A Logistic Normal Mixture Model for Compositional Data Allowing Essential Zeros

John Bear
Statistical Consulting Lab,
University of Arizona, U.S.A.

Dean Billheimer
Statistical Consulting Lab,
University of Arizona, U.S.A.

Abstract

The usual candidate distributions for modeling compositions, the Dirichlet and the logistic normal distribution, do not include zero components in their support. Methods have been developed and refined for dealing with zeros that are rounded, or due to a value being below a detection level. Methods have also been developed for zeros in compositions arising from count data. However, essential zeros, cases where a component is truly absent, in continuous compositions are still a problem.

The most promising approach is based on extending the logistic normal distribution to model essential zeros using a mixture of additive logistic normal distributions of different dimension, related by common parameters. We continue this approach, and by imposing an additional constraint, develop a likelihood, and show ways of estimating parameters for location and dispersion. The proposed likelihood, conditional on parameters for the probability of zeros, is a mixture of additive logistic normal distributions of different dimensions whose location and dispersion parameters are projections of a common location or dispersion parameter. For some simple special cases, we contrast the relative efficiency of different location estimators.

Keywords: composition, subcomposition, essential zero, logistic normal, projection.

1. Introduction

An essential zero in compositional data is a zero component which is not caused by rounding or some other difficulty in measurement, but rather, is genuinely believed to be zero. This is fundamentally a different problem than that addressed by recent work on rounded zeros, or below-detection level zeros, such as in Palarea-Albaladejo and Martín-Fernández (2015) and references therein. Although there are recent workable Bayesian approaches to zeros in compositions from count data, Martín-Fernández, Hron, Templ, Filzmoser, and Palarea-Albaladejo (2014) and references therein, essential zeros in continuous compositions still present a problem.

We develop an approach proposed by Aitchison and Kay (2003) to extend the logistic normal distribution to accommodate essential zeros. Aitchison (1986) and Aitchison and Kay (2003) note that a key feature compositional data is that ratios of the components contain all pertinent information about the composition. Essential zeros complicate this feature in that

they contain no information about the other components of the composition. In addition, an observation containing an essential zero is at the boundary of the simplex and is a composition of smaller dimension.

2. Previous work

In addition to the work mentioned above, there have been other approaches to zeros in compositions. Work by Butler and Glasbey (2008) mapped a latent Gaussian variable to a composition, but seems only to work for two and three-part compositions. An additional concern is that it does not preserve ratios of parts in subcompositions. In contrast, Leininger, Gelfand, Allen, and Silander Jr (2013) have developed a more practical treatment of compositions as coming from a latent Gaussian random variable where the compositional component is zero when the latent Gaussian component is less than or equal to zero. They develop a hierarchical power model with the transformation $X_k = \frac{(\max(0, Z_k))^\gamma}{1 + \sum_{k'=1}^{d}(\max(0, z_{k'}))^\gamma}$ where Z_k is the kth normal component and X_k is the corresponding compositional component. D is the number of parts in the composition, $d = D - 1$, and $X_D = (1 + \sum_{k'=1}^{d}(\max(0, Z'_k))^\gamma)^{-1}$. The corresponding inverse transformation is $Z_k = (X_k/X_D)^{1/\gamma}$ if $X_k > 0$, and $Z_k \leq 0$ (latent) if $X_k = 0$, for $k = 1, 2, \ldots, d$. To estimate parameters they use MCMC. One limitation of their approach is also a limitation of ours: we require one component of the composition to be strictly positive.

Work by Stewart and Field (2011) uses a multiplicative logistic normal mixture model that allows them to consider the univariate log odds for the ith component to be normally distributed where the ith component is not zero. It works well for their applications, in particular regression, but does not capture covariance easily.

Scealy and Welsh (2011) transform compositions into directional data on the hypersphere, and develop a regression model using the Kent distribution, Kent (1982), which tolerates zeros, though they write, "When any of the components of **u** are distributed too close to 0, boundary issues arise and in this case we need to pursue alternative approaches since the fitted Kent model (and the von Mises-Fisher model) may have significant support outside the positive orthant." A further issue with their approach is that their square root transformation does not preserve ratios of parts in subcompositions.

Our goal here is to extend the additive logistic normal distribution to handle essential zeros for continuous data.

3. Motivating example

Suppose we have compositional data on how much Bill spends on rice, lentils, and spices when he buys food. Suppose he buys in bulk, and occasionally the store is out of either the spices or lentils, but they always have plenty of rice. Table 1 shows a set of such compositions where some of the entries, for spices or lentils, are zero. Our goal is to develop a model for data like these by extending the additive logistic normal distribution.

4. Definitions (from Aitchison, 1986)

Definition: The *d-dimensional simplex embedded in D-dimensional real space* is the set of compositions, **x**, defined by

$$\mathcal{S}^d = \{\mathbf{x} = (x_1, \ldots, x_d, x_D) : x_1 > 0, \ldots, x_D > 0; \sum_{i=1}^{D} x_i = 1\},$$

where $d = D - 1$. If $\mathbf{x} = (x_1, x_2, \ldots, x_d, x_D)^T$, then $\mathbf{x}_{-D} = (x_1, x_2, \ldots, x_d)^T$. The *additive logratio transformation,* alr, is defined as follows:

$$\text{alr} : \mathcal{S}^d \to \mathbb{R}^d$$

Table 1: Composition of food expense

	spices	lentils	rice
1	0.16	0.00	0.84
2	0.17	0.00	0.83
3	0.16	0.00	0.84
4	0.00	0.37	0.63
5	0.00	0.37	0.63
6	0.00	0.37	0.63
7	0.12	0.33	0.55
8	0.11	0.34	0.56
9	0.12	0.32	0.56
10	0.10	0.34	0.56
11	0.10	0.33	0.57
12	0.11	0.33	0.55

$$\mathbf{x} \mapsto \mathbf{y} = (\log(x_1/x_D), \log(x_2/x_D), \ldots, \log(x_d/x_D))^T. \tag{1}$$

We define the shorthand $\log(\mathbf{x}_{-D}/x_D) = (\log(x_1/x_D), \log(x_2/x_D), \ldots, \log(x_d/x_D))^T$. Since alr is one-to-one, its inverse exists. It is called the *logistic transformation*, alr^{-1}, defined as

$$\mathrm{alr}^{-1} : \mathbb{R}^d \to \mathcal{S}^d \tag{2}$$
$$\mathbf{y} \mapsto \mathbf{x} = (x_1, x_2, \ldots, x_d, x_D)^T, \text{ where for } (i = 1, \ldots, d),$$
$$x_i = \exp(y_i)/\{\exp(y_1) + \cdots + \exp(y_d) + 1\}$$
$$x_D = 1/\{\exp(y_1) + \cdots + \exp(y_d) + 1\}.$$

5. Simplifying assumption

In this section we outline our method for building a mixture distribution for dealing with compositions containing essential zeros, but leave most of the details about the weights for later. A key simplifying assumption we make throughout is that one of the parts of the composition, the Dth component, is never zero. We allow zeros anywhere else but not in the last component.

In a set of logistic normal data without zeros, the likelihood has been shown to be permutation invariant (Aitchison 1986). In our extension which allows zeros, if some parts are never zero, the likelihood is invariant to the choice of which one of those nonzero parts is chosen as the reference provided the same reference part is used throughout the data set.

Let $\mathbf{x} = (x_1, x_2, \ldots, x_d, x_D)^T$ be a composition with $x_i < 1$ for all $i \in \{1, 2, \ldots, d, D\}$ and $x_D > 0$. For $i \in \{1, 2, \ldots, d\}$, consider two possibilities. Either $x_i = 0$ or $x_i > 0$. Let $W = \{i : i \in \{1, 2, \ldots, d\}, x_i > 0\}$. That is, W is the set of indices for the parts of $\mathbf{x}$ (other than x_D) which are nonzero (positive). For any given composition $\mathbf{x}$, W is the set of all the indices of the nonzero components of $\mathbf{x}$. There are $2^d - 1$ possible sets W. There are $2^d - 1$ and not 2^d because W cannot be empty. If W were empty that would require that $x_D = 1$ in order for $\mathbf{x}$ to be a composition, but we have already said we require all $x_i < 1$ including x_D. Each pattern of zeros corresponds to a different set W. We index them as W_ℓ with $\ell \in \{1, 2, \ldots, 2^d - 1\}$. They are elements of the power set, $W_\ell \in \mathcal{P}(\{1, 2, \ldots, d\})$. Sometimes we refer to these sets with incidence vectors where the ith component $V_{W_\ell i} = 1 \iff x_i > 0$ and $V_{W_\ell i} = 0 \iff x_i = 0$.

Each W_ℓ has some probability of occurrence, $P(W_\ell)$. Although some pattern can be not

present $P(W_\ell) = 0$, the probabilities must sum to one,

$$\sum_{\ell=1}^{2^d-1} P(W_\ell) = 1.$$

We use the probabilities $P(W_\ell)$ as the weights in a mixture distribution. For the other distributions making up our mixture, we use logistic normal distributions $\mathcal{L}(\mathbf{x}; \boldsymbol{\mu}_{W_\ell}, \boldsymbol{\Omega}_{W_\ell})$ derived from a single parent logistic normal distribution $\mathcal{L}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Omega})$. They are in fact projections from the parent. We will call the distributions derived from the parent distribution *subdistributions* once we define them. So the mixture distribution will be denoted as follows, once we define a few more terms,

$$g(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Omega}) = \sum_{\ell=1}^{2^d-1} P(W_\ell)\mathcal{L}(\mathbf{x}; \boldsymbol{\mu}_{W_\ell}, \boldsymbol{\Omega}_{W_\ell}).$$

In the parent distribution, $\mathcal{L}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Omega})$, $\boldsymbol{\mu}$ is a d-part location parameter vector, $\boldsymbol{\mu} \in \mathbb{R}^d$, and $\boldsymbol{\Omega}$ is a $d \times d$ positive definite dispersion matrix. To ease the discussion we will refer to $\boldsymbol{\mu}$ and $\boldsymbol{\Omega}$ as *mean vector* and *variance-covariance matrix* respectively, although they are not moments of the distribution. For the distributions derived from the logistic normal parent distribution, the parameters $\boldsymbol{\mu}_{W_\ell}$ and $\boldsymbol{\Omega}_{W_\ell}$ are defined in terms of the parameters $\boldsymbol{\mu}$, and $\boldsymbol{\Omega}$, and the set of indices of nonzero components of $\mathbf{x}$, W_ℓ, and a selection matrix $\mathbf{B}_{W_\ell}$.

Let $W_\ell \subset \{1, 2, 3, \ldots, d\}$ be a nonempty set of indices (of the nonzero components of $\mathbf{x}$); without loss of generality we can order the indices from least to greatest

$$W_\ell = \{j_1, j_2, \ldots, j_J\} \text{ where } 0 < j_1 < j_2 < \ldots < j_J \leq d.$$

Now we define our $J \times d$ selection matrix, $\mathbf{B}_{W_\ell} = [B_{i,m}]$. For $i \in \{1, 2, \ldots, J\}$, and $m \in \{1, 2, \ldots, d\}$, with $W_\ell = \{j_1, j_2, \ldots, j_J\}$, we define the elements of $[B_{i,m}]$ to be $B_{i,j_i} = 1$ and $B_{i,m \neq j_i} = 0$. For example, let $\mathbf{x} = (.2, 0, .3, 0, .25, .25)$, a 6-part composition, with $x_6 > 0$. The set of nonzero indices is $W_\ell = \{1, 3, 5\}$, and the selection matrix is

$$\mathbf{B}_{W_\ell} = \mathbf{B}_{\{1,3,5\}} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Now it is easy to define $\boldsymbol{\mu}_{W_\ell}$ and $\boldsymbol{\Omega}_{W_\ell}$. We define:

$$\boldsymbol{\mu}_{W_\ell} = (\mathbf{B}_{W_\ell})(\boldsymbol{\mu}).$$

$$\boldsymbol{\Omega}_{W_\ell} = (\mathbf{B}_{W_\ell})(\boldsymbol{\Omega})(\mathbf{B}_{W_\ell}^T).$$

With this structure, the mixture distribution can be more fully specified.

$$g(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Omega}) = \sum_{\ell=1}^{2^d-1} P(W_\ell)\mathcal{L}^{||W_\ell||}(\mathbf{x}; \boldsymbol{\mu}_{W_\ell}, \boldsymbol{\Omega}_{W_\ell}) \text{ where}$$

- $||W_\ell||$ refers to the cardinality of the set W_ℓ.

- $\sum P(W_\ell) = 1$.

- $\boldsymbol{\mu}$ is a d-part vector in $\mathbb{R}^d$.

- $\boldsymbol{\mu}_{W_\ell}$ is a subvector of $\boldsymbol{\mu}$ corresponding to the W_ℓ pattern of zeros.

- $\boldsymbol{\Omega}_{W_\ell}$ is a submatrix of a $d \times d$ positive definite covariance matrix corresponding to the W_ℓ pattern of zeros.

5.1. Multivariate normal foundation

Now we extend the notation for the inverse of the additive logratio transformation, alr^{-1}, from Aitchison (1986). We use the new symbols, $\tilde{\mathrm{alr}}$ and $\tilde{\mathrm{alr}}^{-1}$. We define them in terms of W and D, the maximum index. Let $W \subset \{1, 2, \ldots, d\}$ be a pattern of zeros, i.e., a set of indices of nonzero components of $\mathbf{x}$ and denote them: $W = \{i_1, i_2, \ldots, i_r\}$, and let $j \in \{1, 2, \ldots, d, D\}$.

In our approach there is a tight correspondence between the y_i variables of a multivariate normal vector and the x_i parts of a composition, possibly one containing essential zeros.

$$y_i = \log(x_i/x_D) \text{ for } i = 1, 2, 3, \ldots, d.$$

$$\text{Composition: } \mathbf{x} = (x_1, x_2, x_3, \ldots, x_d, x_D)^T$$
$$\mid \quad \mid \quad \mid \qquad \mid$$
$$\text{alr transformed vector: } \mathbf{y} = \log(\mathbf{x}_{-D}/x_D) = (y_1, y_2, y_3, \ldots, y_d, \odot)^T.$$

When there is an essential zero in the composition in one of the x_i parts, e.g., in x_2, we use $\odot$ as a placeholder so things line up, for example:

$$\text{Composition: } \mathbf{x} = (x_1, 0, x_3, \ldots, x_d, x_D)^T$$
$$\mid \qquad \mid \qquad \mid \qquad \qquad (3)$$
$$\tilde{\mathrm{alr}} \text{ transformed subvector: } \mathbf{y} = \log(\mathbf{x}_{-\{2,D\}}/x_D) = (y_1, \odot, y_3, \ldots, y_d, \odot)^T.$$

When we have an essential zero in the ith part of the composition, we have a selected subvector of the y's which does not contain an element corresponding to y_i. The requirement that $x_D > 0$ is what allows us to maintain this strict correspondence between x_i and y_i.

Now we define $\tilde{\mathrm{alr}}^{-1}$

$$\tilde{\mathrm{alr}}^{-1}(\mathbf{y}, W, D) = (x_1, \ldots, x_d, x_D)^T \text{ where,} \qquad (4)$$

$$x_j = \begin{cases} \exp(y_j)/\{\exp(y_{i_1}) + \exp(y_{i_2}) + \ldots + \exp(y_{i_r}) + 1\} & \text{if } j \in W \\ 0 & \text{if } j \notin W \ \& \ j \in \{1, \ldots, d\} \\ 1/\{\exp(y_{i_1}) + \exp(y_{i_2}) + \ldots + \exp(y_{i_r}) + 1\} & \text{if } j = D. \end{cases}$$

Next we turn to defining an extension to the logistic normal distribution. Let $\mathbf{x} = (x_1, x_2, x_3, \ldots, x_d, x_D)^T$ be a composition, and let $\mathbf{y} = \log(\mathbf{x}_{-D}/x_D) = (y_1, y_2, y_3, \ldots, y_d)^T$. Then $\mathbf{x}$ is defined to have a logistic normal distribution with location $\boldsymbol{\mu}$ and dispersion $\boldsymbol{\Omega}$, written $\mathbf{x} \sim \mathcal{L}(\boldsymbol{\mu}, \boldsymbol{\Omega})$, if $\mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Omega})$. Note also that if $\mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Omega})$, and $\mathbf{B}$ is a selection matrix as mentioned in Section 5, then $\mathbf{By}$ is also multivariate normal with distribution:

$$\mathbf{By} \sim \mathcal{N}(\mathbf{B}\boldsymbol{\mu}, \mathbf{B}\boldsymbol{\Omega}\mathbf{B}^T).$$

And finally, recall that also in Section 5 we used W_ℓ to represent the set of indices of the nonzero components of $\mathbf{x}$. $\mathbf{B}_{W_\ell}$ is the selection matrix that selects those components when we perform these matrix multiplications:

$$\boldsymbol{\mu}_{W_\ell} = (\mathbf{B}_{W_\ell})(\boldsymbol{\mu}),$$

$$\boldsymbol{\Omega}_{W_\ell} = (\mathbf{B}_{W_\ell})(\boldsymbol{\Omega})(\mathbf{B}_{W_\ell}^T).$$

With all these definitions in place we are now in a position to define a logistic normal distribution with essential zeros.
Definition:

Let $\mathbf{x} = (x_1, x_2, x_3, \ldots, x_d, x_D)^T$ be a composition with $\mathbf{x}_D > 0$.
Let $W_\ell = \{i_1, i_2, \ldots, i_r\} \subset \{1, 2, \ldots, d\}$ be a nonempty set of indices of nonzero components of $\mathbf{x}$.

Let $\mathbf{B}_{W_\ell}$ be the corresponding selection matrix.

Let $\mathbf{y} = \log(\mathbf{B}_{W_\ell}\mathbf{x}_{-D}/x_D) = (y_{i_1}, y_{i_2}, \ldots, y_{i_r})^T = \tilde{\mathrm{alr}}(\mathbf{x}, W_\ell, D)$ be the logratios of the nonzero components of $\mathbf{x}$.

If for every set W_ℓ of indices of nonzero components of $\mathbf{x}$, we have $\mathbf{y}\sim\mathcal{N}(\boldsymbol{\mu}_{W_\ell}, \boldsymbol{\Omega}_{W_\ell})$, then $\mathbf{x}$ has a logistic normal distribution with essential zeros, written $\mathbf{x}\sim\mathcal{L}(\boldsymbol{\mu}, \boldsymbol{\Omega})$, with probability density function

$$g(\mathbf{x}; \boldsymbol{\mu}, \Omega) = \sum_{\ell=1}^{2^d-1} P(W_\ell)\mathcal{L}^{||W_\ell||}(\mathbf{x}; \boldsymbol{\mu}_{W_\ell}, \boldsymbol{\Omega}_{W_\ell}),$$

where

$$\sum P(W_\ell) = 1.$$

$\boldsymbol{\mu}$ is a d-part vector in R^d.

$\boldsymbol{\mu}_{W_\ell}$ is a subvector of $\boldsymbol{\mu}$ corresponding to the W_ℓ pattern of zeros.

$\boldsymbol{\Omega}$ is a $d \times d$ positive definite covariance matrix.

$\boldsymbol{\Omega}_{W_\ell}$ is a square submatrix of $\boldsymbol{\Omega}$, corresponding to the W_ℓ pattern of zeros.

For the case where $W = \{1, 2, \ldots, d\}$ the composition $\mathbf{x} = (x_1, x_2, \ldots, x_D)^T$ has the additive logistic normal distribution, $\mathcal{L}^d(\boldsymbol{\mu}, \boldsymbol{\Omega})$.

5.2. Common expectations and variances

The definition of $\tilde{\mathrm{alr}}^{-1}$ enables compositions from different subdistributions to be used to estimate parameters of their shared parent distribution. Let $\mathbf{x}_1 = (x_{11}, x_{21}, \ldots, x_{D1})^T$, and let $\mathbf{x}_2 = (x_{12}, x_{22}, \ldots, x_{D2})^T$ with

$$\mathbf{x}_1\sim\mathcal{L}^{||W_1||}(\boldsymbol{\mu}_{W_1}, \boldsymbol{\Omega}_{W_1}), \text{ and} \tag{5}$$

$$\mathbf{x}_2\sim\mathcal{L}^{||W_2||}(\boldsymbol{\mu}_{W_2}, \boldsymbol{\Omega}_{W_2}). \tag{6}$$

The two sets of nonzero indices, W_1, W_2 need not have any elements in common, nor do they need to have the same number of elements, though $\mathbf{x}_1$ and $\mathbf{x}_2$ both have D elements. Suppose they have an index, m, in common: $m \in W_1 \cap W_2$. By properties of the logistic normal distribution (Aitchison 1986, p. 116), and the definition of $\tilde{\mathrm{alr}}^{-1}$ in Equation 4 we have:

$$E \log(x_{m1}/x_{D1}) = Ey_m = \mu_m = Ey_m = E \log(x_{m2}/x_{D2}). \tag{7}$$

And similarly,

$$\mathrm{Var}[\log(x_{m1}/x_{D1})] = \mathrm{Var}[y_m] = \sigma_m^2 = \mathrm{Var}[y_m] = \mathrm{Var}[\log(x_{m2}/x_{D2})]. \tag{8}$$

Thus, compositions from different subdistributions of the same logistic normal distribution can be used to estimate the parameters of their shared parent distribution.

6. Data blocks

Now that we have a correspondence between multivariate normal variables and compositions with zeros, we could derive a density function using the standard formula for transformed variables, analogous to Aitchison (1986, chapter 6). However, for estimating parameters it is more convenient to work in the space of the transformed variables (multivariate normal projections).

Here we apply the techniques and notation of block matrices and matrix calculus to do some preparation in order to build a likelihood and attack the problem of finding estimators for the

parameters. We discuss two sets of estimators, a general maximum likelihood estimator, and a simpler pair of estimators reminiscent of method of moments estimators.

6.1. Block matrices of compositions

We write a collection of compositional data with zeros, $\mathbf{X}$, as a column of blocks of compositions where each block, $\mathbf{X}_\ell$, has a particular pattern of zeros throughout. That is, for a particular block, $\mathbf{X}_\ell$, and $i \in \{1, 2, \ldots, d\}$, the ith column of $\mathbf{X}_\ell$ is either all positive, or all zero. Let

$$\mathbf{X}_{n \times D} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_b \end{bmatrix}. \tag{9}$$

The dimensions of the blocks are: $\underset{r_1 \times D}{\mathbf{X}_1}$, $\underset{r_2 \times D}{\mathbf{X}_2}$, ..., $\underset{r_b \times D}{\mathbf{X}_b}$ and the sum of their vertical dimensions is $r_1 + r_2 + \ldots + r_b = n$, where n is the number of data points.

We use ℓ to indicate a block, and t to indicate a composition (row) in that block. Next we define the patterns of zeros in each block. Here $i \in \{1, 2, \ldots, D\}$. For $\ell \in \{1, 2, \ldots, b\}$, let $W_\ell \subset \{1, 2, \ldots, d\}$ be the set of indices of strictly positive components of $\mathbf{X}_\ell$. For $\ell \in \{1, 2, \ldots, b\}$,

$$\mathbf{X}_\ell = [x_{ti}], \text{ where } \begin{cases} x_{ti} > 0 \text{ if } i = D, \\ x_{ti} > 0 \text{ if } i \in W_\ell, \\ x_{ti} = 0 \text{ if } i \notin W_\ell \text{ and } i \neq D. \end{cases} \tag{10}$$

6.2. Transformations - ratios and logratios

We have already defined the alr transformation for the case where there are no zeros in (1). Next we extend alr to $\tilde{\text{alr}}$ for a block matrix of compositions, $\underset{r_\ell \times D}{\mathbf{X}}$ which may contain zeros. We do this by defining a selection matrix $\mathbf{B}_{W_\ell}$ corresponding to set W_ℓ. We still have $W_\ell \subset \{1, 2, 3, \ldots, d\}$ being a nonempty set of indices of the nonzero components of $\mathbf{x}$, and without loss of generality we can order the indices from least to greatest:

$$W_\ell = \{j_1, j_2, \ldots, j_J\} \text{ where } 0 < j_1 < j_2 < \ldots < j_J < D. \tag{11}$$

Now we define our $(J + 1) \times D$ selection matrix, $\mathbf{B}_{W_\ell} = [B_{p,m}]$. We use $J + 1$ here because we construct the selection matrix so that the final, D^{th}, component of the data is always selected. This is slightly different than before. Previously we constructed B to conform to the parameters $\boldsymbol{\mu}(d \times 1)$ and $\boldsymbol{\Omega}(d \times d)$.

$$\text{For } p \in \{1, 2, \ldots, J + 1\}, \text{ and } m \in \{1, 2, \ldots, D\}, \text{ with } W_\ell = \{j_1, j_2, \ldots, j_J\}, \tag{12}$$

$$\text{we define the elements of } [B_{p,m}] \text{ to be } B_{p,j_p} = 1 \text{ and } B_{p,m \neq j_p} = 0. \tag{13}$$

$\mathbf{X}_\ell \mathbf{B}_{W_\ell}^T$ is a matrix where each row vector is a composition without zeros.

$$\mathbf{X}_\ell \mathbf{B}_{W_\ell}^T = \begin{bmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \vdots \\ \mathbf{x}_{r_\ell}^T \end{bmatrix} \mathbf{B}_{W_\ell}^T = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1(J+1)} \\ x_{21} & x_{22} & \cdots & x_{2(J+1)} \\ \vdots & \vdots & & \vdots \\ x_{r_\ell 1} & x_{r_\ell 2} & \cdots & x_{r_\ell(J+1)} \end{bmatrix}. \tag{14}$$

$$\text{We define } \tilde{\text{alr}}(\mathbf{X}_\ell, W_\ell, D) = \text{alr}(\mathbf{X}_\ell \mathbf{B}_{W_\ell}^T) = \underbrace{\begin{bmatrix} \text{alr}(\mathbf{x}_1^T) \\ \text{alr}(\mathbf{x}_2^T) \\ \vdots \\ \text{alr}(\mathbf{x}_{r_\ell}^T) \end{bmatrix}}_{r_\ell \times (J+1)} = \underbrace{\begin{bmatrix} \mathbf{y}_1^T \\ \mathbf{y}_2^T \\ \vdots \\ \mathbf{y}_{r_\ell}^T \end{bmatrix}}_{r_\ell \times J}. \tag{15}$$

$$\text{Let } \underset{r_\ell \times J}{\mathbf{Y}_\ell} = \tilde{\text{alr}}(\mathbf{X}_\ell, W_\ell, D) \tag{16}$$

Each row vector in $\mathbf{Y}_\ell$ is a vector of reals, all potentially from the multivariate normal distribution corresponding to the ℓth pattern of zeros. Note that we cannot form a single block matrix, $\mathbf{Y}$, from the collection of $\mathbf{Y}_\ell$ because they can have different numbers of columns.

6.3. Illustration - spices, lentils, and rice

In our example about compositions of spending on spices, lentils, and rice (Table 1), there are three patterns of zeros. Tables 2-4 show the result of applying the $\tilde{\text{alr}}$ transformation.

$\mathbf{X}_1$ corresponds to rows 1-3 and its set of indices is $W_1 = \{1\}$.
$\mathbf{X}_2$ corresponds to rows 4-6 and its set of indices is $W_2 = \{2\}$.
$\mathbf{X}_3$ corresponds to rows 7-12 and its set of indices is $W_3 = \{1, 2\}$.

Table 2: $\mathbf{Y}_1 = \tilde{\text{alr}}(\mathbf{X}_1, \{1\}, 3)$

	log(spices/rice)
1	-1.66
2	-1.59
3	-1.68

Table 3: $\mathbf{Y}_2 = \tilde{\text{alr}}(\mathbf{X}_2, \{2\}, 3)$

	log(lentils/rice)
4	-0.52
5	-0.52
6	-0.51

Table 4: $\mathbf{Y}_3 = \tilde{\text{alr}}(\mathbf{X}_3, \{1, 2\}, 3)$

	log(spices/rice)	log(lentils/rice)
7	-1.56	-0.52
8	-1.64	-0.51
9	-1.52	-0.57
10	-1.72	-0.51
11	-1.77	-0.53
12	-1.58	-0.51

6.4. Means

The matrix $\mathbf{Y}_\ell$ contains rows of compositions with the same pattern of zeros. We refer to the tth row vector of $\mathbf{Y}_\ell$ as $\mathbf{y}_{\ell t}^T$. We refer to the mean as the vector $\bar{\mathbf{y}}_\ell$, and define it as:

$$\underset{J \times 1}{\bar{\mathbf{y}}_\ell} = \frac{1}{r_\ell}(\mathbf{1}_{r_\ell}^T \mathbf{Y}_\ell)^T. \tag{17}$$

Here we are using $\mathbf{1}_{r_\ell}$ to represent an $r_\ell \times 1$ column vector of ones. $\bar{\mathbf{y}}_\ell$ is a column vector. We define it this way because of how we intend to use it in quadratic forms from the multivariate normal density.

7. Simple estimators

7.1. Mean

It is also possible to construct simpler estimators relying on properties of the normal distribution. For the location, if $\underset{n \times D}{\mathbf{X}} = [x_{ti}]$ is a collection of n compositional data points with zeros, and the Dth component always strictly positive, we can define a simple estimator of the mean, $\overset{*}{\boldsymbol{\mu}} = (\overset{*}{\mu}_1, \overset{*}{\mu}_2, \ldots, \overset{*}{\mu}_d)^T$. Let n_i be the number of elements of the ith column of $\mathbf{X}$ that are nonzero. For $i \in \{1, 2, \ldots, d\}$, and $t \in \{1, 2, \ldots, n\}$, define

$$\overset{*}{\mu}_i = \frac{1}{n_i} \sum_{\{t:x_{ti} \neq 0\}} \log(x_{ti}/x_{tD}). \tag{18}$$

By the assumption of normality of the logratios, the estimator $\overset{*}{\boldsymbol{\mu}}$ is unbiased.

In the spices-lentils-rice example, $\overset{*}{\mu} = $ (log(spices/rice): -1.635 , log(lentils/rice): -0.523). For ease of interpretation, we convert the estimator back to a composition with the alr^{-1} transformation giving: (spices: 0.109 , lentils: 0.332, rice: 0.559). That is, our estimate of Bill's mean expenditure is 10.9% on spices , 33.2% on lentils, and 55.9% on rice.

7.2. Variance

Here we show how to find estimators for variances and covariances using maximum likelihood estimators for normal random variables. For a single random composition, $\mathbf{x}$, with components $x_1, x_2, \ldots, x_D$, we substitute $\log(x_i/x_D)$ into the MLE for variances of normal random variables. We use $\overset{*}{\sigma}_{ii}^2$ for the estimator of the variances of the logratios $\log(x_i/x_D)$, for $i \in \{1, 2, \ldots, d\}$, and $t \in \{1, 2, \ldots, n_i\}$.

$$\overset{*}{\sigma}_{ii}^2 = \frac{1}{n_i} \sum_{\{t:x_{ti} \neq 0\}} \left(\log(x_{ti}/x_{tD}) - \overset{*}{\mu}_i \right)^2. \tag{19}$$

If we want an unbiased estimator, we can divide by $(n_i - 1)$ instead of n_i. As with means, the different $\overset{*}{\sigma}_{ii}$ are based on different numbers of observations, n_i.

7.3. Covariance

It only makes sense to talk about estimating the covariance of the variables $\log(x_i/x_D)$ and $\log(x_j/x_D)$ when both x_{ti} and x_{tj} are not 0 so we define $n_{ij} = ||\{t : x_{ti} \neq 0 \ \& \ x_{tj} \neq 0\}||$. That is, n_{ij} is the number of data points where both x_{ti} and x_{tj} are not 0. As we did with variance, we can start with the canonical maximum likelihood formula for estimating covariance among normally distributed variables, and substitute in appropriate logratios.

$$\overset{*}{\sigma}_{ij} = \frac{1}{n_{ij}} \sum_{\{t:x_{ti} \neq 0 \ \& \ x_{tj} \neq 0\}} (\log(x_{ti}/x_{tD}) - \overset{*}{\mu}_i)(\log(x_{tj}/x_{tD}) - \overset{*}{\mu}_j) \tag{20}$$

Note that $\overset{*}{\sigma}_{ij}$ is based on n_{ij} observations, while $\overset{*}{\mu}_i$ and $\overset{*}{\mu}_j$ are based on n_i and n_j observations, respectively. The formula in Equation 20 is based on the maximum likelihood estimator for covariance of normal variables. For unbiased estimators we would divide by $(n_{ij} - 1)$ instead of n_{ij}.

Our estimator for the $d \times d$ variance-covariance matrix is $\overset{*}{\mathbf{\Omega}} = [\overset{*}{\sigma}_{ij}]$. There are two potential problems with this approach. There could be $i, j, i \neq j$, such that whenever $x_i > 0$, $x_j = 0$. In that case we cannot estimate the covariance. Also, irrespective of that, the estimate of the variance-covariance matrix, $\overset{*}{\mathbf{\Omega}}$, might not be positive definite.

In the spices-lentils-rice example,

$$\overset{*}{\mathbf{\Omega}} = \begin{pmatrix} 0.00648 & -0.00096 \\ -0.00096 & 0.00035 \end{pmatrix}$$

which is positive definite.

8. Maximum likelihood estimators

For the case where there are no zeros, the location estimator described earlier is a maximum likelihood estimator (MLE), but in general the estimator we found earlier is not an MLE. From now on we will call that estimator the simple estimator, to contrast it with the MLE, which we derive next.

We start by finding the location MLE given $\mathbf{\Omega}$ for 3-part compositions, show it is unbiased, and then show the relative efficiency of the simple estimator with to the MLE. Assume we have a set of logistic normal compositional data with b different patterns of zeros as in (9).

$$\mathbf{x}_{11}, \ldots, \mathbf{x}_{1r_1} \overset{i.i.d.}{\sim} \mathcal{L}^{||W_1||}(\mathbf{B}_{W_1}\boldsymbol{\mu}, \mathbf{B}_{W_1}\mathbf{\Omega}\mathbf{B}_{W_1}^T) \qquad (\text{rows of } \mathbf{X}_1)$$

$$\mathbf{x}_{21}, \ldots, \mathbf{x}_{2r_2} \overset{i.i.d.}{\sim} \mathcal{L}^{||W_2||}(\mathbf{B}_{W_2}\boldsymbol{\mu}, \mathbf{B}_{W_2}\mathbf{\Omega}\mathbf{B}_{W_2}^T) \qquad (\text{rows of } \mathbf{X}_2)$$

$$\vdots$$

$$\mathbf{x}_{b1}, \ldots, \mathbf{x}_{br_b} \overset{i.i.d.}{\sim} \mathcal{L}^{||W_b||}(\mathbf{B}_{W_b}\boldsymbol{\mu}, \mathbf{B}_{W_b}\mathbf{\Omega}\mathbf{B}_{W_b}^T). \qquad (\text{rows of } \mathbf{X}_b) \qquad (21)$$

In a block of data, as in (21), we use $\mathbf{x}_{\ell t}$ to refer to the tth compositional observation with W_ℓ pattern of zeros. We define $\mathbf{y}_{\ell t} = \tilde{\text{alr}}(\mathbf{x}_{\ell t}, W_\ell, D)$, and to ease notation, we write in terms of $\mathbf{y}_{\ell t}$.

8.1. Likelihood

First we write the full likelihood and log likelihood for D-part compositions, and then restrict ourselves to 3-part compositions. The full likelihood is:

$$L(\boldsymbol{\mu}, \mathbf{\Omega}|r_1, \ldots, r_b, \mathbf{y}_{11}, \ldots, \mathbf{y}_{br_b}) = \tag{22}$$

$$\prod_{\ell=1}^{b} \prod_{t=1}^{r_\ell} \frac{P(W_\ell)}{(2\pi)^{||W_\ell||/2}|\mathbf{B}_{W_\ell}\mathbf{\Omega}\mathbf{B}_{W_\ell}^T|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{y}_{\ell t} - \mathbf{B}_{W_\ell}\boldsymbol{\mu})^T(\mathbf{B}_{W_\ell}\mathbf{\Omega}\mathbf{B}_{W_\ell}^T)^{-1}(\mathbf{y}_{\ell t} - \mathbf{B}_{W_\ell}\boldsymbol{\mu})\right].$$

The constant

$$\prod_{\ell=1}^{b} \prod_{t=1}^{r_\ell} \frac{P(W_\ell)}{(2\pi)^{||W_\ell||/2}|\mathbf{B}_{W_\ell}\mathbf{\Omega}\mathbf{B}_{W_\ell}^T|^{1/2}} \tag{23}$$

is independent of $\boldsymbol{\mu}$, so for purposes of maximizing the likelihood with respect to $\boldsymbol{\mu}$, we can treat it as a single constant, C.

$$L(\boldsymbol{\mu}, \mathbf{\Omega}|r_1, \ldots, r_b, \mathbf{y}_{11}, \ldots, \mathbf{y}_{br_b})$$

$$= C \prod_{\ell=1}^{b} \prod_{t=1}^{r_\ell} \exp\left[-\frac{1}{2}(\mathbf{y}_{\ell t} - \mathbf{B}_{W_\ell}\boldsymbol{\mu})^T(\mathbf{B}_{W_\ell}\mathbf{\Omega}\mathbf{B}_{W_\ell}^T)^{-1}(\mathbf{y}_{\ell t} - \mathbf{B}_{W_\ell}\boldsymbol{\mu})\right] \tag{24}$$

$$= C \exp\Big[-\frac{1}{2}\sum_{\ell=1}^{b}\sum_{t=1}^{r_\ell}(\mathbf{y}_{\ell t}-\mathbf{B}_{W_\ell}\boldsymbol{\mu})^T(\mathbf{B}_{W_\ell}\boldsymbol{\Omega}\mathbf{B}_{W_\ell}^T)^{-1}(\mathbf{y}_{\ell t}-\mathbf{B}_{W_\ell}\boldsymbol{\mu})\Big]. \tag{25}$$

Taking the log gives:

$$\log L(\boldsymbol{\mu},\boldsymbol{\Omega}|r_1,\ldots,r_b,\mathbf{y}_{11},\ldots,\mathbf{y}_{br_b}) \tag{26}$$

$$= \log C - \frac{1}{2}\sum_{\ell=1}^{b}\sum_{t=1}^{r_\ell}(\mathbf{y}_{\ell t}-\mathbf{B}_{W_\ell}\boldsymbol{\mu})^T(\mathbf{B}_{W_\ell}\boldsymbol{\Omega}\mathbf{B}_{W_\ell}^T)^{-1}(\mathbf{y}_{\ell t}-\mathbf{B}_{W_\ell}\boldsymbol{\mu}). \tag{27}$$

For the simple case of three-part compositional data with some zeros in component one, and some zeros in component two, the parent distribution for the transformed data is bivariate normal,

$$\mathcal{N}(\boldsymbol{\mu},\boldsymbol{\Omega}) \text{ where } \boldsymbol{\mu}=\begin{bmatrix}\mu_1\\\mu_2\end{bmatrix}, \text{ and } \boldsymbol{\Omega}=\begin{bmatrix}s_{11}&s_{12}\\s_{12}&s_{22}\end{bmatrix} \tag{28}$$

For the full bivariate normal distribution,

$$A=\boldsymbol{\Omega}^{-1}=\frac{1}{s_{11}s_{22}-s_{12}^2}\begin{bmatrix}s_{22}&-s_{12}\\-s_{12}&s_{11}\end{bmatrix}. \tag{29}$$

For the two univariate normal distributions, the inverses of the variances are: $\frac{1}{s_{11}}$ and $\frac{1}{s_{22}}$.

In these formulas,
y_{1j1} is the jth data point among the univariate data from the first component.
y_{2j2} is the jth data point among the univariate data from the second component.
$\mathbf{y}_{3j}$ is a 2-part vector with data from both components,

$$\begin{bmatrix}y_{1j1}\\\odot\end{bmatrix} \qquad\qquad \begin{bmatrix}\odot\\y_{2j2}\end{bmatrix} \qquad\qquad \begin{bmatrix}y_{3j1}\\y_{3j2}\end{bmatrix}. \tag{30}$$

In the example, the y_{1j1} correspond to elements of Table 2, log(spices/rice). The y_{2j2} correspond to elements of Table 3, log(lentils/rice), and y_{3j1} correspond to elements of Table 4, both log(spices/rice) and log(lentils/rice).

We define the means of these matrices in the usual way.

$$\frac{1}{r_1}\sum_{j=1}^{r_1}y_{1j1}=\bar{y}_{11} \qquad\qquad \frac{1}{r_2}\sum_{j=1}^{r_2}y_{2j2}=\bar{y}_{22} \qquad\qquad \frac{1}{r_3}\sum_{j=1}^{r_3}\mathbf{y}_{3j}=\begin{bmatrix}\bar{y}_{31}\\\bar{y}_{32}\end{bmatrix} \tag{31}$$

Partial derivatives

$$\frac{\partial \log L(\boldsymbol{\mu}|\mathbf{Y},\boldsymbol{\Omega},r_1,r_2,r_3)}{\partial \mu_1}=$$
$$\frac{1}{s_{11}}r_1(\bar{y}_{11}-\mu_1)+\frac{s_{22}}{s_{11}s_{22}-s_{12}^2}r_3(\bar{y}_{31}-\mu_1)+\frac{-s_{12}}{s_{11}s_{22}-s_{12}^2}r_3(\bar{y}_{32}-\mu_2) \tag{32}$$

$$\frac{\partial \log L(\boldsymbol{\mu}|\mathbf{Y},\boldsymbol{\Omega},r_1,r_2,r_3)}{\partial \mu_2}=$$
$$\frac{1}{s_{22}}r_2(\bar{y}_{22}-\mu_2)+\frac{s_{11}}{s_{11}s_{22}-s_{12}^2}r_3(\bar{y}_{32}-\mu_2)+\frac{-s_{12}}{s_{11}s_{22}-s_{12}^2}r_3(\bar{y}_{31}-\mu_1) \tag{33}$$

MLE for location, given $\mathbf{\Omega}$

We set the partial derivatives equal to zero, replace μ with $\hat{\mu}$, and solve. The result is:

$$\hat{\mu}_1|\mathbf{\Omega}, r_1, r_2, r_3 = \frac{(r_1\bar{y}_{11} + r_3\bar{y}_{31})(r_2 + r_3)s_{11}s_{12} - r_1\bar{y}_{11}r_2 s_{12}^2 + (\bar{y}_{22} - \bar{y}_{32})r_2 r_3 s_{11}s_{12}}{(r_1 + r_3)(r_2 + r_3)s_{11}s_{22} - r_1 r_2 s_{12}^2} \quad (34)$$

$$\hat{\mu}_2|\mathbf{\Omega}, r_1, r_2, r_3 = \frac{(r_2\bar{y}_{22} + r_3\bar{y}_{32})(r_1 + r_3)s_{11}s_{12} - r_2\bar{y}_{22}r_1 s_{12}^2 + (\bar{y}_{11} - \bar{y}_{31})r_1 r_3 s_{12}s_{22}}{(r_1 + r_3)(r_2 + r_3)s_{11}s_{22} - r_1 r_2 s_{12}^2} \quad (35)$$

In the case where there are no univariate data from the second component, i.e, $r_2 = 0$, we have:

$$\left.(\hat{\mu}_1|\mathbf{\Omega}, r_1, r_2, r_3)\right|_{r_2=0} = \frac{r_1\bar{y}_{11} + r_3\bar{y}_{31}}{(r_3 + r_1)} = \frac{1}{(r_3 + r_1)}\left[\sum_{j=1}^{r_3} y_{3j1} + \sum_{j=1}^{r_1} y_{1j1}\right]. \quad (36)$$

That shows that when we have $r_2 = 0$, the MLE $(\hat{\mu}_1|\mathbf{\Omega}, r_1, r_2, r_3)$ is equal to our simple estimator for μ_1. Similarly, when $r_1 = 0$, $(\hat{\mu}_2|\mathbf{\Omega}, r_1, r_2, r_3)$ is equal to our simple estimator for μ_2. It also turns out that $\left.(\hat{\mu}_1|\mathbf{\Omega}, r_1, r_2, r_3)\right|_{r_3=0} = \bar{y}_{11}$, and when $r_3 = 0$, the simple estimator is also $\bar{y}_{11}$, so they are equal in that case as well.

8.2. Unbiasedness of conditional MLE for 3-part composition

To show that $\hat{\boldsymbol{\mu}}|\mathbf{\Omega}, r_1, r_2, r_3$ is unbiased, we start by pointing out the expectations of the various means:

$$E[\bar{y}_{11}] = E\left[\frac{1}{r_1}\sum_{j=1}^{r_1} y_{1j1}\right] = \frac{1}{r_1}\sum_{j=1}^{r_1} E\left[y_{1j1}\right] = \mu_1 \quad (37)$$

$$E[\bar{y}_{22}] = E\left[\frac{1}{r_2}\sum_{j=1}^{r_2} y_{2j2}\right] = \frac{1}{r_2}\sum_{j=1}^{r_2} E\left[y_{2j2}\right] = \mu_2 \quad (38)$$

$$E\left[\begin{bmatrix} \bar{y}_{31} \\ \bar{y}_{32} \end{bmatrix}\right] = E\left[\frac{1}{r_3}\sum_{j=1}^{r_3} \mathbf{y}_{3j}\right] = \frac{1}{r_3}\sum_{j=1}^{r_3} E\left[\mathbf{y}_{3j}\right] = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \quad (39)$$

When we take the expectation in expression(34), the term with $(\bar{y}_{22} - \bar{y}_{32})$ vanishes because $E[\bar{y}_{22}] = E[\bar{y}_{32}]$. That leaves only terms with $E[\bar{y}_{11}] = \mu_1$ and $E[\bar{y}_{31}] = \mu_1$, which we can factor:

$$E\left[\hat{\mu}_1|\mathbf{\Omega}, r_1, r_2, r_2\right] = \frac{\mu_1\left[(r_1 + r_3)(r_2 + r_3)s_{11}s_{12} - r_1 r_2 s_{12}^2\right]}{(r_1 + r_3)(r_2 + r_3)s_{11}s_{22} - r_1 r_2 s_{12}^2} = \mu_1 \quad (40)$$

This shows that $\hat{\mu}_1$ is unbiased. By symmetry we get that $\hat{\mu}_2$ is unbiased.

8.3. General maximum likelihood estimators

For the general case of MLE for higher dimensions than shown here, the log likelihood can be differentiated, and the score functions can be solved with a computer algebra system. In addition, the Hessian can be checked to verify the solution is a maximum. We have done this for the case of 3-part compositions and do not anticipate any obstacles to extending the program to handle the general case of D-dimensions.

9. Variances of location estimators

Next we find variances of the two location estimators, the MLE, and the simple estimator. Both are unbiased. A question we need to answer is, what is the efficiency of the simple estimator relative to the MLE. We have been using $\hat{\boldsymbol{\mu}}$ for the MLE. We continue to use $\overset{*}{\boldsymbol{\mu}}$ for the simple estimator (of the location). In our discussion,

$$\text{efficiency}(\overset{*}{\mu}_1, \hat{\mu}_1) = \frac{\text{Var}(\hat{\mu}_1)}{\text{Var}(\overset{*}{\mu}_1)}. \tag{41}$$

9.1. Variances of location estimators

The variances of the MLE and the simple location estimator are derived in the Appendix. They are:

$\text{Var}(\hat{\mu}_1|\mathbf{\Omega}, r_1, r_2, r_3) =$

$$\frac{r_3\left((r_3+r_2)^2 s_{11}^3 s_{22}^2 + r_2^2 s_{11}^2 s_{12}^2 s_{22} - 2r_2(r_3+r_2)s_{11}^2 s_{12}^2 s_{22}\right) + r_2 r_3^2 s_{11}^2 s_{12}^2 s_{22} + r_1 s_{11}\left((r_3+r_2)s_{11}s_{22} - r_2 s_{12}^2\right)^2}{\left((r_3^2 + (r_2+r_1)r_3 + r_1 r_2)s_{11}s_{22} - r_1 r_2 s_{12}^2\right)^2} \tag{42}$$

$\text{Var}(\hat{\mu}_2|\mathbf{\Omega}, r_1, r_2, r_3) =$

$$\frac{r_3\left((r_3+r_1)^2 s_{22}^3 s_{11}^2 + r_1^2 s_{22}^2 s_{12}^2 s_{11} - 2r_1(r_3+r_1)s_{22}^2 s_{12}^2 s_{11}\right) + r_1 r_3^2 s_{22}^2 s_{12}^2 s_{11} + r_2 s_{22}\left((r_3+r_1)s_{22}s_{11} - r_1 s_{12}^2\right)^2}{\left((r_3^2 + (r_2+r_1)r_3 + r_1 r_2)s_{11}s_{22} - r_1 r_2 s_{12}^2\right)^2} \tag{43}$$

$$\text{Var}(\overset{*}{\mu}_1|\mathbf{\Omega}, r_1, r_2, r_3) = \frac{s_{11}}{r_1 + r_3}. \tag{44}$$

$$\text{Var}(\overset{*}{\mu}_2|\mathbf{\Omega}, r_1, r_2, r_3) = \frac{s_{22}}{r_2 + r_3}. \tag{45}$$

9.2. Relative efficiency of location estimators

The first thing we show is that when the covariance element of $\mathbf{\Omega}$ is zero, i.e, $s_{12} = 0$, then $\text{Var}(\hat{\mu}) = \text{Var}(\overset{*}{\mu})$.

$\text{Var}(\hat{\mu}_1|\mathbf{\Omega}, r_1, r_2, r_3) =$

$$\frac{r_3\left((r_3+r_2)^2 s_{11}^3 s_{22}^2 + r_2^2 s_{11}^2 s_{12}^2 s_{22} - 2r_2(r_3+r_2)s_{11}^2 s_{12}^2 s_{22}\right) + r_2 r_3^2 s_{11}^2 s_{12}^2 s_{22} + r_1 s_{11}\left((r_3+r_2)s_{11}s_{22} - r_2 s_{12}^2\right)^2}{\left((r_3^2 + (r_2+r_1)r_3 + r_1 r_2)s_{11}s_{22} - r_1 r_2 s_{12}^2\right)^2} \tag{46}$$

Evaluate at $s_{12} = 0$.

$$\text{Var}(\hat{\mu}_1|\mathbf{\Omega}, r_1, r_2, r_3)\Big|_{s_{12}=0} = \frac{r_3\left((r_3+r_2)^2 s_{11}^3 s_{22}^2\right) + r_1 s_{11}\left((r_3+r_2)s_{11}s_{22}\right)^2}{\left((r_3^2 + (r_2+r_1)r_3 + r_1 r_2)s_{11}s_{22}\right)^2} \tag{47}$$

Factor numerator and denominator.

$$\text{Var}(\hat{\mu}_1|\mathbf{\Omega}, r_1, r_2, r_3)\Big|_{s_{12}=0} = \frac{(r_3+r_2)^2 s_{11}^3 s_{22}^2 (r_3 + r1)}{(r_3+r_1)^2 (r_3+r_2)^2 s_{11}^2 s_{22}^2} = \frac{s_{11}}{r_3 + r_1} = \text{Var}(\overset{*}{\mu}_1). \tag{48}$$

Similarly,

$$\text{Var}(\hat{\mu}_2|\mathbf{\Omega}, r_1, r_2, r_3)\Big|_{s_{12}=0} = \frac{s_{22}}{r_3 + r_2} = \text{Var}(\overset{*}{\mu}_2). \tag{49}$$

We have already shown in Section 8.1.2 that when $r_2 = 0$, $\hat{\mu}_1 = \overset{*}{\mu}_1$, and when $r_1 = 0$, $\hat{\mu}_2 = \overset{*}{\mu}_2$; and when $r_3 = 0$, $\hat{\mu}_1 = \overset{*}{\mu}_1$, and $\hat{\mu}_2 = \overset{*}{\mu}_2$. Next we need to compare the variance of $\overset{*}{\mu}$ with the variance of $\hat{\mu}$ in cases where the estimators are not obviously the same.
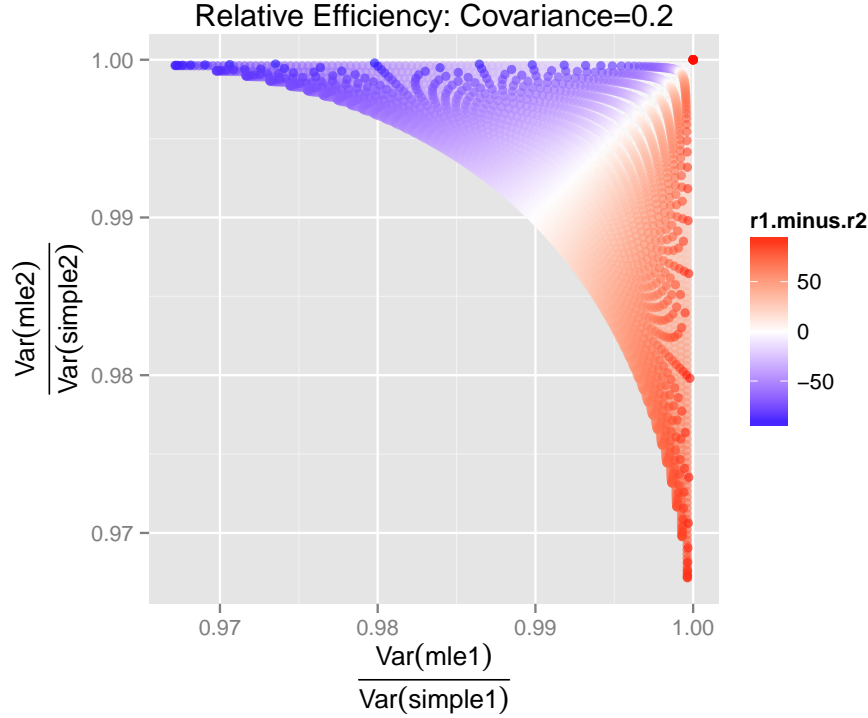
Figure 1: Efficiency of $\overset{*}{\mu}$ relative to $\hat{\mu}$ with low covariance (0.2)

We consider a sample of 100 compositions from a logistic normal distribution with the number of zeros in part 1 ranging from 0 to 100, and similarly for part 2. We calculate the relative efficiency. These are not simulations; they are calculations based on the expressions for the variances of the estimators. We consider all possible combinations of r_1, r_2, r_3 such that $r_1 + r_2 + r_3 = 100$. A larger sample would give roughly the same picture, just with finer granularity. In addition, while we want to understand the effect of the covariance term s_{12} for every possible value between -1 and 1, we get a feel for the space by choosing three values, $s_{12} \in \{0, 0.2, 0.8\}$. For simplicity we choose $s_{11} = s_{22} = 1$.

In all three figures, we plot $\text{Var}(\hat{\mu}_2)/\text{Var}(\overset{*}{\mu}_2)$ versus $\text{Var}(\hat{\mu}_1)/\text{Var}(\overset{*}{\mu}_1)$. In Figure 1 we use a small covariance term, $s_{12} = 0.2$. In Figure 2 we use a large covariance, $s_{12} = 0.8$. In both figures, we shade by the size of r_1 relative to r_2. We already showed in (48) and (49) that when $s_{12} = 0$, the relative efficiency of $\overset{*}{\mu}$ with respect to $\hat{\mu}$ is 1, so there is no plot for $s_{12} = 0$.

Figure 1 shows the relationship between efficiency of $\overset{*}{\mu}_1$ and $\overset{*}{\mu}_2$ and the relative sizes of r_1 and r_2. In the worst case, when $r_1 >> r_2$, the efficiency of $\overset{*}{\mu}_1$ approaches 1, and the efficiency of $\overset{*}{\mu}_2$ falls off toward 0.97. A point to note here is that for a relatively small covariance, 0.2, the simple estimator, $\overset{*}{\mu}$ has a variance almost as small as that of $\hat{\mu}$. We will save discussion of the bands or striations for Figure 3.

Figure 2, which shows efficiency based on a covariance of 0.8, has the same pattern as Figure 1, but with larger variances for $\overset{*}{\mu}$, smaller efficiency. Here the worst cases can have an efficiency of less than 0.5 for either component of $\overset{*}{\mu}$, though when the efficiency of $\overset{*}{\mu}_1$ is that small, the efficiency of $\overset{*}{\mu}_2$ is very near 1.

Figure 3 shows the same points, for a covariance of 0.8, but shaded by the value of r_3. To
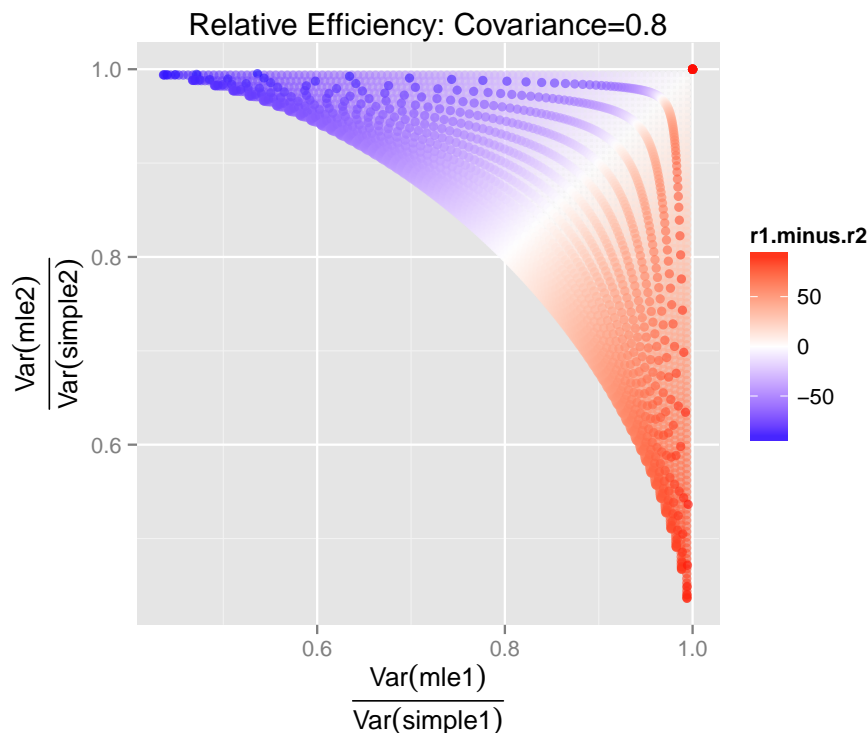
Figure 2: Efficiency of $\overset{*}{\mu}$ relative to $\hat{\mu}$ with high covariance (0.8)

help decipher it, we show a subset of the points in Figure 4.

Figure 4 shows a subset of the points, only the points where $r_3 \in \{1, 2, 3, 4, 61, 62, 63, 64\}$. When r_3 is very small there is a wide range of possibilities for r_1 and r_2. The four leftmost points in the upper left of Figure 4 are points where r_1 is 1 or 2; r_2 is somewhere between 94 and 97, and r_3 is 2, 3, or 4. In these cases, the sample for estimating μ_1 is very small, from 3 to 6 points, some from univariate data and some from the bivariate data. In that case, the MLE has a much smaller variance than the simple estimator. In that same case, there is a much larger sample from univariate data for estimating μ_2, upwards of 90 points, plus a handful of points from the bivariate data. In that case, the difference between the variance of $\overset{*}{\mu}_2$ and $\hat{\mu}_2$ is very small.

Graphs with negative covariances, -0.2, and -0.8 look the same as with positive covariances, and are omitted for the sake of brevity.

9.3. Summary of relative efficiency

Both the simple estimator for the location, $\overset{*}{\boldsymbol{\mu}}$, and the maximum likelihood estimator, $\hat{\boldsymbol{\mu}}$, are unbiased given $\boldsymbol{\Omega}$. The simple estimator's efficiency relative to the MLE tends to decrease as the covariance component of $\boldsymbol{\Omega}$ increases. We say "tends" because even with a covariance of 0.8, there are cases where the efficiency of both components of $\overset{*}{\boldsymbol{\mu}}$ relative to $\hat{\boldsymbol{\mu}}$ is very close to one.

When there are relatively few zeros, and they are balanced, $\overset{*}{\boldsymbol{\mu}}$ has a variance almost as small as $\hat{\boldsymbol{\mu}}$. The more zeros there are, or the more unbalanced their distribution is, the larger the variance of one or more components of the simple estimator.
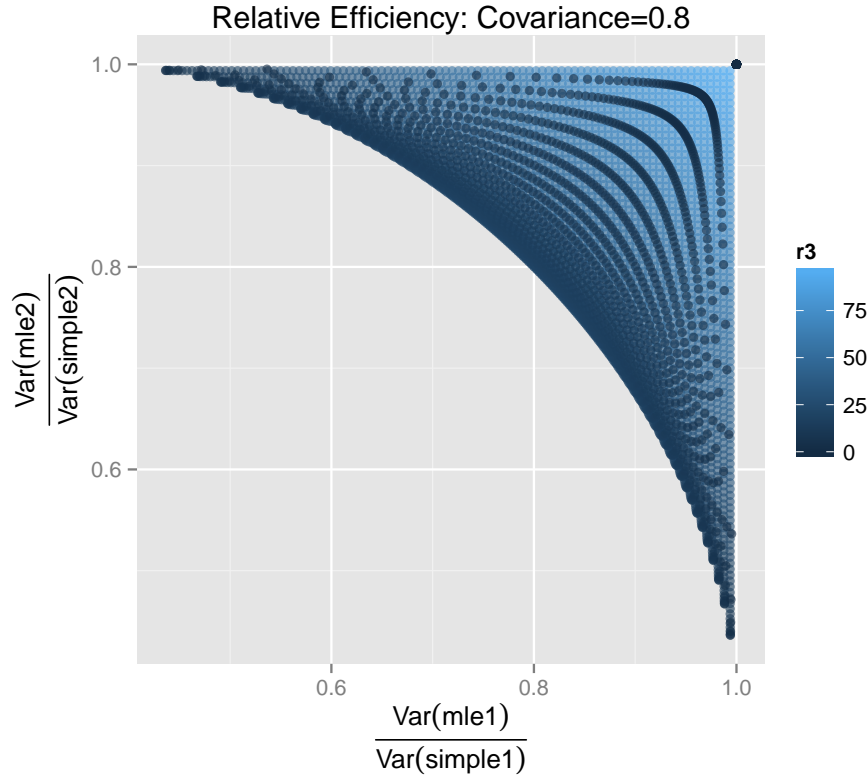
Figure 3: Efficiency of $\overset{*}{\hat\mu}$ relative to $\hat\mu$ with high covariance (0.8) and relative to r_3

10. Subcompositional coherence

One of the reasons for using the logistic normal approach is that, in the base case without zeros, it preserves subcompositional coherence, described by Aitchison and Egozcue (2005) p. 831, as follows, "Subcompositional coherence demands that two scientists, one using full compositions and the other using subcompositions of these full compositions, should make the same inference about relations within the common parts." This implies that the subcomposition of a location estimator equals the location estimator for the subcomposition.

In the presence of zeros, do we maintain this property? It depends on which estimators are used. We have shown that in general when there are zeros, the MLE for the mean is not the same as the simple estimator for the mean. The MLE does not preserve subcompositional coherence when we have zeros. The simple estimators, by construction, do preserve subcompositional coherence provided the same D^{th} component is in both. Thus for inference, there is a choice to be made between maintaining subcompositional coherence and maximizing likelihood.

The issue of the relationship between compositions containing zeros, and subcompositional coherence, has been addressed from other points of view as well. Greenacre (2011) introduced a measure of *subcompositional incoherence* and suggested ways of getting it small enough for practical purposes in the paradigm of correspondence analysis. Scealy and Welsh (2014) argue more generally that although logratio methods for analyzing compositions have their uses, some of the principles that have been used to motivate them, such as subcompositional coherence, should not be taken to be as important as has been argued, e.g., by Aitchison (1994).
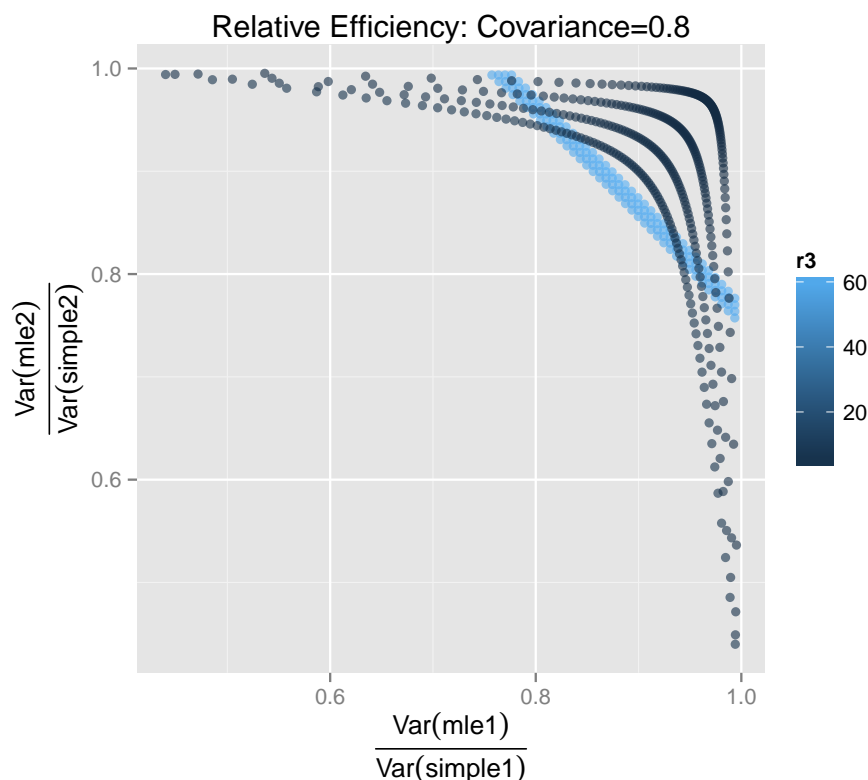
Figure 4: Efficiency of $\overset{*}{\hat{\mu}}$ relative to $\hat{\mu}$ with high covariance (0.8). $r_3 \in \{1:4, 61:64\}$.

11. Discussion

The goal has been to extend the additive logistic normal distribution to cope with essential zeros. We have done that by requiring that the final component of each composition be nonzero, and by projecting compositions with zeros onto smaller dimensional subspaces, thereby addressing the issues of division by zero, and the log of zero. We arrive at a mixture of logistic normals where each distribution has a mean and a covariance parameter which are projections from a common mean and covariance.

We construct two sets of estimators, simple estimators, $\overset{*}{\boldsymbol{\mu}}, \overset{*}{\boldsymbol{\Omega}}$, and maximum likelihood estimators, $\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Omega}}$. These are estimated using all of the compositions in the data, regardless of where the zeros occur, assuming that the Dth component is always nonzero. The simple estimators preserve subcompositional coherence, while the maximum likelihood estimators do not.

There are some limitations to this approach. In addition to the assumption that the Dth part is always nonzero, we assume that each composition has at least one more nonzero part, i.e., the vertices of the simplex are not in the support of the distribution. We assume a common mean and variance. Obviously, for a data set where different zero patterns have different means or variances or both, this model would not be appropriate. It is possible for the simple estimator of the covariance to produce a nonpositive definite matrix. If that happens, one possible approach is to estimate the covariance matrix using only the compositions that do not contain zeros. Another possible approach, once more work is done, would be to use the MLE. Currently, though, we do not have a general software solution for finding the MLE. One last concern is that a data set might have two parts which are never positive at the same time, in which case, the simple estimator for the covariance cannot be found.

In compositional data zeros are a common occurrence. We developed this logistic normal mixture model with the intention of making analysis of such data easier. For future work, we plan to extend existing compositional data methods for inference, graphing, clustering, etc., to work with this distribution.

12. Acknowledgements

13. Appendix

13.1. Variance of location MLE, $\hat{\boldsymbol{\mu}}$

Next we derive the variance of the location MLE, $\hat{\boldsymbol{\mu}}|\boldsymbol{\Omega}, r_1, r_2, r_3$. First we rewrite the expression (34) so that each of the $\bar{y}$ terms stands alone.

$$\hat{\mu}_1|\boldsymbol{\Omega}, r_1, r_2, r_3 =$$
$$\frac{(r_3^2 + r_2 r_3)s_{11}s_{22}\bar{y}_{31} - r_2 r_3 s_{11}s_{12}\bar{y}_{32} + r_2 r_3 s_{11}s_{12}\bar{y}_{22} + ((r_1 r_3 + r_1 r_2)s_{11}s_{22} - r_1 r_2 s_{12}^2)\bar{y}_{11}}{(r_3^2 + (r_2 + r_1)r_3 + r_1 r_2)s_{11}s_{22} - r_1 r_2 s_{12}^2} \tag{50}$$

To find the variance of $\hat{\mu}_1|\boldsymbol{\Omega}, r_1, r_2, r_3$, we need to replace $r_1\bar{y}_{11}$ with $\sum_{j=1}^{r_1} y_{1j1}$; $r_2\bar{y}_{22}$ with $\sum_{j=1}^{r_2} y_{2j2}$; $r_3\bar{y}_{31}$ with $\sum_{j=1}^{r_3} y_{3j1}$; and $r_3\bar{y}_{32}$ with $\sum_{j=1}^{r_3} y_{3j2}$. We also make some other substitutions to simplify the algebra.

Let $k_{31} = (r_3 + r_2)s_{11}s_{22}$. $\tag{51}$

Let $k_{32} = r_2 s_{11}s_{12}$. $\tag{52}$

Let $k_{22} = r_3 s_{11}s_{12}$. $\tag{53}$

Let $k_{11} = (r_3 + r_2)s_{11}s_{22} - r_2 s_{12}^2$. $\tag{54}$

Let $k_{denom} = (r_3^2 + (r_2 + r_1)r_3 + r_1 r_2)s_{11}s_{22} - r_1 r_2 s_{12}^2$. $\tag{55}$

With these in place, we get

$$\hat{\mu}_1|\boldsymbol{\Omega}, r_1, r_2, r_3 = \frac{1}{k_{denom}}\left(k_{31}\sum_{j=1}^{r_3} y_{3j1} - k_{32}\sum_{j=1}^{r_3} y_{3j2} + k_{22}\sum_{j=1}^{r_2} y_{2j2} + k_{11}\sum_{j=1}^{r_1} y_{1j1}\right)$$
$$= \frac{1}{k_{denom}}\left(\sum_{j=1}^{r_3}(k_{31}y_{3j1} - k_{32}y_{3j2}) + k_{22}\sum_{j=1}^{r_2} y_{2j2} + k_{11}\sum_{j=1}^{r_1} y_{1j1}\right). \tag{56}$$

The y_{2j2} are i.i.d. univariate normal; the y_{1j1} are i.i.d. univariate normal; and the $\mathbf{y}_{3j}$ are i.i.d bivariate normal, so the variance of the estimator is:

$$\text{Var}(\hat{\mu}_1|\boldsymbol{\Omega}, r_1, r_2, r_3) =$$
$$\left(\frac{1}{k_{denom}}\right)^2\left[\text{Var}\left(\sum_{j=1}^{r_3}(k_{31}y_{3j1} - k_{32}y_{3j2})\right) + \text{Var}\left(k_{22}\sum_{j=1}^{r_2} y_{2j2}\right) + \text{Var}\left(k_{11}\sum_{j=1}^{r_1} y_{1j1}\right)\right]. \tag{57}$$

$\mathrm{Var}(y_{2j2}) = s_{22}$ and $\mathrm{Var}(y_{1j1}) = s_{11}$, so

$$\mathrm{Var}(\hat{\mu}_1|\boldsymbol{\Omega}, r_1, r_2, r_3) = \left(\frac{1}{k_{denom}}\right)^2 \left[\mathrm{Var}\left(\sum_{j=1}^{r_3}(k_{31}y_{3j1} - k_{32}y_{3j2})\right) + k_{22}^2 r_2 s_{22} + k_{11}^2 r_1 s_{11}\right].$$

(58)

To find the variance of the remaining sum requires the facts that $\mathbf{y}_{3j}$ are i.i.d., and that $\mathrm{Cov}(y_{3j1}, y_{3j2}) = s_{12}$.

$$\begin{aligned}
\mathrm{Var}\left(\sum_{j=1}^{r_3}(k_{31}y_{3j1} - k_{32}y_{3j2})\right) &= \sum_{j=1}^{r_3}\mathrm{Var}(k_{31}y_{3j1} - k_{32}y_{3j2}) \\
&= \sum_{j=1}^{r_3}\left[\mathrm{Var}(k_{31}y_{3j1}) + \mathrm{Var}(k_{32}y_{3j2}) - 2k_{31}k_{32}\mathrm{Cov}(y_{3j1}, y_{3j2})\right] \\
&= \sum_{j=1}^{r_3}\left[k_{31}^2 s_{11} + k_{32}^2 s_{22} - 2k_{31}k_{32}s_{12}\right] \\
&= r_3\left[k_{31}^2 s_{11} + k_{32}^2 s_{22} - 2k_{31}k_{32}s_{12}\right].
\end{aligned}$$

(59)

With that we can write the variance of the MLE, $\hat{\mu}_1$.

$$\mathrm{Var}(\hat{\mu}_1|\boldsymbol{\Omega}, r_1, r_2, r_3) =$$
$$\left(\frac{1}{k_{denom}}\right)^2 \left[r_3\left(k_{31}^2 s_{11} + k_{32}^2 s_{22} - 2k_{31}k_{32}s_{12}\right) + k_{22}^2 r_2 s_{22} + k_{11}^2 r_1 s_{11}\right].$$

(60)

Substituting the values for the k's back in gives:

$$\mathrm{Var}(\hat{\mu}_1|\boldsymbol{\Omega}, r_1, r_2, r_3) =$$
$$\frac{r_3\left((r_3+r_2)^2 s_{11}^3 s_{22}^2 + r_2^2 s_{11}^2 s_{12}^2 s_{22} - 2r_2(r_3+r_2)s_{11}^2 s_{12}^2 s_{22}\right) + r_2 r_3^2 s_{11}^2 s_{12}^2 s_{22} + r_1 s_{11}((r_3+r_2)s_{11}s_{22} - r_2 s_{12}^2)^2}{((r_3^2 + (r_2+r_1)r_3 + r_1 r_2)s_{11}s_{22} - r_1 r_2 s_{12}^2)^2}$$

(61)

Symmetry also gives the variance of $\hat{\mu}_2$ given $\boldsymbol{\Omega}$.

$$\mathrm{Var}(\hat{\mu}_2|\boldsymbol{\Omega}, r_1, r_2, r_3) =$$
$$\frac{r_3\left((r_3+r_1)^2 s_{22}^3 s_{11}^2 + r_1^2 s_{22}^2 s_{12}^2 s_{11} - 2r_1(r_3+r_1)s_{22}^2 s_{12}^2 s_{11}\right) + r_1 r_3^2 s_{22}^2 s_{12}^2 s_{11} + r_2 s_{22}((r_3+r_1)s_{22}s_{11} - r_1 s_{12}^2)^2}{((r_3^2 + (r_2+r_1)r_3 + r_1 r_2)s_{11}s_{22} - r_1 r_2 s_{12}^2)^2}$$

(62)

13.2. Variance of simple location estimator, $\overset{*}{\boldsymbol{\mu}}$

Our simple estimator for the location is $\overset{*}{\boldsymbol{\mu}} = \begin{bmatrix} \overset{*}{\mu}_1 \\ \overset{*}{\mu}_2 \end{bmatrix}$. Here we concern ourselves with $\mathrm{Var}(\overset{*}{\mu}_1)$ and then rely on symmetry to arrive at the variance of $\overset{*}{\mu}_2$.

$$\overset{*}{\mu}_1 = \frac{1}{r_1 + r_3}\left[\sum_{j=1}^{r_1} y_{1j1} + \sum_{j=1}^{r_3} y_{3j1}\right]$$

(63)

$$\begin{aligned}
\mathrm{Var}(\overset{*}{\mu}_1) &= \mathrm{Var}\left(\frac{1}{r_1 + r_3}\left[\sum_{j=1}^{r_1} y_{1j1} + \sum_{j=1}^{r_3} y_{3j1}\right]\right) \\
&= \frac{1}{(r_1 + r_3)^2}\mathrm{Var}\left(\sum_{j=1}^{r_1} y_{1j1} + \sum_{j=1}^{r_3} y_{3j1}\right)
\end{aligned}$$

$$= \frac{1}{(r_1 + r_3)^2}(r_1 s_{11} + r_3 s_{11})$$

$$= \frac{s_{11}}{r_1 + r_3}. \tag{64}$$

By symmetry, $\mathrm{Var}(\overset{*}{\mu}_2) = \dfrac{s_{22}}{r_2 + r_3}. \tag{65}$

References

Aitchison J (1986). *The Statistical Analysis of Compositional Data. Monographs on Statistics and Applied Probability.* Chapman & Hall, Ltd., London (UK). (Reprinted in 2003 with additional material by The Blackburn Press), London (UK).

Aitchison J (1994). *Principles of Compositional Data Analysis*, pp. 73–81. In Anderson, Olkin, and Fang (1994).

Aitchison J, Egozcue JJ (2005). "Compositional Data Analysis: Where Are We and Where Should We Be Heading?" *Mathematical Geology*, **37**(7), 829–850.

Aitchison J, Kay JW (2003). "Possible Solution of Some Essential Zero Problems in Compositional Data Analysis." In *Proceedings of CoDaWork 2003, The First Compositional Data Analysis Workshop.* Universitat de Girona. Departament d'Informàtica i Matemàtica Aplicada. http://ima.udg.edu/Activitats/CoDaWork03/paper_Aitchison_and_Kay.pdf.

Anderson TW, Olkin I, Fang K (eds.) (1994). *Multivariate Analysis and Its Applications.* Institute of Mathematical Statistics, Hayward, CA.

Bacon-Shone J (2008). "Discrete and Continuous Compositions." In *Proceedings of CODA-WORK'08, The 3rd Compositional Data Analysis Workshop, May 27–30, University of Girona, Girona (Spain), CD-ROM.* Universitat de Girona. Departament d'Informàtica i Matemàtica Aplicada.

Billheimer D, Guttorp P, Fagan WF (2001). "Statistical Interpretation of Species Composition." *Journal of the American Statistical Association*, **96**(456), 1205–1214. http://dx.doi.org/10.1198/016214501753381850.

Butler A, Glasbey C (2008). "A Latent Gaussian Model for Compositional Data with Zeros." *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **57**(5), 505–520.

Daunis-i-Estadella J, Martín-Fernández JA (eds.) (2008). *Proceedings of CODAWORK 2008, The 3rd Compositional Data Analysis Workshop.* University of Girona. CD-ROM (ISBN: 978-84-8458-272-4).

Fry JM, Fry TR, Mclaren KR (2000). "Compositional Data Analysis and Zeros in Micro Data." *Applied Economics*, **32**(8), 953–959. http://www.tandfonline.com/doi/pdf/10.1080/000368400322002.

Greenacre M (2011). "Measuring Subcompositional Incoherence." *Mathematical Geosciences*, **43**, 681–693. doi:10.1007/s11004-011-9338-5. URL http://link.springer.com/content/pdf/10.1007/s11004-011-9338-5.pdf.

Kent JT (1982). "The Fisher–Bingham Distribution on the Sphere." *Journal of the Royal Statistical Society, B*, **44**, 71–80.

Leininger TJ, Gelfand AE, Allen JM, Silander Jr JA (2013). "Spatial Regression Modeling for Compositional Data with Many Zeros." *Journal of Agricultural, Biological, and Environmental Statistics*, **18**(3), 314–334.

Martín-Fernández JA, Hron K, Templ M, Filzmoser P, Palarea-Albaladejo J (2014). "Bayesian-Multiplicative Treatment of Count Zeros in Compositional Data Sets." *Statistical Modelling.* doi:10.1177/1471082X14535524.

Martín-Fernández JA, Palarea-Albaladejo J, Olea RA (2011). *Dealing With Zeros*, pp. 43–58. In Pawlowsky-Glahn and Buccianti (2011).

Palarea-Albaladejo J, Martín-Fernández J (2008). "A Modified EM Alr-Algorithm for Replacing Rounded Zeros in Compositional Data Sets." *Computers & Geosciences*, **34**(8), 902–917.

Palarea-Albaladejo J, Martín-Fernández JA (2015). "zCompositions - R Package for Multivariate Imputation of Left-Censored Data under a Compositional Approach." *Chemometrics and Intelligent Laboratory Systems*, **143**, 85–96. doi:10.1016/j.chemolab.2015.02.019.

Palarea-Albaladejo J, Martín-Fernández JA, Olea RA (2014). "A Bootstrap Estimation Scheme for Chemical Compositional Data with Nondetects." *Journal of Chemometrics*, **28**(7), 585–599.

Pawlowsky-Glahn V, Buccianti A (eds.) (2011). *Compositional Data Analysis: Theory and Applications.* Wiley.

Scealy J, Welsh A (2011). "Regression for Compositional Data by Using Distributions Defined on the Hypersphere." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **73**(3), 351–375.

Scealy J, Welsh A (2014). "Colours and Cocktails: Compositional Data Analysis 2013 Lancaster Lecture." *Australian & New Zealand Journal of Statistics*, **56**(2), 145–169.

Stewart C, Field C (2011). "Managing the Essential Zeros in Quantitative Fatty Acid Signature Analysis." *Journal of Agricultural, Biological, and Environmental Statistics*, **16**(1), 45–69.

Affiliation:

John Bear
Statistical Consulting Lab
University of Arizona
888 N Euclid Ave
85719-4824 Tucson, AZ, USA
E-mail: jbear@email.arizona.edu

Dean Billheimer
Statistical Consulting Lab
University of Arizona
888 N Euclid Ave
85719-4824 Tucson, AZ, USA
E-mail: dean.billheimer@arizona.edu