ZASTOSOWANIA MATEMATYKI
V (1960)

J. PERKAL (Wrocław)

# *SUR LA NOTION DE RESSEMBLANCE RELATIVISÉE À LA POPULATION* *

Bien que la similitude (géométrique) soit une notion connue depuis l'antiquité (sculptures, peintures, cartes etc.), elle ne se prête pas, au moins dans son aspect mathématique élémentaire, à saisir les relations de ressemblance entre les individus du monde vivant. Il n'existe même pas deux êtres géométriquement semblables, à moins qu'il ne soient de dimensions identiques. En effet, pour qu'un arbre par exemple soit capable de faire circuler sa sève, la section transversale de son tronc devrait être à peu près proportionnelle à son volume $V$, donc son diamétre à $\sqrt{V}$. Pour la même raison, le diamètre d'un arbre de volume $4V$ devrait être $\sqrt{4} = 2$ fois plus grand. Or pour que la similitude géométrique se présente, ce diamètre, en tant qu'une dimension linéaire, ne peut augmenter que $\sqrt[3]{4} < 2$ fois.

Cependant, les naturalistes se servent notoirement de la notion de ressemblance bien plus générale que l'identité et, sans doute, une telle notion est nécessaire. On peut lui attacher évidemment plusieurs sens différents. Cette communication a pour but de proposer deux tentatives de le faire.

Soient $x, y, z, \ldots$ les individus d'une population $P$. Ces individus sont définis par les grandeurs de $k$ propriétés $c_1, c_2, \ldots, c_k$ et peuvent être traités comme des points d'un espace euclidien á $k$ dimensions. Le symbole $xSy$ désignera que $x$ et $y$ *se ressemblent* (sont semblables). Admettons que

1. Si $x = y$, on a $xSy$ (la ressemblance est plus générale que l'identité).
2. Si $xSy$, on a $ySx$ (la ressemblance est une relation symétrique).
3. Si $xSy$ et $ySz$, on a $xSz$ (la ressemblance est une relation transitive).

La loi de transitivité restreint la notion de ressemblance employée

* Referat wygłoszony na Sympozjum Biometrycznym w Budapeszcie we wrześniu 1959 r.

par les naturalistes. Ils conçoivent, en effet, l'existence des chaînes intermédiaires entre deux individus dissemblables, formées d'individus-chaînons dont chacun ressemble à ses voisins immédiats. En vertu de la propriété 3, une telle possibilité est exclue pour nous.

Deux individus seront considérés comme *identiques*, lorsqu'ils sont indiscernables par les grandeurs de leurs propriétés mesurées $c_1, c_2, \ldots, c_k$, c'est-à-dire que leurs points individuels dans l'espace à $k$ dimensions coïncident. Deux individus seront dits *semblables au sens* $S$ lorsque leurs points individuels se trouvent dans un même ensemble d'une famille $S$ définie d'avance et dont nous dirons qu'elle *induit* la relation $xSy$ (vu que cette induction est univoque). Par exemple, la famille $S$ de toutes les droites passant par l'origine des coordonnées induit la similitude géométrique.

Une famille $S$ et, par suite, la ressemblance $xSy$ induite par elle étant données, il existe une *transformation laissant invariante* la ressemblance $S$, c'est-à-dire une transformation telle que tout point $x$ et son transformé appartiennent à un même ensemble de la famille $S$. Une telle transformation sera désignée par $S(x)$. Pour le cas, par exemple, où $S$ est la similitude géométrique, telle est la transformation $S(x) = a \cdot x$ avec $a$ constant.

Toute grandeur qui est un invariant de la transformation $S(x)$ sera dite un *indice* de la ressemblance $S$. Les quotients de la forme $c_i/c_j$ par exemples ont donc des indices de la similitude géométrique. En outre, les indices

$$\frac{c_1}{c_2}, \frac{c_2}{c_3}, \ldots, \frac{c_{k-1}}{c_k}.$$

constituent un système *complet* d'indices de cette similitude, c'est-à-dire que l'égalité de tous ces indices d'un individu aux indices correspondants d'un autre est une condition nécessaire et suffisante pour leur similitude géométrique.

Or le sens usuel du terme ressemblance se montre relatif. Deux Chinois qualifiés semblables dans la population des habitants de l'Europe peuvent être dissemblables dans une population des Chinois. Il est donc naturel de lier la notion de la ressemblance à la population.

Je vais distinguer ici deux genres de ressemblance relativisée à la population. J'en appellerai la première *la ressemblance par déviations* et la seconde — *la ressemblance naturelle.*

Si une population est homogène, nous pouvons y considérer le *centre*, c'est-à-dire un individu virtuel dont les grandeurs des propriétés sont égales aux moyennes. Il sera désigné par $p$. Deux individus $x$ et $y$ s'appelleront *semblables par déviations* lorsqu'ils dévient du centre $p$ par les grandeurs de leurs propriétés d'une manière proportionnelle. Ainsi, si $x$ et $y$ diffèrent de $p$ par les valeurs $c_i$ (où $i = 1, 2, \ldots, k$) le rapport de ces différences étant constant et indépendant de $i$, $x$ et $y$ se ressemblent bien dans ce sens. Si par exemple un homme ne diffère de centre de la population que par la longueur de son nez, les hommes qui lui ressemblent par déviations ne différeront du centre que par cette propriété.

Une carricature consiste à accentuer les traits caractéristiques du modèle, ce qui peut être atteint en multipliant proportionnellement les différences entre les propriétés du modèle et celles du centre. Alors la carricature ressemble par déviations au modèle.

La famille $S$ d'ensembles qui induit la ressemblance par déviations est celle de toutes les droites passant par le centre $p$ de la population (et non pas par l'origine, comme dans la cas de la similitude géométrique). Deux individus se ressemblent donc par déviations, lorsque la droite unissant leurs points individuels dans l'espace à $k$ dimensions passe par le centre $p$. Désignons par $\bar{c}_1, \bar{c}_2, \ldots, \bar{c}_k$ ses coordonnées. La transformation laissant invariante cette ressemblance est de la forme (vectorielle) $S(x) = a \cdot (x - p)$, c'est-à-dire $S(c_i) = a \cdot (c_i - \bar{c}_i)$ dans sa forme numérique, le coefficient $a$ étant une constante.

Les indices de cette ressemblance sont les quotients

$$\frac{c_i - \bar{c}_i}{c_j - \bar{c}_j},$$

et le système suivant:

$$\frac{c_1 - \bar{c}_1}{c_2 - \bar{c}_2}, \quad \frac{c_2 - \bar{c}_2}{c_3 - \bar{c}_3}, \quad \ldots, \quad \frac{c_{k-1} - \bar{c}_{k-1}}{c_k - \bar{c}_k}$$

est un système complet des ces indices.

Néanmoins, la notion de ressemblance par déviations n'a pas toujours un sens naturel intéressant. Considérons par exemple la population d'enfants de divers âge caracterisés par deux propriétés, à savoir par la taille et le poids. Le centre y dépend de la distribution de l'âge dans la population et, en conséquence, ne présent pas d'intérêt biologique, pas plus que la ressemblance par déviations.

Dans de tels cas c'est la notion de ressemblance naturelle qui permet de mieux saisir l'intuition du naturaliste. Le rôle du centre y est joué non pas par un point, mais par une ligne.

Si l'on subdivise la même population d'enfants en groupes d'âge et l'on trouve les centres de chaque groupe, on voit ces centres se disposer le long d'une ligne (ligne de développement). Mais aussi dans d'autres cas le centre d'une population peut être privé de valeur biologique. Une population est représentée dans l'espace par un nuage des points plus ou moins allongé. Plus sphérique est la forme de ce nuage, plus grand rôle y revient au centre de la population. Par contre, plus ce nuage est allongé, plus grand rôle y joue une ligne. En considérant les populations comme ne s'éloignant pas trop de la distribution normale, on peut admettre que cette ligne est une droite (axe principal de Hotelling [1]).

Deux individus peuvent être alors regardés comme semblables au sens naturel lorsque leurs points individuels son situés sur une droite parallèle à l'axe principal.

Or la détermination numérique de cet axe étant d'habitude fort fastidieuse, je le remplace par une autre droite, à savoir par celle que j'ai appelée l'*axe naturel* (voir [2]) et qui a été considérée aussi par Teissier (voir [3]). J'entend par là l'axe ayant pour composantes les écarts-types $\sigma_1, \sigma_2, \ldots, \sigma_k$ des propriétés $c_1, c_2, \ldots, c_k$ respectivement. En conséquence, j'appelle deux individus *semblables au sens naturel* lorsque les différences entre les valeurs de leurs propriété $c_i$ sont proportionelles aux écarts-types $\sigma_i$ de ces propriétés, le coefficient de proportionnalité étant constant (indépendant de $i$).

La famille induisant la ressemblance naturelle est par suite celle des droites parallèles à l'axe naturel. On peut donc dire que deux individus sont semblables dans ce sens relativement à une population, lorsque la droite unissant leurs points individuels est parallèle à l'axe naturel de cette population. La transformation qui laisse invariante la ressemblance naturelle est la translation de l'espace, parallèle à l'axe naturel. Ses invariants, donc ses indices naturels, sont par conséquent les paramètres arbitraires caractérisants les droites parallèles à l'axe naturel.

Formons un système de ces indices. Normons dans ce but les propriétés $c_1, c_2, \ldots, c_k$ de manière que leur valeur moyenne soit 0 et leur variance soit 1:

$$\gamma_i = \frac{c_i - \bar{c}_i}{\sigma_i} \qquad (i = 1, 2, \ldots, k).$$

Normons les $\gamma_i$ encore une fois à l'aide de leur. moyenne

$$\bar{\gamma} = \frac{1}{k}(\gamma_1 + \gamma_2 + \ldots + \gamma_k)$$

de manière que la valeur moyenne d'une propriété arbitraire de chaque individu soit 0:

$$w_i = \gamma_i - \gamma \qquad (i = 1, 2, \ldots, k).$$

Les $w_i$ ainsi définis constituent un système des indices pour la ressemblance naturelle. Ce système n'est pas complet, car les indices n'y sont pas indépendants, à savoir on a $w_1 + w_2 + \ldots + w_k = 0$. Mais il suffit d'en supprimer l'un quelconque pour que le système devienne complet.

J'ai prouvé (voir [2]) que l'égalité des indices naturels qui viennent d'être définis est en effet une condition nécessaire et suffisante pour la ressemblance naturelle de deux individus.

Mme Milicer (cf. [4]) a obtenu des résultats intéressants en appliquant les indices naturels à l'analyse somatique des boxeurs et des autres sportifs.

Exemple. 50 photographies de têtes humaines, faites par M. Michalski, sont jointes au volume XXI.2 de „Przegląd Antropologiczny". On a pris sur ces têtes des mesurements antropologiques suivants: 1. la longueur de la tête, 2. sa largeur, 3. la largeur de la face, 4. sa longueur entière, 5. sa longueur partielle (sans mâchoire), 6. la longueur du nez et 7. sa largeur.


Fig. 1. Géométrique ZM-334

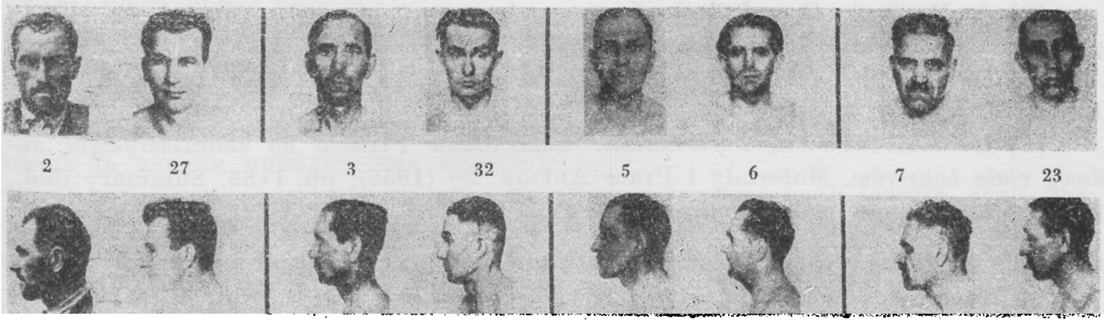


Fig. 2. Par déviations ZM-335

Ces propriétés ont été employées pour calculer les indices de ressemblance géométrique, de ressemblance par déviations et de ressemblance naturelle. La fig. 1 représente trois séries (à savoir 2, 3 et 2) de têtes les plus semblables au sens géométrique parmi les 50 têtes mesurées et photographiées; la fig. 2 en représente quatre séries (couples) de têtes les plus semblables par déviations et la fig. 3 — trois séries (2, 2 et 3) de têtes les plus semblables au sens naturel.

En accord avec leur ressemblance naturelle, les couples formant deux premières séries de la fig. 2, de même que celui composé de deux premiers individus de la troisième série qui y est représentée, se trouvent rangés chacun dans une même classe typologique aussi par M. Michalski, procédant à l'aide de la méthode dite morphologique. Par contre, les


Fig. 3. Naturelle ZM—336

individus appartenant à des mêmes séries des fig. 1 et 2, c'est-à-dire semblables au sens géométrique et par déviations, ont été rangés par M. Michalski dans des classes différentes.

### Travaux cités

[1] H. Hotelling, *Analysis of a complex of statistical variables into principal components*, J. Educ. Psych. 24 (1933), pp. 417-441 et 498-520.

[2] J. Perkal, *O wskaźnikach antropologicznych*, Przegl. Antrop. 19 (1953), pp. 209-221. Résume ibid. 20, pp. 679-680.

[3] G. Teissier, *Allométrie de taille et variabilité chez Maia squinado*, Zool. Exper. 42 (1955), pp. 221-264.

[4] H. Milicerowa, *Zastosowanie wskaźników Perkala do charakterystyki budowy ciała bokserów*, Materiały i Prace Antrop. 20 (1956), pp. 1-85. Summary ibid., pp. 71-77. Résume (russe) ibid., pp. 78-84.

INSTITUT MATHÉMATIQUE DE L'ACADÉMIE POLONAISE DES SCIENCES

*Praca wpłynęła 21. 9. 1959*

J. PERKAL (Wrocław)

## *O POJĘCIU PODOBIEŃSTWA ZE WZGLĘDU NA POPULACJĘ*

### STRESZCZENIE

Podobieństwo geometryczne różni się istotnie od podobieństwa używanego w naukach przyrodniczych. Dwa twory przyrodnicze, np. drzewa różnej wielkości, nie mogą być geometrycznie podobne, gdyż niektóre liniowe wymiary drzewa są proporcjonalne do jego wysokości, inne do pola korony, jeszcze inne do objętości. Pojęcie podobieństwa zależy od populacji, w jakiej to podobieństwo jest rozpatrywane.

W przypadku gdy populacja tworzy chmurę punktów indywidualnych o kształcie podobnym do kuli, środek tej kuli odgrywa ważną przyrodniczo rolę i odchylenia od niego mogą służyć do następującego nowego określenia pojęcia podobieństwa:

Dwa indywidua są *podobne pod względem odchyleń*, jeśli odchylenia poszczególnych cech tych indywiduów od średnich w populacji są proporcjonalne.

W przypadku, gdy kształt chmury punktów indywidualnych populacji jest podłużny, środek chmury przestaje mieć przyrodnicze znaczenie. Jego rolę przejmuje pewna linia (w pierwszym przybliżeniu — prosta) i bliższe intuicjom przyrodnika staje się tzw. podobieństwo przyrodnicze:

Dwa indywidua są *przyrodniczo podobne*, jeśli różnice ich cech są proporcjonalne do standardowych odchyleń tych cech w populacji.

---

Ю. ПЭРКАЛЬ (Вроцлав)

## *О ПОНЯТИИ ПОДОБИЯ ПО ОТНОШЕНИЮ К СОВОКУПНОСТИ*


РЕЗЮМЕ

Геометрическое подобие существенно отличается от общеупотребительного в естествознании, так как оно не может характеризовать двух различных объектов, например двух деревьев (одни измерения дерева пропорциональны длинам, другие площадям, а третьи объемам). Понятие подобия оказывается зависящим от рассматриваемой совокупности.

Когда индивидуализированные точки совокупности образуют в пространстве тучу близкую по форме к сферической, то при помощи отклонений от ее центра, который тогда играет важную роль, можно ввести понятие подобия в следующем новом смысле:

Два индивидуума *похожи* друг на друга *по отклонениям*, если отклонения их примет от примет центра совокупности пропорциональны.

Если же форма тучи продолговатая, то более удобным является подобие называемое естественным подобием. Роль центра в этом случае играет некоторая линия (в первом приближении прямая):

Два индивидуума похожи друг на друга *в естественном смысле*, если разницы их примет пропорциональны, стандартным отклонениям этих примет в совокупности.


---

J. PERKAL (Wrocław)

## *ON THE NOTION OF SIMILARITY WITH RESPECT TO POPULATION*


SUMMARY

Geometrical similarity differs considerably from the similarity used in natural sciences. Two natural objects, for instance two trees of different size, cannot be geometrically similar since some of the linear dimensions of the tree are proportional to its height, others to the area of its crown and still others to its volume. The notion of similarity depends on the population in which the similarity is considered.

In the case where the population forms a cloud of individual points of a shape resembling a sphere, the centre of this sphere plays biologically an important part and deviations from it may be used for the following new definition of the notion of similarity:

Two individuals are *similar* with respect to *deviations* if the deviations of the features of those individuals from the mean features in the population are proportional.

In the case where the shape of the cloud of individual points is oblong, the centre of the cloud ceases to have any biological significance. Its role is taken over by a certain line (in the first approximation — a straight line) and the so called biological similarity becomes nearer to the intuitions of a biologist. It is defined as follows:

Two individuals are *biologically similar* if the differences of their features are proportional to the standard deviations of those features in the population.