phi}(t)$ of $\Phi(t)$. If $\bar{\varphi}(t)$ belongs to $L_2(a, b)$ and is not identically equal to zero in (a, b), then the equality in (3.10) *holds if and only if $x(t) - C = A\bar{\varphi}(t)$ almost everywhere (with nonnegative coefficient A) and* (3.3) *hold at every point of discontinuity, if any, of $\Phi(t)$.*

PROOF. The integral on the left-hand side of (3.10) does not change its value if one replaces $x(t)$ by $x(t) - C$. Therefore we get (3.10) from (3.8).

We note in particular, that if C is chosen as the mean value of $x(t)$ in (a, b), the first factor on the right-hand side of (3.10) reduces to the standard deviation.

4. Applications to distributions. Let us consider the inverse function (Cf. [3] pp. 152–3, [5] p. 189.) $x(F)$ of the cumulative distribution function $F(x)$. Then for instance $x(\alpha)$ may be called the α-quantile. $x(F)$ is a nondecreasing function of F in $(0, 1)$. We assume that it belongs to $L_2(0, 1)$.

EXAMPLE 1. The distance between the α-quantile and the mean EX is given by

$$x(\alpha) - EX = \int_0^1 x(F)\,d\Phi(F), \tag{4.1}$$

where

$$\begin{aligned} \Phi(F) &= -F && \text{when } 0 \leqq F < \alpha, \\ &= 1 - F && \text{when } \alpha \leqq F < 1. \end{aligned} \tag{4.2}$$

Theorem 2 is applicable here, leading us to

$$-\sqrt{\frac{1-\alpha}{\alpha}} \leqq \frac{x(\alpha) - EX}{\sigma} \leqq \sqrt{\frac{\alpha}{1-\alpha}}. \tag{4.3}$$

We note that (4.3) has been derived and used by Hoeffding [6]. It is equivalent to the inequality (cf. [7])

$$\Pr\{X \leqq EX + b\sigma\} \geqq \frac{b^2}{b^2 + 1} \qquad (b > 0). \tag{4.4}$$

In the special case where $\alpha = \frac{1}{2}$, (4.3) tells us that the distance between the median and the mean can never exceed the standard deviation (cf. [8]).

Example 2. The distance between the expected value of the i-th smallest member in the sample and the population mean is given by

$$EX_i - EX = \int_0^1 x(F)\varphi(F)\, dF, \tag{4.5}$$

where

$$\varphi(F) = \frac{n!}{(i-1)!\,(n-i)!} F^{i-1}(1-F)^{n-i} - 1. \tag{4.6}$$

Here again Theorem 2 is applicable. If i equals neither 1 nor n, then the function $\bar{\varphi}(F)$ turns out to be

$$\begin{aligned} \bar{\varphi}(F) &= \varphi(F) && \text{when } 0 \leqq F < F_2, \\ &= \varphi(F_2) && \text{when } F_2 \leqq F < 1, \end{aligned} \tag{4.7}$$

where F_2 is determined by

$$(1 - F_2)\varphi(F_2) = \int_{F_2}^1 \varphi(F)\, dF, \qquad 0 < F_2 < F_1 = \frac{i-1}{n-1}. \tag{4.8}$$

The integral on the right-hand side can be evaluated by means of incomplete Beta function. If n is not very large, the tables of the incomplete Beta function [9] will be satisfactory. Then we can compute the least upper bound of (4.5) for given population standard deviation. The bound is actually attained for a particular distribution $x(F) = A\bar{\varphi}(F)$. This is a mixed distribution with concentrated probability on the value $x = A\varphi(F_2)$ and distributed probability on the interval $(-A, A\varphi(F_2))$.

For $i = n$ (i.e. for the largest member), $\varphi(F)$ is monotone increasing. Consequently, the ordinary Schwarz inequality will suffice to get the least upper bound $(n-1)/\sqrt{2n-1}\cdot\sigma$. (This is naturally a larger bound than the formula (3.6) of the previous paper [1] which is applicable only to symmetrical populations.) The bound is actually attained for a particular distribution $x(F) = A\varphi(F)$.

A few numerical results for the sample median are shown in Table 1. Also shown are the results given by ordinary Schwarz inequality. This illustrates the improvement obtained by the present method.

Example 3. For $x(1-\beta) - x(\alpha)$ we get

$$
(4.9) \qquad \bar{\varphi}(F) = \begin{cases} -\dfrac{1}{\alpha}, & 0 \leqq F < \alpha, \\ = 0, & \alpha \leqq F < 1 - \beta, \\ = \dfrac{1}{\beta}, & 1 - \beta \leqq F < 1, \end{cases}
$$

and finally

$$
(4.10) \qquad x(1 - \beta) - x(\alpha) \leqq \sqrt{\frac{1}{\alpha} + \frac{1}{\beta}}\,\sigma.
$$

We note that by putting $\alpha = \beta$ in (4.10), we obtain $x(1 - \alpha) - x(\alpha) \leqq \sigma\sqrt{2/\alpha}$. This is, for symmetrical distribution, equivalent to the Tchebycheff-Bienaymé inequality. The latter, in the general case, can be derived by applying the above formula after symmetrizing the given distribution.

TABLE 1

Upper bounds for the difference between the expected value of the sample median and the population mean measured by the population standard deviation

$\{EX_{(n+1)/2} - EX\}/\sigma$

Sample size n	Least Upper Bound	Upper Bound Given by Ordinary Schwarz Inequality
3	.27099	.44721
5	.37659	.65465
7	.43918	.79480
9	.48291	.90226
11	.51604	.99019
13	.54221	1.0650
15	.56374	1.1305
17	.58210	1.1888
19	.59776	1.2416

EXAMPLE 4. For the expected value of the difference of two order statistics, (3.10) leads us to

$$
(4.11) \qquad E(X_i - X_j) \leqq \sigma \left\{ \int_0^1 \bar{\varphi}(F)^2 \, dF \right\}^{\frac{1}{2}},
$$

where $\bar{\varphi}(F)$ is to be obtained from

$$
(4.12) \qquad \varphi(F) = \frac{n!}{(i-1)!\,(n-i)!} F^{i-1}(1 - F)^{n-i} - \frac{n!}{(j-1)!\,(n-j)!} F^{j-1}(1 - F)^{n-j}.
$$

Thus, if $i = n, j = n - 1$, then

$$\bar{\varphi}(F) = \varphi(F_2) \qquad 0 \leqq F < F_2, \tag{4.13}$$
$$= \varphi(F) \qquad F_2 \leqq F < 1,$$

where $F_2 = (n - 2)/(n - 1)$.

If $j = n - i + 1$, then

$$\bar{\varphi}(F) = \begin{cases} -\varphi(F_2) & 0 \leqq F < 1 - F_2, \\ \varphi(F) & 1 - F_2 \leqq F < F_2, \\ \varphi(F_2) & F_2 \leqq F < 1, \end{cases} \tag{4.14}$$

where F_2 is determined by

$$(1 - F_2)\varphi(F_2) = \int_{F_2}^{1} \varphi(F)\, dF. \qquad \tfrac{1}{2} < F_2 < 1. \tag{4.15}$$

The case where $i = n, j = 1$ (i.e. the case of the "sample range") is an exception, but it has essentially been discussed already (cf. [10] and [1]; $E(X_n - X_1)/\sigma$ is equal to twice the corresponding value in Table 1 of [1]).

5. Acknowledgements. The author wishes to express his appreciation to Professor Harold Hotelling for the suggestions made during the study. Thanks are also due to Professor Kunihiko Kodaira for suggestive discussions and to Professors Wassily Hoeffding and Herbert Robbins for helpful information.

REFERENCES

[1] SIGEITI MORIGUTI, "Extremal properties of extreme value distributions," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 523–536.

[2] E. F. BECKENBACH, "Convex functions," *Bull. Amer. Math. Soc.*, Vol. 54 (1948), pp. 439–460.

[3] G. H. HARDY, J. E. LITTLEWOOD AND G. PÓLYA, *Inequalities*, Cambridge University Press, 1934.

[4] S. SAKS, *Theory of the Integral*, Hafner Publishing Co., New York, 1937, p. 102.

[5] H. SCHEFFÉ AND J. W. TUKEY, "Non-parametric estimation. I. Validation of order statistics," *Ann. Math. Stat.*, Vol. 16 (1945), pp. 187–192.

[6] W. HOEFFDING, "The large-sample power of tests based on permutations of observations," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 169–192.

[7] M. FRECHET, *Généralités sur les Probabilités. Éléments aléatoires*, Gauthier-Villars, Paris, 1950, p. 137; J. V. USPENSKY, *Introduction to Mathematical Probability*, McGraw-Hill Book Co., 1937, p. 198.

[8] HAROLD HOTELLING AND LEONARD M. SOLOMONS, "The limits of a measure of skewness," *Ann. Math. Stat.*, Vol. 3 (1932), pp. 141–142.

[9] KARL PEARSON, *Tables of the Incomplete Beta function*, Cambridge University Press, 1934.

[10] R. L. PLACKETT, "Limits of the ratio of mean range to standard deviation," *Biometrika*, Vol. 34 (1947), pp. 120–122.

[11] A. W. KIMBALL, "On dependent tests of significance in the analysis of variance," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 600–602.