Associative functions and statistical triangle inequalities

By B. SCHWEIZER (Tucson) and A. SKLAR (Chicago)

Introduction. In the course of our work on statistical metric spaces [16, 17], we have been led to consider a class of real-valued 2-place functions T, whose domain is the closed unit square $[0,1]\times[0,1]$ and which satisfy the following conditions:

(0.1) $T(0,0)=0, \quad T(a,1)=T(1,a)=a.$ (Boundary conditions)

(0.2) $T(a,b)\leqq T(c,d)$, whenever $a\leqq c, b\leqq d.$ (Monotonicity)

(0.3) $T(a,b)=T(b,a).$ (Symmetry)

(0.4) $T(T(a,b),c)=T(a,T(b,c)).$ (Associativity)

These functions arise naturally in the study of generalized triangle inequalities for statistical metric spaces; and, following K. MENGER [14], a function which satisfies the conditions (0.1)—0.4) is called a *triangular norm* (briefly, a *t-norm*).

Knowing whether a given generalized triangle inequality holds or does not hold in some given statistical metric space can often be crucial. For this reason, among others, it is important to know as much as possible about t-norms and, in particular, to have a large repertoire of them at hand. What is required, therefore, is a characterization of t-norms — a characterization which will reveal their mutual relationships and enable us to construct them at will.

In attacking this problem of characterization, it has turned out that the most useful — and intrinsically the most interesting — property of t-norms is their associativity (Condition (0.4)).[1] Now, associativity has been studied in extenso from the algebraic point of view; and even the, by comparison,

[1] This condition states that the t-norm T defines a semigroup on the closed unit interval [0, 1]. The other conditions further imply that this semigroup has a unit, 1, and an annihilator, 0 (Condition (0.1)); and that the semigroup operation is order-preserving (Condition (0.2)) and commutative (Condition (0.3)) [10]. Our central problem may thus be restated as that of characterizing and constructing all semigroups on [0, 1] which have these properties.

neglected function-theoretic aspect of associativity, which departs from the functional equation (0.4), can boast a distinguished roster of investigators [1, 3, 7, 9, 13], headed by ABEL. As a result of their researches, there exists today a means of characterizing in a simple manner, not, it is true, all t-norms, but a large and important class of them. This is the class of *strict t-norms*, which in addition to (0.1) and (0.4) satisfy the following conditions:[2])

(0.5) T is continuous (on $[0,1]\times[0,1]$).

(0.6) $T(a,b)<T(c,b)$, whenever $0<a<c\leqq 1$,
$T(a,b)<T(a,d)$, whenever $0<b<d\leqq 1$. (Strict monotonicity)

Accordingly, in this paper we shall confine ourselves to a study of strict t- norms (Part I) and their applications to statistical metric spaces (Part II).

I. Associative functions

1. Preliminary theorems. The topics discussed in this paper take as their starting point the following known theorems:

Theorem 1. *Let I be an open or half-open (but not closed) interval of the real line and T a 2-place function from $I\times I$ to I. Suppose that T is continuous and strictly increasing in each of its places, i. e., that*

$$T(a,b)<T(c,b), \quad T(a,b)<T(a,d)$$

for all a, b, c, d in I such that $a<c, b<d$. Suppose further that T is associative, i. e., satisfies the functional equation,

$$T(T(a,b),c)=T(a,T(b,c)), \tag{1.1}$$

for all a, b, c in I. Then there exists a 1-place function f, defined, continuous, and strictly monotone on I, such that for all a, b in I, $T(a,b)$ has the representation,

$$T(a,b)=f^*(f(a)+f(b)), \tag{1.2}$$

where f^* is the inverse function of f.

Corollary. *If T satisfies the hypotheses of Theorem 1, then T is symmetric, i. e., $T(a,b)=T(b,a)$ for all a, b in I. Thus every continuous, strictly increasing, associative function is symmetric.*

---

[2]) It is clear that (0.6) implies (0.2). The fact that (0.4), (0.5) and (0.6), taken together, imply (0.3) is a direct consequence of an important theorem of J. ACZÉL [3], which is quoted in Section 1 of this paper.

Theorem 2. *If, for a given T, f and g are both strictly monotone solutions of* (1.2), *then there exists a number λ such that $g = \lambda f$; conversely, if for a given T, f is a (strictly monotone) solution of* (1.2) *and $g = \lambda f$, for some number λ, then g is a solution of* (1.1) *for this same T.*

Theorem 3. *(Converse of Theorem 1.) Let f be a continuous, strictly monotonic 1-place function from the (open or half-open) interval I (= Dom f, to the interval Ran f. Let f^* be the inverse of f. Suppose further that Ran f is closed under addition, i. e., that if x and y are both in Ran f then so is $x+y$. Then the 2-place function T which is given by* (1.2) *is defined, continuous, strictly increasing in each place, and associative on $I \times I$.*

The solution of (1.1), the functional equation of associativity, in the form (1.2) was first obtained — under the additional assumptions of commutativity and differentiability — by ABEL in 1826. It is, in fact, the subject of the first paper published by him in Crelle's journal [1]. Further work along these lines has since been done by L. E. J. BROUWER [7]. É. CARTAN [9], J. ACZÉL [3], M. HOSSZÚ [13] and T. S. MOTZKIN [15]. The form in which we have stated Theorem 1 uses the weakest hypotheses thus far known to be sufficient to guarantee the existence of the representation (1.2) and is due to J. ACZÉL. We refer the reader to his interesting and elegant paper [3] for the proof. Similarly, the statement of Theorem 2 may be found in another paper by ACZÉL [4, p. 353], and its proof in a third [2] (Cf. also, R. CACCIOPPOLI [8]).[3])

Theorem 3 is very much simpler than Theorem 1 — its proof being a mere matter of calculation. On the other hand, whereas the proof of Theorem 1 makes essential use of many of the properties of the real number system, Theorem 3 can readily be modified so as to apply to general abstract semigroups. Our study of non-strict t-norms (the results of which will be presented in detail in a subsequent paper) has led — indeed, forced — us to consider this modification. Here, our starting point is a group of theorems due to AL. C. CLIMESCU [11] on the transformation of semigroups into semigroups. CLIMESCU's results can be extended in several ways. And these extensions, which are of interest in their own right, have, when specialized back to the case of associative functions on the reals, the effect of giving us the conclusion of Theorem 3 under a considerably weakened set

[3]) *Note added in proof:* These theorems, their proofs and many other questions connected with the functional equation of associativity are discussed in detail by J. ACZÉL in his recently published book, *Vorlesungen über Funktionalgleichungen und ihre Anwendungen,* Basel und Stuttgart, 1961.

of hypotheses. They thereby apply to, and yield, a large number of non-strict t-norms.

2. The characterization of strict t-norms. Strict t-norms, as defined in the introduction, satisfy all the hypotheses of Theorem 1, whence we immediately have a representation of these t-norms in the form (1.2). In this case, f is a continuous, strictly monotone function on the half-open interval $(0, 1]$. Moreover, the presence of the boundary conditions (0.1) allows us to determine the behavior of any such f at the endpoints of this interval as follows: The boundary condition $T(a, 1) = a$ yields $f(T(a, 1)) = f(a)$. But

$$f(T(a, 1)) = f[f^*(f(a) + f(1))] = f(a) + f(1),$$

from which it follows that $f(1) = 0$. Next, in order to determine the behavior of f near 0, suppose that $\lim_{a \to 0+} f(a) = A$. Then, upon imposing the condition $\lim_{a \to 0+} T(a, a) = 0$, we obtain $\lim_{a \to 0+} f(T(a, a)) = A$. But, as above,

$$f(T(a, a)) = f[f^*(f(a) + f(a))] = 2f(a).$$

Consequently, $\lim_{a \to 0+} f(T(a, a)) = 2 \lim_{a \to 0+} f(a) = 2A$, whence $A = 2A$. Now A cannot be zero, since f cannot assume the same value twice. Hence A is not finite. The choice of A as $+\infty$ or $-\infty$ is at our disposal and, as a matter of convenience, we shall consistently choose $A = +\infty$. Consequently, the function f appearing in (1.2) decreases steadily from $+\infty$ to 0 as its argument increases from 0 to 1. Correspondingly, f^*, the inverse of f, decreases steadily from 1 to 0 as its argument increases from 0 to $+\infty$. Summarizing, we have

Theorem 4. *If T is a strict t-norm, i.e., a 2-place function satisfying the conditions* (0.1), (0.4), (0.5) *and* (0.6), *then there exists a 1-place function f, defined, continuous and strictly decreasing on the half-open interval* $(0, 1]$, *with* $\lim_{a \to 0+} f(a) = +\infty$, $f(1) = 0$, *and such that for any (a, b) in* $(0, 1] \times (0, 1]$,

$$T(a, b) = f^*(f(a) + f(b)), \tag{2.1}$$

where f^* is the inverse of f.

Given T, we shall call any function f that satisfies all the conditions stated in Theorem 4 an *additive generator* of T. It follows that if f and g are both additive generators of one and the same strict t-norm, then $g = \lambda f$, where λ is a positive constant. Conversely, if f is a function which is defined, continuous and strictly decreasing from $+\infty$ to 0 on the interval $(0, 1]$, then f is an additive generator of a strict t-norm; and if λ is a positive constant then λf is an additive generator of the same t-norm.

Strictly speaking, an additive generator determines its corresponding strict t-norm only on $(0, 1] \times (0, 1]$, not on $[0, 1] \times [0, 1]$. However, this is a matter of little consequence: first of all, the boundary conditions (0. 1) imply that any t-norm assumes the value zero whenever one or the other of its argument is zero; and secondly, any strict t-norm may be extended — either directly by continuity or via the representation (2.1) — so as to assume the correct values on those parts of the boundary of the unit square to which the representation (2. 1) does not apply directly. *Thus a strict t-norm is completely characterized by any one of its additive generators.*

The following theorem is the multiplicative equivalent of Theorem 4:

Theorem 5. *If T is a strict t-norm, then there exists a 1-place function h, defined, continuous and strictly increasing on the closed interval* $[0, 1]$, *with* $h(0) = 0$, $h(1) = 1$, *and such that for any* (a, b) *in* $[0, 1] \times [0, 1]$,

$$T(a, b) = h^*(h(a) \cdot h(b)), \tag{2. 2}$$

where h^* is the inverse of h.

PROOF. In view of Theorem 4, an additive generator, f of T exists. Let h be the function defined by,

$$h = \exp(-f) = e^{-f}. \tag{2. 3}$$

Then we have,

$$f = -\log h, \quad f^* = h^*(e^{-j}), \quad h^* = f^*(-\log), \tag{2. 4}$$

where h^* is the inverse of h, and j is the identity function defined by: $j(x) = x$, for any real number x. It follows that h and h^* are both defined, continuous and strictly increasing on the half-open interval $(0, 1]$. Furthermore, h and h^* may be extended by continuity to the closed interval $[0, 1]$. Both h and h^* are strictly increasing on this interval and we find that $h(0) = h^*(0) = 0$, $h(1) = h^*(1) = 1$, so that each of these functions maps the closed unit interval onto itself. Lastly, in terms of h and h^*, (2. 1) takes the form,

$$\begin{aligned} T(a, b) &= h^*[\exp(-(-\log h(a) - \log h(b)))] \\ &= h^*[\exp(\log h(a) + \log h(b))] \\ &= h^*(h(a) \cdot h(b)), \end{aligned}$$

which is our desired result.

The function f completely determines the function h, and conversely. Therefore a strict t-norm is as completely determined by the latter as by the former. Accordingly, we call h a *multiplicative generator* of its corresponding t-norm. Any two multiplicative generators of the same strict t-norm are positive powers of each other.

The 2-place function Prod, defined by

$$\text{Prod}(a, b) = a \cdot b, \tag{2.5}$$

is a strict t-norm. The function $-\log$ is an additive generator of Prod; the corresponding multiplicative generator is j_1, the restriction of the identity function j to the interval $[0, 1]$. The other additive (respectively, multiplicative) generators of Prod are of the form $-\lambda \log$ (respectively, j_1^{λ}), where $\lambda > 0$. Now, in terms of Prod, (2.2) may be cast into the form

$$T(a, b) = h^*(\text{Prod}(h(a), h(b))), \tag{2.6}$$

and accordingly, Theorem 5 may be rephrased as follows:

Theorem 6. *A 2-place function T on the closed unit square is a strict t-norm if and only if it is derivable from the particular t-norm Prod via* (2.6) *through the intermediary of a multiplicative generator h.*

3. Examples. With the equivalent Theorems 4 and 5 we have achieved the first aim of this paper: the characterization of strict t-norms. Their converses, similarly, enable us to achieve another aim: the construction of t-norms. We can, in fact, construct entire families of t-norms at will. A number of examples follow.

(*a*) Let $h_p = [1-(1-j_1)^p]^{1/p}$, for any $p > 0$, i. e., $h_p(a) = [1-(1-a)^p]^{1/p}$, for $0 \leqq a \leqq 1$, $p > 0$. Then $h_p^* = 1-(1-j_1^p)^{1/p}$, and

$$T_p(a, b) = 1-[(1-a)^p + (1-b)^p - (1-a)^p(1-b)^p]^{1/p}. \tag{3.1}$$

For $p = 1$, we find that $T_1 = \text{Prod}$. The limiting cases: $p \to 0+$, $p \to \infty$ are also of interest. A short calculation yields

$$\lim_{p \to 0+} T_p(a, b) = T_w(a, b),$$

$$\lim_{p \to \infty} T_p(a, b) = \text{Min}(a, b),$$

where T_w is the function on the unit square given by

$$T_w(a, b) = \begin{cases} a, & b = 1, \\ b, & a = 1, \\ 0, & \text{otherwise}; \end{cases} \tag{3.2}$$

and Min is given by

$$\text{Min}(a, b) = \begin{cases} a, & 0 \leqq a \leqq b \leqq 1, \\ b, & 0 \leqq b \leqq a \leqq 1. \end{cases} \tag{3.3}$$

Direct reference to the definition shows that T_w and Min are both t-norms,[4]) though neither is a strict t-norm. It is also easy to show (Cf.

[4]) This fact is also a direct consequence of a theorem due to E. Thorp (Theorem 4 of [20]).

[17], p. 318) that T_w is the minimal, and Min the maximal t-norm.

(*b*) Let $f_p = |\log|^p$, for any $p>0$. Then $f_p^* = \exp(-j^{1/p})$ and

$$T_p(a, b) = \exp[-(|\log a|^p + |\log b|^p)^{1/p}]. \tag{3.4}$$

Again we find that $T_1 = \text{Prod}$, and $\lim_{p\to 0+} T_p = T_w$, $\lim_{p\to\infty} T_p = \text{Min}$.

(*c*) Let $f_p = (1-\log)^p - 1$, for any $p>0$. Then $f_p^* = \exp[1-(1+j)^{1/p}]$ and

$$T_p(a, b) = \exp\{1-[(1-\log a)^p + (1-\log b)^p]^{1/p}\}. \tag{3.5}$$

Again we find that $T_1 = \text{Prod}$, and $\lim_{p\to\infty} T_p = \text{Min}$. However, in this case, for the limit $p\to 0+$, we obtain

$$\lim_{p\to 0+} T_p(a, b) = a\cdot b\cdot\exp(-\log a\cdot\log b)$$

(*d*) Let $f_p = j_1^{-p} - 1$, for any $p>0$. Then $f_p^* = (1+j)^{-1/p}$ and

$$T_p(a, b) = (a^{-p} + b^{-p} - 1)^{-1/p} = \frac{a\cdot b}{(a^p + b^p - a^p\cdot b^p)^{1/p}}. \tag{3.6}$$

In this case we find that

$$T_1(a, b) = \frac{a\cdot b}{a+b-a\cdot b},$$

and that $\lim_{p\to 0+} T_p = \text{Prod}$, $\lim_{p\to\infty} T_p = \text{Min}$.

(*e*) Let $f_p = p/j_1 - p + 1 - j_1$, for any $p>0$. Then $f_p^* = \frac{1}{2}\{1-p-j + [(1-p-j)^2 + 4p]^{1/2}\}$ and

$$\begin{aligned} T_p(a, b) = \frac{1}{2}\Big(a+b-1+p\Big(\frac{1}{a}+\frac{1}{b}+1\Big) + \\ + \Big\{\Big[a+b-1+p\Big(\frac{1}{a}+\frac{1}{b}+1\Big)\Big]^2 + 4p\Big\}^{1/2}\Big). \end{aligned} \tag{3.7}$$

The interest in this example lies in the limiting case $p\to 0+$. It is readily seen that

$$\begin{aligned} \lim_{p\to 0+} T_p(a, b) &= \frac{1}{2}(a+b-1) + \frac{1}{2}([a+b-1]^2)^{1/2} \\ &= \frac{1}{2}(a+b-1+|a+b-1|) \\ &= T_m(a, b), \end{aligned}$$

where T_m is the function on the unit square given by

$$T_m(a, b) = \max(a + b - 1, 0). \tag{3.8}$$

Again T_m is a t-norm, though not a strict t-norm.

4. Geometry of t-norms. We end the first part of this paper with a few remarks of a geometrical nature.

In studying t-norms, it is desirable to have a "picture" at hand. Now, a t-norm determines a surface over the unit square in the xy-plane. From Condition (0.1) it follows that this surface contains (and, when continuous, is bounded by) a skew quadrilateral (indicated by the heavy lines in Figure 1) whose vertices, in order, are the four points: (0, 0, 0), (1, 0, 0), (1, 1, 1)

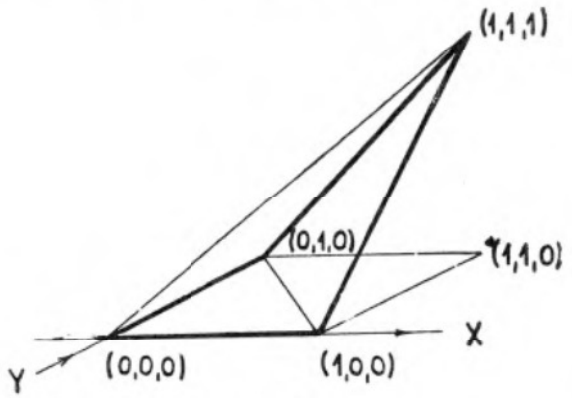


Fig. 1

and (0, 1, 0). Condition (0.2) implies that the height of this surface above a point in the unit square increases, or at least does not decrease, as the x or y coordinates of this point increase. Thus the surface lies entirely in the unit cube. Condition (0.3) implies that this surface is symmetric about the plane $x = y$.

In Figure 1 we have sketched the graphs of the t-norms Min and T_m (see (3.3) and (3.8)), each of which consists of two triangular pieces. In addition, the graph of Prod consists of the portion of the hyperbolic paraboloid $z = xy$ which is bounded by the previously mentioned skew quadrilateral; and the graph of T_w (see (3·2)) consists of the region $[0, 1) \times [0, 1)$ in the xy-plane together with the two sides of this quadrilateral which are not in the xy-plane.

There remains the associativity condition (0.4). This seems to have no simple geometric interpretation as far as the "associative surface" determined by a t-norm is concerned. About the most that seems to be known at present is that, if T is a strict t-norm, then the three systems of curves, $x = \text{const.}$,

$y = \text{const.}$, and $z = T(x, y) = \text{const.}$, form — at least locally — a hexagonal web (Sechseckgewebe) [5, 6, 19] — which one may consider as lying either on the associative surface or in the xy-plane. Nevertheless, this fact raises several intriguing questions. In his classic paper [19], G. THOMSEN connected the property that certain triples of systems of curves on a surface form a Sechseckgewebe with the property that this surface is *isothermal.*[5]) Now, the surface $z = xy$, determined by the t-norm Prod, is isothermal; and every associative surface has at least one triple of systems of curves on it which form a Sechseckgewebe. Thus one may ask: Is every (sufficiently smooth) associative surface isothermal? More generally: What geometric properties of Prod (or, for that matter, any 2-place function) are preserved by the transformation (2.6)?

II. Applications to statistical triangle inequalities

5. Statistical metric spaces. Before applying the results of the previous sections to statistical metric spaces, we briefly review their definition and several of their basic properties.[6])

Definition 1. A *statistical metric space* (briefly, an *SM space*) is an ordered pair $(\mathbb{S}, \mathfrak{J})$, where $\mathbb{S}$ is a non-empty set (whose elements are the *points* of the *SM* space) and $\mathfrak{J}$ is a mapping from $\mathbb{S} \times \mathbb{S}$ into the set of distribution functions — i. e., $\mathfrak{J}$ associates a distribution function F_{pq} with every pair of points p, q in $\mathbb{S}$. The functions F_{pq} are required to satisfy the following conditions:

(5. 1)
I. $F_{pq}(x) = 1$ for all $x > 0$ if and only if $p = q$.
II. $F_{pq}(0) = 0$.
III. $F_{pq} = F_{qp}$.
IV. If $F_{pq}(x) = 1$ and $F_{qr}(y) = 1$, then $F_{pr}(x + y) = 1$.

The number $F_{pq}(x)$ is interpreted as the probability that the distance from p to q is less than x; and, as one readily sees, the conditions I—IV are straightforward generalizations of the corresponding properties of ordinary metrics; in particular, IV is a generalization, albeit a very weak one, of the ordinary triangle inequality.

---

[5]) A surface is said to be isothermal if its lines of curvature form an isothermally orthogonal net, i. e., if there exists a parametrization, say in terms of u, v, such that the lines of curvature are the curves $u = \text{const.}$, $v = \text{const.}$, and such that the first fundamental form, i. e. the square of the element of length, has the from, $ds^2 = \lambda(u, v)(du^2 + dv^2)$.

[6]) For a detailed discussion, see [17].

In this paper we shall be concerned with a different generalized triangle inequality — namely the one due to K. MENGER [14] which stipulates that

$$F_{pr}(x+y) \geqq T(F_{pq}(x), F_{qr}(y)), \tag{5.2}$$

where p, q, r are points in $\mathfrak{S}$, x, y are non-negative real numbers, and T is a t-norm. Since $T(1, 1) = 1$, the inequality (5.2) contains IV as a special case. This, as in [16, 17], leads us to the following:

Definition 2. A statistical metric space $(\mathfrak{S}, \mathcal{J})$ is a *Menger space* if the inequality (5.2) is valid for all points, p, q, r in $\mathfrak{S}$, all $x, y \geqq 0$, and some t-norm T; and we say that $(\mathfrak{S}, \mathcal{J})$ is a Menger space *under* T.

Definition 3. Let W_1 and W_2 be two 2-place functions with a common domain D. Then W_1 is said to be *weaker* than W_2 (and W_2 *stronger* than W_1) if $W_1(a, b) \leqq W_2(a, b)$ for all (a, b) in D and $W_1(a, b) < W_2(a, b)$ for at least one pair (a, b) in D.

It follows that if an SM space is a Menger space under some t-norm T, then it is a Menger space under any t-norm weaker than T. As mentioned in Section 3, T_w is the weakest and Min the strongest possible t-norm. The t-norms T_m and Prod are of intermediate strength, with T_m being weaker than Prod.

An immediate application of the results of the preceding section yields the following:

Theorem 7. *Let T_1 and T_2 be strict t-norms, f_1 an additive generator of T_1, and f_2 an additive generator of T_2. Then T_1 is weaker than T_2 if and only if the composite $f_1 f_2^*$ is a non-linear subadditive function.*[7])

PROOF. Let $s, t \geqq 0$ be given and set $a = f_2^*(s)$, $b = f_2^*(t)$. Then (a, b) lies in the unit square. By hypothesis, $T_1(a, b) \leqq T_2(a, b)$, which by Theorem 4 is equivalent to

$$f_1^*(f_1(a) + f_1(b)) \leqq f_2^*(f_2(a) + f_2(b))$$

or

$$f_1^*[f_1(f_2^*(s)) + f_1(f_2^*(t))] \leqq f_2^*(s+t).$$

Now f_1 is decreasing, so that on applying f_1 to both sides of the above inequality, we obtain

$$f_1[f_2^*(s+t)] \leqq f_1(f_2^*(s)) + f_1(f_2^*(t)),$$

i. e.,

$$f_1 f_2^*(s+t) \leqq f_1 f_2^*(s) + f_1 f_2^*(t).$$

Thus $f_1 f_2^*$ is subadditive. If $f_1 f_2^*$ were linear, then we would have $f_1 f_2^*(x) =$

[7]) This theorem is thus, in a sense, an extension of Theorem 2.

$= ux + v$, for some u, v and all $x \geqq 0$. Setting $x = 0$ yields $v = f_1 f_2^*(0) = -f_1(f_2^*(0)) = -f_1(1) = 0$ and leads to $f_1 = uf_2$. But this, by Theorem 4 ff., implies that T_1 and T_2 are identical and is contrary to hypothesis. This proves the "only if" half of the theorem; the "if" half follows on reserving the steps of the proof.

If an SM space is a Menger space under a strict t-norm, then by applying Theorem 4, the inequality (5.2) can be cast into an interesting form. For, using the representation (2.1) in (5.2), we obtain

$$F_{pr}(x+y) \geqq f^*[f(F_{pq}(x)) + f(F_{qr}(y))]$$

whence, since f is a decreasing function, we have

$$f[F_{pr}(x+y)] \leqq f[F_{pq}(x)] + f[F_{qr}(y)] \tag{5.3}$$

an inequality that bears a striking resemblance to the triangle inequality in an ordinary metric space.

6. Triangular conorms. In studying SM spaces, it is sometimes more convenient to work with the functions $G_{pq} = 1 - F_{pq}$ rather than with the functions F_{pq} themselves. Since $G_{pq}(x)$ is the probability that the distance between p and q is greater than or equal to x, this amounts to working with the tails of the distance distributions rather than with their central portions. Simultaneously, it is convenient to replace any t-norm T that occurs by a function S, which will be called a *triangular conorm* (briefly, a *t-conorm*), and is defined in terms of T by:

$$S(a, b) = 1 - T(1-a, 1-b). \tag{6.1}$$

It is readily verified that S has the same domain and range as T; and that the boundary conditions (0.1) for T transform into

$$S(1, 1) = 1, \quad S(a, 0) = S(0, a) = a, \tag{6.2}$$

whereas the conditions (0.2), (0.3) and (0.4) are satisfied by S as well as by T.

On replacing a and b in (6.1) by $1-a$ and $1-b$, respectively, we obtain

$$T(a, b) = 1 - S(1-a, 1-b). \tag{6.3}$$

Thus there is a one-one correspondence between t-norms and t-conorms — a t-norm determines a unique t-conorm, and conversely. Geometrically, the graphs of a t-norm and its corresponding t-conorm are reflections of each other in the point $(1/2, 1/2, 1/2)$, the center of the unit cube. It follows that if T_1 and T_2 are t-norms whose t-conorms are S_1 and S_2, respectively, and if T_1 is weaker than T_2, then S_1 is stronger than S_2. Consequently, the weakest possible t-conorm corresponds to the strongest possible t-norm,

namely Min. Now, the t-conorm of Min is the function Max, defined by

$$\text{Max}(a, b) = \begin{cases} b, & 0 \leqq a \leqq b \leqq 1, \\ a, & 0 \leqq b \leqq a \leqq 1; \end{cases} \tag{6.4}$$

and since Max is stronger than Min, it follows that any t-conorm is stronger than every t-norm.

A strict t-norm determines a *strict* t-conorm. In particular, the t-norm Prod determines the t-conorm Sum — Prod, defined by

$$[\text{Sum} - \text{Prod}](a, b) = a + b - a \cdot b. \tag{6.5}$$

If T is a strict t-norm, then a substitution of the representation (2.1) into (6.1) yields

$$S(a, b) = 1 - f^*(f(1-a) + f(1-b)).$$

Thus, upon defining the function g by

$$g = f(1 - j_1), \tag{6.6}$$

so that

$$f = g(1 - j_1), \quad g^* = 1 - f^*, \quad f^* = 1 - g^*, \tag{6.7}$$

we have the familiar representation,

$$S(a, b) = g^*(g(a) + g(b)). \tag{6.8}$$

From (6.6), (6.7) and the known properties of the additive generator f, it follows that

$$g(0) = g^*(0) = 0, \quad \lim_{x \to 1} g(x) = \infty, \quad \lim_{x \to \infty} g^*(x) = 1, \tag{6.9}$$

and that both g and g^* are strictly increasing on their respective domains.[8]) Following our previous terminology, we call the function g an *additive generator* of the t-conorm S. Next, let

$$k = \exp(-g), \tag{6.10}$$

so that

$$g = -\log k, \quad g^* = k^*(e^{-j}), \quad k^* = g^*(-\log). \tag{6.11}$$

Then, in terms of k and k^*, (6.8) takes the form

$$\begin{aligned} S(a, b) &= k^*[k(a) + k(b) - k(a) \cdot k(b)] = \\ &= k^*([\text{Sum} - \text{Prod}](k(a), k(b))). \end{aligned} \tag{6.12}$$

Equation (6.12) bears the same relationship to (6.8) as (2.2) and (2.6)

[8]) One can, of course, obtain the representation (6.8) and the properties of g and g^* à la Theorem 1, etc.

bear to (2.1). It should also be noted that the relationship between k and g is the same as that between h and f; and furthermore, we have

$$k = h(1-j_1), \quad h = k(1-j_1), \quad h^* = 1-k^*, \quad k^* = 1-h^*, \tag{6.13}$$

which should be compared with (6.6) and (6.7).

Some of the examples of associative functions given previously take on a simpler form when expressed in terms of S, g, k instead of T, f, h. For instance, the functions k_p corresponding to the functions h_p of Example (*a*) of Section 3 are given by

$$k_p = (1-j_1^p)^{1/p}, \tag{6.14}$$

and the t-conorms, S_p, determined by these k_p via (6.12) are given by

$$S_p(a, b) = (a^p + b^p - a^p \cdot b^p)^{1/p}. \tag{6.15}$$

These latter form a family of particularly simple associative functions. It is also worth noting that, for $p \geqq 1$, the graphs of the functions k_p are the indicatrices of Minkowski geometries.

Returning to the remarks made at the beginning of this section, suppose that T is a t-norm and S the corresponding t-conorm. Then, on substituting (6.3) into the Menger triangle inequality (5.2), we have

$$F_{pr}(x+y) \geqq 1 - S(1-F_{pq}(x), 1-F_{qr}(y)),$$

or

$$1-F_{pr}(x+y) \leqq S(1-F_{qp}(x), 1-F_{qr}(y)),$$

which, when expressed in terms of the tails of the distribution functions involved, becomes simply

$$G_{pr}(x+y) \leqq S(G_{pq}(x), G_{qr}(y)). \tag{6.16}$$

7. Equilateral spaces. The function g^* introduced in (6.7) and (6.9) increases steadily from 0 to 1 as its argument increases from 0 to $+\infty$. It thus has all the properties that are required of a distance distribution function in an SM space (see Definition 1). This observation leads at once to the following:

Lemma. *Suppose that the points p, q, r are the vertices of an equilateral triangle in an SM space, i. e., that $F_{pq} = F_{qr} = F_{pr} = F$. Suppose further that, on the interval $[0, \infty)$, we have $F = g^*$, where g^* is the inverse of an additive generator g of a strict t-conorm S. Then, for the triple of points, p, q, r, the Menger triangle inequality* (5.2) *holds, with strict equality, under the t-conorm S — that is to say, for all $x, y \geqq 0$,*

$$F_{pr}(x+y) = S(F_{pq}(x), F_{qr}(y)). \tag{7.1}$$

PROOF. Using the representation (6.8), we have

$$\begin{aligned} S(F_{pq}(x), F_{qr}(y)) &= g^*[g(F_{pq}(x)) + g(F_{qr}(y))] \\ &= g^*[g(g^*(x)) + g(g^*(y))] \\ &= g^*(x+y) \\ &= F_{pr}(x+y). \end{aligned}$$

The simplest SM spaces are the *equilateral spaces* [17, pp. 321—322], namely those in which all pairs of *distinct* points have the same distance distribution. It follows that any three distinct points in an equilateral SM space form an equilateral triangle. Consequently, we have

Theorem 8. *Let S be a strict t-conorm. Then there exists an equilateral SM space such that, for all triples of distinct points in the space, the Menger triangle inequality holds with strict equality under the t-conorm S, i.e., in the form* (7.1). *Moreover, this result is the best possible, in the sense that, in* (7.1), *the t-conorm S cannot be replaced by any stronger function.*

PROOF. Let $\mathfrak{S}$ be a given set; and let g be an additive generator of the t-conorm S, with inverse g^*. For any pair of points p, q in $\mathfrak{S}$ define F_{pq} by:

$$F_{pq}(x) = \begin{cases} 0, & x \leqq 0, \\ 1, & x > 0, \quad p = q, \\ g^*(x), & x > 0, \quad p \neq q. \end{cases}$$

This construction leads to an SM space which is clearly equilateral. The remainder of the proof is now an immediate consequence of the preceding lemma.

It should be noted that since, in view of Examples (*a*) and (*b*) of Section 3, there exist sequences of t-conorms which converge to the strongest possible t-conorm — the t-conorm of T_w —, the S in Theorem 8 may be arbitrarily strong.

Theorem 8 is a counterpart to some results of E. THORP [20] which, for any given t-norm T, enable one to construct a Menger space for which T is the strongest possible t-norm. THORP's results are obtained in a manner quite different from ours. This is to be expected. For f^*, the inverse of an additive generator of a strict t-norm, is not a distribution function — a fact which makes the considerations of Theorem 8 and the preceding lemma inapplicable to the case of t-norms.

8. Convexity and copulas. In Section 6 we encountered a family of functions whose graphs are indicatrices for Minkowski geometries. Now as is well known, in a Minkowski space, the convexity of the indicatrix is a

necessary and sufficient condition for the validity of the (ordinary) triangle inequality. It is therefore not unnatural to ask whether convexity plays any significant role in the theory of SM spaces and t-norms. The following discussion seems to indicate that the answer is in the affirmative.

Theorem 9. *A strict t-norm T satisfies the inequality*

$$T(a, d) + T(c, b) \leqq T(a, b) + T(c, d), \tag{8.1}$$

for all a, b, c, d such that $0 \leqq a \leqq c \leqq 1$, $0 \leqq b \leqq d \leqq 1$, if and only if any additive generator f of T is convex.

[N. B. The inequality (8.1) says that T is non-decreasing in the sense in which this term is applied to a 2-dimensional distribution function.]

PROOF. Suppose that T satisfies (8.1) and let a, b be any two numbers such that $0 \leqq a \leqq b \leqq 1$. Set $u = f^*\left(\frac{1}{2} f(a)\right)$, $v = f^*\left(\frac{1}{2} f(b)\right)$, so that $a = f^*(2f(u))$, $b = f^*(2f(v))$. Then $0 \leqq u \leqq v \leqq 1$; and using (8.1), we have

$$T(u, v) + T(v, u) \leqq T(u, u) + T(v, v).$$

Now by Theorem 4, $T(u, u) = f^*(2f(u))$, $T(v, v) = f^*(2f(v))$; and $T(u, v) = T(v, u) = f^*(f(u) + f(v))$. Consequently,

$$2f^*(f(u) + f(v)) \leqq f^*(2f(u)) + f^*(2f(v)) = a + b.$$

Therefore,

$$\frac{1}{2}(a + b) \geqq f^*(f(u) + f(v)) = f^*\left(\frac{1}{2} f(a) + \frac{1}{2} f(b)\right).$$

But since f is decreasing, this means

$$f\left(\frac{a + b}{2}\right) \leqq \frac{f(a) + f(b)}{2},$$

whence f, being continuous, is convex.

To prove the converse, suppose that f is convex. Then f^*, being the inverse of a non-increasing, invertible convex function, is convex. Accordingly, for any λ between 0 and 1, and any x, y we have

$$f^*(\lambda x + (1 - \lambda) y) \leqq \lambda f^*(x) + (1 - \lambda) f^*(y),$$
$$f^*((1 - \lambda) x + \lambda y) \leqq (1 - \lambda) f^*(x) + \lambda f^*(y).$$

Adding these two inequalities yields,

$$f^*(\lambda x + (1 - \lambda) y) + f^*((1 - \lambda) x + \lambda y) \leqq f^*(x) + f^*(y). \tag{8.2}$$

Now let a, b, c, d be such that $0 \leqq a \leqq c \leqq 1$, $0 \leqq b \leqq d \leqq 1$; and let $u = f(a)$, $v = f(b)$, $s = f(c)$, $t = f(d)$. Since f is decreasing, we have $s \leqq u$,

$t \leqq v$. There are now two cases: (1) either $s<u$ or $t<v$, or both; (2) $s=u$ and $t=v$. In the first case, on setting $x=u+v$, $y=s+t$ and substituting in (8.2), we obtain

$$f^*(\lambda(u+v)+(1-\lambda)(s+t))+f^*((1-\lambda)(u+v)+\lambda(s+t)) \leqq \leqq f^*(u+v)+f^*(s+t). \tag{8.3}$$

Next, choose $\lambda=\dfrac{u-s}{u-s+v-t}$, whence $1-\lambda=\dfrac{v-t}{u-s+v-t}$. Then, after some simplification, (8.3) reduces to

$$f^*(u+t)+f^*(s+v) \leqq f^*(u+v)+f^*(s+t). \tag{8.4}$$

In the second case, i. e., when $s=u$ and $t=v$, (8.4) is trivial. Thus (8.4) holds whenever $s \leqq u$, $t \leqq v$. But (8.4) is equivalent to

$$f^*(f(a)+f(d))+f^*(f(c)+f(b)) \leqq f^*(f(a)+f(b))+f^*(f(c)+f(d)),$$

which, by Theorem 4, is in turn equivalent to (8.1). This completes the proof of Theorem 9.

In order to discuss the implications of Theorem 9, we need the concept of a "copula" which was introduced by one of us in [18].

Definition 4. A (2-dimensional) *copula* is a 2-place function C, which is defined and continuous on the unit square and satisfies the following conditions:

$$\begin{aligned} &(A) \quad C(0,0)=0,\ C(a,1)=C(1,a)=a. \\ &(B) \quad C(a,b) \leqq C(c,d), \text{ whenever } a \leqq c,\ b \leqq d. \\ &(C) \quad C(a,d)+C(c,b) \leqq C(a,b)+C(c,d) \text{ whenever } a \leqq c,\ b \leqq d. \end{aligned} \tag{8.5}$$

Note that the Conditions (A) and (B) of (8.5) are the same as the Conditions (0.1) and (0.2) which are satisfied by all t-norms. Indeed, many (but not all) copulas are t-norms, and many (but not all) t-norms are copulas. For example, the 2-place function B defined by: $B(x,y)=xy+xy(1-x)(1-y)$ is a copula, but is not associative and hence not a t-norm; while on the other hand, the t-norm T_w is not continuous and hence not a copula.

The following properties of copulas were presented in [18]:[9])

1. Let V be a 2-dimensional distribution function with margins F_1 and F_2, i. e., $F_1(x)=V(x,+\infty)$ and $F_2(x)=V(+\infty,x)$, for all x. Let R_1, R_2 denote the ranges of F_1, F_2, respectively. Then there exists a unique function H, whose domain is $R_1 \times R_2$ (and is therefore a subset of the closed unit square), such that for all x, y,

$$V(x,y)=H(F_1(x),F_2(y)). \tag{8.6}$$

[9]) In [18] these results were stated for the general n-dimensional case.

2. The function H may be extended (generally in more than one way) to a copula C, which, being an extension of H, satisfies

$$V(x, y) = C(F_1(x), F_2(y)). \tag{8.7}$$

3. Let F_1 and F_2 be two 1-dimensional distribution functions and C a copula. Then the 2-place function V defined by (8.7) is a 2-dimensional distribution function whose margins are F_1 and F_2.

4. The weakest (2-dimensional) copula is the t-norm T_m, defined in (3.8), and the strongest copula coincides with the strongest t-norm, Min.[10])

Restated in terms of copulas, Theorem 9 becomes:

Theorem 10. *A strict t-norm T is a copula if and only if any (and hence every) additive generator of T is convex.*

Corollary. *If an additive generator of the t-norm T is convex, then T is not weaker than T_m.*

Looked at from the point of view of SM spaces, Theorem 10 says that an additive generator of a t-norm T is convex if and only if the 2-place function V_{pqr}, defined in terms of the triple (p, q, r) of points in the SM space by

$$V_{pqr}(x, y) = T(F_{pq}(x), F_{qr}(y)),$$

is a *possible joint distribution function for the distances from p to q and from q to r.* If V_{pqr} is in fact this joint distribution function, and if the SM space is a Menger space under T, then the generalized triangle inequality (5.2) simply states that the probability that the distance from p to r is less than $x+y$ is at least as large as the joint probability that the distance from p to q is less than x and the distance from q to r is less than y. Thus assuming (or showing) that V_{pqr} is indeed the joint distribution function for the distances from p to q and from q to r considerably strengthens the interrelationship between the geometric and probabilistic aspects of a statistical metric space.

Bibliography

[1] N. H. Abel, Untersuchungen der Functionen zweier unabhängig veränderlichen Größen x und y wie $f(x, y)$, welche die Eigenschaft haben, daß $f(z, f(x, y))$ eine symmetrische Function von x, y und z ist, *J. Reine Angew. Math.*, **1** (1826), 11—15. (Oeuvres complètes de N. H. Abel, v. 1, Christiania (1881), 61—65.)

[2] J. Aczél, Über eine Klasse von Funktionalgleichungen, *Comment. Math. Helv.* **21** (1948), 247—256.

[10]) The corresponding result for distribution functions is due to M. Fréchet [12].

[3] J. Aczél, Sur les opérations définies pour nombres réels, *Bull. Soc. Math. France*, **76** (1949), 59—64.

[4] J. Aczél, A solution of some problems of K. Borsuk and L. Jánossy, *Acta Phys. Hungar.* **4** (1955), 351—362.

[5] J. Aczél, V. D. Belousov and M. Hosszú, Generalized associativity and bisymmetry on quasigroups, *Acta Math. Hungar.* **11** (1960), 127—136.

[6] W. Blaschke und G. Bol, Geometrie der Gewebe, *Berlin*, 1938.

[7] L. E. J. Brouwer, Die Theorie der endlichen kontinuierlichen Gruppen unabhängig von den Axiomen von Lie, *Math. Ann.* **67** (1909), 246—267.

[8] R. Caccioppoli, L'equazione funzionale $f(x+y)=F(f(x), f(y))$, *Giorn. Math. Battaglini*, **66** (1928), 69—74.

[9] É. Cartan, La théorie des groupes finis et continus et l'Analysis Situs, (Mém. Sci. Math., 42) *Paris*, 1930.

[10] A. H. Clifford, Totally ordered commutative semigroups, *Bull. Amer. Math. Soc.*, **64** (1958), 305—316.

[11] Al. C. Climescu, Sur l'équation fonctionelle de l'associativité, *Bull. École Polytechn. Jassy*, **1** (1946), 1—16.

[12] M. Fréchet, Sur les tableaux de corrélation dont les marges sont données, *Ann. Univ. Lyon*, Sect. A(3), **14** (1951), 53—77.

[13] M. Hosszú, Some functional equations related with the associative law, *Publ. Math. Debrecen*, **3** (1954), 205—214.

[14] K. Menger, Statistical metrics, *Proc. Nat. Acad. Sci. U. S. A.*, **28** (1942), 535—537.

[15] T. S. Motzkin, Sur le produit des espaces métriques, *C. R. du Congrés d'Oslo*, (1936), 137.

[16] B. Schweizer and A. Sklar, Espaces métriques aléatoires, *C. R. Acad. Sci. Paris*, **247** (1958), 2092—2094.

[17] B. Schweizer and A. Sklar, Statistical metric spaces, *Pacific. J. Math.*, **10** (1960), 313—334.

[18] A. Sklar, Fonctions de répartition à n dimensions et leur marges, *Publ. Inst. Statist. Univ. Paris*, **8** (1959), 229—231.

[19] G. Thomsen, Un teorema topologico sulle schiere di curve e una carratterizatione geometrica delle superficie isotermo-asintotiche, *Boll. Unione Mat. Ital. Bologna*, **6** (1927), 80—85.

[20] E. O. Thorp, Best-possible triangle inequalities for statistical metric spaces, *Proc. Amer. Math. Soc.* **11** (1960), 734—740.

(Received September 13, 1960.)