BIOLOGY

Concentration of Entries in Binary Arrays

NUMERICAL methods for the classification of elements into sets commonly begin with an assortment of individuals defined by the possession or non-possession of a number of attributes, and seek to provide final sets of highly similar individuals[1,2]. The attribute-groups which form the basis of such a classification may also be of interest, in which case the system may be transposed and the attributes classified into groups by reference to the individuals[3]. Finally, attempts have been made[4] to extract 'noda', simultaneously defined by a final set of individuals and by a final set of attributes. Existing methods for the extraction of such noda sub-divide individuals and attributes independently and collate the results by reference to a 'two-way table'; but such methods do not provide the most efficient concentration of positive entries. It should in principle be possible to obtain each block substantially independently of the others, concentrating its entries by reference to such individuals and attributes as are involved in it; but, except in the trivial case of the 'black-and-white' table, it is not to be expected that the blocks will then dispose themselves conveniently within the cells of a conventional two-way table. Some may be expected to do so, but some will almost certainly transgress cell-boundaries and/or occupy only part of the fixed cells.

The situation can be improved if, at any stage, the incipient divisions in the two directions can be made comparable, so that division can be effected of either individuals or attributes, whichever gives the greater concentration. This would be possible with existing systems if a probabilistic parameter were in use, so that the two probabilities could be directly compared; but the computation would be extremely heavy, and the purpose of this communication is to suggest an altogether simpler strategy which will serve a similar purpose. Suppose a data-matrix to be divided into two groups of A and α individuals specified by two groups of B and β attributes; the table is now divided into four cells with possible maximum entries of AB, αB, $A\beta$ and $\alpha\beta$; let the actual positive entries be a, b, c and d respectively; and write $N=(A+\alpha)(B+\beta)$ and $n=(a+b+c+d)$. Although we now have the form of a 2×2 table, the usual null hypothesis based on marginal totals is inapplicable, since it may require a cell to contain more unities than there are spaces available (for example, that $a>AB$). We therefore state a new null hypothesis. The proportion of unities to total spaces is n/N: call this k. The hypothesis states that this proportion occurs in each of the four cells, that is, that the unities are disposed in proportion to the spaces available. For the first cell the expected number of unities is ABk, and of zeros $AB(1-k)$; the corresponding observed numbers are a and $(AB-a)$. It is easily shown that the contribution of this cell to a χ^2 is

$$\frac{(a-ABk)^2}{ABk(1-k)}$$

and the remaining three contributions can be written down by symmetry. Since A, B, α and β are given, only one parameter—n—is used in calculating the expectations, and the resulting χ^2 has three degrees of freedom. Mather[5] has shown how such a χ^2 can be partitioned into three additive components, each with one degree of freedom; using his methods we obtain the partition:

For the A/α division $$\frac{\{(a+c)\alpha-(b+d)A\}^2}{k(1-k)A\alpha N}$$

For the B/β division $$\frac{\{(a+b)\beta-(c+d)B\}^2}{k(1-k)B\beta N}$$

For the interaction $$\frac{(a\alpha\beta-bA\beta-c\alpha B+dAB)^2}{k(1-k)A\alpha B\beta N}$$

We note five properties of interest: (i) the maximum value of the total χ^2 is N, achieved by any division of the table into completely full and completely empty cells; (ii) empirical tests suggest that the direct comparison of the A/α and B/β divisions is more profitable than consideration of either the total χ^2 or the interaction; (iii) k must be calculated anew at each successive division, or else the χ^2 will have four degrees of freedom and the rationale of the partition will fail; (iv) χ^2 components for successive divisions are not monotonic, and do not define a stopping rule, which must be based on density and overall content of cells; (v) the system can be immediately generalized to arrays of more than two dimensions.

In any divisive system, since consideration of all dichotomous choices is normally impracticable, a restricted set of divisions must be designated. We have adopted the usual divisive strategy of considering only monothetic splits; but as a test we have enumerated all 961 divisions of a 6×6 table, and have been reassured to find that the highest contributions in both directions were from monothetic divisions. The method has been applied to a 15×27 set of ecological data of 20 per cent overall density, and produced five cells, which had a mean density of 81 per cent and accounted for 70 per cent of the positive entries; it is noteworthy that these cells cannot be contained within a conventional two-way table. This very satisfactory degree of concentration suggests to us that the method will find particular value in the analysis of tables which, by conventional methods, concentrate at an unacceptably low level. It is being tested on larger-scale ecological data, and the results will be submitted for publication to an ecological journal in due course.

We thank Mr. P. Macnaughton-Smith, of the Home Office Research Unit, for his advice.

J. THARU
W. T. WILLIAMS

Department of Botany,
University of Southampton.

[1] Macnaughton-Smith, P., *Home Office Res. Rep.*, 6 (1965).
[2] Williams, W. T., and Lance, G. N., *Nature*, 207, 159 (1965).
[3] Williams, W. T., and Lambert, J. M., *J. Ecol.*, 49, 717 (1961).
[4] Lambert, J. M., and Williams, W. T., *J. Ecol.*, 50, 775 (1962).
[5] Mather, K., *Statistical Analysis in Biology*, fifth ed. (Methuen, London, 1964).

The Nature of Pseudovacuoles in Cyanophyceae

THE nature of the pseudovacuoles, or gas vacuoles, which occur in bacteria and Cyanophyceae, is still enigmatic, although a number of reviews have been devoted to them.

The opinion which still prevails is that put forward by Fritsch[1]. After discussing the various observations and opinions with respect to Cyanophyceae he concluded that "the vacuoles develop at the bottom under anaerobic conditions (by a kind of gas-producing fermentation), and that as a result of the buoyancy thus acquired, the forms in question float to the surface". He doubted "whether material possessed of pseudovacuoles is actually in a healthy condition and whether cell-division continues".

Observations on a strain of *Oscillatoria agardhii*—a typical plankton—in axenic culture[2] (the first in which pseudovacuoles were isolated bacteria-free) showed that this view did not hold true for all cases. *O. agardhii* is strictly photo-autotrophic and damaged by organic nutrients—even by glucose at concentrations as low as 0·01 per cent. This in itself does not exclude the possibility of the gas being produced by a kind of fermentation, but at least it renders it impossible to prove.

It was possible to show that in this alga, pseudovacuoles are not simply gas-bubbles in a colloidal protoplasmic matrix but are instead cell-organelles of definite shape. They are elongated and multiply by fission. When centrifuged they decrease considerably in number, but after a few hours they gradually reappear in similar shape and arrangement as before.