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 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

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

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

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

is one-to-one, with Jacobian $D\mathbf{x}^{(d)}/D(\mathbf{t}; \mathbf{s}_1, \ldots, \mathbf{s}_{k+1}) = t_1^{d_1} \ldots t_{k+1}^{d_{k+1}}$ and with inverse P^{-1} given by $x_{c_{j-1}+r} = t_j s_{jr}$ $(r = 1, \ldots, d_j;\ j = 1, \ldots, k+1)$. 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, $x_1 + \ldots + x_c$. 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

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

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 $\alpha_i > 1$ $(i = 1, \ldots, d+1)$ 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), $\mathbf{t} \perp\!\!\!\perp \mathbf{s}_1 \perp\!\!\!\perp \ldots \perp\!\!\!\perp \mathbf{s}_{k+1}$, with $\mathbf{t}$ of $D^k(\boldsymbol{\gamma})$ form and $\mathbf{s}_j$ of $D^{d_j}(\boldsymbol{\beta}_j \gamma_j)$ form $(j = 1, \ldots, k+1)$ where $P(\boldsymbol{\alpha}^{(d+1)}) = (\boldsymbol{\gamma}; \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_{k+1})$.

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

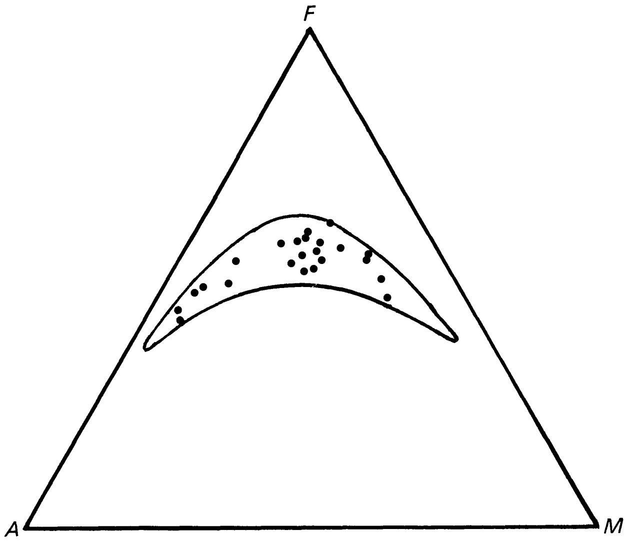

FIG. 1. A concave data set and the 95 per cent prediction region of a fitted additive logistic normal distribution. The points are the subcompositions $C(\mathrm{Na_2O+K_2O, Fe_2O_3, MgO})$ of 23 aphyric Si-poor basalt-benmoreites from the AFM diagram of Fig. 7 of Thomson, Essen and Duncan (1972).

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).

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

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 $x_1 x_2 \dots x_{d+1}$. 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

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

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 $fN^d(\boldsymbol{\mu},\Sigma)$ distribution in $\mathbb{S}^d$ then $\mathbf{y}^{(d)}=f^{-1}(\mathbf{x}^{(d+1)})$ follows a $N^d(\boldsymbol{\mu},\Sigma)$ 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 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

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

Successive perturbations $\mathbf{w}_{[1]}, \mathbf{w}_{[2]}, \ldots$ on an initial composition $\mathbf{x}_{[0]}$ produce a sequence of compositions $\mathbf{x}_{[1]}, \mathbf{x}_{[2]}, \ldots$, related by $\mathbf{x}_{[r]} = \mathbf{w}_{[r]} \circ \mathbf{x}_{[r-1]}$ $(r = 1, 2, \ldots)$ 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 $\boldsymbol{\Omega} = \operatorname{cov}(\log \mathbf{w})$ 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 $f N^d(\boldsymbol{\alpha}+\boldsymbol{\beta} t, \boldsymbol{\Sigma})$ or $f N^d(\boldsymbol{\alpha}+\boldsymbol{\beta} \log t, \Sigma)$. 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 $(i = 1, \ldots, d+1)$, if for simplicity we identify household total expenditure with household income. In expectation terms $\beta_i = e_i - e_{d+1}$ $(i = 1, \ldots, d)$, 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 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 $b \leq c$. 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

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

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 $aN^d(\boldsymbol{\mu}, \mathbf{\Sigma})$ 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

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

where $\mathbf{V}$ is the sample covariance matrix of the transformed vector $\log(\mathbf{x}^{(d)}/x_{d+1})$ and $\hat{\mathbf{\Sigma}}_H$ and $\hat{\mathbf{\Sigma}}_M$ are the maximum likelihood estimates of $\mathbf{\Sigma}$ under the hypothesis and model, against the appropriate upper percentile of $\chi^2\{\frac{1}{2}d(d-1)-1\}$. The estimate $\hat{\mathbf{\Sigma}}_M$ is simply $\mathbf{V}$ but the computation of $\hat{\mathbf{\Sigma}}_H$ 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 $|\mathbf{\Sigma}_H|$, $\mathbf{\Sigma}_H^{-1}$ 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 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 $C(FeO + Fe_2O_3, MnO)$ 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

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

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		
$\mathscr{I}_1$	$C(\mathbf{x}^{(c)}) \perp\!\!\!\perp T(\mathbf{x}^{(c)})$	$\boldsymbol{\beta}_1 = \mathbf{0}$
$\mathscr{I}_2$	$C(\mathbf{x}_{(c)}) \perp\!\!\!\perp T(\mathbf{x}_{(c)})$	$\boldsymbol{\beta}_2 = \mathbf{0}$
Conditional subcompositional independence		
$\mathscr{C}$	$C(\mathbf{x}^{(c)}) \perp\!\!\!\perp C(\mathbf{x}_{(c)}) \mid T(\mathbf{x}^{(c)})$	$\boldsymbol{\Sigma}_{12} = \mathbf{0}$
Neutrality		
$\mathscr{N}_1$ (left)	$C(\mathbf{x}^{(c)}) \perp\!\!\!\perp \mathbf{x}_{(c)}$	$\boldsymbol{\beta}_1 = \mathbf{0}, \boldsymbol{\Sigma}_{12} = \mathbf{0}$
$\mathscr{N}_2$ (right)	$C(\mathbf{x}_{(c)}) \perp\!\!\!\perp \mathbf{x}^{(c)}$	$\boldsymbol{\beta}_2 = \mathbf{0}, \boldsymbol{\Sigma}_{12} = \mathbf{0}$
Partition independence		
$\mathscr{P}$	$\perp\!\!\!\perp \{C(\mathbf{x}^{(c)}), C(\mathbf{x}_{(c)}), T(\mathbf{x}^{(c)})\}$	$\boldsymbol{\beta}_1 = \mathbf{0}, \boldsymbol{\beta}_2 = \mathbf{0}, \boldsymbol{\Sigma}_{12} = \mathbf{0}$

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 $\mathbf{s}_1 \perp\!\!\!\perp \mathbf{s}_2 \mid t$ 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 $\mathbf{s}_2 \perp\!\!\!\perp (\mathbf{s}_1, t)$. 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 $\mathbf{s}_1 \perp\!\!\!\perp t$ and $\mathbf{s}_1 \perp\!\!\!\perp \mathbf{s}_2 \mid t \Leftrightarrow \mathbf{s}_1 \perp\!\!\!\perp (\mathbf{s}_2, t)$ 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 $d \geq 3$ that we have a real distinction between the various concepts.

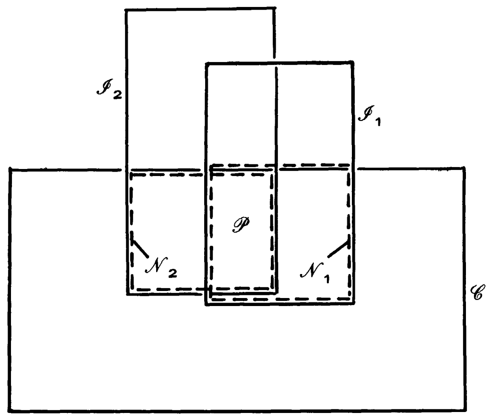

FIG. 2. Diagrammatic representation of the relationships between independence properties for a partition of order one.

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 Σ associated with independence hypotheses

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

Required matrix computations

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

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

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

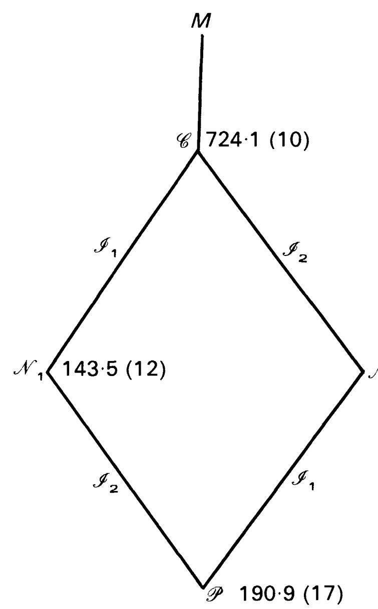

FIG. 3. Lattice for Skye lava analysis showing values of test statistics with associated degrees of freedom in brackets.

the lattice the hypothesis $\mathscr{P}$ is the simplest explanation of the relationship of $C(\mathbf{x}^{(c)})$, $C(\mathbf{x}_{(c)})$ and $T(\mathbf{x}^{(c)})$, 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 $\mathbf{s}_1 \perp\!\!\!\perp \mathbf{s}_2$, 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

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

we can specify partial subcompositional independence as the parametric hypothesis

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

in term of the obvious partitioning of $\boldsymbol{\Sigma}$. 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 $k = 1, \ldots, c$.

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 $C(\mathbf{x}_{(c)}) \perp\!\!\!\perp \mathbf{x}^{(c)}$ for $c = 1, \ldots, d-1$. 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

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

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 $\mathbf{y}_r$ $(r = 1, \ldots, n)$ 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

$$\begin{aligned} &|\hat{\boldsymbol{\Sigma}}_M| = |\mathbf{V}|, \quad |\hat{\boldsymbol{\Sigma}}_{\mathscr{I}_c}| = |\mathbf{V}_{11}| \cdot |\mathbf{V}_{22}|, \\ &|\hat{\boldsymbol{\Sigma}}_{\mathscr{I}^c}| = v_{11} \ldots v_{cc} |\mathbf{V}_{22}|, \quad |\hat{\boldsymbol{\Sigma}}_{\mathscr{I}^{d-1}}| = v_{11} \ldots v_{dd}, \end{aligned} \tag{7·4}$$

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
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

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

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.

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

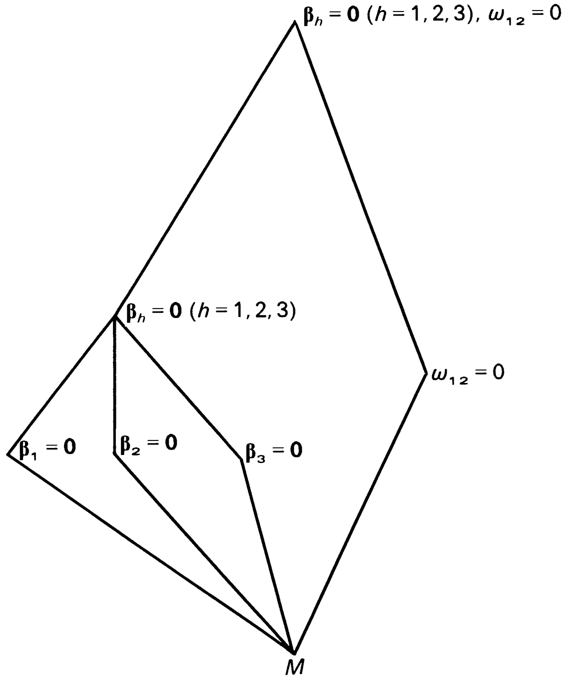

FIG. 4. Lattice for testing subcompositional invariance and neutrality for subcompositional shares.

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 $y_i = \{(x_i/x_{d+1})^\lambda - 1\}/\lambda$ $(i = 1, ..., d)$ 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 $\boldsymbol{\Sigma} = \text{cov}\{\log(\mathbf{x}^{(d)}/x_{d+1})\}$ might seem a hopeful technique until it is realized that trace $(\boldsymbol{\Sigma})$ is not invariant under a permutation of the components $x_1, ..., x_{d+1}$. 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.

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.

Discussion of Professor Aitchison's Paper

Dr J. A. Anderson (University of Newcastle upon Tyne): It gives me great pleasure to be able to propose this vote of thanks to Professor Aitchison. He has been absent from our deliberations for too long and I am very pleased to welcome him back. Indeed, I am pleased to welcome everyone who made journeys of varying difficulty and ingenuity because of the rail strike. Perhaps it emphasizes Professor Aitchison's foresight in leaving these shores before the combined misery of the weather, transport difficulties and the UGC contrived to give us such an inauspicious start to tonight's proceedings.

We have been very fortunate tonight in hearing about a breakthrough in the analysis of compositional data. Professor Aitchison is to be congratulated on producing a family of distributions which enables us not only to test hypotheses about compositional data but also to estimate parameters and to fit distributions. He has produced a class of working models for the practitioner and he has stimulated our interest in a rather neglected field. As Professor Aitchison has noted, there is a natural progression from several strands of his earlier work to the current paper. Like all good ideas, it is obvious once we see it.

It is traditional for the proposer of the vote of thanks to find all the holes in the paper and then proceed to tell the audience all about them. In this case that is rather difficult because I admire the paper tremendously and find it rather difficult to criticize.

A major attraction of the paper is that there is one class of models which is intended for quite general use. In the paper, a distinction is made between intrinsic and extrinsic analyses. In the former, we are concerned solely with compositions. In the latter, the basis is also defined. However, it appears that three kinds of sampling plan are possible. Using the notation of the paper, $\mathbf{x}^{d+1}$ denotes the composition, $\mathbf{w}^{d+1}$ the basis and T is the sum of the elements of $\mathbf{w}^{d+1}$. The three kinds of sampling are:

1. *Direct composition sampling.* Here $\mathbf{x}^{d+1}$ is observed directly so that $\mathbf{z}^{d+1}$ and T are neither available nor defined.
2. *Conditional sampling.* Here, T is selected and then $\mathbf{z}^{d+1}$ observed. Clearly $\mathbf{x}^{d+1}$ is available but as T is fixed, the variation in $\mathbf{z}^{d+1}$ does not reflect genuine random variation. For example, in the Skye lava context, the size of a specimen is at choice.
3. *Complete random sampling.* Here $\mathbf{z}^{d+1}$ is observed from some random mechanism and, hence, $\mathbf{x}^{d+1}$ and T are available as random variables. For example, in a household expenditure survey, the amounts spent under various headings are noted.

It is not clear whether case 2 above is more suitable for an extrinsic or an intrinsic analysis. More seriously, if measurement or observational error is a material factor, then data observed either as case 2 or case 3 above require modification of the basic transformed normal.

Another important point for the practitioner is whether his sample compositions contain any zeros. This is referred to in Section 7.4 and relates to the difficulty of finding $\ln 0$. In some cases there will be no or few zeros and the problem is of little moment. For example, there are no zeros in the Skye lava data which give a good fit for the transformed normal model. However, some examples fitted in Newcastle with moderate numbers of zeros do not fit the transformed normal model so well. Moreover, the fit varies drastically with the treatment of the zeros. One possibility is to invoke again the idea of measurement error leading to zeros by a round-off process. This could lead to consideration of a left-censored distribution. In other situations, the zeros are an intrinsic part of the data, caused perhaps by a mixture of distributions or by contamination. I wonder whether the measures outlined in Section 7.4 are really powerful enough to deal with the full range of these difficulties.

I found the treatment of compositional invariance in Section 4.2 rather restrictive. For example, it does not appear to be appropriate when measurement error is a feature of the data. Suppose we have an unobservable basis $u_1, u_2, \ldots, u_{d+1}$, and $u_i = k_i U$, $i = 1, \ldots, d+1$, where $U = \Sigma_{i=1}^{d+1} U_i$. Suppose further that $\mathbf{k} = (k_1, \ldots, k_{d+1})$ is constant over all sample points, whereas U varies randomly over realizations. Hence the basis $\mathbf{u}$ is compositionally invariant. Now $\mathbf{u}$ is unobservable but suppose that there is an observable basis $\mathbf{w}$, such that $w_i = u_i + e_i$, $i = 1, \ldots, d+1$. The e_i are independent error terms. Since we have perturbed a compositionally invariant basis by independent, measurement errors, we might require the basis $\mathbf{w}$ to be compositionally invariant also. In fact, it is not. To see this, let $x_i = w_i/W$, then $x_i = k_i + (e_i - k_i E)/W$, for all i, where $W = \Sigma_{i=1}^{d+1} w_i$ and $E = \Sigma_{i=1}^{d+1} e_i$.

Clearly x_i and W are not independent as they should be for compositional invariance. I conclude that Professor Aitchison's definition of the latter is not appropriate where measurement error is present.

This is an important paper and my comments add little to a full and extensive discussion. It gives me great pleasure to move this vote of thanks.

Professor A. P. DAWID (University College London): I join with Dr Anderson in the pleasure I feel in welcoming Professor Aitchison back before us. Once again our Society can benefit from that practised elegance which makes the important new ideas he presents appear so obvious that we might have thought of them ourselves—but then we did not!

I have only one quibble with Aitchison's approach, and that concerns his election to treat the $D = d+1$ constituents of a mixture unsymmetrically: he imposes structure only on the first d, leaving the last to be determined "by default". I believe that some additional insight may be gained by retaining a completely symmetrical expression throughout. I shall confine attention to the logistic-normal distribution (Aitchison and Shen, 1980), generated by the additive logistic transformation. I prefer to express this as follows:

Let $\mathbf{Y} \sim N_D(\mathbf{v}, \mathbf{\Phi})$, and define $\mathbf{X}$ by

$$X_i \propto e^{Y_i} \quad (i = 1, 2, ..., D = d+1). \tag{1}$$

The constant of proportionality is not to depend on i, and can be chosen to be $(\Sigma\, e^{Y_i})^{-1}$, summing over i from 1 to D, in which case $\Sigma\, X_i = 1$, and $\mathbf{X}$ has a logistic-normal distribution. But, for given $\mathbf{X}$, equation (1) does not determine the Y's uniquely, but only up to a (fixed or random) additive constant. In particular, the distribution of $\mathbf{X}$ remains the same if the parameters $\mathbf{v}$ and $\mathbf{\Phi}$ are transformed by $v_i \to v_i + a$, $\phi_{ij} \to \phi_{ij} + \delta_i + \delta_j$. So the price of symmetry is some indeterminacy in the value and distribution of $\mathbf{Y}$. I think this price is well worth paying.

The trick in handling (1) is to work only with the *contrasts* in the Y's. These are invariant under addition of an arbitrary constant, and knowledge of all contrasts is equivalent to knowledge of the ratios of the X's and thus to $\mathbf{X}$ itself under the condition $\Sigma\, X_i = 1$, summing over $i = 1, ..., D$. Note that any contrast $\Sigma\, a_i\, Y_i$ (where $\Sigma\, a_i = 0$) in the Y's is equivalent to a *log-contrast* $\Sigma\, a_i \log X_i$ in the X's. So, under a logistic-normal model, all such log-contrasts are normally distributed, and this is a characterization of these distributions.

An important application of these ideas was made by Lindley (1964), who approximated the Dirichlet distribution by noting that, for such a distribution, (1) holds with the Y_i having independent log-gamma distributions. A normal approximation to the log-gamma allows the log-contrasts in Dirichlet proportions (which completely determine those proportions) to be treated as normally distributed.

An extension of this approach handles the case of several proportion vectors $(\mathbf{X}_j)$, where $\mathbf{X}_j = (X_{ij})$ with $\Sigma_i X_{ij} \equiv 1$. We just put $X_{ij} \propto e^{Y_{ij}}$, where all the Y's have a jointly normal distribution, and confine attention to log-contrasts of the form $\Sigma\, a_{ij} \log X_{ij}$, where $\Sigma_i\, a_{ij} = 0$. The $(\mathbf{X}_j)$ might be the various components (amalgamation and subcompositions) of a partition; the model allows complex dependence both within and between the several vectors of proportions. Again, we have a certain amount of acceptable indeterminacy.

We can go on to introduce a size-variable T, jointly distributed with the Y's. For example, $(Y_1, ..., Y_D, \log T)$ could have a multivariate normal distribution, or we could even take $\log T = Y_D$ or $\Sigma\, Y_i$. Although the Y's can be thought of as a basis for the X's, this can be a purely fictitious one, and there is no need to interpret $\Sigma\, e^{Y_i}$ as the size variable.

Professor Aitchison recognizes some of the difficulties associated with his asymmetric definition when he notes that an attempt to construct principal componets from his Y's suffers from a dependence on which constituent is arbitrarily taken as the Dth. In a symmetric formulation, we simply require principal directions in the d-dimensional space of contrasts in the Y's; the principal variables are those which attain a stationary value for $\mathrm{var}(\Sigma\, a_i\, Y_i)$ under the constraints $\Sigma\, a_i = 0$, $\Sigma\, a_i^2 = 1$. Equivalently, a straightforward principal components analysis may be performed on the variables (W_i), where $W_i = Y_i - \bar{Y} = Z_i - \bar{Z}$ with $Z_i = \log X_i$. To avoid singularity, the covariance matrix $\mathbf{\Psi}$ of $\mathbf{W}$ may be transformed by $\psi_{ij} \to \psi_{ij} + a$. This will introduce an extra principal variable $\Sigma\, W_i$, which may be ignored.

Another advantage of the formulation (1) is that, for certain problems, if we proceed as if we knew the Y's, we may discover that the quantities we have to calculate are unaffected by the indeterminacy and so can be calculated from the X's alone. In this case, the *Reduction Principle* (Dawid, 1977) asserts that we should make the same inference from the X's as from the Y's. For example, suppose we have p samples, yielding X_{ij} as the portion of constituent i in sample j. One possible model is $X_{ij} \propto e^{Y_{ij}}$, with

$$Y_{ij} \sim N(v_{ij}, \sigma^2) \quad \text{all independently.}$$

We might wish to test the hypothesis that the samples come from a single population, so that the $(\mathbf{X}_j)$ are identically distributed. Owing to the arbitrary constant in the Y's for each j, this has to be expressed as: v_{ij} has the form $\alpha_i+\beta_j$. In other words, for the Y's this is the hypothesis of *no interaction* between sample and constituent. Assuming, for simplicity, that σ^2 is known, we would test this by referrring $\Sigma_{ij}(Y_{ij}-Y_{i.}-Y_{.j}+Y_{..})^2/\sigma^2$ to $\chi^2_{(D-1)(p-1)}$. But the test statistic is invariant under the indeterminacy transformations $Y_{ij}\to Y_{ij}+c_j$, and so is calculable from the X's. In fact, it is just $\Sigma_{ij}(Z_{ij}-Z_{i.}-Z_{.j}+Z_{..})^2/\sigma^2$, with $Z_{ij}=\log X_{ij}$. So the homogeneity hypothesis is tested by referring this quantity to its null χ^2-distribution.

If the hypothesis is accepted, the estimate of the (common) proportion of constituent i, based on least-squares estimation in Y-space, is found by taking the geometric mean of the p sample proportions, and renormalizing so that the estimates sum to 1.

Another important application is to account for several components of variation in compositional data. Thus for the Skye lava data, we might measure proportions of constituents (i) in several samples (k) taken in various areas (j). We could take $X_{ijk}\propto e^{Y_{ijk}}$, with $\Sigma_i X_{ijk}=1$, and a possible model

$$Y_{ijk}=a_i+b_{jk}+\varepsilon_{ij}+\eta_{ijk}, \tag{2}$$

where $\varepsilon_{ij}\sim N(0,\sigma_\varepsilon^2)$, $\eta_{ijk}\sim(0,\sigma_\eta^2)$, all independently. (Note that the model must include the "indeterminacy constants" (b_{jk}). This is analogous to a similar requirement on log-linear models for multinomial distributions when using a Poisson representation.) We find that the usual estimates of σ_ε^2, σ_η^2, and contrasts in the (a_i) are unaffected by the indeterminacy, and hence are calculable from the X's. Again, we can just replace Y_{ijk} by $\log X_{ijk}$ in the formulae, and proceed *as if* the $\log X_{ijk}$ satisfied the model (2), ignoring the constraints.

I have said more than enough to demonstrate that the ideas that Professor Aitchison has put before us tonight have been a great stimulus to at least one person. I am sure that all of you will agree with me that Professor Aitchison deserves a very warm vote of thanks, which I have great pleasure in seconding.

The vote of thanks was passed by acclamation.

Professor G. J. GOODHARDT (City University Business School): One of the great values to the applied statistician of papers of this kind is that they provide us with a language in which we can set down our previously rather vague problems and difficulties in more precise terms and so communicate them more readily to the mathematical experts. For that, in particular, I welcome this evening's paper, and I would like to set out a problem I have to see if anyone can help.

It arises in the study of consumer choice behaviour where I am concerned to model the mix of brands of a product that people buy over a period. A fruitful model has been to postulate that each consumer has a personal vector of probabilities (adding to one) specifying the probability that they will buy each of the brands conditional on their making some purchase. These probabilities themselves are, of course, unobservable. The data consist of the mixed multinomial distribution of the number of purchases of each brand made in a particular time period. The problem is to specify the distribution of this probability vector over the population of consumers.

In many markets where consumers exercise a simple choice between the available brands a Dirichlet distribution provides a very good description of the data and all the independence properties that that implies make sense. However, in some markets certain groups of similar brands seem to cluster together and pairs of brands within a cluster are more likely both to be bought by the same consumers than pairs in different clusters. Such markets are referred to in the marketing literature as partitioned, and that word has much the same connotation as in Professor Aitchison's paper. I would like to fit to such markets a complex Dirichlet model in which the partition is of Dirichlet form and each of the subcompositions is also of Dirichlet form. In the notation of the paper we have a partition:

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

$$\mathbf{t}\text{ is of form }D^k(\boldsymbol{\gamma}),$$

$$\mathbf{s}_j\text{ is of form }D^{d_j}(\boldsymbol{\beta}_j),\quad j=1,\ldots,k+1.$$

Now, if

$$T(\boldsymbol{\beta}_j)=\gamma_j,\quad j=1,\ldots,k+1$$

then

$$\mathbf{x}^{(d+1)}\text{ is of form }D^d(\boldsymbol{\alpha})$$

and there is no partitioning at all in the marketing sense.

The kinds of problems I will have to address are:
(i) In particular cases, can the more complex, partitioned model be justified compared with the simpler, non-partitioned model?
(ii) A particular partitioning will often be suggested *a priori*, but can we derive a partitioning from the data?
(iii) Are there ways of deciding between alternative *a priori* partitionings? This relates to the so-called "hierarchy of choice". For example, in the canned soup market where consumers choose between, say, Heinz Tomato Soup and Campbell's Chicken Soup, whether the partitioning is by brand or by flavour may be an indication of whether consumers first decide on a brand and then choose a flavour or *vice versa*.

Professor Aitchison may be surprised to see his lofty ideas applied to such mundane problems but the breadth of application is a tribute to the value of his contribution. One of the advantages of working in Consumer Research is that there is a wealth of routinely collected data that is available for more sophisticated analysis, and I am grateful to our author tonight for providing further tools.

Dr R. J. HOWARTH (Imperial College, Department of Geology): Speaking as a geologist, I am particularly pleased that Professor Aitchison should address himself to an area which has long caused us interpretational problems. We are frequently plagued by the facts that we usually have to accept as our statistical sample a suite of specimens gathered from where the rocks happen to be accessible and that the chemical compositions of the major rock-forming minerals or chemical elements are traditionally expressed in terms of percentages. In addition, nature has made the behaviour of rock-forming magma systems very difficult to understand and we often use the variation in chemical compositions as clues to the understanding of fundamental mechanisms. The use of methods such as the AFM ternary diagram, discussed in the paper, is in part as an aid to the visualization of high-dimensional space, but also because it is conceptually easier to relate to mineralogical changes in the rock compositions than inspection of the raw data.

The geologist is often interested in three main aspects of rock composition: description of the overall variation, leading perhaps to questions such as "are these two suites of specimens actually distinct, or could they have come from the same parent magma suite?"; relating chemical composition to the known mineralogy of the rocks; and investigation of the statistical significance of inter-element or other multi-attribute correlations (generally based on the Pearson linear coefficient in geological studies). The author's recent contribution to the investigation of null correlations of proportions should be of great assistance to geological studies, particularly in petrological investigations such as those exemplified by the Skye lavas study.

It is particularly interesting that the confidence bound on the Skye lavas AFM diagram (Fig. 1) projects as a concave enclosing envelope, since in many traditional petrological studies, a petrologist would draw a line through the median line of this envelope and interpret it as representing a "trend" of chemical (compositional) variation from one end-member of extreme composition to another, in terms of a continuous changing suite of rocks representing chemical evolution as a product of magmatic activity. Its representation in terms of a multivariate normal distribution within the simplex will necessitate rethinking of at least some geological hypotheses.

The relatively small number of geologists who have used statistical techniques to study compositional information have generally used empirical approaches such as principal components or cluster analysis methods and, more recently, ridge regression in an attempt to simplify the understanding of large multi-measurement data sets in geological studies. I would like to thank Professor Aitchison for his contribution to the statistical methodology at our disposal and encourage him to take a continuing interest in the problems of the earth sciences.

Professor M. A. STEPHENS (Simon Fraser University): Professor Aitchison will, I am sure, receive congratulations on his paper, and I am pleased to add my own. The subject interests me greatly since I have been working with data of this type, which I have called continuous proportions. These data consisted of activity patterns of students; in the notation of the present paper, component $x(i)$ of a vector $\mathbf{x}$ was the proportion of time spent in activity i. There were many students, and many activities, and interest focused on whether or not the activity patterns differed between groups of students; for example, between the sexes, or between disciplines. In the analysis, I transformed vector $\mathbf{x}$ to a vector $\mathbf{v}$ with $v(i) = \sqrt{(x(i))}$; the constraint that the sum of $x(i)$ is 1 translates into making $\mathbf{v}$ a vector with endpoint on the hypersphere of unit radius. Then distributions on the hypersphere are available to describe the data; for example, for activity patterns I used the p-dimensional von Mises distribution. A very simple ANOVA technique first developed by Watson (1956) can be used to analyse differences between subpopulations. The analysis is illustrated in Stephens (1980, 1982).

So I am one of those statisticians mentioned briefly in Section 8, whose fingers are anxious to repack the package; some new areas for delivery are also mentioned in Stephens (1980). But the two methods of analysis should be complementary, and I hope to make some comparisons using Professor Aitchison's data sets.

In the transformations to multivariate normality, the constraint on the proportions gets submerged, to bob up, presumably, in the covariance structure of the multivariate normal distribution; this can be handled in the analysis, once the multivariate normal can be accepted as the induced distribution. The techniques especially lend themslves to the questions of compositional independence which are raised here, and to regression on an external variable. Regression on a sphere is difficult, and an unexpected bonus of the paper is that it suggests techniques which might be tried also with directional data. One worry is that the transformations seem to have worked almost too well, judging from the goodness-of-fit tests of Section 3; has the hidden constraint affected these tests to give overly good results? It would be interesting to have further details.

Dr N. I. FISHER (C.S.I.R.O. Division of Mathematics and Statistics, Sydney, Australia): I should like to congratulate the speaker on his substantial contribution to a most vexing problem—the analysis and interpretation of compositional data. There are two qualitatively different comments I wish to make.

The first relates to modelling. Clearly, the speaker has been very successful in fitting simple models to normal-transformed data; the counterpart to the simplicity of these models is the complexity of corresponding relationships amongst the untransformed components. This is hardly an original observation. Yet, there are certain aromas rising from the murky potage of compositional data problems which are redolent of some aspects of problems with directional data, and herein lies the point. When attacking these latter problems, one is ultimately better off working within the confines of the original geometry (of the circle, sphere, cylinder, ...), and with techniques particular thereto (vector methods, etc.), in terms of perceiving simple underlying ideas and of modelling them in a natural way. Mapping from, say, the sphere into the plane, and then back, rarely produces these elements, and usually introduces unfortunate distortion. I still hold out some hope that simple models of dependence can be found, peculiar to the simplex (thus revealing myself as a page, if not a knight—see Section 8(vii)). Meanwhile, I shall analyse data with the normal-transform methods.

The second comment concerns a common problem with geochemical data. When a rock sample is analysed for several elements, different techniques may be used for measuring minor or trace elements from those used for major elements; correspondingly, different sorts of precision attach to the various components. One would hope that the methods used to analyse these data would be sturdy enough to be relatively unaffected by such abuses.

Professor C. E. V. LESER (University of Leeds): If this very interesting analysis is to be applied to mutually exclusive and exhaustive commodity groups, it seems surprising that the simplex is not so defined that $x_1+...+x_d = 1$, in other words that the starting point is the "augmented set". The present treatment introduces an asymmetry, which might be justified if we were dealing with proportions of income rather than total expenditure, so that x_{d+1} represents savings; however, x_{d+1} would in this case not necessarily be positive or even non-negative.

As the budget share approach represents an alternative to demand equations with logarithms of commodity group expenditures (or logs of expenditure shares) as dependent variables, it is not clear from the econometric point of view why a logarithmic transformation should then be adopted not only for income but also for budget shares. Some advantages of using budget shares rather than expenditures or their logs, which were mentioned in Leser (1963), are that zero expenditures and combination of commodity groups present no problems.

Dr A. C. ATKINSON (Imperial College, London): Like several other speakers this evening, I would like to welcome Professor Aitchison back to these shores and to congratulate him on an interesting and important paper.

The models that we have heard about this evening are concerned with transformations from the simplex into more tractable spaces. As a couple of the comments suggest, there are other transformations than those involving logarithms and exponentials. Some time ago, in work on the design of experiments which led to optimization over the simplex, I used the trigonometric transformation $x_1 = \sin^2\theta_1$, $x_2 = \cos^2\theta_1 \sin^2\theta_2$, etc. (Atkinson, 1969). This produces a space which is Euclidean, but with the disadvantage that the transformation is not one to one. An advantage of the transformation is that the problem with zero responses does not arise.

Professor Aitchison has called his responses x. A related problem that arises where the x's sum to 1 is in experiments with mixtures. In his first example it might be that the chemical compositions of lava in the island of Skye are measured, but that interest is in some univariate response Y, which might be the hardness of the rock. The problem of suitable models for such data does not seem to have been completely solved. The usual models, which are polynomial in x, are described in detail by Cornell (1981). The models do not really seem to describe the very rapid changes that there may be in the response towards the edge of the simplex. Gunpowder with only one component is a very different substance from a mixture of saltpetre, carbon and sulphur. One suggestion for an alternative model (Draper and St John, 1977; St John and Draper, 1977) is to introduce inverse terms into the model. These lead to the kind of problems with zero components about which we have heard this evening. I would be interested to know whether Professor Aitchison has any advice on how to deal with this problem.

Professor D. R. Cox (Imperial College, London): The paper describes an important and flexible way of representing and understanding distributions of compositional data. It would be interesting to know more about simple stochastic models for such systems. These models would presumably have to be fairly specific to particular applications.

One such is the mixing and blending of textile fibres, say of different colours (Cox, 1954). Here the proportions of interest are the proportions, by weight, surface area or number, of the different colours as measured at a sample of yarn cross-sections. There is a base-line model according to which the fibres of different colours are arranged in independent Poisson processes. Such "ideal" mixing is not achievable and to a certain extent departures from the Poisson model can be interpreted physically.

Rather similar mixing problems arise in several chemical engineering contexts.

The following contributions were received in writing, after the meeting.

Professor J. N. Darroch (Flinders University): In this and other papers Professor Aitchison has greatly expanded both the theory of distributions on the simplex and the range of possible statistical analyses of compositional data.

The independence problems at the core of previous work are those which are fully manifested when $d = 2$, the problem there being to define independence of x_1, x_2, where $x_1 > 0$, $x_2 > 0$, $x_1 + x_2 < 1$. This is preferably done in such a way that, when the distribution of (x_1, x_2) is concentrated near $(0, 0)$, the resulting simplex independence is approximately the same as ordinary, rectangular independence. The main candidates for simplex independence have been neutrality and the more general F-independence, the latter being applicable to the simplex lattice as well as to the simplex. Once defined for $d = 2$, these concepts are easily extended to cover distributions on $\mathbb{S}^d$ for $d > 2$. Work in this area has been discouraged by the realization that there are almost certainly no nice parametric families of distributions on the simplex which can handle dependence as well as independence.

Professor Aitchison has brought new life to the subject by providing many ways of adapting multivariate normality to the simplex and by proposing new definitions of independence, notably subcompositional independence. By conceiving the basic variables to be ratios of x's, instead of x's, he has created new fields for problem formulation and data analysis. However his ideas require $d \geqslant 3$ for their proper manifestation. They have very little application to the case of $d = 2$ and the above-mentioned lack of nice families remains.

It is possibly time to acknowledge that while there is a shortage of nice parametric families on the simplex, that is ones for which the normalization constant of the density function is an explicit, tractable function of the unknown parameters, there is an infinite choice of not-nice parametric families. There is no evidence that Nature confines her attention to nice families and computational constraints no longer compel us to do so.

Mr Bent Jørgensen (Odense University): The problem of zero components has been touched upon by both Professor Aitchison and some of the discussants, and is clearly an important one. One way to model this phenomenon is by taking the basis to have a distribution for which the components have a mass probability at zero. If, in particular, the components are independent, a model for basis independence is obtained.

One interesting model of the latter kind is obtained by taking the distribution of the reciprocal of a component of the basis to have the so-called defective inverse Gaussian distribution (Whitmore, 1978), with probability density function

$$f(x) = \begin{cases} (a^2/2\pi\omega x^3)^{\frac{1}{2}} \exp(-(a-\delta x)^2/2\omega x), & 0<x<\infty, \\ 1-\exp(2a\delta/\omega), & x=\infty. \end{cases}$$

Since the defective inverse Gaussian distribution has a mass probability at infinity, the components have a mass probability at zero. An important property of this model is that it has a scale parameter, so that the distribution of the logarithm has a location parameter. It follows that this model can be analysed analogously to the way described by Professor Dawid, that is, if we have a linear model for the location parameters, as we are only interested in contrasts, we may analyse the composition vector as though it were the basis from which it originated.

Dr J. T. Kent (University of Leeds): I would like to say a few words about a multivariate gamma distribution which can be used to define a generalization of the Dirichlet distribution. Suppose initially that the index parameters α_j are integer multiples of $\frac{1}{2}$ and set $n_j = 2\alpha_j$, $j = 1, \ldots, p+1$. Let $\mathbf{\Sigma} = (\sigma_{jk})$ be a positive definite matrix and consider a sequence $\mathbf{u}_1, \mathbf{u}_2, \ldots$ from the $p+1$ dimensional normal distribution with mean $\mathbf{0}$ and covariance matrix $\mathbf{\Sigma}$. Denote the components of $\mathbf{u}_i$ by u_{ij} and define a $p+1$ dimensional vector $\mathbf{w}^{(p+1)}$ by

$$w_j = \sum_{i=1}^{n_j} u_{ij}^2, \quad j = 1, \ldots, p+1.$$

Each w_j has a gamma distribution with index α_j and scale $(2\sigma_{jj})^{-1}$ so that $\mathbf{w}^{(p+1)}$ can be said to have a multivariate gamma distribution. (The case where all the n_j are equal is well known; see Krishnamoorthy and Parthasarathy, 1951.) Thus the composition based on $\mathbf{w}^{(p+1)}$, $\mathbf{x}^{(p+1)} = C(\mathbf{w}^{(p+1)})$ gives a generalization of the Dirichlet distribution. In particular, if $\mathbf{\Sigma}$ is a multiple of the identity matrix, then $\mathbf{x}^{(p+1)}$ has a standard Dirichlet distribution.

Of course, a family of distributions where the α_j are restricted to discrete values has limited usefulness. Thus it is of interest to note that if $\mathbf{\Sigma}$ is assumed to be a *Markov* matrix (i.e. $\mathbf{\Sigma}^{-1}$ is tri-diagonal), then this multivariate gamma distribution can be defined for all *real-valued* indices $\alpha_j > 0$; see the infinite divisibility properties in Griffiths (1970).

Unfortunately, it does not seem that the generalization of the Dirichlet distribution arising from this construction will be very tractable. Also, the Markov assumption on $\mathbf{\Sigma}$ implies that there is a preferred ordering of the variables.

Professor Tom Leonard (University of Wisconsin–Madison): My comments on this interesting and stimulating paper could be viewed as supplementing the ideas expressed in Section 4.3.

Suppose there exists a basis $\mathbf{w}^{(d+1)} = (w_1, \ldots, w_{d+1})$ where $\boldsymbol{\beta} = (\log w_1, \ldots, \log w_{d+1})^{\mathrm{T}}$ possesses a multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\mathbf{C}$. Rather then working with a specific set of d log-contrasts, it is often simpler to work with $d+1$ multivariate logits (only identified up to the addition of the same scalar to each logit) and to write

$$x_j = e^{\gamma_j} \Big/ \sum_{g=1}^{d+1} e^{\gamma_g} \quad (j = 1, \ldots, d+1). \tag{1}$$

Then the distribution of the logit vector may be written as

$$\boldsymbol{\gamma} = \eta \mathbf{e}_{d+1} + \boldsymbol{\beta}, \tag{2}$$

where $\mathbf{e}_{d+1}$ is the appropriate unit vector, and η is a scalar random variable possessing any distribution and not necessarily independent of $\boldsymbol{\beta}$. This arbitrary choice of distribution for η will not affect the distribution of the composition $\mathbf{x}^{(d+1)}$ as η cancels out in (1).

Note that if $\boldsymbol{\beta}$ satisfies (2) for some η, then (2) will also be satisfied by $\boldsymbol{\beta}+\varepsilon\mathbf{e}_{d+1}$ for any random variable ε. Therefore "basis independence" can be redefined by asking "Is there a random variable ε such that a basis $\boldsymbol{\beta}$ can be expressed as the sum of $\varepsilon\mathbf{e}_{d+1}$ and a random vector with independent elements?"

Under multivariate normal assumptions, this definition is equivalent to enquiring whether the covariance matrix of $\boldsymbol{\beta}$ can be expressed in the form

$$\mathbf{C} = \operatorname{diag}(\sigma_1^2, \ldots, \sigma_{d+1}^2) + \sigma^2\,\mathbf{e}_{d+1}\mathbf{e}_{d+1}^{\mathrm{T}} - \mathbf{e}_{d+1}^{\mathrm{T}} \times \mathbf{c} - \mathbf{c}^{\mathrm{T}} \times \mathbf{e}_{d+1}, \tag{3}$$

where $\mathbf{c} = (c_1, c_2, \ldots, d_{d+1})^{\mathrm{T}}$; for some choices of the variances $\sigma^2, \sigma_1^2, \ldots, \sigma_{d+1}^2$ and the covariances $c_1, \ldots, c_{d+1}$.

The covariance structure (3) implies that with $\mathbf{y} = (\log x_1, \ldots, \log x_{d+1})^{\mathrm{T}}$, any $d \times 1$ vector $\mathbf{z} = \mathbf{A}\boldsymbol{\beta} = \mathbf{A}\mathbf{y}$ of log contrasts possesses covariance structure

$$\operatorname{cov}(\mathbf{z}) = \mathbf{A}\operatorname{diag}(\sigma_1^2, \ldots, \sigma_{d+1}^2)\,\mathbf{A}^{\mathrm{T}}, \tag{4}$$

a similar conclusion to Professor Aitchison's result in (4.1) and (5.1).

There are no constraints on the parameters appearing in (4) and their maximum likelihood estimates may be obtained explicitly under the null hypothesis; they depend upon the observed $\mathbf{z}$'s and the likelihood ratio possesses the property of invariance under appropriate choices of $\mathbf{A}$. It therefore seems that basis independence leads to simpler statistical procedures then complete subcompositional independence. Moreover, I find the concept of basis independence to be more appealing; it is easy enough to generalize this idea, e.g. to independence of partitions.

Professor C. A. B. Smith (Galton Laboratory, University College London): Professor Aitchison is interested in knowing what transformations have been applied to proportions x_i. Fisher (1947) used the following one in an estimation problem, although he does not explicitly say so:

$$x_i = (\exp y_i)/\Sigma_j \exp y_j.$$

(N.B. Fisher's notation is x_i for the y_i used here, and R_2, R_0, etc., for the x_i used here.) The advantage from Fisher's point of view is that the relations

$$\partial x_i/\partial y_j = x_i(\delta_{ij} - x_j)$$

enable him to use the y_i as an intermediate step in the estimation of the x_i. (But there are alternative simpler ways of tackling the problem considered by Fisher.)

Mr Robin Thompson (ARC Unit of Statistics, University of Edinburgh): With regard to the use of principal components I wonder why $\boldsymbol{\Sigma}_1 = \operatorname{cov}\{\log(\mathbf{x}^{(d+1)})\}$ cannot be used to generate the principal components? If the invariance property is necessary then $\boldsymbol{\Sigma}_h = \operatorname{cov}\{\log(\mathbf{x}^{(d+1)}/h)\}$, where $(d+1)h = \Sigma \log x_i$, summing over i from 1 to $d+1$, can be used to generate principal components of $\boldsymbol{\Sigma}_1$ with the required invariance property. Or has Professor Aitchison found the need for other non-linear restrictions on the components?

Professor James E. Mosimann, (National Institutes of Health, Bethesda, Maryland, USA): My reaction to this paper is a mixed one. On the one hand, the paper is stimulating and clearly presented in a fashion which generates interest in the analysis of scale-free dimensionless data. On the other hand, some aspects of the paper represent a step backwards. For example, most of the definitions of independence offered are applicable to any positive random vector, not just to proportions over the simplex. To attach these definitions solely to proportions and the simplex, as done in this paper, is to seriously misdirect attention from the crucial question which both statistician and scientist must face when confronted with dimensionless scale-free observations; namely, what size variables are scientifically pertinent to the investigation? The importance of such a question when scale information is available is revealed by statement (2) below, and the question remains equally important when scale information is lacking (statement (5)). In contrast the particular expression of the data as proportions over the simplex is of virtually no importance in questions of independence (statements (1), (6) and (7)). By not explicitly treating such issues an important point is missed; namely, that previous work on transformations to the normal class in the analysis of size and shape variables (Mosimann, 1975a, 1978, 1979) is just as applicable to the proportions over the simplex as to any type of "shape vector", and that the distribution of proportions in such work is the "additive logistic normal" of this paper. To explain these comments, I would like to show how the present paper is related to published work on size and shape variables.

Using the author's notation with $\mathbf{w}$ of $\mathbb{P}^{d+1}$ a positive vector and $T(\mathbf{w}) = \Sigma w_i$, then the vector of proportions can be denoted by $\mathbf{w}/T(\mathbf{w})$. Similarly if we divide each co-ordinate of $\mathbf{w}$ by the last co-ordinate, w_{d+1}, we obtain a vector of ratios $\mathbf{w}/w_{d+1}$. Both such vectors are "shape vectors" and both $T(\mathbf{w})$ and w_{d+1} are "size variables". (A size variable G is any positive-valued function, homogeneous of degree 1, from $\mathbb{P}^{d+1}$ to $\mathbb{P}^1$ (always onto), and a shape vector is any function from $\mathbb{P}^{d+1}$ to $\mathbb{P}^{d+1}$ (never onto) given by $\mathbf{w}/G(\mathbf{w})$.) If proportions are to make scientific sense, each co-ordinate of $\mathbf{w}$ will be measured in the same scale (length, count, ...) and expressed in the same units (millimetres, dozens, ...). When this is true, a size variable will have the same scale and units as each co-ordinate of $\mathbf{w}$. In contrast, the co-ordinates of the shape vector will be dimensionless and scale-free. These scale-free co-ordinates may be statistically associated with size, and the study of such associations is part of the field of allometry (Huxley, 1932; Reeve and Huxley, 1945; Gould, 1977, and references).

A number of results for size and shape variables (Mosimann, 1970, 1975a, b) are directly related to the present paper. A shape vector of a given type (say, $\mathbf{w}/T(\mathbf{w})$) is linked by an invertible function with any other type of shape vector (say $\mathbf{w}/w_{d+1}$). Therefore (1) if some random shape vector is statistically independent of a random variable h, then every shape vector is also independent of h. Thus we can speak unambiguously of the independence of the class "shape" and size; that is of "isometry" with respect to size. Another important result is (2) That "shape" can be statistically independent of *at most one* size variable. (The only exceptions are scalar multiples of that same size variable.) The choice of a size variable is important. The addition (or deletion) of a measurement from $\mathbf{w}$ results immediately in the question: How is the independence of d-dimensional shape (d-shape) and d-size related to the independence of $(d+1)$-shape and $(d+1)$-size? Such a question led to the definition of "regular sequences" of size variables (Mosimann, 1975a, b). Examples of regular sequences are: $S_j = \Sigma_i^j w_i$ (additive size variables); $C_j = w_j$ (co-ordinate size variables);

$$M_j = \left(\prod_{i=1}^{j} w_i\right)^{1/j}$$

(multiplicative or geometric mean, size variables); $j = 1, ..., d+1$. There are many such sequences. Some important results (Mosimann, 1975a, b) for a regular sequence $G_1, ..., G_{d+1}$ are (3) there is an invertible function linking the vector of contiguous ratios $G_2/G_1, ..., G_{d+1}/G_d$ with any shape vector. (4) The statistical independence of d-shape and the additive ratio S_{d+1}/S_d is precisely the concept of a neutral proportion (from the right), and the mutual independence of the size-ratios $S_2/S_1, ..., S_{d+1}/S_d$ is precisely that of complete neutrality (from the right) as presented (from the left) in Connor and Mosimann (1969). (Definitions of neutrality with respect to any regular sequence (additive, co-ordinate, multiplicative, etc.) follow naturally, but as we see next there can be at most one kind of neutrality at a time). Thus (5) d-shape can be independent of at most one ratio G_{d+1}/G_d where the G's are regular. Thus if $\{H_j\}$ and $\{G_j\}$ are both regular sequences, d-shape independent of G_{d+1}/G_d implies d-shape is not independent of H_{d+1}/H_d. Next (6) the size ratios of $\mathbf{w}$ and of any associated shape vector are the same. Therefore independence properties based on the size ratios of a particular regular sequence are *invariant* over all shape vectors and $\mathbf{w}$ (Mosimann, 1975b, p. 233). They are applicable to any positive random vector. Finally, although a shape vector $\mathbf{w}/G(\mathbf{w})$ is intrinsically constrained so that its size G is 1 (since $G(\mathbf{w}/G(\mathbf{w})) = G(\mathbf{w})/G(\mathbf{w}) = 1$ by homogeneity) the form of the constraint does not affect the independence or lack thereof of its size ratios. Therefore (7) in an essential way *independence concepts based on size-ratios do not depend on the particular constraint* used for a given shape vector.

Now consider the author's definition of "compositional invariance" (additive isometry; shape independent of $T(\mathbf{w}) = S_{d+1}$) in Section 4.2. Here the crucial choice is that of the size variable (see statement (2) above), not that the scale-free data are expressed as proportions (see statement (1)). The definition implies a clear interest in additive size. On the other hand, the definition of the additive logistic normal reflects no such interest. In fact the author's "additive logistic normal" distribution is nothing more than the distribution of the proportion vector $\mathbf{w}/T(\mathbf{w})$ associated with a ratio vector $\mathbf{w}/w_{d+1}$ which has a multivariate lognormal distribution, and no additive neutrality can occur in a proportion vector with this distribution, nor in any of its subcomponents (Mosimann, 1975b, p. 224). The distribution is eminently suitable for testing multiplicative or co-ordinate (or any loglinear size) neutrality and isometry. It has been exploited in a variety of scientific analyses (Mosimann, 1975a, turtle morphology and human limb bones; 1978, schistosome egg counts; 1979, geographic variation in blackbirds). Thus I am myself surprised at the author's surprise in believing (Section 2.3) that the notion of transformation to normality has not emerged in studies of proportions.

The author's multiplicative logistic distribution allows the testing of additive neutrality in a given direction. This distribution results when logs of additive size ratios minus one (of the reverse permutation vector) are assumed to be normal. Caution is necessary since if even slight additive ratio independence should occur in more than one direction, Dirichlet distributions may result (James and Mosimann, 1980). Thus consistency problems abound in such applications, and the search for a good class which contains the Dirichlet and within which additive neutrality can be studied should continue.

It should also be noted that none of the points made above depend in any essential way on $\mathbf{w}$ being an "open" or unconstrained vector. In problems where the researcher cannot specify such a $\mathbf{w}$, there are still invertible functions relating shape vectors and the size ratios from a regular sequence. The invariance of the size ratios still applies.

In closing I want to thank the author for his stimulating paper and to endorse his search for transformations to normality.

Mr R. L. Obenchain (Bell Laboratories. USA): Studying methods for fitting distributions to simplex data was one of the first projects that I undertook when I finished graduate school and joined Bell Laboratories in 1969. Since I had been a student of N. L. Johnson in Chapel Hill and since the aN^d distributions I studied could be viewed as generalizations of his S_B curves, I sent a manuscript to Professor Johnson in 1970. I was initially fascinated by the observation that, unlike Dirichlet distributions, certain pairs of proportions can be positively correlated in the aN^d family. But I ultimately became discouraged by what Aitchison refers to as "the problem of zero components" and, thus, never attempted to publish my simplex work.

I personally view the occurrence of zero components as almost fatal to the fitting of aN^d distributions to simplex data. After all, an aN^d density must approach zero, with "high contact", at the boundary of $\mathbb{S}^d$; observing a zero component is not only an event of probability zero but also an impossible event. And moving zero components too small an amount away from the boundary can cause the fitted density to have a local or even global mode in this region. Professor Aitchison's "sensitivity analysis' can be used to avoid modes near a boundary. But, for data sets with many zeros, choices among alternative approaches seem somewhat *ad hoc*.

Professor R. L. Plackett (University of Newcastle upon Tyne): John Aitchison's paper is characterized by his usual originality and elegance. It is another illustration of the rule that the best developments in statistical theory have come in answer to practical needs.

He points out in Section 4.2 that the compositions arise from actual bases in the form of discrete or continuous measurements. The total size may be a count, for example of pebbles or of expenditure in units of Hong Kong cents. There is consequently some interest in comparing the transformed normal class fN^d with multinomial distributions $\mathscr{M}$ based on probabilities $p_1, p_2, \ldots, p_{d+1}$ which are fixed for a given total size but may vary when the size does. Asymptotically, there is no conflict between fN^d and M because the sample percentages under $\mathscr{M}$ have a singular multivariate normal distribution which is transformed into a non-singular N^d by any of the methods to which he refers in Section 2.3. The same result holds when the multinomial probabilities vary in a fixed distribution, and so is not dependent on $\mathscr{M}$ in particular.

The class fN^d may also have something to offer in the analysis of individuals grouped into families (Altham, 1976; Plackett and Paul, 1978). Several of the models here can be generated from moments of Dirichlet distributions, and a natural extension would be to consider the corresponding moments of fN^d. Explicit moments seem not so readily available as for Dirichlet but the effort could be worthwhile in view of the variety of shapes disclosed by Fig. 1. The applications of such models to contingency tables need to take account of a possible ordering in the categories. No such orderings appear in the examples from geology and consumer demand. Two questions arise: first, whether examples exist in which the categories of a composition are ordered; and secondly, what type of transformed distributions would then supply an appropriate model.

The need to deal with sampling zeros is another challenge to the practical statistician to do better in his modelling. It looks as if a distribution over the boundary of the simplex is required as well as one which accounts for the interior.

Dr D. A. Preece, Dr R. Webster and Dr J. A. Catt (Rothamsted Experimental Station): Relations among the proportions of rock and soil constituents are important to earth scientists. They are used for matching and classifying complex natural materials, and for identifying spatial trends that reflect temporal trends in the processes of rock and soil formation. Earth scientists must therefore have reliable estimates of interrelationships, and are often aware of the limitations of correlation coefficients. Professor Aitchison's paper arouses interest as it seems to offer a new method without these limitations. Nevertheless some aspects worry us, and we should value Professor Aitchison's comments and advice.

The first aspect concerns the errors in original data. Professor Aitchison's Example 1 is fairly typical of rock analyses using X-ray fluorescence spectrometry (XRF). Percentages of ten constituents are recorded, and their sums range from 96·97 to 100·66 per cent with an average of 98·81 per cent. The discrepancies from 100 per cent are in most specimens larger than the total amounts of P_2O_5 and MnO, and they arise for two reasons: errors in the determinations, and the fact that not all the constituents are determined. Further, the determination error for any one constituent using XRF is approximately proportional to its total amount. So the error in the determination of SiO_2 is likely to be larger than the total amount of P_2O_5 or MnO. Other problems arise with particle size analysis (Example 3). Here the percentages of coarse fractions, i.e. sand, are determined by sieving; those for the finer ones, i.e. silt and clay, are obtained by sedimentation and so have different errors. Further, the percentage of one of the constituents, namely silt, is sometimes calculated as the difference between 100 per cent and the sum of the percentages of the others. If this last percentage is small then the errors in the other determinations will make its estimate subject to a large proportionate error. Professor Aitchison's proposed method seems not to allow for such differences in errors; we should like to know whether it can be made to do so, and to what extent any failure to do so is likely to matter in practice.

One of the aims of analysis is to identify real associations while avoiding drawing unwarranted conclusions, especially from negative correlation coefficients that are almost inevitable for major constituents. Correlation coefficients calculated for minor components of mixtures, or even between some minor and major components, may well provide useful information. In fact, any strong correlations among constituents in the Skye lavas (Example 1) are likely to reflect changes in the composition of the molten material with time, some elements becoming enriched together while others are depleted together as its composition changes. In the correlation matrix (Table D1) for Example 1, some correlations stand

Table D1

Correlation coefficients for the 32 Skye lava specimens of Thompson et al. (1972)

SiO_2										
Al_2O_3	+0·09									
Fe_2O_3	−0·82	+0·23								
MgO	−0·29	−0·77	−0·05							
CaO	+0·39	+0·64	−0·32	−0·54						
Na_2O	−0·18	+0·11	+0·53	−0·14	−0·40					
K_2O	+0·63	−0·02	−0·47	−0·45	+0·14	−0·04				
TiO_2	−0·90	−0·04	+0·87	+0·06	−0·42	+0·32	−0·42			
P_2O_5	−0·61	+0·15	+0·78	−0·28	−0·30	+0·47	+0·02	+0·84		
MnO	−0·27	+0·73	+0·53	−0·57	+0·46	+0·02	−0·05	+0·39	+0·54	
	SiO_2	Al_2O_3	Fe_2O_3	MgO	CaO	Na_2O	K_2O	TiO_2	P_2O_5	MnO

out as being large and making sound geological sense. For example, the large positive correlation between Fe_2O_3 and TiO_2 almost certainly results from both these elements occurring mainly in the mineral titaniferous magnetite. The large negative correlation between Al_2O_3 and MgO probably reflects early crystallization of magnesium olivine (which contains no Al_2O_3) from the molten rock material. Other large correlations exist and need to be explained. All these relationships suggest that the correlation matrix, despite its perils, can be a valuable tool for exploring compositional data with many constituents. Again, we should value Professor Aitchison's comments.

We should also be glad if Professor Aitchison could explain his method further, for geologists unfamiliar with advanced mathematics, by giving details of the interpretation of one of his examples as a model for geologists to use with other sets of data.

Dr D. M. TITTERINGTON (University of Glasgow): The inability of the Dirichlet distribution to represent non-spurious correlations among probabilities has caused a problem in developing methods, for the smoothing of multinomial data, more sophisticated than those referred to briefly in Section 7.4. If $\mathbf{r} \in \mathbb{S}^k$, or its closure, denotes a set of relative frequencies,, with expected value $\boldsymbol{\theta}$, and if $\mathbf{c} \in \mathbb{S}^k$ is given then, for $0 < \lambda < 1$,

$$\hat{\boldsymbol{\theta}}(\lambda) = \lambda \mathbf{r} + (1-\lambda)\mathbf{c} \tag{*}$$

represents a smoothing of $\mathbf{r}$ towards $\mathbf{c}$. The value of λ can be chosen to minimize, for instance, a mean squared error criterion and, ultimately, a data-based procedure can be developed with properties superior to those of $\mathbf{r}$. There is, as hinted by the form of (*), a strong Bayesian connection, with $\mathbf{c}$ representing a prior mean for $\boldsymbol{\theta}$ and λ indicating the relative importance of the data. The connection is complete if the prior distribution for $\boldsymbol{\theta}$ is Dirichlet (Fienberg and Holland, 1973).

The problem arises if we want to build in some feeling of prior correlations among the components of $\boldsymbol{\theta}$. If, for instance, our objective is to smooth relative frequencies where the cells correspond to categories ordered in the sense that the probabilities on nearby cells are likely to be similar, then we should like this to be reflected in the prior. An obvious example is histogram data. As mentioned in tonight's paper, Leonard (1973) developed the logistic-Normal approach to the problem and his Bayesian analysis can be modified to give a smoothing formula of the type

$$\hat{\mathbf{w}}(\lambda) = \{\lambda \Sigma^{-1} + (1-\lambda)\Sigma_0^{-1}\}^{-1}\{\lambda \Sigma^{-1}\mathbf{y} + (1-\lambda)\Sigma_0^{-1}\mathbf{w}_0\},$$

where $\hat{w}_i = \log\{\hat{\theta}_i/\hat{\theta}_{k+1}\}$, $y_i = \log\{r_i/r_{k+1}\}$, $i = 1, \ldots, k$, Σ denotes the data covariance matrix and $\mathbf{w}_0$, Σ_0 contain prior means and correlations. A data-based choice for λ can be made, as with (*). This formula is based on the approximation $\mathbf{y} \sim N^k$. If, instead, we consider the frequency data, $\mathbf{n}$, say,, and not the *relative* frequencies, and take $y_i = \sqrt{n_i}$, to be a set of independent $N(w_i, \frac{1}{4})$ random variables, where $w_i \propto \sqrt{\theta_i}$, $i = 1, \ldots, k+1$, then the above formula gives an alternative smoothing procedure with $\Sigma = \frac{1}{4}\mathbf{I}_{k+1}$. Prior correlations can be chosen to reflect the ordering of the categories and a natural choice for $(\mathbf{w}_0)_i$ is

$$(\mathbf{w}_0)_i = \sum_{j=1}^{k+1} n_j/(k+1),$$

for each i.

A study of these techniques has shown that the smoothing certainly helped although, for the examples looked at, a kernel-based method could usually be found which was better (Titterington and Bowman, 1982).

I have been rather lazy in not checking the literature but I wonder if multivariate use of the Normal approximation to the Poisson, using the square root transformation, has proved, or might prove, useful, at least in extrinsic analysis. A nice feature of being able to use Normal-based methods is the availability of techniques for dealing with missing values (Dempster *et al.*, 1977, Section 4.1). The Poisson-based approximation will be slightly more amenable than the logistic-Normal in this context.

The AUTHOR replied briefly at the meeting and subsequently more fully in writing as follows.

I should like to express my sincere thanks to all the discussants for their kind and encouraging remarks. Although statisticians are trained to cope with uncertainty and variability the determination with which so many ingeniously battled their way to the Goldsmiths' Theatre against the combined vagaries of man and nature calls for admiration as well as thanks. Since a number of discussants raise the same or similar issues the most convenient way to reply is by subject matter rather than by discussant.

Historical. Mr Obenchain is too modest in his contribution. His internal Bell Laboratories report of 1970, which he has kindly allowed me to see, contains not only the first explicit definition but also many of the properties of the additive logistic-normal class, as set out, for example, in Aitchison and Shen (1980).

Extrinsic and intrinsic analysis. In questioning the use of these terms Dr Anderson raises a much wider issue, the nature of the often complicated sampling processes (Chayes, 1971, p. 44) whereby compositions or bases are determined. In geostatistics this is certainly an area of study in search of a statistician more geologically competent than myself. For the particular question posed here I think there is a clear answer. The distinction between Dr Anderson's sampling processes 2 and 3 is analogous to that between controlled and natural experimental designs in, for example, regression analysis. Either process allows consideration of compositional invariance through the conditional distribution of composition on basis size and the analysis is properly called extrinsic according to my definition. Even if, in the case of the controlled experiment, interest is in the conditional distribution of basis size on composition, strictly a calibration problem, we can again term the analysis extrinsic since attention is still directed outside the composition.

Measurement error. Several discussants (Drs Anderson, Fisher, Preece *et al.*) raise important questions of how to deal with measurement error and what effect such imprecision may have on inferences. It is a nice feature of compositional data that measurement error is so self-revealing in the breach of the constant-sum constraint. If I had been presenting a paper on statistical diagnosis with the components of the feature vector all subject to a measurement coefficient of variation of 10 per cent I doubt if the question of the effect of this imprecision on the diagnosis would have been raised, though it can be answered (Aitchison and Lauder, 1980).

Imprecision in compositional data can be satisfactorily studied through multiplicative and perturbation error models. If $\mathbf{x}$ and $\mathbf{X}$ denote the vectors of true and observed proportions then the multiplicative error model takes the form

$$X_i = x_i u_i \quad (i = 1, \ldots, d+1), \tag{1}$$

where the u_i are positive error variables such as $\Lambda(0, \sigma_i^2)$, assumed independent of x_i and roughly centred on 1. If the σ_i are all equal then we have essentially equal coefficients of variation for each component measurement, but there is no need to make such an assumption so that the forms of imprecision described by Dr Fisher and by Dr Preece *et al.* can be accommodated. If data are available from repeated measurements on a composition the measurement error can be readily investigated. Even if the data are reported in rescaled form $C(\mathbf{X})$ measurement error can still be easily analysed through the perturbation error model

$$C(\mathbf{X}) = \mathbf{x} \circ \mathbf{v}, \tag{2}$$

where $\mathbf{v} = C(\mathbf{u})$.

Moreover, some forms of independence are equally testable with either true or observed compositions. For example, a reasonable assumption will often be that the components of $\mathbf{u}$ are independent. This implies that the perturbation $\mathbf{v}$ in (2) has complete subcompositional independence, which in turn implies that $C(\mathbf{X})$ has complete subcompositional independence if and only if $\mathbf{x}$ has. A similar approach with a slightly more complicated error model leads to the equivalence of compositional invariance in true and observed bases. In his modelling Dr Anderson adopts an assumption of additive errors. In compositional data analysis this will always lead to problems. Whether errors are additive or multiplicative is, of course, testable given sufficient replicate data, though I suspect it may prove difficult to reach any firm conclusions in practice about the nature of the error.

Zeros. There is little I can add to Section 7.4 and the constructive points raised by several discussants. In each application the nature of each zero must be thoroughly investigated, to determine whether it is a trace, measurement error rounding or an essential zero. To the extent that the answers to these questions determine the approach (sensitivity analysis, conditional modelling, etc.) the methodology could be described as *ad hoc*. I do not see how it could be otherwise.

Where there is a substantial frequency of zeros in any component I see no alternative to modelling with mass probability at zero and conditional distributions. Such an approach has had some practical success in the related problems of lognormal modelling; see, for example, Aitchison (1955). Mr Jørgensen's suggestion of the use of the Whitmore defective inverse Gaussian distribution is of this form. His proposal, however, leaves me with a number of doubts. His model for a basis seems to be defined in terms of the marginals for each component. While this may present no problems as far as basis independence is concerned, there are two questions which require answering before I can see this as a rival model. What *multivariate* inverse Gaussian form would he suggest for a basis with dependent components? What is the distribution of the composition formed from such a basis?

Yet another possible approach to zeros is through Box–Cox transformations. If we are sure that x_{d+1}, say, will always be positive, for example the proportion of expenditure on food, then we may consider modelling through

$$y_i = \{(x_i/x_{d+1})^\lambda - 1\}/\lambda \quad (i = 1, \ldots, d),$$

taking $\mathbf{y}^{(d)}$ to be N^d. This will accommodate zeros and may be adequate for all descriptive purposes. In the consideration of the various independence concepts, however, it completely loses all the tractability advantages stemming from the logarithmic function.

Symmetric approach. Since submitting my paper I have been investigating a symmetric approach with $z^{(d+1)} = \log\{x^{(d+1)}/g(\mathbf{x})\}$, where $g(\mathbf{x}) = (x_1 \ldots x_{d+1})^{1/(d+1)}$ is the geometric mean of the components, so that the constraint $\Sigma x_i = 1$ is replaced by the constraint $\Sigma z_i = 0$. Thus we may consider $\mathbf{z}^{(d+1)}$ to be $N^{(d+1)}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\mathbf{u}_{d+1}^{\mathrm{T}} \boldsymbol{\mu} = 0$ and $\boldsymbol{\Sigma}$ singular of rank d. The choice thus seemed to be between persuading clients to use pseudo-inverses in a symmetric approach or convincing them of the invariance of the asymmetric procedures. I felt that the second was on the whole the simpler choice. Professor Dawid's masterly demonstration of how to avoid the singularity problem in linear modelling of the logistic-normal mean I find convincing. Unfortunately there are many procedures where the inverse of the covariance matrix is necessarily involved, such as in density function estimation and discriminant analysis, so that my dilemma persists. Moreover the symmetric approach does not seem particularly suited to the study of some hypotheses of independence. For example, in the symmetric approach the form of the covariance matrix corresponding to complete subcompositional independence is

$$\left(\mathbf{I}_{d+1} - \frac{1}{d+1}\mathbf{U}_{d+1}\right)\operatorname{diag}(\lambda_1, \ldots, \lambda_{d+1})\left(\mathbf{I}_{d+1} - \frac{1}{d+1}\mathbf{U}_{d+1}\right)$$

which even with some simplification is much more difficult to handle than my simple asymmetric form (5.1).

Professor Leonard's analysis of basis independence, in particular his form (3) for $\operatorname{cov}(\log \mathbf{x}^{(d+1)})$ for a composition with basis independence, seems identical to that in Aitchison (1981a); compare the form on p. 179. I also do not see how his points subsequent to (4) differ in substance from the asymmetric approach to testing in my 1981 paper since the particular contrast-producing matrix $\mathbf{A}$ selected must introduce asymmetry. I chose the particular $\mathbf{A} = [\mathbf{I}_d, -\mathbf{u}_d]$, leading to the simple logratio contrasts. Nor do I understand the appeal of basis independence. Its algebra is marginally simpler, the statistical tests are virtually identical, but the interpretation with basis independence is infinitely more difficult. If Professor Leonard doubts this he should follow the history of pitfalls in the geological literature. I can only repeat the question. If there is no real basis why invent one when a concept equivalent to basis independence can be defined within the composition; and if there is a real basis is there any need to involve compositions in the investigation of basis independence?

Principal component analysis. Professor Dawid and Dr Thompson observe that a symmetric approach using the singular matrix $\operatorname{cov}[\log\{\mathbf{x}^{(d+1)}/g(\mathbf{x}^{(d+1)})\}]$, where $g(\cdot)$ denotes geometric mean, is well suited to principal component analysis. The eigenvectors associated with the d non-zero eigenvalues are exactly the eigenvectors found by my asymmetric method which requires the introduction of the interesting concept of an isotropic covariance structure $\mathbf{H}_d = \mathbf{I}_d + \mathbf{U}_d$ and consideration of

$$[\operatorname{cov}\{\log(\mathbf{x}^{(d)}/x_{d+1})\} - \lambda \mathbf{H}_d]\,\mathbf{a} = \mathbf{0}.$$

I had already come to this conclusion in a paper under consideration by *Biometrika.* Indeed there is a further advantage in the symmetric approach here in that it leads naturally to a method of quantifying the amount of the overall variability retained by the commonly practiced procedure of examining only a subcomposition.

I was delighted by Dr Howarth's disapproval of the practice of some geologists who draw "trend" lines through data such as in Fig. 1 and then impose an interpretation along the trend. Geologists have also applied principal component analysis to such data sets but these all use straight line axes quite unsuited to many of their data sets. Logcontrast principal component analysis applied to Fig. 1 produces a first principal axis which nicely follows the curvature of the data. Such curved data in the simplex do not in themselves indicate any trend: they are no different in nature from a typical elliptical cluster in $\mathbb{R}^2$.

Stochastic models. Professor Cox's example has counterparts in other disciplines, such as the compositions of fossil pollens or foraminifera at different levels of a core sample. One possible approach here is to investigate the process $\mathbf{y}(t) = \log\{\mathbf{x}^{(d)}(t)/x_{d+1}(t)\}$ as a multivariate, possibly Gaussian, process in $\mathbb{R}^d$.

Shape and size analysis. Dr Mosimann seems to chide me for not turning my paper into a general analysis of shape and size problems. There would have been little point in so doing because of the full accounts already given in the papers I cited, Mosimann (1970, 1975a, b), and now excellently summarized in his contribution to the discussion. Where I disagree fundamentally with Dr Mosimann is the extent to which shape and size analysis has helped or hindered the analysis of compositional data. I can only summarize two main points here.

(1) There are many issues concerning compositional data where questions of size play no part, for example, in geochemical compositions where complete or partial subcompositional independence is important or the purpose is discrimination, as in Section 4.3 of Aitchison and Shen (1980). To insist on incorporating such procedures within the scope of size-and-shape analysis is only to complicate what is already simple.

(2) Where size is of interest in compositional data it is invariably additive size which is under consideration, whether the size of a basis as in compositional invariance or the size of a subvector of the composition as in subcompositional invariance. I question Dr Mosimann's implication that shape-and-size analysts have already used, even implicitly, logistic-normal distributions over the simplex for proportions. Just because they use logratio shapes in *multiplicative* size situations seems to me an extremely unconvincing argument since the size constraint does not then confine the shape vector to the simplex. If, as he seems to claim, logistic-normals are "old hat" to size and shape analysis a number of repeatedly reported deficiencies of that theory become even more puzzling: the degeneracy of the only model proposed for the investigation of additive isometry, insistence that there is a shortage of classes of distributions over the simplex other than Dirichlet and its simple generalizations, the non-emergence of testing procedures for hypotheses of neutrality.

Relation to multinomial theory. Professor Plackett's problems in multinomial theory are extremely interesting. To his specific questions about ordered categories the only answer I can presently provide is that the forms of analysis in Section 7.2 involve an ordering of the components $(x_1, x_2, ..., x_{d+1})$ of the composition and for such orderings the logratio transformation

$$y_1 = \log\{x_1/(1-x_1)\}, \quad y_2 = \log\{x_2/(1-x_1-x_2)\}, ...$$

proves useful in compositional data analysis. Perhaps also the involvement of ordering in Dr Titterington's interesting comments on smoothing procedures may be of some relevance to this problem. His linking of missing values and Poisson-based approximations is appealing and should certainly be pursued. I suspect that multinomial theory may have many more contributions to make to compositional data analysis. For example, I have so far been unable to find a model for a two-way compositional distribution yielding tractable distributions for *both* row and column marginal compositions. Does the answer lie somewhere in multinomial theory?

Other transformations. Professor Stephens proposes a transformation that goes from the d-dimensional positive simplex to the positive orthant of the surface of the d-dimensional sphere, a device already advocated to me by other directional data specialists at North American seminars. While this no doubt provides a means of describing variability and so allows comparisons of the type discussed by Professor Stephens and other procedures such as discriminant analysis, the sphere is a difficult space in which to discuss independence and regression, as acknowledged by him. Moreover the fact that the surface of the sphere and the simplex are topologically different does limit the transformation to only part of the sphere, such as the positive orthant, and this limitation can prove a source of difficulty. Dr Atkinson's transformation, if converted into a one-to-one form, would be equivalent to going to the positive orthant of the sphere, followed by a spherical polar transformation, and so is subject to similar problems. Indeed it could be argued that the worker on the sphere has more to gain from visiting the simplex than the other way around. For example, consideration of moving from the sphere to the simplex has led me to suggest the use of logtan-normal, or more properly exponential-inverse-tan normal, distributions on the sphere. For example, for $d = 2$, with ϕ denoting longitude and θ latitude and confining attention to the positive orthant, we might consider joint distributions of (ϕ, θ) for which $(\text{logtan}\,\phi, \text{logtan}\,\theta)$ is $N^2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. The independence of longitude and latitude would correspond to the parameteric hypothesis $\sigma_{12} = 0$ of zero correlation in the transformed variables.

Dirichlet generalizations. A number of discussants (Professors Darroch, Goodhardt; Drs Kent, Mosimann) still see hope in some forms of generalization of the Dirichlet class. For the moment, I feel a more hopeful line of enquiry is to investigate more fully the nature of the differences between logistic-normal and Dirichlet distributions. Aitchison and Shen (1980) showed that for many Dirichlet distributions there is a closely approximating logistic normal distribution; this fact has indeed been exploited in the construction of a test of Dirichlet against logistic-normal form by Shen (1982). Some of the tests of independence such as complete subcompositional independence when they lead to rejection also automatically reject the Dirichlet form.

The case $d = 2$. Professor Darroch picks up my point in Section 5 that some forms of independence are trivially satisfied and so of no interest for the case $d = 2$. There are reasons for this. For $d = 1$ there are no questions of interest about independence: the mathematical dependence $x_2 = 1 - x_1$ ensures statistical dependence. For $d = 2$ the effect of the constraint is still stifling (for example, the covariances between raw proportions are completely determined by the variances of the raw proportions) so that not all concepts of independence will be applicable. Only for $d \geqslant 3$ is it possible to see a wider range of definitions, and, to some extent, concentration on $d = 2$ has prevented this wider view.

For $d = 2$ the main, possibly the only, forms of independence are of the $x_1 \perp\!\!\!\perp C(x_2, x_3)$ form, namely neutrality for particular ordering or subcompositional invariance. As a practical means of investigating data sets for $d = 2$, I would suggest the following procedure. Test the data for Dirichlet form. If the Dirichlet form is rejected then there is non-neutrality of at least one of forms $x_1 \perp\!\!\!\perp C(x_1, x_3)$, $x_2 \perp\!\!\!\perp C(x_1, x_3)$, $x_3 \perp\!\!\!\perp C(x_2, x_3)$. Use the neutrality tests described in Section 7.2 to obtain some indication of the nature of the non-neutrality.

The constraint. A number of discussants seem to have a persistent worry about the effect of the constraint after the transformations. A main objective of the transformations advocated is to ensure a form of modelling which specifically takes full account of the constraint so that the question of continuing constraint effects does not arise. If one accepts a particular transformation as valid then the constraint can be forgotten. One may, of course, question the appropriateness of a particular transformation; but that is a different matter.

Applications. It was interesting to see that the application to activity patterns considered by Professor Stephens is similar to that of the original application by Mr Obenchain in his 1970 report. I have myself used another example, 21 days in the life of a statistician divided into the three activities of work (W), sleep (S) and general grubbing around (G). My purpose was to use the data in the GSW triangle, rather like the curved set in Fig. 1, to show the absurdity of reading trend into such curvatures. A well-known example of this form of analysis is the diaries which a sample of British academics were asked to keep some years ago.

Professor Goodhardt's problems of consumer choice are intriguing. His hypothesis (i) $T(\boldsymbol{\beta}_j) = \gamma_j \, (j = 1, ..., k+1)$ of no partition within Dirichlet modelling falls readily within the scope of standard parametric hypothesis testing, for example through a generalized likelihood ratio test. His more complex problem (iii) would involve the testing of separate families in the sense of Cox (1962) and could prove much more difficult because of the high dimensionality of the parameter vector. His partition-selection problem (ii) would seem to depend largely on a satisfactory resolution of problem (iii).

I think that my asymmetric approach has led Professor Leser into a misunderstanding of x_{d+1}, which is on an equal footing with $x_1, ..., x_d$ as a proportion of total expenditure. The point in using total expenditure rather than income is to avoid having to consider the nastiness of negative saving which can hopefully be incorporated into some form of conditional analysis as outlined in Section 4.2.

I may be misinterpreting the genetic application of Fisher cited by Professor Smith, but it seems to me an early example of the use of the generalized logistic function in multinomial modelling and not an example of compositional data.

I am greatly encouraged by the comments of Dr Howarth and Dr Preece *et al* on the viability of the techniques for geological applications. I hope that my comments under measurement error and principal component analysis have gone some way to reassure them that the system is flexible enough to cope, particularly with measurement errors. I am currently working on an expository paper on geological applications and hope that this may meet the request made by Dr Preece and his colleagues.

On the use of the correlation matrix between *raw* proportions I can only comment that all the evidence of the past two decades suggests that it leads to more problems than it resolves. For example, Miesch (1969) demonstrates that an apparently significant raw correlation between two oxides may have really arisen, through the closure process, from an actual correlation between two other oxides! The use of $\operatorname{cov}(\mathbf{x}^{(d+1)})$, in my view, is a symptom of the barbecue syndrome. The barbecue is a very effective instrument if you are enjoying the wide open spaces ($\mathbb{R}^d$) of, say, North America; but would you continue to use it if you were suddenly confined to an $\mathbb{S}^d$ housing unit in Hong Kong? Some transformation of the barbecue would obviously be desirable. You might end up preparing the ingredients of your problems so that they can be stir-fried in that rather different, but equally effective, instrument, the wok.

References in the Discussion

Aitchison, J. (1955). On the distribution of a positive random variable having a discrete probability mass at the origin. *J. Amer. Statist. Ass.*, **50**, 901–908.

Aitchison, J. and Lauder, I. J. (1979). Statistical diagnosis from imprecise data. *Biometrika*, **67**, 475–483.

Altham, P. M. E. (1976). Discrete variable analysis for individuals grouped into families. *Biometrika*, **63**, 263–269.

Atkinson, A. C. (1969). Constrained maximization and the design of experiments. *Technometrics*, **11**, 616–618.

CORNELL, J. A. (1981). *Experiments with Mixtures: Designs, Models and the Analysis of Mixture Data.* New York: Wiley.

COX, D. R. (1954). Some statistical aspects of mixing and blending. *J. Text. Inst.*, **45**, T113–122.

—— (1962). Further results on tests of separate families of hypotheses. *J. R. Statist. Soc.* B, **24**, 406–424.

DAWID, A. P. (1977). Conformity of inference patterns. In *Recent Developments in Statistics* (J. R. Barra *et al.*, eds), pp. 245–256. Amsterdam: North-Holland.

DEMPSTER, A. P., LAIRD, N. M. and RUBIN, D. B. (1977). Maximum likelihood from incomplete data via the *EM* algorithm (with Discussion). *J. R. Statist. Soc.* B, **39**, 1–38.

DRAPER, N. R. and ST JOHN, R. C. (1977). A mixtures model with inverse terms. *Technometrics*, **19**, 37–46.

FIENBERG, S. E. and HOLLAND, P. W. (1973). Simultaneous estimation of multinomial probabilities. *J. Amer. Statist. Ass.*, **68**, 683–691.

FISHER, R. A. (1947). The fitting of gene frequencies to data on rhesus reactions. *Ann. Eugen.*, **13**, 150–153.

GOULD, S. J. (1977). *Ontogeny and Phylogeny.* Cambridge, Mass.: Belknap Press.

GRIFFITHS, R. C. (1970). Infinitely divisible multivariate gamma distributions. *Sankhya* A, **32**, 393–404.

HUXLEY, J. (1932). *Problems of Relative Growth.* New York: Dial Press.

KRISHNAMOORTHY, A. S. and PARTHASARATHY, M. (1951). A multivariate gamma-type distribution. *Ann. Math. Statist.*, **22**, 549–557.

LESER, C. E. V. (1963). Forms of Engel functions. *Econometrica*, **31**, 694–703.

LINDLEY, D. V. (1964). The Bayesian analysis of contingency tables. *Ann. Math. Statist.*, **35**, 1622–1643.

MIESCH, A. T. (1969). The constant sum problem in geochemistry. In *Computer Applications in the Earth Sciences* (D. F. Merriam, ed.), pp. 161–177. New York: Plenum Press.

MOSIMANN, J. E. and JAMES, F. C. (1979). New statistical methods for allometry with application to Florida red-winged blackbirds. *Evolution*, **33**, 444–459.

MOSIMANN, J. E., MALLEY, J. D., CHEEVER, A. W. and CLARK, C. B. (1978). Size and shape analysis of schistosome egg-counts in Egyptian autopsy data. *Biometrics*, **34**, 341–3456.

PLACKETT, R. L. and PAUL, S. R. (1978). Dirichlet models for square contingency tables. *Commun. Statist.*, **A7**, 939–952.

REEVE, E. C. R. and HUXLEY, J. (1945). Some problems in the study of allometric growth. In *Essays on Growth and Form* (W. E. Le Gros Clark and P. B. Medawar, eds), Oxford: Clarendon.

SHEN, S. M. (1982). A method for discriminating between models describing compositional data. *Biometrika* (in the press).

STEPHENS, M. A. (1980). The von Mises distribution in p dimensions, with applications. Technical Report, Department of Statistics, Stanford University.

——— (1982). Use of the von Mises distribution to analyse continuous proportions. *Biometrika*, **69**, 197–203.

ST JOHN, R. C. and DRAPER, N. R. (1977). Designs in three and four components for mixtures models with inverse terms. *Technometrics*, **19**, 117–130.

TITTERINGTON, D. M. and BOWMAN, A. W. (1982). A comparative study of smoothing procedures for ordered categorical data. Submitted for publication.

WATSON, G. S. (1956). Analysis of dispersion on a sphere. *Mon. Not. Roy. Astron. Soc., Geophys. Suppl.*, **7**, 153–159.

WHITMORE, G. A. (1978). Comments in discussion of paper by Folks and Chikara. *J. R. Statist. Soc.* B, **40**, 285–286.