THE AMERICAN MINERALOGIST, VOL. 50, OCTOBER, 1965

CLASSIFICATION IN A TERNARY DIAGRAM BY MEANS OF DISCRIMINANT FUNCTIONS

Felix Chayes, *Geophysical Laboratory, Carnegie Institution of Washington, Washington, D. C.*

ABSTRACT

For ternary closed data, discriminant functions based on polynomials of the form

$$\sum_{p=1}^{q} (X_i + X_j)^p$$

are analytically equivalent and geometrically identical for $1 \leq i \neq j \leq 3$. For any assigned value of q only a single function need be calculated and it makes no difference which pair of variables is used for the calculation.

Discriminant functions based on polynomials not satisfying this condition will vary with the choice of variables, so that if they are to be computed at all it will usually be necessary to compute them for at least two and sometimes for all three pairs of variables. If the objective is merely efficient classification, however, there will usually be little advantage in such calculations, since in general the efficiency of a discriminant based on a polynomial of the form

$$\sum_{p=1}^{q} (X_i + X_j)^p$$

will not be less than that of a polynomial of the same order but different form in these variables.

Generalization of the basic relation is immediate; for an M-variable closed array discriminant functions based on polynomials of the form

$$\sum_{p=1}^{q} (X_1 + X_2 + \cdots + X_{M-1})^p$$

will be analytically equivalent, the numbering of the variables being arbitrary. No other polynomial of the same order in these variables will yield a more efficient discriminant.

If $M=3$, functions based on all three variables cannot be plotted in the ternary diagram, those based on only one variable plot as straight lines parallel to one of the edges, and those based on any pair of variables plot as straight lines if $q=1$ and curves if $q>1$. Any discriminant function based on one or two variables divides the ternary diagram into two fields, corresponding to the two groups into which the data are classified by the discriminant. An example is described.

Discriminant function analysis has enjoyed considerable vogue in the life and behavioral sciences since shortly after the close of the second war, and is now finding application in the earth sciences as well. Like most forms of multivariate analysis, the technique is not likely to be of much practical use to the naturalist who does not have access to a reasonably capacious electronic computer. To the rapidly increasing number of naturalists who do have access to such equipment, however, it offers two major advantages, at the same time providing a perfectly objective procedure for determining how effectively objects belonging to two different groups can be distinguished from each other by means of a given set of properties measurable in each object, and making possible the simultaneous consideration of large numbers of variables.

For petrographers, traditionally committed to graphical analysis and acutely aware of its limitation in this respect, the ability to treat large numbers of variables simultaneously will no doubt have considerable appeal. Under some circumstances, however, the ability to discriminate in an objective fashion, and more particularly the ability to estimate the

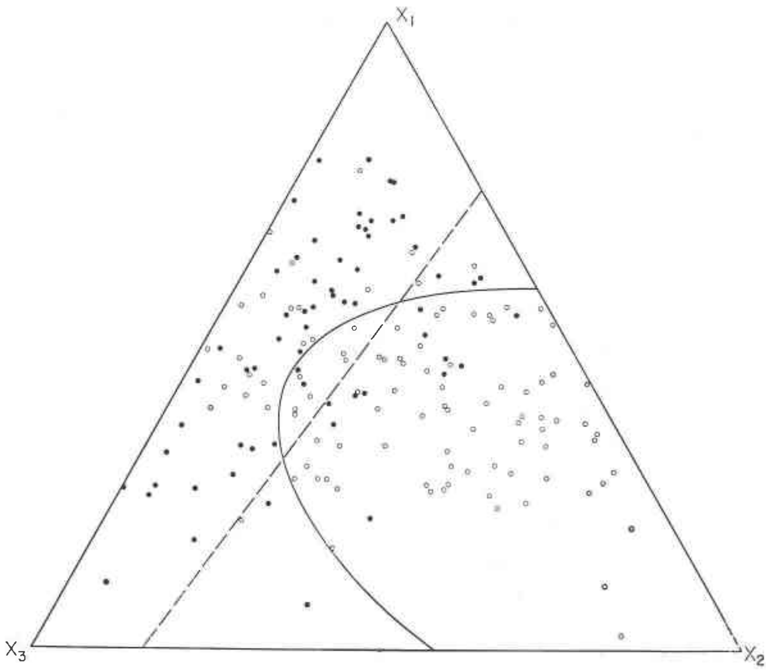

FIG. 1. Items in group A of Table 1 shown by solid circles, those of group B by open circles. The trace of equation (4) is shown by the dashed line, that of equation (6) by the curve.

optimum efficiency with which the discrimination can be made, may be of greater importance. This note presents an example.

In Table 1 are listed recorded values of variables X_1, X_2 and X_3 in each of 63 items of group A and 97 of group B. Each triplet sums to 100, and all are plotted in Fig. 1, those belonging to group A being shown by solid circles. The problem for which a solution is sought is the simple and fundamental one which characterizes all descriptive science, viz., whether, how, and how efficiently, individual items can be assigned to the proper type on the basis of their announced contents of variables $X_1, X_2, \cdots X_m$. In the ternary array a discriminant function computed from any pair of the variables provides a convenient and usually an optimum solution of this problem.

Other solutions or, at any rate, other modes of examining the problem, are of course possible. Intuitively, one suspects that if the groups really are distinct it ought to be possible to divide the diagram into two fields—whether by inspection or by contouring—one occupied almost exclusively by open, the other almost exclusively by solid, circles. A close look at the

GROUP A

X	1	2	3	1	2	3
1	53.21	9.58	37.21	71.35	1.49	27.16
3	21.00	37.29	41.71	40.95	26.68	32.37
5	44.37	15.53	40.10	59.65	33.54	6.81
7	62.09	12.58	25.33	69.14	17.78	13.08
9	62.47	6.09	31.44	54.78	12.33	32.89
11	54.35	27.71	17.94	47.45	14.19	38.36
13	55.09	17.99	26.92	67.48	12.54	19.98
15	59.66	27.47	12.87	68.70	16.80	14.50
17	65.98	14.40	19.62	69.71	11.45	18.84
19	65.09	7.15	27.76	55.40	16.32	28.28
21	57.15	13.86	28.99	56.41	14.26	29.33
23	49.64	10.16	40.20	78.02	8.72	13.26
25	77.76	1.85	20.39	42.72	2.48	54.80
27	25.78	10.53	63.69	24.61	4.74	70.65
29	74.32	13.83	11.85	68.29	13.85	17.86
31	58.81	10.41	30.78	58.59	32.95	8.46
33	32.04	15.20	52.76	27.91	14.81	57.28
35	23.19	22.01	54.80	60.59	15.68	23.73
37	60.25	4.61	35.14	35.66	3.82	60.52
39	7.08	35.27	57.65	17.21	14.26	68.53
41	35.96	24.80	39.24	74.63	13.52	11.85
43	10.56	5.44	84.00	64.24	21.99	13.77
45	51.26	13.13	35.61	53.48	41.55	4.97
47	53.95	11.62	34.43	39.25	22.46	38.29
49	42.24	17.58	40.18	25.56	0.17	74.27
51	44.80	9.44	45.76	44.40	8.33	47.27
53	31.10	3.81	65.09	25.98	4.90	69.12
55	32.72	13.50	53.78	47.79	3.00	49.21
57	32.93	18.01	49.06	40.43	25.61	33.96
59	67.05	13.66	19.29	45.26	37.97	16.77
61	46.42	35.24	18.34	44.04	36.07	19.89
63	50.22	30.16	19.62			

TABLE 1. TERNARY COORDINATES, IN PERCENTAGES, FOR 63 ITEMS BELONGING TO GROUP A AND 97 TO GROUP B

(Variables identified by column headings; row headings give number of first item in each line.)

GROUP B

X	1	2	3	1	2	3
1	38.32	6.32	55.36	25.43	43.58	30.99
3	25.93	45.32	28.75	16.21	34.33	49.46
5	27.11	28.04	44.85	47.28	36.47	16.25
7	25.93	30.17	43.90	37.28	50.27	12.45
9	41.22	49.03	9.75	53.37	37.91	8.72
11	27.32	23.71	48.97	27.15	26.77	46.08
13	29.27	24.18	46.55	38.47	14.10	47.43
15	22.82	64.18	13.00	52.75	38.55	8.70
17	48.32	30.56	21.12	25.14	65.93	8.93
19	36.28	60.33	3.39	51.29	25.88	22.83
21	43.73	16.16	40.11	54.94	2.51	42.55
23	51.28	20.02	28.70	46.11	21.27	32.62
25	26.38	42.63	30.99	32.79	56.04	11.17
27	28.62	47.09	24.29	42.03	37.01	20.96
29	37.10	53.11	9.79	46.75	25.79	27.46
31	32.42	27.31	40.27	54.42	44.27	1.31
33	45.81	29.76	24.43	42.88	50.29	6.83
35	20.71	19.73	59.56	28.11	53.36	18.53
37	46.18	26.73	27.09	39.66	44.36	15.98
39	27.52	58.75	13.73	33.33	23.81	42.86
41	54.87	39.37	5.76	46.37	28.93	24.70
43	29.29	44.47	26.24	41.84	6.65	51.51
45	61.23	31.75	7.02	54.55	10.42	35.03
47	47.63	1.16	51.21	66.62	0.29	33.09
49	63.67	21.88	14.45	61.58	6.05	32.37
51	36.82	55.54	7.64	31.76	42.91	25.33
53	10.31	75.33	14.36	33.27	52.09	14.64
55	35.27	64.07	0.66	49.32	15.01	35.67
57	43.80	9.03	47.17	32.81	33.64	33.55
59	53.51	30.33	16.16	76.64	8.09	15.27
61	66.31	4.62	29.07	58.53	25.10	16.37
63	48.87	13.99	37.14	53.61	35.41	10.98
65	54.18	9.55	36.27	44.55	33.64	21.81
67	23.60	57.92	18.48	24.86	52.04	23.10
69	19.51	74.61	5.88	29.56	50.42	20.02
71	26.32	45.60	28.08	43.96	51.26	4.78
73	41.96	45.08	12.96	40.28	11.07	48.65
75	57.61	18.77	23.62	51.96	47.37	0.67
77	45.72	36.10	18.18	38.63	29.84	31.53
79	2.60	81.67	15.73	35.52	44.61	19.87
81	37.36	18.66	43.98	46.09	6.37	47.54
83	42.71	11.92	45.37	41.56	30.57	27.87
85	33.62	62.37	4.01	38.23	19.10	42.67
87	39.04	37.83	23.13	47.22	20.29	32.49
89	38.47	39.60	21.93	34.49	62.24	3.27
91	40.26	19.07	40.67	54.37	30.74	14.89
93	28.58	67.59	3.83	40.97	25.59	33.44
95	36.09	59.87	4.04	26.92	66.97	6.11
97	42.29	57.21	0.50			

diagram should persuade almost anyone that this is not possible in the present example. There is nevertheless a perceptible concentration of solid circles in the upper left of the triangle and a somewhat weaker concentration of open circles a little below and to the right of its center. In fact, in nearly two-thirds of the specimens located by solid circles $X_1 > 2X_2$ and in nearly three-quarters of those shown by open circles $X_1 < 2X_2$. Those in the first category lie above, and those in the second below, the line $X_1 = 2X_2$ (not shown in the figure). Against the alternative that there is no tendency for points to concentrate in any particular region of the triangle the relative frequency of solid circles above this line indicates a highly significant departure from randomness, but that of the open circles below it does not. Against the superficially similar alternative that points are as likely to fall on one side of the line as on the other, however, the data give a rather stronger result; the distributing of *open* circles provides very strong reason for rejecting this second hypothesis and that of the solid circles is also significant against it at the 1 per cent level, but barely so.

In the absence of substantive information it is of course impossible to determine which—if either—of these descriptive hypotheses is realistic. They are discussed here merely as a reminder that partition tendencies easily strong enough to be highly significant may nevertheless be so weak as to be worthless for purposes of classification or description. If, for instance, only 31 instead of 37 of the 63 solid circles lay above the line $X_1 = 2X_2$, this would still be sufficient to establish, at the 1 per cent level, the hypothesis that there is a tendency for items of group A to be concentrated in the upper part of the triangle. Yet if the condition $X_1 > 2X_2$ were then adopted as a criterion for inclusion of an item in group A, more than half the sample items known on other grounds to be of this type would be excluded from it.[1]

When the potential utility of some particular partition as a descriptive or taxonomic device is at issue, tests of the sort so far described are essentially irrelevant. What is needed is an estimate of the proportion of the *total sample variation* which is properly allocated or "accounted for" by the partition in question. Assurance that a particular partition is the best of its kind would also be extremely valuable; for example, the line $X_1 = 2X_2$, which from inspection seems a fairly good discriminant, is a

[1] In analogous fashion, a properly constructed response surface may well reveal statistically significant tendencies of great interest and importance, yet provide an uninformative or misleading description of the "topography" of both sample and parent frequency distributions. This is almost inevitable when, as in many published maps, the computed surface accounts for only a small part of the total sample variation (see Chayes and Suzuki, 1964).

member of the family of lines $\lambda_1 X_1 + \lambda_2 X_2 + \beta = 0$, with λ_1 taken as 1, λ_2 as -2, and β as 0. To determine by inspection whether this represents the best choice of coefficients would be an endless chore, yet clearly if a better partition could be obtained by some other set of coefficients, that set would be preferable for the purpose at hand. It is thus a considerable advantage that the coefficients of the computed discriminant function are, in a simple and intuitively appealing sense, the best obtainable, and that the efficiency with which the function partitions the sample—*i.e.* the proportion of correct assignments it yields—is easily found. (When, furthermore, the sampling procedure is suitably random, the efficiency of the partition effected by the function is a sound estimate of the efficiency with which it may be expected to partition future samples.)

A full explication of the structure and calculation of the two group discriminant function would be inappropriate,[1] but it will be advantageous to review here as much of the argument as is required to explain the sense in which its coefficients are indeed the best obtainable. From the paired variables in each item we may form a new variable, viz. $z = \lambda_1 X_1 + \lambda_2 X_2$, from these new variables, in turn, an average for each type, $\bar{z}_a$, $\bar{z}_b$, and, finally, the ratio

$$\frac{(\bar{z}_a - \bar{z}_b)^2}{n_a \sigma_a^2 + n_b \sigma_b^2} \tag{1}$$

in which n_a denotes the number and σ_a the standard deviation of the z's of group A. The total variation of z in the array, expressed as a sum of squares of deviations, is the sum of the numerator and denominator of this ratio. The denominator is the "within-group sum of squares" and the numerator is the "between-group sum of squares," or mean-square for difference between groups. The ratio is spoken of as the "distance" between the groups, and the feature which distinguishes the discriminant function from all other linear combinations of the same variables is simply that the λ's which appear in it are those which maximize this distance in the sample. It is in this sense that the coefficients of the function are the best obtainable, and it is to be noted that the property is nearly distribution free—requiring only that the variances of z exist and that there be some tendency toward central concentration in each group—and independent of sampling technique.[2]

[1] A detailed petrographic example will be found in Chayes and Velde (1965), but the reader interested in applying the technique should certainly consult a standard text, *e.g.* Hoel (1962).

[2] Of course, if a discriminant function calculated from one sample is to be used on other samples, sampling technique cannot be ignored. With properly randomized sampling, however, the optimal character of the λ's persists whether the function is used descriptively or predictively.

Selection of the constant β, usually referred to as *the* discriminant and denoted below by $\hat{z}$, implies knowledge or assumption about the parent distributions of z in the two groups. If z is symmetrically distributed about group means with the same variance in both groups, $\hat{z}$ is taken at the midpoint between the group averages, viz.,

$$\hat{z} = (\bar{z}_a + \bar{z}_b)/2 \tag{2}$$

and the probability of misclassification is the same for members of each group. If the variances differ, the same result is obtained by a discriminant located at

$$\hat{z} = (\sigma_b\bar{z}_a + \sigma_a\bar{z}_b)/(\sigma_a + \sigma_b) \tag{3}$$

provided the distributions are "sufficiently normal."

The simplest discriminant, that based on only one variable, is merely a point located between the group means for this variable. For homogeneous sub-group variances it is placed at the mid-point between the means, as in equation (2). If variances differ in the two groups, it is found from equation (3) and will always be closer to the mean for the group in

TABLE 2. PERFORMANCE OF DISCRIMINANTS BASED ON DIFFERENT COMBINATIONS OF VARIABLES

Variables	Misclassifications of items in group		Efficiency, %
	A	B	
X_1	21	36	64.4
X_2	14	27	74.4
X_3	28	35	60.6
(X_i, X_j), $1 \leq i \neq j \leq 3$	16	26	73.7
$(X_i, X_j, X_i^2, X_j^2, X_iX_j)$, $1 \leq i \neq j \leq 3$	13	22	78.1

which the variance is smaller. (All $\hat{z}$ values in this note are those given by equation (3).) In the ternary diagram a discriminant function based on any variable, say X_i, plots as a line parallel to the X_j–X_k edge. For the data of Table 1, the behavior of these one-variable discriminants is summarized in the first three lines of Table 2. Variable X_2 provides a more effective partition than X_1 or X_3, but an efficiency of 74.4 per cent leaves something to be desired, and it is natural to wonder whether a line *not* parallel to an edge of the triangle might not do a better job. The answer to this question is given by the efficiency with which a binary linear discriminant function partitions the data.

For the data of Table 1 the binary linear discriminant functions may be written

$$X_2 - 0.155X_1 - 15.665 = 0 \tag{4a}$$
$$X_2 + 0.134X_3 - 26.942 = 0 \tag{4b}$$
$$X_1 + 0.866X_3 - 72.972 = 0 \tag{4c}$$

in which each line has been "normalized" by the coefficient of its leading term. The trace of these functions in the ternary is shown by the dashed line in Fig. 1. The constant term in each line is the ternary equivalent of an intercept; when $X_1 = 0$, for instance, $X_2 = 15.7$ on the X_2–X_3 edge of the triangle, and when $X_3 = 0$, the trace of the function intersects the X_1–X_2 edge at $X_2 = 26.9$, $X_1 = 73.0$. In similar fashion, the coefficient of the second term in each line is a slope.

Although normalization greatly facilitates geometrical interpretation in the ternary diagram it actually obscures the analytical relation between the functions which leads to their rather surprising geometrical identity. Prior to normalization (and transfer of $\hat{z}$ to the left side of each equation), the set of coefficients and constants is as shown in Table 3.[1] Using column headings to denote entries in any line, and denoting the coefficients and constant of any other line by κ_i, κ_k and $\hat{z}_k$, it becomes apparent after a little study of the table that, with differences small enough to be attributed to rounding,

TABLE 3. COEFFICIENTS (λ) AND CONSTANT ($\hat{z}_l$) OF BINARY LINEAR DISCRIMINANT FUNCTIONS COMPUTED FROM DATA OF TABLE 1

Variables	λ_1	λ_2	λ_3	$\hat{z}_l$
X_1, X_2	67	−433		− 67.83
X_1, X_3	500		433	364.86
X_2, X_3		−500	−67	−134.71

$$\kappa_i = \lambda_i - \lambda_j$$
$$\kappa_k = -\lambda_j$$
$$\hat{z}_k = \hat{z}_l - \lambda_j$$

Had the problem been solved by foresight rather than hindsight it would have been appropriate to remark here that this is precisely what would have been expected from the fact that $X_i + X_j + X_k = 1$. For if λ_i, λ_j, κ_i, and κ_j are indeed the coefficients of the respective discriminant functions, we must have (see Hoel, 1962, p. 179–184)

$$\left.\begin{aligned} \lambda_i S_{ii} + \lambda_j S_{ij} &= d_i \\ \lambda_i S_{ij} + \lambda_j S_{jj} &= d_j \end{aligned}\right\} \tag{5a}$$

[1] Tables 3 and 4 and the relevant discussion presume data in proportions rather than the percentages explicit in Table 1 and implicit in equations (4) and (6).

and

$$\left.\begin{aligned}\kappa_i S_{ii} + \kappa_k S_{ik} &= d_i \\ \kappa_i S_{ik} + \kappa_k S_{kk} &= d_k\end{aligned}\right\} \tag{5b}$$

where

$$d_i = \bar{x}_{ai} - \bar{x}_{bi}$$
$$S_{ii} = n_a\sigma_{ai}^2 + n_b\sigma_{bi}^2$$
$$S_{ij} = n_a\sigma_{ai}\sigma_{aj}\rho_{aij} + n_b\sigma_{bi}\sigma_{bj}\rho_{bij}$$

and similarly for S_{ik}, S_{jj} and S_{kk}.

Because of closure,

$$\sum_{q=i}^{k} S_{pq} = 0$$

whether $p=i, j$, or k, and it is evident that, if $X_i+X_j+X_k=1$, $d_i+d_j+d_k=0$. By solving (5a) for λ_j, (5b) for κ_k, and taking advantage of these relations depending on closure, it may be shown that

$$\kappa_k = \frac{d_iS_{ik} - d_kS_{ii}}{S_{ik}^2 - S_{ii}S_{kk}} = \frac{d_jS_{ii} - d_iS_{ij}}{S_{ij}^2 - S_{ii}S_{jj}} = -\lambda_j$$

and that

$$\kappa_i = \frac{d_i - \lambda_jS_{ij} - \lambda_jS_{ii}}{S_{ii}} = \lambda_i - \lambda_j$$

Finally, for any individual item in either group,

$$z_k = (\lambda_i - \lambda X_j)_i - \lambda_j X_k = z_l - \lambda_j \tag{5c}$$

hence the group averages are related by $\bar{z}_k = \bar{z}_l - \lambda_j$ and the discriminants by $\hat{z}_k = \hat{z}_l - \lambda_j$. (Readers familiar with elementary matrix algebra will find a more compact derivation of these results in the second section of the appendix.)

There is thus only one binary linear discriminant function in a ternary closed array, and which pair of variables happens to be used for its calculation is a matter of no consequence. The relation generalizes immediately; in an M-variable closed array there is only one $(M-1)$th linear discriminant function and this function may be calculated from any of the $(M-1)$ sets of $(M-1)$ variables. Although it is certainly meaningful to ask which of the three individual variables in a ternary closed array provides the best linear discriminant, it is not meaningful to ask which of the three possible pairs of variables does so.

Graphically, all points lying to the left of the dashed line in Fig. 1 are placed in group A, all to the right of it in group B. Analytically, using equation (4a) for illustration, the binary linear discriminant function classifies an item in group A or B depending on whether the amounts of X_1 and X_2 observed in it are such that the quantity $X_2 - 0.1547\ X_1$ is or is not less than 15.655, this number being simply the $\hat{z}$ of equation (3). The efficiency with which the binary linear discriminant reclaims the original dichotomy, shown in line 4 of table 2, is 73.7 per cent. This is no improvement over the 74.4 per cent found for X_2 alone, from which it appears that nothing has been gained by introducing a second variable into the function. (It is to be remembered, however, that although three calculations are required to determine which of the individual variables gives the best partition, it will never be necessary to compute more than one binary linear discriminant.)

Following the procedure outlined in most texts, this is about as far as the analysis can be taken; one may conclude that no linear combination of these variables is more effective than X_2 alone, and that the partition effected by X_2 is rather mediocre. The considerable number of open circles lying just above the central portion of the trace of equation 4 in Fig. 1 suggests, however, that a non-linear partition might be better than the best linear one.

Although I know of no examples of non-linear discriminant functions, the requirement for linearity in the original derivation seems to be purely formal. In exact analogy with the relation between multiple and curvilinear regression, as many variables of the type $X_i \equiv X_1^p X_2^q$ may be added as seem desirable. Each λ_i in the computed function will then be the coefficient of a term of order $(p+q)$ in (X_1, X_2). The discriminant function may thus be made into a polynomial of any required order and degeneracy without change in its basic properties or their interpretation. In principle, at least, no polynomial in X_1 and X_2 will partition the data more efficiently than a discriminant function of the same order.

Even confining the discussion to linear and quadratic terms, however, there are 31 possible functions for any pair of variables, and their relations both with each other and with functions based on either of the other pairs may be extremely complex. A choice based on exhaustive trial, though perhaps not impractical, is certainly inelegant; it is, in fact, almost as unsatisfying as the conventional graphical guess-work for which the discriminant function is here proposed as a substitute.

Fortunately, the difficulty may prove much less serious than it seems. Indeed, if one is willing to move in large steps rather than small ones it practically vanishes. If we add a complete quadratic instead of a single higher order term to the binary linear in (X_i, X_j), the new discriminant function will be analytically equivalent to and geometrically identical with one based on the complete suite of linear and quadratic coefficients in either (X_i, X_k) or (X_j, X_k). Using the same denotations of λ, κ and $\hat{z}$ as before, the conversion relations between the coefficients in (X_i, X_j) and (X_i, X_k) are then

$$\kappa_i = (\lambda_i + \lambda_{ij}) - (\lambda_j + 2\lambda_{jj}) \qquad \kappa_k = -(\lambda_j + 2\lambda_{jj})$$
$$\kappa_{ii} = \lambda_{ii} + \lambda_{jj} - \lambda_{ij} \qquad \kappa_{kk} = \lambda_{jj}$$
$$\kappa_{ik} = 2\lambda_{jj} - \lambda_{ij} \qquad \hat{z}_k = \hat{z}_l - (\lambda_j + \lambda_{jj})$$

where λ_{ij} denotes the coefficient of a cross-product, λ_{jj} that of a term in X_j^2, and similarly for κ_{ik}, κ_{ii}. Coefficients computed directly from the data are shown in Table 4 and a derivation of the conversion relations is given in an appendix.

Thus if one passes directly from the binary linear discriminant to the binary linear+quadratic one, only a single function need be computed at each stage. In principle, and nearly always in practice as well, a function based on X_i, X_j, (X_iX_j), X_i^2 and X_j^2 will not be less efficient than one based on any subset of these terms. (No exceptions to this rule occur in the 33 functions calculated in the course of the work reported here.) The partition obtained by means of the full linear+quadratic will be the same whatever pair of variables is used, and the resulting functions will be geometrically identical in the conventional trilinear coordinates.

TABLE 4. COEFFICIENTS (λ) AND CONSTANTS ($\hat{z}_l$) OF BINARY LINEAR+QUADRATIC DISCRIMINANT FUNCTIONS COMPUTED FROM DATA OF TABLE 1

Variables	λ_1	λ_2	λ_3	λ_{11}	λ_{22}	λ_{33}	λ_{12}	λ_{13}	λ_{23}	$\hat{z}_l$
X_1, X_2	−2154.969	−1413.323		2164.4	561.8		1644.3			−617.334
X_1, X_3	− 221.573		288.323	1081.8		561.8		−520.1		233.930
X_2, X_3		−1943.350	−2175.085		1081.8	2164.4			2684.4	−627.396

Generalization to polynomials of higher order, though possibly not of much practical use, is evident. If each term of one polynomial can be expressed as a linear combination of terms in a second polynomial of the same form and order, discriminant functions based on the two polynomials will be equivalent. In a ternary closed array this relation will always hold for polynomials of the form

$$\sum_{p=1}^{q} (X_i + X_j)^p \qquad \text{for } 1 \leq i \neq j \leq 3;$$

the three possible polynomials of this form and order q will yield discriminant functions which are analytically equivalent and geometrically identical. An analogous relation holds for closed arrays containing more than 3 variables, but with increase in the number of variables the number of terms to be calculated soon passes realistic limits. For $q=2$, for instance, there are only 5 terms if $M=3$ but 20 if $M=6$.

The curved line in Fig. 1 is the trace of the full linear+quadratic discriminant function

$$(X_2 + 1.119X_3) - (0.006X_2^2 + 0.014X_2X_3 + 0.011X_3^2) = 32.264 \quad (6)$$

performance data for which are shown in the last line of Table 2. Thirteen of the solid circles fall below and to the right of the curve, twenty-two of the open ones lie above and to the left of it. The overall efficiency of this discriminant is thus 78.1 per cent, a small but appreciable improvement over that obtained from X_2 alone.

With knowledge of X_2 only, the ratio of correct to incorrect classifications is 2.91:1. With the full linear+quadratic discriminant function based on any pair of variables it rises to 3.57:1, and it may be confidently asserted that no simpler polynomial will yield more efficient partition. An efficiency of 78 per cent is better than nothing, but whether it is good enough to be useful depends entirely on subject matter considerations. Space does not permit a full description of the data, the variables, or the substantive problem, but the reader is entitled to know that much more than an exercise in arithmetic is involved.

The raw data are chemical analyses of Cenozoic basic volcanics in which $H_2O<2$ per cent and $(Fe_2O_3/FeO)<0.6$, and whose norms lack both *ne* and *Q*. Group A contains analyses of specimens associated with basic volcanics the great majority of which are *ne*-normative, group B those associated with basic volcanics the great majority of which are *Q*-normative. The problem is to determine how effectively analyses whose norms lack both *ne* and *Q* can be assigned to the proper parent group by means of their positions in the ternary projection $di'+hy'+ol'=100$. In Table 1 and Fig. 1 di' is denoted by X_1, hy' by X_2 and ol' by X_3. A detailed description of the study and of conclusions drawn from it will be presented in a separate communication. The present note is intended only to show that the discriminant function may be conceptually useful even when, as in a ternary diagram, the number of variables is not large.

Acknowledgment

I am deeply indebted to J. M. Cameron of the National Bureau of Standards for much helpful discussion, and in particular for assistance with the appendix.

References

Chayes, F. and Y. Suzuki (1963) Geological contours and trend surfaces. *Jour. Petrology*, **4**, 307–312.

——— and D. Velde, (1965) On distinguishing basaltic lavas of circumoceanic and oceanic-island type by means of discriminant functions. *Am. Jour. Sci.*, **263**, 206–222.

Hoel, P. G. (1962) *Introduction to Mathematical Statistics.* 3rd ed., John Wiley and Sons, Inc. New York.

APPENDIX. $(M-1)$-VARIABLE CUMULATIVE POLYNOMIAL DISCRIMINANT FUNCTIONS COMPUTED FROM M-VARIABLE CLOSED ARRAYS

GENERAL RELATIONS

There are n_1 items in group 1 and n_2 items in group 2. Each item is characterized by two row vectors,

$$\begin{aligned}\mathbf{u}_a' &= \{u_1, u_2, \cdots, u_W\} \qquad a = 1, 2, \cdots, (n_1+n_2)\\ \mathbf{v}_a' &= \{v_1, v_2, \cdots, v_W\}\end{aligned} \tag{1a}$$

in which each element is measured from its own group mean. The matrices $\mathbf{u}$, $\mathbf{v}$ are the assemblages of vectors $\mathbf{u}_a'$, $\mathbf{v}_a'$, respectively, and the elements of the vectors of mean differences, $\mathbf{d}_u$, $\mathbf{d}_v$, are

$$[d_{uj}] = (\bar{u}_{1j} - \bar{u}_{2j}); \qquad [d_{vj}] = (\bar{v}_{1j} - \bar{v}_{2j}) \qquad j = 1, 2, \cdots, w \tag{2a}$$

in which subscripts 1 and 2 refer to groups. By virtue of these definitions the discriminant functions may be written

$$\mathbf{u}'\mathbf{u}\boldsymbol{\lambda} = \mathbf{d}_u; \qquad \mathbf{v}'\mathbf{v}\boldsymbol{\kappa} = \mathbf{d}_v \tag{3a}$$

In situations of interest in this note the raw values are successive terms of

$$\sum_{p=1}^{q} \left(\sum_{i=1}^{M-1} (\mathrm{X}_{ai}) \right)^p \tag{4a}$$

in which X_{ai} is the observed value of variable i in the ath item, and the X_i's in each item are subject to the closure constraint, viz.

$$\sum_{i=1}^{M} (\mathrm{X}_{ai}) = \sum_{i=1}^{M} (\bar{x}_i) = 1; \qquad \sum_{i=1}^{M} (x_{ai}) = 0 \tag{5a}$$

The numbering of variables in (4a) is arbitrary, with $(M-2)$ variables common to any pair of matrices $\mathbf{u}$, $\mathbf{v}$, and one variable unique to each. For $M=3$, $q=1$, for instance, we might have, for the ath row of the matrices,

$$\mathbf{u}_a' = \{x_{ai}, x_{aj}\}; \qquad \mathbf{v}_a' = \{x_{ai}, x_{ak}\}$$

The purpose of this appendix is to demonstrate the effect of (4a) and (5a) on the relations between equations (3a), and between the discriminants calculated from them. We show first that each κ_j (or λ_j) may be expressed as a linear combination of the λ's (or κ's).

By using (4a) and (5a), as shown below by examples, one may readily form a right multiplier, $\mathbf{T}=\mathbf{T}^{-1}$, such that

$$\mathbf{u}\mathbf{T} = \mathbf{v}; \qquad \mathbf{v}\mathbf{T} = \mathbf{u} \tag{6a}$$

and it then follows from the definitions of $\mathbf{d}_u$, $\mathbf{d}_v$ that

$$\mathbf{T}'\mathbf{d}_u = \mathbf{d}_v; \qquad \mathbf{T}'\mathbf{d}_v = \mathbf{d}_u \tag{7a}$$

Solving (3a) for κ and taking advantage of these transformations;

$$\begin{aligned}\boldsymbol{\kappa} &= (\mathbf{v}'\mathbf{v})^{-1}\mathbf{d}_v \\ &= (\mathbf{T}'\mathbf{u}'\mathbf{u}\mathbf{T})^{-1}\mathbf{T}'\mathbf{d}_u \\ &= \mathbf{T}^{-1}(\mathbf{u}'\mathbf{u})^{-1}\mathbf{T}'^{-1}\mathbf{T}'\mathbf{d}_u \\ &= \mathbf{T}^{-1}(\mathbf{u}'\mathbf{u})^{-1}\mathbf{d}_u \\ &= \mathbf{T}^{-1}\boldsymbol{\lambda} \end{aligned} \tag{8a}$$

and, by an exactly analogous development,

$$\boldsymbol{\lambda} = \mathbf{T}^{-1}\boldsymbol{\kappa} \tag{8b}$$

Thus, if ${}^j\mathbf{t}$ is the jth row of $\mathbf{T}^{-1}$, the coefficients of the jth terms in the discriminant functions are

$$\kappa_j = {}^j\mathbf{t}\boldsymbol{\lambda}; \qquad \lambda_j = {}^j\mathbf{t}\boldsymbol{\kappa} \qquad\qquad 1 \leq j \leq (M-1) \tag{9a}$$

which was to be shown.

The resulting relationship between discriminants calculated from equations (3a) remains to be discussed. Corresponding to (1a) we have the row vectors

$$\begin{aligned}\mathbf{U}_a' &= \{U_1, U_2, \cdots, U_W\} \\ \mathbf{V}_a' &= \{V_1, V_2, \cdots, V_W\}\end{aligned} \qquad a = 1, 2, \cdots, (n_1 + n_2) \tag{10a}$$

in which each element is measured about zero. From these vectors the statistics used to classify the ath item, namely,

$$z_{au} = \mathbf{U}_a'\boldsymbol{\lambda}; \qquad z_{av} = \mathbf{V}_a'\boldsymbol{\kappa} \tag{11a}$$

are to be computed, and it may be shown that, because of (4a) and (5a),

$$\mathbf{V}_a' = \mathbf{U}_a'\mathbf{T} + \mathbf{R}'; \qquad \mathbf{U}_a' = \mathbf{V}_a'\mathbf{T} + \mathbf{R}' \tag{12a}$$

in which $\mathbf{T}$ is as previously defined and $\mathbf{R}'$ is a row vector whose elements are 0's and 1's; specifically, $[r_j] = 1$ if $[V_j]$ is simply a power of the variable unique to $\mathbf{V}_a'$ and zero otherwise. Combining relevant portions of (8a), (11a) and (12a),

$$z_{av} = \mathbf{V}'\boldsymbol{\kappa} = (\mathbf{U}'\mathbf{T} + \mathbf{R}')\mathbf{T}^{-1}\boldsymbol{\lambda} = z_{au} - \mathbf{R}'\boldsymbol{\lambda} \tag{13a}$$

since $\mathbf{R}'\mathbf{T}^{-1} = -\mathbf{R}'$ for all appropriate choices of $\mathbf{R}'$ and $\mathbf{T}$. Thus z_{au} and z_{av} differ only by a constant, $\mathbf{R}'\boldsymbol{\lambda}$, and the same must be true of $\hat{z}_u$ and $\hat{z}_v$ as well, so that

$$z_{au} - \hat{z}_u = z_{av} - \hat{z}_v \tag{14a}$$

from which it is evident that whatever partition of the data is effected by $\hat{z}_u$ will be duplicated exactly by $\hat{z}_v$.

For functions containing coefficients of each term in (4a) constructed from variables subject to (5a), therefore,

a) the discriminant and each coefficient of the discriminant function of order q for any set of $(M-1)$ variables may be obtained as linear combinations of the coefficients and discriminant of any other set, and

b) all discriminants of the same order based upon sets of $(M-1)$ variables partition the data identically.

Application to Functions Used in Text

Using the subscript notation of (4a), in the first function described in the text $q=1$, in the second $q=2$, and in both $M=3$. For the binary linear discriminant function in a closed ternary array;

$$\mathbf{U}_a' = \{X_i, X_j\}; \qquad \mathbf{V}_a' = \{X_i, X_k\} = \{X_i, (1 - X_i - X_j)\}$$
$$\mathbf{u}_a' = \{x_i, x_j\}; \qquad \mathbf{v}_a' = \{x_i, (-x_i - x_j)\}$$

With $\mathbf{R}' = \{0, 1\}$ and

$$\mathbf{T} = \begin{bmatrix} 1 & -1 \\ 0 & -1 \end{bmatrix}$$

the conversions from $\mathbf{U}_a'$, $\mathbf{u}_a'$, $\mathbf{u}$ and $\mathbf{d}_u$ to $\mathbf{V}_a'$, $\mathbf{v}_a'$, $\mathbf{v}$ and $\mathbf{d}_v$ go through correctly. Substituting these values of $\mathbf{T}$ and $\mathbf{R}'$ in (9a) and (13a) we have at once that

$$\kappa_i = \lambda_i - \lambda_j, \qquad \kappa_k = -\lambda_j, \qquad z_v = z_u - \lambda_j$$
$$\lambda_i = \kappa_i - \kappa_k, \qquad \lambda_j = -\kappa_k, \qquad z_u = z_v - \kappa_k$$

which are the results observed in the actual computation and reported in the text. Extension to $M>3$ is immediate. Denoting unique variables by subscripts k and m and shared ones by subscripts $\leq (M-2)$, the general conversion rules are

$$\kappa_i = \lambda_i - \lambda_k; \qquad \lambda_i = \kappa_i - \kappa_m \qquad\qquad 1 \leq i \leq (M-2)$$
$$\kappa_m = -\lambda_k; \qquad z_v = z_u - \lambda_k; \qquad z_u = z_v - \kappa_m$$

For the elements of $\mathbf{U}_a'$, $\mathbf{V}_a'$ in any pair of binary linear+quadratic functions (i.e. functions in which, in eq. (4a), $M=3$, $q=2$) we have

$$U_1 = X_i \qquad U_2 = X_j \qquad U_3 = X_i^2 \qquad U_4 = X_iX_j \qquad U_5 = X_j^2$$
$$V_1 = X_i \qquad V_2 = X_k \qquad V_3 = X_i^2 \qquad V_4 = X_iX_k \qquad V_5 = X_k^2$$

and since j and k are the unique variables we must have $\mathbf{R} = \{0, 1, 0, 0, 1\}$.

Computing average values for each element, subtracting them from the initial values to obtain deviations, v and u, and expressing each v as a function of u, we have that

$$\begin{aligned} v_1 &= u_1 \\ v_2 &= -u_1 - u_2 \\ v_3 &= u_3 \\ v_4 &= u_1 - u_3 - u_4 \\ v_5 &= -2u_1 - 2u_2 + u_3 + 2u_4 + u_5 \end{aligned} \tag{15a}$$

The coefficients of u in each line of (15a) then form a column of the required right multiplier of $\mathbf{u}$, so that

$$\mathbf{T} = \begin{bmatrix} 1 & -1 & 0 & 1 & -2 \\ 0 & -1 & 0 & 0 & -2 \\ 0 & 0 & 1 & -1 & 1 \\ 0 & 0 & 0 & -1 & 2 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Substitution of these values for **T** and **R**′ in (9a) and (13a) gives the following conversion rules

$$\kappa_1 = \lambda_1 - \lambda_2 + \lambda_4 - 2\lambda_5 \qquad \kappa_2 = -\lambda_2 - 2\lambda_5$$
$$\kappa_3 = \lambda_3 - \lambda_4 + \lambda_5 \qquad \kappa_4 = -\lambda_4 + 2\lambda_5$$
$$\kappa_5 = \lambda_5 \qquad z_v = z_u - \lambda_2 - \lambda_5$$

together with a similar set obtained by interchanging κ and λ. Allowing for the difference in subscript notation, these are precisely the rules given in the text and first found by inspection of Table 4.