Logistic-Normal Distributions: Some Properties and Uses
Author(s): J. Aitchison and S. M. Shen
Source: *Biometrika*, Vol. 67, No. 2 (Aug., 1980), pp. 261-272
Published by: Oxford University Press on behalf of Biometrika Trust
Stable URL: http://www.jstor.org/stable/2335470
Accessed: 25-02-2016 18:45 UTC

Biometrika (1980), **67**, 2, *pp*. 261–72
Printed in Great Britain

Logistic-normal distributions: Some properties and uses

BY J. AITCHISON AND S. M. SHEN
Department of Statistics, University of Hong Kong

SUMMARY

The logistic transformation applied to a d-dimensional normal distribution produces a distribution over the d-dimensional simplex which can sensibly be termed a logistic-normal distribution. Such distributions, implicitly used in a number of recent applications, are here given a formal identity and some useful properties are recorded. A main aim is to extend the area of application from the restricted role as a substitute for the Dirichlet conjugate prior class in the analysis of multinomial and contingency table data to the direct statistical description and analysis of compositional and probabilistic data.

Some key words: Compositional data; Directed divergence measure; Dirichlet distribution; Logistic discrimination; Logistic-normal distribution; Log normal distribution; Multiple contingency table; Probabilistic data.

1. DEFINITION

Most statisticians if asked to list the distributions they know over the unit interval $(0, 1)$ and its generalization, the d-dimensional simplex S^d, would start and end with the class of beta distributions and their higher-dimensional counterpart, the Dirichlet distributions. Another useful and richer such class, the logistic-normal distributions, has arisen usually by implication in a few widely differing applications:

(i) for the case $d = 1$ as a Johnson (1949) four-parameter log normal distribution with the two range parameters determining the interval $(0, 1)$;

(ii) in the Bayesian analysis of multinomial and contingency table data in the use of normal approximations to log contrasts by Lindley (1964), by C. Swe in a Liverpool Ph.D. dissertation, and by Bloch & Watson (1967); and as the first stage in the construction of exchangeable prior distributions by Leonard (1973);

(iii) in studies of size and shape in biological allometry, for example by Mosimann (1975), as the distribution of ratios of log normally distributed measurements;

(iv) in statistical diagnosis where classification of the basic cases is subject to uncertainty, as discussed by Aitchison & Begg (1976), who provide an explicit definition of the class of logistic-normal distributions;

(v) in the reconciliation of subjective probability assessments, where Lindley, Tversky & Brown (1979) use normal log-odds models to describe assessments.

The class has, however, gained no clear identity. Our purpose here is to provide such an identity, to describe enough of its properties and a sufficient variety of new applications to encourage further exploration of what, in our view, is a promising tool of statistical analysis.

Let R^d denote d-dimensional real space, P^d the positive orthant of R^d and S^d the d-dimensional positive simplex defined by

$$S^d = \{u \in P^d \colon u_1 + \ldots + u_d < 1\}.$$

For any d vector u and any real-valued function f, let $f(u)$ denote the d vector with ith component $f(u_i)\,(i = 1, \ldots, d)$. Suppose that v follows a multinormal distribution $N_d(\mu, \Sigma)$

over R^d. The exponential transformation from R^d to P^d, namely $w = e^v$, or its inverse the logarithmic transformation $v = \log w$, is the familiar device for defining a corresponding log normal distribution with w distributed as $\Lambda_d(\mu, \Sigma)$, say. In a similar way the logistic transformation from R^d to S^d, or its inverse log ratio transformation:

$$u = e^v \bigg/ \left(1 + \sum_{j=1}^{d} e^{v_j}\right), \quad v = \log(u/u_{d+1}), \tag{1·1}$$

where

$$u_{d+1} = 1 - \sum_{j=1}^{d} u_j, \tag{1·2}$$

can be used to define a logistic-normal distribution over S^d, and we can say briefly that u is $L_d(\mu, \Sigma)$. The density function of $L_d(\mu, \Sigma)$ is then

$$|2\pi\Sigma|^{-\frac{1}{2}} \left(\prod_{j=1}^{d+1} u_j\right)^{-1} \exp\left[-\tfrac{1}{2}\{\log(u/u_{d+1}) - \mu\}^{\mathrm{T}} \Sigma^{-1} \{\log(u/u_{d+1}) - \mu\}\right] \quad (u \in S^d). \tag{1·3}$$

Note that the density function is defined on the strictly positive simplex. This is necessary because of the logarithmic transformation involved.

We first look at properties of the logistic-normal distribution in § 2, compare it in § 3 with its competitor, the Dirichlet distribution, describe applications in § 4, and finally in § 5 draw some conclusions and point the way to further research and applications.

2. PROPERTIES

2·1. *Compositions*

Throughout this section the d-dimensional random vectors u and v have $L_d(\mu, \Sigma)$ and $N_d(\mu, \Sigma)$ distributions. Most of the properties of logistic-normal distributions derive from corresponding properties of multinormal distributions but usually require adjustment to provide useful practical results. For example, normality of the marginal distribution of $(v_1, \ldots, v_{c+1})$ over R^c does not provide a simple result about the distribution of $(u_1, \ldots, u_{c+1})$ over S^{c+1}, that is about $u_1, \ldots, u_{c+1}$ and the residual $1 - u_1 - \ldots - u_{c+1}$, but rather about the relative structure within the subvector $(u_1, \ldots, u_{c+1})$.

This and other properties are most simply expressed in terms of the concept of the composition and subcompositions of a vector. The composition of any positive $(d+1)$ vector w is the d vector u defined by $u_i = w_i/(w_1 + \ldots + w_{d+1})$ $(i = 1, \ldots, d)$, is written $C(w)$, and is an element of S^d. The composition of any subvector of u, such as $(u_1, \ldots, u_{c+1})$ is then a subcomposition of u or of w, and is an element of S^c.

Proofs of the following properties are straightforward and are therefore omitted.

2·2. *The composition of a log normal vector*

If w is $\Lambda_{d+1}(\xi, \Omega)$ then $C(w)$ is $L_d(A\xi, A\Omega A^{\mathrm{T}})$, where the $d \times (d+1)$ matrix $A = [I_d, -e_d]$, I_d is the unit matrix of order d and e_d is a d vector with unit components.

This result is of particular interest in § 4·2 where we study problems concerning the analysis of compositional data.

2·3. *Moment properties*

Although moments of all positive orders $E(u_j^a)$ $(a > 0)$ and the geometric moment $\exp\{E(\log u_j)\}$ exist the integral expressions for them are not reducible to any simple form.

This is no great loss since interest in practice is often more naturally in the ratios u_j/u_k or their logarithms. From normal-log normal theory, with σ_{jk} denoting the (j,k)th element of Σ and with the convention that $\mu_{d+1} = 0$ and $\sigma_{j,d+1} = 0$ $(j = 1, \ldots, d+1)$, we have that

$$E\{\log(u_j/u_k)\} = \mu_j - \mu_k, \quad \operatorname{cov}\{\log(u_j/u_k), \log(u_l/u_m)\} = \sigma_{jl} + \sigma_{km} - \sigma_{jm} - \sigma_{kl},$$

$$E(u_j/u_k) = \exp\{\mu_j - \mu_k + \tfrac{1}{2}(\sigma_{jj} - 2\sigma_{jk} + \sigma_{kk})\},$$

$$\operatorname{cov}(u_j/u_k, u_l/u_m) = E(u_j/u_k)\,E(u_l/u_m)\,\{\exp(\sigma_{jl} + \sigma_{km} - \sigma_{jm} - \sigma_{kl}) - 1\}.$$

2·4. *Class-preserving properties*

The well-known linear transformation property of multinormal distributions, that if v is $N_d(\mu, \Sigma)$ and B is a $c \times d$ matrix then Bv is $N_c(B\mu, B\Sigma B^{\mathrm{T}})$, has the following counterpart in logistic-normal theory. If u is $L_d(\mu, \Sigma)$ then the c vector t, defined by

$$t_i = \prod_{j=1}^{d} (u_j/u_{d+1})^{b_{ij}} \left\{1 + \sum_{i=1}^{c} \prod_{j=1}^{d} (u_j/u_{d+1})^{b_{ij}}\right\}^{-1} \quad (i = 1, \ldots, c), \tag{2·1}$$

is $L_c(B\mu, B\Sigma B^{\mathrm{T}})$. Two special cases of this property are of particular importance.

We first consider the permutation property. In our definition of the logistic-normal distribution over S^d, u_{d+1} was the common divisor in all the ratios of the transformation (1·1). A first application of (2·1) with $c = d$, $b_{ii} = 1$ $(i \neq h)$, $b_{ih} = -1$ $(i = 1, \ldots, d)$, $b_{ij} = 0$ otherwise shows that the d vector t defined by $t_i = u_i$ $(i \neq h)$, $t_h = u_{d+1}$, $t_{d+1} = u_h$ is also of L_d form, with ratio denominator t_{d+1} now effectively the original u_h. This class preservation property is reassuring for any work in the simplex, where obviously it is a matter of no consequence which d of the $d+1$ positive quantities $u_1, \ldots, u_{d+1}$ are chosen to define the simplex of interest.

A second application of (2·1), with B the $d \times d$ permutation matrix associated with the permutation $(1, \ldots, d) \rightarrow (j_1, \ldots, j_d)$ so that $b_{ij_i} = 1$ $(i = 1, \ldots, d)$, $b_{ij} = 0$ otherwise, gives $t_i = u_{j_i}$ $(i = 1, \ldots, d)$. This shows that the logistic-normal form is preserved under a permutation of $u_1, \ldots, u_d$.

Combination of these two preceding results establishes that the logistic-normal class of distributions is closed under the group of permutations of the components $u_1, \ldots, u_d, u_{d+1}$. This is particularly reassuring in many statistical investigations of vector data, where we hope that the analysis is invariant with respect to the ordering of the vector components.

For the subcomposition property, a useful counterpart of the multinormal marginal property can be obtained from (2·1) with B the $c \times d$ matrix defined by $b_{ii} = 1$, $b_{i,c+1} = -1$ $(i = 1, \ldots, c)$, $b_{ij} = 0$ otherwise. Then $t_i = u_i/(u_1 + \ldots + u_{c+1})$ $(i = 1, \ldots, c)$ and the property is simply that the subcomposition $C(u_1, \ldots, u_{c+1})$ is $L_c(B\mu, B\Sigma B^{\mathrm{T}})$.

2·5. *The conditional subcomposition property*

The multinormal conditional property can be adjusted to provide a useful conditional distribution property for subcompositions. Suppose that the subcomposition $C(u_{c+1}, \ldots, u_{d+1})$ is known, expressed most conveniently for our purposes in terms of specified values r_i of u_{c+i}/u_{d+1} $(i = 1, \ldots, d-c)$. The conditional distribution of the subcomposition $C(u_1, \ldots, u_{c+1})$ given the above subcomposition $C(u_{c+1}, \ldots, u_{d+1})$ is then

$$L_c\{\mu_1 - e_c \log r_1 + \Sigma_{12}\Sigma_{22}^{-1}(\log r - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\},$$

where (μ_1, μ_2) is the $(c, d-c)$ partition of μ and $\Sigma_{11}, \Sigma_{12}, \Sigma_{22}$ are the obvious submatrices in the corresponding partition of Σ. From this the conditional distribution of $C(u_1, \ldots, u_c)$ given $C(u_{c+1}, \ldots, u_{d+1})$ can easily be obtained from the subcomposition property of § 2·4.

2·6. *Statistical properties*

For the analysis of compositional data, such as independent vectors $u^{(1)}, \ldots, u^{(n)}$ with each $u^{(i)} \in S^d$ $(i = 1, \ldots, n)$, the class of logistic-normal distributions provides, through its close relationship to the multinormal class, a ready means of tractable statistical analysis. In addition to simple estimation and hypothesis testing of the μ and Σ parameters, tests of logistic-normality, linear modelling of $E(u^{(i)})$ $(i = 1, \ldots, n)$ to take account of experimental design and possible concomitant factors, and all the special multivariate techniques such as discriminant analysis. Bayesian statistical analysis is directly available through the normal-Wishart class of conjugate prior distributions for μ and $\tau = \Sigma^{-1}$. In particular, predictive density functions are easily derived in terms of new and easily defined classes of distributions such as logistic-Student distributions over S^d. We shall see applications of such predictive distributions in § 4.

3. Comparison with the Dirichlet class

3·1. *Closeness*

No study of a proposed class of distributions over the simplex can fail to make comparisons with the well established Dirichlet class, with typical distribution $D_d(\alpha_1, \ldots, \alpha_{d+1})$ or $D_d(\alpha)$ defined by a density function proportional to

$$\prod_{i=1}^{d+1} u_i^{\alpha_i - 1},$$

where $u_{d+1} = 1 - u_1 - \ldots - u_d$ as before.

Aitchison & Begg (1976) suggest that the greater richness of the L_d class with its $\frac{1}{2}d(d+3)$ parameters compared with only $d+1$ for the D_d class, may allow any Dirichlet distribution to be closely approximated by a suitably chosen logistic normal distribution. Their suggestion can be investigated more quantitatively by means of the Kullback & Liebler (1951) measure of directed divergence of a density function q from a target density function p:

$$I(p, q) = \int_{S^d} p(u) \log \frac{p(u)}{q(u)} du.$$

For $p(u)$ of $D_d(\alpha)$ form the closest $q(u)$ of $L_d(\mu, \Sigma)$ form in the sense that $I(p, q)$ is minimized is given by

$$\mu_i = \delta(\alpha_i) - \delta(\alpha_{d+1}), \quad \sigma_{ii} = \varepsilon(\alpha_i) + \varepsilon(\alpha_{d+1}) \quad (i = 1, \ldots, d), \quad \sigma_{ij} = \varepsilon(\alpha_{d+1}) \quad (i \neq j), \qquad (3\cdot1)$$

where $\delta(x) = \Gamma'(x)/\Gamma(x)$ and $\varepsilon(x) = \delta'(x)$ are the digamma and trigamma functions.

For $D_d(\alpha)$ with $d = 1, 2, 3$, and with components of α in the range 5 to 100, the minimized divergences range from 2×10^{-6} for $\alpha = (5, 5)$ to 5×10^{-2} for $\alpha = (5, 5, 5, 100)$.

Some indication of the degree of closeness of these approximations can be provided by directed divergences for more familiar situations. For example, the directed divergence of a $N(\lambda, 1)$ from the $N(0, 1)$ distribution also ranges from 2×10^{-6} to 5×10^{-2} as λ ranges from 0·002 to 0·316.

We can also try to judge success for given $D_d(\alpha)$ by finding a neighbouring $D_d(\beta)$ distribution with the same directed divergence as the minimized logistic-normal divergence. Confining attention to neighbouring distributions with α differing from β in a single component we have the following results. When the components of α are equal the increase in a single α_i never exceeds 0·6. When the components are unequal, the more asymmetrical the component values are, the greater, roughly speaking, is the increase in these single component values, the greatest increases being 0·7 in 5, 1·4 in 20 and 8·6 in 100, the last occurring for $d = 3$ and $\alpha = (5, 5, 5, 100)$.

Whether or not such results are acceptable, approximations must clearly depend on the particular application. We shall see in § 4·1 that minimized-divergence logistic-normal distributions are indeed close to an already widely accepted approximation to Dirichlet distributions.

3·2. *Comparison of properties*

Some comparison of the Dirichlet and logistic-normal classes with respect to the properties of § 2 is required.

The Dirichlet composition property analogous to § 2·2 relates a Dirichlet-distributed composition u to a $(d+1)$ vector with gamma-distributed components (Wilks, 1962, p. 179). The fact that the gamma components are independent and have equal scale parameters indicates that the components of a Dirichlet composition have a special and near-independence structure, with correlations between components arising solely from the division by the common Σw_i in the formation of the composition u. Thus Dirichlet distributions may be too simple to be realistic in the analysis of compositional data where the underlying w_i's are dependent.

The Dirichlet class has simple analogues to the class-preserving properties of § 2·4. Note that the subcomposition property of § 2·4 is not a result about $(u_1, \ldots, u_{c+1}, 1-u_1-\ldots-u_{c+1})$ but about the distribution of $u_1/(u_1+\ldots+u_{c+1}), \ldots, u_c/(u_1+\ldots+u_{c+1})$. In other words the logistic-normal subcomposition property is not a class-preserving property allowing addition of components of a composition. The Dirichlet does possess such a component-additive property (Wilks, 1962, p. 181): for example $(u_1, \ldots, u_{c+1}, 1-u_1-\ldots-u_{c+1})$ is $D_{c+1}(\alpha_1, \ldots, \alpha_{c+1}, \alpha_{c+2}+\ldots+\alpha_{d+1})$. This is, however, a direct consequence of the compositional relationship to independent gamma variables and so, as an advantage over the logistic-normal class, may be buying mathematical elegance at the price of realism. Moreover, if Dirichlet distributions are truly appropriate to an analysis involving additions of compositional data then the logistic-normal distributions that we mistakenly use may yet prove to be satisfactory understudies through the closeness property.

The Dirichlet counterpart of the conditional subcomposition property of § 2·5 reinforces this caution in the use of Dirichlet distributions in the analysis of compositional data. For the Dirichlet distribution the conditional distribution of $C(u_1, \ldots, u_c)$ given $C(u_{c+1}, \ldots, u_{d+1})$ is the same as the unconditional distribution $C(u_1, \ldots, u_c)$. In other words in Dirichlet modelling $C(u_1, \ldots, u_c)$ and $C(u_{c+1}, \ldots, u_{d+1})$ are independent, a very strong assumption to impose, without investigation, on the nature of any compositional data.

Although the Dirichlet class may possess admirable qualities of mathematical tractability in its role as the conjugate prior class for the Bayesian analysis of multinomial and contingency table data, and as an essential tool in the determination of distribution-free statistical tolerance limits it has many disadvantages as a direct describer of patterns of variability. Maximum likelihood estimation of α, for example, requires solving equations involving digamma functions so that Newton–Raphson or some equivalent numerical method is required; and the distributional properties of the estimators must also be approximated. Moreover the absence of a class of conjugate priors makes the possibility of tractable Bayesian analysis and statistical prediction analysis remote.

4. SOME APPLICATIONS

4·1. *Bayesian analysis of contingency tables*

For the $(d+1)$-category multinomial distribution with category probabilities

$$\theta_i \; (i = 1, \ldots, d+1),$$

the conjugate class of distributions on the parameter space S^d is the Dirichlet class. When the multinomial distributions relate to contingency tables then interest is often concerned with contrasts, linear combinations of $\log \theta_i$ such as

$$k_h = \sum_{i=1}^{d+1} c_{hi} \log \theta_i$$

with $\Sigma_i c_{hi} = 0$. From the logistic-normal approximation (3·1) it follows immediately that k_h is approximately distributed as $N\{\Sigma_i c_{hi}\, \delta(\alpha_i),\, \Sigma_i c_{hi}^2\, \varepsilon(\alpha_i)\}$ and moreover that k_g and k_h are approximately distributed binormally with means and variances as determined above and with covariance $\Sigma_i c_{gi} c_{hi}\, \varepsilon(\alpha_i)$.

Bloch & Watson's (1967) approximation to such contrasts, essentially derived from a component by component choice of expressions for means and variances of the logarithm of gamma random variables, uses Stirling's approximation to the log gamma function and so could easily have led to a digamma-trigamma approximation coincident with our own, whose derivation is based on a global approximation to the Dirichlet distribution. Presumably their determination to arrive at approximate means and variances expressible in terms of logarithmic and reciprocal functions was motivated by a wish to avoid digamma and trigamma functions. Our global approximation provides overall support for the component-wise method used by Bloch & Watson. The directed divergence of their approximation from $D_d(\alpha)$ is greater than that of the minimizing logistic-normal distribution (3·1) by less than 0·1% when the components of α are all greater than 2, which will almost always be the case in applications.

4·2. *Analysis of compositional data*

There are many disciplines, for example, sedimentology, petrology, biochemistry, palaeoecology, where interest is in compositional data such as proportions of sand, silt, clay in sediments, of chemical constituents of rocks, of serum proteins in blood, of pollens of different species at different levels in sample borings. For illustrative purposes we here adapt a problem posed by McCammon (1975, p. 162) to demonstrate the simplicity of statistical analysis with logistic-normal tools.

Figure 1 shows in terms of triangular coordinates the sand, silt, clay composition of 17 sediments, 7 of which are identified as nearshore, type I, and the remaining 10 as offshore, type II. Four new samples, all from the same site and hence of the same type, have been analysed and the problem is to assess this unknown type. We adopt $L_2(\mu_1, \Sigma_1)$ and $L_2(\mu_2, \Sigma_2)$ distributions for the nearshore and offshore data. The statistical problem is assumed to be the assessment of a reasonable factor for the conversion of prior odds to posterior odds for type. With such a small data set we adopt a predictive approach to the typing or diagnostic problem, for the advantages argued, for example, by Aitchison, Habbema & Kay (1977).

The unusual feature of this example, in contrast to the more familiar area of application of predictive diagnosis, namely medical diagnosis, is that for the new case we have four replicate observations. For the purposes of predictive diagnosis the predictive problem can then be condensed into obtaining a predictive density function for M and V, the mean vector and matrix of corrected cross-products of the four, in general N, vectors of log ratios. The appropriate theory is summarized by Aitchison & Dunsmore (1975, Table 2·3) and leads to predictive distributions $p(M, V \mid D_i)$ for M and V of Student–Siegel type based on data D_i:

$$\mathrm{St\,Si}_2\left\{n_i-1, m_i, \left(\frac{1}{n}+\frac{1}{N}\right)S_i;\; N-1, (n_i-1)\,S_i\right\},$$

where n_i, m_i and S_i are the number, the mean vector and covariance matrix of the log ratios, of vectors in the data set D_i for type i. In this assessment we have used the vague prior suggested by Aitchison & Dunsmore (1975) for the reasons given by Aitchison (1976). Straightforward computation then gives $p(M, V \mid D_1)/p(M, V \mid D_2) = 0{\cdot}19$ as the converting factor from prior odds to posterior odds on type I. Thus if type I and II are equally likely *a priori* our evidence leads to odds of 5 to 1 in favour of type II.

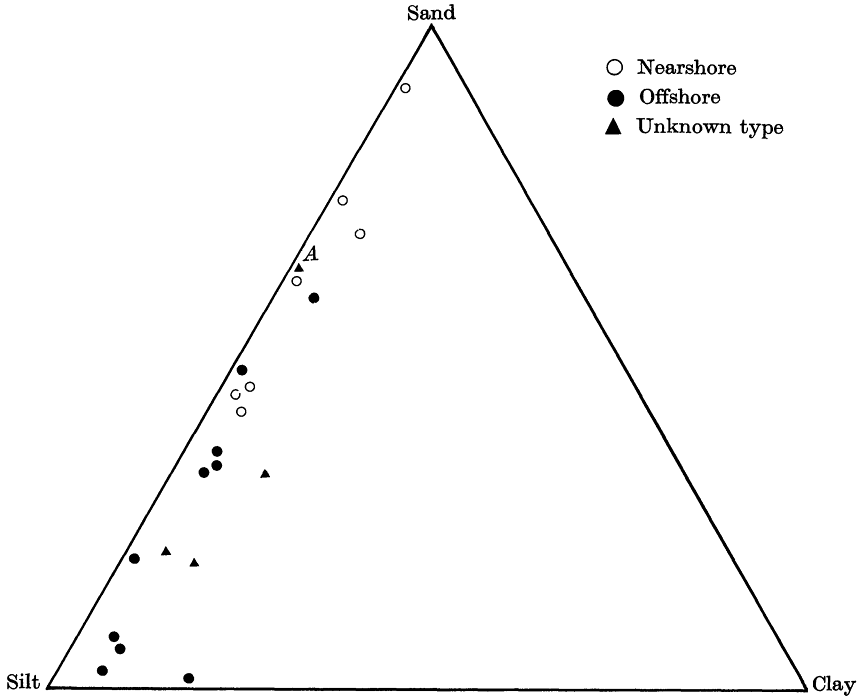

Fig. 1. Sand, silt, clay compositions of 17 sediment specimens of known type and four sediment samples of unknown type.

The predictive method can be contrasted with the estimative method (Aitchison, Habbema & Kay, 1977) which would simply replace the parameters $\mu_1, \Sigma_1, \mu_2, \Sigma_2$ by their estimates. This is equivalent to replacing the previous $p(M, V \mid D_i)$ by a normal-Wishart density function $\mathrm{NoWi}_2(m_i, S_i/N, N-1, S_i)$ as defined by Aitchison & Dunsmore (1975, Table 2·1), leading to the replacement of the odds of 5 to 1 by odds of approximately 80 to 1. Examination of Fig. 1 suggests that these latter odds are extravagant, illustrating the tendency of estimative methods to read too much into the data.

The proposition of the specimen labelled A in Fig. 1 raises the question of whether it is atypical of the identified offshore standards. To examine this its atypicality index, defined by Aitchison & Dunsmore (1975, p. 226) as the probability that a case has a higher predictive density than the case under scrutiny, may be evaluated from formula (11·20) of Aitchison & Dunsmore (1975). The atypicality index is only 0·62 and so the specimen can hardly be regarded as atypical.

This example can be further used to illustrate the nature of the conditioning property. Suppose that for an offshore specimen we wish to study the variability of the composition of (sand, clay) for a given silt/clay ratio. We can proceed as follows. First find the predictive distribution for a new complete vector based on D_2. This will be two-dimensional logistic-Student with 9 degrees of freedom. We can then easily derive the predictive conditional distribution of (sand, clay) for given value v of log (silt/clay) as logistic-Student with 9 degrees of freedom. In more familiar terms the conditional distribution of log (sand/clay) is

$$\mathrm{St}_1\{9, -3{\cdot}29+1{\cdot}66v, 2{\cdot}10+0{\cdot}422(v-2{\cdot}71)^2\}$$

in the notation of Aitchison & Dunsmore (1975, Table 2·2). Figure 2 shows the considerable differences in this logistic-Student distribution for three silt/clay ratios spanning the range of silt/clay ratios observed in the specimens.

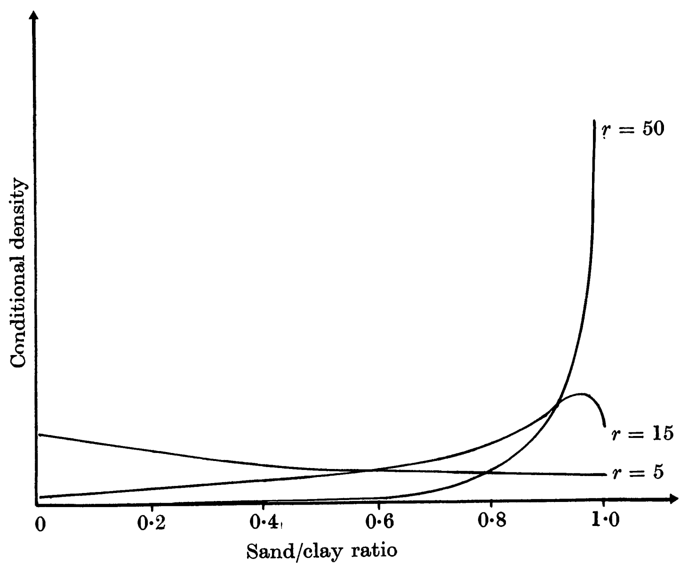

Fig. 2. Conditional density functions of (sand, clay) composition for given value r of silt/clay ratio.

4·3. *Analysis of probabilistic data*

Data in the form of probability vectors arise in a variety of applications such as answers to multiple-choice questions (de Finetti, 1972, p. 30) and in the study of subjective performance in inferential tasks (Taylor, Aitchison & McGirr, 1971). Provided that the probabilities are all positive we then have data in the form of vectors in S^d.

To obtain a simple illustration we presented 24 students with exactly the same diagnostic problem, the differential diagnosis of newmath syndrome (Aitchison, 1974), and asked them to assess subjectively the diagnostic probabilities they attached to each of three possible types. Conditions were identical for all students except that 12, randomly selected, performed the task before, and the remaining 12 after, they had encountered the appropriate statistical tool, Bayes's formula. The diagnostic assessments form two sets of probabilistic data, which can conveniently be presented in triangular coordinates in Fig. 3. A question of interest is then whether there is any significant difference in performance in the after and before groups. Adopting logistic-normal distributions $L_2(\mu_A, \Sigma_A)$ and $L_2(\mu_B, \Sigma_B)$ to describe the variability in the after and before data we can then test differences in terms of standard multinormal tests (Anderson, 1958, Chapter 10) of $\mu_A = \mu_B$ and $\Sigma_A = \Sigma_B$. No significant differences are found.

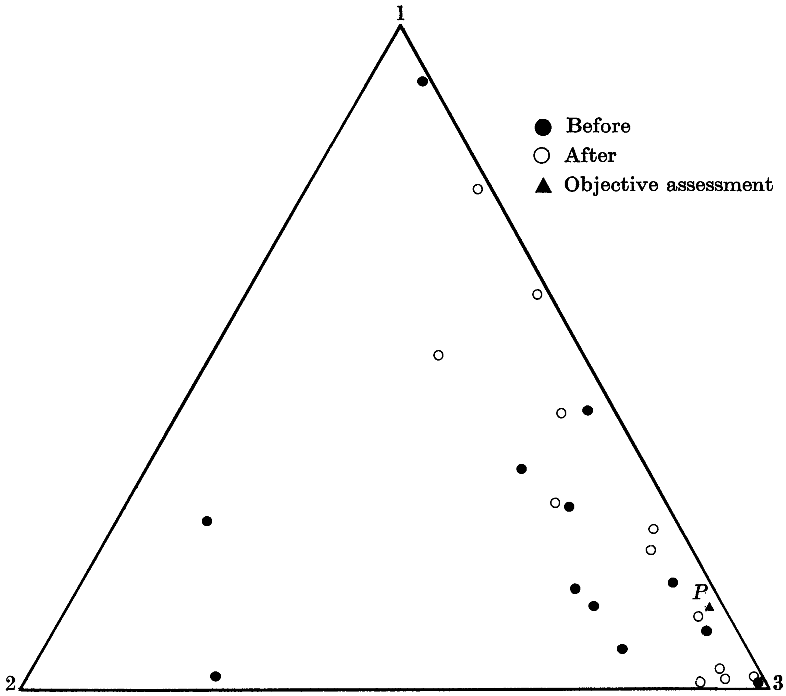

Fig. 3. Diagnostic assessments of 24 subjects.

This particular inference task has an objective answer $P = (0{\cdot}14, 0{\cdot}02, 0{\cdot}84)$ shown on Fig. 3. We can then investigate the extent to which the subjective inferences depart from this by testing the hypotheses $\mu_A = \mu$ and $\mu_B = \mu$, where $\mu = (-1{\cdot}79, -3{\cdot}74)$, the log ratio vector associated with P. Both these tests give significant departure from μ, at the 1% significance level for μ_A and at the 5% significance level for μ_B.

4·4. *An application in logistic discriminant analysis*

For simplicity we confine attention to discrimination between two types. In logistic discriminant analysis the probability that a case with given vector x of diagnostic features is of type t ($t = 1, 2$) is expressed in the parametric form

$$\operatorname{pr}(t = 1 \mid x, \beta) = 1 - \operatorname{pr}(t = 2 \mid x, \beta) = \exp(\beta^{\mathrm{T}} x)/\{1 + \exp(\beta^{\mathrm{T}} x)\},$$

where β denotes the parameter.

Standard practice here is to use the diagnostic training set $D = \{(t_i, x_i) : i = 1, \ldots, n\}$ of n cases of known types t_i and corresponding feature vectors x_i first to estimate β by its maximum likelihood estimate $\hat{\beta}$. To produce diagnostic probabilities for a new case with feature vector x a common procedure is then simply to quote the estimative probabilities $p(t \mid x, \hat{\beta})$ or even to leave these in their transformed or log odds versions, commonly termed the scores $\hat{\beta}^{\mathrm{T}} x$. There is usually little attempt to quantify in a meaningful way the reliability of such a diagnostic assessment other than to make some comment about the possibilities of producing standard errors for the scores. One way of taking account of the unreliability of the estimation process in reaching diagnostic probabilities is through the predictive diagnostic device (Aitchison, Habbema & Kay, 1977) of weighting each possible assessment $p(t \mid x, \beta)$ by a suitable posterior distribution $p(\beta \mid D)$ to obtain diagnostic probabilities $\int p(t \mid x, \beta) p(\beta \mid D)\, d\beta$, where the integral is over the range B which is the set of possible parameters β. Although this device does take account of the unreliability it is sometimes criticized because its presentation of a single set of diagnostic probabilities gives the impression that these are the diagnostic probabilities rather than the result of a weighting process. For more than two types this weighting process is probably the only realistic way of presenting a comprehensible and practically useful overall view. For two types, however, a middle course can be steered which gives an impression of the extent of the unreliability of the diagnostic probability assessments for a new case with feature vector x. The approach follows the Bayesian device of using $p(\beta \mid D)$ as a vehicle for carrying the unreliability of the estimation process. The distribution $p(\beta \mid D)$ in its asymptotic Bayesian form is multivariate normal and so, for a given x, induces a multivariate normal distribution, say $N\{\mu(x), \sigma^2(x)\}$ on the score $\beta^{\mathrm{T}} x$. This in turn induces, through the inverse logistic transformation (1·1), a logistic-normal distribution for the diagnostic probabilities $u_t = p(t \mid x, \beta)$.

For a case with given feature vector x the above argument leads to a predictive diagnostic probability for type I:

$$\int_{-\infty}^{\infty} \frac{e^v}{1+e^v} \phi\{v \mid \mu(x), \sigma^2(x)\}\, dv.$$

For any specified value, say 0·9, of this diagnostic probability there must be a relationship between corresponding $\mu(x)$ and $\sigma^2(x)$ and using Lauder's (1978, formula 3·7b) approximation this can be shown to be

$$\mu(x) = 2{\cdot}20 \sqrt{\{1 + 0{\cdot}346\sigma^2(x)\}}.$$

It is thus possible for cases with the same predictive diagnostic probabilities to have widely different μ and σ^2 values and so different induced logistic-normal distributions for u_1. That such differences can reflect very different reliabilities of the diagnostic probability assertion is easily seen from Fig. 4 which shows the graphs of the logistic-normal distribution function

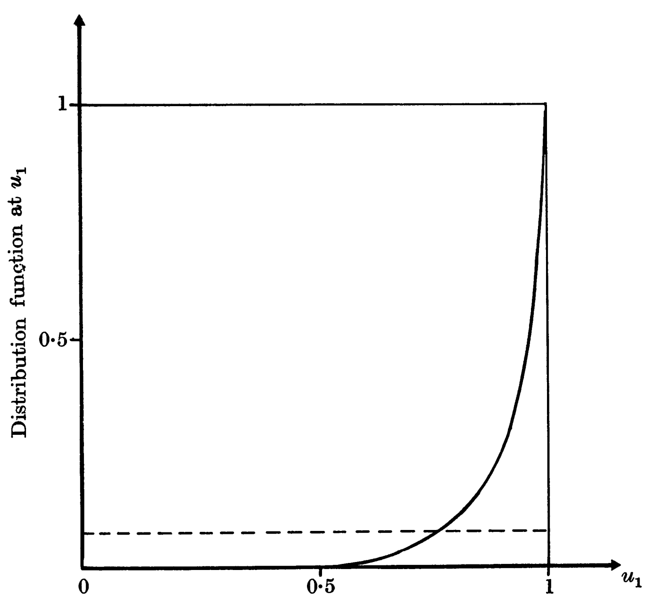

Fig. 4. Logistic-normal distribution functions of the diagnostic probability u_1 for two new cases: case 1, ($\mu = 2{\cdot}9$, $\sigma = 1{\cdot}4$), shown by solid line; case 2, ($\mu = 105$, $\sigma = 81$), broken line.

of u_1 for two actual cases, with $\mu = 2\cdot9$, $\sigma = 1\cdot4$ and $\mu = 105$, $\sigma = 81$, each with predictive diagnostic probabilities of 0·9 for type 1. The first case gives a fairly reliable diagnosis for type I since there is little chance of u_1 being less than 0·5. The second case has obviously a very unreliable diagnosis since the logistic-normal probability distribution is almost entirely concentrated in the neighbourhoods of $u_1 = 0$ and $u_2 = 1$. Thus these induced logistic-normal distributions do give some insight into the diagnostic process.

5. Discussion

Although we have shown that the logistic-normal distributions provide a flexible tool for statistical analysis of a variety of applications a number of problems remain for future consideration.

(i) How can the techniques be adapted to cope with zero components in u vectors?

(ii) Can we develop satisfactory tests of the separate families, Dirichlet and logistic-normal, along the lines of Cox (1962)? In particular, to what extent are current tests of multivariate normality powerful against a Dirichlet alternative?

(iii) To what extent may logistic-normal distributions possess the component-additive property in some approximate form which would allow us to apply logistic-normal analysis to complete vectors of compositional data and to vectors collapsed through addition of components?

(iv) How worthwhile is it to widen the logistic-normal class, from the use of the logarithmic transformation to the complete Box & Cox (1964) class of transformations:

$$v_i = \{(u_i/u_{d+1})^\lambda - 1\}\lambda^{-1} \quad (i = 1, \ldots, d)?$$

We are currently investigating applications in such widely differing areas as petrology, soil compositions, fresh-water ecology and the analysis of subjective performance in inferential tasks. The answers to some of the above questions will clearly be conditioned by the particular needs of such practical problems.

References

Aitchison, J. (1974). Hippocratic inference. *Bull. Inst. Math. Applic.* **10**, 48–53.

Aitchison, J. (1976). Goodness of prediction fit. *Biometrika* **62**, 547–54.

Aitchison, J. & Begg, C. B. (1976). Statistical diagnosis when cases are not classified with certainty. *Biometrika* **63**, 1–12.

Aitchison, J. & Dunsmore, I. R. (1975). *Statistical Prediction Analysis*. Cambridge University Press.

Aitchison, J., Habbema, J. D. F. & Kay, J. W. (1977). A critical comparison of two methods of statistical discrimination. *Appl. Statist.* **26**, 15–25.

Anderson, T. W. (1958). *An Introduction to Multivariate Statistical Analysis*. New York: Wiley.

Bloch, D. A. & Watson, G. S. (1967). A Bayesian study of the multinomial distribution. *Ann. Math. Statist.* **38**, 1423–35.

Box, G. E. P. & Cox, D. R. (1964). The analysis of transformations (with discussion). *J.R. Statist. Soc.* B **26**, 211–52.

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

De Finetti, B. (1972). *Probability, Induction and Statistics*. New York: Wiley.

Johnson, N. L. (1949). Systems of frequency curves generated by methods of translation. *Biometrika* **36**, 149–76.

Kullback, S. & Liebler, R. A. (1951). On information and sufficiency. *Ann. Math. Statist.* **22**, 525–40.

Lauder, I. J. (1978). Computational problems in predictive diagnosis. *Compstat* 1978, 185–92.

Leonard, T. (1973). A Bayesian method for histograms. *Biometrika* **60**, 297–308.

Lindley, D. V. (1964). The Bayesian analysis of contingency tables. *Ann. Math. Statist.* **35**, 1622–43.

Lindley, D. V., Tversky, A. & Brown, R. V. (1979). On the reconciliation of probability assessments (with discussion). *J. R. Statist. Soc.* A **142**, 146–80.

McCammon, R. B. (1975). *Concepts in Geostatistics*. New York: Wiley.

MOSIMANN, J. E. (1975). Statistical problems of size and shape. In *Statistical Distributions in Scientific Work*, Eds. G. P. Patil, S. Kotz and K. Ord, pp. 187–239. Dordrecht: Reidel.
TAYLOR, T. R., AITCHISON, J. & MCGIRR, E. M. (1971). Doctors as decision-makers: a computer-assisted study of diagnosis as a cognitive skill. *Br. Med. J.* **3**, 35–40.
WILKS, S. S. (1962). *Mathematical Statistics*. New York: Wiley.

[*Received March* 1979. *Revised January* 1980]