Zero adjusted Dirichlet regression for compositional data with zero values present

Michail Tsagris
College of engineering and technology, American university of the middle east, Egaila, Kuwait
mtsagris@yahoo.gr

Abstract

Compositional data are met in many different fields, such as economics, archaeometry, ecology, geology and political sciences. The feature that makes them special is that each vector contains positive number which sum to the same constant, usually taken to be 1 for convenience purposes. Regression where the dependent variable is a composition is usually carried out via log-ratio transformation of the composition or via the Dirichlet distribution. However, when there are zero values in the data these two ways are not readily applicable. In this paper we adjust the Dirichlet distribution in order to allow for zero values to be present in the data. To do so, we take into advantage the conditionality (or the marginality) property of the Dirichlet distribution. In addition we check the effect of the zero values on the resulting model. Finally, we propose a new information criterion to be used for model selection and for assessing the goodness of fit.

Keywords: Compositional data, regression, Dirichlet distribution, zero values

1 Introduction

Compositional data are positive multivariate data that sum to the same constant and their sample space is the standard simplex

$$S^d = \left\{ (x_1, ..., x_D)^T \left| x_i \geq 0, \sum_{i=1}^{D} x_i = 1 \right. \right\},$$

where $D$ denotes the number of variables (better known as components) and $d = D - 1$.

A natural candidate distribution for analyzing compositional data is the Dirichlet, since its support is the simplex. Dirichlet regression analysis has been considered in (Gueorguieva et al., 2008) and Hijazi and Jernigan (2009). A more popular type of regression was suggested by Aitchison (2003). By applying a log-ratio transformation, one can map the data for the simplex onto the Euclidean space and there apply the standard multivariate regression techniques. Back transformation of the fitted values will result in the fitted compositional values.

The problem however with Aitchison’s regression and Dirichlet regression is that they do not address the problem of zero values. If zero values are present in the data, the logarithm of the data is not applicable and the Dirichlet cannot be readily applied. For this reason, model based zero value imputation techniques have been developed (Palarea-Albaladejo and Martín-Fernández, 2008). These methods must be applied prior to the regression analysis and assume that zeros are due to roundings or measurement errors.

Other regression techniques on the simplex addressing the problem of zero values can be found in Scealy and Welsh (2011) who performed regression by employing the Kent distribution. Their
approach falls within the category of treating compositional data as directional data through the square root transformation (see also [Stephens (1982)]. Leininger et al. (2013) implemented spatial regression for compositional data with many zeros employing a scaling power transformation and assuming a latent multivariate normal model.

In this paper we take into account the nice properties of the Dirichlet distribution for regression analysis with zero values present. We adjust the Dirichlet log-likelihood to accommodate zero values using both parametrizations of the Dirichlet distribution and compare them as well. We suggest a diagnostic test to see the effect of the zero values on the estimates. In addition, we suggest a information criterion in order to perform model selection regardless of the assumed (parametric or not) regression model.

Aitchison (2003, pg. 119) provided formulas for the relationship between a logistic normal in the full composition and the logistic normal for a composition with less components. However, in our case we have conditioned on some components having zero values. The conditional logistic normal distribution becomes singular as the value of a component approaches zero. That is why we cannot use the logistic normal distribution in the same way to model compositional data with zero values. The conditional Dirichlet distribution on the other hand is not singular and coincides with the marginal distribution in this case.

Section 2 describes the Dirichlet regression with both parametrizations of the Dirichlet distribution. The problem of zeros is briefly discussed in Section 3. In the same Section the adjustment of the log-likelihood based on the two formulations of the Dirichlet distribution is described. The effect of zeros is also discussed and some diagnostics are suggested. In addition, a new information criterion is suggested for model selection regardless of the assumed distribution or modelling technique used for compositional data regression. In Section 4 we compare the two zero adjusted Dirichlet parametrizations using real data examples and finally, conclusions in Section 5 close this paper.

2 Zero values and Dirichlet regression

There are two types of Dirichlet regression, since there are two parametrizations of the Dirichlet distribution.

2.1 Dirichlet regression type I

Dirichlet regression is performed in the form of generalized linear models. Hijazi and Jernigan (2009) used Dirichlet regression to model compositional data linking the parameters linearly with the covariates. However, the log link Gueorguieva et al. (2008) is preferred to the identity for this distribution, because it ensures that the estimated parameters are always positive. We can write the log of the parameters of the Dirichlet distribution ($\mathbf{a}_i$) as a generalised linear function of the design matrix $\mathbf{X}$ which contains the covariates (the first column is the vector of 1s).

$$\log \mathbf{a}_i = \mathbf{x}_i^T \mathbf{b}_i \Rightarrow \mathbf{a}_i = e^{\mathbf{x}_i^T \mathbf{b}_i},$$

where $p$ is the number of independent variables, $n$ denotes the sample size and

$$\mathbf{b}_i = (b_{0i}, b_{1i}, ..., b_{pi})^T, \quad i = 1, ..., D.$$  

(1)
Gueorguieva et al. (2008) developed and discussed diagnostics for over-dispersion and global and local influence. They also discussed different types of residuals, such as standardized residuals, composite residuals and score residuals which are related to over-dispersion diagnostics. The likelihood of the Dirichlet with covariates is (Gueorguieva et al., 2008)

\[
L(b) = \prod_{j=1}^{n} \frac{\Gamma \left( \sum_{i=1}^{D} a_{ij} \right)}{\prod_{i=1}^{D} \Gamma \left( a_{ij} \right)} \prod_{i=1}^{D} y_{ij}^{a_{ij} - 1} \text{ and } a_{ij} = e^{x_j^T b_i},
\]

where \(x_j\) are covariates for the \(j\)-th observational vector and \(b_i\) is the vector of parameters for each component estimated through maximisation of the log-likelihood

\[
\ell = \sum_{j=1}^{n} \log \Gamma \left( \sum_{i=1}^{D} e^{x_j^T b_i} \right) - \sum_{j=1}^{n} \sum_{i=1}^{D} \log \Gamma \left( e^{x_j^T b_i} \right) + \sum_{j=1}^{n} \sum_{i=1}^{D} \left( e^{x_j^T b_i} - 1 \right) \log y_{ij}. \tag{2}
\]

We also have to note that Yee (2011) offers maximisation of (2) in his R-package VGAM.

### 2.2 Dirichlet regression type II

We can also use another parametrization of the Dirichlet distribution which includes the parameter \(\phi = \sum_{i=1}^{D} a_i\).

\[
f(x) = \frac{\Gamma \left( \sum_{i=1}^{D} \phi a_i^* \right)}{\prod_{i=1}^{D} \Gamma \left( \phi a_i^* \right)} \prod_{i=1}^{D} x_i^{\phi a_i^* - 1},
\]

where \(\sum_{i=1}^{D} a_i^* = 1\). The link function used for the parameters (except for \(\phi\)) is

\[
a_i^* = \frac{1}{\sum_{j=1}^{D} e^{x_j^T \beta_i}} \quad a_i^* = \frac{e^{x_j^T \beta_i}}{\sum_{j=1}^{D} e^{x_j^T \beta_i}} \quad \text{for } i = 2, ..., D, \tag{3}
\]

where

\[
\beta_i = (\beta_{0i}, \beta_{1i}, ..., \beta_{pi})^T, \quad i = 1, ..., d. \tag{4}
\]

The corresponding log-likelihood is

\[
\ell = n \log \Gamma(\phi) - \sum_{j=1}^{n} \sum_{i=1}^{D} \log \Gamma(\phi a_i^*) + \sum_{j=1}^{n} \sum_{i=1}^{D} (\phi a_i^* - 1) \log y_{ij}. \tag{5}
\]

The interpretation of the resulting regression parameters is easier than (2) as will be seen later and are comparable with those of the standard linear regression model suggested by Aitchison (2003) since the same link function is used. This is true because they same the number of parameters whose interpretation is the same. In addition, the second formulation requires \(d(p + 1) + 1\) parameters, whereas the first formulation requires \(D(p + 1)\) parameters, where \(p\) denotes the number of independent variables and \(d = D - 1\). Maier (2014) has written an R package which performs maximisation of (5) allowing also the \(\phi\) parameter to vary as a function of the covariates as well.
3 Dirichlet regression with zero values present

In this Section we will show one can perform Dirichlet regression when zero values are present in the data. But at first we will briefly comment on the problem of zero values.

3.1 Zero values and current approaches

When Aitchison (1982) suggested the log-ratio analysis of compositional data he noted that zero values are not allowed. For this reason, he suggested ad-hoc zero replacement strategies (Aitchison, 2003). Some of them have nice properties and are preferred. More recently, Palarea-Albaladejo and Martín-Fernández (2008) developed a model-based imputation of zero values and Martin et al. (2012) suggested a robust version of it. The underlying assumption behind all these strategies is that the zero values are actually unobserved very small quantities which were rounded to zero. An example is when the detection limit of a measurement instrument is not low enough.

The drawback however of these techniques is that the values of the whole compositional vector which has at least one zero value will have to change even slightly. So, when there are many zeros in a dataset many other observed values will have to change as well. From one point of view this could be necessary, since the observed proportions are not the true ones. Imputation though of zero values induces some extra variability to the data.

On the other hand, can somebody, with probability 1, say that in a geological, geochemical, or any other earth related sample of some data where zeros occur that these zero values are rounded? Isn’t it possible that in some rock samples for example, the percentage of gold is actually zero? The famous Glass Identification Data Set (available to download from UC Irvine Machine Learning Repository) is another example. Not all glass types necessarily contain the same materials. Why then one should try to impute the zero values, assuming that a chemical substance exists in a glass type when in fact it simply does not?

Butler and Glasbey (2008) and Leininger et al. (2013) introduce latent variables to model the zero values. The observed zero values are modelled without being imputed and no further change in the values of the other components is necessary.

A third approach adopted by Zadora et al. (2010), Scealy and Welsh (2011) and Stewart and Field (2011) is to handle the zero values naturally. Zadora et al. (2010) models the probability of a zero value separately and Stewart and Field (2011) moves in a similar spirit. The square root transformation Scealy and Welsh (2011) maps the zero values on the surface of a hyper-sphere, thus they are treated as allowable points on the hyper-sphere. In the last three models, no zero value imputation is performed either.

Zadora et al. (2010) mentioned that if the zero values are actually non zero values rounded to zero this should be be taken into account building the model. At some other point they mention "Of course, absolute zero does not exist in analytical chemistry as measurements depend on the detection level of the method used. Elements and components present in trace levels (below the detection limit) will always be undetectable for the specific method, and hence the measurements obtained can be considered as structural zeros for the purposes of the statistical analysis". So they do not argue that zero values are rounded zeros but for the purpose of the analysis they choose not to impute them.

Our approach falls more within the directions of Zadora et al. (2010) and (Scealy and Welsh, 2011). We do not make any assumptions on whether the zeros are rounded or not, we choose to
model the data as they are.

3.2 Zeros values and Dirichlet regression type I

An advantage of the Dirichlet distribution is that its marginals are also Dirichlet and this allows modelling in the presence of zeros (Ng et al., 2011). We will take advantage of this property in order to incorporate zero values in the observed sample. The same is true also for the conditional distributions. That is, if we condition on some components being zero for example, the resulting distribution is a Dirichlet. In this context, there seems to be no separation between the two. Hence, we suggest to use the zero adjusted log-likelihood of the Dirichlet regression

\[ \ell_0 = \sum_{j=1}^{n} \log \Gamma \left( \sum_{i \in C_j} e^{x_i^T b_i} \right) - \sum_{j=1}^{n} \sum_{i \in C_j} \log e^{x_i^T b_i} + \sum_{j=1}^{n} \sum_{i \in C_j} \left( e^{x_i^T b_i} - 1 \right) \log y_{ij}, \]  

(6)

where \( C_j = \{ k \in [1, ..., D] : y_{kj} > 0 \} \). The hypothesis we state here is a simple and a not unrealistic one. The vectors which have zero values come from the same Dirichlet distribution as that of the vectors with no zero values. The same hypothesis is imposed when the second formulation (5) is used (see below). We will call this type of Dirichlet regression zero adjusted Dirichlet regression (ZADR) type I.

3.3 Zeros values and Dirichlet regression type II

The modification of the second type of the Dirichlet log-likelihood (5) to accommodate zero values in the observed data is also feasible and leads to the zero adjusted Dirichlet log-likelihood

\[ \ell_0 = n \log \Gamma (\phi) - \sum_{j=1}^{n} \sum_{i \in C_j} \log \Gamma (\phi a_i^*) + \sum_{j=1}^{n} \sum_{i \in C_j} (\phi a_i^* - 1) \log y_{ij}, \]  

(7)

where \( C_j = \{ k \in [1, ..., D] : y_{kj} > 0 \} \) and \( a_i^* \) is defined in (3).

The advantage of (7) over (6) is that it requires less parameters and their interpretation is easier, since the first one uses the inverse of the additive log-ratio transformation as the link function. We will refer to this type of Dirichlet regression as zero adjusted Dirichlet regression (ZADR) type II. In the next Section we describe a way of of maximising the two log-likelihoods (6) and (7).

3.4 Maximising the log-likelihoods

When a maximisation is performed numerically, there is the question of sensitivity to the initial values. We have seen empirically that this is not an issue. At first we must use the zero free compositional vectors to get estimates which will be termed initial estimated coefficients. We then add the contribution to the log-likelihood of the compositional vectors having zero values in at least one of their components. Thus, the log-likelihood in (6) and (7) is actually the sum of two log-likelihoods, the one due to zero free compositional vectors and the other due to the non-zero free compositional vectors. When the ZADR type II is used we performed the following process to maximise the log-likelihood.

1. Use the zero free compositional vectors and apply the additive log-ratio transformation
\[ z_i = \log \frac{y_i}{y_1}, \text{ where } i = 2, \ldots, D \]

2. Estimate the parameters using least squares
\[ B = (X^T X)^{-1} X^T Z, \]
where \( Z \) is the \( n \times D \) matrix containing the log-ratio transformed composition and \( X \) is the \( n \times p \) design matrix.

3. Use the estimates from step 2 as initial values for maximising the log-likelihood of ZADR type II \((5)\). The estimates are the initial estimated coefficients.

4. Use the initial estimated coefficients as initial values in maximising \((7)\) and obtain the final estimated coefficients.

When the ZADR type I regression is used, we skip step 1 and for step 2 we use random values as initial values. Steps 3 and 4 are the same using \((2)\) and \((7)\) respectively.

When the above described procedure is implemented we have two advantages. At first we reduce the probability of obtaining a local instead of a global maximum of the relevant log-likelihood. Secondly, in this way we have two sets of estimated coefficients, the one based on the zero free composition and the one based on the full composition. This enables us to examine the effect of the zeros, described in the next Section.

### 3.5 Effect of zeros

Can we assume that the conditional (or marginal) distribution of the components which have zero values comes from the same Dirichlet distribution as the composition with no zeros?

#### 3.5.1 Quadratic type diagnostics

To answer these questions we propose the use of a quadratic form diagnostic involving the coefficients and their relevant covariance matrix, of the zero free composition and of the final coefficients. The diagnostic has the following two forms

\[ T_1 = \text{vec} \left( \Delta \hat{B}_I^T \right) \Sigma_1^{-1} \text{vec} \left( \Delta \hat{B}_I \right), \]

\[ T_2 = \left[ \theta, \text{vec} \left( \Delta \hat{B}_{II}^T \right) \right] \Sigma_2^{-1} \left[ \theta, \text{vec} \left( \Delta \hat{B}_{II} \right) \right], \]

where

\[ \text{vec} \left( \Delta \hat{B}_I \right) = \text{vec} \left( \hat{B}_I^{ini} - \hat{B}_I \right) = \left( \hat{b}_1^{ini} - \hat{b}_1, \ldots, \hat{b}_D^{ini} - \hat{b}_D \right), \]

\[ \Sigma_1 = \tilde{V}^{ini} + \tilde{V} \text{ and } \]

\[ T_2 = \left[ \theta, \text{vec} \left( \Delta \hat{B}_{II}^T \right) \right] \Sigma_2^{-1} \left[ \theta, \text{vec} \left( \Delta \hat{B}_{II} \right) \right], \]

6
where

\[
\text{vec} \left( \Delta \hat{B}_{II} \right) = \text{vec} \left( \hat{B}_{II}^{ini} - \hat{B}_{II} \right) = \left( \hat{\beta}^{ini}_1 - \hat{\beta}_1, \ldots, \hat{\beta}^{ini}_d - \hat{\beta}_d \right),
\]

\[\Sigma_2 = \hat{U}^{ini} + \hat{U} \text{ and } \theta = \phi^{ini} - \hat{\phi}\]

for the ZADR type I and type II respectively. The \(\hat{b}^{ini}_i\) and \(\hat{b}_i\) are the coefficients of the zero free composition and the final coefficients of type I regression respectively and are defined in (1). The \(\hat{\beta}^{ini}_i\) and \(\hat{\beta}_i\) are the estimated coefficients of the zero free composition and the final estimated coefficients of type II regression respectively and are defined in (4). In both cases, the covariance matrices of the estimated coefficients of the ZADR type I and type II models are denoted by the letters \(V\) and \(U\) respectively. These are estimated numerically by the second derivative of the log-likelihood of (6) and (6) respectively.

### 3.5.2 A distance based diagnostic

The drawback of \(T_1\) (8) and \(T_2\) (9) is that when there are a few zeros they might not be able to detect any deviation from the Dirichlet assumption. For this reason we also suggest another type of diagnostic which is distance based.

At first we estimate the parameters of the Dirichlet regression for the zero free composition (step 3 of the algorithm for maximising the ZADR models). Then we use them to predict the values of the compositional vectors having zero elements. This means, that we will adjust our fitted compositional vectors by setting zeros in the same elements and re-scale them so that they sum to 1. Then, we will calculate their between distance. To do so we will use a recently suggested metric for probability distributions [Endres and Schindelin (2003) and Österreicher F. and Vajda I. (2003)]

\[
\Delta_{esov}(u, w) = \sqrt{KL(u, M) + KL(w, M)},
\]

where \(M = 0.5 (u + w)\) and \(KL(u, w) = \sum_{i=1}^{k} u_i \log \frac{u_i}{w_i}\) is the Kullback-Leibler divergence [Kullback, 1997]. We will refer to (10) as the ES-OV metric. Summing the ES-OV metric (separately) for all compositional vectors which have at least one zero value, we get the two values of this type of diagnostic

\[
T_{esov} = \sum_{i=1}^{m} \Delta_{type}(y_{0i}, \hat{y}_{0i}),
\]

The \(y_{0i}\) is the \(i\)-th observed compositional vector with zero elements and \(\hat{y}_{0i}\) is its corresponding fitted vector based on the Dirichlet regression \(2\) or \(5\). So we are not using the ZADR models in this one, but the classical Dirichlet regression models. In this way we want to see how well the regression models based on the zero free composition, predicts the compositional vectors containing zeros.

### 3.5.3 Bootstrap p-value of the diagnostics

The distribution of the diagnostics (8), (9) and (11) is rather difficult to derive, so for this reason we propose the use of parametric bootstrap whose steps are as follows
1. Generate data from a Dirichlet distribution using the final estimated parameters. This means, that some vectors will have no zero values and some vectors will.

2. Estimate the parameters of the ZADR models and calculate the diagnostic values \( \mathbb{S} \), \( \mathbb{U} \) and \( \mathbb{M} \) based on the bootstrap sample.

3. Repeat steps 1 − 2 \( B \) times and calculate a bootstrap p-value as

\[
p-value = \frac{\sum_{b=1}^{B} \mathbb{1}\{ T \geq T_{observed} \} }{B + 1},
\]

where \( T_{observed} \) is any of the diagnostic tests based on the observed data, \( b = 1, \ldots, B \) and \( \mathbb{1} \) is the indicator function.

### 3.6 Some comments

Dirichlet regression type I estimates \( D(p + 1) \) parameters, where \( p \) is the number of independent variables, whereas Dirichlet regression type II estimates \( d(p + 1) + 1 \) parameters, where \( d = D - 1 \). Thus \( p \) less parameters. In addition, the parameters of the Dirichlet regression type II are easier to interpret. For this reason we suggest the use of the Dirichlet regression type II with or without zeros.

Dirichlet distribution offers a natural way of handling zeros and allows us to perform regression without having to substitute the zero values with small quantities. The goal is to treat the zero values without changing their values and to some extent the values of the other components in the same compositional vector.

So, when we have a sample for which we assume that it comes from a Dirichlet population and some components have zero values we need to make an extra assumption about the vectors which contain zero values. We assume that these vectors come from the same Dirichlet distribution. This might be viewed as a strong parametric assumption, but in the EM algorithm also assumes that the conditional distribution of the missing data given the observed data is of a known form.

Another point that is worthy to highlight it that as appealing the Dirichlet regression might seem we have to note that the Dirichlet distribution is not a member of the linear exponential family (Gourieroux et al., 1984) and hence any regression model based on it is not robust to distribution misspecification (Murteira and Ramalho, 2013). However, under correct specification, the asymptotic distribution of the estimated parameters is a multivariate normal distribution whose covariance matrix is the inverse of the Fisher’s information matrix (Murteira and Ramalho, 2013). This means that in this case only the estimates are consistent.

The same holds true in our case, since we maximise the Dirichlet log-likelihood in order to obtain the regression parameters, but we need an extra assumption to ensure the property of consistency. That is, the compositional vectors which contain zero values, and are thus modelled by a conditional Dirichlet with less dimensions, come from a Dirichlet with the same parameters as the full Dirichlet. The diagnostics suggested try to evaluate the assumption of this assumption.

A known drawback of the Dirichlet distribution and thus of regression is that the covariance matrix allows only negative correlations. This restrictive correlation is also met in the multinomial logistic regression but even so, the model works fine in practice. A more serious drawback is that when no covariates are present the Dirichlet distribution cannot capture curvature in the simplex (Aitchison, 2003, pg. 59), but we are not in this situation.
The inflated beta distribution (Ospina and Ferrari, 2010) is not related to the conditional Dirichlet distribution we describe here. In the beta distribution (with or without covariates) zero values are allowed and the same is true in the logistic regression. But when we move to higher dimensions, the Dirichlet distribution does not allow a vector of zeros or a vector of 1s (again this is also true for the multinomial logistic regression). Therefore, a generalization of the inflated beta to Dirichlet distribution is not possible.

3.7 An information criterion for compositional regression

As mentioned before, type II regression estimates fewer parameters than the type I regression. So, there are two questions raised, which of the two formulations should one use and which regression model in general should one use for compositional data? A suggestion to the first answer was given before. However, both questions are to be answered via a new information criterion.

The Kullback-Leibler divergence (Kullback, 1997) is a divergence measure of one distribution from another. We exploit this divergence and use it as a measure for compositional data. The divergence suggested here can be used not only with univariate proportions but with compositional data in general, regardless of the model or distribution used. In logistic regression it is called deviance, but since here we do not deal with categorical data or frequencies we cannot calibrate it against a \( \chi^2 \) distribution. However we can call it compositional deviance, since it shows the divergence between the observed and expected compositions. Theil (1967) defined a measure of information inaccuracy as

\[
TII = \sum_{j=1}^{n} q_j \log \frac{q_j}{\hat{q}_j},
\]

where \( q_j \) and \( \hat{q}_j \) denote the observed and expected points of the Lorenz curve. On the grounds of this we can define information criteria based on this divergence, such as the compositional deviance information criterion (CDIC)

\[
CDIC = 2 \sum_{j=1}^{n} \sum_{i=1}^{D} y_{ij} \log \frac{y_{ij}}{\hat{y}_{ij}} + k \log n,
\]

where \( k \) and \( n \) denote the number of parameters and sample size respectively and \( y_{ij} \) and \( \hat{y}_{ij} \) are the observed and fitted compositional values. The word Deviance was chosen because the deviance in the multinomial regression is two times the measure (12) suggested by Theil (1967). The CDIC can be used even if there are zero values since \( 0 \log 0 = 0 \). Theil’s information does not take into account the number of parameters a model contains. The penalty for the number of parameters and the sample size was inspired from the Bayesian Information criterion.

If the two models come from the same family of distributions, such as in our example we can also use the already known information criteria, such as AIC or BIC.

4 Examples with real data

We will show two examples with real data in order to see the two types of zero adjusted Dirichlet regression types in practice and compare them.
4.1 Example 1. Foraminiferal data

This dataset consists of foraminiferal (marine plankton species) compositions measured at 30 different depths (1-30 metres) and can be found in Aitchison (2003). There are 5 compositional vectors having a zero value, either in the third or the fourth component. The data were also analysed by Scealy and Welsh (2011) using the logarithm of the depth as the independent variable. Since there are 4 components we cannot plot them on a triangle. We could plot them in a 3D plot but when printed it would not reveal the structure of the data. For this reason, we will plot them in a bar plot (Larrosa, 2003). Figure 2 presents the data as function of the water depth. It can be seen that the logarithm of the water depth does not really affect the composition. Some columns have three colours and that is because of the zero values.

Table 1: Parameter estimates for the two ZADR models applied to the foraminiferal data. The Bias columns refer to the estimated bias of the parameters calculated using parametric bootstrap. The standard errors appear in the parentheses next to the estimates.

| Model   | Response variable | Constant | Bias       | Slope      | Bias       |
|---------|-------------------|----------|------------|------------|------------|
| ZADR    | Triloba           | 3.353(0.499) | -0.363(0.633) | -0.404(0.192) | 0.106(0.240) |
| type I  | Obesa             | 2.162(0.536) | -0.345(0.637) | -0.305(0.209) | 0.100(0.241) |
|         | Pachyderma        | 0.615(0.503) | -0.307(0.640) | -0.186(0.192) | 0.093(0.242) |
|         | Atlantica         | 0.791(0.551) | -0.274(0.653) | -0.346(0.216) | 0.077(0.254) |
| ZADR    | Obesa             | -1.225(0.348) | -0.001(0.331) | 0.117(0.131) | 0.001(0.125) |
| type II | Pachyderma        | -2.392(0.463) | -0.004(0.505) | 0.087(0.173) | 0.008(0.186) |
|         | Atlantica         | -2.298(0.494) | -0.021(0.485) | -0.046(0.193) | 0.012(0.190) |
| $\phi$  |                   | 15.889(2.473) | -1.76(2.963)  |             |             |

We have chosen the first component (by default) to be the reference component for all the other components in the ZADR type II. Thus, the interpretation of the parameters (see Table 1) of the components is with respect to the first component. The CDIC value for the ZADR type I and II were 30.338 and 26.934 respectively. The difference is small (and roughly equal to $\log 39$) and thus shows that the fit of the two models is almost the same. The difference is that ZADR II needs one (in this case) parameter less and the interpretation of the coefficients is clearer.

Table 2 present the diagnostic values along with their corresponding bootstrap calculated p-values.

This is an example where the zeros do not seem to have a significant effect on the fit of the model. This is more a statistical result we want hold true, since in practice one can still use this model even if this assumption is not true. The caution level though would be increased in the later case.

Table 2: Diagnostic values for the ZADR models applied to the foraminiferal data. The bootstrap calibrated p-values appear inside the parentheses.

| Model      | $T_1$ or $T_2$ | $T_{esow}$  |
|------------|----------------|-------------|
| ZADR type I| 0.655(0.172)  | 1.014(0.203) |
| ZADR type II | 0.850(0.110) | 1.026(0.109) |
Figure 1: Foraminiferal compositions as a function of the water depth. As we move from left to right, we see that the logarithm of the water depth increases.

Figure 2: Fitted foraminiferal compositions as a function of water depth using (a) the ZADR type I (6) and (b) the ZADR type II model (7).

4.2 Example 2. Glacial data

The second example contains information about the weight of 92 glacial tills (in percentage form) along with the total number of pebbles. The data are available at Aitchison (2003). For each of the glacial tills there are four different components, red sandstone, gray sandstone, crystalline and some miscellaneous elements. There are 42 compositional vectors containing either one or two zeros, making a total of 48 zero values. Table 3 contains the estimated regression coefficients and Figure 3 shows this compositional dataset versus the total number of pebbles.
Table 3: Parameter estimates for the two ZADR types applied to the glacial tills data. The *Bias* columns refer to the estimated bias of the parameters calculated using parametric bootstrap. The standard errors appear in the parentheses next to the estimates.

| Model  | Response variable | Constant | Bias       | Slope       | Bias       |
|--------|-------------------|----------|------------|-------------|------------|
| ZADR   | Redsandstone      | 4.967(0.870) | -0.450(1.011) | -0.727(0.146) | 0.070(0.171) |
| type I | Graysandstone     | 4.218(1.105) | -0.432(0.996) | -0.679(0.188) | 0.068(0.167) |
|        | Crystalline       | 2.136(1.046) | -0.375(0.970) | -0.549(0.176) | 0.059(0.163) |
|        | Misc              | 2.976(1.257) | -0.372(1.178) | -0.661(0.212) | 0.058(0.196) |
| ZADR   | Graysandstone     | -2.266(1.014) | -0.029(0.900) | 0.314(0.169)  | 0.006(0.150) |
| type II| Crystalline       | -1.518(1.057) | -0.061(1.037) | -0.037(0.178) | 0.012(0.174) |
|        | Misc              | -1.118(1.305) | -0.061(1.336) | -0.078(0.219) | 0.012(0.224) |
|         | $\phi$            | 3.419(0.315)  | -0.134(0.344) |

Table 4: Diagnostic values for the ZADR models applied to the glacial tills data. The bootstrap calibrated p-values appear inside the parentheses.

| Model    | $T_1$ or $T_2$ | $T_{ESOV}$ |
|----------|----------------|------------|
| ZADR type I | 6.687(0.027)     | 16.049(0.000) |
| ZADR type II | 4.458(0.013)     | 16.398(0.000) |

Figure 3: Percentages by weight of glacial tills for different total number of pebbles expressed in a logarithmic scale.
Figure 4 contains the fitted compositions of glacial tills using both types of ZADR. The CDIC value for the ZADR type I is equal to 69.671, whereas for type II it is equal to 64.949. The difference is again roughly equal to the logarithm of the sample size, showing that the fit of the two models is almost the same. But again here, the ZADR II model is to be preferred since it needs one parameter less.

This is an example where the zeros seem to have a significant effect on the fit of the model. This means, that the Dirichlet assumption for all the data is not satisfied. However, this does not exclude the model from being used by researchers and practitioners. As mentioned before, the caution level is increased, but then a simple question comes to mind. Is the normality assumption always satisfied when one applies techniques which assume normality?

A possible answer as to why this happens in this example could be found in examining the three plots in Figures (3) and (4). We can see that the two models have not fitted the data adequately. So, maybe, the assumption of the Dirichlet distribution might not have been a very good choice or the link function should be different.

### 4.3 A small scale simulation study

We implemented a small scale simulation study to see the asymptotic behaviour of the coefficients of the two ZADR models when zeros are present. We generated data assuming the two models of example 1. In that example there are 30 observations 5 of which have one component with a zero value. Data of sample sizes, (multiples of 30, so $n = 60, 120, 240, 360, 480, 600$) were generated 1000 times and each time we applied the two ZADR models. In all cases there was 1/6 of compositional observations with a zero value in one of its components. The mean squared error (MSE) of each coefficient for each of the two models was calculated and is presented in Tables 5 and 6. There is one covariate only, and thus ZADR model I has 8 coefficients (4 for the constants and 4 for the slopes), whereas ZADR model II has 7 (3 for the constants, 3 for the slopes and one for the $\phi$
Table 5: Mean squared error (MSE) (based on 1000 simulations) for the estimated coefficients of the ZADR I model assuming the coefficients of Example 1 (see Table 1) are the true coefficients.

|          | n=60 | n=120 | n=240 | n=360 | n=480 | n=600 |
|----------|------|-------|-------|-------|-------|-------|
| Constant | 0.177| 0.080 | 0.037 | 0.021 | 0.016 | 0.013 |
| Slope    | 0.024| 0.011 | 0.005 | 0.003 | 0.002 | 0.002 |
|          | 0.176| 0.082 | 0.037 | 0.021 | 0.016 | 0.014 |
| Constant | 0.024| 0.012 | 0.006 | 0.003 | 0.003 | 0.002 |
| Slope    | 0.027| 0.012 | 0.042 | 0.018 | 0.018 | 0.014 |
|          | 0.189| 0.085 | 0.042 | 0.024 | 0.018 | 0.014 |
| Constant | 0.024| 0.012 | 0.006 | 0.004 | 0.003 | 0.002 |
| Slope    | 0.027| 0.012 | 0.006 | 0.018 | 0.018 | 0.014 |

5 Conclusions

In this paper we suggested the zero adjusted Dirichlet regression (ZADR) for modelling compositional data with covariates when zero values are present. The importance of this simple approach is important, since no modification of the data is necessary, such as zero value replacement (or imputation). This means that no extra variance is introduced and most importantly the observed data are not at all changed or distorted to the slightest.

We also suggested diagnostics to assess the effect of the zero values on the assumed model. The assumption that the conditional distribution of the vectors with zero values is the same as that of the zero free vectors needs not always be true. But since we are dealing with parametric models this assumption is sometimes violated, nevertheless the model is kept and used for inference. There is a question however of whether this assumption checking is a good way and if yes, how good it is.

The suggested quadratic form diagnostic, in its two versions $T_1$ and $T_2$, is not capable of detecting every violation of the Dirichlet assumption. It is however intended to be used as a diagnostic test. The distance based diagnostics on the other hand, seem more capable of detecting small deviations as they are based on each compositional vector with zero elements. Simulation studies need to be performed in order to see the abilities of this diagnostic under difficult scenarios.

Finally we provided an information criterion to choose amongst different regression models for compositional data regardless of the fitted distribution. This is a very useful criterion which does not depend upon any parametric assumptions made about the model. The choice of this criterion is reasonable but the performance, or the discriminative power of it needs to be checked. That is, can we identify if a model performs best when we know the true model? Simulation studies will need to be performed to answer this question.

The question though remains, should we treat zeros as rounded and try to impute their missing value or should we treat them naturally? We believe the second option is preferable and one reason
Table 6: Mean squared error (MSE) (based on 1000 simulations) for the estimated coefficients of the ZADR II model assuming the coefficients of Example 1 (see Table 1) are the true coefficients.

|       | n=60 | n=120 | n=240 |
|-------|------|-------|-------|
|       | Constant | Slope | Constant | Slope | Constant | Slope |
| φ     | 0.059  | 0.008 | 0.027   | 0.004 | 0.014   | 0.002 |
| φ     | 0.131  | 0.018 | 0.062   | 0.009 | 0.032   | 0.004 |
| φ     | 0.125  | 0.020 | 0.063   | 0.010 | 0.028   | 0.004 |

|       | n=360 | n=480 | n=600 |
|-------|-------|-------|-------|
|       | Constant | Slope | Constant | Slope | Constant | Slope |
| φ     | 0.021  | 0.003 | 0.016   | 0.002 | 0.013   | 0.002 |
| φ     | 0.021  | 0.003 | 0.016   | 0.002 | 0.014   | 0.002 |
| φ     | 0.023  | 0.003 | 0.018   | 0.003 | 0.014   | 0.002 |
| φ     | 0.024  | 0.004 | 0.018   | 0.003 | 0.014   | 0.002 |

for this is the compositional nature of the data. Changing the value of one components results in change in all the other components. This approach was a simple one which seems to work adequately.

Appendix

The R codes to perform the ZADR type II parametrization are presented below. If anybody wants the relevant codes for ZADR type I, or any other question regarding the codes, or has better or faster codes, please send me an e-mail.

**Dirichlet regression type II**

At first, the function `dirireg2` is the one used in `optim` to maximize the log-likelihood of the Dirichlet regression type II [5].

```
dirireg2=function(param,z=z) {
  ## param contains the parameter values
  ## z contains the compositional data and independent variable(s)
  phi=param[1] ; para=param[-1]
  ## a small check against negative values of phi
  if (phi<0) l=10000
  if (phi>0) {
    ya=z$ya ; xa=z$xa
    ## ya is the compositional data and xa the independent variable(s)
    n=nrow(ya) ; d=ncol(ya)-1  ## sample size and dimensionality of the simplex
    be=matrix(para,ncol=d)  ## puts the beta parameters in a matrix
    mu1=cbind(1,exp(xa%*%be))
    ma=mu1/rowSums(mu1)  ## the fitted values
  }
  return(ma)
}
```

15
The next function requires `dirireg2` in order to perform the Dirichlet regression type II. We could of course have used the functions used in Maier's DirichletReg R package (Maier, 2014). But then this would require installation of the package and the goal was to write codes with as less dependence to any R package as possible.

diri.reg2=function(ya,xa) {
  ## ya is the compositional data
  ya=as.matrix(ya) ; n=nrow(ya)
  ya=ya/rowSums(ya) ; xa=as.matrix(cbind(1,xa))
  ## the line above makes sure ya is compositional data and
  ## then the unit vector is added to the design matrix
  d=ncol(ya)-1 ; z=list(ya=ya,xa=xa) ## dimensionality of the simplex
  rla=log(ya[, -1]/ya[,1]) ## additive log-ratio transformation
  ini=solve(t(xa)%*%xa)%*%t(xa)%*%rla ## initial values based on the logistic normal
  ## the next lines optimize the dirireg2 function and estimate the parameter values
  el=rep(0,2)
  qa=optim(c(20,as.vector(t(ini))),dirireg2,z=z,control=list(maxit=4000))
  el[1]=-qa$value
  qa=optim(qa$par,dirireg2,z=z,control=list(maxit=4000))
  el[2]=-qa$value
  vim=2
  while (el[vim]-el[vim-1]>0.0001) { ## the tolerance value can of course change
    vim=vim+1
    qa=optim(qa$par,dirireg2,z=z,control=list(maxit=4000))
    el[vim]=-qa$value
  }
  qa=optim(qa$par,dirireg2,z=z,control=list(maxit=4000),hessian=T)
  phi=qa$par[1] ; para=qa$par[-1] ## the estimated parameter values
  beta=matrix(para,ncol=d) ## the matrix of the betas
  colnames(beta)=colnames(ya[, -1]) ## names of the matrix of betas
  mu1=cbind(1,exp(xa%*%beta))
  ma=mu1/rowSums(mu1) ## fitted values
  s=sqrt(diag(solve(qa$hessian))) ## std of the estimated betas
  std.phi=s[1] ## std of the estimated phi
  S=matrix(s[-1],ncol=d) ## matrix of the std of the estimated betas
  colnames(S)=colnames(ya[, -1])
  V=solve(qa$hessian) ## covariance matrix of the parameters
  list(log.lik=-qa$value, param=ncol(xa)*d+1,phi=phi,std.phi=std.phi, beta=t(beta),std.errors=t(S),Cov=V,fitted=ma) 
}
Zeros values and Dirichlet regression type II

The next function $\text{mixreg2}$ is necessary to perform the ZADR type II since it will be used in $\text{optim}$ to maximize the corresponding log-likelihood [7].

```r
mixreg2=function(param,z=z) {
  ## separation of phi and the betas
  phi=param[1] ; para=param[-1]
  ## next a condition to avoid negative values of phi
  if (phi<0) f=10000
  if (phi>0) {
    y=z$y ; x=z$x ; D=ncol(y)
    ## y is the compositional data and x the independent variable(s)
    n=nrow(y) ; d=D-1
    ## n is the sample size and d is the dimensionality of the simplex
    ## next we separate the compositional vectors, those which contain
    ## zeros and those without. The same separation is performed for the
    ## independent variable(s)
    y1=y[rowSums(y>0)==D,]
    x1=x[rowSums(y>0)==D,]
    y3=y[rowSums(y>0)!==D,]
    x2=x[rowSums(y>0)!==D,]
    n1=nrow(y1) ; n2=n-n1
    ## n1 is the sample size of the compositional vectors with no zeros
    ## n2 is the sample size of the compositional vectors with zeros
    be=matrix(para,ncol=d)
    ## be is the matrix of the betas
    be=cbind(0,be)
    mu1=exp(x1%*%be)
    mu=mu1/rowSums(mu1) ## fitted values
    ###################
    ## next we find the fitted values for the compositional vectors with zeros
    zeros=rep(0,n2)
    if (n2>1) {
      for (l in 1:n2) {
        na=which(y3[,]==0)
        pa2=be[-na]
        mu2=exp(x2[,]%*%pa2)
        mu2=mu2/sum(mu2)
        zeros[l]=lgamma(phi)-sum(lgamma(phi*mu2))+sum((mu2*phi-1)*log(y3[,]-na))
      }
    } if (n2==1) {
      na=which(y3==0)
      pa2=be[-na]
      mu2=exp(x2%*%pa2)
      mu2=mu2/sum(mu2)
    }
  }
```
zeros = lgamma(phi) - sum(lgamma(phi*mu2)) + sum((mu2*phi-1)*log(y3[l,-na]))

## finally the log-likelihood is the sum of the non zeros part and the zeros part
f = -( n1*lgamma(phi) - sum(lgamma(phi*mu)) + sum((mu*phi-1)*log(y1)) + sum(zeros) )

The next function, `mix.reg2` performs the ZADR II.

```r
mix.reg2=function(y,x) {
  ## y is the compositional data
  ## x is the independent variable(s)
  D = ncol(y); d = D-1; x = as.matrix(cbind(1,x))
  ## d is the dimensionality of the simplex
  ## the unit vector is added in the design matrix
  colnames(x)[1]='constant'
  y = as.matrix(y); y = y/rowSums(y)
  ## makes sure y is compositional data
  ## next we separate the compositional vectors, those which contain
  ## zeros and those without. The same separation is performed for the
  ## independent variable(s)
  y1 = y[rowSums(y>0)==D,]
  x1 = x[rowSums(y>0)==D,]
  y0 = y[rowSums(y>0)!=D,]
  x0 = x[rowSums(y>0)!=D,]
  n2 = nrow(y0)  ## sample size of the zeros part
  mat = matrix(rep(1:D,n2),byrow=T,ncol=D)
  for (i in 1:n2) {
    for (j in 1:D) {
      if (y0[i,j]==0) mat[i,j]=0
    }
  }
  ame = diri.reg2(y1,x1[,,-1])  ## dirichlet regression on the non zeros part
  beta.ini = ame$beta  ## initial beta values
  dis = rep(0,n2)
  esa0 = cbind(1,exp(x0%*%t(beta.ini)))
  esa0 = esa0/rowSums(esa0)  ## fitted values
  esov = function(x,y) sqrt(sum( x*log(2*x/(x+y))+y*log(2*y/(x+y)),na.rm=T))
  for (i in 1:n2) dis[i]=esov(y0[i,mat[i,]],esa0[i,mat[i,]])  ## ES-OV
  test2 = sum(dis)  ## distance type diagnostic
  ##############
  ini.par = c(ame$phi,as.vector(t(beta.ini)))  ## initial parameter values
  ini.S = ame$std.errors
  z = list(y=y,x=x)
  el = NULL
  qa = optim(ini.par,mixreg2,z=z,control=list(maxit=5000))
  el[1] = -qa$value
  qa = optim(qa$par,mixreg2,z=z,control=list(maxit=5000))
  el[2] = -qa$value
  vim = 2
}
```

18
while (el[vim]-el[vim-1]>0.0001) { ## the tolerance value can of course change
        vim=vim+1
        qa=optim(qa$par,mixreg2,z=z,control=list(maxit=5000))
        el[vim]=qa$value
        qa=optim(qa$par,mixreg2,z=z,control=list(maxit=5000),hessian=T)
        phi=qa$par[1] ## final phi value
        Beta=matrix(qa$par[-1],ncol=d) ## final beta values
        s=sqrt(diag(solve(qa$hessian))) ## std errors of the betas
        phi.std=s[1] ## standard error of the phi value
        S=matrix(s[-1],ncol=d) ## matrix containing the standard errors of the betas
        ma=cbind(1,exp(x%*%Beta))
        mu=ma/rowSums(ma) ## fitted values
        colnames(mu)=colnames(y)
        colnames(Beta)=colnames(S)=colnames(y[,-1])
        rrownames(Beta)=rrownames(S)=colnames(ini.S)=colnames(beta.ini)=colnames(x)
        V=solve(qa$hessian) ## final covariance matrix of the betas
        U=ame$Cov ## initial covariance matrix of the betas
        ## below is the quadratic type diagnostic
        test1=( c(phi,as.vector(Beta))-c(ame$phi,as.vector(t(beta.ini))) )%*%solve(V+U)%*% ( c(phi,as.vector(Beta))-c(ame$phi,as.vector(t(beta.ini))) )
        test=c(test1,test2) ## values of the 2 tests
        names(test)=c('test1','test2')
        list(log.lik=-qa$value,param=ncol(x)*d+1,ini.beta=beta.ini,beta=t(Beta),ini.std.errors=ini.S,std.errors=t(S),ini.phi=ame$phi,phi=phi,ini.phi.std=ame$std.phi,phi.std=s[1],test=test,fitted=mu) }

**Bootstrap p-value of the diagnostics**

Finally, we have the code to do the two diagnostics, the quadratic type \[ (9) \] and the distance based diagnostic \[ (11) \].

diagnostic2=function(y,x,phi,Beta,test,R=999) {
        ## y is the compositional data
        ## x is the independent variable(s)
        ## phi and Beta are obvious
        ## test are the values of the two diagnostics
        ## R denotes the number of simulations
        Beta=t(Beta)
        y=as.matrix(y) ; y=y/rowSums(y)
        ## makes sure y is compositional data
        D=ncol(y) ; n=nrow(y) ; x=cbind(1,x)
        ## next split the y and x into zero free and non zero free parts
        y1=y[rowSums(y>0)==D,]
        x1=x[rowSums(y>0)==D,]
        y3=y[rowSums(y>0)!=D,]
        x2=x[rowSums(y>0)!==D,]
        }
n1=nrow(y1) ; n2=n-n1
Beta=cbind(0,Beta)
a1=exp(x1%*%Beta)
a1=a1/rowSums(a1)
a2=matrix(rep(0,n2*D),ncol=D)
mat=matrix(rep(1,n2*D),ncol=D)
for (i in 1:n2) {
  for (j in 1:D) {
    if (y3[i,j]==0) mat[i,j]=0 }
  }
if (n2>1) {
  for (i in 1:n2) {
    pa2=Beta[,y3[i,]>0]
    mu1=exp(x2[i,]%*%pa2)
    mu=mu1/rowSums(mu1) ## fitted values
    k=0
    for (j in 1:D) {
      if (mat[i,j]==1) {
        k=k+1
        a2[i,j]=mu[,k] }
    }
    if (n2==1) {
      pa2=Beta[,y3>0]
      mu1=exp(x2%*%pa2)
      mu=mu1/rowSums(mu1)
      k=0
      for (j in 1:D) {
        if (mat[i,j]==1) {
          k=k+1
          a2[,j]=mu[,k] }
      }
    }
  }
}
## the a3 below contains the non zero and then the zero parts
a3=rbind(a1,a2)
a4=phi*a3
## the x3 below contains the non zero and then the zero parts
x3=rbind(x1,x2)
tes1=tes2=rep(0,R)
esa=matrix(nrow=n,ncol=D)
for (b in 1:R) {
  ## for every observed row it generates one observation
  for (i in 1:n) esa[i,]=rdiric(1,a4[i,])
  ## then it performs the ZADR model on the simulated data
  palind=mix.reg2(esa,x3[,-1])
tes1[b]=palind$test[1]
tes2[b]=palind$test[2] }
## finally calculated the p-value of each diagnostic
p.val1=sum(tes1>test[1])/(R+1)
p.val2=sum(tes2>test[2])/(R+1)
## then it plots the values of of the diagnostics

```r
par(mfrow=c(1,2))
hist(tes1,xlab='Bootstrap test statistics',main=' ')
abline(v=test[1],col=2,lty=2)
hist(tes2,xlab='Bootstrap test statistics',main=' ')
abline(v=test[2],col=2,lty=2)
list(pvalue1=p.val1,pvalue2=p.val2) }
```

References

Aitchison, J. (1982). The statistical analysis of compositional data. *Journal of the Royal Statistical Society. Series B* 44(2), 139–177.

Aitchison, J. (2003). *The Statistical Analysis of Compositional Data*. London (UK): Chapman & Hall (Reprinted with additional material by The Blackburn Press).

Butler A. and 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.

Endres D.M. and Schindelin J.E. (2003) A new metric for probability distributions. *IEEE Transactions on Information Theory* 49(7), 1858–1860.

Gueorguieva R., Rosenheck, R. and Zelterman, D. (2008). Dirichlet component regression and its applications to psychiatric data. *Computational statistics & data analysis* 52(12), 5344–5355.

Gourieroux C., Monfort A. and Trognon A. (1984) Pseudo maximum likelihood methods: Theory. *Econometrica* 52(3), 681–700.

Hijazi R.H. and Jernigan R.W. (2009). Modelling compositional data using Dirichlet regression models. *Journal of Applied Probability and Statistics* 4(1), 77–91.

Kullback S. (1997) *Information theory and statistics*. New York: Dover Publications.

Larrosa, J.M. (2003) A compositional statistical analysis of capital stock. *Proceedings of the 1st Compositional Data Analysis Workshop, Girona, Spain*.

Leininger T.J., Gelfand A.E., Allen J.M. and Silander Jr J.A. (2013). Spatial Regression Modeling for Compositional Data With Many Zeros. *Journal of Agricultural, Biological, and Environmental Statistics* 18(3), 314–334.

Maier J.M. (2014). [DirichletReg: Dirichlet Regression in R](https://CRAN.R-project.org/package=DirichletReg).

Martín-Fernández J.A., Hron, K., Templ M., Filzmoser P. and Palarea-Albaladejo, J. (2012) Model-based replacement of rounded zeros in compositional data: Classical and robust approaches. *Computational Statistics & Data Analysis* 56(9), 2688–2704.

Murteira José M.R. and Ramalho J.J.S. (2013). Regression analysis of multivariate fractional data. *Econometric Reviews* (to appear), 183–196.

Ng K.W., Tian G.L. and Tang M.L. (2011). *Dirichlet and Related Distributions: Theory, Methods and Applications* Chichester: John Wiley & Sons.
Ospina R. and Ferrari S.L.P. (2010) Inflated beta distributions. *Statistical papers* 51(1), 111–126.

Österreicher F. and Vajda I. (2003) A new class of metric divergences on probability spaces and its applicability in statistics. *Annals of the Institute of Statistical Mathematics* 55(3), 639–653.

Palarea-Albaladejo J. and Martín-Fernández J.A. (2008) A modified EM algorithm for replacing rounded zeros in compositional data sets. *Computers & Geosciences* 34(8), 902–917.

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* 73(3), 351–375.

Stephens M.A. (1982) Use of the von Mises distribution to analyse continuous proportions. *Biometrika* 69(1), 197–203.

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

Theil H. (1967) *Economics and information theory* Amsterdam: North-Holland publishing company.

Yee T.W. (2011). [VGAM: Vector Generalized Linear and Additive Models. R package version 0.8-4](https://cran.r-project.org/package=VGAM)

Zadora G, Neocleous T. and Aitken C. (2010) A Two-Level Model for Evidence Evaluation in the Presence of Zeros. *Journal of forensic sciences* 55(2), 371–384.