J. R. Statist. Soc. B (1982),
44, *No.* 2, pp. 139–177

The Statistical Analysis of Compositional Data

By J. AITCHISON

University of Hong Kong

[Read before the ROYAL STATISTICAL SOCIETY at a meeting organized by the RESEARCH SECTION on Wednesday, 13th January, 1982, Professor R. N. CURNOW in the Chair]

SUMMARY

The simplex plays an important role as sample space in many practical situations where compositional data, in the form of proportions of some whole, require interpretation. It is argued that the statistical analysis of such data has proved difficult because of a lack both of concepts of independence and of rich enough parametric classes of distributions in the simplex. A variety of independence hypotheses are introduced and interrelated, and new classes of transformed-normal distributions in the simplex are provided as models within which the independence hypotheses can be tested through standard theory of parametric hypothesis testing. The new concepts and statistical methodology are illustrated by a number of applications.

1. INTRODUCTION

THERE are many practical problems for which the positive simplex

$$\mathbb{S}^d = \{(x_1, ..., x_d): x_i > 0 (i = 1, ..., d), x_1 + ... + x_d < 1\}, \tag{1.1}$$

forms the whole, or a major component, of the sample space. For such problems, concepts of independence must often play an important role in any form of statistical analysis. The simplex, however, has proved to be an awkward space to handle statistically; the difficulties appear to lie in the scarcity of meaningful definitions of independence and of measures of dependence and in the absence of satisfactory parametric classes of distributions on $\mathbb{S}^d$. It is the aim of this paper to introduce a number of concepts of independence in the simplex, to relate these to some existing concepts, and to develop within the framework of rich new parametric classes of distributions appropriate statistical methods of analysis.

To motivate all the concepts introduced and to provide illustrations of the statistical methodology developed we shall use data sets in two very different areas of application, geology and consumer demand analysis. We hope that the expert reader will see these examples for what they are, attempts at providing potential statistical insights into these and similar disciplines rather than presumptuous criticism by a novice of interpretations already placed on the particular data sets.

Geology. The geological literature abounds with problems of the interpretation of chemical, mineral and fossil compositions of rock and sediment specimens. Each composition of each specimen is a set of some three to twenty proportions summing to unity and so can be represented by a point in an appropriately dimensioned simplex. We concentrate on three published geological data sets chosen to illustrate, as simply as possible, various aspects of our analysis.

Example 1: *Skye lavas.* Thompson, Esson and Duncan (1972), in their Table 2, give the chemical compositions of 32 basalt specimens from the Isle of Skye in the form of percentages of 10 major oxides. A typical percentage vector in $\mathbb{S}^9$ is thus

SiO_2	Al_2O_3	Fe_2O_3	MgO	CaO	Na_2O	K_2O	TiO_2	P_2O_5	MnO
46·31	14·18	12·32	12·74	9·62	2·51	0·34	1·53	0·16	0·18

For this example we shall discuss classes of parametric models for describing the experienced

 0035–9246/82/44139$2.00

pattern of variability, investigate the adequacy of such models and test a number of independence hypotheses for such sets of proportions.

Example 2: *Glacial tills in North-Central New York.* As part of a study of the composition of glacial till samples Kaiser (1962) presents, within his Table 1, the percentage compositions in terms of four pebble types, together with the total pebble count, of 93 till samples. Typical sample information thus takes the form

Percentage composition				Total
Red sandstone	Grey sandstone	Crystalline	Miscellaneous	pebbles
67·2	31·5	0·3	1·0	387

In addition to the composition in $\mathbb{S}^3$ we have here an abundance or size associated with each sample. Interest may then be in the extent, if any, to which composition depends on size.

Example 3: *Arctic lake sediments.* Coakley and Rust (1968) give, in their Table 1, the compositions in terms of sand, silt and clay percentages of 39 sediment samples at different water depths in an Arctic lake, with typical entry

Sediment composition in percentages			Water
Sand	Silt	Clay	depth (m)
10·5	55·4	34·1	49·4

Of interest here is the question of quantifying the extent to which water depth is explanatory of compositional pattern.

An appreciation of the difficulty imposed by this confinement of data points, such as the compositions in the above examples, to a simplex is inherent in the comments of Pearson (1897) on spurious correlations, and in geological circles the difficulty has since become known as the constant or bounded sum problem and the problem of closed arrays. As our analysis unfolds we shall cite various attempts to overcome this difficulty, and, in identifying reasons for limited success, we shall discover a means of overcoming most of the difficulties.

Consumer demand analysis. An important aspect of the study of consumer demand is the analysis of household budget surveys, in which attention focuses on expenditures on a number of mutually exclusive and exhaustive commodity groups and their relations to total expenditure, income, type of housing, household composition, and so on.

Example 4: *Hong Kong household expenditure budgets.* The set of household expenditure data available to us is from a pilot selection of 199 Hong Kong households, used as a preparatory study for a large-scale household expenditure survey by the Hong Kong Census and Statistics Department. From this set we have for simplicity selected subgroups of 41 and 42 households in two low-cost housing categories A and B. For each household information is available on number of persons, household composition, total household income, and monthly expenditures in nine commodity/service groups. The contents of these commodity groups are fully defined in the monthly Consumer Price Index Report of the Hong Kong Census and Statistics Department. To keep our illustrative analysis simple we have avoided the problem of zero components by combining two pairs of commodity groups to obtain the following seven: (1) housing, (2) fuel and light, (3) foodstuffs, (4) transport and vehicles, (5) tobacco, alcohol and miscellaneous goods, (6) services, (7) clothing, footwear and durable goods, and by replacing the few remaining zero expenditures in these groups by HK$0.05, half the lowest recordable expenditure.

In the investigation of such data the pattern or composition of expenditures, the proportions of total expenditure allocated to the commodity groups, can be shown to play a central role, and indeed some economists (Working, 1943; Leser, 1976; Deaton, 1978; Deaton and Muellbauer, 1980) have investigated such a budget-share approach. Since each pattern of expenditures is again represented by a point in the simplex, questions such as "To what extent

does the pattern of expenditure depend on the total amount spent?" and "Are there some commodity groups which are given priority in the allocation of expenditure?" obviously require adequate models to describe patterns of variability in the simplex and careful definitions of independence structure in the simplex for their satisfactory resolution.

2. Parametric Classes of Distributions on $\mathbb{S}^d$

2.1. *Fundamental Operations on Compositions*

As a first step towards the introduction of new classes of distributions and independence concepts we establish a suitable terminology and notation for certain mathematical operations in the simplex which help in the study and manipulation of compositional data.

Spaces and vectors. Let $\mathbb{R}^d$ denote d-dimensional real space, $\mathbb{P}^d$ its positive orthant and $\mathbb{S}^d$ its positive simplex (1.1). The symbols, $\mathbf{w}$, $\mathbf{x}$ and $\mathbf{y}$ are reserved for vectors in $\mathbb{P}^d$, $\mathbb{S}^d$ and $\mathbb{R}^d$, respectively, although we shall occasionally have to use other symbols for such vectors. Any vector or point $\mathbf{x}$ in $\mathbb{S}^d$ is termed a *composition* and any collection of such vectors, *compositional data*. We use the symbol x_{d+1} always in the sense

$$x_{d+1} = 1 - x_1 - \ldots - x_d, \tag{2.1}$$

to denote the fill-up value. The notation $\mathbf{x}^{(c)} = (x_1, \ldots, x_c)$ allows focusing on leading subvectors with the dimension of the subvector indicated by the superscript. Thus $\mathbf{x}^{(c)}$ with $c < d$ is a subvector of $\mathbf{x}$ or equivalently $\mathbf{x}^{(d)}$, and $\mathbf{x}^{(d+1)}$ is the augmented $\mathbf{x}$ vector $(x_1, \ldots, x_d, x_{d+1})$. The subvector $(x_{c+1}, \ldots, x_{d+1})$ obtained by deletion of $\mathbf{x}^{(c)}$ from $\mathbf{x}^{(d+1)}$ is denoted by $\mathbf{x}_{(c)}$. We use $T(\mathbf{x}^{(c)})$ to denote the sum $x_1 + \ldots + x_c$ of the elements of any vector or subvector, such as $\mathbf{x}^{(c)}$.

Basis of a composition. In our household expenditure example the d-dimensional budget-share composition $\mathbf{x}^{(d+1)}$ is derived from the actual amounts spent $\mathbf{w}^{(d+1)}$ on the $d+1$ commodity groups through an operation C: $\mathbb{P}^{d+1} \to \mathbb{S}^d$ defined by $\mathbf{x}^{(d+1)} = C(\mathbf{w}^{(d+1)})$ where $x_i = w_i / T(\mathbf{w}^{(d+1)})$ $(i = 1, \ldots, d+1)$. For convenience we term such a vector $\mathbf{w}^{(d+1)} \in \mathbb{P}^{d+1}$, when it exists, the *basis* of the composition $\mathbf{x}^{(d+1)}$.

Subcomposition. Often in the study of geochemical compositions attention is directed towards the relative proportions of a few oxides. For example, a popular diagrammatic representation treats the relative proportions

$$(\mathrm{CaO},\ \mathrm{Na_2O},\ \mathrm{K_2O})/(\mathrm{CaO} + \mathrm{Na_2O} + \mathrm{K_2O})$$

in $\mathbb{S}^2$ as triangular coordinates in a CNK ternary diagram. We can formalize this process of focusing on a subset of components as follows. Any subvector, such as $\mathbf{x}^{(c)}$, of a composition $\mathbf{x}^{(d+1)}$ can play the role of a basis in $\mathbb{P}^c$ for a composition $C(\mathbf{x}^{(c)})$ in $\mathbb{S}^{c-1}$. Such a composition is termed a *subcomposition* $C(\mathbf{x}^{(c)})$ of $\mathbf{x}^{(d+1)}$.

Amalgamation. In a household expenditure enquiry there may be reasons for combining some commodity groups, to form new amalgamated groups. If we suppose that the composition has been ordered in such a way that combinations are between neighbouring components, the formal general process can be set out as follows. Let the integers $c_0, \ldots . c_{k+1}$ satisfy

$$0 = c_0 < c_1 < \ldots < c_k < c_{k+1} = d+1 \tag{2.2}$$

and define

$$t_j = x_{c_{j-1}+1} + \ldots + x_{c_j} \quad (j = 1, \ldots, k+1). \tag{2.3}$$

Then $\mathbf{t}^{(k+1)} \in \mathbb{S}^k$ and so is a k-dimensional composition which we term an *amalgamation* of $\mathbf{x}^{(d+1)}$. It is obvious that the transformation from $\mathbf{x}^{(d+1)}$ to $\mathbf{t}^{(k+1)}$ can be represented by a matrix

some commodity groups, to form new amalgamated groups. If we suppose that the composition has been ordered in such a way that combinations are between neighbouring components, the formal general process can be set out as follows. Let the integers $c_0, \ldots . c_{k+1}$ satisfy

$$0 = c_0 < c_1 < \ldots < c_k < c_{k+1} = d+1 \tag{2.2}$$

and define

$$t_j = x_{c_{j-1}+1} + \ldots + x_{c_j} \quad (j = 1, \ldots, k+1). \tag{2.3}$$

Then $\mathbf{t}^{(k+1)} \in \mathbb{S}^k$ and so is a k-dimensional composition which we term an *amalgamation* of $\mathbf{x}^{(d+1)}$. It is obvious that the transformation from $\mathbf{x}^{(d+1)}$ to $\mathbf{t}^{(k+1)}$ can be represented by a matrix operation $\mathbf{t}^{(k+1)} = \mathbf{A}\mathbf{x}^{(d+1)}$ from $\mathbb{S}^d$ to $\mathbb{S}^k$, where $\mathbf{A}$ consists of 0s and 1s, with a single 1 in each column.

Partition. The amalgamation just discussed involves a separation of the vector $\mathbf{x}^{(d+1)}$ into $k+1$ subvectors. When considering such an amalgamation we may often be interested also in

the $k+1$ subcompositions associated with these subvectors. The jth such subcomposition, $\mathbf{s}_j \in \mathbb{S}^{d_j}$ where $d_j = c_j - c_{j-1} - 1$, has components

$$s_{jr} = x_{c_{j-1}+r}/t_j \quad (r = 1, ..., d_j + 1), \tag{2.4}$$

where the (d_j+1)th component is the fill-up value. An extremely useful feature is that the transformation from $\mathbb{S}^d$ to

$$\mathbb{S}^k \times \prod_{j=1}^{k+1} \mathbb{S}^{d_j} \tag{2.5}$$

specified by

$$P(\mathbf{x}^{(d+1)}) = (\mathbf{t}; \mathbf{s}_1, ..., \mathbf{s}_{k+1}) \tag{2.6}$$

is one-to-one, with Jacobian $D\mathbf{x}^{(d)}/D(\mathbf{t}; \mathbf{s}_1, ..., \mathbf{s}_{k+1}) = t_1^{d_1} ... t_{k+1}^{d_{k+1}}$ and with inverse P^{-1} given by $x_{c_{j-1}+r} = t_j s_{jr}$ $(r = 1, ..., d_j;\ j = 1, ..., k+1)$. We shall refer to $P(\mathbf{x}^{(d+1)})$ as a *partition of order k* of the composition $\mathbf{x}^{(d+1)}$. Thus a partition directs attention to an amalgamation together with its associated subcompositions.

Independence notation. In discussing statistical independence we use the $\perp\!\!\!\perp$ notation of Dawid (1979). Thus $C(\mathbf{x}^{(c)}) \perp\!\!\!\perp C(\mathbf{x}_{(c)})$ denotes independence of the two subcompositions, and $C(\mathbf{x}^{(c)}) \perp\!\!\!\perp C(\mathbf{x}_{(c)}) \mid T(\mathbf{x}^{(c)})$ denotes their conditional independence, given the sum, $x_1 + ... + x_c$. We use $\perp\!\!\!\perp \mathbf{w}^{(d+1)}$ to indicate that $\mathbf{w}^{(d+1)}$ consists of independent components.

2.2. *The Dirichlet Class*

Undoubtedly the only familiar class of distributions on $\mathbb{S}^d$ is the Dirichlet class with typical member $D^d(\boldsymbol{\alpha})$ having density function

$$\prod_{i=1}^{d+1} x_i^{\alpha_i - 1}/\Delta(\boldsymbol{\alpha}) \quad (\mathbf{x}^{(d)} \in \mathbb{S}^d),$$

where $\boldsymbol{\alpha}$ or $\boldsymbol{\alpha}^{(d+1)} \in \mathbb{P}^{d+1}$ is a $(d+1)$-vector parameter and

$$\Delta(\boldsymbol{\alpha}) = \Gamma(\alpha_1) ... \Gamma(\alpha_{d+1})/\Gamma(\alpha_1 + ... + \alpha_{d+1})$$

is the Dirichlet function. A major obstacle to its use in the statistical analysis of compositional data is that it seldom, if ever, provides an adequate description of actual patterns of variability of compositions. The reasons for this are not difficult to find. First, the isoprobability contours of every Dirichlet distribution with $\alpha_i > 1$ $(i = 1, ..., d+1)$ are convex, and so the Dirichlet class must fail to describe obviously concave data patterns such as in Fig. 1. More importantly, the Dirichlet class has so much independence structure built into its definition that it represents, not a convenient modelling class for compositional data but the ultimate in independence hypotheses. This strong independence structure stems from a well-known relationship between the Dirichlet and gamma classes, which can be expressed in the terminology of compositional data as follows.

D1. Any Dirichlet composition in $\mathbb{S}^d$ can be expressed as the composition of a basis of $d+1$ independent gamma-distributed quantities, each with the same scale parameter.

There are many ways of expressing the strong internal independence structure of $D^d(\boldsymbol{\alpha})$ without reference to a conceptual external basis. For our purposes here we can collect most of these into a single general result concerning any partition of a Dirichlet composition.

D2. If $\mathbf{x}^{(d+1)}$ is $D^d(\boldsymbol{\alpha})$ then, for partition (2.6), $\mathbf{t} \perp\!\!\!\perp \mathbf{s}_1 \perp\!\!\!\perp ... \perp\!\!\!\perp \mathbf{s}_{k+1}$, with $\mathbf{t}$ of $D^k(\boldsymbol{\gamma})$ form and $\mathbf{s}_j$ of $D^{d_j}(\boldsymbol{\beta}_j \gamma_j)$ form $(j = 1, ..., k+1)$ where $P(\boldsymbol{\alpha}^{(d+1)}) = (\boldsymbol{\gamma}; \boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_{k+1})$.

We shall show later the relevance of these two properties to various concepts of independence in the simplex.

The realization that the Dirichlet class leans so heavily towards independence has prompted a number of authors (Connor and Mosimann, 1969; Darroch and James, 1974; Mosimann, 1975b; James and Mosimann, 1980; James, 1981) to search for generalizations of

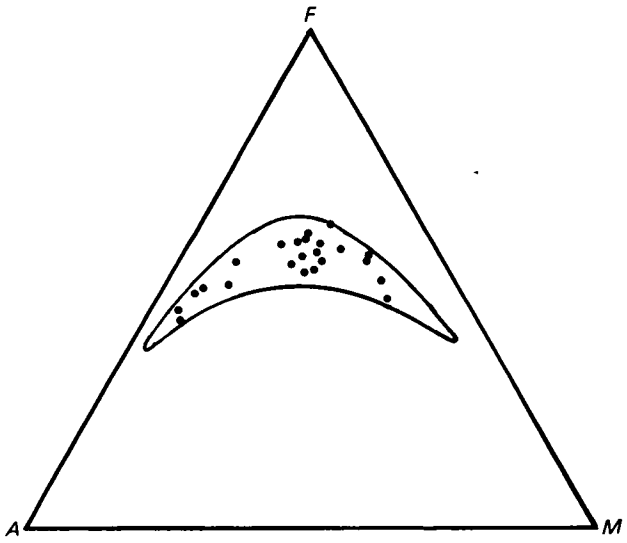

FIG. 1. A concave data set and the 95 per cent prediction region of a fitted additive logistic normal distribution. The points are the subcompositions $C(Na_2O + K_2O, Fe_2O_3, MgO)$ of 23 aphyric Si-poor basalt-benmoreites from the AFM diagram of Fig. 7 of Thomson, Essen and Duncan (1972).

the Dirichlet class with less structure. Their efforts have met with only limited success and it remains an open problem to find a useful parametric class of distributions on $\mathbb{S}^d$ which contains the Dirichlet class but also contains distributions which do *not* satisfy any of the simplex independence properties already appearing in the literature or to be introduced in this paper.

In our view the way out of the impasse is simply to travel by a different route, escaping from the awkward constrictions of $\mathbb{S}^d$ into the wide open spaces of $\mathbb{R}^d$ through suitably selected transformations between $\mathbb{S}^d$ and $\mathbb{R}^d$.

2.3. *Transformed Normal Classes*

The idea of inducing a tractable class of distributions over some awkward sample space from a proven and well-established class over some simpler space is at least a century old. McAlister (1879), faced with the "awkward" sample space $\mathbb{P}^1$, saw that if he considered y in $\mathbb{R}^1$ to be $N(\mu, \sigma^2)$ then the transformation $x = \exp(y)$ would induce a useful "expnormal" distribution $\Lambda(\mu, \sigma^2)$ on $\mathbb{P}^1$: he, of course, expressed the idea in terms of the inverse, logarithmic, transformation and we are stuck with the name lognormal. Over the century there has been a continuing interest in transformations to normality, intensified in recent years following the work of Box and Cox (1964) and the increasing availability of tests of multinormality, as in Andrews, Gnanadesikan and Warner (1973). It seems surprising therefore that the idea of moving from multinormal distributions $N^d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ on $\mathbb{R}^d$ to a class $f\,N^d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ of distributions on $\mathbb{S}^d$ by a suitable transformation f: $\mathbb{R}^d \to \mathbb{S}^d$ has been so slow to emerge. Our surprise must be even greater when one such transformation, the additive logistic transformation a_d: $\mathbb{R}^d \to \mathbb{S}^d$ defined in Table 1, is already heavily exploited in other areas of statistical activity, such as logistic discriminant analysis (Cox, 1966; Day and Kerridge, 1967; Anderson, 1972) and in the analysis of binary data (Cox, 1970).

transformation and we are stuck with the name lognormal. Over the century there has been a continuing interest in transformations to normality, intensified in recent years following the work of Box and Cox (1964) and the increasing availability of tests of multinormality, as in Andrews, Gnanadesikan and Warner (1973). It seems surprising therefore that the idea of moving from multinormal distributions $N^d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ on $\mathbb{R}^d$ to a class $f\,N^d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ of distributions on $\mathbb{S}^d$ by a suitable transformation f: $\mathbb{R}^d \to \mathbb{S}^d$ has been so slow to emerge. Our surprise must be even greater when one such transformation, the additive logistic transformation a_d: $\mathbb{R}^d \to \mathbb{S}^d$ defined in Table 1, is already heavily exploited in other areas of statistical activity, such as logistic discriminant analysis (Cox, 1966; Day and Kerridge, 1967; Anderson, 1972) and in the analysis of binary data (Cox, 1970).

Aitchison and Shen (1980) have identified as the logistic-normal class those distributions induced on $\mathbb{S}^d$ from the class of $N^d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distributions on $\mathbb{R}^d$ by the transformation a_d. The earliest explicit mention of this class we have traced is in a personal communication to Johnson and Kotz (1972, p. 20) by Obenchain, who does not seem to have developed the idea

TABLE 1

Elementary logistic transformations from $\mathbb{R}^d$ to $\mathbb{S}^d$

Name and notation	*Specification*	*Inverse*
Additive a_d	$x_i\left\{1+\sum_{j=1}^{d}\exp(y_i)\right\} = \begin{cases}\exp(y_i) & (i=1,\ldots,d)\\ 1 & (i=d+1)\end{cases}$	$y_i = \log\dfrac{x_i}{x_{d+1}}$
Multiplicative m_d	$x_i\prod_{j=1}^{i}\{1+\exp(y_i)\} = \begin{cases}\exp(y_i) & (i=1,\ldots,d)\\ 1 & (i=d+1)\end{cases}$	$y_i = \log\dfrac{x_i}{1-\sum_{j=1}^{i}x_j}$
Hybrid h_d	$x_1 = \exp(y_1)/\{1+\exp(y_1)\}$	$y_i = \log\dfrac{x_1}{1-x_1}$
	$x_i\left\{1+\sum_{j=1}^{i-1}\exp(y_i)\right\}\left\{1+\sum_{j=1}^{i}\exp(y_j)\right\}$	
	$= \exp(y_i),\ (i=2,\ldots,d)$	$y_i = \log\dfrac{x_i}{\left(1-\sum_{j=1}^{i-1}x_j\right)\left(1-\sum_{j=1}^{i}x_j\right)}$
	$x_{d+1} = 1\Big/\left\{1+\sum_{j=1}^{d}\exp(y_j)\right\}$	$(i=2,\ldots,d)$

further. Aitchison and Shen (1980) cite a number of earlier implicit uses, particularly as a vehicle for the description of prior and posterior distributions of vectors of multinomial probabilities which are naturally confined to a suitably dimensioned simplex. Leonard (1973) started a thorough investigation of this use of the class over simplex parameter spaces. The first use of the class for describing patterns of variability of data appears to be for probabilistic data in a medical diagnostic problem by Aitchison and Begg (1976) and for compositional data by Aitchison and Shen (1980), who discuss a number of useful properties and demonstrate the simplicity of its application in a variety of problems. Our interest here in logistic-normal distributions is in their membership of a wider class of transformed normal distributions on the simplex and their use in relation to the independence concepts of subsequent sections.

The additive logistic transformation a_d is by no means the only transformation from $\mathbb{R}^d$ to $\mathbb{S}^d$, and may be quite unsuited to particular investigations. Table 1 gives two other elementary transformations, the multiplicative logistic m_d and the hybrid logistic h_d. All three transformations a_d, m_d, h_d have Jacobian $D\mathbf{x}/D\mathbf{y}$ given by $x_1 x_2 \ldots x_{d+1}$. We shall see that such elementary transformations can act as the building blocks of much more complicated transformations. An obvious comment is that the exponential function used in the definitions is not an essential feature; it could be replaced by any one-to-one transformation from $\mathbb{R}^1$ to $\mathbb{P}^1$, though there are few transformations as tractable.

There are two main ways of building further useful transformations.

Linear transformation method. The fact that the N^d class on $\mathbb{R}^d$ is closed under the group of non-singular linear transformations implies that any one of the elementary transformed-normal classes on $\mathbb{S}^d$ will have a related closure property (Aitchison and Shen, 1980). In practical terms this means that we could replace $\mathbf{y}^{(d)}$ by $\mathbf{Q}\mathbf{y}^{(d)}$, with $\mathbf{Q}$ non-singular, in any one of the elementary transformations and formally obtain a new transformation but with the assurance that we are remaining within the same class of distributions on $\mathbb{S}^d$. For example, with a_d and

$$q_{ii} = 1\ (i=1,\ldots,d), \quad q_{i,i+1} = -1\ (i=1,\ldots,d-1), \quad q_{ij} = 0$$

otherwise, we obtain a new transformation

$$x_i\left\{1+\sum_{k=1}^{d}\exp\left(\sum_{j=k}^{d}y_j\right)\right\}=\exp\left(\sum_{j=i}^{d}y_j\right),\quad y_i=\log(x_i/x_{i+1})\quad(i=1,\ldots,d).$$

involving ratios of adjacent components of the composition.

Partition transformation method. When a partition $P(\mathbf{x}^{(d+1)})$, as defined in (2.6), is under consideration a relevant transformation from $\mathbb{R}^d$ to $\mathbb{S}^d$ may be constructed as follows. Let

$$f_0\colon \mathbb{R}^d\to\mathbb{S}^k,\quad f_j\colon \mathbb{R}^{d_j}\to\mathbb{S}^{d_j}\quad(j=1,\ldots,k+1)$$

be any $k+2$ suitably dimensioned elementary transformations from Table 1. The compound $\mathbf{f}=(f_0;f_1,\ldots,f_{k+1})$ is a one-to-one transformation

$$\mathbf{f}\colon \mathbb{R}^d=\mathbb{R}^k\times\prod_{j=1}^{k+1}\mathbb{R}^{d_j}\to\mathbb{S}^k\times\prod_{j=1}^{k+1}\mathbb{S}^{d_j}.$$

and the inverse transformation P^{-1} then takes us further on to $\mathbb{S}^d$ to complete a transformation $P^{-1}\mathbf{f}$ from $\mathbb{R}^d$ to $\mathbb{S}^d$. We denote this resultant transformation shortly by $(f_0;f_1,\ldots,f_{k+1})$. The choice of $f_j\,(j=0,\ldots,k+1)$ from among the appropriately dimensioned elementary transformations obviously offers a multitude of transformations from $\mathbb{R}^d$ to $\mathbb{S}^d$. The choice in any particular application should clearly depend on the situation under investigation.

3. Validity of Transformed-Normal Models

Any statistical weapon designed to overcome such a resistant fortress as the simplex is unlikely to gain acceptance before undergoing proving tests as to its suitability to the terrain.

Goodness-of-fit tests. If $\mathbf{x}^{(d+1)}$ follows a $fN^d(\boldsymbol{\mu},\boldsymbol{\Sigma})$ distribution in $\mathbb{S}^d$ then $\mathbf{y}^{(d)}=f^{-1}(\mathbf{x}^{(d+1)})$ follows a $N^d(\boldsymbol{\mu},\boldsymbol{\Sigma})$ distribution in $\mathbb{R}^d$. We can thus test the goodness of fit of any transformed-normal class to a compositional data set by applying the now extensive battery of multivariate normal tests, as for example in Andrews, Gnanadesikan and Warner (1973), to the transformed data set.

For d-dimensional compositional data sets we have applied Kolmogorov–Smirnov and Cramér–von Mises tests in their Stephens (1974) versions to all d marginal distributions, to all $\frac{1}{2}d(d-1)$ bivariate angle distributions, and to the distribution of d-dimensional radii. For the Skye lava compositions we have tested in this way both the additive aN^9 and the multiplicative mN^9 logistic-normal models. For the additive version not a single one of the battery of 92 tests gives a significant indication of non-normality at the 5 per cent significance level; for the multiplicative version only one of the marginal tests gives evidence of any departure from normality, at the 1 per cent significance level. Application of the battery of tests to another 20 data sets of different geological types similarly encourages the view that transformed-normal distributions may have an important practical role to play in the analysis of compositional data.

The ability of transformed-normal distributions to cope with concave data sets in $\mathbb{S}^d$ is illustrated in Fig. 1 where the 95 per cent prediction region of a fitted additive logistic-normal distribution, constructed by transformation of the corresponding elliptical region in $\mathbb{R}^2$, neatly contains the data points.

Two caveats are worth recording. First, testing for multivariate normality and trying to detect outliers are two highly interrelated activities (Gnanadesikan and Kettenring, 1972); delicate judgements may occasionally have to be made between rejection of an apparent

data sets of different geological types similarly encourages the view that transformed-normal distributions may have an important practical role to play in the analysis of compositional data.

The ability of transformed-normal distributions to cope with concave data sets in $\mathbb{S}^d$ is illustrated in Fig. 1 where the 95 per cent prediction region of a fitted additive logistic-normal distribution, constructed by transformation of the corresponding elliptical region in $\mathbb{R}^2$, neatly contains the data points.

Two caveats are worth recording. First, testing for multivariate normality and trying to detect outliers are two highly interrelated activities (Gnanadesikan and Kettenring, 1972); delicate judgements may occasionally have to be made between rejection of an apparent outlier to justify multivariate normal modelling and retention of suspect data with consequently more complex modelling. Secondly, in multivariate normal regression modelling, multivariate normality of the vector residuals, not of the regressand vectors, is the hypothesis under scrutiny. Thus in Example 3 the sediment compositions show significant departure from

logistic-normality, whereas in the appropriate regression analysis on the explanatory water depth, reported later in Section 7.3, the residuals survive such scrutiny.

Genesis models. Many of the natural and sampling processes by which compositions are determined are extremely complex; see, for example, the description by Chayes (1971, p. 44) for some geological sampling. Just as some support for normal and lognormal modelling can be provided by additive and multiplicative central limit theorems so we can postulate a process of random modifications to compositions which lead, through central limit theory arguments, to transformed-normal distributions for compositions. The underlying concept is that of a *perturbation* $\mathbf{w}^{(d+1)} \in \mathbb{P}^{d+1}$, whose effect on a composition $\mathbf{x}^{(d+1)} \in \mathbb{S}^d$ is to produce a perturbed composition

$$\mathbf{w} \circ \mathbf{x} = C(w_1 x_1, ..., w_{d+1} x_{d+1}).$$

Successive perturbations $\mathbf{w}_{[1]}, \mathbf{w}_{[2]}, ...$ on an initial composition $\mathbf{x}_{[0]}$ produce a sequence of compositions $\mathbf{x}_{[1]}, \mathbf{x}_{[2]}, ...$, related by $\mathbf{x}_{[r]} = \mathbf{w}_{[r]} \circ \mathbf{x}_{[r-1]}$ $(r = 1, 2, ...)$ and satisfying

$$\log(x_{rj}/x_{r,d+1}) = \log(x_{0j}/x_{0,d+1}) + \sum_{i=1}^{r} \log(w_{ij}/w_{i,d+1}).$$

It is then clear that suitable conditions on the perturbations could lead, for large r, to approximately additive logistic-normal or aN^d distributions for $\mathbf{x}_{[r]}$.

4. Extrinsic Analysis of Independence

4.1. *Introduction*

We distinguish between two forms of structural analysis of compositional data:

(1) extrinsic analysis, where compositions in $\mathbb{S}^d$ have been derived, or are conceptualized as arising, from bases in $\mathbb{P}^{d+1}$ and interest is in the relation of composition to basis;
(2) intrinsic analysis, where there is no basis and so interest is not directed outside the simplex but in the composition *per se*.

In this section we consider two independence concepts of extrinsic analysis.

One general point should first be made. It will be obvious that most of the independence concepts introduced and their properties could be presented in a weaker moment form involving correlations. Since a main aim is to develop tests of hypotheses within transformed normal models, where independence and zero correlation coincide, we have not considered it worthwhile to interrupt the narrative to draw such fine distinctions when they exist.

4.2. *Compositional Invariance*

In Examples 2 and 4 the compositions arise from actual bases in the form of quantities of different types of pebbles and expenditures in different commodity groups. Questions such as "Is pebble-type composition independent of the abundance of the pebbles?" and "To what extent is the pattern of household expenditure dependent on total expenditure?" direct us towards investigation of the relationship between the composition $\mathbf{x} = C(\mathbf{w})$ and the total size $t = T(\mathbf{w})$ of a basis $\mathbf{w} \in \mathbb{P}^{d+1}$. This leads naturally to the following independence concept.

Definition: compositional invariance of a basis. A basis $\mathbf{w} \in \mathbb{P}^{d+1}$ is *compositionally invariant* if $C(\mathbf{w}) \perp\!\!\!\perp T(\mathbf{w})$.

This concept has appeared under a variety of guises: as the Lukacs condition in a characterization of the Dirichlet distribution (Mosimann, 1962), as additive isometry in the analysis of biological shape and size (Mosimann, 1970, 1975a, b), as proportion invariance in the study of F-independence (Darroch and James, 1974).

The development of a satisfactory parametric test of compositional invariance seems to have been delayed by two model-building deficiencies of the multivariate lognormal class $\Lambda^{d+1}(\boldsymbol{\mu}, \boldsymbol{\Omega})$, a natural first-thought contender for the role of modelling the variability of bases in $\mathbb{P}^{d+1}$.

(1) If $\mathbf{w}$ is $\Lambda^{d+1}(\boldsymbol{\mu}, \boldsymbol{\Omega})$ there is no simple, tractable form for the distribution of $T(\mathbf{w})$ and so investigation of $C(\mathbf{w}) \perp\!\!\!\perp T(\mathbf{w})$ is difficult.

(2) A multivariate lognormal basis $\mathbf{w}$ can be compositionally invariant only if $\mathbf{w}$ has a degenerate, one-dimensional distribution with covariance matrix $\boldsymbol{\Omega} = \mathrm{cov}(\log \mathbf{w})$ a scalar multiple of the matrix $\mathbf{U}_{d+1}$ consisting of unit elements and so of rank 1 (Mosimann, 1975b).

Thus not only from a point of view of tractability but also on logical grounds, study of compositional invariance within multivariate lognormal modelling of the basis is doomed to failure. Since non-degenerate compositional invariance is obviously a logical possibility the message to the practical statistician is clear: he must do better in his modelling. With transformed normal classes the answer is easy. Since interest is in $\mathbf{x} \perp\!\!\!\perp t$ we need not insist on finding an elegant model for the joint distribution of $(t, \mathbf{x})$ but concentrate on the conditional distribution $p(\mathbf{x} \mid t)$ using a transformed normal regression form such as $fN^d(\boldsymbol{\alpha} + \boldsymbol{\beta} t, \boldsymbol{\Sigma})$ or $fN^d(\boldsymbol{\alpha} + \boldsymbol{\beta} \log t, \Sigma)$. Then compositional invariance is simply the parametric hypothesis $\boldsymbol{\beta} = \mathbf{0}$. Moreover, testing this hypothesis on a data set consisting of n bases, and hence of n pairs of corresponding compositions and sizes, is standard methodology in multivariate analysis of dispersion (Morrison, 1976, Chapter 5). This regression approach seems appropriate since we would surely want, in the event of rejecting the hypothesis of compositional invariance, to study the basis further by trying to describe the nature of the dependence of composition on size.

Glacial tills. We have tested compositional invariance for the 93 pebble samples of Example 2 in both the $aN^3(\boldsymbol{\alpha} + \boldsymbol{\beta} t, \boldsymbol{\Sigma})$ and $aN^3(\boldsymbol{\alpha} + \boldsymbol{\beta} \log t, \boldsymbol{\Sigma})$ models with very similar results. Using the generalized likelihood ratio criterion as in Morrison (1976, p. 222) we obtain values 2·74 and 3·05 for the test statistics, each to be compared against 7·81, the upper 5 per cent χ^2 (3) point. Thus there is no evidence against compositional invariance in these glacial tills. Two comments should be made. First, while two of the marginal tests indicate evidence of departure from additive logistic normality the other tests show no such evidence. Secondly, zero components in 14 of the samples were replaced by proportions 0·0005, half the lowest recorded value, before analysis. We shall return to this problem of zeros in Section 7.4.

Household expenditure budgets. Incorporating compositional analysis directly into the analysis of household budgets has many advantages and provides opportunities for new forms of investigation. Modelling as above with $p(\mathbf{x} \mid t)$ of $aN^d(\boldsymbol{\alpha} + \boldsymbol{\beta} \log t, \boldsymbol{\Sigma})$ form has interesting consequences. First, the sometimes troublesome Engel aggregation condition (Brown and Deaton, 1972, p. 1163) that, for each household, total expenditure should equal the sum of all commodity expenditures, is automatically satisfied. Secondly, the hypothesis of compositional invariance, $\boldsymbol{\beta} = \mathbf{0}$, has a direct interpretation in terms of the income elasticities $e_i = \partial \log w_i / \partial \log t$ of demand $(i = 1, ..., d+1)$, if for simplicity we identify household total expenditure with household income. In expectation terms $\beta_i = e_i - e_{d+1}$ $(i = 1, ..., d)$, so that compositional invariance corresponds to equality of all $d+1$ income elasticities. Thirdly, whether or not there is compositional invariance, the modelling can clearly be extended to a full consumer demand analysis by the incorporation of commodity prices and other explanatory variables such as household type and household composition into the mean parameter of the aN^d distribution. Indeed such an extension can be shown to be identical with the Houthakker (1960) indirect addilog model of consumer demand (Brown and Deaton, 1972, equation 115).

There is, however, an important extra flexibility in the present compositional approach, for we are not restricted to the additive logistic transformation but could equally use other forms, for example, directed towards the investigation of whether households place priorities in allocation of expenditures on some commodity groups.

In the above discussion we have identified household total expenditure t with household income s. This is not an essential feature of the modelling since we could approach it through

explanatory variables such as household type and household composition into the mean parameter of the aN^d distribution. Indeed such an extension can be shown to be identical with the Houthakker (1960) indirect addilog model of consumer demand (Brown and Deaton, 1972, equation 115).

There is, however, an important extra flexibility in the present compositional approach, for we are not restricted to the additive logistic transformation but could equally use other forms, for example, directed towards the investigation of whether households place priorities in allocation of expenditures on some commodity groups.

In the above discussion we have identified household total expenditure t with household income s. This is not an essential feature of the modelling since we could approach it through the conditioning

$$p(s, t, \mathbf{x}) = p(s)\, p(t \mid s)\, p(\mathbf{x} \mid s, t)$$

with perhaps the reasonable assumption that $\mathbf{x} \perp\!\!\!\perp s \mid t$ leading to the above focus on $p(\mathbf{x} \mid t)$.

Application of the test of compositional invariance gives observed values of 36·4 and 39·0 for the test statistics for household types A and B respectively, each to be compared against upper $\chi^2(6)$ values, and hence highly significant. Thus for both types A and B the hypothesis of compositional invariance is firmly rejected, not surprisingly when we recall that the hypothesis is equivalent to the equality of the income elasticities for all commodity groups. More interestingly, from the estimated values of β_i the relationship $\beta_i = e_i - e_{d+1}$ provides us with an ordering of the commodity groups in terms of increasing magnitude of income elasticity, that is in conventional economic jargon from necessity to increasing luxury groups. For household type A this ordering is as follows: housing; fuel and light; foodstuffs; transport and vehicles; alcoholic drinks, tobacco and miscellaneous goods; services; clothing, footwear and durable goods. For household type B the ordering is identical except that the groups 4 and 5 are interchanged. While these orderings seem reasonable for Hong Kong it should be clear that any satisfactory analysis must involve the introduction of concomitant explanatory variables such as household size and the use of data from the eventual household expenditure survey rather than from specially selected pilot households. We hope to report on a more detailed analysis elsewhere.

4.3. *Basis Independence*

Even when no basis actually exists a number of authors, conscious of the difficulties of defining independence concepts for compositions, have seen a method of escape through the relating of the compositional property to that of independence of an imaginary basis. Their various forms of this idea can be simply expressed as follows.

Definition: basis independence. A composition $\mathbf{x}^{(d)} \in \mathbb{S}^d$ is said to have *basis independence* if there exists a basis $\mathbf{w}^{(d+1)} \in \mathbb{P}^{d+1}$ with $\perp\!\!\!\perp \mathbf{w}^{(d+1)}$ and such that $\mathbf{x}^{(d)} = C(\mathbf{w}^{(d+1)})$.

Since every Dirichlet-distributed composition has basis independence, by property D1 of Section 2.2, the Dirichlet class has obviously no fruitful rôle to play in the investigation of this independence property.

Attention has concentrated on assessing null correlations, the spurious correlations that would arise in the raw proportions solely from the process of forming proportions from conceptual, independent basis measurements, and subsequently on comparing sample correlations against these null values (Chayes, 1960, 1962, 1971; Mosimann, 1962; Chayes and Kruskal, 1966; Darroch, 1969). Many awkward features and pitfalls of this direct correlational approach have been pointed out: see, for example, Aitchison (1981a) who, after emphasizing the limitations of inferences about bases from compositions imposed by the fact that a composition $\mathbf{x}^{(d+1)}$ determines a basis $\mathbf{w}^{(d+1)} = t\mathbf{x}^{(d+1)}$ only up to a multiplicative factor t, provides an overall test by showing that basis independence is associated with a particularly simple covariance structure of *logratios* of the raw proportions:

$$\operatorname{cov}\{\log(\mathbf{x}^{(d)}/x_{d+1})\} = \operatorname{diag}\{\lambda_1, \ldots, \lambda_d\} + \lambda_{d+1}\mathbf{U}_d, \quad (\lambda_i > 0,\ i = 1, \ldots, d+1), \tag{4.1}$$

where $\mathbf{U}_d$ is the $d \times d$ matrix of units. Even a simplified approach, however, has merit only so long as it proves impossible to provide an equivalent intrinsic concept. Since we have now discovered a simple way of defining the illusive concept of almost-independence within the composition itself we proceed immediately to this new concept.

5. Intrinsic Analysis: Complete Subcompositional Independence

It has long been appreciated that there must be at least one pair of correlated components in any composition $\mathbf{x}^{(d+1)}$. An obvious first problem in studying independence in $\mathbb{S}^d$ is therefore to find a structure which most closely approaches the unattainable goal of $\perp\!\!\!\perp \mathbf{x}^{(d+1)}$. The following definition embodies such a concept.

Definition: complete subcompositional independence. A composition $\mathbf{x}^{(d+1)}$ has *complete*

subcompositional independence if, for each possible partition of $\mathbf{x}^{(d+1)}$, the set of all its subcompositions is independent.

Every Dirichlet composition has complete subcompositional independence, by D2 of Section 2.2. Note also that complete subcompositional independence is automatically satisfied by any composition of dimension $d = 1$ or 2, since partitions involve one-component subvectors such as x_1 which have trivial subcompositions such as $C(x_1) = 1$.

For a composition $\mathbf{x}^{(d+1)}$ with complete subcompositional independence, $C(\mathbf{x}^{(b)}) \perp\!\!\!\perp C(\mathbf{x}_{(c)})$ for $b \leqq c$. Moreover, since every subcomposition based on a two-dimensional subvector such as (x_1, x_2) is a function only of the ratio x_1/x_2, complete subcompositional independence implies independence of every pair of ratios x_i/x_j and x_k/x_l with i, j, k, l all different and, *a fortiori*, of the logratios $\log(x_i/x_j)$ and $\log(x_k/x_l)$. This implication can be fully expressed in terms of the special form for the covariance structure

$$\boldsymbol{\Sigma}_H = \operatorname{cov}\{\log(\mathbf{x}^{(d)}/x_{d+1})\} = \operatorname{diag}(\lambda_1, \ldots, \lambda_d) + \lambda_{d+1}\mathbf{U}_d, \tag{5.1}$$

where $\boldsymbol{\lambda}^{(d+1)}$ has the following interpretations:

$$\lambda_i = \operatorname{cov}\{\log(x_j/x_i), \log(x_k/x_i)\}, \quad \lambda_i + \lambda_j = \operatorname{var}\{\log(x_i/x_j)\} \tag{5.2}$$

where i, j, k are unequal. This attractive form for the covariance structure suggests that additive logistic-normal modelling may be useful. This approach is further encouraged by the easily proved equivalence result, that, for an additive logistic-normal composition, complete subcompositional independence and covariance structure (5.1) are equivalent.

The similarity of (5.1) to (4.1) confirms that we have found an intrinsic counterpart of the doubtful extrinsic concept of basis independence. The difference lies only in the restrictions placed on $\boldsymbol{\lambda}^{(d+1)}$, the positivity in form (4.1) being relaxed to the extent that $\boldsymbol{\lambda}^{(d+1)}$ need only ensure positive-definiteness of form (5.1).

Within this framework of the $aN^d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ class for the composition $\mathbf{x}^{(d+1)}$, testing for complete subcompositional independence becomes testing the parametric hypothesis that the covariance structure is of form (5.1). Note that this hypothesis places $\frac{1}{2}d(d-1)-1$ constraints on the parameters. No exact test of the hypothesis has been found but the familiar Wilks (1938) asymptotic generalized likelihood ratio test gives a reasonable substitute. This compares

$$n\{\log(|\hat{\boldsymbol{\Sigma}}_H|/|\hat{\boldsymbol{\Sigma}}_M|) + \operatorname{trace}(\hat{\boldsymbol{\Sigma}}_H^{-1}\mathbf{V}) - d\}, \tag{5.3}$$

where $\mathbf{V}$ is the sample covariance matrix of the transformed vector $\log(\mathbf{x}^{(d)}/x_{d+1})$ and $\hat{\boldsymbol{\Sigma}}_H$ and $\hat{\boldsymbol{\Sigma}}_M$ are the maximum likelihood estimates of $\boldsymbol{\Sigma}$ under the hypothesis and model, against the appropriate upper percentile of $\chi^2\{\frac{1}{2}d(d-1)-1\}$. The estimate $\hat{\boldsymbol{\Sigma}}_M$ is simply $\mathbf{V}$ but the computation of $\hat{\boldsymbol{\Sigma}}_H$ requires a suitable numerical maximization procedure. We have used a modification of the Marquardt (1963) mixture of Newton–Raphson and steepest ascent methods, exploiting the special forms taken by $|\boldsymbol{\Sigma}_H|$, $\boldsymbol{\Sigma}_H^{-1}$ and the positive-definiteness constraint. The details are tedious and unimportant to our context: any reader interested may obtain a program in BASIC from the author.

Skye lavas. For the Skye lava data of Example 1 with $n = 32$ and $d = 9$ we obtain the value 325 for the test quantity (5.3) to be compared against upper $\chi^2(35)$ values, with consequent sound rejection of the hypothesis of complete subcompositional independence.

6. Intrinsic Analysis: Partition of Order One

6.1. *Introduction*

In their considerations of geochemical compositions geologists almost invariably concen-

constraint. The details are tedious and unimportant to our context: any reader interested may obtain a program in BASIC from the author.

Skye lavas. For the Skye lava data of Example 1 with $n = 32$ and $d = 9$ we obtain the value 325 for the test quantity (5.3) to be compared against upper $\chi^2(35)$ values, with consequent sound rejection of the hypothesis of complete subcompositional independence.

6. Intrinsic Analysis: Partition of Order One

6.1. *Introduction*

In their considerations of geochemical compositions geologists almost invariably concentrate on a few low-dimensional subcompositions, often with some amalgamation and represented in ternary diagrams such as AFM for $C(Na_2O + K_2O, Fe_2O_3, MgO)$. Such partial analyses inevitably raise questions about possible loss of information and one relevant form of analysis is to ask the extent of the dependence of the subcomposition on other aspects of the

complete composition. We suspect that an underlying reason for some of the subcompositional approaches has been the absence of suitable and readily available methodology for their undoubtedly special multivariate problems with a consequential need to project down into dimensions which can be inspected by eye. We hope that transformed multinormal modelling on the simplex will encourage full multivariate analyses of geochemical data. It should also throw some light on the validity of past choices, and the optimization of future choices, of subcompositions. More positively, with this methodology and with the concepts of intrinsic independence about to be introduced, it may be possible for the geologist to formulate his questions about subcompositions more precisely. For example, if he wishes to ask what factors affect the relative proportions of iron and manganese oxides in specimens, part of his investigation must concern the relationship of the subcomposition $C(\mathrm{FeO}+\mathrm{Fe_2O_3}, \mathrm{MnO})$ to the other aspects of the whole composition. There may, of course, be other contributory factors external to the composition such as water content. We shall see later that these could be investigated within a multivariate regression model for compositional data. Here we concentrate only on compositional factors.

As nothing more than an illustration of the analytical possibilities we consider for Example 1 the popular AFM subcomposition, actually used by Thompson, Esson and Duncan (1972); it is then natural to reorder the components, make a division of the complete vector as follows

$$(\mathrm{A} = \mathrm{Na_2O}+\mathrm{K_2O}, \mathrm{F} = \mathrm{Fe_2O_3}, \mathrm{M} = \mathrm{MgO} \mid \mathrm{MnO}, \mathrm{P_2O_5}, \mathrm{TiO_2}, \mathrm{CaO}, \mathrm{Al_2O_3}, \mathrm{SiO_2}) \qquad (6.1)$$

and thus direct interest to this partition of order one of the composition now in $\mathbb{S}^8$.

More generally then our interest is in a partition $(\mathbf{x}^{(c)}, \mathbf{x}_{(c)})$ of $\mathbf{x}^{(d+1)}$ and in the extent of interdependence of the amalgamation $\mathbf{t} = \{T(\mathbf{x}^{(c)}), T(\mathbf{x}_{(c)})\} = (t, 1-t)$ and the associated left and right subcompositions $\mathbf{s}_1 = C(\mathbf{x}^{(c)})$ and $\mathbf{s}_2 = C(\mathbf{x}_{(c)})$. We can form altogether ten independence hypotheses, falling into four types (i) $\mathbf{s}_1 \perp\!\!\!\perp \mathbf{s}_2 \mid t$; (ii) $\mathbf{s}_1 \perp\!\!\!\perp t$; (iii) $\mathbf{s}_1 \perp\!\!\!\perp (\mathbf{s}_2, t)$; (iv) $\mathbf{s}_1 \perp\!\!\!\perp \mathbf{s}_2 \perp\!\!\!\perp t$; types (i)–(iii) each have two other obvious versions. Note that, by D2, the Dirichlet class satisfies all these ten independence properties. Only type (iii), in its versions $\mathbf{s}_1 \perp\!\!\!\perp (\mathbf{s}_2, t)$ and $\mathbf{s}_2 \perp\!\!\!\perp (\mathbf{s}_1, t)$, has been previously studied, following its introduction by Connor and Mosimann (1969) under the name of neutrality. In any particular application only some subset of the ten independence hypotheses is likely to be relevant and it is clearly not practicable to consider here all possible selections of such independence hypotheses. We have therefore chosen to concentrate on six hypotheses; these, we believe, are appropriate to a large number of applications, can be fully illustrated by the application specified above, and display interesting relationships which throw light on the concept of neutrality.

6.2. *Related Concepts of Independence*

For convenience of reference the definitions of the six forms of independence are set out formally in Table 2, their implication relationships are completely summarized in the Venn diagram of Fig. 2, and a lattice of interest in our illustrative application is shown in Fig. 3. Our main purpose in the text is then to motivate the concepts, to describe modelling within which tests can be devised and to provide a rationale for the multiple-hypothesis testing situation of the lattice.

Subcompositional invariance. In the relation of a composition to its basis the concept of compositional invariance, independence of the composition $C(\mathbf{w})$ and the total size $T(\mathbf{w})$ of the basis $\mathbf{w}$ as defined in Section 4.2, plays an important role. There is a simple and useful intrinsic counterpart of this concept for subcompositions, namely subcompositional invariance, defined as independence of a subcomposition from the share of the available unit which is taken up by its components. Thus $\mathbf{s}_1$ has subcompositional invariance, denoted by $\mathscr{I}_1$, when $\mathbf{s}_1 \perp\!\!\!\perp t$. There is, of course, another possible subcompositional invariance associated with the partition, namely $\mathbf{s}_2 \perp\!\!\!\perp 1-t$ or equivalently $\mathbf{s}_2 \perp\!\!\!\perp t$, and denoted by $\mathscr{I}_2$.

TABLE 2
Some forms of independence for the partition $(\mathbf{x}^{(c)}, \mathbf{x}_{(c)})$ *of* $\mathbf{x}^{(d+1)}$

Notation	*Definition*	*Parametric hypothesis*
Subcompositional invariance		
$\mathscr{I}_1$	$C(\mathbf{x}^{(c)}) \parallel T(\mathbf{x}^{(c)})$	$\boldsymbol{\beta}_1 = \mathbf{0}$
$\mathscr{I}_2$	$C(\mathbf{x}_{(c)}) \parallel T(\mathbf{x}_{(c)})$	$\boldsymbol{\beta}_2 = \mathbf{0}$
Conditional subcompositional independence		
$\mathscr{C}$	$C(\mathbf{x}^{(c)}) \parallel C(\mathbf{x}_{(c)}) \mid T(\mathbf{x}^{(c)})$	$\boldsymbol{\Sigma}_{12} = \mathbf{0}$
Neutrality		
$\mathscr{N}_1$ (left)	$C(\mathbf{x}^{(c)}) \parallel \mathbf{x}_{(c)}$	$\boldsymbol{\beta}_1 = \mathbf{0}, \boldsymbol{\Sigma}_{12} = \mathbf{0}$
$\mathscr{N}_2$ (right)	$C(\mathbf{x}_{(c)}) \parallel \mathbf{x}^{(c)}$	$\boldsymbol{\beta}_2 = \mathbf{0}, \boldsymbol{\Sigma}_{12} = \mathbf{0}$
Partition independence		
$\mathscr{P}$	$\parallel \{C(\mathbf{x}^{(c)}), C(\mathbf{x}_{(c)}), T(\mathbf{x}^{(c)})\}$	$\boldsymbol{\beta}_1 = \mathbf{0}, \boldsymbol{\beta}_2 = \mathbf{0}, \boldsymbol{\Sigma}_{12} = \mathbf{0}$

Conditional subcompositional independence. The subcompositional invariances $\mathscr{I}_1$ and $\mathscr{I}_2$ are not concerned with the relationship of $\mathbf{s}_1$ and $\mathbf{s}_2$. A question of some interest concerning the two subcompositions $\mathbf{s}_1$ and $\mathbf{s}_2$, if, for example, $\mathscr{I}_1$ and $\mathscr{I}_2$ do not hold, is whether their dependence on each other may be only through the total amounts t and $1-t$ being assigned to each. This leads naturally to the concept of conditional subcompositional independence defined as $\mathbf{s}_1 \parallel \mathbf{s}_2 \mid t$ and denoted by $\mathscr{C}$. We note that this hypothesis is symmetric in $\mathbf{s}_1$ and $\mathbf{s}_2$ so that $\mathscr{C}$ requires no distinguishing suffices in contrast to $\mathscr{I}_1$ and $\mathscr{I}_2$.

Neutrality. Connor and Mosimann (1969) introduced the concept of neutrality which in our notation may be expressed as $C(\mathbf{x}_{(c)}) \perp\!\!\!\perp \mathbf{x}^{(c)}$. This question of whether the subcomposition on the right is independent of the entire subvector on the left was motivated by a biological problem of whether turtle scutes compete for space along the plastron during their development. The concept has been the source of a number of developments by Darroch and James (1974), Darroch and Ratcliff (1970, 1971, 1978), James (1975), James and Mosimann (1980), Mosimann (1975a, b), but much of the statistical analysis of neutrality has been hampered because until recently no parametric class of distributions on the simplex had been found rich enough to accommodate both neutrality and non-neutrality.

Since there is a one-to-one transformation between $\mathbf{x}^{(c)}$ and $(\mathbf{s}_1, t)$, neutrality as defined above can be expressed as $\mathbf{s}_2 \parallel (\mathbf{s}_1, t)$. We term this neutrality on the right and denote it by $\mathscr{N}_2$, to distinguish it from $\mathscr{N}_1$, neutrality on the left where the independence property $\mathbf{s}_1 \perp\!\!\!\perp (\mathbf{s}_2, t)$ involves the relationship of the subcomposition on the left to the entire subvector on the right. Since $\mathbf{s}_1 \parallel t$ and $\mathbf{s}_1 \parallel \mathbf{s}_2 \mid t \Leftrightarrow \mathbf{s}_1 \parallel (\mathbf{s}_2, t)$ we obtain the very simple relationships $\mathscr{I}_1 \cap \mathscr{C} = \mathscr{N}_1$, $\mathscr{I}_2 \cap \mathscr{C} = \mathscr{N}_2$. These, together with other similar relationships, are recorded in Fig. 2. Subcompositional invariance and conditional subcompositional independence are weaker forms of independence than neutrality and may thus be appropriate forms for investigation in situations where neutrality is rejected.

Partition independence. We have been discussing above various forms of independence involving $\mathbf{s}_1$, $\mathbf{s}_2$ and t, and it is natural to go to the ultimate form $\mathbf{s}_1 \perp\!\!\!\perp \mathbf{s}_2 \perp\!\!\!\perp t$. We term this partition independence, denote it by $\mathscr{P}$, and note the relation $\mathscr{I}_1 \cap \mathscr{N}_2 = \mathscr{P}$ depicted in Fig. 2

to distinguish it from $\mathscr{N}_1$, neutrality on the left where the independence property $\mathbf{s}_1 \perp\!\!\!\perp (\mathbf{s}_2, t)$ involves the relationship of the subcomposition on the left to the entire subvector on the right. Since $\mathbf{s}_1 \parallel t$ and $\mathbf{s}_1 \parallel s_2 \mid t \Leftrightarrow \mathbf{s}_1 \parallel (\mathbf{s}_2, t)$ we obtain the very simple relationships $\mathscr{I}_1 \cap \mathscr{C} = \mathscr{N}_1$, $\mathscr{I}_2 \cap \mathscr{C} = \mathscr{N}_2$. These, together with other similar relationships, are recorded in Fig. 2. Subcompositional invariance and conditional subcompositional independence are weaker forms of independence than neutrality and may thus be appropriate forms for investigation in situations where neutrality is rejected.

Partition independence. We have been discussing above various forms of independence involving $\mathbf{s}_1$, $\mathbf{s}_2$ and t, and it is natural to go to the ultimate form $\mathbf{s}_1 \perp\!\!\!\perp \mathbf{s}_2 \perp\!\!\!\perp t$. We term this partition independence, denote it by $\mathscr{P}$, and note the relation $\mathscr{I}_1 \cap \mathscr{N}_2 = \mathscr{P}$ depicted in Fig. 2

Note that for $d = 1$ all the independence properties introduced are trivially satisfied. For $d = 2$ and partition $(x_1, x_2 \mid x_3)$, satisfaction of $\mathscr{C}$, $\mathscr{I}_2$ and $\mathscr{N}_2$ is again automatic; for the partition $(x_1 \mid x_2, x_3)$ the concepts are identical with $\mathscr{C} = \mathscr{I}_2 = \mathscr{N}_2$. It is only for $d \geqq 3$ that we have a real distinction between the various concepts.

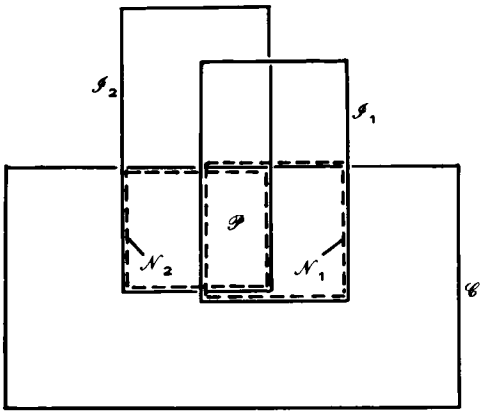

FIG. 2. Diagrammatic representation of the relationships between independence properties for a partition of order one.

6.3. *Modelling and Testing*

The problem we now face is how to model the partition $(t; \mathbf{s}_1, \mathbf{s}_2)$, and hence the original composition, in such a way that the independence hypotheses just discussed become appropriate parametric hypotheses. Since $\mathscr{C}$ involves conditioning on t it is natural to try to accommodate all the hypotheses within a conditional model for $(\mathbf{s}_1, \mathbf{s}_2 \mid t)$. For example, we can adopt additive logistic modelling for $\mathbf{s}_1$ and $\mathbf{s}_2$ with mean vector parameters dependent on t or some transform of t. With

$$\mathbf{y}_1 = a_{c-1}^{-1}(\mathbf{s}_1) = \log\{\mathbf{s}_1^{(c-1)}/s_{1c}\}, \quad \mathbf{y}_2 = a_{d-c}^{-1}(\mathbf{s}_2) = \log\{\mathbf{s}_2^{(d-c)}/s_{2,d-c+1}\}$$

and $z = \log\{t/(1-t)\}$ we can take our model M with conditional model for $(\mathbf{y}_1, \mathbf{y}_2 \mid z)$ of the following form:

$$N^{d-1}\left\{\begin{bmatrix}\boldsymbol{\alpha}_1 + \boldsymbol{\beta}_1 z \\ \boldsymbol{\alpha}_2 + \boldsymbol{\beta}_2 z\end{bmatrix}, \begin{bmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22}\end{bmatrix}\right. \tag{6.2}$$

All the independence hypotheses considered are then easily identified with constraints on the parameters $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\Sigma}_{12}$. For example, $\mathscr{I}_2$ requires $\mathbf{y}_2 \perp\!\!\!\perp z$ and so has parametric counterpart $\boldsymbol{\beta}_2 = \mathbf{0}$; and $\mathscr{N}_2$ requires the further condition $\mathbf{y}_1 \perp\!\!\!\perp \mathbf{y}_2 \mid z$ or $\boldsymbol{\Sigma}_{12} = \mathbf{0}$ and so is identical to the parametric hypotheses $\boldsymbol{\beta}_2 = \mathbf{0}$, $\boldsymbol{\Sigma}_{12} = \mathbf{0}$. All these parametric counterparts are listed for convenience beside the definitions in Table 2.

Since the hypotheses under test impose linear constraints on mean vector and simple restrictions on covariance matrices the generalized likelihood ratio test statistic again takes the form (5.3) with approximate critical values given through asymptotic theory as upper χ^2 percentiles with appropriate degrees of freedom q_H for hypotheses H. The derivation of $\hat{\boldsymbol{\Sigma}}_H$ and q_H for the various hypotheses and of $\hat{\boldsymbol{\Sigma}}_M$ is routine; for easy reference we provide the computational forms in Table 3.

6.4. *Testing a Lattice of Hypotheses: An Application*

If only one of the independence hypotheses already discussed is under scrutiny then the appropriate test procedure set out in Section 6.3 applies. If, however, we have under investigation a number of the hypotheses then we must consider more carefully our strategy, such as order of testing. In the lattice of hypotheses set out in Fig. 3 for the partition (6.1) of the Skye lava compositions, the model is at the highest level with hypotheses at deeper levels corresponding to more and more constraints on the parameters. Viewed from the bottom of

TABLE 3

Maximum likelihood estimates of $\boldsymbol{\Sigma}$ *associated with independence hypotheses*

Hypothesis H or model M	*Maximum likelihood estimate of* $\boldsymbol{\Sigma}$ *with submatrices in the order* $\boldsymbol{\Sigma}_{11}, \boldsymbol{\Sigma}_{12}, \boldsymbol{\Sigma}_{22}$	*Degrees of freedom* q_H
M	$\hat{\boldsymbol{\Sigma}}_{11}, \hat{\boldsymbol{\Sigma}}_{12}, \hat{\boldsymbol{\Sigma}}_{22}$	
$\mathscr{I}_1$	$\mathbf{S}_{11}, \hat{\boldsymbol{\Sigma}}_{12}, \hat{\boldsymbol{\Sigma}}_{22}$	$c-1$
$\mathscr{I}_2$	$\hat{\boldsymbol{\Sigma}}_{11}, \hat{\boldsymbol{\Sigma}}_{12}, \mathbf{S}_{22}$	$d-c$
$\mathscr{C}$	$\hat{\boldsymbol{\Sigma}}_{11}, \mathbf{0}, \hat{\boldsymbol{\Sigma}}_{22}$	$(c-1)(d-c)$
$\mathscr{N}_1$	$\mathbf{S}_{11}, \mathbf{0}, \hat{\boldsymbol{\Sigma}}_{22}$	$(c-1)(d-c+1)$
$\mathscr{N}_2$	$\hat{\boldsymbol{\Sigma}}_{11}, \mathbf{0}, \mathbf{S}_{22}$	$c(d-c)$
$\mathscr{P}$	$\mathbf{S}_{11}, \mathbf{0}, \mathbf{S}_{22}$	$c(d-c)+c-1$

Required matrix computations

$$nS_{ij} = \sum_{r=1}^{n} (y_{ir}-\bar{y}_i)(y_{jr}-\bar{y}_j) \quad (i,j=1,2); \quad nS_{zz} = \sum_{r=1}^{n} (z_r-\bar{z})^2;$$

$$nS_{iz} = \sum_{r=1}^{n} (y_{ir}-\bar{y}_i)(z_r-\bar{z}); \quad \hat{\beta}_i = S_{iz}/S_{zz}, (i=1,2);$$

$$\hat{\Sigma}_{ij} = S_{ij}-\hat{\beta}_i\hat{\beta}_j S_{zz}, (i,j=1,2).$$

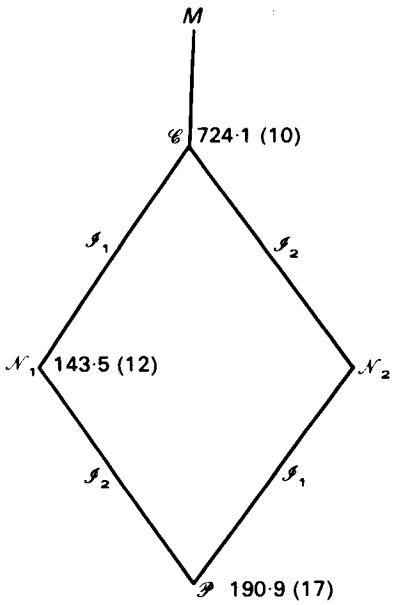

FIG. 3. Lattice for Skye lava analysis showing values of test statistics with associated degrees of freedom in brackets.

the lattice the hypothesis $\mathscr{P}$ is the simplest explanation of the relationship of $C(\mathbf{x}^{(c)})$, $C(\mathbf{x}_{(c)})$ and $(\mathbf{x}^{(c)})$, namely mutual independence. As we move up the lattice, for example to $\mathscr{N}_1$, we have to introduce more parameters, namely $\boldsymbol{\beta}_2$, to provide an explanation of the pattern of variability,

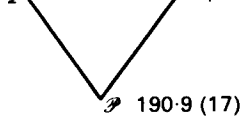

FIG. 3. Lattice for Skye lava analysis showing values of test statistics with associated degrees of freedom in brackets.

the lattice the hypothesis $\mathscr{P}$ is the simplest explanation of the relationship of $C(\mathbf{x}^{(c)})$, $C(\mathbf{x}_{(c)})$ and $T(\mathbf{x}^{(c)})$, namely mutual independence. As we move up the lattice, for example to $\mathscr{N}_1$, we have to introduce more parameters, namely $\boldsymbol{\beta}_2$, to provide an explanation of the pattern of variability, and to $\mathscr{C}$ further parameters, namely $\boldsymbol{\beta}_1$.

For such multiple-hypothesis testing, a sensible approach is to adopt the simplicity postulate of Jeffreys (1961, p. 47): in order to move from a simple explanation, such as $\mathscr{P}$, to a more complex explanation, such as $\mathscr{N}_1$, we require to reject the simpler explanation through

an appropriate significance test. In other words, to justify the introduction of more parameters, we require a mandate, provided by significant rejection, to allow us to move to a higher level in the lattice. Thus our procedure would involve the following steps. First test $\mathscr{P}$ within M. If we cannot reject $\mathscr{P}$ then there is nothing to justify moving from the simple explanation $\mathscr{P}$. If we reject $\mathscr{P}$ then we move up to the next level, testing each of $\mathscr{N}_1$ and $\mathscr{N}_2$ within M. If we cannot reject both then we have a feasible explanation at this level. If we reject both then we move to a test of $\mathscr{C}$ within M, and so on. Note that the tests are all of a hypothesis H within M and the mechanism of these tests has already been described in Section 6.3.

For our geochemical partition the values of the test statistics with their bracketed degrees of freedom are shown at the appropriate nodes of the lattice. All the hypotheses of the lattice are rejected at significance levels well below 0·1 per cent. However we care to interpret the lattice, the C(A, F, M) subcomposition has clearly neither subcompositional invariance nor is it conditionally independent of the complementary subcomposition. Further analysis, not reported here, shows that it is also not (absolutely) independent, defined as $\mathbf{s}_1 \parallel \mathbf{s}_2$, of the complementary subcomposition. Thus any analysis of AFM which subsumes that this subcomposition is independent of other apsects of the composition is surely suspect.

7. Further Aspects of Intrinsic Analysis

7.1. *Partial Subcompositional Independence*

In the extrinsic approach to compositional structure some geologists, for example Sarmanov and Vistelius (1959), consider forms of partial basis independence under such terms as *concretionary* and *metasomatic*. These have satisfactory intrinsic counterparts whose form we can now indicate briefly in terms of a partition $(\mathbf{x}^{(c)}, \mathbf{x}_{(c)})$ or $(t; \mathbf{s}_1, \mathbf{s}_2)$ of order one.

Definition: partial subcompositional independence restricted by $\mathbf{x}^{(c)}$. A composition $\mathbf{x}^{(d+1)}$ has partial subcompositional independence restricted by $\mathbf{x}^{(c)}$ if $\mathbf{s}_1 \parallel \mathbf{s}_2$ and $\mathbf{s}_2$ has complete subcompositional independence within $\mathbb{S}^{d-c}$.

Since the amalgamation $\mathbf{t} = (t, 1-t)$ is not involved in the definition we can investigate such partial subcompositional independence within a model for the joint distribution of $(\mathbf{s}_1, \mathbf{s}_2)$. Taking this to be of transformed normal form $\{a_{c-1}^{-1}(\mathbf{s}_1), a_{d-c}^{-1}(\mathbf{s}_2)\}$ and hence with covariance matrix

$$\mathbf{\Sigma} = \operatorname{cov}\{\log(x_i/x_c)\,(i = 1, \ldots, c-1); \log(x_{c+i}/x_{d+1})\,(i = 1, \ldots, d-c)\} \tag{7.1}$$

we can specify partial subcompositional independence as the parametric hypothesis

$$\mathbf{\Sigma}_{12} = \mathbf{0}, \quad \mathbf{\Sigma}_{22} = \operatorname{diag}(\lambda_{c+1}, \ldots, \lambda_d) + \lambda_{d+1}\,\mathbf{U}_{d-c} \tag{7.2}$$

in term of the obvious partitioning of $\mathbf{\Sigma}$. Such a formulation brings this form of independence within the scope of the test procedures developed in Section 6. Moreover, the fact that partial subcompositional independence is seen as the conjunction of two less stringent hypotheses, unconditional subcompositional independence $\mathbf{s}_1 \perp\!\!\!\perp \mathbf{s}_2$ and complete subcompositional independence of $\mathbf{s}_2$ within $\mathbb{S}^{d-c}$, open up another means of probing compositional structure through a lattice approach.

7.2. *Independence up to Level c*

There are a number of situations, where a specific ordering of the $d+1$ components has been made and already embodied in $\mathbf{x}^{(d+1)}$, and where interest is in considering independence properties for partitions of order one at a sequence of levels c. Since we consider here only independence in the form $\mathscr{C}$, $\mathscr{I}_2$ and $\mathscr{N}_2$ we drop the suffix 2 to allow us to emphasize the level c at which division has been made. Thus $\mathscr{C}_c$, $\mathscr{I}_c$, $\mathscr{N}_c$ denote $\mathscr{C}, \mathscr{I}_2, \mathscr{N}_2$ at level c. We recall the basic relation $\mathscr{C}_c \cap \mathscr{I}_c = \mathscr{N}_c$, and, for any one of these hypotheses, say H_c, define the corresponding concept H^c up to level c as follows.

Definition: independence property up to level c. A composition $\mathbf{x}^{(d+1)}$ has independence property H up to level c if H_k holds for $k = 1, ..., c$.

It follows from the relationship that $\mathscr{C}^c \cap \mathscr{I}^c = \mathscr{N}^c$. For the special case when $c = d-1$ (or equivalently d) we use the term *complete*.

Definition: complete independence property. A composition $\mathbf{x}^{(d+1)}$ possesses the complete independence property H if H^{d-1} holds.

Thus, for example, complete neutrality (Connor and Mosimann, 1969) requires $C(\mathbf{x}_{(c)}) \parallel \mathbf{x}^{(c)}$ for $c = 1, ..., d-1$. The investigation of neutrality at different levels is best pursued in terms of the multiplicative logistic transformation. This approach has been adopted by Aitchison (1981b) to provide a suitable parametric statistical framework within which to test $\mathscr{N}_c \mathscr{N}^c$ and lattices of hypotheses involving these. Adopting a $mN^d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ model we see that the hypotheses $\mathscr{N}_c \mathscr{N}^c$ and $\mathscr{N}^{d-1}$ correspond to the following covariance matrix structures

$$\begin{bmatrix} \boldsymbol{\Sigma}_{11} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{22} \end{bmatrix}, \quad \begin{bmatrix} \operatorname{diag}(\sigma_{11}, ..., \sigma_{cc}) & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{22} \end{bmatrix}, \quad \operatorname{diag}(\sigma_{11}, ..., \sigma_{dd}), \tag{7.3}$$

where $\boldsymbol{\Sigma}_{11}$ is of order $c \times c$. The numbers of constraints imposed by the three hypotheses are $c(d-c)$, $c\{d-\frac{1}{2}(c+1)\}$ and $\frac{1}{2}d(d-1)$. If $\mathbf{V}$ is the estimated covariance matrix associated with d-dimensional vectors $\mathbf{y}_r$ $(r = 1, ..., n)$ defined in the m_d entry of Table 1 then the test statistics (Aitchison, 1981b) associated with $\mathscr{N}_c \mathscr{N}^c$ and $\mathscr{N}^{d-1}$ are again of form (5.3) with

$$\begin{aligned} &|\hat{\boldsymbol{\Sigma}}_M| = |\mathbf{V}|, \quad |\hat{\boldsymbol{\Sigma}}_{\mathscr{N}_c}| = |\mathbf{V}_{11}| \cdot |\mathbf{V}_{22}|, \\ &|\hat{\boldsymbol{\Sigma}}_{\mathscr{N}^c}| = v_{11} \ldots v_{cc} |\mathbf{V}_{22}|, \quad |\hat{\boldsymbol{\Sigma}}_{\mathscr{N}^{d-1}}| = v_{11} \ldots v_{dd}, \end{aligned} \tag{7.4}$$

where $\mathbf{V}_{11}, \mathbf{V}_{22}$ are obvious submatrices of $\mathbf{V}$ in a $(c, d-c)$ partitioning and v_{ij} is the (i,j)th element of $\mathbf{V}$. Note that in all these tests the term trace $(\hat{\boldsymbol{\Sigma}}_H^{-1} \hat{\boldsymbol{\Sigma}}_M) = d$ so that the test statistic reduces to $n \log(|\hat{\boldsymbol{\Sigma}}_H| / |\hat{\boldsymbol{\Sigma}}_M|)$.

Since $\mathscr{I}_c$ and $\mathscr{N}_c$ $(c > 1)$ are quite distinct hypotheses we might expect $\mathscr{I}^c$ and $\mathscr{N}^c$ to be distinct and, since $\mathscr{N}^c \subset \mathscr{I}^c$, to be able to devise a model for which $\mathscr{I}^c$ holds but $\mathscr{N}^c$ does not. We have failed to produce such a model and are beginning to conjecture that, within the framework of transformed normal modelling, $\mathscr{I}^c \equiv \mathscr{N}^c$, though so far we have failed to prove the conjecture.

That there is a distinction between $\mathscr{C}^c$ and $\mathscr{N}^c$ can be readily seen for the case $d = 3$. Since $\mathscr{C}_1$ and $\mathscr{C}_3$ are trivially satisfied for any compositional distribution, model (6.2) with $c = 2$ and $\sigma_{12} = 0$ supports $\mathscr{C}_2$ and hence complete conditional subcompositional independence, whereas $\mathscr{N}_2$ does not hold unless $\beta_2 = 0$. We have not so far found any practical problem to which the idea of $\mathscr{C}^c$ seems relevant and have not therefore pursued the modelling problem further.

Skye lavas. From the strong rejection of complete subcompositional independence and the neutrality hypotheses $\mathscr{N}_2$ there can be little surprise in discovering that tests of neutrality associated with an ordering such as (6.1) of the entire compositional vector lead to rejections. Simply as an illustrative example for numerical comparison therefore we show in Table 4

TABLE 4
Test results for neutrality hypotheses for Skye lavas

Level	*Test statistic*	*Degrees of freedom*
1	58·0	7
2	135·6	13
3	182·7	18

associated with an ordering such as (6.1) of the entire compositional vector lead to rejections. Simply as an illustrative example for numerical comparison therefore we show in Table 4

TABLE 4
Test results for neutrality hypotheses for Skye lavas

Level	*Test statistic*	*Degrees of freedom*
1	58·0	7
2	135·6	13
3	182·7	18
4	227·0	22
5	275·0	25
6	283·6	27
7	290·0	28

values of the test statistics, as described above for testing $\mathcal{N}^c$ up to all possible levels c for this ordering together with the corresponding degrees of freedom at each of the seven levels. Since the comparison is against upper chi-squared values at the degrees of freedom shown, the neutrality hypotheses up to all levels for this ordering are strongly rejected. To those who regard hypothesis-testing as a means towards arriving at a model for subsequent analyses we reiterate the important fact that rejection of all these hypotheses still leaves transformed normal models on the simplex as possible describers of patterns of variability of non-neutral compositional data.

7.3. *Compositional Regression Models*

On finding a subcomposition $\mathbf{s}_1$, such as AFM, dependent on complementary aspects t and $\mathbf{s}_2$ of the complete composition we may wish to assess the conditional distribution $p(\mathbf{s}_1 \mid t, \mathbf{s}_2)$. This aspect of estimation, essentially regression analysis in transformed normal modelling, has already been illustrated for sediment compositions by Aitchison and Shen (1980) and need not be detailed here. Rather we present briefly examples where we may wish to explore the dependence of a composition $\mathbf{x}^{(d+1)} \in \mathbb{S}^d$ on concomitant information z.

Arctic lake sediments. In Example 3 for each Arctic lake composition the associated depth z is provided. There is, through the transformed normality approach, an obvious way of modelling to allow investigation of the dependence of composition on depth, namely to take $p(\mathbf{x}^{(d+1)} \mid z) = f N^d(\mathbf{g}(z), \boldsymbol{\Sigma})$, where the regression function $\mathbf{g}(z)$ can be investigated in the usual multivariate regression form. In our model M we have taken f to be a_2 and $\mathbf{g}(z)$ to include terms in z, z^2, $\log z$ and $(\log z)^2$. We have then worked through a lattice of increasingly complex hypotheses along the lines of Section 6.4, and found that linear regression is certainly rejected, but that hypotheses of the form $\mathbf{g}(z) = \boldsymbol{\alpha} + \boldsymbol{\beta} \log z$, or quadratic regression $\mathbf{g}(z) = \boldsymbol{\alpha} + \boldsymbol{\beta} z + \boldsymbol{\gamma} z^2$ are equally good fits and cannot be rejected. Moreover the residuals based on either of these fitted regressions pass the complete battery of multivariate normal tests.

Household budgets. As another illustration of the simplicity of regression techniques we might extend the model of Section 4.2 to include the possibility of compositional dependence on household size, for example with the regression function of the form

$$\boldsymbol{\alpha} + \boldsymbol{\beta} \log(\text{total expenditure}) + \boldsymbol{\gamma} \log(\text{household size}).$$

If we then investigate the lattice with nodes at $\boldsymbol{\beta} = \mathbf{0}$, $\boldsymbol{\gamma} = \mathbf{0}$, at $\boldsymbol{\beta} = \mathbf{0}$ and at $\boldsymbol{\gamma} = \mathbf{0}$ we find that the hypothesis $\boldsymbol{\gamma} = \mathbf{0}$ is the only one that cannot be rejected. Moreover, fitting of this accepted regression function leaves residuals which survive the battery of goodness-of-fit tests.

7.4. *The Problem of Zero Components*

Throughout the paper attention has been confined to the strictly positive simplex. The reason is the obvious one that we cannot take logarithms of zero. And yet zero components do occur in a number of applications, for example, when a household spends nothing on the commodity group "tobacco and alcohol" or a rock specimen contains "no trace" of a particular mineral. In the absence of a one-to-one monotonic transformation between the real line and its non-negative subset the problem of zeros is unlikely ever to be satisfactorily resolved. A similar problem occurs in lognormal modelling and, as there, *ad hoc* solutions naturally depend on the frequency and nature of the zeros.

If there are only a few zeros of the no-trace type then replacement by positive values smaller than the smallest traceable amounts will allow an analysis. In such circumstances it will always be wise to perform a sensitivity analysis to determine the effect that different zero replacement values have on the conclusions of the analysis. For example, in the investigation of compositional invariance in glacial tills in Section 4.2 we replaced 14 zero proportions by 0·0005 obtaining the value 3·05 for the test statistic in the $aN^3(\boldsymbol{\alpha} + \boldsymbol{\beta} \log t, \boldsymbol{\Sigma})$ modelling. For other replacement values 0·001, 0·00025, 0·00001 and 0·000001 the values of the test statistic are 3·93,

2·46, 2·01 and 1·54 all leading to the same conclusion of no evidence against compositional invariance at the 5 per cent significance level.

If there is a moderate number of real zeros it may be worth considering the device of three-parameter lognormal modelling (Aitchison and Brown, 1957, p. 14), whereby a constant, either known or to be estimated, is added to every observation. One compositional counterpart would be to apply the transformations, not to $\mathbf{x}^{(d+1)}$, but to $C(\mathbf{x}^{(d+1)}+\boldsymbol{\tau}^{(d+1)})$ where $\boldsymbol{\tau}^{(d+1)}$ is either chosen or estimated. For example, for the case $d = 1$ and an additive logistic model we are considering the model with $\log\{(x+\tau_1)/(1+\tau_2-x)$ of $N^1(\mu,\sigma^2)$ form, which is a four-parameter lognormal model of Johnson (1949). Clearly if $\boldsymbol{\tau}^{(d+1)}$ has to be estimated there are substantial estimation and interpretation problems even for small d.

If there is a substantial number of zeros mostly in a few components and if amalgamations of components are ruled out, then some form of conditional modelling separating out the zero may be possible. For example, if the zeros are confined to the last component then the conditional distribution of $C(\mathbf{x}^{(d)})$ on x_{d+1} might be modelled by taking $\log(\mathbf{x}^{(d-1)}/x_d)$ to be $N^{d-1}(\boldsymbol{\alpha}+\boldsymbol{\beta}x_{d+1},\boldsymbol{\Sigma})$ with the marginal distribution of x_{d+1} having a mass probability at zero and $\log\{x_{d+1}/(1-x_{d+1})\}$ following $N^1(\mu,\sigma^2)$ for $x_{d+1}>0$.

7.5. *Partitions of Higher Order*

For a partition of order one we saw there are ten different independence hypotheses and that careful selection of hypotheses relevant to the practical problem is of primary importance to a successful analysis. The choice of relevant hypotheses for a higher order partition $(\mathbf{t};\mathbf{s}_1,\ldots,\mathbf{s}_{k+1})$ is even more crucial and we have no space to discuss it at length here. A brief look at a partition of order 2 should, however, indicate the potentialities of transformed normal modelling.

Suppose that for a partition $(\mathbf{t};\mathbf{s}_1,\mathbf{s}_2,\mathbf{s}_3)$ of order 2 we wish to investigate the extent of subcompositional invariance with respect to the sums t_1,t_2,t_3 and also whether the amalgamation $\mathbf{t}^{(3)}=(t_1,t_2,t_3)$ displays complete neutrality. If we model in terms of the transformed partition $(\mathbf{z};\mathbf{y}_1,\mathbf{y}_2,\mathbf{y}_3)$ of $(m_k;a_{d_1},a_{d_2},a_{d_3})$ type we might use conditional modelling $p(\mathbf{z})\,p(\mathbf{y}_1,\mathbf{y}_2,\mathbf{y}_3\,|\,\mathbf{z})$ with $p(\mathbf{y}_1,\mathbf{y}_2,\mathbf{y}_3\,|\,\mathbf{z})$ of multinormal form

$$N^{d-1}\left\{\begin{bmatrix}\boldsymbol{\alpha}_1+\boldsymbol{\beta}_1\mathbf{z} & \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} & \boldsymbol{\Sigma}_{13}\\ \boldsymbol{\alpha}_2+\boldsymbol{\beta}_2\mathbf{z} & \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} & \boldsymbol{\Sigma}_{23}\\ \boldsymbol{\alpha}_3+\boldsymbol{\beta}_3\mathbf{z} & \boldsymbol{\Sigma}_{31} & \boldsymbol{\Sigma}_{32} & \boldsymbol{\Sigma}_{33}\end{bmatrix}\right\} \tag{7.5}$$

and $p(\mathbf{z})$ of $N^2(\boldsymbol{\gamma},\boldsymbol{\Omega})$ form. It must now be clear that there could be a large number of hypotheses of interest.

As an example of a simple lattice approach we refer to Fig. 4 where forms of hypotheses of total subcompositional invariance, $(\mathbf{s}_1,\mathbf{s}_2,\mathbf{s}_3)\perp\!\!\!\perp\mathbf{t}$ or parametrically $\boldsymbol{\beta}_h=\mathbf{0}$ $(h=1,2,3)$, and of complete neutrality of $\mathbf{t}^{(3)}$, namely $\omega_{12}=0$, are brought together. The testing of such a lattice is straightforward following the lines of Section 6.4. It is also clear that the total subcompositional invariance hypothesis could be broken into interesting hypotheses such as $\boldsymbol{\beta}_1=\mathbf{0}$ at a higher level of the lattice. Note that in the selection of the transformation we used m for the amalgamation since interest was in complete neutrality. Had an objective been to study neutrality within the subcompositions then m transformations could have replaced the a transformations actually used in the modelling.

total subcompositional invariance, $(\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3) \perp\!\!\!\perp \mathbf{t}$ or parametrically $\boldsymbol{\beta}_h = \mathbf{0}$ $(h = 1, 2, 3)$, and of complete neutrality of $\mathbf{t}^{(3)}$, namely $\omega_{12} = 0$, are brought together. The testing of such a lattice is straightforward following the lines of Section 6.4. It is also clear that the total subcompositional invariance hypothesis could be broken into interesting hypotheses such as $\boldsymbol{\beta}_1 = \mathbf{0}$ at a higher level of the lattice. Note that in the selection of the transformation we used m for the amalgamation since interest was in complete neutrality. Had an objective been to study neutrality within the subcompositions then m transformations could have replaced the a transformations actually used in the modelling.

8. Discussion

There remain many loose ends to our transformed normal package. We hope that discussion in the Society tonight will reveal many statistical fingers anxious to tie up, to add to, even to repack, the package and to address it for delivery to new areas of application. The

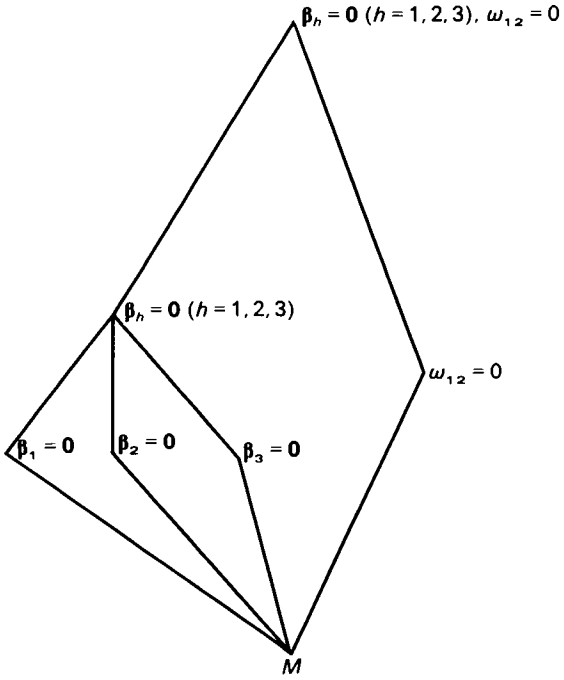

FIG. 4. Lattice for testing subcompositional invariance and neutrality for subcompositional shares.

following collection of random thoughts on the current state of the package is little more than an attempt to draw attention to topics of personal interest.

(i) We have dealt only with one-way compositions. There are problems where the components fall naturally into a two-way classification. It would be of interest to discuss problems of this type and the means of analysing them.

(ii) Of our three elementary transformations in Table 1 we have used only a_d and m_d. Are there any applications where h_d is essential? What other transformations between $\mathbb{R}^d$ and $\mathbb{S}^d$ might find applications? To what extent will it be necessary to widen the class of transformations, as suggested by Aitchison and Shen (1980), through the Box and Cox (1964) approach, with $y_i = \{(x_i/x_{d+1})^\lambda - 1\}/\lambda$ $(i = 1, \ldots, d)$ and λ being estimated from the compositional data?

(iii) Although the immediate relationship to multivariate normality usually ensures the carry-over of existing techniques, such as discriminant analysis, to compositional data some care is needed to check the validity of this transfer. For example, reduction of the compositional dimension through the use of principal components based on $\boldsymbol{\Sigma} = \mathrm{cov}\{\log(\mathbf{x}^{(d)}/x_{d+1})\}$ might seem a hopeful technique until it is realized that trace $(\boldsymbol{\Sigma})$ is not invariant under a permutation of the components $x_1, \ldots, x_{d+1}$. A substantial modification to standard principal component analysis is required to restore the desirable invariance property.

(iv) One embarrassment of the transformed normal approach is the galaxy of possible models it offers. For example, in our discussion of right neutrality $\mathcal{N}_c$ in Section 6.3 either the model (6.2) with $\boldsymbol{\beta}_2 = \mathbf{0}$, $\boldsymbol{\Sigma}_{12} = \mathbf{0}$ or the model $mN^d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ of Section 7.2 with $\boldsymbol{\Sigma}_{12} = \mathbf{0}$ could be used. Although the problem of choice between models here is no different from similar problems in other areas of statistics, tests of these separate classes could prove troublesome because of the dimension of the parameter space. One possible line of investigation might be the examination of how close the models are in the same way as Aitchison and Shen (1980) considered the closeness of logistic-normal and Dirichlet classes.

(v) Conjecture about the potential of the transformed normal approach to the analysis of the structure of geological compositions is a fascinating subject. Geologists, for example Chayes (1971), assure us that the study of correlations in compositions is essential to their understanding and yet it appears difficult to pinpoint their precise hypotheses of interest. There seems little doubt that the package can play a useful rôle in descriptive geostatistics, such as in classification, but can the fundamental hypotheses of compositional structure now be specified within the concepts of this paper?

(vi) Compositional data obviously occur in areas other than the geological and economic applications cited here; for example, in developmental biology if we wish to explore how the shape (composition) of a linear organism relates to size, the model used for the study of compositional invariance will obviously play a rôle. There are also problems with simplex sample spaces where the data are not compositions; for example, probabilistic data in $\mathbb{S}^d$ occur in the analysis of subjective performance of inferential tasks (Aitchison, 1981c). More complex product sample spaces, such as $\mathbb{S}^d \times \mathbb{R}^c$ or $\mathbb{S}^d \times \mathbb{P}^c$, also arise, as in medical diagnosis (Aitchison and Begg, 1976), and succumb to the transformed normal technique.

(vii) There are still many distributional problems to be resolved. For example, although we have in mN^d a model for the investigation of complete right neutrality and in a separate mN^d model applied to the reversal of the vector $\mathbf{x}^{(d+1)}$ a means of investigating left neutrality, we have been unable to find a class of models which will accommodate both forms of neutrality as parametric hypotheses and will also have non-neutral members. Thus the battle of the statistical knights who search for the holy grail of a parametric class which will include the highly structured Dirichlet distributions and all forms of dependent distributions, is obviously not over. We hope, however, that transformed normal distributions may sharpen their lances and encourage the search.

ACKNOWLEDGEMENTS

I had the good fortune of meeting at conferences in Trieste and Sydney during 1980 three seasoned campaigners in the simplex, J. N. Darroch, W. Kruskal and J. E. Mosimann. The first draft of this paper owed much to discussions with them, with other participants at these conferences and with my Hong Kong colleagues during the formative period of some of the ideas. As usual the final version has been greatly improved by the diverse, but always penetrating and constructive, comments of four referees. I am also grateful to Colin Greenfield, Commissioner of Census and Statistics in Hong Kong, for making available the pilot household budget data.

REFERENCES

AITCHISON, J. (1981a). A new approach to null correlations of proportions. *J. Math. Geol.*, **13**, 175–189.

—— (1981b). Distributions on the simplex for the analysis of neutrality. In *Statistical Distributions in Scientific Work* (C. Taillie, G. P. Patil and B. Baldessari, eds), vol. 4, pp. 147–156. Dordrecht, Holland: D. Reidel Publishing Company.

—— (1981c). Some distribution theory related to the analysis of subjective performance in inferential tasks. In *Statistical Distributions in Scientific Work* (C. Taillie, G. P. Patil and B. Baldessari, eds), vol. 5, pp. 363–386. Dordrecht, Holland: D. Reidel Publishing Company.

AITCHISON, J. and BEGG, C. B. (1976). Statistical diagnosis when the cases are not classified with certainty. *Biometrika*, **63**, 1–12.

AITCHISON, J. and BROWN, J. A. C. (1957). *The Lognormal Distribution*. Cambridge University Press.

AITCHISON, J. and SHEN, S. M. (1980). Logistic-normal distributions: some properties and uses. *Biometrika*, **67**, 261–272.

ANDERSON, J. A. (1972). Separate sample logistic discrimination. *Biometrika*, **59**, 19–35.

ANDREWS, D. F., GNANADESIKAN, R. and WARNER, J. L. (1973). Methods for assessing multivariate normality. In *Multivariate Analysis* III (P. R. Krishnaiah, ed) pp. 95–116. New York: Academic Press.

Company.
—— (1981c). Some distribution theory related to the analysis of subjective performance in inferential tasks. In *Statistical Distributions in Scientific Work* (C. Taillie, G. P. Patil and B. Baldessari, eds), vol. 5, pp. 363–386. Dordrecht, Holland: D. Reidel Publishing Company.
AITCHISON, J. and BEGG, C. B. (1976). Statistical diagnosis when the cases are not classified with certainty. *Biometrika*, **63**, 1–12.
AITCHISON, J. and BROWN, J. A. C. (1957). *The Lognormal Distribution*. Cambridge University Press.
AITCHISON, J. and SHEN, S. M. (1980). Logistic-normal distributions: some properties and uses. *Biometrika*, **67**, 261–272.
ANDERSON, J. A. (1972). Separate sample logistic discrimination. *Biometrika*, **59**, 19–35.
ANDREWS, D. F., GNANADESIKAN, R. and WARNER, J. L. (1973). Methods for assessing multivariate normality. In *Multivariate Analysis* III (P. R. Krishnaiah, ed) pp. 95–116. New York: Academic Press.
BOX, G. E. P and COX, D. R. (1964). The analysis of transformations (with discussion). *J. R. Statist. Soc.* B, **26**, 211–252.
BROWN, A. and DEATON, A. S. (1972). Models of consumer behaviour. *Econ. J.*, **82**, 177–268.
CHAYES, F. (1960). On correlations between variables of constant sum. *J. Geophys. Res.*, **65**, 4185–4193.
—— (1962). Numerical correlation and petrographic variation. *J. Geol.*, **70**, 440–452.
—— (1971). *Ratio Correlation*. University of Chicago Press.

CHAYES, F. and KRUSKAL, W. (1966). An approximate statistical test for correlations between proportions. *J. Geol.*, **74**, 692–702.
COAKLEY, J. P. and RUST, B. R. (1968). Sedimentation in an Arctic lake. *J. Sedimentary Petrology*, **38**, 1290–1300.
CONNOR, R. J. and MOSIMANN, J. E. (1969). Concepts of independence for proportions with a generalization of the Dirichlet distribution. *J. Amer. Statist. Assoc.*, **64**, 194–206.
COX, D. R. (1966). Some procedures associated with the logistic qualitative response curve. In *Research Papers in Statistics: Festschrift for J. Neyman* (F. N. David, ed), pp. 57–71. New York: Wiley.
—— (1970). *The Analysis of Binary Data*. London: Methuen.
DARROCH, J. N. (1969). Null correlations for proportions. *J. Math. Geol.*, **3**, 467–483.
DARROCH, J. N. and JAMES, I. R. (1974). F-independence and null correlations of continuous, bounded-sum, positive variables. *J. R. Statist. Soc.* B, **36**, 467–483.
DARROCH, J. N. and RATCLIFF, D. (1970). Null correlations for proportions II. *J. Math. Geol.*, **2**, 307–312.
—— (1971). A characterization of the Dirichlet distribution. *J. Amer. Statist. Assoc.*, **66**, 641–643.
—— (1978). No-association of proportions. *J. Math. Geol.*, **10**, 361–368.
DAWID, A. P. (1979). Conditional independence in statistical theory (with discussion). *J. R. Statist. Soc.* B, **41**, 1–31.
DAY, N. E. and KERRIDGE, D. F. (1967). A general maximum likelihood discriminant. *Biometrics*, **23**, 313–323.
DEATON, A. S. (1978). Specification and testing in applied demand analysis. *Econ. J.*, **88**, 524–536.
DEATON, A. S. and MUELLBAUER, J. (1980). *Economics and Consumer Behavior*. New York: Cambridge University Press.
GNANADESIKAN, R. and KETTENRING, J. R. (1972). Robust estimates, residuals and outlier detection with multiresponse data. *Biometrics*, **28**, 81–124.
HOUTHAKKER, H. S. (1960). Additive preferences. *Econometrica*, **28**, 244–254.
JAMES, I. R. (1975). Multivariate distributions which have beta conditional distributions. *J. Amer. Statist. Assoc.*, **70**, 681–684.
—— (1981). Distributions associated with neutrality properties for random proportions. In *Statistical Distributions in Scientific Work* (C. Taillie, G. P. Patil and B. Baldessari, eds), vol. 4, pp. 125–136. Dordrecht, Holland, D. Reidel Publishing Company.
JAMES, I. R. and MOSIMANN, J. E. (1980). A new characterization of the Dirichlet distribution through neutrality. *Ann. Statist.*, **8**, 183–189.
JEFFREYS, H. (1961). *Theory of Probability*. Oxford University Press.
JOHNSON, N. L. (1949). Systems of frequency curves generated by methods of translation. *Biometrika*, **36**, 149–176.
JOHNSON, N. L. and KOTZ, S. (1972). *Distributions in Statistics. Continuous Multivariate Distributions*. Boston: Houghton Mifflin.
KAISER, R. F. (1962). Composition and origin of glacial till Mexico and Kasoog quadrangles, New York. *J. Sedimentary Petrology*, **32**, 502–513.
LEONARD, T. (1973). A Bayesian method for histograms. *Biometrika*, **59**, 581–589.
LESER, C. E. V. (1976). Income, household size and price changes 1953–1973. *Oxford Bull. Econ. Statist.*, **38**, 1–10.
MCALISTER, D. (1879). The law of the geometric mean. *Proc. R. Soc.*, **29**, 367.
MARQUARDT, D. W. (1963). An algorithm for least-squares estimation of non-linear parameters. *J. SIAM*, **11**, 431–441.
MORRISON, D. F. (1976). *Multivariate Statistical Methods*. New York: McGraw-Hill.
MOSIMANN, J. E. (1962). On the compound multinomial distribution, the multivariate β-distribution and correlations among proportions. *Biometrika*, **49**, 65–82.
—— (1970). Size allometry: size and shape variables with characterizations of the lognormal and generalized gamma distributions. *J. Amer. Statist. Assoc.*, **65**, 630–645.
—— (1975a). Statistical problems of size and shape. I. Biological applications and basic theorems. In *Statistical Distributions in Scientific Work* (G. P. Patil, S. Kotz and J. K. Ord, eds), pp. 187–217, Dordrecht, Holland: D. Reidel Publishing Company.
—— (1975b). Statistical problems of size and shape. II. Characterizations of the lognormal, gamma and Dirichlet distributions. In *Statistical Distributions in Scientific Work* (G. P. Patil, S. Kotz and J. K. Ord, eds), pp. 219–239. Dordrecht, Holland: D. Reidel Publishing Company.
PEARSON, K. (1897). Mathematical contributions to the theory of evolution. On a form of spurious correlations which may arise when indices are used in the measurement of organs. *Proc. R. Soc.*, **60**, 489–498.
SARMANOV, O. V. and VISTELIUS, A. B. (1959). On the correlation of percentage values. *Doklady Akad. Nauk. SSSR*, **126**, 22–25.
STEPHENS, M. A. (1974). EDF statistics for goodness of fit and some comparisons. *J. Amer. Statist. Assoc.*, **69**, 730–737.
THOMPSON, R. N., ESSON, J. and DUNCAN, A. C. (1972). Major element chemical variation in the Eocene lavas of the Isle of Skye, Scotland. *J. Petrology*, **13**, 219–253.
WILKS, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Ann. Math. Statist.*, **9**, 60–62.
WORKING, H. (1943). Statistical laws of family expenditure. *J. Amer. Statist. Assoc.*, **38**, 43–56.