ZASTOSOWANIA MATEMATYKI
APPLICATIONES MATHEMATICAE
XIII, 3 (1973)

J. TABOR (Kraków)

SOME REMARKS ON DISTANCE IN NATURAL SCIENCES

Marczewski and Steinhaus have introduced in [3] the notion of systematic distance of biotopes. This distance has been defined in the following manner.

Let us consider two biotopes A and B. Let us denote the number of species in biotop A by a, the number of species in biotop B by b and the number of common species for both biotopes by w. Then we put

$$\varrho(A, B) = 1 - \frac{w}{a+b-w} = \frac{a+b-2w}{a+b-w}. \tag{1}$$

The distance defined in such a manner satisfies the following condition:

$$\varrho(A, B) \leqslant 1. \tag{2}$$

A more general notion is the distance of functions (see [3], p. 200). Let X be a space with measure m. The distance of functions f, g defined on a set X is given by the following formula:

$$\varrho(f, g) = \frac{\int |f-g|\, dm}{\int \max(|f|, |g|, |f-g|)\, dm}. \tag{3}$$

We are going to propose some definitions of distance, different than those presented above. These definitions could be utilized in natural sciences.

Let us consider a class K of objects and let us assume that we have distinguished some measurable feature X of those objects. The measure of this feature of object A will be denoted by $m(A)$. As such measure we may take e. g. the degree of souring of soil, the level of underground water, the gradient of inclination of a slope, the degree of salinity of water, the number of species in a biotop, etc. Every such measure determines some natural metric (distance) defined by the formula

$$\varrho(A, B) = |m(A) - m(B)|. \tag{4}$$

The metric defined in such a manner usually does not satisfy condition (2). We are going to give some ways of normalization of the metric defined by formula (4).

1. Let us assume that the expression $|m(A)-m(B)|$ is bounded from above. Let c denote the l.u.b. (the supremum if it exists) of the expression $|m(A)-m(B)|$ (for $m(A)$ and $m(B)$ running all values of the feature X).

Let us assume that c is a positive number (the case where $c = 0$ is trivial).

We put

$$\varrho(A, B) = \frac{|m(A)-m(B)|}{c}. \tag{5}$$

It is easy to see that formula (5) defines a metric satisfying condition (2). The distance $\varrho(A, B)$ is equal to 0 iff $m(A)$ is equal to $m(B)$ and $\varrho(A, B)$ is equal to 1 iff $|m(A)-m(B)|$ is equal to c.

2. Let us assume that the measure of feature X is bounded from below. Let d denote the g.l.b. (the infimum if it exists) of this measure. Then we put

$$\varrho(A, B) = \begin{cases} 0 & \text{if } m(A) = m(B) = d, \\ \dfrac{|m(A)-m(B)|}{\max\big(m(A)-d,\, m(B)-d\big)} & \text{in other cases.} \end{cases} \tag{6}$$

We are going to prove that formula (6) defines a metric. To this aim it is sufficient to verify the triangle inequality (the verification of other axioms of a metric is trivial). Let us write $m(A) = x$, $m(B) = y$ and $m(C) = z$.

We are going to consider the following cases:

1. $x = y$ or $y = z$ or $x = z$,
2. $x > y > z$,
3. $z > y > x$,
4. $y > x > z$,
5. $x > z > y$,
6. $y > z > x$,
7. $z > x > y$.

The verification of the triangle inequality in case 1 is trivial. In case 2 we have

$$\varrho(A, B)+\varrho(B, C) = \frac{x-y}{\max(x-d, y-d)} + \frac{y-z}{\max(y-d, z-d)}$$

$$= \frac{x-y}{x-d}+\frac{y-z}{y-d} \geqslant \frac{x-y+y-z}{x-d} = \frac{x-z}{x-d} = \frac{x-z}{\max(x-d, z-d)}$$

$$= \varrho(A, C).$$

Now we prove the following

LEMMA. *If $y > x \geqslant 0$ and $c \geqslant 0$, then*

$$\frac{x}{y} \leqslant \frac{x+c}{y+c}$$

for arbitrary real numbers x, y, c.

Proof. By the assumption we have $cx \leqslant cy$. Then $xy+cx \leqslant xy+cy$, i. e. $x(y+c) \leqslant y(x+c)$. Dividing both sides of this inequality by $y(y+c)$, we obtain

$$\frac{x}{y} \leqslant \frac{x+c}{y+c}.$$

This completes the proof of the lemma.

In order to prove the triangle inequality in case 4, let us observe that

$$\varrho(A, C) = \frac{x-z}{x-y}, \qquad \varrho(B, C) = \frac{y-z}{y-d}, \qquad \varrho(A, B) = \frac{y-x}{y-d}.$$

Using the lemma, we obtain

$$\frac{x-z}{x-d} \leqslant \frac{x-z+y-x}{x-d+y-x} = \frac{y-z}{y-d}.$$

Thus we have

$$\varrho(A, C) = \frac{x-z}{x-d} \leqslant \frac{y-z}{y-d} \leqslant \frac{y-z+y-x}{y-d} = \varrho(B, C)+\varrho(A, B),$$

which verifies the triangle inequality in case 4.

The proof (of the triangle inequality) in cases 3, 5 and 7 is analogous as in case 2. Proof in case 6 is similar to the proof in case 4.

The distance defined by formula (6) satisfies condition (2). This distance, contrary to the distance defined by formula (5), has a heterogeneous degree of differentiation of objects.

If $m(A)$ and $m(B)$ are near d, then even a little difference $m(A)-m(B)$ can imply a large distance of the objects A and B. On the other hand, if $m(A)$ and $m(B)$ are large, then even for large differences $m(A)-m(B)$ the distance $\varrho(A, B)$ can be a little one.

3. Let us assume that the measure of feature X is bounded from above. Let e denote the l.u.b. (the supremum if it exists) of this measure.

We put

$$\varrho(A,B) = \begin{cases} 0 & \text{if } m(A) = m(B) = e, \\ \dfrac{|m(A)-m(B)|}{\max\big(e-m(A),\, e-m(B)\big)} & \text{in other cases.} \end{cases} \tag{7}$$

Formula (7) defines a metric. Proof of the triangle inequality is similar to the proof of the triangle inequality of metric (6). Metric (7) obviously satisfies condition (2). It has a heterogeneous degree of differentiation. If $m(A)$ and $m(B)$ are near to e, then even for little differences $m(A)-m(B)$ the distance $\varrho(A,B)$ can be large.

Remark. We may (if it is convenient) replace c in formula (5) by an arbitrary number $c' > c$, d in formula (6) by an arbitrary number $d' < d$, and e in formula (7) by an arbitrary number $e' > e$. Then formulas obtained in such a manner define metrics satisfying condition (2).

It seems that the distances defined by formulas (6) and (7) could be used in the case where the feature X has the following property:

There exists such a number d that if values of the feature X are near to d, then even their little difference generates considerable differentiation of the habitat, but the values are far — away from d the difference is less important.

The features having such property are, for instance, the degree of salinity of water, the quantity of rainfall, etc.

Now, let us assume that the considered objects are characterized by n features, and that the distance satisfying condition (2) has been defined for every feature. Let $\varrho_i(A,B)$ denote the distance of the objects A and B with respect to the i-th feature.

Let us consider an arbitrary sequence $a_1, a_2, \ldots, a_n$ of positive real numbers such that

$$a_1 + a_2 + \ldots + a_n = 1. \tag{8}$$

We define the distance $\varrho(A,B)$ of the objects A and B by the formula

$$\varrho(A,B) = a_1\varrho_1(A,B) + a_2\varrho_2(A,B) + \ldots + a_n\varrho_n(A,B). \tag{9}$$

It is easy to see that formula (9) defines the metric satisfying condition (2). Besides, this metric satisfies two following conditions:

$$\varrho(A,B) = 0 \quad \text{iff} \quad \varrho_1(A,B) = \varrho_2(A,B) = \ldots = \varrho_n(A,B) = 0,$$

$$\varrho(A,B) = 1 \quad \text{iff} \quad \varrho_1(A,B) = \varrho_2(A,B) = \ldots = \varrho_n(A,B) = 1.$$

This means that the distance of objects is equal to 0 iff those objects are indentical with respect to every feature, and is equal to 1 iff the object A differs from the object B to a maximum (in sense of the distance) with respect to every feature.

Formula (9) defines a synthetic distance of objects. In this formula the coefficient a_i is a measure of an importance of the i-th feature. If we assume that the considered features are equiponderant one to other, i. e. that

$$(10) \qquad a_1 = a_2 = \ldots = a_n = \frac{1}{n},$$

then formula (9) takes the form

$$(11) \qquad \varrho(A, B) = \frac{\varrho_1(A, B) + \ldots + \varrho_n(A, B)}{n}.$$

Example. In the note [2] there was taken, as a feature characterizing ground, 1. an exposition, 2. a decline, 3. a soil reaction (PH), 4. a geological structure, and 5. a kind of the usage. Two little regions A and B (in the note [2] they were denoted by 41 and 45, respectively) were characterized as follows:

(A) 1. 145°, 2. 9°, 3. 6, 4. fleksschiefer, 5. plough-land;

(B) 1. 120°, 2. 12°, 3. 5, 4. inoceram beds, 5. plough-land.

In the researched region we have supremum of the difference of the expositions — 180°, of the declines — 35°, and of PH — 4.5.

Using formula (5), we obtain

$$\varrho_1(A, B) = \frac{25^\circ}{180^\circ} \approx 0.14, \qquad \varrho_2(A, B) = \frac{3^\circ}{35^\circ} \approx 0.09,$$

$$\varrho_3(A, B) = \frac{1}{4.5} \approx 0.22.$$

Features 4 and 5 are unmeasurable. Therefore, $\varrho_4(A, B)$ and $\varrho_5(A, B)$ we have to give by ourself.

Let us put $\varrho_4(A, B) = 0.5$ and $\varrho_5(A, B) = 0$. Using formula (11), we obtain

$$\varrho(A, B) = \frac{0.14 + 0.09 + 0.22 + 0.5 + 0}{5} = 0.19 \approx 0.2.$$

The calculated distances of objects of a given class may be used for classifying those objects. It is possible to do it using e. g. the Czekanowski method (cf., e. g., [1], p. 118-120).

References

[1] F. Fukarek, *Fitosocjologia*, PWRiL, Warszawa 1967.
[2] J. Lach, *Typologia fizjotopów zachodniej części Beskidu Niskiego*, unpublished dissertation, WSP, Kraków 1972.
[3] E. Marczewski i H. Steinhaus, *O odległości systematycznych biotopów*, Zastosow. Matem. 4 (1959), p. 195-203.

Received on 26. 9. 1972

J. TABOR (Kraków)

UWAGI O ODLEGŁOŚCI W NAUKACH PRZYRODNICZYCH

STRESZCZENIE

W pracy podajemy pewne propozycje zdefiniowania odległości w naukach przyrodniczych.

Rozważmy pewną klasę obiektów K i załóżmy, że wyróżniono pewną cechę mierzalną tych obiektów. Miara tej cechy dyktuje w klasie K pewną naturalną metrykę (odległość) zdefiniowaną wzorem (4). Metryka taka nie spełnia na ogół warunku (2). Dla odpowiednich założeń wzory (5), (6) i (7) definiują (za pomocą metryki (4)) metrykę spełniającą już warunek (2). Metryki (6) i (7) wydają się odpowiednie dla cech spełniających następujący warunek:

Istnieje liczba d taka, że jeśli wartości rozważanej cechy są bliskie d, to nawet niewielkie ich różnice powodują znaczne zróżnicowanie środowiska, a im wartości cechy są dalsze od d — tym tego samego rzędu różnice ich wartości grają mniejszą rolę.

Za taką cechę można uważać np. stopień zasolenia wody czy wielkość opadów.

Jeśli obiekty klasy K są scharakteryzowane przez n cech i, ze względu na każdą z tych cech, odległość jest zdefiniowana, to syntetyczną odległość obiektów klasy K można zdefiniować wzorem (9). Współczynnik a_i w tym wzorze jest miarą istotności (ważności) i-tej cechy. Jeśli rozważane cechy uznamy za równorzędne, to wzór (9) można zastąpić wzorem (11).
