properties of a square root transformation regression model

J. L. Scealy[1] and A. H. Welsh[1]

[1]Centre for Mathematics and its Applications- Australian National University, Australia janice.scealy@anu.edu.au

We consider the problem of modelling the conditional distribution of a response given a vector of covariates $\boldsymbol{x}$ when the response is a compositional data vector $\boldsymbol{u}$. That is, $\boldsymbol{u}$ is defined on the unit simplex

$$\Delta^{p-1} = \{(u_1, u_2, \ldots, u_p)^T : \quad u_i \geq 0 \quad (i = 1, 2, \ldots, p), \quad \sum_{i=1}^{p} u_i = 1\}.$$

This definition of the unit simplex differs subtly from that of Aitchison (1982), as we relax the condition that the components of $\boldsymbol{u}$ must be strictly positive. Under this scenario, use of the ratio (or logratio) to compare different compositions is not ideal since it is undefined in some instances, and subcompositional analysis is also not appropriate due to the possibility of division by zero. It has long been recognised that the square root transformation

$$\boldsymbol{y} = \sqrt{\boldsymbol{u}} = (\sqrt{u_1}, \sqrt{u_2}, \ldots, \sqrt{u_p})^T$$

transforms compositional data (including zeros) onto the surface of the $(p-1)$-dimensional hypersphere (see for example Mardia and Jupp (2000), page 160) and that this opens the possibility of using distributions for directional data to model compositional data. One advantage of this approach over the logratio approach is that zero components are catered for naturally in the models. Stephens (1982) modelled $\boldsymbol{y}$ by a p-dimensional von Mises-Fisher distribution and Wang et. al. (2007) recently described an ad hoc approach to modelling $\boldsymbol{y}$. Scealy and Welsh (2011) proposed the Kent distribution to model $\boldsymbol{y}$ and the advantage of this approach is that unlike the von Mises-Fisher distribution, the Kent distribution has a sufficiently general covariance structure. In this talk we first summarise the additive and multiplicative Kent regression models for compositional data proposed by Scealy and Welsh (2011) and describe some of the properties of these models. The two Kent regression models are named 'additive' and 'multiplicative' because in these models the square root of the additive and multiplicative logistic transformations (Aitchison, 1986, pages 113 and 131) is respectively applied to link the mean direction of the Kent distribution with $p-1$ linear functions of $\boldsymbol{x}$. We then compare the logratio approach with the square root transformation approach and show that in certain cases the analysis based on the Kent distribution will work better than the logistic normal compositional models of Aitchison (1986), even when there are no zero components in $\boldsymbol{u}$. In summary, the additive Kent regression model has a permutation equivariance property and the multiplicative Kent regression model has an asymptotic (large concentration) amalgamation property similar to Aitchison (1986, Property 6.12(a), page 132). The square root transformation method can therefore lead to models with nice properties, and similar to Aitchison's logratio transformations, the square root transformation is also scale invariant.

1 The Kent Distribution

We begin by summarising the Kent distribution and its properties. For further details see Kent (1982) and Mardia and Jupp (2000). Let $S^{p-1} = \{\boldsymbol{y} \in R^p : \|\boldsymbol{y}\| = 1\}$ denote the unit hypersphere in R^p. The Kent distribution for general $p \geq 3$ is defined by the density

$$f(\boldsymbol{y}) \quad = \quad c(\kappa, \boldsymbol{\beta})^{-1} \exp[\kappa \boldsymbol{\mu}^T \boldsymbol{y} + \beta_2\{\boldsymbol{\gamma}_2^T \boldsymbol{y}\}^2 + \ldots + \beta_{p-1}\{\boldsymbol{\gamma}_{p-1}^T \boldsymbol{y}\}^2 - (\beta_2 + \ldots + \beta_{p-1})\{\boldsymbol{\gamma}_p^T \boldsymbol{y}\}^2],$$

for $\boldsymbol{y} = (y_1, y_2, \ldots, y_p)^T \in S^{p-1}$, which contains a total of $\frac{p(p-1)}{2}$ location parameters and $p-1$ scalar shape parameters. In total the Kent distribution contains the same number of parameters as the $p-1$ dimensional normal distribution. We assume $\kappa > 0$ is a shape parameter specifically describing the concentration and $\boldsymbol{\beta} = (\beta_2, \ldots, \beta_{p-1})^T \in R^{p-2}$ are also shape parameters, but defined according to

$$\frac{\kappa}{2} > \beta_2 \geq \beta_3 \geq \ldots \geq \beta_{p-1} \geq -(\beta_2 + \beta_3 + \ldots + \beta_{p-1}). \qquad (1)$$

The normalising constant in the density is denoted by $c(\kappa, \boldsymbol{\beta})$, which is a function of the shape parameters only. The $\frac{p(p-1)}{2}$ location parameters are associated with a $p \times p$ orthogonal matrix

$$\boldsymbol{\Gamma} = (\boldsymbol{\mu}, \boldsymbol{\gamma}_2, \ldots, \boldsymbol{\gamma}_p),$$

where $\boldsymbol{\mu} \in S^{p-1}$ is the mean direction and the vectors $\boldsymbol{\gamma}_2, \boldsymbol{\gamma}_3, \ldots \boldsymbol{\gamma}_p$ are the set of principal axes orthogonal to $\boldsymbol{\mu}$. It is useful to further parametrise $\boldsymbol{\Gamma}$ by writing

$$\boldsymbol{\Gamma} = \boldsymbol{H}\boldsymbol{K} = (\boldsymbol{\mu}, \boldsymbol{H}^*\boldsymbol{K}^*),$$

where

$$\boldsymbol{H} = (\boldsymbol{\mu}, \boldsymbol{H}^*) \quad \text{and} \quad \boldsymbol{K} = \begin{pmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & \boldsymbol{K}^* \end{pmatrix}$$

are both $p \times p$ orthogonal matrices. The matrix $\boldsymbol{H}$ is a function of $\boldsymbol{\mu}$ only and the $(p-1) \times (p-1)$ matrix $\boldsymbol{K}^*$ is a function of $(p-1)(p-2)/2$ scalar parameters.

The transformation $\boldsymbol{y}^* = \boldsymbol{\Gamma}^T\boldsymbol{y}$ is called the population standardising transformation because its density is a function of the shape parameters only and hence $\boldsymbol{y}^* = (y_1^*, y_2^*, \ldots, y_p^*)^T \in S^{p-1}$ can be regarded as a standardised form of the Kent distribution. For the Kent distribution,

$$\mathrm{E}(\boldsymbol{y}) = \boldsymbol{\mu}\,\mathrm{E}(y_1^*) \quad \text{and} \quad \mathrm{E}(\boldsymbol{y}\boldsymbol{y}^T) = \boldsymbol{\Gamma}\mathrm{diag}\left(\mathrm{E}(y_1^{*2}), \mathrm{E}(y_2^{*2}), ..., \mathrm{E}(y_p^{*2})\right)\boldsymbol{\Gamma}^T, \tag{2}$$

where $\mathrm{diag}(\boldsymbol{w})$ is the diagonal matrix with the vector $\boldsymbol{w}$ on the diagonal and hence $\boldsymbol{\mu}$ can be interpreted as the mean direction.

2 Regression Models

In this section we summarise the Kent regression models that were proposed by Scealy and Welsh (2011). Suppose we have vector responses $\{\boldsymbol{y}_i \in S^{p-1} : i = 1, \ldots, n\}$ associated with a set of covariates $\{\boldsymbol{x}_i \in R^q : i = 1, \ldots, n\}$. It is assumed that the responses are conditionally independent given the covariates and are distributed according to the Kent distribution. In a regression setting we could have all or only a subset of the model parameters dependent on $\boldsymbol{x}$. The models of Scealy and Welsh assume $\boldsymbol{\mu} = \boldsymbol{\mu}(\boldsymbol{x})$ is a function of $\boldsymbol{x}$ only, with the remaining parameters κ, $\boldsymbol{\beta}$ and $\boldsymbol{K}^*$ being constant and not dependent on $\boldsymbol{x}$. Note that this is arguably the most natural and straight forward approach and is analogous in a sense to modelling the mean of a multivariate normal vector as a function of $\boldsymbol{x}$ and assuming the covariance matrix is constant.

2.1 The Additive Kent Regression Model

The additive Kent regression model links $\boldsymbol{\mu}$ to $p-1$ linear functions of $\boldsymbol{x}$ by using the square root of the additive logistic transformation. That is,

$$\mu_i(\boldsymbol{x}) = \begin{cases} \left(1 + \sum_{j=1}^{p-1} \exp\left(\boldsymbol{a}_j^T\boldsymbol{x}\right)\right)^{-\frac{1}{2}} & i = 1 \\ \exp\left(\frac{\boldsymbol{a}_{i-1}^T\boldsymbol{x}}{2}\right)\left(1 + \sum_{j=1}^{p-1} \exp\left(\boldsymbol{a}_j^T\boldsymbol{x}\right)\right)^{-\frac{1}{2}} & i = 2, 3, ..., p, \end{cases} \tag{3}$$

where $\mu_i(\boldsymbol{x})$ is the ith component of $\boldsymbol{\mu}(\boldsymbol{x})$ and $\boldsymbol{a} = \left(\boldsymbol{a}_1^T, \boldsymbol{a}_2^T, \ldots, \boldsymbol{a}_{p-1}^T\right)^T$ is a $q(p-1)$ dimensional vector of regression coefficients defined on $R^{q(p-1)}$. In a regression context $\boldsymbol{H} = \boldsymbol{H}(\boldsymbol{x})$, and a symmetric choice of $\boldsymbol{H}(\boldsymbol{x})$ is

$$\boldsymbol{H}(\boldsymbol{x}) = \begin{pmatrix} \mu_1 & \boldsymbol{\mu}_L(\boldsymbol{x})^T \\ \boldsymbol{\mu}_L(\boldsymbol{x}) & \boldsymbol{H}_L^*(\boldsymbol{x}) \end{pmatrix} = \begin{pmatrix} \mu_1(\boldsymbol{x}) & \boldsymbol{\mu}_L(\boldsymbol{x})^T \\ \boldsymbol{\mu}_L(\boldsymbol{x}) & \frac{1}{1+\mu_1(\boldsymbol{x})}\boldsymbol{\mu}_L(\boldsymbol{x})\boldsymbol{\mu}_L(\boldsymbol{x})^T - \boldsymbol{I}_{p-1} \end{pmatrix}, \tag{4}$$

where $\boldsymbol{I}_{p-1}$ is the $(p-1) \times (p-1)$ identity matrix and $\boldsymbol{\mu}_L(\boldsymbol{x}) = (\mu_2(\boldsymbol{x}), \mu_3(\boldsymbol{x}), ..., \mu_p(\boldsymbol{x}))^T$. As proved by Scealy and Welsh (2011), this choice of $\boldsymbol{H}(\boldsymbol{x})$ is unique in the sense that it is the only possible

symmetric orthogonal matrix with first column $\boldsymbol{\mu}(\boldsymbol{x})$ and with a symmetric non-indefinite value of $\boldsymbol{H}_L^*(\boldsymbol{x})$.

The additive Kent regression model is defined using (3) and (4). As Scealy and Welsh (2011) proved, this model has a permutation equivariance property. That is, let $\boldsymbol{A} = \begin{pmatrix}\boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_{p-1}\end{pmatrix}^T$ and let $\boldsymbol{y}^{(C)} = \boldsymbol{C}\boldsymbol{y}$ be a permutation of $\boldsymbol{y}$ that leaves y_1 fixed (only $(y_2, y_3, ..., y_p)^T$ gets permuted). Then $\boldsymbol{y}^{(C)}$ follows an additive Kent regression model with $\boldsymbol{A}$ and $\boldsymbol{K}^*$ both permuted, and unchanged shape parameters.

2.2 The Multiplicative Kent Regression Model

The multiplicative Kent regression model links $\boldsymbol{\mu}$ to $p-1$ linear functions of $\boldsymbol{x}$ by using the square root of the multiplicative logistic transformation. That is,

$$\mu_i(\boldsymbol{x}) = \begin{cases} \exp\left(\frac{\boldsymbol{a}_i^T\boldsymbol{x}}{2}\right)\left(\prod_{j=1}^{i}\left(1+\exp\left(\boldsymbol{a}_j^T\boldsymbol{x}\right)\right)\right)^{-\frac{1}{2}} & i = 1, 2, ..., p-1 \\ \left(\prod_{j=1}^{p-1}\left(1+\exp\left(\boldsymbol{a}_j^T\boldsymbol{x}\right)\right)\right)^{-\frac{1}{2}} & i = p. \end{cases} \tag{5}$$

or equivalently

$$\mu_i(\boldsymbol{x}) = \begin{cases} \cos\{\alpha_1(\boldsymbol{x})\} & i = 1 \\ \left(\prod_{j=1}^{i-1}\sin\{\alpha_j(\boldsymbol{x})\}\right)\cos\{\alpha_i(\boldsymbol{x})\} & i \neq 1,\ i \neq p-1,\ i \neq p \\ \left(\prod_{j=1}^{p-2}\sin\{\alpha_j(\boldsymbol{x})\}\right)\cos\{\alpha_{p-1}(\boldsymbol{x})\} & i = p-1 \\ \left(\prod_{j=1}^{p-2}\sin\{\alpha_j(\boldsymbol{x})\}\right)\sin\{\alpha_{p-1}(\boldsymbol{x})\} & i = p, \end{cases} \tag{6}$$

where

$$\alpha_j(\boldsymbol{x}) = \arctan\left(\exp\left(-\frac{(\boldsymbol{a}_j^T\boldsymbol{x})}{2}\right)\right) \quad \text{for} \quad j = 1, 2, \ldots, p-1,$$

and the arctan function is defined on $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$. For the multiplicative Kent regression model, $\boldsymbol{H}(\boldsymbol{x})$ is defined using a non-symmetric matrix as follows. Let $\boldsymbol{R}_{ij}(\alpha)$ be a $p \times p$ plane rotation matrix, that is $\boldsymbol{R}_{ij}(\alpha)$ is the $p \times p$ identity matrix with the (i,i), (i,j), (j,i) and (j,j) elements replaced by $\cos\alpha$, $-\sin\alpha$, $\sin\alpha$ and $\cos\alpha$ respectively. Then let

$$\boldsymbol{H}(\boldsymbol{x}) = \boldsymbol{R}_{(p-1)(p)}\left(\alpha_{p-1}(\boldsymbol{x})\right)\boldsymbol{R}_{(p-2)(p-1)}\left(\alpha_{p-2}(\boldsymbol{x})\right)\cdots\boldsymbol{R}_{12}\left(\alpha_1(\boldsymbol{x})\right). \tag{7}$$

The advantage of this choice is that it contains the smallest number of plane rotation matrix products resulting in first column equal to $\boldsymbol{\mu}(\boldsymbol{x})$ from (5).

The multiplicative Kent regression model has an asymptotic (large κ) amalgamation property similar to Aitchison (1986, Property 6.12 (a), page 132). That is, assume that $\boldsymbol{y}$ follows a $p-$dimensional multiplicative Kent regression model, then for large κ the amalgamation

$$\boldsymbol{y}^a = (y_1, y_2, ..., y_q, \sqrt{y_{q+1}^2 + ... + y_p^2})^T$$

is distributed the same as a $(q+1)$-dimensional multiplicative Kent regression model. See Scealy and Welsh (2011) for further details about asymptotic large κ properties of the Kent regression models.

3 Comparing the Square Root and the Logratio Transformation Approaches

The logratio transformation approach is one of the most widely used techniques for analysing compositional data. In particular, the additive logratio transformation defined by

$$\breve{y}_j = \log\frac{u_{j+1}}{u_1}, \quad j = 1, 2, \ldots, p-1, \tag{8}$$

where $\breve{\boldsymbol{y}} = (\breve{y}_1, \breve{y}_2, \ldots, \breve{y}_{p-1})^T \in R^{p-1}$, is often applied. The standard approach is to then assume $\breve{\boldsymbol{y}}$ has a $p-1$ dimensional normal distribution, which results in $\boldsymbol{u}$ having an additive logistic normal distribution (Aitchison, 1986, page 113) and estimation of model parameters is straightforward.

The square root transformation $\boldsymbol{y} = \sqrt{\boldsymbol{u}}$ transforms compositional data onto the unit hypersphere in R^p and we assume $\boldsymbol{u}$ has a squared Kent distribution. One advantage of this approach is that it handles zeros, however there are two issues which are pointed out by Scealy and Welsh (2011):

(a) The normalising constant $c(\kappa, \boldsymbol{\beta})$ does not exist in a convenient form and involves multidimensional integrals which are difficult to evaluate when p is not small.

(b) The square root transformation maps compositional data directly onto the positive orthant of the hypersphere only. When the components of $\boldsymbol{u}$ are distributed close to the boundaries, the fitted Kent model may have significant support outside the positive orthant.

To overcome issue (b), we could instead assume $\sqrt{\boldsymbol{u}}$ follows a folded Kent distribution on the positive orthant (i.e. if $\boldsymbol{y}$ has a Kent distribution, then $\mid \boldsymbol{y} \mid = (\mid y_1 \mid, \mid y_2 \mid, ..., \mid y_p \mid)^T$ has a folded Kent distribution). However, the folded Kent distribution in $p-$dimensions is difficult to work with. As Scealy and Welsh (2011) mentioned, in many applications we expect the majority of the data to be distributed away from the boundaries 0 and 1, and as a consequence the concentration parameter κ will also be at least moderately large. This justifies the use of large κ approximations to address issues (a) and (b). That is, for large κ the folding can be effectively ignored in estimation and an approximation for the normalising constant is also available (see Scealy and Welsh (2011) for further details).

We now demonstrate a particular situation where the squared Kent distribution will work better than the logistic normal distribution. Suppose we have a compositional data sample and we observe that a component $\sqrt{u_1}$ appears to have a mean close to zero with only a small amount of folding present (the distribution is roughly symmetric looking). In this situation the square root transformation coupled with the Kent distribution will be a reasonable approach and from the large κ results in Scealy and Welsh (2011), it follows that $\sqrt{u_1}$ is approximately normal with small mean and variance. We now undertake a simulation study to compare the logratio approach with the square root transformation approach under this scenario.

First, we simulate 10,000 $\sqrt{u_1}$ observations from the folded normal distribution with mean 0.03 and standard deviation 0.01. Plots 1 and 2 in Figure 1 contain a histogram and boxplot of $\sqrt{u_1}$. We then apply the logratio transformation $\breve{y}_1 = \log\left(\frac{1-u_1}{u_1}\right)$ and plots 3 and 4 contain the resulting histogram and boxplot of $\breve{y}_1$. The treatment of $\breve{y}_1$ as normal is not sensible here since the distribution is highly skewed. The problem is that values of u_1 near zero which are close to each other on the original scale, are quite far apart on the logratio scale. We now consider the scenario where u_1 is not distributed close to zero and repeat the above simulation but use a mean of 0.1 instead of 0.03. Figure 2 contains the resulting histograms and boxplots of $\sqrt{u_1}$ and $\breve{y}_1$. In this case, both distributions look roughly symmetric and either the logratio or the square root transformation may be appropriate. The differences between the two approaches becomes more apparent when the data is distributed closer to the boundaries.

The square root transformation has some nice properties and as we will show, adheres to two out of the three principles of compositional data analysis described by Egozcue (2009). Consider the following definition of the simplex (in the abstract we set $c = 1$)

$$\Delta^{p-1} = \{(u_1, u_2, \ldots, u_p)^T : \quad u_i \geq 0 \quad (i = 1, 2, \ldots, p), \quad \sum_{i=1}^{p} u_i = c\}, \tag{9}$$

where $c > 0$ is a known constant. In this case the square root transformation is

$$\boldsymbol{y} = \frac{1}{\sqrt{u_1 + u_2 + ... + u_p}} \left(\sqrt{u}_1, \sqrt{u}_2, ..., \sqrt{u}_p\right)^T,$$

which is always defined since c is assumed greater than zero. This transformation is permutation invariant (see Egozcue (2009) for a definition) and it is also scale invariant since $\boldsymbol{y}(\lambda \boldsymbol{u}) = \boldsymbol{y}(\boldsymbol{u})$ for $\lambda >$

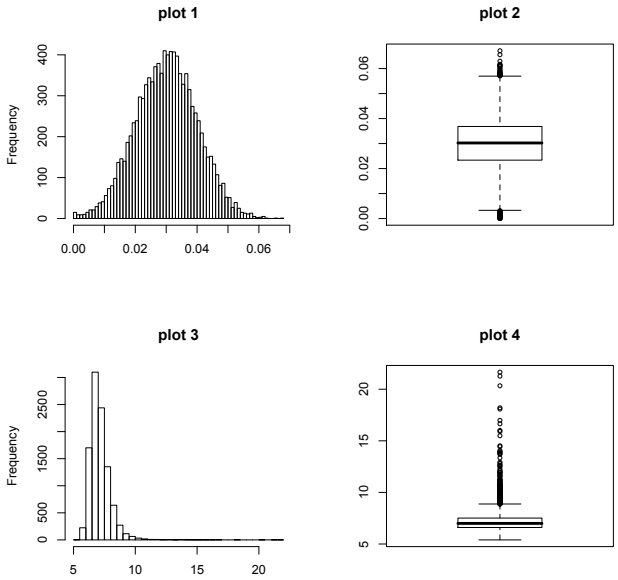

Figure 1: Histograms and boxplots of simulated $\sqrt{u_1}$ and $\breve{y}_1$ values for mean 0.03.

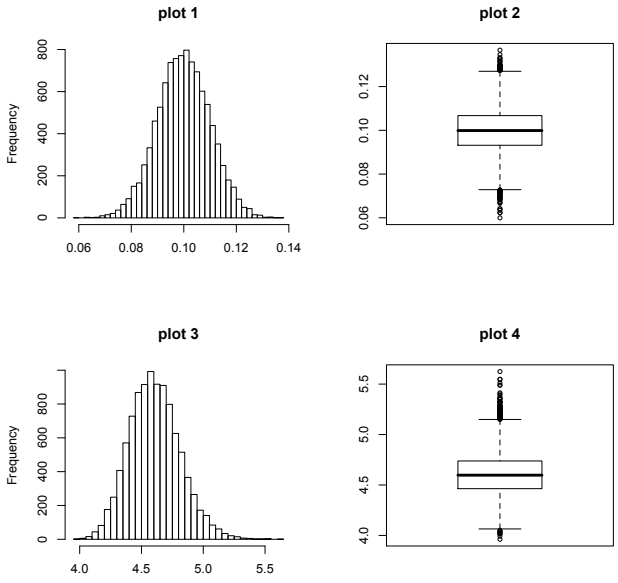

Figure 2: Histograms and boxplots of simulated $\sqrt{u_1}$ and $\breve{y}_1$ values for mean 0.1.

0. The square root transformation therefore satisfies two of the compositional data analysis principles. The third principal called compositional coherence is violated due to the fact that subcompositions are not compatible with our definition of the simplex given at (9). That is, suppose $u_1 = 0$ and $u_2 = 0$, then the subcomposition

$$\left(\frac{u_1}{u_1 + u_2}, \frac{u_2}{u_1 + u_2} \right)^T$$

is undefined and is not in Δ^1 since $u_1 + u_2 = 0$ (recall that $c > 0$ in definition (9)).

For predictive purposes it is of interest to estimate the quantity $\mathrm{E}(\boldsymbol{u})$. In the case of the squared Kent distribution, it follows from (2) that

$$\mathrm{E}(\boldsymbol{u}) = \boldsymbol{\Gamma}^2 \mathrm{E}\left(\boldsymbol{y}^{*2}\right),$$

where the squared operation on the above matrix and vector is a component wise square. This expectation can be estimated directly by plugging in estimates of the model parameters (see Scealy and Welsh (2011) for further details). However, in the case of the logistic normal distribution, estimation of $\mathrm{E}(\boldsymbol{u})$ is not so straightforward. Suppose we apply the additive logratio transformation (8) and assume $\breve{\boldsymbol{y}} \sim N_{p-1}(\boldsymbol{b}, \boldsymbol{\Sigma})$ and let $\hat{\boldsymbol{b}}$ and $\hat{\boldsymbol{\Sigma}}$ be estimates of the mean $\boldsymbol{b}$ and covariance matrix $\boldsymbol{\Sigma}$ respectively. Then, to avoid backtransformation bias, we need to compute the following integrals

$$\mathrm{E}(u_1) = \int_{R^{p-1}} \frac{1}{1 + \sum_{m=1}^{p-1} \exp(\breve{y}_m)} f(\breve{\boldsymbol{y}}) d\breve{\boldsymbol{y}},$$

and for $j = 2, 3, ...p$

$$\mathrm{E}(u_j) = \int_{R^{p-1}} \frac{\exp(\breve{y}_{j-1})}{1 + \sum_{m=1}^{p-1} \exp(\breve{y}_m)} f(\breve{\boldsymbol{y}}) d\breve{\boldsymbol{y}},$$

where $f(\breve{\boldsymbol{y}})$ is the multivariate normal density of $\breve{\boldsymbol{y}}$. Note that an estimate of $\mathrm{E}(\boldsymbol{u})$ is obtained by plugging in $\hat{\boldsymbol{b}}$ and $\hat{\boldsymbol{\Sigma}}$ in $f(\breve{\boldsymbol{y}})$. These multidimensional integrals are difficult to evaluate when p is not small, and we need to use for example Monte Carlo methods to compute these which is a disadvantage.

One other major issue with the logratio transformation is that it does not allow any zero components in the compositions. To apply the logratio approach in this case, we could adopt one of the strategies suggested by Aitchison (1986, pages 266-274), namely collapsing categories, modelling zeros separately or replacing zeros by small positive values. In general collapsing categories is not ideal due to the fact that information is lost during this process. Aitchison and Kay (2004) considered modelling zeros separately, but as these authors stated the computations needed are awkward (for example, some multidimensional integrals need to be computed using a MCMC approach). Zero replacement strategies for rounded zeros have been considered by various authors (for example, see Fry et. al. (2000), Martín-Fernández et. al. (2003) and Palarea-Albaladejo et. al. (2007)). These zero replacement strategies are imputation methods and the drawback is that by treating the zeros as missing values, imputation variance is introduced which will ultimately affect estimates of both $\boldsymbol{b}$ and $\boldsymbol{\Sigma}$. Notice that $\mathrm{E}(\boldsymbol{u})$ in the logistic normal distribution case is a function of both $\boldsymbol{b}$ and $\boldsymbol{\Sigma}$, and the estimator of $\mathrm{E}(\boldsymbol{u})$ relies on having good estimates of both $\boldsymbol{b}$ and $\boldsymbol{\Sigma}$ available.

4 An Example

We now compare the logratio approach with the square root transformation approach using Data 34 from Appendix D of Aitchison (1986). Aitchison (1986) used this dataset to demonstrate a zero replacement strategy since 5 out of 30 observations contained a zero component. Bacon-Shone (1992), Martín-Fernández et. al. (2003) and Scealy and Welsh (2011) also analysed this dataset and used it to compare different methodologies. This dataset consists of foraminifer, a marine micro-organism, at 30 different depths and contains $p = 4$ components.

Figure 3 contains plots of $\sqrt{u_1}, \sqrt{u_2}, \sqrt{u}_3$ and $\sqrt{u}_4$ versus depth (the ordering of the components is the same as given in Appendix D of Aitchison (1986)). Interestingly, there appears to be an outlier for $\sqrt{u_4}$ (observation 24 appears to be far away from the rest of the data). Scealy and Welsh (2011)

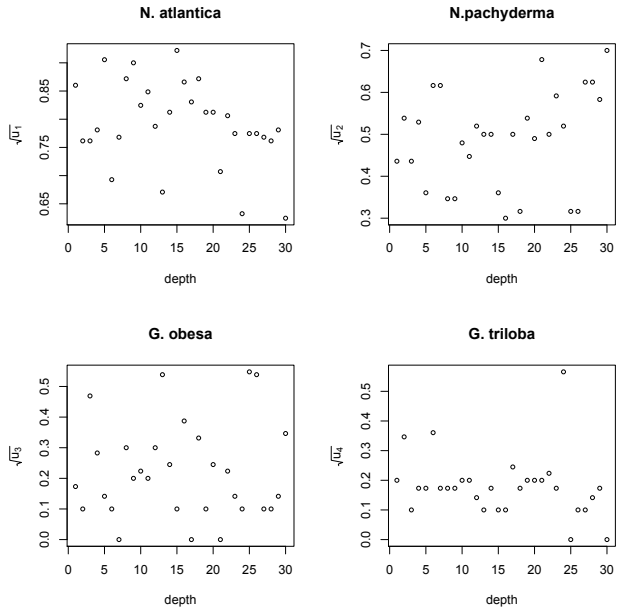

Figure 3: Plots of $\sqrt{u_1}$, $\sqrt{u_2}$, $\sqrt{u_3}$ and $\sqrt{u_4}$ versus depth.

also identified this observation as an outlier, but Aitchison (1986), Bacon-Shone (1992) and Martín-Fernández et. al. (2003) did not. We suspect that the G. obsea and G. triloba proportions may have been swapped by mistake for observation 24. If we make this swap, then there appears to be a relationship between $\sqrt{u_4}$ and depth ($\sqrt{u_4}$ appears to decrease with depth). Both $\sqrt{u_3}$ and $\sqrt{u_4}$ are distributed close to zero and the distributions look roughly symmetric about the means, implying that the square root transformation methodology of Scealy and Welsh (2011) is reasonable for this dataset.

Scealy and Welsh (2011) applied the multiplicative Kent regression model to this example dataset and showed that the model appeared to fit the data well. Here we apply the additive Kent regression model instead and we model the mean direction $\boldsymbol{\mu}(\boldsymbol{x})$ using

$$\mu_1(\boldsymbol{x}) = \frac{1}{\left(1+\sum_{j=1}^{3}\exp\left(a_{j1}+a_{j2}x\right)\right)^{\frac{1}{2}}}, \quad \mu_2(\boldsymbol{x}) = \frac{\exp\left(\frac{a_{11}+a_{12}x}{2}\right)}{\left(1+\sum_{j=1}^{3}\exp\left(a_{j1}+a_{j2}x\right)\right)^{\frac{1}{2}}},$$

$$\mu_3(\boldsymbol{x}) = \frac{\exp\left(\frac{a_{21}+a_{22}x}{2}\right)}{\left(1+\sum_{j=1}^{3}\exp\left(a_{j1}+a_{j2}x\right)\right)^{\frac{1}{2}}}, \quad \text{and} \quad \mu_4(\boldsymbol{x}) = \frac{\exp\left(\frac{a_{31}+a_{32}x}{2}\right)}{\left(1+\sum_{j=1}^{3}\exp\left(a_{j1}+a_{j2}x\right)\right)^{\frac{1}{2}}},$$

where x represents depth and a_{11}, a_{12}, a_{21}, a_{22}, a_{31} and a_{32} are six scalar regression coefficients. We estimate the parameters associated with this additive Kent regression model by applying the estimation techniques in Scealy and Welsh (2011). Note that before fitting the model we corrected the suspected typo in observation 24. After fitting the model, a_{22} did not appear to be significant and so it was removed, and then the model was refitted with the remaining five regression coefficients only.

Figure 4 contains plots of the observed values of $\boldsymbol{u}$ versus depth and the lines on the plots are the estimates of $\mathrm{E}(\boldsymbol{u}|\boldsymbol{x})$ (the correction to observation 24 has also been implemented in the plots). The estimates of $\mathrm{E}(\boldsymbol{u}|\boldsymbol{x})$ are very similar to those obtained by Scealy and Welsh (2011) in the case of the multiplicative Kent regression model and therefore the additive Kent regression model also appears to fit this data.

An alternative approach is to fit linear models to the additive logratio transformed data, however in order to apply this approach we need to somehow eliminate the zeros first. To do this we apply the multiplicative zero replacement strategy suggested by Martín-Fernáadez et. al. (2003), page 262 and

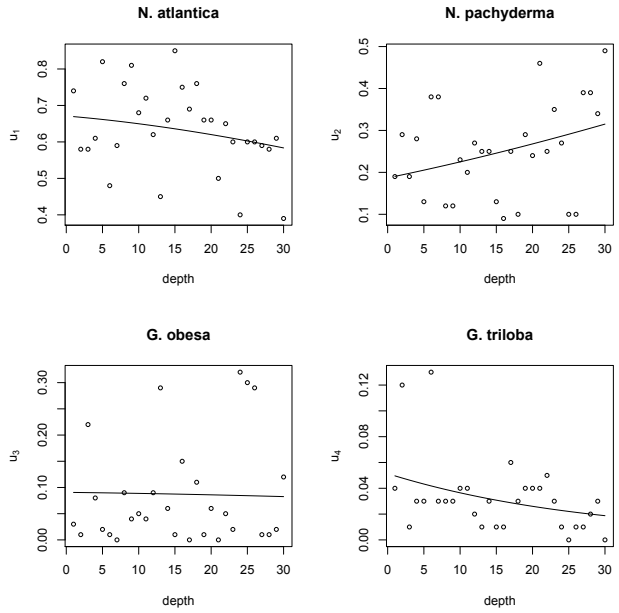

Figure 4: Plots of u_1, u_2, u_3 and u_4 versus depth.

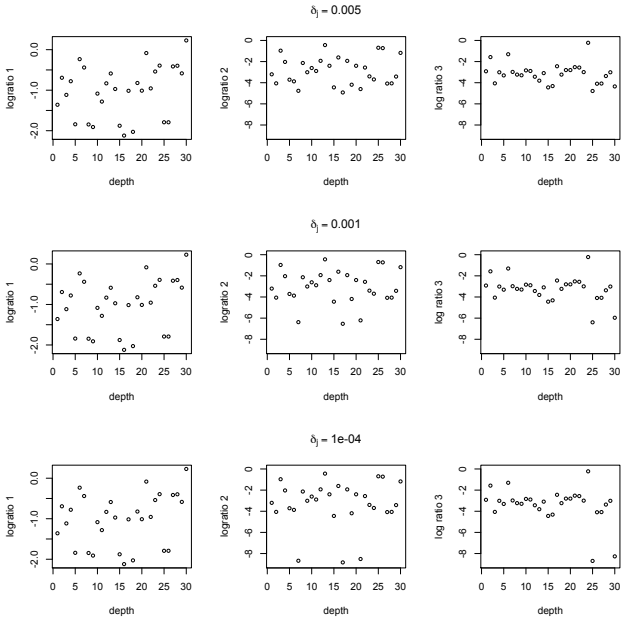

Figure 5: Plots of $\breve{y}_1$, $\breve{y}_2$, $\breve{y}_3$ versus depth.

hence we replace $\boldsymbol{u}$ by $\boldsymbol{r} = (r_1, r_2, ..., r_p)$ using

$$r_j = \begin{cases} \delta_j, & \text{if } u_j = 0, \\ (1 - \sum_{k|u_k=0} \delta_k) u_j, & \text{if } u_j > 0. \end{cases}$$

If we assume that the zero's are a result of rounding error and are not essential zeros, then each δ_j is in the range $(0, 0.005)$.

Figure 5 contains plots of the additive logratios versus depth for three different choices of δ_j (i.e. $\delta_j = 0.005$, $\delta_j = 0.001$ and $\delta_j = 0.0001$). If δ_j is in the range $[0.001, 0.005]$ and if we correct observation 24, then the following linear model is plausible ($\breve{y}_3$ appears to roughly decrease linearly with depth)

$$\breve{y}_3 = b_{31} + b_{32} x + \epsilon_3, \tag{10}$$

where $\epsilon_3 \sim N(0, \sigma^2)$ and b_{31} and b_{32} are scalar regression coefficients. However, the assumption that each $\delta_j \geq 0.001$ is very strong and it is still plausible that the correct values of δ_j are in the range $[0, 0.001]$. That is, some of the five zero components may represent a complete absence or near absence of the species at the depth measured. If this is the case, then model (10) should not be fitted. The advantage of the Kent regression model approach over the logratio approach is that it does not lead to introducing additional uncertainties by having to impute δ_j.

5 Conclusion

As we have shown, the square root transformation approach has some nice properties and will work better than the logratio approach in certain situations such as when the compositional data is distributed close to the boundaries of the simplex. As we demonstrated, the square root transformation model fit the example dataset well and by comparison, the logratio approach was much more problematic to implement due to the presence of zeros.

Future research will involve exploring the folding issue further and to extend the Kent regression model to other scenarios such as the case where we have correlated compositional data or where κ, $\boldsymbol{\beta}$ and $\boldsymbol{K}^*$ also depend on $\boldsymbol{x}$.

References

Aitchison, J. (1982). The Statistical Analysis of Compositional Data (with discussion). *Journal of the Royal Statistical Society Series B 44*, 139–177.

Aitchison, J. (1986). *The Statistical Analysis of Compositional Data.* Monographs on Statistics and Applied Probability. Chapman & Hall Ltd., London (UK).

Aitchison, J. and Kay, J. W. (2004). Possible solutions of some essential zero problems in compositional data analysis. *In: Thió-Henestrosa S, Martín-Fernández JA (eds). Compositional data analysis workshop, Girona, Spain. http://ima.udg.es/Activitats/CoDaWork03/.*

Bacon-Shone, J. (1992). Ranking methods for compositional data. *Applied Statistics 41:3*, 533-537.

Egozcue, J. J. (2009). Reply to "On the Harker Variation Diagrams;..." by J. A. Cortés. *Math Geosci 41*, 829–834.

Fry, J. M., Fry, T. R. L. and McLaren, K. R. (2000). Compositional data analysis and zeros in micro data. *Applied Economics. 32:8*, 953-959.

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

Mardia, K. V. and Jupp, P. E. (2000). *Directional Statistics.* John Wiley & Sons, LTD.

Martín-Fernández, J. A., Barceló-Vidal, C. and Pawlowsky-Glahn, V. (2003). Dealing with zeros and missing values in compositional data sets using nonparametric imputation. *Mathematical Geology 35:3*, 253–278.

Palarea-Albaladejo J., Martín-Fernández, J. A. and Gómez-García, J. (2007) A parametric approach for dealing with compositional rounded zeros. *Mathematical Geology 39*, 625–645.

Scealy, J. L., and Welsh, A. H. (2011). Regression for compositional data by using distributions defined on the hypersphere. *Journal of the Royal Statistical Society Series B to appear.*

Stephens, M. A. (1982). Use of the von Mises Distribution to Analyse Continuous Proportions. *Biometrika 69*, 197-203.

Wang, H., Liu, Q., Mok, H. M. K., Fu, L., and Tse, W. M. (2007). A hyperspherical transformation forecasting model for compositional data. *European Journal of Operational Research 179*, 459-468.