1968

ON TWO CRITERIA OF CLASSIFICATION

I. C. Lerman

Centre de Calcul, Maison des Sciences de l'Homme

I. Introduction

Let E be an N-set of objects $x,y,z, \ldots$; let F denote the set of unordered object pairs: $F = \{(x,y), x \varepsilon E, y \varepsilon E, x \neq y\}$. The basic data is a pre-order on F. $p = (x,y)$ precedes $q = (z,t)$ if x and y are less similar than z and t. Let ω denote the graph of pre-order relation, ω is a subset of the product $F \times F$.

A partition of E determines a pre-order on F with two classes R and S; (x,y) belongs to R if and only if x and y are in the same class of the partition, S is the complement of R in F; that is to say; (z,t) belongs to S if and only if z and t belong to distinct classes, $F = R + S$. Any pair of S precedes any pair of R.

One of the essential purposes of taxonomy is to find one partition of E which "best approximates" the similarities of objects. The degree of the similarity between two objects is expected to be high if these belong to the same class and low if not. More precisely, at the beginning we define a numerical function which measures the accordance between a classification (*e.g.* partition) and the graph ω; this function is called criterion. In fact the criterion will be a mapping of Π into N, where Π denotes the set of all partitions of E. Then the problem is to find the partitions which maximise the value of the criterion.

1) J.P. Benzecri proposed as a criterion the cardinal of the intersection of ω and $S \times R$, in $F \times F$.

$$P \varepsilon \Pi \xrightarrow{a_1} a_1\ (P) = |\omega\ S \times R|$$

2) W.F. de la Vega proposed as a criterion the cardinal of the complement of the symmetric difference between ω and $S \times R$, in $F \times F$:

don't use this algorithm: but if you do use it then you are forced by the analysis to say that the straight line is more or less what one wants. I can also explain the monotonic relationship between the two but that would take some time to go in to. There are obvious intuitive reasons for seeing why this should be the case. It also enables one to distinguish between an actual classification - that is something one can put through a sort of classification engine - and random data. If one could not distinguish these it would be a measure of one's failure.

$$P\in\Pi \xrightarrow{a_2} a_2\ (P) = |F \times F - \omega\Delta\ S \times R|$$

Considering the relations:

$$|\omega \vee S \times R| = |\omega\Delta S \times R| + |\omega\wedge S \times R| = |\omega| + |S \times R| - |\omega\wedge S \times R|$$

We obtain: $a_2\ (P) = |F|^2 - |\omega| + 2(\omega\wedge S \times R| - \frac{1}{2}|S| \times |R|)$

Then the optimizing problem consists in finding classifications which maximise $|\omega\wedge S \times R|$ or $|\omega\wedge S \times R| - \frac{1}{2}|S| \times |R|$ accordingly to the first criterion or to the second one.

Our purpose is to compare these two criteria.

II. The theorem of two criteria comparisons

II.A. Preliminaries: On some enumerative problems connected with the partitions of E.

Definitions and Generalities: We call the decreasing sequence of cardinals of the different classes the partition type. So, $(n_1, n_2, n_3, ---, n_i, ---)$; $n_1 \geq n_2 \geq --- n_i \geq ---$ is the type of one partition of E for which the cardinals of the classes are respectively in the decreasing order $n_1, n_2, ---, n_i$ ---. $N = |E| = \sum_{i=1}^{\infty} n_i$. The partition of E has exactly k classes if the last subscript such as $n_i \neq 0$, is k. A subjective mapping f of E into the collection of numbers $(1, 2, ---, k)$ determines a partition of E with k classes to each of which a number is assigned. f will be called "a partition with labelled classes" or more briefly a labelled partition. Let $\mathfrak{F}$ denote the set of the partitions with labelled classes the type of which $(n_1, n_2, ---, n_k)$ is given; if f belongs to $\mathfrak{F}$, $f^{-1}(i) = n_i$ where $f^{-1}(i)$ denotes the class of number i, that can be also noted E_i.

On the other hand, let Π be the set of the "partitions with unlabelled classes" the type of which $(n_1, n_2, ---, n_k)$, is given. The cardinal of $\mathfrak{F}$ is given by the following formula

$$|\mathfrak{F}| = \frac{N!}{n_1!\, n_2! \cdots n_k!}$$

Generally, the integers $n_1, n_2, \cdots, n_k$ are not mutually distinct. If a partition ω of Π has k_1 classes with the same cardinal $\nu_1, \ldots$, k_2 classes with the same cardinal ν_2, ---, and k_r classes with the same cardinal ν_r ($k_1 + k_2 + \cdots + k_r = k$ and $k_1\nu_1 + k_2\nu_2 + \cdots + k_r\nu_r = N$) we have:

$$|\Pi| = \frac{N!}{k_1!\, k_2! \cdots k_n!\, n_1!\, n_2! \cdots n_k!}$$

For one element P of Π the cardinals of R and S are given by the following formulae

$$|R| = \sum_{i=1}^{k} n_i\,(n_i - 1)/2 \text{ and } |S| = \sum_{i<j} n_i n_j \;;$$

on the other hand $|R| + |S| = |F| = N(N-1)/2$. Then $|R|$ determines $|S|$.

2 - Lemmas

2.1 The proportion in $\mathfrak{F}$ of elements for which a given pair (x,y) is contained in the same class is $\sum_{i=1}^{k} n_i\,(n_i-1)/N(N-1)$.

Our purpose is to determine the cardinal of the following subset of $\mathfrak{F}$ $\{ f \in \mathfrak{F} \mid f(x) = f(y) \}$; which may be expressed as a sum of subsets:

$$\{ f\in \mathfrak{F} \mid f(x) = f(y)\} = \sum_{i=1}^{k} \{f\in \mathfrak{F} \mid f(x)=f(y)=i \}$$

Hence $$|\{ f\in \mathfrak{F} \mid f(x) = f(y)\}| = \sum_{i=1}^{k} |\{f\in \mathfrak{F} \mid f(x)=f(y)=i \}|$$

The subset $\{f(x)=f(y)=i\}$ can be mapped 1 - 1 onto the set of labelled partitions of $E - \{x,y\}$; such that the cardinals of different classes of one partition are respectively

$$n_1, n_2, ---, n_{(i-1)}, n_i - 2, n_{(i+1)}, ---, n_k$$

subject to $n_i \geq 2$

Therefore $|\{f(x) = f(y) = i\}| = (N-2)!/n_1!n_2! --- (n_i-2)!--n_k!$ if $n_i \geq 2$ and 0 if not

Hence

$$|\{ f(x) = f(y)\}| = \sum_{\{i, n_i \geq 2\}} (N-2)!/n_1!n_2! --- (n_i-2)!---n_k!$$

The consideration of the ratio of this cardinal on $|\mathfrak{F}|$ leads to the result above.

We abbreviate the following demonstrations which are analogous to this one.

2.2 The proportion in $\mathfrak{F}$ of elements, for which two objects of a given pair (x,y), belong, respectively, to distinct classes, is,

$$2 \sum_{i<j} n_i n_j / N(N-1)$$

$$|\{f \in \mathfrak{F} \; |f(x) \neq f(y)\}| = \sum_{i \neq j} |\{ f \in \mathfrak{F} \; |f(x)=i, \; f(y) = j \}|$$

The sum includes $k(k-1)$ terms.

The set $\{f(x) = i, \; f(y) = j, \; i \neq j\}$; can be mapped 1 - 1 onto the set of "labelled partitions classes" of $E - (x,y)$; such that the cardinals of classes of one partition are respectively:

$$n_1, n_2, ---, n_{(i-1)}, n_i - 1, n_{i+1}, ---, n_{(j-1)}, n_j - 1, n_{j+1}, ---, n_k$$

This remark is sufficient to find the result.

2.3 Let x, y and z be three given objects of E, the proportion in $\mathcal{F}$ of elements for which x and y are in the same class, x and z in distinct classes is $\sum_{i \neq j} n_i(n_i-1)n_j/N(N-1)(N-2)$

$$|\{f \in \mathcal{F} \mid f(x)=f(y) \neq f(z)\}| = \sum_H |\{f \in \mathcal{F} \mid f(x)=f(y)=i,\ f(z) = j\}|$$

where H is $\{ (i,j) \mid i \neq j,\ n_i \geq 2\}$.

The set $\{f \in \mathcal{F} \mid f(x) = f(y) = i,\ f(z) = j,\ i \neq j$ and $n_i \geq 2\}$ can be mapped 1 - 1 onto the set of "labelled partitions of $E - (x,y,z)$ so that the cardinals of different classes subject to $n_i \geq 2$, are respectively

$$n_1, n_2, ---, n_{(i-1)}, n_i-2, n_{(i+1)}, ---, n_{(j-1)} n_j-1, n_{j+1}, ---, n_k.$$

It is now easy to express the current term of the upper sum and to establish the result.

2.4 Let x, y, z and t, be four objects of E, the proportion in $\mathcal{F}$ of elements for which x and y are in a same class, z and t in distinct classes is

$$[\sum_J n_i (n_i-1)n_j n_\ell + 2 \sum_k n_i (n_i-1) (n_i-2)n_j]/N(N-1)(N-2) (N-3)$$

where $J = \{(i,j,\ell) \mid i \neq j,\ j \neq \ell,\ \ell \neq i,\ n_i \geq 2\}$, $K = \{(i,j) \mid i \neq j,\ n_i \geq 3\}$

The first sum involves $k(k-1)(k-2)$ terms and the second one $k(k-1)$ terms.

An element f of $\mathcal{F}$ satisfies $f(x) = f(y)$ and $f(z) \neq f(t)$ if and only if f answers to one of the following alternatives.

1. $f(x) = f(y) = i$, $f(z) = j$ and $f(t) = \ell$ for an ordered triplet (i,j,t) such as $i \neq j$, $j \neq \ell$, $\ell \neq i$ and $n_i \geq 2$.

or

2.a $f(x) = f(y) = f(z) = i$ and $f(t) = j$

or

b $f(x) = f(y) = f(t) = i$ and $f(z) = j$

for a couple (i,j) such as $i \neq j$ and $n_i \geq 3$.
The cardinal of the subset of $\mathcal{F}$ whose elements satisfy 1 is the one of the set of labelled partitions of $E - (x,y,z,t)$, such as the cardinals of classes are respectively:

$$n_1, n_2, ---, n_{(i-1)}, n_i-2, n_{(i+1)}, ---, n_{(j-1)}, n_j-1, n_{j+1}, ---, n_{(\ell-1)}, n_\ell-1, n_{\ell+1}, ---, n_k.$$

The cardinal of the subset of $\mathcal{F}$ whose elements satisfy 2.a (resp. 2.b) is the same as the set of labelled partitions of $E - (x,y,z,t)$ such that the class cardinals are respectively

$$n_1, n_2, ---, n_{(i-1)}, n_i-3, n_{i+1}, ---, n_{(j-1)}, n_j-1, n_{j+1}, ---, n_k.$$

With these remarks we shall complete the calculation.

2.5 The different ratios, that we have just derived still hold if we consider them in Π instead of in $\mathcal{F}$.
To one element P of Π corresponds, exactly, $k_1!\ k_2!\ ---\ k_r!$ elements of $\mathcal{F}$ by labelling in $k_i!$ ways the k_i classes with the same cardinal ν_i, $i = 1, 2, ---, r$ [cf. IIA.1]. Let us suppose that we have calculated one of the previous proportions (1,2,3 or 4) in Π; we obtain the same proportion in $\mathcal{F}$ by multiplying the two terms of the ratio into $k_2! \times k_2! \times --- \times k_r!$. Because one element P of Π intervenes in the enumeration of a given term of the ratio if and only if the corresponding $k_1!\ k_2!\ ---\ k_r!$ elements in $\mathcal{F}$ intervene in the enumeration of this term of the ratio defined in $\mathcal{F}$. Henceforth we shall work in Π.

2.6 Important Remarks

1. For reasons of symmetry, the different proportions which had been defined, hold whatever the objects that form the pair or the couple of pairs.
2.a. Let G denote the set of couples of pairs such that the two pairs have one common component:

$G = \{((x,y),\ (x,z))$ where x,y and z are mutually distinct$\}$.

The cardinal of G is $N(N-1)\ (N-2)$; $N = |E|$

b. Let H denote the set of couples of pairs such that the two pairs have no common component.

$H - \{((x,y),\ (z,t))$ where x,y,z and t are mutually distinct$\}$.

The cardinal of H is $N(N-1)\ (N-2)\ (N-3)/4$.

II. B. Theorem. If ω is a total order on F, the mean in Π of $|\omega \cap S \times R| - \frac{1}{2}|S| \times |R|$ is null.

If P is an element of Π, there corresponds to it R and S for which we have:

$$\frac{1}{2}|S| \times |R| = \frac{1}{4}\left[\sum_{i=1}^{k} n_i(n_i-1)\right] \times \left[\sum_{i<j} n_i n_j\right] \qquad (1)\ [\text{cf.}\ §\text{IIA.1}]$$

We denote:

$\Phi[(p,q),\ S \times R]$ the indicator function of the set $S \times R$:

$\Phi[(p,q),\ S \times R] = 1$ if $(p,q) \in S \times R$ and 0 if not

$\Phi[(p,q),\ \omega]$ the indicator function of ω.

The mean in Π of $|\omega \cap S \times R|$ may be written as follows:

$$\frac{1}{|\Pi|} \sum_{P \in \Pi} \sum_{(p,q) \in F \times F} \Phi\ [(p,q),\ S \times R]\Phi[(p,q),\ \omega] \quad (2)$$

Let us decompose the sum over $F \times F$ into two sums: The former will be over G and the latter of H. [cf. the above remarks], namely:

$$\frac{1}{|\Pi|} \sum_{P \in \Pi} \sum_{(p,q) \in G} \Phi[(p,q),\ S \times R]\Phi[(p,q),\omega] \quad +$$

$$\frac{1}{|\Pi|} \sum_{P \in \Pi} \sum_{(p,q) \in H} \Phi[(p,q),\ S \times R]\Phi[(p,q),\omega] \qquad (3)$$

If we reverse the sum signs, the expression (3) may be written as

$$\sum_G \Phi[(p,q),\ \omega] \times \frac{1}{|\Pi|} \sum_\Pi \Phi[(p,q),\ S \times R] + \sum_H \Phi[(p,q),\ \omega] \times$$

$$\frac{1}{|\Pi|} \sum_\Pi \Phi[(p,q),\ S \times R] \qquad (4)$$

ω being a total order: $(p,q)\varepsilon\omega \leftrightarrows (q,p)\varepsilon\omega$.

On the other hand: $(p,q)\varepsilon G \rightleftarrows (q,p)\varepsilon G$ and $(p,q)\varepsilon H \leftrightarrows (q,p)\varepsilon H$.

By the lemmas (3) and (4) and according to the above remarks, the expression (4) becomes:

$$\frac{1}{2} N(N-1)(N-2) \sum n_i(n_i-1)n_j / N(N-1)(N-2)$$

$$+ \frac{1}{8} N(N-1)(N-2)(N-3) [\sum_J n_i(n_i-1)n_j n_\ell + 2 \sum_{i \neq j} n_i(n_i-1)(n_i-2)n_j] /$$

$$N(N-1)(N-2)(N-3) \qquad (5)$$

where $J = \{(i,j,\ell), i \neq j,\ j \neq \ell,\ \ell \neq i$ and $n_i \geq 2\}$.

The expression (5) becomes

$$\frac{1}{2} [\sum_{i \neq j} n_i(n_i-1)n_j + \frac{1}{4} (\sum_J n_i(n_i-1)n_j n_\ell + 2 \sum_{i \neq j} n_i(n_i-1)(n_i-2)n_j)] \qquad (6)$$

The last sum, from left to right, may be expressed as follows

$$\sum_{i \neq j} n_i^2(n_i-1)n_j - 2 \sum_{i \neq j} n_i(n_i-1)n_j \ .$$

Simplifying, the expression (6) we obtain:

$[\frac{1}{2}[\frac{1}{4}\sum_{i=1}^{k} n_i(n_i-1)(\sum_I n_j n_\ell + 2n_i \sum_{j\neq i} n_j]$ where $I = \{(j,\ell),\ j\neq i \text{ and } \ell\neq i\}$.

We have: $\sum_I n_j n_\ell + 2n_i \sum_{j\neq i} n_j = \sum_{r\neq s} n_r n_s$

So, the assertion is shown.

III. Type of partition and cardinal of the corresponding set R.

1) Introduction

Remembering that if $(n_1, n_2, ---, n_i, ---)$, where $n_1 \geq n_2 \geq --- \geq n_i \geq ---$, is the type of partition of E, for which there corresponds R and S; we have $|R| = \sum_i n_i(n_i-1)/2 = \frac{1}{2}(\sum_i n_i^2 - N)$ where $N = |E|$.

$|S| = \sum_{i<j} n_i n_j$ is the complement to $|F|$ of $|R|$; $|F| = N(N-1)/2$. It can be easily established that in the set of all partitions of E into two classes, $|R|$, or equivalently Σn_i^2, determines univocally the type of the partition. This elementary result leads us to deal with the following problem: If Σn_i^2 is given as equal to a positive integer M, then the set of partition-types such as $\Sigma n_i^2 = M$ is to be determined. We resolved this problem by a recursive technique. We shall not speak of this problem; we shall denote by $\Psi(N,M)$ the set of partition-types for which: $\Sigma n_i = N$ and $\Sigma n_i^2 = M$, N and M being given.

2) Proposition

The partitions which maximize $|R \times S|$ are those for which $|R|$ is the closest of $N(N-1)/4$. This result is evident. $|R \times S| = |R| \times |S|$ is the product of two cardinals whose sum is fixed; $|R| + |S| = N(N-1)/2$. If $\Psi(N,N(N+1)/2)$ is different from the empty set, there exists, at least one type of partition for which

$|R|$ is equal to $N(N-1)/4$. For example, if N is the square of an integer $(N + \sqrt{N})/2$, $(N - \sqrt{N})/2)$ is the type of a partition which realises $|R| = N(N-1)/4$; beforehand it should be noted that $(N + \sqrt{N})/2$ and $(N - \sqrt{N})/2$ are integers since N and $\sqrt{N}$ have the same parity.

IV On some aspects of comparison of the two criteria

The following propositions are corollaries of the main theorem of paragraph II. Let $\mathcal{R}$ denote the set of partitions for which $|R| = r$, where r is a given integer; this set is not empty if and only if $\Psi(N, 2r + N)$ is not empty.

1. Proposition 1 If ω is a total order on F, the mean in $\mathcal{R}$ of $|\omega \cap S \times R| - \frac{1}{2} |S| \times |R|$ is null.

The sets Π_t, where Π_t is the set of partitions of type t belonging to $\Psi(N, 2r + N)$, form a partition of $\mathcal{R}$:

$$\mathcal{R} = \sum_{t \in T} \Pi_t \text{ (sum of sets), where } T = \Psi(N, 2r + N)$$

The mean of $|\omega \cap S \times R| - \frac{1}{2} |S| \times |R|$ in each set Π_t is null [cf. §II.B]; then the mean of the same quantity in $\mathcal{R}$ is also null.

2. Proposition 2 The two criteria are equivalent in $\mathcal{R}$. This follows from the fact that the difference between the two criteria depends uniquely on N and r.

3. Proposition 3 If ω is a total order on F and the cardinal N of E such as $\Psi(N,N(N+1)/2)$ different from the empty set, then a partition for which

$$|R| < \frac{N(N-1)}{4} \left(1 - \frac{\sqrt{2}}{2}\right) \quad \text{or} \quad |R| > \frac{N(N-1)}{4} \left(1 + \frac{\sqrt{2}}{2}\right)$$

can not be optimal according to the first criterion.

Since $\Psi(N,N(N+1)/2)$ is not empty, there is at least one partition ω for which $|S| \times |R|$ equals $N^2(N-1)^2/2^4$. [cf. III.2]. If t is an arbitrary element of $\Psi(N,N(N+1)/2)$, by applying the theorem II.B. there exists in Π_t at least one partition P_t for which:

$$|\omega \cap S_t \times R_t| - \frac{1}{2}\,|S_t| \times |R_t| \geq 0$$

that is to say

$$|\omega \cap S_t \times R_t| \geq \frac{1}{2^5}N^2(N-1)^2 \qquad (1)$$

where S_t and R_t are referred to P_t.

P_m denoting an optimal partition for the first criterion (maximum of $|\omega \cap S \times R|$), S_m and R_m being related to P_m, we have:

$$|\omega \cap S_m \times R_m| > |\omega \cap S_t \times R_t| > \frac{1}{2^5}N^2(N-1)^2 \qquad (2)$$

But $|S_m \times R_m| \geq |\omega \cap S_m \times R_m|$, then necessarily

$$|S_m \times R_m| > \frac{1}{2^5}N^2(N-1)^2 \qquad (3)$$

In other words a partition for which

$$|S \times R| < \frac{1}{2^5}N^2(N-1)^2 \qquad (4)$$

can not be optimal for the first criterion.

We have: $|S \times R| = |S| \times |R| = [N(N-1)/2 - |R|] \times |R|$.

Putting $|R| = r$ and $N(N-1)/2 = s$, the inequality (4) becomes

$$r(s-r) < \frac{s^2}{2^3}$$

which holds if and only if

$$r < \frac{s}{2}\left(1 - \frac{\sqrt{2}}{2}\right) \quad \text{or} \quad r > \frac{s}{2}\left(1 + \frac{\sqrt{2}}{2}\right) \qquad (5)$$

The condition $\Psi(N,N(N+1)/2) \neq \Phi$ is not a restrictive one, if N is an arbitrary integer, a similar result can be obtained by replacing the right member of (1) by the maximum value of $\frac{1}{2}$ $|S| \times |R|$, in the set of all partitions of E.

4. Proposition 4 The integer N is assumed to satisfy the two following conditions: 1) $\Psi(N,N(N+1)/2)$
2) N is divisible by k: $N = n.k$.

If $k > 6$ or $k = 1$, a partition of E, into k classes with the same cardinal n, cannot be optimal for the first criterion.

If P is a partition of E into k classes with the same cardinal n, we have

$$|R| = kn(n-1)/2 \tag{1}$$

where R is associated to ω.

According to the above proposition, a partition for which $|R| < \frac{N(N-1)}{4}(1 - \frac{\sqrt{2}}{2})$, cannot be optimal for the first criterion.
Therefore if $kn\,(n-1)/2 < \frac{kn(kn-1)}{4}\,(1 - \frac{\sqrt{2}}{2})$ (2)

the partition P cannot be optimal for the first criterion.
The inequality (2) can be written

$$k > 4 + 2\sqrt{2} - \frac{5 + 2\sqrt{2}}{n} \tag{3}$$

(3) is necessarily satisfied for $k > 6$.

On the other hand, according to the above proposition a partition for which $|R| > \frac{N(N-1)}{4}\,(1 + \frac{\sqrt{2}}{2})$ cannot be optimal for the first criterion. Therefore if

$$\frac{kn(n-1)}{2} > \frac{kn(kn-1)}{4}\,(1 + \frac{\sqrt{2}}{2}) \tag{4}$$

the partition P cannot be optimal.

Simplifying the inequality (4) we obtain

$$k < 2(2 - \sqrt{2}) - \frac{3 - 2\sqrt{2}}{n}$$

This inequality is necessarily satisfied for $k = 1$ and $N > 1$.

We illustrate the last proposition with an elementary example: For $N = 49$, the set $\Psi(N, N(N+1)/2) = \Psi(49, 1225)$ is not empty; effectively N is a square of an integer, the type $(N + \sqrt{N})/2$, $(N - \sqrt{N})/2) = (28, 21)$ belongs to $\Psi(49, 1225)$. By the preceding proposition a classification of a set E, whose cardinal is 49, into 7 classes with the same cardinal 7 cannot be optimal for the first criterion.

5) Conclusion We have shown that the two criteria are equivalent in a set of partitions for which $|R|$ is constant. In a set of partitions of given type, the mean of the second criterion is independent of this type. This proposition shows that the second criterion has an intrinsic character. On the other hand, we have shown (proposition 3) and illustrated (proposition 4) an impossibility for some partitions to be optimal for the first criterion. In other words the first criterion is biased. For this reason the second criterion is preferable to the first one.

References

J.P. Benzecri 1) Problemes et methodes de la taxonomie, Rennes (1965b) et cours I.S.U.P., Paris 1965-1966. 2) Ordre lateral entre lois de probabilite sur un ensemble ordonne; applications a l'Analyse Factorielle et aux criteres de classification (I.S.U.P. Paris 1968).

P. Dagnelie, A propos des differentes methodes de classification numerique - *Revue de Statistique Appliquee* - (I.S.U.P.) 1966 - Vol. XIV no.3.

I.C. Lerman 1) Indice de Similarite et Preordonnance associee. Communication a Seminaix, Centre de Calcul - Maison des Sciences de l'Homme - Paris 1967. 2) Sur deux criteres de la classification. Centre de Calcul - M.S.H. Paris 1967. 3) Analyse de probleme de la recherche d'une hierarchie de classifications. Centre de Calcul - M.S.H. Paris 1968.

S. Regnier, Sur quelques aspects mathematiques des problemes de classification automatique. Centre de Calcul - M.S.H. Paris 1965 et revue I.C.C. 1965. Vol. 4 pp. 175-191.

R.N. Shepard, The analysis of proximities: scaling with an unknown distance function, I & II, *Psychometrika* 1962.

R.R. Sokal et P.M.A. Sneath, Principles of numerical taxonomy, W.H. Freeman and Cie, San Francisco and London; 1963.

W.F. de la Vega, Techniques de classification automatique utilisant un indice de ressemblance. Centre de Calcul- M.S.H. Paris 1965. p.48-54.

Discussion

Q. (Jackson) In minimising your criterion function do you examine each of the very many possible partitions or have you certain heuristic methods which you haven't mentioned?

A. I said, we obtain only local optima and not a general optimum. We can't examine all possibilities.

Q. (Jackson) In that case, how do you locate the subset in the locality of which you search for an optimum?

A. We obtain local optimum by an algorithm of transfer, which moves one object at a time until the local optimum is reached.

Comment: (Roux) The algorithm about which M. Lerman spoke begins by dividing the set to be classified into ten or more arbitrarily classes and then taking one element and trying to put it in the better class, *i.e.* that which maximises the function given by M. Lerman.

Q. (Cole) Do you use the concept of linear ordination only in the proof of your theorem or also in the practical algorithm? If you use it in the algorithm does not this destroy the concept of taxonomic distance?

A. Yes, but it's of practical interest. I think we should take the pre-order on F rather than a distance which in taxonomy is always arbitrary.

Q. (Saksena)

1. Could you illustrate with examples that your approach gives better classification than the existing methods of numerical taxonomy?
2. From what you have stated, it appears, that you even doubt the validity of the idea of taxonomic distance as a means of classification; am I right in concluding this?

A.

1. Yes, I have several and I can send you the details.
2. I think that distance is a topological notion. If we can do without using distance, so much the better. In any case I need some parameter to establish my pre-order. I can show that my pre-order is more stable.