South African Statistical Journal
2019, Vol. 53, No. 1, 55–64

A METHOD FOR BAYESIAN REGRESSION MODELLING OF COMPOSITION DATA

Sean van der Merwe
University of the Free State, Bloemfontein, South Africa
e-mail: *vandermerwes@ufs.ac.za*

Many scientific and industrial processes produce data that is best analysed as vectors of relative values, often called compositions or proportions. The Dirichlet distribution is a natural distribution to use for composition or proportion data. It has the advantage of a low number of parameters, making it the parsimonious choice in many cases. This paper considers the case where the outcome of a process is Dirichlet, dependent on one or more explanatory variables in a regression setting. The paper explores some existing approaches to this problem, and then introduces a new simulation approach to fitting such models, based on the Bayesian framework. The paper illustrates the advantages of the new approach through simulated examples and an application in sport science. These advantages include: increased accuracy of fit, increased power for inference, and the ability to introduce random effects without additional complexity in the analysis.

Key words: Bayes, Compositional data, Dirichlet distribution, Proportions, Regression, Simulation.

1. Introduction

When vectors are measured in whole numbers, say based on a classification process, analysis is often based on the multinomial distribution, with the total count being seen as a nuisance parameter. In cases where the total count is not relevant it seems more natural to work directly with the observed proportions. Sometimes the proportions themselves are observed directly, rather than counts and totals. Any situation where the quantity surveyed or analysed does not affect the expected vector (but may affect the precision) falls in this class of problem.

The Dirichlet distribution has been widely accepted in literature for modelling composition data, subject to the constraint that all the correlations between variables are negative (Maier, 2014). A wider class of distributions which allows for positive correlations defined on the same sample space is the logistic-normal (LN) (Aitchison and Shen, 1980). However, the LN distribution has many parameters to estimate due to the unknown covariance parameter matrix. In contrast, the Dirichlet distribution has only P unknown parameters to estimate for P composition components. Further, when the model is expanded to the regression framework then the requirement of negative correlations no longer applies (Maier, 2014).

The only remaining restriction of relevance is that observed values should be strictly positive. The occasional zero will not affect results, but the method cannot handle binary data in the dependent variable vectors. No restrictions apply to the explanatory variables — they may have any structure.

MSC2010 subject classifications. 62J02, 62F15.

As examples of applications, consider the work of Boukal, Ditrich, Kutcherov, Sroka, Dudová and Papáček (2015), who look at the development of insects, or the work of Espin-Garcia, Shen, Qiu, Brhane, Liu and Xu (2014), who look at genetic analysis problems, or the work of Smithson and Verkuilen (2006), who discuss applications in psychology. For an industrial application see de Waal, Coetzer and van der Merwe (2016) where the composition of coal is analysed. Another area in which Dirichlet regression may be useful is in politics, where only the proportion of voters supporting a set of candidates is of interest, possibly dependent on district or demographics. In ecology, preferences of animals for specific types of prey are of interest. To generalise, any situation where people choose between a set of options, and the researcher is interested in the choice and not the number of people making a choice, then the Dirichlet regression model may be of use.

The paper is outlined as follows: In Section 2, a brief introduction on applicable Dirichlet properties is given. After that Section 3 considers existing approaches to the problem of regression with dependent variables that follow the Dirichlet distribution (conditional on explanatory data). The new methodology is then introduced in Section 4, and its usefulness is illustrated via simulated examples in Section 5. As an example based on observed data, movement data arising from a school netball tournament is analysed (Section 6). Conclusions and future work are discussed in Section 7.

2. The Dirichlet distribution

If $\mathbf{Y}$ is distributed Dirichlet$(\alpha_1, \ldots, \alpha_P)$, denoted by $D(\boldsymbol{\alpha})$, then the joint density is given by

$$f(\mathbf{y}) = \frac{\prod_{i=1}^{P} \Gamma(\alpha_i)}{\Gamma(\alpha_0)} \left\{ \prod_{i=1}^{P} y_i^{\alpha_i - 1} \right\}, \quad 0 < y_i < 1, \quad \sum_{i=1}^{P} y_i = 1, \quad \alpha_0 = \sum_{i=1}^{P} \alpha_i.$$

Aitchison (1986) calls this distribution the compositional Dirichlet defined on the specified simplex.

Wilks (1962) and de Groot (1970) provide detailed discussions on many of the properties of the Dirichlet distribution. Some relevant properties follow.

1. The means and covariances are

$$E[Y_i] = \mu_i = \frac{\alpha_i}{\alpha_0}, \quad i = 1, \ldots, P,$$
$$\sigma_{ij} = \begin{cases} \dfrac{-\alpha_i \alpha_j}{\alpha_0^2(1+\alpha_0)} & \text{if } i \neq j, \\ \dfrac{\alpha_i(\alpha_0 - \alpha_i)}{\alpha_0^2(1+\alpha_0)} & \text{if } i = j, \end{cases} \tag{1}$$

   where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_P)$ denotes the mean of the distribution and $\Sigma = (\sigma_{ij})$, $i, j = 1, \ldots, P$, the covariance matrix.

2. The marginal distribution of any subset $\mathbf{W}$ of $\mathbf{Y}$ is again Dirichlet$(\alpha_{\mathbf{W}}, \alpha_0 - \sum_{i \in \mathbf{W}} \alpha_i)$. See Aitchison (1986).

3. In (1), if $\boldsymbol{\alpha}$ is a multiple of $\boldsymbol{\beta}$ for two Dirichlet distributions $D(\boldsymbol{\alpha})$ and $D(\boldsymbol{\beta})$, the means are the same, but the covariance matrices differ.

4. Fitting the Dirichlet distribution to a sample of reasonable size can be done by following the method of moments or method of maximum likelihood (ML) as described in Minka (2012), or by using the Bayesian approach (Van der Merwe and de Waal, 2013). Note that the ML approach is iterative, while the Bayesian approach uses simulation. Both methods are fast in most cases due to the neat convex nature of the likelihood.

3. Dirichlet regression

In Carmargo, Stern and Lauretto (2012) the problem is defined as follows:

Let $Y = (\mathbf{y}_{1\cdot}; \mathbf{y}_{2\cdot}; \ldots; \mathbf{y}_{n\cdot})$ be a sample of vector observations. They use the $i\cdot$ notation to indicate that the vectors are arranged in rows of the matrix Y, for practical convenience. Let $X = (\mathbf{x}_{1\cdot}; \mathbf{x}_{2\cdot}; \ldots; \mathbf{x}_{n\cdot})$ be Q explanatory variables arranged the same way (each column of X is a component and each row of X corresponding with the same row of Y). Recall that $\sum_{j=1}^{P} y_{ij} = 1,\ y_{ij} > 0$; while the values of X could be any real numbers. Their notation is used going forward.

Based on the work of Campbell and Mosimann (1987), Hijazi and Jernigan (2009) and Carmargo et al. (2012) each parameter is modelled as a function of the explanatory variables. In terms of a single observation $\mathbf{y}_{i\cdot} = (y_{i1}, \ldots, y_{iP}) \sim D(\alpha_{i1}, \ldots, \alpha_{iP})$,

$$\alpha_{ij} = x_{i1}\beta_{1j} + \cdots + x_{iQ}\beta_{Qj} = \mathbf{x}_{i\cdot}\boldsymbol{\beta}_{\cdot j}.$$

Thus, the parameters to be estimated are all the $\beta_{kj},\ k = 1, \ldots, Q, j = 1, \ldots, P$, subject to the constraint $\alpha_{ij} > 0\ \forall\ i = 1 \ldots n, j = 1, \ldots, P$. They describe a custom optimisation procedure to estimate these parameters under these constraints. Finally, Carmargo et al. (2012) explain an approach to testing $\beta_{kj} = 0$, which is useful in many problems.

Gueorguieva, Rosenheck and Zelterman (2008) propose using a log-link in each dimension, thus eliminating constraints in the optimisation procedure.

However, each β_{kj} does not have a clean interpretation in the above models as $E[Y_{ij}]$ is a function of all β_{kj}. The difficulty of interpretation is seen as a major drawback by many researchers, and led to the investigation of alternatives.

Maier (2014) applies a multivariate transformation to the parameters of the Dirichlet distribution, arriving at an alternative formulation that has the advantage of modelling the expected value of an observation separately from its precision. He begins by defining a parameter $\phi = \alpha_0$ to denote the precision of an observation. Looking at (1) it can be seen that ϕ is not exactly the precision but acts like the precision in the sense that, for large values, if the value of ϕ doubles while the mean vector is unchanged then the variance halves. Note the relationship $\boldsymbol{\alpha} = \boldsymbol{\mu}\phi$. Maier (2014) applies a log-link function to model ϕ, i.e. $\ln \phi_i = \mathbf{w}_{i\cdot}\boldsymbol{\beta}_{\cdot\phi}$, where W is a matrix of explanatory variables for the precision.

For the purpose of modelling the mean, Maier (2014) uses a multivariate logit link. This involves choosing a base category and setting all coefficients to zero for this category, then using a log-link to model the other categories and rescaling the results so that the means sum to one. In the notation

defined previously, and using dimension 1 as the base, for every observation $i = 1, \ldots, n$,

$$\mu_{ij} = \frac{\exp(\mathbf{x}_{i\cdot}\boldsymbol{\beta}_{\cdot j})}{\sum_{k=1}^{P} \exp(\mathbf{x}_{i\cdot}\boldsymbol{\beta}_{\cdot k})}, \quad j = 2 \ldots P,$$
$$\mu_{i1} = \frac{\exp \mathbf{x}_{i\cdot}\mathbf{0}}{\sum_{k=1}^{P} \exp \mathbf{x}_{i\cdot}\boldsymbol{\beta}_{\cdot k}} = \frac{1}{1 + \sum_{k=2}^{P} \exp \mathbf{x}_{i\cdot}\boldsymbol{\beta}_{\cdot k}}.$$

Maier (2014) explains that using the transformation above results in coefficients that are interpretable as odds ratios if exponentiated.

However, each μ_{ij} is still a function of all β_{kj}, and there are no coefficients for the base dimension. These limitations inhibit interpretation. The next section introduces a new methodology that incorporates the best aspects of the approaches described above.

4. New methodology

The first change is the use of a univariate logit transformation for each mean parameter individually, thus allowing all β_{kj} coefficients to be unrestricted real numbers. The second change that ties in with the first is that this new methodology abandons the idea of a reference category or dimension.

In theory, one dimension is redundant since it is a linear combination of the others, but in practice it is of interest to know the relationship between the explanatory variables and the outcomes in all dimensions. Often the dimensions are equal in the view of the researcher and it is not sensible to relegate one to reference status. It is for this reason that researchers such as Chen and Li (2016) resort to modelling each dimension individually as beta distributed random variables, but that in turn ignores the multivariate nature of the data.

It is desirable to have each β_{kj} relate directly to a single dimension in a way that can be directly interpreted. By modelling all dimensions the researcher moves closer to this ideal.

The third change is a move to the Bayesian framework. This involves introducing vague normal priors on all β parameters. All other parameters are defined in terms of these, so no further priors are necessary at this stage.

Specifically, this method uses the Bayesian simulation framework, which holds many advantages. It allows us to directly quantify uncertainty in both the coefficients and the means. Also, when moving to a predictive framework, construction of predictive densities is relatively straightforward.

The method is still bound by the conditions

$$\sum_{j=1}^{P} \mu_{ij} = 1, \quad \forall\, i = 1, \ldots, n, \tag{2}$$

which impede the standard simulation approach greatly. In order to have the simulation process run smoothly, the researcher must introduce a source of flexibility into the model. This paper adds flexibility by replacing the restriction (2) by a penalty on the likelihood:

$$L^* \propto L \times \exp\left\{ -\frac{1}{\xi} \sum_{i=1}^{n} \left[\left(\sum_{j=1}^{P} \mu_{ij} \right) - 1 \right]^2 \right\}.$$

The hyperparameter ξ must be chosen large enough to allow the simulation procedure to run smoothly, but small enough to have minimal impact on the simulation results. Here, minimal impact implies

mean deviations that can easily be corrected. The valid region for ξ to meet these criteria seems surprisingly large. One may consider ξ as a hyperparameter and choose its value manually, or, more conveniently, apply a prior distribution to ξ and have the value vary as part of the simulation process.

Then, for further simulation flexibility a second penalty parameter (ξ^*) is introduced in the relationship between $\boldsymbol{\alpha}$, $\boldsymbol{\mu}$ and ϕ. For both penalty parameters a simple exponential prior works well in all scenarios tested.

Assuming explanatory data captured in matrices X and W (which may overlap), the model is defined in a hierarchical fashion:

$$\begin{aligned}
\mathbf{y}_{i\cdot}|\mathbf{x}_{i\cdot}, \mathbf{w}_{i\cdot} &\sim D(\boldsymbol{\alpha}_{i\cdot}),\\
\ln \alpha_{ij} &\sim N\left(\ln \mu_{ij} + \ln \phi_i\ ,\ 1/\xi^*\right),\\
\ln \phi_i &= \mathbf{w}_{i\cdot}\boldsymbol{\beta}_{\cdot\phi},\\
\operatorname{logit}(\mu_{ij}) &= \mathbf{x}_{i\cdot}\boldsymbol{\beta}_{\cdot j},\\
\sum_{j=1}^{P} \mu_{ij} &\sim N\left(1, \frac{1}{\xi}\right),\\
\beta_{ij}\ ,\ \beta_{i\phi} &\sim N(0, 10000),\\
\xi^* &\sim Exp(\mu = 100/P),\\
\xi &\sim Exp(\mu = 1000/P).
\end{aligned}$$

Note that in the expressions above, the models for the mean ($\operatorname{logit}(\mu_{ij}) = \mathbf{x}_{i\cdot}\boldsymbol{\beta}_{\cdot j}$) and precision ($\ln \phi_i = \mathbf{w}_{i\cdot}\boldsymbol{\beta}_{\cdot\phi}$) are in linear form for ease of understanding only — these models can be extended as needed by the researcher.

Looking only at the likelihood it might be said that the model is over-parameterised or lacking identifiability; however, the model is identifiable in the Bayesian sense (Rannala, 2002). All parameters are given proper priors, ensuring a proper posterior, and use of informative priors on the penalty parameters ensures that the posterior distributions of the parameters of interest reflect the information in the data. In a practical sense the identifiability of the model is improved over previous models as there is now exactly one coefficient joining each component of the explanatory data to each component of the observation vectors.

This paper implements the model using Gibbs sampling (Gelfand and Smith, 1990) via the OpenBUGS program (http://www.openbugs.net/). Implementation is done indirectly through the R2OpenBUGS package (Sturtz, Ligges and Gelman, 2005) for R (R Core Team, 2018). All pre- and post-calculations are done in R. The MASS and parallel packages supplied with R were also used to facilitate calculations.

Post-simulation, the following corrections are applied to each simulated parameter set k individually to ensure that fitted expected values sum to one for each observation:

$$\begin{aligned}
\mu_{ijk}^{adj} &= \frac{\mu_{ijk}^{sim}}{\sum_{j=1}^{P} \mu_{ijk}^{sim}},\\
\alpha_{ijk}^{adj} &= \mu_{ijk}^{adj}\phi_{ik}^{sim}.
\end{aligned}$$

Table 1. Fit statistics for Scenario A.

Scenario A	Target	Maier approach	New approach
Error	0.00	19.59	18.38
Coverage	0.95	0.87	0.94
Std. Width	0.00	0.70	0.75

5. Simulation study

Since the different methods discussed in the previous sections use different transformations (identity, log, multivariate logit, and univariate logit), the methods are compared on a single scale. In general the researcher fitting Dirichlet regression models is interested in three things: the significance of the coefficients, the direction of any significant relationships, and the accuracy of the model fit on the observed data. The first point of focus is model accuracy.

As a measurement of error, consider the average *sum of composition errors* (SCE), explained by Hijazi and Jernigan (2009). It is the sum of the Aitchison distances (Aitchison, 1986) between estimated compositions and the target values.

Further, it is sensible to calculate the intervals for each expected value individually and then report the average coverage, along with the average width (standardised by dividing by the expected values).

Datasets are constructed from a model exactly in line with the 'alternative' specification of Maier (2014). Models are then fitted using his DirichletReg package, as well as the new methodology. Models are correctly specified in all cases — model misspecification is beyond the scope of this work.

5.1 Simulation Scenario A

Scenario A is a simple analogy to the MANOVA problem. Consider a single explanatory variable that is a factor with three levels. A researcher might be interested in whether the mean vectors differ between the three groups created by the factor levels, under the assumption of constant variance.

Let the observed vectors have three dimensions. Set the coefficients for the first dimension to zero, and then use coefficients of $(-0.9, 0.6, 1.2)$ for the second dimension and $(0.8, -1, 0.5)$ for the third. The inverse multivariate logit transformation is then applied to obtain the 'true' expected values for every observation, and expanded to 20 observations per factor level ($n = 60$ in total). The first step is creating a matrix of binary variables (X) from an expansion of the explanatory factor. The transformation (explained in detail in Maier, 2014) involves multiplying X by each set of coefficients, and then exponentiating to obtain raw expected values, which are then standardised to sum to one for each observation.

The next step is to multiply by a chosen value for ϕ. This paper uses $\phi = 1$ for illustration. The effect of changing this value will be discussed after the results. Multiplying the expected values by ϕ yields a matrix of α_{ij} values, which is used to simulate hundreds of Dirichlet samples in the standard way.

After every sample is modelled and the results summarised, various statistics are produced. See Table 1 for the most important values. It is clear that the new methodology a produces better fit.

Table 2. Fit statistics for Scenario B.

Scenario B	Target	Maier approach	New approach
Error	0.00	19.19	18.81
Coverage	0.95	0.85	0.86
Std. Width	0.00	0.52	0.54

The next question of interest would be the effect of varying the chosen parameter values. It appears that the important parameter is the underlying precision, which is closely related to the ϕ parameter. When the data are measured with high precision (say $\phi \geq 5$) then there is little difference between the fits obtained by the two methods (both methods fit very well). As the precision drops and the underlying relationships become more obfuscated then the accuracy of the previous methodology falls away rapidly, while the new methodology loses accuracy slowly, resulting in the difference observed in Table 1.

Another question of interest would be the effect of increasing the dimensionality of the problem. If the dimension is increased to 8 and the coefficients are chosen as random $U(-1.5, 1.5)$ values, then the relative differences between the methods become even more exaggerated. The new methodology adapts easily to having many categories in the dependent variable.

5.2 Simulation Scenario B

Scenario B is a more complex scenario where a linear term is introduced in every mean vector as well as the precision model, in addition to the factor explained in Section 5.1. The goal is to determine whether the model can identify both relationships simultaneously in all categories.

The explanatory factor (X_1) is given two levels with coefficients $(0, 0)$, $(-0.9, 0.6)$, $(1.8, -1)$ in three dimensions respectively. The explanatory variable with a linear relationship (X_2) is given real values between 4.5 and 7.5 with 40 values per level of the explanatory factor. The linear relationship is created by adding $0.75X_2$ to the third dimension and then correcting the means to add to one, thus creating a positive relationship in the third dimension and implicitly creating negative relationships in the first two dimensions. As X_2 increases, y_1 and y_2 will tend to decrease, while y_3 will tend to increase. The expression for $\log \phi$ used to generate the data is $-1 + 0.5X_2$.

The results are averaged from hundreds of samples and summarised. As indicated in Table 2, the results from the methods are similar, but the new methodology is more accurate.

Concerning inference, the new method shows a major improvement over the existing method. In the precision model the median p-value over simulated samples for the existing method is 0.1%, while for the new method it is approximately 0%. In the second and third dimensions the median p-values for the existing method are 50% and 24%, while the new method yields 1% and 0.1%. The new method also reports a p-value for the first dimension (median 0.4%), while the existing method does not. Thus, for almost all simulations done, the new method correctly identifies the direction of the linear relationships and marks all of them as significant. The existing method only identifies the linear relationship in the precision model.

6. Observed data example from sport science

During a school netball tournament, scholars were tracked accurately as they move across the field. One of the resulting measurements was the proportion of time spent standing/walking/running during the course of a match. The goal is to investigate the relationship between these measurements and the playing position.

A major complicating factor is the fitness and behaviour variation between players. Some players were observed for only one match, while others were observed for up to nine matches. This suggests an unbalanced mixed effects model, with position as fixed effect and player as random effect.

The model is defined explicitly as follows:

$$\begin{aligned}
\mathbf{y}_{i\cdot}|x_i, w_i &\sim D(\alpha_{i1}, \alpha_{i2}, \alpha_{i3}),\\
x_i &= \text{position for observation } i,\\
w_i &= \text{player for observation } i,\\
\ln \alpha_{ij} &\sim N\left(\ln \mu_{ij} + \ln \phi_{x_i}, 1/\xi^*\right),\\
\text{logit}(\mu_{ij}) &= \beta_{x_i j} + \beta_{w_i j},\\
\sum_{j=1}^{3} \mu_{ij} &\sim N\left(1, 1/\xi\right),
\end{aligned}$$

with priors

$$\begin{aligned}
\ln \phi_{x_i} &\sim N(0, 10000),\\
\beta_{x_i j} &\sim N(0, 10000),\\
\beta_{w_i j} &\sim N(0, \tau_j^{-1}),\\
\tau_j &\sim Gamma(0.001, 0.001),\\
\xi^* &\sim Exp(\mu = 100/3),\\
\xi &\sim Exp(\mu = 1000/3).
\end{aligned}$$

For additional clarity, where there is a second subscript or the letter j, these refer to the category: standing=1, walking=2, running=3. The model as implemented has exactly one precision parameter for each playing position, but three intercepts for each playing position corresponding to the three categories. The model has three random effects for each player, corresponding to the three categories. These random effects have a single common variance within each category (three in total).

The results are summarised in terms of posterior means and 95% credibility intervals for each position, by category. Figure 1 shows this result with the positions sorted from highest to lowest expected running proportion. This sorting makes it easier to identify differences between positions. The information in this figure may be of use to netball coaches going forward.

Code with detailed comments for this example is available at
http://seanvdm.co.za/files/ProportionsExample20180704.pdf, along with its accompanying dataset (anonymised): http://seanvdm.co.za/files/NetballData.csv.

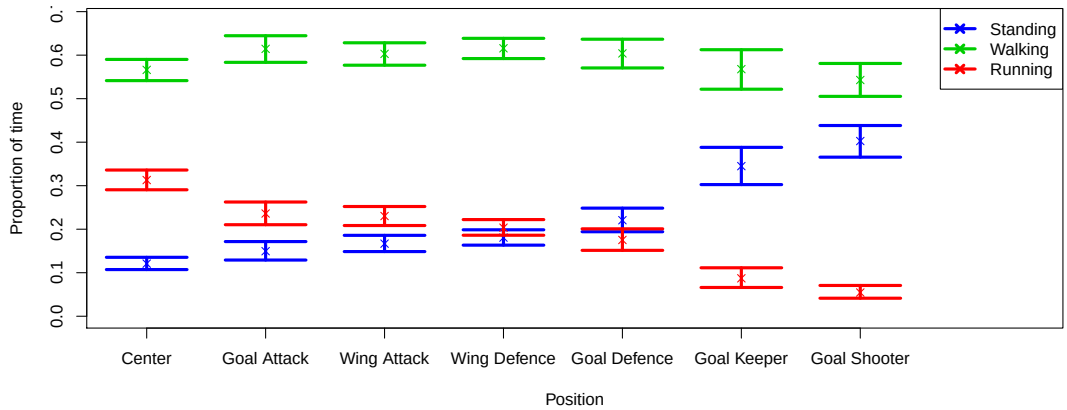

Figure 1. Differences in movement between playing positions in a netball tournament as suggested by Bayesian unbalanced mixed effects model.

7. Conclusions

In this paper the goal was a generic solution to the problem where the outcome of a process is Dirichlet, dependent on one or more explanatory variables in a regression setting. Existing approaches were discussed and a new methodology introduced. The new methodology was directly compared to the latest of the existing approaches and found to perform well. At worst the performance is in line with existing tools, but in many cases the improvement is remarkable, especially when the data has high variance. Advantages of the new methodology were discussed, including ease of interpretation and prediction, with accurate intervals, as well as the ability to introduce random effects.

Acknowledgements

The netball data is from a study by Michael Shaw and Derik Coetzee of the Department of Exercise and Sports Science, University of the Free State, South Africa. The author wishes to thank Robert Schall and Michael von Maltitz for useful advice during the development of this work.

References

Aitchison, J. (1986). *The Statistical Analysis of Compositional Data*. Springer Netherlands. doi: 10.1007/978-94-009-4109-0.

Aitchison, J. and Shen, S. M. (1980). Logistic-normal distributions: some properties and uses. *Biometrika*, **60**, 67, 261–272.

Boukal, D. S., Ditrich, T., Kutcherov, D., Sroka, P., Dudová, P., and Papáček, M. (2015). Analyses of developmental rate isomorphy in ectotherms: Introducing the Dirichlet regression. *PLOS one*, **10**, e0129341.

Campbell, G. and Mosimann, J. E. (1987). Multivariate methods for proportional shape. *In ASA Proceedings of the Section on Statistical Graphics*, volume 1. 10–17.

Carmargo, A. P., Stern, J. M., and Lauretto, M. S. (2012). Estimation and model selection in

Dirichlet regression. *Conference proceedings*, **1143**, 206–213.

Chen, E. Z. and Li, H. (2016). A two-part mixed-effects model for analyzing longitudinal microbiome compositional data. *Bioinformatics*, **32**, 2611–2617.

de Groot, M. H. (1970). *Optimal Statistical Decisions*. McGraw-Hill.

de Waal, D. J., Coetzer, R., and van der Merwe, S. (2016). Classification of multiple Dirichlet observations under a multinomial model. *Chemometrics and Intelligent Laboratory Systems*, **150**, 51–57.

Espin-Garcia, O., Shen, X., Qiu, X., Brhane, Y., Liu, G., and Xu, W. (2014). Genetic association analysis for common variants in the genetic analysis workshop 18 data: a Dirichlet regression approach. *BMC Proceedings*, **8**, S70.

Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, **85**, 398–409.

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

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

Maier, M. J. (2014). Dirichletreg: Dirichlet regression for compositional data in r.
URL: *http://epub.wu.ac.at/4077/*

Minka, T. P. (2012). Estimating a Dirichlet distribution.
URL: *https://tminka.github.io/papers/dirichlet/*

R Core Team (2018). R: A language and environment for statistical computing.
URL: *http://www.R-project.org/*

Rannala, B. (2002). Identifiability of parameters in MCMC Bayesian inference of phylogeny. *Systematic Biology*, **51**, 754–760. doi:10.1080/10635150290102429.

Smithson, M. and Verkuilen, J. (2006). A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychological Methods*, **11**, 54.

Sturtz, S., Ligges, U., and Gelman, A. (2005). R2WinBUGS: a package for running WinBUGS from R. *Journal of Statistical Software*, **12**, 1–16.

Van der Merwe, S. and de Waal, D. J. (2013). Bayesian fitting of Dirichlet type I and II distributions. Technical report, University of the Free State.

Wilks, S. S. (1962). *Mathematical Statistics*. John Wiley & Sons Inc. NY.

Manuscript received 2017-12-14, revised 2018-09-07, accepted 2018-09-17.